blood advances

American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
bloodadvances@hematology.org

A 69 long non-coding RNA signature predicts relapse and acts as independent prognostic factor in pediatric AML

Tracking no: ADV-2024-012667R1

Zhiyao Ren (Ghent University, Belgium) Jolien Vanhooren (Ghent University, Belgium) Charlotte Derpoorter (Ghent University, Belgium) Barbara De Moerloose (Ghent University Hospital, Belgium) Tim Lammens (Ghent University Hospital, Belgium)

Abstract:
Risk stratification using genetics and minimal residual disease (MRD) has allowed to increase the cure rates of pediatric acute myeloid leukemia (pedAML) up to 70% in contemporary protocols. Nevertheless, approximately 30% of patients still experience relapse, indicating a need to optimize stratification strategies. Recently, long non-coding RNA (lncRNA) expression has been shown to hold prognostic power in multiple cancer types. Here, we aimed at refining relapse prediction in pedAML using lncRNA expression. We built a relapse-related lncRNA prognostic signature, named AMLlnc69, using 871 pedAML patients transcriptomes obtained from the Therapeutically Applicable Research To Generate Effective Treatments (TARGET) repository. We identified a 69 lncRNA signature AMLlnc69 that is highly predictive of relapse-risk (c-index = 0.73), with area under the ROC curve (AUC) values for predicting the 1-, 2-, and 3-year relapse-free survival (RFS) of 0.78, 0.77, and 0.77, respectively. The internal validation using a bootstrap method (resampling times = 1000) resulted in a c-index of 0.72 and AUC values for predicting the 1-, 2-, and 3-year RFS of 0.77, 0.76, and 0.76, respectively. Through a Cox regression analysis, AMLlnc69, NPM mutation and WBC at diagnosis were identified as independent predictors of RFS. Finally, a nomogram was build using these two parameters, showing a c-index of 0.80 and 0.71 after bootstrapping (n =1000). In conclusion, the identified AMLlnc69 will, after prospective validation, add important information to guide management of pedAML patients. The nomogram is a promising tool for easy stratification of patients into a novel scheme of relapse-risk groups.

Conflict of interest: No COI declared

COI notes:

Preprint server: No;

Author contributions and disclosures: Z.R, J.V, and T.L drafted the manuscript. Z.R, J.V and T.L designed the figures. All authors critically revised the manuscript and approved the final version.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: The results published here are in whole or part based upon data generated by the Therapeutically Applicable Research to Generate Effective Treatments (https://www.cancer.gov/ccg/research/genome-sequencing/target) initiative, phs000218. The data used for this analysis are available at the Genomic Data Commons (https://portal.gdc.cancer.gov).

Clinical trial registration information (if any):

A 69 long non-coding RNA signature predicts relapse and acts as independent prognostic factor in pediatric AML

Zhiyao Ren[1,2,3], Jolien Vanhooren[1,2,3], Charlotte Derpoorter[1,2,3], Barbara De Moerloose[1,2,3], Tim Lammens[1,2,3]

[1] Department of Internal Medicine and Pediatrics, Ghent University, 9000 Ghent, Belgium

[2] Department of Pediatric Hematology-Oncology and Stem Cell Transplantation, Ghent University Hospital, 9000 Ghent, Belgium

[3] Cancer Research Institute Ghent, 9000 Ghent, Belgium

Corresponding author

Tim Lammens

Department of Pediatric Hematology-Oncology and Stem Cell Transplantation

Ghent University Hospital

9000 Ghent

Belgium

E-mail address: Tim.Lammens@UGent.be

Running title

LncRNA expression in AML: an independent prognostic factor

Key Points

A comprehensive prognostic model of 69 lncRNAs was generated to predict RFS in pedAML and further refining the current risk stratification

A nomogram incorporating the 69 lncRNAs signature, NPM mutation and WBC was developed to stratify patients into relapse-risk groups

Data sharing statement

The results published here are in whole or part based upon data generated by the Therapeutically Applicable Research to Generate Effective Treatments initiative, phs000218. The data used for this analysis are available at the Genomic Data Commons (https://portal.gdc.cancer.gov).

Abstract

Risk stratification using genetics and minimal residual disease (MRD) has allowed to increase the cure rates of pediatric acute myeloid leukemia (pedAML) up to 70% in contemporary protocols. Nevertheless, approximately 30% of patients still experience relapse, indicating a need to optimize stratification strategies. Recently, long non-coding RNA (lncRNA) expression has been shown to hold prognostic power in multiple cancer types. Here, we aimed at refining relapse prediction in pedAML using lncRNA expression. We built a relapse-related lncRNA prognostic signature, named AML^{lnc69}, using 871 pedAML patients transcriptomes obtained from the Therapeutically Applicable Research To Generate Effective Treatments (TARGET) repository. We identified a 69 lncRNA signature AML^{lnc69} that is highly predictive of relapse-risk (c-index = 0.73), with area under the ROC curve (AUC) values for predicting the 1-, 2-, and 3-year relapse-free survival (RFS) of 0.78, 0.77, and 0.77, respectively. The internal validation using a bootstrap method (resampling times = 1000) resulted in a c-index of 0.72 and AUC values for predicting the 1-, 2-, and 3-year RFS of 0.77, 0.76, and 0.76, respectively. Through a Cox regression analysis, AML^{lnc69}, NPM mutation and WBC at diagnosis were identified as independent predictors of RFS. Finally, a nomogram was build using these two parameters, showing a c-index of 0.80 and 0.71 after bootstrapping (n =1000). In conclusion, the identified AML^{lnc69} will, after prospective validation, add important information to guide management of pedAML patients. The nomogram is a promising tool for easy stratification of patients into a novel scheme of relapse-risk groups.

Introduction

The 5-year overall survival (OS) rate for pediatric acute myeloid leukemia (pedAML) is now up to 70% with the application of most contemporary protocols[1]. This improvement has been achieved through treatment intensification, the optimization of transplant procedures and supportive care, and the introduction of risk-adapted treatment strategies[2]. Risk stratification for pedAML depends mainly on the presence or absence of cytogenetic and molecular abnormalities known to be associated with the achievement of complete remission (CR), OS, and relapse-free survival (RFS)[3-5]. The achievement of minimal residual disease (MRD) during treatment has been shown to be another important risk stratification indicator enabling treatment adaptation[6-8]. Nevertheless, pedAML remains a therapeutic challenge, with high (~30%) relapse rates despite intensive therapy[9]. Relapse is a major cause of pedAML treatment failure and an indicator of poor prognosis[10]. Recent coding gene expression analyses revealed the ability of the LSC6 and LSC17 stemness signatures, reflecting the expression of 6 and 17 coding mRNAs, respectively, to predict the event-free survival (EFS) and OS of patients with pedAML[11,12] . Similarly, the LSC47 signature was developed to predict EFS in the context of existing cytogenetic and molecular risk stratification[13]. Recently, the lncScore, a long non-coding RNA (lncRNA)-based predictor of the EFS and OS of patients with pedAML, was developed[14]. However, none of these signatures predicts RFS or has been implemented in a clinical setting. Thus, the further refinement of relapse-based risk stratification for pedAML remains an urgent need.

LncRNAs are transcripts longer than 200 nucleotides that are not translated into proteins and have highly tissue-specific expression[15]. They have been shown to play important roles in normal development and the development of diseases[16,17], including pedAML[18]. The aberrant

expression of key lncRNAs involved in hematopoietic stem cell maintenance and differentiation has been shown to result in the development of hematological malignancies[19,20]. LncRNA expression also has prognostic power for malignancies such as adult AML[21], breast cancer[22], and neuroblastoma[23]. Hence, the inclusion of lncRNA expression in pedAML risk stratification could add value for RFS prediction.

In this study, we aimed to identify an lncRNA signature that predicts pedAML RFS using publicly available RNA sequencing (RNA-seq) data from patients with pedAML from the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) repository, including cases from Children's Oncology Group (COG) studies AAML0531[24] and AAML1031[25,26].

