

Review

DOI: <https://doi.org/10.17816/ecogen637074>

EDN: JFNADD



A New Era of Bioinformatics

Anna Yu. Aksenova¹, Anna S. Zhuk^{1,2,3}, Elena I. Stepchenkova^{1,3},
Viacheslav A. Semenikhin⁴, Mikhail A. Langovoy⁵

¹ Saint Petersburg State University, Saint Petersburg, Russia;

² ITMO University, Saint Petersburg, Russia;

³ Vavilov Institute of General Genetics Russian Academy of Science, Saint Petersburg branch, Saint Petersburg, Russia;

⁴ Matheomics, Skolkovo Innovation Center, Moscow, Russia;

⁵ Center for Artificial Intelligence SPbU, Saint Petersburg, Russia

ABSTRACT

Bioinformatics is a rapidly growing discipline at the interface of biology, computer science, and mathematics. Recent scientific and technological advances in biological and biomedical sciences have led to a rapid increase in data generation. The analysis and interpretation of such data requires powerful computational tools and specialists with deep expertise in various fields, including molecular biology, genetics, programming, and mathematics. Currently, machine learning and deep learning methods are being rapidly integrated into various fields of biology and medicine, significantly transforming bioinformatic solutions and marking the advent of a new era in bioinformatics. The development of new algorithms and efficient data analysis methods using artificial intelligence forms the foundation for the future growth of this field. In this context, the demand for specialists capable of bridging the gap between biological and mathematical disciplines continues to grow, necessitating the adaptation of educational programs. This article reviews recent trends in bioinformatics, including the development of multi-omics approaches and the use of artificial intelligence, and highlights the importance of multidisciplinary education with advanced training in mathematics and statistics to prepare a new generation of scientists capable of driving innovation in this dynamic field.

Keywords: bioinformatics; machine learning; deep learning; omics technologies.

To cite this article

Aksenova AY, Zhuk AS, Stepchenkova EI, Semenikhin VA, Langovoy MA. A New Era of Bioinformatics. *Ecological genetics*. 2025;23(2):211–219. DOI: 10.17816/ecogen637074 EDN: JFNADD

Submitted: 15.10.2024

Accepted: 05.03.2025

Published online: 30.06.2025

Научный обзор

DOI: <https://doi.org/10.17816/ecogen637074>

EDN: JFNADD

Новая эра биоинформатики

А.Ю. Аксенова¹, А.С. Жук^{1,2,3}, Е.И. Степченкова^{1,3}, В.А. Семенихин⁴, М.А. Ланговой⁵¹ Санкт-Петербургский государственный университет, Санкт-Петербург, Россия;² Университет ИТМО, Санкт-Петербург, Россия;³ Институт общей генетики им. Н.И. Вавилова Российской академии наук, Санкт-Петербургский филиал, Санкт-Петербург, Россия;⁴ Матеомика, Инновационный центр Сколково, Москва, Россия⁵ Центр искусственного интеллекта СПбГУ, Санкт-Петербург, Россия

АННОТАЦИЯ

Биоинформатика — это быстро развивающаяся дисциплина на стыке биологии, информатики и математики. Научно-технический прогресс в области биологических и биомедицинских наук за последние годы привел к стремительному росту объемов данных. Для анализа и интерпретации больших данных нужны мощные вычислительные инструменты и специалисты с глубокими знаниями в различных областях, включая молекулярную биологию, генетику, программирование и математику. В настоящее время происходит стремительная интеграция методов машинного и глубокого машинного обучения в различные области биологии и медицины, что в существенной степени меняет формат биоинформатических решений и позволяет говорить о наступлении новой эры в биоинформатике. Разработка новых алгоритмов и способов эффективного анализа данных с использованием искусственного интеллекта является основой для будущего развития этой области. В этой связи спрос на специалистов, способных преодолеть разрыв между биологическими и математическими дисциплинами, продолжает расти, что требует соответствующей адаптации учебных программ. В статье рассматриваются последние тенденции в биоинформатике, такие как развитие мультиомиксных подходов и использование искусственного интеллекта, а также подчеркивается важность многопрофильного образования с углубленным обучением в области математики и статистики для подготовки нового поколения ученых, способных стимулировать инновации в этой динамичной области науки.

Ключевые слова: биоинформатика; машинное обучение; глубокое машинное обучение; омиксные технологии.

Как цитировать

Аксенова А.Ю., Жук А.С., Степченкова Е.И., Семенихин В.А., Ланговой М.А. Новая эра биоинформатики // Экологическая генетика. 2025. Т. 23. № 2. С. 211–219. DOI: 10.17816/ecogen637074 EDN: JFNADD

SUBJECT AND TASKS OF BIOINFORMATICS, AND ITS ROLE IN MEDICINE AND FUNDAMENTAL AND APPLIED BIOLOGY

Bioinformatics is an interdisciplinary area that combines biological sciences, mathematics, statistics, and computer technology to collect, store, analyze, and interpret biological and biomedical data. This rapidly evolving field focuses on the development and implementation of algorithms and computational tools for biological data analysis, particularly in genomics, transcriptomics, proteomics, and structural and systems biology. The modern era in biology is distinguished by the rapid accumulation of vast amounts of data generated by advanced techniques such as next-generation sequencing (NGS), third-generation sequencing (TGS), structural biology, and mass spectrometry. The data generated by these techniques are often too large and complex to be managed using conventional approaches. At the same time, these massive databases may be crucial for understanding how life is organized at the molecular level. They are essential for understanding the complex biological processes that govern the structure and functioning of living systems, ranging from gene expression regulation and protein interactions to the organization of complex intracellular structures and intercellular interactions. Advances in computational techniques are critical for interpreting such complex systems and exploring the fundamental laws that control life. Research in personalized medicine, drug development, systems biology, and agricultural sciences cannot advance without the active development and implementation of bioinformatics approaches for large-scale data analysis and interpretation [1–8].

