Novel/Alternative statistical approaches for genetic association studies

Silvia Pineda Sanjuán (Spanish National Cancer Research Centre)

June 7, 2012

Introduction: Single SNP analyses using logistic regression have traditionally been applied in genetic association studies to assess main effects. However, many complex diseases, such as bladder cancer (BC), are likely to be associated with the combined effects of multiple loci. Problems derived from genetic data are that most single SNP analyses are underpowered to detect small effects and so methods such as least absolute shrinkage and selection operator (LASSO) have emerged as promising alternatives since they can deal both with high-throughput genotyping data and with the correlations between them. We applied the classical and the LASSO approaches to conduct a comprehensive association analysis between common germline variation in genes from the P53 pathway and BC risk.

Material and Methods: We genotyped 1058 cases and 1138 controls from the EPICURO-Spanish Bladder Cancer Case-Control study in order to investigate 184 tagSNPs in 18 genes. All study participants were of white European ancestry. We carried out a classical logistic regression analysis to assess each SNP individually and a lasso-penalized logistic regression analysis to assess multiple SNPs simultaneously.

Results: We found no evidence of association after correction for multiple testing using classical analysis ($P \geq 0.4$) or Lasso-penalized regression ($P \geq 0.3$). However, the latter analysis identified weak potential associations with 4 SNPs in the SERPINB5, TP63, BAK1 and TP73 genes with a selection reproducibility of 90%, 88%, 84% and 81%, respectively.

Discussion: The analysis of these data suggests that common variation in genes in the P53 pathway is not strongly associated with bladder cancer susceptibility. Nevertheless, we propose that small genetic effects may be best identified using non-classical multivariate analytical methods which may be more suited to overcome the problems inherent to high-throughput genetic data analyses.