haematologica
Journal of the Ferrata Storti Foundation

The potential role of intraclonal diversification of immunoglobulin light chain genes in chronic lymphocytic leukemia prognostication

by Jana Nabki, Filippo Vit, Federico Pozzo, Bashar Al Deeban, Mohammad Almasri, Nawar Maher, Chiara Cosentino, Samir Mouhssine, Luca Cividini, Fatemeh Bokaei, Riccardo Dondolin, Joseph Ghanej, Francesca Maiellaro, Silvia Rasi, Antonella Zucchetto, Tamara Bittolo, Riccardo Bomben, Gianluca Gaidano, Valter Gattei and Riccardo Moia

Received: August 28, 2026.
Accepted: November 19, 2026.

Citation: Jana Nabki, Filippo Vit, Federico Pozzo, Bashar Al Deeban, Mohammad Almasri, Nawar Maher, Chiara Cosentino, Samir Mouhssine, Luca Cividini, Fatemeh Bokaei, Riccardo Dondolin, Joseph Ghanej, Francesca Maiellaro, Silvia Rasi, Antonella Zucchetto, Tamara Bittolo, Riccardo Bomben, Gianluca Gaidano, Valter Gattei and Riccardo Moia. The potential role of intraclonal diversification of immunoglobulin light chain genes in chronic lymphocytic leukemia prognostication.
Haematologica. 2026 Nov 26. doi: 10.3324/haematol.2026.301894 [Epub ahead of print]

Publisher's Disclaimer.
E-publishing ahead of print is increasingly important for the rapid dissemination of science. Haematologica is, therefore, E-publishing PDF files of an early version of manuscripts that have completed a regular peer review and have been accepted for publication.
E-publishing of this PDF file has been approved by the authors.
After having E-published Ahead of Print, manuscripts will then undergo technical and English editing, typesetting, proof correction and be presented for the authors' final approval, the final version of the manuscript will then appear in a regular issue of the journal.
All legal disclaimers that apply to the journal also pertain to this production process.

The potential role of intraclonal diversification of immunoglobulin light chain genes in chronic lymphocytic leukemia prognostication

Jana Nabki,[1*] Filippo Vit,[2*] Federico Pozzo,[2*] Bashar Al Deeban,[1] Mohammad Almasri,[1] Nawar Maher,[1,3] Chiara Cosentino,[1] Samir Mouhssine,[1] Luca Cividini,[1] Fatemeh Bokaei,[1] Riccardo Dondolin,[1] Joseph Ghanej,[1] Francesca Maiellaro,[1] Silvia Rasi,[1] Antonella Zucchetto,[2] Tamara Bittolo,[2] Riccardo Bomben,[2#] Gianluca Gaidano,[1#] Valter Gattei[2#], Riccardo Moia[1#]

[1]Division of Hematology, Department of Translational Medicine, Università del Piemonte Orientale and AOU Maggiore della Carità di Novara, Novara, Italy

[2]Clinical and Experimental Onco-Hematology Unit, Centro di Riferimento Oncologico di Aviano (CRO), IRCCS, Aviano, Italy

[3]Università del Piemonte Orientale, Dipartimento di Medicina Traslazionale, e Azienda Ospedaliero Universitaria SS Antonio e Biagio e Cesare Arrigo, Ematologia, Alessandria, Italy

[*]Equally contributed as first authors

[#]Equally contributed as last authors

Corresponding Author: Riccardo Moia, MD, PhD, Division of Hematology, Department of Translational Medicine, University of Eastern Piedmont, Via Solaroli 17, 28100 Novara, Italy. E-mail: riccardo.moia@uniupo.it

Acknowledgments: This work was supported by: Molecular bases of disease dissemination in lymphoid malignancies to optimize curative therapeutic strategies (AIRC5 x 1000 No. 21198), Associazione Italiana per la Ricerca sul Cancro (AIRC) Foundation Milan, Italy; Progetto di Ricerca Sanitaria Finalizzata 2024 (GR-2024-12377449); AIRC IG-32112; PNRR-MAD-2022-12375673 (Next Generation EU, M6/C2_CALL 2022), Italian MoH, Rome, Italy; the AGING Project – Department of Excellence – DIMET, Università del Piemonte Orientale, Novara, Italy; AIL Novara VCO ODV, Novara, Italy.

Contributions

J.N., F.V., F.P., R.B., G.G., V.G., R.M. designed the study, performed analysis, interpreted data, and wrote the manuscript. B.A.D., M.A., N.M., C.C., S.M., L.C., F.B., R.D., J.G., F.M., S.R., A.Z., T.B. performed biological analysis and contributed to data analysis and manuscript preparation.

Data Availability Statement

High-throughput datasets produced for this work will be available from the corresponding author and/or are on the European Genome-phenome Archive (EGA) EGAD50000003183 and EGAS50000002219.

AI use declaration

AI was not used for generating this manuscript.

Disclosure of Conflicts of Interest

RM is on the advisory board and received speaker's honoraria from Johnson & Johnson, Abbvie, AstraZeneca and BeOne. GG received honoraria from Abbvie, AstraZeneca, BeOne, Hikma, Incyte, Johnson & Johnson and Lilly.

TO THE EDITOR:

This study identifies intraclonal diversification (ID) of immunoglobulin light chain (LC) genes as a novel independent prognostic factor in early-stage chronic lymphocytic leukemia (CLL) by exploiting a large consecutive real-life cohort of 664 patients with CLL. Beyond improving risk stratification, LC-ID provides a previously unrecognized molecular framework to investigate CLL clonal growth and evolution, raising the hypothesis that ongoing diversification reflects distinct biological programs influencing clonal expansion, B-cell receptor (BCR) biology, and disease progression.

The mutational status of the immunoglobulin heavy variable (IGHV) genes remains a cornerstone for CLL risk stratification, by distinguishing mutated (M-IGHV) and unmutated (U-IGHV) cases.[1] LC genes also play a role in BCR structure and signaling, as exemplified by the usa of specific LC rearrangements in stereotyped subsets such as IGLV3-21 in subset#2 and IGKV1-39/1D-39 in subset#1,[2] and by the IGLV3-21^{R110} mutation responsible for autonomous signaling and aggressive behavior.[3,4] In addition, the mutational status of LC genes complements IGHV mutational status in refining time to first treatment (TTFT).[5]

Deep sequencing techniques have revealed that CLL clonotypes may evolve through the accumulation of somatic hypermutation (SHM) within IGHV, generating sub-clonotypes through a process known as ID.[6,7] ID within IGHV genes appears to be ubiquitous in CLL, although at different levels,[7,8] and recent evidence indicates that high levels of ID within M-IGHV CLL correlate with prolonged TTFT.[8] In contrast, ID within LC genes has only been investigated in selected stereotyped subsets.[9] Therefore, we aimed to characterize ID within LC genes in a real-world consecutive CLL cohort and to determine the prognostic impact of ID across both IGHV and LC genes.

IGHV and LC genes (κ and λ) libraries were generated from tumor genomic DNA extracted from peripheral blood mononuclear cells at CLL diagnosis, using a two-step PCR amplification protocol, followed by next generation sequencing (NGS) on Illumina platforms. A bioinformatic pipeline was used to quantify IGHV ID levels using the inverse Simpson Index (iSI).[8] Analysis of LC ID levels was based on core packages of the Immcantation framework; pRESTO (v.0.6.2) and Change-O (v.0.4.6). Identical sequences were collapsed while retaining duplicate counts, sequences supported by fewer than 5 reads were excluded, and the remaining unique sequences were aligned to the human IMGT reference using IgBLAST. Clonal clustering was performed based on CDR3 nucleotide sequence similarity using a normalized Hamming distance model with a threshold of 0.1. Within each sample, the clonotype with the highest read count was defined as the dominant clonotype, *bona fide* representing the CLL population. Sub-clonotypes were defined as sequences sharing the same V gene and identical CDR3 nucleotide sequence as the dominant clonotype, differing only by SHM across the FR1–FR3 regions of the V domain. Only samples with productive dominant clonotypes with more than 1000 reads assigned were kept for further analysis. For diversity analysis, only sub-clonotypes representing at least 0.1% of the total reads within the dominant clonotype were considered. LC-ID levels were quantified using iSI calculated from sub-clonal frequencies. The primary endpoint of this study was TTFT analyzed using the cumulative incidence function with death as a competing risk event. This study was approved by the Ethical Committee of AOU Maggiore della Carità di Novara (CE 120/19).

