A utility for whole-genome prediction of complex medical traits

N.A.Kulemin, V.A.Naumov, S.A.Danilenko, E.V.Generozov
Research Institute of Physical Chemical Medicine FMBA Russia,
maveriksvao@gmail.com

The recent development of genotyping and sequencing technologies results in continuous increasing of the amount of human genetic data. Nevertheless is still no clear, whether personal genetic profiling based on testing at multiple loci will be useful in clinical care and public health. It is assumed that by DNA tests in which several genes are evaluated simultaneously will be claimed in personalized medicine. Today, there are a number of commercial companies offer different direct-to-consumer services that are based on their custom fixed panel-based algorithms. The predictive ability of these commercial tests varied considerably largely due to the limited numbers of analyzed SNPs, inefficiency risk prediction algorithms and other factors.

Currently, there is a shortage of bioinformatical instruments allowing estimation of all up-to-date medical risks using full genome or genotyping data. As for example, a popular free utility Promethease (project SNPedia) classified markers for risk groups and does not provide information on the composite medical risks.

In present work we develop the new analytical utility that may help to annotate virtually any set of published genetic variation including SNPs and rare mutations. Using special mathematics, we build the model allowing estimate the composite risk of the diseases or phenotypes based on given amount of variations. Our program uses manually curated database of 150 000 SNPs and mutations from GWAS catalog, OMIM, LOVD and some others open databases. In addition to the known variations our algorithm can predict potentially pathogenic mutations not previously known. As a result, our software gives structured medical conclusions, which includes genome reference information and some text abstracts related to analyzed SNPs.

We tested our pipeline on simulated genotypic dataset and have compared our results with the ones obtained using opened software packages.