

## ARTICLE OPEN



## Disease-specific U1 spliceosomal RNA mutations in mature B-cell neoplasms

Ferran Nadeu<sup>1,2,50</sup>✉, Shimin Shuai<sup>3,50</sup>, Guillem Clot<sup>1,2,4,50</sup>, Laura K. Hilton<sup>5</sup>, Ander Diaz-Navarro<sup>2,6</sup>, Silvia Martín<sup>1,2,48</sup>, Romina Royo<sup>7</sup>, Tycho Baumann<sup>8,49</sup>, Marta Kulis<sup>1,2</sup>, Irene López-Oreja<sup>1,2,8</sup>, Manuel Cossio<sup>1</sup>, Junyan Lu<sup>9</sup>, Viktor Ljungström<sup>10</sup>, Emma Young<sup>11</sup>, Karla Plevova<sup>12,13,14</sup>, Binyamin A. Knisbacher<sup>15,16</sup>, Ziao Lin<sup>15,17</sup>, Cynthia K. Hahn<sup>15,18</sup>, Pablo Bousquets<sup>6</sup>, Miguel Alcoceba<sup>2,19</sup>, Marcos González<sup>2,19</sup>, Enrique Colado<sup>20</sup>, Ángel R. Payer<sup>20</sup>, Marta Aymerich<sup>1,2,8</sup>, María J. Terol<sup>21</sup>, Alfredo Rivas-Delgado<sup>1,8</sup>, Anna Enjuanes<sup>1</sup>, Sílvia Ruiz-Gaspà<sup>1</sup>, Thomas Chatzikonstantinou<sup>22,23</sup>, Daniel Hägerstrand<sup>11</sup>, Cecilia Jylhä<sup>11,24</sup>, Aron Skafason<sup>11</sup>, Larry Mansouri<sup>11</sup>, Kamila Stranska<sup>12,14</sup>, Michael Doubek<sup>12,13</sup>, Ellen J. van Gastel-Mol<sup>25</sup>, Zadie Davis<sup>26</sup>, Renata Walewska<sup>27</sup>, Lydia Scarfò<sup>28,29</sup>, Livio Trentin<sup>30</sup>, Andrea Visentin<sup>30</sup>, Sameer A. Parikh<sup>31</sup>, Kari G. Rabe<sup>32</sup>, Riccardo Moia<sup>33</sup>, Marine Armand<sup>34</sup>, Davide Rossi<sup>35</sup>, Frederic Davi<sup>34</sup>, Gianluca Gaidano<sup>33</sup>, Neil E. Kay<sup>31</sup>, Tait D. Shanafelt<sup>36</sup>, Paolo Ghia<sup>28,29</sup>, David Oscier<sup>27</sup>, Anton W. Langerak<sup>25</sup>, Sílvia Beà<sup>1,2,4,8</sup>, Armando López-Guillermo<sup>1,2,4,8</sup>, Donna Neuberg<sup>37</sup>, Catherine J. Wu<sup>15,18,38,39</sup>, Gad Getz<sup>15,38,40,41</sup>, Sarka Pospisilova<sup>12,13,14</sup>, Kostas Stamatopoulos<sup>11,22,23</sup>, Richard Rosenquist<sup>11,24</sup>, Wolfgang Huber<sup>42</sup>, Thorsten Zenz<sup>43</sup>, Dolors Colomer<sup>1,2,4,8</sup>, José I. Martín-Subero<sup>1,4,44</sup>, Julio Delgado<sup>1,2,4,8</sup>, Ryan D. Morin<sup>5,45</sup>, Lincoln D. Stein<sup>46,47</sup>, Xose S. Puente<sup>2,6</sup> and Elías Campo<sup>1,2,4,8</sup>✉

© The Author(s) 2025

Recurrent mutations in the third base of U1 spliceosomal RNA responsible for marked splicing and expression abnormalities have been described in chronic lymphocytic leukemia (CLL) and some solid tumors. However, the clinical significance of these mutations in large and independent CLL cohorts as well as their presence in other B-cell neoplasms is unknown. Here we characterized U1 mutations in 1670 CLL and 363 mature B-cell lymphomas. We confirmed that the g.3A>C U1 mutation is found in 3.5% of CLL, which conferred rapid disease progression independently of the main biological and clinical prognostic markers of the disease. Additionally, a recurrent g.9C>T mutation was found in 1.5% of CLL causing downstream splicing alterations and associated with adverse prognosis. We also identified a g.4C>T mutation in 10% of diffuse large B-cell lymphomas of the germinal center subtype and a g.7A>G mutation in 30% of EBV-negative Burkitt lymphomas, both of which altered the splicing pattern of multiple genes. This study reveals novel, recurrent, and tumor-specific U1 mutations in mature B-cell neoplasms with biological and prognostic implications, thus establishing U1 as a novel pan-B-cell malignancy driver gene.

*Leukemia* (2025) 39:2076–2086; <https://doi.org/10.1038/s41375-025-02667-7>

## INTRODUCTION

The small nuclear RNA U1, which is present in multiple identical copies in the human genome, is part of the spliceosome complex where it is involved in 5' splice site (5'SS) recognition via base-pairing and has been found to be mutated in cancer [1, 2]. The third base of the gene is a hotspot site with an A>C mutation (g.3A>C) detected in 3.8% of chronic lymphocytic leukemia (CLL) and 5.9% of hepatocarcinoma patients [1]. A distinct g.3A>G mutation was found in 50% of patients with Sonic Hedgehog medulloblastoma [2]. In CLL, the g.3A>C mutation induces global gene splicing and expression changes with more than 1500 differentially spliced introns and 800 differentially expressed genes between U1 mutated and wild-type tumors [1]. Our initial study [1] found the g.3A>C mutation exclusively in CLL with unmutated immunoglobulin heavy variable (IGHV) genes (U-CLL) [3, 4], and the presence of the mutation was associated with a short time to first treatment (TTFT) [1]. However, this study focused on a relatively small cohort of 318 patients [1]. In our CLL

molecular map study of 1148 patients [5], the g.3A>C U1 mutation was found in 3.8% of cases and virtually exclusive of U-CLL, and associated with a specific CLL expression cluster [5]. The comprehensive nature of this study and the heterogeneous cohorts analyzed [5] precluded us from performing a detailed analysis of U1 mutations and their impact on outcome. Altogether, the clinical implications of this g.3A>C U1 mutation, its relationship with other driver alterations, its dynamics through the disease course, and the possible relevance of other (i.e., non-position 3) U1 mutations remain undisclosed. In addition to the findings in CLL, our initial analysis of the Pan-Cancer Analysis of Whole Genomes Consortium (PCAWG) cohort identified a number of U1 mutations in B-cell non-Hodgkin lymphomas, suggesting that these mutations may also be relevant in other lymphoid neoplasms [1]. However, the low number of patients analyzed and lack of histological characterization prevented us from reaching any solid conclusions on the significance of this observation [1].

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

Received: 3 January 2025 Revised: 19 May 2025 Accepted: 13 June 2025

Published online: 30 June 2025

In the current study, we aimed to provide a comprehensive catalog of U1 mutations across a wide spectrum of mature B-cell neoplasms. First, we performed an unbiased characterization of the U1 mutational status in 762 B-cell neoplasms analyzed by whole-genome sequencing (WGS) and complemented with RNA sequencing (RNA-seq) data. This WGS cohort comprised 399 CLL and 363 mature B-cell lymphomas, including diffuse large B-cell lymphomas (DLBCL), Burkitt lymphomas (BL), mantle cell lymphomas (MCL), and follicular lymphomas (FL). Second, we expanded the biologic and clinical characterization of recurrent U1 mutations in immunogenetic- and epigenetic-based subtypes [3, 4, 6, 7] of CLL by integrating genomic and transcriptomic data of 1,670 patients from two independent cohorts. The results presented here below uncover specific U1 mutations in distinct B-cell neoplasms with downstream molecular and clinical consequences.

## METHODS

### Cohorts studied

A total of 1,670 patients with CLL were included in this study. Patients with CLL were divided in two independent cohorts: cohort 1 (C1)-CLL comprised 1,120 CLL patients from our International Cancer Genome Consortium (ICGC) [8], which included 318 patients analyzed previously [1]; and cohort 2 (C2)-CLL comprised 550 patients from three distinct centers [University Hospital Heidelberg ( $n = 258$ ), a multi-center WGS cohort coordinated by Karolinska Institutet on behalf of the European Research Initiative on CLL (ERIC) ( $n = 170$ ), and Dana-Farber Cancer Institute (DFCI,  $n = 122$ )] (Supplementary Table 1) [5, 9, 10]. We also studied 363 previously published B-cell lymphoma patients including 155 DLBCL [11–13], 110 BL [13, 14], 61 MCL [15], and 37 FL [13] (Supplementary Table 2). The study was approved by the Ethics Committee of the Hospital Clínic de Barcelona. Informed consent was obtained from all patients. All methods were performed in accordance with the relevant guidelines and regulations.

### Bioinformatic analyses and functional validations

We developed a novel U1 mutation calling pipeline based on a matched tumor-normal mixture model to call somatic and germline mutations in any of the 11 canonical U1 genes found in the human genome (GRCh38/hg38) from WGS data (Fig. 1a, b, Supplementary Fig. 1, Supplementary Table 3, Supplementary Methods). Analyses of RNA-seq and gene expression microarrays, PCR-based genotyping of U1 mutations (rhAMP assay), and transfection experiments in cell lines were performed as previously described [1, 8] and are detailed in Supplementary Methods.

