Supplementary Methods

Detailed methods are provided in the online version of this paper and include the following

Supplementary-Reagents-Table:

CRISPR gRNA sequences

shRNA sequences

Primers

Antibodies

Kits

Chemical inhibitors

Cell Lines

MOLM13, MV411, HL60 cells were cultured in RPMI1640 medium containing 10%FBS, 1%L-Glutamine and 1%PenStrep. OCIAML3 cells were cultured in MEMα medium containing 20%FBS, 1%L-Glutamine and 1%PenStrep. Cas9 expressing OCIAML3 and HL60 we received from Prof. George Vassiliou, University of Cambridge. MOLM13 and MV411 cells with stable Cas9 expression were generated by using lentiCas9-Blast vector (Addgene#52962). Single cell clones were obtained by culturing cells in methylcellulose media followed by picking single colonies and established in liquid culture. Cas9 activity and expression was confirmed in multiple clones independently.

293T cells were cultured in DMEM medium containing 10%FBS, 1%L-Glutamine and 1%PenStrep. All cell lines were routinely tested negative for mycoplasma contamination.

For treatment with inhibitors Panobinostat-LBH589 (Cat # S1030, Selleckchem) and Chaetocin (Cat # HY-N2019, MedChemExpress), cells were seeded in 6 well plate at density of $5X10^5$/ml in 3 ml media. Inhibitors were diluted to working stock of 10mM with DMSO and were used at final concentration of 4nM Panobinostat and 40nM Chaetocin. Cells were harvested for assays at specific time points.

Primary AML sample assays

Human AML mononuclear cells (blood or bone marrow) were provided by the Cambridge Blood and Stem Cell Biobank (LREC 18/EE/0199). Patients gave written informed consent, and research was performed in accordance with the Declaration of Helsinki under local ethical approval. Suspension cultures (~0.5 million cells/ml) were maintained in StemSpan medium (Cat#09605, Stem Cell technologies), supplemented with 1xStemSpan CD34+ Expansion Supplement (Cat# 02691, Stem Cell technologies), 1uM of UM729 (Cat#72332, Stem Cell technologies) and 1%PenStrep. Inhibitors were used at a final concentration of 4nM Panobinostat and 40nM Chaetocin in DMSO. Viability was assessed using AnnexinV/7-AAD (Cat#640922, Biolegend) as per the manufacturer's recommendation. Flow cytometry was performed on a BD LSRFortessa cell analyzer, and all data were analyzed with FlowJo software (Tree Star). Cells were also used for DuoLink, CUT&RUN and RT-QPCR, these methods are described in the relevant sections.

Cloning of sgRNA into lentiviral vector

The chosen sgRNAs sequences were designed with overhang sequence for BbsI restriction sites. Annealing of the sense 1 µl (100 pmol/µl) and antisense 1 µl (100 pmol/µl) gRNA oligonucleotides for each gRNA was performed in 10x NEB buffer 2: 2 µl in a final volume of 20 µl, with denaturation at 95 °C for 5 min in a heating block and slowly allowed to anneal oligonucleotides until mixture is at room temperature (which takes approximately 1-2 h). The annealed gRNA oligonucleotides can be store at -20 °C for later use or used for ligation, all in one tube by assembling reaction in a PCR tube:

i. 100 ng pkLV2 gRNA expression plasmid

ii. 1 µl of annealed gRNA oligonucleotide

iii. 1 µl BbsI restriction enzyme

iv. 1 µl T4 DNA ligase.

v. 2 µl 10x T4 ligase buffer (to a final concentration of 1x).

vi. Nuclease-free water up to 20 µl total reaction volume.

The reaction was incubated in a thermal cycler for 10 cycles for [5 min at 37 °C, 10 min at 22 °C], hold for 30 min at 37 °C followed by 15 min at 75 °C, then hold 4°C.
Competent *E. coli* bacterial was transformed using 2 μl of the ligation product and colonies were sequenced to confirm the clones using pKLV2 sequencing primer AGATAATTAGAATTAATTTGACTG.

Cloning cDNAs into lentiviral vector

NOTCH-ICN-GFP retroviral construct was a kind gift from Dr. Pieter Van Vlierberghe from Ghent University.
Human CEBPD cDNA was ordered from GenScript (NM_005195.4) ORF clone (Catalog no. OHu17203D). Primers were designed to PCR out the CEBPD cDNA with overhangs to clone with EcoRI and BamHI sites into pLVX-TetOnePuro vector (Clontech). Clones were confirmed by sanger sequencing.

Cloning shRNAs into lentiviral vector (pLKO-TetOn)

shRNA against human HOXA9 and SAFB were cloned into pLKO “all-in-one” system for the inducible shRNA expression as described earlier (Wee et al., 2008; Wiederschain et al., 2009).

***In vivo* transplantation of human leukemia cells**

MOLM13 cells were transduced with lentiviruses expressing the puromycin resistance gene and a doxycycline-inducible shRNA against *HOXA9* or *SAFB* or a control (scrambled) shRNA. Cells were selected by 2μg/ml puromycin. Knockdown was confirmed by RT-QPCR and western blotting. The selected shRNA stable cells were further engineered to express luciferase for bioluminescence studies. GFP positive cells were sorted and maintained as stable sh-luciferase-MOLM13 cells. For *in vivo* studies, 100,000 cells were transplanted via tail vein injection into sub-lethally irradiated (2Gy) 6-8 weeks old female NSG (NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ) mice. Three days post injection, engraftment and disease induction were confirmed by bioluminescence imaging and the baseline was calculated for all 3 cohorts of mice. Disease dissemination was confirmed and subsequently tracked by IVIS bioluminescence imaging (PerkinElmer). In brief, D-luciferin (Cat#122799, Perkin Elmer) was administered by intraperitoneal (IP) injection (10ul/gr of body weight of a 15mg/ml (DPBS) solution) followed by inhalation anaesthesia (isoflurane) and IVIS bioluminescence imaging (seven-and-a-half minutes post luciferin injection for all

animals/imaging sessions). Tumour burden was quantified using Living Image Software (version 4.7.2, PerkinElmer). To induce shRNA expression, mice were provided with a doxycycline containing feed (Cat# A112D72003, Ssniff Spezialdiäten GmbH). All mice were housed in a pathogen-free animal facility and were allowed unrestricted access to food and water. All experiments were conducted under a UK Home Office project (under the Animals (Scientific Procedures) Act 1986, Amendment Regulations (2012)) and following ethical review by the University of Cambridge Animal Welfare and Ethical Review Body.

METHOD DETAILS

Mass spectrometry (MS) based proteomics analyses

Endogenous HOXA9 was immunoprecipitated from MOLM13 cells using 10ug of HOXA9 antibody (details in reagents table) from the nucleus of HOXA9-dependent *MLL-AF9*-rearranged MOLM13 AML cell line (Faber et al., 2009) followed by mass spectrometry. Two replicates of immunoprecipitated samples were analyzed by label free quantification mass spectrometry (LFQMS). Digestion was performed in 2steps; step1) for 30 minutes at 27°C and 800rpm/min Urea buffer plus 5ug/ml Trypsin, followed by 2washes with Urea buffer plus 1mM DTT. Step 2) overnight at room temperature with the pooled supernatants from step1 and washes. Next day, peptides were alkylated for 30 minutes with freshly prepared, light protected Iodoacetamide (IAA, SigmaDAldrich) solution and acidified with Trifluoroacetic, acid, (TFA, Sigma5Aldrich). Desalting was next performed in house prepared staged tips with Solution A (0.1% TFA in H2O) and solution B (50% Acetonitrile (ACN, SigmaDAldrich), 0.1% TFA in H2O), while elution was performed prior to the loading into a Thermo Orbitrap Q Exactive mass spectrometer with DionexRSLC3000UPLC.

Co-Immunoprecipitation

Co-immunoprecipitation was performed in MOLM13 cells using Nuclear Complex Co-IP Kit (Active Motif, Cat#54001) following manufacturer's instructions. The immunoprecipitated samples were processed for run on mass spectrometer as described above or were run on 4-15% gradient SDS-PAGE gels (Biorad) and transferred to PVDF membrane and followed for western blotting in western blotting

section using HRP-conjugated light chain specific secondary antibody (Cat# 211-002-171).

Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME)

RIME was performed as described earlier (Mohammed et al., 2016). Endogenous SAFB or HOXA9 were precipitated from MOLM13 nuclei using 10ug antibodies against SAFB, HOXA9 or IgG controls, after formaldehyde crosslinking. Briefly, wild type MOLM13 or shRNA containing MOLM13 cells (SAFB-sh, HOXA9-sh, Scrambled) were treated with doxycycline (1.5ug/ml) for 5 days. The knockdown and differentiation phenotype were confirmed before fixing the cells with formaldehyde. 60million cells were fixed with 1%formaldehyde and followed the procedure as described in the paper (Mohammed *et al.*, 2016). Antibodies used for RIME were HOXA9 (Atlas Antibodies, Cat# HPA061982), SAFB (Millipore, Cat# 05-588) or Rabbit IgG control (Proteintech, Cat# 30000-0-AP). Three independent replicates were analysed for SAFB or IgG antibodies whereas two replicates were performed for HOXA9 due to limited availability of HOXA9 antibody. We only considered SAFB-associated or HOXA9-associated proteins that were present in all independent replicates for corresponding pull-downs and excluded any protein that occurred in any one of IgG control RIME. The MaxLFQ values were analysed to compare different samples, that includes standard statistical testing combined with quantification accuracy for each of the quantified proteins across different samples (Rappsilber et al., 2002)

MS Data Processing

Raw files were analysed with MaxQuant (version 1.6.6.0) with integrated Andromeda used for searching MS spectra. The following parameters were used: specific trypsin digestion; up to 2 missed cleavages allowed; Precursor ion tolerance: 20 and 4.5 ppm for first and main searches, respectively; Human database (UniProt reference proteome downloaded 18 Dec 2018 containing 21066 proteins) with additional inbuilt MaxQuant contaminant database (containing 246 common contaminants); oxidation (M) and N terminal acetylation as variable modifications; carbamidomethylation (C) as a fixed modification; label-free quantification enabled (MaxLFQ algorithm), LFQ min. ratio count = 2; Match between runs enabled. Proteins were filtered for FDR (< 1%),

reverse sequences and contaminants (both excluded), and three valid values required. Missing values were imputated from a normal distribution (0.3 width, 1.8 SD downshift. Post-processing was carried out using R with bespoke scripts incorporating the “limma” statistical package.

Protein interaction network analyses was performed using STRING online tool (https://string-db.org). The prediction is based on direct (physical) or indirect (functional) associations. STRING version 11.0 was used for this analysis. Number of nodes 79, average local clustering classification 0.716, PPi enriched p value <1.0e-16, number of edges 1095, 48 proteins fall into negative regulation of gene expression category (GO0010629).

Western blotting

Cells were counted and washed twice with cold PBS prior to collection. One million cells were resuspended in 30ul 2XDTT buffer [1.24g DTT, 4ml 20% SDS, 4ml Glycerol, 0.4ml 10% Bromophenol Blue, for total volume of 40ml with water], vortexed and boiled for 5 min. Whole cell lysates were run on 4-15% gradient SDS-PAGE gels (Biorad) and transferred to PVDF-FL membrane. Membranes were blocked in 5%BSA in TBS-tween for 30 min at room temperature (RT), and incubated with primary antibody overnight at 4°C, followed by washes in TBS-Tween and incubation with secondary antibody-fluorescent tag (LI-COR) for 30min. Blots were scanned in LI-COR instrument.

Proximity Ligation Assay (PLA) by Duolink

The PLA was performed using Duolink® In Situ Red Starter Kit Mouse/Rabbit Sigma Aldrich) following manufacturer’s instructions. Briefly, 500,000 cells were fixed onto poly-lysine coated glass slides at 37 ^{o}C for 20 minutes to allow cells to adhere to the slide. Cells were fixed with 4%PFA (10min/RT) followed by permeabilization using 0.15% TX-100 (20min/RT). After subsequent washes, cells were blocked using blocking reagents provided in the kit (overnight/ at 4 degree) and followed the protocol from manufacturer’s instructions. Antibodies used for staining were anti-HOXA9 (Rabbit Polyclonal, Atlas antibodies), anti-SAFB (Mouse monoclonal, Millipore). After last wash, slides were dried overnight in dark and mounted with a coverslip using a

minimal volume of Duolink In Situ Mounting Medium with DAPI. Images were acquired on Airyscan using red channel (AF594) and DAPI.

Lentiviral production

Lentiviral particles were produced in 293T cell lines with co-transfecting packaging plasmids (pMDG2 and psPAX2) with pkLV2 (gRNA expression vector) or pLKO-Tet-on (shRNA vector) using TransIT transfection reagent. The supernatant containing viral particles were collected at 36 hrs post transfection in IMDM medium containing 30%FBS and filtered through 0.35um filters and stored at -80°C.

Generation and validation of CRISPR-KO cell lines

CRISPR knockout was performed using 4-5 distinct sgRNA via lentiviral delivery of guide RNA into cells stably expressing spCas9 (Addgene #52962). The lentiviral plasmid used for cloning sgRNA was pKLV2-U6gRNA5(BbsI)-PGKpuro2AmCherry-W or BFP (Addgene #67977). Briefly, cells were freshly seeded at 1million/ml density in 6 well plate in 3ml medium. One ml of lentivirus was added onto cells (1:4 dilution), with 8-10ug/ml polybrene. Cells were spun at 3000rpm for 1.5hr at 32°C. Following spinoculation, cells were incubated overnight at 37degree. Next day, medium was changed with fresh media. And cells were harvested for subsequent experiment on following day.

Colony forming unit assay (CFU)

CRISPR-ko or shRNA expressing MOLM13 cells were seeded (1000cells per ml) in methylcellulose (MehtoCult H4434, Stem Cell technologies). CFU were scored and photographs were taken using STEMVision instrument (Stem Cell technologies).

Cumulative growth curve

Cells were seeded at 50,000 cells/ml density. For shRNA containing cells, expression was induced with doxycycline (1.5μg/ml). Cells were counted daily or alternate days depending on the experiment. Cells were maintained at initial density every 2 days. Fold accumulation was calculated by comparing to the starting cell number, followed by multiplying the fold change each consecutive day. The data shown as average of

biological replicates (n=3) ± SD. Statistical significance was calculated using prism7 software, 2way ANOVA test. Error bars represent ± SD.

For CEBPδ overexpressing cells, cell density was maintained at initial concentration after each count and fresh dox was added to the culture. Fold accumulation was calculated as mentioned above.

Competition assay

Competition assay was performed between Flow sorted GFP positive NOTCH-ICN or MSCV-control vector expressing MOLM13 cells. Equal number of GFP positive cells were seeded at day0 and GFP percentage was measured upto 5 days by flowcytometry. Data shown as average of biological replicates (n=3) ± SD.

Flow cytometry (FACS) analysis

To monitor the differentiation status, cells were stained with the following antibodies PECy7-αHu-CD11b (Cat#301322, Biolegend), APC-αHu-CD15 (Cat#17-0158-42, eBioscience), FITC-αHu-CD14 (Cat#301804, Biolegend), [1:100 in FACS buffer: PBS with 2%FBS]. To measure apoptosis, cells were washed once with Annexin V binding buffer, and incubated with anti-FITC-AnnexinV (Cat# 640945, Biolegend) and 7-AAD (Cat# 420404, Biolegend) in the AnnexinV binding buffer for 15 min based on manufacturer's instructions. Cells were analysed on a BD-Fortessa instrument. Data were analysed using FlowJo software version 10.6.1.

