POLITECNICO DI TORINO
Repository ISTITUZIONALE

Integrated systems genomic approaches for characterizing uncharacterized proteins

Original
Integrated systems genomic approaches for characterizing uncharacterized proteins / Valadi, J., Seshadri Sundararajan, V., Reddy Bandapalli, O., Benso, A., Suravajhala, P.. - In: FRONTIERS IN GENETICS. - ISSN 1664-8021. - ELETTRONICO. - (2022). [10.3389/fgene.2022.1000825]

Availability:
This version is available at: 11583/2971250 since: 2022-09-13T06:54:19Z

Publisher:
Frontiers Media

Published
DOI:10.3389/fgene.2022.1000825

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

05 November 2026

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 13 September 2022
DOI 10.3389/fgene.2022.1000825

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Richard D. Emes,
University of Nottingham,
United Kingdom

*CORRESPONDENCE
Jayaraman Valadi,
valadi@gmail.com
Prashanth Suravajhala,
prash@am.amrita.edu

SPECIALTY SECTION
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

RECEIVED 22 July 2022
ACCEPTED 22 August 2022
PUBLISHED 13 September 2022

CITATION
Valadi J, Sundararajan VS, Bandapalli OR, Benso A and Suravajhala P (2022), Editorial: Integrated systems genomic approaches for characterizing uncharacterized proteins. *Front. Genet.* 13:1000825. doi: 10.3389/fgene.2022.1000825

COPYRIGHT
© 2022 Valadi, Sundararajan, Bandapalli, Benso and Suravajhala. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Integrated systems genomic approaches for characterizing uncharacterized proteins

Jayaraman Valadi[1,2,3]*, Vijayaraghava Seshadri Sundararajan[2,4], Obul Reddy Bandapalli[2], Alfredo Benso[4] and Prashanth Suravajhala[2,5]*

[1]Department of Computer Science, Vidyashilp University, Maharastra, India, [2]Bioclues.org, Hyderabad, India, [3]Department of Computer Science, FLAME University, Pune, India, [4]PolitoBIOMed Lab—Biomedical Engineering Lab, Turin, Italy, [5]Amrita School of Biotechnology, Amrita Vishwa Vidyapeetham University, Kollam, Kerala, India

KEYWORDS

known unknowns, hypothetical proteins, systems genomics, bioinformatics, annotation

Editorial on the Research Topic
Integrated Systems Genomic Approaches for Characterizing Uncharacterized Proteins

The machinery of life is made possible by proteins. Each protein is characterized by its unique amino acid sequence which determines its shape as well as its function. Over the years, researchers have shown how the shape and functions of relatively small numbers of proteins vary (Adams, 2008) from which functions of these proteins could be ascertained. On the other hand, there are known unknown regions in the genome which are the uncharacterized sequences. The genes which are known but whose functions are not known are called orphan genes and their corresponding protein products are called hypothetical proteins (HPs; Galperin, 2011; Suravajhala and Sundararajan, 2012). In this special issue, authors show how characterizing the uncharacterized sequences can advance the state of the art in different fields, from studying drug resistance in Malaria, to understanding biochemical mechanisms that allow them to adapt to cold-habitats besides investigating bacterial antibiotic resistance using the HPs. The topic entails four articles integrating *bona fide* computational tools for discerning the functions of the known unknown regions in the genome.

Singh and Gupta describe annotation of *Plasmodium falciparum* HPs with an aim to understand the growing drug resistance in Malaria, which, over the years has created an urgent requirement for alternative and more effective antimalarial drugs or vaccines (Singh and Gupta). The availability of complete genomes of various Plasmodium species has opened new opportunities for development of drugs and vaccines, albeit with mixed success. This, the authors, emphasize results in a poor understanding of the parasite

biology and its incomplete functional annotation. While the Plasmodium genomes are yet to be completely annotated with exact mechanisms of pathogenesis in human hosts, a host of proteins are currently not annotated. For example, of 5,389 proteins in Plasmodium falciparum 3D7 strain 1,626 proteins (~30% data) are labeled as "hypothetical". Annotation of these HPs may lead to better understanding of the parasite mechanisms to evade and exploit host anti parasitic mechanisms. To bring the work closer, they employ a computational pipeline through which they identify 266 conserved functional signatures with essential candidates attributing to various biochemical, signaling and metabolic pathways. Furthermore, they could successfully model 11 proteins and perform molecular dynamic simulations to check for stable conformations. Thus, they could deploy a protein-protein interaction (PPI) network revealing 3,299 nodes and 2,750,692 edges. The work attempts to bridge the gap for experimenting candidate HPs (Singh and Gupta).

Ijaq et al. discuss the importance of studying the molecular mechanisms emphasizing the adaptation of a cold-habitat Gram-negative psychrophilic Antarctic bacterium *Pseudomonas spp. Lz4W* that can survive in extreme temperature conditions. The authors used explorative proteome analyses to identify the proteins and pathways that might be aiding for the adaptation and survival of the bacteria. By employing a bioinformatics pipeline which they describe, they studied that Pseudomonas sp. Lz4W genome (CP017432, version 1) which were found to contain 4,493 genes, 4,412 coding sequences (CDS), of which 743 CDS were annotated as HPs. Among them, 61 HPs were found to be expressed consistently at the protein level with 18 functionally assigned which were found to be involved in peptidoglycan metabolism, cell wall organization, ATP hydrolysis, outer membrane fluidity, catalysis aiding in the biological mechanisms of cold adaptations. Their study aimed to provide a deeper understanding of the functional significance of HPs in cold adaptation of Pseudomonas sp. Lz4W. The authors further submitted the data generated from MS proteomics to the ProteomeXchange through PRIDE (PXD029741; Ijaq et al.).

