

На правах рукописи



Иванов Роман Артемович

**ФИЛОСТРАТИГРАФИЧЕСКИЙ И
ФИЛОТРАНСКРИПТОМНЫЙ АНАЛИЗ ГЕНОВ,
СВЯЗАННЫХ С РАЗВИТИЕМ ОНКОЛОГИЧЕСКИХ
ЗАБОЛЕВАНИЙ**

1.5.8. Математическая биология, биоинформатика

АВТОРЕФЕРАТ
диссертации на соискание учёной степени
кандидата биологических наук

Новосибирск – 2025

Работа выполнена в секторе биоинформатики и информационных технологий в генетике Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», г. Новосибирск.

Научный руководитель:

Лашин Сергей Александрович

д.б.н., зав. сектора биоинформатики и информационных технологий в генетике ФГБНУ «ФИЦ Институт цитологии и генетики СО РАН», г. Новосибирск

Официальные оппоненты:

Щербаков Дмитрий Юрьевич, доктор биологических наук, доцент, заведующий лабораторией геносистематики, ФГБУН "Лимнологический институт" СО РАН, г. Иркутск.

Козлов Константин Николаевич, кандидат биологических наук, доцент, ФГАОУ ВО Санкт-Петербургский политехнический университет Петра Великого, г Санкт-Петербург

Ведущее учреждение:

Федеральное бюджетное учреждение науки «Научно-исследовательский институт системной биологии и медицины» Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека, г. Москва

Защита диссертации состоится «____» _____ 202__г. на утреннем заседании диссертационного совета 24.1.239.01 на базе Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», в конференц-зале Института по адресу:
пр. академика Лаврентьева 10, г. Новосибирск, 630090,
тел.: (383) 363-49-06 (1321); факс: (383) 333-12-78;
e-mail: dissov@bionet.nsc.ru

С диссертацией можно ознакомиться в библиотеке ИЦиГ СО РАН и на сайте института www.icgbio.ru.

Автореферат разослан «____» _____ 202__ г.

Учёный секретарь
диссертационного совета,
доктор биологических наук

Хлебодарова Т. М.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования

Злокачественные заболевания остаются одной из основных причин смертности в мире, что обуславливает непрекращающую актуальность фундаментальных исследований механизмов онкогенеза. Наряду с интенсивным изучением молекулярных и клинических аспектов злокачественности, в последние годы существенно возрос интерес к изучению роли эволюции генов в возникновении и прогрессировании злокачественных опухолей [Greaves, Maley, 2012; Merlo и др., 2006; Pepper и др., 2009; Dujon et al., 2021; Gillies, Verduzco, Gatenby, 2012; Maley et al., 2017; Mina et al., 2017]. Анализ происхождения генов и метаболических путей, ассоциированных с онкопатологиями, открывает новые возможности для объяснения причин приобретения генами онкогенных свойств и особенностей поведения опухолевых клеток [Davies, Lineweaver, 2011]. В связи с этим эволюционный подход, основанный на принципах молекулярной филогенетики и сравнительной геномики, становится важным для выявления фундаментальных механизмов, лежащих в основе приобретения злокачественными клетками их отличительных черт.

Одним из распространенных подходов к анализу молекулярной эволюции генов является филостратиграфический анализ, который сопоставляет каждый ген с его самым ранним идентифицируемым предком на таксономическом древе. Это позволяет отнести ген к определённой филострате, описываемой *индексом филостратиграфического возраста* (*Phylostratigraphy Age Index – PAI*), в которой он предположительно возник [Domazet-Lošo, Brajković, Tautz, 2007a; Domazet-Lošo, Tautz, 2008a]. Этот подход показал, что гены, связанные с фундаментальными клеточными процессами, зачастую восходят к ранним эволюционным эпохам, отражая консервативность ключевых биологических функций.

На основе филостратиграфического анализа в работах Domazet-Lošo и соавторов [Domazet-Lošo, Brajković, Tautz, 2007a], а также Drost, Grosse и Quint [Quint и др., 2012b] были разработаны специальные метрики, которые позволяют количественно охарактеризовать вклад генов разного «возраста» в транскриптоме. Первым таким показателем является *индекс возраста транскриптома* (*Transcriptome Age Index – TAI*), который отражает средний эволюционный возраст экспрессируемых в определённых условиях генов. Второй показатель, *индекс дивергенции транскриптома* (*Transcriptome Divergence Index – TDI*), описывает

среднюю степень дивергенции транскриптома, указывая на то, насколько увеличивается активность более консервативных генов или, наоборот, дивергентных [Quint и др., 2012b].

Филостратиграфический анализ был применен и в изучении патологических процессов, в том числе онкологических заболеваний [Domazet-Lošo, Tautz, 2008a]. При этом было показано, что некоторые гены, вовлеченные в развитие онкологических заболеваний, обладают более ранним эволюционным происхождением. Дополнительные доказательства эволюционных различий генов, вовлеченных в онкогенез, были представлены в работе [Zhang и др., 2019a], где с помощью филостратиграфического анализа сетей ко-экспрессии генов в клетках карциномы яичников было показано, что функциональные модули, связанные с опухолевой прогрессией, не являются равномерно распределенными по возрастам составляющих их генов. Авторы обнаружили, что кластеры «древних» генов определяют ключевые метаболические и сигнальные пути, необходимые для поддержания базовых клеточных функций, тогда как более «молодые» гены способствуют тонкой настройке взаимодействий с микроокружением опухоли.

В контексте изучения опухолей, индексы TAI и TDI позволяют выявлять характерные паттерны экспрессии на разных стадиях патологического процесса. В частности, они демонстрируют, как в онкологических заболеваниях смещается равновесие между активностью «древних» и «молодых» генов. Так, было показано, что несмотря на высокую гетерогенность, разные опухоли демонстрируют повторяемый сдвиг транскриптомов к сходным режимам экспрессии, что приводит к ограниченному набору устойчивых клеточных состояний. Филостратиграфический анализ в данном случае указывает, что этот сдвиг обусловлен усиленной активностью генетических программ, сформировавшихся на ранних этапах эволюции, наряду с подавлением функций генов, сформировавшихся в процессе эволюции многоклеточных организмов, что согласуется с идеей о функциональной специализации генов разного возраста в онкогенезе [Chen, He, 2016].

