SUPPLEMENTARY APPENDIX

INTEGRATED MUTATIONAL AND CYTOGENETIC ANALYSIS IDENTIFIES NEW PROGNOSTIC SUBGROUPS IN CHRONIC LYMPHOCYTIC LEUKEMIA

Davide Rossi, M.D., Ph.D.,[1] Silvia Rasi, Ph.D.,[1] Valeria Spina, Ph.D.,[1] Alessio Bruscaggin, Ph.D.,[1] Sara Monti, Ph.D.,[1] Carmela Ciardullo, Ph.D.,[1] Clara Deambrogi, Ph.D.,[1] Hossein Khiabanian, Ph.D.,[2] Roberto Serra, B.Sc.,[3] Francesco Bertoni, M.D.,[4] Francesco Forconi, M.D., Ph.D.,[5,6] Luca Laurenti, M.D.,[7] Roberto Marasca, M.D.,[8] Michele Dal-Bo, Ph.D.,[9] Francesca Maria Rossi, Ph.D.,[9] Pietro Bulian, M.D.,[9] Josep Nomdedeu, M.D.,[10] Giovanni Del-Poeta, M.D.,[13] Valter Gattei, M.D.,[9] Laura Pasqualucci, M.D., Ph.D.,[11,12,14] Raul Rabadan, Ph.D.,[2] Robin Foà, M.D.,[15§] Riccardo Dalla-Favera, M.D.,[11,12,16§] Gianluca Gaidano, M.D., Ph.D.[1§]

[§]R.F, R.D.-F., and G.G. contributed equally to this study

[1]Division of Hematology and [3]Laboratory of Medical Informatics, Department of Translational Medicine, Amedeo Avogadro University of Eastern Piedmont, Novara, Italy; [2]Department of Biomedical Informatics and Center for Computational Biology and Bioinformatics, [11]Institute for Cancer Genetics and the Herbert Irving Comprehensive Cancer Center, [12]Department of Pathology & Cell Biology, and [16]Department of Genetics & Development, Columbia University, New York, USA; [4]Lymphoma and Genomics Research Program, IOR-Institute of Oncology Research and Lymphoma Unit, Oncology Institute of Southern Switzerland, Bellinzona, Switzerland; [5]Cancer Sciences Unit, CRUK Clinical Centre, University of Southampton, Southampton, United Kingdom & [6]Division of Hematology, University of Siena, Siena, Italy; [7]Institute of Hematology, Catholic University of the Sacred Heart, Rome, Italy; [8]Division of Hematology, Department of Oncology and Hematology, University of Modena and Reggio Emilia, Modena, Italy; [9]Clinical and Experimental Onco-Hematology, Centro di Riferimento Oncologico, Aviano, Italy; [10]Department of Hematology and Laboratory, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain; [13]Department of Hematology, Tor Vergata University, Rome, Italy; [14]Institute of Hematology, Department of Hematology and Clinical Immunology, University of Perugia, Perugia, Italy; [15]Division of Hematology, Department of Cellular Biotechnologies and Hematology, Sapienza University, Rome, Italy.

SUPPLEMENTARY APPENDIX CONTENTS

SUPPLEMENTARY METHODS

Patients

The training series was represented by a multicentric cohort of 637 newly diagnosed and previously untreated CLL patients who consecutively presented for initial evaluation at four institutions from June 1996 through June 2011 (Table S1). Regular follow-up (at least three visits/year) was available for 583 (91.5%) patients (Table S1). Features at presentation did not differ between cases with complete follow-up and cases lacking follow-up information. The clinical database was updated in May 2012. After a median observation of 5.6 years in alive patients, 178/583 patients had died (5-year OS: 72.9%; 10-year OS: 50.7%).

The validation series was represented by a cohort of 370 newly diagnosed and previously untreated CLL patients who consecutively presented for initial evaluation from June 1996 through June 2011 (Table S1) and provided with regular follow-up (at least three visits/year). The clinical and genetic features of the validation series are provided in table S1. The clinical database was updated in February 2012. After a median observation of 5.9 years in alive patients, 62/370 patients had died (5-year OS: 89.5%; 10-year OS: 65.8%).

Time-dependent analysis and analysis of clonal evolution were based on a mono-institutional cohort of 257 CLL out of the 637 cases of the training series. Inclusion criteria for time-dependent analysis were: *i)* having at least two years of follow-up after diagnosis; and either *ii)* availability of $\geq$2 sequential samples (patients: 202; sequential samples: 469; median interval between baseline and last sequential sample: 62.8 months) collected at: *a)* diagnosis; *b)* each progression requiring treatment; *c)* last follow-up; or *iii)* availability of the baseline sample collected at presentation in patients treated at diagnosis and persistently in continuous remission after first line treatment (patients: 55).

Mutation analysis

The mutation hotspots of *TP53* (exons 4-9, including splicing sites; RefSeq NM_000546.5), *NOTCH1* (exon 34; RefSeq NM_017617.2), *SF3B1* (exons 14, 15, 16, 18, including splice sites; RefSeq NM_012433.2), *MYD88* (exons 3, 5, including splicing sites; RefSeq NM_002468.4), and *BIRC3* (exons 6-9, including splicing sites; RefSeq NM_001165.4) genes were analyzed by PCR amplification and DNA direct sequencing of high molecular weight genomic DNA. Sequences for all annotated exons and flanking splice sites were retrieved from the UCSC Human Genome database using the corresponding mRNA accession number as a reference. PCR primers, located ~50 bp upstream or downstream to target exon boundaries, were either derived from previously published studies or designed in the Primer 3 program (http://frodo.wi.mit.edu/primer3/) and filtered using UCSC in silico PCR to exclude pairs yielding more than a single product. All PCR primers and conditions are available upon request. Purified amplicons were subjected to conventional DNA Sanger sequencing using the ABI PRISM 3100 Genetic Analyzer (Applied Biosystems), and compared to the corresponding germline sequences using the Mutation Surveyor Version 4.0.5 software package (SoftGenetics) after automated and/or manual curation. Of the evaluated sequences, 99% had a Phred score of 20 or more and 97% had a score of 30 or more. Candidate variants were confirmed from both strands on independent PCR products. The following databases were used to exclude known germline variants: Human dbSNP Database at NCBI (Build 136) (http://www.ncbi.nlm.nih.gov/snp); Ensembl Database (http://www.ensembl.org/index.html); The 1000 Genomes Project (http://www.1000genomes.org/); five single-genome projects available at the UCSC Genome Bioinformatics resource (http://genome.ucsc.edu/). Synonymous variants, previously reported germline polymorphisms and changes present in the matched normal DNA were removed from the analysis.[1–5]

***IGHV-IGHD-IGHJ* rearrangement analysis**

PCR amplification of *IGHV-IGHD-IGHJ* rearrangements was performed on HMW genomic DNA using IGHV leader primers or consensus primers for the IGHV FR1 along with appropriate IGHJ genes, as previously described.[6] PCR products were directly sequenced with the ABI PRISM BigDye Terminator v1.1 Ready Reaction Cycle Sequencing kit (Applied Biosystems) using the ABI PRISM 3100 Genetic Analyzer (Applied Biosystems). Sequences were analyzed using the IMGT databases and the IMGT/V-QUEST tool (version 3.2.17, Université Montpellier 2, CNRS, LIGM, Montpellier, France).

Fluorescence in situ hybridization (FISH)
Probes used for FISH analysis were: *i)* LSID13S319 (13q14 deletion), CEP12 (trisomy 12), LSIp53 (17p13/*TP53* deletion), and LSIATM (11q2-q23/*ATM* deletion) (Abbott, Rome, Italy); and *ii)* the RP11-177O8 (*BIRC3*) BAC clone.[4] The labeled BAC probe was tested against normal control metaphases to verify the specificity of the hybridization. For each probe, at least 400 interphase cells with well-delineated fluorescent spots were examined. Nuclei were counterstained with 4',6'-diamidino-2-phenylindole (DAPI) and antifade reagent, and signals were visualized using an Olympus BX51 microscope (Olympus Italia, Milan, Italy). The presence of 13q14 deletion, trisomy 12, 11q22-q23 deletion, 17p13 deletion, and *BIRC3* deletion was scored when the percentage of nuclei with the abnormality was above our internal cut off (5%, 5%, 7%, 10%, and 10% respectively), defined as the mean plus 3 standard deviations of the frequency of normal control cells exhibiting the abnormality.[4]

Prognostic factors
Fixed exposure variables collected at diagnosis and included in the dataset were age (continuous), gender (female vs male), Rai stage (0-I vs II vs III-IV), *IGHV* mutation status (*IGHV* homology <98% vs *IGHV* homology ≥98%), del13q14 (absent vs present), +12 (absent vs present), del11q22-q23 (absent vs present), del17p13 (absent vs present), mutation status of *TP53* (absent vs present), *NOTCH1* (absent vs present), *SF3B1* (absent vs present), *MYD88* (absent vs present) and *BIRC3* (absent vs present), *BIRC3* deletion (absent vs present), *TP53* disruption (absent vs present), *BIRC3* disruption (absent vs present) and the integrated mutational and cytogenetic model (very-low risk vs low-risk vs intermediate-rsik vs high-risk). Time-varying exposure variables repeatedly assessed at clinically relevant time points (i.e. disease progression or last follow-up) and registered in the dataset for the time-dependent analysis were age (≤65 years vs >65 years), Rai stage (0-I vs II vs III-IV), del11q22-q23 (absent vs present), del17p13 (absent vs present), mutation status of *TP53* (absent vs present), *NOTCH1* (absent vs present), *SF3B1* (absent vs present), *MYD88* (absent vs present) and *BIRC3* (absent vs present), *BIRC3* deletion (absent vs present), *TP53* disruption (absent vs present), *BIRC3* disruption (absent vs present) and the integrated mutational and cytogenetic model (very-low risk vs low-risk vs intermediate-rsik vs high-risk). The date of modification of these covariates during follow-up was also registered.

Proportional hazard regression
The proportional hazard assumption was assessed by plotting the smoothed Schoenfeld residuals against time. Assumption of linearity for continuous variable was assessed by plotting the smoothed martingale residuals.[7] Possible interactions were tested.[8] The bias corrected c-index and calibration slope of the Cox model were calculated through the .632 bootstrap method (1000 resamplings).[8,9] The heuristic shrinkage estimator was calculated using the formula (model likelihood ratio χ^2) - (number of degree of freedom in the model)/(model likelihood ratio χ^2).[8,10] This approach provides an estimate of prediction accuracy of the Cox model to protect against overfitting.

Internal validation
Internal validation was performed using a bootstrapping resampling procedure.[11] In the first step, 1000 bootstrap samples were generated randomly with replacement from the original CLL population. Cox regression was applied to each bootstrap sample with the same covariates as the original modeling. The percentage of bootstrap samples for which each covariate was selected as significant in the model was then calculated. Percent of selection reflects the prognostic importance of a covariate, because it is expected that an important covariate will be selected for the majority of bootstrap samples. In the second step, 1000 additional bootstrap samples were generated randomly with replacement from the original CLL population. Cox regression was applied to each bootstrap sample with the same covariates as the original modeling. For each covariate, the mean standard deviation and confidence intervals were computed for the 1000 bootstrap replications.

Recursive partitioning
Compared to the Cox-fitted model, recursive partitioning for survival data with censoring has the advantage of a more objective and non arbitrary construction of a hierarchical classification of covariates.[12,13] The first step in recursive partitioning analysis was to find the best split of the data into two groups (nodes) by the genetic predictor variable that captures the most information in the variability of OS. The process was recursively repeated, so succeeding steps find the best splits of the data within each of the nodes resulting from prior splits (daughter nodes). The entire dataset was considered as the primary node. Three major steps were utilized to derive the best decision tree: *i)* growing an initial tree under the following constraints and stopping rules: *a)* split criteria of $p<0.05$ according to the log-rank test adjusted for multiple comparisons; *b)* >20 patients in a node in order to be considered for splitting; *c)* >20 patients in a terminal node; *ii)* applying a pruning algorithm based on the complexity parameter

(cp=0.015); and *iii)* cross-validating the best tree size. Ten-fold cross-validation was used to determine the best tree size. The best number of splits was identified as that showing a cross-validation error lower than the smallest cross validation error + the corresponding standard error.

