Unified classification and risk-stratification in Acute Myeloid Leukemia

Yanis Tazi[1,2,3,4], Juan E. Arango-Ossa[1,2†], Yangyu Zhou[1,2†], Elsa Bernard PhD[1,2], Ian Thomas[5], Amanda Gilkes[6], Sylvie Freeman MD. DPhil[7], Yoann Pradat[1], Sean J Johnson[5], Robert Hills DPhil[8], Richard Dillon PhD[9], Max F Levine[1], Daniel Leongamornlert PhD[10], Adam Butler[10], Arnold Ganser MD[11], Lars Bullinger MD[12], Konstanze Döhner MD[13], Oliver Ottmann MD[6], Richard Adams MD[5], Hartmut Döhner MD[13], Peter J Campbell MD PhD[10], Alan K Burnett MD[14], Michael Dennis MD[15], Nigel H Russell MD*[16], Sean M. Devlin PhD*[1], Brian J P Huntly MD. PhD*[17], and Elli Papaemmanuil PhD*[1,2]

[†]Shared second authorship
*Shared senior authorship

Address for correspondence: papaemme@mskcc.org

Affiliations:

1. Computational Oncology Service, Department of Epidemiology & Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

2. Center for Hematologic Malignancies, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

3. Tri-Institutional Computational Biology and Medicine PhD Program, Weill Cornell Medicine of Cornell University and Rockefeller University, New York, NY, USA.

4. The Rockefeller University, New York, NY, USA.

5. Centre for Trials Research, School of Medicine, Cardiff University, UK

6. Department of Haematology, School of Medicine, Cardiff University, UK

7. Institute of Immunology and Immunotherapy, University of Birmingham, UK

8. Nuffield Department of Population Health, University of Oxford, Oxford, UK

9. Department of Medical and Molecular Genetics, King's College, London, UK

10. Cancer, Ageing and Somatic Mutation Programme, Wellcome Sanger Institute, Hinxton, UK

11. Department of Hematology, Hemostasis, Oncology, and Stem Cell Transplantation, Hannover Medical School, Hannover, Germany

12. Department of Hematology, Oncology, and Tumorimmunology, Campus Virchow Klinikum, Berlin, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

13. Department of Internal Medicine III, Ulm University, Ulm, Germany

14. Visiting Professor University of Glasgow, formerly Cardiff University, UK

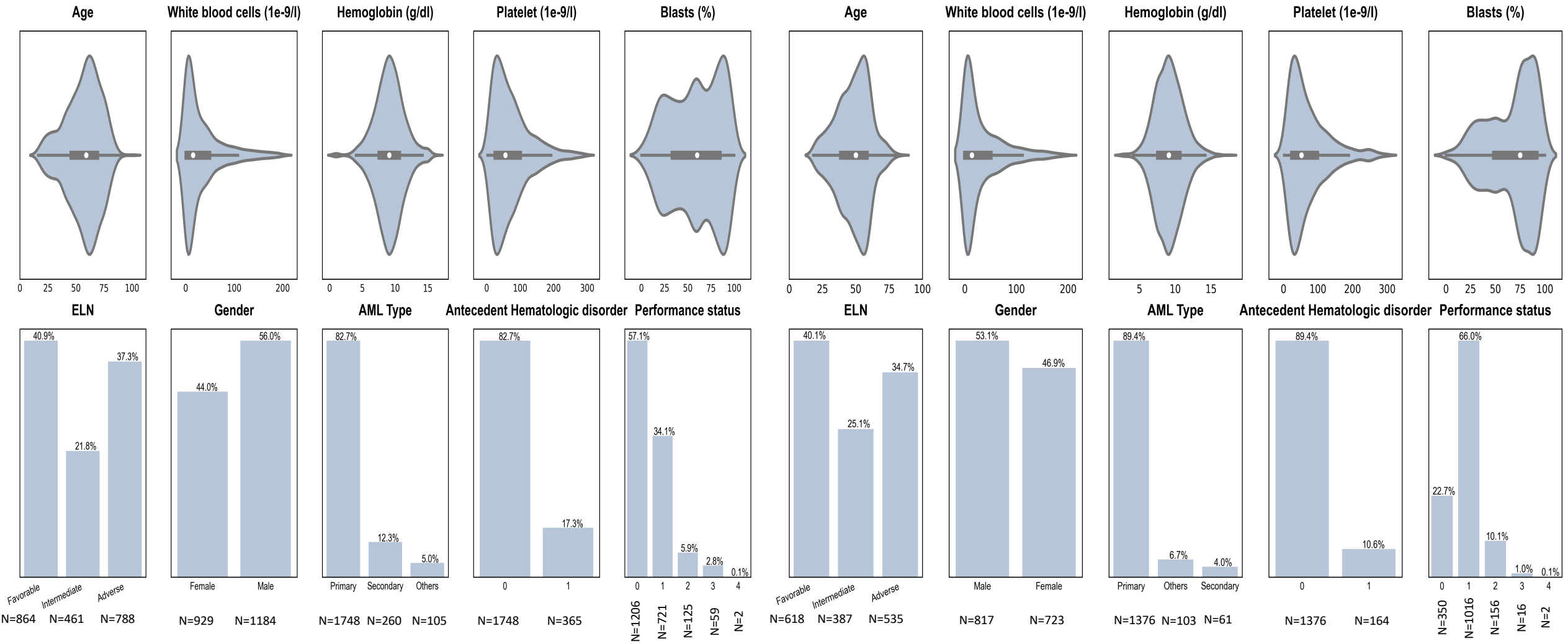
15. The Christie NHS Foundation Trust, Manchester, U.K.

16. Department of Haematology, Nottingham University Hospital, Nottingham, UK

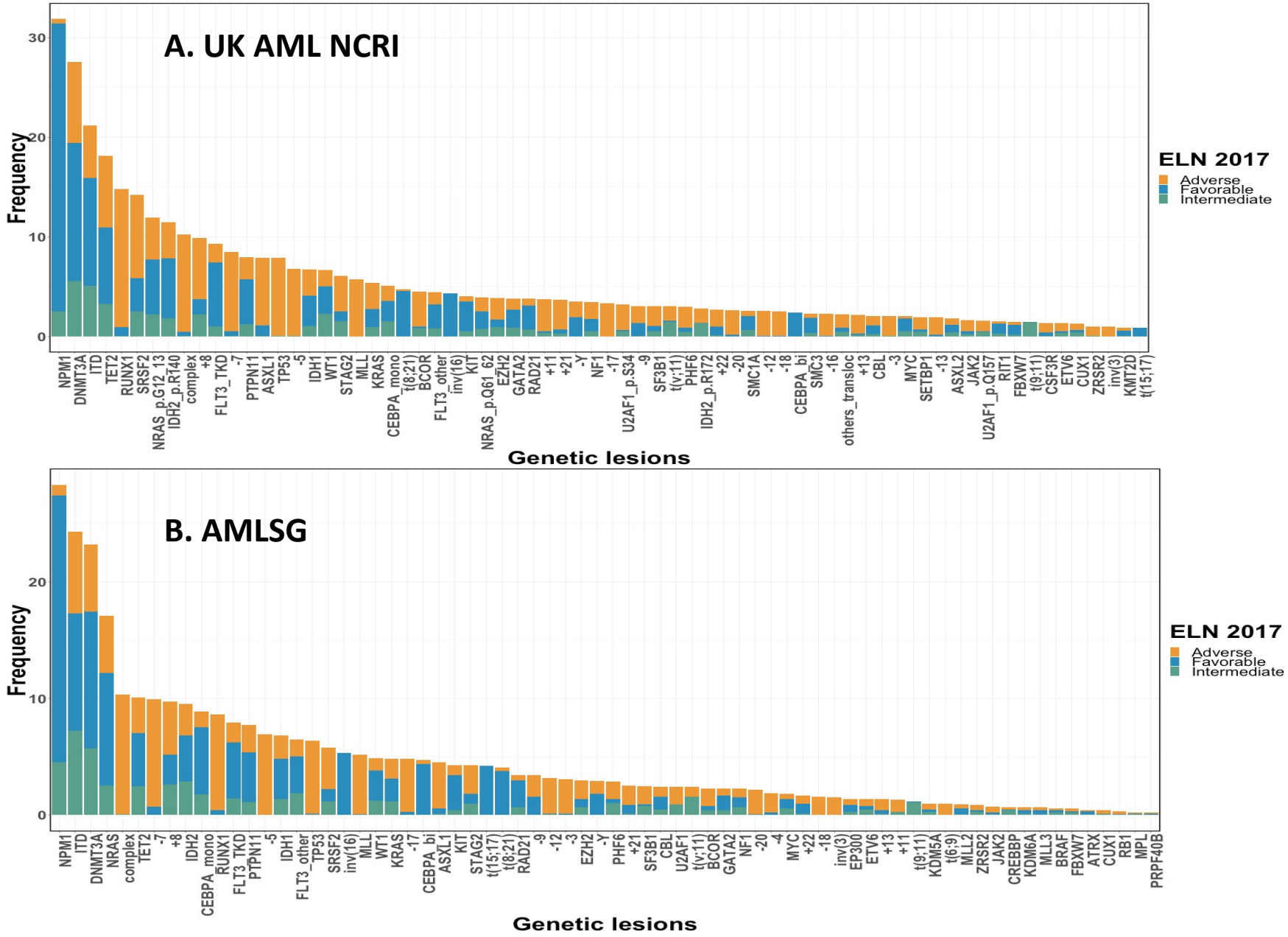
17. Department of Haematology and Wellcome Trust-MRC Cambridge Stem Cell Institute, University of Cambridge, UK

Supplementary Figures

S.Figure 1: Study Cohort. A. Training cohort characteristics for UK AML NCRI Trial study set (n=2,113). B. Validation cohort characteristics for AML SG study set (n=1,540). In all boxplots, the median is indicated by the white dot and the first and third quartiles by the box edges. The lower and upper whiskers extend from the hinges to the smallest and largest values, respectively, no further than 1.5× interquartile range from the hinges.

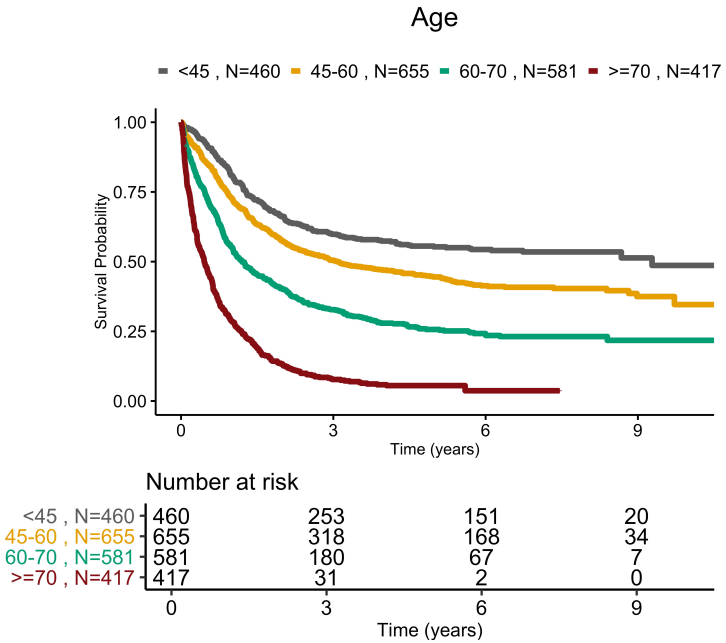


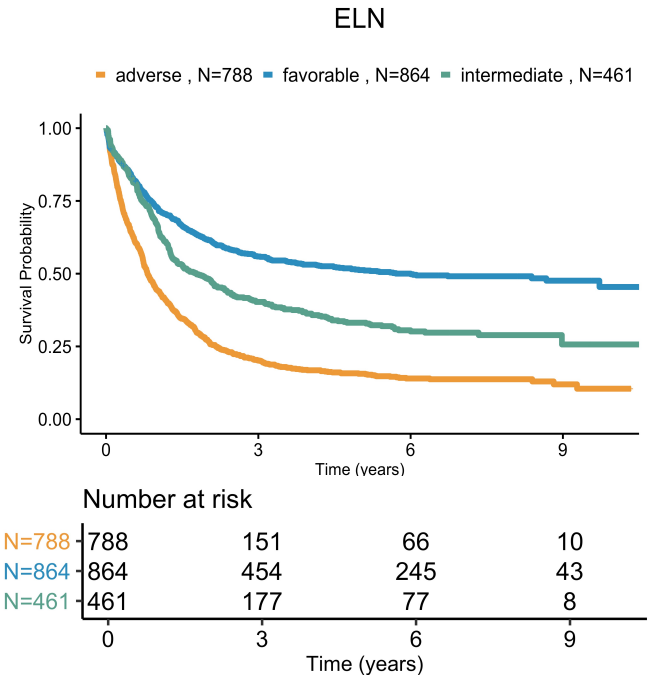
S.Figure 2: Molecular characteristics. Distribution of gene mutations and cytogenetic abnormalities in A. AML NCRI trial study set (2,113 patients and 8,460 driver events) and B. AML SG study set (1,540 patients and 5,043 driver events).



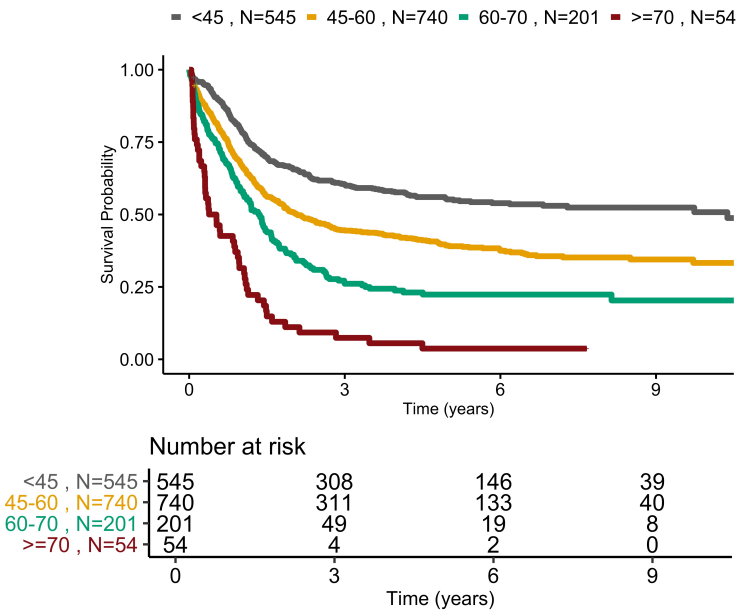
S.Figure 3: Age and ELN^{2017} relationships to survival. Kaplan-Meier curves for overall survival by age and ELN^{2017} stratification in A. NCRI trial study set (n=2,113) and B. AMLSG study set (n=1,540).

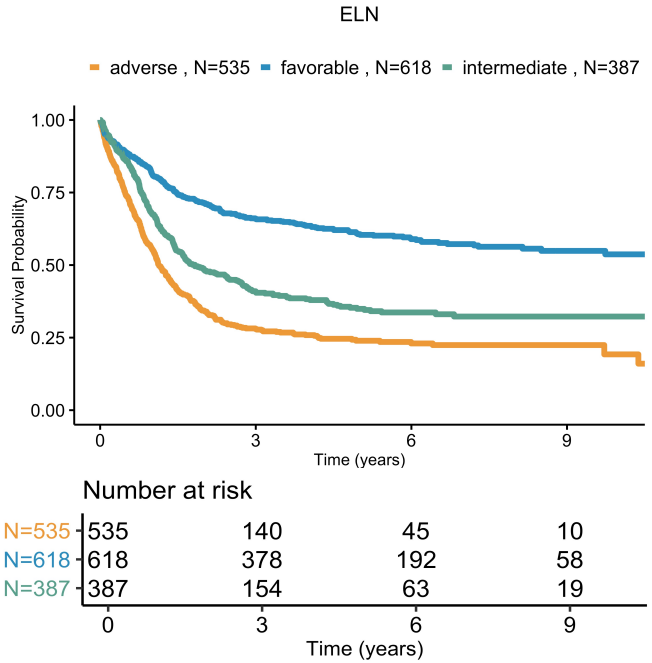
A. UK AML NCRI





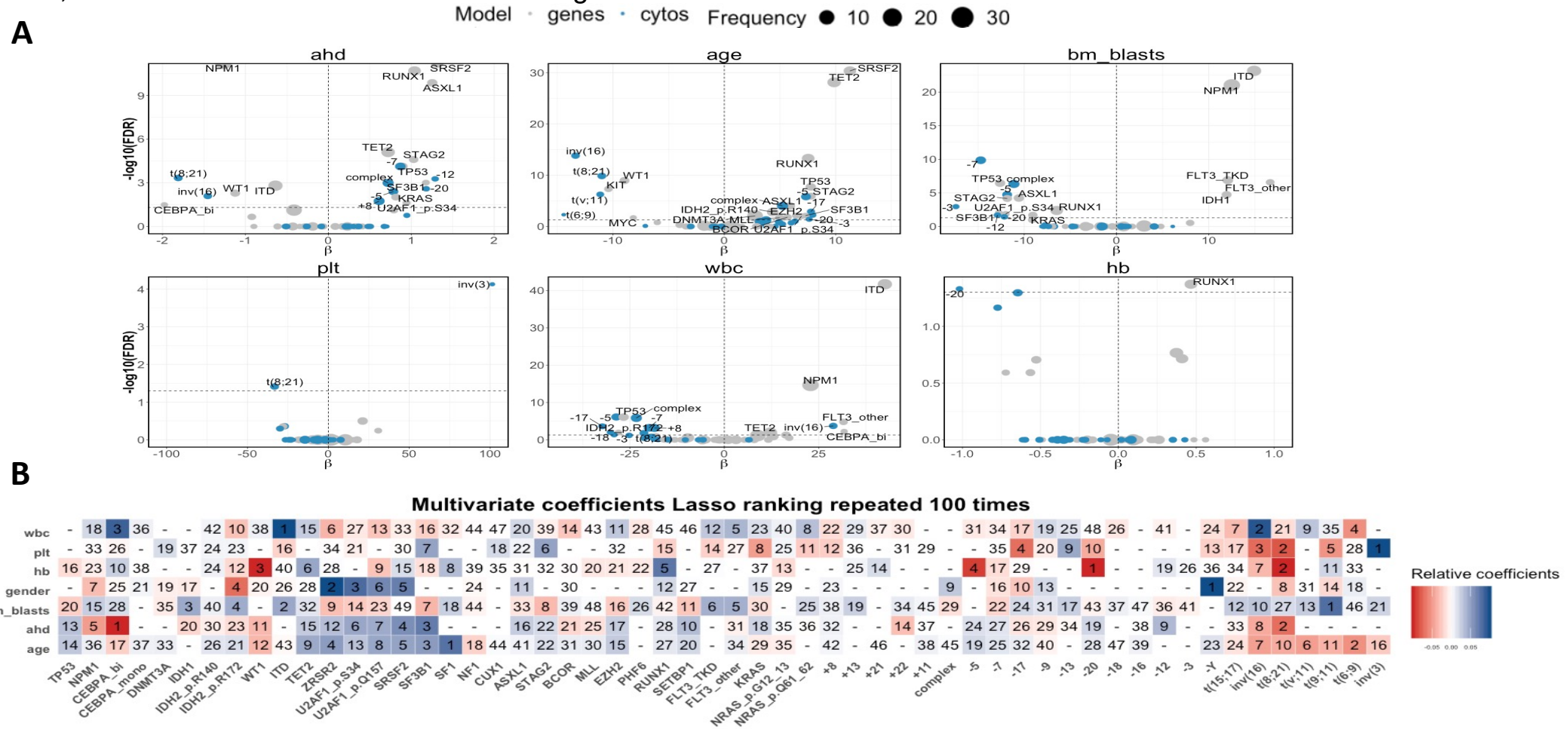
B. AMLSG



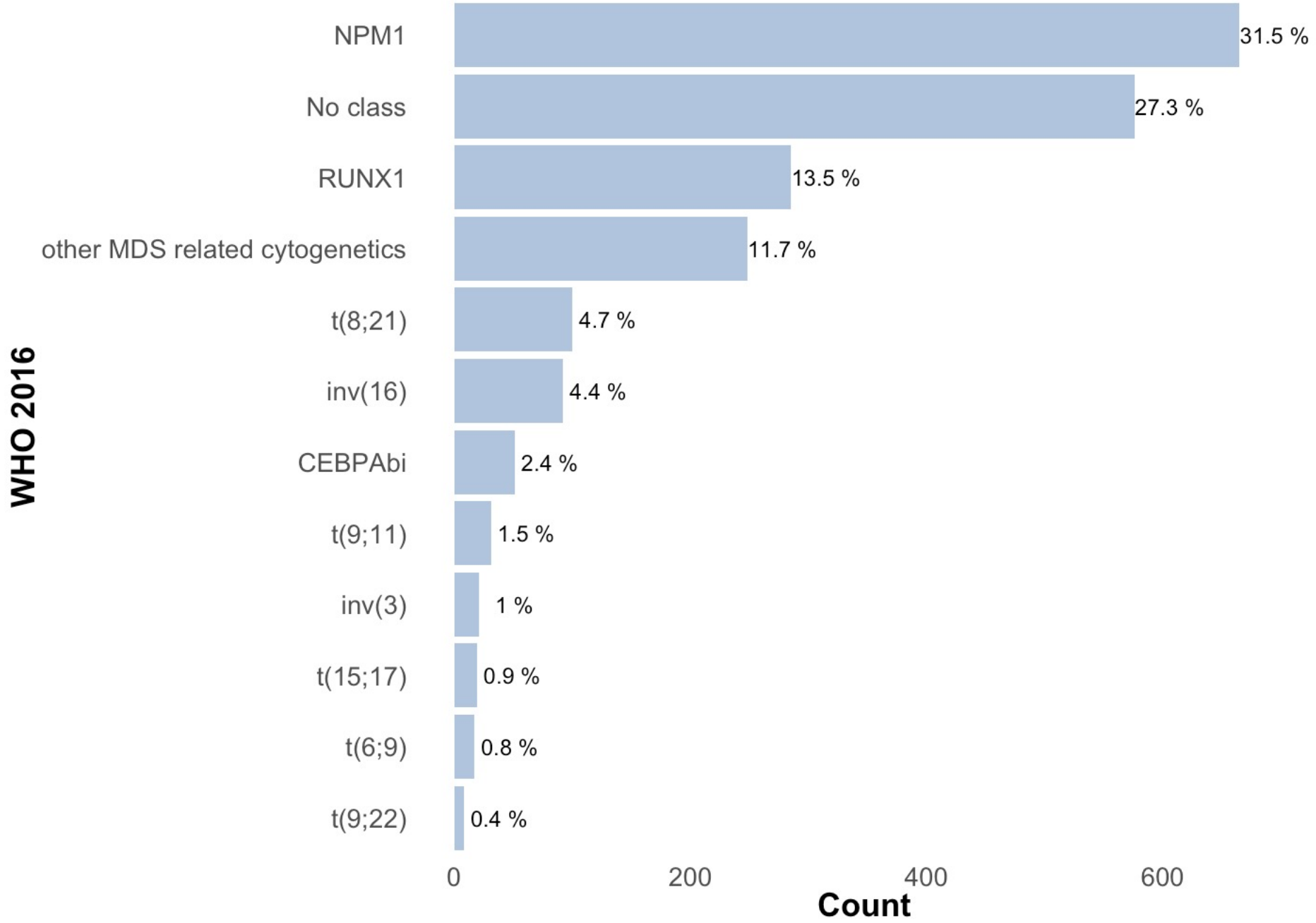


S.Figure 4: Molecular predictors of clinical correlates. A. Volcano plots for univariate regression analysis for each clinical parameter (Antecedent hematologic disorder, Age, Bone marrow blasts, Platelets, White blood cell counts and hemoglobin) in NCRI trial study set (n= 2,113). Genes are colour coded in blue, whilst cytogenetic events in grey. The size of each point corresponds to the frequency of the event. The horizontal dotted curve corresponds to the p-value threshold of 0.05 (on the y-axis) and the vertical one corresponds to β=0 (x axis). Significant predictors are highlighted (p-value less than 0.05).

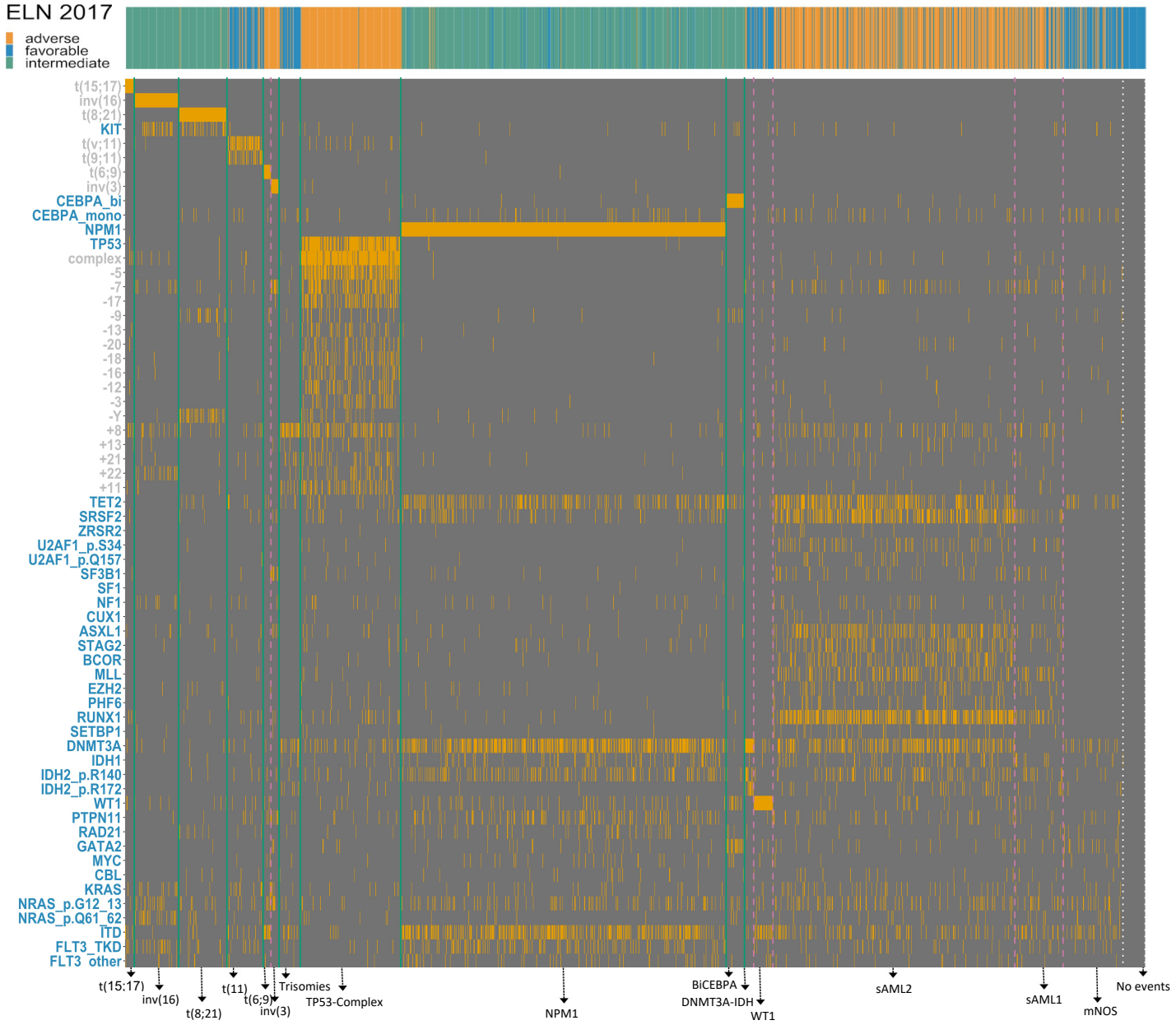
B. Multivariate regression ranking of clinical correlates in NCRI trial study set (n= 2,113). The numbers correspond to the ranking of the absolute value of the β coefficients while the colour intensities represent the absolute coefficient values ranking (red with high intensity for largest negative value and blue with high intensity for largest positive value). The regression process is explained in S.Appendix. Abbreviations: ahd refers to antecedent hematologic disorder, plt to platelet, wbc to white blood cells and hb to hemoglobin.



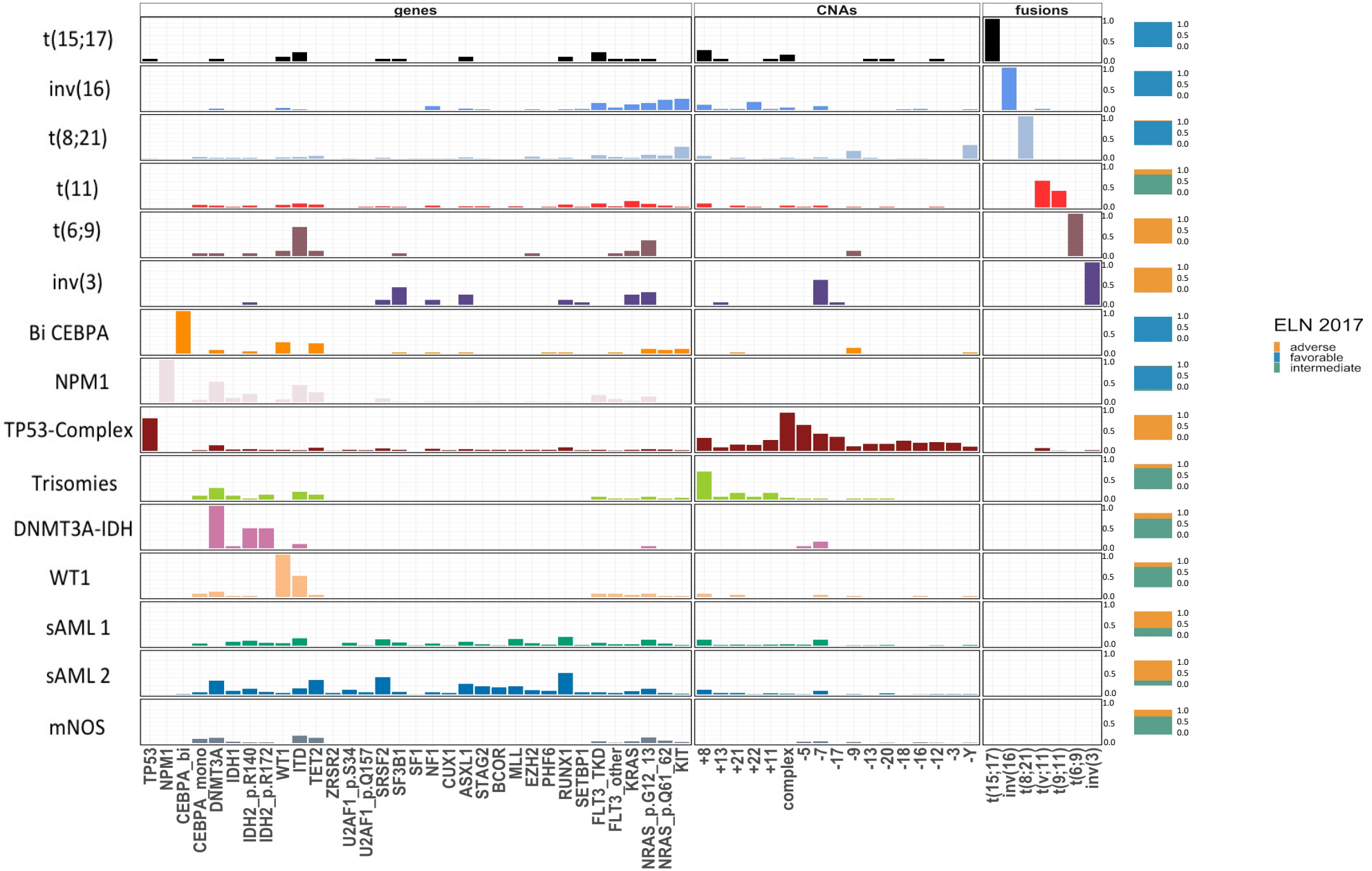
S.Figure 5: Repartition of AML NCRI Cohort (n= 2,113) per WHO 2016 guidelines to include the provisional categories defined by *RUNX1* and t(9;22).



S.Figure 6: Heatmap of Bayesian Dirichlet Process clusters. Rows demarcate distinct genetic lesions, columns represent individual samples. Orange lines in heatmap indicate the presence of a specific genetic lesion. Green and pink dotted lines demarcate major clusters and sub-clusters as derived by the first and second iteration of the Bayesian Dirichlet Process. Dotted white lines demarcate patients in the molecularly not otherwise specified (mNOS) class and the no recurrent molecular findings. The top sidebar indicates ELN^{2017} risk score for each patient (blue for favourable, green for intermediate and orange for adverse). For more details on the class assignment process, refer to S.Appendix.

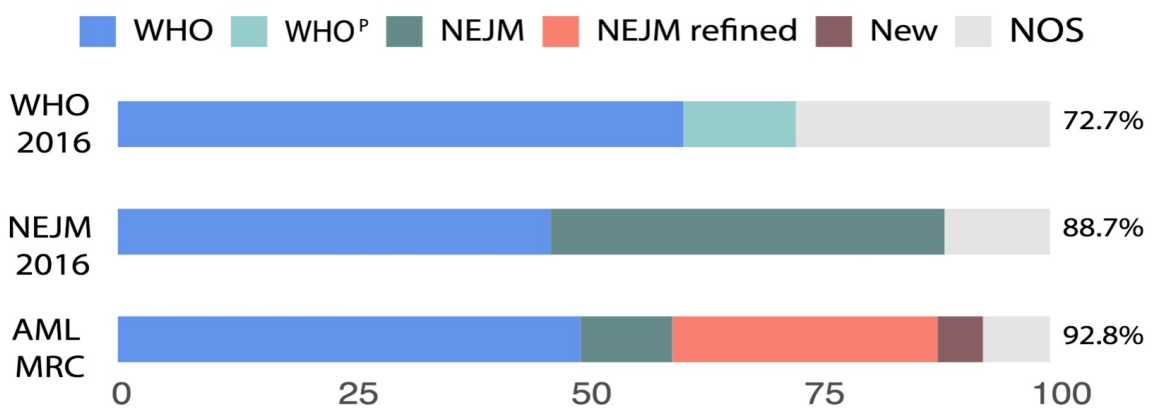


S.Figure 7: Genetic landscape of AML classes. Co-mutation bar plots for acquired gene mutations, chromosomal aneuploidies and fusion genes for each class. The y-axis represents the fraction of patients carrying the driver event (on the x-axis) within each class. For each class vertical sidebar denoting distribution of patients in class across the three (ELN[2017]) risk strata (blue for favorable, green for intermediate and orange for adverse).

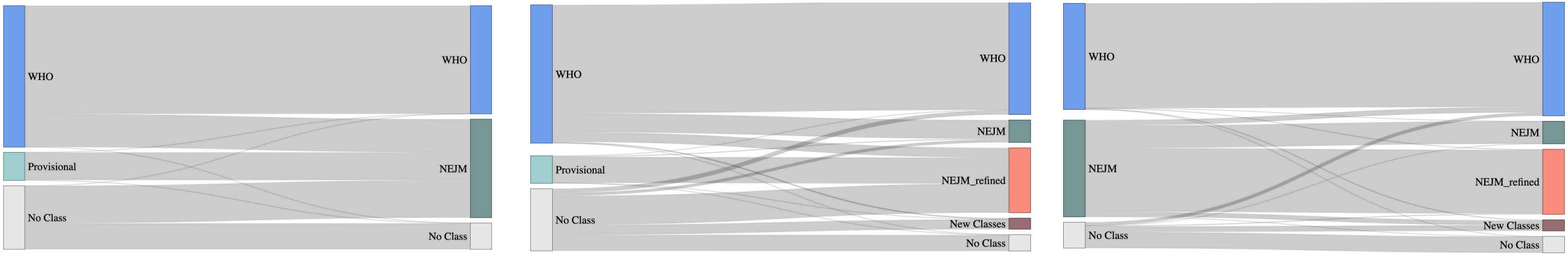


S.Figure 8: Molecular classifications in AML. A. Representation of the AML NCRI cohort (n=2,113 patients) using the WHO classification, the New England Journal of Medicine 2016 classification and the proposed molecular classification in this study. B. Sankey Plots comparing those different classifications. NOS means not otherwise specified.

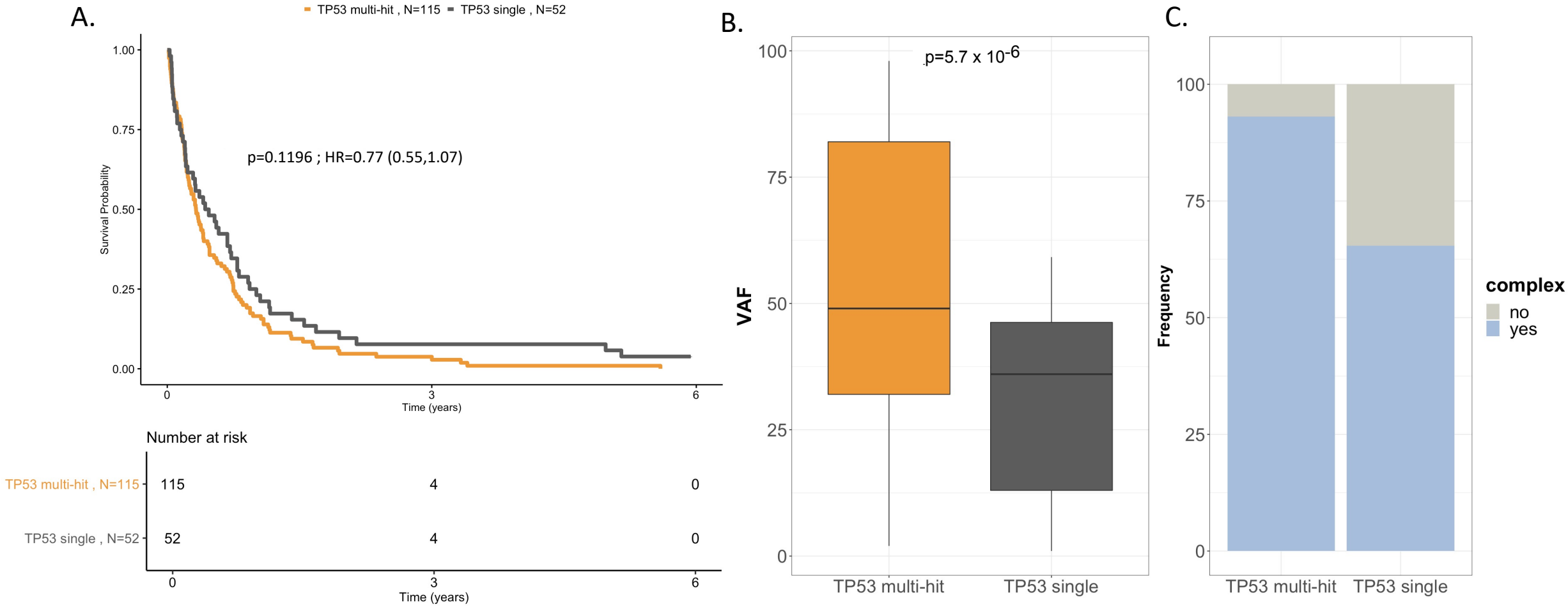
A.



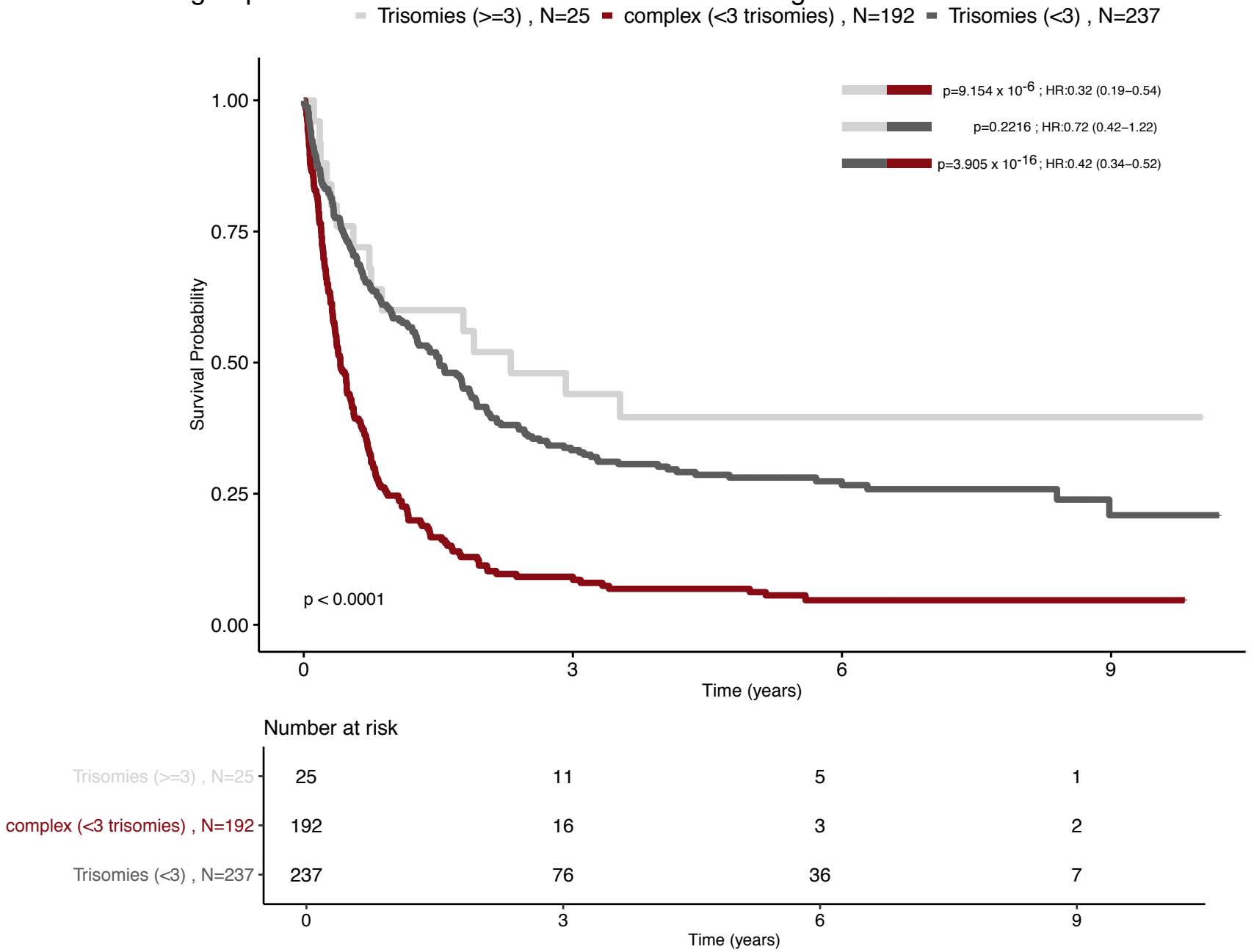
B.



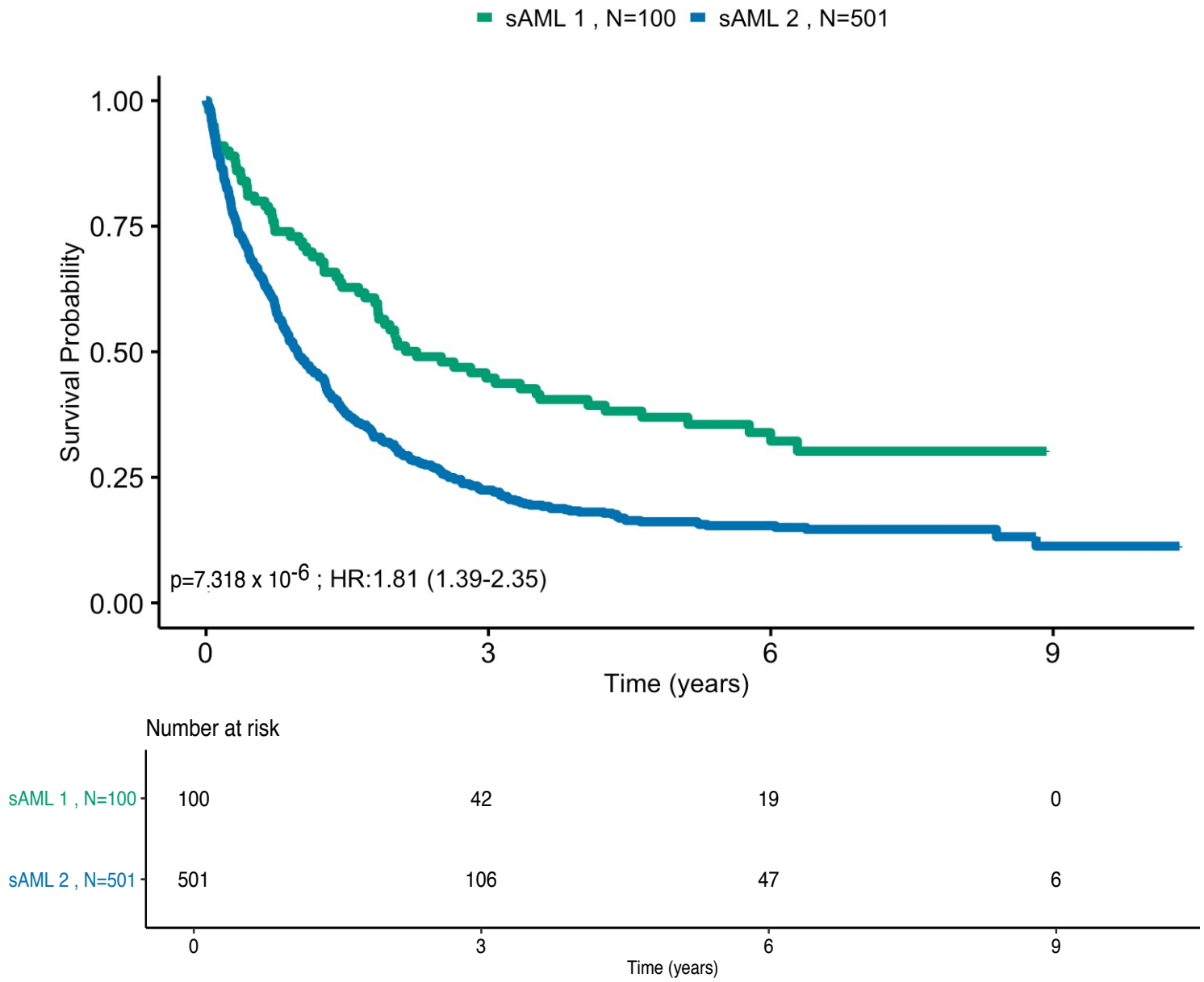
S.Figure 9: A. Kaplan-Meier Curves for OS and associated risk table comparing TP53 single and multi hit in the AML NCRI cohort (N=2,113). Log-rank tests compared the survival distributions between the 2 subgroups. Annotated P values are from two-sided log-rank tests. B. Comparison of variant allele frequency distribution for TP53 single and multi-hit in the AML NCRI cohort (N=2,113). Pvalue was computed with a two-sided Wilcoxon rank-sum test. C. Comparison of frequency of complex karyotype patients for TP53 single and multi-hit in the AML NCRI cohort
(N=2,113). In all boxplots, the median is indicated by the horizontal line and the first and third quartiles by the box edges. The lower and upper whiskers extend from the hinges to the smallest and largest values, respectively, no further than 1.5× interquartile range from the hinges.



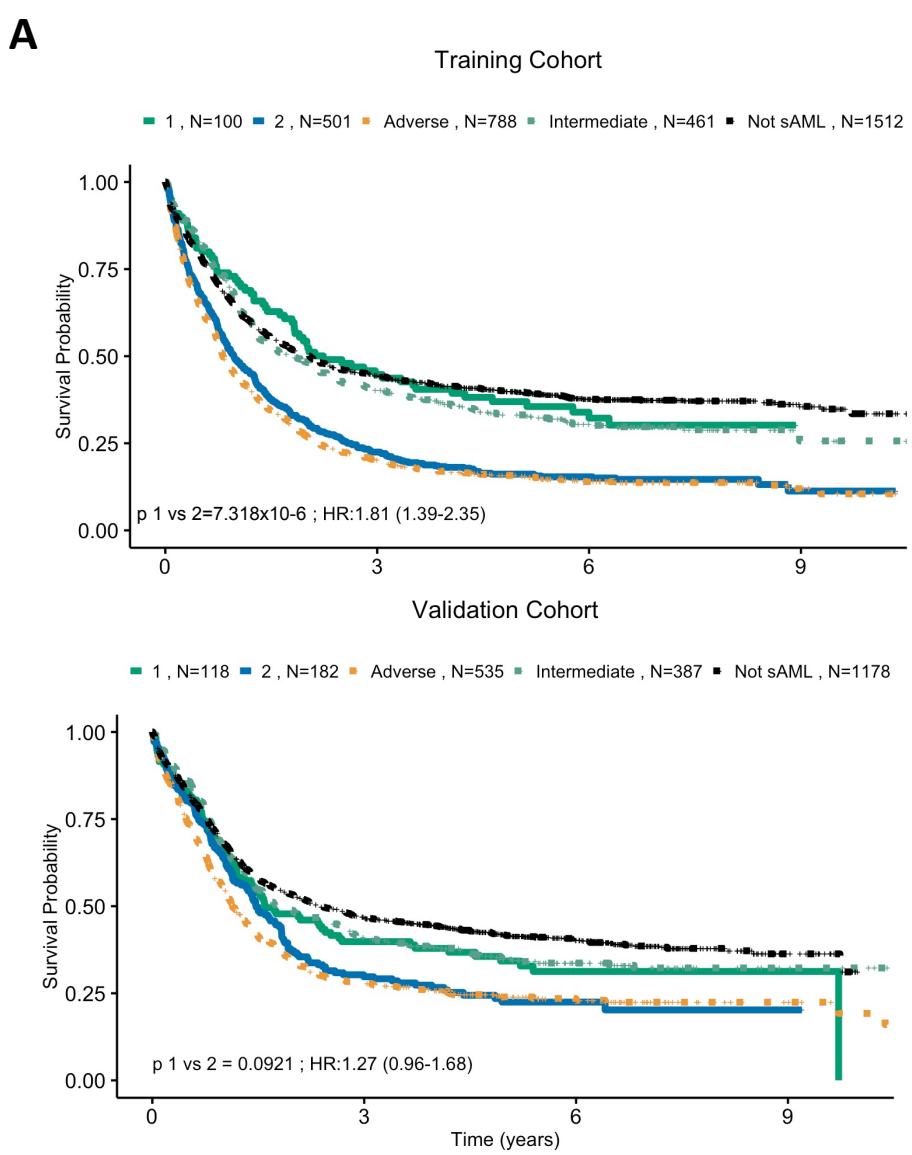
S.Figure 10: Kaplan-Meier and associated risk table for overall survival curves for patients with trisomies (<3) (grey), trisomies (>=3) (lightgrey) and complex karyotype (burgundy) in the AML NCRI cohort (N=2,113). Log-rank tests compared the survival distributions between the 3 possible combinations of 2 subgroups. Annotated P values are from two-sided log-rank tests.

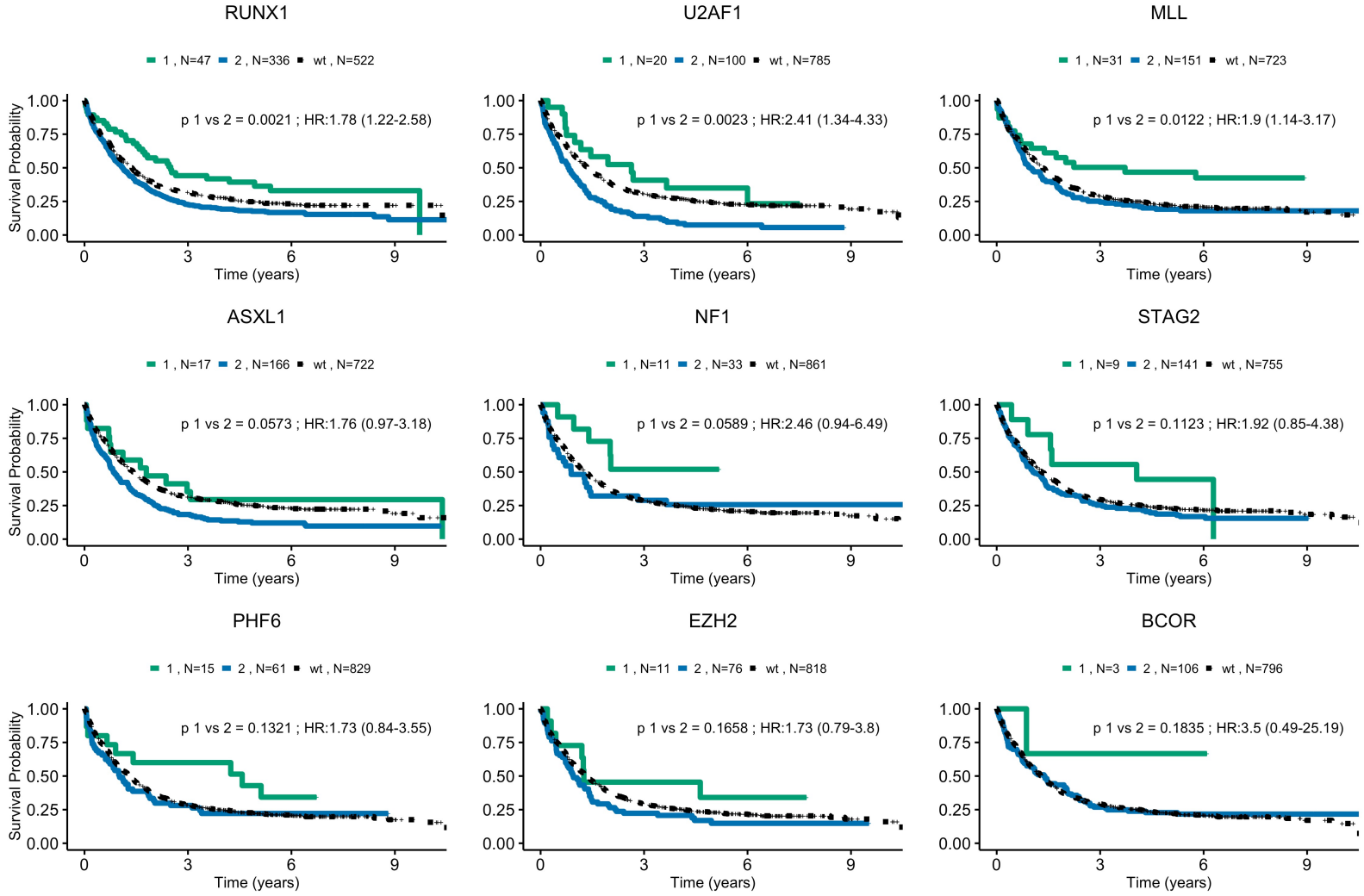


S.Figure 11: Kaplan-Meier and associated risk table for overall survival curves for the 2 secondary AML like classes in the AML NCRI cohort (N=2,113). Log-rank tests compared the survival distributions between the 2 subgroups. Annotated P values are from two-sided log-rank tests.

