Hla-mapper: an application to optimize the mapping of hla sequences produced by massively parallel sequencing procedures

Erick C. Castelli, Michelle A. Paz, Andréia S. Souza, Jaqueline Ramalho, Celso Teixeira Mendes-Junior

PII: S0198-8859(18)30235-0
DOI: https://doi.org/10.1016/j.humimm.2018.06.010
Reference: HIM 10087

To appear in: *Human Immunology*

Received Date: 23 November 2017
Revised Date: 15 June 2018
Accepted Date: 29 June 2018

Please cite this article as: Castelli, E.C., Paz, M.A., Souza, A.S., Ramalho, J., Mendes-Junior, C.T., Hla-mapper: an application to optimize the mapping of hla sequences produced by massively parallel sequencing procedures, *Human Immunology* (2018), doi: https://doi.org/10.1016/j.humimm.2018.06.010

HLA-MAPPER: AN APPLICATION TO OPTIMIZE THE MAPPING OF HLA SEQUENCES PRODUCED BY MASSIVELY PARALLEL SEQUENCING PROCEDURES

Erick C. Castelli [1,2], Michelle A. Paz [1], Andréia S. Souza [1], Jaqueline Ramalho [1], Celso Teixeira Mendes-Junior [3]

[1] São Paulo State University (UNESP), Molecular Genetics and Bioinformatics Laboratory, Experimental Research Unit (UNIPEX), School of Medicine, Botucatu, State of São Paulo, Brazil

[2] São Paulo State University (UNESP), Pathology Department, School of Medicine, Botucatu, State of São Paulo, Brazil.

[3] Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 14040-901, Ribeirão Preto - SP, Brazil.

Contact: Erick C. Castelli (erick.castelli@unesp.br)

Keywords: MHC, HLA, Next Generation Sequencing (NGS), Second Generation Sequencing, variability, polymorphisms; typing; aligners; mapping tool

ABSTRACT

A challenging task when more than one HLA gene is evaluated together by second-generation sequencing is to achieve a reliable read mapping. The polymorphic and repetitive nature of HLA genes might bias the read mapping process, usually underestimating variability at very polymorphic segments, or overestimating variability at some segments. To overcome this issue we developed *hla-mapper*, which takes into account HLA sequences derived from the IPD-IMGT/HLA database and unpublished HLA sequences to apply a scoring system. This comprehends the evaluation of each read pair, addressing them to the most likely HLA gene they were derived from. *hla-mapper* provides a reliable map of HLA sequences, allowing accurate downstream analysis such as variant calling, haplotype inference, and allele typing. Moreover, *hla-mapper* supports whole genome, exome, and targeted sequencing data. To assess the software performance in comparison with traditional mapping algorithms, we used three different simulated datasets to compare the results obtained with *hla-mapper*, BWA MEM, and Bowtie2. Overall, *hla-mapper* presented a superior performance, mainly for the classical HLA class I genes, minimizing wrong mapping and cross-mapping that are typically observed when using BWA MEM or Bowtie2 with a single reference genome.