

## EDITORIAL

# Bioinformatics helping to mitigate the impact of COVID-19 – Editorial

The coronavirus disease 2019 (COVID-19) pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) became known to the world at the end of 2019 [1]. The severity of the pandemic and its worldwide spread provoked an unprecedented effort of the scientific community and a lot of new research was conducted, especially by the medicine, biology, public health, bioinformatics and computer science researchers, that led to the rapid development of several novel vaccines [2].

At the biological level, SARS-CoV-2 and COVID-19 research involves several themes, including high-throughput technologies such as Next-Generation Sequencing for detecting the genome of SARS-CoV-2, databases storing SARS-CoV-2 genomes and variants, bioinformatics software tools and databases for analyzing and storing host–virus interactions [3].

At the medical level and in particular when considering the search for therapeutic strategies, the identification of COVID-19 biomarkers, the discovery of therapeutic targets for drugs and the bioinformatics approaches for drug repurposing, i.e. the use of already available drugs for the COVID-19 disease, are main research themes.

At the epidemiological and public-health level, main research themes regard: the systematic collection and sharing of data about the spread of the infection, such as the number of cases, hospitalized, ICU and deceased patients, that may be helpful to manage the pandemic [4]; the biological tests for testing, and the computational methods for tracing and tracking infected people; the exploitation of the vast clinical data stored into the Electronic Health Records of COVID-19 patients [5]; the analysis of the impact of lockdown measures in various contexts, e.g. at socioeconomic level, that may benefit from sentiment analysis methods; and finally measures to help quarantined people, such as local healthcare service, robotics and virtual assistants.

Finally, those unprecedented research efforts yield an overwhelming volume of scientific publications that require new methods and tools to improve learning from SARS-CoV-2 and COVID-19 literature, such as novel text mining and natural language processing techniques to distill relevant information [6].

This Special Issue aims to collect relevant scientific contributions on methods and applications of bioinformatics and informatics in themes related to COVID-19 and SARS-CoV-2. In particular, the special issue is organized in two main strands: one on *Bioinformatics helping to mitigate the impact of Covid-19* and another one on *Informatics helping to mitigate the impact of Covid-19*.

Here, we present the first-strand *Bioinformatics helping to mitigate the impact of Covid-19* that comprises more than 60 manuscripts, each dealing with one of the following central key issues, as detailed below.

### Bioinformatics tools and resources for SARS-CoV-2 and COVID-19 research

Next-generation sequencing is the central technology for detecting genomes of SARS-CoV-2 that provides the basic data about the virus. Bioinformatics pipelines, biological and host–virus interaction databases, are key tools for computing such data and advancing knowledge on SARS-CoV-2.

In *Next-generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities*, Chiara, D’Erchia, Gissi, Manzari, Parisi, Resta, Zambelli, Picardi, Pavesi, Horner and Pesole discuss next-generation sequencing (NGS), a fundamental technology and method for tracing origins and understanding the evolution of infectious agents, and in particular to reconstruct the genomic sequence of SARS-CoV-2. Authors briefly introduce available platforms and approaches for the sequencing of SARS-CoV-2 genomes and outline current databases for SARS-CoV-2 genomic data. As a result, they provide some useful guidelines for the sharing and deposition of SARS-CoV-2 data and metadata, suggesting the use of efficient and standardized approaches for the production, handling and integration of SARS-CoV-2 sequencing data.

In *Bioinformatics resources for SARS-CoV-2 discovery and surveillance*, Hu, J. Li, Zhou, C. Li, Holmes and Shi discuss the role of next-generation sequencing and available bioinformatics pipelines for the worldwide genomic surveillance of SARS-CoV-2, focusing on the tracking of COVID-19 spread and the analysis of evolution and patterns of SARS-CoV-2 variation on a global scale. The authors review the main bioinformatics resources available for the discovery and surveillance of SARS-CoV-2 and discuss their advantages and disadvantages, highlighting areas needing urgent technical improvements.

