

## Supplementary Information

### **Replication collisions induced by de-repressed S-phase transcription are connected with malignant transformation of adult stem cells**

Ting Zhang<sup>1</sup>, Carsten Künne<sup>1,2</sup>, Dong Ding<sup>1</sup>, Stefan Günther<sup>1,3</sup>, Xinyue Guo<sup>1</sup>,  
Yonggang Zhou<sup>1</sup>, Xuejun Yuan<sup>\*,1</sup> and Thomas Braun<sup>\*,1,4,5</sup>

<sup>1</sup>Department of Cardiac Development and Remodeling, Max-Planck-Institute for Heart and Lung Research, Bad Nauheim, Germany

<sup>2</sup>Bioinformatics Core Unit (BCU), Max Planck Institute for Heart and Lung Research, Bad Nauheim, Germany

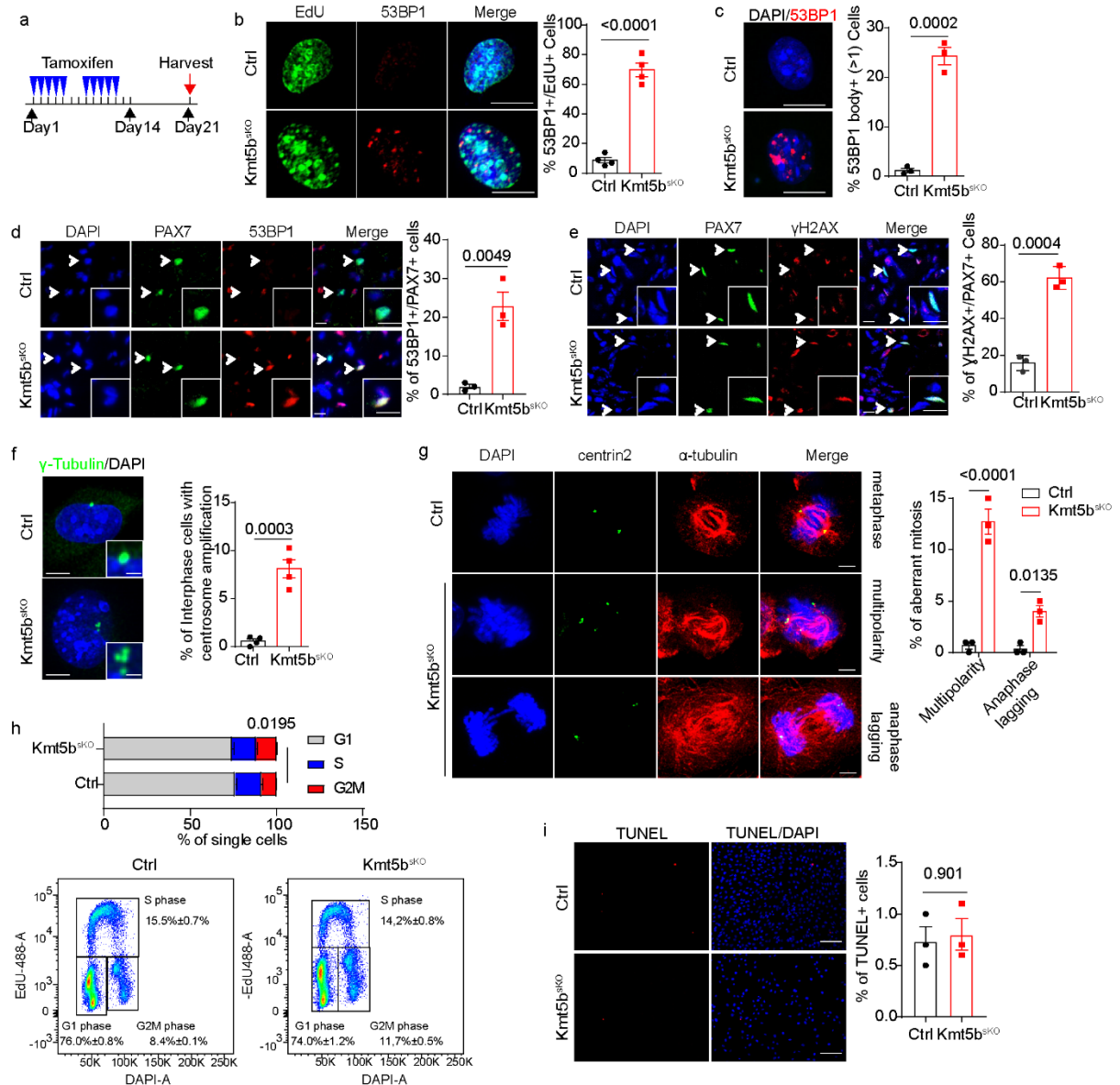
<sup>3</sup>CPI Deep Sequencing Platform, Max Planck Institute for Heart and Lung Research, Bad Nauheim, Germany

<sup>4</sup>German Center for Cardiovascular Research (DZHK), Rhine Main, Germany

<sup>5</sup>German Center for Lung Research (DZL), Giessen, Germany

\*Corresponding authors: Xuejun Yuan and Thomas Braun; [xuejun.yuan@mpi-bn.mpg.de](mailto:xuejun.yuan@mpi-bn.mpg.de),  
[thomas.braun@mpi-bn.mpg.de](mailto:thomas.braun@mpi-bn.mpg.de)

<b>Supplementary Information</b>	<b>Page</b>
Supplementary Figure 1	3
Supplementary Figure 2	5
Supplementary Figure 3	7
Supplementary Figure 4	9
Supplementary Figure 5	11
Supplementary Figure 6	13
Supplementary Figure 7	15
Supplementary Figure 8	17
Supplementary Figure 9	19
Supplementary Figure 10	21



**Supplementary Figure 1: *Kmt5b* prevents aberrant mitosis and genomic instability in proliferating MuSC.**

(a) Schematic description of tamoxifen administration in *Ctrl* and *Kmt5b<sup>SKO</sup>* mice. Mice were analyzed at least 7 days after the last tamoxifen injection.

(b) Immunofluorescence staining for EdU and 53BP1 in *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs after 4 days in culture (n=4 mice). DNA was stained by DAPI. Scale bar: 10 μm. Quantification of 53BP1<sup>+</sup>/EdU<sup>+</sup> is on the right.

(c) Images of 53BP1 nuclear bodies in nuclei of *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs in G1 phase (EdU<sup>-</sup> with 2N DNA content based on DAPI intensity) (n=3 mice). DNA was stained by DAPI. Scale bar: 10 μm. Quantification of G1-phase MuSCs with >1 53BP1<sup>+</sup> nuclear bodies is on the right.

