Cell Reports, Volume *43*

Supplemental information

ZBTB7A is a modulator of KDM5-driven transcriptional networks in basal breast cancer

Benedetto DiCiaccio, Marco Seehawer, Zheqi Li, Andriana Patmanidis, Triet Bui, Pierre Foidart, Jun Nishida, Clive S. D'Santos, Evangelia K. Papachristou, Malvina Papanastasiou, Andrew H. Reiter, Xintao Qiu, Rong Li, Yijia Jiang, Xiao-Yun Huang, Anton Simeonov, Stephen C. Kales, Ganesha Rai, Madhu Lal-Nag, Ajit Jadhav, Myles Brown, Jason S. Carroll, Henry W. Long, and Kornelia Polyak

SUPPLEMENTARY FIGURES

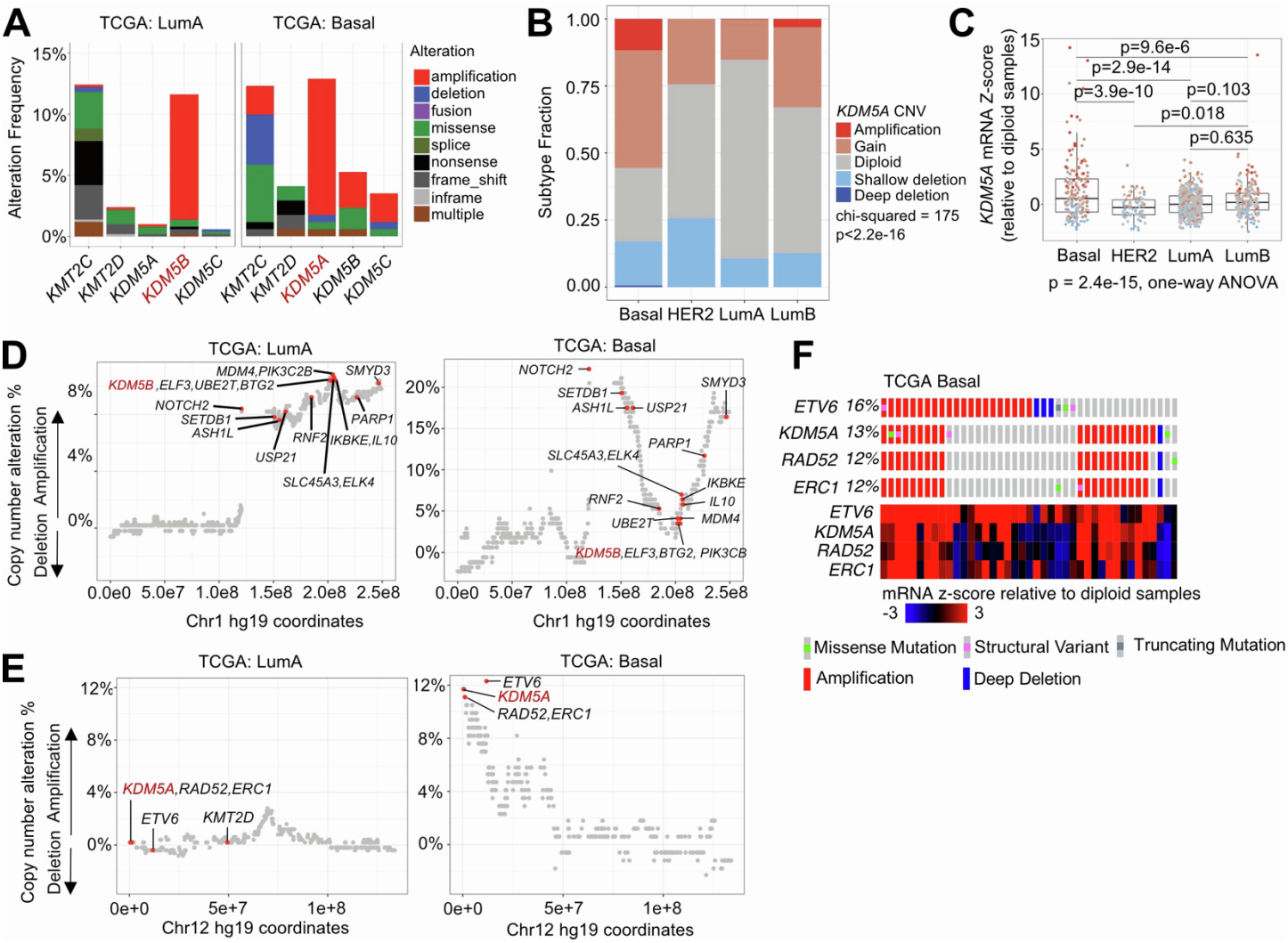

Figure S1. *KDM5A* is commonly amplified in basal breast cancer.
(A) Stacked bar chart depicting the frequencies of genetic alterations in genes encoding H3K4 methyltransferases (KMT2 family) and H3K4 demethylases (KDM5 family) in the LumA and basal PAM50 breast cancer subtypes in the TCGA cohort.
(B) Stacked bar chart showing fraction of tumors with *KDM5A* copy number variation (CNV) within each PAM50 breast cancer subtype. Chi- squared = 175, p<2.2e-16.
(C) Box plot showing distribution of *KDM5A* mRNA Z-scores across the PAM50 breast cancer subtypes. One-way ANOVA p=2.38e-15. Pairwise t-tests with Bonferroni corrections are shown.
(D-E) Amplification/deletion frequencies of individual genes mapped to their positions across chromosomes 1 or 12 in LumA and Basal breast cancer. Frequently amplified cancer-associated genes, based on OncoKB, are labelled.
(F) Oncoprint of *ETV6*, *KDM5A*, *RAD52*, and *ERC1* across basal breast cancer samples.

