Comprehensive characterization of pediatric acute myeloid leukemia reveals novel molecular features and age-specific interactions

Hamid Bolouri[1#*], Jason E Farrar[2#], Timothy Triche Jr[3#], Rhonda E Ries[4#], Emilia L Lim[5]; Todd A Alonzo[6,7]; Yussanne Ma[5]; Richard Moore[5]; Andrew Mungall[5]; Marco A Marra[5]; Jinghui Zhang[8]; Xiaotu Ma[8]; Yu Liu[8]; Yanling Liu[8]; Jaime M Guidry Auvil[9]; Tanja M Davidsen[9]; Patee Gesuwan[9]; Leandro C Hermida[9]; Bodour Salhia[10]; Stephen Capone[3]; Giridharan Ramsingh[3]; Christian Michel Zwaan[11]; Sanne Noort[11]; Stephen R Piccolo[12,13]; E Anders Kolb[14]; Alan S Gamis[15]; Malcolm A Smith[16]; Daniela S Gerhard[9]; and Soheil Meshinchi[4*]

[1]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA; [2]Winthrop P Rockefeller Cancer Institute, University of Arkansas for Medical Sciences and Arkansas Children's Research Institute, Little Rock, AR; [3]Jane Anne Nohl Division of Hematology, USC/Norris Comprehensive Cancer Center, Los Angeles, CA; [4]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA; [5]Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, BC, Canada; [6]Keck School of Medicine, University of Southern California, Los Angeles, CA; [7]Children's Oncology Group, Monrovia, CA; [8]Division of Computational Biology, St Jude Children's Research Hospital, Memphis, TN; [9]Office of Cancer Genomics, National Cancer Institute, Bethesda, MD; [10]Department of Translational Genomics, Keck School of Medicine, University of Southern California, Los Angeles, CA; [11]Dept of Pediatric Oncology, Erasmus MC-Sophia Children's Hospital, Rotterdam, [12]Department of Biology, Brigham Young University, Provo, UT; [13]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT; [14]Nemours Center for Cancer and Blood Disorders, Alfred I. DuPont Hospital for Children, Wilmington, DE; [15]Division of Hematology/Oncology/Bone Marrow Transplantation, Children's Mercy Hospitals and Clinics, Kansas City, MO; [16]Cancer Therapy Evaluation Program, National Cancer Institute, Bethesda, MD

# Equal contribution
* corresponding author

Abstract

We present the molecular landscape of pediatric acute myeloid leukemia (AML), characterizing nearly 1,000 participants in Children's Oncology Group (COG) AML trials. The COG/NCI TARGET AML initiative assessed cases by whole-genome, targeted DNA, mRNA, and miRNA sequencing, and CpG methylation profiling. Validated DNA variants revealed diverse, infrequent mutations, with fewer than 40 genes mutated in >2% of cases. We find that somatic structural variants, including novel gene fusions, and focal *MBNL1*, *ZEB2,* and *ELF1* deletions are common in younger patients, whereas short sequence variants predominate in adults. Mutations of *DNMT3A* and *TP53,* common in adults, are conspicuously absent from virtually all pediatric patients. Pediatric AML harbors novel *GATA2, FLT3*, and *CBL* mutations, recurrent *MYC*-ITD, frequent *NRAS*, *KRAS,* and *WT1* mutations, and recurrent promoter hypermethylation of activating NK cell ligands. Across all age groups, we find distinct molecular alterations with clinical implications, suggesting a need for age-specific targeted therapies.

Acute leukemia is the most common form of childhood cancer[1], and its incidence is increasing. Despite constituting only 20% of pediatric acute leukemia, AML is overtaking acute lymphoblastic leukemia as the leading cause of childhood leukemic mortality, in part because current prognostic schemas classify a large proportion of children who will ultimately succumb to their disease as low- and intermediate-risk. Additionally, with the exception of investigational tyrosine kinase inhibitors for *FLT3*-activated AML, targeted therapies are not used in pediatric AML. Both problems stem from an inadequate understanding of the biology of childhood AML.

AML is a molecularly heterogeneous group of diseases that affects patients of all ages. Recent genome-scale studies have elucidated novel, potentially targetable mutations prevalent in adult *de novo* AML[2-4]. However, the relevance of these findings to childhood AML remains unclear, since several of the most common adult mutations appear much less prevalent or absent in pediatric AML[5,6].

To date, no comprehensive characterization of pediatric AML has been described. In this study, we report the initial results of the TARGET (Therapeutically Applicable Research to Generate Effective Treatments) AML initiative, a collaborative COG/NCI project to comprehensively characterize the mutational, transcriptional and epigenetic landscape of a large, well-annotated cohort of pediatric AML samples obtained at diagnosis from patients treated on COG clinical trials.

Results

Overview of cohort characteristics

A total of 1023 children enrolled in COG studies are included in the TARGET AML dataset. Comprehensive clinical data are available for 993 patients. Of these, 815 subjects were profiled for somatic mutations at presentation with AML: 197 by whole-genome sequencing (WGS) and 800 by targeted capture sequencing (TCS), at read depths averaging 500x, for validation of mutations identified by WGS. The WGS discovery cohort of diagnostic and remission (germline comparison) specimens were selected from patients treated on recent COG studies who achieved an initial remission to induction chemotherapy. These trials randomized type or timing of induction therapy (CCG-2961)[7] and the addition of gemtuzumab ozogamicin in a single arm pilot (AAML03P1)[8] or randomized fashion

(AAML0531)[9]. Specimens for TCS validation were obtained from 800 patients, including 182 from the WGS discovery cohort and an additional 618 from COG AAML0531. A complete listing of cases and their characterization is available in the TARGET sample matrix (https://ocg.cancer.gov/programs/target/data-matrix). The age at presentation of TARGET AML participants ranged from 8 days to 29 years (median 10 years, **Fig 1a**). We observed distinctive differences among infants, children and adolescents/young adults (AYA) in the distribution of cytogenetic abnormalities and risk-group classification (**Fig 1a**), as established by clinically-evaluated structural abnormalities and mutations (summarized in **Fig 1b**). Notably, of these abnormalities, only 5 mutations and 5 structural aberrations occur in more than 5% of patients.

We validated each class of somatic DNA sequence alteration discovered by WGS through secondary assays. WGS single nucleotide variants (SNVs) and short insertions and deletions (indels), were confirmed by TCS. WGS-detected copy number alterations were confirmed by GISTIC2.0 scores from SNP arrays, and WGS-detected structural changes (such as translocations and inversions) were confirmed by RNA-seq and clinical data. Across all variant types, we find >70% concordance between at least two assays in 684 samples with paired remission DNA. These variants are referred to as verified variants hereon. A high-level overview of the multiplatform-verified somatic DNA variants in 684 patients is presented in **Fig 1c**. Roughly a quarter of patients possess normal karyotype, yet nearly all revealed at least one recurrent verified somatic DNA alteration, and at least 12 common cancer-associated cellular processes in pediatric AML are recurrently impacted (**Fig 1d, Table S1**).

