

```

#####
# Code to accompany Ng et al
# Author: Charlotte KY Ng
# Last updated: 21 Feb 2014
#####

# This script will execute all the simulation and classification
# performed in the manuscript

# All the scripts required to run these analyses can be found
# https://dl.dropboxusercontent.com/u/15115364/Ng%20et%20al.zip
# A copy of ComBAT is included

# R libraries CMA, superpc, ROCR are required
# The second part of the script calls a perl script, which requires
# Acme::Tools perl library.

# PART 1 - simulation
# This script has 6 main parts
# 1.0: download the data from http://compbio.dfci.harvard.edu/pubs/sbtpaper/data.zip
# 1.1: ComBAT
# 1.2: spiking in and classification
# 1.3: read the results in batches
# 1.4: merge the batches
# 1.5: generate plots and tables

# WARNING:
# Running this code as is, without parallelisation will take a very long time
# Parallelising the R commands in parts 2 and 3 is highly recommended
# Running this code in its entirety will also require significant
# amount of disk space.

# PART 2 - breast cancer classification

# This script has 2 main parts
# 2.0: pre-processing the data obtained from GSE25055 and GSE25065
# 2.1: running the classification

#####
#####
##### CODE BEGINS HERE #####
#####

##### Part 1.0: download the required datasets #####
#####

mkdir -p data
cd data

wget http://compbio.dfci.harvard.edu/pubs/sbtpaper/data.zip
unzip data.zip

cd ..

#####
##### Part 1.1: pre-process data in ComBAT #####
#####

R CMD BATCH --no-save --no-restore Rscripts/Run_Combat.R Rscripts/Run_Combat.Rout

#####
##### Part 1.2: do the main processing #####
#####

# create output and temp directories
mkdir -p dldaCMA
mkdir -p superPC

```

```

mkdir -p temp

### the for-loops set random seed

### equalprop spike_value=0.5, 100 iterations

for i in `seq 1101 2 1199`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=0.5" Rscripts/ConvergentPheno_perturb.R Rscripts/ConvergentPheno_0.5_${i}.Rout
done

for i in `seq 2101 2 2199`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=0.5" Rscripts/ConvergentPheno_perturb.R Rscripts/ConvergentPheno_0.5_${i}.Rout
done

### equalprop spike_value=1.0, 100 iterations

for i in `seq 1201 2 1299`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.0" Rscripts/ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.0_${i}.Rout
done

for i in `seq 2201 2 2299`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.0" Rscripts/ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.0_${i}.Rout
done

### equalprop spike_value=1.5, 100 iterations

for i in `seq 1301 2 1399`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.5" Rscripts/ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.5_${i}.Rout
done

for i in `seq 2301 2 2399`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.5" Rscripts/ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.5_${i}.Rout
done

### equalprop spike_value=0.5, 100 iterations, superPC

for i in `seq 1101 2 1199`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"superpc\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=0.5" Rscripts/ConvergentPheno_perturb_superpc.R Rscripts/
ConvergentPheno_superpc_0.5_${i}.Rout
done

for i in `seq 2101 2 2199`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"superpc\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=0.5" Rscripts/ConvergentPheno_perturb_superpc.R Rscripts/
ConvergentPheno_superpc_0.5_${i}.Rout
done

### equalprop spike_value=1.0, 100 iterations, superPC

for i in `seq 1201 2 1299`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"superpc\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.0" Rscripts/ConvergentPheno_perturb_superpc.R Rscripts/
ConvergentPheno_superpc_1.0_${i}.Rout
done

for i in `seq 2201 2 2299`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"superpc\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.0" Rscripts/ConvergentPheno_perturb_superpc.R Rscripts/
ConvergentPheno_superpc_1.0_${i}.Rout
done

### equalprop spike_value=1.5, 100 iterations, superPC

for i in `seq 1301 2 1399`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"superpc\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.5" Rscripts/ConvergentPheno_perturb_superpc.R Rscripts/
ConvergentPheno_superpc_1.5_${i}.Rout
done