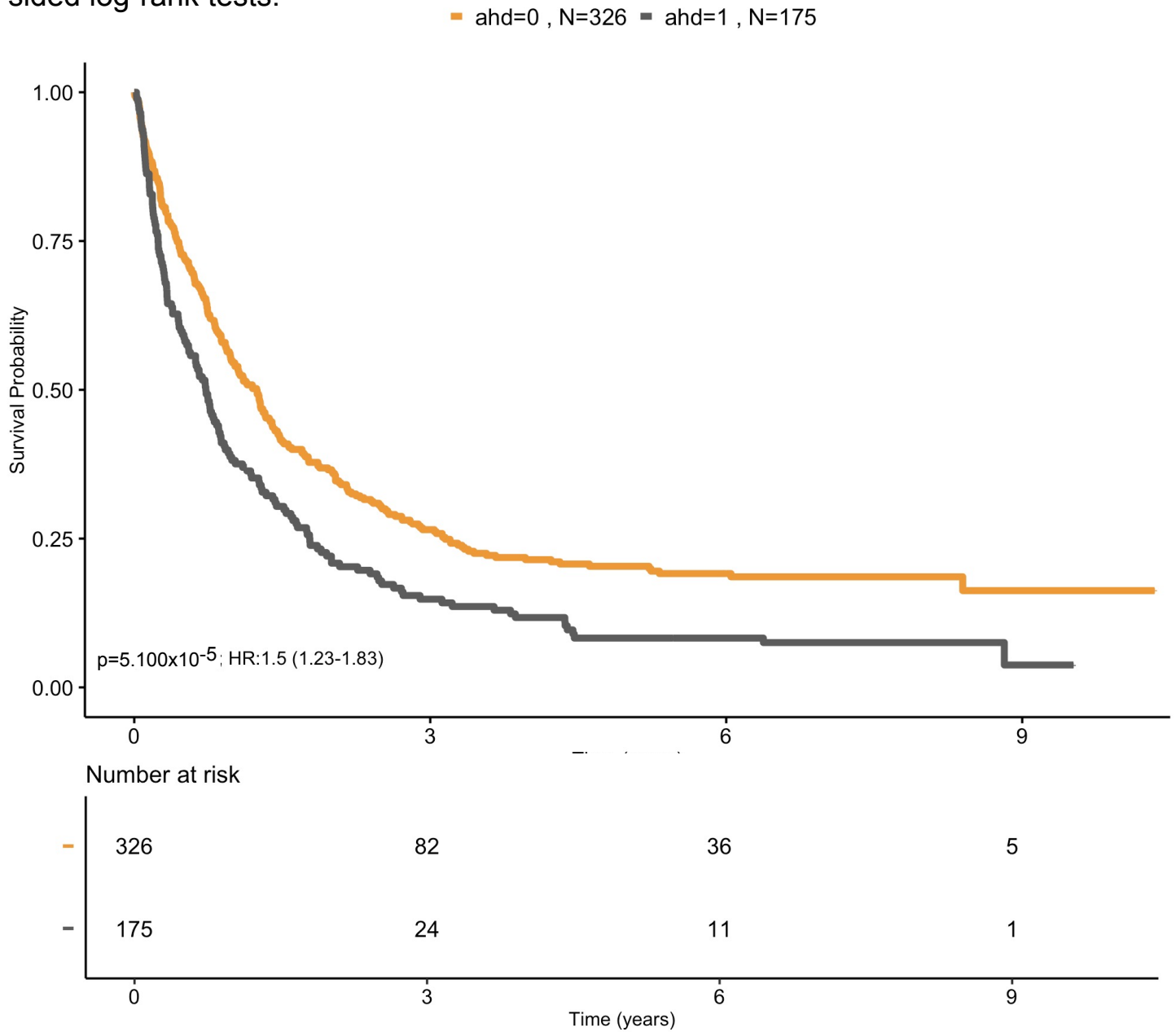


S.Figure 12: Prognostic relevance of mutation number in sAML Like subgroups. A. Kaplan-Meier Curves for overall survival comparing AML class with sAML1 (green plain line) and sAML2 (blue plain line) mutations to ELN^{2017} risk groups in AML NCRI cohort (training, n=2,113) and AMLSG (validation cohort, n=1,540). B. Kaplan- Meier curves for overall survival comparing outcomes associated with specific gene mutations in the context of sAML1 and sAML2 subgroup in AML NCRI cohort (n=2,113). Log-rank tests compared the survival distributions between the 2 subgroups. Annotated P values are from two-sided log-rank tests.

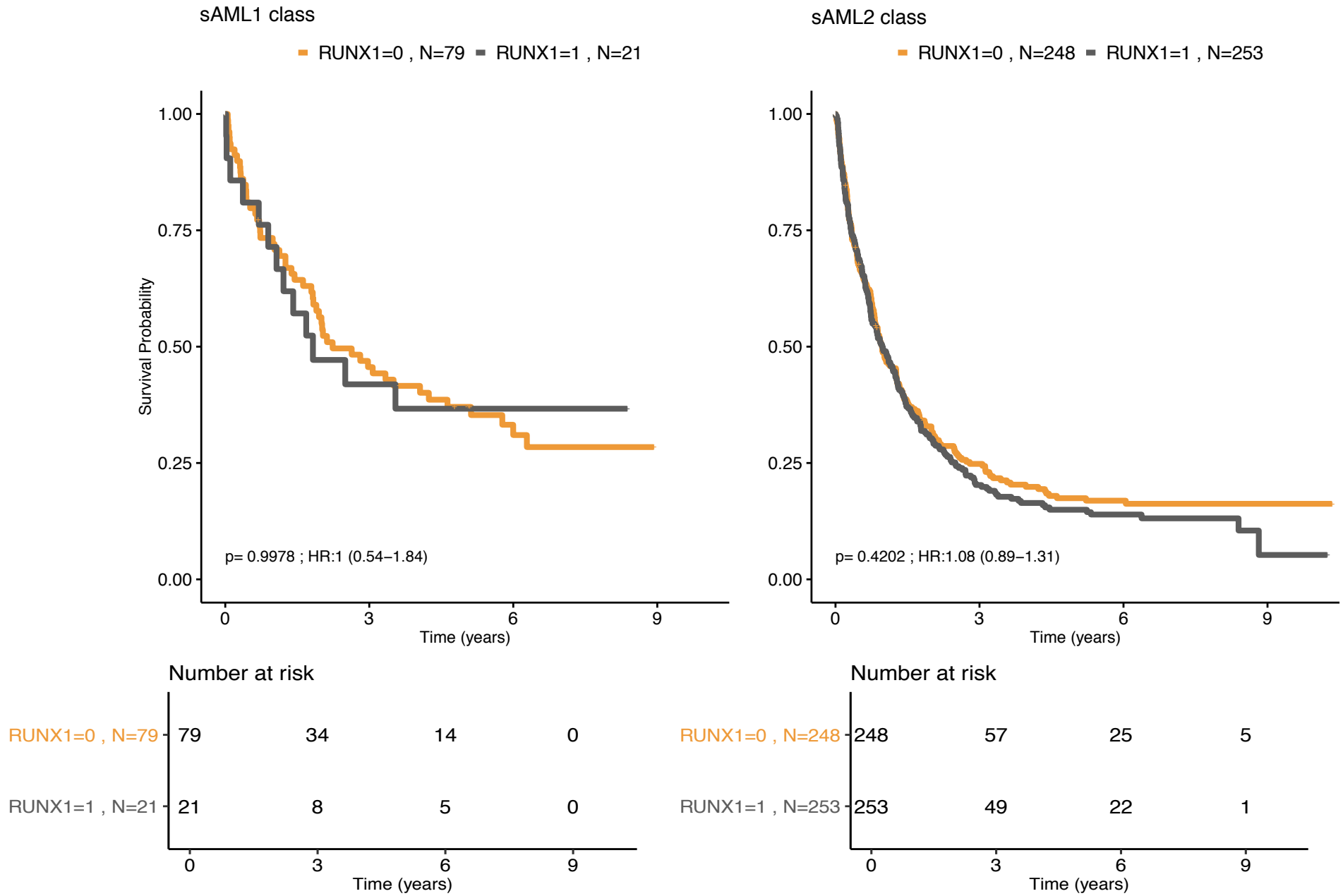




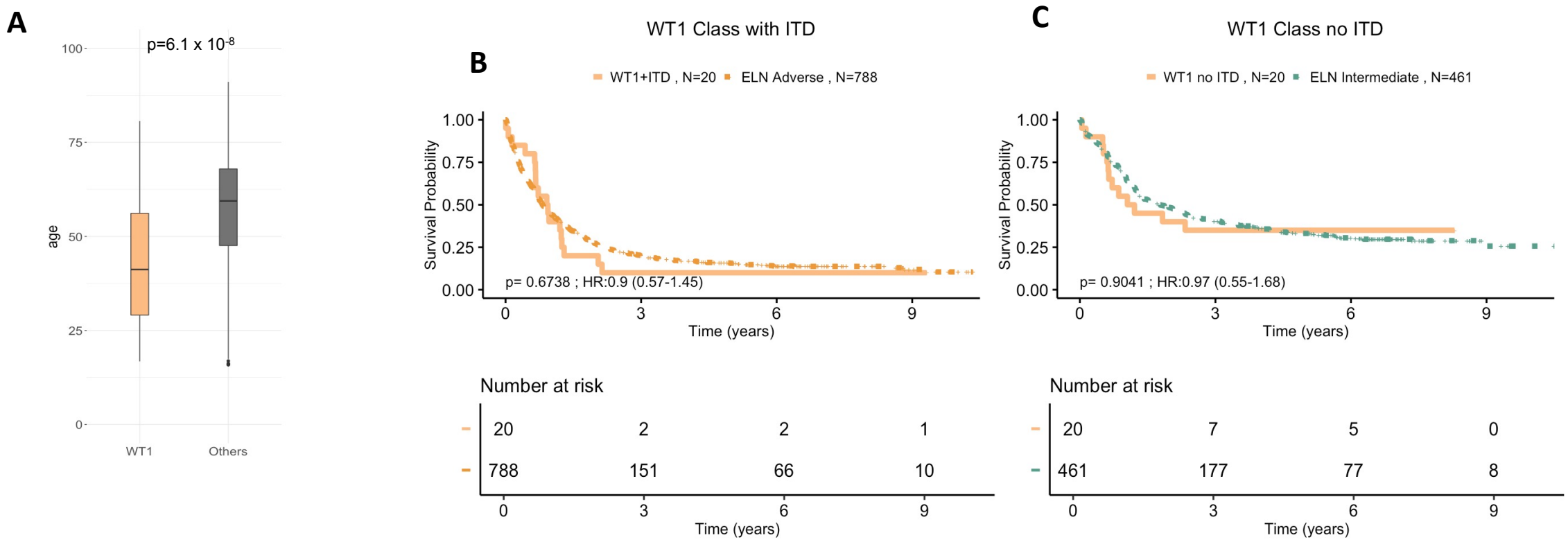
S.Figure 13: Prognostic relevance of AHD in sAML2. Kaplan-Meier curves for overall survival and associated risk table comparing outcomes in sAML2 patients (training, n=501) on the basis of AHD status. ahd=0 corresponds to patients without antecedent hematologic disorder while ahd=1 corresponds to patient with antecedent hematologic disorder. Log-rank tests compared the survival distributions between the 2 subgroups. Annotated P values are from two-sided log-rank tests.



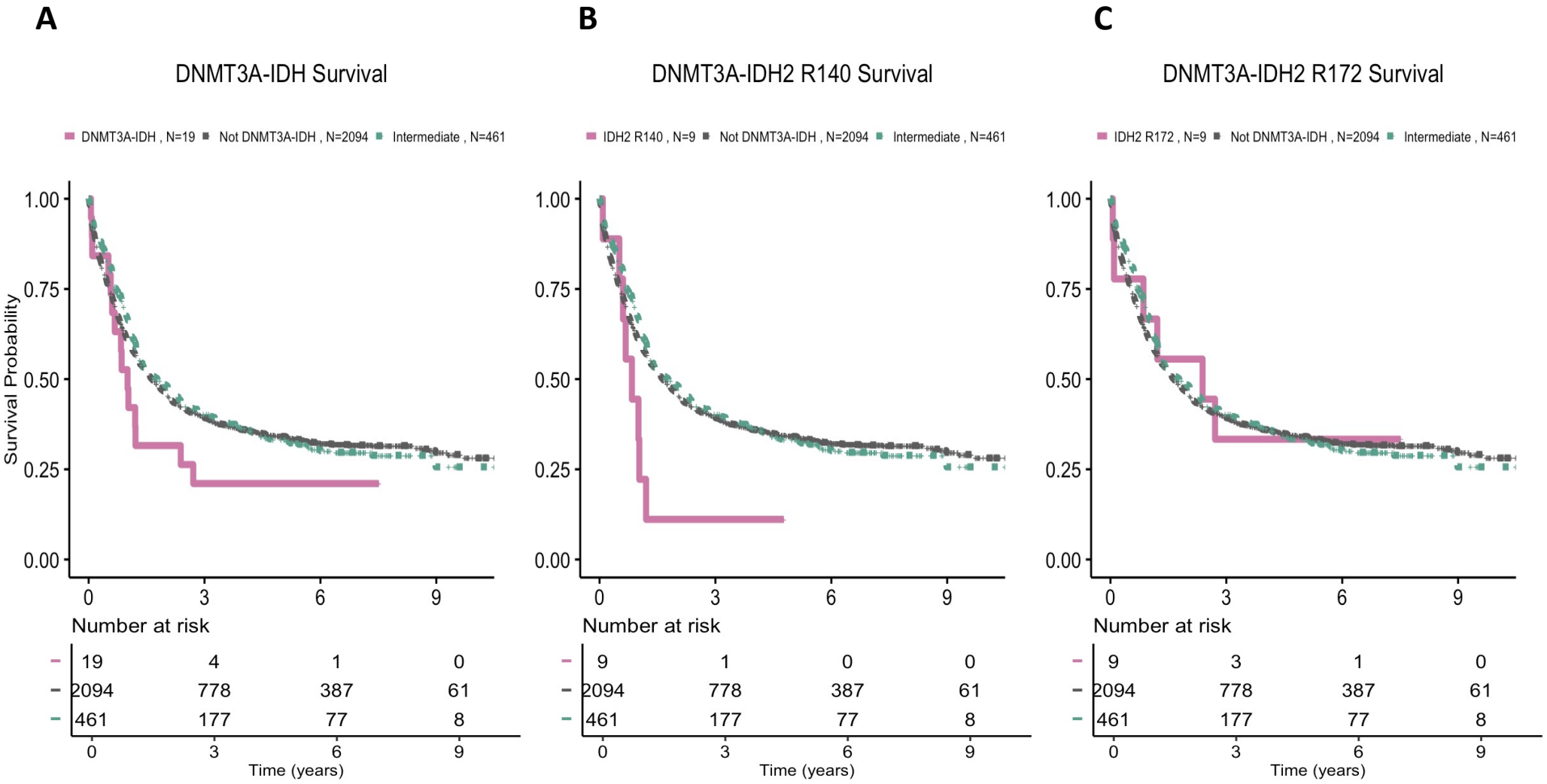
S.Figure 14: Kaplan-Meier and associated risk table for RUNX1 mutation for the 2 secondary AML like classes in the AML NCRI cohort (N=2,113). Log-rank tests compared the survival distributions between the 2 subgroups. Annotated P values are from two-sided log-rank tests. Annotated P values are from two-sided log-rank tests.



S.Figure 15: WT1 Class. A. Distribution of age for patients in WT1 class (N=40) compared to otherAML (N=2,113). Pvalue was computed with a two-sided Wilcoxon rank-sum test. In all boxplots, the median is indicated by the white dot and the first and third quartiles by the box edges. The lower and upper whiskers extend from the hinges to the smallest and largest values, respectively, no further than 1.5× interquartile range from the hinges. B. Kaplan-Meier Curves for overall survival and associated risk table comparing WT1 class with $FLT3^{ITD}$ to ELN^{2017} adverse risk group inAML NCRI cohort (n=2,113). Log-rank tests compared the survival distributions between the 2 subgroups. C. Kaplan-Meier Curves for overall survival and associated risk table for patients in WT1 class without $FLT3^{ITD}$ to ELN^{2017} intermediate risk group inAML NCRI cohort (n=2,113). Log-rank tests compared the survival distributions between the 2 subgroups. Annotated P values are from two-sided log-rank tests.

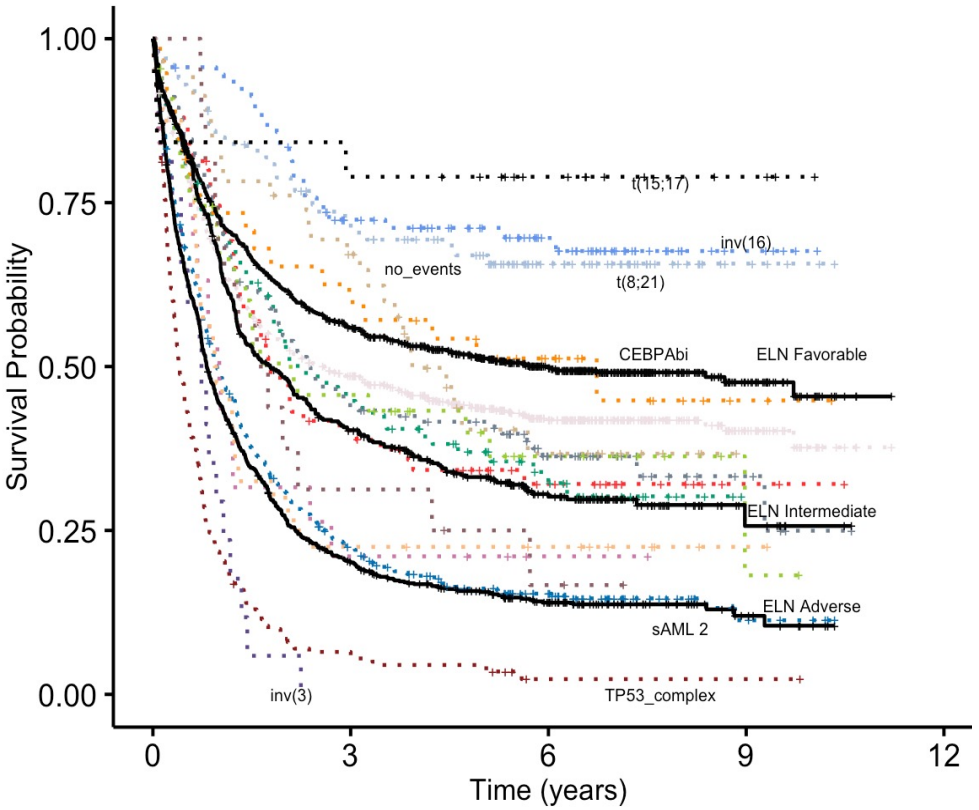


S.Figure 16: DNMT3A-IDH class. Kaplan-Meier curves for overall survival and associated risk tables comparing
A. AML class DNMT3A IDH to ELN^{2017} intermediate risk groups and other AML patients in AML NCRI cohort (training, n=2,113).
B. AML subclass DNMT3A IDH2 ^{R140} to ELN^{2017} intermediate risk groups and other AML patients in AML NCRI cohort (training, n=2,113).
C. AML subclass DNMT3A IDH2 ^{R172} to ELN^{2017} intermediate risk groups and other AML patients in AML NCRI cohort (training, n=2,113).

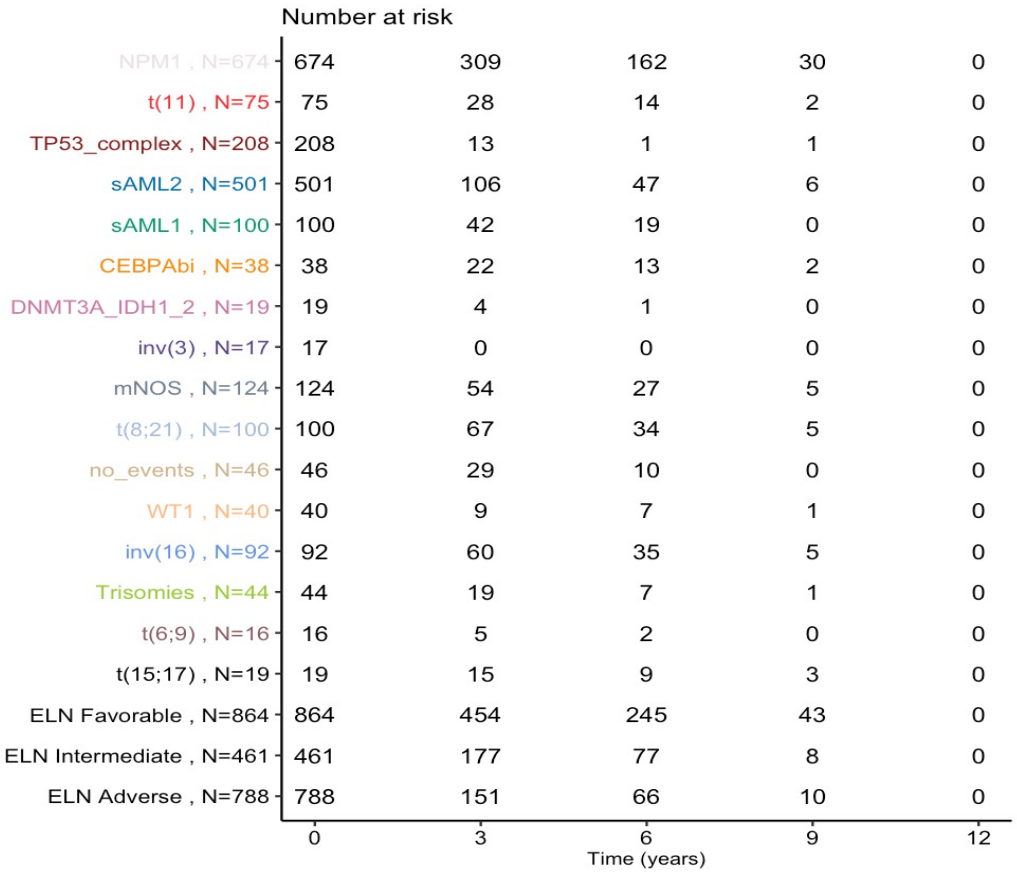


S.Figure 17: Overall survival K-M curves by AML class. A. Kaplan-Meier curves for overall survival comparing all AML classes and ELN^{2017} risk strata in AML NCRI cohort (n=2,113). B. Risk stratification table by AML class and ELN^{2017} score in AML NCRI cohort (n=2,113). C. 10 years 95% confidence interval survival estimates by AML class and ELN^{2017} score in AML NCRI cohort (n=2,113).

A.



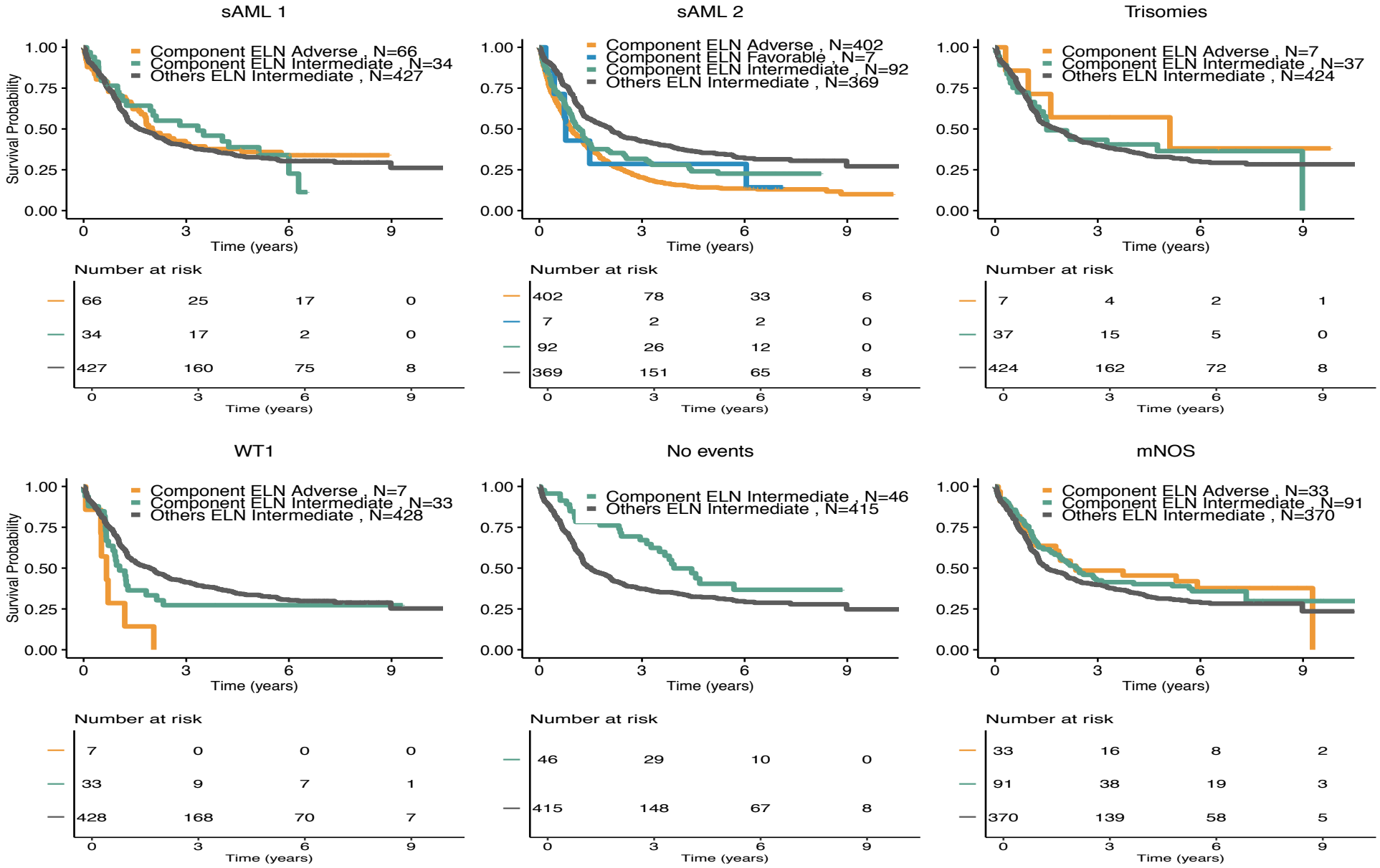
B.



C.

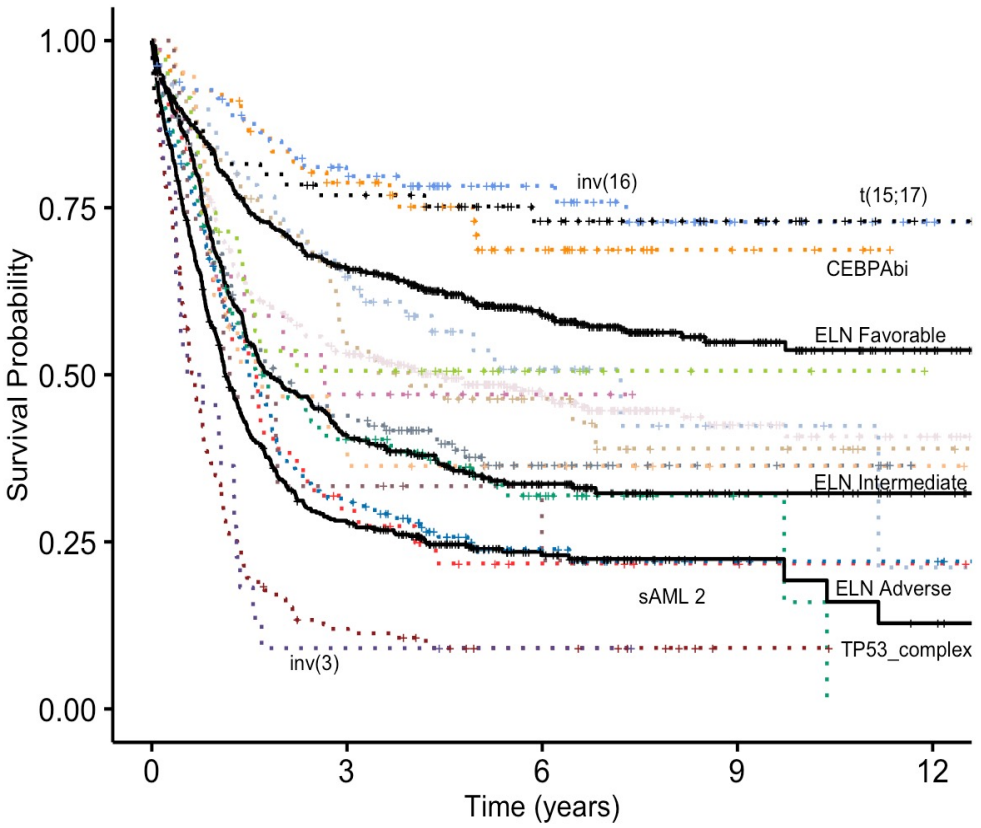
NPM1	0.38 (0.32-0.45)
t(11)	0.32 (0.23-0.45)
TP53-complex	0.02 (0.01-0.07)
sAML2	0.11 (0.07-0.17)
sAML1	0.3 (0.22-0.42)
CEBPAbi	0.45 (0.3-0.68)
DNMT3A-IDH	0.21 (0.09-0.5)
inv(3)	0
mNOS	0.25 (0.13-0.47)
t(8;21)	0.66 (0.57-0.76)
No events	0.37 (0.24-0.56)
WT1	0.22 (0.13-0.4)
inv(16)	0.68 (0.58-0.79)
Trisomies	0.18 (0.04-0.77)
t(6;9)	0.17 (0.05-0.54)
t(15;17)	0.79 (0.63-1)
ELN Favorable	0.45 (0.4-0.51)
ELN Intermediate	0.26 (0.19-0.34)
ELN Adverse	0.1 (0.07-0.15)

S.Figure 18: Kaplan-Meier and associated risk tables for overall survival curves for the sAML 1, sAML 2, trisomies, WT1, no event and mNOS subgroups, separated by ELN^{2017} scores.

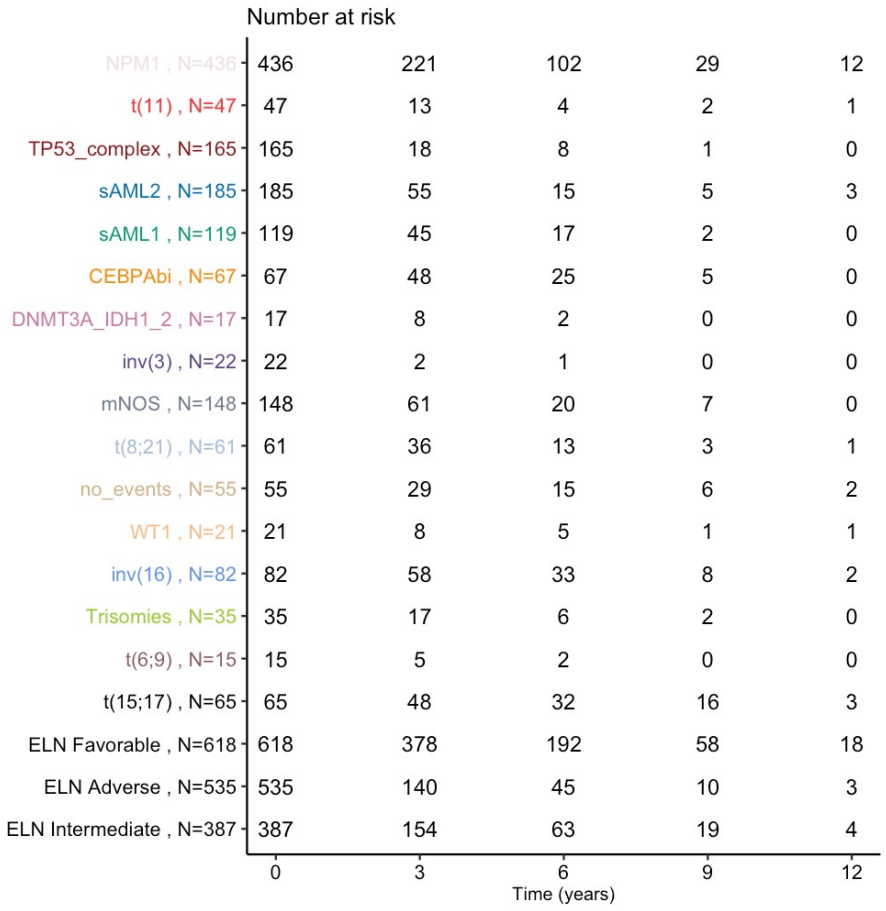


S.Figure 19: Validation of class outcomes in the AML SG Cohort (n=1,540). A. Kaplan-Meier curves for overall survival comparing all AML classes and ELN^{2017} risk strata in AML SG cohort (n=1,540). B. risk stratification table by AML class and ELN^{2017} score in AML SG cohort (n=1,540). C. 10 years 95% confidence interval survival estimates by AML class and ELN^{2017} score in AML SG cohort (n=1,540).

A.



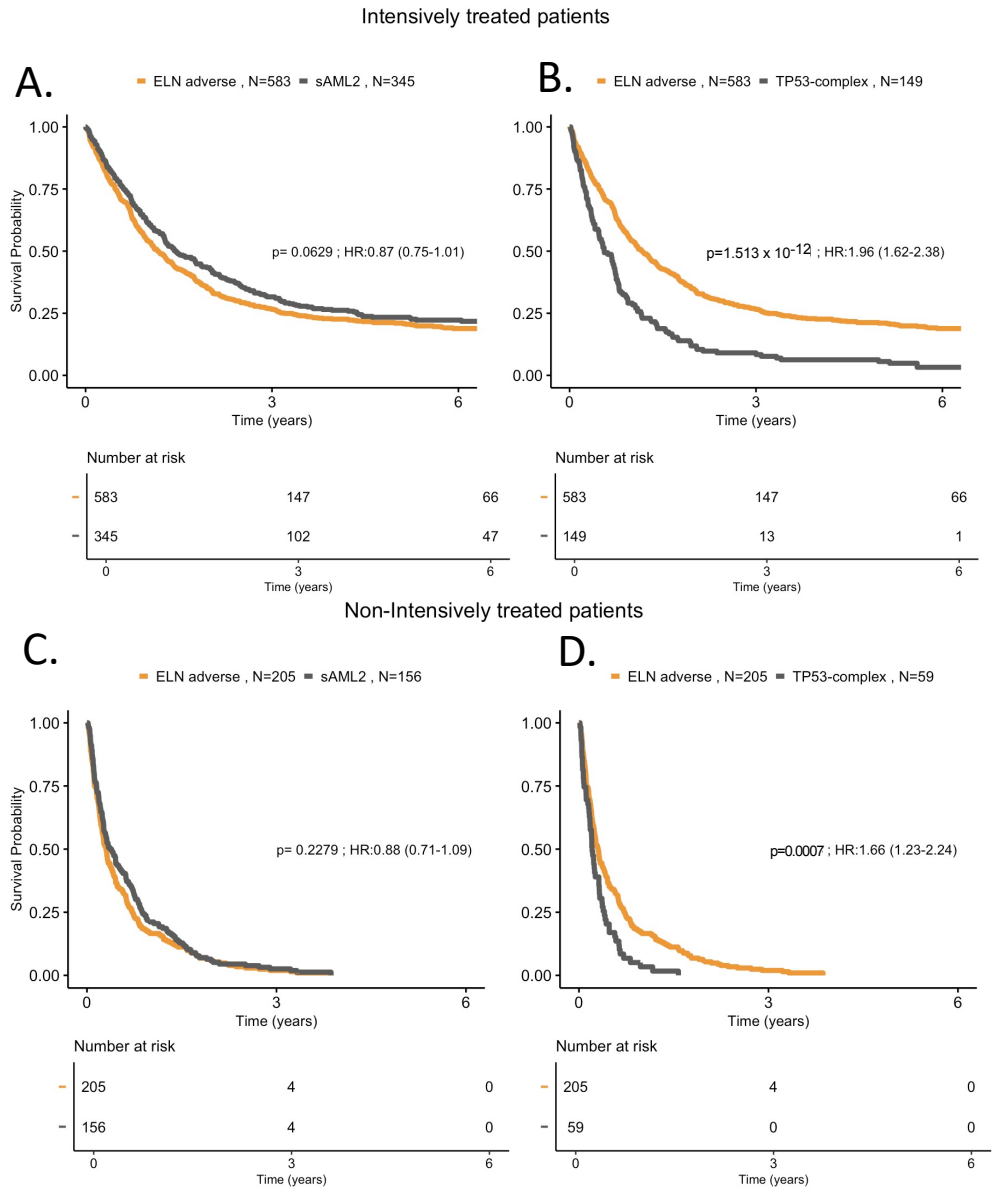
B.



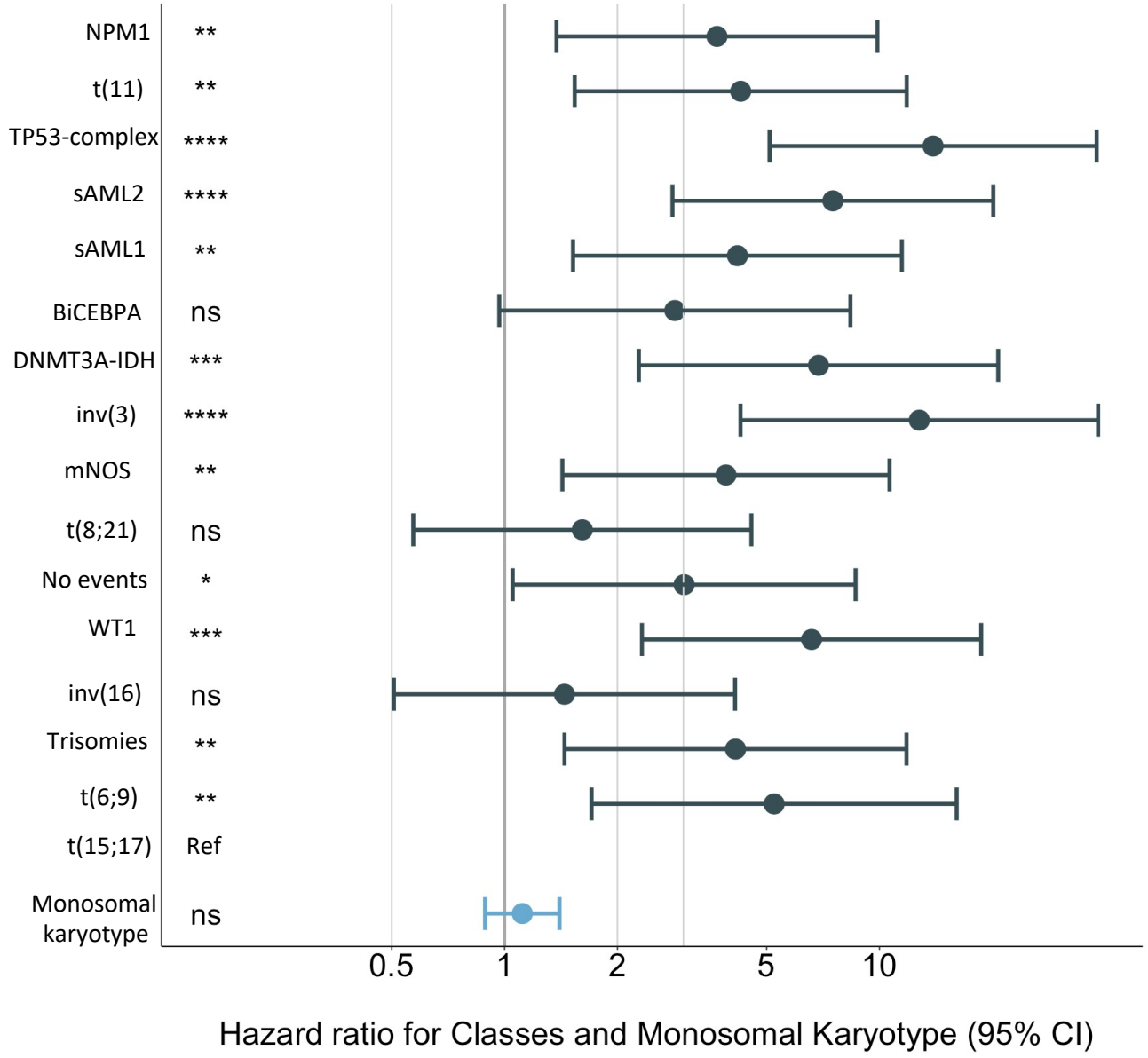
C.

NPM1	0.41 (0.35-0.48)
t(11)	0.22 (0.12-0.38)
TP53-complex	0.09 (0.06-0.15)
sAML2	0.22 (0.16-0.3)
sAML1	0.16 (0.04-0.66)
CEBPAbi	0.69 (0.58-0.82)
DNMT3A-IDH	0.47 (0.28-0.78)
inv(3)	0.09 (0.02-0.34)
mNOS	0.36 (0.29-0.46)
t(8;21)	0.42 (0.27-0.66)
No events	0.39 (0.27-0.57)
WT1	0.36 (0.2-0.65)
inv(16)	0.73 (0.62-0.85)
Trisomies	0.51 (0.36-0.7)
t(6;9)	0.22 (0.08-0.65)
t(15;17)	0.73 (0.63-0.85)
ELN Favorable	0.54 (0.49-0.59)
ELN Adverse	0.19 (0.14-0.27)
ELN Intermediate	0.32 (0.28-0.38)

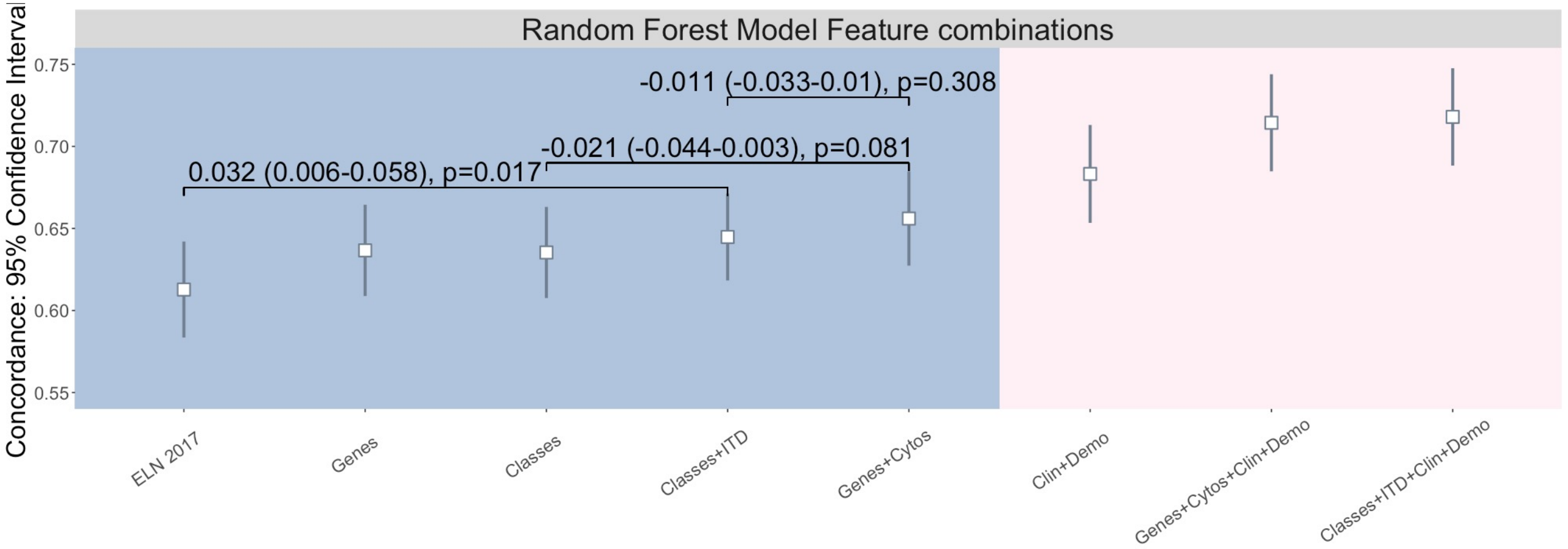
S.Figure 20: Kaplan-Meier curves for overall survival and associated risk tables comparing
A. AML class sAML2 to ELN^{2017} adverse risk group in AML NCRI cohort on the subset of intensively treated patients (n=1,755). B. AML class TP53-complex to ELN^{2017} adverse risk group in AML NCRI cohort on the subset of intensively treated patients (n=1,755). C. AML class sAML2 to ELN^{2017} adverse risk group in AML NCRI cohort on the subset of non-intensively treated patients (n=358). D. AML class TP53-complex to ELN^{2017} adverse risk group in AML NCRI cohort on the subset of non-intensively treated patients (n=358). Annotated P values are from two-sided log-rank tests.



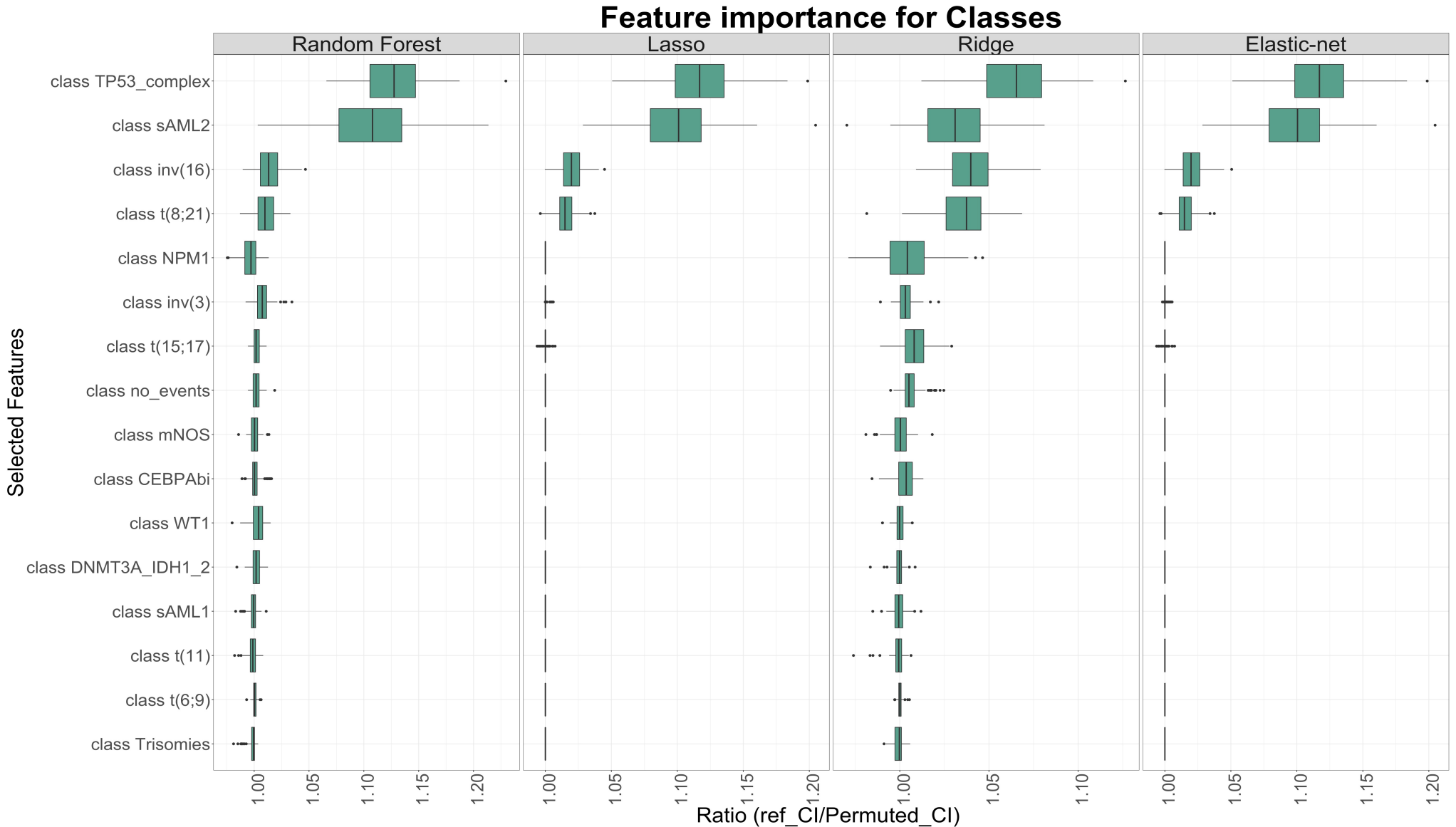
S.Figure 21: Forest plot multivariate Cox regression of classes and monosomal karyotype as defined by Breems et al. in NCRI trial study set (n= 2,113). Dots and lines represent estimated hazard ratios and 95% CI, respectively. Dots and lines represent estimated hazard ratios and 95% CI, respectively. ****P < 0.0001, ***P < 0.001, **P < 0.01, ns, not significant. P > 0.05, Wald test.



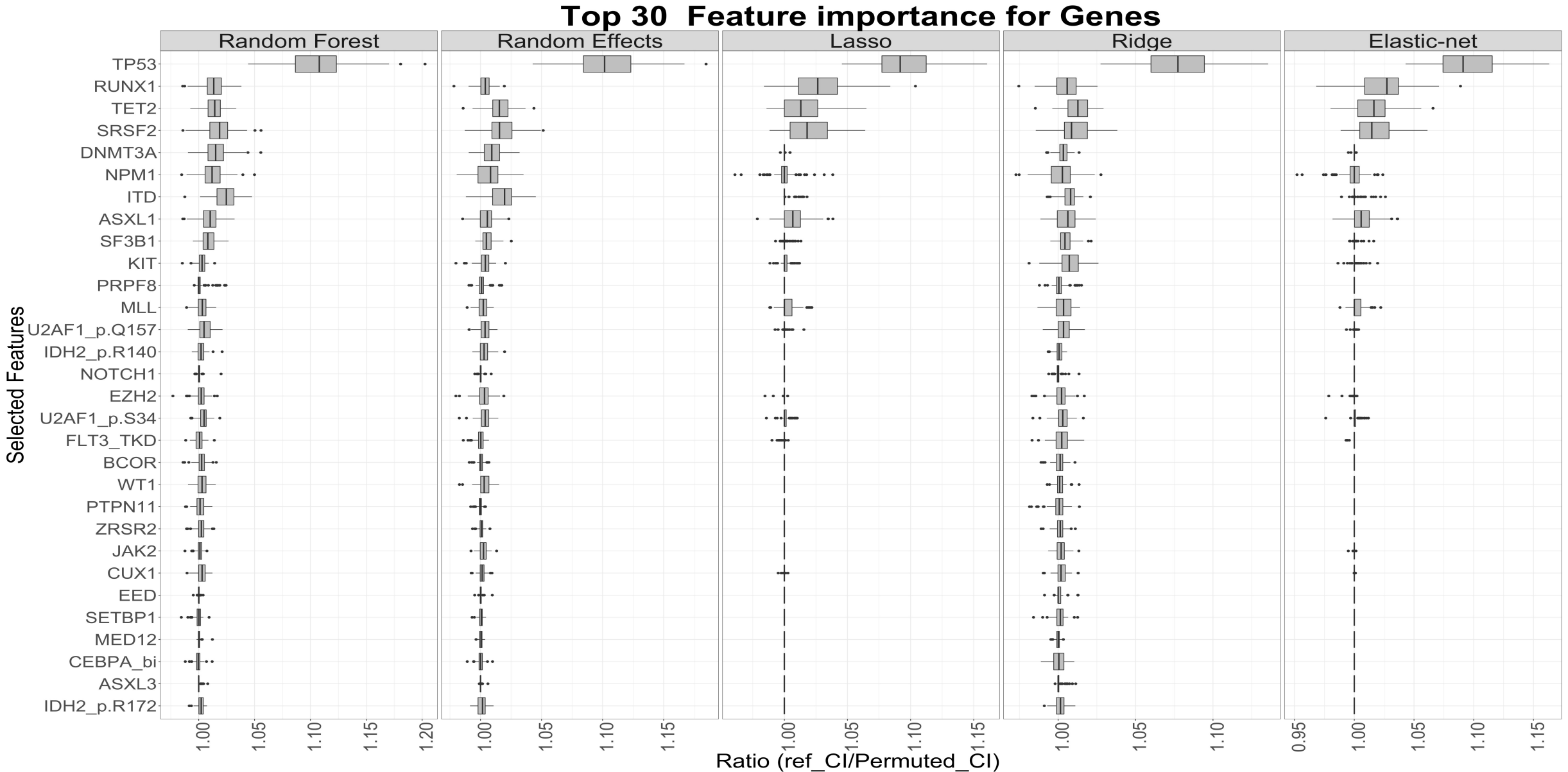
S.Figure 22: Estimates of the concordance index (C-index) derived from Random Forest Survial model that consider ELN^{2017} strata; gene mutations; molecular classes; molecular classes + FLT3$^{ITD;}$ genetic data (gene mutations and cytogenetics); clinical and demographic; genetic, clinical and demographic ; classes, FLT3^{ITD}, clinical and demographic features in the AML NCRI trial (n=2,113). The centers of the error bars represent the mean; the lower and upper whiskers represent the 95% CIs. Annotated P values are from two-sided t score test.



