Human models of NUP98-KDM5A megakaryocytic leukemia in mice contribute to uncovering new biomarkers and therapeutic vulnerabilities

Sophie Cardin*, Mélanie Bilodeau*, Mathieu Roussy, Léo Aubert, Thomas Milan, Loubna Jouan, Alexandre Rouette, Louise Laramée, Patrick Gendron, Jean Duchaine, Hélène Decaluwe, Jean-François Spinella, Stéphanie Mourad, Françoise Couture, Daniel Sinnett, Élie Haddad, Josette-Renée Landry, Jing Ma, R. Keith Humphries, Philippe P. Roux, Josée Hébert, Tanja A. Gruber, Brian T. Wilhelm and Sonia Cellot

(*Co-first authors)

SUPPLEMENTAL DATA

SUPPLEMENTAL METHODS

Lentiviral transduction

Cord blood-derived $CD34^+$ human hematopoietic stem/progenitor cells (e.g. CB-$CD34^+$ cells) were prestimulated for 16-20h in expansion media prior to lentiviral transduction in 96-well plates coated with RetroNectin (Takara Bio USA, Inc., cat. no. T100B). CB-$CD34^+$ cells (1-4 $x10^4$ cells per well) were transduced with lentiviral particles encoding *NUP98-KDM5A* or an empty vector at a multiplicity of infection (MOI) of ≈50 in expansion media supplemented with 1µg/ml polybrene (hexadimethrine bromide, MilliporeSigma, USA, cat. no. H9268), for 16 hours.

Clonogenic progenitor cell assay

Transduced cells (unsorted) were seeded in Methocult H4034 optimum **(**Stem Cell Technologies Inc., cat. no. 04034) at a density of 200 cells/ml in 35 mm dishes. Colony quantification and morphological assessment were performed manually at day 14 post-seeding with an EVOS FL Auto imaging system (Thermo Fisher Scientific, cat. no. AMAFD1000).

Xenotransplantation

NOD-*scid* IL2Rg^{null} (NSG) mice from Jackson Laboratories (Bar Harbor, ME, USA) were used as primary recipients for xenotransplantation. NOD.Cg-$Prkdc^{scid}$ $Il2rg^{tm1Wjl}$ Tg(CMV-IL3,CSF2,KITLG)1Eav/MloySzJ mice (NSG-SGM3, Jackson Laboratories) were used as secondary recipients to enhance engraftment of AMKL primary xenograft cells. At day 6-7 of culture, unsorted transduced cells (approx. 8-10 x 10^4 or 70% of the well) were injected intravenously into a single

immunodeficient (NSG) recipient mouse sub-lethally irradiated (1 well/mouse, 3 to 28 mice/experiment, 6 independent experiments, 7 CB units). Primary AMKL xenograft cells (2.2 x 10^6 bulk cells) were injected in secondary NSG-SGM3 mice. All mouse recipients were sub-lethally irradiated (whole-body irradiation with 2Gy X-rays, CP160 irradiator, Faxitron X-Ray Corporation, USA) 6 to 24 hours prior to the xenotransplantation. Following xenotransplantation, the percentage of human $CD45^+$ cells in mouse bloodstream was monitored monthly by flow cytometry. Mice were maintained up to 64 weeks post-transplantation or until showing advanced signs of leukemia (reduced mobility, paleness, hunchback and/or dyspnea). Upon sacrifice, bone marrow (femurs, tibias and pelvic bones) and spleen cells were harvested in RPMI 1640 (Thermo Fisher Scientific, cat. no. 22400-089) supplemented with 1% FBS (Seradigm, cat.no. TXPLCA1400-500) and characterized by flow cytometry and histopathology. Approximately 1-5 X 10^5 leukemic cells were either enriched by flow cytometry (infiltration ≤ 85%) or directly lysed in Trizol (infiltration ≥ 85%) for RNA extraction and sequencing. One tibia per mouse was fixed in 10% neutral buffered formalin for histopathology.

NUP98 rearranged patient-derived xenograft (PDX) model was generated by intravenous injection of primary NUP98-BPTF AMKL cells collected at the stage of disease progression[1] and propagated by serial transplantation in NSG mice (600 000-2x10^6 cells per mouse).

Flow cytometry

The staining buffer for flow cytometry was comprised of 2% FBS and 1mM EDTA (Thermo Fisher Scientific, cat. no. 15575-020) in PBS. Cell surface proteins were blocked for non-specific binding (mouse gamma globulin, dilution 1/1000, Jackson ImmunoResearch Laboratories, Inc., PA, USA, cat no.015-000-002), stained with directly conjugated primary antibodies listed in Table S4, and washed with staining

buffer according to standard procedures. Analytic flow cytometry was conducted with a LSRFortessa (BD Biosciences, USA) and a FACSCanto II (BD Biosciences) cytometers. Automated cell counting was performed with LSRFortessa cytometer equipped with a high throughput sampler (BD Biosciences), by recording the number of events in a fixed sample volume. Cell were sorted with a FACSAria II flow cytometer (BD Biosciences), either at the flow cytometry platform of CHU Sainte-Justine or of the Institute for Research in Immunology and Cancer (Montreal, Canada).

Histopathology and microscopy

Peripheral blood smears, cytospin preparations of bone marrow and spleen cells, and organ touch preparations (e.g. touch preps of spleen, kidney, lung, liver, lymph nodes) were stained with Giemsa according to standard protocols. Formalin-fixed tibias were decalcified prior to paraffin embedding. Four micron thick tissue sections were deparaffinised prior to hematoxylin phloxin saffron staining and mounted with permanent mounting media, according to standard procedures. Cytospin and touch prep imaging were conducted with an Axio-Imager Z1 microscope (Zeiss, Germany) equipped with a 63X objective (1.4 Plan-Apochromat DIC III, Zeiss) and a color camera (Canon 5DMKII), at the bioimaging platform of the Institute for Research in Immunology and Cancer (Montreal, Canada). Cytospin, touch prep and tibia section imaging were also conducted with a DM6 upright microscope (Leica Microsystems,Wetzlar, Germany) equipped with 10X (HC PL FLUOTAR 10x/0.32, Leica Microsystems), 63X (HC PL APO 63x/1.40-0.60 oil, Leica Microsystems) and 100X objectives and a color camera, at the Platform for Imaging by Microscopy of CHU Sainte-Justine.

Molecular studies

RNA sequencing and variant calling

Quantification of total RNA was performed using a QuBit (Applied Biosystems, Thermo Fisher Scientific) and 500 ng of total RNA was used for Illumina sequencing library preparation. The quality of total RNA was assessed with the BioAnalyzer Nano (Agilent) and all samples had a RIN above 8 and sample purity was also assessed by Nanodrop using 260/280 and 260/230 ratios. Library preparation was done with the KAPA mRNAseq stranded kit (KAPA Biosystems, Thermo Fisher Scientific, Cat. no. KK8420) and quantified by QuBit and BioAnalyzer. All libraries were normalized and pooled to equimolar concentration by qPCR using the KAPA library quantification kit (KAPA Biosystems, Thermo Fisher Scientific, Cat. no. KK4973). Sequencing was performed on the Illumina HiSeq2000 or Nextseq500 with 200 and 150 cycles paired-end runs, respectively. Sequences were trimmed to remove sequencing adapters and low quality 3' bases using Trimmomatic version 0.35[2] and then aligned to the reference human genome version GRCh38 (gene annotation from Gencode version 26) using STAR version 2.5.1b[3]. Gene expression were estimated directly from STAR mapping as readcount values as well as computed using RSEM version 1.2.28[4] in order to obtain transcript level expression. Library preparation, sequencing, and data processing were performed at the Institute for Research in Immunology and Cancer's Genomics Platform (Montreal, Canada).

Variants were called using Freebayes (https://arxiv.org/abs/1207.3907) and then annotated using snpEff (http://snpeff.sourceforge.net/). Only mutations in regions that are covered at a 5X depth were kept. A list of 95 leukemia-related genes (see Table S6) was used to filter potential candidate variants relevant to the disease. Additional annotations related to the variants (coding or non-coding, alternative allele frequency

in public databases, *in silico* prediction on protein damage) was performed using the web interface wANNOVAR in order to prioritize variants[5]. Using these annotations, only rare and predicted damaging variations were retained.

Differential expression analysis was performed using the DESeq2 R package (version 1.16.1). Count tables were generated using HTSeq-count on BAM alignment files. Raw data count distributions were adjusted by regularised log transformation to generate heatmaps. Genes upregulated by at least two-fold in N5A AMKL samples (xenograft models and patients) compared to CB-CD34$^+$ cells were submitted to Gene Set Enrichment Analysis (GSEA) using annotated gene sets publicly available in the Molecular Signature Database (MSigDB)[6,7]. Distribution of selected gene expression values in bone marrow-derived human AML samples was generated with datasets extracted from the National Cancer Institute (NCI) TARGET database (https://ocg.cancer.gov/programs/target/data-matrix)[8].

Validation Cohort – Gene expression data

Annotated gene expression data used as validation cohort were generated from pediatric acute megakaryoblastic leukemia samples and analysed as described previously[9]. Briefly, Illumina paired-end sequencing reads were aligned to the GRCh37-lite genome build using an inhouse pipeline (StrongArm, unpublished). 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[10]. A gene was considered "expressed" if the FPKM value was >= 0.35 based on the distribution of FPKM values.

Enrichment analyses of CD3$^+$ GFP$^+$ cells

Human CD3$^+$ cells were flow-sorted from 2 independent NUP98-KDM5A AMKL xenograft models and transcriptome analyses were performed as described. Genes overexpressed in CD3$^+$ cells as compared to the transcriptome of AMKL models (Fold Change ≥ 5, mean FPKM ≥ 10 in CD3$^+$ cells; 725 GENES) were submitted to GO-term enrichment analyses using goseq (v1.34.1) and biomaRt (v2.38.0) packages.

PCA analyses from transcriptome

Principal component analyses were performed using readcounts data from in-house and TARGET cohorts with the DESeq2 package[11]. Readcount data were fitted against a negative binomial distribution with the DESeq function, transformed using varianceStabilizingTransformation (VST) function and the output was used along with the plotPCA function to calculate principal components to be plotted. Data from figure 3B were left batch-uncorrected. For figures with both in-house and TARGET cohorts, batch corrections were performed using the *removeBatchEffect* function from the limma package (v3.38.3). Specifically, three batches were corrected: in-house data, TARGET-1 and TARGET-2 previously identified via a PCA analysis of uncorrected values.

Statistical Analyses – Figure 5B

For each indicated gene, RPKM data from TARGET were compared between FAB categories using a Kruskal-Wallis test. Pairwise comparisons were performed using a Mann-Whitney rank sum test corrected by *Benjamini–Hochberg* for each pairwise comparisons.

Data visualization

Visualizations were generated in the R statistical environment (v3.4.1 - 3.5.1) using the BPG (v5.9.1 – 5.9.6) and ggplot2 (v3.1.0) package along with Inkscape (v0.91) and Illustrator (vCS6).

RT-PCR

Reverse transcription was conducted with 10ng total RNA and the "High Capacity cDNA Reverse Transcription Kit" (Applied Biosystems, ThermoFisher Scientific, cat. no. 4368814) according to the manufacturer's recommendations. cDNA products (2-3 ul) were amplified by PCR using primers specific to *N5A* breakpoint[12], *NEO1*, and *KDM5B* (positive control) (Table S7) and with Taq DNA Polymerase (Thermo Fisher Scientific, cat. no. 18038042). PCR cycling conditions were as follow: 94°C 3min for 1 cycle; 95°C 45s, 60°C 30s, 72°C 60s for 40 cycles; 72°C 7min for 1 cycle, and 4°C holding.

ChIP-Seq

Distribution of H3K4me3 and H3K27me3 histone modifications was determined by ChIP-Seq with chromatin extracts from two N5A cell lines or control CB-CD34^{+} cells. Cells were crosslinked with formaldehyde 1% for 7 minutes at room temperature. Cross-linking was quenched with 0.125nM glycine (final concentration) for 5 minutes at room temperature. Fixed cells were washed with PBS, resuspended in 1 mL of a 1st lysis buffer (0.25% Triton, 10mM Tris pH 8.0, 10mM EDTA, 0.5mM EGTA, 1X protease inhibitor) for 5 minutes, and in 1 mL of a 2nd lysis buffer for 30 minutes (200mM NaCl, 10mM Tris pH 8.0, 1mM EDTA, 0.5mM EGTA, 1X protease inhibitor). The lysate was resuspended in the sonication buffer (0.5% SDS, 0.5% Triton, 10mM Tris pH 8.0, 140mM NaCl, 1mM EDTA, 0.5mM EGTA, 1X protease inhibitor), sonicated with a Covaris sonicator (Duty Factor 10%, 200 cycles per burst, 105W, 4 minutes) and centrifuged at 13000 rpm for 5 minutes. The supernatant was used for immunoprecipitation.

