Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/5/7/eaaw4304/DC1

Supplementary Materials for

Mutant CEBPA directly drives the expression of the targetable tumor-promoting factor CD73 in AML

Janus S. Jakobsen, Linea G. Laursen, Mikkel B. Schuster, Sachin Pundhir, Erwin Schoof, Ying Ge, Teresa d'Altri, Kristoffer Vitting-Seerup, Nicolas Rapin, Coline Gentil, Johan Jendholm, Kim Theilgaard-Mönch, Kristian Reckzeh, Lars Bullinger, Konstanze Döhner, Peter Hokland, Jude Fitzgibbon, Bo T. Porse*

*Corresponding author. Email: bo.porse@finsenlab.dk

Published 10 July 2019, *Sci. Adv.* **5**, eaaw4304 (2019)
DOI: 10.1126/sciadv.aaw4304

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/5/7/eaaw4304/DC1)

Data file S7 (Microsoft Excel format). Materials lists, related to Materials and Methods.

Supplementary Materials and Methods

Bone Marrow Transplantations

For competitive BMT, cells were counted, GFP/YFP expression analyzed by LSR-II (BD), and input samples recorded after mixing cells 1:1, before transplantation of $2x10^5$ GFP/YFP+ cells.

For survival assays sublethally irradiated recipients were transplanted with 10,000 transduced and sorted GFP^+ Lp30 cells along with $5x10^5$ cells/mouse irradiated (2000 cGy) bone marrow cells as support.

For survival assays with CRISPRi/lentivirus transduced cells, cells were sorted for mCherry expression on AriaIII (BD) and $1x10^5$ cells transplanted. The expanded pool of cells was harvested from the BM when sufficient mCherry+ blast cells could be detected in the PB by flow cytometry (five weeks), and subjected directly to a second round of transduction with sgRNA-GFP, and sorted for double-positive cells on AriaIII (BD) after seven days, and $5x10^3$ cells per mouse were used for BMT along with $5x10^5$ carrier cells (see above).

For any survival assay where mice did not succumb from leukemia after 2x median survival time had passed, PB cells were analyzed by flow cytometry, and for all mice analyzed, no contribution of transduced cells to the PB could be detected, and mice were euthanized.

For gene expression analysis, cells were sorted two days post shRNA/retrovirus transduction for GFP expression, and $3x10^5$ cells for knockdown constructs and $1x10^5$ for scrambled were used for BMT of recipient mice. BM cells were harvested after three weeks of leukemia expansion for further analysis

For the treatment experiment, a fresh Lp30 leukemia was established, and BM cells were harvested four weeks post BMT, counted and serially diluted in irradiated carrier cells to 100 cells transplanted per mouse. Mice were treated intra peritoneal (i.p.) with targeting antibody/control + inhibitor/mock. CD73 antibody 2C5 or control Ig NIP228 (both mIgG2a) (Medimmune) were used at 100 µg/mouse and A2AR inhibitor, SCH58261 (Sigma) at 20 µg/mouse, PBS as mock.

Sorting of murine myeloid progenitors

Flow cytometry-based cell sorting of GM progenitor cells was done as previously described (*52*). Briefly, BM cells were collected from tibia, femur and ilium and enriched using CD117 magnetic microbeads (Miltenyi Biotech). Cells were subsequently stained using the following conjugated antibodies (all from eBioscience unless otherwise stated): CD150-APC (clone TCF15-12F12.2; Biolegend), CD41-FITC (clone MWReg30), CD105-PE (clone Mj7/18), Sca1-PerCPCy.5 (clone D7), FcgRII/III-A700 (clone 93), cKit-A780 (clone 2B8), Ter119- PECy7 (clone Ter119, BioLegend) and a lineage cocktail consisting of Gr1- PECy5 (clone RB6-8C5) B220-PECy5 (clone RA3-6B2), CD3e-PECy5 (clone 145-2C1) and Mac1-PECy5 (clone M1/70, BioLegend). CD45.1-eF450 (clone A20), CD45.2-biotin (clone 104), SA-QD655 (BD). Donor derived cells (CD45.2+) of the following previously defined immunophenotypes (*10, 53*) were sorted on a FACSAriaI or –III: GMPs: Live cells (7-AAD-, Invitrogen) Lin^- $cKit^+$ $Sca\text{-}1^-$ $CD41^-$ $FcgRII/II^+$. Populations relevant for this study were sorted as in Fig. S1A.

Sorting of murine transduced cells

Retrovirally transduced cells were analyzed or sorted for GFP/YFP expression two days post transduction, with dead cells excluded by 7-AAD staining (1 µg/mL) on LSR-II/AriaI/AriaIII (BD). Lentivirally transduced cells were sorted for GFP/mCherry seven days after transduction, using DAPI (0.5 µg/mL) to discriminate dead cells on AriaIII (BD).

Assessment of CD73 expression on murine GMPs

BM from WT/Lp30 transplanted mice were stained with antibodies and analyzed on an LSR-II (BD). For analyzing hematopoietic progenitors CD41-FITC, Ter119/CD3e/Gr1-PeCy5, Sca-1-APC, CD73-PECy7, FcgRII/III-Alexa Flour700, c-kit-APC780, CD45.1-eF450, CD45.2-biotin (eBioscience), B220-PeCy5, CD105-PE, SA-QD655 (BD) and Mac1-PeCy5, CD150-BV650 (BioLegend) were used. Gating strategy as in S1A. Data was analyzed using the FlowJo software.

