

## ARTICLE OPEN



## CHRONIC LYMPHOCYTIC LEUKEMIA

## Different prognostic impact of recurrent gene mutations in chronic lymphocytic leukemia depending on IGHV gene somatic hypermutation status: a study by ERIC in HARMONY

Larry Mansouri<sup>1,42</sup>, Birna Thorvaldsdottir<sup>1,42</sup>, Lesley-Ann Sutton<sup>1,42</sup>, Georgios Karakatsoulis<sup>2,3</sup>, Manja Meggendorfer<sup>4</sup>, Helen Parker<sup>5</sup>, Ferran Nadeu<sup>6,7</sup>, Christian Brieghel<sup>8</sup>, Stamatia Laidou<sup>2</sup>, Riccardo Moia<sup>9</sup>, Davide Rossi<sup>10,11</sup>, Mark Catherwood<sup>12</sup>, Jana Kotaskova<sup>13,14,15</sup>, Julio Delgado<sup>6,7,16,17</sup>, Ana E. Rodríguez-Vicente<sup>18,19,20</sup>, Rocío Benito<sup>18,19,20</sup>, Gian Matteo Rigolin<sup>21</sup>, Silvia Bonfiglio<sup>22</sup>, Lydia Scarfo<sup>22</sup>, Mattias Mattsson<sup>23</sup>, Zadie Davis<sup>24</sup>, Ajay Gogia<sup>25</sup>, Lata Rani<sup>25</sup>, Panagiotis Baliakas<sup>23</sup>, Hassan Foroughi-Asl<sup>1</sup>, Cecilia Jylhä<sup>1</sup>, Aron Skafason<sup>1</sup>, Inmaculada Rapado<sup>26,27</sup>, Fatima Miras<sup>26</sup>, Joaquín Martínez-López<sup>26,27</sup>, Javier de la Serna<sup>26,27</sup>, Jesús María Hernández Rivas<sup>18,19,20</sup>, Patrick Thornton<sup>28</sup>, María José Larráyo<sup>29,30</sup>, María José Calasanz<sup>29,30</sup>, Viktória Fésüs<sup>31</sup>, Zoltán Mátrai<sup>32</sup>, Csaba Bödör<sup>31</sup>, Karin E. Smedby<sup>33</sup>, Blanca Espinet<sup>34</sup>, Anna Puiggros<sup>34</sup>, Ritu Gupta<sup>25</sup>, Lars Bullinger<sup>35</sup>, Francesc Bosch<sup>36</sup>, Bárbara Tazón-Vega<sup>36</sup>, Fanny Baran-Marszak<sup>37</sup>, David Oscier<sup>24</sup>, Florence Nguyen-Khac<sup>38</sup>, Thorsten Zenz<sup>39</sup>, Maria Jose Terol<sup>40</sup>, Antonio Cuneo<sup>21</sup>, María Hernández-Sánchez<sup>18,19,20</sup>, Sarka Pospisilova<sup>13,14,15</sup>, Ken Mills<sup>12</sup>, Gianluca Gaidano<sup>9</sup>, Carsten U. Niemann<sup>8</sup>, Elias Campo<sup>6,7,16,17</sup>, Jonathan C. Strefford<sup>5</sup>, Paolo Ghia<sup>22,43</sup>, Kostas Stamatopoulos<sup>1,2,43</sup> and Richard Rosenquist<sup>1,41,43</sup>

© The Author(s) 2022, corrected publication 2023

Recent evidence suggests that the prognostic impact of gene mutations in patients with chronic lymphocytic leukemia (CLL) may differ depending on the immunoglobulin heavy variable (IGHV) gene somatic hypermutation (SHM) status. In this study, we assessed the impact of nine recurrently mutated genes (*BIRC3*, *EGR2*, *MYD88*, *NFKBIE*, *NOTCH1*, *POT1*, *SF3B1*, *TP53*, and *XPO1*) in pre-treatment samples from 4580 patients with CLL, using time-to-first-treatment (TTFT) as the primary end-point in relation to IGHV gene SHM status. Mutations were detected in 1588 (34.7%) patients at frequencies ranging from 2.3–9.8% with mutations in *NOTCH1* being the most frequent. In both univariate and multivariate analyses, mutations in all genes except *MYD88* were associated with a significantly shorter TTFT. In multivariate analysis of Binet stage A patients, performed separately for IGHV-mutated (M-CLL) and unmutated CLL (U-CLL), a different spectrum of gene alterations independently predicted short TTFT within the two subgroups. While *SF3B1* and *XPO1* mutations were independent prognostic variables in both U-CLL and M-CLL, *TP53*, *BIRC3* and *EGR2* aberrations were significant predictors only in U-CLL, and *NOTCH1* and *NFKBIE* only in M-CLL. Our findings underscore the need for a compartmentalized approach to identify high-risk patients, particularly among M-CLL patients, with potential implications for stratified management.

*Leukemia* (2023) 37:339–347; <https://doi.org/10.1038/s41375-022-01802-y>

## INTRODUCTION

Chronic lymphocytic leukemia (CLL) is clinically heterogeneous, ranging from an indolent condition without urgent need for treatment, to an aggressive disease characterized by rapid progression, resistance to therapy, and poor overall survival. The Rai and Binet staging systems are used in clinical practice to assess prognosis in patients with CLL [1, 2]. However, in early-stage disease, which constitute the great majority of new diagnoses, these systems are unable to predict which cases will progress to a more aggressive disease [3]. That said, there are a number of molecular markers with prognostic and/or predictive impact that should be assessed in all patients prior to treatment initiation

[4, 5]. These include genomic aberrations detected by fluorescence in situ hybridization (FISH), such as del(17p), del(11q), trisomy 12 and del(13q), and the immunoglobulin heavy variable (IGHV) gene somatic hypermutation (SHM) status [6–9]. In addition, *TP53* sequence analysis is essential, as patients with *TP53* mutations, irrespective of the presence of del(17p), experience a more adverse outcome, even with targeted therapies, particularly in the relapsed/refractory setting [5, 10–12].

Over the past decade, next-generation sequencing (NGS) studies have led to the discovery of recurrently mutated genes in CLL, such as *NOTCH1*, *SF3B1*, *BIRC3*, *XPO1*, *POT1*, *NFKBIE* and *EGR2*, that are associated with poor clinical outcome [13–21]. On

A full list of author affiliations appears at the end of the paper.

Received: 11 October 2022 Revised: 8 December 2022 Accepted: 15 December 2022  
Published online: 24 December 2022

**Table 1.** Clinicobiological characteristics of the studied cohort.

	All cases	Binet A	Binet A M-CLL	Binet A U-CLL
Gender				
Male	2890 (63%)	1992 (59%)	1183 (58%)	708 (61%)
Female	1690 (37%)	1370 (41%)	862 (42%)	445 (39%)
Median age at diagnosis (years)	64.6	64.9	64.6	64.9
IGHV gene SHM status				
M-CLL	2454 (57%)	2045 (64%)		
U-CLL	1878 (43%)	1153 (36%)		
Unknown	239	164		
Recurrent aberrations <sup>a</sup>				
del(13q)	1856 (42%)	1465 (45%)	775 (57%)	302 (27%)
trisomy12	566 (13%)	380 (12%)	72 (5%)	212 (19%)
del(11q)	495 (11%)	277 (8%)	12 (1%)	225 (20%)
del(17p)	236 (5%)	148 (5%)	42 (3%)	78 (7%)
No recurrent abnormalities	1306 (29%)	1005 (31%)	469 (34%)	319 (28%)
Unknown	121	87	35	17
Binet stage				
A	3362 (74%)			
B	825 (18%)			
C	387 (8%)			
Unknown	6			
Treatment status <sup>b</sup>				
Treated	2680 (59%)	1570 (47%)	640 (32%)	875 (76%)
Untreated	1900 (41%)	1792 (53%)	1405 (68%)	278 (24%)

SHM somatic hypermutation, M-CLL CLL with mutated IGHV genes, U-CLL CLL with unmutated IGHV genes.

<sup>a</sup>According to the Döhner classification [6].

