Correcting signal biases and detecting regulatory elements in STARR-seq data.

Supplemental Material

Table of contents

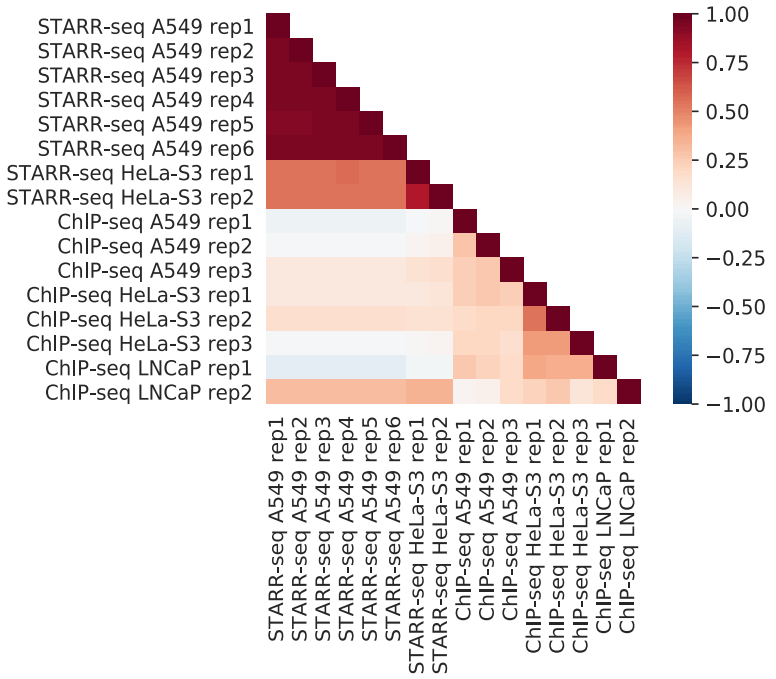


Figure S1 | Pearson's correlations of 1 bp signals in STARR-seq input and ChIP-seq control libraries along Chromosome 1.

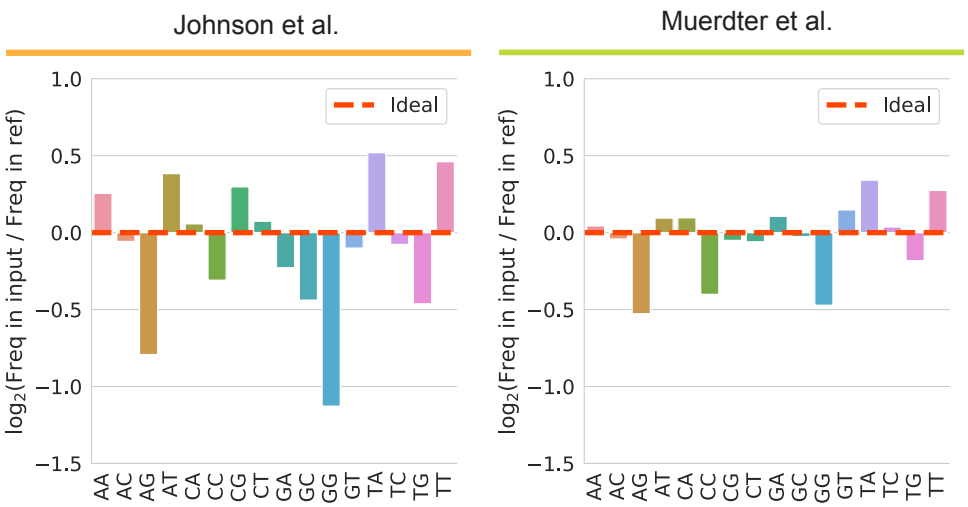


Figure S2 | DNA structure bias in the terminal positions of STARR-seq fragments. The frequencies of observed dimers starting from one bp external to the 5' ends of fragments was compared to that in reference autosomes (hg38) excluding excluding gap, centromere, and telomere that are available in UCSC Gap and Centromere table browser and ENCODE blacklist regions.

