

Auxin-inducible degron 2 system deciphers functions of CTCF domains in transcriptional regulation

Judith Hyle^{1, *}, Mohamed Nadhir Djekidel^{2, *}, Justin Williams¹, Shaela Wright¹, Ying Shao³, Beisi Xu^{2, #}, Chunliang Li^{1, #}

¹Department of Tumor Cell Biology, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, USA.

²Center for Applied Bioinformatics, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, USA.

³Department of Computational Biology, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, USA.

* Contributed equally

Correspondence should be addressed to Dr. Chunliang Li (chunliang.li@stjude.org) or Dr. Beisi Xu (beisi.xu@stjude.org)

Fig. S1. Characterization of protein degradation of CTCF^{AID2} cells and impact on survival.

(a) Immunoblot analysis from CTCF^{AID2} cells following 0-24 hours 10 μ M 5-Ph-IAA treatment; CTCF^{AIDmClover3} was detected by a CTCF antibody and EGFP^{AID2} was detected by a miniAID antibody.

(b) Immunoblot analysis of CTCF^{AID2} cells treated with a 1 μ M 5-Ph-IAA titration for 6 hours. (

(c) Immunoblot analysis of CTCF^{AID2} cells co-treated with 1 μ M 5-Ph-IAA and proteasome inhibitor MG132 for 5 hours.

(d) Immunoblot analysis of SEM^{WT} and CTCF^{AID1} cells treated with 500 μ M IAA, 1 μ g/ml doxycycline, or a combination of both for 24 hours. The lower CTCF band represents untagged endogenous CTCF. The higher CTCF band represents the miniAID-tagged endogenous CTCF (CTCF^{AIDmClover3}).

(e) Growth assay of SEM^{WT} cells treated with DMSO, 500 μ M IAA, 1 μ g/ml doxycycline, or a combination of both over 4 days. Samples were set up in triplicate and cell counts for each replicate were collected daily. Bars represent cell numbers (x millions).

(f) Growth assay of CTCF^{AID1} cells treated with DMSO, 500 μ M IAA, 1 μ g/ml doxycycline, or a combination of both over 4 days. Samples were set up in triplicate and cell counts for each replicate were collected daily. Bars represent cell numbers (x millions).

(g) Growth assay of SEM^{WT}, SEM^{OsTIR1(F74G)} and CTCF^{AID2} cells treated with DMSO or 1 μ M 5-Ph-IAA over 4 days. Samples were set up in triplicate and cell counts for each replicate were collected daily. Bars represent cell numbers (x millions).

Fig. S2. Characteristics of CTCF lost and resistant peaks after AID2 depletion.

(a) CTCF binding enrichment heatmap centered at CTCF peaks showing the binding intensity of CTCF peaks before and after 5-Ph-IAA treatment. Two main groups are identified: i) 27,262 CTCF

peaks that significantly reduced in intensity but were not completely lost, and ii) 19,291 CTCF peaks that were completely lost after 5-Ph-IAA treatment. An additional 67 peaks identified in AID2-treated cells are not shown in the heatmap due to low confidence.

(b) Meta plot showing the mean CTCF peak intensity in untreated cells of the lost and significantly reduced peaks.

(c) Stacked Bar-plot showing the genomic distribution of CTCF peaks from panel a.

(d) Genomic distribution of CTCF peaks from c) with a more stringent cutoff ($FDR \leq 0.05$, $(|\log_2(FC)| \geq 1)$).

(e) Boxplot plot showing the genomic distance of lost and significantly reduced CTCF peaks to nearby domain boundaries identified by HiChIP.

Fig. S3. QC analysis of HiC and HiChIP.

(a) Stratum-adjusted correlation coefficient at 100-kb resolution of six SEM Hi-C replicates [12], two publicly available SEM Hi-C data sets (GSM4120559, GSM4120560), and CTCF HiChIP with and without 6 hours 10 μ M 5-Ph-IAA treatment (2 replicates each). The values plotting was median across all chromosomes calculated by HiCRep R package. The higher value indicates more reproducible events.

(b) Venn-diagram overlapping HiC (called by HiCCUPS from Juicer pipeline, see method) and CTCF HiChIP loops with and without 10 μ M 5-Ph-IAA treatment for 6 hours (called by FitHiChIP, see method).

(c) Stacked Bar-plot showing the loops based on anchors distribution for retained (368), lost (6,852) or new (355) loops. Each loop anchor was assigned to one out of three groups, with the

following annotation, as Promoter (TSS +/- 2kb), Enhancer (SEM H3K27ac peak summit from GSM1934089) or CTCF.

(d) Boxplot with jitter for log₂(fold change) of CTCF ChIP signal (+5-Ph-IAA versus -5-Ph-IAA) at CTCF peaks either overlap lost loops anchors or retained loops anchors. T-tests were used for the statistical test.