Random survival forest

A random survival forest was generated by drawing 1000 bootstrap samples from the original dataset and growing a tree for each bootstrapped sample.[14] Log-rank was utilized as splitting rule. At each branch, a random set of genetic variables were chosen as candidates to split the branch into two other branches, and the variable maximizing the log-rank statistic was used for splitting. The number of variables assessed at each branch was 4/5 of the total number of variables. Trees were grown under the constraint that a terminal node should have at least 1 unique event (death). The most important variables were identified as those that most frequently split the branches near the tree trunks.[15] There were no prespecified assumptions regarding variables, and randomization was introduced into this model by both random bootstrap sampling of patients from the original cohort and random sampling of variables for each tree branch. Importance of a variable was assessed by minimal depth from the tree trunk. The most predictive genetic variables for the cohort were defined as those whose minimal depth (averaged over the forest) was smaller than the mean minimal depth determined under the null hypothesis of no effect.

Amalgamation

Within a tree, any two terminal nodes of patients arising from the same parent node must be significantly different with respect to survival. The same is not necessarily true for clusters of patients arising from different parent nodes. To achieve better efficiency, an amalgamation algorithm was used to merge terminal nodes that showed homogenous survival.[12] At each step of the amalgamation algorithm, log-rank statistics testing the hypothesis that the survival of two nodes differed versus the hypothesis that they were the same was calculated, and the one with the greatest significance level was searched. If this maximal significance level was greater than 0.05, the two corresponding nodes were joined to form one new terminal node. The same procedure was then applied on the set of the new terminal nodes. This recursive amalgamation algorithm was stopped when no significance level greater than 0.05 was found.

Relative survival

Estimates of the expected survival were calculated utilizing Italian life tables stratified by age, sex and calendar year that were obtained from the Human Mortality Database (http://www.mortality.org/, accessed June 18 2012).[16] The major advantage of relative survival is that it provides a measure of the excess mortality experienced by CLL patients, irrespective of whether the excess mortality is directly or indirectly attributable to the disease.

SUPPLEMENTARY FIGURE LEGENDS

Fig S1. **Molecular complexity of newly diagnosed CLL.** Panel A. Heat map showing the mutual relationship of gene mutations and cytogenetic lesions in newly diagnosed CLL from the training series. Rows correspond to identical lesions, and columns represent individual patients color-coded based on the molecular status (white: absence of the lesion; red: presence of the lesion). Panel B. Prevalence of gene mutations and cytogenetic lesions in newly diagnosed CLL from the training series. Panel C. Circos plot representing the relative frequency and pairwise co-occurrence of gene mutations and cytogenetic lesions in newly diagnosed CLL from the training series. The length of the arc corresponds to the frequency of the genetic lesions. The width of the ribbon corresponds to the percentage of patients in which the two connected genetic lesions co-occur. Pairwise associations and anti-associations between genetic lesions were calculated with the use of Fisher's exact test and corrected for multiple hypothesis testing. The asterisk indicates statistically significant associations ($p<0.05$).

Fig S2. **Circos plots showing the pairwise co-occurrence of mutational and cytogenetic lesions in the training series.** The length of the arc corresponds to the frequency of the genetic lesions. The width of the ribbon corresponds to the percentage of patients in which the two connected genetic lesions co-occurred.

Fig S3. **Kaplan-Meier estimates of overall survival (OS) according to the FISH cytogenetic model in the training series.** Cases harboring del17p13 irrespective of co-occurring cytogenetic lesions are represented by the red line. Cases harboring del11q22-q23 in the absence of del17p13 are represented by the purple line. Cases harboring +12 in the absence of del17p13 and del11q22-q23 are represented by the yellow line. Cases harboring a normal FISH karyotype are represented by the green line. Cases harboring del13q14 deletion in the absence of other cytogenetic abnormalities are represented by the blue line. Nr=not reached.

Fig S4. **Kaplan-Meier estimates of overall survival (OS) and treatment free survival according to the integrated mutational and cytogenetic model in early stage CLL from the training series.** Panel A. Overall survival (OS). Panel B. Probability of progressive disease requiring treatment according to IWCLL-NCI guidelines as indicated by treatment free interval. Cases harboring *TP53* and/or *BIRC3* disruption (*TP53* DIS/*BIRC3* DIS) independent of co-occurring genetic lesions are represented by the red line. Cases harboring *NOTCH1* mutations (*NOTCH1* M) and/or *SF3B1* mutations (*SF3B1* M) and/or del11q22-q23 in the absence of *TP53* and *BIRC3* disruption are represented by the yellow line. Cases harboring +12 in the absence of *TP53* disruption, *BIRC3* disruption, *NOTCH1* mutations, *SF3B1* mutations and del11q22-q23, and cases wild type for all genetic lesions (Normal) are represented by the green line. Cases harboring del13q14 as the sole genetic lesion are represented by the blue line. Nr=not reached.

Fig S5. **Kaplan-Meier estimates of overall survival (OS) according to the integrated mutational and cytogenetic model in CLL diagnosed in 2005 or afterwards (panel A) and in CLL treated with fludarabine-cyclophosphamide-rituximab.** Cases harboring *TP53* and/or *BIRC3* disruption (*TP53* DIS/*BIRC3* DIS) independent of co-occurring genetic lesions are represented by the red line. Cases harboring *NOTCH1* mutations (*NOTCH1* M) and/or *SF3B1* mutations (*SF3B1* M) and/or del11q22-q23 in the absence of *TP53* and *BIRC3* disruption are represented by the yellow line. Cases harboring +12 in the absence of *TP53* disruption, *BIRC3* disruption, *NOTCH1* mutations, *SF3B1* mutations and del11q22-q23, and cases wild type for all genetic lesions (Normal) are represented by the green line. Cases harboring del13q14 as the sole genetic lesion are represented by the blue line.

Fig S6. **Kaplan-Meier estimates of overall survival (OS) according to the integrated mutational and cytogenetic model in the validation series.** Cases harboring *TP53* and/or *BIRC3* disruption (*TP53* DIS/*BIRC3* DIS) independent of co-occurring genetic lesions are represented by the red line. Cases harboring *NOTCH1* mutations (*NOTCH1* M) and/or *SF3B1* mutations (*SF3B1* M) and/or del11q22-q23 in the absence of *TP53* and *BIRC3* disruption are represented by the yellow line. Cases harboring +12 in the absence of *TP53* disruption, *BIRC3* disruption, *NOTCH1* mutations, *SF3B1* mutations and del11q22-q23,

and cases wild type for all genetic lesions (Normal) are represented by the green line. Cases harboring del13q14 as the sole genetic lesion are represented by the blue line. Nr=not reached.

Fig S7. Comparison of the variant allele frequency of *TP53*, *NOTCH1*, *SF3B1*, *MYD88* and *BIRC3* mutations across sequential samples. Modifications of the mutant allele frequency during disease course was estimated by next generation sequencing in 79 paired sequential samples from 56 patients (median depth of coverage: 719x; range: 479-998x). The red box plots represent the baseline samples. The yellow box plots represent the sequential samples. The bottom and top of the box represents the 25th and 75th percentile. The band near the middle of the box represents the median. The ends of the whiskers represent the smallest and the largest value not classified as outliers. Circles and stars represent outlier values.

Fig S8. Clonal fluctuation of genetic lesions in three CLL patients. Patient 1 lost the c.1996A>G p.K666E mutation of *SF3B1* but concomitantly developed a new c.2098A>G p.K700E mutation of *SF3B*1. Patient 24 lost the c.637C>T p.R213* mutation of *TP53* but concomitantly developed a new c.743G>A p.R248Q mutation of *TP53*. Patient 22 lost the c.7544_7545delCT p.P2515fs*4 mutation of *NOTCH1*, but concomitantly developed the c.2098A>G p.K700E mutation of *SF3B1*. Genetic lesions are reported on the vertical axis. Follow-up is reported on the horizontal axis. Circles are color-coded based on the molecular status (white: absence of the lesion; red: presence of the lesion). Arrows represent clinically relevant time points. FC, fludarabine-cyclophosphamide; FCR, fludarabine-cyclophosphamide-ristuximab; BA, bendamustine-alemtuzumab, R-CHOP, rituximab-cyclophosphamide-doxorubicin-vincristine-prednisone; R-DHAP, rituximab-dexamethasone-high dose cyterabine-cisplatin.

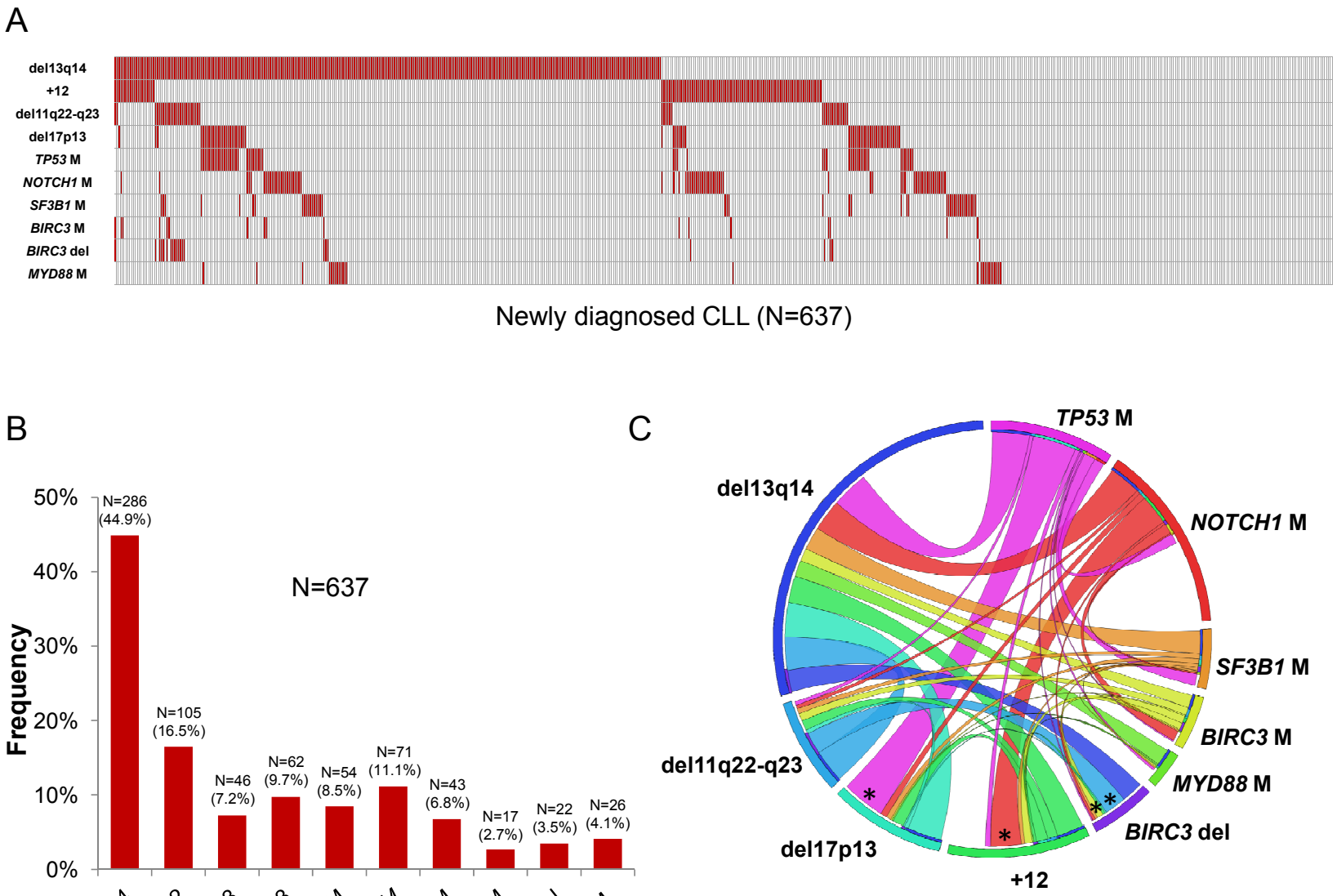

A
del13q14
+12
del11q22-q23
del17p13
TP53 M
NOTCH1 M
SF3B1 M
BIRC3 M
BIRC3 del
MYD88 M
Newly diagnosed CLL (N=637)
B
Frequency
50%
40%
30%
20%
10%
0%
N=637
N=286 (44.9%)
N=105 (16.5%)
N=46 (7.2%)
N=62 (9.7%)
N=54 (8.5%)
N=71 (11.1%)
N=43 (6.8%)
N=17 (2.7%)
N=22 (3.5%)
N=26 (4.1%)
del13q14
+12
del11q22-q23
del17p13
TP53 M
NOTCH1 M
SF3B1 M
BIRC3 M
BIRC3 del
MYD88 M
C
TP53 M
del13q14
NOTCH1 M
SF3B1 M
BIRC3 M
del11q22-q23
MYD88 M
BIRC3 del
del17p13
+12