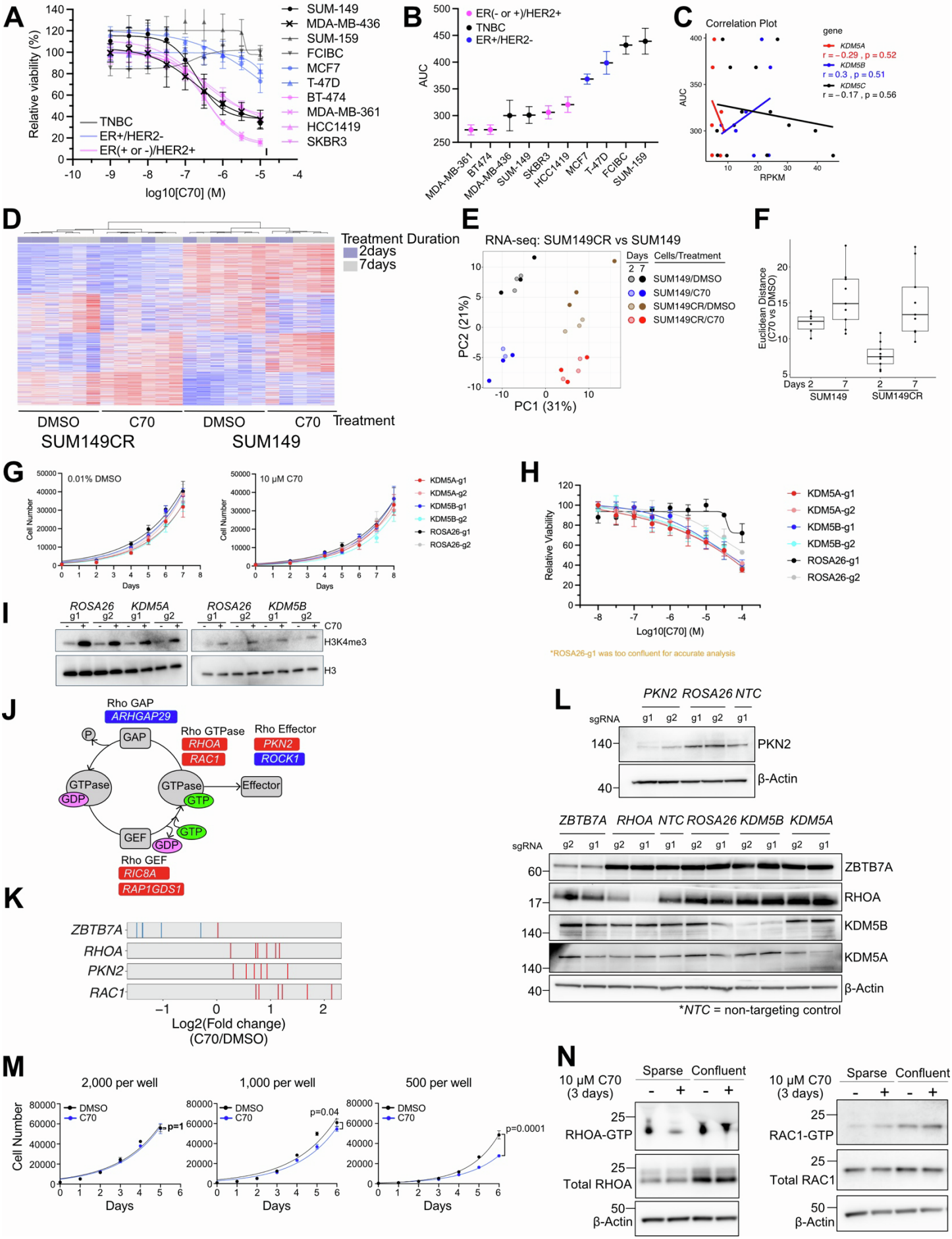
A
Relative viability (%)
log10[C70] (M)
SUM-149
MDA-MB-436
SUM-159
FCIBC
MCF7
T-47D
BT-474
MDA-MB-361
HCC1419
SKBR3
TNBC
ER+/HER2-
ER(+ or -)/HER2+
B
AUC
ER(- or +)/HER2+
TNBC
ER+/HER2-
C
Correlation Plot
gene
KDM5A
r = −0.29 , p = 0.52
KDM5B
r = 0.3 , p = 0.51
KDM5C
r = −0.17 , p = 0.56
RPKM
D
Treatment Duration
2days
7days
DMSO
C70
DMSO
C70
Treatment
SUM149CR
SUM149
E
RNA-seq: SUM149CR vs SUM149
PC2 (21%)
PC1 (31%)
Days
Cells/Treatment
SUM149/DMSO
SUM149/C70
SUM149CR/DMSO
SUM149CR/C70
F
Euclidean Distance (C70 vs DMSO)
Days
SUM149
SUM149CR
G
0.01% DMSO
10 μM C70
Cell Number
Days
KDM5A-g1
KDM5A-g2
KDM5B-g1
KDM5B-g2
ROSA26-g1
ROSA26-g2
H
Relative Viability
Log10[C70] (M)
*ROSA26-g1 was too confluent for accurate analysis
I
ROSA26
KDM5A
KDM5B
g1
g2
C70
H3K4me3
H3
J
Rho GAP
ARHGAP29
GAP
Rho GTPase
RHOA
RAC1
Rho Effector
PKN2
ROCK1
GTPase
GDP
GTP
Effector
GEF
Rho GEF
RIC8A
RAP1GDS1
K
ZBTB7A
RHOA
PKN2
RAC1
Log2(Fold change)
(C70/DMSO)
L
PKN2
ROSA26
NTC
sgRNA
PKN2
β-Actin
ZBTB7A
RHOA
NTC
ROSA26
KDM5B
KDM5A
*NTC = non-targeting control
M
2,000 per well
1,000 per well
500 per well
Cell Number
DMSO
C70
p=1
p=0.04
p=0.0001
Days
N
10 μM C70
(3 days)
Sparse
Confluent
RHOA-GTP
Total RHOA
β-Actin
RAC1-GTP
Total RAC1