Cell Cycle analyses

Cell cycle analyses was performed using Click-iT Plus EdU cell proliferation kit (Invitrogen, Cat# C10645) following manufacturer's instructions. Cells were analysed on a BD-Fortessa instrument. Data were analysed using FlowJo software version 10.6.1.

RNA isolation and qPCR

Total RNA was extracted from cells using RNeasy Plus mini (QIAGEN) kit following manufacturer's instructions. For RNA-seq, libraries were prepared using Illumina kit (NEBNext Ultra II Directional RNA Library Prep Kit for Illumina, Cat# E7760S) with RiboZero depletion following manufacturer's instructions. Libraries were sequenced on Hiseq platform with PE150 reads length.

For RT-QPCR, 1ug of RNA was reverse transcribed into cDNA using Taqman Reverse Transcription Reagents kit (applied biosystem, Cat# N8080234). QPCR was performed using an Mx3000P qPCR System (Agilent) detection system using primers (sequences listed in primers table) together with SYBR green master mix (Brilliant III Ultra Fats SYBR green QPCR master mix, Agilent, Cat# 600882).

CUT&RUN

CUT&RUN was performed using CUTANA CUT&RUN kit (Epicypher) following manufacturer's instructions. 100,000 MOLM13 cells were used for the assay for each antibody. Antibodies are listed in the table. For Input, same number of cells were used and treated with MNase similar to antibody containing sample and processed as other samples. Libraries were prepared using 2ng of DNA using NEBNext UltraII DNA library preparation kit (Cat# E7645S, Illumina) and Dual index primer-Multiplex Oligos (Cat# E7600S, Illumina). The samples were amplified for 12 cycles using KAPA HiFI 2X master mix (Cat# 07958935001, Roche). After amplification library was size selected using Ampure beads. The library size and concentration were analyzed using Tapestation reagents-D5000 (Cat# 5067-5588, 5067-5589, Agilent) in 4200 TapeStation system (Agilent) before pooling the samples. The samples were sequenced at Novaseq SP50bpPE.

QUANTIFICATION AND STATISTICAL ANALYSES

ChIP-Seq and CUT&RUN analyses

For ChIP-seq data, single/paired end reads were mapped against human reference genome (Hg38) using Bowtie2 (Langmead and Salzberg, 2012). Prior to that read were quality checked and adapter sequences were trimmed using TrimGalore package (https://github.com/FelixKrueger/TrimGalore). Uniquely mapped reads were

retained and peaks were called MACS2 using the Input as control (Zhang et al., 2008). Peaks were further filtered by removing blacklisted regions from Encode (Amemiya et al., 2019). Peaks were classified into intergenic, intragenic and promoter bound using an in-house script. Any peak overlapping 1kb upstream and 500bp downstream window of TSS is classified as promoter bound. Motif analysis was performed using HOMER (Heinz et al., 2010).

For CUT&RUN, paired end reads were processed using recently published CUT&RUNTools package version 2.0 [https://github.com/fl-yu/CUT-RUNTools-2.0] (Zhu et al., 2019). Reads were aligned to the human reference genome (Hg38). Peaks were called after filtering PCR duplicates using MACS2 with narrow-Peak setting against Input using FDR 5% (Zhang *et al.*, 2008). Black-listed regions were further filtered from the called peaks. Peaks were merged as we observed strong correlation between the replicates. We used ≤120bp fragments generated from CUTnRUN pipeline (which is likely to contain TF binding sites) for motif analyses using HOMER. Broad peaks (CBX3) were called using SICER with parameters window size-200 bp (W200), gap-600 bp (G600) at FDR 1% against Input.

Additional motif analysis was performed using the MEME Suite 5.5.0 (Bailey et al., 2015). *De novo* motif discovery was conducted on peak summits (±100bp) from different peak-calling results using the XSTREME tool from MEME Suite (Grant and Bailey, 2021), using the following parameters (--time 240 --streme-totallength 4000000 --meme-searchsize 100000 --fdesc description --dna --evt 0.05 --minw 6 --maxw 15 --align center --meme-mod zoops). Motif enrichment analysis was carried out on the same summit sequences using the SEA tool from MEME Suite (Grant and Bailey, 2021), with the following parameters (-thresh 10.0 --align center). JASPAR2022 CORE non-redundant database v2 was used throughout motif analyses. In all instances of motif analysis, every peak set was run against a matched set of controls of the same number of fragments, lengths and chromosome composition to the Cut&Run/ChIP-Seq peak set. The canonical HOXA9 motif (MA0594.2) was used to run FIMO (Grant and Bailey, 2011) on summits regions to find instances of enrichment of the motif in the peak sequences.

Overlap between different peak sets was done by using Bedtools package (Quinlan and Hall, 2010), using a minimum of one base pair overlap between any two peaks. Intersection between HOXA9/SAFB co-occupied regions across human patients and cell-lines was estimated and visualised using the Intervene package (version 0.6.5)

UpSet module (Khan and Mathelier, 2017). Gene ontology analyses was performed using Enrichr online tool (Kuleshov et al., 2016).

Softwares used

Figures from CUT&Run and ChIP-seq were prepared using Easeq online tool (Lerdrup et al., 2016). Statistical analyses and other figures were prepared using Prism software version 7.0. Venn diagrams were made using online tool Venny (https://www.stefanjol.nl/venny). Arivis software used for image analyses for Duolink pictures.

RNA-Seq analysis

Paired end RNA-seq reads were quality filtered similar to ChIP-Seq reads and were mapped using STAR version=2.7.10a (Dobin et al., 2013) against the human genome (Hg38). Uniquely-mapping, high confidence reads were retained using the following flags (NH:i:1 and MAPQ=255). Read counts were quantified with HTSeq version 1.99.2 (Anders et al., 2015) and differential expression analysis was carried out with these counts using Bioconductor package DESeq2 (Love et al., 2014). Volcano plots were plotted using an R package EnhancedVolcano (https://github.com/kevinblighe/EnhancedVolcano). Gene ontology analyses was performed using the R package 'clusterProfiler' version 4.4.4 (Yu et al. 2012).

To further refine the list of targets genes directly regulated by HOXA9 and SAFB, we integrated HOXA9/SAFB CUT&RUN-seq data with RNA-seq following *HOXA9/SAFB* perturbation, using Bedtools suit v2.30.0 (Quinlan and Hall, 2010), with a larger window size (±50kb) to assign non-promoter bound peaks to genes.

S/MAR feature motif search

For searching S/MAR features motif on HOXA9-SAFB cobound sequences, FIMO-MEME suit online tool was used. In brief, presence of OriC was determined by detecting presence of ATTA or ATTTA or ATTTTA motif, AT richness by presence of two WWWWWW (where W is A or T) motifs intervened by 8–12 nt, Kinked DNA by the presence of TAN3TGN3CA or TAN3CAN3TG or TGN3TAN3CA or TGN3CAN3TA or CAN3TAN3TG or CAN3TGN3TA motif (where N is any nucleotide), Curved DNA by presence of AAAAN7AAAAN7AAAA or TTTTN7TTTTN7TTTT or TTTAAA (where

N is any nucleotide), TG richness by the presence of TGTTTTG or TGTTTTTTG or TTTTGGGG motifs, MAR signature by presence of a bipartite sequence containing AATAAYAA and AWWRTAANNWWGNNNC (where W is A or T, Y is pyrimidine, R is purine and N is any nucleotide) and Topoisomerase II binding site by the presence of RNYNNCNNGYNGKTNYNY or GTNWAYATTNATNNR (where W is A or T, Y is pyrimidine, R is purine and N is any nucleotide) consensus (Narwade et al., 2019).

Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci Rep *9*, 9354. 10.1038/s41598-019-45839-z.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169. 10.1093/bioinformatics/btu638.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21. 10.1093/bioinformatics/bts635.

Faber, J., Krivtsov, A.V., Stubbs, M.C., Wright, R., Davis, T.N., van den Heuvel-Eibrink, M., Zwaan, C.M., Kung, A.L., and Armstrong, S.A. (2009). HOXA9 is required for survival in human MLL-rearranged acute leukemias. Blood *113*, 2375-2385. 10.1182/blood-2007-09-113597.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589. 10.1016/j.molcel.2010.05.004.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res *44*, W90-97. 10.1093/nar/gkw377.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359. 10.1038/nmeth.1923.

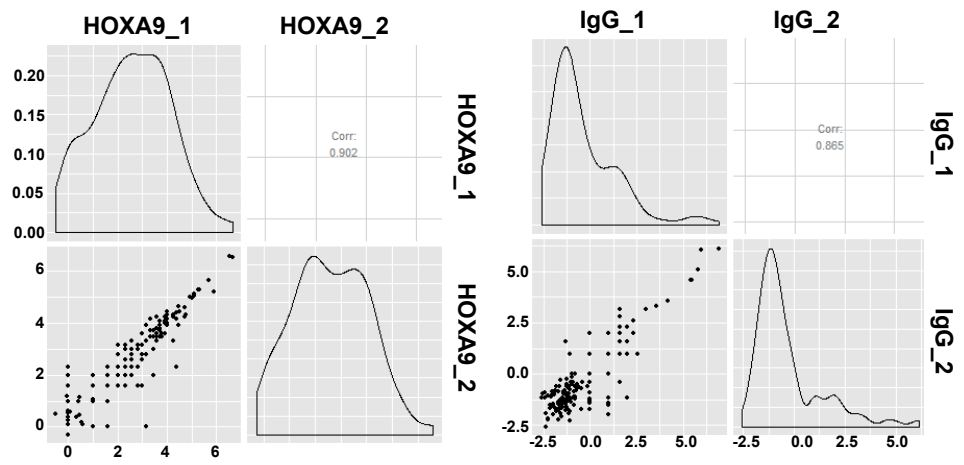
Lerdrup, M., Johansen, J.V., Agrawal-Singh, S., and Hansen, K. (2016). An interactive environment for agile analysis and visualization of ChIP-sequencing data. Nat Struct Mol Biol *23*, 349-357. 10.1038/nsmb.3180.

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

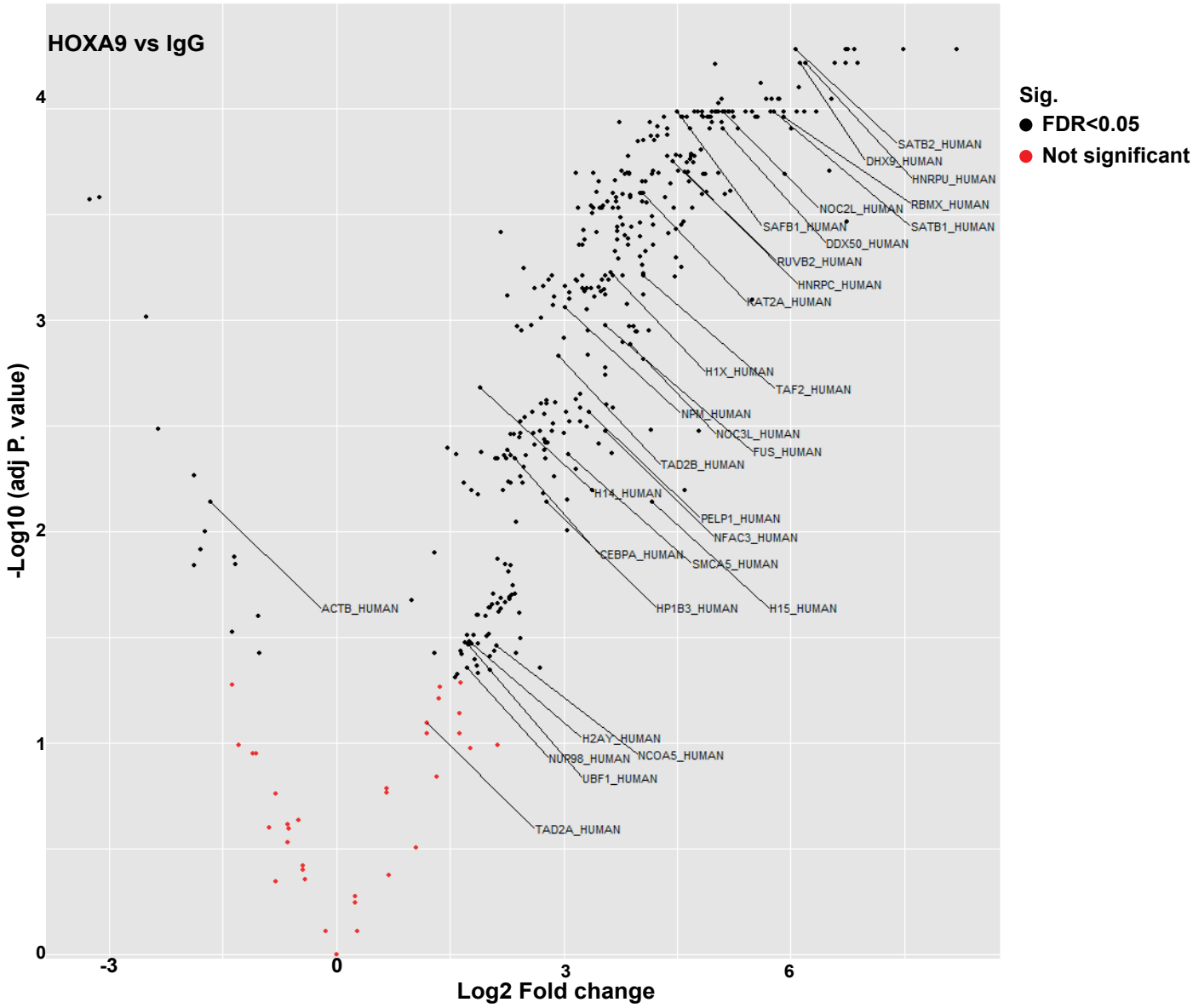
Mohammed, H., Taylor, C., Brown, G.D., Papachristou, E.K., Carroll, J.S., and D'Santos, C.S. (2016). Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. Nat Protoc *11*, 316-326. 10.1038/nprot.2016.020.

Narwade, N., Patel, S., Alam, A., Chattopadhyay, S., Mittal, S., and Kulkarni, A. (2019). Mapping of scaffold/matrix attachment regions in human genome: a data mining exercise. Nucleic Acids Res *47*, 7247-7261. 10.1093/nar/gkz562.
Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842. 10.1093/bioinformatics/btq033.
Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. Genome Res *12*, 1231-1245. 10.1101/gr.473902.
Wee, S., Wiederschain, D., Maira, S.M., Loo, A., Miller, C., deBeaumont, R., Stegmeier, F., Yao, Y.M., and Lengauer, C. (2008). PTEN-deficient cancers depend on PIK3CB. Proc Natl Acad Sci U S A *105*, 13057-13062. 10.1073/pnas.0802655105.
Wiederschain, D., Wee, S., Chen, L., Loo, A., Yang, G., Huang, A., Chen, Y., Caponigro, G., Yao, Y.M., Lengauer, C., et al. (2009). Single-vector inducible lentiviral RNAi system for oncology target validation. Cell Cycle *8*, 498-504. 10.4161/cc.8.3.7701.
Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137. 10.1186/gb-2008-9-9-r137.
Zhu, Q., Liu, N., Orkin, S.H., and Yuan, G.C. (2019). CUT&RUNTools: a flexible pipeline for CUT&RUN processing and footprint analysis. Genome Biol *20*, 192. 10.1186/s13059-019-1802-4.