```

```

for i in `seq 2301 2 2399`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"superpc\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.5" Rscripts/ConvergentPheno_perturb_superpc.R Rscripts/
ConvergentPheno_superpc_1.5_${i}.Rout
done

##### unequal proportions, spike_value = 0.5, 200 iterations

for i in `seq 1401 2 1799`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=0.5 nmechs.vector=2:5 randommechs=T" Rscripts/ConvergentPheno_perturb.R
Rscripts/ConvergentPheno_0.5_randommechs_${i}.Rout
done

##### unequal proportions, spike_value = 1.0, 200 iterations

for i in `seq 2401 2 2799`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.0 nmechs.vector=2:5 randommechs=T" Rscripts/ConvergentPheno_perturb.R
Rscripts/ConvergentPheno_1.0_randommechs_${i}.Rout
done

##### unequal proportions, spike_value = 1.5, 200 iterations

for i in `seq 3401 2 3799`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.5 nmechs.vector=2:5 randommechs=T" Rscripts/ConvergentPheno_perturb.R
Rscripts/ConvergentPheno_1.5_randommechs_${i}.Rout
done

##### random training/test, spike_value=0.5, 200 iterations

for i in `seq 11401 2 11799`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=0.5 nmechs.vector=2:5 randommechs=T randomvalidation=T" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_0.5_randomvalidation_${i}.Rout
done

## random training/test, spike_value=1.0, 200 iterations

for i in `seq 12401 2 12799`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.0 nmechs.vector=2:5 randommechs=T randomvalidation=T" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.0_randomvalidation_${i}.Rout
done

## random training/test, spike_value=1.5, 200 iterations

for i in `seq 13401 2 13799`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.5 nmechs.vector=2:5 randommechs=T randomvalidation=T" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.5_randomvalidation_${i}.Rout
done

##### overlapping signatures, spike_value=0.5, overlap=0, 100 iterations

for i in `seq 7301 2 7499`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=0.5 nmechs.vector=2:5 overlappingsigs.vector=0" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_0.5_overlaps_${i}.Rout
done

##### overlapping signatures, spike_value=0.5, overlap=0.01, 100 iterations

for i in `seq 4301 2 4499`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=0.5 nmechs.vector=2:5 overlappingsigs.vector=0.01" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_0.5_overlaps_${i}.Rout
done

##### overlapping signatures, spike_value=0.5, overlap=0.05, 100 iterations

for i in `seq 5301 2 5499`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=0.5 nmechs.vector=2:5 overlappingsigs.vector=0.05" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_0.5_overlaps_${i}.Rout
done

```

done

#### overlapping signatures, spike\_value=0.5, overlap=0.1, 100 iterations

```
for i in `seq 6301 2 6499`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=0.5 nmechs.vector=2:5 overlappingsigs.vector=0.1" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_0.5_overlaps_${i}.Rout
done
```

#### overlapping signatures, spike\_value=0.5, overlap=0.2, 100 iterations

```
for i in `seq 8301 2 8499`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=0.5 nmechs.vector=2:5 overlappingsigs.vector=0.2" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_0.5_overlaps_${i}.Rout
done
```

#### overlapping signatures, spike\_value=0.5, overlap=0.5, 100 iterations

```
for i in `seq 9301 2 9499`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=0.5 nmechs.vector=2:5 overlappingsigs.vector=0.5" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_0.5_overlaps_${i}.Rout
done
```

#### overlapping signatures, spike\_value=0.5, overlap=0.9, 100 iterations

```
for i in `seq 10301 2 10499`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=0.5 nmechs.vector=2:5 overlappingsigs.vector=0.9" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_0.5_overlaps_${i}.Rout
done
```

#### overlapping signatures, spike\_value=1.0, overlap=0, 100 iterations

```
for i in `seq 7501 2 7699`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.0 nmechs.vector=2:5 overlappingsigs.vector=0" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.0_overlaps_${i}.Rout
done
```

#### overlapping signatures, spike\_value=1.0, overlap=0.01, 100 iterations

```
for i in `seq 4501 2 4699`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.0 nmechs.vector=2:5 overlappingsigs.vector=0.01" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.0_overlaps_${i}.Rout
done
```

#### overlapping signatures, spike\_value=1.0, overlap=0.05, 100 iterations

```
for i in `seq 5501 2 5699`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.0 nmechs.vector=2:5 overlappingsigs.vector=0.05" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.0_overlaps_${i}.Rout
done
```

#### overlapping signatures, spike\_value=1.0, overlap=0.1, 100 iterations

```
for i in `seq 6501 2 6699`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.0 nmechs.vector=2:5 overlappingsigs.vector=0.1" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.0_overlaps_${i}.Rout
done
```

#### overlapping signatures, spike\_value=1.0, overlap=0.2, 100 iterations

```
for i in `seq 8501 2 8699`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.0 nmechs.vector=2:5 overlappingsigs.vector=0.2" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.0_overlaps_${i}.Rout
done
```