(d) & (e) Immunofluorescence staining for (d) 53BP1 and (e) γH2AX in Pax7<sup>+</sup> MuSCs (indicated by arrowheads) in TA muscles of *Ctrl* and *Kmt5b<sup>SKO</sup>* mice, 7-days after CTX injections. Scale bar: 10 μm. Quantification of (d) 53BP1<sup>+</sup>/PAX7<sup>+</sup> and (e) γH2AX<sup>+</sup>/PAX7<sup>+</sup> MuSCs detected in 10 sections from each animal in the right panel (n=3 mice).

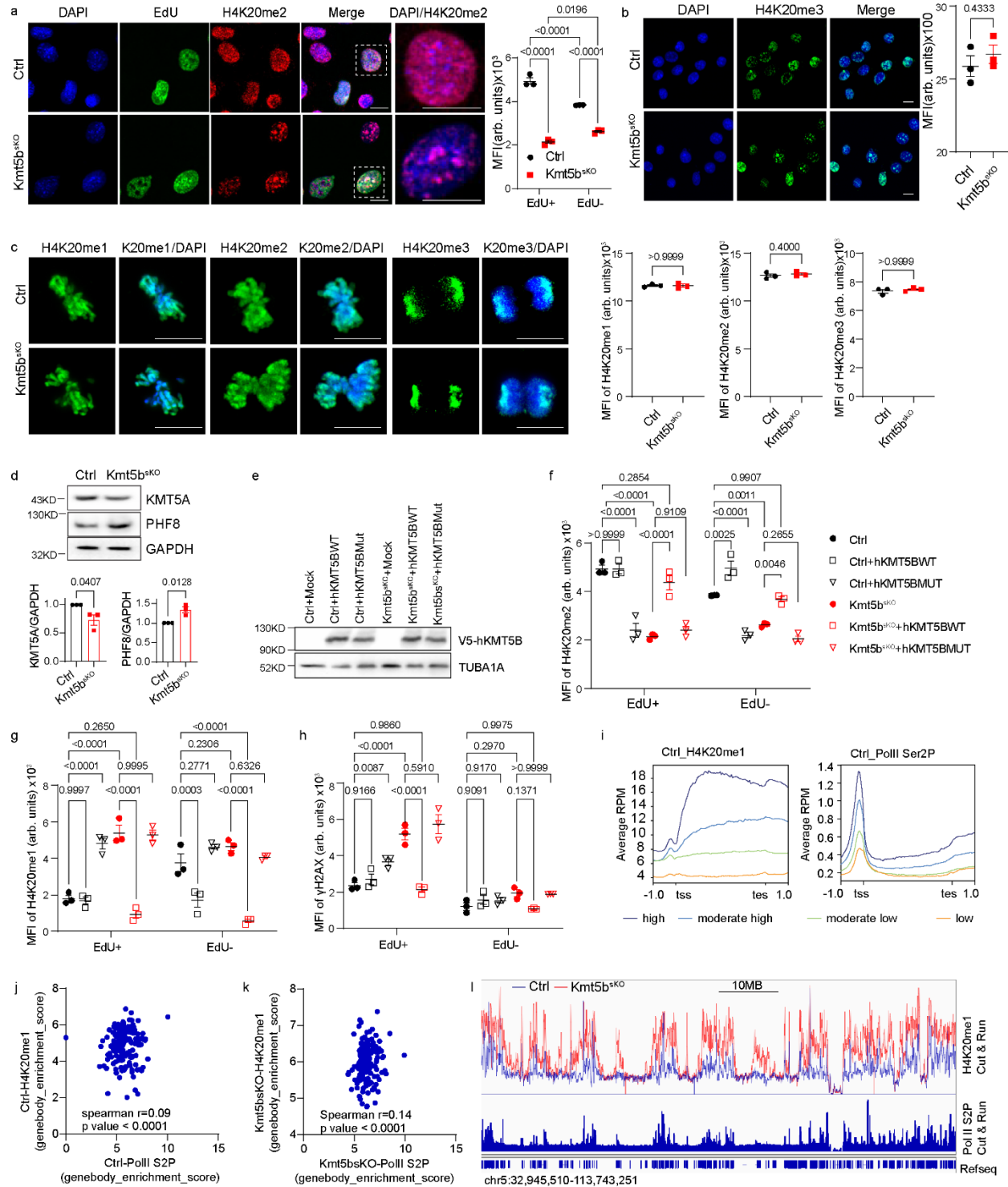
(f) Immunofluorescence analysis of  $\gamma$ -tubulin demonstrating amplified centrosome in *Kmt5b<sup>SKO</sup>* MuSCs. Scale bar: 5  $\mu\text{m}$ . Scale bars in the magnified images represent 1  $\mu\text{m}$ . Quantification of interphase MuSCs with amplified centrosome ( $>2$ ) in *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs (n=4 mice) is on the right.

(g) Immunofluorescence staining for centrin2 and  $\alpha$ -tubulin revealing multipolar metaphases and bridging anaphases in *Kmt5b<sup>SKO</sup>* MuSCs in comparison to *Ctrl* MuSCs after 4 days in growth medium (n=3 mice). Scale bar: 2  $\mu\text{m}$ . Quantification of cells with aberrant mitosis is on the right.

(h) Flow cytometry-based cell cycle analysis by EdU/DAPI staining in *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs after 4 days in culture (n=3 mice). Quantification of percentages of cells in different cell cycle phase is shown in the upper panel.

(i) TUNEL assay of 4-days cultured *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs (n=3 mice). Scale bar: 100  $\mu\text{m}$ . Quantification of TUNEL<sup>+</sup> cells is on the right.

