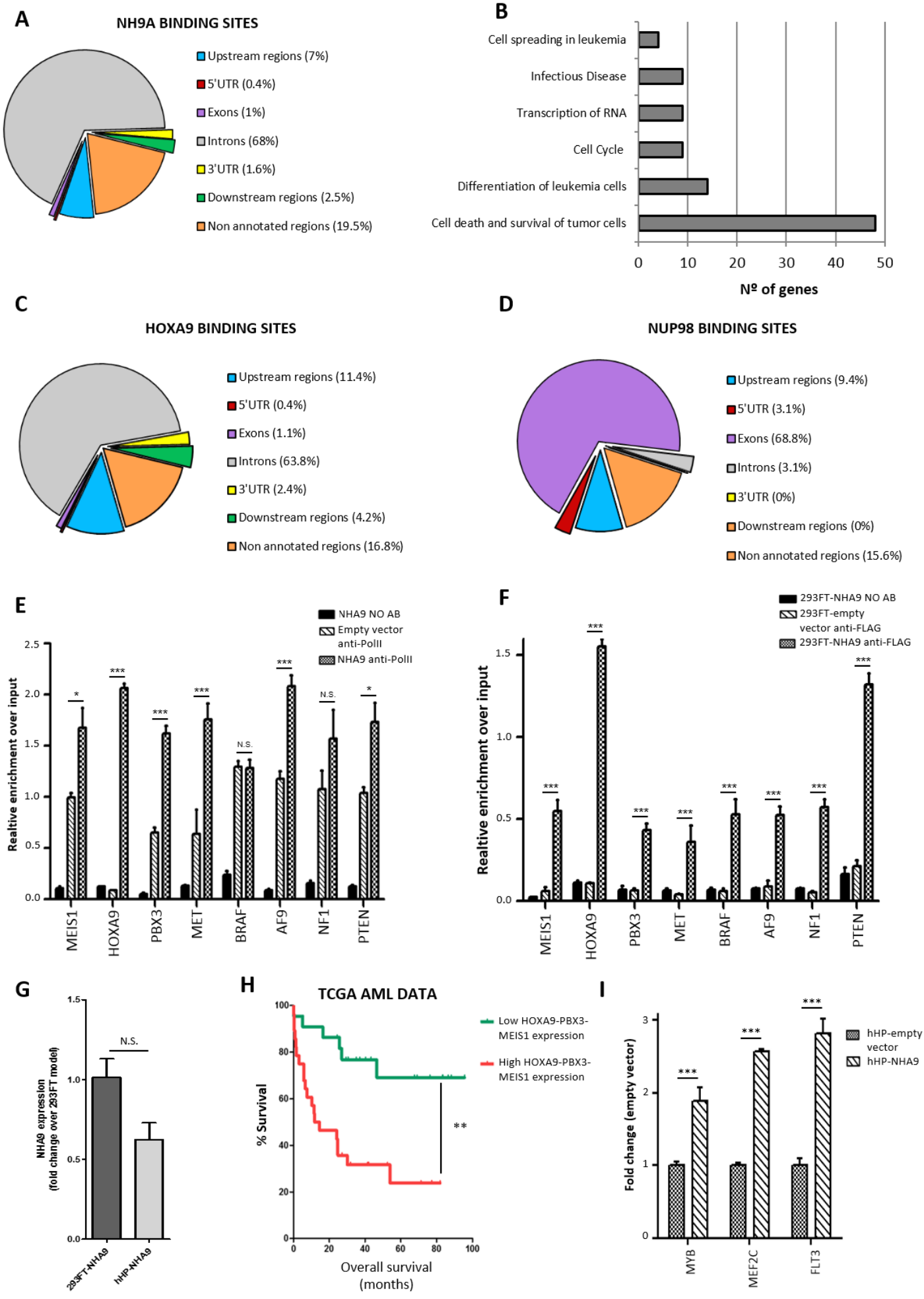
Supplementary information for:

The molecular pathogenesis of the *NUP98-HOXA9* fusion protein in Acute Myeloid Leukemia

Ana Rio-Machin[1,2]; Gonzalo Gomez-Lopez [3]; Javier Muñoz[4]; Fernando Garcia-Martinez[4]; Alba Maiques-Diaz[1]; Sara Alvarez[1]; Rocio N Salgado[1]; Mahesh Shrestha[5]; Raúl Torres[6]; Claudia Haferlach[7]; Maria Jose Larrayoz[8]; Maria Jose Calasanz[8]; Jude Fitzgibbon[2] and Juan C. Cigudosa[1]

[1] Molecular Cytogenetics Group, Human Cancer Genetics Programme, Centro Nacional Investigaciones Oncologicas (CNIO), Madrid, Spain
[2] Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, UK
[3] Bioinformatics Unit, Centro Nacional Investigaciones Oncologicas (CNIO), Madrid, Spain
[4] Proteomics Unit, Centro Nacional Investigaciones Oncologicas (CNIO), Madrid, Spain. ProteoRed-ISCIII.
[5] Division of Experimental Hematology and Cancer Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA
[6] Viral Vector Facility, Fundacion Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain
[7] MLL Münchner Leukämielabor, München, Germany
[8] Department of Genetics and Center for Applied Medical Research (CIMA), University of Navarra, Pamplona, Spain

