# ATAC-seq Processing/Sequence Alignment Guide

The purpose of this guide is to provide each step in obtaining ATAC-seq alignments from fastq files. Only steps 1-4 of the ATAC-seq pipeline will be executed and file cleanup will need to be performed manually. For each step, this guide will provide information on what files can be deleted. This information is important for maintaining low disk space usage as data is duplicated in each step of the pipeline.

### Step 1: Directory Setup

Make the directory that will store all ATAC-seq fastq (or fastq.gz) files that will be processed.

```
mkdir <YourDirectoryName>
```

### Step2: Fastq File Renaming

Rename all fastq files to fit the following format:

```
<name>_R[1|2]_001.fastq[.gz]
```

*Example Filenames:*

```
example1_R1_001.fastq
example1_R2_001.fastq
example2_R1_001.fastq.gz
example2_R2_001.fastq.gz
```

*Example shell script code for renaming fastq files obtained from SRATools fastq-dump:*

If getting split paired end files using SRA Tools fastqdump where the file format will be “SRAID_1.fastq”, the following script will rename all the files in the directory to the correct format:

```
for file in /PathToYourFastQDirectory/*
do
     NEWFILE="${file/_/_R}"
     NEWFILE="${NEWFILE/.fastq/_001.fastq}"
     mv "$file" $NEWFILE
done
```

### Step3: Cloning the ATAC-seq pipeline

Clone the pipeline using HTTPS into the fastq directory:

```
cd /PathToYourFastQDirectory/
git clone https://github.com/UcarLab/ATAC-seq.git
```

Once the repository is cloned, move all contents cloned in the “ATAC-seq” directory to the fastq directory

```
mv /PathToYourFastQDirectory/ATAC-seq/* /PathToYourFastQDirectory/
```

### Step 4: Installing local python libraries

The ATAC-seq pipeline is currently configured to use the “python/2.7.11” module. This module is for running the pyadapter_trim.py python script, which has the following dependencies:

```
Levenshtein
```

```
Bio
```

Currently the “python/2.7.11” module does not have the Levenshtein library installed. There are a few options for installing this library:

1) Ask IT to install this library for the “python/2.7.11” module
2) Build and locally install the library

*Example local python library installation: Levenshtein library*

1) Download the library: https://pypi.org/project/python-Levenshtein/
2) Unzip the downloaded file:

```
tar -xvzf <PathToDownloadedFile>
```

3) Load the python module:

```
module load python/2.7.11
```

4) Build the library:

```
cd <pathToUnzippedDirectory>
python setup.py build
```

5) Copy the built library to the “auyar” directory of the ATAC-seq pipeline:

```
cp –r ./build/lib.linux-x86_64-2.7/Levenshtein
/PathToYourFastQDirectory/auyar/
```

**Step 5: Run step 1 of the ATAC-seq pipeline**

```
./make_atac_seq_shifted_bam_1_fastqc.sh /PathToYourFastQDirectory/
```

Note: Be sure to include trailing backslash(/).

**Step 6: Run Step 2 of the pipeline**

```
./make_atac_seq_shifted_bam_2_trimmomatic.sh /PathToYourFastQDirectory/
```

Note: Be sure to include trailing backslash(/).

The next steps will use the fastq files found within the “adapterTrimmed” directory:

```
/PathToYourFastQDirectory/working/trimmomatic/adapterTrimmed
```

*Files that can be removed:*

If the trimmomatic step is successful, the fastq files found within the “trimmomatic” directory can be removed:

```
/PathToYourFastQDirectory/working/trimmomatic/
```

If the fastq files were downloaded from a public database, such as SRA from ncbi, these files can be removed given that the trimmomatic code was successful.

```
/PathToYourFastQDirectory/
```

**Step 7: Run step 3 of the pipeline**

```
./make_atac_seq_shifted_bam_3_bwa.sh
```

Step 3 of the pipeline aligns the reads to the reference genome. These alignments are contained in SAM files which can be found in the “bwa” directory:

```
/PathToYourFastQDirectory/working/trimmomatic/adapterTrimmed/bwa
```

*Files that can be removed:*

All fastq files in the “adapterTrimmed” directory can now be safely removed.

```
/PathToYourFastQDirectory/working/trimmomatic/adapterTrimmed/
```

**Step 8: Run step 4 of the pipeline**

```
./make_atac_seq_shifted_bam_4_shift_sam.sh
```

Step4 will generate sorted BAM files and sorted shifted BAM files.

*Files that can be removed:*

All SAM files can now be removed from the “bwa” directory:

```
/PathToYourFastQDirectory/working/trimmomatic/adapterTrimmed/bwa
```

**Notable Output Files:**

*Alignments:*

```
/PathToYourFastQDirectory/working/trimmomatic/adapterTrimmed/bwa/*.bam
/PathToYourFastQDirectory/working/trimmomatic/adapterTrimmed/bwa/*_shifted.bam
```

*Quality Control Files:*

```
/PathToYourFastQDirectory/working/fastQC/*_fastqc.html
/PathToYourFastQDirectory/working/fastQC/*_fastqc.zip
```

*Alignment Summary Files:*

```
/PathToYourFastQDirectory/working/trimmomatic/adapterTrimmed/bwa/*_insertSize.pdf
/PathToYourFastQDirectory/working/trimmomatic/adapterTrimmed/bwa/*_insertSize.txt
/PathToYourFastQDirectory/working/trimmomatic/adapterTrimmed/bwa/*_metrics.txt
```