The development of bioinformatics has been largely driven by the advancement of next-generation sequencing, which is becoming increasingly accessible and integrated into routine clinical practice. Technological advances, combined with lower sequencing costs and expanded applications in a variety of fields, have significantly increased the use of these methods. In the near future, NGS and TGS technologies are expected to play a key role in shaping healthcare and become the standard for biomedical research.

In clinical diagnosis and personalized medicine, there has been substantial global increase in research based on whole genome or targeted DNA sequencing (including exomes and individual gene panels) and transcriptome sequencing. These techniques are becoming increasingly accessible to a wide range of researchers and clinical laboratories [9, 10]. NGS makes it possible to get unparalleled insights into genetic variations in human populations and investigate the mechanisms underlying hereditary diseases and cancer [11, 12]. Genome-wide association studies (GWAS) using NGS or microarray hybridization

data enable identifying correlations between genetic variants and traits or diseases [13, 14]. These advancements make it possible to identify specific molecular markers of various diseases and consider their cumulative impact, allowing for personalized treatment based on patients' individual characteristics [15]. Moreover, NGS facilitates the development of non-invasive diagnostic techniques, such as liquid biopsy, which allows monitoring disease progression and response to treatment, and non-invasive prenatal testing (NIPT) [16–18]. In addition, NGS plays a fundamental role in the development of personalized cancer immunotherapy (cancer vaccines) by identifying neoantigens expressed in tumors [19–21]. This approach not only maximizes therapeutic impact on cancer cells, but also minimizes potential side effects of broader immunotherapies that can affect healthy cells. Moreover, the development of techniques such as Chromium (10x Genomics), C1 (Fluidigm), and Seek One (Seek Gene Biotechnology) has enabled simultaneous acquisition of sequencing data for thousands of single cells (scDNA-seq and scRNA-seq) [22]. Therefore, this approach refines genetic testing to the level of single-cell analysis for assessing heterogeneity in cell populations and identifying unique events in individual cells.

Other techniques, such as ChIP-seq, ATAC-seq, and Methyl-seq and their combinations with other omics technologies, allow studying the regulation of gene expression, chromatin dynamics, and various epigenetic mechanisms [23–27]. Specialized NGS methods are being integrated into clinical practice as evidence accumulates and the clinical significance of specific mechanisms in the development of diseases is confirmed [15, 28, 29]. The molecular mechanisms of disease development are becoming much better understood at the genetic and epigenetic levels as genome sequencing and data analysis techniques advance, paving the way for a new era of precision medicine and human life extension. The new assembly T2T-CHM13, a continuous sequence of the human genome without gaps, including previously unstudied regions such as centromeres and telomeres [30, 31], has accelerated the development of genome analysis methods, including studies on functional role of repetitive sequences and the search for various structural variants. Owing to the intensive development of this field, the number of omics data collected worldwide has increased dramatically. In particular, large genomic data analysis centers generate tens to hundreds of terabytes of new data per day. By 2025, the volume of genetic data accumulated globally is expected to exceed that of information technology giants such as *YouTube* and *X (Twitter)* [32–34].

Bioinformatics is critical for understanding the structural and functional properties of proteins and peptides. Advances in mass spectrometry and other proteomic technologies enable generating complex arrays of data

on protein interactions and modifications and investigating their structure [35–37]. The interpretation of these data aids in the investigation of various protein complexes and the understanding of the intricate networks of interactions between proteins, as well as proteins and nucleic acids within cells [38–42]. This information is critical for developing new drugs, studying disease mechanisms, and identifying biomarkers [43–48]. On a higher level, systems biology combines multiple layers of biological data (genomics, transcriptomics, proteomics, and metabolomics) to create complex models of biological systems. Bioinformatics tools are critical for modeling these complex systems and predicting their behavior under various scenarios.

As data accumulated, extensive multipurpose biological databases were created. These include NCBI services offering various databases and data analysis tools (<https://www.ncbi.nlm.nih.gov/>), UCSC genome browser, which allows visualizing genomes and contains various analysis tools (<https://genome.ucsc.edu/>) [49], Ensembl [50], EMBL-EBI (<https://www.ebi.ac.uk/>) [51], UniProt [52], Protein Data Bank [53], KEGG [54], and Enzyme Database (BRENDA) [55]. These and many other bioinformatics resources are used to annotate genomes, investigate gene function and regulation, track protein functions, metabolic pathways, and genetic interactions, and uncover new patterns by comparing biological information from various sources.

