Comprehensive molecular and clinical characterization of *NUP98* fusions in pediatric acute myeloid leukemia

Eline J. M. Bertrums,[1,2,3*] Jenny L. Smith,[4*] Lauren Harmon,[5*] Rhonda E. Ries,[4] Yi-Cheng J. Wang,[6,7] Todd A. Alonzo,[6,7] Andrew J. Menssen,[8] Karen M. Chisholm,[9] Amanda R. Leonti,[4] Katherine Tarlock,[4,10] Fabiana Ostronoff,[11] Era L. Pogosova-Agadjanyan,[4] Gertjan J. L. Kaspers,[1,12,13] Henrik Hasle,[14] Michael Dworzak,[15,16] Christiane Walter,[17] Nora Mühlegger,[15] Cristina Morerio,[18] Laura Pardo,[8] Betsy Hirsch,[19] Susana Raimondi,[19] Todd M. Cooper,[10] Richard Aplenc,[20] Alan S. Gamis,[21] Edward A. Kolb,[22] Jason E. Farrar,[23] Derek Stirewalt,[4] Xiaotu Ma,[24] Tim I. Shaw,[24] Scott N. Furlan,[4] Lisa Eidenschink Brodersen,[8] Michael R. Loken,[8] Marry M. van den Heuvel-Eibrink,[1,25] C. Michel Zwaan,[1,2,13] Timothy J. Triche Jr.,[5,6,26] Bianca F. Goemans[1#] and Soheil Meshinchi[4,7,10#]

[1]Princess Máxima Center for Pediatric Oncology, Utrecht, the Netherlands; [2]Department of Pediatric Oncology/Hematology, Erasmus Medical Center – Sophia Children's Hospital, Rotterdam, the Netherlands; [3]Oncode Institute, Utrecht, the Netherlands; [4]Fred Hutchinson Cancer Research Center, Clinical Research Division, Seattle, WA, USA; [5]Department of Epigenetics, Van Andel Institute, Grand Rapids, MI, USA; [6]Department of Translational Genomics, University of Southern California, Los Angeles, CA, USA; [7]Children's Oncology Group, Monrovia, CA, USA; [8] Hematologics Inc., Seattle, WA, USA; [9]Department of Laboratories, Seattle Children's Hospital, Seattle, WA, USA; [10]Division of Hematology and Oncology, Seattle Children's Hospital, Seattle, WA, USA; [11]Intermountain Blood and Marrow Transplant and Acute Leukemia Program, Intermountain Healthcare, Salt Lake City, UT, USA; [12]Emma Children's Hospital, Amsterdam UMC, Vrije Universiteit Amsterdam, Pediatric Oncology, Amsterdam the Netherlands; [13]Dutch Childhood Oncology Group, Den Haag, the Netherlands; [14]Department of Pediatrics, Aarhus University Hospital, Aarhus, Denmark; [15]Children's Cancer Research Institute, Medical University of Vienna, Vienna, Austria; [16]St. Anna Kinderspital, Department of Pediatrics, Medical University of Vienna, Vienna, Austria; [17]Department of Pediatric Hematology and Oncology, University Hospital Essen, Essen, Germany; [18]Laboratory of Human Genetics, IRCCS Istituto Giannina Gaslini, Genoa, Italy; [19]Department of Pathology, St. Jude Children's Research Hospital, Memphis, TN, USA; [20]Division of Oncology and Center for Childhood Cancer Research, Children's Hospital of Philadelphia, Philadelphia, PA, USA; [21]Division of Hematology/Oncology, Children's Mercy Kansas City, Kansas City, MO, USA; [22]Nemours Alfred I. duPont Hospital for Children, Wilmington, DE, USA; [23]Arkansas Children's Research Institute and Department of Pediatrics, Hematology/Oncology Section, Department of Pediatrics, University of Arkansas for Medical Sciences, Little Rock, AR, USA; [24]Computational Biology Department, St. Jude Children's Research Hospital, Memphis, TN, USA; [25]Utrecht University, Utrecht, the Netherlands and [26]Department of Pediatrics, Michigan State University College of Human Medicine, Grand Rapids, MI, USA

**EJMB, JLS and LH contributed equally as first authors.*
#BFG and SM contributed equally as last authors.

Correspondence: Eline J. M. Bertrums
e.j.m.bertrums@prinsesmaximacentrum.nl

S. Meshinchi
smeshinc@fredhutch.org

Received: June 27, 2022.
Accepted: February 14, 2023.
Early view: February 23, 2023.

https://doi.org/10.3324/haematol.2022.281653

Supplementary Methods

Patient samples

Patients enrolled in the COG trials CCG-2961, AAML03P1, AAML0531 and AAML1031 were eligible for this study. Details of these studies have been previously described[1-4]. In total, 3,493 patients were included in these studies, of which 2,235 were eligible for inclusion due to availability of comprehensive *NUP98* fusion, molecular, and clinical data (**Supplementary Table 1, 2**). Eligible patients for each study included 13% (121/901) of all patients enrolled in CCG-2961, 34% (116/339) in AAML03P1, 84% (854/1022) in AAML0531 and 93% (1144/1231) in AAML1031. For the remaining patients, these data were unavailable. Eligible patients for each analysis performed in this manuscript are depicted in **Supplementary Fig. 1**. Sixteen of the 32 total NUP98-KDM5A patients in this study have been previously described by Noort et al, Haematologica (2021)[5]. In addition, we sent out an I-BFM AML study group proposal to include pediatric AML patients with a *NUP98*-X translocation from other study groups. Consent, in accordance with the Declaration of Helsinki, was obtained from all study participants. The Fred Hutchinson Cancer Research Center Institutional Review Board and the COG Myeloid Biology Committee approved and oversaw the conduct of this study. Adult AML patients from the Beat AML study, The Cancer Genome Atlas AML (TCGA LAML), and Southwestern Oncology Group (SWOG) AML studies were included as comparators for *NUP98* fusion analysis and details were reported accordingly in references[6-11].

Transcriptome sequencing

Pediatric patients with *de novo* AML (N=1,482) enrolled on COG trials CCG-2961, AAML03P1, AAML0531, and AAML1031, were included for RNA-sequencing (RNA-seq) when biological samples were available; samples included in transcriptome analyses are reported in **Supplementary Table 2**. Total RNA from diagnostic peripheral blood or bone marrow was extracted and purified using the QIAcube automated system with AllPrep DNA/RNA/miRNA Universal Kits (QIAGEN, Valencia, CA).

Libraries were prepared for 75-bp strand-specific paired-end sequencing using the ribodepletion v2.0 protocol by the British Columbia Genome Sciences Center (BCGSC, Vancouver, BC). Libraries were sequenced on the Illumina HiSeq 2000/2500 and aligned to the hg19 (GRCh37-lite) reference genome using BWA v0.5.7 with default parameters, except the addition of "-s" option, and duplicate reads were marked with Picard Tools. Gene level quantification was performed using the BCGSC-pipeline v1.1 with Ensembl v69 annotations. MicroRNA-sequencing was completed as previously reported[12].

Adult AML RNA-seq protocols were described previously for the Beat AML Study and TCGA LAML cohort[6, 7]. RNA-seq from SWOG AML was completed as follows: RNA was extracted from diagnostic specimens collected from participants on trials S9031, S9333, S0112, and S0106 with the AllPrep DNA/RNA Mini kit (Qiagen). Libraries were prepared for paired-end 75-bp sequencing using RiboErase (Roche, Wilmington, MA) and KAPA Stranded RNA-Seq Library Preparation Kit (KAPA Biosystems, Wilmington, MA) and reads were sequenced on Illumina NovaSeq 6000 instruments (Illumina, San Diego, CA) at Fred Hutchinson Cancer Research Center (Seattle, WA).

Screening of *NUP98* fusions

The *NUP98* fusions were detected by either karyotype or combined fusion detection algorithms STAR-fusion v1.8.1, TransAbyss v1.4.10, and CICERO v0.1.8[13-15] completed on RNA-seq. Patients' fusion annotations from prior studies were incorporated for additional coverage of cryptic fusions where available for protocols AAML0531, AAML03P1, and CCG-2961[16]. Differences in *NUP98* fusion detection per COG trial cohort were as follows. AAML1031 was screened by RNA-seq and checked by reverse transcription PCR (RT-PCR). Most patients from AAML0531 were screened by RNA-seq. For AAML03P1, all patients were screened for *NUP98-KDM5A*, and in addition all patients with *FLT3*-ITD were screened for other *NUP98* fusions. Lastly, all patients from CCG-2961 were screened for *NUP98-NSD1*, while only a small (unselected) subset was screened for other *NUP98* fusions. The majority (94%) of *NUP98*-translocated patients had RNA-seq evidence of their fusion. STAR-fusion was run using default parameters with the pre-made GRCh37 resource library with Gencode v19 annotations (https://data.broadinstitute.org/Trinity/CTAT_RESOURCE_LIB/). The TransAbyss software was executed with the GRCh37-lite reference genome with the following parameters included: fusion breakpoint reads ≥ 1, flanking pairs and spanning reads ≥ 2 counts. CICERO fusion detection was performed with default parameters with GRCh37-lite. Fusions detected computationally were verified using Fusion Inspector v.1.8.1 (Broad Institute, Cambridge, MA) and visualized on IGV[17-20] and BAMBINO[21]. Beat AML (N=440) and SWOG AML (N=206) transcriptome sequence reads were analyzed using STAR-fusion v1.8.1 with the same reference resource library and parameters as above[13]. TCGA LAML (N=179) RNA-seq fusion data were downloaded from supplementary materials[7].

