DOI 10.18699/SBB-2022-190

Identification of virus-host interaction mechanisms based on text-mining approach

Biziukova N.*, Ivanov S., Filimonov D., Tarasova O., Poroikov V.
Institute of Biomedical Chemistry, Moscow, Russia
** nad.smol@gmail.com*

Key words: viral diseases, virus-host interactions, molecular mechanisms, text-mining

Motivation and Aim: Infections cause significant proportion of deaths worldwide [1]. The COVID-19 pandemic has clearly shown the urgent need to turn efforts towards the development of effective antiviral drugs able to affect human proteins that interact with viruses directly or indirectly. Due to the high mutability of viruses, the development of antiviral drugs is a great challenge. Therefore, it is necessary to have a deep understanding of the molecular mechanisms of interaction between the virus and the host. In our study, we focused on development of a workflow that is intended to retrieve the latest information on experimentally proven interactions between the most common viruses and human leading to immune system reaction and identification of key targets involved in the mechanisms of viral infections affecting immune system, including SARS-CoV-2, herpesviruses, hepatitis B, C, influenza, and HIV [2–4].
Methods and Algorithms: The texts of scientific publications were automatically extracted from NCBI PubMed using Python 3.10 scripts and Bio library Entrez module. Protein names recognition was performed using our web-applications previously developed [5, 6]. Fig. 1 represents the interface of the tool.

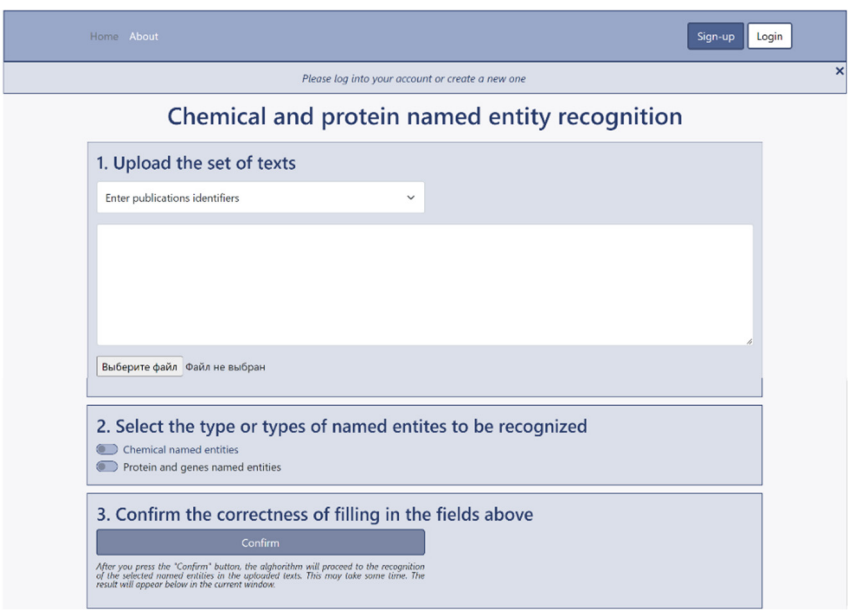

Fig. 1. Interface of web-application SigNER

Named entity recognition is based on combined utilization of conditional random fields (CRF) algorithm and Naïve Bayes classifier. We used a set of rules to implement relation extraction. We manually analyzed a set of texts in order to find patterns between related entities. To build patterns, we considered only phrases that are commonly used for indicating association between two or more terms. As a result, we identified a set of proteins involved in virus-host interactions and immune suppression mechanisms shared by the studied viruses.

Detected proteins/genes named were then matched with differentially expressed genes. We presume that changes in the expression of these genes may underlie different responses to viral infections [7]. Therefore, we collected freely available transcriptomic data and analyzed them. Analysis was performed using R and BioConductor package.

Results: We constructed requests to PubMed in order to compile text corpus dedicated to mechanisms of action between human immune system and the following viruses: SARS-CoV-2, herpesviruses, hepatitis B, C, influenza, and HIV. We performed named entity recognition using the developed workflow. Then, we created a set of rules that allows to indicate relationships between proteins/genes named entities within the text. For example, APOBEC3 family members and SAMHD1 were identified as restriction factors for DNA- and RNA-viruses [8, 9]. These relationships were used to identify mechanisms of host response to viruses and immune suppression, caused by the viruses. In order to test the performance accuracy of our method, identification of differentially expressed genes was also carried out by transcriptomic data analysis.

Conclusion: Presented approach allowed us to collect data on genes involved in immune system reactions on different viral infections. In addition to evidence based on literature sources, the workflow also provides experimental data. Our approach can be used for design of novel drugs inhibiting virus-mediated immune suppression, and it can open the ways for drug repurposing.

Acknowledgements: This study is supported in the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021-2030), project No. 122030100170-6.

References

1. Piret J., Boivin G. Pandemics throughout history. *Front Microbiol*. 2021;11:631736.
2. Fragkou P.C. et al. Host immune responses and possible therapeutic targets for viral respiratory tract infections in susceptible populations: a narrative review. *Clin Microbiol Infect*. 2022;S1198743X22001495.
3. Piconese S. et al. Viral hepatitis, inflammation, and cancer: A lesson for autoimmunity. *J Autoimmun*. 2018;95:58-68.
4. Danastas K. et al. Herpes Simplex Virus Type 1 Interactions with the Interferon System. *Int J Mol Sci*. 2020;21(14):5150.
5. SigNER URL [http://www.way2drug.com/signer/]. Accessed on 24.04.2022.
6. Naïve-Bayes CNER URL [http://www.way2drug.com/cner/]. Accessed on 24.04.2022.
7. Ivanov S. et al. A computational analysis of transcriptional profiles from CD8(+) T lymphocytes reveals potential mechanisms of HIV/AIDS control and progression. *Comput Struct Biotechnol J*. 2021;19:2447-2459.
8. Milewska A. et al. APOBEC3-mediated restriction of RNA virus replication. *Sci Rep*. 2018;8:5960.
9. Sze A. et al. SAMHD1 host restriction factor: a link with innate immune sensing of retrovirus infection. *J Mol Biol*. 2013;425:4981-4994.