Integrative Genomic Analysis of Pediatric Myeloid Related Acute Leukemias Identifies Novel Subtypes and Prognostic Indicators

Maarten Fornerod¹, Jing Ma³*, Sanne Noort²*, Yu Liu⁴, Michael P. Walsh³, Lei Shi⁵, Stephanie Nance⁶, Yanling Liu⁷, Yuanyuan Wang⁶, Guangchun Song⁹, Tamara Lamprecht³, John Easton⁷, Heather L. Mulder⁷, Donald Yergeau⁷, Jacquelyn Myers⁶, Jennifer L. Kamens⁶, Esther A. Obeng⁶, Martina Pigazzi⁹,¹⁶, Marie Jarosova¹⁰, Charikleia Kelaidi¹¹, Sophia Polychronopoulou¹¹, Jatinder K. Lamba¹², Sharyn D. Baker¹³, Jeffrey E. Rubnitz⁶, Dirk Reinhardt on behalf of Berlin-Frankfurt-Munster Study Group (BFM)¹⁴, Marry M. van den Heuvel-Eibrink on behalf of the Dutch Children’s Oncology Group (DCOG)²,¹⁵, Franco Locatelli on behalf of Associazione Italiana di Ematologia e Oncologia Pediatria (AIEOP)¹⁶, Henrik Hasle on behalf of the Nordic Society for Pediatric Hematology and Oncology (NOPHO)¹⁷, Jeffery M. Klco³, James R. Downing³, Jinghui Zhang⁷, Stanley Pounds⁵, C. Michel Zwaan on behalf of the Dutch Children’s Oncology Group (DCOG)²,¹⁵, and Tanja A. Gruber on behalf of St. Jude Children’s Research Hospital Study Group (SJCRH)⁸,¹⁸*

*These authors contributed equally

¹Co-senior authors

¹Department of Cell Biology, Erasmus Medical Center, Rotterdam, Netherlands

²Department of Pediatric Oncology Hematology, Erasmus Medical Center-Sophia Children’s Hospital, Rotterdam, the Netherlands

³Department of Pathology, St. Jude Children’s Research Hospital, Memphis, TN

⁴Pediatric Translational Medicine Institute, Shanghai Children’s Medical Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China

⁵Department of Biostatistics, St. Jude Children’s Research Hospital, Memphis, TN

⁶Department of Oncology, St. Jude Children’s Research Hospital, Memphis, TN

⁷Department of Computational Biology, St. Jude Children’s Research Hospital, Memphis, TN

⁸Department of Pediatrics, Stanford University School of Medicine, Stanford, CA

⁹Women and Child Health Department, Hematology-Oncology Clinic and Lab, University of Padova, IRP, Padova, Italy

¹⁰Department of Internal Medicine Hematology and Oncology Center of Molecular Biology and Gene Therapy, Masaryk University Hospital, Brno, Czech Republic
Department of Pediatric Hematology and Oncology Aghia Sophia Children's Hospital, Athens, Greece

Department of Pharmacotherapy and Translational Research, College of Pharmacy, University of Florida, Gainesville, FL

Division of Pharmaceutics, College of Pharmacy and Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio

Department of Pediatrics, University Hospital Essen, Essen, Germany

Department of Pediatric Oncology, Princess Máxima Center for Pediatric Oncology, Utrecht, Netherlands

Department of Pediatric Hematology Oncology, IRCCS Ospedale Pediatrico Bambino Gesù, Sapienza, University of Rome, Italy

Department of Pediatrics, Aarhus University, Aarhus, Denmark

Stanford Cancer Institute, Stanford University School of Medicine, Stanford CA

Correspondence Addressed To:

Maarten Fornerod, Ph.D. Tanja Andrea Gruber, M.D., Ph.D.
Department of Cell Biology Department of Pediatrics
Erasmus Medical Center Stanford University
Doctor Molewaterplein 40, 3015 GD 1000 Welch Road Suite 300
Rotterdam, Netherlands Palo Alto, CA 94304
m.fornerod@erasusmc.nl tagruber@stanford.edu
Abstract
Genomic characterization of pediatric patients with acute myeloid leukemia (AML) has led to the discovery of somatic mutations with prognostic implications. While gene expression profiling can differentiate subsets of pediatric AML, its clinical utility in risk stratification has remained limited. Here we evaluate gene expression, pathogenic somatic mutations and outcome in a cohort of 435 pediatric patients with a spectrum of pediatric myeloid-related acute leukemias for biologic subtype discovery. This analysis revealed 63 patients with varying immunophenotypes that span a T-lineage and myeloid continuum designated as Acute Myeloid/T-Lymphoblastic Leukemia (AMTL). Within AMTL, two patient subgroups distinguished by \textit{FLT3-ITD} and PRC2 mutations have different outcomes, demonstrating the impact of mutational composition on survival. Across the cohort variability in outcomes of patients within iso-mutational subsets is influenced by transcriptional identity and the presence of a stem cell-like gene expression signature. Integration of gene expression and somatic mutations lead to improved risk stratification.

Statement of Significance
Immunophenotype and somatic mutations play a significant role in treatment approach and risk stratification of acute leukemia. We conducted an integrated genomic analysis of pediatric myeloid malignancies and found that a combination of genetic and transcriptional readouts was superior to immunophenotype and genomic mutations in identifying biologic subtypes and predicting outcomes.
**Introduction**

Acute myeloid leukemia (AML) comprises a heterogeneous group of malignancies that are linked by the presence of blasts displaying morphologic and immunophenotypic features of myeloid cell differentiation. These characteristics served as the initial approach to subdivide AML into distinct clinical entities (1). Morphology and immunophenotype, however, are limited in biologic, prognostic, and therapeutic significance. The identification of cytogenetic alterations and molecular lesions has allowed newer classification schemes to be developed with the most recent widely used approach being the World Health Organization classification of AML (2). Although the latter classification scheme divides AML into many distinct clinical, morphologic and/or molecular subtypes, from a clinical perspective most current therapeutic pediatric protocols stratify patients into favorable, intermediate, and poor prognostic groups (3). Therapy in these groups is based on the relative risk of relapse, with poor prognostic groups proceeding to allogeneic hematopoietic stem cell transplantation in first remission when a suitable donor is available.

