HemaSphere ❀ eha
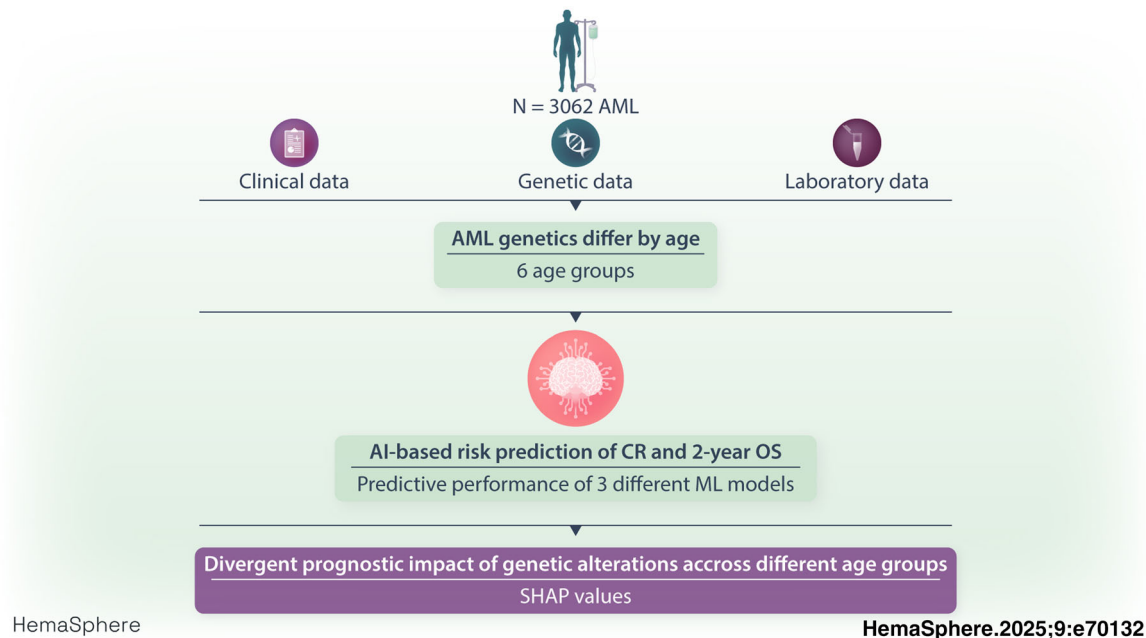
Age-stratified machine learning identifies divergent prognostic significance of molecular alterations in AML

Jan-Niklas Eckardt[1,2] | Waldemar Hahn[3,4] | Rhonda E. Ries[5,6] |
Szymon D. Chrost[1] | Susann Winter[1] | Sebastian Stasik[1] |
Christoph Röllig[1] | Uwe Platzbecker[7] | Carsten Müller-Tidow[8] |
Hubert Serve[9] | Claudia D. Baldus[10] | Christoph Schliemann[11] |
Kerstin Schäfer-Eckart[12] | Maher Hanoun[13] | Martin Kaufmann[14] |
Andreas Burchert[15] | Johannes Schetelig[1] | Martin Bornhäuser[1,16,17] |
Markus Wolfien[3,4] | Soheil Meshinchi[5,6,^] | Christian Thiede[1,^] |
Jan Moritz Middeke[1,2,^]

Graphical Abstract

N = 3062 AML

Clinical data | Genetic data | Laboratory data

AML genetics differ by age
6 age groups

AI-based risk prediction of CR and 2-year OS
Predictive performance of 3 different ML models

Divergent prognostic impact of genetic alterations accross different age groups
SHAP values

ARTICLE

HemaSphere ⬡ eha

Age-stratified machine learning identifies divergent prognostic significance of molecular alterations in AML

Jan-Niklas Eckardt[1,2] | Waldemar Hahn[3,4] | Rhonda E. Ries[5,6] |
Szymon D. Chrost[1] | Susann Winter[1] | Sebastian Stasik[1] |
Christoph Röllig[1] | Uwe Platzbecker[7] | Carsten Müller-Tidow[8] |
Hubert Serve[9] | Claudia D. Baldus[10] | Christoph Schliemann[11] |
Kerstin Schäfer-Eckart[12] | Maher Hanoun[13] | Martin Kaufmann[14] |
Andreas Burchert[15] | Johannes Schetelig[1] | Martin Bornhäuser[1,16,17] |
Markus Wolfien[3,4] | Soheil Meshinchi[5,6,^] | Christian Thiede[1,^] |
Jan Moritz Middeke[1,2,^]

Correspondence: Jan-Niklas Eckardt (jan-niklas.eckardt@uniklinikum-dresden.de)

Abstract

Risk stratification in acute myeloid leukemia (AML) is driven by genetics, yet patient age substantially influences therapeutic decisions. To evaluate how age alters the prognostic impact of genetic mutations, we pooled data from 3062 pediatric and adult AML patients from multiple cohorts. Signaling pathway mutations dominated in younger patients, while mutations in epigenetic regulators, spliceosome genes, and *TP53* alterations became more frequent with increasing age. Machine learning models were trained to identify prognostic variables and predict complete remission and 2-year overall survival, achieving area-under-the-curve scores of 0.801 and 0.791, respectively. Using Shapley (SHAP) values, we quantified the contribution of each variable to model decisions and traced their impact across six age groups: infants, children, adolescents/young adults, adults, seniors, and elderly. The highest contributions to model decisions among genetic variables were found for alterations of *NPM1*, *CEBPA*, inv (16), and t(8;21) conferring favorable risk and alterations of *TP53*, *RUNX1*, *ASXL1*, del(5q), -7, and -17 conferring adverse risk, while *FLT3*-ITD had an ambiguous role conferring favorable treatment responses yet poor overall survival. Age significantly modified the prognostic value of genetic alterations, with no single alteration consistently predicting outcomes across all age groups. Specific alterations associated with aging such as *TP53*, *ASXL1*, or del(5q) posed a disproportionately higher risk in younger patients. These results challenge uniform risk stratification models and highlight the need for context-sensitive AML treatment strategies.

[1]Department of Internal Medicine I, University Hospital Carl Gustav Carus, TUD Dresden University of Technology, Dresden, Germany

[2]Else Kröner Fresenius Center for Digital Health, TUD Dresden University of Technology, Dresden, Germany

[3]Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Dresden/Leipzig, Dresden, Germany

[4]Institute for Medical Informatics and Biometry, TUD Dresden University of Technology, Dresden, Germany

[5]Translational Science and Therapeutics Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

[6]Division of Hematology and Oncology, Seattle Children's Hospital, Seattle, Washington, USA

[7]Department of Hematology, Cellular Therapy, Hemostaseology and Infectious Disease, University of Leipzig Medical Center, Leipzig, Germany

[8]Department of Medicine V, University Hospital Heidelberg, Heidelberg, Germany

[9]Department of Medicine 2, Hematology and Oncology, Goethe University Frankfurt, Frankfurt, Germany

[10]Department of Hematology and Oncology, University Hospital Schleswig Holstein, Kiel, Germany

[11]Department of Medicine A, University Hospital Münster, Münster, Germany

[12]Department of Internal Medicine V, Paracelsus Medizinische Privatuniversität and University Hospital Nürnberg, Nürnberg, Germany

[13]Department of Hematology, University Hospital Essen, Essen, Germany

INTRODUCTION

Adjusted for age, acute myeloid leukemia (AML) affects 4.3 per 100,000 persons in the United States every year.[1] Despite recent advances in personalized and targeted therapies, patient outcomes still remain largely unsatisfactory.[2,3] Age plays a major role in both AML biology and patient management. The incidence of AML increases with age, posing a major challenge for healthcare systems in light of the uniformly aging population in industrialized countries.[4,5] The median age at initial diagnosis is estimated to be 68 years.[6] In patients over 65 years of age, incidence rates rise to 20.1 per 100,000 persons.[1] Meanwhile, age remains a major criterion in treatment decisions, as many intensive chemotherapy-based regimens or allogeneic stem cell transplantation (ASCT) are less tolerable for older patients due to frailty or comorbidities.[7] A better understanding of disease biology, based on numerous studies of the individual impact of cytogenetic and molecular genetic alterations on prognostication, has led to refined treatment recommendations. Given the advancements in molecular analysis, guidelines such as the European Leukemia Net (ELN) recommendations[8–10] have been continuously improved over the last decade. Nevertheless, following these guidelines often mirrors decision-making along the nodes of a binary decision tree. First, the absence or presence of a particular mutation is noted. Second, if the mutation is classified within the guidelines as, for example, favorable and other risk-defining mutations are absent, the patient is assigned to the favorable risk group. Third, this risk stratification is used to inform treatment decisions. While we do not aim to replace validated clinical decision-making models, we aim to demonstrate their limitations and advocate for more personalized models incorporating multidimensional patient variables in a data-driven manner.

Given the disparity between tolerability of intensive regimens, these treatment recommendations differ between pediatric and adult patients.[10–12] Nevertheless, the age-dependent impact of cytogenetic aberrations or molecular alterations is often neglected in risk stratification, despite evident discrepancies in the molecular landscape of pediatric and adult AML.[13,14] Studies in pediatric AML often fail to compare results to adult AML and vice versa. Studies in adult AML seldom stratify age groups beyond the common threshold of 65 years of age, where patients are deemed less eligible for intensive therapy.[15] As patient age modifies disease biology, we hypothesize that risk stratification based on cytogenetic and molecular alterations may be context-sensitive regarding patient age.

MATERIALS AND METHODS

Patient data

To identify the age-dependent impact of genetic alterations, we analyzed a total of 3062 patients with pediatric or adult AML from the German Study Alliance Leukemia (SAL), the Fred Hutchinson Cancer Research Center Seattle, as well as publicly available data from Bottomly et al.[16] and TARGET data originally published by the National Cancer Institute.