#### overlapping signatures, spike\_value=1.0, overlap=0.5, 100 iterations

```

for i in `seq 9501 2 9699`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.0 nmechs.vector=2:5 overlappingsigs.vector=0.5" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.0_overlaps_${i}.Rout
done

#### overlapping signatures, spike_value=1.0, overlap=0.9, 100 iterations

for i in `seq 10501 2 10699`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.0 nmechs.vector=2:5 overlappingsigs.vector=0.9" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.0_overlaps_${i}.Rout
done

#### overlapping signatures, spike_value=1.5, overlap=0, 100 iterations

for i in `seq 7701 2 7899`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.5 nmechs.vector=2:5 overlappingsigs.vector=0" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.5_overlaps_${i}.Rout
done

#### overlapping signatures, spike_value=1.5, overlap=0.01, 100 iterations

for i in `seq 4701 2 4899`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.5 nmechs.vector=2:5 overlappingsigs.vector=0.01" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.5_overlaps_${i}.Rout
done

#### overlapping signatures, spike_value=1.5, overlap=0.05, 100 iterations

for i in `seq 5701 2 5899`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.5 nmechs.vector=2:5 overlappingsigs.vector=0.05" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.5_overlaps_${i}.Rout
done

#### overlapping signatures, spike_value=1.5, overlap=0.1, 100 iterations

for i in `seq 6701 2 6899`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.5 nmechs.vector=2:5 overlappingsigs.vector=0.1" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.5_overlaps_${i}.Rout
done

#### overlapping signatures, spike_value=1.5, overlap=0.2, 100 iterations

for i in `seq 8701 2 8899`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.5 nmechs.vector=2:5 overlappingsigs.vector=0.2" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.5_overlaps_${i}.Rout
done

#### overlapping signatures, spike_value=1.5, overlap=0.5, 100 iterations

for i in `seq 9701 2 9899`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.5 nmechs.vector=2:5 overlappingsigs.vector=0.5" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.5_overlaps_${i}.Rout
done

#### overlapping signatures, spike_value=1.5, overlap=0.9, 100 iterations

for i in `seq 10701 2 10899`; do
R CMD BATCH --no-save --no-restore "--args outdir=\"dldaCMA\" tempdir=\"temp\" my.mccvrep=50 s=
$i spike_value=1.5 nmechs.vector=2:5 overlappingsigs.vector=0.9" Rscripts/
ConvergentPheno_perturb.R Rscripts/ConvergentPheno_1.5_overlaps_${i}.Rout
done

#####
##### Part 1.3: Read the results in batches #####
#####

### move the resultant files to their subdirectories
### this is a bit convoluted but as the data files are very