Immunophenotype analysis methods

Flow cytometry data was analyzed for immunophenotypic markers from 93 *NUP98-NSD1,* 30 *NUP98-KDM5A,* and 20 *NUP98-X* patients. Mean fluorescence was measured for each patient and each marker. The presence and absence of all markers that were measured with fluorophores PE or FITC (HLA_DR, CD11B, CD13, CD33, CD36, CD49D, CD56, CD64, and CD117) was defined as Mean Fluorescence Index (MFI) greater than 14.84. This is a conservative threshold of three standard deviations above the mean autofluorescence in an unselected group of pediatric AML patients. This threshold covers the autofluorescence in 99% of cases (i.e., 1% of unstained cases pass the cutoff). Presence or absence of markers measured with the fluorophore APC (CD34) was defined as MFI greater than 20, as has been used in past analyses. Marker assignments were then hand-validated by experts at Hematologics, Inc. (Seattle, WA).

Differential expression, clustering, and Gene-set Enrichment Analysis

Differential expression analyses were completed in the R v4.0.2 statistical environment. Differences in gene expression were identified with trimmed mean of M-values (TMM) normalized counts per million (CPM) using Limma voom v.3.44.3 and edgeR v3.30.3 packages. DEGs were considered significant with absolute log2 fold-change > 1 and Benjamini–Hochberg adjusted p-values <0.05. DEGs per *NUP98* fusion subtype are listed in **Supplementary Tables 2-7**.

Unsupervised hierarchical clustering was performed using Euclidean distance matrices derived from log2 TMM-normalized CPM expression matrices, with a count of 1 added to avoid taking the log of zero, with the ward.D2 linkage algorithm using the stats R package. Samples were clustered based on the expression of highly variable genes across the dataset (988 heterogenous AML samples), selected using the mean versus dispersion parametric model trend (total 6858 genes selected) using SeqGlue v0.1. Heatmaps were constructed with ComplexHeatmap v2.4.3.

Unsupervised uniform manifold approximation and projection (UMAP)[22] was completed with term frequency–inverse document frequency (TF-IDF) transformed counts. TF-IDF transformation (SeqGlue v0.1) was carried out on the normalized counts matrix (total of 38,247 genes included in TF-IDF transformation); gene counts were size-factor normalized by the geometric mean of the total read counts. Input genes for the UMAP model were selected by identifying genes that showed the highest dispersion (CV^2) across a range of mean expression using a parametric model with the non-transformed counts matrix. Input genes were further refined by jackstraw principal component analysis using the jackstraw v1.3[23] package. UMAP was carried out with UWOT v0.1.5[24] and clusters were assigned by the Leiden clustering algorithm[25] applied to the UMAP reduced dimensional data (**Supplementary Table 8**). UMAP parameters used: cosine distance metric with a size of n_neighbors=15 using the "annoy" nearest neighbors method with 200 trees for constructing the nearest neighbor index and search_k=15000 nodes.

Gene-set enrichment analysis (GSEA) was completed with log2(x+1) TMM normalized CPM. GSEA was performed using the GAGE v2.30.0 R-package[26], which tests for differential expression of gene-sets by contrasting all possible combinations of fusion-positive to reference samples. Gene-sets from the

KEGG pathway database were used and non-redundant gene-sets were extracted for further analysis and identification of genes that most contribute to pathway enrichment.

Gene-set enrichment scores per patient were calculated using the single sample GSEA (ssGSEA) method[27] (GSVA v1.32.0), which transforms the normalized count data from a gene by sample matrix to a gene-set by sample matrix[28]. Transcription factor and microRNA regulatory target genes-sets were curated from Pathway Commons v11 database (www.pathwaycommons.org), and curated miRNA targets from the Molecular Signatures database (MSigDB v7.2, gsea-msigdb.org). Transcription factor motif enrichment was completed with RcisTarget v1.10.0 with the hg19 transcript start site (TSS) centered motifs +/- 5kbp v9 database[29].

DNA methylation analysis methods

DNA methylation was measured using 334,934 high-quality CpG probes shared by specimens run on HumanMethylation450 & HumanMethylationEPIC platforms. The methylation data was analyzed using RcppML[30], singlet[31] and sesame[32]. Non-negative matrix factorization (NMF) was performed at an optimal rank (ascertained by 5-fold cross-validation with automatic rank determination based on reconstruction error). Data from the HumanMethylation450 and HumanMethylationEPIC platforms were merged, mapped to human chromosomes 1-22, and compressed into 11 nonnegative factors. A multivariate linear model with empirical Bayes shrinkage was then used to test association of each factor with *HOX*-activating fusions (*NSD1, HOXA9, HOXA13, HOXD13, and PRRX1*) and epigenetic "reader-like" fusions (*KDM5A, BPTF, BRWD3, DDX10, HMGB3, KAT7, PHF15, PHF23, SET* and *TOP1*), with or without abnormal chr13. Benjamini-Hochberg correction was applied to the resulting matrix of p-values (predictor by factor). Hypermethylation signatures significant at an FDR of less than 0.1 were plotted. Locus-level (CpG) weights for actors associated with one or more biological features were then tested for enrichment against chromHMM state[33], histone mark ChIPseq (HM), JASPAR transcription factor binding sites (TFBS) [Castro-Mondragon, 2022], and CpG island locations (CGI) by selecting the highest 2% of weights (98th percentile; robust from 90th to 99.9th) as driving features, using the full array of shared CpG loci as the background distribution. The factor-level heatmap of sample clustering was likewise plotted on normalized (x/max(x)) NMF signal strengths across *NUP98* fusion samples as well as pediatric, adolescent, and young adult normal bone marrow (NBM) samples from the Heimfeld lab at FHCRC and from AllCells, which were included in the NMF model fit and regression analyses as referents.

Statistical methods

Data were current as of March 31, 2019. The Kaplan-Meier method was used to estimate overall survival (OS, defined as time from study entry to death) and event-free survival (EFS, time from study entry until failure to achieve CR during induction, relapse, or death). Relapse risk (RR) was calculated by cumulative incidence methods defined as time from the end of induction I for patients in CR to relapse or death, where deaths without a relapse were considered competing events. Patients who withdrew from therapy due to relapse, persistent central nervous system (CNS) disease, or refractory disease with >20% bone marrow blasts by the end of induction I were defined as induction I failures. The significance of predictor variables was tested with the log-rank statistic for OS, EFS and with Gray's statistic[34] for RR. All estimates were reported with two times the Greenwood standard errors. Children lost to follow-up were censored at their date of last known contact. Cox proportional hazards models were used to estimate the hazard ratio (HR) for defined groups of patients in univariate and multivariable analyses of OS and EFS. Competing risk regression models were used to estimate HRs for univariate and multivariable analyses of RR. *NUP98* translocation partner, cytogenetic/mutational risk group, age group, white blood cell count (WBC) and hematopoietic stem cell transplantation (HSCT) status were used as covariates. Three cytogenetic/mutational risk groups were defined: standard risk, low risk and high risk, based on the COG risk group stratification[4].

Comparison of clinical characteristics between different subgroups of *NUP98*-translocated patients and the reference cohort was carried out. The chi-squared test was used to test the significance of observed differences in proportions, and Fisher's exact test was used when data were sparse. Differences in medians were compared by the Mann-Whitney test. A *P*-value <0.05 was considered statistically significant. Measurable residual disease (MRD) was defined at the end of course one using flowcytometry with a cut-off of 0.1% detection of disease. The I-BFM patients were excluded from survival analyses due to variation in study groups and treatment protocols.

Data Availability

RNA-seq and DNA methylation array data on primary patient samples, as well as associated clinical/outcome data, are deposited in Genomic Data Commons (GDC, https://portal.gdc.cancer.gov/) and the Target Data Matrix (https://ocg.cancer.gov/programs/target/data-matrix) under project ID "TARGET-AML". Access to protected files hosted on the Sequence Read Archive (SRA), such as raw sequencing data in bam or fastq format, are available through dbGaP TARGET: Acute Myeloid Leukemia study (Accession: phs000465.v20.p8). Additional DNA methylation data are hosted on the Gene Expression Omnibus (GEO) under accessions GSE190931 and GSE124413. The Beat AML Study controlled access RNA-seq data were downloaded from the Genomic Data Commons (GDC) portal and are available through the Functional Genomic Landscape of Acute Myeloid Leukemia study on dbGaP (Accession: phs001657.v1.p1). TCGA LAML RNA-seq fusion data were accessed from the GDC Data Portal (https://gdc.cancer.gov/about-data/publications/laml_2012)[7].

References

1. Lange BJ, Smith FO, Feusner 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. 2008;111(3):1044-1053.

2. Cooper TM, Franklin J, Gerbing RB, 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. 2012;118(3):761-769.

3. Aplenc R, Meshinchi S, Sung L, et al. Bortezomib with standard chemotherapy for children with acute myeloid leukemia does not improve treatment outcomes: a report from the Children's Oncology Group. Haematologica. 2020;105(7):1879-1886.

4. Pollard JA, Loken M, Gerbing RB, et al. CD33 Expression and Its Association With Gemtuzumab Ozogamicin Response: Results From the Randomized Phase III Children's Oncology Group Trial AAML0531. J Clin Oncol. 2016;34(7):747-755.

5. Noort S, Wander P, Alonzo TA, et al. The clinical and biological characteristics of NUP98-KDM5A in pediatric acute myeloid leukemia. Haematologica. 2020;106(2):630-634.

6. Tyner JW, Tognon CE, Bottomly D, et al. Functional genomic landscape of acute myeloid leukaemia. Nature. 2018;562(7728):526-531.