## RESULTS

### U1 mutations in 762 mature B-cell neoplasm whole genomes

We first re-analyzed the WGS data of 762 patients with B-cell neoplasms, 399 CLL and 163 mature B-cell lymphomas, to characterize the distribution of mutations along the U1 gene. We identified 155 U1 mutations in 133/762 (17.5%) samples. Among them, 150 mutations were somatic and 5 were present in the germline of the patients and included in downstream analyses since they affected the 5'SS recognition sequence of U1 (positions 3–10) (Supplementary Table 4). The high purity of the normal DNA of the 5 patients with U1 germline mutations suggested that these mutations represented true germline events. Based on the calculated percentage of cells carrying each mutation, virtually all U1 mutations occurred in heterozygosity in a single canonical U1 gene, which could be identified for most somatic and germline mutations (see Supplementary Table 4 for details on the somatic origin of the mutations, mutated U1 genes, and corresponding genomic coordinates).

The incidence of U1 mutations differed between entities: 56 mutations in 51/399 (12.8%) CLL, 59 mutations in 46/155 (29.7%) DLBCL, 27 in 24/110 (21.8%) BL, 5 in 5/61 (8.2%) MCL, and 8 in 7/37 (18.9%) FL (Fig. 1c). We found mutations in 51 distinct positions of the 164-bp transcribed portion of the U1 gene, but the majority of

mutations (98/155, 63.2%) were found between positions 3 and 10, the regions responsible for 5'SS recognition via base-pairing (Fig. 1a). Of note, 80.4% of the mutations identified in CLL, 54.2% in DLBCL, and 63% in BL were located within the 5'SS recognition sequence of U1 (Fig. 1c). In line with this, 42/399 (10.5%) CLL, 27/155 (17.4%) DLBCL, and 17/110 (15.5%) BL carried mutations in the U1 5'SS recognition sequence, which contrast with only 2/61 (3.3%) MCL and 2/37 (5.4%) FL (Fig. 1d). We observed substantial differences in the mutated sites of CLL, DLBCL, and BL. In CLL, the most frequently mutated site was position 3 with 29 tumors (7.3% of the whole CLL WGS cohort) carrying the previously identified g.3A>C mutation [1]. We also identified two additional recurrent mutations in CLL, a C>T mutation at position 9 (g.9C>T) in 8 (2%) and a T>A mutation at position 5 (g.5T>A) in 3 (0.75%) (Fig. 1d). DLBCL carried recurrent mutations in position 3 (g.3A>C,  $n = 7$ , 4.5%), position 4 (g.4C>T,  $n = 13$ , 8.4%; g.4C>G,  $n = 2$ , 1.3%), position 8 (g.8C>T,  $n = 3$ , 1.9%) and position 9 (g.9C>T,  $n = 3$ , 1.9%), while BL in position 4 (g.4C>T,  $n = 4$ , 3.6%) and position 7 (g.7A>G,  $n = 12$ , 10.9%) (Fig. 1d).

Overall, the 5'SS recognition sequence of U1 is recurrently mutated in CLL, DLBCL and BL, with distinct mutations enriched in specific entities. Given the reported downstream effects associated with mutations in position 3 [1, 2], the novel mutations identified between positions 4–10 in DLBCL, BL and CLL could also alter the 5'SS recognition and binding of U1.

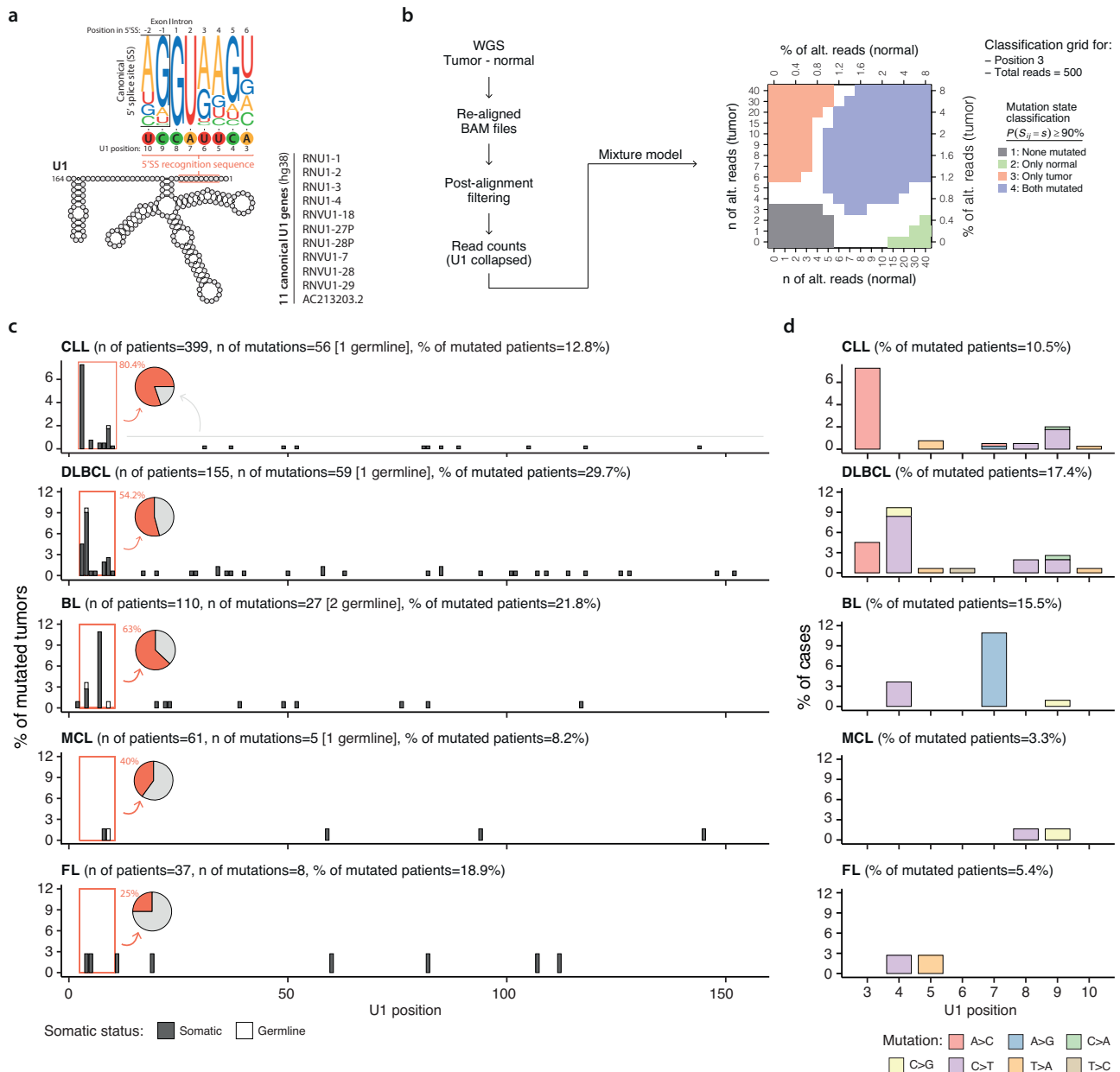
### Downstream effect of U1 mutations in DLBCL subtypes

Based on the cell of origin [16], our DLBCL cohort of 155 patients included 53 activated B-cell like (ABC)-DLBCL, 78 germinal center B-cell like (GCB)-DLBCL, and 11 unclassified tumors, while this information was not available for the remaining 13 patients (Supplementary Table 2). U1 mutations found in the 5'SS recognition sequence (positions 3–10) showed a non-significant enrichment in GCB-DLBCL (16/78 mutated, 20.5%) compared to ABC-DLBCL (5/53, 9.4%) ( $p = 0.14$ ) (Fig. 2a). This trend was confirmed on the identified hotspot positions 3 and 4 of the U1 gene. Within GCB-DLBCL, g.3A>C and g.4C>T/g.4C>G mutations accounted for 5/78 (6.4%) and 10/78 (12.8%) of tumors, respectively, which contrasted with the frequency of the g.3A>C (0%) and g.4C>T (2/53; 3.8%) in ABC-DLBCL ( $p = 0.08$  and  $p = 0.12$ , respectively) (Fig. 2b). The U1 g.4C>T mutation in GCB-DLBCL co-occurred significantly with *HIST1H1E* and *TBL1XR1* mutations, while both g.3A>C and g.4C>T co-occurred with *SOC1* ( $q < 0.1$ , Fig. 2c, Supplementary Fig. 2).

We next conducted differential splicing analyses in the RNA-seq data of GCB-DLBCL (4g.3A>C, 7g.4C>T, and 44 wild-type). The analysis for g.3A>C in GCB-DLBCL revealed no significant difference for the base composition of 5'SS (Supplementary Fig. 3a, b), whereas the analysis of g.4C>T revealed 1,902 introns differentially spliced ( $q < 0.1$  and absolute log2 effect size  $> 1$ ) with a significant enrichment of adenine at position 5 of the 5'SS in introns with increased excision in g.4C>T GCB-DLBCL (Fig. 2d, e, Supplementary Table 5). Differential gene expression analyses in GCB-DLBCL also revealed that the g.3A>C mutation was not associated with significant expression changes (0 genes with  $q < 0.1$  and absolute log2 fold-change  $> 1$ ; Supplementary Fig. 3c), while the g.4C>T mutation was associated with up-regulation of 47 genes and down-regulation of 634 genes in mutated tumors under the same threshold (Fig. 2f, Supplementary Table 6).