Protein G Dynabeads™ were resuspended in immunoprecipitation buffer (1% Triton, 10mM Tris pH 8.0, 150mM NaCl, 2mM EDTA) and incubated overnight on a rotator at 4°C with 3µg of antibodies recognizing a specific histone mark (anti-histone H3, Abcam, USA, cat no. ab1791; anti-histone H3K4me3, Abcam, cat.no. ab8580; anti-histone H3K27me3, MilliporeSigma, cat. no. 07-449). Sonicated chromatin was incubated with 250µL of beads during 4 hours on a rotator at 4°C. The beads were washed once in 'Low Salt' Buffer (0.5% NP40, 15mM KCl, 10mM Tris pH 8.0, 1mM EDTA), once in 3 different 'High Salt' buffers (0.5% Triton, 10mM Tris pH 8.0, 100mM (2) or 400mM (3) or 500mM (4) NaCl), twice in LiCl Buffer (0.5% NP40, 250mM LiCl, 10mM Tris pH 8.0, 1mM EDTA), and finally in TE. To perform the elution step, ChIPed material was incubated overnight at 65°C with shaking (1200rpm). To reverse crosslink and purify DNA, an RNase and proteinase K digestion was performed, and DNA was extracted by a classic phenol-chloroform protocol. TruSeq ChIP Library Preparation Kit (Illumina, cat. no. IP-202-9001) was used to construct DNA libraries. HiSeq 2000 was used to sequence samples (parameter: 100bp paired-end) according to Illumina protocols. Quality of the raw sequence data was checked by FastQC (version 0.11.4) and Trimmomatic (version 0.32). DNA fragments were mapped on *hg19* genome reference using BWA (version 0.7.10). PCR duplicates were removed by Samtools (version 1.2). EaSeq (version 1.04) was used to further analyze data.

Cell surface proteomic analysis

Patient-derived xenotransplantation (PDX) of NUP98-BPTF bone marrow cells were produced in NSG mice. Harvested splenic megakaryoblastic cells of two secondary recipients were sorted on CD41-APC and pooled to perform cell surface proteomic analysis using a previously described procedure[13]. Proteins were then selected based on their detection in at least two replicates of the same biological condition, with a minimum of 2 unique peptides. These proteins were considered as identified with high-confidence. We

then manually sorted all identified proteins such that we kept those that contain at least one cell surface-exposed domain, which could be potentially biotinylated. To do so, we analyzed proteins identified with "high-confidence" by querying them against the UniProt online database (http://www.uniprot.org/). The final cell membrane protein list was compared to the "Hallmark gene set" using Metascape[14,15] .

Pharmacological inhibition assays

Xenograft cells were maintained in optimised cultures conditions in a serum free media supplemented with cytokines, 500 nM SR1[16] (Cedarlane) and 750 nM UM729[17] (STEMCELL Technologies). CMK and ML-2 cells were maintained RPMI 1640 media supplemented with 20% fetal bovine serum (Wisent Inc.). M07e cells were maintained RPMI 1640 media supplemented with 10% fetal bovine serum and 10ng/ml human IL-3 (Gibco). Pharmacological inhibition assays were conducted by the High-Throughput Screening Core Facility from The Institute for Research in Immunology and Cancer (IRIC, Montreal). Dose-response curves (10 data points, 1nM to 10μM, 3 fold dilutions) were generated by seeding 1000-5000 cells in 384-well plate format in quadruplicate and read-out was conducted with Cell Titer Glo (Promega) after 6 days of incubation at 37°C/5% CO_2 with inhibitors or vehicle (0.1% DMSO). Percentage of viability was calculated as follow: 100 x (mean luminescence compound/mean luminescence DMSO vehicle). IC_{50} concentrations were determined using a nonlinear regression analysis in GraphPad Prism version 7.03. Clofarabine (A10228), INCB018424 (ruxolitinib, A11041) and tofacitinib citrate (CP-690550) were purchased from AdooQ Bioscience.

REFERENCES

1 Roussy, M. *et al.* NUP98-BPTF gene fusion identified in primary refractory acute megakaryoblastic leukemia of infancy. *Genes, chromosomes & cancer* **57**, 311-319, doi:10.1002/gcc.22532 (2018).

2 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).

3 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

4 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).

5 Yang, H. & Wang, K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc* **10**, 1556-1566, doi:10.1038/nprot.2015.105 (2015).

6 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

7 Mootha, V. K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-273, doi:10.1038/ng1180 (2003).

8 Bolouri, H. *et al.* The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat Med* **24**, 103-112, doi:10.1038/nm.4439 (2018).

9 de Rooij, J. D. *et al.* Pediatric non-Down syndrome acute megakaryoblastic leukemia is characterized by distinct genomic subsets with varying outcomes. *Nat Genet* **49**, 451-456, doi:10.1038/ng.3772 (2017).

10 Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* **31**, 46-53, doi:10.1038/nbt.2450 (2013).

11 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

12 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, doi:10.1016/j.ccr.2012.10.007 (2012).

13 Goyette, M. A. *et al.* The Receptor Tyrosine Kinase AXL Is Required at Multiple Steps of the Metastatic Cascade during HER2-Positive Breast Cancer Progression. *Cell Rep* **23**, 1476-1490, doi:10.1016/j.celrep.2018.04.019 (2018).

14 Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications* **10**, 1523, doi:10.1038/s41467-019-09234-6 (2019).

15 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425, doi:10.1016/j.cels.2015.12.004 (2015).

16 Boitano, A. E. *et al.* Aryl hydrocarbon receptor antagonists promote the expansion of human hematopoietic stem cells. *Science (New York, N.Y.)* **329**, 1345-1348, doi:10.1126/science.1191536 (2010).

17 Pabst, C. *et al.* Identification of small molecules that support human leukemia stem cell activity ex vivo. *Nat Methods* **11**, 436-442, doi:10.1038/nmeth.2847 (2014).

Table S1. Characteristics of bone marrow-derived AML samples from pediatric patients.

Patient sample ID	Age at presentation	Immunophenotype	Karyotype	Genetic Alteration	Percentage of blasts (%)	Clinical presentation
pAMKL-1	17 months	$CD45^{dim}$, CD34 -, CD33+, CD41+, CD42b+, CD61+, CD71+, HLA_DR+, CD4+, MPO-	48,XY,t(6;13)(p21;q14),+der(6)t(6;13),+21[25]	*NUP98-KDM5A*[a]	50.6	Bruising, epistaxis,pancytopenia, CNS 1
pAMKL-2	30 months	$CD45^{dim}$, CD34-, CD33+, CD41+, CD42b+, CD61+, CD71+, CD117-	49,XX,t(1;12)(p32,p13),+4,+del(6)(q?21q?23),del(13)(q12q22),+21[19]/46,XX,t(1;12)(p32;p13),der(19)t(19;21)(q13;q21),add(21)(q11)[13]/ 46, XX[42]	*NUP98-KDM5A*[a]	60	Fatigue, pallor, anorexia, bruising, splenomegaly, pancytopenia, CNS 1
pAMKL-3D[b]	8 months	$CD45^{dim}$, CD34-, CD33-, CD42b+, MPO-, CD117-, TdT-, CD19-, CD10-, CD3-	47,XY,+6, t(11;17(p15;q23)[7]/48,idem, +19[2]/49,idem, +7,+19[9]/46,XY[2]. Ish t(11;17)(5'NUP98+;3'NUP98+)[4]	*NUP98-BPTF*[a]	N/A	Bruising, petechiæ, conjunctival hemorrhage, mild hepatosplenomegaly, pancytopenia, CNS1
pAMKL-3P[b]	13 months	$CD45^{dim}$, CD34-, CD41+, CD61+	49,XY,+6,+7,t(11;17)(p15;q23),+19[25] (initial progression) 49,XY,+6,+7,(t11;17)(p15;q23),+19 [20]/49, idem,t(7;14)[3]/49, idem,t(3;4),t(10;21)[2] (uncontrolled disease)	*NUP98-BPTF*[a]	N/A	N/A
pAMKL-4	2 years	CD45-, CD34+,CD41+,CD42b+, CD61+, CD117 +, PAS+,TdT-	46,XY, t(1;16)(q21;p11,2).ish t(1;16)	*CBFA2T3-GLIS2*[a]	93	Fatigue, recurrent fever, anorexia, anemia and thrombocytopenia, CNS 1
pAMKL-5	18 months	CD34-,CD41+,CD61+, CD42b+	N/A (referred case)	*RMB15-MKL1*[a]	49	Bony pain, anemia, thrombocytopenia, CNS 2
pAML-6	16 years	$CD45^{dim}$, CD34+, CD33+, CD13+, CD11a+, CD117+, CD123+, HLA-DR+	46, XX	*NUP98-NSD1*[a] *FLT3 ITD*[c]	48.8% in PB	Fatigue, diffuse adenopathy, fever, night sweats, headache, respiratory distress, hyperleukocytosis, anemia, thrombocytopenia, CNS 2

[a] Fusion genes were identified by RNA sequencing.[b] pAMKL-3D and pAMKL-3P, diagnostic and disease progression sample, respectively (Roussy et al, *Genes Chromosomes Cancer*, 2018). [c] *FLT3* status identified by targeted PCR. AML, acute myeloid leukemia; AMKL, acute megakaryoblastic leukemia; N/A, not available; CNS, central nervous system infiltration.

Table S2. Frequency of leukemia subtypes in *NUP98-KDM5A* human xenograft models.

		CTL	N5A
EXP1	AMKL	0	1
(ctl n=3; N5A n=3)	AML-O	0	0
	B-ALL	0	1
	T-ALL	0	0
	Death unrelated to leukemia	3	1
	Total leukemia λ	**0/3**	2/3
EXP2	AMKL	0	1
(ctl n=4; N5A n=12)	AML-O	0	0
	B-ALL	0	2
	T-ALL	0	0
	Death unrelated to leukemia	4	9
	Total leukemia λ	**0/4**	**3/12**
EXP3	AMKL	0	0
(ctl n=3; N5A n=9)	AML-O	0	1
	B-ALL	0	0
	T-ALL	0	1
	Death unrelated to leukemia	3	7
	Total leukemia λ	**0/3**	**2/9**
EXP4	AMKL	0	2
(ctl n=6; N5A n=8)	AML-O	0	0
	B-ALL	0	3
	T-ALL	0	0
	Death unrelated to leukemia	6	3
	Total leukemia λ	**0/6**	**5/8**
EXP5	AMKL	0	1
(ctl n=2; N5A n=9)	AML-O	0	4
	B-ALL	0	1
	T-ALL	0	0
	Death unrelated to leukemia	2	3
	Total leukemia λ	**0/2**	***6/9***
EXP6	AMKL	0	1
(ctl n=12; N5A n=28)	AML-O	0	6
	B-ALL	0	0
	T-ALL	0	0
	Death unrelated to leukemia	12	21
	Total leukemia λ	***0/12***	***7/28***

EXP, experiment; transplanted cells transduced with ctl (control) or N5A (*NUP98-KDM5A*) vector; AMKL, acute megakaryoblastic leukemia; AML-O, acute myeloid leukemia other (non-AMKL); B- or T-ALL, B- or T-cell acute lymphoblastic leukemia; λ, frequency.

Table S3. Characteristics of leukemias in primary NUP98-KDM5A xenograft recipients.