Recently, there has been a significant increase in the use of machine learning (ML) to identify patterns in complex NGS data for addressing various pharmacogenomics and oncogenetics issues [56]. Artificial intelligence (AI) technologies, a powerful tool for improving the accuracy and speed of data interpretation, are at the cutting edge of science. The integration of machine learning and AI methods facilitates the extraction of useful information, revolutionizing the analysis of omics data. It enables identifying new genetic variants relevant to disease progression, predicting disease risk, and discovering new biomarkers, which facilitates the development of personalized medicine and targeted therapeutic approaches. For example, AI-based algorithms are increasingly used in large-scale searches for new drug targets and diagnostic tools. As a result, numerous pharmaceutical companies are expanding their use of AI approaches in biomedical data processing and analysis.

ROLE OF FUNDAMENTAL MATHEMATICAL KNOWLEDGE IN THE TRAINING OF HIGHLY QUALIFIED BIOINFORMATICS SCIENTISTS

As bioinformatics becomes more important, so does the demand for qualified experts who can bridge the gap between biology and data science. The pharmaceutical

and biotechnology industries are in need of professionals who can interpret genomic, transcriptomic, and proteomic data and investigate biomolecule structures for drug development and precision medicine. Bioinformaticians are also required by research institutes and laboratories to implement scientific projects and manage increasingly complex data sets. The majority of biological research today relies on bioinformatics tools to identify biomarkers, analyze next-generation sequencing findings, and model disease mechanisms.

The interdisciplinary nature of bioinformatics presents unique educational challenges, highlighting the need for specialized training programs. In addition to the fundamentals of molecular biology, bioinformatician training requires a strong mathematics background. This is due to the fact that biological data are analyzed using a variety of statistical and computational methods. The competence and correctness with which specific mathematical methods and software tools are used determines both the reliability of the analysis and the usefulness of information extracted from complex biological data. A thorough understanding of mathematics, especially statistics, probability, linear algebra, combinatorics, and graph theory, is necessary for modeling biological systems, managing large data sets, and developing predictive models.

Moreover, the complexity of biological data requires expertise in machine learning, data visualization, and programming. Programming for research purposes is typically done in languages such as Python, R, and SQL, whereas larger projects (including commercial bioinformatics software) may require C/C++, Java, or even specialized programming languages. Bioinformatics training also typically requires familiarity with cloud computing platforms for data storage and processing. All of these factors must be considered in bioinformatics education for combining theoretical and practical training in data analysis, learning advanced algorithms for interpreting omics data, and developing software. This is due to the fact that developing and optimizing new algorithms for biological data processing and analysis is one of the most important tasks of bioinformatics.

Rapid technological advancements in bioinformatics necessitate continuous education and training. This means that experts must be perpetual learners to keep up with new tools, methods, and techniques, such as deep learning approaches for analyzing omics data or advances in quantum computing. Bioinformatics education programs must evolve in tandem with these trends to ensure that graduates have up-to-date skills. The incorporation of AI and machine learning into bioinformatics is gaining traction. AI technologies have proven particularly useful in analyzing large and complex data sets in genomic analysis, single-cell transcriptome sequencing, spatial transcriptomics, and multi-omics technologies

[57–61]. Deep machine learning is currently indispensable in drug development, proteomics analysis, and protein structure studies [62, 63]. DeepMind's AlphaFold2 (and a new version, AlphaFold3, available to research institutes since November 2024) has revolutionized protein structure prediction owing to its amazing accuracy in determining 3D structures from amino acid sequences. AlphaFold2 and AlphaFold3 predict protein structure using deep learning methods, specifically neural networks trained on large data sets of known protein structures [64–68]. Demis Hassabis and John Jumper were awarded the Nobel Prize in Chemistry in 2024 for developing the AlphaFold2 algorithms, highlighting the innovative nature and significance of this technique. David Baker, who made significant contributions to the computer design of protein molecules, shared the award with them [69, 70].

Neural networks have also shown promise in designing genome editing experiments using CRISPR/Cas9 [71–73]. As these techniques gain popularity in medicine, agriculture, and ecology, bioinformaticians will play an increasingly important role in ensuring their safe and effective use.

The increasing volume and complexity of data sets inevitably raises the demand for machine learning and deep learning experts. This trend is evident in personalized medicine, where machine learning models can assist in treatment selection based on a patient's unique genomic profile [74–79]. Multi-omics approaches that combine genomics, transcriptomics, proteomics, and metabolomics data will surely necessitate the development of new computational tools and bioinformatics methods for interpreting these complex data sets and discovering meaningful biological relationships. Bioinformatics tools are becoming more powerful and mathematically complicated, therefore training new experts requires a thorough understanding of mathematical concepts and statistical methods. Bioinformatics training programs in universities must place a greater emphasis on mathematics and statistics to provide students with the foundational skills required to traverse the field's increasing complexity.

Importantly, the two most prominent machine learning approaches of the last 30 years, deep neural networks and support vector machines (SVM), were originally proposed and developed in the 1960s by Soviet experts in applied mathematics and mathematical statistics. Here, we will mention only the fundamental works on the first learning neural networks [80], the first deep neural networks [81], and pattern recognition [82]. Classical, strictly justified mathematical problem-solving methods remain relevant for a long time. For example, the training of modern large neural networks and large language models is based on the optimal control theory proposed in the 1950s by Pontryagin et al. [83], and the

backpropagation method proposed by Galushkin [84]. Geoffrey Hinton earned the 2024 Nobel Prize in Physics for his work in training deep neural networks by applying and enhancing these methods.