We carried out analyses of microRNA, mRNA, and/or DNA methylation in 412 subjects. A summary of the assays performed and case-assay overlap is presented in **Fig. S1**. We compared our verified variants to those of 177 adult AML cases from the Cancer Genome Atlas (TCGA) project[2], stratified by the age groupings outlined above (**Table S2**). The TARGET and TCGA discovery cohorts both contained numerous AYA patients. Importantly, our conclusions regarding the molecular characteristics of this age group are identical when analyzing either or both cohorts (**Fig. S2**).

Somatic gene mutations in pediatric AML

Like adult AML, pediatric AML has one of the lowest rates of mutation among 28 molecularly well-characterized cancers (**Fig. 2a**). However, the landscape of somatic variants in pediatric AML is markedly different from that reported in adults[2,3], with significant variation across age groups (**Figs. 2b, S3-S4, Table S3**). Indeed, the number of coding SNVs, within and across cohorts, is best predicted by age (**Fig. 2c**, $p < 10^{-15}$) and by cytogenetic subgroup. In contradistinction to the higher prevalence of small sequence variants in older patients, recurrent structural alterations, fusions, and focal copy number aberrations are more common in younger patients (**Fig. 2d-e**, $p < 10^{-3}$, see below). Patients with *CBFA2T3-GLIS2*, *KMT2A,* or *NUP98* fusions tend to have fewer mutations ($p < 10^{-9}$), with subgroups demonstrating inferior clinical outcome (**Fig. S5**). Patients with core binding factor rearrangements tend to have more mutations than expected for their age ($p < 10^{-15}$), yet more favorable outcomes. The mutational spectrum of coding SNVs (**Fig. S4**) accumulates C→T transitions with age ($p < 10^{-3}$), further favoring C→A transversions in t(8;21) ($p < 10^{-2}$) and aberrant karyotype ($p < 10^{-2}$) patients.

After adjustment for cytogenetics and multiple comparisons, *NRAS* ($p < 10^{-3}$) and *WT1* ($p < 10^{-3}$) are mutated significantly more often in younger patients, while *DNMT3A* ($p < 10^{-23}$), *IDH1|2* ($p < 10^{-4}$), *RUNX1* ($p < 10^{-4}$), *TP53* ($p < 10^{-4}$), and *NPM1* ($p < 0.03$) are mutated significantly more frequently in older patients. *KRAS, CBL, GATA2, SETD2,* and *PTPN11* mutations appear to be more common in younger patients

(0.05<p<0.1, adjusted, **Figs. S6-S7**). Additionally, previously unreported *MYC* internal tandem duplications appear exclusively in children (**Fig. S6**). These observations are replicated in comparisons to an independent ECOG cohort (**Fig. S8a**) of 384 adult AML patients[10]. Since gene fusions have characteristic cooperating mutations[11], we devised a weighted resampling scheme to compare mutation frequencies in 584 TARGET and 131 TCGA AML cases while controlling for karyotypic associations. The results (**Fig. S8b**) confirm the generality of the pediatric-adult differences identified above.

For genes such as *CBL, GATA2, WT1, MYC* and *FLT3*, not only the frequency, but also the sites of mutation differ between children and adults (**Figs. S6 and 3a**), with multiple, frequently recurrent, alterations distinct from those identified in adult AML. *RAS*-related mutations (mutant *KRAS, NRAS, PTPN11*, or *NF1*) are common, particularly with *KMT2A* fusions (**Fig. S9, Tables S3-S5**). *WT1* mutations are also more common, more varied, and appear to arise earlier (**Fig. 3a-b**) in younger patients. These differences are likely to have significant clinical impact: we have previously shown that novel *FLT3* mutations are functional and yield poor responses to standard therapy[12]. Additionally, the impact of *FLT3*-ITDs on clinical outcome, a well-described adverse prognostic factor, is significantly modulated by co-occurring variants, including *WT1* and *NPM1* mutations and *NUP98* translocations. As shown in **Figs. 3c and S10-S11**, three independent, large-scale studies demonstrate that *FLT3*-ITD accompanied by *NPM1* mutations is associated with favorable outcome in pediatric patients, while *FLT3*-ITD with *WT1* mutations or *NUP98-NSD1* fusions yields poorer outcomes than *FLT3*-ITD alone. A propensity for clonal diversification is evident, with a majority of patients presenting with 2 or more clones when assessed by variant allele frequencies (**Fig. S12**).

We found no coding mutations of *DNMT3A* in any pediatric AML patient, despite their frequency in adults. Spontaneous deamination of 5-methylcytosine is strongly associated with aging, and *DNMT3A* contains a CpG dinucleotide yielding hotspot R882 mutations by C-to-T deamination[13]. *DNMT3A* also directly interacts with *TP53*[14], itself impacted far more frequently in adults. Mutations of *DNMT3A* or *TP53* drive clonal hematopoiesis in many apparently healthy adults[15] but are rare in children, as are the *IDH1* and *IDH2* mutations with which they often co-occur.

The spectrum of somatic structural DNA changes in pediatric AML

Many pediatric AML cases harbor chromosomal abnormalities distinct from those previously reported in adult patients (**Fig. 4a**). Among the 197 cases assayed by WGS, we identified 14 novel focal deletions involving *MBNL1*, a splicing regulator, or *ZEB2*, a key regulator of normal[16] and leukemic[17] hematopoiesis (**Fig. S13**). Despite occurring on separate chromosomes, in regions devoid of other deletions, *MBNL1:ZEB2* co-deletions occur far more often than expected (p<10[-13]). Half of these accompany *KMT2A-MLLT3* fusions (p=0.035, **Fig. S9, Tables S4-S5**). Another 15 novel, validated focal deletions specifically impact *ELF1*, an ETS-family transcriptional regulator of *MEIS1*[18]. A statistically significant difference in *ELF1* mRNA expression exists between *ELF1*-deleted and intact samples (p<0.01), with 63 genes differentially expressed between the two groups (p<0.01, **Fig. S14**). Among other novel recurrent copy losses, we note four heterozygous deletions of a region containing the *IL9R* gene (**Table S4**) co-occurring with *KIT* mutations and t(8;21).

Consistent with our previous findings regarding *NUP98-NSD1* fusions[19], adult-pediatric genomic differences also extend to translocations. An expansive catalog of gene fusions, many observed primarily or exclusively in pediatric cases, underscores the disproportionate impact of structural variants in younger patients (**Fig. 4b-c**).

Patterns of mutational cooperation are not limited to patients with recurrent structural alterations (**Tables S4-S5**). Mutant *GATA2* is frequently seen in children with normal karyotype (NK) AML. Both *GATA2* ($p<10^{-9}$) and *CSF3R* ($p<10^{-6}$, **Fig. S15**) mutations co-occur with mutations of *CEBPA*[20]. *GATA2* and *CEBPA* are key regulators of hematopoiesis[21,22], both interacting with *RUNX1* in normal hematopoiesis and leukemogenesis[23]. As with *FLT3/NUP98-NSD1/WT1* interactions, these findings show prognostic interactions in pediatric AML outcome (**Fig. S15b**). *RUNX1* mutations and *RUNX1-RUNX1T1* gene fusions are significantly exclusive of *GATA2* and *CEBPA* mutations (p=0.006, **Fig. S16, Table S6**). All four are significantly exclusive of *KMT2A* rearrangements ($p<10^{-15}$), *CBFB-MYH11* gene fusions ($p<10^{-11}$), and *ETV6* aberrations (p=0.01).

