The microRNA miR-196b acts as a tumor suppressor in Cdx2-driven acute myeloid leukemia

Vijay P. S. Rawat,[1] Maria Götze,[1*] Avinash Rasalkar,[1] Naidu M. Vegi,[1] Susann Ihme,[1] Silvia Thoene,[2] Alessandro Pastore,[3] Deepak Bararia,[3] Hartmut Döhner,[4] Konstanze Döhner,[4] Michaela Feuring-Buske,[1,4] Leticia Quintanilla-Fend[5] and Christian Buske[1]*

[*]VPR and MG contributed equally to the work

[1]Institute of Experimental Cancer Research, Comprehensive Cancer Center, University Hospital Ulm, Ulm; [2]Institute of Clinical Chemistry and Pathobiochemistry, Technical University of Munich, School of Medicine, Munich; [3]Department of Medicine III, LMU University Hospital, Munich; [4]Department of Internal Medicine III, University Hospital Ulm, Ulm and [5]Institute for Pathology, Eberhard-Karls University of Tübingen, Tübingen Germany

Correspondence:
CHRISTIAN BUSKE - christian.buske@uni-ulm.de
VIJAY P. S. RAWAT - vijay.rawat@uni-ulm.de

doi:10.3324/haematol.2019.223297

Supplemental Data

The microRNA miR-196b acts as tumor suppressor in Cdx2 driven acute myeloid leukemia

Vijay P. S. Rawat[1#*], Maria Götze[1#], Avinash Rasalkar[1], Naidu M. Vegi[1], Susann Ihme[1], Silvia Thoene[2], Alessandro Pastore[3], Deepak Bararia[3], Hartmut Döhner[4], Konstanze Döhner[4], Michaela Feuring-Buske[1,4], Leticia Quintanilla-Fend[5] and Christian Buske[1*]

 # VPR and MG contributed equally to the work

[1]Institute of Experimental Cancer Research, Comprehensive Cancer Center, University Hospital Ulm, Germany

[2]Institute of Clinical Chemistry and Pathobiochemistry, Technical University of Munich, School of Medicine, Germany

[3]Department of Medicine III, LMU University Hospital, Munich, Germany

[4]Department of Internal Medicine III, University Hospital Ulm, Germany

[5]Institute for Pathology, Eberhard-Karls-University of Tübingen, Germany

[*]Corresponding Authors:

Christian Buske, MD

E-Mail: christian.buske@uni-ulm.de

Phone: +49 731 500 65800

Fax: +49 731 500 65822

Address: Institute of Experimental Cancer Research

CCC Ulm, University of Ulm

Albert-Einstein-Allee 11

89091 Ulm, Germany

Vijay P.S. Rawat, PhD

E-Mail: vijay.rawat@uni-ulm.de

Phone: +49 731 500 65805

Fax: +49 731 500 65822

Address: Institute of Experimental Cancer Research

CCC Ulm, University of Ulm

Albert-Einstein-Allee 11

89091 Ulm, Germany

Supplementary Materials and Methods:

BM transduction

Murine hematopoietic stem and progenitor cells (HSPCs) were obtained after 5-Fluorouracil (5-FU) treatment of the donor mice and were cultured in cell culture medium consisting of DMEM supplemented with 15% FBS, murine cytokines from Immunotools (6 ng/ml mIL3, 10 ng/ml mIL6 and 50 ng/ml mSCF) for 48h as previously described [1, 2] and retrovirally transduced with different constructs (miR-196b-*GFP*, *Cdx2*-*YFP*, *Cdx2*/miR-196b and empty *GFP*/*YFP* vector controls). Briefly, viral particles were produced by transfected E86 cells as described before [1, 3] or by 293T Phoenix Eco cells using TransIT-LT1 transfection reagent from Mirus Bio LLC (Madison, USA). For the latter case, after viral particle production 6-well suspension culture dishes were coated with RetroNectin (TAKARA Clontech, France) according to manufactures instructions. Viral particles were spun down on the RetroNectin coated plates for 45 min at 4°C with 1258 rcf. In parallel, for miR-196b-*GFP*/*Cdx2*-*YFP* co-expression, bone marrow (BM) cells in 500 µl culture medium were spin infected with 500 µl retroviral particle containing medium plus protamine sulfate (5ug/ml). Afterwards, BM cells plus an additional ml cell culture medium were layered on top of the virus bound RetroNectin coated wells at a concentration of $1x10^6$ cells/ml. After 72h, cells were harvested using the enzyme free dissociation buffer (Invitrogen), sorted one day later and used for *in vitro* assays or kept in culture for 48h for RNA-Seq or BM transplantation.

Human AML cell transduction

For the retroviral transduction of the OCI-AML3 AML cell line viral particles containing the empty vector or miR-196b were produced by 293T Phoenix Ampho cells. OCI-AML3 cells were 3-times spin infected in the presence of 8 µg/ml polybrene after 0-8-16h with parallel two-time cell layering on virus bound RetroNectin coated wells at the 0h and 16h time points.

Cells were harvested 8h after the last infection, 96h later FACS purified and 6-8 h later used for CFC assay or kept in culture for 48h for qRT-PCR analysis. For the lentiviral transduction of OCI-AML3, NB4 and Kasumi-1 AML cell lines, viral particles were produced by Lenti-X 293T cells, using TransIT-LT1 transfection reagent from Mirus Bio LLC (Madison). AML cell lines were infected for 48h at a concentration of 1-2x10^6/2 ml culture medium in the presence of 6-8 µg/ml polybrene with subsequent addition of culture medium as required. NB4 cells were washed and resuspended in fresh culture medium with 1.5-2 µg puromycin. After another 48h NB4 cells were purified using fluorescence-activated cell sorting (FACS) and used for proliferation assay or qRT-PCR analysis. OCI-AML3 and Kasumi-1 cells were either both puromycin and/or FACS purified. OCI-AML3 cells used for transplantation were not puromycin treated, but FACS purified 72h post transduction and transplanted four days post sort.

Transplantation

For murine transplantation experiments transduced and according to YFP or GFP expression highly FACS purified BM was intravenously injected into 8-16 weeks old lethally (1200 cGy) irradiated B6C3 [(C57BL/6J × C3H/HeJ) F1 (B6C3)] mice together with carrier cells (**Supplementary Table S*W*5**). For human transplantation experiments lentivirally transduced and highly FACS purified OCI-AML3 cells (no puromycin selection) were intravenously injected into 12-15 weeks old sublethally (325 cGy) irradiated and intraperitoneally immunoglobulin (IVIG) treated NSG [NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ] mice (0.5x10^6 cells per mouse). The animal studies were approved by the Ethics Committee of the University of Ulm and abided by the tenets of the revised World Medical Association Declaration of Helsinki (http://www.wma.net/e/policy/b3.htm).

Delta Colony Forming Unit-Spleen (ΔCFU-S) assay

The ΔCFU-S assay was performed to determine the impact of miR-196b overexpression on Cdx2 transforming activity. Highly FACS purified successfully transduced murine HSPCs (10000-15000 cells day 0 equivalent per mouse) were cultured in BM medium (DMEM) supplemented with cytokines (6 ng/ml mIL3, 10 ng/ml mIL6, 50 ng/ml mSCF and 15% FBS) for 7 days and subsequently transplanted into lethally irradiated recipient mice without helper cells. The frequency of ΔCFU-S cells was quantified by determining Telleyesniczky's based fixed colony numbers on the spleen 12 days post injection.

Quantitative Real-Time PCR (qRT-PCR)

For the detection of miRNA and mRNA, total RNA was isolated using the TRIzol reagent according to the manufacturer's protocol (Invitrogen). RNA was reversed transcribed into cDNA using specific miRNA primers from the TaqMan MicroRNA Reverse Transkription Kit (Applied Biosystems) or random hexamers from the Primescript RT-PCR kit (TAKARA) for mRNA reverse transcription. For the determination of overexpression, RNA was DNaseI (ThermoFisher) treated before reverse transcription. Expression levels were assayed by qRT-PCR with specific qRT-PCR primer/probe assays (**Supplementary Table S*W*7**). *RNU6b* and *RNU19* served as endogeneous control (housekeeping gene: HKG) for reverse transcribed miRNA. *TBP* or *Hprt* were used as HKG for human or murine cDNA, respectively (**Supplementary Table S*W*7**). Fold expression relative to the control sample was calculated using the formula: $2^{-\Delta\Delta CT}$ and fold expression relative to the HKG was calculated using the formula: $2^{-\Delta CT4}$.

Identification of miR-196b transcripts MF139050 and MF139051 in human hematopoietic cells

Using bioinformatic tools, different forward primers (FP) were designed at ~50 bp (basepair) intervals upstream of the miR-196b coding region and used together with the reverse primer (RP) located in exon 2 of *HOXA9*, which is a highly conserved sequence among known "Expressed Sequence Tags" (ESTs; CB127847) for conventional qPCR (**Supplementary Table S*W*7**). The final primer pair consisting of the FP located 154 bps upstream and the RP located 5782 bps downstream of the coding region for miR-196b resulted in two PCR products, firstly visualized by standard agarose based gel electrophoresis, secondly Sanger sequenced and thirdly confirmed via qRT-PCR.

Human samples and cell lines

Mononuclear cells (MNC) from diagnostic bone marrow (BM) or peripheral blood (PB) from 25 adult cytogenetically normal (CN) AML patients with the genotype NPM1c$^+$ (n=12) and NPM1wt/FLT3mt$^+$ (n=13) were analyzed (**Supplementary Table S*W*3**). The diagnosis of AML was performed according to the French-American-British criteria and the World Health Organization classification. The study was approved by the ethics committees of all participating institutions, and informed consent was obtained from all patients before they entered the study in accordance with the Declaration of Helsinki (http://www.wma.net/en/30publications/10policies/b3/index.html). BM CD34$^+$ and BM MNCs from healthy donors were purchased (Lonza) for comparative expression analyses. OCI-AML3, NB4 and Kasumi-1 AML cell lines were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ). HEK293T (CCL11268), 293T Phoenix Eco and 293T Phoenix Ampho cells were purchased from the American Type Culture Collection

(ATCC). Lenti-X 293T cells were purchased from Clontech. All cell lines were cultured according to the supplier's instructions.

Colony Forming Cell (CFC) assay

CFC assays were performed by plating 300-500 transduced murine hematopoietic stem and progenitor cells (HSPCs) per well in methylcellulose, supplemented with murine cytokines (MethoCult GF M3434, Stem Cell Technologies) colonies were counted 7-9 days post plating. All culturing was performed at 37°C in a humidified incubator under 5% CO2. For the murine cell re-plating experiments 1000-2500 cells/well were used obtained by harvesting all of the cells present in the previous CFC assay. For AML cell lines 500 transduced OCI-AML3, NB4 and 2000 transduced Kasumi-1 cells were plated in methylcellulose (H4330, Stem Cell Technologies) and counted 12-13 days post plating.

Luciferase assay

To prove promoter activity of the bioinformatically determined human promoter sequence of miR-196b, a luciferase assay was performed. The 802 bps long promoter fragment (**Supplementary Table S*W*8**) was successfully cloned into the pGL3-Basic vector with the firefly reporter gene (Promega) via the KpnI and XhoI restriction sites (primers: **Supplementary Table S*W*8**). The correct sequence was confirmed by complete nucleotide sequencing. $5x10^4$ HEK293T cells were seeded in 24-well plates in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% heat-inactivated fetal bovine serum (FBS) 16 h before transfection. Transient transfection was carried out using Lipofectamine 2000 reagent (Invitrogen) according to the manufacture's protocol. As internal control for co-transfections, the pRL-null construct encoding a Renilla luciferase gene (Promega) was used [5]. Firefly and Renilla luciferase activities were determined 24 h post transfection with the dual

luciferase reporter assay system (Promega). Samples were lysed in 100 µl of passive lysis buffer, and 20 µl of lysates were used to measure activity. Firefly luciferase readings by a luminometer were normalized against internal control renilla luciferase and calculated as fold difference against the activity obtained for the reporter plasmid pGL3-Basic transfection.

Plasmids

For retroviral gene transfer into primary BM cells, miR-196b transcripts (MF139050 and MF139051) were cloned into Murine Stem Cell Virus (MSCV) based vectors. Vectors with MSCV upstream of the *EGFP* or *EYFP* gene and the internal ribosomal entry site (IRES) are named MIG and MIY, respectively. The LMP vector was kindly provided by Scott Lowe (New York, USA) and was used for the overexpression of miR-196b ~127 bps upstream and downstream of the coding region (primers: **Supplementary Table S*W*8**), alternatively the Ph1-*GFP* plasmid from Addgene (Cambridge) was used. The MIY vector was used for the overexpression of MF139050 and MF139051 as well as for *Cdx2* [6]. Empty vectors were used as control. For the lentiviral gene transfer of miR-196b into the human AML cell lines, the SMART vector shMIMIC human lentiviral microRNA hsa-miR-196b-5p hEF1a- TurboRFP from Dharmacon was used and compared to the SMARTvector non-targeting hEF1a-TurboRFP vector. The correct sequences of the constructs were confirmed by complete nucleotide sequencing, and expression was proven by qRT-PCR.

Promoter prediction and transcription factor binding motif analysis

For the prediction of a promoter like region upstream (1.1kb) of the human miR-196b precursor sequence was analyzed by using online prediction programs such as Promoter 2.0 (www.cbs.dtu.dk/services/Promoter/) and GPMiner (http://gpminer.mbc.nctu.edu.tw/index.php). For the prediction of potential transcription factor

(TF) binding sites of the miR-196b promoter binding site prediction tools such as Alibaba2.1, TFBind and Jaspar were used as depicted in Figure S1G. The results of all programs were overlapped to predict the most probable TFs.

Flow Cytometric Analysis (FACS)

Immunophenotypic analysis was performed as previously described [7]. Briefly, single-cell suspensions were prepared from murine BM and PB. Red blood cells (RBCs) were lysed with 1% ammonium chloride solution, washed and resuspended in Flow Cytometric Analysis (FACS) staining buffer. Cells were incubated with a purified rat anti-mouse CD16/32 blocking antibody (Biolgend, Fell, Germany) and stained with: Gr1- APC-Cy7 or PE, Mac1- PE or APC, Ckit- APC or AlexaFluor700, Sca1- PE-Cy7, B220- APC, CD19- PE, CD4- APC, CD8- PE. For BM staining of NSG mice, cells were blocked with human serum and stained for CD45 – APC-H7. Cells were analyzed using a FACS FORTESSA LSR II (Becton Dickinson, Heidelberg, Germany) or a FACS Calibur Flow Cytometer (Becton Dickinson, Heidelberg, Germany).

Morphology of cells and histopathology analysis

The morphology of the cells from leukemic mice was analysed on cytospins stained with modified Wright-Giemsa. For histological analyses, sections of selected organs were prepared and stained at the Institute for Pathology Laboratory (Tübingen, Germany) using standard protocols as previously described [6, 8]

RNA-Sequencing analysis

RNA of murine highly purified HSPCs transduced with the different constructs after 96h transduction was isolated using the column based Direct-zol™ RNA Products kit (Zymo

Research) to perform RNA-sequencing (RNA-Seq). RNA of human functionally validated leukemic stem cells (LSCs) and $CD34^-$ cells was isolated using the miRNeasy Mini Kit (4.13) according to manufacturer's specifications. For both RNA-Seq experiments, sequencing libraries were prepared using Illumina TruSeq™ RNA Kit (Illumina, Inc.). All samples were run on the Illumina HiSeq2000 platform and sequenced as paired end reads. RNA-Seq data was analyzed using Basepair software (https://www.basepairtech.com/) with a pipeline that included the following steps: (1) Reads were aligned to the transcriptome derived from UCSC genome assembly mm10 for murine BM samples and hg19 for human stem cell subpopulations, using STAR [9] with default parameters. (2) Read counts for each transcript were measured using featureCounts [10]. (3) Differentially expressed genes were determined using DESeq2 [11] and cut off parameters of read count >10, $P<0.0046$ and P-adjusted (FDR, corrected for multiple hypotheses testing) < 0.3 were used for pathway and target gene analysis of the murine samples. Pathway, Ontology and cancer signature analysis was performed using Enrichr [12, 13] and the desktop application GSEA v3.0. For human samples expression counts from STAR in log-transformed transcript per kilobase million (TPM) format were evaluated.

microRNA-Sequencing

For the microRNA-Sequencing (miRNA-Seq) of functionally validated leukemic stem cells (LSCs) and healthy BM, miRNAs were isolated with the miRNeasy Mini Kit (4.13) according to manufacturer's specifications. Sequencing libraries were prepared using Illumina TruSeq™ small RNA sample preparation kit and technical replicates were sequenced as single end reads on Illumina HiSeq 2000. The raw single-end reads were adapter trimmed using "cutadapt" to remove the adapter sequence. Reads missing the adapter sequence were excluded from further analysis. Only reads that were at least 16 bps long after trimming were retained for further analysis. miRNA-Seq data was analyzed using Basepair software

(https://www.basepairtech.com/) with a pipeline that included the following steps: (1) Reads were aligned to the transcriptome derived from UCSC genome assembly hg19 using STAR [9] with default parameters. (2) Read counts for each transcript were measured using featureCounts [10]. (3) For single miRNA expression analysis of miR-196b, normalized count values from STAR in TPM were determined.

