Supplementary Information

Supplementary methods

Targeted sequencing

Panel information for targeted-capture sequencing

We applied targeted-capture sequencing using a panel designed for the genetic study of myeloid malignancies. A total of 445 genes were included in this panel, as previously described [1]. This panel also included 1,428 SNPs probes to detect genome-wide copy number changes and allelic imbalances.

Sequencing method

Genomic DNA (50 ng or 200 ng) was enriched for target regions by liquid-phase hybridization using the SureSelect custom kit (Agilent Technologies) according to the manufacturer's protocol optimized for automated sample processing, as previously described [2]. The purified library was subjected to high-throughput sequencing analysis with HiSeq 2500, NovaSeq 6000 (Illumina), or DNBSEQ-G400RS (MGI) using 125 bp or 150 bp paired-end mode. We also sequenced 106 blood samples from subjects without hematological diseases and used them as normal controls to exclude sequencing errors.

Mutation calling

Sequencing reads were aligned to the human genome reference (hg19) using Burrows-Wheeler Aligner, version 0.7.8, with default parameter settings [3]. Mutation calling was performed through our established pipeline (genomon pipeline 2.6.3, https://github.com/Genomon-Project), as previously reported [2, 4-6] using the following parameters.

Adopt variants fulfilling the following criteria:

(i) Mapping Quality score ≥ 20

(ii) Base Quality score ≥ 15

(iii) Number of total reads ≥ 100

(iv) Number of variant reads ≥ 4

(v) Variant allele frequency ≥ 0.02

(vi) Fisher's p-value for specific presentation compared with normal 1 controls < 0.02

Following candidates were excluded:

(i) Synonymous and ambiguous (unknown) variants

(ii) Variants which were read only from one direction

(iii) Single nucleotide substitutions in which other mutations were called at the same position and their variant allele frequency was ≥ 0.1.

Mapping errors were removed by visual inspection on the Integrative Genomics Viewer (IGV) browser (http://software.broadinstitute.org/software/igv/). Structural variants were also called using genomon pipeline 2.6.3.

Curation of oncogenic variants

The significant variants that fulfilled the quality filter noted above were further assessed for oncogenicity based on an in-house curation program. The curation policy was determined individually for each gene based on previous reports and databases after exclusion of variants registered in public SNPs databases (the 1000 genomes project as of 2014 Aug, ESP6500, Human Genome Variation Database) and call errors using EB call [7] and in-house blacklist of error calls.

Copy number and allelic imbalance

We included 1,428 SNPs probes to allow for the detection of copy number changes and allelic imbalances. This technique, called CNACS, is implemented in the program available at

https://github.com/papaemmelab/toil_cnacs. Manual inspection of the results was conducted to discriminate call errors. Total copy number (TCN) of 2.22 or larger was assumed CN-gains, and TCN <1.88 was assumed CN-loss. Copy number neutral LOH was called when B-allele frequency was <0.90 with TCN between 1.88 and 2.22. Arm-level copy number alterations (CNA) were called when the total length of the affected region within the arm was > 1 M bp for 17p and >3 M bp for the other arms. For gains in chromosome 8, CNA events in the long and short arms were counted together.

Calculation of mutation clone size

The variant allele frequency (VAF) of point mutations was adjusted to account for copy number alterations or allelic imbalances, and adjusted VAF values (adjVAF) were calculated, which represent the fraction of cells having relevant point mutations. Details of the adjustment calculation have been described previously [8].

Statistical analysis

For comparison of patient background or response, Fisher's exact test was applied for categorical data (including mutation profile), and the t-test was applied for parametric data. Mutation precedence was evaluated with the Bradley-Terry model. For survival analysis, the time of the first marrow examination that detected *NUP98*-rearrangement was assumed to be t =0, and the time of the first marrow test that diagnosed acute myeloid leukemia (AML) was assumed to be t =0 for non-*NUP98*-rearranged controls. For event-free survival (EFS), death from any cause, relapse, and failure to achieve remission with the first induction were assumed to be events. The probabilities of overall survival (OS) and EFS were estimated using the Kaplan-Meier method, and groups were compared using the log-rank test. Cumulative incidence curves and Fine and Gray competing risk regression models were applied for

relapse/non-relapse mortality (NRM) analysis, where relapse and NRM were assumed to be competing risks for each other. Cox proportional hazard regression was used for multivariate analysis of OS. To construct a simple predictive model for OS, significant and sub-significant factors (P <0.10) identified by univariate analysis with an incidence of >20% in the relevant cohort were subjected to multivariate analysis. The weight of each covariate in the multivariate analysis was determined based on the coefficients of the Cox proportional hazard model (log2(hazard ratio)). The total score for each patient was calculated by summing the assigned scores for present factors. The patients were then stratified into three risk groups based on their total score. All statistical analyses were performed using R (version 4.4.2). In addition, the R packages tidyverse_2.0.0, ggplot2_3.5.1, maftools_2.22.0, survminer_0.5.0, forestmodel_0.6.2, cmprsk_2.2-12, tidycmprsk_1.1.0, BradleyTerry2_1.1-2, and patchwork_1.3.0 were used for specific analysis and graphical presentation.