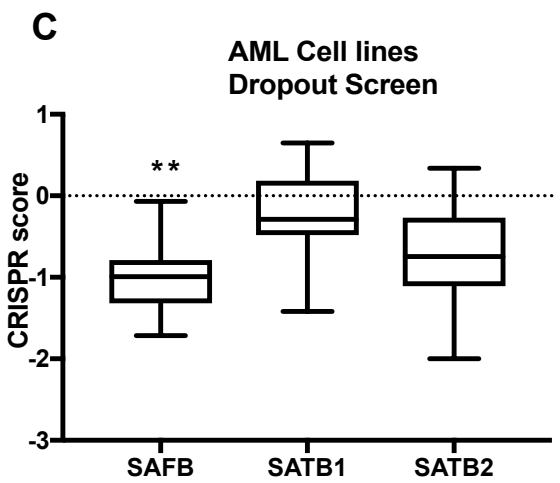
Supplementary Figure 1 A



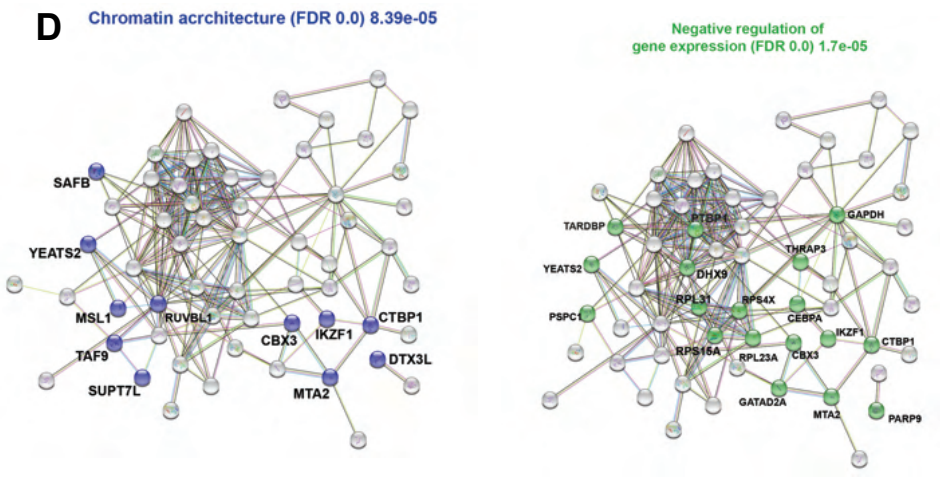
B



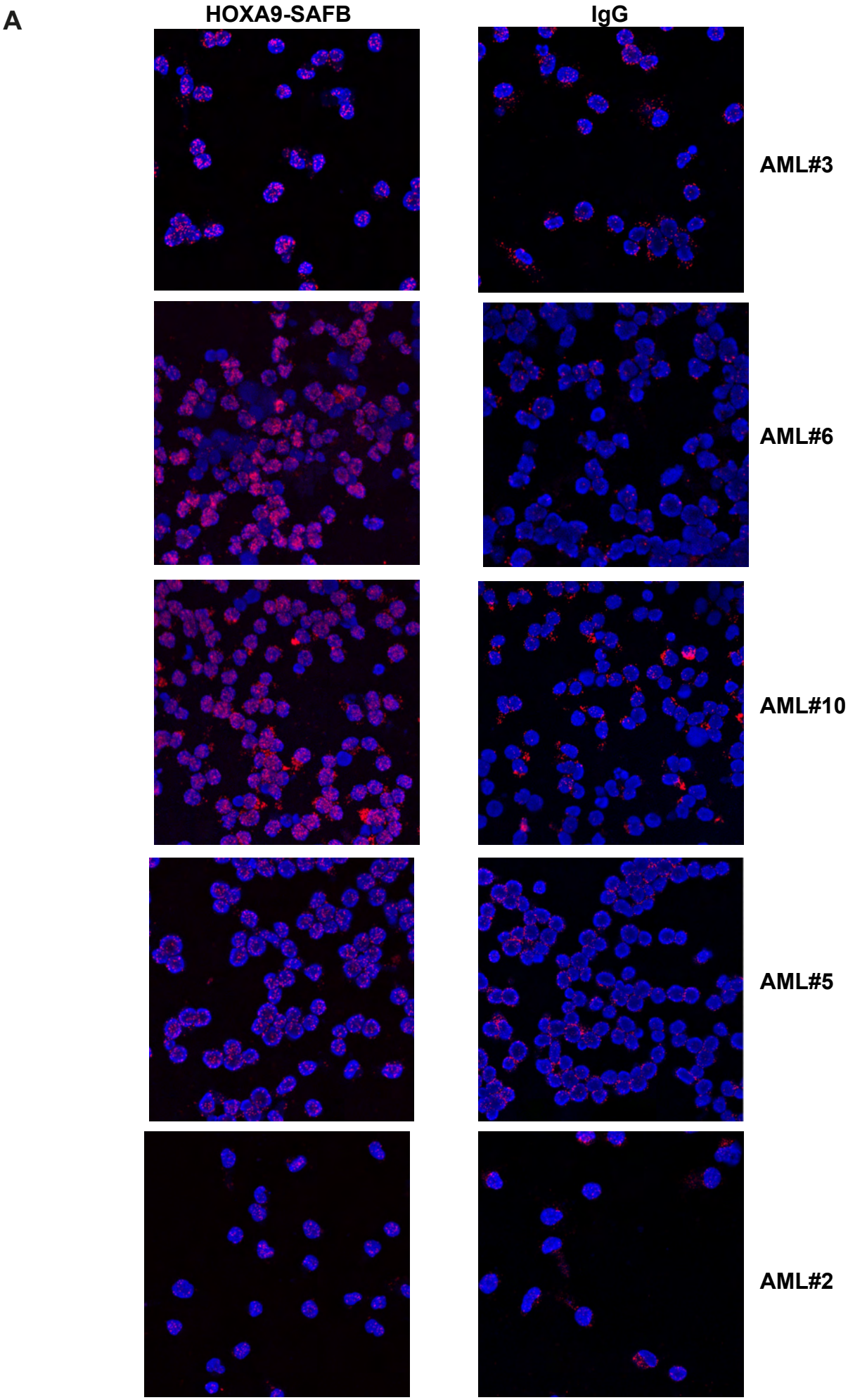
C



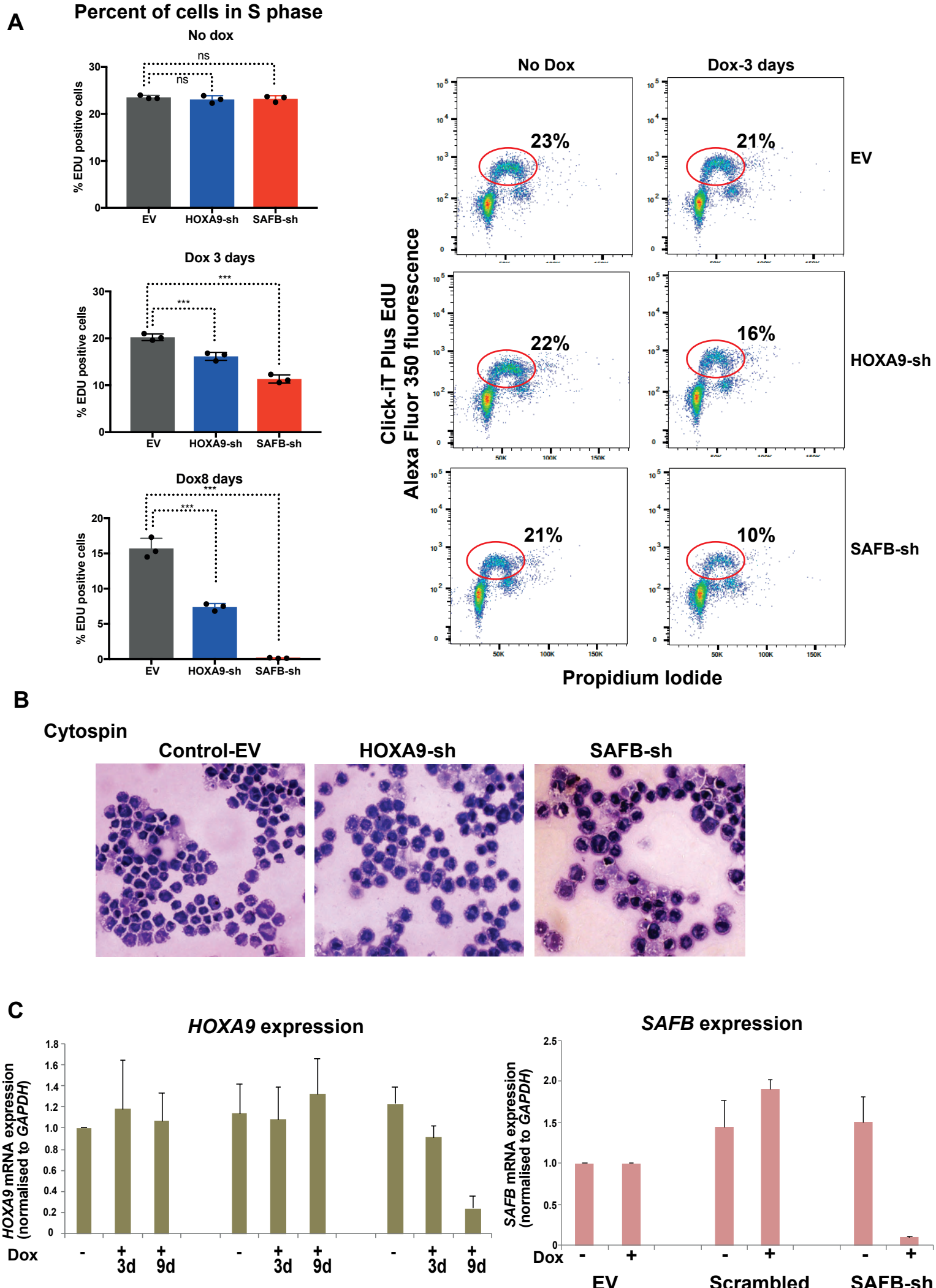
D



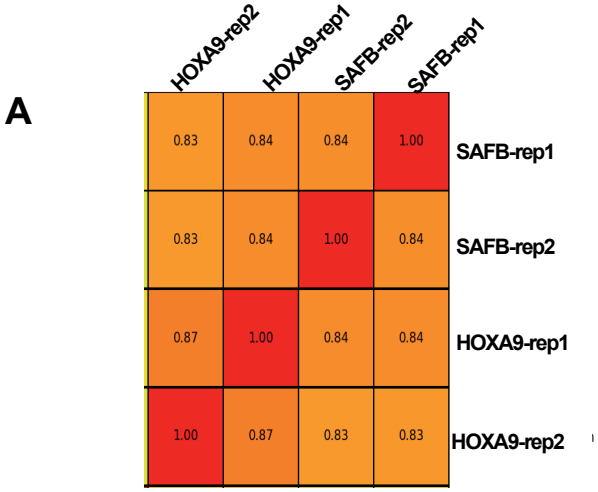
Supplementary Figure2

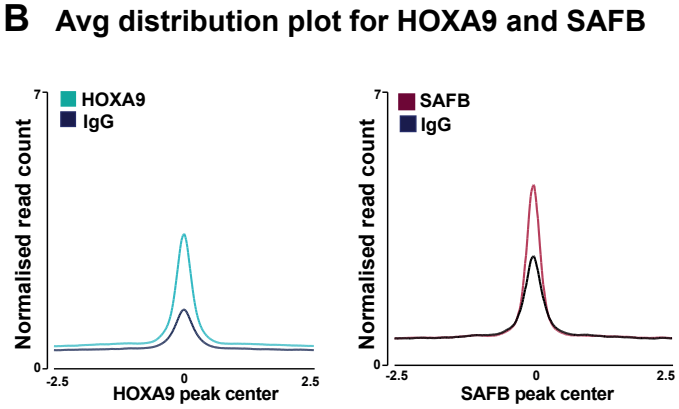


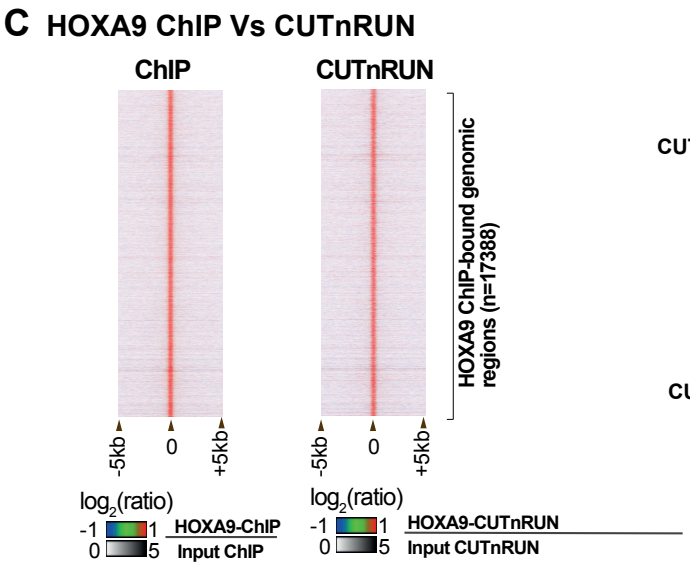
Supplementary Figure 3

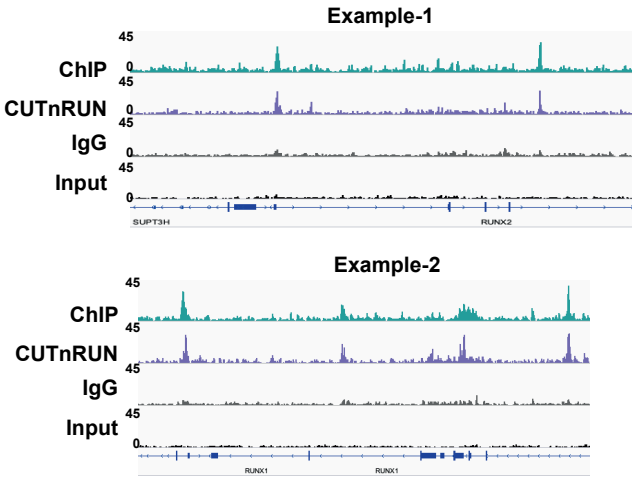


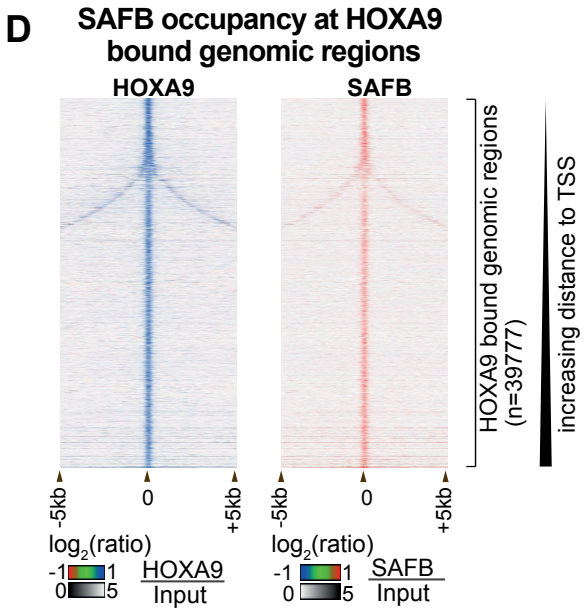
Supplementary Figure 4









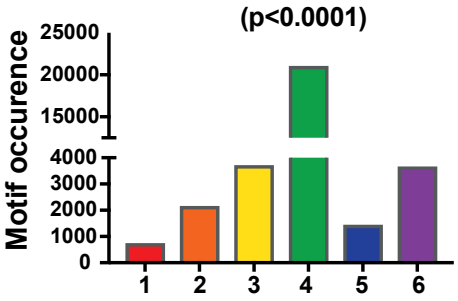


E Motif enrichment

Rank	Motif	Best Match	P-value
1		SpiB (ETS)	1e-2508
2		AP-1 (bZIP)	1e-2106
3		RUNX1 (Runt)	1e-645
4		CEBPA	1e-605
5		NFY (CCAAT)	1e-594

