



OPEN ACCESS

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

*CORRESPONDENCE
Bruno César Feltes,
✉ bruno.feltes@ufrgs.br

RECEIVED 03 July 2023
ACCEPTED 10 July 2023
PUBLISHED 14 July 2023

CITATION

Feltes BC, Ligabue-Braun R and Dorn M (2023), Editorial: Computational and integrative approaches for developmental biology and molecular evolution. *Front. Genet.* 14:1252328. doi: 10.3389/fgene.2023.1252328

COPYRIGHT

© 2023 Feltes, Ligabue-Braun and Dorn. 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: Computational and integrative approaches for developmental biology and molecular evolution

Bruno César Feltes^{1*}, Rodrigo Ligabue-Braun² and Márcio Dorn^{3,4,5}

¹Department of Biophysics, Institute of Biosciences, Federal University of Rio Grande do Sul, Porto Alegre, Brazil, ²Department of Pharmacosciences, Federal University of Health Sciences of Porto Alegre, Porto Alegre, Brazil, ³Department of Theoretical Informatics, Institute of Informatics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil, ⁴Center of Biotechnology, Federal University of Rio Grande do Sul, Porto Alegre, Brazil, ⁵National Institute of Science and Technology: Forensics, Porto Alegre, Brazil

KEYWORDS

systems biology, bioinformatics, omics, machine learning, developmental biology, evolution

Editorial on the Research Topic

Computational and integrative approaches for developmental biology and molecular evolution

1 Introduction

Developmental Biology is a broad field where each type of biological data at the molecular, tissue, or phenotype level is pertinent to uncover an organism's origins and developmental process. The final panorama is that each layer of information requires distinct mindsets for accurate data interpretation.

As Theodosius Dobzhansky expressed, “*Nothing in biology makes sense except in the light of evolution,*” (Dobzhansky, 1973) and Developmental Biology flawlessly reflects Dobzhansky's statement. Because every piece of evolutionary characteristics are relevant to understanding an organism's developmental process, it is common to use model organisms to compare data or look for indications to best invest time, resources, and efforts. When a molecular process is conserved in multiple organisms throughout evolution, the probability of being essential for organism survival increases. Therefore, multiple data layers are required to rigorously understand a developmental process. The final panorama is an escalation of data complexity. As complexity increases, the need to apply computational approaches grows proportionally. There are numerous ways to apply computational methods to Developmental Biology (Figure 1), each adapted to the question guiding the study and the type of data available. The same is true for diseases that take place during the developmental process.

For example, Network Systems Biology (NSB) is one of the most interdisciplinary subareas of Bioinformatics. NSB became an invaluable tool for understanding biological systems, for the formulation of original hypotheses, for comparison between different organisms, and for interpreting a massive amount of data (Vidal et al., 2011). This possibility arises because

network topological proprieties and organization principles, such as clustering, percolation, and centralities, can be applied to numerous biological problems, regardless of the background study (Albert, 2005). However, despite this flexibility, how they can or should be utilized differs from data to data (e.g., gene expression, genomics, proteomics, epigenomics, etc.). Since the significant limitation of NSB is data availability, the continuous effort to create new databases, specially curated ones, directly impacts NSB studies' success. New network topological parameters are also necessary for a more accurate selection of potential targets.

In silico structural analysis at the protein level is also a common integrative approach, combined with data from the wet lab or as a stand-alone analysis. Analyzing a single protein structure can be considered an entire universe of complexity because numerous investigations need to be applied to infer a molecular behavior or a possible structural feature. In this sense, there is a demand for more computational methods to analyze proteins essential to the developmental process. Due to the numerous limitations of experimental techniques at the wet lab, many proteins, such as

HOX proteins, have no complete structural data. Therefore, the precision of protein structural modeling tools, especially using *ab initio* calculations, significantly increases the success of protein structure research (Delarue and Koehl, 2018).

Some widespread computational approaches, such as gene expression analyses and phylogenetic studies, are also essential in developmental biology research. Although the applicability of these studies is straightforward, they are not without challenges. Statistical analyzes are crucial in both cases, and the creation of new tools implementing robust statistical treatments for multi-level and meta-analyzes are decisive in generating accurate data. In particular, phylogenetic studies need new tools to deal with massive amounts of sequence.