DNA methylation subtypes in pediatric AML

By combining DNA methylation and mRNA expression results in 395 TARGET and TCGA AML cases, we identified dozens of genes with recurrent transcriptional silencing via promoter hypermethylation across both TARGET and TCGA AML patients (**Fig. 5a, Fig S17**). A number of samples exhibited widespread silencing of genes by aberrant promoter hypermethylation, and this group is enriched for younger patients with *WT1* mutations (p=0.0012, **Fig. 5a** boxed region). Aberrant *Wnt*/β-catenin signaling is required for the development of leukemic stem cells[24], and one or more of *DKK1, SMAD1, SMAD5, SFRP4, SFRP5, AXIN2, WIF1, FZD3, HES1,* or *TLE1* is deleted or aberrantly methylated in most AML cases, with methylation predominant in younger cases. Furthermore, repression of NK cell ligands occurs in many younger patients. As with other recurrently silenced genes (e.g. *CDKN2A*), the NK ligands *ULBP1/2/3* and *MICA* are expressed at low levels in healthy cells, but appear to trigger condition-specific clearance of damaged hematopoietic progenitors (**Fig S18**; Capone, *et al* in review). Their repression is most common in children and adolescents, perhaps reflecting different stages of immune development[24].

We applied non-negative matrix factorization (NMF) to CpG methylation data from 313 TARGET and TCGA AML patients for whom we had genome-scale DNA methylation data. By cross-validation, we determined that 31 signatures best captured the information in the results, after controlling via *in silico* purification for differences in cellularity. Unsupervised clustering of the resulting DNA methylation signatures largely separated patients by age within karyotypic subtypes (**Fig 5b**), but also revealed several signatures which did not associate strongly with common recurrent mutations, gene fusions, cytogenetic aberrations, or other known prognostic factors. Two such signatures (signatures 2 and 13) predicted significantly (p < 0.05) poorer event-free survival in both pediatric and adult patients with above-median scores, after stratifying by cohort and adjusting for both TP53 mutation status and white blood cell count (**Figure S19**). The clinical significance of these signatures will require larger sample sizes to evaluate.

The pediatric AML transcriptome is shaped by diverse miRNAs

To characterize miRNA expression patterns in pediatric AML, we analyzed miRNA expression in 152 cases using miRNA sequencing. Unsupervised clustering of the data revealed 4 discrete subgroups that were correlated with specific genomic alterations (**Fig 6a** and **S20**). For instance, we observed that abundant miR-10a expression accompanied *NPM1* mutations, in agreement with previous reports[25]. Further, Cox proportional hazards analyses identified multiple miRNAs impacting clinical outcome (**Figs S21-S23, Table S7**). Among these was miR-155, which we have previously reported to be associated with poor survival[26].

Differential expression analyses using Wilcoxon tests revealed miRNAs that are differentially expressed between pediatric and adult AML (**Fig. 6b**). Of note, miR-330 was the most over-expressed in pediatric samples when compared to adult samples, and has previously been shown to have oncogenic potential in AML[27].

Several age-associated miRNAs harbor binding sites within, and have expression levels anti-correlated with, putative target genes that may be involved in RNA and protein processing (**Fig. 6c**), suggesting that miRNAs could contribute to leukemogenesis through the dysregulation of transcripts and proteins. These findings are in agreement with our previous report showing that miRNA expression patterns are associated with treatment resistance and metabolic dysregulation in pediatric AML[28]. In particular, *let-7b*, which is more abundantly expressed in older patients, has *EIF2S3* (a regulator of protein synthesis) as a putative target (**Fig. 6d**). In addition, *let-7b* expression is associated with shorter time to relapse (multivariate p<0.05, **Fig 6e)** in pediatric patients.

Discussion

This study establishes the prevalence of, and coincident relationships among, recurrent somatic genetic and epigenetic abnormalities in pediatric AML from a large cohort of patients. We observe many features in common between pediatric and adult AML: both share a low overall mutation rate in comparison to other cancers, a long tail of infrequently affected genes, and overlap among recurrently impacted genes. However, pediatric AML exhibits distinctive and critically important differences as well. We and others have previously reported on the presence and clinical impact of novel fusion genes in pediatric AML[19,29]. As this study comprehensively illustrates, the impact of fusion transcripts is both broad and age-dependent in pediatric AML. Recognition and comprehensive testing for these alterations are key first steps in the development of new and potentially novel modes of targeted therapy[30].

Recurrent copy loss also represents a unique aspect of pediatric in contrast to adult AML. Regional (e.g. chromosomal arm- and band-level) copy loss differs substantially, but surprisingly, focal areas of copy loss are also more common in children, specifically impacting *ZEB2*, *MBNL1*, and *ELF1*. *MBNL1* is upregulated by the *KMT2A-AF9* fusion protein[31], and genes involved in post-transcriptional processing (*SETD2, U2AF1, DICER1*) harbor the sole recurrent mutation in many *KMT2A*-rearranged cases, suggesting a functional role for altered splicing in pediatric leukemogenesis. Alterations in *ZEB2* have been identified as cooperating events in murine *CALM-AF10* leukemia models[32] while *ZEB2* knockout mice develop myelofibrosis[33], suggesting a fundamental role for this gene in the pathogenesis of AML.

Many of the genes characteristically mutated in AML are altered at widely variable frequencies across age groups; several (including *FLT3* and *WT1*) are impacted by pediatric specific variants and hotspots.

Clinical tests for a handful of genomic alterations are widely used to further risk-stratify patients and determine the treatment regimens. However, the current practice of considering the effect of each somatic alteration in isolation is inadequate and potentially misleading: interactions among sequence variants can have dramatic clinical consequences. In pediatric AML, the co-occurrence of *FLT3*-ITD and *NPM1* mutations is associated with superior outcomes (**Figs. 3c, S10 and S11**), in striking contrast to the inferior outcomes reported in adults with this combination. In the TCGA AML cohort, over half the subjects with somatic *FLT3* and *NPM1* mutations also possessed somatic *DNMT3A* mutations[2]. Subsequent studies established the generality of this result[3], and revealed that *DNMT3A* mutations are

early clonal events[34], which often co-operate with later *NPM1* and *FLT3* mutations to promote chemoresistance and mutagenesis[35] while conferring inferior outcomes[36]. Similarly, the co-occurrence of *FLT3*-ITD with *WT1* mutations or *NUP98-NSD1* fusions accompanies frequent induction failure and dismal outcomes in children with AML (multivariate $p<10^{-4}$, **Figs. 3c, S10 and S11**).