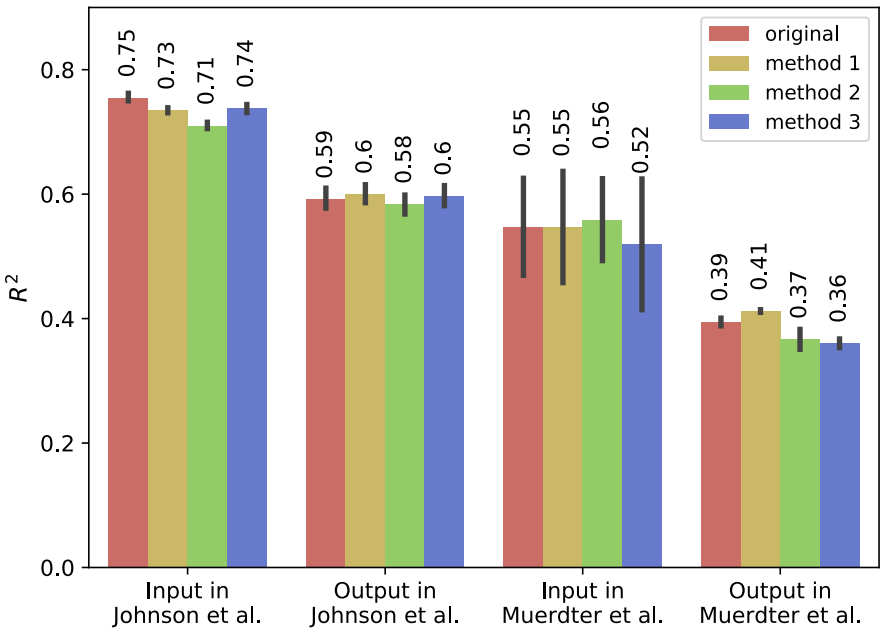


Figure S3 | Robustness of the structured sampling approach. We slightly modified structured sampling strategy (method 1-3) and calculated R^2 values. We used following modified approaches; Method 1, sampling 60% of a training set from high signal regions; Method 2, sampling 40% of a training set from high signal regions; Method 3, sampling regions that do not have high signal with two bins instead of four bins.

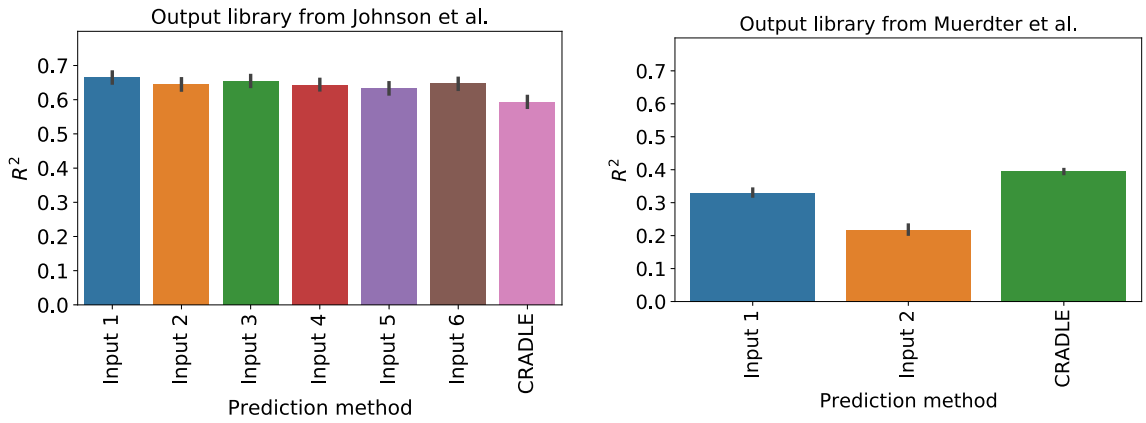


Figure S4 | R^2 values of output libraries in Johnson et al. and Muedter et al., using input library and CRADLE- predicted signal. In Johnson et al., we used 0hr-dex-treated output library. The error bar indicates variance between replicates of the output library. The number following 'Input' is the replicate number.