CONCLUSION

In conclusion, bioinformatics is a rewarding professional path with a variety of opportunities in academic and applied research, biotechnology industry, healthcare, and entrepreneurship. The future of bioinformatics is incredibly promising, and as data accumulates and new problems emerge, bioinformaticians will continue to play a significant role in scientific progress. It is critical that the education system adapts to the changing need for highly qualified specialists in this field. The new era of bioinformatics, with the increasing use of machine and deep learning in various fields of biology and medicine, must be followed by a new era of education. We encourage non-mathematical students who want to specialize in bioinformatics to systematically improve their skills and proficiency in classical mathematical tools, particularly linear algebra, discrete mathematics, probability theory, and statistics. Universities should emphasize the mathematical component of bioinformatics by engaging relevant experts in teaching.

ADDITIONAL INFO

Author contributions: A.Yu. Aksenova, concept and design of the manuscript, collection and processing of literary data, writing the text, final editing; A.S. Zhuk — analysis of literary data, writing the text, final editing; E.I. Stepchenkova — participation in the development of the manuscript concept, final editing; V.A. Semenikhin — analysis of literary data, writing the text, intermediate editing; M.A. Langovoy — collection and processing of literary data, writing the text, final editing. The authors approved the version for publication and agreed to take responsibility for all aspects of the work, ensuring proper consideration and resolution of issues related to the accuracy and integrity of any part of it.

Acknowledgments: The authors would like to thank the staff of the RC MCT and the RC "Biobank" of St. Petersburg State University. The authors are grateful to Kirill V. Volkov for his critical comments on the manuscript.

Funding sources: This work was supported by Saint Petersburg State University, project No. 125021902561–6.

Disclosure of interests: The authors declare that there are no relationships, activities, or interests in the past three years related to third parties (commercial and non-commercial) whose interests may be affected by the content of this article.

Statement of originality: In creating this work, the authors did not use previously published information.

Generative AI: The text of this article is not the result of generative artificial intelligence.

Provenance and peer-review: This work was submitted to the journal on its own initiative and reviewed according to the standard procedure. Two external reviewers, and a member of the editorial board participated in the review.

ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

Вклад авторов. А.Ю. Аксенова — концепция и дизайн манускрипта, сбор и обработка литературных данных, написание текста, внесение окончательной правки; А.С. Жук — анализ литературных данных, написание текста, внесение окончательной правки; Е.И. Степченкова — участие в разработке концепции рукописи, внесение окончательной правки; В.А. Семенихин — анализ литературных данных, написание текста, внесение промежуточных правок; М.А. Ланговой — сбор и обработка литературных данных, написание текста, внесение окончательной правки. Авторы одобрили версию для публикации, а также согласились нести ответственность за все аспекты работы, гарантируя надлежащее рассмотрение и решение вопросов, связанных с точностью и добросовестностью любой ее части.

Благодарности. Авторы благодарят сотрудников РЦ РМИКТ и РЦ «Биобанк» СПбГУ. Авторы благодарны Кириллу Владимировичу Волкову за критические замечания, высказанные относительно данной рукописи.

Источники финансирования. Работа выполнена при поддержке ФГБОУ ВО «Санкт-Петербургский государственный университет» (проект № 125021902561-6).

Раскрытие интересов. Авторы заявляют об отсутствии отношений, деятельности и интересов за последние три года, связанных с третьими лицами (коммерческими и некоммерческими), интересы которых могут быть затронуты содержанием статьи.

Оригинальность. При создании настоящей работы авторы не использовали ранее опубликованные сведения.

Генеративный искусственный интеллект. Текст настоящей статьи не является результатом работы генеративного искусственного интеллекта.

Рассмотрение и рецензирование. Настоящая работа подана в журнал в инициативном порядке и рассмотрена по обычной процедуре. В рецензировании участвовали два внешних рецензента и член редакционной коллегии.

