In silico identification of the effects of genetic variants in transcription factors recognition sites in regulatory regions of candidate genes for reproductive disorders in cattle

Diniz, LAF[1]; Fonseca, PAS[1]; Paiva, AE[1]; Santos, FC[1]; Rosse, IC[1]; Moura, GS[2]; Santos, DJA[3]; Oliveira, G[4,5]; Andrade, VJ[2]; Vale-Filho, VR[2]; Silva, MVGB[3]; Carvalho, MRS[1]

[1] Departamento de Biologia Geral, Universidade Federal de Minas Gerais, Belo Horizonte; [2] Escola de Veterinária, Universidade Federal de Minas Gerais, Belo Horizonte; [3] Embrapa Gado de Leite, Juiz de Fora; [4] Grupo de Genômica e Biologia Computacional, Centro de Pesquisas René Rachou (CPqRR) - FIOCRUZ, Belo Horizonte ; [5] Vale Technology Institute, Belém, PA, Brazil

Identify genetic variants with high functional potential among all the variants identified in a whole genome sequencing is a hard task. Variants in regulatory regions have a great functional potential in multifactorial traits. The frequency of reproductive disorders increase with the intensity of selection for production traits in bovine herds, suggesting a regulatory component shared in both processes. In the present study, we propose a pipeline for identification of variants with high functional potential in regulatory regions (5'-UTR and promoter), using as example a dataset composed by variants identified in a bull affected by reproductive disorders. To reach this aim, the variants present in the affected animal were compared with variants identified in three healthy Gir bulls. After this step, only the variations observed exclusively in the affected animal were kept for the next analyses. The exclusive variants were subjected to a functional annotation using the software NGS-SNP. The variants mapping on 5'-UTR and upstream regions were evaluated respective to their co-location with QTLs for reproductive traits. This step was developed using scripts in R and Pearl, developed in house. Three variants in 5'-UTR and six variants in upstream regions were identified in candidate genes for male reproductive traits. The selection of the variants with the highest functional impact was based on the following criteria, using these bioinformatics tools: 1) genomic context (MatInspector and NCBI); 2) evolutionary conservation (ECR Browser and Mulan); 3) alterations in transcription factor recognition sites (Mulan and MatInspector); 4) alterations in the probability of recognition of a transcription factor-binding site (Cister); 5) ratio of co-expression in male reproductive tissues of the target gene and the transcription factors (BioGPS); and, 6) involvement of the candidate gene and transcription factors in biological processes related to reproductive disorders (scientific literature). At the end of these analyses, three genetic variants with high functional potential were identified, mapped in the upstream region of three important positional and functional candidate genes for reproductive disorders in bovines (*DGAT1*, *ACTN1* and *INHBA*). Using this approach it was possible to filter variants from a whole genome sequencing (over 3 million), associate those variants with QTLs for male reproductive traits (over 180.000) and, finally, select those with the highest functional potential in important candidate genes for reproductive phenotypes (3 variants). Furthermore, suggesting that the approach proposed is efficient to identify variants with high functional potential for complex traits.