SUPPLEMENTARY METHODS

Genomic Characterization

The targeted sequencing processing of the IWG cohort has been previously described in detail.[1,2] Genomic and clinical data can be explored and downloaded through the cBioPortal platform (https://www.cbioportal.org/),[3] under the Data Set denoted as *Myelodysplastic Syndromes (MDS IWG, IPSSM, NEJM Evidence 2022)*. This includes the samples included in the IPSS-M study, as well as the additional samples of the present study.

Thirty patient diagnostic samples were also profiled with whole genome sequencing (WGS). We used an analysis framework to characterize likely acquired somatic events from unmatched WGS data.[4] Briefly, CNVkit was used to define chromosomal copy number alterations (CNVs) based on logR segmentation and bi-allelic frequencies (BAF), using an unrelated panel of normals as reference.[5] ThetA2 was used to estimate purity values.[6] Single nucleotide variants and indels were ascertained with Mutect2 and Strelka.[7,8] The results in cancer related genes and hotspots were matched using the established annotation of COSMIC and OncoKB databases.[9,10] Brass, GRIDSS and SvABA, were integrated to detect structural variants.[11,12]

Clustering Analysis

Unsupervised clustering analysis was performed using Hierarchical Dirichlet Processes (HDP), with the R package available at https://github.com/nicolaroberts/hdp. A binary matrix for the presence or absence of molecular alterations was used as input to HDP, as previously described.[13,14] Molecular features included recurrent cytogenetic alterations, copy-neutral loss of heterozygosity events, and gene mutations. A total of 228 binary features observed in at least 5 patients were included. *TP53* multi-hit vs. monoallelic mutations, bi-allelic *TET2* vs. other *TET2* mutations (**Figure S2-S3**), *U2AF1*^{157} vs. *U2AF1*^{34} (**Figure S4**), and *IDH2*^{140} vs. *IDH2*^{172} were encoded as distinct features.[1,2,14,15]

The clustering analysis was performed in a stepwise manner, as illustrated in **Figure S5**. A first run of HDP was performed from the full cohort of 3,233 patients using a Gaussian base dirichlet distribution with hyperparameters for concentration set to rate=0.5 and shape=2. This resulted in 15 initial groups. We then examined the distribution of posterior probabilities of cluster assignment per patient across and within groups. The probabilities of assignment from two groups (groups 4-5 in **Figure S5**) were convoluted. We then therefore performed a second run of HDP only on the 1,051 patients initially classified within those two groups. We extracted three groups from the re-run that were enriched in patients with low genomic complexity (*del(5q), SF3B1, DDX41, CCUS-like*). The two

runs of HDP were then combined and manually inspected and curated to derive a rule-based hierarchical classification tree. The confusion matrix between the raw clusters and the processed ones is provided in **Figure S5**.

Clonality Analysis

Variant allele fraction (VAF) were corrected for allelic imbalances (adjusted VAF) using the estimation of total copy-number and major/minor allele specific copy-number provided by CNACS, as previously described.[16] The VAF of mutations occurring in males on the X chromosome were corrected by a factor two.

The adjusted VAF and total coverage depth were used to compute pairwise gene precedences from patients with 2 or more gene mutations, as previously described.[13,14] The tumor cell fraction (TCF) was defined as twice the adjusted VAF. In brief, a mutation A was considered to be dominant to a mutation B from the same patient if (i) TCF(A)+TCF(B)>1 i.e. the pigeon-hole principle was violated and the mutations were likely to be in nested clones (not parallel subclones); and (ii) TCF(A) was significantly higher than TCF(B) using a fisher exact test accounting for depth (expected number of mutant and wildtype reads given the observed TCF and depth) and using a 0.01 p-value threshold for significance.

Following the above methodology, a matrix of precedence recording the number of times across the cohort each mutation A was dominant to each mutation B, and vice versa. This precedence matrix was used as input to a Bradley Terry model to estimate the relative order of mutation acquisition. We used the *BradleyTerry2* R package. Only genes with at least 10 informative precedences across all other genes were considered. Bradley Terry analysis was performed within each molecular group for which at least 3 genes were informative.

Statistical Analysis

Distributions of categorical and continuous variables between two subgroups were compared using the Fisher's exact test and the Wilcoxon rank sum test, respectively. When more than two subgroups, distribution of continuous variables were compared using the Kruskal-Wallis test. Two-sided p-values <0.05 were reported as significant with the following notations ***p<0.001, **p<0.01, *p< 0.05. P-values were corrected for multiple testing with the Benjamini-Hochberg procedure when appropriate.

For overall survival (OS) and leukemia-free survival (LFS) analysis, the time-to-event was measured from the time of sample collection to the time of the event of interest (death from any cause for OS, leukemic transformation, or death from any cause for LFS), or last follow-up. OS and LFS probabilities were estimated using the Kaplan-Meier method and comparisons between subgroups were conducted using the log-rank test. Cumulative incidence functions were used to estimate the rate of leukemic transformation. Comparisons of cumulative incidence functions between subgroups were conducted using the Grey's test.

All statistical analyses were conducted using the R statistical platform version 4.3.0.

References

1. Bernard E, Nannya Y, Hasserjian RP, et al. Implications of TP53 allelic state for genome stability, clinical presentation and outcomes in myelodysplastic syndromes. Nat Med 2020;26(10):1549–56.

2. Bernard E, Tuechler H, Greenberg PL, et al. Molecular International Prognostic Scoring System for Myelodysplastic Syndromes. *NEJM Evidence*. 2022;1(7):EVIDoa2200008.

3. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012;2(5):401–4.

4. Gutiérrez-Abril J, Leongamornlert D, Shukla N, et al. Unmatched Whole-Genome Sequencing As a Clinical Tool for Hematological Neoplasms with Significant Utility in Cases with Tumor-in-Normal Contamination. *Blood* 2022; 140 (Supplement 1): 1197–1199.

5. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comput Biol 2016;12(4):e1004873.

6. Oesper L, Mahmoody A, Raphael BJ. THetA: inferring intra-tumor heterogeneity from high-throughput DNA sequencing data. Genome Biol 2013;14(7):R80.

7. Van der Auwera GA, O'Connor BD. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. O'Reilly Media; 2020.

8. Kim S, Scheffler K, Halpern AL, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nat Methods 2018;15(8):591–4.

9. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res 2019;47(D1):D941–7.

10. Suehnholz SP, Nissan MH, Zhang H, et al. Quantifying the Expanding Landscape of Clinical Actionability for Patients with Cancer. Cancer Discov 2023; doi: 10.1158/2159-8290.CD-23-0467. Epub ahead of print.

11. Cameron DL, Schröder J, Penington JS, et al. GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. Genome Res 2017;27(12):2050–60.

12. Wala JA, Bandopadhayay P, Greenwald NF, et al. SvABA: genome-wide detection of structural variants and indels by local assembly. Genome Res 2018;28(4):581–91.

13. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med 2016;374(23):2209–21.

14. Tazi Y, Arango-Ossa JE, Zhou Y, et al. Unified classification and risk-stratification in Acute Myeloid Leukemia. Nat Commun 2022;13(1):4622.

15. Awada H, Nagata Y, Goyal A, et al. Invariant phenotype and molecular association of biallelic mutant myeloid neoplasia. Blood Adv 2019;3(3):339–49.

16. Yoshizato T, Nannya Y, Atsuta Y, et al. Genetic abnormalities in myelodysplasia and secondary acute myeloid leukemia: impact on outcome of stem cell transplantation. Blood 2017;129(17):2347–58.

Figure S1 | Patterns of co-occurrence and mutual-exclusivity of molecular alterations from 3,233 MDS patients. The upper part of the heatmap in shades of blues represents the number of occurrences of each pairwise association of molecular alterations. The lower part of the heatmap represents the odds ratio, with darker brown color associated with increased strength of mutual-exclusivity and darker green color associated with increased strength of co-occurrence. FDR: false discovery rate; FWER: family-wise error rate.

Co-mutation patterns across gene mutations and copy-number alterations were consistent with prior studies (Papaemmanuil et al. 2013 PMID 24030381; Haferlach et al. 2014 PMID 24220272; Nazha et al., 2021 PMID 34406850), to include co-occurrence between *TP53* multi-hit mutations and aneuploidies, mutual-exclusivity among splicing factor mutations, or between *IDH1/2* and *TET2* mutations.

We also confirmed less commonly reported pairwise interactions, such as *CSNK1A1* mutations and deletion(5q) (Smith et al. 2015 PMID 26688096), *SF3B1* and *ZBTB33* mutations (Beauchamp et al. 2021 PMID 34568833) or *ETNK1* mutations and derivative (der) (1;7)(q10;p10) (Okuda et al. 2021 DOI: 10.1182/blood-2021-149556).

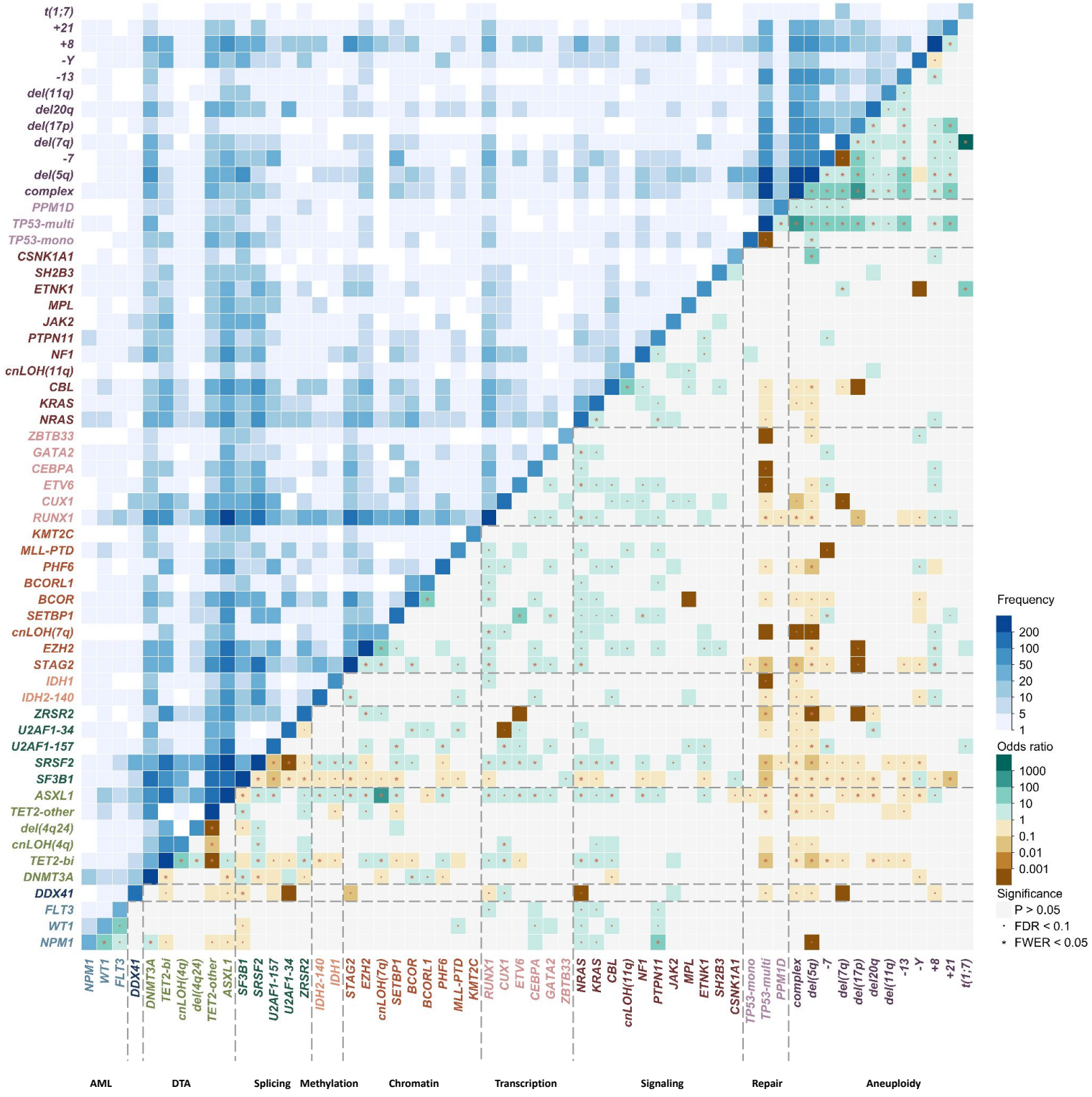
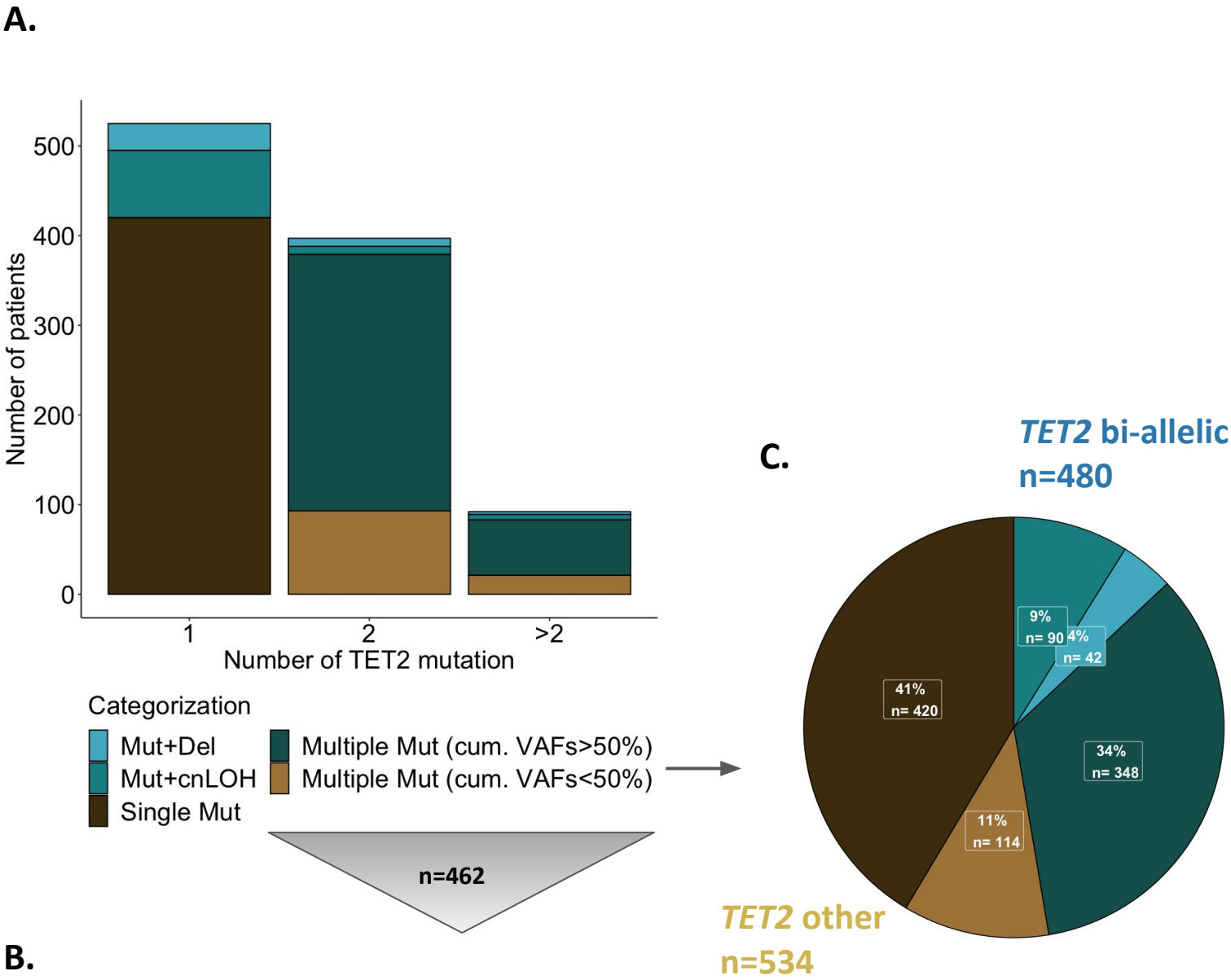

t(1;7)
+21
+8
-Y
-13
del(11q)
del20q
del(17p)
del(7q)
-7
del(5q)
complex
PPM1D
TP53-multi
TP53-mono
CSNK1A1
SH2B3
ETNK1
MPL
JAK2
PTPN11
NF1
cnLOH(11q)
CBL
KRAS
NRAS
ZBTB33
GATA2
CEBPA
ETV6
CUX1
RUNX1
KMT2C
MLL-PTD
PHF6
BCORL1
BCOR
SETBP1
cnLOH(7q)
EZH2
STAG2
IDH1
IDH2-140
ZRSR2
U2AF1-34
U2AF1-157
SRSF2
SF3B1
ASXL1
TET2-other
del(4q24)
cnLOH(4q)
TET2-bi
DNMT3A
DDX41
FLT3
WT1
NPM1
AML
DTA
Splicing
Methylation
Chromatin
Transcription
Signaling
Repair
Aneuploidy
Frequency
200
100
50
20
10
5
1
Odds ratio
1000
100
10
1
0.1
0.01
0.001
Significance
P > 0.05
FDR < 0.1
FWER < 0.05

Figure S2 | Categorization of *TET2* mutations as "*TET2* bi-allelic" or "*TET2* other".

A. Number of *TET2* mutations per patient categorized according to their chromosomal and variant allele fraction (VAF) status. Among the 1,014 *TET2*-mutant patients, 525 (51%) had a single *TET2* mutation, 397 (40%) had two mutations, and 92 (9%) had more than two mutations. Loss of heterozygosity (LOH), either through haploid LOH (i.e. deletion, Del) or copy neutral LOH (cnLOH) at the *TET2* locus were observed in 20% (105 of 525) of cases with a single *TET2* mutation and in 6% (27 of 489) of cases with two or more mutations. Cum: cumulative.

B. Positive correlation between the VAFs of the first and second *TET2* mutations from 462 cases with multiple *TET2* mutations in the absence of LOH (R^2=0.46, p<0.0001). Of note, this correlation was less than the one observed in other genes subject to bi-allelic inactivation (e.g. R^2=0.77 in the case of *TP53,* Bernard et al. *Nature Medicine* 2020 PMID: 32747829), and 114 cases had VAFs of the first two *TET2* mutations compatible with subclonal mosaicism. Therefore, we separated cases with multiple *TET2* mutations between those where the sum of the first two VAFs was below 50% (n=114, light brown) or above 50% (n=348, dark green).

C. Categorization of *TET2*-mutant patients as having "*TET2* bi-allelic" or "*TET2* other" mutation allelic state. Consistent with published methodology (Awada et al. *Blood Advances* 2019 PMID: 30709865), we integrated the number of *TET2* mutations, their cumulative VAFs, and LOH status at the locus to stratify 1,014 *TET2*-mutant patients into two groups: (i) "*TET2* bi-allelic" (47%, 480 of 1,014), where 73% (348 of 480) had multiple mutations with cumulative VAFs above 50% in the absence of LOH, and 27% had LOH (90 of 480 with cnLOH and 42 of 480 with deletion); and (ii) "*TET2* other" (53%, 534 of 1,014), representing likely mono-allelic mutation, where 79% (420 of 534) had a single mutation (with VAF below 50%) and 21% had multiple mutations with cumulative VAFs below 50%.


A.
Number of patients
500
400
300
200
100
0
1
2
>2
Number of TET2 mutation
Categorization
Mut+Del
Mut+cnLOH
Single Mut
Multiple Mut (cum. VAFs>50%)
Multiple Mut (cum. VAFs<50%)
n=462
C.
TET2 bi-allelic
n=480
9%
n= 90
4%
n= 42
41%
n= 420
34%
n= 348
11%
n= 114
TET2 other
n=534

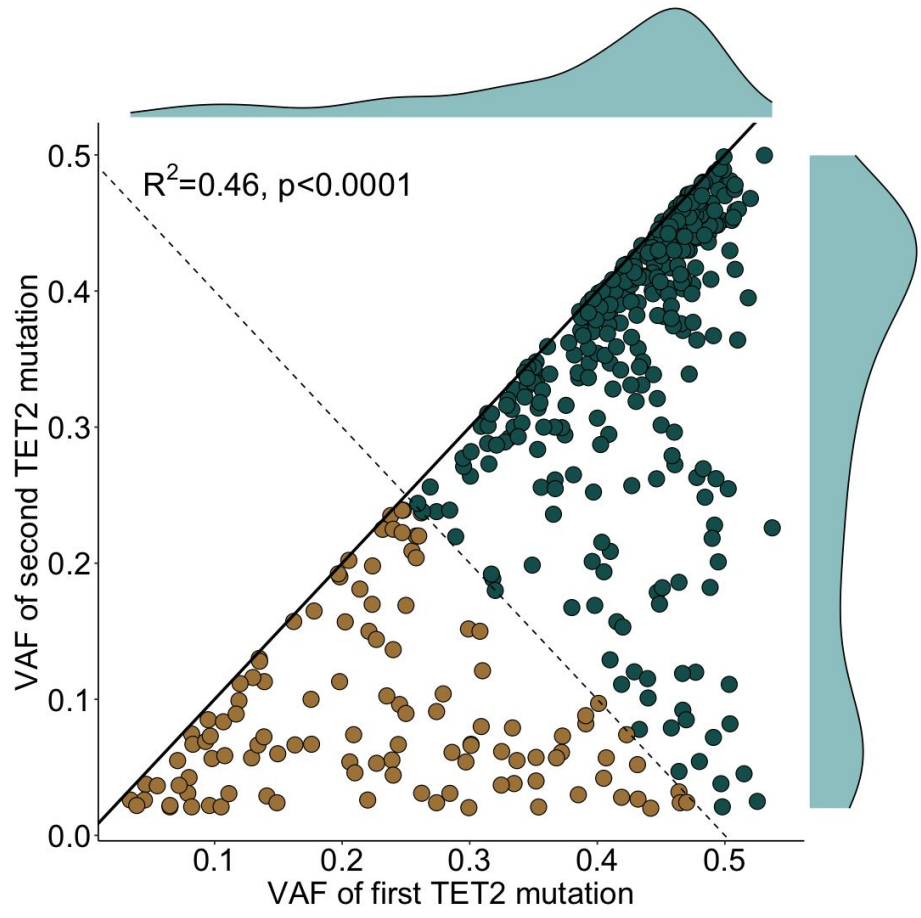

B.
R2=0.46, p<0.0001
VAF of second TET2 mutation
0.5
0.4
0.3
0.2
0.1
0.0
0.1
0.2
0.3
0.4
0.5
VAF of first TET2 mutation

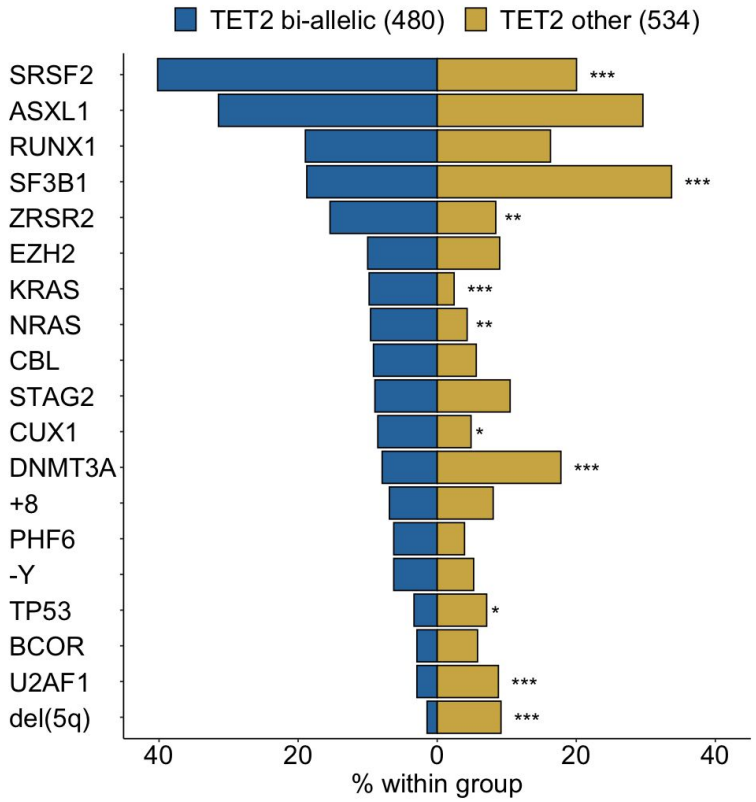
Figure S3 | Implications of the allelic state of *TET2* mutations for co-mutation and genotype-phenotype analysis.

A. Frequency distribution of gene mutations and cytogenetic alterations in patients classified as *TET2 bi-allelic* (blue, n=480) or *TET2 other* (gold, n=534). ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$, two-sided Fisher's exact test with Benjamini–Hochberg multiple testing correction.

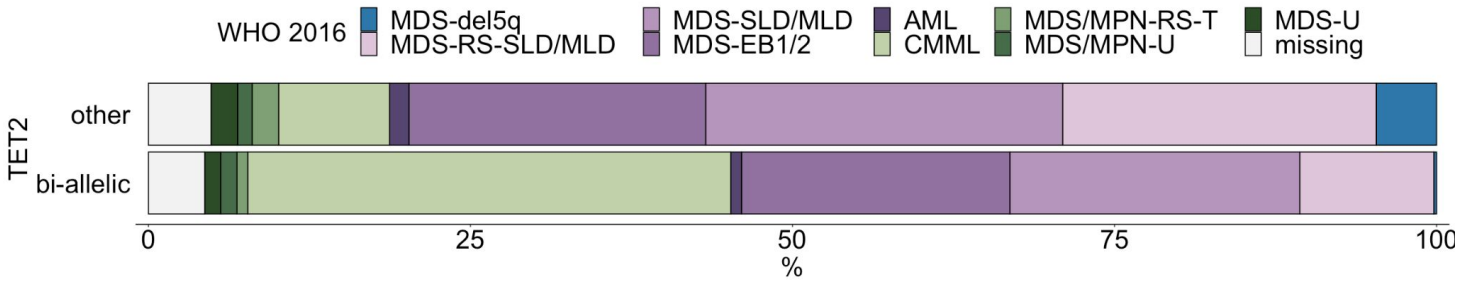
B. Distribution of WHO 2016 diagnosis per *TET2 bi-allelic* or *TET2 other* category. CMML (light green) represented 38% (180 of 480) of the *TET2 bi-allelic* group compared to 9% (46 of 534) of the *TET2 other* group.

C. Left and middle panels: Kaplan-Meier probability estimates of overall survival (left) and leukemia free survival (middle) per *TET2 bi-allelic* or *TET2 other* category. P-values are from the log-rank test. Right panel: cumulative incidence curves for the rate of leukemic transformation per *TET2 bi-allelic* or *TET2 other* category. P-value is from the Gray's test.

A.


TET2 bi-allelic (480)
TET2 other (534)
SRSF2
ASXL1
RUNX1
SF3B1
ZRSR2
EZH2
KRAS
NRAS
CBL
STAG2
CUX1
DNMT3A
+8
PHF6
-Y
TP53
BCOR
U2AF1
del(5q)
***
***
**
***
**
*
***
*
***
***
40
20
0
20
40
% within group

B.


WHO 2016
MDS-del5q
MDS-RS-SLD/MLD
MDS-SLD/MLD
MDS-EB1/2
AML
CMML
MDS/MPN-RS-T
MDS/MPN-U
MDS-U
missing
TET2
other
bi-allelic
0
25
50
75
100
%

C.

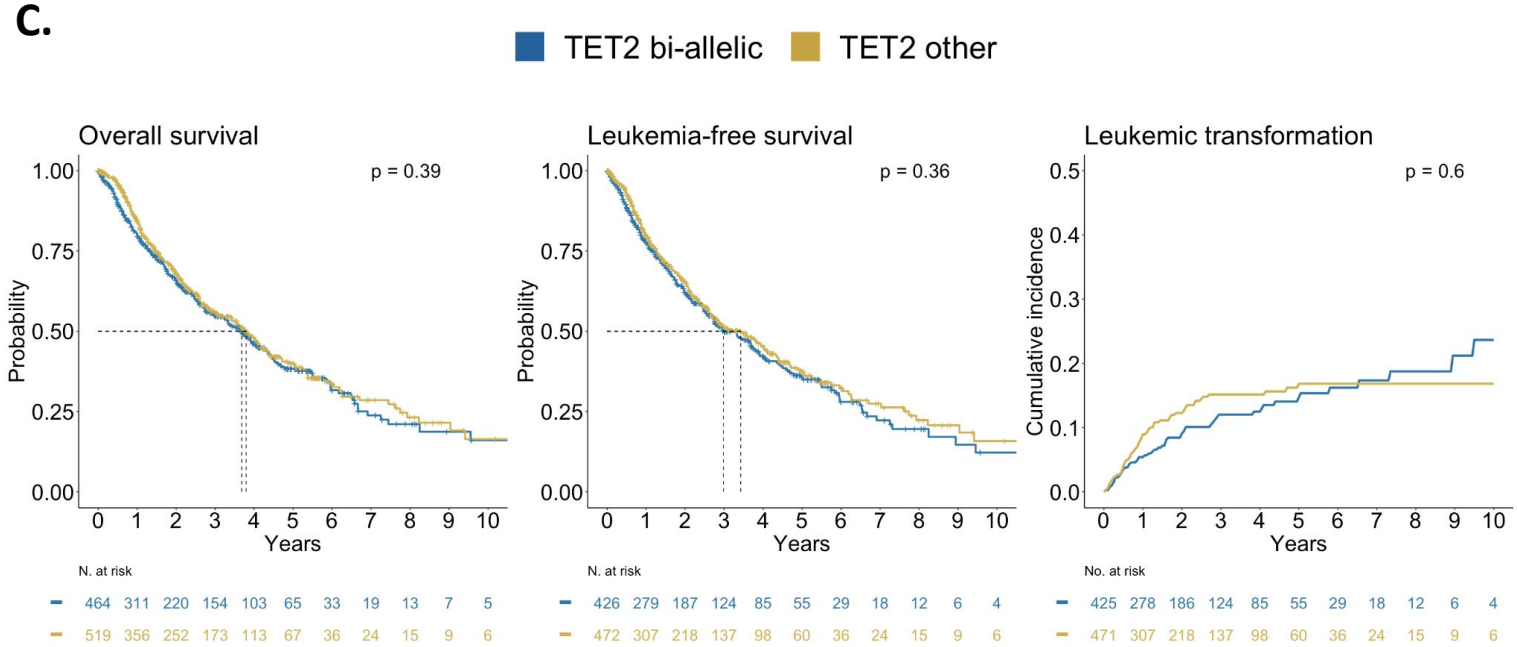

TET2 bi-allelic
TET2 other
Overall survival
p = 0.39
Leukemia-free survival
p = 0.36
Leukemic transformation
p = 0.6
Probability
Cumulative incidence
Years
N. at risk
No. at risk
464 311 220 154 103 65 33 19 13 7 5
519 356 252 173 113 67 36 24 15 9 6
426 279 187 124 85 55 29 18 12 6 4
472 307 218 137 98 60 36 24 15 9 6
425 278 186 124 85 55 29 18 12 6 4
471 307 218 137 98 60 36 24 15 9 6

Figure S4 | Comparisons between mutated *SF3B1* or *U2AF1* hotspots.

A. Frequency distribution of gene mutations and cytogenetic alterations in patients with mutated *SF3B1* K666 hotspot (blue, n=92) or *SF3B1* K700 hotspot (gold, n=421). ***p< 0.001, two-sided Fisher's exact test with Benjamini–Hochberg multiple testing correction.

B. Kaplan-Meier probability estimates of overall survival per mutated *SF3B1* K666 hotspot (blue), *SF3B1* K700 hotspot (gold), or other *SF3B1* driver mutations (grey). P-value is from the log-rank test.

C. Distribution of the WHO 2016 diagnosis subtypes for patients with mutated *U2AF1* Q157 hotspot (n=159) or *U2AF1* S34 hotspot (n=112).

D. Left and middle panels: Kaplan-Meier probability estimates of overall survival (left) and leukemia free survival (middle) per mutated *U2AF1* Q157 or S34 hotspot category. P-values are from the log-rank test. Right panel: cumulative incidence curves for the rate of leukemic transformation per mutated *U2AF1* Q157 or S34 hotspot category. P-value is from the Gray's test.

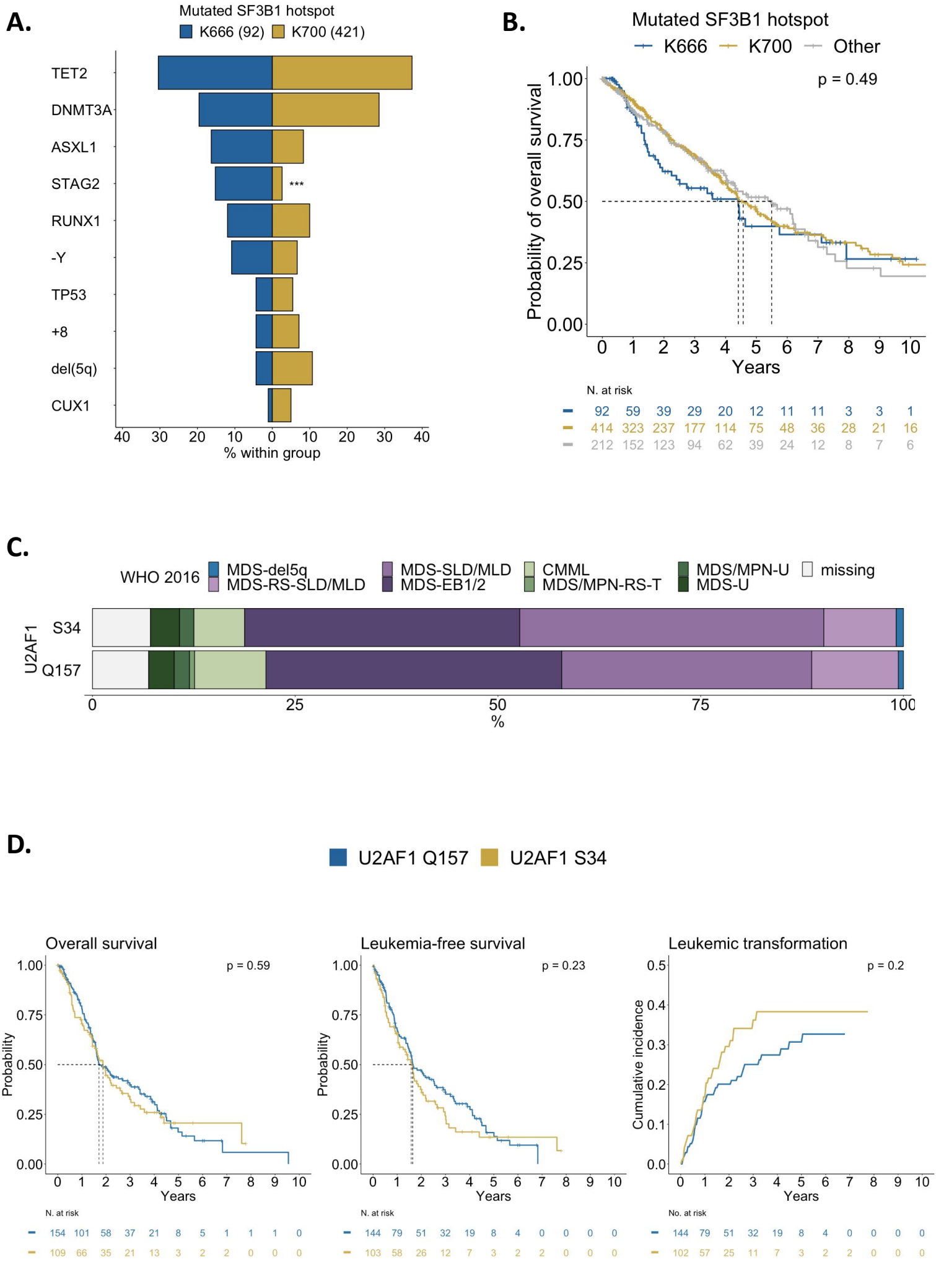

A.
Mutated SF3B1 hotspot
K666 (92)
K700 (421)
TET2
DNMT3A
ASXL1
STAG2
***
RUNX1
-Y
TP53
+8
del(5q)
CUX1
40 30 20 10 0 10 20 30 40
% within group
B.
Mutated SF3B1 hotspot
K666
K700
Other
p = 0.49
Probability of overall survival
1.00
0.75
0.50
0.25
0.00
0 1 2 3 4 5 6 7 8 9 10
Years
N. at risk
92 59 39 29 20 12 11 11 3 3 1
414 323 237 177 114 75 48 36 28 21 16
212 152 123 94 62 39 24 12 8 7 6
C.
WHO 2016
MDS-del5q
MDS-RS-SLD/MLD
MDS-SLD/MLD
MDS-EB1/2
CMML
MDS/MPN-RS-T
MDS/MPN-U
MDS-U
missing
U2AF1
S34
Q157
0
25
50
75
100
%
D.
U2AF1 Q157
U2AF1 S34
Overall survival
p = 0.59
Probability
1.00
0.75
0.50
0.25
0.00
0 1 2 3 4 5 6 7 8 9 10
Years
N. at risk
154 101 58 37 21 8 5 1 1 1 0
109 66 35 21 13 3 2 2 0 0 0
Leukemia-free survival
p = 0.23
Probability
1.00
0.75
0.50
0.25
0.00
0 1 2 3 4 5 6 7 8 9 10
Years
N. at risk
144 79 51 32 19 8 4 0 0 0 0
103 58 26 12 7 3 2 2 0 0 0
Leukemic transformation
p = 0.2
Cumulative incidence
0.5
0.4
0.3
0.2
0.1
0.0
0 1 2 3 4 5 6 7 8 9 10
Years
No. at risk
144 79 51 32 19 8 4 0 0 0 0
102 57 25 11 7 3 2 2 0 0 0

Figure S5 | Unsupervised clustering analysis and derivation of MDS molecular groups.

A. Stepwise analytical steps to derive MDS molecular groups from iterative hierarchical dirichlet process (HDP) analysis and manual curation.

B. Top: Posterior probability distribution per component (i.e. cluster) of the maximum probability per sample (plain color) and of the second maximum probability per sample (shaded color). **Middle**: zoom into components 4 and 5, with the distribution probabilities of pertaining to those components across all components. **Bottom**: Results of the the HDP re-run on components 4 and 5 with the extraction of three different new components (s1, s2, s3).

C. Concordance matrix comparing the assignments to HDP components with the processed/curated molecular subtypes.

A.

Step 1

Hierarchical Dirichlet Process (HDP)
https://github.com/nicolaroberts/hdp

Initialisation Parameters:
Base Dirichlet distribution = Gaussian
Shape hyperparameter = 2
Rate hyperparameter = 0.5
Number of starting clusters = 15

Gibbs sampling Parameters:
Number of burn-in iterations = 5000
Number of posterior samples = 350
Number of iterations between collected samples = 20
Number of chains = 3

Step 2

HDP post processing

Evaluation of Posterior Probability Distribution
Distribution of the first and second maximum probabilities per data item (i.e. sample) per component (i.e. cluster)

Selection of components for second iterative HDP run
Second HDP run on low complexity components (i.e. SF3B1, del(5q), DDX41, CCUS-like)

Step 3

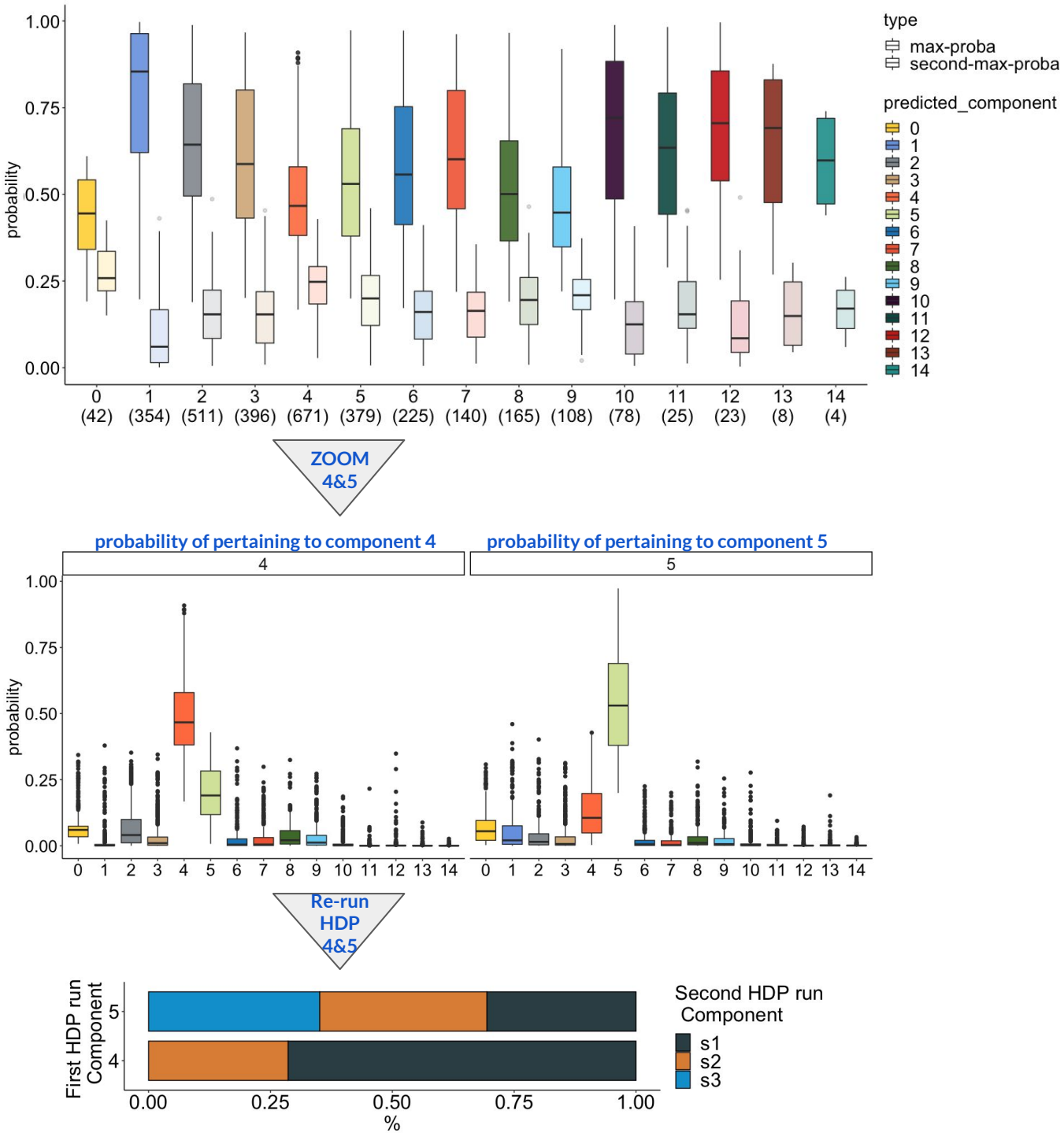
HDP manual curation

Rule-based hierarchical classification tree
Extraction of the strongest signals of mutual-exclusivity and co-occurrence

Comparisons of the raw clustering results and processed results

B.