Figure S2. CRISPR viability screen to identify modulators of KDM5 inhibitor sensitivity. Related to Figure 2.

(A) Dose-response curves to the KDM5 inhibitor C70 across a panel of breast cancer cell lines of the indicated subtype. Mean -/+ standard deviation is shown (n = 6).

(B) Area under the curve calculated from panel A renormalized to the lowest C70 concentration. Error bars are 95% confidence intervals. Error was propagated from error bars in panel A when calculating the AUC of Fig. S3A)

(C) Plot depicting Pearson correlation between C70-sensitivity and expression level of KDM5A/B/C across cell lines. AUC = area under the curve for % viability DMSO vs C70. RPKM values were obtained from the cancer cell line encyclopedia.

(D) Heatmap of RNA-seq in SUM149 and SUM149CR cells -/+ 10 μM C70 for 2 or 7 days. Rows and columns are ordered based on hierarchical clustering. Values are row normalized Z-scores.

(E) PCA plot of RNA-seq data in SUM149 and SUM149CR -/+ 10 μM C70 for 2 or 7 days.

(F) Euclidean distance between PC1/PC2 coordinates of treated and untreated samples in panel D.

(G,H) Growth curves (G) and IC50 plots (H) illustrating the viable cell numbers of SUM149 ROSA26, *KDM5A* or *KDM5B* knockout cell lines growing in DMSO or 10 μM C70.

(I) Immunoblot analysis for H3K4me3 in SUM149 ROSA26, *KDM5A* or *KDM5B* knockout cell lines after growth for three days in DMSO or 10 μM C70.

(J) Diagram of the Rho-GTPase cycle with significant hits from the SUM149 CRISPR screen labelled. Red = enriched in C70, blue = enriched in DMSO.

(K) Log2(fold change) of individual sgRNAs from the CRISPR screen in SUM149 -/+ 10 μM C70. Positive log2(fold change) indicates enrichment in C70.

(L) Immunoblot validating deletion of the indicated genes with individual gRNAs targeting top hits from the CRISPR screen in SUM149.

(M) Growth curves of SUM149 cells seeded at varying densities and treated with -/+ 10 μM C70. Cells were seeded at 2,000, 1,000, or 500 cells per well in a 96-well plate. Data are mean -/+ standard deviation (n=6). Statistical analysis for growth curves was performed using two-way repeated measures ANOVA followed by Bonferroni's multiple comparisons test; the significance at the last time point compared to is shown.

(N) Levels of Active RHOA-GTP and RAC1-GTP levels -/+ 10 μM C70 (3 days) in SUM149 cells. Active RHOA and RAC1 pulldowns were performed with GST-Rhotekin-RBD and GST-Pak1-PBD beads, respectively.