In *Computational strategies to combat COVID-19: useful tools to accelerate SARS-CoV-2 and coronavirus research*, Franziska Hufsky et al. present bioinformatics tools that have been explicitly developed for SARS-CoV-2 with the aim to provide key tools for the detection, understanding and treatment of COVID-19. The reviewed tools include detection of SARS-CoV-2, analysis

of sequencing data, tracking and containment of the COVID-19 pandemic, study of coronavirus evolution, discovery of potential drug targets and related therapeutic strategies. All analyzed tools are available online and free to use and for each tool the authors describe a use case and discuss the contribution to the SARS-CoV-2 research.

In *A review on viral data sources and search systems for perspective mitigation of COVID-19*, Bernasconi, Canakoglu, Masseroli, Pinoli and Ceri discuss the data integration activities needed for accessing and searching SARS-CoV-2 genome sequences and metadata stored in main viral sequences databases. The authors review some host-pathogen integrated datasets and underline possible integrative surveillance mechanisms, e.g. based on the time-space distribution of common virus variants. They observe that while organizations already managing virus databases are offering novel specific SARS-CoV-2 data and services, novel specific approaches and resources to face COVID-19 are appearing, providing better accessibility of viral sequence data, integration with clinical data and with the genotype of the human host.

The role of pathway enrichment analysis (PEA) in finding possible targets present in biological pathways of host cells that are targeted by SARS-CoV-2 is discussed in *Comprehensive pathway enrichment analysis workflows: COVID-19 case study*. To guide bioinformaticians in the choice of the many available PEA methods and software tools, Agapito, Pastrello and Jurisica highlight how to choose the most suitable PEA methods based on the type of SARS-CoV-2/COVID-19 data to analyze.

In *Web tools to fight pandemics: the COVID-19 experience*, Mercatelli, Holding and Giorgi focus on the state of the art of COVID-19 online resources and review the most popular web tools for the analysis of COVID-19 data, focusing on the epidemiology, genomics, interactomics and pharmacology fields.

## COVID-19 biomarkers, drug targets and bioinformatics approaches for drug repurposing

The identification of COVID-19 biomarkers, the discovery of therapeutic targets for drugs and the bioinformatics approaches for drug repurposing are key research topics to address for facing the COVID-19 disease. The research in these fields is mainly driven by the SARS-CoV-2 proteins structure, protein dynamics produced by computer simulations, variants and mutations of the virus.

In *A review of COVID-19 biomarkers and drug targets: resources and tools*, Caruso, Scala, Cerulo and Ceccarelli present a review of tools and resources to identify biomarkers and drug targets in COVID-19, through the automatic analysis of a consolidated corpus of 27 570 papers. Using latent Dirichlet analysis, authors extracted topics associated with computational methods for biomarker identification and drug repurposing, which include machine learning and artificial intelligence for disease characterization, vaccine development and therapeutic target identification.

In *Bioinformatics resources facilitate understanding and harnessing clinical research of SARS-CoV-2*, Ahsan, Liu, Feng, Zhou, Ma, Bai and Chen review some bioinformatics resources, the status of drug development and various resources for enabling research toward effective treatment of COVID-19, including phylogenetic characteristics, genomic conservation and interaction data. The authors review several SARS-CoV-2-related tools and databases, focusing on bioinformatics approaches for target prioritization

and drug repurposing. They present a web-portal named OverCOVID that provides a detailed description of SARS-CoV-2 basics and shares a collection of bioinformatics resources and information that may contribute to better understanding of SARS-CoV-2 and to therapeutic advances.