Finally, machine learning (ML) approaches are gaining territory in almost all biological and Biomedical Sciences. Such tools' vast applicability can be adapted to virtually all Big-Data problems. Either being applied to expression (Ang et al., 2016), genomic (Libbrecht and Noble, 2015), epigenetics (Holder et al., 2017) or biomedical image data (Kan, 2017), new ML algorithms and elegant application

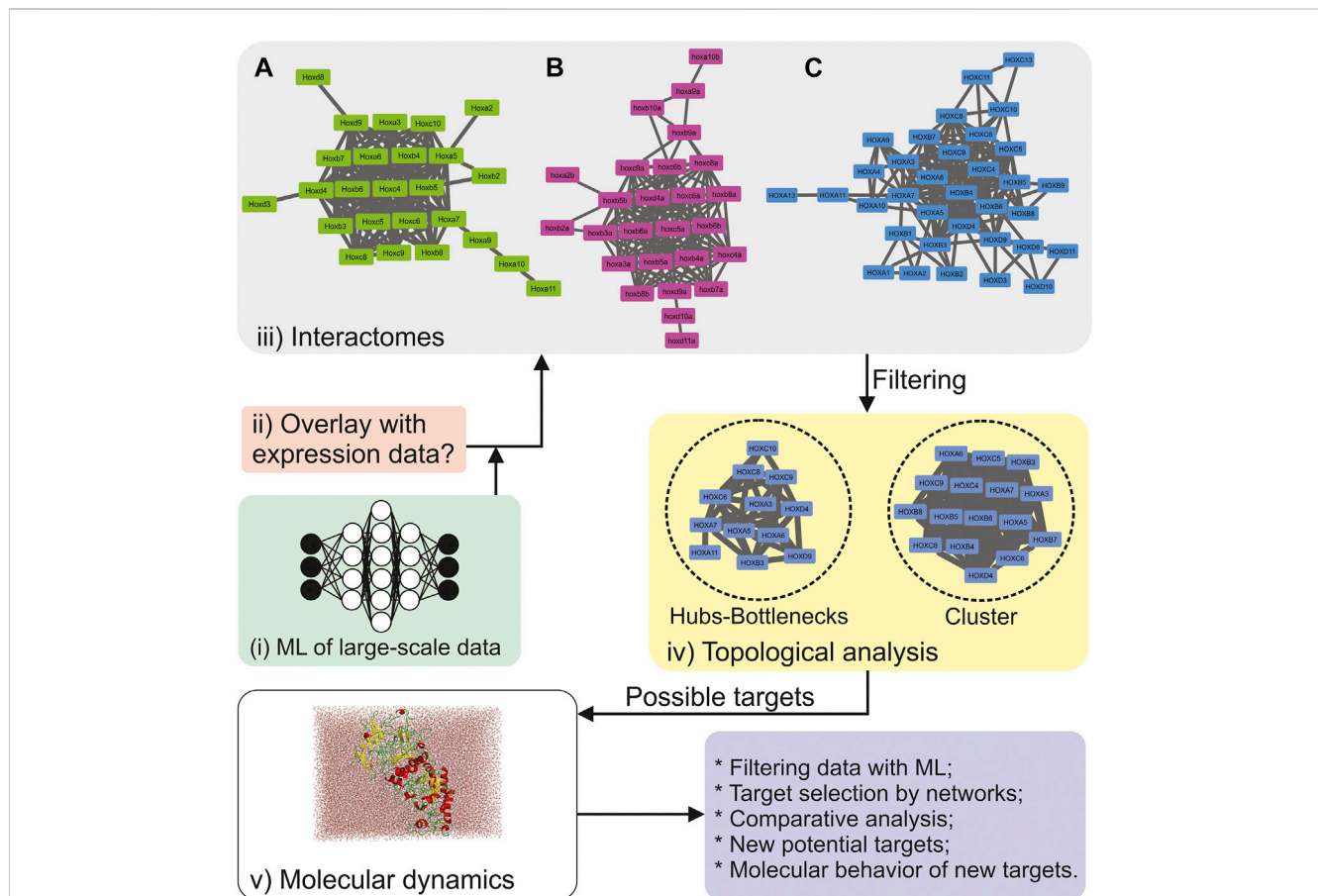


FIGURE 1

Example of an integrative computational approach to explore biological data. The concept is to integrate different computational approaches into one single pipeline. (A) Protein-Protein Network (PPN) for *Mus musculus*. From the 39 Hox proteins, Hoxa1, Hoxa4, Hoxa13, Hoxb9, Hoxb13, Hoxc11, Hoxc12, Hoxc13, Hoxd1, Hoxd10, Hoxd11, Hoxd12, and Hoxd13 were not connected to the network; (B) PPN for the 36 hox genes in *Danio rerio*. Hoxa1, hoxa11, hoxa13, hoxb1, hoxc10, hoxd3, hoxl2, and hoxd13 were not connected to the network, whereas hoxa9 and hoxc3 were not found by STRING; (C) PPN for *Homo sapiens*. From the 39 HOX proteins, HOXB13, HOXC12, HOXD1, HOXD12, and HOXD13 were not connected to the network. The primary networks were built in the STRING v.11 (Szklarczyk et al., 2019) meta-search engine, using the following parameters (10 Oct 2020): all search options enabled, except for *textmining* and *gene-fusion*, the confidence score of 0.4, and no more than 20 interactions in the 1st shell. The final networks were edited in Cytoscape 3.6.1 (Shannon, 2003). Clusters were obtained using MCODE (Bader and Hogue, 2003), and centralities were calculated by CentiScaPe 2.2 (Scardoni et al., 2009). Protein image was created using Pymol (DeLano, 2002), system was build using GROMACS 2018.1 (Abraham et al., 2015).

protocols are necessary for this day and age of science. In addition to developing new algorithms, the major challenge of this field is how they are tested. Input data is perhaps the major player in this equation since training the algorithm with up-to-date curated quality data would, undoubtedly, generate more accurate results than training them with older datasets (Feltes et al., 2019; Feltes et al., 2021).