### Downstream effect of U1 mutations in BL subtypes

BL cohort ( $n = 110$ ) included 34 patients with Epstein-Barr virus (EBV)-negative and 71 EBV-positive tumors (EBV status was not available for 5 patients, Supplementary Table 2) [14]. Focusing on mutations located at the 5'SS recognition motif of U1, we observed a significant enrichment of U1 mutations in EBV-negative (11/34 mutated, 32.4%) compared to EBV-positive BL

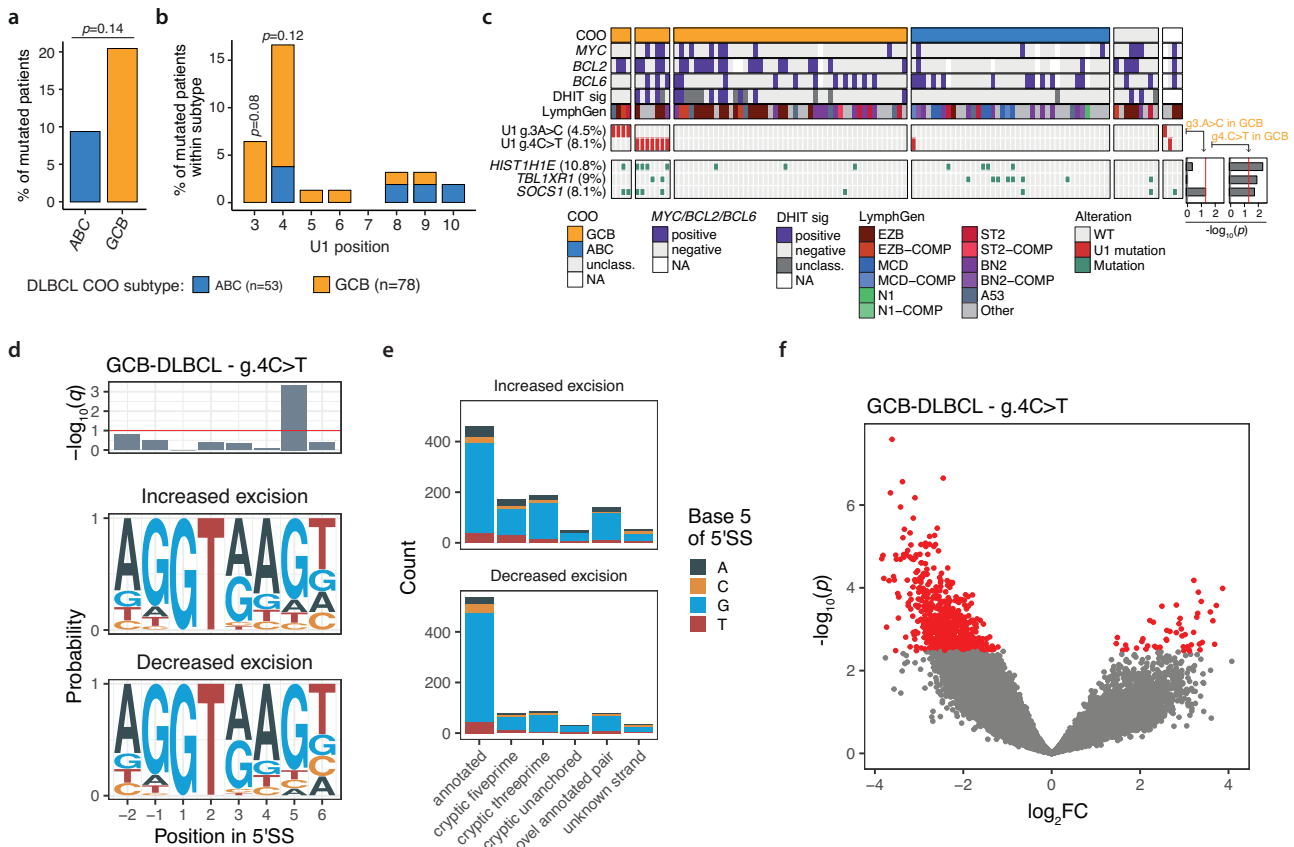


**Fig. 1** U1 mutations in 762 mature B-cell neoplasm whole genomes. **a** Representation of the U1 gene and 5' splice site (SS). **b** Schema of the bioinformatic algorithm to call U1 mutations from WGS (left) and classification grid for position 3 and tumors with 500 total number of reads in the tumor and normal sample (right). The four classification areas of the grid change depending on the position and total number of reads. Tumors with a large number of alternative (alt.) reads in the tumor and a very low number in the normal are considered to carry a somatic mutation (state 3, red area). Regions of uncertainty are represented in white. **c** Distribution of mutations along the U1 gene. The region highlighted in red (positions 3–10) corresponds to the 5'SS recognition sequence of U1. **d** Percentage of tumors carrying mutations within position 3–10 depicted according to the nucleotide change.

(5/71, 7.0%) ( $p = 0.002$ ) (Fig. 3a). Specifically, the g.7A>G mutation was detected in 10/34 (29.5%) EBV-negative compared to only 1/71 (1.4%) EBV-positive BL ( $p < 0.001$ ). Contrarily, all four g.4C>T mutations were present in EBV-positive BL (5.6% of tumors) (Fig. 3b). We observed that EBV-negative BL carrying U1 mutations usually harbored mutations in genes of the BCR/PI3K signaling pathway, while EBV-positive BL harboring U1 mutations seemed to lack mutations in epigenetic regulator genes (Fig. 3c, Supplementary Fig. 4).

In EBV-negative BL, a differential splicing analysis identified 6,970 introns differentially spliced between U1 g.7A>G mutated and wild-type tumors. The differentially spliced introns had a

significant enrichment of cytosine at position 2 of the 5'SS, which belongs to the highly conserved GT dinucleotide motif (Fig. 3d, e, Supplementary Table 7). In addition, we found 65 genes significantly differentially expressed between g.7A>G mutated and wild-type EBV-negative BL, although only 10 were protein-coding genes (Fig. 3f, Supplementary Table 8). Intriguingly, the g.4C>T mutation had no significant impact on the distribution of 5'SS in EBV-positive BL (Supplementary Fig. 5a, b). A gene expression analysis in EBV-positive BL revealed 79 genes to be significantly differentially expressed between g.4C>T mutated and wild-type tumors, with only 14/79 protein-coding genes (Supplementary Fig. 5c).



**Fig. 2** U1 mutations in DLBCL subtypes. **a** Percentage of DLBCL carrying U1 mutations (positions 3–10) according to the cell-of-origin (COO) subtypes. **b** Percentage of DLBCL carrying specific U1 mutations within each subtype. **c** Integration of U1 mutations (middle) with the main DLBCL genetic alterations and LymphGen subtypes (top) and associated driver mutations (bottom). **d** 5' splice site for introns with increased or decreased excision in GCB-DLBCL with g.4C>T U1 mutation. Top bar charts show  $q$  values from  $\chi^2$  tests for base composition difference and red lines indicate the  $q = 0.1$  cutoff. **e** Category of mis-splicing events in GCB-DLBCL with g.4C>T U1 mutation. The number of introns is colored by the fifth base of 5' splice sites. **f** Volcano plot of differential expression results for GCB-DLBCL with g.4C>T U1 mutation. Significant genes with  $q < 0.1$  and absolute  $\log_2$  fold-change  $> 1$  are in red.

### Biologic and clinical features associated with g.3A>C U1 mutations in CLL

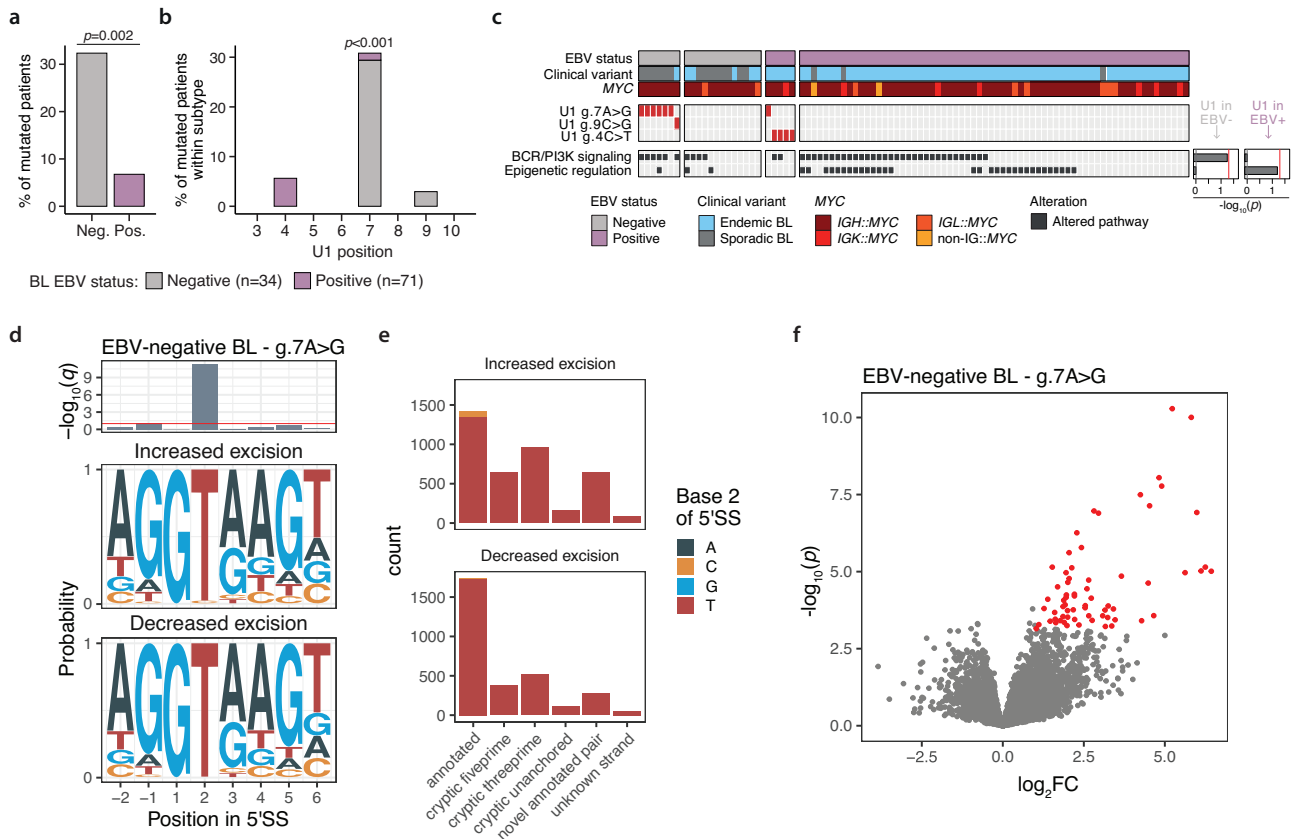
We analyzed the clinico-biological characteristics of CLL patients carrying the g.3A>C U1 mutation using two independent cohorts, C1-CLL and C2-CLL with 1120 and 550 patients respectively (1670 total). Note that C1-CLL is a representative population-based cohort while C2-CLL is biased towards specific aggressive CLL subgroups (i.e., higher percentage of U-CLL and B-cell receptor stereotypy as well as advanced clinical stages; Supplementary Table 1).