```

```
### large, having too many in the same directory makes the file
### system very unresponsive.
```

```
cd dldaCMA
```

```
mkdir -p equalprop_0.5_01 equalprop_0.5_02 equalprop_0.5_03 equalprop_0.5_04
mkdir -p equalprop_1.0_01 equalprop_1.0_02 equalprop_1.0_03 equalprop_1.0_04
mkdir -p equalprop_1.5_01 equalprop_1.5_02 equalprop_1.5_03 equalprop_1.5_04
```

```
mv ConvergentPheno.0.5*.11[01234][1234567890].*RData equalprop_0.5_01
mv ConvergentPheno.0.5*.11[56789][1234567890].*RData equalprop_0.5_02
mv ConvergentPheno.0.5*.21[01234][1234567890].*RData equalprop_0.5_03
mv ConvergentPheno.0.5*.21[56789][1234567890].*RData equalprop_0.5_04
```

```
mv ConvergentPheno.1*.12[01234][1234567890].*RData equalprop_1.0_01
mv ConvergentPheno.1*.12[56789][1234567890].*RData equalprop_1.0_02
mv ConvergentPheno.1*.22[01234][1234567890].*RData equalprop_1.0_03
mv ConvergentPheno.1*.22[56789][1234567890].*RData equalprop_1.0_04
```

```
mv ConvergentPheno.1.5*.13[01234][1234567890].*RData equalprop_1.5_01
mv ConvergentPheno.1.5*.13[56789][1234567890].*RData equalprop_1.5_02
mv ConvergentPheno.1.5*.23[01234][1234567890].*RData equalprop_1.5_03
mv ConvergentPheno.1.5*.23[56789][1234567890].*RData equalprop_1.5_04
```

```
for i in 01 02 03 04 05 06 07 08; do mkdir -p randommechs_0.5_${i}; done
for i in 01 02 03 04 05 06 07 08; do mkdir -p randommechs_1.0_${i}; done
for i in 01 02 03 04 05 06 07 08; do mkdir -p randommechs_1.5_${i}; done
```

```
mv ConvergentPheno.0.5*.14[01234][1234567890].*RData randommechs_0.5_01
mv ConvergentPheno.0.5*.14[56789][1234567890].*RData randommechs_0.5_02
mv ConvergentPheno.0.5*.15[01234][1234567890].*RData randommechs_0.5_03
mv ConvergentPheno.0.5*.15[56789][1234567890].*RData randommechs_0.5_04
mv ConvergentPheno.0.5*.16[01234][1234567890].*RData randommechs_0.5_05
mv ConvergentPheno.0.5*.16[56789][1234567890].*RData randommechs_0.5_06
mv ConvergentPheno.0.5*.17[01234][1234567890].*RData randommechs_0.5_07
mv ConvergentPheno.0.5*.17[56789][1234567890].*RData randommechs_0.5_08
```

```
mv ConvergentPheno.1*.24[01234][1234567890].*RData randommechs_1.0_01
mv ConvergentPheno.1*.24[56789][1234567890].*RData randommechs_1.0_02
mv ConvergentPheno.1*.25[01234][1234567890].*RData randommechs_1.0_03
mv ConvergentPheno.1*.25[56789][1234567890].*RData randommechs_1.0_04
mv ConvergentPheno.1*.26[01234][1234567890].*RData randommechs_1.0_05
mv ConvergentPheno.1*.26[56789][1234567890].*RData randommechs_1.0_06
mv ConvergentPheno.1*.27[01234][1234567890].*RData randommechs_1.0_07
mv ConvergentPheno.1*.27[56789][1234567890].*RData randommechs_1.0_08
```

```
mv ConvergentPheno.1.5*.34[01234][1234567890].*RData randommechs_1.5_01
mv ConvergentPheno.1.5*.34[56789][1234567890].*RData randommechs_1.5_02
mv ConvergentPheno.1.5*.35[01234][1234567890].*RData randommechs_1.5_03
mv ConvergentPheno.1.5*.35[56789][1234567890].*RData randommechs_1.5_04
mv ConvergentPheno.1.5*.36[01234][1234567890].*RData randommechs_1.5_05
mv ConvergentPheno.1.5*.36[56789][1234567890].*RData randommechs_1.5_06
mv ConvergentPheno.1.5*.37[01234][1234567890].*RData randommechs_1.5_07
mv ConvergentPheno.1.5*.37[56789][1234567890].*RData randommechs_1.5_08
```

```
for i in 01 02 03 04 05 06 07 08; do mkdir -p randomvalidation_0.5_${i}; done
for i in 01 02 03 04 05 06 07 08; do mkdir -p randomvalidation_1.0_${i}; done
for i in 01 02 03 04 05 06 07 08; do mkdir -p randomvalidation_1.5_${i}; done
```

```
mv ConvergentPheno.0.5*.114[01234][1234567890].*RData randomvalidation_0.5_01
mv ConvergentPheno.0.5*.114[56789][1234567890].*RData randomvalidation_0.5_02
mv ConvergentPheno.0.5*.115[01234][1234567890].*RData randomvalidation_0.5_03
mv ConvergentPheno.0.5*.115[56789][1234567890].*RData randomvalidation_0.5_04
mv ConvergentPheno.0.5*.116[01234][1234567890].*RData randomvalidation_0.5_05
mv ConvergentPheno.0.5*.116[56789][1234567890].*RData randomvalidation_0.5_06
mv ConvergentPheno.0.5*.117[01234][1234567890].*RData randomvalidation_0.5_07
mv ConvergentPheno.0.5*.117[56789][1234567890].*RData randomvalidation_0.5_08
```

```
mv ConvergentPheno.1*.124[01234][1234567890].*RData randomvalidation_1.0_01
mv ConvergentPheno.1*.124[56789][1234567890].*RData randomvalidation_1.0_02
mv ConvergentPheno.1*.125[01234][1234567890].*RData randomvalidation_1.0_03
mv ConvergentPheno.1*.125[56789][1234567890].*RData randomvalidation_1.0_04