Isolation and validation of leukemic stem cell subpopulations

For the identification of functionally validated LSCs, $CD34^+$ cells from AML patients (**Supplementary Table S*E*14**) were highly purified for hematopoietic stem cell markers according to Goardon et al. 2011 [14] (lymphoid-primed multi-potential progenitor LMPP: $CD34^+CD38^-CD45RA^+CD90^-$ and granulocyte-macrophage progenitor GMP: $CD34^+CD38^+CD123^+CD110^-CD45RA^+$) by FACS and transplanted into sublethally irradiated non-obese diabetic/severe combined immunodeficient Gamma (NSG) mice. The cell fractions that resulted in engraftment (>1%) and a leukemic phenotype of the NSG mice were defined as functionally validated LSCs and used for miRNA/RNA-Seq compared to $CD34^-$ cells, which did not give rise to human leukemic cell engraftment. In addition, corresponding healthy BM counterparts were prepared and used for miRNA-Seq. Healthy BM subpopulations of 3 individuals were pooled resulting in two biological replicates.

Statistical analysis

Data were evaluated using the PRISM Graph pad software (La Jolla, California, USA). For statistical significance we first checked if the values of all experimental arms are normally distributed by using the recommended D'Agostino Pearson normality test for >3 biological replicates. Based on the outcome of the normality test we performed either parametric or

nonparametric interference statistics. In case of comparing the assay specific readout values of two independent experimental conditions, we used the t-test. In case of comparing more than two conditions the ANOVA was used. Differences with *P* values less than 0.05 were considered statistically significant. ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$. Kaplan–Meier plots for survival were analysed with the PRISM Graph pad software (La Jolla, California, USA) and significance was tested via the log-rank test. Microsoft Excel 2010 was used to determine the coefficient of determination and P-value in the regression analysis.

.

References:

1. Rawat VP, Cusan M, Deshpande A, et al. Ectopic expression of the homeobox gene Cdx2 is the transforming event in a mouse model of t(12;13)(p13;q12) acute myeloid leukemia. Proc Natl Acad Sci USA . 2004;101(3):817-22.

2. Rawat VP, Thoene S, Naidu VM, et al. Overexpression of CDX2 perturbs HOX gene expression in murine progenitors depending on its N-terminal domain and is closely correlated with deregulated HOX gene expression in human acute myeloid leukemia. Blood. 2008;111(1):309-19.

3. Rawat VP, Humphries RK, Buske C. Beyond Hox: the role of ParaHox genes in normal and malignant hematopoiesis. Blood. 2012;120(3):519-27.

4. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative CT method. Nature Protocols. 2008;3(6):1101-8.

5. Behre G, Smith LT, Tenen DG. Use of promoterless Renilla luciferase vector as an internal control plasmid for transient co-transfection assays of Ras-mediated transcription activation. Biotechnique. 1999;26(1):24-68.

6. Pineault N, Buske C, Feuring-Buske M, et al. Induction of acute myeloid leukemia in mice by the human leukemia-specific fusion gene NUP98-HOXD13 in concert with Meis1. Blood. 2003;101(11):4529-38.

7. Deshpande AJ, Cusan M, Rawat VP, et al. Acute myeloid leukemia is propagated by a leukemic stem cell with lymphoid characteristics in a mouse model of CALM/AF10-positive leukemia. Cancer Cell. 2006;10(5):363-74.

8. Buske C, Feuring-Buske M, Abramovich C, et al. Deregulated expression of HOXB4 enhances the primitive growth activity of human hematopoietic cells. Blood. 2002;100(3):862-8.

9. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.

10. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-30.

11. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

12. Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14:128.

13. Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44(W1):W90-7.

14. Goardon N, Marchi E, Atzberger A, et al. Coexistence of LMPP-like and GMP-like Leukemia Stem Cells in Acute Myeloid Leukemia. Cancer Cell. 2011;19(1):138-52.

15. Laurenti E, Doulatov S, Zandi S, et al. The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. Nature immunology. 2013;14(7):756-63.

16. Ng SWK, Mitchell A, Kennedy JA, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. Nature. 2016;540:433-437.

17. Network TCGAR. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. N Eng Journal of Medicine. 2013;368(22):2059-74.

18. Tyner JW, Tognon CE, Bottomly D, et al. Functional genomic landscape of acute myeloid leukaemia. Nature. 2018;562(7728):526-31.

Supplementary Tables:

Supplementary Table S*W*1: **Cell line abbreviations used from the UCSC genome browser in Supplementary Figure S1H.**

Abbreviation	**Cell line**
1	H1-hESC
A	A549
G	GM12878
H	HeLa-S3
I	IMR90
K	K562
L	HepG2
M	MCF-7
S	SK-N-SH, SK-N-SH_RA
U	HUVEC
a	AG04449, AG04450, AG09309, AG09319, AG10803, AoAF
b	BE2_C, BJ
c	Caco-2
d	Dnd41
e	ECC-1
f	Fibrobl
g	GM06990, GM08714, GM10847, GM12801, GM12864, GM12865, GM12872, GM12873, GM12874, GM12875, GM12891, GM12892, GM15510, GM18505, GM18526, GM18951, GM19099, GM19193, GM19238, GM19239, GM19240, Gliobla

h	HA-sp, HAc, HBMEC, HCFaa, HCM, HCPEpiC, HCT-116, HEEpiC, HEK293, HEK293-T-REx, HFF, HFF-Myc, HL-60, HMEC, HMF, HPAF, HPF, HRE, HRPEpiC, HSMM, HSMMtube, HVMF
m	MCF10A-Er-Src
n	NB4, NH-A, NHDF-Ad, NHDF-neo, NHEK, NHLF, NT2-D1
o	Osteobl
p	PANC-1, PBDE, PBDEFetal, PFSK-1, ProgFib
r	RPTEC, Raji
s	SAEC, SH-SY5Y, SK-N-MC
t	T-47D
u	U2OS, U87
w	WERI-Rb-1, WI-38

Supplementary Table S*W*2: Regression analysis of miR-196b, MF139050, MF139051 and *HOXA9* in healthy BM and CN-AML.

Comparison group	**Healthy BM**	**CN-AML Total**	**NPM1c$^+$**	**NPM1^{wt}/FLT3mt$^+$**
MF139050 vs miR-196b	R^2 = **0.9818** P = 0.08624 n = 3	R^2 = **0.2049** P = 0.02307 n = 25	R^2 = **0.31892** P = 0.05574 n = 12	R^2 = **0.4045** P = 0.01945 n = 13
MF139051 vs miR-196b	R^2 = **0.9265** P = 0.00207 n = 6	R^2 = **0.0031** P = 0.79707 n = 24	R^2 = **0.0304** P = 0.5880 n = 12	R^2 = **0.1446** P = 0.22271 n = 12*
MiR-196b vs *HOXA9*	R^2 = **0.9546** P = 0.00078 n = 6	R^2 = **0.0200** P = 0.50942 n = 25	R^2 = **0.0985** P = 0.32047 n = 12	R^2 = **0.0277** P = 0.60531 n = 12$^{\#}$
MF139050 vs *HOXA9*	*Excluded from analysis since no expression in BM CD34$^+$*	R^2 = **0.6537** P < 0.001 n = 25	R^2 = **0.4323** P = 0.02015 n = 12	R^2 = **0.7646** P < 0.001 0. n = 13
MF139051 vs *HOXA9*	R^2 = **0.8565** P = 0.00514 n = 6	R^2 = **0.0849** P = 0.16718 n = 24	R^2 = **0.0195** P = 0.66533 n = 12	R^2 = **0.6592** P = 0.00134 n = 12*

* 1 out of 13 samples with undetectable MF139051 expression was excluded for the regression analysis; # 1 out of 13 samples was identified as outlier and not considered in this comparison. CN-AML: Cytogenetically normal acute myeloid leukemia, NPM1c$^+$: Cytoplasmic nucleophosmin 1, FLT3 mutated (mt)$^+$: FLT3-ITD/FLT3-TKD = fms-like tyrosine kinase receptor-3 /internal tandem duplications /tandem kinase domain.

Supplementary Table S*W*3: Characteristics of CN-AML patients for bone marrow bulk analysis.

Patient No.	Karyotype	Age	Sex	Mutations
1	46,XY	46	M	$FLT3^{wt}/NPM1c^{+}$
2	46,XY	44	M	$FLT3^{wt}/NPM1c^{+}$
3	46,XY	43	M	$FLT3^{wt}/NPM1c^{+}$
4	46,XX	55	F	$FLT3\text{-}TKD^{+}/NPM1c^{+}$
5	46,XY	48	M	$FLT3^{wt}/NPM1c^{+}$
6	46,XX	51	F	$FLT3\text{-}TKD^{+}/NPM1c^{+}$
7	46,XX	34	F	$FLT3^{wt}/NPM1c^{+}$
8	46,XX	52	F	$FLT3\text{-}ITD^{+}/NPM1c^{+}$
9	46,XX	50	F	$FLT3^{wt}/NPM1c^{+}$
10	46,XX	44	F	$FLT3^{wt}/NPM1c^{+}$
11	46,XX	52	F	$FLT3\text{-}ITD^{+}/NPM1c^{+}$
12	46,XX	48	F	$FLT3\text{-}ITD^{+}/NPM1c^{+}$
13	46,XY	56	M	$FLT3\text{-}ITD^{+}/NPM1^{wt}$
14	46,XY	40	M	$FLT3\text{-}ITD^{+}/NPM1^{wt}$
15	46,XY	39	M	$FLT3\text{-}ITD^{+}/FLT3\text{-}TKD^{+}/\ NPM1^{wt}$
16	46,XX	51	F	$FLT3\text{-}ITD^{+}/NPM1^{wt}$
17	46,XY	31	M	$FLT3\text{-}ITD^{+}/NPM1^{wt}$
18	46,XX	61	F	$FLT3\text{-}ITD^{+}/NPM1^{wt}$
19	46,XX	34	F	$FLT3\text{-}ITD^{+}/NPM1^{wt}$
20	46,XY	40	M	$FLT3\text{-}ITD^{+}/NPM1^{wt}$
21	46,XX	22	F	$FLT3\text{-}ITD^{+}/NPM1^{wt}$
22	46,XX	66	F	$FLT3\text{-}ITD^{+}/NPM1^{wt}$
23	46,XX	53	F	$FLT3\text{-}ITD^{+}/NPM1^{wt}$
24	46,XX	75	F	$FLT3\text{-}ITD^{+}/NPM1^{wt}$

25	46,XY	83	M	FLT3-ITD^{+}/NPM1^{wt}

FLT3 mutated (mt)$^{+}$: FLT3-ITD/FLT3-TKD = fms-like tyrosine kinase receptor-3 /internal tandem duplications /tandem kinase domain, NPM1c^{+}: cytoplasmic nucleophosmin 1, F = female, M = male.

Supplementary Table S*W*4: **qRT-PCR expression values of *HOXA9*, mature miR-196b and MF139051 in normal BM CD34$^+$ stem and progenitor cells, normal total BM and CN-AML patients.**

Sample	No.	ΔC_T MF139051 (*TBP*)	ΔC_T miR-196b (*RNU6b*)	ΔC_T *HOXA9* (*TBP*)	2^-ΔC_T MF139051 (*TBP*)	2^-ΔC_T miR-196b (*RNU6b*)	2^-ΔC_T *HOXA9* (*TBP*)
BM CD34$^+$	1	8.26	-1.821	-0.55	0.0033	3.534	1.467
BM CD34$^+$	2	9.23	-1.046	-0.29	0.0017	2.065	1.226
BM CD34$^+$	3	8.19	-2.746	-0.51	0.0034	6.708	1.425
BM MNC	1	9.38	0.437	0.61	0.0015	0.739	0.656
BM MNC	2	11.3	2.694	1.25	0.0004	0.155	0.421
BM MNC	3	10.51	2.588	1.07	0.0007	0.166	0.475
NPM1c$^+$	1	8.78	-1.945	-2.77	0.0023	3.850	6.803
NPM1c$^+$	2	8.63	-0.757	-5.11	0.0025	1.690	34.653
NPM1c$^+$	3	9.80	-2.277	-3.79	0.0011	4.848	13.845
NPM1c$^+$	4	10.33	-2.117	-3.1	0.0008	4.337	8.595
NPM1c$^+$	5	3.54	-1.784	-7.67	0.0860	3.445	203.573
NPM1c$^+$	6	5.16	-0.262	-3.91	0.0280	1.199	15.015
NPM1c$^+$	7	9.62	-3.487	-3.14	0.0013	11.212	8.794
NPM1c$^+$	8	12.20	-2.890	-6.00	0,0002	7.412	63.924
NPM1c$^+$	9	13.46	-1.471	-7.59	0.0001	2.773	192.205
NPM1c$^+$	10	9.56	-2.792	-4.35	0.0013	6.925	20.366
NPM1c$^+$	11	12,05	-1.603	-6.03	0.0002	3.038	65.152
NPM1c$^+$	12	12,08	-0.330	-6.58	0.0002	1.257	95.701

$NPM1^{wt}/FLT3mt^{+}$	13	6.84	-1.153	-2.14	0.0087	2.224	4.420
$NPM1^{wt}/FLT3mt^{+}$	14	Undetermined	-0.227	-1.69	Undetermined	1.170	3.223
$NPM1^{wt}/FLT3mt^{+}$	15	5.01	-1.862	-5.77	0.0310	3.634	54.644
$NPM1^{wt}/FLT3mt^{+}$	16	7.39	0.666	-4.07	0.0060	0.630	16.769
$NPM1^{wt}/FLT3mt^{+}$	17	8.16	-0.322	-3.04	0.0035	1.250	8.238
$NPM1^{wt}/FLT3mt^{+}$	18	5.21	0.975	-6.65	0.0271	0.509	100.617
$NPM1^{wt}/FLT3mt^{+}$	19	5.45	-1.890	-3.72	0.0228	3.707	13.190
$NPM1^{wt}/FLT3mt^{+}$	20	7.65	-1.753	-3.38	0.0050	3.370	10.387
$NPM1^{wt}/FLT3mt^{+}$	21	7.23	2.097	-5.88	0.0066	0.234	58.721
$NPM1^{wt}/FLT3mt^{+}$	22	8.48	-2.525	-4.81	0.0028	5.758	27.973
$NPM1^{wt}/FLT3mt^{+}$	$23^{\#}$	11.25	4.799	4.80	0.0004	0.036	0.036
$NPM1^{wt}/FLT3mt^{+}$	24	0.82	-1.009	-9.37	0.5662	2.012	6661.167
$NPM1^{wt}/FLT3mt^{+}$	25	9.37	-1.549	-3.55	0.0015	2.927	11.703

$^{\#}$ Patient identified as outlier in the regression analysis of miR-196b versus HOXA9; BM: bone marrow, MNC: Mononuclear Cell, FLT3 mutated $(mt)^{+}$: FLT3-ITD/FLT3-TKD = fms-like tyrosine kinase receptor-3 /internal tandem duplications /tandem kinase domain, $NPM1c^{+}$: cytoplasmic nucleophosmin 1.

Supplementary Table S*W*5: Characteristics of mice.

Mouse no.	Retroviral construct	Transduced cells* transplanted + carrier cells ($x10^6$)	Day of sacrifice post transplantation	PB WBC $x10^6$/ml	PB RBC $x10^9$/ml	Spleen weight (mg)	BM blasts >20%/ Histopathology#	Engraftment % at day of sacrifice in BM		
								YFP	GFP	YFP/GFP
Cdx2										
1	Cdx2	0.1+0.3	158	50.50	4.41	714	Yes	95.8	0	0
2	Cdx2	0.1+0.3	190	NA	NA	1087	Yes	98.2	0	0
3	Cdx2	0.1+0.3	267	37.62	5.65	420	Yes	93.8	0	0
4	Cdx2	0.1+0.3	339	16.94	8.64	409	Yes	97.6	0	0
5	Cdx2	0.11+0.3	101	NA	NA	1235	Yes	93	0	0
6	Cdx2	0.11+0.3	193	151	3.54	1034	Yes	99.7	0	0
7	Cdx2	0.11+0.3	302	NA	NA	NA	Yes	95.6	0	0
8	Cdx2	0.15+0.45	99	102.36	0.56	504	Yes	94.7	0	0
Cdx2/miR-196b										
9	Cdx2/miR-196b	0.15+0.45	186	139.28	1.59	1317	Yes	0	0.8	99.2
10	Cdx2/miR-196b	0.15+0.45	242	NA	NA	NA	Yes	0.1	0.9	97.5
11	Cdx2/miR-196b	0.1+0.3	249	13.38	8.03	432	Yes	2.2	0.1	49.3

12	Cdx2/miR-196b	0.1+0.3	339	32.50	2.54	1759	Yes	0.1	0.5	98.2
miR-196b										
14	miR-196b	1+1	245	NA	NA	NA	AML#	0	86.5	0
15	miR-196b	1+1	217	NA	NA	NA	AML M0#	0	7.67	0
16	miR-196b	1+1	314	NA	NA	465	AML#	0	97.08	0

*Transduced cells were highly purified by cell sorting before transplant; NA: not analyzed, # according to histopathological analysis (data not shown), BM: bone marrow, PB: Peripheral Blood.