Figure S1

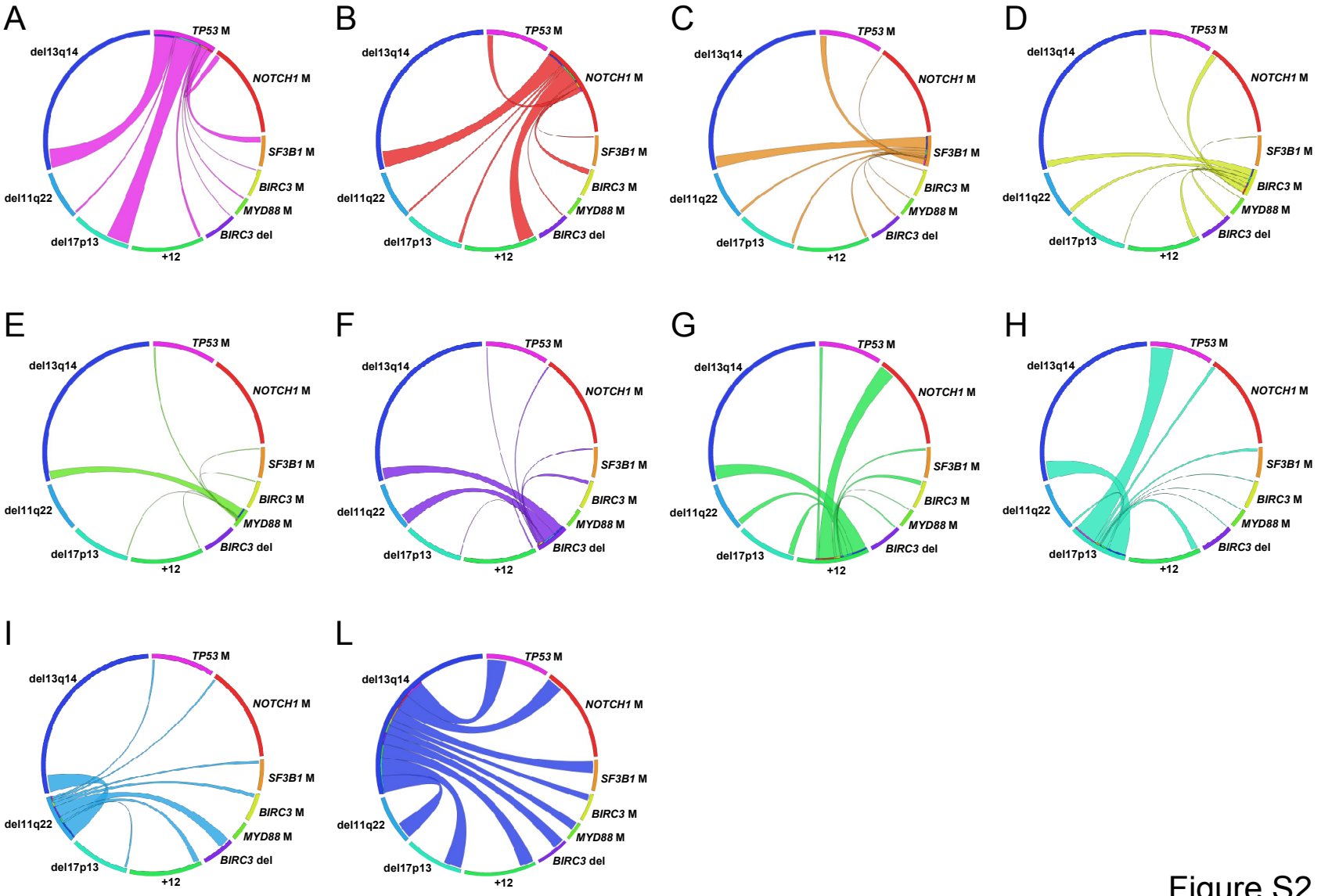
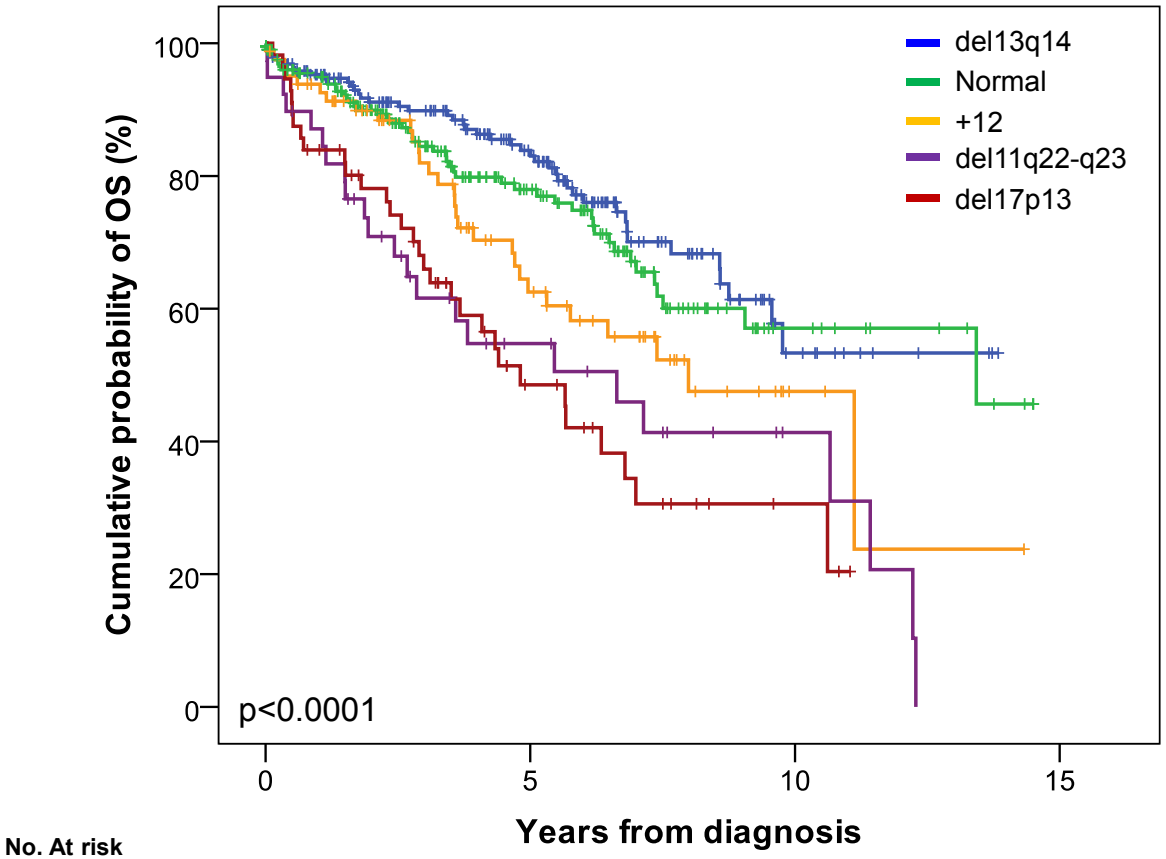

A
B
C
D
E
F
G
H
I
L
TP53 M
NOTCH1 M
SF3B1 M
BIRC3 M
MYD88 M
BIRC3 del
+12
del17p13
del11q22
del13q14

Figure S2



No. At risk

	194	99	11	0
	212	79	12	0
	82	32	3	0
	39	14	4	0
	56	16	3	0

	Events	Total	Median	95% CI
	44	194	nr	-
	50	212	13.4	10.2-16.6
	30	82	7.8	5.7-10.2
	23	39	8.8	2.8-10.4
	31	56	4.8	3.2-6.4

