Published in final edited form as:

Nat Genet. 2017 March ; 49(3): 451–456. doi:10.1038/ng.3772.

Pediatric Non-Down Syndrome Acute Megakaryoblastic Leukemia is Characterized by Distinct Genomic Subsets with Varying Outcomes

Jasmijn D.E. de Rooij[1,*], **Cristyn Branstetter**[2,*], **Jing Ma**[3], **Yongjin Li**[4], **Michael P. Walsh**[3], **Jinjun Cheng**[3], **Askar Obulkasim**[1], **Jinjun Dang**[2], **John Easton**[4], **Lonneke J. Verboon**[1], **Heather L. Mulder**[4], **Martin Zimmermann**[5], **Cary Koss**[2], **Pankaj Gupta**[4], **Michael Edmonson**[4], **Michael Rusch**[4], **Joshua Yew Suang Lim**[6], **Katarina Reinhardt**[7], **Martina Pigazzi**[8], **Guangchun Song**[3], **Allen Eng Juh Yeoh**[6,9], **Lee-Yung Shih**[10], **Der-Cherng Liang**[11], **Stephanie Halene**[12], **Diane S. Krause**[13], **Jinghui Zhang**[4], **James R. Downing**[3], **Franco Locatelli**[14,†], **Dirk Reinhardt**[7,†], **Marry M. van den Heuvel-Eibrink**[1,15,†], **C. Michel Zwaan**[1,†], **Maarten Fornerod**[1,†], and **Tanja A. Gruber**[2,3,†]

[1]Department of Pediatric Oncology, Erasmus MC-Sophia Children's Hospital, Rotterdam, Netherlands [2]Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN, USA [3]Department of Pathology, St. Jude Children's Research Hospital, Memphis, TN, USA [4]Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN, USA [5]Department of Pediatric Hematology/Oncology, Medical School Hannover, Hannover, Germany [6]Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore [7]Department of Paediatric Oncology, University of Duisburg-Essen, Essen, Germany [8]Department of Women's and Children's Health, University of Padova, Padova, Italy [9]Cancer Science Institute, National University of Singapore, Singapore [10]Chang Gung Memorial Hospital-Linkou, Chang Gung University, Taiwan [11]Department of Pediatrics, Mackay Memorial Hospital, Taipei, Taiwan [12]Section of Hematology, Department of Internal Medicine and Yale Comprehensive Cancer Center, Yale University School of Medicine, New Haven, CT, USA [13]Department of Laboratory Medicine, Yale University, New Haven, CT USA [14]Department of Pediatric Hematology Oncology, University of Pavia, IRCCS Ospedale Pediatrico Bambino Gesù,

Correspondence should be addressed to F.L. (franco.locatelli@opbg.net), D.R. (dirk.reinhardt@uk-essen.de), M.VDH-E. (m.m.vandenheuvel-eibrink@prinsesmaximacentrum.nl), C.M.Z. (c.m.zwaan@erasmusmc.nl), M.F. (m.fornerod@erasmusmc.nl) or T.A.G. (tanja.gruber@stjude.org).

*These authors contributed equally

†Co-corresponding authors on behalf of AIEOP, BFM, DCOG, and SJCRH study groups

AUTHOR CONTRIBUTIONS

T.A.G. and M.F. designed all experiments. J.C., H.L.M., and J.E. constructed libraries and sequenced samples. J.M. led the sequencing analysis. J.M., Y.L., M.P.W., M.R., G.S., A.O., M.F., and J.Z. performed computational data analyses. J.D.E.dR., C.B., and L.J.V. performed validation experiments. C.B., J.M., and T.A.G. manually filtered SNV/Indel calls on unpaired samples. C.K. and J.D. performed functional work on the HOX fusions. J.D.E.dR, M.Z., and M.F. performed outcome analysis. J.L., K.R., M.P., A.E.J.Y., L-Y.S., D-C.L., S.H., D.S.K., F.L., D.R., M.VDH-E., C.M.Z., and T.A.G. provided annotated patient samples. J.D.E.dR., C.B., J.R.D., F.L., D.R., M.VDH-E., and C.M.Z. performed critical reading and contributed to the writing of the manuscript. M.F. and T.A.G. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Rome, Italy [15]Department of Pediatric Oncology, Princess Maxima Center for Pediatric Oncology, Utrecht, Netherlands

Abstract

Acute Megakaryoblastic Leukemia (AMKL) is a subtype of acute myeloid leukemia (AML) in which cells morphologically resemble abnormal megakaryoblasts. While rare in adults, AMKL accounts for 4–15% of newly diagnosed childhood AML[1–3]. AMKL in patients without Down syndrome (non-DS-AMKL) is frequently associated with poor outcomes. Previous efforts have identified chimeric oncogenes in a significant number of cases, including *RBM15-MKL1*, *CBFA2T3-GLIS2, KMT2A* gene rearrangements and *NUP98-KDM5A*[4–6]. The etiology of 30–40% of cases, however, remains unknown. To better understand the genomic landscape of non-DS-AMKL, we performed RNA and exome sequencing on specimens from 99 patients (75 pediatric and 24 adult). We demonstrate that pediatric non-DS-AMKL is a heterogeneous malignancy that can be divided into seven subgroups with varying outcomes. These subgroups are characterized by chimeric oncogenes with cooperating mutations in epigenetic and kinase signaling genes. Overall, these data shed light on the etiology of AMKL and provide useful information for treatment tailoring.

The earliest insight into the genomic alterations that drive non-DS-AMKL occurred with the recognition of a recurrent t(1;22) found in infants[7]. Ten years after this initial report, the *RBM15* and *MKL1* genes involved in this translocation were identified[4,8]. Subsequent to this, a high resolution study of DNA copy number abnormalities (CNAs) and loss of heterozygosity (LOH) on pediatric *de novo* AML identified a significant number of alterations in non-DS-AMKL cases, suggesting that additional gene rearrangements may be present in this population[9]. To define the functional consequences of these structural variations, diagnostic leukemia specimens from 14 pediatric patients previously underwent RNA and exome sequencing[5]. This effort identified chimeric transcripts encoding fusion proteins in 12 of 14 cases, including the novel recurrent *CBFA2T3-GLIS2* fusion which correlates with a poor prognosis. In parallel, a separate non-DS-AMKL cohort was evaluated by PCR and split-signal fluorescence *in situ* hybridization (FISH) for fusion events associated with myeloid malignancies including *NUP98* and *KMT2A* gene rearrangements (*KMT2A*r)[6]. In this cohort, *NUP98* and *KMT2Ar* were identified in 15% and 10% of cases respectively. To gain a more comprehensive understanding of the genomic alterations that lead to non-DS-AMKL, specimens from 99 patients (75 pediatric and 24 adult cases) were subjected to RNA and/or exome sequencing (Supplementary Tables 1–3). Combined with the 14 cases previously described, the pediatric cohort described in this manuscript yields 89 cases, the largest of this rare malignancy to undergo next generation sequencing to date.