Among C1-CLL, we identified the g.3A>C mutation in 30/1120 (2.7%) tumors using a previously validated rhAMP assay (Supplementary Table 9) [1]. The frequency of this mutation was similar in unmatched samples analyzed before and after therapy [27/1,023 (2.6%) vs 3/97 (3.1%),  $p = 0.74$ , respectively] (Fig. 4a). In line with this, the analysis of sequential pair samples of 33 patients (4 mutated, 29 wildtype) analyzed before therapy and at relapse showed that the g.3A>C mutational status did not change after chemoimmunotherapy. Among patients analyzed prior to any therapy, the g.3A>C U1 mutation showed a trend for enrichment in patients diagnosed with CLL compared to monoclonal B-cell lymphocytosis (MBL) (26/850 (3.1%) vs 1/173 (0.6%),  $p = 0.069$ ). U1 mutations were equally distributed among Binet stages ( $p = 0.56$ ). All mutated tumors but one were U-CLL ( $p < 0.001$ ) and highly enriched within the naïve-like epigenetic subgroup (n-CLL,  $p < 0.001$ ) (Fig. 4b). Of note, g.3A>C U1 mutated CLL accounted for 6.4% and 7.2% of U-CLL and n-CLL subtypes, respectively. The

sole M-CLL C1-CLL carrying the g.3A>C mutation was classified as an epigenetically intermediate-CLL subtype, belonged to B-cell receptor stereotyped subset #2, and expressed the IGLV3-21 gene with the R110 mutation (IGLV3-21<sup>R110</sup>), which is known for its aggressive behavior independently of their IGHV gene somatic hypermutation (SHM) status [17–21]. With respect to B-cell receptor stereotyped subsets, the U1 mutated CLL had a trend to a higher frequency of stereotypy than U1 wild-type tumors ( $p = 0.096$ ) (Fig. 4c). To study the potential co-occurrence or mutual exclusivity of U1 mutations with known CLL driver alterations, we integrated the g.3A>C U1 with the mutational status of 28 gene and 21 copy number alterations from 693 C1-CLL patients (600 CLL and 93 MBL) [8, 22]. None of the U1 mutated tumors carried mutations in *SF3B1* or *BIRC3*, only one carried concomitant *NOTCH1* and *ATM* mutations, and two carried *TP53* mutations. In contrast, U1 mutations were statistically more likely to co-occur with *NFKB1E*, as previously described [1], and *POT1* mutations ( $q < 0.1$ ) (Fig. 4d). Of note, 4 CLL with the g.3A>C U1 mutation harbored del(13q) as a sole driver alteration, and one tumor lacked all CLL drivers analyzed. Altogether, g.3A>C U1 mutant CLL harbored a similar number of driver alterations as U1 wild-type U-CLL (Fig. 4e).

We validated these results in the C2-CLL cohort. We identified the g.3A>C mutation in 30/550 (5.5%) tumors by rhAMP assay, WGS, and/or RNA-seq. As observed in C1-CLL, this mutation was similarly found in pre- and post-treatment samples [20/412 (4.9%) and 7/89 (7.9%), respectively;  $p = 0.3$ ]. Among





**Fig. 3** U1 mutations in BL subtypes. **a** Percentage of BL carrying U1 mutation within each subtype based on EBV status. **b** Percentage of BL carrying specific U1 mutation within each subtype based on EBV status. **c** Integration of U1 mutations in BL (middle) with EBV status, clinical variant, and MYC status (top) as well as the associated mutated pathways (bottom). **d** 5' splice site for introns with increased or decreased excision in EBV-negative BL with g.7A>G U1 mutation. Top bar charts show  $q$  values from  $\chi^2$  tests for base composition difference and red lines indicate the  $q = 0.1$  cutoff. **e** Category of mis-splicing events in EBV-negative BL with g.7A>G U1 mutation. The number of introns is colored by the second base of 5' splice sites. **f** Volcano plot of differential expression results for EBV-negative BL with g.7A>G U1 mutation. Significant genes with  $q < 0.1$  and absolute  $\log_2$  fold-change  $> 1$  are in red.

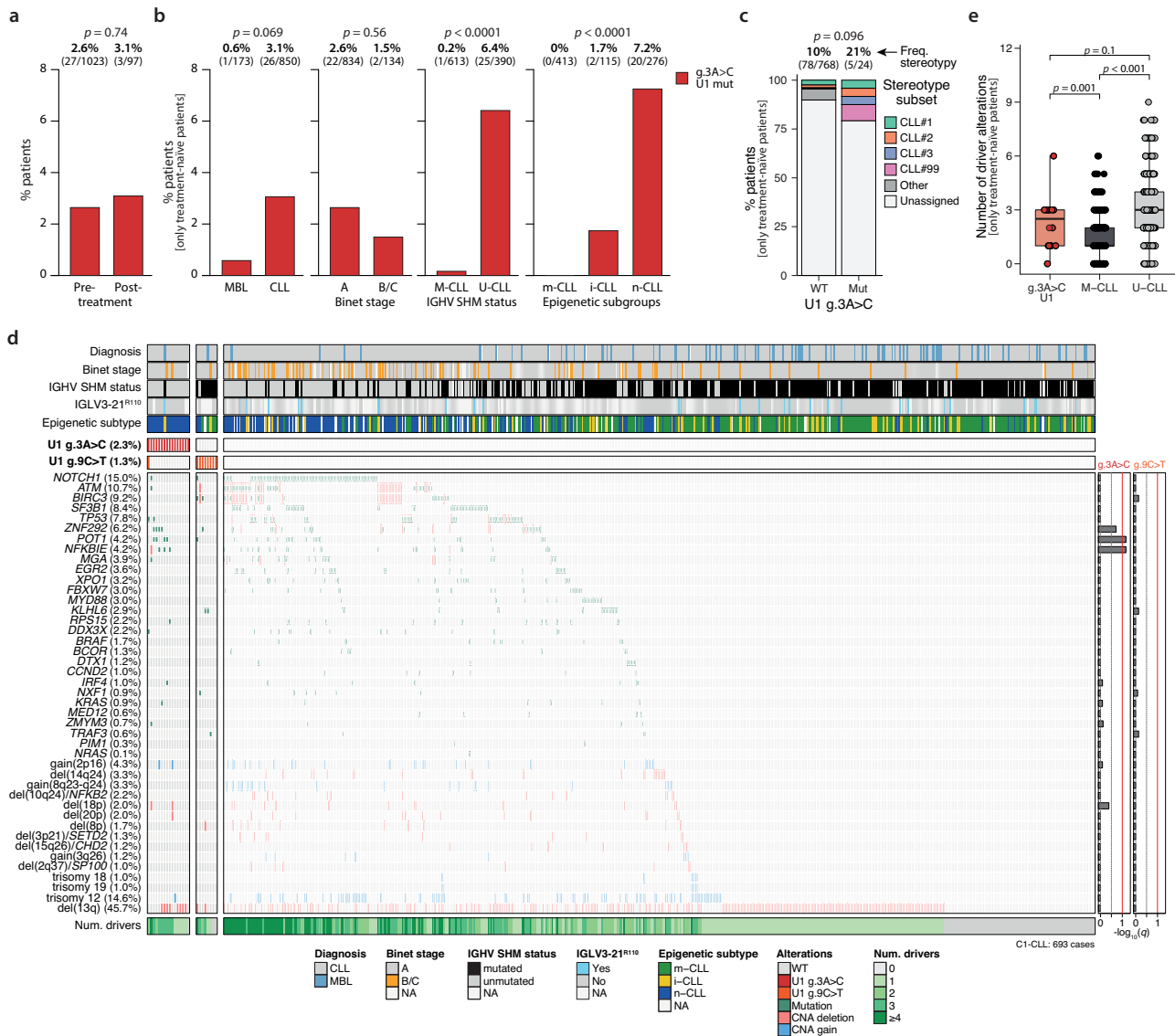
pretreatment samples, g.3A>C U1 mutations were significantly enriched in U-CLL [17/181 (9.4%) U-CLL; 3/219 (1.4%) M-CLL;  $p < 0.001$ ] and in the naïve-like CLL subgroup [5/54, (9.3%) naïve-like; 1/24 (4.2%) intermediate; 1/91 memory-like (1.1%);  $p = 0.04$ ] (Supplementary Fig. 6a). Of note, 2/3 M-IGHV carrying the g.3A>C U1 mutation belonged to subset #2, while the remaining non-stereotyped CLL carried the IGLV3-21<sup>R110</sup> mutation. Despite the higher incidence of stereotyped tumors in C2-CLL (68%), g.3A>C U1 mutated CLL also showed a higher frequency of stereotypy than U1 wild-type tumors (93% vs 61%,  $p = 0.02$ , Supplementary Fig. 6b). Similar to the results in C1-CLL, we found an association between U1 mutations and the presence of *NFKBIE* mutations ( $q < 0.1$ ) and none of the patients analyzed before therapy carried concomitant *SF3B1* and U1 mutations (Supplementary Fig. 6c). Intriguingly, among C2-CLL samples analyzed at relapse post-treatment or at an unknown treatment stage, we observed 3 CLL carrying concomitant *SF3B1* and U1 mutations (Supplementary Fig. 6d). However, we could not confirm that both mutations co-occurred in the same CLL cells based on their clonality. Overall, similar clinical and genetic associations were observed when both cohorts were analyzed together (Supplementary Table 10).