In the first cohort obtained from the German SAL, clinical, laboratory, and genetic data were obtained from 1606 adult patients that were treated within previously conducted multicentric prospective clinical trials (AML96 [NCT00180115],[17] AML2003 [NCT00180102],[18] AML60+ [NCT00180167],[19] and SORAML [NCT00893373][20]) registered in the SAL registry (NCT03188874). Table S1 shows an overview of trial protocols. In the original clinical trial protocols, only patients ≥18 years of age were eligible. Acute promyelocytic leukemia (APL, syn. AML-M3) patients were excluded from these trials. All patients gave their written informed consent according to the revised Declaration of Helsinki.[21] All studies were previously approved by the Institutional Review Board of the TUD Dresden University of Technology (EK 98032010). Complete remission (CR) and overall survival (OS) were defined according to the revised ELN criteria.[10]

In the second cohort obtained from the Fred Hutchinson Cancer Center Seattle, clinical, laboratory, and genetic data were available for 578 pediatric patients from the Children's Oncology Group (COG) protocol, AAML0531 (NCT01407757).[22] All subject samples were obtained by member COG institutions after written consent was obtained from the parents/guardians of minors upon enrolling in the trial. The genomic study was overseen by the Institutional Review Board at Fred Hutchinson Cancer Research Center (protocol 9365). Data on selected clinical (e.g., age, presenting hematological indices and cytogenetic classification) and molecular (e.g., KIT, RAS genes, NPM1, WT1, CEBPA mutations, and FLT3-ITD allelic ratios) features were clinically available before additional genomic analyses were performed.

Both cohorts were pooled with publicly available data from Bottomly et al.[16] (n = 728) and 150 patients from TARGET-AML,[23] both accessed via cbioportal.com in October 2023 and comprising clinical data, as well as whole genome or whole exome sequencing data. Notably, all pediatric patients and adult patients from the SAL cohort were treated with intensive chemotherapy, while 15.9% of patients within the cohort derived from Bottomly et al. were treated with non-intensive regimens. Baseline demographic and clinical characteristics of the cohorts are displayed in Table S2.

Molecular analyses

For the SAL cohort, biomaterial was obtained from bone marrow aspirates or peripheral blood before treatment initiation. Next-generation sequencing (NGS) was conducted using the TruSight Myeloid Sequencing Panel (Illumina, San Diego, CA, USA). A comprehensive list of covered alterations is provided in Table S3. Pooled samples were sequenced paired-end and a 5% variant allele frequency mutation calling cut-off was used with human genome build hg19 as a reference, as previously described in detail.[24] Additionally, high resolution fragment analysis for FLT3-ITD,[25] NPM1,[26] and CEBPA[27] was performed as described previously. For cytogenetics, standard techniques for chromosome banding and fluorescence in situ hybridization were used.

For the cohort from the Fred Hutchinson Cancer Center Seattle, targeted capture sequencing was performed for a panel of over 400 cancer-related genes. Clinical sequencing of prognostically relevant alterations such as FLT3-ITD, NPM1 4bp-insertion, and mutations in CEBPA was performed as previously described.[13]

As only NGS panel data were available for SAL patients, we focused on common genes listed in Table S3 and cytogenetics available for all cohorts. Finally, all patients were categorized into six age

[14]Department of Hematology, Oncology and Palliative Care, Robert Bosch Hospital, Stuttgart, Germany

[15]Department of Hematology, Oncology and Immunology, Philipps-University Marburg, Marburg, Germany

[16]German Cancer Consortium (DKTK), Partner Site Dresden, German Cancer Research Center (DKFZ), Heidelberg, Germany

[17]National Center for Tumor Diseases (NCT), NCT/UCC Dresden, a partnership between DKFZ, Faculty of Medicine and University Hospital Carl Gustav Carus, TUD Dresden University of Technology, and Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Germany

^These authors contributed equally to this work.

groups: infants (0–2 years), children (3–14 years), adolescents and young adults (AYA, 15–39 years), adults (40–64 years), seniors (65–74 years), and elderly (75+ years). Age ranges were guided by definitions of the American Society of Pediatrics[28] and the National Cancer Institute.[29] The conventional binarization of adult AML patients at 65 years was included, however, given the advances in patient management and supportive therapies, patients beyond 65 years of age may still be eligible for intensive therapies yielding improved outcomes.[30–32] To account for this paradigm shift, an additional binarization was introduced at 75 years. We acknowledge that further categories are conceivable, but additional subdivision leads to smaller group sizes, which destabilize the model training process since there are fewer subjects per group to train on.

Statistical analysis

All tests were carried out as two-sided tests. Statistical significance was determined using a significance level α of 0.05. Multiple hypothesis testing was adjusted for using the Benjamini–Hochberg method.[33] Univariate analyses for binary outcomes (CR rate) were carried out via LR to obtain odds ratios (OR) and 95% confidence intervals (95% CI). Time-to-event analyses (OS) were carried out using Cox proportional hazard models to obtain hazard ratios (HR) and 95% CI. Statistical analysis was conducted using R version 4.2.3 and RStudio version 2023.12.0-369.

Explainable Machine Learning (ML)

Binary classifications were applied for the achievement of CR after intensive induction therapy, as well as for OS binarized at the 2-year mark (2-year OS). Patients with missing CR status were excluded for CR prediction ($n = 191$), and patients that were censored before 2 years of follow-up were excluded for 2-year OS prediction ($n = 204$). Missing values were imputed using a k-nearest neighbor (KNN) imputer with $k = 5$. The dataset was separated into 80% training and 20% test data, stratified by age groups and the outcome of interest. To reduce model dimensionality, which may otherwise destabilize the training process (see "curse of dimensionality"[34]), features with a frequency of <1% were excluded. ML models used include Random Forest (RF), Logistic Regression (LR), and XGBoost (XGB). LR is a linear model that predicts binary outcomes based on a weighted combination of input variables and assumes a linear relationship between predictors and the log-odds of the outcome. RF is an ensemble learning method that builds multiple decision trees during training and outputs the mode of classifications or the average of predictions, effectively capturing nonlinear relationships and interactions between variables. XGB is a gradient-boosted decision tree method that sequentially builds trees to minimize prediction errors, offering high flexibility in handling complex, high-dimensional data. Hyperparameter optimization was performed for each model using a 10-fold cross-validation on the training set. Model performance was measured using the area under the receiver-operating characteristic (AUROC), as well as the Matthew's correlation coefficient (MCC)[35] on the test set. For each non-deterministic model (RF and XGB), we report the average classification performance and the range of 100 runs to provide the most robust result instead of reporting highly-performant outlier runs that may occur due to random variation in the training process. Explainability (i.e., the causal effect of a feature's contribution to model inference) was quantified using Shapley (SHAP) values. These are derived from game theory, where they provide a measure of how much credit is allocated to each individual model feature at the input level, given a specific output for binary endpoints.

Therefore, they allow to measure and rank the importance of each feature, in our case genetic alterations, for model performance. We provide the SHAP values for the model with most average results (according to MCC) of 100 runs to ensure a balanced representation of feature importance. To understand how each feature influenced the ML model's decisions across different age groups, we first calculated the SHAP values for each feature within each age group. Next, we summed the SHAP values for each feature across all age groups. Finally, we expressed the contribution of each age group as a percentage of the total SHAP values for that feature.

RESULTS

Altered genetic pathways differ by age groups

Among all 3062 patients in our pooled cohort, infants constituted 4.1% ($n = 126$), children 13.2% ($n = 408$), AYA 19.2% ($n = 589$), adults 39.3% ($n = 1204$), seniors 16.9% ($n = 518$), and elderly 7.1% ($n = 217$). For a comprehensive overview of the genetic landscape per age group, molecular alterations were grouped according to the affected cellular functions (Figure 1A) in accordance with previous definitions.[36] Alterations in signaling pathways were most frequently found in younger patients and their frequency decreased with age. More than half of the pathogenic alterations found in infants (70.9%) and children (52.7%) affected signaling pathways, whereas only 21.1% of seniors and 16.5% of elderly patients carried such alterations. Vice versa, the frequency of alterations affecting epigenetic regulators or the spliceosome, as well as TP53 alterations increased with age. Alterations in transcription factors were most frequently found in children (22.1%) and AYA (20.9%), while alterations of NPM1 were most frequent in adults (16.4%) and AYA (12.5%). Alterations in genes of the cohesin complex were equally distributed across age groups. Figure 1B shows the age distribution for individual molecular alterations. The lowest median age was found for KIT alterations at 18 years (interquartile range [IQR]: 11–51), while alterations of SRSF2 had the highest median age at 68 years (IQR: 59–74). The median age for patients with NPM1 alterations was 55 years (IQR: 44–64), similar to those with FLT3-ITD at 53 years (IQR: 35–64), while patients with TP53 alterations had a median age of 63 years (IQR: 54–69). Cytogenetics were categorized into ELN2022 risk groups.[10] The highest frequency of favorable risk cytogenetics was found in children and decreased with age, while adverse risk cytogenetics were frequently found in infants followed by seniors and elderly patients (Figure S1A). The cytogenetic alterations t(8;21) and inv(16) or t(16;16) were more frequent in younger patients with a median age of 15 years (IQR: 10–41) and 23 years (IQR: 12–48), respectively. Contrastingly, del (7q), -7, -5, -17, and del(5q) were associated with a median age above 55 years, with del(5q) at the upper end of the spectrum with a median age of 65 years (IQR: 55–71; Figure S1B). Cases of therapy-associated AML (tAML) were rare, accounting for 3.4% ($n = 54/1606$) in the SAL cohort and 4.8% ($n = 35/728$) in the cohort initially published by Bottomly et al., while information on tAML status was missing for all pediatric cases, although the proportion of tAML is reportedly very low in this age group.[37]