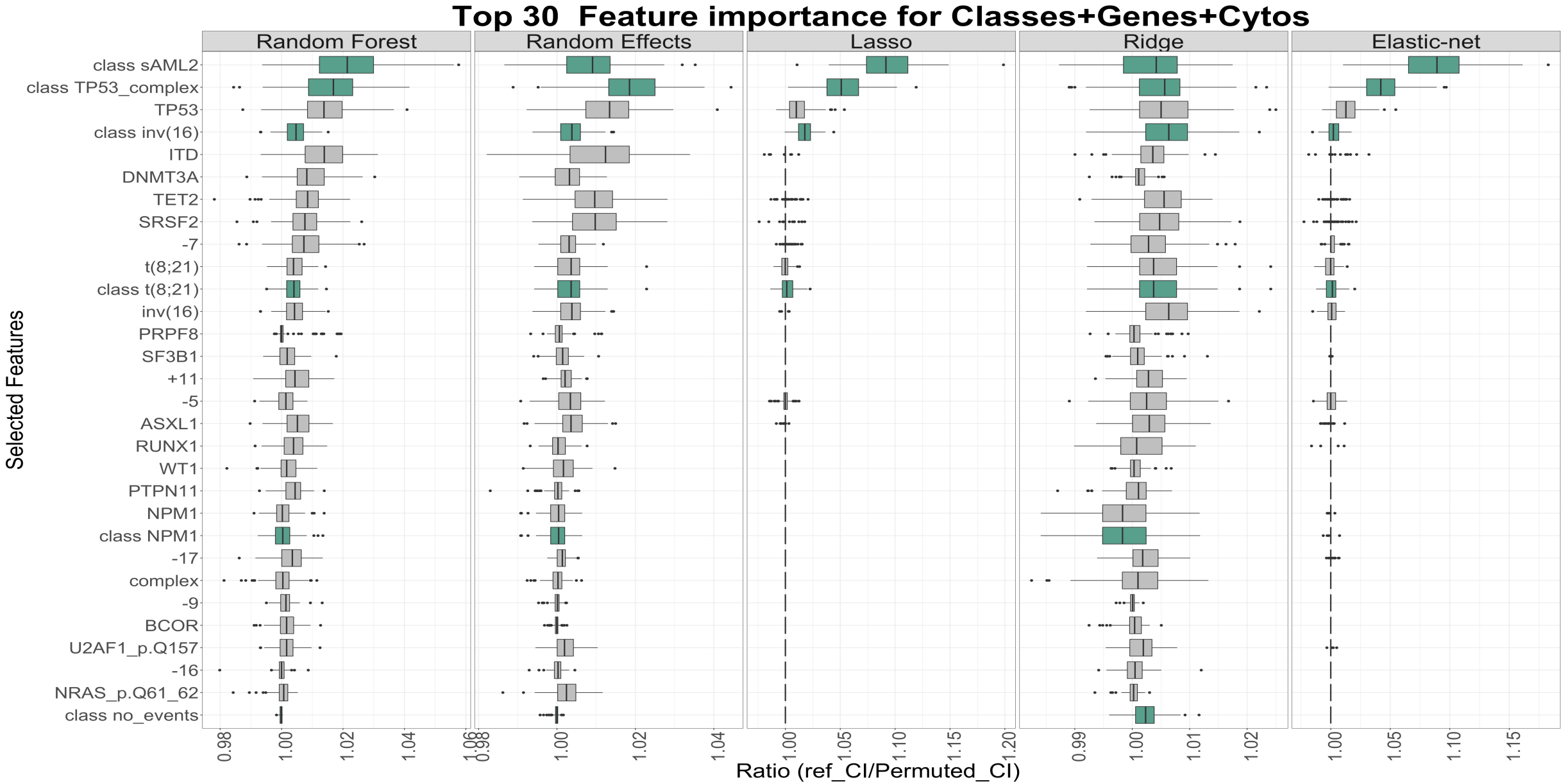
S.Figure 23: Feature Importance in the class based model in AML NCRI Trial (n=2,113).
The y axis corresponds to the different features evaluated by the models and ordered by importance. The x axis corresponds to the reshuffling ratio metric for each feature (metric = reference C-Index / permuted C-Index). The higher the ratio, the more sensitive the model is to that particular feature. The results are stratified by algorithms (Random Forest, Lasso, Ridge and Elastic-Net). For more details, please refer to S.Appendix. We omitted Random Effects algorithms for this class based model as the C-index ratio was constant across features. In all boxplots, the median is indicated by the white dot and the first and third quartiles by the box edges. The lower and upper whiskers extend from the hinges to the smallest and largest values, respectively, no further than 1.5× interquartile range from the hinges.



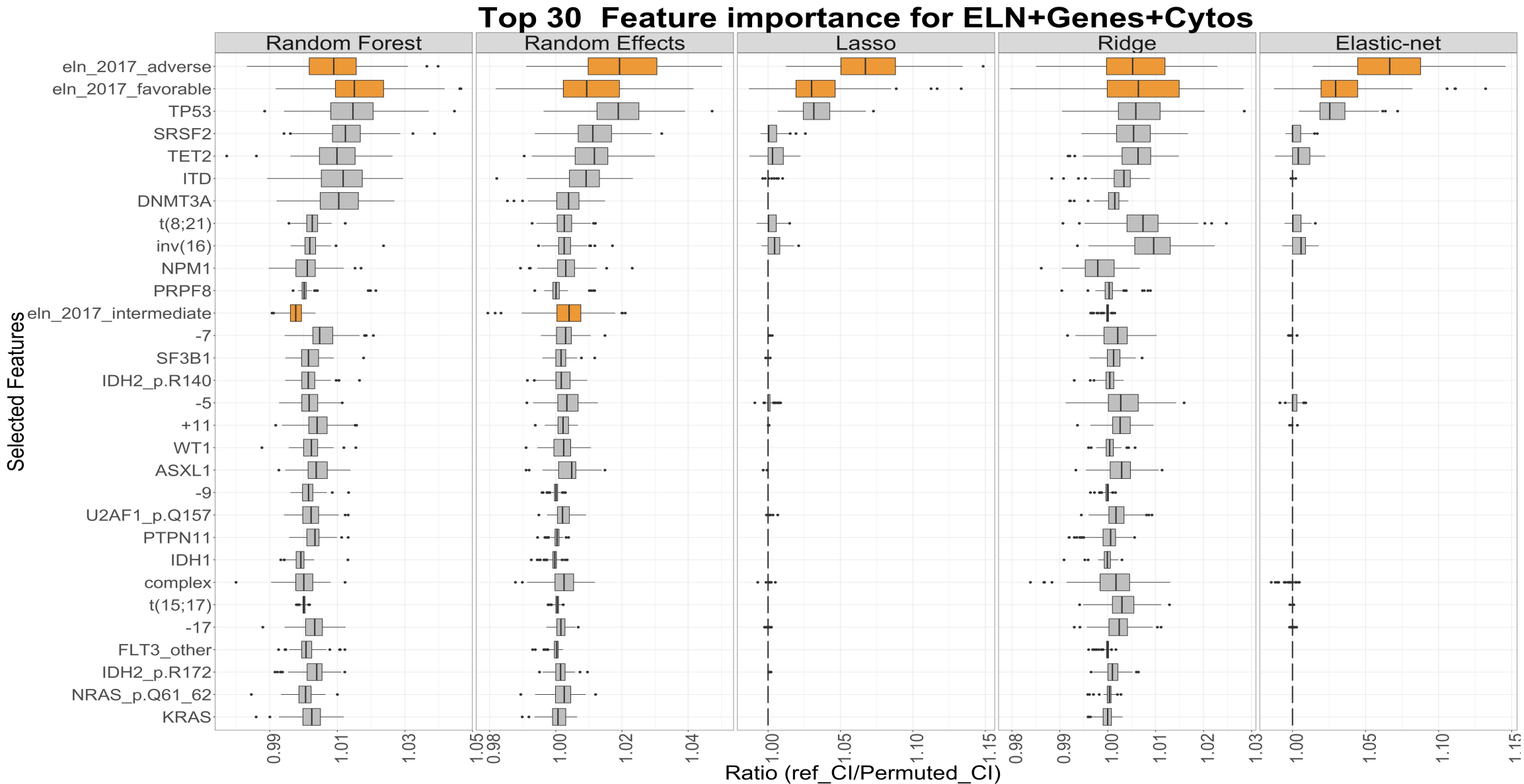
S.Figure 24: Feature Importance in the genes based model (top 30) in AML NCRI Trial (n=2,113).
The y axis corresponds to the different features evaluated by the models and ordered by importance. The x axis corresponds to the reshuffling ratio metric for each feature (metric = reference C-Index / permuted C-Index). The higher the ratio, the more sensitive the model is to that particular feature. The results are stratified by algorithms (Random Forest, Random Effects, Lasso, Ridge and Elastic-Net). For more details, please refer to S.Appendix. In all boxplots, the median is indicated by the white dot and the first and third quartiles by the box edges. The lower and upper whiskers extend from the hinges to the smallest and largest values, respectively, no further than 1.5× interquartile range from the hinges.



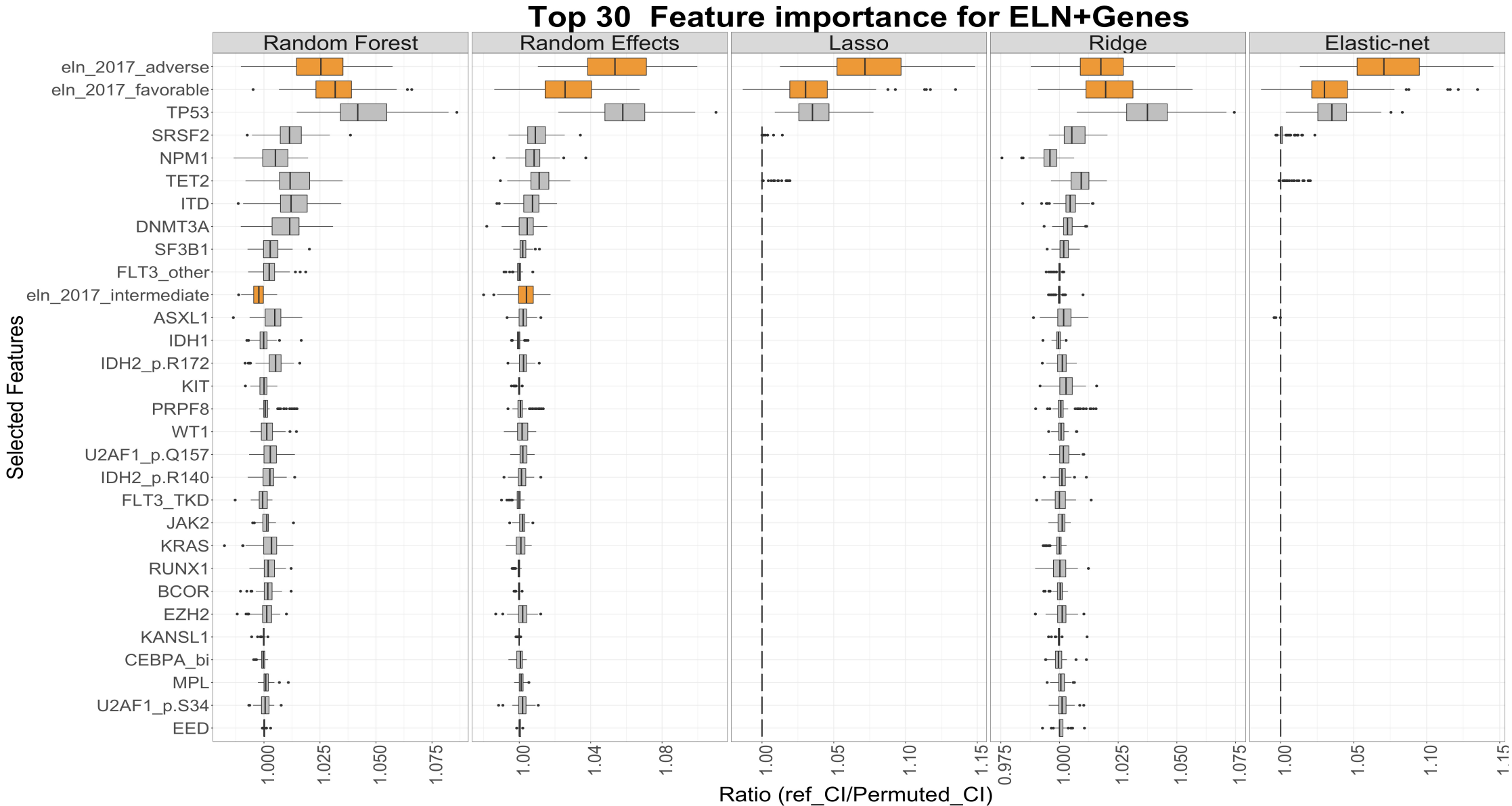
S.Figure 25: Feature Importance in the class + genes + cytogenetics based model (top 30) in AML NCRI Trial (n=2,113).
The y axis corresponds to the different features evaluated by the models and ordered by importance. The x axis corresponds to the reshuffling ratio metric for each feature (metric = reference C-Index / permuted C-Index). The higher the ratio, the more sensitive the model is to that particular feature. The results are stratified by algorithms (Random Forest, Random Effects, Lasso, Ridge and Elastic-Net). For more details, please refer to S.Appendix. In all boxplots, the median is indicated by the white dot and the first and third quartiles by the box edges. The lower and upper whiskers extend from the hinges to the smallest and largest values, respectively, no further than 1.5× interquartile range from the hinges.



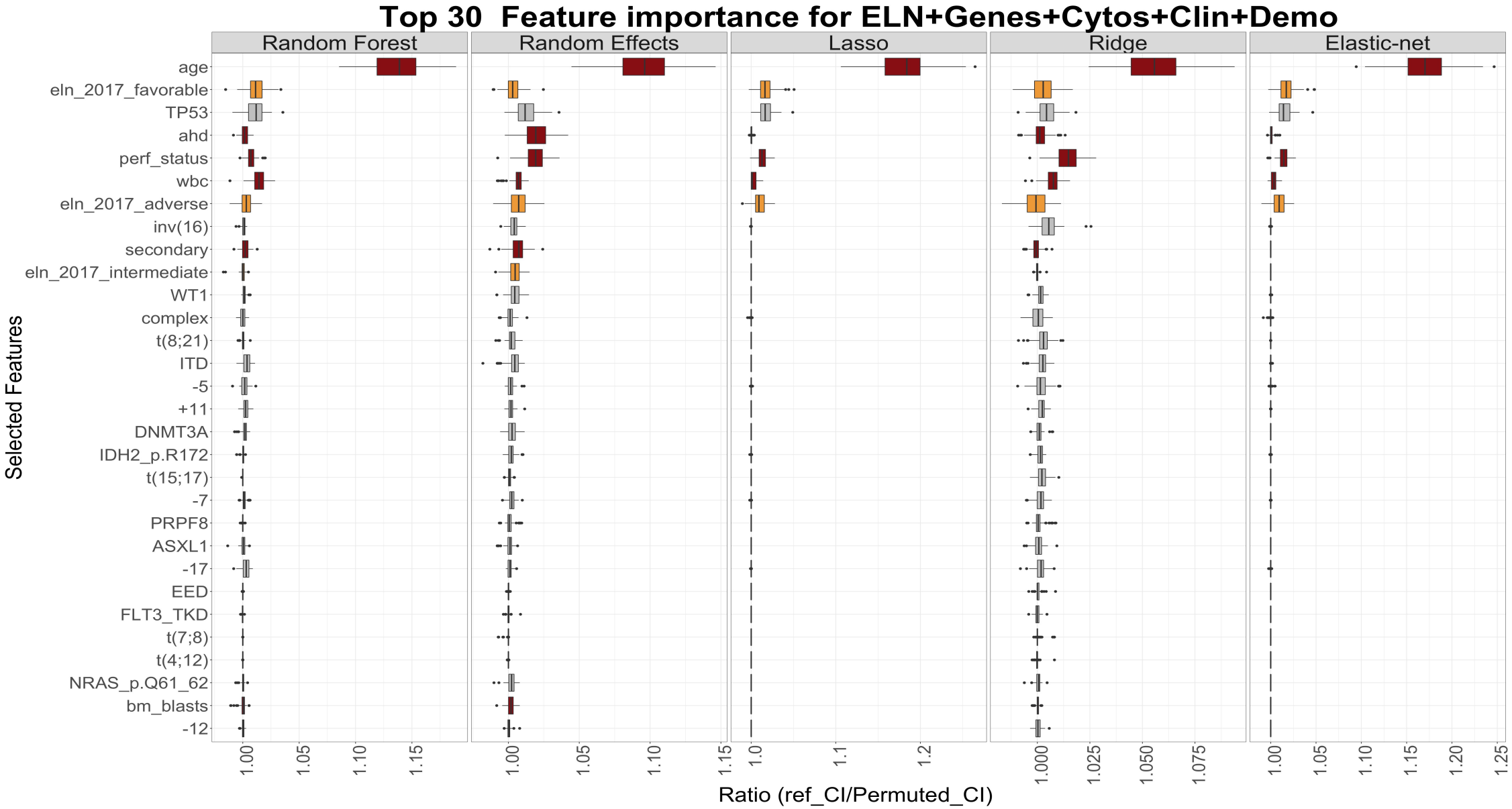
S.Figure 26: Feature Importance in the ELN + genes + cytos based model (top 30) in AML NCRI Trial (n=2,113).
The y axis corresponds to the different features evaluated by the models and ordered by importance. The x axis corresponds to the reshuffling ratio metric for each feature (metric = reference C-Index / permuted C-Index). The higher the ratio, the more sensitive the model is to that particular feature. The results are stratified by algorithms (Random Forest, Random Effects, Lasso, Ridge and Elastic-Net). For more details, please refer to S.Appendix. In all boxplots, the median is indicated by the white dot and the first and third quartiles by the box edges. The lower and upper whiskers extend from the hinges to the smallest and largest values, respectively, no further than 1.5× interquartile range from the hinges.



S.Figure 27: Feature Importance in the ELN + genes based model (top 30) in AML NCRI Trial (n=2,113).
The y axis corresponds to the different features evaluated by the models and ordered by importance. The x axis corresponds to the reshuffling ratio metric for each feature (metric = reference C-Index / permuted C-Index). The higher the ratio, the more sensitive the model is to that particular feature. The results are stratified by algorithms (Random Forest, Random Effects, Lasso, Ridge and Elastic-Net). For more details, please refer to S.Appendix. In all boxplots, the median is indicated by the white dot and the first and third quartiles by the box edges. The lower and upper whiskers extend from the hinges to the smallest and largest values, respectively, no further than 1.5× interquartile range from the hinges.

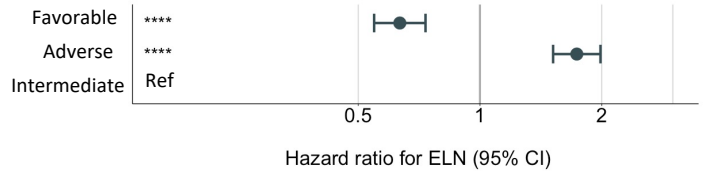


S.Figure 28: Feature Importance in the full Model (ELN + Genes + Cytos + Clin + Demo) in AML NCRI Trial (n=2,113).
The y axis corresponds to the different features evaluated by the models and ordered by importance. The x axis corresponds to the reshuffling ratio metric for each feature (metric = reference C-Index / permuted C-Index). The higher the ratio, the more sensitive the model is to that particular feature. The results are stratified by algorithms (Random Forest, Random Effects, Lasso, Ridge and Elastic-Net). For more details, please refer to S.Appendix. In all boxplots, the median is indicated by the white dot and the first and third quartiles by the box edges. The lower and upper whiskers extend from the hinges to the smallest and largest values, respectively, no further than 1.5× interquartile range from the hinges.

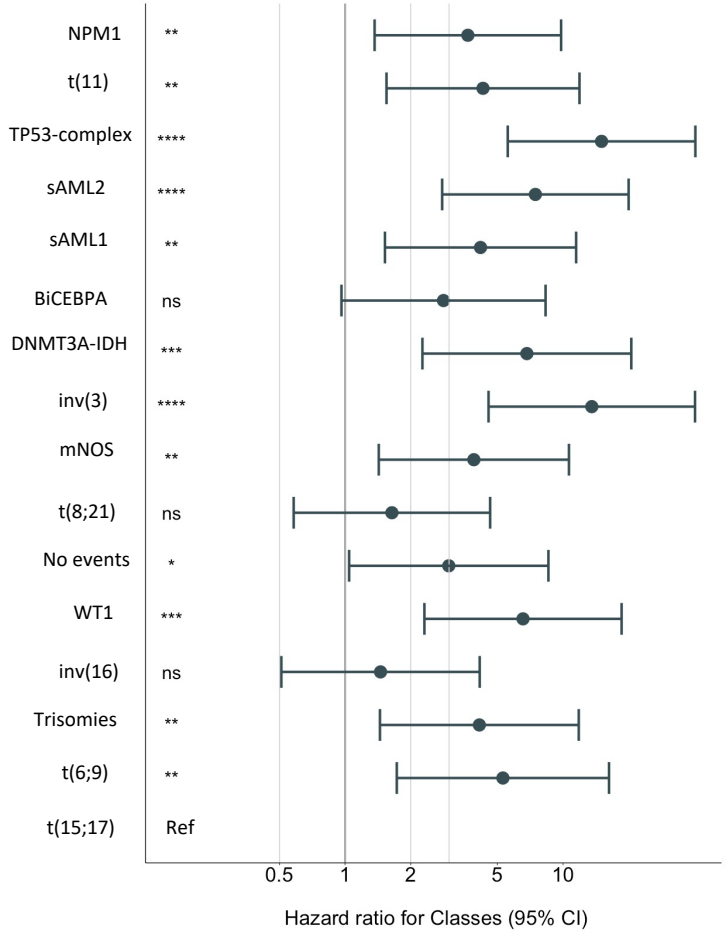


S.Figure 29: Forest plot multivariate Cox regression of A. ELN^{2017} risk categories and B. Classes in NCRI trial study set (n= 2,113). Dots and lines represent estimated hazard ratios and 95% CI, respectively. ****P < 0.0001, ***P < 0.001, **P < 0.01, ns, not significant. P > 0.05, Wald test.

A.

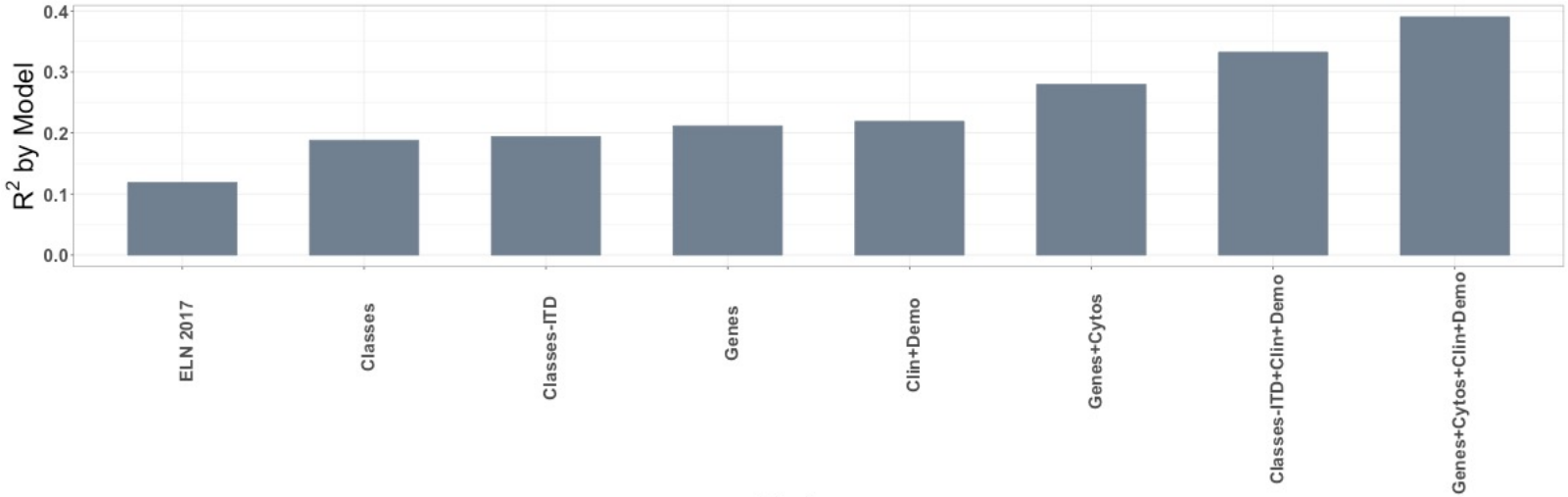


B.

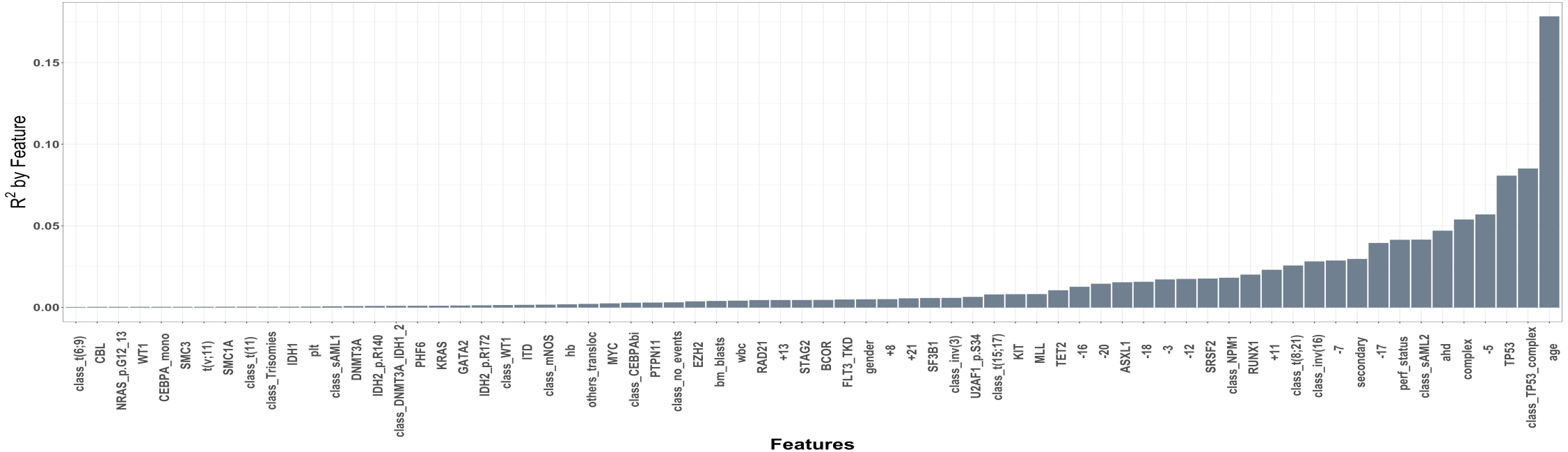


S.Figure 30: Explained variation and randomness using Nagelkerke's pseudo R^2 in the AML NCRI Trial (n=2,113). A. Explained variation and randomness using different subset of the covariates to include: ELN^{2017}, classes, classes + ITD, genes, clinical data, genes + cytogenetics, classes + clinical data, genes + cytogenetics + clinical data**.** B. Explained variation and randomness for each covariate.
Pseudo R^2 are relative measures indicating how well a model/feature explains the data.

A



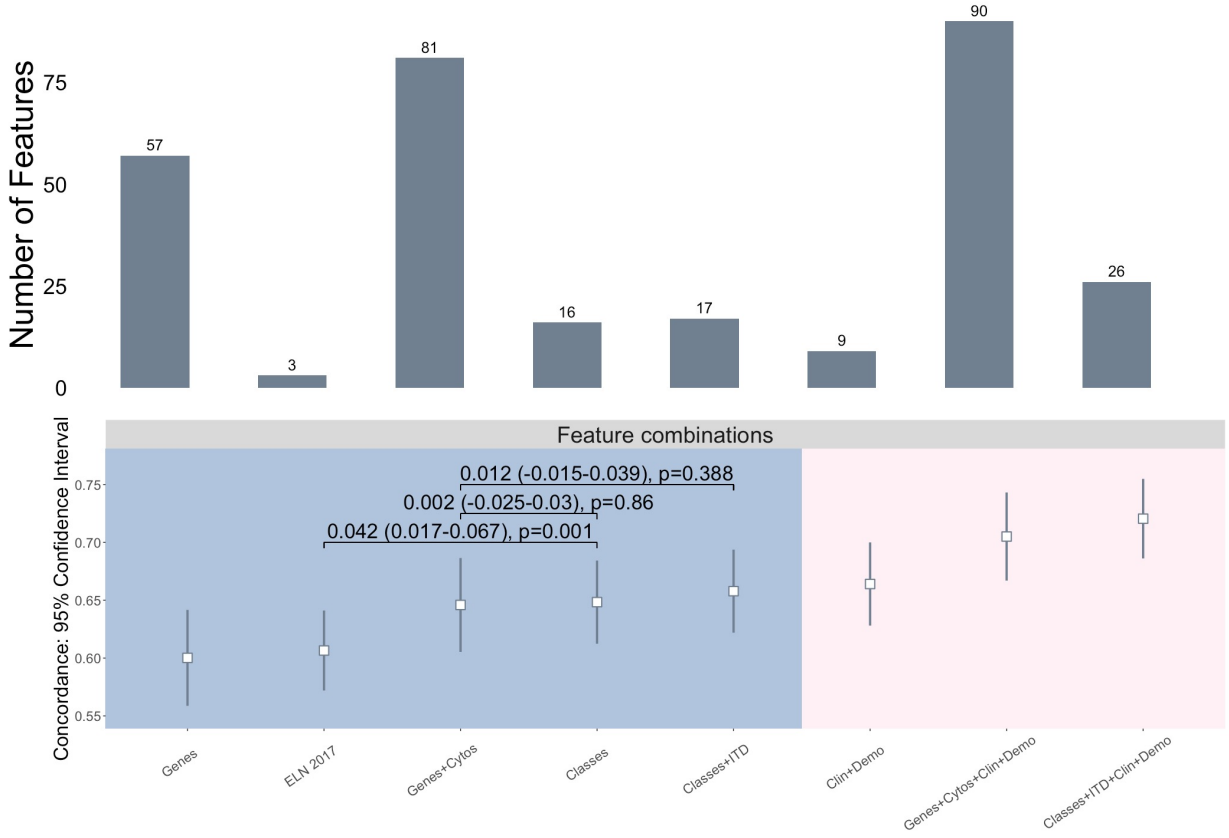
B



S.Figure 31: Prognosis Evolution Validation of C-Index in the AML SG Cohort (1,540 patients).
A. Bar plots of number of features for the different models that were evaluated to include: ELN[2017], classes, classes + ITD, genes, clinical data, genes + cytogenetics, classes + clinical data, genes + cytogenetics + clinical data**.**
B. Concordance Index (C-Index) measured using a Cox Ridge model on different subsets of the features (ELN[2017], classes, classes + ITD, genes, clinical data, genes + cytogenetics, classes + clinical data**)** with internal 5 fold cross-validation for the regularization parameter λ. We randomly reshuffled the data (n=1,540) and used 75% for training and cross validation and the remaining 25% for testing. We used 100 Bootstraps iterations on the test set to produce the box plots with 95%confidence interval and we evaluated the pvalues by comparing the differences in C-index distributions on the displayed models. We trained different models (penalized Cox Models, Random Forest, Cox Boosting, Cox Random Effects and Support Vector Machines) and here we display the results for the Ridge Cox Model. For more details, please refer to S.Appendix. The bar plots represent the number of features used to evaluate the C-Index on the different subsets of features. The centers of the error bars represent the mean; the lower and upper whiskers represent 95% CIs. Annotated P values are from two-sided t score test.

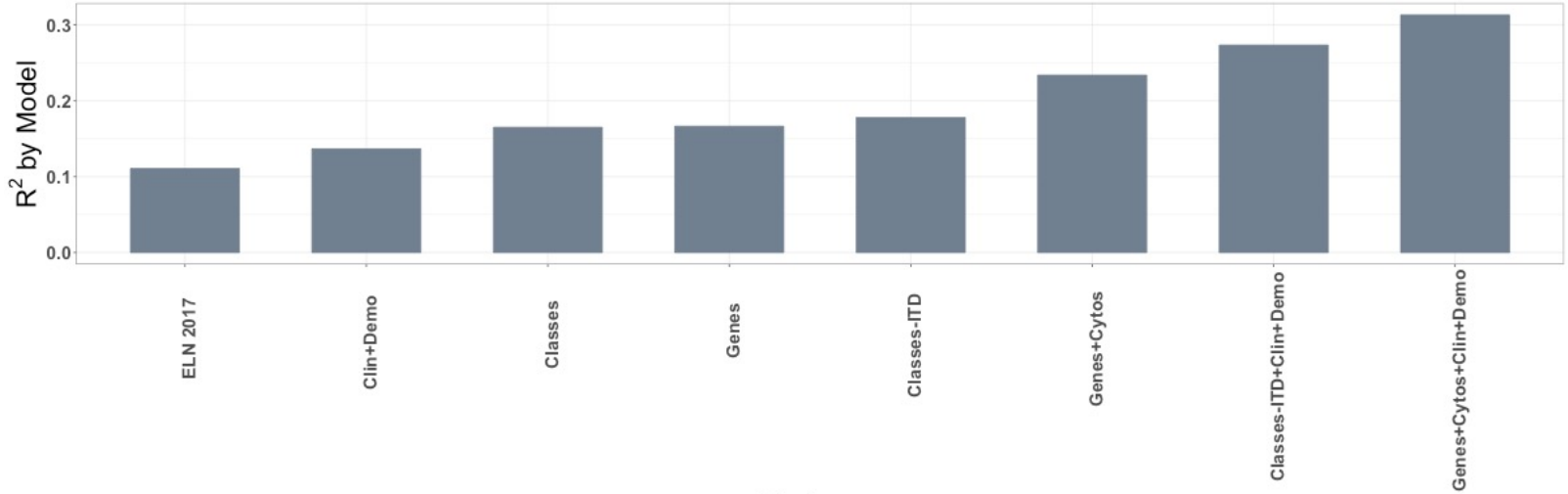
A

B

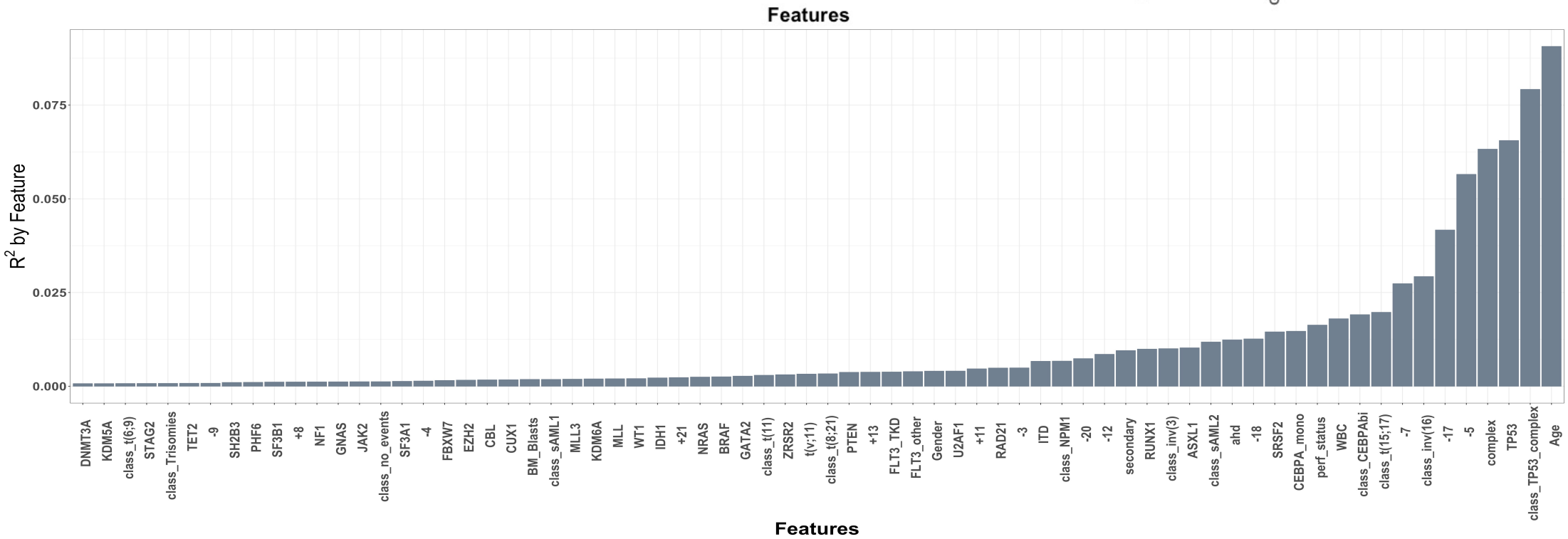


S.Figure 32: Explained variation and randomness using Nagelkerke R^2 in the validation AML SG cohort (n=1,540). A. Explained variation and randomness using different subset of the covariates to include: ELN^{2017}, classes, classes + ITD, genes, clinical data, genes + cytogenetics, classes + clinical data, genes + cytogenetics + clinical data**.** B. Explained variation and randomness for each covariate.
Pseudo R^2 are relative measures indicating how well a model/feature explains the data.

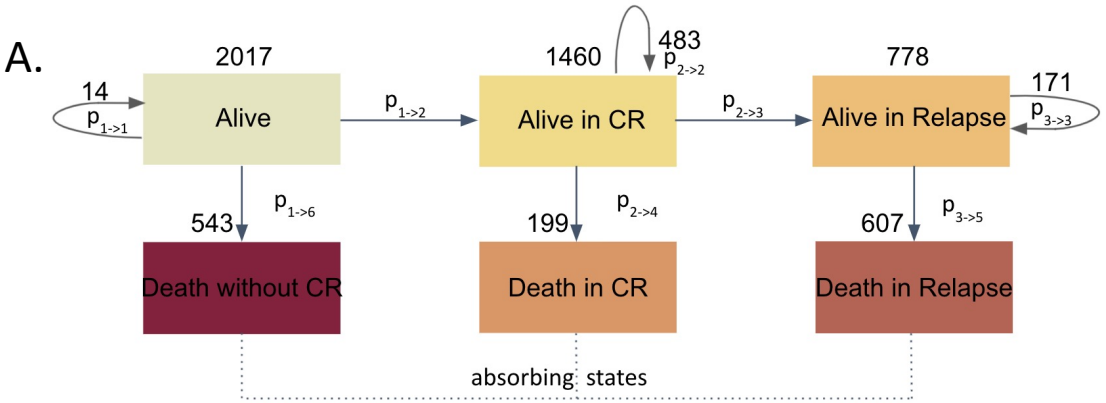
A

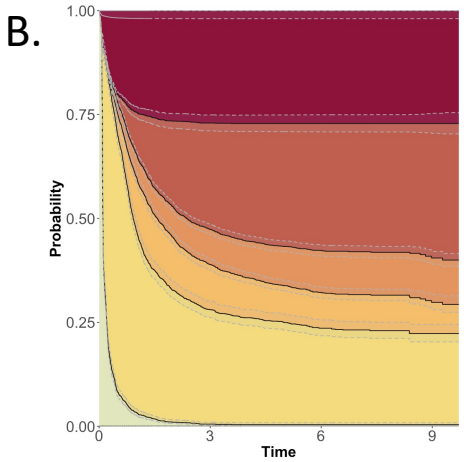


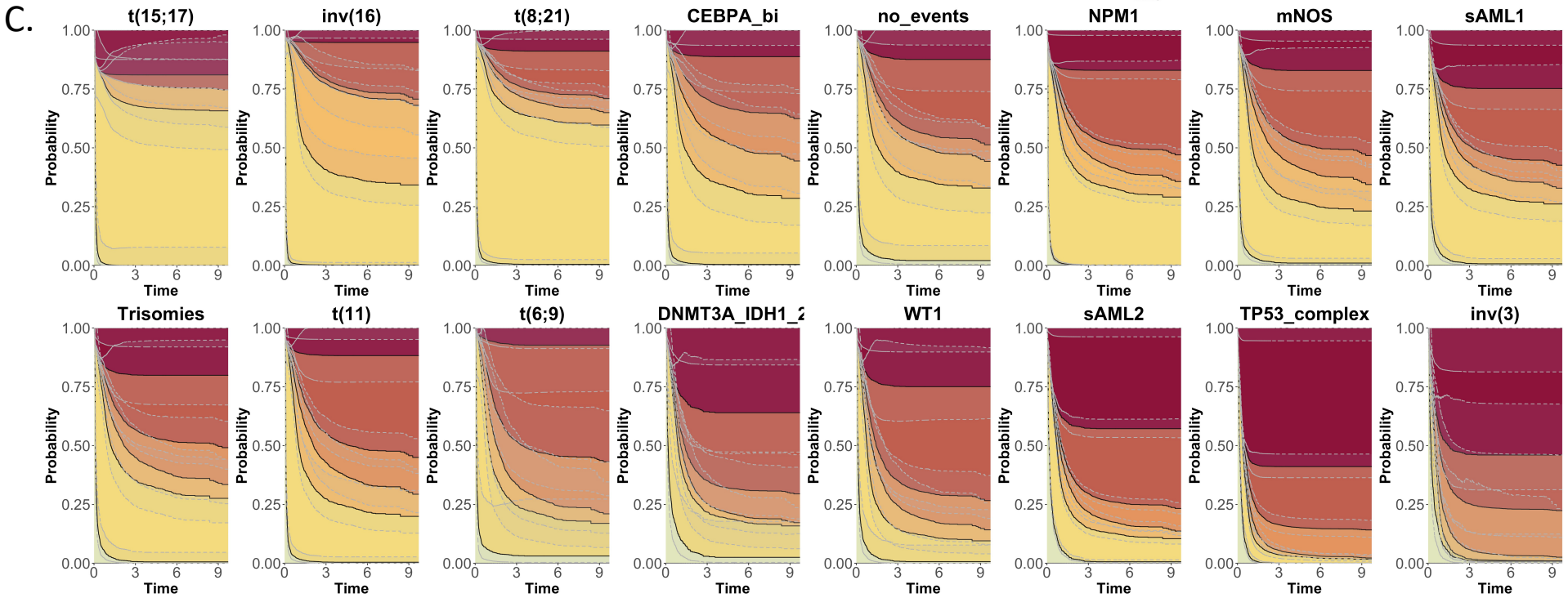
B



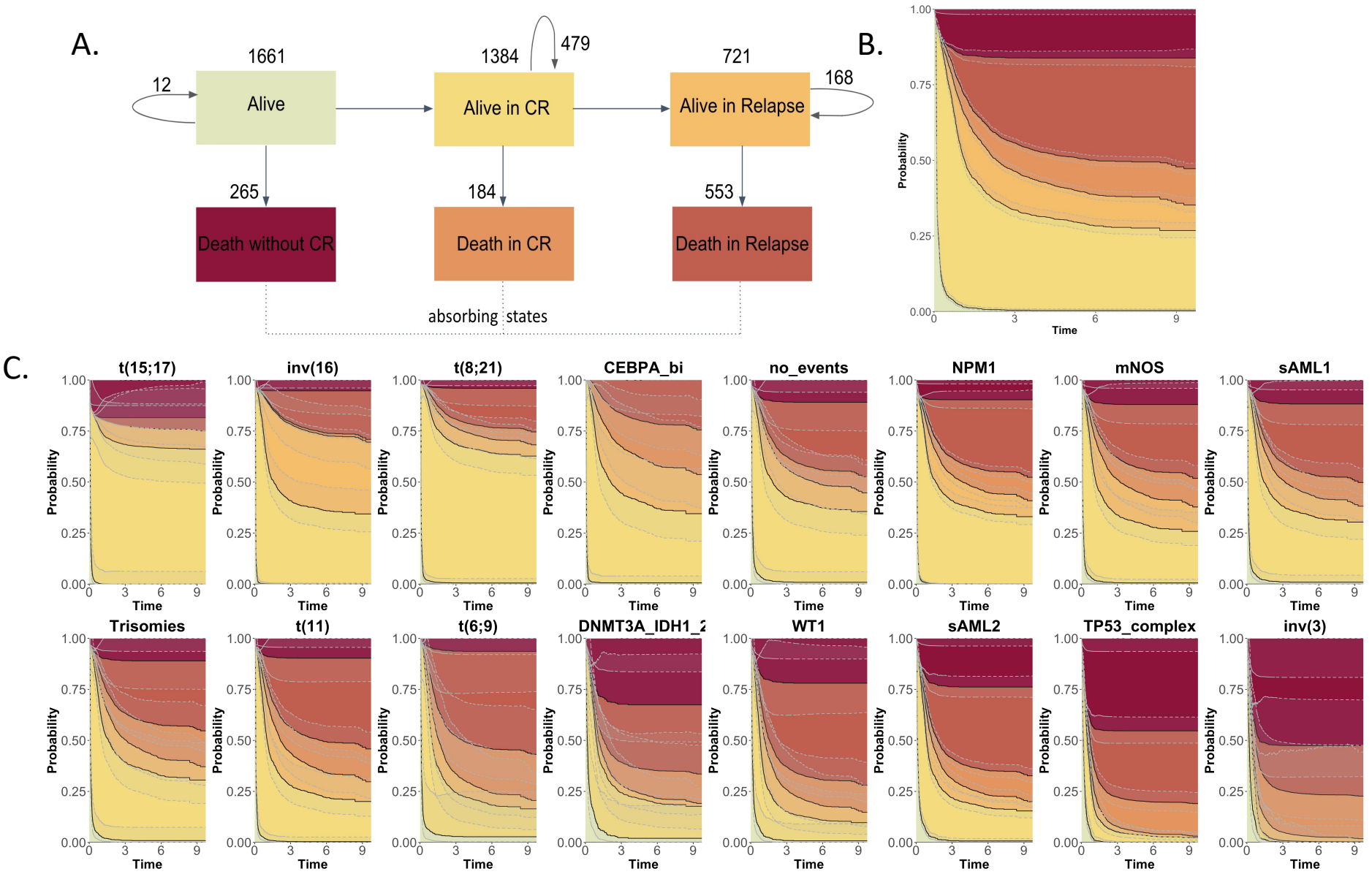
S.Figure 33: Multi-state model for disease progression in the AML NCRI cohort (n=2,017).
A. Representation of patient transitions (in numbers) across clinical endpoints (alive (meaning received induction chemotherapy); alive in complete remission; alive in relapse; death without complete remission; death in complete remission; death in relapse). B. Non-parametric multi-state transition probability with 95% confidence bands for the AML NCRI cohort for patients that received intensive treatment (n=1,661). C. Stacked transition probabilities with 95% confidence bands for each class (y-axis) across time (x-axis).







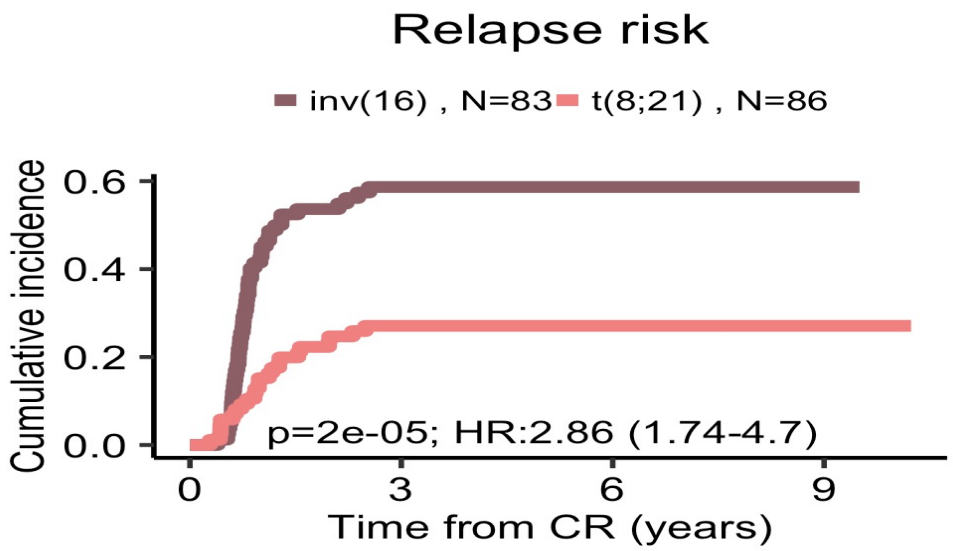
S.Figure 34: Multi-state model for disease progression for patients that received intensive treatment (n=1,661).
A. Representation of patient transitions (in numbers) across clinical endpoints (alive (meaning received induction chemotherapy); alive in complete remission; alive in relapse; death without complete remission; death in complete remission; death in relapse). B. Non-parametric multi-state transition probability with 95% confidence bands for the AML NCRI cohort for patients that received intensive treatment (n=1,661). C. Stacked transition probabilities with 95% confidence for each class (y-axis) across time (x-axis).



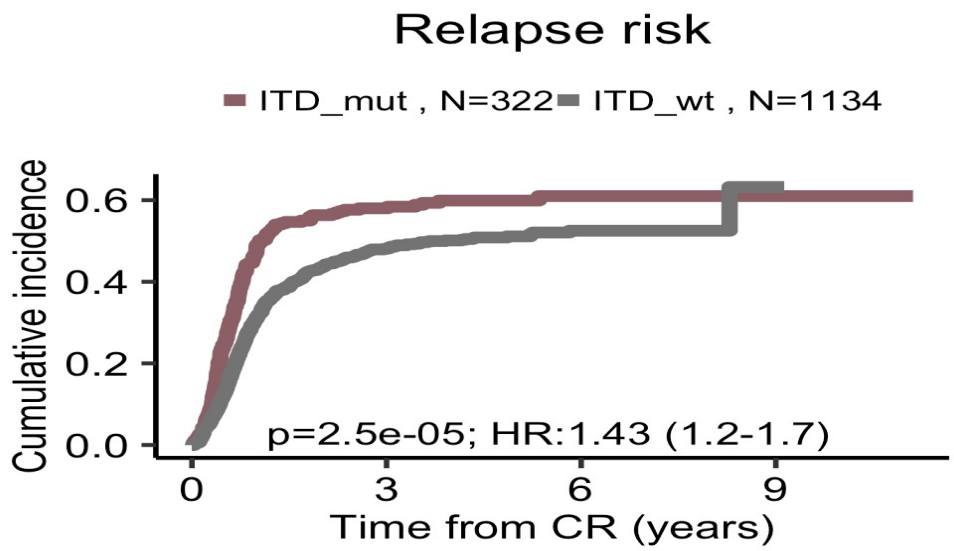
S.Figure 35: Cumulative incidence and risk outcomes for different subsets in the AML NCRI Trial (n=2,017).

A. Cumulative incidence curves comparing inv(16) and t(8;21) classes in the AML NCRI Trial (n=2,017). Pvalues were computed with two-sided Gray's test.
B. Cumulative incidence curves comparing patients with and without $FLT3^{ITD}$ in the AML NCRI Trial (n=2,017). Pvalues were computed with two-sided Gray's test.

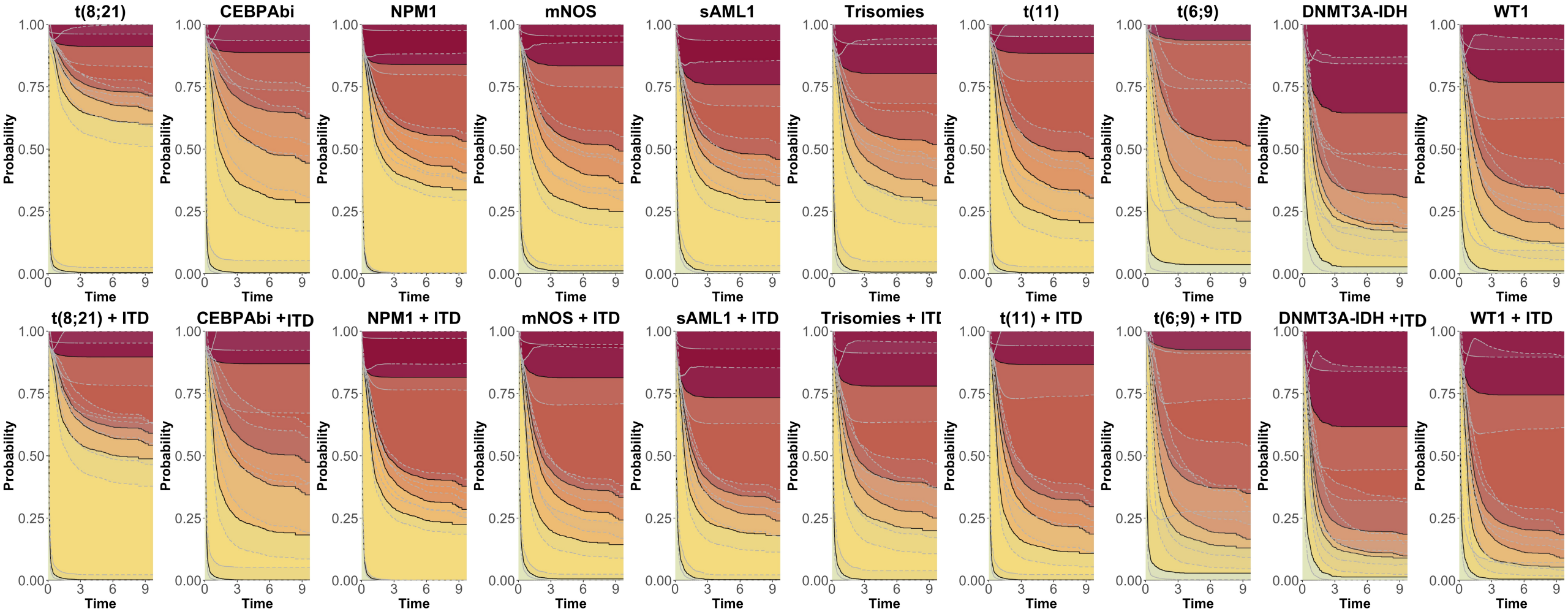
A



B

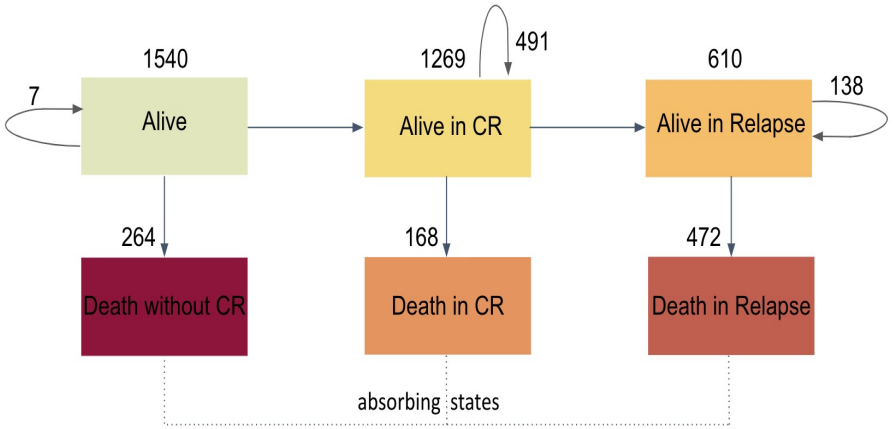


S.Figure 36: Multi-state semi-parametric Cox model incorporating FLT3^{ITD} shift for relevant components in the AML NCRI Trial (n=2,017). This is a semi-parametric Cox multi-state transition probability plot with 95% confidence bands for selected classes with and without the presence of FLT3^{ITD}. The model contains 6 possible states: alive (meaning received induction chemotherapy); alive in complete remission; alive in relapse; death without complete remission; death in complete remission; death in relapse. For more details about the semi-parametric transition hazard model, please refer to S.Appendix. The bold lines represent the death states.