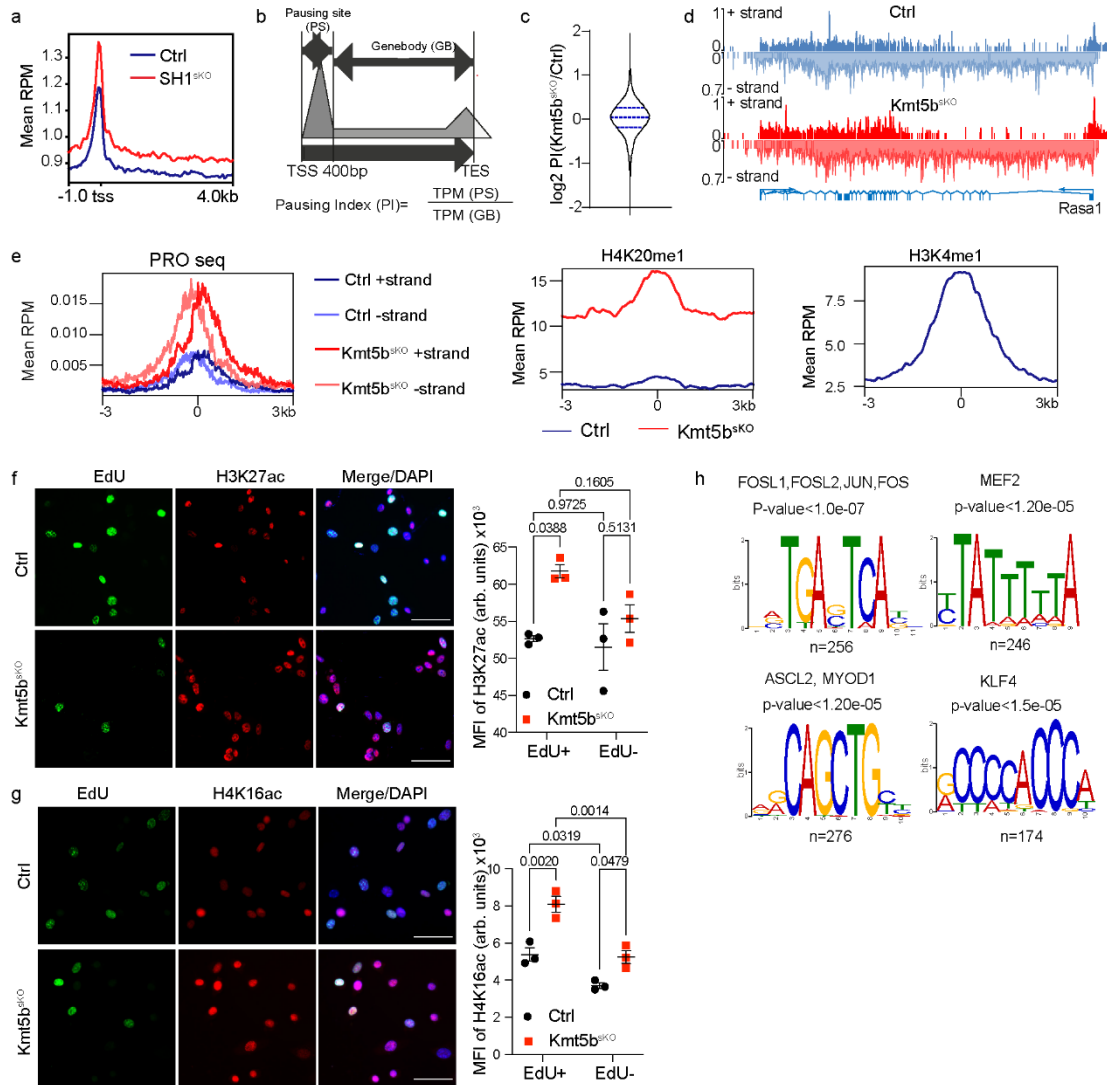
Data are shown as mean  $\pm$  SEM. Two-sided p-values for unpaired Student's t-tests are indicated in the figure. Source data are provided in the Source Data file.



## Supplementary Figure 2: Loss of *Kmt5b* alters the extent of H4K20 methylation in MuSC.

(a) Immunofluorescence analysis of EdU incorporation and H4K20me2 level in *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs. DNA was stained by DAPI. Magnified images of DAPI/H4K20me2 staining in selected EdU<sup>+</sup> MuSCs are shown. Scale bar: 10  $\mu$ m. Median of fluorescence intensity (MFI, arbitrary units) of H4K20me2 in EdU<sup>+</sup>/EdU<sup>-</sup> cells is shown on the right (n=3 mice). Two-way ANOVA.

(b) Immunofluorescence analysis of H4K20me3 in *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs after 4-days in culture. DNA was stained by DAPI. Scale bar: 10  $\mu$ m. Median of fluorescence intensity (MFI, arbitrary units (arb. units)) of H4K20me3 is shown on the right (n=3 mice). Unpaired two-sided Student's t-test. (c) Immunofluorescence analysis of H4K20me1, H4K20me2, and H4K20me3 in *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs during metaphase. DNA was stained by DAPI. Scale bars: 10  $\mu$ m. Quantification of MFI is shown on the right (n=3 mice). Unpaired two-sided Student's t-test. (d) Western blot analysis of KMT5A and PHF8 in *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs (n=3 mice). GAPDH was used as internal control. Unpaired two-sided Student's t-test. (e) Western blot analysis of overexpressed wildtype (WT) and catalytically inactive (MUT) hKMT5B with V5 tag in *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs (n=3 mice). TUBA1A served as the loading control. (f-h) Immunofluorescence analysis (MFI) of (f) H4K20me2, (g) H4K20me1, and (h)  $\gamma$ H2AX in EdU<sup>+</sup>/EdU<sup>-</sup> *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs after transduction with empty, wildtype (WT) and catalytically-inactive (MUT) hKMT5B vectors (n=3 mice). Two-way ANOVA. (i) Average levels of H4K20me1 (left panel) and Pol II S2P (right panel) in four groups of genes by abundance of H4K20me1 CUT&RUN signals within gene bodies of all coding genes in proliferating *Ctrl* MuSC. (j, k) Enrichment of H4K20me1 and Pol II S2P binding on bodies of genes coding for small non-coding RNAs, including rRNAs, snRNAs, snoRNAs, and scaRNAs in *Ctrl* (j) and *Kmt5b<sup>SKO</sup>* (k) MuSCs. Spearman's correlation coefficient r-values are indicated. Two-sided P-values indicate the significance of the correlation determined by r values and sample size. (l) IGV snapshots of a representative region on mouse chromosome chr5, showing CUT&RUN signals for H4K20me1 in *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs (upper panel) and Pol II S2P in *Ctrl* MuSCs (lower panel). Data are shown as mean  $\pm$  SEM. P values are indicated in the figure. Source data are provided in the Source Data file.



**Supplementary Figure 3: PRO-seq analysis of *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs and identification of transcription factor binding sites in enhancer regions showing increased eRNA transcription.**

(a) Average profiles of PRO-seq at the TSS and 4kb downstream of 8020 genes with increased accumulation of H4K20me1 and Pol II S2P. TSS: transcription start site

(b) Illustration of the calculation for the polymerase II pausing index using the ratios of densities of PRO-seq signals at pausing sites (TSS to TSS +500bp) vs gene bodies (TSS +500bp to TES). TSS: transcription start site; TES: transcription termination site.