Эта гипотеза также была подтверждена филотранскриптомным анализом широкого ряда солидных опухолей, который показал, что переход нормального клеточного состояния к злокачественному сопровождается сдвигом в экспрессии генов: активируются эволюционно «древние» гены, участвующие в процессах, возникших у одноклеточных организмов, в то время как экспрессия генов,

связанных с многоклеточными регуляторными процессами, подавляется [Trigos и др., 2017].

Таким образом, филотранскриптомный анализ дополняет традиционный филостратиграфический анализ, предоставляя количественные метрики для исследования не только происхождения генов, но и их реальной функциональной активности в контексте эволюции и патологии. Однако, несмотря на значительный прогресс, в данной области сохраняются существенные пробелы. В частности, остается недостаточно изученной динамика филотранскриптомных показателей (TAI/TDI) на различных стадиях клинического развития опухолей. Большинство существующих работ анализируют опухоли без учёта стадийности, что ограничивает понимание эволюционной перестройки транскриптома в процессе прогрессирования заболевания.

Цели и задачи работы

Целью работы является филостратиграфический и филотранскриптомный анализ генов, вовлечённых в патогенез онкологических заболеваний человека.

Для достижения этой цели были поставлены следующие задачи:

1. Разработка программных средств для филостратиграфического и филотранскриптомного анализа генных сетей, а также для автоматизированной обработки и анализа данных экспрессии генов из опухолевых образцов, полученных на различных клинических стадиях;
2. Выявление белок-кодирующих генов, дифференциально экспрессирующихся на разных стадиях развития злокачественных опухолей;
3. Исследование макро- (PAI) и микроэволюционных (DI) характеристик генов, дифференциально экспрессирующихся в тканях опухолей разных органов человека, с использованием филостратиграфического анализа;
4. Исследование филотранскриптомных характеристик (TAI и TDI) транскриптомов на разных стадиях развития опухолей человека;
5. Анализ эволюционных характеристик генов, входящих в генные сети онкологических заболеваний человека из баз данных KEGG PATHWAY и WikiPathways.

Научная новизна

В рамках данной кандидатской диссертации был проведен комплексный анализ эволюционных характеристик генов, дифференциально экспрессирующихся на различных клинических стадиях опухолей, с использованием методов филостратиграфии и филотранскриптомного анализа. Впервые была применена методика филостратиграфии для определения индекса филостратиграфического возраста генов, проявляющих дифференциальную экспрессию на различных стадиях патологического развития ряда карцином человека. На основе интеграции индексов возраста (TAI) и дивергенции (TDI) транскриптома была разработана методика филотранскриптомного анализа, предназначенная для количественной оценки эволюционной динамики экспрессии генов на различных стадиях развития онкологических заболеваний. С использованием современных технологий был разработан веб-сервис – Orthoweb 2, обеспечивающий расчёт эволюционных индексов на основе ортологов и гомологии последовательностей, а также интеграцию и визуализацию генных сетей из внешних источников. Orthoweb 2 является на сегодняшний день единственным доступным программным комплексом, позволяющим проводить филостратиграфический анализ как отдельных генов, так и генных сетей, не требуя ни знания языков программирования, ни установки ресурсоёмких программ, такие как Cytoscape. Важным усовершенствованием является реализованная способность к интеграции в эволюционный анализ генных сетей дополнительных данных, включая функциональную аннотацию генов (GO), сведения об однонуклеотидных полиморфизмах (SNP) и данные об экспрессии генов.

Теоретическая и практическая ценность

Orthoweb 2 представляет собой программный комплекс для проведения филостратиграфического анализа как отдельных генов, так и генных сетей. Сервис интегрирует разнородные данные, включая функциональную аннотацию, однонуклеотидные полиморфизмы (SNP) и данные об экспрессии генов, в рамках эволюционного анализа. Это позволяет изучать закономерности эволюции геномов и генных сетей, в том числе в контексте сложных биологических процессов, таких как онкогенез. Результаты, полученные с использованием Orthoweb 2, например, распределения значений TAI в клиническом развитии ряда карцином, способствуют развитию представлений об эволюционной природе злокачественных заболеваний,

раскрывая взаимодействие генетических программ, сформировавшихся на разных этапах эволюции, в процессе прогрессирования опухолей.

Положения, выносимые на защиту

1. Разработан программный комплекс Orthoweb 2, предназначенный для расчета эволюционных характеристик генов – филостратиграфических и филотранскриптомных индексов на основе транскриптомных данных, примененный для анализа генов дифференциально экспрессирующихся на различных стадиях развития карцином человека.
2. Эволюционный возраст дифференциально экспрессирующихся генов, вовлеченных у человека в развитие карциномы лёгких, печени, простаты, мочевого пузыря и молочной железы, различается на разных стадиях развития опухоли – на ранних и поздних стадиях преобладают эволюционно «молодые» гены, на промежуточных – эволюционно «древние».

Апробация работы

Материалы работы вошли в отчёты по гранту РФФИ (№20-04-00885). Результаты работы были представлены на всероссийских и международных научных конференциях и школах молодых учёных в Новосибирске (BGRS/SB 13, 14 и SBB-2023, Марчуковские научные чтения) и Москве (Постгеном2024).

Объём и структура диссертации

Работа состоит из введения, обзора литературы, описания материалов и методов, результатов, заключения, выводов и списка литературы. Работа изложена на 144 страницах, содержит 25 рисунков, 6 таблиц и 3 приложения. Список литературы включает 277 ссылок.

Публикации

По материалам диссертации опубликовано 7 работ, из них 3 статьи в рецензируемых научных журналах, входящих в перечень ВАК, 4 тезиса конференций, а также одно авторское свидетельство

Личный вклад автора

Основная часть работы выполнена автором самостоятельно. Автор принимал непосредственное участие в постановке задач, разработке конвейера программ, проведении вычислительных экспериментов, анализе данных и полученных результатов.