Supplementary Table S*W*6: Percent engraftment of OCI-AML3 cells transduced with miR-196b and vector control in the BM of sacrificed NSG mice.

SMART vector control		SMART miR-196b		Experiment number
% RFP$^+$	% CD45$^+$	%RFP+	% CD45+	
66.6	70.4	34.2	38.7	1
46.8	50.1	26.8	37.7	1
55.2	54.4	34.7	36.2	2
10.4	10.9	0	0.6	3
18.0	18.4	9.3	10.2	3
16.7	17.7	7.6	8.8	3

Supplementary Table S*W*7: List of primer/probe assays for qRT-PCR.

Gene name	Probe I.D.
hsa-miR-196b	002215
RNU19	001003
RNU6b	001093
MF139050	814305 A10
MF139051	814305 B1
TBP	Hs 99999910_m1
Hprt	Mm01545399_m1

Cdx2	Mm01212280
HOXA9	Hs00365956_m1
PBX1	Hs00231228_m1
MAPK1	Hs01046830_m1
MAPK3	Hs00385075_m1
MSI2	Hs01592569_m1
HOXA7	Hs00600844_m1

Supplementary Table S*W*8: miR-196b primer and promoter sequences

Row no.	**Description**	**Forward Primer 5'-3'**	**Reverse Primer 5'-3'**
1	miR-196b transcript amplification	CTGGGGCACTCTGTTGCACT	AGTCGAGCCACCTCGTACCTG
2	miR-196b promoter	ACCTTCCACTTTATCCCGTTCACCACTTTTACAACAGGAGGACTAACCCGAGCCCCTGCAATTACTTTAGGCATCTATTTAAATATTACCTAGACGGTCGTAATTTGTCTGGGCCCTATAGCCCTGGTGCCGTAAGGTTTGTCGGCTTTTGTTCAGTTTTATGGCTTGCTAGTATATCTGGATTGTGGCTGTCTTGGACCAGTGATTTCAGTTGAGAGGGGAGCTACATAGACAGAGGAAGCCAAACAGGATTTCTTTCGGGAGCCCCAGGGAGGCTTGCAGAGGCCGGTTATTTTGCGGGTGTCTGGGTCACAGATGACCCTCAATCTGAATCGGGGAGCAGGCCTGGCCCACAGCGCATCTCTAGCCGCTGGAGGCAGCGGTTAGGTGGACCCGGGCTGGACTGCATGGGGGCCTCCCCCTCCCCCAGTACGGCCTCCCCCACCTGGCGGCACTCCTTAGGGCCACGCGTTTCCTGCTTGCCAGAGTGGGGGGGGCGGGGGCTAAGGGGAGGGGGCGCAGGAGCGCGCGCCCCGGCGGGCGGGCAGCTAGGAGGGAGAGGGGGAGAGCGAGAGGCGGGAGGCCGGGGCCAGACAGGGAGCCGGGATCTGTACCGTCCCTGGCGCTTTCGAGCCTTCCACGGCTACCGCCTCTTGGCGCCGGCTCGCTGGGCTGCAAGATTTGGGAAAGCGGCGCGCACACTGCTTCTCGGAGCTGCTCTCGTCTCGCCTGCTCCCCGCCTTTATTCCTCTCTCCCTGCCTTTCCTCCCTCTCTTTTCCTCCCTCTCCCTCCCAGGAATAC

3	Cloning of the miR-196b promoter	AACTTGGTACCACCTTCCACTTTATCCCGTTCA	AACTTCTCGAGGTATTCCTGGGAGGGAGAGG
4	Cloning of the murine coding region for miR-196b	ACCTCGAGGTGGTGTCTGGTACAGGTTGC	ACGAATTCAGGAAGACCCGAAGCTCCTC

Supplementary Table S*E*1-14 are generated in Microsoft Excel (*E*): RNA-Seq data, pathway analysis, miRNA target analysis and characteristics of CN-AML patients used for LSC analysis.

Uploaded as separate Excel file.

Supplementary Figures

Supplementary Figure S1. miR-196b transcript and promoter characterization. (A) Schematic representation of the human HOXA9-10 genomic region to scale. Genes are represented as grey block arrows from their transcription start site (+1) to their poly-A (pA) sequence. A yellow block arrow head indicates the known stem loop coding sequence for miR-196b. The miR-196b precursor transcripts MF139050 and MF139051 are indicated by thick black lines. The blue dotted line indicates spliced out nucleotides. Primers used are presented by arrows, blue arrows resemble the primer set used for final transcript identification. (B) Alignment of MF139050 and MF139051 using Clustal Omega. (C) Evolutionary Conserved Sequence alignment of the genomic MF139050 and MF139051 sequence with mouse, dog and rhesus macaque. The ECR browser (hg19) provided species specific sequences for the genomic location of MF139050 and MF139051, which were used for alignment with Clustal Omega. (D) Fold expression of miR-196b relative to the housekeeping gene (HKG) *RNU6b* and normalized to the empty vector control (ctrl) was determined by qRT-PCR after transcript overexpression in HEK293T cells (n=3). Significance was determined by Kruskal-Wallis with Dunn's post hoc test, comparing the values to the vector ctrl. (E) Sequence alignment and homology of the newly identified miR-196b and the classical HOXA9 promoter (Sequence from SWITCHGEAR GENOMICS Prod. ID: S720614) with the UCSC Genome Browser (GRCh37/hg19). UCSC genes are displayed in blue bars: dark blue depicts coding and bright blue non-coding transcript sequences. CpG islands are shown in light green. "100 Vert. Cons" track corresponds to sequence conservation across 100 vertebrates. Conservation of individual organisms are displayed below (F) Evolutionary Conserved Sequence alignment of the genomic human miR-196b promoter region with frog, chicken, mouse and dog. The ECR browser (hg19) provided species specific sequences for the genomic location of the miR-196b

promoter, which was used for alignment with Clustal Omega. (G) Transcription factor (TF) binding analysis of the miR-196b promoter. Depicted are six trapezoids harboring the names of different online search platforms, which were used to predict TFs that can bind to the miR-196b promoter. The results were overlapped to predict the most probable TFs. TF in bold are ChIP-Seq validated from ENCODE provided by the UCSC genome browser. (H) Sequence alignment of ChIP-Seq validated TFs binding to the miR-196b promoter in the UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly. UCSC genes are displayed in blue bars: dark blue depicts coding and bright blue non-coding transcript sequences. CpG islands are shown in light green. A gray box encloses each peak cluster of TF occupancy, with the darkness of the box being proportional to the maximum signal strength observed in any cell line (cell line abbreviations can be found in Supplementary Table S*W*1) contributing to the cluster. The HGNC (acronym for a nomenclature committee) gene name for the TF is shown to the left of each cluster. Within a cluster, a green highlight indicates the highest scoring site. Stars (*) in the Clustal Omega alignment of panel B), C) and F) indicate identical bases, dashed line indicates missing bases and repeats are indicated by lower-case letters.

Supplementary Figure S2. Expression of miR-196b and *HOXA9* in normal and leukemic human BM. (A) Correlation of miR-196b transcript MF139051 with mature miR-196b (left) and mature miR-196b with *HOXA9* (right) ΔC_T-values in bone marrow (BM) $CD34^+$ (n=3) and mononuclear cells (MNCs; n=3). Light gray: BM MNC; Dark gray: BM $CD34^+$. Presented are the coefficient of determination (R^2) and P-value. (B) Expression levels of miR-196b (left) or MF139051 (right) and HOXA9 in healthy and CN-AML BM determined by qRT-PCR relative to the housekeeping gene *RNU6b* or *TBP*. Data were depicted in box-plots using the Tukey method for plotting whiskers and outliers. BM CD34+ (n=3); BM MNC (n=3); Total CN-AML (combined expression values of $NPM1c^+$ and $FLT3mt^+$; n=25 (left) and n=24 (right); $NPM1c^+$

(cytoplasmic nucleophosmin 1; n=12); NPM1wt/FLT3mt^{+} (FLT3 mutated (mt): FLT3-ITD/FLT3-TKD = fms-like tyrosine kinase receptor-3 /internal tandem duplications /tandem kinase domain; n=13 (left) and n=12 (right). Significance was determined by Kruskal-Wallis with Dunn's post hoc test, comparing the values to BM CD34^{+}.

Supplementary Figure S3. Expression of miR-196b and *HOXA9* in LSC subpopulations. (A) Log-transformed transcript per kilobase million normalized count values (TPM) of miR-196b in CN-AML leukemic stem cells (LSC^{+}): leukemic granulocyte-macrophage progenitor (LGMP; n=5) and leukemic lymphoid-primed multi-progenitor (LLMPP; n=5) versus healthy bone marrow (BM) subpopulations: GMP (n=2 pools) and LMPP (n=2 pools). Significance was tested by Mann-Whitney *U* test. Data were depicted in box-plots using the Tukey method for plotting whiskers and outliers. (B) Relative quantile normalized log2 transformed microarray expression values of pre-miR-196b (ILMN_3309089) and *HOXA9* (ILMN_1702479) in healthy progenitor cells (PROG: MLP: n=4, GMP: n=5 and MPP: n=5) cells from cord blood (CB; GSE42414) [15] compared to relative variance stabilized, quantile normalized and log2 transformed microarray expression values of pre-miR-196b (ILMN_3309089) and *HOXA9* (ILMN_1702479) in LSCs (n=138; GSE76009) [16]. Human LSCs were defined by phenotype (CD34/CD38) and *in vivo* engraftment. Samples were scanned on a BeadArray™ Reader using BeadScan software (v3.2). Data were analyzed using GEO2R software. Significance in healthy PROG CB was determined by 2-tailed Student's *t* test and in LSCs by Mann-Whitney *U* test. Data were depicted in box-plots using the Tukey method for plotting whiskers and outliers.

Supplementary Figure S4. Correlation of expression for miR-196b and *HOXA9* in leukemic BM. (A) Correlation of miR-196b transcript MF139051 (left) and mature miR-196b (right) with *HOXA9* ΔC_T-values in bulk BM of CN-AML subtypes. Blue: NPM1c^{+}

($n_{MF139051}$=12; $n_{miR\text{-}196b}$=12); Green: $NPM1^{wt}/FLT3mt^{+}$ ($n_{MF139051}$=12*; $n_{miR\text{-}196b}$=12[#]). *1 out of 13 samples with undetectable MF139051 expression was excluded for the regression analysis. [#] 1 out of 13 samples was identified as outlier and not considered in this comparison. All expression values including the outliers' can be found in the Supplementary Table S*W*4. (B) Correlation of miR-196b with *HOXA9* normalized expression counts in TPM in leukemic stem cells. Red: leukemic GMP (n=5); Green: leukemic LMPP (n=5). For the correlation plots a trend line with coefficient of determination (R^2) and the P-value are depicted.

Supplementary Figure S5. Impact of miR-196b overexpression on clonogenic and proliferative potential of murine HSPCs. (A) Fold expression of miR-196b relative to the housekeeping gene (HKG) *RNU19* or *RNU6b* and normalized to the empty vector control (ctrl.) was determined by qRT-PCR after Ph1-miR-196b transduction in murine hematopoietic stem and progenitor cells (HSPCs; n=3). (B) Colony (top) and cell number (bottom) of HSPCs transduced with miR-196b or the empty vector ctrl. using a Ph1 (n=5) vector system after 7-9 days in methylcellulose. Colonies were classified in granulocyte/macrophage (GM), megakaryocyte (M) and granulocyte (G) colony forming units. (C) Fold expression of miR-196b and *Cdx2* in cells from 1st CFC transduced with LMP-miR-196b (n=3) and/or MIY-*Cdx2* (n=3). Expression values, determined by qRT-PCR, are relative to the HKG *RNU19*, *RNU6b* or *Hprt* and are normalized to the empty vector ctrl. (D) Yield of cells in % generated by 300-500 initially plated cells constitutively expressing *Cdx2*/miR-196b referred to the yield in the *Cdx2* arm (n=3) in re-plating experiments in methylcellulose. (E) Representative cytospins stained with May-Grünwald-Giemsa after third plating with a 600x original magnification. (F) Percent of cells after 7 days in culture (n=3). p indicates plating. Error bars indicate SEM. For all panels, statistical significance was determined by Mann-Whitney *U* test.

Supplementary Figure S6. Impact of miR-196b overexpression on clonogenic and proliferative potential of human AML cells. (A) Fold expression of miR-196b relative to the housekeeping gene (HKG) *RNU6b* and normalized to the vector control (ctrl) after lentiviral miR-196b transduction of OCI-AML3 96h after end of transduction was determined by qRT-PCR (n=8). (B) Cell proliferation curve of OCI-AML3 cells transduced with lentiviral miR-196b or the vector ctrl. Cell counts were taken at day 3 (n=6) and day 6 (n=9) after seeding. (C) Total number of highly proliferative colonies (>5000 cells) in methylcellulose (left) and yield of cells in % generated by 500 initially plated OCI-AML3 cells constitutively expressing miR-196b compared to the vector ctrl. (n=8). (D) RNA-Seq expression of *CDX2*, *HOXA9* and *HOXA7* genes in AML cell lines from the Cancer Cell Line Encyclopedia database (https://portals.broadinstitute.org/ccle) (left) and fold expression of *CDX2* relative to the HKG *TBP* determined by qRT-PCR in CDX2 positive (OCI-AML3, NB4) and negative (Kasumi-1) AML cell lines (n=3) (right). (E) Fold expression of miR-196b relative to the HKG *RNU6b* and normalized to the vector ctrl. after lentiviral miR-196b transduction of NB4 96h after end of transduction was determined by qRT-PCR (n=3). (F) Cell proliferation curve of NB4 cells transduced with lentiviral miR-196b or the vector ctrl. (n=3). Cell counts were counted every 48h until day 6 after seeding. (G) Total number of colonies (left) and yield of cells in % generated by 500 initially plated NB4 cells constitutively expressing miR-196b compared to the vector ctrl. (n=3) (right). (H) Fold expression of miR-196b in Kasumi-1 relative to the HKG *RNU6b* and normalized to the vector ctrl. after lentiviral miR-196b transduction of Kasumi-1 96h after end of transduction was determined by qRT-PCR (n=3). (I) Cell proliferation curve of Kasumi-1 cells transduced with lentiviral miR-196b or the vector ctrl. Cells were counted every 48h until day 6 after seeding: day 2 (n=2), day 4 (n=4), day 6 (n=4). (J) Total number of colonies in Kasumi-1 cell line transduced with miR-196b and vector ctrl. (n=3) (left) and yield of cells in % generated by 2000 initially plated Kasumi-1 cells constitutively expressing miR-

196b compared to the vector ctrl. (n=3) (right). (K) Engraftment of OCI-AML3 cells was quantified by the percentage of RFP^+ (left) and $CD45^+$ (right) cells in the bone marrow (BM) 4 weeks post injection (n=3, in total 6 mice per arm) relative to the engraftment of vector ctrl transduced cells. Absolute engraftment values are provided in Supplementary Table S*W*6. Error bars indicate SEM. Statistical significance for all panels was determined by Mann-Whitney *U* test.

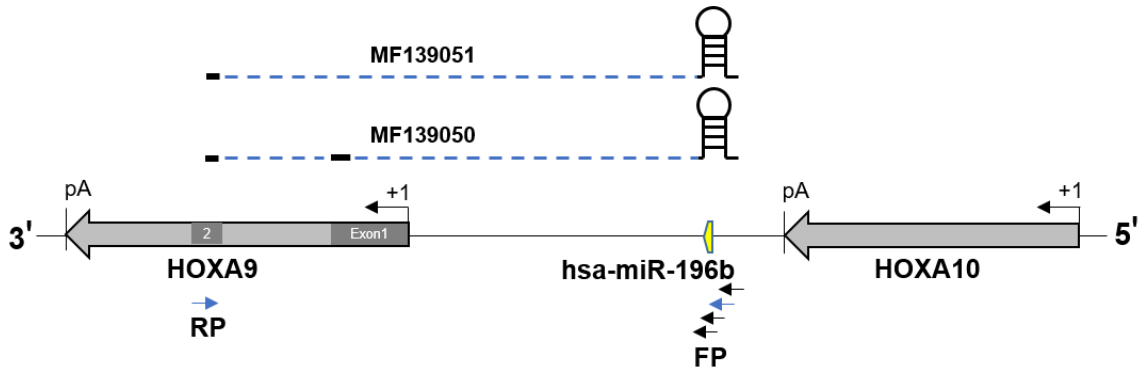
Supplementary Figure S7. RNA-Seq: DEG, pathway and miR-196b target gene expression analysis in murine HSPCs transduced with Cdx2, miR-196b or co-transduced with Cdx2 and miR196. (A) Heatmap of hierarchically clustered differentially expressed genes (DEGs) for LMP empty vector control (ctrl.) vs miR-196b (n=3) in murine hematopoietic stem and progenitor cells (HSPCs) 48h post sort. (B) Venn-diagram of all downregulated (DN) DEGs (blue) in miR-196b overexpressing cells compared to a miR-196b target gene list from "TargetScan2017" provided by the Enrichr platform (mint green). (C) Top 15 miRNA (based on P-value) enriched targets of all DN DEGs in miR-196b overexpressing murine HSPCs ($P<0.0017$; $FDR<0.1$) determined by Enrichr, "TargetScan microRNA 2017 analysis". Depicted are enriched miRNAs with $P<0.001$. (D) Heatmap of hierarchically clustered DEGs for miR-196b vs Cdx2/miR-196b in murine HSPCs 48h post sort (n=3). (E) Pathway enrichment analysis of DN miR-196b target DEGs in Cdx2/miR-196b overexpressing murine HSPCs. Depicted are top 10 enriched pathways with $P<0.0022$ using the Reactome database on the Enrichr platform. Enriched pathways are arranged according to P-values. (F) Significant enrichment of all DN genes in Cdx2/miR-196b in an oncogenic signature with the public gene matrix: "c6.all.v6.1.symbols [oncogenic]" determined by GSEA. Red "1" indicates the phenotype "BM Cdx2". Blue "0" indicates the phenotype "BM Cdx2/miR-196b". Depicted are nominal P-value and NES: Normalized enrichment score.