Figure S3

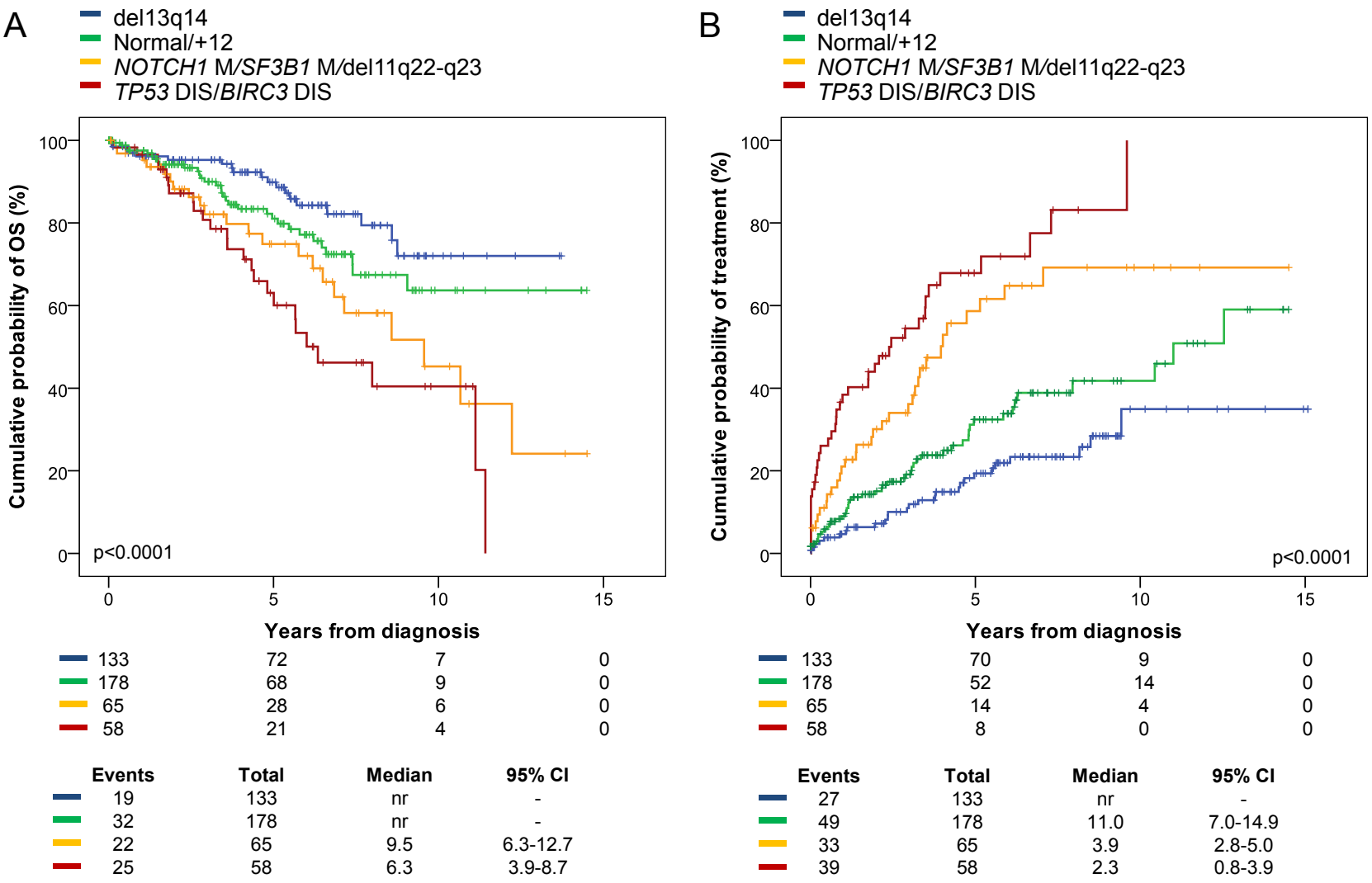

A
del13q14
Normal/+12
NOTCH1 M/SF3B1 M/del11q22-q23
TP53 DIS/BIRC3 DIS
Cumulative probability of OS (%)
100
80
60
40
20
0
p<0.0001
0
5
10
15
Years from diagnosis
133 72 7 0
178 68 9 0
65 28 6 0
58 21 4 0
Events Total Median 95% CI
19 133 nr -
32 178 nr -
22 65 9.5 6.3-12.7
25 58 6.3 3.9-8.7
B
del13q14
Normal/+12
NOTCH1 M/SF3B1 M/del11q22-q23
TP53 DIS/BIRC3 DIS
Cumulative probability of treatment (%)
100
80
60
40
20
0
p<0.0001
0
5
10
15
Years from diagnosis
133 70 9 0
178 52 14 0
65 14 4 0
58 8 0 0
Events Total Median 95% CI
27 133 nr -
49 178 11.0 7.0-14.9
33 65 3.9 2.8-5.0
39 58 2.3 0.8-3.9

Figure S4

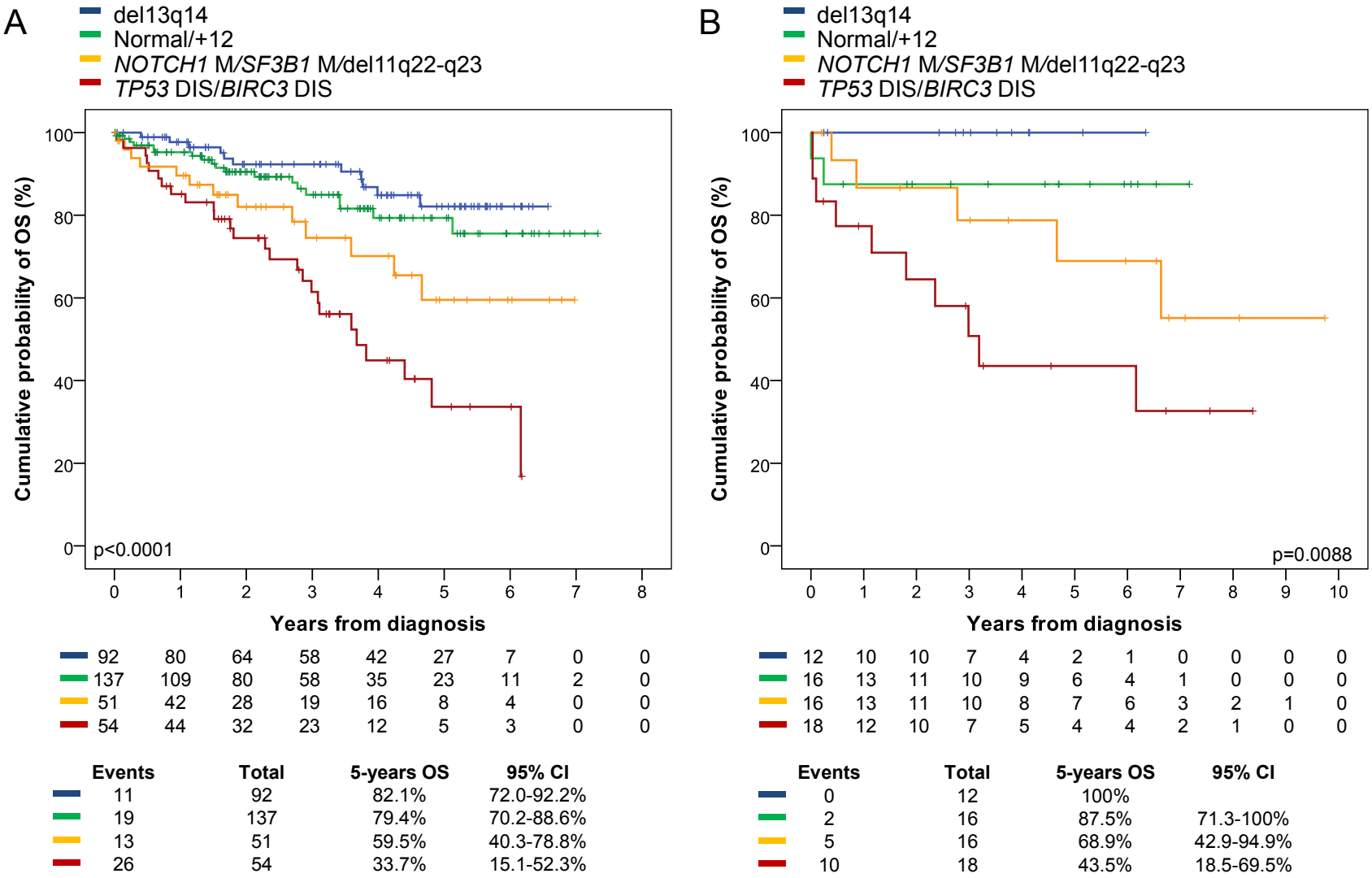
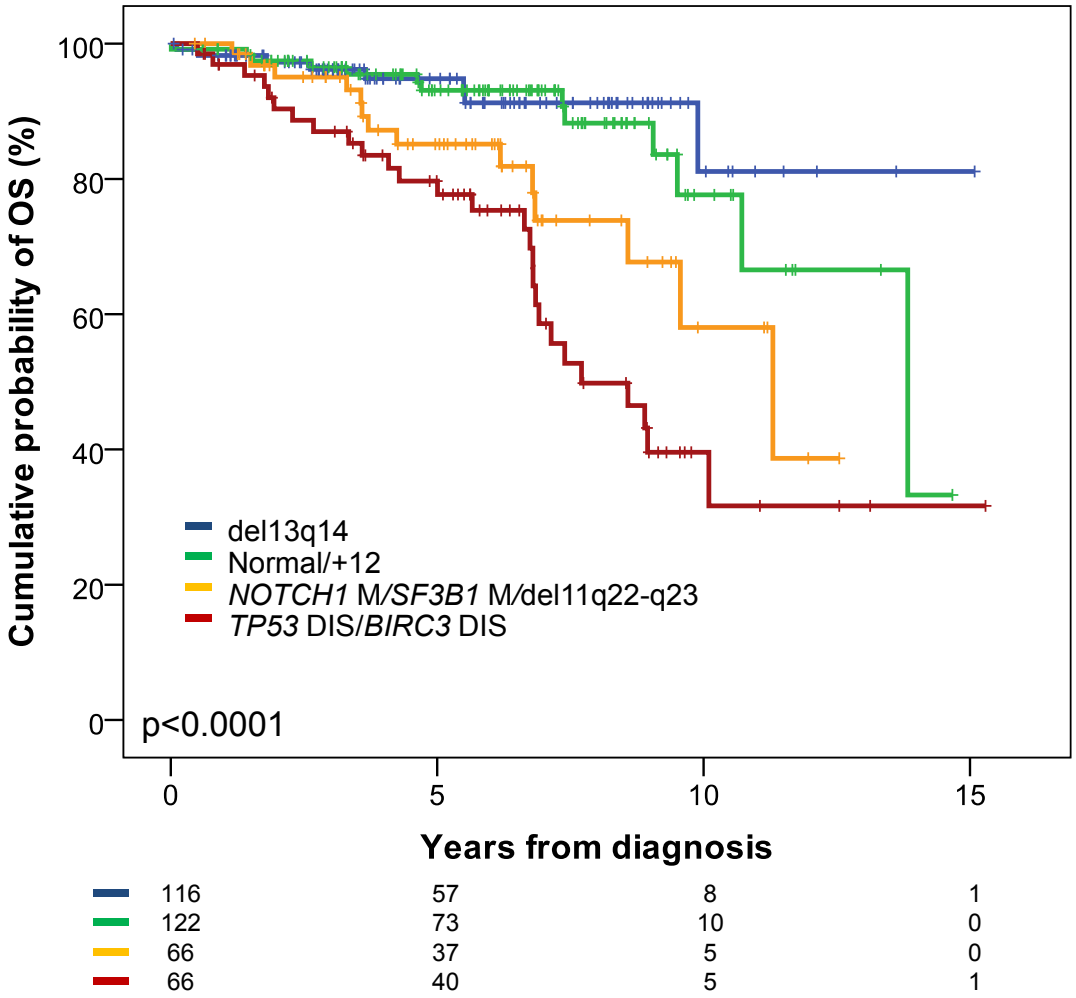
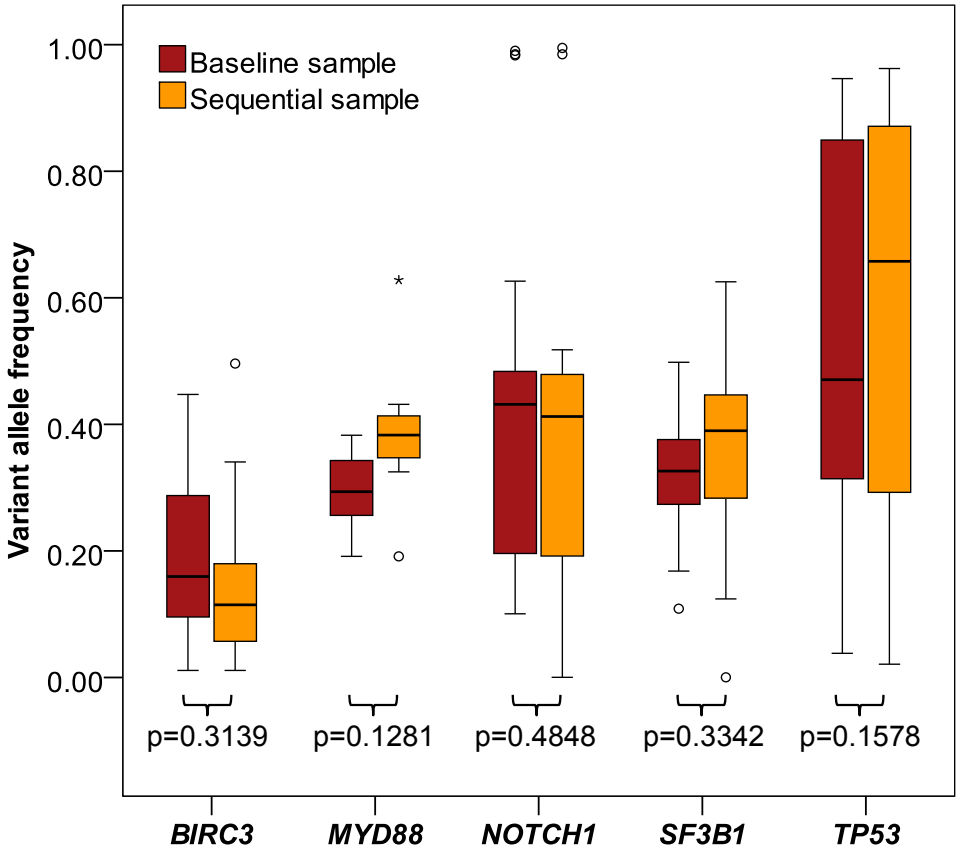