Liu et al. studied the role of HPs in *Mycobacterium tuberculosis* and BCG strains with a focus on the mortality rate associated with Tuberculosis. The authors perform a comparative proteomic profiling of intracellular and extracellular virulent strain M. tb and bacille Calmette–Guérin (BCG) from infected THP-1 cells through incubation and post incubation studies. They identify a total of 1,557 proteins, which were divided into four groups for comparison of M. tb versus BCG under 7H9 culture, M. tb one-type versus M. tb second-type, BCG one-type versus BCG second-type, and other combinations. Among them, 8 proteins were found to be associated with ESAT6, related to virulence besides 5 uncharacterized proteins that were differentially expressed. Further, pathway enrichment analysis revealed that glyoxylate and dicarboxylate metabolism pathways were majorly involved in the biological process and these findings, the authors claim to have valuable data for further exploration of molecular mechanisms for M. tb virulence and BCG immune response (Liu et al.).

Bhat et al., reviewed in silico approaches in the form of bioinformatics. They discussed the applications of bioinformatics in collecting the experimental data, organizing them in several databases available online like nucleotide database, protein databases, GenBank and others and employing them as reference for experimental evaluation and validation in the form of genomic, transcriptomic, proteomics, epigenomics and metabolomics data. The application of characterizing the uncharacterized proteins include the discovery of new metabolic pathways which can be targeted for therapeutic purposes, development of vaccines, and novel potential drug targets for drug discovery. They further describe mutational databases and regulatory genome databases, including Human Splicing Finder (HSF), Exonic Splicing Enhancer Mutation taster, aiding in better understanding of the intronic mutations that can be further validated experimentally (Bhat et al.).

Abbasi et al. provided an *in silico* approach to characterize the HPs of six different *Clostridium difficile* strains. A multi-strain, spore-forming, Gram-positive, opportunistic enteropathogen bacteria, *Clostridium difficile* (*C. difficile*) is primarily linked to nosocomial infections, which cause severe diarrhea and colon inflammation. The report suggests that the excessive use of antibiotics has led to the development of antibiotic resistance in *C. difficile* strains. The authors have emphasized the need for functional annotation of the HPs that were found in the genome of *Clostridium difficile* using various bioinformatics tools in which they analyzed complete genomes, including contigs identification, coding sequences, phages, CRISPR-Cas9 systems, antimicrobial resistance determination, membrane helices, instability index, secretory nature, conserved domain, and determined its physicochemical properties along with vaccine target properties like comparative homology analysis, allergenicity, antigenicity determination along with structure prediction and binding-site analysis to annotate its possible functions in vaccine development and potential drug targets. They identified ca. 32–38% and 23–27% of HPs were localized in the cytoplasm and cytoplasmic membrane, respectively. While 99.7% of the identified HPs across all the selected strains were non-homologous to humans, and show no virulence factors, they further exploited the antigenicity of the HPs and successfully model highly antigenic proteins using structure prediction and their binding sites. They underpin the most recent developments in the adage of next generation sequencing (NGS) with vivid curation tools which help understand and discover novel drugs for ascertaining the known unknown regions in bacteria (Abbasi et al.).

In the realm of biology, the known unknown regions (Logan, 2009) could be better ascertained using hypothome (~ an interactome of hypothetical proteins, which we proposed earlier; Desler et al., 2014) with multidisciplinary subjects such as systems genomics, bioinformatics aided functional genomics and third generation sequencing technologies with a focal interest on identifying candidate regions associated with causality. There is a room for exploring more tools and fundamental processes elucidating the mechanisms underpinning small molecular interactions and deploying methods for ascertaining the function of HPs. In conclusion, multi-integration is the key challenge.

Author contributions

JV and PS wrote the first draft. OB and AB reviewed sections of Abbasi et al., with VS and AB contributing to other reviews. All authors approved the manuscript before being proofread by PS.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adams, J. (2008). The proteome: Discovering the structure and function of proteins. *Nat. Educ.* 1 (3), 6.

Desler, C., Zambach, S., Suravajhala, P., and Rasmussen, L. J. (2014). Introducing the hypothome: A way to integrate predicted proteins in interactomes. *Int. J. Bioinform. Res. Appl.* 10 (6), 647–652. doi:10.1504/IJBRA.2014.065247

Galperin, M. Y. (2001). Conserved 'hypothetical' proteins: New hints and new puzzles. *Comp. Funct. Genomics* 2 (1), 14–18.

Logan, D. C. (2009). Known knowns, known unknowns, unknown unknowns and the propagation of scientific enquiry. *J. Exp. Bot.* 60 (3), 712–714. doi:10.1093/jxb/erp043

Suravajhala, P., and Sundararajan, V. S. (2012). A classification scoring schema to validate protein interactors. *Bioinformation* 8 (1), 34–39. doi:10.6026/97320630008034