

Научный обзор

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

EDN: JFNADD

## Новая эра биоинформатики

А.Ю. Аксенова<sup>1</sup>, А.С. Жук<sup>1,2,3</sup>, Е.И. Степченкова<sup>1,3</sup>, В.А. Семенихин<sup>4</sup>, М.А. Ланговой<sup>5</sup><sup>1</sup> Санкт-Петербургский государственный университет, Санкт-Петербург, Россия;<sup>2</sup> Университет ИТМО, Санкт-Петербург, Россия;<sup>3</sup> Институт общей генетики им. Н.И. Вавилова Российской академии наук, Санкт-Петербургский филиал, Санкт-Петербург, Россия;<sup>4</sup> Матеомика, Инновационный центр Сколково, Москва, Россия<sup>5</sup> 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](https://doi.org/10.17816/ecogen637074) EDN: JFNADD

Review

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

EDN: JFNADD

# A New Era of Bioinformatics

Anna Yu. Aksenova<sup>1</sup>, Anna S. Zhuk<sup>1,2,3</sup>, Elena I. Stepchenkova<sup>1,3</sup>,  
Viacheslav A. Semenikhin<sup>4</sup>, Mikhail A. Langovoy<sup>5</sup>

<sup>1</sup> Saint Petersburg State University, Saint Petersburg, Russia;

<sup>2</sup> ITMO University, Saint Petersburg, Russia;

<sup>3</sup> Vavilov Institute of General Genetics Russian Academy of Science, Saint Petersburg branch, Saint Petersburg, Russia;

<sup>4</sup> Matheomics, Skolkovo Innovation Center, Moscow, Russia;

<sup>5</sup> Center for Artificial Intelligence SPbU, Saint Petersburg, Russia

## ABSTRACT

Potato 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

## ПРЕДМЕТ И ЗАДАЧИ БИОИНФОРМАТИКИ, ЕЕ ЗНАЧЕНИЕ ДЛЯ МЕДИЦИНЫ, ФУНДАМЕНТАЛЬНОЙ И ПРИКЛАДНОЙ БИОЛОГИИ

Биоинформатика — это междисциплинарная область, объединяющая биологические науки, математику, статистику и компьютерные технологии для сбора, хранения, анализа и интерпретации биологических и биомедицинских данных. Эта интенсивно развивающаяся область включает в себя разработку и применение алгоритмов и вычислительных инструментов для анализа биологической информации, особенно в таких областях, как геномика, транскриптомика, протеомика, а также структурная и системная биология. Современная эра в биологии характеризуется лавинообразным накоплением огромного объема данных, генерируемых с помощью таких передовых методов, как секвенирование следующего поколения (NGS, Next Generation Sequencing) и секвенирование третьего поколения (TGS, Third Generation Sequencing) и методов структурной биологии и масс-спектрометрии. Данные, получаемые с помощью этих методов, зачастую слишком объемны и сложны, чтобы ими можно было оперировать с использованием традиционных подходов. Вместе с тем в огромных массивах таких данных кроется тайна организации жизни на молекулярном уровне: они дают ключ к пониманию сложных биологических процессов, определяющих организацию и функционирование живых систем, от регуляции экспрессии генов и взаимодействия белков, до организации сложных внутриклеточных структур и межклеточных взаимодействий. Развитие вычислительных методов играет ключевую роль в расшифровке таких сложных систем, а также в исследовании фундаментальных принципов, управляющих жизнью. Продвижение исследований в области персонализированной медицины, разработки лекарств, системной биологии и сельскохозяйственных наук в настоящее время невозможно без активного развития и внедрения биоинформатических методов для анализа и интерпретации больших объемов данных [1–8].

Развитие биоинформатики в значительной степени было обусловлено развитием технологий высокопроизводительного секвенирования, которые становятся все более доступными и интегрируются в рутинную клиническую практику. Технологический прогресс в совокупности со снижением стоимости секвенирования и расширением сфер его применения в различных областях, привел к бурному росту использования этих методов. Ожидается, что в ближайшем будущем технологии NGS и TGS начнут играть ключевую роль в формировании здравоохранения и станут стандартом для биомедицинских исследований.

