#### Genotype Calling and Haplotyping in Parent-Offspring Trios

Wei Chen1,\*, Bingshan Li1, Serena Sanna2, Carlo Sidore1,2,3,

Fabio Busonero1,2, Yun Li4 and Goncalo Abecasis1,\*

*1 Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI; 2 Istituto di Ricerca Genetica e Biomedica, Centro Nazionale di Ricerca, Monserrato, Sardinia, Italy, 3 Dipartimento di Scienze Biomediche, Università di Sassari, Cagliari, Italy, 4 Department of Genetics, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC*

Emerging sequencing technologies allow common and rare variants to be systematically assayed across the human genome in many individuals. In order to improve variant detection and genotype calling, raw sequence data is typically examined across many individuals. Here, we describe a method for genotype calling in settings where sequence data is available for unrelated individuals and parent-offspring trios and show that modeling trio information can greatly increase the accuracy of sequence based genotypes and haplotypes. Our method considers both linkage disequilibrium (LD) patterns and the constraints imposed by the family structure in assigning individual genotypes and haplotypes and should facilitate genotype calling and haplotype inference for many ongoing sequencing projects. Using simulations, we show trios provide higher genotype calling accuracy across the frequency spectrum, both overall and at hard-to-call heterozygous sites. In addition, trios provide greatly improved phasing accuracy —improving the accuracy of downstream analyses (such as genotype imputation) that rely on phased haplotypes. We evaluated our method using data generated by an ongoing sequencing study focused on a Sardinian population. Our results show that our method reduces the genotyping error rate by 50% compared to analysis using existing methods that ignore family structure.