Mutation analysis of *SF3B1*

Mutational analysis of *SF3B1* (exons 1–25, including splicing sites; RefSeq NM_012433.2) was performed on PCR amplimers obtained from genomic DNA by a combination of Sanger sequencing (performed on an ABI PRISM 3100 Genetic Analyzer, Applied Biosystems) and targeted next generation sequencing (performed on a Genome Sequencer Junior, 454 Life Sciences, Roche, Branford, CT; mean coverage ~200×). Sanger sequences were compared to the corresponding germline RefSeq using Mutation Surveyor Version 2.41 (SoftGenetics, State College, PA) after both automated and manual curation. Sequencing reads obtained by next generation sequencing were mapped on RefSeq using the Amplicon Variant Analyzer software package (Roche). All sequence variants identified by Sanger sequencing or next generation sequencing were subsequently confirmed by Sanger sequencing from both strands on independent amplimers. Synonymous mutations, germline polymorphisms known from databases (dbSNP132, Ensembl Database, UCSC Genome Browser), and changes present in matched normal DNA were removed from the analysis. Molecular studies were performed in blind with respect to clinical data. All PCR primers and conditions are available upon request. The prediction of functional effects of the amino acid substitutions was performed by using the PolyPhen-2 algorithm (Software version 2.1, http://genetics.bwh.harvard.edu/pph2).[1]

Analysis of FISH karyotype and of *IGHV*, *TP53* and *NOTCH1* mutations

FISH analysis was performed as reported using probes LSI13 and LSID13S319, CEP12, LSIp53, and LSIATM (Abbott, Rome, Italy).[2] For each probe, at least 400 interphase cells with well-delineated fluorescent spots were examined. The presence of 13q14 deletion, trisomy 12, 11q22–q23 deletion and 17p13 deletion was scored when the percentage of nuclei with the abnormality was above our internal cut off (5%, 5%, 7%, and 10% respectively), defined as the mean plus 3 standard deviations of the frequency of normal control cells exhibiting the abnormality. *IGHV* mutational status was investigated as previously reported.[3] Sequences were aligned to the ImMunoGeneTics sequence directory and considered mutated if identity to corresponding germline genes was <98%.[4,5] *TP53* (exons 4–8; RefSeq NM_000546.4) and *NOTCH1* (exons 26, 27 and 34; RefSeq NM_017617.2) were analysed by Sanger sequencing as reported.[2,6] The regions of *TP53* and *NOTCH1* that have been analysed represent the mutation hotspots in CLL.[2,6]

Copy number analysis

Genome-wide DNA profiles were obtained from high molecular weight genomic DNA of CLL patients using the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA, USA), following the manufacturer's instructions. The bioinformatics pipeline used for the identification of copy number alterations is described in detail in Supplementary References 7 and 8.

Gene expression profile analysis

Gene expression profile analysis of purified normal B cell subpopulations and CLL samples was performed using Affymetrix HG-U133_plus2 arrays as part of an independent study (GEO database GSE12195). The probes used in Fig. 2D are the following: 228758_at, 203140_at, and 215990_s_at (for *BCL6*); 219841_at and 224499_s_at (for *AICDA*); 203684_s_at and 203685_at (for *BCL2*); 204562_at and 216986_s_at (for *IRF4*); and 201070_x_at, 201071_x_at, 211185_s_at, and 214305_s_at (for *SF3B1*).

Statistical analysis

Over-all survival was measured from date of diagnosis to date of death (event) or of last follow-up (censoring). Treatment free survival was measured from date of diagnosis to date of progressive and symptomatic disease requiring treatment according to IWCLL-NCI guidelines (event), death, or last follow up (censoring).[9] Survival was estimated by the Kaplan-Meier method.[10] The crude association between *SF3B1* mutations and survival was estimated by log-rank analysis.[10] The independence of *SF3B1* mutations as a predictor of CLL OS was estimated after controlling for confounding variables by multivariate Cox regression analysis.[11–13] The following variables were included in multivariate analysis: *SF3B1* mutations (present vs absent), age (>65 years vs ≤65 years), Rai stage (III–IV vs 0–II), *IGHV* identity ≥98% (present vs absent), 11q22–q23 deletion (present vs absent), *TP53* disruption by mutation and/or deletion (present vs absent), and ZAP70 expression (≥20% vs <20%). None of the covariates violated the proportional hazard assumption as documented by plotting the smoothed Schoenfeld residuals, and by performing a correlation test between time and residuals.[12–14] The assumption of effect additivity of predictors was not violated, as documented by a global test of additivity including interactions between *SF3B1* mutations and other covariates.[12,13] None of the covariates showed colinearity.[12,13] The prediction accuracy of the multivariate model was verified by assessing model discrimination and calibration.[12,13,15]