REFERENCES

1. Alser M, Lindegger J, Firtina C, et al. From molecules to genomic variations: Accelerating genome analysis via intelligent algorithms and architectures. *Comput Struct Biotechnol J*. 2022;20:4579–4599. doi: 10.1016/j.csbj.2022.08.019
2. Tan YC, Kumar AU, Wong YP, Ling APK. Bioinformatics approaches and applications in plant biotechnology. *J Genet Eng Biotechnol*. 2022;20(1):1–13. doi: 10.1186/S43141-022-00394-5/TABLES/2
3. Naqvi RZ, Mahmood MA, Mansoor S, et al. Omics-driven exploration and mining of key functional genes for the improvement of food and fiber crops. *Front Plant Sci*. 2023;14:1273859. doi: 10.3389/FPLS.2023.1273859/PDF
4. Srivastava R. Applications of artificial intelligence multiomics in precision oncology. *J Cancer Res Clin Oncol*. 2023;149:503–510. doi: 10.1007/S00432-022-04161-4/METRICS
5. Pezoulas VC, Hazapis O, Lagopati N, et al. Machine learning approaches on high throughput ngs data to unveil mechanisms of function in biology and disease. *Cancer Genom Proteom*. 2021;18(5):605–626. doi: 10.21873/CGP.20284
6. Sadee W, Wang D, Hartmann K, Toland AE. Pharmacogenomics: Driving personalized medicine. *Pharmacol Rev*. 2023;75(4):789–814. doi: 10.1124/PHARMREV.122.000810
7. Uesaka K, Oka H, Kato R, et al. Bioinformatics in bioscience and bioengineering: recent advances, applications, and perspectives. *J Biosci Bioeng*. 2022;134(5):363–373. doi: 10.1016/J.JBIOSEC.2022.08.004
8. Jamialahmadi H, Khalili-Tanha G, Nazari E, Rezaei-Tavirani M. Artificial intelligence and bioinformatics: A journey from traditional techniques to smart approaches. *Gastroenterol Hepatol Bed Bench*. 2024;17(3):241–252. doi: 10.22037/GHFB.V17I3.2977
9. Riess O, Sturm M, Menden B, et al. Genomes in clinical Care. *NPJ Genomic Med*. 2024;9:20. doi: 10.1038/s41525-024-00402-2
10. Mosele F, Remon J, Mateo J, et al. Recommendations for the use of next-generation sequencing (NGS) for patients with metastatic cancers: A report from the ESMO Precision Medicine Working Group. *Ann Oncol*. 2020;31(11):1491–1505. doi: 10.1016/j.annonc.2020.07.014
11. Morganti S, Tarantino P, Ferraro E, et al. Next generation sequencing (NGS): A revolutionary technology in pharmacogenomics and personalized medicine in cancer. In: Ruiz-Garcia E, Astudillo-de la Vega H, editors. *Translational research and onco-omics applications in the era of cancer personal genomics. Advances in experimental medicine and biology*. Vol. 1168. Springer, Cham; 2019. P. 9–30. doi: 10.1007/978-3-030-24100-1_2
12. Edsjö A, Gisselsson D, Staaf J, et al. Current and emerging sequencing-based tools for precision cancer medicine. *Mol Aspects Med*. 2024;96:101250. doi: 10.1016/J.MAM.2024.101250
13. Abdellaoui A, Yengo L, Verweij KJH, Visscher PM. 15 years of GWAS discovery: Realizing the promise. *Am J Hum Genet*. 2023;110(2):179–194. doi: 10.1016/j.ajhg.2022.12.011
14. Defo J, Awany D, Ramesar R. From SNP to pathway-based GWAS meta-analysis: Do current meta-analysis approaches resolve power and replication in genetic association studies? *Brief Bioinform*. 2023;24(1):bbac600. doi: 10.1093/bib/bbac600
15. Yadav D, Patil-Takbhathe B, Khandagale A, et al. Next-generation sequencing transforming clinical practice and precision medicine. *Clin Chim Acta*. 2023;551:117568. doi: 10.1016/J.CCA.2023.117568
16. Roberto TM, Jorge MA, Francisco GV, et al. Strategies for improving detection of circulating tumor DNA using next generation sequencing. *Cancer Treat Rev*. 2023;119:102595. doi: 10.1016/J.CTRV.2023.102595
17. Shegekar T, Vodithala S, Juganavar A. The emerging role of liquid biopsies in revolutionising cancer diagnosis and therapy. *Cureus*. 2023;15(8):e43650. doi: 10.7759/CUREUS.43650
18. Jenkins M, Seasely AR, Subramaniam A. Prenatal genetic testing 2: Diagnostic tests. *Curr Opin Pediatr*. 2022;34(6):553–558. doi: 10.1097/MOP.0000000000001174
19. Schäfer RA, Guo Q, Yang R. ScanNeo2: A comprehensive workflow for neoantigen detection and immunogenicity prediction from diverse genomic and transcriptomic alterations. *Bioinformatics*. 2023;39(11):btad659. doi: 10.1093/bioinformatics/btad659
20. Xie N, Shen G, Gao W, et al. Neoantigens: Promising targets for cancer therapy. *Signal Transduct Target Ther*. 2023;8:9. doi: 10.1038/s41392-022-01270-x