<sup>b</sup>At follow-up.

the other hand, recurrent mutations in *MYD88* have been correlated with a favorable outcome, though not conclusively [22, 23]. Based on these findings, several studies have devised prognostic models that incorporate clinicobiological factors as well as various combinations of recurrent gene mutations, aimed at improving risk stratification in CLL [24–26]. However, several questions remain unanswered. While *NOTCH1*, *SF3B1* and *BIRC3* mutations have been investigated in larger patient cohorts [17, 25, 27, 28], the prognostic role of the remaining gene mutations is less well studied. Furthermore, the frequencies and clinical impact of gene mutations differ between the poor-prognostic IGHV-unmutated (U-CLL) subgroup and the more favorable-prognostic IGHV-mutated CLL (M-CLL) subgroup [29–32]. In fact, in recent studies, a number of genetic markers were demonstrated to affect outcome in U-CLL and M-CLL differently, indicating that a more compartmentalized approach based on the IGHV gene SHM status might be necessary when developing prognostic models [31, 32].

In this study, we investigated the prognostic roles of nine genes recurrently mutated in CLL, as well as their relative impact in U-CLL and M-CLL, with a particular focus on early-stage patients and time-to-first-treatment (TTFT) as the primary clinical endpoint. This analysis was performed in a well-annotated series of pre-treatment samples from 4580 patients with CLL, to our knowledge the largest series analyzed thus far, consolidated in the context of a European multicenter effort coordinated by the European Research Initiative on CLL (ERIC) as partner of the HARMONY Alliance. Our novel data highlight that different spectra of genetic mutations predict clinical outcome in U-CLL and M-CLL, respectively, with *SF3B1* and *XPO1* mutations being the most significant prognostic markers irrespective of subgroup assignment.

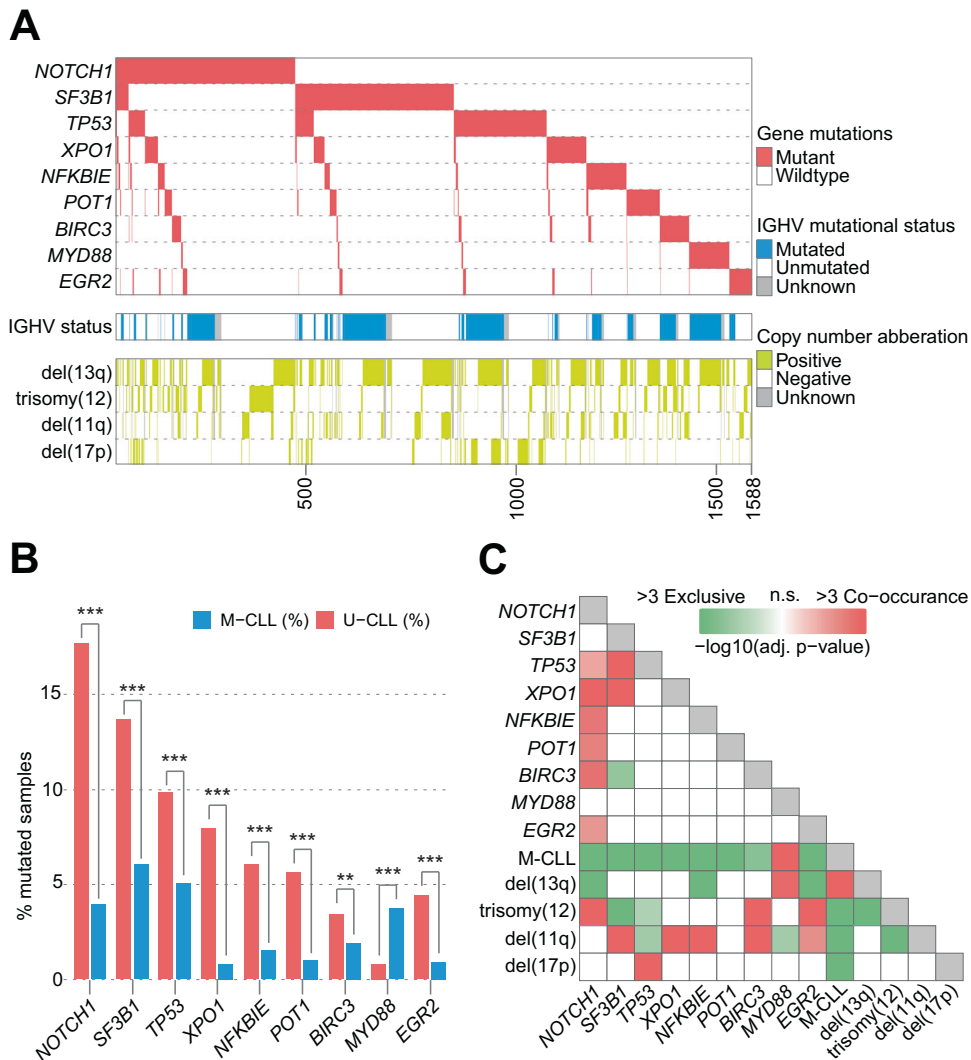
## MATERIALS AND METHODS

### Patients

Our cohort included 4580 patients with CLL from 26 European centers (Supplementary Table S1). Only pre-treatment samples were analyzed; the median time from diagnosis to sample collection was 3 months (data available for 3991 samples). The majority of cases were non-trial patients from referral centers, while the cohort also included patients from the UK LRF CLL4 clinical trial [33] ( $n = 499$ ). The median age at diagnosis was 64.6 years (interquartile range, 56.7–71.7 years). Clinicobiological characteristics for the cohort are shown in Table 1. All cases were diagnosed according to the iwCLL guidelines [4]. Informed consent was obtained according to the Helsinki declaration and the study was approved by the local Ethics Review Committees.

### Mutational analysis

All cases were assessed for mutations within the coding sequences of the *BIRC3*, *EGR2*, *MYD88*, *NFKBIE*, *NOTCH1*, *POT1*, *SF3B1*, *TP53*, and *XPO1* genes. The *ATM* gene was initially included, however, mutation data was only available for 3611 cases and considering the difficulties to interpret *ATM* mutation without germline DNA, it was excluded from further analysis. Mutational screening was performed by NGS for the majority of the cohort (80%), with targeted gene panels, covering all or hotspot exons of each gene, as the most frequent technique employed. Sanger sequencing was performed in 18% of cases and the remaining methods combined, mostly targeting hotspot mutations, accounted for 2% of analyses (Supplementary Table S1). For mutational data generated using NGS analysis, sequence alignment, annotation and variant calling were performed at each center using a  $\geq 5\%$  variant allele frequency (VAF) threshold to classify mutated cases. Variant filtering was performed by including nonsynonymous variants and small insertions/deletions within the coding sequences and by excluding variants with a population frequency  $>0.0001$  in the gnomAD database unless variant was included as a somatic variant in the COSMIC database.



**Fig. 1 Overview of 1588 CLL cases carrying mutations in recurrently mutated genes. A** OncoPrint of detected gene mutations (in order of frequency), IGHV gene somatic hypermutation (SHM) status and chromosomal aberrations. **B** Relative distribution of recurrent gene mutations stratified by IGHV gene SHM status, \*\* denotes a  $p$  value  $<0.01$ , while \*\*\* represents a  $p$  value  $<0.001$ . **C** Co-occurrence of recurrently mutated genes and chromosomal aberrations.

The mutation frequencies were compared in samples analyzed by NGS and Sanger sequencing, examining both hotspot mutations (exemplified by *SF3B1* p.K700E, *NOTCH1* p.P2514fs and *EGR2* p.H383N mutations) or entire genes without hotspot positions (exemplified by *TP53* and *XPO1*), without significant differences in mutation rates when comparing the methods (Supplementary Table 2).

IGHV gene SHM status, available for 4332/4580 (94.6%) patients, was determined using PCR amplification and sequence analysis of IGHV-IGHD-IGHJ gene rearrangements and employing a 98% identity cut-off to germline to define M-CLL ( $<98\%$  identity) and U-CLL ( $\geq 98\%$  identity) [9]. Chromosomal aberrations [data available for 4459/4580 (97.4%) cases] were detected using FISH with targeted probes for chromosomes 13q, 11q, 17p and 12 or Affymetrix 250K SNP-arrays and classified according to the Döhner hierarchical model [6].

### Statistical analysis

The chi-square test was employed for comparing the clinicobiological variables and  $p$  values were Yates corrected. Co-occurrence and exclusivity matrices were constructed using two-sided Fisher's exact tests and  $p$  values were corrected using Benjamini–Hochberg for multiple testing. TTFT was available for 4543/4580 (99.2%) cases and calculated from the diagnostic date until date of first treatment; median follow-up was 3.6 years for the entire cohort and 6.5 years for untreated patients. Survival curves were constructed using the Kaplan–Meier method, pairwise

comparisons using the Cox–Mantel log-rank test determined differences between subgroups and  $p$  values were adjusted using the Benjamini–Hochberg method. The Cox proportional hazards model was employed to assess the prognostic strength of each marker in multi-variable analysis. Stepwise variable selection using Akaike's information criterion was applied in the Cox regression model to examine the significant risk factors associated with TTFT, and the proportion of the explainable log-likelihood was used to examine their relative importance. All statistical analyses were performed using R (version 4.1.2) and R Studio software. Plots were created using ggplot2 v3.3.5, ComplexHeatmap v2.11.1 and G3viz packages [34].