In *A review on drug repurposing applicable to COVID-19*, Dotolo, Marabotti, Facchiano and Tagliaferri present a review of different drug repurposing strategies useful to face COVID-19 pandemic, i.e. strategies for discovering new applications of existing drugs to COVID-19, that may reduce costs and provide shorter time application. Authors categorize computational drug repurposing approaches into network, structure and artificial intelligence approaches. Network-based approaches, further categorized into clustering and propagation approaches, allow the identification of proteins that are functionally associated with COVID-19, evidencing novel drug-disease or drug-target relationships useful for new therapies. Structure-based approaches study how chemical compounds can interact with the macro molecular targets, finding new possible applications for existing drugs. Finally, artificial intelligence approaches are evaluated less relevant at the moment, due to the scarcity of data to learn models.

In *The impact of structural bioinformatics tools and resources on SARS-CoV-2 research and therapeutic strategies*, Waman, Sen, Varadi, Daina, Wodak, Zoete, Velankar and Orengo review recent structural bioinformatics tools and discuss the impact of structure-based studies on SARS-CoV-2 research, with focus on the differences between SARS-CoV-2 and SARS-CoV, the SARS-CoV-2 residues involved in receptor-antibody recognition, the variants in host proteins that affect susceptibility to infection, and the computational analyses enabling structure-based drug and vaccine development.

In *SARS-CoV-2 3D database: understanding the coronavirus proteome and evaluating possible drug targets*, Alsulami, Thomas, Jamasb, Beaudoin, Moghul, Bannerman, Copoiu, Vedithi, Torres and Blundell propose a new database containing 3D models of the SARS-CoV-2 proteome, including models of protomers and oligomers, protein-ligand docking, interactions of SARS-CoV-2 proteins with human proteins, impacts of mutations and experimental structures. The resulting SARS-CoV-2 3D database provides information for drug discovery, useful to evaluate targets and design new possible therapeutics.

## Knowledge extraction from SARS-CoV-2 and COVID-19 literature

The unprecedented rate of SARS-CoV-2 and COVID-19 publications strongly accelerated the development of text mining and natural language processing techniques to analyze scientific literature.

In *Text mining approaches for dealing with the rapidly expanding literature on COVID-19*, Wang and Lo tackle the problem of extracting recent knowledge from the overwhelming COVID-19 literature using text mining applications and discuss the corpora, models and systems that have been introduced for COVID-19. They analyzed 39 systems that support search, discovery, visualization and summarization of the COVID-19 literature, and categorized them through qualitative description, performance assessment and user interface. The authors note that some systems, in addition to standard functions such as search and discovery, provide new functions such as summary of multiple documents or connections between scientific articles and clinical trials.

In *How do we share data in COVID-19 research?* A systematic review of COVID-19 datasets in PubMed Central Articles, Zuo, Chen, Ohno-Machado and Xu review more than 100 datasets about COVID-19 that were reported into several scientific articles available from PubMed Central. Starting from 12 324 COVID-19 full-text articles published until 31 May 2020, the authors extracted the links to 128 datasets that were manually reviewed using 10 variables. Although the analysis was performed in an initial stage of the pandemic, the authors found 128 unique dataset links. The largest portion (53.9%) are epidemiological datasets and most datasets (84.4%) were available for immediate download. The study found that GitHub was the most used repository and evidenced a great heterogeneity in the way the datasets are mentioned, shared, and updated.

### Key infrastructures, technologies and applications for managing the COVID-19 pandemic

In addition to bioinformatics research, COVID-19 pandemic has given an impulse to the development and adaption of several informatics techniques, including computational methods for tracing and tracking infected people; collaborative data infrastructures for COVID-19 research; sentiment analysis methods for monitoring the impact of lockdown measures; artificial intelligence methods and robotics applications to support remote patients assistance (e.g. quarantined people).

In *Health informatics and EHR to support clinical research in the COVID-19 pandemic: an overview*, Dagliati, Malovini, Tibollo and Bellazzi discuss the role of Electronic Health Records (EHR) that are primarily used to support day-by-day clinical activities, to enable global scale research on COVID-19. Authors review collaborative data infrastructures to support COVID-19 research, including studies on effectiveness of drugs and therapeutic strategies, and discuss the data sharing and governance issues emerged with the COVID-19 pandemic, that may prevent a full exploitation of EHR data, especially when considering international collaborations. The authors underline the data management, interoperability and governance issues, the modelling of healthcare processes and the management of data privacy regulations as primary aspects to boost collaborative research.

