SUPPLEMENTAL MATERIALS

Promoter-distal RNA Polymerase II binding discriminates active from inactive CCAAT/enhancer-binding protein beta binding sites

Daniel Savic[1], Brian S. Roberts[1], Julia B. Carleton[2], E. Christopher Partridge[1], Michael A. White[3], Barak A. Cohen[3], Gregory M. Cooper[1], Jason Gertz[2,4] and Richard M. Myers[1,4]

(1) HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA

(2) Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA

(3) Washington University at St. Louis, Center for Genome Sciences and Systems Biology, Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA

(4) Corresponding authors (Richard M. Myers and Jason Gertz)

TABLE OF CONTENTS

SUPPLEMENTAL TABLES

Supplemental Table 1. Genome location (hg19), barcode and DNA sequence information for all CEBPB binding sites and scrambled control sequences.

Supplemental Table 2. The number of total CEBPB binding sites across distinct categories of sites and the number, as well as percentage, of active binding events (driving activity above the 95th percentile cutoff of scrambled control sequences) are given for HepG2 and K562 cells.

Supplemental Table 3. Genomic annotation analysis of CEBPB binding site CRE-seq activity.

Supplemental Table 4. Genomic annotation analysis of cell type-specific CEBPB binding site CRE-seq activity.

SUPPLEMENTAL FIGURES

Supplemental Figure 1. CEBPB binding site analysis. (A) Venn diagram of reproducible CEBPB binding events, found in two independent ChIP-seq replicate experiments in HepG2 (shown in blue) and K562 (shown in red) cells. The number of sites shared in both cell types is shown in gray. (B) Bar plots of the percentage of CEBPB sites in HepG2 and K562 cells that harbor a canonical CEBPB binding motif is given in green, while sites devoid of a CEBPB motif are shown in red.

Supplemental Figure 2. Read fraction comparisons of DNA and RNA. (A) Comparisons of read count measurements for all identical barcodes across biological replicates in HepG2 cells. Data for DNA-derived (plasmid) barcodes are shown at the left while RNA-derived barcodes are given at the right. (B) Comparisons of read count measurements for all identical barcodes across biological replicates in K562 cells. Data for DNA-derived (plasmid) barcodes are shown at the left while RNA-derived barcodes are given at the right.

Supplemental Figure 3. Read tag counts comparisons of DNA and RNA. Pair-wise correlations (R^2) of read tag counts for DNA and RNA is given across biological replicates and for both HepG2 and K562 cells. In addition, correlations are given for read tag counts obtained directly from the plasmid pool (by sequencing the plasmid pool directly without transfection). Strong correlation is observed for biological replicates of RNA measurements within the same cell line. DNA as well as plasmid pool measurements are strong across replicate experiments and across cell lines, suggesting that both cell lines were transfected with the same CRE-seq plasmid pool. Strong correlations are shown in dark blue, lower correlations are given in lighter blue and low correlation is shown in red.

Supplemental Figure 4. Analysis of luciferase activity in CRE-seq activity categories. (A) A plot HepG2 luciferase activities in groups binned by HepG2 CRE-seq activities. Binding sites were ranked on CRE-seq activity (x-axis) and grouped evenly (6 per group) into low activity sites (6 sites exhibiting the lowest HepG2 CRE-seq activity, labeled LOW), middle activity sites (labeled MID), and high activity sites (6 sites with the strongest HepG2 CRE-seq activity, labeled HIGH). Group labels are shown on the x-axis and log2 transformed luciferase activities are shown on the y-axis. Significant differences in activity are observed for both high and middle activity groups when compared to the activities of the low activity group (two-sided t-test;