Steps 1-2



C. **Steps 3**

HDP Components

Processed Molecular subtypes	0	1	10	11	12	13	14	2	3	6	7	8	9	s1	s2	s3
DDX41	0	2	0	1	0	0	0	29	3	2	0	0	2	17	50	1
AML-like	0	2	48	1	0	0	0	0	2	2	2	9	0	0	0	0
TP53-complex	2	263	0	21	1	3	2	1	1	10	2	3	2	0	5	9
der(1;7)	0	0	0	0	15	0	0	0	0	0	0	0	0	1	0	0
-7/SETBP1	0	8	0	0	1	0	0	2	4	119	5	1	4	4	9	0
del(5q)	2	12	1	0	0	0	0	6	5	2	0	6	1	45	34	108
EZH2-ASXL1	1	0	0	0	0	0	0	6	11	6	89	0	10	3	3	0
IDH-STAG2	0	0	0	0	0	0	0	17	217	4	12	4	10	11	13	0
BCOR/L1	0	1	0	0	1	0	0	7	14	6	0	76	4	2	3	0
bi-TET2	0	4	0	0	0	0	0	347	36	2	4	1	2	6	9	0
U2AF1-157	1	1	0	0	0	0	0	1	1	14	4	0	40	0	10	0
U2AF1-34	27	2	0	0	0	0	0	1	1	0	0	30	0	3	4	0
SRSF2	0	5	0	0	2	0	1	25	27	4	1	0	2	3	1	0
ZRSR2	0	0	0	0	0	0	0	9	0	0	6	1	2	14	11	0
SF3B1	1	6	3	0	0	1	0	11	3	6	2	17	1	339	66	0
CCUS-like	3	9	2	0	1	1	0	9	3	4	1	7	9	144	25	6
mNOS	5	18	7	1	2	3	1	32	52	29	8	2	17	0	66	8

Figure S6 | Prevalence of MDS molecular groups and molecular characteristics.

A. Proportion of patients from the 3,233 categorized within each molecular group. The absolute number of patients is indicated on top of each bar.

B. Summary of the molecular complexity of each molecular group, as indicated by the median number of cytogenetic alterations per patient (leftmost dot) and the median number of mutated gene per patient (rightmost dot) calculated within each group.

C. Distribution of the values of maximum variant allele fraction per patient within each molecular group. Median values are indicated by a thick black line.

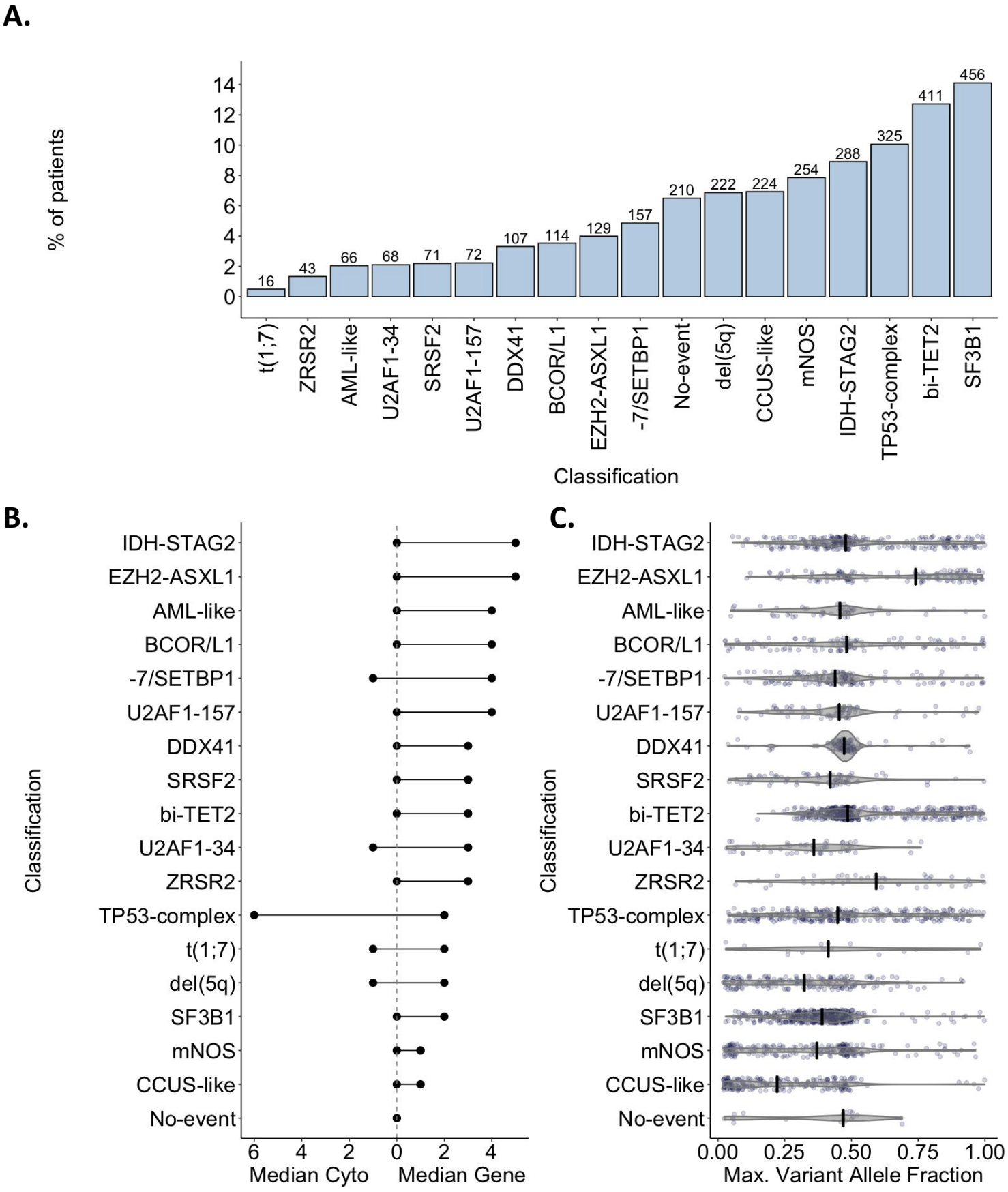

A.
% of patients
14
12
10
8
6
4
2
0
16
43
66
68
71
72
107
114
129
157
210
222
224
254
288
325
411
456
t(1;7)
ZRSR2
AML-like
U2AF1-34
SRSF2
U2AF1-157
DDX41
BCOR/L1
EZH2-ASXL1
-7/SETBP1
No-event
del(5q)
CCUS-like
mNOS
IDH-STAG2
TP53-complex
bi-TET2
SF3B1
Classification
B.
Classification
IDH-STAG2
EZH2-ASXL1
AML-like
BCOR/L1
-7/SETBP1
U2AF1-157
DDX41
SRSF2
bi-TET2
U2AF1-34
ZRSR2
TP53-complex
t(1;7)
del(5q)
SF3B1
mNOS
CCUS-like
No-event
6
4
2
0
2
4
Median Cyto
Median Gene
C.
Classification
IDH-STAG2
EZH2-ASXL1
AML-like
BCOR/L1
-7/SETBP1
U2AF1-157
DDX41
SRSF2
bi-TET2
U2AF1-34
ZRSR2
TP53-complex
t(1;7)
del(5q)
SF3B1
mNOS
CCUS-like
No-event
0.00
0.25
0.50
0.75
1.00
Max. Variant Allele Fraction

Figure S7 | Heatmap representing the molecular composition of MDS molecular groups. The groups are ordered by size. Each row represents a genetic alteration and each column represents a patient.

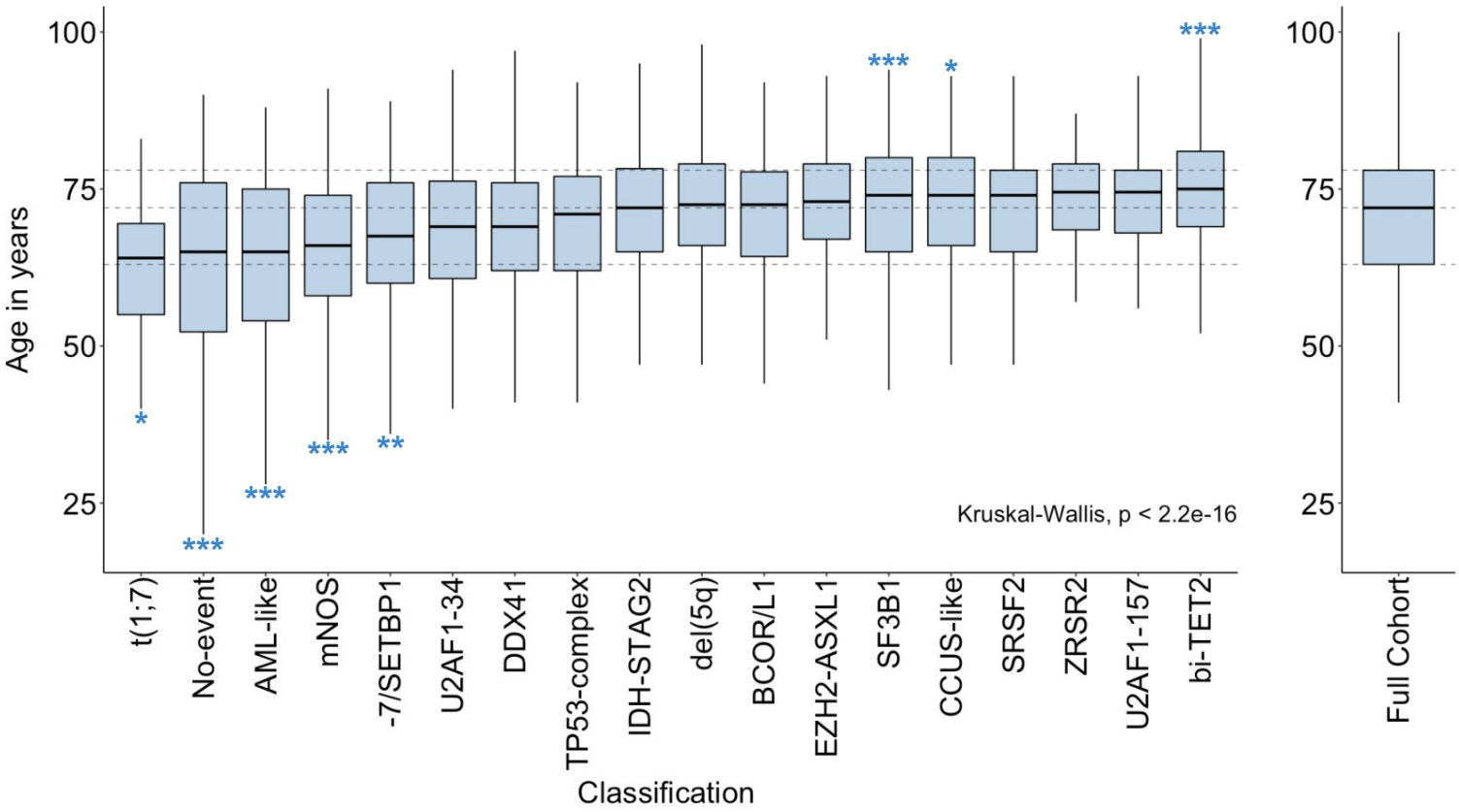

U2AF1^157 2.3%
SRSF2 2.2%
U2AF1^34 2.1%
AML-like 2.0%
ZRSR2 1.3%
der(1;7) 0.5%
EZH2-ASXL1 4.0%
DDX41 3.3%
BCOR/L1 3.5%
-7/SETBP1 4.9%
del(5q) 6.8%
CCUS-like 6.9%
mNOS 7.9%
IDH-STAG2 8.9%
TP53-complex 10.1%
bi-TET2 12.8%
SF3B1 14.1%
NPM1
WT1
FLT3
MYC
DDX41
DNMT3A
TET2 bi-allelic
cnLOH(4q)
del(4q24)
TET2 other
ASXL1
SF3B1
SRSF2
U2AF1-157
U2AF1-34
ZRSR2
IDH2-140
IDH1
STAG2
EZH2
cnLOH(7q)
SETBP1
BCOR
BCORL1
PHF6
MLL-PTD
KMT2C
RUNX1
CUX1
ETV6
CEBPA
GATA2
ZBTB33
NRAS
KRAS
CBL
cnLOH(11q)
NF1
PTPN11
JAK2
MPL
ETNK1
SH2B3
CSNK1A1
IRF1
RAD50
NFE2
TP53 mono-allelic
TP53 multi-hit
PPM1D
complex
del(5q)
-7
del(7q)
-17/17p
del(20q)
del(11q)
-13
+8
+21
-Y
t(1;7)
inv(3)
Diagnosis
Blast
IPSSM
Diagnosis
MDS
MDS/MPN
AML
Blast
[0,5)
[5,10)
[10,20)
[20,30]
missing
IPSSM
Very Low
Low
Moderate Low
Moderate High
High
Very High

Figure S8 | Demographics of MDS molecular groups.

A. Distribution of the age at diagnosis of patients within each molecular group. The distribution for the full cohort from 3,233 patients is depicted on the right for comparison. The global p-value is from the Kruskal-Wallis test. Stars at the top (elevated distribution) or the bottom (lower distribution) of each boxplot are the results of Wilcoxon rank-sum tests between a molecular group and the rest of the cohort with Benjamini–Hochberg multiple testing correction. ***p<0.001, **p<0.01, *p< 0.05.

B. Male (M) / Female (F) proportion within each molecular group. The proportion for the full cohort from 3,233 patients is depicted on the right for comparison. Stars in gold (enriched for females) or green (enriched for males) colors at the top of each bar are the results of Fisher's exact tests between a molecular group and the rest of the cohort with Benjamini–Hochberg multiple testing correction. ***p<0.001, **p<0.01, *p< 0.05.

A.


Age in years
100
75
50
25
Kruskal-Wallis, p < 2.2e-16
t(1;7)
No-event
AML-like
mNOS
-7/SETBP1
U2AF1-34
DDX41
TP53-complex
IDH-STAG2
del(5q)
BCOR/L1
EZH2-ASXL1
SF3B1
CCUS-like
SRSF2
ZRSR2
U2AF1-157
bi-TET2
Full Cohort
Classification

B.

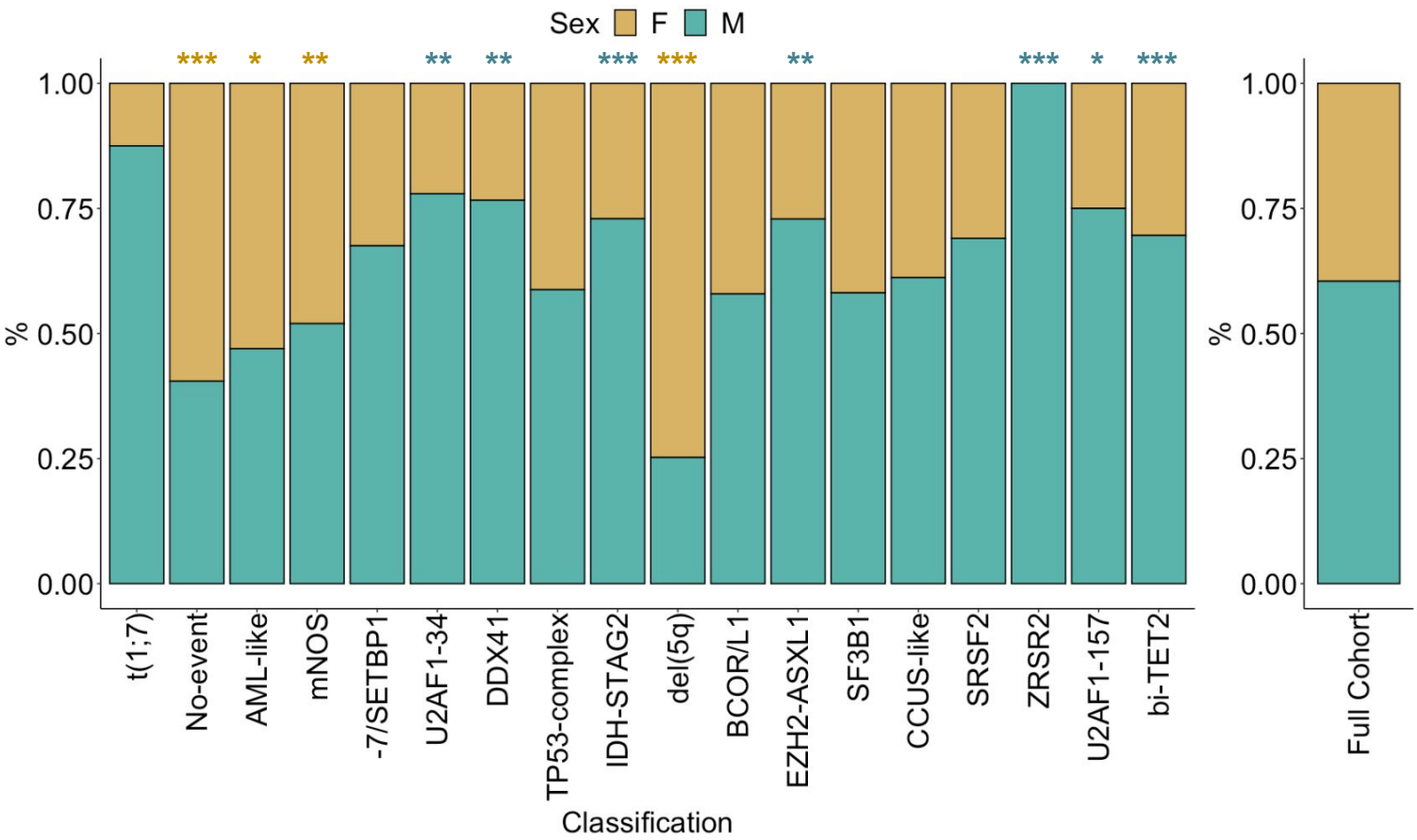
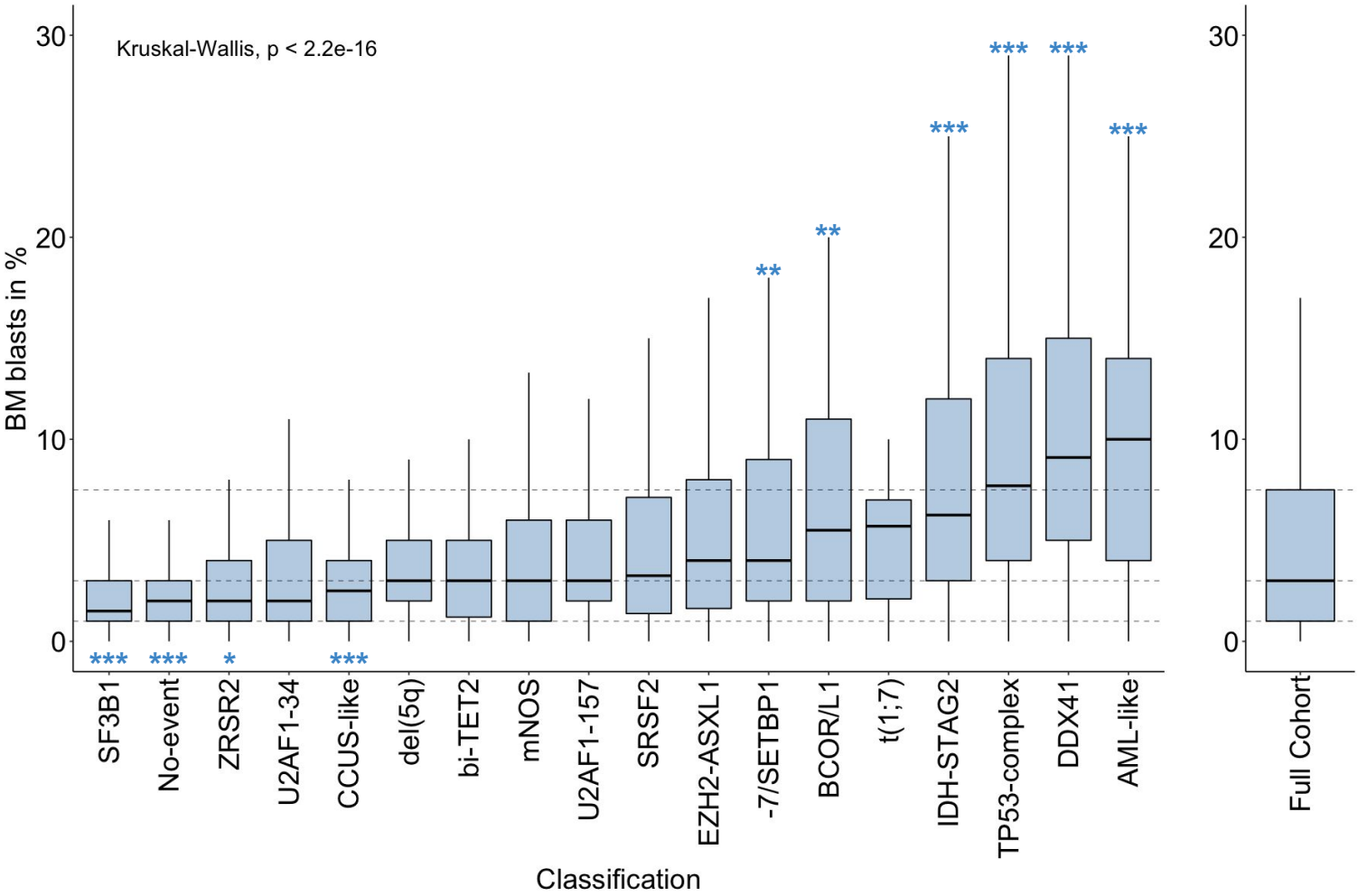

Sex F M
%
1.00
0.75
0.50
0.25
0.00
t(1;7)
No-event
AML-like
mNOS
-7/SETBP1
U2AF1-34
DDX41
TP53-complex
IDH-STAG2
del(5q)
BCOR/L1
EZH2-ASXL1
SF3B1
CCUS-like
SRSF2
ZRSR2
U2AF1-157
bi-TET2
Full Cohort
Classification

Figure S9 | Association between MDS molecular groups and hematological phenotypes.

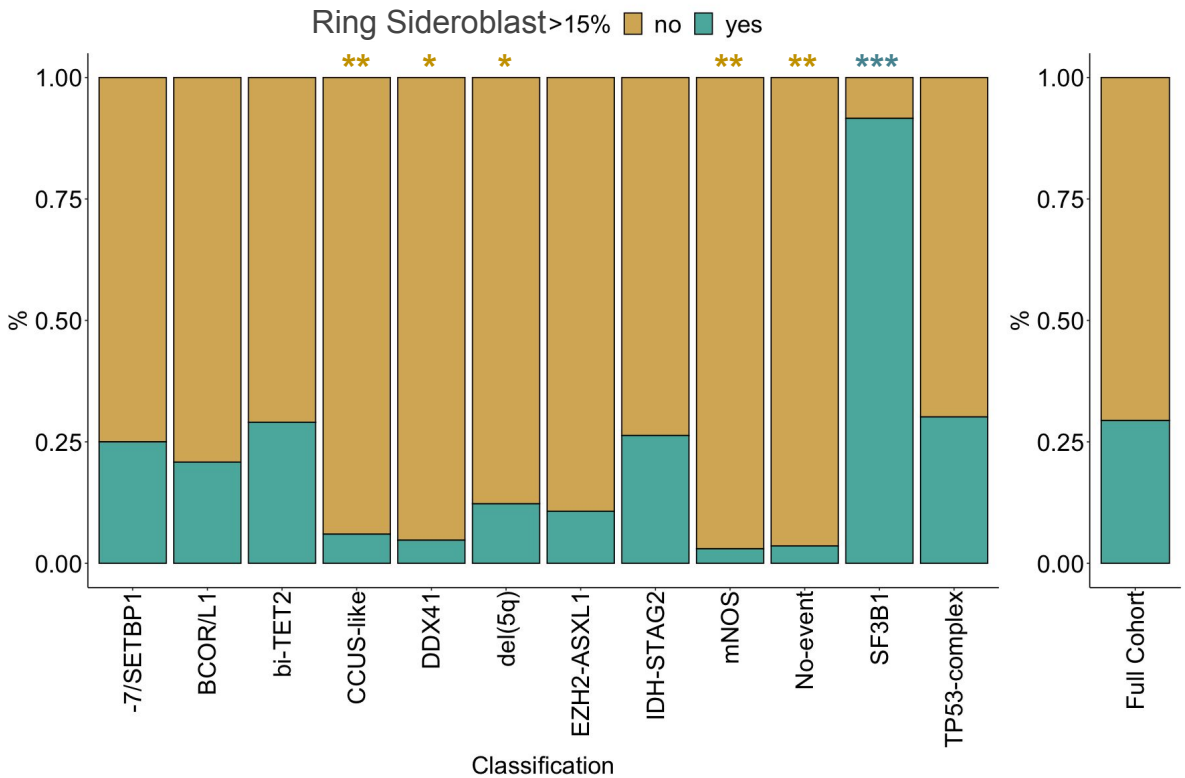
A-E. Distribution of the percentage of bone marrow blast (**A.**), percentage of ring sideroblast binarized at 15% (**B.**), hemoglobin level (**C.**), platelet count (**D.**), and white blood cell (WBC) count (**E.**) of patients within each molecular group. The distributions for the full cohort from 3,233 patients is depicted on the right for comparison. Statistical tests are depicted in the same way as explained in Figure S8A (**A.**, **C-E.**) and Figure S8B (**B.**).

For (**B.**) the analysis is restricted to the molecular groups with at least 20 cases with available ring sideroblast data.

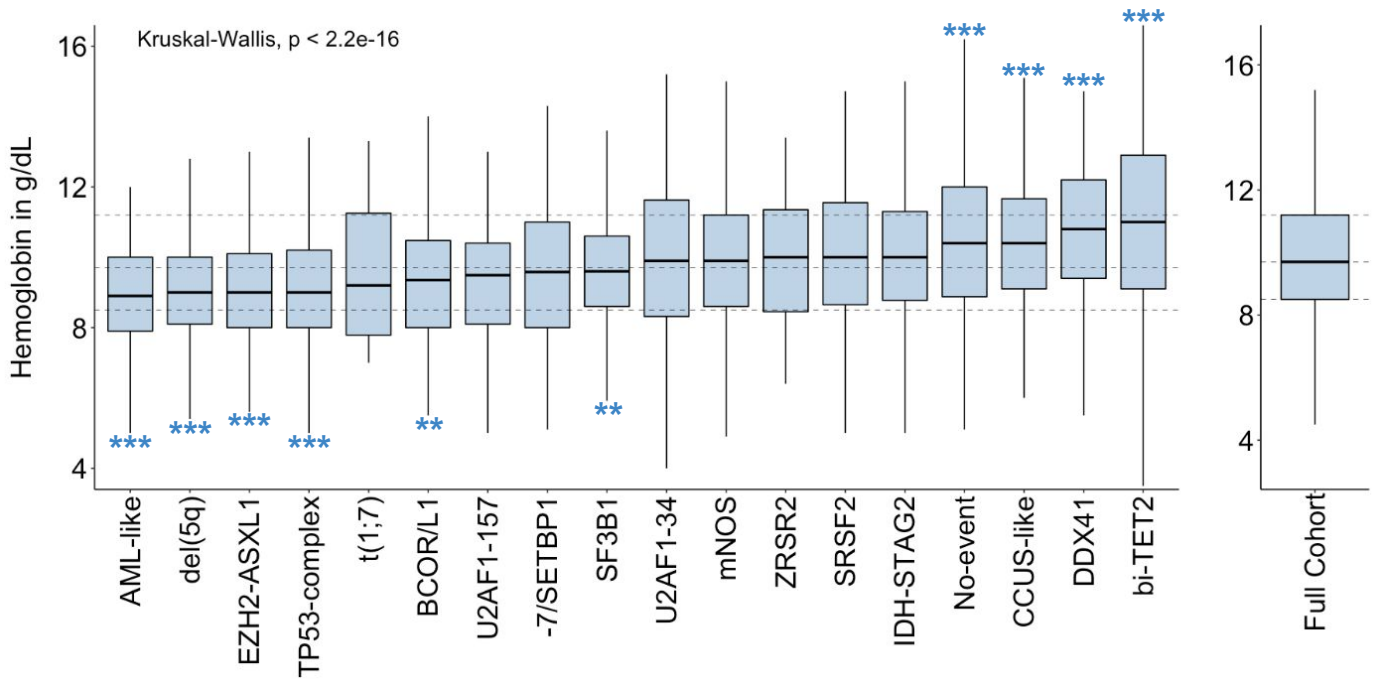
A.


Kruskal-Wallis, p < 2.2e-16
BM blasts in %
30
20
10
0
*** *** * *** ** ** *** *** ***
SF3B1
No-event
ZRSR2
U2AF1-34
CCUS-like
del(5q)
bi-TET2
mNOS
U2AF1-157
SRSF2
EZH2-ASXL1
-7/SETBP1
BCOR/L1
t(1;7)
IDH-STAG2
TP53-complex
DDX41
AML-like
Full Cohort
Classification

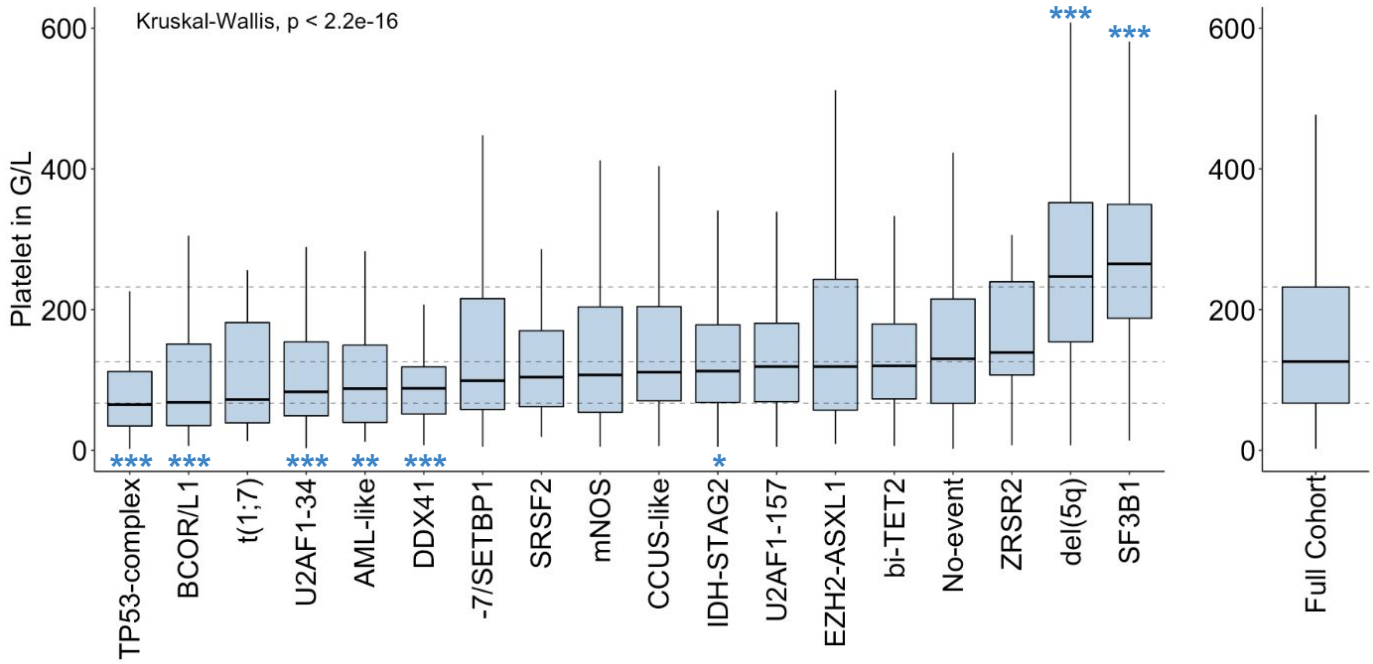
B.


Ring Sideroblast>15% no yes
** * * ** ** ***
1.00
0.75
0.50
0.25
0.00
%
-7/SETBP1
BCOR/L1
bi-TET2
CCUS-like
DDX41
del(5q)
EZH2-ASXL1
IDH-STAG2
mNOS
No-event
SF3B1
TP53-complex
Full Cohort
Classification

C.


Kruskal-Wallis, p < 2.2e-16
Hemoglobin in g/dL
16
12
8
4
AML-like
del(5q)
EZH2-ASXL1
TP53-complex
t(1;7)
BCOR/L1
U2AF1-157
-7/SETBP1
SF3B1
U2AF1-34
mNOS
ZRSR2
SRSF2
IDH-STAG2
No-event
CCUS-like
DDX41
bi-TET2
Full Cohort

D.


Kruskal-Wallis, p < 2.2e-16
Platelet in G/L
600
400
200
0
TP53-complex
BCOR/L1
t(1;7)
U2AF1-34
AML-like
DDX41
-7/SETBP1
SRSF2
mNOS
CCUS-like
IDH-STAG2
U2AF1-157
EZH2-ASXL1
bi-TET2
No-event
ZRSR2
del(5q)
SF3B1
Full Cohort

E.

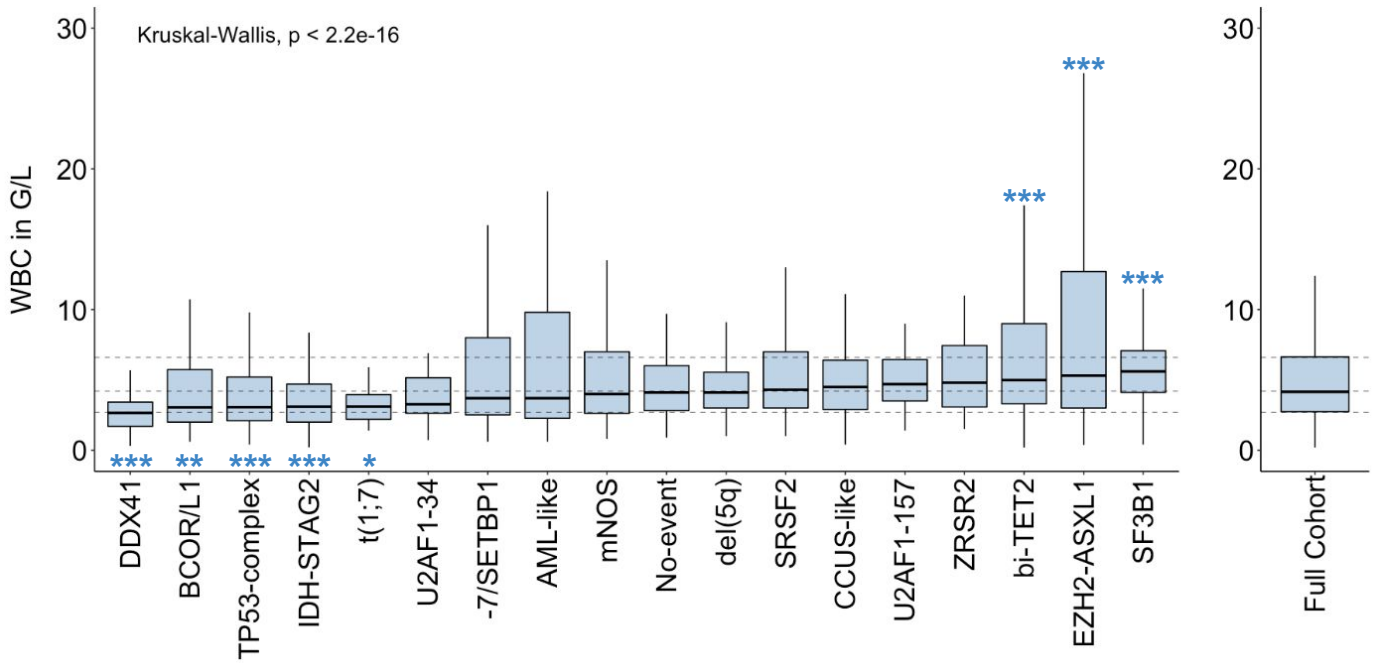

Kruskal-Wallis, p < 2.2e-16
WBC in G/L
30
20
10
0
DDX41
BCOR/L1
TP53-complex
IDH-STAG2
t(1;7)
U2AF1-34
-7/SETBP1
AML-like
mNOS
No-event
del(5q)
SRSF2
CCUS-like
U2AF1-157
ZRSR2
bi-TET2
EZH2-ASXL1
SF3B1
Full Cohort

Figure S10 | Association between MDS molecular groups and overall survival. Kaplan-Meier probability estimates of overall survival (OS) separating within each facet patients that are categorized in the molecular group (red curve) versus not (grey curve). P-values are from the log-rank test.

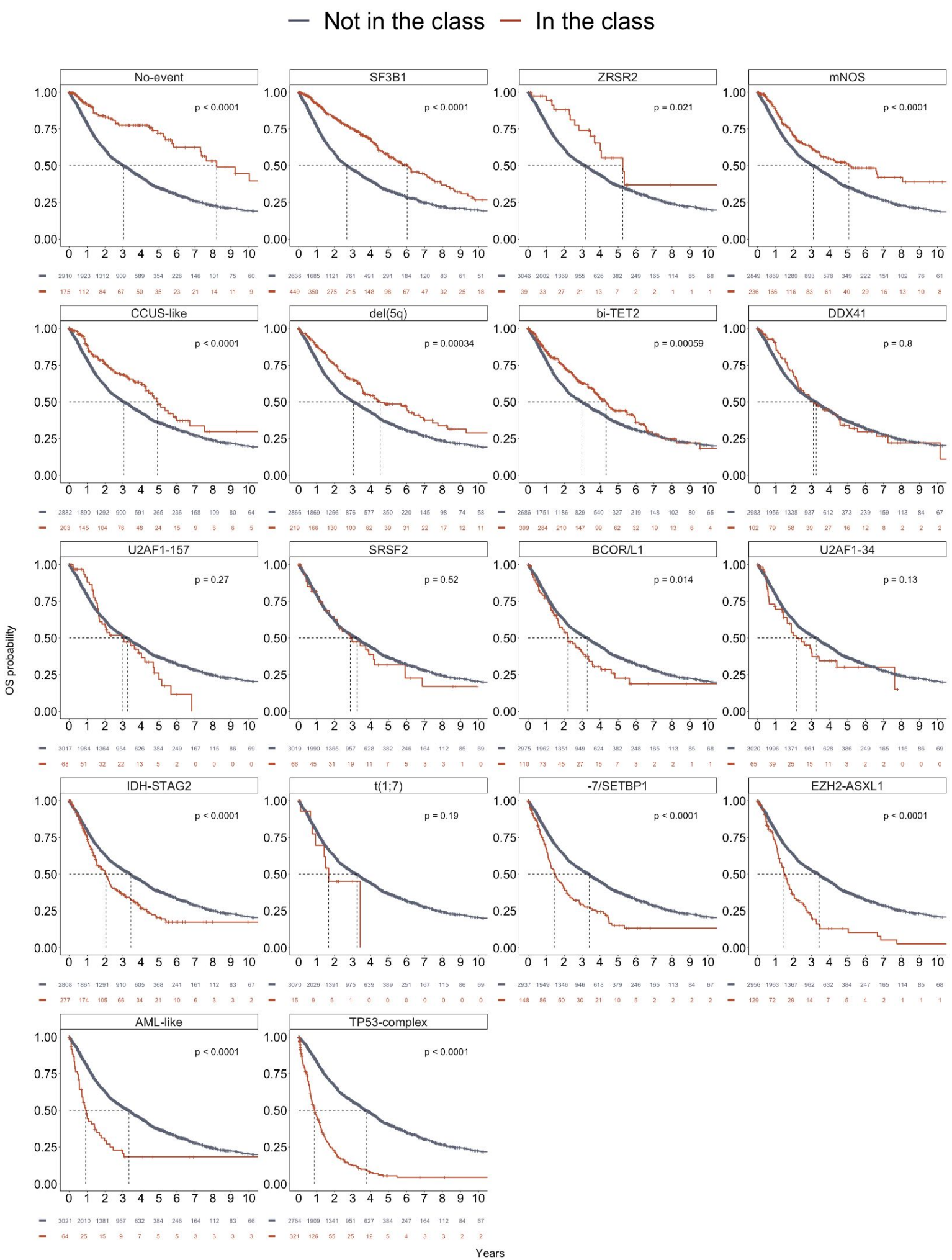

Not in the class
In the class
No-event
p < 0.0001
SF3B1
p < 0.0001
ZRSR2
p = 0.021
mNOS
p < 0.0001
CCUS-like
p < 0.0001
del(5q)
p = 0.00034
bi-TET2
p = 0.00059
DDX41
p = 0.8
U2AF1-157
p = 0.27
SRSF2
p = 0.52
BCOR/L1
p = 0.014
U2AF1-34
p = 0.13
IDH-STAG2
p < 0.0001
t(1;7)
p = 0.19
-7/SETBP1
p < 0.0001
EZH2-ASXL1
p < 0.0001
AML-like
p < 0.0001
TP53-complex
p < 0.0001
OS probability
Years

Figure S11 | Association between MDS molecular groups and leukemic transformation. Cumulative incidence curves of AML transformation separating within each facet patients that are categorized in the molecular group (red curve) versus not (grey curve). P-values are from the Gray's test.

