**Required Format for Shiloh Toolkit Input Files**

The toolkit runs on CLIP data in either **CSV** or **BED** formats. Be sure to include the required fields and to have the exact header names needed.

**CSV -** A comma separated file with the following header names (other columns can be present and the program will ignore them).

*Required Fields*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column Name | **Chr** | **Strand** | **Start** | **End** |
| Description | The chromosome the CLIP cluster aligned to. This can be expressed either as a number/character (1 or X) or with the prefix “chr” (chr1 or chrx). | The strand this alignment corresponds to (+ or -). | The starting genomic coordinate for this CLIP cluster. | The ending genomic coordinate for this CLIP cluster. |

*Optional Fields (Can be used for filtering and/or scaling purposes)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column Name | **ReadCount** | **UniqueReads** | **T2Cfraction** | **ConversionSpecificity** |
| Description | The number of reads that map to this CLIP cluster. | The number of unique reads that map to this CLIP cluster. | The fraction of diagnostic T2C fractions in the cluster. | log(# of reads with T-to-C conversions / # of reads with other conversions) |

**BED -** A tab separated file containing the following fields.

*Required Fields*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column Name | **chrom** | **chromStart** | **chromEnd** | **GeneName** |
| Description | The chromosome the CLIP cluster aligned to. Naming convention should match that of the annotation .csv file used. | The end of the feature with respect to the reference chromosome. | The end of the feature with respect to the reference chromosome. | The name of the gene the BED row maps to.  Note: If you do not have the GeneName column you may run the program with the “-bedGTF” parameter in order to try to map bed rows to genes. This is somewhat time consuming. |