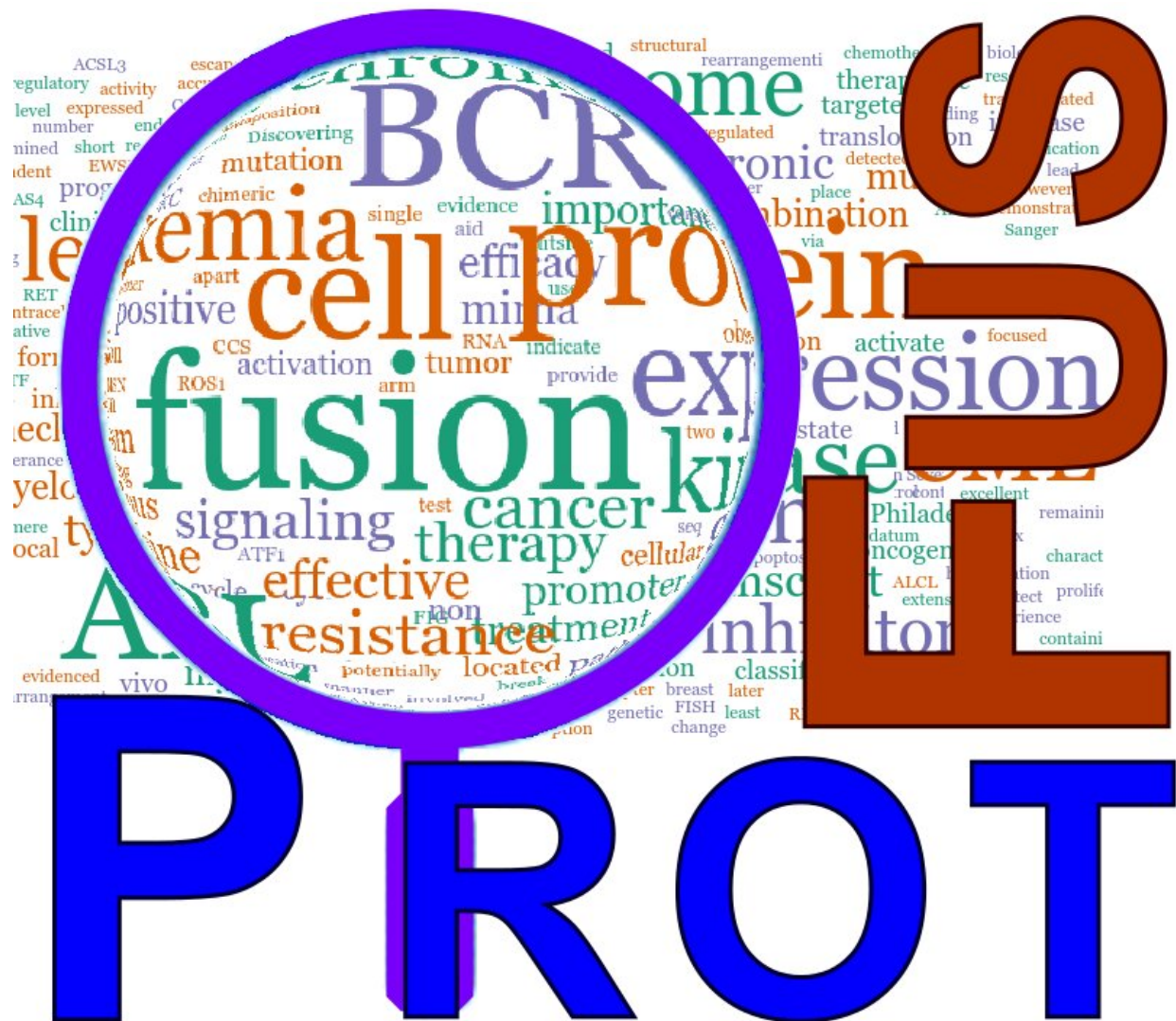# New tool mines scientific texts for fusion protein facts

August 22 2019

ProtFus finds protein interactions of fusions using text mining. Credit: Tagore et al, pcbi.1007239.

A new computational tool called ProtFus screens scientific literature to validate predictions about the activity of fusion proteins—proteins encoded by the joining of two genes that previously encoded two separate proteins. Somnath Tagore in the Frenkel-Morgenstern Lab at Bar-Ilan University, Israel, and colleagues present ProtFus in *PLOS Computational Biology*.

Different kinds of fusion proteins can arise naturally in the human body, sometimes leading to cancer. Understanding interactions between fusion proteins and other proteins can help improve personalized cancer treatment. However, the number of scientific papers discussing these interactions is growing rapidly, and there is no standard format for presenting this information. Thus, organizing and keeping abreast of this knowledge poses a major challenge.

ProtFus addresses this challenge by using computational strategies—such as text mining and machine learning—to analyze scientific literature from the online search engine PubMed. It is able to identify fusion proteins that may go by multiple names, and it can identify experimentally verified interactions between fusion proteins and other proteins. When applied to a test set of 1,817 fusion proteins, ProtFus identified 2,908 interactions across 18 cancer types that had been published in scientific texts in PubMed.

ProtFus also builds on a tool previously developed by the researchers in order to predict a given fusion protein's interactions based on the known properties of its two parental proteins. ProtFus takes a fusion protein of interest, uses the previously developed tool (Chimeric protein-Protein-Interactions, or ChiPPI) to predict its interactions, and then validates these interactions by means of a PubMed search.

"Our findings demonstrate the potential for text mining of large-scale scientific articles using a novel big-data infrastructure, with real-time updating from articles published daily," says Dr. Milana Frenkel-Morgenstern, corresponding author of the study. "ProtFus can promote studying alterations of protein networks for individual cancer patients in a fully personalized manner," highlights Tagore, the first author and previous postdoc in the lab (currently a postdoc at Columbia University, New York).

  More information: Tagore S, Gorohovski A, Jensen LJ, Frenkel-Morgenstern M (2019) ProtFus: A Comprehensive Method Characterizing Protein-Protein Interactions of Fusion Proteins. *PLoS Comput Biol* 15(9): e1007239. doi.org/10.1371/journal.pcbi.1007239

Provided by Public Library of Science