21. Kiyotani K, Chan HT, Nakamura Y. Immunopharmacogenomics towards personalized cancer immunotherapy targeting neoantigens. *Cancer Sci*. 2018;109(3):542–549. doi: 10.1111/CAS.13498
22. See P, Lum J, Chen J, Ginhoux F. A single-cell sequencing guide for immunologists. *Front Immunol*. 2018;9:415498. doi: 10.3389/FIMMU.2018.02425/BIBTEX
23. Choi H, Kim H, Chung H, et al. Application of computational algorithms for single-cell RNA-Seq and ATAC-Seq in neurodegenerative diseases. *Brief Funct Genom*. 2025;24: elae44. doi: 10.1093/BFGP/ELAE044
24. Lee J-W, Cho J-Y. Comparative epigenetics of domestic animals: Focusing on DNA accessibility and its impact on gene regulation and traits. *J Vet Sci*. 2025;26(1):24259. doi: 10.4142/JVS.24259
25. Cox OH, Seifuddin F, Guo J, et al. Implementation of the Methyl-Seq platform to identify tissue- and sex-specific DNA methylation differences in the rat epigenome. *Epigenetics*. 2024;19:2393945. doi: 10.1080/15592294.2024.2393945
26. Li S-J, Gao X, Wang Z-H, et al. Cell-free DNA methylation patterns in aging and their association with inflamm-aging. *Epigenomics*. 2024;16(10):715–731. doi: 10.1080/17501911.2024.2340958
27. Hubert J-N, Iannuccelli N, Cabau C, et al. Detection of DNA methylation signatures through the lens of genomic imprinting. *Sci Rep*. 2024;14:1694. doi: 10.1038/s41598-024-52114-3
28. Lee H, Martinez-Agosto JA, Rexach J, Fogel BL. Next generation sequencing in clinical diagnosis. *Lancet Neurol*. 2019;18(5):426. doi: 10.1016/S1474-4422(19)30110-3
29. Gibbs SN, Peneva D, Cuyun Carter G, et al. Comprehensive review on the clinical impact of next-generation sequencing tests for the management of advanced cancer. *JCO Precis Oncol*. 2023;7:715. doi: 10.1200/PO.22.00715
30. Nurk S, Koren S, Rhie A, et al. The complete sequence of a human genome. *Science*. 2022;376(6588):44–53. doi: 10.1126/SCIENCE.ABJ6987
31. Hoyt SJ, Storer JM, Hartley GA, et al. From telomere to telomere: the transcriptional and epigenetic state of human repeat elements. *Science*. 2022;376(6588):eabk3112. doi: 10.1126/science.abk3112
32. Stephens ZD, Lee SY, Faghri F, et al. Big Data: Astronomical or genomic? *PLOS Biol*. 2015;13:e1002195. doi: 10.1371/JOURNAL.PBIO.1002195
33. Katz K, Shutov O, Lapoint R, et al. The sequence read archive: A decade more of explosive growth. *Nucleic Acids Res*. 2022;50(D1):D387–D390. doi: 10.1093/NAR/GKAB1053
34. Danielewski M, Szalata M, Nowak JK, et al. History of biological databases, their importance, and existence in modern scientific and policy context. *Genes*. 2025;16(1):100. doi: 10.3390/GENES16010100/S1
35. Fedorov II, Protasov SA, Tarasova IA, Gorshkov MV. Ultrafast proteomics. *Biochem*. 2024;89:1349–1361. doi: 10.1134/S0006297924080017/FIGURES/4
36. Anderton CR, Uhrig RG. The promising role of proteomes and metabolomes in defining the single-cell landscapes of plants. *New Phytol*. 2025;245(3):945–948. doi: 10.1111/NPH.20303
37. Godoy Sanches PH, Clemente De Melo N, Porcari AM, Miguel De Carvalho L. Integrating molecular perspectives: strategies for comprehensive multi-omics integrative data analysis and machine learning applications in transcriptomics, proteomics, and metabolomics. *Biology*. 2024;13(11):848. doi: 10.3390/BIOLOGY13110848
38. Wu S, Zhang S, Liu CM, et al. Recent advances in mass spectrometry-based protein interactome studies. *Mol Cell Proteom*. 2025;24(1):100887. doi: 10.1016/j.mcpro.2024.100887
39. Dang V, Voigt B, Marcotte EM. Progress toward a comprehensive brain protein interactome. *Biochem Soc Trans*. 2025;53(1):303–314. doi: 10.1042/BST20241135
40. Rahmati S, Emili A. Proximity labeling: precise proteomics technology for mapping receptor protein neighborhoods at the cancer cell surface. *Cancers*. 2025;17(2):179. doi: 10.3390/cancers17020179
41. Edwards AN, Hsu KL. Emerging opportunities for intact and native protein analysis using chemical proteomics. *Anal Chim Acta*. 2025;1338:343551. doi: 10.1016/J.ACA.2024.343551
42. Goel RK, Bithi N, Emili A. Trends in co-fractionation mass spectrometry: a new gold-standard in global protein interaction network discovery. *Curr Opin Struct Biol*. 2024;88:102880. doi: 10.1016/J.SBI.2024.102880
43. Kim SG, Hwang JS, George NP, et al. Integrative metabolome and proteome analysis of cerebrospinal fluid in Parkinson's disease. *Int J Mol Sci*. 2024;25(21):11406. doi: 10.3390/IJMS252111406/S1
44. Wu D, Zhang L, Ding F. Current status and future directions of application of urine proteomics in neonatology. *Front Pediatr*. 2024;12:1509468. doi: 10.3389/FPED.2024.1509468/BIBTEX
45. Kliuchnikova AA, Ilgisonis EV, Archakov AI, et al. Proteomic markers of aging and longevity: A systematic review. *Int J Mol Sci*. 2024;25(23):12634. doi: 10.3390/IJMS252312634/S1
46. Nalla LV, Kanukolanu A, Yeduvaka M, Gajula SNR. Advancements in single-cell proteomics and mass spectrometry-based techniques for unmasking cellular diversity in triple negative breast cancer. *Proteomics — Clin Appl*. 2025;19(1):e202400101. doi: 10.1002/PRCA.202400101
47. Pomella S, Melaiu O, Cifaldi L, et al. Biomarkers identification in the microenvironment of oral squamous cell carcinoma: A systematic review of proteomic studies. *Int J Mol Sci*. 2024;25(16):8929. doi: 10.3390/IJMS25168929/S1
48. Zhang Z, Huang J, Zhang Z, et al. Application of omics in the diagnosis, prognosis, and treatment of acute myeloid leukemia. *Biomark Res*. 2024;12:60. doi: 10.1186/s40364-024-00600-1
49. ar do Perez G, Barber GP, Benet-Pages A, et al. The UCSC genome browser database: 2025 update. *Nucleic Acids Res*. 2025;53(D1):D1243–D1249. doi: 10.1093/NAR/GKAE974
50. Dyer SC, Austine-Orimoloye O, Azov AG, et al. Ensembl 2025. *Nucleic Acids Res*. 2025;53(D1):D948–D957. doi: 10.1093/NAR/GKAE1071
51. Rodriguez-Tomé P, Stoehr PJ, Cameron GN, Flores TP. The European Bioinformatics Institute (EBI) databases. *Nucleic Acids Res*. 1996;24(1):6–12. doi: 10.1093/NAR/24.1.6
52. Consortium TU, Bateman A, Martin M-J, et al. UniProt: The universal protein knowledgebase in 2025. *Nucleic Acids Res*. 2025;53(D1):D609–D617. doi: 10.1093/NAR/GKAE1010
53. Zardecki C, Dutta S, Goodsell DS, et al. PDB-101: Educational resources supporting molecular explorations through biology and medicine. *Protein Sci*. 2022;31(1S):129–140. doi: 10.1002/PRO.4200
54. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30. doi: 10.1093/NAR/28.1.27
55. Chang A, Jeske L, Ulbrich S, et al. BRENDA, the ELIXIR core data resource in 2021: New developments and updates. *Nucleic Acids Res*. 2021;49(D1):D498–D508. doi: 10.1093/NAR/GKAA1025
56. Mondello A, Dal Bo M, Toffoli G, Polano M. Machine learning in onco-pharmacogenomics: a path to precision medicine with many challenges. *Front Pharmacol*. 2024;14:1260276. doi: 10.3389/fphar.2023.1260276
57. Erfanian N, Heydari AA, Feriz AM, et al. Deep learning applications in single-cell genomics and transcriptomics data analysis. *Biomed Pharmacother*. 2023;165:115077. doi: 10.1016/J.BIOPHA.2023.115077

