OPEN ACCESS Freely available online

ISSN: 2153-0602
Journal of Data Mining in Genomics & Proteomics

Commentary

Deep Learning and the New Intelligence of Biological Data

Junmei Sekeres*

Department of Computational Biology, Tsinghua University, Beijing, China

DESCRIPTION

Deep learning has emerged as a powerful computational approach for interpreting the enormous and complex datasets generated by modern omics technologies. Genomics, transcriptomics, proteomics, metabolomics and epigenomic each describe different layers of biological activity. When these datasets are combined, they can contain millions of measurements and intricate relationships that are difficult to analyze with conventional techniques. Deep learning provides flexible models capable of recognizing complex patterns, learning useful representations and generating predictions from high-dimensional biological information. Omics datasets are particularly suitable for deep learning because they contain many interacting variables. A genomic sequence may contain subtle patterns associated with gene regulation, while transcriptomic measurements can reveal coordinated changes across thousands of genes. Protein abundance, molecular modifications and metabolite concentrations add further layers of information. Deep neural networks can process these variables simultaneously and learn relationships that may not be apparent when each molecular feature is considered separately.

One important application is genomic sequence analysis. Deep learning models can examine Deoxy Ribonucleic Acid (DNA) sequences to identify patterns associated with promoters, enhancers, transcription factor binding sites and other regulatory elements. Convolutional neural networks can recognize local sequence motifs, while more advanced architectures can capture relationships between distant genomic regions. These approaches can also support the prediction of potential effects caused by genetic variants by examining how sequence changes alter learned molecular patterns. Transcriptomic analysis provides another major application. Gene expression datasets commonly contain thousands of genes measured across numerous biological samples or individual cells. Deep learning can compress these high-dimensional measurements into informative representations that preserve important biological variation. Auto encoders, for example, can learn compact representations of expression profiles, whileneural classification models can distinguish molecular patterns associated with different cellular states or biological conditions.

Single-cell omics has further increased the importance of deep learning. A single-cell dataset can contain expression measurements from thousands or millions of individual cells, each with substantial biological and technical variation. Deep learning models can help identify cellular populations, characterize transitional states, remove certain technical effects and reconstruct relationships among cells. These capabilities can contribute to more detailed representations of tissue organization and cellular diversity. Proteomic and metabolomic datasets also benefit from deep learning. Protein measurements can contain information about abundance, structure, interactions and molecular modifications, while metabolomic profiles describe changes in biochemical compounds. Neural networks can integrate these features to recognize molecular signatures and predict relationships among proteins, metabolites and biological pathways. Such models may be particularly useful when molecular relationships are nonlinear and cannot be adequately represented by simple statistical assumptions.

Multi-omics integration is one of the most promising applications of deep learning. Each omics layer provides a different perspective on biological systems, but these layers are interconnected. A genomic alteration may influence transcription, which can change protein abundance and subsequently affect metabolic activity. Deep learning architectures can combine different data types through shared representations or specialized processing branches. This can create integrated models that capture relationships across several molecular levels. Attention-based architectures have introduced additional possibilities for omics analysis. These models can identify relationships among distant features and assign greater importance to information that contributes strongly to a prediction.

Correspondence to: Junmei Sekeres, Department of Computational Biology, Tsinghua University, Beijing, China, Email: junmei.sekeres@tsinghua.edu.cn

Received: 30-May-2026, Manuscript No. JDMGP-26-32481; **Editor assigned:** 02-Jun-2026, PreQC No. JDMGP-26-324821; **Reviewed:** 16-Jun-2026, QC No. JDMGP-26-324821; **Revised:** 22-Jun-2026, Manuscript No. JDMGP-26-32481; **Published:** 30-Jun-2026, DOI: 10.35248/ 2153-0602.26.17.420

Citation: Sekeres J (2026). Deep Learning and the New Intelligence of Biological Data. J Data Mining Genomics Proteomics. 17:420.

Copyright: © 2026 Sekeres J. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

OPEN ACCESS Freely available online

CONCLUSION

Neural networks can also over fit when models contain many parameters but training data are insufficient. Careful pre-processing, regularization, suitable validation strategies and independent datasets are therefore essential for developing reliable models. Interpretability is another major consideration. A highly accurate model may still provide limited insight into the biological mechanisms underlying its predictions. Explainable artificial intelligence approaches can identify influential genes, genomic regions, proteins or metabolites and help connect computational outputs with biological processes.