Salmon

(https://combine-lab.github.io/salmon/getting_started/)

```
$ conda activate salmon
```

Obtain transcriptome:

```
$ mkdir salmon_example
$ cd salmon_example/
Curl
ftp://ftp.ensembl.org/pub/current_fasta/homo_sapiens/cdna/Homo_sapien
s.GRCh38.cdna.all.fa.gz -o homo.fa.gz
```

Build an index on transcriptome:

```
$ salmon index -t homo.fa.gz -i homo_index
```

Job file:

```
#!/bin/bash

#SBATCH --nodes=1
#SBATCH --ntasks=1
#SBATCH --job-name=Salmon1.job
#SBATCH --cpus-per-task=8
#SBATCH --mem=32GB
#SBATCH --time=48:00:00
#SBATCH --account=mangul_341
#SBATCH --mail-type=all
#SBATCH --mail-user=your email

module purge
module load gcc/11.2.0
module load conda
eval "$(conda shell.bash hook)"
conda activate salmon

salmon quant -i homo_index -l A \
         -1 filename_R1.fastq.gz \
         -2 filename_R2.fastq.gz \
         -p 8 --validateMappings -o quants/filename_quant
```

If the job failed with the error message similar to this, use binary instead of installing it with conda

```
/var/spool/slurm/d/job10650478/slurm_script: line 22: 10888
Segmentation fault      salmon quant -i homo_index -l A -1
RS_2_SW_R1.fastq.gz -2 RS_2_SW_R2.fastq.gz -p 8 --validateMappings -o
quants/RS_2_SW_quant.
```

Download Salmon binary from
https://github.com/COMBINE-lab/salmon/releases

Decompress it

```
tar xzvf salmon-1.9.0_linux_x86_64.tar.gz
```

Job file:

```
#!/bin/bash

#SBATCH --nodes=1
#SBATCH --ntasks=1
#SBATCH --job-name=Salmon1.job
#SBATCH --cpus-per-task=8
#SBATCH --mem=32GB
#SBATCH --time=48:00:00
#SBATCH --account=mangul_341
#SBATCH --mail-type=all
#SBATCH --mail-user=your email

module purge
module load gcc/11.2.0

/project/mangul_341/keruipen/tools/salmon-1.9.0_linux_x86_64/bin/salm
on quant -i homo_index -l A \
         -1 RS_2_SW_R1.fastq.gz \
         -2 RS_2_SW_R2.fastq.gz \
         -p 8 --validateMappings -o quants/RS_2_SW_quant
```

GEDIT

Get .sf files from Salmon, only “Name” and “TPM” columns are needed for the following steps
Manual change column “TPM” to the sample name

Combine all the files

```
Paste *.sf > combined.tsv
```

Generate a file that has “Name” as the first column and TPM from each sample as the following columns (the example has 19 samples in total)

```
Cut -f 1,4,9,14,19,24,29,34,39,44,49,54,59,64,69,74,79,84,89,94
combined.tsv > TPMonly.tsv
```

Download the gene symbol conversion from
http://uswest.ensembl.org/biomart/martview/2208876ea2ad95c5ed4765793cc2f080
Select "Ensembl Genes 107", "Human genes (GRCh38.p13)"
Under "Attributes", select "Transcript stable ID" and "Gene name"
Hit "Results", Export the tsv file (mart_export.txt)

Download GEDIT from GitHub
https://github.com/BNadel/GEDIT

Create the following python script to convert gene symbols, name this script "GeneSymbolConversion.py"

```
from sys import *
EnsToGS = {}

for line in open(argv[1], "r"):
  parts = line.strip().split("\t")
  if len(parts) < 2:
      continue
  EnsToGS[parts[0]] = parts[1]

first = True
for line in open(argv[2], "r"):
  if first:
    print line.strip()
    first = False
  parts = line.strip().split("\t")
  StrippedID = parts[0].split(".")[0]
  if StrippedID in EnsToGS:
    print EnsToGS[StrippedID] +"\t" + "\t".join(parts[1:])
```

Convert the gene symbol [Note that it used python2]
```
python2 GeneSymbolConversion.py mart_export.txt TPMonly.tsv > converted.tsv
```

Run GEDITv2.0 (Caution: it is written in python2)
Download Python 2.7.18 from
https://www.python.org/downloads/release/python-2718/

Run the following commands to install packages in python2
```
sudo pip install statistics
pip install numpy
pip install scipy
```

Use HPCA_Recommended.csv or SkinSignaturesV1.0.tsv as references

```
python2 GEDIT2.py -mix converted.tsv  -ref
ReferenceMatrices/HPCA_Recommended.csv

python2 GEDIT2.py -mix converted.tsv  -ref
ReferenceMatrices/SkinSignaturesV1.0.tsv

After using the tool, consider upgrading via the 'pip install
--upgrade pip' command
```