Sorting of human myeloid progenitors

Sorting was carried out as previously described (*54*). Briefly, wBM cells were enriched for CD34 expression and subsequently stained using the following conjugated antibodies (all from eBioscience unless otherwise stated): A lineage cocktail consisting of CD3-bio (clone UCHT1), CD7-bio (clone eBio124-1D1), CD14-bio (clone 61D3), CD19-bio (clone HIB19), CD56-bio (Biolegend, clone HCD56), CD61-bio (Serotec, clone PM13/3), CD235a-bio (clone HIR2), CD15-bio (clone HI98) and SA-PECy5 (Biolegend), plus the following unique conjugates: CD38-FITC (Biolegend, clone HIT2), CD123-PE (clone 6H6), CD34-PECy5.5 (Biolegend, clone 581), CD11b-PECy7 (clone M1/70), CD90-APC (clone eBio5E10), CD33-Alexa700 (clone WM53), CD45RA-eF605 (clone HI100). Populations relevant for this study were sorted as depicted in Fig. S3A.

Identification and quantification of $CD73^+$ human GMPs

Human BM samples were analyzed on an LSR-II (BD), and cells were stained with the following antibodies: CD38-FITC (Biolegend, clone HIT2), CD19-PE (BD Biosciences, Clone HIB19), CD34-PE-CF594 (BD Biosciences, clone 581), CD123-PE-eF650NC (Clone 6H6), CD3-APC (BD Biosciences, Clone UCHT1), CD33-Alexa700 (Clone WM53), CD11b-PE-Cy7 (BD Biosciences, clone ICRF44), CD45RA-BV605 (Biolegend, clone HI100), CD73-BV421 (BD Biosciences, clone AD2). Gating strategy is indicated in Supplementary Fig. S5A

A full list of antibodies used can be found in Data File S7

Mass spectrometry

FACS sorted cells were washed twice with ice-cold PBS. 100,000 cells for each experimental condition were subjected to sample preparation according to (*55*). Cells were lysed using 20 µl of lysis buffer (consisting of 6 M Guanidinium Hydrochloride, 10 mM TCEP, 40 mM CAA, 100 mM Tris pH8.5). Samples were boiled at 95^{o}C for 5 minutes, after which they were sonicated on high for 3x 10 seconds in a Bioruptor sonication water bath (Diagenode) at 4^{o}C. Samples were diluted 1:3 with 10% Acetonitrile, 25 mM Tris pH 8.5, LysC (MS grade, Wako) was added in a 1:50 (enzyme to protein) ratio, and samples were incubated at 37^{o}C for 4hrs. Samples were further diluted to 1:10 with 10%

Acetonitrile, 25 mM Tris pH 8.5, trypsin (MS grade, Promega) was added in a 1:100 (enzyme to protein) ratio and samples were incubated overnight at 37^{o}C. Enzyme activity was quenched by adding 2% trifluoroacetic acid (TFA) to a final concentration of 1%. Prior to mass spectrometry analysis, the peptides were desalted on in-house packed C18 Stagetips (*56*). For each sample, two discs of C18 material (3M Empore) were packed in a 200 µl tip, and the C18 material activated with 40 µl of 100% Methanol (HPLC grade, Sigma), then 40 µl of 80% Acetonitrile, 0.1% formic acid. The tips were subsequently equilibrated 2x with 40ul of 1%TFA, 3% Acetonitrile, after which the samples were loaded using centrifugation at 4,000x rpm. After washing the tips twice with 100 µl of 0.1% formic acid, the peptides were eluted into clean 500 µl Eppendorf tubes using 80% Acetonitrile, 0.1% formic acid. The eluted peptides were frozen on dry ice and concentrated in an Eppendorf Speedvac, and re-constituted in 1% TFA, 2% Acetonitrile for Mass Spectrometry (MS) analysis.

For each sample, peptides were loaded onto a 2cm C18 trap column (ThermoFisher 164705), connected in-line to a 50cm C18 reverse-phase analytical column (Thermo EasySpray ES803) using 100% Buffer A (0.1% Formic acid in water) at 750bar, using the Thermo EasyLC 1000 HPLC system in a single-column setup and the column oven operating at 45°C. Peptides were eluted over a 200 minute gradient ranging from 5 to 48% of 100% acetonitrile, 0.1% formic acid at 250 nl/min, and the Orbitrap Fusion (Thermo Fisher Scientific) was run in a three second MS-OT, ddMS2-IT-HCD top speed method. Full MS spectra were collected at a resolution of 120,000, with an AGC target of 4×10^5 or maximum injection time of 50ms and a scan range of 400–1500m/z. Ions were isolated in a 1.6m/z window, with an AGC target of 1×10^4 or maximum injection time of 35ms, fragmented with a normalized collision energy of 30 and the resulting MS2 spectra were obtained in the ion trap. Dynamic exclusion was set to 60 seconds, and ions with a charge state <2, >7 or unknown were excluded. MS performance was verified for consistency by running complex cell lysate quality control standards, and chromatography was monitored to check for reproducibility.