7. Ley TJ, Miller C, Ding L, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013;368(22):2059-2074.

8. Anderson JE, Kopecky KJ, Willman CL, et al. Outcome after induction chemotherapy for older patients with acute myeloid leukemia is not improved with mitoxantrone and etoposide compared to cytarabine and daunorubicin: a Southwest Oncology Group study. Blood. 2002;100(12):3869-3876.

9. Petersdorf SH, Rankin C, Head DR, et al. Phase II evaluation of an intensified induction therapy with standard daunomycin and cytarabine followed by high dose cytarabine for adults with previously untreated acute myeloid leukemia: a Southwest Oncology Group study (SWOG-9500). Am J Hematol. 2007;82(12):1056-1062.

10. Godwin JE, Kopecky KJ, Head DR, et al. A double-blind placebo-controlled trial of granulocyte colony-stimulating factor in elderly patients with previously untreated acute myeloid leukemia: a Southwest oncology group study (9031). Blood. 1998;91(10):3607-3615.

11. List AF, Kopecky KJ, Willman CL, et al. Benefit of cyclosporine modulation of drug resistance in patients with poor-risk acute myeloid leukemia: a Southwest Oncology Group study. Blood. 2001;98(12):3212-3220.

12. Smith JL, Ries RE, Hylkema T, et al. Comprehensive Transcriptome Profiling of Cryptic CBFA2T3-GLIS2 Fusion-Positive AML Defines Novel Therapeutic Options: A COG and TARGET Pediatric AML Study. Clin Cancer Res. 2020;26(3):726-737.

13. Haas BJ, Dobin A, Li B, et al. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. Genome Biol. 2019;20(1):213.

14. Robertson G, Schein J, Chiu R, et al. De novo assembly and analysis of RNA-seq data. Nat Methods. 2010;7(11):909-912.

15. Tian L, Li Y, Edmonson MN, et al. CICERO: a versatile method for detecting complex and diverse driver fusions using cancer RNA sequencing data. Genome Biol. 2020;21(1):126.

16. Bolouri H, Farrar JE, Triche T, Jr., et al. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. Nat Med. 2018;24(1):103-112.

17. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-26.

18. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178-192.

19. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant Review with the Integrative Genomics Viewer. Cancer Res. 2017;77(21):e31-e34.

20. Robinson JT, Thorvaldsdóttir H, Turner D, Mesirov JP. igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). bioRxiv. 2020:2020.2005.2003.075499.

21. Edmonson MN, Zhang J, Yan C, et al. Bambino: a variant detector and alignment viewer for next-generation sequencing data in the SAM/BAM format. Bioinformatics. 2011;27(6):865-866.

22. McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:180203426. 2018.

23. Chung NC, Storey JD. Statistical significance of variables driving systematic variation in high-dimensional data. Bioinformatics. 2015;31(4):545-554.

24. McInnes LH, John. Melville, James. UMAP: Uniform Manifold Approximation and Projection for Dimensopn Reduction. arXiv; 2020.

25. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep. 2019;9(1):5233.

26. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics. 2009;10:161.

27. Barbie DA, Tamayo P, Boehm JS, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009;462(7269):108-112.

28. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7.

29. Aibar S, González-Blas CB, Moerman T, et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods. 2017;14(11):1083-1086.

30. DeBruine ZJ, Melcher K, Triche TJ. Fast and robust non-negative matrix factorization for single-cell experiments. bioRxiv. 2021:2021.2009.2001.458620.

31. DeBruine. singlet: Non-negative Matrix Factorization for single-cell analysis. R package version 0.0.99 ed. github; 2022.

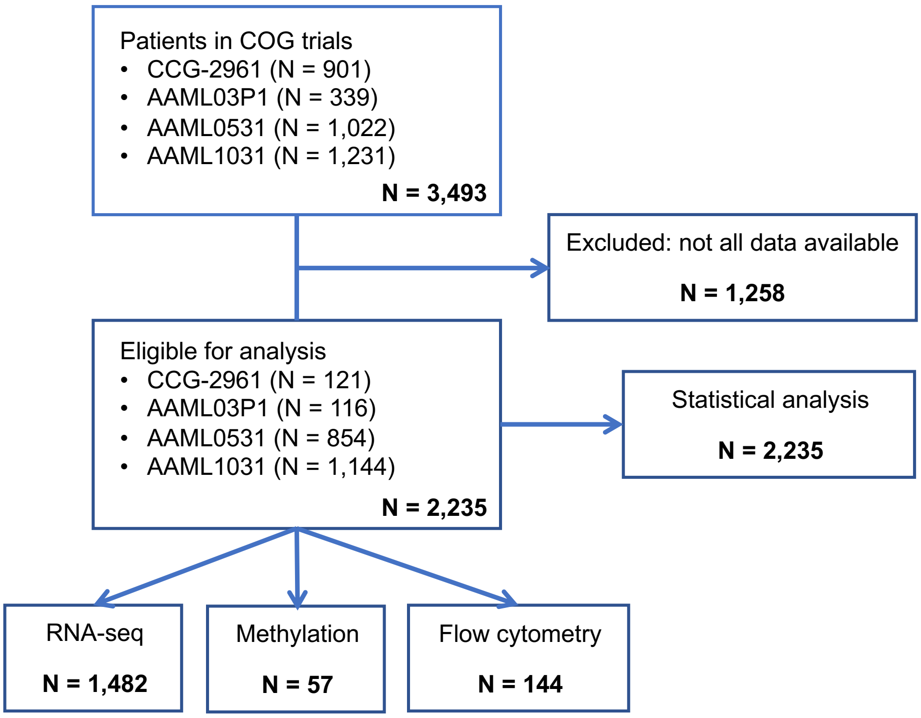
32. Zhou W, Triche TJ, Jr., Laird PW, Shen H. SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. Nucleic Acids Res. 2018;46(20):e123.

33. Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. Nat Protoc. 2017;12(12):2478-2492.

34. Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. The Annals of statistics. 1988:1141-1154.

Supplementary Data

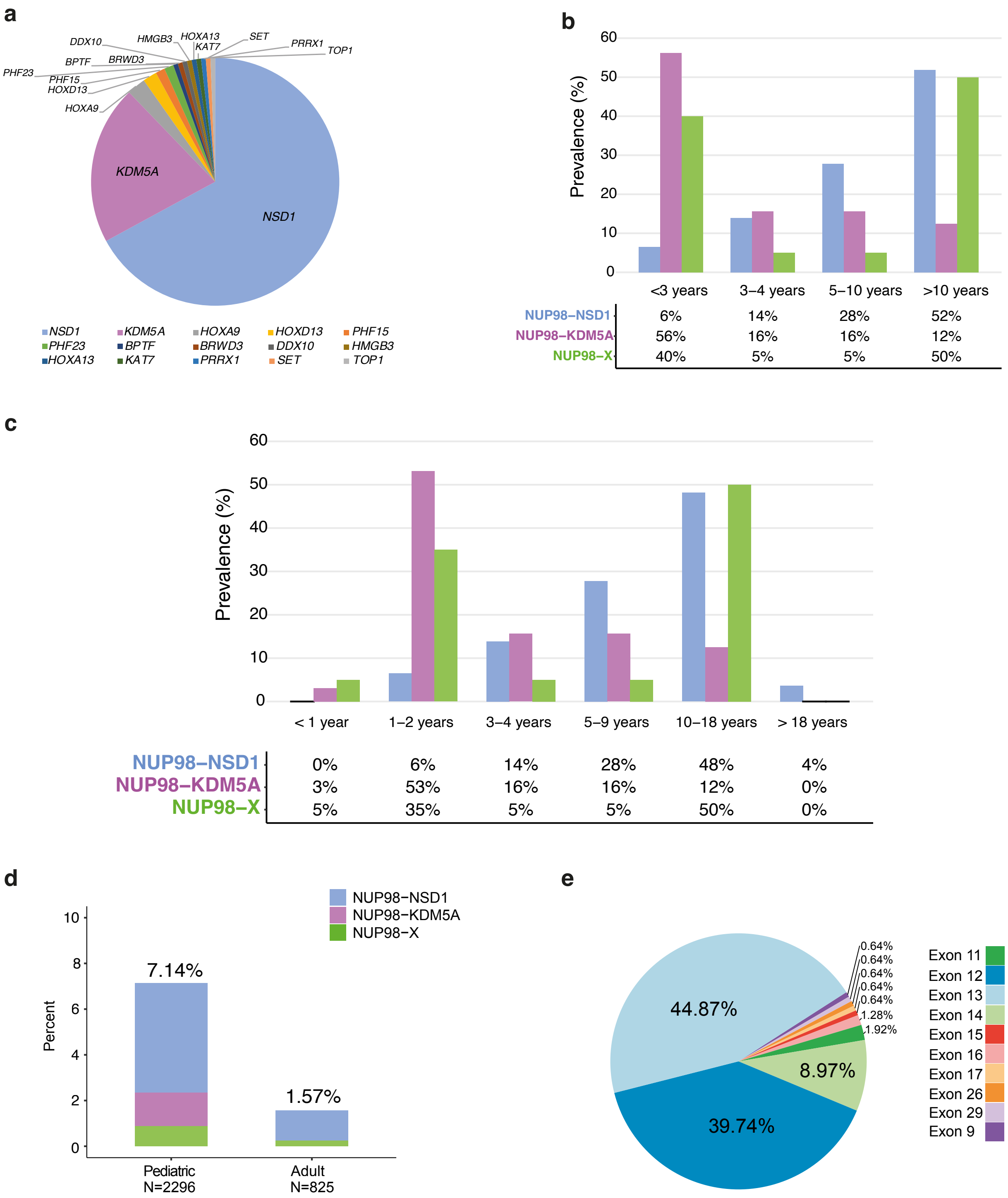
Supplementary Figure 1



Supplementary Figure 1. Patient inclusion per analysis.

