**APA-Scan User Manual**

Naima Ahmed Fahmi, Wei Zhang

1. **Download**

APA-Scan is downloadable directly from [github](https://github.com/compbiolabucf/APA-Scan). Users need to have python (version 3.0 or higher) installed in their machine.

1. **Required Softwares**
   1. [Python](https://www.python.org/downloads/) (v3.0 or higher)
   2. [Samtools (v 0.1.8)](https://sourceforge.net/projects/samtools/files/samtools/0.1.8/)\* [This specific version is mandatory]
2. **Required python packages**
   1. [Pandas](https://pandas.pydata.org/)
   2. [Bio](https://biopython.org/)
   3. [Scipy](https://www.scipy.org/)
   4. [Numpy](https://www.numpy.org/)
   5. [Peakutils](https://peakutils.readthedocs.io/en/latest/)
3. **Running APA-Scan**

APA-Scan can handle both human and mouse data for detecting potential APA truncation sites. The tool is designed to follow the format of [Refseq annotation](https://www.ncbi.nlm.nih.gov/refseq/) and genome file from [UCSC Genome Browser](https://genome.ucsc.edu/). Users need to have the following two files in the parent directory in order to run APA-Scan:

* Refseq annotation (in .txt format)
* Genome fasta file (downloaded from UCSC genome browser)

APA-Scan comprises of two python scripts:

* APA-scan.py
* Make-plots.py

**Run APA-scan.py**

****

**Example:**

****

**Options:** (\*denotes mandatory fields)

|  |  |
| --- | --- |
| -s/-S\* | Species name. APA-Scan can handle human and mouse in the current version. Users have to specify **h** for human and **m** for mouse. |
| input1\_dir\* | Required field, directory of input1 RNA-seq data |
| input2\_dir\* | Required field, directory of input2 RNA-seq data |
| -o/-O | Denotes output directory. Its an optional field. If -o is not specified, the results will be generated inside of ‘Output’ folder. |
| -p/-P | P denotes whether the user gives the 3’end-seq data or not. If -p is initialized, the next two fields after -p will be the directories of 3’end data for two samples. If -p is not specified, APA-Scan will automatically determines APA events according to its algorithm. |

**Run Make-plots.py**

****

Make-plots.py will ask the user to insert the region of interest in a specific way:

**Chrom:GeneName:RegionStart-RegionEnd**

**Parameters Explanation:**

Chrom: Chromosome Name. Example: chr1

GeneName: denotes the gene ID or gene Name. Example: Tceb1

RegionStart: Start of the untranslated region

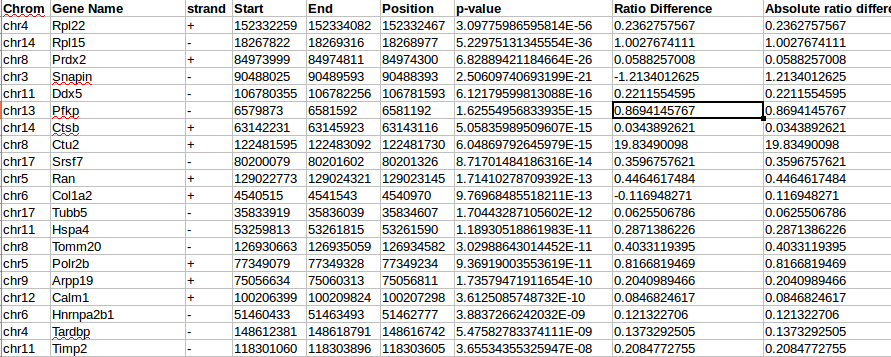
RegionEnd: End of the untranslated region

Example:

**chr1:Tceb1:16641724-16643478**

1. **Results**

APA-Scan will generate a spreadsheet in the output directory, with the potential transcript splice site for each region. The result file contains the following fields(see image below) as long as all other information necessary to compute the association among two samples.



Make-Plots.py will generate a visual representation of the results shown above, for each of the region entered. The plot will illustrate the most significant transcript cleavage site with a red vertical bar on top of RNA-seq read data (see figure below). It will also show the UTR truncation(annotated and unannotated) at the bottom panel.

