# How to call genotype likelihood using Next Generation Sequencing data?

# Here we will be using bowtie + samtools to align and call genotype.

# If you have multiple files, I highly recommend running one file first. If the file works, then batch run all your files.

## Quality Control:

First, let us check the quality of the sequencing result using fastQC.

Let’s say your file is located in <input\_directory>. To make our life easier, let’s create an atlas.

*in=<input\_directory>*

*out=<output\_directory>*

To run fastQC, first let us load the module (I’m on a cluster, but you can also download fastQC from [here](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

*module load fastqc/0.11.2*

To run, type:

*fastqc $in/<input\_file> --outdir=$out*

Now we will have our results in the output directory. The document from [here](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/Help/3%20Analysis%20Modules/) will help you understand the result.

If you see adaptor contamination especially towards the end. This is because if you template is not long enough, the adapter sequences will be sequences right after the template ends. If this is the case, we will need to cut the adatpters with cutadapt.

If you don’t have cutadapt, install using:

*pip install cutadapt*

To trim illumina single-ended indexed libraries:

*cutadapt -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNATCTCGTATGCCGTCTTCTGCTTG -o trimmed.1.fastq reads.1.fastq > -o trimmed.1.log*

# note to self: Rui’s mmPCR adapters has an additional ‘A’ at the beginning. So the read will be *AGATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNATCTCGTATGCCGTCTTCTGCTTG*

To run cutadapt on all files in a folder, use the script **run\_cutadapt.sh.**

If you don’t understand why adapters need to be trimmed, look [here](http://www.med.unc.edu/pharm/calabreselab/files/tufts-sequencing-primer).