Methods

TARGET data acquisition

RNA-seq data and corresponding clinical information on patients with pedAML were retrieved from the TARGET repository (https://ocg.cancer.gov/programs/target) and downloaded from the Genomic Data Commons (https://portal.gdc.cancer.gov). We included pedAML cases from COG studies AAML0531[24] and AAML1031[25,26]. During enrollment in those clinical trials, which were conducted in accordance with the Declaration of Helsinki, participants provided written informed consent to the use of their data for correlative biological studies. We excluded cases evaluated by low-depth sequencing and sequencing data obtained at non-diagnostic time points, such as after treatment or relapse. As we focused on relapse, only patients whose first event was relapse and those who were censored were included. Patients from the AAML1031 study were allocated to the discovery cohort (n = 871) and those from the AAML0531 study were allocated to the validation cohort (n = 158;

Supplemental Figure 1). Detailed information on each of the included patients is provided in Supplemental Table 1, and the value labels used in Supplemental Table 1 are provided in Supplemental Table 2. Gene annotation was performed using Homo_sapiens.GRCh38.110 from the Ensembl project (http://www.ensembl.org)[27].

lncRNA prognostic signature construction

Univariate Cox regression analysis was used to identify lncRNAs associated significantly with RFS ($p < 0.05$). Then, least absolute shrinkage and selection operator (LASSO) Cox regression was used to further filter for lncRNAs associated strongly with RFS, and to estimate their coefficients for linear predictor establishment. During this process, the optimal parameter 'λ' was obtained through cross-validation to maintain equilibrium between model deviation and variance. The caret, tidyverse, tibble, data.table, survival, survminer, glmnet, pbapply, and magrittr R packages were used for these analyses. Next, an outcome-oriented method was applied using the surv_cutpoint function of the survminer R package to determine an optimal cutoff value that maximized survival differences between low- and high-risk lncRNA signature groups. Finally, the risk score distribution was plotted and RFS status mapping was performed to further analyze the influence of the selected signature on RFS.

Outcome analysis and validation

Using the survival and survminer R packages and the Kaplan–Meier (KM) method, RFS and OS probabilities were estimated. Standard errors were calculated using the Greenwood formula, and the data were compared using the log-rank test. Receiver operating characteristic (ROC) curves were drawn and the areas under the curves (AUCs) were calculated using the survminer and timeROC R package to assess model performance at 1, 2,

and 3 years. Concordance (c)-indices were calculated for signature assessment using the concordance function[28]. The bootstrap method, based on 1000-fold resampling, was used for internal validation[29,30]. Univariate and multivariate Cox regression analyses were performed to obtain independent prognostic values for the lncRNA signature, and c-indices of these values were calculated. Finally, a nomogram was generated using the regplot, survival, rms, ggDCA, and timeROC R packages[31]. The coefficients of lncRNAs in the model generated from discovery cohort data were used for external validation.

Comparison of clinical characteristics

Fisher's exact test and the χ^2 test were used to assess distributional differences in categorical data between the low- and high-risk groups. A heatmap of these differences was generated using the ComplexHeatmap R package. The normality of data distributions was assessed using the D'Agostino–Pearson, Anderson–Darling, Shapiro–Wilk, and Kolmogorov–Smirnov tests. Non-normally distributed quantitative data were evaluated using the Mann–Whitney *U* test. Box plots of quantitative results were generated using GraphPad Prism 9.

Functional and biological pathway enrichment analyses

Gene set enrichment analysis (GSEA) was performed using the limma, org.Hs.eg.db, clusterProfiler, and enrichplot R packages ($p < 0.05$)[32]. The hallmark, gene ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets were downloaded from MSigDB (http://www.broadinstitute.org/msigdb)[33] and used for this purpose.

Results

Prognostic lncRNA signature establishment

The characteristics of patients in the discovery and validation cohorts are summarized in Supplemental Table 3, and RFS in the two cohorts is illustrated in Supplemental Figure 2. Univariate Cox regression analysis of discovery cohort data yielded 1751 relapse-related lncRNAs (Supplemental Table 4). To improve the prediction accuracy and avoid overfitting, a LASSO Cox regression analysis was performed to construct a linear prognostic model (AML^{lnc69}), in which 69 lncRNAs were retained (Figure 1A, Supplemental Figure 3, Supplemental Table 5,). AML^{lnc69} risk group information, survival data, and expression levels of the lncRNAs used in model construction are provided for all patients in the discovery cohort in Supplemental Table 6. An optimal cutoff value was calculated using the surv_cutpoint function in the survminer R package and an outcome-oriented method that enabled patient assignment to low- and high-risk AML^{lnc69} groups. The risk score distribution is shown in Figure 1B. The RFS status map for all patients with pedAML shows that the proportion of patients with relapse increases with the AML^{lnc69} score (Figure 1C).

Signature validation

KM survival curves further illustrated that the RFS rate was significantly lower for high-risk AML^{lnc69} patients than low-risk AML^{lnc69} patients in the discovery cohort (Figure 2A). In addition, the AML^{lnc69} signature was significantly predictive of OS (Figure 2B). To further substantiate the predictive value of the model, we performed ROC analyses with the calculation of AUCs for the prediction of 1-, 2-, and 3-year RFS; these values were 0.78, 0.77, and 0.77, respectively, with a corresponding c-index of 0.73 (Figure 2C). Internal validation yielded a c-index of 0.72 and AUCs for the prediction of 1-, 2-, and 3-year RFS of 0.77, 0.76, and 0.76, respectively (Supplemental Table 7). As some patients underwent hematopoietic stem cell transplantation (HSCT) during their first CR periods, which might confound the survival analysis, separate KM RFS curves were generated for patients who did and did not

undergo this procedure. These curves illustrated that the model remains valid irrespective of HSCT performance (Figure 3A, B).

The potential of the LSC6[11], LSC17[12], and LSC47[13] mRNA signatures and the lncScore[14] for pedAML risk stratification has been demonstrated. As these signatures were generated for the prediction of OS and EFS, we tested their predictive value for RFS in the discovery cohort. Our analyses confirmed this value. KM plots showed that all three models successfully stratified patients based on RFS. However, acceptable AUCs (~0.65) for 1-, 2-, and 3-year RFS prediction were obtained only for the LSC47 signature and lncScore, and these values were significantly lower than those obtained for the AML^{lnc69} (Supplemental Figure 4).

Next, AML^{lnc69} was validated with an independent validation cohort (*n* = 158 patients with pedAML; Supplemental Figure 1). Patients in the validation cohort were categorized into low- and high-risk groups based on the AML^{lnc69}, yielding AUCs of 0.66, 0.68, and 0.70 for 1-, 2-, and 3-year RFS prediction, respectively (Supplemental Figure 5A, B). As the benefit of adding gemtuzumab ozogamicin was a randomized question in this study cohort, we evaluated if AML^{lnc69} was predictive for RFS on each of the randomization arms. Interestingly, KM survival analysis illustrated that AML^{lnc69} remained predictive, irrespective of gemtuzumab ozogamicin randomization arm (Supplemental Figure 5C, D).

Subsequently, we plotted the distribution of cytogenetic/molecular risk for the AML^{lnc69} low- and high-risk groups in the discovery cohort. Importantly, a significant proportion of conventionally stratified standard-risk patients could be reclassified as high-risk according to the AML^{lnc69} (Figure 3C). Furthermore, KM plots showed that the AML^{lnc69} could be used to reclassify patients in all (low, standard, and high) traditional risk groups based on

cytogenetic/molecular characteristics (Figure 3D–F, Supplemental Figure 6). The frequency distributions of patient characteristics according to HSCT administration and AML^{lnc69} risk group are shown in Supplemental Table 8.

Altogether, these results illustrated that AML^{lnc69} use could improve traditional cytogenetic/molecular risk stratification, better defining low-risk and high-risk patients in terms of RFS.