References

1. Ikoma Y, Nakamura N, Kaneda Y, Takamori H, Seki T, Hiramoto N, et al. Impact of myelodysplasia-related gene mutations and residual mutations at remission in venetoclax/azacitidine for AML. Leukemia. 2025;39:1362-1367.
2. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature. 2011;478:64-69.
3. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754-1760.
4. Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia. 2014;28:241-247.
5. Yoshizato T, Dumitriu B, Hosokawa K, Makishima H, Yoshida K, Townsley D, et al. Somatic Mutations and Clonal Hematopoiesis in Aplastic Anemia. N Engl J Med. 2015;373:35-47.
6. Suzuki H, Aoki K, Chiba K, Sato Y, Shiozawa Y, Shiraishi Y, et al. Mutational landscape and clonal architecture in grade II and III gliomas. Nat Genet. 2015;47:458-468.
7. Shiraishi Y, Sato Y, Chiba K, Okuno Y, Nagata Y, Yoshida K, et al. An empirical Bayesian framework for somatic mutation detection from cancer genome sequencing data. Nucleic Acids Res. 2013;41:e89.
8. Yoshizato T, Nannya Y, Atsuta Y, Shiozawa Y, Iijima-Yamashita Y, Yoshida K, et al. Genetic abnormalities in myelodysplasia and secondary acute myeloid leukemia: impact on outcome of stem cell transplantation. Blood. 2017;129:2347-2358.

Supplemental Figure Legends

Supplementary Fig. 1 ICC 2022 of the control AML cohort.

Supplementary Fig. 2 Genetic characteristics of adult AML patients in this study.

A Oncoprint depicting the genomic alterations identified across the cohort. Alterations are shown on the y-axis and individual patient samples on the x-axis. Different colors represent types of alterations. Only genomic alterations with ≥6 events are shown. **B** Correlation matrix of genomic alterations in the entire cohort. Green squares indicate co-occurrence, while brown squares denote mutual exclusivity between alterations. Color intensity reflects the strength of the association. The analysis was performed using the top 25 most frequent genomic events. **C** Diagonal plot comparing the mutated cell fractions of *NUP98* rearrangements and co-occurring mutations. The diagonal line indicates equal mutated cell fractions. Deviations from the diagonal reflect differences in clonal abundance; higher mutated cell fractions suggest earlier occurrence.

Supplementary Fig. 3 Survival of adult AML patients with *NUP98* rearrangements.

A Impact of *NUP98* rearrangements and ELN 2022 risk classification on non-relapse mortality (NRM) and relapse. **B-C** Kaplan–Meier estimates of (**B**) overall survival (OS) and (**C**) event-free survival (EFS) in AML patients with different *NUP98* fusion partners. **D-E** Kaplan–Meier estimates of (**D**) OS and (**E**) EFS in AML patients with *NUP98::HOXA9* versus *NUP98:*:non-*HOXA9*. Event rates were compared using the log-rank test.

Supplementary Fig. 4 Impact of co-mutations on survival in adult AML patients with and without *NUP98* rearrangements.

A-B Kaplan–Meier estimates of OS and EFS in adult AML patients (**A**) with and (**B**) without

NUP98 rearrangements, stratified by the presence of co-occurring mutations. **C** Kaplan–Meier estimates of EFS stratified by risk scores based on age and *FLT3*-ITD status in the *NUP98*-rearranged cohort. Event rates were compared using the log-rank test.

Supplementary Fig. 5 Impact of *NUP98* rearrangements and ELN 2022 risk classification on survival.

A Kaplan–Meier estimates of OS in the entire cohort based on ELN 2022 risk stratification. **B-C** Kaplan–Meier estimates of (**B**) OS and (**C**) EFS in the *NUP98*-rearranged cohort according to ELN 2022 risk stratification. **D-E** Kaplan–Meier estimates of OS showing the impact of allo-HSCT in AML patients under 70 years of age (**D**) with or (**E**) without *NUP98* rearrangements. Event rates were compared using the log-rank test.

Supplementary Table 1. List of primers used for PCR.