Of the 93 patients for whom sufficient RNA was available, 5.5% adult and 72.4% pediatric cases carried a structural variation (SV) predicted to lead to a fusion product by RNAseq (Supplementary Table 4). Ten additional fusion events in the 19 patients that lacked RNA for sequencing could be recognized by RT-PCR, FISH, or southern blotting (Figure 1, Supplementary Tables 1,4). In the pediatric cohort, the most frequent previously reported

fusion events include *CBFA2T3-GLIS2* (18.6%), *KMT2A*r (17.4%), *NUP98-KDM5A* (11.6%), and *RBM15-MKL1* (10.5%). Previously described low frequency non-DS-AMKL fusions identified in this expanded cohort include a case of *NIPBL-HOXB9* and a novel, but analogous *NIPBL-HOXA9* fusion[5]. Similarly, a case carrying *GATA2-HOXA10* was identified, which is comparable to the *GATA2-HOXA9* fusion that has been reported in a single case[5]. Chimeric transcripts not previously described include several fusions involving genes within the HOX cluster (*EWSR1-HOXB8*, *PLEK-HOXA11-AS*, *BMP2K-HOXD10, and EP300-HOXA7*). Collectively, fusions involving a *HOX* cluster gene (*HOX*r) occurred in 14% of patients within this cohort (Table 1, Supplementary Tables 1,4). Many of the *HOX*r are predicted to lead to an in-frame functional fusion protein (Table 1, Supplementary Table 4). As proof of principal, we evaluated three of these fusion events and found all three to enhance self-renewal as determined by an *in vitro* colony replating assay (Supplemental Figure 1). Several fusions, however, involve non-coding RNA species and are predicted to result in a loss of function of these regulatory transcripts (Supplementary Table 4). 3.5% of cases carried non-recurrent fusion proteins involving hematopoietic transcription factors and epigenetic regulators, including *MN1-FLI1*, *BCR-ABL1*, and *MAP2K2-AF10* (Supplementary Tables 1,4). 3.5% of cases were found to have chimeric transcripts involving the cohesin gene *STAG2* which were all predicted to lead to a truncation in the protein (Supplementary Tables 1,4). In 21% of the pediatric cases, no potentially oncogenic fusion event could be detected. However, in 50% of these latter cases, a truncating mutation in exon 2 or 3 of *GATA1* was found, amounting to 10.1% of the pediatric cohort overall (further discussed below, Figure 1 and Supplementary Tables 1,3–7).

To determine if these fusion events contribute significantly to gene expression patterns, samples with greater than 60% purity were subjected to unsupervised clustering using the top 100 most variant genes by standard deviation (Figure 2a,b, Supplementary Tables 8,9). Confirming the strength of the fusions in determining gene expression signatures, samples clustered according to fusion status; specifically *KMT2A*r, *HOX*r, *NUP98-KDM5A*, and *CBFA2T3-GLIS2* cases formed distinct clusters. When analyzing expression of the *HOX* gene cluster, we found upregulation of the *HOX* gene involved in the fusion construct (Supplemental Figure 2a), often accompanied by upregulation of adjacent *HOX* genes. To determine whether this upregulation has downstream effects on HOX protein targets such as *PIM1*, we evaluated expression of a gene set defined by *HOXA9* overexpression in hematopoietic cells[10,11]. This demonstrated the highest association for the majority of target genes with the *HOX*r subgroup, providing further evidence of a *HOX* gene expression signature (Figure 2c and Supplementary Figure 2b). Combined with *KMT2A*r and *NUP98-KDM5A*, chimeric oncogenes known to upregulate *HOX* cluster genes, roughly half of pediatric non-DS-AMKL patients carry a *HOX* gene expression program. These cases were distinct from those carrying the cryptic *CBFA2T3-GLIS2* inversion which clustered away from all other non-DS-AMKL, as previously shown[5].

In addition to RNAseq data, 68 pediatric patients had DNA available for whole exome sequencing (WES), of which 30 had paired germline material (Supplementary Tables 1–2, 5). To identify single nucleotide variants and insertion/deletion events (SNV/Indels) in cases lacking WES, RNAseq data was also interrogated for these mutational events (Supplementary Table 5). To identify high confidence somatic calls, unpaired samples

underwent a rigorous filtering process as described in the online methods (Supplementary Figures 3,4 and Supplementary Tables 6,7). Of 83 pediatric cases at diagnosis for which SNV/Indel analysis was available, the most highly recurrent mutations occurred in *GATA1* (13.3%), *JAK* kinase *or STAT* genes (16.9%), Cohesin or *CTCF* genes (18.1%), and RAS pathway genes (15.7%) (Figure 3a, Supplementary Tables 1,5,6). Additionally, 18.1% of patients carried mutations in a cytokine receptor gene, the most frequent of which was the thrombopoietin receptor *MPL* (n=11) that plays a role in normal megakaryoblast growth and survival[12]. In contrast, in the adult cohort (n=24) which lacked recurrent fusion genes, the most highly recurrent mutations were in *TP53* (20.8%), cohesin genes (16.7%), splicing factor genes (16.7%), *ASXL* genes (16.7%) and *DNMT3A* (12.5%) (Supplementary Table 7). Paired exome specimens and single nucleotide polymorphism (SNP) arrays were available in 40 specimens for copy number alterations analysis to identify additional cooperating mutations (Figure 3b, Supplementary Table 10). The tumor suppressor RB1 gene was found to be recurrently targeted with focal deletion events (Supplementary Table 11). Combined with SNV/Indels and structural variations, *RB1* mutational events occurred in 14.3% of the pediatric cohort and 8.3% of the adult cohort. Confirming previous reports, gains in chromosome 19 and 21 were also recurrent in the pediatric cohort as determined by WES and/or cytogenetics in 24% and 39.2% of cases, respectively (Supplementary Tables 1,10)[13,14].