HIGH versus LOW p < 3.5e-4, MID versus LOW p <0.02). (B) A plot K562 luciferase activities in groups binned by K562 CRE-seq activities. Binding sites were ranked on CRE-seq activity (x-axis) and grouped evenly (6 per group) into low activity sites (6 sites exhibiting the lowest K562 CRE-seq activity, labeled LOW), middle activity sites (labeled MID), and high activity sites (6 sites with the strongest K562 CRE-seq activity, labeled HIGH). Group labels are shown on the x-axis and log2 transformed luciferase activities are shown on the y-axis. Significant differences in activity are observed for the HIGH activity group when compared to the activities of the low activity group (two-sided t-test; HIGH versus LOW p <0.03). For HepG2 cells, 6 out of 6 HIGH sites, 6 out of 6 MID sites and 0 out of 6 LOW sites fell above the 95% cutoff, while for K562 cells, 6 out of 6 HIGH sites, 4 out of 6 MID sites and 0 out of 6 LOW sites fell above the 95% cutoff.

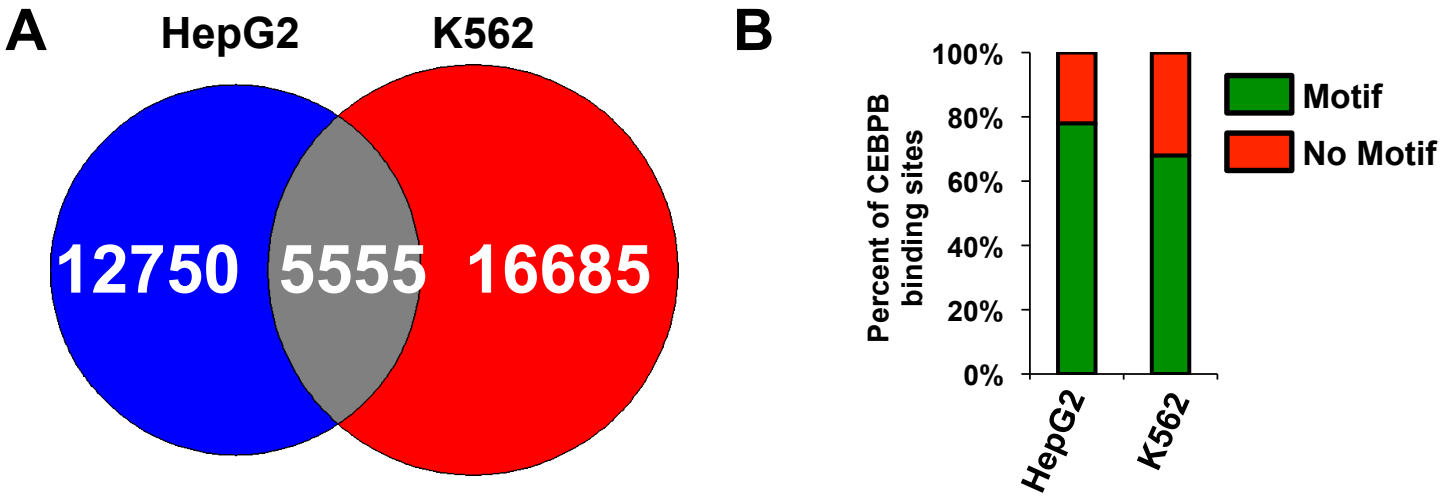
Supplemental Figure 5. Shared CEBPB sites with and without RNAP2 co-occupancy. (A) HepG2 CRE-seq activity for CEBPB sites shared between HepG2 and K562 cells is shown as box plots. The log2-transformed HepG2 CRE-seq activity (RNA/DNA barcode counts) is displayed on the y-axis. Shared binding sites that are coincident with promoter-distal RNAP2 are shown in dark blue and the scrambled control sequences for these RNAP2-associated binding sites are displayed in dark gray. Shared binding sites devoid of promoter-distal RNAP2 binding are displayed in light blue and the scrambled control sequences for these RNAP2-devoid sites are shown in light gray. The percentage of RNAP2-associated and RNAP2-devoid shared sites above the 95th percentile of associated scrambled control sites in HepG2 cells are displayed on the plot. (B) K562 CRE-seq activity for CEBPB sites shared between HepG2 and K562 cells is shown as box plots. The log2-transformed K562 CRE-seq activity (RNA/DNA barcode counts) is displayed on the y-axis. Shared binding sites that are coincident with promoter-distal RNAP2 are shown in dark red and the scrambled control sequences for these

