

Efficient two-stage genome-wide association designs based on False Positive Report Probabilities

Peter Kraft

Depts. of Epidemiology and Biostatistics, Harvard School of Public Health, USA

Despite recent advances, very-high-throughput (VHT) technologies capable of genotyping hundreds of thousands of SNPs in individual samples remain prohibitively expensive for the large studies necessary to screen substantial sections of the genome for variants with modest effects on disease risk. I present a two-stage strategy, where a portion of available samples are genotyped with VHT technology, and a small number of the most promising variants are genotyped in the remaining samples with standard techniques. The sample sizes in the first and second stages and the corresponding significance levels are chosen to limit the False Positive Report Probability (FPRP), while maximizing the number of Expected True Positives (ETPs). I show that for a fixed budget, the multi-stage strategy has greater power (a larger number of ETPs) than the single-stage strategy (where all subjects are genotyped using expensive VHT technology). Furthermore, concentrating on the FPRP leads to considerable savings relative to strategies designed to control the family-wise error (e.g. Bonferroni correction). The FPRP and number of ETPs can also accommodate researchers' prior beliefs about the number of causal loci and the magnitude of their effects. I show that the expected number of false positives does not change if the true number and effects of causal loci differs from the specified prior, thus limiting the amount of resources spent chasing "false leads." Consideration of the FPRP under different priors suggests it is better to design scans that cover a fraction of the genome well, rather than all of it poorly.