Supplementary Figure S8. Identification of relevant miR-196b targets. (A) Venn-diagram of downregulated (DN; blue) and upregulated (UP; red) differentially expressed genes (DEGs) in Cdx2/miR-196b overexpressing murine hematopoietic stem and progenitor cells (HSPCs) compared to the miR-196b target gene list form TargetScan by Enrichr (mint green). For the generation of the Venn-diagram the "Venn Diagram Plotter" was used. (B) Venn-diagram of DN miR-196b targets in Cdx2/miR-196b overexpressing murine HSPCs using 2 different miR-196b target lists. Depicted is the overlap (green) of miR-196b DN targets genes identified with a miR-196b target gene list form TargetScan 2017 provided by Enrichr (blue) and miR-196b DN target genes identified with a miR-196b target list comprising 4 different databanks (miRTarBase 7.0; TarBase 6.0-8.0; TargetScan 7.1; miRDB) (mint green). For the generation of the Venn-diagram the "Venn Diagram Plotter" was used. (C) Evolutionary conserved miRNA binding sites of 5 predicted non-Hox miR-196b targets in the 3' untranslated region (UTR) are presented using "TargetScanHuman 7.1". The white box represents the conserved miR-196b-5p binding site in the 3' UTR of the depicted gene. (D-E) Expression values of selected miR-196b targets in (D) highly purified functionally validated CN-AML LSCs (LLMPP: n=5; LGMP: n=5) from our own RNA-Seq data and in (E) bulk BM of CN-AML patients from the RNA-Seq TCGA database[17]. Only patients that were analyzed for miRNA and mRNA were considered (n=86). Data points in black indicate HOX genes, in grey stem cell regulators, in blue genes that are involved in proliferation and in green non-Hox DN DEGs in Cdx2 transformed cells only. In case of > 1 RNA variant an average value was calculated. (F) Expression of selected genes based on published data from the Vizome database using RNA-Seq in the following cohorts: CN-AML (n=325), NPM1 mutated BM samples ($NPM1^+$; n=108), t(8;21) (n=11) and healthy BM MNCs (n=19)[18]. Statistical significance was determined by Kruskal-Wallis with Dunn's post hoc test, comparing the values to BM MNC and t(8;21). Error bars indicate SEM. (G) Fold expression of miR-196b targets relative to the

housekeeping gene *TBP* and normalized to the empty vector control, whose value 1 is depicted as dashed line, after retroviral (left) and lentiviral (right) miR-196b transduction of OCI-AML3 96h after end of transduction determined by qRT-PCR (n=3-5). Statistical significance was determined by Mann-Whitney *U* test. Error bars indicate SEM.

Supplementary Figure S1

A)



B)

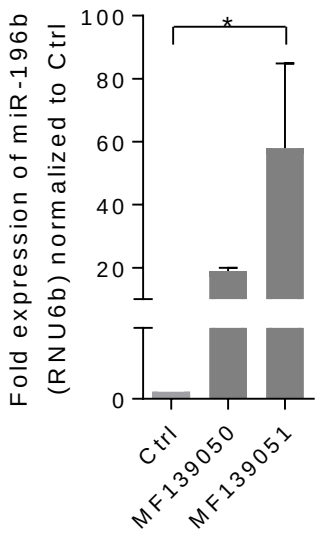
```
MF139050      CTGGGGCACTCTGTTGCACTGGCGGGCGCAGGTTGCCTAGGGGCTGGGCTGGGCCGGGCC   60
MF139051      CTGGGGCACTCTGTTGCACTGGCGGGCGCAGGTTGCCTAGGGGCTGGGCTGGGCCGGGCC   60
              ************************************************************

MF139050      AGGCGCGATGGCAGGGTTCTCTCCTTGGCGGCGGCGGCAGCGGCGGAGGCGGCGGCGGCG   120
MF139051      AGGCGCGATGGCAGGGTTCTCTCCTTGGCGGCGGCGGCAGCGGCGGAGGCGGCGGCGGCG   120
              ************************************************************

MF139050      GCGGGCGAGGCAGCCCTTCGCGGGCAGCACCAGAACTGGTCGGTGATTTAGGTAGTTTCC   180
MF139051      GCGGGCGAGGCAGCCCTTCGCGGGCAGCACCAGAACTGGTCGGTGATTTAGGTAGTTTCC   180
              ************************************************************

MF139050      TGTTGTTGGGATCCACCTTTCTCTCGACAGCACGACACTGCCTTCATTACTTCAGTTGAA   240
MF139051      TGTTGTTGGGATCCACCTTTCTCTCGACAGCACGACACTGCCTTCATTACTTCAGTTGAA   240
              ************************************************************

MF139050      ATCGTCTCCAGTTGATAGAGAAAAACAACCCAGCGAAGGCGCCTTCTCTGAAAACAATGC   300
MF139051      ATCGTCTCCAGATA----------------------------------------------   254
              ********** *

MF139050      TGAGAATGAGAGCGGCGGAGACAAGCCCCCCATCGATCCCAATAACCCAGCAGCCAACTG   360
MF139051      ------------------------------------------ACCCAGCAGCCAACTG   270
                                                         ****************

MF139050      GCTTCATGCGCGCTCCACTCGGAAAAAGCGGTGCCCCTATACAAAACACCAGACCCTGGA   420
MF139051      GCTTCATGCGCGCTCCACTCGGAAAAAGCGGTGCCCCTATACAAAACACCAGACCCTGGA   330
              ************************************************************

MF139050      ACTGGAGAAAGAGTTTCTGTTCAACATGTACCTCACCAGGGACCGCAGGTACGAGGTGGC   480
MF139051      ACTGGAGAAAGAGTTTCTGTTCAACATGTACCTCACCAGGGACCGCAGGTACGAGGTGGC   390
              ************************************************************

MF139050      TCGACT   486
MF139051      TCGACT   396
              ******
```

Supplementary Figure S1

C)

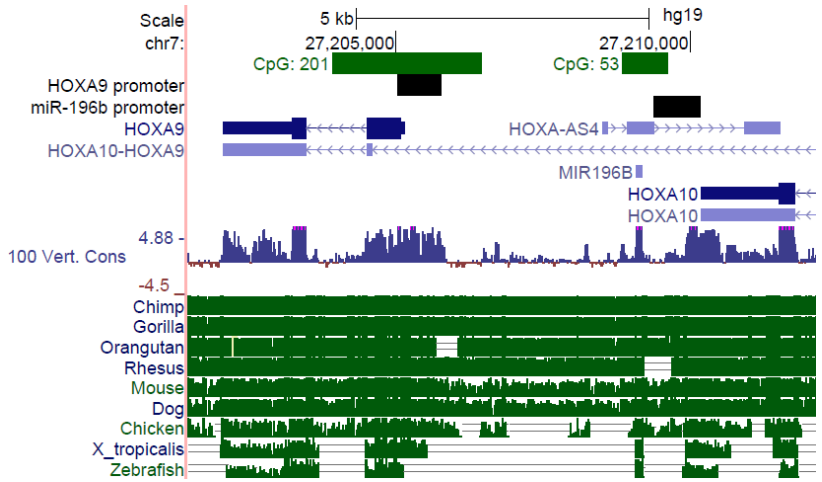
```
Mouse           AGCCGGGCCACCTCGTACCTGCGGTCCCGTGTGAGGTACATGTTAAACAGAAACTCCTTC    60
Dog             AGTCGGGCCACCTCGTACCTGCGGTCCCTGGTAAGGTACATGTTGAACAGAAACTCTTTC    60
RhesusMacaque   AGTCGGGCCACCTCGTACCTGCGGTCCCTGGTGAGGTACATGTTGAACAGAAACTCTTTC    60
MF139050        AGTCGAGCCACCTCGTACCTGCGGTCCCTGGTGAGGTACATGTTGAACAGAAACTCTTTC    60
MF139051        AGTCGAGCCACCTCGTACCTGCGGTCCCTGGTGAGGTACATGTTGAACAGAAACTCTTTC    60
                ** ** ************************  ** ************ ************ ***

Mouse           TCCAGTTCCAGCGTCTGGTGTTTTGTGTAGGGGCATCGCTTCTTCCGAGTGGAGCGAGCA    120
Dog             TCCAGTTCCAGCGTCTGGTGTTTGGTGTAGGGGCAGCGCTTCTTTCTCGTGGAGCGCGCA    120
RhesusMacaque   TCCAGTTCCAGGGTCTGGTGTTTTGTATAGGGGCAGCGCTTTTTCCGAGTGGAGCGCGCA    120
MF139050        TCCAGTTCCAGGGTCTGGTGTTTTGTATAGGGGCACCGCTTTTTCCGAGTGGAGCGCGCA    120
MF139051        TCCAGTTCCAGGGTCTGGTGTTTTGTATAGGGGCACCGCTTTTTCCGAGTGGAGCGCGCA    120
                *********** ************ ** ******** ***** ** *  ******** ***

Mouse           TGTAGCCAGTTGGCAGCCGGGTTATTGGGATCGATGGGGGGGCTTGTCTCCGCCGCTCTCA    180
Dog             TGGAGCCAATTGGCAGCAGGGTTGCTGGGATCGATGGGGGGGCTTGTCTCCGCCGCTCTCA    180
RhesusMacaque   TGAAGCCAGTTGGCCGCTGGGTTATTGGGATCGATGGGGGGCTTGTCTCCGCCGCTCTCA    180
MF139050        TGAAGCCAGTTGGCTGCTGGGTTATTGGGATCGATGGGGGGCTTGTCTCCGCCGCTCTCA    180
MF139051        TGAAGCCAGTTGGCTGCTGGGTTA----------------------------------    144
                ** ***** ***** ** *****

Mouse           TTCTCGGCATTGTTTTCGGAGAAGGCGCCTTCGCTGGGTTGTTTTTCTCTATCAACTGGA    240
Dog             TTCTCAGCATTGTTTTCGGAGAAGGCGCCTTCGCTGGGCTGTTTTTCTCTATCAACTGGA    240
RhesusMacaque   TTCTCAGCATTGTTTTCGGAGAAGGCGCCTTCGCTGGGTTGTTTTTCTCTATCAA-----    235
MF139050        TTCTCAGCATTGTTTTCAGAGAAGGCGCCTTCGCTGGGTTGTTTTTCTCTATCAACTGGA    240
MF139051        ----------------------------------------------------TCTGGA    150

Mouse           GATGCTTTCAACTGAAGTAATGAAGGCAGTGTCGTGCTGTCGAGAGAAAGGTGGATCCCA    300
Dog             GACGATTTCAACTGAAGTAATGAAGGCAGTGTCGTGCTGTCGAGAGAAAGGTGGATCCCA    300
RhesusMacaque   ----------------TAATGAAGGCAGTGTCGTGCTGTCGAGAGAAAGGTGGATCCCA    278
MF139050        GACGATTTCAACTGAAGTAATGAAGGCAGTGTCGTGCTGTCGAGAGAAAGGTGGATCCCA    300
MF139051        GACGATTTCAACTGAAGTAATGAAGGCAGTGTCGTGCTGTCGAGAGAAAGGTGGATCCCA    210
                          ****************************************

Mouse           ACAACAGGAAACTACCTAAATCACCGACCAGTTCTGATGCTGCGGGCGTAGGGTTGCCTC    360
Dog             ACAACAGGAAACTACCTAAATCACCGACCAGTTCTGGTGCTGCCCGCGCGGGGCTGCCTC    360
RhesusMacaque   ACAACAGGAAACTACCTAAATCACCGACCAGTTCTGGTGCTGCCCGCGCAGGGCTGCCTC    338
MF139050        ACAACAGGAAACTACCTAAATCACCGACCAGTTCTGGTGCTGCCCGCGAAGGGCTGCCTC    360
MF139051        ACAACAGGAAACTACCTAAATCACCGACCAGTTCTGGTGCTGCCCGCGAAGGGCTGCCTC    270
                ********************************** ******  ***  *** ******

Mouse           GCCCCTCGCCGCCTCCGCTGCTGCCACTGTCGCCGCCAAGGAAAGAGCCCTGCGATT---    417
Dog             GCCCGCCGCCGCCGCCGCCACCGCCGCCGCCGCCAAGGGAGAACCCTGCGATCGCGCCCG    420
RhesusMacaque   GCCCGCCGCCGATGCCGCCGCCGCCGCCGCCGCTGCCGTCGCCGCCGCCGCCAGGAGAGA    398
MF139050        GCccgccgccgccgccgcctccgccgctgccgccgccgc------cAAGGAGAGAACCCT    414
MF139051        GCccgccgccgccgccgcctccgccgctgccgccgccgc------cAAGGAGAGAACCCT    324
                ****  *****   ****  * *** * * ***                  *

Mouse           ------GTGCCCAGTCGAGCCCTCCTCCGCAACCTGTACCAGAC--ACCACCCTA---AC    466
Dog             GCCCGGC----------------------CCCTTGCCAACCTGCGCCCGCGCTGCC-AC    456
RhesusMacaque   ACCCTGCGATCGCGCCCGGCCCAGCGCAGCCCCTAGGCAACCTGCGCCCGCCGCTGCAAC    458
MF139050        GCCATCGCGCCTGGCCCGGCCCAGCCCAGCCCCTAGGCAACCTGCGCCCGCCAGTGCAAC    474
MF139051        GCCATCGCGCCTGGCCCGGCCCAGCCCAGCCCCTAGGCAACCTGCGCCCGCCAGTGCAAC    384
                                      *  *          ** *       **

Mouse           CCCAT-------       471
Dog             TGCGTGTCCCGC       468
RhesusMacaque   GGAATGCCCCA-       469
MF139050        AGAGTGCCCCAG       486
MF139051        AGAGTGCCCCAG       396
                    *
```

D)



Supplementary Figure S1

E)



