COPY NUMBER VARIATION DISCOVERY IN CANCHIM CATTLE USING DATA FROM SNP GENOTYPING ARRAYS

André Robles Gonçalves, Fernanda Cristina de Paiva Pereira, Luciana Correia de Almeida Regitano , Poliana Fernanda Giachetto

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

Copy number variations (CNVs) are gains and losses of genomic sequences between 2 or more individuals from a population. In humans, it has been shown that CNVs are involved 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 and complex disorders such as autism, schizophrenia and Parkinson´s disease. In livestock species, it has been observed that CNVs can affect a wide range of phenotypic traits of economic importance, including meat and milk production, meat quality and disease and parasite resistance. Several techniques have been developed to detected CNVs in genomes, including whole-genome resequencing and SNP genotyping arrays. Data from cattle genotyping using SNP arrays for genome-wide association studies, is widely available at Embrapa. The objective os this study was to use an open source tool, CNstream, for CNV identification from cattle genotyping using Illumina SNP arrays. Were used data from 400 bovines (Canchim cattle), part of an Embrapa breeding program focused on meat quality, genotyped with the BovineHD BeadChip (Illumina). A total of 5,789 CNVs were detected, being 4,536 deletions and 1,253 duplications. CNVs length varied from 2,251 to 99,494 bp. Several authors have noted a large variability between CNV-calling algorithms, as well as a substantial false positive and false negative rates associated with the methods. The same set of samples was analysed with PennCNV, one of the most used CNV methods for Illumina platform and we have found a few different results, which can be attributed to methodological limitations of the tools that were used. We suggest the adoption of more than one tool for CNV discovery in a pipeline, and the use of overlapped CNVs between them as a more reliable CNV prediction. Supported by: Embrapa

Keywords: CNV discovery, Canchim, SNP array, CNstream

Concentration area: Genomics Evolution