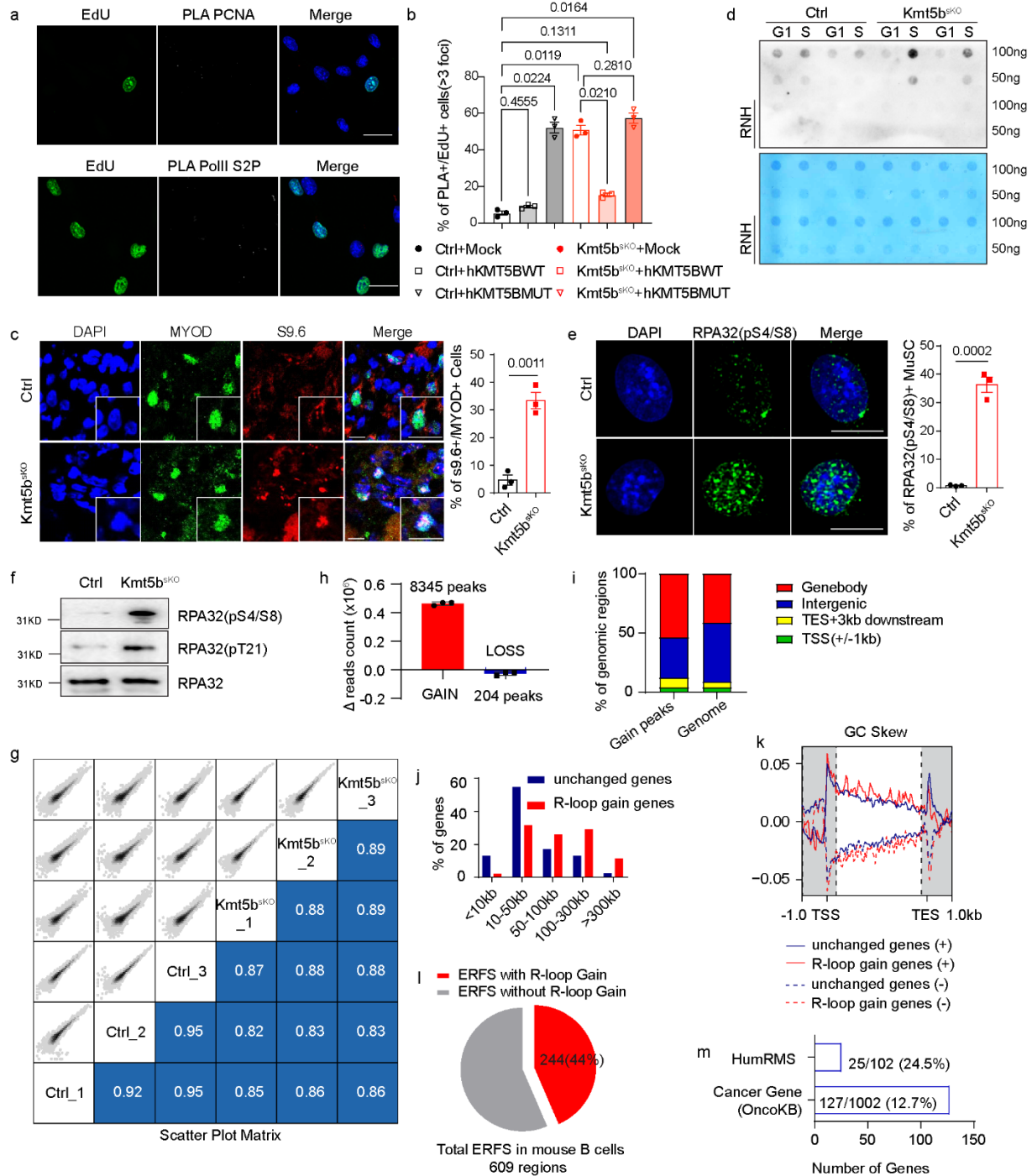
(c) Violin plot of the log<sub>2</sub> pausing index (*Kmt5b<sup>SKO</sup>*/*Ctrl*). The blue lines indicate the upper/lower adjacent values defining third/first quartiles and the median.

(d) Genome browser track of PRO-seq signals in the *Ras1* gene locus in *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs. Positive and negative values on the y-axis represent library-normalized reads mapping to the plus and minus strands, respectively.

(e) Average profiles of PRO-seq, H4K20me1 CUT&RUN around enhancer regions (±3kb from the center) in *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs. H3K4me1 ChIP-seq signals (BioProject: PRJNA412267) were used to identify enhancer regions.

(f, g) Immunofluorescence analysis of (f) H3K27ac and (g) H4K16ac in EdU<sup>+</sup>/EdU<sup>-</sup> *Ctrl* and *Kmt5b*<sup>skO</sup> MuSCs after 4-days in culture. DNA was stained by DAPI. Scale bar: 50 μm. Median of fluorescence intensity (MFI, arbitrary units (arb. units)) of (f) H3K27ac and (g) H4K16ac in EdU<sup>+</sup>/EdU<sup>-</sup> cells is shown on the right (n=3 mice). Data are shown as mean ± SEM. Two-way ANOVA.

(h) Identification of enriched transcription factor binding motifs by MEME and DMEME within the center of enhancers regions with increased eRNA transcription after loss of *Kmt5b*. Corresponding transcription factors were identified by TOMTOM. Source data are provided in the Source Data file.