Supplementary Figure 1



Supplementary Figure 2

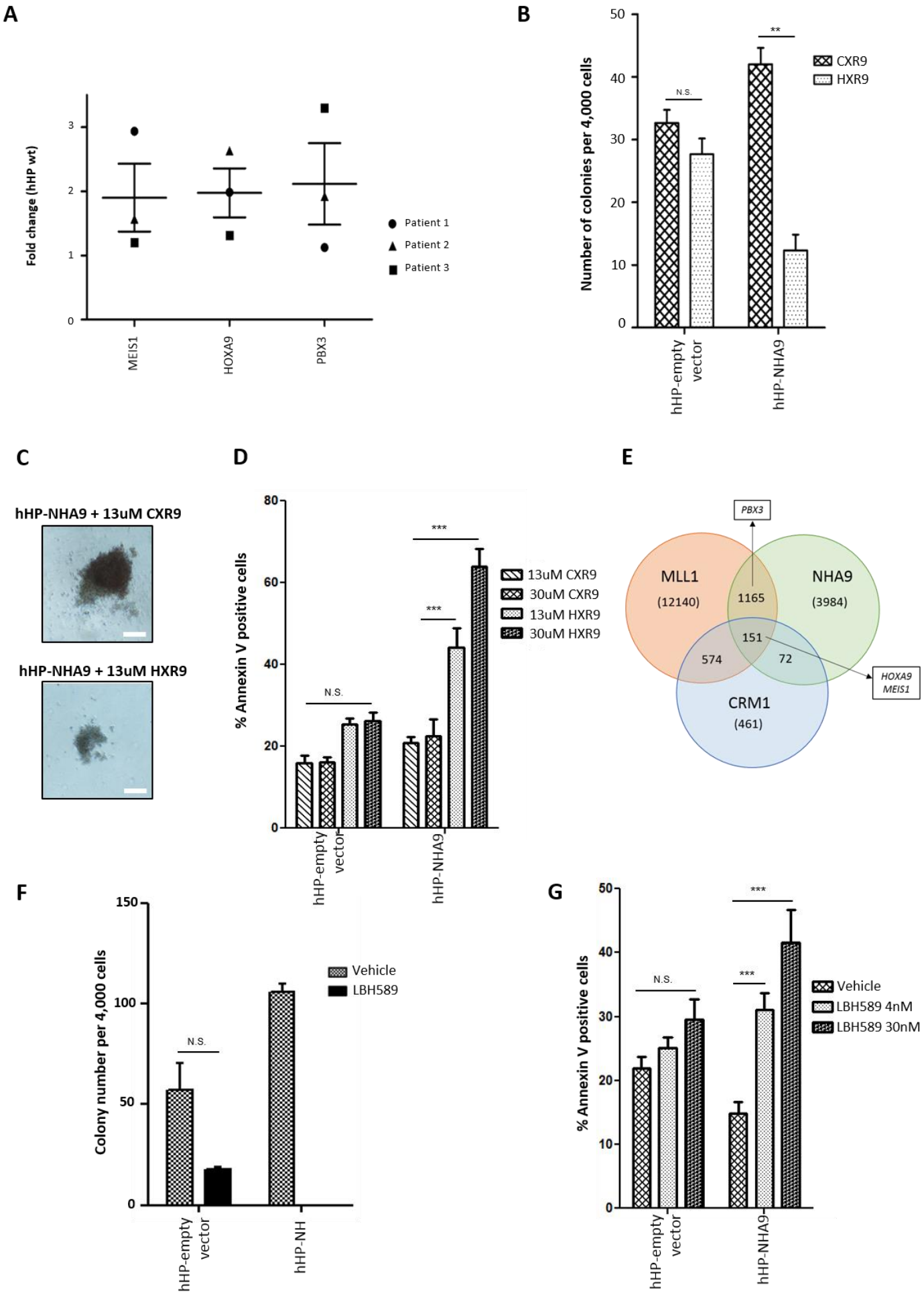


Figure S1.