ML predicts patient outcome and quantifies prognostic impact

Three different ML models (RF, LR, and XGB) were compared for their predictive performance regarding CR and 2-year OS prediction. For each model, 100 runs were performed. For CR prediction, the best performing model was RF with an average AUROC of 0.801 (range:
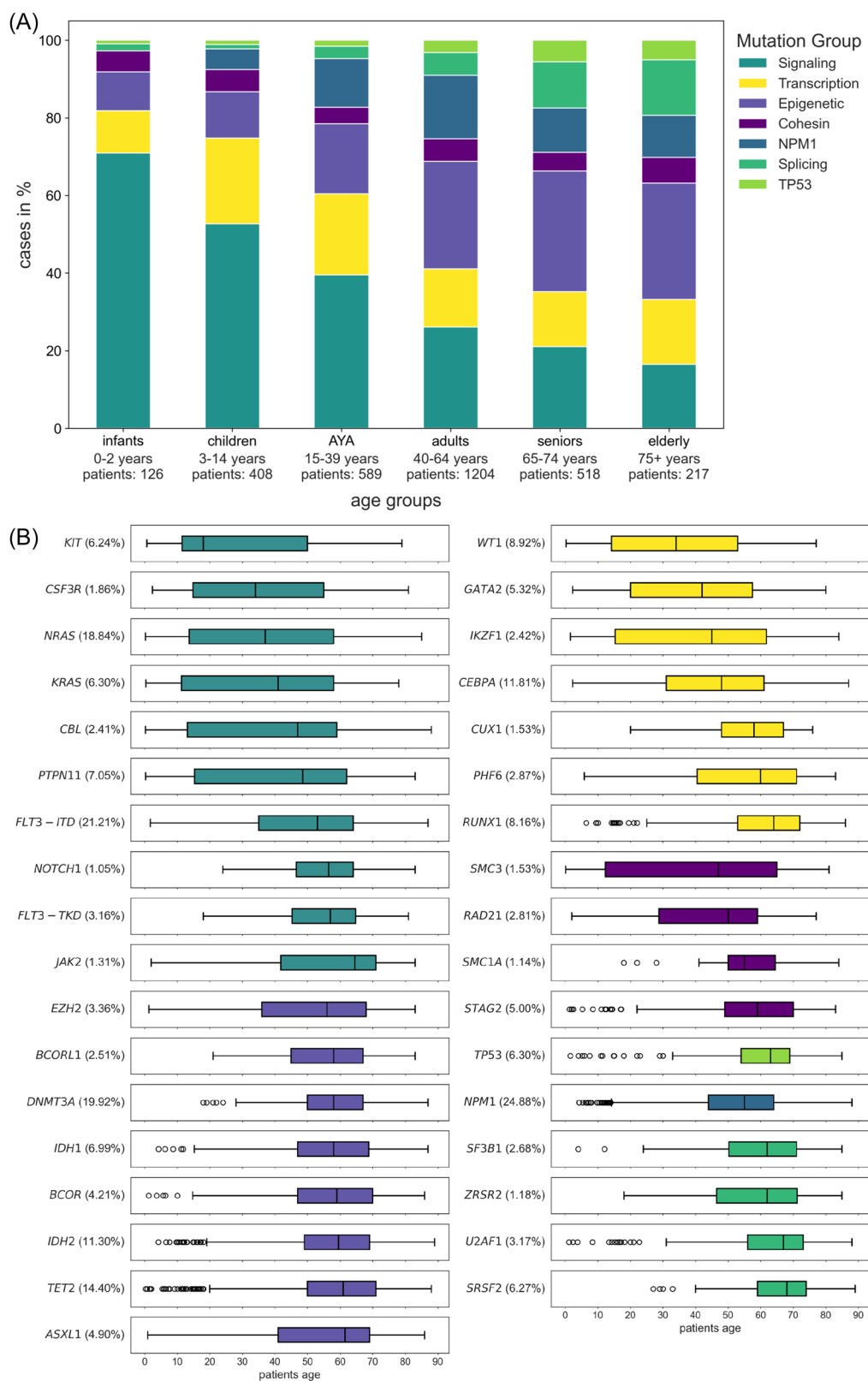
FIGURE 1 Distribution of molecular alterations across age groups. Genetic alterations are grouped according to their function, and their distribution is compared between age groups (A). Alterations of signaling pathways were more prevalent in younger patients and decreased with age, while alterations of epigenetic regulators, the spliceosome, and *TP53* were rare in younger patients and continuously increased with age. Boxplots show the age distribution of each individual alteration (B). Boxplot: bold vertical line, median; box, interquartile range (IQR, i.e., 25th to 75th percentile); lower whisker, Q1 − 1.5 * IQR; upper whisker, Q3 + 1.5 * IQR; dots, outliers.

0.796–0.807; Figure S2A) and an average MCC of 0.446 (range: 0.421–0.477; Figure S2C). XGB achieved an average AUROC of 0.773 (range: 0.768–0.779; Figure S2A) and average MCC of 0.387 (range: 0.361–0.414; Figure S2C), while LR had an AUROC of 0.791 and MCC of 0.408 (since LR is a deterministic model, no ranges can be given). For the prediction of 2-year OS, XGB outperformed RF and LR, achieving an average AUROC of 0.791 (range: 0.788–0.800; Figure S2B) and average MCC of 0.437 (range: 0.0.417–0.465; Figure S2D). RF showed an average AUROC of 0.769 (range: 0.764–0.775; Figure S2B) and average MCC of 0.421 (range: 0.393–0.440; Figure S2D), while LR had an AUROC of 0.780 and MCC of 0.402.

An individual variable's influence on model decision was assessed using SHAP values. The impact on model decisions for the best performing model for CR prediction, RF, is shown in Figure 2A, where SHAP values > 0 indicate a positive impact on CR achievement and SHAP values < 0 indicate a negative impact on CR achievement (i.e., a higher likelihood of induction failure). Variables in Figure 2 are sorted from most influential (top) to least influential (bottom). Each single dot represents a single patient. We found age to be the most influential variable for our model. Higher feature values for age (red) were associated with unfavorable model predictions (SHAP < 0). However, a proportion of patients with higher age were also predicted to have favorable outcomes (SHAP > 0), while patients with younger age (blue dots) were predominately predicted to have favorable outcomes (SHAP > 0). Genetic alterations were treated as binary variables (0 = wildtype, 1 = mutated/altered). Hence, the presence of an alteration (red dots) either indicates treatment response (SHAP > 0) or treatment failure (SHAP < 0) with differing magnitudes. Absence of an alteration (wildtype) is indicated by blue dots. Achievement of CR was found to be more likely for patients with alterations of NPM1, CEBPA, FLT3-ITD, KIT, GATA2, inv(16), and t(8;21). Treatment failure was associated with alterations of TP53, RUNX1, ASXL1, SRSF2, U2AF1, TET2, PHF6, and SF3B1, as well as -7, del(5q), -17, and trisomy 8. Beeswarm plots for XGB and LR regarding CR prediction are shown in Figures S3 and S4, respectively.

This analysis was repeated for prediction of 2-year OS. Figure 2B shows the corresponding beeswarm plot for the best performing model in 2-year OS prediction, which was XGB. Again, patient age was selected as the most influential variable in model decisions. As patient age decreased (blue), the likelihood of favorable outcomes (SHAP > 0) increased, and vice versa, as patient age increased (red), the likelihood of adverse outcomes (SHAP < 0) increased. However, in comparison to CR prediction (Figure 2A), heterogeneity in outcomes was more commonly indicated by a larger proportion of blue dots on the negative (%) and red dots on the positive sides (%) of the null. As for genetic alterations, NPM1 was again selected as the most important genetic variable, with patients harboring an alteration being more likely to survive beyond the 2-year mark. Interestingly, the effect of FLT3-ITD was found to be reversed compared to CR prediction, as patients bearing the alteration were now more likely to experience adverse outcomes. Further variables associated with favorable outcomes were t(8;21), inv(16), and -Y, as well as alterations of CEBPA, IDH2, GATA2, and STAG2, while alterations of DNMT3A, TP53, NRAS, WT1, KRAS, ASXL1, RUNX1, and U2AF1 as well as -7 and del(5q) were associated with higher risk. Beeswarm plots for RF and LR are shown in Figures S5 and S6, respectively.

Genetic alterations and their impact on patient outcome are shaped by age groups

To highlight differences in magnitude between individual variables' impact on RF's predictions for CR, we next compared mean SHAP values per variable for the overall cohort. Figure 3A ranks mean SHAP values for the most influential variables from top to bottom. Patient age was the most influential variable with a SHAP value of 0.217, surpassing genetic alterations. Among genetic alterations, NPM1, TP53, RUNX1, inv(16), and -7 had substantial influences on model decisions, while the remaining variables exhibited progressively smaller effects. Next, we analyzed the contributions of each variable to the overall SHAP values, broken down by age groups, to illustrate the differential impact of variables across age demographics. Figure 3B displays these differing contributions as percentages of mean SHAP values per variable and age group. For instance, patient age itself is highly influential on model decisions, as reflected by a large mean SHAP value (Figure 3A), which was primarily derived from contributions from patients in the senior and elderly groups, accounting for 29.9% and 33.5% of the overall effect, respectively (Figure 3B). For alterations of NPM1, the largest contributions to the mean SHAP value were derived from patients in the AYA (14.6%), adult (28.8%), and senior (20.2%) groups, consistent with age distributions for NPM1 alterations (Figure 1). Interestingly, some alterations exhibited a contrasting pattern: Although alterations of TP53 were more prevalent in older patients (Figure 1), their contribution to model decisions predominantly came from younger patient groups (children 16.4%, AYA 23.4%, adults 21.0%). A similar pattern was found for -7, del(5q), -17 and, though to a smaller degree, for alterations of RUNX1 and ASXL1. Conversely, inv(16) and t(8;21) were more frequently found in younger patients (Figure S1), yet the highest contribution to mean SHAP values were observed for patients in the senior (inv(16): 22.0%; t(8;21): 28.6%) and elderly (inv(16): 25.3%; t(8;21): 28.3%) age groups. Mean SHAP values and age-group-wise contributions for CR prediction with XGB and LR are displayed in Figures S7 and S8, respectively.

For features selected for 2-year OS, the influence of age surpassed that of genetic alterations by far, represented by a mean SHAP value of 0.300 (Figure 4A). Alterations of NPM1 dominated the impact of genetic alterations with a mean SHAP value of 0.100, while FLT3-ITD, alterations of DNMT3A, t(8;21), altered CEBPA, inv(16), and the remaining alterations had a progressively smaller impact on model decisions. The individual contributions to a variable's mean SHAP value were again analyzed per age group (Figure 4B). Patient age itself was most influential as a prognostic marker in infants (19.4% contribution to the mean SHAP value), children (15.9%), seniors (21.5%), and AYA (22.9%), while it played only a minor role in AYA and adults. For alterations of NPM1, the contributions to its mean SHAP value again were consistent with its age distribution (Figure 1). FLT3-ITD, in contrast to CR prediction (Figure 2A) now an adverse risk marker for 2-year OS (Figure 2B), also showed a consistent pattern with its age distribution (Figure 1). However, similar to the pattern observed for CR prediction, some genetic variables exhibited more pronounced effects (either as positive or negative risk markers) when found outside their age distribution. This was most pronounced for alterations of ASXL1 deriving 45.1% of its mean SHAP value only from patients (2.4% [n = 3] of 126 infant patients) in the infant group (Figure 4B), while its median age exceeded 60 years (Figure 1). Similar effects were found for -7, where 55.6% of its mean SHAP value came from infants, children, and AYA, similar to del(5q) with 43.1% and alterations of U2AF1 with 34.1%, even though these alterations typically occurred in individuals with a median age beyond 60 years (Figures 1 and S1). Mean SHAP values and age-group-wise contributions for 2-year OS prediction with RF and LR are displayed in Figures S9 and S10, respectively.

