

Association Mapping Case Study

Instructors: Tamar Sofer, Kari North, and Misa Graff

July 2017

Overview

- ▶ In this module we will present and discuss methods and softwares for genetic association studies.
- ▶ We will focus on studies with complex population structure.
- ▶ We will follow the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) as a case study.

Overview

We will...

1. Present the HCHS/SOL.
2. Present a simulated data set that mimics the HCHS/SOL (simpler).
3. Learn how to run
 - ▶ Linear mixed model GWAS.
 - ▶ Logistic mixed model GWAS.
 - ▶ GEE based GWASs.
4. Discuss and practice possible ways to deal with heterogeneity within the study.
 - ▶ In both the mixed models and the GEE frameworks.
5. Learn to perform generalization analysis.
6. Run admixture mapping GWAS.
7. Discuss and practice GxE analyses.
 - ▶ In both the mixed models and the GEE frameworks.

See module schedule for details!

Resources

The class site contains

- ▶ PDF slide sets.
- ▶ Data sets.
- ▶ Instructions for software installation.
- ▶ Published manuscripts of interest.

Instruction team: Tamar Sofer



- ▶ Research scientist at the University of Washington, Department of Biostatistics.
 - ▶ Future position at Harvard Medical School/Brigham and Women's Hospital.
- ▶ Worked at the Genetic Analysis Center of the HCHS/SOL.

Instruction team: Kari North



- ▶ Professor of Epidemiology at UNC.
- ▶ Leads the Population Architecture using Genomics and Epidemiology (PAGE) study.

Instruction team: Mariaelisa (Misa) Graff



- ▶ Research Assistant Professor of Epidemiology at UNC.
- ▶ Worked on many genetic epidemiology studies.

Now let's download and install softwares!

- ▶ First, download and install the latest R version.
- ▶ Install Bioconductor R packages:

```
source("https://bioconductor.org/biocLite.R")  
biocLite("GWASTools")  
biocLite("GENESIS")  
biocLite("gdsfmt")
```

- ▶ If you cannot install the latest R version, you may have to install GENESIS package manually to have all functionalities!

Now let's download and install softwares!

- ▶ Install R packages from GitHub:

```
install.packages("devtools")  
library("devtools")  
install_github("tamartsi/generalize@Package_update",  
              subdir = "generalize")  
install_github("tamartsi/MetaCor")
```

Now let's download and install softwares!

- ▶ Other R packages.

```
install.packages("mvtnorm")  
install.packages("ggplot2")  
install.packages("gridExtra")  
install.packages("RColorBrewer")
```

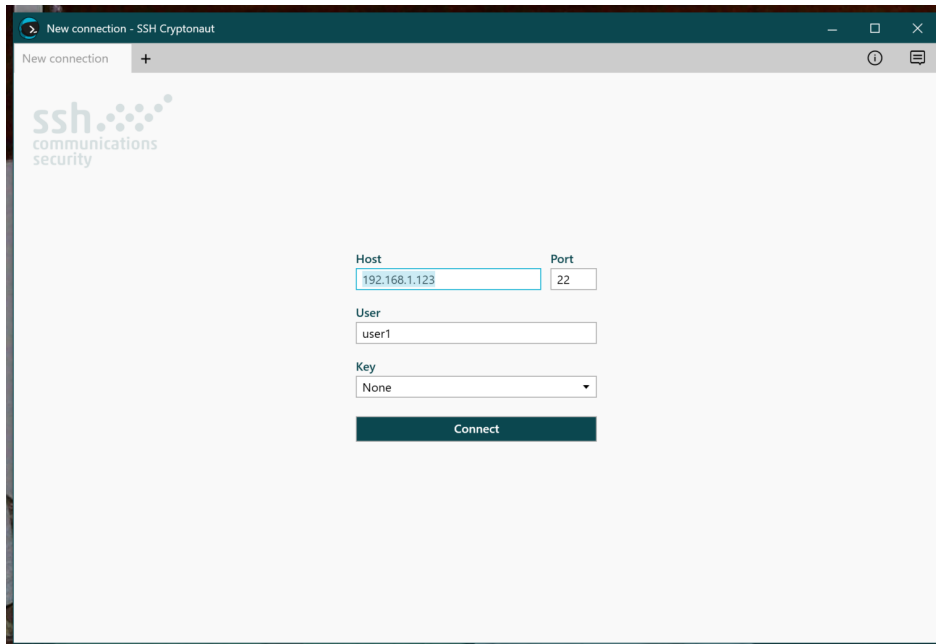
Download data sets

- ▶ Download data sets from the module website
`https://www.biostat.washington.edu/suminst/SISG2017/modules/SM1712`.
- ▶ And save all data sets in the same folder.
- ▶ When using R code, we will always start by setting our working directory using a command that looks like:

```
dir <- "/mycomputer/variousfolders/module12_folder"  
# put your own!
```

- ▶ Save this command somewhere, with YOUR working directory (the path for the folder you use) for an easy use in the next sessions.