A total of 664 consecutive patients with CLL were included in the analysis, and their characteristics were consistent with a real-world cohort of unselected newly diagnosed CLL (Table 1 and Figure S1A), as also documented by shorter survival of patients harboring U-IGHV and *TP53* disruption (Figure S1B, C) and shorter TTFT in Binet A patients with unmutated IGHV and

LC (Figure S1D, E). Median follow-up was 13 years. An optimal statistical cut-point of IGHV-ID levels (iSI =1.288) was determined using a recursive partitioning approach to predict TTFT in early-stage CLL (Binet A, N=428; p=0.002; Figure S1F). This iSI cutoff is very close to the 1.23 cutoff identified by Vit *et al*. and significantly stratified TTFT in early-stage CLL (p<0.001) therefore supporting previous findings.[8] More precisely, IGHV High-ID patients (iSI≥1.288; N=259) had a 13-year treatment risk of 32.4% compared to 53.5% for IGHV Low-ID patients (iSI<1.288; N=169) (Figure 1A).

Overall, High-ID prevalence was superimposable in M and U-IGHV patients (p=0.93) (Figure S1G). In Binet A patients, High-ID refined the prognostic value of IGHV mutational status in predicting TTFT both in M-IGHV, with a 13-year treatment risk of 19.9% vs 38.1% in Low-ID patients (p=0.009), and in U-IGHV CLL, with a 13-year treatment risk of 58.6% vs 81.8% in Low-ID (p=0.013) (Figure 1B). In a multivariate regression model, High-ID levels in IGHV genes (sHR 0.59, 95% CI 0.42-0.81; p=0.001) were independently associated with a lower risk of treatment when adjusted for the International Prognostic Score for Early-stage CLL (IPS-E)[10] variables (namely U-IGHV, lymphocyte count >15×10^3/μl and palpable lymph nodes), and for LC genes mutational status.[5] The association remained highly stable (91.1%) across 1000 bootstraps (Figure 1C). In addition, borderline-mutated IGHV (BL-IGHV) cases showed an intermediate outcome in terms of treatment requirement in Binet A (p<0.001)[11](Figure S1H) and ID also stratified the risk of treatment requirement within the BL-IGHV subgroup (p=0.02) (Figure S1I).

Central to this analysis, an optimal cut-point of LC genes ID levels for predicting TTFT in early-stage CLL was determined as an iSI=1.007 (p=0.006) (Figure S1J). LC High-ID CLL (iSI≥1.007; N=239) had a 13-year treatment risk of 32.2% compared to 50.7% for LC Low-ID CLL (iSI<1.007; N=176) (p<0.001) (Figure 1D).

Similar to IGHV genes, LC High-ID prevalence in the whole cohort was superimposable in patients with with mutated (M-LC) or unmutated (U-LC) LC genes (p=0.5) (Figure S1K). In Binet A CLL, LC High-ID levels refined the prognostic value of LC mutational status in predicting TTFT in U-LC CLL with a 13-year treatment risk of 53.6% vs 81.3% in Low-ID (p=0.012) and demonstrated a significant trend in M-LC patients with a 13-year treatment risk of 23.0% vs 31.0% in Low-ID (p=0.047) (Figure 1E). In multivariate analysis, LC High-ID independently associated with lower risk of treatment (sHR 0.61, 95% CI 0.43-0.85; p=0.004) when adjusted for IPS-E[10] variables and for LC genes mutational status. The association showed a stability of 83.6% across 1000 bootstraps (Figure 1F).

ID was subsequently correlated with baseline clinical and molecular features of CLL in all Binet stages. Distinct distributions of ID status were observed across both heavy and light chains, including κ and λ subtypes (Figures S2A-C). IGLV3-21^{R110} cases (N=28) were split into 19 High-ID and 9 Low-ID while subset #2 was enriched in cases with LC High-ID (N=10/12) (Figure S2C,D). Preliminary descriptive data that require validation suggest that subset #2 cases retain aggressive behavior despite High-ID. This association could be due the IGLV3-21^{R110} mutation that leads to autonomous BCR signaling and is conserved across sub-clonotypes (Figure S2E).[12] Moreover, High-ID in both IGHV and LC demonstrated significant negative correlations with the dominant clonotype percentage (p<0.001 and p<0.001, respectively), which reflects the CLL clonotype occupancy in the B-cell compartment, as well as patient's lymphocyte count (p=0.042 and p=0.0123, respectively) (Figures S2F-I). ID may reflect a biological process not explained by the assessed molecular drivers, supported by the lack of significant correlations after multiple comparisons adjustment (Figure 2A).

Paired analysis of iSI levels between IGHV and LC genes demonstrated significantly higher levels of ID in IGHV (median iSI=1.330) versus LC genes (median iSI=1.008, $p<0.001$) (Figure 2B). In Binet A patients, both IGHV High-ID and LC High-ID independently predicted TTFT (Figure 2C). According to combined ID levels of IGHV and LC genes, concordant High-ID patients (N=136) had a significantly lower risk of treatment (27.1% at 13-year) compared to other groups of patients, namely concordant Low-ID (N=79; 55.5% at 13-year, $p<0.001$), LC High-ID + IGHV Low-ID (N=74; 52.8% at 13-year; $p=0.007$), and IGHV High-ID + LC Low-ID (N=87; 48.1% at 13-year, $p=0.009$) (Figure 2D). Importantly, the presence of concordant High-ID in both IGHV and LC genes predicted significantly lower treatment risk (sHR 0.52, 95% CI 0.35-0.77; $p<0.001$) that was independent when adjusted for IPS-E[10] variables and for mutational status of LC genes. Bootstrap stability was 91.8% across 1000 resamples. (Figure 2E). The addition of concordant High-ID status and LC mutational status improved the time-dependent AUC at 13 years of 0.81 vs 0.77 of the IPS-E model (95% CI 0.007–0.061; $p=0.013$)

Based on these results, ID emerges as a recurrent process in CLL that extends to LC genes, adding further complexity to the BCR role in disease pathogenesis. This study provides the first evidence that High-ID levels in both heavy and LC immunoglobulin genes independently translate into prolonged TTFT in early-stage patients with CLL, complementing the IPS-E variables and LC mutational status. The reduction in both clonal dominance and lymphocyte counts in High-ID patients raises the hypothesis that High-ID may characterize CLL populations with lower clonal dominance and reduced proliferative fitness.[13] In contrast, Low-ID clonotypes appear to expand without further evolution, potentially being more susceptible to microenvironmental stimuli leading to shorter TTFT.[14] Future external validation will be required to elucidate ID mechanisms and connection to the tumor microenvironment, alongside whether ID prognostic relevance

extends to patients receiving fixed-duration therapies, in which IGHV still modulates treatment choice.[15]

REFERENCES

1. Moia R, Gaidano G. Prognostication in chronic lymphocytic leukemia. Semin Hematol. 2024;61(2):83-90.
2. Agathangelidis A, Chatzidimitriou A, Gemenetzi K, et al. Higher-order connections between stereotyped subsets: implications for improved patient classification in CLL. Blood. 2021;137(10):1365-1376.
3. Nadeu F, Royo R, Clot G, et al. IGLV3-21R110 identifies an aggressive biological subtype of chronic lymphocytic leukemia with intermediate epigenetics. Blood. 2021;137(21):2935-2946.
4. Stamatopoulos B, Smith T, Crompot E, 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(20):5048-5057.
5. Nabki J, Al Deeban B, Sium AM, et al. Immunoglobulin light chain mutational status refines IGHV prognostic value in identifying chronic lymphocytic leukemia patients with early treatment requirement. Leukemia. 2025;39(3):643-649.
6. Gurrieri C, McGuire P, Zan H, et al. Chronic lymphocytic leukemia B cells can undergo somatic hypermutation and intraclonal immunoglobulin VHDJH gene diversification. J Exp Med. 2002;196(5):629-639.
7. Bagnara D, Tang C, Brown JR, et al. Post-transformation IGHV-IGHD-IGHJ mutations in chronic lymphocytic leukemia b cells: implications for mutational mechanisms and impact on clinical course. Front Oncol. 2021;11:640731.