Flowchart depicting the patient samples that are used per analysis performed. For RNA-sequencing (RNA-seq), methylation and flowcytometry analyses, patients with available data were included.

Supplementary Figure 2



Supplementary Figure 2. *NUP98* fusion partners, their prevalence and breakpoints

a) Prevalence of different *NUP98* fusion gene partners within our cohort of *NUP98*-translocated pediatric AML patients. **b,c)** *NUP98*-translocated subgroup frequencies within age categories. **d)**

Prevalence of *NUP98* fusions in adult and pediatric AML. **e)** Distribution of *NUP98* exon breakpoint junctions across all *NUP98* fusions identified by RNA-sequencing (N=156).

Supplementary Figure 3

a

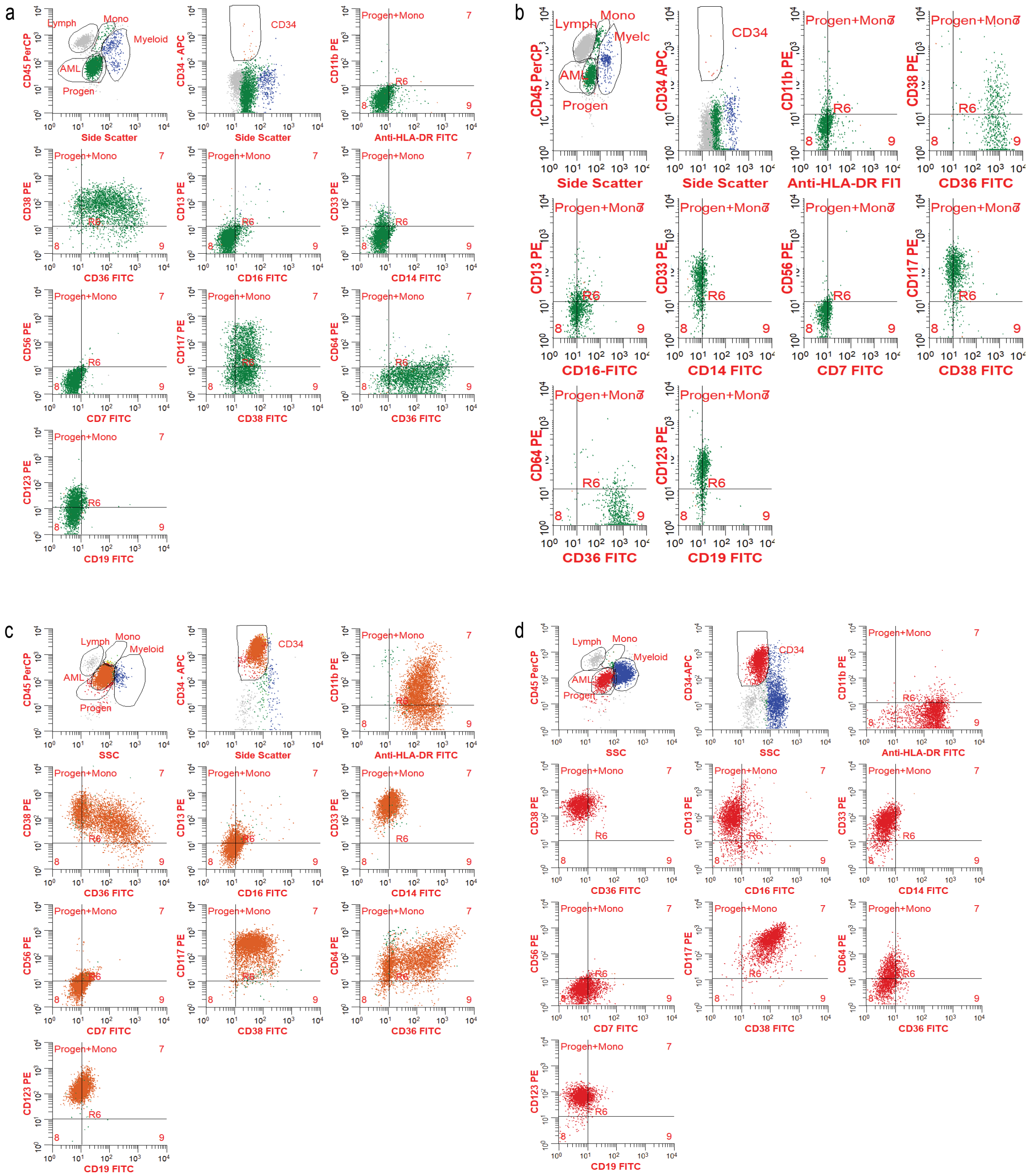
NUP98 Exon Junctions Identified	Number of Patients	Percent of Patients	Ensembl Transcript ID	RefSeq Accession
Exon13	70	44.87	ENST00000324932	NM_016320
Exon12	62	39.74	ENST00000324932	NM_016320
Exon14	14	8.97	ENST00000324932	NM_016320
Exon11	3	1.92	ENST00000324932	NM_016320
Exon16	2	1.28	ENST00000324932	NM_016320
Exon9	1	0.64	ENST00000324932	NM_016320
Exon29	1	0.64	ENST00000324932	NM_016320
Exon26	1	0.64	ENST00000324932	NM_016320
Exon17	1	0.64	ENST00000324932	NM_016320
Exon15	1	0.64	ENST00000324932	NM_016320

b

NUP98-Rearranged Groups	NUP98 Exon Junctions Identified	Number of Patients in NUP98 Group	Percent of Patients in NUP98 Group	Ensembl Transcript ID	RefSeq Accession
NUP98-KDM5A	Exon13	20	62.5	ENST00000324932	NM_016320
NUP98-KDM5A	Exon14	12	37.5	ENST00000324932	NM_016320
NUP98-NSD1	Exon13	45	43.27	ENST00000324932	NM_016320
NUP98-NSD1	Exon12	55	52.88	ENST00000324932	NM_016320
NUP98-NSD1	Exon16	1	0.96	ENST00000324932	NM_016320
NUP98-NSD1	Exon17	1	0.96	ENST00000324932	NM_016320
NUP98-NSD1	Exon26	1	0.96	ENST00000324932	NM_016320
NUP98-NSD1	Exon29	1	0.96	ENST00000324932	NM_016320
NUP98-X	Exon13	5	25	ENST00000324932	NM_016320
NUP98-X	Exon12	7	35	ENST00000324932	NM_016320
NUP98-X	Exon14	2	10	ENST00000324932	NM_016320
NUP98-X	Exon11	3	15	ENST00000324932	NM_016320
NUP98-X	Exon16	1	5	ENST00000324932	NM_016320
NUP98-X	Exon15	1	5	ENST00000324932	NM_016320
NUP98-X	Exon9	1	5	ENST00000324932	NM_016320

Supplementary Figure 3. Frequencies of *NUP98* breakpoint junctions. a) Prevalence of *NUP98* breakpoints per exon for all *NUP98* translocated patients. **b)** Frequencies of *NUP98* breakpoints per exon within each *NUP98*-translocated subtype (*NUP98-KDM5A*, *NUP98-NSD1*, and *NUP98*-X).