A
del13q14
Normal/+12
NOTCH1 M/SF3B1 M/del11q22-q23
TP53 DIS/BIRC3 DIS
Cumulative probability of OS (%)
100
80
60
40
20
0
p<0.0001
0 1 2 3 4 5 6 7 8
Years from diagnosis
92 80 64 58 42 27 7 0 0
137 109 80 58 35 23 11 2 0
51 42 28 19 16 8 4 0 0
54 44 32 23 12 5 3 0 0
Events Total 5-years OS 95% CI
11 92 82.1% 72.0-92.2%
19 137 79.4% 70.2-88.6%
13 51 59.5% 40.3-78.8%
26 54 33.7% 15.1-52.3%
B
del13q14
Normal/+12
NOTCH1 M/SF3B1 M/del11q22-q23
TP53 DIS/BIRC3 DIS
Cumulative probability of OS (%)
100
80
60
40
20
0
p=0.0088
0 1 2 3 4 5 6 7 8 9 10
Years from diagnosis
12 10 10 7 4 2 1 0 0 0 0
16 13 11 10 9 6 4 1 0 0 0
16 13 11 10 8 7 6 3 2 1 0
18 12 10 7 5 4 4 2 1 0 0
Events Total 5-years OS 95% CI
0 12 100%
2 16 87.5% 71.3-100%
5 16 68.9% 42.9-94.9%
10 18 43.5% 18.5-69.5%

Figure S5



	Events	Total	Median	95% CI
del13q14	8	116	nr	-
Normal/+12	13	122	13.8	9.3-18.2
NOTCH1 M/*SF3B1* M/del11q22-q23	14	66	11.2	8.1-14.4
TP53 DIS/*BIRC3* DIS	27	66	7.7	5.5-9.8

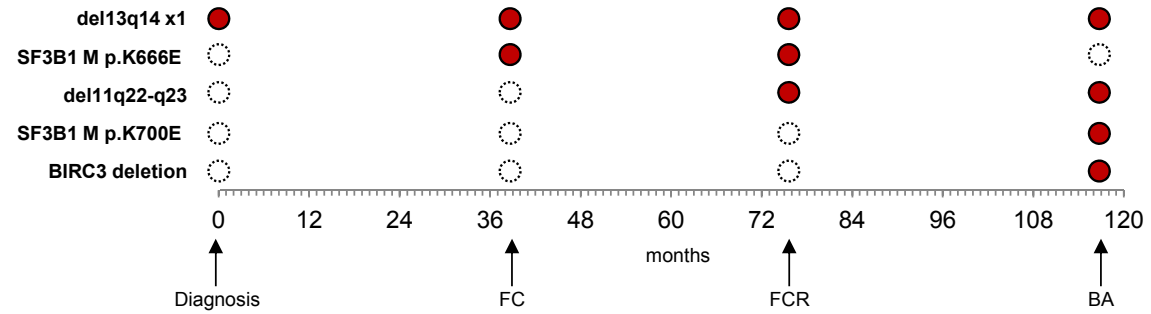
Figure S6



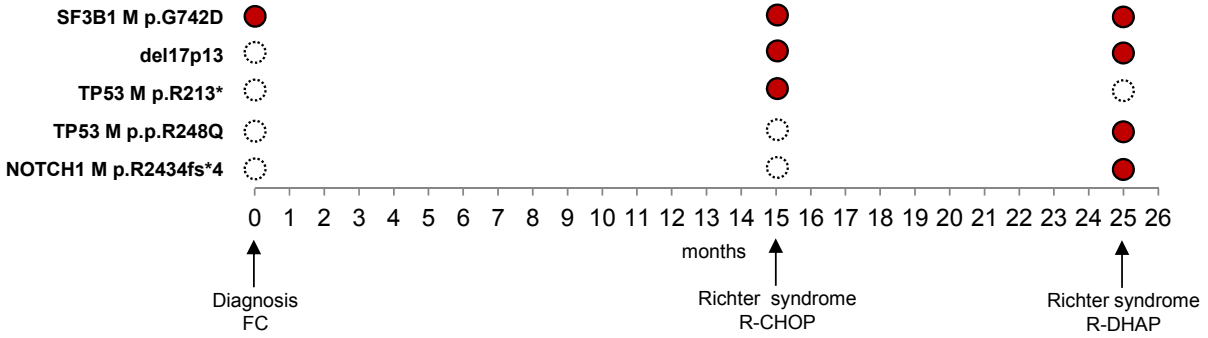
	Variant allele frequency	
	Baseline sample	Sequential sample
Gene	Median	Median
BIRC3	0.15	0.11
MYD88	0.29	0.38
NOTCH1	0.43	0.41
SF3B1	0.32	0.39
TP53	0.47	0.65

Figure S7

Patient 1



Patient 24



Patient 22

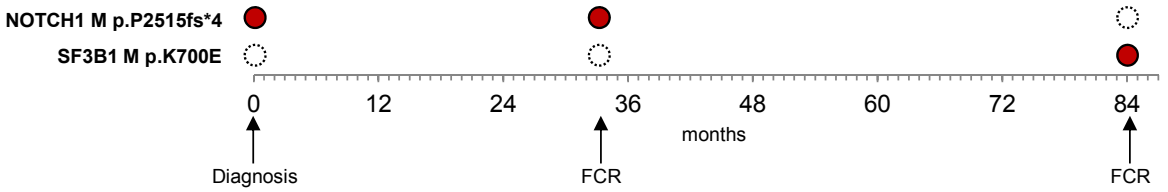


Figure S8

Table S1. Clinical features at presentation of the CLL cohorts [a]

	Training series									Validation series			
	Molecular analysis			Survival analysis			Clonal evolution analysis						
	N	Total	%	N	Total	%	N	Total	%	N	Total	%	p [b]
Age >65 years	326	637	51.2%	312	583	53.5%	177	257	68.9%	189	370	51.1%	0.4631
Male gender	372	637	58.4%	341	583	58.5%	143	257	55.6%	207	370	55.9%	0.4387
Rai stage													
0-I	477	637	74.9%	434	583	74.4%	186	257	72.4%	279	370	75.4%	0.7386
II	73	637	11.5%	69	583	11.8%	27	257	10.5%	64	370	17.3%	0.0177
III-IV	87	637	13.7%	80	583	13.7%	44	257	17.1%	27	370	7.3%	0.0022
***IGHV* homology ≥98%**	247	628	39.3%	234	575	40.7%	90	250	36.0%	147	368	39.9%	0.8189
del13q14	286	637	44.9%	257	583	44.1%	128	257	49.8%	194	370	52.4%	0.0119
+12	105	637	16.5%	97	583	16.6%	51	257	19.8%	46	370	12.4%	0.0764
del11q22-q23	46	637	7.2%	42	583	7.2%	19	257	7.4%	41	370	11.1%	0.0386
del17p13	62	637	9.7%	56	583	9.6%	26	257	10.1%	25	370	6.8%	0.1243
FISH stratification													
del13q14	217	637	34.1%	194	583	33.3%	95	257	37.0%	148	370	40.0%	0.0349
Normal	226	637	35.5%	212	583	36.4%	77	257	30.0%	115	370	31.1%	0.0941
+12	89	637	14.0%	82	583	14.1%	41	257	16.0%	43	370	11.6%	0.1381
del11q22-q23	43	637	6.8%	39	583	6.7%	18	257	7.0%	39	370	10.5%	0.0346
del17p13	62	637	9.7%	56	583	9.6%	26	257	10.1%	25	370	6.8%	0.1243
***TP53* mutation**	54	637	8.5%	50	583	8.6%	27	257	10.5%	33	370	8.9%	0.8550
***NOTCH1* mutation**	71	637	11.1%	67	583	11.5%	29	257	11.3%	37	370	10.0%	0.4715
***SF3B1* mutation**	43	637	6.8%	41	583	7.0%	17	257	6.6%	24	370	6.5%	0.7445
***MYD88* mutation**	26	637	4.1%	24	583	4.1%	9	257	3.5%	-	-	-	
***BIRC3* mutation**	17	637	2.7%	14	583	2.4%	7	257	2.7%	18	370	4.9%	0.0396
***BIRC3* deletion**	22	637	3.5%	20	583	3.4%	9	257	3.5%	11	370	2.9%	0.6401

[a] *IGHV*, immunoglobulin heavy variable gene; FISH, fluorescence in situ hybridization

[b] Comparison between the Survival analysis cohort of the Training series and the Validations series

Table S2. *TP53*, *NOTCH1*, *SF3B1*, *BIRC3* and *MYD88* somatic mutations in the training series

Sample ID	Gene	Nucleotode change	Amino acid change	COSMIC^	RefSeq
14450	BIRC3*	c.1101_1132del32	p.G367fs*6	N	NM_001165.4
14263	BIRC3*	c.1270G>T; c.1183_1352del4894	p.E424*; p.V395fs*78	N; N	NM_001165.4
3878	BIRC3*	c.1279_1280insA	p.I427fs*11	N	NM_001165.4
3714	BIRC3*	c.1638_1639insA	p.Q547fs*12	N	NM_001165.4
5889	BIRC3*	c.1663_1666del4	p.R555fs*12	N	NM_001165.4
14281	BIRC3	c.1281_1285del5	p.I427fs*9	N	NM_001165.4
9696	BIRC3	c.1282delA	p.R428fs*19	N	NM_001165.4
12684	BIRC3	c.1284_1288del5	p.R428fs*8	N	NM_001165.4
6070	BIRC3	c.1313_1314delAA	p.E438fs*13	N	NM_001165.4
12915	BIRC3	c.1633G>T	p.E545*	N	NM_001165.4
12857	BIRC3	c.1639delC	p.Q547fs*21	N	NM_001165.4
5114	BIRC3	c.1641delA	p.Q547fs*21	N	NM_001165.4
10471	BIRC3	c.1658_1661del4	p.E553fs*22	N	NM_001165.4
10755	BIRC3	c.1660G>T	p.E554*	N	NM_001165.4
12914	BIRC3	c.1672A>G	p.K558E	N	NM_001165.4
12904	BIRC3	c.1748_1780del33	p.P583fs*12	N	NM_001165.4
8667	BIRC3	c.1798C>G	p.R600G	N	NM_001165.4
5114	MYD88	c.649G>T	p.V217F	Y	NM_002468.4
5096	MYD88	c.649G>T	p.V217F	Y	NM_002468.4
4848	MYD88	c.649G>T	p.V217F	Y	NM_002468.4
7835	MYD88	c.649G>T	p.V217F	Y	NM_002468.4
12664	MYD88	c.649G>T	p.V217F	Y	NM_002468.4
14525	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
14534	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
10628	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
3560	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
10894	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
8436	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
10879	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
10637	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
3501	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
3722	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
5170	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
3965	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
4048	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
8762	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
4767	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
12639	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
6525	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
10865	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
12924	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
12689	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
12671	MYD88	c.794T>C	p.L265P	Y	NM_002468.4
6727	NOTCH1**	c.6485_6847del363	p.P2162del122	N	NM_017617.2
12857	NOTCH1**	c.6802_6803delGA	p.E2268fs*86	N	NM_017617.2
12685	NOTCH1**	c.6987_6988insG	p.S2330fs*25	Y	NM_017617.2
12907	NOTCH1**	c.7006_7007insC	p.L2336fs*19	Y	NM_017617.2
7398	NOTCH1**	c.7023_7024ins4	p.S2342fs*13	Y	NM_017617.2
6426	NOTCH1**	c.7247_7274del28	p.P2416fs*11	N	NM_017617.2
12884	NOTCH1**	c.7250_7251insCAC	p.Q2417>HT	N	NM_017617.2
7140	NOTCH1**	c.7389_7390CG>T	p.P2463fs*15	Y	NM_017617.2
3878	NOTCH1**	c.7391delC	p.A2464fs*14	Y	NM_017617.2
7140	NOTCH1**	c.7411_7429del19	p.S2471fs*1	N	NM_017617.2
3985	NOTCH1**	c.7433delC	p.T2478fs*6	N	NM_017617.2
5675	NOTCH1**	c.7446delC	p.F2482fs*2	N	NM_017617.2
12124	NOTCH1	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
6070	NOTCH1	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
8667	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
3494	NOTCH1	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
11730	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
5303	NOTCH1	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
11724	NOTCH1	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
9273	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
5368	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
3724	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
9930	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
5889	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
5825	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2

