HAMFinder: Combining information to predict HLA-associated mutations with a Bayesian regression model

Daniel Habermann[1], Daniel Hoffmann[1]

Abstract

Motivation
The human leucocyte antigen system (HLA) is of paramount importance to combat viral infections by presenting peptides on the cell surface via MHC I. In this way, CD8+ cytotoxic T-Lymphocytes exert a strong selection pressure towards virus variants that escape that immune recognition pathway, e.g. through point mutations that decrease binding of the respective peptide to MHC I.

Reliably identifying HLA-associated mutations is important for understanding viral evolution, but experimental methods like binding assays are prohibitively expensive for large-scale use and fail to recognize other mechanisms of immune escape like proteasomal processing.
One step in finding these mutations is through the statistical analysis of sequence data. However, existing methods are based on nullhypothesis significance testing and do not make use of all the available information and therefore have poor real-world performance.

Results
Here, we present a Bayesian regression model that is easily extensible to include information from different sources (e.g. epitope prediction software) and makes use of recent advances in Bayesian inference, e.g. by using a sparsifying prior. We show that including this kind of information improves predictive performance considerably over state-of-the-art methods.

Availability and Implementation
The source code of this software is available at `http://gogs.uni-due.de/habermann/Escape.git` under a permissive MIT license.

Supplementary information
`google.de`[Supplementary data] are available at *Bioinformatics* online.

Keywords
human leucocyte anigen system, multiple sequence alignment, escape mutations, Bayesian inference, sparsity, horseshoe, epitope prediction

[1] *Department of Biology, University of Duisburg-Essen, Essen, Germany*

Contents