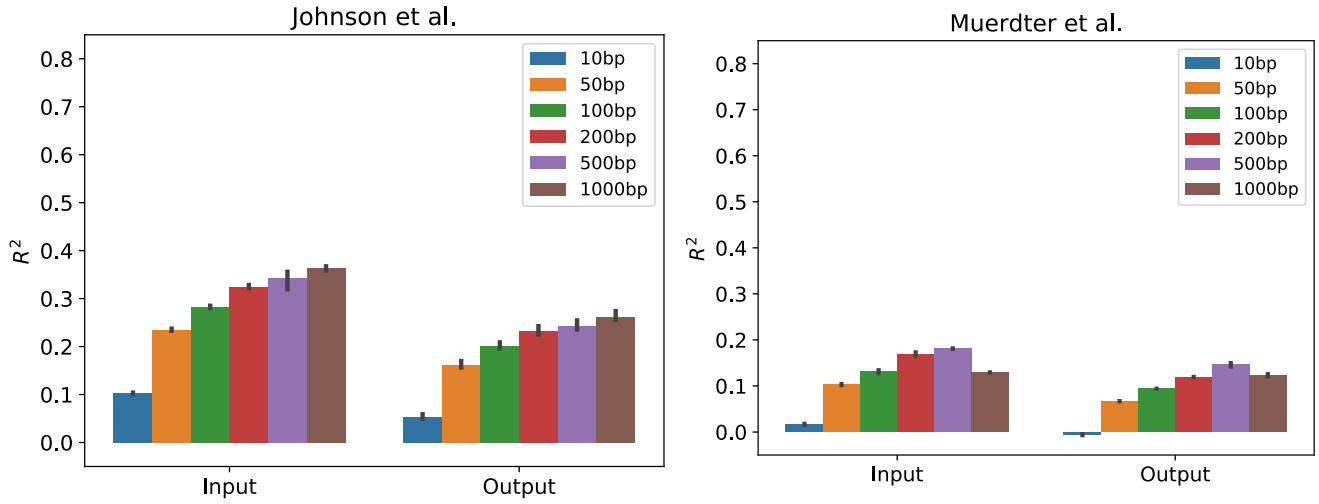


Figure S5 | Predicting biases with GC content model. We binned genome with non-overlapping sliding windows with six different sizes (10bp-1000bp) and used GC content in each bin as a covariate in fitting Poisson GLM. We used Chr1 to calculate R^2 values.