Label-free quantitative proteomics analysis:

The raw files were analyzed using MaxQuant version 1.5.2.8 (Cox and Mann, 2008) and standard settings. Briefly, label-free quantitation (LFQ) was enabled with a requirement of 2 unique peptides per protein, and iBAQ quantitation was also enabled during the search. Variable modifications were set as

Oxidation (M), Acetyl (protein N-term), Gln->pyro-Glu and Glu->pyro-Glu. Fixed modifications were set as Carbamidomethyl (C), false discovery rate was set to 1% and "match between runs" was enabled. The resulting protein groups file was processed with an in-house developed tool (PINT) (*57*), which imputes missing LFQ values with adjusted iBAQ values. Briefly, the distributions of iBAQ intensities for each sample are adjusted to overlap with the LFQ intensity distributions using median-based adjustment, enabling the direct imputation of missing LFQ values with adjusted iBAQ values for those proteins that did not have LFQ values across all the samples. Simultaneous filtering for reverse hits, contaminants and only those proteins observed in two biological replicates (N=2), this resulted in a final list of a total 6,997 samples identified across all samples.

Mapping and quantification of RNA-seq data (murine and human AML):

The raw sequencing data were trimmed with Trimmomatic (v 0.32) with the parameters: ILLUMINACLIP:fastqc.fa:2:30:10 HEADCROP:X LEADING:22 SLIDINGWINDOW:4:22 MINLEN:25 where fastqc.fa is a fasta file containing all the adapters that are distributed with FastQC v0.11.2 and where X is an individual number (human samples: median=10, range=8-10, mouse samples: median=1, range=1-5) to remove the reverse transcriptase (RT) bias (*58*) in the individual libraries.

The trimmed data were mapped to the genome (hg19, mm9) using TopHat2 (v2.0.9) using the parameters --b2-very-sensitive --library-type=fr-unstranded and supplying the Ensemble GTF file to the --GTF argument (release 66 for both human and mouse). The mapped reads were summarized into gene counts using the featureCounts() function from the Rsubread R package (V1.16.1) (*59*) using the same GTF file described above and with default parameters apart from isPairedEnd=TRUE.

Filtering and differential expression:

Genes with at least 2 fragments uniquely mapping in at least 1 library we tested for differential expression. A gene was deemed differentially expressed if it was called as significant with a False Discovery Rate (FDR) corrected p-value < 0.01 by both edgeR and DESeq. Specifically the exactTest() function from EdgeR (v 3.12.0) (*60*) was used after TMM normalization and estimation of dispersion

using both estimateCommonDisp() and estimateTagwiseDisp(). The DESeq() function from DESeq (v 1.18.0) (*61*) was used with default parameters. For illustration of individual gene expression, edgeR's FPKM function was used and the output displayed as log2 values (as shown in Fig. 3D).

Analysis of orthologous differential expression:

Orthologous genes were extracted with the biomaRt R package (*62*) using Ensembl gene_ids and a minimum homology of 50% was required by specifying "mmusculus_homolog_perc_id_r1 > 50". Differentially expressed homologous genes were extracted by overlapping the orthologous ensemble genes ids with the differentially expressed ensemble gene ids defined above. Enrichment for overlap in differentially expressed genes were calculated using the Fisher´s test.

Mapping and quantification of RNA-seq data (Nt5e knockdown):

Raw fastq files were mapped onto the mouse genome (mm10) using STAR (*63*). Transcript level quantification was performed with Isolator (*64*). Differential expression analysis was performed with Sleuth (*65*), after processing of the raw fastq files with kallisto (*66*).

GO analysis RNA-seq on Nt5e knockdown cells:

For *Nt5e* knock-down samples, we created gene sets using the 1% most up- and downregulated genes, compared to the *shScr* control. Next, using a hypergeometric test, we calculated the significance of the overlap between gene sets of up- and downregulated genes against various genes signatures extracted from the MSigDB molecular signature database (www.broadinstitute.org/gsea/msigdb/), or compiled by us manually.

Adenosine receptor signature:

A list of genes shown to be repressed by active signaling via the adenosine A2 receptor (ADORA2A) in leukocytes was assembled by a comprehensive literature survey (full list of references used can be provided by request). Genes included in the signature are listed in Data File S6

GSEA analysis of A2A signature and core transcriptional program:

For TCGA AML data analysis, RNA-Seq data for AML patients of all cytogenetic risk groups were downloaded from the TCGA AML data portal (*2*). Raw reads where mapped to the human genome (build GRCh37) with Kallisto, and isoforms expression values one gene value. Subsequently, Gene set enrichment analysis (GSEA; http://www.broadinstitute.org/gsea/index.jsp) was performed on five biallelic *CEBPA* mutant AML vs all other AMLs in the dataset, using default parameters, and permutations by gene sets were conducted to assess statistical significance. Core CEBPA-mutant program genes were filtered for presence in the data set (upregulated found: 20, downregulated found: 71), for the A2A signature, 30 genes were found.

ChIP-seq data analysis

Mapping of sequences to the mouse genome (mm9) was performed as previously described (*6*). Briefly, we used Bowtie 2 (default parameters) for the mapping and determined genomic regions enriched for CEBPA TFs (WT and p30) using Macs2. To increase the specificity of the enriched regions, we used a control sample prepared using a control anti-body ('IgG' control). Finally, Irreproducible Discovery Rate (IDR) was applied to filter out irreproducible regions between the two replicates.

