

PgmNr 1282: Ultra-rapid detection of identity-by-descent tracts for biobank-scale inference

Ruhollah Shemirani¹, Gillian M Belbin², Eimear E Kenny², Chris R Gignoux³, Jose Luis Ambite¹, on behalf of the PAGE-II Consortium

1 Information Sciences Institute, University of Southern California

2 The Charles Bronfman Institute for Personalized Medicine, [Icahn School of Medicine at Mount Sinai](#).

3 Colorado Center for Personalized Medicine, University of Colorado, Aurora, CO

The identification of identity-by-descent (IBD) tracts is a standard step in many pipelines for genome-wide studies in both population and medical genetics. With novel mega-scale initiatives ongoing from biobanks and nation-wide resources there is an ongoing need for efficient methods to detect these IBD segments for identification of cryptic relatedness, population linkage, and demographic inference. While IBD detection has been a topic of vigorous development, existing algorithms scale poorly, exhibiting emergent quadratic or higher order run time as study size becomes large and IBD segments become commonplace. This makes integration into large studies such as across health systems or national biobank initiatives intractable. Here, we present a novel algorithm for mega-scale IBD detection via partitioning and locality sensitive hashing, allowing close-to-linear time scaling and enabling IBD segment detection genome-wide in studies of hundreds of thousands of samples and beyond. Our algorithm, called I-LASH (Ibd by LocAlity Sensitive Hashing), provides a pipeline-ready approach to IBD detection, uses standard file formats, and contains flexible tuning parameters for arrays or sequencing. By leveraging parallelism and our computationally efficient design I-LASH allows entire mega-scale studies to be analyzed on a standard multi-core machine. We benchmark I-LASH on simulated datasets and across reference datasets including the ethnically diverse >51,000 individuals genotyped in the PAGE-II Consortium. While maintaining similar precision and recall metrics to the standard GERMLINE method, I-LASH provides a >40-fold speedup compared to GERMLINE with 50,000 samples. We anticipate that I-LASH will empower IBD inference for the next generation of medical and population genetic studies.