8. Vit F, Bittolo T, Zucchetto A, et al. Measurement of immunoglobulin intraclonal diversification refines the clinical impact of IGHV mutational status in chronic lymphocytic leukemia. Leukemia. 2025;39(8):1905-1914.
9. Kostareli E, Sutton LA, Hadzidimitriou A, et al. Intraclonal diversification of immunoglobulin light chains in a subset of chronic lymphocytic leukemia alludes to antigen-driven clonal evolution. Leukemia. 2010;24(7):1317-1324.
10. Condoluci A, Terzi di Bergamo L, Langerbeins P, et al. International prognostic score for asymptomatic early-stage chronic lymphocytic leukemia. Blood. 2020;135(21):1859-1869.
11. Angotzi F, Bevilacqua A, Serafin A, et al. Outcomes of patients with CLL and borderline IGHV mutational status: a systematic review and meta analysis. Biomark Res. 2026;14(1):42.
12. Gemenetzi K, Psomopoulos F, Carriles AA, et al. Higher-order immunoglobulin repertoire restrictions in CLL: the illustrative case of stereotyped subsets 2 and 169. Blood. 2021;137(14):1895-1904.
13. Packham G, Krysov S, Allen A, et al. The outcome of B-cell receptor signaling in chronic lymphocytic leukemia: proliferation or anergy. Haematologica. 2014;99(7):1138-1148.
14. Ten Hacken E, Burger JA. Microenvironment interactions and B-cell receptor signaling in chronic lymphocytic leukemia: implications for disease pathogenesis and treatment. Biochim Biophys Act. 2016;1863(3):401-413.
15. Eichhorst B, Ghia P, Bosch F, et al. EHA Guidelines on management of chronic lymphocytic leukemia and Richter transformation. Hemasphere. 2026;10(6):e70403.

Table 1. Patient characteristics

Characteristic	Value
Median age at diagnosis (IQR)	70 (62-77)
Sex	
Male	396 (59.6%)
Female	268 (40.4%)
Median lymphocyte count (IQR)	9.8 x 10^3/µl (5.7-17.2)
Median hemoglobin level (IQR)	13.7 g/dl (12.3-14.8)
Median platelet count (IQR)	193 x 10^3/µl (151-237)
Light chain	
Kappa	370 (55.7%)
Lambda	214 (32.2%)
Undefined	80 (12%)
IGHV	
Mutated	391 (58.9%)
Unmutated	265 (39.9%)
na	8 (1.2%)
LC	
Mutated (<99%)	374 (56.3%)
Unmutated (≥99%)	272 (41%)
na	18 (2.7%)
Binet staging system	
Binet A	497 (74.8%)
Binet B	92 (13.9%)
Binet C	60 (9.0%)
na	15 (2.3%)
***TP53* disrupted**	
yes	74 (11.2%)
no	584 (88.8%)
na	6
del(17p)	
yes	51 (8.2%)
no	568 (91.8%)
na	45
***TP53* mutated**	
yes	60 (9.2%)
no	593 (90.8%)
na	11
Trisomy 12	
yes	118 (19.1%)
no	501 (80.9%)
na	45
del(13q)	
yes	296 (47.8%)
no	323 (52.2%)
na	45
del(11q)	
yes	55 (8.9%)
no	564 (91.1%)
na	45
Stereotyped cases	
Subset #1	16 (2.4%)

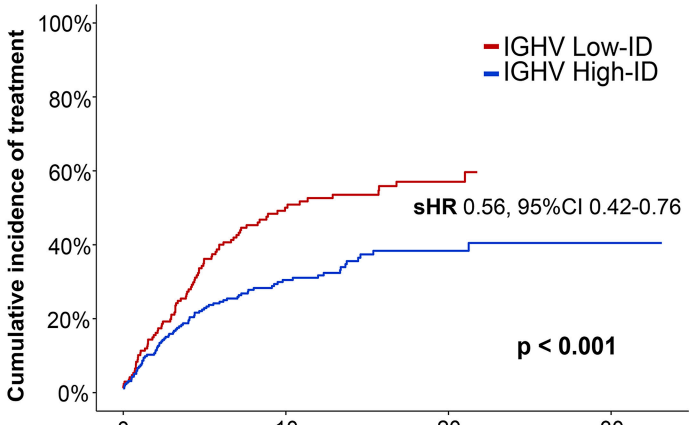
Subset #2	15 (2.3%)
Subset #4	8 (1.2%)
Subset #8	5 (0.8%)

FIGURE LEGENDS

Figure 1. Impact of ID groups and immunoglobulin mutational status on TTFT. (A) Cumulative incidence function of treatment comparing groups defined by IGHV ID levels (High-ID in blue vs. Low-ID in red) and accounting for death as a competing risk event. **(B)** Cumulative incidence function of treatment of IGHV ID levels according to IGHV mutational status (M-IGHV High-ID: represented by the dark blue line, M-IGHV Low-ID: represented by the dark red line, U-IGHV High-ID: represented by the light blue line, U-IGHV Low-ID: represented by the light red/pink line) with death treated as a competing event. **(C)** Forest plot summarizing multivariate Fine-Gray regression analysis of treatment risk while accounting for death as a competing risk event demonstrating the independent prognostic value of IGHV High-ID. **(D)** Cumulative incidence function of treatment accounting for death as a competing risk event and comparing groups defined by LC ID status (LC High-ID in blue vs. LC Low-ID in red). **(E)** Cumulative incidence function of treatment of LC ID levels according to LC mutational status (M-LC High-ID: represented by the dark blue line, M-LC Low-ID: represented by the dark red line, U-LC High-ID: represented by the light blue line, U-LC Low-ID: represented by the light red/pink line) with death treated as a competing event. **(F)** Forest plot summarizing multivariate Fine-Gray regression analysis of treatment risk accounting for death as a competing risk event and demonstrating independent prognostic value of LC High-ID. The "Numbers at Risk" table is shown at the bottom of each curve in the panel. Sub-distribution hazard ratio (sHR), 95% confidence interval (CIs), and Fine–Gray *p* values are indicated adjacent to the curves. Pairwise comparisons between strata are summarized in a table of Holm-adjusted Fine-Gray *p* values, with color-coding matching the corresponding curves

Figure 2. Correlation of ID in both IGHV and LC genes with CLL baseline characteristic and TTFT. (A) Spearman's correlation matrix depicting associations between High-ID status (LC High-ID, IGHV High-ID), mutational status (U-IGHV, U-LC) and CLL biological and molecular baseline features. To account for multiple testing, p-values were corrected using the Benjamini-Hochberg (BH) method. The color scale represents the statistical significance of these associations, calculated as the -log10 of the p-value. Red indicates a significant positive association and blue indicates a significant negative association, with the color intensity corresponding to the level of significance. **(B)** Boxplot illustrating the distribution of iSI for IGHV genes and their corresponding LC genes in patients with CLL. Grey lines connect the paired iSI measurements in each chain of the same patient. **(C)** Cumulative incidence function of treatment risk accounting for death as a competing risk event according to concordant or discordant ID status in IGHV and LC genes. Concordant High-ID in dark blue, concordant Low-ID in red, IGHV Low-ID + LC High-ID in grey, and IGHV High-ID + LC Low-ID in yellow. Pairwise comparisons between strata are summarized in a table of Holm-adjusted Fine-Gray *p* values, with color-coding matching the corresponding curves. **(D)** Forest plot summarizing multivariate Fine-Gray regression analysis of treatment risk accounting for death as a competing risk event and demonstrating the independent prognostic value of combined High-ID status of IGHV and LC genes.