In total, 30951 unique set of CEBPA enriched regions were identified out of 22525 and 25196 regions identified in WT and p30, respectively using BEDTools (multiIntersectBed) (*67*). Out of 30951, 2430 and 1385 regions were specific to WT and p30, respectively, while the remaining 27136 regions were common (defined based on CEBPA fold change (log2) of two between WT and p30).

Normalized (TPM) H3K4me1, H3K4me3 and H3K27ac signals were computed at 30,951 CEBPA-enriched regions by taking a window of 500 bp up- and down-stream to the center. ngs.plot (*68*) was used to plot histone enrichment relative to the center of CEBPA-enriched regions as aggregation plots and heat maps. We associated each CEBPA-enriched region to its closest UCSC gene annotation (mm9 genome assembly) as its potential regulatory target.

HLF bound regions in multipotent progenitor LSK cells ($Lin^{-}Sca\text{-}1^{+}c\text{-}Kit^{+}$) (N=3153) were retrieved from GEO, GSM1709416 (*69*), whereas regions bound by ERG, PU.1 and FLI1 in HPC-7 cells were retrieved from GSE22178 (*19*). Significance levels at which the three classes of CEBPA-enriched regions (WT, p30 and common) overlap with HLF bound regions were computed using Fisher's exact test as described in (*6*).

Gene ontology analysis and motif enrichment analysis was carried out as previously described (*6*).

External Human HiC ChIP-seq and RNA-seq data

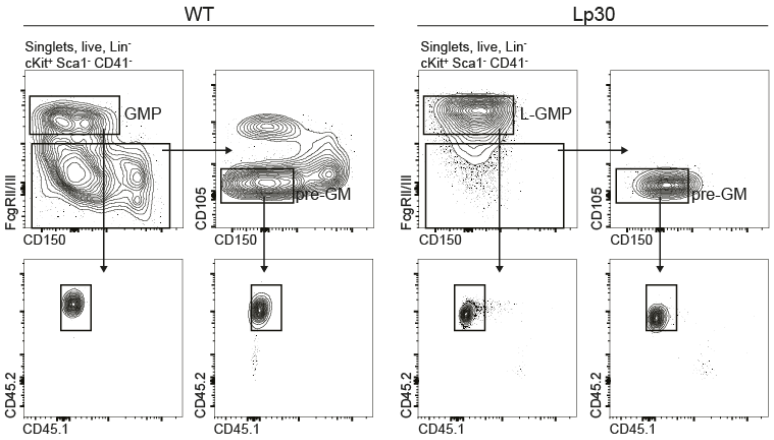
Human HiC data from CD8+ Tcells where downloaded from EGA European Genome Archive (EGAS00001001911, (*30*)). Position in the proximity of the *NT5E* locus with an interaction score above 5 and a *NT5E* TSS anchor was identified and repositioned to Hg19 (*Homo Sapiens*, GRCh37) using UCSC tools (*70*). Fastq files for CD34+ HSPC H3K4me1 data (GSM486707, (*29*) and CD8+ T-cell H3K2ac and H3K4me1 data (EGAD00001000936, (*31*)) were acquired from the GEO and BLUEPRINT repositories, respectively, and mapped to Hg19 using Bowtie v. 0.12.8 with standard settings for unique mapping (*71*). HiC positions and external ChIP-seq tracks were visualized in the UCSC browser with our human H3K27ac ChIP-seq data.

CEBPA and *NT5E* RNA-seq data were acquired for the TCGA AML data set at cBioportal (*72*) and five biallelic CEBPA mutant AML patients manually identified. Array expression data were downloaded from Leukemia Gene Atlas (*73*) for the Kohlmann (GSE 15434, (*27*)) and Verhaak (*26*) datasets, for both of which bi- and monoalleic CEBPA mutant AML were not separated.

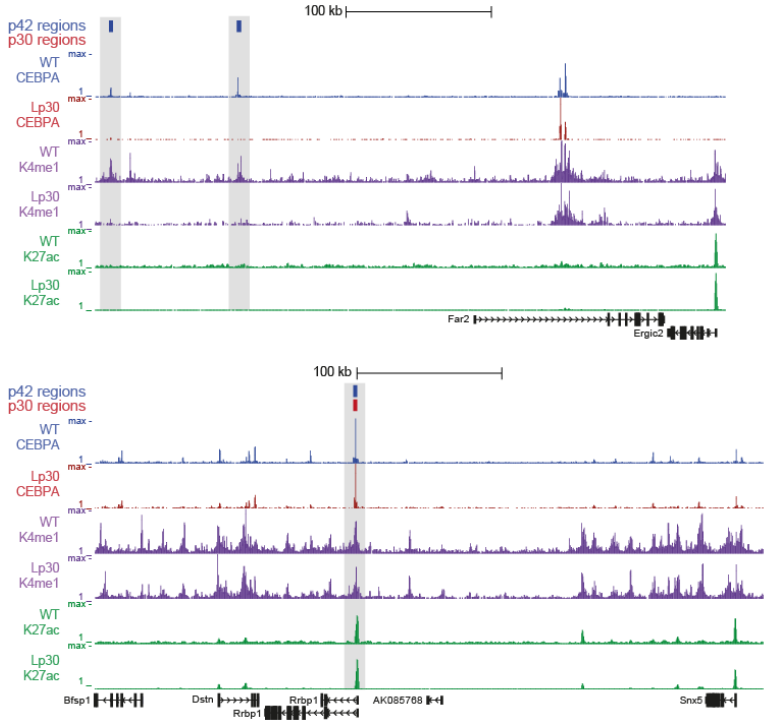
INV(16) ChIP-seq analysis from (*28*) was visualized in the UCSC browser using the provided track-hubs. INV(16) fusion products CBF A1329 and MYH Novus from cell lines as well as CBF A1329 and MYH Novus Patient tracks were displayed for the *Nt5e* TSS locus in hg18.