Of the ten cases carrying *GATA1* truncating mutations in the pediatric cohort, (Supplementary Table 12), none had physical stigmata consistent with DS. Karyotypes were available for all patients and found to be negative for constitutional trisomy 21. Four pediatric cases had matched germline available with an average coverage of 110X. With the exception of one case that had evidence of low level tumor contamination in the germline specimen, *GATA1* mutant calls were absent from these remission samples, including a case with 160X coverage, arguing against mosaicism in the hematopoietic compartment of these patients. One patient had non-hematopoietic tissue available for analysis and was found to be germline mosaic for trisomy 21 (Supplementary Table 12)[15]. The strong association of *GATA1* truncations in DS patients suggests cooperativity between amplification of the Down syndrome critical region (DSCR) on chromosome 21 and the *GATA1* mutant. We therefore evaluated cases for amplification of the DSCR using SNP arrays and exome read depth in paired samples. Nine of ten *GATA1* mutant cases had amplifications in the DSCR (Supplementary Table 12). In all cases, GATA1 mutations and chromosome 21 amplifications were in the major clone and thus the order of acquisition could not be determined (Supplementary Table 12). Across the entire cohort, amplifications of chromosome 21 were found to be one of the most highly recurrent copy number alterations (39.2% overall; Figure 3b, Supplementary Tables 1,10). Candidate genes in this region that play a role in megakaryopoiesis include *UBASH3A*, *LINC00478/MONC, DYRK1A, and ERG among others*[16–20]. *GATA*1 mutant cases comprised a distinct subset at the gene expression level (Figure 2a,b), and this signature was strongly correlated with that found in DS-AMKL (Supplementary Figure 5a). Confirming cooperativity with chromosome 21, the *GATA1* mutant subset significantly overexpressed chromosome 21 genes, even in comparison to other samples carrying extra copies of this chromosome (Supplementary Figure 5b and Table 13). Combined with RNA-seq data, this led us to divide our cohort into

seven subsets based on genomic lesions for further analysis: *CBFA2T3-GLIS2*, *RBM15-MKL1*, *NUP98-KDM5A*, *KMT2A*r, *HOX*r, *GATA1*, and "Other" which is a subset comprised of cases not falling into any of the aforementioned categories (Figure 1).

Cooperating mutations as identified by WES and SNP arrays revealed a significant association between subgroups and recurrent mutations ($p=2.8\times10^{-8}$, global test) (Figure 3c, Supplementary Table 14)[21]. *NUP98-KDM5A* cases carried mutations in *RB1* almost without exception and demonstrated a decrease in expression of this gene (Supplementary Figure 6 and Table 11); *KMT2A*r often associated with RAS pathway lesions as has been previously described (p=0.09 for enrichment, Fisher's exact test) while JAK pathway and cohesin mutations were commonly identified in *GATA1* mutant cases (p=0.003 and p=0.04 respectively)[22,23]. *HOX*r cases were found to be significantly enriched in activating *MPL* mutations (p=0.01). To further evaluate the functional consequences of *HOX*r and *MPL* mutations, we introduced the wild type *MPL* gene or one of two *MPL* mutations along with a *HOX*r into murine bone marrow for colony forming assays (Supplementary Figure 7a). Expression of both a *HOX*r and either wild type or mutant *MPL* failed to alter colony numbers or immunophenotype (Supplementary Figure 7a and data not shown). When cells were removed from cytokine containing media, however, a growth advantage was identified in cells containing both a *HOX*r and a *MPL* mutation (Supplementary Figure 7b). Activated JAK-STAT signaling as determined by phosphorylated JAK2 and STAT5 was found in HOXr cells containing a MPL mutation in contrast to wild type MPL or empty vector, providing one mechanism for this growth advantage (Supplementary Figure 8).

Clinical outcomes for DS-AMKL are uniformly excellent, whereas studies on non-DS-AMKL have shown discrepant results, with the majority reporting inferior survival rates compared to other AML subtypes[2,3,24–26]. Furthermore, the recommendation for allogeneic stem cell transplant (SCT) in first complete remission for non-DS-AMKL patients is not uniform among pediatric cooperative groups. To understand the association between genomic subgroups and patient outcome in pediatric non-DS-AMKL, we first utilized the global test to evaluate if the subgroups correlated with different probabilities of survival. Overall survival (pOS) did not differ statistically between patients treated across continents (p=0.8; data not shown). We observed statistically significant associations with *CBFA2T3-GLIS2,* which carries the strongest negative association, and *GATA1* mutant cases which carry the strongest positive association ($p=1.7\times10^{-5}$ for pOS, $p=3.4\times10^{-5}$ for event free survival (pEFS), see Supplementary Figure 9)[21]. Kaplan-Meier estimates of pOS, pEFS, and cumulative incidence of relapse or primary resistance (pCIR, $p_{gray}=1.4\times10^{-4}$) confirmed this trend (Figure 4a–c, Supplementary Table 15). Specifically, *CBFA2T3-GLIS2* and *KMT2A*r were found to have significantly inferior pEFS and pOS. *NUP98-KDM5A* cases also demonstrated a trend towards poor outcomes, however due to small numbers this failed to reach statistical significance. Conversely, *GATA1* and *HOX*r subgroups carried significantly superior outcomes. Of note, all *GATA1* mutant cases that lacked a fusion gene were cured, mimicking the excellent outcomes observed in DS-AMKL[24]. Hence these patients are not only biologically but also clinically similar, suggesting that they may benefit from the less intensive chemotherapy regimens given to DS-AMKL patients[15,27]. In contrast, the two patients carrying a *GATA1* mutation and a poor prognosis fusion gene were non-survivors. Based on these results, we recommend all pediatric non-DS-AMKL patients be tested for the

presence of *GATA1* mutations, *CBFA2T3-GLIS2, KMT2A*r, and *NUP98-KDM5A*. Chimeric oncogenes that define two of these subsets, *CBFA2T3-GLIS2* and *NUP98-KDM5A*, are missed by conventional karyotyping and therefore require split-signal FISH or RT-PCR for detection. Although SCT was not associated with improved pEFS or pOS, a decrease in relapsed free survival (RFS) was found for those patients receiving this treatment modality (HR 0.28 p=0.044 see Supplementary Table 16). Therefore patients carrying *CBFA2T3-GLIS2* or *KMT2A*r that have inferior outcomes may benefit from allogeneic SCT in first complete remission. While *NUP98-KDM5A* outcomes did not reach statistical significance, their pEFS and pOS warrant close monitoring and consideration of allogeneic SCT as well. In contrast, patients lacking these lesions have outcomes on par with or superior to other subtypes of pediatric AML and transplant in first remission should be reserved for those showing a poor response to induction therapy (e.g. high levels of minimal residual disease).

In summary, pediatric non-DS-AMKL is a heterogeneous malignancy comprised of distinct subsets as defined by next generation sequencing with varying outcomes. We have identified a previously unrecognized subtype characterized by diverse rearrangements in the *HOX* loci that share similar gene expression signatures, cooperating mutations, and clinical outcomes. Identification of key genomic events in newly diagnosed non-DS-AMKL patients is important for risk stratification as these lesions have therapeutic implications. Allogeneic SCT in first complete remission should be considered for patients carrying a poor prognosis fusion event including *CBFA2T3-GLIS2*, *KMT2A*r, and *NUP98-KDM5A*.

ONLINE METHODS

Patient Samples—Specimens were provided from multiple institutions. All samples were obtained with patient or parent/guardian provided informed consent under protocols approved by the Institutional Review Board at each institution. Samples were deidentified prior to nucleic acid extraction and analysis.