Categorical variables were compared by chi-square test and exact tests when appropriate. All statistical tests were two-sided. Statistical significance was defined as p value <.05. The analysis was performed with the Statistical Package for the Social Sciences (SPSS) software v.18.0 (Chicago, IL) and with R statistical package 2.13.0 (http://www.r-project.org).

REFERENCES

1. Adzhubei IA, Peshkin L, Ramensky VE, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4)248–249.
2. Rossi D, Cerri M, Deambrogi C, et al. The prognostic value of TP53 mutations in chronic lymphocytic leukemia is independent of Del17p13: implications for over-all survival and chemorefractoriness. *Clin Cancer Res*. 2009;15(3):995–1004.
3. 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.
4. Hamblin TJ, Davis Z, Gardiner A, et al. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood*. 1999;94(6):1848–1854.
5. Damle RN, Wasil T, Fais F, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood*. 1999;94(6):1840–1847.
6. 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.
7. Pasqualucci L, Trifonov V, Fabbri G, et al. Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat Genet*. 2011; doi: 10.1038/ng.892.
8. Rinaldi A, Mian M, Kwee I, et al. Genome-wide DNA profiling better defines the prognosis of chronic lymphocytic leukaemia. *Br J Haematol*. 2011; doi: 10.1111/j.1365–2141.2011.08789.x.

9. Hallek M, Cheson BD, Catovsky D, et al. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. *Blood*. 2008;111(12):5446–5456.
10. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Am Stat Assoc*. 1958;53:457–481.
11. Cox DR. Regression models and life tables. *J R Stat Assoc*. 1972; 34:187–220.
12. 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.
13. Harrell FE Jr. *Regression modeling strategies*. New York, NY: Springer-Verlag; 2001: 465–522.
14. Schoenfeld D. Partial residuals for the proportional hazard regression model. *Biometrika*. 1982;69(1):239–241.
15. van Howelingen JC, le Cessie S: Predictive value of statistical models. *Stat Med*. 1900;9(1):1303–1325.

Table S1. Clinical and biological characteristics of the fludarabine-refractory CLL cohort[a]

	All (n=59)		***SF3B1* mutated (n=10)**		***SF3B1* wt (n=49)**		**p**
	Number	**%**	**Number**	**%**	**Number**	**%**	
Age >65 years	37	62.7	6	60.0	31	63.0	1.000
Male	40	67.8	7	70.0	33	67.3	1.000
Rai stage III-IV	28	47.5	7	70.0	21	42.9	.168
Number of prior therapies							.264
0	26	44.1	4	40.0	22	44.9	
1	24	40.7	6	60.0	18	36.7	
>1	9	15.3	0	0	9	18.4	
Treatment regimen at refractoriness							.750
FCR	17	28.8	4	40.0	13	26.5	
FR	3	5.1	0	0	3	6.1	
FC	19	32.2	2	20.0	17	34.7	
F	20	33.8	4	40.0	16	32.7	
IGHV identity ≥98%	48	81.4	8	80.0	40	81.6	1.000
CD38 ≥30%	34	57.6	6	60.0	28	57.1	1.000
ZAP70 ≥20%	39	66.1	6	60.0	33	67.3	.721
TP53 disruption	23	39.0	1	10.0	22	44.9	.072
NOTCH1 mutations	14	23.7	1	10.0	13	26.5	.425
11q22-q23 deletion	15	25.4	3	30.0	12	24.5	.704
Trisomy 12	16	27.1	0	0	16	32.7	**.049**
13q14 deletion	31	52.5	6	60.0	25	51.1	.734
Normal FISH	10	16.9	5	50.0	5	10.2	**.008**

[a] wt, wild type; FCR, fludarabine, cyclophosphamide, rituximab; FR, fludarabine, rituximab; FC, fludarabine, cyclophosphamide; F, fludarabine; *IGHV*, immunoglobulin heavy variable gene; FISH, fluorescence in situ hybridization