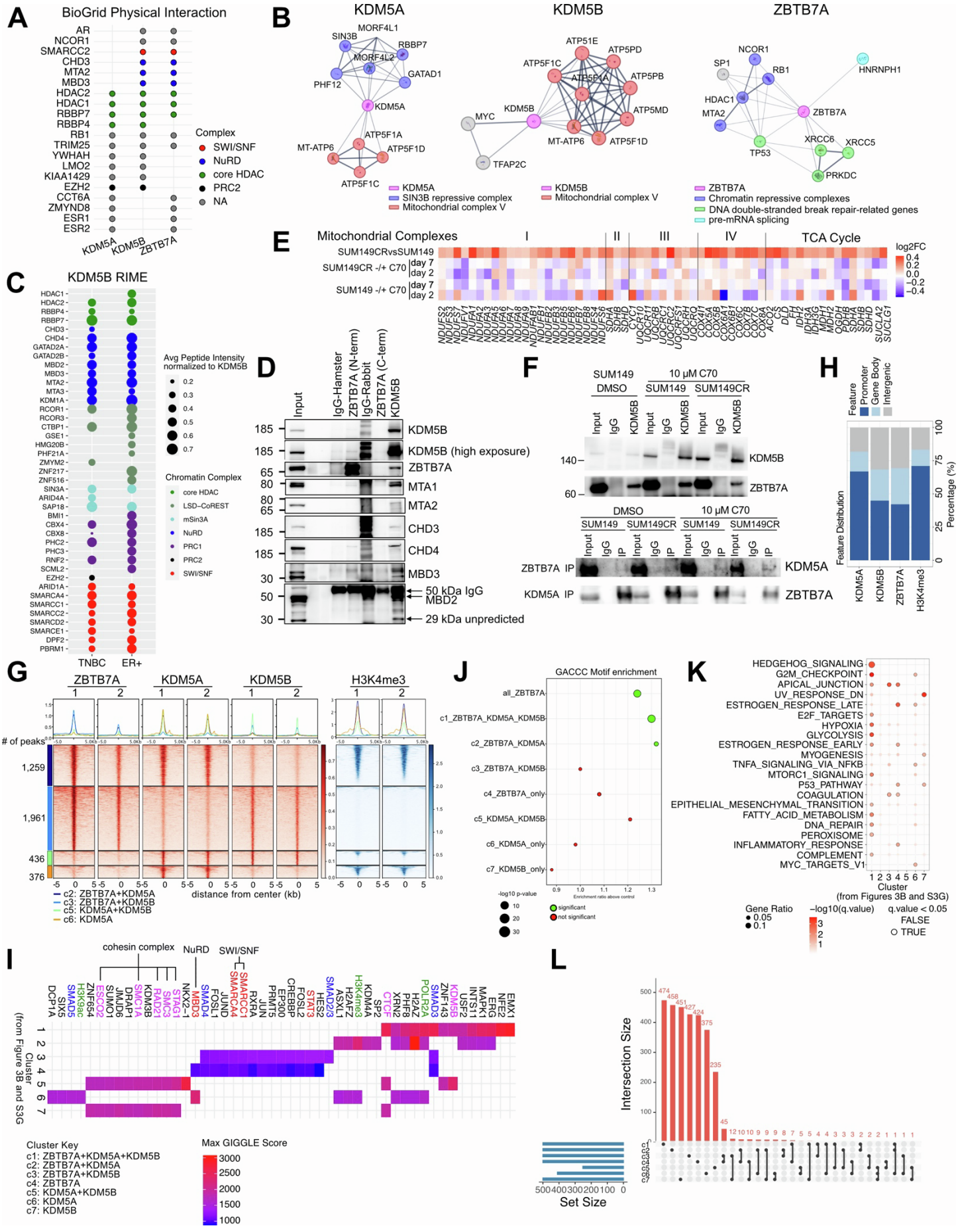
A
BioGrid Physical Interaction
AR
NCOR1
SMARCC2
CHD3
MTA2
MBD3
HDAC2
HDAC1
RBBP7
RBBP4
RB1
TRIM25
YWHAH
LMO2
KIAA1429
EZH2
CCT6A
ZMYND8
ESR1
ESR2
KDM5A
KDM5B
ZBTB7A
Complex
SWI/SNF
NuRD
core HDAC
PRC2
NA
B
KDM5A
MORF4L1
SIN3B
RBBP7
MORF4L2
GATAD1
PHF12
KDM5A
ATP5F1A
MT-ATP6
ATP5F1D
ATP5F1C
KDM5A
SIN3B repressive complex
Mitochondrial complex V
KDM5B
ATP51E
ATP5PD
ATP5F1C
ATP5PB
ATP5F1A
ATP5MD
KDM5B
MYC
MT-ATP6
ATP5F1D
TFAP2C
KDM5B
Mitochondrial complex V
ZBTB7A
NCOR1
SP1
RB1
HNRNPH1
HDAC1
MTA2
ZBTB7A
XRCC6
XRCC5
TP53
PRKDC
ZBTB7A
Chromatin repressive complexes
DNA double-stranded break repair-related genes
pre-mRNA splicing
C
KDM5B RIME
HDAC1
HDAC2
RBBP4
RBBP7
CHD3
CHD4
GATAD2A
GATAD2B
MBD2
MBD3
MTA2
MTA3
KDM1A
RCOR1
RCOR3
CTBP1
GSE1
HMG20B
PHF21A
ZMYM2
ZNF217
ZNF516
SIN3A
ARID4A
SAP18
BMI1
CBX4
CBX8
PHC2
PHC3
RNF2
SCML2
EZH2
ARID1A
SMARCA4
SMARCC1
SMARCC2
SMARCD2
SMARCE1
DPF2
PBRM1
TNBC
ER+
Avg Peptide Intensity normalized to KDM5B
0.2
0.3
0.4
0.5
0.6
0.7
Chromatin Complex
core HDAC
LSD–CoREST
mSin3A
NuRD
PRC1
PRC2
SWI/SNF
D
Input
IgG-Hamster
ZBTB7A (N-term)
IgG-Rabbit
ZBTB7A (C-term)
KDM5B
185
KDM5B
185
KDM5B (high exposure)
65
ZBTB7A
80
MTA1
80
65
MTA2
185
CHD3
185
CHD4
30
MBD3
50
50 kDa IgG
MBD2
30
29 kDa unpredicted
E
Mitochondrial Complexes
I
II
III
IV
TCA Cycle
SUM149CRvsSUM149
SUM149CR -/+ C70
day 7
day 2
SUM149 -/+ C70
day 7
day 2
log2FC
0.4
0.2
0.0
−0.2
−0.4
NDUFS2
NDUFS3
NDUFV1
NDUFA1
NDUFA2
NDUFA3
NDUFA5
NDUFA6
NDUFA7
NDUFA8
NDUFA9
NDUFAB1
NDUFB2
NDUFB3
NDUFB5
NDUFB6
NDUFB7
NDUFB8
NDUFB9
NDUFS4
NDUFS6
SDHA
SDHB
SDHC
UQCR10
UQCR11
UQCRB
UQCRC1
UQCRC2
UQCRFS1
UQCRH
UQCRQ
COX4I1
COX5A
COX6A1
COX6B1
COX6C
COX7B
COX7C
COX8A
ACO2
CS
DLD
IDH1
IDH3A
IDH3G
MDH1
MDH2
OGDH
SDHA
SDHB
SDHD
SUCLA2
SUCLG1
F
SUM149
10 μM C70
DMSO
SUM149
SUM149CR
Input
IgG
KDM5B
140
KDM5B
60
ZBTB7A
DMSO
10 μM C70
SUM149
SUM149CR
SUM149
SUM149CR
Input
IgG
IP
ZBTB7A IP
KDM5A
KDM5A IP
ZBTB7A
H
Feature
Promoter
Gene Body
Intergenic
Feature Distribution
Percentage (%)
100
75
50
25
0
KDM5A
KDM5B
ZBTB7A
H3K4me3
G
ZBTB7A
KDM5A
KDM5B
H3K4me3
1
2
# of peaks
1,259
1,961
436
376
-5
0
5
distance from center (kb)
c2: ZBTB7A+KDM5A
c3: ZBTB7A+KDM5B
c5: KDM5A+KDM5B
c6: KDM5A
J
GACCC Motif enrichment
all_ZBTB7A
c1_ZBTB7A_KDM5A_KDM5B
c2_ZBTB7A_KDM5A
c3_ZBTB7A_KDM5B
c4_ZBTB7A_only
c5_KDM5A_KDM5B
c6_KDM5A_only
c7_KDM5B_only
0.9
1.0
1.1
1.2
1.3
Enrichment ratio above control
-log10 p-value
10
20
30
significant
not significant
K
HEDGEHOG_SIGNALING
G2M_CHECKPOINT
APICAL_JUNCTION
UV_RESPONSE_DN
ESTROGEN_RESPONSE_LATE
E2F_TARGETS
HYPOXIA
GLYCOLYSIS
ESTROGEN_RESPONSE_EARLY
MYOGENESIS
TNFA_SIGNALING_VIA_NFKB
MTORC1_SIGNALING
P53_PATHWAY
COAGULATION
EPITHELIAL_MESENCHYMAL_TRANSITION
FATTY_ACID_METABOLISM
DNA_REPAIR
PEROXISOME
INFLAMMATORY_RESPONSE
COMPLEMENT
MYC_TARGETS_V1
1 2 3 4 5 6 7
Cluster
(from Figures 3B and S3G)
Gene Ratio
0.05
0.1
−log10(q.value)
3
2
1
q.value < 0.05
FALSE
TRUE
I
cohesin complex
NuRD
SWI/SNF
DCP1A
SIX5
SMAD5
H3K9ac
ZNF654
ESCO2
SUMO1
JMJD6
DRAP1
SMC1A
KDM3B
RAD21
SMC3
STAG1
NKX2–1
MBD3
SMAD4
FOSL1
JUND
SMARCA4
SMARCC1
RXRA
JUN
PRM15
EP300
CREBBP
FOSL2
STAT3
HES2
SMAD2/3
ASXL1
H2AFZ
H3K4me3
KDM4A
SP2
CTCF
XRN2
PHF8
H2AZ
POLR2A
SMAD3
ZNF143
KDM5B
USF2
INTS11
MAPK1
ERG
NFE2
EMX1
Cluster
(from Figure 3B and S3G)
1 2 3 4 5 6 7
Cluster Key
c1: ZBTB7A+KDM5A+KDM5B
c2: ZBTB7A+KDM5A
c3: ZBTB7A+KDM5B
c4: ZBTB7A
c5: KDM5A+KDM5B
c6: KDM5A
c7: KDM5B
Max GIGGLE Score
3000
2500
2000
1500
1000
L
Intersection Size
500
400
300
200
100
0
474
458
451
427
424
375
235
45
12
10
10
9
9
9
8
7
5
5
4
4
3
3
3
2
2
2
1
1
1
1
500
400
300
200
100
0
Set Size