## RESULTS

### Frequency of recurrent gene mutations

For the nine genes analyzed, we detected at least one mutation in 1588/4580 (34.7%) patients (Fig. 1A and Supplementary Fig. S1A, B). Of these, mutations were found in a single gene in 1221/1588 (76.9%) patients, while 321 cases carried mutations in 2 genes, and 46 cases had mutations in 3–4 genes. Except for *MYD88*, mutations were significantly more frequent in U-CLL (Fig. 1B), advanced-stage patients (except for *NFKBIE* mutations) and patients requiring treatment (Supplementary Table S2). In contrast,

*MYD88* mutations were enriched in M-CLL, while no difference was seen when comparing early-stage versus late-stage patients or patients remaining untreated versus those requiring treatment (Fig. 1A and Supplementary Table S3).

### Clinicobiological profiles associated with recurrent gene mutations

*NOTCH1* mutations were the most common, detected in 448/4580 (9.8%) patients, with the highest frequency among U-CLL (17.6%) of all genes studied (Fig. 1A, B). *NOTCH1*-mutated patients had a relatively high proportion of co-occurring mutations in other genes [181/448 (40.4%) cases], most often in *TP53* ( $n = 45$ ) or *XPO1* ( $n = 41$ ). As most centers performed targeted analysis of exon 34 (86%), the previously reported 2 base-pair frameshift deletion (p.2514fs) [13, 14] was detected in 329/448 (73.4%) mutated samples, while no other hotspot was observed within exon 34 (the locations of all coding mutations identified in all genes are illustrated in Supplementary Fig. S2, while hotspot mutations for all genes are summarized in Supplementary Table S4). As expected, *NOTCH1* mutations were significantly associated with trisomy 12 (Fig. 1C and Supplementary Fig. S3).

*SF3B1* mutations were detected in 428/4580 (9.3%) cases and were the most frequent mutation in M-CLL (148/2454, 6.0%), advanced-stage patients at diagnosis (206/1212, 17.0%) and patients requiring treatment (368/2680, 13.7%). As centers predominantly sequenced exons 14–16 (80%), the majority of mutations (>98%) were localized to these exons, with p.K700E (191/423, 45.2%), p.G742D (54/423, 12.8%) and p.K666E/M/N/Q/R/T (39/423, 9.2%) being the most common hotspot mutations [16, 17, 25, 35–37]. Mutations in other genes were observed in 151/423 (35.7%) *SF3B1*-mutated cases with *TP53* mutations, the most common co-occurring event (50/423, 11.8%). Interestingly, *BIRC3* and *SF3B1* mutations appeared almost mutually exclusive, co-occurring in only 3/423 (0.7%) patients (Fig. 1C). Furthermore, *SF3B1* mutations were negatively associated with trisomy 12, only found in 3.9% of trisomy 12 cases (Fig. 1C and Supplementary Fig. S3 and Supplementary Table S3).

*TP53* mutations were identified in 322/4580 (7.0%) patients with 35/322 (10.9%) cases carrying multiple *TP53* mutations [31/35 (89%) cases carried two mutations, the remaining four cases had three mutations]. Mutations frequently coincided with del(17p) [137/311, 44.1% (no data on genomic aberrations available for 11 cases)] and were the second most frequent mutation in M-CLL [122/2454 (5.0%)] (Fig. 1B, C and Supplementary Fig. S3). *TP53* mutations were detected along the entire coding sequence with p.R248Q/W ( $n = 16$ ) being the most common mutation.

*XPO1* mutations were identified in 176/4580 (3.8%) cases with most mutations detected in U-CLL (88%), patients requiring treatment (90%) and patients carrying mutations in a second gene (97/176, 55.1%). Additional mutations were most commonly found in *NOTCH1* ( $n = 41$ ) or *SF3B1* ( $n = 36$ ) while, on the contrary, only 14 patients carried a second mutation in *TP53*. Patients with *XPO1* mutations often carried del(11q) (20.8% of cases) (Fig. 1A, C and Supplementary Fig. S3). Hotspot mutations at p.E571A/G/K/Q/V were detected in 154/175 (88%, mutation data missing for 1 patient) cases followed by p.D624G detected in 7% of patients.

*NFKBIE* mutations were detected in 150/4580 (3.3%) patients. Among these, 125/150 (83.3%) carried the previously reported frameshift 4-bp deletion (p.Y254fs) [18] while an additional three cases displayed a p.Y254\* mutation. Besides *MYD88*, *NFKBIE* was the only other recurrently mutated gene with a similar distribution of mutated cases in early versus late-stage patients, while a significant difference was noted when comparing *NFKBIE* mutations stratified for IGHV gene SHM status or need for treatment ( $p < 0.001$  for both comparisons) (Supplementary Table S3).

*POT1* mutations were identified in 142/4580 (3.1%) cases; no mutational hotspot was evidenced with the most common mutations, p.S38C/G/N/R, observed in only 9/142 (6.3%) cases.

Most *POT1*-mutated patients showed favorable cytogenetics [isolated del(13q) or no recurrent copy-number aberration (71.9% of all cases); Fig. 1A and Supplementary Fig. S3].

*BIRC3* mutations were observed in 122/4580 (2.7%) cases with 25/122 (20%) carrying the p.Q547fs mutation. The only other mutation found in >5 cases was p.E433fs ( $n = 6$ ). *BIRC3* mutated cases had the highest co-occurrence of del(11q) (35.6% of cases). In fact, a large proportion of patients (70.3%) carrying *BIRC3* mutations harbored unfavorable cytogenetics [trisomy 12, del(11q) or del(17p)] (Fig. 1A and Supplementary Fig. S3).

*MYD88* mutations were detected in 114/4580 (2.5%) patients, with the p.L265P and p.V217F mutations identified in 78/114 (68.4%) and 14/114 (12.3%) cases, respectively. *MYD88* mutations were strongly associated with young age at diagnosis (median 59.9 years,  $p < 0.001$ ) and favorable cytogenetics, with 62.5% of cases carrying only del(13q) (Supplementary Fig. S3). Additionally, *MYD88*-mutated cases rarely [16/114 (14.0%) cases] carried a mutation in a second gene (Fig. 1C).

