

GWAS practicum with PLINK and Ricopili

Presenter: Daniel Howrigan

Data group leader – Neale Lab

July 2020

Practicum Format

For each section

- **Part 1: What I do..**
 - Share my screen
 - Tutorial walk-through
- **Part 2: What you do..**
 - Gallery view (questions/answers/share your screen)
 - Run commands / Break time
- More experienced users:
- https://personal.broadinstitute.org/howrigan/workshop/hapmapEA_QC_script.sh
- https://personal.broadinstitute.org/howrigan/workshop/GWAS_QC_Howrigan_July2020.pdf

Practicum overview

- Getting started
 - Script and command line
 - Logging onto the Broad servers
 - Broad dotkit and .my.bashrc
 - Creating web-accessible directories
- Using PLINK and R
 - Copying sample data
 - Running PLINK
 - Running R
 - Viewing graphs
- Ricopili
 - Downloading Ricopili
 - Installing Ricopili
 - Running Ricopili pre-imputation QC
 - Running Ricopili PCA

Practicum overview

- Getting started
 - Script and command line
 - Logging onto the Broad servers
 - Broad dotkit and .my.bashrc
 - Creating web-accessible directories
- Using PLINK and R
 - Copying sample data
 - Running PLINK
 - Running R
 - Viewing graphs
- Ricopili
 - Downloading Ricopili
 - Installing Ricopili
 - Running Ricopili pre-imputation QC
 - Running Ricopili PCA

Getting started: Scripts and command line

- **Scripts**
 - Where you write/edit your commands
 - Scripts follow specific formats based on program
 - Examples: shell script (.sh), Python script (.py), R script (.R)
- **Command line**
 - Where you run/execute commands and scripts
 - Navigating directories/servers

Getting started: Logging onto the Broad server

- Make sure you are logged into the VPN (Virtual Private Network)
 - Cisco AnyConnect on your home computer
 - DUO app with Broad Institute log in
 - 2-factor authentication (2FA)

Logging onto the Broad servers

```
ssh [username]@login.broadinstitute.org
```

Getting started: Logging onto the Broad server

Viewing home directory

```
ls
ls -l
ls -ltrh
ls -ltrah
ls -ltrah --color='auto'
```

Checking group permissions

```
groups
ls /stanley/genetics/
ls -ltrah /humgen/atgu1
```

Getting started: Broad dotkit and .my.bashrc

Viewing the Broad dotkit

```
use  
use -l  
use -l | grep -C 4 'R-'
```

Installing a program from the dotkit

```
use UGER  
use  
unuse UGER  
use
```

Getting started: Broad dotkit and .my.bashrc

Viewing .my.bashrc file

```
cat .my.bashrc
```

Editing .my.bashrc file with nano editor

```
nano .my.bashrc
```

Adding some start up commands in .my.bashrc

```
use UGER  
use PLINK  
use R-3.3
```

```
alias lsa='ls -ltrah --color=auto'
```

ctrl+x to exit

Re-run .my.bashrc

```
. .my.bashrc
```

Getting started: creating web-accessible directories

Creating password-protected web accessible directory:

Return to home directory (just in case)

```
cd ~
```

Create private_html directory

```
mkdir -p private_html/  
mkdir -p private_html/atgu_workshop
```

Create file in directory

```
echo 'hello_world' > private_html/hello_world.txt
```

View file in web browser:

[https://internal.broadinstitute.org/~\[username\]/](https://internal.broadinstitute.org/~[username]/)

```
chmod 755 ~/
```

update permissions

Run commands

Q & A

Break time

Practicum overview

- Getting started
 - Script and command line
 - Logging onto the Broad servers
 - Broad dotkit and .my.bashrc
 - Creating web-accessible directories
- Using PLINK and R
 - Copying sample data
 - Running PLINK
 - Running R
 - Viewing graphs
- Ricopili
 - Downloading Ricopili
 - Installing Ricopili
 - Running Ricopili pre-imputation QC
 - Running Ricopili PCA

Using PLINK and R: copying sample data

Working in an interactive node (more memory!)

Calling an interactive node

```
ish -l h_vmem=8G
```

Create atgu_workshop directory

```
mkdir -p ~/atgu_workshop/  
cd ~/atgu_workshop
```

Create soft links to sample PLINK data

```
ln -s /web/personal/howrigan/workshop/hapmapEA.bed hapmapEA.bed  
ln -s /web/personal/howrigan/workshop/hapmapEA.bim hapmapEA.bim  
ln -s /web/personal/howrigan/workshop/hapmapEA.fam hapmapEA.fam
```

Using PLINK and R: copying sample data

List files

```
ls -l
```

View files

```
less hapmapEA.fam ## press 'q' to exit  
more hapmapEA.bim ## press 'q' to exit  
zless hapmapEA.bed ## press 'q' to exit
```

Copy over additional R scripts

```
cp /web/personal/howrigan/workshop/*R .
```

Using PLINK and R: Running PLINK

Run a GWAS (chisq test) in PLINK

```
plink --bfile hapmapEA --assoc --out hapmapEA
```

Run a GWAS (logistic regression) in PLINK

```
plink --bfile hapmapEA --logistic --out hapmapEA
```

View .assoc files

```
less hapmapEA.assoc ## press 'q' to exit
```

```
less hapmapEA.logistic.assoc ## press 'q' to exit
```

