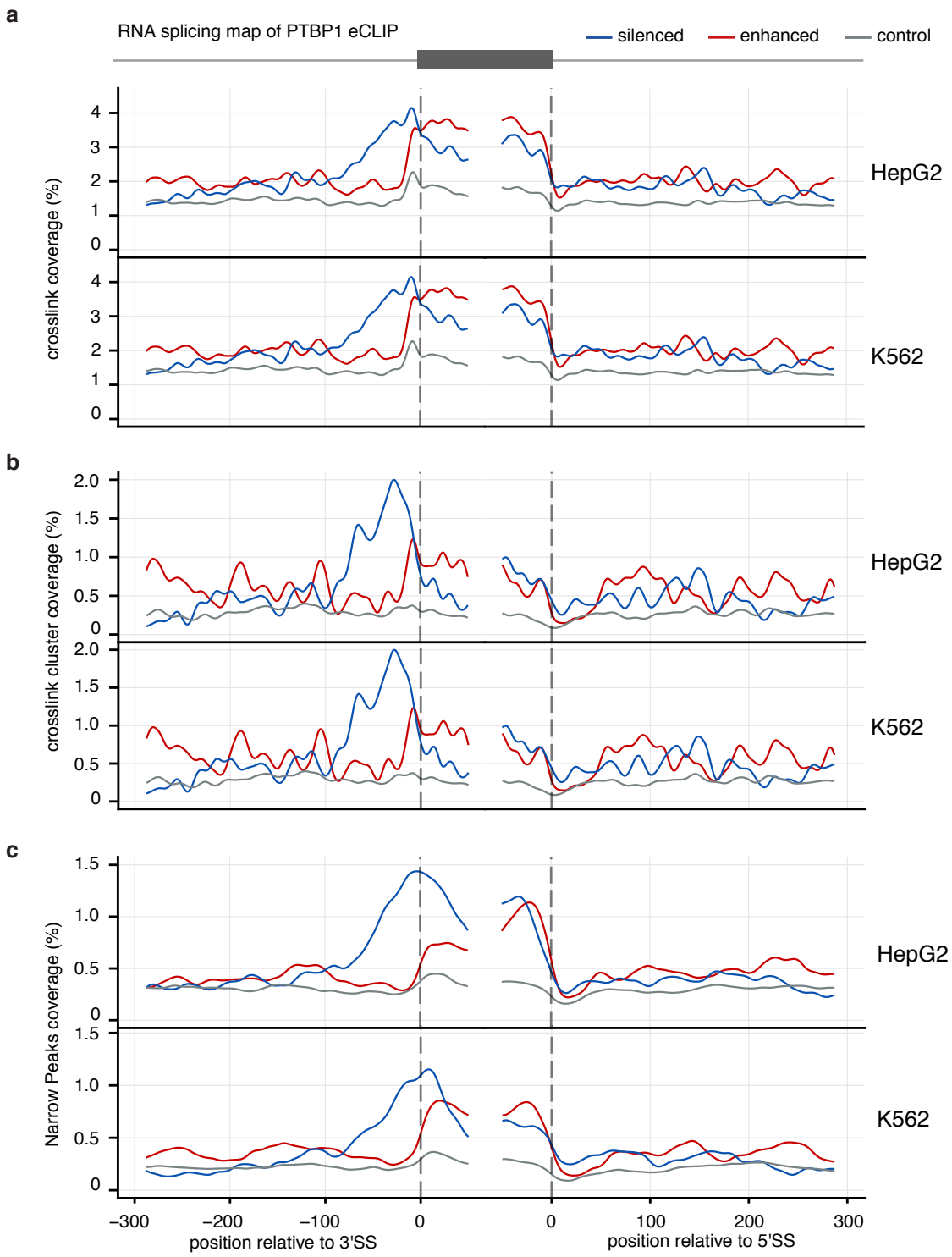
Supplemental Material: Annu. Rev. Biomed. Data Sci. 2018. 1:235–61
https://doi.org/10.1146/annurev-biodatasci-080917-013525
Data Science Issues in Studying Protein–RNA Interactions with CLIP Technologies
Chakrabarti et al.

Supplemental Figure 1: Analysing eCLIP RNA maps using data from the same cell lines

RNA maps are used to explore the sensitivity and specificity of PTBP1 eCLIP data in both K562 and HepG2 cells from the ENCODE website. The regulated exons are defined using RNA-seq data following PTBP1 CRISPR knockout in K562 cells also from the ENCODE website where we observed greater signal compared to the shRNA followed by RNA-seq data. We identified the skipped exons detected using rMATS (101) using only junction counts and a p-value threshold of 0.05 and FDR threshold of 0.1. Silenced and enhanced exons were defined using an inclusion level difference threshold of 0.05; control exons were selected as those with a p-value greater than 0.1, an FDR value greater than 0.1, an inclusion level of less than 0.9, and an inclusion level difference less than 0.05.

In a) we show the raw data. In b) we use peaks identified using iCount (using a 3 nucleotide peak calling half-window and 7 nucleotide clustering window). In c) We use the eCLIP peaks defined using the narrowPeaks algorithm and available from the ENCODE website.

The code to reproduce this figure is available at https://github.com/jernejule/clip-data-science


a
RNA splicing map of PTBP1 eCLIP
silenced
enhanced
control
crosslink coverage (%)
HepG2
K562
b
crosslink cluster coverage (%)
HepG2
K562
c
Narrow Peaks coverage (%)
HepG2
K562
−300
−200
−100
0
0
100
200
300
position relative to 3'SS
position relative to 5'SS