Supplementary Figures

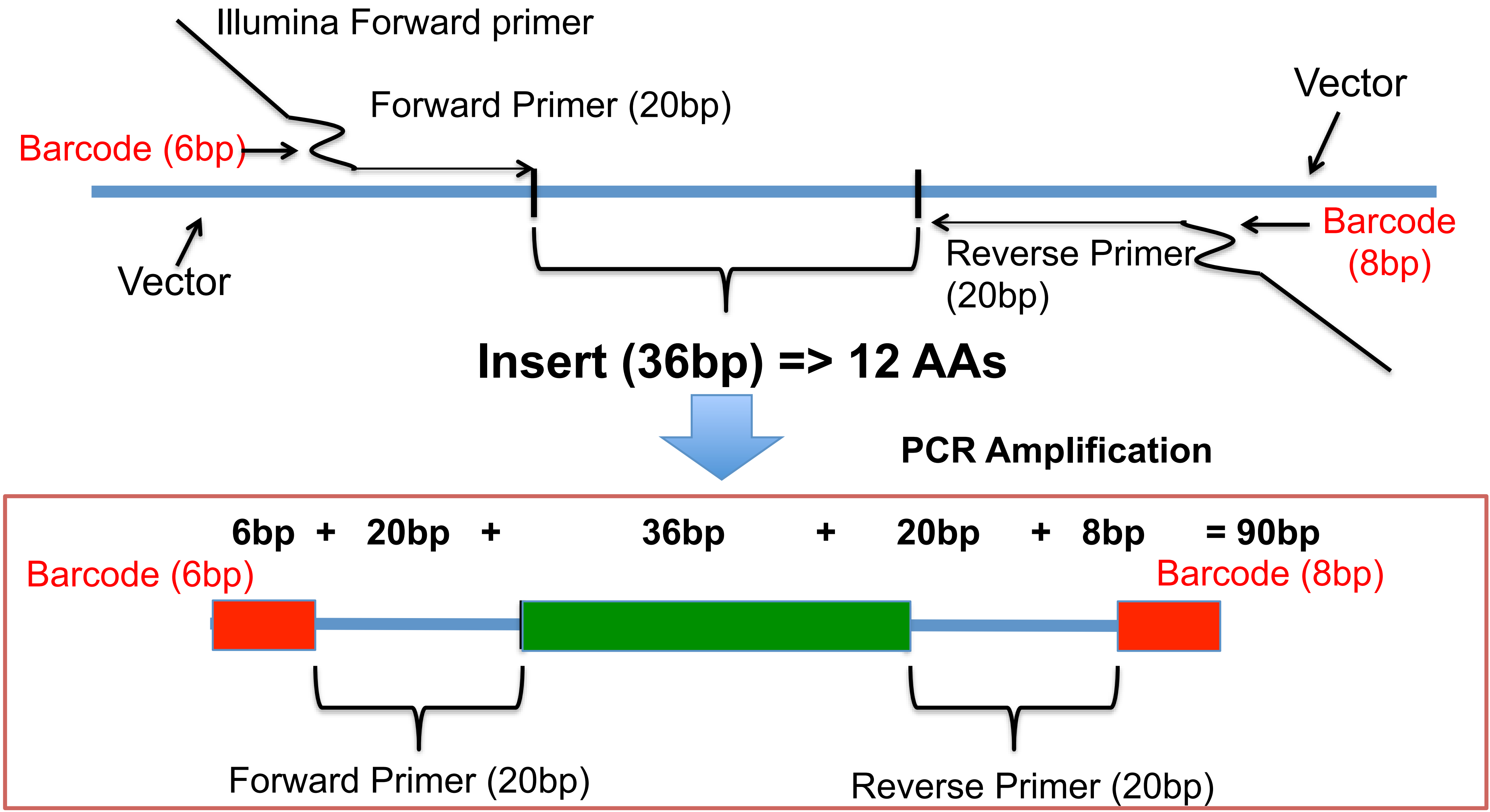
Supplementary Figure 1. An example of an Illumina Sequencing strategy compatible with MUSI.
Supplementary Figure 2. Histogram of AUC differences among MUSI, MEME, and Uniprobe (305 genomic datasets).
Supplementary Figure 3. Histogram of AUC on 305 genomic datasets (MUSI = red, MEME = green, and Uniprobe = blue).
Supplementary Figure 4. Performance comparison between MUSI and BEEML on 167 transcription factors from Berger et al. (a) Histogram of AUC differences between MUSI and BEEML (b) AUCs for MUSI and BEEML (c) Time comparison between BEEML and MUSI. For BEEML we have used matrices from Uniprobe for optimization. For MUSI, we have used our filtered data set (See *Materials and Methods*)
Supplementary Figure 5. Testing MUSI with noisy data. 100, 200, 300, 500, and 1000 random peptides have been added to the set of phage display peptides binding to SRC SH3 domain. MUSI identifies these noisy data in each example and assigns them to an additional unspecific motif (PWM3). The fraction after the comma in the last column indicates the number of random peptides correctly assigned to the flat logo.