RNAP2-associated binding sites are displayed in dark gray. Shared binding sites devoid of promoter-distal RNAP2 binding are displayed in light red and the scrambled control sequences for these RNAP2-devoid sites are shown in light gray. The percentage of RNAP2-associated and RNAP2-devoid shared sites above the 95th percentile of associated scrambled control sites in K562 cells are displayed on the plot.

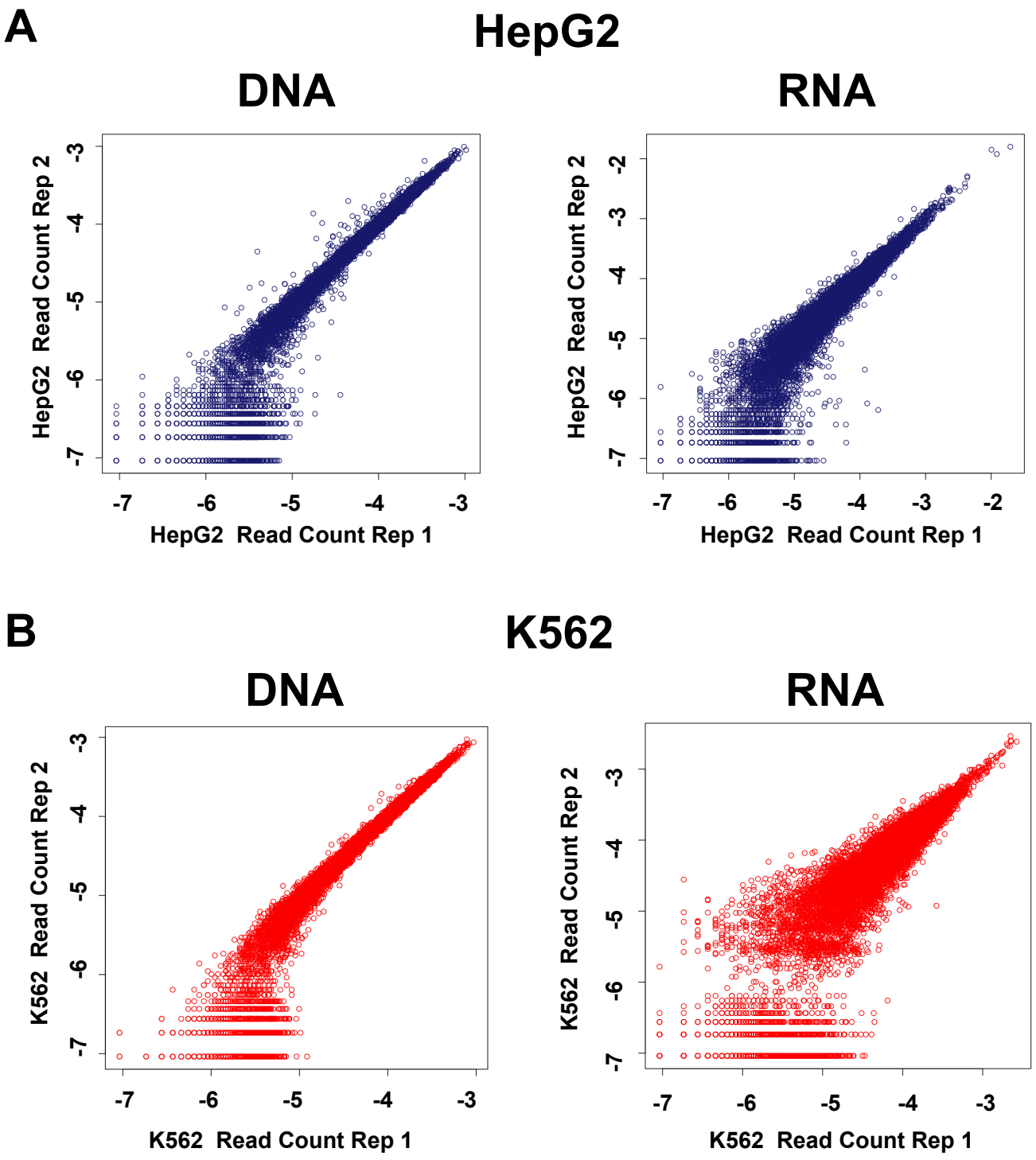
Supplemental Figure 6. RNAP2 ChIP-seq signal at active and inactive sites. (A) Box plots display RNAP2 HepG2 ChIP-seq signal enrichments (y-axis) at active HepG2 CEBPB sites above 95th percentile of control sequences (labeled Active sites) and inactive HepG2 CEBPB sites with activity below 95th percentile of control sequences sites in HepG2 cells (labeled Inactive sites). (B) Box plots display RNAP2 K562 ChIP-seq signal enrichments at active K562 CEBPB sites above 95th percentile of control sequences (labeled Active sites) and inactive K562 CEBPB sites with activity below 95th percentile of control sequences sites in K562 cells (labeled Inactive sites). RNAP2 occupancy at CEBPB sites was assessed through normalized read counts (reads per million total sequenced reads) and is displayed on the y-axis. Significantly higher signal (p= 3.211e-10 in HepG2 cells and p=4.059e-08 in K562 cells) was observed at active sites in both cell lines. P-values are shown on each graph.

