Dependencies:

Bash

Java (version 1.8.0_171 or more recent):
Download Location: https://java.com/en/download/
Command: java -jar

SAMTools:
Download Location: https://github.com/samtools/samtools/releases
Command: samtools

HOMER
Download Location: http://homer.ucsd.edu/homer/
Command: findMotifsGenome.pl
Command: annotatePeaks.pl

In addition, install the appropriate reference packages (e.g. hg19, hg38) in HOMER.

Python:
Download Location: https://www.python.org/downloads/
Command: python (This can be configured in the GUI)

Python Libraries:
1. numpy
2. pandas
3. sklearn
4. matplotlib
5. keras
6. tensorflow

These packages can be installed using *pip* or *conda*:

```
pip install numpy pandas scikit-learn matplotlib keras tensorflow
```

```
conda install --upgrade numpy pandas scikit-learn matplotlib keras tensorflow
```

Dependencies for Building PEASTools from Source

htsjdk.samtools
Download Location: https://github.com/samtools/htsjdk

apache math commons
Download Location: http://commons.apache.org/proper/commons-math/

Step 1: Data Preparation:

To prepare the data and encode it for model prediction, you will need the following files:

1. The paired-end ATAC-seq BAM file
2. Peaks called from this BAM file. (Recommended peak calling strategy: MACS2 using "-f BAMPE –nomodel"
3. FASTA (.fa) file for the corresponding reference
4. A directory with chromosome separated fasta (.fa) files
5. A specified output directory

With the CoRE-ATAC source code and the above files/directories, you should have all you need to begin!

Next, run the FeatureExtractor.sh file (use chmod (e.g., chmod +x FeatureExtractor.sh) to make it executable if necessary) by providing the following arguments:

1. The absolute path to the BAM file.
2. The absolute path to the peak file
3. The absolute path to the output directory
4. The reference genome (e.g., hg19 or hg38)
5. The absolute path to the FASTA file
6. The absolute path to directory containing separated chromosome fasta files
7. The path of the CoRE-ATAC source code
8. Whether or not to keep duplicate reads (used for snATAC-seq data at the moment): Provide "TRUE" without quotes for the final argument to keep duplicates (ignore/provide any other input to omit duplicates)

This script will take a few hours to complete, varying based on the size of the BAM file.

Step 2: Predicting *cis*-RE Functional Annotations:

Once the ATAC-seq data has been encoded, all the information needed for predicting cis-RE function (i.e., Promoter, Enhancer, Insulator, and "Other") is now available.

Download the latest pretrained model from https://github.com/UcarLab/CoRE-ATAC

Next run the python file: **CoREATAC_PredictionTool.py** with the following execution:
python3 CoREATAC_PredictionTool.py <arg 1> <arg 2> <arg 3>

Providing the following input arguments:

1. The absolute path to the output directory specified in step 1
2. The absolute path to the pre-trained model (.h5) file
3. The output file

The output file will be a tab delimited text file with the following columns:

1. The chromosome of the ATAC-seq peak
2. The start position of the ATAC-seq peak
3. The end position of the ATAC-seq peak
4. Promoter probability
5. Enhancer probability
6. Insulator probability
7. Other probability