(A) Relative enrichment levels of NHA9 binding peaks identified by ChIP-seq (+5/-5kb from a TSS) in different genomic regions. We used PAVIS (Huang *et al.*, 2013) tool to determine the relative location relationship between query peaks and annotated genomic loci. **(B)** Main functions of the NHA9 target genes (Ingenuity Pathway Analysis output (http://www.ingenuity.com/products/ipa)) **(C & D)** Localization of HOXA9 (C) and NUP98 (D) binding sites identified by ChIP-seq, (+5/-5kb from a TSS) using PAVIS tool. The identified HOXA9 target regions are located more than 1 kb upstream/downstream of the TSS, whereas NUP98 binding sites were mostly located within promoters, both in agreement with previous studies.[21,22] **(E)** RNA Polymerase II (PolII) qChIP fold enrichment on the eight selected NHA9 target regions. Average of 3 experiments is shown **(F)** NHA9 qChIP fold enrichment on the eight selected NHA9 target enhancer regions in the NHA9 expressing-HEK293FT cells. The average of 3 experiments is shown. Statistical significance for relative enrichment and proliferation was determined at $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***), using a t-test with Bonferroni correction. N.S corresponds to non-significant comparisons. Error bars represent SEM. **(G)** qRT-PCR analysis to compare *NHA9* expression levels in 293FT cells and hHP transduced with pMSCV-NHA9 retroviruses. Non-significant differences in *NHA9* expression levels were detected between both cellular models. **(H)** TCGA samples were classified as High (n = 20, higher quartile) and Low (n = 20, lower quartile) HOXA9-PBX3-MEIS1 expression based on RNA-seq data available. Kaplan–Meier survival analysis was then applied to these groups. Statistical significance was determined by the Log-rank (Mantel-Cox) test ($p = 0.0013$) and Gehan_Breslow-Wilcoxon Test ($p = 0.0016$) **(I)** Expression analysis of three target genes of the complex *MEIS1-HOXA9-PBX3*: *MYB*, *MEF2C* and *FLT3* in the *hHP-NHA9* cellular model. . The expression of the endogenous human housekeeping gene GAPDH was used to normalize the data, which are expressed as the mean of $2^{-\Delta Ct}$ values obtained for each sample. Error bars represent SEM. Statistical significance of relative level of expression was determined using a t-test with Bonferroni correction.

Figure S2.

(A) Expression analysis by qRT-PCR of *MEIS1, HOXA9* and *PBX3* in three patient samples harboring NHA9. The expression of the endogenous human housekeeping gene GAPDH was used to normalize the data, which are expressed as the mean of $2^{-\Delta Ct}$ values obtained for each sample after normalization based on the hHP-empty vector model. Error bars represent SEM. Statistical significance of relative level of expression was determined using a t-test with Bonferroni correction. **(B)** Colony-forming assay with hHP-NHA9 cells and hHP-empty vector cells treated with 13uM of HXR9/CXR9 peptides. The average number of colonies per dish from two independent experiments after 10 days is shown (only colonies with > 50 cells/colony were counted). **(C)** Morphology of colonies of hHP-NHA9 cells treated with HXR9/CXR9 peptides. Scale bar, 100µm **(D)** hHP-NHA9 and hHP-empty vector cell apoptosis assay after the treatment with HXR9/CXR9 peptides for 24 hours. Cells were stained with annexin V and analyzed using flow cytometry. The plot shows the average fold changes of three independent experiments compared to vehicle. Error bars represent SEM. Statistical significance for annexin inmmunostaining levels was determined using a t-test with Bonferroni correction. **(E)** Venn diagram that represents the overlapping of all NHA9 potential target genes identified in our ChIP-seq experiments, based on the closest annotated Transcrption Start Sites (TSS), with both MLL1 and CRM1 target gene lists provided by Xu *et al.* and Oka *et al.* respectively. Significant ChIP-seq peaks were established at FDR ≤ 5%. **(F)** Colony-forming assay with hHP-NHA9 cells and hHP-emtpy vector cells treated with low doses of LBH589 (4nM) showed a complete abrogation of the ability of hHP-NHA9 cells to form colonies The average number of colonies per dish from two independent experiments after ten days is shown (only colonies with > 50 cells/colony were counted). Error bars represent SEM. **(G)** hHP-NHA9 and hHP-empty vector cell apoptosis assay after the treatment with LBH589 peptide for 24 hours, showed a significant apoptosis induction. Cells were stained with annexin V and analyzed using flow cytometry. The plot shows the average fold changes of three independent experiments compared to vehicle. Statistical significance for relative enrichment and proliferation was determined at $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***), using a t-test with Bonferroni correction. N.S corresponds to non-significant comparisons. Error bars represent SEM.

Table S1:
NUP98-HOXA9 target miRNA

NHA9 TARGET miRNA	FDR
hsa-miR-181A2	0.00
hsa-miR-181b	0.00
hsa-miR-586	0.00
hsa-miR-548F1	0.00
hsa-miR-581	0.00
hsa-miR-550-1	0.00
hsa-miR-194-1	0.00
hsa-miR-128-1	0.00
hsa-miR-615	0.00
hsa-miR-297	0.00
hsa-miR-125B2	0.00
hsa-miR-576	0.00
hsa-LET7I	0.00
hsa-miR-181A1	0.01
hsa-miR-1245	0.02
hsa-miR-30A	0.02
hsa-miR-626	0.04

Table S2: NUP98 target genes

NUP98 TARGET GENES*	FDR
HOXA9	
BIRC3	
PBX3	
LMO2	
RARA	
PGK1	
PML	
HOXA7	
BIRC2	
LTBP1	
MAPKAP1	
FCGR2A	
ESCO2	
IGFBP4	
PIAS1	
STOML1	
KCNJ3	

