



**Application Number / Title:** 21413 - Phasing human populations

**Applicant PI:** Professor John Hickey

**Applicant Institution:** University of Edinburgh, [Genetics and Genomics, The Roslin Institute, Easter Bush, Edinburgh, EH25 9RG, United Kingdom](#)

**Keywords provided by the Applicant PI to describe the research project:**  
Imputation Phasing Haplotyping

**Application Lay Summary:**

**1a:** Genotype phasing and imputation is an important step in many genetic studies. Genotyping technology normally produce unphased genotype and do not provide the underlying haplotypes. Inferring these haplotypes can improve many genetic studies as they provide additional information. Accurately inferred haplotypes can also improve genotype imputation.

Genetic studies benefit from genotype imputation as the inclusion of imputed markers increases the power of many of the tests involved in such studies. We propose to investigate whether our genotype phasing and imputation software can outperform existing methods when imputing on human datasets.

**1b:** Our intention is to improve the available phasing and imputation methods for human datasets and so maximize the benefit to researchers investigating the association between health and genetics.

We have developed AlphaImpute, a tool for phasing and imputing missing genotypes. To date AlphaImpute has been successfully used on animal data but we believe our methods have the potential to outperform existing methods on human data. By optimising our software for use in a human context we hope to improve its performance further and, by doing so, we will maximise its contribution to human health genetics research.

**1c:** Our software, AlphaImpute, can phase and impute large numbers of genomes with high accuracy. We have tested AlphaImpute on many animal species and now wish to test it on a human dataset.

We propose that we would first test the performance of AlphaImpute on the genetic data of the full cohort. This testing is likely to identify areas where we could improve our method to better account for the unique challenges of a human dataset.

**1d:** We propose that we would test the performance of AlphaImpute on the genetic data of the full cohort. By using the full cohort we would maximise the usefulness of our imputation to other researchers.

#### PROJECT EXTENTION

Building upon our current UK Biobank project for phasing and imputing data, this proposal concerns downstream research using accurately imputed genotypes to improve the association between genotypes and phenotypes. Our long-term aim is to improve the identification of causal mutations that impact the population health, using genomic information to enable biological insight in complex diseases with societal and economic implications.

After accurately phasing and imputing the UK Biobank genotypes we will use funBayesB, a recently developed method for genome-wide association analysis. funBayesB is a single-step Bayesian method that uses known structural and functional annotation of the SNP markers to determine the most likely causal mutations. Access to UK Biobank phenotypic data will allow us to validate funBayesB, and will contribute to unraveling the complexities between genome and diseases.

For this stage of the propose we would require the following phenotypes:

- Standing height (Field ID 50)
- Body mass index (Field ID 21001)

The outputs from this work will be published in peer review open source journals, we will provide a list of most likely causal mutations associated with these phenotypes and funBayesB will be freely available to academics.