With the development of genome-wide gene expression profiling, array-based comparative genomic hybridization methodologies, and next generation sequencing technologies, the field has gained a greater understanding of the molecular features involved in the occurrence of pediatric myeloid malignancies. Several pathologic lesions have been found to have prognostic implications contributing to a continuous refinement of risk stratification over time in the context of modern therapy. We previously applied an integrated analysis to a large cohort of pediatric acute megakaryoblastic leukemia (AMKL) that underwent next generation sequencing with the goal of identifying biologically and clinically relevant subtypes so that we could gain a greater understanding of the biology of the disease as well as inform clinical decision making (4). In that study, using gene expression profiling coupled with somatic variants and outcome data, we were able to identify distinct molecular subtypes with varying outcomes. These results led to a
recommendation to limit high-risk designation to a subset of patients, which has already been instituted in the ongoing multi-institutional AML16 trial for newly diagnosed pediatric AML patients (NCT03164057) and several other collaborative group protocols. Here we apply a similar approach to a cohort of 435 pediatric patients with a spectrum of myeloid-related malignancies to provide a comprehensive view of this clinical entity and propose a refined classification scheme with clinical utility. Using this approach, we identify a previously undescribed subtype that spans a T-lineage and myeloid continuum, as well as new prognostic mutational events within previously described subtypes. Further, we demonstrate that mutational events, transcriptional profile, and evidence of a primitive hematopoietic progenitor gene expression signature all associate independently with outcome. The most significant association occurs when all three of these factors are combined, arguing in favor of subgroup classification by comprehensive molecular profiling to optimize risk stratification in pediatric AML.

**Results**

*Genomic Landscape of Normal and Complex Karyotype Pediatric AML*

The Children's Oncology Group (COG)–National Cancer Institute (NCI) TARGET AML initiative molecularly characterized 993 pediatric AML cases including 197 specimens that underwent comprehensive whole genome sequencing(5). Of these, 94 carried one of three oncogenic fusions known to be strong drivers of leukemogenesis: \textit{RUNX1-RUNX1T1}, \textit{CBFB-MYH11} and \textit{KMT2A} rearrangements (\textit{KMT2Ar}). Among all other somatic alterations detected, only ten occurred in more than 5% of subjects, all of which had been previously described. This suggested that low-frequency molecular subsets may exist that require larger cohorts to fully elucidate. To address this limitation, we selected 122 pediatric AML normal, non-complex, and
complex karyotype specimens from five cooperative study groups (SJCRH, DCOG, NOPHO, AIEOP, and BFM) that lacked RUNX1-RUNX1T1, CBFB-MYH11 and KMT2A by clinical testing for whole genome (WGS) and/or exome (WES) and RNA (RNAseq) sequencing to enrich for cases that carry low-frequency events (Supplementary Tables 1,2 and Figure 1A). Structural variations (SV), copy number alterations (CNA), single nucleotide variations (SNV) and indels were determined by our established pipelines, as well as an evaluation for regulatory rearrangements driving oncogene overexpression through enhancer hijacking (Supplementary Tables 3-9 and Supplementary Figures 1,2)(6). When considering exonic SNV/Indels, CNA, and SV calls mutational burden ranged from 1 to 101 somatic events including a case with TP53 associated chromothripsis that carried 89 lesions in total (Supplementary Table 9 and Supplementary Figures 1,3). In addition to known AML somatic mutations in genes such as CEBPA, GATA2, NPM1, WT1, FLT3, NRAS, KRAS, ETV6, RAD21, SMC1A, STAG1, STAG2, STAG3, SMC3, and rearrangements in NUP98 and KAT6A, we identified rare events in known oncogenic drivers. These include internal tandem duplications (ITD) in GATA2, RUNX1, and CEBPA, as well as the repositioning of a distal ZEB2 enhancer, MYC enhancer, or ETV6 enhancer to ectopically activate BCL11B, MECOM, and MNX1 loci respectively (Supplementary Table 6, Supplementary Figures 4,5). Interestingly, 15 AML cases (12.3%) carrying loss of function mutations in polycomb repressive complex 2 (PRC2) genes were found to resemble an early T-cell precursor acute lymphoblastic leukemia (ETP-ALL) gene expression profile (GEP) by gene set enrichment analysis (Supplementary Figure 6). ETP-ALL exhibits aberrant expression of stem cell and myeloid markers and has been shown to have a GEP consistent with transformation of a stem cell progenitor(7,8). Further, mixed phenotype acute leukemias (MPAL) with T and myeloid lineage characteristics have been previously suggested to be in this spectrum of immature leukemias(9). We therefore hypothesized that these PRC2-mutated AML cases represented the myeloid end of this continuum. To provide global transcriptional context to these ETP-like AMLs and evaluate a comprehensive cohort encompassing a range of
pediatric myeloid malignancies, we integrated results from previously published AML (N=169), MPAL (N=80), AMKL (N=45), and ETP-ALL (N=19) datasets that had RNAseq and either WES or WGS available for a total of 435 cases (Supplementary Table 10 and Figure 1A)(4,5,7-9).