Leukemia subtype	Xenograft ID	Mouse ID	Latency (wks)	Brittle bones	Blasts infiltration in Bone Marrow (%)	Splenomegaly	Spleen weight (mg)	Blasts infiltration in Spleen (%)	Blast Immunophenotype
AMKL									
	xAMKL-1	**C362**	**34**	**yes**	**31,9**	**Mild**	**N/A**	**0,1**	**hCD45^{low}GFP$^+$CD34$^-$CD61$^+$**
	xAMKL-2	**D922**	**37**	**yes**	**15.9**	**Mild**	**N/A**	**0.8**	**hCD45^{low}GFP$^+$CD34$^-$CD41$^+$CD61$^+$**
	xAMKL-3	**E745**	**48**	**yes**	**34,7**	**Mild**	**76**	**0,34**	**hCD45^{low}GFP$^+$CD34$^-$CD41$^+$CD61$^+$**
	xAMKL-4	E771	63	yes	55,2	Mild	76	3,5	hCD45^{low}GFP$^+$CD34$^-$CD41$^+$CD61$^+$
	xAMKL-5	**E890**	**71**	**yes**	**37,1**	**Mild**	**69**	**N/A**	**hCD45^{low}GFP$^+$CD34$^-$CD41$^+$CD61$^+$**
	xAMKL-6 *	F821	46	partially	5,3 5,4	Severe	330	1,4 47,1	hCD45^{low+}GFP$^+$CD34$^-$CD41$^+$CD36$^+$ / hCD45$^+$GFP$^+$CD34$^-$CD41$^-$CD36$^+$
AML-O									
	xAML-O-1	**E385**	**40**	**yes**	**77,3**	**Mild**	**70**	**43,2**	**hCD45$^+$GFP$^+$CD34$^-$CD33$^-$CD41$^-$CD61$^-$CD117$^+$CD36$^+$**
	xAML-O-2	**E737**	**32**	**yes**	**95,3**	**Moderate**	**273**	**46,3**	**hCD45$^+$GFP$^+$CD34$^-$CD61$^-$CD71$^+$CD117$^{+/-}$CD36$^+$**
	xAML-O-3	**F823**	**34**	**yes**	**89,1**	**Severe**	**342**	**77,6**	**hCD45$^+$GFP$^+$CD34$^-$CD33$^{+/-}$CD41$^-$CD71$^+$CD117$^{+/-}$CD36$^+$**
	xAML-O-4	E892	64	yes	52,5	Mild	64	0,48	hCD45$^{low/+}$GFP$^+$CD34$^-$CD41$^-$CD36$^+$
	xAML-O-5	E884	56	yes	55,3	Severe	531	23	hCD45$^+$GFP$^+$CD34$^-$CD61$^-$CD117$^-$CD36$^+$
	xAML-O-6	E891	67	yes	62,2	Moderate	176	10,8	hCD45$^+$GFP$^+$CD34$^-$CD33$^+$CD117$^-$CD36$^+$
	xAML-O-7	F827	46	yes	19,7	Severe	630	21,2	hCD45$^+$GFP$^+$CD34$^-$CD33$^-$CD41$^-$CD36$^+$
	xAML-O-8	E857	67	yes	30,6	Mild	61	1,07	hCD45$^+$GFP$^+$CD34$^-$CD41$^-$CD33$^+$CD117$^-$CD36$^+$
	xAML-O-9	F788	27	yes	72,2	Moderate	261	90,4	hCD45$^+$GFP$^+$CD34$^-$CD61$^-$CD117$^{+/-}$CD36$^+$
	xAML-O-10	F832	27	yes	54,3	no	32	10,5	hCD45$^+$GFP$^+$CD34$^-$CD61$^-$CD71$^+$CD117$^{+/-}$CD36$^+$
	xAML-O-11	F802	34	yes	27,6	Mild	109	37,1	hCD45$^+$GFP$^+$CD34$^-$CD33$^+$CD41$^-$CD71$^+$CD117$^{+/-}$CD36$^+$
B-ALL									
	xB-ALL-1	**C384**	**50**	**yes**	**95,2**	**Severe**	**N/A**	**91,2**	**hCD45$^+$GFP$^-$CD19$^+$CD10$^+$CD20$^-$CD38$^+$CD34$^{+/-}$**
	xB-ALL-2	**D918**	**37**	**yes**	**93,4**	**Severe**	**N/A**	**49,7**	**hCD45$^+$GFP$^+$CD19$^+$CD10$^+$CD20$^-$CD38$^+$CD34$^{+/-}$**
	xB-ALL-3	**D917**	**40**	**yes**	**31,6**	**Severe**	**N/A**	**59,9**	**hCD45$^+$GFP$^+$CD19$^+$CD10$^+$CD20$^-$CD38$^+$CD34$^{+/-}$**
	xB-ALL-4	**E755**	**39**	**yes**	**77,3**	**Severe**	**762**	**76,8**	**hCD45$^+$GFP$^-$CD19$^+$CD34$^{+/-}$**
	xB-ALL-5	E742	59	yes	87,1	Severe	794	82,0	hCD45$^+$GFP$^+$CD19$^+$CD34$^{+/-}$
	xB-ALL-6	E743	59	yes	99,8	Moderate	205	99,7	hCD45$^+$GFP$^-$CD19$^+$CD34$^{+/-}$
	xB-ALL-7	E854	67	partially	16,8	Severe	310	1,2	hCD45$^+$GFP$^+$CD19$^+$CD34$^{+/-}$
T-ALL									
	xT-ALL-1	E389	61	yes	61,6	Severe	601	29,8	hCD45$^+$GFP$^+$CD3$^+$CD4$^+$CD8$^-$HLA-DR$^+$

AMKL, acute megakaryoblastic leukemia; AML-O, acute myeloid leukemia others (non-AMKL); B- or T-ALL, B- or T-cell acute lymphoblastic leukemia; wks, weeks; N/A, not available. *For sample AMKL-6, 2 blast populations were detected. Samples in bold characters were used for RNA sequencing.

Table S4. Primary antibody list.

Target Protein	Conjugation	Host, isotype, clone	Company	Cat. No.	Dilution for FACS
CD3	APC	Mouse IgG2a, κ, clone HIT3a	Biolegend	300312	3.0 : 100
CD4	PE-Cy7	Rat IgG2b, κ , clone A161A1	Biolegend	357409	2.0 : 100
CD7	PE	Mouse BALB/c IgG1, κ , clone M-T701	BD Biosciences	340581	2.0 : 100
CD8	PerCP	Mouse IgG1, κ , clone SK1	Biolegend	344707	3.5 : 100
CD10/MME	BV421	Mouse (BALB/c) IgG1,k, clone HI10a	BD Biosciences	562902	4.0 : 100
CD15	BV421	Mouse IgG1, κ; clone W6D3	Biolegend	323039/323040	2.5 : 100
CD19	APC	Mouse IgG1, κ, clone HIB19	BD Pharmingen	555415	2.5 : 100
CD19	BV421	Mouse IgG1, κ, clone HIB19	Biolegend	302234	3.0 : 100
CD19	PE	Mouse IgG1, κ, clone HIB19	Biolegend	302208	2.0 : 100
CD20	PerCP-Cy5.5	Mouse IgG2b, κ, clone 2H7	Biolegend	302326	2.5 : 100
CD33	PE-Cy7	Mouse IgG1, κ, clone WM-53	ThermoFisher Scientific /eBioscience	25-0338-42	1.5 : 100
CD34	APC	Mouse IgG1, κ, clone 581	BD Biosciences	555824	3.0 : 100
CD34	BV421	Mouse IgG1, κ, clone 581	BD Biosciences	562577	3.0 : 100
CD34	PE	Mouse IgG1, κ, clone 8G12	BD Biosciences	348057	2.0 : 100
CD36	APC	Mouse IgG2a, κ, clone 5-271	Biolegend	336208	1.25 : 100
CD36	PE	Mouse IgG2a, κ, clone 5-271	Biolegend	336206	2.0 : 100
CD38	PE-Cy7	Mouse IgG1, κ, clone HIT2	ThermoFisher Scientific /eBioscience	25-0389-42	1.2 : 100
CD41/ITGA2B	APC	Mouse IgG1, κ, clone H1P8	Biolegend	303710	3.0 : 100
CD41/ITGA2B	APC-Cy7	Mouse IgG1, κ, clone H1P8	Biolegend	303716	2.0 : 100
CD45/PTPRC	APC-Cy7	Mouse IgG1, κ, clone HI30	Biolegend	304014	3.5 : 100
CD45/PTPRC	PerCP	Mouse IgG1, κ, clone HI30	Biolegend	304026	4.0 : 100
CD45/PTPRC	PE	Rat IgG2b, κ, clone 30-F11	Biolegend	103106	0.1 : 100
CD61/ITGB3	PerCP-Cy5.5	Mouse IgG1, κ, clone VI-PL2	BD Biosciences	564173	3.0 : 100
CD62P/SELP	PE-Cy7	Mouse IgG1, κ, clone AK4	Biolegend	304922	2.0 : 100
CD71	PE-Cy7	Mouse IgG2a, κ, clone CY1G4	Biolegend	334112	2.0 : 100
CD117/KIT	PE	Mouse IgG1, κ , clone 104D2	ThermoFisher Scientific /eBioscience	12-1178-42	2.0 : 100
CD235a/Glycophorin-A	PE-Cy7	Mouse IgG2b, κ, clone GA-R2, HIR2	BD Biosciences	563666	1.25 : 100
HLA-DR	PE	Mouse BALB/c IgG2a, κ , clone L243	BD Biosciences	347367	2.0 : 100

Table S5. Descriptive list of RNAseq samples.

Sample ID	Fusion gene	Sample description	GEO dataset (GSE123485) ID
Patients_AMKL			
pAMKL-1	*NUP98-KDM5A*	BM[b]	pAMKL-1_12H093
pAMKL-2	*NUP98-KDM5A*	BM[b]	pAMKL-2_14H014
pAMKL-3D (diagnostic)[a]	*NUP98-BPTF*	BM[b]	pAMKL-3D
pAMKL-3P (progression)[a]	*NUP98-BPTF*	BM[b]	pAMKL-3P_BM_019
pAMKL-4	*CBFA2T3-GLIS2*	BM[b]	pAMKL-4_15H014
pAMKL-5	*RMB15-MKL1*	BM[b]	pAMKL-5_BM_007
Patient_AML			
pAML-6	*NUP98-NSD1*	BM[b]	pAML-6_BM_005
Xenograft_AMKL			
xAMKL-1	*NUP98-KDM5A*	FACS sorted $CD3^-GFP^+$ BM cells	xAMKL-1_C362
xAMKL-2	*NUP98-KDM5A*	FACS sorted $CD3^-GFP^+$ BM cells	xAMKL-2_D922
xAMKL-3	*NUP98-KDM5A*	FACS sorted $CD3^-GFP^+$ BM cells	xAMKL-3_E745
xAMKL-5	*NUP98-KDM5A*	FACS sorted GFP^+ BM cells	xAMKL-5_E890
Xenograft_AMKL-associated T-cells			
xAMKL-1_associated CD3+ Cells	N/A	FACS sorted $CD3^+$ BM cells	xAMKL-1_C362_CD3
xAMKL-2_associated CD3+ Cells	N/A	FACS sorted $CD3^+$ BM cells	xAMKL-2_D922_CD3
Xenograft_AML-O			
xAML-O-1	*NUP98-KDM5A*	BM[b]	xAML-1_E385
xAML-O-2	*NUP98-KDM5A*	BM[b]	xAML-2_F737
xAML-O-3	*NUP98-KDM5A*	BM[b]	xAML-3_F823
Xenograft_B-ALL			
xB-ALL-1	*NUP98-KDM5A*	BM[b]	xB-ALL-1_C384
xB-ALL-2	*NUP98-KDM5A*	FACS sorted $CD19^+$ spleen cells	xB-ALL-2_D918
xB-ALL-3	*NUP98-KDM5A*	FACS sorted $CD19^+$ spleen cells	xB-ALL-3_D917
xB-ALL-4	*NUP98-KDM5A*	FACS sorted $CD19^+$ spleen cells	xB-ALL-4_E755
CB-$CD34^+$ cells			
CB-$CD34^+$ -1	none	MACS sorted $CD34^+$ CB cells	CB-$CD34^+$-1
CB-$CD34^+$ -2	none	MACS sorted $CD34^+$ CB cells	CB-$CD34^+$-2
CB-$CD34^+$ -3	none	MACS sorted $CD34^+$ CB cells	CB-$CD34^+$-3
CB-$CD34^+$ -4	none	MACS sorted $CD34^+$ CB cells	CB-$CD34^+$-4

[a]pAMKL-3D and pAMKL-3P, diagnostic and disease progression sample, respectively (Roussy et al, Genes Chromosomes Cancer, 2018). [b]Unsorted bone marrow (BM) cells. AMKL, acute megakaryoblastic leukemia; AML-O, acute myeloid leukemia others (non-AMKL); B-ALL, B-cell acute lymphoblastic leukemia; CB, cord blood; FACS and MACS, fluorescence and magnetic -activated cell sorting, respectively. N/A, not available.