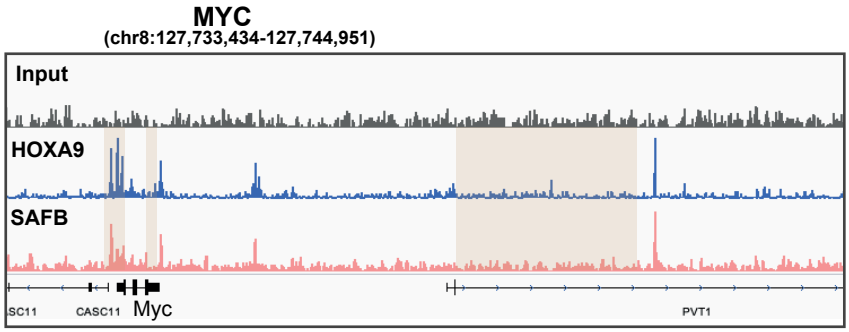
Supplementary Figure 5

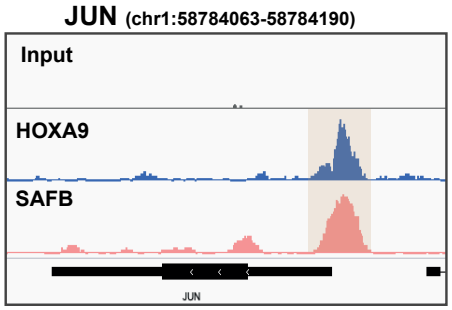
A S/MAR features on HOXA9-SAFB co-bound genomic regions

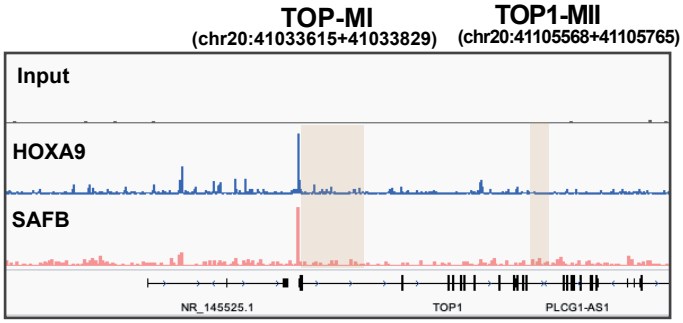


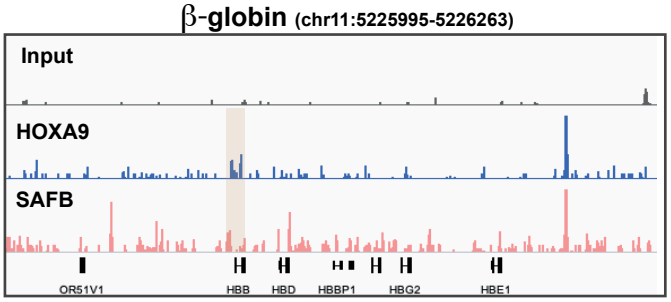
1 AT richness
2 TG richness
3 Kinked DNA
4 Curved DNA
5 MAR signature bipartite seq
6 Topoisomaerase-II binding

B HOXA9-SAFB occupancy at known classical S/MAR (shaded area)