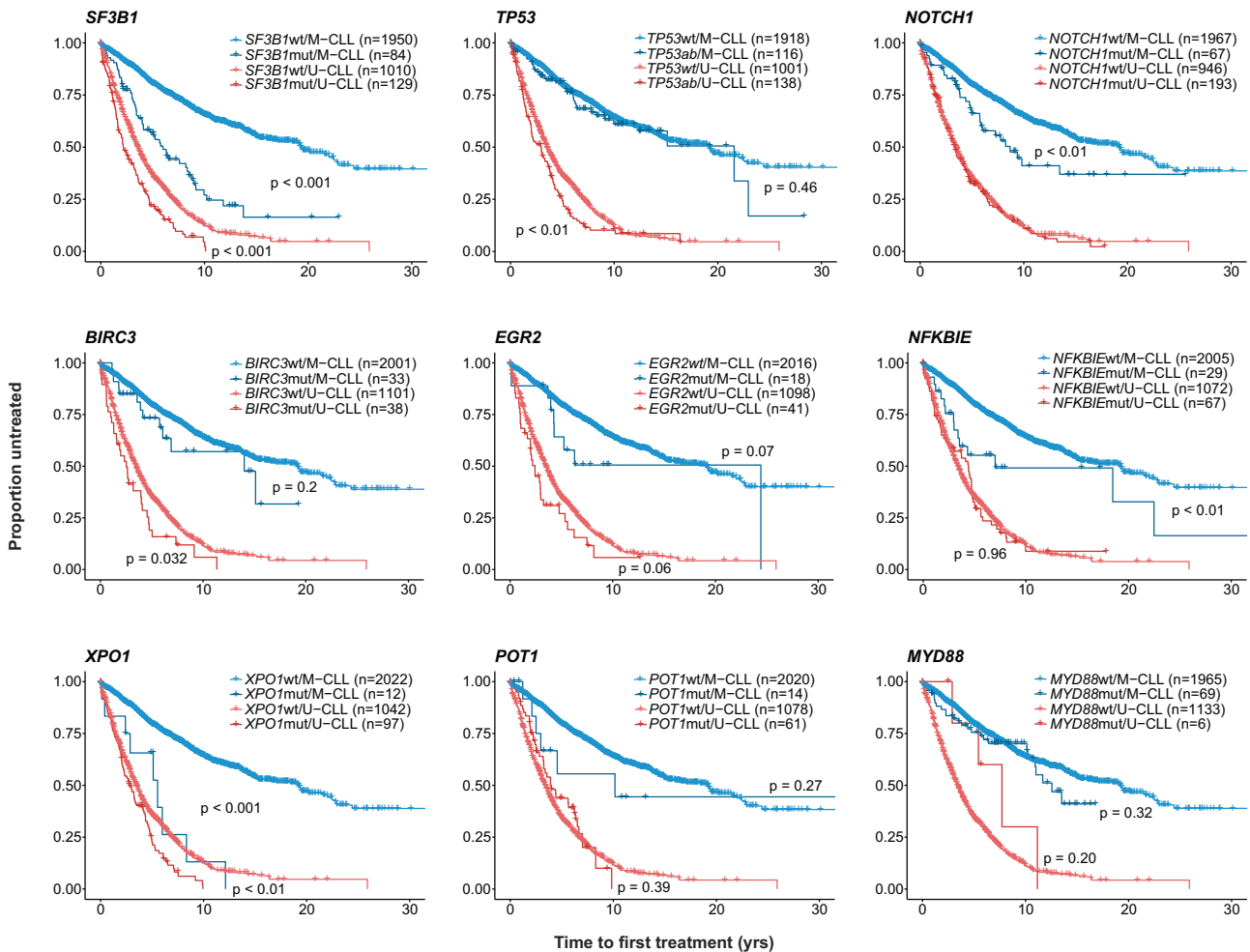
Finally, *EGR2* mutations were the least common, observed in only 106/4580 (2.3%) cases. Two hotspot mutations were detected; p.H384N found in 56/106 (52.8%) cases, followed by the p.E356K mutation detected in 30/106 (28.3%) cases. *EGR2* mutations were associated with unfavorable cytogenetics (59% of cases) (Fig. 1C and Supplementary Fig. S3 and Supplementary Table S3).

### Recurrent gene mutations and clinical outcome

We assessed the clinical impact of each gene mutation individually using TTFT as the primary endpoint. Patients carrying del(17p) and/or mutations in *TP53* were grouped as *TP53* aberrant (*TP53ab*) as they exhibit a similarly poor outcome [10]: 421 patients were classified as *TP53ab* of whom 44% carried *TP53* mutations, 24% harbored del(17p), while 33% had biallelic *TP53* aberrations.

Except for *MYD88*, all mutated genes were associated with significantly worse outcome ( $p < 0.001$ ); *MYD88*-mutated patients displayed no significant difference in TTFT versus *MYD88* wildtype patients ( $p = 0.092$ , Supplementary Fig. S4). Multiple *TP53* mutations (excluding del(17p)) had no added effect on TTFT, nor was there a significant difference when comparing patients carrying a *TP53* mutation with or without del(17p). We also investigated if carrying mutations in multiple genes affected clinical outcome. In this analysis, cases harboring *MYD88*-mutations were grouped with cases wildtype for all other genes. Regardless of the gene mutated, patients with  $\geq 2$  mutations had a significantly worse TTFT compared to those with a single mutation ( $p < 0.001$ ); while no differences were identified when comparing patients with 2 gene mutations versus those with  $\geq 3$  (Supplementary Fig. S5). When stratifying patients based on IGHV gene SHM status, no added effect was found in U-CLL when comparing cases with a single mutation to those with  $\geq 2$  mutations. In contrast, multiple mutations were associated with a worse outcome in M-CLL, however, only 65/2454 (2.6%) cases carried mutations in more than 1 gene.

As advanced-stage disease at diagnosis is an indication for treatment, we performed the analysis in Binet stage A patients and obtained similar results, whereby mutations in all genes, except *MYD88*, resulted in significantly shorter TTFT ( $p < 0.001$  for all, Supplementary Fig. S6). We repeated the analysis in the Binet stage A patient cohort, this time stratifying the patients based on IGHV gene SHM status. Notably, only mutations in *SF3B1* and *XPO1* could further discriminate patients with a statistically significant worse outcome in both U-CLL and M-CLL (Fig. 2). *TP53* and *BIRC3* aberrations were significantly correlated with TTFT in U-CLL only, while, *NOTCH1* and *NFKBIE* mutations were significantly associated with shorter TTFT only in M-CLL (Fig. 2). Finally, mutations in *EGR2*, *POT1* and *MYD88* conferred no additional prognostic information to patients stratified by IGHV gene SHM status (Fig. 2).



**Fig. 2 Time-to-first-treatment (TTFT) in Binet A CLL patients carrying recurrent gene mutations stratified based on IGHV gene SHM status.** Pairwise comparisons were performed using the Cox-Mantel log-rank test. U-CLL, CLL with unmutated IGHV genes, M-CLL, CLL with mutated IGHV genes.

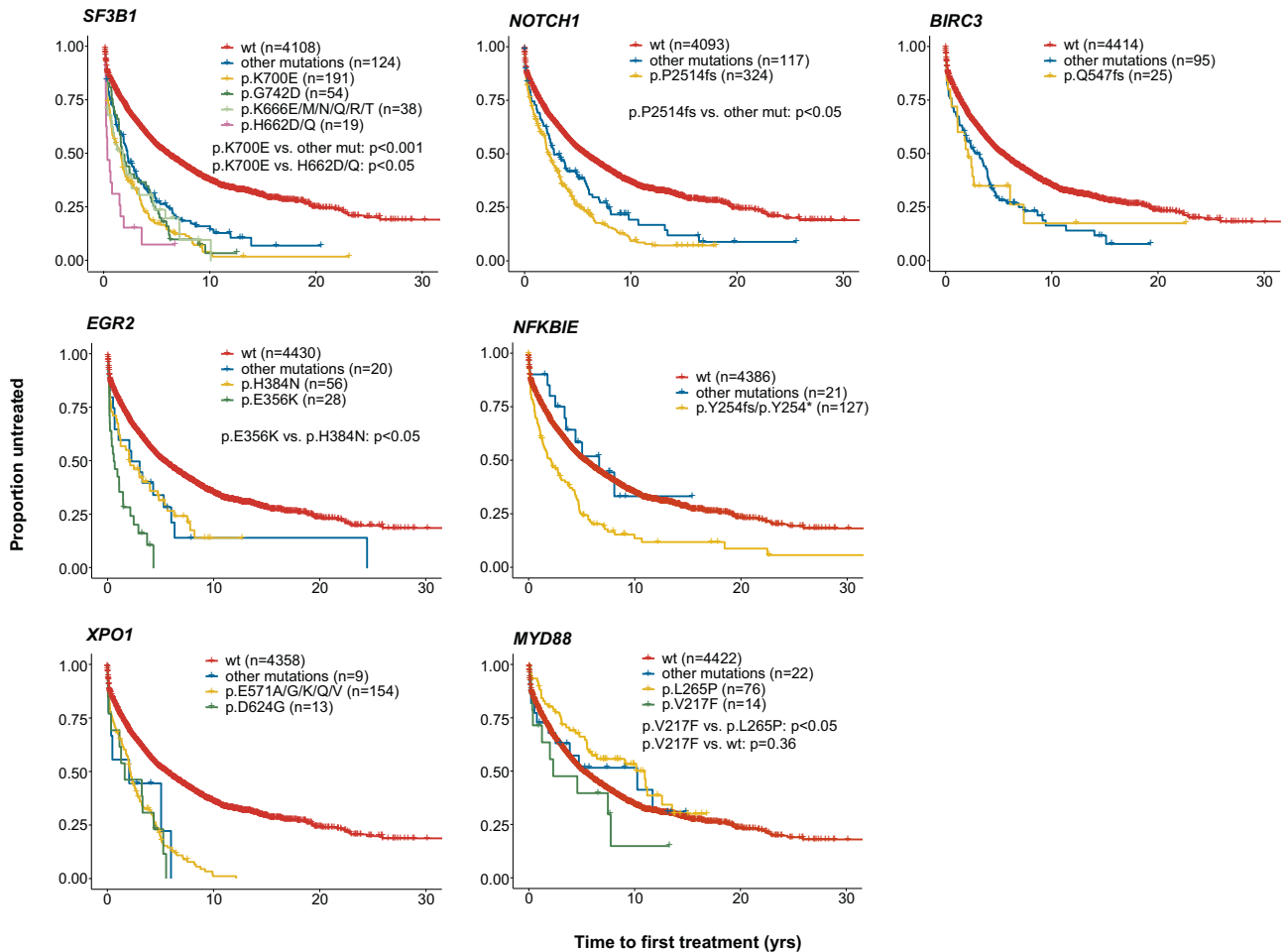
We next investigated how hotspot versus non-hotspot mutations within the same gene affected TTFT. As there were no hotspot mutations detected in *TP53* or *POT1*, these genes were omitted from the analysis. Although all non-hotspot *NOTCH1* mutations combined resulted in a worse outcome when compared to *NOTCH1* wildtype patients ( $p < 0.001$ ), cases carrying the hotspot p.Y254fs mutation had a particularly poor outcome when compared to non-hotspot mutations ( $p < 0.05$ , Fig. 3). A similar trend was observed for *SF3B1* mutations whereby patients with hotspot mutations (p.K700E, p.G742D, p.K666E/M/N/Q/R/T, p.H662D/Q and p.I704F/N/S) appeared to have the worst outcome, although the numbers for some hotspot mutations were low (Fig. 3). Among patients with *EGR2* mutations, those carrying the p.E356K mutation showed a tendency toward worse outcome compared to all other *EGR2*-mutated cases, although not statistically significant, while for *NFKBIE*-mutated patients, only the hotspot p.Y254fs mutation significantly affected clinical outcome (Fig. 3). No differences were detected when comparing variants in *XPO1* or *BIRC3*, while for *MYD88* mutated patients, neither hotspot nor non-hotspot mutations differed significantly from wildtype patients (Fig. 3).