**Molecular Classifier of Pediatric Myeloid Malignancies Agnostic of Immunophenotype**

t-SNE visualization using a 381 gene list derived from the top 100 most variably expressed transcripts within each of the five sequencing datasets revealed a clear molecular classifier, identifying groups that had consistent mutational compositions but were agnostic of immunophenotype (Figure 1B,C and 2, Supplementary Table 10-13 and Supplementary Figure 7). A bootstrap hierarchical clustering procedure defined subgroups with an overall reproducibility of 97.4% and highly concordant with the tSNE transcriptional subgroups (Adjusted Rand Index=0.72, Supplementary Table 14); indicating the subgroups identified by tSNE are statistically meaningful. This classifier allowed the distinction of 63 cases with an ETP-ALL GEP comprising a mixture of AML (N=12/63, 19%), AUL (N=1/63, 1.6%), MPAL (N=31/63 49.2%), and ETP-ALL (N=19/63, 30.2%) leukemias (bootstrap reproducibility=93.6%, Figure 1B). All but one MPAL case within this subgroup co-expressed T lineage antigens in addition to either myeloid and/or B lineage antigens (Figure 1B and Supplementary Table 10). Expression of *MPO* and *CD3E* confirmed that the reported immunophenotypes of these cases were correct (Supplementary Figure 8). A separate validation cohort of 399 pediatric AML cases with microarray data confirmed the presence of this entity with 23 cases identified (Supplementary Figure 9 and Supplementary Tables 15 and 16). A five gene classifier consisting of *CD3G*, *COCH*, *SLC35D2*, *SPTLC3*, and *TOR4A* was able to predict these cases in both the discovery and validation cohorts (area under the curve 0.977 and 0.88 respectively, see Supplementary Table 27). A molecularly distinct subtype of acute leukemia termed acute myeloid/T-lymphoblastic leukemia (AMTL), with shared myeloid and T-lineage features has been previously proposed by Gutierrez and Kentsis(10). In support of this entity, they noted shared
gene mutations in prior sequencing reports of T-lineage and AML studies including WT1, PHF6, RUNX1, and BCL11B. Consistent with this, transcriptionally defined AMTL cases in our discovery cohort carried mutations in these genes and were found to fall into one of two subgroups: a group characterized by FLT3-ITD (N=26/63, 41.3%) and a second group enriched for loss of function alterations in one of three core PRC2 complex genes including EZH2, SUZ12, and EED or a splicing factor mutation that leads to inclusion of a cryptic exon resulting in truncated EZH2 transcripts predicted to undergo nonsense mediated decay (N=37/63, 58.7%) (Figure 3A and Supplementary Figure 10)(11). Both subsets were found to carry cooperating events in transcription factors (WT1, NOTCH1, ETV6, PHF6, RUNX1, IKZF1, BCL11B TLX3); unique to PRC2 cases were activating events in RAS (NRAS, KRAS, NF1) and JAK/STAT (JAK1, JAK3, IL7R, SH2B3) signaling cascades, as well as loss of function mutations in genes that play a role in G1 checkpoint arrest (RB1, CCDN3, CDKN1B, CDKN2A/B) (Figure 3A). In particular, network analyses identified a strong association between transcription factors associated with T-lineage differentiation (NOTCH1, PHF6, BCL11B, TLX3, TAL1, IKZF2), PRC2 loss of function mutations, and JAK/STAT pathway alterations whereas FLT3-ITD cases were enriched for RUNX1 and WT1 transcription factors (Supplementary Figure 11)(12). A comparison of overall survival clearly demonstrated that outcomes of the iso-transcriptional AMTL subset are influenced by the mutational spectrum. Irrespective of whether the patient received AML, ALL, or a hybrid treatment approach, FLT3-ITD positive AMTL were associated with a favorable outcome, whereas those with PRC2 mutations had a poor prognosis (p=8x10^{-4}, Figure 3B, Supplementary Table 10). Consistent with this, AMTL cases in our AML validation cohort for which mutational data was available (N=16/23, 69.6%) were similarly comprised of FLT3-ITD positive (N=8/16, 50%) and negative cases (N=8/16, 50%); a subset of the negative cases (N=3/8) had copy number data available which confirmed deletional events in PRC2 genes in all three cases and an association with poor outcomes (p=0.01, Supplementary Figure 12 and Supplementary Table 16). PRC2 loss of function mutations were also present in a
subset of core binding factor cases (N=8/61, 13.1%) (Supplementary Table 12). To determine if the presence of PRC2 mutations confers a poor prognosis in these patients as well, we evaluated outcomes in pediatric core binding factor AML cases from two previously published cohorts and found an inferior event free survival in patients carrying both KIT activating mutations and PRC2 loss of function mutations (N=5/142, 3.5%) (p=0.026 Supplementary Table 17 and Supplementary Figure 13)(5,13). In alignment with these data, prior studies have shown chemoresistance as a result of PRC2 loss in AML and T lineage ALL models(14,15).

Outcomes of Iso-Mutational Subsets are Influenced by Transcriptional Identity

The favorable prognosis of AMTL cases carrying FLT3-ITD included those with cooperating WT1 mutations, several of which were classified as AML by immunophenotype (N=10/26, 38.5% of FLT3-ITD AMTL cases carried WT1 mutations, two of which were AML). Historically pediatric AML patients with FLT3-ITD and a WT1 mutation have been reported to have a dismal prognosis(16). A significant number of these FLT3-ITD/WT1 double mutant cases were also found to associate within a different transcriptional cluster, AML MK-V (N=14/25, 56% in MK-V; N=10/25, 40% in AMTL; N=1/25, 4% in MK-IV; Figure 1B,C). In contrast to AMTL, FLT3-ITD/WT1 double mutant patients that fell into AML MK-V transcriptional cluster had an extremely poor outcome consistent with prior reports (Figure 3C). Thus, the presence of these somatic events alone is insufficient to distinguish high risk status. A comparison of differentially expressed genes between AMTL FLT3-ITD/WT1 and AML MK-V FLT3-ITD/WT1 identified significant upregulation of genes within the HOX locus in AML MK-V cases (Figure 3D). Although the mutational spectrum is known to influence the transcriptional profile of leukemia, the cell that acquires the mutations (“cell of origin”) may also be reflected. To look at this further we evaluated expression of the HOX locus in a normal hematopoietic progenitor dataset and found elevated expression of the HOX genes upregulated in our AML MK-V FLT3-ITD/WT1 patients in both hematopoietic stem cell (HSC) and common myeloid progenitor (CMP).
compartments compared to lymphoid restricted progenitors (LP) (Figure 3D), suggesting the differential HOX expression between the two subsets may reflect a stem cell like state(17).

We therefore identified gene expression signatures for the different hematopoietic subsets and looked for enrichment of those signatures in our two subsets of FLT3-ITD/WT1 patients to determine whether the correlation of stem cell associated genes extended beyond the HOX locus. This analysis confirmed an enrichment in our AML MK-V cluster cases for HSC as well as CMP signatures in contrast to AMTL cases that have a greater enrichment for LP signatures (Figure 3E). We hypothesize that this reflects a more primitive cell of transformation in AML MK-V FLT3-ITD/WT1 cases that retain a stem cell progenitor like state contributing to chemotherapy resistance. To assess whether this phenomenon is restricted to FLT3-ITD/WT1 genotypes, we applied the same analysis to KMT2Ar cases that fell into AML MK-V and the 11q23 rearranged transcriptional cluster (Figure 1C, Supplementary Figure 9). Consistent with the inferior event free survival (EFS) of AML MK-V KMT2Ar cases compared to those in the 11q23 rearranged transcriptional cluster in both discovery and validation cohorts, we found a more pronounced enrichment for HSC and CMP signatures in AML MK-V KMT2Ar suggesting a more primitive stem cell like state (Supplementary Figures 14-16).

Leukemia Stemness is Unevenly Distributed Across Myeloid Leukemias

Ng and colleagues previously developed a 17 gene transcriptional score related to stemness, derived from functionally defined leukemia stem cells of adult AML patients, which was predictive of prognosis (LSC17)(18). More recently, a six-gene LSC score has been developed with significant prognostic value in pediatric AML (pLSC6)(19). To determine if the more primitive nature of AML MK-V FLT3-ITD/WT1 cases was reflected in this score we compared pLSC6 in AML MK-V and AMTL FLT3-ITD/WT1 patients (Figure 3F). Consistent with enrichment of more primitive hematopoietic progenitor gene expression signatures, AML MK-V
FLT3-ITD/WT1 patients had a higher pLSC6 score (p=0.038). To evaluate this more comprehensively across the cohort, we determined the pLSC6 score in normal hematopoietic progenitor subsets to define thresholds of low (lineage committed cells), intermediate (multipotent progenitors), and high values (pluripotent progenitors) (Figure 4A-B). Imposing these thresholds on our cohort, we identified a subset of patients with intermediate and high scores which was significantly associated with an inferior overall survival (N=302/435 69.4% low pLSC6, N=119/435 27.4% intermediate pLSC6, N=14/435 3.2% high pLSC6 p=9.3x10⁻⁷ discovery cohort; and N=262/399 65.7% low pLSC6, N=124/399 31.1% intermediate pLSC6, N=13/399 3.2% high pLSC6 p=2.1x10⁻⁶ validation cohort) (Figure 4C, Supplementary Figure 17). While several subsets had uniform pLSC6 scores, such as CBFA2T3-GLIS2, RUNX1-RUNXI1T1, CBFB-MYH11, and MNX1 rearranged cases, other subsets had variable scores demonstrating heterogeneity in leukemia “stemness” (for example KMT2Ar cases), highlighting pLSC6 as an independent variable in addition to mutational type and overall transcriptional signature (Figure 4D, Supplementary Table 17 and Supplementary Figures 18-20).