Table S2. Clinical and biological characteristics of the consecutive series of newly diagnosed and previously untreated CLL[a]

	All		*SF3B1* mutated		*SF3B1* wt		p
	Number	%	Number	%	Number	%	
Age >65 years	183/301	60.8	13/17	76.5	170/284	59.9	.173
Male	163/301	54.2	13/17	76.5	150/284	52.8	.057
Rai stage III-IV	33/301	11.0	7/17	41.2	26/284	9.2	**.001**
IGHV identity ≥98%	100/294	34.0	8/17	47.1	92/277	33.2	.242
CD38 ≥30%	81/298	27.2	7/17	41.2	74/281	26.3	.259
ZAP70 ≥20%	77/253	30.0	8/13	61.5	69/240	28.7	**.025**
TP53 disruption	30/301	10.0	1/17	5.9	29/284	10.2	1.000
NOTCH1 mutations	34/301	11.3	1/17	5.9	33/284	11.6	.704
11q22-q23 deletion	21/301	7.0	2/17	11.8	19/284	6.7	.336
Trisomy 12	58/301	19.3	1/17	5.9	57/284	20.1	.211
13q14 deletion	157/301	52.2	8/17	47.1	149/284	52.5	.665
Normal FISH	89/301	29.6	8/17	47.1	81/284	28.5	.104

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

Table S3. Clinical and biological characteristics of the RS cohort[a]

	Number (n=33)	%
Clinical features at RS diagnosis		
Age >65 years	19	57.6
Male	22	66.7
ECOG PS >1	13	39.3
Ann Arbor stage III-IV	33	100
Rai stage III-IV	14	42.4
B symptoms	13	39.3
Tumor size >5 cm	24	72.7
Platelets <100 x 10^9/L	7	21.2
LDH >1.5 ULN	17	51.5
Prior CLL therapies >1	7	21.2
Pathologic features at RS diagnosis		
Non-GC phenotype	32	96.9
EBV infection	0	0
Genetic features		
TP53 disruption	18	54.5
c-MYC aberrations	5	15.1
NOTCH1 mutations	13	39.4
IGHV identity ≥98%	24	72.7

[a] ULN, upper limit of normal; GC, germinal center; *IGHV*, immunoglobulin heavy variable gene

Table S4. *SF3B1* mutations in CLL and RS

Sample ID	Disease phase	Nucleotide change [c]	Amino acid change [d]	Affected domain	Conserved site [e]	PolyPhen-2 [f]	Score	COSMIC v54 [g]
7040 [a]	CLL diagnosis	c.2044A>G	p.K666E	HEAT4	No	Damaging	1.000	No
11772 [a]	CLL diagnosis	c.2044A>G	p.K666E	HEAT4	No	Damaging	1.000	No
9094	CLL diagnosis	c.2046G>T	p.K666N	HEAT4	No	Damaging	1.000	No
4602 [a]	CLL diagnosis	c.2146A>G	p.K700E	HEAT5	Yes	Damaging	1.000	Yes
4681 [a]	CLL diagnosis	c.2146A>G	p.K700E	HEAT5	Yes	Damaging	1.000	Yes
7561 [a]	CLL diagnosis	c.2146A>G	p.K700E	HEAT5	Yes	Damaging	1.000	Yes
10676 [a]	CLL diagnosis	c.2146A>G	p.K700E	HEAT5	Yes	Damaging	1.000	Yes
11196 [a]	CLL diagnosis	c.2146A>G	p.K700E	HEAT5	Yes	Damaging	1.000	Yes
11197 [a]	CLL diagnosis	c.2146A>G	p.K700E	HEAT5	Yes	Damaging	1.000	Yes
11489 [a]	CLL diagnosis	c.2146A>G	p.K700E	HEAT5	Yes	Damaging	1.000	Yes
11785 [a]	CLL diagnosis	c.2146A>G	p.K700E	HEAT5	Yes	Damaging	1.000	Yes
3950 [a]	CLL diagnosis	c.2267G>A	p.G740E	-	Yes	Damaging	0.949	No
4845 [a]	Fludarabine-refractory CLL [b]	c.1938A>T	p.R630S	HEAT3	Yes	Damaging	1.000	No
7425 [a]	Fludarabine-refractory CLL [b]	c.2034C>A	p.H662Q	HEAT4	Yes	Damaging	1.000	No
7228	Fludarabine-refractory CLL [b]	c.2034C>A	p.H662Q	HEAT4	Yes	Damaging	1.000	No
12627	Fludarabine-refractory CLL	c.2032C>G	p.H662D	HEAT4	Yes	Damaging	1.000	No
7915 [a]	Fludarabine-refractory CLL	c.2044A>G	p.K666E	HEAT4	No	Damaging	1.000	No
12571	Fludarabine-refractory CLL	c.2146A>G	p.K700E	HEAT5	Yes	Damaging	1.000	Yes
12631	Fludarabine-refractory CLL	c.2146A>G	p.K700E	HEAT5	Yes	Damaging	1.000	Yes
14220_R [a]	Fludarabine-refractory CLL	c.2146A>G	p.K700E	HEAT5	Yes	Damaging	1.000	Yes
3981 [a]	Fludarabine-refractory CLL [b]	c.2146A>G	p.K700E	HEAT5	Yes	Damaging	1.000	Yes
5565 [a]	Fludarabine-refractory CLL [b]	c.2143_2148delCAGAAA	p.delQ699_K700	HEAT5	Yes	na	na	No
8343	Richter syndrome	c.2056C>G	p.Q670E	HEAT4	Yes	Damaging	0.999	No
7509 [a]	Richter syndrome	c.2146A>G	p.K700E	HEAT5	Yes	Damaging	1.000	Yes