Figure S3. ZBTB7A and KDM5A/B interactomes and chromatin patterns. Related to Figure 3.
(A) List of physical interactions reported in BioGrid that have experimental evidence for KDM5A, KDM5B, and ZBTB7A. Common interactors between ZBTB7A and KDM5A/KDM5B are shown.
(B) The edges indicate that the proteins are part of a physical complex, and the data is based on known interactions based on experiments or curated databased. The edge thickness indicates edge confidence. The interactions are non-exhaustive, in which the parameters were set to display a maximum of 10 interactions, and all interaction scores had to be above 0.4.
(C) KDM5B-associated proteins identified by qPLEX-RIME. The average peptide intensity normalized to KDM5B within TNBC (SUM149 and SUM159) and luminal ER+ (MCF7 and T47D) cell lines tested are shown.
(D) Immunoblot for NuRD subunits in ZBTB7A and KDM5B immunoprecipitants. Both N-terminal and C-terminal targeting antibodies were tested against ZBTB7A (labelled). Only the N-term targeting antibody successfully immunoprecipitated ZBTB7A.
(E) Heatmap depicting log2(FC) from RNA-seq in SUM149 and SUM149CR cells for the specified OXPHOS-related genes. The represented genes drive the leading edge of the GSEA analysis in Figure 1B.
(F) Immunoblot analysis of ZBTB7A and KDM5B or KDM5A in total cell lysates (input), control IgG and KDM5B, KDM5A or ZBTB7A immunoprecipitants in SUM149 and SUM149CR cells treated with DMSO or 10 μM C70 for 3 days.
(G) Heatmap of ChIP-seq for ZBTB7A, KDM5A, KDM5B, and H3K4me3. Peaks are clustered based on the intersection of peak calls among the three proteins.
(H) Feature distribution plot of KDM5A, KDM5B, ZBTB7A, and H3K4me3 ChIP-seq peaks in SUM149 cells.
(I) Factors with significant binding overlap with the identified ChIP-seq peaks in **Figures 3B, S3G**. Overlap is assessed via the Giggle score computed from CISTROME DB's toolkit. The maximum Giggle score is reported per comparison.
(J) Bubble plot showing enrichment ratios above controls of GACCC motif separated by cluster from Figure 3B and S3G.
(K) Overlap of the top 500 predicted target genes within each ChIP-seq cluster from **Figures 3B, S3G** with Hallmark pathways. Top genes were identified based on regulatory potential scores from BETA-minus.
(L) Overlap between the top 500 predicted target genes across cluster from **Figures 3B, S3G**.

