Human Intrigue: Meta-analysis approaches for big questions with big data while shaking up the peer review process

Carly A. Bobak, Meghan Muse, Kristine A. Giffin
Program in Quantitative Biomedical Sciences, Dartmouth College
Hanover, NH 03755, USA
Email: Carly.A.Bobak.GR, Meghan.E.Muse.GR, Kristine.A.Giffin@Dartmouth.edu

Derek A. Williamson
HigherMe CEO
Toronto, ON, Canada
Email: derek@higherme.com

Casey S. Greene
Center for AI Health, University of Colorado's School of Medicine
Aurora, CO 80045
Email: Casey.S.Greene@cuanschutz.edu

Jason H. Moore
Department of Computational Biomedicine
Cedars-Sinai Medical Center
Los Angeles, CA 90048
Email: jason.moore@cshs.org

Dennis P. Wall
Pediatrics (Systems Medicine) and Biomedical Data Science, Stanford University
Stanford, CA 94305
Email: dpwall@stanford.edu

Scientific innovation has long been heralded the collaborative effort of many people, groups, and studies to drive forward research. However, the traditional peer review process relies on reviewers acting in a silo to critically judge research. As research becomes more cross-disciplinary, finding reviewers with appropriate expertise to provide feedback on an entire paper is increasingly difficult. We sought to pilot a crowd peer review process that allowed reviewers to interact with one another in the spirit of collaborative science. We focused this session on manuscripts using meta-analysis, to fully embrace the importance of collaborative and open scientific research in the field of biocomputing. Our pilot study found that researchers enjoy a more collaborative peer review process and felt that the process led to higher quality feedback for submitting authors than traditional review offers.

Keywords: Crowd peer review, collaborative scientific review, meta-analysis, publicly available data, methods development

1. Introduction

Peer review has long been a staple of the scientific community, dating back as early as the 9th century when Al-Ruhawi sought peer feedback on his book on physician ethics[1]. For centuries the peer review process relied on the physical delivery of manuscripts to established colleagues for feedback, until it was officially implemented in the early 20th century by *Science*, *The Journal of the American Medical Association (JAMA)*, and *American Practitioner* as in its current form as a method to both identify manuscripts that should be accepted/rejected from publications, but also to address and revise gaps in manuscripts to improve before subsequent publication[1].

While peer review is recognized as an important institution in maintaining the integrity of the scientific community, it has flaws that are widely accepted. 'Reviewer 2' jokes are often shared around academic circles based on the common occurrence of one reviewer being unnecessarily difficult, or even insulting in their manuscript feedback.[2] Moreover, despite the many rounds of revisions that authors, reviewers, and editors conduct on a manuscript, blatant mistakes still slip through the cracks. One example of this includes an ecology paper, which was published containing the thought "should we cite the crappy Gabor paper here".[3] Other studies have shown that the peer review process introduces bias and systemic discrimination and can hold authors hostage in a community where 'publish or perish' reigns true[1,4]. It was perhaps best stated by R. Smith in 1999 that peer review is "slow, expensive, highly subjective, prone to bias, easily abused, and poor at detecting gross defects."[5]

One of the major influences on these problems is 'reviewer fatigue'. As the number of scientific publications increases every year, so does the reliance on volunteer reviewers – particularly those who are most renowned or are at the top of their field. In most cases, reviewers are not compensated or credited for their reviews or contributions. As such, it can be difficult for editors to identify qualified reviewers who have the appropriate time to commit to conducting reviews. This leads to a lack of response to editors, poor quality or rushed reviews, resorting to unqualified reviewers, and longer publishing times[1,2].

Moreover, as many areas of science embrace multi-disciplinary methods, it is increasingly difficult to identify reviewers who are qualified to provide feedback on entire manuscripts. In many cases, editors resort to finding one reviewer to represent one field of research, and hence lack any diversity of opinions regarding the quality of the science.

To address these gaps in the peer review process, Benjamin List has explored the idea of a selected crowd-based review process in his chemistry publication *Synlett.* List sought to recreate the idea of scientific discussion in his peer reviews – where he hypothesized that reviewers who could discuss with each other would provide better feedback and recommendations. To facilitate this process, rather than focus on two or three reviewers per manuscript, he instead recruits many reviewers to engage with the manuscript over a shorter time period. Each reviewer sees both the current manuscript version, as well as the other reviewers anonymized feedback which they can comment on to engage in scientific discussion[2,6].

