



Lab 4: PLINK Setup (30 Points)

Objective

Introduce students to the PLINK CLI tool.

Part 1: Installation (10 Points)

- Download and install Plink from this [link](#). Note that the installation setup may differ based on your platform. It is better to use Linux-based environments such as Ubuntu/macOS since some commands may differ on Windows.
- Add Plink to your Path environment variable.
 - It is found in ~/.bashrc in Linux-based platforms or ~/.zshrc if you are running oh-my-zsh on your terminal
 - It is found in the Windows environment variables found in the advanced properties of This PC

```
→ Bioinformatics plink --version  
PLINK v1.90b7.2 64-bit (11 Dec 2023)
```

Part 2: Basic Commands (15 Points)

- Download the [Dataset of 156 Qataris](#).
- Extract the dataset. What formats do you see?
- Convert the files in the current format to PED/MAP format using:

```
plink --bfile your_input_filename --recode --out your_output_filename
```

- Upon executing the previous command, observe the terminal output. Specify the number of variants and the number of samples.

- Describe the columns of the ped and map files while exploring each file's head (first 5 rows).
- Perform one of the quality controls, Missing Call Rate, found in this [link](#)
You can run the command based on your file format

.bed,.bim,.fam:

```
plink --bfile <prefix> --geno <threshold> --recode --make-bed -out <filename>
```

PED/MAP:

```
plink --file <prefix> --geno <threshold> --recode --make-bed -out <filename>
```

Try different thresholds and report the number of variants removed based on the thresholds used.

Reporting (5 points)

- You are required to write a report discussing the commands, and file formats and include any screenshots of the output of such commands
- Write your comments on how the data quality control affected the dataset

References

- [Dataset of 156 Qataris](#)
- [PLINK file formats](#)