B12 Confirming Statistical Phased Single Nucleotide Polymorphisms (SNPs) Haplotype Data of 74 Microhaplotypes (MH) Across a Global Set of Populations by Massively Parallel Sequencing (MPS)

Fabio Oldoni, PhD, The George Washington University, Washington, DC 20007; Leena Yoon, Tysons, VA 22102; Sathya Prakash Harihar, The George Washington University, Department of Forensic Science, Washington, DC 20007; Aishwaryaa Subramanian, The George Washington University, Washington, DC 20052; Drew A. Bader, AB, Washington, DC 20007; Sharon C. Wootton, PhD, South San Francisco, CA 94080; Robert Lagacé, BS, Thermo Fisher Scientific, South San Francisco, CA 94080; Ryo Hasegawa, BS, Foster City, CA 94404; Joseph P. Chang, BS, Thermo Fisher Scientific, South San Francisco, CA 94080; Kenneth Kidd, PhD, Yale University School of Medicine, New Haven, CT 06520; Daniele S. Podini, PhD, The George Washington University, Department of Forensic Science, Washington, DC 20007*

Learning Overview: After attending this presentation, attendees will be able to understand the potential of using Massively Parallel Sequencing (MPS) technology for acquiring haplotype data of short DNA regions containing microhapolotypes (MHs).

Impact on the Forensic Science Community: This presentation will impact the forensic science community by demonstrating the utility of using an emerging sequencing approach to generate accurate haplotype information.

Microhaplotypes are loci of two or more single nucleotide polymorphisms (SNPs) within a short distance from each other (< 300 nucleotides) with three or more allelic combinations.[1] These multi-SNP loci have small amplicon size, no stutter peak, and lower mutation rate than short tandem repeats (STRs), which make them candidate markers for human identification, mixture deconvolution, and ancestry prediction.[2] Although different approaches including fast TaqMan® assay and conventional Sanger sequencing can be used for SNP typing, these methods do not allow determining the *cis/trans* relationship between individual SNP alleles (i.e., phase) within the same amplicon. Consequentially, statistical Bayesian-based software such as PHASE, designed to reconstruct haplotype patterns from genotype data, while accurate on a population level has the potential to make mistakes on an individual level and for rare alleles.[3] Also if there are other SNPs within the target region that vary, these will be undetected using such approach as it focuses only on the selected SNPs that define the locus. Conversely, high-throughput sequencing technology enables the specific cloning and sequencing of each individual DNA strand, and thus distinguishing the parental haplotypes at a given locus, while also detecting other potentially discriminating SNP variations within the entire region sequenced. For forensic implementation of MHs, allele population frequencies are critical as they vary across populations and significantly more than STRs. Here, the authors focused on the comparison of statistically phase-inferred haplotypes to MPS determined haplotypes generated from 74 MH loci genotyped on a large set of population samples.

In this study, the authors selected a worldwide population set of 539 samples representative of Africa (40 Sandawe, 38 Hausa), Europe (48 Danes, 53 Khanty), South Central Asia (119 Laotians, 30 Keralites), East Asia (54 Koreans, 42 Atayal), Native America (94 Mexican Pimas) and Oceania (21 Papua New Guineans). All population samples were genotyped using the TaqMan® assay and SNP haplotypes of 74 MH loci computationally inferred by PHASE. To determine the exact phase of parental SNP haplotypes at each MH locus, the authors genotyped all DNA samples using a newly-developed MPS multiplex panel of 74 MH markers implemented on the Ion Chef™ and Ion S5™ (Thermo Fisher Scientific) platform.[4] The MPS panel of 74 MH loci totalling 230 SNPs was specifically developed for enhancing ancestry prediction and mixture deconvolution capabilities.

As expected, PHASE provided accurate haplotype prediction at the individual level when SNPs were homozygous within a MH allele and when no more than one site was heterozygous. A few differences were identified between statistically phased haplotypes and MPS determined haplotypes in haplotypes with multiple heterozygous SNPs within an individual. These were due either to missing/mistyped SNPs by TaqMan® assay or incorrect phase estimation by PHASE.

The preliminary findings indicate that overall statistical phasing provides accurate haplotype reconstruction for population allele frequency inference, particularly for common alleles. This supports the fact that computational haplotype phasing is a valuable and inexpensive screening approach as it can use databases that are already available (e.g., ALFRED).[5] However, going forward these are likely to be replaced by the confirmatory MPS methodology in light of its increasing cost-effectiveness and more importantly as it allows the detection of rare haplotypes caused by variations within the target region of SNPs not originally used to define the locus.

Reference(s):

1. Kidd K.K., Pakstis A.J., Speed W.C., Lagacé R., Chang J., Wootton S., Haigh E., Kidd J.R.. Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics. *Forensic Science International: Genetics* 12 (2014) 215–224.
2. Kidd KK, Speed WC, Pakstis AJ, Podini DS, Lagacé R, Chang J, Wootton S, Haigh E, Soundararajan U. Evaluating 130 microhaplotypes across a global set of 83 populations. *Forensic Science International: Genetics* 29 (2017) 29-37.
3. Stephens M., Smith N.J., Donnelly P. A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* 68 (2001) 978–989.
4. Oldoni F, Hart R, Long K, Maddela K, Cisana S, Schanfield M, Wootton S, Chang J, Lagace R, Hasegawa R, Kidd K, Podini D, Microhaplotypes for ancestry prediction (2017) *Forensic Science International Genetics Supplement Series* 6: e513-e515
5. https://alfred.med.yale.edu

Microhaplotypes, Massively Parallel Sequencing, Phased and Sequenced Haplotypes