
# DEVELOPMENT OF A PIPELINE FOR CNV DETECTION AND ANALYSIS USING DATA FROM SNP ARRAYS

*Fernanda Cristina de Paiva Pereira, Roberto Hirochi Herai, André Robles Gonçalves, Luciana Correia de Almeida Regitano, Poliana Fernanda Giachetto*

Embrapa Informática Agropecuária, Embrapa Informática Agropecuária, Embrapa Informática Agropecuária, Embrapa Pecuária Sudeste, Embrapa Informática Agropecuária

A copy number variation (CNV) is defined as a genomic region where the DNA copy number differs between two or more individuals from a population. In humans, some of these CNVs have been shown to be important in both normal phenotypic variability and disease susceptibility, and several publications have reviewed the effects of CNVs on gene expression and association with important human diseases. In livestock species, the characterization of this genetic variation is an important step toward linking genes or genomic regions with phenotypic traits, particularly those of economic importance. At Embrapa, we have used SNP chips for bovine genotyping in genome wide association studies, with focus on meat quality and endo-and ectoparasite resistance. The objective of this study was to develop a bioinformatics pipeline for CNV identification and analysis from data generated by high density SNP chips genotyping from Illumina platform, which has been extensively used to genotyping cattle at Embrapa. Were used data from 400 bovines (Canchim cattle), part of an Embrapa breeding program, genotyped with the BovineHD BeadChip (Illumina). The pipeline was based on PennCNV tool for CNV identification, ANNOVAR tool for CNV annotation and programs and scripts developed in Perl to conversion of input files and visualization of output files. Analyzing the data set above in our pipeline, a total of 5,684 CNVs in 192 Canchim DNA samples was detected, with an average size of 578,159bp. We are now identifying the number of CNV regions (CNVRs), by merging overlapping CNVs, and identifying the overlapping genes and GO terms enriched among CNVs. The next steps involve the visualization of identified CNVs in a Genome Browser and the inclusion of the pipeline on Galaxy platform, to be widely used by the scientific community of the area.

**Keywords**: CNV, bovine, PennCNV, ANNOVAR
**Concentration area**: Genomics Evolution