In subgroup analysis, we dichotomized for female (47.0%, n = 1439) and male (53.0%, n = 1623) patients within our cohorts. For CR prediction, RF showed better performance in male patients with an AUROC of 0.82 and an MCC of 0.47 compared to

female patients who had an AUROC of 0.78 and an MCC of 0.42. Similarly, for 2-year OS prediction, XGBoost (XGB) performed better in male patients, achieving an AUROC of 0.79 and an MCC of 0.46, compared to an AUROC of 0.74 and an MCC of 0.36 in
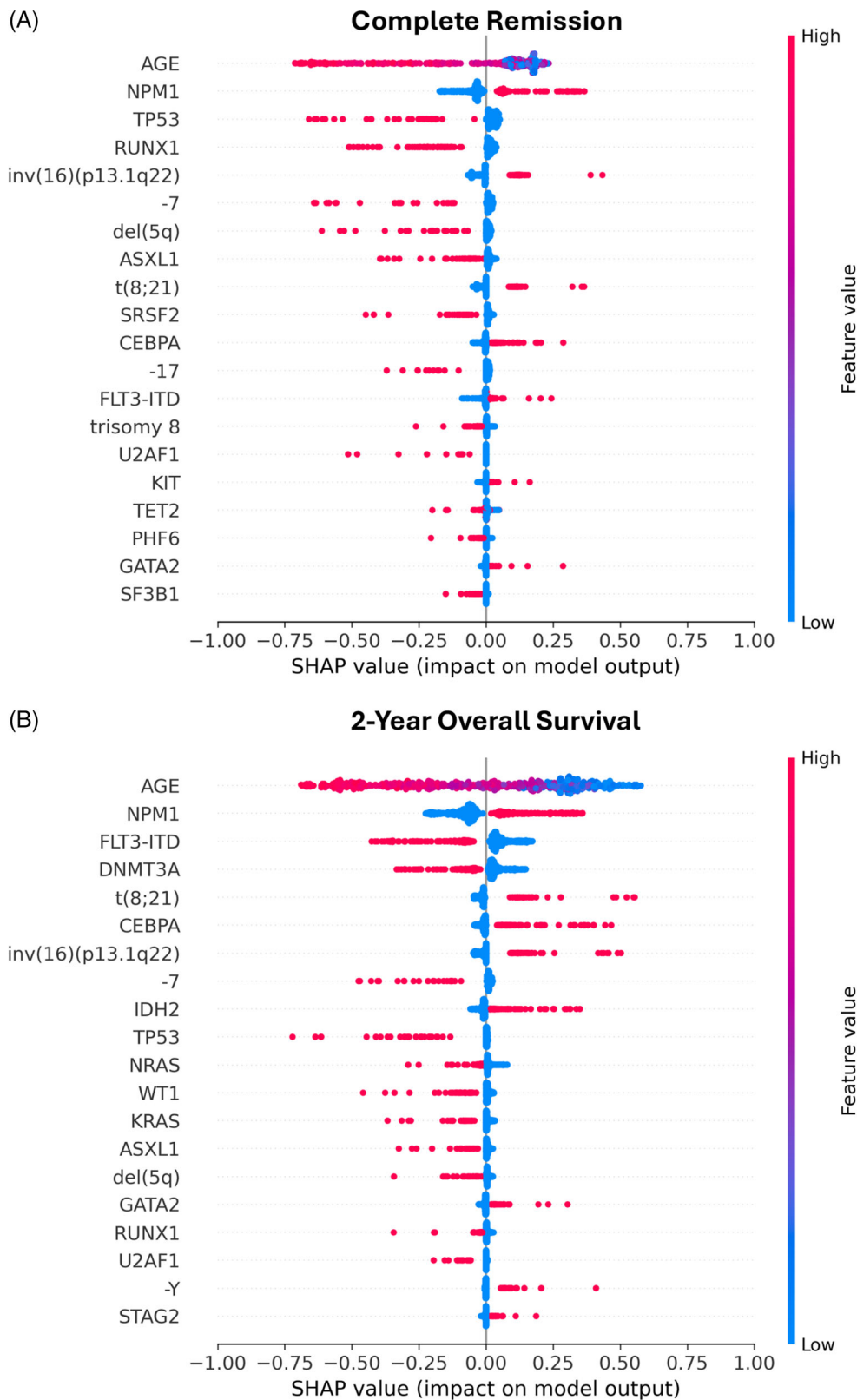

FIGURE 2 (See caption on next page).

FIGURE 2 SHAP beeswarm plot for prediction of complete remission after intensive induction therapy with Random Forest (A) and 2-year overall survival with XGBoost (B). The influence of individual variables on model predictions is displayed for the best performing models regarding complete remission (CR) (Random Forest; A) and 2-year overall survival (OS) prediction (XGBoost; B). Automatically selected variables are listed from most influential (top) to least influential (bottom). SHAP values (x-axis) represent the impact of each variable on model decisions. Positive SHAP values (>0) indicate a favorable prediction (i.e., achievement of CR), while negative values (<0) indicate an unfavorable prediction (i.e., treatment failure). Each single dot represents a single patient. Age was treated as a continuous variable. Hence, high feature values (red) represent older patients, while lower feature values (blue) represent younger patients. Genetic alterations were treated as binary variables: present/mutated (red) versus absent/wildtype (blue). The magnitude of the effect per patient is indicated by the SHAP value, with deviations from 0 showing either a positive (>0) or negative (<0) effect on treatment response. A positive impact on CR prediction (A) was found for alterations of *NPM1*, *CEBPA*, *FLT3*-ITD, *KIT*, *GATA2*, inv(16), and t(8;21). Conversely, a negative impact (i.e., prediction of treatment failure) was observed for alterations of *TP53*, *RUNX1*, *ASXL1*, *SRSF2*, *U2AF1*, *TET2*, *PHF6*, and *SF3B1*, as well as -7, del(5q), -17, and trisomy 8. A positive impact on 2-year OS prediction (B) was found for alterations of *NPM1*, *CEBPA*, *IDH2*, *GATA2*, and *STAG2*, as well as t(8;21), inv(16), and -Y. A negative impact on model predictions (i.e., death before the 2-year mark) was observed for alterations of *DNMT3A*, *TP53*, *NRAS*, *WT1*, *KRAS*, *ASXL1*, *RUNX1*, *U2AF1*, and *FLT3*-ITD, as well as -7 and del(5q).
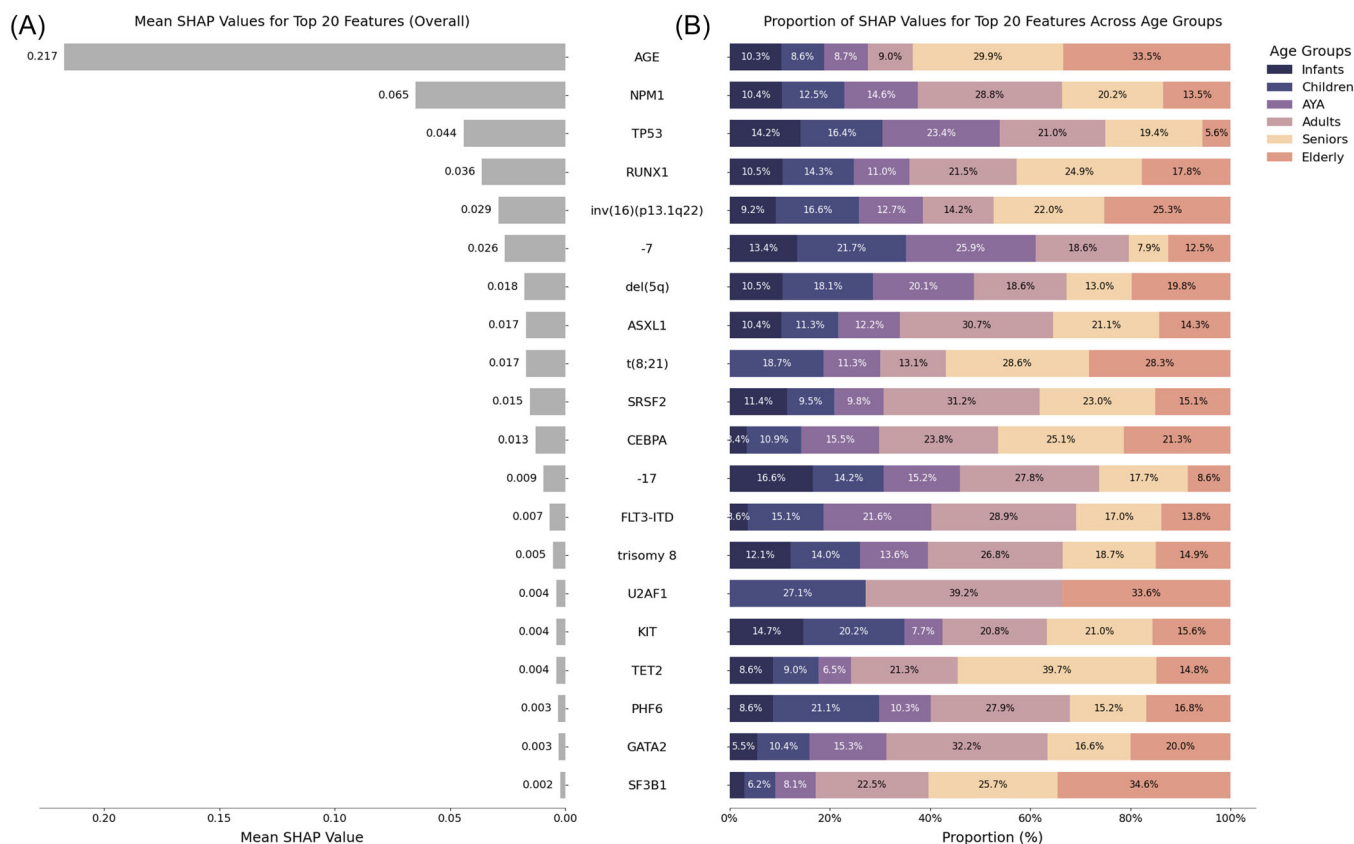