#### Multivariate analysis reveals different impact of molecular markers in U-CLL and M-CLL

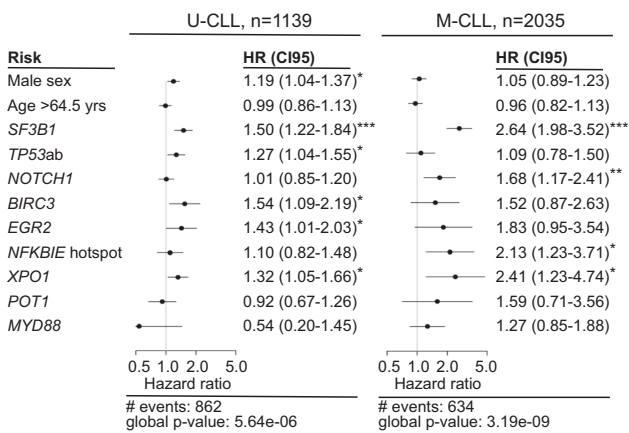
To assess the relationships between the different gene aberrations and their relevance for TTFT in CLL, we first performed a

multivariate model focusing on the nine investigated genes, age at diagnosis as well as sex in the entire cohort. As before, *TP53ab* rather than *TP53* mutations were included as a variable and *NFKBIE*-mutated cases were defined as those carrying only the hotspot p.Y254fs/p.Y254\* mutation based on the association with poor outcome as described above. All variables other than age at diagnosis and *MYD88* mutations, were found to be significant (Supplementary Table S5). In a second model, which also included IGHV gene SHM status, del(11q), trisomy 12 and Binet stage, *SF3B1*, *TP53*, *EGR2*, *NFKBIE* and *XPO1* mutations remained as significant variables, while *NOTCH1*, *POT1* and *BIRC3* mutations were no longer significant (Supplementary Table S6).

Next, we studied Binet A cases only using the same multivariate models and obtained very similar results (Supplementary Tables S5 and S6). We repeated the analysis focusing on the 9 genes (including also age at diagnosis and sex) separately for U-CLL and M-CLL. Notably, *SF3B1* and *XPO1* mutations were the only independent variables in the multivariate analysis of both U-CLL and M-CLL with *SF3B1* showing the highest hazard ratio (Fig. 4). Additionally, male sex, *TP53*, *BIRC3* and *EGR2* aberrations were significant factors for U-CLL, while mutations in *NOTCH1* and *NFKBIE* were significant markers in M-CLL. Using the extended model, including del(11q) and trisomy 12, revealed that *SF3B1* and *XPO1* were again the only independent genes in the multivariate analysis of both U-CLL and M-CLL (Supplementary Table S7).



**Fig. 3 Clinical impact of hotspot versus non-hotspot mutations in recurrently mutated genes assessed using time-to-first-treatment (TTFT) as clinical endpoint.** Pairwise comparisons were performed using the Cox–Mantel log-rank test.



**Fig. 4 Multivariate analysis in Binet stage A CLL patients with unmutated IGHV genes (U-CLL) and mutated IGHV genes (M-CLL).** CI95, 95% confidence interval; \* indicates a  $p$  value <0.05, \*\* denotes a  $p$  value <0.01, while \*\*\* represents a  $p$  value <0.001.

Finally, by investigating the relative impact of the risk factors associated with TTFT, *SF3B1* mutations was the most significant marker followed by *XPO1* mutations in Binet stage A M-CLL patients, while for Binet stage A U-CLL cases, *SF3B1* mutations was the most significant factor by far (Supplementary Fig. S7).

**DISCUSSION**

Over the last 10 years, whole-exome and/or whole-genome sequencing studies have dissected the genomic landscape of CLL, demonstrating that only a few recurrent gene mutations are present in more than 5–10% of cases (*ATM*, *NOTCH1*, *SF3B1*, *TP53*), while the majority are seen in a minor proportion of cases (<1–5%) [20, 21, 32]. To date, recurrent mutations in more than 25 genes have been associated with clinical outcome, most of which affect key signaling pathways and cellular processes [16, 20, 21, 25, 26, 38]. Most of these gene mutations have been linked to a clinically more aggressive disease with shorter TTFT and overall survival. However, while some of these gene mutations have been more comprehensively studied in relatively large cohorts, others are less well explored. Considering the relatively low frequency of mutations in many CLL-related genes, large-scale studies are necessary to offer robust insights into the frequency of these mutations and their impact on prognosis [39].

Another relevant aspect concerns the asymmetric distribution of genomic aberrations between U-CLL and M-CLL, with an enrichment of poor-prognostic alterations in the former [31, 32, 40]. Moreover, the prognostic impact of gene mutations may differ depending on IGHV gene SHM status [29, 31, 32], prompting the suggestion that separate prognostic indices are warranted for U-CLL and M-CLL.

In this study, we analyzed nine CLL-related genes in a cohort of more than 4500 CLL cases to gain comprehensive insight into the impact of different gene mutations on TTFT, particularly in relation to IGHV gene SHM status. The median age at diagnosis was 64.5

years, which is lower than expected and possibly explained by the fact that most contributing institutes are referral centers where patients tend to be younger at admission. As anticipated, a higher prevalence of *NOTCH1*, *SF3B1* and *TP53* aberrations was confirmed, while mutations in the remaining genes were less frequent (between 2–4% of patients). The vast majority of mutated patients carried mutations in a single gene, whereas a minor proportion harbored  $\geq 2$  mutations. In line with previous studies, several genes showed prominent mutation hotspots (e.g., *EGR2*, *MYD88*, *NFKBIE*, *BIRC3*, *NOTCH1*, *SF3B1*, *XPO1*), however this was not observed in other genes (e.g., *POT1*, *TP53*) [18, 19, 41–44]. As previously reported, *NOTCH1* mutations were associated with trisomy 12, *BIRC3* mutations with del(11q) and *MYD88* mutations with sole del(13q) [22, 25, 45, 46]. Notably, certain lesions appeared almost mutually exclusive, such as *BIRC3* and *SF3B1* mutations, and *TP53* aberrations and *XPO1* mutations, implying that acquisition of either of these events is sufficient to develop a more aggressive disease course [47]. For all genes investigated, except for *MYD88*, there was a striking enrichment of mutations in U-CLL compared to M-CLL, particularly for *NOTCH1*, *XPO1*, *NFKBIE* and *EGR2*. Nevertheless, for all genes investigated, mutations were also detected in a minor proportion of M-CLL patients (1–6% of cases).

In univariate analysis of TTFT, all gene mutations, except for *MYD88*, were associated with a significantly worse outcome, both in the entire study cohort and when restricting the analysis to Binet stage A patients. In multivariate analysis in Binet stage A patients, all recurrently mutated genes, again excepting *MYD88*, were independently associated with shorter TTFT. However, when performing the same analysis in U-CLL and M-CLL separately, only *SF3B1* and *XPO1* mutations retained significance in both subgroups. More specifically, *NOTCH1* and *NFKBIE* mutations were significant only in M-CLL, while *TP53*, *BIRC3* and *EGR2* aberrations were significant only in U-CLL. To evaluate the relative importance of the different gene mutations, we also performed stepwise variable selection in M-CLL and U-CLL separately. This analysis revealed that *SF3B1* and *XPO1* mutations had the highest relative impact on TTFT in M-CLL and *SF3B1* and *TP53* aberrations in U-CLL (Supplementary Fig. 7).