In TARGET, TCGA, and ECOG AML cases, *WT1* mutations were mutually exclusive with those of *ASXL1* and *EZH2* ($p < 10^{-3}$). *WT1* recruits *EZH2* to its targets[37], and *WT1* mutations have been linked to hypermethylation of *EZH2* target genes[38]. Mutant *ASXL1* abolishes *EZH2*-mediated silencing of HOX genes[39]. *EZH2* resides on a recurrently deleted region of chromosome 7, and decreased *EZH2* activity is associated with treatment resistant AML[40]. In pediatric AML, mutant *WT1* and *EZH2* were both of exclusively clonal or near-clonal origin, with nearly a quarter of TARGET cases harboring mutations affecting one or the other. Aberrant *WT1*, *EZH2*, or *ASXL1* predicted induction failure in TARGET AML cases (multivariate $p<0.05$, adjusted for interactions with *FLT3* alterations, *NUP98-NSD1*, and *KMT2A* fusions) and were mutually exclusive with *KMT2A* fusions ($p < 10^{-5}$). Many of these patients possess apparently normal karyotypes, yet less than 20% achieve long-term remission with standard treatment, highlighting the importance of molecular stratification to achieve better outcomes. It is tempting to speculate that early events such as *WT1* mutations and *NUP98-NSD1* fusions in children may play a similar role to that observed for DNMT3A mutations[13] in adults, with significant implications for risk stratification in AML across age groups.

Our data also demonstrate that DNA-methylation and miRNA expression profiles both accompany and complement DNA alterations, and can stratify pediatric AML patients in terms of both overall and progression-free survival. These findings suggest a need to update pediatric AML clinical risk categories beyond current classifications, with important implications for clinical practice.

Despite incremental improvements with increasingly intensified regimens, modern outcomes in pediatric AML have plateaued, with only 60% of patients achieving long term survival. As many as 10% of children will die from direct complications of treatment. Survivors suffer unacceptably high rates of long-term morbidities resulting from anthracycline exposure or sequelae of hematopoietic stem cell transplantation. As illustrated herein, pediatric AML is a collection of molecularly diverse diseases with similar phenotypes, which may explain repeated failure of randomized clinical trials improve outcomes in recent years. No single treatment strategy is likely to be effective for all pediatric AML subtypes. Recent studies have promulgated a shift towards comprehensive, molecularly based classification schemas in AML[3]. The time has come to develop targeted therapies that address the specific vulnerabilities of pediatric AML subtypes. The TARGET AML dataset will serve as a foundation for development of pediatric-specific classification schemas and, hopefully, the development of personalized treatment strategies.

Acknowledgements

Figure Legends

Figure 1. An overview of the TARGET AML study. (**a**) TARGET cases represent patients ranging from 8 days to 29 years of age with a median age of 10 years. The distribution by risk strata (as defined in the treating study) and cytogenetic classification is shown adjacent to each age group analyzed (Infant, <3 years; Child, 3 to <15 years; Adolescent/Young Adult (AYA), 15 to <40 years. (**b**) Clinically established molecular aberrations in the cohort (n=993) include FLT3 internal tandem duplications and point mutations within the activating loop (FLT3.ITD and FLT3PM), cytogenetic abnormalities including *KMT2A* fusions, t(8;21), inversion 16, trisomies 8 and 21, loss of sex chromosomes, and monosomies 5 and 7. Selected mutational data available include *WT1*, *CEBPA*, and *KIT* mutations. (**c**) At least one variant impacting a gene recurrently altered in pediatric AML was identifiable in 684 patients with multi-platform validated variants. Junction, protein fusions (see methods); chromCNV, chromosomal arm/band level copy variant; focalCNV, gene level copy variant; indel, small insertion/deletion; SNV, single nucleotide variant. (**d**) The overall impact of gene alterations on cellular processes and pathways.

Figure 2. Age-related differences in mutational and structural alterations in AML. (a) The mutational burden of SNV and indels is low in pediatric AML (in blue), with a median of 10 mutations/case across the 197 sample WGS cohort. This places pediatric AML, along with other pediatric malignancies (rhabdoid tumor, Ewing sarcoma, medulloblastoma) and adult AML (in red) among the least mutated of human cancers. (**b**) While pediatric and adult AML share an overlapping spectrum of commonly mutated genes, there are marked age-dependent differences in the prevalence of mutations, with numerous genes mutated at higher rates in the TARGET over TCGA cohort or vice versa. *FLT3* mutations are plotted in 3 categories: internal tandem duplication (ITD; FLT3.ITD), activation loop domain (FLT3.C), and novel, childhood-specific changes (FLT3.N). (**a, inset**) A clear pattern of waxing or waning mutation rates spanning the developmental spectrum is evident in *KRAS* and *NPM1*, among several additional genes. (**c**) Childhood AML, like adult AML, develops in the overall context of a low somatic mutational burden, often < 1 somatic, protein-coding change per megabase, with progressively greater mutational overhead and fewer recurrent cytogenetic alterations with advancing age. For color key, see legend on (e). (**d**) Enumeration of the ratio of recurrently impacted AML genes by either mutation or structural alteration highlights these changes across time. While both pediatric and adult AML are highly heterogeneous diseases, a loess fit illustrates that the burden of structural variation tends to exceed that of sequence variants in infancy and early childhood. Adolescents & young adults (AYA) exhibit roughly equal burden from structural and sequence variants, while in older adults, sequence variants often contribute a larger share of the mutational burden than do structural alterations. For color key, see legend on (e). (**e**) These findings are highlighted by illustration of peaks of incidence of common translocations in AML across ages. Using a sliding-window approach to account for uneven sampling by age, we fit a loess curve to the proportion of cases affected by fusions involving recurrent partner genes; *KMT2A* fusions are most common in infants, while core binding factor fusions tend to affect older children.

Figure 3. Biological and prognostic interactions between *WT1*, *NPM1*, *FLT3*-ITD and *NUP98-NSD1* alterations. (**a**) *WT1* mutations appear not only more frequently (TCGA – 7.3%, 13 alterations among 177 patients; TARGET – 18.4%, 150 alterations among 815 patients), but also impact novel sites in childhood AML, suggesting a distinctive biology to these alterations in pediatric compared to adult-onset AML. (**b**) In support of age-related functional differences in the impact of typical myeloid driver genes, mutation-based inference of clonality in 197 TARGET AML cases with WGS and 177 TCGA AML cases

demonstrates that the likelihood of subclonal mutations in characteristic driver genes, such as *WT1*, varies with age. (**c**) Given the strong co-association of the well-established adverse risk factor, *FLT3* internal tandem duplication (ITD) with *NUP98* translocations, *WT1* and *NPM1* mutations, we tested the contribution of these variants to outcome. In 963 TARGET patients with complete data for *FLT3*-ITD, *NPM1* and *WT1* mutation and *NUP98-NSD1* fusion available for analysis, those with *FLT3* ITD plus *WT1* and/or *NUP98-NSD1* fusion (n=73) exhibit markedly inferior event-free (p=<0.0001) and overall survival (see **Fig. S10**). Additionally, the salutary effect of *NPM1* mutations is evident in this cohort, with pediatric patients possessing *FLT3* ITD and *NPM1* mutations (n=27) faring better than *FLT3*-ITD negative patients (n=791). To assess the generality of these findings, we separately analyzed the outcomes of children treated on AAML0531 and CCG-2961 with complete molecular data (including TARGET and additional non-TARGET cases) as well as an independent cohort of patients treated on European cooperative group trials (see online methods). The interactions identified in the TARGET cohort appear broadly applicable across each of these studies.

