Supplementary methods

Immunohistochemical analysis

To detect adipogenesis, cells were washed with PBS and fixed in 4% formaldehyde for 1 hour at RT, washed with PBS and then stained for 1 hr with fresh and filtered Oil-Red O solution (Sigma-Aldrich # O0625) composed of 3 parts of a 0.5% stock solution in isopropanol and 2 parts of distilled water. Then cells were washed three times with distilled water.

To detect osteogenesis, cells were washed with PBS, fixed with ice-cold 70% ethanol and incubated with filtered 2% (p/v in distilled water) Alizarin red solution (Sigma # A5533) for 15 min. Then cells were washed three times with distilled water.

To detect chondrogenesis, cells were washed with PBS and fixed in 4% formaldehyde for 1 hr at RT and then washed with PBS. Cells were incubated with Alcian blue solution (1 g/L in 0.1 M HCl, Sigma-Aldrich # B8438) for 6 hrs at room temperature and then extensively washed with PBS. To measure Alcian blue deposition, dry wells were incubated with 1 mL of 6 N Guanidine HCl for 1 hour and then the absorbance was measured with a spectrophotometer measured between 600 and 650 nm.

Labeling with CellTrace Violet

WT iMSCs were grown for 48h in complete culture medium followed by incubation for 20 minutes at 37°C in a PBS containing 5µM CellTrace Violet (Invitrogen #C34557) (staining solution). After incubation, the staining solution was removed and cells were washed twice with fresh culture medium and incubated for further 10 minutes in complete culture medium, at 37°C. Afterwards cells were harvested, counted and mixed in a 1:1 ratio with unstained MLL4^{Q4092X} MSCs. The WT/ MLL4^{Q4092X} iMSCs mix was seeded on coverslips for immunofluorescence assays.

3D Imaging cluster data analyses

In order to automatically detect and quantify PcG and TrxG complex proteins in fluorescence cell image z-stacks we developed an algorithm that implements a method derived from previous works [78] [79] [80], with variants and adaptations. The algorithm performs the 2D segmentation of cell nuclei and the detection of Protein Bodies (PBs) for each slice of the stack, followed by the 3D reconstruction and identification of nuclei and PBs. It then measures the volume of nuclei and the number and volume of the PBs and the relative positioning of PBs in the nucleus. The algorithm has been implemented in MATLAB following this scheme:

%Initializes threshold

t_{max} = -1;

for each slice n of the stack

I_{dapi+} = (I_{dapi} + ***medfilt2***($I_{PBfluo,n}$[4,4]))/2

%Performs 2D nuclei segmentation

[$nuclei_n$, $avg_{PBfluo,n}$]= ***nuclei_seg***($I_{dapi+,n}$,$I_{PBfluo,n}$)

$nuclei\_vol(:,:,n) = nuclei_n(:,:)$

$I_{avg,n}$ = ***imfilter***($I_{PBfluo,n}$,***fspecial***('average', [3,3]))

%Identifies PB regions with highest intensity

$I_{filt,n} = I_{PBfluo,n} - I_{avg,n}$

%Evaluates threshold for PBs detection

t_{temp} = ***isodata_thresh***($I_{filt,n}$,$avg_{PBfluo,n}$)

$t_{max} = max(t_{max}, t_{temp})$

endfor

for each slice n of the stack

%Performs 2D PBs detection

$PBs_n = I_{filt,n} > t_{max}$

$PBs\_vol(:,:,n) = PBs_n(:,:)$

endfor

nuclei_CC = **bwconncomp**(nuclei_vol)

nuclei_L = **labelmatrix**(nuclei_CC)

compute volume for each nucleus object in nuclei_CC

exclude nuclei whose volume is less than 10% of mean volumes

$\{NCL\}_M$ <- identified 3D nuclei

for each nucleus m in $\{NCL\}_M$

    %3D positions of detected PBs within the nucleus NCL_m

    NCL_m.PBs = PBs_vol .* $\{NCL\}_M$

    NCL_m.PBs = **bwareaopen**(NCL_m.PBs,17,6)

    PBs_CC = **bwconncomp**(NCL_m.PBs)

    compute number of PBs in PBs_CC

    compute volume for each PB in PBs_CC

    compute distances of any PB from the nuclear periphery

    compute distances of any PB from the nuclear centroid

endfor

I_{dapi} is the image showing the fluorescence of the nucleus while I_{PBfluo} is the image showing the fluorescence of PBs. *PBfluo* stands for BMI11 or RING1B in case of PcG and BRD4 or MLL4 in case of TrxG. In order to better enhance nucleus areas we added to I_{dapi} the image obtained by performing a median filtering of I_{PBfluo}, producing the image I_{dapi+}.

