

---

## **Supplementary information**

---

# **A systematic evaluation of the design and context dependencies of massively parallel reporter assays**

---

In the format provided by the  
authors and unedited

## Supplemental Note 1

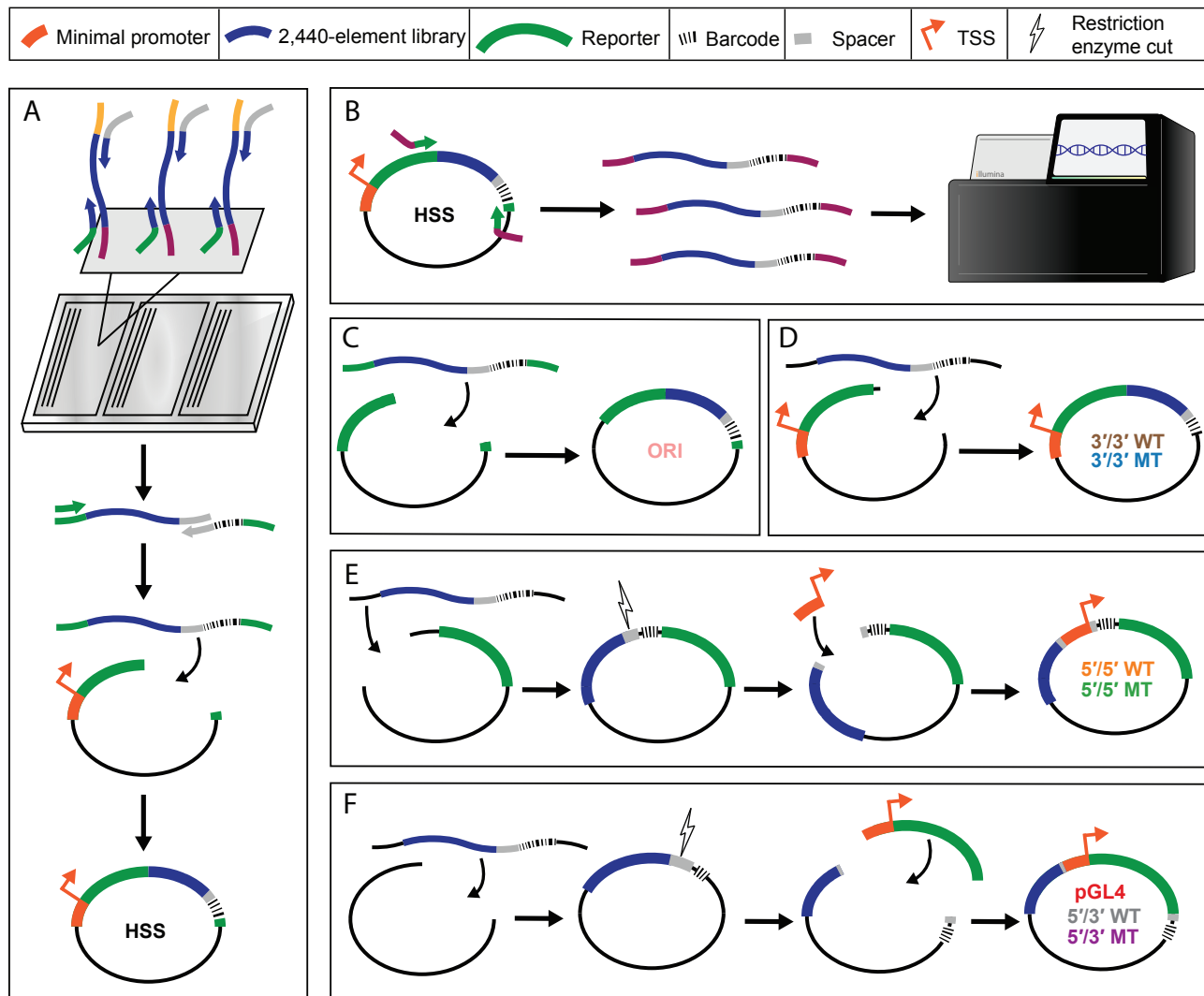
First, although assays like STARR-seq wherein the enhancer serves as the barcode are more straightforward to implement, our understanding of how position (3' of the promoter, rather than 5' as in a more conventional reporter assay vector) or the fact that the sequence is serving as the barcode, influences results, remains incomplete. Arnold *et al.* notably benchmarked STARR-seq against 142 conventional luciferase assays ( $r = 0.83$ ), but STARR-seq has yet to be systematically compared to other MPRAs<sup>16</sup>. Second, although we previously showed differences between episomal vs. integrated MPRAs<sup>17</sup>, it is not clear how these differences rank relative to those resulting from other design choices. Third, although the orientation-independence of enhancers has been evaluated in *Drosophila*<sup>16,25,26</sup>, to our knowledge the robustness of this assumption has not previously been systematically tested in a mammalian system. Finally, the typical choice to test <200 bp fragments, each corresponding to a putative enhancer, is entirely based on technical limitations of massively parallel DNA synthesis, rather than on any principled understanding of the actual size of enhancers. The consequences of this choice for the results obtained remain largely unquantified.

## Supplemental Note 2

Potential reasons for inter-feature correlations include: i) several paralogous factors may bind to nearly identical motifs, leading to overlapping sequence-based binding predictions, ii) several factors may bind in a complex, leading to a redundant ChIP-seq signal, and iii) an upstream factor might lead to the deposition or removal of multiple epigenetic marks simultaneously, leading to a correlated or anti-correlated signal. Of note, another potential confounder is that the criteria used to select elements relied on features including FOXA1/2, HNF4A, EP300, H3K27ac, CHD2, RAD21, and SMC3 ChIP peaks<sup>17</sup>, which could exaggerate the contributions of these factors, or bias the contributions of other features, in models attempting to predict general enhancer activity in HepG2 cells.

## Supplemental Note 3

To test the possibility that barcode counts include short transcripts initiating within the candidate enhancer itself, we repeated the HSS and ORI experiments using two different 5' primers -- our original primer, which hybridizes immediately upstream of the barcode, or a new primer, which hybridizes upstream of the candidate enhancer. The median number of barcode counts for all replicates was greater than 100 (**Supplemental Figure 17A**). Estimated activities in HSS and ORI experiments with these distinct primer sets were correlated, which is modestly reassuring but does not completely rule out this potential confounder of this assay design, particularly as these are preselected enhancer candidates that might initiate transcription poorly (**Supplemental Figure 17A-B**; HSS: Pearson  $r = 0.89$ ; Spearman  $\rho = 0.90$ ; ORI: Pearson  $r = 0.84$ ; Spearman  $\rho = 0.83$ ; for comparison, Pearson  $r$  for biological replicates ranged from 0.92-1.0). Of note, there was a significant batch effect between these vs. our original experiments, which were performed months earlier with different reagent kits and at a later cell passage, highlighting the importance of controlling and documenting technical variables whenever possible (**Supplemental Figure 17C-D**).



### Supplemental Figure 1. Cloning of MPRA vector libraries.

**A)** We amplified an existing array library from Inoue *et al.*<sup>18</sup>, adding homology to the STARR-seq vector on the 5' end and a spacer sequence for cloning on the 3' end. The library was put into a second PCR reaction adding barcodes and homology to STARR-seq to the 3' end, and cloned sequences into the STARR-seq vector.

**B)** We amplified enhancers from the STARR-seq vector library, adding flow cell adapters to the sequences, and sequenced the library on a NextSeq Mid 300 cycle kit to associate enhancer sequences to their corresponding barcodes.

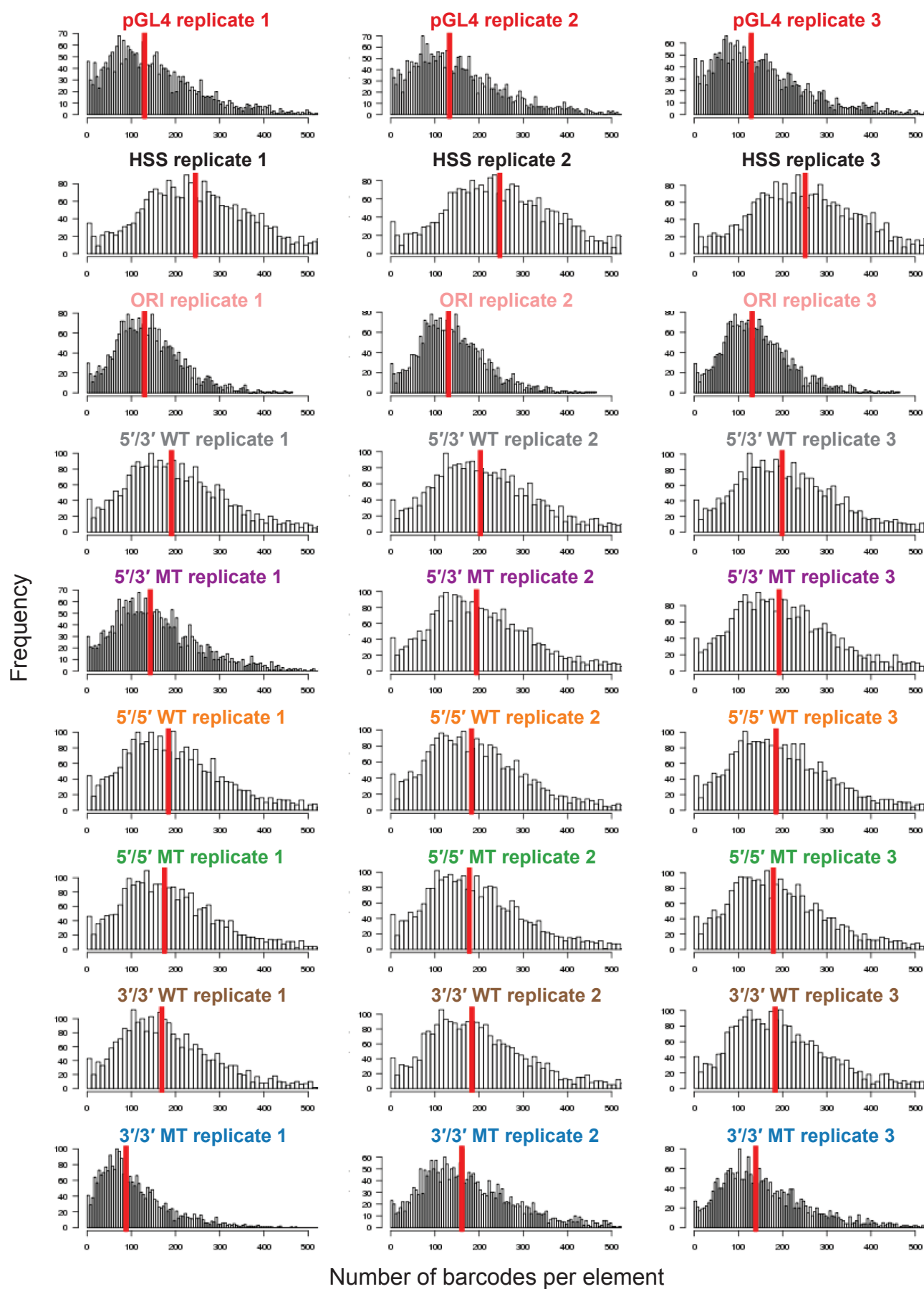
**C)** The STARR-seq ORI assay was cloned by amplifying enhancers from the STARR-seq library while adding homology to the STARR-seq ORI vector, and cloned into the linearized ORI backbone.

**D)** We cloned the 3'/3' lenti- and mutant lentiMPRA libraries by first linearizing the pLSmP backbone without removing the minimal promoter and GFP. The enhancer library and barcodes were then amplified from the STARR-seq vector adding homology to the linearized backbone, and cloned into the vector.

**E)** To clone the 5'/5' lenti- and mutant lentiMPRA libraries, we linearized the pLSmP vector such that the minimal promoter was removed. We then linearized plasmids at the spacer sequence and cloned in the minimal promoter such that it was between the enhancer and barcode sequences.

**F)** The pGL4.23c episomal assay and the 5'/3' lenti- and mutant lentiMPRA libraries were cloned in the same manner as the 5'/5' vector library, except the pGL4.23c and pLSmP backbones were initially linearized to remove both the minimal promoter and reporter gene; after cloning in our library, the minimal promoter and reporter were inserted into the spacer sequence such that enhancers were upstream and barcodes downstream of the reporter.

Figure S2



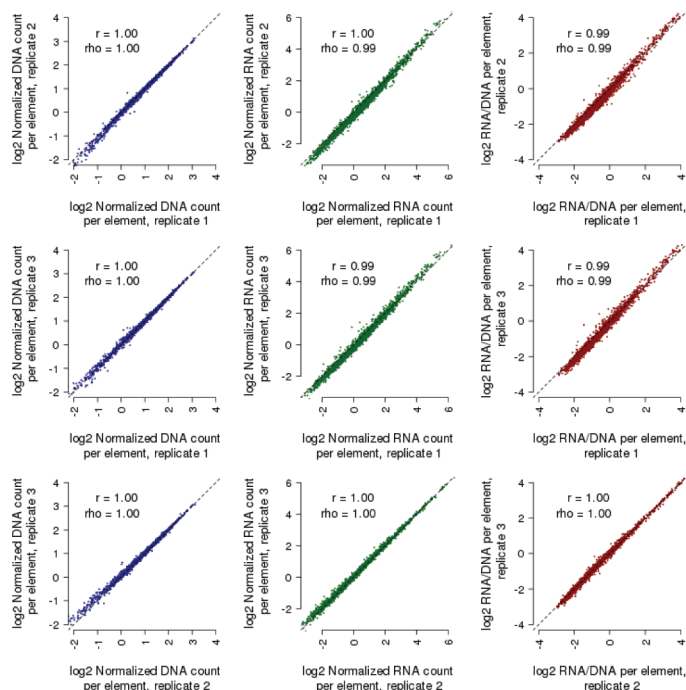


**Supplemental Figure 2. Histograms of barcodes per element for each replicate of the 9 MPRA methods.** Shown are histograms indicating the number of observed barcodes per element, for each of the 2,440 elements tested in each of the 3 replicates for each of the 9 MPRA methods. Shown with a vertical red line is the median number of barcodes per element.