Most previous studies have analyzed the impact of gene mutations in relation to different clinical endpoints (e.g., TTFT, progression-free survival and overall survival) [25, 44, 48]. International prognostic schemes have also been developed to identify patients with high-risk disease (CLL-IPI) or high risk of progression (IPS-E) using similar endpoints [49, 50]. In contrast to these studies, which consider CLL patients independent of IGHV gene SHM status, we recently provided preliminary evidence that different markers were relevant in U-CLL versus M-CLL; i.e., *TP53* abnormalities, del(11q) and/or *SF3B1* mutations in U-CLL and *TP53* abnormalities, trisomy 12 and stereotyped subset #2 membership in M-CLL [31]. In the present study, we confirm and significantly extend these initial findings, further underscoring the varying impact of gene mutations in relation to IGHV gene SHM status. In other words, if the aim is to identify CLL patients with the highest risk of progressive disease and in early need of treatment, different sets of genetic biomarkers should be used in U-CLL and M-CLL. For M-CLL, one should assess *SF3B1*, *XPO1*, *NOTCH1* and *NFKBIE*, while *TP53* aberrations do not appear to have any impact on TTFT in this subgroup. Conversely, in U-CLL, *SF3B1*, *TP53*, *XPO1*, *BIRC3* and *EGR2* appear to be the most relevant to analyze for identifying high-risk patients.

One of the limitations of our study is the multicenter data collection, where different sequencing techniques/targeted approaches have been applied with varying sensitivity. That said, in a recent multicenter study performed by ERIC, high concordance was observed for NGS-based gene panels above a VAF of 5% or more, a cutoff which we also applied in the present study [51]. For a minor proportion of cases, approaches that only

targeted hotspot mutations [e.g., *NFKBIE* (8% of cases) and *MYD88* (4% of cases)], or hotspot exons (e.g., *SF3B1* and *NOTCH1*) could potentially underestimate the true frequency of mutations. Here, we found no significant differences in mutation rates when comparing NGS and Sanger sequencing for either hotspot or non-hotspot genes. Additionally, in the case of *NFKBIE*, only the hotspot mutations were clinically relevant, whereas for the remaining genes the significance, if any, of non-hotspot mutations remains unknown. However, we cannot fully exclude that mutations with low frequencies (VAF 5–15%) were missed using Sanger sequencing. In addition, the majority of samples (>80%) derived from non-purified PBMC samples (<20% of samples were purified), which could potentially have affected the ability to detect minor subclones, although the vast majority of unpurified CLL samples have a high tumor percentage (>80–90%). Another caveat concerns the differences in treatment given at the different centers, precluding a meaningful analysis of the impact of gene mutations on overall survival. Moreover, most patients have received chemo(immuno)therapy, hence, in the coming years it will be important to conduct similar large-scale, real-world analysis of the impact of treatment in patients using targeted treatments.

In conclusion, we reinforce *SF3B1* mutations as a key biomarker with a very strong negative impact on TTFT in both M-CLL and U-CLL and also highlight *XPO1* as an additional highly relevant gene in both subgroups. We also demonstrate that *TP53* aberrations are clinically relevant in U-CLL, yet they appear to have no or a limited effect on TTFT in M-CLL. From a clinical perspective, our results may assist in identifying high-risk patients within the heterogeneous M-CLL subgroup with potential implications for stratified management and treatment decisions, also in prospective clinical trials. Since most clinical laboratories are currently applying NGS-based gene panel sequencing to detect *TP53* mutations, often including other CLL-related genes (that are usually not reported), it would be informative to extend the analysis to other genes to identify high-risk patients in routine diagnostics. Finally, future efforts to develop prognostic schemes including gene mutations and other established prognostic factors should apply a more compartmentalized approach, hopefully paving the way for personalized medicine approaches for patients belonging to the M-CLL and U-CLL subgroups.

#### DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

#### REFERENCES

- Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN, Pasternack BS. Clinical staging of chronic lymphocytic leukemia. *Blood*. 1975;46:219–34.
- Binet JL, Lepoprier M, Dighiero G, Charron D, D'Atis P, Vaugier G, et al. A clinical staging system for chronic lymphocytic leukemia: prognostic significance. *Cancer*. 1977;40:855–64.
- Baliakas P, Mattsson M, Stamatopoulos K, Rosenquist R. Prognostic indices in chronic lymphocytic leukaemia: where do we stand how do we proceed? *J Intern Med*. 2016;279:347–57.
- Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Dohner H, et al. iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. *Blood*. 2018;131:2745–60.
- Eichhorst B, Robak T, Montserrat E, Ghia P, Niemann CU, Kater AP, et al. Chronic lymphocytic leukaemia: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2021;32:23–33.
- Dohner H, Stilgenbauer S, Benner A, Leupolt E, Krober A, Bullinger L, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med*. 2000;343:1910–6.
- Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood*. 1999;94:1848–54.