Next Generation Sequencing—RNA and DNA library construction for RNA and whole exome DNA sequencing were done as per manufacturer's instructions using the Illumina True-seq RNA sample preparation V2 and Nextera rapid capture exome kits, respectively. Sequencing was completed on the Illumina HiSeq 2000 as per manufacturer's instructions. Analysis of RNA and whole-exome sequencing data which includes mapping, coverage and quality assessment, SNV/Indel detection, tier annotation for sequence mutations, and prediction of deleterious effects of missense mutations have been described previously[5,28]. Open reading frames predictions of fusion transcripts detected by RNAseq were validated by RT-PCR followed by Sanger sequencing of the purified PCR products.

Exome Filtering—To identify high confidence somatic calls, unpaired samples underwent a vigorous filtering process including the removal of low quality calls and known polymorphisms. Rare variants (defined as a mutant allele frequency of <0.1% in the non-cancer NHLBI ESP cohort) were retained for further analysis. Known recurrent somatic variants present in the catalogue of somatic mutations in cancer database (COSMIC) were designated as high confidence lesions. Remaining calls were evaluated by damage-

prediction algorithms. Those with mutations occurring in a conserved domain of a cancer consensus gene and predicted to be damaging were designated as intermediate confidence lesions.

Gene Expression Analysis—Transcript expression levels were estimated as Fragments Per Kilobase of transcript per Million mapped fragments (FPKM); gene FPKMs were computed by summing the transcript FPKMs for each gene using Cuffdiff2[29,30]. A gene was considered "expressed" if the FPKM value was >= 0.35 based on the distribution of FPKM values. Genes that were not expressed in any sample group were excluded from the downstream analysis. Hierarchical clustering and t-SNE were performed using the top 100 most variant genes. Prior to the analysis we excluded sex specific genes, sno, miRNAs and genes whose expression are correlated with inflammatory responses (see Supplementary Table 11).

All statistical analyses were performed in R statistical environment.

Unsupervised clustering: Expression levels of genes were estimated as Fragments Per Kilobase of transcript per Million mapped fragments (FPKM); FPKMs were computed by summing the transcript FPKMs for each gene using Cuffdiff2[29,30]. A gene was considered "expressed" if the FPKM value was ≥0.35 based on the distribution of FPKM gene expression levels. Genes that were not expressed in any sample group were excluded from the final data matrix for downstream analysis (log2 FKBM value −1.514573) leaving 18905 genes. To avoid disturbances from gene expression from normal bone marrow cells and based on the distribution of tumor purity, samples were included with a blast count percentage of >60%. To further eliminate contaminating gene expression, technical replicas from a sample which was either blast cell purified (SJAMLM7010964_D1, blast percentage 93%) or unpurified (SJAMLM7010964_D1, blast percentage 78%) were compared and genes with high expression in the unpurified sample (n = 25) only were eliminated (Supplementary Table 11). In addition sex specific genes (n = 41) and small RNA encoding genes (n = 279) were removed. The latter to remove variance based on RNA extraction protocol (Supplementary Table 11). The final data matrix is composed of 48 samples and 18563 genes. We performed hierarchical clustering (HC) using the top 100 most variant genes (Supplementary Table 12). A HC tree, with Spearman correlation as distance metric and Ward linkage, was constructed. The distributed stochastic neighborhood embedding method implemented in the R-package tSNE was used to visualize the similarities between samples in two-dimensional space.

Analysis of chromosome 21 expression levels: Genes on chr21 (chr21) and those not on chr21 (non-chr21) were first saved. Lowly expressed genes (mean log2 FKBM value <= −1) were removed. For each gene mean expression levels in AMKL subgroups were calculated, and subsequently log fold changes between chr21 and non-chr21 were determined. P value was calculated using Wilcoxon rank sum test. Chr21 amplification status was derived from karyotype and/or CNV data. Differential chr21 gene expression between the non-DS *GATA1*s subgroup samples (n=5, all carrying one or more extra copies of chr21) and other non-DS AMKL samples with one or more extra copies of chr21 (n=13) (Supplementary Table 13) was performed on RNAseq count data using the R-package edgeR. Similar

filtering was performed, i.e. genes with a count per million value of > 2 across >= 4 samples included.

Comparison of gene expression in DS and Non-DS AMKL: DS-AMKL logFC gene expression data was taken from Klusmann et al[31]. Subgroup-specific RNA expression in non-DS-AMKL was calculated using RNAseq count data aggregated on genes from samples from unique patients with a blast count > 60%, comparing each subgroup against all others using the R-package edgeR[32]. Prior to the analysis, sex-specific genes were removed. Data were merged and aggregated on gene symbol and gene expression values. Finally, most significantly differentially expressed genes in both data sets (FDR < 0.2) were correlated.

Association of gene expression with HOXA9 target genes: First, up-regulated targets of HOXA9 were taken from Dorsam et al[10]. Then, association between the expression values of these genes, with the exception of those that were not expressed in either sample, and AMKL subgroups was quantified using the global test, where group label were used as response variable.

Associations of AMKL subgroups with cooperating mutations—High confidence SNVs/Indels from the initial diagnostic samples excluding mutations in GATA1 were combined with structural alterations excluding those identifying genomic subgroups were subjected to mutational frequency analysis (see Supplementary Tables 4, 6, 9, and 15). One hypermutated sample (SJAMLM7060_D) was excluded from this analysis. Genes mutated in >4 cases (*RB1*, *MPL*, *CTCF*, *JAK2* and *NRAS*) were identified. Based on these five genes, five non-overlapping proximal gene sets were constructed from mutated genes, covering 44% of identified cooperating mutations. Global association with AMKL subgroup was calculated using the global test[21]. Enrichment of gene set mutations in AMKL subgroups was determined by one-sided Fisher's exact tests. The Circos plot was constructed using Martin Krzywinski's table viewer (http://mkweb.bcgsc.ca/tableviewer/).

Colony Forming Assay—All experiments involving mice were reviewed and approved by the Institutional Animal Care and Use Committee. Bone marrow from 4–6 week old female C57BL/6 mice was harvested, lineage depleted, and cultured in the presence of recombinant murine SCF (rmSCF, Peprotech, 50ng/ml), IL-3 (rmIL3, Peprotech, 50ng/ml), and IL-6 (rmIL6, Peprotech, 50ng/ml) for 24 hours prior to transduction on RetroNectin (Takara Bio Inc.) coated plates. Cultured supernatants containing ecotropic envelope pseudotyped retroviral vectors were produced by transient transfection of 293T cells as previously described[33]. Murine bone marrow cells were harvested 48 hours following transduction, sorted for vector-encoded mCherry or GFP expression, and plated on methylcellulose containing IL-3, IL-6, SCF and erythropoietin (EPO) (Stem Cell technologies, Vancouver, BC) as per manufacturer's instructions. Colonies were counted after 7 days of growth at 37°C, harvested and serially replated.