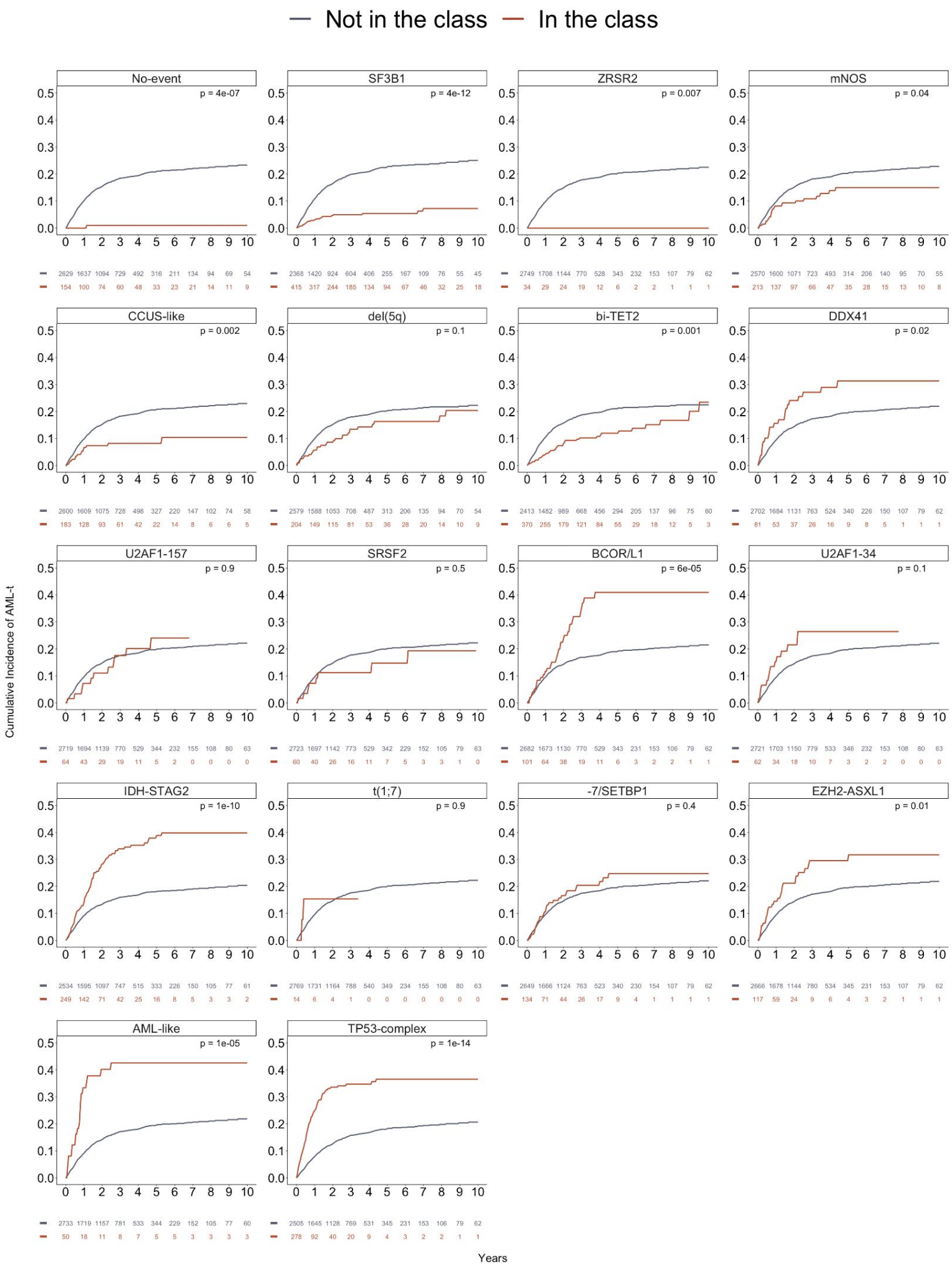

Not in the class
In the class
No-event
p = 4e-07
SF3B1
p = 4e-12
ZRSR2
p = 0.007
mNOS
p = 0.04
CCUS-like
p = 0.002
del(5q)
p = 0.1
bi-TET2
p = 0.001
DDX41
p = 0.02
U2AF1-157
p = 0.9
SRSF2
p = 0.5
BCOR/L1
p = 6e-05
U2AF1-34
p = 0.1
IDH-STAG2
p = 1e-10
t(1;7)
p = 0.9
-7/SETBP1
p = 0.4
EZH2-ASXL1
p = 0.01
AML-like
p = 1e-05
TP53-complex
p = 1e-14
Cumulative Incidence of AML-t
Years

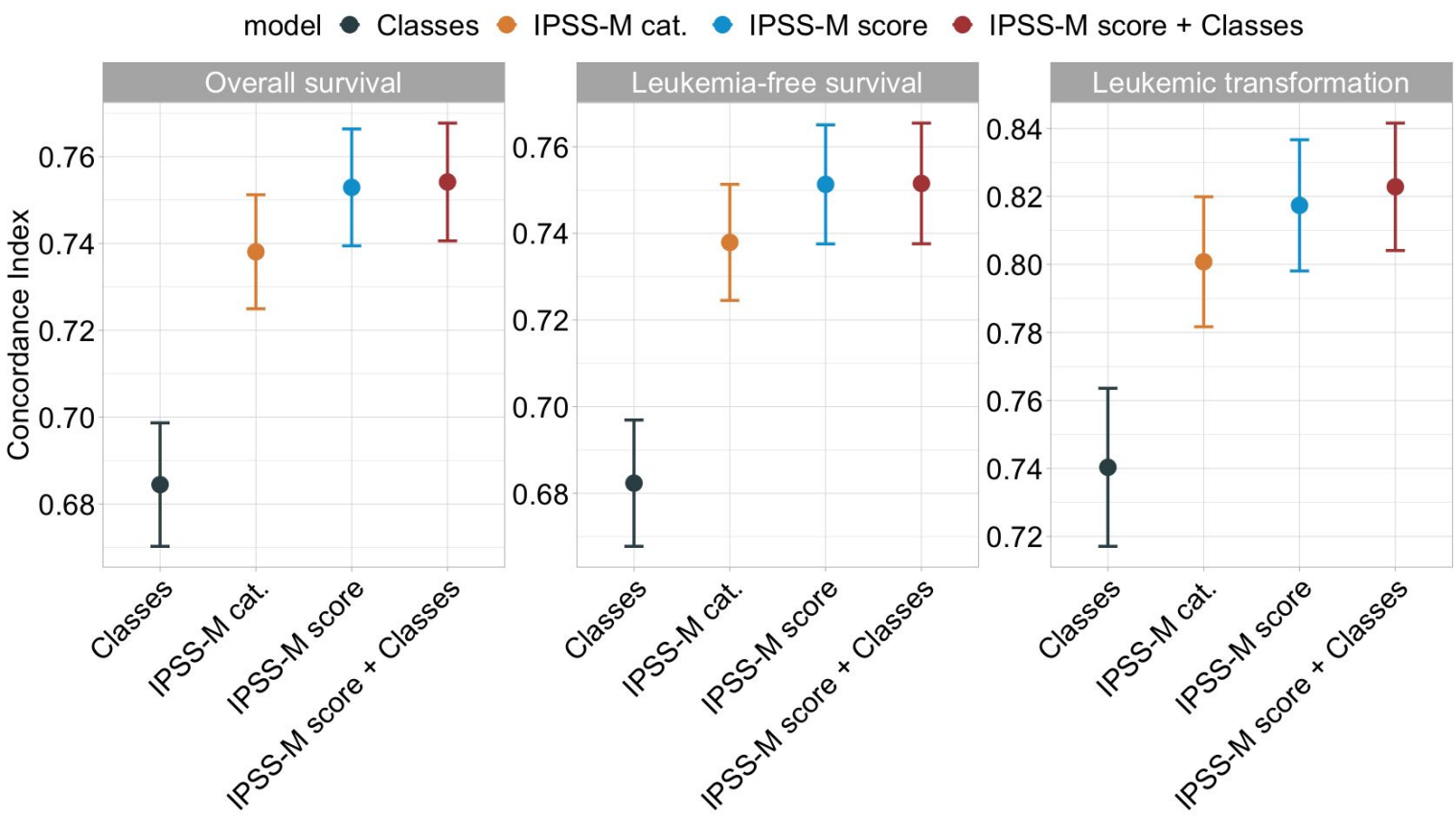
Figure S12 | Comparison of the predictive performances for patient outcomes for molecular group based models versus IPSS-M models.

A. Model discrimination as measured by the concordance index across three endpoints (left: overall survival; middle: leukemia free survival; right: leukemic transformation) obtained with different predictive models: a molecular group only model (back), a model using the IPSS-M risk category (orange) or the IPSS-M risk score (blue), and a model combining the IPSS-M risk score and the molecular group (red).

The predictive performances for clinical outcomes of a model considering molecular groups were inferior to the IPSS-M (e.g. 0.68 and 0.75 concordance index for OS, respectively). Adding molecular groups to the IPSS-M did not improve prognostication, consistent with previous studies comparing class-based versus gene-based models (Bersanelli et al., JCO 2021).

B. Kaplan-Meier probability estimates of overall survival (OS) stratified by IPSS-M risk category within each molecular group. P-values are from the log-rank test.

A.


model
Classes
IPSS-M cat.
IPSS-M score
IPSS-M score + Classes
Overall survival
Leukemia-free survival
Leukemic transformation
Concordance Index
0.76
0.74
0.72
0.70
0.68
0.84
0.82
0.80
0.78
Classes
IPSS-M cat.
IPSS-M score
IPSS-M score + Classes

B.

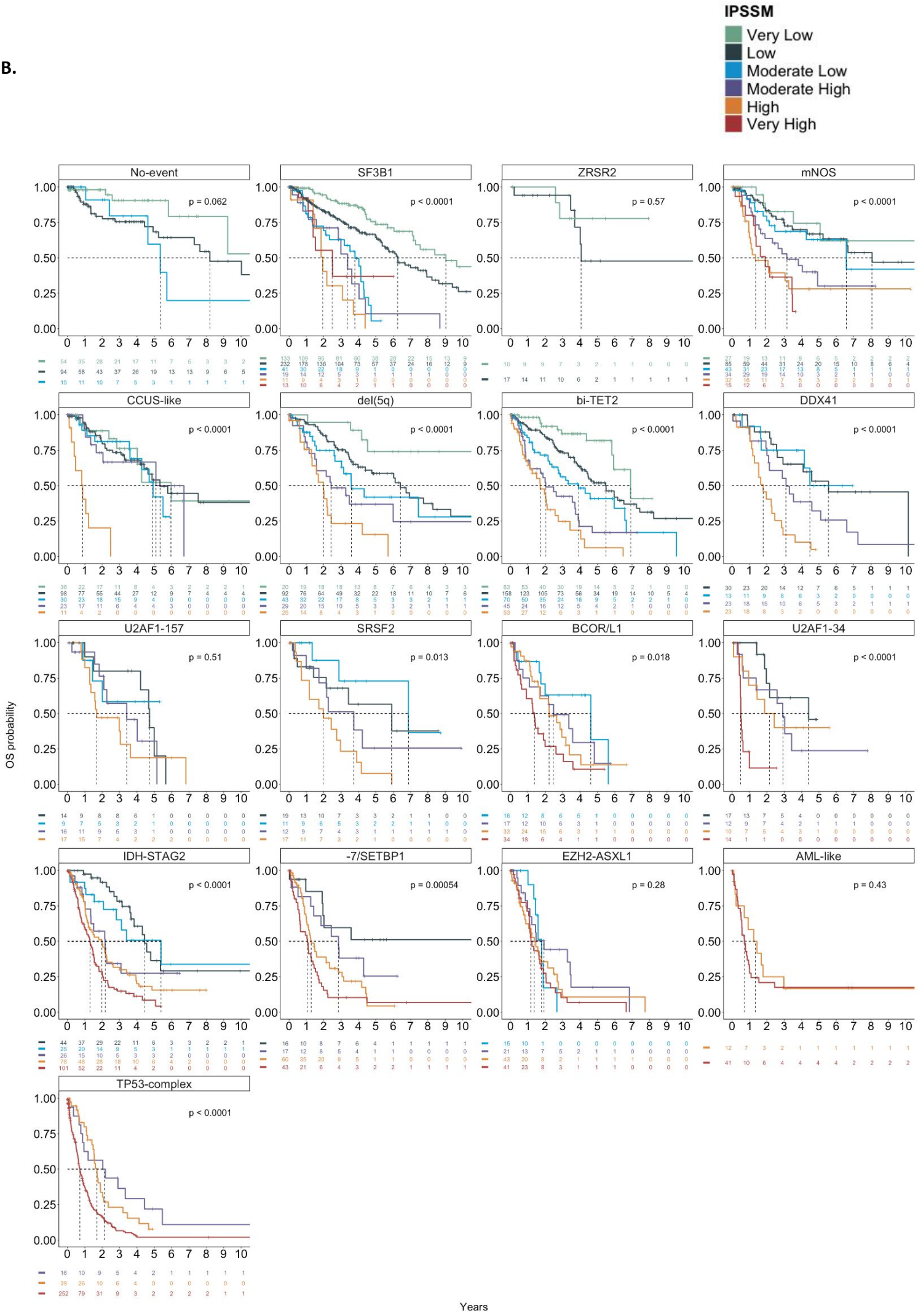

IPSSM
Very Low
Low
Moderate Low
Moderate High
High
Very High
No-event
p = 0.062
SF3B1
p < 0.0001
ZRSR2
p = 0.57
mNOS
p < 0.0001
CCUS-like
p < 0.0001
del(5q)
p < 0.0001
bi-TET2
p < 0.0001
DDX41
p < 0.0001
U2AF1-157
p = 0.51
SRSF2
p = 0.013
BCOR/L1
p = 0.018
U2AF1-34
p < 0.0001
IDH-STAG2
p < 0.0001
-7/SETBP1
p = 0.00054
EZH2-ASXL1
p = 0.28
AML-like
p = 0.43
TP53-complex
p < 0.0001
OS probability
Years

Figure S13 | Summary molecular and clinical metrics of the *DDX41* group. Two multi-panel figures, molecular (**A.**) and clinical (**B.**), are provided.

A. Molecular summary metrics comparing cases categorized within the group (blue) versus other groups (grey).

Top row. Boxplot distribution of the number of mutated genes per patient (left) and representation of the most frequent molecular alterations within the group (right).

Middle row. Boxplot and violin plots representing the distribution of the maximum values of variant allele fraction (VAF) (left) and of VAF adjusted for ploidy (middle left) per patient. Distribution of the VAF of all mutations identified in patients without (middle right) or with (right) adjustment for ploidy.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Bottom row. Density distribution of the VAF adjusted for ploidy of mutations from the most commonly mutated genes within the group.

B. Clinical summary metrics comparing cases categorized within the group (blue) versus all other groups (grey).

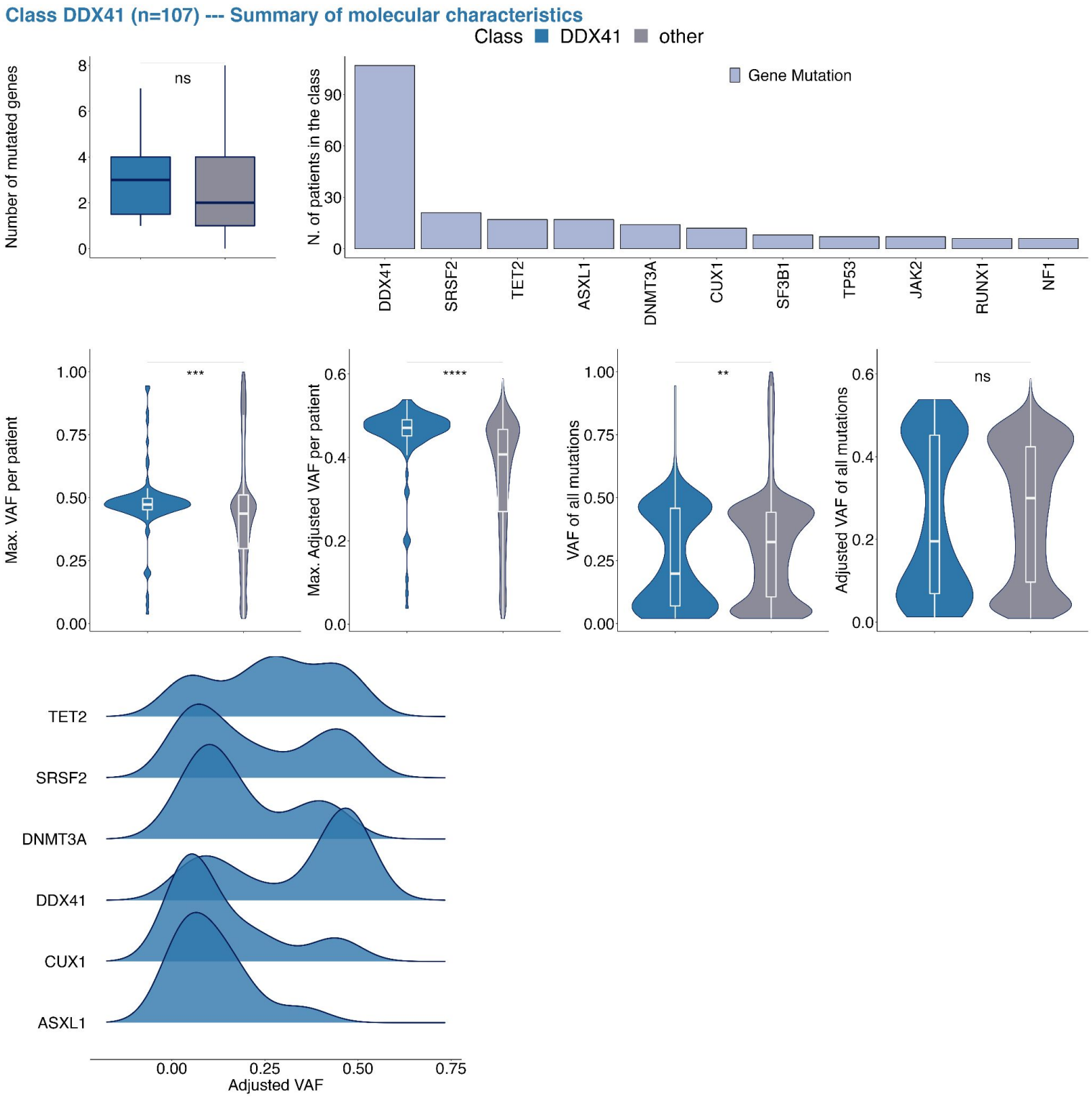
Top row. Distribution of the age at diagnosis, percentage of bone marrow blast, hemoglobin level, platelet count, white blood cell (WBC) count, and monocyte count.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Middle row. Left panel: Male (M) / Female (F) proportion. Right panels: Distribution of the WHO 2016 diagnosis subtypes (top) and of IPSS-M risk categories (bottom).

Bottom row. Left and middle panels: Kaplan-Meier probability estimates of overall survival (left) and leukemia free survival (middle). P-values are from the log-rank test. Right panel: cumulative incidence curves for the rate of leukemic transformation. P-value is from the Gray's test.

A.


Class DDX41 (n=107) --- Summary of molecular characteristics
Class DDX41 other
Number of mutated genes
ns
N. of patients in the class
Gene Mutation
DDX41
SRSF2
TET2
ASXL1
DNMT3A
CUX1
SF3B1
TP53
JAK2
RUNX1
NF1
Max. VAF per patient
***
Max. Adjusted VAF per patient
****
VAF of all mutations
**
Adjusted VAF of all mutations
ns
TET2
SRSF2
DNMT3A
DDX41
CUX1
ASXL1
Adjusted VAF

B.

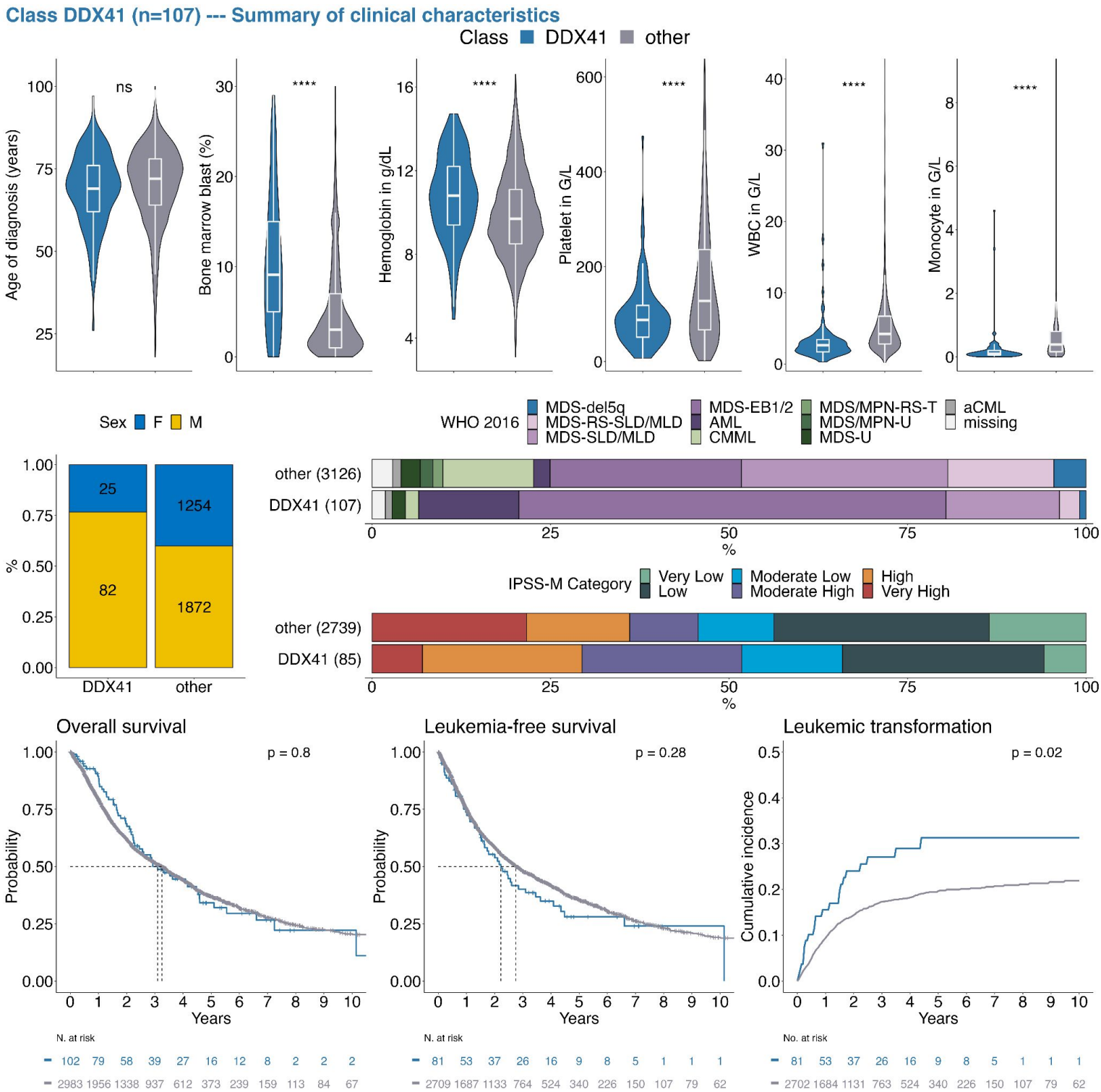

Class DDX41 (n=107) --- Summary of clinical characteristics
Class DDX41 other
Age of diagnosis (years)
ns
Bone marrow blast (%)
****
Hemoglobin in g/dL
****
Platelet in G/L
****
WBC in G/L
****
Monocyte in G/L
****
Sex F M
25
82
1254
1872
DDX41
other
% 0.00 0.25 0.50 0.75 1.00
WHO 2016
MDS-del5q
MDS-RS-SLD/MLD
MDS-SLD/MLD
MDS-EB1/2
AML
CMML
MDS/MPN-RS-T
MDS/MPN-U
MDS-U
aCML
missing
other (3126)
DDX41 (107)
0 25 50 75 100
%
IPSS-M Category
Very Low
Low
Moderate Low
Moderate High
High
Very High
other (2739)
DDX41 (85)
0 25 50 75 100
%
Overall survival
p = 0.8
Probability 0.00 0.25 0.50 0.75 1.00
0 1 2 3 4 5 6 7 8 9 10
Years
N. at risk
102 79 58 39 27 16 12 8 2 2 2
2983 1956 1338 937 612 373 239 159 113 84 67
Leukemia-free survival
p = 0.28
Probability 0.00 0.25 0.50 0.75 1.00
0 1 2 3 4 5 6 7 8 9 10
Years
N. at risk
81 53 37 26 16 9 8 5 1 1 1
2709 1687 1133 764 524 340 226 150 107 79 62
Leukemic transformation
p = 0.02
Cumulative incidence 0.0 0.1 0.2 0.3 0.4 0.5
0 1 2 3 4 5 6 7 8 9 10
Years
No. at risk
81 53 37 26 16 9 8 5 1 1 1
2702 1684 1131 763 524 340 226 150 107 79 62

Figure S14 | Summary molecular and clinical metrics of the *AML-like* group. Two multi-panel figures, molecular (**A.**) and clinical (**B.**), are provided.

A. Molecular summary metrics comparing cases categorized within the group (blue) versus other groups (grey).

Top row. Boxplot distribution of the number of mutated genes per patient (left) and representation of the most frequent molecular alterations within the group (right).

Middle row. Boxplot and violin plots representing the distribution of the maximum values of variant allele fraction (VAF) (left) and of VAF adjusted for ploidy (middle left) per patient. Distribution of the VAF of all mutations identified in patients without (middle right) or with (right) adjustment for ploidy.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Bottom row. Density distribution of the VAF adjusted for ploidy of mutations from the most commonly mutated genes within the group (left). Scatter plot representing the VAF of *NPM1* and *DNMT3A* mutations from patients harboring both of those mutations (right).

B. Clinical summary metrics comparing cases categorized within the group (blue) versus all other groups (grey).

Top row. Distribution of the age at diagnosis, percentage of bone marrow blast, hemoglobin level, platelet count, white blood cell (WBC) count, and monocyte count.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Middle row. Left panel: Male (M) / Female (F) proportion. Right panels: Distribution of the WHO 2016 diagnosis subtypes (top) and of IPSS-M risk categories (bottom).

Bottom row. Left and middle panels: Kaplan-Meier probability estimates of overall survival (left) and leukemia free survival (middle). P-values are from the log-rank test. Right panel: cumulative incidence curves for the rate of leukemic transformation. P-value is from the Gray's test.

C. Clinical summary metrics of subgroups of *AML-like* comparing cases with AML-defining features (*NPM1* mutation, inv(3)) (gold) versus other cases with 2 or more mutations in (*WT1*, *FLT3*, *MLL* PTD, *MYC*) (green).

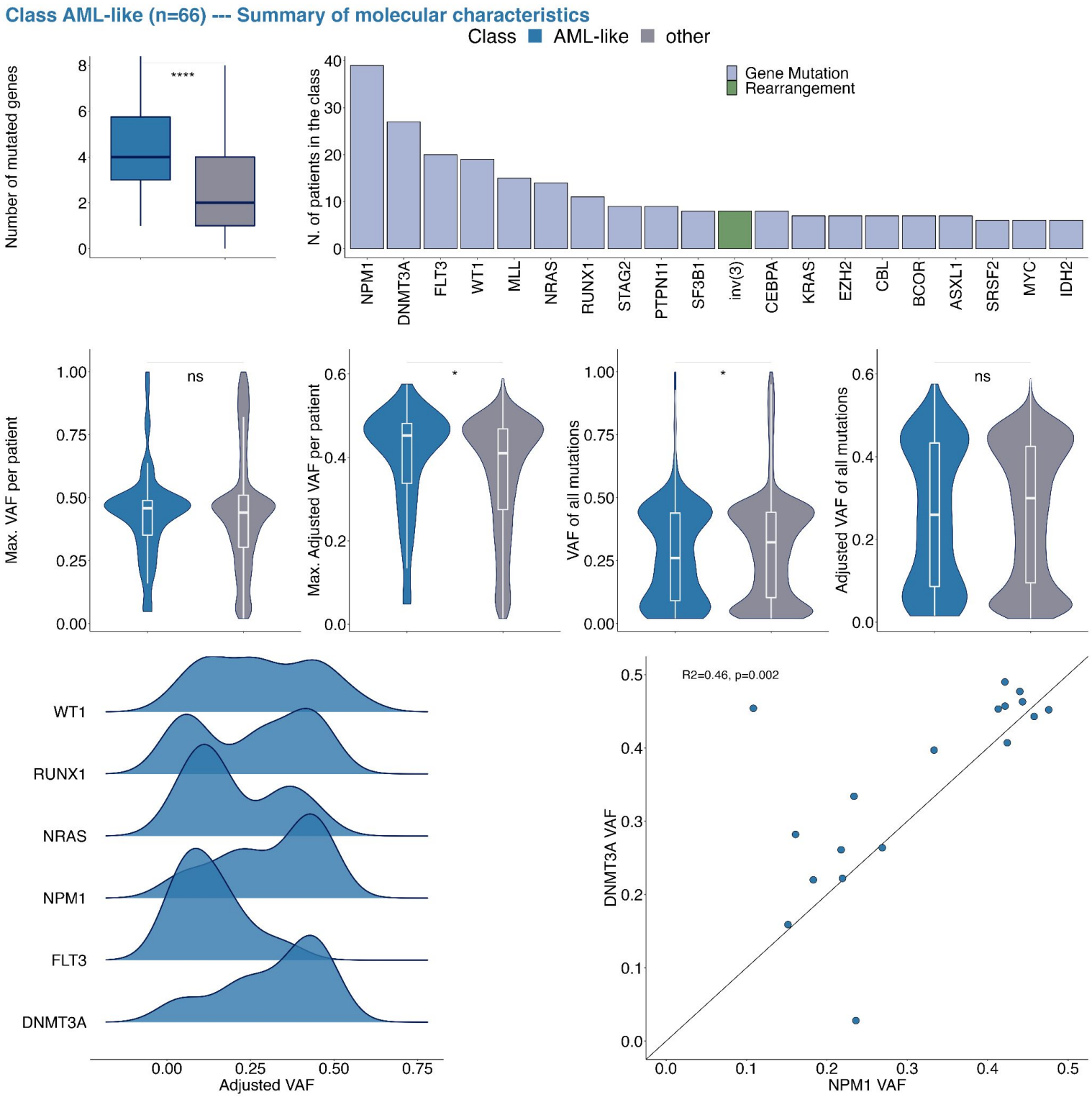
Top row. Distribution of the age at diagnosis, percentage of bone marrow blast, hemoglobin level, platelet count, white blood cell (WBC) count, and monocyte count.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

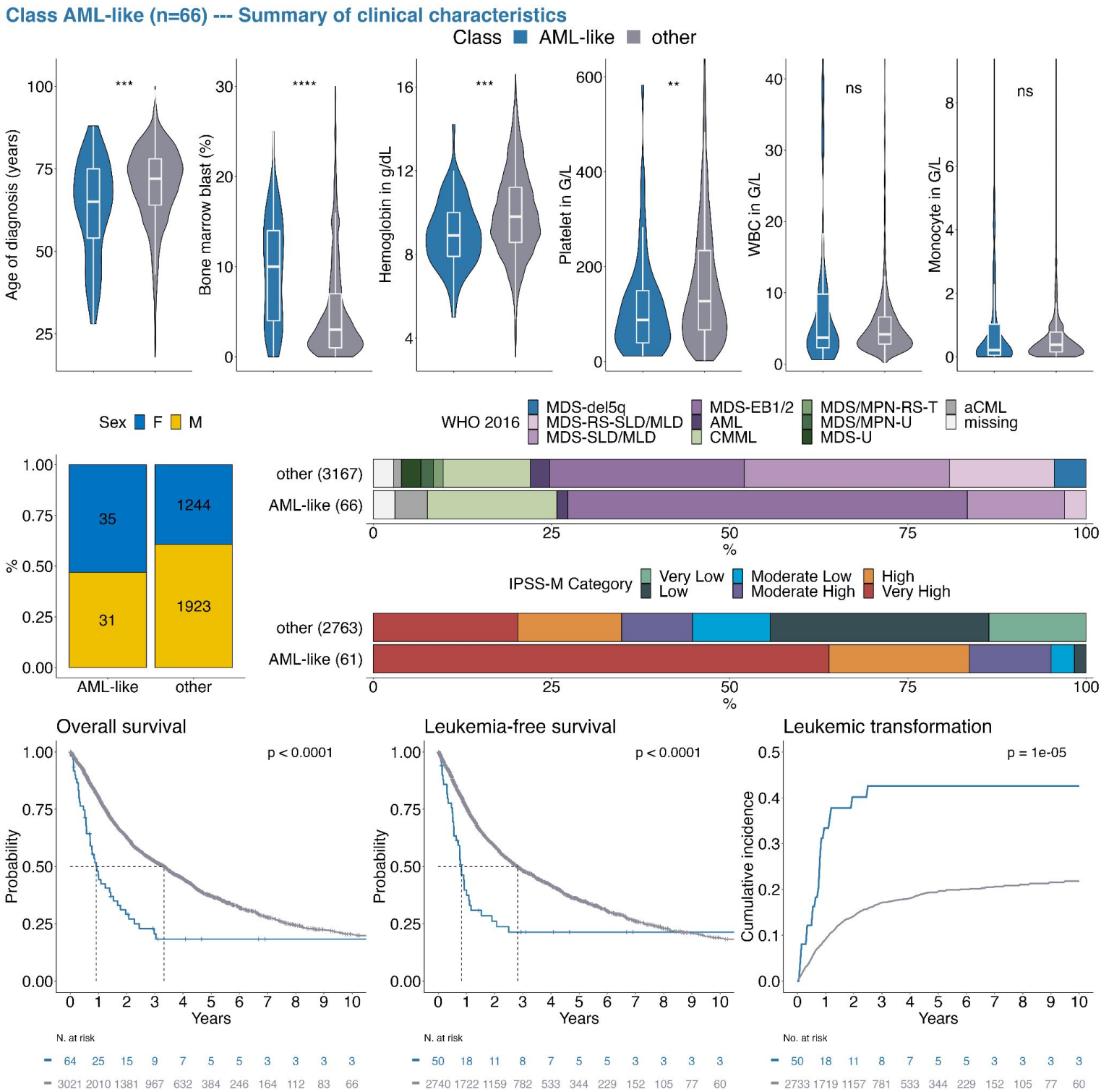
Middle row. Left panel: Male (M) / Female (F) proportion. Right panels: Distribution of the WHO 2016 diagnosis subtypes (top) and of IPSS-M risk categories (bottom).

Bottom row. Left and middle panels: Kaplan-Meier probability estimates of overall survival (left) and leukemia free survival (middle). P-values are from the log-rank test. Right panel: cumulative incidence curves for the rate of leukemic transformation. P-value is from the Gray's test.

A.


Class AML-like (n=66) --- Summary of molecular characteristics
Class AML-like other
Number of mutated genes
****
N. of patients in the class
Gene Mutation
Rearrangement
NPM1
DNMT3A
FLT3
WT1
MLL
NRAS
RUNX1
STAG2
PTPN11
SF3B1
inv(3)
CEBPA
KRAS
EZH2
CBL
BCOR
ASXL1
SRSF2
MYC
IDH2
Max. VAF per patient
ns
Max. Adjusted VAF per patient
*
VAF of all mutations
*
Adjusted VAF of all mutations
ns
WT1
RUNX1
NRAS
NPM1
FLT3
DNMT3A
Adjusted VAF
R2=0.46, p=0.002
DNMT3A VAF
NPM1 VAF

B.


Class AML-like (n=66) --- Summary of clinical characteristics
Class AML-like other
Age of diagnosis (years)
Bone marrow blast (%)
Hemoglobin in g/dL
Platelet in G/L
WBC in G/L
Monocyte in G/L
***
****
***
**
ns
ns
Sex F M
35
31
1244
1923
AML-like
other
WHO 2016
MDS-del5q
MDS-RS-SLD/MLD
MDS-SLD/MLD
MDS-EB1/2
AML
CMML
MDS/MPN-RS-T
MDS/MPN-U
MDS-U
aCML
missing
other (3167)
AML-like (66)
IPSS-M Category
Very Low
Low
Moderate Low
Moderate High
High
Very High
other (2763)
AML-like (61)
%
Overall survival
Leukemia-free survival
Leukemic transformation
p < 0.0001
p < 0.0001
p = 1e-05
Probability
Cumulative incidence
Years
N. at risk
No. at risk
64 25 15 9 7 5 5 3 3 3 3
3021 2010 1381 967 632 384 246 164 112 83 66
50 18 11 8 7 5 5 3 3 3 3
2740 1722 1159 782 533 344 229 152 105 77 60
50 18 11 8 7 5 5 3 3 3 3
2733 1719 1157 781 533 344 229 152 105 77 60

C.

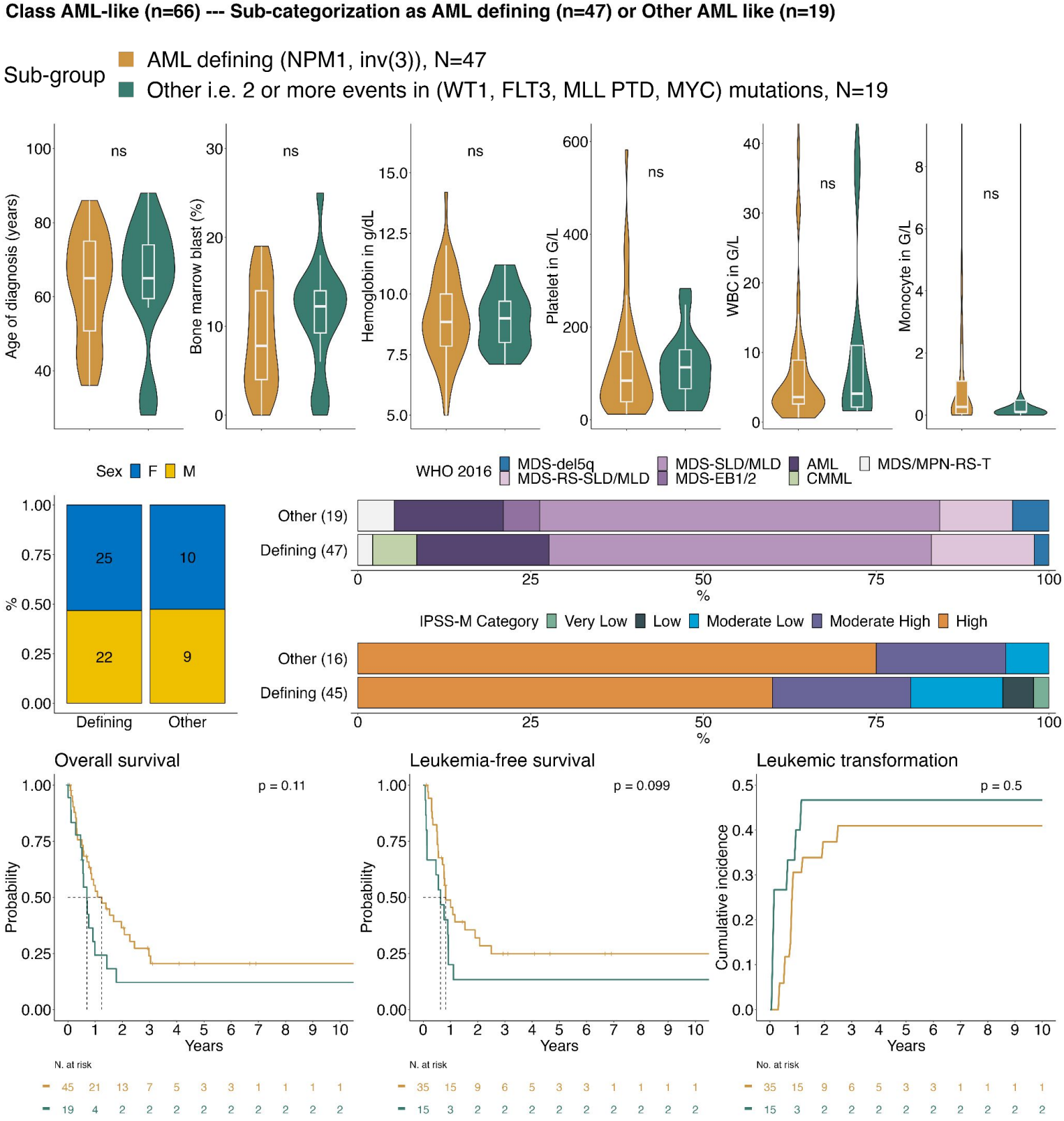

Class AML-like (n=66) --- Sub-categorization as AML defining (n=47) or Other AML like (n=19)
Sub-group
AML defining (NPM1, inv(3)), N=47
Other i.e. 2 or more events in (WT1, FLT3, MLL PTD, MYC) mutations, N=19
Age of diagnosis (years)
Bone marrow blast (%)
Hemoglobin in g/dL
Platelet in G/L
WBC in G/L
Monocyte in G/L
ns
Sex F M
25
10
22
9
Defining
Other
%
WHO 2016
MDS-del5q
MDS-RS-SLD/MLD
MDS-SLD/MLD
MDS-EB1/2
AML
CMML
MDS/MPN-RS-T
Other (19)
Defining (47)
IPSS-M Category
Very Low
Low
Moderate Low
Moderate High
High
Other (16)
Defining (45)
Overall survival
p = 0.11
Leukemia-free survival
p = 0.099
Leukemic transformation
p = 0.5
Probability
Cumulative incidence
Years
N. at risk
No. at risk
45 21 13 7 5 3 3 1 1 1 1
19 4 2 2 2 2 2 2 2 2 2
35 15 9 6 5 3 3 1 1 1 1
15 3 2 2 2 2 2 2 2 2 2

Figure S15 | Summary molecular and clinical metrics of the *TP53-complex* group. Two multi-panel figures, molecular (**A.**) and clinical (**B.**), are provided.