Using the two cohorts, we validated the previously identified effect on expression and mis-splicing caused by the g.3A>C U1 mutation [1] in 468 C1-CLL patients analyzed by microarray-based gene expression profiling [8] and 75 C2-CLL cases by RNA-seq (Supplementary Results, Supplementary Figs. 7, 8, Supplementary Tables 11–14).

### g.9C>T U1 mutations in CLL

We found a recurrent g.9C>T mutation in the WGS of 7 CLL (Fig. 1d). To further characterize this mutation, we first genotyped it using a custom rhAMP assay in 1,051 C1-CLL tumors (note that they included the 152 C1-CLL studied by WGS as controls; 3g.9C>T mutated) (Supplementary Table 9). Altogether, we identified 14/1,051 (1.4%) CLL carrying the g.9C>T mutation. In contrast to the g.3A>C mutation, the g.9C>T was enriched in M-CLL [12/615 (2%) M-CLL vs 2/417 (0.5%) U-CLL,  $p = 0.055$ ]. A similar trend was observed regarding the epigenetic subtypes [7/432 (1.6%) memory-like, 1/119 (0.84%) intermediate, and 3/300 (1%) naïve-like CLL]. Once integrated with the driver genomic landscape of CLL (Fig. 4d), the g.9C>T mutation did not significantly co-occur nor was mutually exclusive with any of the studied CLL driver alterations. Only one tumor carried both g.3A>C and g.9C>T U1 mutations, and none of the g.9C>T CLL carried *SF3B1* mutations (Fig. 4d). Of note, two CLL carried the g.9C>T mutation as the sole putative driver alteration, while in two additional tumors, g.9C>T was accompanied by a single *TRAF3* or *KLHL6* mutation, respectively (Fig. 4d).

We next performed a differential splicing analysis and identified 505 differentially spliced introns between g.9C>T ( $n = 4$ ) vs wild-type ( $n = 288$ ) U1 CLL (Fig. 5a, b, Supplementary Table 15). Despite the small sample size, we observed, as expected, significant differences in the -1 position of the 5'SS in which introns with increased excision in g.9C>T U1 mutated CLL were enriched for adenine (Fig. 5a), specifically enriched in cryptic 5'SS and annotated junctions (Fig. 5b). To confirm the impact of g.9C>T



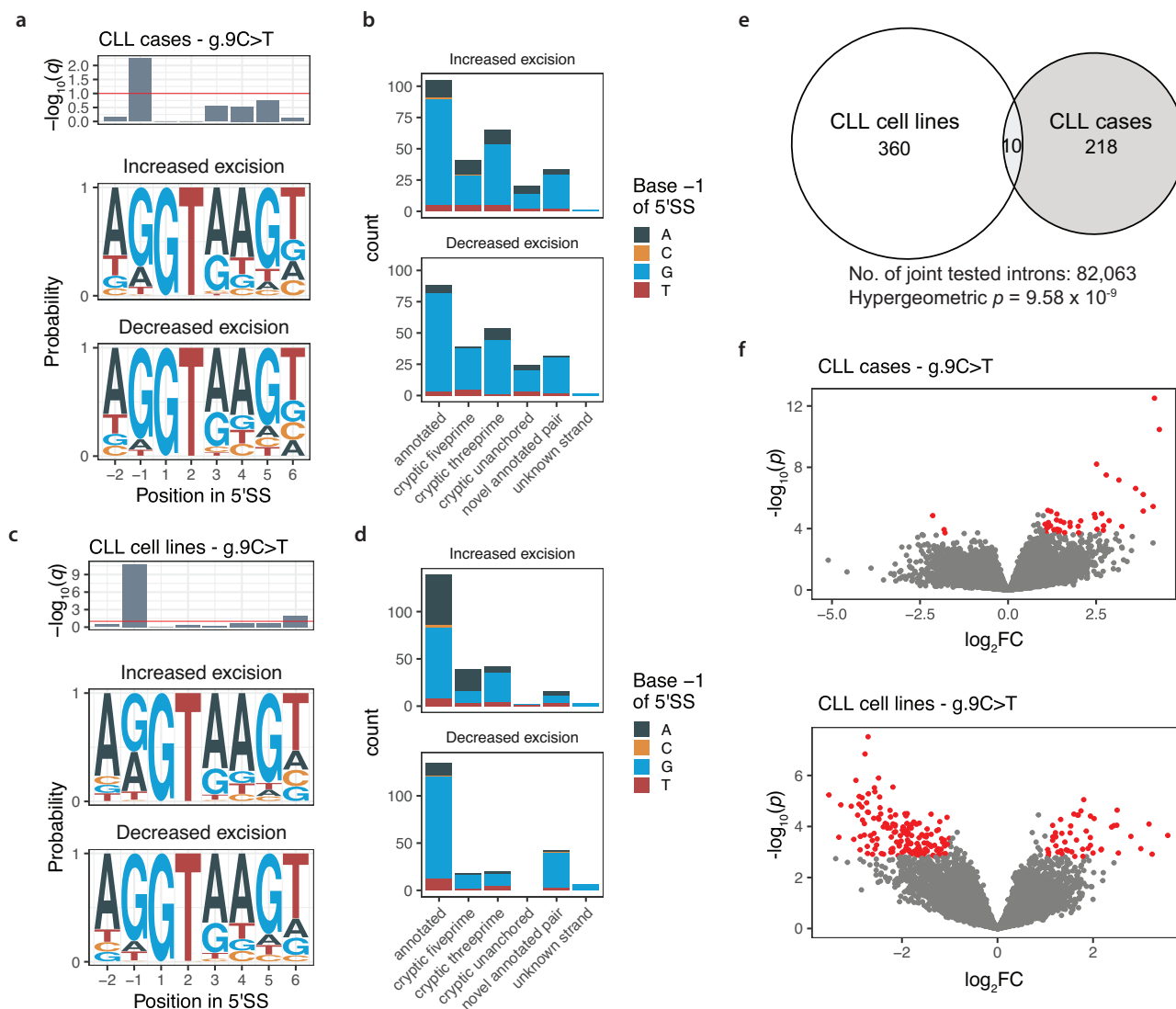
**Fig. 4** Clinic and biological characteristics of the 1,120 C1-CLL patients according to the presence of the g.3A>C U1 mutation. **a** Bar plots showing the distribution of U1 g.3A>C mutated CLL among samples analyzed before and after the initiation of therapy. **b** Bar plots showing the distribution of U1 g.3A>C mutated tumors according to diagnosis (MBL vs CLL), Binet stages, IGHV gene SHM status, and epigenetic subtypes. m-CLL, memory-like CLL; i-CLL, intermediate CLL; n-CLL, naïve like CLL. **c** Stereotyped subsets in U1 wild-type and mutated CLL. **d** Oncoprint showing the co-occurrence of g.3A>C and g.9C>T U1 mutations with known CLL driver alterations and clinic-biological variables. The total number of previously recognized driver alterations (Num. drivers) is shown. The bar plot on the right represents the  $q$  value of the two-sided Fisher's exact tests applied to study the co-occurrence/independence of g.3A>C (left) and g.9C>T (right) U1 mutations with the known driver alterations. **e** Distribution of known driver alterations in CLL with or without the g.3A>C mutation stratified based on their IGHV gene SHM status.

on splicing, we introduced exogenous U1 genes with or without the g.9C>T in three CLL cell lines (JVM3, HG3, and MEC1). Although we did not observe differences in proliferation, viability, or apoptosis in g.9C>T vs wild-type cell lines, the same transcriptome analysis using cell-line RNA-seq data confirmed the downstream effect of the g.9C>T mutation observed in primary tumors (462 differentially spliced introns; Fig. 5c, d, Supplementary Table 16). We also found more than the expected number of overlaps between differentially spliced introns in primary CLL samples and cell lines (hypergeometric test  $p = 9.58 \times 10^{-9}$ ; Fig. 5e), despite the fact that different sequencing technologies were used for the two transcriptome experiments (poly-A and total RNA-seq, respectively). In contrast with the g.3A>C mutation, which results in the differential expression of several hundred genes in primary CLL and cell lines, differential

gene expression analyses in g.9C>T CLL revealed that this mutation had a smaller impact on gene expression since only 45 and 199 genes were significantly differentially expressed in primary CLL and cell lines, respectively, with only one gene in common (Fig. 5f, Supplementary Tables 17, 18).

#### Clinical impact of U1 mutations in CLL

In the C1-CLL, the g.3A>C U1 mutation was associated with a shorter TTFT in univariate analysis across the entire CLL cohort as well as in an analysis restricted to early stage (Binet A) patients ( $p = 0.008$  and  $p = 0.003$ , respectively) (Supplementary Fig. 9a, b). Despite this, the g.3A>C U1 mutation was not associated with a shortened overall survival (OS) ( $p = 0.59$ ) (Supplementary Fig. 9c). We next stratified patients according to their g.3A>C status, IGHV gene SHM status, and presence of *TP53*, *SF3B1*, *NOTCH1*, and/or



**Fig. 5** Downstream effect of g.9C>T U1 mutations in primary CLL and cell lines. **a** 5' splice site for introns with increased or decreased excision in C1-CLL tumors with g.9C>T U1 mutation. **b** Category of mis-splicing events in C1-CLL with g.9C>T U1 mutation. The number of introns is colored by the -1 base of 5' splice sites. **c** 5' splice site for introns with increased or decreased excision in CLL cell lines with g.9C>T U1 mutation. **d** Category of mis-splicing events in CLL cell lines with g.9C>T U1 mutation. For **(a)** and **(c)**, top bar charts show  $q$  values from  $\chi^2$  tests for base composition difference and red lines indicate the  $q = 0.1$  cutoff. For **(b)** and **(d)**, the number of introns is colored by the -1 base of 5' splice sites. **e** Euler plot showing the overlap of introns that are differentially spliced in primary CLL tumors and cell lines. Only introns tested in both datasets ( $n = 82,063$ ) are included. **f** Volcano plots of differential expression results for CLL tumors and cell lines with g.9C>T U1 mutation. Significant genes with  $q < 0.1$  and absolute  $\log_2$  fold-change  $> 1$  are in red.