Figure 4. Chromosomal alterations in AML. AML at any age is often associated with numerical partial or whole chromosome abnormalities, however the patterns of gain (**a**, outward projection) and loss (inward projection) differ between the TARGET and TCGA AML cohorts, with losses of 5q, 7, and 17 predominant in adults, while gains of 4, 6, 19 and losses of 9, X, and Y are more characteristic of AML in younger patients. (**b**) Recurrent fusion gene neighborhoods identified from TARGET and TCGA AML subjects include well-recognized AML fusion targets involving core binding factors (CBF families) and 11q23 (*KMT2A*), along with novel and more recently recognized fusions of nucleoporin genes (NUP), *GLIS2*, and ETS transcription factors, and finally, a group of partners not defined by major categories. Gene symbol font size illustrates the overall prevalence of involvement, and connecting line widths reflect fusion frequencies. (**c**) Chromosomal translocations result in characteristic fusion genes, both novel and recurrent, which tend to affect specific age ranges, including a large group of fusion genes observed exclusively in pediatric cases. Age-specific enrichment for each fusion pair is indicated by white-red shading while the colored ribbon below reflects the fusion family of each pair, as in (**b**).
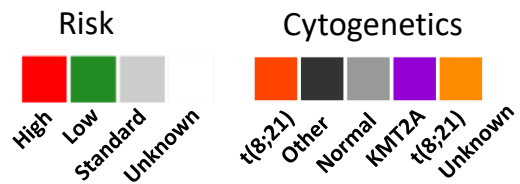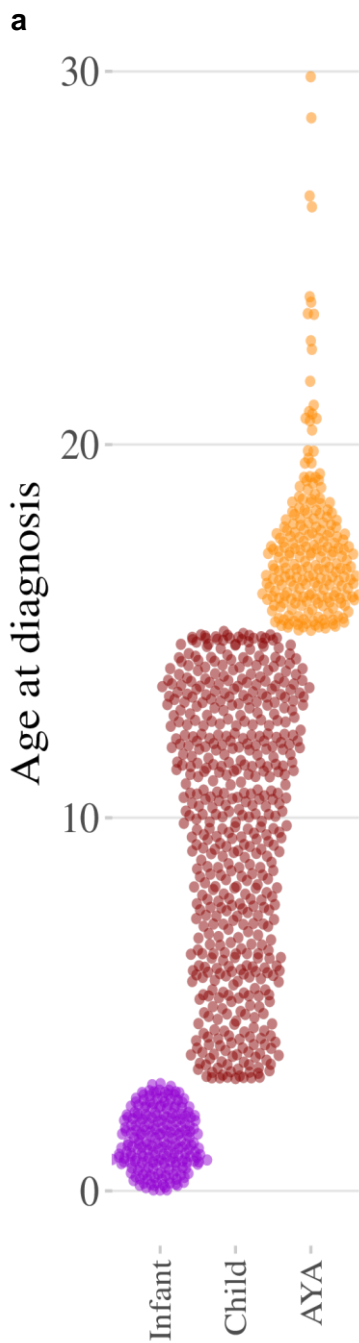
Figure 5. Comparison of DNA methylation in adult and pediatric AML. (**a**) Integration of mRNA expression data with promoter methylation data identifies numerous genes subject to methylation-mediated silencing across age groups, with a prominent group of childhood and AYA subjects showing silencing across dozens of genes (boxed region). Recurrent silencing of genes involved in innate immune response (*ULBP1/2/3*) and WNT signaling (*DKK1*), as highlighted at far right, is particularly notable. (**b**) Unsupervised hierarchical clustering of a methylation metaclassifier largely separates adult and pediatric patients within relatively cytogenetic subgroups (e.g. t(8;21) and inv(16)) and reveals DNA methylation signatures which do not correspond directly to known recurrent alterations, some (signatures 2 and 13) demonstrating potential prognostic significance.

Figure 6. miRNAs differentially regulate distinct age and molecular sub-groups in AML (**a**) Interrogation of miRNA expression patterns in childhood AML identifies major clusters with correlation to known molecular features (cytogenetics, *FLT3*-ITD). **(b)** Age-related differences in miRNA expression are evident between adult (n=162) and pediatric AML (n=152). Volcano plot indicates differentially expressed miRNAs between adult and pediatric cases. (Wilcoxon test, Benjamin-Hochberg adjusted P<0.05; Threshold indicated by dashed red line). **(c)** A predicted miRNA:mRNA target relationship involving *let-7b*, which is less abundant in most pediatric cases than in adult cases. **(d)** Abundant *let-7b* expression is associated with shorter time to relapse in pediatric AML.
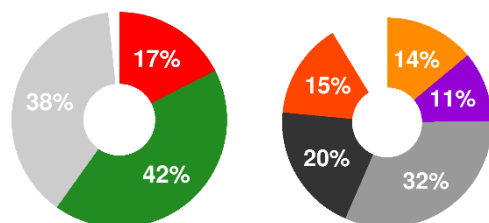
References

1. Steliarova-Foucher, E., *et al.* International incidence of childhood cancer, 2001-10: a population-based registry study. *Lancet Oncol* (2017).
2. Cancer Genome Atlas Research, N. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* **368**, 2059-2074 (2013).
3. Papaemmanuil, E., *et al.* Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* **374**, 2209-2221 (2016).
4. Patel, J.P., *et al.* Prognostic Relevance of Integrated Genetic Profiling in Acute Myeloid Leukemia. *N Engl J Med* (2012).
5. Ho, P.A., *et al.* Leukemic mutations in the methylation-associated genes DNMT3A and IDH2 are rare events in pediatric AML: a report from the Children's Oncology Group. *Pediatr Blood Cancer* **57**, 204-209 (2011).
6. Farrar, J.E., *et al.* Genomic Profiling of Pediatric Acute Myeloid Leukemia Reveals a Changing Mutational Landscape from Disease Diagnosis to Relapse. *Cancer Res* **76**, 2197-2205 (2016).
7. Lange, B.J., *et al.* Outcomes in CCG-2961, a children's oncology group phase 3 trial for untreated pediatric acute myeloid leukemia: a report from the children's oncology group. *Blood* **111**, 1044-1053 (2008).
8. Cooper, T.M., *et al.* AAML03P1, a pilot study of the safety of gemtuzumab ozogamicin in combination with chemotherapy for newly diagnosed childhood acute myeloid leukemia: a report from the Children's Oncology Group. *Cancer* **118**, 761-769 (2012).
9. Gamis, A.S., *et al.* Gemtuzumab ozogamicin in children and adolescents with de novo acute myeloid leukemia improves event-free survival by reducing relapse risk: results from the randomized phase III Children's Oncology Group trial AAML0531. *J Clin Oncol* **32**, 3021-3032 (2014).
10. Patel, J.P., *et al.* Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med* **366**, 1079-1089 (2012).
11. Faber, Z.J., *et al.* The genomic landscape of core-binding factor acute myeloid leukemias. *Nat Genet* (2016).
12. Tarlock, K., *et al.* Discovery and Functional Validation of Novel Pediatric Specific FLT3 Activating Mutations in Acute Myeloid Leukemia: Results from the COG/NCI Target Initiative. *Blood* **126**, 87-87 (2015).
13. Ley, T.J., *et al.* DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* **363**, 2424-2433 (2010).
14. Wang, Y.A., *et al.* DNA methyltransferase-3a interacts with p53 and represses p53-mediated gene expression. *Cancer Biol Ther* **4**, 1138-1143 (2005).
15. Genovese, G., *et al.* Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *N Engl J Med* (2014).
16. Goossens, S., *et al.* The EMT regulator Zeb2/Sip1 is essential for murine embryonic hematopoietic stem/progenitor cell differentiation and mobilization. *Blood* **117**, 5620-5630 (2011).
17. Goossens, S., *et al.* ZEB2 drives immature T-cell lymphoblastic leukaemia development via enhanced tumour-initiating potential and IL-7 receptor signalling. *Nat Commun* **6**, 5794 (2015).
18. Xiang, P., *et al.* Identification of E74-like factor 1 (ELF1) as a transcriptional regulator of the Hox cofactor MEIS1. *Exp Hematol* **38**, 798-798, 808 e791-792 (2010).
19. Ostronoff, F., *et al.* NUP98/NSD1 and FLT3/ITD coexpression is more prevalent in younger AML patients and leads to induction failure: a COG and SWOG report. *Blood* **124**, 2400-2407 (2014).
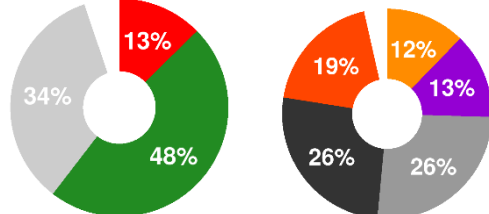
20. Maxson, J.E.*, et al.* CSF3R mutations have a high degree of overlap with CEBPA mutations in pediatric AML. *Blood* (2016).