A. Molecular summary metrics comparing cases categorized within the group (blue) versus other groups (grey).

Top row. Boxplot distribution of the number of mutated genes per patient (left) and representation of the most frequent molecular alterations within the group (right).

Middle row. Boxplot and violin plots representing the distribution of the maximum values of variant allele fraction (VAF) (left) and of VAF adjusted for ploidy (middle left) per patient. Distribution of the VAF of all mutations identified in patients without (middle right) or with (right) adjustment for ploidy.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Bottom row. Density distribution of the VAF adjusted for ploidy of mutations from the most commonly mutated genes within the group (left). Relative order of gene mutation acquisition inferred through Bradley-Terry analysis. The numbers (n) indicate the number of informative pairwise precedences per gene (right).

B. Clinical summary metrics comparing cases categorized within the group (blue) versus all other groups (grey).

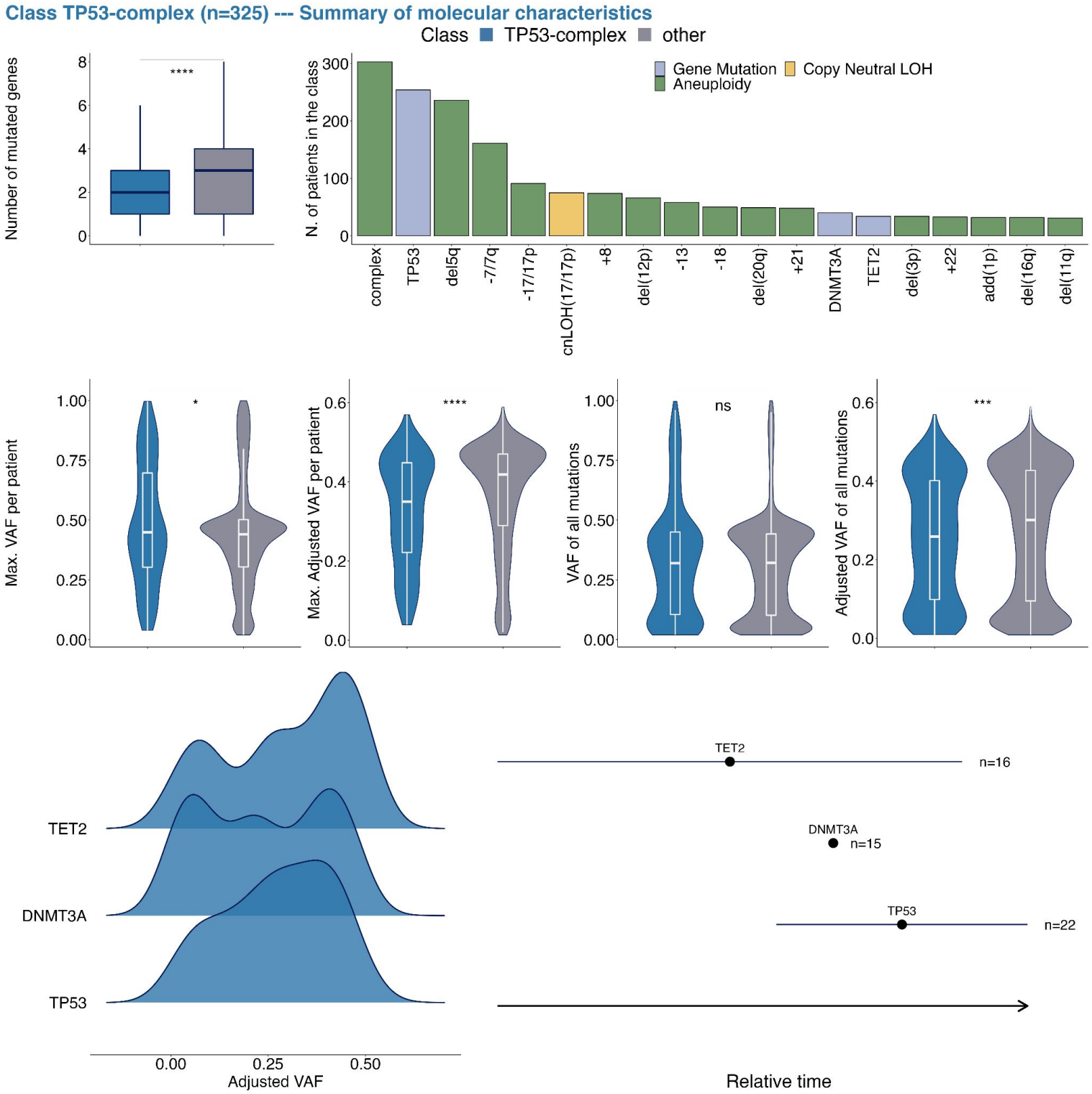
Top row. Distribution of the age at diagnosis, percentage of bone marrow blast, hemoglobin level, platelet count, white blood cell (WBC) count, and monocyte count.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Middle row. Left panel: Male (M) / Female (F) proportion. Right panels: Distribution of the WHO 2016 diagnosis subtypes (top) and of IPSS-M risk categories (bottom).

Bottom row. Left and middle panels: Kaplan-Meier probability estimates of overall survival (left) and leukemia free survival (middle). P-values are from the log-rank test. Right panel: cumulative incidence curves for the rate of leukemic transformation. P-value is from the Gray's test.

A.


Class TP53-complex (n=325) --- Summary of molecular characteristics
Class TP53-complex other
Number of mutated genes
****
N. of patients in the class
Gene Mutation
Copy Neutral LOH
Aneuploidy
complex
TP53
del5q
-7/7q
-17/17p
cnLOH(17/17p)
+8
del(12p)
-13
-18
del(20q)
+21
DNMT3A
TET2
del(3p)
+22
add(1p)
del(16q)
del(11q)
Max. VAF per patient
*
Max. Adjusted VAF per patient
****
VAF of all mutations
ns
Adjusted VAF of all mutations
***
TET2
DNMT3A
TP53
Adjusted VAF
Relative time
TET2
n=16
DNMT3A
n=15
TP53
n=22

B.

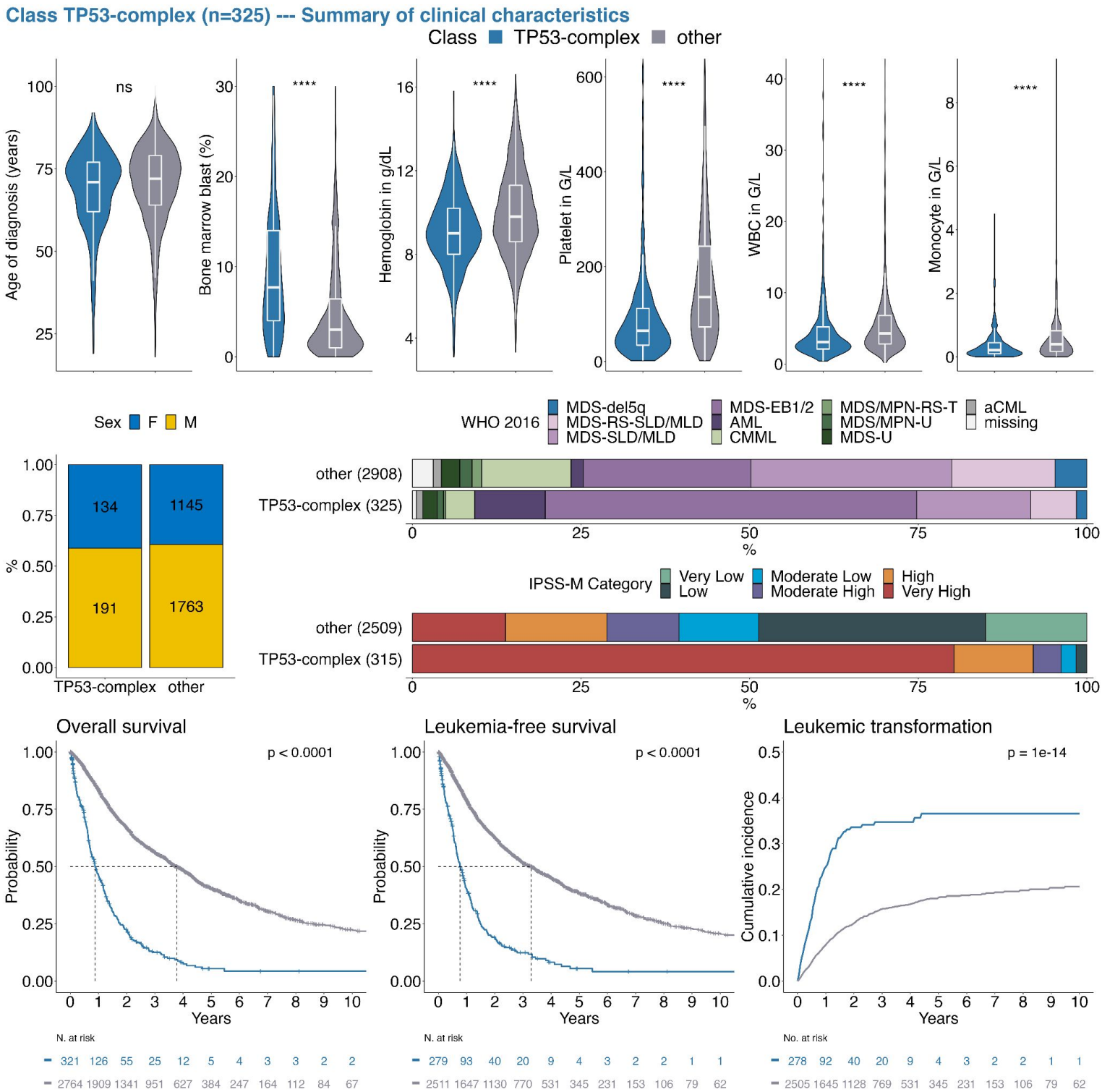

Class TP53-complex (n=325) --- Summary of clinical characteristics
Class TP53-complex other
Age of diagnosis (years)
ns
Bone marrow blast (%)
****
Hemoglobin in g/dL
****
Platelet in G/L
****
WBC in G/L
****
Monocyte in G/L
****
Sex F M
134
1145
191
1763
TP53-complex other
WHO 2016
MDS-del5q
MDS-RS-SLD/MLD
MDS-SLD/MLD
MDS-EB1/2
AML
CMML
MDS/MPN-RS-T
MDS/MPN-U
MDS-U
aCML
missing
other (2908)
TP53-complex (325)
IPSS-M Category
Very Low
Low
Moderate Low
Moderate High
High
Very High
other (2509)
TP53-complex (315)
Overall survival
p < 0.0001
Leukemia-free survival
p < 0.0001
Leukemic transformation
p = 1e-14
Probability
Cumulative incidence
Years
N. at risk
321 126 55 25 12 5 4 3 3 2 2
2764 1909 1341 951 627 384 247 164 112 84 67
279 93 40 20 9 4 3 2 2 1 1
2511 1647 1130 770 531 345 231 153 106 79 62
No. at risk
278 92 40 20 9 4 3 2 2 1 1
2505 1645 1128 769 531 345 231 153 106 79 62

Figure S16 | Summary molecular and clinical metrics of the *del(5q)* group. Two multi-panel figures, molecular (**A.**) and clinical (**B.**), are provided.

A. Molecular summary metrics comparing cases categorized within the group (blue) versus other groups (grey).

Top row. Boxplot distribution of the number of mutated genes per patient (left) and representation of the most frequent molecular alterations within the group (right).

Middle row. Boxplot and violin plots representing the distribution of the maximum values of variant allele fraction (VAF) (left) and of VAF adjusted for ploidy (middle left) per patient. Distribution of the VAF of all mutations identified in patients without (middle right) or with (right) adjustment for ploidy.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Bottom row. Density distribution of the VAF adjusted for ploidy of mutations from the most commonly mutated genes within the group (left).

B. Clinical summary metrics comparing cases categorized within the group (blue) versus all other groups (grey).

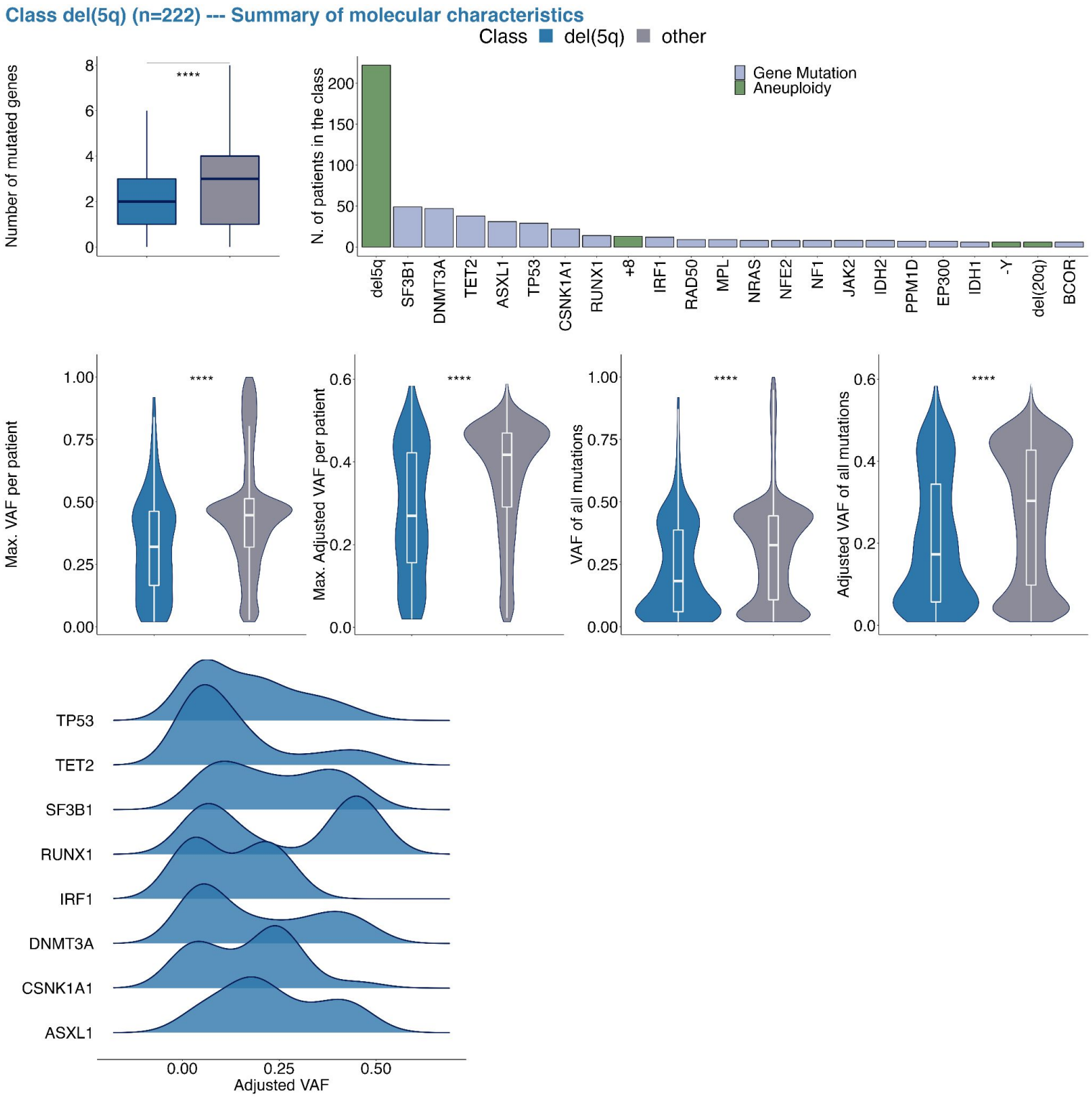
Top row. Distribution of the age at diagnosis, percentage of bone marrow blast, hemoglobin level, platelet count, white blood cell (WBC) count, and monocyte count.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Middle row. Left panel: Male (M) / Female (F) proportion. Right panels: Distribution of the WHO 2016 diagnosis subtypes (top) and of IPSS-M risk categories (bottom).

Bottom row. Left and middle panels: Kaplan-Meier probability estimates of overall survival (left) and leukemia free survival (middle). P-values are from the log-rank test. Right panel: cumulative incidence curves for the rate of leukemic transformation. P-value is from the Gray's test.

A.


Class del(5q) (n=222) --- Summary of molecular characteristics
Class del(5q) other
Number of mutated genes
****
N. of patients in the class
Gene Mutation
Aneuploidy
del5q
SF3B1
DNMT3A
TET2
ASXL1
TP53
CSNK1A1
RUNX1
+8
IRF1
RAD50
MPL
NRAS
NFE2
NF1
JAK2
IDH2
PPM1D
EP300
IDH1
-Y
del(20q)
BCOR
Max. VAF per patient
Max. Adjusted VAF per patient
VAF of all mutations
Adjusted VAF of all mutations
TP53
TET2
SF3B1
RUNX1
IRF1
DNMT3A
CSNK1A1
ASXL1
Adjusted VAF

B.

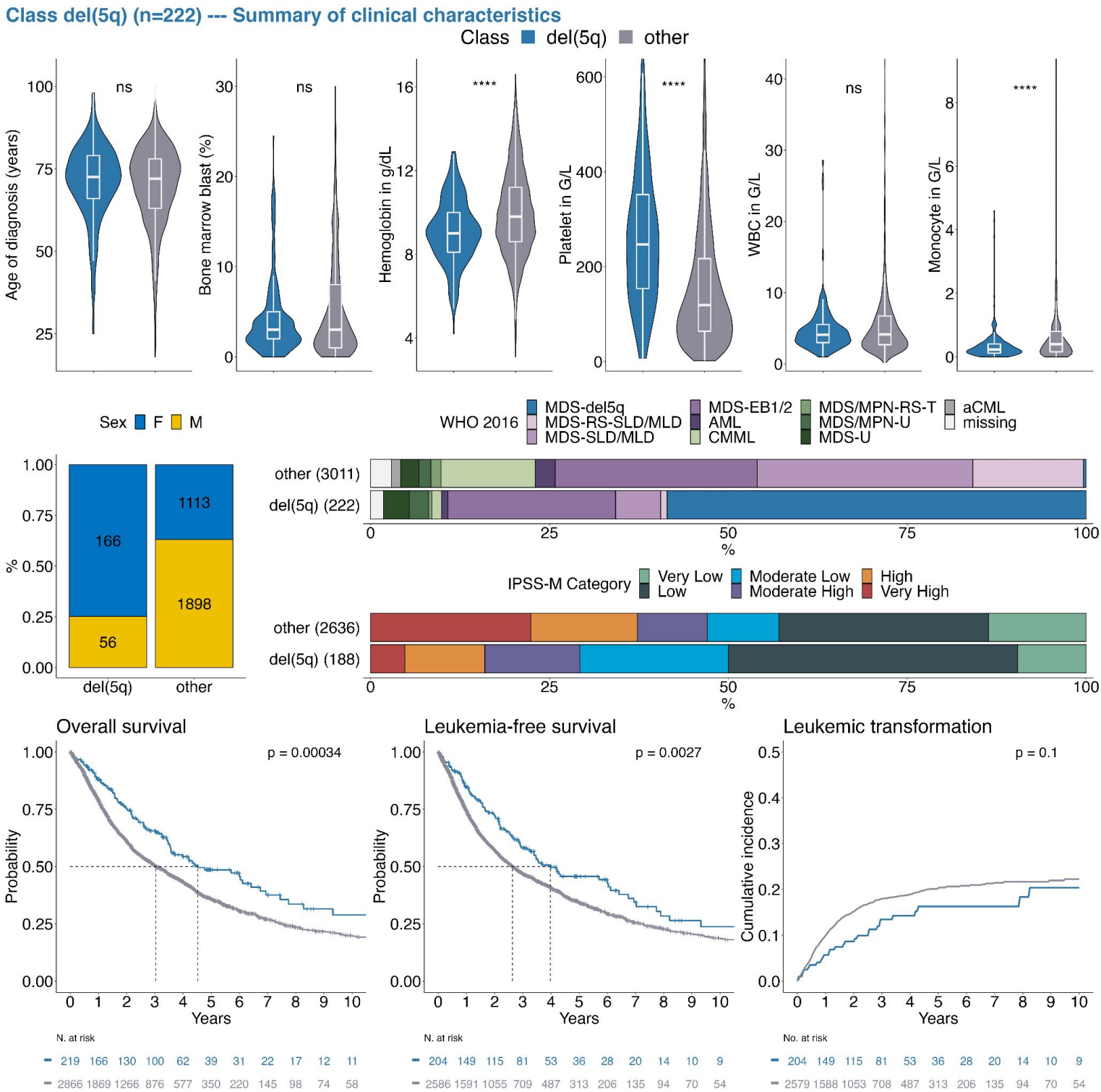

Class del(5q) (n=222) --- Summary of clinical characteristics
Class del(5q) other
Age of diagnosis (years)
Bone marrow blast (%)
Hemoglobin in g/dL
Platelet in G/L
WBC in G/L
Monocyte in G/L
ns
****
Sex F M
WHO 2016
MDS-del5q
MDS-RS-SLD/MLD
MDS-SLD/MLD
MDS-EB1/2
AML
CMML
MDS/MPN-RS-T
MDS/MPN-U
MDS-U
aCML
missing
other (3011)
del(5q) (222)
IPSS-M Category
Very Low
Low
Moderate Low
Moderate High
High
Very High
other (2636)
del(5q) (188)
del(5q)
other
Overall survival
Leukemia-free survival
Leukemic transformation
p = 0.00034
p = 0.0027
p = 0.1
Probability
Cumulative incidence
Years
N. at risk
No. at risk

Figure S17 | Summary molecular and clinical metrics of the *SF3B1* group. Two multi-panel figures, molecular (**A.**) and clinical (**B.**), are provided.

A. Molecular summary metrics comparing cases categorized within the group (blue) versus other groups (grey).

Top row. Boxplot distribution of the number of mutated genes per patient (left) and representation of the most frequent molecular alterations within the group (right).

Middle row. Boxplot and violin plots representing the distribution of the maximum values of variant allele fraction (VAF) (left) and of VAF adjusted for ploidy (middle left) per patient. Distribution of the VAF of all mutations identified in patients without (middle right) or with (right) adjustment for ploidy.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Bottom row. Density distribution of the VAF adjusted for ploidy of mutations from the most commonly mutated genes within the group (left). Relative order of gene mutation acquisition inferred through Bradley-Terry analysis. The numbers (n) indicate the number of informative pairwise precedences per gene (right).

B. Clinical summary metrics comparing cases categorized within the group (blue) versus all other groups (grey).

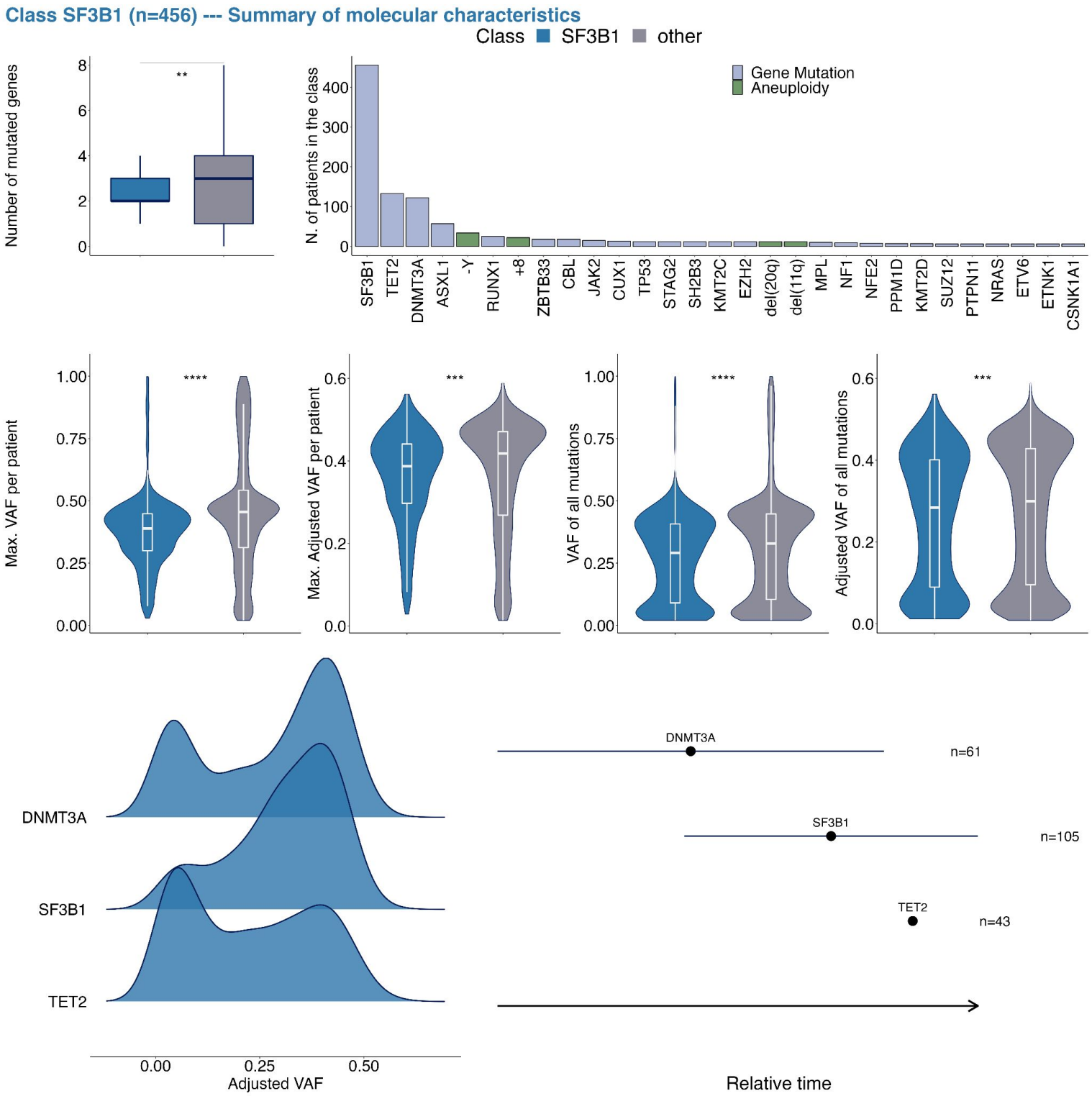
Top row. Distribution of the age at diagnosis, percentage of bone marrow blast, hemoglobin level, platelet count, white blood cell (WBC) count, and monocyte count.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Middle row. Left panel: Male (M) / Female (F) proportion. Right panels: Distribution of the WHO 2016 diagnosis subtypes (top) and of IPSS-M risk categories (bottom).

Bottom row. Left and middle panels: Kaplan-Meier probability estimates of overall survival (left) and leukemia free survival (middle). P-values are from the log-rank test. Right panel: cumulative incidence curves for the rate of leukemic transformation. P-value is from the Gray's test.

A.


Class SF3B1 (n=456) --- Summary of molecular characteristics
Class SF3B1 other
Number of mutated genes
**
N. of patients in the class
Gene Mutation
Aneuploidy
SF3B1
TET2
DNMT3A
ASXL1
-Y
RUNX1
+8
ZBTB33
CBL
JAK2
CUX1
TP53
STAG2
SH2B3
KMT2C
EZH2
del(20q)
del(11q)
MPL
NF1
NFE2
PPM1D
KMT2D
SUZ12
PTPN11
NRAS
ETV6
ETNK1
CSNK1A1
Max. VAF per patient
****
Max. Adjusted VAF per patient
***
VAF of all mutations
****
Adjusted VAF of all mutations
***
DNMT3A
SF3B1
TET2
0.00
0.25
0.50
Adjusted VAF
DNMT3A
n=61
SF3B1
n=105
TET2
n=43
Relative time

B.

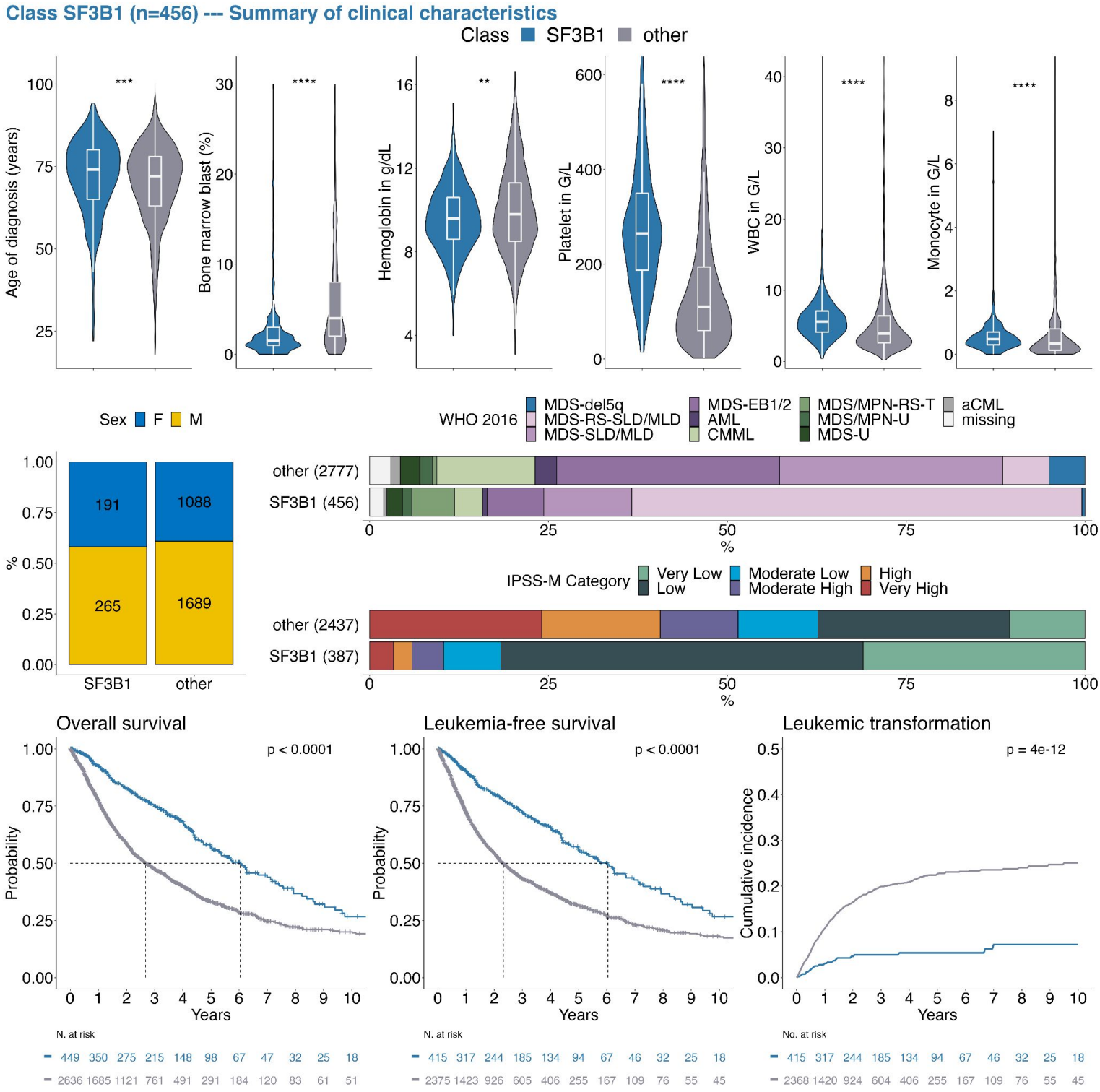

Class SF3B1 (n=456) --- Summary of clinical characteristics
Class SF3B1 other
Age of diagnosis (years)
Bone marrow blast (%)
Hemoglobin in g/dL
Platelet in G/L
WBC in G/L
Monocyte in G/L
***
****
**
****
****
****
Sex F M
191
1088
265
1689
SF3B1
other
%
WHO 2016
MDS-del5q
MDS-RS-SLD/MLD
MDS-SLD/MLD
MDS-EB1/2
AML
CMML
MDS/MPN-RS-T
MDS/MPN-U
MDS-U
aCML
missing
other (2777)
SF3B1 (456)
IPSS-M Category
Very Low
Low
Moderate Low
Moderate High
High
Very High
other (2437)
SF3B1 (387)
Overall survival
p < 0.0001
Probability
Years
N. at risk
449 350 275 215 148 98 67 47 32 25 18
2636 1685 1121 761 491 291 184 120 83 61 51
Leukemia-free survival
p < 0.0001
Probability
Years
N. at risk
415 317 244 185 134 94 67 46 32 25 18
2375 1423 926 605 406 255 167 109 76 55 45
Leukemic transformation
p = 4e-12
Cumulative incidence
Years
No. at risk
415 317 244 185 134 94 67 46 32 25 18
2368 1420 924 604 406 255 167 109 76 55 45

Figure S18 | Comparison of cases with complex karyotype with or without mutated *TP53*.

A. Distribution of the number of cytogenetic alterations and number of mutated genes per patient for cases with complex karyotype (CK) and with wildtype (WT) *TP53* or mutated (MUT) *TP53*.

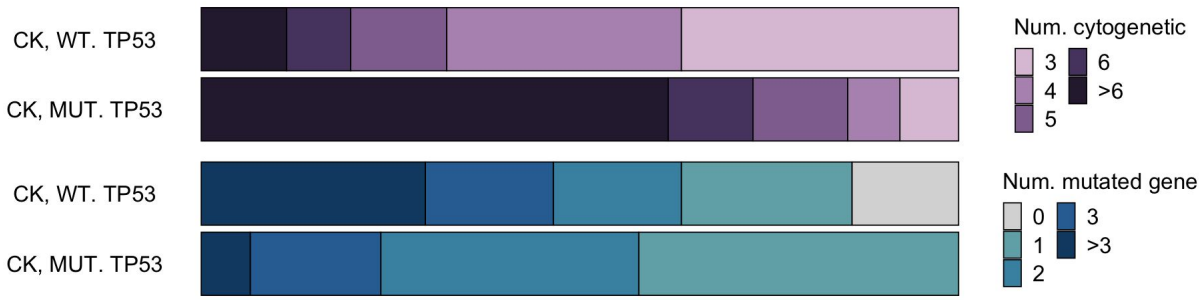
B. Frequency distribution of gene mutations and cytogenetic alterations in patients with complex karyotype and mutated *TP53* (blue, n=232) or complex karyotype and wildtype *TP53* (gold, n=71).
***p< 0.001, **p< 0.01, *p< 0.05, two-sided Fisher's exact test with Benjamini–Hochberg multiple testing correction.

C. Kaplan-Meier probability estimates of overall survival (left) and leukemia free survival (middle). P-values are from the log-rank test. Right panel: cumulative incidence curves for the rate of leukemic transformation. P-value is from the Gray's test.

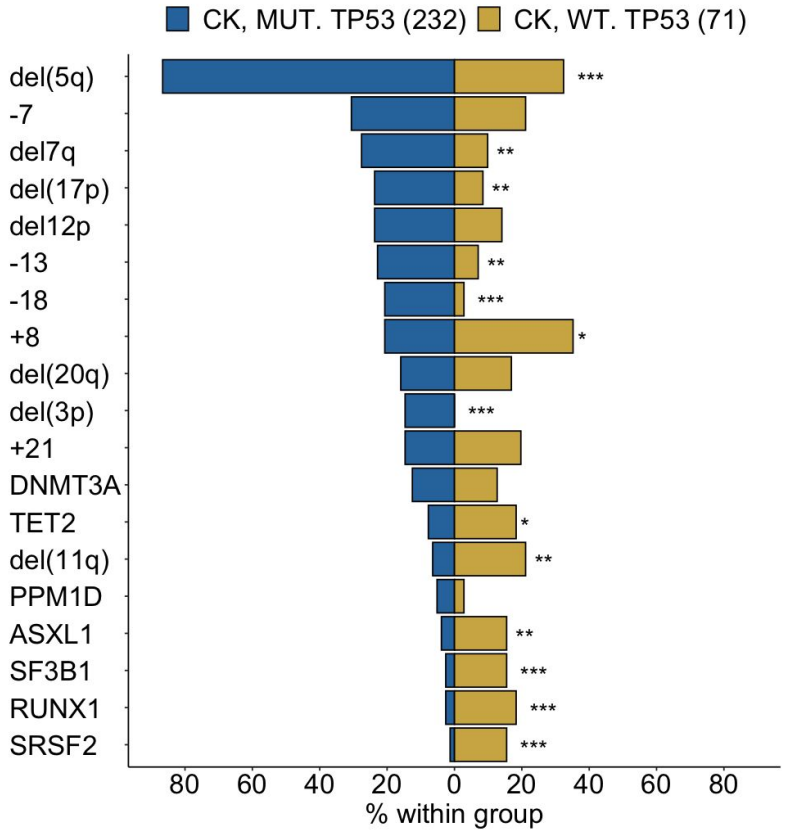
CK, MUT. TP53 n=232

CK, WT. TP53 n=71

A.



B.



C.

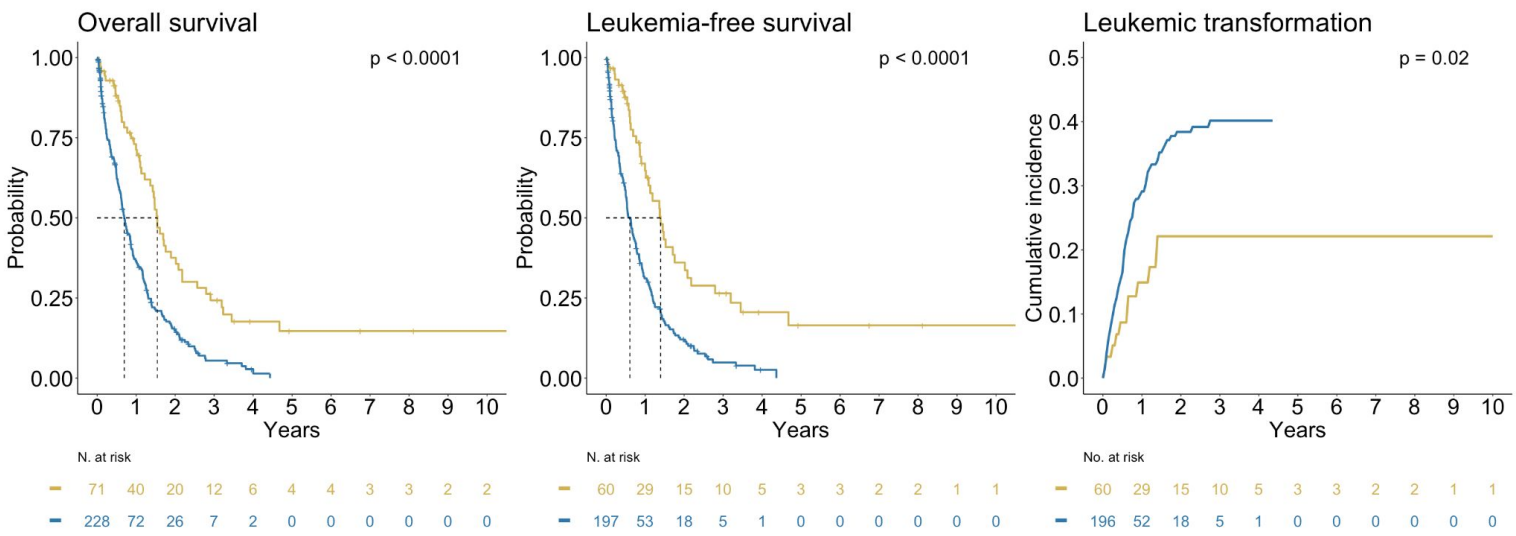
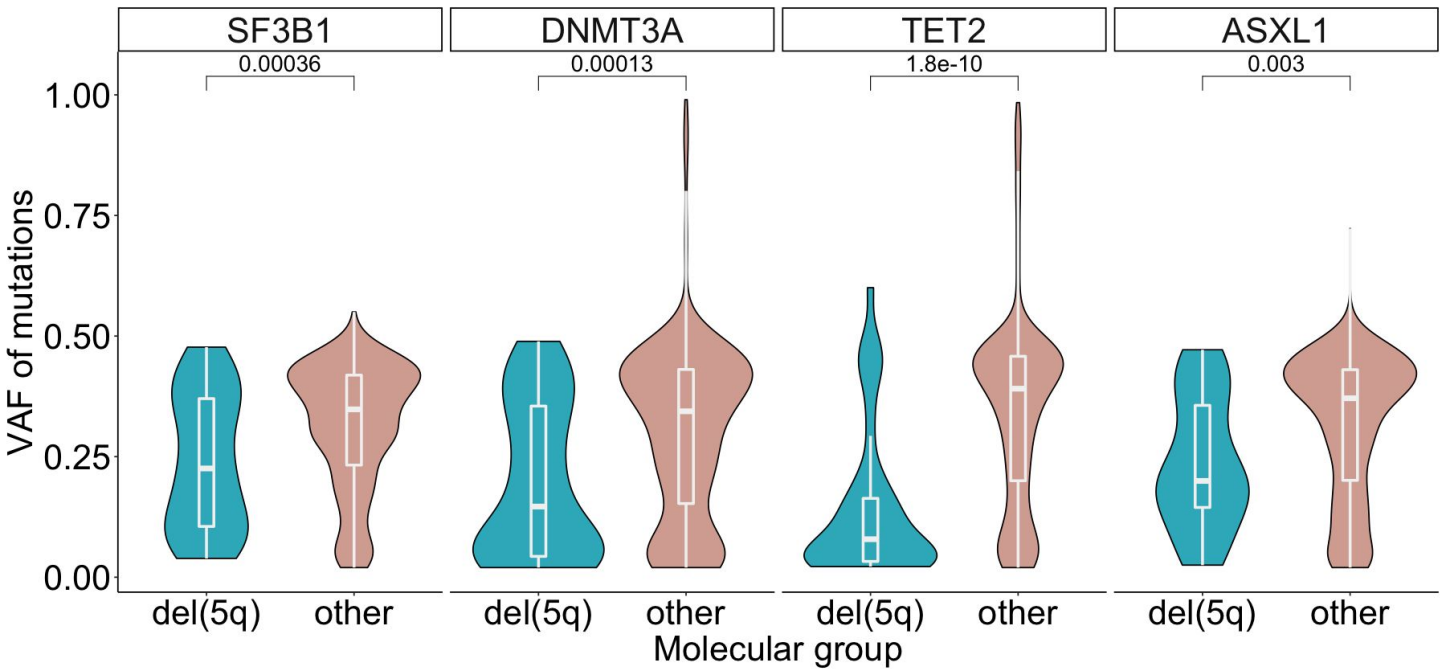


Figure S19 | Comparison of the variant allele fraction of *SF3B1* and *DTA* mutations between the *del(5q)* or other molecular groups.