The function *nuclei_seg* performs a partition of cell image I_{dapi+} in nuclei regions and background implementing a region based segmentation algorithm [n], and computes avg_{PBfluo}, the mean intensity value of the nuclei regions in the image I_{PBfluo}.

In order to better enhance PB areas we subtract from the original I_{PBfluo} image its smoothed version obtained by applying an averaging filter of size 3, producing the image I_{filt}.

The function *isodata_thresh* implements the ISODATA classification algorithm [n] and uses relevant values computed by *nuclei_seg* function in order to extract PBs from the nuclei regions. It sets the initial threshold value of ISODATA method as avg_{PBfluo}.

For each slice of the stack, the algorithm separates PBs from nuclei regions by means of a thresholding operation using the maximum of the threshold values estimated by the function *isodata_thresh* applied to all the images $I_{filt,n}$.

PBs_vol and *nuclei_vol* are 3D arrays that contain the positions of the detected PBs and nuclei from all slices.

The 3D reconstructions of nuclei are obtained through the connected components algorithm (*bwconncomp* MATLAB function, using a connectivity of 26). 3D nuclei are then labeled by applying the *labelmatrix* MATLAB function so they can easily separate each from the others.

The algorithm computes the volume of each 3D reconstruction, discarding objects whose volume is less than 10% of mean volumes which are just noise.

The algorithm uses the *bwareaopen* function in order to discard too small (less than 17 pixels) detected PB objects which are probably just noise.

3D reconstructions of PBs are obtained through the connected components algorithm (*bwconncomp* MATLAB function, using a connectivity of 6).

The algorithm computes the number of PBs, the volume of any PB and the distances of the centroid of each PB from the nuclear periphery and the nuclear centroid.

Protein extraction and Western Blot analysis

For histone modifications, nuclear protein extracts were obtained as follows. Cells were washed twice with cold PBS, harvested by scraping in 1 ml cold PBS and centrifuged for 5 minutes at 1500 rpm. Pellet was resuspended in acid buffer (10mM Hepes pH 8, 10mM KCl, 0,1mM MgCl2, 0,1mM EDTA pH 8, 2mM PMSF, 0,1mM DTT) in order to have 10^7 cells/ml. Cells were left at 4°C

10 minutes and then were centrifuged for 10 minutes at 5000 rpm at 4°C. The supernatant (the cytosolic extracts) was discarded and the pellet was resuspended in 0,2N HCl in order to have $4x10^7$ cells/ml and left O/N at 4°C on rotating wheel. The day after, proteins were recovered by centrifugation for 10 minutes at 4000rpm at 4°C. Supernatant was recovered and protein concentration was measured with Bradford assay (Biorad #5000006) according to manufacturer's instructions.

Nuclear protein extracts were prepared in extraction buffer (Tris-HCl 50 mM pH8; NaCl 137.5 mM; EDTA 5 mM; 10% Glycerol; 0.5% Triton; 0.5% SDS). Harvested cell pellets were lysed by the addition of 6× v/v ice-cold extraction buffer for 15 min at 4 °C. The supernatant containing the cytoplasmic fraction was collected by centrifugation for 5 min at 100 xg, at 4 °C and stored. After two washes in extraction buffer, nuclear pellets were resuspended in 6× v/v ice-cold extraction buffer containing 0.5% SDS and sonicated. Lysates were cleared by centrifugation for 10 min at 21000 x g at 4°C and supernatant was collected. Western Blots were performed using the antibodies listed in Supplementary Table 4.

Recombinant protein purification

pET-mCherry-MLL4-PrLD/ PrLD ΔQ was subcloned from pET mCherry-MED1-IDR, a gift from Richard A. Young laboratory. Briefly, the MLL4 PrLD region (from amino acid 3560 to 4270) was PCR amplified and cloned between the BglII and SalI sites in the pET mCherry-MED1 IDR. The MLL4 PrLD ΔQ region was obtained by overlap-extension PCR and cloned between the BglII and SalI sites in the pET mCherry-MED1 IDR. The Protein purification was done using a standard protocol as follows: bacterial pellet was resuspended in 25mL of Ni-NTA Lysis Buffer (LB) (50 mM TrisHCl pH 7.5, 500mM NaCl), and sonicated. The lysate was cleared by centrifugation at 12,000g for 20 minutes at 4°C and added to Ni-NT Agarose (Qiagen, ID: 30210) pre-equilibrated with Ni-NTA LB. Tubes containing the agarose lysate slurry were rotated at

