Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/7/9/eabd7974/DC1

# Supplementary Materials for

## Functional antagonism of chromatin modulators regulates epithelial-mesenchymal transition

Michela Serresi*, Sonia Kertalli, Lifei Li, Matthias Jürgen Schmitt, Yuliia Dramaretska, Jikke Wierikx, Danielle Hulsman, Gaetano Gargiulo*

*Corresponding author. Email: michela.serresi@mdc-berlin.de (M.S.); gaetano.gargiulo@mdc-berlin.de (G.G.)

Published 24 February 2021, *Sci. Adv.* **7**, eabd7974 (2021)
DOI: 10.1126/sciadv.abd7974

**The PDF file includes:**



**Other Supplementary Material for this manuscript includes the following:**

(available at advances.sciencemag.org/cgi/content/full/7/9/eabd7974/DC1)

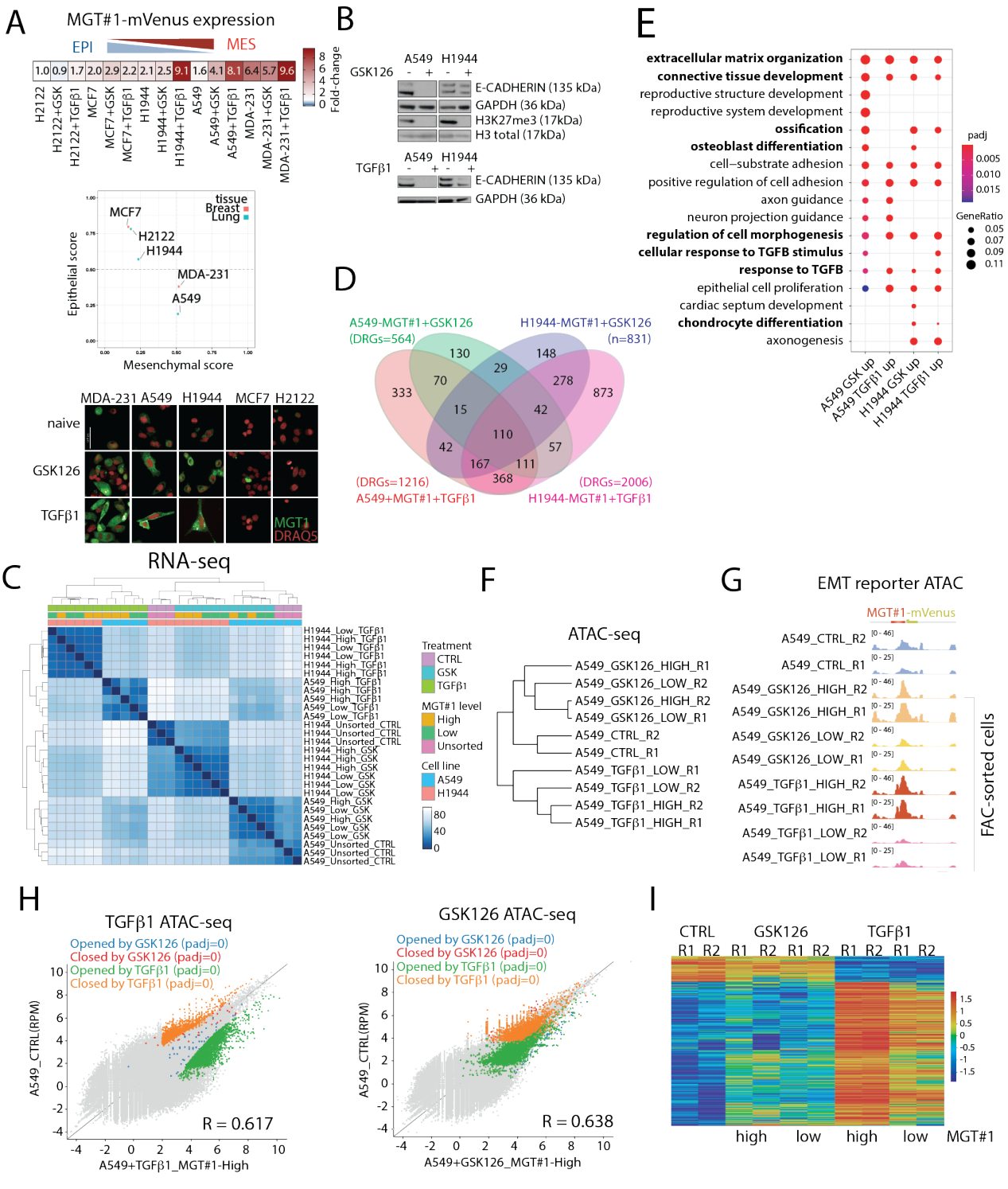

A
MGT#1-mVenus expression
EPI
MES
1.0 0.9 1.7 2.0 2.9 2.2 2.1 2.5 9.1 1.6 4.1 8.1 6.4 5.7 9.6
H2122
H2122+GSK
H2122+TGFβ1
MCF7
MCF7+GSK
MCF7+TGFβ1
H1944
H1944+GSK
H1944+TGFβ1
A549
A549+GSK
A549+TGFβ1
MDA-231
MDA-231+GSK
MDA-231+TGFβ1
8 6 4 2 0
Fold-change
tissue
Breast
Lung
MCF7
H2122
H1944
MDA-231
A549
Epithelial score
Mesenchymal score
MDA-231 A549 H1944 MCF7 H2122
naive
GSK126
TGFβ1
MGT1
DRAQ5
B
A549 H1944
GSK126
E-CADHERIN (135 kDa)
GAPDH (36 kDa)
H3K27me3 (17kDa)
H3 total (17kDa)
A549 H1944
TGFβ1
E-CADHERIN (135 kDa)
GAPDH (36 kDa)
D
A549-MGT#1+GSK126
(DRGs=564)
H1944-MGT#1+GSK126
(n=831)
130
148
29
70
278
333
873
15
42
110
42
57
167
111
368
(DRGs=1216)
A549+MGT#1+TGFβ1
(DRGs=2006)
H1944-MGT#1+TGFβ1
E
extracellular matrix organization
connective tissue development
reproductive structure development
reproductive system development
ossification
osteoblast differentiation
cell–substrate adhesion
positive regulation of cell adhesion
axon guidance
neuron projection guidance
regulation of cell morphogenesis
cellular response to TGFB stimulus
response to TGFB
epithelial cell proliferation
cardiac septum development
chondrocyte differentiation
axonogenesis
A549 GSK up
A549 TGFβ1 up
H1944 GSK up
H1944 TGFβ1 up
padj
0.005
0.010
0.015
GeneRatio
0.05
0.07
0.09
0.11
C
RNA-seq
H1944_Low_TGFβ1
H1944_High_TGFβ1
H1944_Low_TGFβ1
H1944_Low_TGFβ1
H1944_High_TGFβ1
H1944_High_TGFβ1
A549_High_TGFβ1
A549_High_TGFβ1
A549_High_TGFβ1
A549_Low_TGFβ1
A549_Low_TGFβ1
H1944_Unsorted_CTRL
H1944_Unsorted_CTRL
H1944_Unsorted_CTRL
H1944_High_GSK
H1944_High_GSK
H1944_High_GSK
H1944_Low_GSK
H1944_Low_GSK
H1944_Low_GSK
A549_High_GSK
A549_Low_GSK
A549_High_GSK
A549_Low_GSK
A549_Low_GSK
A549_Unsorted_CTRL
A549_Unsorted_CTRL
A549_Unsorted_CTRL
Treatment
CTRL
GSK
TGFβ1
MGT#1 level
High
Low
Unsorted
Cell line
A549
H1944
80
40
0
F
ATAC-seq
A549_GSK126_HIGH_R1
A549_GSK126_LOW_R2
A549_GSK126_HIGH_R2
A549_GSK126_LOW_R1
A549_CTRL_R2
A549_CTRL_R1
A549_TGFβ1_LOW_R1
A549_TGFβ1_LOW_R2
A549_TGFβ1_HIGH_R2
A549_TGFβ1_HIGH_R1
G
EMT reporter ATAC
MGT#1-mVenus
A549_CTRL_R2
A549_CTRL_R1
A549_GSK126_HIGH_R2
A549_GSK126_HIGH_R1
A549_GSK126_LOW_R2
A549_GSK126_LOW_R1
A549_TGFβ1_HIGH_R2
A549_TGFβ1_HIGH_R1
A549_TGFβ1_LOW_R2
A549_TGFβ1_LOW_R1
FAC-sorted cells
H
TGFβ1 ATAC-seq
Opened by GSK126 (padj=0)
Closed by GSK126 (padj=0)
Opened by TGFβ1 (padj=0)
Closed by TGFβ1 (padj=0)
A549_CTRL(RPM)
R = 0.617
A549+TGFβ1_MGT#1-High
GSK126 ATAC-seq
Opened by GSK126 (padj=0)
Closed by GSK126 (padj=0)
Opened by TGFβ1 (padj=0)
Closed by TGFβ1 (padj=0)
A549_CTRL(RPM)
R = 0.638
A549+GSK126_MGT#1-High
I
CTRL
GSK126
TGFβ1
R1 R2 R1 R2 R1 R2 R1 R2 R1 R2
1.5
1
0.5
0
-0.5
-1
-1.5
high low high low MGT#1