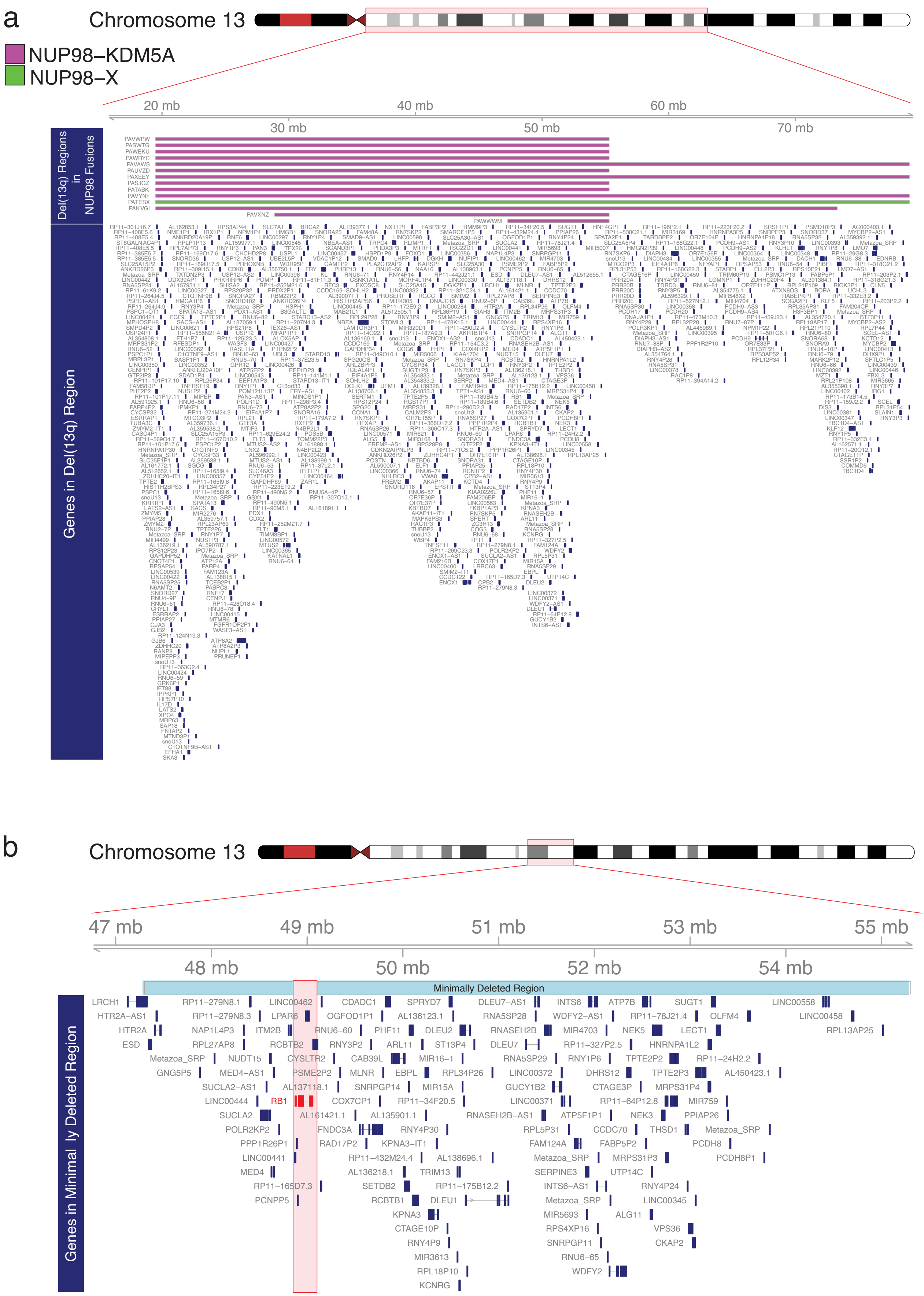
Supplementary Figure 4



Supplementary Figure 4. *NUP98* fusion immunophenotypes defined by the identity of the fusion partner and co-occurring mutations. Representative examples of the immunophenotype at diagnosis in patients with either *NUP98-KDM5A* or *NUP98-NSD1* fusions. **a,b)** Leukemias harboring *NUP98-KDM5A* were defined by a lack of cell surface CD34, CD11b, CD13, and CD64 but consistently expressed CD36 and CD33. CD38 and CD123 were also frequently decreased or absent. **c,d)** All

NUP98-NSD1 leukemias consistently expressed the immature markers CD34 and CD117. When co-occurring with *FLT3*-ITD mutations, tumors also typically expressed the monocytic markers CD36 and CD64; however, the expression of these markers was not seen when *FLT3*-ITD was absent.

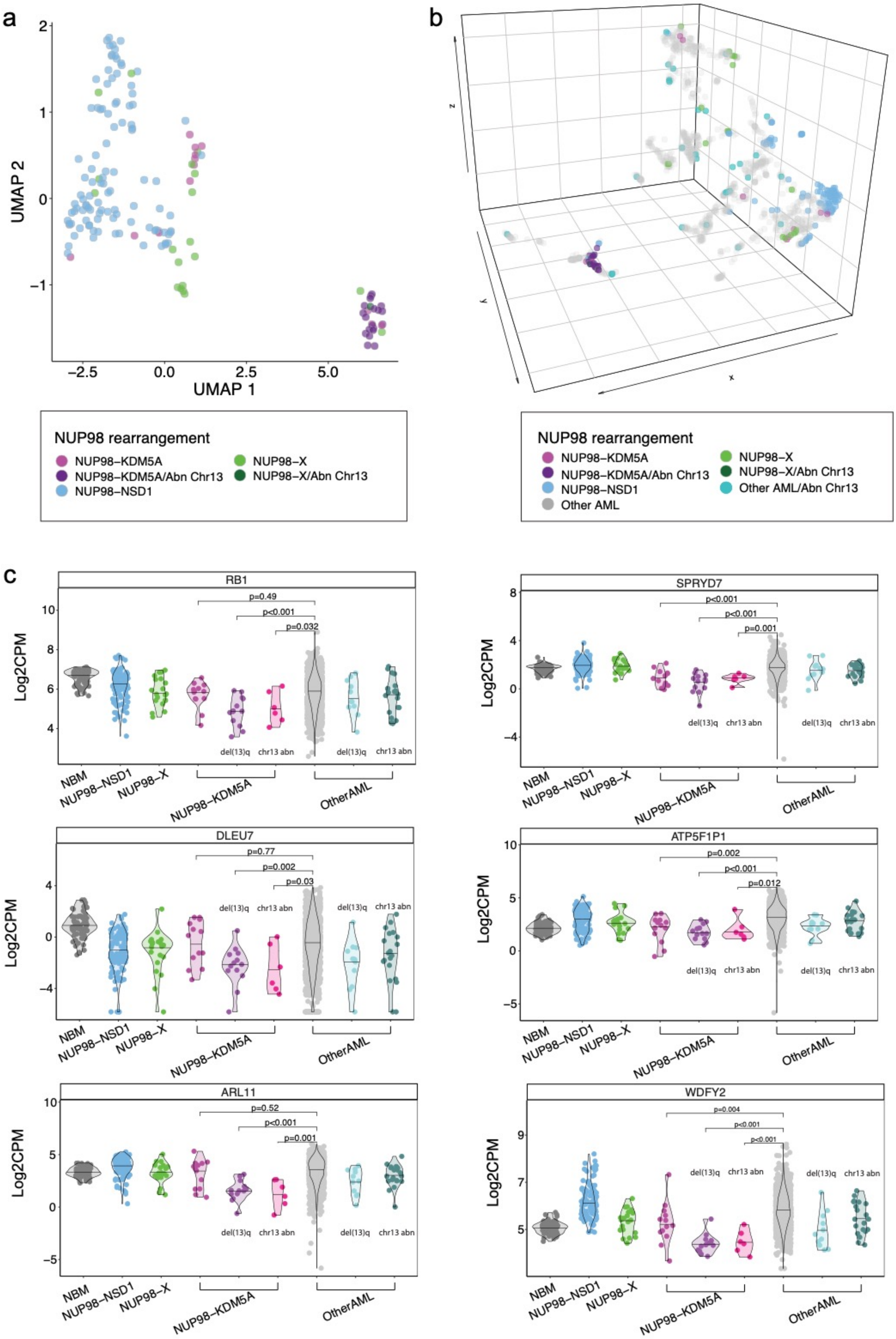
Supplementary Figure 5



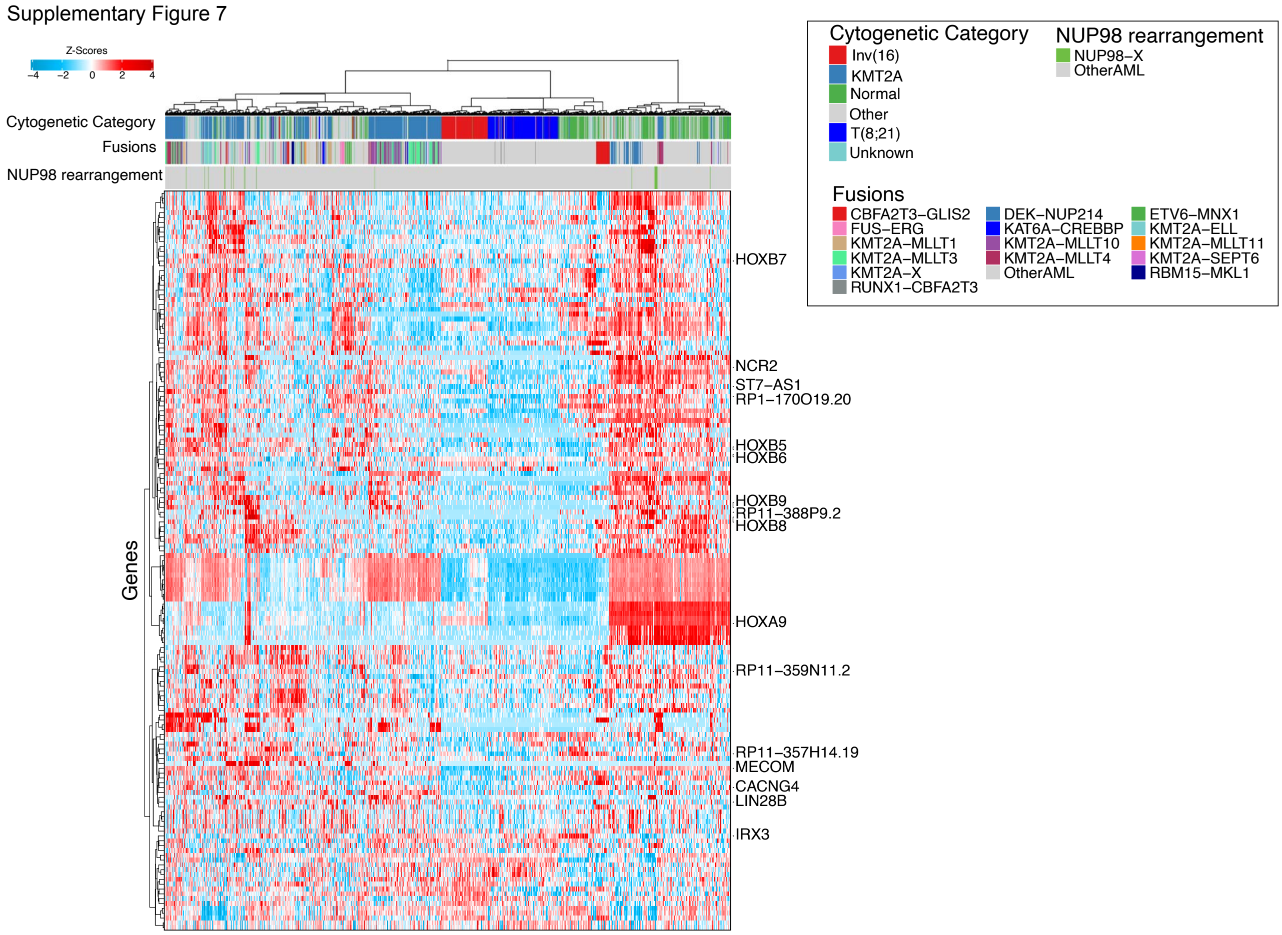
Supplementary Figure 5. Genomic positions of del(13q) alterations in *NUP98*-translocated cases.

a) Ideogram and genome track depicting the location of del(13q) alterations identified in *NUP98-KDM5A* patients (N=13) and a single *NUP98-SET* case (N=1). **b)** Representation of the minimally deleted region found in *NUP98*-translocated patients in 13q14.2 to 13q14.3, and the genes which reside in this locus, including *RB1*.