Supplementary Figures

S1A



S1B



S1C

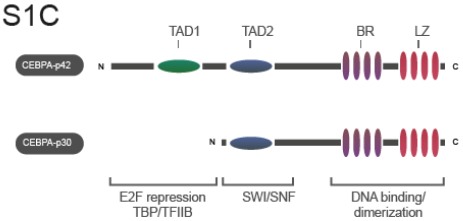


Fig. S1. Distinct enhancer-binding of CEBPA in WT GMPs versus p30 L-GMPs, related to Fig. 1. (S1A) Sorting strategy used for isolation of murine myeloid progenitors. **(S1B)** Representative examples of ChIP-seq tracks showing overlap of H3K4me1 and H3K27Ac with p42-specific (top) and common (bottom) CEBPA binding in WT GMPs and Lp30 L-GMPs, respectively. **(S1C)** Schematic representation of the domain structure of the p30 and p42 CEBPA isoforms.

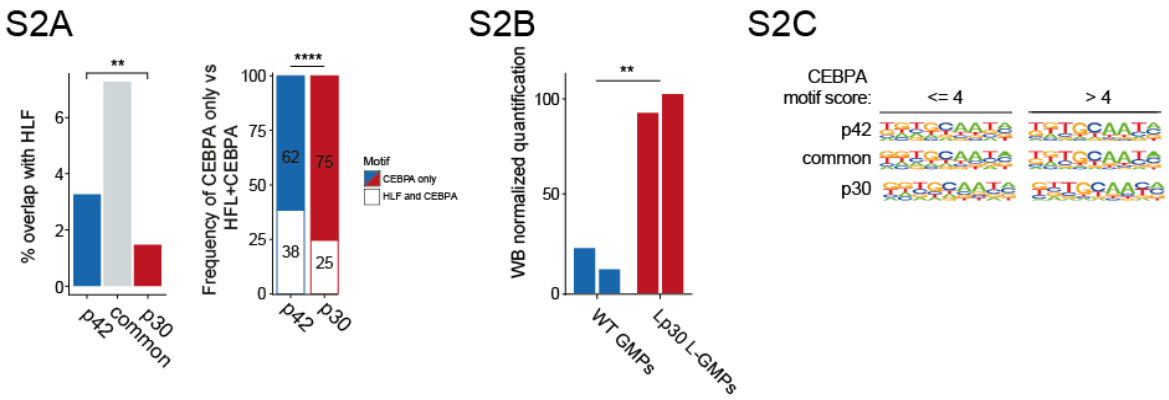


Fig. S2. Motif enrichment analysis of p42, common and p30 CEBPA-bound regions, related to Fig. 2. (S2A) Overlap of p42, common and p30 regions sets with HLF binding in hematopoietic progenitor cells. All regions set to 1 kb. Difference between p42 and p30 region overlap tested (Fisher's exact test, **, P≤0.01). **(S2B)** Western band intensity quantification (relating to Fig. 2F) (n=2, Student's t-test. **, P≤0.01). **(S2C)** Compiled motif logos of CEBPA motifs with low (</= 4) or high (> 4) match score for p42, common and p30 region sets (relating to Fig. 2H).