Благодарности. Автор выражает благодарность и глубокую признательность научному руководителю д.б.н. Лашину С.А., а также д.б.н. Афонникову Д.А., к.б.н. Землянской Е.В., к.б.н. Клименко А.И. и к.б.н. Казанцеву Ф.В. за консультации и плодотворные научные дискуссии, а также академику Колчанову Н.А. и всем сотрудникам Отдела Системной Биологии ИЦиГ СО РАН.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Глава 1. Обзор литературы

Первая глава диссертации представляет собой обзор литературы. В разделе 1.1 рассматривается, как применение концепций эволюционной биологии позволяет глубже понять природу различных заболеваний, включая злокачественные, анализируя их с позиции эволюционных процессов и адаптаций. В разделе 1.2 описаны методы определения режима отбора на основе сравнения последовательностей, направленного на изучение влияния отбора на уровне генов и популяций. Раздел 1.3 посвящен методологии и применении филостратиграфического анализа в изучении эволюционной истории генов, а также подробному описанию методологии анализа филогенетических индексов. В разделе 1.4 описываются программные инструменты, используемые в филостратиграфии. Раздел 1.5 содержит описание баз данных и биоинформационических ресурсов, применяемых для анализа онкологических заболеваний. В разделе 1.6 обзора литературы обсуждаются молекулярно-генетические аспекты опухолей, где особое внимание уделяется ключевым признакам злокачественных заболеваний и нарушениям сигнальных путей в новообразованиях. В заключении по обзору литературы подчеркивается значимость применения филостратиграфического метода в изучении онкологических заболеваний.

Глава 2. Материалы и методы

В разделе 2.1 представлены данные, использованные для анализа экспрессии генов в злокачественных опухолях. Основным источником данных является база данных The Cancer Genome Atlas (TCGA), которая предоставляет информацию о количестве транскриптов, картированных на каждый ген, полученную путем секвенирования транскриптомов из образцов опухолевых и здоровых тканей.

В исследовании были использованы данные по экспрессии генов в различных типах злокачественных опухолей на разных стадиях клинического развития: печеночной карциномы, светлоклеточной карциномы почек, аденокарциномы кишечника, карциномы молочной железы, аденокарциномы предстательной железы, карциномы щитовидной железы, карциномы уретерии мочевого пузыря и эндометриоидной карциномы тела матки. Для каждого типа опухоли были также проанализированы образцы соответствующих нормальных тканей, полученных от тех же пациентов, что позволяет минимизировать влияние индивидуальной вариабельности. Всего в исследовании было использовано 5058 образцов опухолей.

Образцы нормальных тканей (NAT – Normal Adjacent Tissues) были взяты из гистологически нормальных тканей, прилегающих к краям опухоли, и использовались в качестве контроля. Несмотря на возможные различия в уровнях экспрессии некоторых генов между NAT и истинно здоровыми тканями, использование NAT как контрольной группы было обосновано необходимостью избежать групповых эффектов, связанных с использованием данных из других источников.

В разделе 2.2 представлено подробное описание программных средств и подходов, применяемых для реализации вычислительного конвейера cancEv, направленного на анализ транскриптомных данных. Основными инструментами, использованными в исследовании, стали языки программирования Python и R, а также bash-скрипты.

Конвейер cancEv включает несколько ключевых этапов, начиная с выгрузки и фильтрации данных. На первом этапе данные экспрессии генов извлекаются из TCGA с использованием пакета R *recount3*, после чего проводится фильтрация некодирующих участков и генов с низким уровнем экспрессии, что позволяет сохранить около 80% генов для дальнейшего анализа. Также осуществляется отбор образцов, относящихся к первичным опухолям и нормальным тканям, что позволяет сосредоточиться на сравнении этих типов тканей.

На втором этапе проводится анализ дифференциальной экспрессии с помощью программы *DESeq2*, что позволяет выявить дифференциально экспрессирующиеся гены в опухолевых и нормальных тканях. Пороговые значения для отбора дифференциально экспрессирующихся генов были установлены на уровне $\log FC > 1$ и <-1 , что соответствует изменению экспрессии в два раза, при этом порог ложноположительных результатов был установлен на уровне менее 5%, оцененные

по p-value с поправкой Бенджамина-Хохберга. Результаты включают списки дифференциально экспрессирующихся генов и матрицы уровней экспрессии, а также визуализацию в виде графиков volcano plot.

Третий этап включает расчет эволюционных индексов с использованием разработанного нами веб-сервиса Orthoweb 2. Для реализации серверной части веб-инструмента Orthoweb 2 был использован фреймворк *Spring* на языке программирования Java, а для клиентской части – фреймворки *vue/webix*. Визуализация сетей осуществлялась с помощью библиотеки *cytoscape.js*. Для хранения данных из базы KEGG и промежуточных результатов анализа применялась документоориентированная СУБД MongoDB. Результаты анализа сохранялись в базе данных на основе PostgreSQL, состоящей из пяти таблиц, включая организмы, гены и индексы PAI и DI.

Orthoweb 2 использовался для расчета индексов филостратиграфического возраста (PAI – Phylostratigraphy Age Index). Расчет проводился как по группам ортологов KEGG, так и по таблицам идентичности с шагом 10%. Порог оценки Смита-Ватермана был установлен на ноль для работы с полногеномными данными. Также оценивался индекс дивергенции (DI) при уровне параметра $dN/dS = 2$, что соответствует анализу ортологов из семейства гоминидов.

На четвертом этапе проводится анализ филотранскриптомных индексов и тестирование паттерна песочных часов при помощи программного пакета myTAI, который позволяет вычислять индекс возраста транскриптома

$$TAI_s = \frac{\sum_{i=1}^n ps_i e_{is}}{\sum_{i=1}^n e_{is}}$$
 и индекс дивергенции транскриптома $TDI_s = \frac{\sum(e_{is}*ds_i)}{\sum e_{is}}$

где ps_i – целое число, отражающее возраст для гена с индексом i , e_{is} – уровень экспрессии гена с индексом i в образце s , ds_i – результат вычисления показателя dN/dS , отражающего отношения несинонимичных и синонимичных замен при сравнении последовательностей для гена i и его ортолога, а n – общее число генов

Профили TAI и TDI визуализировались и оценивались с помощью теста FlatLineTest, а также редукционного теста песочных часов и обратного теста песочных часов для выявления значимых паттернов [Drost и др., 2018].