*ATM* alterations. Patients carrying the g.3A>C mutation had a similar TTFT to those carrying *TP53*, *SF3B1*, *NOTCH1*, and/or *ATM* alterations (Fig. 6a). A multivariable model including these variables together with IGHV gene SHM status and disease stage showed that the g.3A>C U1 mutation was independently associated with a shorter TTFT ( $p = 0.005$ , Fig. 6b). We also observed a significant prognostic impact of the g.3A>C U1 mutation independently of the number of driver alterations ( $p = 0.02$ ), a surrogate of the genomic complexity of the tumors (Fig. 6c).

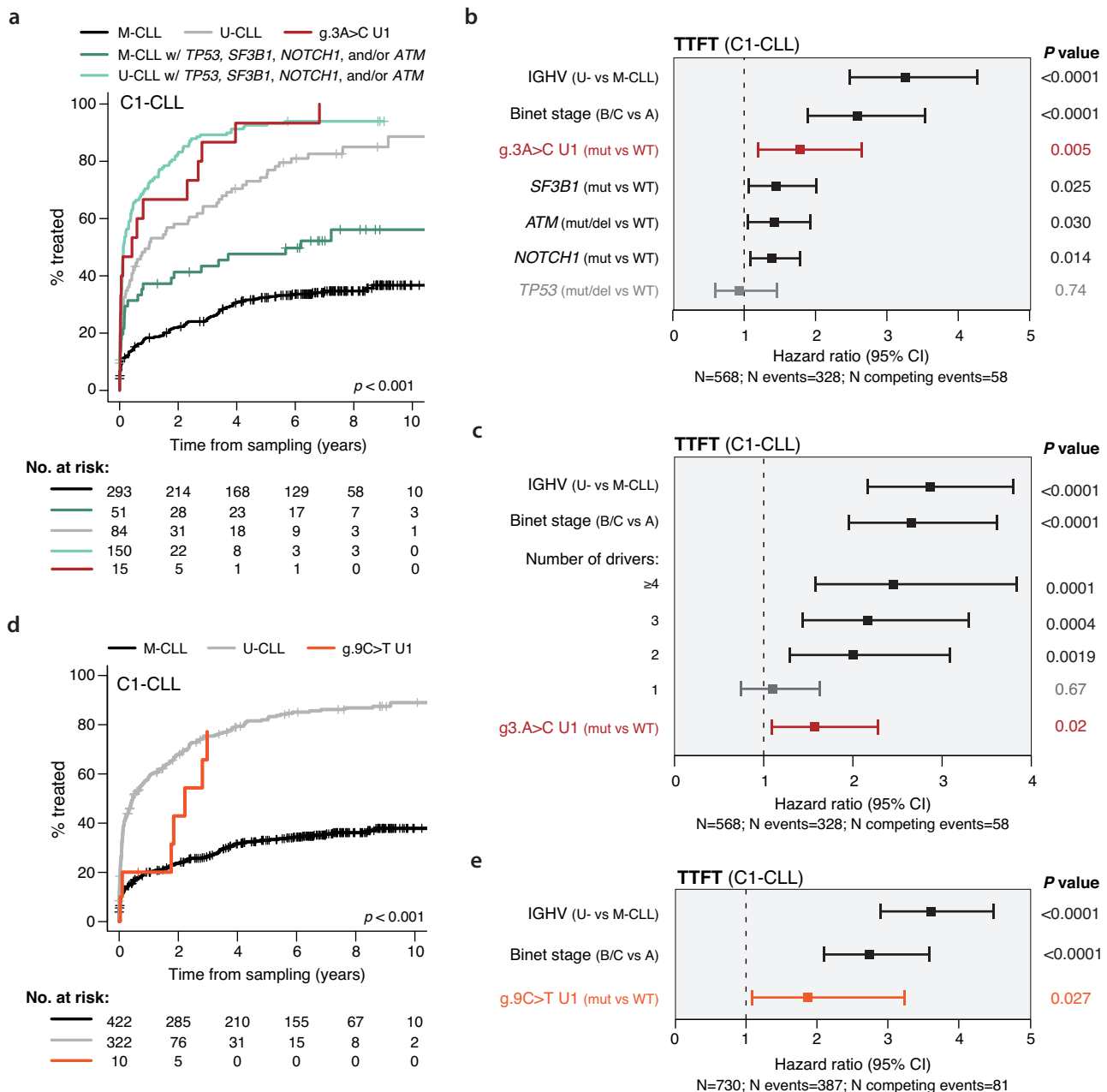
In the more aggressive C2-CLL, the g.3A>C U1 mutation showed a trend for association with a shorter TTFT in a univariate analysis ( $p = 0.057$ , 389 treatment-naïve patients at time of sampling and with available clinical data included in the analysis) (Supplementary Fig. 10a). However, this mutation did not retain independent prognostic value in a multivariate model including the IGHV gene SHM status, Binet stage, and main driver alterations (*TP53*, *SF3B1*,

*NOTCH1*, and *ATM*). Note that only *TP53* alterations, but not *SF3B1*, *NOTCH1*, nor *ATM* aberrations had independent prognostic value in this cohort (Supplementary Fig. 10b, c).

Regarding the g.9C>T mutation, we observed a shorter TTFT in the C1-CLL cohort that was independent of the IGHV gene SHM status and Binet stage ( $p = 0.027$ , Fig. 6d, e, Supplementary Fig. 11a, b). Nonetheless, its independent prognostic value was not confirmed when considering other driver alterations (*TP53*, *SF3B1*, *NOTCH1*, *ATM*, g.3A>C U1), which is probably due to the small number of patients carrying this g.9C>T U1 mutation (Supplementary Fig. 11c).

## DISCUSSION

We previously identified a novel noncoding recurrent A>C mutation affecting the third base of the U1 spliceosomal RNA (g.3A>C) in distinct cancer types including CLL, in which the



**Fig. 6 Clinical implications of U1 mutations in CLL.** **a** TTFT curves of C1-CLL patients according to the presence of g.3A>C U1 mutation, *TP53*, *SF3B1*, *NOTCH1* and/or *ATM* alterations, and IGHV gene SHM status (U-CLL, M-CLL). **b** Forest plot showing a multivariable model for TTFT including the IGHV gene SHM status, Binet stage, presence of g.3A>C U1, *TP53*, *NOTCH1*, *ATM* and *SF3B1* alterations. **c** Multivariable model for TTFT including IGHV gene SHM status, Binet stage, number of previously recognized driver alterations, and g.3A>C U1 mutation. **d** TTFT curves of C1-CLL patients according to the presence of g.9C>T U1 mutations and IGHV gene SHM status. **e** Multivariable model of TTFT including IGHV gene SHM status, Binet stage, and g.9C>T U1 mutations.

mutation was observed in 3.8% patients and caused specific downstream transcriptomic changes [1, 2]. Here we provided a comprehensive characterization of U1 mutations in 1670 patients with CLL to better understand their biological and clinical consequences. Furthermore, we investigated the presence and effect of U1 mutations in other mature B-cell neoplasms by studying 363 tumor whole genomes.

Reanalysis of the whole genome of 399 CLL led to the identification of a new recurrent mutation in position 9 of the U1 gene (g.9C>T), which caused significant downstream splicing, but not expression, changes in primary samples and transduced CLL cell lines. Although the incidence of the g.9C>T U1 mutation

was relatively low (1.4%), this mutation defined a subgroup (2%) of M-CLL with a short TTFT, thus indicating the heterogeneity of this subgroup of patients while also highlighting the need to further dissect its biological background. This WGS-based analysis also confirmed that g.3A>C is the most frequent U1 mutation in CLL [1]. After studying 1670 CLL from two independent cohorts, the g.3A>C mutation was detected in 60 (3.6%) CLL confirming its significant enrichment within the U-CLL (7.4%) and naïve-like epigenetic (n-CLL, 10.9%) subtypes, and not significantly co-occurring with several CLL driver alterations of prognostic relevance. We did, however, confirm a significant association of these mutations with alterations in the less-common driver gene



*NFKBIE* [1] in both cohorts analyzed and found, for the first time, a significant association with *POT1*. Although enriched in U-CLL/n-CLL, four g.3A>C mutated tumors were classified as M-CLL. All four of these CLL belonged to stereotyped subset #2 and/or expressed IGLV3-21<sup>R110</sup>, known markers of aggressive disease independent of their IGHV gene SHM status [19–21, 23–25]. Hence, all g.3A>C U1 mutations were identified in aggressive forms of CLL. The presence of the g.3A>C mutation did not impact the OS but was associated with shorter TTFT independently of the disease stage, IGHV gene SHM status, and presence of previously recognized driver alterations in the C1-CLL. The occurrence of the g.3A>C and g.9C>T mutations in treatment-naïve patients and its association with an earlier need for therapy supports the driver role of U1 mutations in CLL. This finding expands the central role of RNA metabolism in CLL [26].