Supplementary Figure 6. Testing MUSI with longer sequences. Originally, Src SH3 domain binders from phage-display experiments followed by next generation sequencing have a length of 12 amino acids. We have randomly inserted these peptides within random peptides of length 10, 15, 20, 25, 30, 35 and 40. As shown, MUSI can still detect multiple specificity with these random flanking regions, until alignment breaks (around 40-mers).

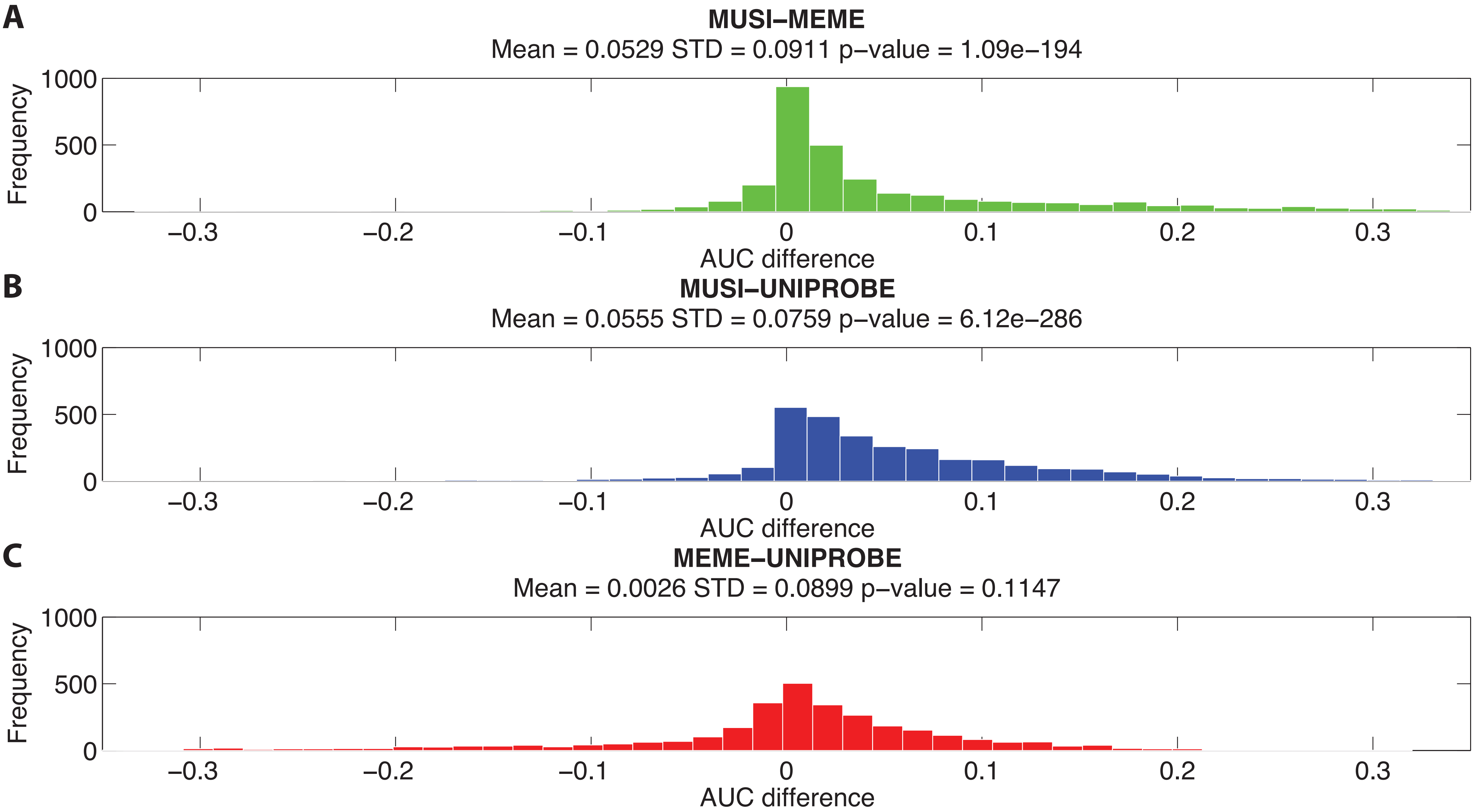
Supplementary Table 1. Cross-validation results (10-fold) for the 305 genomic datasets with different values for parameters P, D and D' used to determine the optimal number of PWMs (see Materials and Methods)

