BMC Bioinformatics

ORAL PRESENTATION   **Open Access**

A computational pipeline for diagnostic biomarker discovery in the human pathogen Trypanosoma cruzi

Santiago J Carmona[1*], Paula Sartor[2], Maria Susana Leguizamón[2], Oscar Campetella[1], Fernán Agüero[1]
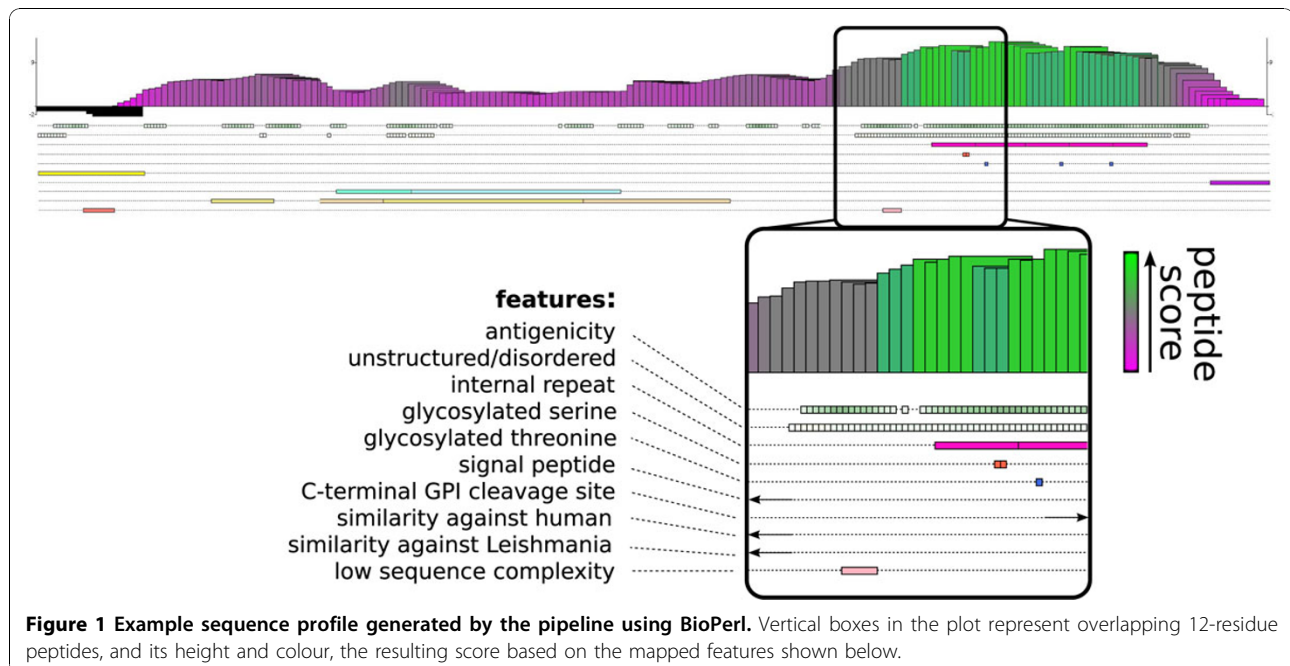
Background

The protozoan parasite *Trypanosoma cruzi* is the causative agent of Chagas' disease, endemic in 18 countries in Central and South America. Transmission also occurs in non-endemic countries by way of blood transfusion and organ transplantation. Diagnosis of American trypanosomiasis is based on the detection of antibodies directed against *T. cruzi* antigens. In this work we mined the *T. cruzi* genome sequence to identify new peptidic diagnostic biomarkers.

Methods

An integrative bioinformatic strategy was adopted to prioritize peptidic antigens with low cross-reactivity in the genome of *T. cruzi*. A computational pipeline was developed to assess a set of molecular properties on

Figure 1 Example sequence profile generated by the pipeline using BioPerl. Vertical boxes in the plot represent overlapping 12-residue peptides, and its height and colour, the resulting score based on the mapped features shown below.

* Correspondence: sjcarmona@gmail.com
Full list of author information is available at the end of the article

each protein from the reference *T. cruzi* genome, such as subcellular localization or expression level (by mass spec. evidence, number of gene copies and synonymous codon usage bias). At a higher resolution, a set of local properties were evaluated, such as repetitive motifs, disorder (structured vs natively unstructured regions), trans-membrane spans, glycosylation sites, polymorphisms (conserved vs. divergent regions), predicted B-cell epitopes, sequence similarity against human proteins and *Leishmania* (potential cross-reacting species) (Figure 1). A scoring function based on these properties was used to rank each of the ~10 million 12-residue overlapping peptides in which the ~ 22,000 *T. cruzi* proteins can be virtually fragmented. Experimental validation of predicted epitopes was performed with peptide microarrays, screened using pooled sera from human chagasic patients and controls.

Results

We show that our integrative method outperforms alternative antigen prioritizations based on individual properties (such as B-cell epitope predictors alone). Our genome-wide prioritization uncovered more than 300 promising biomarker candidates. 200 high-scoring peptides corresponding mostly to hypothetical proteins were selected for immunological validation, along with 40 peptides derived from previously validated B-cell epitopes and an additional set of 40 low-scoring peptides as controls. Preliminary results based on microarray images revealed that ~25% (49/200) of the candidate peptides reacted specifically against the positive sera pools assayed.

Conclusion

The developed bioinformatic approach proved to be successful, leading from a genome-wide prioritization to the identification of novel peptidic antigens with diagnostic potential. Moreover, the algorithm may be used to prioritize biomarkers in other pathogen species.

Author details
[1]Instituto de Investigaciones Biotecnológicas, Universidad de San Martín, San Martín, Argentina. [2]Departamento de Microbiología, Facultad de Medicina, Universidad de Buenos Aires, Buenos Aires, Argentina.

Published: 7 December 2010