*NUP98 target genes that are common to NHA9 are highlighted in grey

Table S3: Highly enriched motifs in NHA9 target genomic regions

HOX genes

Possible NHA9 binding motive

Motif Found (DREME program)	Known transcription factor	Positive sequences matching the motif	P-value	E-value (score MEME-ChIP)
	Hoxd13 Hoxd10	1013/4471	1.8e-35	1.2e-30
	Hoxa9 Hoxc9 CDX2	330/4471	5.9e-7	0.037
	Hoxa10 Irx3 CDX1	696/4471	3e-12	2e-7
	Unknown	1421/4471	8.5e-24	5.9e-19
	TAL1::GATA1 complex	1450/4471	3.7e-21	2.5e-16
	NFATC2 RELA STAT1	809/4471	9e-8	0.0058
	ARID3A	525/4471	9.4e-13	6.1e-8

Table S4: Overrepresented transcription factors binding sites in the region adjacent to CA/gTTT. MEME-ChIP_SpaMO results. (Whitington *et al.*, 2011)

Transcription Factor	Cluster	e-value	Best gap (bp from primary motif)	Best orientation
HOXB7	UP00206_1	6.15e-4	32	Upstream/opposite strand
ZSCAN4	UP00042_1	1.17e-3	7	Downstream/same strand
HOXD11	HOXD11_DBD_1	1.23e-3	129	All/both palindromic
ZSCAN26	UP00082_2	1.80e-3	54	Upstream/secondary palindromic
ETV2	ETV2_DBD	2.13e-3	125	Upstream/secondary palindromic
POU2F3	UP00179_1	2.14e-3	93	Upstream/secondary palindromic
POU2F4	UP00179_1	2.94e-3	93	Upstream/same strand
ESR1	MA0112.2	3.49e-3	94	Upstream/same strand
HOXD12	HOXD12_DBD_2	4.27e-3	16	Downstream/primary palindromic
HOXC11	HOXD11_DBD_1	5.37e-3	129	All/both palindromic
TFAP2A	TFAP2A_DBD_5	7.24e-3	128	All/both palindromic
TFAP2B	TFAP2A_DBD_5	7.50e-3	128	Upstream/primary palindromic

Table S5: Primers used to clone the NHA9, HOXA9 and NUP98 coding regions into the retroviral vector

CLONING IN pMSCV-IRES-GFP	NUP98-HOXA9	5'-ATGTTTAACAAATCATTTGGAACACCCT-3'
		5'-TCACTCGTCTTTTGCTCGGTCTTTGTTG-3'
	HOXA9	5'-ATGGCCACCACTGGGGCCCTGGGCAACT-3'
		5'-TTACTCGTCTTTTGCTCGGTCTTTGTTG-3'
	NUP98	5'-ATGTTTAACAAATCATTTGGAACACCCT-3'
		5'-TCACTGTCCTTTTTTCTCTACCTGAGGT-3'

Table S6: qChIP primers

			GENOMIC LOCATION (GRCh37)	
	GENE	PRIMER SEQUENCE	Start	End
ENHANCERS	HOXA9 (ENSG00000078399)	5'- GCTTACAATACCTCCTCCATCAA-3' 5'- CTCTAAACCTCAGGCCACATC -3'	chr7:27203007	chr7:27203097
	MEIS1 (ENSG00000143995)	5'- TGCCATTCATACCCTGCTATAC -3' 5'- TTGGAATCCTAGTGGATGTTTCT -3'	chr2:66691633	chr2:66691731
	PBX3 (ENSG00000167081)	5'- GCTACTCCAGGAACACTAAACC -3' 5'- CGGTCACAGTCAGAGAGAAATG -3'	chr9:128607073	chr9:128607215
	MET (ENSG00000105976)	5'- GGGCATTCTGCTCCTGTTAT -3' 5'- ATCCCAAGCGGAATATACTAACC -3'	chr7:116315894	chr7:116316021
	BRAF (ENSG00000157764)	5'- CACAAGGGCACCTGTGAATA -3' 5'- TGGATTGCATCATAACCCTGAA -3'	chr7:140620646	chr7:140620756
	AF9 (ENSG00000171843)	5'- CCTTAGGCACCCACATGTATATT -3' 5'- GCCTCTGATTAGCCTGTGTATG -3'	chr9:20382495	chr9:20382611
	NF1 (ENSG00000196712)	5'- GGTACAGGTCTATATGTGTGTCTAAA -3' 5'- AGAGTCTGCCATTAACCTTGTAA -3'	chr17:29485729	chr17:29485823
	PTEN (ENSG00000171862)	5'- GTACATGACCTTGGACGAGTTAT -3' 5'- CTCCATTCGACCCTCACAAA -3'	chr10:89684650	chr10:89684756
PROMOTERS	HOXA9 (ENSG00000078399)	5'- CGGCACGATCCCTTTACAT -3' 5'- CCCGTCCAGCAGAACAATAA -3'	chr7:27165707	chr7:27165628
	MEIS1 (ENSG00000143995)	5'- TGTAAGACGCGACCTGTTATG -3' 5'- GCGTGTGTAAAGTGTGTGTTG -3'	chr2:66435079	chr2:66435176
	PBX3 (ENSG00000167081)	5'- TCTTTCTTCTCCTCCCTCGT -3' 5'- CAGCATCCTGGATTGATCGT -3'	chr9:125747365	chr9:125747477
	AFF3 (ENSG00000144218)	5'- CTCTTTAGGAGCCACGATGATAC -3' 5'- CTGGGTGTCGACTTCAAACT -3'	chr2:100105546	chr2:100105601
	PTEN (ENSG00000171862)	5'- ATGTGGCGGGACTCTTTATG -3' 5'- GCGGCTCAACTCTCAAACT -3'	chr10:87863559	chr10:87863677
	BIRC3 (ENSG00000023445)	5'- GCAGAGCTTTCCACCATGAA -3' 5'- GGTCAACTTCTCCAGGCTACTA -3'	chr11:102324064	chr11:102324165
	FILIP1L (ENSG00000168386)	5'- TGAAGTCCCTGAAGCAACAA -3' 5'- CTGTCGGGCTAACAAAGTCA -3'	chr3:99876417	chr3:99876493
	SMAD1 (ENSG00000170365)	5'- TGTCTCGTTCCTTTCCCTTTAC -3' 5'- CTAACAAAGACAGGGAGCAGAG -3'	chr4:145482768	chr4:145482873

