Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Commentary

Repurposing Biomedical Informaticians for COVID-19

Daniel N. Sosa, Binbin Chen, Amit Kaushal, Adam Lavertu, Jake Lever, Stefano Rensi, Russ Altman

PII: S1532-0464(21)00002-2
DOI: https://doi.org/10.1016/j.jbi.2021.103673
Reference: YJBIN 103673

To appear in: Journal of Biomedical Informatics

Received Date: 6 October 2020
Revised Date: 20 December 2020
Accepted Date: 2 January 2021

Please cite this article as: Sosa, D.N., Chen, B., Kaushal, A., Lavertu, A., Lever, J., Rensi, S., Altman, R., Repurposing Biomedical Informaticians for COVID-19, Journal of Biomedical Informatics (2021), doi: https://doi.org/10.1016/j.jbi.2021.103673

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Elsevier Inc. All rights reserved.
Repurposing Biomedical Informaticians for COVID-19

Daniel N. Sosa\textsuperscript{a}, Binbin Chen\textsuperscript{b,c}, Amit Kaushal\textsuperscript{a,c}, Adam Lavertu\textsuperscript{a}, Jake Lever\textsuperscript{d}, Stefano Rensi\textsuperscript{d}, Russ Altman\textsuperscript{a,b,c,d,*}

\textsuperscript{a}Department of Biomedical Data Science, Stanford University, Stanford, CA, USA
\textsuperscript{b}Department of Genetics, Stanford University, Stanford, CA, USA
\textsuperscript{c}Department of Medicine, Stanford University, Stanford, CA, USA
\textsuperscript{d}Department of Bioengineering, Stanford University, Stanford, CA, USA

Abstract

The COVID-19 pandemic is an unprecedented challenge to the biomedical research community at the intersection of great uncertainty due to the novelty of the virus and extremely high stakes due to the large global death count. The global quarantine shut-downs complicated scientific matters because many laboratories were closed down unless they were actively doing COVID-19 related research, making repurposing of activities difficult for many biomedical researchers. Biomedical informaticians, who have been primarily able to continue their research through remote work and video conferencing, have been able to maintain normal activities. In addition to continuing ongoing studies, there has been great grass roots interest in helping in the fight against COVID-19. In this commentary, we describe several projects that arose from this desire to help, and the lessons that the authors learned along the way. We then offer some insights into how these lessons might be applied to make scientific progress be more efficient in future crisis scenarios.

Keywords:
COVID-19, Structural Bioinformatics, Biomedical Literature, Translational Research

1. Introduction

As of this writing on October 1st, 2020, COVID-19 has killed more than 1 million people globally. This pandemic has presented an unprecedented challenge to the scientific community due to the urgency with which we need preventative or palliative treatments and the great uncertainty surrounding this novel virus and disease. To complicate matters, the quarantine shut-downs that have been put in place across the globe prevented researchers from going into lab if their research was not actively COVID-19-related. In such cases, it was difficult for wet lab biologists to repurpose their resources for studying COVID-19, which may have been a missed opportunity in times of a global pandemic.

By contrast, biomedical informaticians have been largely able to continue normal research activities since their research is primarily computational. Many of us have had a strong desire to help however we can in this pandemic. In this commentary, we
describe the stories of how some of us in a computational pharmacology lab translated this desire to make an impact into projects that harnessed our bioinformatics research toolkits. We describe the challenges faced along the way and the key lessons learned from our experiences so far.

2. Anecdotes of Repurposed Biomedical Informaticians

Binbin, a recent Genetics PhD and soon-to-be MD graduate, saw immediate parallels between his previous work on applying artificial intelligence for designing therapeutic cancer vaccines [1] and developing vaccine candidates for the new SARS-CoV-2 virus. This work focused on using AI tools built upon immunology knowledge and data to provide principled “educated guesses” for immunogenic fragments, which can increase the chances of finding an effective vaccine. In the preprint he published online in February 2020, Binbin's work highlighted the importance of spike protein as a vaccine candidate and 405 viral antigens for vaccine efficacy [2]. Currently, his team is collaborating with immunologists and pathologists at Stanford to refine this antigen list for either COVID-19 vaccine or diagnosis.

Stefano, a Research Engineer in the Bioengineering Department, has previously worked at the intersection of chemoinformatics and natural language processing. Having heard much conflicting news about SARS-CoV-2 from reports in early January 2020, Stefano used a text mining tool he had recently developed, Docs2Graph, to find answers [3]. This tool uses natural language processing (NLP) and semantic technologies to help researchers “connect the dots” across different publications by synthesizing, organizing, and summarizing the relationships between drugs, genes, proteins, and diseases described in scientific abstracts. He used this information to identify a promising therapeutic target, TMPRSS2, and virtually screen libraries of compounds that had been previously tested in humans to identify several potential inhibitors, which were also tested in experimental assays [4].