AML^{lnc69} is an independent prognostic factor

To evaluate whether the AML^{lnc69} is an independent prognostic factor, we first performed Cox regression analysis and calculated c-indices for the discovery cohort. Due to the requirement for complete data, 704 patients were included in this analysis. Univariate Cox regression analysis indicated that age, white blood cell (WBC) count at the time of diagnosis, presence of core binding factor (CBF), KMT2A, t(6;11)(q27;q23), t(9;11)(p22;q23), t(10;11)(p11.2;q23), trisomy 21, nucleophosmin (NPM) mutation, MRD at the end of induction course 1, MRD at end of induction course 2, and the AML^{lnc69} signature were prognostic factors (Figure 4A, Supplemental Table 9). In this analysis, AML^{lnc69} performed well in distinguishing low- and high-risk individuals in all subgroups (defined according to genetics and MRD; Supplemental Figure 7), although some of these subgroups were small ($n < 5$) and these results should be interpreted with caution. Next, significant variables were examined further in a multivariate Cox analysis, which showed that the AML^{lnc69} signature, NPM mutation, and WBC count at the time of diagnosis were independent prognostic factors (Figure 4B). C-indices indicated that AML^{lnc69} had better RFS-predictive value than did NPM mutation and WBC count at the time of diagnosis (Supplemental Figure 8). Furthermore, AML^{lnc69} was an independent prognostic factor in the validation cohort, as demonstrated by univariate [hazard ratio (HR), 3.89; 95% confidence interval (CI), 2.04–7.44; $p < 0.001$] and

multivariate (HR, 3.56; 95% CI, 1.85–6.84; $p < 0.001$) Cox regression analyses (Supplemental Table 10).

Based on the multivariate Cox regression analysis results, a nomogram was established using the AML^{lnc69}, NPM mutation, and WBC count at the time of diagnosis (Figure 4C). Decision curve analysis indicated that the RFS-predictive effect of the three factors combined was superior to those of the individual factors (Supplemental Figure 9A). AUCs for the prediction of 1-, 2-, and 3-year RFS were 0.76, 0.75, and 0.75, respectively, and the c-index was 0.80 (Supplemental Figure 9B). Internal validation yielded a c-index of 0.71 and AUCs for 1-, 2-, and 3-year RFS prediction of 0.76, 0.75, and 0.75, respectively. Calibration curves showed good consistency of these three nomogram predictions with actual observations (Figure 4D). Table 1 provides an overview of all combinations possible and 1-, 2, and 3-year cumulative pedAML relapse risk prediction according to the nomogram. As an example, a patient with pedAML in the AML^{lnc69} high-risk group (100 points) with no NPM mutation (60 points) and a WBC count < 50 at the time of diagnosis (0 point) would have a 76.8% cumulative risk of relapse at 3 years.

Comparison of characteristics and biological pathway enrichment

Differences in clinical characteristics between the low- and high-risk AML^{lnc69} groups were examined (Supplemental Table 11). Relative to the AML^{lnc69} low-risk group, the AML^{lnc69} high-risk group contained significantly more cases with t(6;11)(q27;q23), t(9;11)(p22;q23), t(10;11)(p11.2;q23), t(11;19)(q23;p13.1), trisomy 21, KMT2A, MRD at the end of courses 1 and 2, and HSCT, as well as greater cytogenetic complexity, a younger age at diagnosis, a higher WBC count at the time of diagnosis, and a higher percentage of leukemic bone-marrow blasts (Figure 5).

To explore the biological pathways associated with the AML^{lnc69} signature, we performed GSEA using the cancer GO, KEGG, and hallmark gene sets. The five most significant results are shown in Supplemental Figure 10, and the clustered pathways are provided in Supplemental Tables 12–14 encompasses all the clustered pathways. The GSEA analysis revealed the significant involvement of the Hedgehog and KRAS pathways and epithelial–mesenchymal transition (EMT) in AML^{lnc69}-based high risk.

Discussion

The survival of patients with pedAML has increased steadily in recent decades, due in large part to the incorporation of risk stratification based on parameters such as cytogenetic and molecular abnormalities and MRD. Currently, pedAML relapse occurs in approximately one-third of patients, which is a major obstacle in treatment and adversely impacts OS[9,10]. AML relapse is attributable primarily to the poor responsiveness of therapy-resistant leukemic stem cells (LSCs) to common chemotherapeutic agents[34]. Treatments for pedAML relapse include HSCT and reinduction regimens, but their inevitable generation of side effects remains a challenge[9].

No robust prognostic model for the prediction of pedAML relapse is currently available. Several models based on EFS have been developed for pedAML risk stratification. However, due to the inclusion of relapse, induction failure, and death at first event, the LSC6, LSC17, and LSC47 signature and lncScore predicted RFS in the discovery cohort inefficiently. Moreover, we argue that induction failure should be included in analyses as a binary variable, regardless of its temporal span. For these reasons, we excluded induction failure and death, and focused solely on RFS.

In this study, we demonstrated that the AML^{lnc69} prognostic signature strengthens risk prediction, even in multivariate analysis including established risk-stratifying factors. Among the lncRNAs constituting the signature, MIR100HG (ENSG00000255015)[35], MIR17HG (ENSG00000215417)[36], MALAT1 (ENSG00000251562)[37-39], and ZFAS1 (ENSG00000177410)[40-43] were previously shown to be associated with AML; the associations of the other 65 lncRNAs with AML are novel. In contrast to the use of median values as dichotomization standards in most studies, we used an outcome-oriented method to identify optimal cutoff values in this study, thereby maximizing the distinction between low- and high-risk groups in terms of RFS. The good RFS-predicting performance of AML^{lnc69} was demonstrated through KM survival plots and ROC curves. Internal validation is crucial for the estimation of a model's generalizability[44]. Compared with other internal validation methods, bootstrap analysis not only enables the use of the entire sample for validation, but also provides nearly unbiased estimates of model performance[45]. The favorable results of bootstrap-based internal validation in this study provide strong evidence for the reliability of model construction. As some high-risk patients underwent HSCT to improve RFS[46], we split patients in the discovery cohort into HSCT and no-HSCT groups. The model retained its predictive value in both scenarios. Moreover, external validation must be performed to determine a model's reproducibility and generalizability to other samples[47]. Thus, data from the AAML0531 study, distinct from the discovery cohort, was employed for external validation in this study. Although this cohort was small and significantly more heterogeneous than the discovery cohort, the use of the AML^{lnc69} to predict RFS in this cohort was successful. The AUCs obtained in this study were satisfactory, indicating the reproducibility of AML^{lnc69} use in actual practice. The AML^{lnc69} remained predictive independently of gemtuzumab ozogamicin administration in the AAML0531 cohort; no such analysis could be

performed for sorafenib or bortezomib administration in the AAML1031 cohort due to the lack of information. With an ever-growing number of pedAML therapeutics available (including bcl-2 protein family and DNA methyltransferase inhibitors), the evaluation of the predictive value of AML^{lnc69} in this evolving therapeutic landscape would be of interest.

Recently, Farrar et al[14] constructed a lncRNA signature (lncScore) for pedAML with a completely different set of lncRNAs than used in the present study, which might be explained by some notable differences between the studies. First, lncScore was built based on EFS, while AML^{lnc69} was developed using RFS, which thus limits the direct comparison of the performance of both signatures. Second, Farrar et al[14] initially sought to identify lncRNAs that were differentially expressed (DE) in the bone marrow of patients with pedAML and that of healthy individuals to construct a model based on the upregulation of these lncRNAs in pedAML. However, the healthy individuals were actually post-induction patients, whose bone marrow may differ substantially from normal[48]. Third, the selection of DE lncRNAs may have led to the overlooking of some lncRNAs that actually impact prognosis. Thus, instead of performing differential expression analysis, we directly employed the entire lncRNAome for model construction. In addition, Farrar et al[14] constructed a regression model based on EFS, whereas we focused on RFS, and they included samples with low-depth sequencing, whereas we excluded such samples due to the usually low expression of lncRNAs. Despite these differences in composition, construction criteria, and endpoints, however, the two signatures are associated with very similar pathways in GSEAs (data not shown).