Fig. S4. QC analysis and differential gene analysis of SLAM-seq.

(a) Box plot of the percentage conversion rate of A>G (orange box, negative strand) and T>C (blue box, positive strand) quantified in each SLAM-seq experiment by SlamDunk using either AID¹ or AID² for CTCF depletion.

(b) Volcano plot representing the differential gene distribution for each time point in the AID¹ and AID² SLAM-seq time course. The horizontal dotted line represents a *p-value* of 0.05, and vertical lines represent the log₂FC of -1.0 and 1.0. TCCPM = T>C CPM reads.

Fig. S5. Characterization of CTCF-independent effect in CTCF^{AID1} system.

(a) MA-plot of SEM^{WT} cells following AID1 treatment (500 μM IAA; 1 μg/ml doxycycline) for 24 hours. The log₂ average mean expression is shown on the x-axis and Log₂(FC) is shown on the y-axis. The most significant differentially expressed genes ($|\log_2(\text{FC})| \geq 1$, $\text{FDR} \leq 0.05$) are shown as red dots.

(b) Differential expression of the *ASNS*, *DDIT3* and *TRIB3* genes by RNA-seq (FKPM) in response to 24 hours AID1 treatment in SEM^{WT} cells.

(c) GSEA plot of the top three enriched pathways following 24 hours of AID1 treatment in SEM^{WT} cells.

(d) Scatter plot showing the correlation between the 24h transcriptional changes of treated SEM^{WT} (WT) and AID1-treated CTCF^{AID1}. The x-axis shows the log₂(FC) of SEM^{WT} cells after IAA treatment for 24 hours. The y-axis represents the log₂(FC) of CTCF^{AID1} cells after IAA treatment for 24 hours.

(e) Scatter plot showing the low correlation between 24h transcriptional change of treated AID1 SEM^{WT} (WT) and AID2-treated (1 μM 5-Ph-IAA) CTCF^{AID2} cells. The x-axis shows the log₂(FC) of SEM^{WT} cells after IAA treatment for 24 hours. The y-axis represents the log₂(FC) of CTCF^{AID2} cells after 5-Ph-IAA treatment for 24 hours.

(f) Scatterplots showing the correlation between 24h transcriptional change of AID1-treated CTCF^{AID1} cells and AID2-treated CTCF^{AID2} cells. The y-axis represents the log₂(FC) of CTCF^{AID1} cells after IAA treatment for 24 hours. The y-axis represents the log₂(FC) of CTCF^{AID2} cells after 5-Ph-IAA treatment for 24 hours. d-f) Both Spearman and R² correlation values are shown.

Fig. S6. Characterization of CTCF binding signature of WT and mutants.

(a) Heatmap of normalized CTCF signal centered at CTCF peaks identified from 2 replicates of HA ChIP-seq of CTCF^{AID2/WT}, CTCF^{AID2/dZF1}, CTCF^{AID2/dZF10}, and CTCF^{AID2/dRBR} cells following 10 μM 5-Ph-IAA treatment for 24 hours and 18 hours of 1 μg/mL doxycycline. A combined count of 27,896 peaks was identified in all conditions.

(b) Genomic distribution of all CTCF peaks from CTCF^{AID2/WT}. A total of 27,923 peaks were called.

(c) Genomic distribution of CTCF binding peaks decreased (≥ 2 fold, FDR ≤ 0.05) comparing CTCF^{AID2/dZF10} to CTCF^{AID2/WT}.

(d) Genomic distribution of CTCF binding peaks increased (≥ 2 fold, FDR ≤ 0.05) comparing CTCF^{AID2/dZF10} to CTCF^{AID2/WT}.

(e) Genomic distribution of CTCF binding peaks decreased (≥ 2 fold, $\text{FDR} \leq 0.05$) comparing CTCF^{AID2/dZF1} to CTCF^{AID2/WT}.

(f) Genomic distribution of CTCF binding peaks increased (≥ 2 fold, $\text{FDR} \leq 0.05$) comparing CTCF^{AID2/dZF1} to CTCF^{AID2/WT}.

(g) Genomic distribution of CTCF binding peaks decreased (≥ 2 fold, $\text{FDR} \leq 0.05$) comparing CTCF^{AID2/dRBR} to CTCF^{AID2/WT}.

(h) Genomic distribution of CTCF binding peaks increased (≥ 2 fold, $\text{FDR} \leq 0.05$) comparing CTCF^{AID2/dRBR} to CTCF^{AID2/WT}.

Fig. S7. Correlation analysis of DNA-binding and transcriptome changes in CTCF^{AID2} cells.