Violin plots and boxplots represent the distribution of variant allele fraction (VAF) of *SF3B1* or *DTA* (*DNMT3A*, *TET2*, *ASXL1*) mutations from patients classified within the *del(5q)* (green) or other molecular subgroups (red). Number of patients with *SF3B1* or *DTA* mutations are indicated below the plot. P-values are from the Wilcoxon rank-sum test. The VAFs of *SF3B1* and *DTA* mutations were significantly lower within the *del(5q)* group compared to other molecular groups (22% vs. 35% and 15% vs. 38%, respectively).



N.	49	684	47	463	38	976	31	837

Figure S20 | Summary molecular and clinical metrics of the *bi-TET2* group. Two multi-panel figures, molecular (**A.**) and clinical (**B.**), are provided.

A. Molecular summary metrics comparing cases categorized within the group (blue) versus other groups (grey).

Top row. Boxplot distribution of the number of mutated genes per patient (left) and representation of the most frequent molecular alterations within the group (right).

Middle row. Boxplot and violin plots representing the distribution of the maximum values of variant allele fraction (VAF) (left) and of VAF adjusted for ploidy (middle left) per patient. Distribution of the VAF of all mutations identified in patients without (middle right) or with (right) adjustment for ploidy.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Bottom row. Density distribution of the VAF adjusted for ploidy of mutations from the most commonly mutated genes within the group (left). Relative order of gene mutation acquisition inferred through Bradley-Terry analysis. The numbers (n) indicate the number of informative pairwise precedences per gene (right).

B. Clinical summary metrics comparing cases categorized within the group (blue) versus all other groups (grey).

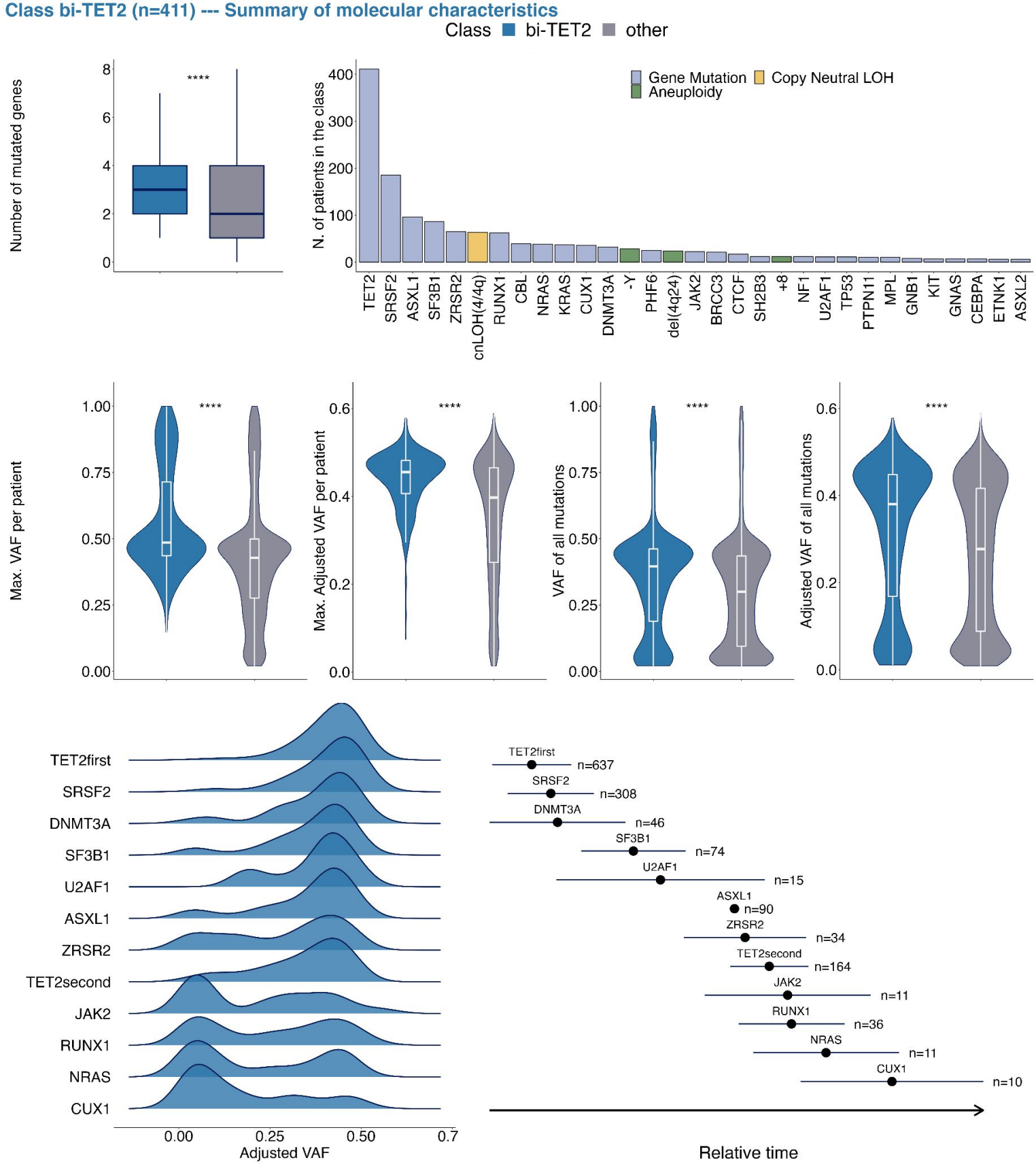
Top row. Distribution of the age at diagnosis, percentage of bone marrow blast, hemoglobin level, platelet count, white blood cell (WBC) count, and monocyte count.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Middle row. Left panel: Male (M) / Female (F) proportion. Right panels: Distribution of the WHO 2016 diagnosis subtypes (top) and of IPSS-M risk categories (bottom).

Bottom row. Left and middle panels: Kaplan-Meier probability estimates of overall survival (left) and leukemia free survival (middle). P-values are from the log-rank test. Right panel: cumulative incidence curves for the rate of leukemic transformation. P-value is from the Gray's test.

A.


Class bi-TET2 (n=411) --- Summary of molecular characteristics
Class bi-TET2 other
Number of mutated genes
****
N. of patients in the class
Gene Mutation
Copy Neutral LOH
Aneuploidy
TET2
SRSF2
ASXL1
SF3B1
ZRSR2
cnLOH(4/4q)
RUNX1
CBL
NRAS
KRAS
CUX1
DNMT3A
-Y
PHF6
del(4q24)
JAK2
BRCC3
CTCF
SH2B3
+8
NF1
U2AF1
TP53
PTPN11
MPL
GNB1
KIT
GNAS
CEBPA
ETNK1
ASXL2
Max. VAF per patient
Max. Adjusted VAF per patient
VAF of all mutations
Adjusted VAF of all mutations
TET2first
SRSF2
DNMT3A
SF3B1
U2AF1
ASXL1
ZRSR2
TET2second
JAK2
RUNX1
NRAS
CUX1
Adjusted VAF
n=637
n=308
n=46
n=74
n=15
n=90
n=34
n=164
n=11
n=36
n=11
n=10
Relative time

B.

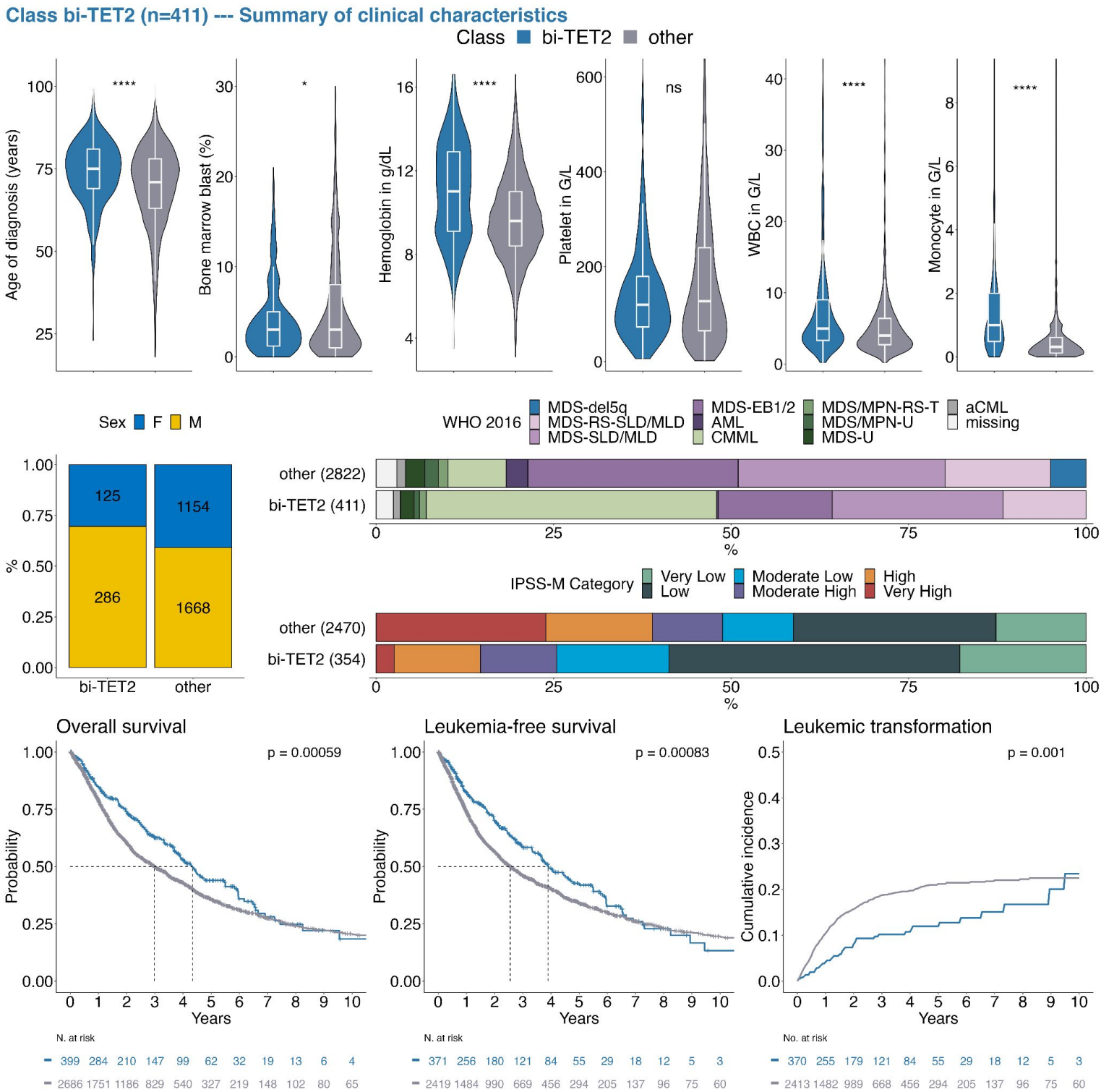

Class bi-TET2 (n=411) --- Summary of clinical characteristics
Class bi-TET2 other
Age of diagnosis (years)
Bone marrow blast (%)
Hemoglobin in g/dL
Platelet in G/L
WBC in G/L
Monocyte in G/L
****
*
ns
Sex F M
125
1154
286
1668
bi-TET2
other
%
WHO 2016
MDS-del5q
MDS-RS-SLD/MLD
MDS-SLD/MLD
MDS-EB1/2
AML
CMML
MDS/MPN-RS-T
MDS/MPN-U
MDS-U
aCML
missing
other (2822)
bi-TET2 (411)
IPSS-M Category
Very Low
Low
Moderate Low
Moderate High
High
Very High
other (2470)
bi-TET2 (354)
Overall survival
p = 0.00059
Leukemia-free survival
p = 0.00083
Leukemic transformation
p = 0.001
Probability
Cumulative incidence
Years
N. at risk
399 284 210 147 99 62 32 19 13 6 4
2686 1751 1186 829 540 327 219 148 102 80 65
371 256 180 121 84 55 29 18 12 5 3
2419 1484 990 669 456 294 205 137 96 75 60
No. at risk
370 255 179 121 84 55 29 18 12 5 3
2413 1482 989 668 456 294 205 137 96 75 60

Figure S21 | Genotype-phenotype association analysis and evolutionary trajectories within the *bi-TET2* group.

A. Scatter plots comparison within the *bi-TET2* group of the cancer cell fraction (CCF) of mutations of the dominant *TET2* mutation with (i) the second *TET2* mutation for cases with multiple *TET2* mutations (top left), (ii) *SRSF2* mutation (top right), (iii) *SF3B1* mutation (bottom left), and (iv) *ZRSR2* mutation (bottom right). Of note, the median CCF of the dominant *TET2* mutation was 87% (IQR 76-94), while for the 306 cases with multiple *TET2* mutations the second allele CCF was 76% (IQR 54-88).

B. Boxplot and violin plots comparing the distribution of monocyte counts of patients categorized as *bi-TET2* across the different mutational status for splicing factor mutations (*SRSF2*, *SF3B1*, *ZRSR2*, or none)**.**
****p<0.0001, *p<0.01, *p< 0.05, Wilcoxon rank-sum test.

C. Multiple multivariable linear least square regression analysis using the presence of gene mutations as explanatory features and phenotype variables as target variables. White blood cell and monocyte counts were log-transformed. The heatmap represents the fitted coefficients with darker purple denoting positive effects and darker red negative effects. The coefficients of determination (R^2) are indicated as dot-lines.
**p<0.0001, *p<0.01, +p< 0.01, Z-score test.

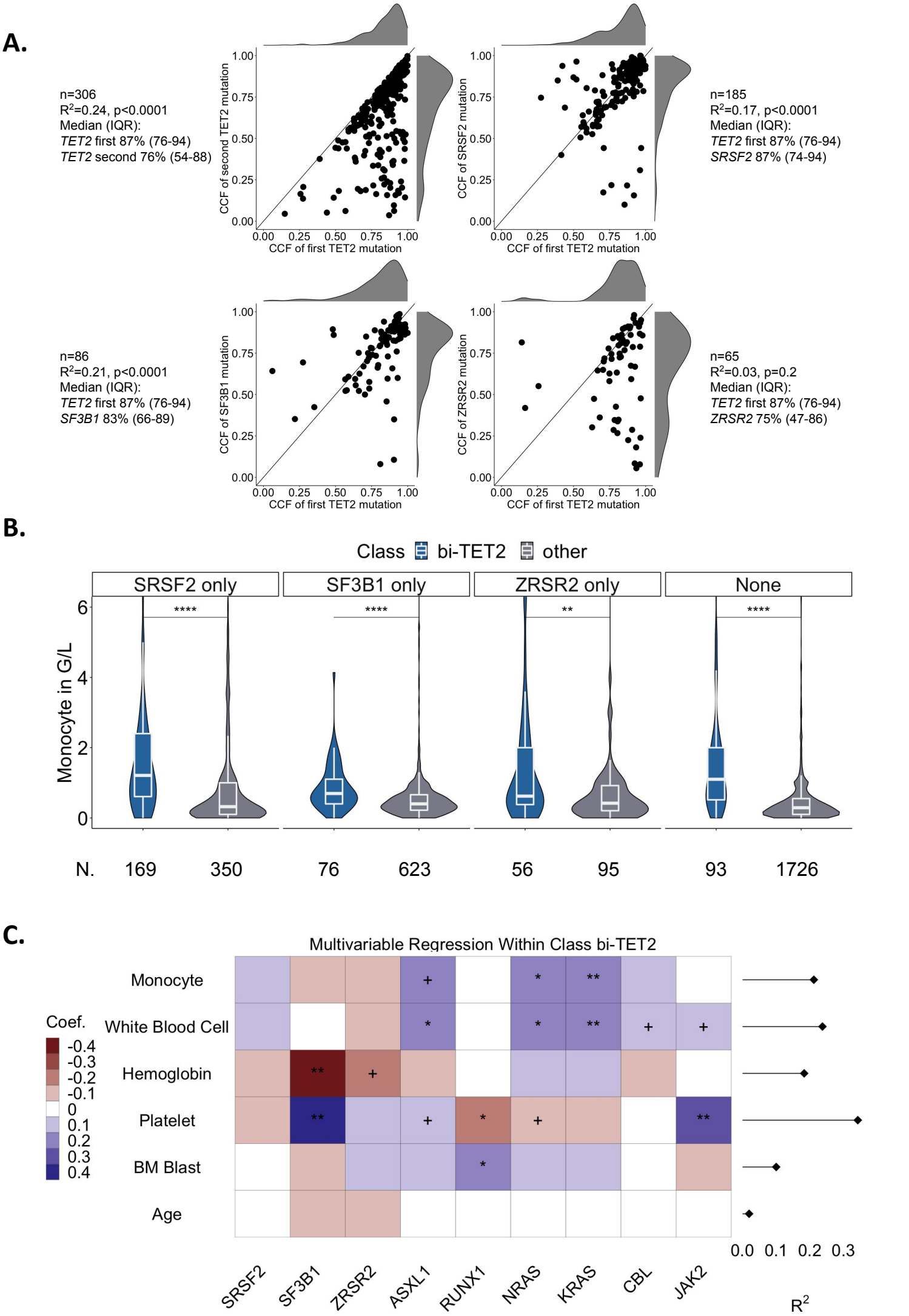

A.
n=306
R2=0.24, p<0.0001
Median (IQR):
TET2 first 87% (76-94)
TET2 second 76% (54-88)
CCF of second TET2 mutation
CCF of first TET2 mutation
n=185
R2=0.17, p<0.0001
Median (IQR):
TET2 first 87% (76-94)
SRSF2 87% (74-94)
CCF of SRSF2 mutation
n=86
R2=0.21, p<0.0001
Median (IQR):
TET2 first 87% (76-94)
SF3B1 83% (66-89)
CCF of SF3B1 mutation
n=65
R2=0.03, p=0.2
Median (IQR):
TET2 first 87% (76-94)
ZRSR2 75% (47-86)
CCF of ZRSR2 mutation
B.
Class bi-TET2 other
SRSF2 only
SF3B1 only
ZRSR2 only
None
Monocyte in G/L
N. 169 350 76 623 56 95 93 1726
C.
Multivariable Regression Within Class bi-TET2
Coef.
Monocyte
White Blood Cell
Hemoglobin
Platelet
BM Blast
Age
SRSF2
SF3B1
ZRSR2
ASXL1
RUNX1
NRAS
KRAS
CBL
JAK2
R2

Figure S22 | Summary molecular and clinical metrics of the *CCUS-like* group. Two multi-panel figures, molecular (**A.**) and clinical (**B.**), are provided.

A. Molecular summary metrics comparing cases categorized within the group (blue) versus other groups (grey).

Top row. Boxplot distribution of the number of mutated genes per patient (left) and representation of the most frequent molecular alterations within the group (right).

Middle row. Boxplot and violin plots representing the distribution of the maximum values of variant allele fraction (VAF) (left) and of VAF adjusted for ploidy (middle left) per patient. Distribution of the VAF of all mutations identified in patients without (middle right) or with (right) adjustment for ploidy.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Bottom row. Density distribution of the VAF adjusted for ploidy of mutations from the most commonly mutated genes within the group (left).

B. Clinical summary metrics comparing cases categorized within the group (blue) versus all other groups (grey).

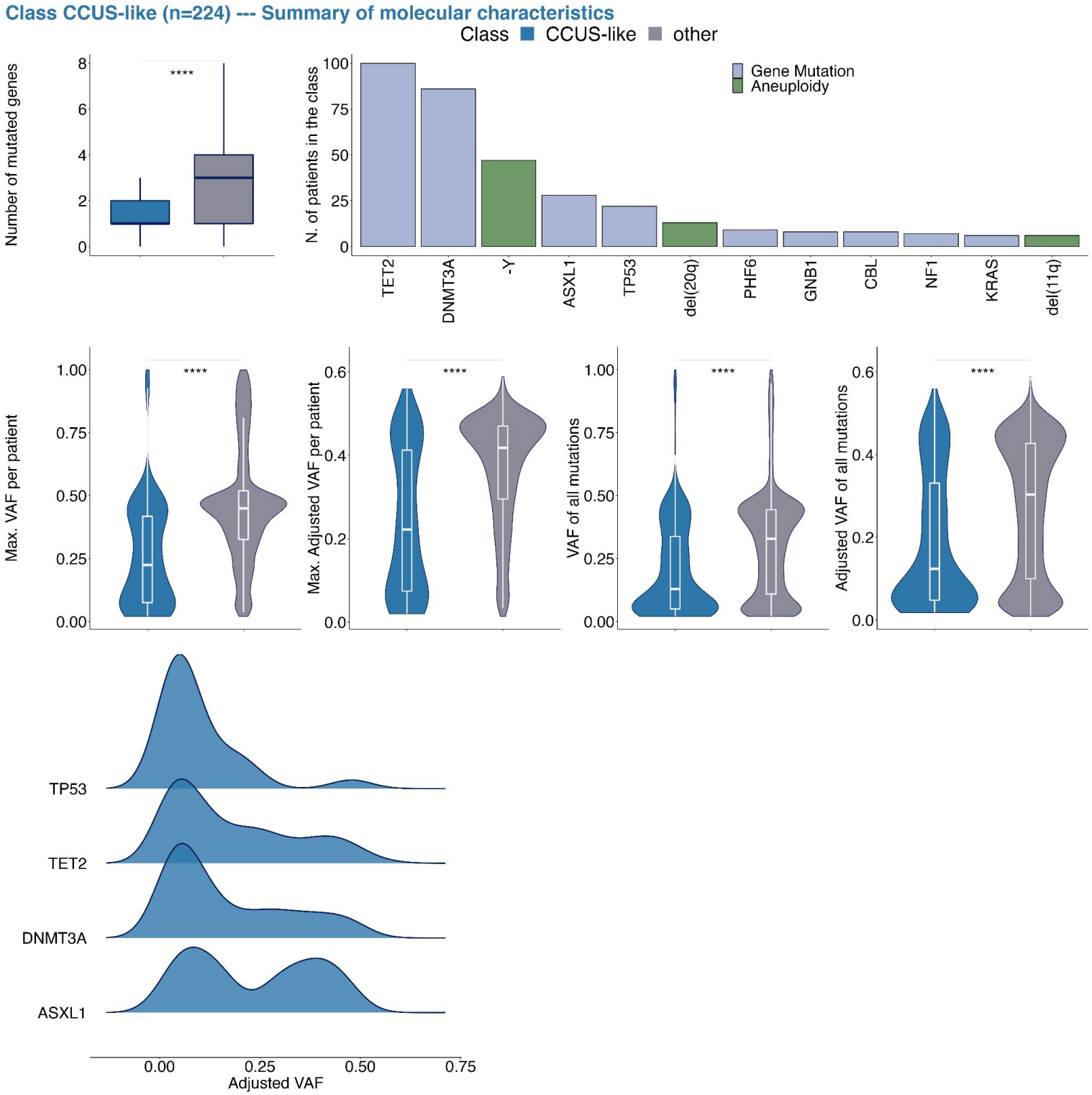
Top row. Distribution of the age at diagnosis, percentage of bone marrow blast, hemoglobin level, platelet count, white blood cell (WBC) count, and monocyte count.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Middle row. Left panel: Male (M) / Female (F) proportion. Right panels: Distribution of the WHO 2016 diagnosis subtypes (top) and of IPSS-M risk categories (bottom).

Bottom row. Left and middle panels: Kaplan-Meier probability estimates of overall survival (left) and leukemia free survival (middle). P-values are from the log-rank test. Right panel: cumulative incidence curves for the rate of leukemic transformation. P-value is from the Gray's test.

A.


Class CCUS-like (n=224) --- Summary of molecular characteristics
Class CCUS-like other
Number of mutated genes
****
N. of patients in the class
Gene Mutation
Aneuploidy
TET2
DNMT3A
-Y
ASXL1
TP53
del(20q)
PHF6
GNB1
CBL
NF1
KRAS
del(11q)
Max. VAF per patient
Max. Adjusted VAF per patient
VAF of all mutations
Adjusted VAF of all mutations
TP53
TET2
DNMT3A
ASXL1
Adjusted VAF

B.

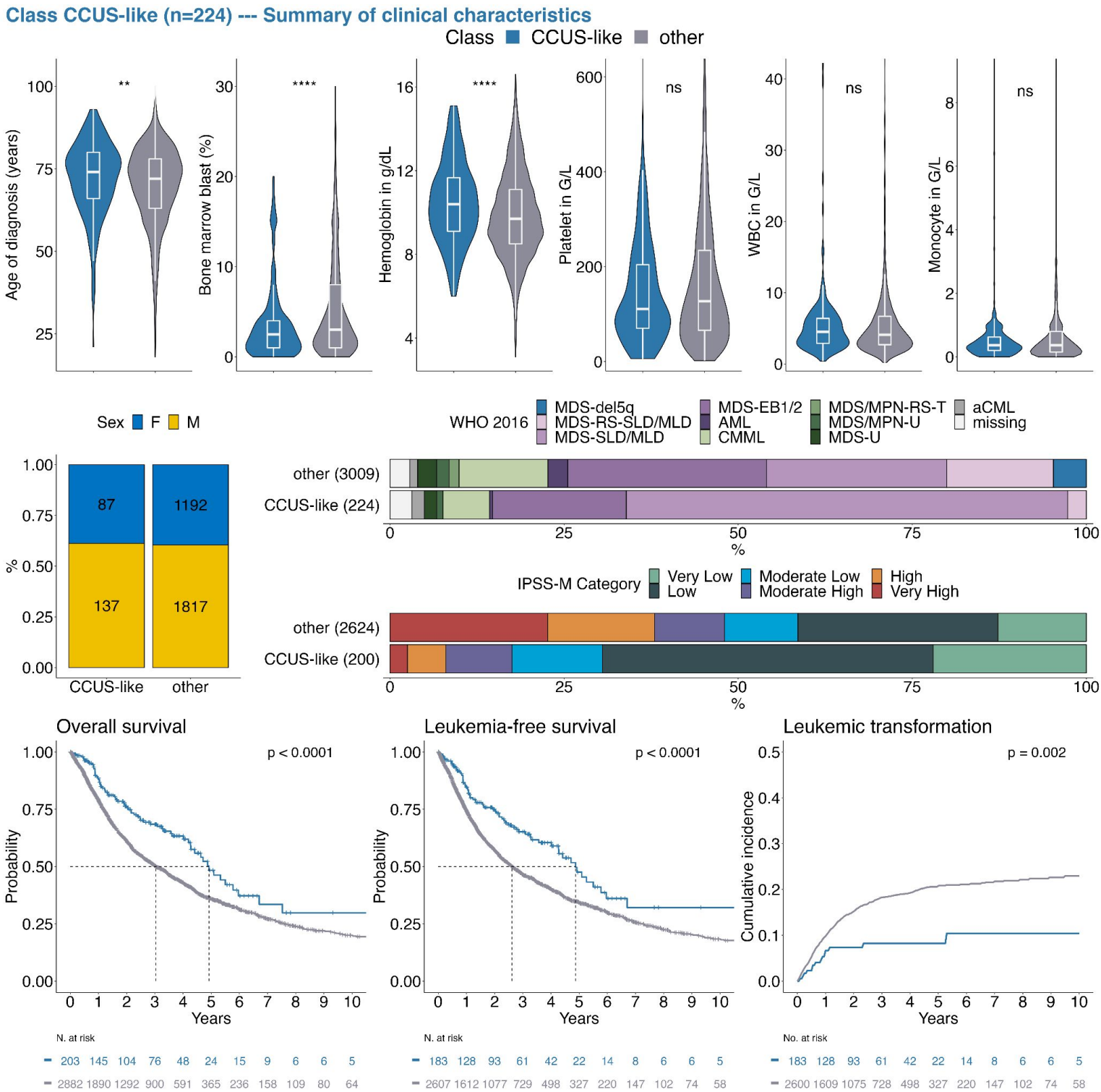

Class CCUS-like (n=224) --- Summary of clinical characteristics
Class CCUS-like other
Age of diagnosis (years)
Bone marrow blast (%)
Hemoglobin in g/dL
Platelet in G/L
WBC in G/L
Monocyte in G/L
**
****
****
ns
ns
ns
Sex F M
87
1192
137
1817
CCUS-like
other
%
WHO 2016
MDS-del5q
MDS-RS-SLD/MLD
MDS-SLD/MLD
MDS-EB1/2
AML
CMML
MDS/MPN-RS-T
MDS/MPN-U
MDS-U
aCML
missing
other (3009)
CCUS-like (224)
IPSS-M Category
Very Low
Low
Moderate Low
Moderate High
High
Very High
other (2624)
CCUS-like (200)
Overall survival
p < 0.0001
Leukemia-free survival
p < 0.0001
Leukemic transformation
p = 0.002
Probability
Cumulative incidence
Years
N. at risk
203 145 104 76 48 24 15 9 6 6 5
2882 1890 1292 900 591 365 236 158 109 80 64
183 128 93 61 42 22 14 8 6 6 5
2607 1612 1077 729 498 327 220 147 102 74 58
No. at risk
183 128 93 61 42 22 14 8 6 6 5
2600 1609 1075 728 498 327 220 147 102 74 58

Figure S23 | Summary molecular and clinical metrics of the *-7/SETBP1* group. Two multi-panel figures, molecular (**A.**) and clinical (**B.**), are provided.

A. Molecular summary metrics comparing cases categorized within the group (blue) versus other groups (grey).

Top row. Boxplot distribution of the number of mutated genes per patient (left) and representation of the most frequent molecular alterations within the group (right).

Middle row. Boxplot and violin plots representing the distribution of the maximum values of variant allele fraction (VAF) (left) and of VAF adjusted for ploidy (middle left) per patient. Distribution of the VAF of all mutations identified in patients without (middle right) or with (right) adjustment for ploidy.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Bottom row. Density distribution of the VAF adjusted for ploidy of mutations from the most commonly mutated genes within the group (left). Relative order of gene mutation acquisition inferred through Bradley-Terry analysis. The numbers (n) indicate the number of informative pairwise precedences per gene (right).

B. Clinical summary metrics comparing cases categorized within the group (blue) versus all other groups (grey).

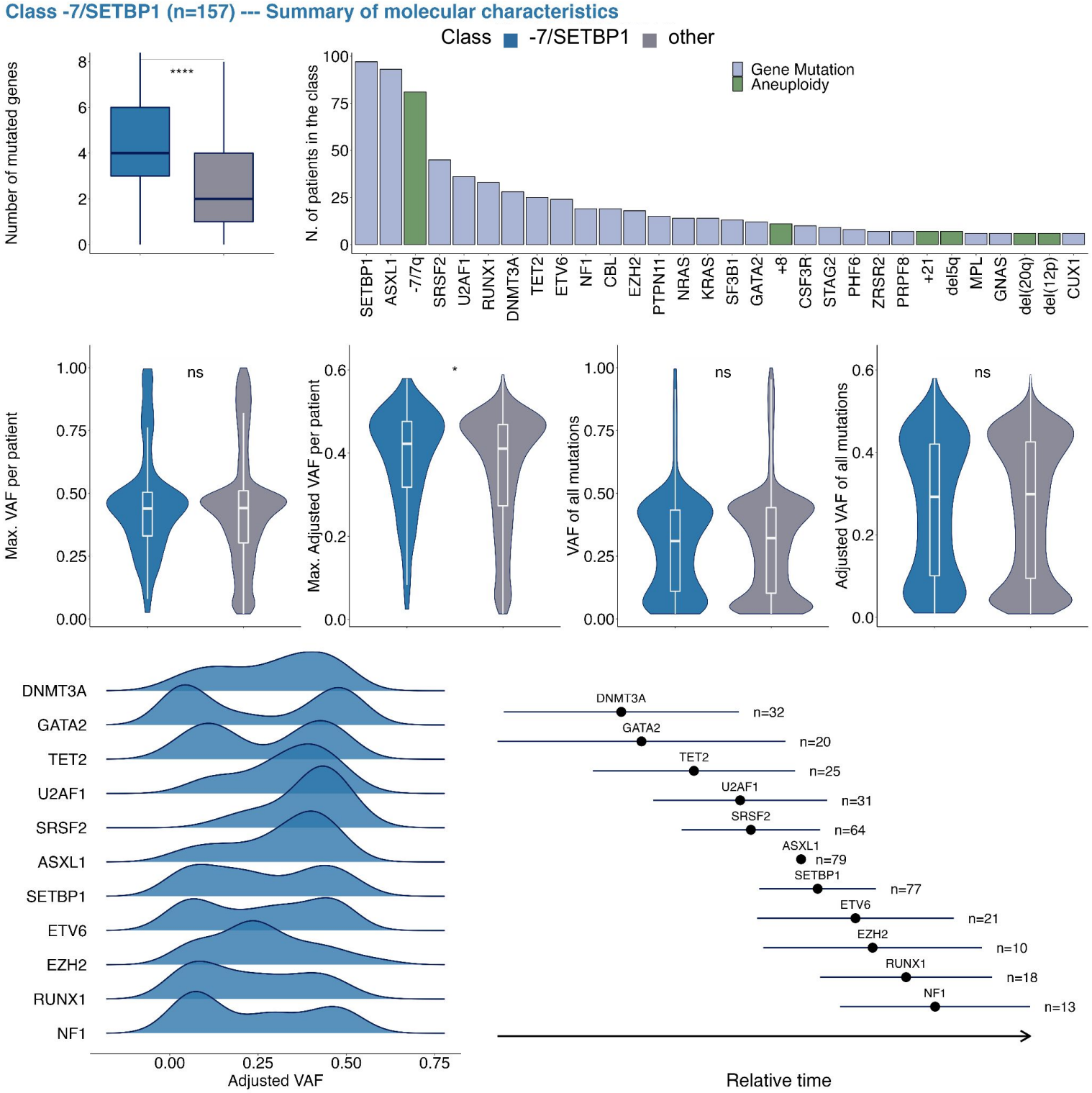
Top row. Distribution of the age at diagnosis, percentage of bone marrow blast, hemoglobin level, platelet count, white blood cell (WBC) count, and monocyte count.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Middle row. Left panel: Male (M) / Female (F) proportion. Right panels: Distribution of the WHO 2016 diagnosis subtypes (top) and of IPSS-M risk categories (bottom).

Bottom row. Left and middle panels: Kaplan-Meier probability estimates of overall survival (left) and leukemia free survival (middle). P-values are from the log-rank test. Right panel: cumulative incidence curves for the rate of leukemic transformation. P-value is from the Gray's test.

A.


Class -7/SETBP1 (n=157) --- Summary of molecular characteristics
Class -7/SETBP1 other
Number of mutated genes
****
N. of patients in the class
Gene Mutation
Aneuploidy
SETBP1
ASXL1
-7/7q
SRSF2
U2AF1
RUNX1
DNMT3A
TET2
ETV6
NF1
CBL
EZH2
PTPN11
NRAS
KRAS
SF3B1
GATA2
+8
CSF3R
STAG2
PHF6
ZRSR2
PRPF8
+21
del5q
MPL
GNAS
del(20q)
del(12p)
CUX1
Max. VAF per patient
ns
Max. Adjusted VAF per patient
*
VAF of all mutations
ns
Adjusted VAF of all mutations
ns
DNMT3A
GATA2
TET2
U2AF1
SRSF2
ASXL1
SETBP1
ETV6
EZH2
RUNX1
NF1
Adjusted VAF
n=32
n=20
n=25
n=31
n=64
n=79
n=77
n=21
n=10
n=18
n=13
Relative time

B.

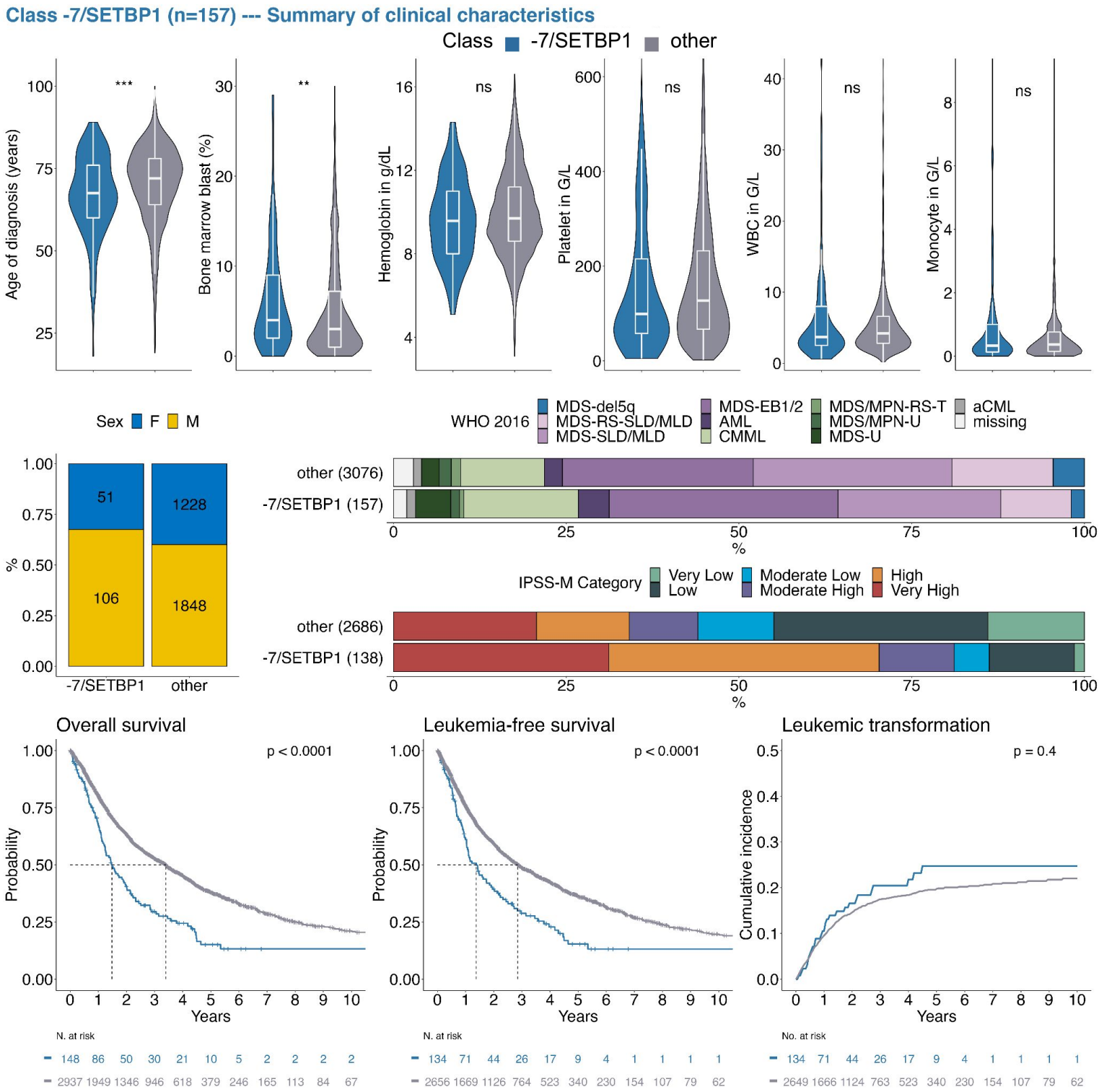

Class -7/SETBP1 (n=157) --- Summary of clinical characteristics
Class -7/SETBP1 other
Age of diagnosis (years)
Bone marrow blast (%)
Hemoglobin in g/dL
Platelet in G/L
WBC in G/L
Monocyte in G/L
***
**
ns
Sex F M
51
1228
106
1848
-7/SETBP1
other
WHO 2016
MDS-del5q
MDS-RS-SLD/MLD
MDS-SLD/MLD
MDS-EB1/2
AML
CMML
MDS/MPN-RS-T
MDS/MPN-U
MDS-U
aCML
missing
other (3076)
-7/SETBP1 (157)
IPSS-M Category
Very Low
Low
Moderate Low
Moderate High
High
Very High
other (2686)
-7/SETBP1 (138)
Overall survival
p < 0.0001
Probability
Years
N. at risk
148 86 50 30 21 10 5 2 2 2 2
2937 1949 1346 946 618 379 246 165 113 84 67
Leukemia-free survival
p < 0.0001
134 71 44 26 17 9 4 1 1 1 1
2656 1669 1126 764 523 340 230 154 107 79 62
Leukemic transformation
p = 0.4
Cumulative incidence
No. at risk
134 71 44 26 17 9 4 1 1 1 1
2649 1666 1124 763 523 340 230 154 107 79 62

Figure S24 | Summary molecular and clinical metrics of the *EZH2-ASXL1* group. Two multi-panel figures, molecular (**A.**) and clinical (**B.**), are provided.

A. Molecular summary metrics comparing cases categorized within the group (blue) versus other groups (grey).

Top row. Boxplot distribution of the number of mutated genes per patient (left) and representation of the most frequent molecular alterations within the group (right).

Middle row. Boxplot and violin plots representing the distribution of the maximum values of variant allele fraction (VAF) (left) and of VAF adjusted for ploidy (middle left) per patient. Distribution of the VAF of all mutations identified in patients without (middle right) or with (right) adjustment for ploidy.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Bottom row. Density distribution of the VAF adjusted for ploidy of mutations from the most commonly mutated genes within the group (left). Relative order of gene mutation acquisition inferred through Bradley-Terry analysis. The numbers (n) indicate the number of informative pairwise precedences per gene (right).