Fig. S4

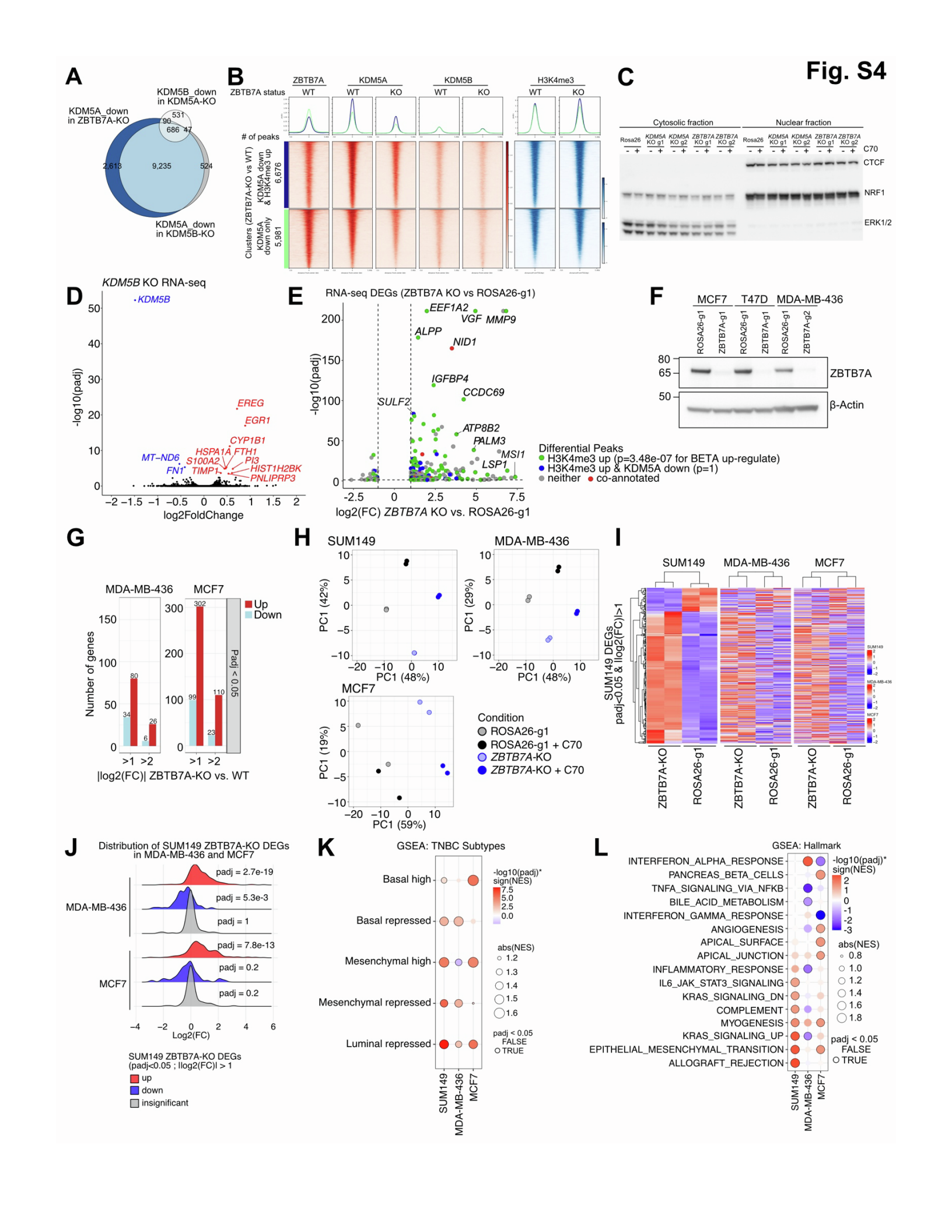
A
KDM5B_down in KDM5A-KO
KDM5A_down in ZBTB7A-KO
KDM5A_down in KDM5B-KO
531
99
686
47
2,618
9,235
524
B
ZBTB7A
KDM5A
KDM5B
H3K4me3
ZBTB7A status
WT
KO
# of peaks
Clusters (ZBTB7A-KO vs WT)
KDM5A down & H3K4me3 up
6,676
KDM5A down only
5,981
C
Cytosolic fraction
Nuclear fraction
Rosa26
KDM5A KO g1
KDM5A KO g2
ZBTB7A KO g1
ZBTB7A KO g2
C70
CTCF
NRF1
ERK1/2
D
KDM5B KO RNA-seq
KDM5B
EREG
EGR1
CYP1B1
HSPA1A
FTH1
MT-ND6
S100A2
PI3
FN1
TIMP1
HIST1H2BK
PNLIPRP3
-log10(padj)
log2FoldChange
E
RNA-seq DEGs (ZBTB7A KO vs ROSA26-g1)
EEF1A2
VGF
MMP9
ALPP
NID1
IGFBP4
CCDC69
SULF2
ATP8B2
PALM3
MSI1
LSP1
-log10(padj)
log2(FC) ZBTB7A KO vs. ROSA26-g1
Differential Peaks
H3K4me3 up (p=3.48e-07 for BETA up-regulate)
H3K4me3 up & KDM5A down (p=1)
neither
co-annotated
F
MCF7
T47D
MDA-MB-436
ROSA26-g1
ZBTB7A-g1
ZBTB7A-g2
ZBTB7A
β-Actin
G
MDA-MB-436
MCF7
Up
Down
Number of genes
Padj < 0.05
|log2(FC)| ZBTB7A-KO vs. WT
H
SUM149
MDA-MB-436
MCF7
PC1 (42%)
PC1 (29%)
PC1 (19%)
PC1 (48%)
PC1 (59%)
Condition
ROSA26-g1
ROSA26-g1 + C70
ZBTB7A-KO
ZBTB7A-KO + C70
I
SUM149
MDA-MB-436
MCF7
SUM149 DEGs padj<0.05 & |log2(FC)|>1
ZBTB7A-KO
ROSA26-g1
J
Distribution of SUM149 ZBTB7A-KO DEGs in MDA-MB-436 and MCF7
MDA-MB-436
MCF7
padj = 2.7e-19
padj = 5.3e-3
padj = 1
padj = 7.8e-13
padj = 0.2
padj = 0.2
Log2(FC)
SUM149 ZBTB7A-KO DEGs (padj<0.05 ; |log2(FC)| > 1
up
down
insignificant
K
GSEA: TNBC Subtypes
Basal high
Basal repressed
Mesenchymal high
Mesenchymal repressed
Luminal repressed
-log10(padj)* sign(NES)
abs(NES)
padj < 0.05
FALSE
TRUE
SUM149
MDA-MB-436
MCF7
L
GSEA: Hallmark
INTERFERON_ALPHA_RESPONSE
PANCREAS_BETA_CELLS
TNFA_SIGNALING_VIA_NFKB
BILE_ACID_METABOLISM
INTERFERON_GAMMA_RESPONSE
ANGIOGENESIS
APICAL_SURFACE
APICAL_JUNCTION
INFLAMMATORY_RESPONSE
IL6_JAK_STAT3_SIGNALING
KRAS_SIGNALING_DN
COMPLEMENT
MYOGENESIS
KRAS_SIGNALING_UP
EPITHELIAL_MESENCHYMAL_TRANSITION
ALLOGRAFT_REJECTION
-log10(padj)* sign(NES)
abs(NES)
padj < 0.05
FALSE
TRUE
SUM149
MDA-MB-436
MCF7