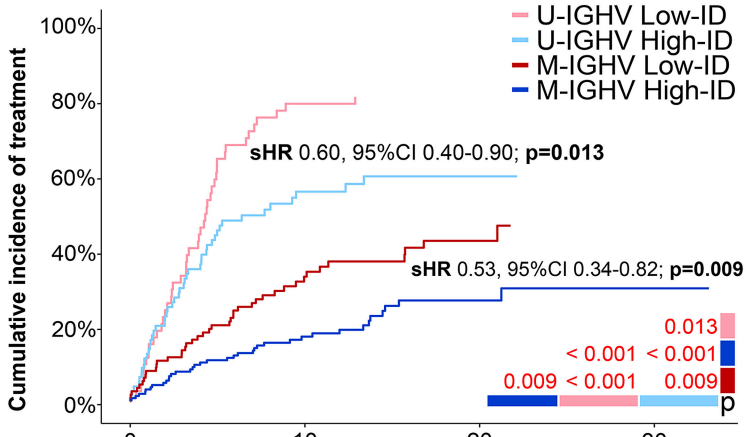
A



Numbers at Risk

IGHV Low-ID	169	35	4	0
IGHV High-ID	259	80	15	3

B

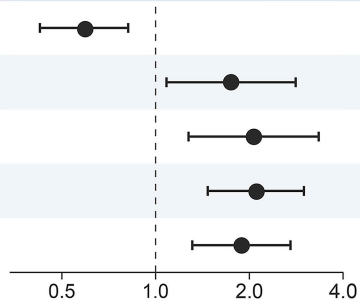


Numbers at Risk

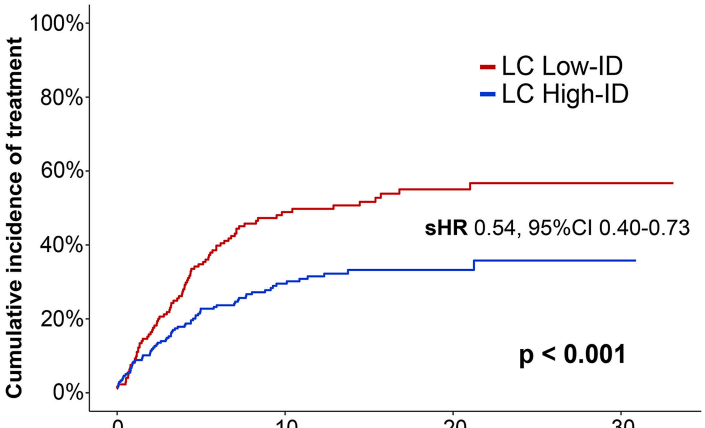
U-IGHV Low-ID	56	1	0	0
U-IGHV High-ID	83	14	1	0
M-IGHV Low-ID	113	34	4	0
M-IGHV High-ID	176	66	14	3

C

Variable	N	Subdistribution hazard ratio	sHR	95% CI	p-value	Bootstrap CI	Bootstrap stability
IGHV High-ID	414		0.59	0.43–0.82	0.001	0.42–0.81	91.1%
U-LC	414		1.75	1.08–2.82	0.022	1.11–2.84	62.2%
U-IGHV	414		2.06	1.28–3.34	0.003	1.27–3.24	83.7%
Lymphocyte count >15 × 10³/μL	414		2.10	1.47–2.99	<0.001	1.51–3.07	98.1%
Palpable lymph nodes	414		1.89	1.31–2.71	<0.001	1.30–2.78	93.9%



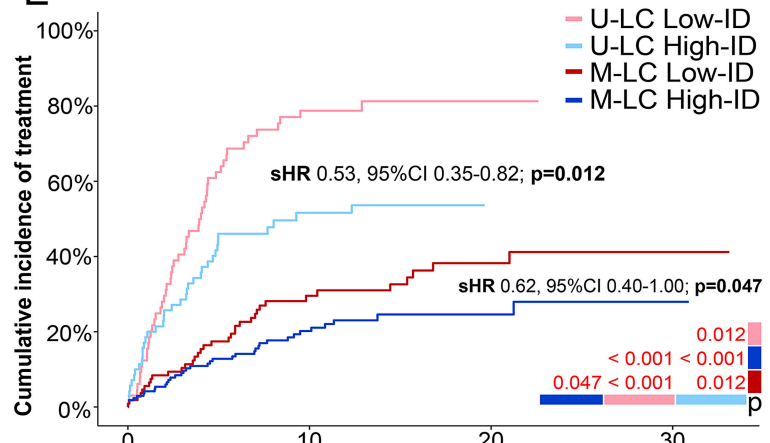
D



Numbers at Risk

LC Low-ID	176	39	10	2
LC High-ID	239	72	9	1

E

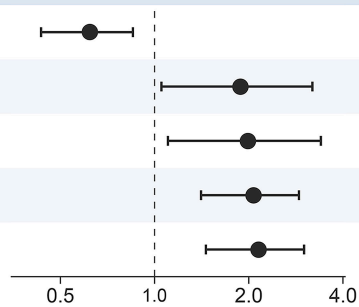


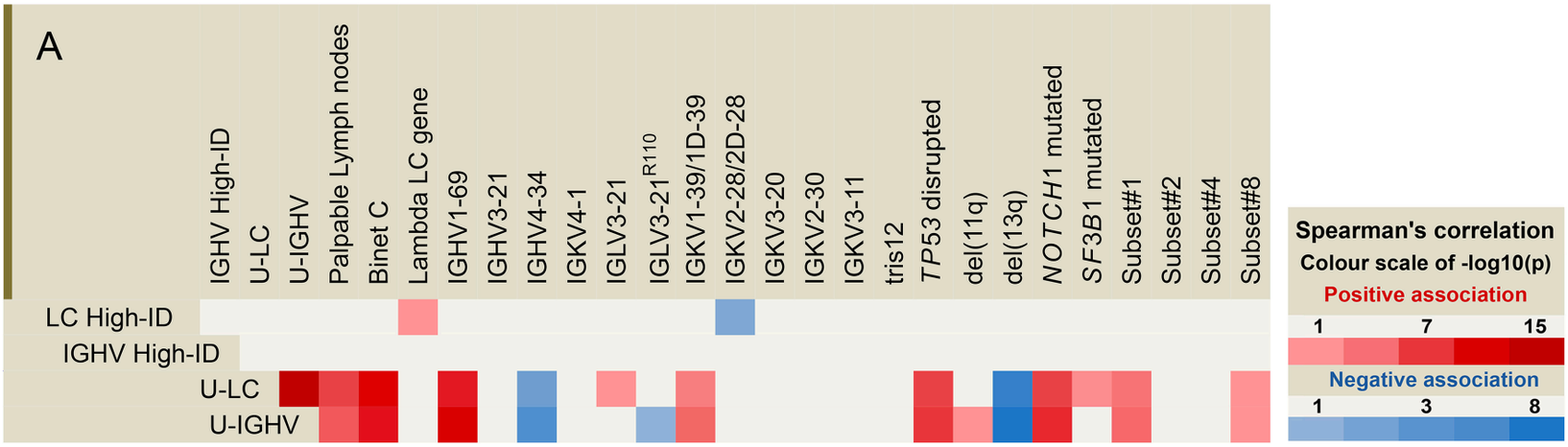
Numbers at Risk

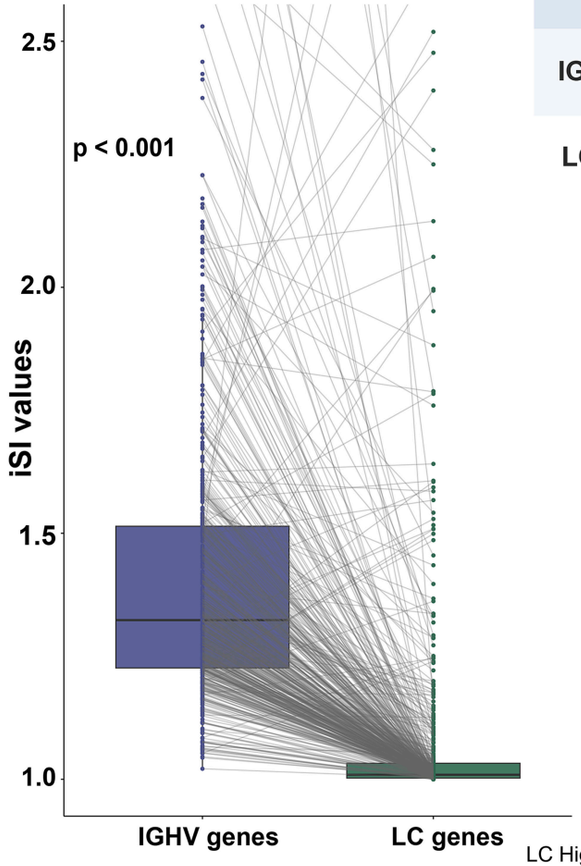
U-LC Low-ID	66	5	1	0
U-LC High-ID	70	9	0	0
M-LC Low-ID	110	34	9	2
M-LC High-ID	169	63	9	1