[a] For these patients, paired normal DNA was available and confirmed the somatic origin of the mutation
[b] In these patients, the time of fludarabine-refractoriness was concomitant with clinical diagnosis
[c] Numbering according to GenBank accession No. NM_012433.2
[d] Numbering according to GenBank accession No. NP_036565.2
[e] Position conserved among SF3B1 orthologues
[f] na, not applicable, since the PolyPhen-2 algorithm predicts only the impact of amino acid substitutions
[g] Mutations listed in the Catalog of Somatic Mutations in Cancer (COSMIC) database v54 release (http://www.sanger.ac.uk/genetics/CGP/cosmic/)

Table S5. Distribution of genetic lesions according to *IGHV* mutation status[a]

	IGHV identity ≥98%		*IGHV* identity <98%		p
	Number	**%**	**Number**	**%**	
SF3B1 mutations	8/100	8.0	9/194	4.6	.242
TP53 disruption	13/100	13.0	17/194	8.8	.256
NOTCH1 mutations	26/100	26.0	8/194	4.1	<.001
ATM deletion	15/100	15.0	6/194	3.1	<.001

[a] *IGHV*, immunoglobulin heavy variable gene

Table S6. Univariate and multivariate analysis for overall survival in newly diagnosed and previously untreated CLL[a]

	Univariate analysis				Multivariate analysis			
Characteristics	**HR**	**LCI**	**UCI**	**p**	**HR**	**LCI**	**UCI**	**p**
SF3B1 wild type	-	-	-		-	-	-	
SF3B1 mutated	3.33	1.51	7.32	.003	3.02	1.24	7.35	.015
Age ≤ 65 years	-	-	-		-	-	-	
Age >65 years	2.71	1.60	4.58	<.001	3.17	1.68	5.97	<.001
Rai stage 0-II	-	-	-		-	-	-	
Rai stage III-IV	5.60	3.45	9.10	<.001	3.33	1.71	6.48	<.001
IGHV identity <98%	-	-	-		-	-	-	
IGHV identity ≥98%	1.92	1.22	3.00	.004	1.38	0.79	2.40	.252
No 11q22-q23 deletion	-	-	-		-	-	-	
11q22-q23 deletion	3.20	1.72	5.97	<.001	1.87	0.81	4.26	.137
TP53 wild type	-	-	-		-	-	-	
TP53 disruption	3.77	2.19	6.50	<.001	3.14	1.71	5.79	<.001
ZAP70 <20%	-	-	-		-	-	-	
ZAP70 ≥20%	1.52	0.90	2.58	.116	0.99	0.55	1.76	.977