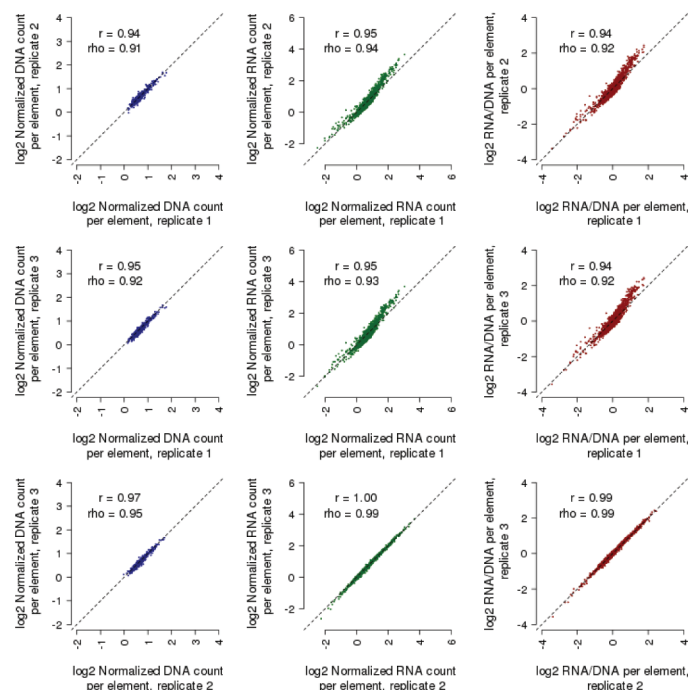
Figure S3

A

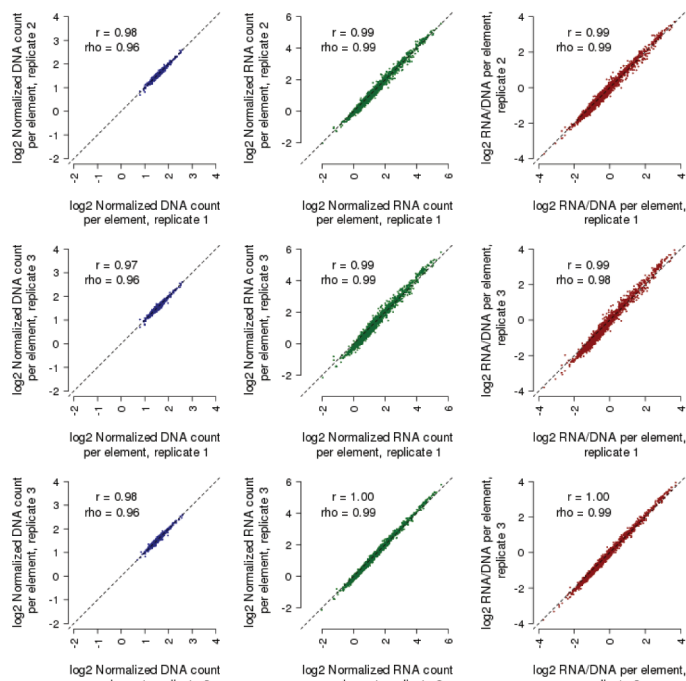
pGL4



HSS



ORI



5'/3' WT

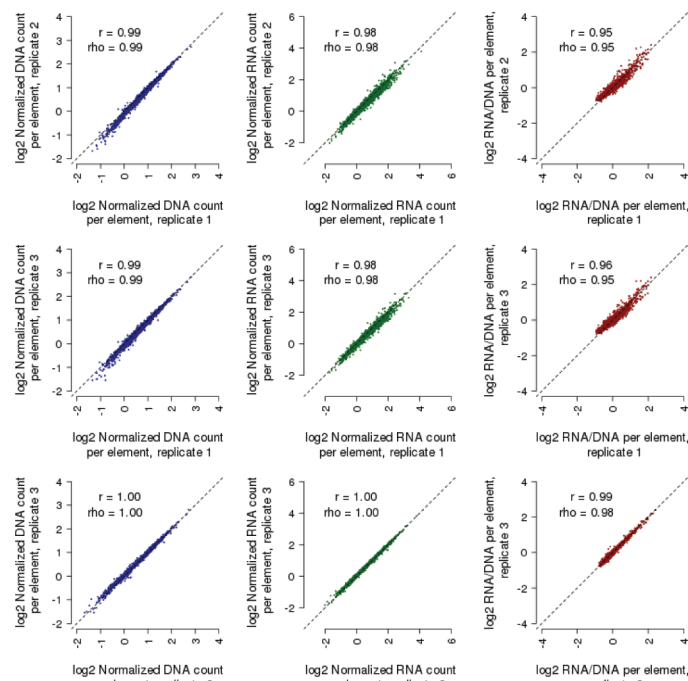
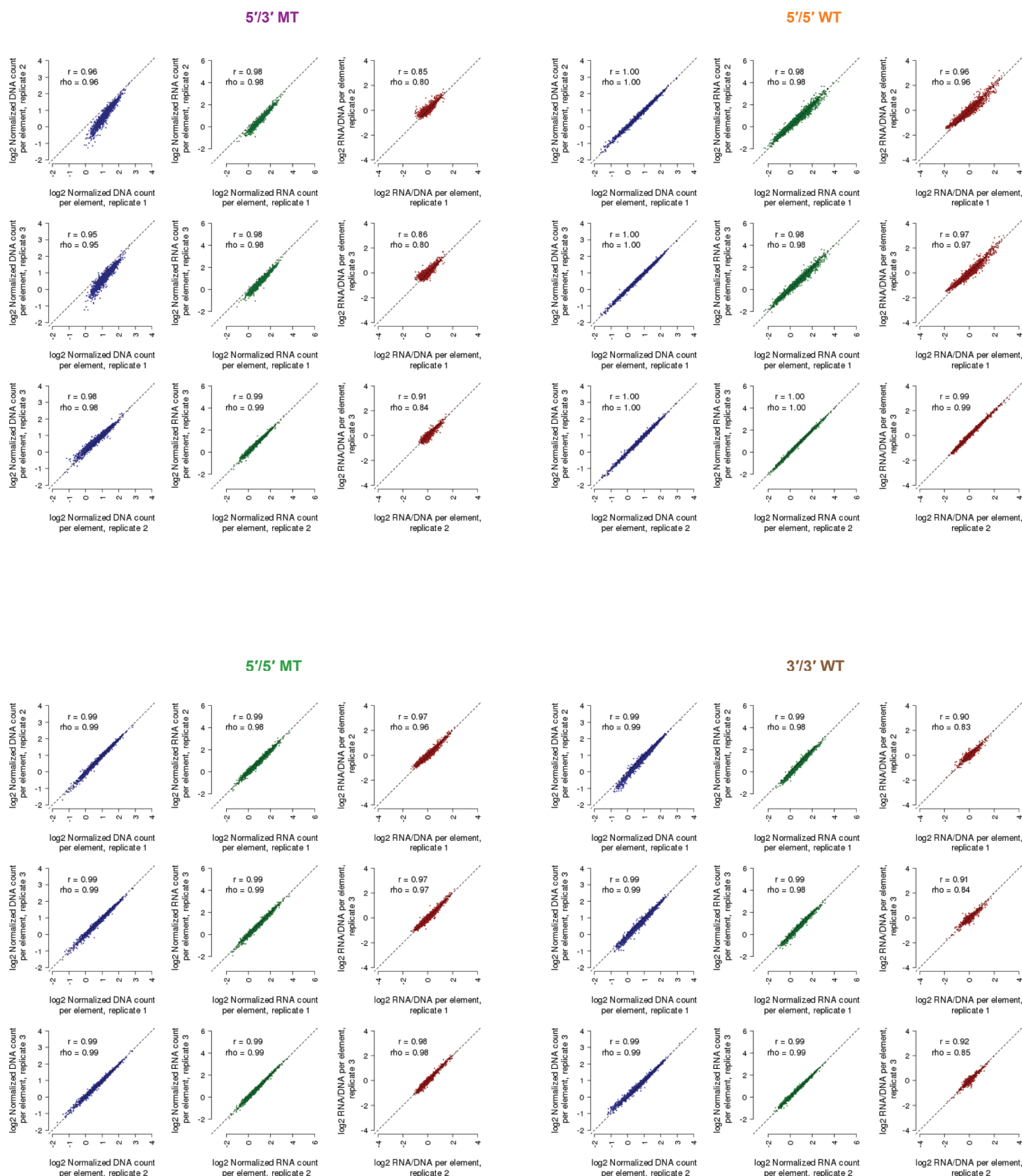
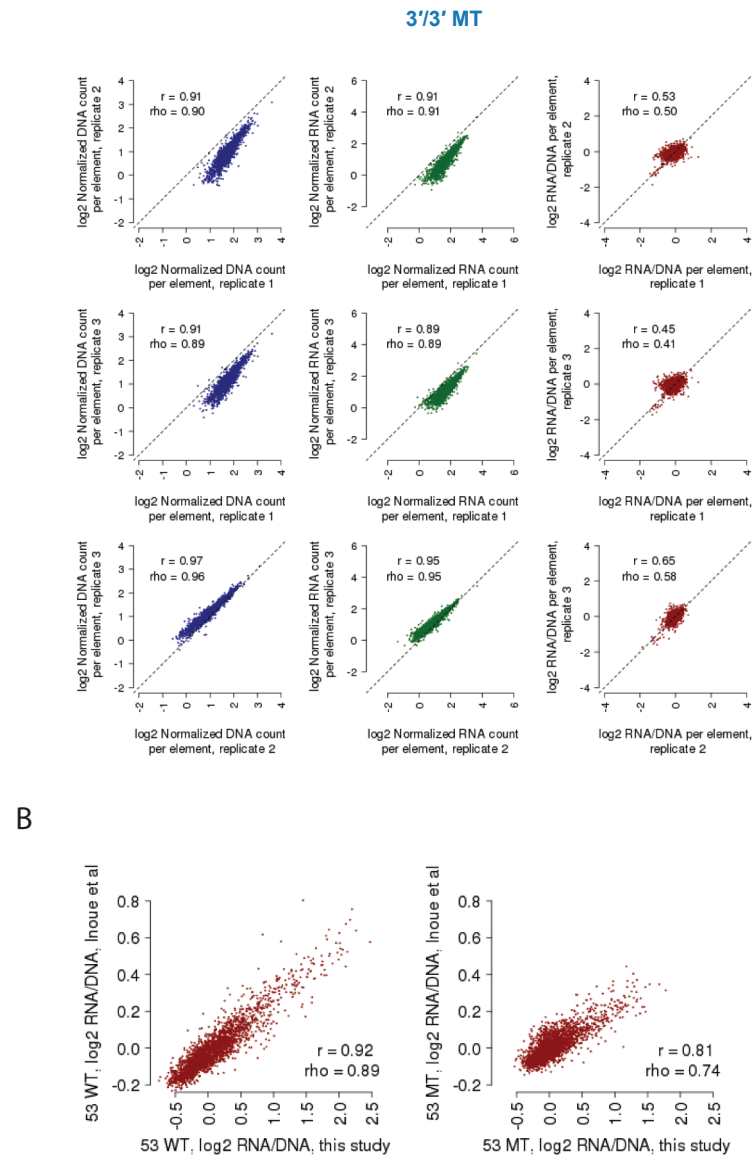


Figure S3



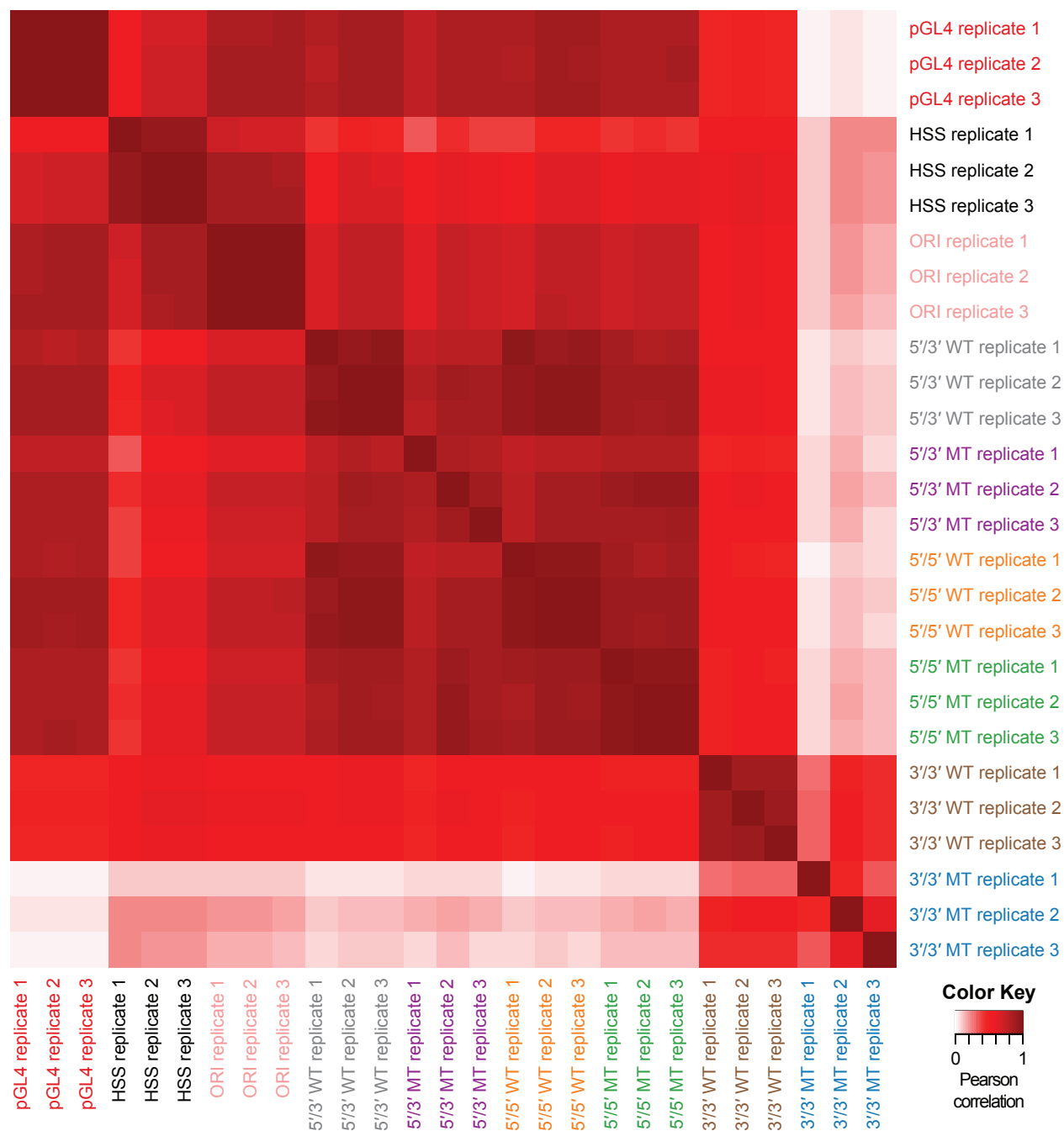


**Supplemental Figure 3. Reproducibility among replicates for each of the 9 MPRA methods.**

**A)** Shown are scatter plots displaying the relationship between observed DNA counts (blue), RNA counts (green), and RNA/DNA ratios (red) for all pairwise comparisons among replicates, for each of the 9 MPRA methods tested. Also indicated is the Pearson ( $r$ ) and Spearman ( $\rho$ ) correlation values. Candidate enhancers supported by fewer than 10 barcodes were filtered out prior to this analysis to reduce the impact of technical noise.

**B)** Scatter plots displaying the relationship between 5'/3' WT and 5'/3' MT ratios from this study and the indicated datasets from Inoue *et al.*<sup>18</sup>. Also indicated are the Pearson ( $r$ ) and Spearman ( $\rho$ ) correlation values.

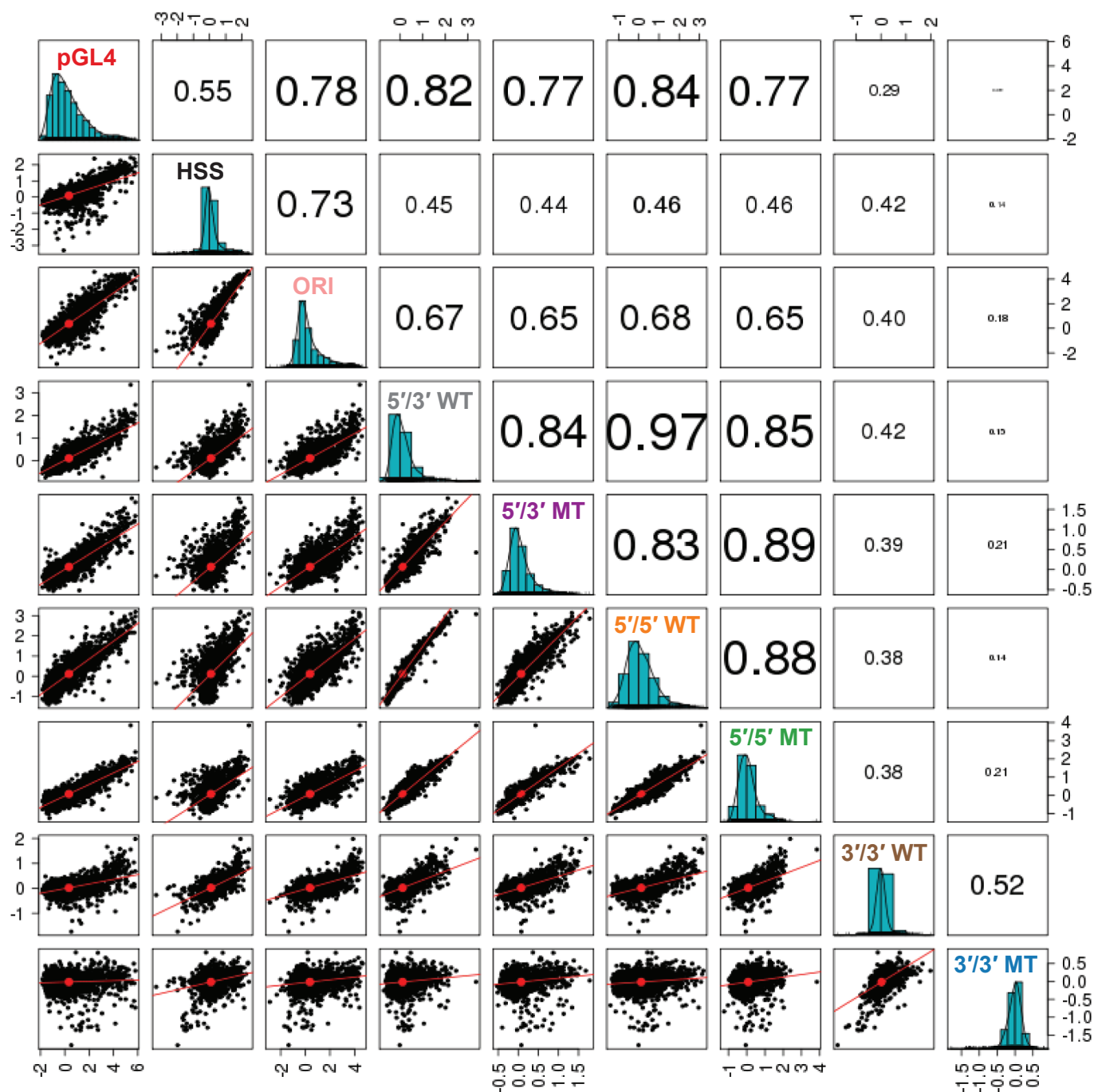
Figure S4



**Supplemental Figure 4. Reproducibility among RNA/DNA ratios among individual replicates for each of the 9 MPRA methods.**

Shown is a heatmap displaying the pairwise Pearson correlation among individual replicates, using RNA/DNA ratios from each of the 3 replicates for each of the 9 MPRA methods tested.

