# iVar Instructions

# iVar Instructions

Link to the full iVar manual

**Standard Analysis:**

Most commands given below are examples, and may need to be modified given the specific iVar installation directory and path of the files. Examples given correspond to paths on the lab server in the V8 lab.

**Step 0) Activate the iVar conda environment**

- In the terminal, type the following command:

```
conda activate ivar
```

- Your terminal prompt should now show `(ivar)` at the beginning of the line.

**Step 1) Create the iVar sample sheet**

- Navigate to the ivar scripts folder in the terminal.
  `cd /data/ivar/scripts`
- Edit the “generate_sample_sheet.sh" file. Switch the path in line 3 to the data folder containing the fastq.gz files from the sequencing run. For example:
  `find /data/analysis/20191212_run -name "*.fastq.gz"`

**Note:** If you are analyzing fastq files, switch `”*.fastq.gz"` to `"*.fastq"`. Also, the script will list all files of the fastq type in the specified folder. If there are index read fastq files present in the same folder, it is easiest to copy the sample fastq files into a new folder, and point the script to the new folder.

- Save the file, and navigate to the pipeline folder in the terminal. Run the script as shown below:
  `cd /data/ivar/pipeline`
  `bash ../scripts/generate_sample_sheet.sh`

- Open the “samples.tsv” file in the pipeline folder that the script generated and verify that it is correct. Samples that have been sequenced more than once should be indicated with the same sample ID, but consecutive library numbers. For example, BEN42128_L1 and BEN42128_L2.

**Step 2) Setup the config.json in the ivar/pipeline folder:**

- Open the config.json file in a text editor
- Make sure that the path to the reference fasta file and the bed file containing the primer positions are correct. Change the “out_dir” path to the desired output folder.

```
{
  "reference": "/data/ivar/db/DRC-2018-MAN-REF.fa",
  "bed_file": "/data/ivar/db/drc_primers.bed",
  "out_dir": "/data/analysis/OUTPUT_FOLDER",
    "consensus_threshold": "50",
  "samples_path": "samples.tsv",
  "library_delimiter": "_"
}
```

**Step 3) Run the pipeline!**

- In the terminal, make sure that you are still in the pipeline folder.
- [OPTIONAL] Run the following command to do a test run of the pipeline `snakemake -n`

The output should indicate every job in the Snakefile (except 'all') is triggered as many times as there are samples in the folder. This is an example output for 32 samples:

```
Job counts:
    count   jobs
    32  align_reads
    1   all
    32  call_consensus
    32  merge_multiple_libraries
```

```
    32  trim_reads
    129
This was a dry-run (flag -n). The order of jobs does not reflect the
order of execution.
```

- When you are ready, run the pipeline using the following command. Substitute the number of available CPU cores in the command to speed up the data analysis. You can monitor the progress in a second terminal window using `htop`.

```
snakemake --cores 40
```

- Enjoy the show! Snakemake will notify you once all jobs are finished. All files will be deposited in the output folder.

## Setup of new reference and primer bed files

### Step 1) Setup of a new fasta reference file

- Copy the new reference fasta file into the db folder (`/data/ivar/db/`), navigate to the db folder in the terminal, and run the following command in the terminal:

```
bwa index [name of file].fasta
```

- bwa will index the reference fasta file and create the index files in the same folder. For example, for `DRC-2018-MAN-REF.fa` the following files are created:

```
DRC-2018-MAN-REF.fa
DRC-2018-MAN-REF.fa.sa
DRC-2018-MAN-REF.fa.pac
DRC-2018-MAN-REF.fa.fai
DRC-2018-MAN-REF.fa.bwt
DRC-2018-MAN-REF.fa.ann
DRC-2018-MAN-REF.fa.amb
```

**IMPORTANT:** If you change the reference fasta file, you will also have to change the corresponding primer bed file, which lists the locations of the amplicon primers in relation to the reference fasta file.

**Step 2) Generate a new primer bed file**

**Note:** If you only updated the reference fasta, but will keep using the same amplicon primers, you do NOT need to supply a new primer fasta file. In that case, skip the below step and continue with the previous primer fasta file (the default is `drc_primers.fa`).

- [Optional] Copy a fasta file containing all new primer sequences to the db folder (`/data/ivar/db/`).
- Navigate to the ivar scripts folder in the terminal.
  `cd /data/ivar/scripts`
- Edit the “generate_sample_sheet.sh" file. Switch the paths in line 1 to the (bwa indexed, see step 1) reference fasta file, as well as a fasta file containing all amplicon primer sequences (the default is `drc_primers.fa`).

`bwa mem -k 5 -T 16 db/REFERENCE.fa db/NEW-PRIMERS.fa`

- Save the file, and run the script as shown below:

`bash generate_bed_file.sh`

- The script will create the necessary .bed and .bam files in the `db` folder. For example using `drc_primers.fa`, will generate the following files:

```
/drc_primers.bam
/drc_primers.bed
/drc_primers.fa
```

**Step 3) Update the config.json file**

- Remember to update the reference fasta and bed files in the config.json before running iVar again.

```
{
  "reference": "/data/ivar/db/NEW-REFERENCE.fa",
  "bed_file": "/data/ivar/db/NEW-PRIMERS.bed",
  "out_dir": "/data/analysis/20191212_run_output",
  "consensus_threshold": "50",
  "samples_path": "samples.tsv",
  "library_delimiter": "_"
}
```