Supplementary Figure 1

Illumina Sequencing strategy for sequencing the phage encapsulated DNA



Supplementary Figure 2



A **MUSI–MEME**

Mean = 0.0529 STD = 0.0911 p–value = 1.09e–194

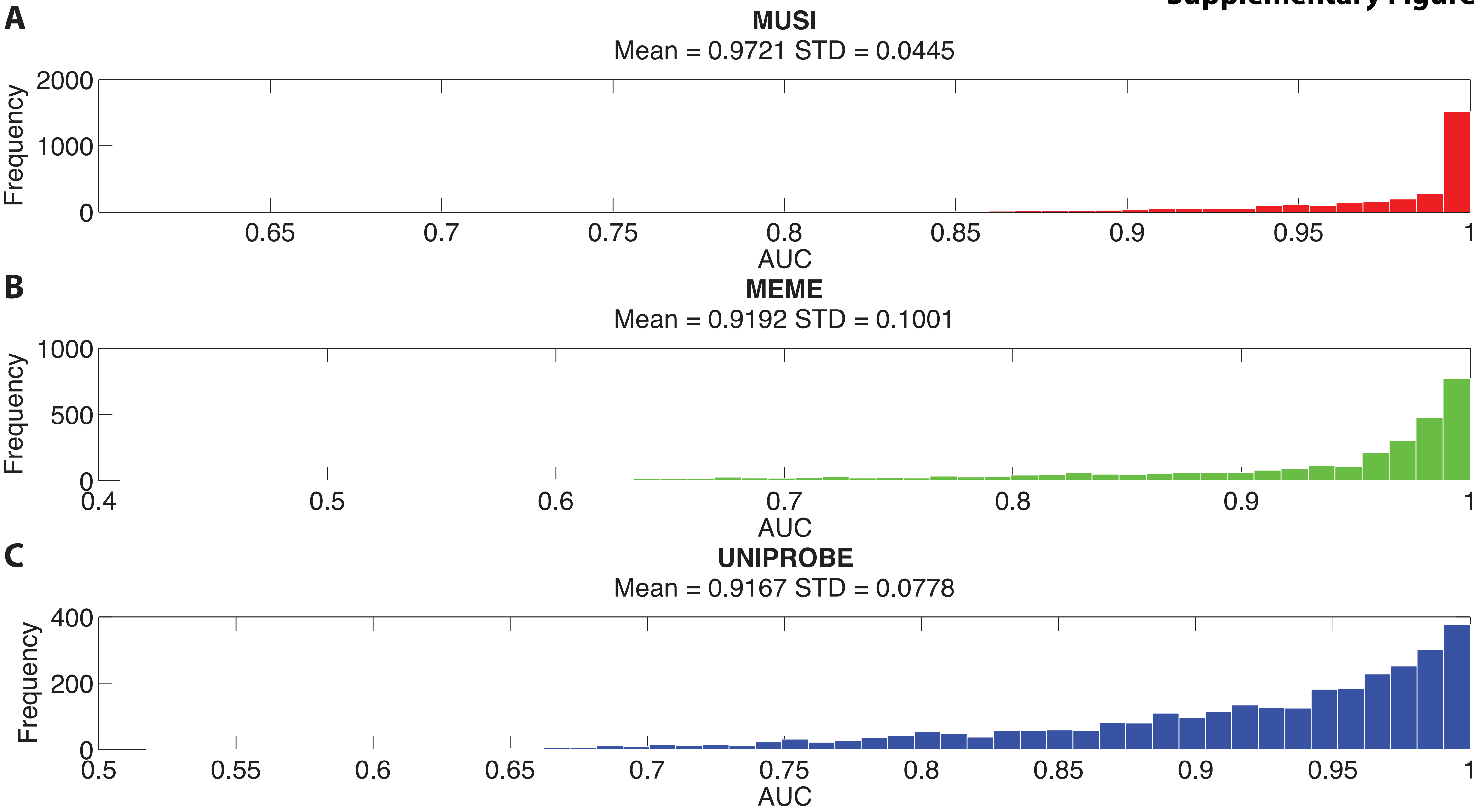
B **MUSI–UNIPROBE**

Mean = 0.0555 STD = 0.0759 p–value = 6.12e–286

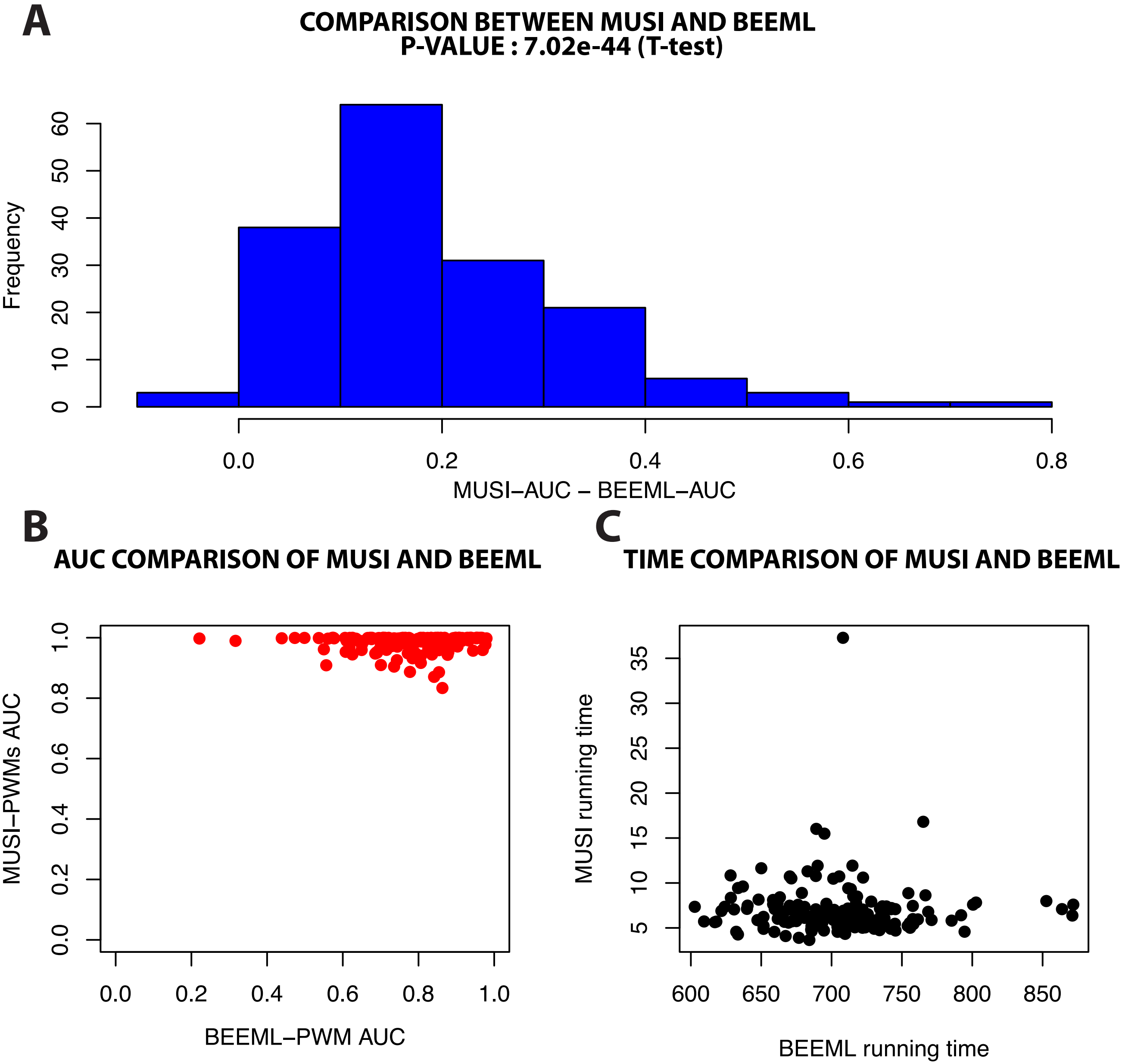
C **MEME–UNIPROBE**