**Fig. S1: EZH2 inhibition by GSK126 in lung cancer cells promotes EMT reporter expression and the activation of an endogenous mesenchymal program with limited chromatin remodeling.**

**(A)** Above, heatmap of MGT#1 gene expression in cancer cells lines treated as indicated. Gene expression was normalized by *GAPDH* and number of integrations relative to the H2122 cell line (see Methods). Middle, EMT score-based cell line classification on gene expression as in Fig. 1C (see Methods). Below, live cell imaging of mVenus (EMT reporter) and DRAQ5 (nuclear dye) in the indicated cell lines and conditions. **(B)** Immunoblot of epithelial marker (e-cadherin), trimethyl-H3K27 and H3 markers of A549-MGT#1 and H1944-MGT#1 samples from RNA-seq upon treatment with GSK126 and TGF- β1 (see Methods). **(C)** Heatmap with hierarchical clustering showing correlation of the RNA-seq samples by Euclidian distance. The color intensity inversely correlates to the distance between pairs of samples, calculated from the normalized expression data. Column color bars denote the indicated cell lines, treatment conditions and MGT#1 expression levels. **(D)** Venn diagram depicting the overlap of the significant differentially regulated genes (DESeq2, padj < 0.05, log2FC +/- 1.5) in the indicated cell lines and conditions compared to the respective unsorted controls. **(E)** Dotplot visualization of the top upregulated gene sets, commonly induced by GSK126 and TGF- β1 in A549 and H1944 cell lines. Color-code and node size indicate significance and gene ratio, respectively. **(F)** Hierarchical clustering of ATAC-seq profiling of the top 250k accessible DHS for the indicated conditions. Reads were corrected for total count only in probes per million reads, for probe length, log transformed and transformed by size factor and percentile normalization. **(G)** IGV view of the ATAC-seq profile of the MGT#1 at its endogenous location for the indicated conditions. **(H)** Scatter plot of the ATAC-seq normalized reads count for 2M DHS. Colors indicate selected significant accessible sites as determined by DESeq2 (padj<0.05 for all conditions) and padj refers to gene set enrichment analysis for the indicated conditions. Reads were corrected for total count only in probes per million reads, for probe length, log transformed and transformed by size factor and percentile normalization. **(I)** Heatmap of the ATAC-seq normalized reads count for significant accessible sites as determined by DESeq2 (padj<0.05 for the TGF- β conditions).

