AO 1680

Exomim: a web tool to improve prioritization of genetic variants

Delva Pereira Leão; Pedro B. Marin; Junior Figueiredo Barros; Julio Cesar Santos dos Anjos; Claudio R. Geyer; Ursula da Silveira Matte - UFRGS

The Whole-Exome Sequencing (WES) is one of the main applications of next-generation sequencing (NGS) increasingly used into medical practice. A turning point into WES analysis is the prioritization of genetic variants. Such complexity has stimulated the development of methodologies to assist this process. The majority of available tools do not comprise phenotype information, required to perform an integrated analysis. We propose the development of Exomim tool, a web tool that comprises four distinc services to analyze .vcf files: (1)file storage; (2)annotation of variants; (3)variant prioritization and (4)gene listing. We implemented the annotation service based on Variant Effect Preditor. To implement the gene listing service we used the Online Mendelian Inheretance in Man (OMIM) API to create a local database to access information of each OMIM record, indexing was obtained through Medical Subject Headings (MeSH) terms. Combining OMIM and MeSH we developed an efficient search engine that outputs a list of genes associated with input phenotype. We obtained an initial validation of this service using a simple approach. Phenotypes of five diseases of interest were selected and searched against OMIM website and gene listing service; our goal was to obtain the target gene to each genetic condition. We performed this search using And/Or grammatical conjuctions and in all cases the given gene was found. We also used an analog approach to perform a similarity analysis measured by Jaccard's coefficient J(A,B) = 0.798 . Together these results indicate that our strategy is in agreement with OMIM information and that our service accomplishes the initial objective. The prioritization service concatenate the annotation information required by the user through selectable fields, with phenotype information of gene listing service to filter from a .vcf file a final output of variants that meets user's criteria. To validate this module we analyzed 35 .vcf files obtained from different gene panels of NGS. Taking together a proper filter set and phenotype information, the service was capable of define the causal variant to each case this dataset. We also conducted a performance analysis of this service; prioritization of 250000 genetic variants was carried out in less than 160 seconds and consumed approximately 900 MB of memory. At this point we are working to improve our models and we hope that once fulfilled, this work will contribute to future analysis of case-only WES.
Key Words: Bioinformatica; Next-generation Sequencing; Genetic diseases