List and his team ran small trials with this method where they had manuscripts that underwent both the traditional and crowd review process. The crowd review method has many merits. One, it reduces the workload for any individual reviewer and hence mitigates reviewer fatigue. Two, it allows reviewers to focus on sections of a manuscript they are qualified to give feedback on. Three, due to the nature of the scientific discussion, the feedback tends to be stronger and reach an informative consensus among reviewers. Four, a greater diversity of reviewers can be included, improving the overall quality of a final manuscript. Finally, and perhaps most compellingly for authors, the turnaround time for feedback is much faster than in the traditional review process - as short as 4 or 5 days using List's model[2,6].

Due to the increasingly interdisciplinary nature of biocomputing, we sought to investigate the potential of a crowd peer review process as part of a session at the Pacific Symposium of Biocomputing for 2022. As part of this process, we pledged to create a platform that would allow reviewers to engage in discussion about manuscripts through a comment section to allow debate back and forth, seek a wide diversity of reviewers for each submission, and formally acknowledge reviewer contributions in published proceedings. We focused our session on meta-analytics methods and applications for big data as these approaches are often impacted by publication bias, in part due to the nature of traditional peer review.

2. The Crowd Peer Review Process

We set out to establish a minimum viable product (MVP) for the purposes of crowd peer review. We recruited 26 reviewers including graduate level students, established faculty, and industry professionals to participate in the crowd review, and averaged 4.83 ($sd = 1.47$) reviewers per submitted manuscript. Reviewers participated within a Google Workspace, where they were instructed to comment on submitted manuscripts, interact with other comments, and fill out a form with overall quantitative scores and recommendations.

2.1 Reviewer's Feedback

All reviewers were invited to an anonymous feedback survey, where they were asked to rank their agreement with various statements about the crowd peer review process on a five-point scale (1: Strongly Disagree, 3: Neutral, 5: Strongly Agree). Of the 26 reviewers, 13 elected to participate in the feedback survey.

All 13 reviewers indicated that this was their first experience with a crowd review process. When participants were asked if they agreed to participate because of the peer review process, five respondents indicated that they strongly agreed with this statement, although most were neutral ($\bar{x} = 3.62, sd = 1.33$). Surprisingly, only one reviewer indicated that they agreed to participate because they would receive in-proceedings accreditation for their feedback ($\bar{x} = 2.85, sd = 1.14$). However, reviewers overwhelming indicated they would participate in a crowd review process again ($\bar{x} = 4.69, sd = 0.48$) and that the crowd review process was overall more enjoyable than traditional review ($\bar{x} = 4, sd = 0.91$).

Reviewers also felt that the crowd review process better allowed them to focus on areas of their expertise ($\bar{x} = 4.31, sd = 0.85$) and reduced the overall burden of review ($\bar{x} = 4.08, sd = 1.19$). Reviewers indicated that being able to see other's reviews increased the confidence of their overall feedback ($\bar{x} = 4.15, sd = 0.80$), but did not believe that seeing other reviews biased their feedback ($\bar{x} = 2.08, sd = 0.95$). Overall, reviewers were supportive and collaborative, with most strongly agreeing that other reviewers had provided high quality feedback ($\bar{x} = 4.69, sd = 0.48$).

When asked what features could be included to further improve the crowd review process, reviewers indicated they were most interested in being able to up- or downvote feedback (n=7), 'tag-in' other colleagues who may have relevant feedback (n=6), add a poll inline (n=5), and be able to toggle on/off the ability to see other reviewers' feedback (n=5).

2.2 Conclusions

Overall, reviewers indicated that there is an appetite for a more collaborative review process, and that this process provided high quality feedback to submitting authors. While the sample size of our feedback is small, other cross-disciplinary fields may want to consider crowd peer review as part of the scientific publication process.

3. Meta-Analysis in Biocomputing

Our 2022 session accepted three general categories of papers. First, we had papers proposing novel computational methods appropriate for the meta-analysis of 'omics level data. Second, had papers using many publicly available datasets to create and improve upon methodologies for studying specific biocomputational problems, and lastly, we had papers which aimed to study the structure of publicly available data itself to make it more appropriate for meta-analysis.

3.1 Novel Methods for Meta-Analysis of 'Omics Data

Nouira et. al propose a multi-task Lasso for the multivariate analysis of admixed GWAS data. They note that GWAS data is limited due to differences in population structure as well as linkage disequilibrium. They consider a model where each task corresponds to a subpopulation (or dataset in a meta-analysis context) and each group corresponds to a linkage disequilibrium block. They demonstrate that this method is efficient, robust, and outperforms state-of-the-art methodologies in both simulated and real data[7].

