**III. 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.

**Project 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 [1], HapMap project [2], the 1000 Genomes Project [3],

the Personal Genome Education Project (PGeD) [4].

Many large human genome projects provide ethics education, such as the Human Genome Project, HapMap Project, the 1000 Genomes Project, and the Personal Genome Education Project. Ethics education and enforcement of laws and regulations play an important role in genetic privacy protection. A wide range of topics including ethics, regulations, and laws regarding protection and preservation of genetic privacy have been established and disseminated by scientific researchers. The reason the researcher community is making a huge effort to enhance and broaden ethics education is that not only it has significant impact on research, education, and clinical practice of personal genomics but also it helps to achieve of a goal of genetic privacy protection for individuals. As indicated by recent privacy and identity infringement work, protecting genomic anonymity becomes next to impossible because researchers increasingly combine patient data with many types of data from social media posts to entries in genealogy databases. Therefore, it is important for scientific community to train students, researchers, medical practitioners to build ethics about genetic counseling, prenatal screening, and incidental findings. Read the following articles and answer the following questions.

1. What potential genetic privacy risks are there in these large scale genome projects?
2. As a community, how can we enforce or develop ethics guidelines and laws to ensure sharing data toward reproducible research while protecting individuals’ privacy?

[1] National Human Genome Research Institute. All about the human genome project (HGP). 2015. http://www.genome.gov/10001772

[2] The International HapMap Consortium. Integrating ethics and science in the International HapMap Project. Nat. Rev. Genet. 2004; 5:467–475. [PubMed: 15153999]

[3] Rodriguez LL, Brooks LD, Greenberg JH, et al. The complexities of genomic identifiability. Science. 2013; 339:275–276. [PubMed: 23329035]

[4] Personal Genetics Education Project (pgEd). 2016. http://www.pged.org/

**Project 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) [8].

Health Insurance Portability and Accountability Act of 1996 (HIPAA) is United States legislation that provides data privacy and security provisions for safeguarding medical information. The HIPAA privacy rule was established to protect patients’ individually identifiable health information and HIPAA security rule was established to protect patient data security. The main purpose of HIPAA is to guarantee that workers who lose or change their jobs obtain continuous health insurance coverage, and to reduce the administrative burdens. The privacy rule also guarantees patients have the right to request their own protected health information from healthcare providers covered by HIPAA. However, some newly born commercial sequencing and genetic screening companies (for example, 23 and Me) are not covered by HIPAA, even they generate identifiable health information. The HIPAA privacy rule only applies to organizations who are considered HIPAA-covered entities. It means who held or transmitted by a covered entity is protected. In addition, with the goal of ensuring broad and responsible sharing of genomic data, the National Institutes of Health (NIH) issued the genomic data sharing policy, legal and ethical issues in genetic research. In order to mitigate the risk of sharing genetic data publicly, three types of access control models to genomic data has been widely used. The first basic access control model allows users to download data only after approval has been granted and under defined condition. An alternative model for access control uses a trust-but-verify approach, in which users cannot download data without restrictions but, on the basis of their privileges, may execute certain types of queries, which are recorded and audited by the system. Another model of access control is allowing the original participants to grant access to their data instead of delegating this responsibility to a data access committee [9].

Read the following articles and answer these questions:

1. What types of data are protected and what types of data are not protected by the current HIPAA?
2. What modifications or supplements would you envision to complement the HIPAA rule to ensure protection of human genetic privacy?

[5] <https://searchhealthit.techtarget.com/definition/HIPAA>

[6] <https://www.hhs.gov/hipaa/index.html>

[7] <https://www.genome.gov/10000006/issues-in-genetics/>

[8] http://www.ncbi.nlm.nih.gov/gap

[9] Shi, Xinghua, and Xintao Wu. "An overview of human genetic privacy." Annals of the New York Academy of Sciences 1387.1 (2017): 61-72.

**Project 3.**

Sharing genomic data, of course, is vital for biomedical research which can translate the research findings into clinical practice. However, genetic privacy of research participants and regular individuals is a growing concern for sharing data publicly. Because adversary attacks could be used to breach genetic privacy by disclosing personal identities or confidential traits of the individuals. For example, a genome-wide association study (GWAS) is a commonly used technique to identify associations between genetic regions (loci) and traits (including diseases). De-identification is insufficient for protecting genetic privacy. Homer’s attack [10] showed that publicly released GWAS statistics can be used to estimate a GWAS participant’s disease status from knowing his/her genotypes at certain risk factors. This study motivated the NIH to switch the genotype and phenotype data from public domain to controlled access through the database of Genotypes and Phenotypes (dbGaP). A beacon project [11] showed that first-degree relatives and individuals can be identified in the 1000GP. They also demonstrated that they can reidentify a single genome from PGP3 participants by querying the existing beacons 1000 times. Our previous work [12-14] further showed that personal traits and identities on both GWAS participants and regular individuals can be detected even in a differentially private GWAS statistics. In summary, a variety of complementary methods for privacy protection are necessary and relevant ethics education and enforcement of pertinent laws and regulations need to strengthen.

Read the following articles and describe the disclosure and protection of genetic privacy for genomic

study participants and regular individuals. What techniques will be possible to breach and ensure the privacy of GWAS participants, and regular individuals like 1000 Genomes Project individuals.

[10]. Homer N, Szelinger S, Redman M, et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet. 2008; 4:e1000167. [PubMed: 18769715]

[11]. Shringarpure SS, Bustamante CD. Privacy risks from genomic data-sharing beacons. Am. J. Hum. Genet. 2015; 97:631–646. [PubMed: 26522470]

[12]. Wang, Y., Wu, X., Shi, X. 2013 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2013. Using aggregate human genome data for individual identification; p. 410-415.

[13]. Wang, Y., Wu, X., Shi, X. Technical Report. University of Arkansas; Fayetteville, AR: Infringement of individual privacy via mining GWAS statistics. Report No. DPL-2014-004

[14]. Wang, Y., Wen, J., Wu, X., et al. Proceedings of the 2nd International Conference on Big Data Computing and Communication (BIGCOM 2016). Lect. Notes Comput. Sci. Vol. 9784. Springer; 2016. Infringement of individual privacy via mining differentially private GWAS statistics; p. 355-366.