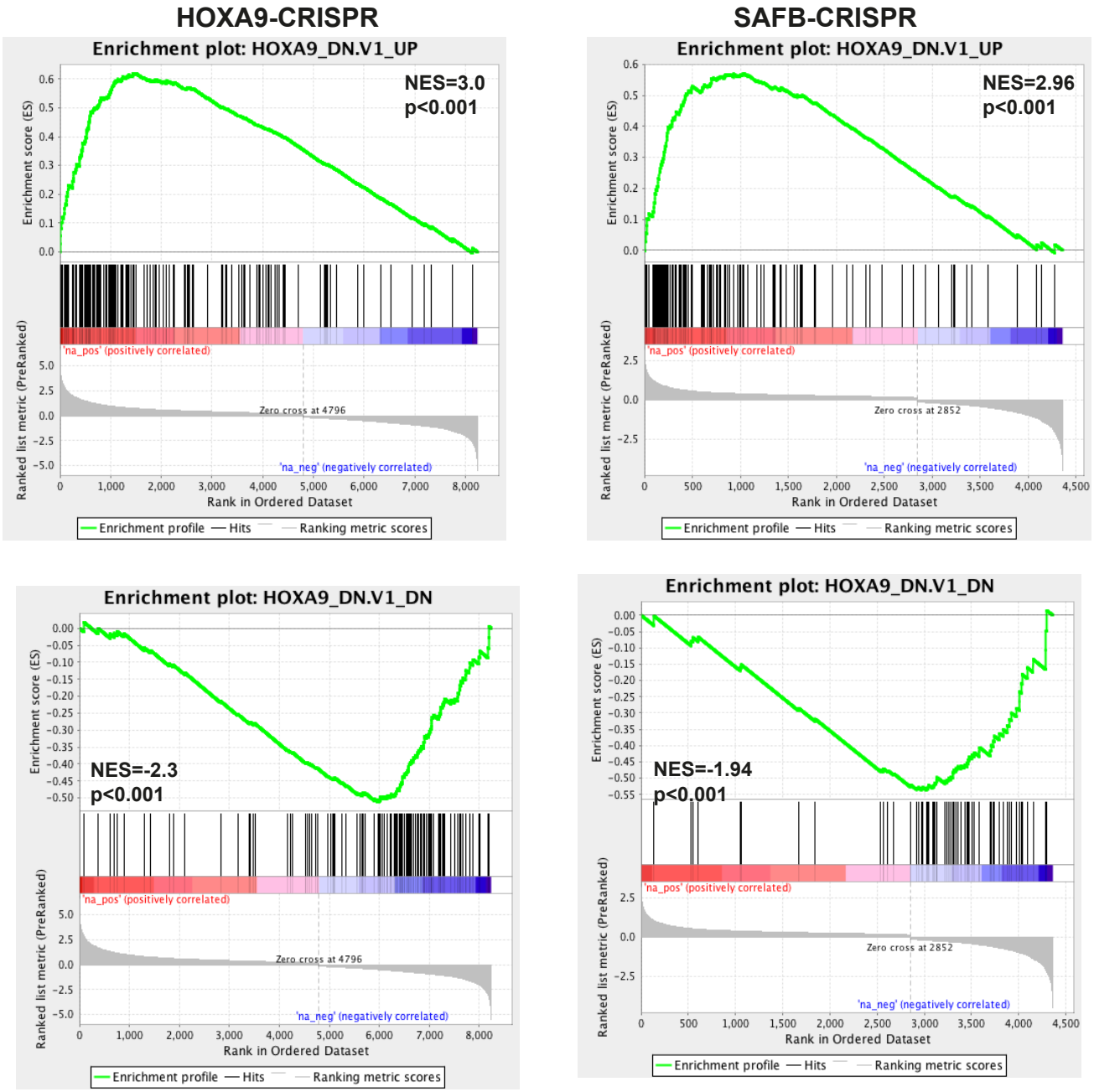




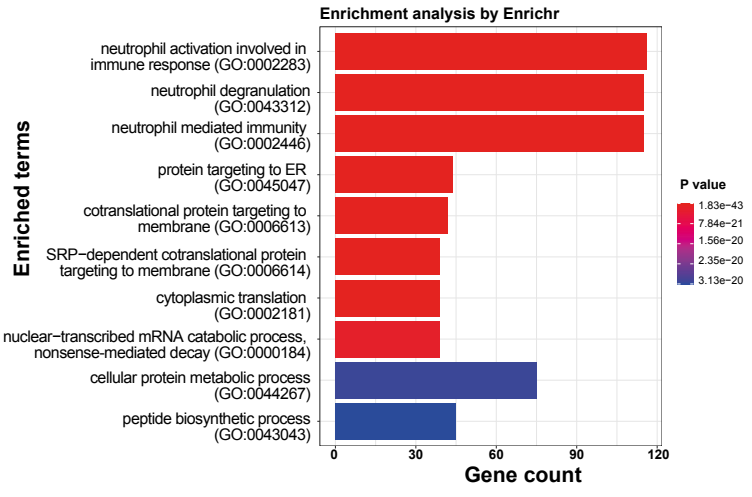


Supplementary Figure 6

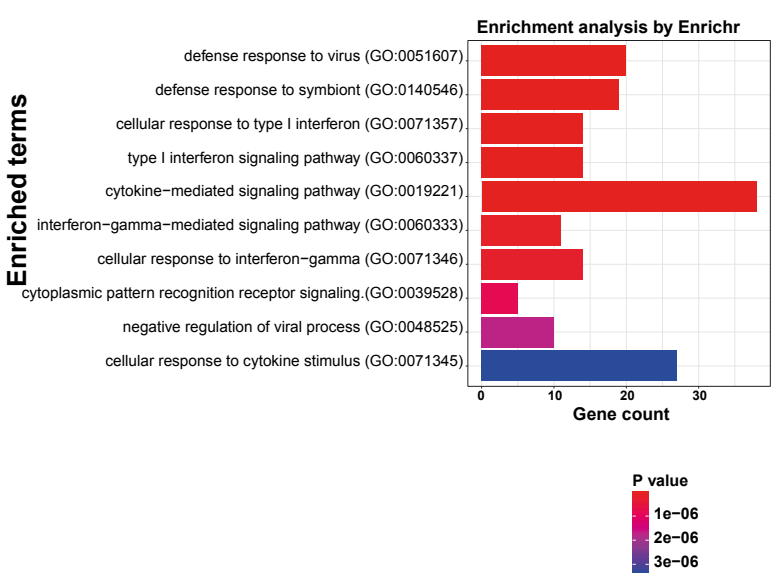
A



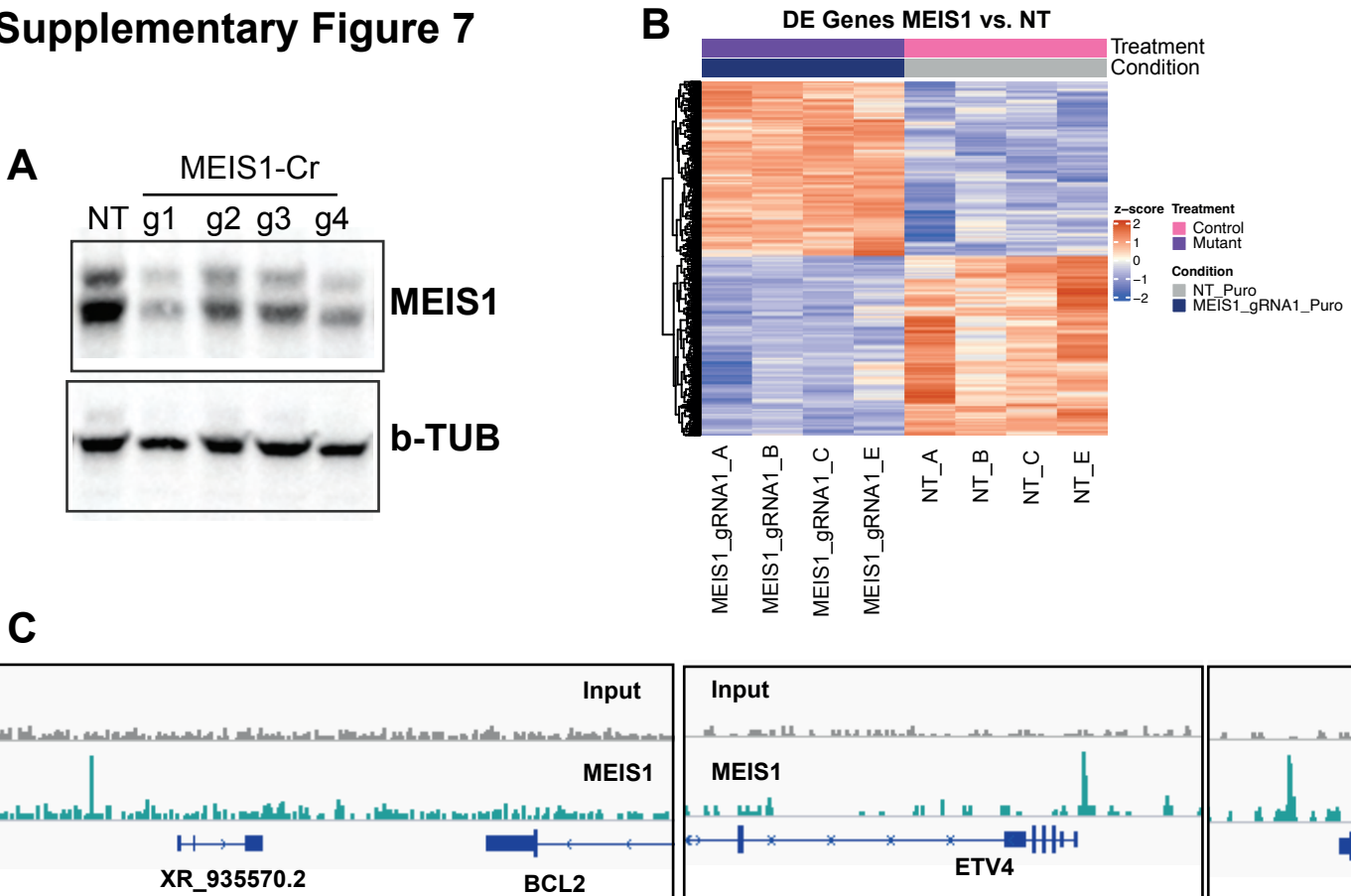
B

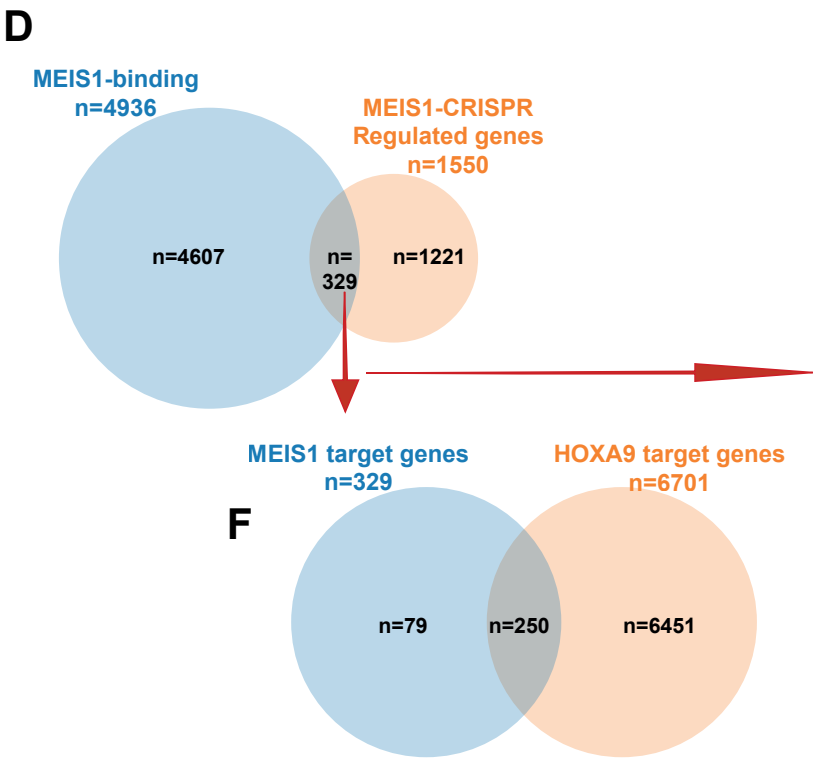


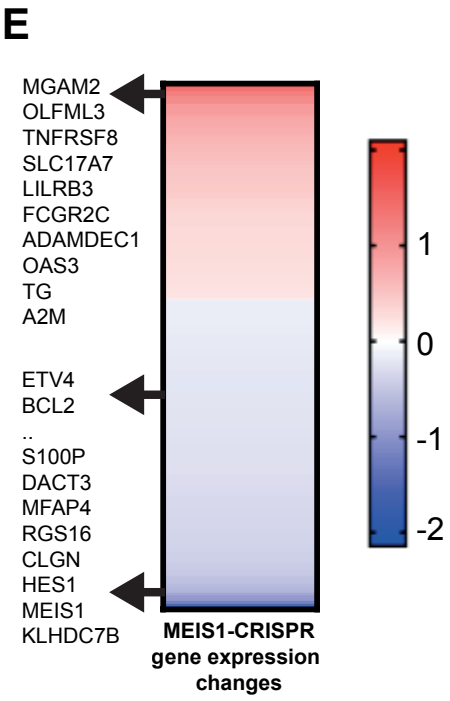
C



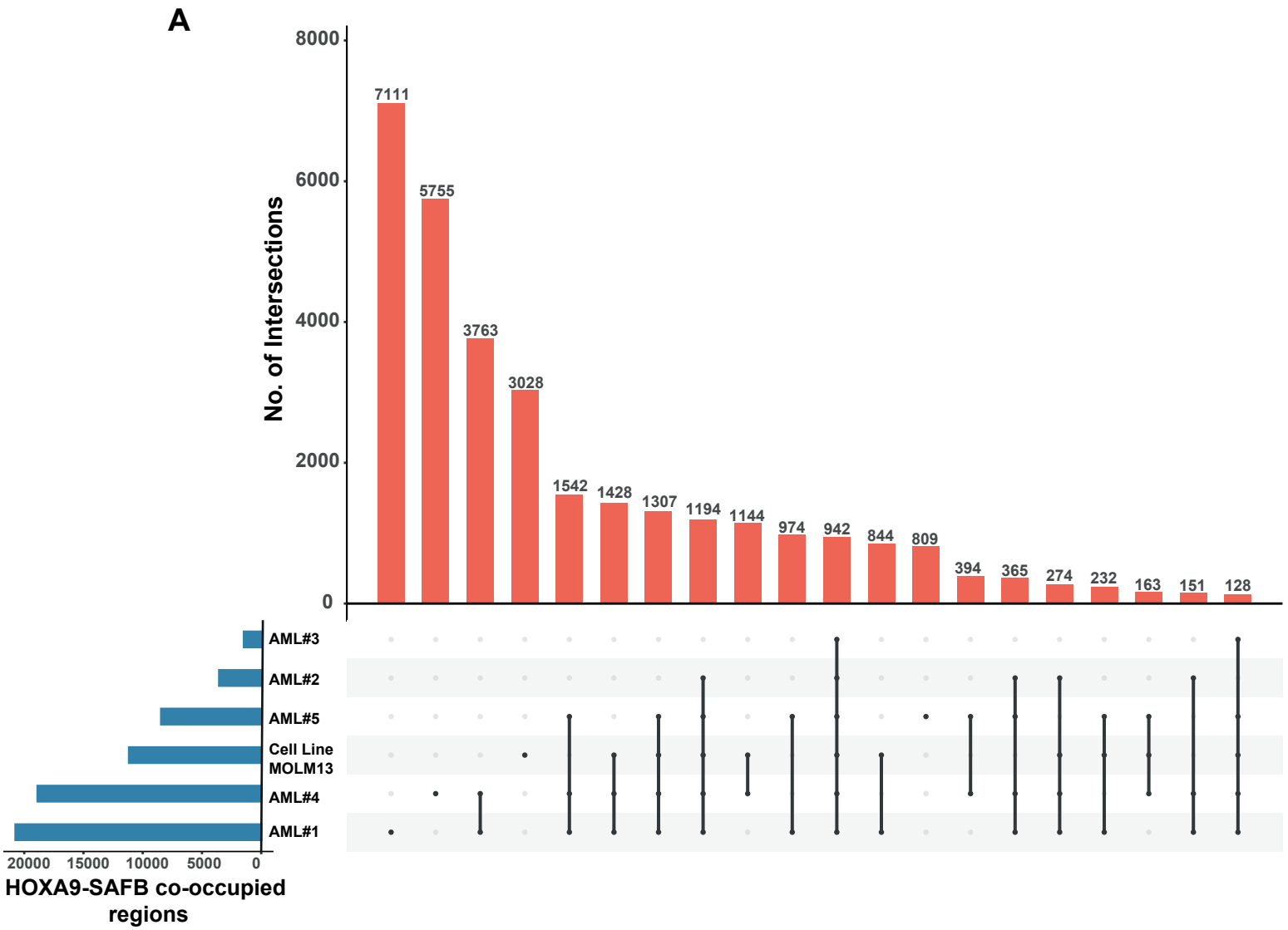
Supplementary Figure 7



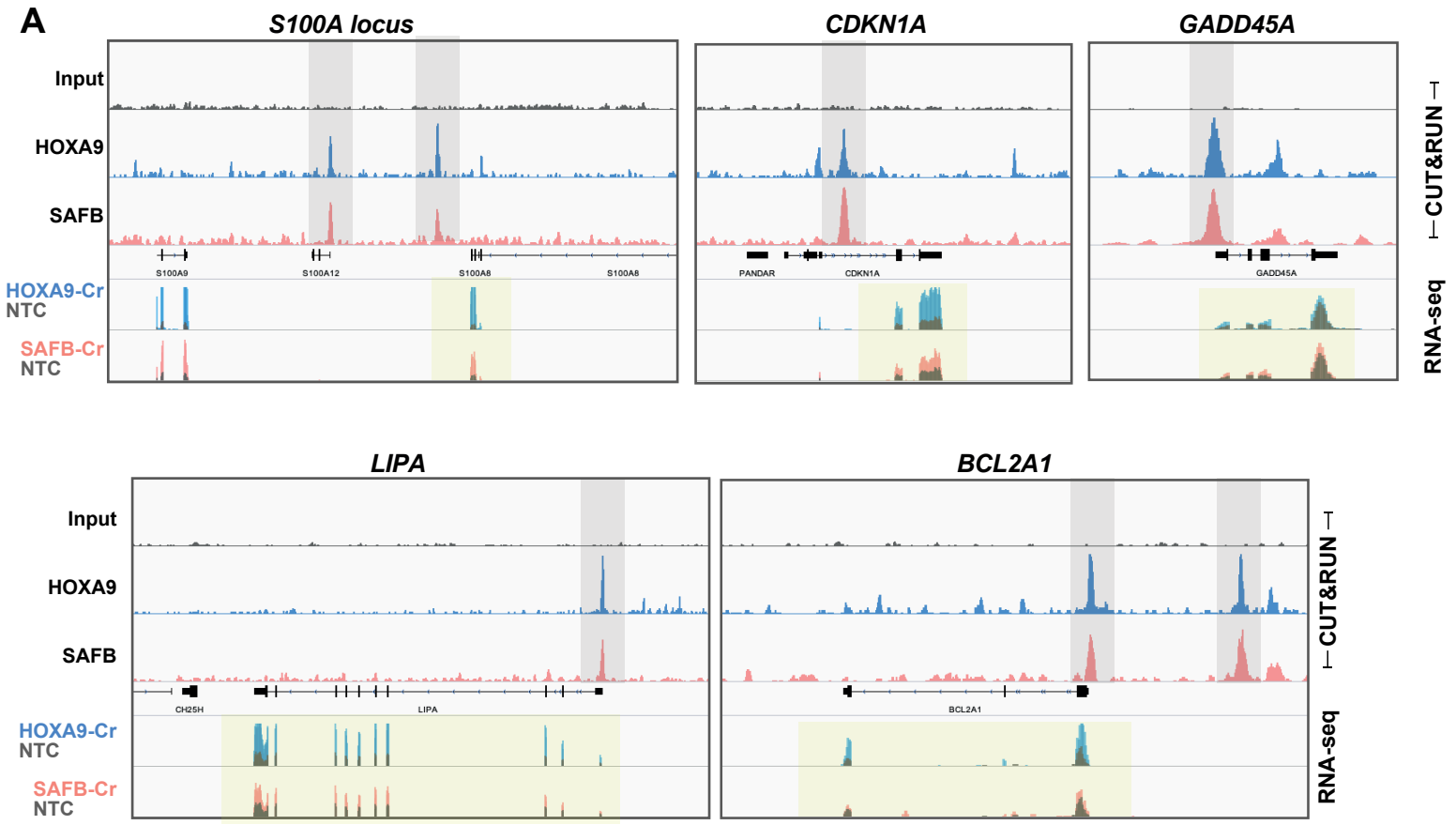


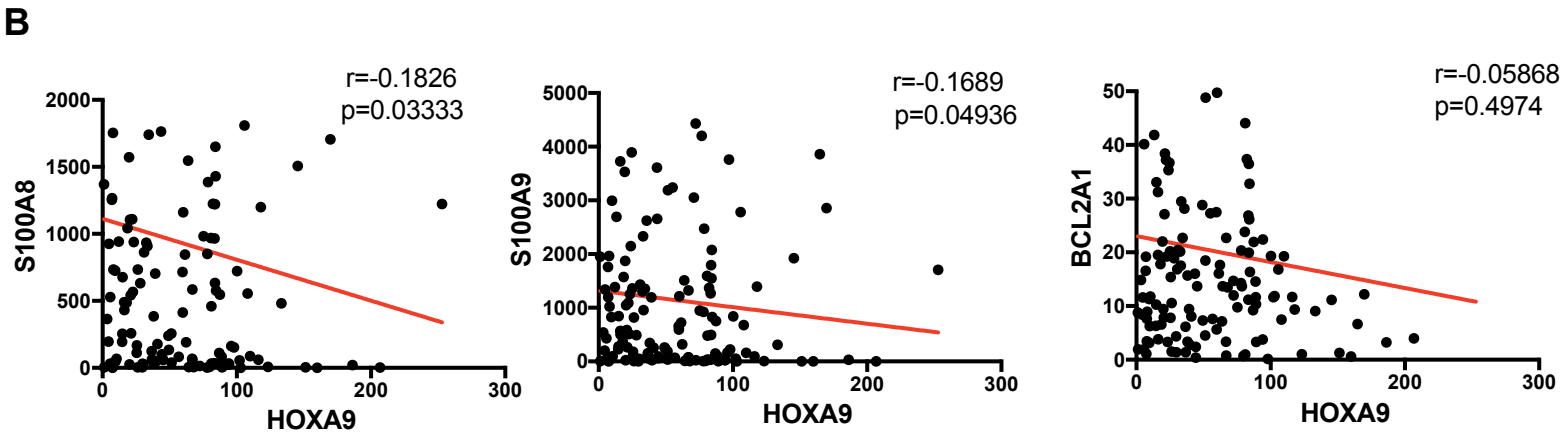


Supplementary Figure 8