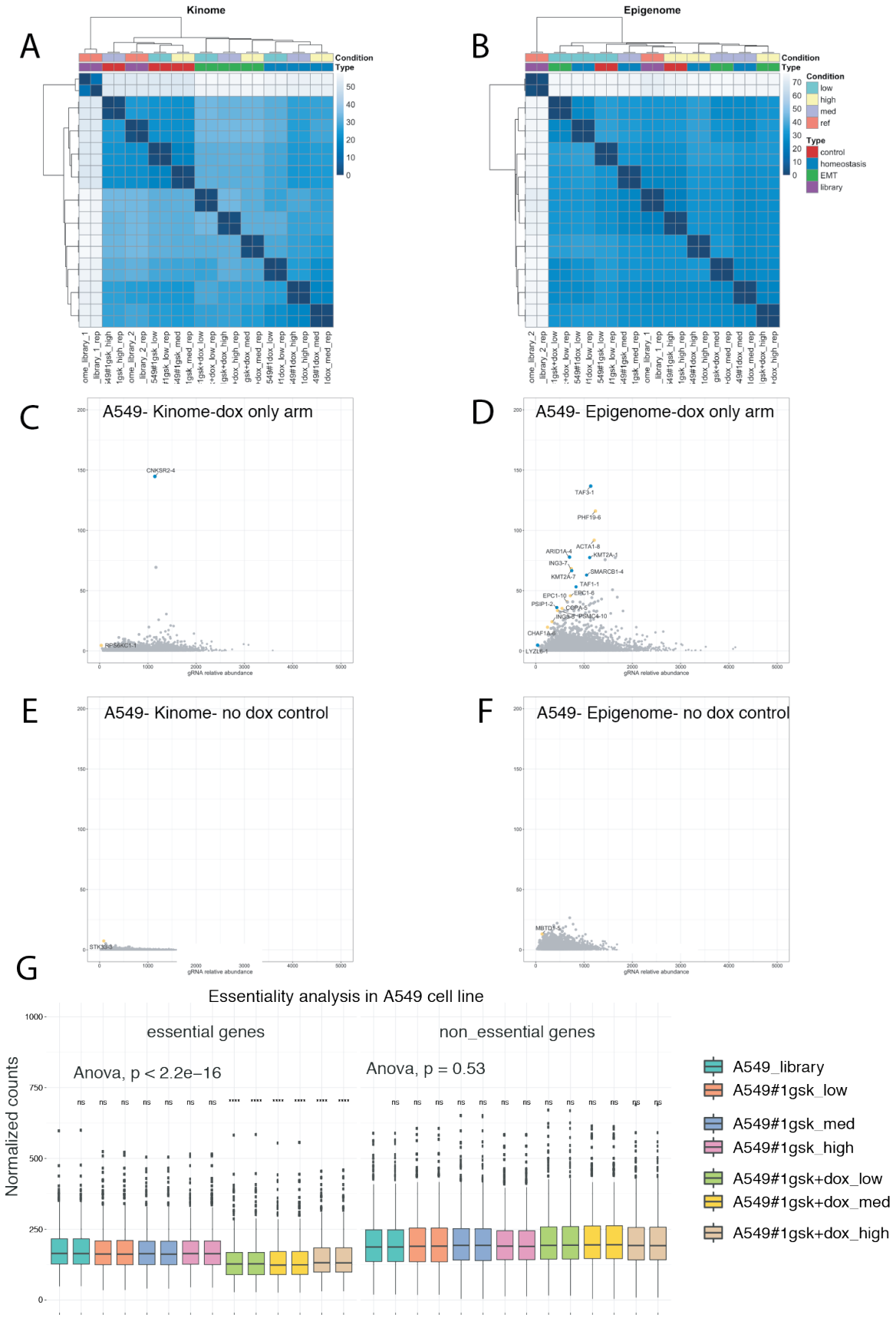



**Fig. S2: phenotypic CRISPR interference screens quality controls.**

**(A)** Heatmap of samples correlation in A549-MGT#1 kinome screen. Sample-to-sample Euclidean distance is calculated on ~6000 gRNAs. The color intensity inversely correlates to the distance between pairs of samples, as calculated from the normalized read counts of all gRNAs. Samples were clustered using complete linkage. Samples were furthered annotated by experimental conditions and MGT#1 expression level. **(B)** Heatmap of samples correlation in A549-MGT#1 epigenome screen. The settings for this analysis are consistent with (A). **(C)** Scatter plot of differential enrichment analysis for gRNAs from the dox-only arm of the A549-MGT#1 kinome screen to highlight the role for GSK126 to amplify the identification of hits. All the settings and scales are consistent with Fig.1F for comparison. **(D)** Scatter plot of differential enrichment analysis for gRNAs in the dox-only arm of the A549-MGT#1 epigenome screen. All the settings and scales are consistent with Fig.1F for comparison. Note that GSK126 numerically increases the hits. **(E)** Scatter plot of differential enrichment analysis for the kinome gRNAs in a representative control (i.e. without dox-induced KRAB-dCas9 activation). All the settings and scales are consistent with Fig.1F for comparison. **(F)** Scatter plot of differential enrichment analysis for the epigenome gRNAs in a representative control (i.e. without dox-induced KRAB-dCas9 activation). All the settings and scales are consistent with Fig.1F for comparison. **(G)** Boxplot of gRNA reads level for the indicated groups and comparisons. Significance is calculated by Wilcoxon rank-sum test.

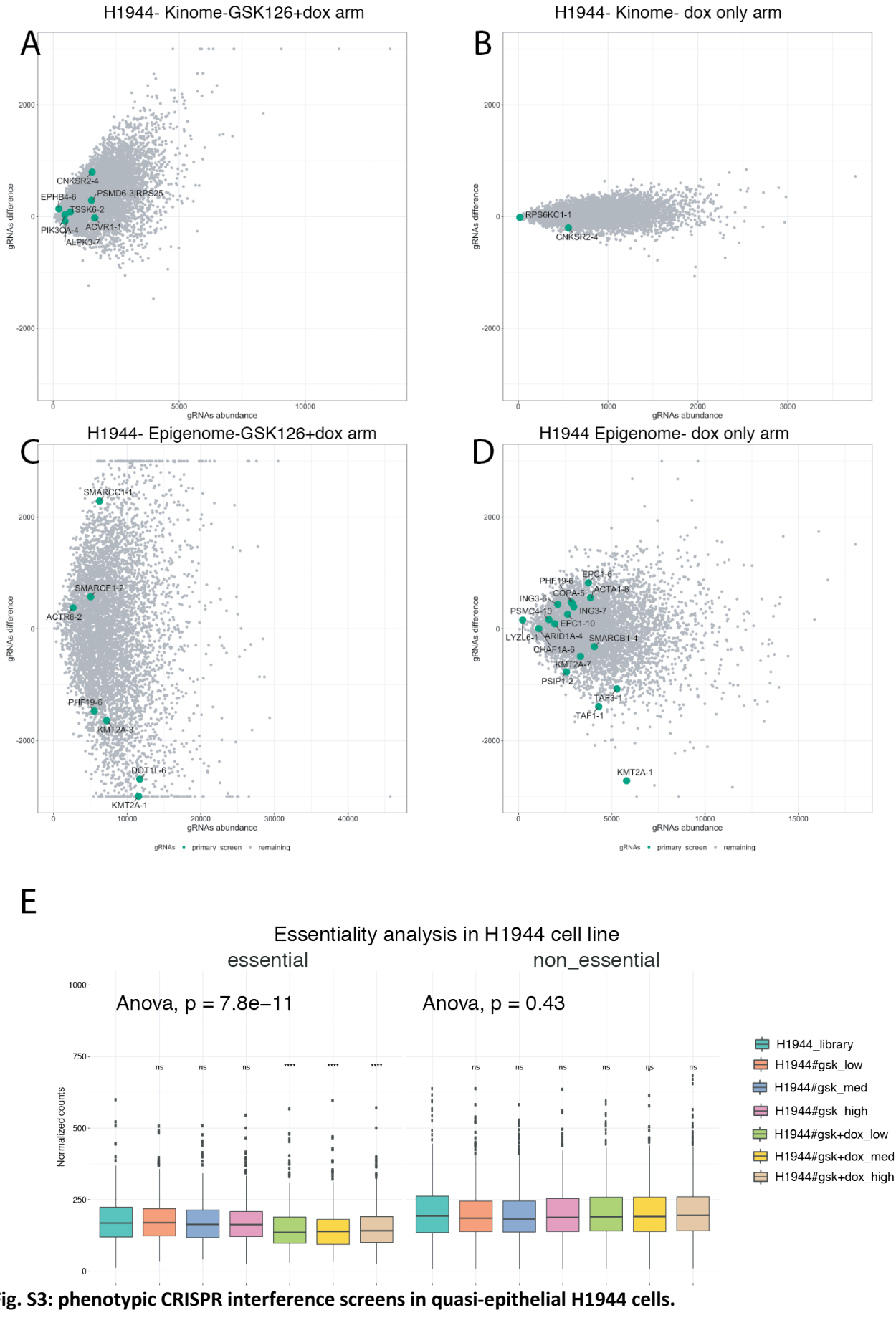



**Fig. S3: phenotypic CRISPR interference screens in quasi-epithelial H1944 cells.**

**(A)** MA plot of gRNA abundance (X-axis) and fold-change (Y-axis) in the GSK126- and dox-arm of the H1944-MGT#1 kinome screen. Selected hits from the A549-MGT#1 screen are colored to highlight that individual hits may be context dependent. **(B)** MA plot of gRNA abundance (X-axis) and difference in gRNA abundance (Y-axis) in the dox-only arm of the H1944-MGT#1 kinome screen to highlight that GSK126 increases the number of hits revealed by the screen. **(C)** MA plot of gRNA abundance (X-axis) and fold-change (Y-axis) in the GSK126- and dox-arm of the H1944-MGT#1 epigenome screen. Selected hits from the A549-MGT#1 screen are colored to highlight that individual hits may be context dependent and the scale is maintained in all analysis to highlight that chromatin factors are dominant over kinases as regulators of EMT homeostasis also in the H1944 screens. **(D)** MA plot of gRNA abundance (X-axis) and difference in gRNA abundance (Y-axis) in the dox-only arm of the H1944-MGT#1 epigenome screen to highlight that GSK126 increases the number of hits revealed by the screen. **(E)** Boxplot of gRNA reads level for the indicated groups and comparisons. Significance is calculated by Wilcoxon rank-sum test.