4261	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
5726	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
12174	NOTCH1	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
4341	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
5984	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
4844	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
5765	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
3975	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
3701	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
3648	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
4233	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
6892	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
3706	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
3392	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
6624	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
11815	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
10267	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
4800	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
10320	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
5565	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
13692	NOTCH1	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
14260	NOTCH1	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
3121	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
3979	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
12926	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
12858	NOTCH1	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
12915	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
12847	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
12875	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
12864	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
12887	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
12922	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
12831	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
12839	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
12916	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
12834	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
12832	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
12684	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
10471	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
12686	NOTCH1	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
12660	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
7831	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
12810	NOTCH1**	c.7544_7545delCT	p.P2515fs*4	Y	NM_017617.2
14532	NOTCH1	c.7321C>T	p.Q2441*	Y	NM_017617.2
14524	NOTCH1	c.7544_7545delCT	p.2515fs*4	Y	NM_017617.2
14547	NOTCH1	c.7544_7545delCT	p.2515fs*4	Y	NM_017617.2
4845	SF3B1***	c.1890A>T	p.R630S	Y	NM_012433.2
7228	SF3B1***	c.1986C>A	p.H662Q	Y	NM_012433.2
4919	SF3B1	c.1986C>A	p.H662Q	Y	NM_012433.2
11772	SF3B1***	c.1996A>G	p.K666E	Y	NM_012433.2
7040	SF3B1***	c.1996A>G	p.K666E	Y	NM_012433.2
9094	SF3B1***	c.1998G>T	p.K666N	Y	NM_012433.2
5565	SF3B1***	c.2094_2101del6	p.delQ699_K700	Y	NM_012433.2
11480	SF3B1***	c.2098A>G	p.K700E	Y	NM_012433.2
4602	SF3B1***	c.2098A>G	p.K700E	Y	NM_012433.2
7561	SF3B1***	c.2098A>G	p.K700E	Y	NM_012433.2
4681	SF3B1***	c.2098A>G	p.K700E	Y	NM_012433.2
11197	SF3B1***	c.2098A>G	p.K700E	Y	NM_012433.2
10676	SF3B1***	c.2098A>G	p.K700E	Y	NM_012433.2
3981	SF3B1***	c.2098A>G	p.K700E	Y	NM_012433.2
11785	SF3B1***	c.2098A>G	p.K700E	Y	NM_012433.2
11196	SF3B1***	c.2098A>G	p.K700E	Y	NM_012433.2
4938	SF3B1	c.2111T>A	p.I704N	Y	NM_012433.2
3726	SF3B1	c.2219G>A	p.G740E	Y	NM_012433.2
3950	SF3B1***	c.2219G>A	p.G740E	Y	NM_012433.2
9201	SF3B1	c.2225G>A	p.G742D	Y	NM_012433.2
5965	SF3B1	c.2225G>A	p.G742D	Y	NM_012433.2
10642	SF3B1	c.2225G>A	p.G742D	Y	NM_012433.2
6434	SF3B1	c.2225G>A	p.G742D	Y	NM_012433.2
14526	SF3B1	c.1866G>T	p.E622D	Y	NM_012433.2

14529	SF3B1	c.1866G>T	p.E622D	Y	NM_012433.2
7003	SF3B1	c.1996A>G	p.K666E	Y	NM_012433.2
9022	SF3B1	c.2008C>G	p.Q670E	Y	NM_012433.2
9614	SF3B1	c.2097_2104delAAAGTT	p.delK700_V701	N	NM_012433.2
10755	SF3B1	c.2098A>G	p.K700E	Y	NM_012433.2
10356	SF3B1	c.2098A>G	p.K700E	Y	NM_012433.2
4570	SF3B1	c.2098A>G	p.K700E	Y	NM_012433.2
4463	SF3B1	c.2098A>G	p.K700E	Y	NM_012433.2
12919	SF3B1	c.2098A>G	p.K700E	Y	NM_012433.2
12845	SF3B1	c.2098A>G	p.K700E	Y	NM_012433.2
10470	SF3B1	c.2098A>G	p.K700E	Y	NM_012433.2
12646	SF3B1	c.2098A>G	p.K700E	Y	NM_012433.2
12689	SF3B1	c.2098A>G	p.K700E	Y	NM_012433.2
3451	SF3B1	c.2110_2111insCCA	p.T703insP	N	NM_012433.2
5125	SF3B1	c.2110A>T	p.I704F	N	NM_012433.2
8990	SF3B1	c.2110A>T	p.I704F	N	NM_012433.2
9767	SF3B1	c.2110A>T	p.I704F	N	NM_012433.2
12837	SF3B1	c.2110A>T	p.I704F	N	NM_012433.2
14549	SF3B1	c.2225G>A	p.G742D	Y	NM_012433.2
4764	TP53	c.150delT	p.I50fs*73	Y	NM_000546.5
12682	TP53	c.217G>A	p.V73M	Y	NM_000546.5
12889	TP53	c.290_306del17	p.V97fs*46	N	NM_000546.5
5125	TP53	c.329G>A	p.R110H	Y	NM_000546.5
9582	TP53	c.329G>T	p.R110L	Y	NM_000546.5
3560	TP53	c.334G>A	p.G112S	Y	NM_000546.5
12926	TP53	c.334G>A	p.G112S	Y	NM_000546.5
8046	TP53	c.365_366delTG	p.V122fs*26	Y	NM_000546.5
4831	TP53	c.377A>G	p.Y126C	Y	NM_000546.5
4232	TP53	c.404G>T	p.C135F	Y	NM_000546.5
12833	TP53	c.422G>A	p.C141Y	Y	NM_000546.5
10553	TP53	c.452C>G	p.P151R	Y	NM_000546.5
4274	TP53	c.455_456insC	p.P152fs*17	N	NM_000546.5
7812	TP53	c.472C>A; c.658T>C	p.R158S; p.Y220H	N; Y	NM_000546.5
7911	TP53	c.473G>A	p.R158H	Y	NM_000546.5
4845	TP53	c.536A>G	p.H179R	Y	NM_000546.5
5538	TP53	c.536A>G	p.H179R	Y	NM_000546.5
4997	TP53	c.536A>T	p.H179L	Y	NM_000546.5
10470	TP53	c.550_553delGATA	p.D184fs*62	Y	NM_000546.5
9930	TP53	c.569_581del13	p.P190fs*53	N	NM_000546.5
12879	TP53	c.614A>G	p.Y205C	Y	NM_000546.5
12920	TP53	c.625_626delAG	p.R209fs*6	Y	NM_000546.5
7916	TP53	c.646G>A	p.V216M	Y	NM_000546.5
5167	TP53	c.673-2A>G	splice site	N	NM_000546.5
3724	TP53	c.701A>G	p.Y234C	Y	NM_000546.5
4270	TP53	c.701A>G	p.Y234C	Y	NM_000546.5
12883	TP53	c.701A>G c.747G>T	p.Y234C; p.R249S	Y; Y	NM_000546.5
7269	TP53	c.710G>T	p.M237I	Y	NM_000546.5
10472	TP53	c.715A>G	p.N239D	Y	NM_000546.5
6342	TP53	c.716A>T	p.N239I	N	NM_000546.5
9022	TP53	c.721T>G	p.S241A	Y	NM_000546.5
3204	TP53	c.733G>A	p.G245S	Y	NM_000546.5
14532	TP53	c.733G>T	p.G245C	Y	NM_000546.5
11325	TP53	c.734G>T	p.G245V	Y	NM_000546.5
12857	TP53	c.740delA	p.N247fs*98	Y	NM_000546.5
7912	TP53	c.742C>T	p.R248W	Y	NM_000546.5
12891	TP53	c.743G>A	p.R248Q	Y	NM_000546.5
6270	TP53	c.747G>T	p.R249S	Y	NM_000546.5
14549	TP53	c.790delC	p.L264fs*81	Y	NM_000546.5
12646	TP53	c.794T>C	p.L265P	Y	NM_000546.5
5967	TP53	c.796G>A	p.G266R	Y	NM_000546.5
11480	TP53	c.797G>A	p.G266E	Y	NM_000546.5
12828	TP53	c.797G>A	p.G266E	Y	NM_000546.5
6012	TP53	c.817C>A	p.R273S	Y	NM_000546.5
3955	TP53	c.817C>T	p.R273C	Y	NM_000546.5
10628	TP53	c.824G>A	p.C275Y	Y	NM_000546.5
5141	TP53	c.824G>A	p.C275Y	Y	NM_000546.5
9878	TP53	c.843C>G	p.D281E	Y	NM_000546.5
4602	TP53	c.848G>A	p.R283H	Y	NM_000546.5
12858	TP53	c.854_855insT	p.E285fs*20	Y	NM_000546.5
12124	TP53	c.856G>A	p.E286K	Y	NM_000546.5

5565	TP53	c.869G>A	p.R290H	Y	NM_000546.5
14542	TP53	c.911_915del4	p.T304fs*91	N	NM_000546.5
5570	TP53	c.918+1G>T	splice site	N	NM_000546.5

* Mutations previously reported in Rossi D, et al. Blood 2012; 119:2854-2862
** Mutations previously reported in Rossi D, et al. Blood 2012; 119:521-529
*** Mutations previously reported in Rossi D, et al. Blood 2011; 118:6904-6908
^Genes annotated in the COSMIC database v60; Y, yes; N, no

Table S3. Time-fixed analysis of the impact of genetic lesions on OS in the training series [a]