Глава 3. Результаты

В разделе 3.1 описана разработка веб-сервиса Orthoweb 2, расширяющего возможности анализа индексов филостратиграфического и филотранскриптомного возраста. В разделе 3.2 приводится детальное описание разработанного конвейера

cancEv для комплексного анализа данных экспрессии генов, полученных из проектов базы данных TCGA, который включает этапы филостратиграфического и филогенетического анализа. Раздел 3.3 содержит результаты анализа дифференциальной экспрессии генов в злокачественных опухолях разных клинических стадий. Результаты филостратиграфического анализа списков дифференциально экспрессирующихся генов представлены в разделе 3.4. Результаты расчета индексов возраста транскриптома (TAI/TDI) изложены в разделе 3.5.

Orthoweb 2 (Рис.1), основанный на предшествующем приложении Orthoscape, представляет собой автономное веб-приложение, разработанное для анализа макро- и микроэволюционных характеристик генов с использованием эволюционных индексов PAI и DI.

Ключевым нововведением стала реализация двух методов оценки индексов возраста генов. Первый метод основывается на использовании данных групп ортологии KEGG (KEGG Orthology, KO). Второй метод оценки использует таблицы сходства последовательностей (Best Similarity Table), что позволяет определять индекс возраста генов на основе сходства аминокислотных последовательностей и результатов алгоритма Смита-Ватермана.

Важной новой функцией является возможность анализа транскриптомов на основе индексов TAI и TDI, где данные экспрессии представляются в виде таблицы с идентификаторами генов и значениями их активности в различных условиях.

The screenshot shows the Orthoweb 2 landing page. At the top, there is a circular phylogenetic tree icon and a welcome message: "Welcome to OrthoWeb 2. On this page you can launch the evolutionary analysis of gene sets." Below the header, there is a search bar labeled "Work ID:" containing "19_11_46_40_dhmgbui", a "Use ID converter" link, and a help icon. A horizontal line separates this from the main parameter setup area. The setup area is divided into several sections:

- Setup parameters or use the defaults**
- The type of orthology**: Radio buttons for "KEGG Orthology groups" (selected) and "Best Similarity Table".
- KO groups filtering**: Radio buttons for "All genes" (selected) and "Only same label".
- The thresholds to filter orthologous genes**: Input fields for "Identity" (0.5) and "SW Score" (0).
- dN/dS setup**: Input field for "dN/dS level" (2).
- Organisms**: An input field containing "hsa:5553; hsa:5554; hsa".
- Additional parameters**: Checkboxes for "DI analysis" (selected), "GO analysis", "SNP analysis", and "Use online database".
- Choose the type of input data**: Radio buttons for "Form" (selected), "Gene list file", and "Network file".
- Genes**: An input field containing "hsa:5553; hsa:5554; hsa".

Рисунок 1. Страница Orthoweb 2, на которой можно настроить параметры запуска и провести анализ списков генов.

Orthoweb 2 также поддерживает эволюционный анализ генных сетей, импортированных из баз данных KEGG Pathway и WikiPathways, а также предоставленных пользователем. Результаты анализа визуализируются с учетом индекса PAI, что позволяет оценить возраст генов в контексте их функциональных взаимодействий. Реализован импорт пользовательских генных сетей в формате tsv.

В систему также интегрированы режимы анализа терминов генной онтологии (GO) и обогащённости SNP. Для обеспечения актуальности информации Orthoweb 2 подключён к онлайн-базам данных, через API ресурса *geneontology.org*, что обеспечивает доступ к агрегированным сведениям из TAIR, FlyBase, UniProt и других ресурсов.

Анализ обогащения списков дифференциально экспрессирующихся генов терминами GO выявил универсальные для различных стадий и локализаций опухолей изменения (рис. 2). Было показано, что во всех случаях усиливается экспрессия генов, регулирующих клеточный цикл (включая переходы G1/S, G2/M и митоз), что отражает усиленное пролиферативное поведение опухолевых клеток. Повышена также активность генов систем reparации ДНК, прежде всего отвечающих за негомологичное соединение концов (*PRKDC*, *MAD2L2*, *POLQ*), что сопровождается высокой вероятностью ошибок и накоплением мутаций. Наблюдается повышение экспрессии генов, участвующих в транскрипции, трансляции и эпигенетической регуляции, что отражает усиленную метаболическую активность опухоли и потребность в синтезе РНК и белков для пролиферации. Одновременно во всех образцах снижена экспрессия генов, связанных с апоптозом и другими формами запрограммированной клеточной гибели, что указывает на подавление механизмов клеточной гибели опухолевых клеток. Похожая тенденция снижения прослеживается и для генов межклеточной адгезии и клеточной подвижности. Филостратиграфический анализ дифференциально экспрессирующихся генов выявил два выраженных пика их представленности. Первый пик приходится на ранние этапы эволюции, соответствующие переходу от одноклеточных к многоклеточным организмам (PAI = 3). К генам этого периода относятся *TNFSF11*, *NOS1*, *TP53I3*, *CDX2* и *BRAF*, вовлечённые в процессы межклеточных взаимодействий, хемотаксиса и передачи сигналов. Второй пик наблюдается среди генов более позднего происхождения, включая филостраты позвоночных (PAI = 6), костных позвоночных (PAI = 7) и плацентарных млекопитающих (PAI = 9). Эти гены участвуют в регуляции клеточного цикла,

контроле апоптоза, ангиогенеза и воспалительных процессов; среди них *FZD3*, *BMP4*, *GPX3* и *NPM1*.