mv ConvergentPheno.1*.126[01234][1234567890].*RData randomvalidation_1.0_05
mv ConvergentPheno.1*.126[56789][1234567890].*RData randomvalidation_1.0_06
mv ConvergentPheno.1*.127[01234][1234567890].*RData randomvalidation_1.0_07
mv ConvergentPheno.1*.127[56789][1234567890].*RData randomvalidation_1.0_08
```

```
mv ConvergentPheno.1.5*.134[01234][1234567890].*RData randomvalidation_1.5_01
mv ConvergentPheno.1.5*.134[56789][1234567890].*RData randomvalidation_1.5_02
mv ConvergentPheno.1.5*.135[01234][1234567890].*RData randomvalidation_1.5_03
mv ConvergentPheno.1.5*.135[56789][1234567890].*RData randomvalidation_1.5_04
mv ConvergentPheno.1.5*.136[01234][1234567890].*RData randomvalidation_1.5_05
mv ConvergentPheno.1.5*.136[56789][1234567890].*RData randomvalidation_1.5_06
mv ConvergentPheno.1.5*.137[01234][1234567890].*RData randomvalidation_1.5_07
mv ConvergentPheno.1.5*.137[56789][1234567890].*RData randomvalidation_1.5_08

for i in 01 02 03 04 05 06 07 08 09 `seq 10 28`; do mkdir -p overlap_0.5_${i}; done
for i in 01 02 03 04 05 06 07 08 09 `seq 10 28`; do mkdir -p overlap_1.0_${i}; done
for i in 01 02 03 04 05 06 07 08 09 `seq 10 28`; do mkdir -p overlap_1.5_${i}; done

mv ConvergentPheno.0.5*.43[01234][1234567890].*RData overlap_0.5_01
mv ConvergentPheno.0.5*.43[56789][1234567890].*RData overlap_0.5_02
mv ConvergentPheno.0.5*.44[01234][1234567890].*RData overlap_0.5_03
mv ConvergentPheno.0.5*.44[56789][1234567890].*RData overlap_0.5_04
mv ConvergentPheno.0.5*.53[01234][1234567890].*RData overlap_0.5_05
mv ConvergentPheno.0.5*.53[56789][1234567890].*RData overlap_0.5_06
mv ConvergentPheno.0.5*.54[01234][1234567890].*RData overlap_0.5_07

mv ConvergentPheno.0.5*.54[56789][1234567890].*RData overlap_0.5_08
mv ConvergentPheno.0.5*.63[01234][1234567890].*RData overlap_0.5_09
mv ConvergentPheno.0.5*.63[56789][1234567890].*RData overlap_0.5_10
mv ConvergentPheno.0.5*.64[01234][1234567890].*RData overlap_0.5_11
mv ConvergentPheno.0.5*.64[56789][1234567890].*RData overlap_0.5_12
mv ConvergentPheno.0.5*.73[01234][1234567890].*RData overlap_0.5_13
mv ConvergentPheno.0.5*.73[56789][1234567890].*RData overlap_0.5_14

mv ConvergentPheno.0.5*.74[01234][1234567890].*RData overlap_0.5_15
mv ConvergentPheno.0.5*.74[56789][1234567890].*RData overlap_0.5_16
mv ConvergentPheno.0.5*.83[01234][1234567890].*RData overlap_0.5_17
mv ConvergentPheno.0.5*.83[56789][1234567890].*RData overlap_0.5_18
mv ConvergentPheno.0.5*.84[01234][1234567890].*RData overlap_0.5_19
mv ConvergentPheno.0.5*.84[56789][1234567890].*RData overlap_0.5_20
mv ConvergentPheno.0.5*.93[01234][1234567890].*RData overlap_0.5_21

mv ConvergentPheno.0.5*.93[56789][1234567890].*RData overlap_0.5_22
mv ConvergentPheno.0.5*.94[01234][1234567890].*RData overlap_0.5_23
mv ConvergentPheno.0.5*.94[56789][1234567890].*RData overlap_0.5_24
mv ConvergentPheno.0.5*.103[01234][1234567890].*RData overlap_0.5_25
mv ConvergentPheno.0.5*.103[56789][1234567890].*RData overlap_0.5_26
mv ConvergentPheno.0.5*.104[01234][1234567890].*RData overlap_0.5_27
mv ConvergentPheno.0.5*.104[56789][1234567890].*RData overlap_0.5_28

mv ConvergentPheno.1.*.45[01234][1234567890].*RData overlap_1.0_01
mv ConvergentPheno.1.*.45[56789][1234567890].*RData overlap_1.0_02
mv ConvergentPheno.1.*.46[01234][1234567890].*RData overlap_1.0_03
mv ConvergentPheno.1.*.46[56789][1234567890].*RData overlap_1.0_04
mv ConvergentPheno.1.*.55[01234][1234567890].*RData overlap_1.0_05
mv ConvergentPheno.1.*.55[56789][1234567890].*RData overlap_1.0_06
mv ConvergentPheno.1.*.56[01234][1234567890].*RData overlap_1.0_07

mv ConvergentPheno.1.*.56[56789][1234567890].*RData overlap_1.0_08
mv ConvergentPheno.1.*.65[01234][1234567890].*RData overlap_1.0_09
mv ConvergentPheno.1.*.65[56789][1234567890].*RData overlap_1.0_10
mv ConvergentPheno.1.*.66[01234][1234567890].*RData overlap_1.0_11
mv ConvergentPheno.1.*.66[56789][1234567890].*RData overlap_1.0_12
mv ConvergentPheno.1.*.75[01234][1234567890].*RData overlap_1.0_13
mv ConvergentPheno.1.*.75[56789][1234567890].*RData overlap_1.0_14

mv ConvergentPheno.1.*.76[01234][1234567890].*RData overlap_1.0_15
mv ConvergentPheno.1.*.76[56789][1234567890].*RData overlap_1.0_16
mv ConvergentPheno.1.*.85[01234][1234567890].*RData overlap_1.0_17
mv ConvergentPheno.1.*.85[56789][1234567890].*RData overlap_1.0_18
mv ConvergentPheno.1.*.86[01234][1234567890].*RData overlap_1.0_19
mv ConvergentPheno.1.*.86[56789][1234567890].*RData overlap_1.0_20