В клинической диагностике и персонализированной медицине по всему миру наблюдается активный рост исследований, базирующихся на секвенировании полных геномов или целевых участков ДНК (включая экзоны

и панели отдельных генов), а также секвенировании транскриптомов, причем эти методы становятся все более доступными для широкого круга исследователей и клинических лабораторий [9, 10]. Возможности NGS позволяют получить беспрецедентное представление о генетических вариациях в популяциях людей, исследовать механизмы наследственных заболеваний и механизмы развития рака [11, 12]. Полногеномные исследования ассоциаций (GWAS) с использованием данных NGS или данных, полученных с помощью гибридизации на микрочипах, позволяют выявлять корреляции между генетическими вариантами и признаками или заболеваниями [13, 14]. Все это в совокупности позволяет выявлять специфические молекулярные маркеры различных заболеваний и учитывать их совокупность, что дает возможность проводить лечение с учетом индивидуальных особенностей пациентов [15]. Кроме того, NGS способствует развитию неинвазивных диагностических подходов, например жидкостная биопсия, которая позволяет отслеживать прогрессирование заболевания и ответ на лечение, и неинвазивное пренатальное тестирование (NIPT) [16–18]. Помимо этого, NGS играет принципиальную роль в разработке персонализированной иммунотерапии онкологических заболеваний (вакцины от рака), базирующейся на выявлении неоантигенов, экспрессируемых в опухолях [19–21]. Такой подход не только обеспечивает максимальное терапевтическое воздействие на раковые клетки, но и минимизирует потенциальные побочные эффекты, связанные с более широкими иммунотерапевтическими методами, которые могут затрагивать здоровые клетки. В дополнение к этому, развитие таких технологий, как Chromium (10x Genomics), C1 (Fluidigm) и Seek One (Seek Gene Biotechnology), позволило параллельно получать данные секвенирования для тысяч одиночных клеток (scDNA-seq и scRNA-seq) [22]. Таким образом, этот подход повышает разрешение генетических тестов до уровня анализа отдельных клеток, что позволяет исследовать гетерогенность клеточных популяций и выявлять уникальные события, возникающие в отдельных клетках.

Другие методы, например ChIP-seq, ATAC-seq и Methylation-seq и их различные комбинации с другими омиксными технологиями, позволяют изучать регуляцию экспрессии генов, динамику хроматина и различные эпигенетические механизмы [23–27]. Специализированные методы NGS внедряются в клиническую практику по мере накопления данных и подтверждения клинической значимости тех или иных механизмов в развитии заболеваний [15, 28, 29]. Молекулярные механизмы развития заболеваний становятся намного понятней на генетическом и эпигенетическом уровне благодаря расширяющимся возможностям методов секвенирования генома и анализа данных, что открывает путь к новой эре точной медицины и продлению жизни человека. Новая сборка T2T-CHM13, представляющая собой непрерывную последовательность генома человека, без пробелов, включая ранее неизученные

регионы, такие как центромеры и теломеры [30, 31], дала новый виток развитию методов анализа генома, включающих изучение функциональной роли повторяющихся последовательностей и поиска различных структурных вариантов. Благодаря интенсивному развитию этой области во всем мире наблюдается взрывной рост объемов собираемых омиксных данных. В частности, крупные центры анализа геномных данных генерируют десятки и сотни терабайт новых данных в сутки. Ожидается, что уже в 2025 г. объем накопленных в мире геномных данных превысит масштабы таких гигантов индустрии информационных технологий как Youtube и X(Twitter) [32–34].

Биоинформатика играет важнейшую роль в изучении структурно-функциональных свойств белков и пептидов. Достижения в области масс-спектрометрии и других протеомных методов позволяют генерировать сложные массивы данных о взаимодействиях и модификациях белков, изучать их структуру [35–37]. Интерпретация этих данных помогает исследовать различные белковые комплексы и понять сложные сети взаимодействий между белками, а также белками и нуклеиновыми кислотами внутри клеток [38–42]. Эта информация является ключевой для разработки новых лекарств, изучения механизмов заболеваний и выявления биомаркеров [43–48]. Системная биология, работающая на более высоком уровне, объединяет различные слои биологических данных (геномика, транскриптомика, протеомика, метаболомика) для создания комплексных моделей биологических систем. Инструменты биоинформатики незаменимы для моделирования этих сложных систем и предсказания их поведения в различных условиях.

По мере накопления данных были созданы обширные и многофункциональные биологические базы данных, такие как сервисы NCBI, который содержит различные базы и инструменты анализа данных (<https://www.ncbi.nlm.nih.gov/>), ресурс UCSC genome browser, позволяющий визуализировать геномы, и содержащий различные инструменты анализа (<https://genome.ucsc.edu/>) [49], Ensembl [50], EMBL-EBI (<https://www.ebi.ac.uk/>) [51], UniProt [52], Protein Data Bank [53], KEGG [54], Enzyme Database (BRENDA) [55]. Эти и многие другие ресурсы биоинформатики используют для аннотирования геномов, изучения функций генов и их регуляции, отслеживания функций белков, метаболических путей и генетических взаимодействий, а также позволяют проводить перекрестное сопоставление биологической информации из различных источников и выявлять новые закономерности.