Figure S4. Gene expression changes induced by *ZBTB7A* KO and its impact on ZBTB7A and KDM5A/B peak sets. Related to Figure 4.

(A) Venn diagram comparing KDM5A peaks that decrease in the *ZBTB7A* KO or *KDM5B* KO along with KDM5B peaks that decrease in the *KDM5A* KO SUM149 cells.

(B) ChIP-seq signal intensity across all regions with a significant reduction in KDM5A signal (padj < 0.05) upon *ZBTB7A* deletion. Peaks are separated into those associated with a significant increase in H3K4me3 (padj < 0.05) upon *ZBTB7A* KO or not.

(C) Immunoblot analysis of NRF1 in nuclear and cytoplasmic fraction in the indicated cell lines and treatment. CTCF and ERK1/2 were used as controls for loading and subcellular localization.

(D) Volcano plot of differentially expressed genes (padj < 0.05; |log2(FC)| > 1) in SUM149 *ZBTB7A* KO vs ROSA26-g1. The nearest genes to differential peaks are annotated. P-values are based on binding and expression target analysis (BETA), indicating if the differential peak set is significantly associated with up- or down-regulated genes. Dashed gray lines indicate adjusted p value and fold change cut offs used for defining DGEs in RNA-seq.

(E) Volcano plot depicting genes differentially expressed between ZBTB7A KO vs. ROSA26-g1 SUM149 cells and are associated with H3K4me3 and/or KDM5A peaks showing the indicated changes.

(F) Immunoblot for ZBTB7A in parental and *ZBTB7A* KO MCF7 and MDA-MB-436 cell lines.

(G) Number of up and down-regulated DEGs upon *ZBTB7A* KO in MDA-MB-436 and MCF7 cells.

(H) PCA plots from RNA-seq of indicated cell lines with *ZBTB7A* KO or ROSA26-g1 -/+ 10 μM c70 for 7 days.

(I) Heatmap of SUM149 DEGs (padj < 0.05 and |log2(FC)| > 1) upon *ZBTB7A* KO. The DEGs are compared across SUM149, MDA-MB-436, and MCF7 cell lines. Rows are ordered based on hierarchical clustering of the SUM149 data. Values are row normalized Z-scores within each cell line.

(J) Log2(FC) distribution of SUM149 *ZBTB7A* KO DEGs (differentially expressed genes) in the MDA-MB-436 and MCF7 cell lines. P-values are based on GSEA using the SUM149 DEGs as the reference gene sets. A random selection of 102 insignificant genes from SUM149 were used as a control.

(K,L) Gene-set enrichment of differentially expressed genes comparing *ZBTB7A* -g1 and ROSA26-g1 within the SUM149, MDA-MB-436, and MCF7 cell lines. Enrichment of TNBC subtypes (K) and Hallmark gene sets (L).