Surprisingly, the standard-risk group (as defined in the TARGET cohort based on cytogenic and molecular characteristics) had worse RFS than did the high-risk group. Aplenc et al.[26]

reported that approximately 78% of patients in the AAML1031 study were allocated to the low-risk group, with the remainder allocated to the high-risk group. However, a large number of these patients were reassigned to the standard-risk group in the TARGET database with the application of additional cytogenetic and molecular risk stratification criteria. In addition, 'AML^{lnc69}-Low_Risk group-High' patients, the majority of whom underwent HSCT, had better RFS than did 'AML^{lnc69}-Low_Risk group-Low/Standard' patients in this study. A portion of patients that were initially classified as 'Risk group-Low/Standard', missing HSCT, would have been reclassified as high risk, potentially explaining the superior RFS in 'AML^{lnc69}-Low_Risk group-High' patients relative to that of patients in the other two groups. Due to the retrospective nature of this study and the complexity of evolving risk stratification, however, complete clarification of these observed differences is difficult. Nevertheless, the results clearly suggest that all stratification schemes, including that used in the AAML1031 study and traditional cytogenetic and molecular risk stratification, have limitations resulting in the under- or overestimation of the pedAML relapse risk. Importantly, the AML^{lnc69} successfully stratified patients in the three TARGET risk groups (low, standard, and high) based on their RFS.

The AML^{lnc69} was shown to be an independent predictor of RFS in the discovery and validation cohorts. With the continued undertaking of large -omics studies and identification of novel risk-stratifying parameters, such as UBTF-TD[49] and GLIS2-fusions[50,51], the assessment of the predictive value of AML^{lnc69} in prospective clinical studies[52,53] documenting those molecular aberrations would be of great interest. Based on the results of the multivariate Cox regression analysis of discovery cohort data, we developed a nomogram for the intuitive prediction of the RFS of patients with pedAML that includes the AML^{lnc69}, NPM mutation, and WBC count at the time of diagnosis. Relative to the isolated application

of each factor, the nomogram provides for more-refined estimation of 1-, 2-, and 3-year cumulative relapse risks. With further independent validation, this tool could offer valuable insights for prognostic assessment and therapeutic decision making in the clinical context.

The GSEA performed in this study revealed that Hedgehog- and KRAS-associated pathways and EMT played important roles in the AML^{lnc69}-based estimation of high relapse risk. The Hedgehog signaling pathway plays a fundamental role in LSC quiescence and may be an effective target for the prevention of pedAML relapse[54,55]. Similarly, KRAS contributes to the emergence of stemness traits[56], and KRAS mutations are frequent in patients with pedAML and associated with worse outcomes[57-59]. EMT is a well-known cellular program that is crucial for the relapse and metastasis of various tumors, and it plays a key role in the progression and relapse of AML[60-62].

In conclusion, we generated a comprehensive prognostic model including 69 lncRNAs for the prediction of the RFS of patients with pedAML. To our knowledge, this study was the first comprehensive evaluation of relationships between lncRNAs and RFS in this population. We provide evidence that our model could further refine current risk stratification. Its application requires the design of a microarray or targeted RNA sequencing panel, which is currently the standard of practice in many clinical laboratories handling oncological diagnosis[63]. Although the AML^{lnc69} is associated with several known predictive markers, it incorporates more information than provided by these individual factors. Thus, AML^{lnc69} use may avoid the need to perform numerous molecular and cellular assays, as is currently done for the full risk classification of patients with pedAML, and thereby be a great asset in resource-limited circumstances. Furthermore, samples from patients' bone marrow or peripheral blood, routinely collected during standard pedAML diagnosis, can be used. This lack of need for

additional sample collection makes this method straightforward, cost effective, and easily implementable. Further validation of the AML^{lnc69} with large independent cohorts is needed to definitively confirm its clinical value.

Contributions

Z.R, J.V, and T.L drafted the manuscript. Z.R, J.V and T.L designed the figures. All authors critically revised the manuscript and approved the final version.

None of the authors has a relevant conflict of interest.

References

1. Zwaan CM, Kolb EA, Reinhardt D, et al. Collaborative Efforts Driving Progress in Pediatric Acute Myeloid Leukemia. *J Clin Oncol*. 2015;33(27):2949-2962.
2. Elgarten CW, Aplenc R. Pediatric acute myeloid leukemia: updates on biology, risk stratification, and therapy. *Current Opinion in Pediatrics*. 2020;32(1):57-66.
3. Shiba N. Comprehensive molecular understanding of pediatric acute myeloid leukemia. *Int J Hematol*. 2023;117(2):173-181.
4. de Rooij JD, Zwaan CM, van den Heuvel-Eibrink M. Pediatric AML: From Biology to Clinical Management. *J Clin Med*. 2015;4(1):127-149.
5. Creutzig U, Zimmermann M, Reinhardt D, et al. Changes in cytogenetics and molecular genetics in acute myeloid leukemia from childhood to adult age groups. *Cancer*. 2016;122(24):3821-3830.
6. Loken MR, Alonzo TA, Pardo L, et al. Residual disease detected by multidimensional flow cytometry signifies high relapse risk in patients with de novo acute myeloid leukemia: a report from Children's Oncology Group. *Blood*. 2012;120(8):1581-1588.
7. Schuurhuis GJ, Heuser M, Freeman S, et al. Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. *Blood*. 2018;131(12):1275-1291.
8. Tierens A, Bjorklund E, Siitonen S, et al. Residual disease detected by flow cytometry is an independent predictor of survival in childhood acute myeloid leukaemia; results of the NOPHO-AML 2004 study. *Br J Haematol*. 2016;174(4):600-609.
9. Egan G, Tasian SK. Relapsed pediatric acute myeloid leukaemia: state-of-the-art in 2023. *Haematologica*. 2023;108(9):2275-2288.
10. Hoffman AE, Schoonmade LJ, Kaspers GJ. Pediatric relapsed acute myeloid leukemia: a systematic review. *Expert Rev Anticancer Ther*. 2021;21(1):45-52.
11. Elsayed AH, Rafiee R, Cao X, et al. A six-gene leukemic stem cell score identifies high risk pediatric acute myeloid leukemia. *Leukemia*. 2020;34(3):735-745.
12. Duployez N, Marceau-Renaut A, Villenet C, et al. The stem cell-associated gene expression signature allows risk stratification in pediatric acute myeloid leukemia. *Leukemia*. 2019;33(2):348-357.
13. Huang BJ, Smith JL, Farrar JE, et al. Integrated stem cell signature and cytomolecular risk determination in pediatric acute myeloid leukemia. *Nat Commun*. 2022;13(1):5487.
14. Farrar JE, Smith JL, Othus M, et al. Long Noncoding RNA Expression Independently Predicts Outcome in Pediatric Acute Myeloid Leukemia. *J Clin Oncol*. 2023:JCO2201114.
15. Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet*. 2014;15(1):7-21.
16. Esteller M. Non-coding RNAs in human disease. *Nature Reviews Genetics*. 2011;12(12):861-874.
17. Batista PJ, Chang HY. Long Noncoding RNAs: Cellular Address Codes in Development and Disease. *Cell*. 2013;152(6):1298-1307.
18. Neyazi S, Ng M, Heckl D, Klusmann JH. Long noncoding RNAs as regulators of pediatric acute myeloid leukemia. *Mol Cell Pediatr*. 2022;9(1):10.
19. Delas MJ, Sabin LR, Dolzhenko E, et al. lncRNA requirements for mouse acute myeloid leukemia and normal differentiation. *Elife*. 2017;6.
20. Nobili L, Lionetti M, Neri A. Long non-coding RNAs in normal and malignant hematopoiesis. *Oncotarget*. 2016;7(31):50666-50681.
21. Garzon R, Volinia S, Papaioannou D, et al. Expression and prognostic impact of lncRNAs in acute myeloid leukemia. *Proceedings of the National Academy of Sciences of the United States of America*. 2014;111(52):18679-18684.