F

Variable	N	Subdistribution hazard ratio	sHR	95% CI	p-value	Bootstrap CI	Bootstrap stability
LC High-ID	408		0.61	0.43–0.85	0.004	0.43–0.84	83.6%
U-LC	408		1.83	1.05–3.20	0.033	1.05–3.27	56.6%
U-IGHV	408		1.94	1.10–3.40	0.021	1.11–3.37	61.2%
Lymphocyte count >15 × 10³/μL	408		2.02	1.41–2.89	<0.001	1.40–3.06	96.0%
Palpable lymph nodes	408		2.09	1.46–3.01	<0.001	1.43–3.06	96.1%

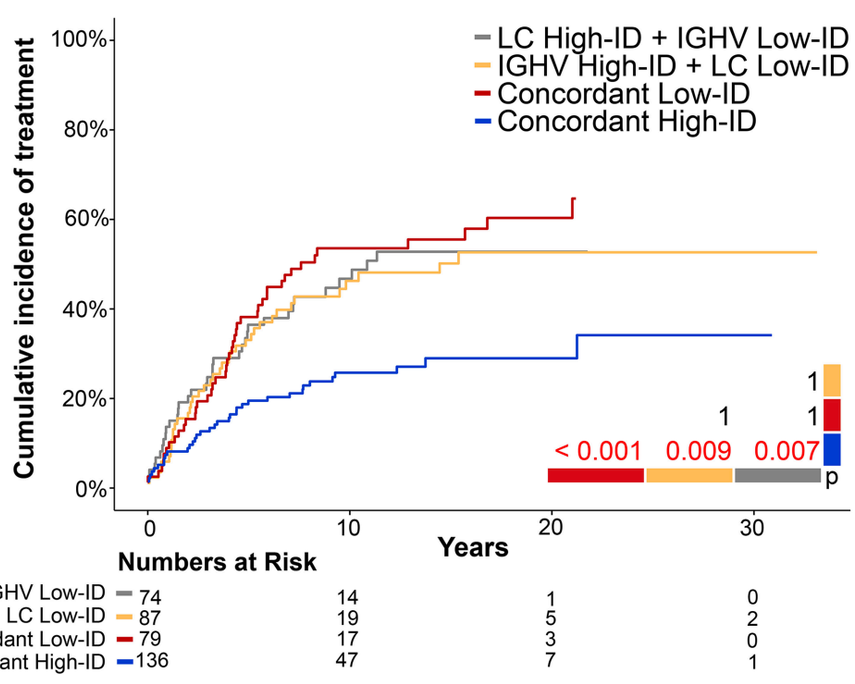






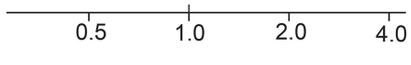
C

Variable	N	Subdistribution hazard ratio	sHR	95% CI	p-value
IGHV High-ID	376		0.63	0.46–0.87	0.005
LC High-ID	376		0.63	0.46–0.87	0.006



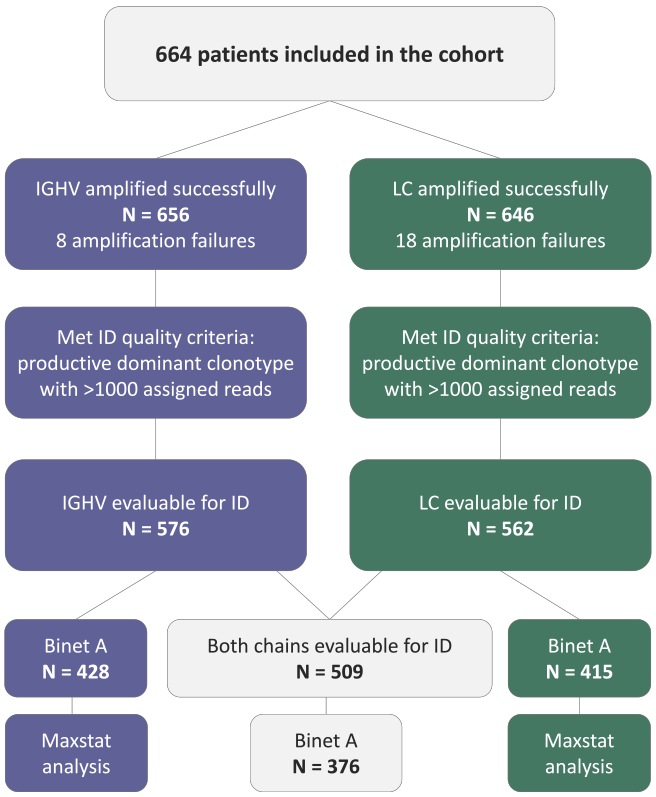
E

Variable	N	Subdistribution hazard ratio	sHR	95% CI	p-value	Bootstrap CI	Bootstrap stability
Concordant High-ID	375		0.52	0.35–0.77	<0.001	0.34–0.78	91.8%
U-LC	375		1.73	0.94–3.17	0.077	1.02–3.43	43.3%
U-IGHV	375		2.02	1.09–3.75	0.025	1.01–3.64	60.7%
Lymphocyte count >15 × 10³/µL	375		2.07	1.43–2.99	<0.001	1.43–3.10	96.8%
Palpable lymph nodes	375		1.79	1.22–2.61	0.003	1.23–2.73	85.7%