### Supplementary Figure 4: Increased R-loop formation and DNA damage responses in MuSCs after loss of *Kmt5b*.

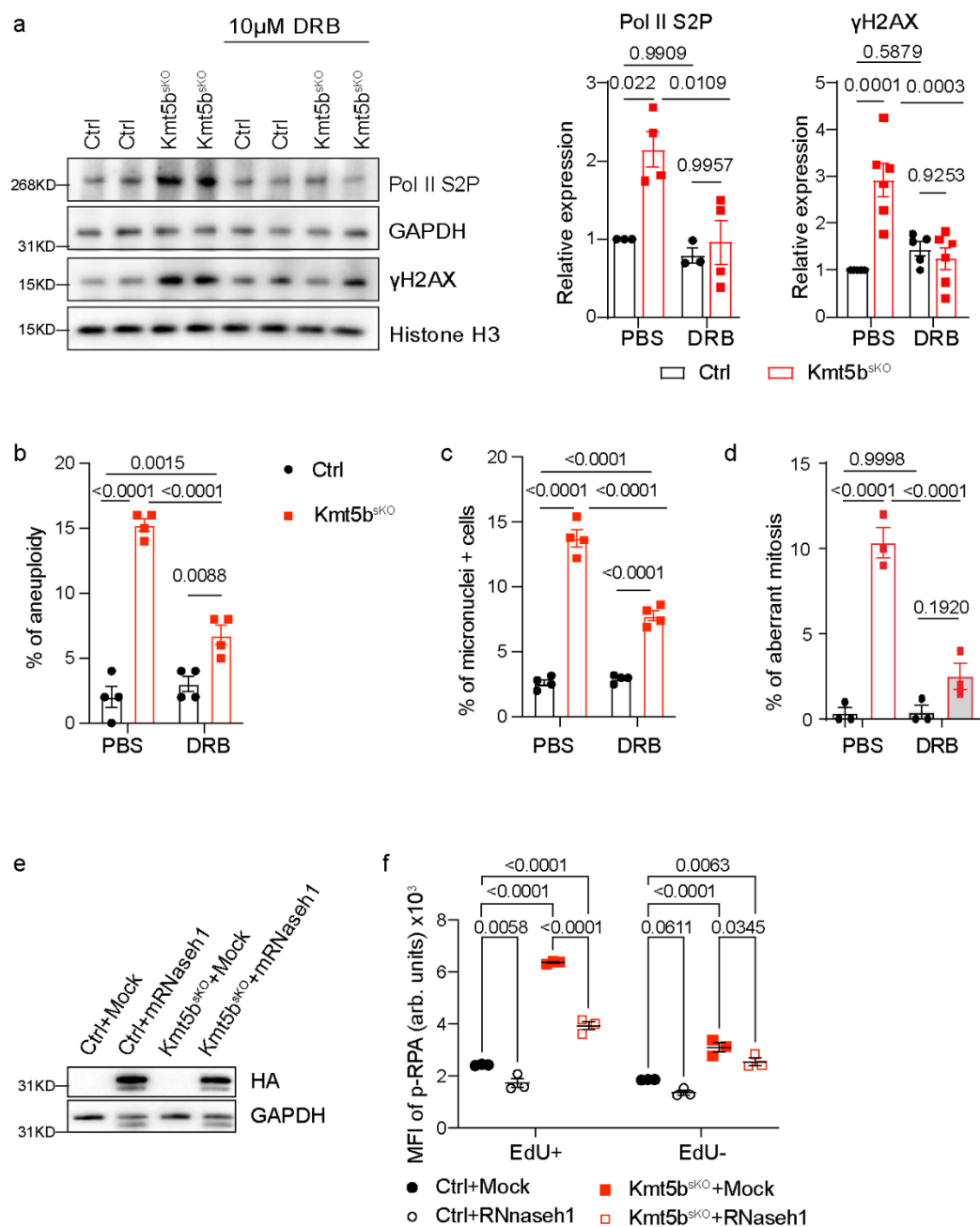
(a) Representative images of controls (PCNA or PolII S2P) using a single antibody in the PLA assay together with EdU labeling. Scale bar: 20  $\mu$ m.

(b) PLA of PCNA and PolII S2P in EdU<sup>+</sup> *Ctrl* and *Kmt5b*<sup>shKO</sup> MuSCs after transduction of empty vector, wildtype (WT) and catalytically-inactive (MUT) hKMT5B (n=3 mice). Data represent mean  $\pm$  SEM. One-way ANOVA.

- (c) Immunofluorescence staining for MYOD and S9.6 in MuSCs from regenerating TA muscles of *Ctrl* and *Kmt5b<sup>SKO</sup>* mice, 7 days after CTX injection (n=3 mice). DNA was stained by DAPI. Scale bar: 20  $\mu$ m. Quantification of S9.6<sup>+</sup>/MYOD<sup>+</sup> cells is shown on the right. Data represent mean  $\pm$  SEM. Unpaired two-sided Student's t-test.
- (d) Dot blot of R-loops in sorted G1 and S-phase cells from *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs.
- (e) Immunofluorescence analysis of RPA32 pS4/S8 in *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs after 4 days in culture (n=3 mice). Quantification of RPA32 pS4/S8<sup>+</sup> MuSCs is on the right. Scale bar: 10  $\mu$ m. Data represent mean  $\pm$  SEM. Unpaired two-sided Student's t-test.
- (f) Western blot analysis of RPA32 pS4/S8 and RPA32 pT21 in *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs after 4-days in culture. Total RPA32 served as internal control (n=2 mice).
- (g) Spearman rank correlations between pairwise comparisons of DRIP-seq peaks from replicates. The upper-left section of the diagram provides smoothed scatter plot data whereas the lower-right section displays rank correlation values.
- (h) Relative differences of uniquely mapped reads between increased (5134) and decreased (210) R-loop peaks when comparing *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSC. Newly acquired R-loops dominate over losses (n=3 mice). Data represent mean  $\pm$  SEM.
- (i) Distribution of newly acquired DRIP-seq peaks in percentages across several genomic regions after *Kmt5b* inactivation.
- (j) Distribution of gene length of 1508 coding genes with newly acquired R-loops and enrichment of H4K20me1 and Pol II S2P in comparison with coding genes without changes in R-loop formation.
- (k) GC skews over gene bodies of genes with newly acquired R-loops and enrichment of H4K20me1 and Pol II S2P in comparison with coding genes without changes in R-loop formation.
- (l) Pie chart showing the ratio of fragile early replication sites (ERFS) with newly acquired R-loops. The number of ERFS determined in B-cells is indicated.
- (m) Representation of genes with newly acquired R-loops corresponding to orthologous mutated cancer genes annotated by OncoKB in human rhabdomyosarcomas. Percentages for each category are indicated. Source data are provided in the Source Data file.



**Supplementary Figure 5: Heatmaps of cancer-related genes with newly acquired R-loops.**  
**(a-d)** Gene ontology and KEGG pathway analysis of cancer-related genes with newly acquired R-loops. Genes with newly acquired R-loops fall in different categories including cellular response to DNA damage response (a), cell migration (b), pathway in cancer (c), and RAS signaling (d). The color key in the heatmaps reflects the frequency of R-loop formation.