8. Damle RN, Wasil T, Fais F, Ghiotto F, Valetto A, Allen SL, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood*. 1999;94:1840–7.
9. Rosenquist R, Ghia P, Hadzidimitriou A, Sutton LA, Agathangelidis A, Baliakas P, et al. Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: updated ERIC recommendations. *Leukemia*. 2017;31:1477–81.
10. Zenz T, Eichhorst B, Busch R, Denzel T, Habe S, Winkler D, et al. TP53 mutation and survival in chronic lymphocytic leukemia. *J Clin Oncol*. 2010;28:4473–9.
11. Campo E, Cymbalista F, Ghia P, Jager U, Pospisilova S, Rosenquist R, et al. TP53 aberrations in chronic lymphocytic leukemia: an overview of the clinical implications of improved diagnostics. *Haematologica*. 2018;103:1956–68.
12. Malcikova J, Tausch E, Rossi D, Sutton LA, Soussi T, Zenz T, et al. ERIC recommendations for TP53 mutation analysis in chronic lymphocytic leukemia—update on methodological approaches and results interpretation. *Leukemia*. 2018;32:1070–80.
13. Puente XS, Pinyol M, Quesada V, Conde L, Ordonez GR, Villamor N, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2011;475:101–5.
14. Fabbri G, Rasi S, Rossi D, Trifonov V, Khiabani H, Ma J, et al. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. *J Exp Med*. 2011;208:1389–401.
15. Villamor N, Conde L, Martinez-Trillos A, Cazorla M, Navarro A, Bea S, et al. NOTCH1 mutations identify a genetic subgroup of chronic lymphocytic leukemia patients with high risk of transformation and poor outcome. *Leukemia*. 2013;27:1100–6.
16. Jeromin S, Weissmann S, Haferlach C, Dicker F, Bayer K, Grossmann V, et al. SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients. *Leukemia*. 2014;28:108–17.
17. Rossi D, Brusca A, Spina V, Rasi S, Khiabani H, Messina M, et al. Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. *Blood*. 2011;118:6904–8.
18. Mansouri L, Sutton LA, Ljungstrom V, Bondza S, Arngarden L, Bhoi S, et al. Functional loss of I $\kappa$ B $\beta$  leads to NF- $\kappa$ B deregulation in aggressive chronic lymphocytic leukemia. *J Exp Med*. 2015;212:833–43.
19. Young E, Noerenberg D, Mansouri L, Ljungstrom V, Frick M, Sutton LA, et al. EGR2 mutations define a new clinically aggressive subgroup of chronic lymphocytic leukemia. *Leukemia*. 2017;31:1547–54.
20. Landau DA, Tausch E, Taylor-Weiner AN, Stewart C, Reiter JG, Bahlo J, et al. Mutations driving CLL and their evolution in progression and relapse. *Nature*. 2015;526:525–30.
21. Puente XS, Bea S, Valdes-Mas R, Villamor N, Gutierrez-Abril J, Martin-Subero JI, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2015;526:519–24.
22. Martinez-Trillos A, Pinyol M, Navarro A, Aymerich M, Jares P, Juan M, et al. Mutations in TLR/MYD88 pathway identify a subset of young chronic lymphocytic leukemia patients with favorable outcome. *Blood*. 2014;123:3790–6.
23. Baliakas P, Hadzidimitriou A, Agathangelidis A, Rossi D, Sutton LA, Kminkova J, et al. Prognostic relevance of MYD88 mutations in CLL: the jury is still out. *Blood*. 2015;126:1043–4.
24. Rossi D, Rasi S, Spina V, Brusca A, Monti S, Ciardullo C, et al. Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. *Blood*. 2013;121:1403–12.
25. Baliakas P, Hadzidimitriou A, Sutton LA, Rossi D, Minga E, Villamor N, et al. Recurrent mutations refine prognosis in chronic lymphocytic leukemia. *Leukemia*. 2015;29:329–36.
26. Brieghel C, da Cunha-Bang C, Yde CW, Schmidt AY, Kinalis S, Nadeu F, et al. The number of signaling pathways altered by driver mutations in chronic lymphocytic leukemia impacts disease outcome. *Clin Cancer Res*. 2020;26:1507–15.
27. Weissmann S, Roller A, Jeromin S, Hernandez M, Abaigar M, Hernandez-Rivas JM, et al. Prognostic impact and landscape of NOTCH1 mutations in chronic lymphocytic leukemia (CLL): a study on 852 patients. *Leukemia*. 2013;27:2393–6.
28. Rossi D, Rasi S, Fabbri G, Spina V, Fangazio M, Forconi F, et al. Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia. *Blood*. 2012;119:521–9.
29. Best OG, Gardiner AC, Davis ZA, Tracy I, Ibbotson RE, Majid A, et al. A subset of Binet stage A CLL patients with TP53 abnormalities and mutated IGHV genes have stable disease. *Leukemia*. 2009;23:212–4.
30. Burns A, Alsolami R, Becq J, Stamatopoulos B, Timbs A, Bruce D, et al. Whole-genome sequencing of chronic lymphocytic leukaemia reveals distinct differences in the mutational landscape between IgHV(mut) and IgHV(unmut) subgroups. *Leukemia*. 2018;32:573.
31. Baliakas P, Moysiadis T, Hadzidimitriou A, Xochelli A, Jeromin S, Agathangelidis A, et al. Tailored approaches grounded on immunogenetic features for refined prognostication in chronic lymphocytic leukemia. *Haematologica*. 2019;104:360–9.
32. Knisbacher BA, Lin Z, Hahn CK, Nadeu F, Duran-Ferrer M, Stevenson KE, et al. Molecular map of chronic lymphocytic leukemia and its impact on outcome. *Nat Genet*. 2022;54:1664–74.
33. Catovsky D, Richards S, Matutes E, Oscier D, Dyer M, Bezares RF, et al. Assessment of fludarabine plus cyclophosphamide for patients with chronic lymphocytic leukaemia (the LRF CLL4 Trial): a randomised controlled trial. *Lancet*. 2007;370:230–9.
34. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32:2847–9.
35. Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med*. 2011;365:2497–506.
36. Oscier DG, Rose-Zerilli MJ, Winkelmann N, Gonzalez de Castro D, Gomez B, Forster J, et al. The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial. *Blood*. 2013;121:468–75.
37. Quesada V, Conde L, Villamor N, Ordonez GR, Jares P, Bassaganyas L, et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat Genet*. 2012;44:47–52.
38. Ljungstrom V, Cortese D, Young E, Pandzic T, Mansouri L, Plevova K, et al. Whole-exome sequencing in relapsing chronic lymphocytic leukemia: clinical impact of recurrent RPS15 mutations. *Blood*. 2016;127:1007–16.
39. Mansouri L, Thorvaldsdottir B, Laidou S, Stamatopoulos K, Rosenquist R. Precision diagnostics in lymphomas—recent developments and future directions. *Semin Cancer Biol*. 2022;84:170–83.
40. Burns A, Alsolami R, Becq J, Stamatopoulos B, Timbs A, Bruce D, et al. Correction: Whole-exome sequencing of chronic lymphocytic leukaemia reveals distinct differences in the mutational landscape between IgHV(mut) and IgHV(unmut) subgroups. *Leukemia*. 2019;33:2342.
41. Del Giudice I, Rossi D, Chiaretti S, Marinelli M, Tavolaro S, Gabrielli S, et al. NOTCH1 mutations in +12 chronic lymphocytic leukemia (CLL) confer an unfavorable prognosis, induce a distinctive transcriptional profiling and refine the intermediate prognosis of +12 CLL. *Haematologica*. 2012;97:437–41.
42. Taylor J, Sendino M, Gorelick AN, Pastore A, Chang MT, Penson AV, et al. Altered nuclear export signal recognition as a driver of oncogenesis. *Cancer Discov*. 2019;9:1452–67.
43. Sutton LA, Ljungstrom V, Mansouri L, Young E, Cortese D, Navrkalova V, et al. Targeted next-generation sequencing in chronic lymphocytic leukemia: a high-throughput yet tailored approach will facilitate implementation in a clinical setting. *Haematologica*. 2015;100:370–6.
44. Blakemore SJ, Clifford R, Parker H, Antoniou P, Stec-Dziedzic E, Larrayoz M, et al. Clinical significance of TP53, BIRC3, ATM and MAPK-ERK genes in chronic lymphocytic leukaemia: data from the randomised UK LRF CLL4 trial. *Leukemia*. 2020;34:1760–74.
45. Balatti V, Bottoni A, Palamarchuk A, Alder H, Rassenti LZ, Kipps TJ, et al. NOTCH1 mutations in CLL associated with trisomy 12. *Blood*. 2012;119:329–31.
46. Rose-Zerilli MJ, Forster J, Parker H, Parker A, Rodriguez AE, Chaplin T, et al. ATM mutation rather than BIRC3 deletion and/or mutation predicts reduced survival in 11q-deleted chronic lymphocytic leukemia: data from the UK LRF CLL4 trial. *Haematologica*. 2014;99:736–42.
47. Walker JS, Hing ZA, Harrington B, Baumhardt J, Ozer HG, Lehman A, et al. Recurrent XPO1 mutations alter pathogenesis of chronic lymphocytic leukemia. *J Hematol Oncol*. 2021;14:17.
48. Tausch E, Beck P, Schlenk RF, Jebaraj BJ, Dolnik A, Yosifov DY, et al. Prognostic and predictive role of gene mutations in chronic lymphocytic leukemia: results from the pivotal phase III study COMPLEMENT1. *Haematologica*. 2020;105:2440–7.
49. International CLLIPIwg. An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. *Lancet Oncol*. 2016;17:779–90.
50. Condoluci A, Terzi di Bergamo L, Langerbeins P, Hoehstetter MA, Herling CD, De Paoli L, et al. International prognostic score for asymptomatic early-stage chronic lymphocytic leukemia. *Blood*. 2020;135:1859–69.
51. Sutton LA, Ljungstrom V, Enjuanes A, Cortese D, Skafason A, Tausch E, et al. Comparative analysis of targeted next-generation sequencing panels for the detection of gene mutations in chronic lymphocytic leukemia: an ERIC multicenter study. *Haematologica*. 2021;106:682–91.

## ACKNOWLEDGEMENTS

The European Research Initiative on CLL (ERIC) is a partner in the HARMONY Alliance, the EHA Scientific Working group on CLL and the ELN Workpackage 7 on CLL. This work was in part supported by; Associazione Italiana per la Ricerca sul Cancro—AIRC, Milano, Italy (Investigator Grant #20246 and Special Program on Metastatic Disease—5 per mille #21198); ERA NET TRANSCAN-2 Joint Transnational Call for Proposals: JTC 2014 (project #143 GCH-CLL) and JTC 2016 (project #179 NOVEL), project code (MIS) 5041673; Bando della Ricerca Finalizzata 2018, Ministero della Salute, Roma, Italy (progetto RF-2018–12368231); “la Caixa” Foundation (Health Research 2017 Program HR17-00221); the American Association for Cancer Research (2021 AACR-Amgen Fellowship in Clinical/Translational Cancer Research, 21-40-11-NADE), the European