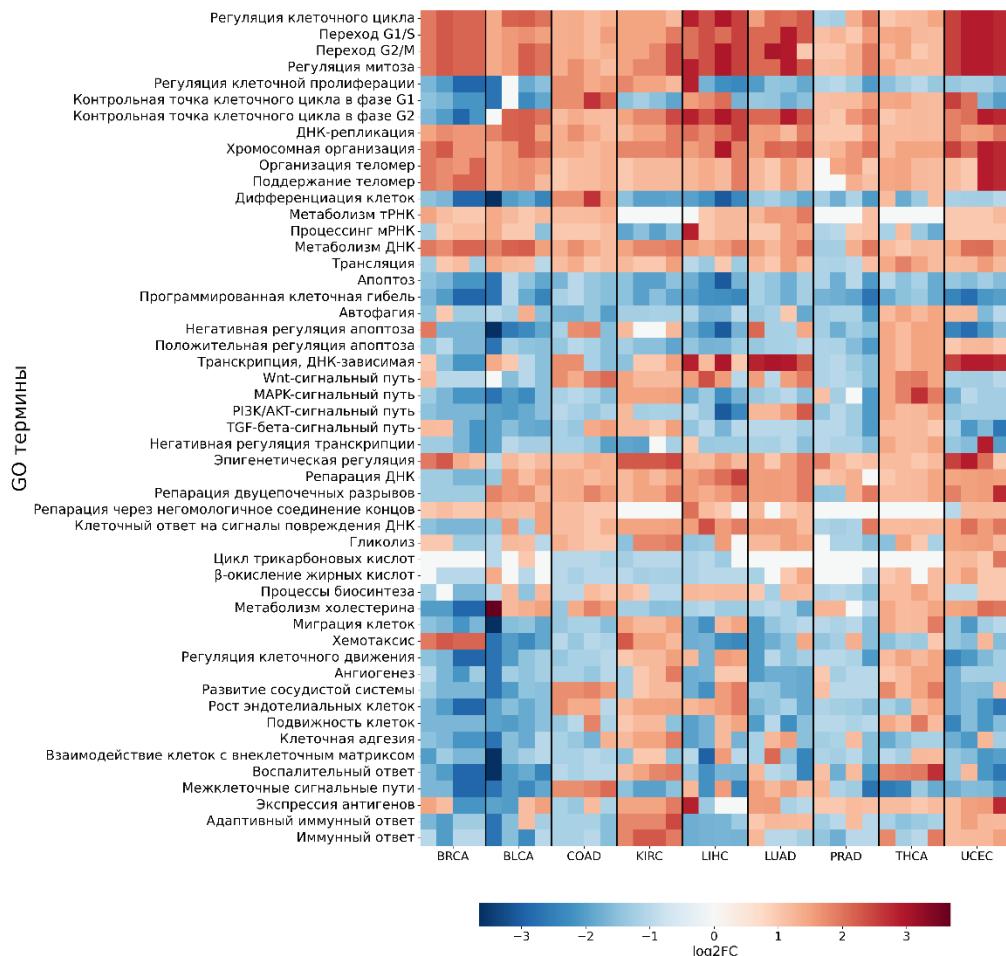


Рисунок 2. Изменение уровней экспрессии генов, обогащенных терминами ГО биологических процессов, ассоциированных с ключевыми признаками злокачественности

Сравнительный анализ распределения значений PAI показал статистически значимое увеличение доли эволюционно «молодых» генов ($PAI = 4-9$) среди ДЭГ по сравнению с общим набором кодирующих генов человека ($p < 1 \times 10^{-5}$), что подчёркивает их важную роль в адаптации опухолевых клеток. Применение точного теста Фишера, критерия Колмогорова–Смирнова и бутстреп-анализа подтвердило достоверность выявленных различий, а значения χ^2 для выборок ДЭГ значительно превышали показатели случайных выборок.

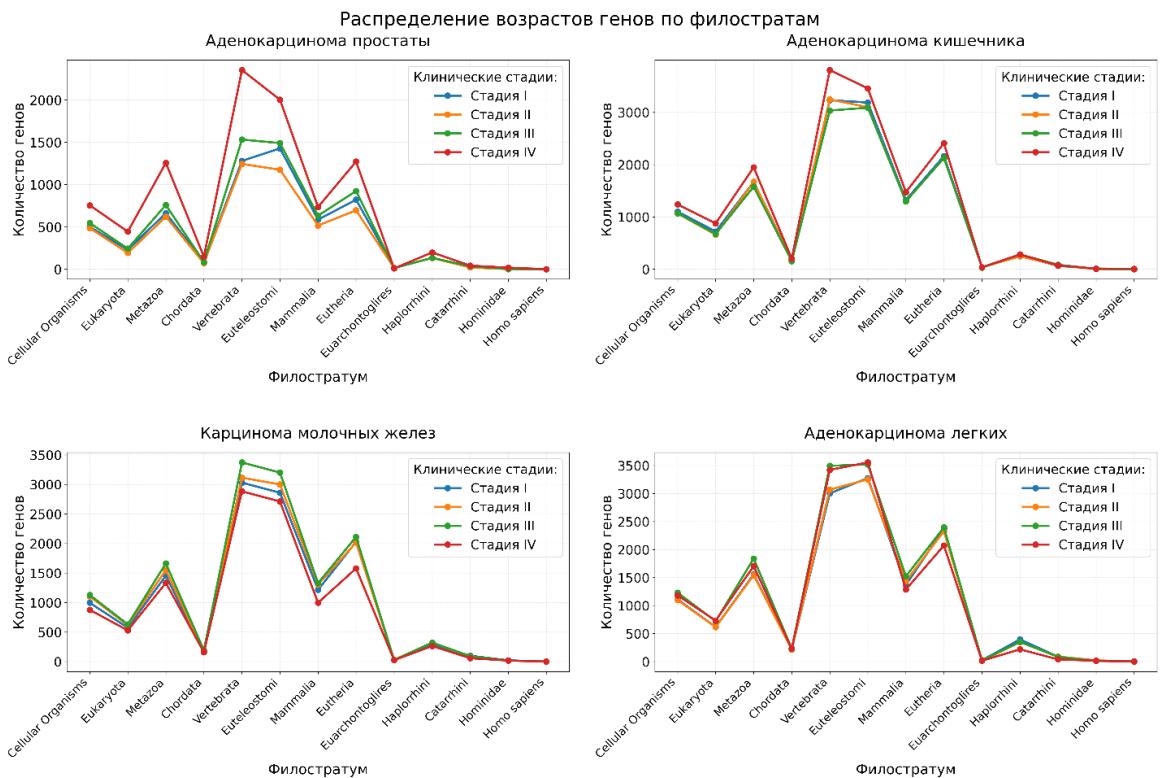


Рисунок 3. Распределение дифференциально экспрессирующихся генов (ДЭГ) по филостратам на различных стадиях патологического развития опухолей аденокарциномы кишечника, лёгкого, предстательной железы и карциномы молочной железы.