В последнее время наблюдается активный рост использования методов машинного обучения (ML) для обнаружения закономерностей в сложных данных NGS для решения различных проблем фармакогеномики и онкогенетики [56]. На переднем крае науки находятся технологии искусственного интеллекта (ИИ) — мощного инструмента для повышения точности и скорости интерпретации данных. Интеграция методов машинного обучения и ИИ

облегчает извлечение значимой информации и, таким образом, революционизирует анализ омиксных данных, позволяя выявлять новые генетические варианты, значимые для течения болезни, прогнозировать риск развития заболеваний и обнаруживать новые биомаркеры, что способствует развитию персонализированной медицины и ускоряет разработку подходов для целевой терапии. Например, алгоритмы на основе ИИ в настоящее время незаменимы для масштабного поиска новых лекарственных мишеней и диагностических инструментов. Соответственно, многие фармацевтические компании переходят на повсеместное использование методов ИИ в обработке и анализе биомедицинских данных.

## ЗНАЧЕНИЕ ФУНДАМЕНТАЛЬНЫХ МАТЕМАТИЧЕСКИХ ЗНАНИЙ ДЛЯ ПОДГОТОВКИ ВЫСОКОКВАЛИФИЦИРОВАННЫХ БИОИНФОРМАТИКОВ

По мере роста значимости биоинформатики растет и спрос на квалифицированных специалистов, способных преодолеть разрыв между биологией и наукой о данных. Фармацевтическая и биотехнологическая отрасли нуждаются в специалистах, способных интерпретировать геномные, транскриптомные и протеомные данные, а также изучать структуры биомолекул для разработки лекарств и точной медицины. Академическим институтам и исследовательским лабораториям также требуются биоинформатики для реализации научных проектов и управления все более сложными массивами данных. Большинство биологических исследований сегодня опираются на биоинформатические инструменты для выявления биомаркеров, анализа данных высокопроизводительного секвенирования и моделирования механизмов заболеваний.

Междисциплинарный характер биоинформатики создает уникальные образовательные проблемы, что подчеркивает необходимость развития специализированных программ обучения. В частности, в подготовке биоинформатики решающее значение играет не только знание основ молекулярной биологии, но и хорошее математическое образование. Это обусловлено тем, что для анализа биологических данных используются различные статистические и вычислительные методы. Достоверность анализа, равно как и количество значимой информации, извлекаемой из сложных биологических данных, зависит от умения и правильности применения тех или иных математических алгоритмов и программных инструментов. Глубокое понимание математики, особенно в таких областях, как статистика, теория вероятностей, линейная алгебра, комбинаторика и теория графов, необходимо для моделирования биологических систем, управления большими массивами данных и разработки прогностических моделей.

Сложность биологических данных также требует знаний в области машинного обучения, визуализации данных и программирования. Программирование решений научно-исследовательских задач обычно осуществляется на таких языках, как Python, R и SQL, в то время как для более масштабных проектов (в том числе в области коммерческого биоинформатического программного обеспечения) могут требоваться, к примеру, C/C++, Java или даже специализированные языки программирования. Биоинформатическая подготовка также стандартно включает в себя знакомство с облачными вычислительными платформами для хранения и обработки данных. Образование в области биоинформатики должно учитывать все это, сочетая теоретическую подготовку с практическими занятиями по анализу данных, изучению передовых алгоритмов для интерпретации омиксных данных и разработке программного обеспечения. Это обусловлено тем, что одной из важнейших задач биоинформатики является разработка новых алгоритмических решений для обработки и анализа биологических данных, а также оптимизация этих процессов.

