

DESHCV: Hepatitis C virus customized web-based software for biomedical text mining

Samuel K Kwofie¹, Aleksandar Radovanovic², Vijayaraghava S Sundararajan¹, Monique Maqungo¹, Alan Christoffels¹ and Vladimir B Bajic^{2*}

Address: ¹South African National Bioinformatics Institute, University of the Western Cape, Private Bag- X17, Modderdam Road, Bellville, Cape Town, South Africa, Samuel@sanbi.ac.za.

²Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia.

The skyrocketing chronicity and global infection rate of Hepatitis C Virus (HCV) necessitate the need to unlock the molecular etiology underlying the pathophysiology of HCV related diseases such as liver cancer. The plethora of essential molecular data in the corpus of published biomedical literature could be leveraged to augment efforts towards discovery of novel anti-viral drugs, cellular receptors and appropriate predictive biomarkers.

We report here, a biomedical text mining tool solely focused on Hepatitis C Virus (HCV) developed by mining text related to HCV. In this report, Dragon Exploratory System on Hepatitis C Virus (DESHCV): a biomedical text mining and relationship exploring knowledgebase has been developed based on the Dragon Exploratory Systems (DES). Biomedical concepts pertaining to HCV proteins and their name variants have been organized into catalogues of ontologies and incorporated into the pre-compiled lexicons of biological concepts existing in the DES. These concepts are cross-referenced to external databases such as Gene Ontologies (GO), UNIPROT and Entrez Gene. A list of 32,895 abstracts was retrieved via PubMed database using specific keywords queries relating to HCV. These abstracts were computationally processed based on concepts recognition of terms in the following ontologies: human genes and proteins, metabolites and enzymes, pathways, chemicals with pharmacological effects, and disease concepts.

The web query interface of DESHCV enable users to retrieve HCV related information using specified concepts, keywords and phrases. Each concept search generate text analyzed association networks and hypothesis which could be tested to identify potentially novel relationship between various concepts. The obtained information could serve as enriched baseline data for further exploration towards gaining insights into plausible therapeutic and diagnostic discoveries. DESHCV is therefore an enhanced and integrated online literature-based discovery resource freely available via <http://apps.sanbi.ac.za/DESHCV/>.