58. Athaya T, Ripan RC, Li X, Hu H. Multimodal deep learning approaches for single-cell multi-omics data integration. *Brief Bioinform.* 2023;24(5):bbad313. doi: 10.1093/BIB/BBAD313
59. Gulati GS, D'Silva JP, Liu Y, et al. Profiling cell identity and tissue architecture with single-cell and spatial transcriptomics. *Nat Rev Mol Cell Biol.* 2024;26:11–31. doi: 10.1038/s41580-024-00768-2
60. Rivero-Garcia I, Torres M, Sánchez-Cabo F. Deep generative models in single-cell omics. *Comput Biol Med.* 2024;176:108561. doi: 10.1016/J.COMPBIOMED.2024.108561
61. Kang M, Ko E, Mersha TB. A roadmap for multi-omics data integration using deep learning. *Brief Bioinform.* 2022;23(1):bbab454. doi: 10.1093/BIB/BBAB454
62. Pun FW, Ozerov IV, Zhavoronkov A. AI-powered therapeutic target discovery. *Trends Pharmacol Sci.* 2023;44(9):561–572. doi: 10.1016/j.tips.2023.06.010
63. Mann M, Kumar C, Zeng W-F, Strauss MT. Artificial intelligence for proteomics and biomarker discovery. *Cell Syst.* 2021;12(8):759–770. doi: 10.1016/j.cels.2021.06.006
64. Wang L, Wen Z, Liu S-W, et al. Overview of AlphaFold2 and breakthroughs in overcoming its limitations. *Comput Biol Med.* 2024;176:108620. doi: 10.1016/j.compbiomed.2024.108620
65. Zhang H, Lan J, Wang H, et al. AlphaFold2 in biomedical research: facilitating the development of diagnostic strategies for disease. *Front Mol Biosci.* 2024;11:1414916. doi: 10.3389/FMOLB.2024.1414916
66. Varga JK, Schueler-Furman O. Who binds better? Let AlphaFold2 decide! *Angew Chemie. Int Ed.* 2023;62(28):e202303526. doi: 10.1002/anie.202303526
67. Bertoline LMF, Lima AN, Krieger JE, Teixeira SK. Before and after AlphaFold2: An overview of protein structure prediction. *Front Bioinform.* 2023;3:1120370. doi: 10.3389/FBINF.2023.1120370
68. Borkakoti N, Thornton JM. AlphaFold2 protein structure prediction: Implications for drug discovery. *Curr Opin Struct Biol.* 2023;78:102526. doi: 10.1016/J.SBI.2022.102526
69. Leman JK, Weitzner BD, Lewis SM, et al. Macromolecular modeling and design in rosetta: Recent methods and frameworks. *Nat Methods.* 2020;17:665–680. doi: 10.1038/S41592-020-0848-2
70. Baek M, DiMaio F, Anishchenko I, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science.* 2021;373(6557):871–876. doi: 10.1126/science.abj8754
71. Zhang G, Luo Y, Dai X, Dai Z. Benchmarking deep learning methods for predicting CRISPR/Cas9 SgRNA on- and off-target activities. *Brief Bioinform.* 2023;24(6):bbad333. doi: 10.1093/BIB/BBAD333
72. Sherkatghanad Z, Abdar M, Charlier J, Makarenkov V. Using traditional machine learning and deep learning methods for on- and off-target prediction in CRISPR/Cas9: A review. *Brief Bioinform.* 2023;24(3):bbad131. doi: 10.1093/BIB/BBAD131
73. Lee M. Deep learning in CRISPR-cas systems: A review of recent studies. *Front Bioeng Biotechnol.* 2023;11:1226182. doi: 10.3389/fbioe.2023.1226182
74. Sun D, Chen W, He J, et al. A novel method for screening malignant hematological diseases by constructing an optimal machine learning model based on blood cell parameters. *BMC Med Inform Decis Mak.* 2025;25:72. doi: 10.1186/s12911-025-02892-1
75. Shan R, Li X, Chen J, et al. Interpretable machine learning to predict the malignancy risk of follicular thyroid neoplasms in extremely unbalanced data: retrospective cohort study and literature review. *JMIR cancer.* 2025;11:e66269–e66269. doi: 10.2196/66269
76. Ayhan B, Ayan E, Atsü S. Detection of dental caries under fixed dental prostheses by analyzing digital panoramic radiographs with artificial intelligence algorithms based on deep learning methods. *BMC Oral Health.* 2025;25:216. doi: 10.1186/s12903-025-05577-3
77. Kovács KA, Kerepesi C, Rapcsák D, et al. Machine learning prediction of breast cancer local recurrence localization, and distant metastasis after local recurrences. *Sci Rep.* 2025;15:4868. doi: 10.1038/s41598-025-89339-9
78. Guo L, Wang W, Xie X, et al. Machine learning-based models for genomic predicting neoadjuvant chemotherapeutic sensitivity in cervical cancer. *Biomed Pharmacother.* 2023;159:114256. doi: 10.1016/J.BIOPHA.2023.114256
79. Zhao Y, Fu Z, Barnett EJ, et al. Genome data based deep learning identified new genes predicting pharmacological treatment response of attention deficit hyperactivity disorder. *Transl Psychiatry.* 2025;15:46. doi: 10.1038/s41398-025-03250-5
80. Ivakhnenko AG, Lapa VG. *Cybernetic predictive devices.* Kyiv: Naukova Dumka; 1965. 214 p. URL: <https://gwern.net/doc/ai/1966-ivakhnenko.pdf>
81. Ivakhnenko AG. Polynomial theory of complex systems. In: *IEEE Trans. Syst. Man Cybern.* 1971. Vol. 1. P. 364–378. doi: 10.1109/TSMC.1971.4308320
82. Vapnik VN, Chervonenkis AJ. On one class of learning algorithms for pattern recognition. *Automation and Remote Control.* 1964;25:937–945. (In Russ.)
83. Boltzansky VG, Gamkrelidze RV, Pontryagin LS. To the theory of optimal processes. *Reports of the USSR Academy of Sciences.* 1956;110:7–10. (In Russ.)
84. Galushkin AI. *Synthesis of multilayer systems of pattern recognition.* Moscow: Energia; 1974. (In Russ.)