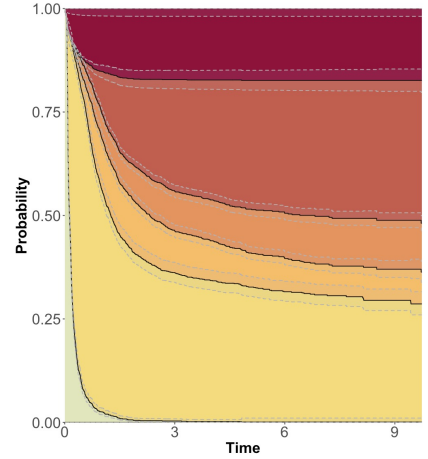


S.Figure 37: Validation of multi-state model in the validation AML SG Cohort (n=1,540). A. Multi-state transitions with number of patients in each possible state in the AML SG Cohort (n=1,540). The model contains 6 possible states: alive (meaning received induction chemotherapy); alive in complete remission; alive in relapse; death without complete remission; death in complete remission; death in relapse. B. Non-parametric multi-state transition probability with 95% confidence bands for the overall AML SG Cohort (n=1,540). C. Multi-state semi-parametric transition probabilities with 95% confidence bands for each class in the AML SG Cohort (n=1,540). The bold lines represent the death states. For more details about the semi-parametric transition hazard model, please refer to S.Appendix.

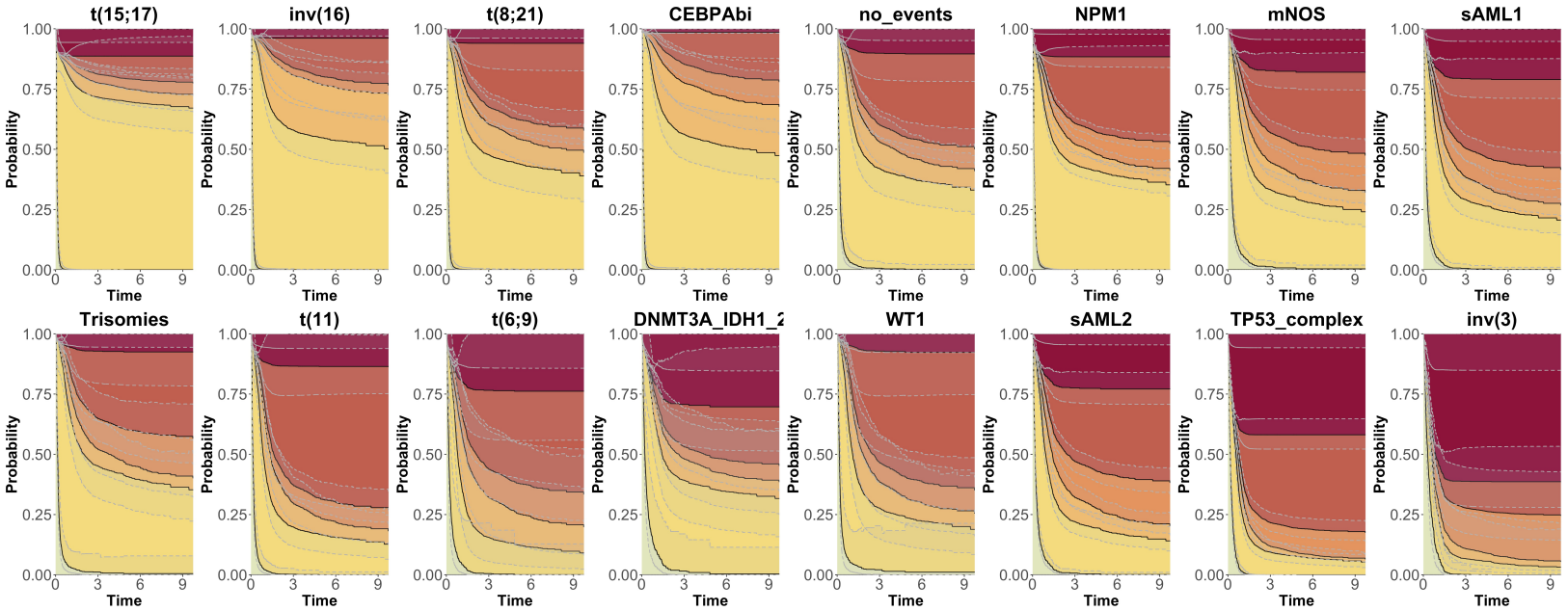
A.



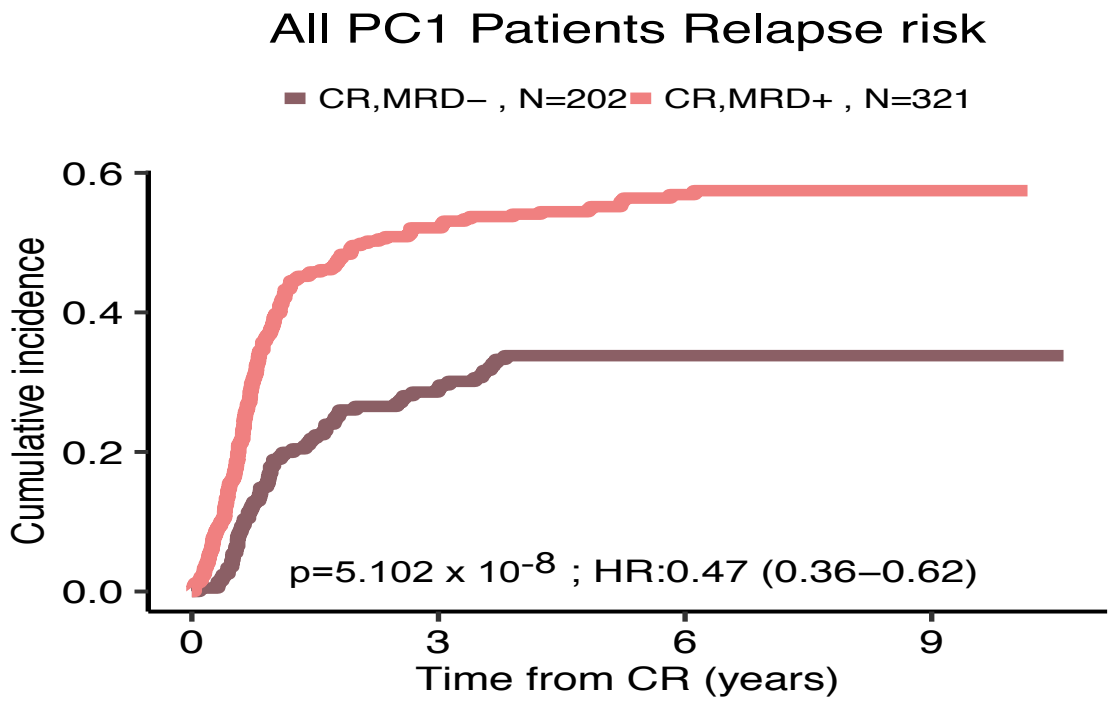
B.

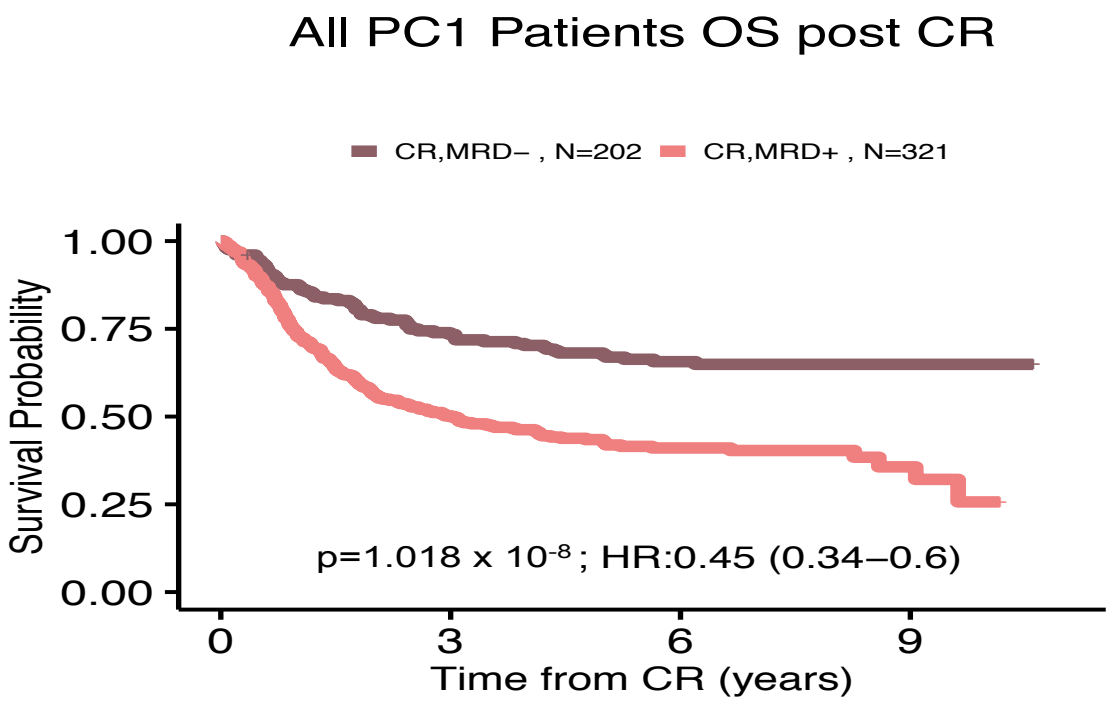


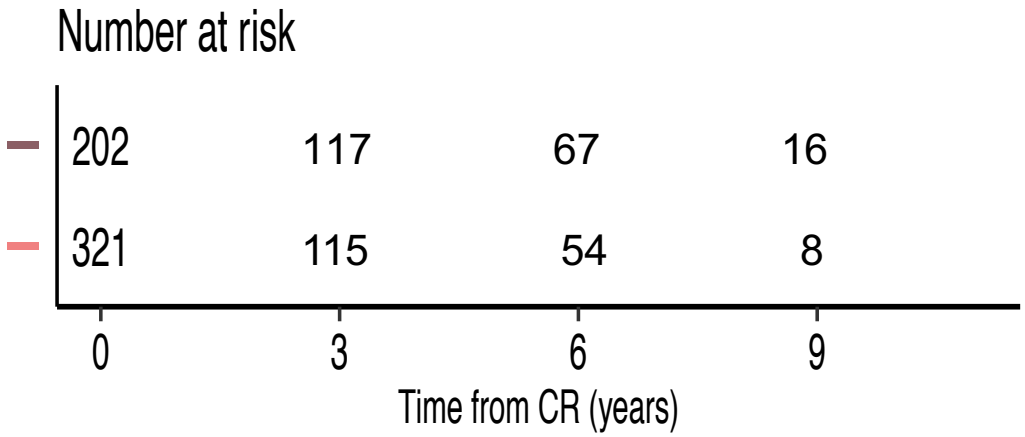
C.

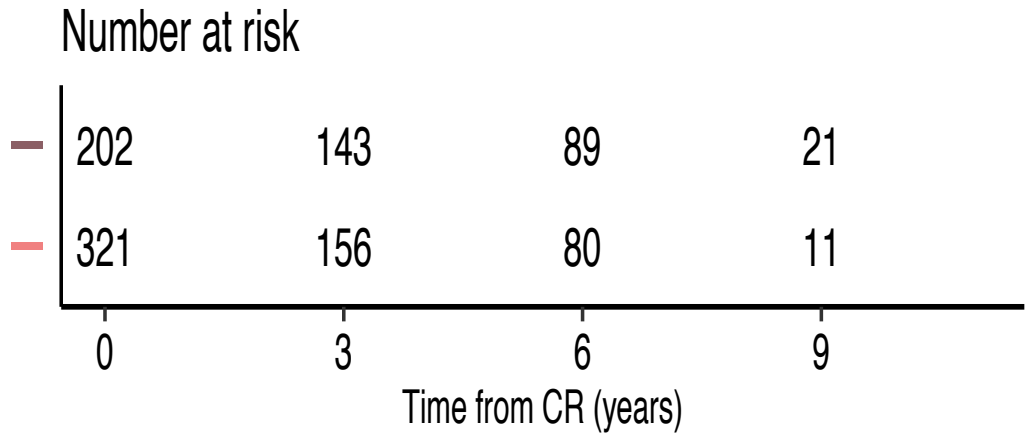


S. Figure 38: Kaplan-Meier overall survival curves, cumulative incidence of relapse and associated risk tables for patients that attained CR in AML17 trial subset, stratified by MRD status post course 1 (n=523). Two-sided Pvalues were computed with Gray's test and log-rank test.

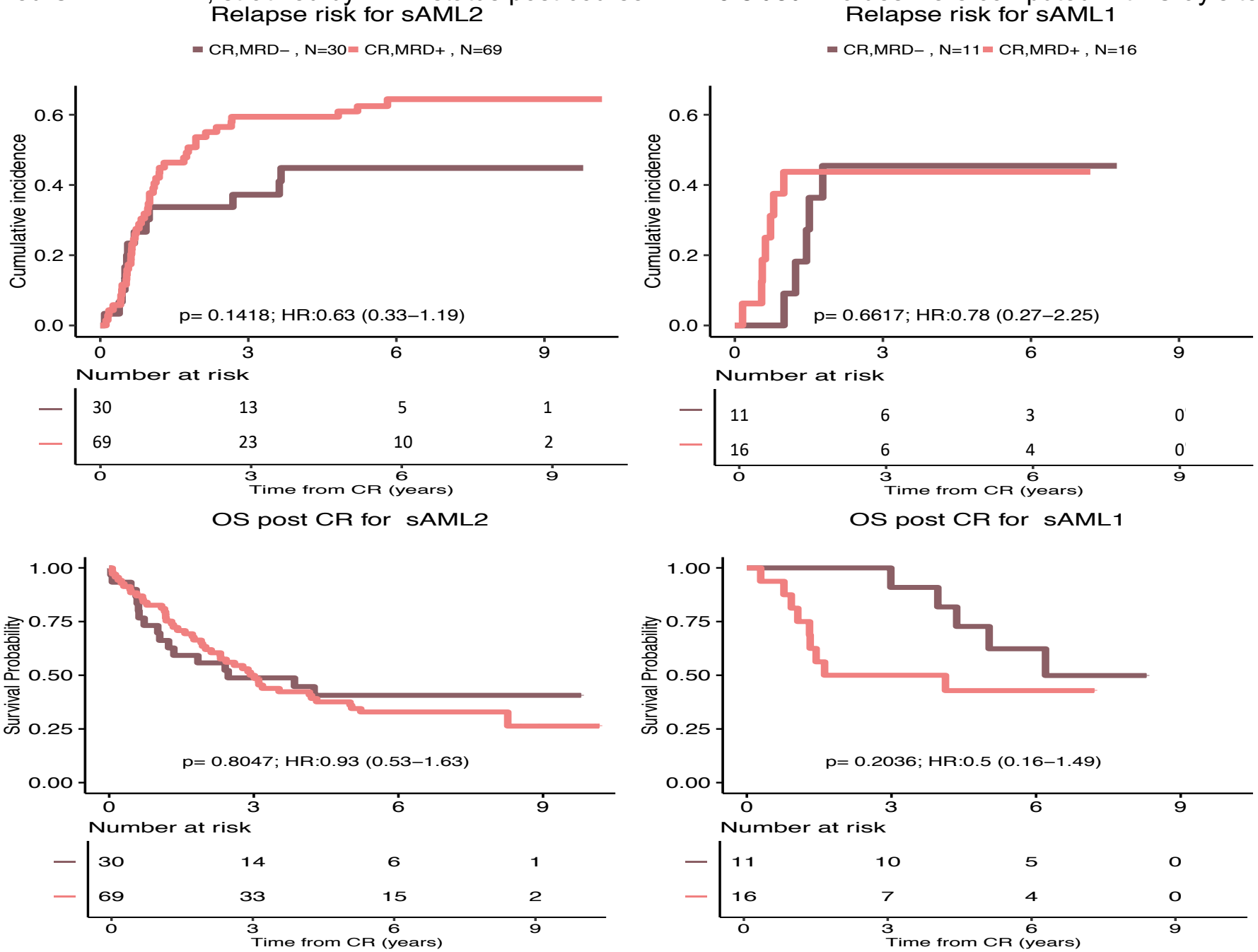






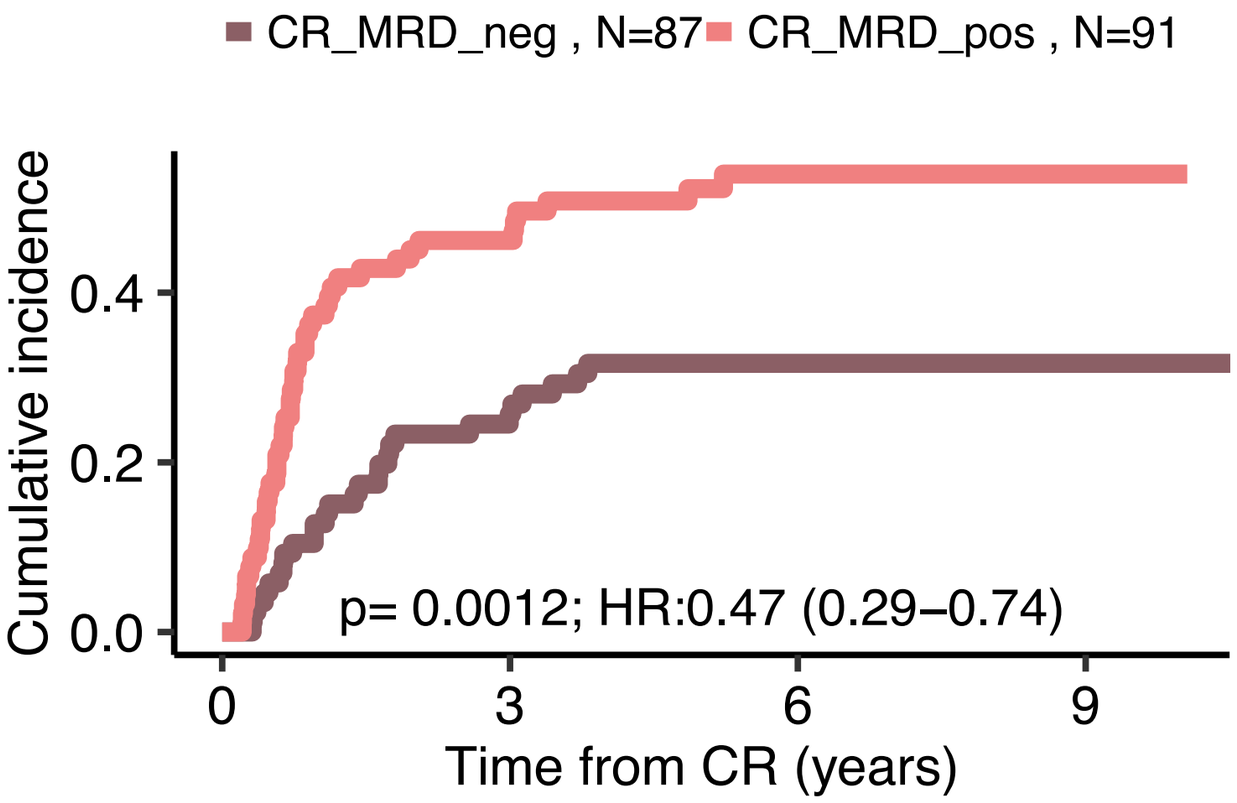


S. Figure 39: Kaplan-Meier overall survival curves, cumulative incidence of relapse and associated risk tables for the sAML2, sAML1 subgroups that attained CR in AML17, stratified by MRD status post course 1. Two-sided Pvalues were computed with Gray's test and log-rank test.

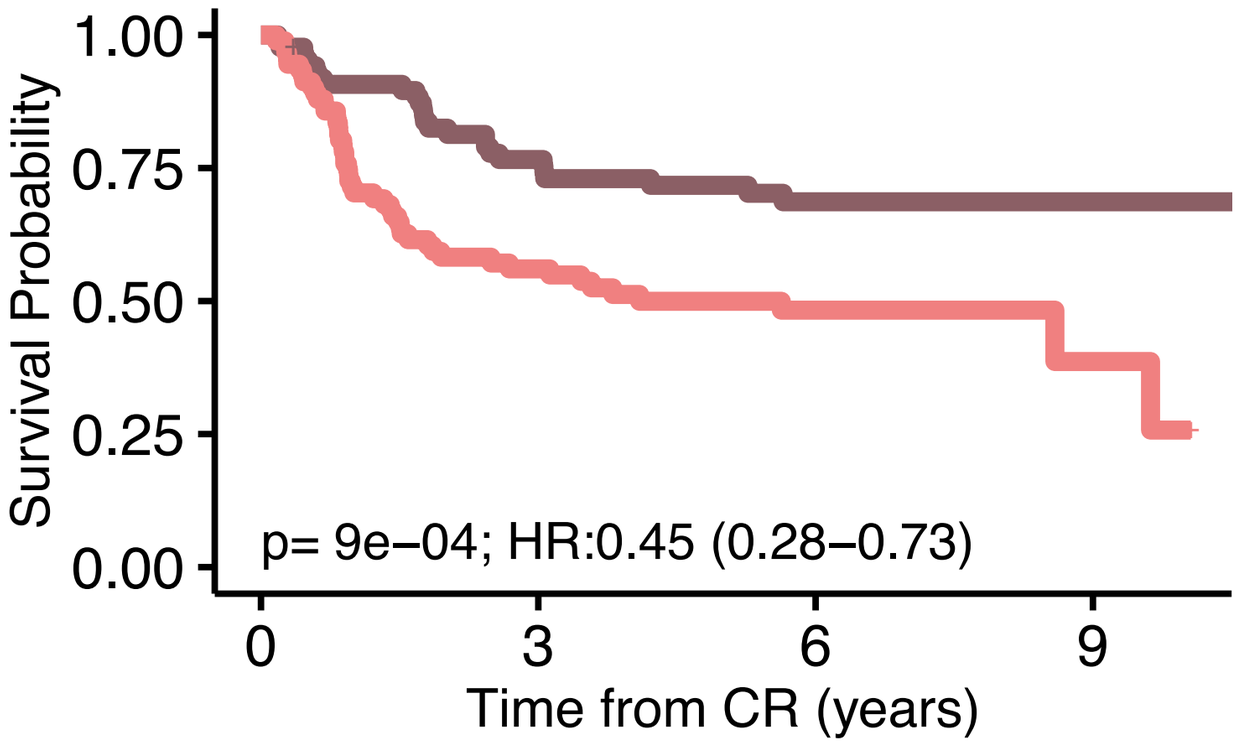


S.Figure 40: Summary by class for cumulative incidence and Kaplan-Meier curves for survival post complete remission on AML 17 NCRI Trial Cohort with post course 1 minimal residual disease (MRD) analysis analysis (n=523). The first panel represents the cumulative incidence of relapse for patients in that specific class stratified by MRD status (p values were computed with Gray test) and the second panel represents the Kaplan-Meier curve for survival post course 1 complete remission stratified by MRD status. We omitted classes with less than 3 patients with MRD positive or negative. Panels where comparator group has less than 10 patients have been removed for robust statistical interpretation. Two-sided Pvalues were computed with Gray's test and log-rank test.

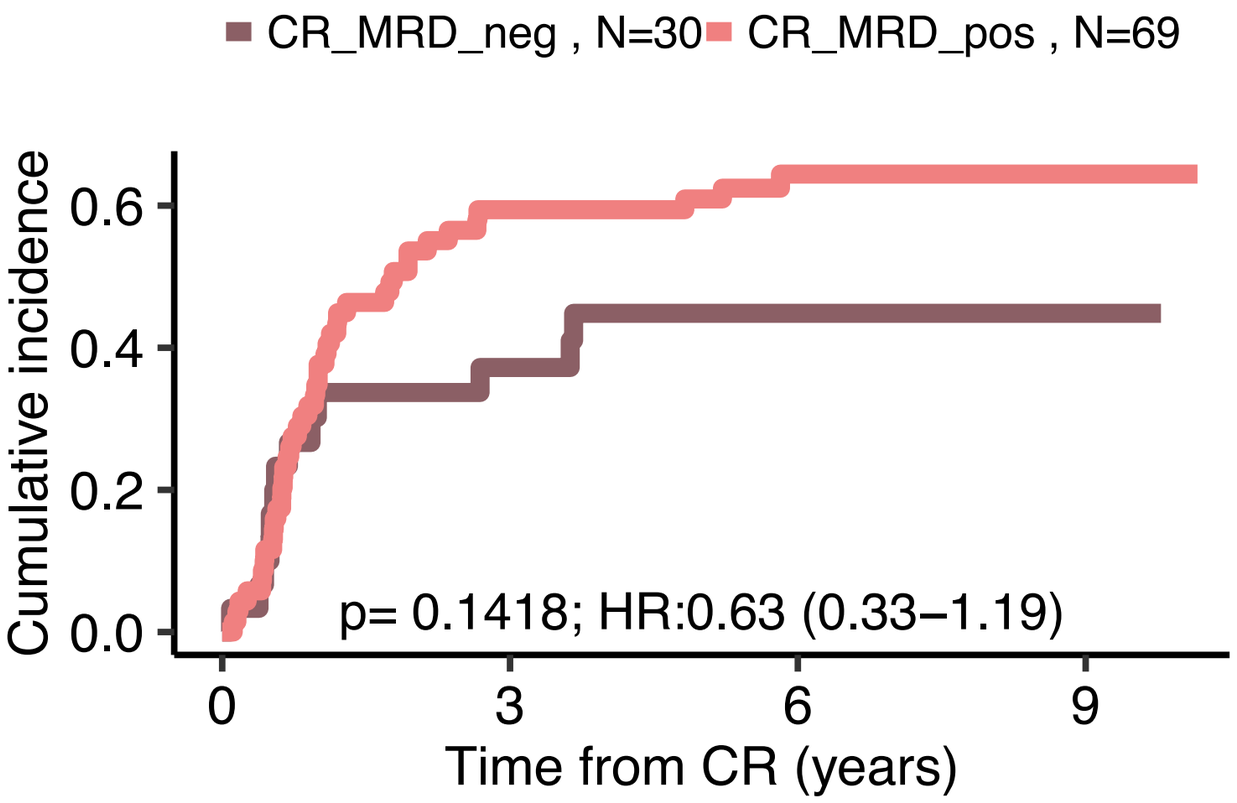
Relapse risk for NPM1



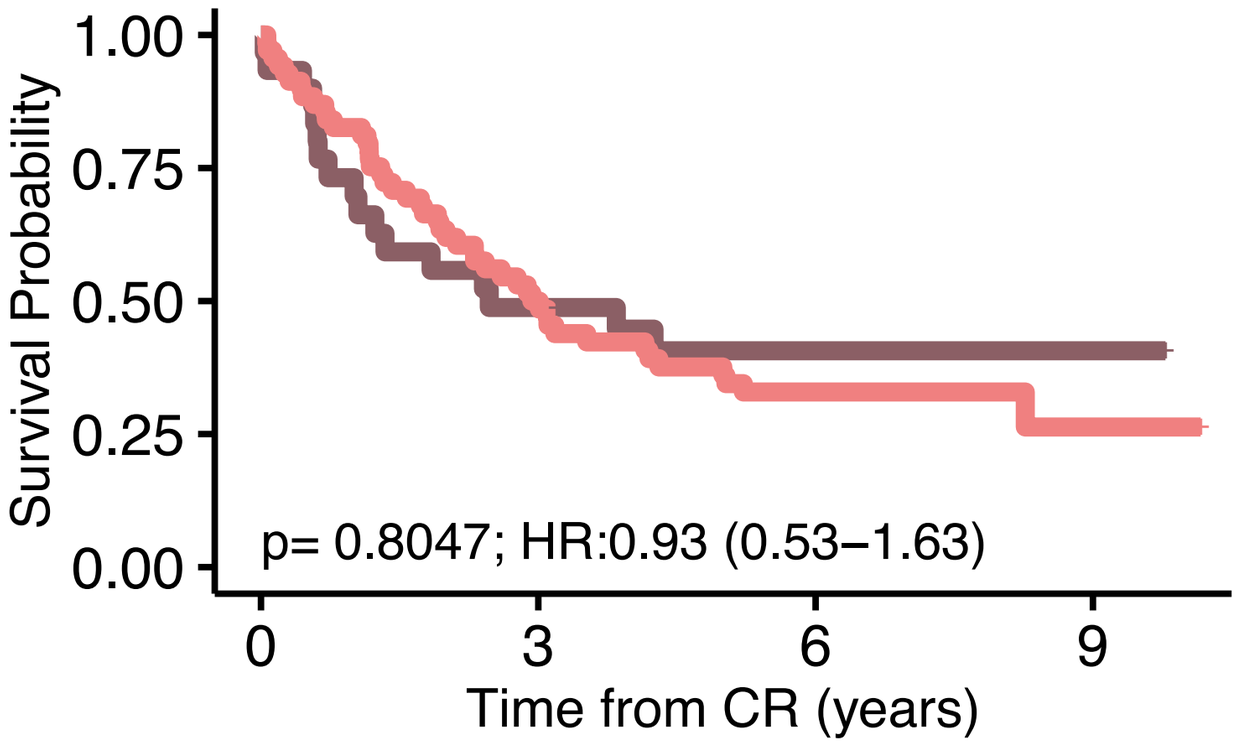
OS post CR for NPM1



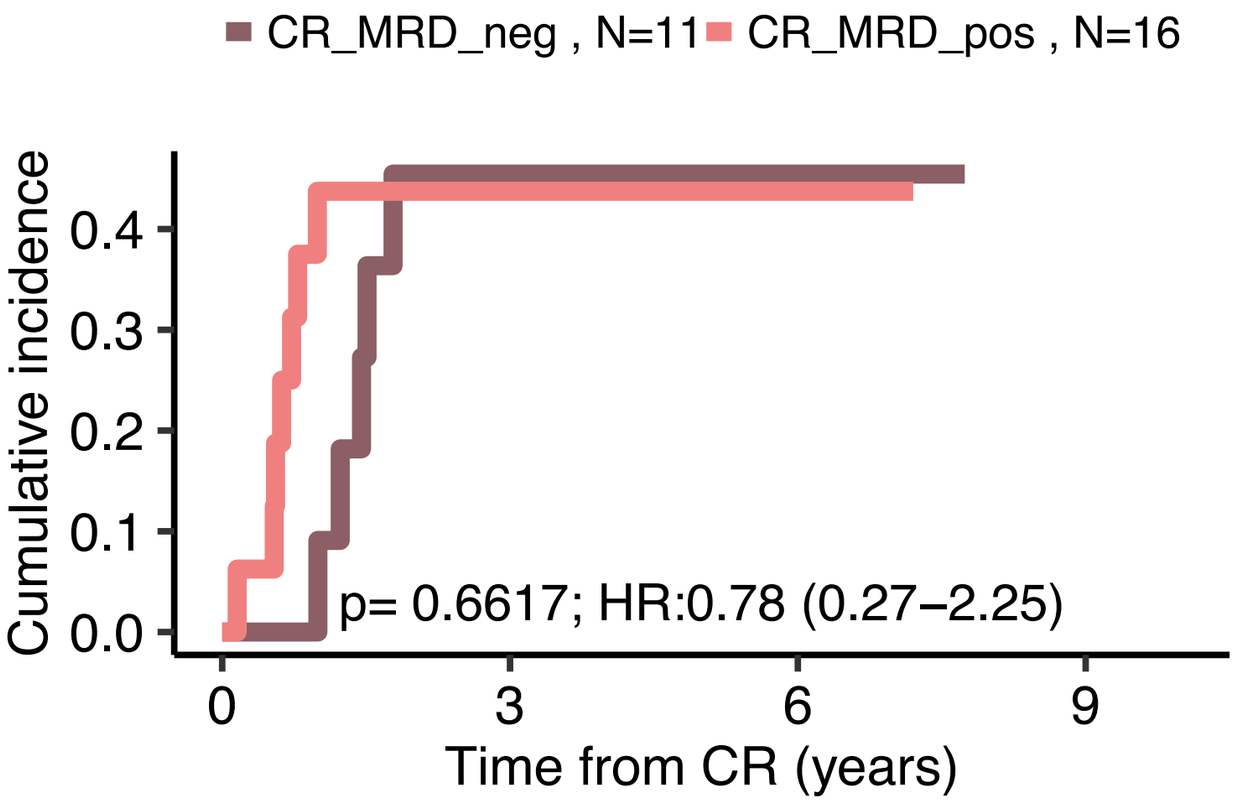
Relapse risk for sAML2



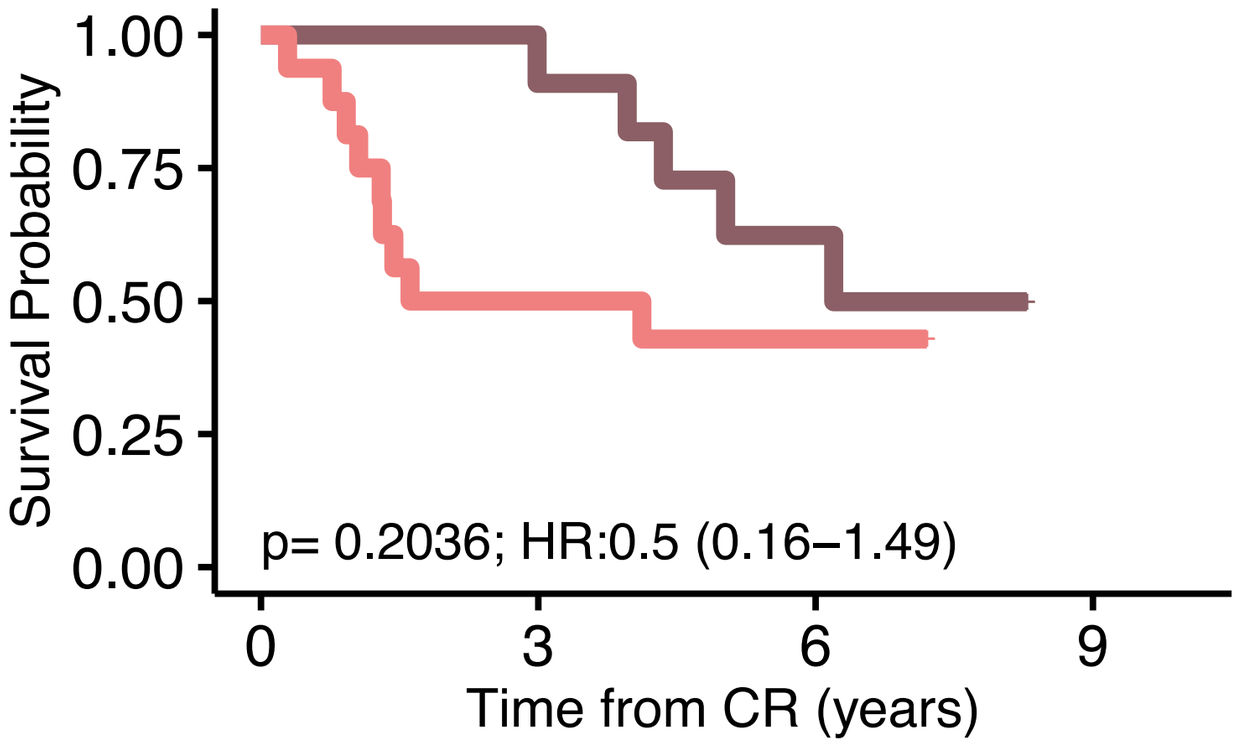
OS post CR for sAML2



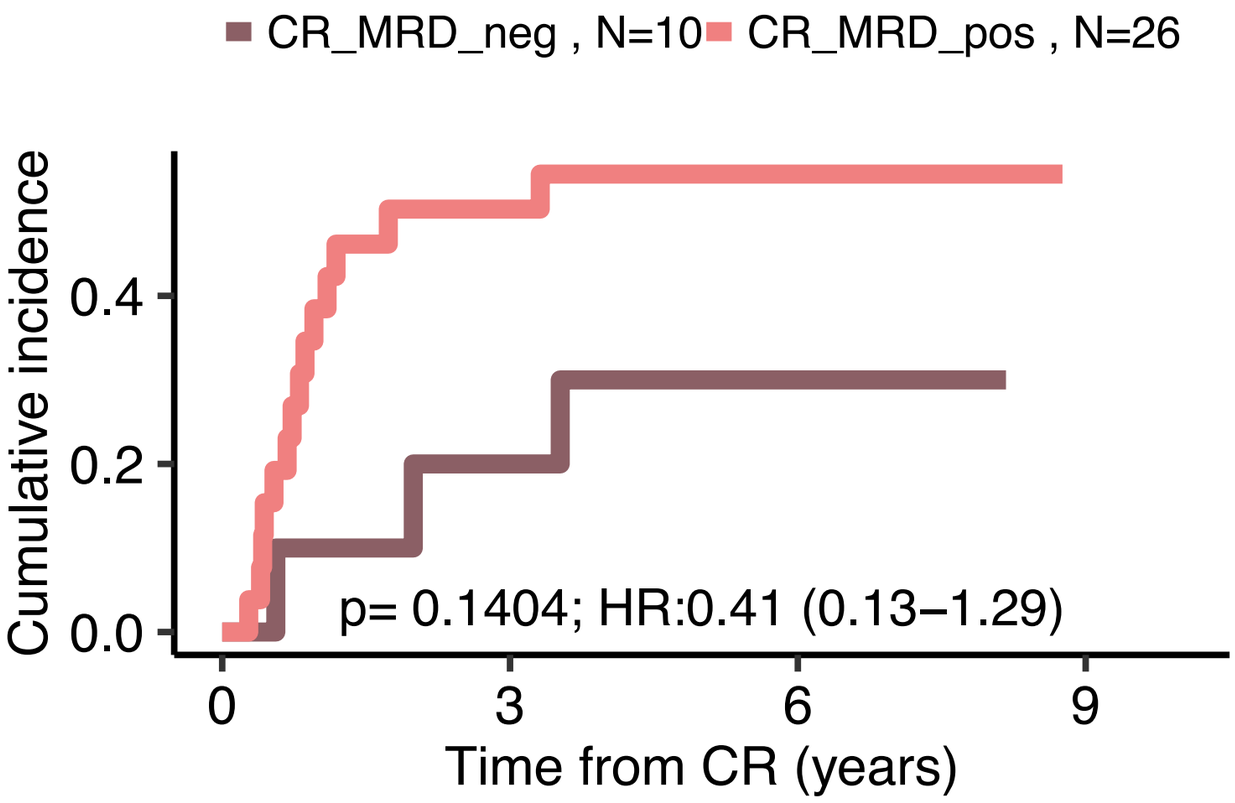
Relapse risk for sAML1



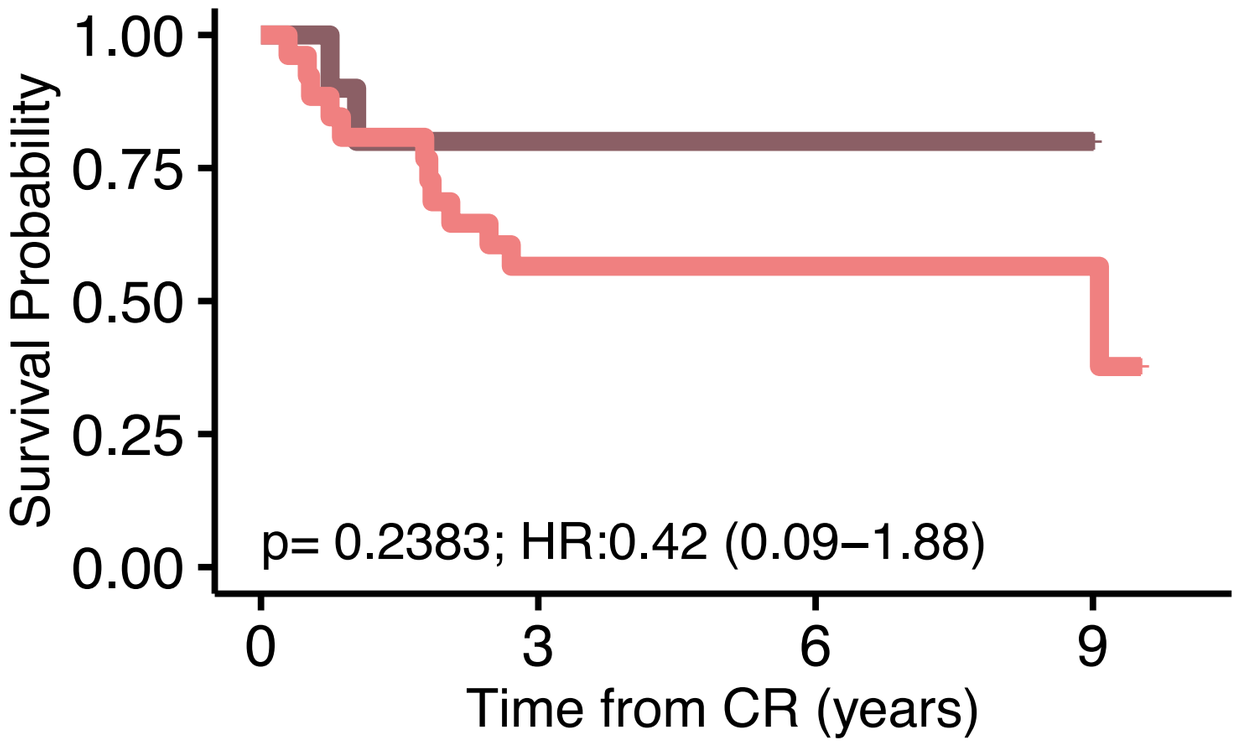
OS post CR for sAML1



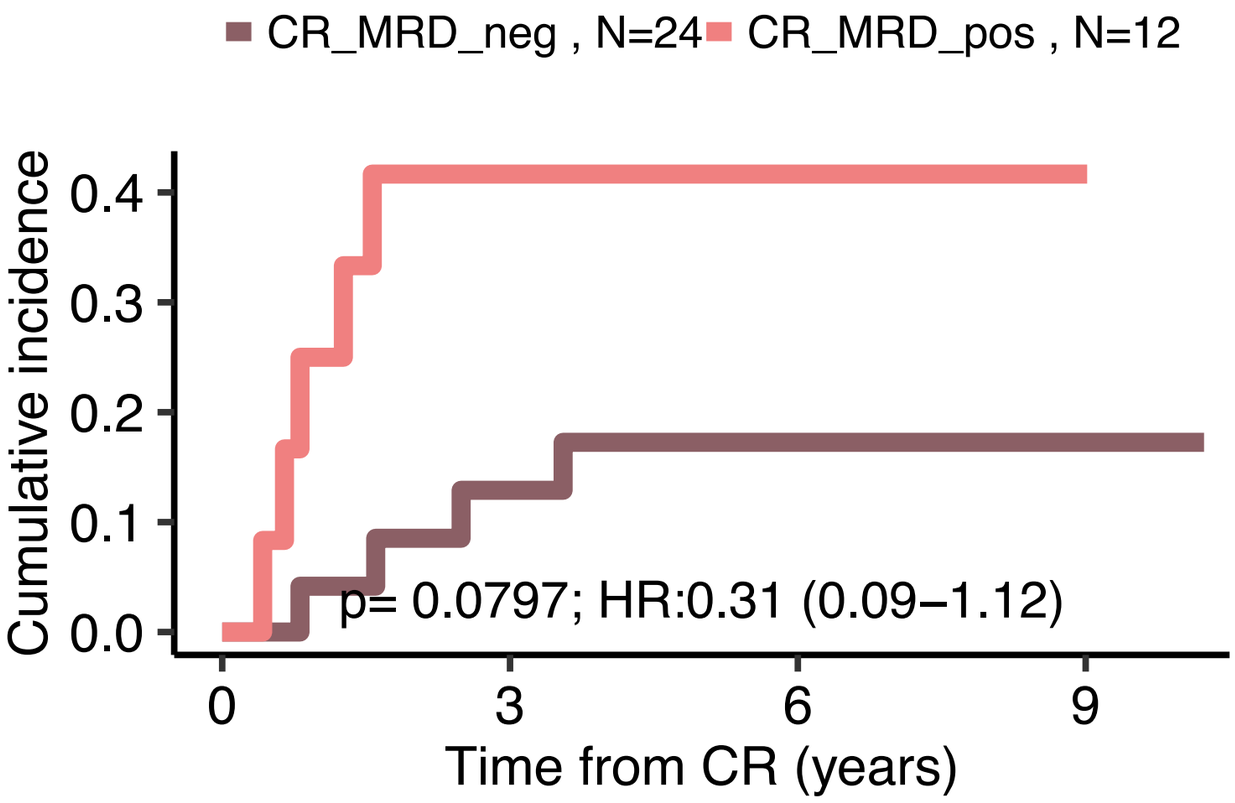
Relapse risk for mNOS



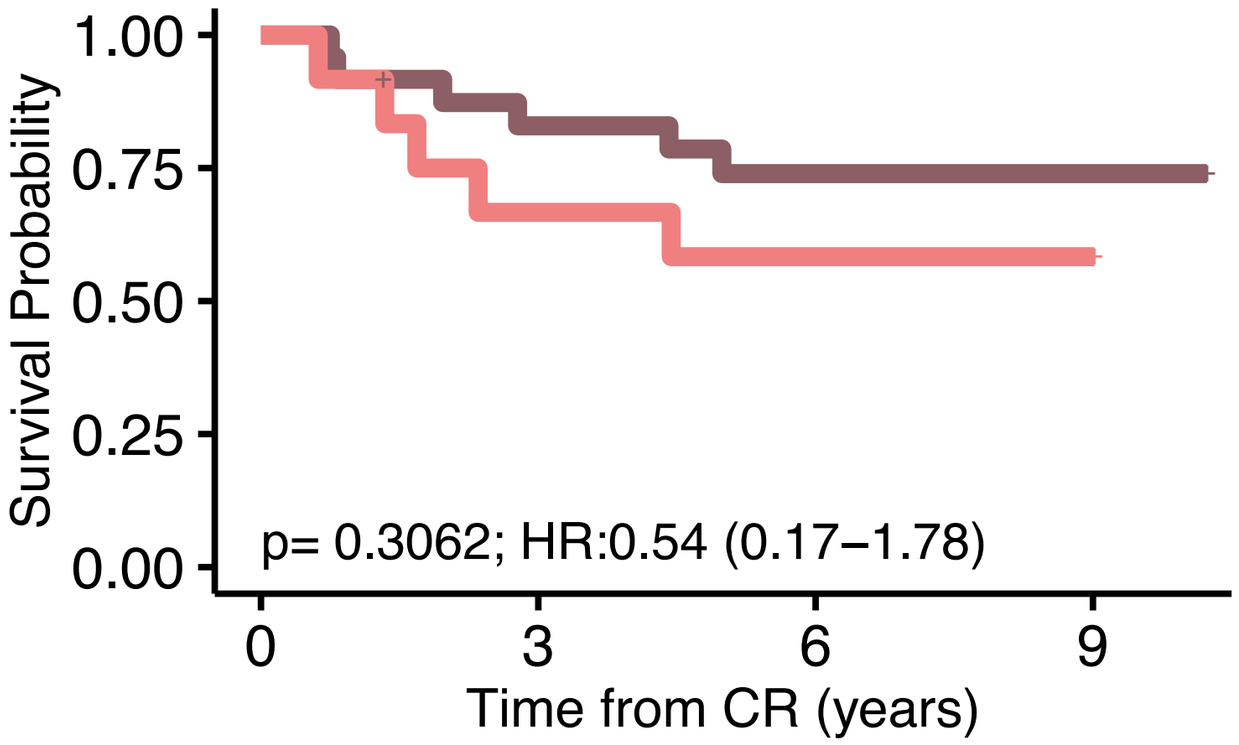
OS post CR for mNOS



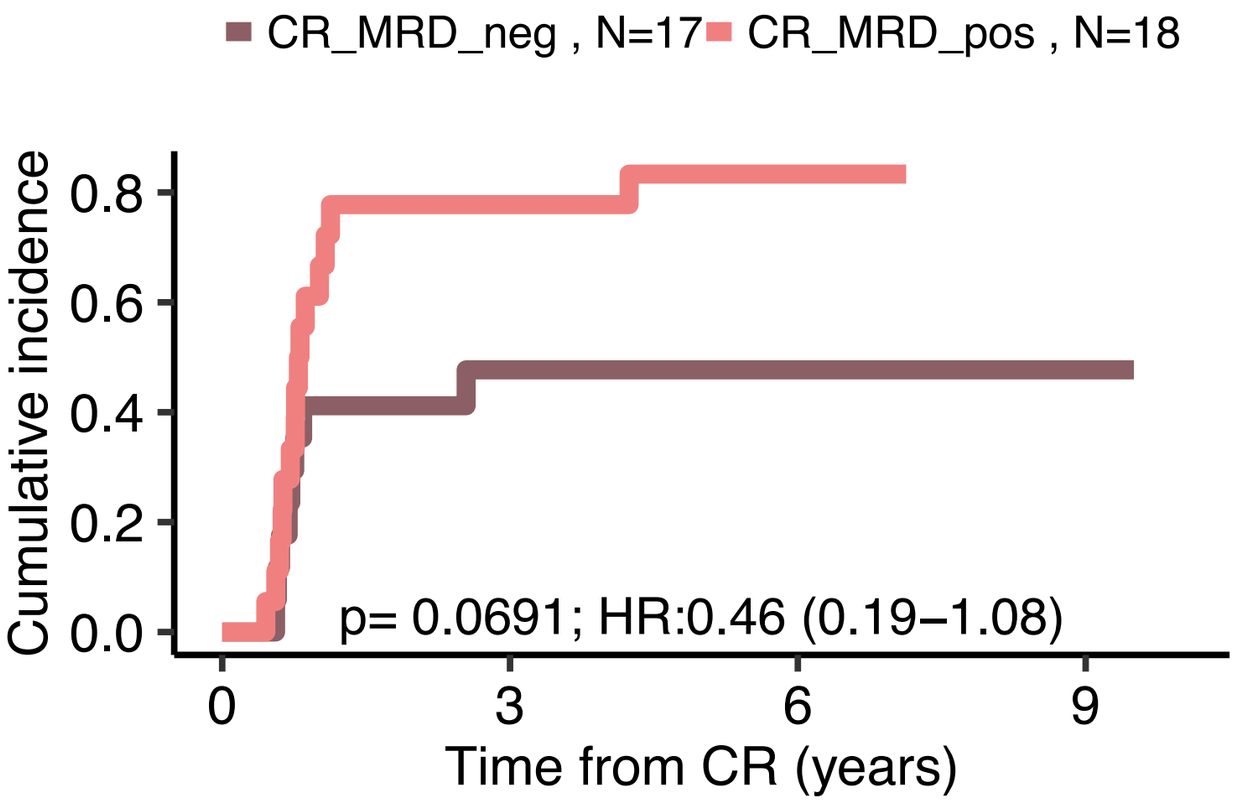
Relapse risk for t(8;21)



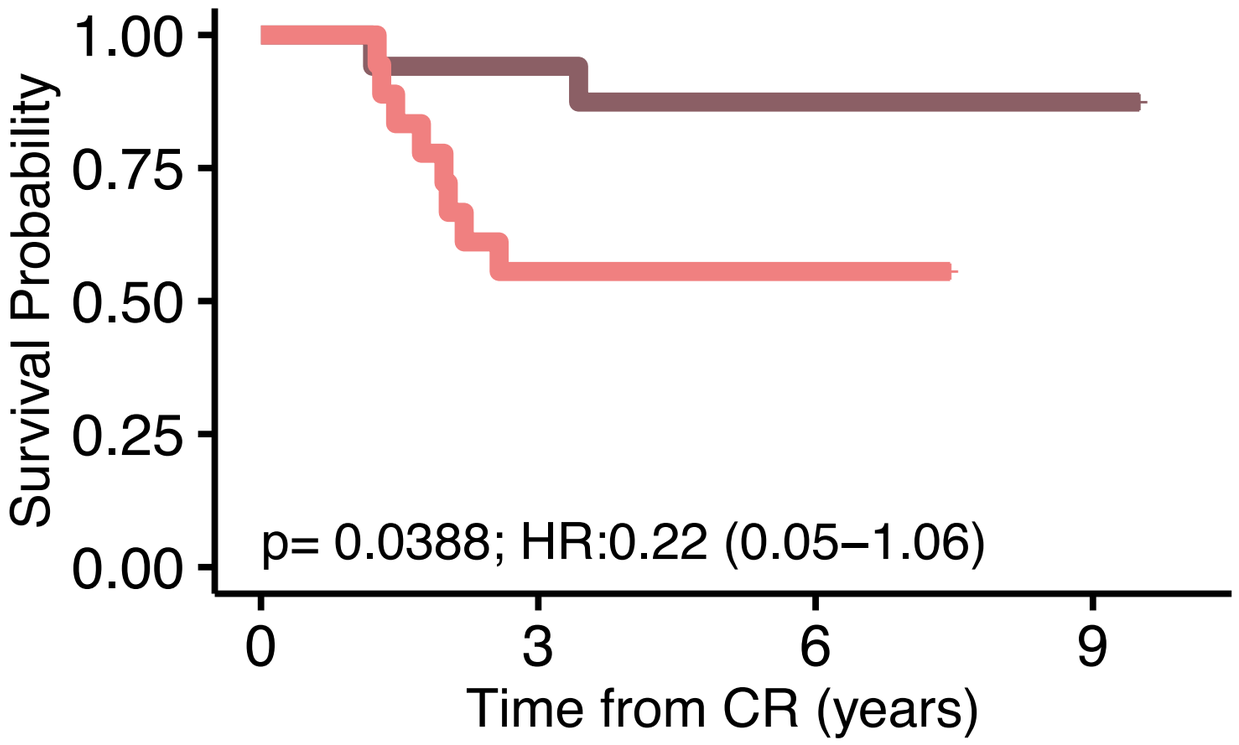
OS post CR for t(8;21)



Relapse risk for inv(16)

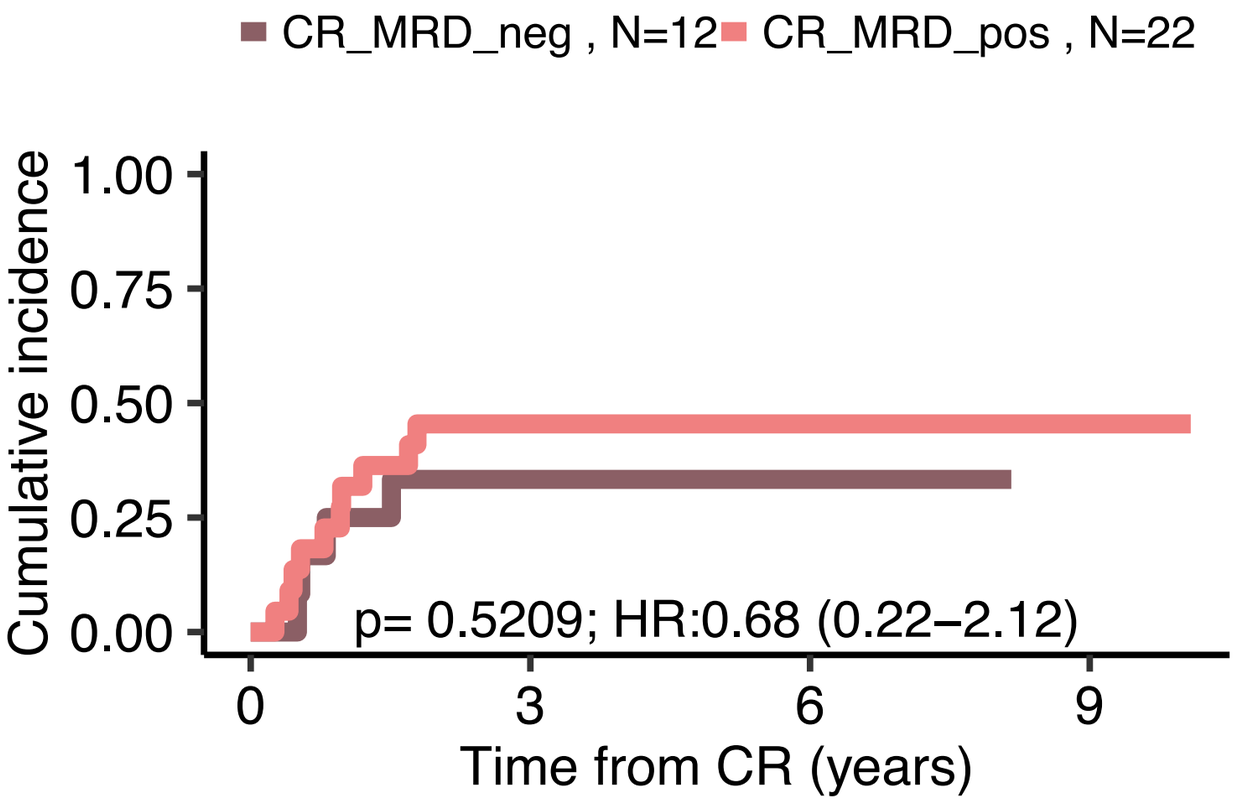


OS post CR for inv(16)

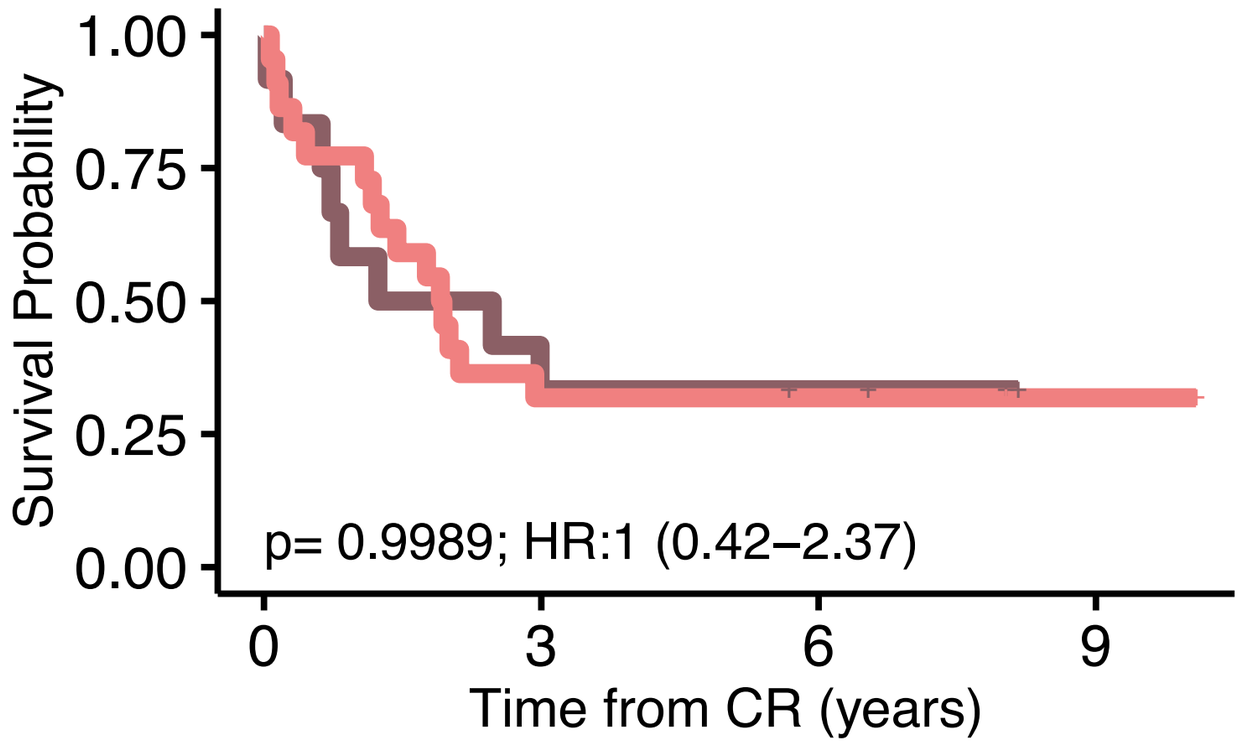


S.Figure 41: Summary by genes mutations and cytogenetics abnormalities for cumulative incidence and Kaplan-Meier curves for survival post complete remission on AML 17 NCRI Trial Cohort with post course 1 minimal residual disease (MRD) analysis (n=523). The first panel represents the cumulative incidence of relapse for patients in that specific event stratified by MRD status (p values were computed with Gray test) and the second panel represents the Kaplan-Meier curve for survival post course 1 complete remission stratified by MRD status. We omitted events with less than 5 patients with MRD positive or negative. Panels where comparator group has less than 10 patients have been removed for robust statistical interpretation. Two-sided Pvalues were computed with Gray's test and log-rank test.