FIGURE 3 Age-group-wise impact on model inference for prediction of complete remission with Random Forest. For Random Forest (RF), individual variables impacting model decisions were ranked according to their mean SHAP values from highest to lowest (A). The individual contributions for a given feature were traced back to the different age groups. The effect size derived from each age group is shown as percentage of the individual contribution per age group to the variable's mean SHAP value (B), indicating differential prognostic impacts per age group for each single variable. We observed that no single variable had a uniform contribution to the overall SHAP values across all age groups, but rather that contributions to model explainability varied largely between age groups per variable.

female patients. In both CR and 2-year OS predictions, age remained the most influential variable regardless of sex. Regarding genetic alterations, we found *NPM1*, *RUNX1*, -7, *TP53*, del(5q), and *ASXL1* to be most influential for CR prediction in female patients (Figure S11), while *NPM1*, *RUNX1*, *TP53*, inv(16), *SRSF2*, and *CEBPA* were the most influential variables for male patients (Figure S12). For 2-year OS, the most influential genetic variables in female patients (Figure S13) were *FLT3*-ITD, *NPM1*, *DNMT3A*, *CEBPA*, *IDH2*, and t(8;21), while for male patients (Figure S14) *NPM1*, *FLT3*-ITD, *TP53*, inv(16), *CEBPA*, and -7 were most influential on model decisions. Notably, no individual genetic variable had a consistent effect across all age groups. Instead, the effect sizes varied by age.

Age-dependency of prognostic markers is confirmed in univariable analyses

To confirm the age-dependency of the observed variables selected by our models for CR and 2-year OS predictions, we next conducted univariable analyses for each alteration per age group.
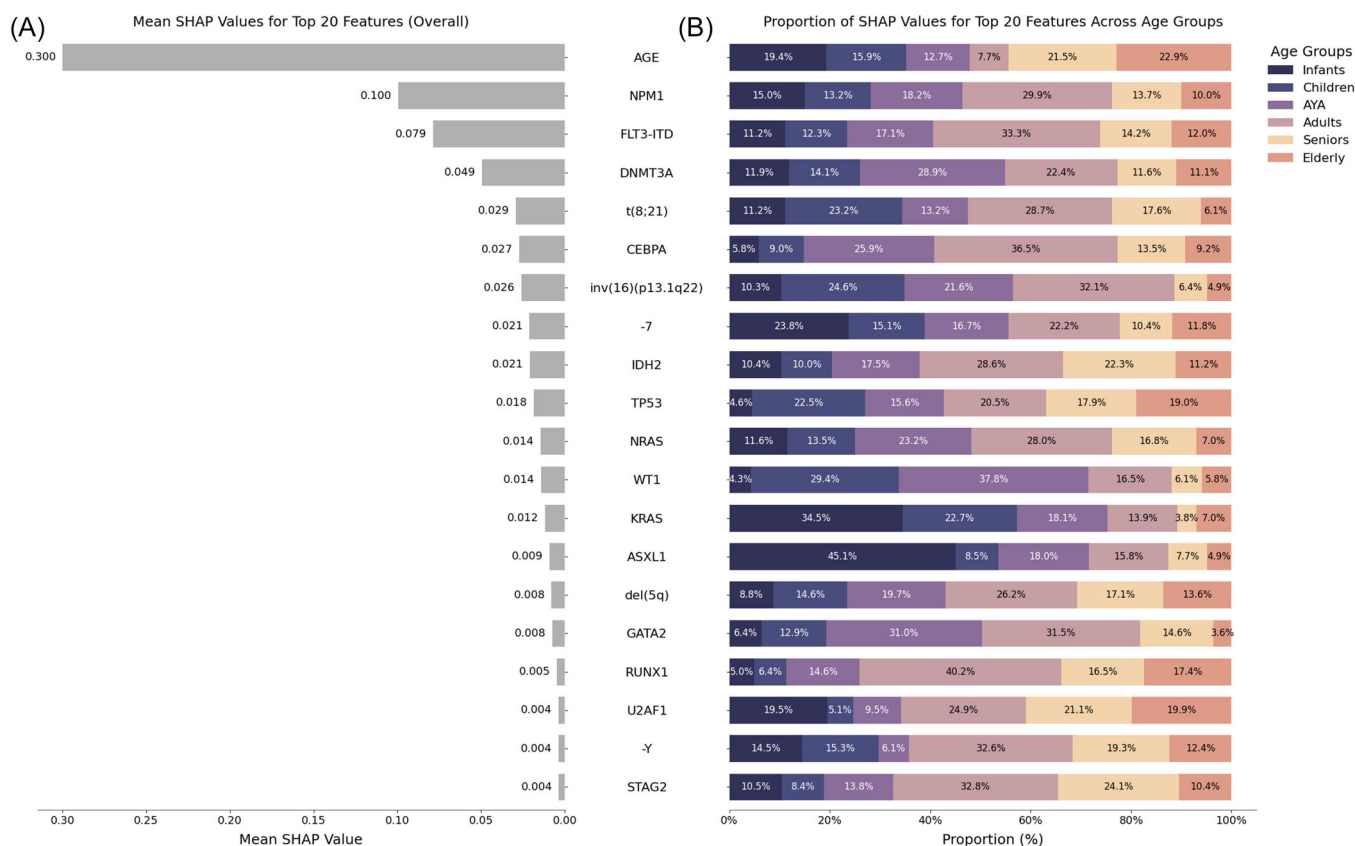
FIGURE 4 **Age-group-wise impact on model inference for prediction of 2-year overall survival with XGBoost.** For XGBoost (XGB), individual variables impacting model decisions were ranked according to their mean SHAP values from highest to lowest **(A)**. The individual contributions for a given feature were traced back to the different age groups. The effect size derived from each age group is shown as percentage of the individual contribution per age group to the variable's mean SHAP value **(B)**, indicating differential prognostic impacts per age group for each single variable. Analogous to Figure 3, contributions to mean SHAP values were not uniform across age groups, indicating differing magnitudes of impact on model decisions per variable across all age groups.

OR and HR for achievement of CR and 2-year OS with 95% CI are visualized in Figure 5. Detailed numerical values for OR/HR, 95% CI, *p* values, adjusted *p* values, as well as the number of patients per age group and variable are shown in Tables S4 and S5. Generally, for both CR and 2-year OS, we observed the impact of genetic alterations to be dependent on age groups, as no single alteration exhibited a uniform favorable or adverse effect across all six age groups (Figure 5). After adjusting for multiple hypotheses testing using the method by Benjamini and Hochberg,[33] favorable treatment response (i.e., significantly higher odds of achieving CR) was only found for adult and senior patients with *NPM1* alterations, children with *FLT3*-ITD, and adult patients with *CEBPA* alterations. Conversely, treatment failures (i.e., significantly lower odds of achieving CR) were observed in AYA, adults, and seniors with altered *TP53*, adults and seniors with altered *RUNX1* and *U2AF1*, as well as AYA and adults with either -7 or del(5q), and adults with -17 (Table S4). Regarding 2-year OS, favorable outcomes, i.e., significantly lower HRs, were found for AYA, adults, and seniors with alterations of *NPM1*, seniors with alterations of *IDH2*, children and adults with inv(16)/t(16;16), as well as AYA and adults with t(8;21). Adverse risk (i.e., significantly higher HRs) was observed for infants, AYA, adults, seniors, and elderly with *TP53* alterations, infants and adults with altered *ASXL1*, children with alterations in either *WT1*, *KRAS*, or *FLT3*-ITD, adults and seniors with *U2AF1* alterations, adults with *RUNX1*

alterations, -7 in children, AYA, adults, and seniors, and del(5q) in adults, seniors, and elderly.

DISCUSSION

Current treatment recommendations, such as ELN2022,[10] function essentially as decision trees, assuming that once a genetic variable is identified, its impact on personalized risk is context-insensitive. This implies that an alteration has a fixed effect irrespective of patient context, with limited consideration of co-occurring risk-modifying alterations. Given the simplicity of this model, designed for practical clinical decision-making, we hypothesized that incorporating patient age with genetic alterations could offer additional insights into risk modification.