22. Liu DX. Identification of a prognostic LncRNA signature for ER-positive, ER-negative and triple-negative breast cancers. *Breast Cancer Research and Treatment*. 2020;183(1):95-105.
23. Sathipati SY, Sahu D, Huang HC, Lin YC, Ho SY. Identification and characterization of the lncRNA signature associated with overall survival in patients with neuroblastoma. *Scientific Reports*. 2019;9.
24. Gamis AS, Alonzo TA, Meshinchi S, et al. Gemtuzumab ozogamicin in children and adolescents with de novo acute myeloid leukemia improves event-free survival by reducing relapse risk: results from the randomized phase III Children's Oncology Group trial AAML0531. *J Clin Oncol*. 2014;32(27):3021-3032.
25. Pollard JA, Alonzo TA, Gerbing R, et al. Sorafenib in Combination With Standard Chemotherapy for Children With High Allelic Ratio FLT3/ITD+ Acute Myeloid Leukemia: A Report From the Children's Oncology Group Protocol AAML1031. *J Clin Oncol*. 2022;40(18):2023-2035.
26. Aplenc R, Meshinchi S, Sung L, et al. Bortezomib with standard chemotherapy for children with acute myeloid leukemia does not improve treatment outcomes: a report from the Children's Oncology Group. *Haematologica*. 2020;105(7):1879-1886.
27. Cunningham F, Allen JE, Allen J, et al. Ensembl 2022. *Nucleic Acids Res*. 2022;50(D1):D988-D995.
28. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361-387.
29. Efron B. 1977 RIETZ LECTURE - BOOTSTRAP METHODS - ANOTHER LOOK AT THE JACKKNIFE. *Annals of Statistics*. 1979;7(1):1-26.
30. Steyerberg EW, Harrell FE, Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 2016;69:245-247.
31. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol*. 2015;16(4):e173-180.
32. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550.
33. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739-1740.
34. Stelmach P, Trumpp A. Leukemic stem cells and therapy resistance in acute myeloid leukemia. *Haematologica*. 2023;108(2):353-366.
35. Emmrich S, Streltsov A, Schmidt F, Thangapandi VR, Reinhardt D, Klusmann JH. LincRNAs MONC and MIR100HG act as oncogenes in acute megakaryoblastic leukemia. *Mol Cancer*. 2014;13:171.
36. Yan J, Yao L, Li P, Wu G, Lv X. Long non-coding RNA MIR17HG sponges microRNA-21 to upregulate PTEN and regulate homoharringtonine-based chemoresistance of acute myeloid leukemia cells. *Oncol Lett*. 2022;23(1):24.
37. Jin J, Fu L, Hong P, Feng W. MALAT-1 regulates the AML progression by promoting the m6A modification of ZEB1. *Acta Biochim Pol*. 2023;70(1):37-43.
38. Sheng XF, Hong LL, Li H, Huang FY, Wen Q, Zhuang HF. Long non-coding RNA MALAT1 modulate cell migration, proliferation and apoptosis by sponging microRNA-146a to regulate CXCR4 expression in acute myeloid leukemia. *Hematology*. 2021;26(1):43-52.
39. Hu N, Chen L, Wang C, Zhao H. MALAT1 knockdown inhibits proliferation and enhances cytarabine chemosensitivity by upregulating miR-96 in acute myeloid leukemia cells. *Biomed Pharmacother*. 2019;112:108720.

40. Su L, Kong H, Wu F, et al. Long non-coding RNA zinc finger antisense 1 functions as an oncogene in acute promyelocytic leukemia cells. *Oncol Lett*. 2019;18(6):6331-6338.
41. Guo H, Wu L, Zhao P, Feng A. Overexpression of long non-coding RNA zinc finger antisense 1 in acute myeloid leukemia cell lines influences cell growth and apoptosis. *Exp Ther Med*. 2017;14(1):647-651.
42. Li Q, Wang J. Long noncoding RNA ZFAS1 enhances adriamycin resistance in pediatric acute myeloid leukemia through the miR-195/Myb axis. *RSC Adv*. 2019;9(48):28126-28134.
43. Gan S, Ma P, Ma J, et al. Knockdown of ZFAS1 suppresses the progression of acute myeloid leukemia by regulating microRNA-150/Sp1 and microRNA-150/Myb pathways. *Eur J Pharmacol*. 2019;844:38-48.
44. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating.: New York: Springer; 2019.
45. Harrell FE, Jr. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis, 2nd Edition. : New York: Springer; 2015.
46. Cooper TM, Ries RE, Alonzo TA, et al. Revised Risk Stratification Criteria for Children with Newly Diagnosed Acute Myeloid Leukemia: A Report from the Children's Oncology Group. *Blood*. 2017;130.
47. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J*. 2021;14(1):49-58.
48. dbGap. TARGET: Acute Myeloid Leukemia (AML). dbGaP Study Accession: phs000465v19p8. https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000465.v19.p8
49. Umeda M, Ma J, Huang BJ, et al. Integrated Genomic Analysis Identifies UBTF Tandem Duplications as a Recurrent Lesion in Pediatric Acute Myeloid Leukemia. *Blood Cancer Discov*. 2022;3(3):194-207.
50. Conneely SE, Stevens AM. Acute Myeloid Leukemia in Children: Emerging Paradigms in Genetics and New Approaches to Therapy. *Curr Oncol Rep*. 2021;23(2):16.
51. Aid Z, Robert E, Lopez CK, et al. High caspase 3 and vulnerability to dual BCL2 family inhibition define ETO2::GLIS2 pediatric leukemia. *Leukemia*. 2023;37(3):571-579.
52. Karlsson L, Cheuk D, De Moerloose B, et al. Characteristics and outcome of primary resistant disease in paediatric acute myeloid leukaemia. *Br J Haematol*. 2023;201(4):757-765.
53. Kaspers GJL, Wijnen NE, Koedijk JB, et al. Chip-AML22 Master Protocol: An Open-Label Clinical Trial in Newly Diagnosed Pediatric De Novo Acute Myeloid Leukemia (AML) Patients Including a Linked Phase II Trial with Quizartinib in FLT3-ITD/ NPM1wt Patients - a Study By the NOPHO-DB-SHIP Consortium. *Blood*. 2023;142(Supplement 1):1532-1532.
54. Pession A, Lonetti A, Bertuccio S, Locatelli F, Masetti R. Targeting Hedgehog pathway in pediatric acute myeloid leukemia: challenges and opportunities. *Expert Opin Ther Targets*. 2019;23(2):87-91.
55. Lainez-Gonzalez D, Serrano-Lopez J, Alonso-Dominguez JM. Understanding the Hedgehog Signaling Pathway in Acute Myeloid Leukemia Stem Cells: A Necessary Step toward a Cure. *Biology (Basel)*. 2021;10(4).
56. Chippalkatti R, Abankwa D. Promotion of cancer cell stemness by Ras. *Biochem Soc Trans*. 2021;49(1):467-476.
57. Bolouri H, Farrar JE, Triche T, Jr., et al. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat Med*. 2018;24(1):103-112.

58. Mustafa Ali MK, Williams MT, Corley EM, AlKaabba F, Niyongere S. Impact of KRAS and NRAS mutations on outcomes in acute myeloid leukemia. *Leuk Lymphoma*. 2023;64(5):962-971.
59. Ball BJ, Hsu M, Devlin SM, et al. The prognosis and durable clearance of RAS mutations in patients with acute myeloid leukemia receiving induction chemotherapy. *Am J Hematol*. 2021;96(5):E171-E175.
60. Dongre A, Weinberg RA. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat Rev Mol Cell Biol*. 2019;20(2):69-84.
61. Chen SC, Liao TT, Yang MH. Emerging roles of epithelial-mesenchymal transition in hematological malignancies. *J Biomed Sci*. 2018;25(1):37.
62. Carmichael CL, Wang J, Nguyen T, et al. The EMT modulator SNAI1 contributes to AML pathogenesis via its interaction with LSD1. *Blood*. 2020;136(8):957-973.
63. Tiwari M. Microarrays and cancer diagnosis. *J Cancer Res Ther*. 2012;8(1):3-10.

Figure legends

Figure 1. **Identification of the 69 relapse-related lncRNAs prognostic signature for pedAML patients.** (A) Coefficients of the 69 lncRNAs originated from the LASSO Cox regression. (B) Risk score distribution and RFS status map of pedAML patients. LncRNA, long non-coding RNA; LASSO, least absolute shrinkage and selection operator; RFS, relapse-free survival; pedAML, pediatric acute myeloid leukemia.

