

Métodos Estatísticos na Análise de Dados Genômicos

Deisy Morselli Gysi

Universidade Federal do Paraná (UFPR)

Av. Cel. Francisco Heráclito dos Santos, 210, Curitiba – PR

deisygysi@hotmail.com

Wesley Bertoli da Silva

Universidade Federal de São Carlos (UFSCar)

Rod. Washington Luís, km 235, São Carlos – SP

wesleybertolli@hotmail.com

RESUMO

Este trabalho refere-se a um estudo caso-controle sobre Artrite Reumatoide no qual se tem disponível uma grande quantidade de marcadores genéticos do tipo *Single Nucleotide Polymorphism* (SNP). A abordagem proposta consiste em investigar quais destes SNP's apresentam associação com a Artrite Reumatoide, o que possibilitaria a identificação de SNP's com potencial conexão com esta doença. Com esse intuito, consideramos um modelo adaptado de regressão logística cujas covariáveis são: a quantidade de alelos de risco presentes em DRB1, o SNP de interesse, as seis primeiras componentes principais associadas a esses SNP's e a interação entre o SNP e a quantidade de alelos de risco em DRB1, sem considerar interações entre SNP's. No cromossomo 6 foram encontrados 166 SNP's associados com a Artrite Reumatoide (valor-p $\leq 10^{-4}$). Tais SNP's estão localizados na região HLA (*Human Leukocyte Antigen*), o que era esperado, uma vez que a Artrite Reumatoide é uma doença autoimune multifatorial.

Palavras-chave: Marcadores Genéticos, Regressão Logística, Componentes Principais.

Área principal: Estatística Aplicada, Modelos de Regressão Logística.

ABSTRACT

This work is a case-control study of Rheumatoid Arthritis which has available a large amount of genetic Single Nucleotide Polymorphism (SNP) markers. The proposed approach is to investigate which of these SNP's are associated with Rheumatoid Arthritis enabling the identification of SNP's with potential connection with this disease. For this purpose, we consider an adapted logistic regression model with covariates: the amount of risk alleles presents in DRB1, the SNP under investigation, the first six associated principal components and the interaction between the SNP and DRB1, without considering interactions between SNP's. On chromosome 6 were found a number of 166 SNP's associated with Rheumatoid Arthritis (p-value $\leq 10^{-4}$). Such SNP's are located in HLA (Human Leukocyte Antigen) region, which was expected since Rheumatoid Arthritis is an autoimmune multi-factorial disease.

Keywords: Genetic Markers, Logistic Regression, Principal Components.

Main Area: Applied Statistics, Logistic Regression Models.