Set up a Linux/Unix environment

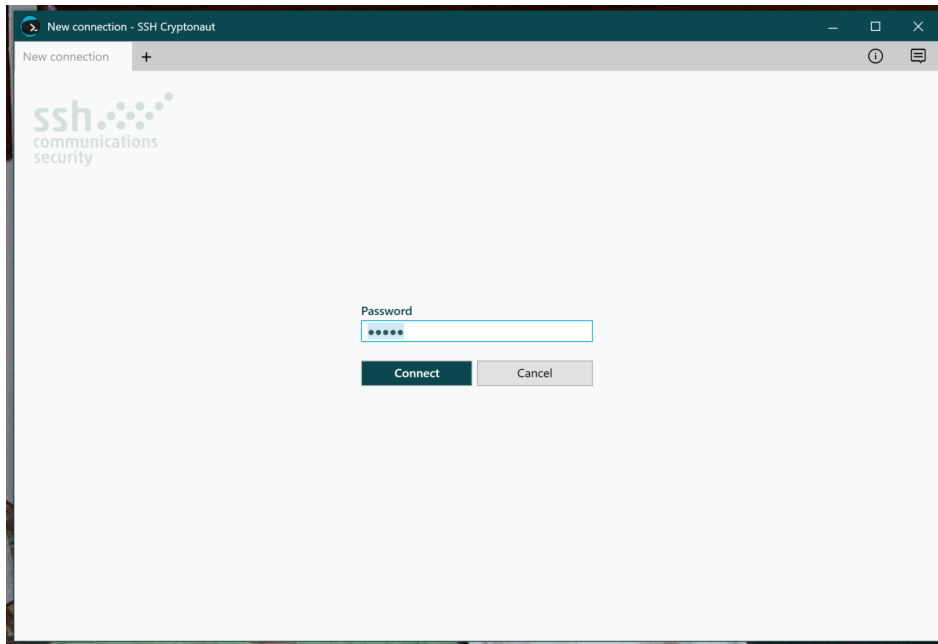


➤ Fill i:

- IP address to Host
- User ID= user#
- Key = None

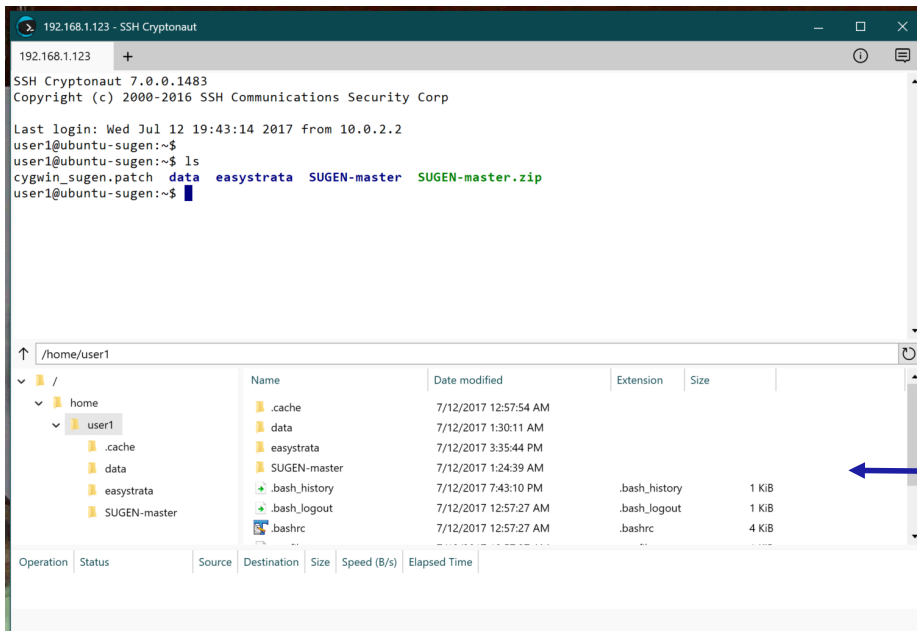
➤ Then click 'Connect'