Figure 2. **Validation of AML^{lnc69}.** (A, B) Kaplan–Meier survival curve of pedAML patients' RFS (A) and OS (B) in the low- and high-risk groups. (C) ROC curves and AUCs for 1-, 2-, and 3-year RFS. PedAML, pediatric acute myeloid leukemia; RFS, relapse-free survival; OS, overall survival; ROC, receiver operating characteristic; AUC, area under the curve.

Figure 3. **The performance of AML^{lnc69} in terms of whether or not undergoing HSCT in 1st CR and cytogenetic/molecular risk.** (A, B) Kaplan–Meier survival curves of RFS for the pedAML patients receiving HSCT (A) and without receiving HSCT in the first CR (B). (C) The distribution of cytogenetic/molecular risk between low- and high-risk groups based on AML^{lnc69}. (D-F) Kaplan–Meier survival curve of pedAML patients' RFS based on AML^{lnc69} in low-, standard- and high-risk groups categorized by cytogenetic/molecular risk stratification. HSCT, hematopoietic stem cell transplantation; CR, complete remission; pedAML, pediatric acute myeloid leukemia; RFS, relapse-free survival.

Figure 4. **Independent prognostic analysis of AML^{lnc69}.** (A, B) Forest plots of univariate (A) and multivariate (B) independent Cox regression analyses of AML^{lnc46} and other characters. (C) Nomogram model of the combined AML^{lnc46}, NPM mutation and WBC at diagnosis for 1-, 2-, and 3-year relapse risk in pedAML patients. (D) Calibration plot comparing nomogram-predicted and actual RFS at 1-, 2-, and 3-year. PedAML, pediatric acute myeloid leukemia; RFS, relapse-free survival.

Figure 5. **Comparision of significant characters between AML^{lnc69} low- and high-risk groups.** (A) Heatmap comparing the distribution of significant characters. (B-D) Comparision of age (B), WBC at diagnosis (C) and bone marrow leukemic blast (D). WBC, white blood cell.

Tables

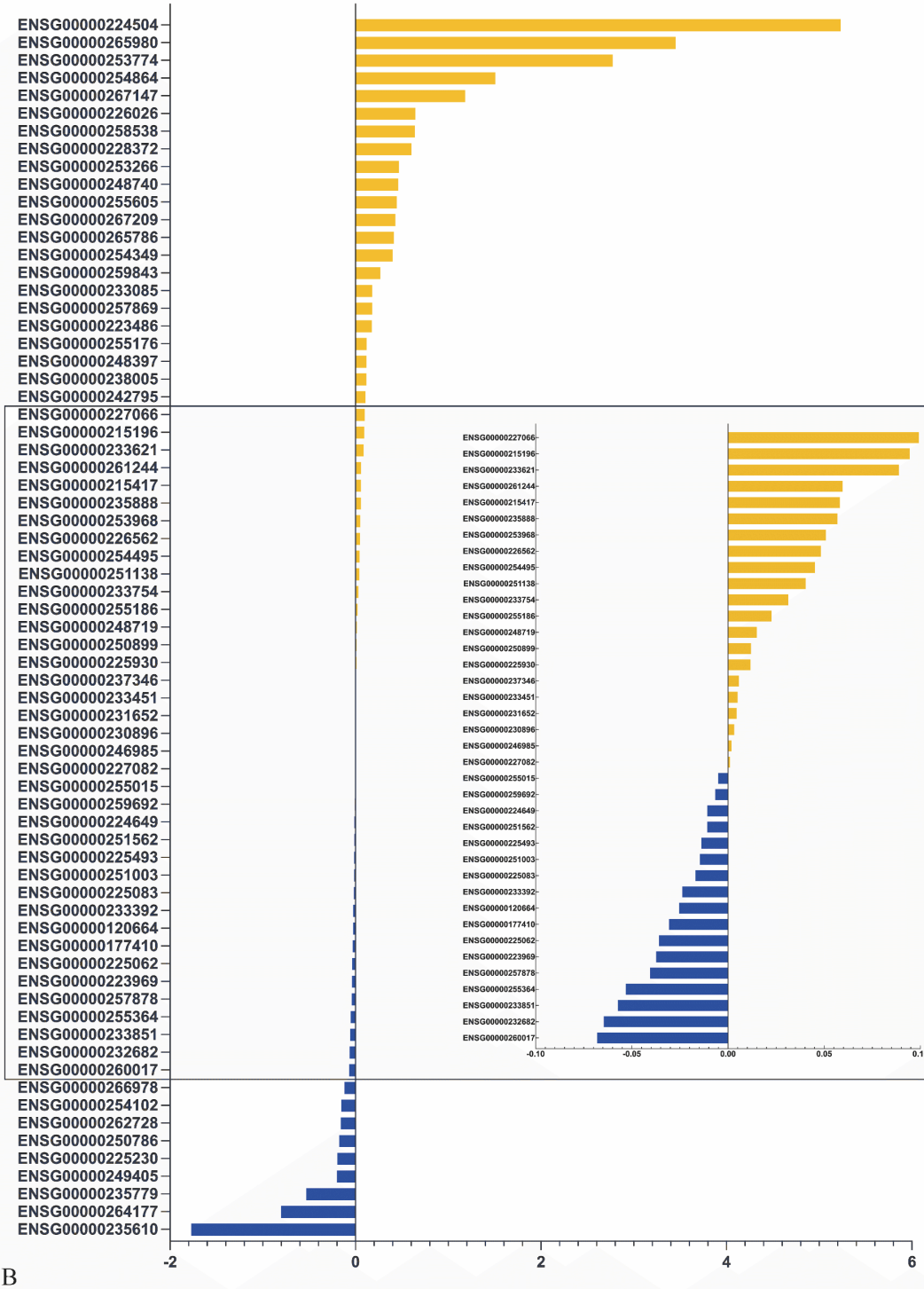
TABLE 1 **Examples of 1-, 2-, and 3-year relapse risk prediction for pedAML patients using the nomogram prediction model.**

AML^{lnc69}	NPM mutation	WBC at diagnosis	1-Year Risk, % (95% CI)	2-Year Risk, % (95% CI)	3-Year Risk, % (95% CI)
High	No	≥50	56.8 (48.3-63.8)	80.5 (72.9-86.0)	83.6 (76.4-88.7)
High	Yes	≥50	29.8 (14.2-42.6)	49.9 (26.1-66.0)	53.4 (28.5-69.7)
High	No	<50	49.2 (42.1-55.5)	73.3 (66.1-79.0)	76.8 (69.8-82.2)
High	Yes	<50	24.9 (11.7-36.2)	42.8 (21.7-58.2)	46.1 (23.8-61.9)
Low	No	≥50	18.1 (13.8-22.3)	32.3 (25.6-38.4)	35.1 (28.0-41.4)
Low	Yes	≥50	8.1 (3.7-12.4)	15.2 (7.2-22.5)	16.7 (7.9-24.6)
Low	No	<50	14.9 (11.9-17.9)	27.1 (22.5-31.4)	29.5 (24.7-34.0)
Low	Yes	<50	6.6 (3.0-10.0)	12.5 (6.0-18.5)	13.7 (6.6-20.3)

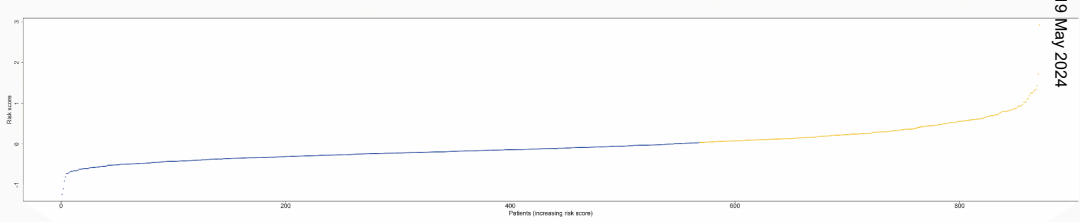
Abbreviations: WBC: white blood cell.

