**How to run gistic and mutsig for different data sets**

1) For testing: On cluster, there is a “Gistic2” folder. If we have already merged all segList files from multiple patients of a cohort together to form one single tsv file (/home/yliu40/Gistic2/ gistic.tsv), you can take the steps below to run gistic:

1)cd /home/yliu40/Gistic2/2.0

2)./my\_run.sh (it looks like below)

./gistic2 -b ../result -seg ../gistic.tsv -refgene ../hg19.UCSC.add\_miR.140312.refgene.mat -genegistic 1 -smallmem 0 -rx 0 -broad 1 -brlen 0.7 -conf 0.99 -armpeel 1 -savegene 1 -gcm extreme -v 30 -maxseg 46000 -ta 0.3 -td 0.3 -cap 1.5 -js 4

**Note: We need to make sure the result folder exists and the parameters required by gistic can be altered.**

2) Use the “runGistic” and “runMutSig” functions in the CGLTools (both functions require single file (gistic.tsv and merged\_50\_patients.maf).

(a)In order to run the mutsig, we need to load the module as below:

module load R/4.1.0

export R\_LIBS="/risapps/iacs/rhel7/R4.1.0\_Libs"

(b)Then enter R environment by typing in R in the command window and load the CGLTools that John packaged by type in:

library("CGLTools")

(c)Run gistic and mutsig by using the commands below:

runGistic("/rsrch4/home/genomic\_med/yliu40/CGLTools/SingleFileData/gistic.tsv", "/rsrch4/home/genomic\_med/yliu40/CGLTools/result", gsPath="/risapps/rhel7/gistic/2.0",build = c("hg19", "hg38"), genegistic = 1, smallmem = 0,rx = 0,broad = 1,brlen = 0.7,conf = 0.99,armpeel = 1,savegene = 1,gcm = "extreme",v = 30, maxseg = 46000, ta = 0.3, td = 0.3, cap = 1.5, js = 4)

runMutSig("/rsrch4/home/genomic\_med/yliu40/CGLTools/SingleFileData/merged\_50\_patients.maf","/rsrch4/home/genomic\_med/yliu40/CGLTools/result", n\_category=5,skip\_permu="false",maxper=1e5,remove\_dup="true",mutsigPath="/rsrch3/scratch/iacs/ngs\_pipe/apps/MutSig2CV/mutsig2cv")

**Note: the gistic.tsv, merged\_50\_patients.maf and the output directory “/home/genomic\_med/yliu40/CGLTools/result” must exist.**

**When you run the mutsig function, the error below may occur:**

Error: Could not find version 9.8 of the MATLAB Runtime.move\_dup="true",mutsigPath="/rsrch3/scratch/iacs/ngs\_pipe/apps/MutSig2CV/mutsAttempting to load libmwmclmcrrt.so.9.8.

Please install the correct version of the MATLAB Runtime.

Contact your vendor if you do not have an installer for the MATLAB Runtime.

This issue can be addressed by coping and pasting the following runMutSig function into R, in which the matlab version was specified as R2020a.

runMutSig <- function(maf, odir, n\_category = 5, skip\_permu = "false", maxper = 1e5,

remove\_dup = "true", mutsigPath = "/rsrch3/scratch/iacs/ngs\_pipe/apps/MutSig2CV/mutsig2cv"){

wd <- getwd()

on.exit(setwd(wd), add = TRUE)

setwd(mutsigPath)

params <- file.path(odir, basename(tempfile()))

on.exit(unlink(params), add = TRUE)

para <- matrix(c("number\_of\_categories\_to\_discover", n\_category, "skip\_permutations", skip\_permu,

"maxperm", maxper, "remove\_duplicate\_patients", remove\_dup), ncol = 2, byrow = TRUE)

write.table(para, file = params, sep = "\t", col.names = FALSE, row.names = FALSE, quote = FALSE)

cmd <- paste("module load matlab/R2020a;", file.path(mutsigPath, "MutSig2CV"), maf, odir, params)

system(cmd)

return(odir)

}

3)If the mutect, pindel and exomecn folders and the files that should be put into each folder are ready, we need to run **cohortrun**.

Under this circumstance, the gistic and mutsig can be run by taking the steps below.

1. library("CGLTools")
2. After copy and paste the full content of “cohortrun\_10\_05\_2022.R” into R.
3. We assume the mutect, pindel and cn data are in: /rsrch4/home/genomic\_med/yliu40/CGLTools/MultipleFileData/results/ROPR0011.

Finally type in the script below to run gistic and mutsig, cohortrun(dirname="/rsrch4/home/genomic\_med/yliu40/CGLTools/MultipleFileData/results/ROPR0011")

The program will generate two additional folders named “gistic” and “mutsig” to save the respective results.

If we don’t have the mutect, pindel and exomecn files ready in their own folders, we need to generate the folders and files from meta data, we need to run the metarun by taking the steps below.

1. meta <-fread("/rsrch4/home/genomic\_med/yliu40/CGLTools/MultipleFileData/metadata/UVM\_metadata\_09282022\_full.tsv")
2. copy and paste the full content of “metarunCombined.R” into R to generate the three folders and files and run

“gistic” and “mutsig”.

Note: If we directly run “metarunCombine.R” the mutsig failed with the error below:

Mutation file is missing a column named one of the following:

gene

Hugo\_Symbol

Gene\_name

Error using MutSig\_2CV\_v3\_11\_core (line 169)

Mutation file missing gene column.

But if we use the samplerun.R result (mutect and exomecn folder with files), run cohortrun\_10\_05\_2022.R, no problem at all.

cohortrun, metarun and samplerun (only requires mutect, pindel and exomecn files).

**cohortrun\_10\_05\_2022.R**: can perform the gistic and mutsig for the entire cohort of a study.

**samplerun\_10\_05\_2022.R**: according to the columns of “mutect”, “pindel” ,“cn” and “exp” (the location of mutect, pindel , cn and RNA expression files) in the meta file (/home/yliu40/data/UVM/metadata/UVM\_metadata\_09282022\_full.tsv) read mutect, pindel, cn and expression files into corresponding folder (mutect and pindel🡪 “mutect” folder; cn files-> “exomecn” folder and RNA expression 🡪 “htseqcount” folder.

**metarun\_10\_05\_2022.R**: the combination of “amplerun” and “cohortrun”.