### Supplementary Figure 6: Inhibition of S-phase transcription ameliorates aberrant mitosis and genome instability

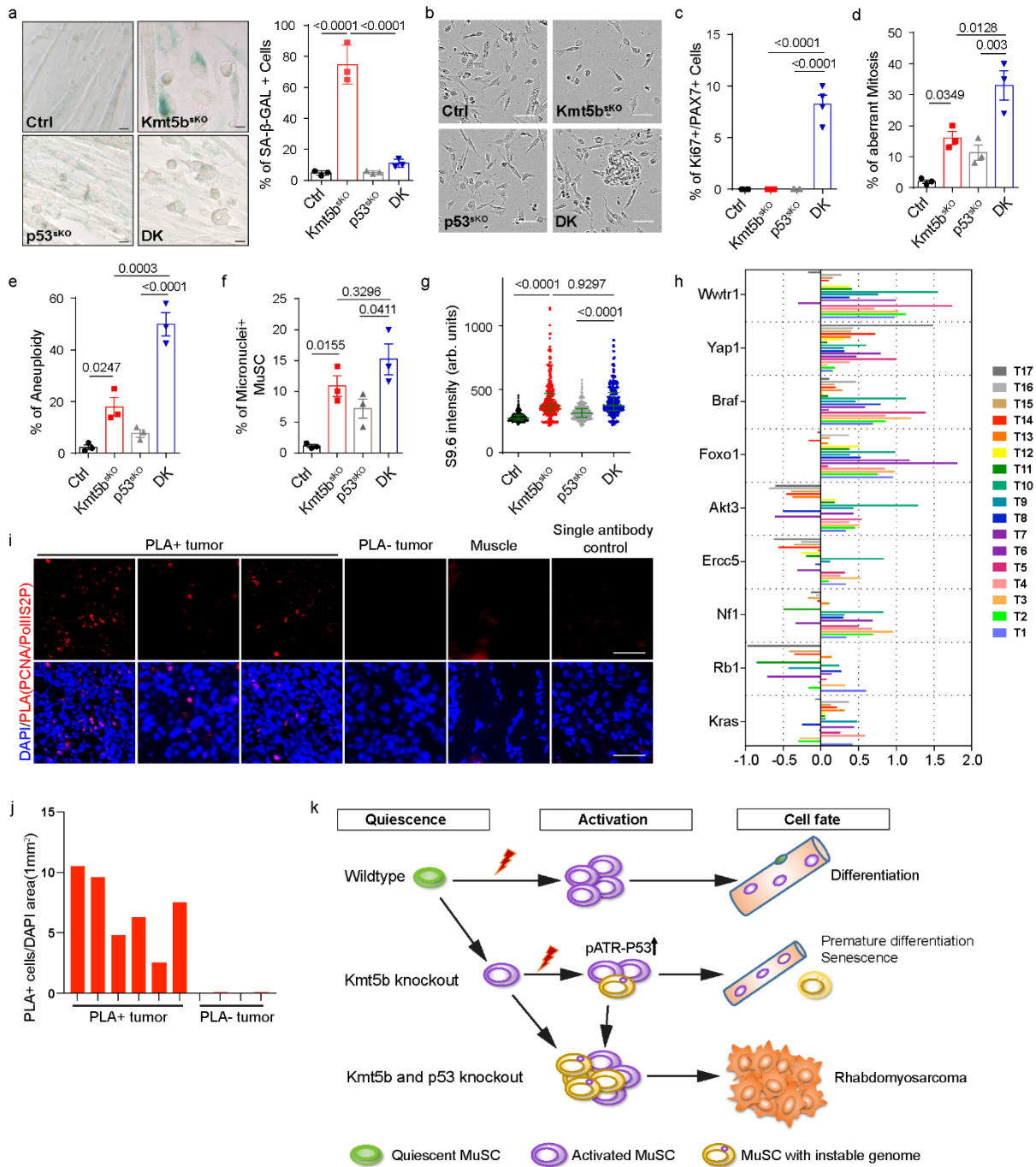
(a) Western blot analysis of Pol II S2P and  $\gamma$ H2AX in mock or DRB treated *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs. GAPDH and Histone H3 served as loading controls. Band intensities were quantified and normalized values are on the right (PolII S2P *Ctrl* n=3 mice, *Kmt5b<sup>SKO</sup>* n=4 mice;  $\gamma$ H2AX *Ctrl* n=5 mice; *Kmt5b<sup>SKO</sup>* n=6 mice).

(b, c, d) Percentages of (b) cells with aneuploidy (n=4 mice), (c) micronuclei (n=4 mice) and (d) aberrant mitosis (n=3 mice) in mock and DRB-treated *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs.

(e) Western blot analysis of transduced HA-tagged mouse RNaseh1 in *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs. GAPDH served as endogenous control (n=3 mice).

(f) Immunofluorescence analysis (MFI) of p-RPA (S33) in EdU<sup>+</sup>/EdU<sup>-</sup> *Ctrl* and *Kmt5b<sup>SKO</sup>* MuSCs with or without transduction of mouse *RNaseh1* (n=3 mice).

Data are shown as mean  $\pm$  SEM. Two-way ANOVA. Source data are provided in the Source Data file.

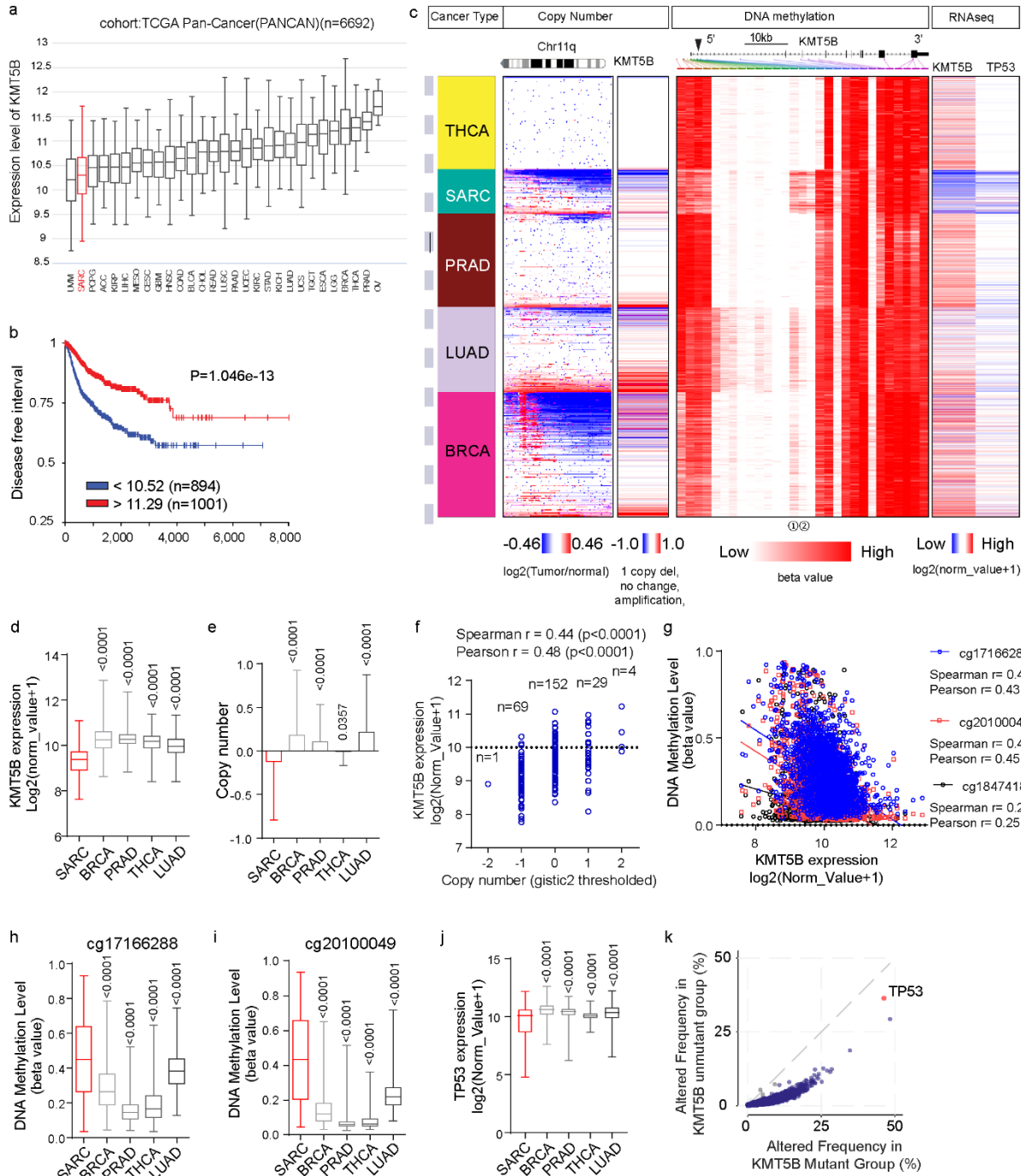


**Supplementary Figure 7 Inactivation of *p53* in *Kmt5b<sup>SKO</sup>* MuSCs prevents cellular senescence and enhances genomic instability.**