Transcriptional Identity, Mutations, and Stemness All Contribute to Outcome

To evaluate the relative contribution of each of the factors identified in our study to carry an association with survival, we utilized a Cox proportional hazard model to look at associations with overall survival. Transcriptional identity, oncogenic drivers and leukemia stemness were all independently found to associate with outcome (Figure 4C, 5A, Supplementary Table 19 and Supplementary Figures 17, 19-20). The greatest association occurred when all three of these factors were combined (p=1.06x10⁻¹² discovery cohort and p=1.19x10⁻⁷ validation cohort). The impact of individual factors on outcome associations was variable in our discovery cohort, with CBFA2T3-GLIS2, ETS family rearrangements (FUS-ERG, EWSR1-ERG, FUS-FEV, FUS-FLI1, MN1-FLI1, and EWSR1-FEV), and high pLSC6 score having the greatest negative association with outcome, while CEBPA mutations (mono and biallelic) and low pLSC6 carried the greatest
positive association with outcome (Supplementary Tables 20 and 21 and Supplementary Figure 21). Within biologic subgroups identified in pediatric AML certain factors carried greater weight than others (Table 1). Utilizing these rules for risk stratification, we compared outcomes in our discovery and validation cohorts for our proposed genomic classification (low, intermediate, and high risk) to those of the ongoing multi-institutional AML16 prospective clinical trial for newly diagnosed pediatric AML patients (NCT03164057) (Figure 5B-C validation cohort; Supplementary Figures 22-25 discovery and combined cohorts; and Supplementary Tables 16, 22 and 23). For a given risk classification, we defined and computed the risk classification utility (RCU) which considers estimate outcomes for each risk group (outcome discrimination index) and the proportion of patients designated as high or low risk given that intermediate risk designates a patient lacking definitive high or low risk characteristics and thus represents a patient whose status is unknown (Supplementary Table 24). A bootstrap procedure was then used to quantify the statistical variability and significance of comparisons of the RCU with the two classification schemes (Supplementary Table 25). In both the discovery and validation cohorts as well as in a combined analysis, our proposed classification was found to have a statistically significant greater RCU for EFS than AML16 (p=0.036 discovery cohort, p=0.018 validation cohort, and p=0.036 combined cohorts) (Figure 5C and Supplementary Figure 25). In particular, the proposed classification was superior at identifying high risk patients within the intermediate and low risk groups resulting in a lower proportion of intermediate risk patients that had an improved EFS which brings the proposed stratification closer to the ideal state – one in which there are only two risk groups: patients that have an event (high risk) and those that do not (low risk) (Figure 5B).

Discussion
Gene expression, genomic classification, and leukemia stemness have all been shown to impact prognosis to varying extents in both adult and pediatric AML(5,18-23). Few studies to date, however, and none in pediatric AML, integrate all three of these aspects to determine the relative contribution to outcomes. Through this comprehensive approach and by including pediatric acute leukemias with myeloid characteristics we were able to identify a previously undescribed subtype, AMTL, which spans a T-lineage and myeloid continuum as well as new prognostic mutational events within previously described subtypes such as PRC2 mutations in core binding factor leukemias. Recently two groups have reported on acute leukemias with T-lineage markers such as cytoplasmic CD3 and/or CD2 that carry BCL11B enhancer hijacking events similar to several cases within the AMTL subgroup (24,25). Unique to our study is the identification of AMTL cases that are devoid of T-lineage markers by flow cytometry and the distinction of the two subsets within AMTL that have differing outcomes. It has been previously shown through murine modeling that T-lymphoid progenitors retain a broad lineage potential when transformed with oncogenes, and specifically have the ability to differentiate into myeloid leukemia while retaining a lymphoid epigenetic memory, consistent with our findings(26). In this study by Riemke and colleagues, a cohort of adult AML patients were found to resemble the murine T-lymphoid progenitor derived myeloid leukemias by gene expression. This population however had a negative association with ETP-ALL by GSEA and the mutation profile of these patients was predominated by mutations not found in pediatric AMTL including NPM1, IDH2, and DNMT3A. This difference may be a result of distinct oncogenic events that are acquired by a T-lymphoid progenitor as opposed to a difference in the cell of origin.

The existence of patients with FLT3-ITD/WT1 in AMTL that had superior outcomes, in contrast with previously published results, led us to compare outcomes of these patients across transcriptional subsets. The inferior overall survival of FLT3-ITD/WT1 double mutant patients was restricted to those within the MK-V cluster. Of note, the vast majority of patients within this
study were treated prior to the implementation of FLT3 inhibitors (2/291 AML patients in the discovery cohort for whom treatment details were known received a FLT3 inhibitor at diagnosis, both of whom had events and are deceased; Supplementary Table 10). While we cannot determine whether FLT3 inhibition would improve outcomes of MK-V FLT3-ITD/WT1 patients in our study, results from COG AAML1031 suggest that this targeted treatment approach can improve outcome in FLT3-ITD/WT1 patients with the caveat that the transcriptional identity in this study are unknown(4). The absence of FLT3 inhibition in our cohort allowed us to identify iso-mutational groups where disease outcome clearly associates with transcriptional identity and iso-transcriptional groups where outcome clearly associates with mutational status. This finding has broad implications on variant interpretation in the era of precision medicine as the impact on prognosis is not limited to the presence or absence of a given mutation. Furthermore, the incorporation of stem cell associated signatures also allowed us to distinguish patients that have the same genomic classification but differing outcomes (Table 1). The highest power and outcome associations occur when all three of these factors are combined arguing in favor of comprehensive diagnostics to optimize risk stratification in pediatric AML. A multivariate analysis to evaluate the prognostic informativeness of WT1 and FLT3-ITD mutational events after considering transcriptional identity, key driver mutation, and pLSC6 score support this conclusion: neither EFS nor OS was significantly associated with the presence of FLT3-ITD, a WT1 alteration, or the combination of these two after adjustment for pLSC6 score as a numeric predictor, transcriptional identity as a stratification factor, or driver mutation as a stratification factor (Supplementary Figure 26, Supplementary Table 26). Further, neither EFS nor OS was significantly associated with FLT3-ITD, WT1, or the presence of both in models that considered only these variables as predictors (Supplementary Table 26).