Table S6. List of 95 leukemia-related genes considered for calling variants from RNA sequencing data.

ABL1	*DGKH*	*IL7R*	*NT5C2*	*SH2B3*
ABL2	*DNMT3A*	*JAK1*	*NTRK3*	*SMC1A*
AKT3	*DYRK1A*	*JAK2*	*NUP98*	*SMC3*
ARID5B	*EBF1*	*JAK3*	*PAG1*	*SRSF2*
ASXL1	*EP300*	*KDM5A*	*PAX5*	*STAG2*
ASXL2	*EPOR*	*KDM6A*	*PDGFRB*	*TERT*
ATRX	*ETV6*	*KIF2B*	*PHF6*	*TET2*
BCOR	*EZH2*	*KIT*	*PIP4K2A*	*TP53*
BMI1	*FBXW7*	*KMT2A*	*PRPF40B*	*TP63*
BRAF	*FLT3*	*KMT2C*	*PTEN*	*TSLP*
CBL	*GATA1*	*KMT2D*	*PTK2B*	*TYK2*
CBLB	*GATA2*	*KMT2E*	*PTPN11*	*U2AF1*
CDKN2A	*GATA3*	*KRAS*	*PTPRJ*	*U2AF2*
CDKN2B	*GNAS*	*MAPK1*	*RAD21*	*WT1*
CEBPA	*IDH1*	*MPL*	*RB1*	*ZRSR2*
CEBPE	*IDH2*	*MYC*	*RUNX1*	
CREBBP	*IKZF1*	*NF1*	*SETD2*	
CRLF2	*IKZF2*	*NOTCH1*	*SF1*	
CSF1R	*IKZF3*	*NPM1*	*SF3A1*	
CUX1	*IL2RB*	*NRAS*	*SF3B1*	

Table S7. Primer list.

RT-PCR assay	Primer name	sequence	Primer name	sequence	Product size (bp)
NUP98-KDM5A breakpoint	NUP98-KDM5A_F2	GTAAACCAGCACCTGGGACTCTTGG	NUP98-KDM5A_R2	GCCCCTGCTTCTTTGCACAGTTTAT	740
NEO1	NEO1_F	GGTTCTTCCAGATCCTGAGGTG	NEO1_R	TGACCCACTTCACAGTTGGAG	413
KDM5B	KDM5B_F	TGTCACAGTGGAATATGGAGCTGAC	KDM5B_R	CATCACTGGCATGTTGTTCAAATTC	142

Table S8. Expression values in FPKM (RNAseq) of genes encoding cell surface proteins that are differentially expressed by a least 10-fold in leukemic bone marrow cells derived from mouse xenograft models (xAMKL) and patients (pAMKL-1 and -2) presenting with NUP98-KDM5A (N5A) AMKL as compared to normal CB-CD34+ cells, and expressed at low levels (≤5 FPKM) in CB-CD34+ cells.

	CB-CD34+	N5A xAMKL-				N5A xAML-O-			xB-ALL-				NUP98r pAMKL				non-NUP98r pAMKL		NUP98r pAML	NUP98r pAMKL validation cohort
Gene Name	n=4	1	2	3	5	1	3	2	1	2	3	4	1	2	3D	3P	4	5	6	n=7
RHAG	4.26	1164.20	430.40	362.93	370.12	0.00	0.00	0.00	0.00	0.20	0.15	1.03	266.93	341.55	334.04	784.33	511.54	218.75	1.34	376.88
MPIG6B	4.26	206.13	249.39	254.63	217.05	0.00	0.03	0.06	1.99	5.59	1.38	2.48	163.63	74.37	64.54	227.10	7.68	48.06	8.21	166.12
SELP	0.69	321.65	318.11	132.96	239.13	0.19	0.12	0.16	0.51	0.30	0.18	0.68	48.32	34.06	12.62	13.76	0.05	0.52	6.07	52.47
NEO1	0.33	125.39	63.43	106.11	116.35	0.39	1.51	0.72	0.20	0.23	0.26	1.01	144.67	47.66	34.73	92.21	268.10	8.38	0.21	36.44
CD96	1.83	123.50	125.65	95.56	107.14	0.05	0.08	0.04	25.05	7.36	9.58	41.56	127.71	49.80	18.06	29.17	129.02	54.96	112.91	38.81
LTK	1.00	105.21	150.96	53.61	104.74	16.49	20.17	8.95	13.46	5.73	9.80	26.21	162.38	21.93	3.09	7.18	16.31	5.84	1.81	57.77
GP9	2.40	111.32	108.81	100.07	107.56	0.15	0.04	0.11	0.08	0.13	0.62	0.36	49.23	26.87	10.02	24.77	11.06	113.09	4.02	27.31
KEL	2.64	67.09	56.27	54.81	67.15	0.22	0.03	0.09	0.21	0.14	0.46	0.88	88.88	52.73	59.79	113.44	1.52	33.88	0.49	65.96
EPOR	3.24	23.42	29.36	36.38	26.52	3.43	2.73	2.39	0.82	4.00	1.95	0.68	188.83	17.97	48.12	116.40	0.38	30.63	1.47	24.71
TRPM4	1.46	74.01	46.20	38.56	15.74	63.23	35.81	15.88	8.08	3.45	5.51	3.23	26.26	10.14	2.09	5.28	0.73	3.24	0.10	19.27
FRAS1	0.06	19.67	34.43	53.24	48.26	0.01	0.00	0.00	0.01	0.06	0.01	0.08	43.78	22.27	0.26	1.64	11.48	5.62	0.00	13.89
XK	1.75	36.26	27.25	21.91	44.67	0.00	0.04	0.04	0.00	0.03	0.01	0.05	30.94	42.11	25.14	36.84	26.72	20.34	3.38	33.99
SMIM1	2.39	22.74	20.23	13.28	25.38	0.16	0.16	0.63	0.31	0.10	0.11	0.11	42.29	39.63	28.02	23.02	6.12	16.72	1.13	10.50
PTGER3	0.52	24.75	14.43	32.23	58.59	0.09	0.13	0.20	0.03	0.00	0.00	0.20	39.28	7.88	13.72	40.36	0.00	19.10	0.01	18.91
PDZK1IP1	1.38	31.82	9.93	15.33	26.79	0.00	0.00	0.00	0.00	0.06	0.16	0.19	41.54	16.75	4.93	5.05	0.03	0.51	0.45	14.47
SLC13A3	0.08	25.14	20.99	33.68	27.06	0.05	0.03	0.01	0.04	0.03	0.04	0.41	13.50	6.19	3.91	10.35	0.26	5.53	0.00	10.05
PTCH1	0.41	7.43	11.51	35.80	11.82	0.02	0.06	0.31	18.73	16.82	8.45	5.26	19.14	13.98	6.47	4.43	9.81	3.66	0.06	7.66
KCNJ4	0.02	21.35	22.14	11.98	36.33	4.57	0.57	0.10	26.26	34.74	2.62	18.42	18.98	8.87	0.00	0.05	0.00	0.00	0.00	10.50
MFSD2B	1.51	20.75	13.76	8.69	10.01	0.09	0.02	0.03	0.04	0.10	0.09	0.08	22.12	19.95	19.96	34.35	1.79	26.39	1.25	11.27
TRHDE	0.00	21.63	8.22	7.05	15.74	0.00	0.00	0.01	0.00	0.00	0.00	0.01	12.92	33.97	5.19	13.79	0.00	0.00	0.00	6.16
PCDH9	1.05	24.04	13.68	10.02	3.93	0.00	0.17	0.01	11.83	21.44	63.49	38.88	14.81	6.67	5.88	0.88	1.50	1.69	0.04	20.40
AMHR2	0.81	17.01	13.28	13.03	19.87	0.00	0.00	0.00	0.03	0.00	0.01	0.13	7.66	7.66	3.28	10.19	0.03	5.33	0.14	2.70
NCAM1	0.52	0.00	3.24	0.32	0.07	109.70	85.47	32.71	0.00	0.05	0.01	0.00	1.04	0.50	1.04	0.16	652.11	2.29	0.03	0.24
ITGA2B	21.67	1328.46	944.06	1234.69	1214.23	0.43	1.23	1.23	0.12	0.51	0.48	2.23	567.02	266.15	264.60	2688.02	407.53	587.24	46.20	361.09
ITGB3	10.21	80.85	68.19	32.03	56.88	0.00	0.01	0.00	0.03	0.10	0.11	0.27	25.68	15.94	16.06	32.90	17.63	62.54	7.78	22.30

NUP98-KDM5A (N5A) acute megakaryoblastic leukemia xenograft models (xAMKL-); N5A acute myeloid leukemia (AML)-monocytoid xenograft models (N5A xAML-O-); N5A B-cell acute lymphoblastic leukemia xenograft models (xB-ALL-); AMKL patient samples (pAMKL) with or without NUP98 rearrangements (NUP98r or non-NUP98r, respectively); pAMKL-1 and -2, N5A; pAMKL-3D and -3P, NUP98-BPTF diagnostic and disease progression samples; pAMKL-4, CBFA2T3-GLIS2; pAMKL-5, RBM15-MKL1; pAML-6, NUP98-NSD1. NUP98r pAMKL validation cohort (de Rooij, J.D.E., *Nature Genetics*, 2017). Megakaryocyte lineage/AMKL markers used for diagnosis are indicated in blue.

Table S10. Hallmark Get Sets enrichment associated with proteins expressed at the cell surface of primary NUP98-BPTF AMKL PDX cells.

