

# How to study genetics while preserving privacy?

Jean-Philippe Vert  
Jean-Philippe.Vert@ens.fr

Collecting genetic informations about individuals by DNA sequencing or genotyping is now cheap and increasingly popular<sup>1</sup>. Sharing this information is useful to identify correlations between genetic factors and particular phenotypes like the occurrence of a disease, the response to a treatment, or an addiction. It also raises privacy concerns: would you like to release publicly your genome and the fact that you have a particular disease or addiction? Following the notion of *differential privacy* introduced 10 years ago in cryptography [1, 2], some groups have recently proposed methods and algorithms to share genetic information sufficient to perform statistical association studies, while preserving the identify of participants [3, 4]. The goal of this project is to clarify the notion of differential privacy and the techniques to achieve it in the context of genetic studies. It requires no prior knowledge in algorithms or biology, but basic notions of probability and statistics; motivated students may even want to implement and test some algorithms on real genetic data.

## References

- [1] C. Dwork. Differential privacy. In *Proceedings of the International Colloquium on Automata, Languages and Programming (ICALP)*, pages 1-12. 2006.
- [2] C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211-407, 2014.
- [3] C. Uhler C. et al. Privacy-preserving data sharing for genome-wide association studies. *J. Priv. Confid.*, 5, 137?166, 2013.

---

<sup>1</sup>23andme.com

- [4] S. Simmons and B. Berger. Realizing privacy preserving genome-wide association studies. *Bioinformatics*, 32(9):1293-1300, 2016.