PolyA Atlas BED file 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

More information about the PolyA atlases can be found here:
https://www.polyasite.unibas.ch/atlas

The following file format specifications were taken directly from the above website.

File specifications

The columns represent:

first - chromosome name

second and third - start and end positions of the poly(A) site cluster, respectively

fourth - unique cluster ID, composed of the chromosome name, the representative poly(A) site of the cluster and the strand. Note that this ID format is inspired by UCSC's position format, which uses 1-based coordinates instead of the 0-based bed coordinates used in the second and third columns. Thus, to convert the position of the representative site to bed coordinates, subtract 1.

fifth - average expression (tags per million, tpm) across all samples

sixth - strand on which the cluster is encoded

seventh - percentage of samples that support the particular cluster

eighth - number of different 3' end sequencing protocols that support the particular cluster

ninth - average expression (tags per million, tpm) across all samples

tenth - two letter code for the cluster annotation (in order of decreasing priority: TE, terminal exon; EX, exonic; IN, intronic; DS, 1,000 nt downstream of an annotated terminal exon; AE, anti-sense to an exon; AI, anti-sense to an intron; AU, 1,000 nt upstream in anti-sense direction of a transcription start site; IG, intergenic)

eleventh - information about the poly(A) signal(s) that are present upstream of the poly(A) site, including the motif, the location with respect to the cleavage site and the genomic coordinate