Supplementary Figure 6

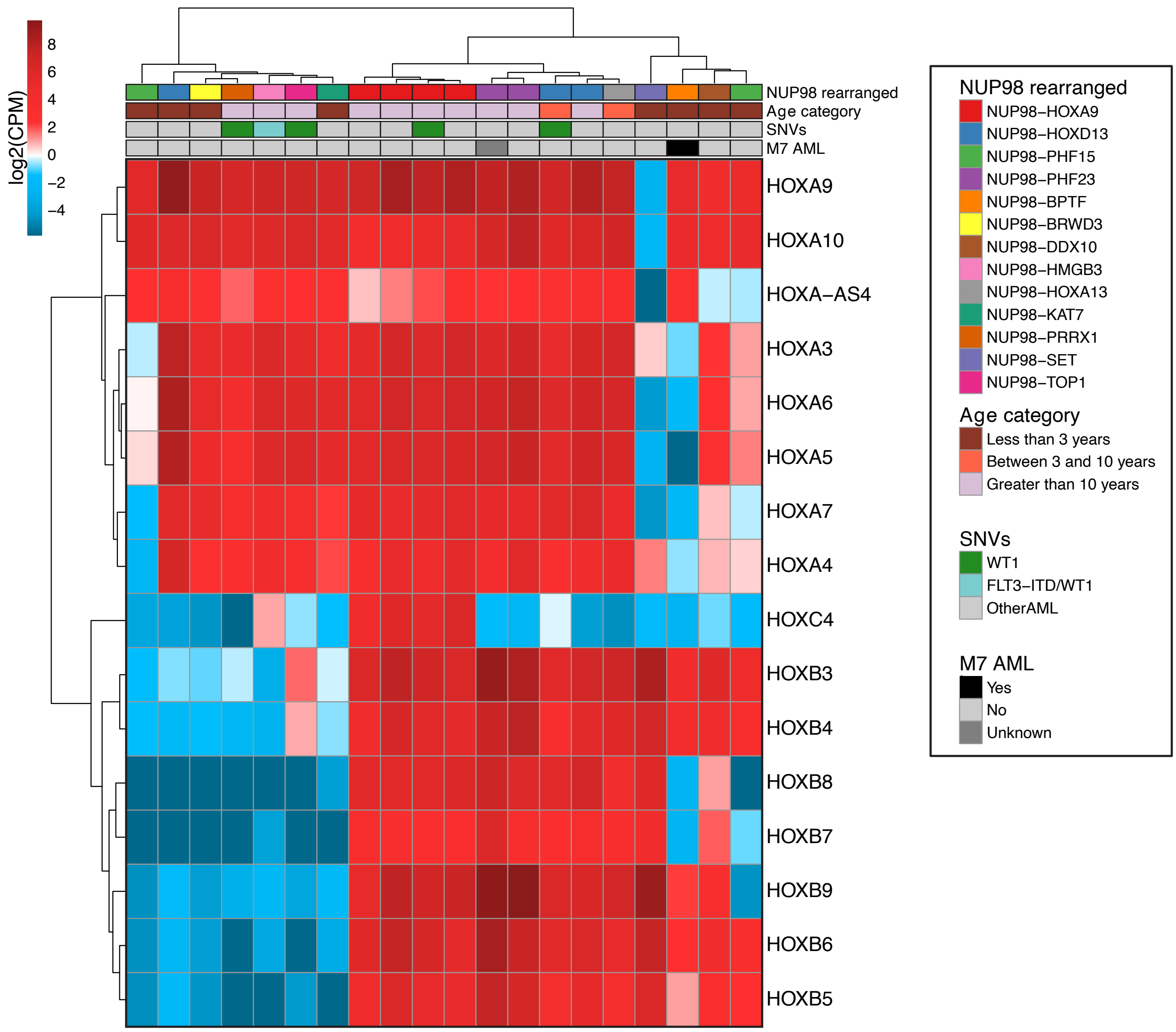


Supplementary Figure 6. Gene Expression patterns of *NUP98*-translocated patients with chr13 abnormalities. a) Unsupervised clustering by uniform manifold approximation and projection (UMAP) of *NUP98*-rearranged AML patients (N=156) illustrating *NUP98-KDM5A* cases cluster based on the presence of chr13 abnormalities (deletions, monosomy 13, or chr13 translocations). **b)** UMAP clustering with *NUP98*-translocated cases and a heterogenous AML reference cohort (other AML). *NUP98-KDM5A* cases with chr13 abnormalities (dark purple) are shown in comparison to patients with del(13q) but lacking *NUP98* fusions (teal). **c)** Down-regulated genes which reside in the minimally deleted region del(13)(q14.2q14.3) in *NUP98-KDM5A* (dark purple) compared to a heterogenous reference cohort of patients lacking chr13 deletions and *NUP98* translocations (grey). AML without *NUP98* translocations but harboring del(13q) alterations are also depicted (teal). Violin plots display the median (center), and points represent the expression of individual samples.



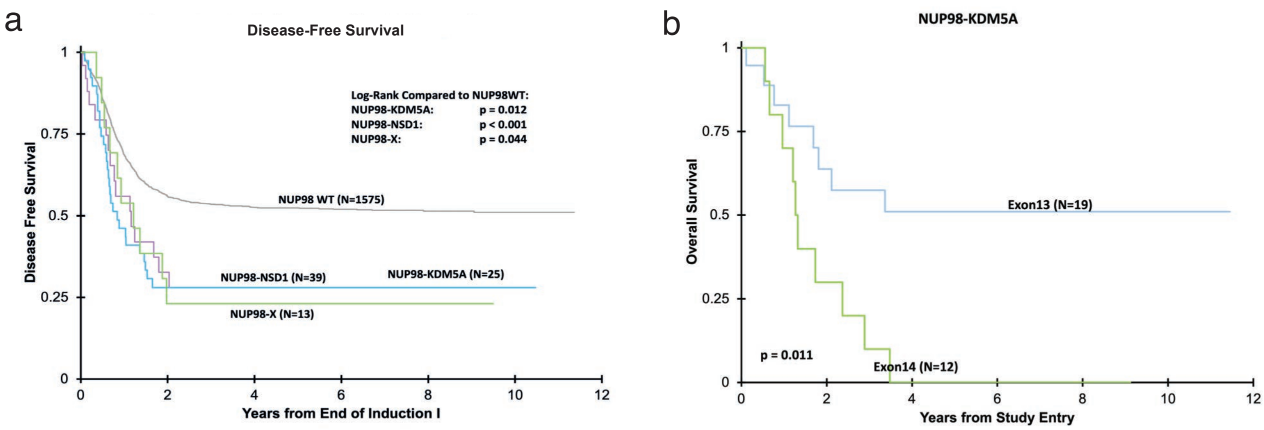
Supplementary Figure 7. Unsupervised hierarchical clustering of *NUP98*-X. Clustering of *NUP98*-X (N=20) and the reference AML cohort with various fusions and mutations (N=1,326) based on genes found to be differentially expressed in the *NUP98*-X cohort.

Supplementary Figure 8



Supplementary Figure 8. Gene expression patterns in *NUP98*-X. Expression of HOX homeobox genes found to be upregulated in *NUP98*-X compared to the reference AML cohort.

Supplementary Figure 9



Supplementary Figure 9. Outcome in *NUP98*-translocated patients. Outcome for patients in complete remission (CR) after induction 1 was examined for **a)** disease-free survival (DFS) compared to the reference cohort. **b)** Overall survival (OS) within *NUP98-KDM5A* by *NUP98* exon breakpoint junction.

