

# LAPA BED file format specifications

Prepared by Fairlie Reese  
January 25, 2022  
Mortazavi Lab, University of California, Irvine

## Contact Information

Fairlie Reese  
2300 Biological Sciences III  
University of California Irvine  
Irvine, CA 92697-2300  
Telephone: (949) 824-8393  
Email: [freese@uci.edu](mailto:freese@uci.edu)

## Poly(A) site BED file format

Poly(A) BED files output from LAPA have the following file format; also detailed in the LAPA documentation: <https://lapa.readthedocs.io/en/latest/output.html>

- Chromosome: Chromosome of the poly(A) cluster.
- Start: Start position of poly(A) cluster.
- End: End position of poly(A) cluster.
- polyA\_site: Exact poly(A) site (peak) of the cluster.
- count: number of reads supporting the cluster (reads that end in the cluster).
- Strand: Strand of the poly(A) cluster.
- Feature: Genomics feature overlapping with the cluster (obtained from GTF file).
- gene\_id: The gene containing the poly(A) clusters.
- tpm: TPM of the cluster calculated by  $\text{count} / \text{sum}(\text{count}) * 1,000,000$ .
- gene\_count: Total number reads in all the clusters of this gene calculated by  $\text{sum}(\text{count}_i)$  where  $i$  is in gene.
- usage: Percentage use of specific poly(A) clusters of the gene calculated by  $\text{count} / \text{gene\_count}$
- fracA: Number of A bp following the poly(A)-site
- Signal: poly(A) signal sequence and the region containing poly(A) signal
- annotated\_site: End position of the 3' UTR based on the GTF if poly(A) cluster is located in 3' UTR.

## TSS site BED file format

- Chromosome: Chromosome of the tss cluster.
- Start: Start position of tss cluster.
- End: End position of tss cluster.
- tss\_site: Exact tss site (peak) of the cluster.
- count: number of reads supporting the cluster (reads that end in the cluster).
- Strand: Strand of the tss cluster.
- Feature: Genomics feature overlapping with the cluster (obtained from GTF file).
- gene\_id: The gene containing the tss clusters.
- tpm: TPM of the cluster calculated by  $\text{count} / \text{sum}(\text{count}) * 1,000,000$ .
- gene\_count: Total number reads in all the clusters of this gene calculated by  $\text{sum}(\text{count}_i)$  where  $i$  is in gene.
- usage: Percentage use of specific tss clusters of the gene calculated by  $\text{count} / \text{gene\_count}$
- annotated\_site: End position of the 3' UTR based on the GTF if tss cluster is located in 3' UTR.