Haplotype Phasing with Single Genome Amplification

Christine Lo, Nitin Udpa, and Vineet Bafna

University of California, San Deigo, La Jolla, CA 92093, USA
cylo@cs.ucsd.edu

Abstract. Knowing haplotypes is important for many genomic applications such as association studies, inferring evolutionary history, and computing recombination rates. With Single Genome Amplification (SGA) technology, we are able to get long, but sparse haplotypes spanning the entire chromosome. In SGA, whole metaphase chromosomes are extracted from a human cell and diluted. The molecules are then placed in different tubes such that each tube contains, on average, one copy of each chromosome. Single genome amplification is then performed on each tube, followed by genotyping/sequencing. Here we discuss a computational method to find a consensus haplotype given the output data of the SGA protocol. We show that with SGA technology, we can achieve accurate, but sparse haplotypes that span the whole chromosome. We also show that population haplotype data and sequencing data can be used to impute or "fill in the gaps" of SGA haplotypes and provide denser, chromosome level haplotypes.

Keywords: Haplotype, Phasing, Single Genome Amplification, Whole Chromosome Haplotype