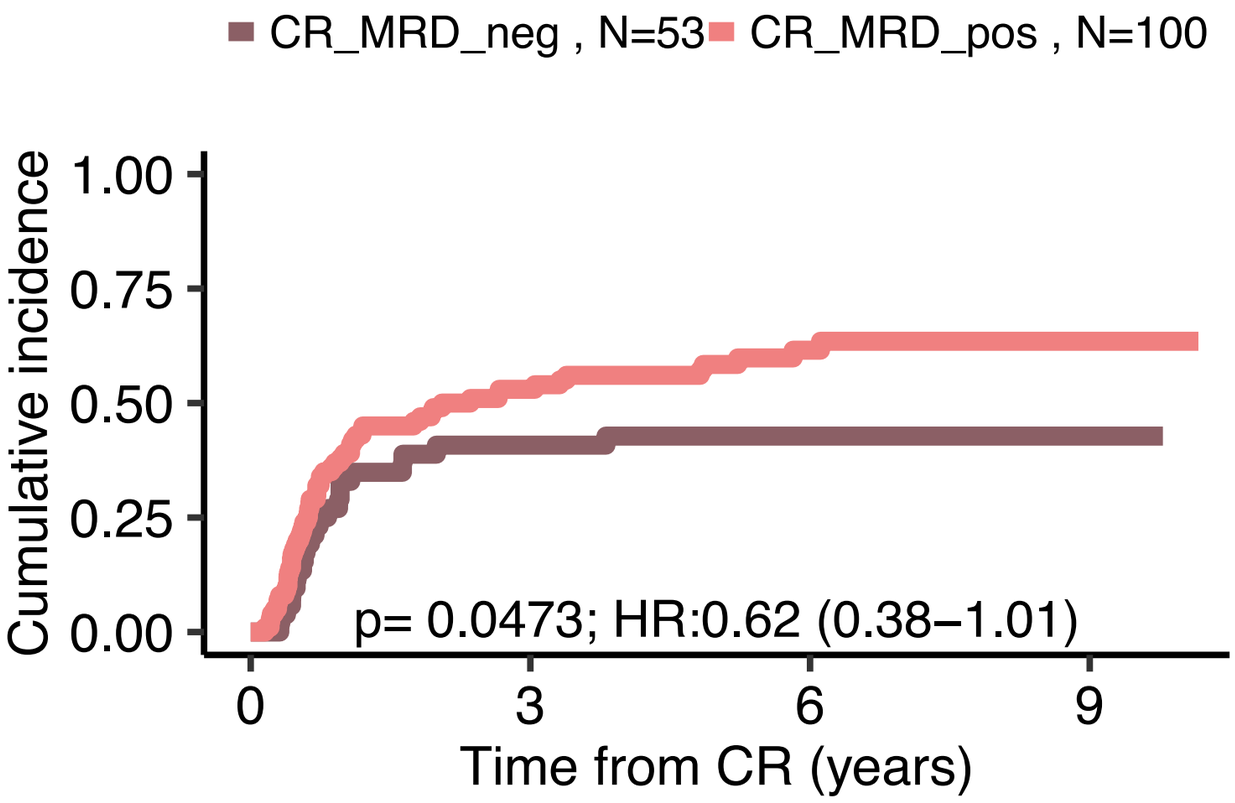
Relapse risk for ASXL1



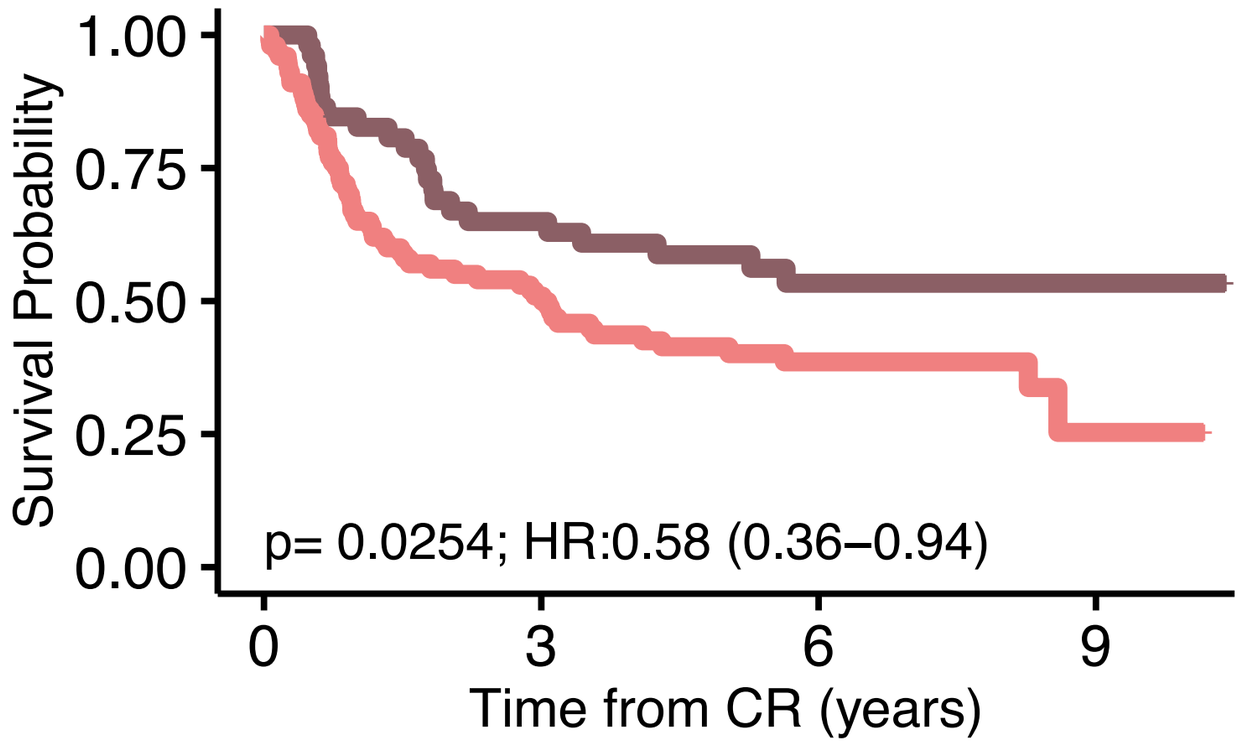
OS CR for ASXL1



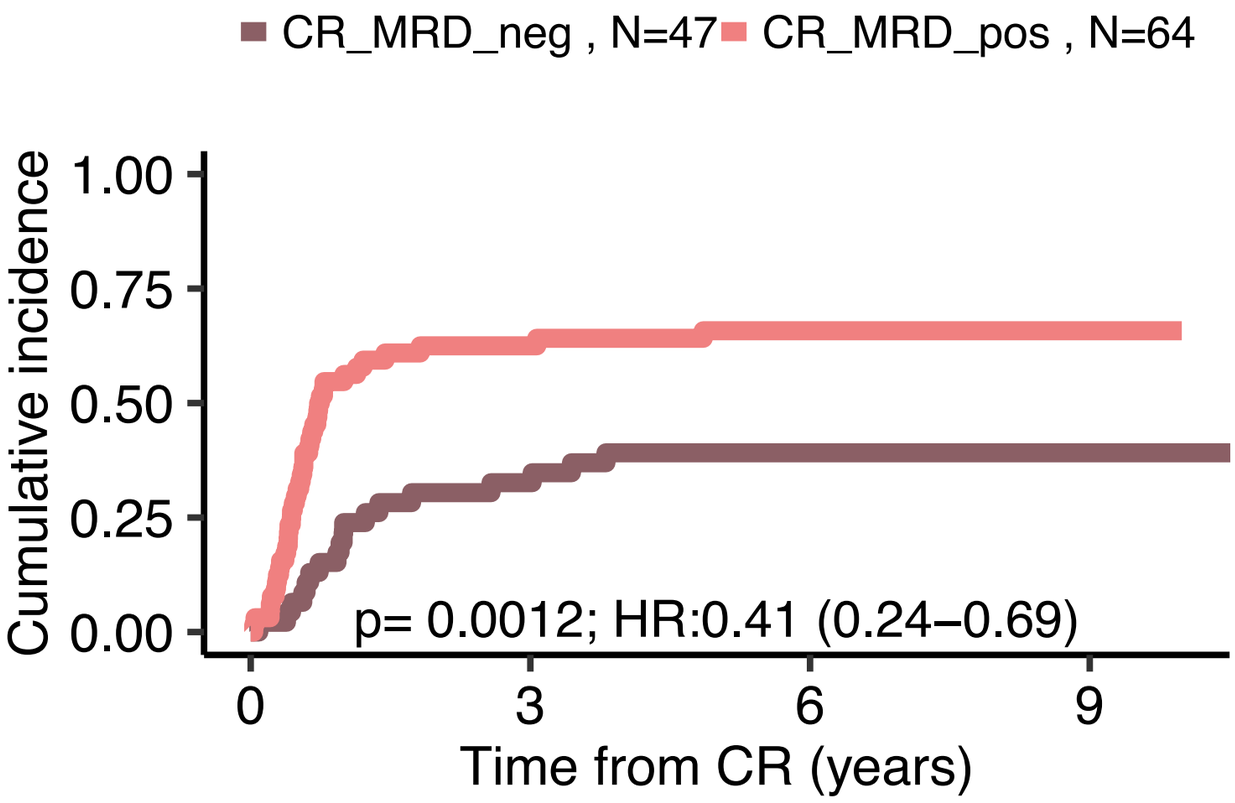
Relapse risk for DNMT3A



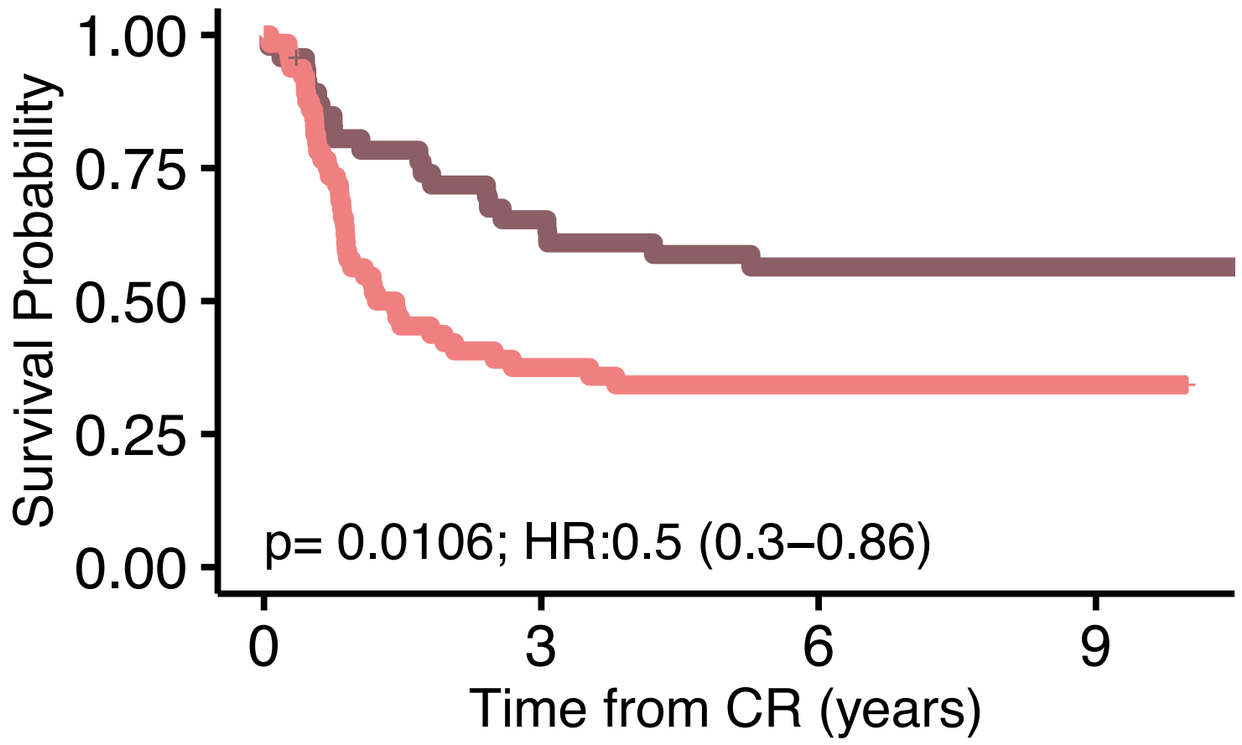
OS CR for DNMT3A



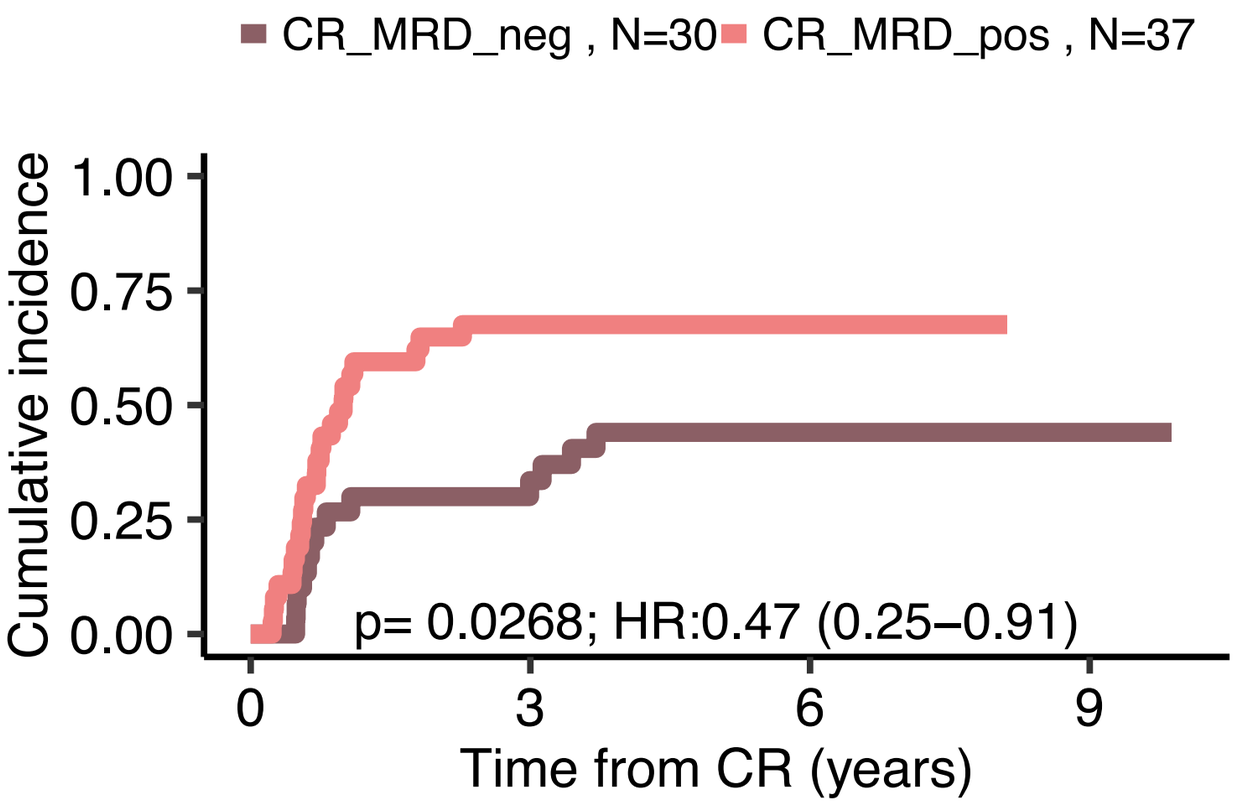
Relapse risk for ITD



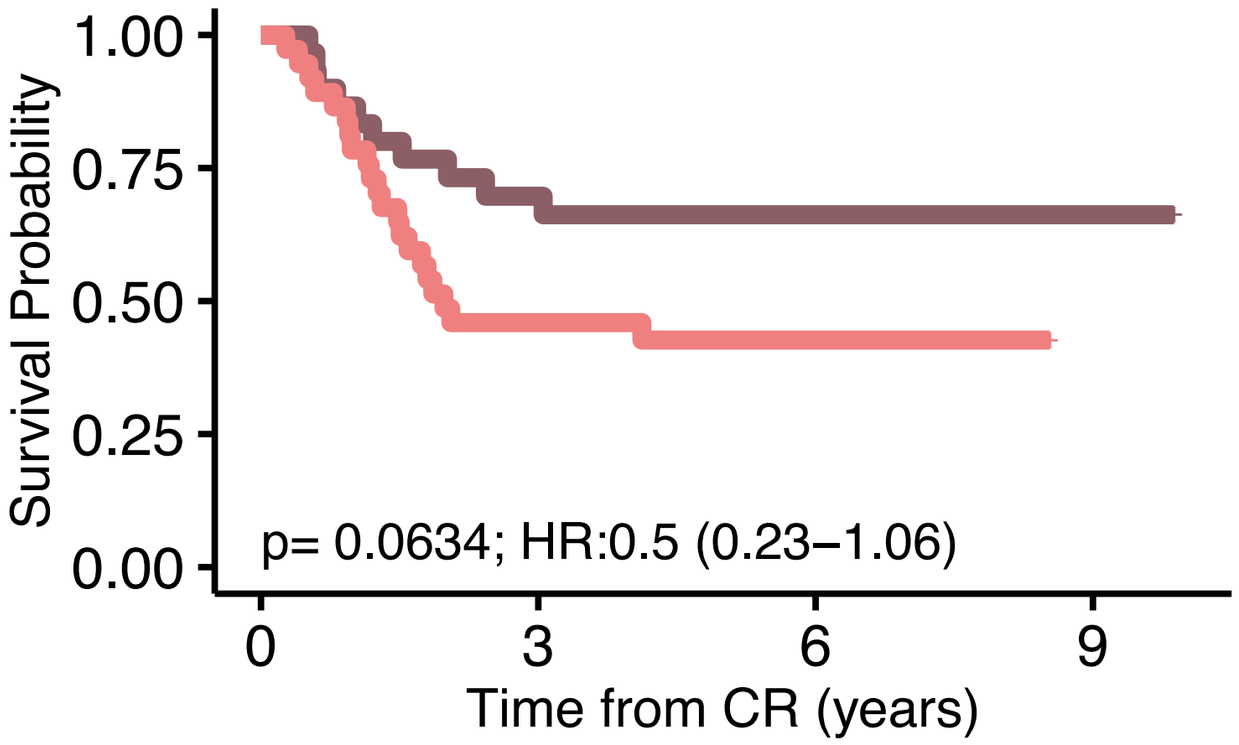
OS CR for ITD



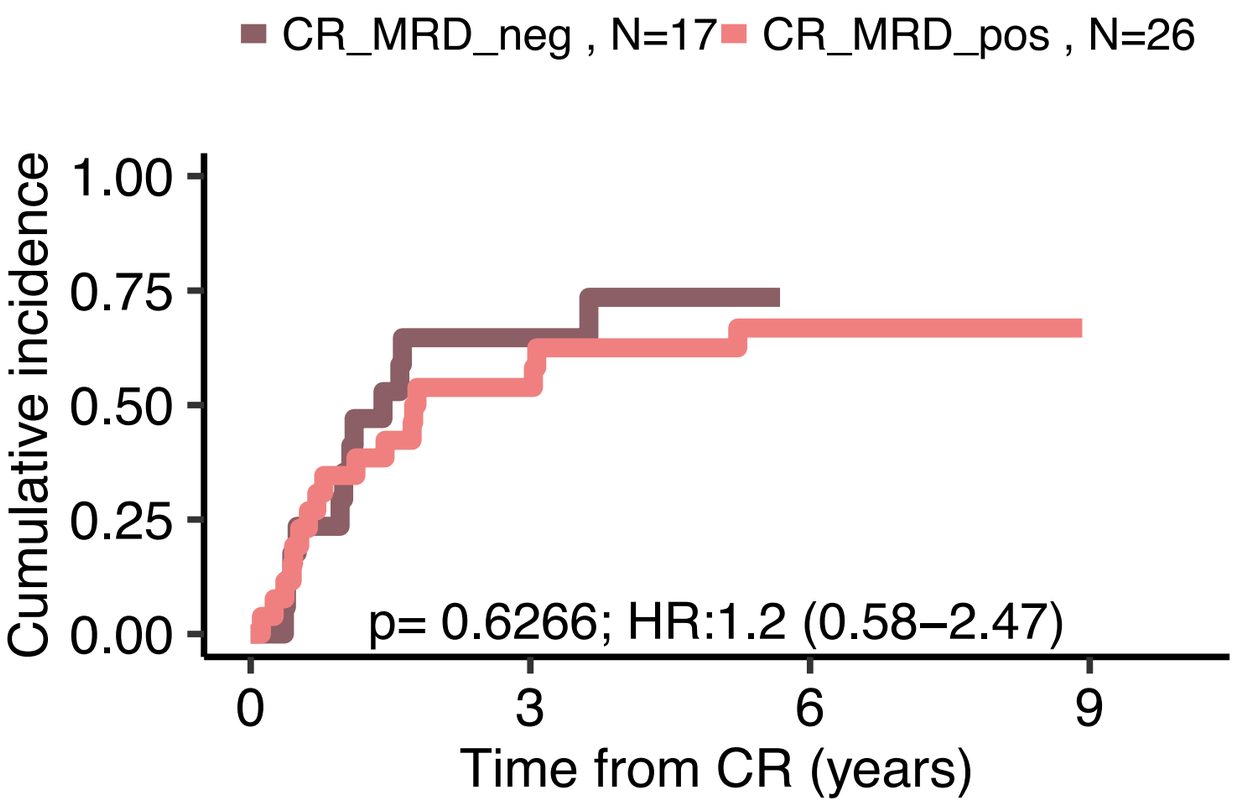
Relapse risk for FLT3_TKD



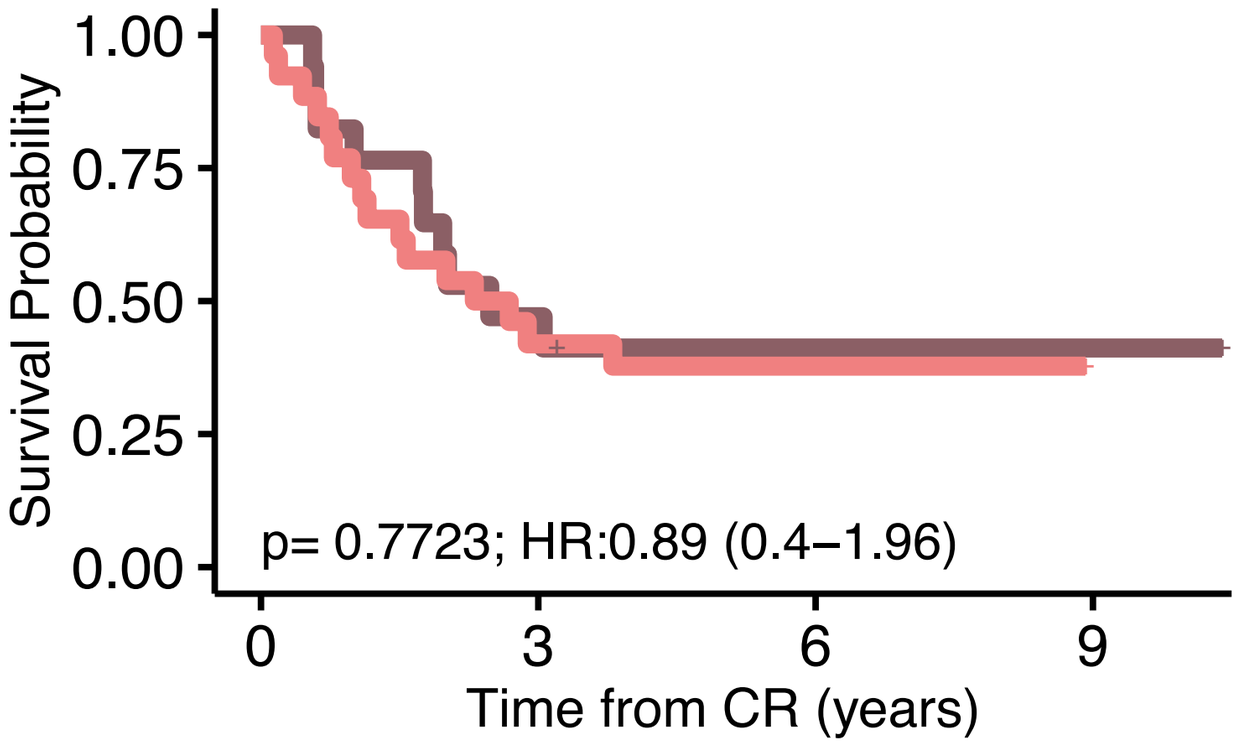
OS CR for FLT3_TKD



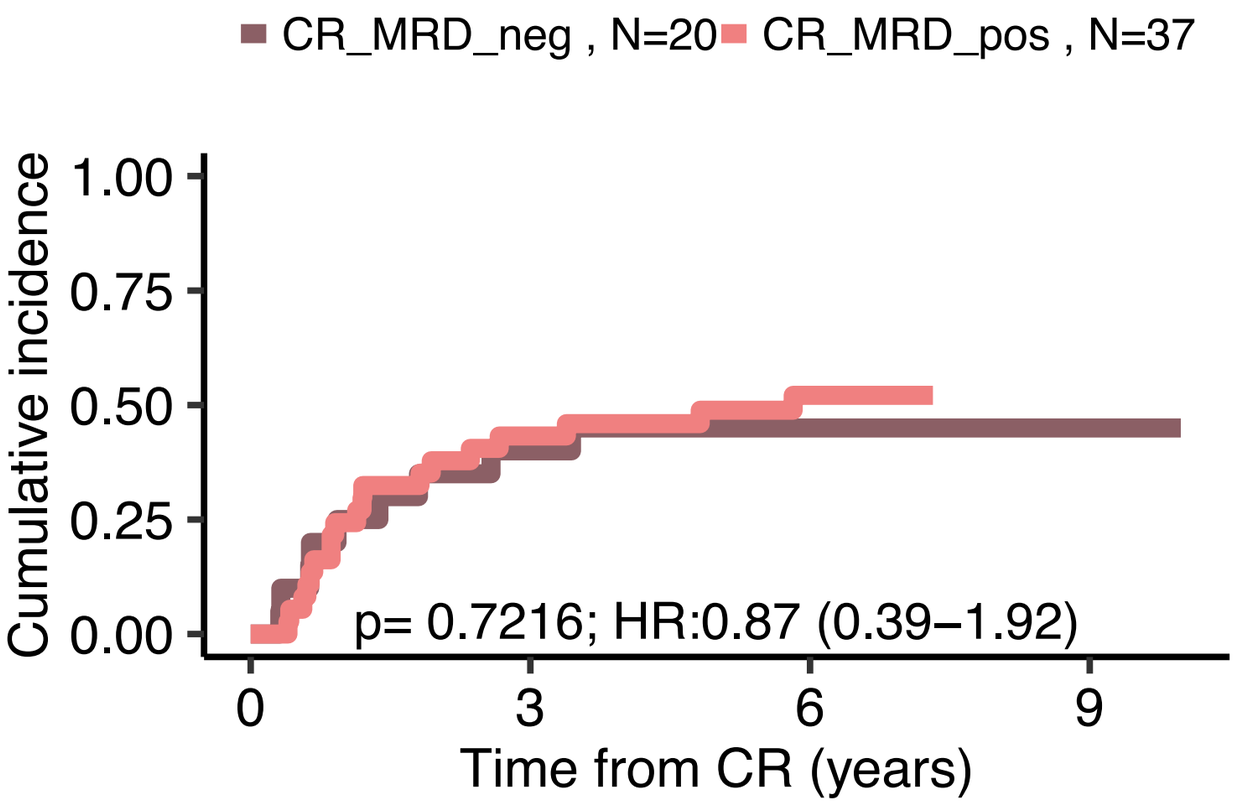
Relapse risk for IDH1



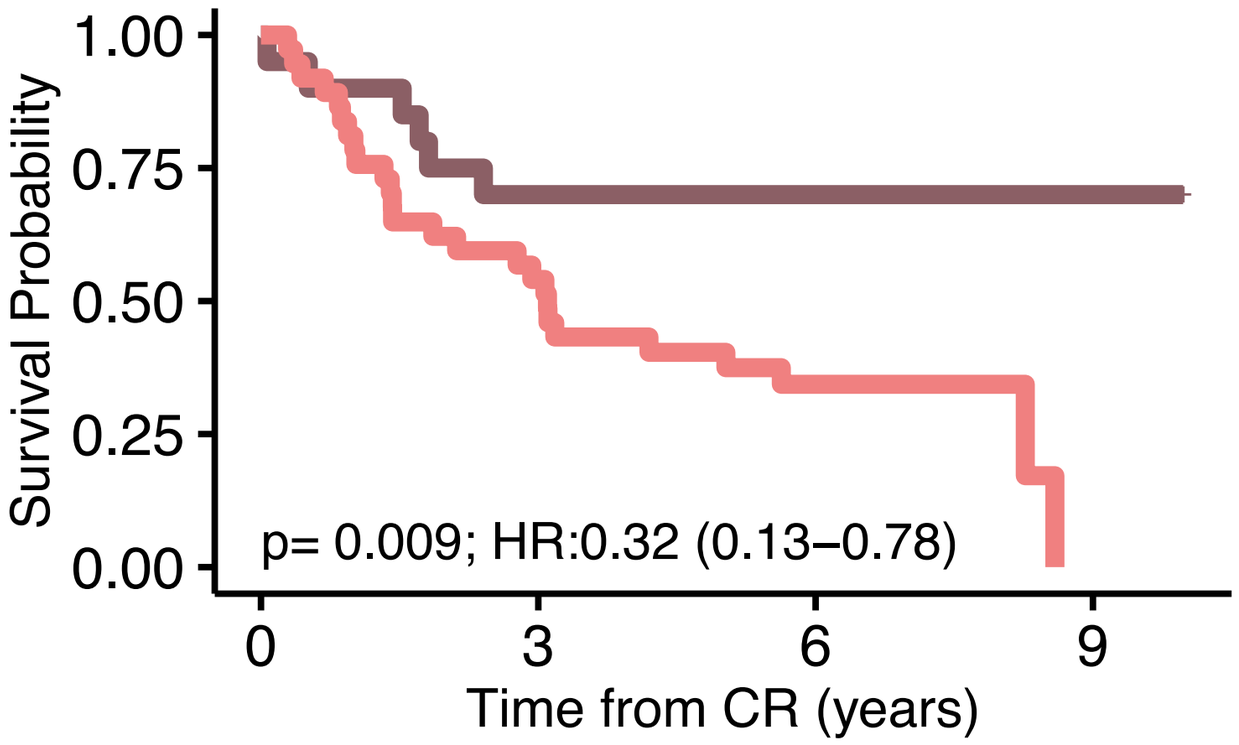
OS CR for IDH1



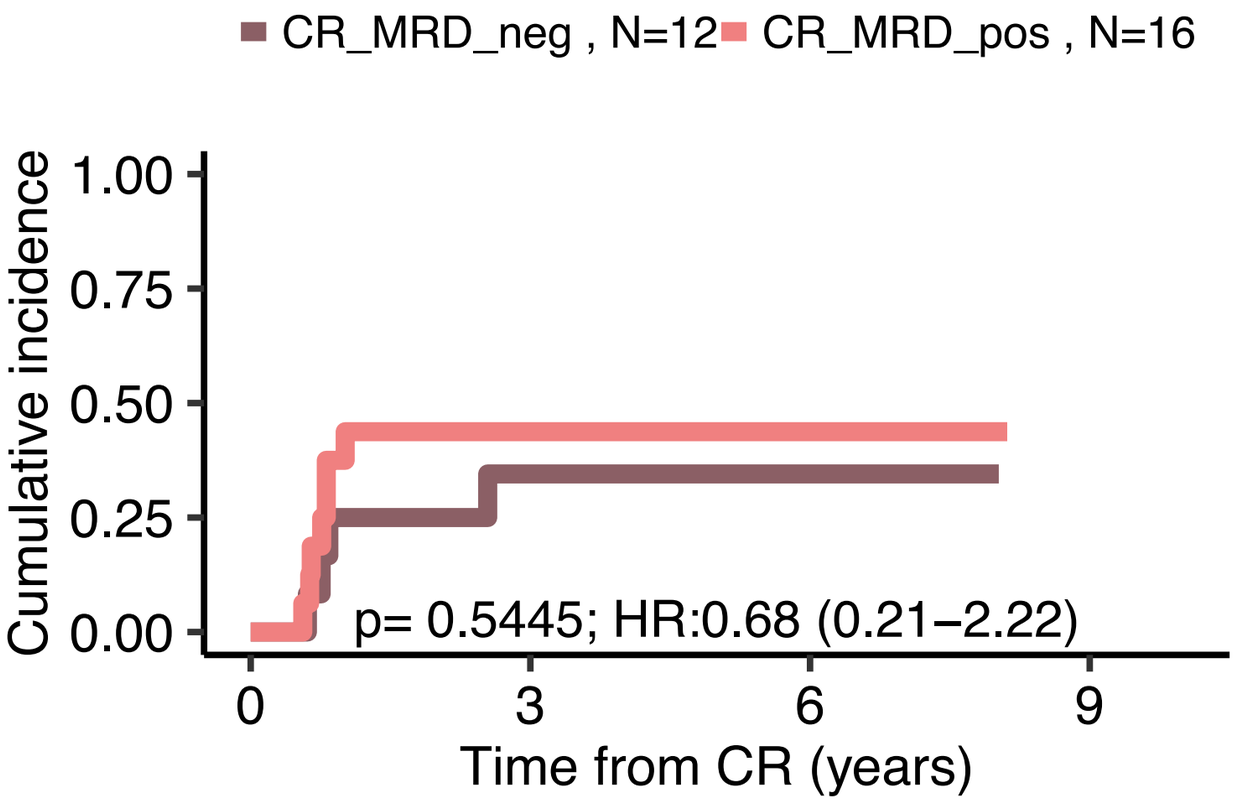
Relapse risk for IDH2_p.R140



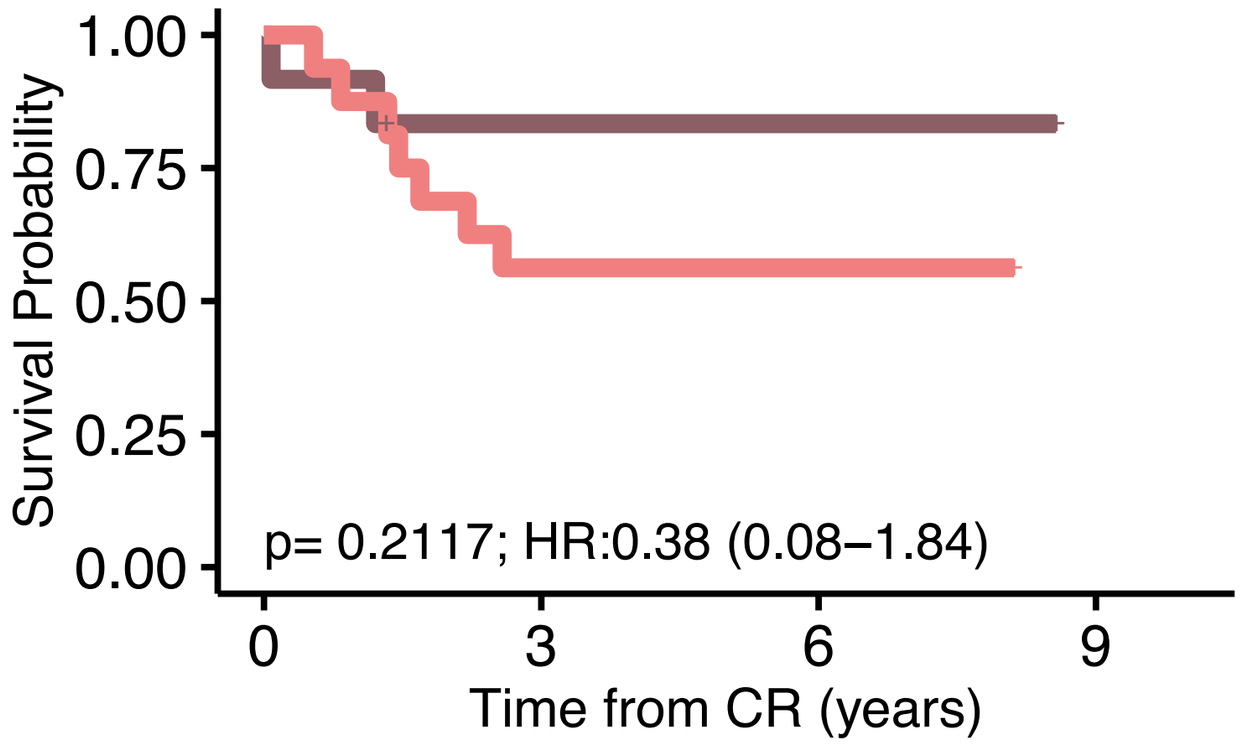
OS CR for IDH2_p.R140



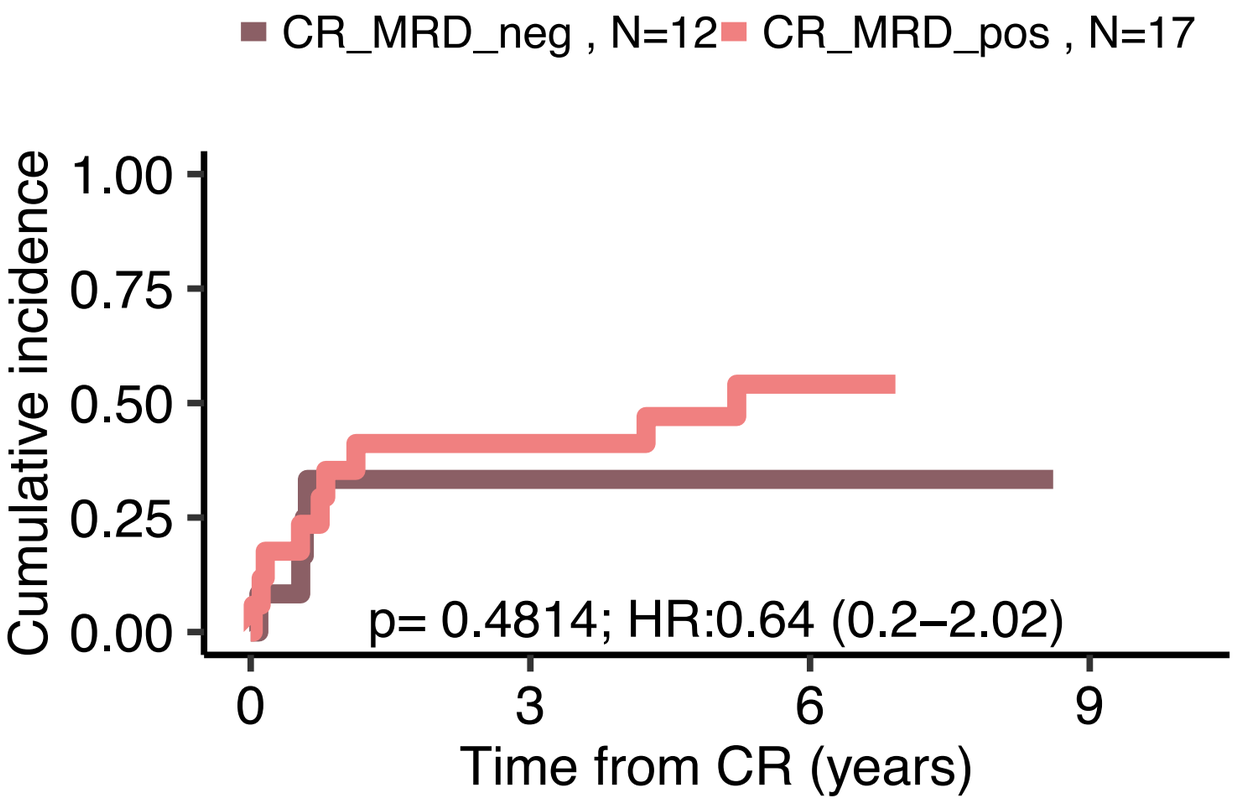
Relapse risk for KIT



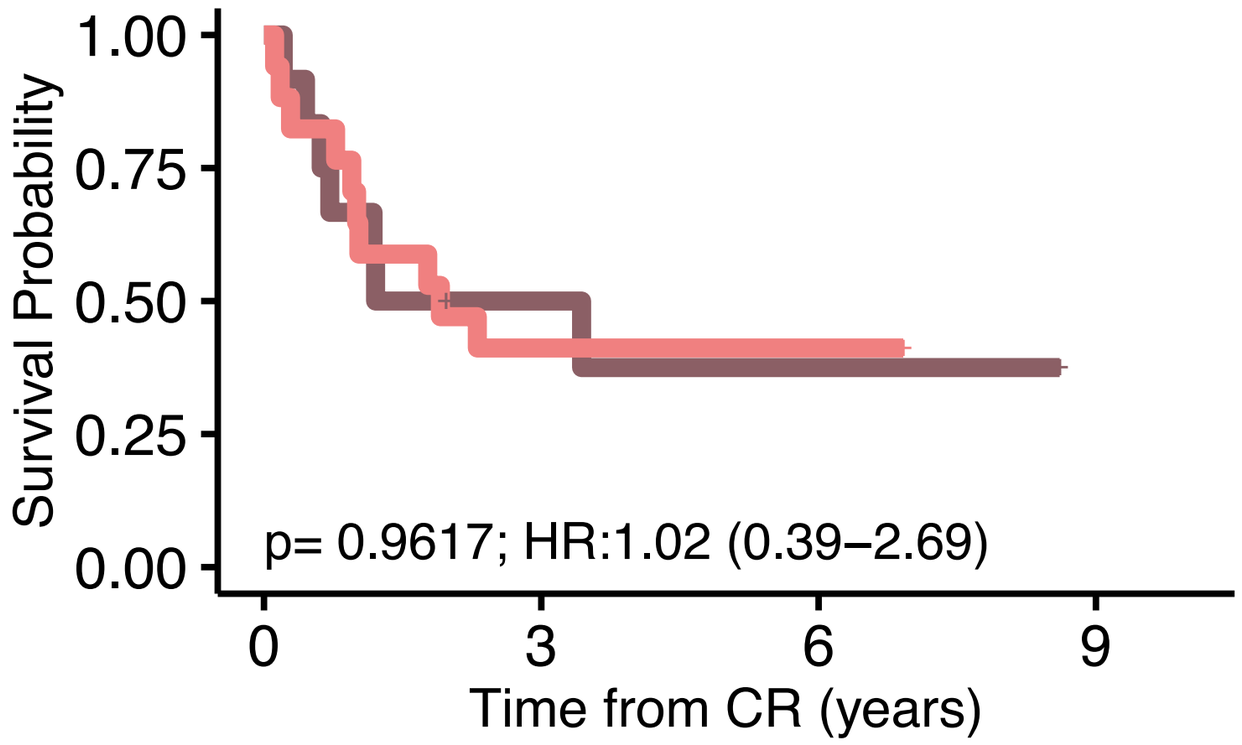
OS CR for KIT



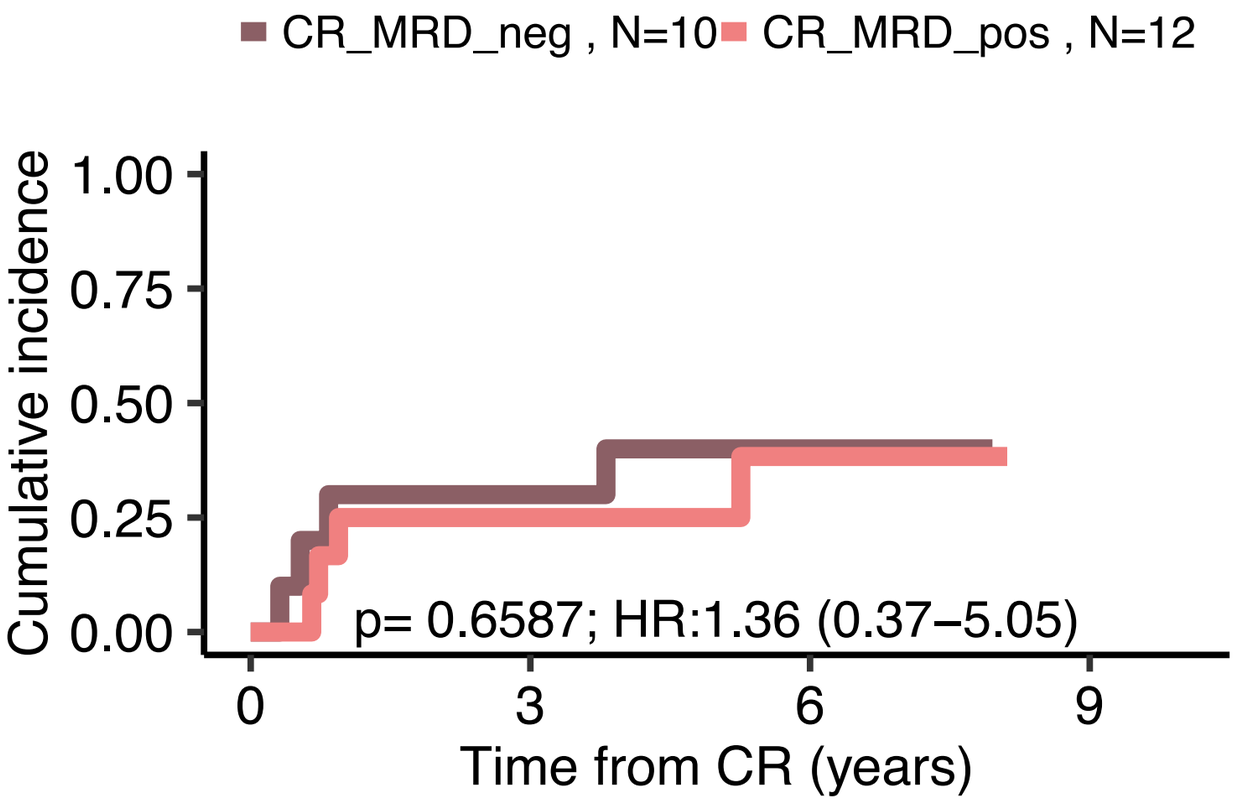
Relapse risk for KRAS



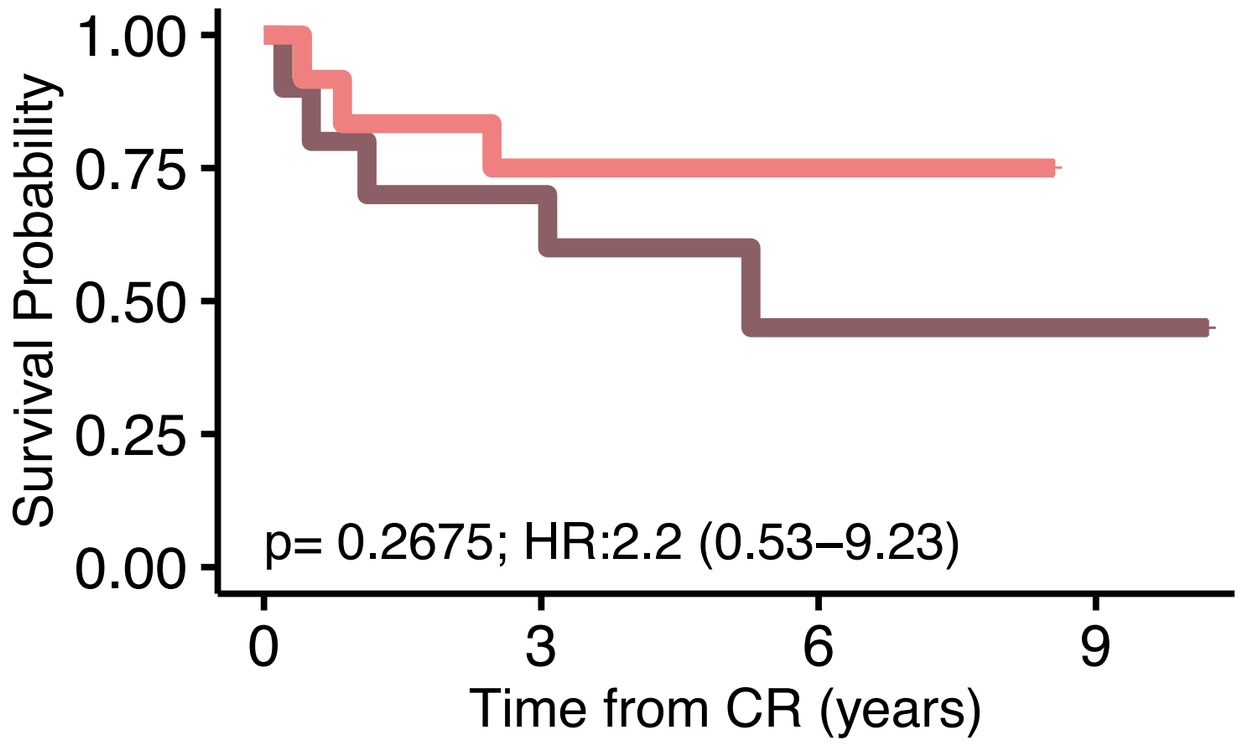
OS CR for KRAS



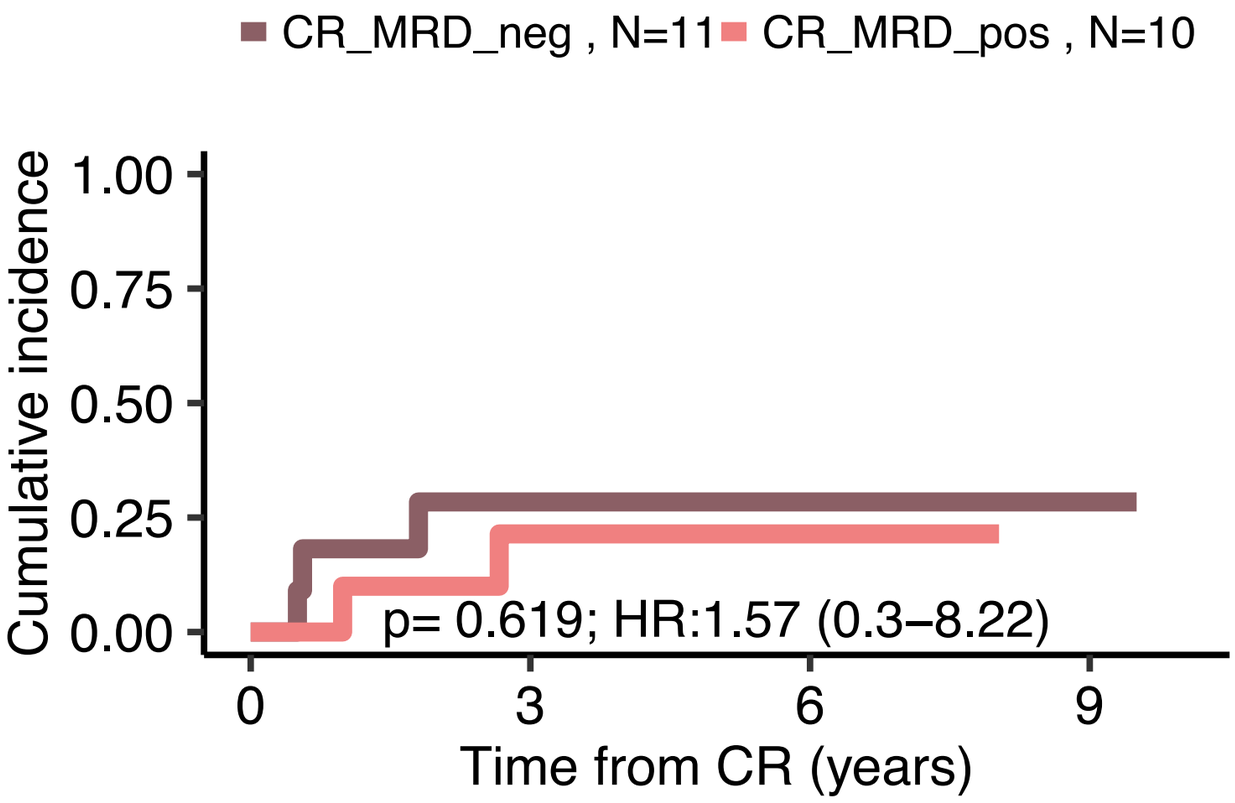
Relapse risk for MYC



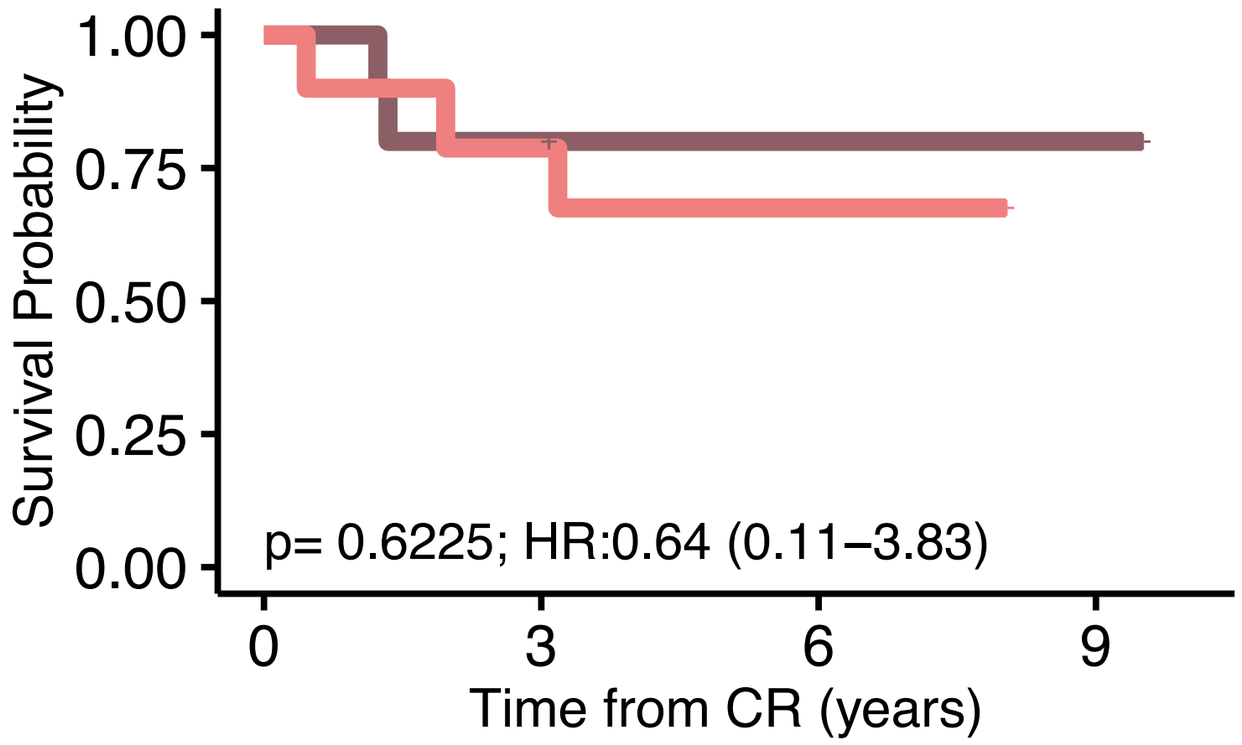
OS CR for MYC



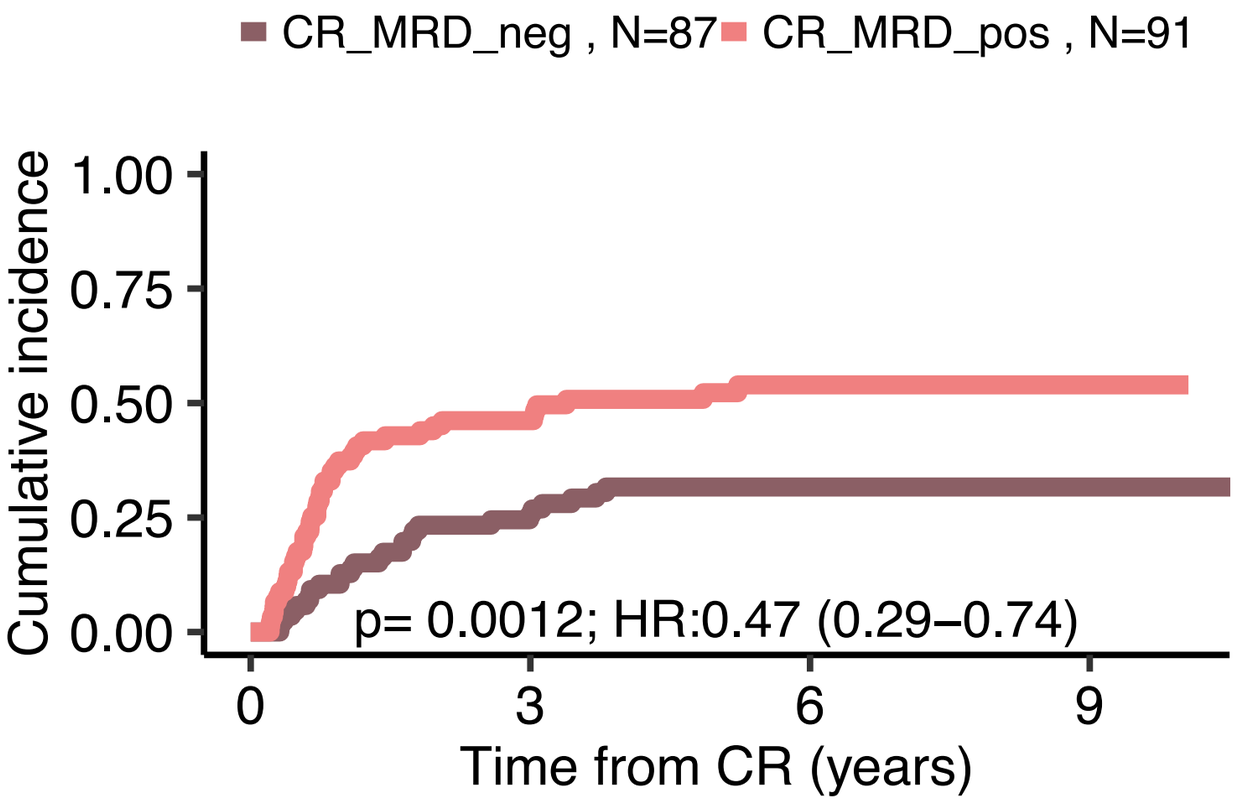
Relapse risk for NF1



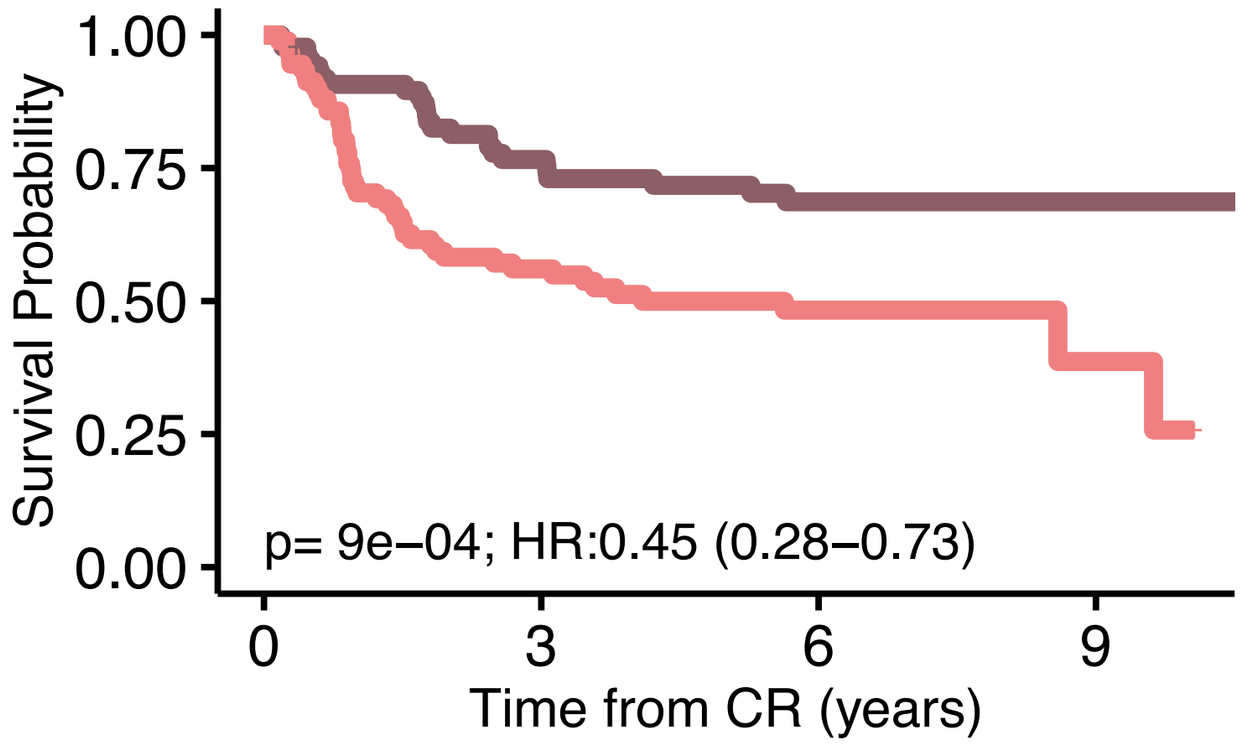
OS CR for NF1



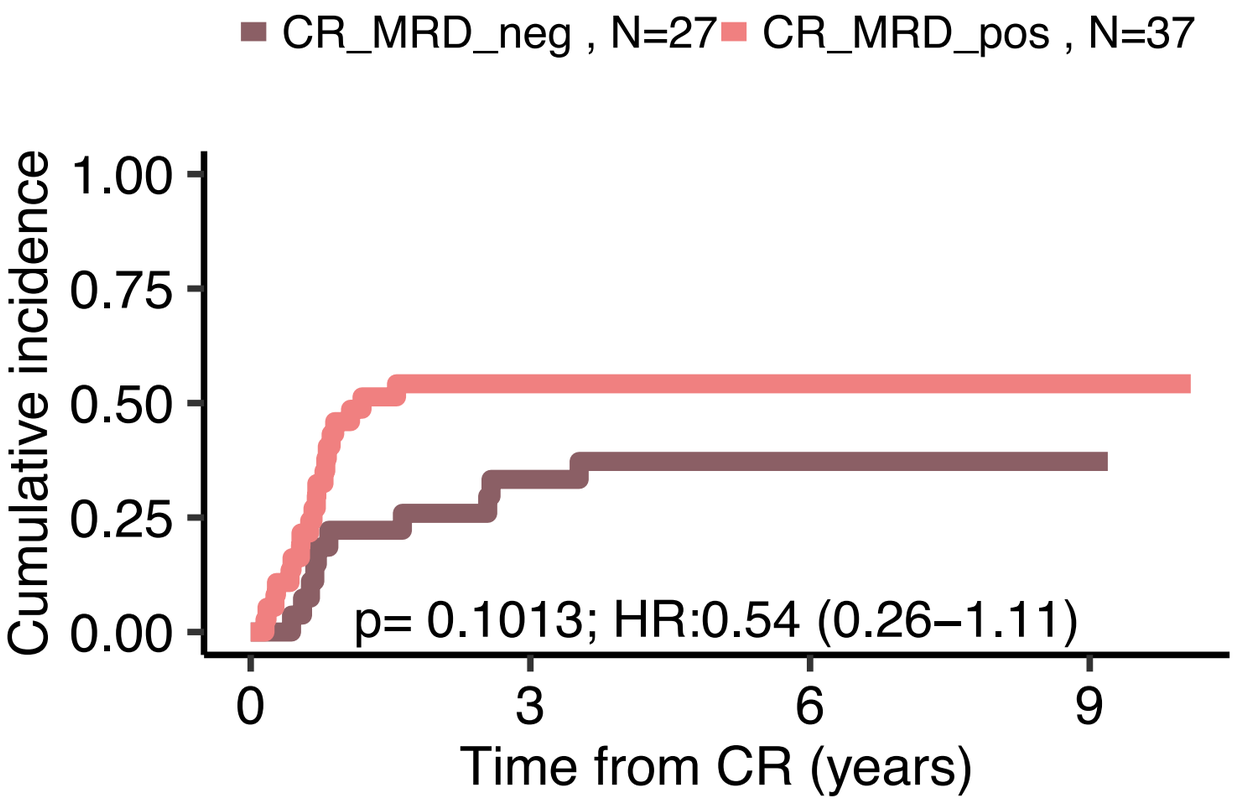
Relapse risk for NPM1



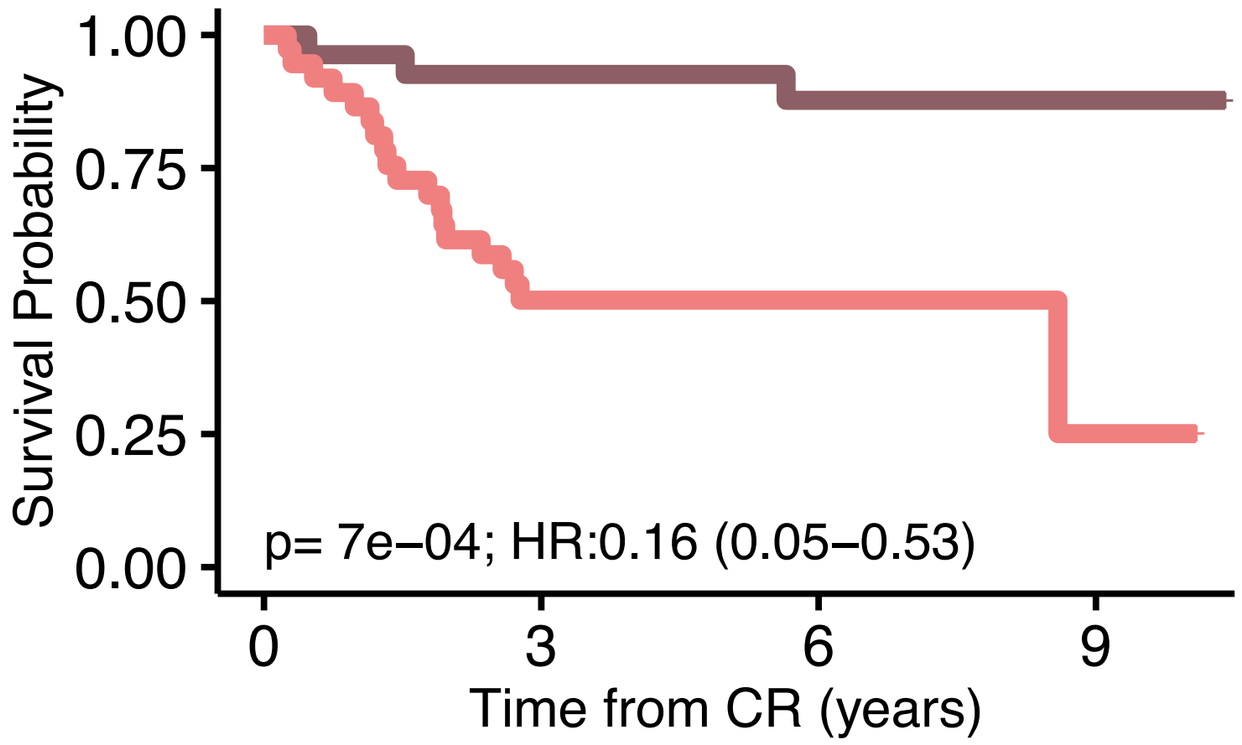
OS CR for NPM1



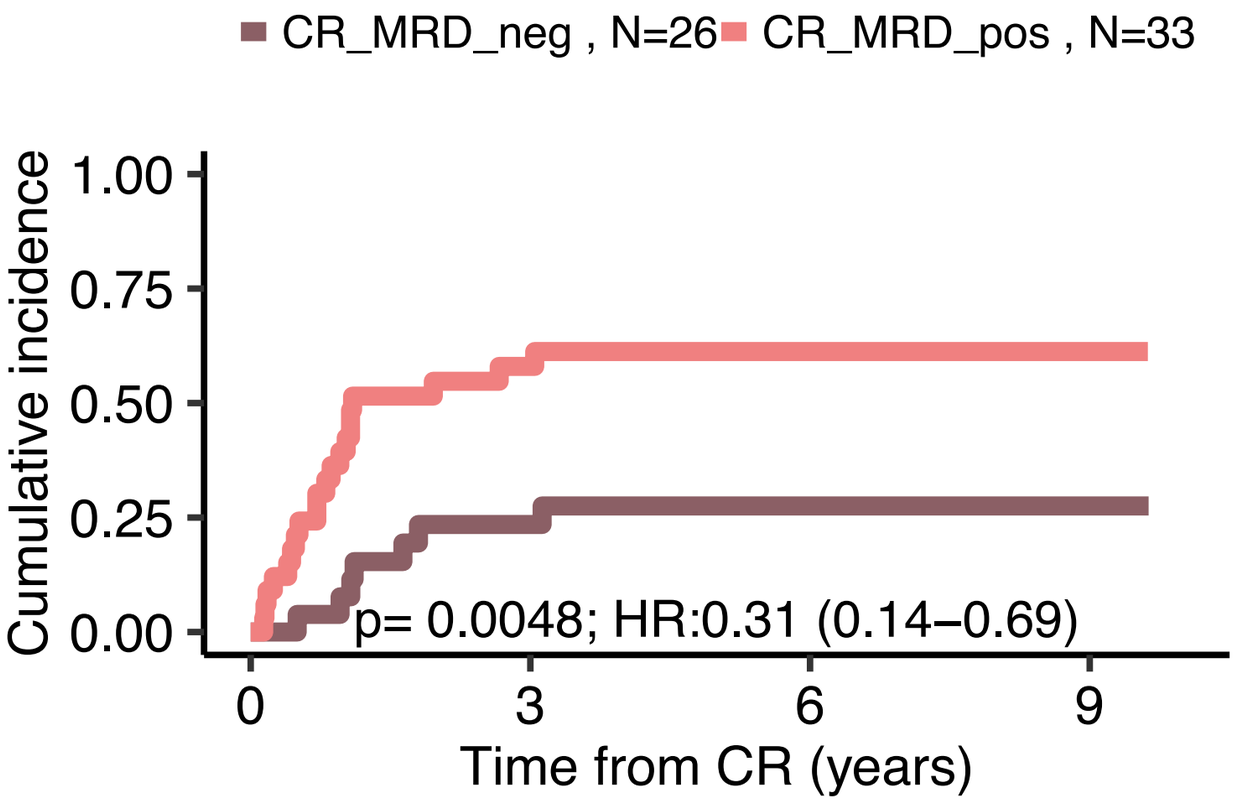
Relapse risk for NRAS_p.G12_13



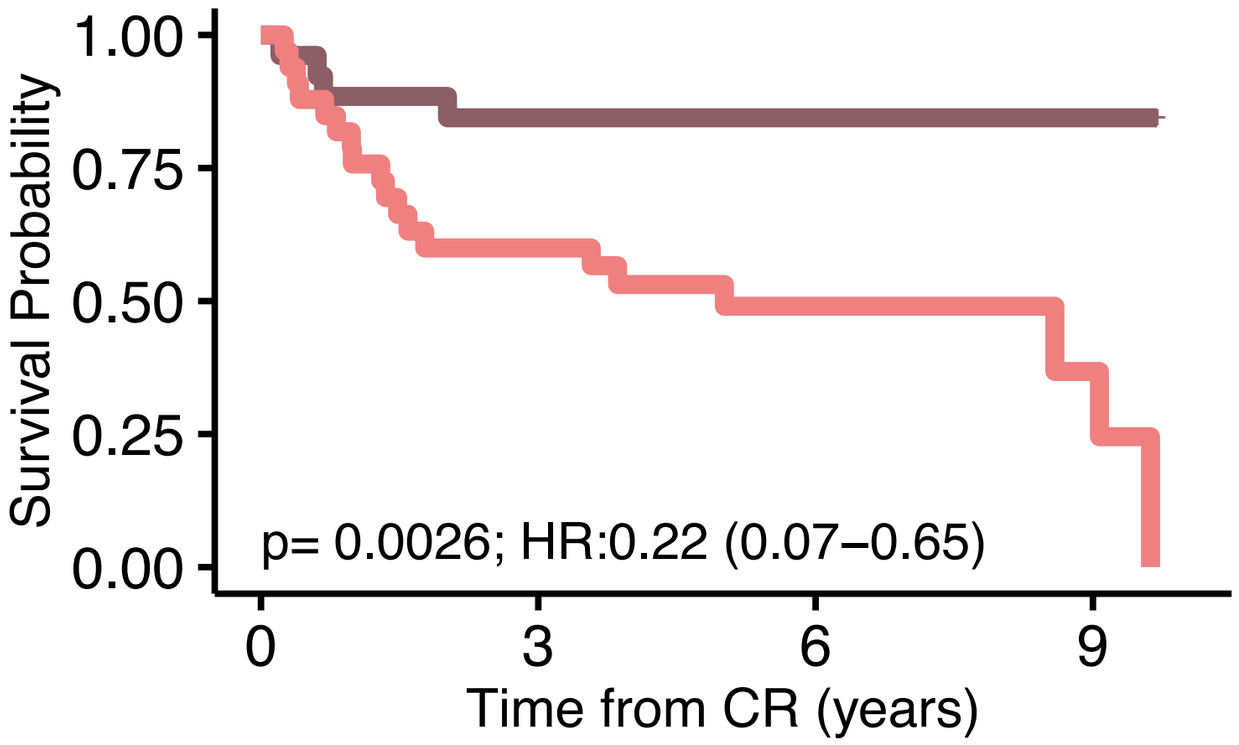
OS CR for NRAS_p.G12_13



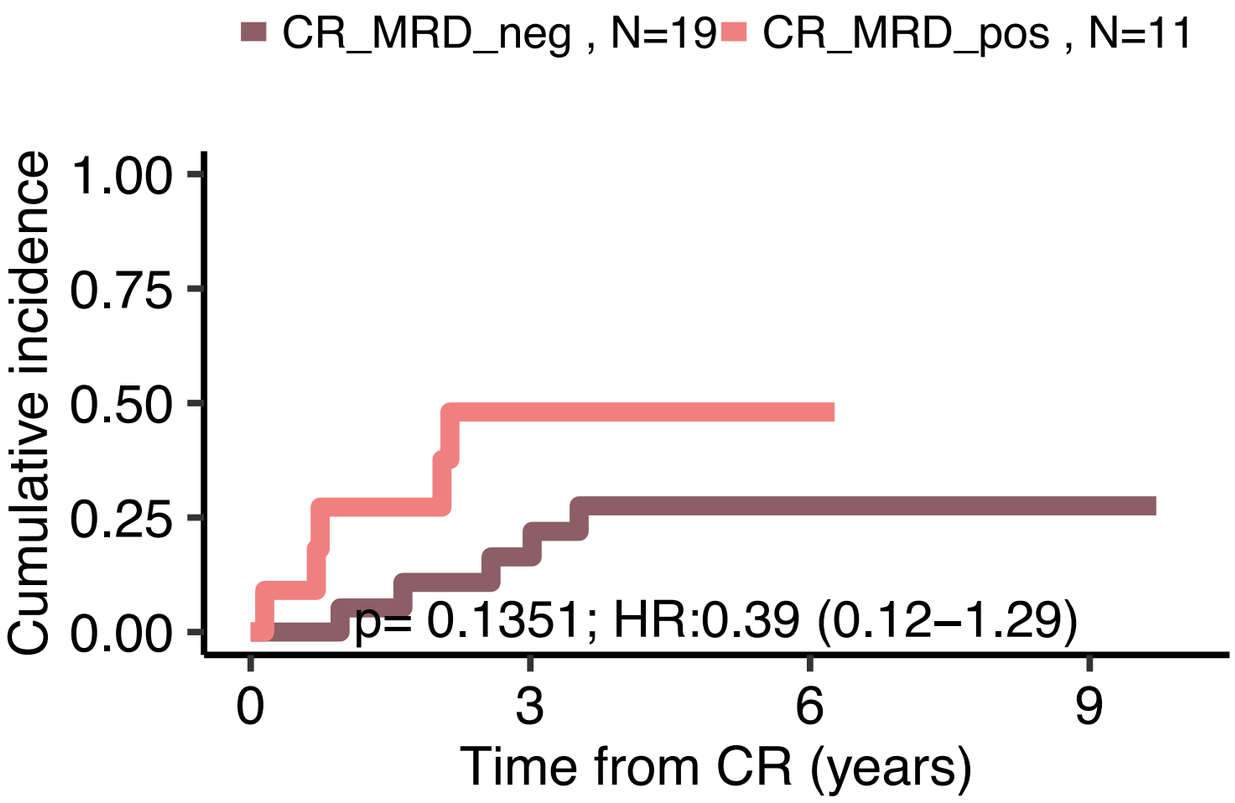
Relapse risk for PTPN11



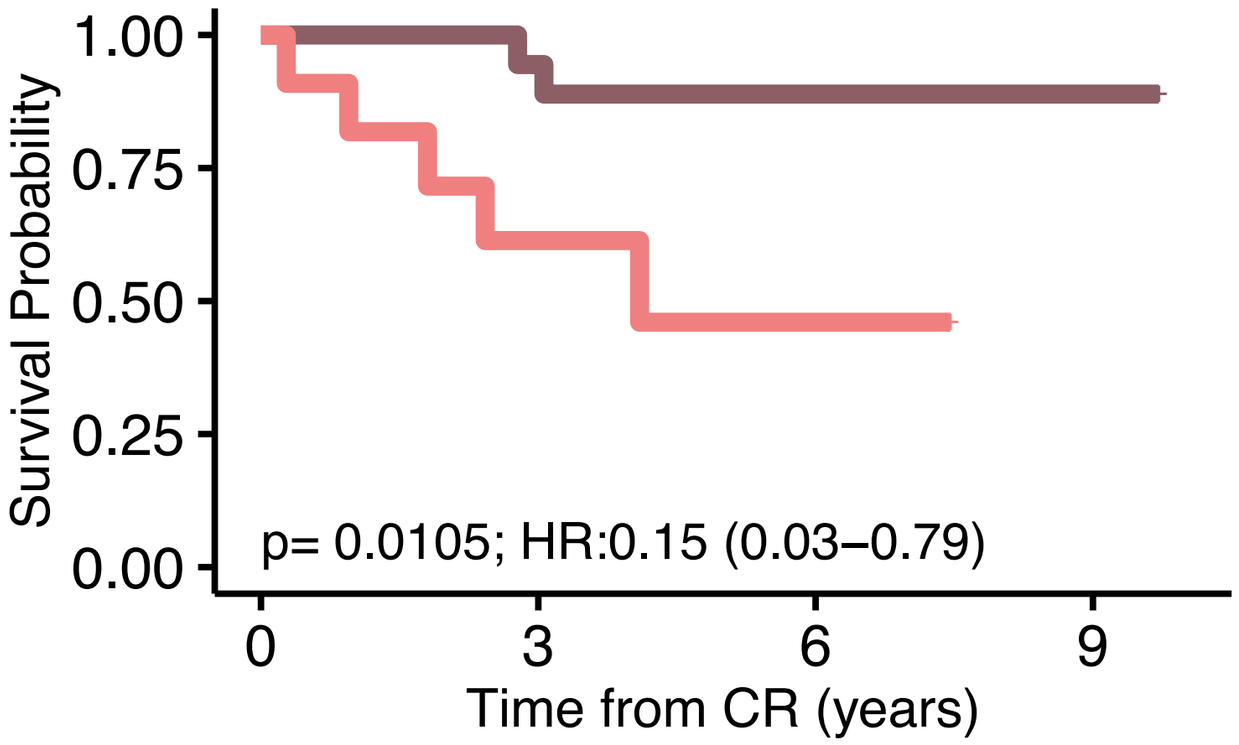
OS CR for PTPN11



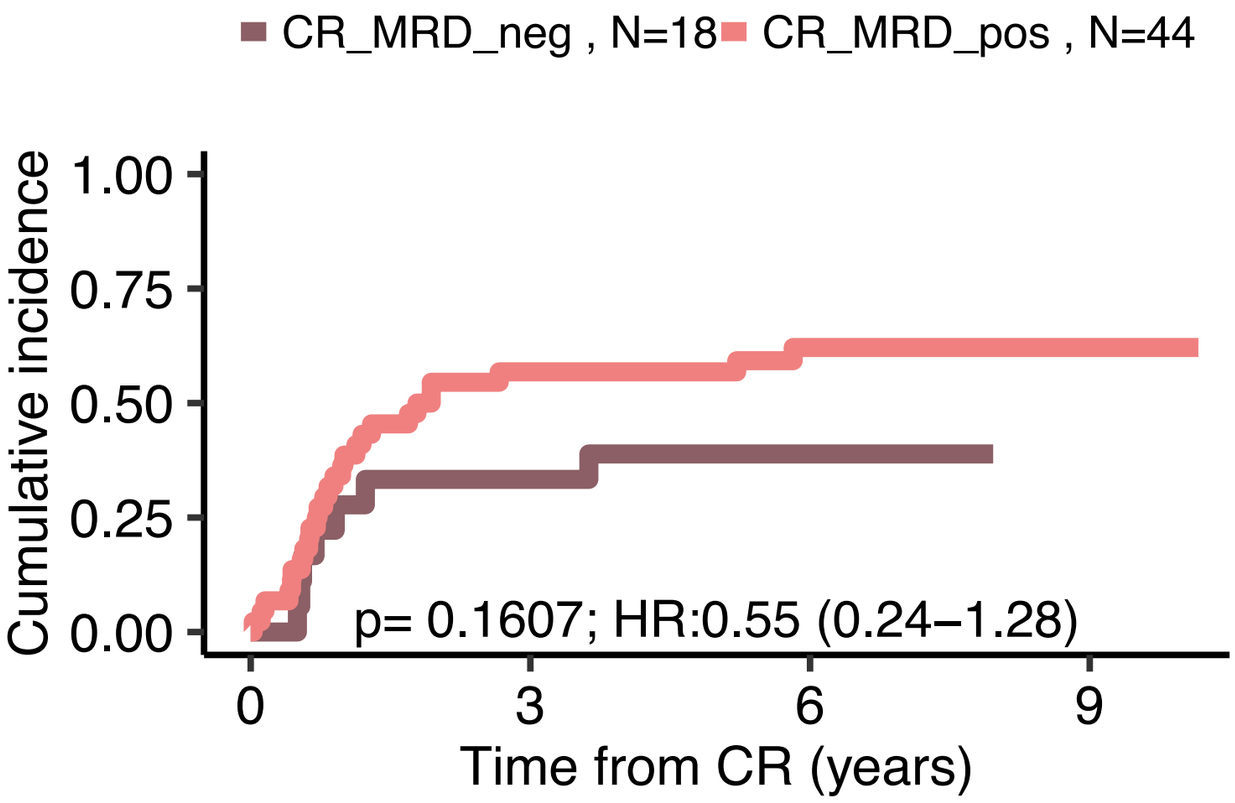
Relapse risk for RAD21



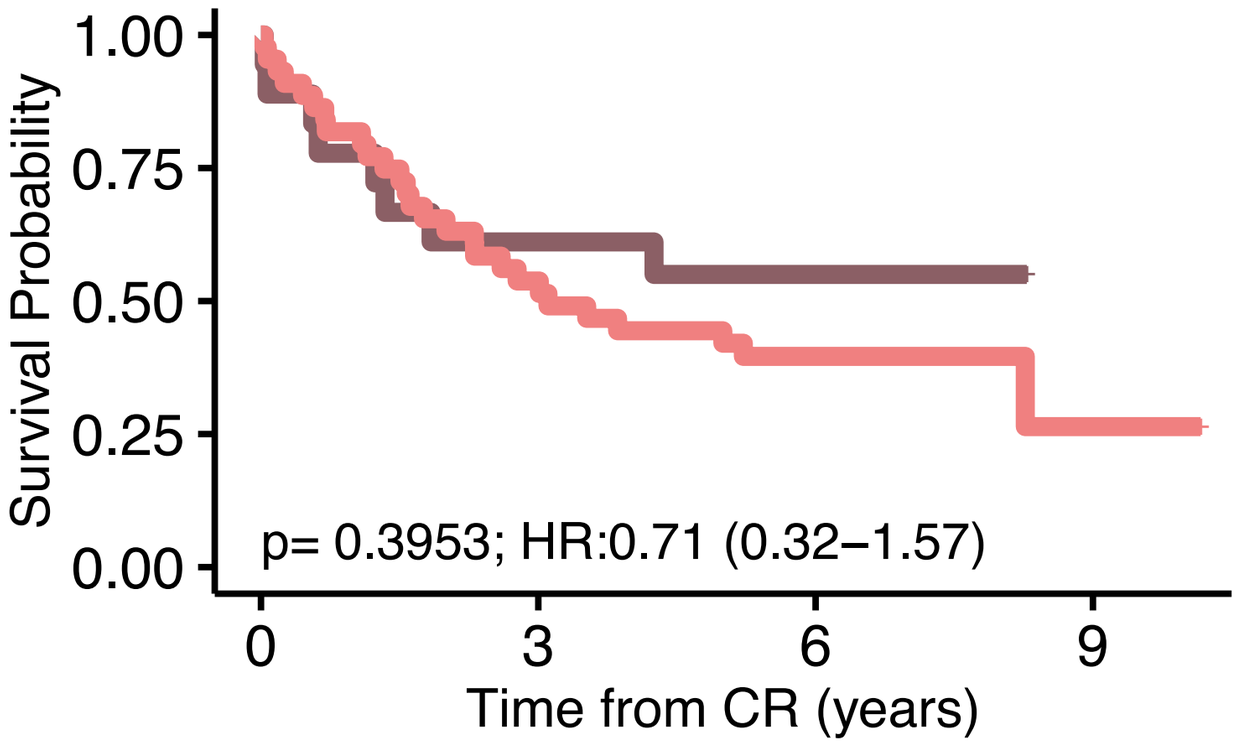
OS CR for RAD21



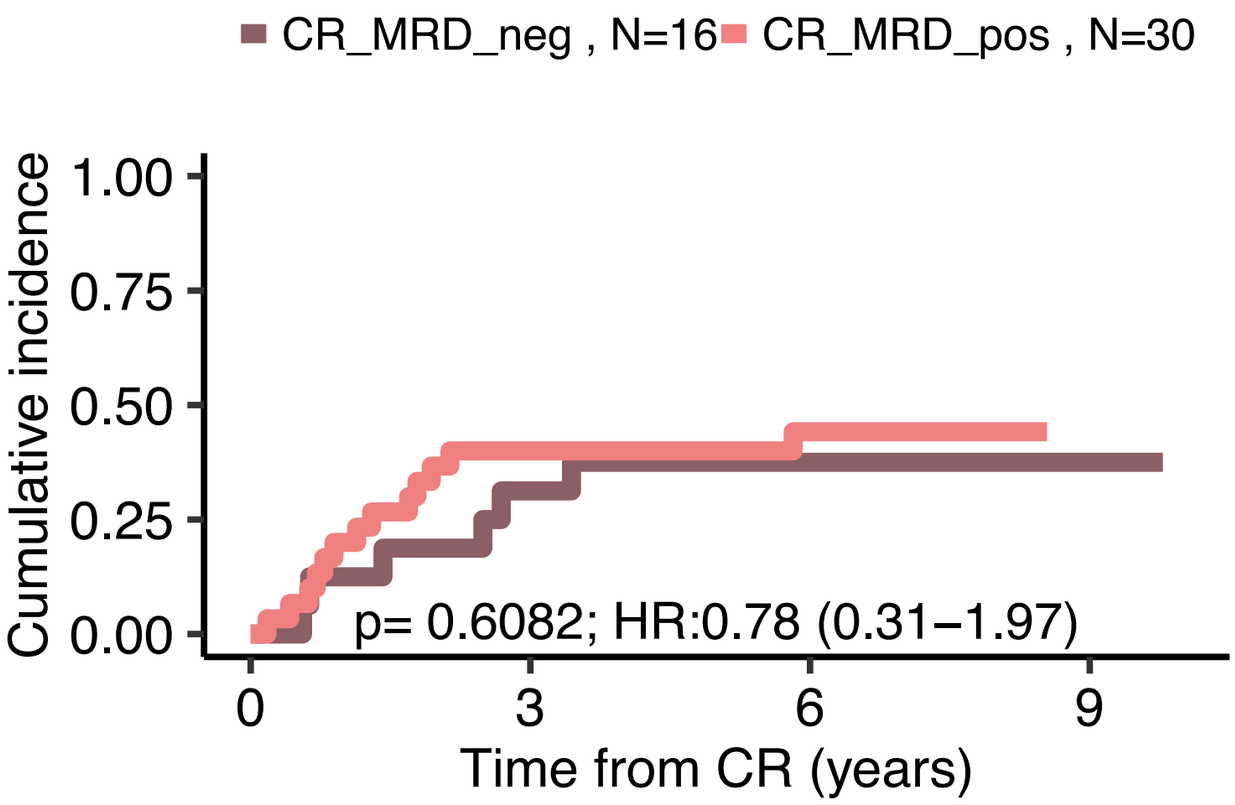
Relapse risk for RUNX1



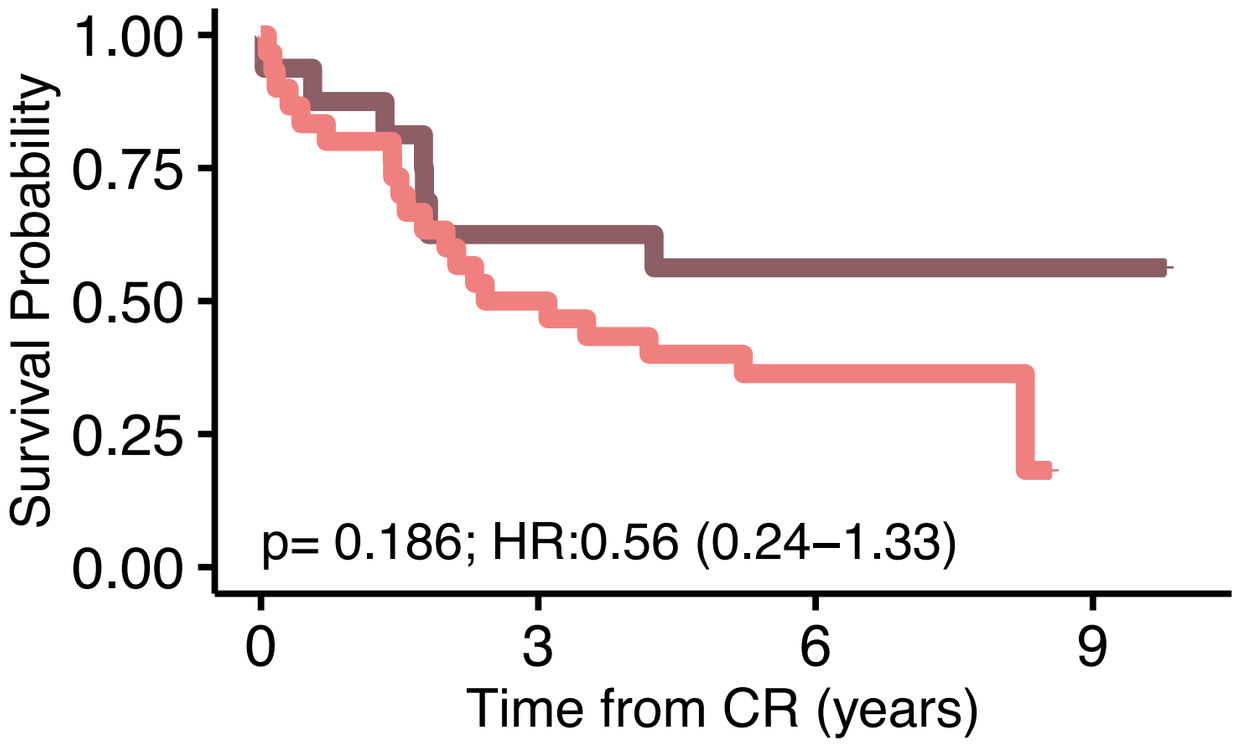
OS CR for RUNX1



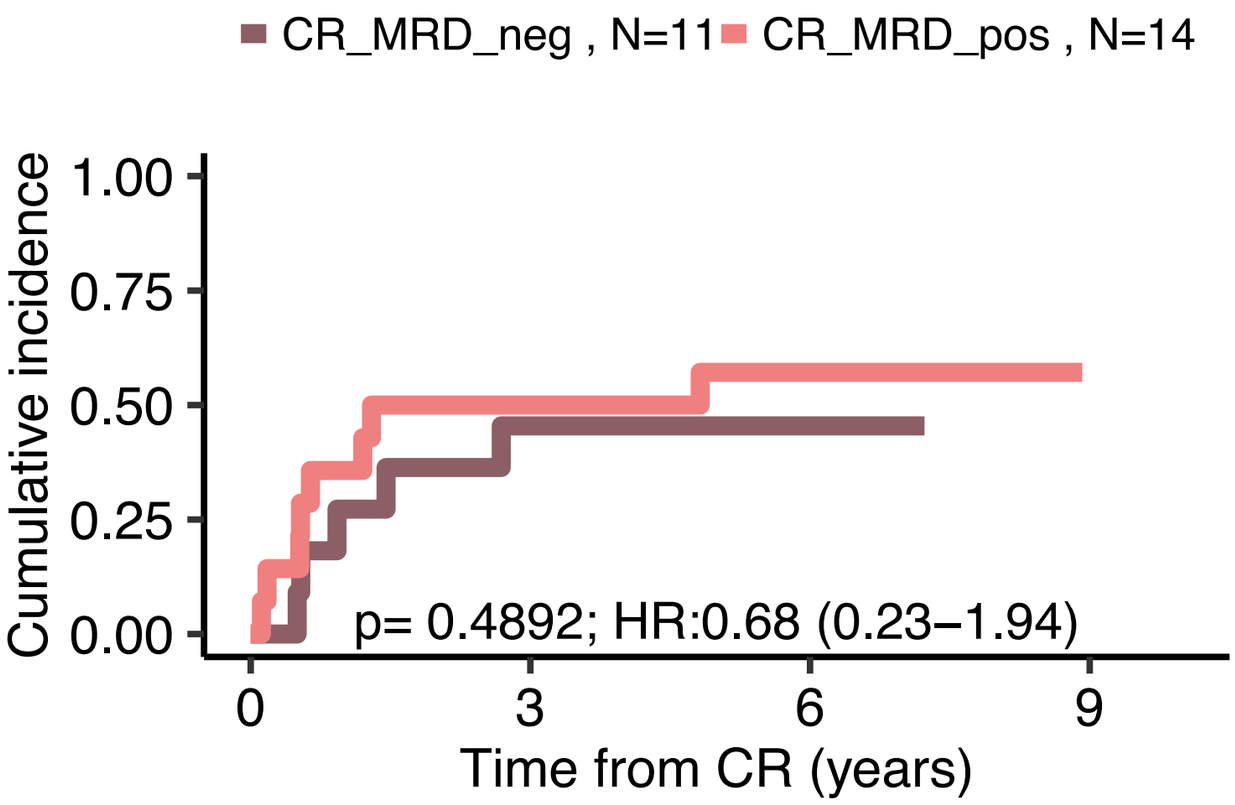
Relapse risk for SRSF2



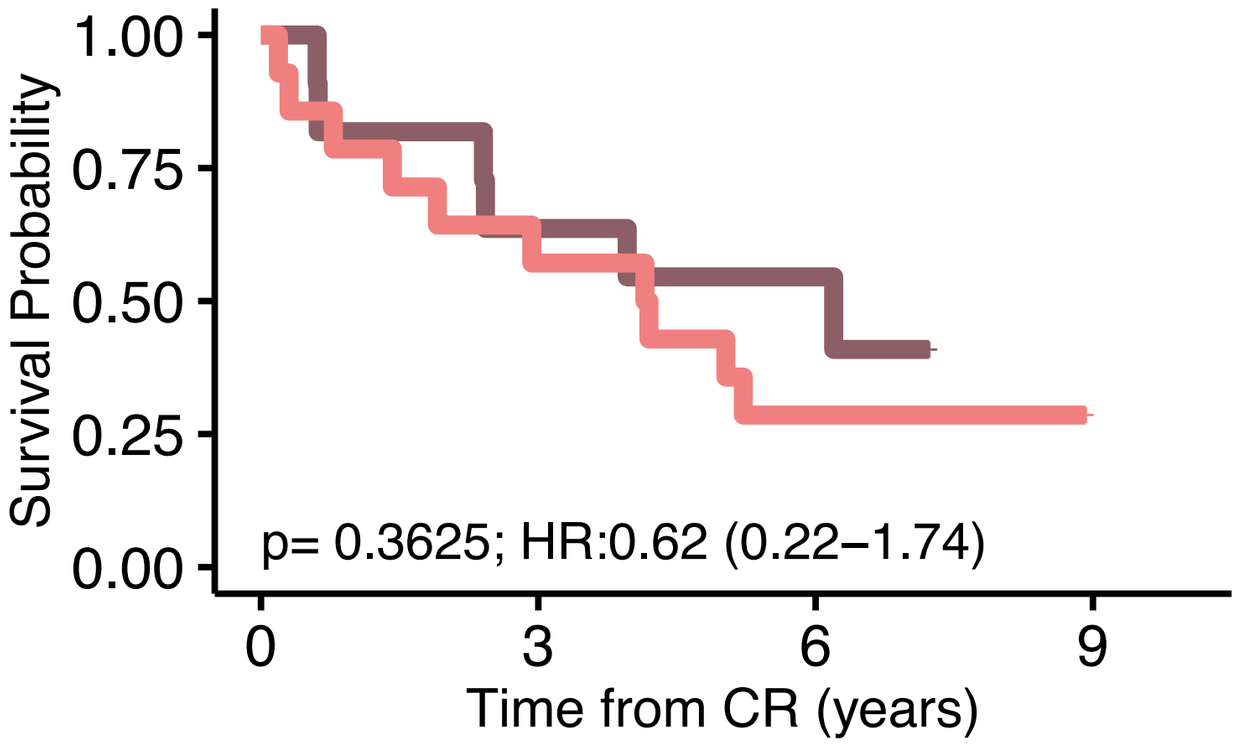
OS CR for SRSF2



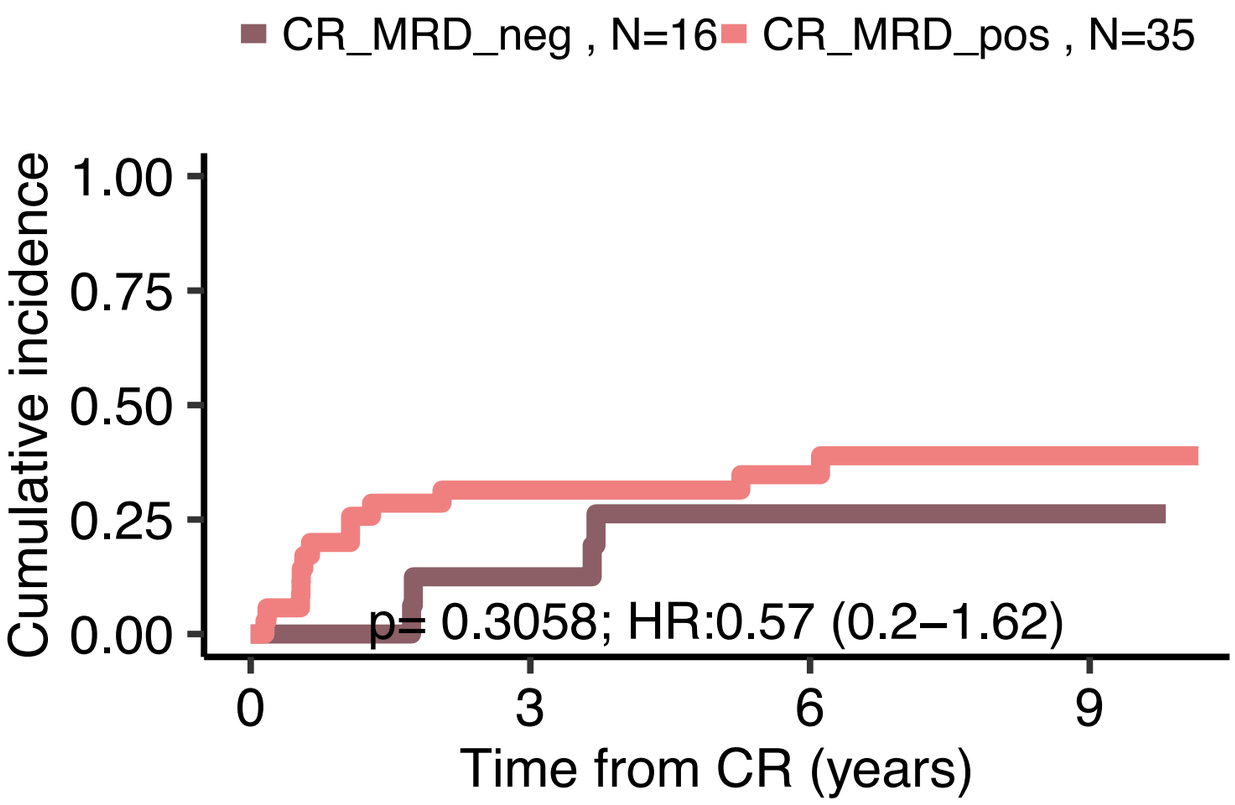
Relapse risk for STAG2



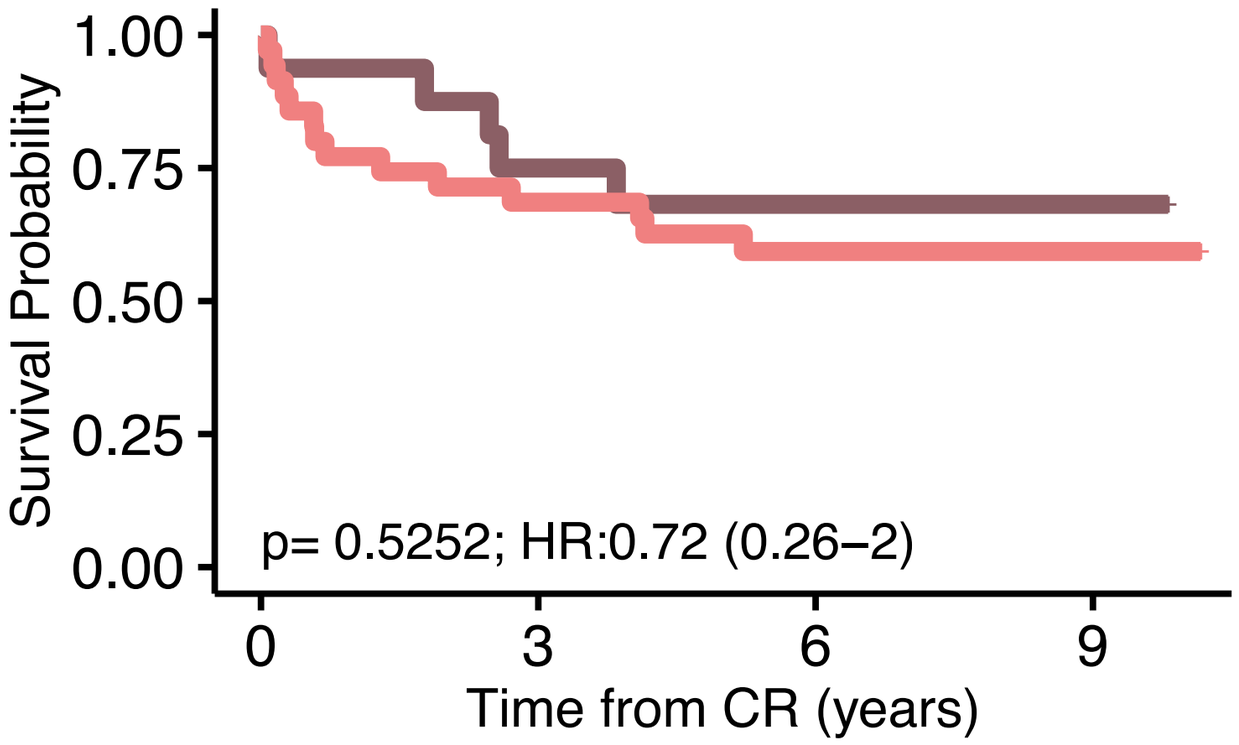
OS CR for STAG2



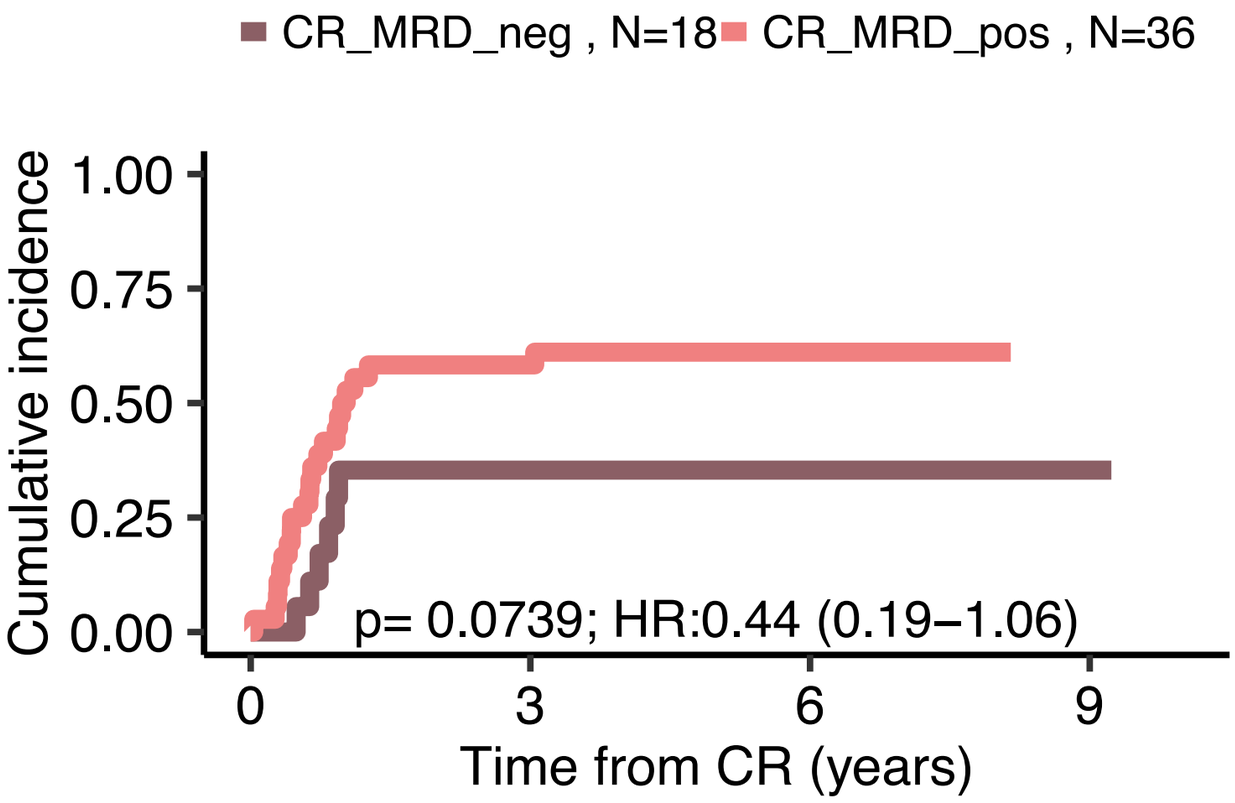
Relapse risk for TET2



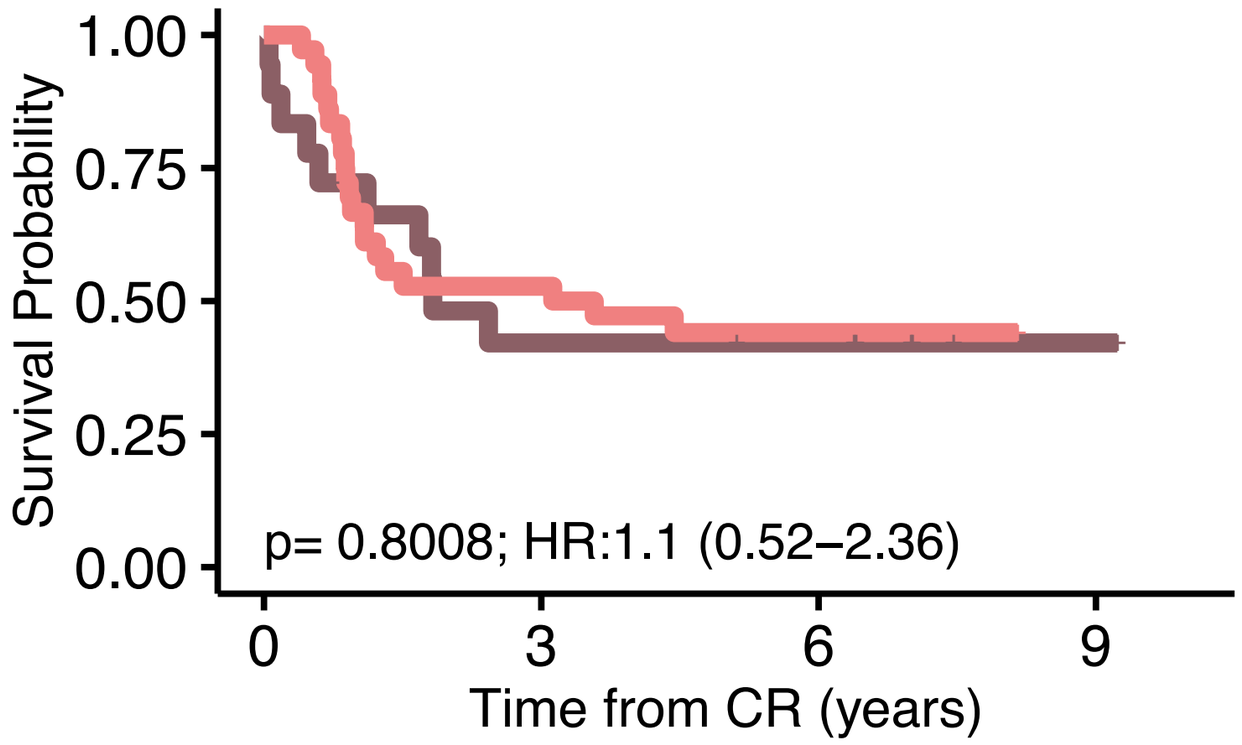
OS CR for TET2



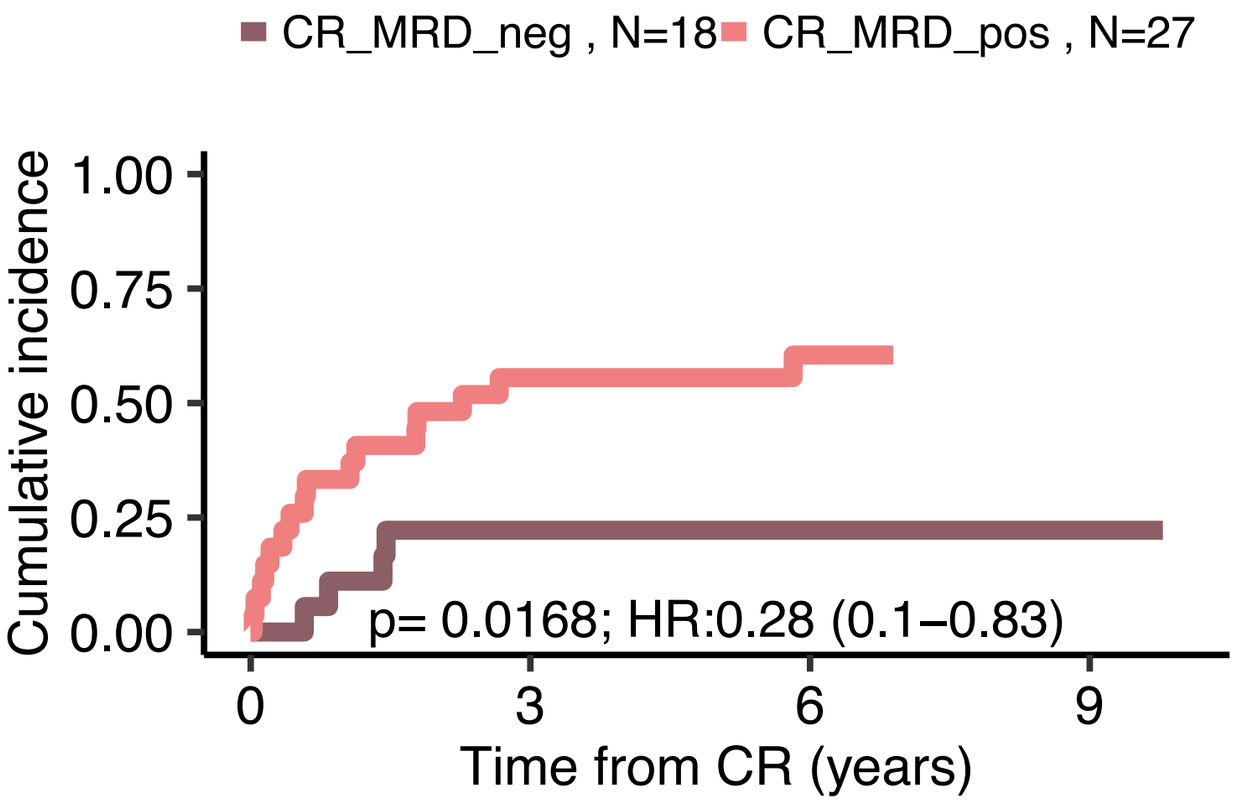
Relapse risk for WT1



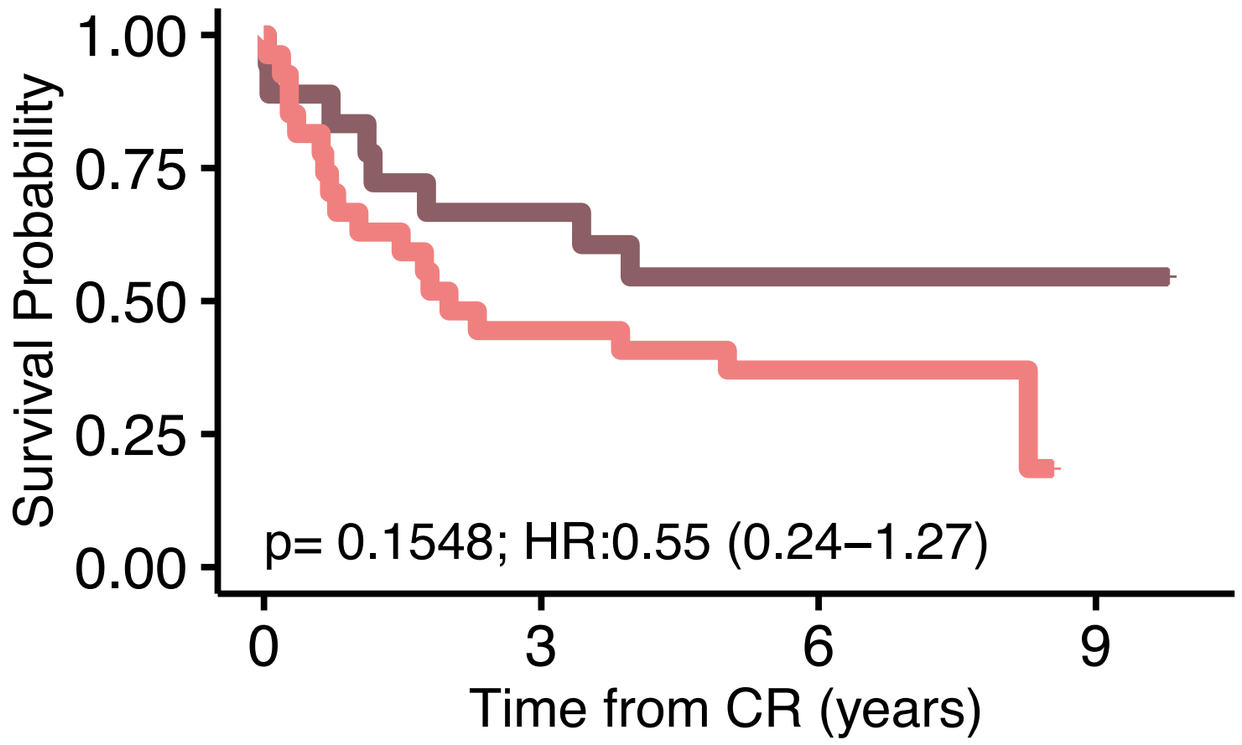
OS CR for WT1



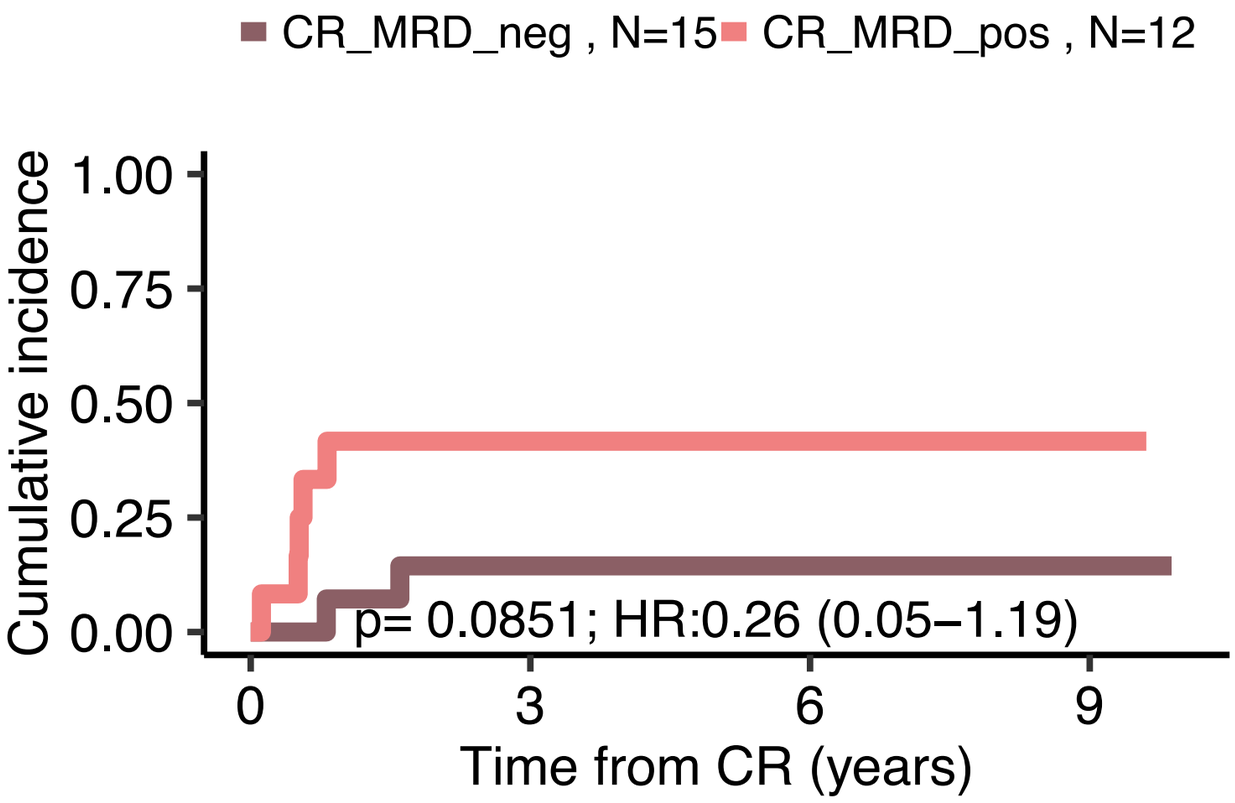
Relapse risk for +8



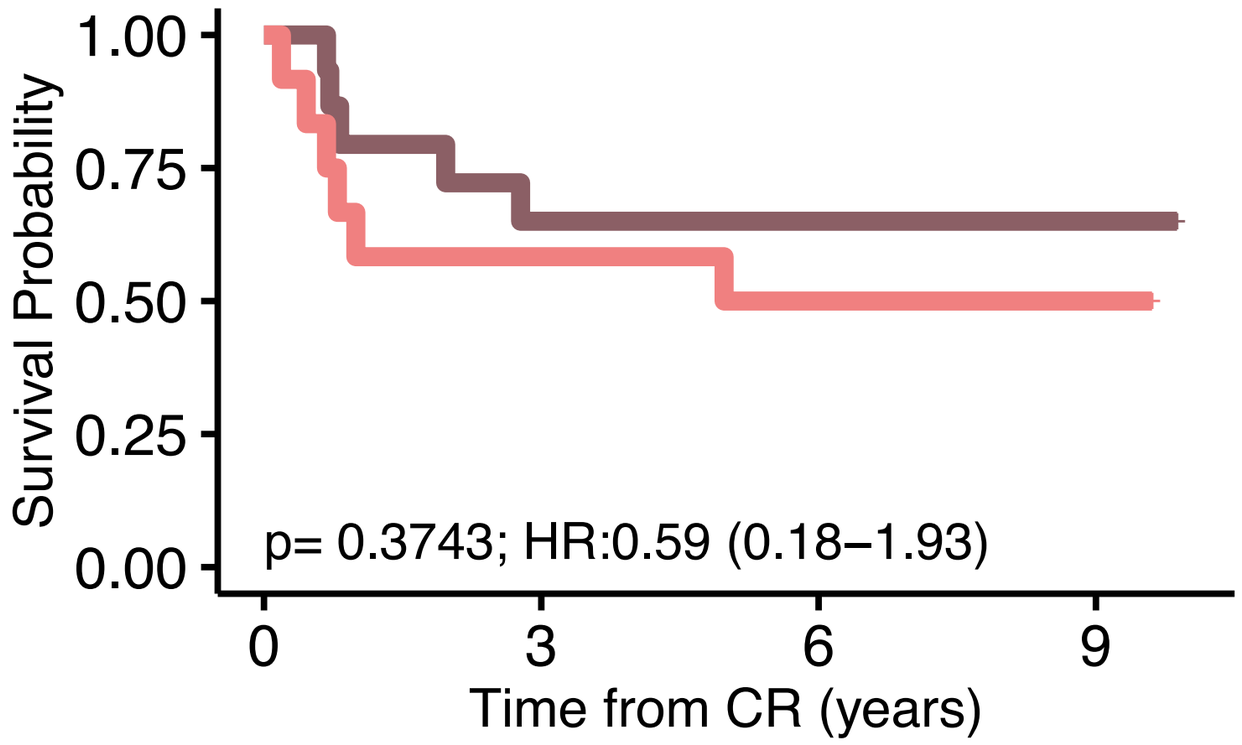
OS CR for +8



Relapse risk for –9



OS CR for –9



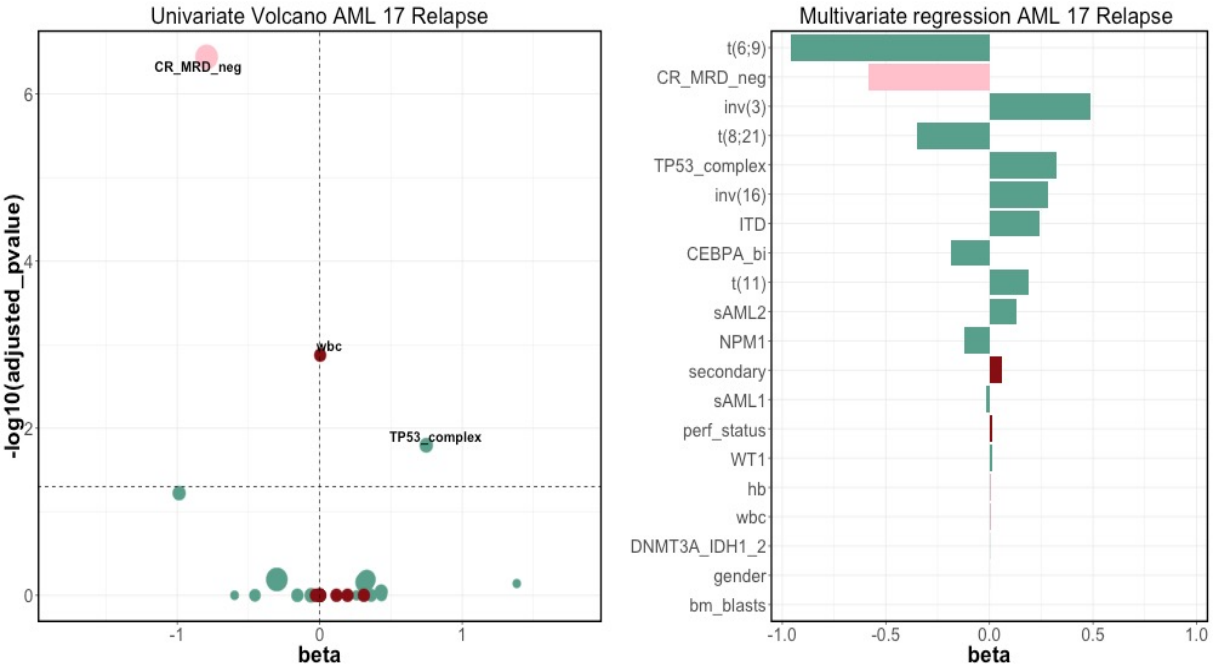
S.Figure 42: Univariate and Multivariate Regression plots for model containing classes, ITD, clinical information as well as minimal residual disease (MRD) status for relapse (panel A) and survival post complete remission (panel B) endpoints in the AML 17 NCRI Trial Cohort with post course 1 MRD analysis (n=523).

