**4.1 Projects on Genomic Data Analysis**

A set of hands-on projects will be designed to accompany with the corresponding course modules

with applications on real human genomic data. We will design a set of lightweight and intense course

projects adapted to the expertise and backgrounds of the targeted students. Specifically, we plan to

develop the following projects based on our previous experiences of analyzing human genomic data

from 1000GP [26, 82, 98] and related projects.

1) The goal of the first project is to train the students to gain a high level understanding of the

sequencing technology, the knowledge of human genome sequence data, and the experiences with

commonly used computational and bioinformatics tools for common practices in human genomics

including variant calling, genotyping, and phasing. We will use latest data release from the openly

accessed 1000GP as a user case [82]. The 1000GP is currently the most comprehensive and largest

sequencing project on human populations.

2) The second project aims to offer real world experiences for applying computational methods

to identifying genetic variation associated with traits/phenotypes in GWAS settings. We will pro-

vide the 1000GP data as an example, using population labels as phenotype labels. Many human

phenotypes have population specificity and previous studies [108] including our prior work [138, 140],

have used population labels as phenotypes and found genetic variation associated with population

differentiation. We will provide an option of replacing the 1000GP data with other genomic data

with personal traits. For example, the users of this project module can use a personal trait of interest

(e.g. "Acne") in the openly accessed Personal Genome Project or use the GWAS dataset they have

access to.

3) The third project aims for practicing computational approaches for meta-analysis in GWAS. We

will use population labels as phenotypes on the 1000GP [82], HapMap [32], the two large scale studies

based on two representative technologies, next generation sequencing and genotyping microarrays

respectively. Again, the users can opt out to GWAS datasets such as data from various studies on

human height [72, 76, 130].

4) The fourth project will be performing eQTL analysis, using variant genotypes from the 1000GP

[82], and gene expression data from GEUVADIS consortium [105] on the 1000GP individuals. As

described in our review article [71], we will apply two major types of eQTL methods, classical

correlation/regression methods and machine learning methods for this project.

5) The fifth project is targeted for epistasis analysis. Like efforts in our previous work [124, 138,

140], we will use the 1000GP genotype data, with population labels as phenotypes for categorical

phenotypes, and gene expression data as examples of cases with quantitative phenotypes.

**4.3 Projects on Ethics, Regulations, and Techniques for Privacy Protection**

These course projects will be mainly focused on educating students and bringing their awareness

of the evolution, dynamics, and current status of ethics, regulations, and techniques for protecting

genetic privacy. Three projects will be designed as follows.

1) The first project will ask the students to pick one of the large human genomic studies as a user

case and address ethic issues in this specific project. These genomic study user cases include but

reach beyond the Human Genome Project [2], HapMap project [29], the 1000 Genomes Project [78],

the Personal Genome Education Project (PGeD [?].

2) The second project will prompt the students to study the recent regulations such as HIPAA privacy rule, describe situations and types of health information that is protected or exempted from this

rule. Pertinent genetic privacy issues will be reviewed regarding to open (e.g. PGP) and controlled

access of human genomic data (e.g. dbGaP).

3) The third project will describe the disclosure and protection of genetic privacy for genomic

study participants and regular individuals. Specific details of these techniques will be required to

be provided for assessing the privacy of GWAS participants, and regular individuals like 1000GP

samples.