B. Clinical summary metrics comparing cases categorized within the group (blue) versus all other groups (grey).

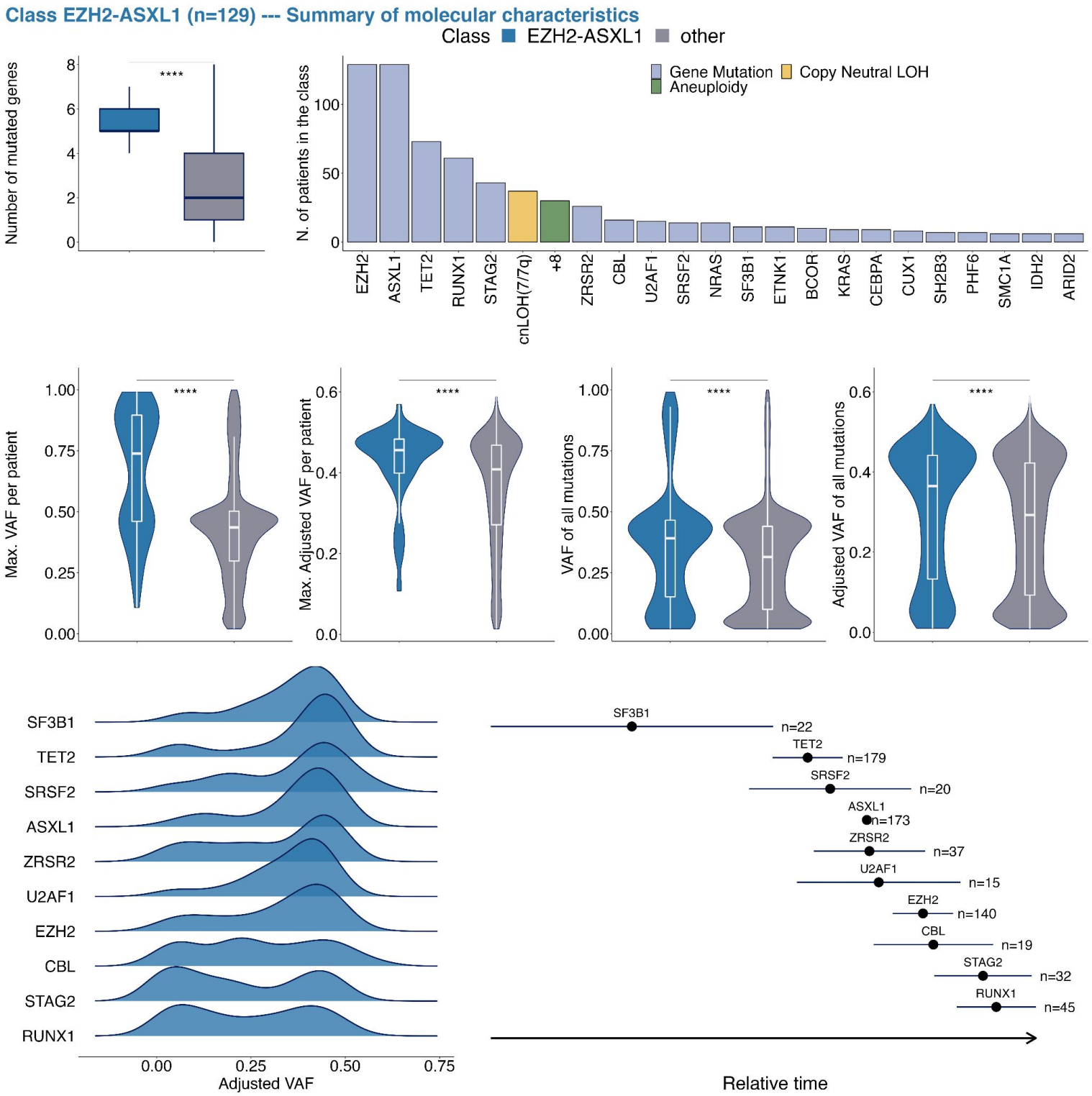
Top row. Distribution of the age at diagnosis, percentage of bone marrow blast, hemoglobin level, platelet count, white blood cell (WBC) count, and monocyte count.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Middle row. Left panel: Male (M) / Female (F) proportion. Right panels: Distribution of the WHO 2016 diagnosis subtypes (top) and of IPSS-M risk categories (bottom).

Bottom row. Left and middle panels: Kaplan-Meier probability estimates of overall survival (left) and leukemia free survival (middle). P-values are from the log-rank test. Right panel: cumulative incidence curves for the rate of leukemic transformation. P-value is from the Gray's test.

A.


Class EZH2-ASXL1 (n=129) --- Summary of molecular characteristics
Class EZH2-ASXL1 other
Number of mutated genes
****
N. of patients in the class
Gene Mutation
Copy Neutral LOH
Aneuploidy
EZH2
ASXL1
TET2
RUNX1
STAG2
cnLOH(7/7q)
+8
ZRSR2
CBL
U2AF1
SRSF2
NRAS
SF3B1
ETNK1
BCOR
KRAS
CEBPA
CUX1
SH2B3
PHF6
SMC1A
IDH2
ARID2
Max. VAF per patient
Max. Adjusted VAF per patient
VAF of all mutations
Adjusted VAF of all mutations
SF3B1
TET2
SRSF2
ASXL1
ZRSR2
U2AF1
EZH2
CBL
STAG2
RUNX1
Adjusted VAF
n=22
n=179
n=20
n=173
n=37
n=15
n=140
n=19
n=32
n=45
Relative time

B.

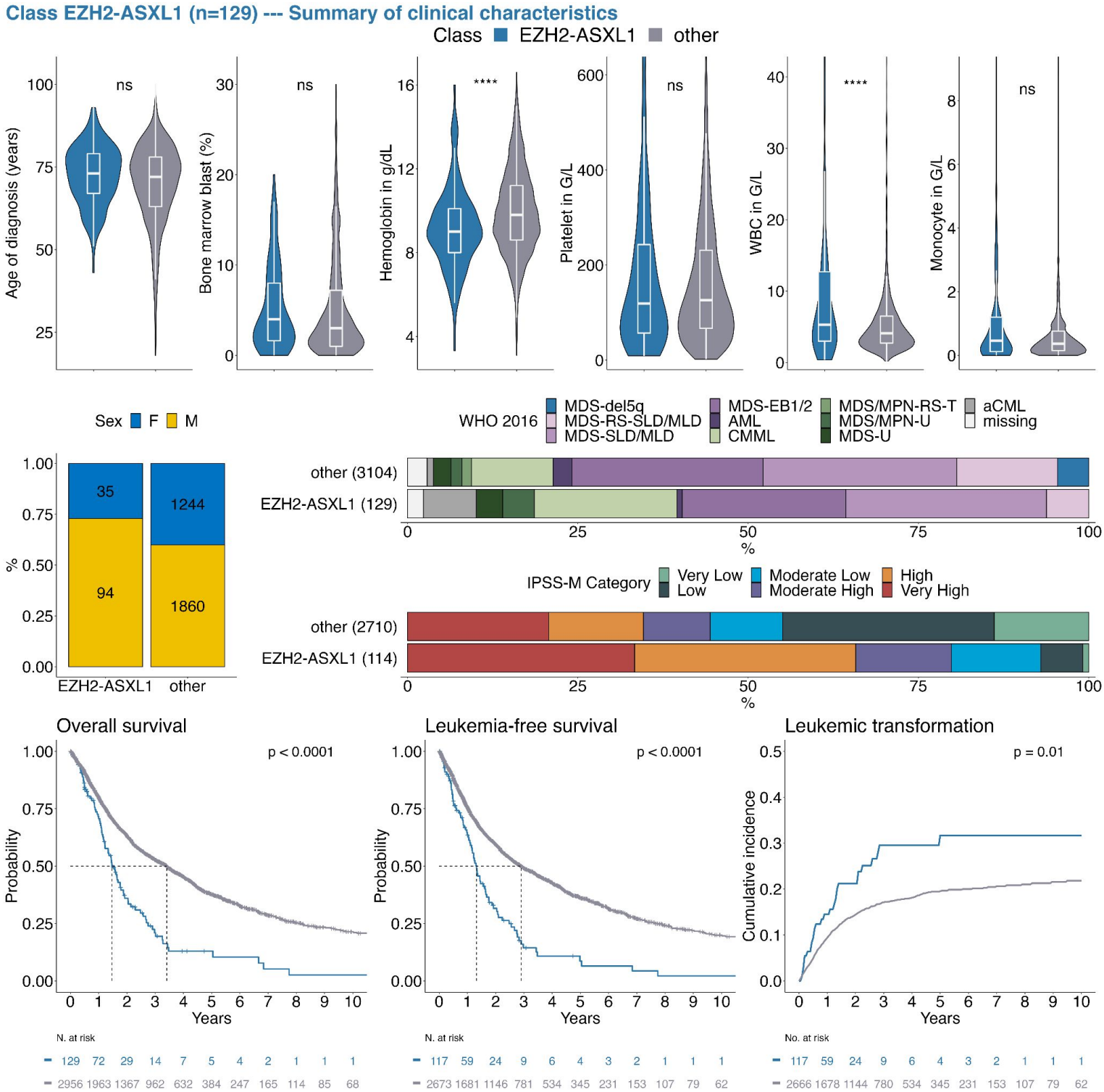

Class EZH2-ASXL1 (n=129) --- Summary of clinical characteristics
Class EZH2-ASXL1 other
Age of diagnosis (years)
Bone marrow blast (%)
Hemoglobin in g/dL
Platelet in G/L
WBC in G/L
Monocyte in G/L
ns
****
Sex F M
35
1244
94
1860
EZH2-ASXL1
other
WHO 2016
MDS-del5q
MDS-RS-SLD/MLD
MDS-SLD/MLD
MDS-EB1/2
AML
CMML
MDS/MPN-RS-T
MDS/MPN-U
MDS-U
aCML
missing
other (3104)
EZH2-ASXL1 (129)
IPSS-M Category
Very Low
Low
Moderate Low
Moderate High
High
Very High
other (2710)
EZH2-ASXL1 (114)
Overall survival
p < 0.0001
Leukemia-free survival
p < 0.0001
Leukemic transformation
p = 0.01
Probability
Cumulative incidence
Years
N. at risk
129 72 29 14 7 5 4 2 1 1 1
2956 1963 1367 962 632 384 247 165 114 85 68
117 59 24 9 6 4 3 2 1 1 1
2673 1681 1146 781 534 345 231 153 107 79 62
No. at risk
117 59 24 9 6 4 3 2 1 1 1
2666 1678 1144 780 534 345 231 153 107 79 62

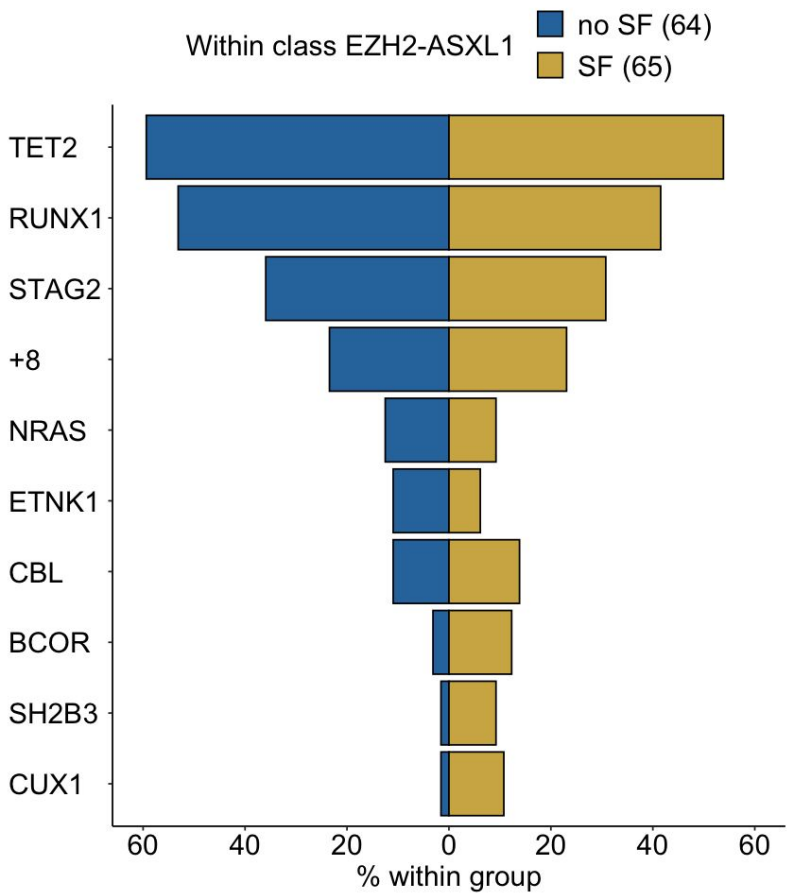
Figure S25 | Effect of splicing factor mutation status within the *EZH2-ASXL1* group.

A. Frequency distribution of gene mutations and cytogenetic alterations in patients categorized within the *EZH2-ASXL1* group and with (gold, n=65) or without (blue, n=64) splicing factor (SF) mutations. No significant differences in the proportion of molecular alterations were observed between the two subsets.

B. Kaplan-Meier probability estimates of overall survival from cases categorized within the *EZH2-ASXL1* group and with or without splicing factor (SF) mutations. P-value is from the log-rank test.

Figure S25 | Effect of splicing factor mutation status within the *EZH2-ASXL1* group.

A.



B.

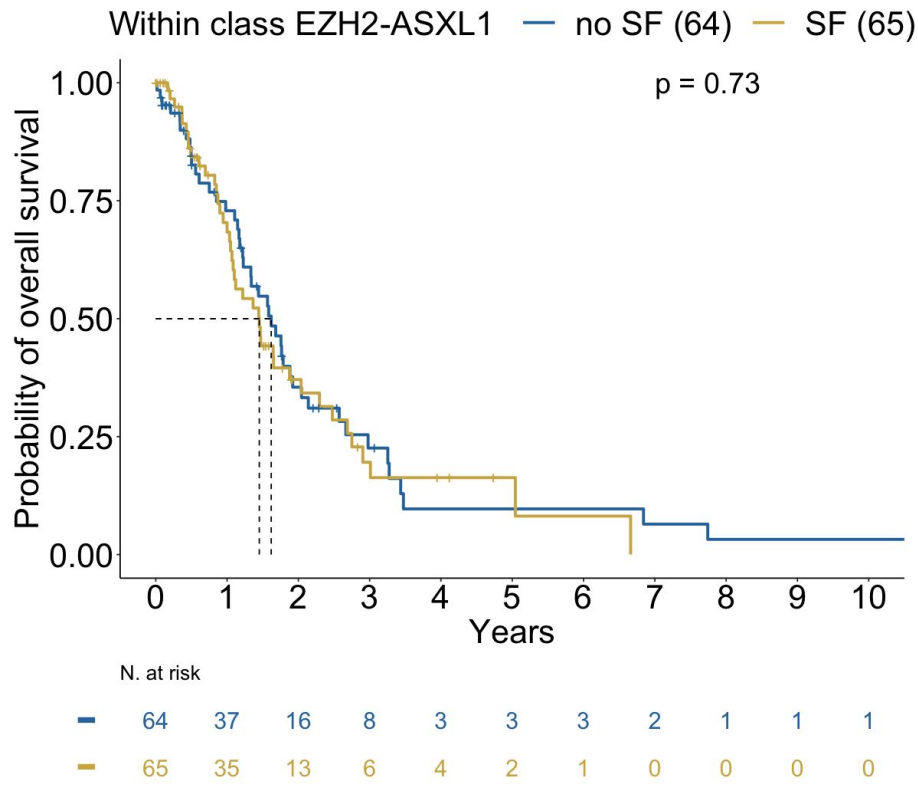


Figure S26 | Summary molecular and clinical metrics of the *IDH-STAG2* group. Two multi-panel figures, molecular (**A.**) and clinical (**B.**), are provided.

A. Molecular summary metrics comparing cases categorized within the group (blue) versus other groups (grey).

Top row. Boxplot distribution of the number of mutated genes per patient (left) and representation of the most frequent molecular alterations within the group (right).

Middle row. Boxplot and violin plots representing the distribution of the maximum values of variant allele fraction (VAF) (left) and of VAF adjusted for ploidy (middle left) per patient. Distribution of the VAF of all mutations identified in patients without (middle right) or with (right) adjustment for ploidy.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Bottom row. Density distribution of the VAF adjusted for ploidy of mutations from the most commonly mutated genes within the group (left). Relative order of gene mutation acquisition inferred through Bradley-Terry analysis. The numbers (n) indicate the number of informative pairwise precedences per gene (right).

B. Clinical summary metrics comparing cases categorized within the group (blue) versus all other groups (grey).

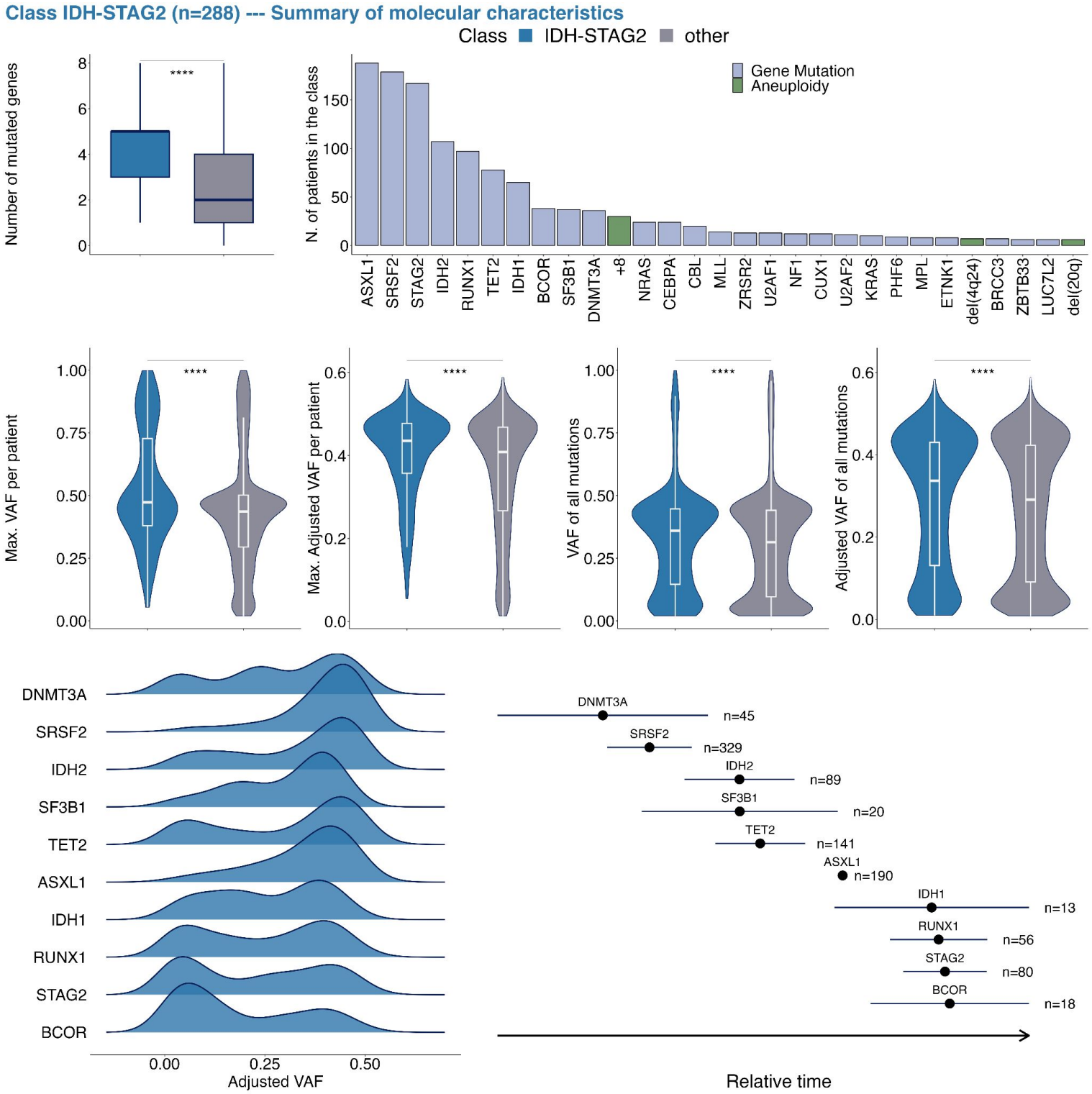
Top row. Distribution of the age at diagnosis, percentage of bone marrow blast, hemoglobin level, platelet count, white blood cell (WBC) count, and monocyte count.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Middle row. Left panel: Male (M) / Female (F) proportion. Right panels: Distribution of the WHO 2016 diagnosis subtypes (top) and of IPSS-M risk categories (bottom).

Bottom row. Left and middle panels: Kaplan-Meier probability estimates of overall survival (left) and leukemia free survival (middle). P-values are from the log-rank test. Right panel: cumulative incidence curves for the rate of leukemic transformation. P-value is from the Gray's test.

A.


Class IDH-STAG2 (n=288) --- Summary of molecular characteristics
Class IDH-STAG2 other
Number of mutated genes
****
N. of patients in the class
Gene Mutation
Aneuploidy
ASXL1 SRSF2 STAG2 IDH2 RUNX1 TET2 IDH1 BCOR SF3B1 DNMT3A +8 NRAS CEBPA CBL MLL ZRSR2 U2AF1 NF1 CUX1 U2AF2 KRAS PHF6 MPL ETNK1 del(4q24) BRCC3 ZBTB33 LUC7L2 del(20q)
Max. VAF per patient
Max. Adjusted VAF per patient
VAF of all mutations
Adjusted VAF of all mutations
DNMT3A
SRSF2
IDH2
SF3B1
TET2
ASXL1
IDH1
RUNX1
STAG2
BCOR
Adjusted VAF
n=45
n=329
n=89
n=20
n=141
n=190
n=13
n=56
n=80
n=18
Relative time

B.

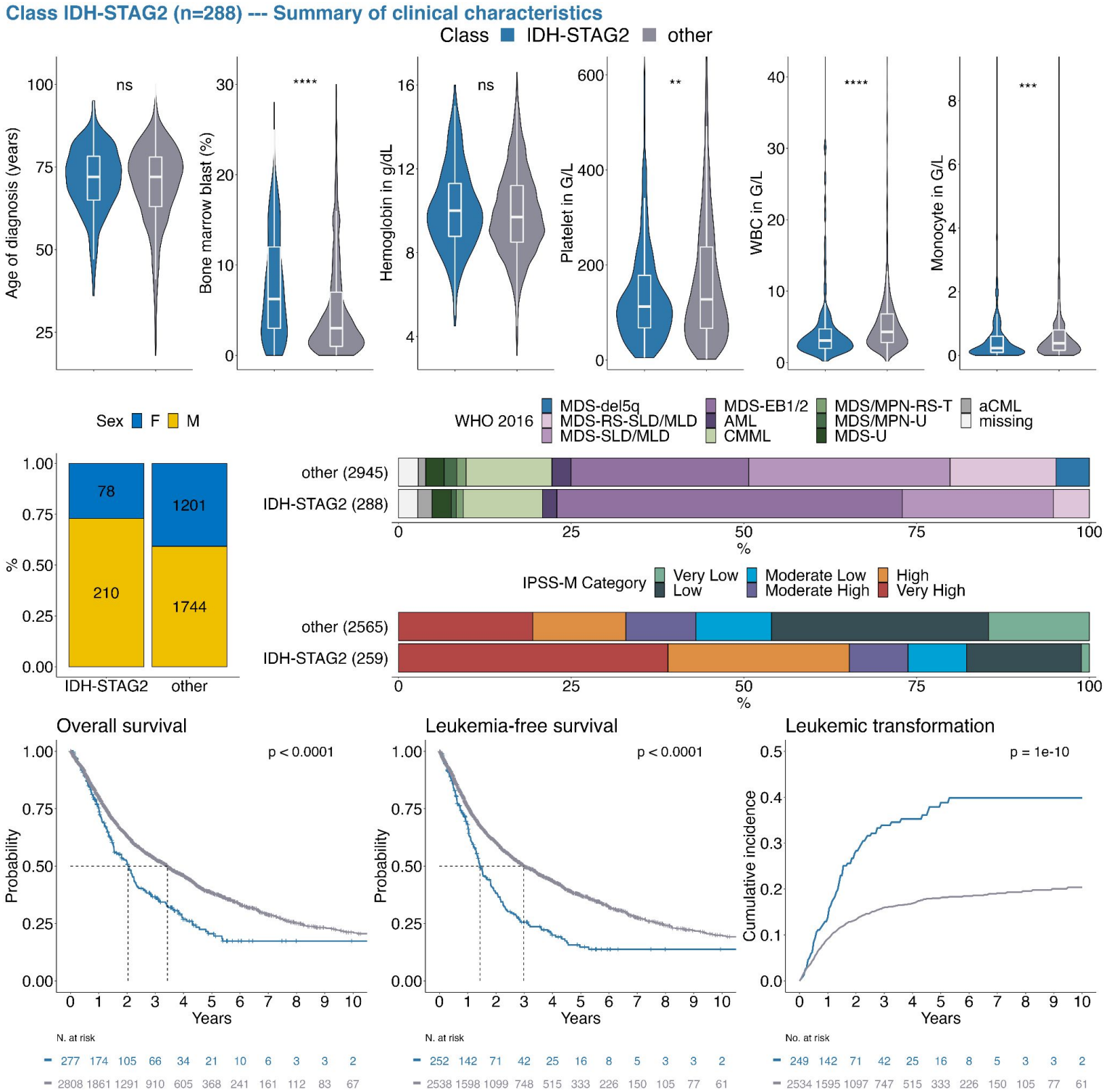

Class IDH-STAG2 (n=288) --- Summary of clinical characteristics
Class IDH-STAG2 other
Age of diagnosis (years)
Bone marrow blast (%)
Hemoglobin in g/dL
Platelet in G/L
WBC in G/L
Monocyte in G/L
ns
****
ns
**
****
***
Sex F M
78
1201
210
1744
IDH-STAG2
other
% 1.00 0.75 0.50 0.25 0.00
WHO 2016
MDS-del5q
MDS-RS-SLD/MLD
MDS-SLD/MLD
MDS-EB1/2
AML
CMML
MDS/MPN-RS-T
MDS/MPN-U
MDS-U
aCML
missing
other (2945)
IDH-STAG2 (288)
0 25 50 75 100
%
IPSS-M Category
Very Low
Low
Moderate Low
Moderate High
High
Very High
other (2565)
IDH-STAG2 (259)
0 25 50 75 100
%
Overall survival
p < 0.0001
Leukemia-free survival
p < 0.0001
Leukemic transformation
p = 1e-10
Probability
Cumulative incidence
0 1 2 3 4 5 6 7 8 9 10
Years
N. at risk
277 174 105 66 34 21 10 6 3 3 2
2808 1861 1291 910 605 368 241 161 112 83 67
N. at risk
252 142 71 42 25 16 8 5 3 3 2
2538 1598 1099 748 515 333 226 150 105 77 61
No. at risk
249 142 71 42 25 16 8 5 3 3 2
2534 1595 1097 747 515 333 226 150 105 77 61

Figure S27 | Summary molecular and clinical metrics of the *BCOR/L1* group. Two multi-panel figures, molecular (**A.**) and clinical (**B.**), are provided.

A. Molecular summary metrics comparing cases categorized within the group (blue) versus other groups (grey).

Top row. Boxplot distribution of the number of mutated genes per patient (left) and representation of the most frequent molecular alterations within the group (right).

Middle row. Boxplot and violin plots representing the distribution of the maximum values of variant allele fraction (VAF) (left) and of VAF adjusted for ploidy (middle left) per patient. Distribution of the VAF of all mutations identified in patients without (middle right) or with (right) adjustment for ploidy.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Bottom row. Density distribution of the VAF adjusted for ploidy of mutations from the most commonly mutated genes within the group (left). Relative order of gene mutation acquisition inferred through Bradley-Terry analysis. The numbers (n) indicate the number of informative pairwise precedences per gene (right).

B. Clinical summary metrics comparing cases categorized within the group (blue) versus all other groups (grey).

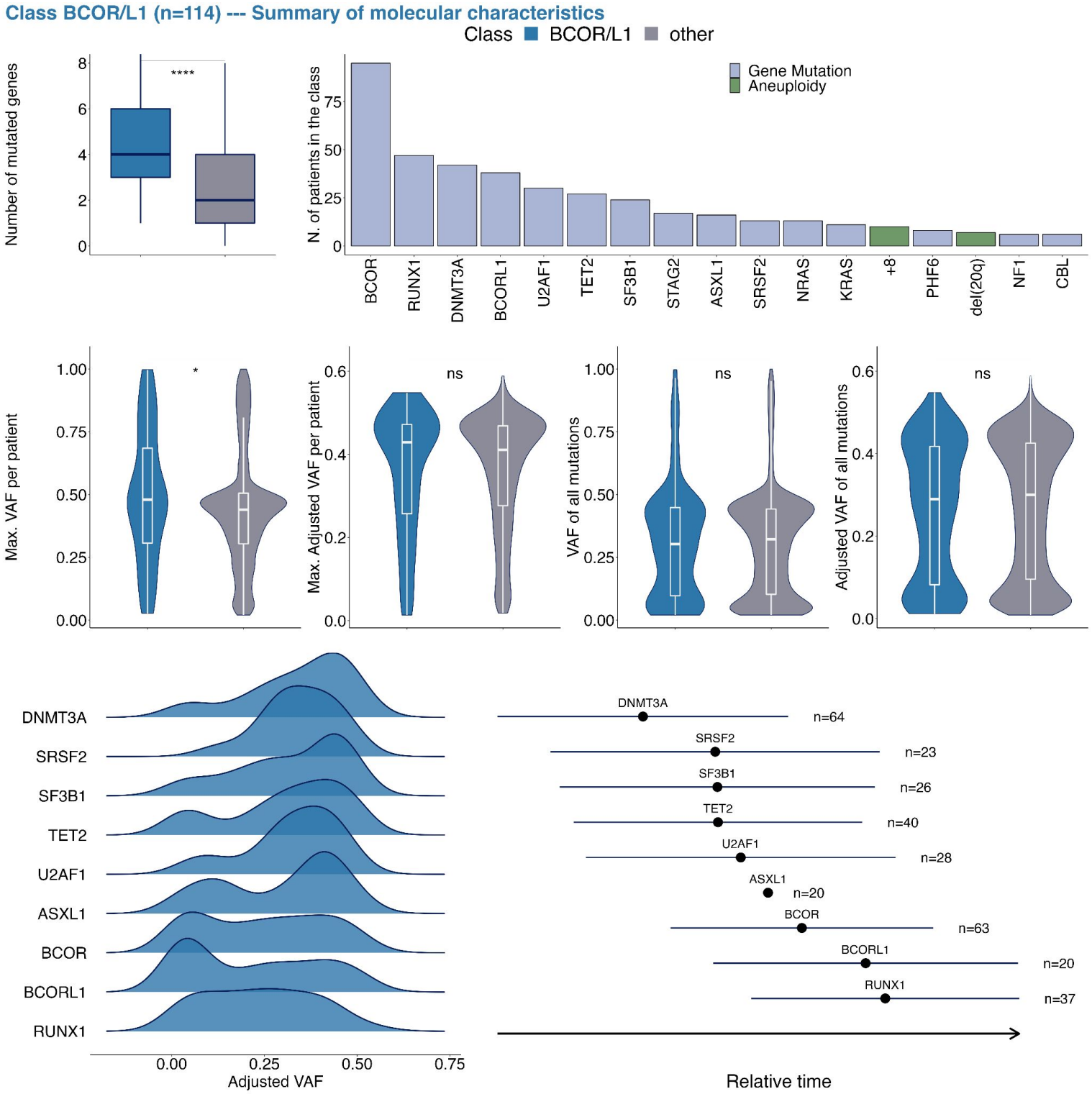
Top row. Distribution of the age at diagnosis, percentage of bone marrow blast, hemoglobin level, platelet count, white blood cell (WBC) count, and monocyte count.

****p<0.0001,***p<0.001, **p<0.01, *p< 0.05, ns: not significant, Wilcoxon rank-sum test.

Middle row. Left panel: Male (M) / Female (F) proportion. Right panels: Distribution of the WHO 2016 diagnosis subtypes (top) and of IPSS-M risk categories (bottom).

Bottom row. Left and middle panels: Kaplan-Meier probability estimates of overall survival (left) and leukemia free survival (middle). P-values are from the log-rank test. Right panel: cumulative incidence curves for the rate of leukemic transformation. P-value is from the Gray's test.

A.


Class BCOR/L1 (n=114) --- Summary of molecular characteristics
Class BCOR/L1 other
Number of mutated genes
****
N. of patients in the class
Gene Mutation
Aneuploidy
BCOR
RUNX1
DNMT3A
BCORL1
U2AF1
TET2
SF3B1
STAG2
ASXL1
SRSF2
NRAS
KRAS
+8
PHF6
del(20q)
NF1
CBL
Max. VAF per patient
*
Max. Adjusted VAF per patient
ns
VAF of all mutations
ns
Adjusted VAF of all mutations
ns
DNMT3A
SRSF2
SF3B1
TET2
U2AF1
ASXL1
BCOR
BCORL1
RUNX1
Adjusted VAF
n=64
n=23
n=26
n=40
n=28
n=20
n=63
n=20
n=37
Relative time

B.

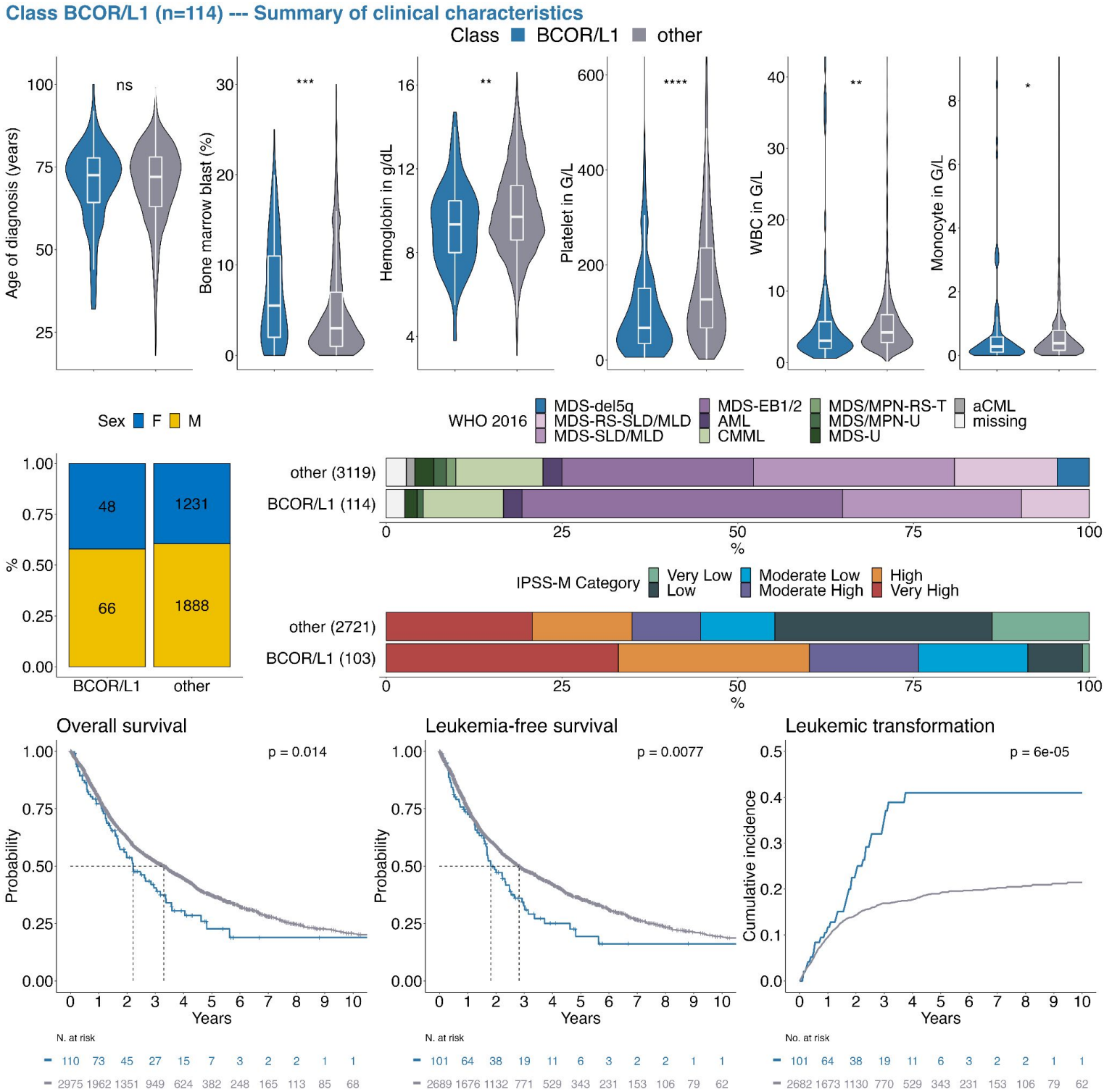
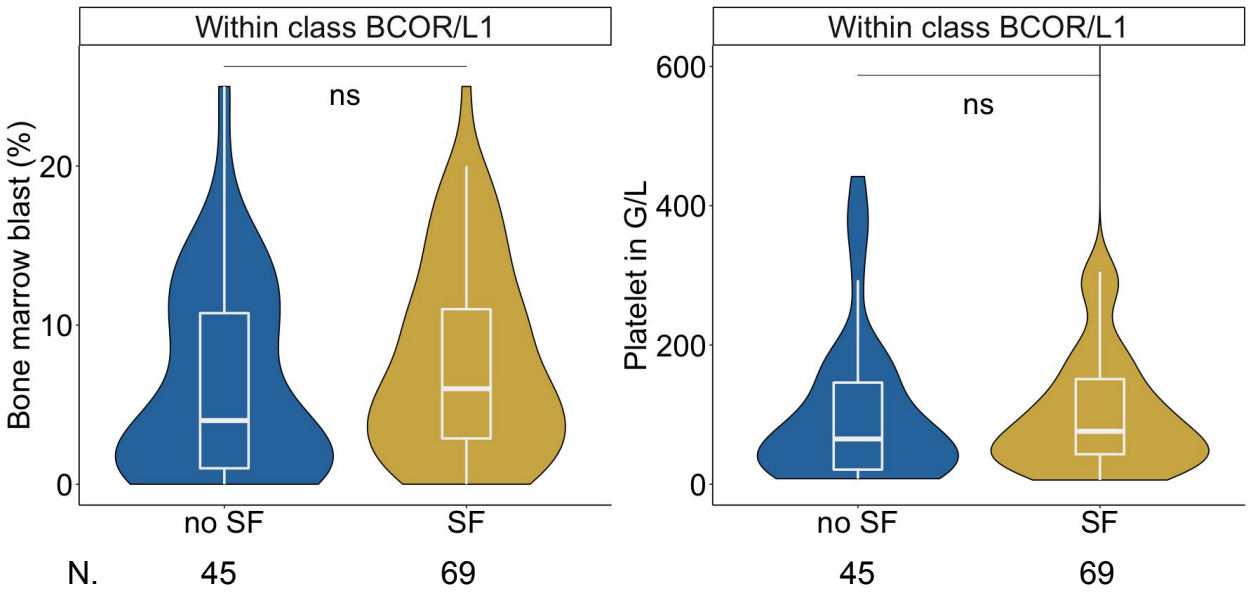

Class BCOR/L1 (n=114) --- Summary of clinical characteristics
Class BCOR/L1 other
Age of diagnosis (years)
ns
Bone marrow blast (%)
***
Hemoglobin in g/dL
**
Platelet in G/L
****
WBC in G/L
**
Monocyte in G/L
*
Sex F M
48
1231
66
1888
BCOR/L1
other
WHO 2016
MDS-del5q
MDS-RS-SLD/MLD
MDS-SLD/MLD
MDS-EB1/2
AML
CMML
MDS/MPN-RS-T
MDS/MPN-U
MDS-U
aCML
missing
other (3119)
BCOR/L1 (114)
IPSS-M Category
Very Low
Low
Moderate Low
Moderate High
High
Very High
other (2721)
BCOR/L1 (103)
Overall survival
p = 0.014
Leukemia-free survival
p = 0.0077
Leukemic transformation
p = 6e-05
Probability
Cumulative incidence
Years
N. at risk
110 73 45 27 15 7 3 2 2 1 1
2975 1962 1351 949 624 382 248 165 113 85 68
101 64 38 19 11 6 3 2 2 1 1
2689 1676 1132 771 529 343 231 153 106 79 62
No. at risk
101 64 38 19 11 6 3 2 2 1 1
2682 1673 1130 770 529 343 231 153 106 79 62

Figure S28 | Effect of splicing factor mutation status within the *BCOR/L1* group.

A. Boxplots and violin plots representing the distribution of the percentage of bone marrow blast for patients categorized within the *BCOR/L1* group and with (gold) or without (blue) splicing factor (SF) mutations.

B. Kaplan-Meier probability estimates of overall survival from cases categorized within the *BCOR/L1* group and with or without splicing factor (SF) mutations. P-value is from the log-rank test.

A.


Within class BCOR/L1
ns
Bone marrow blast (%)
20
10
0
no SF
SF
N.
45
69
Within class BCOR/L1
ns
Platelet in G/L
600
400
200
0
no SF
SF
45
69

B.