This Research Topic presents some integrative approaches applied to developmental biology experiments from different backgrounds.

2 The research topic

Fagny et al. employed systems biology to integrate genomic, transcriptomic, and epigenomic data to elucidate the regulatory relationships between transcription factors, enhancers, and potential target genes in maize and identify regulatory key factors specific to leaves at the seedling stage and husks at flowering. By combining different omic data, they reconstructed tissue-specific-associated transcriptomic factors regulatory networks and uncovered genes crucial to tissue-specific differentiation.

Zhou et al. combined two RNA-seq datasets of human placentas from term and preterm birth to investigate co-altered circadian transcripts-associated long intergenic non-coding RNAs (lincRNAs). Using a diverse set of transcriptomic analyses, they uncovered nine core molecular clock genes deregulated by the decrease of five circadian lincRNAs in the placenta, affecting a myriad of crucial biological processes that could be linked to preterm birth.

Shang et al. analyzed cardiac development-associated transcriptomic data from three Single-Cell Tagged Reverse Transcription (STRT-Seq) datasets. Using a transcriptomic analysis workflow, they explored lineage-specific changes in mouse and humans regarding gene expression, subpopulation composition, and developmental features in cardiac tissues. They described an evolutionary conservation of cell populations and molecular profiles during heart development in both species.

Tang et al. reviewed experimental technologies, public data, and predictive models associated with synthetic lethal pairs. The knowledge of synthetic lethal pairs is deeply interconnected to developmental biology since it revolves around how the impairment of two genes can lead to cellular or organism death, which does not happen when one of these genes is still viable. The authors outlined important perspectives and critical discussions on the subject that can aid future research.

Finally, Wu et al. provided a comprehensive review of the mathematical model of local ancestry inference (LAI) applied to genomic data. In this sense, its application, historical aspects, different benchmarks, and the strengths and limitations of LAI are discussed and outlined.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

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

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1 (2), 19–25. doi:10.1016/j.softx.2015.06.001
- Albert, R. (2005). Scale-free networks in cell biology. *J. Cell Sci.* 118, 4947–4957. doi:10.1242/jcs.02714
- Ang, J. C., Mirzal, A., Haron, H., and Hamed, H. N. A. (2016). Supervised, unsupervised, and semi-supervised feature selection: A review on gene selection. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 13, 971–989. doi:10.1109/TCBB.2015.2478454
- Bader, G. D., and Hogue, C. W. V. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinforma.* 4, 2–27. doi:10.1186/1471-2105-4-2
- DeLano, W. L. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* 40, 82–92.
- Delarue, M., and Koehl, P. (2018). Combined approaches from physics, statistics, and computer science for *ab initio* protein structure prediction: Ex univitate VirES (unity is strength)? *F1000Research* 7, 1125. doi:10.12688/f1000research.14870.1
- Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teach.* 35, 125–129. doi:10.2307/4444260
- Feltes, B. C., Chandelier, E. B., Grisci, B. I., and Dorn, M. (2019). CuMiDa: An extensively curated microarray database for benchmarking and testing of machine learning approaches in cancer research. *J. Comput. Biol.* 26, 376–386. doi:10.1089/cmb.2018.0238
- Feltes, B. C., Poloni, J. D. F., and Dorn, M. (2021). Benchmarking and testing machine learning approaches with BARRA:CuRDa, a curated RNA-seq database for cancer research. *J. Comput. Biol.* 28, 931–944. doi:10.1089/cmb.2020.0463
- Holder, L. B., Haque, M. M., and Skinner, M. K. (2017). Machine learning for epigenetics and future medical applications. *Epigenetics* 12, 505–514. doi:10.1080/15592294.2017.1329068
- Kan, A. (2017). Machine learning applications in cell image analysis. *Immunol. Cell Biol.* 95, 525–530. doi:10.1038/icb.2017.16
- Libbrecht, M. W., and Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* 16, 321–332. doi:10.1038/nrg3920
- Scardoni, G., Petterlini, M., and Laudanna, C. (2009). Analyzing biological network parameters with CentiScaPe. *Bioinformatics* 25, 2857–2859. doi:10.1093/bioinformatics/btp517
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi:10.1093/nar/gky1131
- Vidal, M., Cusick, M. E., and Barabási, A. L. (2011). Interactome networks and human disease. *Cell* 144, 986–998. doi:10.1016/j.cell.2011.02.016