```
### Installation of tools
sh install_tools.sh
```

```
##DON'T RUN BELOW CODE if you already installed BWA, freebayes, bedtools

#creat bin and source directories
mkdir  ~/work/bin
mkdir  ~/work/src

##Installation
cd  ~/work/src
wget https://downloads.sourceforge.net/project/bio-bwa/bwa-0.7.16a.tar.bz2
tar -xvf  bwa-0.7.16a.tar.bz2
cd bwa-0.7.16a
make
cp bwa  ~/work/bin/

cd  ~/work/src
git clone --recursive git://github.com/ekg/freebayes.git
cd freebayes
make
cp bin/freebayes  ~/work/bin/

cd  ~/work/src
wget https://github.com/arq5x/bedtools2/releases/download/v2.26.0/bedtools-
2.26.0.tar.gz
tar -zxvf bedtools-2.26.0.tar.gz
cd bedtools2
make
cp bin/*  ~/work/bin/

cd  ~/work/

###If samtools is not installed
cd  ~/work/src
wget https://github.com/samtools/samtools/releases/download/1.5/samtools-1.5.tar.bz2
tar -xvf  samtools-1.5.tar.bz2
cd samtools-1.5
pwd
##copy the path and assign it as prefix
./configure --prefix=/storage/home/rxv923/work/src/samtools-1.5
make
make install
cp bin/samtools ~/work/bin/
```

```
######Once you finished installation set environment variables

###NOTE: Be extra cautious when you are editing this file.
### You are loading the binaries to your unix environment by default
vim ~/.bashrc
##Press Insert and paste the below line at end of file
export PATH="/storage/home/rxv923/work/bin:$PATH"
## Press Esc and :wq
source ~/.bashrc
```

```
#Download the dataset from github. We provided a BAM file and would like to convert
#it into FASTQ for this session.

git clone https://github.com/PSUGenomix/BG_retreat_workshop_2017.git

cd  BG_retreat_workshop_2017

###Convert BAM to FASTQ
bedtools bamtofastq -i Example_bamFile_singleEnd_reads_45bp_sg11.bam \
-fq yeast.end1.fq

### Workflow

#### Index
##Generate index for the reference genome. This is a one time step and the index can
##be reused.
bwa index sg11_all_chromosomes.fa
#bwa index -a bwtsw ref.fa (specific parameters for large genomes > 2GB)

#### Alignment
bwa mem sg11_all_chromosomes.fa yeast.end1.fq \
      -R '@RG\tID:BGR2017\tSM:yeast1' > yeast_aln.sam
###When you copy paste this line retype the single quotes (') in -R option

#### SAM to BAM
samtools view -Sbh  yeast_aln.sam > yeast_aln.bam

#### Sorting BAM
samtools sort yeast_aln.bam  yeast_aln_sorted

#### Index BAM
samtools index yeast_aln_sorted.bam

##The data set we are handling is single end, in case of paired end we perform an
##extra step to remove ##PCR duplicates. This step is optional and depends on type of
##analysis.
#samtools rmdup yeast_aln_sorted.bam yeast_aln_rmdup.bam

#### Variant call
freebayes -f sg11_all_chromosomes.fa \
      yeast_aln_sorted.bam >yeast_variants.vcf

##### Analysis

#### Bedtools intersect

##Intersection Genes with mutation sites
bedtools intersect -a sg11_reference_gene_annotation.gff  \
                   -b yeast_variants.vcf -wa > genes_with_mutations.bed

##gene annotation
head sg11_reference_gene_annotation.gff

##mutation calls
tail  yeast_variants.vcf

##Different ways to view the output
bedtools intersect -a sg11_reference_gene_annotation.gff \
                   -b yeast_variants.vcf  | head
```

```
##Returns original mutation call (as in vcf file) next to the intersect output in
each line
bedtools intersect -a sg11_reference_gene_annotation.gff \
                   -b yeast_variants.vcf -wb | head


##Find genes without SNV
bedtools intersect -a sg11_reference_gene_annotation.gff \
                   -b yeast_variants.vcf -v > genes_without_mutations.bed

##Genes with highest number of SNV
bedtools intersect -a sg11_reference_gene_annotation.gff \
           -b yeast_variants.vcf | cut -f 9 | sort| uniq -c | sort -rn | head

#### Bedtools window

##Look for mutations in the flanking regions of the gene and the gene included
bedtools window -w 500 -a sg11_reference_gene_annotation.gff \
                       -b yeast_variants.vcf > mutations_gene_plus_500bpwindow.bed

##Subtract the gene regions to leave with only the mutations in the flanking regions
bedtools intersect -a mutations_gene_plus_500bpwindow.bed \
      -b sg11_reference_gene_annotation.gff -v >  mutations_within_500bpwindow.bed

## -w looks at 500 bp upstream and downstream of the gene
bedtools window -w 500 -a sg11_reference_gene_annotation.gff \
       -b yeast_variants.vcf | head

## if we want to set unequal windows then -l for upstream and -r for downstream
should be set accordingly
bedtools window -l 500 -r 100  -a sg11_reference_gene_annotation.gff \
       -b yeast_variants.vcf | head

#### Bedtools coverage
bedtools coverage  -a sg11_reference_gene_annotation.gff \
       -b yeast_aln_sorted.bam > gene_coverage.bed

#### Bedtools getfasta
cat gene_extract.bed
chr1  1807  2169  YAL068C

bedtools getfasta -fi sg11_all_chromosomes.fa \
                  -bed gene_extract.bed -fo YAL068C.fa -name
```