IgTools: a toolkit for analysis of antibody repertoire from immunosequencing and mass spectra data

Yana Safonova 1, Alexander Shlemov 1, Ekaterina Starostina 1, Andrey Bzikadze 1, Alla Lapidus 1, Pavel A. Pevzner 1,2

1 Center for Algorithmic Biotechnology, Institute of Translational Biomedicine, St. Petersburg State University, St. Petersburg, Russia
2 Dept. of Computer Science and Engineering, University of California at San Diego, USA

Abstract

Reconstruction and analysis of antibody repertoire is an important part of various immunological studies, e.g., design of antibody drugs. Here we present IgTools, a first multipurpose toolkit for construction and investigation of antibody repertoire from immunosequencing and mass spectra data:

▶ IgRepertoireConstructor, an algorithm for antibody repertoire construction;
▶ AntEvolo, an algorithm for construction of clonal trees and evolutionary analysis of antibody repertoires;
▶ IgAnalyzer, a tool for statistical analysis of antibody repertoire;
▶ IgSimulator, a versatile simulator of antibody repertoire.

IgRepertoireConstructor and IgSimulator are freely available at Github. IgAnalyzer and AntEvolo will be released at the end of 2015.

Antibody sequencing

Modern sequencing technologies (e.g., Illumina MiSeq) allow biologists to perform full-length scanning of antibody repertoire. For example, overlapping paired-end Illumina MiSeq reads cover variable regions of antibody.

IgRepertoireConstructor

IgRepertoireConstructor takes as an input immunosequencing Illumina reads and constructs full-length repertoire.

▶ VJ Finder: V, J labeling, cleaning and cropping input reads
▶ Construction & clusterization of SW graph: representation of cleaned reads as a Smith-Waterman graph and finding of groups of highly similar antibodies
▶ Antibody construction: reconstruction of antibody sequences from the constructed groups, cleaning amplification errors

Quality assessment

Repertoire constructed by IgRepertoireConstructor can be validated using mass spectra and converted into a database for identification of mass spectra using some standard tools, e.g., MS-GF+.

MassSpectraAnalyzer takes as an input mzIdentML files and computes similarity between constructed repertoire and mass spectra.

IgSimulator is a tool for simulation of antibody repertoire and Ig-seq library. IgSimulator is designed for testing and benchmarking tools for reconstruction of Ig repertoires.

Tool generates:
▶ FASTA file with reference Ig repertoire
▶ FASTQ reads covering variable regions of simulated antibodies using ART read simulator
▶ Ideal repertoire corresponding to Ig-seq library

AntEvolo

Some B-cells are grouped into “families” due to clonal expansion and somatic hypermutations. These families can represented as clonal trees:

Construction of clonal trees from antibody repertoire allows biologists to understand evolutionary processes of B-cells and trace history of SHMs.

AntEvolo algorithm searches for B-cell families in antibody repertoire and constructs clonal tree for each family. Also AntEvolo performs analysis of SHMs: detection of mutations shared between branches of clonal tree, computes SHM characteristics.

IgAnalyzer

IgAnalyzer is a tool for statistical analysis of antibody repertoire. It computes:

▶ Statistics showing diversity of recombination events (distribution of V(D)J recombinations, palindromic insertions, endonuclease removals, CDR3 counting etc)
▶ Similarity of recombination events in several antibody repertoires

Availability

▶ User support: igtools_support@googlegroups.com

Acknowledgments

We are grateful to Anton Bankevich for the help in preparation of this poster. We thank Roche and Genentech companies and Dr. Chudakov’s laboratory for providing us with immunosequencing data. This work was supported by the Russian Science Foundation (grant 14-50-00069).