Jake, a postdoctoral researcher studying NLP and machine learning for processing biomedical literature, applied his previous experience in building resources for collating literature-extracted biological associations on cancer genetics and mutations associated with drug dynamics and kinetics. In the context of COVID-19, Jake sought resources to help researchers quickly digest the flood of SARS-CoV-2 and related coronavirus literature, and found that a comprehensive, up-to-date central repository of categorized research was not to be found. By repurposing several existing pipelines for managing large volumes of biomedical literature, leveraging Kaggle's CORD-19 literature dataset [5], and building a custom topic ontology through manual curation effort, Jake built a system to display current coronavirus research organized by content. This system also uses machine learning to automatically update the database as new papers are published [6].

Dan is a PhD student in Biomedical Informatics whose research focuses on extracting and discovering new knowledge from pharmacologically-relevant biomedical literature. When Kaggle's CORD-19 challenge began, Dan encountered and joined a global crowd-sourced team, CoronaWhy, whose focus was on addressing some of Kaggle's challenges [7]. Through his work with CoronaWhy, Dan was able to combine
his technical training with his previous background in management to lead a global team of 20 volunteers representing a variety of disciplines to pursue various projects concerning COVID-19 therapeutics. The work in this group spanned projects including automating detection of contradictory claims in pharmacological literature about drug efficacy and analyzing Twitter sentiment around drug-related publications in CORD-19.

Adam and Amit are a PhD candidate in Biomedical Informatics and an internal medicine physician on the front line of COVID-19, respectively, who teamed up on a couple of important problems related to contact tracing. First, after reaching out to public health officials, they identified a key bottleneck to enable effective contact tracing at scale. Manually reading and processing the large volume of faxed COVID-19 reports to identify cases that were high-risk for further community spread is a laborious process that has attracted national attention [8]. Drawing on computer vision methods, the two built a pipeline to process handwritten faxes when they arrive at the county public health office and automatically send alerts if the report indicates high risk for community spread. Currently, they are currently working with a local public health partner to deploy this method, which could significantly alleviate a large manual burden. Secondly, Adam and Amit pursued a project investigating whether smartphone-based contact tracing could be effective with adoption by only a subset of the population [9]. The two are developing solutions to this pressing problem.

3. Discussion and Lessons Learned

In undertaking these new research directions, the authors were able to repurpose their technical training to address needs during the pandemic. Real-world problems introduced by the novelty of the pandemic, the deluge of scientific research, and the complexity of modern workflows necessitate the applications of structural informatics, machine learning, natural language processing, and computer vision. As biomedical informaticians we are flexible and ready to apply these general-purpose tools to problems with demonstrable need. From the authors’ collective experiences, a few salient points became clear about the scientific response to pandemic and related crisis situations.

First, we need a robust system through which people with addressable problems related to pandemic response can request assistance from researchers. In many cases, problems, such as delays caused by the processing of faxed documents, can be solved through automation and application of modern informatics methods. As Adam and Amit learned, finding these needs is often non-obvious and may even require front-line ethnographic interviews to ultimately discover key bottlenecks. Solutions that improve clinical workflows might be more impactful opportunities for applying modern informatics methods than pursuing low-hanging machine learning applications, which might be impractical to implement in real-world settings for a variety of reasons such as regulations, data availability, or clinical infrastructure. Creating a system to short-circuit the process of identifying important needs by connecting front-line users to a “network of scientists” would greatly improve the efficiency of finding and solving addressable needs at scale in pandemic-like scenarios. Without such coordinated efforts, time may be wasted by researchers working on problems with little or no impact.
Second, we need a vehicle through which the scientific response to identified needs is well-coordinated. In this time of crisis, many people are finding themselves with a surplus of free time and a strong desire to offer their skills to fight the pandemic however they can. Given the easy access to educational resources about modern data science skills, many citizen scientists are eager and well-equipped to apply their learning to well-defined problems. By having a coordinated response, perhaps adopting a model similar to the CoronaWhy group, those citizen scientists in conjunction with academic researchers across the globe can leverage their skills by forming an open-science network of informatics problem-solvers with a low barrier to entry. In Dan’s experience, he found that a volunteer-based, mission-driven model such as CoronaWhy organically fostered a very positive and agile environment of mutual support and member development across the organization. Such organizations do require great care in coordination, however, and it will be important to establish ground rules early on to address such challenges as: who is creating new projects and ensuring they will have high impact (i.e. have front-line stakeholders in the loop early on), how will the organization be structured to balance agility with effective coordination, and what systems will be in place for onboarding new members and directing them to the right projects within the organization. With the immediacy of COVID-19, the CoronaWhy group found itself navigating these difficult questions on the fly. In laying the groundwork in preparation for future pandemics or even generally considering new models for global-scale scientific collaboration, careful consideration of these points will ensure that every member does work that is valued and impactful in accordance with their ability and interests.

