Enabling genomic data sharing without violating patient privacy

The growing stockpiles of genomic data found in biomedical repositories and patient records promise to be an invaluable resource for improving our understanding of human diseases. This data can be used by biomedical researchers to generate hypotheses and validate new results. Recent work, however, has shown that sharing this data--even when aggregated to produce p-values, regression coefficients, or other study statistics--may compromise patient privacy. This raises a fundamental question: how do we protect patient privacy while still making the most out of their data?

In this talk we will explore two different approaches for protecting genomic data. The first approach uses a privacy preserving technique known as differential privacy. This technique, which works by slightly perturbing the data, protects patient privacy while still allowing researchers access to their genomic data. Unfortunately, many existing differentially private techniques for genome-wide analysis have limitations in terms of accuracy, computational efficiency, and their ability to deal with heterogeneous populations. I will share recent work we have done to help overcome these bottlenecks, work which moves privacy preserving genome-wide analysis closer to real world applicability.