(a)-(c) Schematic diagram of the integrative analysis of RNA-seq and differential binding of HA-CTCF-ZF mutants ChIP-seq. For each gene, we considered all the peaks within 50kb of TSS or TES. The RNA-seq and ChIP-seq fold-changes were converted to z-score, then multiplied together for each gene-peak pair to get a combined score. Pareto optimization was performed to determine correlated peak-gene pairs. Gene cutoff: DB: $|\log_2(\text{FC})| \geq 1$ and $\text{FDR} \leq 0.05$; DE: $|\text{FC}| \geq 2$ and $\text{FDR} \leq 0.05$. The DE/DB pairs that passed $|\log_2(\text{FC})| \geq 1$ $\text{FDR} \leq 0.05$ were highlighted in orange.

(d) Boxplot showing the absolute differential binding $\log_2(\text{FC})$ of the peaks of each HA-CTCF-ZF mutant identified by the integrative analysis. P-values were calculated using the Wilcoxon rank test.

(e) Venn-diagram illustrating the DE/DB pair genes in CTCF^{AID2/dZF1}, CTCF^{AID2/dRBR} and CTCF^{AID2/dZF10} compared with CTCF^{AID2/WT} had little overlap.

Fig. S1

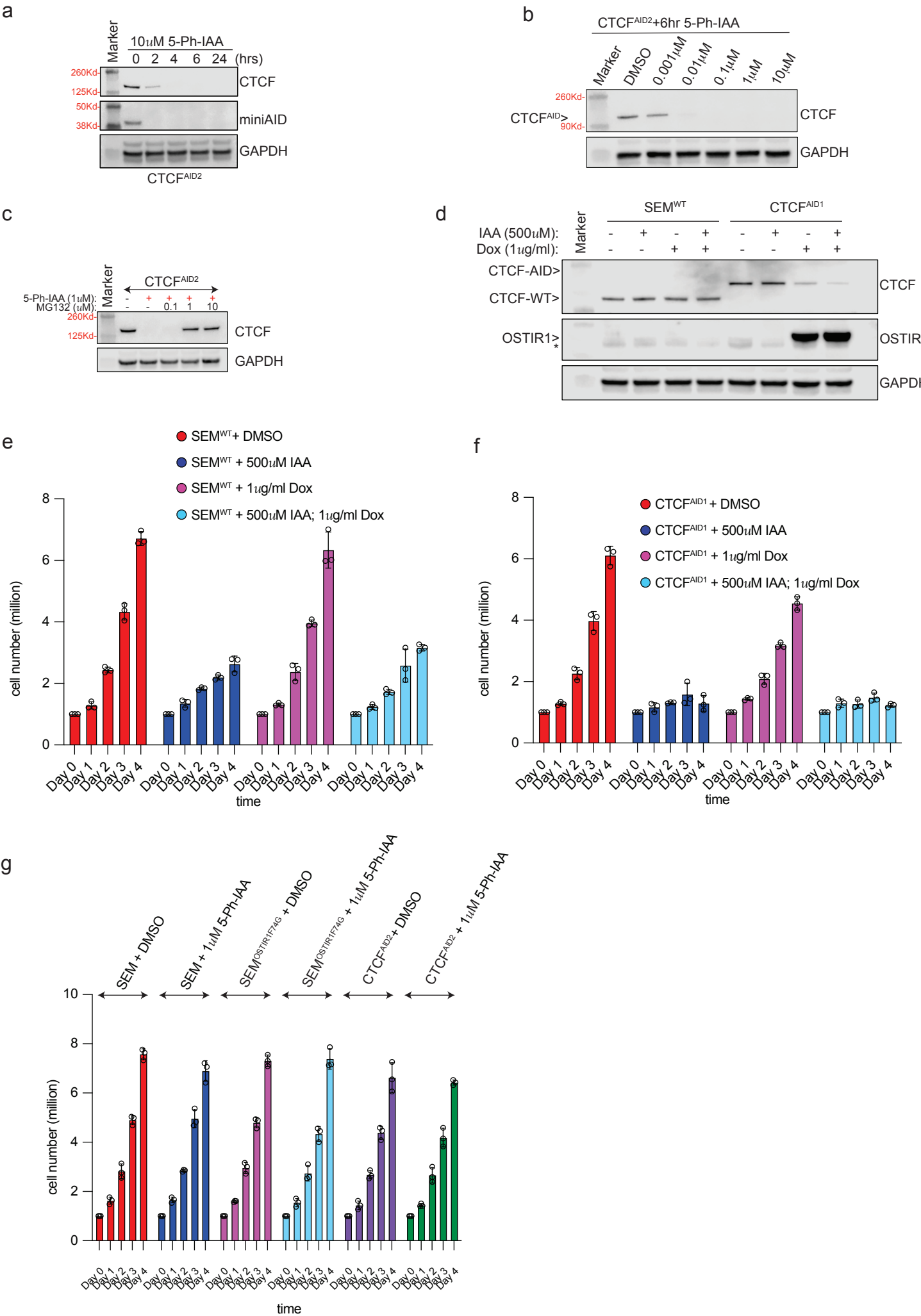


Fig. S2

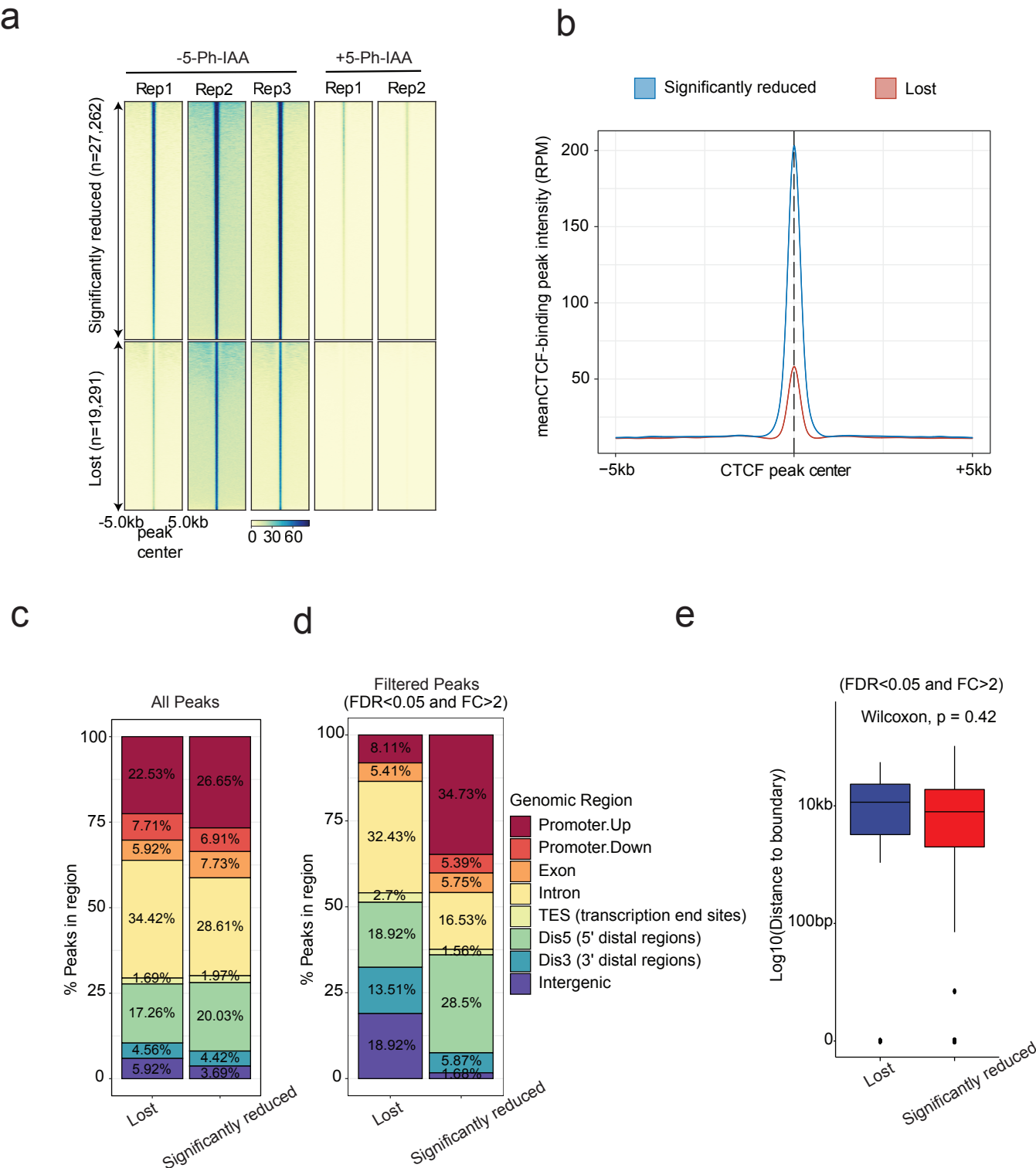


Fig. S3

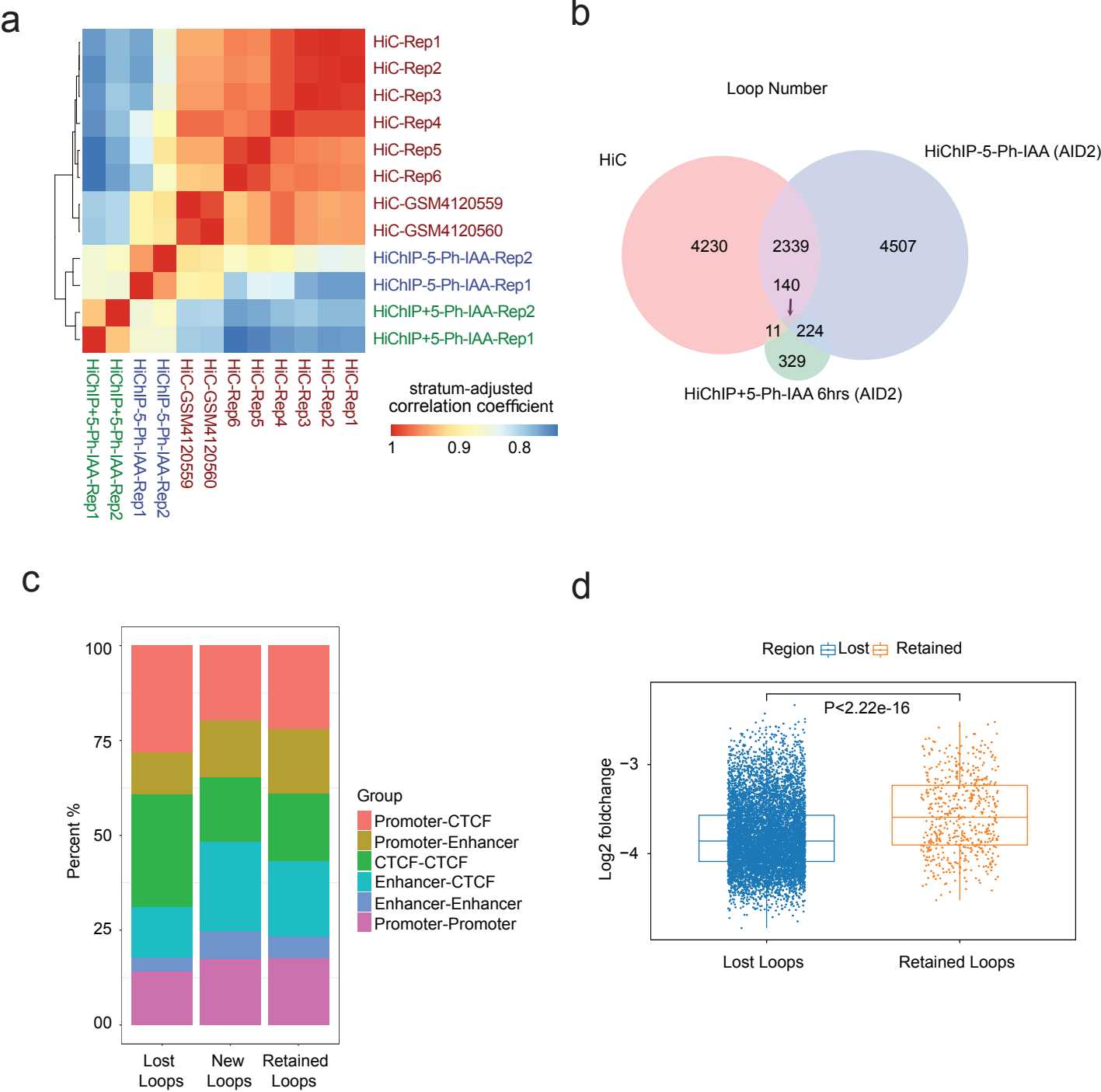
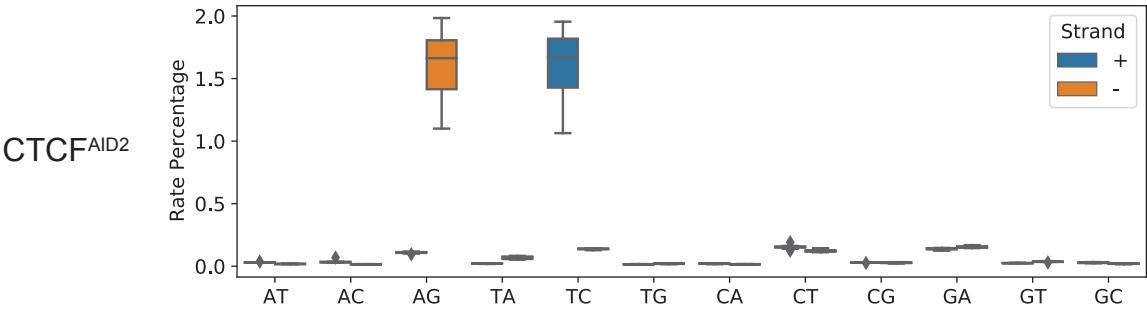
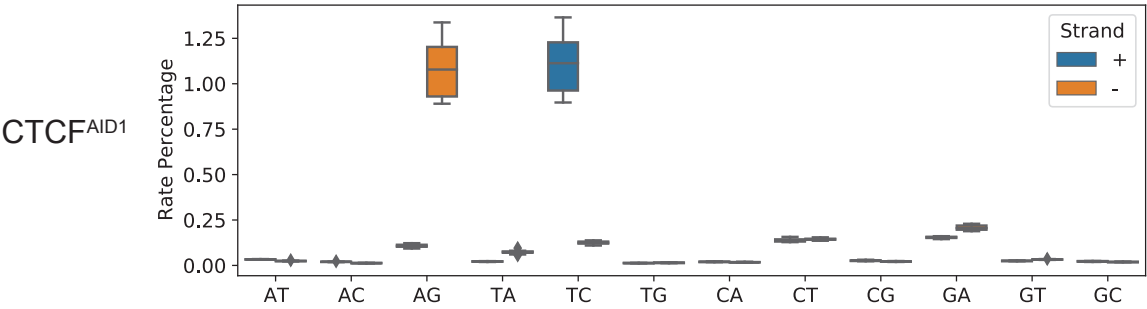