A. The first panel represents a univariate Cox model for all the covariates mentioned above for the relapse endpoint. The second panel represents a multivariate Cox model for all the covariates mentioned above for the relapse endpoint (β coefficients were averaged over 100 iterations).

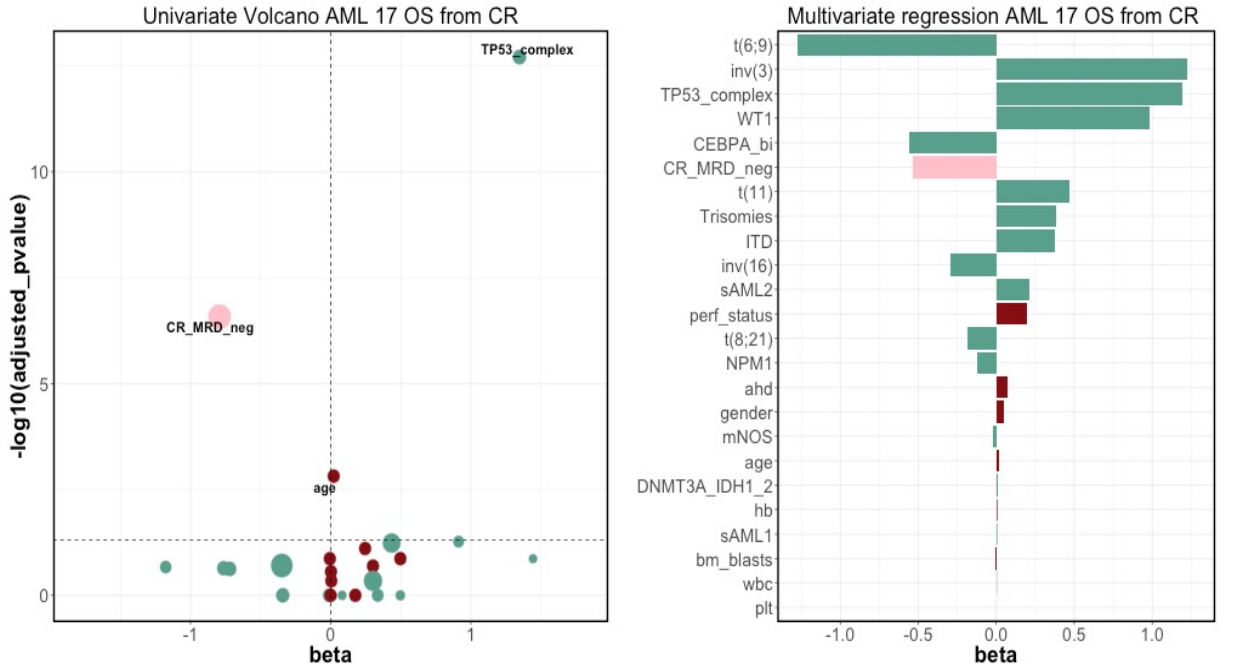
B. The first panel represents a univariate Cox model for all the covariates mentioned above for the death from complete remission endpoint. The second panel represents a multivariate Cox model for all the covariates mentioned above for the survival post complete remission endpoint (β coefficients were averaged over 100 iterations).

The horizontal dotted curve corresponds to the pvalue threshold of 0.05 and the vertical one correspond to β=0 on the x axis. We highlighted the predictors that have a significant effect (pvalue greater than the threshold :0.05 here). The colors correspond to different types for the covariates: green for classes and ITD, pink for MRD status, red for clinical data. Wald test pvalues are adjusted to correct for multiple comparisons.

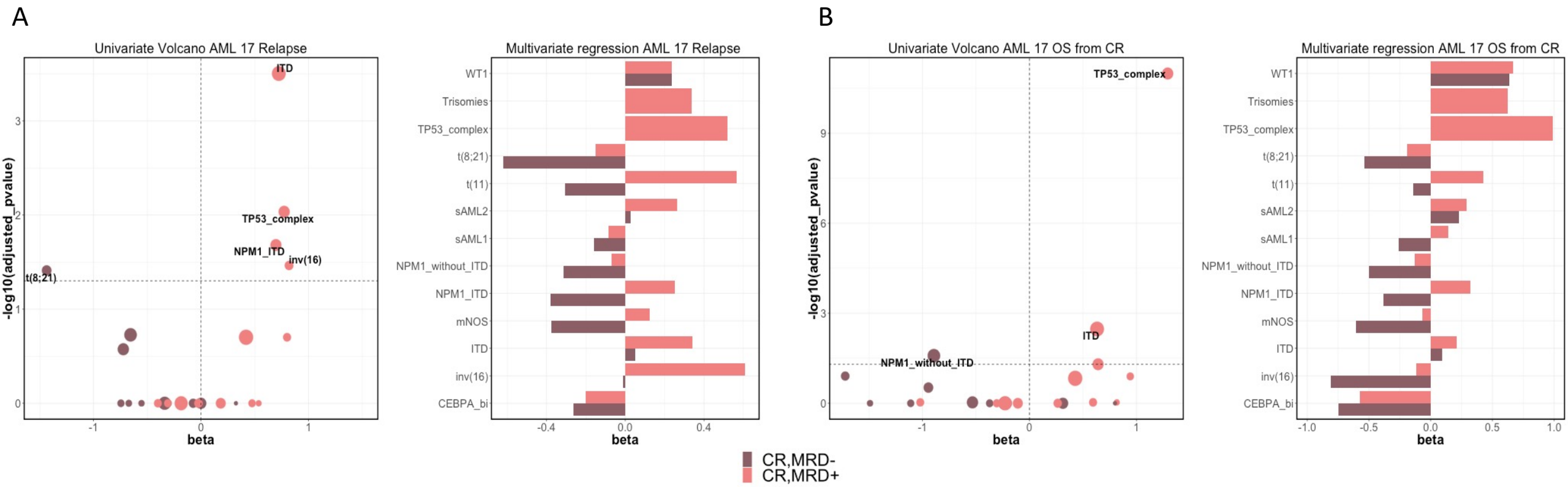
A



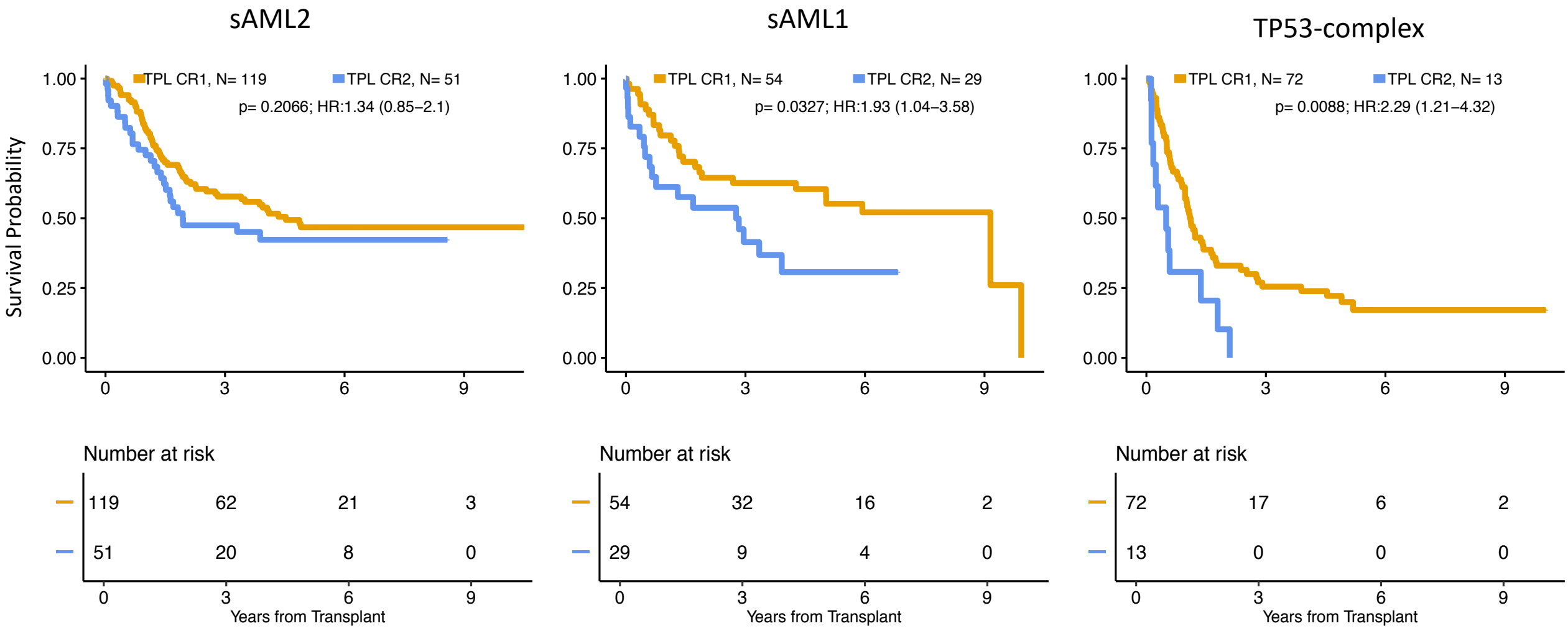
B



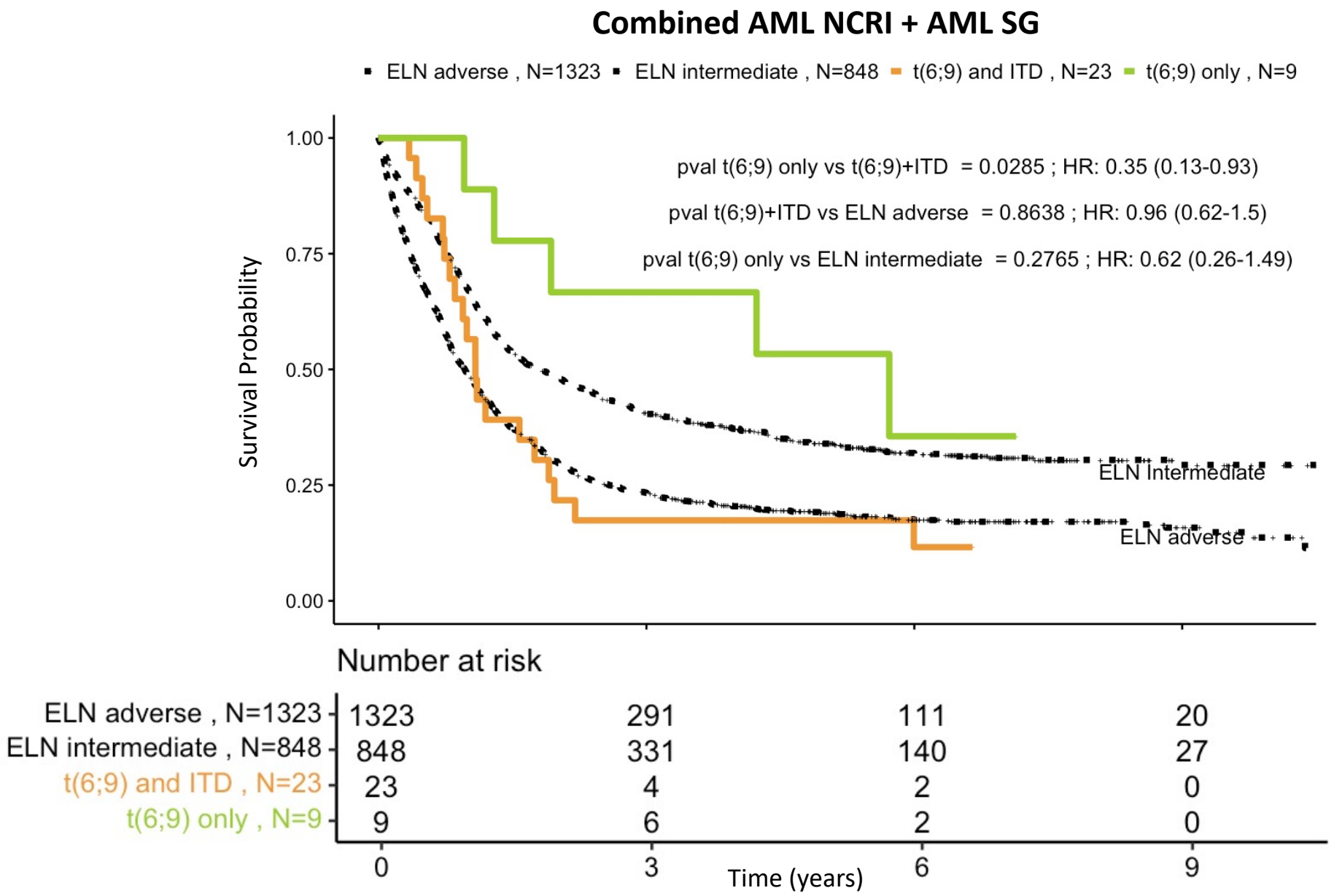
S.Figure 43: Univariate and Multivariate Regression plots for the classes, ITD stratified by minimal residual disease (MRD) status on Relapse (panel A) and survival post complete remission (panel B) endpoints in the AML 17 NCRI Trial Cohort with post course 1 MRD analysis (n=523).
A. The first panel represents a univariate Cox model for all the classes stratified by MRD status for the relapse endpoint. The second panel represents a multivariate Cox model for all the covariates mentioned above stratified by MRD status for the relapse endpoint (β coefficients were averaged over 100 iterations). We further stratified NPM1 class based on ITD status.
B. The first panel represents a univariate Cox model for all the classes stratified by MRD for the death from complete remission endpoint. The second panel represents a multivariate Cox model for all the covariates mentioned above stratified by MRD for the death from complete remission endpoint (β coefficients were averaged over 100 iterations). We further stratified NPM1 class based on ITD status.
The horizontal dotted curve corresponds to the pvalue threshold of 0.05 and the vertical one corresponds to β=0 on the x axis. We highlighted the predictors that have a significant effect (pvalue greater than the threshold :0.05 here). The two colors are defined in the legend corresponding to the stratification of each covariate based on MRD status. Wald test pvalues are adjusted to correct for multiple comparisons.



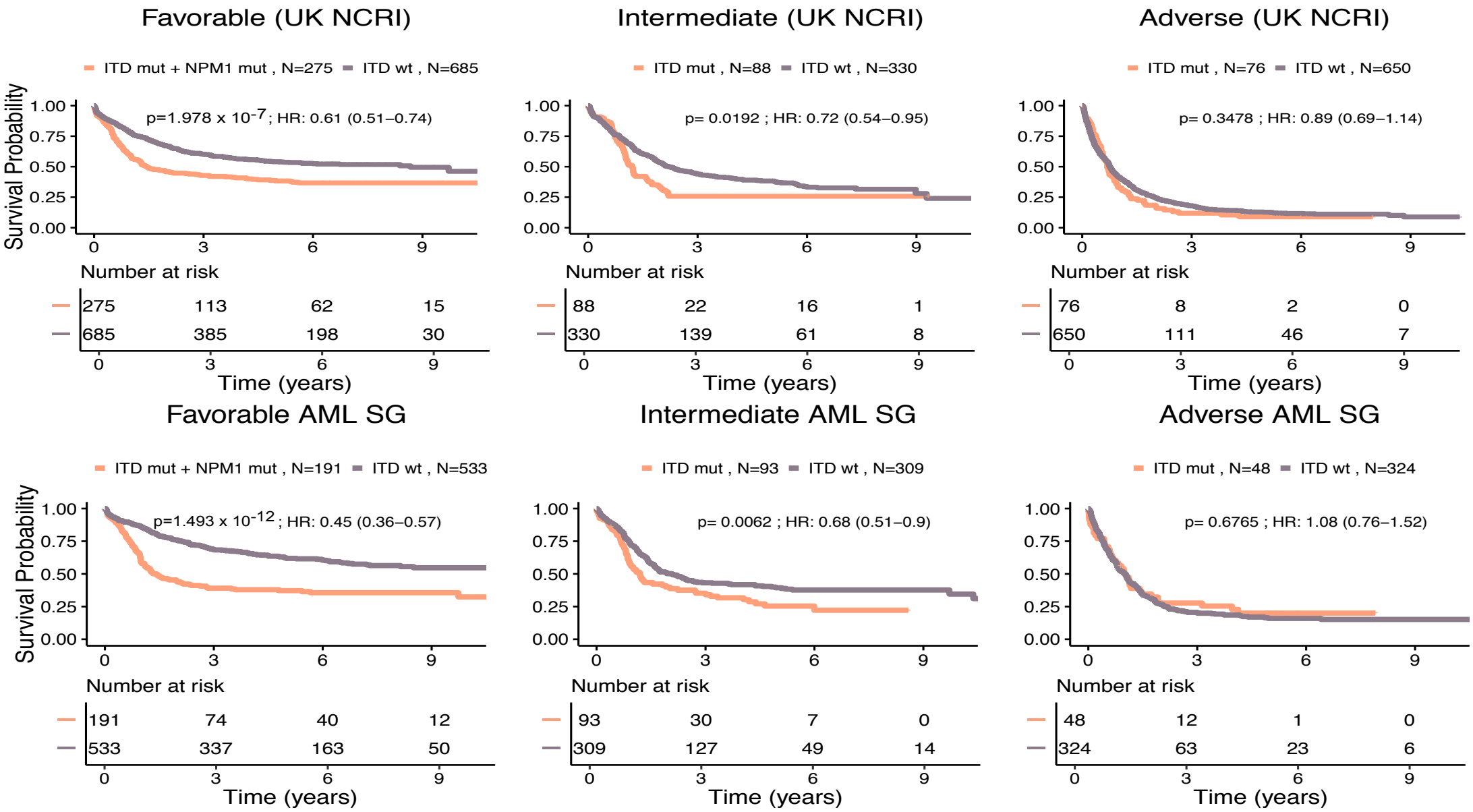