21. Quintana-Bustamante, O.*, et al.* Overexpression of wild-type or mutants forms of CEBPA alter normal human hematopoiesis. *Leukemia* **26**, 1537-1546 (2012).

22. Vicente, C., Conchillo, A., Garcia-Sanchez, M.A. & Odero, M.D. The role of the GATA2 transcription factor in normal and malignant hematopoiesis. *Crit Rev Oncol Hematol* **82**, 1-17 (2012).

23. Ng, K.P.*, et al.* Runx1 deficiency permits granulocyte lineage commitment but impairs subsequent maturation. *Oncogenesis* **2**, e78 (2013).

24. Wang, Y.*, et al.* The Wnt/beta-catenin pathway is required for the development of leukemia stem cells in AML. *Science* **327**, 1650-1653 (2010).

25. Marcucci, G.*, et al.* MicroRNA expression in cytogenetically normal acute myeloid leukemia. *N Engl J Med* **358**, 1919-1928 (2008).

26. Ramamurthy, R.*, et al.* miR-155 expression and correlation with clinical outcome in pediatric AML: A report from Children's Oncology Group. *Pediatr Blood Cancer* **63**, 2096-2103 (2016).

27. Fooladinezhad, H., Khanahmad, H., Ganjalikhani-Hakemi, M. & Doosti, A. Negative regulation of TIM-3 expression in AML cell line (HL-60) using miR-330-5p. *Br J Biomed Sci* **73**, 129-133 (2016).

28. Lim, E.L.*, et al.* Comprehensive Sequence Analysis of Relapse and Refractory Pediatric Acute Myeloid Leukemia Identifies miRNA and mRNA Transcripts Associated with Treatment Resistance - a Report from the COG/NCI-Target AML Initiative. *Blood* **126**, 687-687 (2015).

29. Gruber, T.A.*, et al.* An Inv(16)(p13.3q24.3)-encoded CBFA2T3-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia. *Cancer Cell* **22**, 683-697 (2012).

30. Liang, K.*, et al.* Therapeutic Targeting of MLL Degradation Pathways in MLL-Rearranged Leukemia. *Cell* **168**, 59-72 e13 (2017).

31. Itskovich SS, C.J., Mulloy JC, Disney MD, Kumar AR. MBNL1 As a New Therapeutic Target in MLL-Fusion Gene Leukemia. in *American Society of Hematology Annual Conference* 462 (Orlando, Fl., 2015).

32. Caudell, D.*, et al.* Retroviral insertional mutagenesis identifies Zeb2 activation as a novel leukemogenic collaborating event in CALM-AF10 transgenic mice. *Blood* **115**, 1194-1203 (2010).

33. Li, J.*, et al.* The EMT transcription factor Zeb2 controls adult murine hematopoietic differentiation by regulating cytokine signaling. *Blood* (2016).

34. Shlush, L.I.*, et al.* Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* **506**, 328-333 (2014).

35. Guryanova, O.A.*, et al.* DNMT3A mutations promote anthracycline resistance in acute myeloid leukemia via impaired nucleosome remodeling. *Nat Med* **22**, 1488-1495 (2016).

36. Loghavi, S.*, et al.* Clinical features of de novo acute myeloid leukemia with concurrent DNMT3A, FLT3 and NPM1 mutations. *J Hematol Oncol* **7**, 74 (2014).

37. Xu, B.*, et al.* Tumor suppressor menin represses paired box gene 2 expression via Wilms tumor suppressor protein-polycomb group complex. *J Biol Chem* **286**, 13937-13944 (2011).

38. Sinha, S.*, et al.* Mutant WT1 is associated with DNA hypermethylation of PRC2 targets in AML and responds to EZH2 inhibition. *Blood* **125**, 316-326 (2015).

39. Abdel-Wahab, O.*, et al.* ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. *Cancer Cell* **22**, 180-193 (2012).

40. Gollner, S.*, et al.* Loss of the histone methyltransferase EZH2 induces resistance to multiple drugs in acute myeloid leukemia. *Nat Med* (2016).

a Somatic mutations per megabase across cancer types: Rhabdoid, Pediatric AML, Ewing's, Thyroid, Medulloblastoma, Adult AML, Carcinoid, Neuroblastoma, CLL, Prostate, Low grade glioma, Breast, Pancreas, Myeloma, Renal clear cell, Renal papillary, Ovarian, Glioblastoma, Cervical, Lymphoma, Head and neck, Colorectal, Esophageal, Stomach, Bladder, Lung adeno, Lung squamous, Melanoma.

b Fraction of samples mutated (Pediatric vs Adult) for genes: NRAS, KIT, KRAS, WT1, FLT3.N, CBL, PTPN11, GATA2, ASXL2, CEBPA, SETD2, MYC, CSF3R, FLT3.ITD, IKZF1, ETV6, EZH2, ASXL1, FLT3.C, TET2, TP53, RUNX1, IDH2, IDH1, NPM1, DNMT3A.

KRAS: infant 28/118, child 43/380, AYA 13/174, adult 8/129.
NPM1: infant 3/118, child 40/380, AYA 32/174, adult 44/129.

c Mutations per megabase: Infants (< 3), Children (3−14), AYA (15−39), Adults (40+). Fraction of cases stacked bars.

d Ratio of structural to mutational burden (100:1, 10:1, 3:1, 1:1, 1:3, 1:10, 1:100).

e Fusion frequency across age groups (Infants, Children, AYA, Adults, Elderly).