WGS analyses of patients with DLBCL, BL, MCL, and FL revealed that 5'SS U1 mutations are common in DLBCL and BL but seem uncommon in MCL and FL. Striking differences were observed regarding the mutations found in each entity and their effect on splicing. The g.3A>C mutation and a novel g.4C>T mutation were identified in 4.5% and 8.4% of patients with DLBCL, respectively, with a trend towards a higher incidence in GCB-DLBCL (6.4% and 10.3%, respectively) compared to ABC-DLBCL (0% and 3.8%, respectively). In contrast to CLL, the g.3A>C was not associated with alterations in the distribution of 5'SS in GCB-DLBCL tumors. A different picture emerged for the g.4C>T mutation where we observed significant splicing and expression changes associated with this mutation. In BL, the g.4C>T mutation was also observed in 3.6% of the patients. However, the most frequent recurrent mutation was a novel g.7A>G mutation found in 10.9% BL. All g.4C>T mutations were found in EBV-positive BL (5.6%), whereas g.7A>G mutations were virtually exclusive of EBV-negative BL and accounted for 29.5% of the patients. The g.7A>G mutation was associated with significant changes in the splicing pattern of BL. This mutation seems very specific to BL because it was not seen in any other type of cancer in the PCAWG cohort [1]. Overall, these results uncover novel U1 mutations in mature B-cell lymphomas of germinal center origin revealing entity- and sub-entity-specific patterns of mutations and downstream changes in splicing and expression of specific genes. These differences could be attributed to the distinct sets of genes expressed in each entity or sub-entity, which serve as substrate upon which U1 mutations influence their splicing and expression. Therefore, similar to *SF3B1* mutations [27], one could speculate that the same U1 mutation could be associated with different biological and clinical features in different diseases. The high frequency of U1 mutations in the different subsets of tumors suggest that U1 mutations may play a driving role in DLBCL and BL [14, 28–30]. Nonetheless, further analyses with larger series are needed to completely characterize their biological and clinical consequences.

In summary, this study expands our understanding of the biological and clinical consequences of U1 mutations in CLL and reveals novel U1 mutations in DLBCL and BL. U1 mutations and their downstream effects were specific to different entities and subtypes with distinct mutations enriched in CLL (g.3A>C and g.9C>T), DLBCL (g.4C>T), and BL (g.7A>G). Based on its downstream effects, mutation prevalence among distinct B-cell neoplasms, and prognostic value in CLL, U1 represents a new pan-B-cell malignancy driver gene.

## DATA AVAILABILITY

Data used in this work were obtained from previously published sources, including, among others, data downloaded from the dbGaP web site (phs000235.v15.p3), which are cited in the Methods section and Supplementary Methods.

## REFERENCES

- Shuai S, Suzuki H, Diaz-Navarro A, Nadeu F, Kumar SA, Gutierrez-Fernandez A, et al. The U1 spliceosomal RNA is recurrently mutated in multiple cancers. *Nature*. 2019;574:712–6.
- Suzuki H, Kumar SA, Shuai S, Diaz-Navarro A, Gutierrez-Fernandez A, De Antonellis P, et al. Recurrent noncoding U1 snRNA mutations drive cryptic splicing in SHH medulloblastoma. *Nature*. 2019;574:707–11.
- 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.
- 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.
- 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.
- Kulis M, Heath S, Bibikova M, Queirós AC, Navarro A, Clot G, et al. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat Genet*. 2012;44:1236–42.
- Oakes CC, Seifert M, Assenov Y, Gu L, Przekopowicz M, Ruppert AS, et al. DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat Genet*. 2016;48:253–64.
- Puente XS, Beà S, Valdés-Mas R, Villamor N, Gutiérrez-Abriol J, Martín-Subero JI, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2015;526:519–24.
- Dietrich S, Oleś M, Lu J, Sellner L, Anders S, Velten B, et al. Drug-perturbation-based stratification of blood cancer. *J Clin Invest*. 2018;128:427–45.
- 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.
- Morin RD, Mungall K, Pleasance E, Mungall AJ, Goya R, Huff RD, et al. Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. *Blood*. 2013;122:1256–65.
- Arthur SE, Jiang A, Grande BM, Alcaide M, Cojocar R, Rushton CK, et al. Genome-wide discovery of somatic regulatory variants in diffuse large B-cell lymphoma. *Nat Commun*. 2018;9:4001.
- ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature*. 2020;578:82–93.
- Grande BM, Gerhard DS, Jiang A, Griner NB, Abramson JS, Alexander TB, et al. Genome-wide discovery of somatic coding and noncoding mutations in pediatric endemic and sporadic Burkitt lymphoma. *Blood*. 2019;133:1313–24.
- Nadeu F, Martín-García D, Clot G, Díaz-Navarro A, Duran-Ferrer M, Navarro A, et al. Genomic and epigenomic insights into the origin, pathogenesis, and clinical behavior of mantle cell lymphoma subtypes. *Blood*. 2020;136:1419–32.
- Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000;403:503–11.
- Tobin G, Thunberg U, Johnson A, Thorn I, Soderberg O, Hultdin M, et al. Somatic mutations in Ig VH3-21 genes characterize a new subset of chronic lymphocytic leukemia. *Blood*. 2002;99:2262–4.
- Stamatopoulos K, Belessi C, Moreno C, Boudjograh M, Guida G, Smilevska T, et al. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: pathogenetic implications and clinical correlations. *Blood*. 2007;109:259–70.
- Maity PC, Bilal M, Koning MT, Young M, van Bergen CAM, Renna V, et al. IGLV3-21\*01 is an inherited risk factor for CLL through the acquisition of a single-point mutation enabling autonomous BCR signaling. *Proc Natl Acad Sci USA*. 2020;117:4320–7.
- Nadeu F, Royo R, Clot G, Duran-Ferrer M, Navarro A, Martín S, et al. IGLV3-21R110 identifies an aggressive biological subtype of chronic lymphocytic leukemia with intermediate epigenetics. *Blood*. 2021;137:2935–46.
- Syrykh C, Pons-Brun B, Russiñol N, Playa-Albinyana H, Baumann T, Duran-Ferrer M, et al. IGLV3-21R110 mutation has prognostic value in patients with treatment-naïve chronic lymphocytic leukemia. *Blood Adv*. 2023;7:7384–91.
- Nadeu F, Clot G, Delgado J, Martín-García D, Baumann T, Salaverria I, et al. Clinical impact of the subclonal architecture and mutational complexity in chronic lymphocytic leukemia. *Leukemia*. 2018;32:645–53.
- Strefford JC, Sutton L-A, Baliakas P, Agathangelidis A, Malčíková J, Plevova K, et al. Distinct patterns of novel gene mutations in poor-prognostic stereotyped subsets of chronic lymphocytic leukemia: the case of SF3B1 and subset #2. *Leukemia*. 2013;27:2196–9.
- Sutton L-A, Young E, Baliakas P, Hadzidimitriou A, Moysiadi T, Plevova K, et al. Different spectra of recurrent gene mutations in subsets of chronic lymphocytic leukemia harboring stereotyped B-cell receptors. *Haematologica*. 2016;101:959–67.
- Stamatopoulos B, Smith T, Crompot E, Pieters K, Clifford R, Mraz M, et al. The light chain IGLV3-21 defines a new poor prognostic subgroup in chronic lymphocytic leukemia: results of a multicenter study. *Clin Cancer Res*. 2018;24:5048–57.

26. Ramsay AJ, Rodríguez D, Villamor N, Kwarcia A, Tejedor JR, Valcárcel J, et al. Frequent somatic mutations in components of the RNA processing machinery in chronic lymphocytic leukemia. *Leukemia*. 2013;27:1600–3.
27. Wan Y, Wu CJ. SF3B1 mutations in chronic lymphocytic leukemia. *Blood*. 2013;121:4627–34.
28. Schmitz R, Wright GW, Huang DW, Johnson CA, Phelan JD, Wang JQ, et al. Genetics and pathogenesis of diffuse large B-cell lymphoma. *N Engl J Med*. 2018;378:1396–407.
29. Chapuy B, Stewart C, Dunford AJ, Kim J, Kamburov A, Redd RA, et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med*. 2018;24:679–90.
30. López C, Kleinheinz K, Aukema SM, Rohde M, Bernhart SH, Hübschmann D, et al. Genomic and transcriptomic changes complement each other in the pathogenesis of sporadic Burkitt lymphoma. *Nat Commun*. 2019;10:1459.

## ACKNOWLEDGEMENTS

The authors are indebted to the Genomics Core Facility of the Fundació de Recerca Clínic Barcelona-Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS) for the technical support as well as the Hematopathology Collection registered at the Biobank of the Hospital Clínic - IDIBAPS for sample procurement. This study was supported by the “la Caixa” Foundation (CLLEvolution - LCF/PR/HR17/52150017 [HR17-00221LCF] and CLLSYSTEMS - LCF/PR/HR22/52420015 [HR22-00172] Health Research 2017 and 2022 Programs, to ECa), the Ministry of Science and Innovation (MCIN)/AEI/10.13039/501100011033 and the European Regional Development Fund (FEDER) “Una manera de hacer Europa” (PID2021-123054OB-I00, to ECa), the National Institute of Health “Molecular Diagnosis, Prognosis, and Therapeutic Targets in Mantle Cell Lymphoma” (P01CA229100 to ECa), Generalitat de Catalunya Suport Grups de Recerca AGAUR (2017-SGR-1142 and 2021-SGR-01172 to ECa), and CERCA Programme-Generalitat de Catalunya (to ECa). FN acknowledges research support from 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), the Lady Tata Memorial Trust (International Award for Research in Leukaemia 2021-2022, LADY\_TATA\_21\_3223), and Fundación Española de Hematología y Hemoterapia (FEHH) (FEHH\_BECA524\_01 co-financed by AstraZeneca). SS was supported by the National Natural Science Foundation of China (Grant No. 82270239 and 32200487), Shenzhen Medical Research Funds (Grant No. A2303044), the Shenzhen Municipal Science and Technology Innovation Council (Grant No. JCYJ20220530115207016), and Center for Computational Science and Engineering at Southern University of Science and Technology (Shenzhen, China). AD-N is supported by the Department of Education of the Basque Government (grant number PRE\_2017\_1\_0100). AR-D was supported by the Fundación Española de Hematología y Hemoterapia (FEHH). KP, KStr, MD, and SP acknowledge financial support from the Ministry of Health of the Czech Republic under the project of the conceptual development of research organization FNBr 65269705 and the project AZV NU21-08-00237, further, they acknowledge support from the Ministry of Education, Youth and Sports of the Czech Republic within the European Regional Development Fund Project “A-C-G-T” No. CZ.02.1.01/0.0/0.0/16\_026/0008448 and the National Institute for Cancer Research, Programme EXCELES, No. LX22NPO5102, co-funded by the European Union – Next Generation EU. GGa acknowledges support from AIRC 5 ×1000 project No. 21198, Milan, Italy. PG acknowledges support from Associazione Italiana per la ricerca sul cancro (AIRC) under 5 per Mille 2018—ID. 21198 program—P.I. Roberto Foà, G.L. Paolo Ghia, G. RRoS is supported by the Science for Life Laboratory (SciLifeLab), the Swedish Cancer Society, the Swedish Research Council, the Knut and Alice Wallenberg Foundation, Karolinska Institutet, Karolinska University Hospital, and Radiumhemmets Forskningsfonder, Stockholm, Sweden. For patients analyzed at Karolinska Institutet, sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure (NGI) Sweden and SciLifeLab. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. SAP, KGR, and NEK acknowledge that the conduct of the Mayo Clinic research was supported in part by the Henry J. Predolin Foundation. XSP is supported by Ministerio de Ciencia e Innovación (PID2020-117185RB-I00) and Fundación Asociación Española Contra el Cáncer (AECC). ECa is an Academia Researcher of the “Institució Catalana de Recerca i Estudis Avançats” (ICREA) of the Generalitat de Catalunya. This work was partially developed at the Centre Esther Koplowitz (CEK, Barcelona, Spain).