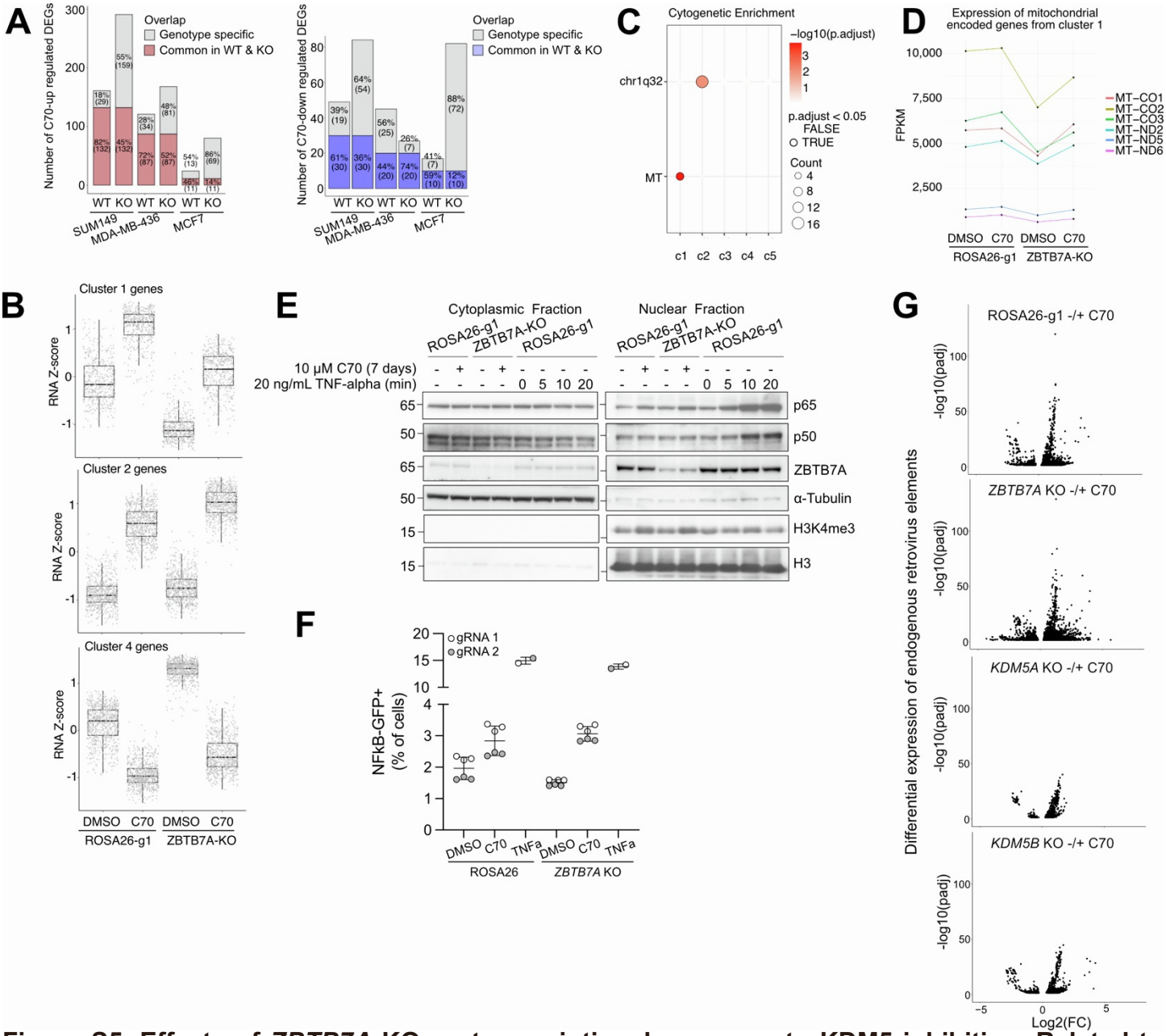

Figure S5. Effects of *ZBTB7A* KO on transcriptional response to KDM5 inhibition. Related to Figure 5.

(A) Number of up and down-regulated DEGs (padj < 0.05 and |log2(FC)| > 1) after 10 μM C70 treatment for 7 days within the indicated wild-type parental and *ZBTB7A* KO cell lines. DEGs common between both the WT and *ZBTB7A* KO are shaded in red (up-regulated) or blue (down-regulated).

(B) Box plot of RNA Z-scores within each cluster from **Figure 5D**.

(C) Enrichment of MSigDB's positional gene sets corresponding to human chromosome cytogenetic bands on each RNA-seq cluster from **Figure 5D**.

(D) FPKM values of mitochondrial encoded genes enriched in **Figure S4D**.

(E) Nuclear localization of NF-kB (p65 and p50). Immunoblot on cytoplasmic/nuclear fractions of SUM149 ROSA26-g1 and *ZBTB7A* KO cells -/+ 10 μM C70 for 7 days. 20 ng/mL TNF-alpha for the indicated time points was used as a positive control.

(F) NF-kB promoter activity assessed via a GFP-reporter construct driven by an NF-kB minimal promoter. Percent of GFP positive cells assessed via flow cytometry. Median + SD is shown.

(G) Differential endogenous retrovirus (ERV) expression after 7 days of 10 μM C70 treatment in the indicated SUM149 wild-type (ROSA26-g1) and knockout (*ZBTB7A* KO, *KDM5A* KO, *KDM5B* KO) cell lines. Results were calculated from RNA-seq data with the ERVmap pipeline. ERVs with -log10(padj) > 40 and |log2(FC)| > 1.5 are labelled.