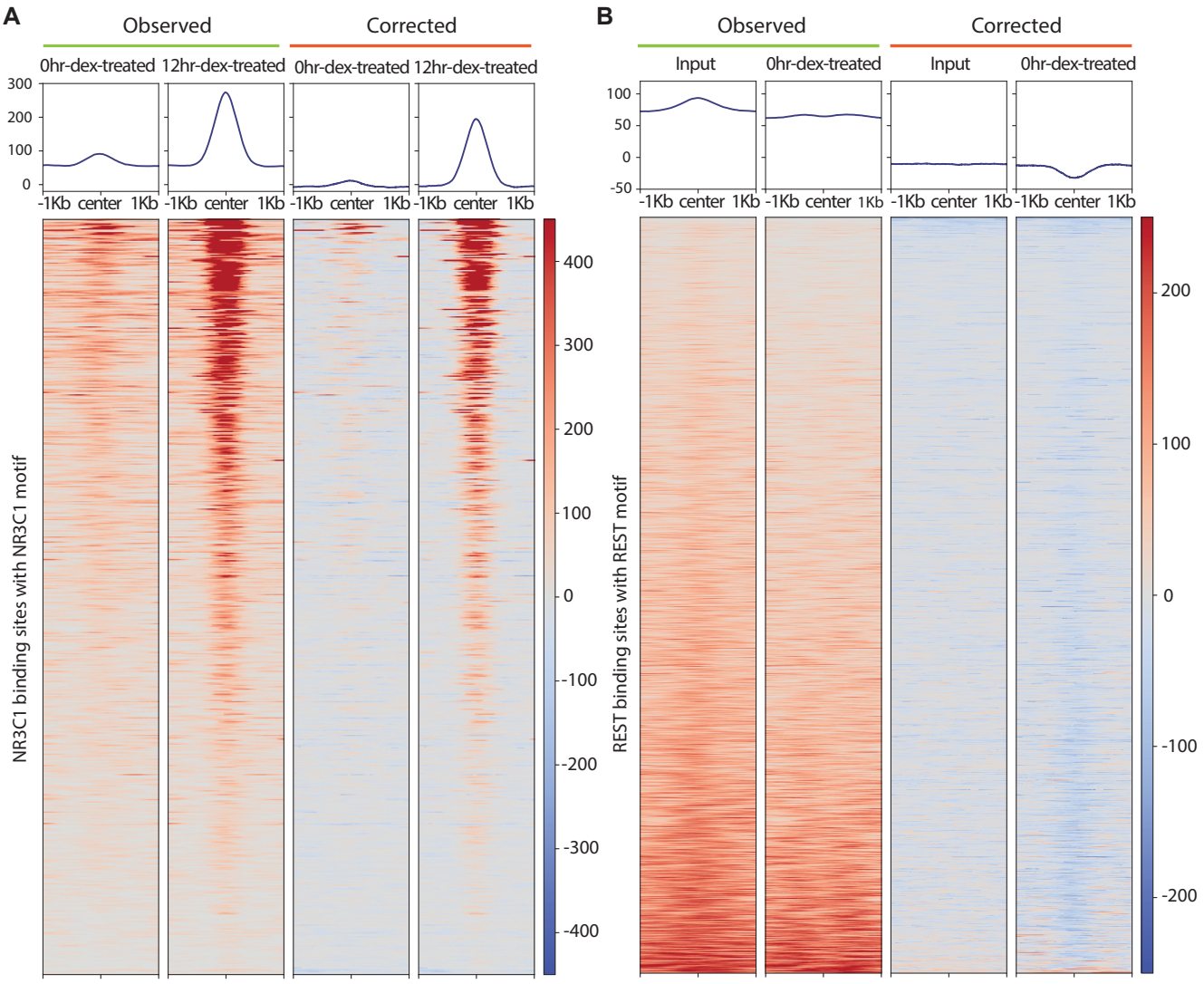


Figure S6 | Comparison of observed and corrected signal in Johnson et al. data. (A)-(B) The top line plots are the mean signal in each genomic position. (A) Observed and corrected signal of 0hr-dex-treated and 12hr-dex-treated output library for GR-binding sites that have GR motif (n=611). (B) Observed and corrected signal of input library and 0hr-dex-treated output library for REST-binding sites that have REST motif (n=2004)

