

# A novel noninvasive prenatal paternity test using microhaplotypes

Jaqueline Yu Ting Wang<sup>1</sup>, Anatoly Yambartsev<sup>2</sup>, Renato Puga<sup>3</sup>, Martin R. Whittle<sup>4</sup>, André Fujita<sup>2</sup>, Helder Takashi Imoto Nakaya<sup>1</sup>,

*1 University of São Paulo*

*2 Institute of Mathematics and Statistics - USP*

*3 Hospital Israelita Albert Einstein*

*4 Genomic Engenharia Molecular*

## Abstract

Paternity tests are usually done by analyzing DNA samples from the alleged father, the mother, and the child. To perform this exam before the birth, invasive methods such as amniocentesis and chorionic villus sampling are usually used. Fortunately, the discovery of fetal DNA (fetal cell-free DNA, fcfDNA) in maternal plasma and serum, and the development of techniques to analyze this fcfDNA have allowed researchers to reduce this risk for both fetus and mother. Although paternity tests that analyze Short Tandem Repeats (STRs) from fcfDNA are possible, they are not reliable because DNA degradation often occurs. SNPs (Single Nucleotide Polymorphisms) have been demonstrated to be good candidates for human identification and they can be obtained from small DNA fragments (even from degraded DNA). To increase the number of possible genotypes and decrease the amount of analyzed SNPs, our analyzes focus on microhaplotypes. Microhaplotypes are chromosomal segments smaller than 200 base pairs (bp) containing two or more SNPs that form at least three distinct haplotypes. Since fcfDNA has approximately 145 bp, this is sufficient to contain microhaplotypes that can be sequenced using Next Generation Sequencing (NGS) technology. The aim of this project is to determine the probability of paternity using SNPs within microhaplotypes. Microhaplotypes were chosen based on previous literature review. The haplotypes frequencies were calculated based on the ethnic groups from 1000 Genomes database. To accomplish this objective, raw DNA sequence data from three DNA samples were analyzed: the alleged father, the mother, and the maternal plasma (mixture of mother and fetus cell-free DNA). Then, using a script developed based on SAMtools and Perl programming language, we obtained the genotypes of the alleged father and mother, for each microhaplotype. By combining genotypic information, population frequencies, and fetal fractions (plasma), we developed a method to calculate the probability of paternity in cases of non-exclusion.

Funding: Genomic Engenharia Molecular