Investigation of mutations in the *HBB* gene using the 1000 GENOMES databank

Tânia Carlice-dos-Reis, Jaime Viana, Fabiano Moreira Cordeiro, Greice de Lemos Cardoso, João Guerreiro, Sidney Santos, Ândrea Ribeiro-dos-Santos

Laboratory of Human and Medical Genetics, Institute of Biological Sciences, Federal University of Pará, Belém, PA, 66.075-110, Brazil; Federal Rural University of the Amazon, Capanema Campus, PA, 66.077-830, Brazil; Research Center of Oncology, Federal University of Pará, Belém, PA, 66.073-005, Brazil.

Sickle-cell disease is one of the most common monogenic diseases worldwide, caused by mutations in the HBB gene (β-globin). Due to its high prevalence, various strategies have been developed to better understand its molecular mechanisms. In silico analysis has been increasingly used to investigate genotype-phenotype relationship of many diseases, and the sequences deposited in the 1,000 Genomes database, of healthy individuals, appears to be an excellent approach for this analysis. This study aims to analyze the variations of the HBB gene in the 1,000 Genomes database, as well as investigate the pattern of pathogenicity, and describe the mutations frequencies in the different population groups. The computational tool SnpEff was used to select HBB mutation identified among 2,504 samples from 1,000 genomes. Nucleotide mutations, amino acid changes, allelic and population frequencies, and type of mutations were visualized using the IGV software. The pathogenicity of each amino acid change was investigated using the databases CLINVAR, dbSNP and five different predictors (POLYPHEN, SIFT, PROVEAN, PANTHER e MUTPRED). Pathogenic mutations of HBB, according to the predictors, that were not identified on CLINVAR database were 3D modeling in the PDB database, to infer the effect of the mutation on protein function. Were found 20 different types of mutations in 209 individuals, where 173 subjects had missense mutations. The African population group presented the highest number of mutated individuals 153 individuals, and European presented the least (9 individuals). According to the results, 70% of the mutations were pathogenic. The constructed 3D model allows to visualize residues that have undergone mutation and location of each of the protein. It is concluded that approximately 8.3% of phenotypically healthy individuals from the database 1,000 Genomes have some mutation in the HBB, of which 70% are pathogenic. The mutations are unequally distributed among the populations, being the most affected the African population (73.2% of subjects) and European population the less affected (4.3% of subjects). Pathogenic mutations with greater allele frequencies (rs334, rs33930165 and rs33950507) are known to cause sickle-cell disease and β-thalassemia.