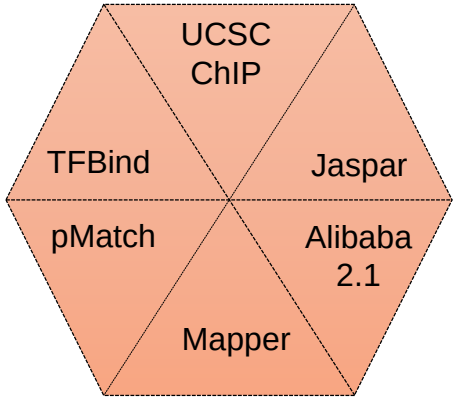
F)

```
Frog                ------------------------------------------------------------  0
Chicken             ------------------------------------------------------------  0
Mouse               GTACGCCTAGGAGGAGAGAGT---GAAAAGGAGGGAAGGGGGACGAGAAAGAGGGAGGAAA  58
Dog                 GTAAATCTGGGAGGGGGGGCG---GGGAGGGAAGGAAGGTAAAG--GAGAG--AGAGGAAA  53
Human               GTATTCCTgggagggagagggaggaaaaagagagggaggaaagg--cagggagagaggaat  58
RhesusMacaque       GTATGCCTTGGAGGGAGAGGGAGGAAAAGAGAGCGAGGAAAGG--CAGGAAGAGAGGAAT  58

Frog                ------------------------------------------------------------  0
Chicken             ----------------------------------------GTTCCCCCCCCCCCGAGGGC  21
Mouse               TAAAAAGGAAGGGGGGAAAAAACTAGAAGGGGGGGGGGGGGAGACCCAGCGGTGTGA-GGC  117
Dog                 ATGGAAGGAGAGGGG---------AAAAGCGAGAAGAAAGGAGAGCGAGCGGTGTGCGTGC  105
Human               aaaggcggggagCAG---------GCGAGACGAGAGCAGCTCCGAGAAGCAGTGTGCGCGC  110
RhesusMacaque       AAAGGCGGGGAGCAG---------GCGAGACGAGAGTAGCTCCGAGAAGCAGTGTGCACGC  110

Frog                ------------------------------------------------------------  0
Chicken             GGCTTTCGCCTCTCCATGCAGATGGGAAGGC----------CGTCCGCTGCGATGTTACA  71
Mouse               GGGTTCCCCAGAT-TCTGCAGCTTTGCCAAACAGAGCCTC----------CCTGGACGA  165
Dog                 TGGTTCCCCAAAT-CTTGCTCCTCGGCGAGCCGGCGCCAGGCGGCCGCAGCCCCGGGCGA  164
Human               CGCTTTCCCAAAT-CTTGCAGCCCAGCGAGCCGGCGCCAAGAGGCGGTAGCCGTGGAAGG  169
RhesusMacaque       TGCTTTCCCAAAT-CTTGCAGCCCAGCGAGCCGGCGCCAAGCGGCGGTAGCCGTGTAAGG  169

Frog                ------------------------------------------------------------  0
Chicken             TCATAAAAAGCTAACGGCTCCCTCTA-AGCCCGGGTTCAAAGTCTCCCCCCCCTCCTCCC  130
Mouse               CCAGAAAACGCAGGGGACAGCAACGAAGAGCCGGGTCCTGTGGCTGGCCGCGGCCTC--T  223
Dog                 CTCGAAAGCGCCTGGGACGGCACC---GAGCCCGGCTCCCGGTCTGGTCGCGGCCTC--C  219
Human               CTCGAAAGCGCCAGGGACGGTACA---GATCCCGGCTCCCTGTCTGGCCCCGGCCTC--C  224
RhesusMacaque       CTCGAAAGTGCCAGGGACAGTACA---GAGCCCGGCTGCCTGTCTGGCCCCGGCCTC--C  224

Frog                ------------------------------------------------------------  0
Chicken             GGCACACACATACCCGGGAAGAAAGCTGGAGCTGTCGCTCCTCTGCCAGCCCTCGTTGTG  190
Mouse               TCTACTCAC----------------------TCCCCGCAGCCGCTCGCTCTCCCCCTCTTC  262
Dog                 GGCGCTC-G----------------------CTCCCCGCGCGGCTCGCTCTCCCCCTCTCT  257
Human               CGCCTCT-C----------------------GCTCTCCCC-CTCTCCCTCC----------  251
RhesusMacaque       CGCCGCT-C----------------------GCTCTCCCGCCTCTGCCTCC----------  252

Frog                ------------------------------------------------------------  0
Chicken             CCTTCCCTTTGTGCGAGCCCCCGCTAGCTAAGCATGCAGCGAAGGA--------------  236
Mouse               TCCCTCCTAGGCGCCCGCCCGCTGAGGCGCGCGCTCCCGCGACCCTTCCCCCACC-----  317
Dog                 CCCTCCCAGCCAGCCAGCCGGGGCGCGCGCTCCCGCGCCCCCTCCCCCTGC-----  312
Human               --------TAGCTgcccgcccgccgggcgcgcgcgcTCCTGCGCCCCCTCCCCTTAGCCC--  302
RhesusMacaque       -------TAGCTGCCCTCCCGCCGGGGCGCGCGCTCCCGCGCCCCCTCCCCCTAGCCCTC  305

Frog                ------------------------------------------------------------  0
Chicken             ----GGGAGGATGC-AAAGGCGGCCCGGGGGGAAT---AAGCGCTCCCGCATGCAGCCAG  288
Mouse               TAGTGTTCCCCCCCCCCCTCCGGCGAGCAGGAAACGCGTGGCCTAGCGGAGGACCGCCAG  377
Dog                 CGTCGCGCCCCCCCCCACCTCCCGCGAGCAGGAAACGCGTGGCCCCGCGGAGGGCGACCAG  372
Human               -----CCGCCCCCCCCACTCTGGCAAGCAGGAAACGCGTGGCCCTAAGGAGTGCCGCCAG  357
RhesusMacaque       CCCCCACACACACAACTCTGGCCAGCAGGAAAAGCGTGGCCCTACGGAGGGCCGCCAG  365

Frog                ------------------------------------------------------------  0
Chicken             GTAGGGAACGGCTCCGTCCGAGGGGGCCCGGTGGCACGGAGGCGGCTCTCACCTCCCCCC  348
Mouse               GTGGGGGAGG----------------CCGGGCAGGGGGAGGGGGTGAC-CCTGGACAGC  419
Dog                 GTGGGGGAGG----------------CCGTGCTGGGGGAGGGGGCTTGCAGCCGGCAGC  415
Human               GTGGGGGAGG----------------CCGTACTGGGGGAGGGGGAGGCCCCCATGCCAGT  400
RhesusMacaque       GTGGGGGAGG----------------CCGTACTGGGGGAGGGGGCGGCCCCCATGCCGC  408
```

```
Frog                ------------------------------------------------------------  0
Chicken             TCCCCATGCGCTCACCTTCACGGTTTCTTCCGAGCCCTGAGATAAAGCTACGGTTTGTCA  408
Mouse               TCGGC-CCAGGTCCACTTAACCGCTGTTGC-AGGCA-----CTAGTGGTGCTCCGTGAGT  472
Dog                 A---------GTCCACCTAACTGCTGCCGC-GGACG-----CTAGTGACGCGCCCTGGGC  460
Human               CCAGC-CCGGGTCCACCTAACCGCTGCCTCCAGCGG-----CTAGAGATGCGCTGTGGGC  454
RhesusMacaque       CCAGC-CCAGGTCCACCTAACGGCTGCCTCCAGCGG-----CTAGGGATGCGCTGTGGGC  462

Frog                ------------------------------------------------------------  0
Chicken             CCGTCCTCGGCAGCTAAAGATCTCCGCACGCTCGTAGGTGTGTGATTAAAAGGGGGGTGG  468
Mouse               GGGCCCTCCTCCGC-GGA---TTTCGGGCGTGGGGTTCTTCTGGGACCTGGATAC-CCGCA  527
Dog                 CCAGCCTGCTCCCCTGAT---TTCTAGGCTGGGGGTCATCTAAGACCAGGGGAC-CCACA  516
Human               CAGGCCTGCT---CCCCG---ATTCAGATTGAGGGTCATCTGTGACCCAGACAC-CCGCA  507
RhesusMacaque       CAGGCCTGCT---CCC-C---ATTCAGATTGAGGGTCATCTGAGACTCAGACAC-CCGCA  514

Frog                ----------------------------------------------------------GCTTCCT  7
Chicken             GGGGGGTGGGGGGGAATAAAAACCTCTTCCCTACCCTCAAGAAATAGTAATTTGCTTCCT  528
Mouse               AAATAACTGGCTCCTGCGAGCCTCTCTGGGGCTCTCTAAAGAAATCCTGTTTAGCTTCCT  587
Dog                 AAACAACCGGCGTCTGCGAGCCTCCCTGGGGCTCCCGAAAGAAATCCTGTTTGGCCTCCT  576
Human               AAATAACCGGCCTCTGCAAGCCTCCCTGGGGCTCCCGAAAGAAATCCTGTTTGGCTTCCT  567
RhesusMacaque       AAATAACCGGCCTCTGCAAGCCTCCCTGGGGCTCCCGAAAGAAATCCTGTTTGGCTTCCT  574
                                                                        ** ****

Frog                ------------------------------------------------------------  0
Chicken             CTGTCTAAACAGTTCCTTGTACAACTGAAACCACTCGTCTAGGGTTGCAGCAAAGCAGAT  67
Mouse               CAGTCTGTGCAGCTCCTCAGTCAACTGAAACCACTCGTTCAAGACAGCAACAATCCAGAT  588
Dog                 CTGTCTATGCCGCTCTCCTCTCAACTGAAACCACTGGTCCAAGACAGCCACAATCCAGAT  647
Human               CTGTCTATGCAGCTCCCCTCTCAACTGAAACCACTGGTCCAAGACAGCCACAATCCAGAT  636
RhesusMacaque       CTGTCTATGTAGCTCCCCTCTCAACTGAAATCACTGGTCCAAGACAGCCACAATCCAGAT  627
Human               CTGTCTATGCAGCTCCCCTCTCAACTGAAATCACTGGTCCAAGACAGCCACAATCCAGAT  634
RhesusMacaque       CTGTCTATGCAGCTCCCCTCTCAACTGAAATCACTGGTCCAAGACAGCCACAATCCAGAT  634
                    * ****    * **     ********* **** **  * *   **  ***  *****

Frog                ------------------------------------------------------------  0
Chicken             ATGCGAGCAAGTCATAAAACTACACAAAAGCCGATCACCCTTCCAGCTCCATTGCTATAA  127
Mouse               ATACTAGCAAGTCATAAAACTGAACAAAAGCCGACAAACCTTAGAGCACCATCGCTATAT  648
Dog                 ATACTAGCAAGTCATAAAACTGAACAAAAGCCGACAAACCTTAAGGCACCAGGGCTATAG  707
Human               ATACTAGCAAGTCATAAAACTGAACAAAAGCCGACAAACCTTACGGCACCAGGGCTATAG  696
RhesusMacaque       ATACTAGCAAGCCATAAAACTGAACAAAAGCCGACAAACCTTACGGCACCAGGGCTATAG  687
Human               ATACTAGCAAGCCATAAAACTGAACAAAAGCCGACAAACCTTACGGCACCAGGGCTATAG  694
RhesusMacaque       ATACTAGCAAGCCATAAAACTGAACAAAAGCCGACAAACCTTACGGCACCAGGGCTATAG  694
                    ** * ****** *********  **********  * ****   ** ***  ******

Frog                ------------------------------------------------------------  0
Chicken             GGCCCAGACAAATTACGACTGTCTAGGTAATATTTAAATAGATTCCTAAAGTGATTGCAC  187
Mouse               GGCCCAGACAAATTACGACCGTCTAGGTAATATTTAAATAGATTCCTAAAGTAATTGCAG  708
Dog                 GGCCCAGACAAATTACGACCGTCTAGGTAATATTTAAATAGATGCCTAAAGTAATTGCAG  767
Human               GGCCCAGACAAATTACGACCGTCTAGGTAATATTTAAATAGATGCCTAAAGTAATTGCAG  756
RhesusMacaque       GGCCCAGACAAATTACGACCGTCTAGGTAATATTTAAATAGATGCCTAAAGTAATTGCAG  747
Human               GGCCCAGACAAATTACGACCGTCTAGGTAATATTTAAATAGATGCCTAAAGTAATTGCAG  747
RhesusMacaque       GGCCCAGACAAATTACGACCGTCTAGGTAATATTTAAATAGATGCCTAAAGTAATTGCAG  754
                    ****************** ********************** ******* ******

Frog                ------------------------------------------------------------  0
Chicken             TGGAGGAAGGGGGGGGGGGGTGCCAACTTCTCCTGTTGTAAAAGTTCCTGAAAGATAAAA  247
Mouse               GGGGTTTAGGCAAC-------------TATTCCTGTTGTAAAAGTTCTCGAAGACATAAA  755
Dog                 GGGGCTCGGGCTAG-------------CCCTCCTGTTGTAAAAGTGGTGGCAGGATAAA  814
Human               GGGGCGCGGGCTAG-------------TCCTCCTGTTGTAAAAGTGGTGGACGGGATAAA  803
RhesusMacaque       GG-GCTCGGGTTAG-------------TCCTCCTGTTGTAAAAGTGGTGAACGGGATAAA  793
RhesusMacaque       GGGGCTCGGGTTAG-------------TCCTCCTGTTGTAAAAGTGGTGAACGGGATAAA  801
                      *     **                ***************     *     ***

Frog                TTAGACCTA  256
Chicken             GTAGAACCT  764
Mouse               ATGGAAGGT  823
Dog                 GTGGAAGGT  812
Human               GTGGAAGGT  802
RhesusMacaque       GTGGAAGGT  810
                     * **
```

Supplementary Figure S1

G)



<u>≥ 4 prediction tools:</u>

SP1; CEBPA; **EGR1**; NFKAPPAB; MZF1;

TFAP2A; AML1

<u>= 3 prediction tools:</u>

CEBPB; ARNT; PBX1; IRF1; NF1; E2; HNF1,

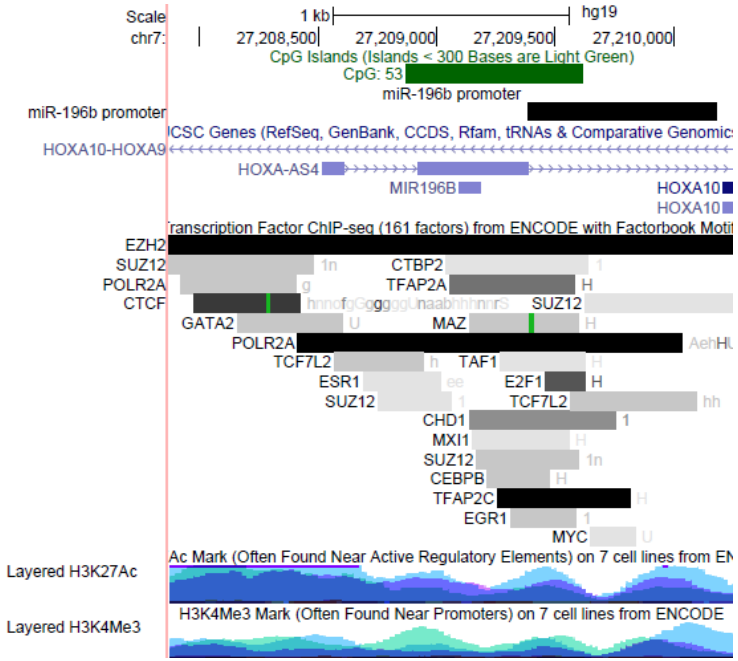
REL; CHOP; LYF1; NHLH1; YY1; ELK1;

COMP1; PAX2; CREB1

<u>= Mapper/UCSC</u>

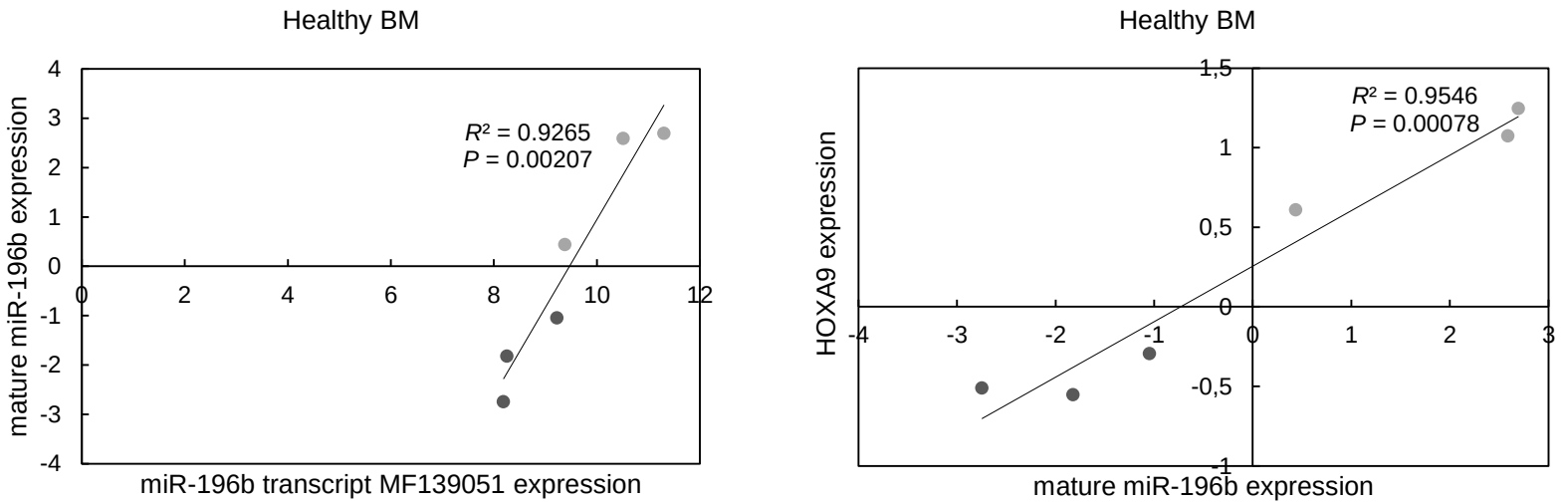
E2F1; **MAZ**

H)