Supplemental Figure 7. eRNA GRO-seq signal at active and inactive sites. (A) Box plots display eRNA GRO-seq signal enrichments (shown on the y-axis) at active K562 CEBPB sites above 95th percentile of control sequences (labeled Active sites) and inactive K562 CEBPB sites with activity below 95th percentile of control sequences sites in K562 cells (labeled Inactive sites). Significantly higher signal (p= 3.211e-10) was observed at active sites. The p-value is shown on the graph.

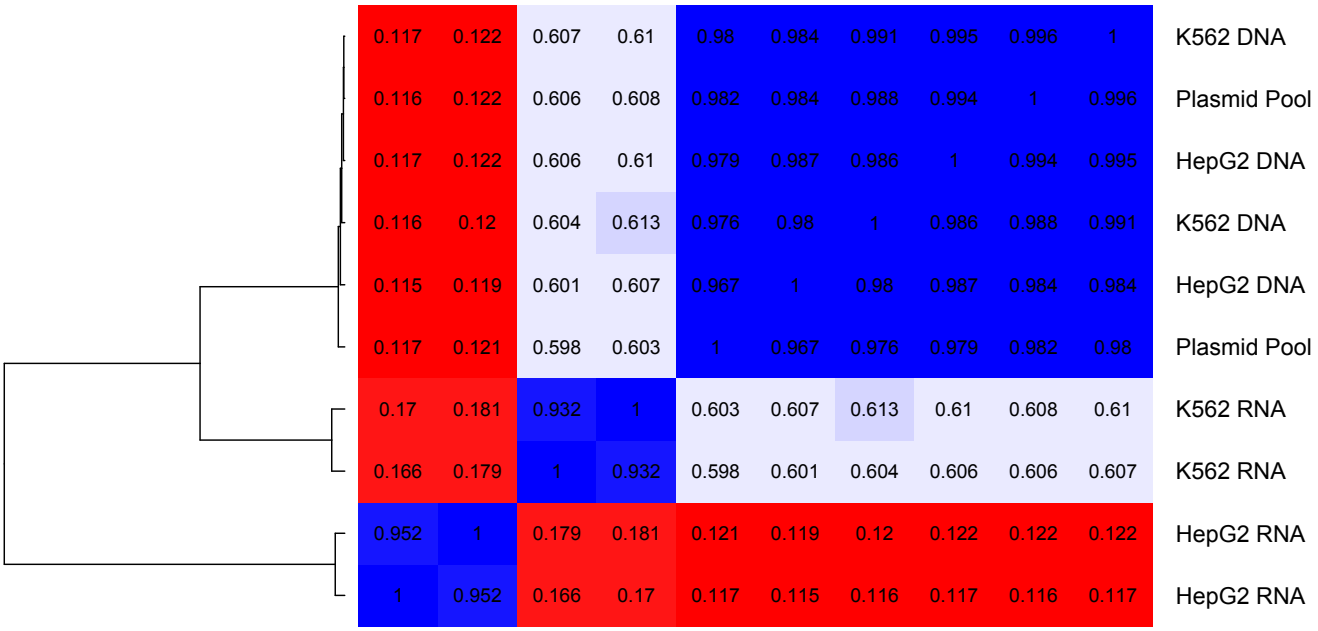
Supplemental Figure 1



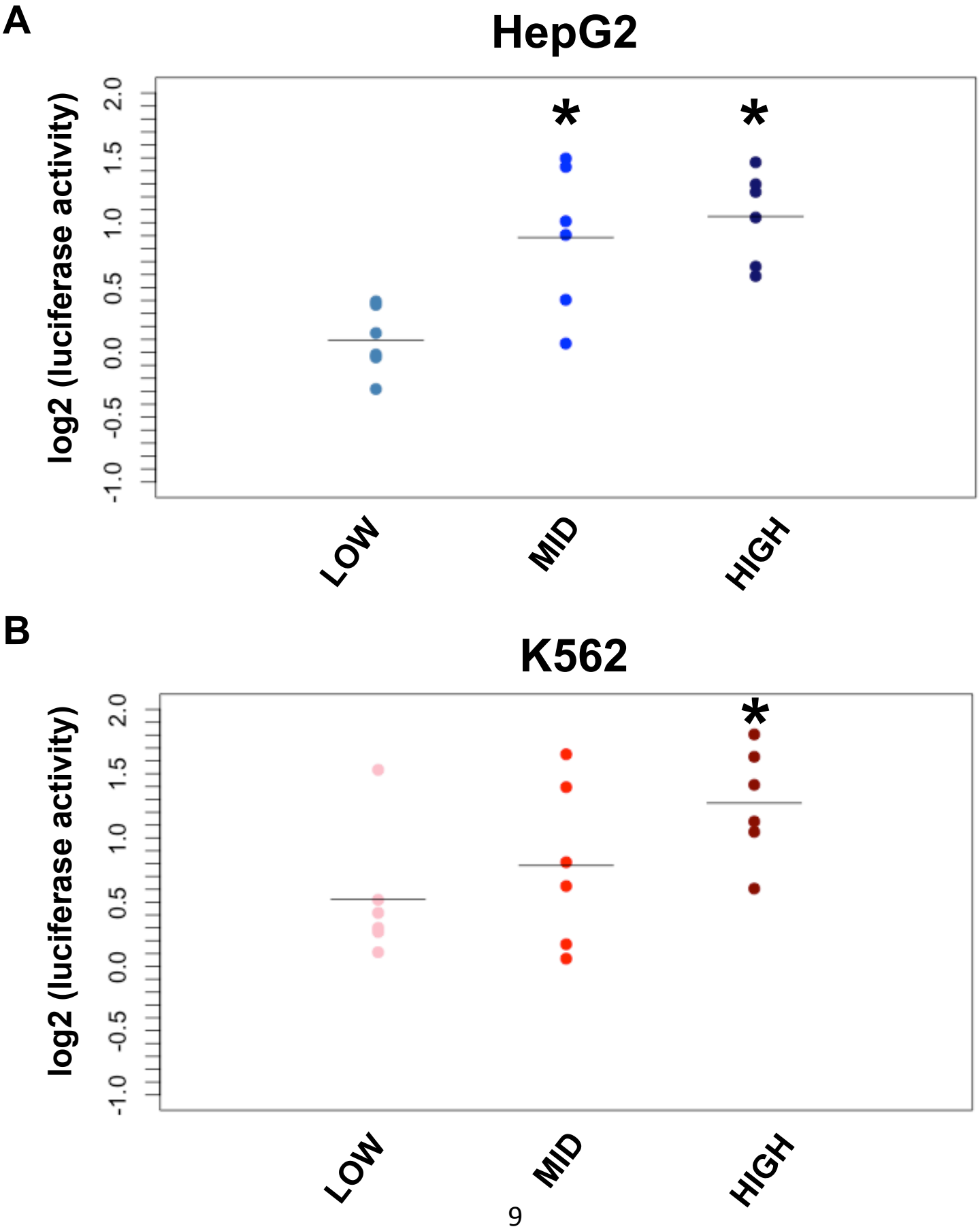
Supplemental Figure 2