Анализ индексов дивергенции (DI) выявил преобладание эволюционно консервативных генов среди ДЭГ. Данний результат указывает на консервативность их нуклеотидных последовательностей и молекулярных функций в ходе эволюции, что может быть обусловлено их ключевой ролью в фундаментальных клеточных процессах.

В разделе 3.6, посвящённом филогенетическому анализу индексов возраста транскриптома (TAI) и индексов дивергенции транскриптома (TDI) на различных стадиях онкогенеза, была проведена комплексная оценка изменения этих индексов на различных клинических стадиях заболевания (Рис. 5). Для каждого из типов опухолей были выделены ранняя стадия (стадия I), средние стадии (стадии II и III) и поздняя стадия (стадия IV).

В целом для большинства типов карцином значения TAI оказались ниже, чем в нормальных тканях, что отражает активацию консервативных онкогенных программ, тогда как более высокие значения TAI в контроле связаны с вовлечением «молодых» адаптивных систем, включая иммунный ответ.

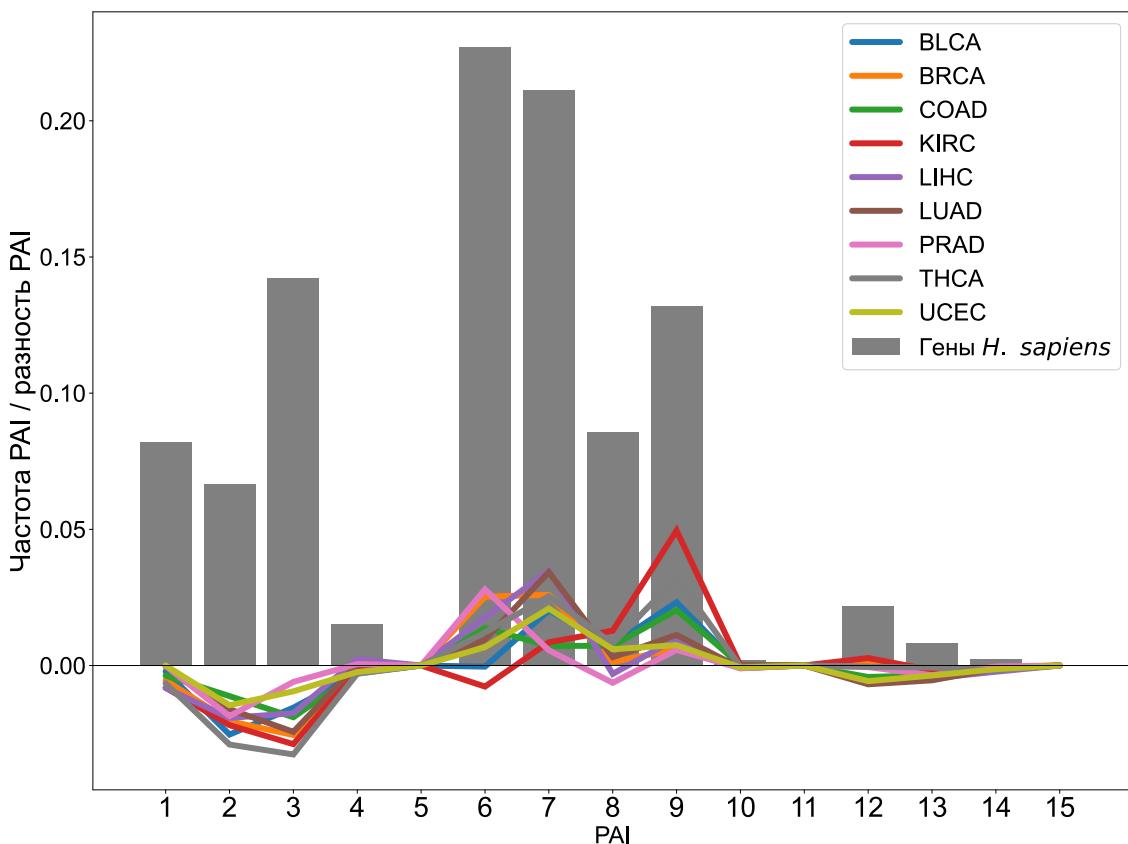


Рисунок 4 Серыми столбцами показано распределение частот встречаемости филостратиграфических индексов всех белок-кодирующих генов человека, участвующих в анализе. Цветными линиями показано отклонение частот филострат в списке дифференциально экспрессирующихся генов для соответствующего типа злокачественного новообразования от распределения среди всех белок-кодирующих генов человека

При анализе динамики по стадиям выявлены статистически значимые паттерны, соответствующие модели «песочных часов». Для карциномы печени, молочной железы и мочевого пузыря в распределении ТАИ отмечалось усиление экспрессии эволюционно «молодых» генов на ранних и поздних стадиях и преобладание эволюционно «древних» генов на промежуточных. В аденокарциномах лёгкого и предстательной железы аналогичный паттерн наблюдался по индексу TDI, где наблюдалось снижение степени консервативности транскриптома на ранних и поздних стадиях и её усиление на средних. В опухолях кишечника, напротив, зафиксирован обратный паттерн «песочных часов». В разделе 3.7, посвящённом филостратиграфическому анализу генных сетей, были изучены эволюционные характеристики генов, формирующих ключевые сигнальные пути, вовлечённые в развитие и прогрессирование злокачественных новообразований (Рис. 6). Для анализа использовались данные из баз KEGG Pathways и WikiPathways,

что позволило охватить как тканеспецифичные, так и универсальные процессы, определяющие молекулярную основу онкогенеза.