Supplementary Table 1. Clinical characteristics and outcome of pediatric AML patients with and without *NUP98* translocations							
	No NUP98 fusion N = 2075	*NUP98-NSD1 N=108*	*P-value[1]*	*NUP98-KDM5A N = 32*	*P-value[1]*	*NUP98-X N = 20*	*P-value[1]*
Median age (range)	10.0 (0.01-29.8)	10.2 (1.19-19.89)	0.228	2.7 (0.98-15.92)	<0.001	7.9 (0.43-16.9)	0.300
Sex male N (%)	1056 (50.9)	70 (64.8)	0.005	18 (56.3)	0.547	10 (50.0)	0.937
Age category N (%)							
<3 years	502 (24.2)	7 (6.5)	<0.001	18 (56.3)	<0.001	8 (40.0)	0.116
3-10 years	529 (25.5)	45 (41.7)	<0.001	10 (31.3)	0.459	2 (10.0)	0.113
>10 years	1044 (50.3)	56 (51.9)	0.755	4 (12.5)	<0.001	10 (50.0)	0.978
FAB M6/M7 N (%)	107 (5.5)	3 (2.9)	0.259	15 (46.9)	<0.001	2 (10.5)	0.283
CNS disease, N (%)	385 (18.9)	19 (18.1)	0.830	1 (3.1)	0.023	5 (25.0)	0.564
WBC x 10^3 ul median (range)	23.9 (0.2-918.5)	169 (1.1-860)	<0.001	11.4 (1.8-237.3)	0.008	14.65 (3.5-445.7)	0.701
Blasts, % (range)							
BM	69 (0-100)	81 (20-98)	<0.001	42 (4-99)	0.007	54 (20-91)	0.521
PB	41 (0-99)	68.9 (0-100)	<0.001	9.5 (0-93)	<0.001	35 (0-93)	0.909
Chromosomal aberrations							
Normal	474 (23.5)	55 (57.3)	<0.001	6 (20.7)	0.727	1 (5.0)	0.061
t(6;9)	43 (2.1)	0 (0)	0.261	0 (0)	1.000	0 (0)	1.000
t(8;21)	294 (14.5)	0 (0)	<0.001	0 (0)	0.016	0 (0)	0.100
inv(16)	224 (11.1)	0 (0)	<0.001	0 (0)	0.066	0 (0)	0.156
Monosomy 5/del5q	25 (1.2)	4 (4.2)	0.040	0 (0)	1.000	0 (0)	1.000
Del7q	30 (1.5)	1 (1.0)	1.000	0 (0)	1.000	0 (0)	1.000
Monosomy 7	49 (2.4)	0 (0)	0.168	0 (0)	1.000	0 (0)	1.000
Trisomy 8	108 (5.3)	18 (18.8)	<0.001	4 (13.8)	0.070	1 (5.0)	1.000
Chromosome 13, N (%)							
Abnormal chr13[2]	47 (2.3)	0 (0)	0.271	19 (65.3)	<0.001	1 (5.0)	0.383
Chr13 deletion (del13q)	18 (0.9)	0 (0)	1.000	13 (43.3)	<0.001	1 (5.0)	0.173
Molecular genetics							
FLT3-ITD	301 (14.7)	80 (74.1)	<0.001	1 (3.1)	0.075	2 (10.0)	0.756
WT1	192 (9.6)	47 (43.5)	<0.001	1 (3.1)	0.358	5 (25.0)	0.039
NPM1	189 (9.3)	0 (0)	0.001	0 (0)	0.111	1 (5.0)	0.512
CEBPA	126 (6.2)	0 (0)	0.008	0 (0)	0.258	0 (0)	0.630
SCT yes, N (%)	353(17.0)	39 (36.1)	<0.001	8 (25.0)	0.230	2 (10.0)	0.630
CR end course 1, N (%)	1575 (78.0)	39 (38.2)	<0.001	25 (80.6)	0.729	13 (65.0)	0.176
MRD+ end course 1, N (%)	477 (27.3)	57 (73.1)	<0.001	14 (51.9)	0.005	4 (22.2)	0.793
Survival							
5-y O[3], % (± 2SE)	*64 (± 2%)*	*36 (± 10%)*	*<0.001*	*30 (± 18%)*	*<0.001*	*35 (± 21%)*	*0.009*
5-y EFS[3], % (± 2SE)	*47 (± 2%)*	*17(± 7%)*	*<0.001*	*25(± 16%)*	*0.010*	*35 (± 21%)*	*0.333*
5-y RR[4], % (± 2SE)	*42 (± 3%)*	*64 (± 16%)*	*0.001*	*68 (± 21%)*	*0.010*	*69 (± 28%)*	*0.071*

[1] P-value represents a comparison with the reference cohort. [2]Including del13q, monosomy 13, and translocations involving chromosome 13. [3]Time from study entry. [4]Time from end of induction 1.
AML: acute myeloid leukemia; BM: bone marrow; CNS: central nervous system; CNV: copy number variation; CR: complete remission (measured by morphology); CR1: first complete remission; EFS: event-free survival; FAB: French-American-British classification; MRD+: measurable residual disease positivity (measured by flow cytometry); OS: overall survival; RR: relapse rate; SCT: stem cell transplantation; SE: standard error; WBC: white blood cell count; y: year. Not all data was available from all included patients, percentages are adjusted to unknown values.

Supplementary Table 2. RNA-sequencing sample manifest

Separate Excel file.

Supplementary Table 3. Characteristics of *NUP98*-other translocated pediatric AML patients

Patient	*Study Group*	*Sex*	*Age (y)*	*NUP98 translocation by karyotype*	*Abnormal chr13 by karyotype*	*WBC (x10^9/l)*	*X =*	*Molecular genetics*	*SCT in CR1*	*Outcome*
PAXJFS	COG	M	0.4	t(11;17)(p15;q23)		14	*BPTF*		No	Relapse, died
PAMYMA	COG	M	1.2	t(X;11)(q13;p15.1)		237	*BRWD3*		N/A	Induction failure, died
PAUYZY	COG	M	1.3	der(11)ins(11;11)(p15; q21q23)		6.1	*DDX10*		No	Relapse, died
PANLXM	COG	F	10.3	cryptic - partner telomeric		11.2	*HMGB3*	*FLT3, WT1*	No	Relapse, died
PAVCNU	COG	F	5.6	t(7;11)(p15;p15)		4.8	*HOXA13*		No	Death, died
PARGDB	COG	F	16.8	t(7;11)(p15;p15)		79	*HOXA9*	*WT1*	No	Relapse, died
PAXAFS	COG	M	13.6	t(7;11)(p15;p15)		26.8	*HOXA9*		Yes	Relapse, died
PARIEG	COG	F	13.4	t(7;11)(p15;p15)		50.2	*HOXA9*		No	Relapse, died
PARDRM	COG	F	12.5	t(7;11;9)(p15;p15;q22)		286	*HOXA9*		Yes	Relapse, died
PATELT	COG	F	12	t(2;11)(q31;p15)		46.7	*HOXD13*		N/A	Censored, alive
PAUPDK	COG	M	4.1	t(2;11)(q31;p15)		444.7	*HOXD13*	*WT1*	No	Death without remission, died
PATETC	COG	M	1.7	t(2;11)(q31;p15)		9.8	*HOXD13*		No	Censored, alive
PASSBI	COG	M	1.2	t(11;17)(p15;q21)		9.1	*KAT7*		No	Censored, alive
PAXFSI	COG	M	2	t(5;11)(q31;p15.5)		9.7	*PHF15*		No	Censored, alive
PAWRUF	COG	F	1.4	t(5;11)(q31;p15)		8.9	*PHF15*		No	Censored, alive
PARSAN	COG	M	14.8	cryptic		5.3	*PHF23*		No	Relapse, died
PAVCPM	COG	F	13.5	t(11;17)(p15;p13)		3.5	*PHF23*		No	Relapse, died
PAWNBB	COG	M	16.9	t(1;11)(q23;p15)		65	*PRRX1*	*FLT3, WT1*	No	Relapse, died
PATESX	COG	F	2.3	cryptic	del(13)(q12q22)	21.1	*SET*		No	Censored, alive
PASPIX	COG	F	16.3	t(11;20)(p15;q11.2)		15.3	*TOP1*	*NPM1, WT1*	No	Censored, alive
IBFM01	BFM-Austria	M	3.3	t(11;20)(p15;q11)		9.25	*TOP1*		No	Relapse, died
IBFM02	BFM-Italy[1]	F	5.2	t(9;11)(p22;p15)		207	*LEDGF*		N/A	Toxicity, died
IBFM03	BFM-Italy[2]	M	11.8	inv(11)(p15q22)		29.7	*DDX10*		Yes	Relapse
IBFM04	BFM-Germany	F	2.3	t(7;11)(p13;p15)		152	*HOXA13*	*N/A*	No	Alive
IBFM05	BFM-NL	F	12.9	t(11;20)(p15;q12).ish t(11;20)		214	*TOP1*	*WT1*	Yes, in CR2	Infection, died
IBFM06	NOPHO-DBH	F	9.8	t(11;20)(p15;q11)		248	*TOP1*	*RUNX1*	Yes	Alive

BFM: Berlin-Frankfurt-Münster; COG: Children's Oncology Group; CR: complete remission; F: female; M: male; N/A: data not available; NL: Netherlands; SCT: stem cell transplantation; WBC: white blood cell count; y: years. 1) C.Morerio et al, Leukemia Res 2005. 2) C.Morerio et al, Cancer Genet Cytogenet 2006.

Supplementary Table 4. *NUP98*-X fusion genomic breakpoints and corresponding exon junctions. The NUP98 transcript identifiers are ENST00000324932 and NM_016320.