Fig. S4

a



b

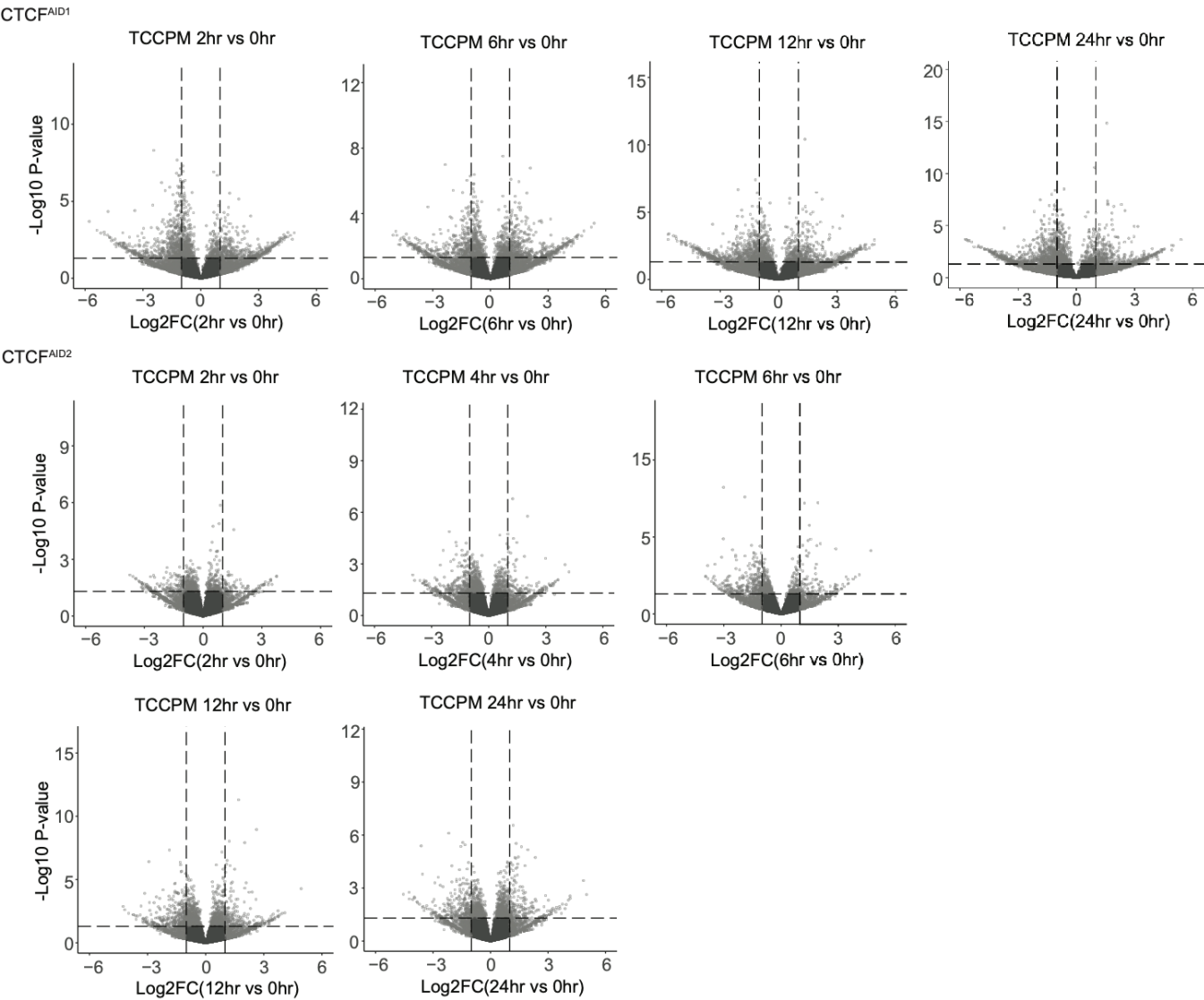


Fig. S5

