

The importance of incorporating transcript information
in CLIP-seq data analysis
Supplementary Material

Michael Uhl, Van Dinh Tran, and Rolf Backofen

March 9, 2020

Supplementary tables

Table 1: Exon overlap statistics of ENCODE eCLIP datasets (see Additional File 1 in .xlsx format).

Table 2: Length statistics for called peaks by CLIPper, PEAKachu, and PureCLIP on YBX3 K562 replicate 1 eCLIP data. Peaks were called as described in supplementary methods section "Peak caller setup". Introns for determining overlapping sites were selected based on the set of exons extracted, as described in methods section "Data preparation and exon overlap statistics". A site is counted as intron-spanning if it completely overlaps with an intronic region.

Metric	CLIPper	PEAKachu	PureCLIP
# sites	132,842	11,537	54,308
# sites > 500 nt	0	471	0
# intron-spanning sites	4	1,096	0
Minimum length	1	18	1
Maximum length	263	22,875	25
Mean length	37.9	112.4	1.6
Median length	34	48	1
25th percentile	19	42	1
75th percentile	51	64	2

Table 3: Dataset statistics for the 5 eCLIP sets used for genomic and transcript context comparison. \log_2 FCT: \log_2 fold change threshold used for filtering initial sites. Positives: number of positive training instances. Negatives: number of negative training instances. Only exonic sites (overlapping $\geq 90\%$ with exons) nearby exon borders were selected and filtered by given \log_2 fold change.

RBP	Cell type	\log_2 FCT	Positives	Negatives
FXR1	K562	2	8672	8672
NIP7	HepG2	2	3791	3791
RPS3	HepG2	3.2	9937	9937
SERBP1	K562	1	1485	1485
YBX3	K562	3	8631	8631

Table 4: Performance results for 5 RBP eCLIP sets with genomic and transcript context. We report average accuracies obtained by 10-fold cross validation together with standard deviations (apart from GraphProt).

Methods	RBP	Cell line	Genomic context	Transcript context
DeepBind	FXR1	K562	79.15 \pm 2.73	88.53 \pm 1.08
	NIP7	HepG2	77.77 \pm 2.04	87.01 \pm 1.76
	RPS3	HepG2	80.25 \pm 1.33	88.45 \pm 1.33
	SERBP1	K562	78.36 \pm 2.66	87.81 \pm 2.15
	YBX3	K562	79.09 \pm 2.28	88.21 \pm 0.79
GraphProt	FXR1	K562	74.84	87.08
	NIP7	HepG2	77.76	87.74
	RPS3	HepG2	75.22	87.55
	SERBP1	K562	77.11	89.95
	YBX3	K562	79.45	90.66
GraphProt2	FXR1	K562	82.39 \pm 0.90	89.38 \pm 0.80
	NIP7	HepG2	81.58 \pm 0.95	89.79 \pm 0.93
	RPS3	HepG2	83.06 \pm 0.72	90.17 \pm 0.60
	SERBP1	K562	82.29 \pm 2.57	91.82 \pm 2.27
	YBX3	K562	81.57 \pm 0.89	90.71 \pm 0.82