Levy et al. explore the use of mixed effect machine learning models to better account for batch effects in spatial 'omics applications. They show that classifiers should consider accounting for repeated measurements at the class and batch level, possibly using an MEML structure, to achieve optimal non-linear performance in studying pathology images for colon cancer[8].

Khatri et al, our guest speakers, have been proponents of the use of meta-analysis in clinical facing diagnostics. Their meta-analytic framework has identified robust gene expression biomarkers for the diagnosis and prediction of presence of sepsis[12,13], bacterial vs viral causes of infectious disease[14], and severity of infections (predicting risk of 30-day mortality[15], predicting severe vs

non-severe viral infections[16]). They've also proposed a 3 gene signature for the diagnosis of Tuberculosis that is being translated in a point-of-care cartridge by Cepheid and has been validated in a prospective cohort across 4 countries in less than 5 years[17-20]. Their work continues to demonstrate that meta-analysis and the use of publicly available data has enormous potential and is an important tool in translation biocomputational research into clinical settings.

3.2 Using Publicly Available Data in Methods Development

Romano et al. sought to improve Quantitative Structure Activity Relationship (QSAR) modeling, which is already a meta-analysis approach regularly used in toxicological research. To do this, they augmented traditional QSAR with publicly available multi-modal datasets and analyzed that data using a graph convolutional neural network model (GCN). They demonstrate that not only does this approach substantially improve upon traditional QSAR modeling, but that it is also more interpretable[9].

Schiebout et al. developed a novel single cell RNA seq method, Cell-typing using variance Adjusted Mahalanobis distances with Multi-Labeling (CAMML), for both customizable cell-typing and quantification of cell differentiation. They leveraged several publicly available single cell RNA sequencing datasets with annotated cell types to inform their model. They show that this method is comparable to gold-standard methods for single cell typing, but that it preserves important information for cells which exhibit features of multiple cell types, as is the case with undifferentiated cells[10].

3.3 Studying the Structure of Publicly Available Data

Rubel et al. note that many different biological pathway databases exist and are widely used in meta-analysis. They observe that similar pathways representing the same biological phenotype exist across many databases, and that these representations have structural variation which may impact biological findings. In their study, they employed the use of undirected graphlets (small sub-networks) to study the structural similarity of pathways across many publicly available databases. They show that database structural similarity exists, but that it can be accounted for to uncover pathway specific topology[11].

3.4 Conclusions

In conclusion, meta-analysis remains an important cornerstone in the field of biocomputing, not only for informing reproducible biological findings, but also in the development of new methodologies to drive forward biological discovery in the era of big data.

Acknowledgements

We want to thank the many researchers who participated in the crowd peer review process and provided invaluable feedback both to submitting authors, as well as to the session organizers. Our deepest thanks go to: Abhimanyu, Max Bluestone, Christian Darabos, Amy T. Farnham, Julian Gullett, Ariana Haidari, Yehudit Hasin, Preethi Krishnan, Damien Lekkas, Joshua J. Levy, Almut

Luetge, Michael P. Mariani, Lauren McDonnell, Brett McKinney, Quang P. Nguyen, Vishal Oza Shachi Patel, Xin Ran, Iben Rickett, Catherine Sherman, Mirjana Stevanovic, Alexander Titus, Van Truong, Yang Xu, Fan Zhang, and Ronnie J. Zipkin. A special thanks goes to Alexander Titus and Michelle Holko for their assistance and guidance in developing the minimum viable product for peer crowd review.

