



## OPEN ACCESS

## EDITED AND REVIEWED BY

Sarah H. Elsea,  
Baylor College of Medicine, United States

## \*CORRESPONDENCE

Jiuyong Xie,  
✉ xiej@umanitoba.ca  
Robert Friedman,  
✉ bob.network.science@gmail.com

<sup>1</sup>Morgan Chu Visiting Professor at the Irell and Manella Graduate School, Beckman Research Institute at the City of Hope, Duarte, CA, United States

RECEIVED 11 May 2023

ACCEPTED 23 May 2023

PUBLISHED 02 June 2023

## CITATION

Xie J and Friedman R (2023), Editorial: Evolution in Neurogenomics.  
*Front. Genet.* 14:1220750.  
doi: 10.3389/fgene.2023.1220750

## COPYRIGHT

© 2023 Xie and Friedman. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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: Evolution in Neurogenomics

Jiuyong Xie<sup>1\*†</sup> and Robert Friedman<sup>2\*</sup>

<sup>1</sup>Department of Physiology and Pathophysiology, Max Rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, Canada, <sup>2</sup>Department of Biological Sciences (Retired), University of South Carolina, Columbia, SC, United States

## KEYWORDS

neurological disorders, genome, transcriptome, diagnosis, prognosis, machine learning

## Editorial on the Research Topic Evolution in Neurogenomics

To encourage further study of neurogenomics and disease by large scale DNA sequencing methods, we have selected five articles in our Research Topic entitled Evolution in Neurogenomics. This Research Topic includes neurodevelopmental disorders and disease, and approaches at the genomic, epigenomic, transcriptomic and epi-transcriptomic levels. In particular, [Kim et al.](#) explored clinical phenotypes and genetic variants by whole exome sequencing, a “massively parallel DNA-sequencing” method ([Rabbani et al., 2014](#)), in a cohort of pediatric patients with a diverse array of movement disorders. These are disorders that defy strict classification by conventional methodology. They successfully showed the potential of whole exome sequencing as a genetic diagnostic tool for yet another complex neurological disorder ([Retterer et al., 2016](#)).

Our second selection of the Research Topic is [Akter et al.](#) who surveyed the clinically relevant “copy number variants” across a large number of patients, an underrepresented population in Bangladesh, with neurodevelopmental disorders. They sampled these variants by chromosomal microarray analysis and droplet digital polymerase chain reaction, an approach that is relatively less precise as a diagnostic tool, yet it showed applicability for clinical use and at a relatively low cost. For a more definable pathology, [Hu et al.](#) applied a meta-analytical method, including a large number of studies and an overall sample of over 18,000 individuals of Chinese ancestry, and identified three single nucleotide polymorphisms that show association with risk of ischemic stroke.

The next two studies focused on glioma. For the first of these studies, [Zhang et al.](#) constructed and validated a risk score model for prognosis of low-grade gliomas. They based their analysis on 14 genes that are chromatin regulators, and included data from single-cell RNA-seq along with clinical data from The Cancer Genome Atlas. For the next study, [Zhang et al.](#) reported 12 m6A regulatory genes as putative biomarkers for the prognosis of glioma, a finding that included 1,600 samples. Both are disease association studies that include different methodologies, while the challenge is in synthesizing their findings for advancement of knowledge in neurology and biomedicine.

Likewise, current research in neurogenomics is showing a great potential for association of genetic features with disease, including by genome- and transcriptome-based analyses of genetic variation, long-read RNA sequencing ([Gao et al., 2023](#)), and spatial/temporal *in situ* genome and transcriptome mapping ([Longo et al., 2021](#); [Payne et al., 2021](#)). The studies of our Research Topic are exemplars of a heterogeneity of techniques for surveying genetic variation, such as by the use of microarray analysis or whole exome sequencing. Of special

interest are the DNA sequencing methods that generate very large samples of data, such as in whole exome sequencing. These techniques are both efficient and applicable for biological research since DNA sequencing is trending towards no cost (Wang et al., 2015), while the traditional cost of software-based analysis and human labor is becoming replaceable by advanced machine learning methods (Yang et al., 2020; Bansal et al., 2022).

Furthermore, models such as the Galactica large language model (Taylor et al., 2022) can facilitate the conversion of human readable data and tables to a machine-readable format, and allow for automated construction of biological databases, especially powerful in the case of organization of large samples of data from current DNA sequencing methods. The Galactica model also has a capability for associating biological data, whether processed or not, with scientific knowledge as composed in the common natural languages (Chen, 2022; Taylor et al., 2022). This can lead to a pooling of knowledge across biomedical journals and facilitates a broad and systematic survey of the literature. Moreover, it has a promise for automation of the process of scientific communication and that of data from large scale DNA sequencing (Whang et al., 2023).