Affymetrix SNP Arrays—Affymetrix SNP 6.0 array genotyping was performed for 14 of 15 AMKL cases in the discovery cohort, and array normalization and DNA copy number alterations identified as previously described[34–37]. To differentiate inherited copy number alterations from somatic events in leukaemia blasts from patient's lacking matched normal

DNA, identified putative variants were filtered using public copy number polymorphism databases and a St. Jude database of SNP array data from several hundred samples[38,39].

Copy Number Alteration Detection Using Whole Exome Sequencing—Samtools mpileup command was used to generate an mpileup file from matched normal and tumor BAM files with duplicates removed[40]. VarScan2 was then used to take the mpileup file to call somatic CNAs after adjusting for normal/tumor sample read coverage depth and GC content[41]. Circular Binary Segmentation algorithm implemented in the DNAcopy R package was used to identify the candidate copy number alterations for each sample[37]. B-allele frequency info for all high quality dbSNPs heterozygous in the germline sample was also used to assess allele imbalance.

Western Blot Analysis

Cells transduced with both a fusion gene and a MPL construct were flow sorted and grown on cytokine supplemented methylcellulose and serially replaced for three weeks. Cells were then harvested and grown in liquid media in the absence of cytokines. 48 hours after withdrawal of cytokines cells were harvested, washed with PBS and lysed with RIPA buffer. Equivalent amounts of protein were separated by Mini-protean TGX Stain-Free Precast Gel (Bio-Rad). The primary antibodies used were Phospho-Jak2 (Tyr1007/1008, clone C80C3, catalog# 3776 Cell Signaling Technology), Jak2 (clone D2E12, catalog# 3230 Cell Signaling Technology), Phospho-Stat5 (Tyr694, clone C11C5 catalog# 9359 Cell Signaling Technology) and Stat5 (catalog# 9363 Cell Signaling Technology). The intensity of the detected signals was measured by using Image Lab.

Survival Analysis—Kaplan-Meier curves for probability of overall survival (pOS), event free survival (pEFS) and cumulative incidence of relapse or non-response (pCIR) were constructed using the R-package survival and IBM SPSS 20.0. Gray's test statistic and p value for CIR were calculated using the *cuminc* function in the R-package cmprsk. Summary statistics for each group are presented in Supplementary Table 15. Events in pEFS calculations were defined as relapse, death in remission by any cause, and non-response which was included as an event at the date of diagnosis. For pCIR, only relapse and non-response were included. No significant differences were present in cumulative incidences of competing risk (p=0.7). For multi-variant analysis, the Cox proportional hazards model was used to obtain the estimates and the 95%-confidence interval of the relative risk for prognostic factors. HSCT was included as time-dependent covariable. Computations were performed using SAS (Statistical Analysis System Version 9.3; SAS Institute, Cary, NC).

Data Availability—Whole exome, RNA–seq and SNP microarray data have been deposited at the European Genome-phenome Archive (EGA) under accession EGAS00001002183.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank all the patients and their parents who allowed their leukemic samples to be stored and studied. We thank the Tissue Resources Laboratory, the Flow Cytometry and Cell Sorting Core, and the Clinical Applications of Core Technology Laboratories of the Hartwell Center for Bioinformatics and Biotechnology of St. Jude Children's Research Hospital. J.D.E.dR. was funded by Stichting Kinderoncologisch Centrum Rotterdam (KOCR). A.O. and L.J.V. were funded by KIKA (Children Cancer free foundation), M.F. was supported by the Dutch Cancer Society KWF. F.L. was supported by the Italian Association for Research on Cancer (Associazione Italiana Ricerca sul Cancro, Special Grant "5xmille"-9962). This work was funded by The St. Jude Children's Research Hospital – Washington University Pediatric Cancer Genome Project, the American Lebanese and Syrian Associated Charities of St. Jude Children's Research Hospital, and the Eric Trump Foundation.

References

1. Pagano L, et al. Acute megakaryoblastic leukemia: experience of GIMEMA trials. Leukemia. 2002; 16:1622–6. [PubMed: 12200673]
2. Athale UH, et al. Biology and outcome of childhood acute megakaryoblastic leukemia: a single institution's experience. Blood. 2001; 97:3727–32. [PubMed: 11389009]
3. Barnard DR, Alonzo TA, Gerbing RB, Lange B, Woods WG. Comparison of childhood myelodysplastic syndrome, AML FAB M6 or M7, CCG 2891: report from the Children's Oncology Group. Pediatr Blood Cancer. 2007; 49:17–22. [PubMed: 16856158]
4. Ma Z, et al. Fusion of two novel genes, RBM15 and MKL1, in the t(1;22)(p13;q13) of acute megakaryoblastic leukemia. Nat Genet. 2001; 28:220–1. [PubMed: 11431691]
5. Gruber TA, et al. An Inv(16)(p13.3q24.3)-encoded CBFA2T3-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia. Cancer Cell. 2012; 22:683–97. [PubMed: 23153540]
6. de Rooij JD, et al. NUP98/JARID1A is a novel recurrent abnormality in pediatric acute megakaryoblastic leukemia with a distinct HOX gene expression pattern. Leukemia. 2013; 27:2280–8. [PubMed: 23531517]
7. Baruchel A, Daniel MT, Schaison G, Berger R. Nonrandom t(1;22)(p12–p13;q13) in acute megakaryocytic malignant proliferation. Cancer Genet Cytogenet. 1991; 54:239–43. [PubMed: 1884357]
8. Mercher T, et al. Involvement of a human gene related to the Drosophila spen gene in the recurrent t(1;22) translocation of acute megakaryocytic leukemia. Proc Natl Acad Sci U S A. 2001; 98:5776–9. [PubMed: 11344311]
9. Radtke I, et al. Genomic analysis reveals few genetic alterations in pediatric acute myeloid leukemia. Proc Natl Acad Sci U S A. 2009; 106:12944–9. [PubMed: 19651601]
10. Dorsam ST, et al. The transcriptome of the leukemogenic homeoprotein HOXA9 in human hematopoietic cells. Blood. 2004; 103:1676–84. [PubMed: 14604967]
11. Hu YL, Passegue E, Fong S, Largman C, Lawrence HJ. Evidence that the Pim1 kinase gene is a direct target of HOXA9. Blood. 2007; 109:4732–8. [PubMed: 17327400]
12. Kaushansky K, et al. Promotion of megakaryocyte progenitor expansion and differentiation by the c-Mpl ligand thrombopoietin. Nature. 1994; 369:568–71. [PubMed: 8202159]
13. Dastugue N, et al. Cytogenetic profile of childhood and adult megakaryoblastic leukemia (M7): a study of the Groupe Francais de Cytogenetique Hematologique (GFCH). Blood. 2002; 100:618–26. [PubMed: 12091356]
14. Nimer SD, MacGrogan D, Jhanwar S, Alvarez S. Chromosome 19 abnormalities are commonly seen in AML, M7. Blood. 2002; 100:3838. author reply 3838–9. [PubMed: 12411327]
15. Reinhardt D, et al. GATA1-mutation associated leukemia in children with trisomy 21 mosaic. Klin Padiatr. 2012; 224:153–5. [PubMed: 22513796]
16. Loughran SJ, et al. The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. Nat Immunol. 2008; 9:810–9. [PubMed: 18500345]
17. Goyama S, et al. UBASH3B/Sts-1-CBL axis regulates myeloid proliferation in human preleukemia induced by AML1-ETO. Leukemia. 2016; 30:728–39. [PubMed: 26449661]

