



☒ **IMPLEMENTAÇÃO DE 4 MÉTODOS DE ANÁLISE DE ASSOCIAÇÃO GENÔMICA AMPLA
BASEADA EM CONJUNTO DE GENES**

Aline Taise **Guerreiro**¹; Roberto Hiroshi **Higa**²

Nº 15604

RESUMO - No contexto de estudos de associação genômica ampla (GWAS), métodos para a análise de enriquecimento de um conjunto de genes (GSEA) analisam conjuntos de SNPs associados a genes que compartilham mesma função biológica, localização cromossômica ou via regulatória e buscam identificar SNPs que estão relacionados com variações no fenótipo em estudo. Neste trabalho foram implementados quatro diferentes métodos de GSEA no contexto de GWAS considerando adaptações para aplicação em espécies animais de interesse para a agricultura (fenótipos quantitativos): GSEA-SNP, teste da razão de SNPs (SRT), Florestas Aleatórias em dois passos (2way-RF) e a utilização de modelos lineares mistos (MLM). Para avaliar o poder desses métodos na detecção de conjunto de genes associados a fenótipos quantitativos, utilizou-se um conjunto de dados, constituído de genótipos simulados por reamostragem a partir de um conjunto de genótipos reais e fenótipos simulados com os SNPs pertencentes a um mesmo conjunto de genes de vias biológicas anotadas para bovinos, extraídos da base de dados KEGG. Nesse teste, o conjunto de genes alvo foi ranqueado como o décimo primeiro do total de 255 pelo método GSEA-SNP, o quarto pelo método SRT, o sexto pelo método 2way-RF e o segundo pelo método MLM. No momento, esses métodos estão sendo avaliados utilizando conjuntos de dados simulados adicionais para avaliação da consistência dos resultados obtidos.

Palavras-chaves: GWAS, SNPs, Random Forests, Mixed Models.

1 Autor, Bolsista CNPq (PIBIC): Graduação em Estatística, UNICAMP, Campinas-SP; alinetgue@gmail.com

2 Orientador: Pesquisador da Embrapa Informática Agropecuária, Campinas-SP; roberto.higa@embrapa.br



9º Congresso Interinstitucional de Iniciação Científica – CIIC 2015
10 a 12 de agosto de 2015 – Campinas, São Paulo

ABSTRACT- *In the context of genome-wide association studies (GWAS), methods for gene set enrichment analysis (GSEA) analyze SNPs sets associated with genes that share the same biological function, chromosomal location, or pathway and they aim to identify SNPs that are related to variations in the phenotype under study. In this work, four different GSEA methods have been implemented in the context of GWAS considering adaptations for use in animal species of interest for agriculture (quantitative phenotypes): GSEA-SNP, SNP ratio test (SRT), Random Forests in two stages (2way-RF) and mixed linear models (MLM). To assess the power of these methods in the detection of gene sets associated with quantitative phenotypes, we used a data set consisting of simulated genotypes by resampling from a set of real genotypes and simulated phenotypes SNPs belonging to the same gene sets on biological pathway annotated for cattle, derived from the KEGG database. The gene sets used were extracted from the KEGG database, considering only for pathways of bovine. In this test, the target gene set was ranked in the eleventh position among 255 pathways by GSEA-SNP method, the fourth by the SRT method, the sixth by 2way-RF method and the second by the MLM method. At present, these methods are being evaluated using additional simulated data sets to evaluate the consistency of obtained results.*

Key-words: GWAS, SNPs, Random Forests, Mixed Models.