Figure S5

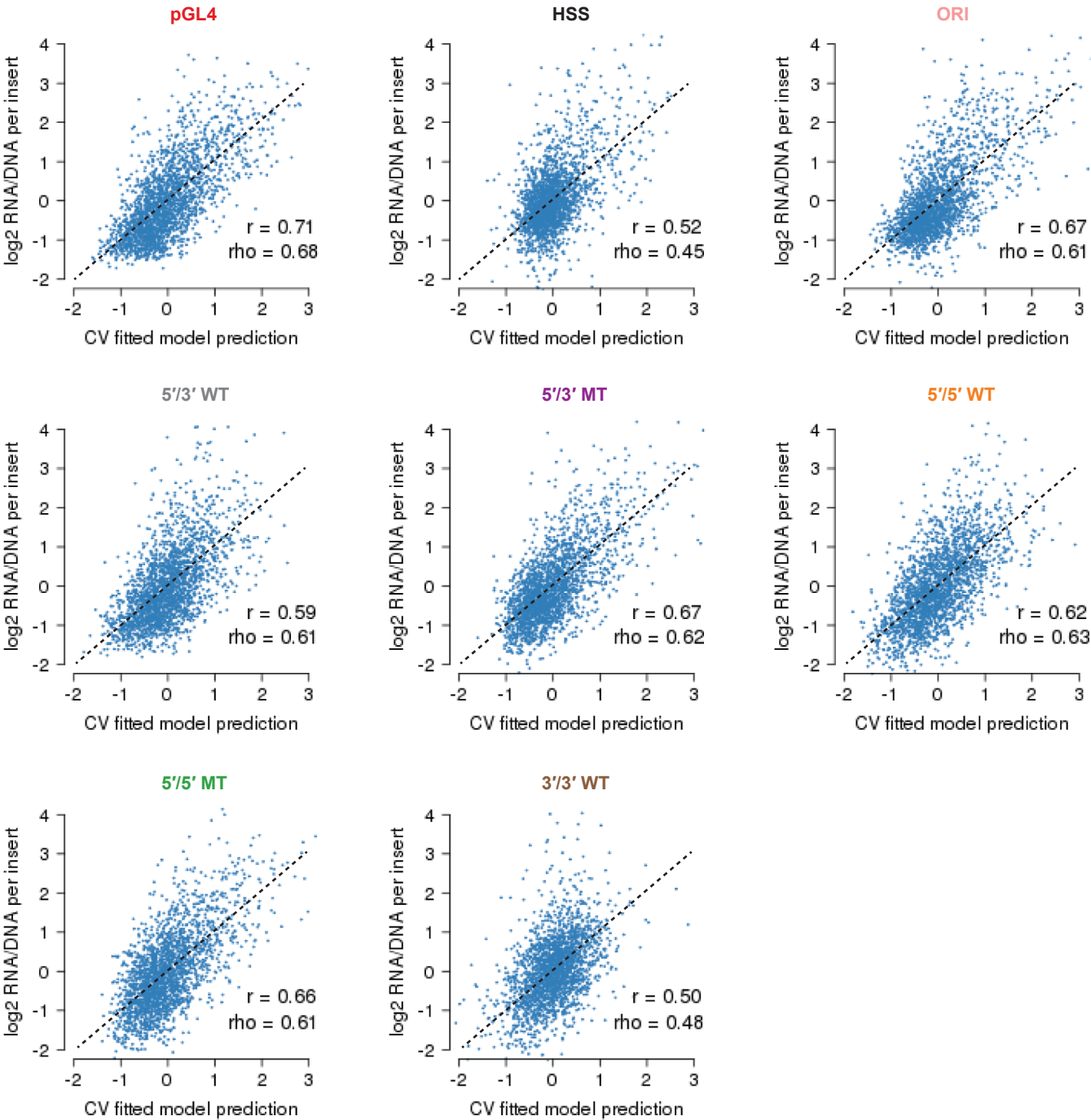


### Supplemental Figure 5. Reproducibility of averaged RNA/DNA ratios.

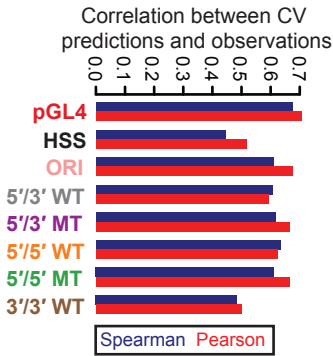
This figure is equivalent to **Figure 2B**, except it displays the scatter matrix values using Spearman correlation values instead of Pearson correlation values.

Figure S6

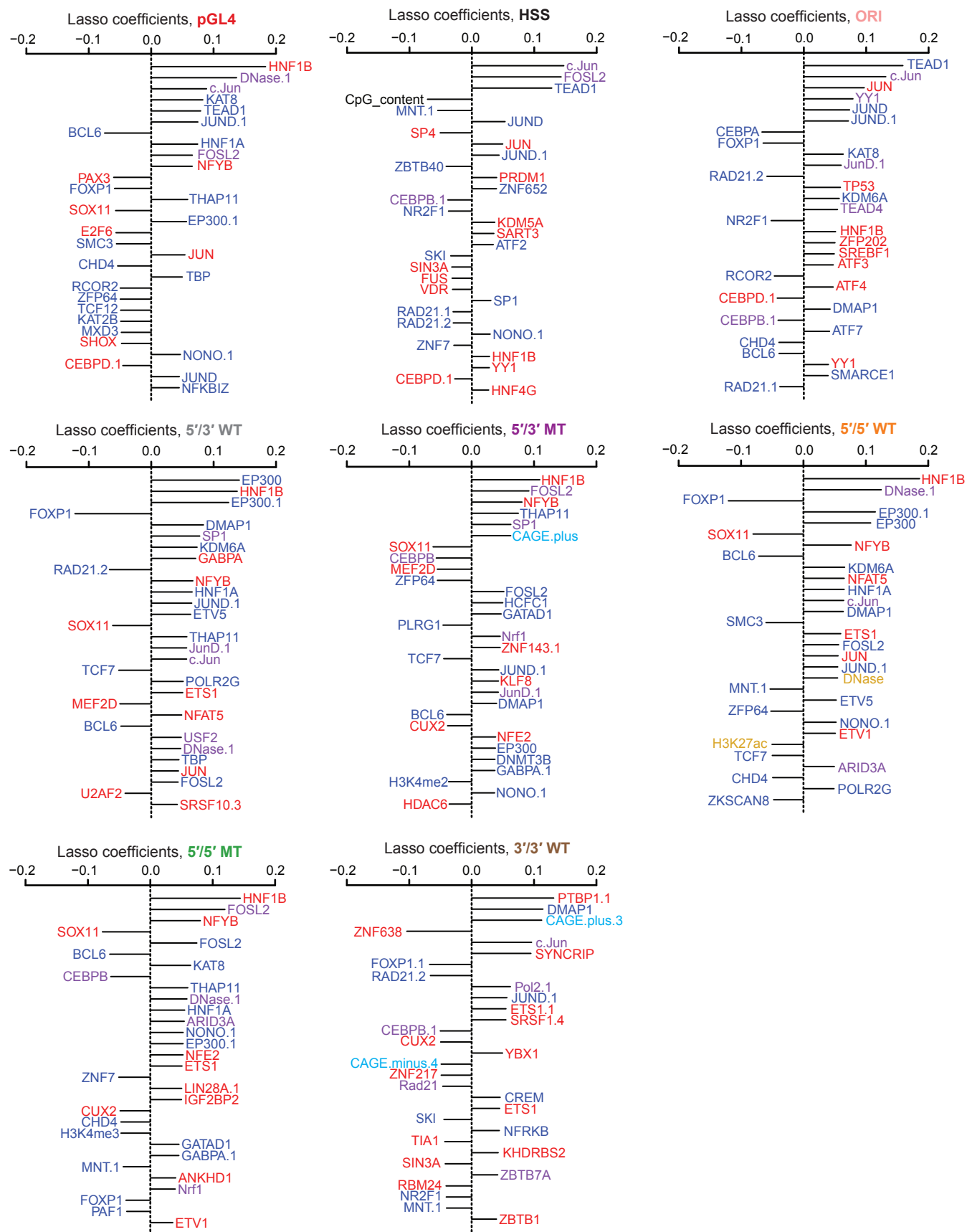
A



B



C







**Supplemental Figure 6. Coefficients and performance of the models to predict element efficacy among the MPRA methods.**

**A)** Scatter plots displaying the relationship between the 10-fold cross-validated predictions derived from lasso regression models and the observed RNA/DNA ratios, for each of 8 MPRA methods tested (excluding 3'/3' MT due to poor inter-replicate reproducibility). Also indicated are the Pearson ( $r$ ) and Spearman ( $\rho$ ) correlation values.

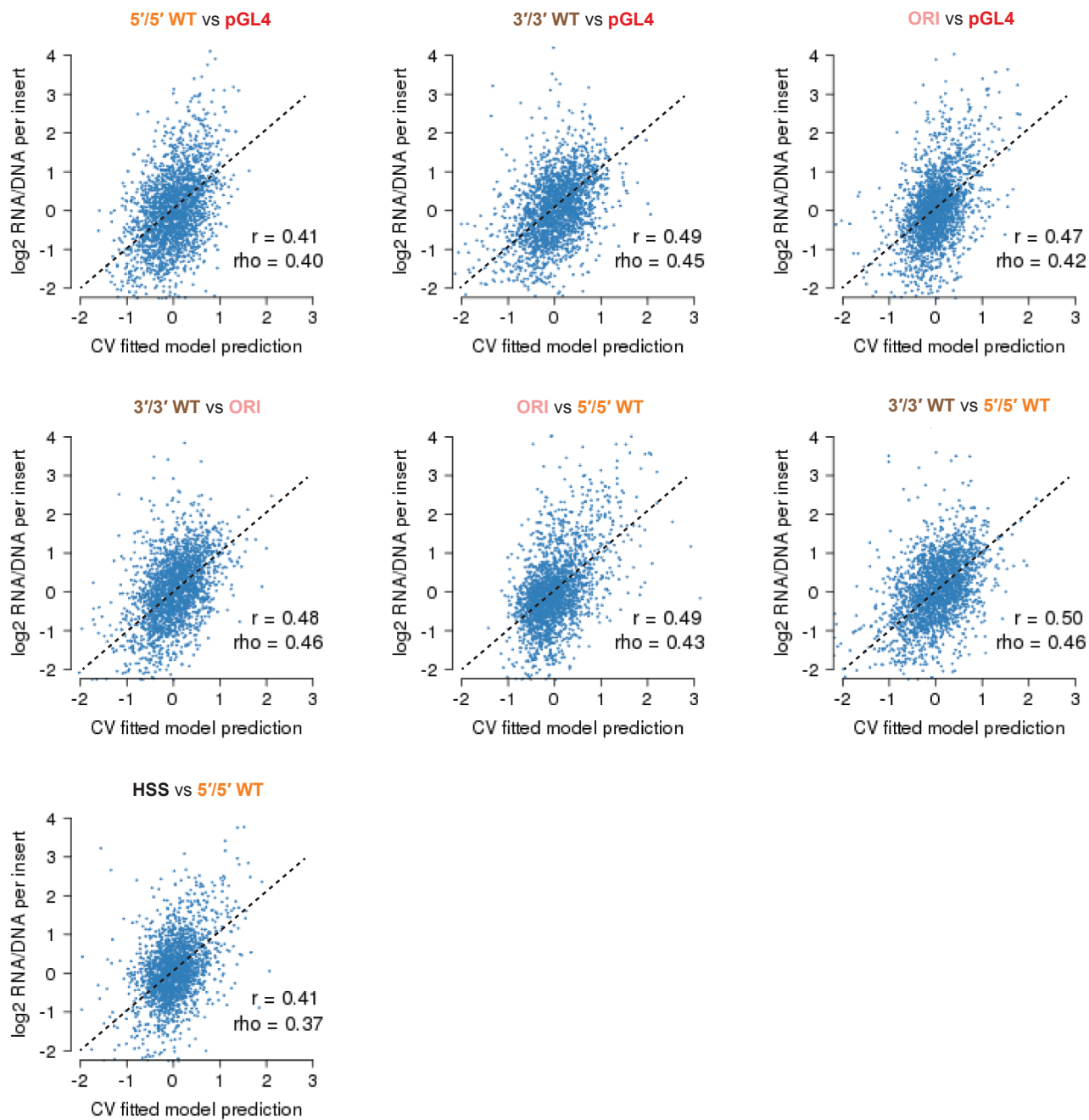
**B)** Pearson and Spearman correlation coefficients for 10-fold cross-validated predictions derived from lasso regression models and the observed RNA/DNA ratios, for each of the 8 MPRA methods tested.

**C)** The top 30 coefficients derived from a lasso regression model trained on the full dataset derived from the eight indicated MPRA methods. Features with the extension ".1", ".2", etc allude to redundant features or replicate samples.

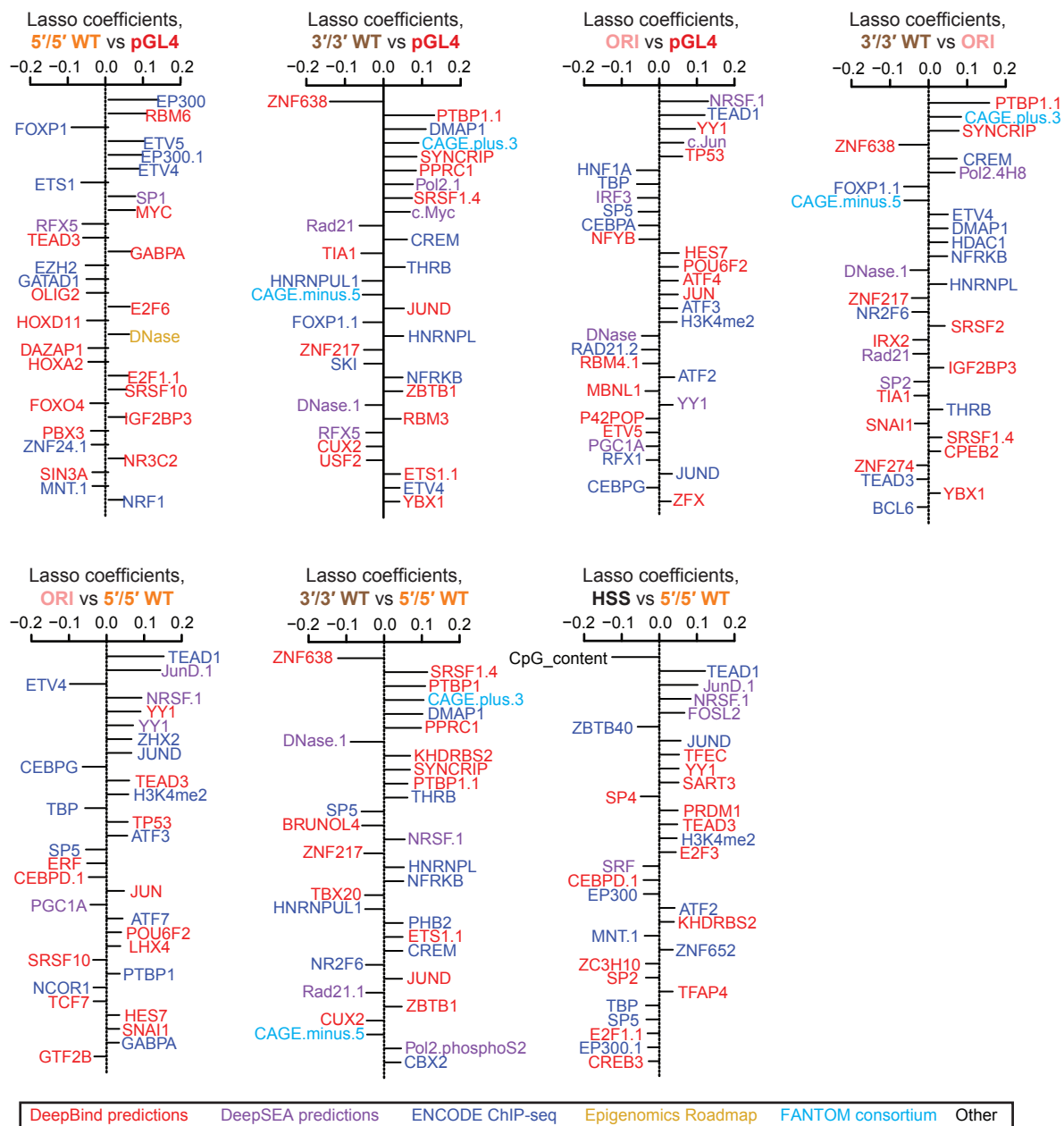
**D)** Figure has been rotated to fit to sheet. Pearson correlation matrix between the union of all top 30 features from (C), shown as rows, and other features sharing a Pearson correlation either  $\leq -0.8$  or  $\geq 0.8$ , shown as columns. Feature names are colored according to the origin of the feature as shown in the boxed key above. Hierarchical clustering was used to group features exhibiting similar correlation patterns.