S.Figure 44: Kaplan-Meier overall survival curves and associated risk tables comparing patients who have been transplanted in CR1 to patients transplanted in CR2 for the selected classes. Annotated P values are from two-sided log-rank tests.



S.Figure 45: Kaplan-Meier curves for overall survival and associated risk table comparing patients with t(6;9) with and without ITD with patients in intermediate and adverse ELN^{2017} on the combined (n=3,653) training AML NCRI (n=2,113) and validation AML SG cohort (n=1,540). Dotted curves represent the ELN^{2017} risk categories. Annotated P values are from two-sided log-rank tests.



S.Figure 46: Kaplan-Meier overall survival curves and associated risk tables comparing each of the proposed risk strata (Favorable^{P}, Intermediate^{P}, Adverse^{P}) by the presence of NPM1 and FLT3^{ITD} status for the Favorable^{P} and by FLT3^{ITD} status for the Intermediate^{P} and Adverse^{P} in the training AML NCRI cohort (n=2,113) and the validation AML SG cohort (n=1,540). Annotated P values are from two-sided log-rank tests.



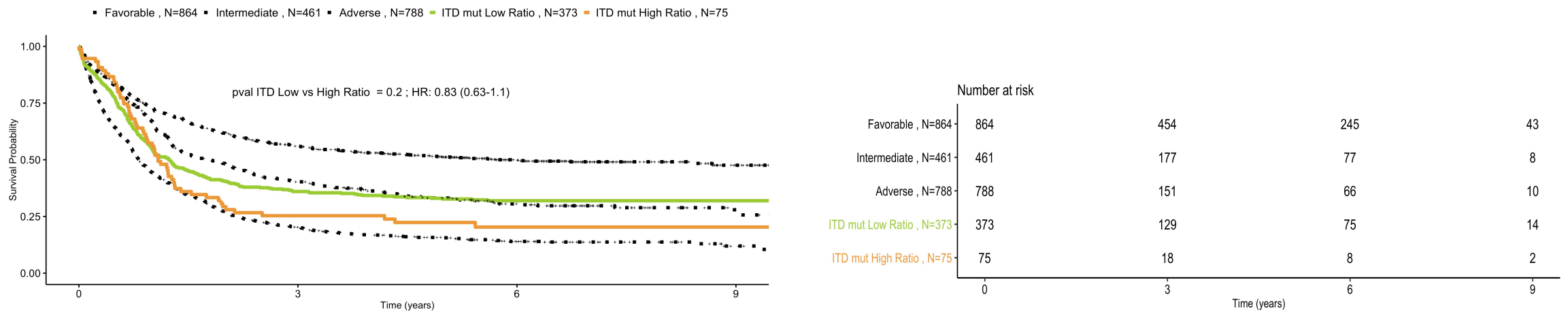
S.Figure 47: Kaplan-Meier curves for overall survival and associated risk tables comparing ITD clinical ratio for all ITD mutated patients and for the subset of patients with both NPM1 and ITD mutations on the AML NCRI cohort (n=2,113).

A. Kaplan-Meier curves for overall survival comparing ITD ratio (low is less than 50 and high is more than 50) for ITD mutated patients on the AML NCRI cohort (n=2,113). Dotted curves represent the ELN^{2017} risk categories.

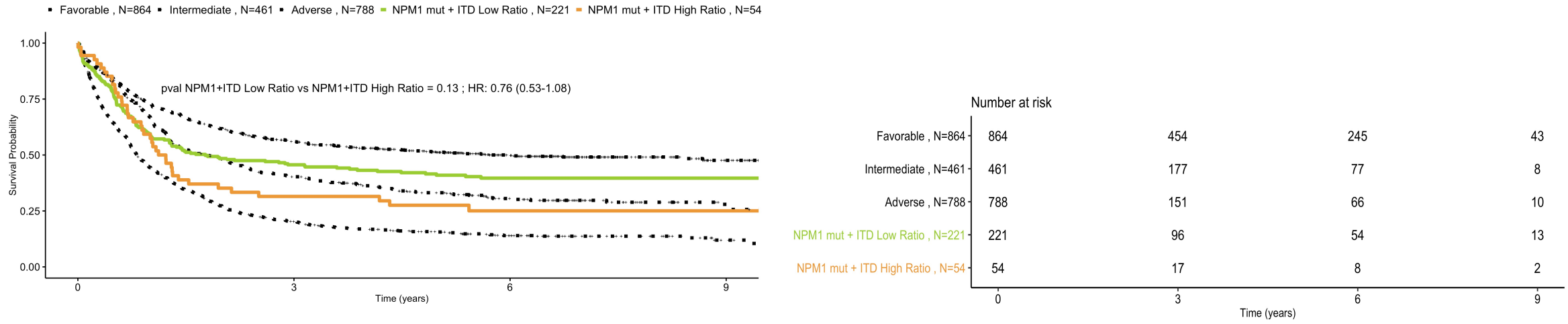
B. Kaplan-Meier curves for overall survival comparing ITD ratio (low is less than 50 and high is more than 50) for patients that have both NPM1 and ITD mutations on the AML NCRI cohort (n=2,113). Dotted curves represent the ELN^{2017} risk categories.

P values were computed using log-rank test to compare ITD ratio survival differences. Annotated P values are from two-sided log-rank tests.

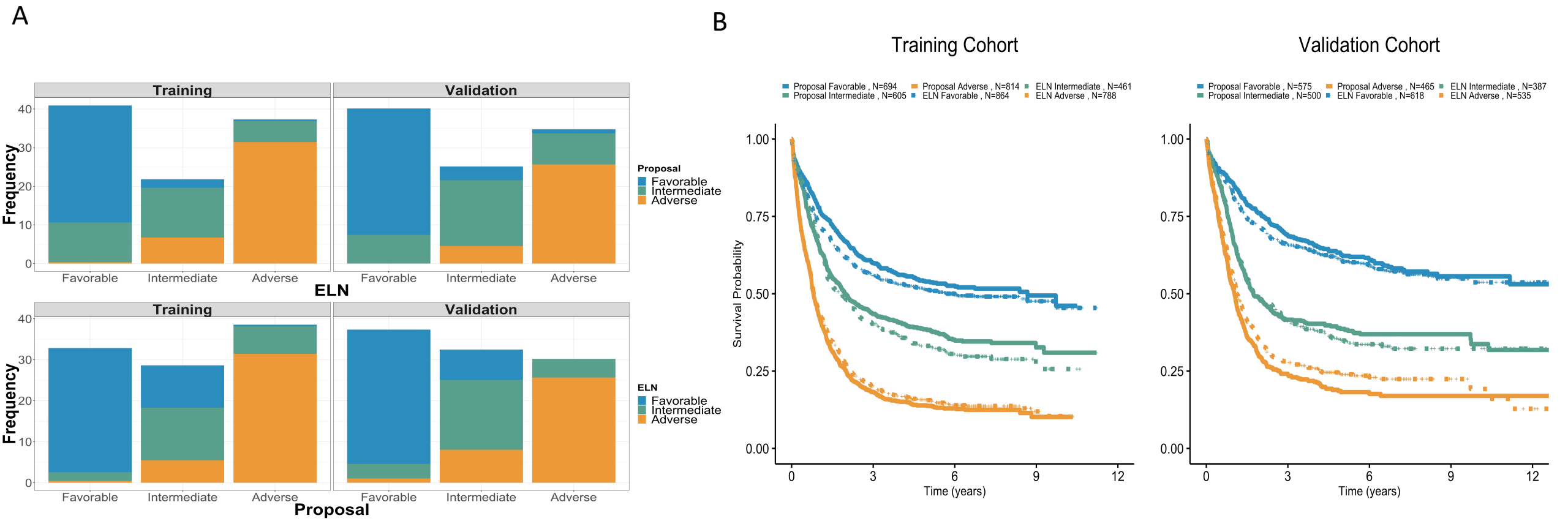
A



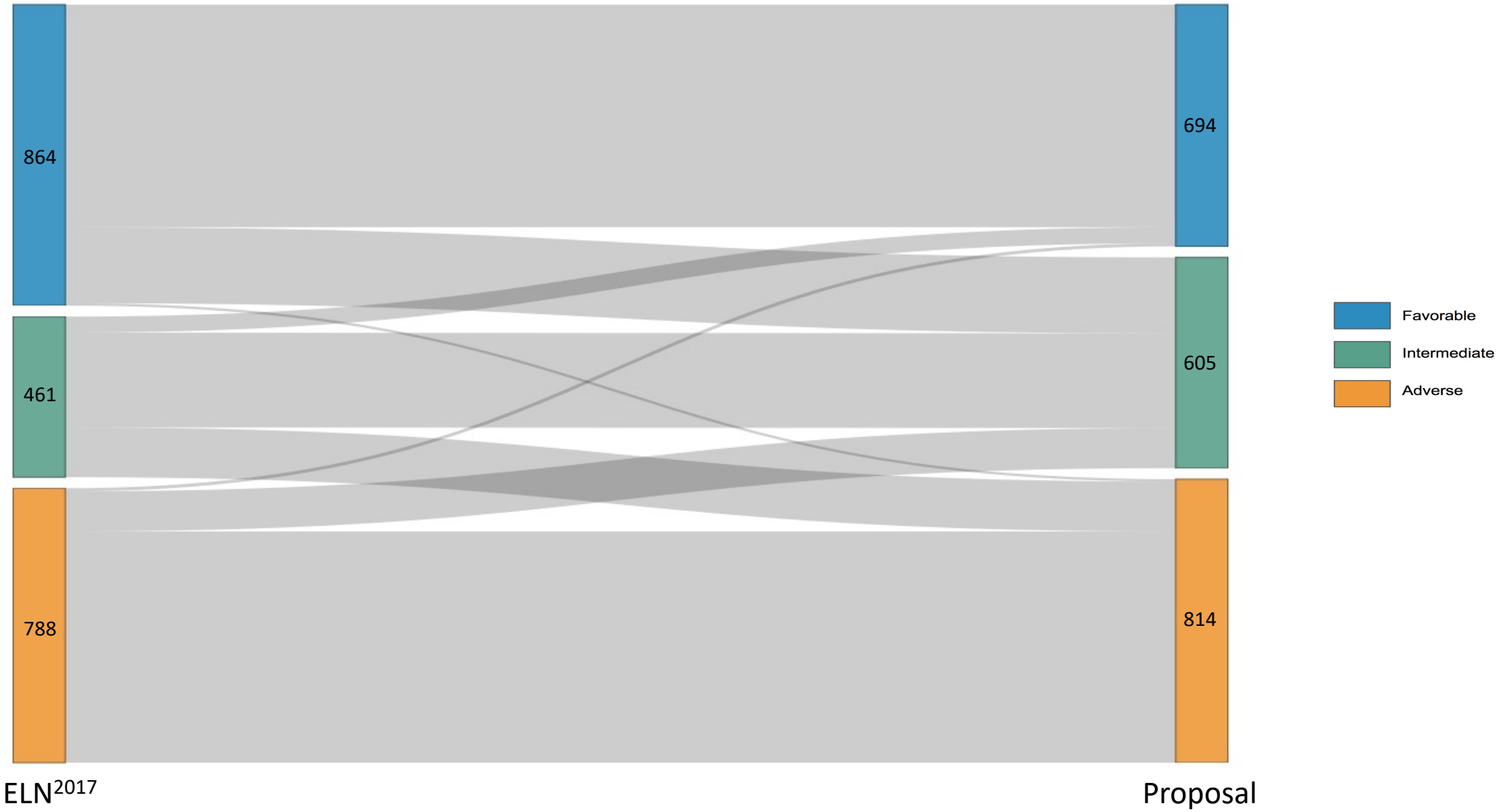
B



S.Figure 48: Comparison of ELN^{2017} and new risk proposal in the training AML NCRI (n=2,113) and validation AML SG Cohort (n=1,540).
A. Top panel : Bar plots of frequencies of ELN^{2017} risk categories stratified by proposal categories in the training (n=2,113) and validation cohort (n=1,540). Bottom panel : Bar plots of frequencies of proposal risk categories stratified by ELN^{2017} categories in the training (n=2,113) and validation cohort (n=1,540).
B. Kaplan-Meier curves for overall survival comparing the ELN^{2017} risk categories and the new risk proposal categories in the training (n=2,113) and validation cohort (n=1,540). Dashed curves are the three ELN^{2017} risk categories and plain curves are the three risk proposal categories.

