Noninvasive prenatal paternity determination by SNPs and microhaplotypes

Jaqueline Y. T. Wang[1,2]; Renato D. Puga[3]; Martin R. Whittle[4]; André Fujita[2,5]; Helder T. I. Nakaya[1,2].

1 - School of Pharmaceutical Sciences - Department of Clinical Chemistry - University of São Paulo; 2 - Bioinformatics Graduate Program - University of São Paulo; 3 - Hospital Israelita Albert Einstein; 4 - Genomic Engenharia Molecular; 5 - Institute of Mathematics and Statistics - Department of Computer Science - University of São Paulo.

Invasive procedures like amniocentesis and chorionic villus sampling can be risky for pregnancy and may result in miscarriages. However, when information from the fetus is needed to investigate congenital abnormalities (such as sex-linked disorders and aneuploidy) or perform paternity tests, this risk is usually acceptable. Fortunately, the discovery of fetal DNA (fetal cell-free DNA, fcfDNA) in maternal plasma and the development of techniques to analyse this fcfDNA have allowed researchers to reduce this risk to fetus and mother. Microhaplotypes are chromosomal segments smaller than 200 bp, with at least three distinct haplotypes (alleles). Their average heterozigosity has to be greater than that of any of the Single Nucleotide Polymorphisms (SNPs) contained within them. Since the fcfDNA has a size of 166bp, it is sufficient to contain microhaplotypes which 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. Raw sequencing data from three DNA samples are analysed: the alleged father, the mother and the maternal plasma (mixture of mother and fetus cell-free DNA). We performed sequencing quality control with FastQC and verified sequencing coverage with BEDtools. An optimum sequencing coverage is vital to infer the fetal fraction in the maternal plasma sample. Next, variant calling was performed with SAMtools, BCFtools and VCFtools. Finally, SNP annotation was done with ANNOVAR. All analysis steps were performed sequentially by using a pipeline written in Perl programming language. Microhaplotypes were chosen based on previous literature and human frequencies were calculated using data from 1000 Genomes. Combining genotype information, populational frequencies and fetal fractions, we will develop a method to calculate the probability of paternity in non-exclusion cases. Twenty microhaplotypes believed to be very informative for paternal investigation have been chosen based on their heterozigosity. The microhaplotypes frequencies have been calculated based on all the ethnic groups from 1000 genomes data. Currently, 30 sets of the three types of samples have been processed and annotated.