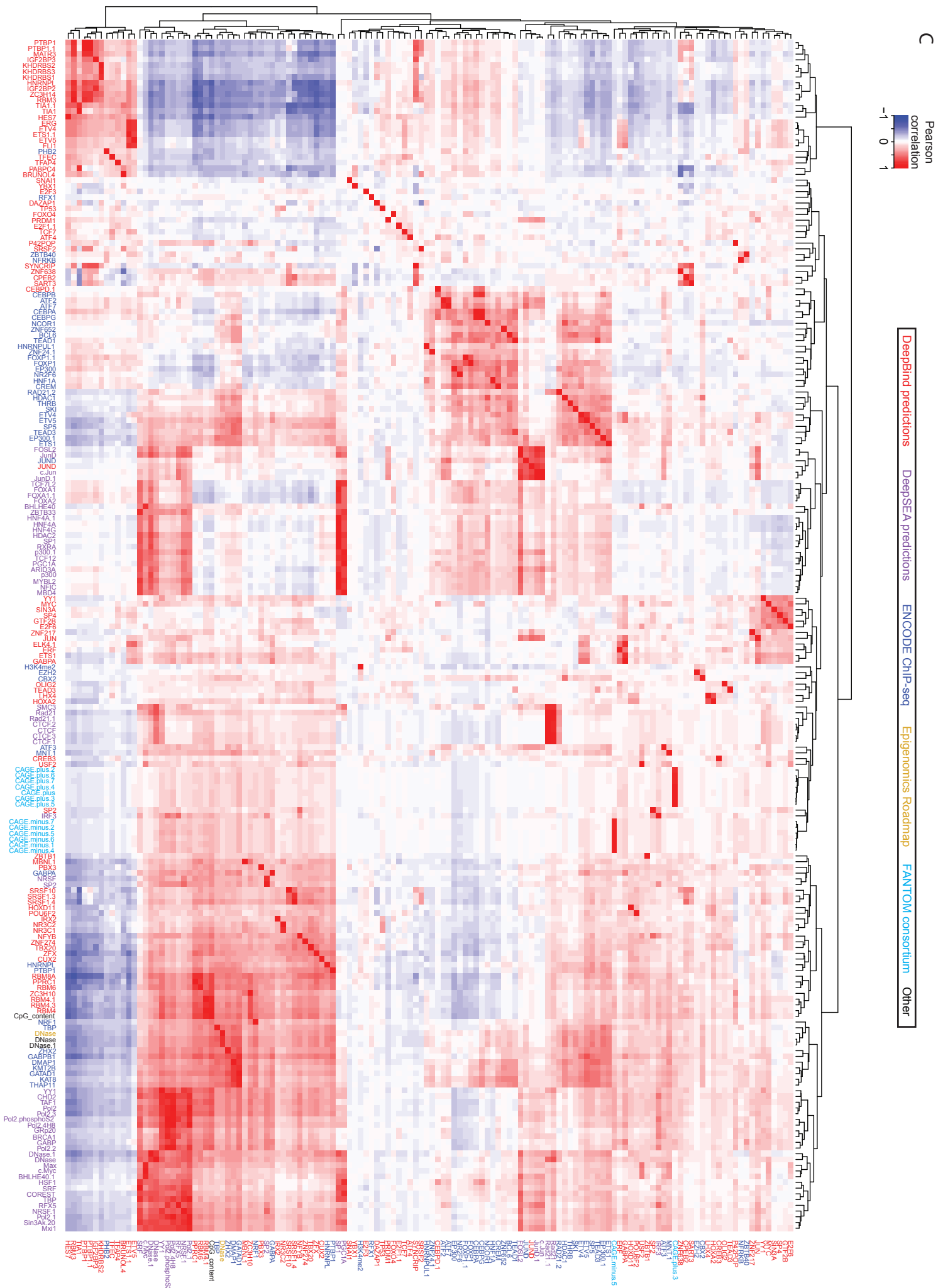
Figure S7

A



B



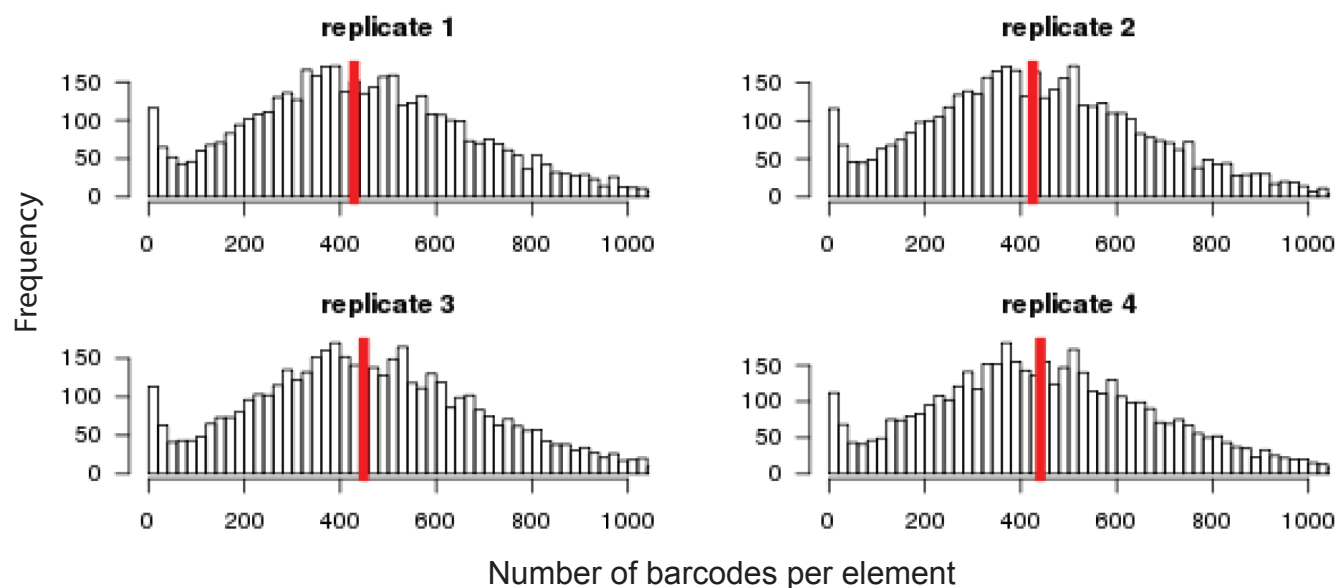


**Supplemental Figure 7. Coefficients and performance of the models to predict differential element efficacy among the MPRA methods.**

**A)** Scatter plots displaying the relationship between the 10-fold cross-validated predictions derived from lasso regression models and the observed RNA/DNA ratios, for each of the 7 indicated differential comparisons tested. Also indicated are the Pearson ( $r$ ) and Spearman ( $\rho$ ) correlation values.

**B)** The top 30 coefficients derived from lasso regression models trained on the full dataset to predict observed differences in the indicated pairs of MPRA methods. Features with the extension “.1”, “.2”, etc. allude to redundant features or replicate samples.

**C)** Figure has been rotated to fit to sheet. Pearson correlation matrix between the union of all top 30 features from (B), shown as rows, and other features sharing a Pearson correlation either  $\leq -0.8$  or  $\geq 0.8$ , shown as columns. Feature names are colored according to the origin of the feature as shown in the boxed key above. Hierarchical clustering was used to group features exhibiting similar correlation patterns.



**Supplemental Figure 8. Histograms of barcodes per element for each replicate of the the library testing orientation.**

Shown are histograms indicating the number of observed barcodes per element, for each of the 2,336 elements tested in each of the 4 replicates for the library testing orientation (i.e., forward and reverse). Shown with a vertical red line is the median number of barcodes per element.

Figure S9

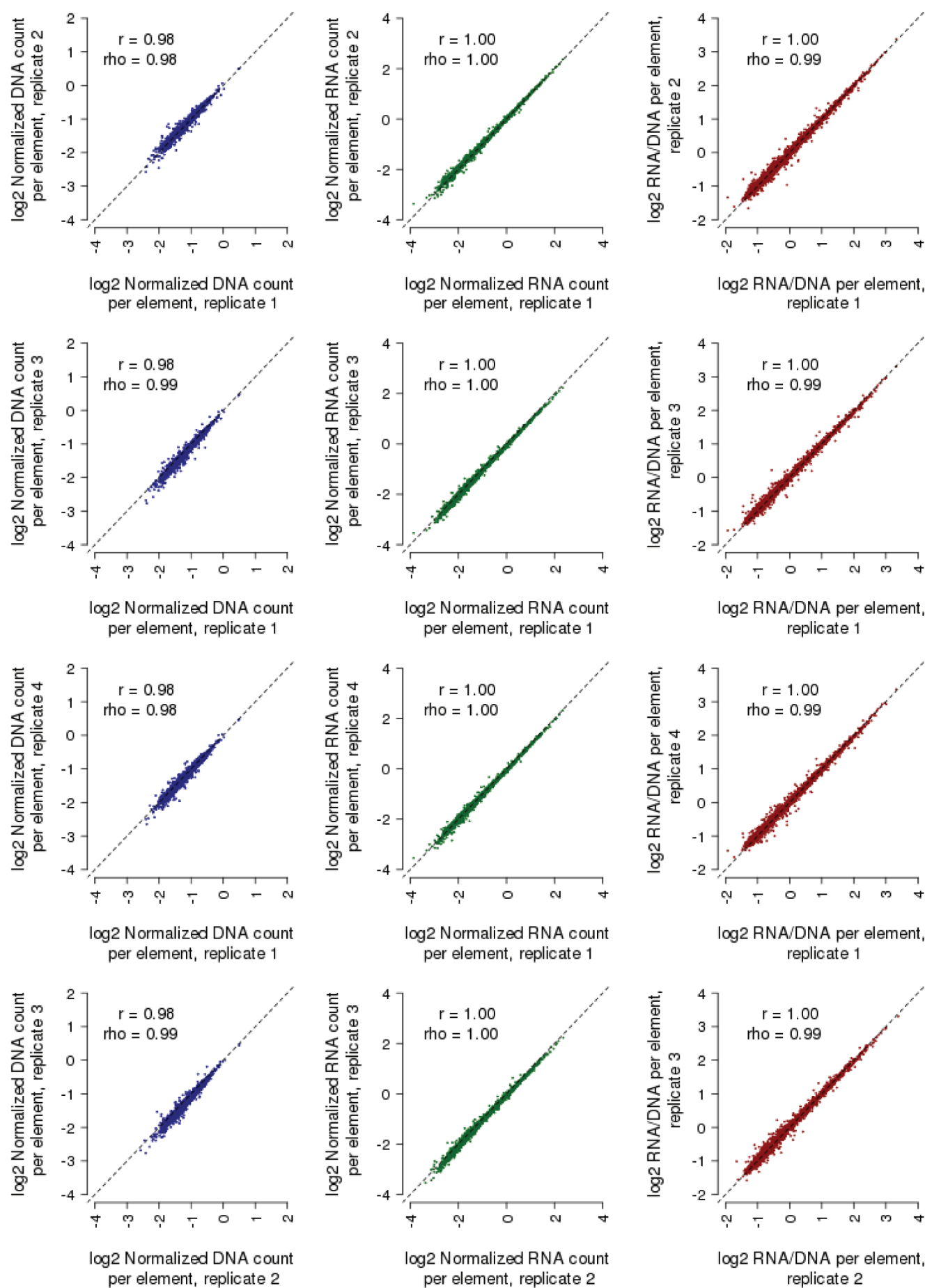
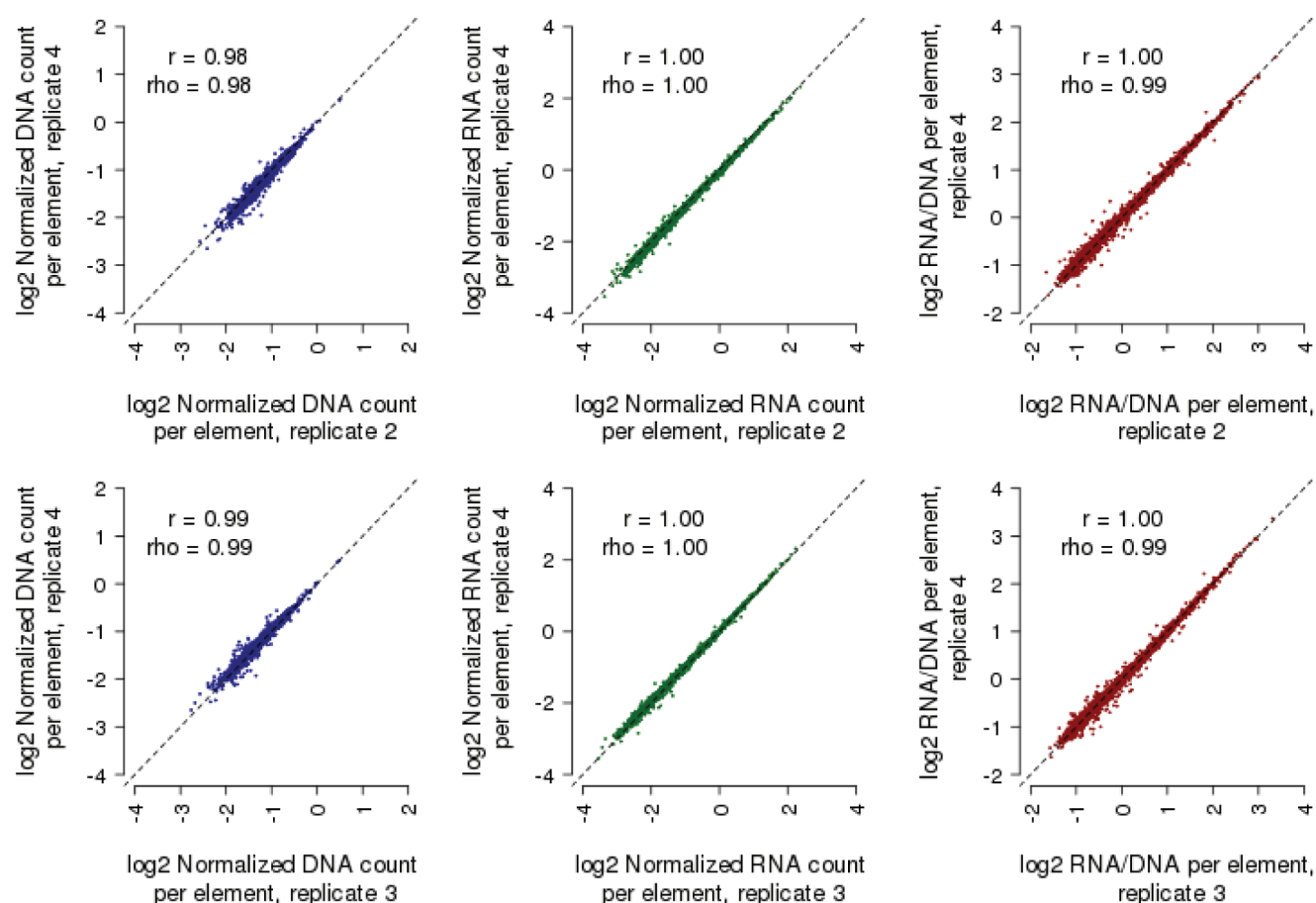


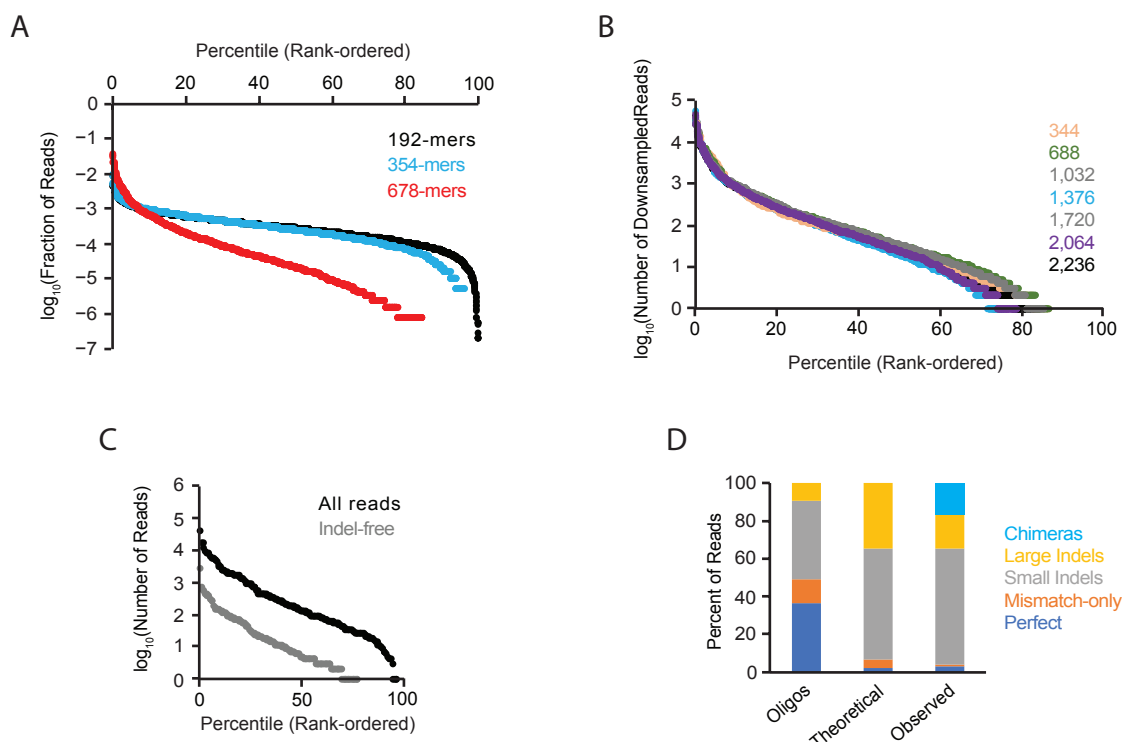


Figure S9



### Supplemental Figure 9. Reproducibility among replicates for the library testing orientation.

Shown are scatter plots displaying the relationship between observed DNA counts (blue), RNA counts (green), and RNA/DNA ratios (red) for all pairwise comparisons among the four replicates, including data from both orientations (i.e., forward and reverse) tested. Also indicated are the Pearson ( $r$ ) and Spearman ( $\rho$ ) correlation values. Candidate enhancers supported by fewer than 10 barcodes were filtered out prior to this analysis to reduce the impact of technical noise.