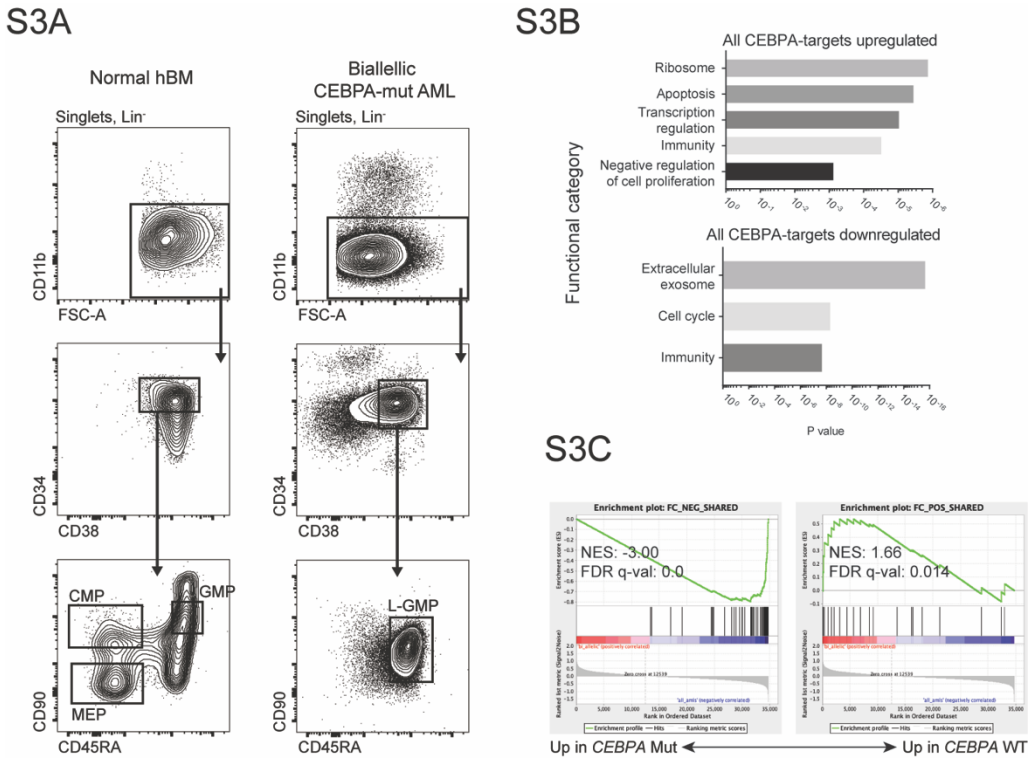
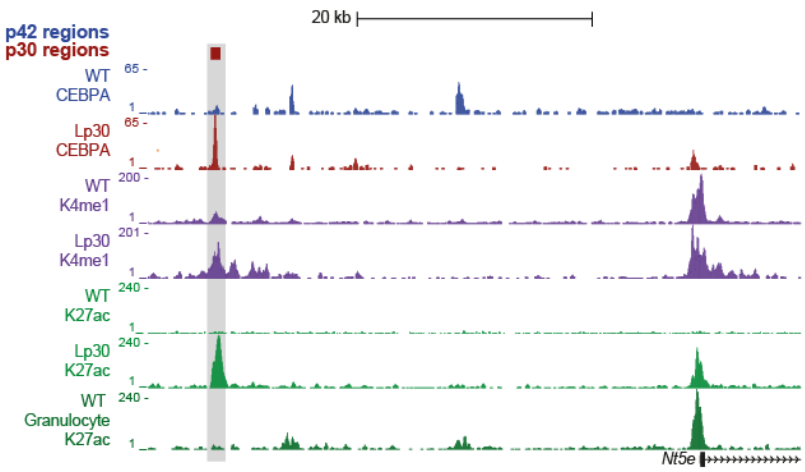


Fig. S3. Gene expression analysis of murine and human *CEBPA* mutant AML, related to Fig. 3. (S3A) Sorting strategy used for isolation of human myeloid progenitors**. (S3B)** GO-term analysis (DAVID) of up- and down-regulated genes in murine p30 AML compared to WT GMPs, which are associated with (nearest neighbor) CEBPA-bound enhancers. **(S3C)** Plot and statistical values from GSEA analysis testing genes of the core AML program (82 downregulated vs. wt/healthy left hand panel; 20 upregulated vs. wt/healthy, right hand panel) in biallelic CEBPA mutated AML versus all other AML subtypes, TCGA AML expression data set.

S4A



S4B

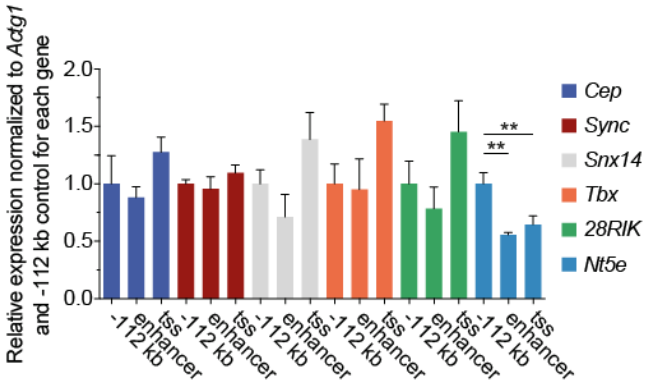


Fig. S4. CEBPA-p30 directly activates the expression of *Nt5e*, related to Fig. 4. (S4A) *Nt5e* locus ChIP-seq tracks of CEBPA binding, H3K4me1 and H3K27Ac in WT GMPs and Lp30 L-GMPs as well as analysis of H3K27ac levels in chromatin from sorted granulocytes (WT). **(S4B)** Relative expression levels of genes close to *Nt5e* in Lp30 cells after transduction with dCas9-KRAB and the indicated gRNAs. Normalized to *Actg1* and -112 kb control for each gene. Mean +/- SD, n=3. Statistics were determined with one-way ANOVA corrected for multiple comparisons (Dunnett's) between -112 kb control gRNA and enhancer or TSS targeting gRNAs. **, P≤0.01.