mv ConvergentPheno.1.*.95[01234][1234567890].*RData overlap_1.0_21

mv ConvergentPheno.1.*.95[56789][1234567890].*RData overlap_1.0_22
mv ConvergentPheno.1.*.96[01234][1234567890].*RData overlap_1.0_23
mv ConvergentPheno.1.*.96[56789][1234567890].*RData overlap_1.0_24
mv ConvergentPheno.1.*.105[01234][1234567890].*RData overlap_1.0_25
mv ConvergentPheno.1.*.105[56789][1234567890].*RData overlap_1.0_26
mv ConvergentPheno.1.*.106[01234][1234567890].*RData overlap_1.0_27
```

```

mv ConvergentPheno.1.*.106[56789][1234567890].*RData overlap_1.0_28

mv ConvergentPheno.1.5*.47[01234][1234567890].*RData overlap_1.5_01
mv ConvergentPheno.1.5*.47[56789][1234567890].*RData overlap_1.5_02
mv ConvergentPheno.1.5*.48[01234][1234567890].*RData overlap_1.5_03
mv ConvergentPheno.1.5*.48[56789][1234567890].*RData overlap_1.5_04
mv ConvergentPheno.1.5*.57[01234][1234567890].*RData overlap_1.5_05
mv ConvergentPheno.1.5*.57[56789][1234567890].*RData overlap_1.5_06
mv ConvergentPheno.1.5*.58[01234][1234567890].*RData overlap_1.5_07

mv ConvergentPheno.1.5*.58[56789][1234567890].*RData overlap_1.5_08
mv ConvergentPheno.1.5*.67[01234][1234567890].*RData overlap_1.5_09
mv ConvergentPheno.1.5*.67[56789][1234567890].*RData overlap_1.5_10
mv ConvergentPheno.1.5*.68[01234][1234567890].*RData overlap_1.5_11
mv ConvergentPheno.1.5*.68[56789][1234567890].*RData overlap_1.5_12
mv ConvergentPheno.1.5*.77[01234][1234567890].*RData overlap_1.5_13
mv ConvergentPheno.1.5*.77[56789][1234567890].*RData overlap_1.5_14

mv ConvergentPheno.1.5*.78[01234][1234567890].*RData overlap_1.5_15
mv ConvergentPheno.1.5*.78[56789][1234567890].*RData overlap_1.5_16
mv ConvergentPheno.1.5*.87[01234][1234567890].*RData overlap_1.5_17
mv ConvergentPheno.1.5*.87[56789][1234567890].*RData overlap_1.5_18
mv ConvergentPheno.1.5*.88[01234][1234567890].*RData overlap_1.5_19
mv ConvergentPheno.1.5*.88[56789][1234567890].*RData overlap_1.5_20
mv ConvergentPheno.1.5*.97[01234][1234567890].*RData overlap_1.5_21

mv ConvergentPheno.1.5*.97[56789][1234567890].*RData overlap_1.5_22
mv ConvergentPheno.1.5*.98[01234][1234567890].*RData overlap_1.5_23
mv ConvergentPheno.1.5*.98[56789][1234567890].*RData overlap_1.5_24
mv ConvergentPheno.1.5*.107[01234][1234567890].*RData overlap_1.5_25
mv ConvergentPheno.1.5*.107[56789][1234567890].*RData overlap_1.5_26
mv ConvergentPheno.1.5*.108[01234][1234567890].*RData overlap_1.5_27
mv ConvergentPheno.1.5*.108[56789][1234567890].*RData overlap_1.5_28

mkdir -p varspike_1.0_01 varspike_1.0_02 varspike_1.0_03 varspike_1.0_04
mkdir -p varspike_0.5_01 varspike_0.5_02 varspike_0.5_03 varspike_0.5_04
mkdir -p varspike_1.5_01 varspike_1.5_02 varspike_1.5_03 varspike_1.5_04

mv ConvergentPheno.var*.1[01234][1234567890].*RData varspike_1.0_01
mv ConvergentPheno.var*.1[56789][1234567890].*RData varspike_1.0_02
mv ConvergentPheno.var*.2[01234][1234567890].*RData varspike_1.0_03
mv ConvergentPheno.var*.2[56789][1234567890].*RData varspike_1.0_04

mv ConvergentPheno.var*.3[01234][1234567890].*RData varspike_0.5_01
mv ConvergentPheno.var*.3[56789][1234567890].*RData varspike_0.5_02
mv ConvergentPheno.var*.4[01234][1234567890].*RData varspike_0.5_03
mv ConvergentPheno.var*.4[56789][1234567890].*RData varspike_0.5_04

mv ConvergentPheno.var*.5[01234][1234567890].*RData varspike_1.5_01
mv ConvergentPheno.var*.5[56789][1234567890].*RData varspike_1.5_02
mv ConvergentPheno.var*.6[01234][1234567890].*RData varspike_1.5_03
mv ConvergentPheno.var*.6[56789][1234567890].*RData varspike_1.5_04

cd ..

cd superpc

mkdir -p equalprop_0.5 equalprop_1.0 equalprop_1.5

mv ConvergentPheno_superpc_0.5*RData equalprop_0.5
mv ConvergentPheno_superpc_1.5*RData equalprop_1.5
mv ConvergentPheno_superpc_1.NA*RData equalprop_1.0

cd ..

### read the simulation results in batches to create summary results files

R CMD BATCH --no-save --no-restore '--args n="equalprop_0.5_01"' Rscripts/
ConvergentPheno_read.R Rscripts/ConvergentPheno_read_equalprop_0.5_01.Rout
R CMD BATCH --no-save --no-restore '--args n="equalprop_0.5_02"' Rscripts/
ConvergentPheno_read.R Rscripts/ConvergentPheno_read_equalprop_0.5_02.Rout
R CMD BATCH --no-save --no-restore '--args n="equalprop_0.5_03"' Rscripts/
ConvergentPheno_read.R Rscripts/ConvergentPheno_read_equalprop_0.5_03.Rout
R CMD BATCH --no-save --no-restore '--args n="equalprop_0.5_04"' Rscripts/
ConvergentPheno_read.R Rscripts/ConvergentPheno_read_equalprop_0.5_04.Rout