Target fusion	Primer name	Sequence (5'-3')
NUP98::MEOX2	*NUP98*_e12_F1	TGTTTGGGAACAACCAACCT
	*MEOX2*_R	CGCCTCAGTCTGGTGAGATA
NUP98::HOXA6	*NUP98*_e12_F2	GACTCTTGGAACTGGGCTTG
	*HOXA6*_R	CTGCGTGGAATTGATGAGC
NUP98::HOXD11	*NUP98*_e11_F1	GGACTCTTGGAACTGGGCTT
	*HOXD11*_R	AGCATCCGAGAGAGTTGAAGTC
NUP98::HOXD13	*NUP98*_e11_F2	GCTGGACAGGCATCTTTGTT
	*HOXD13*_R	AAGCTGTCTGTGGCCAACC
NUP98::GSX2	*NUP98*_e8_F	GGTAATACCAGCACCATAGGACAG
	*GSX2*_R	GGATGTGAGGGAGGAGGCCC

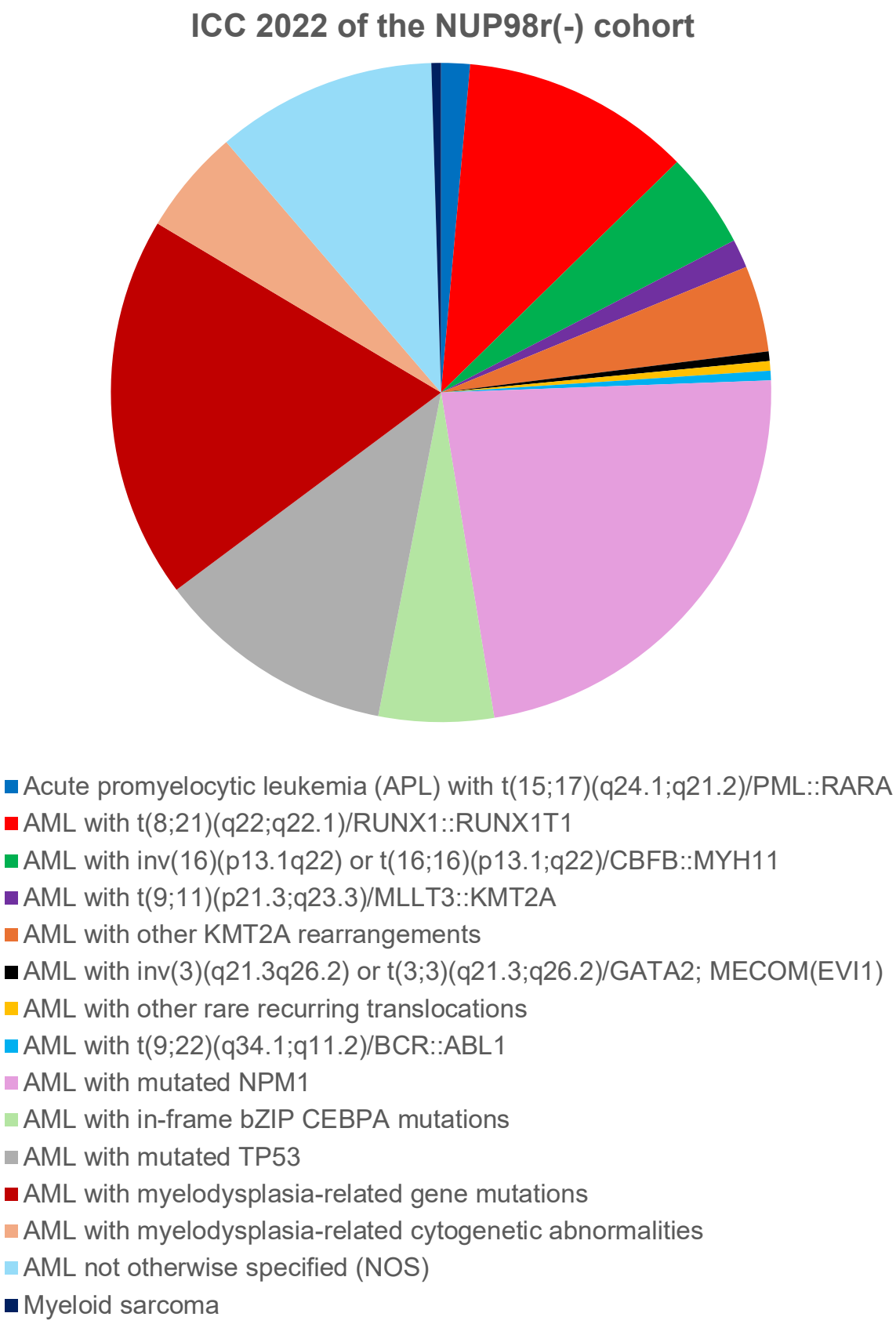
Supplementary Table 2. Karyotype of the *NUP98*-rearranged AML cohort.