Cytogenetics: Normal, Other, inv(16), t(8;21), ETS, NUP, GLIS2, KMT2A.

a

Study cohort
- TARGET (blue)
- TCGA (red)

b

NUP
- DEK
- NUP214
- HOXD13
- KDM5A
- NUP98
- HOXA9
- NSD1
- PHF23
- HMGB3

ETS
- MNX1
- FEV
- ETV6 — INO8D
- FUS
- ERG
- FLI1
- HNRNPH1

CBF
- AFF3
- RUNX1
- CBFA2T2
- MECOM
- RUNX1T1
- MYH11
- CBFB

Fusion families

11q23
- PICALM
- MLLT1
- ZBTB16
- FLNA SEPT9
- MLLT10
- FRYL
- AFF1
- RARA
- BTBD18
- SEPT6
- CT45A2
- KMT2A
- MLLT3
- TET1
- EPS15
- CEP170B
- ARHGAP26
- MLLT11
- ELL
- FNBP1
- ABI1
- MLLT4

Other
- GOSR1
- ZBTB7A
- MLF1
- LRRC37B
- ZNF207
- TNK1
- NPM1
- NF1
- RPS11
- FLT3LG
- TFG
- KAT6A — EP300
- GPR128
- CREBBP
- ZEB2 — BCL11B
- BIRC6
- RBM15
- LTBP1
- LRBA
- MKL1
- SH3D19

GLIS2
- CBFA2T3 — GLIS2

c

	CBFA2T3-GLIS2	KAT6A-EP300	KMT2A-AFF1	MNX1-ETV6	ETV6-INO8D	RBM15-MKL1	RUNX1-CBFA2T2	ZBTB16-RARA	ZEB2-BCL11B	FNBP1-KMT2A	FRYL-KMT2A	KMT2A-ARHGAP26	KMT2A-BTBD18	KMT2A-CEP170B	KMT2A-CT45A2	KMT2A-EPS15	KMT2A-RARA	KMT2A-TET1	FUS-FEV	KMT2A-MLLT10	KMT2A-SEPT6	KMT2A-MLLT1	NUP98-KDM5A	KMT2A-ELL	KMT2A-SEPT9	KMT2A-MLLT3	DEK-NUP214	NUP98-PHF23	NUP98-NSD1	NUP98-HOXD13	HNRNPH1-ERG	RUNX1-RUNX1T1	KMT2A-ABI1	KMT2A-MLLT4	KMT2A-FLNA	KMT2A-MLLT11	RUNX1-AFF3	MLF1-NPM1	FUS-ERG	CBFB-MYH11	NUP98-HMGB3	NUP98-HOXA9	KAT6A-CREBBP	MLLT10-PICALM	LRBA-SH3D19	FUS-FLI1	GPR128-TFG	TNK1-ZBTB7A	RUNX1-MECOM	NF1-LRRC37B	GOSR1-ZNF207	FLT3LG-RPS11	BIRC6-LTBP1
Infants	8	2	2	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20	3	3	5	7	1	31			2	1		2	1	1				1	23			1										
Children	1		1																	11	2	2	4	6	1	31	11	3	27	3	2	104	2	12	1	1	1	3	3	67	2	2	2	4									
AYA																			1	5		3	1	4		8	1		6			39		4				1	1	34	1	1	2	4	1	1							
Adults																				4			2			2			2			4		2				10				1					2	1	1	1	1	1	1