Table S7: Primers used to clone the enhancer regions of HOXA9, MEIS1 and PBX3 in the luciferase vector.

LUCIFERASE ASSAY CLONING	HOXA9	5’-CACGG TACCGGGTTTCCCAGCTTTTTCTC-3’
		5’-TTCCTCGAGGAGAAGGGGGACAAGAGGAC-3’
	PBX3	5’-CACGGTACCTGTTCTGAGACTGCCCATTG-3’
		5’-TTCCTCGAGA AATTGCTGCACAGGTGAGA-3’
	MEIS1	5’-CACGGTACCCCCCTTTAACTTGAGAATCAGC-3’
		5’-TTCCTCGAGTTCAATGCAGCAGCTGAACTAT-3’3’

Table S8: RT-qPCR primers

HOXA9	5’-GTGGTTCTCCTCCAGTTGATAG-3’
	5’-AGTTGGCTGCTGGGTTATT-3’
MEIS1	5’-CTAACACACCCTTACCCTTCTG-3’
	5’-TCTATCATGGGCTGCACTATTC-3’
PBX3	5’-CATCACAGTGTCACAGGTATCC-3’
	5’-CAGCATAGAGGTTGGCTTCTT-3’
MYB	5’-TGTAGGGAGAGGTGGCATAA-3’
	5’-GTCTCTTGTGTGCCTGGTAAA-3’
MEF2C	5’-GCAACAGCAACACCTACATAA -3’
	5’-GTAGAAGGCAGGGAGAGATTTG-3’
HOXA7	5’-AGTTCCACTTCAACCGCTAC-3’
	5’-CCTTCGTCCTTATGCTCTTTCT-3’
BIRC3	5’-CAAGCCAGTTACCCTCATCTAC-3’
	5’-CTGAATGGTCTTCTCCAGGTTC-3’
BIRC2	5’-CTAGTCTGGGATCCACCTCTAA-3’
	5’-TGTTCCAAGGTGGGAGATAATG-3’
LMO2	5’-GATGACAATGCGGGTGAAAG-3’
	5’-GTCAGAGTTGATGAGGAGGTATC-3’
RARA	5’-CGAACAACAGCTCAGAACAAC-3’
	5’-GGCGAACTCCACAGTCTTAAT-3’
PML	5’-TGTACGCCTTCTCCATCAAAG-3’
	5’-GACTCCATCTTGATGACCTTCC-3’