GroupID	Category	Term	Description	LogP	Log(q-value)	InTerm_InList	Genes	Symbols
1_Summary	Hallmark Gene Sets	M5924	HALLMARK MTORC1 SIGNALING	-12,2141	-10,241	23/200	488,811,871,928,1595,1717,2023,2181,3309,3689,40 74,5052,5478,6510,6513,6515,6533,7037,7167,7184, 8140,9415,60481,6275,6576,10994,26355	ATP2A2,CALR,SERPINH1,CD9,CYP51A1,DHCR7,ENO1,ACSL3,HSPA5,ITGB2,M6PR,PRD X1,PPIA,SLC1A5,SLC2A1,SLC2A3,SLC6A6,TFRC,TPI1,HSP90B1,SLC7A5,FADS2,ELOVL5, S100A4,SLC25A1,ILVBL,FAM162A
1_Member	Hallmark Gene Sets	M5924	HALLMARK MTORC1 SIGNALING	-12,2141	-10,241	23/200	488,811,871,928,1595,1717,2023,2181,3309,3689,40 74,5052,5478,6510,6513,6515,6533,7037,7167,7184, 8140,9415,60481	ATP2A2,CALR,SERPINH1,CD9,CYP51A1,DHCR7,ENO1,ACSL3,HSPA5,ITGB2,M6PR,PRD X1,PPIA,SLC1A5,SLC2A1,SLC2A3,SLC6A6,TFRC,TPI1,HSP90B1,SLC7A5,FADS2,ELOVL5
1_Member	Hallmark Gene Sets	M5891	HALLMARK HYPOXIA	-2,62987	-2,072	10/200	2023,3309,6275,6513,6515,6533,6576,7167,10994,26 355	ENO1,HSPA5,S100A4,SLC2A1,SLC2A3,SLC6A6,SLC25A1,TPI1,ILVBL,FAM162A
2_Summary	Hallmark Gene Sets	M5936	HALLMARK OXIDATIVE PHOSPHORYLATION	-9,59234	-7,920	20/200	22,292,481,498,506,529,1537,2110,2806,4259,4718,5 250,6834,7416,7417,7419,8402,8604,10935,11331,70 8,3735,5245,5478,5707,10549	ABCB7,SLC25A5,ATP1B1,ATP5F1A,ATP5F1B,ATP6V1E1,CYC1,ETFDH,GOT2,MGST3,N DUFC2,SLC25A3,SURF1,VDAC1,VDAC2,VDAC3,SLC25A11,SLC25A12,PRDX3,PHB2,C1 QBP,KARS,PHB,PPIA,PSMD1,PRDX4
2_Member	Hallmark Gene Sets	M5936	HALLMARK OXIDATIVE PHOSPHORYLATION	-9,59234	-7,920	20/200	22,292,481,498,506,529,1537,2110,2806,4259,4718,5 250,6834,7416,7417,7419,8402,8604,10935,11331	ABCB7,SLC25A5,ATP1B1,ATP5F1A,ATP5F1B,ATP6V1E1,CYC1,ETFDH,GOT2,MGST3,N DUFC2,SLC25A3,SURF1,VDAC1,VDAC2,VDAC3,SLC25A11,SLC25A12,PRDX3,PHB2
2_Member	Hallmark Gene Sets	M5926	HALLMARK MYC TARGETS V1	-4,38202	-3,488	13/200	708,1537,2806,3735,5245,5250,5478,5707,7416,7419 ,10549,10935,11331	C1QBP,CYC1,GOT2,KARS,PHB,SLC25A3,PPIA,PSMD1,VDAC1,VDAC3,PRDX4,PRDX3,PH B2
3_Summary	Hallmark Gene Sets	M5905	HALLMARK ADIPOGENESIS	-7,96471	-6,469	18/200	348,483,948,1468,1537,1717,4259,5733,6510,6576,7 905,8884,10162,10935,23344,23788,55177,92840	APOE,ATP1B3,CD36,SLC25A10,CYC1,DHCR7,MGST3,PTGER3,SLC1A5,SLC25A1,REEP5, SLC5A6,LPCAT3,PRDX3,ESYT1,MTCH2,RMDN3,REEP6
3_Member	Hallmark Gene Sets	M5905	HALLMARK ADIPOGENESIS	-7,96471	-6,469	18/200	348,483,948,1468,1537,1717,4259,5733,6510,6576,7 905,8884,10162,10935,23344,23788,55177,92840	APOE,ATP1B3,CD36,SLC25A10,CYC1,DHCR7,MGST3,PTGER3,SLC1A5,SLC25A1,REEP5, SLC5A6,LPCAT3,PRDX3,ESYT1,MTCH2,RMDN3,REEP6
4_Summary	Hallmark Gene Sets	M5930	HALLMARK EPITHELIAL MESENCHYMAL TRANSITION	-7,18999	-5,819	17/200	290,355,813,871,960,1004,3673,3678,3685,3688,369 0,3693,3956,4853,5479,7040,7057	ANPEP,FAS,CALU,SERPINH1,CD44,CDH6,ITGA2,ITGA5,ITGAV,ITGB1,ITGB3,ITGB5,LGA LS1,NOTCH2,PPIB,TGFB1,THBS1
4_Member	Hallmark Gene Sets	M5930	HALLMARK EPITHELIAL MESENCHYMAL TRANSITION	-7,18999	-5,819	17/200	290,355,813,871,960,1004,3673,3678,3685,3688,369 0,3693,3956,4853,5479,7040,7057	ANPEP,FAS,CALU,SERPINH1,CD44,CDH6,ITGA2,ITGA5,ITGAV,ITGB1,ITGB3,ITGB5,LGA LS1,NOTCH2,PPIB,TGFB1,THBS1
5_Summary	Hallmark Gene Sets	M5950	HALLMARK ALLOGRAFT REJECTION	-5,03733	-3,763	14/200	355,567,924,958,961,2149,2589,3133,3383,3683,368 9,4065,5788,7040	FAS,B2M,CD7,CD40,CD47,F2R,GALNT1,HLA-E,ICAM1,ITGAL,ITGB2,LY75,PTPRC,TGFB1
5_Member	Hallmark Gene Sets	M5950	HALLMARK ALLOGRAFT REJECTION	-5,03733	-3,763	14/200	355,567,924,958,961,2149,2589,3133,3383,3683,368 9,4065,5788,7040	FAS,B2M,CD7,CD40,CD47,F2R,GALNT1,HLA-E,ICAM1,ITGAL,ITGB2,LY75,PTPRC,TGFB1
6_Summary	Hallmark Gene Sets	M5897	HALLMARK IL6 JAK STAT3 SIGNALING	-4,7593	-3,564	9/87	355,928,948,952,960,3676,3690,5770,7040	FAS,CD9,CD36,CD38,CD44,ITGA4,ITGB3,PTPN1,TGFB1
6_Member	Hallmark Gene Sets	M5897	HALLMARK IL6 JAK STAT3 SIGNALING	-4,7593	-3,564	9/87	355,928,948,952,960,3676,3690,5770,7040	FAS,CD9,CD36,CD38,CD44,ITGA4,ITGB3,PTPN1,TGFB1
7_Summary	Hallmark Gene Sets	M5892	HALLMARK CHOLESTEROL HOMEOSTASIS	-4,43082	-3,488	8/74	214,552,928,1595,1717,2222,9415,50814	ALCAM,AVPR1A,CD9,CYP51A1,DHCR7,FDFT1,FADS2,NSDHL
7_Member	Hallmark Gene Sets	M5892	HALLMARK CHOLESTEROL HOMEOSTASIS	-4,43082	-3,488	8/74	214,552,928,1595,1717,2222,9415,50814	ALCAM,AVPR1A,CD9,CYP51A1,DHCR7,FDFT1,FADS2,NSDHL
8_Summary	Hallmark Gene Sets	M5910	HALLMARK PROTEIN SECRETION	-4,41492	-3,488	9/96	102,476,3482,4074,9341,10067,10972,11079,81542	ADAM10,ATP1A1,IGF2R,M6PR,VAMP3,SCAMP3,TMED10,RER1,TMX1
8_Member	Hallmark Gene Sets	M5910	HALLMARK PROTEIN SECRETION	-4,41492	-3,488	9/96	102,476,3482,4074,9341,10067,10972,11079,81542	ADAM10,ATP1A1,IGF2R,M6PR,VAMP3,SCAMP3,TMED10,RER1,TMX1
9_Summary	Hallmark Gene Sets	M5947	HALLMARK IL2 STAT5 SIGNALING	-4,40458	-3,488	13/199	214,659,960,975,3480,3482,3655,3685,6403,6510,65 15,9144,253558	ALCAM,BMPR2,CD44,CD81,IGF1R,IGF2R,ITGA6,ITGAV,SELP,SLC1A5,SLC2A3,SYNGR2, LCLAT1
9_Member	Hallmark Gene Sets	M5947	HALLMARK IL2 STAT5 SIGNALING	-4,40458	-3,488	13/199	214,659,960,975,3480,3482,3655,3685,6403,6510,65 15,9144,253558	ALCAM,BMPR2,CD44,CD81,IGF1R,IGF2R,ITGA6,ITGAV,SELP,SLC1A5,SLC2A3,SYNGR2, LCLAT1
10_Summary	Hallmark Gene Sets	M5934	HALLMARK XENOBIOTIC METABOLISM	-4,38202	-3,488	13/200	348,355,488,948,1312,1969,2110,3735,6510,6533,10 653,10857,60481	APOE,FAS,ATP2A2,CD36,COMT,EPHA2,ETFDH,KARS,SLC1A5,SLC6A6,SPINT2,PGRMC1 ,ELOVL5
10_Member	Hallmark Gene Sets	M5934	HALLMARK XENOBIOTIC METABOLISM	-4,38202	-3,488	13/200	348,355,488,948,1312,1969,2110,3735,6510,6533,10 653,10857,60481	APOE,FAS,ATP2A2,CD36,COMT,EPHA2,ETFDH,KARS,SLC1A5,SLC6A6,SPINT2,PGRMC1 ,ELOVL5
11_Summary	Hallmark Gene Sets	M5937	HALLMARK GLYCOLYSIS	-4,38202	-3,488	13/200	960,1468,2023,2806,3309,5315,5478,7167,9123,1016 5,26355,50814,54982	CD44,SLC25A10,ENO1,GOT2,HSPA5,PKM,PPIA,TPI1,SLC16A3,SLC25A13,FAM162A,NS DHL,CLN6
11_Member	Hallmark Gene Sets	M5937	HALLMARK GLYCOLYSIS	-4,38202	-3,488	13/200	960,1468,2023,2806,3309,5315,5478,7167,9123,1016 5,26355,50814,54982	CD44,SLC25A10,ENO1,GOT2,HSPA5,PKM,PPIA,TPI1,SLC16A3,SLC25A13,FAM162A,NS DHL,CLN6
12_Summary	Hallmark Gene Sets	M5935	HALLMARK FATTY ACID METABOLISM	-4,0604	-3,201	11/158	948,2110,2182,3320,3956,10961,50814,51170,51703 60481,92840,213,5052,54982,347734	CD36,ETFDH,ACSL4,HSP90AA1,LGALS1,ERP29,NSDHL,HSD17B11,ACSL5,ELOVL5,REEP 6,ALB,PRDX1,CLN6,SLC35B2
12_Member	Hallmark Gene Sets	M5935	HALLMARK FATTY ACID METABOLISM	-4,0604	-3,201	11/158	948,2110,2182,3320,3956,10961,50814,51170,51703 60481,92840	CD36,ETFDH,ACSL4,HSP90AA1,LGALS1,ERP29,NSDHL,HSD17B11,ACSL5,ELOVL5,REEP 6
12_Member	Hallmark Gene Sets	M5949	HALLMARK PEROXISOME	-3,39007	-2,647	8/104	213,2182,5052,51170,51703,54982,60481,347734	ALB,ACSL4,PRDX1,HSD17B11,ACSL5,CLN6,ELOVL5,SLC35B2
13_Summary	Hallmark Gene Sets	M5932	HALLMARK INFLAMMATORY RESPONSE	-3,76085	-2,992	12/200	488,490,958,1604,2811,3383,3678,3690,3732,6541,6 578,131566	ATP2A2,ATP2B1,CD40,CD55,GP1BA,ICAM1,ITGA5,ITGB3,CD82,SLC7A1,SCARF1,DCBL D2
13_Member	Hallmark Gene Sets	M5932	HALLMARK INFLAMMATORY RESPONSE	-3,76085	-2,992	12/200	488,490,958,1604,2811,3383,3678,3690,3732,6541,6 578,131566	ATP2A2,ATP2B1,CD40,CD55,GP1BA,ICAM1,ITGA5,ITGB3,CD82,SLC7A1,SCARF1,DCBL D2
14_Summary	Hallmark Gene Sets	M5939	HALLMARK P53 PATHWAY	-3,76085	-2,992	12/200	355,975,1969,2149,3732,6275,6520,7040,9538,23654 ,26355,55240	FAS,CD81,EPHA2,F2R,CD82,S100A4,SLC3A2,TGFB1,EI24,PLXNB2,FAM162A,STEAP3
14_Member	Hallmark Gene Sets	M5939	HALLMARK P53 PATHWAY	-3,76085	-2,992	12/200	355,975,1969,2149,3732,6275,6520,7040,9538,23654 ,26355,55240	FAS,CD81,EPHA2,F2R,CD82,S100A4,SLC3A2,TGFB1,EI24,PLXNB2,FAM162A,STEAP3
15_Summary	Hallmark Gene Sets	M5915	HALLMARK APICAL JUNCTION	-3,76085	-2,992	12/200	947,1004,3383,3384,3673,3688,5175,5788,5819,681(,9019,140885	CD34,CDH6,ICAM1,ICAM2,ITGA2,ITGB1,PECAM1,PTPRC,NECTIN2,STX4,MPZL1,SIRPA
15_Member	Hallmark Gene Sets	M5915	HALLMARK APICAL JUNCTION	-3,76085	-2,992	12/200	947,1004,3383,3384,3673,3688,5175,5788,5819,681(,9019,140885	CD34,CDH6,ICAM1,ICAM2,ITGA2,ITGB1,PECAM1,PTPRC,NECTIN2,STX4,MPZL1,SIRPA
16_Summary	Hallmark Gene Sets	M5902	HALLMARK APOPTOSIS	-3,34406	-2,626	10/161	355,552,952,960,2149,2934,3148,3482,7417,10134	FAS,AVPR1A,CD38,CD44,F2R,GSN,HMGB2,IGF2R,VDAC2,BCAP31
16_Member	Hallmark Gene Sets	M5902	HALLMARK APOPTOSIS	-3,34406	-2,626	10/161	355,552,952,960,2149,2934,3148,3482,7417,10134	FAS,AVPR1A,CD38,CD44,F2R,GSN,HMGB2,IGF2R,VDAC2,BCAP31
17_Summary	Hallmark Gene Sets	M5945	HALLMARK HEME METABOLISM	-3,17598	-2,525	11/200	682,2995,3321,3792,4259,6513,6521,7037,7779,2305 2,129642	BSG,GYPC,IGSF3,KEL,MGST3,SLC2A1,SLC4A1,TFRC,SLC30A1,ENDOD1,MBOAT2
17_Member	Hallmark Gene Sets	M5945	HALLMARK HEME METABOLISM	-3,17598	-2,525	11/200	682,2995,3321,3792,4259,6513,6521,7037,7779,2305 2,129642	BSG,GYPC,IGSF3,KEL,MGST3,SLC2A1,SLC4A1,TFRC,SLC30A1,ENDOD1,MBOAT2
18_Summary	Hallmark Gene Sets	M5906	HALLMARK ESTROGEN RESPONSE EARLY	-3,17598	-2,525	11/200	960,1717,2222,3480,6513,6566,8140,23052,23223,25 800,60481,493,928,2030,4057,5733	CD44,DHCR7,FDFT1,IGF1R,SLC2A1,SLC16A1,SLC7A5,ENDOD1,RRP12,SLC39A6,ELOVL 5,ATP2B4,CD9,SLC29A1,LTF,PTGER3
18_Member	Hallmark Gene Sets	M5906	HALLMARK ESTROGEN RESPONSE EARLY	-3,17598	-2,525	11/200	960,1717,2222,3480,6513,6566,8140,23052,23223,25 800,60481	CD44,DHCR7,FDFT1,IGF1R,SLC2A1,SLC16A1,SLC7A5,ENDOD1,RRP12,SLC39A6,ELOVL 5
18_Member	Hallmark Gene Sets	M5907	HALLMARK ESTROGEN RESPONSE LATE	-3,17598	-2,525	11/200	493,928,960,1717,2030,2222,4057,5733,6566,8140,6 0481	ATP2B4,CD9,CD44,DHCR7,SLC29A1,FDFT1,LTF,PTGER3,SLC16A1,SLC7A5,ELOVL5
19_Summary	Hallmark Gene Sets	M5922	HALLMARK UNFOLDED PROTEIN RESPONSE	-3,14931	-2,519	8/113	811,3309,7184,8140,9114,10525,28972,58477	CALR,HSPA5,HSP90B1,SLC7A5,ATP6V0D1,HYOU1,SPCS1,SRPRB
19_Member	Hallmark Gene Sets	M5922	HALLMARK UNFOLDED PROTEIN RESPONSE	-3,14931	-2,519	8/113	811,3309,7184,8140,9114,10525,28972,58477	CALR,HSPA5,HSP90B1,SLC7A5,ATP6V0D1,HYOU1,SPCS1,SRPRB
20_Summary	Hallmark Gene Sets	M5942	HALLMARK UV RESPONSE DN	-3,08279	-2,471	9/144	490,493,3480,3690,4052,4363,4853,6541,8455	ATP2B1,ATP2B4,IGF1R,ITGB3,LTBP1,ABCC1,NOTCH2,SLC7A1,ATRN
20_Member	Hallmark Gene Sets	M5942	HALLMARK UV RESPONSE DN	-3,08279	-2,471	9/144	490,493,3480,3690,4052,4363,4853,6541,8455	ATP2B1,ATP2B4,IGF1R,ITGB3,LTBP1,ABCC1,NOTCH2,SLC7A1,ATRN