Hematology Association (EHA Junior Research Grant 2021, RG-202012-00245), and the Lady Tata Memorial Trust (International Award for Research in Leukaemia 2021-2022, LADY\_TATA\_21\_3223); the Hellenic Precision Medicine Network in Oncology; project ODYSSEAS (Intelligent and Automated Systems for enabling the Design, Simulation and Development of Integrated Processes and Products) implemented under the “Action for the Strategic Development on the Research and Technological Sector”, funded by the Operational Programme “Competitiveness, Entrepreneurship and Innovation” (NSRF 2014-2020) and co-financed by Greece and the European Union, with grant agreement no: MIS 5002462; MH CZ—DRO (FNBr, 65269705), NV19-03-00091 and the project National Institute for Cancer Research (Programme EXCELES, ID Project No. LX22NPO5102)—Funded by the European Union—Next-Generation EU; Instituto de Salud Carlos III (ISCIII), “PI21/00983”, co-funded by the European Union; the EU’s Horizon 2020 research and innovation program under grant agreement No. 739593, by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the K21\_137948, FK20\_134253, TKP2021-EGA-24 and TKP2021-NVA-15 funding schemes, and Elixir Hungary; Instituto de Salud Carlos III (ISCIII), “PI21/00983”, co-funded by the European Union; Fondo di Ateneo per la Ricerca (FAR) 2019, 2020 and 2021 of the University of Ferrara (GMR; AC), Associazione Italiana contro le Leucemie-Linfomi e Mieloma ONLUS Ferrara (AC; GMR), BEAT Leukemia Foundation Milan Italy (AC); the Danish Cancer Society and the CLL-CLUE project under the frame of ERA PerMed; Cancer Research UK (ECRIN-M3 accelerator award C42023/A29370, Southampton Experimental Cancer Medicine Centre grant C24563/A15581, Cancer Research UK Southampton Centre grant C34999/A18087, and programme C2750/A23669); the Swedish Cancer Society (19 0425 Pj 01 H), the Swedish Research Council (2020-01750), the Knut and Alice Wallenberg Foundation (2016.0373), Region Stockholm (ALF/FoU-962423), and Radiumhemmets Forskningsfonder (194133), Stockholm; and Lion’s Cancer Research Foundation, Uppsala.

## AUTHOR CONTRIBUTIONS

LM, BT and LS summarized, analyzed and interpreted data and wrote the paper; BT, GK, HF-A and AS performed bioinformatics/statistical analysis; MM, HP, FN, CB, SL, RM, DR, MC, JK, JD, AER-V, RB, GMR, SB, L, C, MaMa, ZD, AG, LR, PB, CJ, IR, FM, JM-L, JdIS, JMHR, PT, MJL, MJC, VF, ZM, CsBö, KES, BE, AP, RG, LB, FB, BT-V, FB-M, DO, FN-K, TZ, MJT, AC, MH-S, SP, KM, GG, CUN, EC and JCS provided samples and clinical data; PG, KS and RR designed the study, interpreted data and wrote the paper; all authors edited and approved the paper for submission.

## FUNDING

Open access funding provided by Karolinska Institute.

## COMPETING INTERESTS

PG: Honoraria/advisory board: AbbVie, Acerta/AstraZeneca, Adaptive, ArQule/MSD, BeiGene, CelGene/Juno, Gilead, Janssen, Loxo/Lilly, Sunesis. Research funding: AbbVie, Gilead, Janssen, Novartis, Sunesis; LS: advisory board AbbVie, AstraZeneca, Janssen; RR: honoraria/advisory board: Abbvie, AstraZeneca, Janssen, Illumina and Roche; KS: honoraria/advisory board: AbbVie, Acerta/AstraZeneca, Gilead, Janssen. Research funding: AbbVie, Gilead, Janssen; PB: honoraria from Abbvie, Gilead and Janssen. Research funding from Gilead; GG: Advisory Board/Speaker’s bureau: Janssen, Abbvie, AstraZeneca, Beigene; LB: honoraria/advisory board: Abbvie, Amgen, Astellas, BMS/Celgene, Daiichi Sankyo, Gilead, Hexal, Janssen, Jazz Pharmaceuticals, Menarini, Novartis, Pfizer, Sanofi, Seattle Genetics. Research funding: Bayer, Jazz Pharmaceuticals; GMR: honoraria from Abbvie, AstraZeneca, Gilead and Janssen. Research funding from Gilead: CB: Honoraria/advisory board: AbbVie, Janssen. The other authors declare no competing financial interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41375-022-01802-y>.

**Correspondence** and requests for materials should be addressed to Richard Rosenquist

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons

Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022, corrected publication 2023

<sup>1</sup>Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden. <sup>2</sup>Centre for Research and Technology Hellas, Institute of Applied Biosciences, Thessaloniki, Greece. <sup>3</sup>Department of Mathematics, University of Ioannina, Ioannina, Greece. <sup>4</sup>MLL Munich Leukemia Laboratory, Munich, Germany. <sup>5</sup>Cancer Genomics, School for Cancer Sciences, Faculty of Medicine, University of Southampton, Southampton, UK. <sup>6</sup>Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain. <sup>7</sup>Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain. <sup>8</sup>Department of Hematology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark. <sup>9</sup>Division of Hematology, Department of Translational Medicine, Università del Piemonte Orientale, Novara, Italy. <sup>10</sup>Division of Hematology, Oncology Institute of Southern Switzerland, Bellinzona, Switzerland. <sup>11</sup>Laboratory of Experimental Hematology, Institute of Oncology Research, Bellinzona, Switzerland. <sup>12</sup>Patrick G Johnston Centre for Cancer Research, Queen’s University Belfast, Belfast, UK. <sup>13</sup>Department of Internal Medicine—Hematology and Oncology, University Hospital Brno, Brno, Czech Republic. <sup>14</sup>Faculty of Medicine, Masaryk University, Brno, Czech Republic. <sup>15</sup>Central European Institute of Technology, Masaryk University, Brno, Czech Republic. <sup>16</sup>Hospital Clínic de Barcelona, Barcelona, Spain. <sup>17</sup>Universitat de Barcelona, Barcelona, Spain. <sup>18</sup>Cancer Research Center (IBMCC) CSIC—University of Salamanca, Salamanca, Spain. <sup>19</sup>Instituto de Investigación Biomédica (IBSAL), Salamanca, Spain. <sup>20</sup>Department of Hematology, University Hospital of Salamanca, Salamanca, Spain. <sup>21</sup>Hematology—Department of Medical Sciences, University of Ferrara, Ferrara, Italy. <sup>22</sup>Università Vita Salute San Raffaele and IRCCS Ospedale San Raffaele, Milano, Italy. <sup>23</sup>Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden. <sup>24</sup>Molecular Pathology Department, University Hospitals Dorset, Bournemouth, UK. <sup>25</sup>All India Institute of Medical Sciences (AIIMS), New Delhi, India. <sup>26</sup>Hospital Universitario 12 Octubre, Madrid, Spain. <sup>27</sup>Spanish National Cancer Research (CNIO), Madrid, Spain. <sup>28</sup>Haematology Department, Beaumont Hospital, Dublin, Ireland. <sup>29</sup>Hematological Diseases Laboratory, CIMA LAB Diagnostics, University of Navarra, Pamplona, Spain. <sup>30</sup>IdiSNA, Navarra Institute for Health Research, Pamplona, Spain. <sup>31</sup>HCEMM-SE Molecular Oncohematology Research Group, Department of Pathology and Experimental Cancer Research, Semmelweis University, Budapest, Hungary. <sup>32</sup>Central Hospital of Southern Pest—National Institute of Hematology and Infectious Diseases, Budapest, Hungary. <sup>33</sup>Clinical Epidemiology Division, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden. <sup>34</sup>Molecular Cytogenetics Laboratory, Pathology Department, Hospital del Mar and Translational Research on Hematological Neoplasms Group, Hospital del Mar Research Institute (IMIM), Barcelona, Spain. <sup>35</sup>Department of Hematology, Oncology and Cancer Immunology, Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin, Germany. <sup>36</sup>Department of Hematology, Hospital Universitari Vall d’Hebron (HUVH), Experimental Hematology, Vall d’Hebron Institute of Oncology (VHIO), Department of Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain. <sup>37</sup>Service d’hématologie Biologique Hôpital Avicenne Assistance Publique des Hôpitaux de Paris, Bobigny, France. <sup>38</sup>Sorbonne Université, Service d’Hématologie Clinique, Hôpital Pitié-Salpêtrière, APHP, Paris, France. <sup>39</sup>Department of Oncology and Haematology, University Hospital and University of Zurich, Zurich, Switzerland. <sup>40</sup>Department of Hematology, INCLIVA Research Institute, University of Valencia, Valencia, Spain. <sup>41</sup>Clinical Genetics, Karolinska University Hospital, Solna, Sweden. <sup>42</sup>These authors contributed equally: Larry Mansouri, Birna Thorvaldsdottir, Lesley-Ann Sutton. <sup>43</sup>These authors jointly supervised this work: Paolo Ghia, Kostas Stamatopoulos, Richard Rosenquist. ✉email: [richard.rosenquist@ki.se](mailto:richard.rosenquist@ki.se)