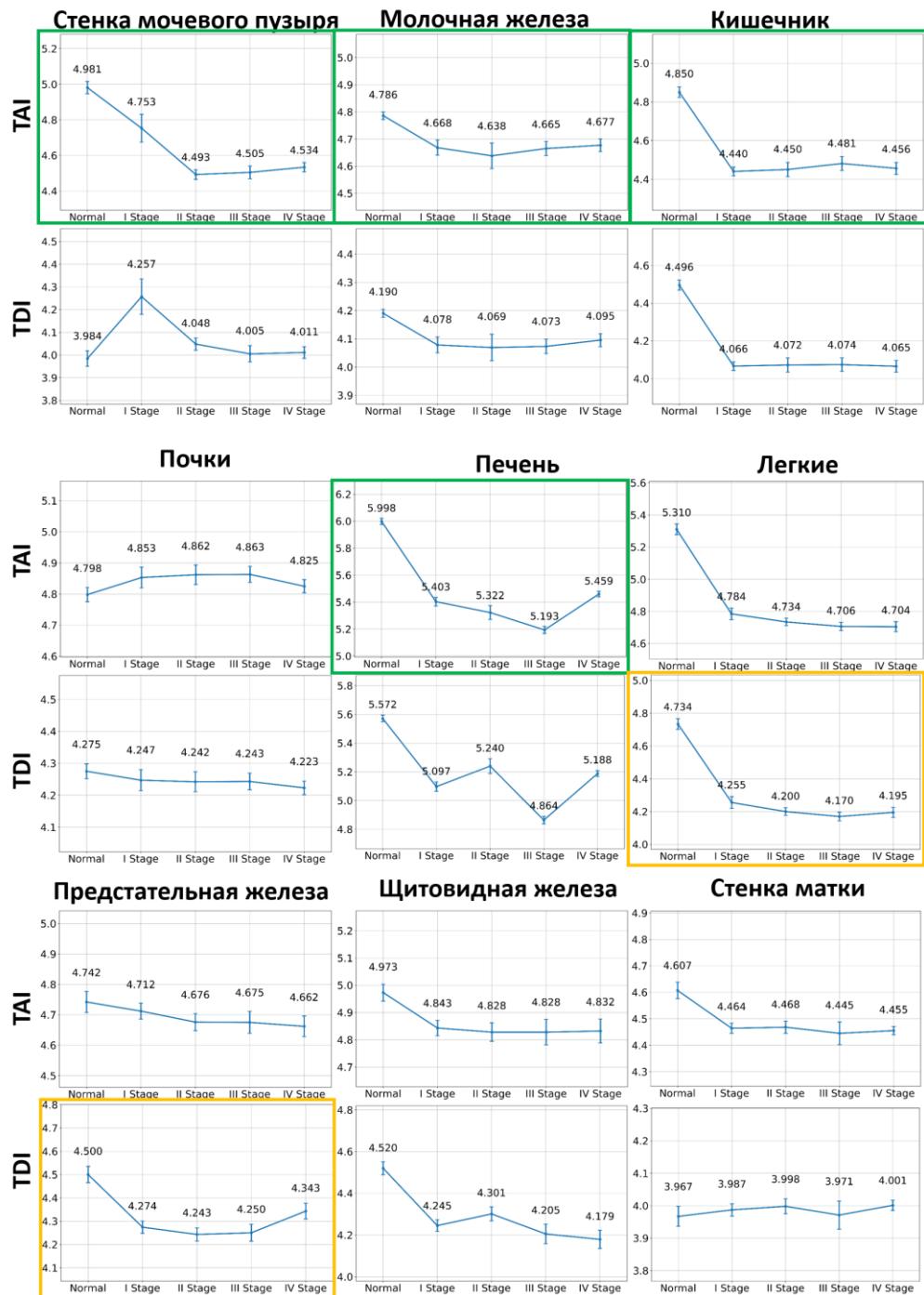


Рисунок 5. Распределение значений индексов возраста транскриптома (TAI) и индексов дивергенции транскриптома (TDI). Зеленым выделены распределения TAI, для которых при помощи теста «песочных часов» был выявлен статистически значимый паттерн «песочных часов» ($p\text{-value} < 0.05$), жёлтым – статистически значимые паттерны «песочных часов» в распределении TDI.

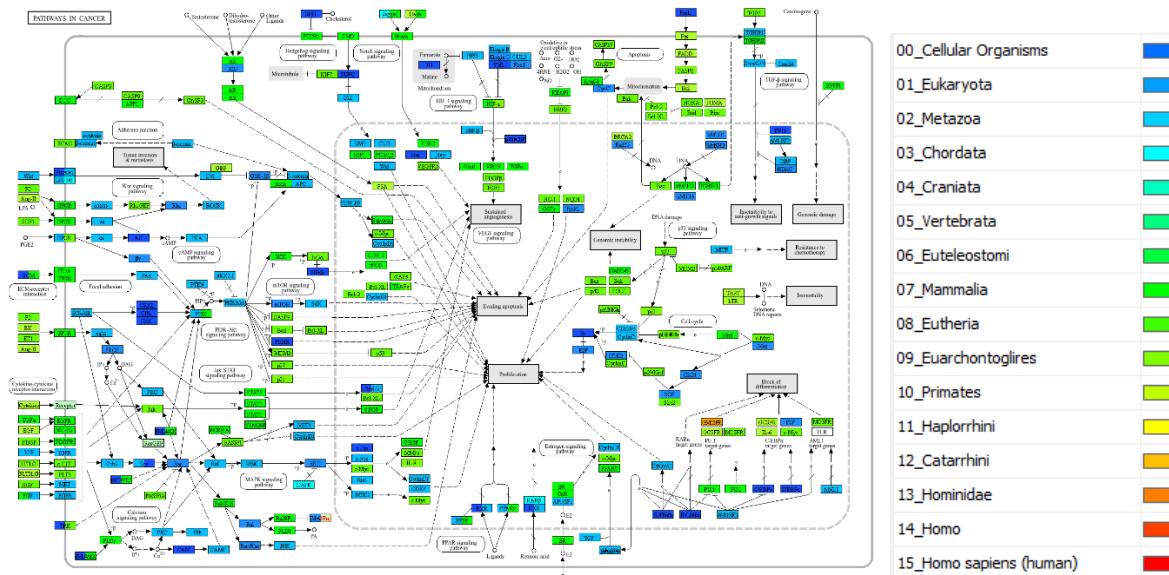


Рисунок 6. Результаты филостратиграфического анализа генов сети сигнальных путей онкологических заболеваний из базы данных KEGG PATHWAY hsa: 05200 - Pathways in cancer.