Set up a Linux/Unix environment



- Enter Password=user#
- click 'Connect' again

Set up a Linux/Unix environment



- See command line
user1@ubuntu-sugen:~\$
- View files in folder
user1@ubuntu-sugen:~\$ls

GUI interface of your space –
can drag and drop items into here

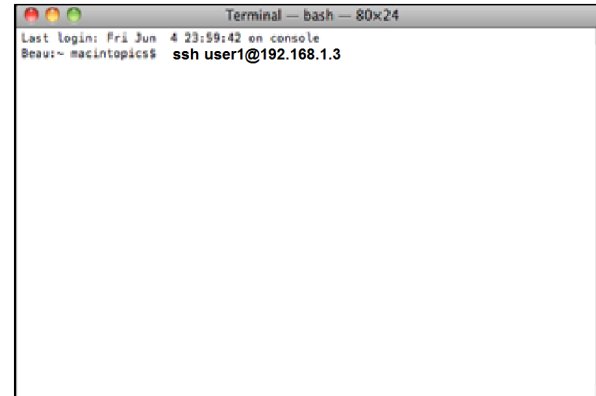
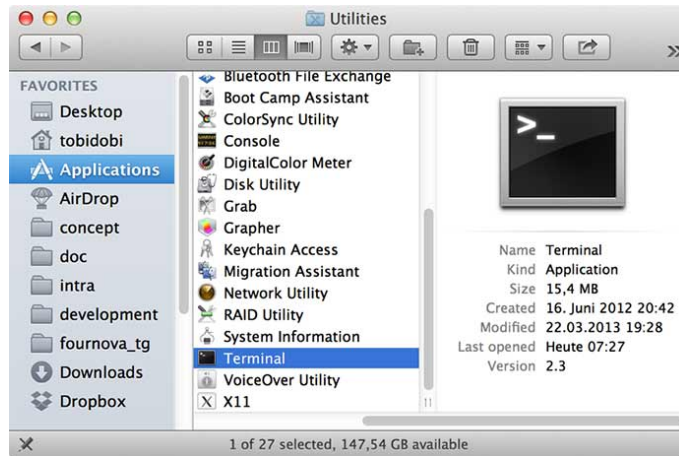
MAC: Set up a Linux/Unix environment

For server shell:

Go to Utilities-> Applications -> Terminal

Log into the server with the following command:

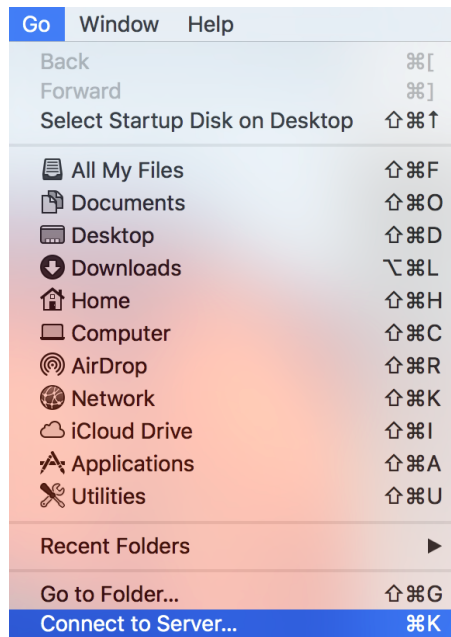
```
$ ssh user1@192.168.1.3
```



MAC: Set up a Linux/Unix environment

For file drag and drop:

Go to Finder-> Go -> Connect to Server



Provide server path with your user#:
smb://192.168.1.3/user1



SUGEN Installation

- SUGEN can be downloaded here: <https://github.com/dragontaoran/SUGEN#>
 - It is already in your folder for convenience.
- To install:
 - Unzip the package.

```
$ unzip SUGEN-master.zip
```
 - Go to the SUGEN directory.

```
$ cd SUGEN-master/
```
 - Install SUGEN. When successful, an executable called "SUGEN" will be generated in SUGEN-master/.
 - On some servers you can go directly to the 'make' command, but if you get a compiler error you can use this patch and rerun 'make'.

```
$ patch -p1 <../cygwin_sugen.patch
```

```
$ make
```

Plotting using EasyStrata R package

- We will do some plotting using EasyStrata.
- Install EasyStrata (version 8.6)

```
install.packages("EasyStrata")
```

- Copy folder and its contents from server called “easystrata” to your laptop:
C:\easystrata