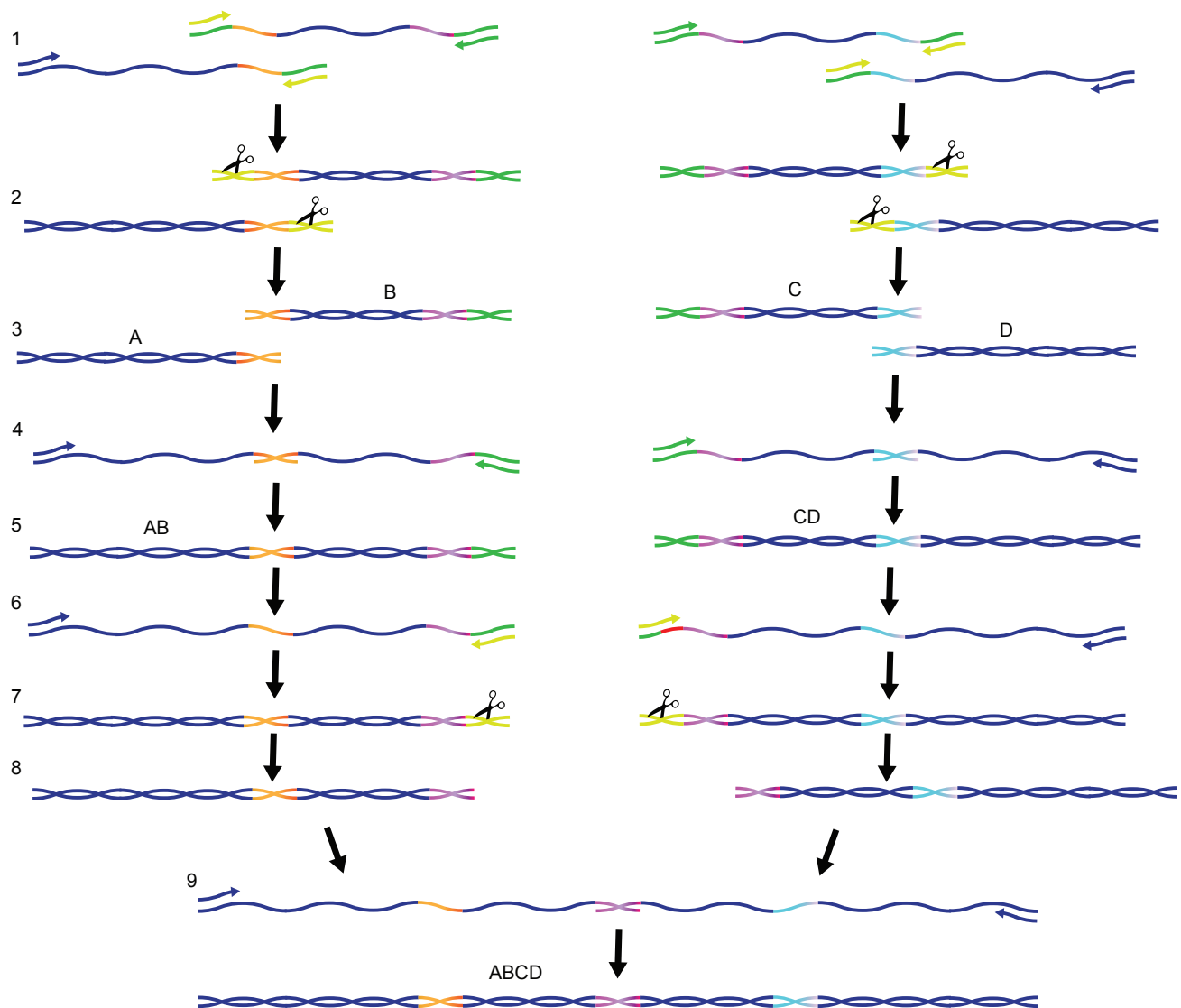
**Supplemental Figure 10. Evaluation of Multiplex Pairwise Assembly (MPA) and Hierarchical Multiplex Pairwise Assembly (HMPA) library quality.**

**A)** Plot of the uniformity of indel-free reads for 2,336 x 192mers (amplified off Agilent 230mer array), 2,336 x 354mers (after Multiplex Pairwise Assembly), and 2,236 x 678mers (after a single Hierarchical Multiplex Pairwise Assembly). The x-axis is rank-ordered according to the most to least abundant from left to right. The y-axis is the fraction of either indel-free reads (for 192mers and 354mers) or all reads (for 678mers).

**B)** Uniformity for various HMPA reactions, with total number of target sequences ranging from 344 to 2,236. The sequencing reads were down-sampled to normalize for the total number of targets ( $\# \text{ reads} = 581 * \# \text{ targets}$ ). The Y-axis is the number of down-sampled reads from a given target sequence.

**C)** One sub-library (of 172 targets), was sequenced on a PacBio Sequel. The plot shows the uniformity for all aligning reads (black) and indel-free reads (grey).

**D)** Composition of the sub-library of 172 targets. The first column shows the breakdown of oligos by Illumina sequencing. The second column shows the theoretical breakdown of 678mers based on each target consisting of four independent oligos. The third column shows the breakdown of 678mers based on PacBio sequencing.



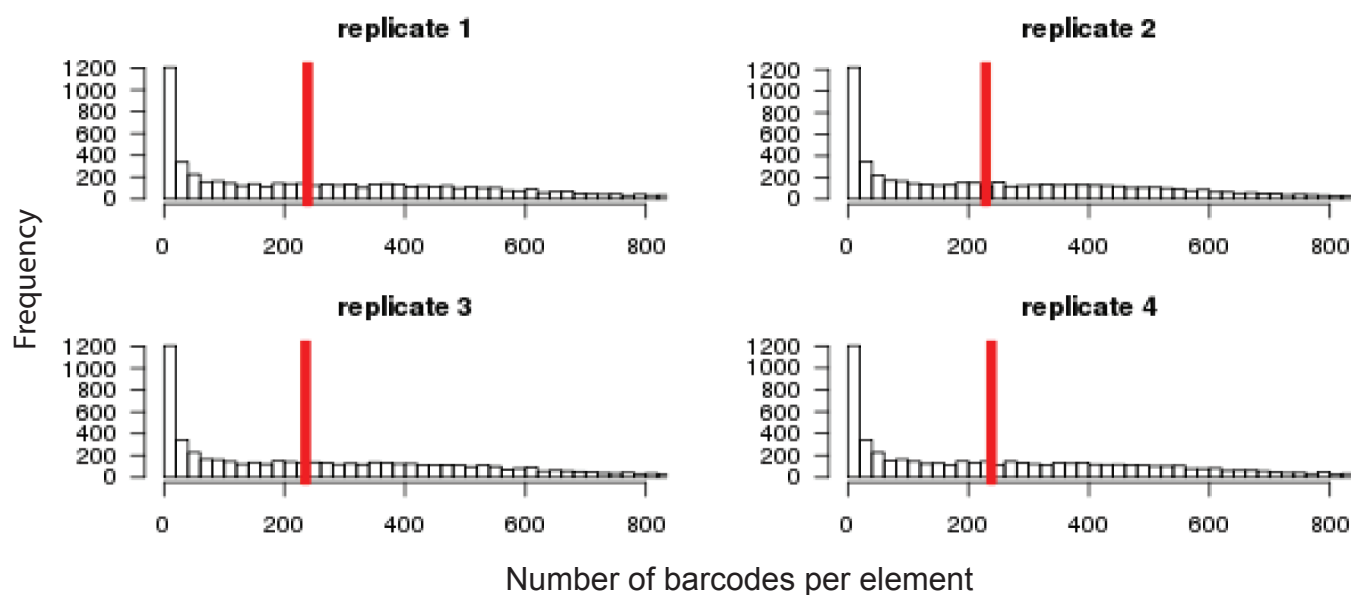
**Supplemental Figure 11. Hierarchical Multiplex Pairwise Assembly (HMPA) strategy.**

**1.** To generate a library of 678 bp enhancers, we ordered each enhancer as four oligonucleotides to be assembled (fragments “A”, “B”, “C”, and “D”). To assemble fragments “AB” and “CD”, sequences were designed such that the 3’ end of fragments A and C had 30 bps of homology to the 5’ ends of fragments B and D, respectively (shown in red and orange). To remove adapter sequences from these ends of the fragments, uracil primers (shown in yellow) were used to incorporate uracils into the adapters during qPCR amplification.

**2.** The resulting fragments were treated with USER enzyme (scissors) and put into an end-repair reaction, **3.** effectively removing the adapters.

**4.** Fragments were assembled in a qPCR reaction by allowing the fragments to first anneal to one another for 5 cycles of PCR without primers, and then adding primers targeting the 5’ end of fragments A and C, and the 3’ ends of fragments B and D, **5.** resulting in fragments “AB” and “CD”.

**6-9.** Adapter sequences were removed from the 3’ end of AB fragments and the 5’ end of CD fragments, and the final 678 bp “ABCD” enhancer sequences were assembled using the aforementioned assembly reaction.

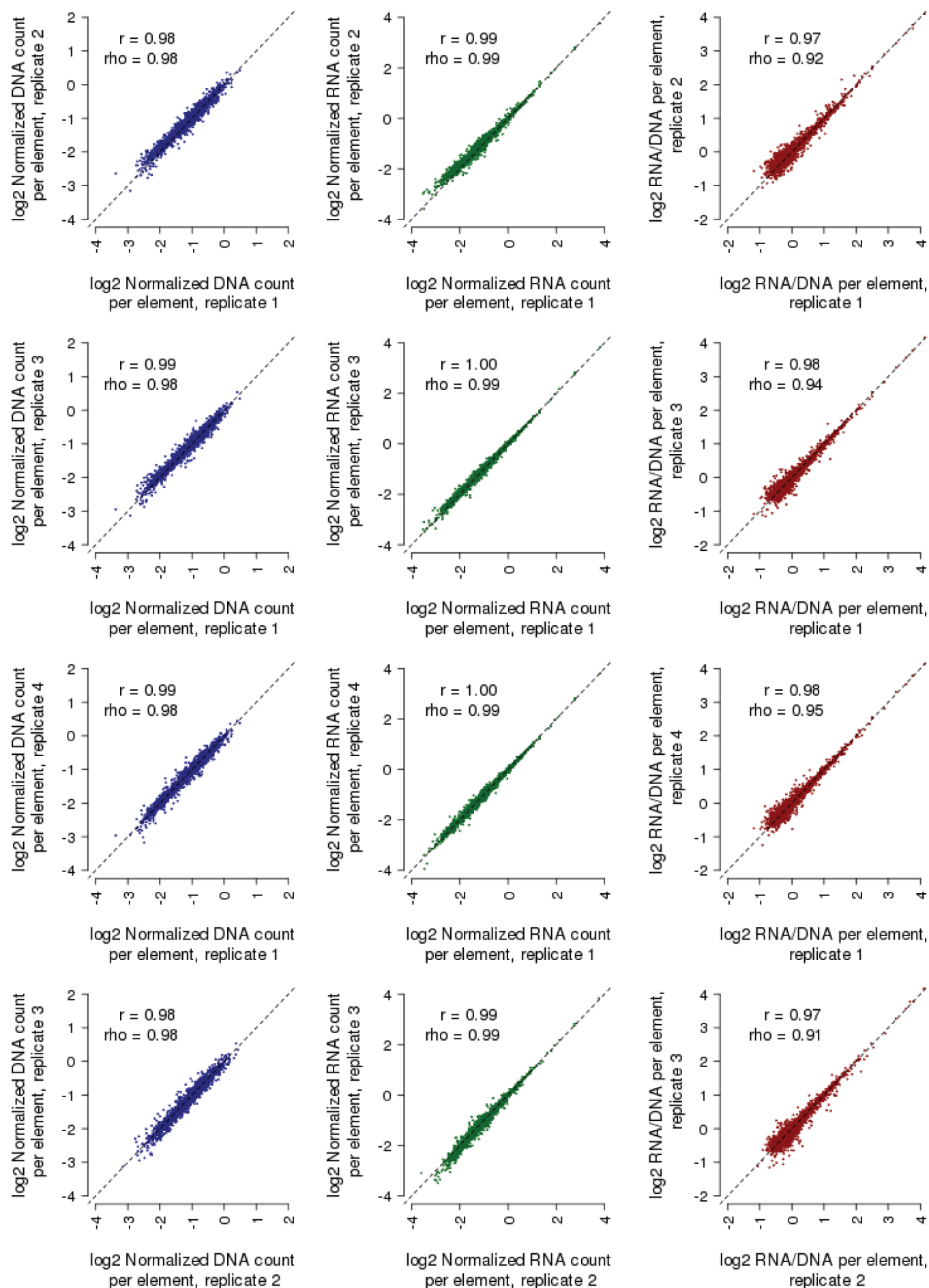


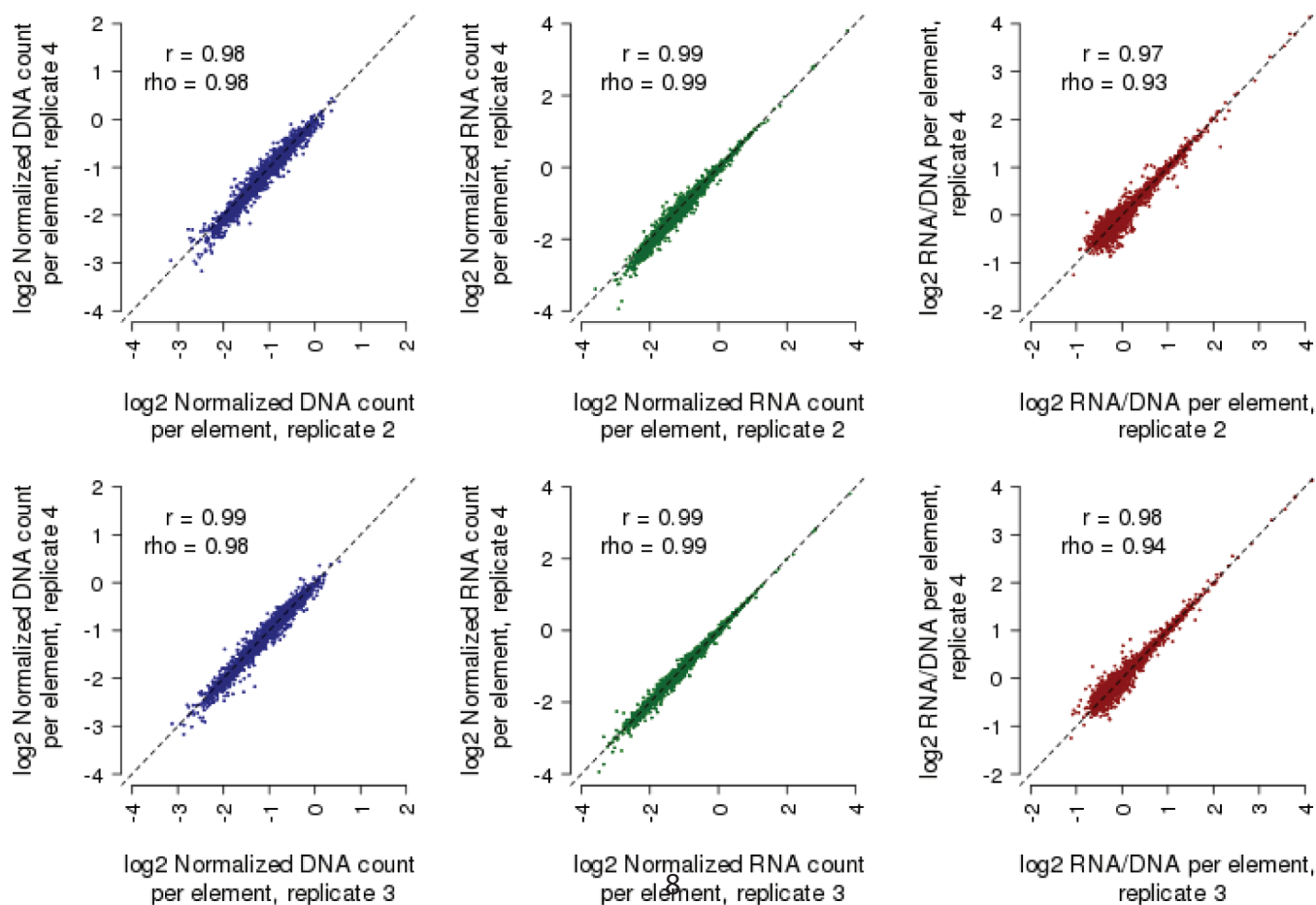
**Supplemental Figure 12. Histograms of barcodes per element for each replicate of the the library testing element size.**

Shown are histograms indicating the number of observed barcodes per element, for each of the 2,336 elements tested in each of the 4 replicates for the library testing size classes (i.e., short, medium, and long). Shown with a vertical red line is the median number of barcodes per element.