The benefit of risk-adapted indications for hematopoietic stem cell transplantation in pediatric AML has recently been shown by the BFM study group, with significantly higher event-free
survival and higher rates of hematopoietic stem cell transplants through improvements in genetic risk stratification (27). In a disease entity where the chemotherapy approach has remained largely unchanged over time with a limited number of novel therapeutic agents on the horizon, risk stratification, refined allograft indications and supportive care continue to be major factors that have led to the improvement in outcome over time (28). It is therefore imperative that risk stratification be optimized to the maximum extent in order to cure more pediatric AML patients. The vast majority of pathogenic calls and transcriptional information necessary to use our integrated approach can be obtained from paired whole exome and RNA sequencing which has been increasingly adopted in the clinical setting, arguing in favor of the feasibility of this approach (29-31). Targeted capture panels that detect SNV/Indels and copy number changes in combination with a fusion detection assays are less comprehensive but also able to detect the vast majority of oncogenic lesions described in this study. In pediatric AML, all patients enrolled on the St. Jude AML16 study are already receiving CLIA certified whole genome, exome, and RNA sequencing on diagnostic blasts. While next generation sequencing approaches are becoming increasingly standardized and prevalent in the field, bioinformatic analyses and interpretation of mutational impact within a case based on transcriptional identity and leukemia stemness will require additional expertise to implement. To enhance the clinical applicability of this study we developed a panel of five genes whose expression can distinguish AMTL cases that can be combined with the previously developed six gene pLSC6 classifier - key determinants in our risk stratification model. In combination with key mutational events this allows one to follow a hierarchical decision-making tree to stratify a patient (Figure 6).

The cell of origin of leukemia is defined as the normal hematopoietic cell from which the disease develops through the acquisition of mutations. A subset of cells termed leukemia stem cells are felt to propagate the disease over time and studies have shown that similar to normal hematopoiesis, a hierarchical structure exists in leukemia with the most primitive clone being
identifiable through functional assays(32). Given the differentiation spectrum seen in leukemias, it can be a challenge to infer the cell of origin in bulk tumor populations. Despite this potential limitation, we found significant enrichment of specific hematopoietic progenitor cell signatures within subsets of patients that correlated with the functionally derived LSC score. Specifically, greater enrichment for more primitive progenitor cell signatures was found in patients with higher LSC6 scores. Our data are consistent with a model whereby a cell of origin acquires oncogenic driver mutations and these two factors both contribute to the transcriptional identity of the leukemia and the stemness all of which influence outcome (Figure 5D).

In summary comprehensive next generation sequencing of pediatric AML can be utilized beyond pathogenic mutation calls to optimize risk stratification. Incorporation of transcriptional identity and leukemia stemness in clinical decision-making will further improve the identification of patients that may benefit from stem cell transplant in first remission and those that can be cured with chemotherapy alone.

**Methods**

**Cohort**

Specimens sequenced in this study were provided from multiple institutions and collaborative groups. All samples were obtained with patient or parent/guardian provided written informed consent under protocols approved by the Institutional Review Board at each institution. Studies were conducted in accordance with the International Ethical Guidelines for Biomedical Research Involving Human Subjects. Samples were deidentified prior to nucleic acid extraction and analysis. Whole genome, whole exome, RNA sequencing and analysis for structural variations, single nucleotide variations, indels, and copy number alterations were performed as previously described(4,7). TARGET AML, Early T-Cell Precursor, Mixed Phenotype Acute Leukemia, and Acute Megakaryoblastic Leukemia cohorts have been previously published and were obtained
with permission from dbGAP and/or St. Jude Children’s Research Hospital(4,5,7-9). Transcript expression levels for gene expression analyses were estimated from RNA sequencing data as Fragments Per Kilobase of transcript per Million mapped fragments (FPKM) as previously described(4). Data for samples sequenced in this study has been deposited to the St. Jude Cloud (www.stjude.cloud) and EGAS00001004701(33).

**RNA-Seq read mapping, gene expression summary and batch correction**

RNA reads were mapped using our StrongARM pipeline, described previously.(13) Paired-end reads from RNA-seq were aligned to the following four database files using BWA: (i) the human GRCh37-lite reference sequence, (ii) RefSeq, (iii) a sequence file representing all possible combinations of non-sequential pairs in RefSeq exons and, (iv) the AceView database flat file downloaded from UCSC representing transcripts constructed from human ESTs. Additionally, they were mapped to the human GRCh37-lite reference sequence using STAR. The mapping results from databases (ii)–(iv) were aligned to human reference genome coordinates. The final BAM file was constructed by selecting the best of the five alignments.

Reads from aligned bam files were assigned to genes and counted using HTSeq with the GENCODE human release 15 gene annotation.(34) The gene count matrix was used to generate FPKM (fragment per kilobase per million) gene expression data matrix using gene length information. A gene was called as “expressed” in a given sample if it had a FPKM value \( \geq 0.01 \) based on the distribution of FPKM gene expression values and genes not expressed in any sample were excluded from downstream analysis. The gene expression data were further quantile-normalized using the normalizeBetweenArrays function available from the Limma R package.(35) The detected batch effect due to data source of St. Jude vs. TARGET was corrected using the ComBat method available from the R package SVA.(36)

**381 Gene Classifier**
For construction of the 381 gene classifier, the top-100 most variant genes from each of the 5 datasets (this manuscript, ETP ALL, MPAL, TARGET AML, and AMKL) were combined using log2 transformed FPKM values and median adjusted deviation. This procedure effectively eliminated remaining batch effects (Supplemental Figure 27). Visualization was performed using t-SNE using a perplexity value of 10 and 10,000 iterations. t-SNE coordinates from the run with the lowest final error (out of ten runs) were selected for further analysis.

**Hematopoietic Stem Cell Progenitor Gene Expression Analysis**

Single cell hematopoietic stem cell progenitor (HSCP) counts, SPRING plot coordinates and population assignments were taken from Pellin et al. For comparing HSCP and leukemia gene expression, single cell counts per gene were summed up for each of the 11 different HSCP populations, normalized to the number of cells in each population and log2 transformed. Resulting gene expression values were scaled together with log2FPKM expression values of the 435 leukemias using the normalize between arrays function of limma (method quantile). pLSC6 scores and spearman correlation coefficients were calculated using these values. For some analyses multi-lymphoid progenitors (MLP) and pre-B/NK values were averaged to generate lymphoid progenitor (LP) values. pLSC6 high, medium and low cutoff values were based on HSCP population values, with the most primitive populations designated as high (populations 1, 2, 3, 7, 9, 10 and 11 from Pellin et al.) and more committed populations designated medium (populations 4,5,6, and 8), and values lower than these low. Exact cutoff values were calculated using linear extrapolation.