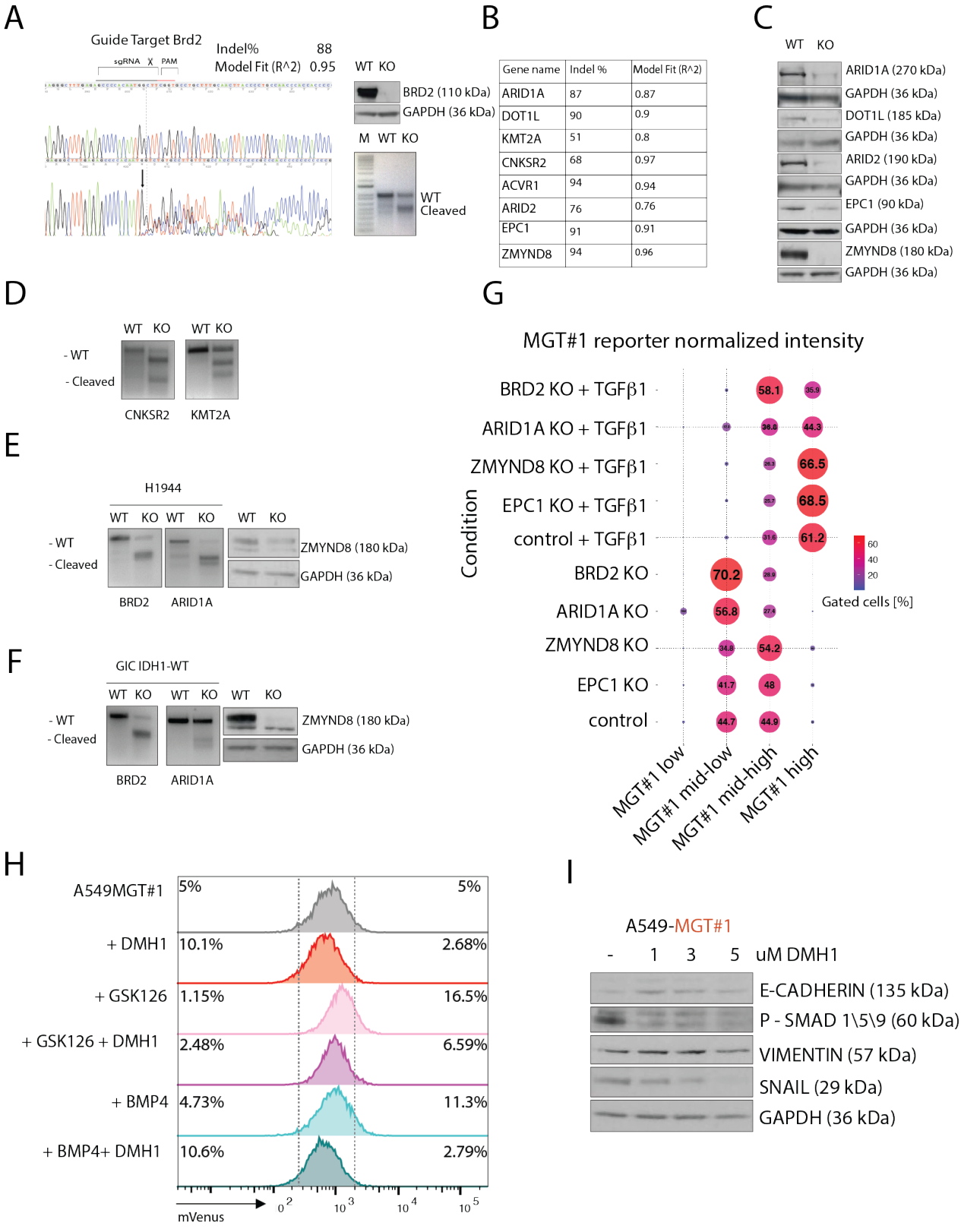



**Fig. S4: generation of knock-out lung cancer reporter cell lines by CRISPR/Cas9**

**(A)** For each selected positive regulator of MGT#1 activation in the CRISPR interference, the upper panel indicates the gRNA sequence and the lower and right panels reports on the validation by TIDE/ICE and

immunoblotting in A549 cells (polyclonal and clonal population), respectively. **(B)** Table summarizing the % of indel calculated by ICE analysis for each KO cell line generated. **(C)** Validation of KO cell lines by western blot. **(D)** Validation of CNKSR2 and KMT2A KO by T7E1 assay. **(E)** Validation of ARID1A and BRD2 KO in H1944 by T7E1 assay and western blot for ZMYND8. **(F)** Validation of ARID1A and BRD2 KO in GIC-IDH1-WT-hGIC by T7E1 assay and western blot for ZMYND8. **(G)** Bubble plot of FACS analysis validation of EMT suppressors/enhancers performed in prolonged exposure of TGF-β1 (10d). **(H)** FACS analysis of MGT#1-reporter in A549-MGT#1 pre-treated with GSK126 and BMP4 and after co-treatment with 1μM DMH1 (see Methods). **(I)** Immunoblot of epithelial (E-cadherin) and mesenchymal (Vimentin, Snail) markers in wild-type and DMH1-treated A549 -MGT#1 cells.