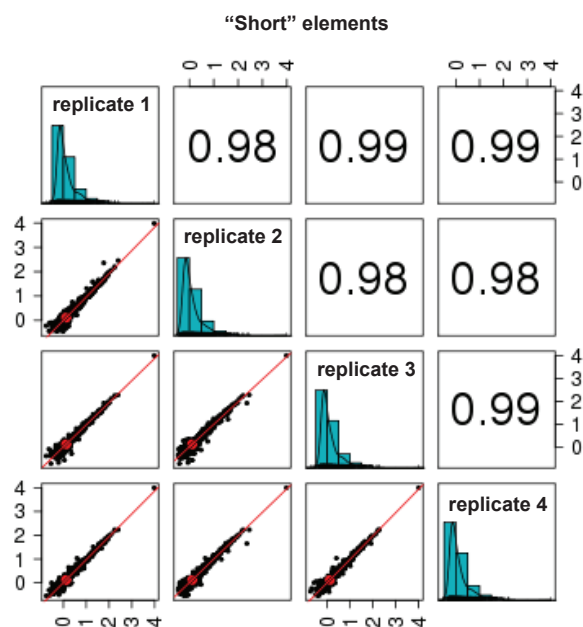
Figure S13

A

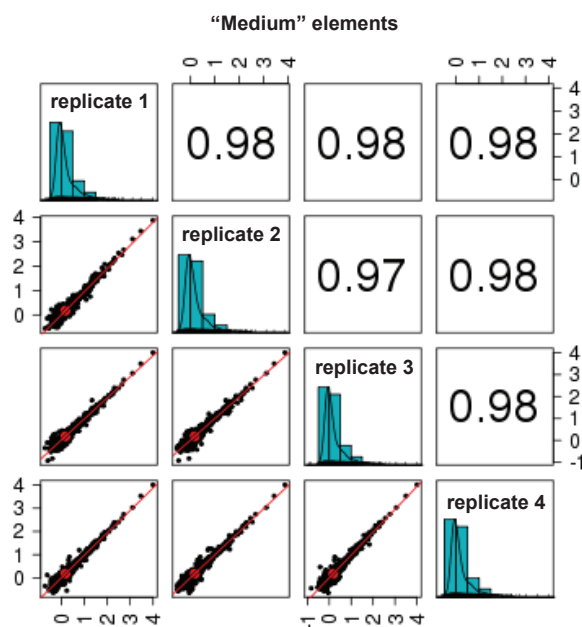




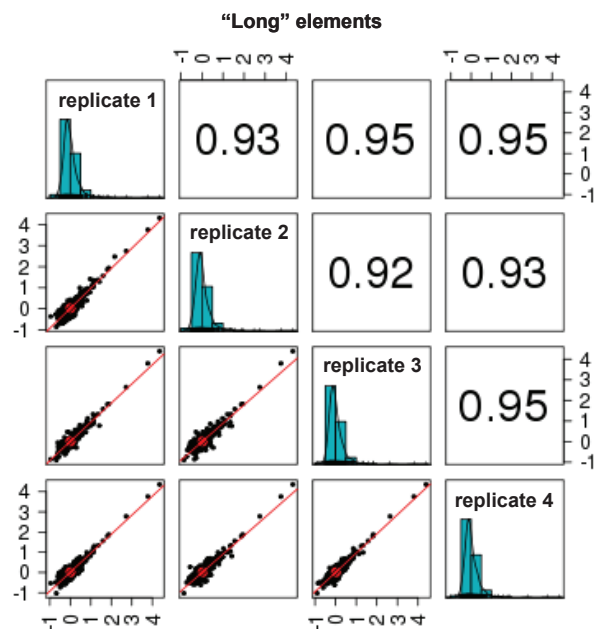
B



C



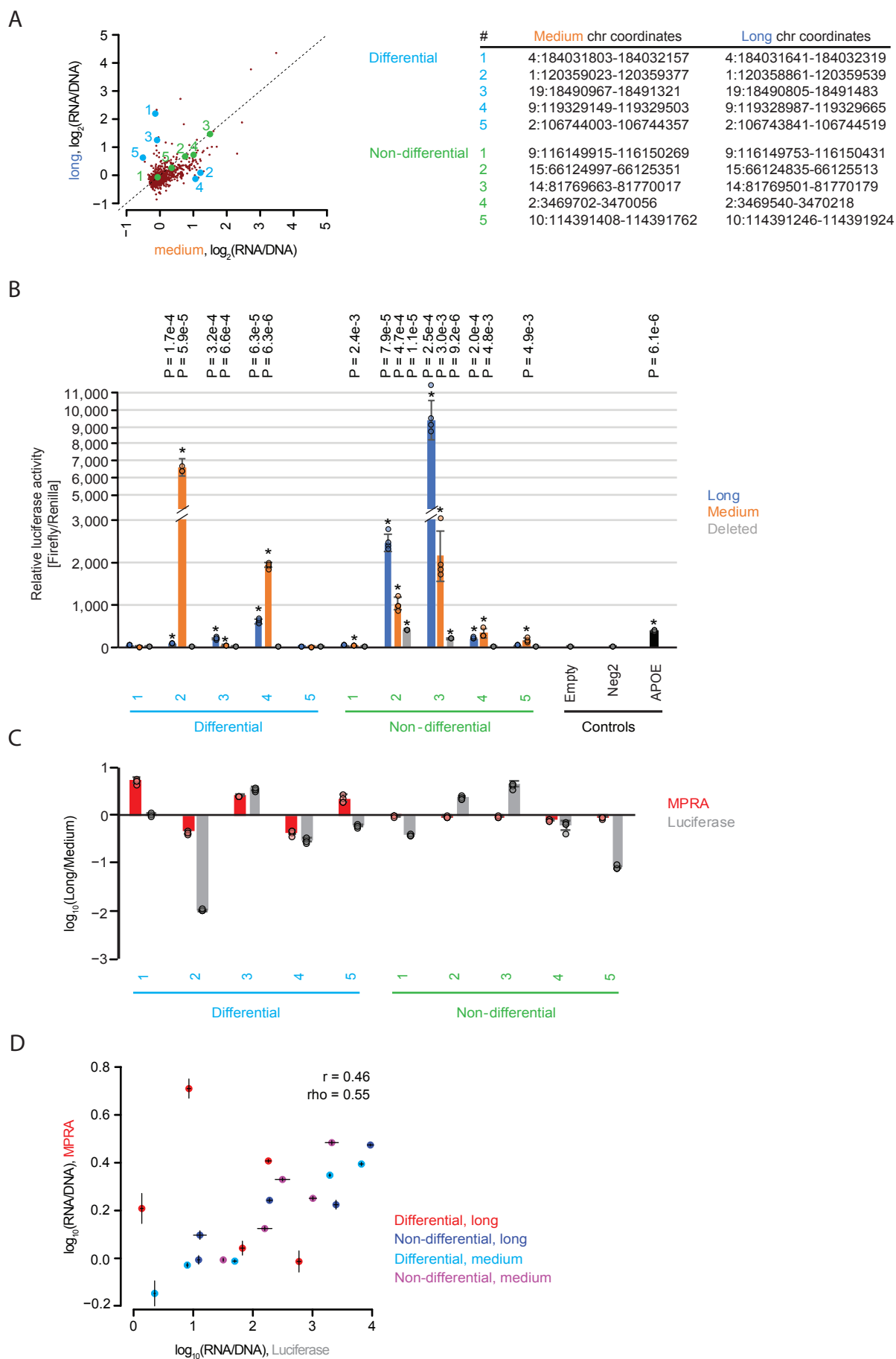
D



**Supplemental Figure 13. Reproducibility among replicates for testing of libraries of the same enhancers with different amounts of sequence context.**

**A)** Shown are scatter plots displaying the relationship between observed DNA counts (blue), RNA counts (green), and RNA/DNA ratios (red) for all pairwise comparisons among the four replicates, including data from all size classes (i.e., short, medium, and long) tested. Also indicated are the Pearson ( $r$ ) and Spearman ( $\rho$ ) correlation values. Candidate enhancers supported by fewer than 10 barcodes were filtered out prior to this analysis to reduce the impact of technical noise.

**B-D)** Pearson correlation between pairs of replicates for the subset of elements tested in the "short" (B), "medium" (C), or "long" (D) size class.





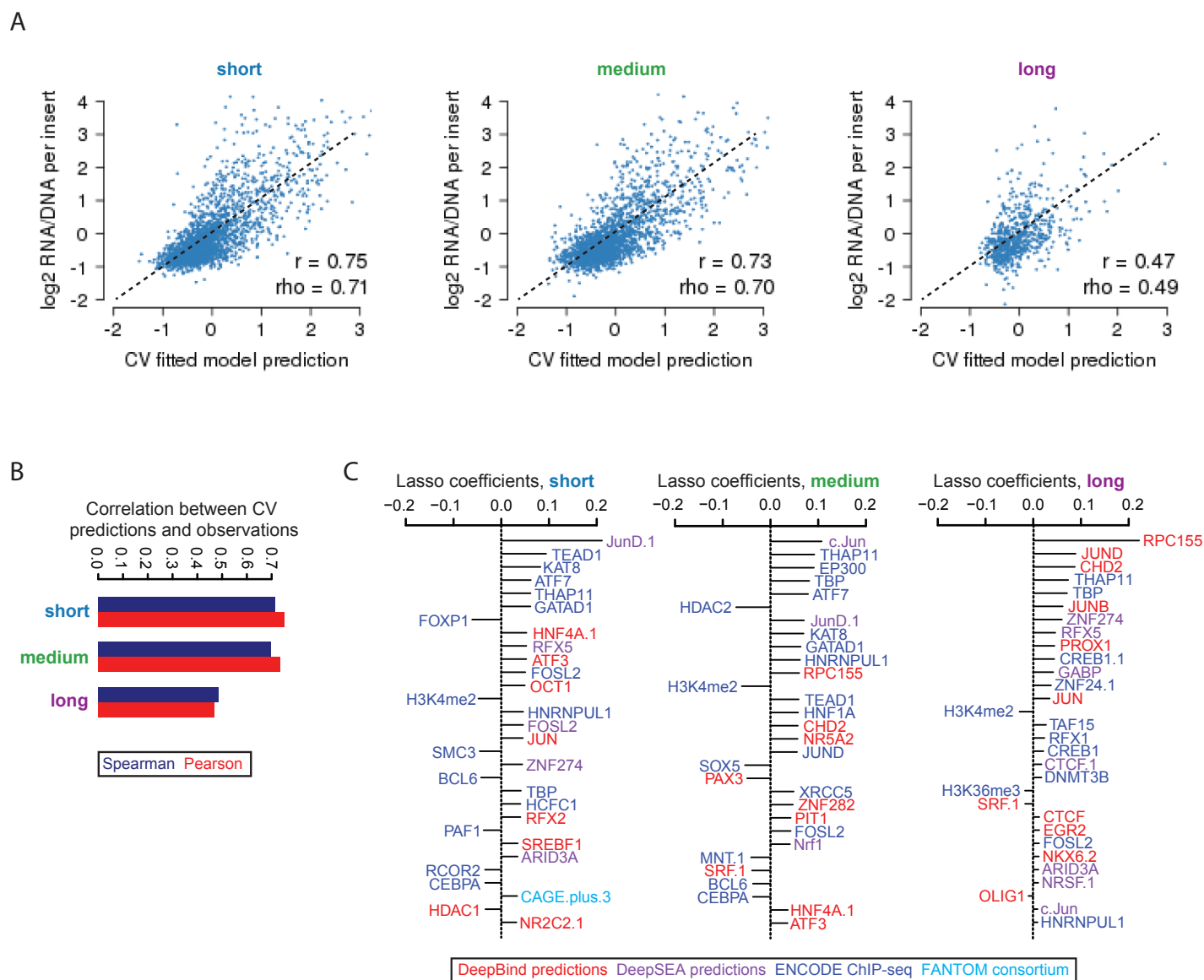
**Supplemental Figure 14. Experimental evaluation of differential size activity with luciferase assays.**

**A)** Scatter plots of the average activity score of each element, comparing medium vs. long versions of each element (as shown in **Figure 5C**). Five colored elements were chosen that exhibited non-differential (green) or differential (cyan) activity between the two size classes. The chromosomal coordinates are shown (human genome build “hg19”) for each element.

**B)** The five non-differential and differential elements were cloned into the pGL4.23 vector and measured with a luciferase assay. A “deleted” version, in which the “long” sequence was tested after deleting its internal “medium” sequence, was also tested, along with an empty vector, negative control (Neg2), and positive control (APOE).  $n = 4$  biologically independent cell cultures. Error bars indicate the standard error around the mean for four replicates, with significance assessed using a one-sided t-test comparing against Neg2 activity ( $*p < 0.01$ ).

**C)** Summary of differential activity between long and medium sizes, as assessed by both the MPRA data shown in (A) and Luciferase data shown in (B).  $n = 4$  biologically independent cell cultures. Error bars indicate the standard error around the mean for four replicates.