				Univariate analysis				Multivariate analysis				Internal bootstrapping validation			
												Bootstrap parameters (mean)			
	Events	Total	5-year OS	HR	LCI	UCI	p	HR	LCI	UCI	p	HR	LCI	UCI	Bootstrap selection
No del13q14	104	326	70.2%	1.00	-	-	0.2349	1.00	-	-	0.6777	1.00	-	-	28.5%
del13q14	74	257	76.0%	0.83	0.62	1.12		0.93	0.68	1.29		0.95	0.68	1.32	
No +12	138	486	75.7%	1.00	-	-	0.0200	1.00	-	-	0.0326	1.00	-	-	75.8%
+12	40	97	59.4%	1.52	1.07	2.16		1.52	1.04	2.22		1.51	1.05	2.29	
No del11q22-q23	154	541	74.4%	1.00	-	-	0.0011	1.00	-	-	0.0188	1.00	-	-	81.9%
del11q22-q23	24	42	54.8%	2.06	1.33	3.17		1.83	1.10	3.02		1.89	1.12	3.18	
No *NOTCH1* mutation	145	516	75.2%	1.00	-	-	0.0027	1.00	-	-	0.0131	1.00	-	-	84.2%
***NOTCH1* mutation**	33	67	56.3%	1.79	1.22	2.61		1.65	1.11	2.46		1.69	1.13	2.54	
No *SF3B1* mutation	160	542	73.8%	1.00	-	-	0.0098	1.00	-	-	0.0159	1.00	-	-	78.1%
***SF3B1* mutation**	18	41	60.3%	1.91	1.17	3.11		1.87	1.12	3.11		1.92	1.14	3.23	
No *MYD88* mutation	174	559	72.0%	1.00	-	-	0.1728	1.00	-	-	0.5205	1.00	-	-	22.8%
***MYD88* mutation**	4	24	95.8%	0.50	0.19	1.35		0.72	0.26	1.96		0.73	0.28	1.98	
No *TP53* disruption	136	509	76.5%	1.00	-	-	<0.0001	1.00	-	-	<0.0001	1.00	-	-	99.9%
***TP53* disruption**	42	74	50.7%	2.54	1.79	3.59		2.49	1.72	3.59		2.54	1.75	3.70	
No *BIRC3* disruption	161	553	74.0%	1.00	-	-	0.0019	1.00	-	-	0.0396	1.00	-	-	70.6%
***BIRC3* disruption**	17	30	52.5%	1.91	1.17	3.11		1.90	1.03	3.50		2.00	1.07	3.74	

[a]OS, overall survival; HR, hazard ratio; LCI, 95% lower confidence interval; UCI, 95% upper confidence interval
Shrinkage coefficient: 0.86
Discrimination: bias-corrected c-index: 0.639; optimism: 0.016
Calibration: bias-corrected calibration slope: 0.875; optimism: 0.125

Table S4. Variable importance measure for the random survival forest model in the training series

Top variables	Minimal depth
TP53 disruption	0.512
BIRC3 disruption	1.766
SF3B1 mutation	1.864
NOTCH1 mutation	1.880
del11q22-q23	1.902
+12	2.150

Variables included in the analysis: *TP53* disruption, *BIRC3* disruption, *SF3B1* mutation, *NOTCH1* mutation, del11q22-q23, +12, del13q14
Mean minimal depth threshold: 2·103

Table S5. Distribution of clinical and biological features across subgroups defined by the mutational and cytogenetic model in the training series [a]

	del13q14			Normal/+12			*NOTCH1* M/*SF3B1* M/del11q22-q23			*TP53* DIS/*BIRC3* DIS			
	N	**Total**	**%**	**N**	**Total**	**%**	**N**	**Total**	**%**	**N**	**Total**	**%**	**p**
Median age (yr)	66.5 (median)			66.6 (median)			67.6 (median)			67.9 (median)			0.7931
Male gender	81	155	52.3	127	228	55.7	56	99	56.6	77	101	76.2	0.0009
Rai stage III-IV	11	155	7.1	24	228	10.5	17	99	17.2	28	101	27.7	<0.0001
***IGHV* homology ≥98%**	29	152	19.1	75	224	33.5	66	98	67.3	64	101	63.4	<0.0001

[a] *IGHV*, immunoglobulin heavy variable gene

Table 6S. Time-fixed univariate and multivariate analysis of OS in the validation series [a]

	Univariate analysis				Multivariate analysis			
	HR	**LCI**	**UCI**	**p**	**HR**	**LCI**	**UCI**	**p**
Age (in year units) [b]	1.04	1.01	1.06	0.0033	1.05	1.02	1.07	0.0004
Female	1.00	-	-	0.6591	1.00	-	-	0.7026
Male	1.12	0.67	1.85		0.90	0.53	1.53	
Rai stage								
0-I	1.00	-	-		1.00	-	-	
II	1.47	0.76	2.86	<0.0001[c]	1.18	0.59	2.36	0.0259[c]
III-IV	4.54	2.51	8.20		2.31	1.25	4.26	
***IGHV* homology <98%**	1.00	-	-	<0.0001	1.00	-	-	<0.0001
***IGHV* homology ≥98%**	6.80	3.62	12.78		4.83	2.48	9.40	
Integrated mutational and cytogenetic model								
Very-low risk	1.00	-	-		1.00	-	-	
Low-risk	1.31	0.54	3.16	<0.0001[c]	1.24	0.50	3.03	0.0053[c]
Intermediate-risk	2.94	1.23	7.02		2.02	0.83	4.90	
High-risk	5.17	2.35	11.39		3.57	1.55	8.21	

[a]OS, overall survival; HR, hazard ratio; LCI, 95% lower confidence interval; UCI, 95% upper confidence interval; *IGHV*, immunoglobulin heavy variable gene
[b]Analyzed as a continuous variable
[c]p for trend
Shrinkage coefficient: 0.90
Discrimination: bias-corrected c-index: 0.782; optimism: 0.023
Calibration: bias-corrected calibration slope: 0.859; optimism: 0.15

Table S7. Distribution of FISH cytogenetic subgroups across the strata defined by the mutational and cytogenetic model in the training series

		Integrated mutational and cytogenetic model								
		del13q14		Normal/+12		*NOTCH1* M/*SF3B1* M/del11q22-q23		*TP53* DIS/*BIRC3* DIS		
		N	%	N	%	N	%	N	%	Total
FISH model	**del13q14**	155	79.9%	0	0.0%	28	14.4%	11	5.7%	194
	Normal	0	0.0%	171	80.7%	31	14.6%	10	4.7%	212
	+12	0	0.0%	57	69.5%	19	23.2%	6	7.3%	82
	del11q22-q23	0	0.0%	0	0.0%	21	53.8%	18	46.2%	39
	del17p13	0	0.0%	0	0.0%	0	0.0%	56	100%	56

Table S8. Newly developed genetic lesions in CLL patients showing clonal evolution

Patient ID	Sample ID	Timepoint	Months since diagnosis	*TP53* mutation	*NOTCH1* mutation	*SF3B1* mutation	*MYD88* mutation	*BIRC3* mutation	+12	del17p13	del11q22-q23	del13q14	*BIRC3* deletion
1	3715	diagnosis and first treatment	0.0	-	-	-	-	-	-	-	-	35% (x1)	-
1	5001	second treatment	38.1	-	-	c.1996A>G p.K666E	-	-	-	-	-	40% (x1)	-
1	7915	third treatment	75.6	-	-	c.1996A>G p.K666E	-	-	-	-	10%	49% (x1)	-
1	13425	fourth treatment	116.8	-	-	c.2098A>G p.K700E	-	-	-	-	90%	92% (x1)	58%
2	4932	diagnosis	0.0	-	-	-	-	-	-	-	-	12% (x1)	-
2	14285	first treatment	68.4	-	-	-	c.649G>T p.V217F	-	-	-	-	30% (x1) 19% (x2)	-
3	9630	diagnosis	0.0	-	-	-	-	-	69%	-	-	74% (x1)	-
3	14461	first treatment	34.5	c.817C>T p.R273C	-	-	-	-	50%	70%	-	82% (x1)	-
4	14263	diagnosis and first treatment	0.0	-	-	-	-	c.1270G>T p.E424*; c.1183_1352del4894 p.V395fs*78	-	-	-	10% (x1)	-
4	14264	second treatment	60.2	-	-	-	-	c.1270G>T p.E424*; c.1183_1352del4894 p.V395fs*78	-	-	29%	10% (x1)	33%
4	14265	third treatment	74.0	-	-	-	-	c.1270G>T p.E424*; c.1183_1352del4894 p.V395fs*78	-	-	35%	10% (x1)	39%
4	6550	fourth treatment	103.6	-	-	-	-	c.1270G>T p.E424*; c.1183_1352del4894 p.V395fs*78	-	-	45%	10% (x1)	84%
4	8080	fifth treatment	114.0	-	-	-	-	c.1270G>T p.E424*; c.1183_1352del4894 p.V395fs*78	-	-	60%	22% (x1)	86%
4	14324	sixth treatment	141.8	-	-	-	-	c.1270G>T p.E424*; c.1183_1352del4894 p.V395fs*78	-	-	80%	42% (x1)	85%
5	5537	diagnosis	0.0	-	-	-	-	-	-	-	-	10% (x2)	-
5	14573	first treatment	80.8	c.523C>G p.R175G	-	-	-	-	-	-	-	30% (x2)	-
6	9872	diagnosis	0.0	-	-	-	-	-	10%	-	-	36% (x1)	-
6	14495	last follow-up	31.8	-	-	-	-	-	10%	-	-	10% (x1) 19% (x2)	-
7	9696	diagnosis	0.0	-	-	-	-	c.1282delA p.R428fs*19	55%	-	55%	69% (x1)	51%
7	14462	last follow-up	34.1	-	-	-	-	c.1282delA p.R428fs*19; c.1609delG	37%	-	74%	66% (x1)	61%

8	4716	diagnosis	0.0	-	-	-	-	p.E537fs*31	70%	-	-	-	-
8	13443	first treatment	86.5	-	c.7544_7545delCT p.P2515fs*4	-	-	c.1639delC p.Q547fs*21	83%	-	-	-	-
9	5726	diagnosis and first treatment	0.0	-	c.7544_7545delCT p.P2515fs*4	-	-	-	-	-	-	10% (x1)	-
9	14235	second treatment	24.1	-	c.7544_7545delCT p.P2515fs*4	-	-	-	-	24%	14%	33% (x1)	-
9	9958	third treatment	46.0	c.408A>C p.Q136H e c.784G>C p.G262R	c.7544_7545delCT p.P2515fs*4	-	-	-	-	52%	40%	53% (x1)	-
10	4718	diagnosis	0.0	-	-	-	-	-	10%	-	-	72% (x1)	-
10	6756	first treatment	30.5	-	-	-	-	-	10%	-	-	70% (x1)	-
10	12276	second treatment	78.4	-	-	-	-	-	10%	17%	-	80% (x1)	-
10	14470	third treatment	87.3	c.743G>A p.R248Q	-	c.1997A>C p.K666T	-	-	56%	73%	-	50% (x1)	-
11	3703	diagnosis	0.0	-	-	-	-	-	55%	-	-	61% (x1)	-
11	14311	first treatment	70.8	-	-	-	-	c.1281_1290del10 p.I427fs*1	68%	-	-	65% (x1)	-
11	9482	second treatment	89.8	-	-	-	-	c.1281_1290del10 p.I427fs*1	44%	-	-	77% (x1)	-
12	4341	diagnosis	0.0	-	c.7544_7545delCT p.P2515fs*4	-	-	-	74%	-	-	-	-
12	5391	Richter syndrome transformation	0.0	c.818G>A p.R273H	c.7544_7545delCT p.P2515fs*4	-	-	-	10%	-	-	-	-
13	4715	diagnosis	0.0	-	-	-	-	-	22%	-	-	-	-
13	4880	first treatment	5.7	-	-	-	-	-	77%	-	-	-	-
13	14236	second treatment	30.6	-	-	-	-	-	80%	-	-	-	-
13	11678	third treatment	73.5	-	-	-	-	-	76%	-	-	-	-
13	13721	fourth treatment	88.4	-	-	-	-	-	81%	13%	-	-	-
14	9245	diagnosis	0.0	-	-	-	-	-	-	-	-	42% (x2)	-
14	14459	last follow-up	55.4	c.731G>A p.G244D	-	-	-	-	-	58%	-	94% (x2)	-
15	7274	diagnosis and first treatment	0.0	-	-	-	-	-	-	-	-	10% (x1)	-
15	14453	second treatment	29.3	-	-	c.1986C>A p.H662Q	-	-	-	-	74%	79% (x1)	77%
16	7146	diagnosis	0.0	-	-	-	-	-	-	-	-	97% (x1)	-
16	12957	first treatment	56.2	-	-	-	-	-	-	10%	-	50% (x1) 48% (x2)	-
17	4326	diagnosis	0.0	-	-	-	-	-	-	-	-	-	-
17	14279	last follow-up	46.8	-	-	-	-	-	-	-	35%	-	35%
18	5564	diagnosis and first treatment	0.0	-	-	-	-	-	-	-	-	19% (x1)	-
18	14071	second treatment	77.2	-	-	c.1866G>T p.E622D	-	-	-	-	-	29% (x1)	-
19	7033	diagnosis	0.0	-	-	-	-	-	-	-	-	75% (x1)	-
19	14382	first treatment	26.1	-	-	-	-	-	-	-	-	93% (x1)	-
19	14334	second treatment	59.5	-	-	-	-	-	-	-	10%	59% (x1)	10%
20	5074	diagnosis	0.0	-	-	-	-	-	-	-	-	-	-
20	9278	first treatment	31.5	-	-	-	-	-	-	-	-	-	-
20	14325	second treatment	58.7	-	-	-	-	-	-	10%	10%	-	-

