HaploCYP: a software for *CYP2D6* genotyping and phenotype prediction

Pereira MA[1], Cardenas RGCCL[2], Mateo ECC[1], Ferreira ACS[1], Freire MCM[1]

[1]Setor de Pesquisa e Desenvolvimento, Grupo Hermes Pardini, Vespasiano, Brasil; [2]Torchmed - Software Development, Belo Horizonte, Brasil; [3]Laboratório Progenética, Grupo Hermes Pardini, Rio de Janeiro, Brasil

Cytochrome P450 2D6 (*CYP2D6*) gene (MIM#124030) is one of the most polymorphic pharmacogenes with more than 109 allelic variants reported up to date by the Human Cytochrome P450 (*CYP*) Allele Nomenclature Database. The polymorphisms include SNPs, indels, copy number variations (CNVs), conversions and gene rearrangements. Due to this genetic polymorphism, *CYP2D6* exhibit notable inter-individual variability in enzyme activity and individuals can be divided into four phenotypic groups: poor (PM), intermediate (IM), extensive (EM) and ultrarapid (UM) metabolizers. Since *CYP2D6* genotype assignment and phenotype prediction are complex and of utmost importance into clinical practice, this work aimed to develop and validate a user-friendly software for *CYP2D6* genotyping and phenotype prediction using Sanger sequencing and CNV data. HaploCYP combines a set of python modules, BLAST tool and a MySQL database system in a web interface running on an Apache web server. The workflow consists of: BLAST alignment, polymorphism detections, genotype annotation and phenotype prediction. CNV information and fasta files from Sanger sequencing of *CYP2D6* gene are given as input. Variants are detected through BLASTn alignment with *CYP2D6**1 reference sequence (Accession Number: AY545216.1). Mutation nomenclature and haplotypes are defined according to *CYP* Allele Nomenclature Committee, PharmGKB and LOVD databases. The haplotypes that best represents the set of polymorphisms and CNV information are defined and reported following the star-allele nomenclature system. In some cases, more than two haplotypes can be reported. At this point, the user needs to review the haplotypes proposed and choose the correct genotype. Then, the phenotype is predicted as PM, IM, EM or UM, based on *CYP2D6* diplotypes and the activity score system recommend by the Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for codeine therapy. Validation included simulation (sequences from NCBI) and ten real data with known *CYP2D6* genotypes. HaploCYP was able to genotype correctly all the simulation datasets and nine real data. For one sample, the program was not capable to do genotyping annotation since two genotypes were equally likely. However, the expertise of the user was enough to solve this genotyping problem. A higher dataset will be used to test accuracy, specificity and sensibility. HaploCYP simplify and facilitate the genotyping and phenotype prediction process of *CYP2D6* into clinical diagnosis, where speed and precision is of high importance. In addition, HaploCYP can be used for others *CYP* locus and any pharmacogenes that it is necessary to identify pharmarcogenomic variants to individualize drug prescription and analyze drug efficacy and safety.

Financial Support: Hermes Pardini Group, FINEP