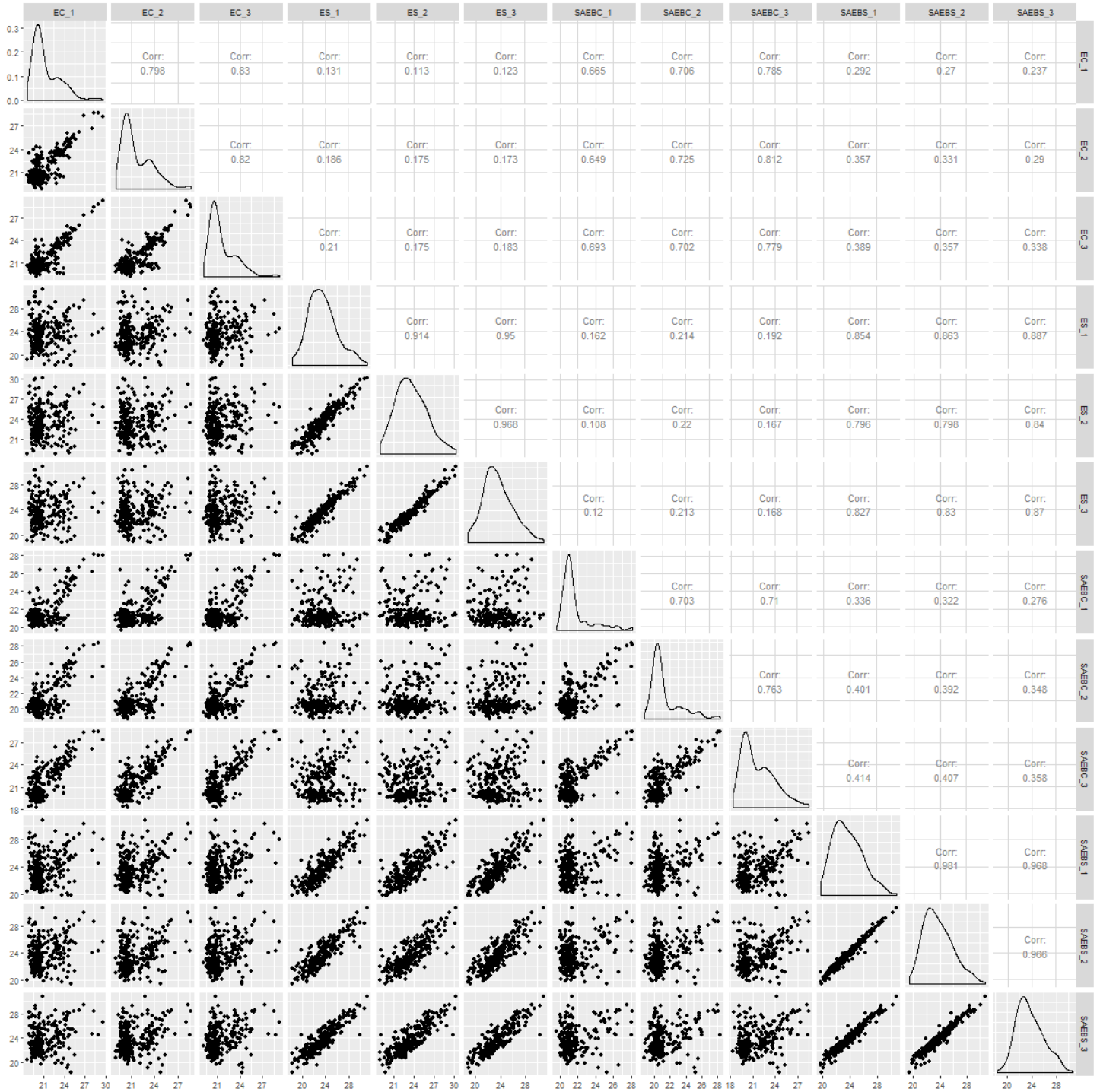


Supplementary Figure 9



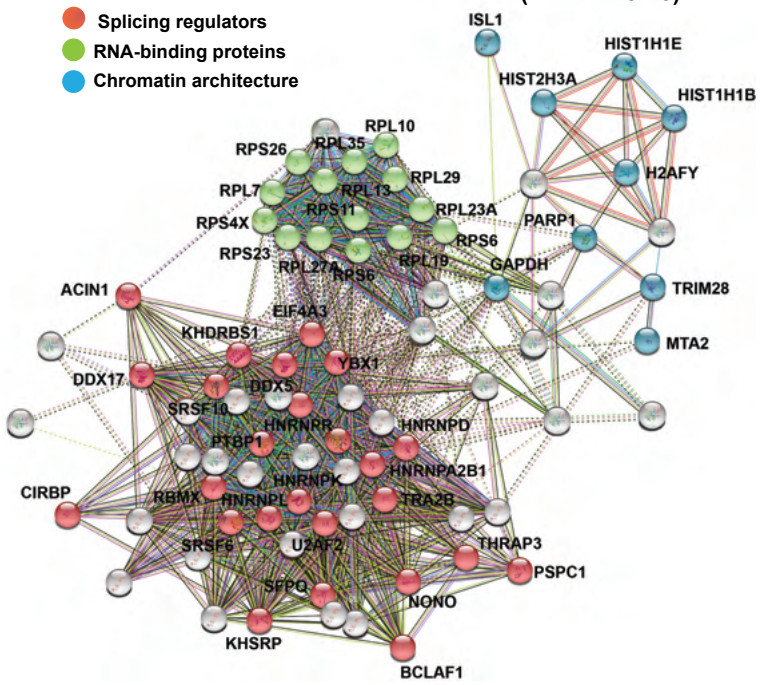


Supplementary Figure 10

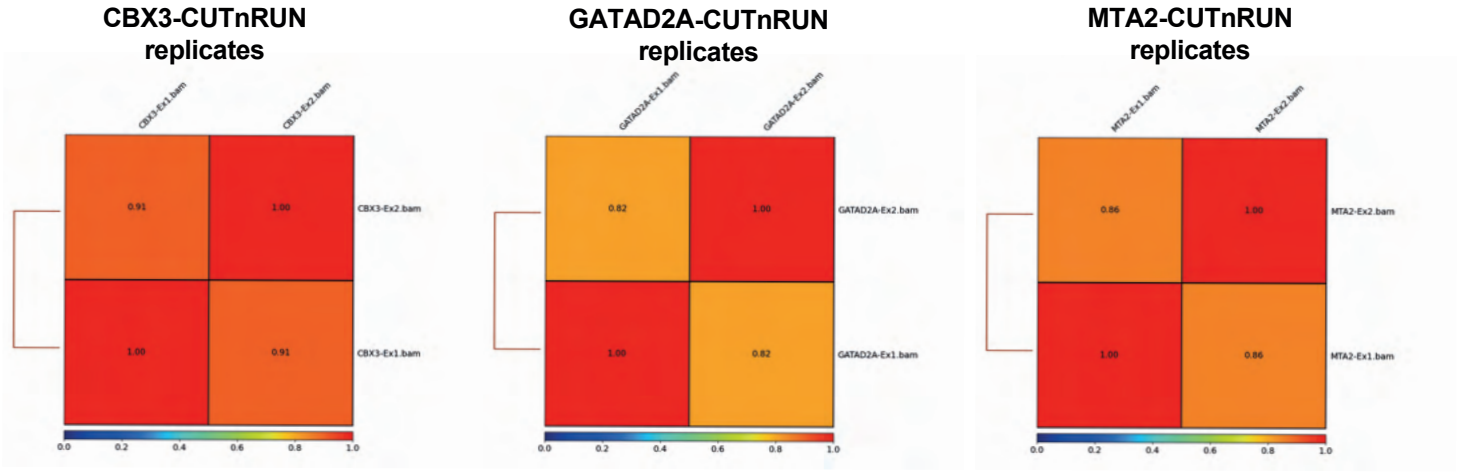


Supplementary Figure 11

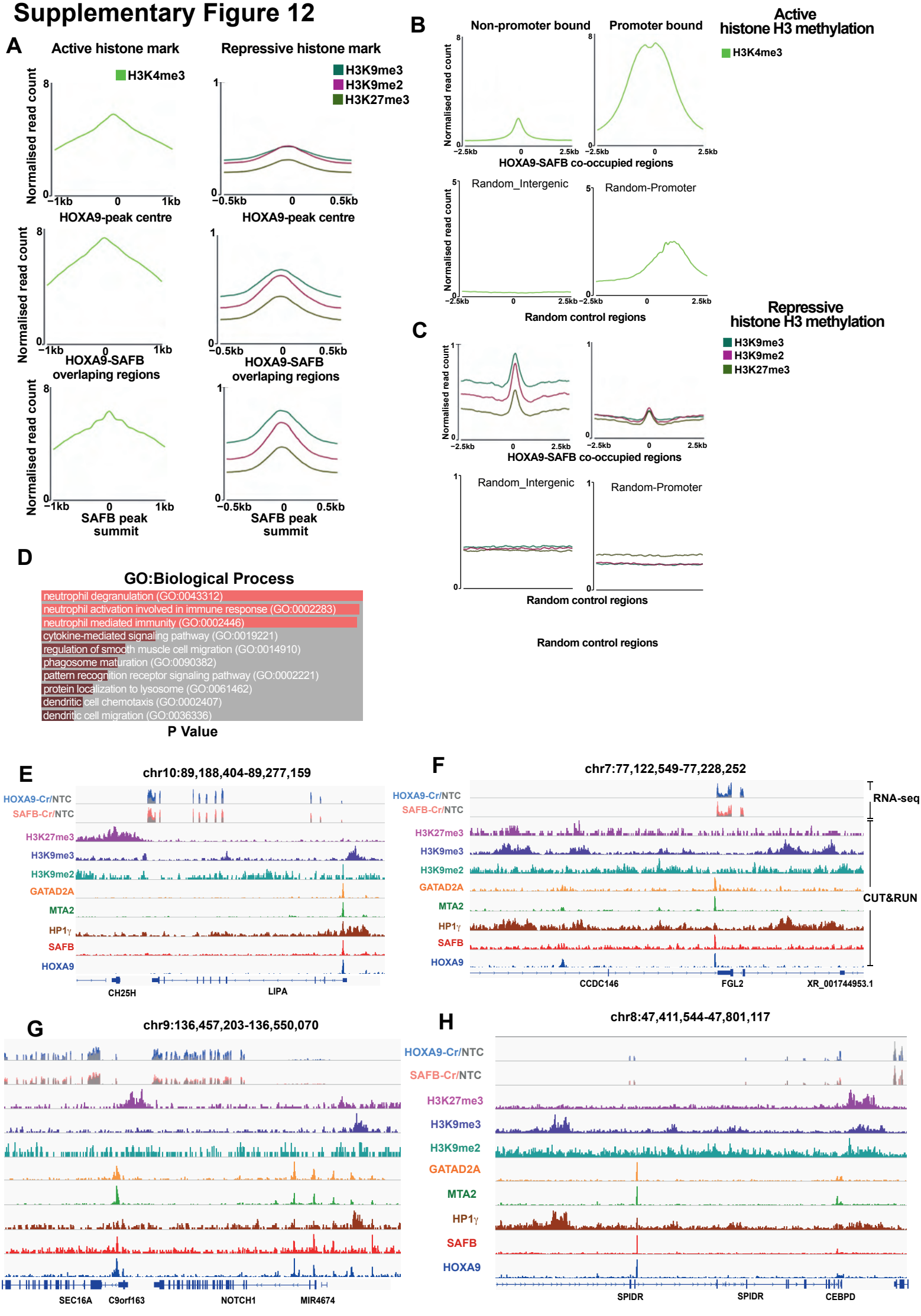
A



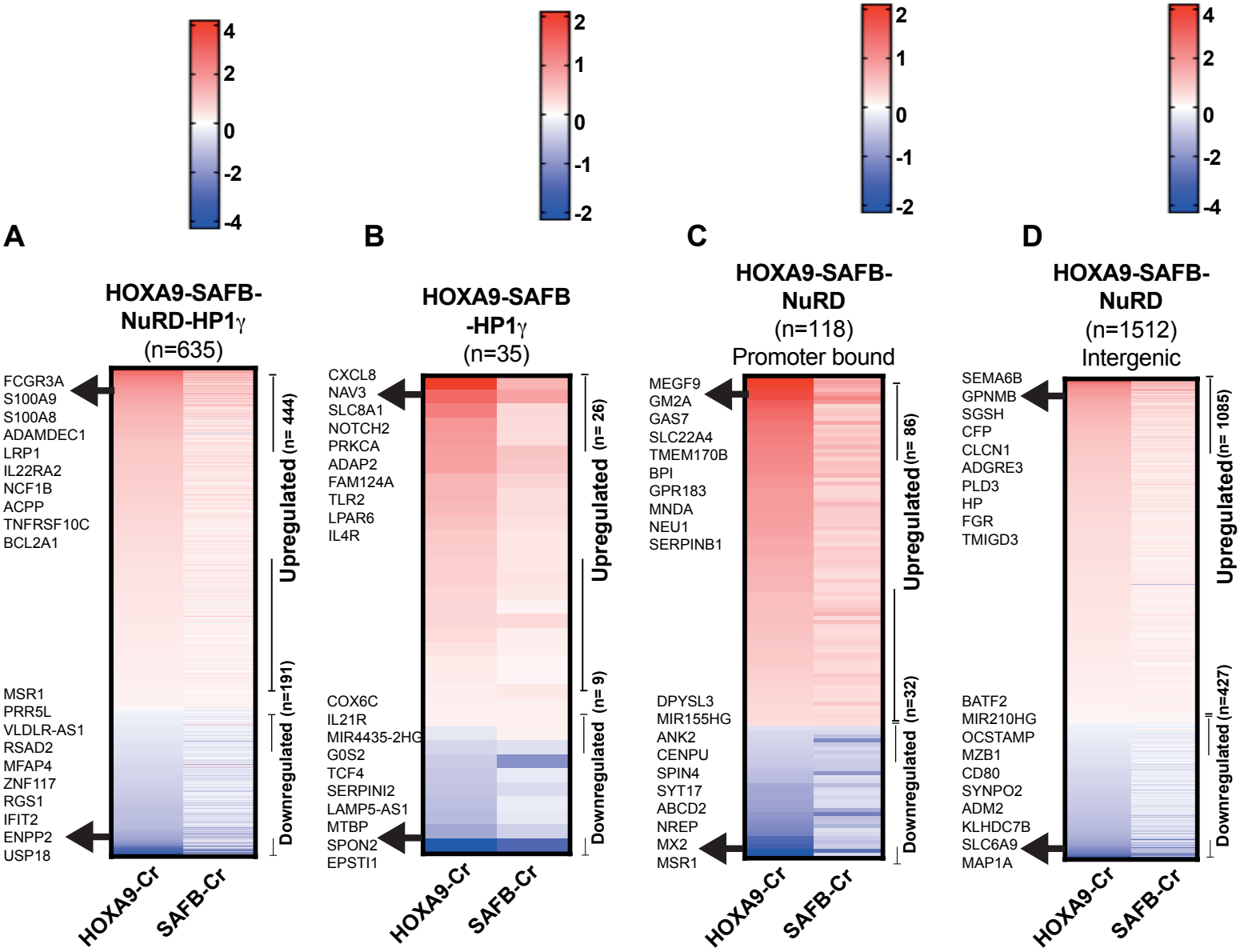
B



Supplementary Figure 12



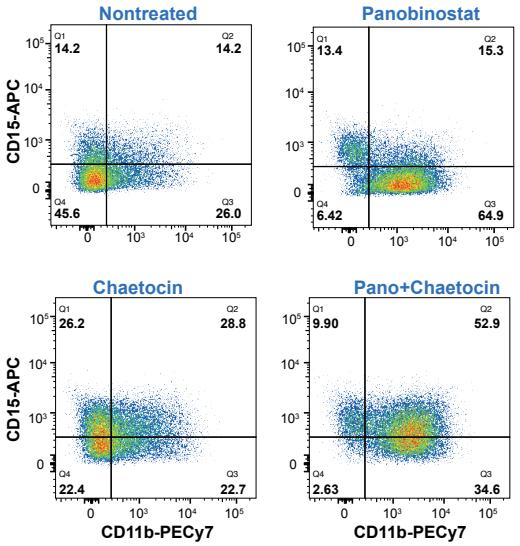
Supplementary Figure 13



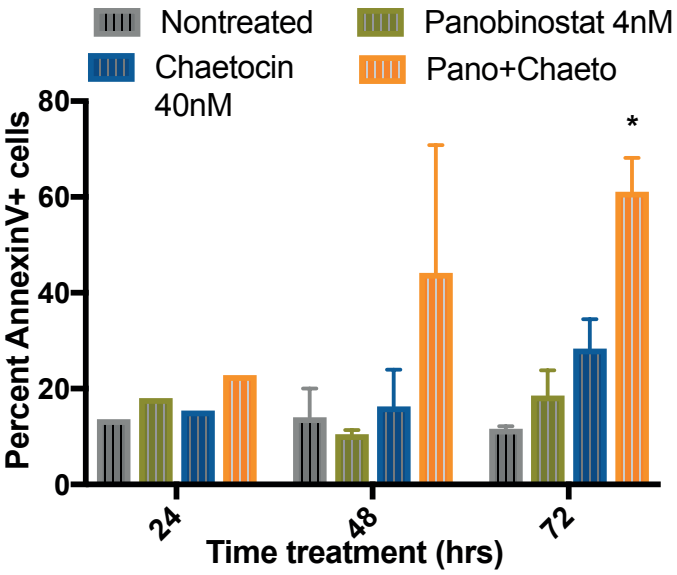
Supplementary Figure 14

MOLM13

A. Differentiation (72hr)