Case #	Karyotype	Fusion partner
1	46,XX,t(7;11)(p15;p15),inv(9)(p12q13)[16]/46,XX,inv(9)(p12q13)[4]	HOXA9
15	NA	HOXA9
17	46,XY,t(7;11)(p15;p15)[20]	HOXA9
20	46,XX,t(7;11)(p15;p15),inv(9)(p12q13)[20]	HOXA9
23	46,XY,t(7;11)(p15;q15)[20]	HOXA9
29	46,XX,t(7;11)(p15;p15)[18]/46,XX[2]	HOXA9
34	46,XY,t(7;11)(p15;p15)[20]	HOXA9
35	46,XY,inv(1)(p22q12),-4,add(6)(p21),?t(7;11)(p15;p15),add(10)(p11.2),add(12)(q24.1),-14,add(16)(q11.2),-21,+mar1,+mar2,+mar3[11]/46,XX[9]	HOXA9
36	46,XX,t(7;11)(p15;p15)[18]/47,idem,+8[2]	HOXA9
37	46,XY,t(7;11)(p15;p15)[20]	HOXA9
42	46,XY,t(7;11)(p15;p15)	HOXA9
43	46,XY,t(7;11)(p15;p15)[20]	HOXA9
3	46,XY	NSD1
5	46,XY,inv(9)(p12q13)[20]	NSD1
6	46,XY	NSD1
9	47,XY,+8[2]/46,XY[18]	NSD1
11	46,XX	NSD1
12	46,XY	NSD1
13	NA	NSD1
14	46,XX	NSD1
19	47,XY,+8[10]	NSD1
21	46,XY	NSD1
24	46,XY	NSD1
26	46.XX	NSD1
27	46,XY,add(1)(q21),t(7;8)(q11.2;q24)[7]/46,XY[13]	NSD1
30	46,XX,+6[20]	NSD1
38	46,XY,del(9)(q?) [1]/46,XY[19]	NSD1
39	46,XX,t(2;5)(p12;q35)[20]	NSD1
41	48,XY,+15,+21 with numerical multiple abnormalities[18]	NSD1
44	46,XX	NSD1
2	46,XY,t(7;11)(p15;p15)	HOXA6
4	46,XX,t(2;11)(q31;p15)[20]	HOXD8
7	46,XY,t(1;11)(q23;p15)[5]/46,idem,del(12)(p?)[15]	PRRX1
8	46,XY	TNRC18
10	46,XY,t(2;11)(q31;p15)[20]	EVX2 (HOXD13)
16	46,XX,t(7;11)(p15p15)[20]	HOXA11
18	46,XX,t(4;11)(q12;p15)[16]/46,XX[4]	GSX2
22	46,XX,t(11;20)(p15;q11.2)[4]/46,XX,del(9)(q13q34),t(11;20)(p15;q11.2)[15]/46,XX[1]	TOP1
25	46,XX	PRRX2
28	44,X,-Y,-7,-20,-21,+mar1,+mar2[13]/46,XY[3]	MEOX2
31	47,XX,add(6)(p21),+8,inv(11)(p15q22)[19]/46,XX[1]	DDX10
32	46,XX,t(11;12)(p15;q13)[20]	HOXC13
33	46,XY	KDM5A
40	46,XX,t(7;11)(p15;p15)[20]	HOXA11