In a large multicenter cohort of both pediatric and adult AML patients, we identified substantial differences in the prevalence of genetic alterations affecting pathways and cell functions between younger and older patients. Alterations in signaling pathways were more commonly found in younger patients, whereas alterations in epigenetic regulators, the spliceosome, and *TP53* increased with age, consistent with previous reports.[13,38,39] We trained three ML models to predict CR and 2-year OS and extract prognostically relevant variables. With an AUROC of 0.801, RF achieved the best performance of our models for CR prediction, while XGB outperformed RF
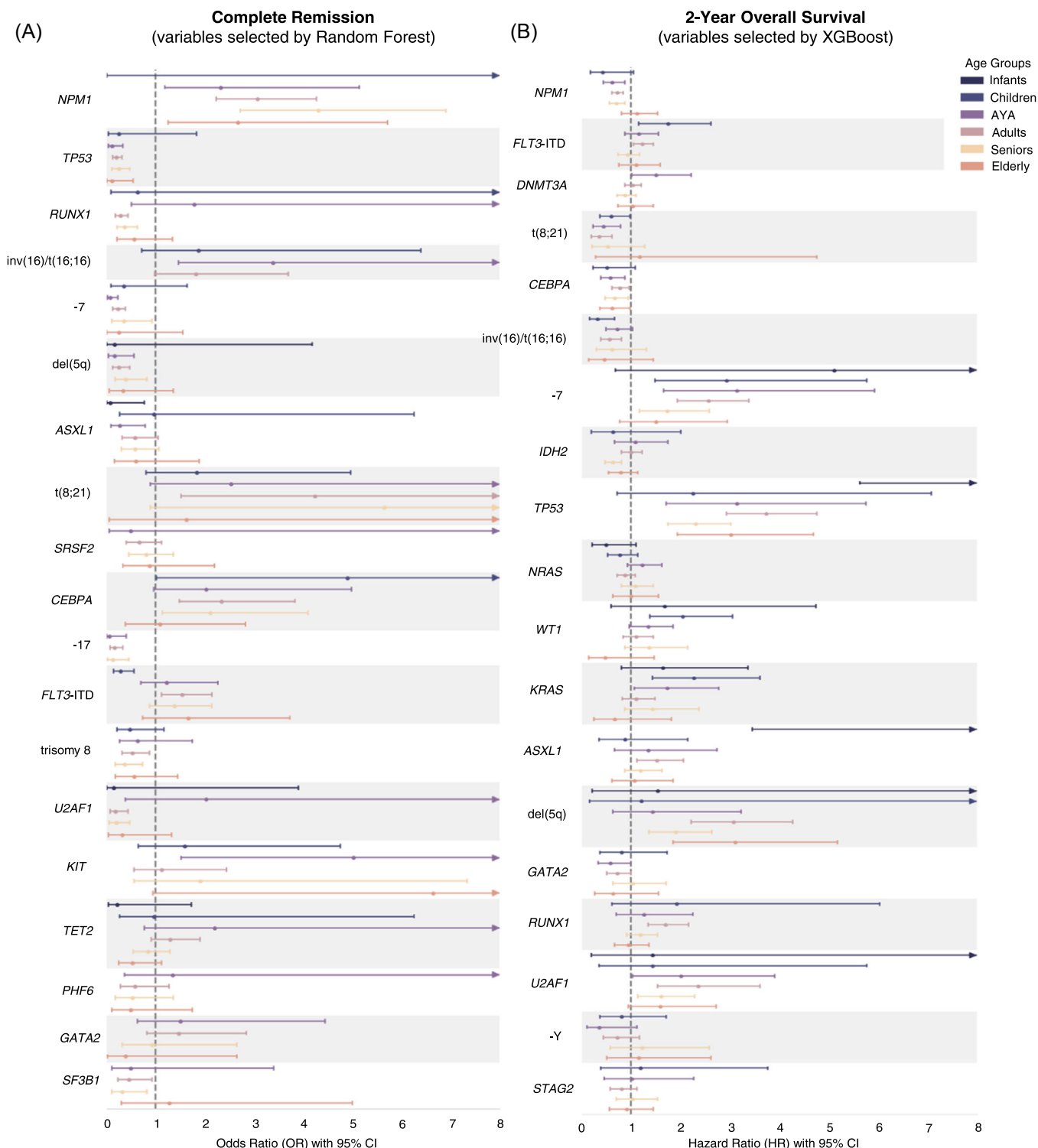
FIGURE 5 **Forest plots for univariable analyses of genetic alterations per age group.** Univariable analyses per age group for complete remission **(A)** and 2-year OS **(B)** regarding genetic alterations that were previously selected by machine learning models confirmed the age-dependency reflected in SHAP values. No single alteration was uniformly associated with favorable or adverse effects, rather individual alterations were found to be prognostic depending on the assignment of the patient bearing the alteration to one of the age groups. Odds ratios (OR) and hazard ratios (HR) are displayed as point estimates with 95% confidence intervals (CI). Arrowheads indicate that the upper limit of the CI is out of bounds of the plot. If a combination of genetic alteration and age group had a non-significant effect with an OR/HR out of bounds of the plot, it is not shown. Patient numbers as well as numerical values for OR/HR, 95% CI, p values, and adjusted p values are displayed in Tables S4 and S5 for CR and 2-year OS, respectively.

and LR with an AUROC of 0.773 in predicting 2-year OS. Several recent studies also demonstrated ML's capability of accurately predicting patient outcomes in AML leveraging a wealth of available patient data from clinical trials, genetic testing, and knowledge banks, albeit none of these studies combined pediatric and adult patients.[40–45] To make model decisions explainable, we used SHAP values quantifying the impact of a variable on model outputs. The magnitude of the prognostic effect was traced back to age groups, unveiling differential contributions to model decisions depending on the age-related context in which a genetic alteration was found. Previous studies usually dichotomized adult patients into younger and older at thresholds of 60, 65, or 75 years of age, reporting differential prognostic effects for alterations such as FLT3-ITD, NPM1, DNMT3A, and TP53, among others.[46–49] Our model further differentiates between six age groups, indicating varying effects of differing magnitudes per genetic variable. In line with previous studies, alterations of NPM1 were a top contributor to positive model decisions for both CR and 2-year OS. The positive prognostic impact of NPM1 mutations are well described in adult AML,[50] while NPM1 mutations in children are rare,[51] although similar favorable outcomes[52] have been reported. However, recent reports suggest that these favorable effects do not necessarily extend to older patients[53,54] and are further abrogated in individuals with co-occurring mutations in DNMT3A and FLT3-ITD.[55] The prevalence of both DNMT3A and RUNX1 alterations increases with age and an unfavorable prognostic impact has been described across all age groups.[48,56–61] A similarly uniform but positive prognostic impact has been reported for core binding factor AML,[62] showing high CR rates even in older patients receiving intensive treatment while OS decreases with age.[63] While the majority of autonomously selected variables in our models are in accordance with ELN recommendations[10] regarding effect directionality, there were two outliers: FLT3-ITD influenced model decisions positively toward CR achievement, while it had a negative impact on overall survival. Recently, studies showed a differing impact of FLT3-ITD dependent on patient age[64] and co-occurring alterations,[65] while high risk disease was reported for a subgroup of patients with altered NPM1, DNMT3A, and FLT3-ITD.[55] Further, alterations of GATA2 contributed toward both increased odds of treatment response and better overall survival. A recent meta-analysis of 13 cohorts[66] found alterations of GATA2 to be associated with poor outcomes in MDS, but not in AML. Previous reports suggest an adverse effect of high levels of GATA2 expression in AML.[67,68] This highlights the need to approach risk stratification in a more nuanced manner contrary to current simplified binary risk allocation models: A genetic alteration can predict prognosis depending on the context in which it is found, not merely on its absence or presence. The frequency with which a genetic alteration occurred expectedly differed with age. However, the risk associated with the specific alteration was occasionally found to be disassociated from its age distribution. This may reflect patterns associated with higher risk disease and may prompt closer monitoring of younger AML patients with aging-related patterns. Additionally, data-driven analysis with ML models addresses a key limitation of conventional univariable analysis, which relies on manual a priori variable selection and often fails to capture complex, nonlinear relationships between features. This capability becomes increasingly crucial as affordable genetic sequencing and omics techniques become more integrated into clinical practice.

A notable limitation of our study is the restriction to genetic alterations covered by the myeloid panel used in the SAL studies, which was necessary to merge all available datasets. This highlights the trade-off between the granularity of genetic information and the applicability of ML models, which require large datasets for training. However, WGS/WES data were not available for the SAL cohort, necessitating a compromise by focusing on alterations covered in the myeloid panel. Nevertheless, these alterations comprise the most commonly used items in routine diagnostics, allowing for a widespread implementation of our findings. Importantly, models such as ELN generate mutually exclusive patient subgroups based on predefined (evidence-based) assumptions of risk associated with certain variables (and therefore essentially function as manual clustering). Clustering frameworks such as the ELN model allow for straightforward comparisons between versions (e.g., ELN2017 vs. ELN2022) through Kaplan–Meier or Sankey analyses, as patients can be easily (and manually) assigned to groups. In contrast, our supervised models start by predicting clinical outcomes (CR, 2-year OS) and retrospectively determine which variables contributed most to those predictions, allowing for a more context-sensitive evaluation of risk factors on a patient-level. Given this fundamental methodological difference, direct performance comparisons between our supervised approach and clustering frameworks such as ELN's model are methodologically not directly comparable; however, their results may complement each other. Moreover, 15.9% of patients in the cohort derived from Bottomly et al.[16] were treated with non-intensive therapy, while all other patients including those in the SAL, TARGET, and Fred Hutch cohort were treated with intensive chemotherapy. Balancing homogeneity to reduce potential confounding with the heterogeneity required to ensure generalizability and sufficient sample size is a well-known challenge in medical research, particularly for rare diseases such as AML. Increasing the sample size may introduce the possibility of further substratification by treatment modalities such as type of intensive induction therapy without risking stability in the training process or generalizability of results in a limited sample as ours. A more comprehensive analysis, ideally covering a broader range of genetic alterations and gene expression data, is desirable to further characterize the interplay between age, AML biology, and patient outcomes, also extending to different treatment regimens (intensive, non-intensive, targeted agents) as an adjustable variable to potentially predict how an individual patient may fare given a certain therapy. Ideally, this may also include predicting other clinically relevant endpoints such as MRD negativity or relapse-free survival. Further, information on comorbidities influencing patient outcomes and potentially modulating disease biology or limiting survival should be incorporated when available. This is particularly relevant for older patients, as comorbidities that limit overall survival are expected to be more prevalent and, potentially, contribute more strongly to disease and outcome modification. To this end, international collaborative efforts in data collection and harmonization are needed to provide a representative database for adaptive risk stratification model development, as manual feature selection in current models in clinical routine likely underestimates inter-variable connections. For routine clinical applicability, future models need to seamlessly integrate into routine workflows without posing additional manual labor to healthcare providers. Integration in clinical information systems and electronic health records may be an option if patient privacy is preserved. In line with this, our analysis represents only a step forward from previous studies that dichotomized age, whereas we use six age groups entailing both pediatric and adult AML. While one could argue for different age group thresholds, no single cut-off accurately reflects biological variation. We selected our age group thresholds based on accepted cut-offs from academic societies and the scientific literature to capture as much heterogeneity as possible.[28–32] Future work may include subgroup analysis for different treatment modalities such as ASCT.

In summary, our study underscores the need for improved risk stratification in AML by contextualizing prognostic genetic alterations selected by highly accurate ML models predicting both CR and 2-year OS. Our findings not only improve upon the risk stratification for

individual genetic variables but also leave simplified age dichotomizations behind and prompt a close monitoring of younger patients with genetics related to aging.

