FASTQ to BAM / BAM to FASTQ

To convert FASTQ files to aligned BAM files

1. **Install Conda via miniconda**: https://waylonwalker.com/install-miniconda/ / https://dev.to/waylonwalker/installing-miniconda-on-linux-from-the-command-line-4ad7
2. **Make a new environmen**t to work in (let's call this one snowflakes)

```
conda create --name snowflakes
```

3. **Activate your environment**

```
conda activate snowflakes
source activate snowflakes #this one always works
```

4. **Trim FASTQ files for adaptors**

By Trimmomatic:

```
conda install -c bioconda trimmomatic

#find adaptors folder installed with the software, and move to working directory

#read documentation here:

#commands for single-end and paried-end differ. Find example here: https://datacarpentry.org/wrangling-genomics/03-trimming/index.html
```

By cutAdapt:

```
conda install -c bioconda cutadapt

#For 5' adaptor
cutadapt -a ADAPTORSEQ -o output.fastq input.fastq
```

5. **Install STAR**

```
conda install -c bioconda star
STAR --help
```

6**. Download latest reference genome build** (.fasta / .fa) **and genome annotation** file (.gtf) from Encode: *Make sure to download both the files from the same source, and do not mix and match (i.e., one from UCSC and one from ENCODE do not align)*

7**. Check number of available threads**

```
lscpu
#multiply number of sockets by number of cores by number of threads
```

8**. Make directory to store index**

```
mkdir starIndex
```

9. Run command to make index from genome

```
STAR --runThreadN 10 --runMode genomeGenerate --genomeDir STARIndex --genomeFastaFiles GRCh38_latest_genomic.fna --sjdbGTFfile gencode.v
38.annotation.gtf

# --genomeDir = directory to store indexed files
# --genomeFastaFiles = reference genome fasta file
# --sjdbGTFfile = GTF annotation file --> optional, but useful
```

10. Align trimmed fastq files

```
STAR --runThreadN 30 --genomeDir STARIndex --readFilesIn trimmed_fastq_file.fastq --outFilterIntronMotifs RemoveNoncanonical --outFileNa
mePrefix mapped_ --outSAMtype BAM SortedByCoordinate

# --genomeDir = directory where indexed files are stored --> different from above!
# --readFilesIn = trimmed fastq files
# --outFilterIntronMotifs RemoveNoncanonical = Removing any SNV variations
# --outFileNamePrefix = prefix to add to mapped files
# --outSAMtype BAM = SAM or BAM output file (here BAM and SortedByCoordinate)
```

To convert aligned BAM files back to FASTQ

1. Install `samtools` : http://www.sthda.com/english/wiki/install-samtools-on-unix-system
2. Run the following:
    a. For one file

```
samtools fastq bamfile.bam > name_of_fastq_file.fastq
```

    b. For multiple files (processing one after the other)

```
for file in *.bam; do samtools fastq $file > ${file//.bam/.fastq}; done
```