**Statistical Analysis**

All analyses were done in R. Survival and global test analyses were performed as previously described. Treatment details for patients are included in Supplementary Tables 1 and 10.
The integrative statistical model was evaluated using the Global Test assuming interaction between the explanatory variables (38). Transcriptional identity and key oncogenic driver were defined as categorical and leukemia stemness (pLSC6) as a continuous variable, and assuming interaction between these three exploratory variables. Individual associations are shown in Supplementary Figure 18, and main contributing covariates clarified in Supplementary Table 15 by using pLSC6 as a categorical variable (low versus medium/high).

Validation Cohort

A pediatric AML microarray gene expression cohort of 443 cases was constructed based on previously published data (39-41). AML M5 cases with t(15;17) were excluded from this cohort prior to assembly because this subclass was absent from the discovery cohort and has excellent therapy options and disease outcome. Of these, 44 were also included in the discovery cohort and functioned as controls for the equivalence of the RNAseq and microarray measured gene expression. 399 cases which did not overlap were used for gene expression validation of results obtained in the discovery cohort. For 386 of these cases, disease outcome data were available (Supplementary Table 15), and were used for outcome validation analyses.

Key oncogenic driver determination was based on a combination of clinical testing and/or laboratory testing from the cohorts as previously published (see Supplementary Table 15 column K). Cases in which mutational status were unknown were removed from analyses as appropriate.

Transcriptional identity of validation cohort cases was determined by co-clustering of microarray mRNA expression values of overlapping classifier genes (n=249) of single cases with the complete RNA seq cohort using Spearman correlation distance-based t-SNE, exactly as done for the RNA seq cohort clustering. For overlapping genes, probe sets with highest specificity and selectivity (www.geneannot.com) were used, omitting probe sets recognizing more than
one gene. For robustness assessment of transcriptional identity calls we made use of the stochastic initial seeding of the t-SNE algorithm by performing 10 clustering repeats. Cases with clustering inconsistency in more than 2 of the 10 runs (25/443 cases, 5.6%) were not assigned a transcriptional identity label. Transcriptional identity of 95% (41/43) of microarray profiled cases also present in the RNA seq cohort were identical. In 9/327 cases, the transcriptional identity calls were inconsistent with oncogenic driver determination (2.8%), similar to the discovery cohort.

Transcriptional identity was further confirmed by clustering of the validation cohort using a classifier derived from the microarray expression values only. For this, the batch effect of AML02 and Rotterdam cohort expression values was removed using ComBat function of the sva R package (Supplementary Figure 11A). Clustering visualization was by t-SNE using a 350 gene set consisting of the highest variant probe sets by least median square (Supplementary Figure 11B), where only probe sets recognizing single genes were used and sex specific and hemoglobin genes were removed.

pLSC6 scores of the validation cohort were calculated as previously described using log2 intensity values of Affimetrrix probe sets 209543_s_at (CD34), 220668_s_at (DNMT3B), 220377_at (FAM30A), 212070_at (GPR56), 203373_at (SOC2), and 206310_at (SPINK2). (39) Of the 44 cases with both microarray and RNA Seq data, pLSC6 values were highly correlated (r = 0.82). pLSC6 categories of low, medium and high were determined by matched RNA Seq expression value pLSC6 quantiles (0%, 66.21%, 96.78% and 100%). 82% of overlapping cases (36 / 44) were assigned the same pLSC6 category using this method.

In the validation cohort, association between Transcriptional identity, Oncogenic driver, pLSC6 score and overall survival was modeled using a cox regression implementation in the global test, accounting for interactions between the three variables (Supplementary Table 19). 293
cases had overall survival data and could be assigned both a transcriptional identity and oncogenic driver label. Sparse Transcriptional identities (3 cases or less) were removed, leaving 8 Transcriptional identity and 14 Oncogenic driver covariates, while pLSC6 was used as a continuous variable. Main covariates contributing (cases >1) to the global association are reported (Supplementary Table 20), with pLSC6 categorized as Medium/High versus Low. Because pLSC6 was developed using the AML02 validation cohort, association with overall survival was independently assessed excluding the AML02 cases from the validation cohort (Supplementary Figure 19).

**AMTL Five Gene Classifier**

A five gene classifier to identify the AMTL subtype was developed as follows. First, using the RNA-seq cohort, the expression of each gene was summarized by computing median expression for each transcriptional subgroup and using the Wilcoxon test to compare medians across each pair of subgroups. The genes for which AMTL had the greatest or least median expression were selected and then ranked by the maximum of the Wilcoxon test p-values comparing AMTL to other subgroups. The top 14 genes in this list were then considered as candidate predictor variables for a logistic regression predicting the AMTL versus non-AMTL class using the bestglm procedure in R. The bestglm procedure defined the model as logit(Prob(AMTL)) = -0.78 + 1.01 × CD3G -0.85 ×x COCH-1.20 SLC35D2 + 0.81 SPTLC3 – 0.93 TOR4A (Table A). The model classified AMTL with an area under the curve of 0.977. In 1,000 rounds of leave out 10% cross-validation, this model building procedure (median calculation, pairwise Wilcoxon tests, bestglm) achieved an average AUC of 0.973 with a range of 0.952 to 0.983. See Supplementary Table 27 for AMTL logistic regression classifier model terms, estimates, confidence intervals, and p-values. We then went on to validate this five gene classifier in our validation cohort. The Affymetrix microarrays (U133 v2.0) included 6 probe-sets that measured the expression of the five genes in the classifier (gene symbol, probe set IDs:
COCH, 205229_s_at; CD3G, 206804_at; SLC35D2, 213082_s_at; SLC35D2, 213083_at; TOR4A, 219620_x_at; and SPTLC3, 220456_at).

A principal components analysis of the two probe-sets measuring SLC35D2 gave similar coefficients for 213082_s_at (0.74) and 213083_at (0.67). Thus, for each subject, the expression of SLC35D2 was computed as the simple arithmetic average of the expression of these two probe-sets. The other four genes were measured by one probe-set each. For each subject, a score was computed as the dot product of the microarray expression of the five genes with the coefficients from the RNA-seq cohort’s logistic regression model. This score classified the AMTL/non-AMTL in the independent microarray cohort (those without RNA-seq data) with an AUC of 0.88.