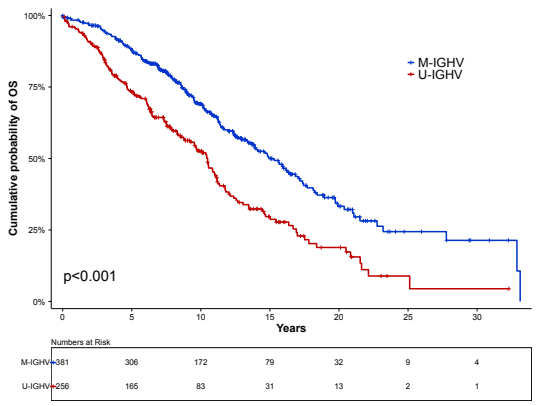


Supplementary

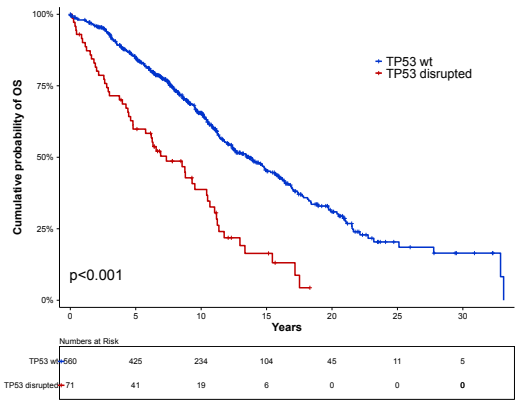
A



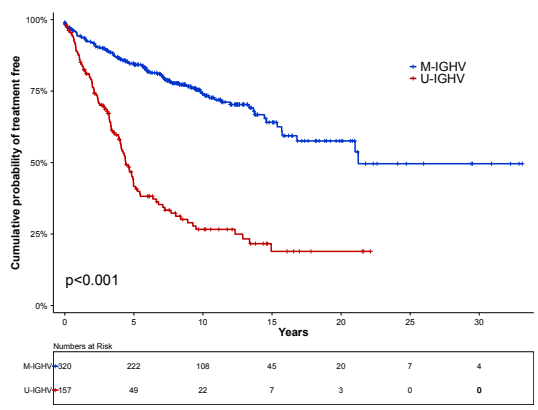
B



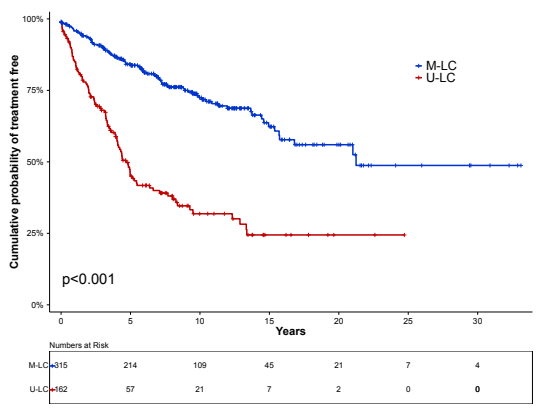
C



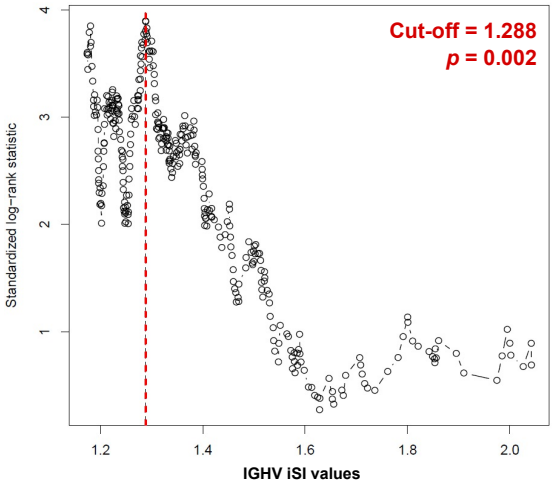
D



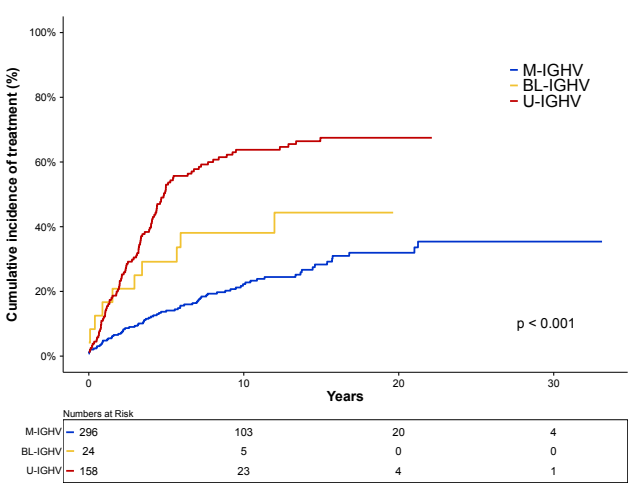
E



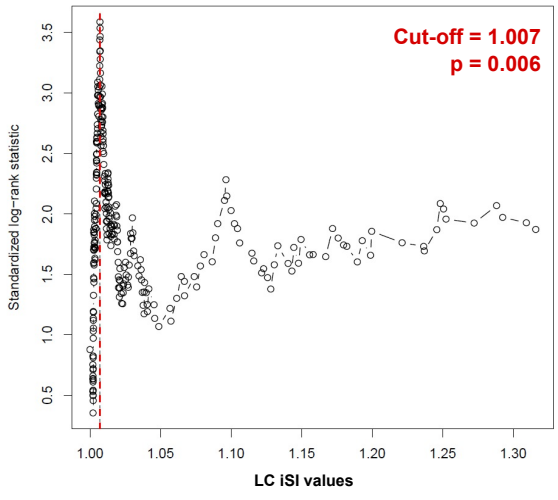
F



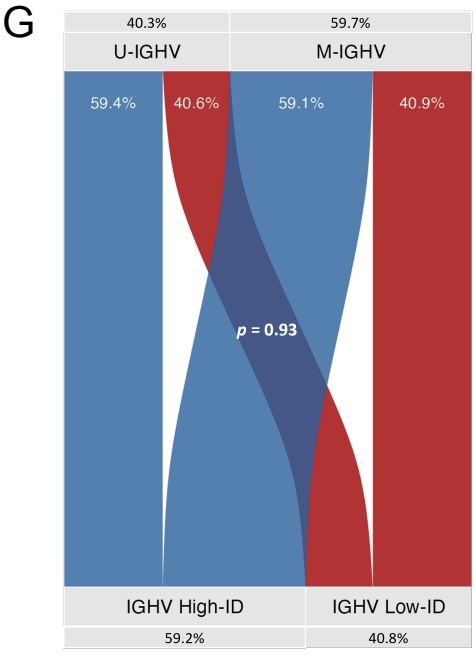
H



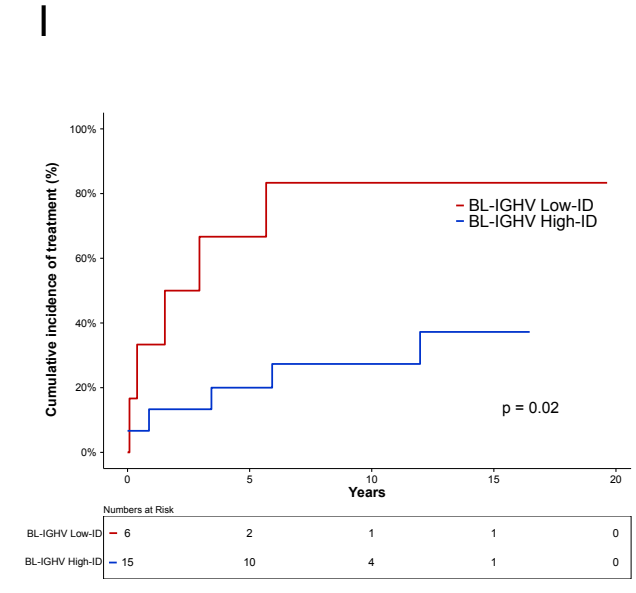
J



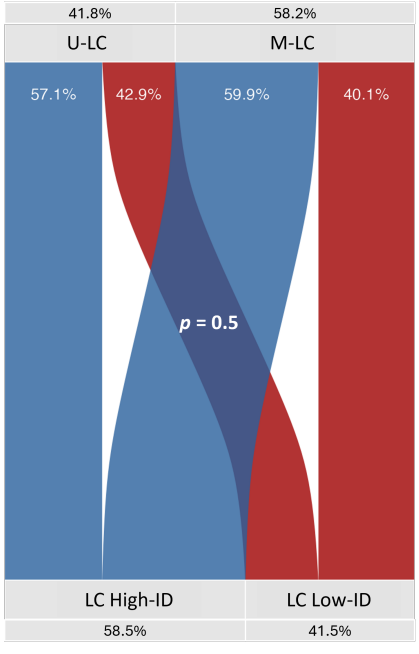
G



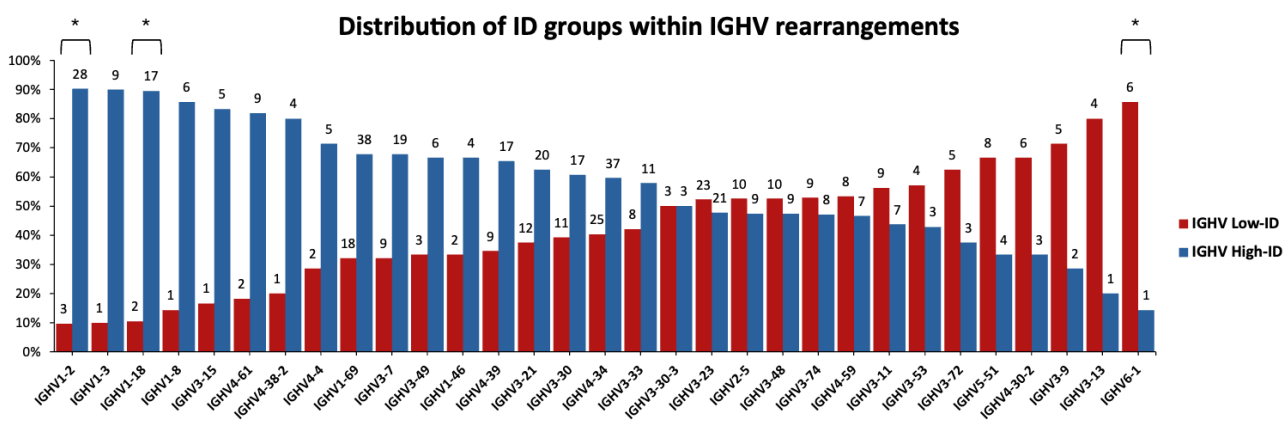
I



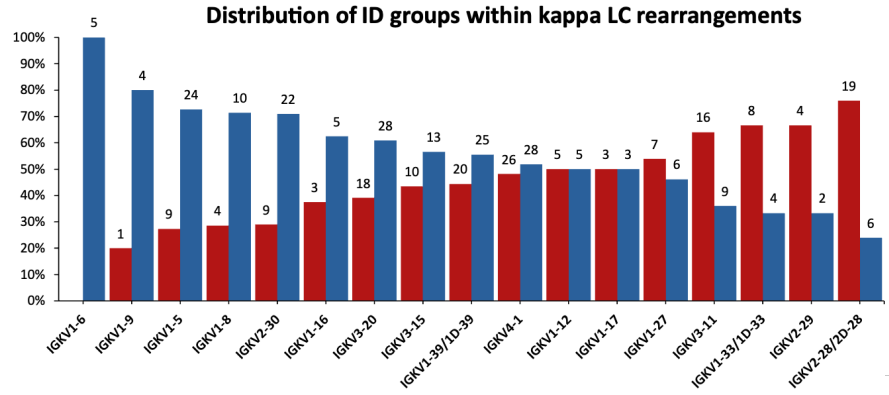
K



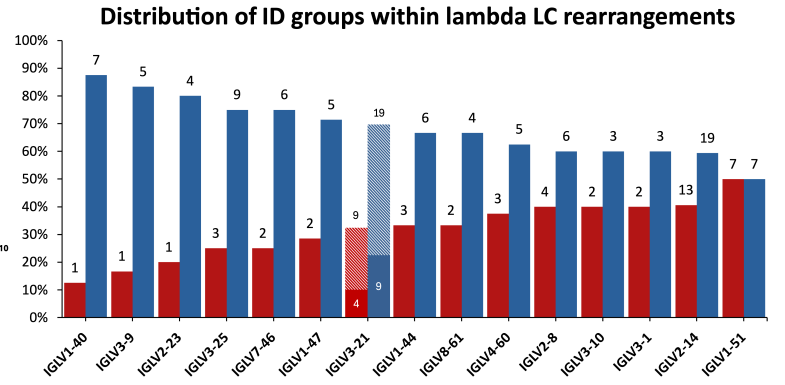
A


Distribution of ID groups within IGHV rearrangements
IGHV Low-ID
IGHV High-ID
IGHV1-2
IGHV1-3
IGHV1-18
IGHV1-8
IGHV3-15
IGHV4-61
IGHV4-38-2
IGHV4-4
IGHV1-69
IGHV3-7
IGHV3-49
IGHV1-46
IGHV4-39
IGHV3-21
IGHV3-30
IGHV4-34
IGHV3-33