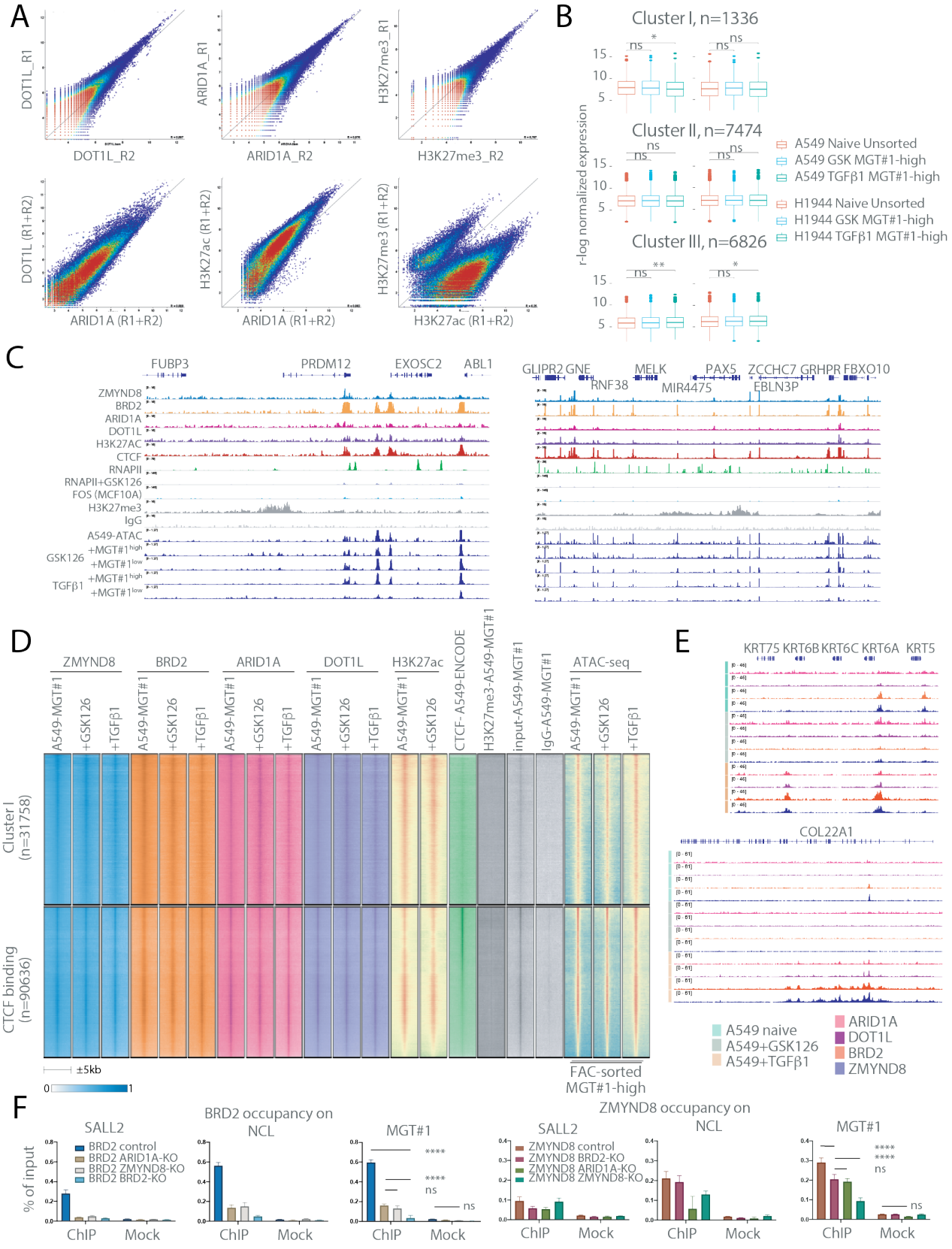



**Figure S5. Quality controls for ChIP-seq**

**(A)** For each pairwise comparison, the scatter plot analysis shows ChIP-seq reads quantitated and corrected for sequencing dept. Note the high concordance between biologically independent ChIP-seq. **(B)** Boxplot

of Cluster I-II-III normalized gene expression distribution. Color-coding depicts the indicated cell lines and treatment conditions. Significance is calculated by Wilcoxon rank-sum test. **(C)** IGV view of individual ChIP-seq binding at selected genomic loci in which all profiles can be assessed. **(D)** Heatmap of the binding for the indicated ChIP-seq and ATAC-seq experiments at Cluster I chromatin region. Below, the binding for indicated ChIP-seq and ATAC-seq experiments at A549 CTCF binding sites. **(E)** IGV view of individual ChIP-seq at selected genomic loci in the A549 cells with the indicated conditions. **(F)** Bar plot representation of the relative BRD2 and ZMYND8 occupancy on the indicated target loci as assessed by ChIP-qPCR. Significance was calculated by 2-way ANOVA and Sidak post-hoc test.