## AUTHOR CONTRIBUTIONS

FN designed the study, collected and analyzed data, interpreted data, wrote the manuscript, and supervised the research. SS analyzed RNA-seq data, interpreted data, and wrote the manuscript. GC implemented the statistical framework for UI

mutational calling from WGS, analyzed and interpreted data, and wrote the manuscript. LKH, RRoY, and RDM analyzed and interpreted data. AD-N, PB, and XSP conducted in vitro analyses, and interpreted data. SM performed and interpreted rhAMP experiments. TB, JL, VL, EY, KP, BAK, ZL, CKH, May, MG, ECo, ARP, MA, MJT, AR-D, TC, DH, CJ, AS, LM, KStr, MD, EJG-M, ZD, RW, LS, LT, AV, SAP, KGR, RM, Mar, DR, FD, GGa, NEK, TS, PG, DO, AWL, SB, AL-G, DN, CJW, GGe, SP, KSta, RRoS, WH, TZ, JIM-S, JD, and RDM provided samples and/or data. MK, IL-O, MC, AE, and DC performed and interpreted experiments. SR-G provided logistical assistance. XSP and LDS interpreted data, and contributed to the conception of the study. ECa designed the study, interpreted data, wrote the manuscript, and supervised the research. All authors read, commented on, and approved the manuscript.

## COMPETING INTERESTS

FN received honoraria from AbbVie, AstraZeneca, Janssen, and SOPHiA GENETICS for speaking at educational activities, and research funding from Gilead. LS received honoraria from AbbVie, AstraZeneca, BeiGene, Johnson&Johnson, Lilly, and Merck. AWL received research support and/or speaker fee from Johnson & Johnson and Roche-Genentech. PG received honoraria from AbbVie, ArQule/MSD, AstraZeneca, BeiGene, Celgene/Juno/BMS, Janssen, Lilly/LoxoSanofi and Roche, and research support from AbbVie, AstraZeneca, and Janssen. RRoS received honoraria from AbbVie, AstraZeneca, Illumina, Janssen, and Roche. DN has stock ownership in Madrigal Pharmaceuticals. GGa has received honoraria from AbbVie, AstraZeneca, Beigene, Hikma, Incyte, Johnson & Johnson, Lilly for speaking at educational activities and advisory boards. CJW holds equity in BioNTech, Inc. and receives research support from Pharmacyclics. GGe receives research funds from IBM and Pharmacyclics; is an inventor on patent applications related to MuTect, ABSOLUTE, MutSig, MSMutSig, MSIDetect, POLYSOLVER and TensorQTL; and is a founder and consultant of and holds privately held equity in Scorpion Therapeutics. LDS received support from the Province of Ontario and the Canadian Institutes for Health Research, grant #202003PJT-438161-CBT-OICR-168885. XSP is a cofounder of and holds an equity stake in DREAMgenics, and received honoraria from AstraZeneca and BeiGene for speaking at educational activities. EC has been a consultant for Takeda; received honoraria from Janssen, ThermoFisher and Roche for speaking at educational activities; and is an inventor on a Lymphoma and Leukemia Molecular Profiling Project patent “Method for subtyping lymphoma subtypes by means of expression profiling” (PCT/US2014/64161) not related to this project. FN and EC licensed the use of the protected IgCaller algorithm for SOPHiA GENETICS and Diagnóstica Longwood. The other authors have no competing interests to declare.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41375-025-02667-7>.

**Correspondence** and requests for materials should be addressed to Ferran Nadeu or Elias Campo.

**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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

<sup>1</sup>Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain. <sup>2</sup>Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Barcelona, Spain. <sup>3</sup>Department of Human Cell Biology and Genetics, School of Medicine, Southern University of Science and Technology, Shenzhen, China. <sup>4</sup>Facultat de Medicina i Ciències de la Salut, Universitat de Barcelona, Barcelona, Spain. <sup>5</sup>Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada. <sup>6</sup>Departamento de Bioquímica y Biología Molecular, Instituto Universitario de Oncología, Universidad de Oviedo, Oviedo, Spain. <sup>7</sup>Barcelona Supercomputing Center (BSC), Barcelona, Spain. <sup>8</sup>Hospital Clínic de Barcelona, Barcelona, Spain. <sup>9</sup>Medical Faculty Heidelberg, Heidelberg University, Heidelberg, Germany. <sup>10</sup>Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden. <sup>11</sup>Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden. <sup>12</sup>Department of Internal Medicine - Hematology and Oncology, Faculty of Medicine, Masaryk University and University Hospital Brno, Brno, Czech Republic. <sup>13</sup>Institute of Medical Genetics and Genomics, Faculty of Medicine, Masaryk University and University Hospital Brno, Brno, Czech Republic. <sup>14</sup>Central European Institute of Technology (CEITEC), Masaryk University, Brno, Czech Republic. <sup>15</sup>The Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>16</sup>The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan, Israel. <sup>17</sup>Harvard University, Cambridge, MA, USA. <sup>18</sup>Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. <sup>19</sup>Biología Molecular e Histocompatibilidad, IBSAL-Hospital Universitario, Centro de Investigación del Cáncer-IBMCC (USAL-CSIC), Salamanca, Spain. <sup>20</sup>Servicio de Hematología y Hemoterapia, Hospital Universitario Central de Asturias; Instituto Universitario de Oncología del Principado de Asturias (IUOPA); and Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), Oviedo, Spain. <sup>21</sup>Servicio de Hematología, Hospital Clínico Universitario, INCLIVA, Universidad de Valencia, Valencia, Spain. <sup>22</sup>Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece. <sup>23</sup>Hematology Department and HCT Unit, G. Papanicolaou Hospital, Thessaloniki, Greece. <sup>24</sup>Clinical Genetics and Genomics, Karolinska University Hospital, Stockholm, Sweden. <sup>25</sup>Department of Immunology, Erasmus MC, University Medical Center, Rotterdam, Netherlands. <sup>26</sup>Department of Molecular Pathology, University Hospitals Dorset, Bournemouth, UK. <sup>27</sup>Department of Haematology, University Hospitals Dorset, Bournemouth, UK. <sup>28</sup>Università Vita Salute San Raffaele, Milano, Italy. <sup>29</sup>Strategic Research Program on CLL, IRCCS San Raffaele Hospital, Milano, Italy. <sup>30</sup>Department of Medicine, Hematology unit, University of Padua, Padua, Italy. <sup>31</sup>Division of Hematology, Mayo Clinic, Rochester, MN, USA. <sup>32</sup>Department of Quantitative Health Sciences, Division of Clinical Trials and Biostatistics, Mayo Clinic, Rochester, MN, USA. <sup>33</sup>Division of Hematology, Department of Translational Medicine, University of Eastern Piedmont, Novara, Italy. <sup>34</sup>Department of Laboratory Medicine and Pathology, Hospital Pitie-Salpetriere, Sorbonne University, Paris, France. <sup>35</sup>Department of Hematology, Oncology Institute of Southern Switzerland and Institute of Oncology Research, Bellinzona, Switzerland. <sup>36</sup>Stanford School of Medicine, Stanford, CA, USA. <sup>37</sup>Department of Data Science, Dana-Farber Cancer Institute, Boston, MA, USA. <sup>38</sup>Harvard Medical School, Boston, MA, USA. <sup>39</sup>Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. <sup>40</sup>Center for Cancer Research, Massachusetts General Hospital, Boston, MA, USA. <sup>41</sup>Department of Pathology, Massachusetts General Hospital, Boston, MA, USA. <sup>42</sup>European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. <sup>43</sup>Department of Medical Oncology and Hematology, University Hospital and University of Zurich, Zurich, Switzerland. <sup>44</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. <sup>45</sup>Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, BC, Canada. <sup>46</sup>Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada. <sup>47</sup>Computational Biology Program, Ontario Institute for Cancer Research, Toronto, ON, Canada. <sup>48</sup>Present address: Instituto de Investigación Biomédica de Salamanca (IBSAL), Salamanca, Spain. <sup>49</sup>Present address: Hospital Universitario 12 de Octubre, Madrid, Spain. <sup>50</sup>These authors contributed equally: Ferran Nadeu, Shimin Shuai, Guillem Clot. <sup>51</sup>email: nadeu@recerca.clinic.cat; ecampo@clinic.cat.