

A new bioinformatics pipeline to identify schistosomiasis vaccine candidates from a phage-display assay

João Vicente de Morais Malvezzi, Sergio Verjovski-Almeida

Instituto Butantan

Abstract

Some species of *Schistosoma* spp. are haemoparasites that cause schistosomiasis, affecting more than two hundred million people worldwide. In Brazil, the etiological agent is *S. mansoni*. Since the publication of its transcriptome in 2003, little success has been achieved in tests of vaccine candidates, and new methods for recognizing new targets are required. In the present study, we have applied a phage immunoprecipitation followed by sequencing (PhIP-Seq) and employs self-healed rhesus macaques antibodies against *S. mansoni* to capture antigenic peptides from a phage-display library constructed by us with synthetic oligonucleotides that encode all fragments of all *S. mansoni* proteins having known sequences, with the aim of identifying new epitopes for a vaccine against schistosomiasis. We developed a bioinformatics pipeline to analyze the data that will be obtained by PhIP-Seq. For each FASTQ the pipeline applies steps of quality control, preprocessing, read mapping against the oligonucleotides library, and count the reads mapped to each oligonucleotide sequence to generate a table of read counts for each sample. The count tables are processed with a script in R which will identify enriched peptides for each sample. To find enriched peptides in the immunoprecipitated phages compared with peptides abundance in the total (non-precipitated), three scripts were created using different statistical approaches: 1) Zero-Inflated Generalized Poisson (ZIGP) distribution, 2) Negative Binomial (NB) distribution, and 3) Z-score. To test these scripts, we used a real PhIP-Seq dataset from a phage-display library of human-virus protein fragments containing 96, 179 peptides screened with 10 human serum samples with two replicates each. The ZIGP script identified 1, 757 peptides as enriched, whereas NB script and Z-score script identified 1, 243 and 2, 369, respectively. ZIGP and NB scripts identified 1, 126 peptides in common, while only 30 peptides were enriched by the three methods. These results show that the NB is more stringent, returning fewer peptides in comparison to the other methods. Noteworthy is that a more robust group of enriched peptides is acquired when using the intersected result between NB and ZIGP when compared with those enriched peptides returned by Z-score method. Given that, this pipeline will be applied to search for antigens of *S. mansoni* recognized by the antibodies from twelve self-healed rhesus macaques sera. We expect to find groups of macaques that developed antibodies for different antigens or even the same antibodies at different times after infection and healing. Probably those proteins targeted by the macaques' antibodies are critical for the survival of the parasite. Thus, we expect to identify several new parasite antigens candidates that can be tested in future studies to generate a vaccine against schistosomiasis.

Funding: FAPESP