Mean = 0.0026 STD = 0.0899 p–value = 0.1147

Supplementary Figure 3



Supplementary Figure 4

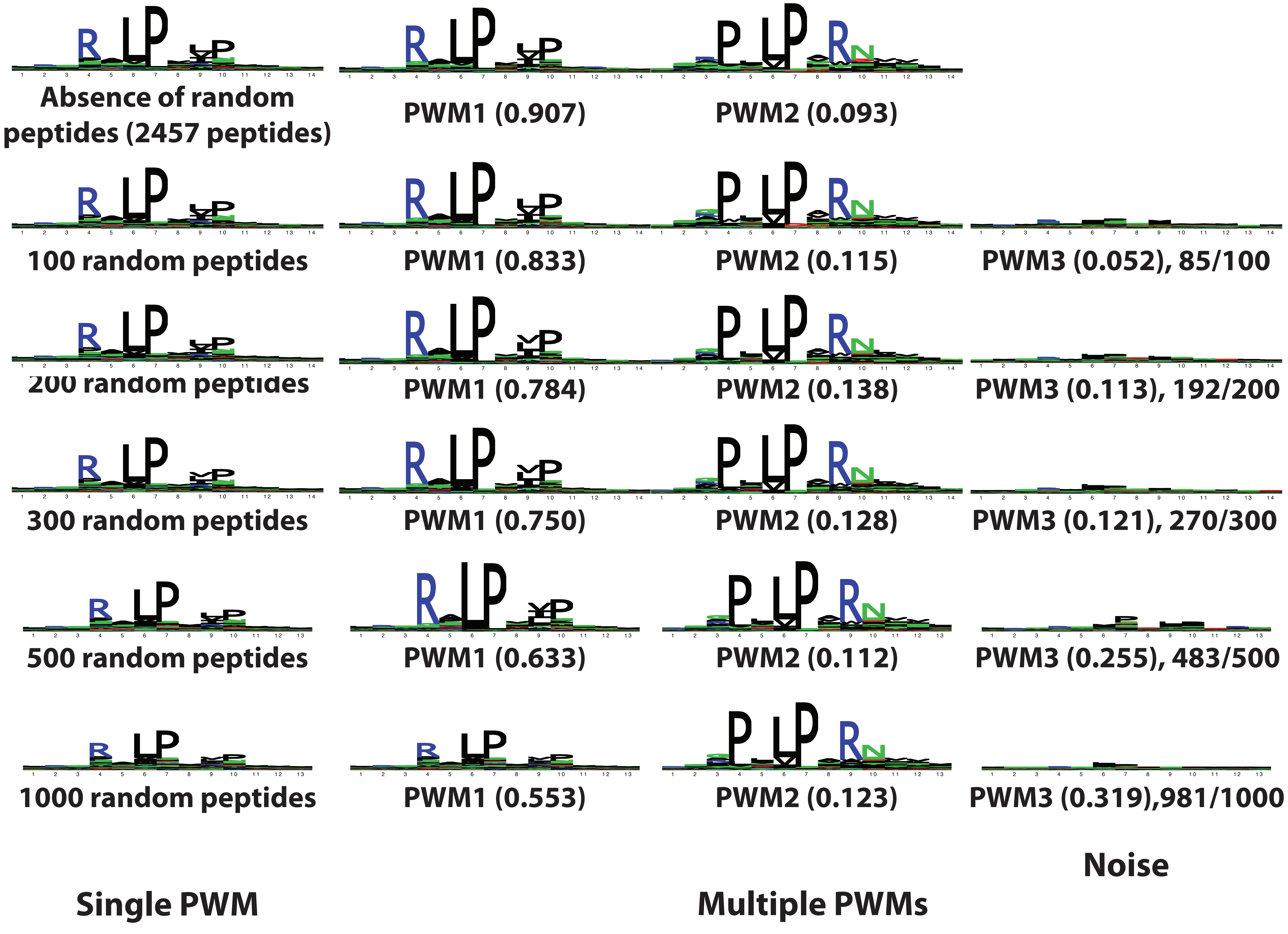


COMPARISON BETWEEN MUSI AND BEEML
P-VALUE : 7.02e-44 (T-test)

AUC COMPARISON OF MUSI AND BEEML

TIME COMPARISON OF MUSI AND BEEML

Supplementary Figure 5



Absence of random peptides (2457 peptides)

PWM1 (0.907)

PWM2 (0.093)

100 random peptides

PWM1 (0.833)

PWM2 (0.115)

PWM3 (0.052), 85/100

200 random peptides

PWM1 (0.784)

PWM2 (0.138)

PWM3 (0.113), 192/200

300 random peptides

PWM1 (0.750)

PWM2 (0.128)

PWM3 (0.121), 270/300

500 random peptides

PWM1 (0.633)

PWM2 (0.112)

PWM3 (0.255), 483/500

1000 random peptides

PWM1 (0.553)

PWM2 (0.123)