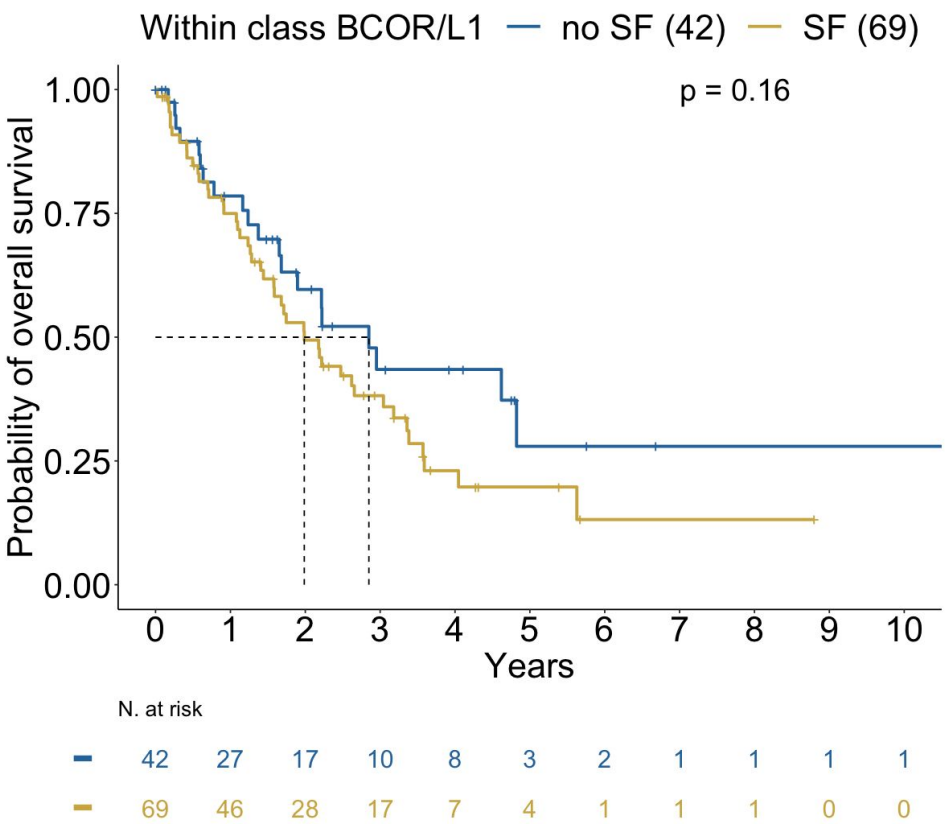

Within class BCOR/L1
no SF (42)
SF (69)
p = 0.16
Probability of overall survival
1.00
0.75
0.50
0.25
0.00
0
1
2
3
4
5
6
7
8
9
10
Years
N. at risk
42 27 17 10 8 3 2 1 1 1 1
69 46 28 17 7 4 1 1 1 0 0

Figure S29 | Characteristics of 193 patients without identified driver alterations by panel sequencing or cytogenetic.

A. Distribution of the age at diagnosis for cases without (“no”, blue) or with (“yes”, grey) identified driver alterations by the assay.

B. Male (M) / Female (F) proportion for cases without or with identified driver alterations.

C. Kaplan-Meier probability estimates of overall survival for cases without or with identified driver alterations. P-value is from the log-rank test.

D-E. Distribution of the WHO 2016 diagnosis subtypes (top) and of IPSS-M risk categories (bottom) for cases without or with identified driver alterations.

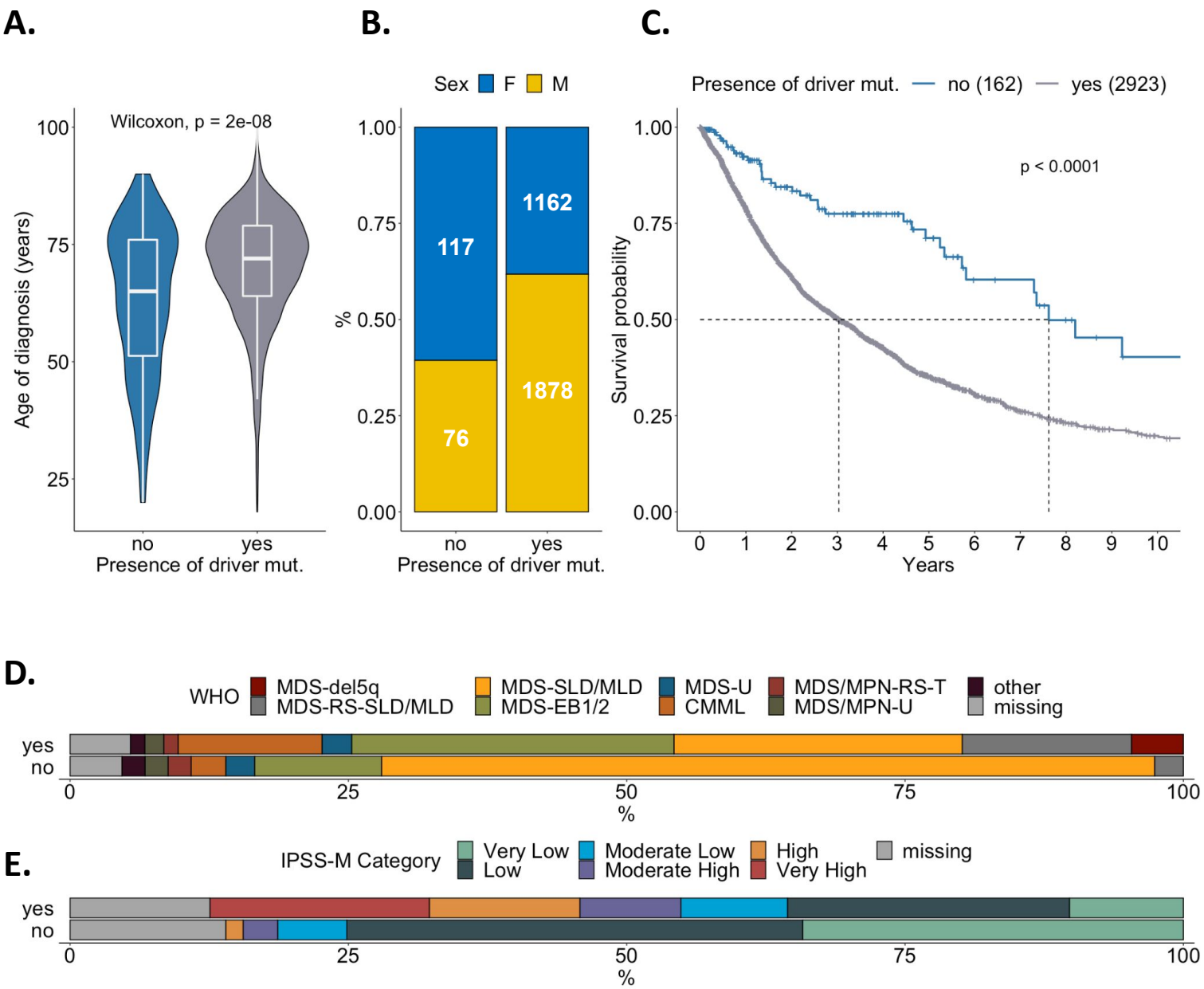

A.
B.
C.
Wilcoxon, p = 2e-08
Age of diagnosis (years)
100
75
50
25
no
yes
Presence of driver mut.
Sex F M
1.00
0.75
0.50
0.25
0.00
%
117
76
1162
1878
no
yes
Presence of driver mut.
Presence of driver mut. no (162) yes (2923)
p < 0.0001
Survival probability
1.00
0.75
0.50
0.25
0.00
0 1 2 3 4 5 6 7 8 9 10
Years
D.
WHO
MDS-del5q
MDS-RS-SLD/MLD
MDS-SLD/MLD
MDS-EB1/2
MDS-U
CMML
MDS/MPN-RS-T
MDS/MPN-U
other
missing
yes
no
0
25
50
75
100
%
E.
IPSS-M Category
Very Low
Low
Moderate Low
Moderate High
High
Very High
missing
yes
no
0
25
50
75
100
%

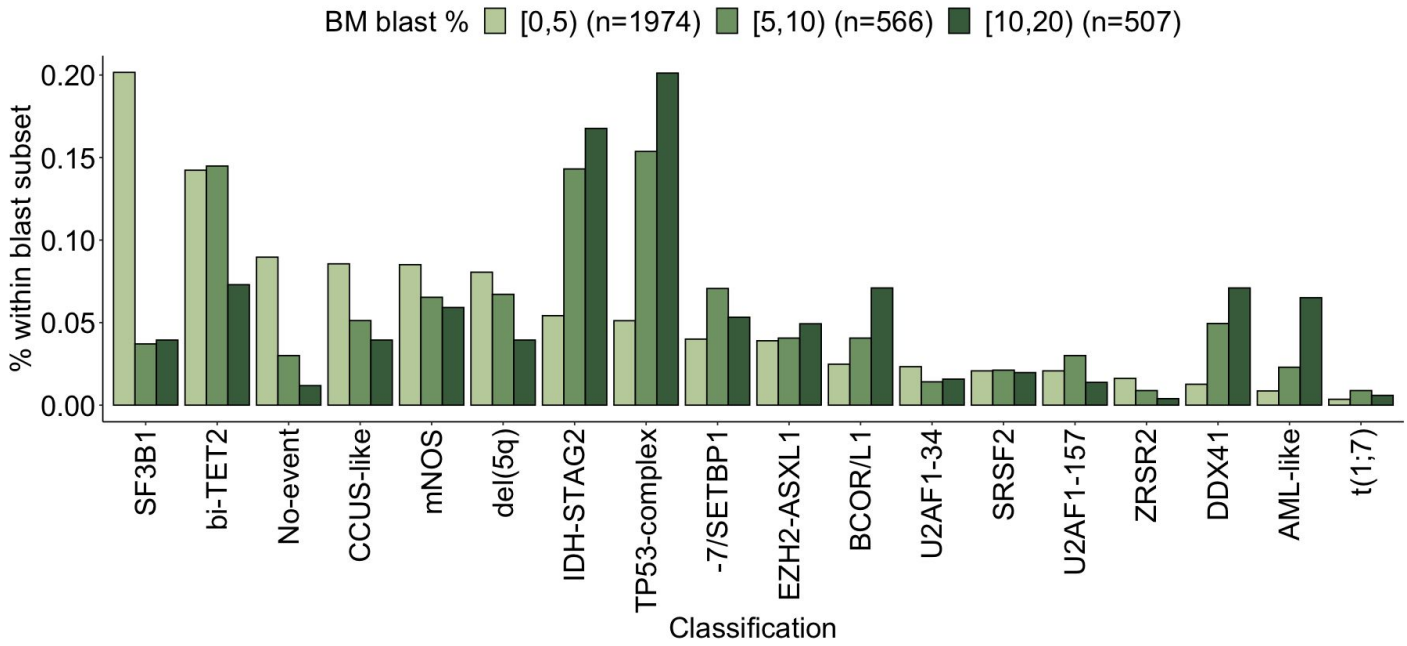
Figure S30 | Relationship between MDS molecular groups and bone marrow blast strata.

A. Proportion of molecular groups within each of the three bone marrow blast stratum (<5% in light green, 5-9% in green, and 10-19% in dark green). For example: *SF3B1* and *bi-TET2* are the two most frequent groups in the <5% blast subset, while *TP53-complex* and *IDH-STAG2* are the two most frequent groups in the 10-19% blast subset.

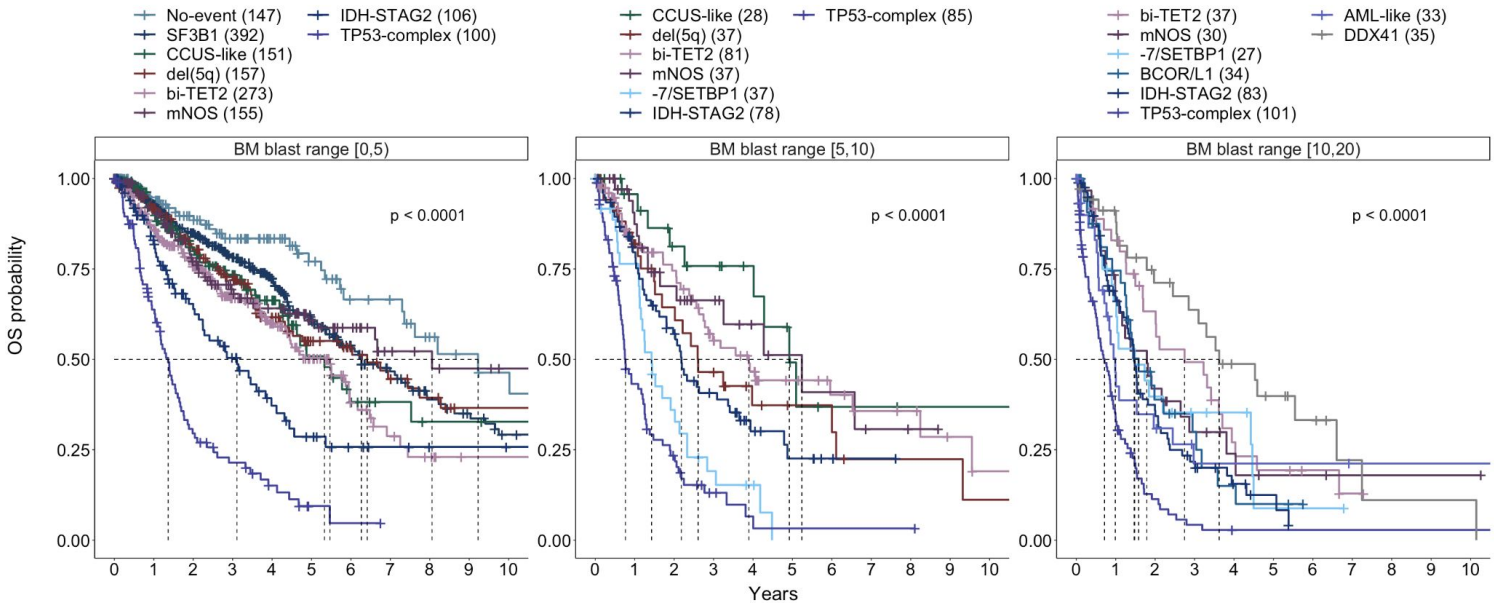
B. Kaplan-Meier probability estimates of overall survival (OS) within each bast stratum (<5% left, 5-9% middle, 10-19% right) stratified per molecular groups.

C. Multivariable Cox proportional hazard regression for overall survival including as features the molecular groups and the bone marrow blast strata.

A.



B.



C.

Multivariable Cox model for overall survival

Variable		N	Hazard ratio	p
Class	No-event	164	Reference	
	ZRSR2	34	1.29 (0.66, 2.53)	0.461
	mNOS	199	1.32 (0.87, 1.98)	0.188
	CCUS-like	180	1.37 (0.90, 2.08)	0.136
	SF3B1	416	1.38 (0.96, 1.98)	0.078
	del(5q)	195	1.59 (1.08, 2.34)	0.019
	bi-TET2	355	1.76 (1.22, 2.53)	0.002
	DDX41	65	1.83 (1.14, 2.95)	0.012
	SRSF2	51	2.20 (1.32, 3.66)	0.002
	U2AF1-157	57	2.58 (1.59, 4.19)	<0.001
	BCOR/L1	73	2.75 (1.75, 4.33)	<0.001
	IDH-STAG2	189	2.79 (1.91, 4.09)	<0.001
	U2AF1-34	52	2.88 (1.74, 4.76)	<0.001
	-7/SETBP1	119	4.21 (2.85, 6.22)	<0.001
	t(1;7)	11	4.22 (1.87, 9.53)	<0.001
	EZH2-ASXL1	101	5.42 (3.61, 8.12)	<0.001
	TP53-complex	218	6.39 (4.43, 9.21)	<0.001
	AML-like	30	6.39 (3.78, 10.80)	<0.001
BM.blast	[0,5)	1879	Reference	
	[5,10)	550	1.45 (1.26, 1.67)	<0.001
	[20,30]	80	1.91 (1.45, 2.53)	<0.001

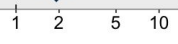
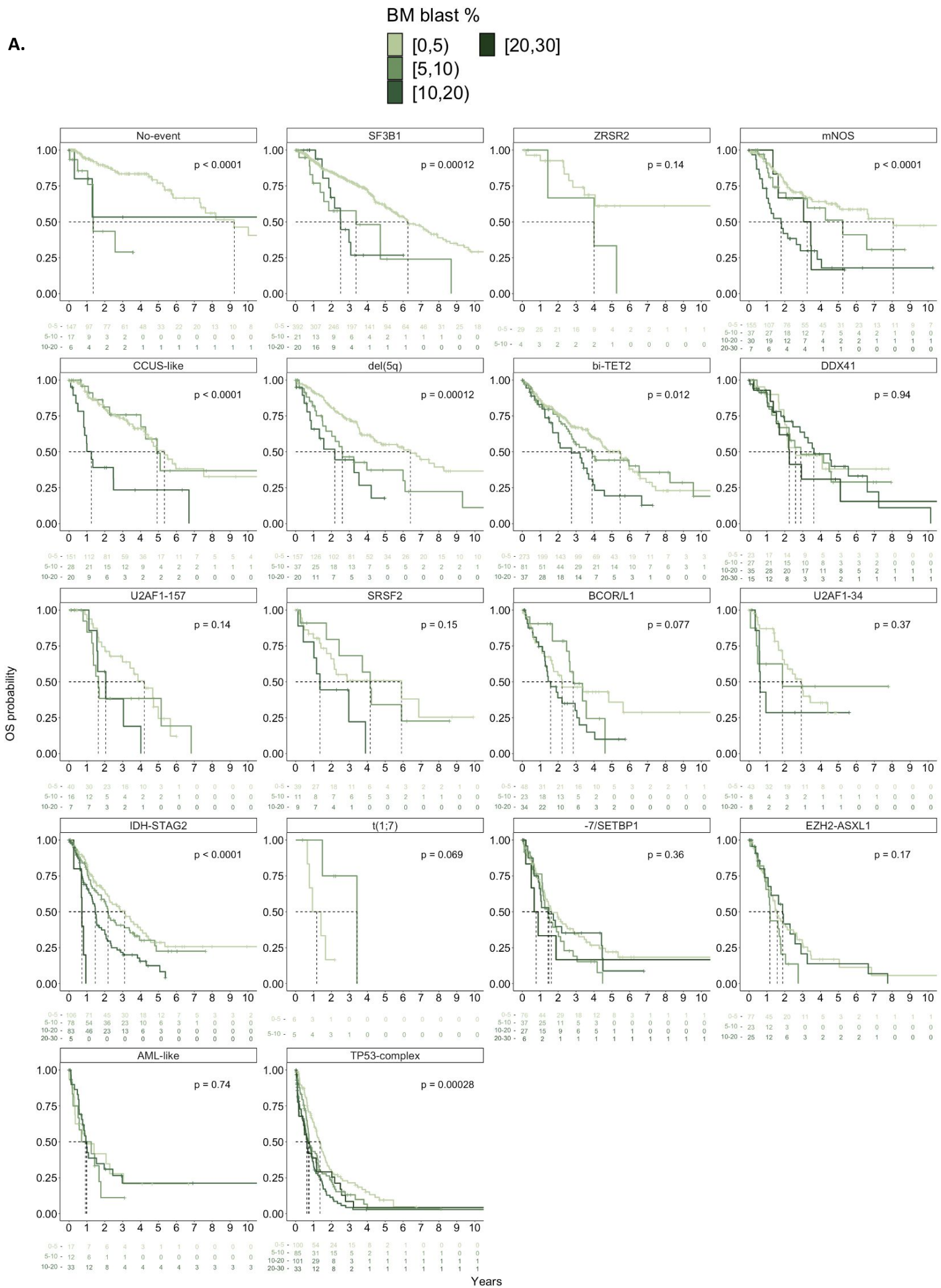


Figure S31 | Effect of bone marrow blast counts on overall survival per molecular group.

A-B. Kaplan-Meier probability estimates of overall survival (OS) (**A.**) and cumulative incidence curves of AML transformation (**B.**) across each molecular group (facets) stratified by bone marrow blast strata (<5% in light green, 5-9% in green, and 10-19% in dark green, 20-30% in very dark green). P-values in A. are from the log-rank test. P-values in B. are from the Gray's test.

A.


BM blast %
[0,5)
[5,10)
[10,20)
[20,30]
No-event
p < 0.0001
SF3B1
p = 0.00012
ZRSR2
p = 0.14
mNOS
p < 0.0001
CCUS-like
p < 0.0001
del(5q)
p = 0.00012
bi-TET2
p = 0.012
DDX41
p = 0.94
U2AF1-157
p = 0.14
SRSF2
p = 0.15
BCOR/L1
p = 0.077
U2AF1-34
p = 0.37
IDH-STAG2
p < 0.0001
t(1;7)
p = 0.069
-7/SETBP1
p = 0.36
EZH2-ASXL1
p = 0.17
AML-like
p = 0.74
TP53-complex
p = 0.00028
OS probability
Years

B.

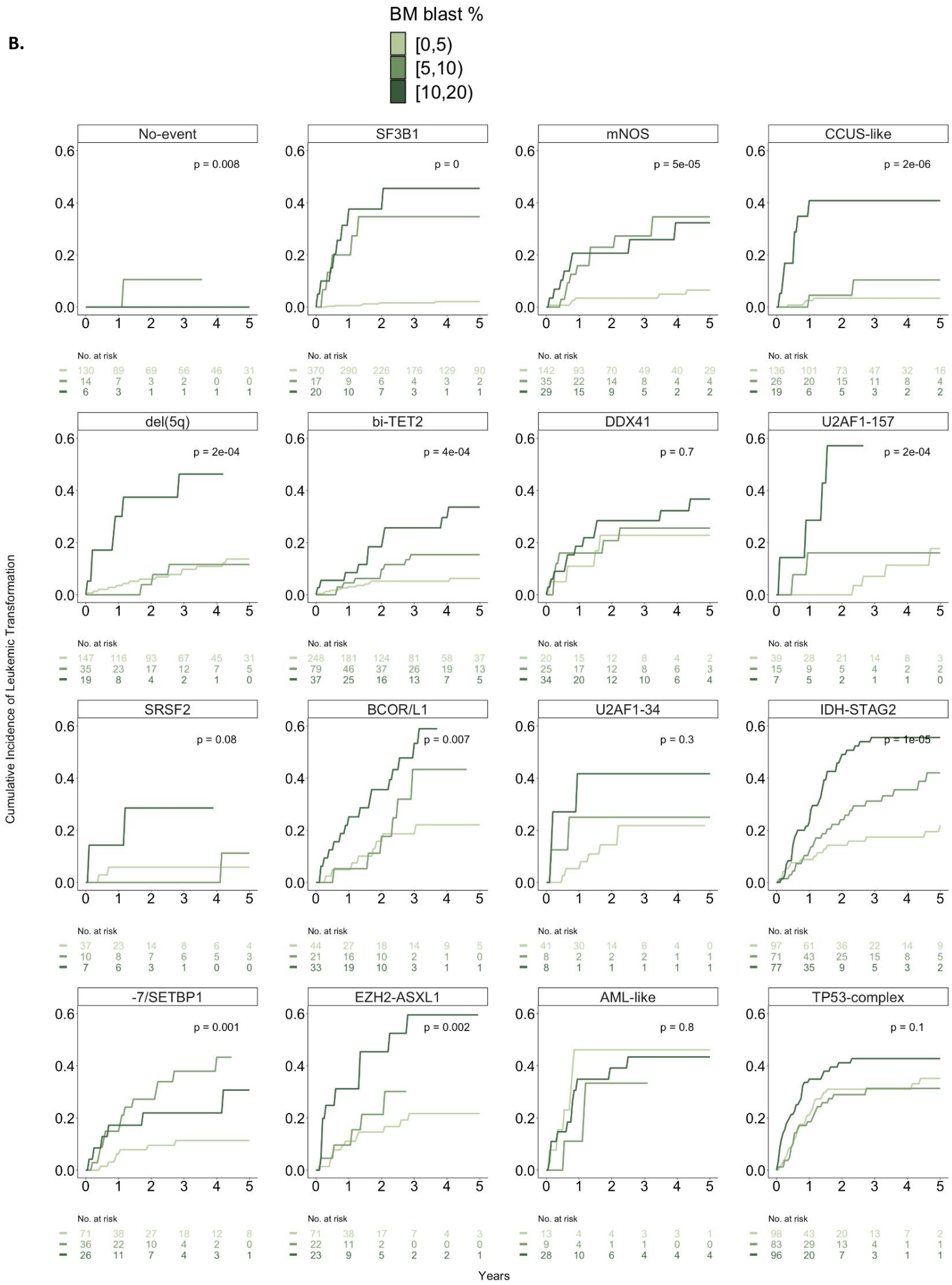

BM blast %
[0,5)
[5,10)
[10,20)
No-event
p = 0.008
SF3B1
p = 0
mNOS
p = 5e-05
CCUS-like
p = 2e-06
del(5q)
p = 2e-04
bi-TET2
p = 4e-04
DDX41
p = 0.7
U2AF1-157
p = 2e-04
SRSF2
p = 0.08
BCOR/L1
p = 0.007
U2AF1-34
p = 0.3
IDH-STAG2
p = 1e-05
-7/SETBP1
p = 0.001
EZH2-ASXL1
p = 0.002
AML-like
p = 0.8
TP53-complex
p = 0.1
No. at risk
Cumulative Incidence of Leukemic Transformation
Years

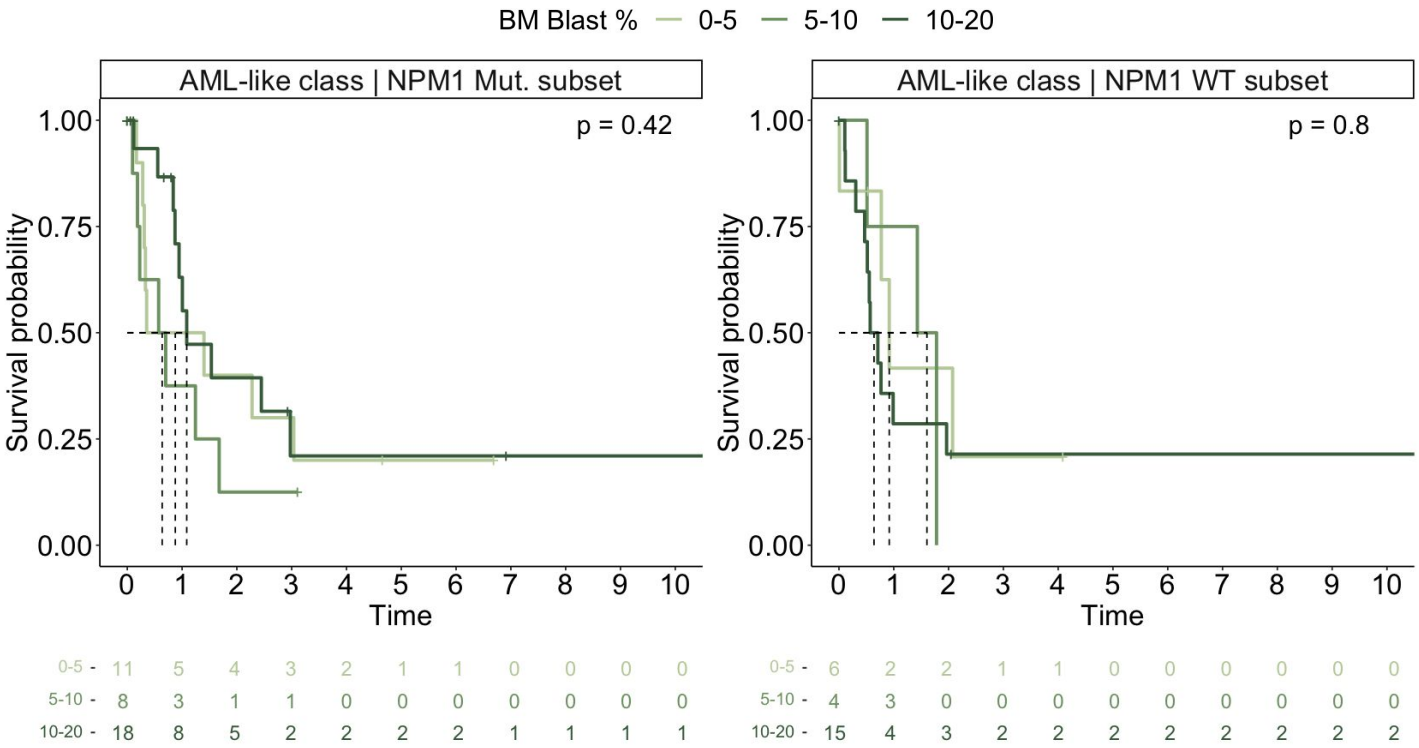
Figure S32 | Blast stratification in subsets of *AML-like* and *TP53-complex* groups.

A. Kaplan-Meier probability estimates of overall survival within the *AML-like* molecular group stratified by bone marrow blast strata (shades of green) and with (left) or without (right) *NPM1* mutations.

B. Kaplan-Meier probability estimates of overall survival within the *TP53-complex* molecular group stratified by bone marrow blast strata (shades of green) and with (left) or without (right) *TP53* mutations.
P-values are from the log-rank test.

Figure S32 | Bast stratification in subsets of *AML-like* and *TP53-complex* groups.

A.



B.

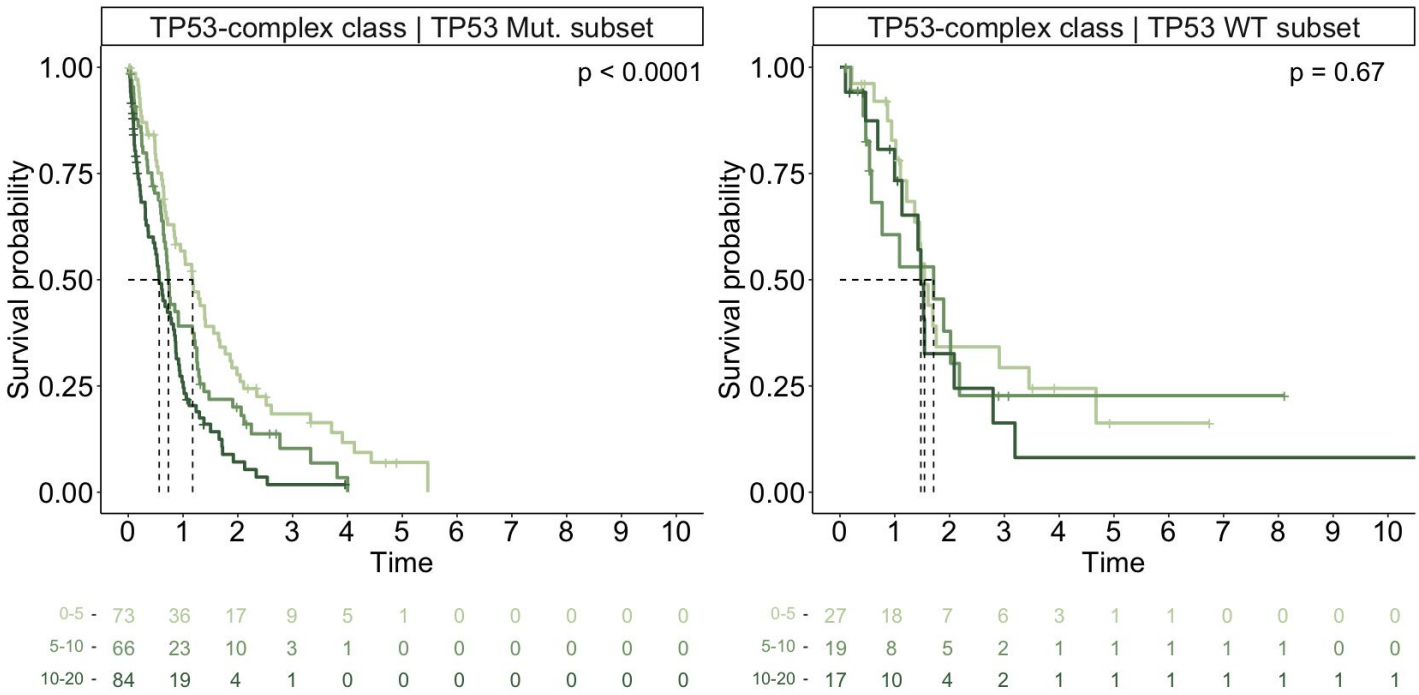


Figure S33 | Within group comparison of molecular characteristics between primary and s/t-MDS.

A-B. Distribution of the maximum value of variant allele fraction (VAF) per patient (**A.**) and of the VAF of all mutations (**B.**) for primary MDS (red) or secondary/therapy-related [s/t]-MDS (blue) within molecular groups. The comparison was restricted to the 4 most frequent groups in the s/t-MDS subset (*TP53-complex, -7/SETBP1, SF3B1, IDH-STAG2*). P-values are from the Wilcoxon rank-sum test.

C. Frequency distribution of gene mutations and cytogenetic alterations within the primary (red) or s/t-MDS (blue) subsets per molecular groups. No significant differences were observed between primary and s/t-MDS within the molecular groups.

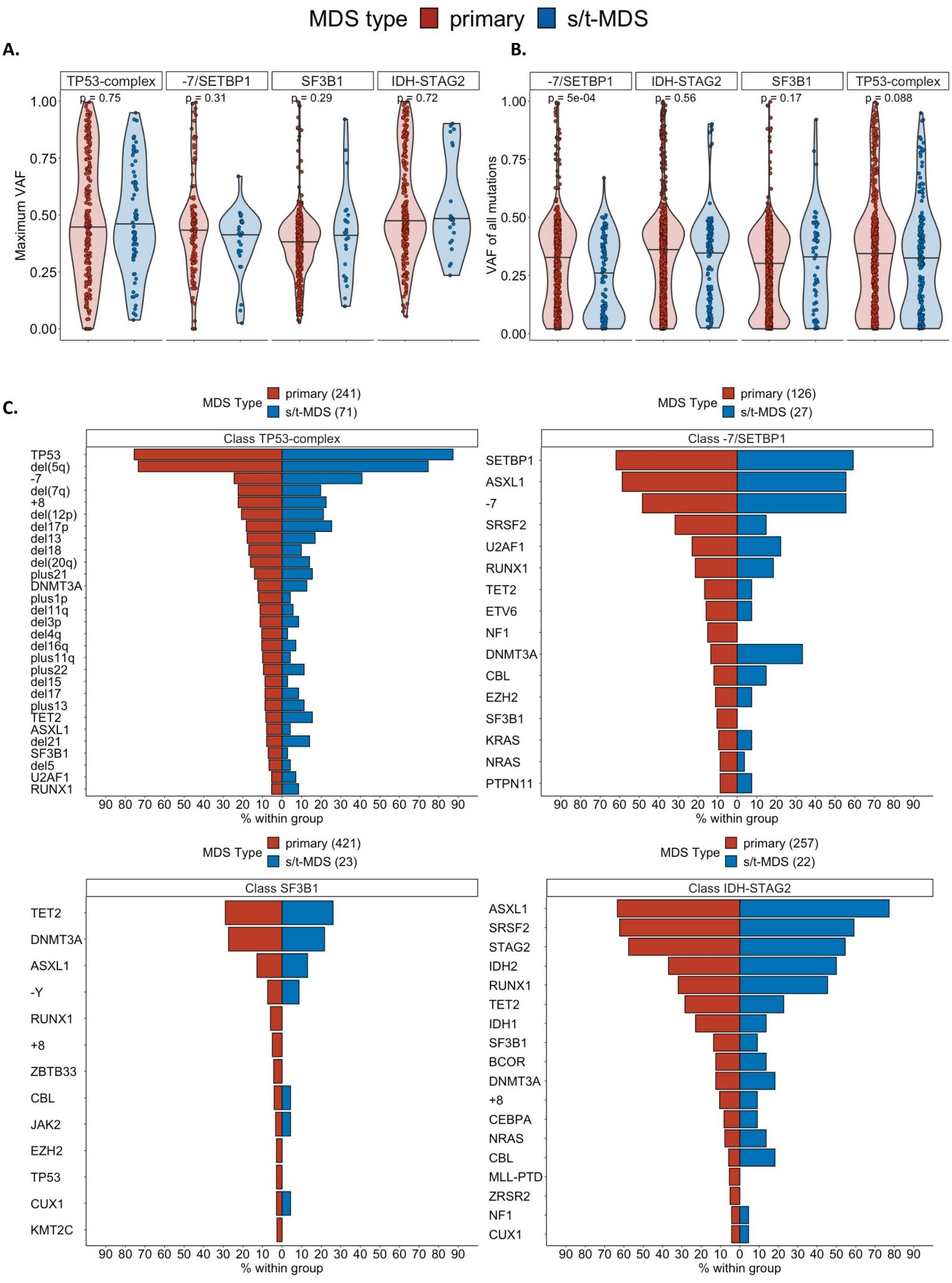

MDS type primary s/t-MDS
A.
TP53-complex
-7/SETBP1
SF3B1
IDH-STAG2
p = 0.75
p = 0.31
p = 0.29
p = 0.72
Maximum VAF
B.
-7/SETBP1
IDH-STAG2
SF3B1
TP53-complex
p = 5e-04
p = 0.56
p = 0.17
p = 0.088
VAF of all mutations
C.
MDS Type primary (241) s/t-MDS (71)
Class TP53-complex
TP53
del(5q)
-7
del(7q)
+8
del(12p)
del17p
del13
del18
del(20q)
plus21
DNMT3A
plus1p
del11q
del3p
del4q
del16q
plus11q
plus22
del15
del17
plus13
TET2
ASXL1
del21
SF3B1
del5
U2AF1
RUNX1
% within group
MDS Type primary (126) s/t-MDS (27)
Class -7/SETBP1
SETBP1
ASXL1
-7
SRSF2
U2AF1
RUNX1
TET2
ETV6
NF1
DNMT3A
CBL
EZH2
SF3B1
KRAS
NRAS
PTPN11
% within group
MDS Type primary (421) s/t-MDS (23)
Class SF3B1
TET2
DNMT3A
ASXL1
-Y
RUNX1
+8
ZBTB33
CBL
JAK2
EZH2
TP53
CUX1
KMT2C
% within group
MDS Type primary (257) s/t-MDS (22)
Class IDH-STAG2
ASXL1
SRSF2
STAG2
IDH2
RUNX1
TET2
IDH1
SF3B1
BCOR
DNMT3A
+8
CEBPA
NRAS
CBL
MLL-PTD
ZRSR2
NF1
CUX1
% within group

Figure S34 | Within group comparison of clinical characteristics between primary and s/t-MDS.

Distribution of the percentage of bone marrow blast, hemoglobin level, platelet count, and white blood cell (WBC) count for primary MDS (red) or secondary/therapy-related [s/t]-MDS (blue) within molecular groups. The comparison was restricted to the 4 most frequent groups in the s/t-MDS subset (*TP53-complex, -7/SETBP1, SF3B1, IDH-STAG2*). P-values are from the Wilcoxon rank-sum test.

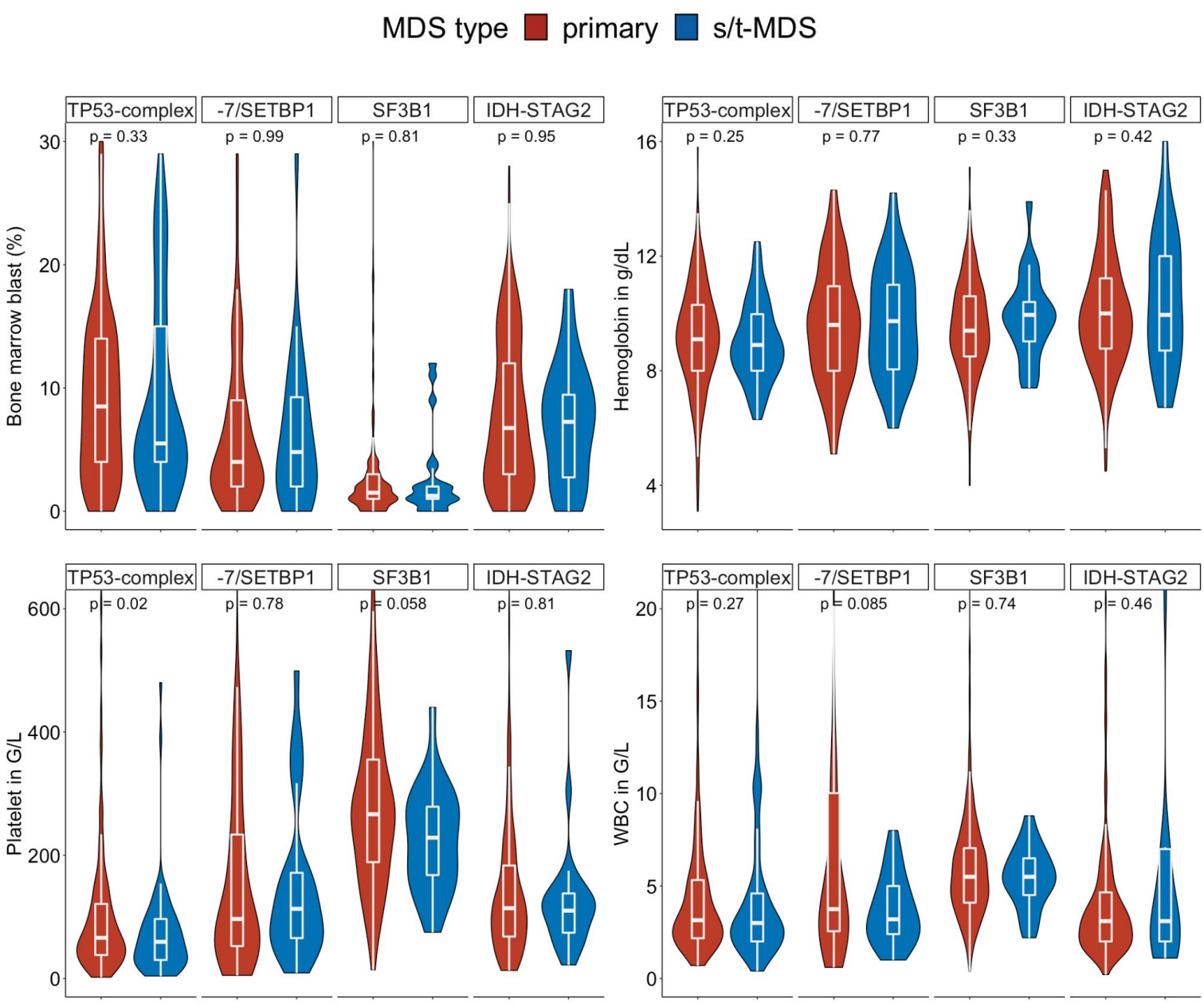

MDS type
primary
s/t-MDS
TP53-complex
-7/SETBP1
SF3B1
IDH-STAG2
Bone marrow blast (%)
p = 0.33
p = 0.99
p = 0.81
p = 0.95
Hemoglobin in g/dL
p = 0.25
p = 0.77
p = 0.33
p = 0.42
Platelet in G/L
p = 0.02
p = 0.78
p = 0.058
p = 0.81
WBC in G/L
p = 0.27
p = 0.085
p = 0.74
p = 0.46

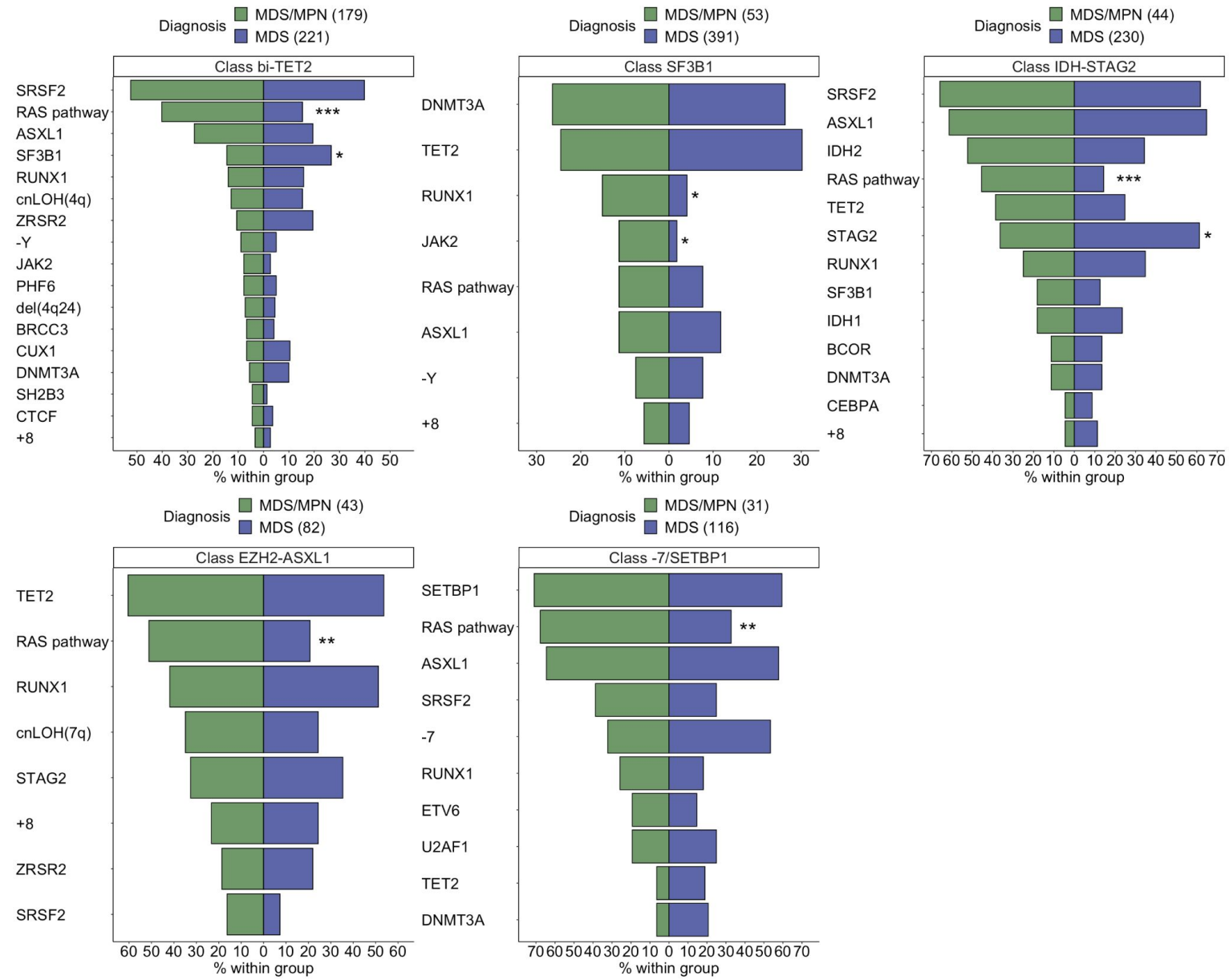
Figure S35 | Within group comparison of molecular characteristics between MDS and MDS/MPN.