18. Emmrich S, et al. LincRNAs MONC and MIR100HG act as oncogenes in acute megakaryoblastic leukemia. Mol Cancer. 2014; 13:171. [PubMed: 25027842]
19. Malinge S, et al. Increased dosage of the chromosome 21 ortholog Dyrk1a promotes megakaryoblastic leukemia in a murine model of Down syndrome. J Clin Invest. 2012; 122:948–62. [PubMed: 22354171]
20. Salek-Ardakani S, et al. ERG is a megakaryocytic oncogene. Cancer Res. 2009; 69:4665–73. [PubMed: 19487285]
21. Goeman JJ, van de Geer SA, de Kort F, van Houwelingen HC. A global test for groups of genes: testing association with a clinical outcome. Bioinformatics. 2004; 20:93–9. [PubMed: 14693814]
22. Andersson AK, et al. The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. Nat Genet. 2015; 47:330–7. [PubMed: 25730765]
23. Yoshida K, et al. The landscape of somatic mutations in Down syndrome-related myeloid disorders. Nat Genet. 2013; 45:1293–9. [PubMed: 24056718]
24. Creutzig U, et al. AML patients with Down syndrome have a high cure rate with AML-BFM therapy with reduced dose intensity. Leukemia. 2005; 19:1355–60. [PubMed: 15920490]
25. Inaba H, et al. Heterogeneous cytogenetic subgroups and outcomes in childhood acute megakaryoblastic leukemia: a retrospective international study. Blood. 2015; 126:1575–84. [PubMed: 26215111]
26. Schweitzer J, et al. Improved outcome of pediatric patients with acute megakaryoblastic leukemia in the AML-BFM 04 trial. Ann Hematol. 2015; 94:1327–36. [PubMed: 25913479]
27. Kudo K, et al. Mosaic Down syndrome-associated acute myeloid leukemia does not require high-dose cytarabine treatment for induction and consolidation therapy. Int J Hematol. 2010; 91:630–5. [PubMed: 20237876]
28. Zhang J, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature. 2012; 481:157–63. [PubMed: 22237106]
29. Trapnell C, et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 2013; 31:46–53. [PubMed: 23222703]
30. Trapnell C, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010; 28:511–5. [PubMed: 20436464]
31. Klusmann JH, et al. miR-125b-2 is a potential oncomiR on human chromosome 21 in megakaryoblastic leukemia. Genes Dev. 2010; 24:478–90. [PubMed: 20194440]
32. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139–40. [PubMed: 19910308]
33. Soneoka Y, et al. A transient three-plasmid expression system for the production of high titer retroviral vectors. Nucleic Acids Res. 1995; 23:628–33. [PubMed: 7899083]
34. Mullighan CG, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature. 2007; 446:758–64. [PubMed: 17344859]
35. Pounds S, et al. Reference alignment of SNP microarray signals for copy number analysis of tumors. Bioinformatics. 2009; 25:315–21. [PubMed: 19052058]
36. Lin M, et al. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. Bioinformatics. 2004; 20:1233–40. [PubMed: 14871870]
37. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics. 2004; 5:557–72. [PubMed: 15475419]
38. McCarroll SA, et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat Genet. 2008; 40:1166–74. [PubMed: 18776908]
39. Iafrate AJ, et al. Detection of large-scale variation in the human genome. Nat Genet. 2004; 36:949–51. [PubMed: 15286789]
40. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–9. [PubMed: 19505943]
41. Koboldt DC, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012; 22:568–76. [PubMed: 22300766]

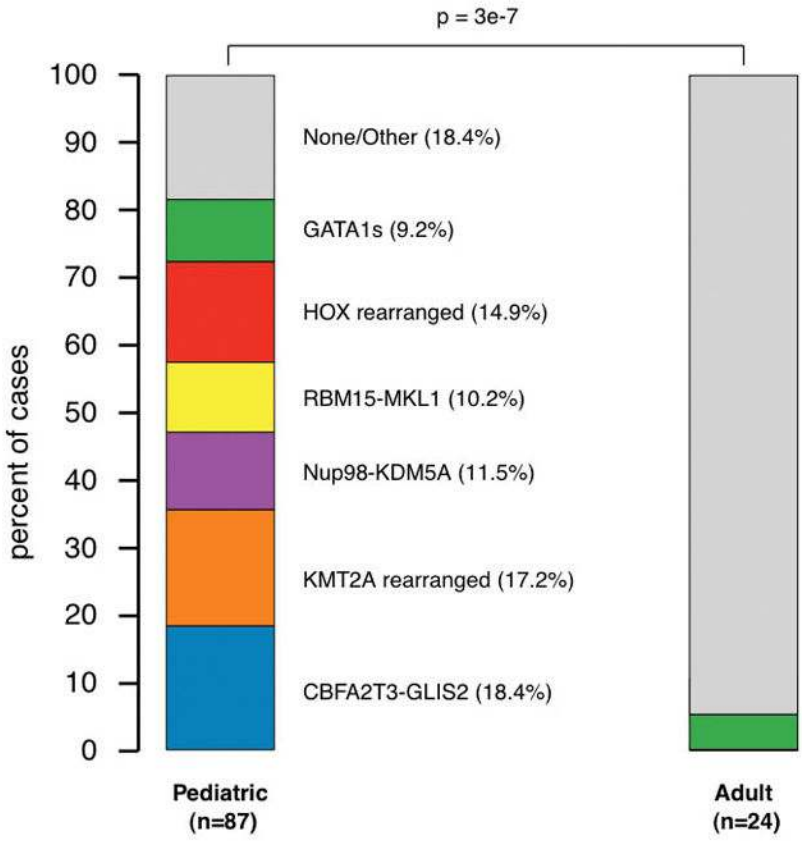

Figure 1.

Pediatric and Adult Non-DS-AMKL are Genomically Distinct. Distributions of recurrent chromosome translocations and *GATA1* mutations in pediatric and adult non-DS-AMKL. *p* value according to Pearson's Chi squared test.