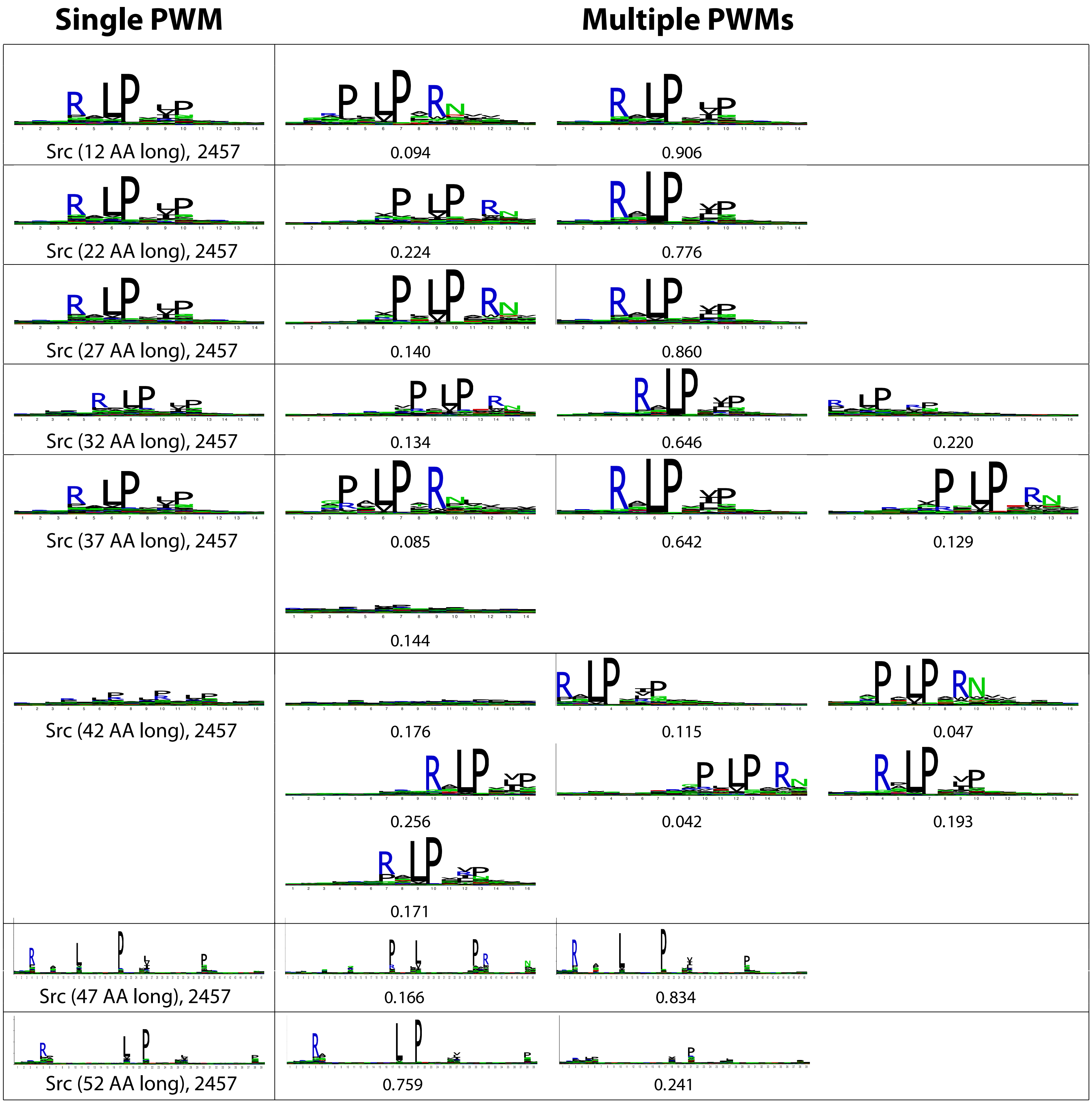
PWM3 (0.319),981/1000

Noise

Single PWM

Multiple PWMs

Supplementary Figure 6



Supplementary table 1

P	D	D'	AUC average	AUC Std
0.01	0.45	0.6	0.974511471	0.036966397
0.01	0.45	0.63	0.973908659	0.037237625
0.01	0.45	0.66	0.973300414	0.038864712
0.01	0.5	0.6	0.974139188	0.037530562
0.01	0.5	0.63	0.973552735	0.037785663
0.01	0.5	0.66	0.972950179	0.039377807
0.01	0.55	0.6	0.972704217	0.039603117
0.01	0.55	0.63	0.972108195	0.039879698
0.01	0.55	0.66	0.971510674	0.041543022
0.02	0.45	0.6	0.974229405	0.037409953
0.02	0.45	0.63	0.97382704	0.037236109
0.02	0.45	0.66	0.973093038	0.039081974
0.02	0.5	0.6	0.974146004	0.037455768
0.02	0.5	0.63	0.97376012	0.03727218
0.02	0.5	0.66	0.973016802	0.039119573
0.02	0.55	0.6	0.97356215	0.037734767
0.02	0.55	0.63	0.973186675	0.037564106
0.02	0.55	0.66	0.972426717	0.039554497