Using PLINK and R: Running R script

View Rscript

```
cat QQplot.R
```

Run Rscript

```
Rscript QQplot.R hapmapEA.assoc  
Rscript QQplot.R hapmapEA.assoc.logistic
```

Move plots to ~/private_html

```
mv hapmapEA.assoc_QQ.png ~/private_html/atgu_workshop/  
mv hapmapEA.assoc.logistic_QQ.png ~/private_html/atgu_workshop/
```

Run commands

Q & A

Break time

Practicum overview

- Getting started
 - Script and command line
 - Logging onto the Broad servers
 - Broad dotkit and .my.bashrc
 - Creating web-accessible directories
- Using PLINK and R
 - Copying sample data
 - Running PLINK
 - Running R
 - Viewing graphs
- Ricopili
 - Downloading Ricopili
 - Installing Ricopili
 - Running Ricopili pre-imputation QC
 - Running Ricopili PCA

Ricopili: Rapid Imputation Consortium Pipeline

- Developed at ATGU by Stephan Ripke
 - An analyst, not a software developer
 - Multi-layered script wrapper calling many different programs
 - Vulnerable to small changes in server environment

Website: <https://sites.google.com/a/broadinstitute.org/ricopili/>

4 main modules:

- Pre-imputation QC (preimp_dir)
- PCA (pcaer)
- Imputation (impute_dirsub)
- Analysis (postimp_navi)

Ricopili: Downloading ricopili

Change back to home directory

```
cd ~
```

Create atgu_workshop directory

```
mkdir -p ricopili/  
cd ricopili
```

Download latest release of Ricopili scripts

```
wget https://sites.google.com/a/broadinstitute.org/ricopili/download//rp_bin.2019_Oct_15.001.tar.gz
```

Un-tar the file bundle

```
tar -zxvf rp_bin.2019_Oct_15.001.tar.gz
```

View directory of scripts

```
lsa rp_bin/
```

Ricopili: Installing ricopili

First create scratch directory

```
mkdir -p sloc
```

Change into rp_bin/

```
cd rp_bin
```

Install ricopili (run rp_config script)

```
./rp_config
```

Ricopili: Installing ricopili

View “new” file created

```
cat rp_config.custom.txt
```

Copy over template from /workshop resources

```
cp /web/personal/howrigan/workshop/rp_config.custom.txt .
```

Edit template to customize

```
nano rp_config.custom.txt
```

```
## edit these lines (replace [fill in here]):  
# path_to_scratchdir           /home/unix/[username]/ricopili/sloc/  
# rp_user_initials             [user 2-letter initials]  
# rp_user_email                [user broad email]  
# rp_logfiles                   /home/unix/[username]/ricopili
```

Run rp_config again

```
./rp_config
```

Ricopili: Installing ricopili

Add additional lines to \$PATH

- Copy text from Ricopili output
- Remove #
- Paste into command line

Check .my.bashrc

```
cat ~/.my.bashrc
```

Add additional LaTeX update

```
## add line:PATH=/psych/ripke/share/latex_2019/bin/x86_64-linux:$PATH;
```

Run .my.bashrc

```
. ~/.my.bashrc
```

Run rp_config YET again (NOTE the change in script execution)

```
rp_config
```

Run commands

Q & A

Break time

Ricopili: Running ricopili pre-imputation QC

Go back to original data

```
cd ~/atgu_workshop
```

Run pre-imputation QC module

```
preimp_dir --dis scz --popname eur --out hapmapEA
```

View name file (but don't make any changes)

```
cat scz.names
```

Run pre-imputation QC module again

```
preimp_dir --dis scz --popname eur --out hapmapEA
```

Check job submission

```
qstat -u [username]  
qstat -j [jobid]
```

Ricopili: Running ricopili pre-imputation QC

When finished, copy .pdf to web-accessible directory

```
cp ~/atgu_workshop/qc/scz_scz1_eur_[rp_user_initials]-qc1.pdf  
~/private_html/atgu_workshop/
```

View in browser

```
https://internal.broadinstitute.org/~[username]/atgu_workshop/scz_scz1_eur_[initials]-qc1.pdf
```

Ricopili: Running ricopili PCA

Make sure you are in the qc directory

```
cd ~/atgu_workshop/qc
```

Run PCA module

```
pcaer --out hapmapEA scz_scz1_eur_[rp_user_initials]-qc1.bim
```

Check job submission

```
qstat -u [username]
```

```
qstat -j [jobid]
```

When completed, copy the unzipped .pdf to /private_html

```
gunzip -c
```

```
~/atgu_workshop/qc/hapmapEA.menv.mds.2ds.pdf.gz >
```

```
~/private_html/atgu_workshop/hapmapEA.menv.mds.2ds.pdf
```

Final Q & A

EXTRA:

Creating publicly accessible directory:

1) Fill out form at BITS

```
https://broad.service-now.com/com.glideapp.servicecatalog\_cat\_item\_view.do?sysparm\_id=26ef16064ff222005c1423d18110c771
```

2) Purpose: “new hire training”

3) Acknowledgement: Yes

Create file in directory

```
echo 'hello_world' > /web/personal/[username]/hello_world.txt
```

View file in web browser:

```
https://personal.broadinstitute.org/[username]/
```

```
chmod 755 /web/personal/[username]/
```