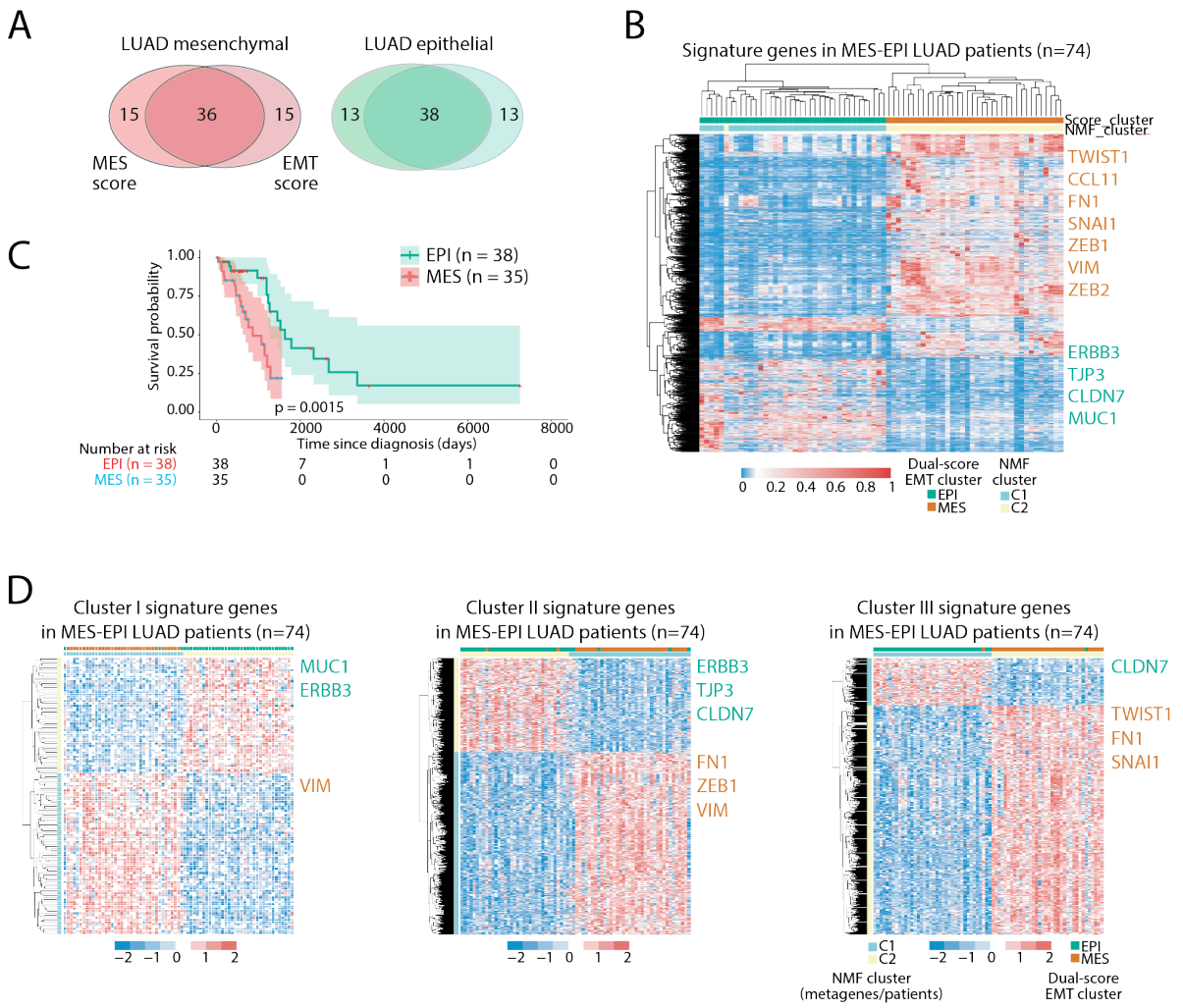



**Fig. S6. Cluster I, II and III genes separate TCGA LUAD patients according to their EMT status**

**(A)** Venn diagram depicting overlap between highly epithelial or mesenchymal LUAD biopsies identified by a dual score analysis. MES and EMT scoring classify LUAD patients based on gene expression of mesenchymal and epithelial-mesenchymal markers, respectively (see methods). **(B)** Heatmap showing clustering of epithelial and mesenchymal LUAD biopsies identified in A by non-negative matrix factorization based on all expressed genes (NMF; see methods). Selected EPI and MES metagenes are highlighted to the

right. **(C)** Overall survival of LUAD patients with EPI or MES gene expression profiling. **(D)** Heatmap showing clustering of epithelial and mesenchymal LUAD biopsies by non-negative matrix factorization based on Cluster I (Left), II (middle) and III (right) genes.

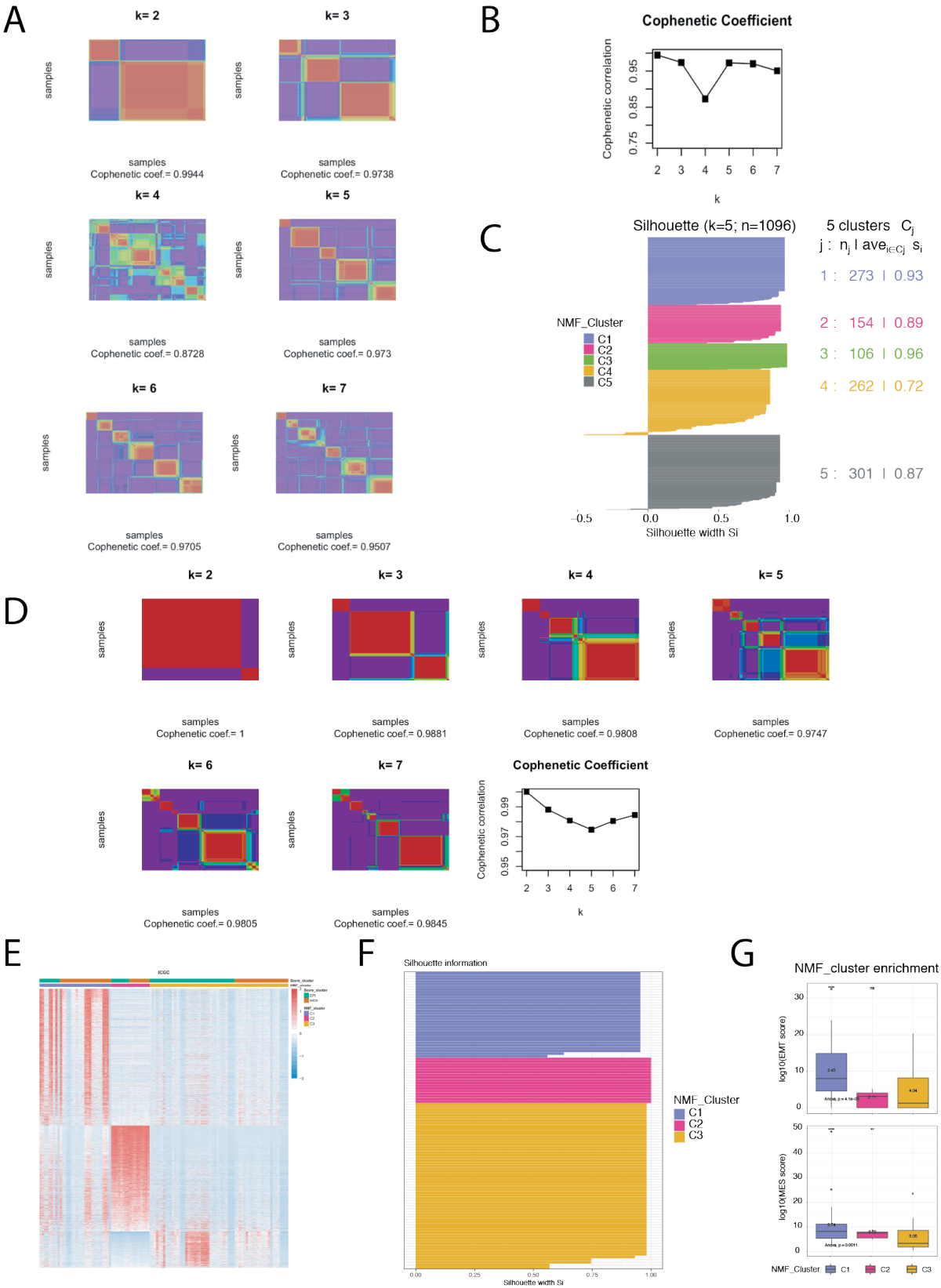


**Fig. S7. Epithelial-mesenchymal classification by NMF consensus clustering in TCGA and ICGC cohorts**

**(A)** Cluster I, II and III genes from Fig. 4a separate TCGA patients according to their EMT status. Consensus clustering matrix of 1,096 TCGA samples. **(B)** Cophenetic coefficient for k = 2 to k = 7. **(C)** Silhouette plot for identification of core TCGA samples. Note the generally high consensus for samples fitting the respective clusters. **(D)** Validation of NMF clustering analysis on TCGA patients by NMF consensus clustering matrix of ICGC samples. Based on cophenetic coefficient consensus for k = 2 to k = 7 and given the narrower set of cancer types in the validation cohort, we chose k=3 as minimal validation. **(E)** Heatmap showing 112 epithelial and mesenchymal ICGC cohorts biopsies defined by the dual scoring analysis are clustered into three clusters by NMF consensus based on the 3,798 signature genes defined by NMF clustering on TCGA cohorts (see methods). **(F)** Silhouette plot for identification of ICGC core samples. In the k=3 setting, note the generally high consensus for samples fitting the C1, C2 and C4 clusters of TCGA cohorts and that C3 in TCGA cohort is largely high- and low-grade gliomas, which are absent in the ICGC cohort. **(G)** Box plot showing the EMT (left) and MES (right) scores for each NMF cluster in the three NMF clusters for ICGC validation cohort.

