

2024 PhD Projects

Project title	Studying the Genetic and Phenotypic Diversity of South Asia
Principal supervisor	Kaustubh Adhikari
Second supervisor	
Discipline	Statistics
Research area/keywords	Statistical genetics, Machine Learning, Multivariate analysis
Suitable for	Full time applicants, Part time applicants

Project background and description

South Asia is one of the world's most populated regions, with enormous diversity in appearance genetics. In recent years, the field of human genetics has grown immensely in terms of data production and analysis, now routinely analysing data from thousands of people over millions of genetic markers [1] and thousands of variables [2]. Yet, South Asians are one of the least-studied groups of people [3]. Even in the developed nations with large cohorts for genetic research, such as the UK, the South Asian diaspora is massively underrepresented [4].

The proposed PhD project will address this imbalance by exploring various datasets, including diverse public datasets, to explore the genetic and phenotypic diversity of South Asia. The project will develop multivariate statistical models for the analysis of large numbers of genetic markers (genotypes) and physical characteristics (phenotypes), such as skin and eye colour [5]. Latest methods that are suitable for high-dimensional data, spanning both AI and machine learning domains [6,7] and classical statistical procedures [8,9], will be explored. Such methods will be useful in understanding the genetic and population history of the region [10], discovery of new genes associated with new phenotypes [5], and also better prediction of physical characteristics from genetic data, useful for forensic reconstructions [11].

There will also be the scope of studying ancient human samples by studying genetic data from ancient DNA obtained from prehistoric human remains [10], to better understand the demographic history of this region.

As an applied Statistics project, it will involve both theoretical and computational work. The candidate should have a suitable knowledge of statistics with a suitable degree. Computational (R / Python etc.) and programming experience would be useful. No prior knowledge of genetics or biology is required – all necessary training will be provided – though some understanding of the context might be useful. A degree in the area of statistical genetics, for example, would be useful, though this is not essential.

An associated theme of the work would be about conversations around increasing the diversity of genetics research [4], discussing the gains brought about in science by broadening the perspective. There will be substantial scope for the student to engage in outreach and public engagement, if interested.

Background reading/references

1. Thompson, S. G., Willeit, P., UK Biobank comes of age. *Lancet*, (2015). 386(9993): p. 509-10.

2. Claes P, et al., genome-wide mapping of global-to-local genetic effects on human facial shape. *Nat Genet* (2018); 50:414–23.
3. Sirugo, G. et al., The Missing Diversity in Human Genetic Studies. *Cell* 177, (2019).
4. The GRACE project. <https://open.ac.uk/research-projects/grace/>
5. Adhikari, K., et al., A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features. *Nat Commun*, (2016). 7: p. 10815.
6. Giardina A, Paria SS, Adhikari K, A naive method to discover directions in the StyleGAN2 latent space. *arXiv:2203.10373* (2022).
7. Yelmen B, Jay F, An Overview of Deep Generative Models in Functional and Evolutionary Genomics. *Annual Review of Biomedical Data Science* (2023), 6:173-189.
8. Turley, P. et al., Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* 50, 229–237 (2018).
9. Lloyd-Jones, L.R., Zeng, J., Sidorenko, J. et al., Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat Commun*, 10, 5086 (2019).
10. Narasimhan, V.M. et al., The formation of human populations in South and Central Asia. *Science*, 365, eaat7487 (2019).
11. Palmal, S., Adhikari, K. et al., Prediction of eye, hair and skin colour in Latin Americans. *Forensic Science International: Genetics*, Volume 53, (2021).