The hope is that our Research Topic will stimulate neurogenomic studies that accurately map the genome to pathologies of the brain and the related tissues of the nervous system, while fusing the more traditional approaches with recent techniques of machine learning, as in the sequence-based deep learning models (Vaswani et al., 2017; Poplin et al., 2018; Baid et al., 2023; Liao et al., 2023). These recent techniques extend beyond the practice of data organization and systematic collection of knowledge, but also to application in algorithmic discovery (Fawzi et al., 2022), leading to the potential for automation of machine learning method selection—another area of

common interest and applicability to disease association studies. Our Research Topic is composed of studies that are particularly adapted to these approaches for automation of scientific research, leading to an acceleration of scientific discovery. This prediction is not only limited to that of biomedical research, but it is also applicable to the vast collections of data in clinical medicine that is not yet centralized nor fully accessible for machine-based retrieval methods (Topol, 2019).

## Author contributions

JX and RF wrote the editorial and revised it together. All authors contributed to the article and approved the submitted version.

## 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

- Baid, G., Cook, D. E., Shafin, K., Yun, T., Llinares-Lopez, F., Berthet, Q., et al. (2023). DeepConsensus improves the accuracy of sequences with a gap-aware sequence transformer. *Nat. Biotechnol.* 41, 232–238. doi:10.1038/s41587-022-01435-7
- Bansal, H., Goyal, A., and Choudhary, A. (2022). A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning. *Decis. Anal. J.* 3, 100071. doi:10.1016/j.dajour.2022.100071
- Chen, Y., and Zhang, D. (2022). Integration of knowledge and data in machine learning. Available at: <https://arxiv.org/abs/2202.10337> (Accessed February 15, 2022).
- Fawzi, A., Balog, M., Huang, A., Hubert, T., Romera-Paredes, B., Barekatin, M., et al. (2022). Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature* 610, 47–53. doi:10.1038/s41586-022-05172-4
- Gao, Y., Wang, F., Wang, R., Kutschera, E., Xu, Y., Xie, S., et al. (2023). Espresso: Robust discovery and quantification of transcript isoforms from error-prone long-read RNA-seq data. *Sci. Adv.* 9, eabq5072. doi:10.1126/sciadv.abq5072
- Liao, W. W., Asri, M., Ebler, J., Doerr, D., Haukness, M., Hickey, G., et al. (2023). A draft human pangenome reference. *Nature* 617, 312–324. doi:10.1038/s41586-023-05896-x
- Longo, S. K., Guo, M. G., Ji, A. L., and Khavari, P. A. (2021). Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nat. Rev. Genet.* 22, 627–644. doi:10.1038/s41576-021-00370-8
- Payne, A. C., Chiang, Z. D., Reginato, P. L., Mangiameli, S. M., Murray, E. M., Yao, C. C., et al. (2021). *In situ* genome sequencing resolves DNA sequence and structure in intact biological samples. *Science* 371, eaay3446. doi:10.1126/science.aay3446
- Poplin, R., Chang, P. C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., et al. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* 36, 983–987. doi:10.1038/nbt.4235
- Rabbani, B., Tekin, M., and Mahdih, N. (2014). The promise of whole-exome sequencing in medical genetics. *J. Hum. Genet.* 59, 5–15. doi:10.1038/jhg.2013.114
- Retterer, K., Juusola, J., Cho, M. T., Vitazka, P., Millan, F., Gibellini, F., et al. (2016). Clinical application of whole-exome sequencing across clinical indications. *Genet. Med.* 18, 696–704. doi:10.1038/gim.2015.148
- Taylor, R., Kardas, M., Cucurull, G., Scialom, T., Hartshorn, A., Saravia, E., et al. (2022). Galactica: A large language model for science. Available at: <https://arxiv.org/abs/2211.09085> (Accessed November 16, 2022).
- Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nat. Med.* 25, 44–56. doi:10.1038/s41591-018-0300-7
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Adv. neural Inf. Process. Syst.* 30. doi:10.48550/arXiv.1706.03762
- Wang, E., Zaman, N., McGee, S., Milanese, J. S., Masoudi-Nejad, A., and O'Connor-McCourt, M. (2015). Predictive genomics: A cancer hallmark network framework for predicting tumor clinical phenotypes using genome sequencing data. *Semin. Cancer Biol.* 30, 4–12. doi:10.1016/j.semcancer.2014.04.002
- Whang, S. E., Roh, Y., Song, H., and Lee, J. G. (2023). Data collection and quality challenges in deep learning: A data-centric ai perspective. *VLDB J.* 1–23. doi:10.1007/s00778-022-00775-9
- Yang, A., Zhang, W., Wang, J., Yang, K., Han, Y., and Zhang, L. (2020). Review on the application of machine learning algorithms in the sequence data mining of DNA. *Front. Bioeng. Biotechnol.* 8, 1032. doi:10.3389/fbioe.2020.01032