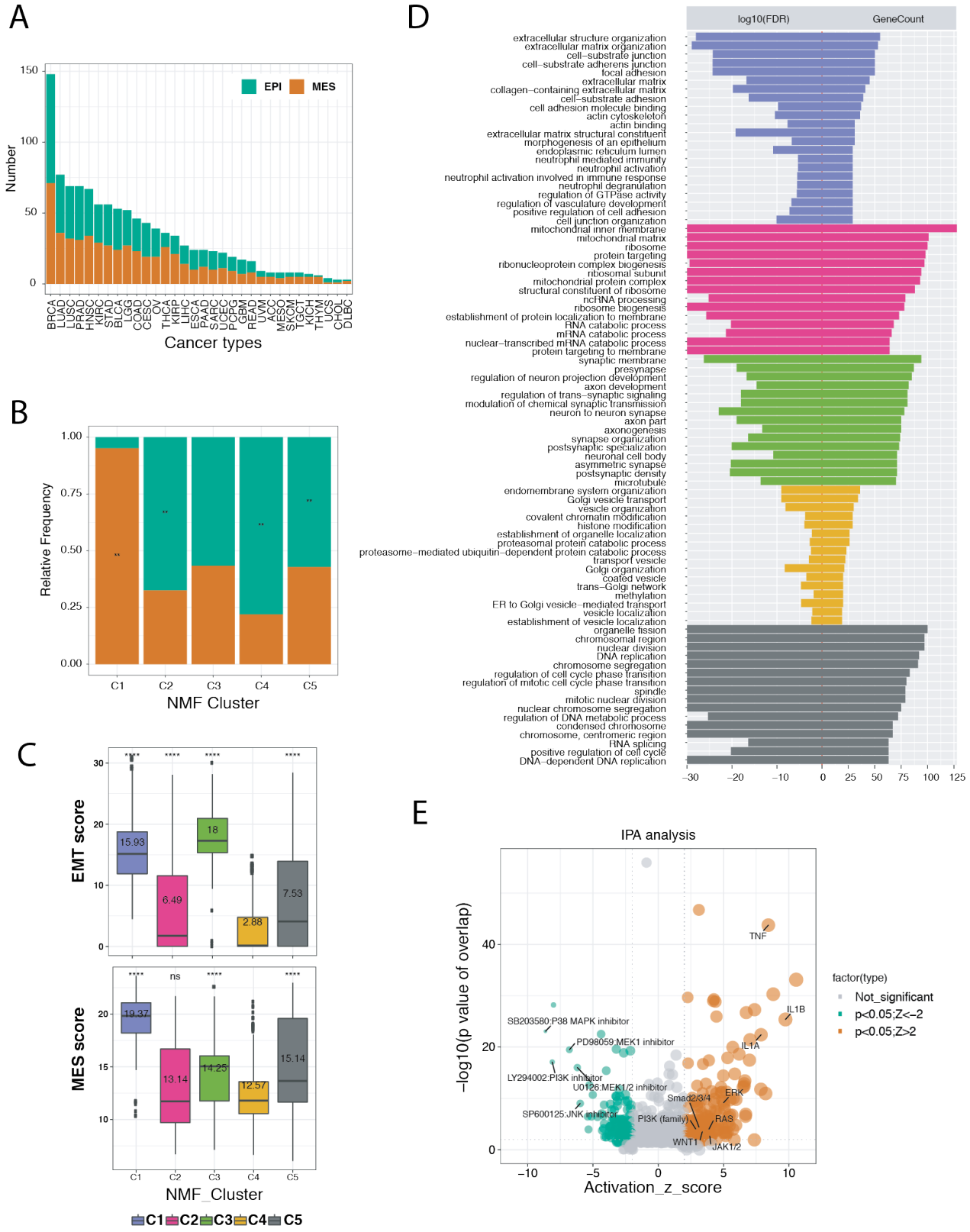



**Fig S8. Tissue of origin and identity of samples in the NMF clusters**

**(A)** Number of epithelial (EPI) or mesenchymal (MES) biopsies as defined by dual-scoring and divided by cancer type. **(B)** Enrichment of MES and EPI patients independent of cancer types in each NMF clusters (Fisher's exact test). **(C)** Box plot showing EMT and MES scores for each NMF cluster. Reference: C4. **(D)**

GO terms enriched in the five NMF clusters as Fig. 5B. **(E)** Upstream regulator analysis of chemical drugs and compounds by IPA on differentially-regulated genes between epithelial and mesenchymal LUAD+KIRC+STAD patients. Selected significant terms are highlighted.

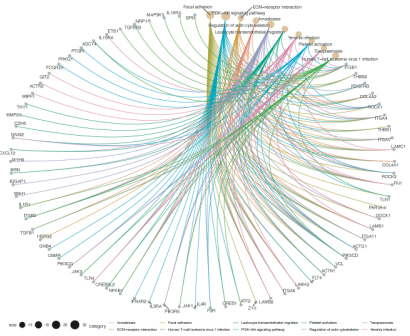
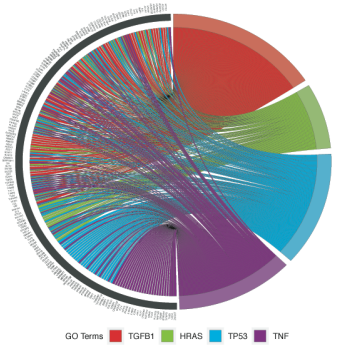
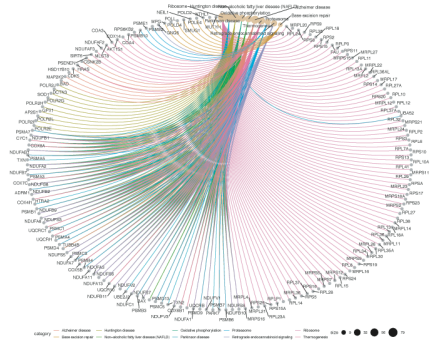
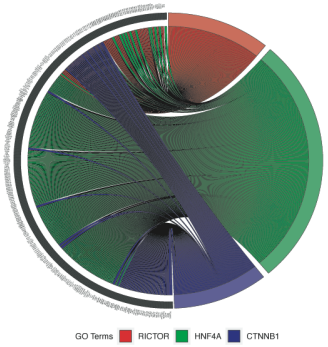
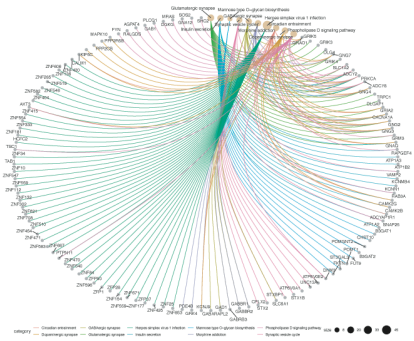
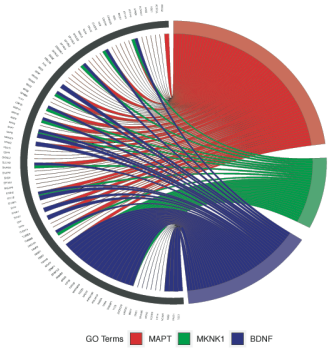
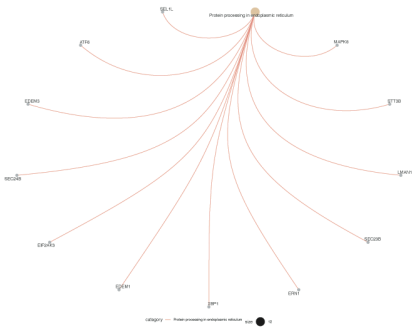
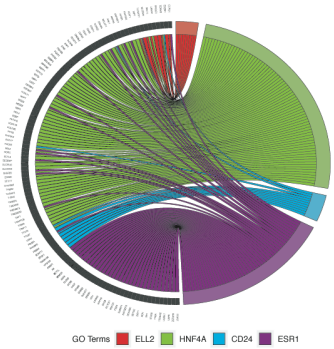
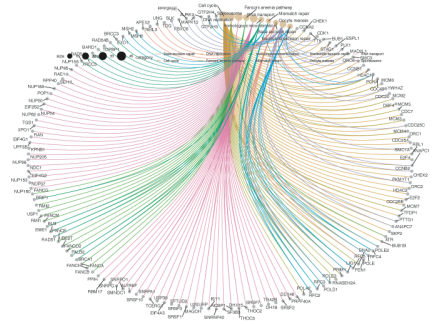
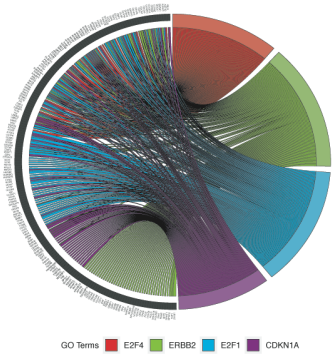
| NMF | KEGG | Upstream Regulators |
| --- | --- | --- |
| C1 |  |   |
| C2 |  |   |
| C3 |  |   |
| C4 |  |   |
| C5 |  |   |