Enrichment (%): 0 – 100

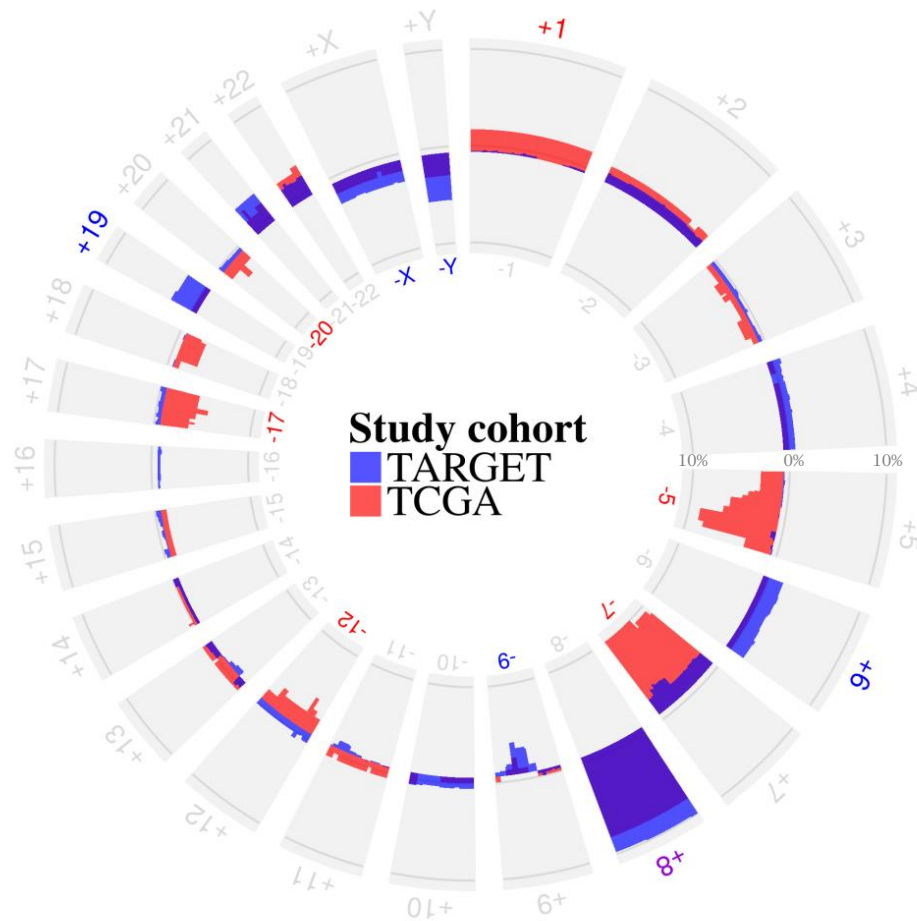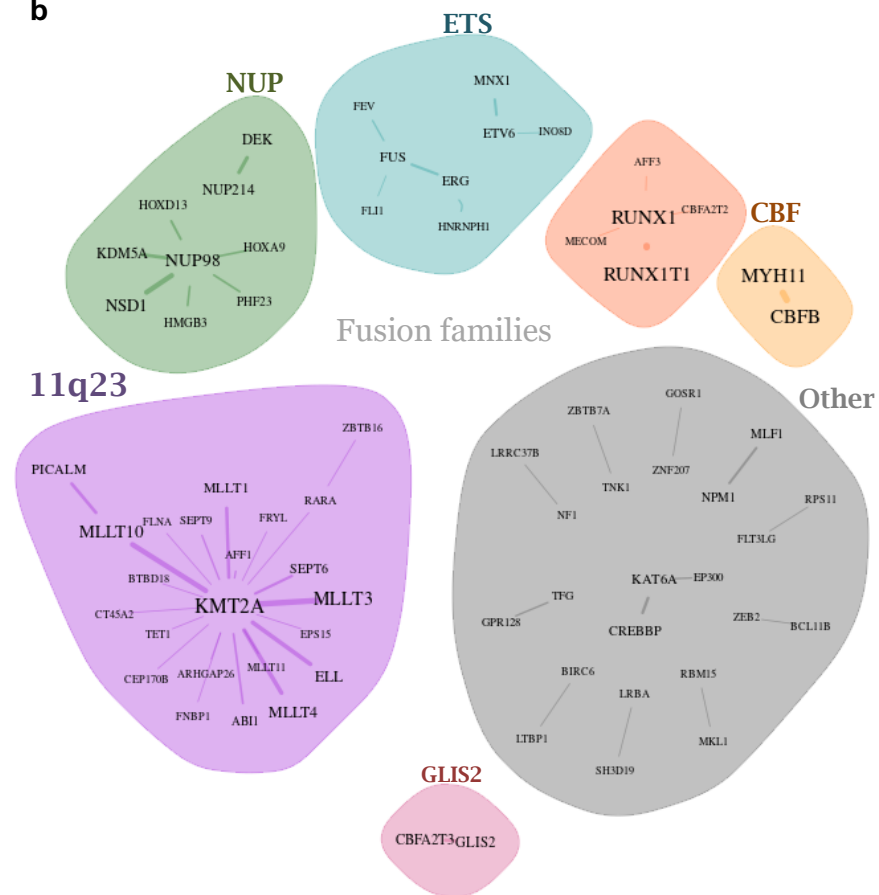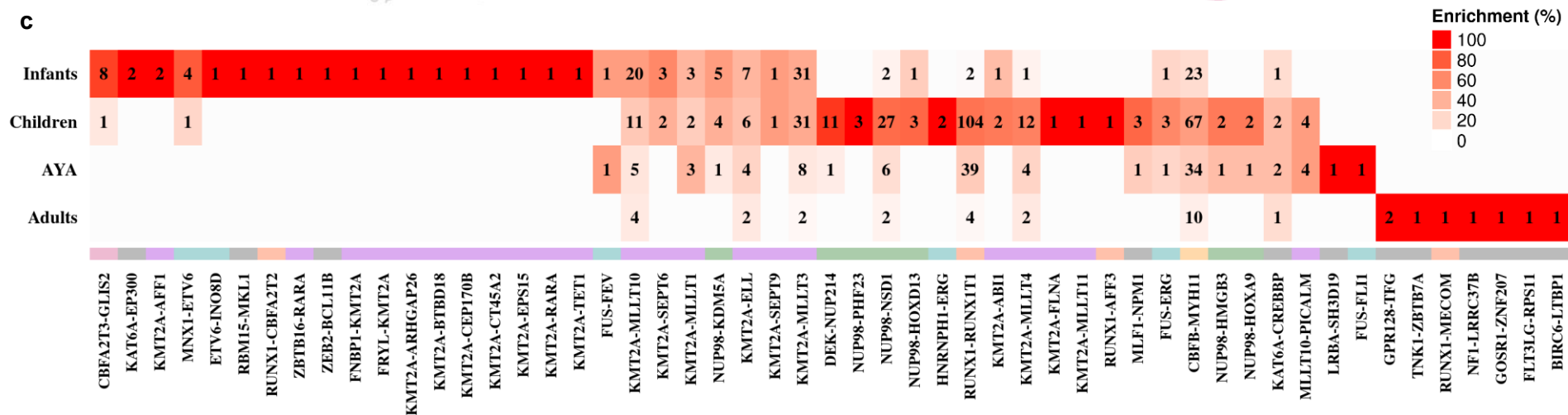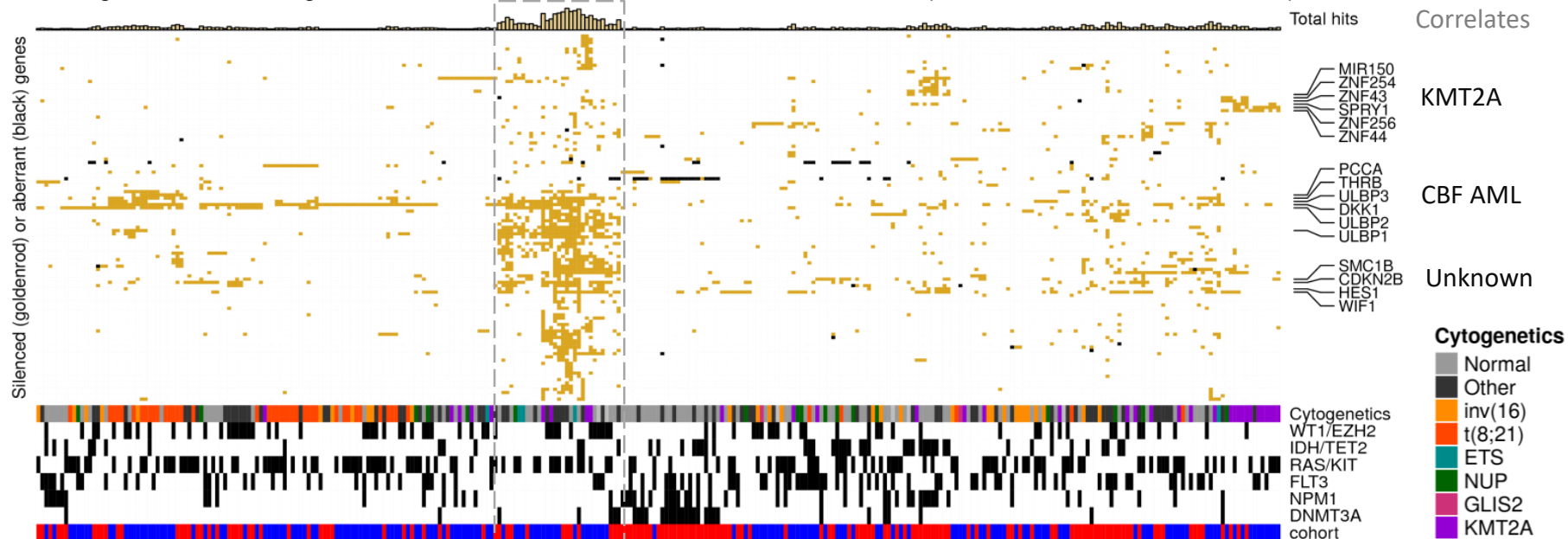**a** Silencing and aberrations of genes with recurrent loss of function in TARGET & TCGA AML (313 of 395 evaluable cases affected)

a

t(8;21)
Inv(16)
FLT3ITD
FLT3mut
NPM1
CEBPA
WT1

miR-21

miR-10a

3
0
-3

b

-log2(adjusted p-value)

150

100

50

miR-30c-5p

miR-222-3p

let-7b-3p

let-7f-5p

miR-25-5p

-2   0   2

log2 fold change(TARGET/TCGA)

c

hsa-let7b-3p

miRNA expression (RPM)

300

200

100

TARGET   TCGA

Let7b-3p vs. EIF2S3INM_001415

Spearman corr. = − 0.67

P-value < 0.001

miRNA Exp (RPM)

150

100

50

0   20   40   60   80   100 120

mRNA Exp (FPKM)

d

hsa-let7b-3p EFS

Percent survival

100

80

60

40

20

0

High expression (n=122)
Low expression (n=30)

P=0.0419

0   1000   2000   3000   4000

EFS (Days)