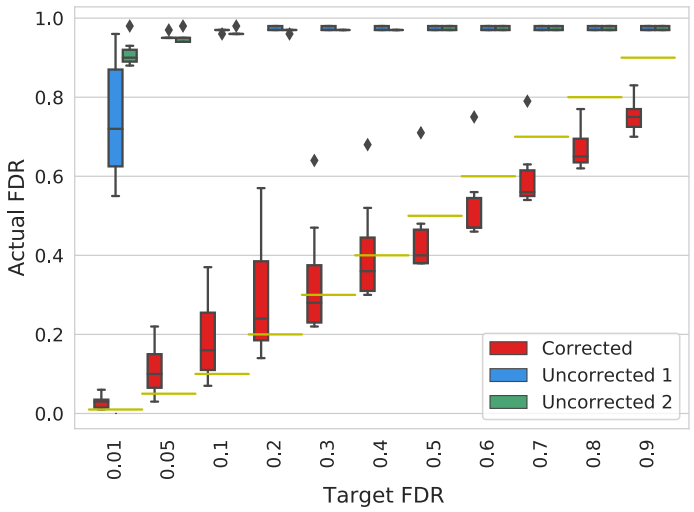
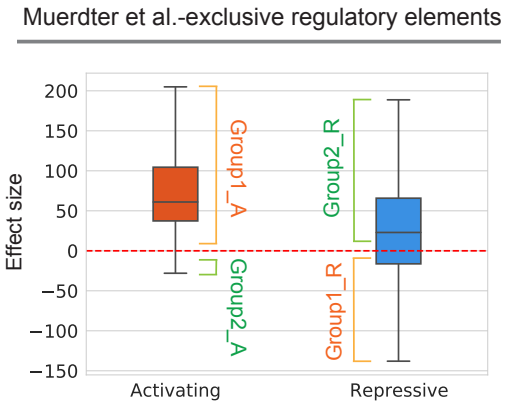


Figure S7 | Determination of actual FDR for regulatory elements detected by CRADLE using simulated STARR-seq data. The relationship between parameterized target FDR values and actual FDR values calculated using simulated corrected (red) and uncorrected STARR-seq signals. To detect regulatory elements with uncorrected signals, two statistical approaches were used (see methods; 'Uncorrected 1', blue; 'Uncorrected 2', green). The yellow line indicates target FDR. Whiskers extend 1.5 times the interquartile range. Center lines show the medians.



Moitfs enriched in Group1_A compared to Group2_A

Rank	Motif	Name	P value
1		HNF1B	10^{-355}
2		EPAS1	10^{-335}
3		GRHL2	10^{-246}
4		NFkB-p65	10^{-229}

Moitfs enriched in Group1_R compared to Group2_R

Rank	Motif	Name	P value
1		IRF3	10^{-52}
2		IRF2	10^{-52}
3		ISRE	10^{-43}
4		IRF8	10^{-38}

Moitfs enriched in Group2_A compared to Group1_A

Rank	Motif	Name	P value
1		PDX1	10^{-3}
2		IRF2	10^{-2}
3		ISRE	10^{-2}

Moitfs enriched in Group2_R compared to Group1_R

Rank	Motif	Name	P value
1		NFkB-p65	10^{-26}
2		BACH1	10^{-15}
3		JUN	10^{-15}

Figure S8 | CRADLE more accurately estimates regulatory element effect sizes. CRADLE effect sizes were plotted for regulatory elements exclusively called by Muerdter et al. Activating and repressive regulatory elements were subsetted according to the sign of their effect size into Group1 and Group2. Group1 subsets includes activating regulatory elements with a positive effect size (Group1_A) and repressive regulatory elements with a negative effect size (Group1_R). In contrast, Group 2 subsets includes activating regulatory elements with a negative effect size (Group2_A) and repressive regulatory elements with a positive effect size (Group2_R). Motif enrichment analysis was performed for each group relative to its partner.