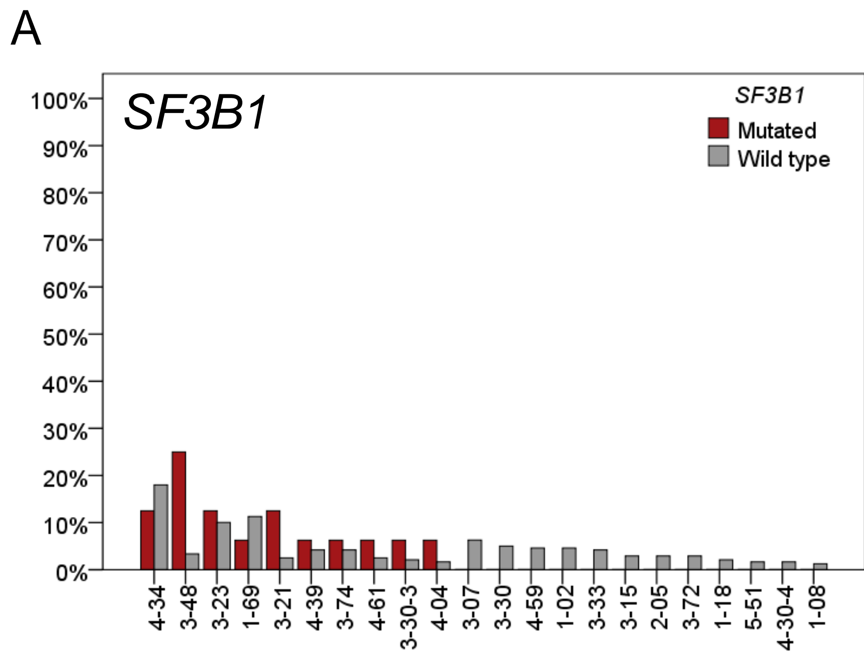
[a] HR, hazard ratio; LCI, 95% lower confidence interval; UCI, 95% upper confidence interval; *IGHV*, immunoglobulin heavy variable gene

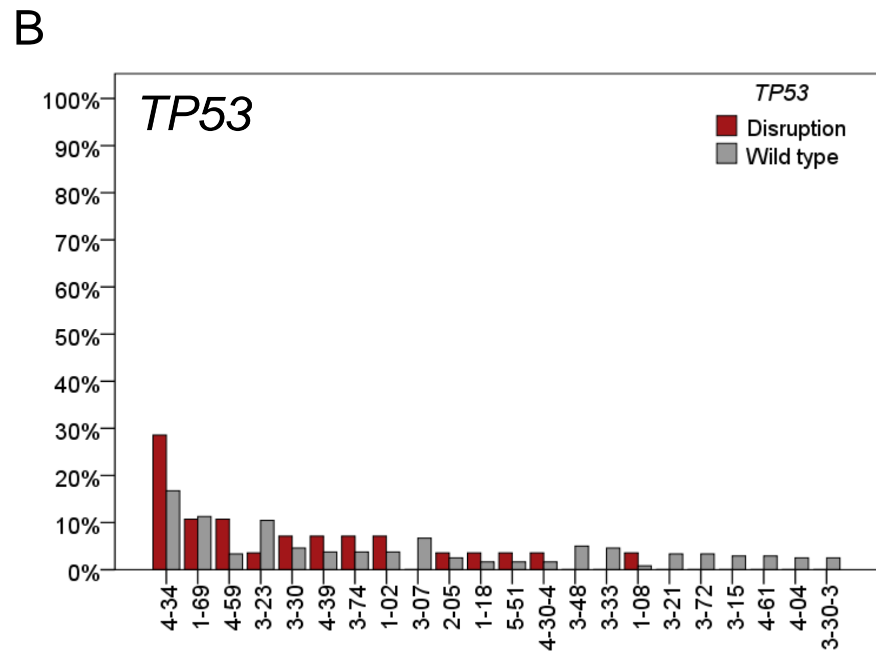
Table S7. *SF3B1* mutations in mature B-cell neoplasia