(a) SA-β-GAL staining of cultured *Ctrl*, *Kmt5b<sup>SKO</sup>*, *p53<sup>SKO</sup>*, and *DK* MuSCs (n=3 mice). Quantification of SA-β-GAL positive cells is on the right. Scale bar: 10 μm.

(b) Representative bright field images of *Ctrl*, *Kmt5b<sup>SKO</sup>*, *p53<sup>SKO</sup>*, and *DK* MuSCs after 7 days in culture (n=3 mice). Scale bar: 50 μm.

- (c) Immunofluorescence-based quantification of Ki67<sup>+</sup>/PAX7<sup>+</sup> MuSCs on TA muscle cryosections from *Ctrl*, *Kmt5b<sup>sKO</sup>*, *p53<sup>sKO</sup>*, and *DK* littermates after tamoxifen injection (n=4).
- (d) Quantification of aberrant mitoses (multipolar metaphases and bridging anaphases) in cultured *Ctrl*, *Kmt5b<sup>sKO</sup>*, *p53<sup>sKO</sup>* and *DK* MuSCs (n=3 mice).
- (e) Quantification of aneuploid MuSCs from *Ctrl*, *Kmt5b<sup>sKO</sup>*, *p53<sup>sKO</sup>* and *DK* mice detected by DAPI-stained metaphase spreads (n=3 mice).
- (f) Quantification of *Ctrl*, *Kmt5b<sup>sKO</sup>*, *p53<sup>sKO</sup>*, and *DK* MuSCs with micronuclei, based on  $\alpha$ -tubulin/DAPI immunofluorescence staining (n=3 mice).
- (g) Quantification of S9.6 staining intensity (*arbitrary units (a.u.)*) in nuclei of cultured *Ctrl* (n=336 cells), *Kmt5b<sup>sKO</sup>* (n=344 cells), *p53<sup>sKO</sup>* (n=355 cells), and *DK* (n=346 cells) MuSCs. Data represent median  $\pm$  interquartile ranges. n= the cell number over 3 independent experiments.
- (h) Copy number variation assay (CNV-qPCR) of representative oncogenes and tumor suppressors with gain or loss of R-loops after *Kmt5b* inactivation (Fig.4f) in tumors from *DK* mice. *Tert* served as reference gene.
- (i) PLA assay (PCNA/PolII S2P) on sections of tumors from *DK* mice. (n= 3 independent experiments) Scale bar: 50  $\mu$ m.
- (j) Quantification of PLA<sup>+</sup> cells on tumor sections per DAPI area.
- (k) Simplified model illustrating the synergism of *Kmt5b* and *p53* in MuSCs for preventing formation of eRMS.
- Data are shown as mean  $\pm$  SEM. One-way ANOVA. Source data are provided in the Source Data file.



**Supplementary Figure 8: Human sarcomas are characterized by low *Kmt5b* expression due to copy number loss and DNA hyper-methylation.**

(a) Average expression levels of *KMT5B* in primary tumors of different TCGA Pan-Cancer types. (b) Kaplan-Meier analysis of the disease-free interval of patients from the TCGA PANCAN project with high (n=1001) or low (n=894) levels of *KMT5B* expression. logrank test:  $p=1.04e-13$ .

(c) Xena visual spreadsheet showing genetic alterations of chr11q, gene copy number (gistic2 threshold), DNA methylation, and RNA-seq of *KMT5B* and *TP53* in primary tumors of SARC, BRCA, PRAD, THCA and LUAD from the TCGA PANCAN project. Each row corresponds to a single sample. The first column shows the scale of samples. Rows are sorted by the CNV (copy

number variation) pattern of chr11q and sub-sorted in subsequent columns. Scale bar: 100 samples. ① cg17166288, ② cg20100049. cg numbers refer to the probe numbers and relevant genomic region.

(d) *KMT5B* expression levels based on RNA-seq in primary tumors of SARC (n=250), BRCA (n=1064), PRAD (n=488), THCA (n=495), and LUAD (n=497).

(e) Average copy number of the *KMT5B* gene locus in primary tumors of SARC (n=250), BRCA (n=1065), PRAD (n=488), THCA (n=495), and LUAD (n=497). Data are presented as mean  $\pm$  SD.

(f) Correlation of *KMT5B* copy numbers (gistic2 thresholded) with *KMT5B* gene expression in TCGA SARC samples. The Spearman and Pearson correlation coefficient  $r$  is indicated. P-values indicate the significance of the correlation determined by  $r$  values and sample size.

(g) Correlation of DNA methylation levels of three *KMT5B* promoter regions with gene expression levels in 5 types of cancer. The Spearman and Pearson correlation coefficient  $r$  is indicated. P-values are ( $< 0.0001$ ) and indicate the significance of the correlation determined by  $r$  values and sample size.