Risk Classification Utility

For a given risk classification, censored event time endpoint (such as event-free survival or overall survival), and cohort outcome data set, we defined and computed the risk classification utility (RCU) as follows. We computed the proportion of patients assigned to low, intermediate, and high-risk groups and the Kaplan-Meier estimates of outcome for each risk group (Figure 4B-C, Supplementary Figures 24 and 25). Then, for each observed event-time, we plotted the utility curve as Kaplan-Meier survival estimate of the low-risk group versus that of the high-risk group (Figure 4C, Supplementary Figure 26). An ideal utility curve is a flat line at y=1; in this case, there is some time-point at which the Kaplan-Meier estimate of high-risk patients is 0 and that of low risk patients is 1. A utility curve along the line y=x could reasonably be obtained by completely random assignment of patients into low or high-risk groups. The outcome discrimination index was defined and computed as twice the area above the line y=x and below the utility curve. The outcome discrimination index is 1 if the utility curve is ideal and 0 if the utility curve does not have any point above the line y=x that can be obtained by random risk classification assignments. We defined and computed the meaningful classification proportion
as the proportion of patients designated as high or low risk because intermediate risk typically designates a patient lacking definitive high risk or low risk characteristics (Figure 4C, Supplementary Figure 27). Finally, the risk classification utility was defined and computed as the product of the meaningful classification proportion and the outcome discrimination index (Figure 4C, Supplementary Figure 27). The risk classification utility (RCU) equals 1 if and only if all patients have a meaningful classification and the outcome discrimination is 1.

A bootstrap procedure was used to quantify the statistical variability and significance of comparisons of RCU of four risk classification schemes. The RCU of each risk classification scheme was computed for the discovery cohort and 100,000 bootstraps of the discovery cohort, the validation cohort, and the combined cohort was determined (Supplementary Table 25).
Acknowledgements

The authors thank St. Jude Tissue Resources Laboratory, Flow Cytometry and Cell Sorting Core, and the Hartwell Center for Biotechnology and Bioinformatics. This work was supported by grants from the American Cancer Society (T.A.G. RSG-16-046-01), Hyundai Hope on Wheels (T.A.G.), Dutch Cancer Society KWF (M.F.), KiKa Children Cancer-free Foundation (S.N.), Fondazione Airc (Associazione Italiana Ricerca sul Cancro) IG 20562 and CARIPARO grant 17/04 (M.P.), the National Institutes of Health (J.K.L., S.P., R01-CA132946) and the American Lebanese Syrian Associated Charities (ALSAC) of St. Jude Children’s Research Hospital.

Author Contributions


Competing Interests

The authors have no competing interests.

Materials and Correspondence

Correspondence and material requests should be addressed to Tanja A. Gruber and Maarten Fornerod.
References


<table>
<thead>
<tr>
<th>Subtype</th>
<th>Immunophenotypes Across the Entire Cohort (N)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Proposed Risk Status Based on Overall Survival (Reference)&lt;sup&gt;b, c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>AML(12), MPAL(30), AUL(1), ETP(19)</td>
<td>FLT3-ITD mutation present: LOW PRC2 mutation present: HIGH</td>
<td></td>
</tr>
<tr>
<td>AML(28), MPAL(3)</td>
<td>pLSC6 medium or high: HIGH pLSC6 low: LOW (42)</td>
<td></td>
</tr>
<tr>
<td>AML(27)</td>
<td>PRC2/Kit double mutant present: HIGH PRC2/Kit double mutant absent and pLSC6 low: LOW (43)</td>
<td></td>
</tr>
<tr>
<td>AML(34)</td>
<td>PRC2/Kit double mutant present: HIGH PRC2/Kit double mutant absent and pLSC6 low: LOW (43)</td>
<td></td>
</tr>
<tr>
<td>AML(3), AMKL&lt;sup&gt;d&lt;/sup&gt;, AUL(1)</td>
<td>HIGH (44)</td>
<td></td>
</tr>
<tr>
<td>AML(10), AMKL(1) MPAL(2)</td>
<td>HIGH (45)</td>
<td></td>
</tr>
<tr>
<td>AML(2), AMKL(11)</td>
<td>HIGH (4,46)</td>
<td></td>
</tr>
<tr>
<td>AML(56), AMKL(10), MPAL(9), AUL(2), ETP(1)</td>
<td>MK-V: HIGH AMKL: HIGH (4) pLSC6 medium or high: HIGH MK-V absent and AMKL absent, and pLSC6 low: INTERMEDIATE (43)</td>
<td></td>
</tr>
<tr>
<td>AML(3), AMKL(6), MPAL(1)</td>
<td>pLSC6 medium or high: HIGH pLSC6 low: LOW (4)</td>
<td></td>
</tr>
<tr>
<td>AML(2), AMKL(13)</td>
<td>pLSC6 medium or high: HIGH pLSC6 low: LOW (4)</td>
<td></td>
</tr>
<tr>
<td>AML(17), AMKL(6), ETP(3)</td>
<td>HIGH (4,47-49)</td>
<td></td>
</tr>
<tr>
<td>AML(25)</td>
<td>pLSC6 medium or high: HIGH pLSC6 low: LOW (50,51)</td>
<td></td>
</tr>
<tr>
<td>AML(5)</td>
<td>HIGH (52)</td>
<td></td>
</tr>
<tr>
<td>AML(28), MPAL(18), ETP(6)</td>
<td>WT1 and MK-V present: HIGH pLSC6 medium or high: HIGH pLSC6 low and AMTL absent: INTERMEDIATE</td>
<td></td>
</tr>
<tr>
<td>AML(39), AMKL(6), MPAL(24)</td>
<td>pLSC6 medium or high: HIGH pLSC6 low: INTERMEDIATE (4)</td>
<td></td>
</tr>
</tbody>
</table>

*aNumber in parentheses indicate the number of cases across the discovery cohort with indicated immunophenotype. Genomic subtypes not identified in AML cases are not included in this table.

*bOutcomes approaching 80% overall survival or greater are designated as low risk, survival less than 40% were designated as high risk. Literature support of previously described subtypes and risk status are indicated in parentheses.

*cMinimal residual disease is considered an independent risk factor and residual levels of disease following induction chemotherapy warrants escalation of risk status.

*dReported in the literature<sup>(53)</sup>

*eFLT3-ITD cases that are NOT included in the other subtypes (see Supplementary Figure 26).
Figure Legends

Figure 1. Transcriptional Identities Correlate with Key Oncogenic Driver Events and are Agnostic of Immunophenotype. (A) Study Design. 122 normal, non-complex, and complex karyotype pediatric specimens were selected. Exclusion criteria for sequencing include FAB M3 (APML), FAB M7 (AMKL), core binding factor leukemia (RUNX1-RUNX1T1, CBFB-MYH11), and KMT2Ar cases. Cases underwent whole genome, exome, and RNA sequencing. Data was combined with four other pediatric datasets including FAB M7, early T cell precursor ALL, the TARGET AML dataset, and pediatric mixed phenotype acute leukemia for a total of 435 cases(4-7-9). 10 additional KMT2Ar AML cases were sequenced to increase the cohort size. Transcriptional clusters as identified tSNE, somatic calls, and outcome correlates were utilized to identify biologic subtypes as previously described(4). (B) RNA sequencing of 435 cases of pediatric AML, AMKL, MPAL, and ETP were combined and batch corrected. tSNE visualization utilizing the top 100 differentially expressed genes within each dataset. Immunophenotype of cases as determined by flow cytometry at diagnosis is shown. (C) Key oncogenic driver mutations as determined by next generation sequencing.