4°C for 1 hour. The agarose beads were collected by centrifugation for 5 min at 200 g and were transferred to the gravity columns. The protein-bound beads were further washed with the Ni-NTA LB containing 10mM Imidazole. Protein was then eluted with Ni-NTA LB containing 50/100/250 mM imidazole. The proteins were further purified over the gel filtration chromatography (Superdex 200 Increase 10/300 GL, GE Healthcare #28990944) and equilibrated with Buffer D (50mM Tris-HCl pH 7.5, 125mM NaCl, 1mM DTT, 10% glycerol). Peak fractions were pooled, aliquoted and concentrated using Pierce™ Protein Concentrator PES, 10K MWCO (Thermo Scientific™ #88527). The eluted fractions containing protein were finally analyzed by Coomassie stained gel.

In vitro phase separation assay was performed as previously described [9]. Briefly, recombinant protein was added to Buffer D containing 10% Polyethylene glycol (PEG) 8000 (Sigma # 1546605) at varying concentrations with indicated final salt and 1.6- hexanediol (Santa Cruz #sc-237791). The protein solution was immediately spotted into a glass slide and then covered with a coverslip. The solution was allowed to mix for 5 minutes at room temperature followed by imaging acquisition. Images of formed droplets were acquired using a Zeiss Axio Observer inverted microscope with an AxioCam 503 mono D camera and a Plan-Apochromatic 100x/1.4 oil-immersion objective equipped with a prism for DIC (Zeiss).

Images were analyzed with FIJI (http://fiji.sc/). We determine the intensity signal inside and outside the droplets by setting a threshold on the minimum intensity observed at the lower tested concentration where droplets formed, then this threshold was applied to every condition. The saturation concentration was quantified as previously described in Wang et al., [81]. We measured the fluorescence intensity inside the droplets (Idroplet) and the fluorescence intensity outside the droplets (Imedia) by summing with respectively the intensity of each pixel inside and outside droplets. The amount of condensed protein for a given candidate under a certain concentration is defined by the ratio of Idroplet to Imedia. If no droplets are present for

a certain condition the ratio is set to zero. Condensed protein appears only above the saturation concentration.

Quantification of Histone PTMs by LC-MS/MS

Histone samples (20ug) were suspended in 50 mM NH4HCO3 and subjected to chemical derivatization and digestion as previously described [82]. Briefly, propionic anhydride solution was freshly prepared by mixing propionic anhydride with acetonitrile in the ratio 1:3 (v/v), creating the propionylation mix. Next, propionylation mix was added to the histone sample in the ratio of 1:4 (v/v), immediately followed by NH4OH with a ratio of 1:5 (v/v) to adjust the pH to ~8.0. Samples were incubated for 15 min at 37°. Propionylation was repeated a second time after drying samples in a SpeedVac centrifuge. Samples were dried, dissolved in 50 mM NH4HCO3 and digested overnight with trypsin at an enzyme:sample ratio of 1:20. The digested peptides were then treated with an additional round of propionylation for the derivatization of peptide N-termini. After drying in a SpeedVac, the samples were desalted by C18 stage-tip, lyophilized, and resuspended in 20 µl of 0.1% formic acid for LC-MS/MS analysis.

Samples were analyzed using an EASY nLC 1200 ultra-high pressure liquid chromatography system (Thermo Scientific) coupled to an Orbitrap Fusion mass spectrometer (Thermo Scientific). Briefly, 1 µg of sample was loaded on a 25 cm long Acclaim PepMap RSLC C18 column (Thermo Fisher Scientific, 2µm particle size, 100Å pore size, id 75 µm) heated at 40°C. Mobile phase A was 0.1% formic acid, mobile phase B was 80% acetonitrile/0.1% formic acid (v/v). The gradient was as follows: from 2 to 34%B over 45 min, from 34 to 90% B in 5 min, and 90%B for 10 min at a flow rate of 300 nl/min. MS acquisition was performed using a data-independent acquisition (DIA) mode, with a 50 m/z isolation window within the 300-1100 m/z mass range).

To quantify histone PTMs, raw files obtained from the LC-MS runs were processed using EpiProfile , a software tool that performs extracted ion chromatography (XIC) of histone with a peak extraction mass tolerance set to 10 ppm. Once the peak area was extracted, the relative abundance of a given PTM was calculated by dividing its intensity by the sum of all modified and unmodified peptides sharing the same amino acid sequence.