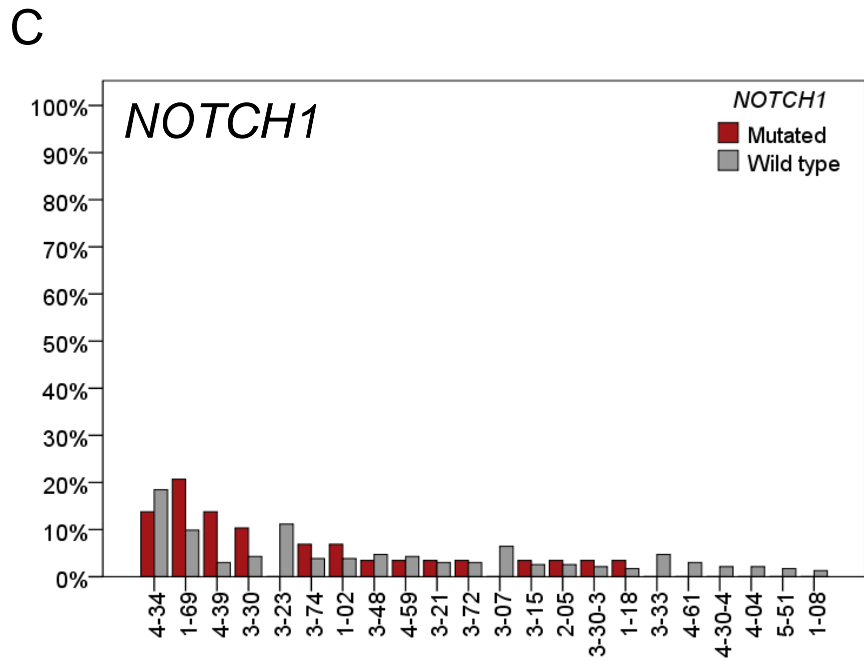
	Number	%
Fludarabine-refractory CLL	10/59	16.9
CLL diagnosis	17/301	5.6
Richter syndrome	2/33	6.0
Diffuse large B-cell lymphoma	0/20	0
Follicular lymphoma	0/20	0
Mantle cell lymphoma	0/20	0
Extranodal marginal zone lymphoma	0/21	0
Splenic marginal zone lymphoma	0/18	0
Hairy cell leukemia	0/17	0
Multiple myeloma	0/20	0

Figure S1. Distribution of genetic lesions in relation to *IGHV* gene usage

The bar graphs represent the prevalence of *SF3B1* mutations (Panel A; present: red bars; absent: grey bars), *TP53* disruption (Panel B; present: red bars; absent: grey bars), *NOTCH1* mutations (Panel C; present: red bars; absent: grey bars), and *ATM* deletion (Panel D; present: red bars; absent: grey bars) within CLL subgroups defined by *IGHV* gene usage (only *IGHV* gene subgroups including >5 CLL cases are represented).







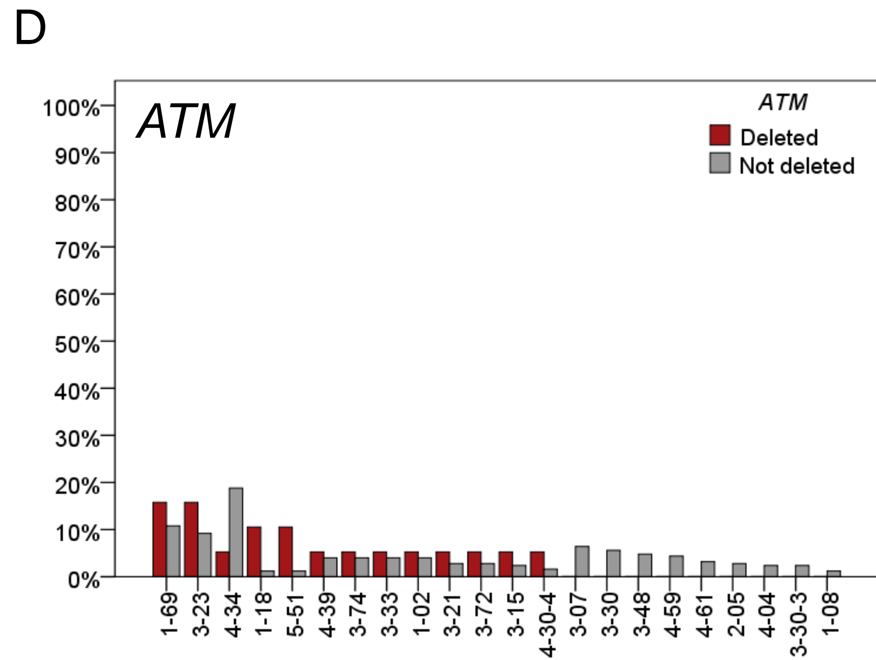


Fig. S1