

Principal Investigator

Dr. Alkes Price

Address

Harvard School of Public Health, Epidemiology, 655 Huntington Avenue,
Building 2 Room 211, Boston, MA02115, United States

Summary of research

Key words: Haplotype, Phasing, Imputation

Application Lay Summary:

1a: Most genetic assays produce unphased diploid genotypes, i.e., data in which the maternal and paternal contributions are unknown. Inferring the pattern of inheritance (phase) across large chromosomal segments is an important step in conducting genome-wide association studies to understand disease genetics, as it enables imputation of untyped markers. We aim to develop new statistical methods to perform faster and more accurate phasing and imputation by harnessing shared genetic material found in close and distant relatives in very large data sets. We will apply these methods to UK Biobank genotypes and contribute the results for use by other researchers.

1b: The target quality of the phased and imputed data we aim to produce should exceed that achievable with existing techniques and thus will be of immediate interest to researchers performing genome-wide association studies of health-related outcomes.

1c: We will begin by applying fast search heuristics to identify likely identity-by-

descent (IBD) segments among all pairs of individuals. We will do so by filtering to long segment pairs with few or no opposite homozygous sites and then evaluating an approximate likelihood ratio of IBD vs. non-IBD. This procedure will yield long partially phased genomic segments, which we will refine by comparing against one another. We will evaluate performance using gold standard phasing from parent-child trios.

1d: We will analyze the full cohort.