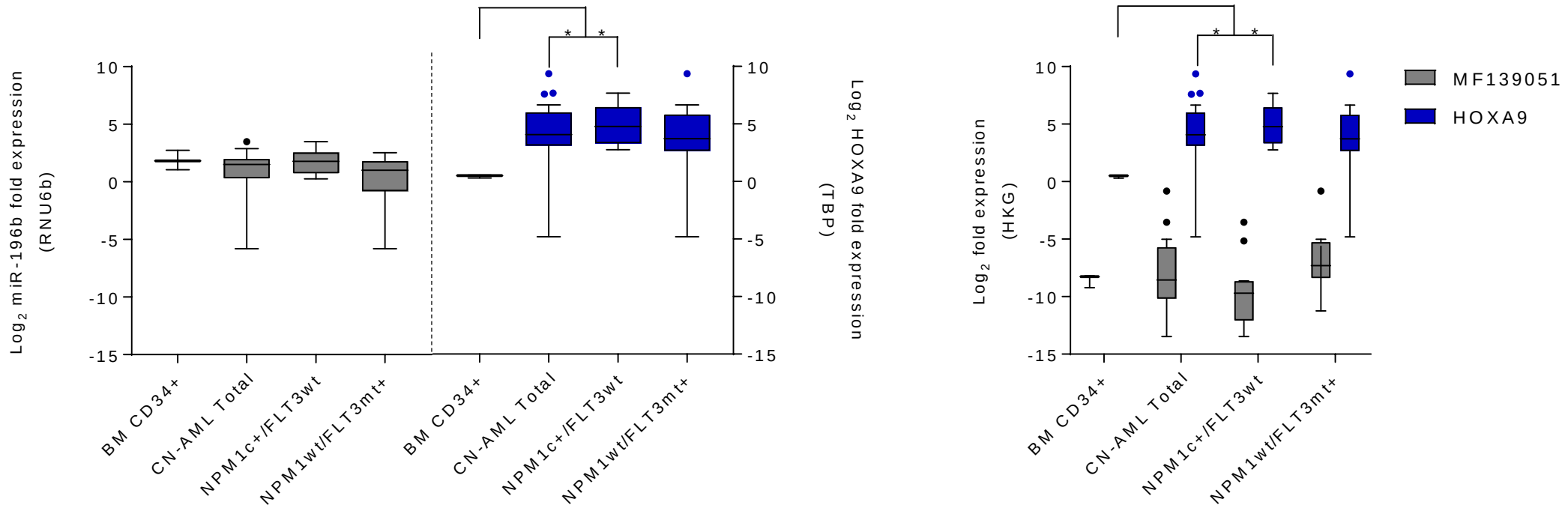


Supplementary Figure S2

A)

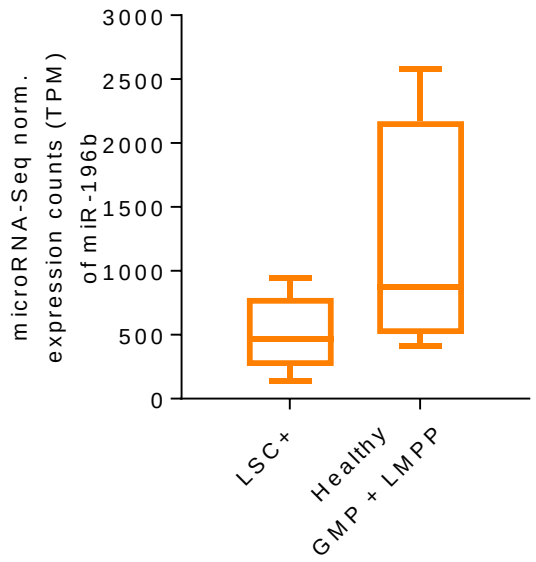


B)

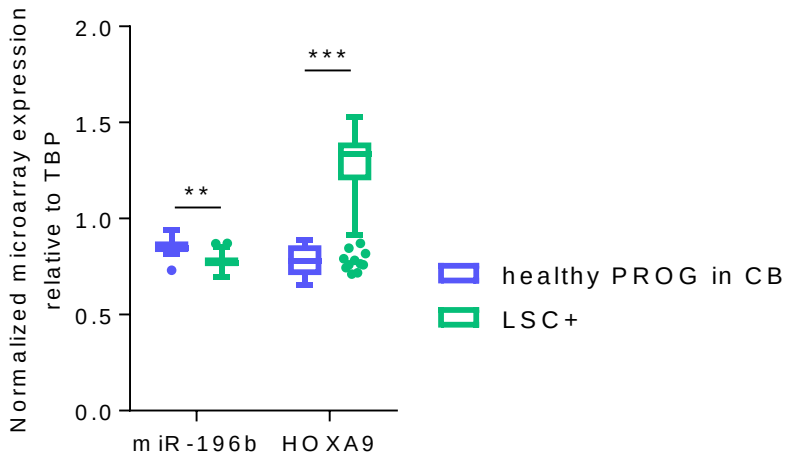


Supplementary Figure S3

A)

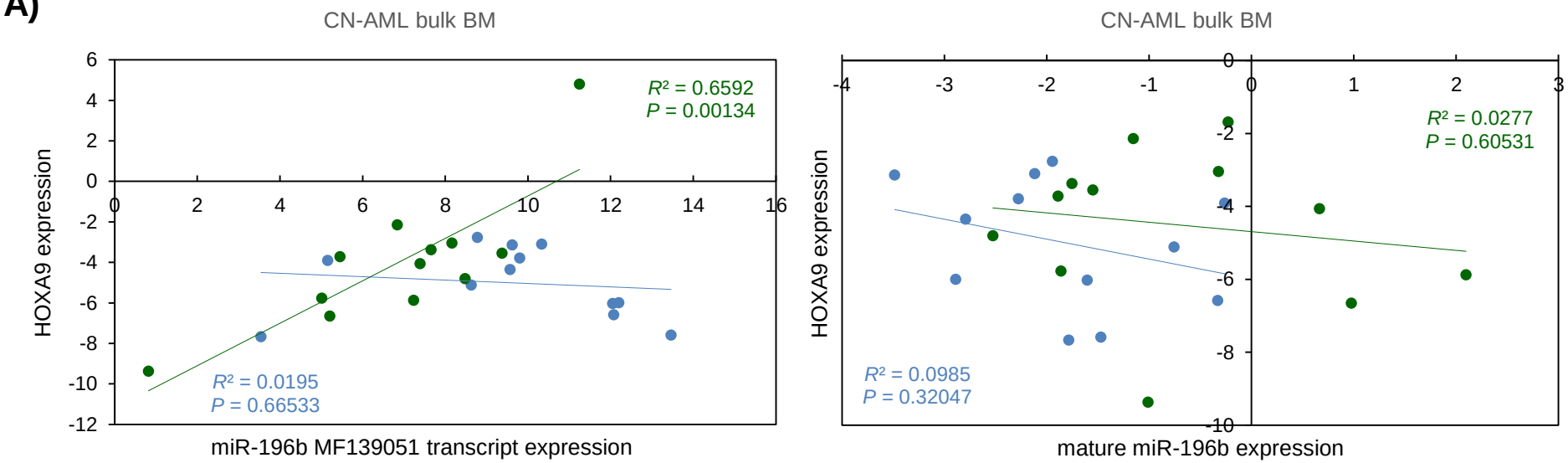


B)

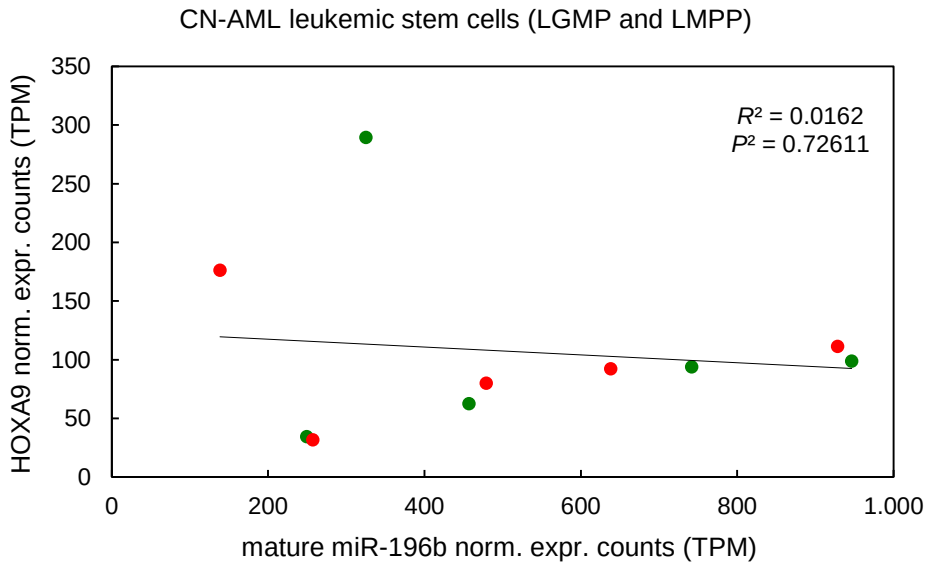


Supplementary Figure S4

A)

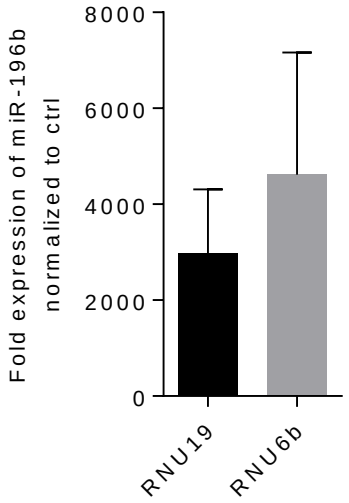


B)

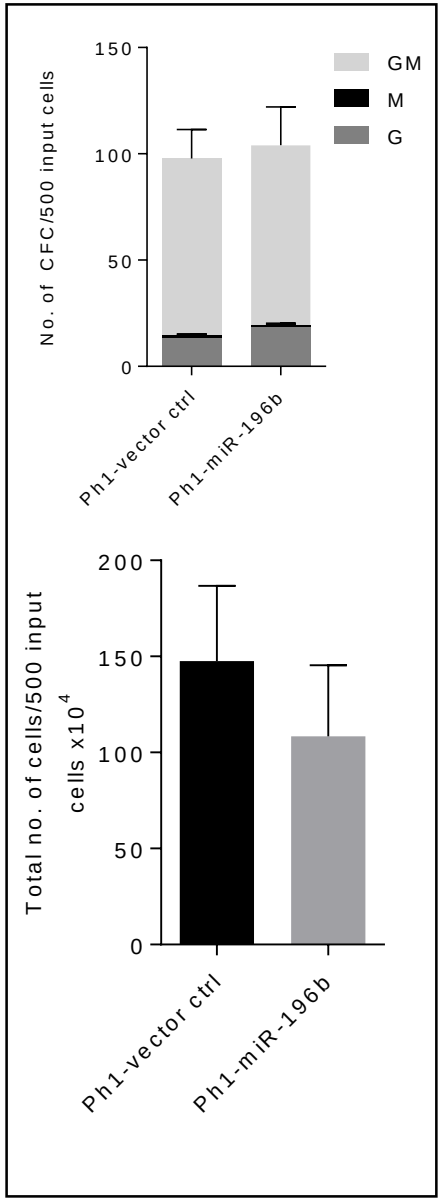


Supplementary Figure S5

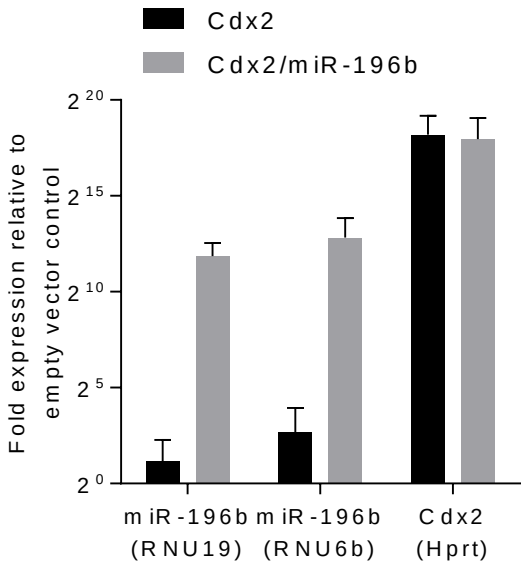
A)



B)

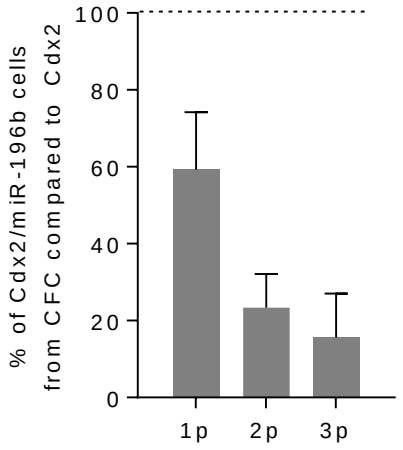


C)

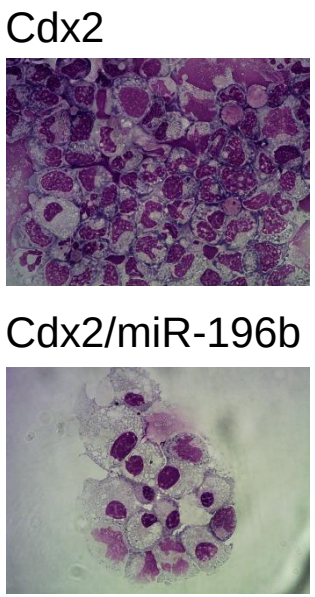


Supplementary Figure S5

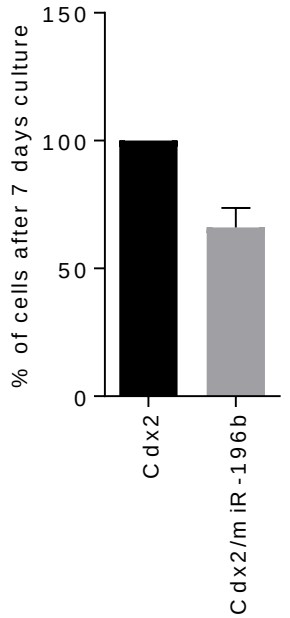
D)



E)

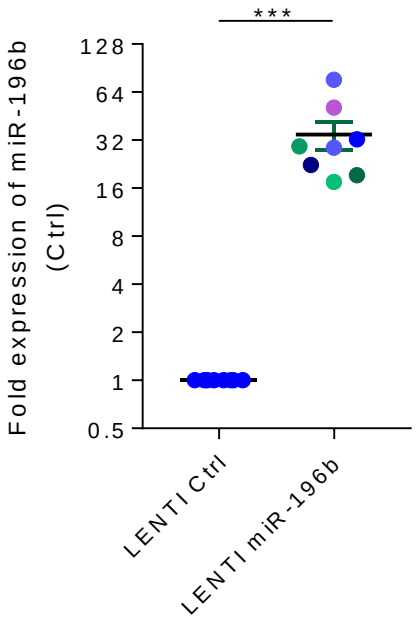


F)

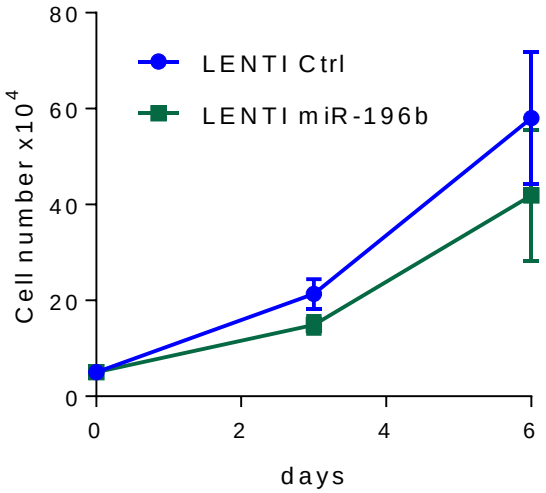


Supplementary Figure S6

A)

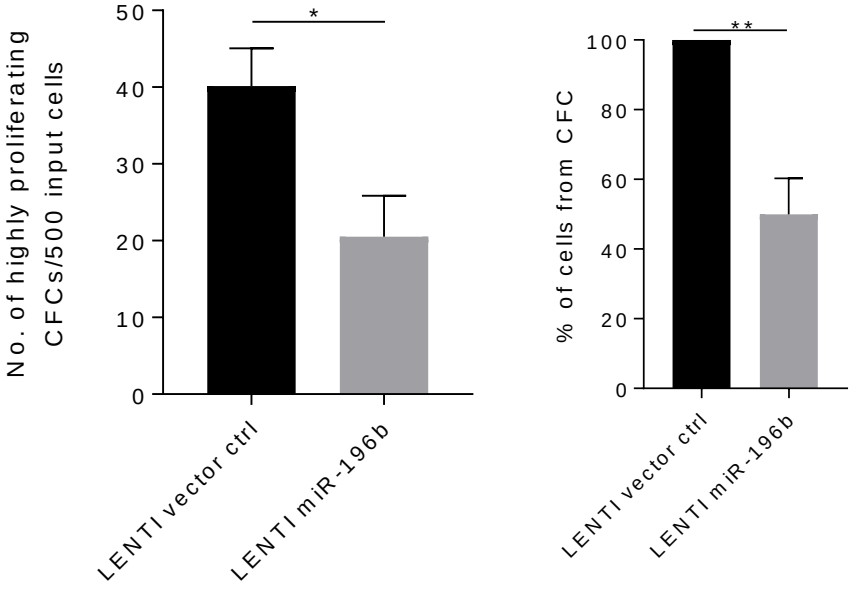


B)