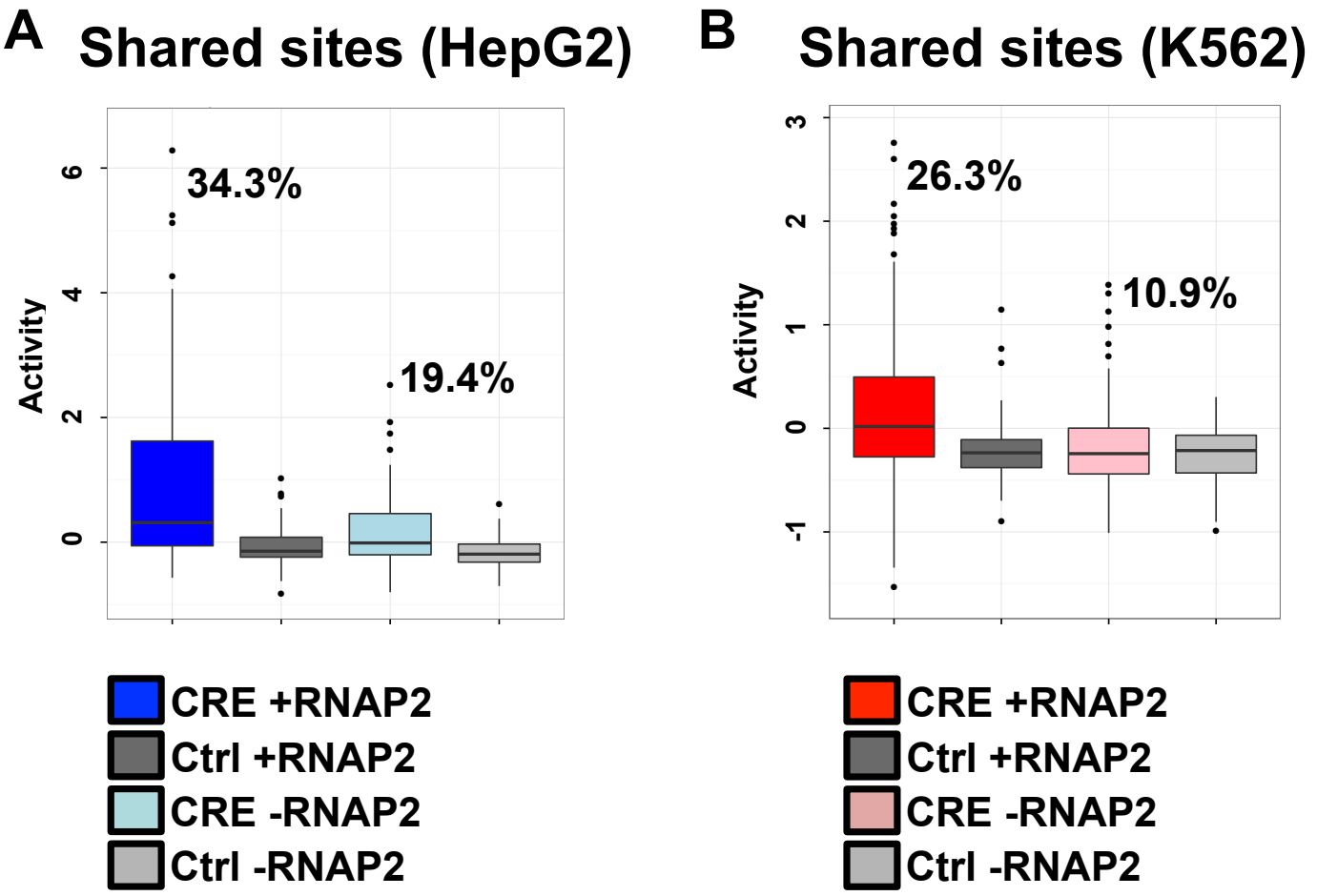
Supplemental Figure 3



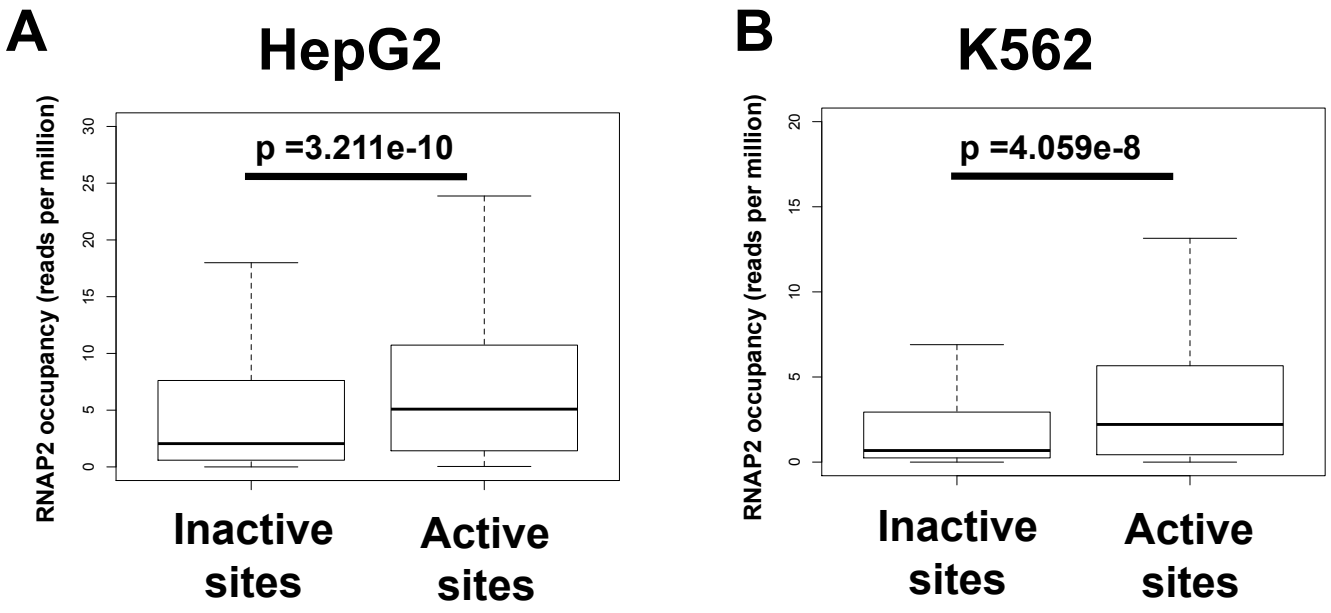
Supplemental Figure 4



Supplemental Figure 5



Supplemental Figure 6



Supplemental Figure 7