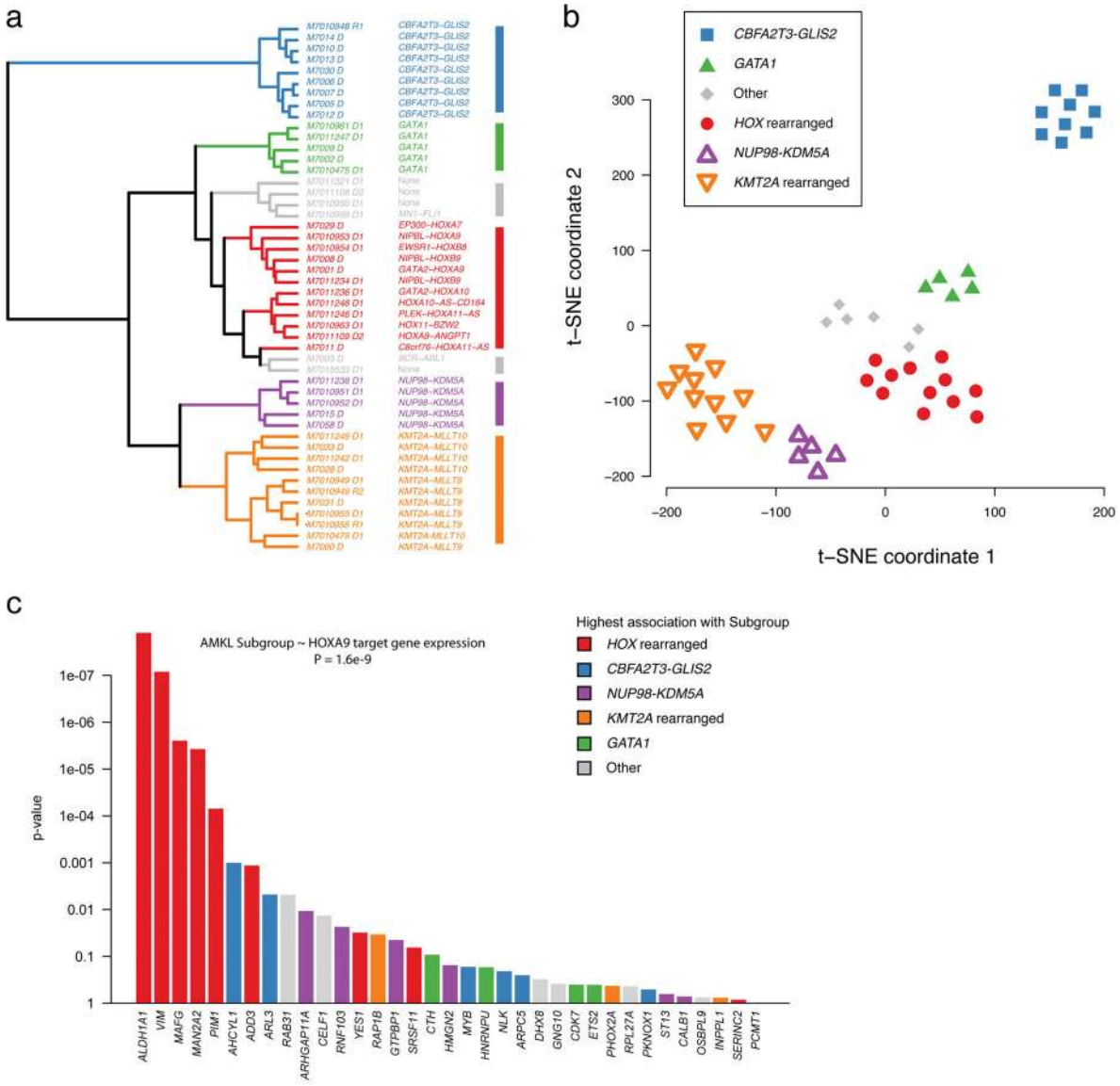

Figure 2.

Gene Expression Analysis Confirms Genomic Subgroups. (a) Unsupervised clustering of patients using the expression values of the 100 most variant genes. (b) t-SNE visualization. (c) Expression of myeloid HOXA9 target genes most highly associates with gene expression in *HOX*r AMKL. Global association between AMKL subgroup and HOXA9 target gene expression was estimated using global test[21]. Contributions of genes to the overall association are indicated by the height of the bars. Bars are ranked with the most significantly correlated genes on the left, and colored according to the subgroup with the highest association. Samples with less than 60% tumor purity were not included in this analysis; as a result no RBM15-MKL1 cases are represented.

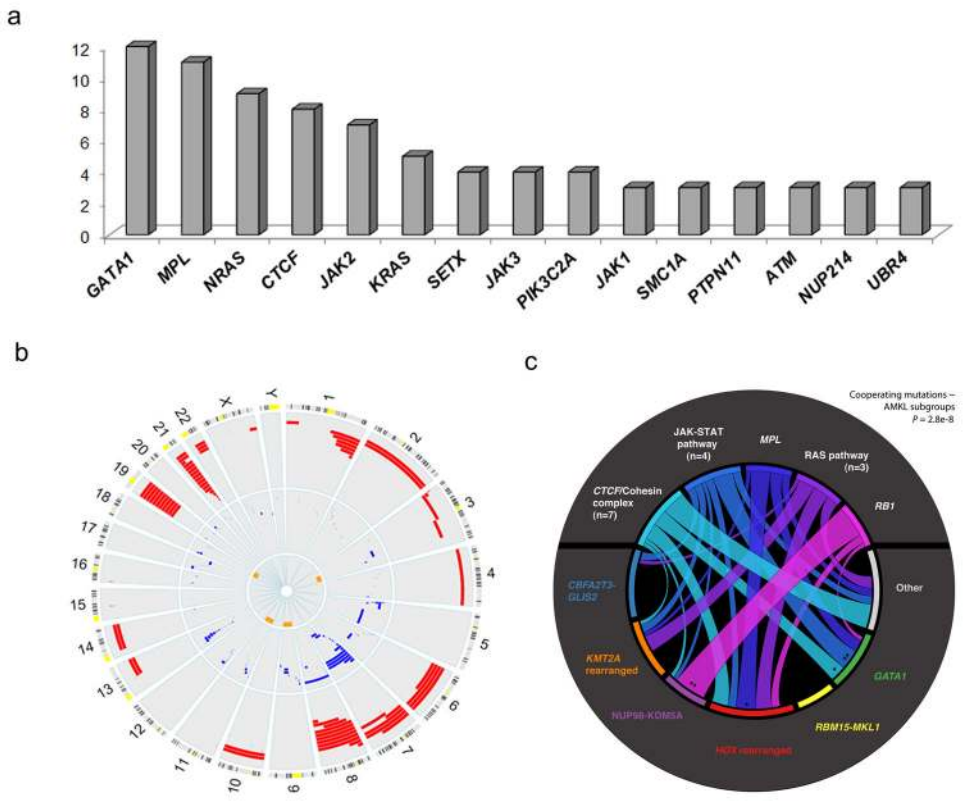

Figure 3.