Figure S1

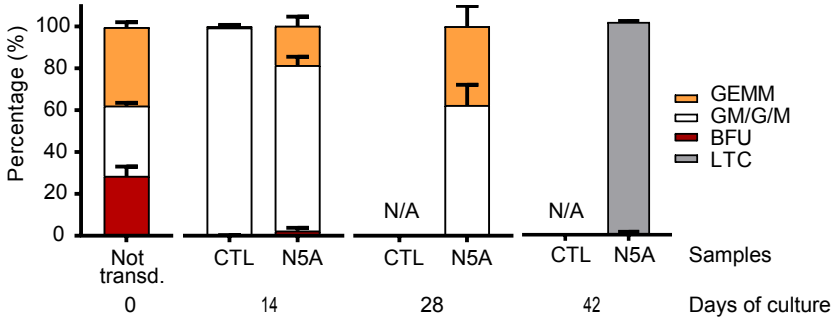


Figure S1. Classification of clonogenic progenitors generated from NUP98-KDM5A cell lines at early time points of development. CFU assays were performed with freshly isolated cord blood $CD34^+$ cells (not transduced cells, day 0, n=2) and control (CTL) or NUP98-KDM5A (N5A)-transduced cell lines on day 14 (n=6/condition), 28 (CTL, n=6; N5A, n=12) and 42 (CTL, n=2; N5A, n=10) of culture. At day 42, white colonies displaying undifferentiated morphology were classified as long-term colonies (LTC). Values are presented as mean percentage ± SEM. N/A, progenitor activity exhausted for CTL cell lines.

Figure S2

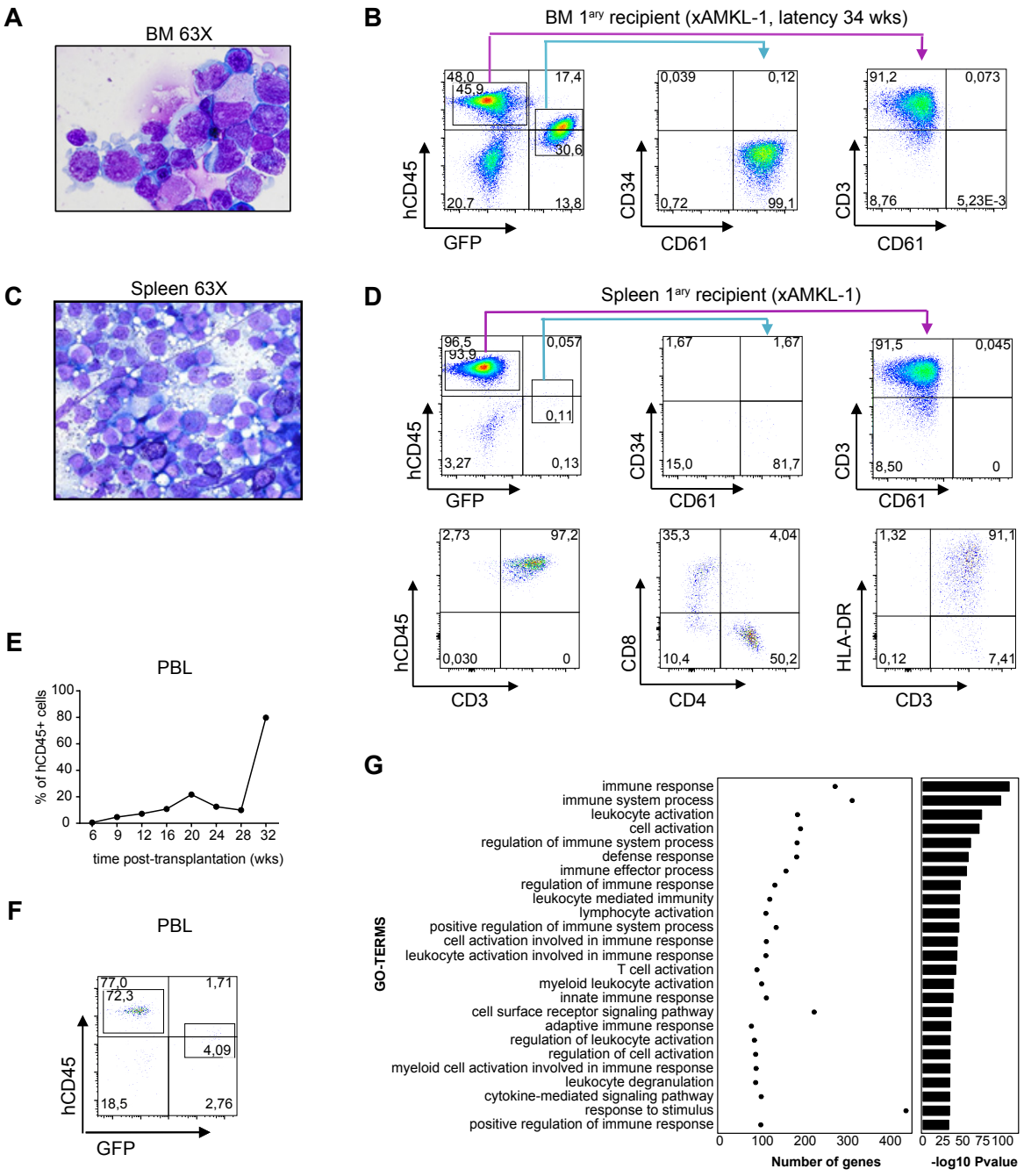


Figure S2. Phenotypic characterisation of NUP98-KDM5A (N5A) acute megakaryoblastic leukemia (AMKL) xenograft model. (**A**) Giemsa-stained cytospin showing bone marrow infiltration with leukemic blasts in a N5A primary xenograft recipient (xAMKL-1) presenting advance signs of acute leukemia at 34 weeks (wks) post-transplantation. (**B**) Flow cytometry analyses showing high infiltration of hCD45^{low}CD34^{-}CD61^{+} cells typical of AMKL, along with CD45^{hi} CD3^{+} activated T-cells, in the bone marrow (BM) of a N5A primary xenograft recipient (xAMKL-1). (**C**) Giemsa-stained cytospin showing low infiltration of megakaryoblasts in the spleen. (**D**) Top panels, flow cytometry analyses revealing low infiltration of hCD45^{low}CD34^{-}CD61^{+} megakaryoblasts in the spleen, along with CD45^{hi} CD3^{+} activated T-cells. Bottom panels, detection of T-cell surface biomarkers by flow cytometry in the hCD45^{+}CD61^{-}CD3^{+} population isolated from the spleen of xAMKL-1 mouse. **(E)** Percentage of human hematopoietic (hCD45^{+}) cells in the peripheral blood (PBL) of xAMKL-1 recipient, from week 6 post-transplantation up to sacrifice. **(F)** FACS profile showing human hematopoietic cells (hCD45^{+}), expressing (or not) GFP, in the PBL of xAMKL-1 mouse. **(G)**. Top 25 of significantly enriched Gene Ontology (GO)-terms for genes overexpressed in RNAseq from CD3^{+} T-cells (n=2) compared to AMKL models are shown. The central dotmap shows the number of CD3-overexpressed genes enriched for each category, while the right panel show the –log10 p-value of association.