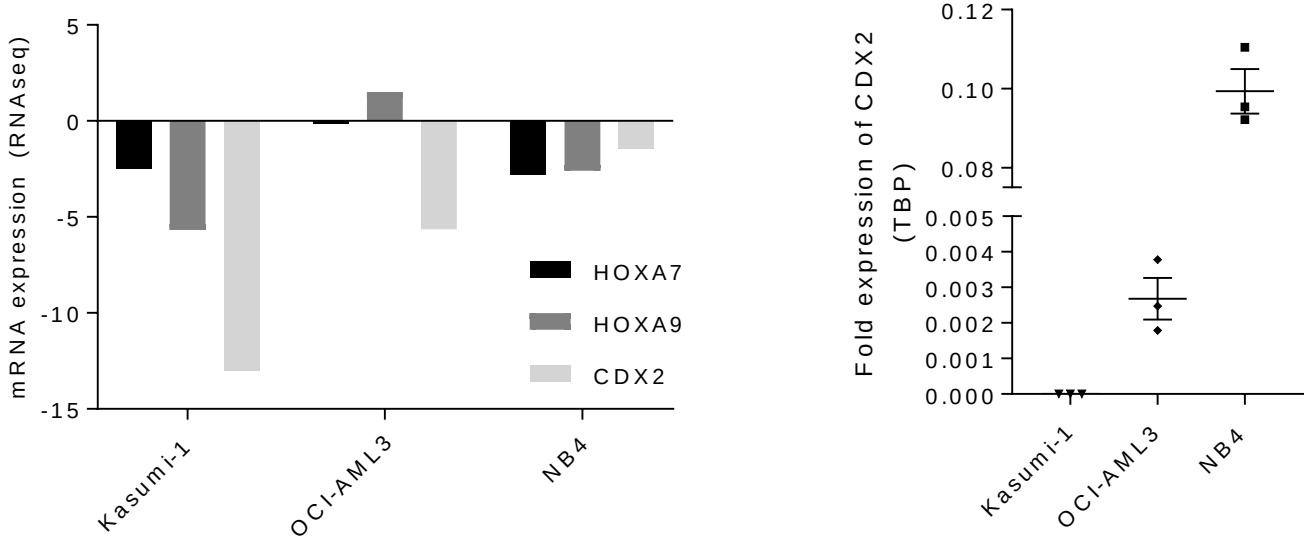


Supplementary Figure S6

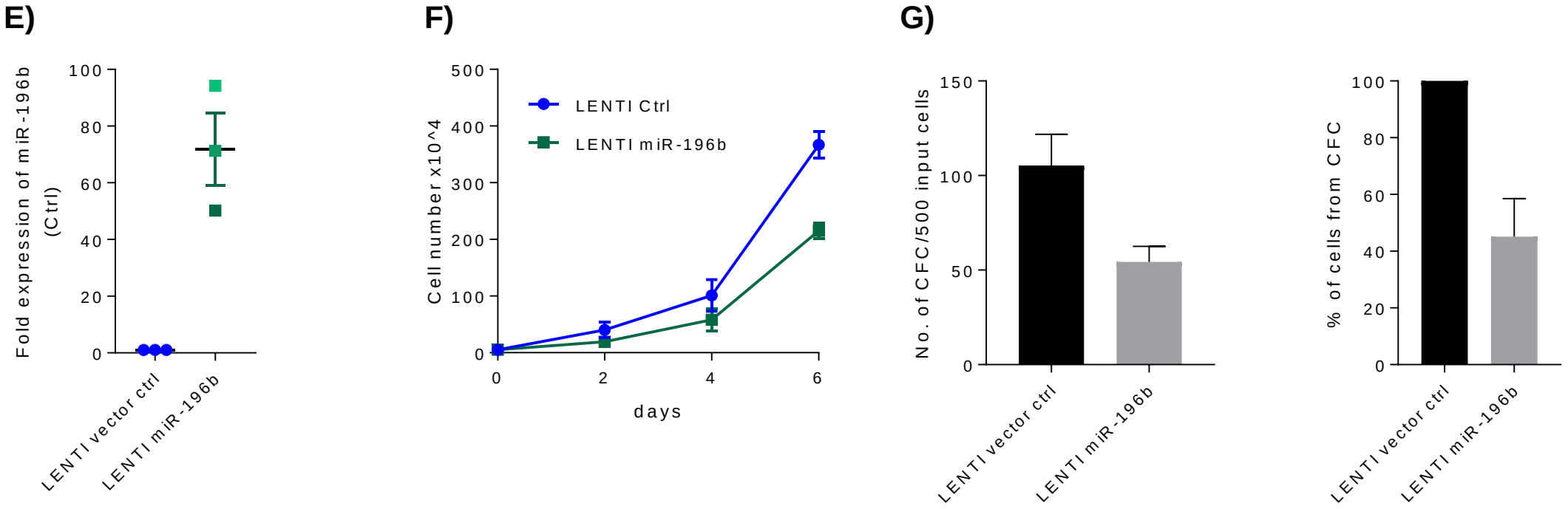
C)



D)

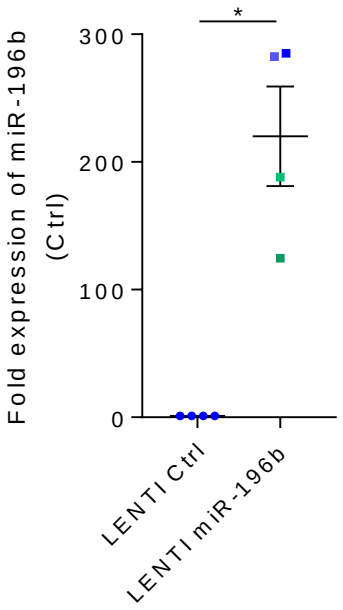


Supplementary Figure S6

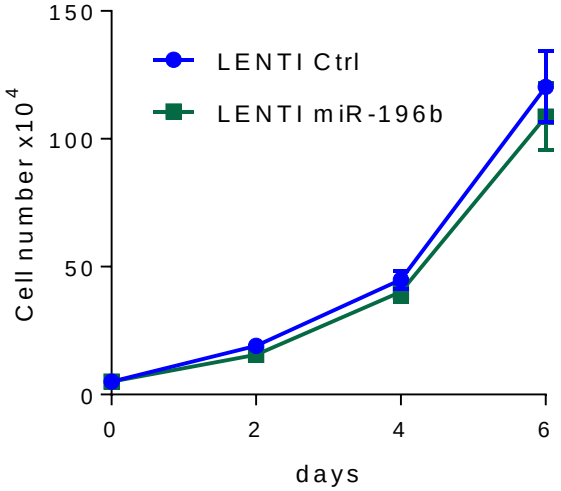


Supplementary Figure S6

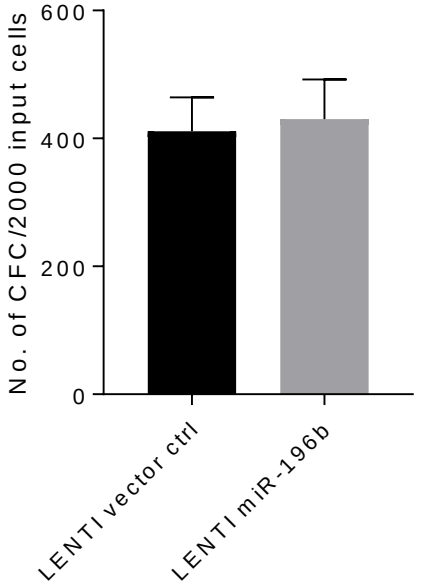
H)

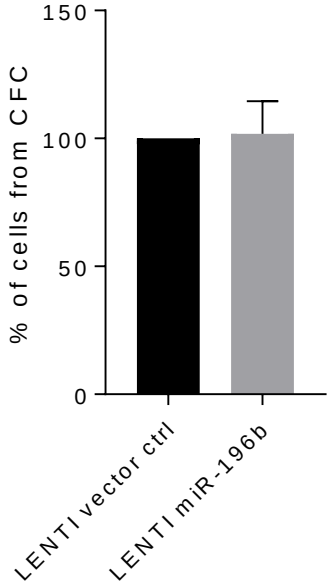


I)



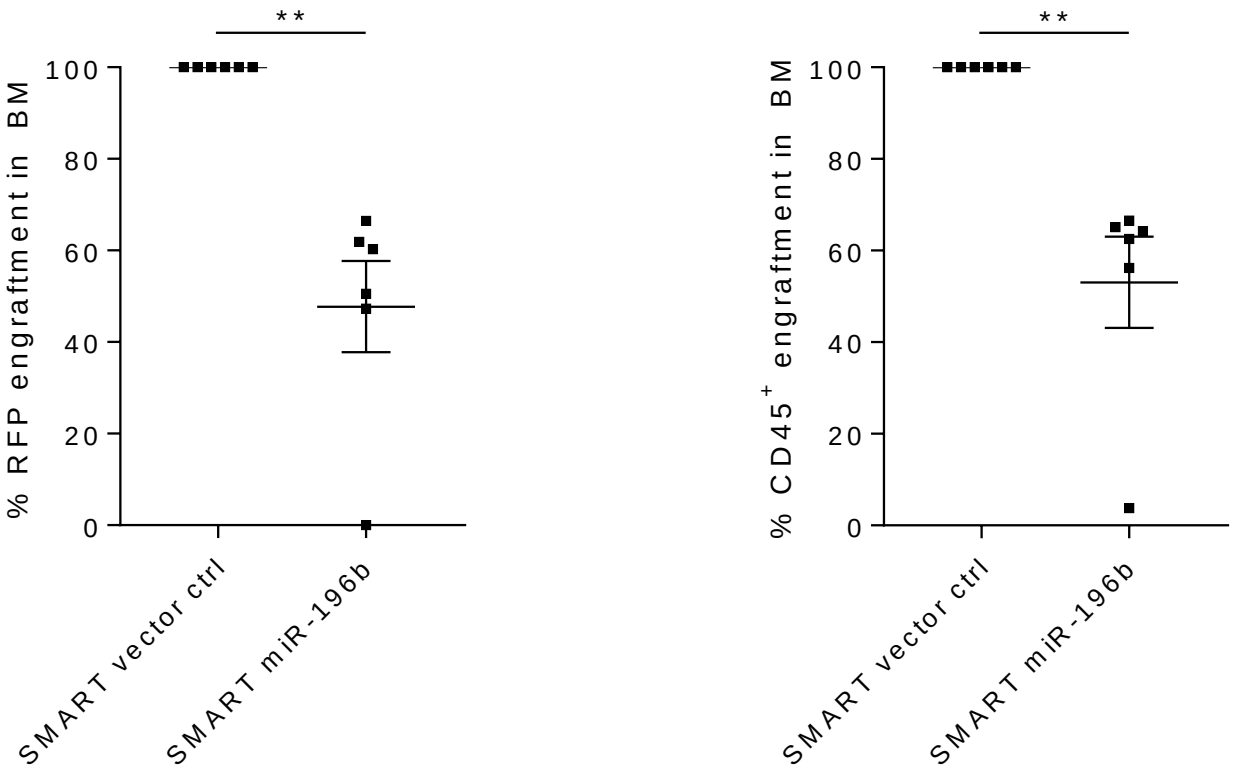
J)



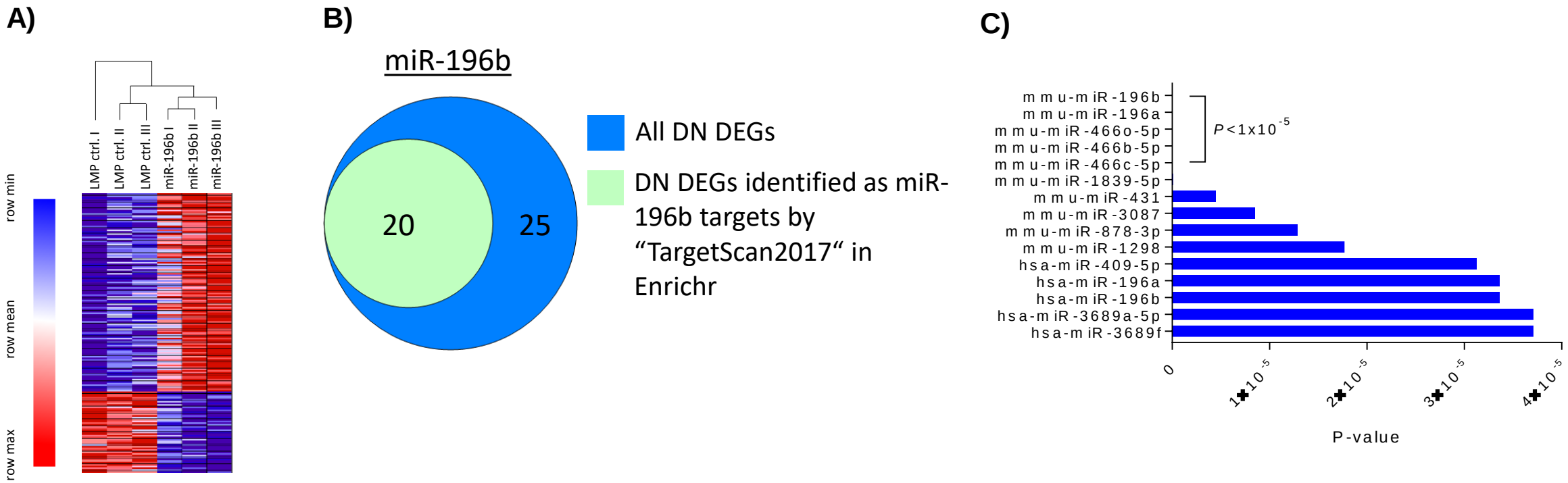


Supplementary Figure S6

K)

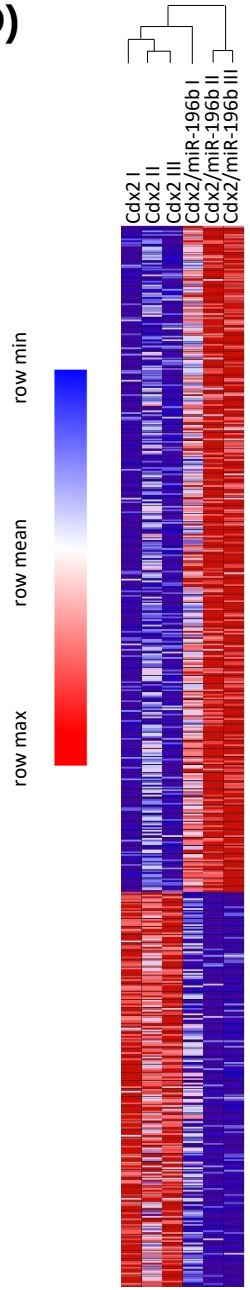


Supplementary Figure S7

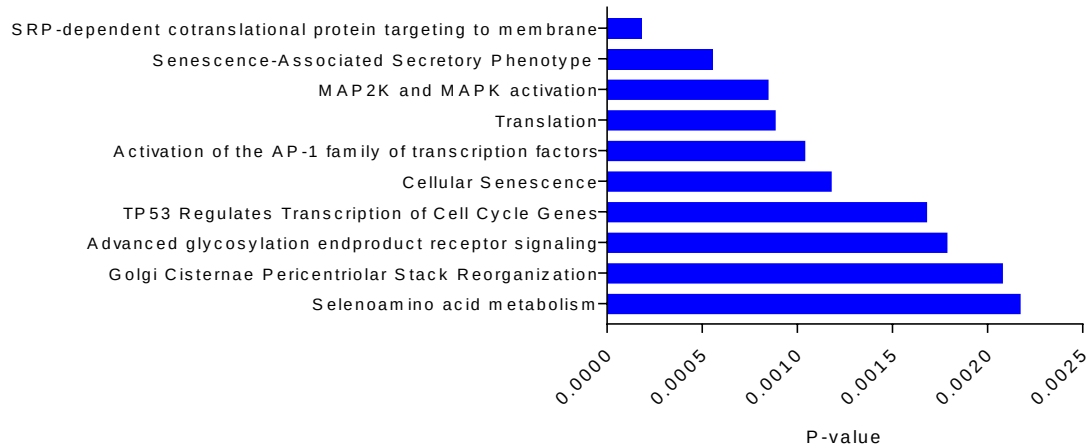


Supplementary Figure S7

D)



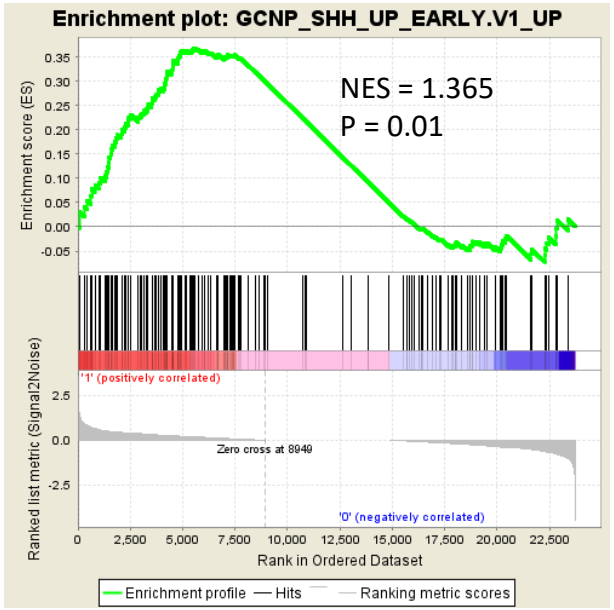
E)