AUTHOR CONTRIBUTIONS

Jan-Niklas Eckardt: Conceptualization (lead); data curation (equal); formal analysis (lead); investigation (lead); methodology (lead); project administration (lead); resources (equal); software (equal); validation (equal); visualization (equal); writing—original draft preparation (lead). **Waldemar Hahn**: Data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); software (lead); validation (equal); visualization (equal); writing—review and editing (equal). **Rhonda E. Ries**: Data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); resources (equal); validation (equal); writing—review and editing (equal). **Szymon D. Chrost**: Data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); software (lead); validation (equal); visualization (equal); writing—review and editing (equal). **Susann Winter**: Formal analysis (equal); investigation (equal); project administration (equal); validation (equal); writing—review and editing (equal). **Sebastian Stasik**: Data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); resources (equal); validation (equal); writing—review and editing (equal). **Christoph Röllig**: Formal analysis (equal); investigation (equal); resources (equal); validation (equal); writing—review and editing (equal). **Uwe Platzbecker**: Formal analysis (equal); investigation (equal); resources (equal); validation (equal); writing—review and editing (equal). **Carsten Müller-Tidow**: Formal analysis (equal); investigation (equal); resources (equal); validation (equal); writing—review and editing (equal). **Hubert Serve**: Formal analysis (equal); investigation (equal); resources (equal); validation (equal); writing—review and editing (equal). **Claudia D. Baldus**: Formal analysis (equal); investigation (equal); resources (equal); validation (equal); writing—review and editing (equal). **Christoph Schliemann**: Formal analysis (equal); investigation (equal); resources (equal); validation (equal); writing—review and editing (equal). **Kerstin Schäfer-Eckart**: Formal analysis (equal); investigation (equal); resources (equal); validation (equal); writing—review and editing (equal). **Maher Hanoun**: Formal analysis (equal); investigation (equal); resources (equal); validation (equal); writing—review and editing (equal). **Martin Kaufmann**: Formal analysis (equal); investigation (equal); resources (equal); validation (equal); writing—review and editing (equal). **Andreas Burchert**: Formal analysis (equal); investigation (equal); resources (equal); validation (equal); writing—review and editing (equal). **Johannes Schetelig**: Formal analysis (equal); investigation (equal); resources (equal); validation (equal); writing—review and editing (equal). **Martin Bornhäuser**: Formal analysis (equal); investigation (equal); resources (equal); validation (equal); writing—review and editing (equal). **Markus Wolfien**: Data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); software (supporting); validation (equal); writing—review and editing (equal). **Soheil Meshinchi**: Conceptualization (equal); formal analysis (equal); funding acquisition (equal); investigation (equal); resources (equal); supervision (equal); validation (equal); writing—review and editing (equal). **Christian Thiede**: Conceptualization (equal); formal analysis (equal); funding acquisition (equal); investigation (equal); resources (equal); supervision (equal); validation (equal); writing—review and editing (equal). **Jan Moritz Middeke**: Conceptualization (equal); formal analysis (equal); funding acquisition (equal); investigation (equal); resources (equal); supervision (equal); validation (equal); writing—review and editing (equal).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data pertaining to the SAL study group cohort and Seattle cohort can be inquired from the corresponding author upon reasonable request. The results published here are in part based upon publicly available data generated by the Therapeutically Applicable Research to Generate Effective Treatments (https://www.cancer.gov/ccg/research/genome-sequencing/target) initiative, phs000218, specifically the AML subgroup (phs000465), and data available at the Genomic Data Commons (https://portal.gdc.cancer.gov).

ORCID

Jan-Niklas Eckardt http://orcid.org/0000-0002-3649-2823
Waldemar Hahn http://orcid.org/0009-0007-7934-930X
Rhonda E. Ries http://orcid.org/0000-0002-3702-7923
Susann Winter http://orcid.org/0000-0002-4228-4537
Sebastian Stasik http://orcid.org/0000-0003-0271-4894
Christoph Röllig http://orcid.org/0000-0002-3791-0548
Uwe Platzbecker http://orcid.org/0000-0003-1863-3239
Carsten Müller-Tidow http://orcid.org/0000-0002-7166-5232
Hubert Serve http://orcid.org/0000-0001-8472-5516
Claudia D. Baldus http://orcid.org/0000-0002-0748-834X
Martin Kaufmann http://orcid.org/0009-0000-1885-5451
Martin Bornhäuser http://orcid.org/0000-0002-5916-3029
Markus Wolfien http://orcid.org/0000-0002-1887-4772
Soheil Meshinchi http://orcid.org/0000-0002-6276-4423
Christian Thiede http://orcid.org/0000-0003-1241-2048
Jan Moritz Middeke http://orcid.org/0000-0003-3250-293X

SUPPORTING INFORMATION

Additional supporting information can be found in the online version of this article.

REFERENCES

1. Shallis RM, Wang R, Davidoff A, Ma X, Zeidan AM. Epidemiology of acute myeloid leukemia: recent progress and enduring challenges. *Blood Rev*. 2019;36:70-87. doi:10.1016/j.blre.2019.04.005
2. Kantarjian H, Kadia T, DiNardo C, et al. Acute myeloid leukemia: current progress and future directions. *Blood Cancer J*. 2021;11(2):41. doi:10.1038/s41408-021-00425-3

3. Bhansali RS, Pratz KW, Lai C. Recent advances in targeted therapies in acute myeloid leukemia. *J Hematol Oncol*. 2023;16(1):29. doi:10.1186/s13045-023-01424-6

4. Appelbaum FR, Gundacker H, Head DR, et al. Age and acute myeloid leukemia. *Blood*. 2006;107(9):3481-3485. doi:10.1182/blood-2005-09-3724

5. Yi M, Li A, Zhou L, Chu Q, Song Y, Wu K. The global burden and attributable risk factor analysis of acute myeloid leukemia in 195 countries and territories from 1990 to 2017: estimates based on the global burden of disease study 2017. *J Hematol Oncol*. 2020;13(1):72. doi:10.1186/s13045-020-00908-z

6. Song X, Peng Y, Wang X, et al. Incidence, survival, and risk factors for adults with acute myeloid leukemia not otherwise specified and acute myeloid leukemia with recurrent genetic abnormalities: analysis of the surveillance, epidemiology, and end results (SEER) database, 2001–2013. *Acta Haematol*. 2018;139(2):115-127. doi:10.1159/000486228

7. Venditti A, Cairoli R, Caira M, et al. Assessing eligibility for treatment in acute myeloid leukemia in 2023. *Expert Rev Hematol*. 2023;16(3):181-190. doi:10.1080/17474086.2023.2185603

8. Döhner H, Estey EH, Amadori S, et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood*. 2010;115(3):453-474. doi:10.1182/blood-2009-07-235358

9. Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*. 2017;129(4):424-447. doi:10.1182/blood-2016-08-733196

10. Döhner H, Wei AH, Appelbaum FR, et al. Diagnosis and management of AML in Adults: 2022 recommendations from an International Expert Panel on behalf of the ELN. *Blood*. 2022;140(12):1345-1377. doi:10.1182/blood.2022016867.

11. Creutzig U, van den Heuvel-Eibrink MM, Gibson B, et al. Diagnosis and management of acute myeloid leukemia in children and adolescents: recommendations from an international expert panel. *Blood*. 2012;120(16):3187-3205. doi:10.1182/blood-2012-03-362608.

12. Rubnitz JE, Kaspers GJL. How I treat pediatric acute myeloid leukemia. *Blood*. 2021;138(12):1009-1018. doi:10.1182/blood.2021011694.

13. Bolouri H, Farrar JE, Triche T, 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. doi:10.1038/nm.4439.

14. Chaudhury S, O'Connor C, Cañete A, et al. Age-specific biological and molecular profiling distinguishes paediatric from adult acute myeloid leukaemias. *Nat Commun*. 2018;9:5280. doi:10.1038/s41467-018-07584-1.

15. Shallis RM, Boddu PC, Bewersdorf JP, Zeidan AM. The golden age for patients in their golden years: the progressive upheaval of age and the treatment of newly-diagnosed acute myeloid leukemia. *Blood Rev*. 2020;40:100639. doi:10.1016/j.blre.2019.100639.

16. Bottomly D, Long N, Schultz AR, et al. Integrative analysis of drug response and clinical outcome in acute myeloid leukemia. *Cancer Cell*. 2022;40(8):850-864.e9. doi:10.1016/j.ccell.2022.07.002.

17. Röllig C, Thiede C, Gramatzki M, et al. A novel prognostic model in elderly patients with acute myeloid leukemia: results of 909 patients entered into the prospective AML96 trial. *Blood*. 2010;116(6):971-978. doi:10.1182/blood-2010-01-267302.

18. Schaich M, Parmentier S, Kramer M, et al. High-dose cytarabine consolidation with or without additional amsacrine and mitoxantrone in acute myeloid leukemia: results of the prospective randomized AML2003 trial. *J Clin Oncol*. 2013;31(17):2094-2102. doi:10.1200/JCO.2012.46.4743.

19. Röllig C, Kramer M, Gabrecht M, et al. Intermediate-dose cytarabine plus mitoxantrone versus standard-dose cytarabine plus daunorubicin for acute myeloid leukemia in elderly patients. *Ann Oncol*. 2018;29(4):973-978. doi:10.1093/annonc/mdy030.

20. Röllig C, Serve H, Hüttmann A, et al. Addition of sorafenib versus placebo to standard therapy in patients aged 60 years or younger with newly diagnosed acute myeloid leukaemia (SORAML): a multicentre, phase 2, randomised controlled trial. *Lancet Oncol*. 2015;16(16):1691-1699. doi:10.1016/S1470-2045(15)00362-9.

21. World Medical Association declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013;310(20):2191-2194. doi:10.1001/jama.2013.281053.

22. 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. doi:10.1200/JCO.2014.55.3628.

23. National Cancer Institute. cBioPortal for cancer genomics: pediatric acute myeloid leukemia (TARGET, 2018). Accessed June 20, 2024. https://www.cbioportal.org/study/summary?id=aml_target_2018_pub.

24. Stasik S, Schuster C, Ortlepp C, et al. An optimized targeted next-generation sequencing approach for sensitive detection of single nucleotide variants. *Biomol Detect Quantif*. 2018;15:6-12. doi:10.1016/j.bdq.2017.12.001.

25. Thiede C, Steudel C, Mohr B, et al. Analysis of FLT3-activating mutations in 979 patients with acute myelogenous leukemia: association with FAB subtypes and identification of subgroups with poor prognosis. *Blood*. 2002;99(12):4326-4335. doi:10.1182/blood.v99.12.4326.

26. Thiede C, Koch S, Creutzig E, et al. Prevalence and prognostic impact of NPM1 mutations in 1485 adult patients with acute myeloid leukemia (AML). *Blood*. 2006;107(10):4011-4020. doi:10.1182/blood-2005-08-3167.