AUTHORS' INFO

***Anna Yu. Aksenova**, Cand. Sci. (Biology);
address: 7/9 Universitetskaya emb., Saint Petersburg, 199034, Russia;
ORCID: 0000-0002-1601-1615; eLibrary SPIN: 4914-7675;
e-mail: a.aksenova@spbu.ru

Anna S. Zhuk, Cand. Sci. (Biology), Assistant Professor;
ORCID: 0000-0001-8683-9533; eLibrary SPIN: 2223-5306;
e-mail: ania.zhuk@gmail.com

Elena I. Stepchenkova, Cand. Sci. (Biology);
ORCID: 0000-0002-5854-8701; eLibrary SPIN: 9121-7483;
e-mail: stepchenkova@gmail.com

ОБ АВТОРАХ

***Аксенова Анна Юрьевна**, канд. биол. наук;
адрес: Россия, 199034, Санкт-Петербург, Университетская наб., д. 7/9;
ORCID: 0000-0002-1601-1615; eLibrary SPIN: 4914-7675;
e-mail: a.aksenova@spbu.ru

Жук Анна Сергеевна, канд. биол. наук, доцент;
ORCID: 0000-0001-8683-9533; eLibrary SPIN: 2223-5306;
e-mail: ania.zhuk@gmail.com

Степchenkova Елена Игоревна, канд. биол. наук;
ORCID: 0000-0002-5854-8701; eLibrary SPIN: 9121-7483;
e-mail: stepchenkova@gmail.com

* Corresponding author / Автор, ответственный за переписку

Viacheslav A. Semenikhin; ORCID: 0000-0001-6923-0363;
eLibrary SPIN: 2251-5652; e-mail: vasemenikhin@hse.ru

Mikhail A. Langovoy, Dr. rer. nat.;
ORCID: 0000-0002-7593-0830; eLibrary SPIN: 6905-9451;
e-mail: mikhail@langovoy.com

Семенихин Вячеслав Алексеевич; ORCID: 0000-0001-6923-0363;
eLibrary SPIN: 2251-5652; e-mail: vasemenikhin@hse.ru

Ланговой Михаил Анатольевич, Dr. rer. nat.;
ORCID: 0000-0002-7593-0830; eLibrary SPIN: 6905-9451;
e-mail: mikhail@langovoy.com