**Fig. S9. Functional analysis of NMF clusters**

Left, gene-concept networks from the KEGG analysis; the yellow dots represent the pathways shown in Fig. 5B and the grey dots represent genes. Right, upstream regulators (legend) and their downstream targets as defined by Ingenuity Pathway Analysis on metagenes associated with each NMF cluster.

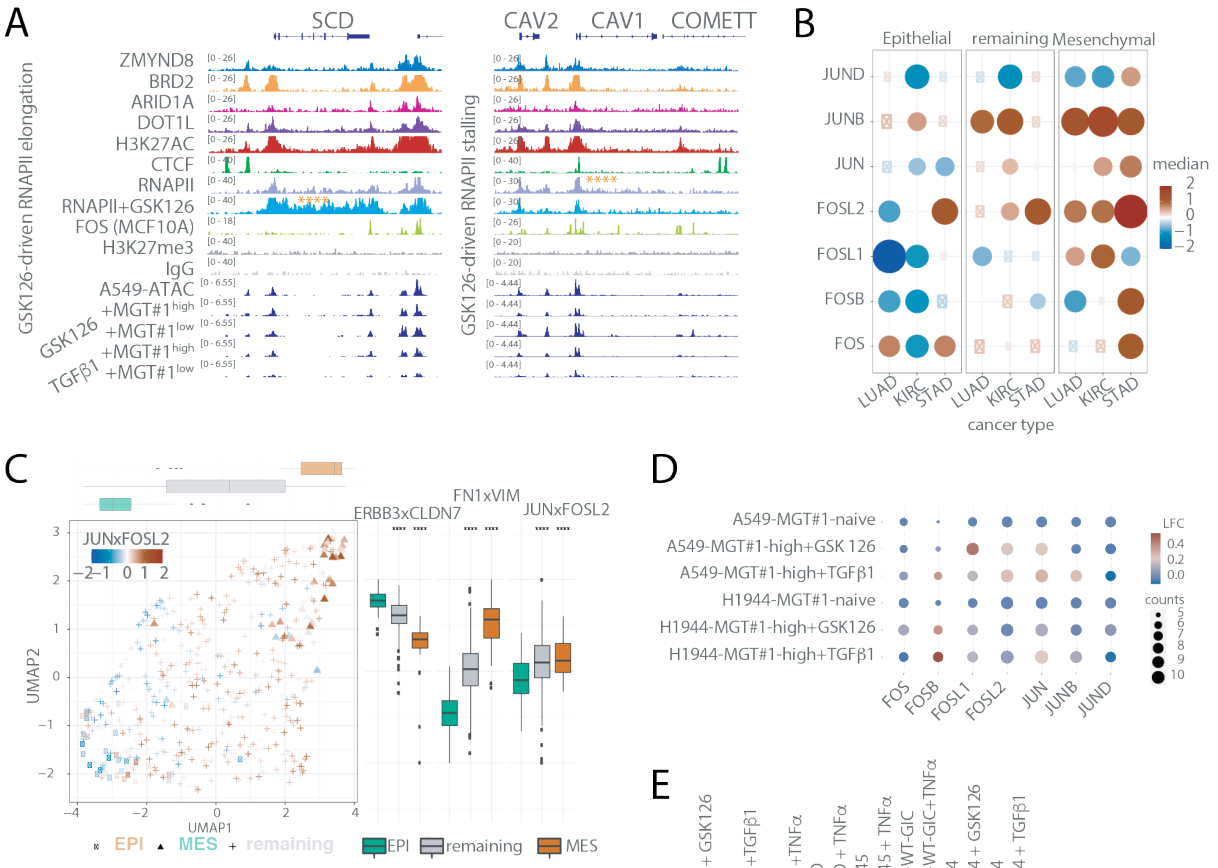



**Fig. S10. Extended analysis of AP-1 characterization in LUAD, KIRC and STAD**

**(A)** IGV view of individual ChIP-seq and ATAC-seq profiles at selected genomic loci, in A549 cells, in the indicated conditions. **(B)** Bubble plot showing expression data for the indicated genes and cancer types. Bubble color and size denote relative expression and fold-changes, respectively. **(C)** UMAP clustering of LUAD patients based on signature genes from NMF cluster C1. Color code denotes the fold-changes in the JUNxFOSL2 score. The box plot above represents the distribution of patients within the UMAP and is the same as in Fig. 5E. The box plot to the right represents the distribution of the indicated scores within the indicated subset of LUAD patients. **(D)** Bubble plot showing expression data for the indicated genes and conditions. Bubble color and size indicate fold-change compared to control and normalized expression per sample, respectively. **(E)** Immunoblotting for c-JUN and vinculin in the indicated in naïve and treated cell lines as indicated (see Methods). **(F)** IGV view of individual ChIP-seq binding at selected genomic loci in A549 cells with the indicated treatment conditions.

**SUPPLEMENTARY TABLES**

**Table S1. Source data for Fig. 1 and Supplementary Fig. 2-3.** For each cell line, condition and library, normalized readcounts for the kinome and epigenome CRISPR interference screens are indicated in columns and the sgRNAs in rows. X-axis=sum, Y-axis=difference, stastistics= -log10(padj).

**Table S2. Source data for Fig. 4A.** K-means clusters I, II and III based on ZMYND8 binding.

**Table S3. Source data for Fig. 5.** TCGA cancers and patients.

**Table S4. Source data for Fig. 5A.** NMF clustering outcome. The first tab includes the patients' TCGA identifier, the NMF cluster, cancer type and the classification of each patient by our dual scoring system.

**Table S5. Source data for Fig. 5C and Supplementary Fig. 9**

**Table S6. Source data for Fig. 6D.** Relative directional RNAPII enrichment score.