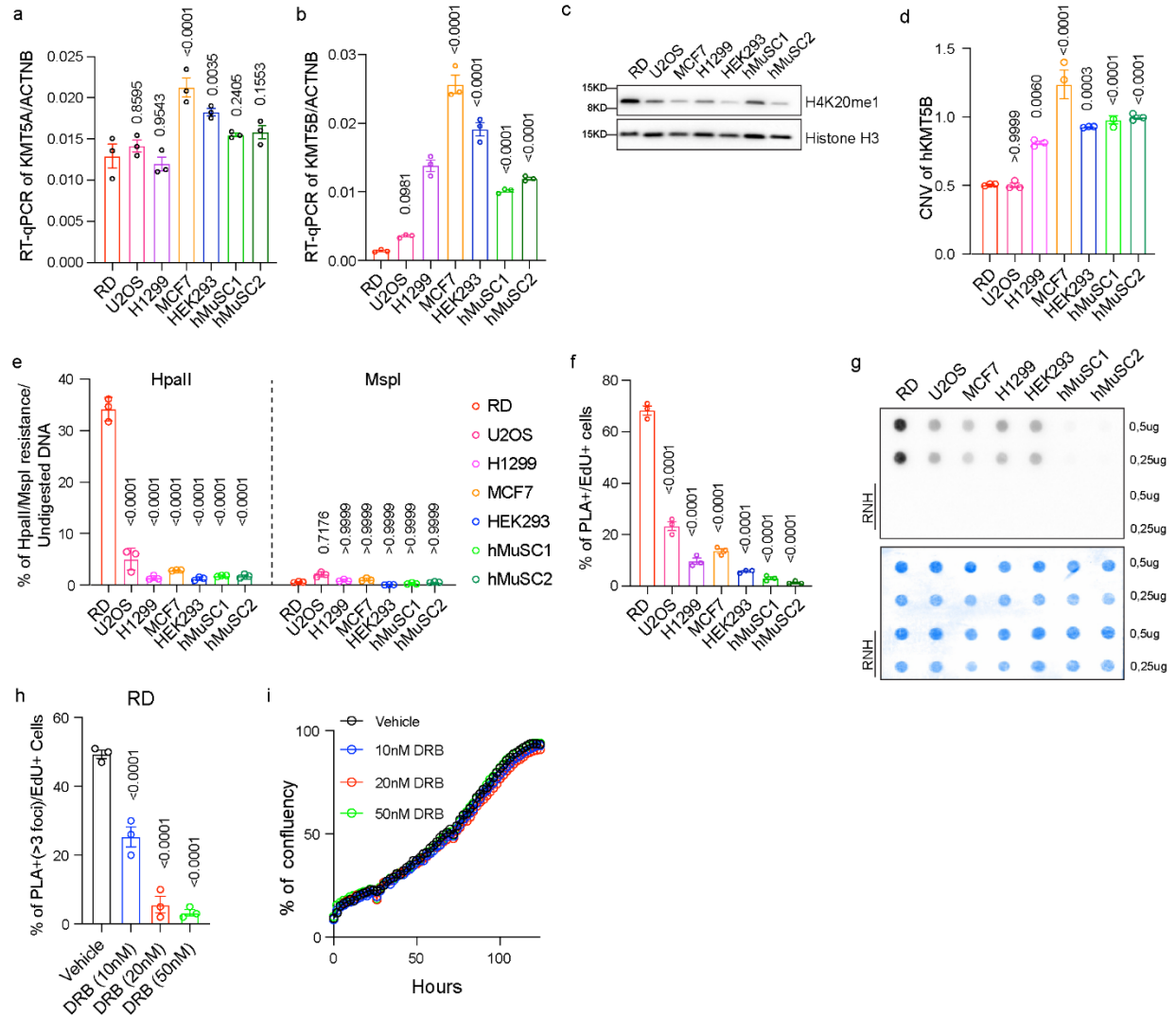
(h, i) DNA methylation levels of *KMT5B* promoter region, cg17166288 (h) and cg20100049 (i) in SARC (n=250), BRCA (n=764), PRAD (n=488), THCA (n=495), and LUAD (n=440).

(j) Expression of *TP53* based on RNA-seq data in SARC (n=250), BRCA (n=1064), PRAD (n=488), THCA (n=495), and LUAD (n=497).

(k) Mutations of *TP53* correlate with mutations of *KMT5B/KMT5B* in TCGA PANCAN.

Box plots in this figure depict the center line as median, box limits as upper and lower quartiles, and whiskers as minimum to maximum values. One-way ANOVA. Source data are provided in the Source Data file.

SARC, Sarcoma; THCA, Thyroid carcinoma; PRAD, Prostate adenocarcinoma; LUAD, Lung adenocarcinoma; BRCA, Breast invasive carcinoma



**Supplementary Figure 9: Analysis of *KMT5B* expression in various cancer cell lines.**

(a,b) RT-qPCR analysis of *KMT5A* (a) and *KMT5B* (b) in RD, U2OS, H1299, MCF7, HEK293, and two human myoblasts cell lines (n=3 independent experiments). *ACTN* served as the reference gene.

(c) Western blot analysis of H4K20me1 in RD, U2OS, H1299, MCF7, HEK293, and two human myoblasts cell lines. Histone H3 served as the internal control.

(d) Copy number variation (CNV) assay of *KMT5B* in RD, U2OS, H1299, MCF7, HEK293, and two human myoblasts cell lines (n=3 independent experiments). *RPR1* served as the reference gene.

(e) Methylation status of the *hKMT5B* gene promoter accessed by HpaII/MspI-qPCR analysis (n=3 independent experiments).

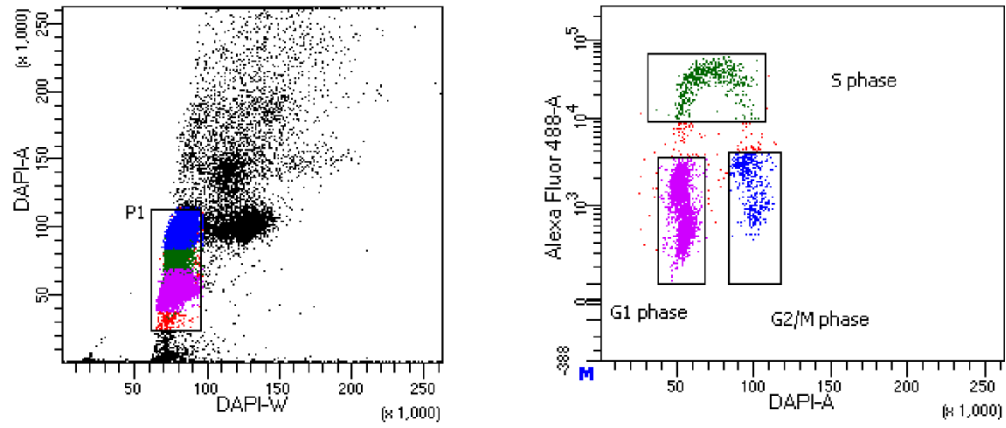
(f) Quantification of PLA (PCNA:PolIII S2P)<sup>+</sup>/EdU<sup>+</sup> cells in different cancer and normal human cell lines (n=3 independent experiments).

(g) Dot blot analysis of R-loops by S9.6 antibody using genomic DNA from different cancer and normal human cell lines.

(h) Quantification of PLA (PCNA:PolIII S2P)<sup>+</sup>/EdU<sup>+</sup> cells in vehicle and different concentration of DRB (10, 20, 50nM) treated RB cells (n=3 independent experiments).

(i) Proliferation assay of RB cells treated with vehicle and different concentration of DRB (10, 20, 50nM).

Data are shown as mean  $\pm$  SEM. One-way ANOVA. Source data are provided in the Source Data file.



**Supplementary Figure 10: Detailed gating strategy for G0/G1, S, and G2/M phases using CD FACS Diva v8 software.**