Проведённый филостратиграфический анализ сигнальных сетей, ассоциированных с онкогенезом, показал, что распределения значений PAI у генов, кодирующих их компоненты, неоднородны и отражают эволюционную стратификацию функций. Белковые лиганды и рецепторы поверхности клетки (внешние сигналы: ростовые факторы, цитокины) представлены продуктами эволюционно более «молодых» генов (PAI = 6–8; Vertebrata–Mammalia). К таким генам относятся *TGFA*, *EGFR*, *EGF*, *KITLG*, *FLT3LG*, *HGF*, а также семейства *PDGF* и *PDGFR*. В то же время гены ряда ключевых ростовых факторов возникли значительно раньше, на этапе становления многоклеточных организмов (PAI = 3; Metazoa), например, семейства *IGF* и *FGF*. Напротив, каскад передачи сигнала от рецепторов к ядру реализуется через высококонсервативные протеинкиназы, кодируемые генами раннего происхождения (PAI = 1–3, Cellular organism–Metazoa). Сюда входят семейства малых ГТФаз RAS, серин/треониновых киназ RAF и митоген-активируемых киназ MAPK, функции которых на протяжении эволюции сохранялись практически неизменными.

Показано, что такие пути, как WNT/β-catenin и Hedgehog представляют собой системы, сформированные преимущественно эволюционно «древними» генами, тогда как JAK-STAT и PI3K-Akt включают гены различного эволюционного возраста. Подобная стратификация отражает эволюционную пластичность опухолевых сетей, позволяя новым рецепторным системам интегрироваться в

консервативные сигнальные механизмы, что обеспечивает их функциональную гибкость и устойчивость в процессе онкогенеза.

ВЫВОДЫ

1. Разработаны программные комплексы *Orthoweb* 2 для филостратиграфического и филотранскриптомного анализа генных сетей и *cancEv* для анализа транскриптомных данных из базы The Cancer Genome Atlas (TCGA) на разных клинических стадиях опухолей.
2. Выявлено, что в карциномах лёгкого, печени, мочевого пузыря, простаты и молочной железы, независимо от клинической стадии опухоли, повышена экспрессия генов, ассоциированных с клеточной пролиферацией и репарацией ДНК, и снижена активность проапоптотических генов.
3. Показано, что в тканях карцином доля дифференциально экспрессирующихся генов, относящихся к «эволюционно молодым» (PAI = 6, 7, 9), достоверно выше, чем их доля в общем составе генома человека.
4. Показано, что в карциномах лёгкого, печени, мочевого пузыря и простаты, а также в протоковой карциноме молочной железы на ранних и поздних стадиях преобладает экспрессия эволюционно «молодых» генов, тогда как на промежуточных – эволюционно «древних».
5. Установлено, что ключевые сигнальные пути, вовлечённые в онкогенез, различаются по эволюционному возрасту своих компонентов: пути WNT/β-catenin и Hedgehog содержат преимущественно эволюционно «древние» гены (PAI = 1–2), путь JAK–STAT представлен в основном «молодыми» генами (PAI = 6), тогда как в ERK–MAPK и PI3K/Akt присутствуют как эволюционно «древние» гены, кодирующие цитоплазматические белки, так и более «молодые», кодирующие регуляторные и рецепторные белки.

Список работ, опубликованных по теме диссертации.

Статьи в рецензируемых журналах, входящих в перечень ВАК:

1. Иванов Р.А., Лашин С.А. Внутриопухолевая гетерогенность: модели возникновения и эволюции злокачественных опухолей // Вавиловский журнал генетики и селекции. – 2023. – Т. 27, № 7. – С. 815–819. – <https://doi.org/10.18699/VJGB-23-94>. Scopus (SJR 0.33), WoS CC (IF 0.98), РИНЦ, Белый список.

2. **Иванов Р.А.**, Мухин А.М., Казанцев Ф.В., Мустафин З.С., Афонников Д.А., Матушкин Ю.Г., Лашин С.А. Orthoweb: программный комплекс для эволюционного анализа генных сетей // Вавиловский журнал генетики и селекции. – 2024. – Т. 28, № 8. – С. 874–881. – <https://doi.org/10.18699/vjgb-24-95>. Scopus (SJR 0.33), WoS CC (IF 0.98), РИНЦ, Белый список.
3. **Ivanov R.A.**, Afonnikov D.A., Matushkin Y.G., Lashin S.A. Evolutionary transcriptomics of cancer development // International Journal of Molecular Sciences. – 2025. – Vol. 26, № 11. – P. 5041. – 10.3390/ijms26115041. Scopus (SJR 1.27), WoS CC (IF 4.9), РИНЦ, Белый список

Авторские свидетельства:

1. Лашин С.А., Мустафин З.С., Казанцев Ф.В., Мухин А.М., **Иванов Р.А.**, Матушкин Ю.Г., Афонников Д.А. Программа для расчёта филостратиграфических индексов (Ортовеб), Свидетельство № 2024618826, 17 апреля 2024

Тезисы конференций:

1. **Иванов Р.А.**, Лашин С.А. Филостратиграфический анализ генов онкологических заболеваний человека // 14th International young scientists school «System Biology and Bioinformatics» (SBB-2023): Тезисы докладов. – Новосибирск, 2023. – С. 12.
2. **Ivanov R.**, Mustafin Z., Lashin S. Phylostratigraphic analysis of human cancers transcriptomic data // BGRS/SB-2022: 13th International Multiconference on “Bioinformatics of Genome Regulation and Structure/Systems Biology”: Abstracts. – Novosibirsk, 2022. – P. 792.
3. **Ivanov R.**, Lashin S., Afonnikov D., Matushkin Yu. Integration of phylostratigraphic and phylotranscriptomic analysis in the study of differentially expressed genes in cancer tissues // BGRS/SB-2024: 14th International Multiconference on “Bioinformatics of Genome Regulation and Structure/Systems Biology”: Abstracts. – Novosibirsk, 2024. – P. 801.
4. **Иванов Р.А.**, Афонников Д.А., Матушкин Ю.Г., Лашин С.А. Эволюционный подход к изучению дифференциальной экспрессии в тканях карцином различных стадий // VI Международная конференция «ПОСТГЕНОМ–2024»: Сборник тезисов докладов. – Москва, 2024. – С. 17.