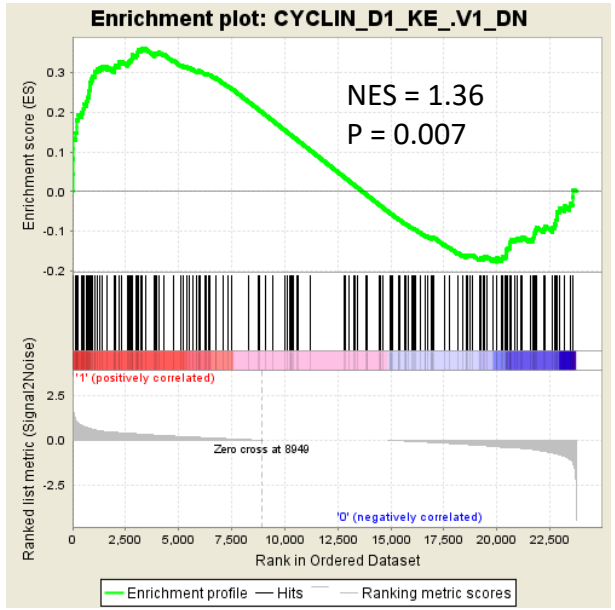
Supplementary Figure S7

F)

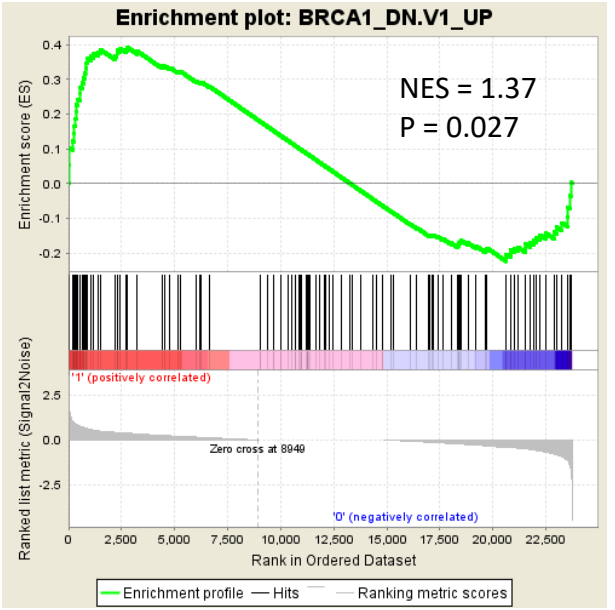
Cdx2 Cdx2/miR-196b

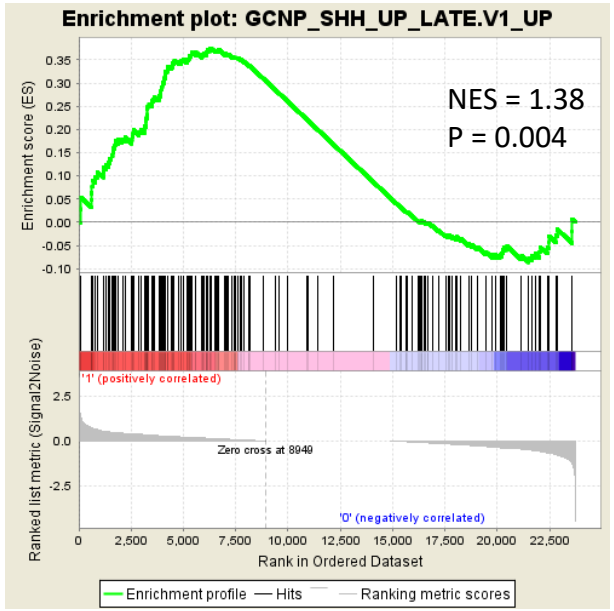


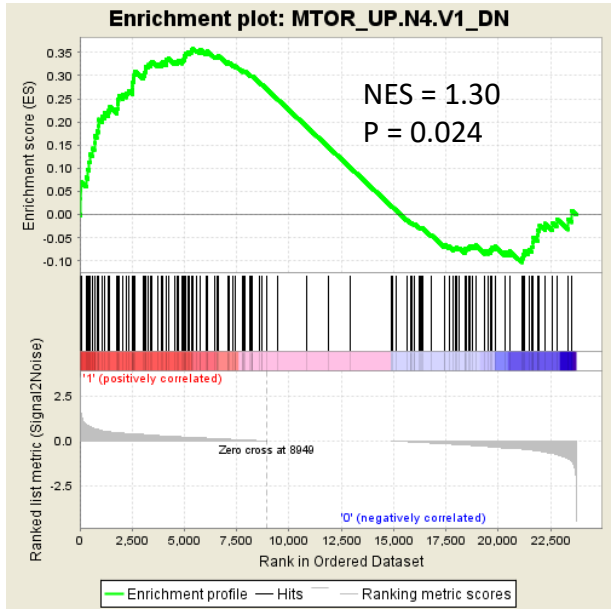
Cdx2 Cdx2/miR-196b

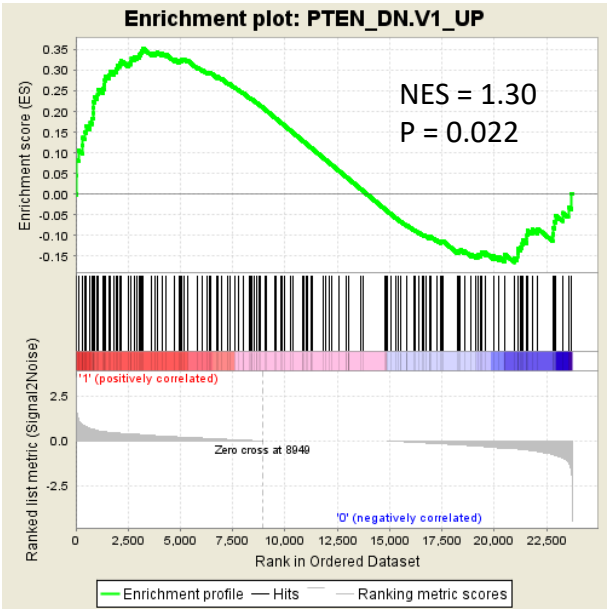


Cdx2 Cdx2/miR-196b



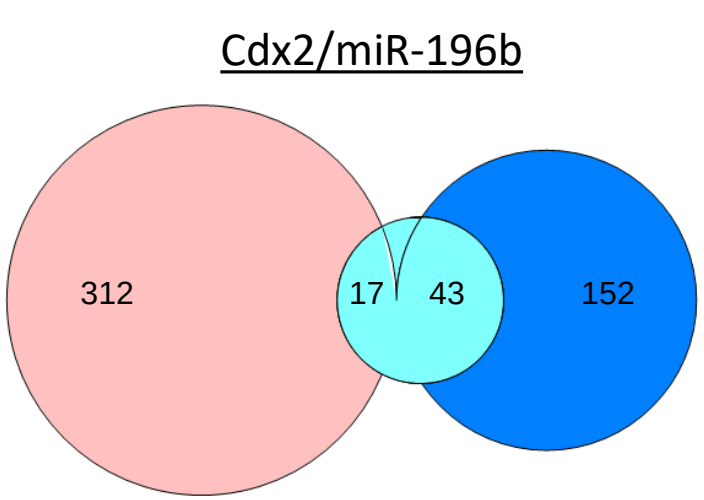






Supplementary Figure S8

A)

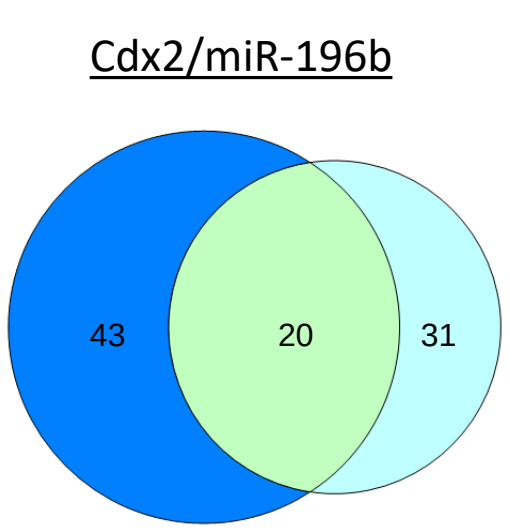


All UP DEGs

All DN DEGs

DEGs identified as miR-196b targets by Enrichr platform

B)



miR-196b DN targets identified by Enrichr platform

miR-196b DN targets identified by a list, comprising miR-196b targets from different online prediction tools

Most relevant miR-196b DN target genes:
Gca, Hoxb8, Nxpe3, Dennd6a, Eps15, Tgfbr3, St6galnac3, Cdyl, Tmx1, Epha7, Ssr1, Hoxa7, Wdr37, Slc4a7, Ing5, Atp8b4, Celf2, Sdcbp, Nme4, Bcat1

Supplementary Figure S8

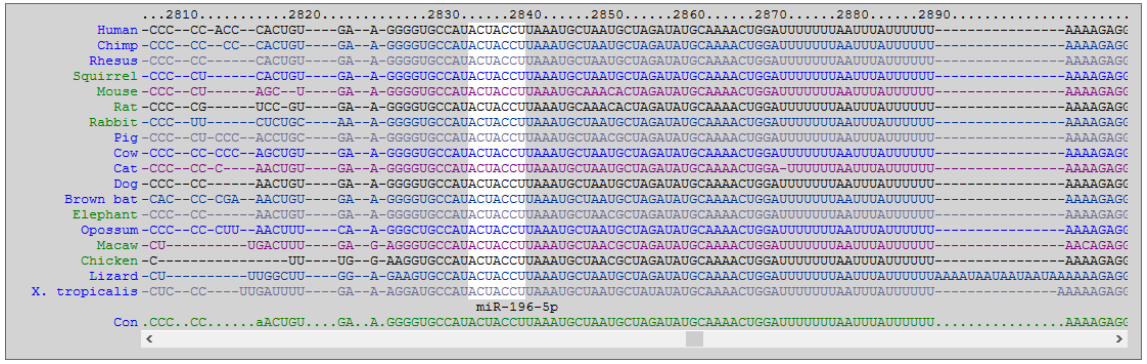
C)

MAPK8

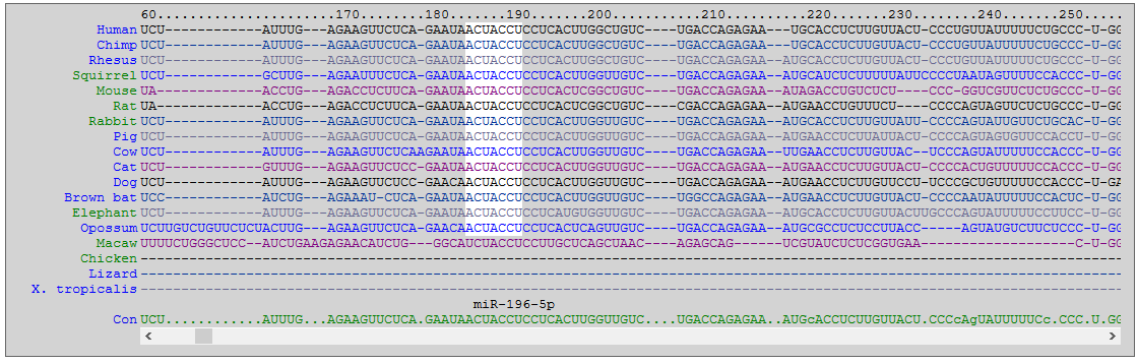
PBX1

MAPK1

MSI2

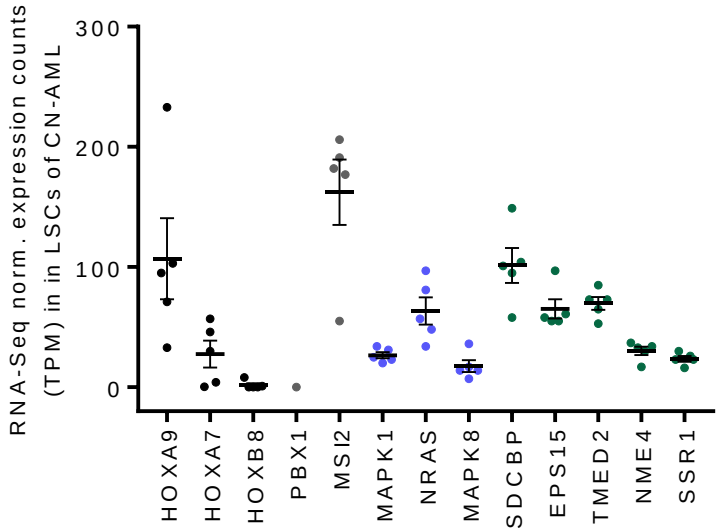


NRAS

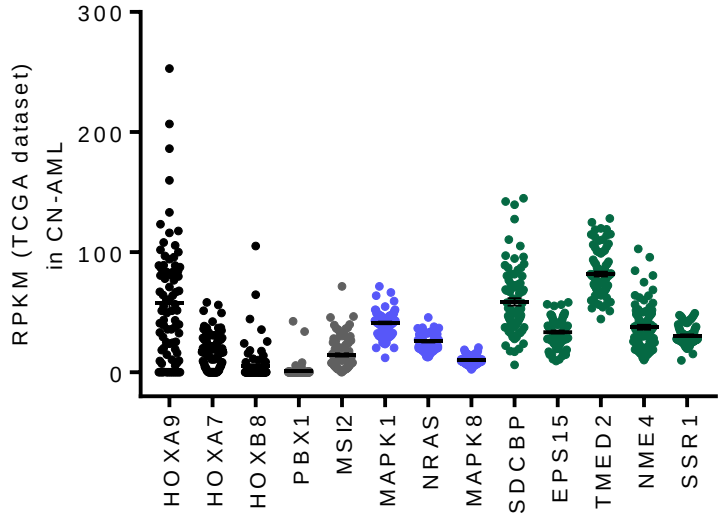


Supplementary Figure S8

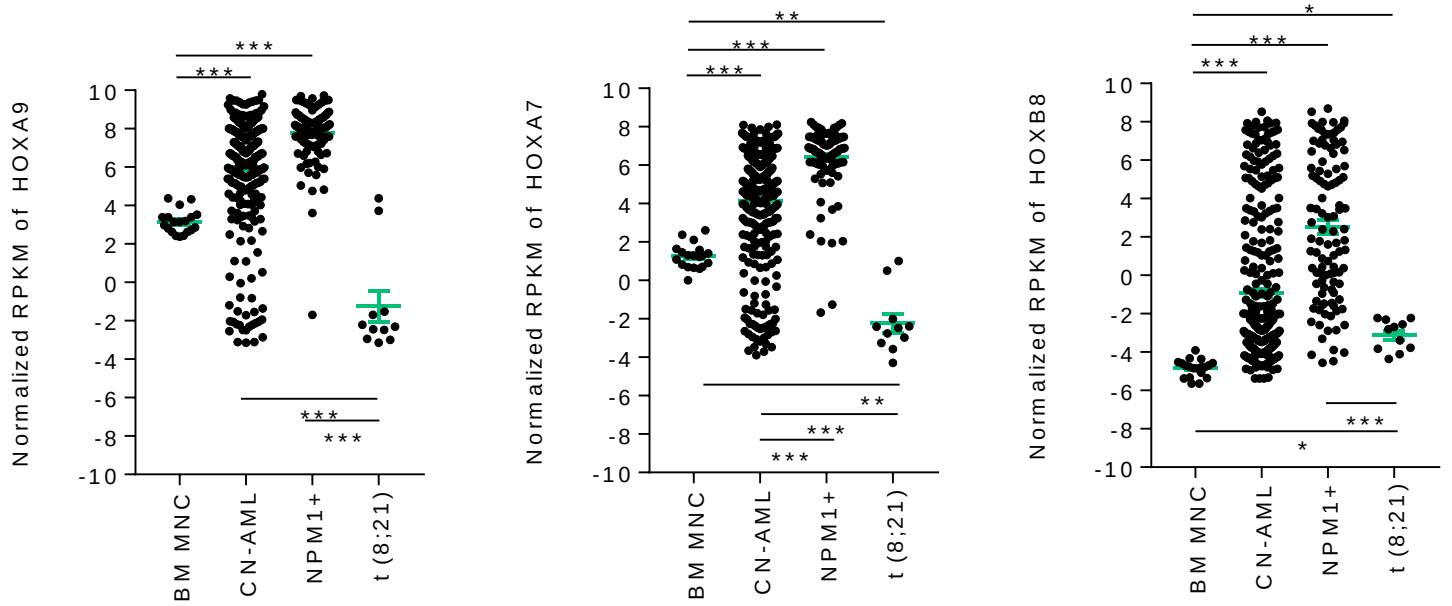
D)



E)



F)



Supplementary Figure S8

G)