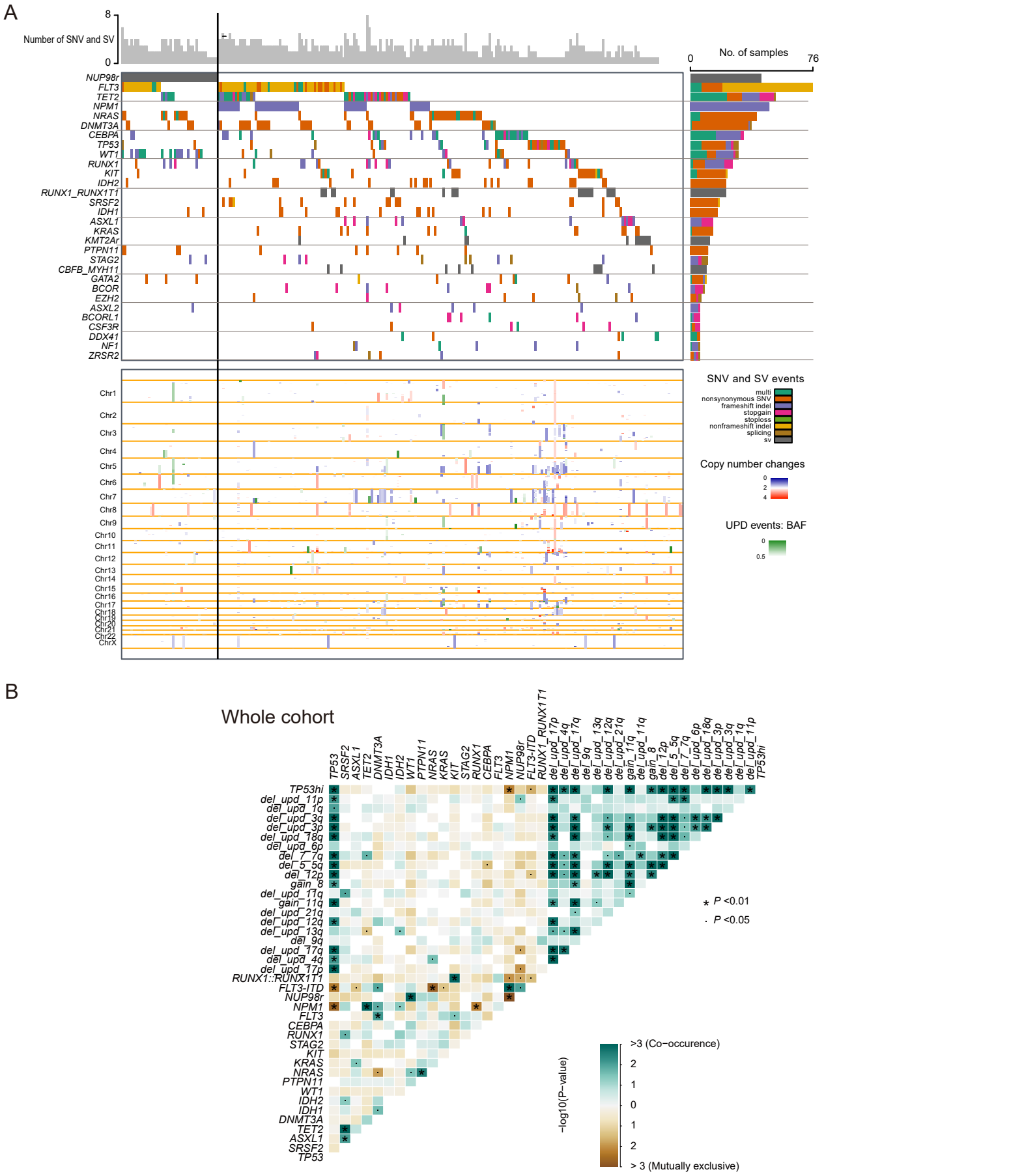
Supplementary Table 3. The impact of age, sex, fusion partners, and co-mutations on OS and EFS in the *NUP98*-rearranged cohort by univariate analysis.

			OS	EFS
		n	P-value (log-rank test)	P-value (log-rank test)
Age	>median (51 year)	21	0.019	0.610
Sex	Male	18	0.540	0.499
Fusion partner	*HOXA9*	12	0.170	0.081
	NSD1	18		
	Others	14		
Fusion partner	*HOXA9*	12	0.062	0.061
	non-*HOXA9*	32		
Co-mutation	*FLT3-ITD*	17	0.000	0.011
	NRAS	11	0.034	0.025
	WT1	17	0.140	0.614
	TET2	6	0.082	0.398
	DNMT3A	6	0.791	0.910
	RUNX1	5	0.909	0.434