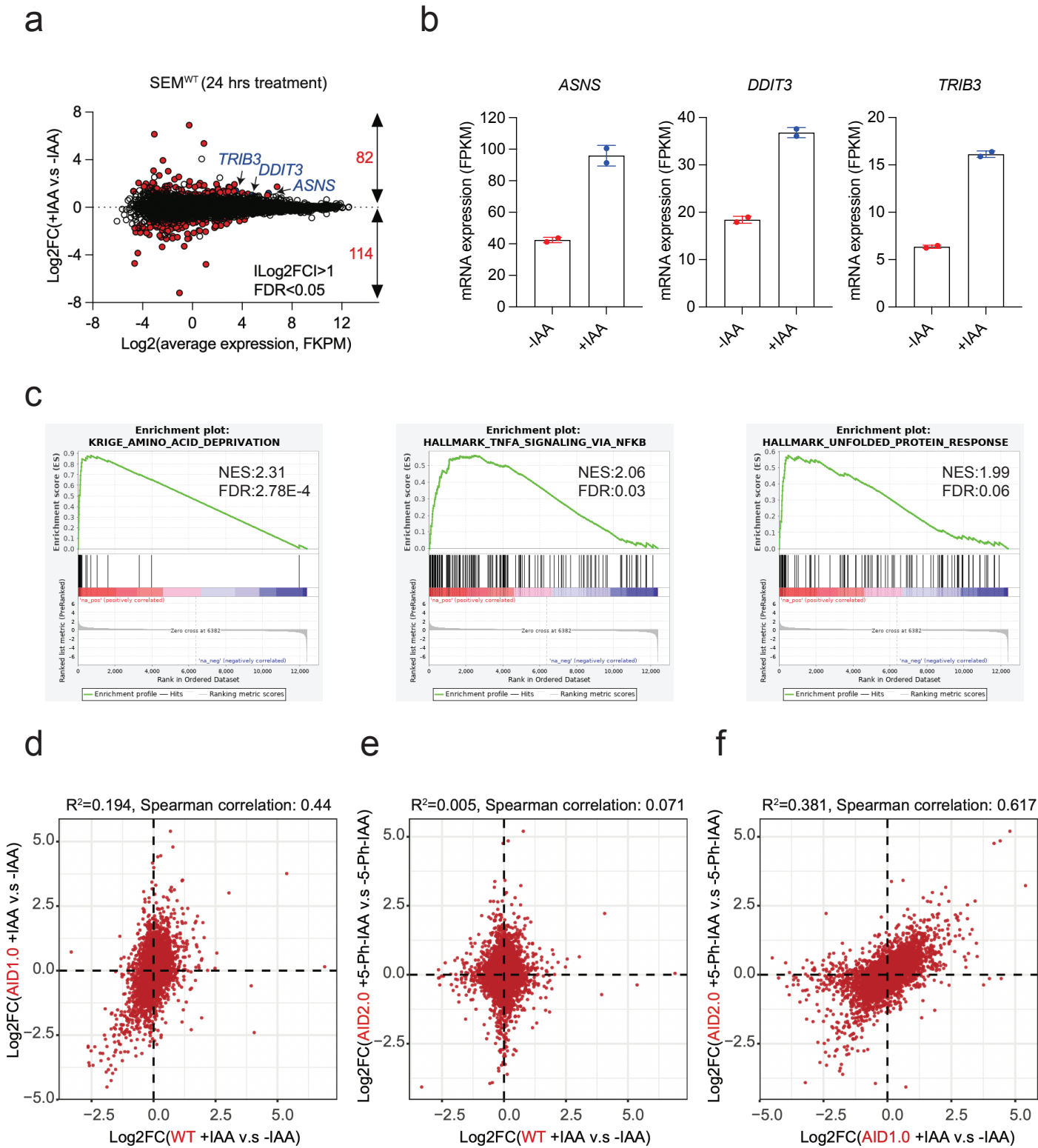
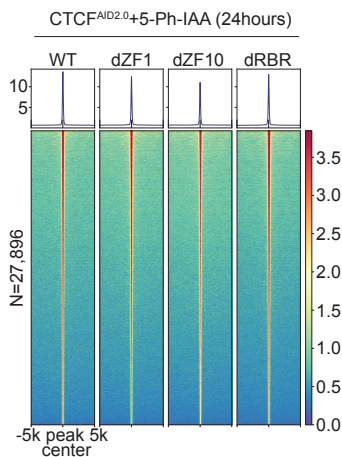