Figure 2. Mutational Composition of Pediatric Myeloid Malignancies. Waterfall plot of major mutational events across the entire cohort. See Supplementary Table 12 for genomic details and Supplementary Table 13 for genes and fusion events included in each of the groupings on the y-axis.

Figure 3. Acute Myeloid/T-Lymphoblastic Leukemia. 63 cases spanning AML, AUL, MPAL, and ETP immunophenotypes shared a common transcriptional identity (see Figure 2A). (A) Mutational spectrum of AMTL cases. (B) Outcomes of AMTL patients according to FLT3-ITD and PRC2 transcriptional identity. (C) Outcomes of FLT3-ITD/WT1 double mutant cases based on AMTL and MK-V transcriptional identity (see Figure 1). (D) Expression of HOX locus genes
in normal hematopoietic progenitor subsets and \textit{FLT3-ITD/WT1} cases from AMTL and MK-V transcriptional clusters. Hematopoietic Stem Cells (HSC), Common Myeloid Progenitors (CMP), Lymphoid Progenitors (LP). (E) Enrichment of gene expression signatures from HSC, CMP, and LP in \textit{FLT3-ITD/WT1} cases from AMTL and MK-V transcriptional clusters. (F) pLSC6 score of \textit{FLT3-ITD/WT1} cases from AMTL and MK-V transcriptional clusters.

**Figure 4. Leukemia Stemness is Associated with Overall Survival.** (A) pLSC6 score was determined in normal hematopoietic progenitor subsets as previously described (17,19). (B) pLSC6 scores from normal hematopoietic progenitors were used to define thresholds of low (lineage committed cells), intermediate (multipotent progenitors), and high values (pluripotent progenitors) in our cohort. (C) Imposing pLSC6 thresholds on our cohort found a subset of patients with intermediate and high scores which was significantly associated with overall survival ($p=9.3 \times 10^{-7}$). (D) tSNE visualization of the cohort with pLSC6 levels indicated. Thresholds of low (grey), medium (blue), and high (red) as determined in panel A.

**Figure 5. Oncogenic Driver Events, Transcriptional Identity, and Leukemia Stemness all Contribute to Outcome in Pediatric Myeloid Related Acute Leukemias.** (A) Integrative Cox proportional hazard model to look at associations with overall survivals in the discovery cohort.(38) Each bar represents the -log10 p-value of covariate(s) and their association with survival. The covariate(s) used in the model to calculate the p-value are indicated below the graph with a check mark. Immunophenotype as a single covariate failed to reach statistical significance. (B) Probability of event free survival (pEFS) of an ongoing multi-institutional prospective pediatric AML trial (AML16) and the proposed classification scheme based on this manuscript for the validation cohort. See Supplementary Figures 22 and 23 for results of each independent cohort. (C) Performance of the proposed genomic classification relative to that utilized in an ongoing prospective up-front pediatric AML study (NCT03164057) in terms of discrimination capability (left panel) and percentage of high or low risk classified patients (upper
right panel) culminating in a risk classification utility score (right lower panel) for the validation cohort. See Supplementary Figures 24 and 25 for results of each independent cohort. (D) Working Model. Mutational events in distinct hematopoietic progenitor subsets lead to transformation and both components contribute to the transcriptional identity and leukemia stemness. Chemotherapy sensitivity and therefore outcomes are a composite of these factors.

**Figure 6. Hierarchical Decision-Making Tree for Proposed Risk Stratification.** *T-MPAL -* mixed phenotype acute leukemias with T lineage markers. MPAL cases co-expressing B-lineage markers contained ZNF384, Ph+, Ph-like, and KMT2Ar oncogenes and should be treated with ALL directed therapy unless they prove non-responsive to this approach. **FLT3-ITD cases that are not AMTL and lack high risk and low risk features such as NUP98r, monosomy 7, NPM1, CEBPA, etc.*
Figure 1
Figure 2

Association for Cancer Research.
Figure 3

A

B

C

D

E

F

Downloaded from https://bloodcancerjournals.aacrjournals.org/ by guest on September 19, 2021. Copyright 2021 American Association for Cancer Research.
Figure 4

A. Schematic diagram showing the distribution of hematopoietic progenitor cells. Key markers include HSC, MEP, GMP, CMP, Ly1, and Ly2.

B. A frequency histogram illustrating the distribution of pLSC6 values with three categories: low, medium, and high.

C. Kaplan-Meier survival analysis for different pLSC6 categories: low, medium, and high. The p-values for each category are indicated.

D. t-SNE coordinate plot showing the relationships between different cell types and their expression of various genes and mutations, such as CBFA2T3-GLIS2, RUNX1-r, CBFB-MYH11, and FLT3ITD.
Figure 5
Figure 6

- KAT6Ar, FLT3-ITD**, RBM15-MKL1, Other
  - Yes -> Med/High pLSC6
  - No -> IR

- CEBPA, NPM1, GATA1, HOxR
  - Yes -> HR
  - No -> LR

- RUNX1-RUNX1T1, RUNX1-CBFA2T3, CBFB-MYH11
  - Yes -> PRC2/ckIT or Med/High pLSC6
  - No -> LR

- CBFA2T3-GLIS2, DEK-NUP214, FLT3-ITD/WT1, Monosomy 7, NUP98r, MNX1r, ETSr
  - Yes -> HR
  - No -> AMTL

- AMTL
  - Yes -> FLT3-ITD
  - No -> KMT2Ar

- KMT2Ar
  - Yes -> HR
  - No -> AMKL

- AMKL
  - Yes -> HR
  - No -> MK-V

- MK-V
  - Yes -> HR
  - No -> Med/High pLSC6

- Med/High pLSC6
  - Yes -> HR
  - No -> IR
Integrative Genomic Analysis of Pediatric Myeloid Related Acute Leukemias Identifies Novel Subtypes and Prognostic Indicators

Maarten Fornerod, Jing Ma, Sanne Noort, et al.

Blood Cancer Discov Published OnlineFirst September 9, 2021.

Updated version
Access the most recent version of this article at:
doi: 10.1158/2643-3230.BCD-21-0049

Supplementary Material
Access the most recent supplemental material at:
http://bloodcancerdiscov.aacrjournals.org/content/suppl/2021/09/04/2643-3230.BCD-21-0049.DC1

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://bloodcancerdiscov.aacrjournals.org/content/early/2021/09/08/2643-3230.BCD-21-0049.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.