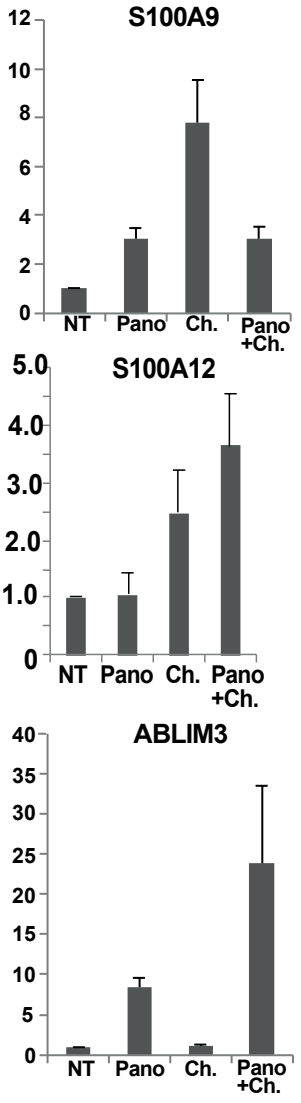


B. Apoptosis

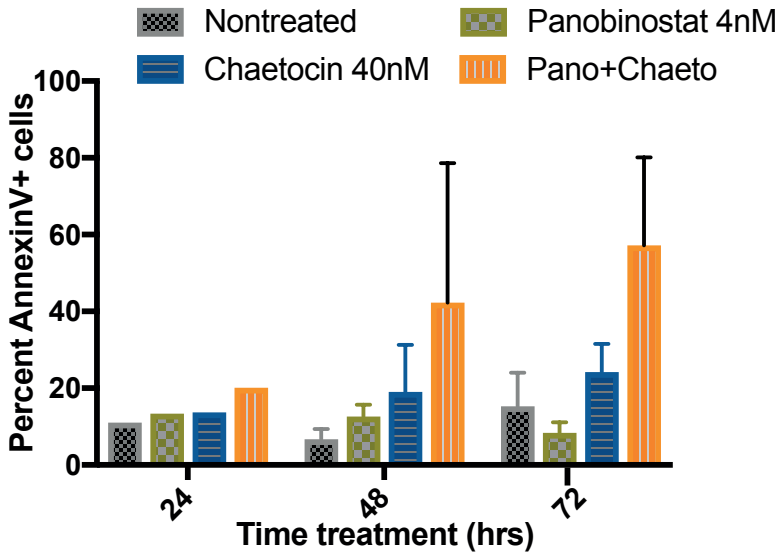


C. Gene expression

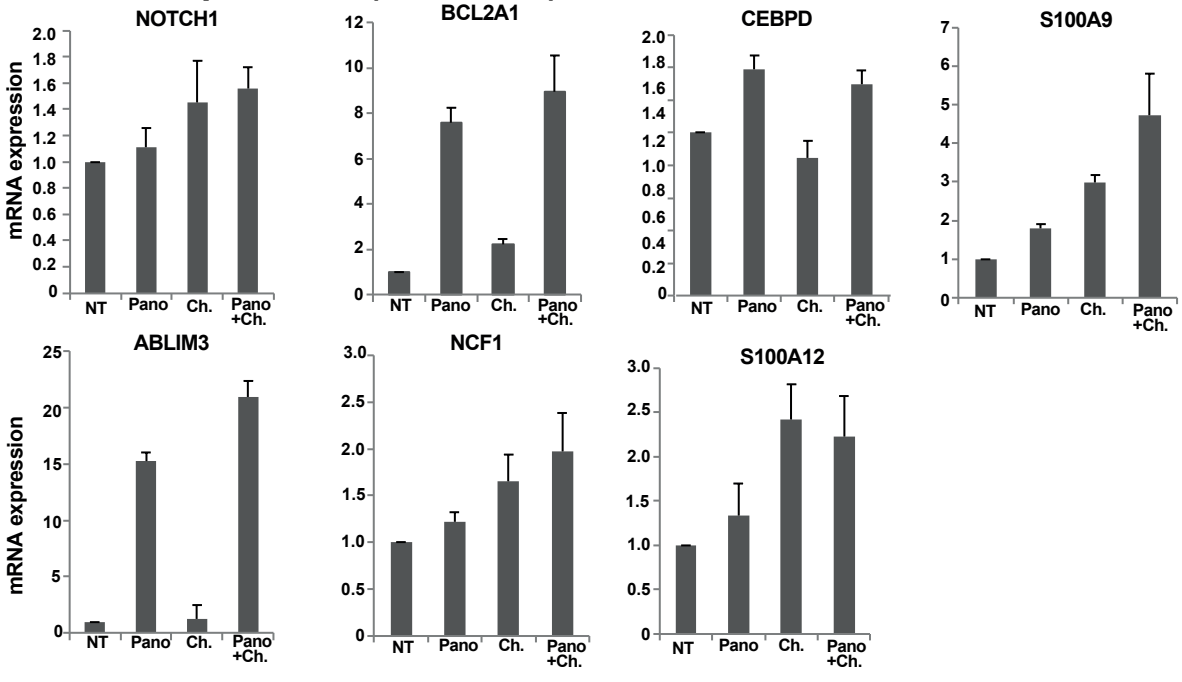
MOLM13



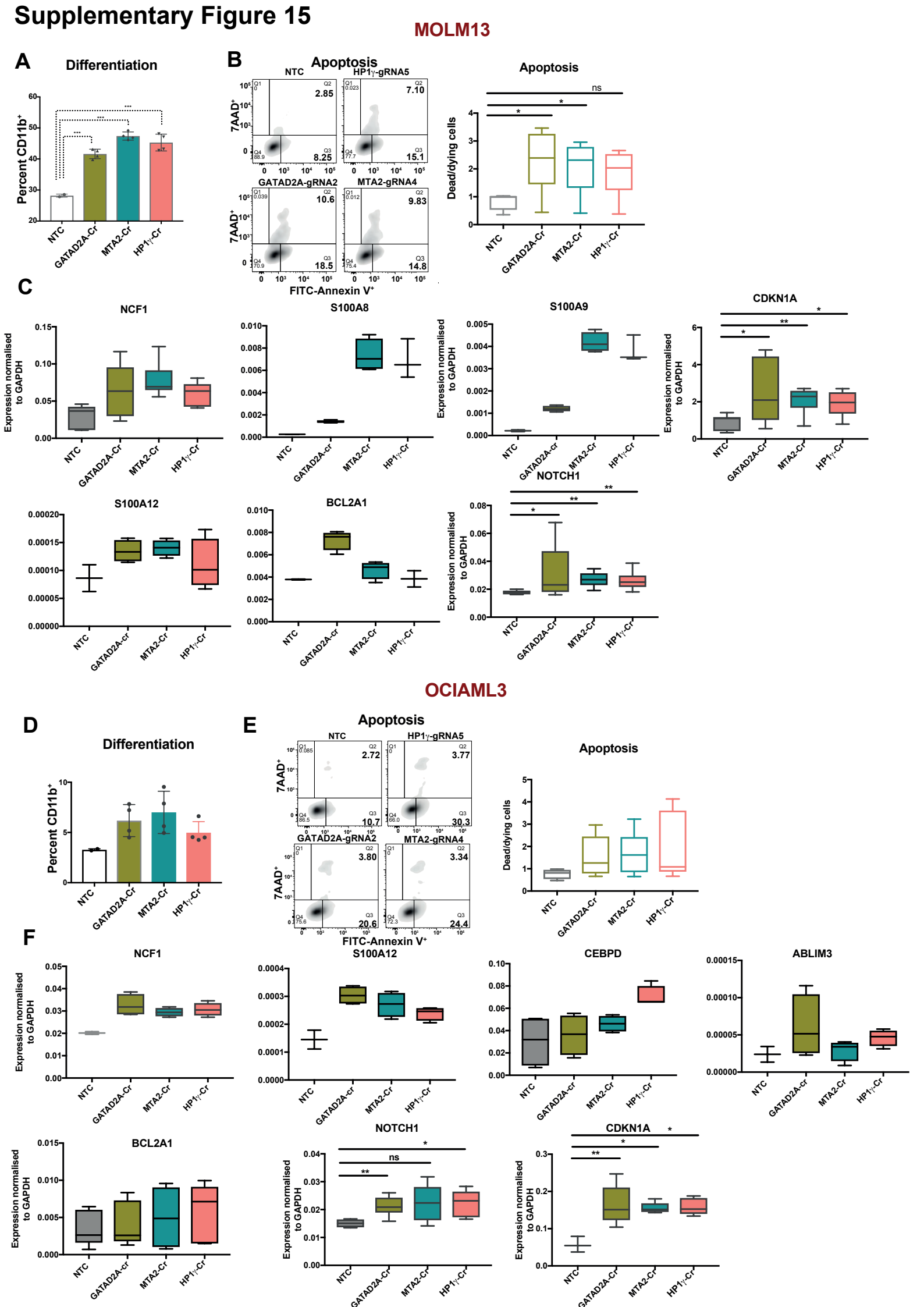
D. Apoptosis (OCIAML3)



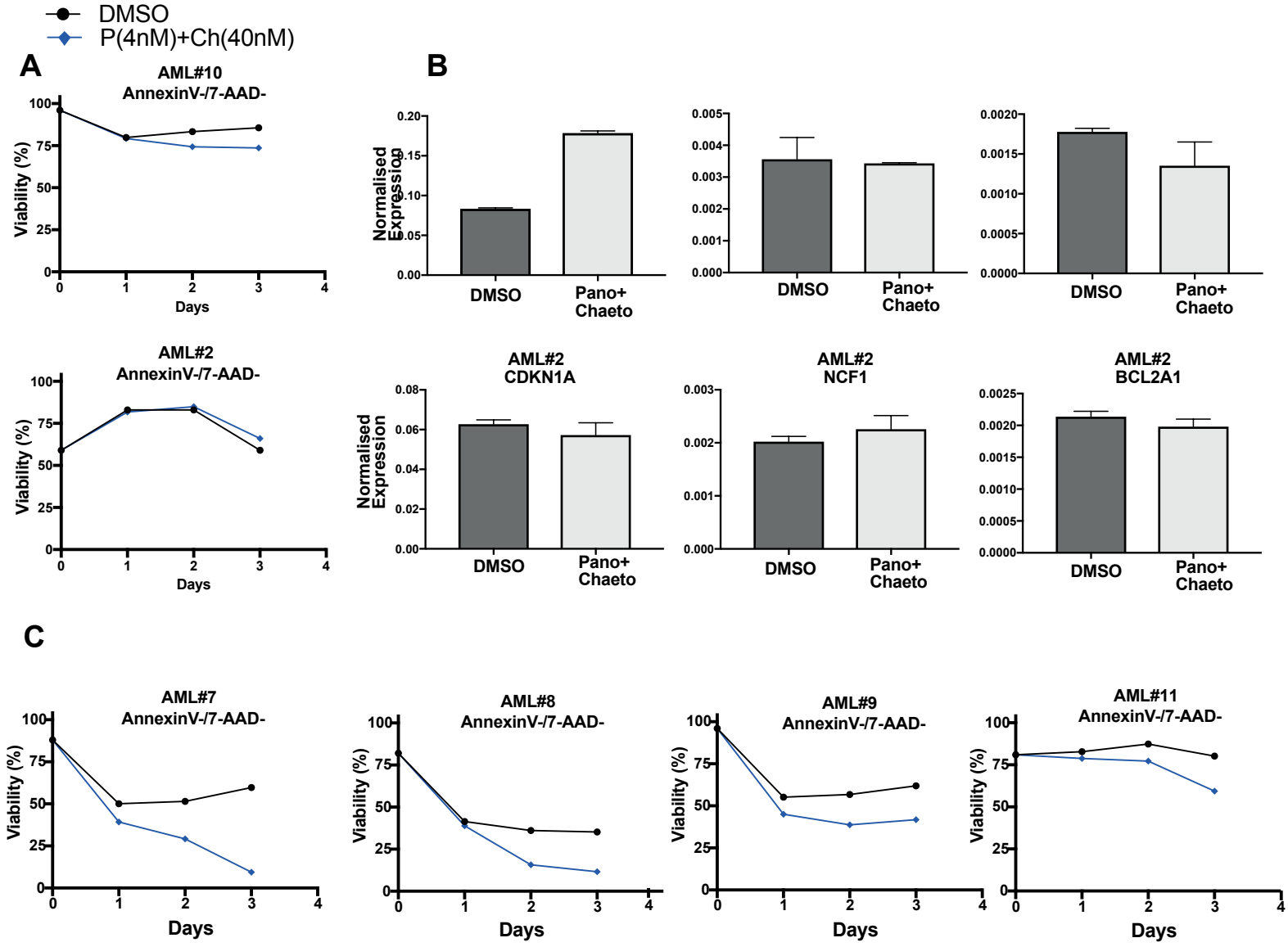
E. Gene expression (OCIAML3)



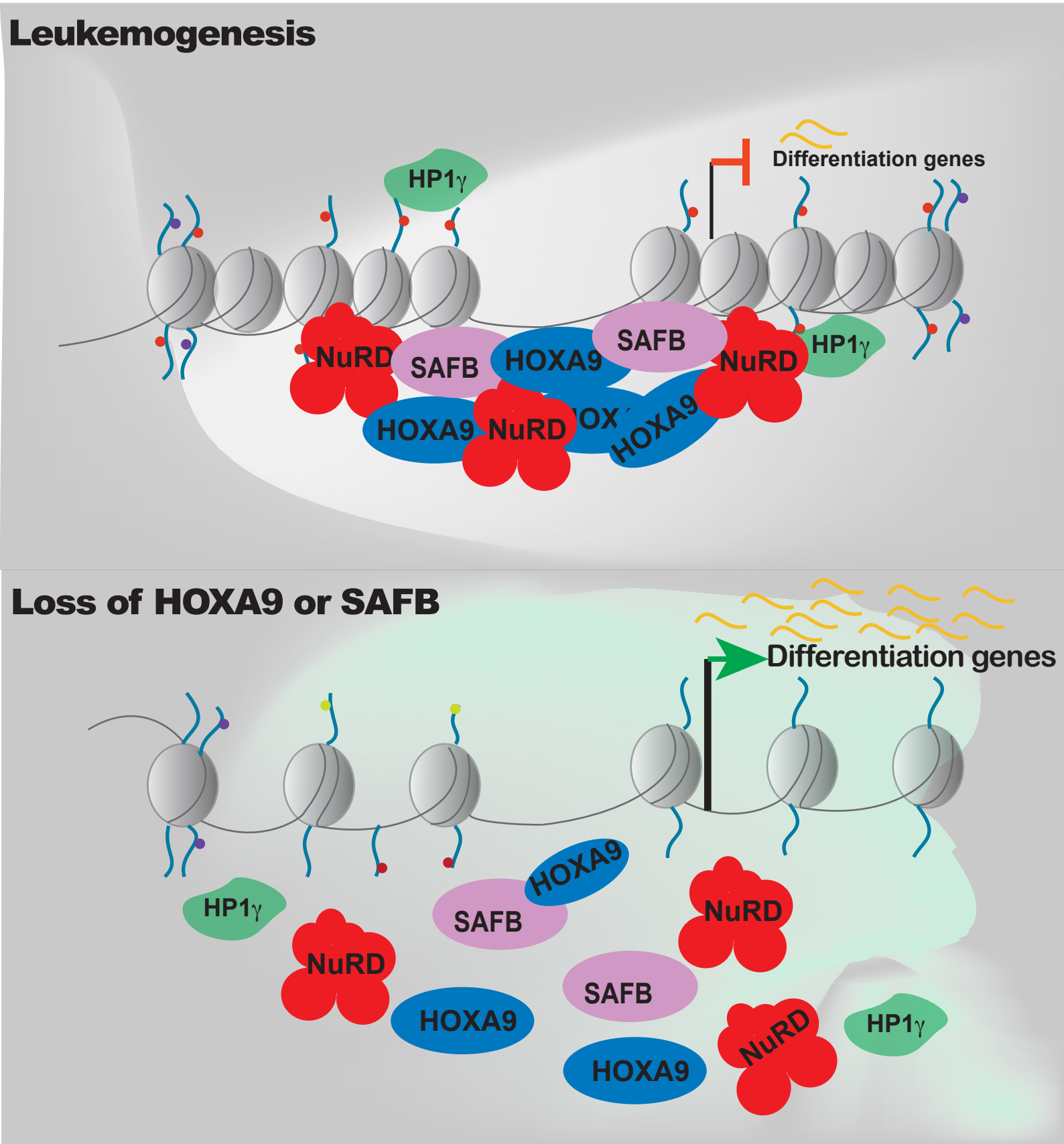
Supplementary Figure 15



Supplementary Figure 16



Supplementary Figure 17



Supplementary Figure Legend

Supplementary Figure 1.