Supplementary Figure 1

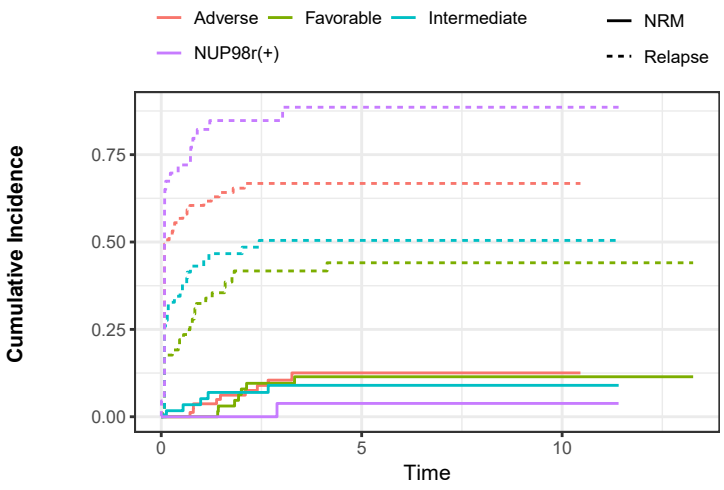


Supplementary Figure 2



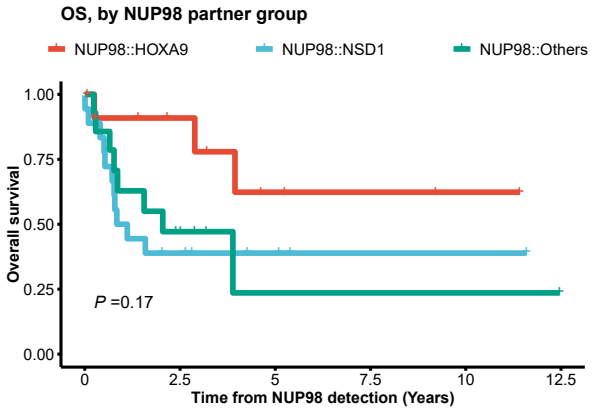
Supplementary Figure 3

A



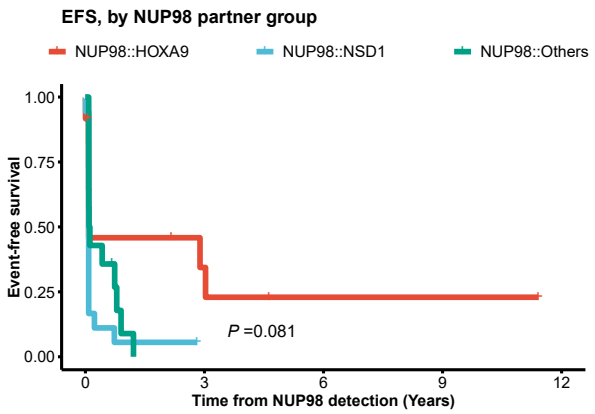
At Risk			
Adverse	86	6	2
Favorable	69	17	5
Intermediate	58	16	3
NUP98	44	1	1
Events			
Adverse	0	65	65
Favorable	0	36	36
Intermediate	0	34	34
NUP98	0	38	38

B



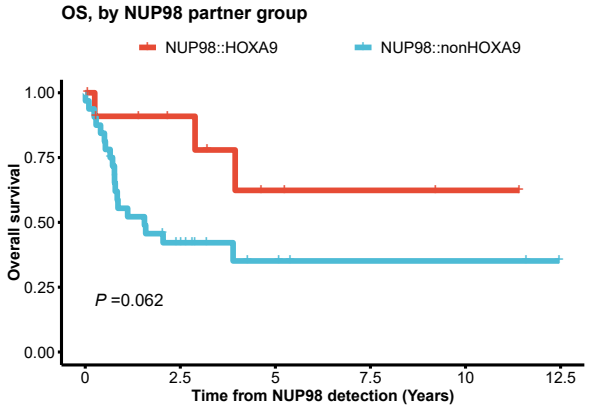
Number at risk					
12	7	3	2	1	0
18	6	3	1	1	0
14	5	1	1	1	0

C



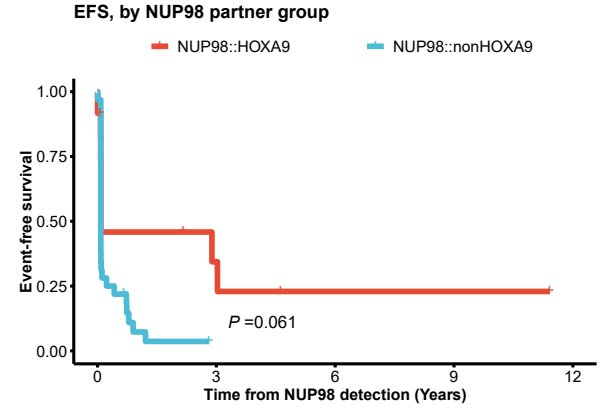
Number at risk					
12	3	1	1	0	
18	0	0	0	0	
14	0	0	0	0	