21	6782	diagnosis	0.0	-	-	-	-	-	-	-	-	-	-
21	14467	last follow-up	62.2	-	-	-	-	-	-	-	12%	-	10%
22	3648	diagnosis	0.0	-	c.7544_7545delCT p.P2515fs*4	-	-	-	-	-	-	-	-
22	14257	first treatment	33.1	-	c.7544_7545delCT p.P2515fs*4	-	-	-	-	-	-	-	-
22	9080	second treatment	84.1	-	-	c.2098A>G p.K700E	-	-	-	-	-	-	-
23	5842	diagnosis	0.0	-	-	-	-	-	28%	-	-	-	-
23	14367	last follow-up	75.2	-	c.7544_7545delCT p.P2515fs*4	-	-	-	39%	-	-	-	-
24	10642	diagnosis and first treatment	0.0	-	-	c.2225G>A p.G742D	-	-	-	-	-	-	-
24	4692	Richter syndrome transformation	15.3	c.637C>T p.R213*	-	c.2225G>A p.G742D	-	-	-	33%	-	-	-
24	4304	Richter syndrome relapse	25.8	c.743G>A p.R248Q	c.7295_7344dupl50 p.R2434fs*4	c.2225G>A p.G742D	-	-	-	30%	-	-	-
25	5076	diagnosis and first treatment	0.0	-	-	-	-	-	63%	-	-	-	-
25	6734	Richter syndrome transformation	21.5	-	-	-	-	-	56%	10%	-	-	-
25	7227	Richter syndrome relapse	30.3	-	c.7321C>T p.Q2441*	-	-	-	41%	14%	-	-	-
25	9921	Richter syndrome relapse	47.0	c.716A>C p.N239T	c.7321C>T p.Q2441*	-	-	-	60%	21%	-	-	-
26	4293	diagnosis and first treatment	0.0	-	-	-	-	-	-	-	35%	-	-
26	4614	second treatment	4.9	-	-	c.2110A>T p.I704F	-	-	-	-	40%	-	-
26	7181	Richter syndrome transformation	45.1	-	-	c.2110A>T p.I704F	-	-	-	-	89%	-	-
27	10872	diagnosis	0.0	-	-	-	-	-	-	-	-	-	-
27	14376	last follow-up	25.3	-	-	-	-	-	-	-	12%	-	-
28	9483	diagnosis	0.0	-	-	-	-	-	-	-	-	-	-
28	14571	last follow-up	37.0	-	-	-	-	-	-	10%	-	-	-
29	5866	diagnosis	0.0	-	-	-	-	-	-	-	-	44% (x1)	-
29	14340	first treatment	56.0	-	-	-	-	-	-	-	-	10% (x1) 8% (x2)	-
30	11178	diagnosis	0.0	-	-	-	-	-	-	-	-	-	-
30	12644	Richter syndrome transformation	11.9	-	c.7544_7545delCT p.P2515fs*4	-	-	-	-	-	-	-	-
31	6503	diagnosis	0.0	-	-	-	-	-	-	-	-	10% (x1)	-
31	14314	last follow-up	30.0	-	-	c.2098A>G p.K700E	-	-	-	-	-	60% (x1)	-
32	6288	diagnosis	0.0	-	-	-	-	-	-	-	-	14% (x1)	-
32	9321	first treatment	31.7	-	-	-	-	c.1638_1639insA p.Q547fs*12	-	-	-	61% (x1)	-
33	4997	diagnosis and first treatment	0.0	c.536A>T p.H179L	-	-	-	-	-	92%	-	87% (x1)	-
33	6867	Richter syndrome transformation	24.5	c.536A>T p.H179L	-	-	-	c.1162A>G p.M388V	-	88%	-	72% (x1)	-
34	7275	diagnosis	0.0	-	-	-	-	-	16%	-	-	-	-
34	9289	first treatment	11.0	c.817C>T p.R273C	-	-	-	-	10%	-	-	-	-

35	4652	diagnosis and first treatment	0.0	-	-	-	-	-	77%	-	-	34% (x1) 29% (x2)	-	
35	9244	second treatment	42.4	-	-	-	-	-	79%	-	33%	37% (x1) 30% (x2)	29%	
36	4982	diagnosis	0.0	-	-	-	-	-	-	-	-	-	-	
36	12739	last follow-up	73.8	-	-	c.2098A>G p.K700E	-	-	-	-	-	-	-	

Table S9. Prevalence of newly developed genetic lesions in sequentially assessed CLL

	N	Total	%
del13q14 x 1	0	202	0.0%
del13q14 x 2	4	202	1.9%
+12	0	202	0.0%
del11q22-q23	10	202	5.0%
del17p13	10	202	5.0%
***TP53* mutation**	9	202	4.5%
***NOTCH1* mutation**	5	202	2.5%
***SF3B1* mutation**	8	202	4.0%
***MYD88* mutation**	1	202	0.5%
***BIRC3* mutation**	5	202	2.5%
***BIRC3* deletion**	7	202	3.4%
***TP53* disruption**	13	202	6.4%
***BIRC3* disruption**	11	202	5.4%

Table S10. Time-dependent analysis of the impact of genetic lesions on OS [a]

	Univariate analysis				Multivariate analysis			
	HR	**LCI**	**UCI**	**p**	**HR**	**LCI**	**UCI**	**p**
No *TP53* disruption	1.00	-	-	<0.0001	1.00	-	-	<0.0001
***TP53* disruption**	3.63	2.25	5.86		3.30	1.99	5.47	
No *BIRC3* disruption	1.00	-	-	0.0077	1.00	-	-	0.0166
***BIRC3* disruption**	2.45	1.27	4.72		2.59	1.19	5.64	
No *NOTCH1* mutation	1.00	-	-	0.0007	1.00	-	-	0.0247
***NOTCH1* mutation**	2.52	1.47	4.30		1.95	1.09	3.49	
No *SF3B1* mutation	1.00	-	-	0.0048	1.00	-	-	0.0766
***SF3B1* mutation**	2.42	1.31	4.49		1.82	0.94	3.54	
No del13q14	1.00	-	-	0.0113	1.00	-	-	0.0850
del13q14	0.56	0.36	0.88		0.66	0.41	1.06	
No del11q22-q23	1.00	-	-	0.0026	1.00	-	-	0.1961
del11q22-q23	2.55	1.39	4.69		1.59	0.79	3.23	
No *MYD88* mutation	1.00	-	-	0.2370	1.00	-	-	0.5494
***MYD88* mutation**	0.30	0.04	2.19		0.55	0.07	3.97	
No +12	1.00	-	-	0.1510	1.00	-	-	0.6536
+12	1.44	0.88	2.37		1.14	0.65	1.99	

[a]OS, overall survival; HR, hazard ratio; LCI, 95% lower confidence interval; UCI, 95% upper confidence interval

Table S11. Time-dependent univariate and multivariate analysis of OS [a]

	Univariate analysis				Multivariate analysis			
	HR	**LCI**	**UCI**	**p**	**HR**	**LCI**	**UCI**	**p**
Age ≤ 65 years	1.00	-	-	0.0002	1.00	-	-	0.0001
Age >65 years [b]	3.73	1.86	7.45		4.02	1.95	8.26	
Female	1.00	-	-	0.1670	1.00	-	-	0.2090
Male [c]	1.37	0.88	2.14		1.35	0.84	2.16	
Rai stage [b]								
0-I	1.00	-	-		1.00	-	-	
II	2.20	1.04	4.66	<0.0001[d]	1.36	0.62	2.97	<0.0001[d]
III-IV	6.78	4.20	10.96		4.51	2.69	7.58	
***IGHV* homology <98**	1.00	-	-		1.00	-	-	0.6366
***IGHV* homology ≥98%** [c]	1.63	1.05	2.54	0.0299	1.12	0.69	1.81	
Integrated mutational and cytogenetic model [b]								
Very-low risk	1.00	-	-		1.00	-	-	
Low-risk	2.10	1.00	4.21	<0.0001[d]	1.64	0.76	3.53	0.0003[d]
Intermediate-risk	4.01	1.90	8.46		2.19	1.00	4.92	
High-risk	6.77	3.41	13.43		3.56	1.67	7.58	

[a]OS, overall survival; HR, hazard ratio; LCI, 95% lower confidence interval; UCI, 95% upper confidence interval; *IGHV*, immunoglobulin heavy variable gene
[b]Time-varying variables
[c]Time-fixed variables
[d]p for trend

SUPPLEMENTARY REFERENCES

1. Pospisilova S, Gonzalez D, Malcikova J, et al. ERIC recommendations on TP53 mutation analysis in chronic lymphocytic leukemia. *Leukemia*. 2012;26(7):1458–1461.
2. Rossi D, Bruscaggin A, Spina V, et al. Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. *Blood*. 2011;118(26):6904–6908.
3. Rossi D, Rasi S, Fabbri G, et al. Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia. *Blood*. 2012;119(2):521–529.
4. Rossi D, Fangazio M, Rasi S, et al. Disruption of BIRC3 associates with fludarabine chemorefractoriness in TP53 wild-type chronic lymphocytic leukemia. *Blood*. 2012;119(12):2854–2862.
5. Fabbri G, Rasi S, Rossi D, et al. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. *J Exp Med*. 2011;208(7):1389–1401.
6. Rossi D, Spina V, Cerri M, et al. Stereotyped B-cell receptor is an independent risk factor of chronic lymphocytic leukemia transformation to Richter syndrome. *Clin Cancer Res*. 2009;15(13):4415–4422.
7. Schoenfeld D. Partial residuals for the proportional hazard regression model. *Biometrika*. 1982;69(1):239–241.
8. Harrell FE Jr, Lee K, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361–387.
9. Efron B, Tibshirani R. Improvements on cross-validation: the .632_bootstrap method. *JASA*. 1997;92(4328):548–560.
10. Van Howelingen JC, Le Cessie S. Predictive value of statistical models. *Stat Med*. 1990;9(11):1303–1325.
11. Chen CH, George SL. The bootstrap and identification of prognostic factors via Cox's proportional hazards regression model. *Stat Med*. 1985;4(1):39–46.
12. Ciampi A, Negassa A, Lou Z. Tree-structured prediction for censored survival data and the Cox model. *J Clin Epidemiol.* 1995;48(5):675–689.
13. Segal MR. Regression trees for censored-data. *Biometrics*. 1988;44(1):35–47.
14. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS., Random survival forests. *Ann Appl Stat*. 2008;2(3):841–860.
15. Ishwaran H, Kogalur UB, Gorodeski EZ, Minn AJ, Lauer MS. High dimensional variable selection for survival data. *J Am Stat Assoc*. 2010;105(489):205–217.
16. Dickman PW, Sloggett A, Hills M, Hakulinen T. Regression models for relative survival. *Stat Med*. 2004;23(1):51–64.