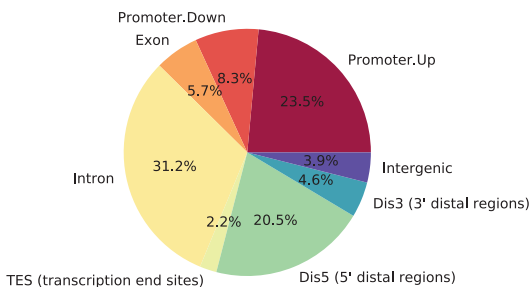
Fig. S6

a



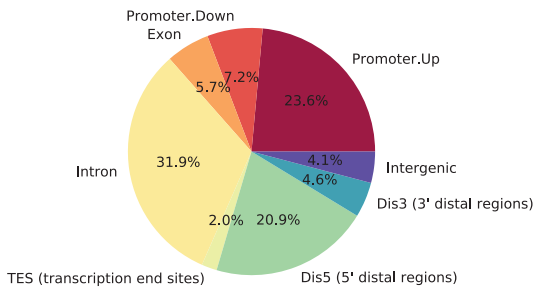
b

27923 regions, AllRegions (All regions from the analysis)



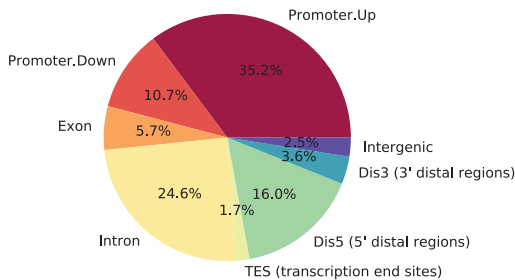
c

dZF10 vs WT
4084 regions, Down2 (>2fold, FDR<0.05)



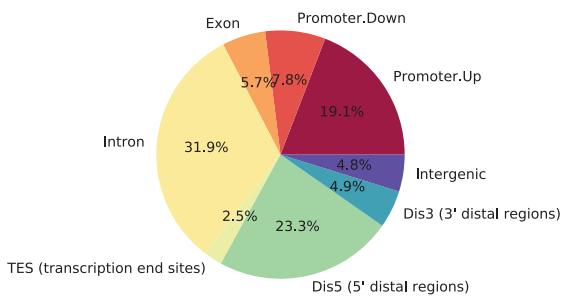
d

dZF10 vs WT
525 regions, Up2 (>2fold, FDR<0.05)



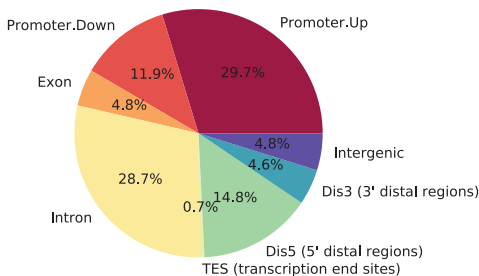
e

dZF1 vs WT
1875 regions, Down2 (>2fold, FDR<0.05)



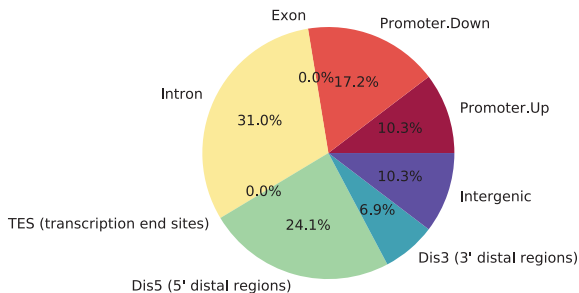
f

dZF1 vs WT
582 regions, Up2 (>2fold, FDR<0.05)



g

dRBR vs WT
29 regions, Down2 (>2fold, FDR<0.05)



h

dRBR vs WT
102 regions, Up2 (>2fold, FDR<0.05)

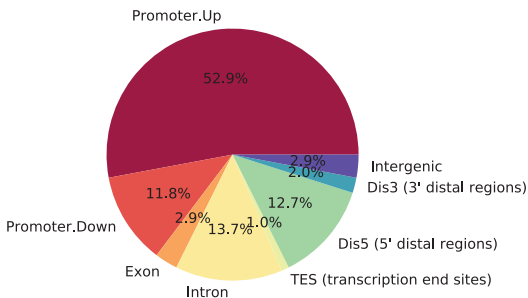


Fig. S7