References

1. Haffar, S., Bazerbachi, F. & Murad, M. H. Peer Review Bias: A Critical Review. *Mayo Clinic Proceedings* **94**, 670–676 (2019).
2. Van Gemmeren, M. & List, B. How and Why Crowd Reviewing Works. *Synlett* (2020). doi:10.1055/s-0040-1705955
3. Culumber, Z. W., Bautista-Hernández, C. E., Monks, S., Arias-Rodriguez, L. & Tobler, M. Variation in Melanism and Female Preference in Proximate but Ecologically Distinct Environments. *Ethology* **120**, 1090–1100 (2014).
4. Gender Bias and the Peer Review Process. Available at: https://www.wiley.com/network/researchers/being-a-peer-reviewer/gender-bias-and-the-peer-review-process. (Accessed: 30th January 2021)
5. Smith, R. Opening up BMJ peer review. *British Medical Journal* **318**, 4–5 (1999).
6. List, B. Crowd-based peer review can be good and fast. *Nature* **546**, 9 (2017).
7. Nouira, A. & Azencott, C.A. Multitask group Lasso for Genome Wide association Studies in admixed populations. In *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2022.* (2022).
8. Levy, J.J., Bobak, C.A., Nasir-Moin, M., Veziroglu, E.M., Palisoul, S.M., Barney, R.E., Salas, L.A., Christensen, B.C., Tsongalis, G.J., & Vaickus, L.J. Mixed Effects Machine Learning Models for Colon Cancer Metastasis Prediction using Spatially Localized Immuno-Oncology Markers. In *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2022.* (2022).
9. Romano, J.D., Hao, Y., & Moore, J.H. Improving QSAR Modeling for Predictive Toxicology using Publicly Aggregated Semantic Graph Data and Graph Neural Networks. In *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2022.* (2022).
10. Schiebout, C.T. & Frost, H.R. CAMML: Multi-Label Immune Cell-Typing and Stemness Analysis for Single-Cell RNA-sequencing. In *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2022.* (2022).
11. Rubel, T, Singh, P., & Ritz, A. Stop Bickering! Reconciling Signaling Pathway Databases with Network Topologies. In *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2022.* (2022).
12. Sweeney, Timothy E., Wong, S.A., & Khatri, P. "A comprehensive time-course–based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set." *Science translational medicine* 7.287 (2015)
13. Sweeney, Timothy E., & Purvesh Khatri. "Benchmarking sepsis gene expression diagnostics using public data." *Critical care medicine* 45.1 (2017)
14. Sweeney, Timothy E., Hector R. Wong, & Purvesh Khatri. "Robust classification of bacterial and viral infections via integrated host gene expression diagnostics." *Science translational medicine* 8.346 (2016)

15. Sweeney, T.E., Perumal, T.M., Henao, R., Nichols, M., Howrylak, J.A., Choi, A.M., Bermejo-Martin, J.F., Almansa, R., Tamayo, E., Davenport, E.E., Burnham, K.L., Hinds, C.J., Knight, J.C., Woods, C.W., Kingsmore, S.F., Ginsburg, G.S., Wong, H.R., Parnell, G.P., Tang, B., Moldawer, L.L., Moore, F.E., Omberg, L.O., Khatri, P., Tsalik, E.L., Mangravite, L.M., & Langley, R.J. A community approach to mortality prediction in sepsis via gene expression analysis. *Nature communications*, *9*(1), pp.1-10. (2018).
16. Zheng, H., Rao, A. M., Dermadi, D., Toh, J., Jones, L. M., Donato, M., Su, Y., Dai, C.L., Kornilov, S.A. and Karagiannis, M., Marantos, T., Hasin-Brumshtein, Y., He, Y.D., Giamarello-Bourboulis, E.J., Heath, J.R., & Khatri, P. Multi-cohort analysis of host immune response identifies conserved protective and detrimental modules associated with severity across viruses. *Immunity*, *54*(4). (2021).
17. Sweeney, T.E., Braviak, L., Tato, C.M. & Khatri, P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. *The Lancet Respiratory Medicine*, *4*(3), pp.213-224. (2016).
18. Warsinske, H., Vashisht, R. & Khatri, P., 2019. Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures. *PLoS medicine*, *16*(4). (2019).
19. Warsinske, H.C., Rao, A.M., Moreira, F.M., Santos, P.C.P., Liu, A.B., Scott, M., Malherbe, S.T., Ronacher, K., Walzl, G., Winter, J. and Sweeney, T.E., Croda, J., Andrews, J.R., & Khatri, P. Assessment of validity of a blood-based 3-gene signature score for progression and diagnosis of tuberculosis, disease severity, and treatment response. *JAMA network open*, *1*(6). (2018).
20. Södersten, E., Ongarello, S., Mantsoki, A., Wyss, R., Persing, D.H., Banderby, S., Strömqvist Meuzelaar, L., Prieto, J., Gnanashanmugam, D., & Khatri, P. and Schumacher, S.G. Diagnostic Accuracy Study of a Novel Blood-Based Assay for Identification of Tuberculosis in People Living with HIV. *Journal of clinical microbiology*, *59*(3). (2021).