IGHV3-30-3
IGHV3-23
IGHV2-5
IGHV3-48
IGHV3-74
IGHV4-59
IGHV3-11
IGHV3-53
IGHV3-72
IGHV5-51
IGHV4-30-2
IGHV3-9
IGHV3-13
IGHV6-1

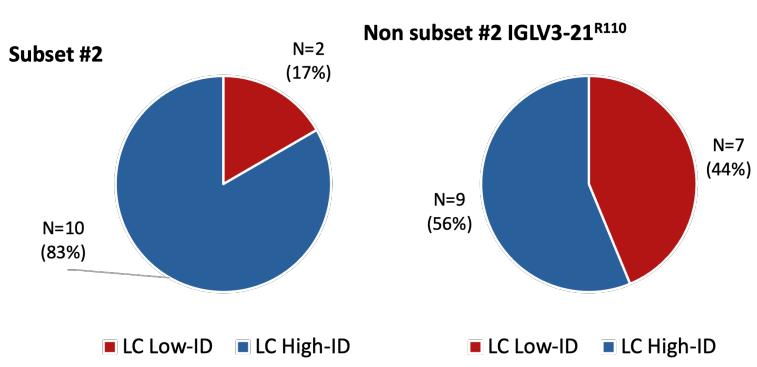
B


Distribution of ID groups within kappa LC rearrangements
LC Low-ID
LC High-ID
IGLV3-21R110
IGKV1-6
IGKV1-9
IGKV1-5
IGKV1-8
IGKV2-30
IGKV1-16
IGKV3-20
IGKV3-15
IGKV1-39/1D-39
IGKV4-1
IGKV1-12
IGKV1-17
IGKV1-27
IGKV3-11
IGKV1-33/1D-33
IGKV2-29
IGKV2-28/2D-28

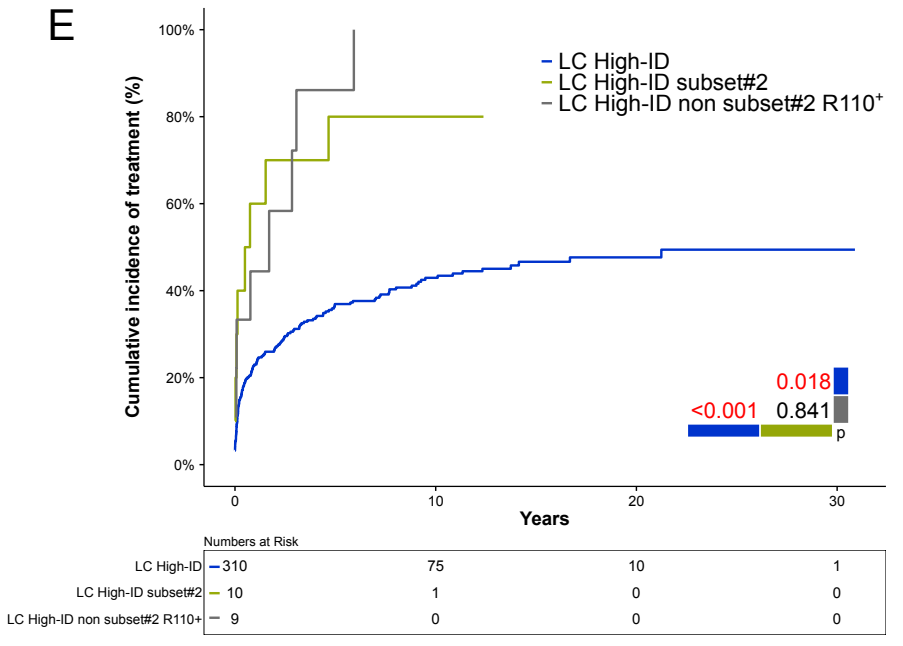
C


Distribution of ID groups within lambda LC rearrangements
IGLV1-40
IGLV3-9
IGLV2-23
IGLV3-25
IGLV7-46
IGLV1-47
IGLV3-21
IGLV1-44
IGLV8-61
IGLV4-60
IGLV2-8
IGLV3-10
IGLV3-1
IGLV2-14
IGLV1-51

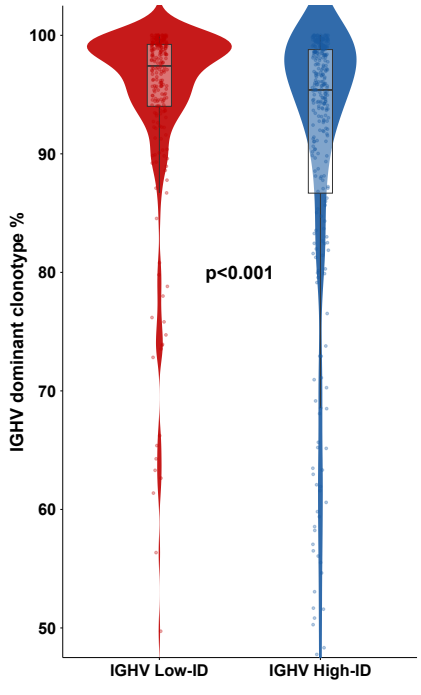
D


Subset #2
N=2
(17%)
N=10
(83%)
LC Low-ID
LC High-ID
Non subset #2 IGLV3-21R110
N=7
(44%)
N=9
(56%)
LC Low-ID
LC High-ID