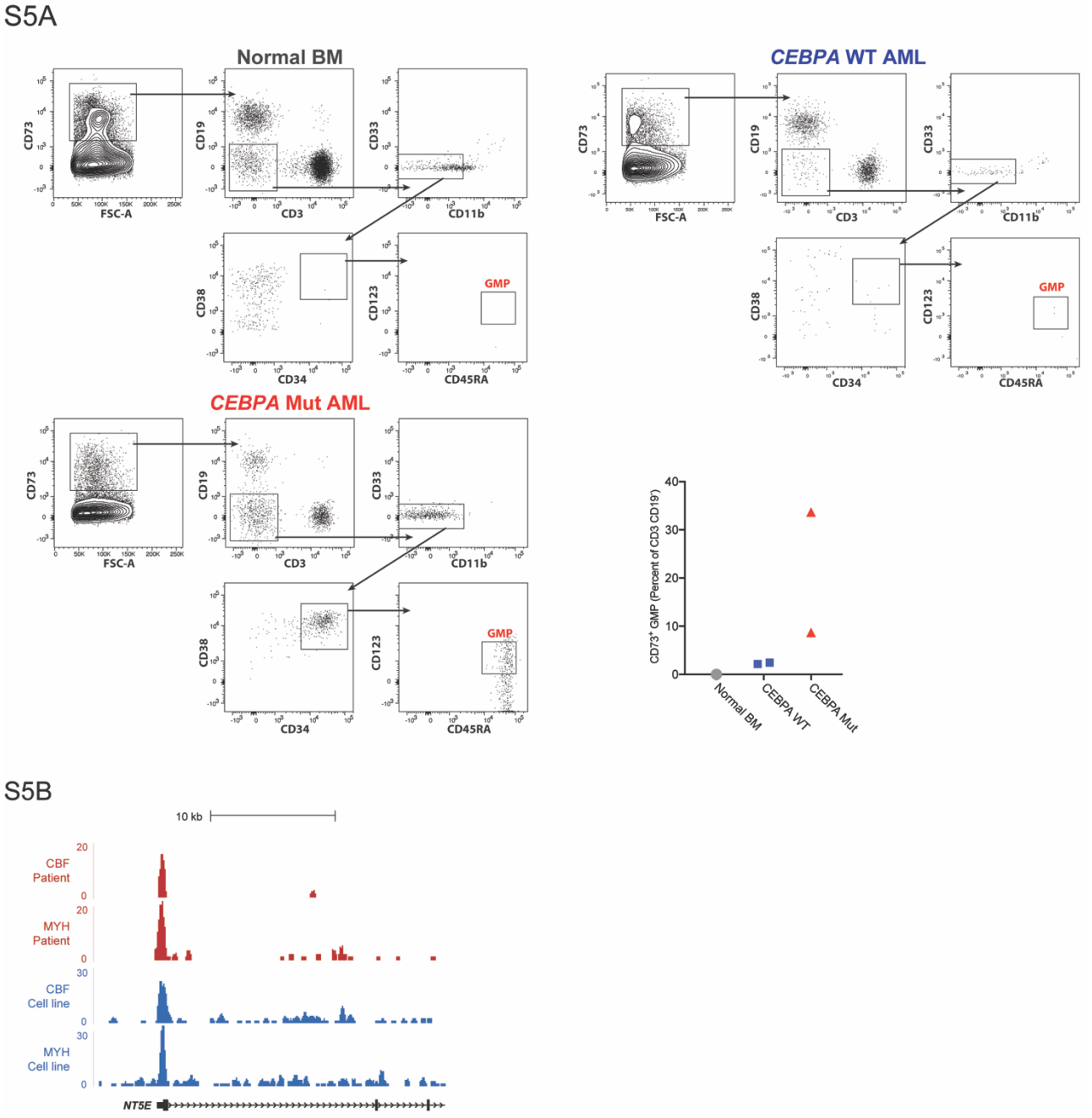


Fig. S5. CD73-positive GMPs in human AML and CBF-MYH binding of the *NT5E* promoter in INV(16) AML, related to Fig. 5. (S5A) Representative example of FACS gating strategy used to identify CD73+ GMPs in normal BM and BM from AML patients with or without bi-allelic CEBPA mutations. Quantification of CD73+ GMPs as percent of all CD73+, CD3-, CD19- cells is shown in the lower right panel. **(S5B)** ChIP-seq analysis from Mandoli et al., 2014 *(28)*, visualized in the genome browser using the provided track-hubs. Tracks of CBF as well as MYH visualized for both patient samples and cell lines of INV(16) AML.

