

Multiple Hypothesis Testing in a Genome Wide Association Study of Bovine Tuberculosis ^[1]

Burak KARACAÖREN ¹ 

^[1] Part of this study was presented as poster at XIVth Congress of Medical Biology and Genetics at Muğla/Turkey (27-30 October 2015) and published in congress abstract book

¹ Department of Animal Science, Faculty of Agriculture, Akdeniz University, TR-07059 Antalya - TURKEY

Article Code: KVFD-2016-15883 Received: 14.04.2016 Accepted: 02.09.2016 Published Online: 09.09.2016

Citation of This Article

Karacaören B: Multiple hypothesis testing in a genome wide association study of bovine tuberculosis. *Kafkas Univ Vet Fak Derg*, 23, 87-94, 2017. DOI: 10.9775/kvfd.2016.15883

Abstract

Genome-wide association studies (GWAS) have been used to detect single nucleotide polymorphisms (SNPs) related to various animal traits. The outcome of GWAS is based on quality of the both phenotypic and genotypic datasets. False positive (or negative) associations can be obtained due to multiple hypothesis testing procedures, quality control measures, or an undetected population structure. The objectives of this study were to 1) investigate different multiple hypothesis testing procedures with different quality measures and 2) to detect and correct ancestral stratification using different single SNPs models of the bovine tuberculosis GWA data set. Based on a regression model, SNPs from chromosomes 2, 7, 8 and 13 were detected at a significance level of $P < 0.001$ without correction for multiple hypothesis testing. However, after Bonferroni correction, Hochberg's method and permutation test for multiple hypothesis correction genomic signals, it became non-significant. Only a false discovery rate approach detected weak signals (at level of 0.54) from chromosomes 2, 8, and 13. We used a model that took into account the effect of linkage disequilibrium to the multiple hypothesis testing procedures by combining adjacent SNPs test statistics with windows sizes of 2, 4 and 6. We detected strong genomic signals from chromosomes 13, 8, 6 and 2 at windows size 6. The results of this study showed that multiple hypothesis testing procedures are related to false positive genomic signals. It is difficult to suggest universally acceptable multiple hypothesis testing and QC measures and their thresholds due to sources of variations between species and within populations. However, additional analytical approaches and studies are needed to evaluate the effects of linkage disequilibrium on the multiple hypothesis testing procedures and QC measures (especially for minor allele frequencies) to GWAS under various scenarios including, but not limited to, level of heritability, linkage disequilibrium, population structure, and population size.

Keywords: Genome wide association analyses, Multiple hypothesis testing, Quality control procedures

Sığır Tüberkülozu İçin Çoklu Hipotez Düzeltmesi İle Genom Tabanlı İlişki Analizi

Özet

Genom tabanlı ilişki çalışmaları (GTİÇ) kullanılarak çiftlik hayvanlarının verimleri ile ilişkili tekil nükleotid polimorfizmler (TNP) belirlenebilmektedir. GTİÇ'den elde edilecek sonuçlar hem fenotip hem de genotip veri setlerinin kalitesine bağlı olacaktır. Populasyon tabakası, çoklu hipotez düzeltim yöntemleri ve kalite kontrol süreçleri yanlış pozitif (veya negatif) ilişki sonuçlarına yol açabilir. Bu çalışmanın amaçları: bir sığır tüberküloz GTİÇ veri setine 1) değişik kalite ölçütleri ve çoklu hipotez düzeltme yöntemlerinin 2) bazı TNP regresyon yöntemleri ile atasal tabakaların belirlenmesi ve düzeltilmesinin etkilerinin incelenmesidir. Çoklu hipotez düzeltmesi olmadan TNP regresyon modeli ile 2, 7, 8 ve 13. kromozomdan önemli TNP'ler ($P < 0.001$) için belirlendi. Ama çoklu hipotez düzeltmesi Bonferroni düzeltmesi, Hochberg yöntemi ve permutasyon ile gerçekleştirildiğinde genomik sinyallerin önemsiz çıktığı gözlemlendi. Sadece yanlış keşif oranı yöntemi 0.54 seviyesinde zayıf genomik sinyalleri 2, 8 ve 13. kromozomdan belirledi. Çoklu hipotez testlerinde dengesiz bağıntıyı, 2, 4 ve 6 TNP pencere büyüklüğü için modele tanıttık. Pencere büyüklüğü 6 olunca 2, 6, 8 ve 13. kromozomlardan güçlü genomik sinyaller tespit ettik. Bu çalışmanın sonuçları çoklu hipotez test yöntemlerinin yanlış genomik sinyallerin keşfedilmesinde önemli olduğunu ortaya koydu ve doğruladı. Hem türler arası hem de populasyonlar içi varyasyon kaynakları nedeniyle evrensel kalite kontrol ölçütleri önermek oldukça zordur. Bununla birlikte kalıtım derecesi seviyeleri, dengesiz bağıntı, populasyon yapısı ve populasyon büyüklüğü dahil farklı senaryoların varlığında dengesiz bağıntının çoklu hipotez test yöntemlerine etkileri farklı kalite kontrol ölçütleri kullanılarak (özellikle farklı minör alel sıklığı seviyelerinde) GTİÇ için analitik olarak incelenmelidir.

Anahtar sözcükler: Genom tabanlı ilişki analizi, Çoklu hipotez test yöntemleri, Kalite kontrol yöntemleri



İletişim (Correspondence)



+90 242 2274400/2507



burakkaracaoren@akdeniz.edu.tr