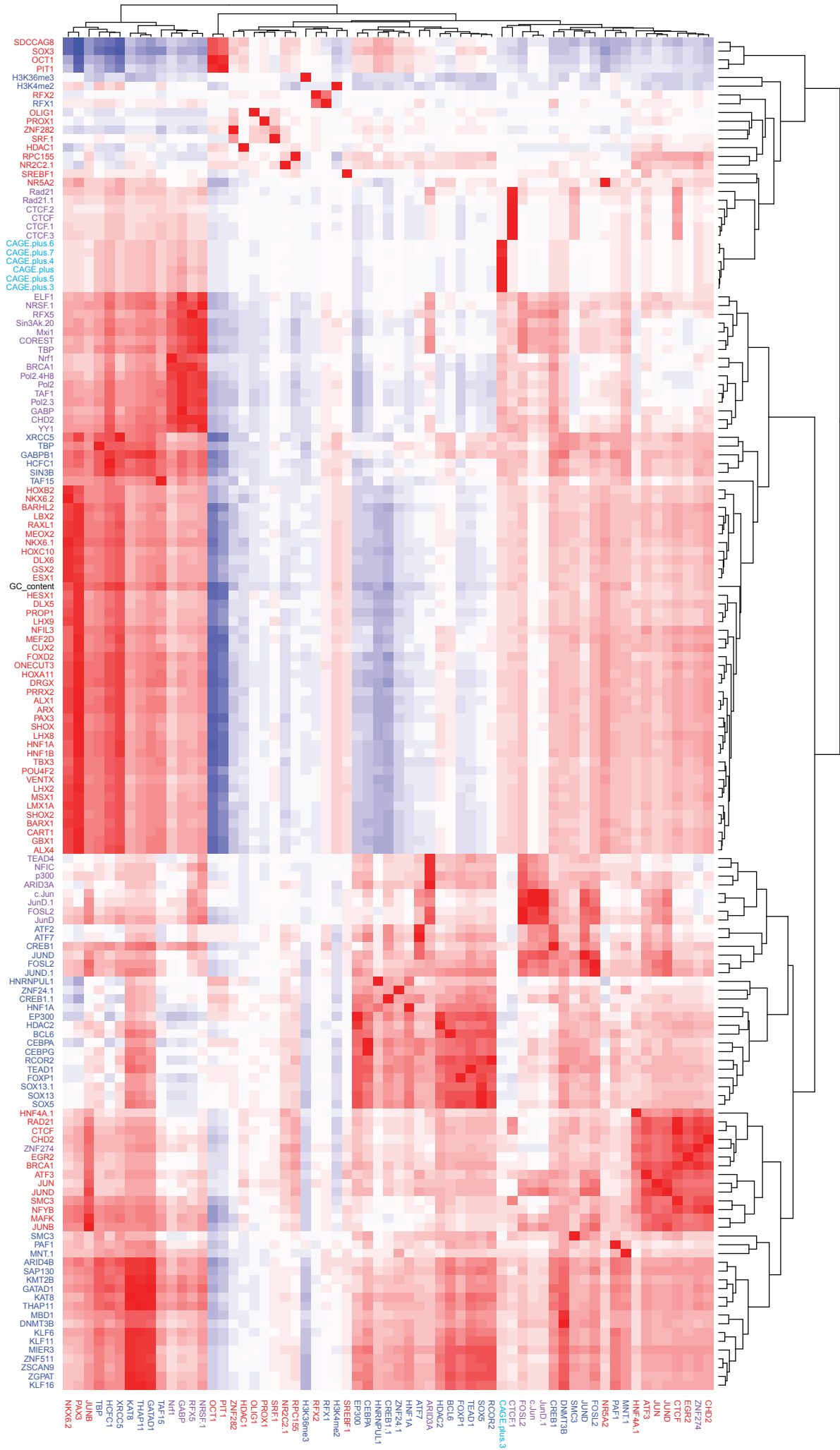
**D)** Scatter plot of the relationship between Luciferase and MPRA measurements. Error bars indicate the standard deviation among four independent replicates.



D

Pearson  
correlation  
-1 0 1

DeepBind predictions    DeepSEA predictions    ENCODE ChIP-seq    Epigenomics Roadmap    FANTOM consortium    Other



**Supplemental Figure 15. Coefficients and performance of the models to predict element efficacy among the three size classes.**

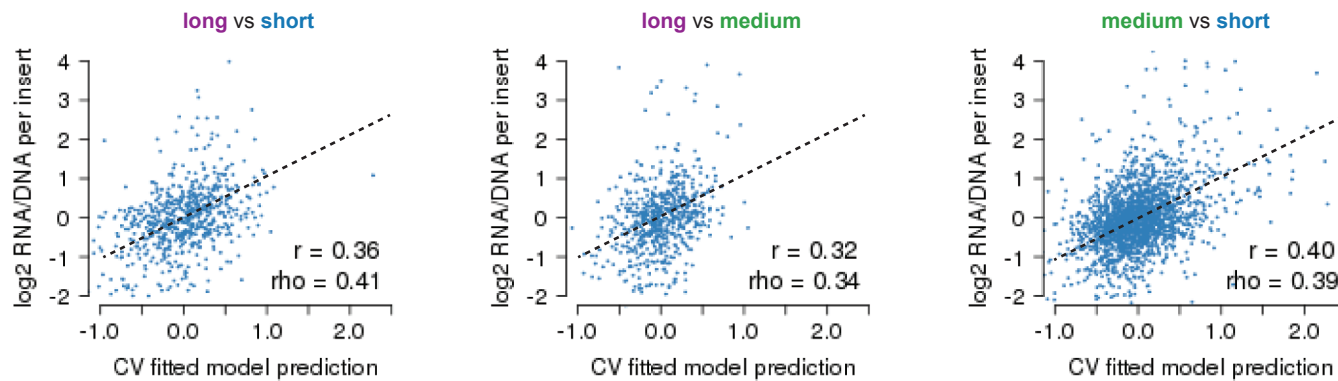
**A)** Scatter plots displaying the relationship between the 10-fold cross-validated predictions derived from lasso regression models and the observed RNA/DNA ratios, for each of the 3 size classes tested. Also indicated are the Pearson ( $r$ ) and Spearman ( $\rho$ ) correlation values.

**B)** Summary of these  $r$  and  $\rho$  values for each of the comparisons shown in (A).

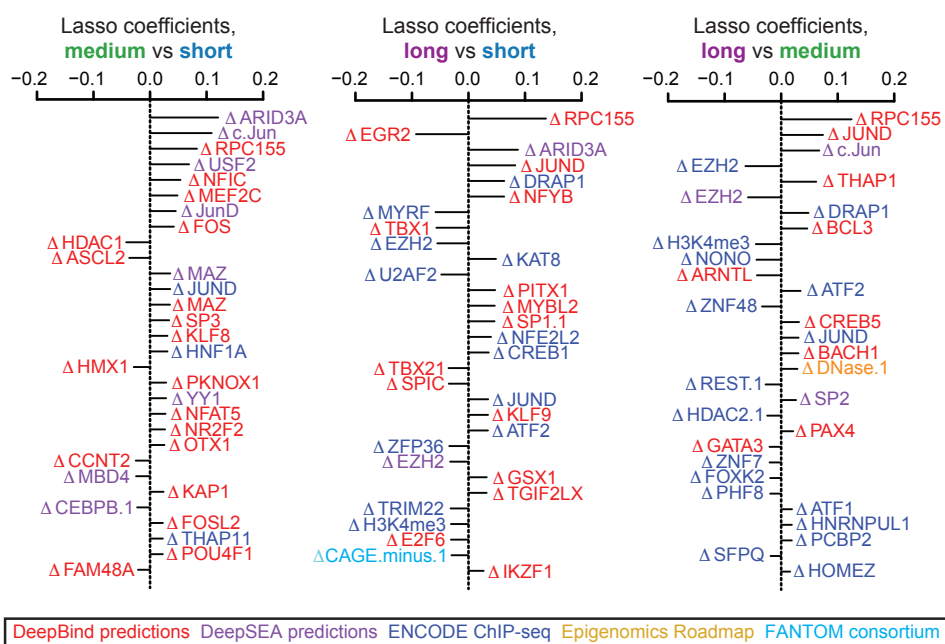
**C)** The top 30 coefficients derived from lasso regression models trained on the full dataset to predict observed values from the indicated size classes. Features with the extension “.1”, “.2”, etc allude to redundant features or replicate samples.

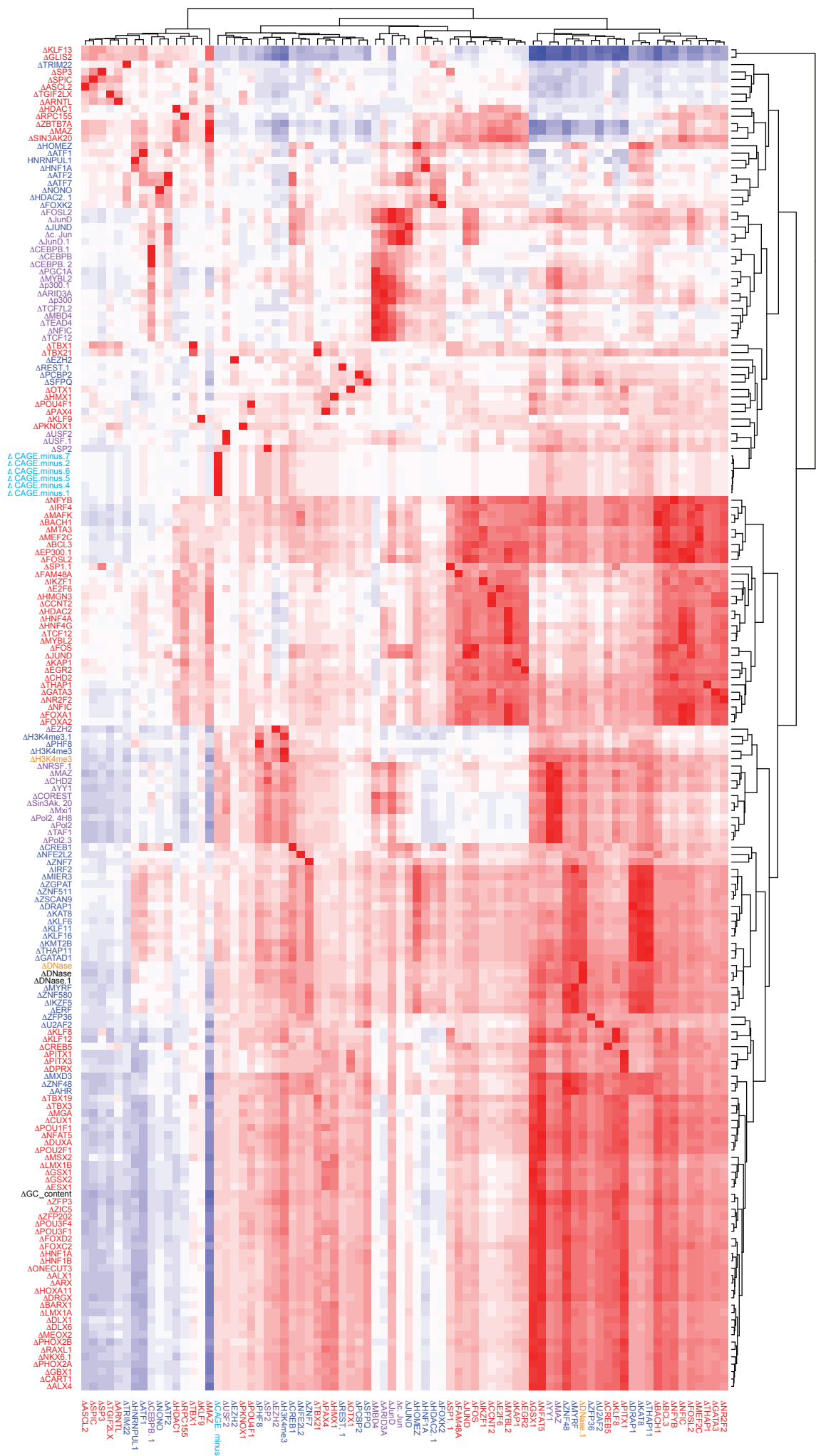
**D)** Figure has been rotated to fit to sheet. Pearson correlation matrix between the union of all top 30 features from (C), shown as rows, and other features sharing a Pearson correlation either  $\leq -0.8$  or  $\geq 0.8$ , shown as columns. Feature names are colored according to the origin of the feature as shown in the boxed key above. Hierarchical clustering was used to group features exhibiting similar correlation patterns.

A



B





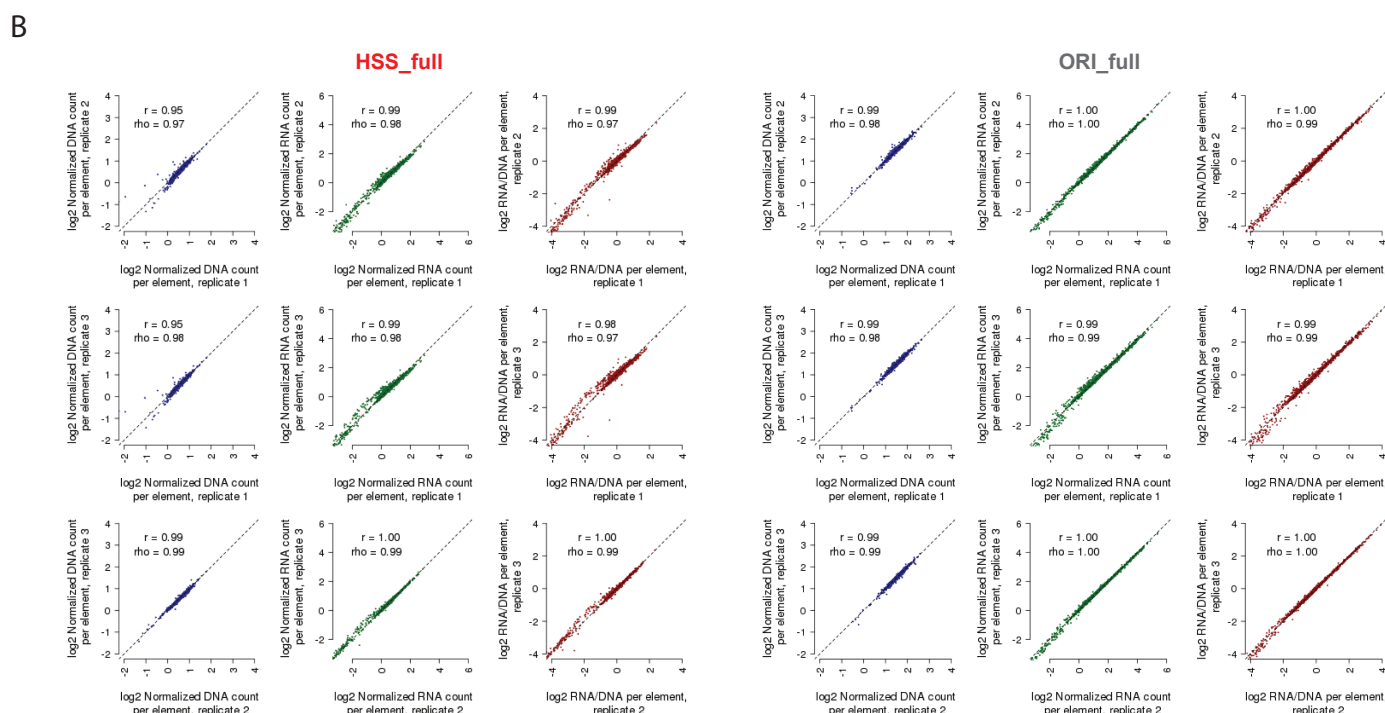
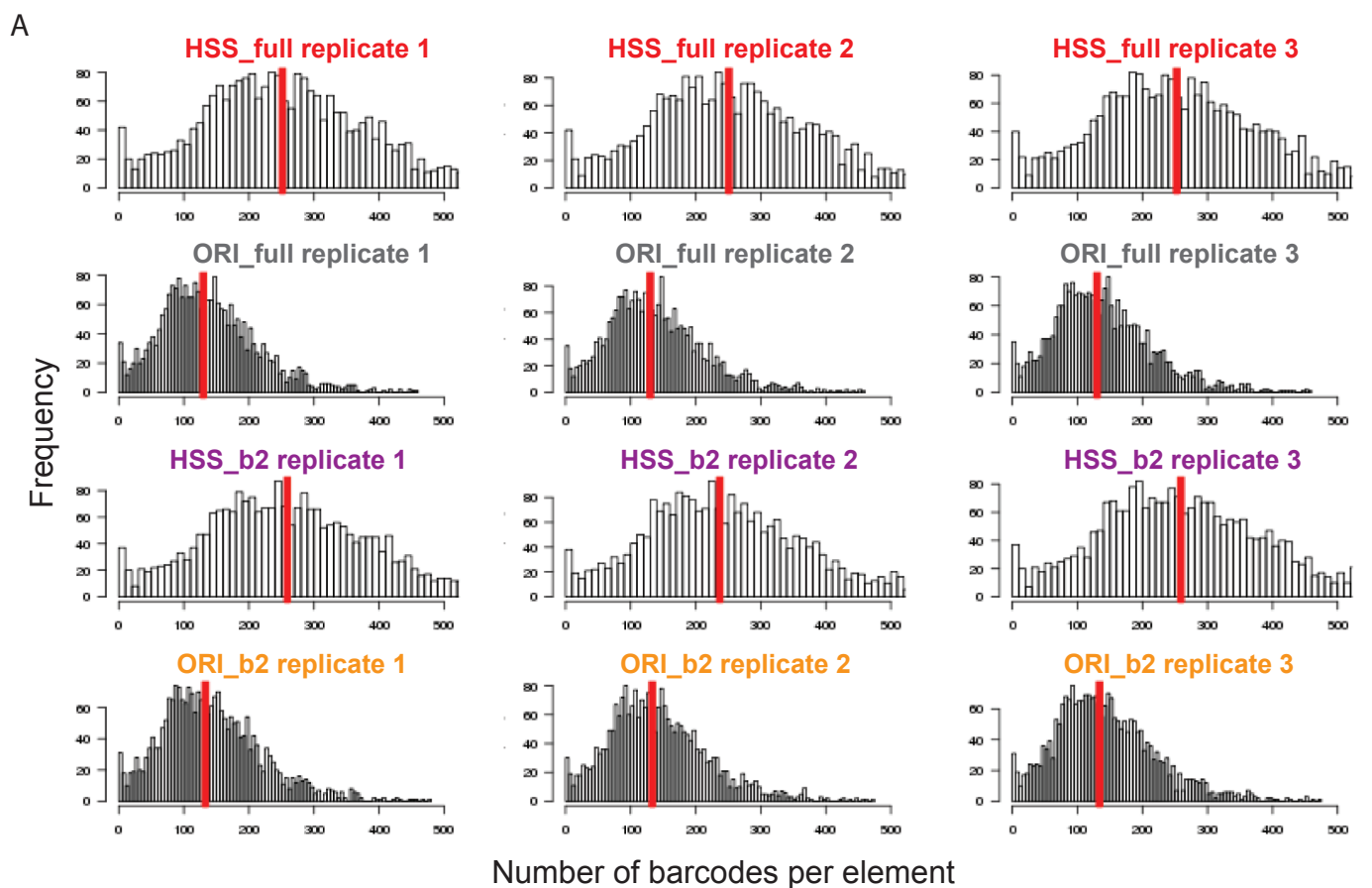
**Supplemental Figure 16. Coefficients and performance of the models to predict differential element efficacy among the three size classes.**

**A)** Scatter plots displaying the relationship between the 10-fold cross-validated predictions derived from lasso regression models and the observed RNA/DNA ratios, for each of differential size comparisons displayed. Also indicated are the Pearson ( $r$ ) and Spearman ( $\rho$ ) correlation values.