A) Correlation between replicates (n=2) from mass spectrometric analyses of immunoprecipitated HOXA9 from MOLM13 cells. The histograms show the distribution of protein abundances in each sample separately.
B) Volcano plot displaying label free quantitative MS results of HOXA9 pull down in MOLM13 cells. The plot shows log2 ratios of averaged peptide MS intensities between HOXA9-IP and control-IP (IgG) eluate samples (x axis) plotted against the negative log10 p values (y axis) calculated across the replicate data sets (one tailed Student's t test, n=2 replicates). Note maximum upper values were set for the x and y axis to accommodate all detected proteins in the plot. The full dataset of specific co-precipitating proteins is given in Table S1.
C) Box plot represents depletion of SAFB, SATB1, SATB2 by CRISPR in a dataset across 14 human acute myeloid leukemia (AML) cell lines (Wang et al., 2017).
D) String network analysis demonstrating HOXA9 interacting proteins (SAFB-dependent) involved in chromatin architecture, (left panel, blue) or involved in negative regulation of gene expression (right panel, green).

Supplementary Figure 2.

A) Proximity ligation assay using Duolink in primary AML cells from 5 individual patient samples showing the interaction between HOXA9 and SAFB in situ. Antibodies against HOXA9 (rabbit polyclonal), and SAFB (mouse monoclonal) was used. Rabbit and mouse IgG were used as negative control.

Supplementary Figure 3.

A) The bar graph shows the percentage of S phase cells measured by Click-iT Plus EdU cell proliferation kit in shRNA expressing MOLM13 cells. A time course experiment was performed analysing cells at day3 and day8 of doxycycline treatment (1.5μg/ml).

B) Cytospins showing signs of myeloid differentiation in HOXA9 and SAFB shRNA cells.
C) RT-QPCR shows the expression of HOXA9 or SAFB after shRNA induction in MOLM13 cells.

Supplementary Figure 4.

A) Correlation matrix comparing peak overlap between the replicates of HOXA9 or SAFB enrichment from CUT&RUN sequencing in MOLM13 cells.
B) Average distribution plot shows the signal (intensity on Y-axis as normalised read count) for HOXA9 (blue) or SAFB (red) relative to IgG measured by CUT&RUN-sequencing in MOLM13 cells.
C) Heatmap comparison HOXA9 signal measured from ChIP or CUT&RUN using same antibody in MOLM13 cells. The Y-axis represents individual position of regions centred at HOXA9 peaks ±5kb around peak centre.
Right panel: Two representative loci are shown for HOXA9 enrichment by ChIP and CUT&RUN in MOLM13 cells.
D) Heatmap shows the enrichment signal for SFAB at HOXA9 occupied genomic regions in MOLM13 cells by CUT&RUN.
E) Motif enrichment for HOXA9-SAFB co-occupied regions obtained from CUT&RUN by HOMER. Short fragments (120bp) were used for motif search.

Supplementary Figure 5.

A) The bar graph shows presence of S/MAR features motif (Narwade et al., 2019) on the sequences obtained from HOXA9-SAFB co-bound genomic regions. Details are given in the method section.
B) Representative genomic loci showing HOXA9-SAFB occupancy at known classical S/MAR (shaded area).

Supplementary Figure 6.

A) GSEA enrichment plot shows HOXA9 signature enriched in differential gene expression data from HOXA9 or SAFB perturbation in MOLM13 cells.

B) Gene ontology analyses for upregulated genes in response to HOXA9 or SAFB perturbation in MOLM13 cells and possess proximal occupancy of HOXA9-SAFB (within 50kb to TSS). The gene count is plotted on X-axis. The colours on the bar represents the significance p-value. A bar represents the colour code is shown on the right side of the plot.
C) Gene ontology analyses for downregulated genes similar to as described in B.

Supplementary Figure 7.

We probed the transcriptional changes in response to *MEIS1* perturbation using CRISPR followed by RNA-sequencing in MOLM13 cells (Supplementary Figure 7A). The *MEIS1* perturbation led to differential expression of 1551 genes including *BCL2, ETV4* and *HES1* ($p<0.01$, FDR 0.1, Supplementary Figure 7B, supplemental table S5). Next the genomic binding of endogenous MEIS1 was analyzed by CUT&RUN in MOLM13 cells and identified 4936 annotated genes around MEIS1 genomic binding sites against our negative control (input). Exemplar tracks are shown in Supplemental Figure 7C. Integrating the transcriptomics (*MEIS1*-perturbation) and MEIS1 genomic binding in MOLM13 cells revealed 329 genes as direct target of MEIS1 (Supplementary Figure 7D). Among these 59% (194 genes) were downregulated including *BCL2, HES1, CCND1,* and *ETV4*, whereas 41% (135 genes) were upregulated upon *MEIS1* knockdown (Supplementary Figure 7E, supplemental table S5). Among these MEIS1 target genes, 75% (250/329) overlapped with HOXA9-regulated genes (Supplementary Figure 7F, supplemental table S5).

A) Western blot for MEIS1 to confirm the knockdown efficiency of gRNAs in MOLM13-Cas9 cells. β-Tubulin used as loading control.
B) Heatmap representing the RNA sequencing data from MEIS1 perturbation in MOLM13 compared to non-targeting control. Differentially expressed genes are shown as each row and columns represents the replicates from MEIS1 gRNAs or non-targeting gRNAs. ($p<0.01$, FDR 0.1, list of all DEG are given in supplemental table S5).
C) Genome browser track shows the peak co-localization of MEIS1 at selected loci BCL2, ETV4 and HES1 in the Hg38 genome, obtained from CUT&RUN sequencing in MOLM13 cells.

D) Venn diagram (top panel) shows the overlap between genes commonly dysregulated upon *MEIS1* perturbation in MOLM13 cells (orange) and genes that are linked to MEIS1 bound genomic regions in MOLM13 cell line (blue). List of these genes provided in Table S5. Venn diagram (lower panel) shows overlap between MEIS1 target genes (n=329) and HOXA9 target genes (n=6701) identified in MOLM 13 cells. Total of 250 genes were shared by HOXA9 and MEIS1. List of these genes are provided in Table S5.

E) The heatmap shows the expression Log2 fold change in (MEIS1-Cr Vs NT) MOLM13 cells for genes present in the vicinity of MEIS1 bound regions (n=329).

Supplementary Figure 8.

A) The UpSet plot showing the overlap within primary AML samples and the MOLM13 cell line to assess the intersection of HOXA9-SAFB bound genomic regions. The bar chart on lower left corner depicts the HOXA9-SAFB co-bound regions in individual primary AML cells or in MOLM13 cells.

Supplementary Figure 9.

A) Representative loci showing HOXA9 (blue) or SAFB (pink) enrichment from CUT&RUN in MOLM13 cells. Grey shaded area shows colocalization of HOXA9 and SAFB. Lower part of the tracks shows the transcripts signal measured from RNA-sequencing in MOLM13 cells with CRISPR knockdown of HOXA9 (blue) or SAFB (pink) relative to nontargeting control (NTC, grey).

B) Correlation between *HOXA9* and *S100A8* mRNA expression (left panel), *S100A9* mRNA expression (middle panel) and *BCL2A1* mRNA expression (right panel) in human AML primary samples (n=165, TCGA data set).

Supplementary Figure 10.

The correlation matrix showing high concordance between replicates (n=3) from mass spectrometric analyses of immunoprecipitated SAFB from MOLM13 cells. The histograms show the distribution of protein abundances in each sample separately. EC=Empty vector MOLM13 cells-IgG; ES= Empty vector MOLM13 cells-SAFB-IP; SAFB-C=SAFB-shRNA MOLM13 cells-IgG; SAFB-S=SAFB-shRNA MOLM13 cells-SAFB-IP.

Supplementary Figure 11.

A) The network displays the prediction of protein-protein interaction of HOXA9-SAFB interacting proteome. p value FDR =2.11e-28. K-mean clustering shows 3 clusters, Red (mainly contains splicing factor and RNA binding proteins), Green (Ribosomal proteins), Blue (chromatin bound gene repressors).
B) CUT&RUN for GATAD2A, MTA2 and HP1γ in MOLM13 cells in replicates (n=2). Correlation matrix comparing peak overlap between the technical replicates of CBX3, or GATAD2A, or MTA2 enrichment from CUT&RUN sequencing in MOLM13 cells.

Supplementary Figure 12.

A) Tracks of the histone modifications average signal (intensity on Y-axis as normalised read count) from CUT&RUN in MOLM13 cells centred at HOXA9-peaks (top panel) or HOXA9-SAFB-co-occupied regions (middle panel), SAFB-peaks (lower panel).
B) Average signal of indicated histone modification (intensity on Y-axis as normalised read count) active-H3K4me3 (top panel) centred at non-promoter or promoter bound HOXA9-SAFB-co-occupied regions determined by CUT&RUN in MOLM13 cells. Lower panel shows the average signal for H3K4me3 at random genomic regions of matched controls (matched for number, length and chromosome number) to HOXA9-SAFB bound genomic regions; intergenic (left panel), random promoters (right panel).

C) Average signal of indicated histone modification (intensity on Y-axis as normalised read count) repressive- H3K9me2, H3K9me3, H3K27me3, (top panel) centred at non-promoter or promoter bound HOXA9-SAFB-co-occupied regions determined by CUT&RUN in MOLM13 cells. Lower panel shows the average signal for H3K9me2, H3K9me3, H3K27me3 at random genomic regions of matched controls (matched for number, length and chromosome number) to HOXA9-SAFB bound genomic regions; intergenic (left panel), random promoters (right panel).

D) Gene ontology analyses for genes upregulated (n=444) in response to HOXA9 or SAFB perturbation in MOLM13 cells and possess proximal occupancy of HOXA9-SAFB, NuRD and HP1γ (within 50kb to TSS).

E) Genome browser track shows the peak co-localization of HOXA9, SAFB, NuRD complex (MTA2, GATAD2A), HP1γ and repressive histone modifications (H3K27me3, H3K9me2, H3K9me3) at selected loci LIPA in the Hg38 genome, obtained from CUT&RUN sequencing in MOLM13 cells. Upper two tracks shows the transcripts signal obtained from RNA sequencing in MOLM13 cells after *HOXA9* (blue) or *SAFB* (pink) perturbation. Transcripts signal for HOXA9 or SAFB- CRISPR samples are shown relative to non-targeting control (grey).

F) Genome browser track at loci FGL2.

G) Genome browser track at loci NOTCH1/C9orf163.

H) Genome browser track at loci CEBPD/SPIDR

Supplementary Figure 13.

A) The heatmap shows the expression Log2 fold change in (HOXA9-Cr Vs NT) or (SAFB-Cr vs NT) MOLM13 cells for genes present in the vicinity of HOXA9-SAFB/NuRD/HP1γ co-bound regions (n=635). All peaks were intergenic.

B) The heatmap shows the expression Log2 fold change in (HOXA9-Cr Vs NT) or (SAFB-Cr vs NT) MOLM13 cells for genes present in the vicinity of HOXA9-SAFB/HP1γ co-bound regions (n=35). All peaks were intergenic.

C) The heatmap shows the expression Log2 fold change in (HOXA9-Cr Vs NT) or (SAFB-Cr vs NT) MOLM13 cells for genes present in the vicinity of HOXA9-SAFB/NuRD co-bound regions (n=118). Peaks were promoter bound.

D) The heatmap shows the expression Log2 fold change in (HOXA9-Cr Vs NT) or (SAFB-Cr vs NT) MOLM13 cells for genes present in the vicinity of HOXA9-SAFB/NuRD co-bound regions (n=1512). These are Intergenic peaks.

Supplementary Figure 14.

A) Scatter plot shows flow-cytometric analyses of CD11b-PECy7 and CD15-APC surface expression in MOLM13 cells treated with Pano (4nM), Chaetocin (40nM) alone or in combination. Representative plots of 3 independent biological replicates are shown.

B) The bar graph shows percentage of apoptotic MOLM13 cells (AnnexinV positive) after treatment with Panobinostat, or Chaetocin alone or in combination in a time course experiment. The data are shown as average of biological replicates (n=3) ± SD. Statistical significance was calculated using 2way ANOVA test, * $p<0.01$.

C) RT-QPCR to measure expression of selected target genes in MOLM13 cells treated with drugs alone or in combination for 48hrs. NT: nontreated, Pano:Panobinostat, Ch.: Chaetocin, Pano+Ch.: Panobinostat+Chaetocin. The data shown here is representative of 3 independent biological replicates.

D) The bar graph shows percentage of apoptotic OCIAML3 cells (AnnexinV positive) after treatment with Panobinostat, or Chaetocin alone or in combination in a time course experiment. The data are shown as average of biological replicates (n=3) ± SD.

E) RT-QPCR to measure expression of selected target genes in OCIAML3 cells treated with drugs alone or in combination for 48hrs. NT: nontreated, Pano:Panobinostat, Ch.: Chaetocin, Pano+Ch.: Panobinostat+Chaetocin. The data shown here is representative of 3 independent biological replicates.

Supplementary Figure 15.

A) Left panel shows a bar graph of the differentiation measured by CD11b surface expression in MOLM13 cells after CRISPR-kd of *GATAD2A* (p=0.0038), *MTA2* (p=0.0095) and HP1γ (p=0.0048). The data are shown as mean±SD, n=3. Statistical significance was calculated using the t-test.

B) Density plots for apoptosis (left panel) measured by AnnexinV and 7AAD positive cells after CRISPR-kd of *GATAD2A, MTA2* and HP1γ. Representative plots of 3 independent biological replicates are shown. Right panel shows the bar graph for percentage of apoptosis (AnnexinV positive) in MOLM13 cells after CRISPR knockdown of GATAD2A, MTA2, CBX3. The data are shown as average of biological replicates (n=3) ± SD. Statistical significance was calculated using 2way ANOVA test, * p<0.05, **p<0.01.

C) RT-QPCR to measure expression of selected target genes in MOLM13 cells after CRISPR knockdown of GATAD2A, MTA2, CBX3. The data are shown as average of biological replicates (n=3) ± SD, a median line represents the frequency of expression.

D) Bar graph of the differentiation measured by CD11b surface expression in OCIAML3 cells after CRISPR-kd of *GATAD2A*, *MTA2* and HP1γ.

E) Same as B in OCIAMl3 cells.

F) Same as above in C, in OCIAML3 cells.

Supplementary Figure 16.

A) Percent viability determined by apoptosis measurement (AnnexinV staining) in primary AML cells (n=2) after treatment with combination of Panobinostat (4nM)+Chaetocin (40nM) or DMSO.

B) RT-QPCR expression levels of selected target genes in same primary AML cells treated with drugs in combination or DMSO for 48hrs. Pano+Ch.: Panobinostat+Chaetocin. The data shown here is average of triplicates of QPCR values.

C) Percent viability determined by apoptosis measurement (AnnexinV staining) in 4 more primary AML cells after treatment with combination of Panobinostat (4nM)+Chaetocin (40nM) or DMSO. No material was available to perform RT-QPCR from these samples.

Supplementary Figure 17.

The schematic model shows that the novel HOXA9-SAFB-chromatin repressive complex is involved in maintaining the differentiation blockade that is critical for the leukemic state in acute myeloid leukemia.

Narwade, N., Patel, S., Alam, A., Chattopadhyay, S., Mittal, S., and Kulkarni, A. (2019). Mapping of scaffold/matrix attachment regions in human genome: a data mining exercise. Nucleic Acids Res *47*, 7247-7261. 10.1093/nar/gkz562.

Wang, T., Yu, H., Hughes, N.W., Liu, B., Kendirli, A., Klein, K., Chen, W.W., Lander, E.S., and Sabatini, D.M. (2017). Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. Cell *168*, 890-903 e815. 10.1016/j.cell.2017.01.013.