HAMDetector: 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 unsatisfactory 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
Supplementary data are available at *Bioinformatics* online.

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

[1] *Bioinformatics & Computational Biophysics, Faculty of Biology, University of Duisburg-Essen, 45117 Essen, Germany*

Contents

1. Introduction

1.1 The HLA system

One way how the human immune system is able to recognize intracellular viral infections is through the human leucocyte antigen system: In cells with active protein biosynthesis, proteins are continuously synthesized and also degraded by a process called proteasomal degradation, which cleaves proteins into linear peptides of varying length. A small subset of these peptides is presented on the cell surface via receptors called MHC class I. The genomic region encoding for MHC I is known to be highly polymorphic, with more than 20000 different HLA alleles described today. The resulting gene products differ in their binding properties, which means that cells from different individuals present a highly diverse set of peptides on their surface. Cytotoxic T cells are selected during maturation to only weakly bind to peptide/MHC I complexes when the peptide originated from proteins of the usual proteome, but might be able to strongly bind to complexes of MHC I with peptides which are generated from of a viral protein. Upon activation, T cells induce cytolytic activity and recruit other immune cells.

1.2 HLA escape

In this way, the HLA system exerts a strong selection pressure towards virus variants that escape T cell recognition, for example through a point mutation that results in reduced binding of an immunogenic peptide to MHC I or through a set of mutations that alters the viral protein in such a way that it is cleaved into different peptides that are not recognized by the host's T cell repertoire.

The evolutionary events are complex and occur not only on the level of individuals, where a virus adapts to specific features of the host, but also on the population level, because HLA alleles differ in their frequency across geographic regions, as they are inherited according to standard Mendelian rules.

Upon transmission, HLA escape mutations typically quickly revert to their wild type as they usually reduce viral replicative capacity (if a mutation would increase viral replicative capacity regardless of the presence of a given HLA allele, it would probably already be the wild type). Kawashima et al. describe an escape mutation that is selected by HLA allele B53, does not strongly affect viral replicative capacity and therefore slowly enriches over time in Japan, where B53 commonly occurs.

How quickly a given escape mutation is selected upon transmission in a host depends on the magnitude of the reduction in viral replicative capacity, on the strength of selection pressure and also on the genetic background, e.g. some escape mutations require compensatory mutations which partly attenuate the negative impact on viral replicative capacity.

Studying HLA escape therefore provides an unique opportunity to gain insight into viral evolution, on the host level but also on the population level. Unfortunately, identifying HLA escape mutations is difficult in practice.

1.3 Identifying HLA-escape mutations

There are several experimental methods available to study HLA escape: Recombinant MHC-I molecules can be used in binding assays: Upon complex formation with a peptide, a change in conformation can be detected with conformation-specific antibodies. This method is relatively fast, but only measures binding affinity of a peptide to MHC-I and does not account for antigen processing or immunodominance, which describes the observation that a peptide may be presented via MHC-I on the cell surface but does not induce an immune response. An experimental setup that resembles the conditions in-vitro more closely but is also more time-consuming is to measure CD8+ T cell responses instead. This is usually done by stimulating PBMCs with prototype and variant peptides and measuring the secretion of IFN-gamma by intracellular cytokine staining and fluorescence-assisted cell sorting. To analyse CD8+ T cell responses against endogenously processed antigens it is necessary to generate cell-lines stably expressing the antigen in question and adding antigen-specific CD8+ T cells. This method scales poorly as it requires transfection of cell lines and antigen-specific expansion of CD8+ T cells.