**B)** The top 30 coefficients derived from lasso regression models trained on the full dataset to predict observed values from the differential size comparisons indicated. Features with the extension “.1”, “.2”, etc allude to redundant features or replicate samples.

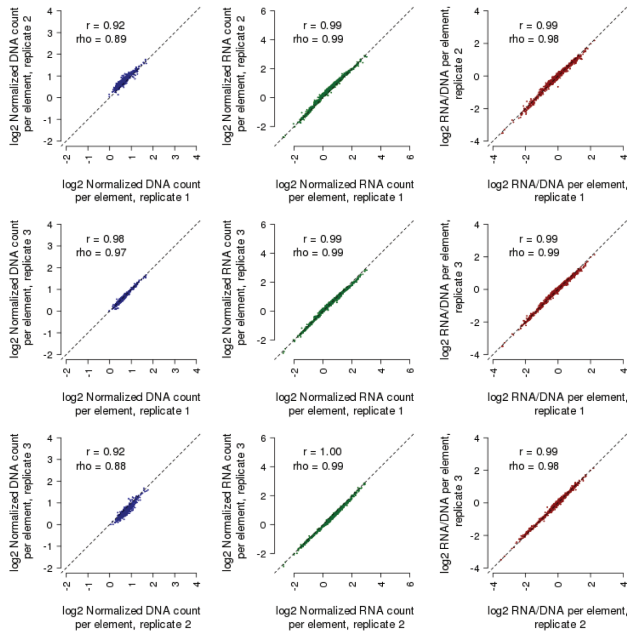
**C)** Figure has been rotated to fit to sheet. Pearson correlation matrix between the union of all top 30 features from (B), shown as rows, and other features sharing a Pearson correlation either  $\leq -0.8$  or  $\geq 0.8$ , shown as columns. Feature names are colored according to the origin of the feature as shown in the boxed key above. Hierarchical clustering was used to group features exhibiting similar correlation patterns.

Figure S17

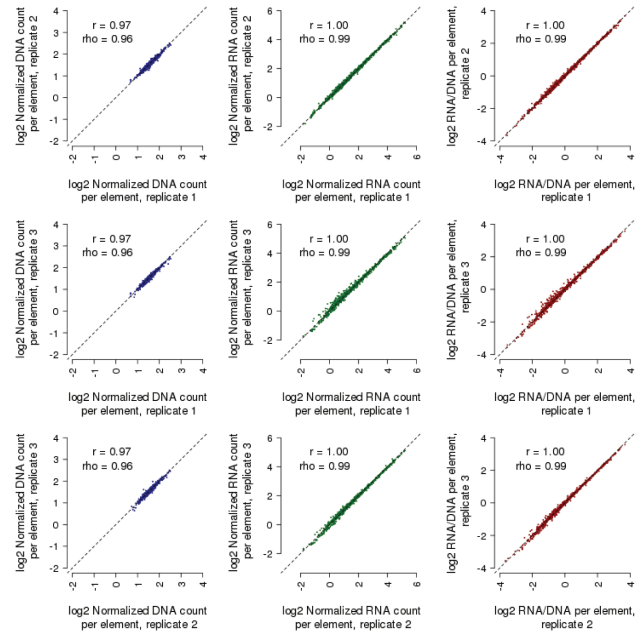




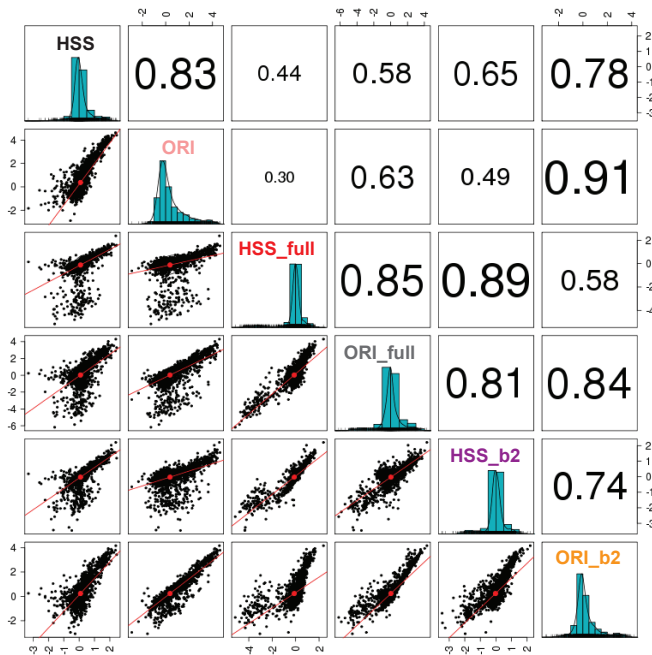
HSS\_b2



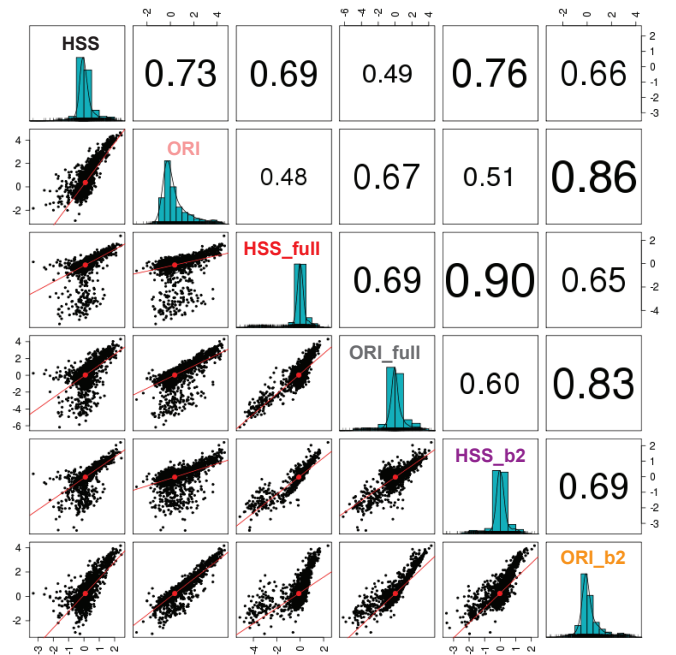
ORI\_b2



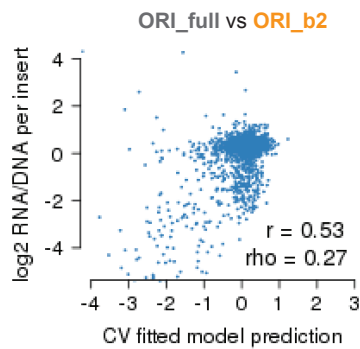
C



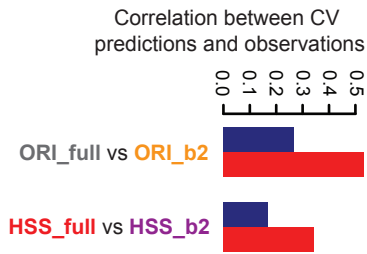
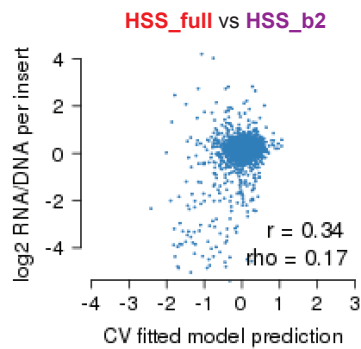
D



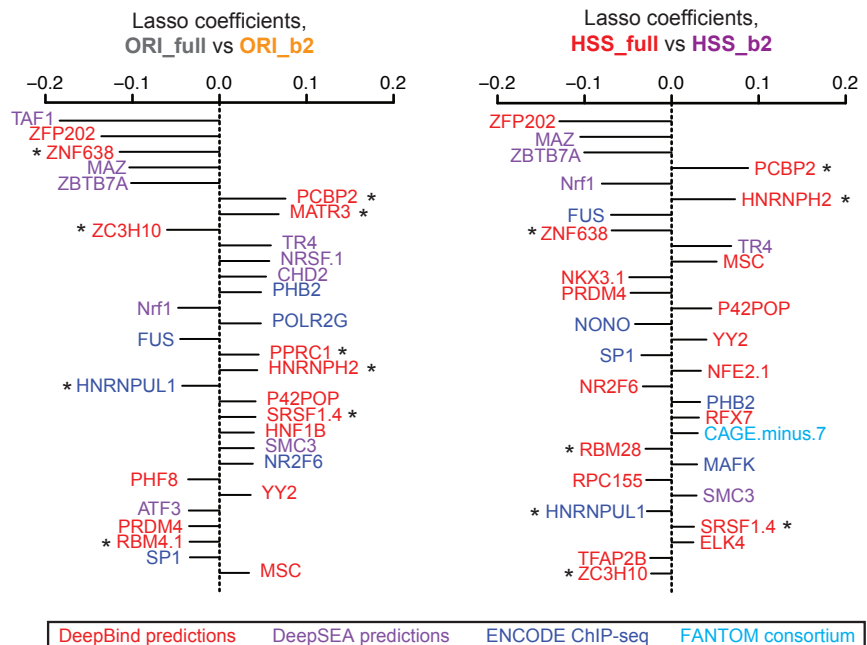
E

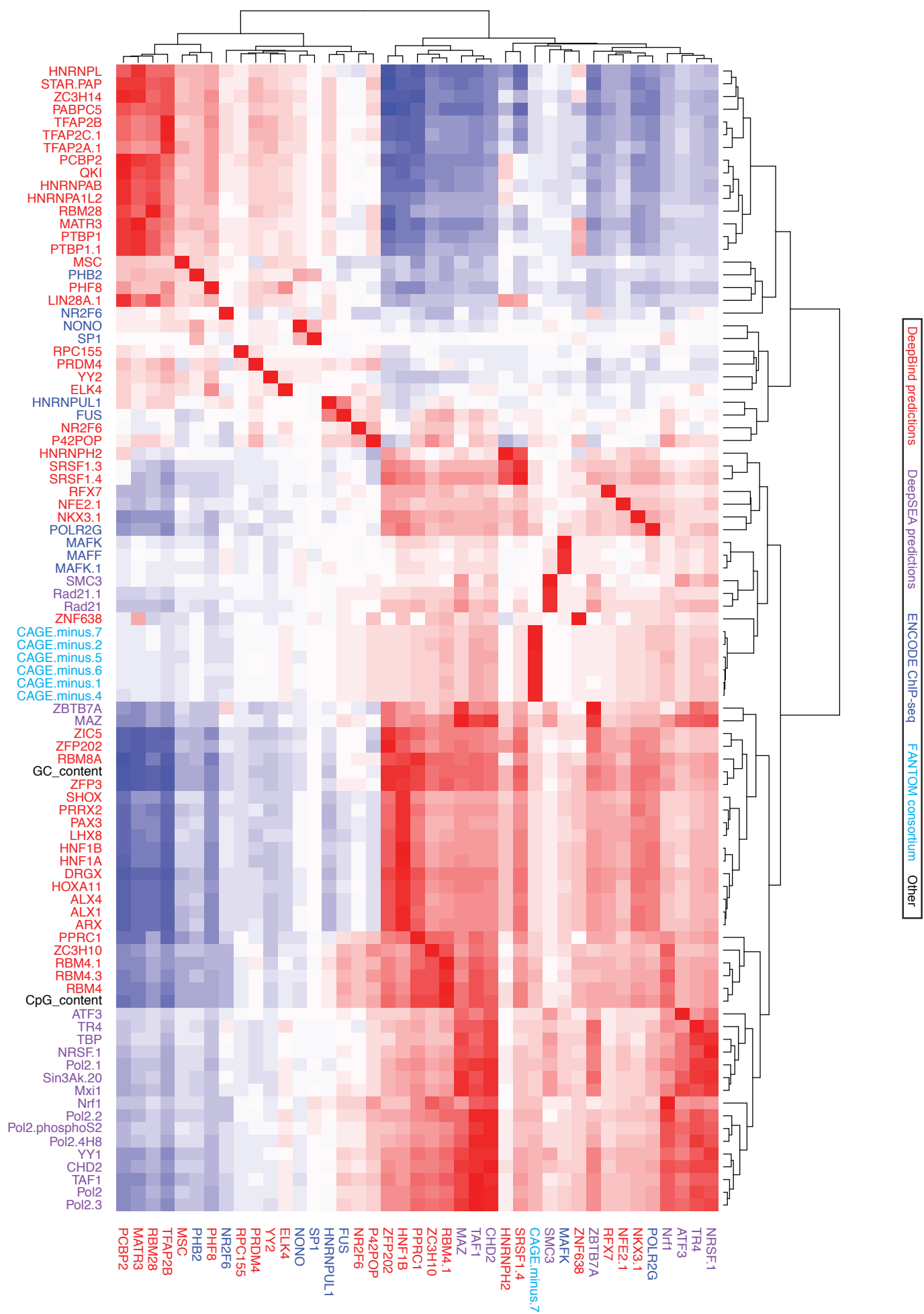


F



G





**Supplemental Figure 17. Evaluation of potential confounding artifact of internal transcriptional initiation in STARR-seq assays.**

**A)** Shown are histograms indicating the number of observed barcodes per element, for each of the 2,440 elements tested in each of the 3 replicates for each of the 4 MPRA methods indicated. HSS\_full and ORI\_full allude to assays in which full length mRNA (i.e., not arising from internal transcriptional initiation within the regulatory element tested) was measured. HSS\_b2 and ORI\_b2 (b2 for “batch 2”) allude to assays that were performed identically to HSS and ORI (as shown in **Figure 1**) but serve as an internal control for HSS\_full and ORI\_full, respectively, to mitigate batch effects. Shown with a vertical red line is the median number of barcodes per element.

**B)** Shown are scatter plots displaying the relationship between observed DNA counts (blue), RNA counts (green), and RNA/DNA ratios (red) for all pairwise comparisons among replicates, for each of the 4 additional MPRA methods tested. Also indicated is the Pearson ( $r$ ) and Spearman ( $\rho$ ) correlation values. Candidate enhancers supported by fewer than 10 barcodes were filtered out prior to this analysis to reduce the impact of technical noise.

**C-D)** Scatter matrix displaying scatter plots corresponding to each of the 15 pairs of possible inter-assay comparisons (lower diagonal elements). Shown on the diagonal is a histogram of the  $\log_2(\text{RNA/DNA})$  ratios, averaged among replicate samples, for the original HSS and ORI (as shown in **Figure 2B**) as well as HSS\_full, ORI\_full, HSS\_b2, and ORI\_b2. Also shown are Pearson (C) and Spearman (D) correlation values among each pair of comparisons, with the size of the text proportional to the magnitude of the correlation coefficient (upper diagonal elements).

**E)** Scatter plots displaying the relationship between the 10-fold cross-validated predictions derived from lasso regression models and the observed RNA/DNA ratios, for each of differential comparisons displayed. Also indicated are the Pearson ( $r$ ) and Spearman ( $\rho$ ) correlation values.

**F)** Summary of these  $r$  and  $\rho$  values for each of the comparisons shown in (E).

**G)** The top 30 coefficients derived from lasso regression models trained on the full dataset to predict observed values from the differential comparisons indicated. Features with the extension “.1”, “.2”, etc allude to redundant features or replicate samples, and asterisks (\*) allude to annotated RNA-binding proteins.

**H)** Figure has been rotated to fit to sheet. Pearson correlation matrix between the union of all top 30 features from (G), shown as rows, and other features sharing a Pearson correlation either  $\leq -0.8$  or  $\geq 0.8$ , shown as columns. Feature names are colored according to the origin of the feature as shown in the boxed key above. Hierarchical clustering was used to group features exhibiting similar correlation patterns.