```











```
R CMD BATCH --no-save --no-restore '--args n="varspike_0.5_03"' Rscripts/ConvergentPheno_read.R
Rscripts/ConvergentPheno_read_varspike_0.5_03.Rout
R CMD BATCH --no-save --no-restore '--args n="varspike_0.5_04"' Rscripts/ConvergentPheno_read.R
Rscripts/ConvergentPheno_read_varspike_0.5_04.Rout
```

```
R CMD BATCH --no-save --no-restore '--args n="varspike_1.0_01"' Rscripts/ConvergentPheno_read.R
Rscripts/ConvergentPheno_read_varspike_1.0_01.Rout
R CMD BATCH --no-save --no-restore '--args n="varspike_1.0_02"' Rscripts/ConvergentPheno_read.R
Rscripts/ConvergentPheno_read_varspike_1.0_02.Rout
R CMD BATCH --no-save --no-restore '--args n="varspike_1.0_03"' Rscripts/ConvergentPheno_read.R
Rscripts/ConvergentPheno_read_varspike_1.0_03.Rout
R CMD BATCH --no-save --no-restore '--args n="varspike_1.0_04"' Rscripts/ConvergentPheno_read.R
Rscripts/ConvergentPheno_read_varspike_1.0_04.Rout
```

```
R CMD BATCH --no-save --no-restore '--args n="varspike_1.5_01"' Rscripts/ConvergentPheno_read.R
Rscripts/ConvergentPheno_read_varspike_1.5_01.Rout
R CMD BATCH --no-save --no-restore '--args n="varspike_1.5_02"' Rscripts/ConvergentPheno_read.R
Rscripts/ConvergentPheno_read_varspike_1.5_02.Rout
R CMD BATCH --no-save --no-restore '--args n="varspike_1.5_03"' Rscripts/ConvergentPheno_read.R
Rscripts/ConvergentPheno_read_varspike_1.5_03.Rout
R CMD BATCH --no-save --no-restore '--args n="varspike_1.5_04"' Rscripts/ConvergentPheno_read.R
Rscripts/ConvergentPheno_read_varspike_1.5_04.Rout
```

```
R CMD BATCH --no-save --no-restore '--args n="equalprop_0.5"' Rscripts/
ConvergentPheno_read_superpc.R Rscripts/ConvergentPheno_read_superpc_ncomps_0.5.Rout
R CMD BATCH --no-save --no-restore '--args n="equalprop_1.0"' Rscripts/
ConvergentPheno_read_superpc.R Rscripts/ConvergentPheno_read_superpc_ncomps_1.0.Rout
R CMD BATCH --no-save --no-restore '--args n="equalprop_1.5"' Rscripts/
ConvergentPheno_read_superpc.R Rscripts/ConvergentPheno_read_superpc_ncomps_1.5.Rout
```

```
#####
##### Part 1.4: Merge the results #####
#####
```

```
R CMD BATCH --no-save --no-restore "--args n='equalprop_0.5' index=c('01','02','03','04')"
Rscripts/ConvergentPheno_read_merge.R \
Rscripts/ConvergentPheno_read_merge_equalprop_0.5.Rout
R CMD BATCH --no-save --no-restore "--args n='equalprop_1.0' index=c('01','02','03','04')"
Rscripts/ConvergentPheno_read_merge.R \
Rscripts/ConvergentPheno_read_merge_equalprop_1.0.Rout
R CMD BATCH --no-save --no-restore "--args n='equalprop_1.5' index=c('01','02','03','04')"
Rscripts/ConvergentPheno_read_merge.R \
Rscripts/ConvergentPheno_read_merge_equalprop_1.5.Rout
R CMD BATCH --no-save --no-restore "--args n='randommechs_0.5'
index=c('01','02','03','04','05','06','07','08')" Rscripts/ConvergentPheno_read_merge.R \
Rscripts/ConvergentPheno_read_merge_randommechs_0.5.Rout
R CMD BATCH --no-save --no-restore "--args n='randommechs_1.0'
index=c('01','02','03','04','05','06','07','08')" Rscripts/ConvergentPheno_read_merge.R \
Rscripts/ConvergentPheno_read_merge_randommechs_1.0.Rout
R CMD BATCH --no-save --no-restore "--args n='randommechs_1.5'
index=c('01','02','03','04','05','06','07','08')" Rscripts/ConvergentPheno_read_merge.R \
Rscripts/ConvergentPheno_read_merge_randommechs_1.5.Rout
```

```
R CMD BATCH --no-save --no-restore "--args n='randomvalidation_0.5'
index=c('01','02','03','04','05','06','07','08')" Rscripts/ConvergentPheno_read_merge.R \
Rscripts/ConvergentPheno_read_merge_randomvalidation_0.5.Rout