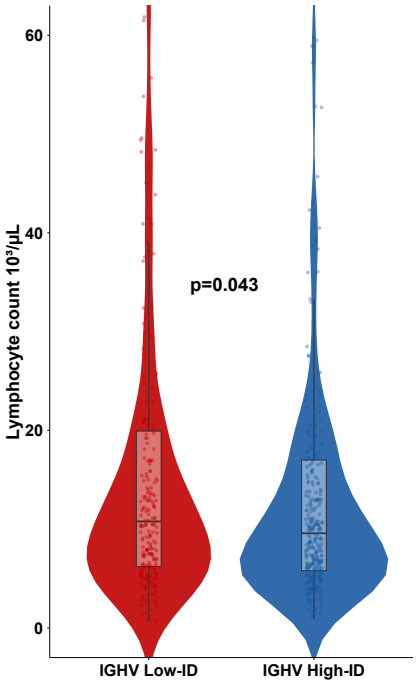
E


LC High-ID
LC High-ID subset#2
LC High-ID non subset#2 R110+
Cumulative incidence of treatment (%)
Years
0.018
<0.001
0.841
p
Numbers at Risk
LC High-ID 310 75 10 1
LC High-ID subset#2 10 1 0 0
LC High-ID non subset#2 R110+ 9 0 0 0

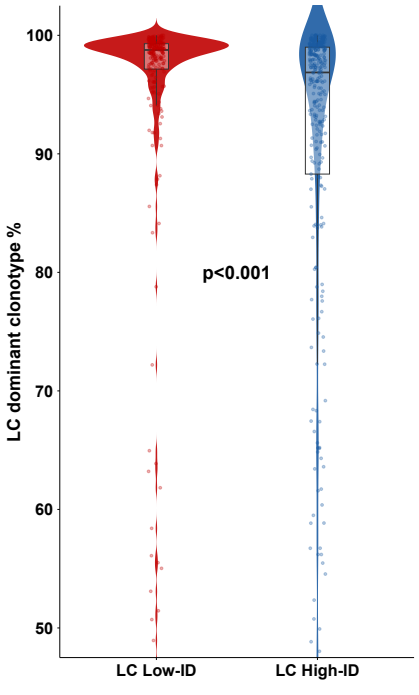
F


IGHV dominant clonotype %
p<0.001
IGHV Low-ID
IGHV High-ID

G


Lymphocyte count 10³/µL
p=0.043
IGHV Low-ID
IGHV High-ID

H


LC dominant clonotype %
p<0.001
LC Low-ID
LC High-ID

I

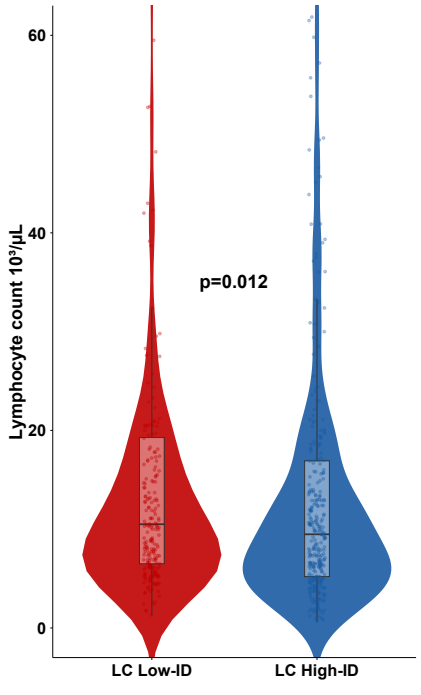

Lymphocyte count 10³/µL
p=0.012
LC Low-ID
LC High-ID

Figure S1. Prognostic impact of IGHV, *TP53*, light-chain (LC) mutational status, and intraclonal diversification in CLL. (A) Flow diagram illustrating the number of patients included at each stage of the study, including intraclonal diversification (ID) assessment. **(B–C)** Kaplan–Meier curves for overall survival (OS) according to **(B)** IGHV mutational status (M-IGHV in blue vs U-IGHV in red) and **(C)** *TP53* status (wild type [wt] in blue vs *TP53* disrupted, defined as *TP53* mutation and/or del(17p) in red). **(D–E)** Kaplan–Meier curves for time to first treatment (TTFT) in patients with Binet A CLL according to **(D)** IGHV mutational status (M-IGHV in blue vs U-IGHV in red) and **(E)** LC gene mutational status (M-LC, identity <99% in blue; U-LC, identity ≥99% in red). Log-rank *p* values are shown in each panel. **(F)** Recursive partitioning analysis to identify the optimal IGHV inverse Simpson Index (iSI) cut-off for predicting TTFT in Binet A CLL (N=428). The optimal cut-off determined using the maximally selected rank statistic (maxstat test) is represented with a red dashed line (1.288) used to stratify patients into IGHV High-ID and IGHV Low-ID. The corresponding p value is shown in the figure. **(G)** Alluvial plot showing the distribution of IGHV ID groups according to IGHV mutational status in the entire cohort (N=576). Flows represent the proportion of patients transitioning between IGHV High-ID (blue) and IGHV Low-ID (red) categories and IGHV mutational status, classified as mutated (M-IGHV) or unmutated (U-IGHV). Relative proportions of patients within each group and subgroup and a *p* value for the between-group comparison are indicated on the figure. **(H)** Cumulative incidence function of treatment accounting for death as a competing risk event and comparing groups defined by IGHV mutational status classified as mutated (M-IGHV) or unmutated (U-IGHV) or borderline mutated (BL-IGHV) defined by a germline identity between 97.00% and 97.99% (M-IGHV in blue, U-IGHV in red, and BL-IGHV in yellow). A Fine–Gray p value is indicated adjacent to the curve. **(I)** Cumulative incidence function of treatment accounting for death as a competing risk event and comparing groups defined by BL-IGHV ID status (BL-IGHV High-ID in blue

vs. BL-IGHV Low-ID in red). A Fine–Gray p value is indicated adjacent to the curve. **(J)** Recursive partitioning analysis to identify the optimal LC iSI cut-off for predicting TTFT in Binet A CLL (N=415). The optimal cut-off determined using the maxstat test is represented with a red dashed line (1.007) used to stratify patients into LC High-ID and LC Low-ID groups. The corresponding *p* value is shown in the figure. **(K)** Alluvial plot showing the distribution of LC ID groups according to LC mutational status in the entire cohort (N=562). Flows represent the proportion of patients between LC High-ID (blue) and LC Low-ID (red) categories and LC mutational status, classified as mutated (M-LC) or unmutated (U-LC). The relative proportions within each category and subgroup are indicated, along with the *p* value for the comparison between groups.

Figure S2. Distribution of intraclonal diversification (ID) groups across immunoglobulin gene rearrangements and their association with clonal dominance, lymphocyte count, and IGLV3-21^{R110} status. (A) Bar plot depicting the proportional distribution of IGHV ID groups (IGHV High-ID in blue and IGHV Low-ID in red) within IGHV gene rearrangements with at least 5 occurrences. Absolute patient numbers for each subgroup are provided above their respective bars. (* : $p < 0.05$). **(B–C)** Bar plots showing the proportional distribution of LC ID categories (LC High-ID in blue and LC Low-ID in red) within LC gene rearrangements with at least 5 occurrences. **(B)** Kappa LC gene rearrangements. **(C)** Lambda LC gene rearrangements. Absolute patient numbers for each subgroup are shown above their respective bars (* : $p < 0.05$). **(D–E)** Distribution and prognostic impact of IGLV3-21^{R110} carriers across LC High-ID group (N=329). **(D)** Pie charts illustrating the proportions of LC High-ID (blue) and LC Low-ID (red) cases within subset #2 (left) and non-subset #2 IGLV3-21^{R110} (right). **(E)** Cumulative incidence function of treatment accounting for death a competing risk event and comparing groups of LC High-ID patients (blue) and LC High-ID subset #2 patients (green)

and LC High-ID non-subset #2 IGLV3-21^{R110} patients (grey). Numbers at Risk are below the figure. Pairwise comparisons between strata are summarized in a table of Holm-adjusted Fine-Gray p values, with color-coding matching the corresponding curve. **(F–G)** Violin plots comparing patients according to IGHV ID group (N=576); IGHV Low-ID (red) or IGHV High-ID (blue). **(F)** Percentage of the dominant clonotype, as determined by IGHV sequencing. **(G)** Absolute lymphocyte count ($\times 10^3/\mu l$). *p* values for between-group comparisons are shown in each panel. **(H–I)** Violin plots comparing patients according to LC ID group (N=562); LC Low-ID (red) or LC High-ID (blue). **(H)** Percentage of the dominant clonotype, as determined by LC sequencing. **(I)** Absolute lymphocyte count ($\times 10^3/\mu l$). *p* values for between-group comparisons are shown in each panel.