Cooperating Mutations in Pediatric Non-DS-AMKL. (a) Recurrent genes in diagnostic and relapsed specimens targeted by SNV/Indel mutations. Genes for which three or more cases carried a lesion are shown. (b) Frequency of copy number alterations for cases SNP array data and/or paired whole exome sequencing data. The outer track indicates the chromosomal location. Amplification events are shown in red, deletions in blue, and copy neutral loss of heterozygosity are shown in orange. Total number of cases carrying the event are shown, tracks do not correspond to a patient sample. (c) Non-random associations between genomic AMKL subgroup and cooperating mutation. Circos plot showing co-occurrence in patients at diagnosis between grouped (n) cooperating mutations (top) and AMKL subgroup (bottom). n, number of genes within cooperating gene sets: *CTCF*/Cohesin: *CTCF, STAG2, STAG3, SMC1A, NIPBL, SMC3, RAD21*; JAK-STAT: *JAK1, JAK2, JAK3, STAT5B*; RAS: *NRAS, KRAS, PTPN11*. Global association p value of 2.8×10^{-8} (i.e. probability of random distribution) is estimated according to global test using a multinomial regression model[21]. Individual associations: *, p <0.05; **, p<0.01, Fisher exact test.

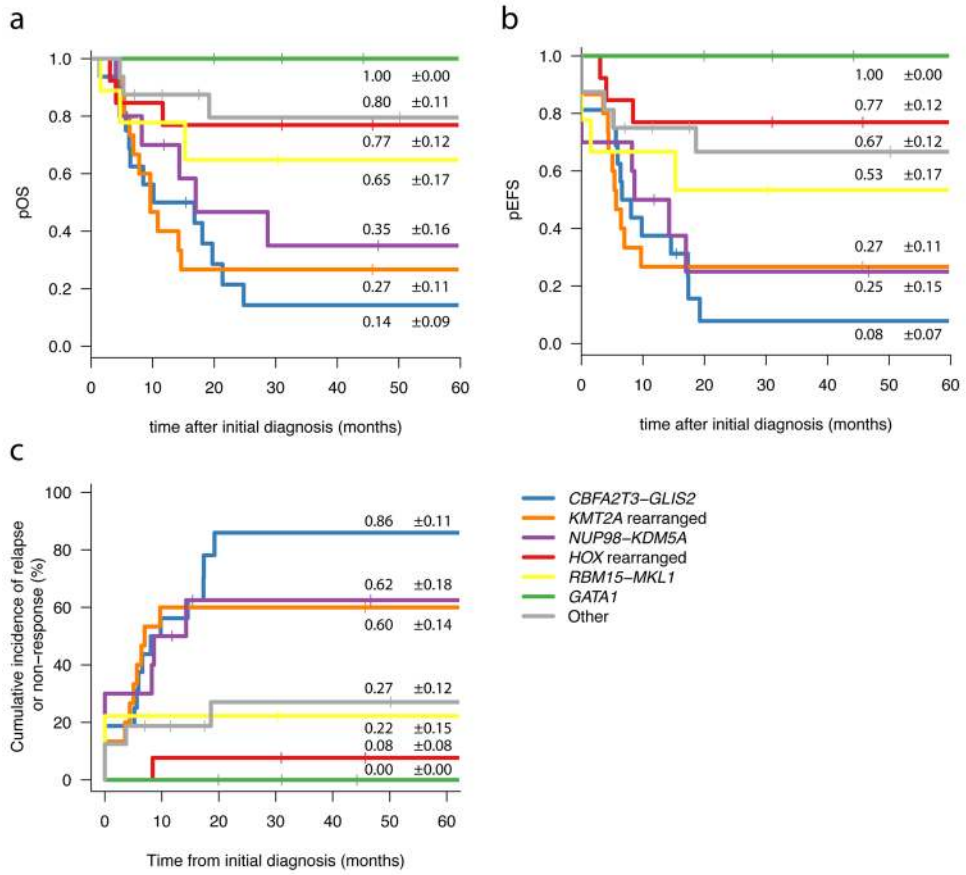

Figure 4.

Clinical Outcomes in Pediatric Non-DS-AMKL. (a) Probability of overall survival of pediatric non-DS-AMKL patients stratified according to the molecular subgroup. *CBFA2T3-GLIS2* (n=16); *KMT2A*r (n=15); *NUP98-KDM5A* (n=10); *HOX*r (n=13); *RBM15-MKL1* (n=9); *GATA1* (n=8); Other (n=16). Medium follow up time: 89 months. (b) Probability of event free survival of pediatric non-DS-AMKL patients stratified according to molecular subgroup. (c) Probability of cumulative incidence of relapse or primary resistance.

Table 1

HOX Cluster Gene Rearrangements Identified in Pediatric AMKL Patients

Chimeric Transcript	Junction	Breakpoint	Patient Sample	Predicted Product
GATA2-HOXA9	e4-e2	TTCA**G/AT**AAC	SJAMLM7001_D1	Protein Coding
GATA2-HOXA10	e4-e2	TTCA**G/GC**AAT	SJAMLM7011236_D1	Protein Coding
NIPBL-HOXA9	e6-e1	ACAA**G/TT**GAT	SJAMLM7010953_D1	Protein Coding
NIPBL-HOXB9	e6-e2	ACAA**G/CC**AAC	SJAMLM7008_D1SJAMLM7011234_D1	Protein Coding
EP300-HOXA7	e31-5′UTR	TGGG**A/TT**CAA	SJAMLM7029_D1	Protein Coding
EWSR1-HOXB8	e12-5′UTR	TT**GAT/CCC**CCA	SJAMLM7010954_D1	Protein Coding
BMP2K-HOXD10	e15-e2	TTCA**G/AG**GAA	SJAMLM7011322_D1	Protein Coding
HOXA11-BZW2	e1–e2	CTCC**A/AA**ATT	SJAMLM7010963_D1	Protein Coding
C8orf76-HOXA11-AS	e1–e1	ATCAG/GAGGT	SJAMLM7011_D1	Non-coding RNA
PLEK-HOXA11-AS	e1–5′UTR	AGAAG/GAGGT	SJAMLM7010964_D1	Non-coding RNA
HOXA9-ANGPT1	3′UTR-intergenic	AGGGT/GGAAA	SJAMLM7011109_D2	Unknown
HOXA10-AS-CD164	5′UTR-e3	TCCAG/ATGAG	SJAMLM7011248_D1	Non-coding RNA