R CMD BATCH --no-save --no-restore "--args n='randomvalidation_1.0'
index=c('01','02','03','04','05','06','07','08')" Rscripts/ConvergentPheno_read_merge.R \
Rscripts/ConvergentPheno_read_merge_randomvalidation_1.0.Rout
R CMD BATCH --no-save --no-restore "--args n='randomvalidation_1.5'
index=c('01','02','03','04','05','06','07','08')" Rscripts/ConvergentPheno_read_merge.R \
Rscripts/ConvergentPheno_read_merge_randomvalidation_1.5.Rout
R CMD BATCH --no-save --no-restore "--args n='overlap_0.5'
index=c('01','02','03','04','05','06','07','08','09',seq(10,28))" Rscripts/
ConvergentPheno_read_merge.R \
Rscripts/ConvergentPheno_read_merge_overlap_0.5.Rout
R CMD BATCH --no-save --no-restore "--args n='overlap_1.0'
index=c('01','02','03','04','05','06','07','08','09',seq(10,28))" Rscripts/
ConvergentPheno_read_merge.R \
Rscripts/ConvergentPheno_read_merge_overlap_1.0.Rout
R CMD BATCH --no-save --no-restore "--args n='overlap_1.5'
index=c('01','02','03','04','05','06','07','08','09',seq(10,28))" Rscripts/
ConvergentPheno_read_merge.R \
Rscripts/ConvergentPheno_read_merge_overlap_1.5.Rout
```

```
R CMD BATCH --no-save --no-restore "--args n='varspike_0.5' index=c('01','02','03','04')"
Rscripts/ConvergentPheno_read_merge.R Rscripts/ConvergentPheno_read_merge_varspike_0.5.Rout
R CMD BATCH --no-save --no-restore "--args n='varspike_1.0' index=c('01','02','03','04')"
Rscripts/ConvergentPheno_read_merge.R Rscripts/ConvergentPheno_read_merge_varspike_1.0.Rout
R CMD BATCH --no-save --no-restore "--args n='varspike_1.5' index=c('01','02','03','04')"
Rscripts/ConvergentPheno_read_merge.R Rscripts/ConvergentPheno_read_merge_varspike_1.5.Rout
```

```
rm -rf temp ## Files in temp can be removed after merging all the files
```

```
#####
##### Part 1.5: plot figures and generate tables #####
#####
```

```
R CMD BATCH --no-save --no-restore Rscripts/ConvergentPheno_plot.R Rscripts/
ConvergentPheno_plot.Rout
```

```
#####
##### Part 2.0: download the required datasets #####
#####
```

```
mkdir -p data
cd data
```

```
ftp ftp://ftp.ncbi.nlm.nih.gov/geo/series/GSE25nnn/GSE25055/matrix/
GSE25055_series_matrix.txt.gz
ftp ftp://ftp.ncbi.nlm.nih.gov/geo/series/GSE25nnn/GSE25065/matrix/
GSE25065_series_matrix.txt.gz
```

```
gunzip GSE25055_series_matrix.txt.gz
gunzip GSE25065_series_matrix.txt.gz
```

```
cd ..
```

```
R CMD BATCH --no-save --no-restore Rscripts/preprocess_clinicalData.R Rscripts/
preprocess_clinicalData.Rout
```

```
#####
##### Part 2.1: the main part #####
#####
```

```
R CMD BATCH --no-restore --no-save '--args split.by="Tstage" validation="testset"' Rscripts/
Clinical_makeSig.R Rscripts/Clinical_makeSig_01.Rout
```

```
R CMD BATCH --no-restore --no-save '--args split.by="Nstage" validation="testset"' Rscripts/
Clinical_makeSig.R Rscripts/Clinical_makeSig_02.Rout
```

```
R CMD BATCH --no-restore --no-save '--args split.by="age" validation="testset"' Rscripts/
Clinical_makeSig.R Rscripts/Clinical_makeSig_03.Rout
```

```
R CMD BATCH --no-restore --no-save '--args split.by="Tstage" validation="L00CV"' Rscripts/
Clinical_makeSig.R Rscripts/Clinical_makeSig_04.Rout
```

```
R CMD BATCH --no-restore --no-save '--args split.by="Nstage" validation="L00CV"' Rscripts/
Clinical_makeSig.R Rscripts/Clinical_makeSig_05.Rout
```

```
R CMD BATCH --no-restore --no-save '--args split.by="age" validation="L00CV"' Rscripts/
Clinical_makeSig.R Rscripts/Clinical_makeSig_06.Rout
```