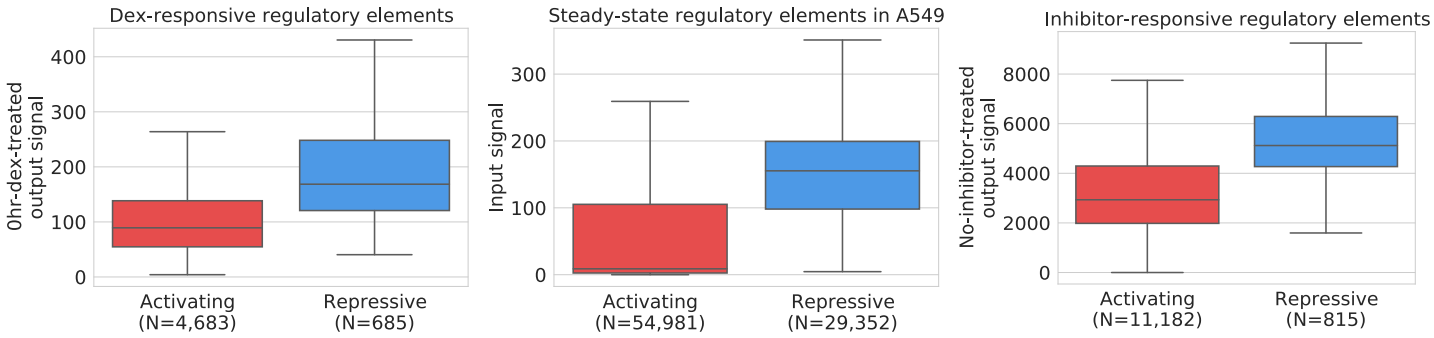


Figure S9 | Distribution of control library signals for regulatory elements detected by CRADLE in Johnson et al. and Muerdter et al. studies. Whiskers extend 1.5 times the interquartile range. Center lines show the medians.