SUPPLEMENTARY METHODS

CELLULAR MODELS

The *NUP98-HOXA9* coding region sequence was PCR amplified from a patient cDNA, the *HOXA9* coding region from a control cDNA, and the *NUP98* coding region from the vector pMP250. These three coding regions were cloned into pMSCV-IRES-GFP retroviral vectors. The primers used for cloning are listed in Table S5. For all three we included a C-terminal FLAG as a tag. Retroviral vectors pMSCV-MLL-AF9 and pMSCV-AML1-ETO were kindly provided by James Mulloy's lab. Retrovirus for each construct were produced in HEK293T cells by co-transfecting viral plasmids along with packaging plasmid using the calcium-phosphate protocol, as described previously (Wunderlich *et al.,* 2009). Constitutively expressing cellular models were generated by retroviral transduction of HEK93FT cell line and cord blood-isolated hHP. GFP-positive cells were sorted using a BD Influx™ cell sorter (BD Biosciences, San Jose, CA, USA). Human Hematopoietic Progenitors (hHP; CD34+) were purified using immunomagnetic beads (Miltenyi Biotech, Bergisch Gladbach, Germany) from human umbilical cord blood (CB). They were cultured in Iscove's modified Dulbecco's media (IMDM) containing 20% BIT 9500 Serum substitute (StemCell Technologies, Vancouver, British Columbia, Canada), 1% penicillin-streptomycin (Invitrogen-Thermo Fisher Scientific, Waltham, Massachusetts, USA) 0.2% β-mercaptoethanol, 10 ng/ml SCF, megakaryocyte growth and development factor (TPO), FLT3 ligand (FLT3L), IL-3, and IL-6, as described previously (Wunderlich *et al.,* 2009). HEK293T and HEK293FT cells were cultured in Dulbecco's modified Eagle's medium (DMEM), supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin (Invitrogen).

CHROMATIN IMMUNPRECIPITATION-DEEP SEQUENCING (ChIP-seq)

ChIP was performed as previously described (Goyama *et al.,* 2013) using the antibodies anti-FLAG M2 (F1804, Sigma-Aldrich, Saint Louis, Missouri , USA) and anti-HA tag antibody-ChIP Grade (ab9110, abcam, Cambridge, UK). Libraries were constructed using the ChIP-seq sample preparation kit (Illumina, San Diego, California, USA) and were sequenced in a Genome Analyzer IIx (GA2, Illumina, San Diego, California, USA) single 36-base read run. Alignment to the human sequence assembly (GRCh37/hg19, Feb 2009) was performed using BWA with default settings, permitting alignments with 1 mismatch in 40 base reads. Peak detection was performed using MACS 2.0 following the developer's technical recommendations (p-value cutoff for peak detection = 10^{-5}). The MACS pipeline was used to identify differential peaks from pair-wise comparisons. Significant ChIP-seq peaks were established at FDR ≤ 5%. (Data deposited in GEO http://www.ncbi.nlm.nih.gov/geo/, accession number: GSE62587). Quantitative ChIP (qChIP) was performed by ChIP followed by quantitative PCR (qChIP) using SYBR Green dye (Applied Biosystems, Foster City, CA, USA). The primers used are listed in Table S6. Enhancer regions were validated by qChIP using anti-H3K4me1 antibody (ab8895, abcam, Cambridge, UK) and anti-RNA Polymerase II antibody (ab26721, abcam, Cambridge, UK). Binding of p300 and HDAC1 was analysed using *anti-p300* (sc-585, Santa Cruz Biotechnology, Dallas, Texas, USA) and *anti-HDAC1* (PA1-860, Thermo Fisher Scientific, Waltham, Massachusetts, USA) antibodies.

LUCIFERASE ASSAYS

Following amplification of the three selected genomic enhancer regions of *HOXA9, PBX3* and *MEIS1,* luciferase constructs were made by subcloning in the *pGL3-Promoter* vector (Promega, Fitchburg, Wisconsin, USA). PCR products were obtained using the primers listed in Table S7. The following specific bacterial artificial chromosomes (BACs) were used as a PCR template: RP11–197I05 (*HOXA9*), RP11-638B02 (*PBX3*) and RP11-678O18 (*MEIS1*). Luciferase constructs containing the enhancer region of *HOXA9, PBX3* and *MEIS1* were co-transfected into HEK293FT cells with the expression vector pMSCV-NHA9, together with Renilla vector for the purpose of normalization. Luciferase activity was determined 48 h after reporter plasmid transfection in all cases using the *Dual Luciferase Reporter Assay System* (Promega, Fitchburg, Wisconsin, USA) as previously described (Rio-Machin *et al.,* 2013). A significant increase in luciferase activity induced by NHA9 expression was observed in each case. Data are presented as the mean value from two separate experiments with n=3 for each experiment. Error bars represent SEM.

GENE ESPRESSION ANALYSIS

Total RNA was extracted using the miRNeasy Mini Kit (Qiagen, Valencia, CA, USA). TaqMan® Gold RT-PCR Kit (Applied Biosystems, Foster City, CA, USA) was used for the reverse-transcription.

RNAs from three different clones of the *hHP-NHA9* cellular model and primary samples from NHA9 patients were hybridized using Array SurePrint G3 Human Gene Expression 8x60K v2 (Agilent Technologies, Santa Clara, California, USA). Microarray background subtraction was carried out using the *normexp* method. Loess normalization was applied within arrays and quantile normalization between arrays. Differentially expressed genes were identified by applying linear models using the limma package in R (Smyth *et al.,* 2005. Bioconductor project, http://www.bioconductor.org). RT-qPCR was performed on 384-well plates, with each cDNA included in triplicate, using quantitative PCR with SYBR Green dyes (Applied Biosystems, Foster City, CA, USA) on a 7500 Fast Real-Time PCR System. The primers used are listed in Table S8.