D



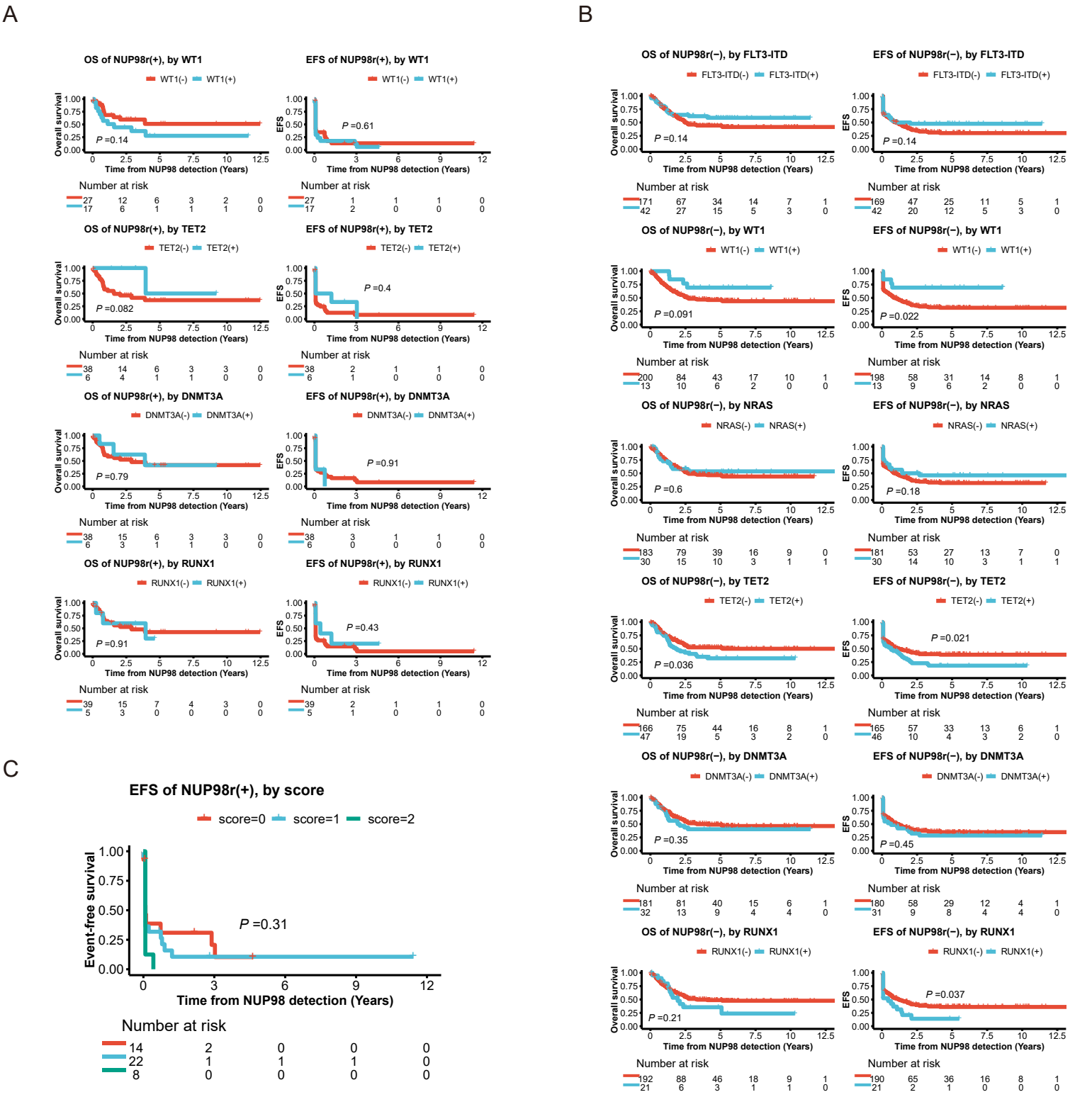
Number at risk					
12	7	3	2	1	0
32	11	4	2	2	0

E



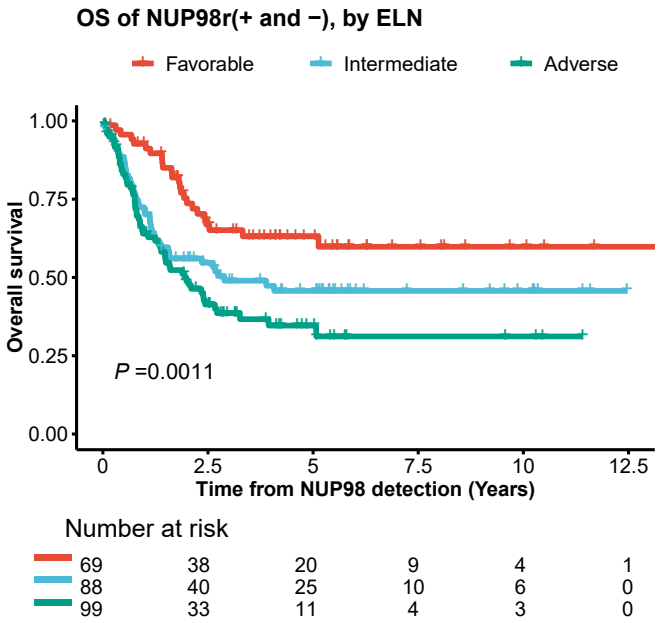
Number at risk					
12	3	1	1	0	
32	0	0	0	0	