In *Robots as intelligent assistants to face COVID-19 pandemic*, Seidita, Lanza, Pipitone and Chella discuss the role that an emerging technology such as robotics may have in the management and fight against the COVID-19 pandemic. Authors analyzed scientific articles and industrial initiatives underlining how robotics was used to face the pandemic, its level of readiness, what are the expectations from robots and what remains to do. Authors reviewed what is offered by research groups in terms of robot support for therapies and for prevention actions and discussed the maturity of robotics in dealing with situations like COVID-19.

In *HVIDB: a comprehensive database for human-virus protein-protein interaction*, X. Yang, Lian, Fu, Wuchty, S. Yang and Zhang present HVIDB (Human-Virus Interaction DataBase), an annotated human-virus protein-protein interaction (PPI) database that contains experimentally verified human-virus PPIs about 35 virus families, experimentally verified 3D complex structures of

human-virus PPIs, and integrates machine learning models to predict interactions between human host and viral proteins.

Although research on SARS-CoV-2 and COVID-19 is continuously evolving, we hope this special issue will represent an authoritative and valuable resource for researchers. The Editors are grateful to both the Editor-in-Chief and the Publisher for having sustained this project, for their timely help, and for having supported them in the day-to-day needs. A special thank is addressed to all the authors and reviewers, whose competence and effort allowed the realization of this special issue.

Mario Cannataro

Data Analytics Research Center,  
Department of Medical and Surgical Sciences, University  
"Magna Graecia" of Catanzaro, Italy, e-mail: cannataro@unicz.it

Andrew Harrison

Department of Mathematical Sciences, University of Essex,  
United Kingdom, e-mail: harry@essex.ac.uk

Mario Cannataro is a Full Professor of computer engineering at the University 'Magna Graecia' of Catanzaro, Italy, and the Director of the Data Analytics Research Center. His current research interests include bioinformatics, health informatics, artificial intelligence, data mining and parallel computing. He published three books and more than 300 papers in international journals and conference proceedings. Mario Cannataro is a Member of the Board of Directors of ACM SIGBio, and a Senior Member of ACM, IEEE, BITS (Bioinformatics Italian Society) and SIBIM (Italian Society of Biomedical Informatics).

Andrew Harrison is a Senior Lecturer in Data Science at the University of Essex, UK. His current research interests include the analysis of Hi-C experiments as well as large-scale meta-analysis of gene expression experiments. He has published over 50 papers on a broad range of topics from data science, applied statistics, bioinformatics, analysis of high-throughput biological experiments and astrophysics.

### References

1. Ben H, Hua G, Peng Z, et al. Characteristics of sars-cov-2 and covid-19. *Nat Rev Microbiol* 2020; **19**:141–54.
2. Forni G, Mantovani A. Covid-19 vaccines: where we stand and challenges ahead. *Cell Death Diffe* 2021; **28**:626–39.
3. Ray M, Sable MN, Sarkar S, et al. Essential interpretations of bioinformatics in covid-19 pandemic. *Meta Gene* 2021; **27**:100844.
4. Dong E, Du H, Gardner L. An interactive web-based dashboard to track covid-19 in real time. *Lancet Infect Dis* 2020; **10**:533–4.
5. Weber G, et al. International comparisons of harmonized laboratory value trajectories to predict severe covid-19: leveraging the 4ce collaborative across 342 hospitals and 6 countries: a retrospective cohort study medRxiv preprint 2020.12.16.20247684. 2021.
6. Murillo J, Villegas LM, Ulloa-Murillo LM, et al. Recent trends on omics and bioinformatics approaches to study sars-cov-2: a bibliometric analysis and mini-review. *Comput Biol Med* 2021; **128**:104162.