Finally, it is imperative that science is communicated and disseminated well. Preprint servers have certainly helped increase the speed and visibility by which new results become public, however their role in pandemic scenarios has become somewhat unclear. Because preprint servers host non-peer reviewed manuscripts, they have traditionally been viewed as simply repositories of works in progress with the large caveat that they have not yet met the scientific gold standard of peer review. With time being of the essence during the COVID-19 pandemic, this expectation of uncertain preprint quality was taken for granted to the point where some of the authors in this commentary even encountered difficulties posting a draft manuscript to one preprint server because that server had blocked manuscripts concerning COVID-19 and potential therapeutic opportunities out of an abundance of caution in safeguarding against misinformation. In this instance, the preprint server took the role of a gatekeeper, breaking with traditional expectations of the purpose of preprint servers.

Because of all the new research being conducted, peer reviewed or otherwise, there is an urgent need for this information to be organized in a manner that is intuitive to end users so they can quickly discover research in the rapidly evolving scientific body of literature, as Jake noted. Systems for organizing new scientific findings and even doing meta-scientific processing, such as detecting contradictions as new results emerge, may help make sense of the inundation of science.

Another problem is that in many cases, several research groups and bottom-up initiatives end up solving the same problems independently. While some redundancy contributes to producing robust scientific findings, having a joint body of academic, industry, and government partners, such as the CAIAC initiative [10], that can...
coordinate all of these efforts across a variety of domain applications can help reduce unnecessary research redundancy and create a more efficient response to common problems. By addressing the need for properly communicating science within the scientific community and to specific end users, we can create a much more effective scientific response when time is of the essence.

4. Conclusion

Being poised at the intersection of data science, biology, and medicine, biomedical informatics has a unique flexibility to address a variety of problems in times of public health crises. Computational labs are able to perform in silico studies to simulate and virtually screen potential therapeutics, which can jump-start downstream experimental and clinical screens for feasible drug repurposing opportunities in global emergency scenarios such as the COVID-19 pandemic. We have described six ad hoc projects by which the authors were able to contribute to some of the challenges presented by this pandemic, and we discuss key considerations for a more coordinated research response, which could greatly improve the efficiency of science in future crisis scenarios. By reimagining scientific infrastructure throughout the biomedical research enterprise, a coordinated system may also transform our efficiency in responding to routine scientific challenges.

5. Acknowledgements

RBA and JL are supported by NIH GM102365, LM05652 and the Chan Zuckerberg Biohub. RBA and AK are supported by a Coulter Foundation Award. SR is supported by NIH TR002515. BC is supported by an NIH/Stanford MSTP training grant, AL by a National Science Foundation Graduate Research Fellowship, and DS by a National Library of Medicine Training Grant T15 LM007033.

References

1. Chen, B., Khodadoust, M. S., Olsson, N., Wagar, L. E., Fast, E., Liu, C. L., Muftuoglu, Y., Sworder, B. J., Diehn, M., Levy, R., Davis, M. M., Elias, J. E., Altman, R. B., and Alizadeh, A. A. (2019). Predicting HLA class II antigen presentation through integrated deep learning. Nature Biotechnology, 37(11), 1332?1343. https://doi.org/10.1038/s41587-019-0280-2

2. Fast, E., Altman, R. B., and Chen, B. (2020). Potential T-cell and B-cell Epitopes of 2019-nCoV. BioRxiv, 2020.02.19.955484.
3. Buhler, K. (2020). What A Stanford Researcher’s Fight Against Covid-19 Can Tell Us About The Future Of Drug Discovery. Retrieved August 18, 2020, from https://www.forbes.com/sites/konstantinebuhler/2020/07/07/what-a-stanford-researchers-fight-against-covid-19-can-tell-us-about-the-future-of-drug-discovery/\#410403d91e01

4. Rensi, S., B Altman, R., Liu, T., Lo, Y.-C., McInnes, G., Derry, A., and Keys, A. (2020). Homology Modeling of TMPRSS2 Yields Candidate Drugs That May Inhibit Entry of SARS-CoV-2 into Human Cells. https://doi.org/10.26434/chemrxiv.12009582.v1

5. Lu Wang, L., Lo, K., Chandrasekhar, Y., et al (2020). CORD-19: The Covid-19 Open Research Dataset. ArXiv. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7251955/

6. Lever, J. (2020). CoronaCentral. Retrieved December 18, 2020, from https://coronacentral.ai/

7. Help Fight Coronavirus | CoronaWhy. (2020). Retrieved December 18, 2020, from https://www.coronawhy.org/

8. Kliff, S. and Sanger-Katz, M. (2020). Bottleneck for U.S. Coronavirus Response: The Fax Machine. The New York Times. https://www.nytimes.com/2020/07/13/upshot/coronavirus-response-fax-machines.html

9. Kaushal, A. and Altman, R. B. (2020). Can Contact Tracing Work At COVID Scale? | Health Affairs. Retrieved October 1, 2020, from https://www.healthaffairs.org/do/10.1377/hblog20200630.746159/full/

10. Wiggers, K. (2020). The Future Society, UNESCO, and others partner on pandemic decision-making platform | VentureBeat. Retrieved October 1, 2020, from https://venturebeat.com/2020/07/09/the-future-society-unesco-and-others-announce-partnership-to-build-pandemic-decision-making-platform/

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:
