CNV calling and its characterization in the Brazilian population

Ana Claudia Martins Ciconelle[1], Júlia Maria Pavan Soler[1],

1 Instituto de Matemática e Estatística - IME/USP

Abstract

A copy number variation (CNV) occurs when the number of copies of a particular region of the DNA differs from two in autosomes or one/two in allosomes and has an important role in the genetic variability in humans. The effects of CNVs to human diseases are not yet known, although several diseases have been associated to this kind of polymorphism, such as uric acid and nervous system disorders. Motivated by the unknown influence of CNVs on anthropometric measurements and cardiovascular phenotypes and in collaboration with the Laboratory of Genetics and Molecular Cardiology at the Heart Institute/InCor-FMUSP, the primary aim in this project is to estimate the CNV from SNP array platforms and understand its transmission rate in family data and its association with complex phenotypes. This project also aims to understand the CNV distribution in Brazilian population and between family members. A pipeline was proposed for CNV calling from SNP Array data by reviewing softwares and packages that are available in the literature. Using the database from the Baependi Heart project, we analyzed the genotype and phenotype data from 80 families to identify the CNVs and to understand their association with height. From the genetic data, the CNVs were estimated from a combination of statistical techniques and algorithms including quantile normalization, classification methods for genotyping of SNPs and hidden Markov chains. After an exploratory data analysis, polygenic linear mixed models were used to estimate the association of CNVs with the chosen phenotypes. Our results suggest that the Brazilian population have a similar number of CNV per person (around 55 CNVs) in comparison with other populations. However, from the total of 64.107 CNVs identified, 147 CNVs are common among our samples, but rare in worldwide populations, one example being a CNV in the NEGR1 gene, which is present in 89% of our sample. Based on family data, the intraclass correlation coefficient for CNVs was estimated between 30% to 60% showing a high similarity on values from the same family. Association analysis between CNVs and different phenotypes are being performed.

Funding: CNPq, CAPES