Figure 1

A



B



C

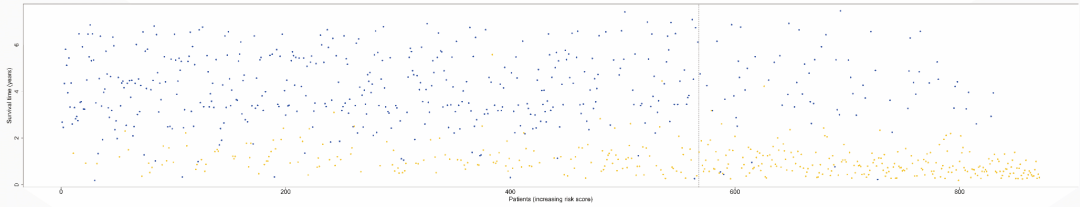


Figure 2

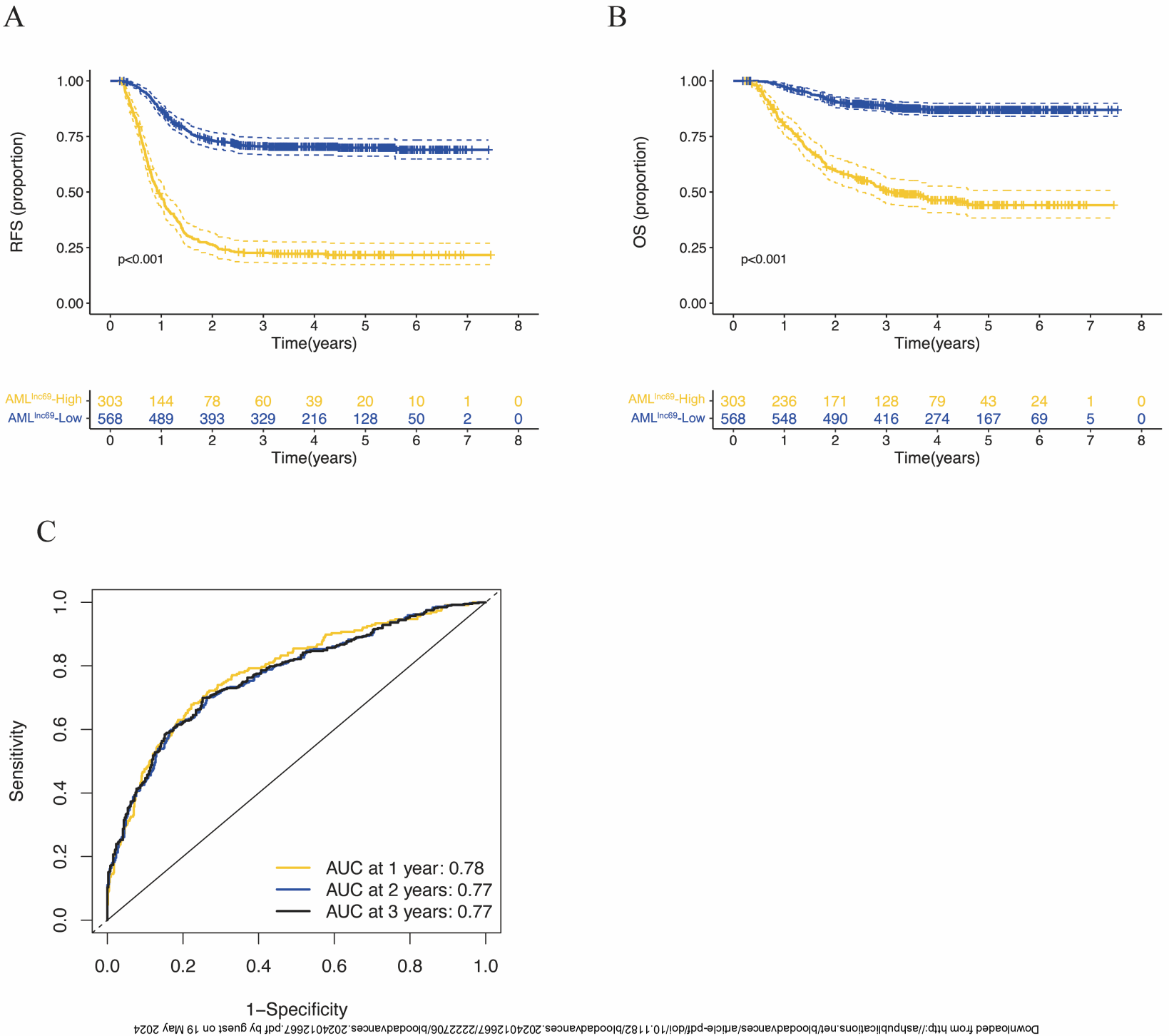


Figure 3

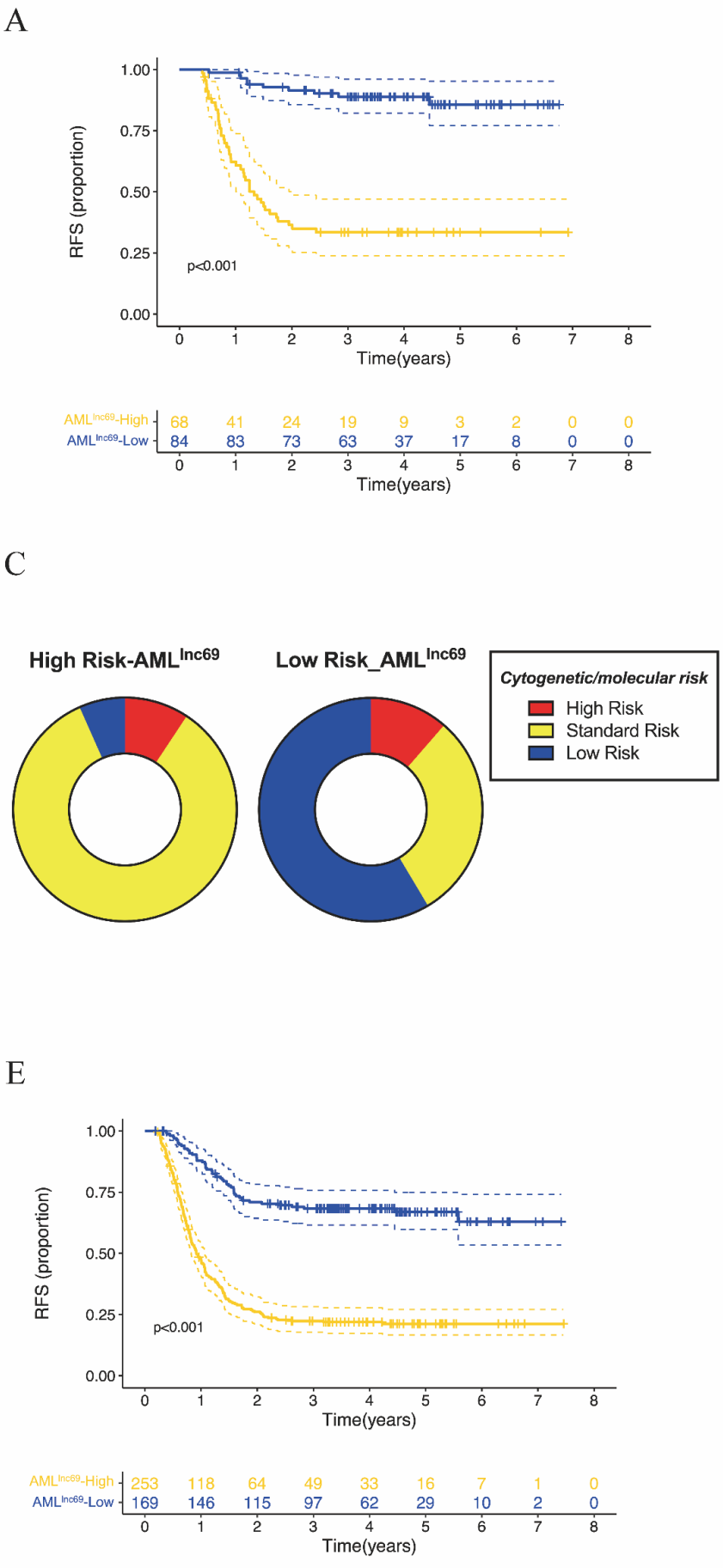


Figure 4

