

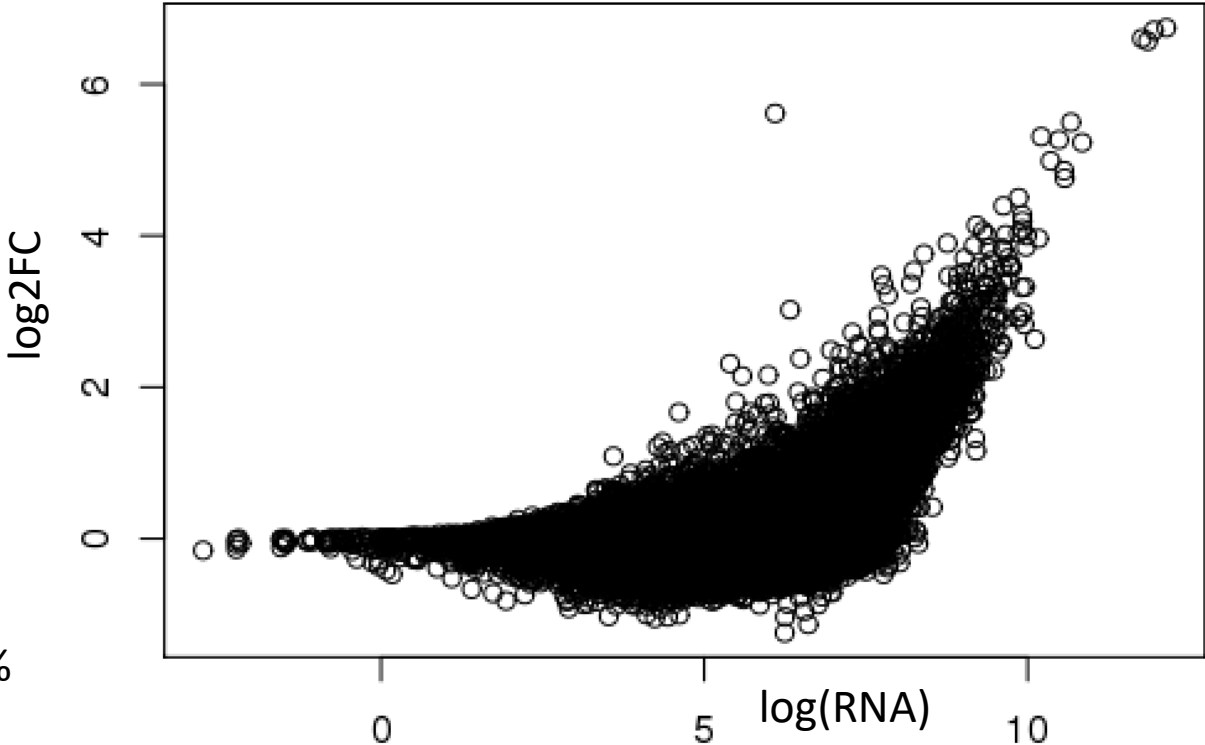
Data from Tewhey et al Cell 2016, Table S1

RNA = concatenation of **C.A.expr.mean** and **C.B.expr.mean**

log2FC = concatenation of **C.A.log2FC** and **C.B.log2FC**  
(RNA/plasmid) log2FC

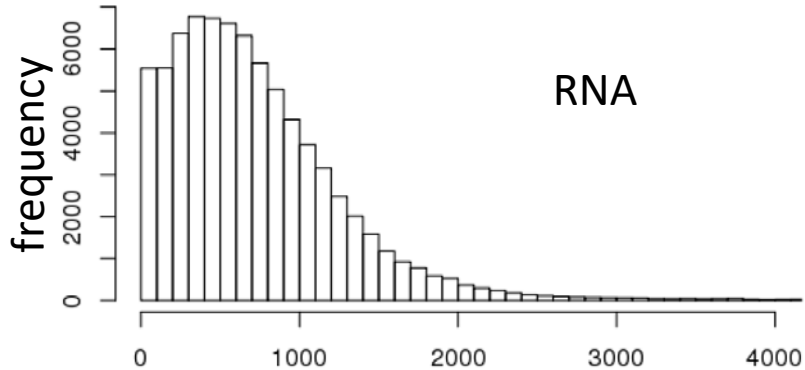
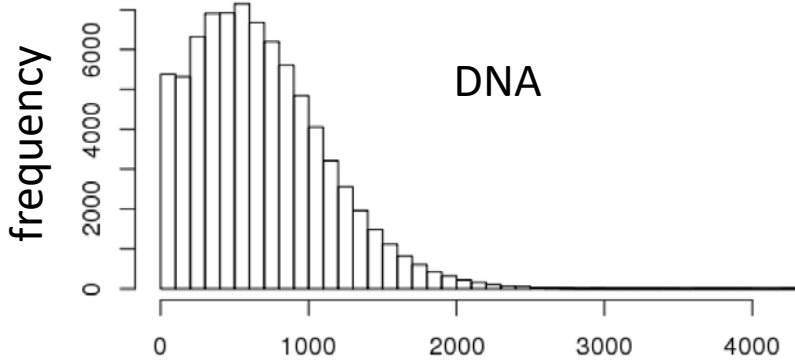
The pearson correlation of RNA and log2FC is only 0.35.