Быстрый темп технологических изменений в биоинформатике требует постоянного обучения и подготовки. Это означает, что специалисты в этой области должны учиться всю жизнь, чтобы успевать овладевать новыми инструментами, методами и методиками, такими как подходы глубокого обучения для анализа омиксных данных или достижения в области квантовых вычислений. Образовательные программы в области биоинформатики должны развиваться вместе с этими тенденциями, чтобы выпускники обладали самыми современными навыками. Интеграция ИИ и машинного обучения в биоинформатику ускоряется. Технологии ИИ особенно успешно используются для анализа больших и сложных массивов данных в таких приложениях, как геномный анализ, секвенирование транскриптомов одиночных клеток, пространственная транскриптомика и мультиомиксные технологии [57–61]. Глубокое машинное обучение в настоящее время незаменимо для поиска новых лекарств, протеомного анализа и изучения структуры белков [62, 63]. AlphaFold2 (и новая версия AlphaFold3, ставшая доступной для академических институтов в ноябре 2024 г.), разработанный компанией DeepMind, произвел революцию в предсказании структуры белков, достигнув поразительной точности в определении 3D-структур по аминокислотным последовательностям. Для предсказания структуры белков AlphaFold2 и AlphaFold3 используют методы глубокого обучения, в частности нейронные сети, обучаясь на больших массивах данных об известных белковых структурах [64–68]. Признание инновационности и значимости такого подхода подчеркивается тем, что за разработку алгоритмов AlphaFold2 Демису Хассабису и Джону Джамперу присуждена Нобелевская премия по химии в 2024 г. Дэвид Бейкер, внесший значительный вклад в компьютерный дизайн белковых молекул, разделил с ними награду [69, 70].

Нейронные сети также показали определенную эффективность в разработке экспериментов по редактированию генома с использованием CRISPR/Cas9 [71–73]. По мере того, как подобные инструменты будут все шире использоваться в медицине, сельском хозяйстве и экологии, биоинформатики будут играть все более значимую роль в обеспечении безопасного и эффективного применения этих технологий.

Увеличение объема и усложнение массивов данных неизбежно повышает спрос на специалистов, владеющих методами машинного обучения и глубокого обучения. Эта тенденция четко прослеживается в области персонализированной медицины, где модели машинного обучения помогут подобрать лечение на основе уникального геномного профиля пациента [74–79]. Несомненно, мультиомиксные подходы, плотно интегрирующие данные геномики, транскриптомики, протеомики и метаболомики, потребуют разработки новых вычислительных инструментов и методов биоинформатики для интерпретации этих сложных наборов данных и поиска значимых биологических связей. Инструментарий биоинформатики становится все более продвинутым и математически сложным, и глубокое понимание математических концепций и статистических методов начинает играть первостепенную роль в подготовке новых кадров. Программы биоинформатической подготовки в вузах должны уделять все больше внимания математике и статистике, чтобы вооружить студентов основополагающими навыками, необходимыми для работы в условиях растущей сложности этой области.

Нельзя не упомянуть также и тот факт, что два наиболее популярных метода машинного обучения последних 30 лет, а именно глубокие нейросети и SVM (Support Vector Machines, метод опорных векторов), были изначально предложены и разработаны в 1960-е годы советскими специалистами по прикладной математике и математической статистике. Упомянем здесь только основополагающие работы по первым обучающимся нейросетям [80], первым глубоким нейросетям [81] и распознаванию образов [82]. Классические, строго обоснованные, математические методы решения задач долго сохраняют свою актуальность. К примеру, в основе обучения современных больших нейросетей и больших языковых моделей лежат принципы теории оптимального управления, разработанные в 1950-е годы группой Л.С. Понтрягина [83], и метод обратного распространения ошибки А.И. Галушкина [84]. Применяя и развивая эти методы, Джеффри Хинтон получил свои результаты в области обучения глубоких нейросетей, за которые был удостоен Нобелевской премии по физике 2024 г.

## ЗАКЛЮЧЕНИЕ

В заключение следует отметить, что образование в области биоинформатики открывает путь к полноценной карьере с разнообразными возможностями в области академических и прикладных исследований,

биотехнологической промышленности, здравоохранения и предпринимательства. Будущее биоинформатики вероятно перспективно, и по мере накопления данных и появления новых задач биоинформатики будут продолжать вносить ключевой вклад в научный прогресс. Исключительно важно, чтобы образовательная система адекватно реагировала на изменяющиеся потребности в высококвалифицированных специалистах в этой области. Вслед за новой эрой биоинформатики, характеризующейся ростом применения методов машинного и глубокого машинного обучения в различных областях биологии и медицины, новая эра должна наступить и в образовательной сфере. Мы рекомендуем студентам нематематических специальностей, планирующим специализацию в области биоинформатики, систематически повышать свою квалификацию и уровень владения классическими математическими инструментами, особенно в таких дисциплинах, как линейная алгебра, дискретная математика, теория вероятностей и статистика. Университетам следует акцентировать внимание на математической составляющей учебных программ по биоинформатике, привлекая соответствующих экспертов к процессу преподавания.

## 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 достоверностью 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).

## СПИСОК ЛИТЕРАТУРЫ | 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

**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.

## 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. Semенихин — 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.

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, Seasey 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.** Boltvansky 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.)

## 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

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

**Семенikhин 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

## 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

**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

\* Автор, ответственный за переписку / Corresponding author