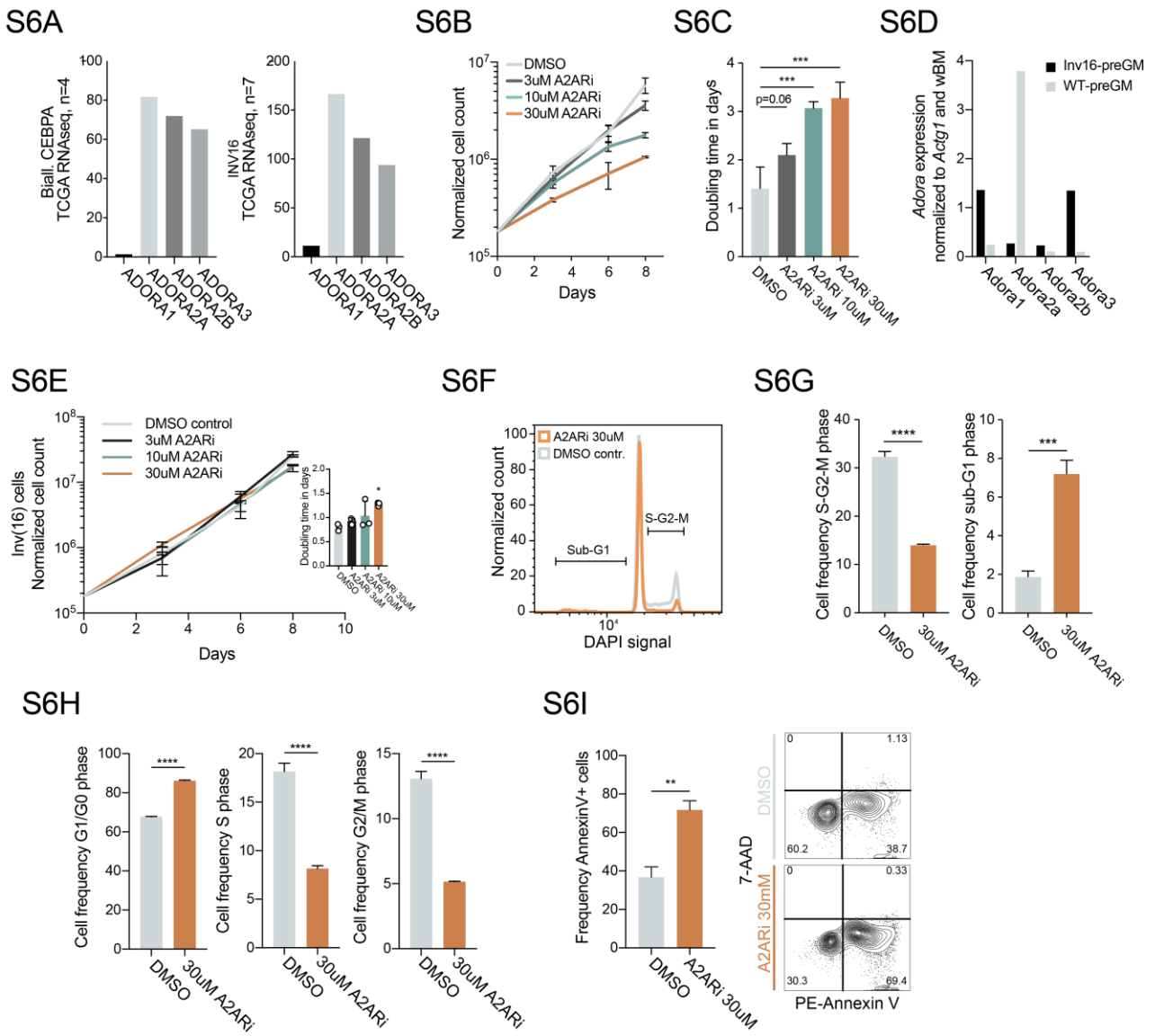


Fig. S6. A2AR-dependent adenosinergic autocrine survival signaling promoted by CD73, related to Fig. 6. (S6A) Expression levels of human *ADORA1, 2A, 2B and 3* in datasets extracted from the cBioPortal (TCGA data, RNA-seq). Left, biallelic *CEBPA* mutant patient data (n=4). Right, INV(16) patient data (n=7). **(S6B)** *In vitro* growth curve of Lp30 cells with 3, 10 or 30 μM of SCH-58261 or DMSO control. Cells were counted and re-plated at fixed cell density at each time point at day 3, 6 and 8. Mean +/- SD (n=3). **(S6C)** Doubling times for each growth condition in (B). **(S6D)** Normalized expression levels of murine *Adora1, 2a, 2b and 3* (assessed by RT-qPCR) in pre-GMs sorted from recipients transplanted with *Cbfb-Myh11* cKit^{D816Y} donor BM cells and WT controls (n=3 for each phenotype). **(S6E)** *In vitro* growth curve of Inv(16) cells cultured with either 3, 10 or 30 μM of SCH-58261 or DMSO control. Cells were counted and re-plated at fixed cell density at each time point at day 3, 6 and 8. Mean +/- SD (n=3 for each condition). **(S6F)** DAPI stain of fixed Lp30 cells plated with 30 μM SCH-58261 or DMSO control for 24h, representative plot of three replicates, frequency of DAPI+ singlets, normalized to maximum cell count (modal). **(S6G)** Left panel, quantification of DAPI stain S-G2-M (left panel) and Sub-G1(right panel) gate from F Mean +/- SD (n=3 for each condition). **(S6H)** Cell cycle analysis of p30 cells treated with 30 μM SCH-58261 or DMSO control (n=3). **(S6I)** Proportion of Annexin-V positive p30 cells treated with 30 μM SCH-58261 or DMSO control for 72h (n=3). Representative FACS plots are shown in the right panel. Statistics were determined by Student's two-tailed *t* test. *, P≤0.05; **, P≤0.01, ****, P≤0.000

Data file S1. CEBPA-peak coordinates, enriched motifs in CEBPA regions, related to Fig. 1.

Data file S2. RNA-seq data, related to Figs. 2 and 3.

Data file S3. MS data, related to Figs. 2 and 3.

Data file S4. Gene signatures, related to Fig. 3.

Data file S5. RNA-seq data, related to Fig. 6.

Data file S6. Gene signatures, related to Fig. 6.

Data file S7. Materials lists, related to Materials and Methods.