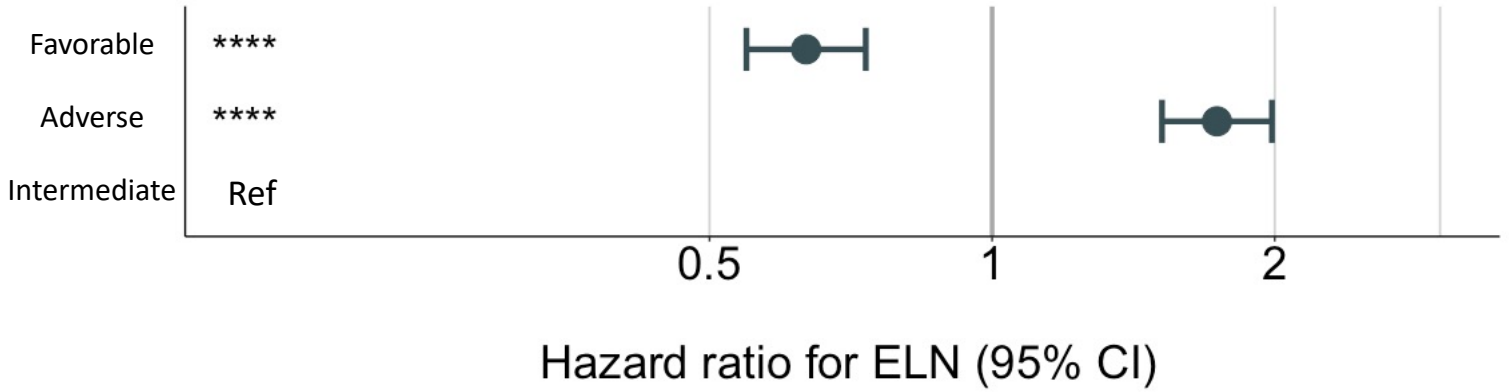


S. Figure 49: Sankey plot comparing the proportion of the ELN^{2017} risk groups shifting in the proposed risk groups in the AML NCRI cohort (n=2,113). Sankey Plots comparing ELN and risk proposal frequencies.

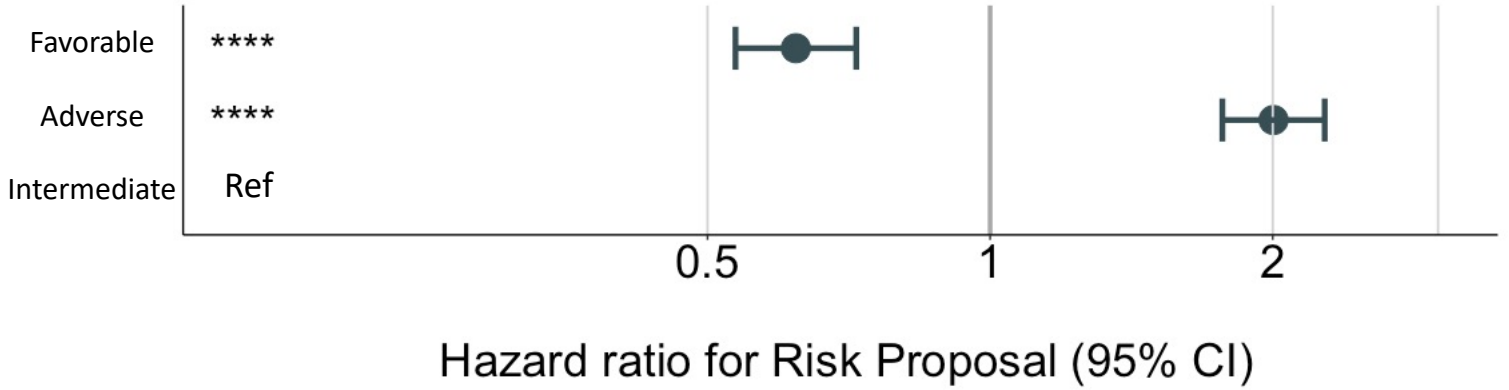


S.Figure 50: Forest plot multivariate Cox regression of A. ELN[2017] risk categories and B. risk proposal in NCRI trial study set (n= 2,113).Dots and lines represent estimated hazard ratios and 95% CI, respectively. ****P < 0.0001, ***P < 0.001, **P < 0.01, ns, not significant. P > 0.05, Wald test.

A.

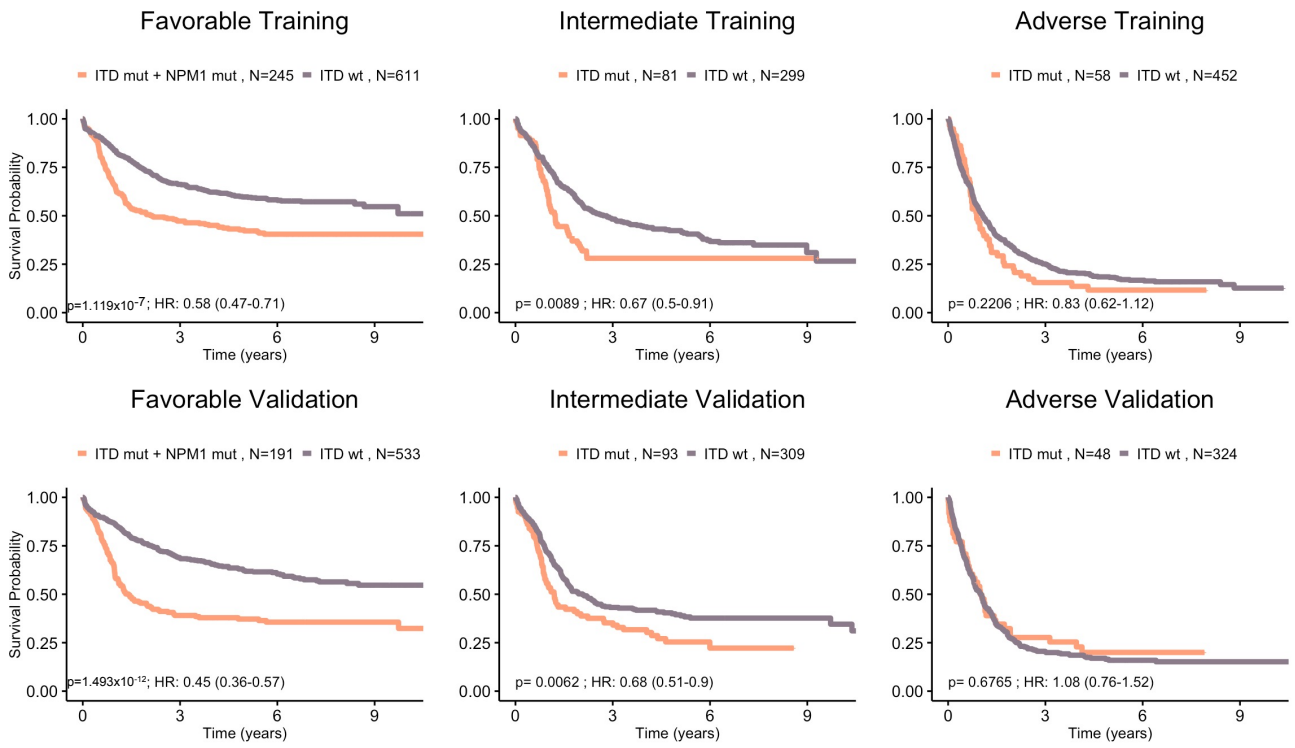


B.

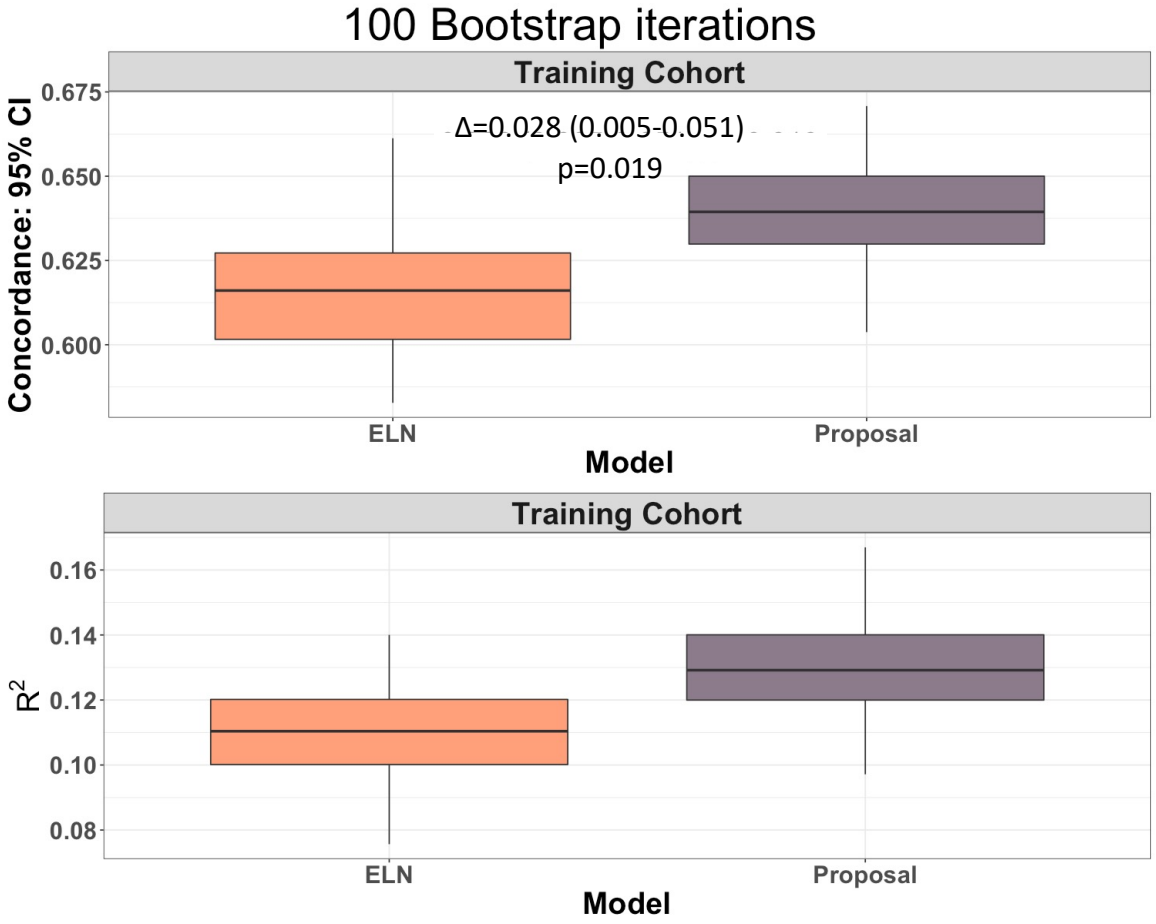


S. Figure 51: New risk proposal based on the AML classes on the subset of intensively treated patients in the AML NCRI cohort (n=1,755). A. Kaplan-Meier overall survival curves comparing each of the proposed risk strata (Favorable^{P}, Intermediate^{P}, Adverse^{P}) by FLT3^{ITD} status for the training AML NCRI cohort for intensively treated patients (n=1,755) validate the rationale for the FLT3^{ITD} shift in risk. P values were computed using two-sided log-rank test. B. The estimated improvement in the concordance index (C-index) and pseudo-variance explained (R^2) for the two classifiers in the training AML NCRI Cohort (n=1,755). 95% confidence intervals were generated by bootstrap resampling for the C-index. In all boxplots, the median is indicated by the horizontal line and the first and third quartiles by the box edges. The lower and upper whiskers extend from the hinges to the smallest and largest values, respectively, no further than 1.5× interquartile range from the hinges. Annotated P values are from two-sided t score test.

A



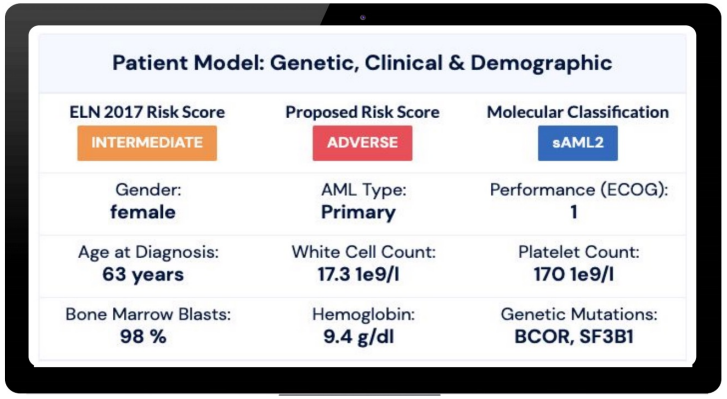
B



S. Figure 52: Example presentation of personalized clinical decision support tool for relative contribution of the covariates in all possible transitions.

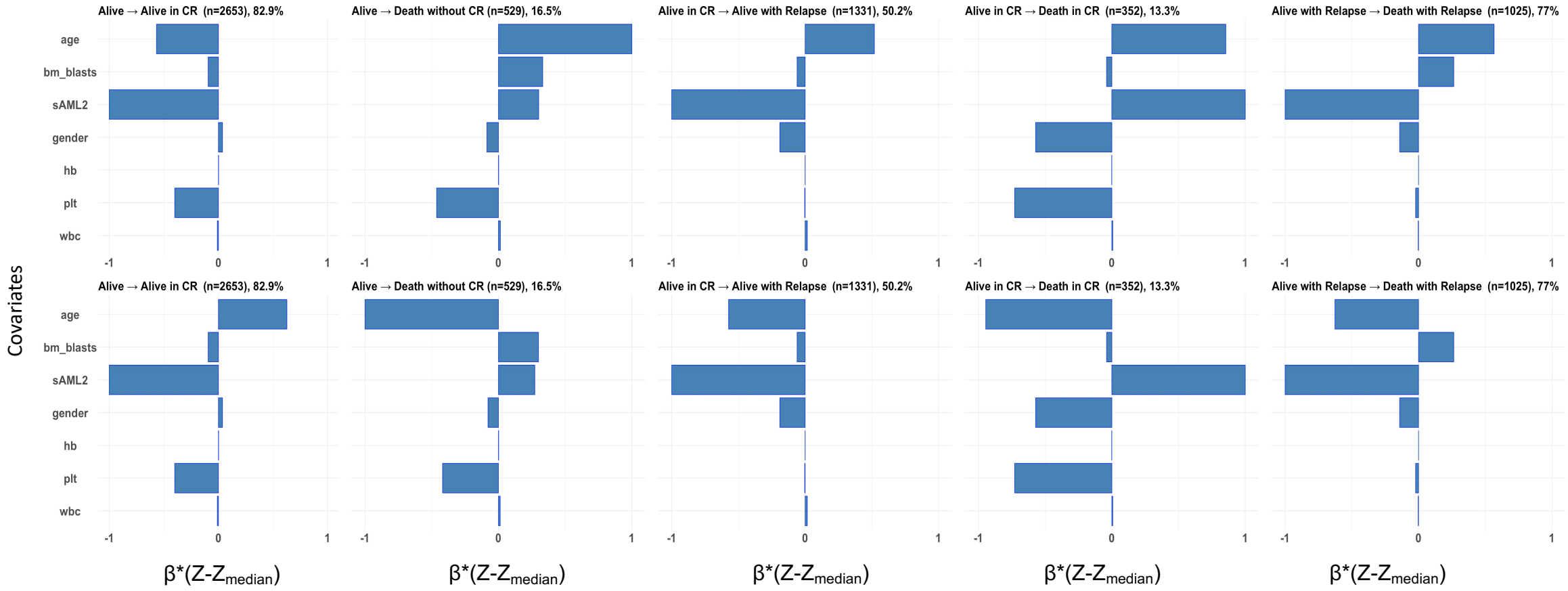
The calculator is derived using the multi-state models that consider data from (n=3,201 total patients, UK NCRI and AMLSG) all intensively treated. Input parameters to include cytogenetic, genetic, clinical and demographic are considered to display each patient's ELN^{2017} score alongside the proposed by this study molecular class and proposed risk group. Personalized estimates on the basis of age and gender are also enabled.

To contrast the contribution of the covariates, we also computed the relative contribution for a younger (40 year old) patient with the same covariates. The relative contributions correspond to the relative $\beta^*(Z-Z_{median})$ values where β is the coefficient derived from the multi-state Cox model for that particular transition for that specific covariate named Z and Z_{median} is that same covariate for a median patient in the combined cohort. For more details, refer to S.Appendix.

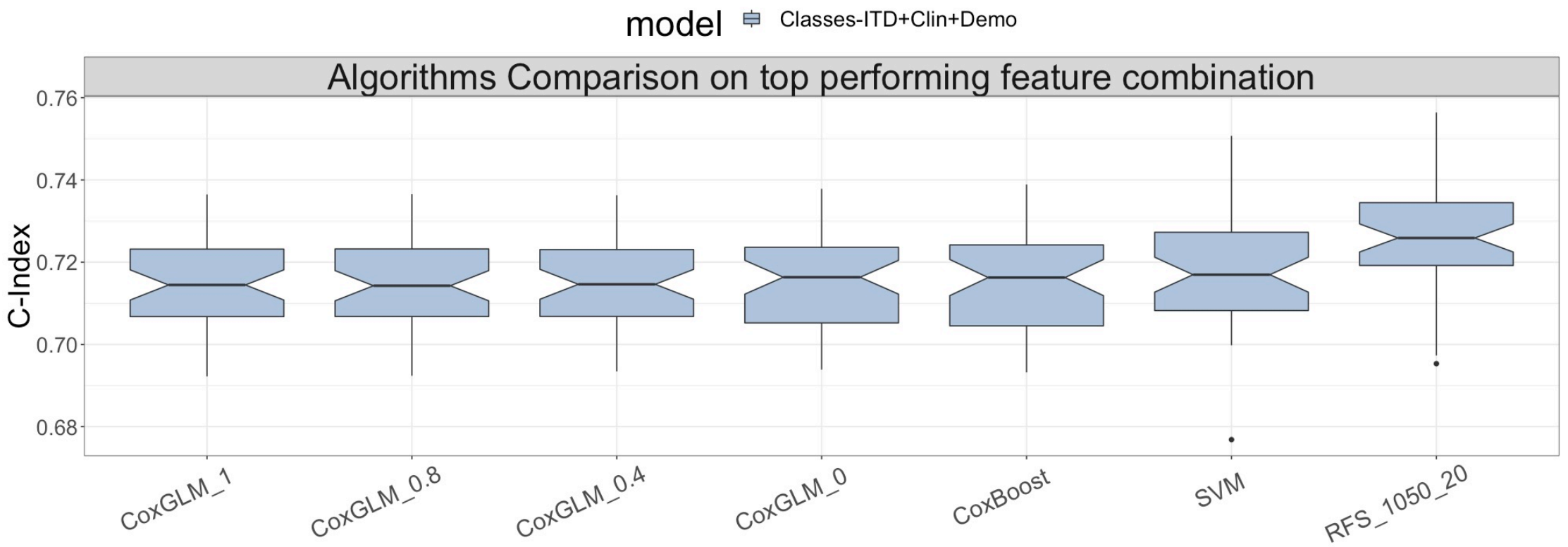


63 y old

40 y old

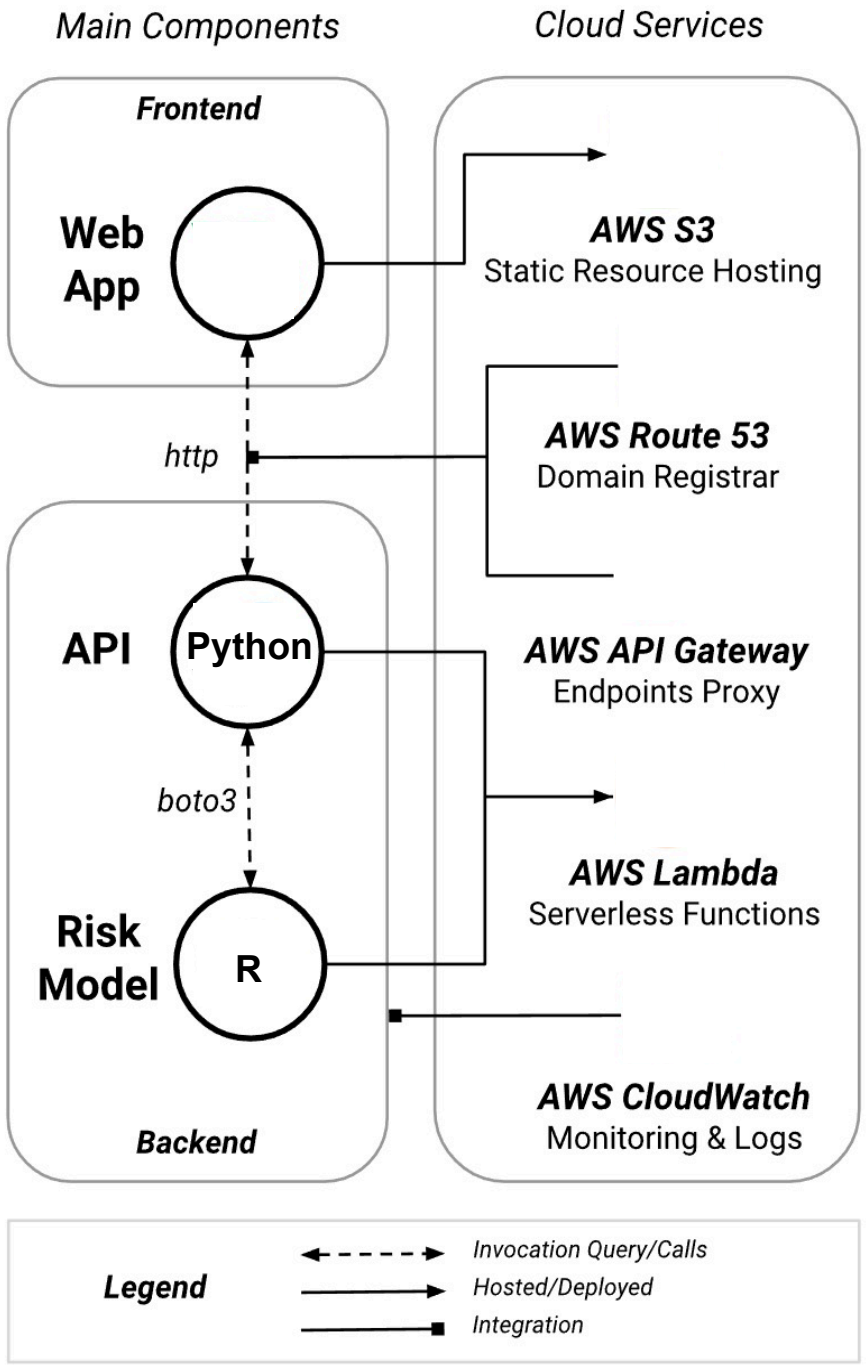


S. Figure 53: Comparison of different statical models C-index distributions (50 repetitions,N=2,113) on the top performing feature combinations. In all boxplots, the median is indicated by the horizontal line and the first and third quartiles by the box edges. The lower and upper whiskers extend from the hinges to the smallest and largest values, respectively, no further than 1.5× interquartile range from the hinges.



S. Figure 54: Risk calculator web infrastructure.

Details of the services infrastructure and the cloud implementation using Amazon Web Services.



Supplementary Tables

S.Table 1: Baseline cohort Characteristics for training (AML NCRI) and validation (AML SG) cohorts.

S.Table 1a	
Training Cohort	**AML NCRI**
Variable	**Distribution in the cohort**
Sample size	**n**
Total	2113
AML 17	1305
AML 16	455
Other trials*	353
Follow up time	**Days median (range)**
AML 17	2290 (36-3897)
AML 16	1468 (126-2664)
Others*	2368 (625-4091)
Type of AML	**Total (AML17,AML 16, Other trials*)**
primary	1748 (1144,334,270)
secondary	260 (103,97,60)
other	105 (58,24,23)
Gender	**Total (AML17,AML 16,others)**
Female	929 (597,186,146)
Male	1184 (708,269,207)
Age	**median (range)**
AML 17	52 (16-101)
AML 16	70 (54-91)
Other trials*	69 (17-91)
ELN 2017 risk group	**Total (AML17,AML 16,others)**
Favorable	864 (566,162,136)
Intermediate	461 (323,71,67)
Adverse	788 (416,222,150)
WBC Count *(10e9/l)*	**median (range)**
AML 17	14 (0.3-456)
AML 16	20 (0-398)
Other trials*	22 (0.5-432)
Platelet Count *(10e9/l)*	**median (range)**
AML 17	60 (2-2013)

S.Table 1b	
Validation Cohort	**AML SG**
Variable	**Distribution in the cohort**
Sample size	**n**
Total	1540
HD98A	627
HD98B	173
07-04	740
Follow up time	**Days median(range)**
HD98A	2693 (30-5384)
HD98B	2709 (1774-5185)
07-04	1781 (35-3080)
Type of AML	**Total (HD98A,HD98B,07-04)**
primary	1376 (572,148,656)
secondary	61 (22,9,30)
other	103 (33,16,54)
Gender	**Total (HD98A,HD98B,07-04)**
Female	723 (306,76,341)
Male	817(321,97,399)
Age	**median(range)**
HD98A	47 (18-65)
HD98B	66 (58-84)
07-04	49 (18-61)
ELN 2017 risk group	**Total (HD98A,HD98B,07-04)**
Favorable	618 (291,48,279)
Intermediate	387 (157,50,180)
Adverse	535 (179,75,281)
WBC Count *(10e9/l)*	**median(range)**
HD98A	14 (0.2-427)
HD98B	8 (0.4-303)
07-04	15 (0.2-533)
Platelet Count *(10e9/l)*	**median(range)**
HD98A	50 (2-746)

AML 16	54 (5-706)
Other trials*	51 (3-937)
Blasts *(%)*	**median**
AML 17	60
AML 16	59
Other trials*	59
Treatment	**Total (AML17,AML 16,others)**
Intensive	1755 (1305,289,161)
Non-intensive	358 (0,166,192)
Transplant type	**AML17**
ALLO	215
AUTO	2
HAPLO	2
MUD/UD	297
Other types	64

HD98B	64 (6-445)
07-04	55 (5-916)
Blasts *(%)*	**median**
HD98A	75
HD98B	75
07-04	75
Transplant type	**Total (HD98A,HD98B,07-04)**
ALLO-HSCT-MRD	302 (119,10,173)
AUTO	95 (87,3,5)
HAPLO	14 (0,0,14)
ALLO-HSCT-MUD	409 (125,2,282)
Other types	

*Other trials include 5 patients from AML11 , 3 from AML12 , 21 from AML14 ,135 from AML15 and 189 from AML Li1

S.Table 2: Repartition of age in each of the classes in the AML NCRI cohort.

Classes	n	Median Age	Min Age	Max Age
WT1	40	41.2	16.8	80.7
t(6;9)	16	41.6	16.9	68.1
biCEBPA	38	43.1	19	84.4
t(11)	75	43.7	15.9	100.7
inv(16)	92	44.3	20.6	78.5
t(8;21)	100	46.4	16	77.8
inv(3)	17	53	16.3	76.6
t(15;17)	19	53.2	29.7	70.6
mNOS	124	54	19.3	84.4
Trisomies	44	55.5	17	79.9
NPM1	674	58.8	16.1	86.4
sAML1	100	59.6	15.9	83.5
No events	46	60.3	22.8	76.5
DNMT3A-IDH	19	61.7	41.7	83.5
TP53-Complex	208	62.5	19.5	91.1
sAML2	501	66.4	23.1	90.8
TOTAL	2113	59.2	15.9	100.7

S.Table 3: Detailed description of the features considered for each prognostic model in the study.

Models	Number of features	Features
Model 1 : ELN 2017 Risk strata	3	favorable,intermediate,adverse risk strata
Model 2 : Genes	84	ASXL1 , ASXL2 , ASXL3 , ATRX , BAGE3 , BCOR , BRAF , CBFB , CBL , CDKN2A , CEBPA_bi , CEBPA_mono , CNTN5 , CREBBP , CSF1R , CSF3R , CTCF , CUL2 , CUX1 , DNMT3A , EED , ETV6 , EZH2 , FBXW7 , ITD , FLT3_TKD , FLT3_other , GATA1 , GATA2 , GNAS , GNB1 , IDH1 , IDH2_p.R140 , IDH2_p.R172 , JAK2 , JAK3 , KANSL1 , KDM6A , KIT , KMT2C , KMT2D , KMT2E , KRAS , LUC7L2 , MED12 , MLL , MPL , MYC , NF1 , NFE2 , NOTCH1 , NPM1 , NRAS_other , NRAS_p.G12_13 , NRAS_p.Q61_62 , PDS5B , PHF6 , PPFIA2 , PRPF8 , PTEN , PTPN11 , PTPRF , PTPRT , RAD21 , RIT1 , RUNX1 , S100B , SETBP1 , SF1 , SF3B1 , SMC1A , SMC3 , SMG1 , SPP1 , SRSF2 , STAG2 , STAT5B , SUZ12 , TET2 , TP53 , U2AF1_p.S34 , U2AF1_p.Q157 , WT1 , ZRSR2
Model 3 : Classes	16	NPM1 , t(11) , TP53_complex , sAML2 , sAML1 , CEBPA_bi , DNMT3A_IDH1_2 , inv(3) , mNOS , t(8;21), no_events , WT1 , inv(16) , Trisomies , t(6;9) , t(15;17)
Model 4 : Classes+ITD	17	NPM1 , t(11) , TP53_complex , sAML2 , sAML1 , CEBPA_bi , DNMT3A_IDH1_2 , inv(3) , mNOS , t(8;21), no_events , WT1 , inv(16) , Trisomies , t(6;9) , t(15;17) , ITD
Model 5 : Genes+Cytogenetics	154	ASXL1 , ASXL2 , ASXL3 , ATRX , BAGE3 , BCOR , BRAF , CBFB , CBL , CDKN2A , CEBPA_bi , CEBPA_mono , CNTN5 , CREBBP , CSF1R , CSF3R , CTCF , CUL2 , CUX1 , DNMT3A , EED , ETV6 , EZH2 , FBXW7 , ITD , FLT3_TKD , FLT3_other , GATA1 , GATA2 , GNAS , GNB1 , IDH1 , IDH2_p.R140 , IDH2_p.R172 , JAK2 , JAK3 , KANSL1 , KDM6A , KIT , KMT2C , KMT2D , KMT2E , KRAS , LUC7L2 , MED12 , MLL , MPL , MYC , NF1 , NFE2 , NOTCH1 , NPM1 , NRAS_other , NRAS_p.G12_13 , NRAS_p.Q61_62 , PDS5B , PHF6 , PPFIA2 , PRPF8 , PTEN , PTPN11 , PTPRF , PTPRT , RAD21 , RIT1 , RUNX1 , S100B , SETBP1 , SF1 , SF3B1 , SMC1A , SMC3 , SMG1 , SPP1 , SRSF2 , STAG2 , STAT5B , SUZ12 , TET2 , TP53 , U2AF1_p.S34 , U2AF1_p.Q157 , WT1 , ZRSR2 , +8 , +11 , +13 , +21 , +22 , -20 , -3 , -5 , -7 , -9 , -12 , -13 , -16 , -17 , -18 , minusy , t(v;11) , t(10;21) , t(12;13) , t(12;17) , t(12;22) , t(13;19) , t(15;16) , t(15;17) , t(16;17) , t(16;21) , t(17;19) , t(17;21) , t(1;12) , t(1;14) , t(1;16) , t(1;17) , t(1;19) , t(1;3) , t(1;4) , t(1;5) , t(1;6) , t(2;17) , t(2;3) , t(2;5) , t(2;7) , t(2;9) , t(3;16) , t(3;21) , t(3;5) , t(3;7) , t(3;9) , t(4;12) , t(4;21) , t(4;9) , t(5;12) , t(5;17) , t(5;9) , t(6;9) , t(7;16) , t(7;17) , t(7;8) , t(8;10) , t(8;13) , t(8;16) , t(8;17) , t(8;21) , t(9;11) , t(9;13) , t(9;17) , t(9_22) , complex , others_transloc , inv(3) , inv(16)
Model 6 : Clinical+Demographic	9	ahd , perf_status , bm_blasts , secondary , wbc , hb , plt , gender , age
Model 7 : Genes+Cytogenetics+Clinical+Demographic	163	ASXL1 , ASXL2 , ASXL3 , ATRX , BAGE3 , BCOR , BRAF , CBFB , CBL , CDKN2A , CEBPA_bi , CEBPA_mono , CNTN5 , CREBBP , CSF1R , CSF3R , CTCF , CUL2 , CUX1 , DNMT3A , EED , ETV6 , EZH2 , FBXW7 , ITD , FLT3_TKD , FLT3_other , GATA1 , GATA2 , GNAS , GNB1 , IDH1 , IDH2_p.R140 , IDH2_p.R172 , JAK2 , JAK3 , KANSL1 , KDM6A , KIT , KMT2C , KMT2D , KMT2E , KRAS , LUC7L2 , MED12 , MLL , MPL , MYC , NF1 , NFE2 , NOTCH1 , NPM1 , NRAS_other , NRAS_p.G12_13 , NRAS_p.Q61_62 , PDS5B , PHF6 , PPFIA2 , PRPF8 , PTEN , PTPN11 , PTPRF , PTPRT , RAD21 , RIT1 , RUNX1 , S100B , SETBP1 , SF1 , SF3B1 , SMC1A , SMC3 , SMG1 , SPP1 , SRSF2 , STAG2 , STAT5B , SUZ12 , TET2 , TP53 , U2AF1_p.S34 , U2AF1_p.Q157 , WT1 , ZRSR2 , +8 , +11 , +13 , +21 , +22 , -20 , -3 , -5 , -7 , -9 , -12 , -13 , -16 , -17 , -18 , minusy , t(v;11) , t(10;21) , t(12;13) , t(12;17) , t(12;22) , t(13;19) , t(15;16) , t(15;17) , t(16;17) , t(16;21) , t(17;19) , t(17;21) , t(1;12) , t(1;14) , t(1;16) , t(1;17) , t(1;19) , t(1;3) , t(1;4) , t(1;5) , t(1;6) , t(2;17) , t(2;3) , t(2;5) , t(2;7) , t(2;9) , t(3;16) , t(3;21) , t(3;5) , t(3;7) , t(3;9) , t(4;12) , t(4;21) , t(4;9) , t(5;12) , t(5;17) , t(5;9) , t(6;9) , t(7;16) , t(7;17) , t(7;8) , t(8;10) , t(8;13) , t(8;16) , t(8;17) , t(8;21) , t(9;11) , t(9;13) , t(9;17) , t(9_22) , complex , others_transloc , inv(3) , inv(16) , ahd , perf_status , bm_blasts , secondary , wbc , hb , plt , gender , age
Model 8 : Classes+ITD+Clinical+Demographic	26	NPM1 , t(11) , TP53_complex , sAML2 , sAML1 , CEBPA_bi , DNMT3A_IDH1_2 , inv(3) , mNOS , t(8;21), no_events , WT1 , inv(16) , Trisomies , t(6;9) , t(15;17) , ITD , ahd , perf_status , bm_blasts , secondary , wbc , hb , plt , gender , age

S.Table 4: Comparison of coefficients for a Cox Proposal Hazard model and a Cox Proportional Hazard weighted model for the classes, ELN and the risk propoosal. Pvalues are computed with Wald test.

		COX PH	COX PH	COX PH WEIGHTED	COX PH WEIGHTED
Model	**Group**	**Hazard ratio [95%CI]**	**p-val**	**Hazard ratio [95%CI]**	**p-val**
Risk Proposal	**Favorable**	0.62 [0.54,0.72]	2.9 e-10	0.63 [0.54,0.75]	4.4 e-8
	Adverse	2.00 [1.77,2.27]	0	1.98 [1.73,2.26]	0
	Intermediate	Ref	Ref	Ref	Ref
ELN	**Favorable**	0.63 [0.55,0.73]	9.8 e-10	0.68 [0.58,0.79]	8.2 e-7
	Adverse	1.74 [1.52,1.99]	1.1 e-15	1.79 [1.56,2.06]	1.1 e-16
	Intermediate	Ref	Ref	Ref	Ref
Classes	**NPM1**	1.29 [0.82,2.05]	0.27	1.36 [0.83,2.21]	0.22
	t(11)	1.52 [0.89,2.58]	0.12	1.40 [0.80,2.43]	0.23
	TP53-complex	5.3 [3.31,8.52]	4.3 e-12	5.23 [3.17,8.62]	8.9 e-11
	sAML2	2.64 [1.67,4.18]	3.6 e-5	2.60 [1.60,4.22]	1.2 e-4
	sAML1	1.48 [0.89,2.47]	0.13	1.42 [0.83,2.43]	0.2
	DNMT3A-IDH	2.41 [1.23,4.75]	0.01	2.38 [1.18,4.79]	0.02
	inv(3)	4.79 [2.49,9.23]	2.9 e-6	4.26 [2.37,7.65]	1.2 e-6
	mNOS	1.38 [0.84,2.27]	0.21	1.38 [0.81,2.36]	0.23
	t(8;21)	0.58 [0.33,1.02]	0.06	0.54 [0.30,0.97]	0.04
	no events	1.06 [0.59,1.91]	0.85	0.89 [0.49,1.61]	0.7
	WT1	2.32 [1.31,4.10]	4.0 e-3	2.26 [1.24,4.11]	7.7 e-3
	inv(16)	0.51 [0.29,0.92]	0.03	0.45 [0.24,0.82]	9.5 e-3
	Trisomies	1.46 [0.81,2.63]	0.21	1.60 [0.85,3.02]	0.14
	t(6;9)	1.87 [0.92,3.79]	0.08	1.55 [0.84,2.88]	0.16
	t(15;17)	0.35 [0.12,1.04]	0.06	0.45 [0.13,1.50]	0.19
	BiCEBPA	Ref	Ref	Ref	Ref

S.Table 5: Cohort characteristics for merged AML NCRI and AMLSG subsets that were used for HSCT analysis.

Variable	Distribution in the cohort
Sample Size	**n**
Total	2244
AML NCRI	1095
AML SG	1149
Follow up time	**Days median (range)**
Total	2213 (35-5384)
AML NCRI	2291 (36-3897)
AML SG	2217 (35-5384)
Type of AML (Total, AML NCR, AML SG)	**Total**
primary	2034, 985, 1049
secondary	98, 69, 29
others	112, 41, 71
Gender (Total, AML NCRI, AML SG)	**Total**
Female	1080, 514, 566
Male	1164, 581, 583
Age	**median (range)**
Total	50 (16-101)
AML NCRI	51 (16-101)
AML SG	49 (18-80)
ELN 2017 risk group (Total, AML NCR, AML SG)	**Total**
Favorable	1070, 521, 549
Intermediate	556, 273, 283
Adverse	618, 301, 317
WBC Count (10e9/l)	**median (range)**

Total	14 (0.2-456)
AML NCRI	14 (0.3-456)
AML SG	14 (0.2-303)
Platelet Count (10e9/l)	**median (range)**
Total	56 (2-2013)
AML NCRI	60 (2-2013)
AML SG	53 (5-746)
Blasts *(%)*	**median**
Total	70
AML NCRI	60
AML SG	75

S.Table 6: Detailed state by state distribution of transplanted patients that achieved complete remission in combined training and validation cohort (n=2,244).

	to							
from	**CR1**	**TPL in CR1**	**TPL in CR2**	**Death no TPL**	**Death from TPL in CR1**	**Death from TPL in CR2**	**no event**	**Total**
CR1		n=759	n=436	n=507			n=542	n=2244
TPL in CR1					n=347		n=412	n=759
TPL in CR2						n=236	n=200	n=436
Death no TPL							n=507	n=507
Death from TPL in CR1							n=347	n=347
Death from TPL in CR2							n=236	n=236

S.Table 7: HSCT characteristics for AML NCRI and AMLSG subsets by ELN 2017 and proposed risk score.

	AML NCRI cohort N=1,095 (568 | 321 | 206)	AML SG cohort N=1,149 (481 | 438 | 230)
	Total (No transplant | Transplant in CR1 | Transplant in CR2)	Total (No transplant | Transplant in CR1 | Transplant in CR2)
ELN Favorable	521 (326 | 80 | 115)	549 (288 | 149 | 112)
ELN Intermediate	273 (109 | 113 | 51)	283 (97 | 121 | 65)
ELN Adverse	301 (133 | 128 | 40)	317 (96 | 168 | 53)
Proposal Favorable	441 (279 | 66 | 96)	517 (258 | 142 | 117)
Proposal Intermediate	344 (147 | 130 | 67)	375 (139 | 163 | 73)
Proposal Adverse	310 (142 | 125 | 43)	257 (84 | 133 | 40)

S.Table 8: Panel of 32 genes sufficient to deliver complete classification and risk stratification.

Class defining genes	Independent genes
NPM1	KIT
TP53	KRAS
WT1	NRAS
CEBPA	PTPN11
DNMT3A	GATA2
IDH1	MYC
IDH2	MPL
ZRSR2	SMC3
U2AF1	**Total ,N=8**
SRSF2	
SF3B1	
ASXL1	
STAG2	
BCOR	
RUNX1	
EZH2	
MLL	
PHF6	
SF1	
NF1	
CUX1	
SETBP1	
FLT3	
TET2	
Total ,N=24	

S.Table 9: Hyperparameter selection for BDP processes in study.

Hyperparameter names	Hyperparameter search range	BDP1	BDP2
Burnin iterations	3000 to 7000	7000	5000
Number of chains	3 to 7	3	3
Number of iterations between collected samples	20	20	20
Cosine similarity threshold	0.9	0.9	0.9
Number of posterior samples	150-1000	250	150
Initial number of clusters	3 to 30	17	5
Base distribution	Uniform dirichlet, Gaussian, binomial	Gaussian	Gaussian
Concentration parameter αA	0.1 to 100	0.5	2
Concentration parameter αB	0.1 to 100	1.5	6

Supplementary Appendix

1. Study Participants

1.1 AML NCRI Cohort (Validation)

The following AML Study Group (AML 17 NCRI) institutions and investigators participated in this study:

Dominic Culligan, M.D, Aberdeen Royal Infirmary, UK; Kiran Tawana, M.D, Addenbrookes University Hospital, UK; Ranjit Dasgupta, M.D, Arrowe Park Hospital, UK; Andres Virchis, M.D, Barnet General Hospital, UK; Mary Mcmullin, Prof, Belfast City Hospital, UK; Manoj Raghavan, M.D, Birmingham Heartlands Hospital, UK; Paul Cahalin, M.D, Blackpool Victoria Hospital, UK; Sam Ackroyd, M.D, Bradford Royal Infirmary, UK; Priyanka Mehta, M.D, Bristol Haematology and Oncology Centre, UK; Adam Rye, M.D, Cheltenham General Hospital, UK; Emma Welch, M.D, Chesterfield Royal Hospital, UK; Salaheddin Tueger, M.D, Countess of Chester Hospital, UK; Hannah Hunter, M.D, Derriford Hospital, UK; Atchamamba Bobbili, M.D, Doncaster Royal Infirmary, UK; Roderick Neilson, M.D, Falkirk and District Royal Hospital, UK; Earnest Heartin, M.D, Glan Clwyd Hospital, UK; Adam Rye, M.D, Gloucestershire Royal Hospital, UK; Norbert Blesing, M.D, Great Western Hospital, UK; Kavita Raj, M.D, Guys & St Thomas Hospital, UK; Jirri Pavlu, M.D, Hammersmith Hospital, UK; Richard Kaczmarski, M.D, Hillingdon Hospital, UK; Debo Ademokun, M.D, Ipswich Hospital, UK; Angela Wood, M.D, James Cook University Hospital, UK; Cesar Gomez, M.D, James Paget Hospital, UK; Paresh Vyas, M.D, John Radcliffe Hospital, UK; Jindriska Lindsay, M.D, Kent and Canterbury Hospital, UK; Mark Kwan, M.D, Kettering General hospital, UK; Ghulam Mufti, Prof, Kings College Hospital, UK; Kate Hodgson, M.D, Leicester Royal Infirmary, UK; Charlotte Kallmeyer, M.D, Lincoln County Hospital, UK; Eleni Tholouli, M.D, Manchester Royal Infirmary, UK; Maadh Aldouri, M.D, Medway Maritime Hospital, UK; Moez Dungarwalla, M.D, Milton Keynes Hospital, UK; Simon Bolam, M.D, Musgrove Park Hospital, UK; Abraham Jacob, M.D, New Cross Hospital, UK; S Tauro, M.D, Ninewells Hospital and Medical Centre, UK; Matthew Lawes, M.D, Norfolk and Norwich University Hospital, UK; Nicki Panoskaltsis, M.D, Northwick Park Hospital, UK; Jenny Byrne, M.D, Nottingham University Hospital and Paediatrics, UK; Sateesh Nagumantry, M.D, Peterborough City Hospital, UK; Paul Moreton, M.D, Pinderfields Hospital, UK; Darshayani Furby, M.D, Poole Hospital, UK; Mansour Ceesay, M.D, Princess Royal Hospital, UK; Mary Ganczakowski, M.D, Queen Alexandra Hospital, Portsmouth, UK; Charles Craddock, Prof, Queen Elizabeth Hospital, Birmingham, UK; Peter Coates, M.D, Queen Elizabeth Hospital, Kings Lynn, UK; Sabia Rashid, M.D, Queen Elizabeth Hospital, Woolwich, UK; Paul Greaves, M.D, Queens Hospital, UK; Caroline Duncan, M.D, Raigmore Hospital, UK; Pratap Neelakantan, M.D, Royal Berkshire Hospital, UK; Rachel Hall, M.D, Royal Bournemouth Hospital, UK; Bryson Pottinger, M.D, Royal Cornwall Hospital, UK; Juanah Addada, M.D, Royal Derby Hospital, UK; Asim Khwaja, M.D, Royal Free Hospital, UK; Mohammad Mohsin, M.D, Royal Hallamshire Hospital, UK; Andrea Corcoz, M.D,

Royal Shrewsbury, UK; Louise Hendry, M.D, Royal Surrey Hospital, UK; Timothy Corbett, M.D, Royal Sussex County Hospital, UK; Josephine Crowe, M.D, Royal United Hospitals Bath NHS Foundation Trust, UK; Craig Taylor, M.D, Russells Hall Hospital, UK; Rowena Thomas-Dewing, M.D, Salford Royal Hospital, UK; Jonathan Cullis, M.D, Salisbury Hospital NHS Foundation, UK; Yasmin Hasan, M.D, Sandwell Hospital, UK; Unmesh Mohite, M.D, Singleton Hospital, UK; Deborah Richardson, M.D, Southampton General Hospital, UK; Gail Loudon, Southern General Hospital, New Victoria, UK; Jamie Cavenagh, St Bartholomew's Hospital, UK; Dr Richard Kelly, St James University Hospital, UK; Jamie Wilson, M.D, St Richard's Hospital, UK; Srinivas Pillai, M.D, Stafford Hospital, UK; Helen Eagleton, M.D, Stoke Mandeville Hospital, UK; Shikha Chattree, M.D, Sunderland Royal Hospital, UK; Marc Drummond, M.D, The Beatson WOS Cancer Centre, UK; Mike Dennis, M.D, The Christie NHS Foundation Trust, UK; Arpad Toth, M.D, The Royal Liverpool Hospital, UK; Mark Ethell, M.D, The Royal Marsden Hospital, UK; Gail Jones, M.D, The Royal Victoria Infirmary Freeman Hospital, UK; Deborah Turner, M.D, Torbay District General Hospital, UK; Asim Khwaja, Prof, University College London Hospitals, UK; Vikram Singh, M.D, University Hospital Aintree, UK; Julie Gillies, M.D, University Hospital Ayr, UK; Beth Harrison, M.D, University Hospital Coventry, Walsgrave, UK; William Gordon, M.D, University Hospital Crosshouse, UK; Naheed Mir, M.D, University Hospital Lewisham, UK; Srinivas Pillai, M.D, University Hospital of North Staffordshire, UK; Maria Szubert, M.D, University Hospital of North Tees and Hartlepool, UK; Jonathan Kell, M.D, University Hospital of Wales, UK; Kerri Davidson, M.D, Victoria Hospital, NHS Fife, UK; Toby Nicholson, M.D, Whiston Hospital, UK; Salim Shafeek, M.D, Worcestershire Royal Hospital, UK; Lee Bond, M.D, York Hospital, UK; Ernest Heartin, M.D, Ysbyty Gwynedd Hospital, UK; Ruth Spearing, M.D, Christchurch Hospital, New Zealand; John Carter, M.D, Wellington Hospital, New Zealand; Marianne Tong Severinsen, M.D, Aalborg University Hospital, Denmark; Ingolf Mølle, M.D, Aarhus University Hospital, Denmark; Ulrik Malthe Overgaard, M.D, Herlev Hospital, Denmark; Claus Werenberg Marcher, M.D, Odense Hospital, Denmark; Ulrik Malthe Overgaard, M.D, Rigshospitalet, Denmark

1.2 AML SG Study Contributors

The following AML Study Group (AMLSG) institutions and investigators participated in this study:

Peter Brossart, M.D., Universitätsklinikum Bonn, Bonn Germany; Bernd Hertenstein M.D., Henrike Thomssen M.D., Klinikum Bremen-Mitte, Bremen, Germany; Rainer Haas M.D., Andrea Kuendgen M.D., Universitätsklinikum Düsseldorf, Düsseldorf, Germany; Peter Reimer M.D., Mohammed Wattad M.D., Kliniken Essen Süd, Ev. Krankenhaus Essen-Werden gGmbH, Essen, Germany; Carsten Schwaenen M.D., Klinikum Esslingen, Esslingen, Germany; Hans Günter Derigs M.D., Klinikum Frankfurt-Höchst GmbH, Frankfurt, Germany; Michael Lübbert M.D., Universitätsklinikum Freiburg, Freiburg, Germany; Alexander Burchardt M.D., Matthias Rummel M.D., Universitätsklinikum Gießen, Gießen, Germany; Volker Runde M.D., Wilhelm-Anton-Hospital, Goch, Germany; Gerald Wulf M.D., Lorenz Trümper M.D., Universitätsklinikum Göttingen, Göttingen, Germany; Walter Fiedler M.D., Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany; Hans Salwender M.D., Asklepios Klinik Altona, Hamburg, Germany; Elisabeth Lange M.D., Evangelisches Krankenhaus

Hamm, Hamm, Germany; Andrea Sendler M.D., Klinikum Hanau, Hanau, Germany; Arnold Ganser M.D., Jürgen Krauter, M.D., Brigitte Schlegelberger M.D., Medizinische Hochschule Hannover, Hannover, Germany; Hartmut Kirchner M.D. KRH Klinikum Siloah, Hannover, Germany; Uwe Martens M.D., SLK-Kliniken GmbH Heilbronn, Heilbronn, Germany; Michael Pfreundschuh M.D., Gerhard Held M.D., Universitätsklinikum des Saarlandes, Homburg, Germany; David Nachbaur M.D., Günter Gastl M.D., Universitätsklinikum Innsbruck, Innsbruck, Austria; Mark Ringhoffer M.D., Martin Bentz M.D., Städtisches Klinikum Karlsruhe gGmbH, 4 Karlsruhe, Germany; Heinz A. Horst M.D., Michael Kneba M.D., Universitätsklinikum Schleswig-Holstein–Campus Kiel, Kiel, Germany; Stephan Kremers M.D., Caritas-Krankenhaus Lebach, Lebach, Germany; Andreas Petzer M.D., Krankenhaus der Barmherzigen Schwestern Linz, Linz, Austria; Gerhard Heil M.D., Klinikum Lüdenscheid, Lüdenscheid, Germany; Thomas Kindler M.D., Matthias Theobald M.D., Universitätsklinikum Mainz, Mainz, Germany; Katharina Götze M.D., Christian Peschel M.D., Klinikum rechts der Isar der Technischen Universität München, München, Germany; Sabine Struve M.D. Klinikum Schwabing, München, Germany; Peter Schmidt M.D., Städtisches Klinikum Neunkirchen, Neunkirchen, Germany; Ali-Nuri Hünerlitürkoglu M.D., Lukaskrankenhaus GmbH Neuss, Neuss; Germany; Claus-Henning Köhne M.D., Klinikum Oldenburg, Oldenburg, Germany; Axel Matzdorff M.D., Caritas-Klinik St. Theresia, Saarbrücken, Germany; Richard Greil M.D., Gudrun Russ M.D., Universitätsklinikum der Paracelsus Medizinischen Universität Salzburg, Salzburg, Austria; Jochen Greiner M.D., Diakonie-Klinikum Stuttgart, Stuttgart, Germany; Heinz Kirchen M.D., Krankenhaus der Barmherzigen Brüder, Trier, Germany; Hans-Gernot Biedermann M.D., Kreisklinik Trostberg, Trostberg, Germany; Helmut Salih M.D., Lothar Kanz M.D., Universitätsklinikum Tübingen, Tübingen, Germany; Hartmut Döhner M.D., Konstanze Döhner M.D., Richard F. Schlenk M.D, Universitätsklinikum Ulm, Ulm, Germany; Wolfgang Brugger M.D., Schwarzwald-Baar Klinikum Villingen-Schwenningen GmbH, Villingen-Schwenningen, Germany; Elisabeth Koller M.D., Hanuschkrankenhaus, Wien, Austria; Aruna Raghavachar M.D., Helios-Klinikum Wuppertal, Wuppertal, Germany.