Figure S3

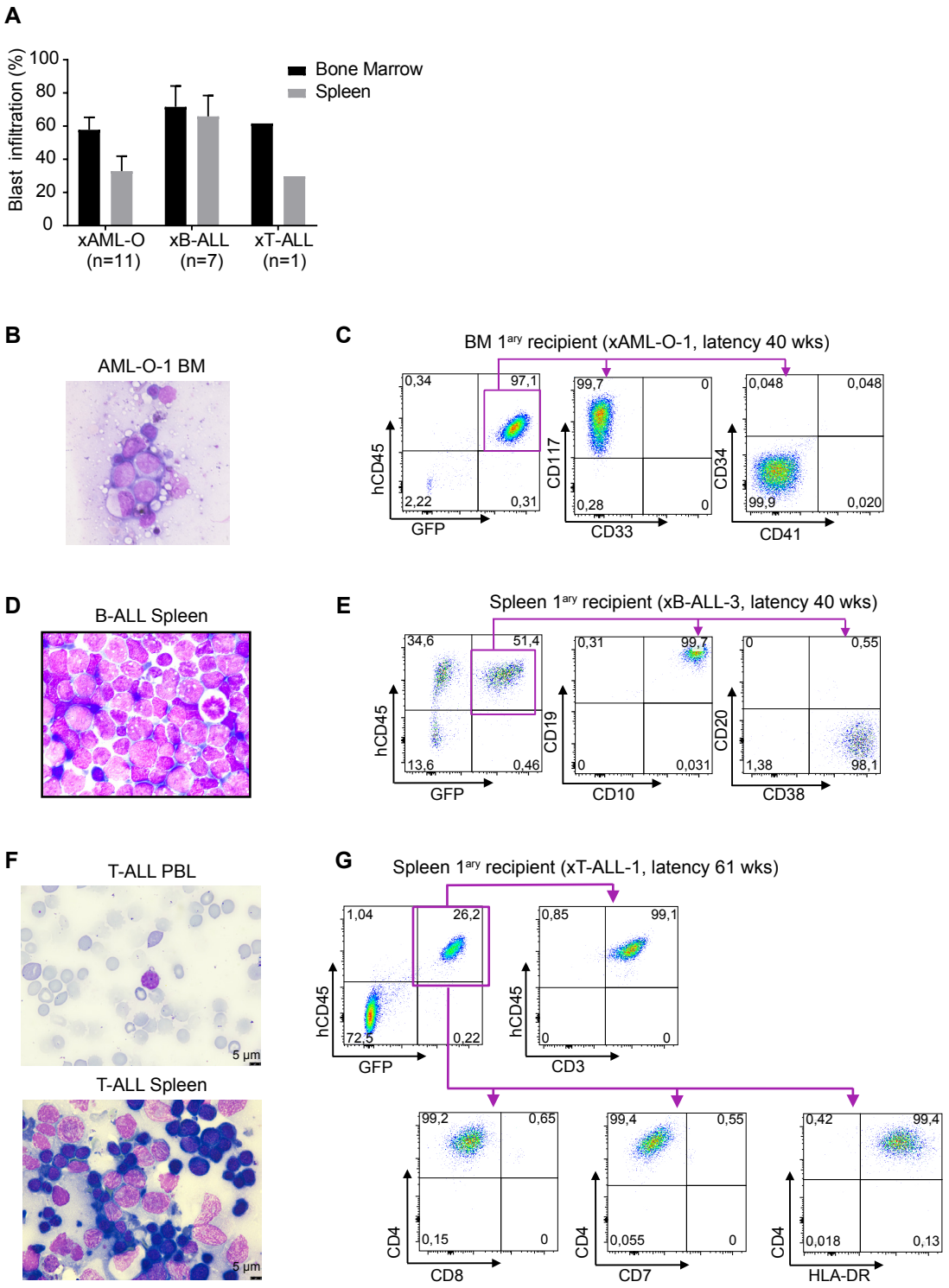
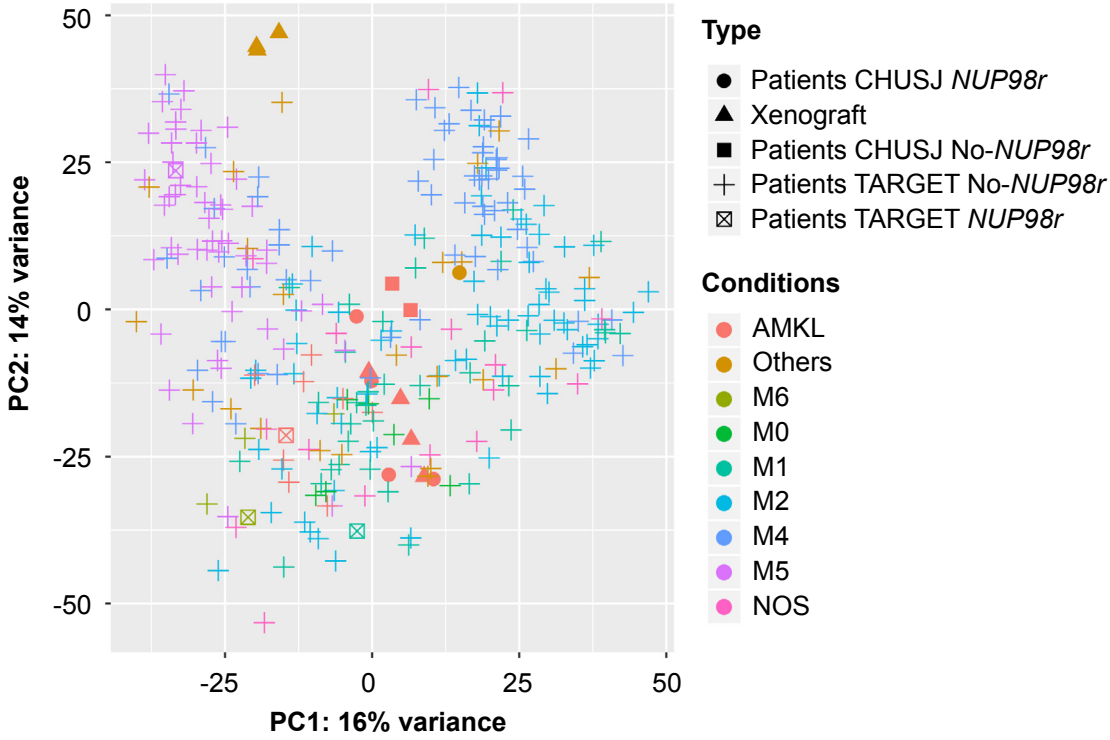


Figure S3. Acute myeloid (AML) and lymphoblastic (ALL) leukemia in NUP98-KDM5A (N5A) xenograft models. (**A**) Average blast infiltration percentage in bone marrow (BM) and spleen from primary xenograft recipients. xAML-O, other acute myeloid leukemia (non-AMKL), B-cell (xB-ALL) and T-cell (xT-ALL) acute lymphoblastic leukemia primary xenograft recipients. (**B**) AML-O in a N5A xenograft recipient characterized by the presence of myeloid blasts on Giemsa-stained BM cytospin preparation **(C)** and expression of CD117 detected by flow cytometry. (**D**) B-ALL in a N5A xenograft recipient characterized by the presence of immature lymphocytes on Giemsa-stained spleen cytospin preparation and (**E**) human GFP$^+$ cells expressing CD19, CD10, and CD38 detected by flow cytometry. Wks, weeks. Characterization of T-cell acute lymphoblastic leukemia xenograft model (xT-ALL-1) based on (**F**) Giemsa-stained spleen touch-prep showing leukemic blasts and on (**G**) detection of T-cell specific cell surface biomarkers by flow cytometry. Additional details for N5A leukemia in primary xenograft recipients are summarized in Table S3.

Figure S4

A



B

Gene_Symbol	xAMKL-1	xAMKL-2	xAMKL-3	xAMKL-5	xAML-O-1	xAML-O-3	xAML-O-2	xB-ALL-1	xB-ALL-3	xB-ALL-2	xB-ALL-4	CB-CD34$^+$-1	CB-CD34$^+$-2	CB-CD34$^+$-3	CB-CD34$^+$-4
PTPRC (CD45)	54	67	61	61	104	121	136	11	40	45	18	157	141	157	88
LYZ	89	1659	645	112	40524	33607	38441	10	11	7	9	55	18	1140	300
CD68	61	109	115	64	283	285	504	12	17	30	8	66	181	6	8
CD33	18	25	44	15	109	95	91	2	6	0	0	22	25	18	43
ITGAM (CD11b)	1	6	8	2	84	73	53	2	5	1	5	13	14	3	8
FUT4 (CD15)	5	6	9	5	37	65	76	7	5	5	7	7	7	6	9
CD36	14	8	3	15	80	104	66	0	0	0	0	15	12	12	7
ANPEP (CD13)	1	1	2	1	1	2	28	2	0	5	1	39	110	21	38
CD14	0	1	1	0	8	6	6	0	0	0	0	6	157	1	1
CD34	1	0	1	0	1	0	0	143	108	173	22	156	91	92	136
MPO	1	94	16	0	569	8	1446	212	5	2	276	52	32	754	298
KIT (CD117)	1	2	14	1	66	26	9	1	0	0	0	32	29	92	35
TFRC (CD71)	60	87	100	83	32	47	87	34	23	20	25	122	105	196	94

Figure S4. Overexpression of NUP98-KDM5A fusion in CB-CD34$^+$ cells induces acute monocytic leukemia and multilineage leukemia subtypes in xenograft models. (A) Principal component analysis showing the expression signatures of acute megakaryoblastic leukemia (AMKL) or AML-others xenografts (xAML-O) clustering with AMKL or acute monocytic leukemia profiles derived from pediatric patients, respectively. Top 500 genes as measured by variance were used to calculate the principal components. **(B)** Heatmap illustrating RNAseq gene expression (FPKM) of selected myeloid markers in leukemic cells derived from NUP98-KDM5A xenograft models compared to normal cord blood cells (CB-CD34$^+$, n=4), suggesting that a subset of xAML-O are distinct and expressing monocyte markers. xAMKL or xB-ALL, acute megakaryoblastic leukemia or B-cell acute lymphoblastic leukemia xenograft, respectively.

Figure S5

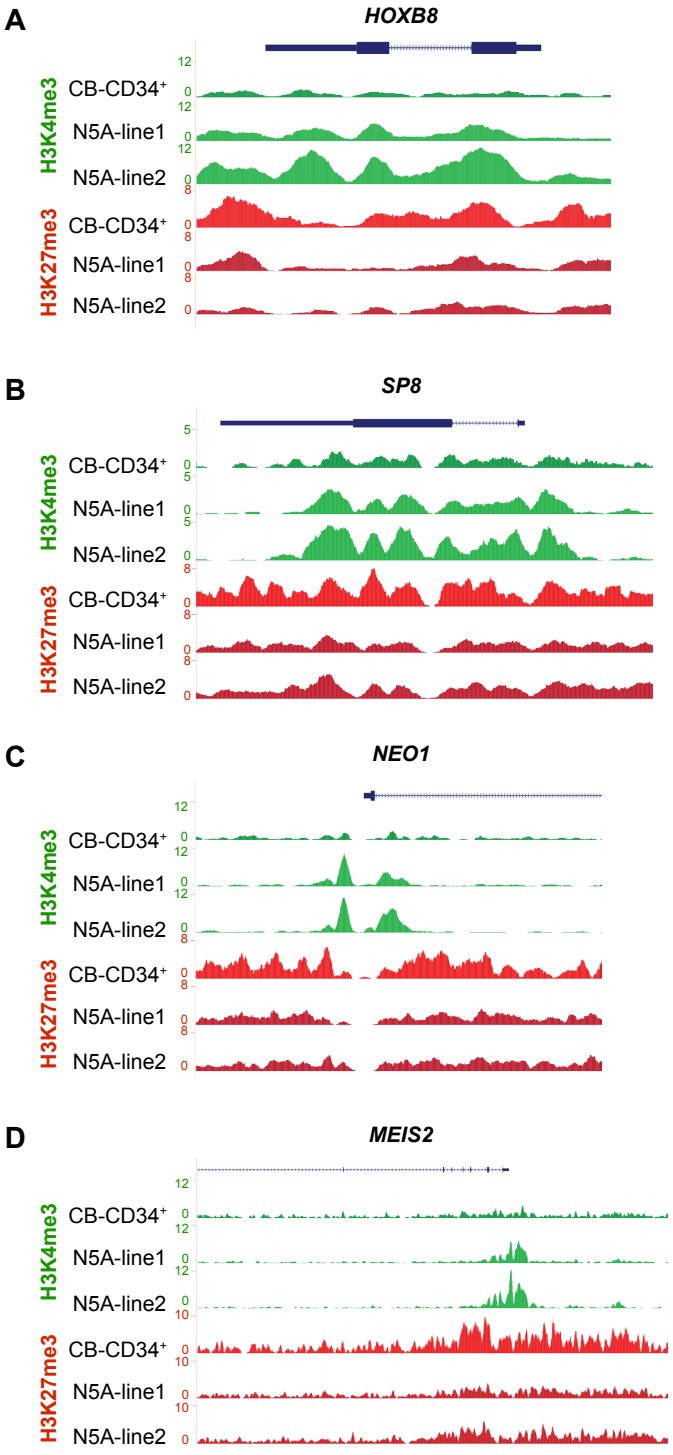
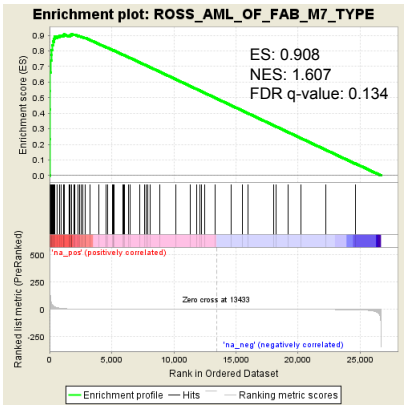


Figure S5. Enrichment of H3K4me3 and H3K27me3 histone modifications along selected loci. (**A-D**) Enrichment of H3K4me3 and H3K27me3 histone modifications along selected loci, as determined using ChIP-seq and chromatin extract from NUP98-KDM5A (N5A) overexpressing cell lines (n=2) and normal cord blood CD34$^+$ (CB-CD34$^+$) cells. The selected genes are upregulated in N5A acute megakaryoblastic leukemia xenograft models compared to CB-CD34$^+$ cells. Images adapted from UCSC genome browser (http://genome.ucsc.edu).