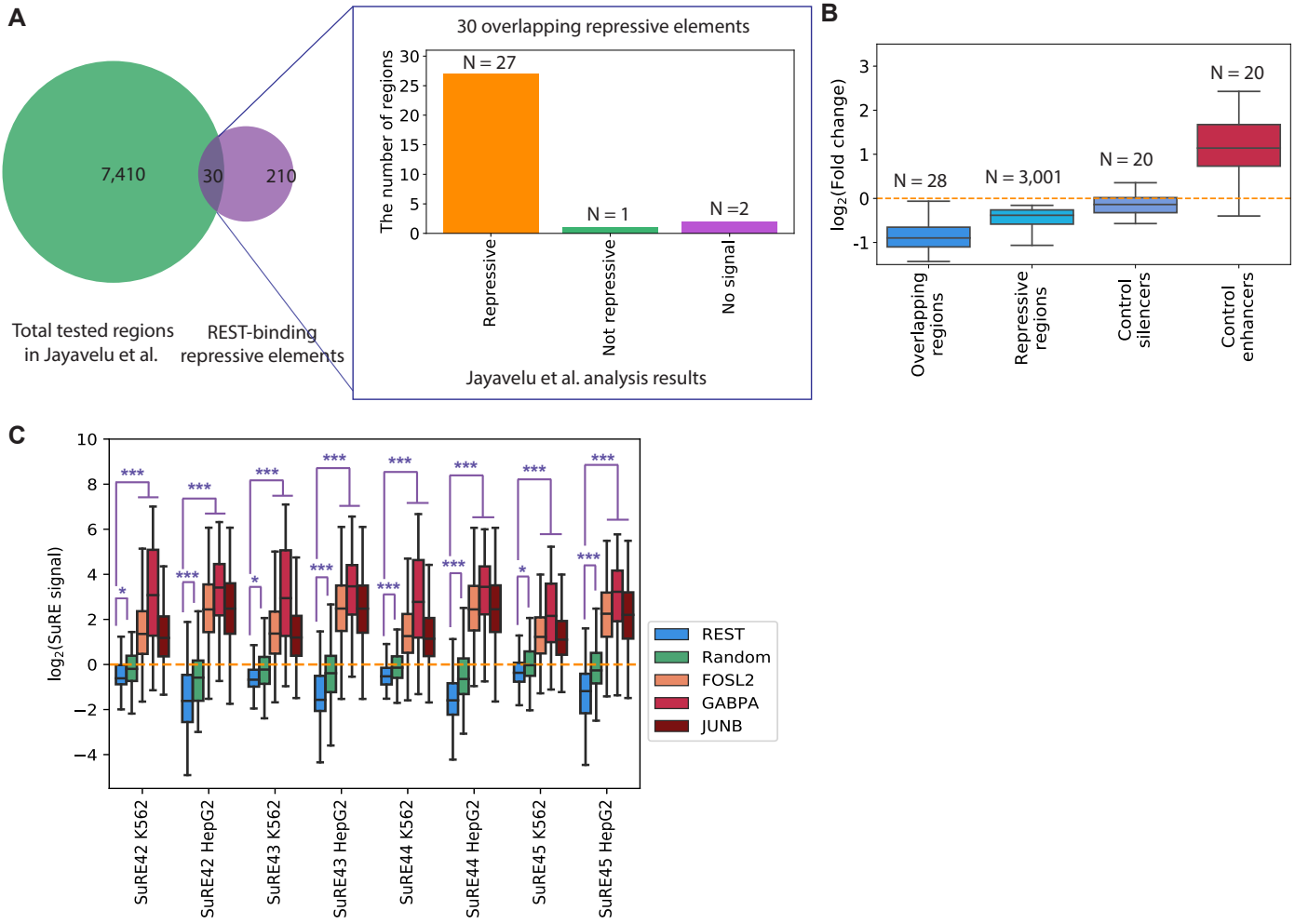


Figure S10 | Validation of REST-binding A549 steady-state repressive regulatory elements identified by CRADLE. (A) Venn diagram showing the intersection of tested regions in Jayavelu et al. and steady-state REST-binding repressive elements in A549 cells. Among the 30 elements in the intersection, 27 elements were previously reported to be repressive. (B) The distribution of previously-reported fold changes for the elements in the intersection as well as repressive and control regions in the prior study. The two elements without coverage in the intersection were not included. (C) Whole genome survey of regulatory elements (SuRE) signals from HepG2 and K562 cells were compared in subsets of regulatory elements identified by CRADLE in A549 cells. These regulatory elements included activating regulatory elements that contained either a FOSL2, GABPA, or JUNB motif and were bound by the corresponding TF in A549, repressive elements that likewise contained a REST motif and were bound by REST, or a set of randomly generated regions. Whiskers extend 1.5 times the interquartile range. Center lines show the medians. Wilcoxon rank-sum test was used for statistical testing. P-values from: ***, P-value < 0.0001; *, 0.001 < P-value < 0.01.

Supplemental Table

Supplemental Table 1 (separate file)
Inhibitor-responsive regulatory elements in Muerdter et al. data that were detected by CRADLE. Regulatory element type indicates activating and repressive regulatory elements for '1' and '-1', respectively.

Supplemental Table 2 (separate file)
Motifs enriched in inhibitor-responsive repressive regulatory elements in Muerdter et al., exclusively detected by CRADLE.

Supplemental Table 3 (separate file)
Motifs enriched in inhibitor-responsive repressive regulatory elements exclusively reported by Muerdter et al.

Supplemental Table 4 (separate file)
Motifs enriched in inhibitor-responsive repressive regulatory elements shared by CRADLE and Muerdter et al.

Supplemental Table 5 (separate file)
Regulatory elements in untreated A549 cells in Johnson et al. data that were detected by CRADLE. Regulatory element type indicates activating and repressive regulatory elements for '1' and '-1', respectively.

Supplemental Table 6 (separate file)
Dex-responsive regulatory elements in A549 cells in Johnson et al. data that were detected by CRADLE. Regulatory element type indicates activating and repressive regulatory elements for '1' and '-1', respectively.

Supplemental Table 7 (separate file)
Motif enriched in repressive regulatory elements in untreated A549 cells in Johnson et al. data that were detected by CRADLE.

Supplemental Table 8 (separate file)
Motifs enriched in dex-responsive activating regulatory elements in Johnson et al. data exclusively detected by CRADLE.

Supplemental Table 9 (separate file)
Motifs enriched in dex-responsive activating regulatory elements shared by CRADLE and Johnson et al.

Supplemental Table 10 (separate file)
Motifs enriched in dex-responsive repressive regulatory elements in Johnson et al. data exclusively detected by CRADLE.

Supplemental Code

CRADLE code is provided in Supplemental code as a separate file.