27. Taube F, Georgi JA, Kramer M, et al. CEBPA mutations in 4708 patients with acute myeloid leukemia: differential impact of bZIP and TAD mutations on outcome. *Blood*. 2022;139(1):87-103. doi:10.1182/blood.2020009680.

28. Hardin AP, Hackell JM, Committee on Practice and Ambulatory Medicine, et al. Age limit of pediatrics. *Pediatrics*. 2017;140(3):e20172151. doi:10.1542/peds.2017-2151.

29. Adolescents and Young Adults (AYAs) with Cancer - NCI. May 2015. Accessed April 29, 2024. https://www.cancer.gov/types/aya.

30. Estey E. Acute myeloid leukemia and myelodysplastic syndromes in older patients. *J Clin Oncol*. 2007;25(14):1908-1915. doi:10.1200/JCO.2006.10.2731.

31. Abdallah M, Xie Z, Ready A, Manogna D, Mendler JH, Loh KP. Management of acute myeloid leukemia (AML) in older patients. *Curr Oncol Rep*. 2020;22(10):103. doi:10.1007/s11912-020-00964-1.

32. Choi JH, Shukla M, Abdul-Hay M. Acute myeloid leukemia treatment in the elderly: a comprehensive review of the present and future. *Acta Haematol*. 2023;146(6):431-457. doi:10.1159/000531628.

33. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc Ser B*. 1995;57(1):289-300.

34. Altman N, Krzywinski M. The curse(s) of dimensionality. *Nat Methods*. 2018;15:399-400. doi:10.1038/s41592-018-0019-x.

35. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom*. 2020;21(1):6. doi:10.1186/s12864-019-6413-7.

36. DiNardo CD, Cortes JE. Mutations in AML: prognostic and therapeutic implications. *Hematology*. 2016;2016(1):348-355. doi:10.1182/asheducation-2016.1.348.

37. Brown CA, Youlden DR, Aitken JF, Moore AS. Therapy-related acute myeloid leukemia following treatment for cancer in childhood: a population-based registry study. *Pediatr Blood Cancer*. 2018;65(12):e27410. doi:10.1002/pbc.27410.

38. Zjablovskaja P, Florian MC. Acute myeloid leukemia: aging and epigenetics. *Cancers*. 2019;12(1):103. doi:10.3390/cancers12010103.

39. Chiara Caprioli C, Federico Lussana L, Silvia Salmoiraghi S, et al. Clinical significance of chromatin-spliceosome acute myeloid leukemia: a report from the Northern Italy Leukemia Group (NILG) randomized trial 02/06. *Haematologica*. 2021;106(10):2578-2587. doi:10.3324/haematol.2020.252825.

40. Gal O, Auslander N, Fan Y, Meerzaman D. Predicting complete remission of acute myeloid leukemia: machine learning applied to gene expression. *Cancer Inform*. 2019;18:1176935119835544. doi:10.1177/1176935119835544.

41. Hernández Sánchez A, Martínez Elicegui J, Sobas M, et al. Machine learning provides individualized prediction of outcomes after first complete remission in adult AML patients - results from the harmony platform. *Blood*. 2023;142(Suppl 1):62. doi:10.1182/blood-2023-183029.

42. Wang X, Sun H, Dong Y, et al. Development and validation of a cuproptosis-related prognostic model for acute myeloid leukemia patients using machine learning with stacking. *Sci Rep*. 2024;14(1):2802. doi:10.1038/s41598-024-53306-7.

43. Sanchis JAD, Pons-Suñer P, Alvarez N, et al. A supervised machine learning model to predict therapy response and mortality at 90 days after acute myeloid leukemia diagnosis. 2023. doi:10.1101/2023.06.26.23291731.

44. Eckardt JN, Röllig C, Metzeler K, et al. Prediction of complete remission and survival in acute myeloid leukemia using supervised machine learning. *Haematologica*. 2023;108(3):690-704. doi:10.3324/haematol.2021.280027.

45. Gerstung M, Papaemmanuil E, Martincorena I, et al. Precision oncology for acute myeloid leukemia using a knowledge bank approach. *Nat Genet*. 2017;49(3):332-340. doi:10.1038/ng.3756.

46. Juliusson G, Jädersten M, Deneberg S, et al. The prognostic impact of FLT3-ITD and NPM1 mutation in adult AML is age-dependent in the population-based setting. *Blood Adv*. 2020;4(6):1094-1101. doi:10.1182/bloodadvances.2019001335.

47. Prassek VV, Rothenberg-Thurley M, Sauerland MC, et al. Genetics of acute myeloid leukemia in the elderly: mutation spectrum and clinical impact in intensively treated patients aged 75 years or older. *Haematologica*. 2018;103(11):1853-1861. doi:10.3324/haematol.2018.191536.

48. Marcucci G, Metzeler KH, Schwind S, et al. Age-related prognostic impact of different types of DNMT3A mutations in adults with primary cytogenetically normal acute myeloid leukemia. *J Clin Oncol*. 2012;30(7):742-750. doi:10.1200/JCO.2011.39.2092.

49. Tsai CH, Hou HA, Tang JL, et al. Genetic alterations and their clinical implications in older patients with acute myeloid leukemia. *Leukemia*. 2016;30(7):1485-1492. doi:10.1038/leu.2016.65.

50. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic classification and prognosis in acute myeloid leukemia. *N Engl J Med*. 2016;374(23):2209-2221. doi:10.1056/NEJMoa1516192.

51. Braoudaki M, Papathanassiou C, Katsibardi K, Tourkadoni N, Karamolegou K, Tzortzatou-Stathopoulou F. The frequency of NPM1 mutations in childhood acute myeloid leukemia. *J Hematol Oncol*. 2010;3(1):41. doi:10.1186/1756-8722-3-41.

52. Xu LH, Fang JP, Liu YC, Jones AI, Chai L. Nucleophosmin mutations confer an independent favorable prognostic impact in 869 pediatric patients with acute myeloid leukemia. *Blood Cancer J*. 2020;10(1):1-10. doi:10.1038/s41408-019-0268-7.

53. Ostronoff F, Othus M, Lazenby M, et al. Prognostic significance of NPM1 mutations in the absence of FLT3-internal tandem duplication in older patients with acute myeloid leukemia: a SWOG and UK National Cancer Research Institute/Medical Research Council report. *J Clin Oncol*. 2015;33(10):1157-1164. doi:10.1200/JCO.2014.58.0571.

54. Khan AM, Nanja Reddy S, Aly MMM, et al. Comprehensive age-stratified impact of NPM1 mutation in acute myeloid leukemia. *Blood*. 2022;140(suppl 1):1433-1434. doi:10.1182/blood-2022-165696.

55. Bezerra MF, Lima AS, Piqué-Borràs MR, et al. Co-occurrence of DNMT3A, NPM1, FLT3 mutations identifies a subset of acute myeloid leukemia with adverse prognosis. *Blood*. 2020;135(11):870-875. doi:10.1182/blood.2019003339.

56. Venugopal S, DiNardo CD, Loghavi S, et al. Differential prognostic impact of RUNX1 mutations according to frontline therapy in patients with acute myeloid leukemia. *Am J Hematol*. 2022;97(12):1560-1567. doi:10.1002/ajh.26724.

57. Greif PA, Konstandin NP, Metzeler KH, et al. RUNX1 mutations in cytogenetically normal acute myeloid leukemia are associated with a poor prognosis and up-regulation of lymphoid genes. *Haematologica*. 2012;97(12):1909-1915. doi:10.3324/haematol.2012.064667.

58. Lew-Derivry L, Marceau-Renaut A, Fenwarth L, et al. Prognostic impact of RUNX1 mutations and deletions in pediatric acute myeloid leukemia: results from the French ELAM02 study group. *Leukemia*. 2023;37(8):1723-1726. doi:10.1038/s41375-023-01931-y.

59. Shivarov V, Gueorguieva R, Stoimenov A, Tiu R. DNMT3A mutation is a poor prognosis biomarker in AML: results of a meta-analysis of 4500 AML patients. *Leuk Res*. 2013;37(11):1445-1450. doi:10.1016/j.leukres.2013.07.032.

60. Ragoonanan D, Harden A, Elgehiny A, et al. DNMT3A mutations should be considered in the risk stratification for pediatric and adolescent and young adult patients with acute myeloid leukemia. *Blood*. 2021;138:1308. doi:10.1182/blood-2021-151043.

61. Ley TJ, Ding L, Walter MJ, et al. DNMT3A mutations in acute myeloid leukemia. *N Engl J Med*. 2010;363(25):2424-2433. doi:10.1056/NEJMoa1005143.

62. Borthakur G, Kantarjian H. Core binding factor acute myelogenous leukemia-2021 treatment algorithm. *Blood Cancer J*. 2021;11(6):114. doi:10.1038/s41408-021-00503-6.

63. Mosna F, Borlenghi E, Litzow M, et al. Long-term survival can be achieved in a significant fraction of older patients with core binding factor acute myeloid leukemia treated with intensive chemotherapy. *Haematologica*. Oct 10, 2025;110(3):608-620. doi:10.3324/haematol.2024.285448.

64. Shimony S, Fell G, Chen EC, et al. FLT3-ITD does not predict inferior prognosis in acute myeloid leukemia patients aged ≥60 years. *Blood Adv*. 2023;7(18):5354-5358. doi:10.1182/bloodadvances.2023009748.

65. Tarlock K, Gerbing RB, Ries RE, et al. Prognostic impact of cooccurring mutations in *FLT3*-ITD pediatric acute myeloid leukemia. *Blood Adv*. 2024;8(9):2094-2103. doi:10.1182/bloodadvances.2023011980.

66. Han X, Liu W, Kang Z, Li D. Prognostic significance of GATA2 in patients with MDS/AML: a systematic review and meta-analysis. *Ann Hematol*. 2024;103(12):4943-4952. doi:10.1007/s00277-024-05899-2.

67. Luesink M, Hollink IHIM, van der Velden VHJ, et al. High GATA2 expression is a poor prognostic marker in pediatric acute myeloid leukemia. *Blood*. 2012;120(10):2064-2075. doi:10.1182/blood-2011-12-397083.

68. Vicente C, Vazquez I, Conchillo A, et al. Overexpression of GATA2 predicts an adverse prognosis for patients with acute myeloid leukemia and it is associated with distinct molecular abnormalities. *Leukemia*. 2012;26(3):550-554. doi:10.1038/leu.2011.235.