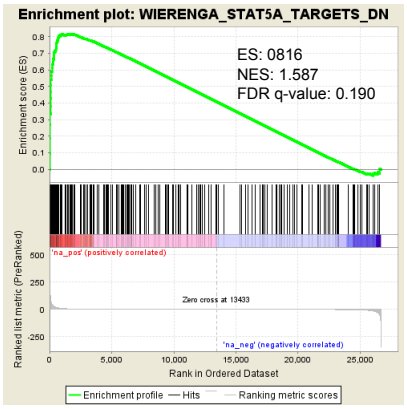
Figure S6

A



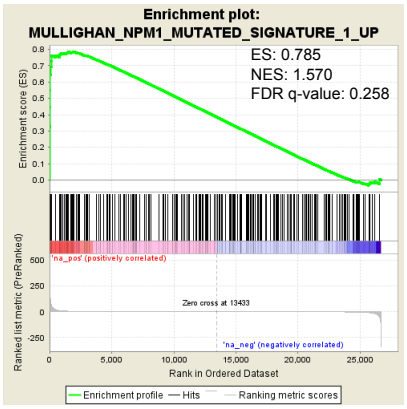
	Probe	Rank in gene list	Rank metric score	Running ES score	Core enrich-ment
1	*NEO1*	9	340,88	0,2318	Yes
2	*MYL4*	15	283,624	0,4247	Yes
3	*DNM3*	30	175,102	0,5434	Yes
4	*RHAG*	34	174,116	0,6619	Yes
5	*ITGA2B*	68	115,996	0,7396	Yes
6	*GJA4*	166	54,29	0,7729	Yes
7	*KEL*	179	51,466	0,8075	Yes
8	*PROS1*	207	45,399	0,8374	Yes
9	*TPM1*	258	38,404	0,8617	Yes
10	*PCDH9*	352	29,04	0,878	Yes
11	*ALDH1A1*	425	23,244	0,8911	Yes
12	*SERPINI1*	618	16,253	0,895	Yes
13	*TAL1*	805	12,015	0,8962	Yes
14	*GATA1*	948	10,143	0,8977	Yes
15	*TEK*	1120	8,557	0,8971	Yes
16	*RYR3*	1129	8,503	0,9026	Yes
17	*PTGS1*	1173	8,006	0,9065	Yes
18	*CD164*	1584	5,57	0,8948	Yes
19	*TFR2*	1625	5,323	0,897	Yes
20	*ARMC8*	1633	5,276	0,9003	Yes
21	*SDPR*	1740	4,891	0,8996	Yes
22	*TNIK*	1746	4,872	0,9028	Yes
23	*DNAJC6*	1774	4,781	0,905	Yes
24	*SLC39A4*	1779	4,776	0,9081	Yes

B



	Probe	Rank in gene list	Rank metric score	Running ES score	Core enrich-ment
1	*EXOC3L4*	19	210,487	0,0984	Yes
2	*SELP*	26	189,63	0,1874	Yes
3	*CD96*	31	174,961	0,2696	Yes
4	*RBPMS2*	49	141,913	0,3358	Yes
5	*SLC22A23*	80	100,028	0,3818	Yes
6	*GP9*	84	97,213	0,4274	Yes
7	*LCN2*	99	86,018	0,4674	Yes
8	*VWF*	103	81,109	0,5055	Yes
9	*GP6*	108	77,634	0,5418	Yes
10	*PDLIM1*	160	55,036	0,5658	Yes
11	*GJA4*	166	54,29	0,5912	Yes
12	*ITGB5*	195	47,614	0,6126	Yes
13	*PROS1*	207	45,399	0,6335	Yes
14	*CAMK1*	267	37,342	0,6489	Yes
15	*SLC9A9*	281	35,608	0,6651	Yes
16	*PBX1*	332	30,656	0,6777	Yes
17	*CAPN11*	338	30,331	0,6918	Yes
18	*NFIB*	363	28,313	0,7042	Yes
19	*PLXDC2*	417	23,741	0,7134	Yes
20	*EPHX2*	502	20,308	0,7198	Yes
21	*CXCR2P1*	512	19,799	0,7287	Yes
22	*MPL*	538	19,018	0,7368	Yes
23	*PLEKHG3*	550	18,483	0,745	Yes
24	*SCPEP1*	565	17,99	0,753	Yes
25	*DENND2C*	568	17,971	0,7614	Yes
26	*PDIA5*	588	17,408	0,7688	Yes
27	*BAMBI*	605	16,793	0,7761	Yes
28	*NINJ2*	650	15,274	0,7817	Yes
29	*CLCN4*	684	14,482	0,7872	Yes
30	*C3orf52*	700	14,069	0,7933	Yes
31	*C15orf52*	712	13,904	0,7994	Yes
32	*TSPAN9*	728	13,4	0,8052	Yes
33	*CD84*	870	11,097	0,8051	Yes
34	*LGALS12*	871	11,066	0,8103	Yes
35	*ZNF185*	942	10,192	0,8124	Yes
36	*ALOX12*	981	9,864	0,8156	Yes
37	*IL7*	1251	7,375	0,809	Yes
38	*FCER1A*	1294	7,041	0,8107	Yes
39	*SLC35D3*	1345	6,718	0,812	Yes
40	*ARHGEF3*	1421	6,339	0,8121	Yes
41	*NBEAL2*	1424	6,325	0,815	Yes
42	*CD9*	1501	6,006	0,815	Yes
43	*SLC37A1*	1621	5,349	0,813	Yes
44	*SH3BGRL2*	1639	5,254	0,8148	Yes
45	*RGS18*	1696	5,069	0,8151	Yes
46	*SDPR*	1740	4,891	0,8158	Yes

C



	Probe	Rank in gene list	Rank metric score	Running ES score	Core enrich-ment
1	*HOXB9*	1	518,149	0,3266	Yes
2	*HOXB6*	25	190,296	0,4458	Yes
3	*PTCH1*	35	169,431	0,5522	Yes
4	*HOXB5*	43	149,299	0,6461	Yes
5	*GP6*	108	77,634	0,6926	Yes
6	*PDE3A*	146	58,941	0,7284	Yes
7	*SEPP1*	157	56,216	0,7635	Yes
8	*RTN2*	498	20,392	0,7634	Yes
9	*HIST1H3H*	791	12,2	0,7601	Yes
10	*HOXA3*	819	11,863	0,7666	Yes
11	*CD84*	870	11,097	0,7717	Yes
12	*HOXB3*	904	10,64	0,7771	Yes
13	*HOXA10*	1044	9,14	0,7776	Yes
14	*HOXA9*	1217	7,635	0,7759	Yes
15	*BMPR1A*	1333	6,791	0,7759	Yes
16	*HOXB2*	1379	6,574	0,7783	Yes
17	*RAB38*	1450	6,196	0,7796	Yes
18	*HOXA6*	1457	6,18	0,7832	Yes
19	*MCHR1*	1521	5,901	0,7846	Yes
20	*MAP3K5*	1719	4,959	0,7802	Yes
21	*SDPR*	1740	4,891	0,7826	Yes
22	*TXNIP*	1824	4,601	0,7823	Yes
23	*HOXA5*	1866	4,447	0,7836	Yes
24	*CHRM5*	1906	4,317	0,7848	Yes
25	*CCL23*	1965	4,112	0,7852	Yes

Figure S6. Gene set enrichment analysis (GSEA) of NUP98-KDM5A acute megakaryoblastic leukemia (AMKL) expression signature. (**A-C**) Enrichment plots for selected gene sets showing significant correlation with genes upregulated in NUP98-KDM5A AMKL patients and xenograft models compared to cord blood CD34$^+$ cells. Genes in the leading edge subsets are listed below enrich-ment plots.

Figure S7

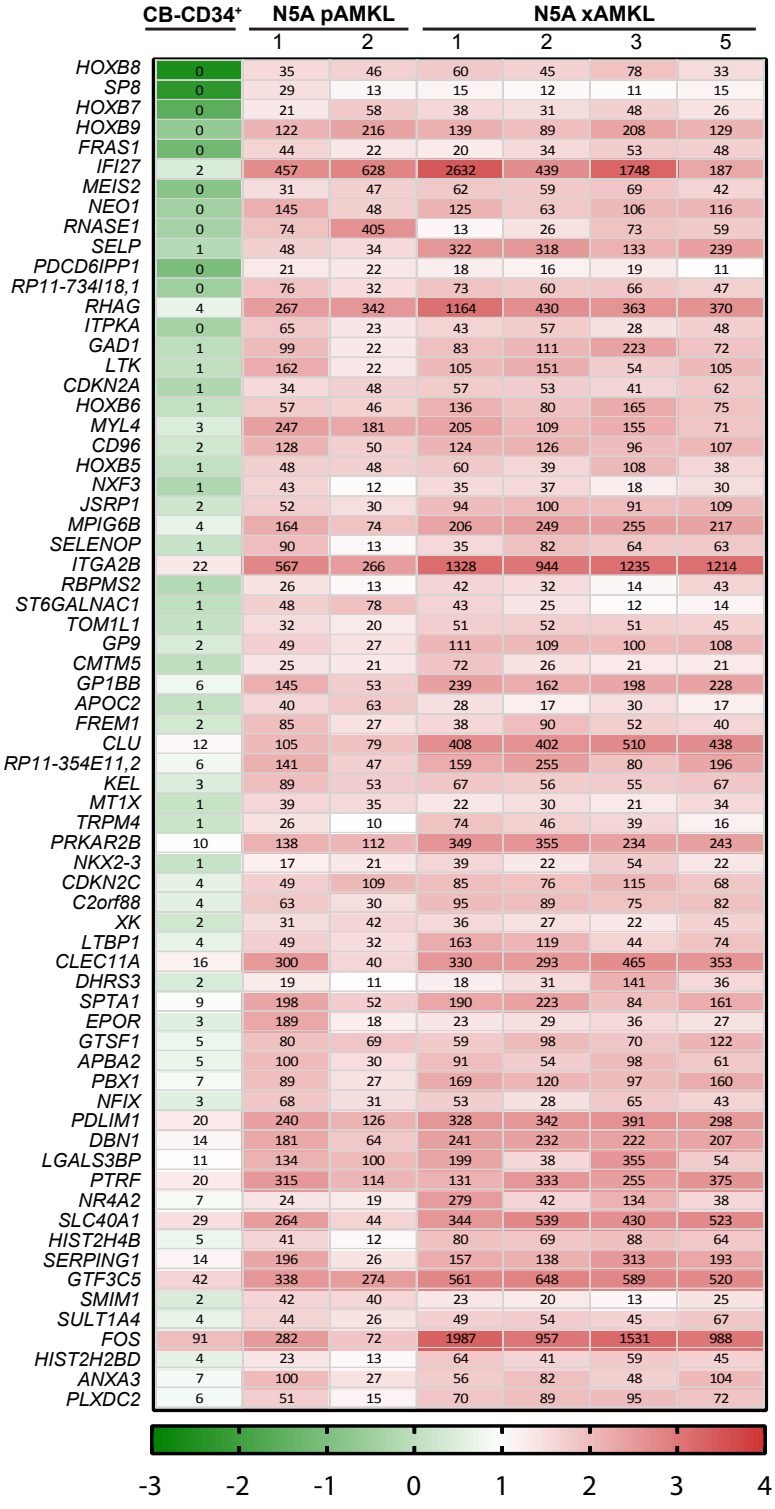


Figure S7. Genes differentially expressed by at least tenfold in a sampling of bone marrow cells derived from 2 patients and 4 xenograft models presenting NUP98-KDM5A (N5A) AMKL, as compared to normal cord blood CD34⁺ cells. Genes with expression values of 0 FPKM for all samples were discarded. Genes with expression values of ≥10 FPKM for all N5A AMKL samples and fold changes ≥10 compared to CB-CD34⁺ samples are displayed. CB-CD34⁺, cord blood CD34⁺ cells. FPKM values are represented by a logarithmic color scale (Log_{10}).