DRUG ASSAYS

NHA9 expressing-hHP cells were exposed to *HXR9* (and *CXR9* as control) (synthesized by Biosynthesis, http://www.biosyn.com/) and Panobinostat (*LB589,* purchased from Selleck Chemicals, Houston, Texas, USA). Cell viability was assessed in triplicate with different doses of drug and at different time points using WST-1 *cell proliferation reagent* (Roche, Basel, Switzerland), according to the manufacturer's instructions. Aliquots of 8 x 10^3 sorted cells were placed in 35 mm culture dishes in 1.5 ml of MethoCult M4230 (StemCell Technologies, Vancouver, British Columbia, Canada) containing 20ng/ul of each of human recombinant SCF, IL-3, IL-6 and GM-CSF, and 3U/ml of EPO for colony-forming assays, as described previously (Li *et al.,* 2013). For apoptosis assay after treatment, cells were stained with FITC annexin V (BD Biosciences) and DAPI and analyzed on FACSCanto II (BD Biosciences, San Jose, CA, USA) as described previously (Li *et al.,* 2013).

CO-IMMUNOPRECIPITATION ANALYSIS

Co-IP with NHA9 was performed for *p300 and HDAC1* in NHA9 expressing-HEK293FT cells, as previously described (Bunda *et al.,* 2013). In accordance with the manufacturer's instructions, antibodies were added to the protein supernatants: *anti-p300* (Santa Cruz) and *anti-HDAC1* (Thermo Scientific). Input lysates and immunoprecipitated proteins were loaded onto SDS-PAGE gels (9% gel) and then immunoblotted. Following transfer, PVDF membranes (Millipore, Billerica, Massachusetts, USA) were incubated with *anti-FLAG M2* (Sigma-Aldrich, San Luis, Misuri, USA) primary antibody to detect NHA9, with *anti-GAPDH* (monoclonal antibodies, CNIO) as an endogenous control.

STATISTICS

Statistical significance for relative enrichment in qChIP experiments, level of expression in qRT-PCRs, luciferase activity, relative cell viability, annexin positive cells and number of colonies was determined at $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***), calculated using a t-test with Bonferroni correction**.** Data are expressed as mean ± standard error of the mean (SEM).

MASS SPECTROMETRY ANALYSIS

We used the 293FT-NH, 293FT-HOXA9 and 293FT-NUP98 models to conduct the corresponding immunopreticiptations using anti-FALG M2 antibody (Sigma Aldrich), followed by mass spectrometry. Two (NH and HOXA9) and three (NUP98) independent experiments (i.e. biological replicates) were performed. As negative controls, we processed in parallel cell cultures transduced with the retroviral empty vector flanked by FLAG-tag (Figure S3A).

Sample Preparation: Proteins in the pull-downs were subjected to label free proteome analysis. Samples were digested by means of the standard FASP protocol (Wiśniewski *et al.,* 2009). Briefly, samples were resuspended in UT buffer (8M urea in 100 Mm Tris-HCl, pH=8.01). Proteins were then reduced with 10 mM DTT, alkylated using 50 mM iodoacetamide for 20 min in the dark and the excess of reagents was washed out with UA twice. Proteins were digested with endoproteinase Lys-C (Wako, Richmond, VA, USA) during 6 hours in a wet chamber (1:50 enzyme to substrate ratio). Finally, samples were diluted in 50 mM ammonium bicarbonate to reduce the urea concentration to 1M and subsequently digested with Trypsin Gold (Promega, Fitchburg, Wisconsin, USA) overnight at 37 °C. Resulting peptides were further desalted and concentrated using homemade reversed phase micro-columns filled with Poros Oligo R3 beads (Life Technologies-Thermo Fisher Scientific, Waltham, Massachusetts, USA). The samples were dried using the Speed-Vac and dissolved in 30 µL of 0.1% formic acid (FA).

LC-MS/MS analysis: Desalted peptides were separated by reversed-phase chromatography using a nanoLC Ultra system (Eksigent Technologies, Dublin, CA, USA), directly coupled with a LTQ-Orbitrap Velos instrument (Thermo Fisher Scientific, Waltham, Massachusetts, USA) via nanoelectrospray source (ProxeonBiosystem). Peptides were loaded onto the column (Dr. Maisch, Reprosil-Pur C18 GmbH 3 □m, 200x0.075 mm), with a previous trapping column step (Prot Trap Column 0.3 x 10 mm, ReproSil C18-AQ, 5 µm, 120Å, SGE), during 10 min with a flow rate of 2.5 □l/min of loading buffer (0.1% FA). Elution from the column was made with a 120 min linear gradient (buffer A: 2% ACN, 0.1%FA; buffer B: 100% ACN, 0.1%FA) at 300 nL/min. The peptides were directly electrosprayed into the mass spectrometer using a PicoTip emitter (360/20 OD/ID µm tip ID 10 µm, New Objective) a 1.4 kV spray voltage with a heated capillary temperature of 325°C and S-Lens of 60%. Mass spectra were acquired in a data-dependent manner, with an automatic switch between MS and MS/MS scans using a top 20 method with a threshold signal of 800 counts. MS spectra were acquired with a resolution of 60000 (FWHM) at 400 m/z in the Orbitrap, scanning a mass range between 350 and 1500 *m/z*. Peptide fragmentation was performed using collision induced dissociation (CID) and fragment ions were detected in the linear ion trap. The normalized collision energy was set to 35%, the Q value to 0.25 and the activation time to 10 ms. The maximum ion injection times for the survey scan and the MS/MS scans were 500 ms and 100 ms respectively and the ion target values were set to 1E6 and 5000, respectively for each scan mode.