Supplementary Figure 4

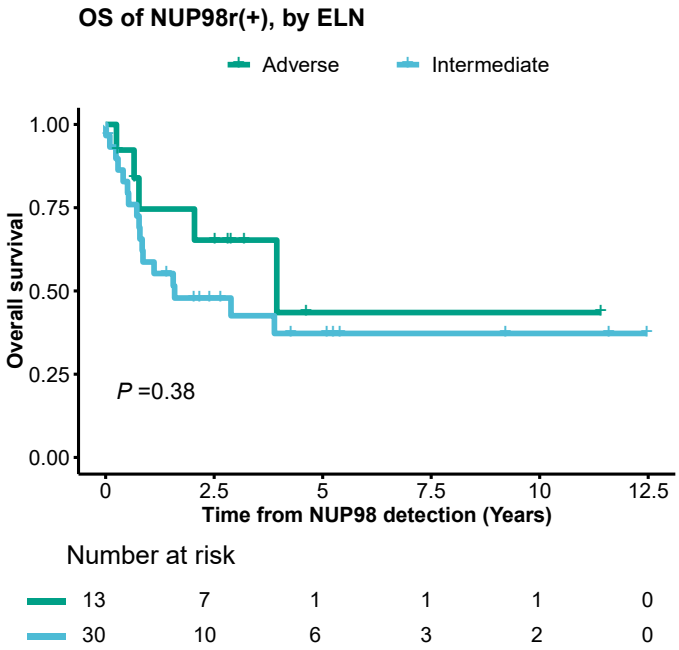


Supplementary Figure 5

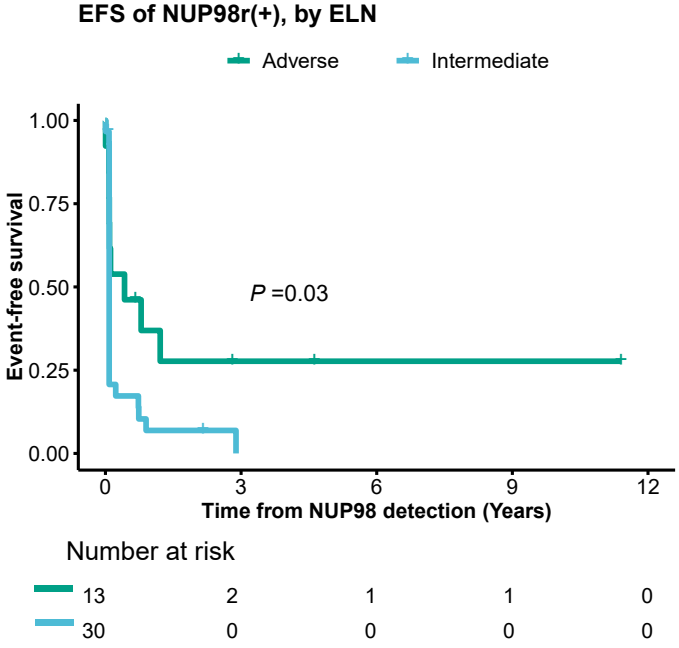
A



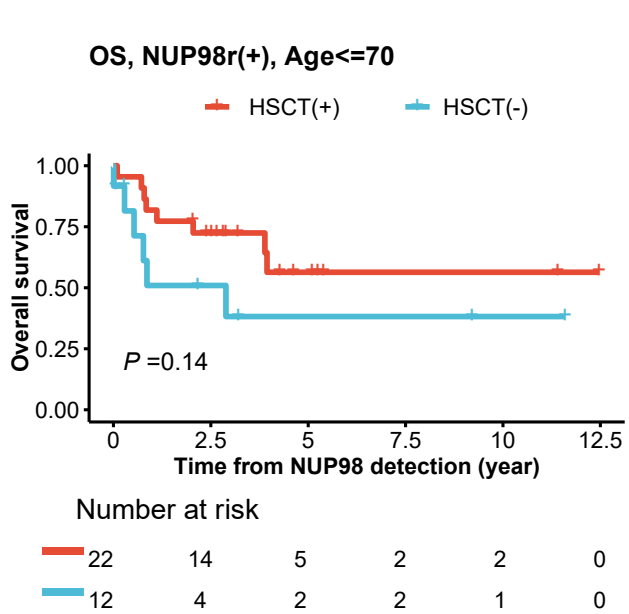
B



C



D



E