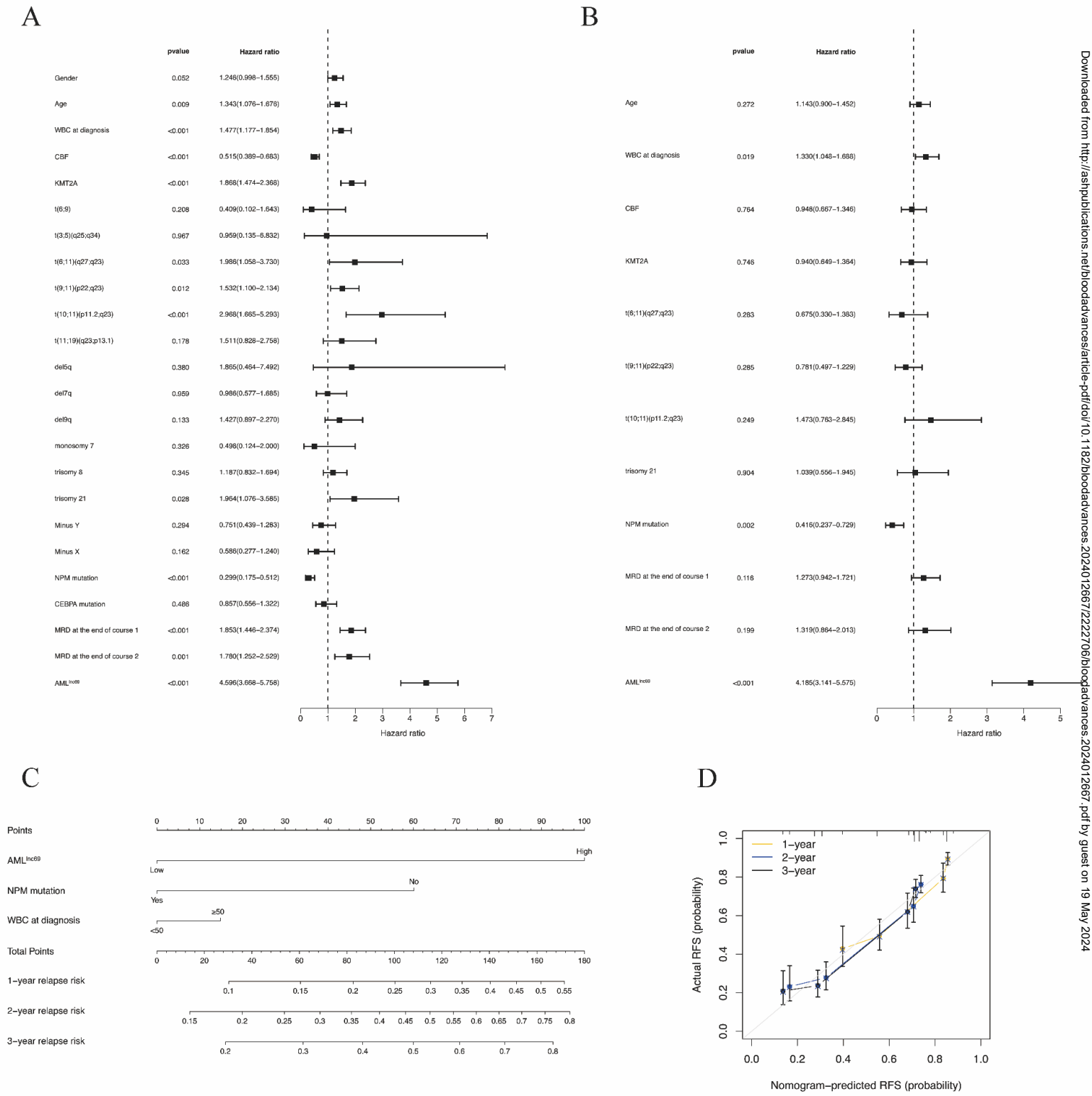
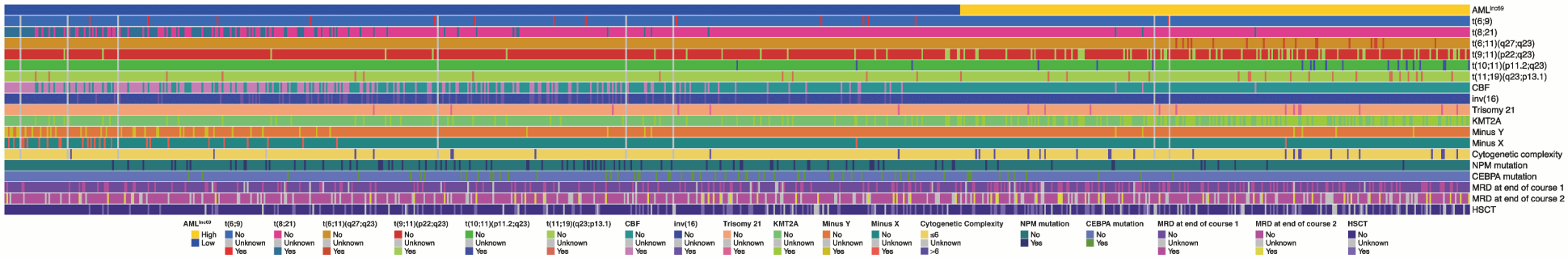
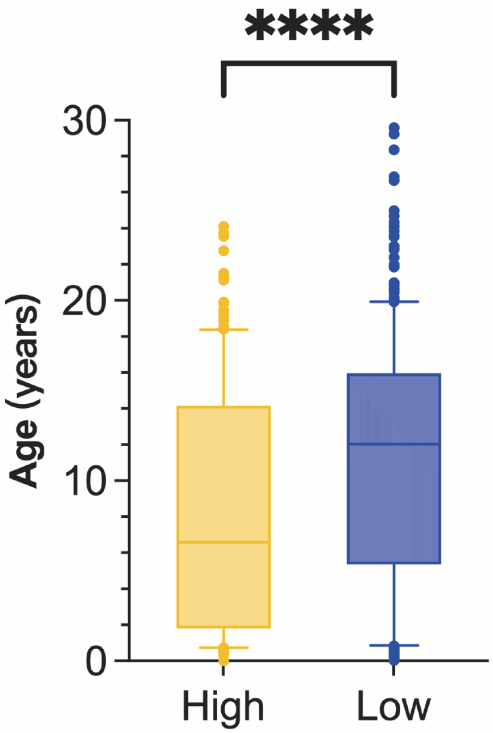


Figure 5

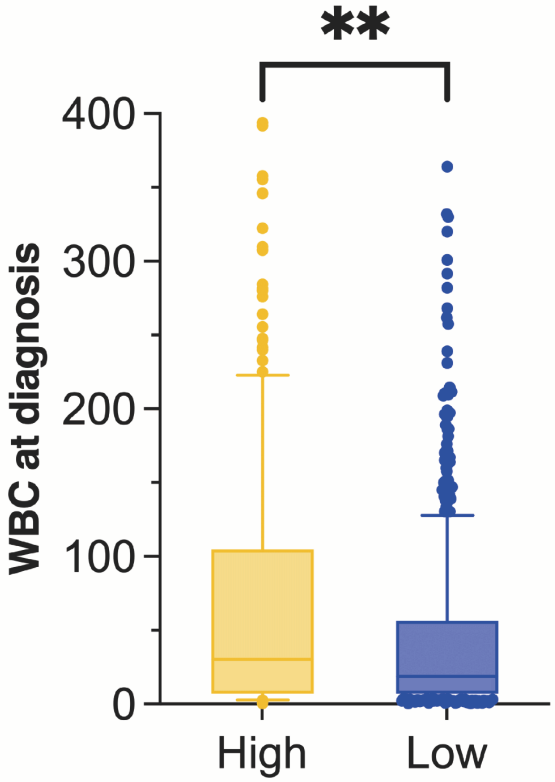
A



B



C



D