Data analysis: Raw files were analyzed with MaxQuant (Cox *et al.,* 2008) against a human canonical Uniprot database (20,187 sequences) supplemented with common contaminants and the recombinant proteins used in this study. Oxidation of methionines and protein N-terminal acetylation were set as variable modifications whereas carbamidomethylation of cysteines was considered as fixed modification. A maximum of two missed-cleavages were allowed in the Andromeda search engine. Other parameters were set to default values. Results were filtered to 1% FDR using a target-decoy approach. Quantification was performed with the LFQ algorithm (Cox *et al.,* 2014). To define a list of specific interactors, MaxQuant output tables (*i.e.* proteingroups.txt) were processed with Perseus for subsequent statistical analysis. A Two-sample T-test was performed in which the protein LFQ values of baits and empty vectors were compared. Statistical results were filtered to FDR=5% (number of permutations = 250). In addition, only those proteins found to be significant and with a T-test Difference > 1 were considered.

Alternatively, data were processed with Proteome Discoverer (Thermo, version 1.4) against a human Uniprot database supplemented with common contaminants and the recombinant proteins used in this study (25,578 sequences). Oxidation of methionines was set as variable modification whereas carbamidomethylation of cysteines was considered as fixed modification. Precursor tolerance was set to 10 ppm and fragment tolerance to 0.5 Da; a maximum of two missed-cleavages were allowed in the Sequest HT search engine. Results were validated with Percolator using a target-decoy approach and filtered to 1% FDR. Other filters included minimal peptide length of 6 amino acids and rank 1 peptides. In case that identified peptides were shared by two or more proteins (homologs or isoforms), they were reported by Proteome Discoverer as one protein group. Further analyses were performed with Microsoft Office Excel. To define a list of proteins of specific interactors for each of the three baits the following criteria were used: was (i) the protein was not identified in any of the empty vectors and (ii) the protein was identified in at least two replicates of the baits.

Proteomics data have been deposited on the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (data set identifier PXD001828)

STUDY APPROVAL

Informed consent was received from participants prior to inclusion in the study according to the Declaration of Helsinki (October 2008). Study protocol for primary samples was approved by the Research Ethics Committee of the Spanish National Institute of Health (CEI-ISCIII).

REFERENCES FOR SUPPLEMENTARY FIGURES AND METHODS:

- Bunda S, Kommaraju K, Heir P and Ohh M. SOCS-1 mediates ubiquitylation and degradation of GM-CSF receptor. *PLoS One* 2013; 8:e76370.
- Cox J. & Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* 2008; 26, 1367–1372.
- Cox J *et al.* MaxLFQ allows accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction. *Molecular & Cellular Proteomics* 2014; M113.031591.
- Goyama S, Schibler J, Cunningham L, Zhang Y, Rao Y, Nishimoto N *et al*. Transcription factor RUNX1 promotes survival of acute myeloid leukemia cells. *Journal of Clinical Investigation* 2013; 123:3876-3888.
- Huang W, Loganantharaj R, Schroeder B, Fargo D and Li L. PAVIS: a tool for Peak Annotation and Visualization. *Bioinformatics* 2013; 29(23):3097-9.
- Li Z, Zhang Z, Li Y, Arnovitz S, Chen P, Huang H *et al*. PBX3 is an important cofactor of HOXA9 in leukemogenesis. *Blood* 2013; 121:1422-1431.
- Rio-Machin A, Ferreira BI, Henry T, Gomez-Lopez G, Agirre X, Alvarez S *et al*. Downregulation of specific miRNAs in hyperdiploid multiple myeloma mimics the oncogenic effect of IgH translocations occurring in the non-hyperdiploid subtype. *Leukemia* 2013; 27:925-931.
- Smyth GK, Michaud J and Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 2005; 21:2067-2075 Smyth GK, Michaud J and Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 2005; 21:2067-2075.
- Whitington T, Frit MC, Johnson J and Bailey TL. Inferring transcription factor complexes from ChIP-seq data. *Nucleic Acids Research* 2011; 39(15):e98.
- Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nature Methods* 2009; 359–362.
- Wunderlich M, Mulloy JC. Model systems for examining effects of leukemia-associated oncogenes in primary human CD34+ cells via retroviral transduction. *Methods in Molecular Biology* 2009; 538:263-285.