MICROHAPLOTYPES: A POWERFUL NEW CLASS OF FORENSIC MARKERS

Kenneth K. Kidd[1], Andrew J. Pakstis[1], William C. Speed[1], Robert Lagacé[2], Joseph Chang[2], Sharon Wootton[2], Eva Haigh[1], Judith R. Kidd[1]

[1]Department of Genetics, Yale University School of Medicine
[2]Human Identification Group, Thermo Fisher Scientific

In mass disaster and missing person cases, scientists often need to link unknowns to a family. We are proposing that haplotypes of SNPs can become the markers of choice for such tasks. Microhaplotypes (two or more SNPs within a span of <200bp) are a powerful new type of forensic marker that can be genotyped by NGS (Now Generation Sequencing) since 200bp is within the span of current desktop sequencing platforms.

We have now fully documented the population genetics globally of over 45 microhaplotype loci selected to define multiple alleles in over 2500 individuals from 54 different populations. All have multiple alleles (haplotypes) and almost all have global average heterozygosities >0.5 (many over 0.6) and over 74% of individual heterozygosities are greater than the 0.5 maximum possible for any single SNP. Ongoing research on 14 selected microhaplotypes comprised of 4 SNPs is yielding nearly 97% of heterozygosities >0.5 and a global mean of over 0.7. These high heterozygosities make microhaplotypes a highly efficient type of forensic marker for typing by sequencing. Large numbers of microhaplotype loci can be multiplexed at affordable costs allowing high statistical power, much greater than even the expanded set of CODIS STR polymorphisms. The microhaplotypes we have fully characterized also give low match probabilities and allow ancestry inference. We have easily identified in HapMap and 1000 Genomes an additional several dozen such loci now being studied on our population samples. The phase-known data from a forensic sample will detect mixtures for these multialleleic loci allowing deconvolution not possible with single SNPs. Thus, microhaplotype loci constitute a statistically powerful new type of genetic marker ready for forensic applications using existing sequencing methods.