Luciferase Assay

YAP/TAZ activity was determined by luciferase reporter assays. WT and MLL4^{Q4092X} MSCs were nucleofected with either the YAP/TAZ responsive luciferase reporter plasmid (8xGTIIC-luciferase, from Addgene #34615) or the empty vector (pGL4.27[luc2P/minP/Hygro]) and the pGL4-CMV-Renilla luciferase vector as a normalization control in a 30:1 ratio. Empty vector and normalizer vector were kindly provided by Dr. Robert Clarke. 1x10^6 cells were nucleofected using an Amaxa Nucleofector (program U-23, Lonza) and homemade buffer (KCl 5 mM, MgCl2 15mM, Glucose 1M, K2HPO4 120mM). After 24 hours of incubation, nucleofected cells were re-seeded as sparse (5000 cells/cm^2). After 24 hours, Firefly and Renilla luciferase activity was measured using the Dual-Luciferase® Reporter Assay System following manufacturer's instructions.

Total RNA isolation

WT and MLL4^{Q4092X} MSCs were seeded as sparse (5000 cells/cm^2) condition and collected 48h after plating, either untreated or at 8 and 24 hours after ATR inhibition (VE-822 treatment) (medchem express # HY-13902/cs-1861). Cells were directly lysed on plates with TRIzol (Thermo Fisher cat. #15596026), and total RNAs were extracted according to the manufacturer's instructions.

Gene expression analysis

Quantitative real-time PCR analysis was performed with SuperScript III One-Step SYBR Green kit (Invitrogen #11746). Relative gene expression levels were determined using the comparative Ct method, normalizing data on endogenous GAPDH expression levels. The oligonucleotides used in this work for gene expression analysis are listed in Supplementary Table 3.

RNA-seq library preparation

Contaminating genomic DNA was removed by DNase (Qiagen cat. #79254) digestion. RNA quality and concentration were assessed using the 2100 Bioanalyzer (Agilent cat. #G2939BA) and the Qubit fluorometer (Thermo Fisher cat. # Q33226), respectively. 3'-RNA-seq libraries were prepared by using the QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen cat. #015.24) and starting from 500 ng of total RNA. Libraries were sequenced as single reads of 50 bp with the Illumina HiSeq2500. Three independent biological replicates were performed for each cell line and time point and sequenced as independent libraries.

Differential Gene expression analysis

Raw reads from fastq files were first checked for their quality using FastQC and trimmed using Trimmomatic [83]. Trimmed reads were aligned to the reference human genome assembly hg38/GRCh38 using STAR with default parameters. Resulting bam files were converted to bed format by using bedtools with the command "bamtobed". For annotation of the reads to the genome, HOMER was employed with the following command 'analyzeRepeats.pl rna hg38 -count 3utr –rpkm'. Genes were considered expressed with rpkm > 1 and used for subsequent analysis. Differential expression analysis was performed using DESeq2 within the HOMER environment.

Computational Analysis of Gene Expression Data

Correlation heatmaps and trajectories of gene expression data were performed by using Clust and visualized using the Multi Experiment Viewer (MeV). Volcano Plots were performed within the R environment.

Gene Ontology (GO) term analysis

Differentially expressed genes in wild-type (WT) compared to MLL4^{Q4092X} MSCs and identified gene clusters from this study and publicly available gene lists [84 85] were used as input for GO term and Reactome pathway analysis with EnrichR. Results were plotted using GraphPad PRISM.

Supplementary references

78. Gregoretti, F., Cesarini, E., Lanzuolo, C., Oliva, G. & Antonelli, L. An Automatic Segmentation Method Combining an Active Contour Model and a Classification Technique for Detecting Polycomb-group Proteinsin High-Throughput Microscopy Images. *Methods Mol Biol* **1480**, 181-97 (2016).
79. Cesarini, E. *et al.* Lamin A/C sustains PcG protein architecture, maintaining transcriptional repression at target genes. *J Cell Biol* **211**, 533-51 (2015).
80. Bianchi, A. *et al.* Dysfunctional polycomb transcriptional repression contributes to lamin A/C-dependent muscular dystrophy. *J Clin Invest* **130**, 2408-2421 (2020).
81. Wang, J. *et al.* A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688-699 e16 (2018).

82. Sidoli, S., Bhanu, N.V., Karch, K.R., Wang, X. & Garcia, B.A. Complete Workflow for Analysis of Histone Post-translational Modifications Using Bottom-up Mass Spectrometry: From Histone Extraction to Data Analysis. *J Vis Exp* (2016).

83. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-20 (2014).

84. Zanconato, F. *et al.* Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. *Nat Cell Biol* **17**, 1218-27 (2015).

85. Provenzano, P.P., Inman, D.R., Eliceiri, K.W. & Keely, P.J. Matrix density-induced mechanoregulation of breast cell phenotype, signaling and gene expression through a FAK-ERK linkage. *Oncogene* **28**, 4326-43 (2009).