Patient	*NUP98 fusion*	*Breakpoint*	*Exon breakpoint NUP98*	*Partner gene*	*Exon breakpoint partner gene*	*Ensembl Transcript ID partner gene*	*RefSeq Accession partner gene*
PARDRM	*NUP98-HOXA9*	11:3765739|7:27204586	12	*HOXA9*	1	ENST00000343483	NM_152739
PARGDB	*NUP98-HOXA9*	11:3765739|7:27204586	12	*HOXA9*	1	ENST00000343483	NM_152739
PARIEG	*NUP98-HOXA9*	11:3765739|7:27204586	12	*HOXA9*	1	ENST00000343483	NM_152739
PAXAFS	*NUP98-HOXA9*	11:3774546|7:27204586	11	*HOXA9*	1	ENST00000343483	NM_152739
PATELT	*NUP98-HOXD13*	11:3744387|2:176959208	16	*HOXD13*	2	ENST00000392539	NM_000523
PATETC	*NUP98-HOXD13*	11:3765739|2:176959208	12	*HOXD13*	2	ENST00000392539	NM_000523
PAUPDK	*NUP98-HOXD13*	11:3765739|2:176959208	12	*HOXD13*	2	ENST00000392539	NM_000523
PAXFSI	*NUP98-PHF15*	11:3756421|5:133871548	13	*PHF15*	2	ENST00000395003	NM_015288
PARSAN	*NUP98-PHF23*	11:3756421|17:7140086	13	*PHF23*	4	ENST00000320316	NM_024297
PAVCPM	*NUP98-PHF23*	11:3756421|17:7140086	13	*PHF23*	4	ENST00000320316	NM_024297
PAWRUF	*NUP98-PHF15*	11:3756421|5:133871548	13	*PHF15*	2	ENST00000395003	NM_015288
PAMYMA	*NUP98-BRWD3*	11:3765739|X:79973258	12	*BRWD3*	19	ENST00000373275	NM_153252
PAVCNU	*NUP98-HOXA13*	11:3765739|7:27238061	12	*HOXA13*	2	ENST00000222753	NM_000522
PASSBI	*NUP98-KAT7*	11:3784132|17:47869248	9	*KAT7*	2	ENST00000259021	NM_007067
PATESX	*NUP98-SET*	11:3774546|9:131453449	11	*SET*	2	ENST00000372692	NM_001122821
PASPIX	*NUP98-TOP1*	11:3756421|20:39713102	13	*TOP1*	8	ENST00000361337	NM_003286
PAUYZY	*NUP98-DDX10*	11:3752621|11:108559663	14	*DDX10*	7	ENST00000322536	NM_004398
PAWNBB	*NUP98-PRRX1*	11:3774546|1:170688867	11	*PRRX1*	2	ENST00000367760	NM_006902
PANLXM	*NUP98-HMGB3*	11:3746435|X:150151833	15	*HMGB3*	1	ENST00000325307	NM_005342
PAXJFS	*NUP98-BPTF*	11:3752808|17:65944422	14	*BPTF*	23	ENST00000306378	NM_182641

Supplementary Table 5. Chromosome 13 abnormalities identified in *NUP98-KDM5A* by karyotype.				
Patient	*NUP98 translocation*	*Deletion chr13*	*Monosomy 13*	*Translocation 13*
PAVXNZ	*NUP98-KDM5A*	del(13)(q12.3q14.3)		t(1;13)(p12;q12)
PASWTG	*NUP98-KDM5A*	del(13)(q12q14)		
PATABK	*NUP98-KDM5A*	del(13)(q12q14)		
PASJGZ	*NUP98-KDM5A*	del(13)(q12q14)		
PAUVZD	*NUP98-KDM5A*	del(13)(q12q14)		
PAVWPW	*NUP98-KDM5A*	del(13)(q12q14)		
PAWEKU	*NUP98-KDM5A*	del(13)(q12q14)		
PAWRYC	*NUP98-KDM5A*	del(13)(q12q14)		
PAKVGI	*NUP98-KDM5A*	del(13)(q12q21)		
PAVAWS	*NUP98-KDM5A*	del(13)(q12q22)		
PAVYNF	*NUP98-KDM5A*	del(13)(q12q22)		
PAXEEY	*NUP98-KDM5A*	del(13)(q12q22)		
PAWWWM	*NUP98-KDM5A*	del(13)(q14.2q14.3)		
PAWPLE	*NUP98-KDM5A*		Monosomy 13	t(13;22) (q21;p11.2)
PARKLC	*NUP98-KDM5A*		Monosomy 13	
PASDTY	*NUP98-KDM5A*			t(10;13)(p11.2;q21)
PAWJIM	*NUP98-KDM5A*			t(13;17)(q22;q25)
PAVAFA	*NUP98-KDM5A*			t(2;13)(q31;q14)
PARXMP	*NUP98-KDM5A*			t(6;13)(q 23;q12)
PATKMB	*NUP98-KDM5A*			
PARDLW	*NUP98-KDM5A*			
PANGTF	*NUP98-KDM5A*			
PARDYG	*NUP98-KDM5A*			
PARMHD	*NUP98-KDM5A*			
PATLFJ	*NUP98-KDM5A*			
PATKJB	*NUP98-KDM5A*			
PAUYCB	*NUP98-KDM5A*			
PAVULK	*NUP98-KDM5A*			
PAWDNM	*NUP98-KDM5A*			
PAWPDC	*NUP98-KDM5A*			
PAKERZ	*NUP98-KDM5A*			
PAEMCF	*NUP98-KDM5A*			

Supplementary Table 6. *NUP98* fusion immunophenotypes identified by multidimensional flow cytometry.													
						***NUP98-KDM5A* (N=31)**							
	CD34-	*HLA-DR+*	*11b-*	*CD38 dim/-*	*CD36 het/+*	*CD13-*	*CD33-*	*CD14+*	*Some CD56+*	*Some CD7+*	*CD17-*	*CD64-*	*123 dim/-*
#	28	15	29	19	25	26	8	0	7	2	14	13	15
%	**90.3**	48.4	**93.5**	**61.3**	**80.6**	**83.9**	25.8	0.0	22.6	6.5	45.2	**81.3**	**83.3**
N	31	31	31	31	31	31	31	31	31	31	31	16	18
						***NUP98*-X (N=20)**							
	CD34-	*HLA-DR+*	*11b-*	*CD38 dim/-*	*CD36 het/+*	*CD13-*	*CD33-*	*CD14+*	*Some CD56+*	*Some CD7+*	*CD17-*	*CD64-*	*123 dim/-*
#	7	12	20	3	4	5	2	0	3	1	5	3	3
%	35.00	60.0	100.00	15.00	20.00	25.00	10.00	0.00	15.00	5.00	25.00	37.50	33.33
N	20	20	20	20	20	20	20	20	20	20	20	8	9
						***NUP98-NSD1* (N=92)**							
	CD34-	*HLA-DR+*	*11b-*	*CD38 dim/-*	*CD36 het/+*	*CD13-*	*CD33-*	*CD14+*	*Some CD56+*	*Some CD7+*	*CD17-*	*CD64-*	*123 dim/-*
#	15	90	29	16	42	13	6	2	4	19	6	17	0
%	**16.30**	**97.83**	31.52	17.39	45.65	14.13	6.52	2.17	4.35	20.65	6.59	37.78	0.00
N	92	92	92	92	92	92	92	92	92	92	91	45	54
						***NUP98-NSD1 FLT3*-ITD pos (N=67)**							
	CD34-	*HLA-DR+*	*11b-*	*CD38 dim/-*	*CD36 het/+*	*CD13-*	*CD33-*	*CD14+*	*Some CD56+*	*Some CD7+*	*CD17-*	*CD64-*	*123 dim/-*
#	11	65	11	13	37	11	5	2	3	15	4	10	0
%	16.42	97.01	**16.42**	19.40	**55.22**	16.42	7.46	2.99	4.48	22.39	5.97	**29.41**	0
N	67	67	67	67	67	67	67	67	67	67	67	34	41
						***NUP98-NSD1 FLT3*-ITD neg (N=23)**							
	CD34-	*HLA-DR+*	*11b-*	*CD38 dim/-*	*CD36 het/+*	*CD13-*	*CD33-*	*CD14+*	*Some CD56+*	*Some CD7+*	*CD17-*	*CD64-*	*123 dim/-*
#	2	23	16	3	3	2	1	0	1	4	2	7	0
%	8.70	100.00	69.57	13.04	13.04	8.70	4.35	0.00	4.35	17.39	9.09	63.64	0
N	23	23	23	23	23	23	23	23	23	23	22	11	13
#: number of patients; %: percentage of patients; het: heterogeneous; N: number of assessed patients													

Supplementary Table 7. RCIS-Target analysis results

Separate Excel file.

Supplementary Table 8. Univariable and multivariable analyses for OS and RR of *NUP98*-translocated AML		
Univariable analyses		
Variables	Overall survival *HR (95% CI), p-value*	Relapse risk *HR (95% CI), p-value*
NUP98-NSD1	2.17 (1.68-2.8); <0.001	2.04 (1.34-3.11); 0.001
NUP98-KDM5A	2.26 (1.43-3.56); 0.001	1.99 (1.21-3.26); 0.007
NUP98-X	2.05 (1.19-3.55); 0.010	1.86 (1.06-3.27); 0.031
Multivariable analyses		
NUP98 fusion partner		
NUP98-NSD1	1.46 (1.1-1.94); 0.009	1.74 (1.1-2.76); 0.018
NUP98-KDM5A	1.83 (1.13-2.96); 0.015	1.42 (0.84-2.42); 0.193
NUP98-*X*	1.75 (1.01-3.04); 0.046	1.43 (0.82-2.51); 0.208
Low risk cytogenetics	0.37 (0.3-0.45); <0.001	0.47 (0.39-0.56); <0.001
High risk cytogenetics	1.20 (0.99-1.45); 0.069	0.61 (0.47-0.8); <0.001
WBC $\geq$ 100 (x10^3/ul)	1.09 (0.91-1.29); 0.354	1.31 (1.08-1.59); 0.006

Overall survival (OS; from study entry) and relapse risk (RR; from end of induction 1) for different *NUP98*-translocated subgroups in univariable and multivariable analysis. Shown are Hazard ratio (HR) with a 95% confidence interval (95% CI) and p-value. In univariable analyses, the reference is the reference cohort with non-*NUP98* translocated patients. In multivariable analysis, cytogenetic risk group and white blood cell count (WBC) are taken into account, references are non-*NUP98*-translocated patients, standard risk cytogenetics and WBC <100 (x10^3/ul), respectively.