A. Frequency distribution of gene mutations and cytogenetic alterations within the MDS (purple) or MDS/MPN (green) subsets per molecular groups. The comparison was restricted to the most frequent groups in the MDS/MPN subset (*bi-TET2, SF3B1, IDH-STAG2, EZH2-ASXL1, -7/SETBP1*). P-values are from the Wilcoxon rank-sum test. ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$, two-sided Fisher's exact test with Benjamini–Hochberg multiple testing correction.

RAS pathway mutations included mutations in *N/KRAS*, *CBL*, *NF1*, and *PTPN11*. *RAS* pathway mutations were observed in 68%, 51%, 45% and 40% of patients with MDS/MPN within groups -7/*SETBP1*, *EZH2-ASXL1*, *IDH-STAG2* and *bi-TET2*, respectively, compared to 33%, 21%, 14% and 15% of patients with MDS within the same groups. In the *SF3B1* group, *JAK2* mutations were associated with MDS/MPN over MDS (11% vs. 2% of patients, respectively)

B. Distribution of the variant allele fraction (VAF) of *RAS* pathway mutations for MDS (purple) or MDS/MPN (green) within molecular groups. P-values are from the Wilcoxon rank-sum test.

A.


Diagnosis
MDS/MPN (179)
MDS (221)
Class bi-TET2
SRSF2
RAS pathway
ASXL1
SF3B1
RUNX1
cnLOH(4q)
ZRSR2
-Y
JAK2
PHF6
del(4q24)
BRCC3
CUX1
DNMT3A
SH2B3
CTCF
+8
***
*
50 40 30 20 10 0 10 20 30 40 50
% within group
Diagnosis
MDS/MPN (53)
MDS (391)
Class SF3B1
DNMT3A
TET2
RUNX1
JAK2
RAS pathway
ASXL1
-Y
+8
*
*
30 20 10 0 10 20 30
% within group
Diagnosis
MDS/MPN (44)
MDS (230)
Class IDH-STAG2
SRSF2
ASXL1
IDH2
RAS pathway
TET2
STAG2
RUNX1
SF3B1
IDH1
BCOR
DNMT3A
CEBPA
+8
***
*
70 60 50 40 30 20 10 0 10 20 30 40 50 60 70
% within group
Diagnosis
MDS/MPN (43)
MDS (82)
Class EZH2-ASXL1
TET2
RAS pathway
RUNX1
cnLOH(7q)
STAG2
+8
ZRSR2
SRSF2
**
60 50 40 30 20 10 0 10 20 30 40 50 60
% within group
Diagnosis
MDS/MPN (31)
MDS (116)
Class -7/SETBP1
SETBP1
RAS pathway
ASXL1
SRSF2
-7
RUNX1
ETV6
U2AF1
TET2
DNMT3A
**
70 60 50 40 30 20 10 0 10 20 30 40 50 60 70
% within group

B.

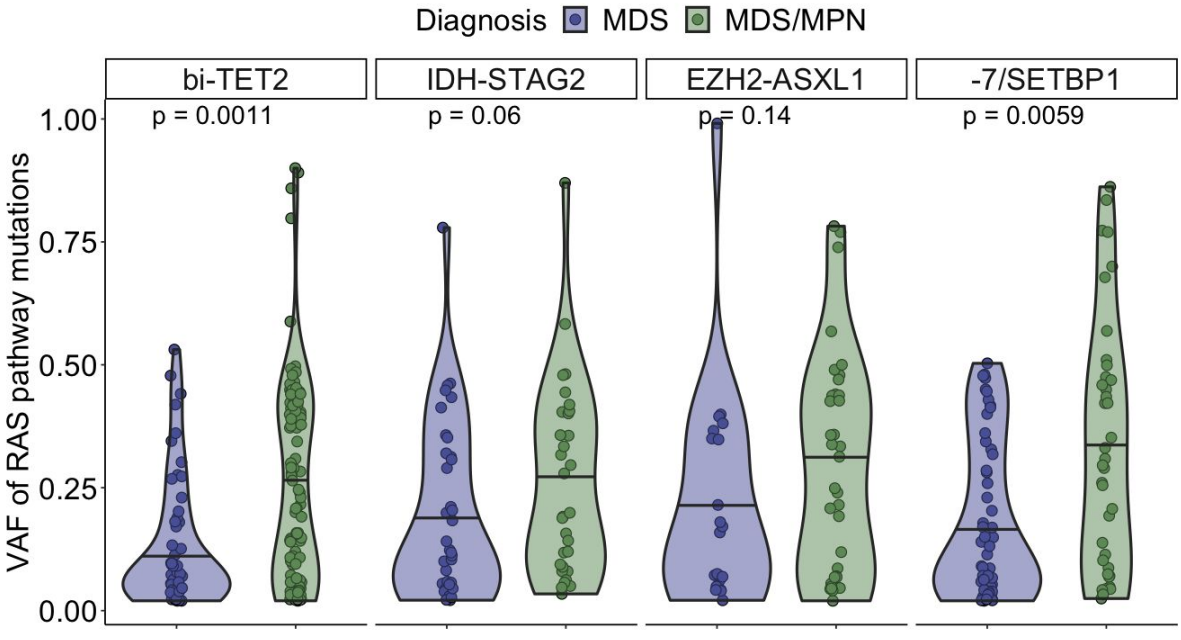

Diagnosis
MDS
MDS/MPN
bi-TET2
IDH-STAG2
EZH2-ASXL1
-7/SETBP1
p = 0.0011
p = 0.06
p = 0.14
p = 0.0059
VAF of RAS pathway mutations
1.00
0.75
0.50
0.25
0.00

Figure S36 | Multivariable models for prediction of leukocytosis, monocytosis, or absolute monocyte counts.

A. Multivariable logistic regression for the prediction of leukocytosis (left) and monocytosis (right) from the presence of gene mutations.

B-C. Multiple multivariable linear least square regression analysis using the variant allele fraction of gene mutations as explanatory features and white blood cell or monocyte count (log-transformed) as the target variables. The heatmap represents the fitted coefficients with darker purple denoting positive effects and darker red negative effects. The coefficients of determination (R^2) are indicated as dot-lines. The regressions were performed independently on the MDS (top) or MDS/MDS (bottom) subsets, as well as using all cases categorized as *bi-TET2*, *IDH-STAG2*, *EZH2-ASXL1*, or *-7/SETBP1* (**B.**), or only restricted to *bi-TET2* (**C.**).

**p<0.0001, *p<0.01, +p< 0.01, Z-score test.

A.

Leukocytosis

Mutation	N.	Odds ratio	p
RAS.pathway	500	6.34 (4.59, 8.77)	<0.001
EZH2	201	2.90 (1.74, 4.76)	<0.001
SRSF2	526	1.90 (1.26, 2.85)	0.002
TET2.biallelic	463	1.86 (1.24, 2.79)	0.003
ASXL1	812	1.78 (1.25, 2.54)	0.001
RUNX1	413	1.57 (1.06, 2.31)	0.023
IDH2	136	1.30 (0.63, 2.54)	0.457
TET2.other	509	1.05 (0.65, 1.65)	0.845
U2AF1	262	0.86 (0.47, 1.49)	0.611
SETBP1	108	0.76 (0.37, 1.47)	0.438
SF3B1	708	0.55 (0.32, 0.91)	0.026
STAG2	272	0.34 (0.18, 0.59)	<0.001
ZRSR2	163	0.33 (0.12, 0.76)	0.016

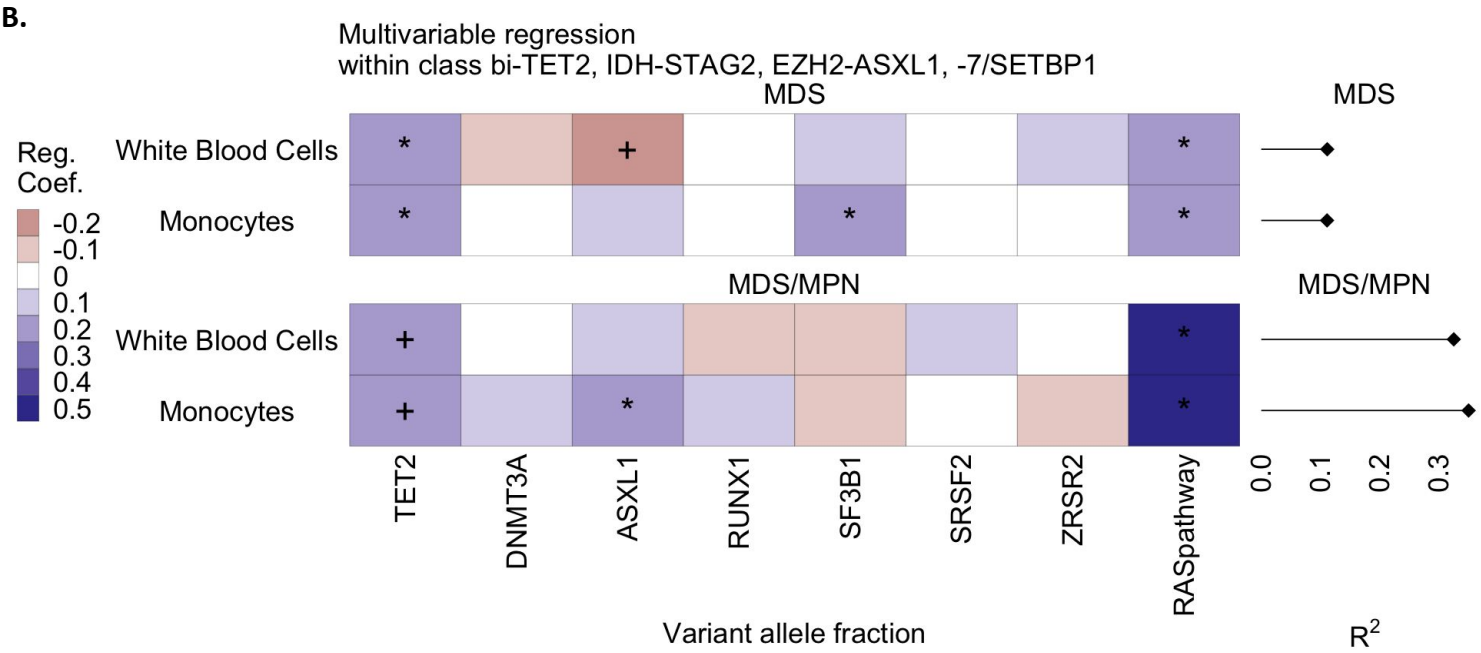
0.2 0.5 1 2 5

Monocytosis

Mutation	N.	Odds ratio	p
RAS.pathway	500	3.96 (3.10, 5.05)	<0.001
EZH2	201	0.95 (0.60, 1.47)	0.82
SRSF2	526	1.75 (1.30, 2.35)	<0.001
TET2.biallelic	463	4.66 (3.54, 6.14)	<0.001
ASXL1	812	1.35 (1.04, 1.75)	0.03
RUNX1	413	1.26 (0.92, 1.70)	0.14
IDH2	136	1.57 (0.92, 2.59)	0.09
TET2.other	509	1.14 (0.83, 1.57)	0.41
U2AF1	262	1.00 (0.65, 1.48)	0.98
SETBP1	108	0.95 (0.54, 1.60)	0.85
SF3B1	708	0.89 (0.66, 1.19)	0.43
STAG2	272	0.20 (0.11, 0.33)	<0.001
ZRSR2	163	1.36 (0.88, 2.06)	0.16

0.2 0.5 1 2 5

B.



C.

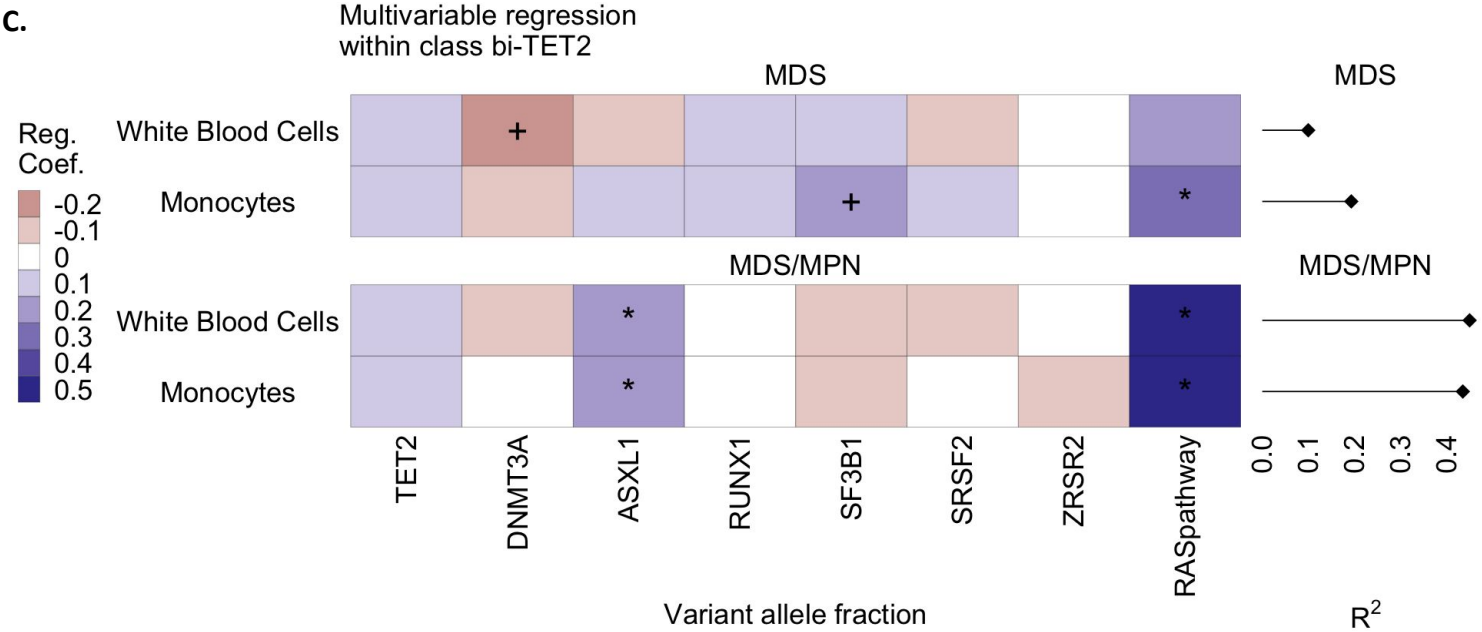
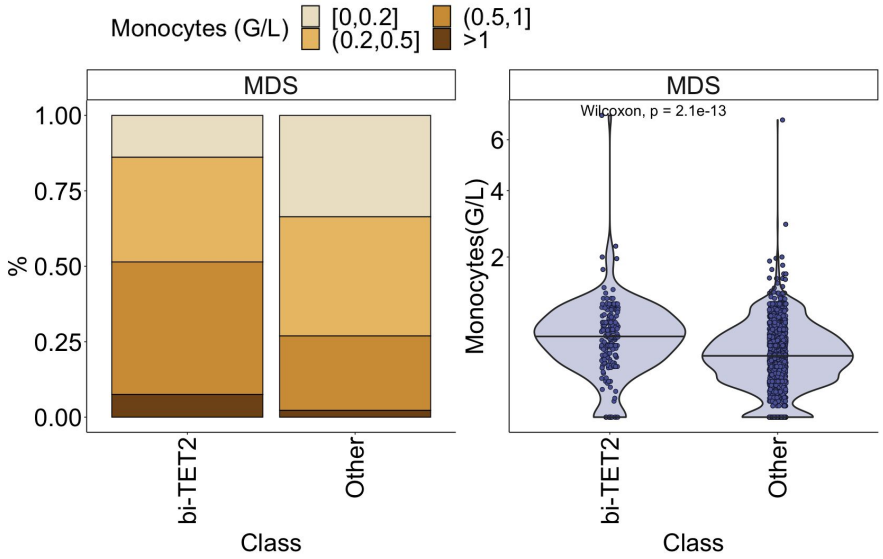


Figure S37 | Association between molecular groups, VAF of RAS pathway mutation, and monocyte counts.

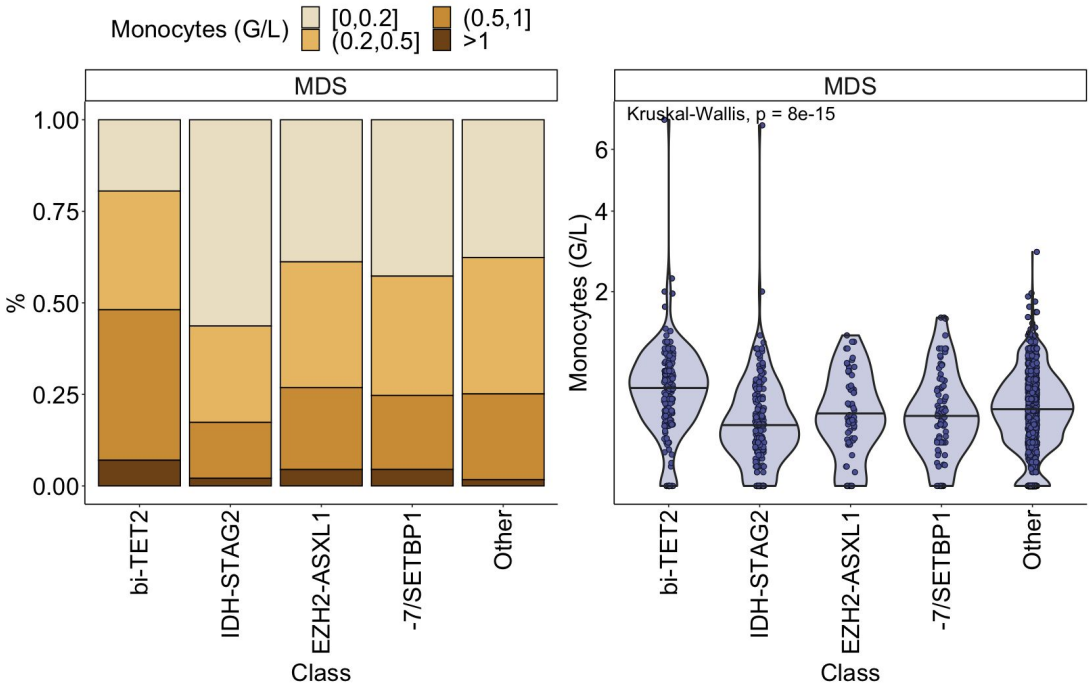
A-B. Stacked barplots (left) and violin plots (right) representing the levels of monocyte count of MDS patients categorized in the *bi-TET2* molecular group versus any other groups (**A.**) or versus the *IDH-STAG2, EZH2-ASXL1, -7/SETBP1* and other molecular groups (**B.**).

C. Violin plots representing the level of monocyte counts as a function of the variant allele fraction (VAF) of RAS pathway mutations, both for MDS patients (left) or MDS/MPN patients (right) categorized as *bi-TET2, IDH-STAG2, EZH2-ASXL1* or *-7/SETBP1*.

A.


Monocytes (G/L)
[0,0.2]
(0.2,0.5]
(0.5,1]
>1
MDS
MDS
Wilcoxon, p = 2.1e-13
%
Monocytes(G/L)
bi-TET2
Other
Class

B.


Monocytes (G/L)
[0,0.2]
(0.2,0.5]
(0.5,1]
>1
MDS
MDS
Kruskal-Wallis, p = 8e-15
%
Monocytes (G/L)
bi-TET2
IDH-STAG2
EZH2-ASXL1
-7/SETBP1
Other
Class

C.

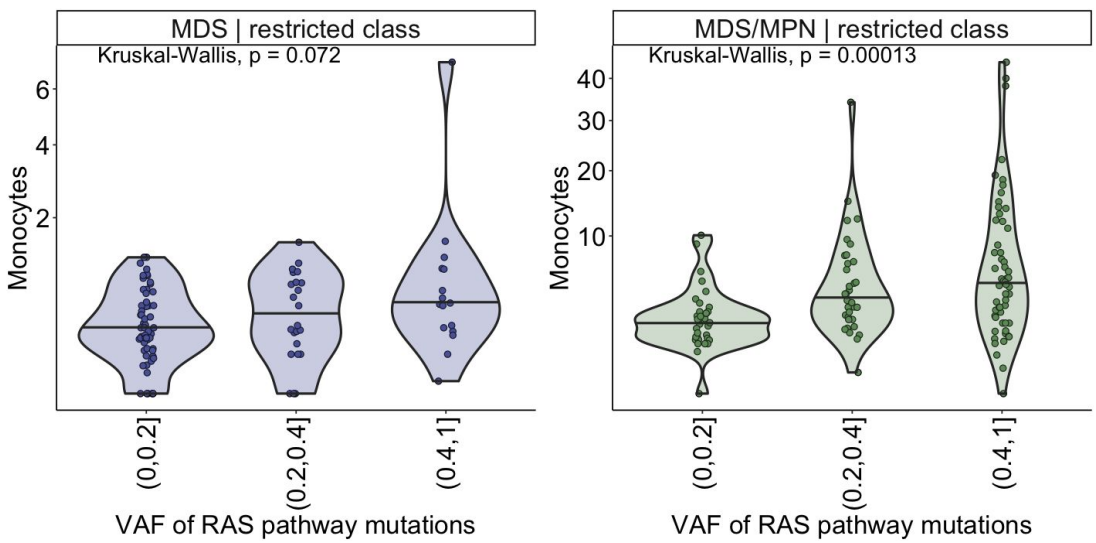

MDS | restricted class
Kruskal-Wallis, p = 0.072
MDS/MPN | restricted class
Kruskal-Wallis, p = 0.00013
Monocytes
(0,0.2]
(0.2,0.4]
(0.4,1]
VAF of RAS pathway mutations

Figure S38 | Within group comparison of clinical characteristics between MDS and MDS/MPN.

Distribution of the percentage of the age at diagnosis, bone marrow blast percentage, hemoglobin level, and platelet count for primary MDS (purple) or MDS/MPN (green) within molecular groups (*bi-TET2, IDH-STAG2, EZH2-ASXL1, -7/SETBP1*). P-values are from the Wilcoxon rank-sum test.

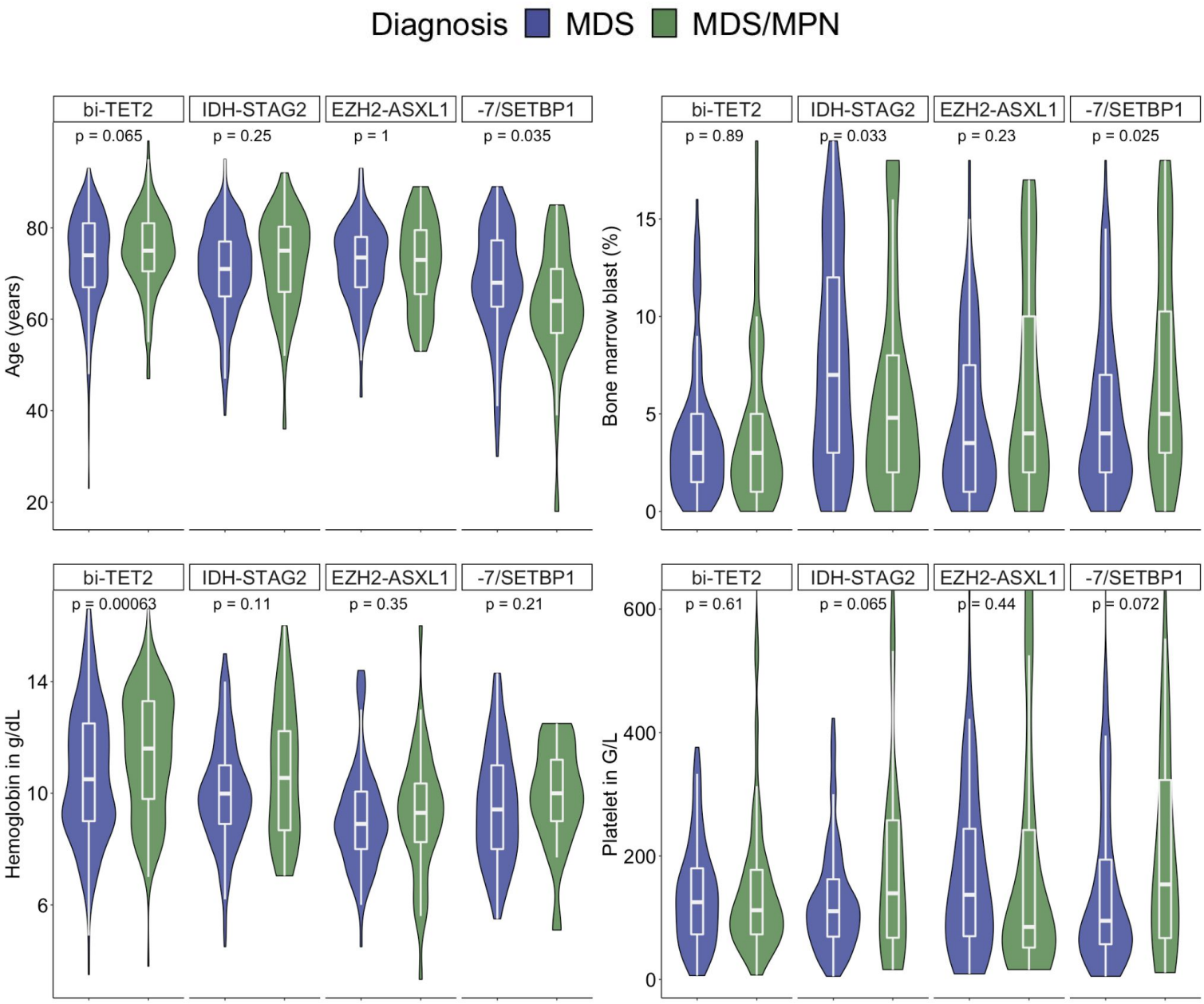
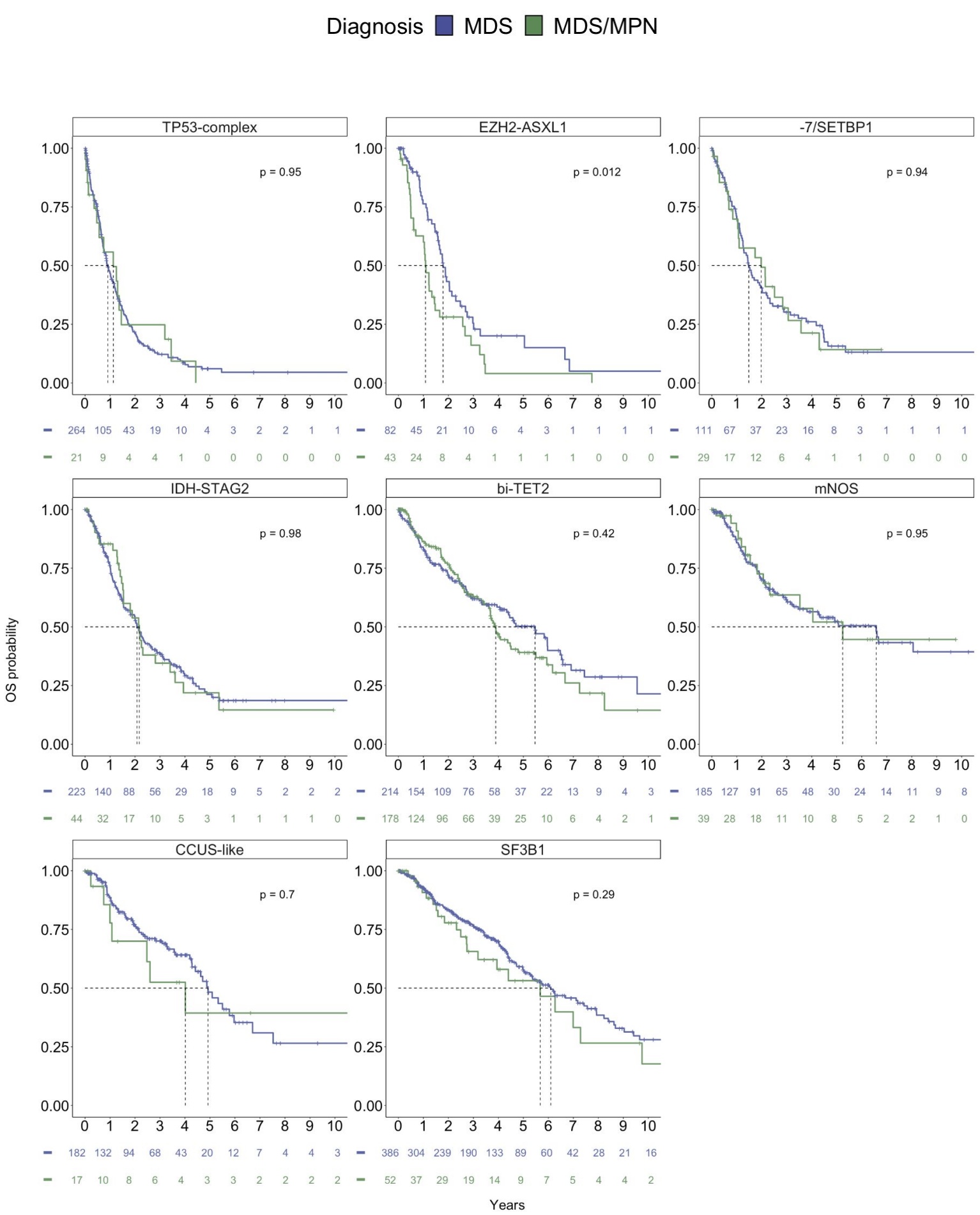

Diagnosis
MDS
MDS/MPN
bi-TET2
IDH-STAG2
EZH2-ASXL1
-7/SETBP1
p = 0.065
p = 0.25
p = 1
p = 0.035
Age (years)
20
40
60
80
p = 0.89
p = 0.033
p = 0.23
p = 0.025
Bone marrow blast (%)
0
5
10
15
p = 0.00063
p = 0.11
p = 0.35
p = 0.21
Hemoglobin in g/dL
6
10
14
p = 0.61
p = 0.065
p = 0.44
p = 0.072
Platelet in G/L
0
200
400
600

Figure S39 | Overall survival per MDS or MDS/MPN diagnostic subtypes across main molecular groups.

Kaplan-Meier probability estimates of overall survival for MDS (purple) or MDS/MPN (green) within each molecular group. P-values are from the log rank test.


Diagnosis MDS MDS/MPN
TP53-complex
p = 0.95
EZH2-ASXL1
p = 0.012
-7/SETBP1
p = 0.94
IDH-STAG2
p = 0.98
bi-TET2
p = 0.42
mNOS
p = 0.95
CCUS-like
p = 0.7
SF3B1
p = 0.29
OS probability
Years

BOX. Genetically-derived subgroups of MDS that have been defined in this study and may inform future classification schemas.

Molecular subgroup	% of all patients	Clinical outcomes	Distinctive feature(s)
DDX41-mutant	3.3	Frequent excess blasts and high risk of leukemic evolution, but favorable prognosis compared to other MDS with excess blasts; prognosis not affected by blast count or co-mutations.[3]	Typically germline mutation in one *DDX41* allele with somatic mutation in the second allele.[3] Hypoplastic bone marrow with excess blasts. Genomic diagnosis is crucial in the transplantation setting for donor selection and prevention of acute GVHD with post-transplant cyclophosphamide.
AML-like	2.0	Biologically resembles AML, except for less than 20% bone marrow blasts.	Some patients may be considered for intensive AML-like therapy.
TP53-complex*	10.1	Extremely aggressive disease with a median survival <1 year.	Poorly responsive to any currently available treatment including allogeneic stem cell transplantation. Enriched in therapy-related cases.
der(1;7)	0.5	Aggressive disease with short survival.	Frequent in the Asian population. Very high risk of life-threatening infection requiring careful prevention and monitoring during clinical course.[4]
-7/*SETBP1*-mutant	4.9	Aggressive disease with short survival; prognosis not stratified by blast count.	Somatic *SETBP1* mutations, often associated with monosomy 7/deletion 7q. Disease phenotypes include MDS, atypical chronic myeloid leukemia (aCML), CMML.[5,6] Enriched in therapy-related cases.
del(5q)*	6.9	Well established MDS subtype whose outcomes are determined by co-mutation patterns (e.g., SF3B1, TP53, and RUNX1).	Responsive to lenalidomide with amelioration of anemia and cytogenetic remission in a portion of patients. Higher risk of treatment resistance and progression in subjects with co-mutation in *TP53*.[7]
EZH2-ASXL1-mutant	4.0	Poor survival, despite the fact that most patients do not have excess blasts; prognosis not stratified by blast count.	High molecular complexity, with most patients carrying ≥5 mutated genes.
IDH-STAG2-mutant	8.9	Aggressive disease with a very high risk of leukemic transformation.	Patients with *IDH1* or *IDH2* mutation may benefit from treatment with ivosidenib or enasidenib, respectively.[8,9]
BCOR/L1-mutant MDS	3.5	Short survival and high risk of leukemic evolution.	Pronounced thrombocytopenia.[10]
Biallelic *TET2* mutation	12.7	Favorable outcomes in most patients.	Older patients, monocytosis overlapping with CMML.[11] Hematologic phenotype determined by co-mutations (anemia in

			SF3B1 mutated patients, thrombocytosis in *SF3B1* or *JAK2* mutated patients).
U2AF1-mutant (*37 & 154)*	4.3	Aggressive diseases with relatively poor survival and high risk of leukemic evolution.	Inhibitors of the spliceosome and modulators of splicing are being considered as potential treatments.[12,13]
SRSF2-mutant	2.2		
ZRSR2-mutant	1.3	Indolent clinical course.	Male patients with refractory macrocytic anemia and no excess of blasts.[14]
SF3B1-mutant*	14.1	Indolent clinical course with low risk of leukemic transformation. Progressive worsening of anemia over time with the development of transfusion requirements.	Ring sideroblast phenotype, ineffective erythropoiesis with iron overloading, specifically responsive to luspatercept treatment.[15,16]
CCUS-like	6.9	Mild phenotype and favorable outcomes in most patients.	Overlap with CCUS, with the only difference being the presence of significant dysplasia (≥10% versus <10% dysplastic cells).[17,18]
mNOS	7.9	Mild phenotype and favorable outcomes in most patients.	Rare putative oncogenic events, often occurring in patients with excess blasts.
No recurrent genetic event in myeloid genes	6.5	Long median overall survival, very low risk of leukemic transformation.	Enrichment for *UBA1* mutations and VEXAS-like clinical phenotypes.[19]

* Included as genetically-defined subgroups within the current classifications of myeloid neoplasms, i.e., the International Consensus Classification and the 5th edition of the WHO classification[1,2]

1. Arber DA, Orazi A, Hasserjian RP, et al. International Consensus Classification of myeloid neoplasms and acute leukemias: integrating morphologic, clinical, and genomic data. *Blood*. 2022;140(11):1200-1228.
2. Khoury JD, Solary E, Abla O, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: myeloid and histiocytic/dendritic neoplasms. *Leukemia*. 2022;36(7):1703-1719.
3. Makishima H, Saiki R, Nannya Y, et al. Germ line DDX41 mutations define a unique subtype of myeloid neoplasms. *Blood*. 2023;141(5):534-549.
4. Okuda R, Ochi Y, Saiki R, et al. Clonal evolution of der(1;7)(q10;p10) myeloid neoplasms [Abstract]. *Blood*. 2023;142(Supplement 1):1000-1000.
5. Piazza R, Valletta S, Winkelmann N, et al. Recurrent SETBP1 mutations in atypical chronic myeloid leukemia. *Nat Genet*. 2013;45(1):18-24.
6. Makishima H, Yoshida K, Nguyen N, et al. Somatic SETBP1 mutations in myeloid malignancies. *Nat Genet*. 2013;45(8):942-946.
7. Sperling AS, Guerra VA, Kennedy JA, et al. Lenalidomide promotes the development of TP53-mutated therapy-related myeloid neoplasms. *Blood*. 2022;140(16):1753-1763.

8. DiNardo CD, Roboz GJ, Watts JM, et al. Final phase I substudy results of ivosidenib in patients with mutant IDH1 relapsed/refractory myelodysplastic syndrome. *Blood Adv*. 2024.
9. Stein EM, Fathi AT, DiNardo CD, et al. Enasidenib in patients with mutant IDH2 myelodysplastic syndromes: a phase 1 subgroup analysis of the multicentre, AG221-C-001 trial. *Lancet Haematol*. 2020;7(4):e309-e319.
10. Abuhadra N, Mukherjee S, Al-Issa K, et al. BCOR and BCORL1 mutations in myelodysplastic syndromes (MDS): clonal architecture and impact on outcomes. *Leuk Lymphoma*. 2019;60(6):1587-1590.
11. Awada H, Nagata Y, Goyal A, et al. Invariant phenotype and molecular association of biallelic TET2 mutant myeloid neoplasia. *Blood Adv*. 2019;3(3):339-349.
12. Lee SC, Dvinge H, Kim E, et al. Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. *Nat Med*. 2016;22(6):672-678.
13. Su M, Fleisher T, Grosheva I, et al. Targeting SRSF2 mutations in leukemia with RKI-1447: A strategy to impair cellular division and nuclear structure. *iScience*. 2024;27(4):109443.
14. Fleischman RA, Stockton SS, Cogle CR. Refractory macrocytic anemias in patients with clonal hematopoietic disorders and isolated mutations of the spliceosome gene ZRSR2. *Leuk Res*. 2017;61:104-107.
15. Malcovati L, Stevenson K, Papaemmanuil E, et al. SF3B1-mutant MDS as a distinct disease subtype: a proposal from the International Working Group for the Prognosis of MDS. *Blood*. 2020;136(2):157-170.
16. Fenaux P, Platzbecker U, Mufti GJ, et al. Luspatercept in patients with lower-risk myelodysplastic syndromes. *N Engl J Med*. 2020;382(2):140-151.
17. Malcovati L, Galli A, Travaglino E, et al. Clinical significance of somatic mutation in unexplained blood cytopenia. *Blood*. 2017;129(25):3371-3378.
18. Weeks LD, Niroula A, Neuberg D, et al. Prediction of risk for myeloid malignancy in clonal hematopoiesis. *NEJM Evid*. 2023;2(5).
19. Sirenko M, Bernard E, Creignou M, et al. Molecular and clinical presentation of UBA1-mutated myelodysplastic syndromes. *Blood*. 2024:doi:10.1182/blood.2023023723.