Only 43% of the top 10% of tested elements by RNA are in the top 10% of tested elements by log2FC. Only 44% of the top 20% of tested elements by RNA are in the top 20% of tested elements by log2FC.



The variability in DNA (= concatenation of **C.A.ctrl.mean** and **C.B.ctrl.mean**) is similar to the variability in RNA

There are sequence features predictive of DNA (synthesis, hybridization, or transfection efficiency)  
gkmSVM AUROC=0.97

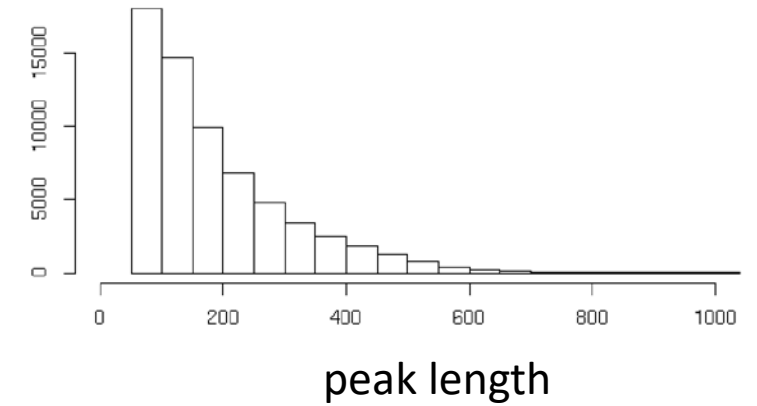
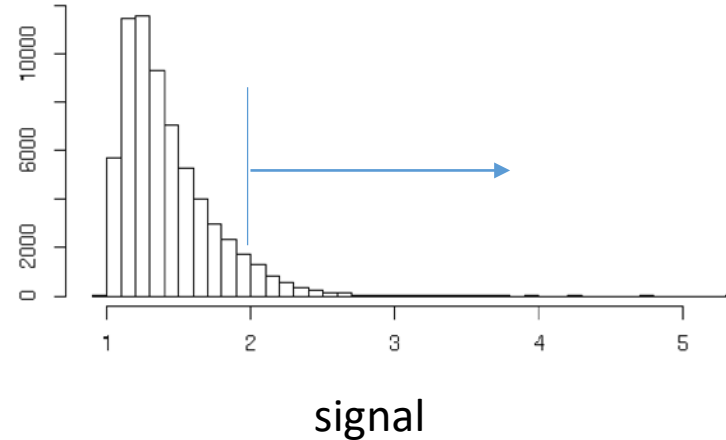


Kevin and Lijia's GM12878 data, peaks from:

```
rsync -avzP ENCODE@beaker.uchicago.edu::ENCODE_GM12878_peaks <local_directory>
```

Password: line-cooperative-guide-cook

took top 4000 peaks  
mapped to hg19



train gkmSVM on

1. DHS: GM12878 DHS peaks vs random (GC/rpt matched) (AUROC=.939)
2. MPRA R/D: top 1.1k log2FC (RNA/DNA) vs bottom 1.2k log2FC (RNA/DNA) (AUROC=.921)
3. MPRA R: top 4000k Kevin/Lijia vs random (GC/rpt matched) (AUROC=.607)

Cor(weights DHS, weights MPRA R/D) = 0.287

Cor(weights DHS, weights MPRA R) = 0.003

Cor(weights MPRA R/D, weights MPRA R) = 0.005

could check with K562 peaks