На правах рукописи

КОЛПАКОВ ФЕДОР АНАТОЛЬЕВИЧ

Компьютерное моделирование биологических систем и анализ биомедицинских данных

1.5.8 – Математическая биология, биоинформатика

Автореферат
диссертации на соискание ученой степени
доктора биологических наук

Новосибирск 2023

Работа выполнена в Федеральном государственном бюджетном научном учреждении «Федеральный исследовательский центр информационных и вычислительных технологий» (ФИЦ ИВТ), г. Новосибирск.

Научный консультант:

Поройков Владимир Васильевич
член-корреспондент РАН, доктор биологических наук, профессор,
Федеральное государственное бюджетное научное учреждение «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича».

Официальные оппоненты:

Копылов Филипп Юрьевич
доктор медицинских наук, профессор, директор,
Институт персонализированной кардиологии, Первый Московский государственный медицинский университет имени И.М. Сеченова Министерства здравоохранения Российской Федерации.

Пальянов Андрей Юрьевич
доктор физико-математических наук, исполняющий обязанности директора, Федеральное государственное бюджетное учреждение науки Институт систем информатики им. А.П. Ершова Сибирского отделения Российской академии наук.

Самсонова Мария Георгиевна
доктор биологических наук, профессор Высшей школы прикладной математики и вычислительной физики, заведующая научно-исследовательской лабораторией «Математическая биология и биоинформатика»,
Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский политехнический университет Петра Великого».

Ведущее учреждение:
Федеральное государственное бюджетное учреждение науки Институт общей генетики им. Н.И. Вавилова Российской академии наук.

Защита состоится **5 февраля 2024 г. в 15.00** на заседании диссертационного совета 24.1.101.01 на базе Федеральное государственное бюджетное учреждение науки Институт проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН) по адресу: 127051, г. Москва, Большой Каретный переулок, д.19 стр. 1.

С диссертацией можно ознакомиться в библиотеке ФГБУ науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН), а также на сайте ИППИ РАН.

Автореферат разослан “___“ _____ 2023 г.

Ученый секретарь диссертационного совета
доктор биологических наук Казенников О.В.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы

С началом XXI века активное развитие получила системная биология – междисциплинарная наука, исследующая сложные взаимодействия в живых системах (Kitano H., 2002). В системной биологии предполагается определенный цикл проведения исследований, состоящий из теории, аналитического или компьютерного моделирования для формулировки гипотез о системе, экспериментальной проверки гипотез, и затем использование полученных данных для описания клеточных процессов, позволяющее улучшить компьютерную модель или уточнить теорию (Covert et al., 2004; Kholodenko et al., 2005). Поскольку конечной целью является построение портретной модели взаимодействий в сложной системе, экспериментальные методики, которые используются в системной биологии, должны быть максимально детализированными. Для этого применяются такие экспериментальные подходы как транскриптомика, метаболомика, протеомика и другие высокопроизводительные омиксные технологии, используемые для получения численных данных. Таким образом, успех в развитии высокопроизводительных автоматизированных биомедицинских технологий привел к ренессансу моделирования биологических систем на молекулярно-клеточном уровне.

Используя данный подход, было построено множество моделей, как на уровне отдельных клеток, так и на уровне отдельных органов, физиологических систем и целых организмов. Однако, при построении больших моделей биологических систем выявились ограничения и области, которые не охвачены существующими подходами в системной биологии:

- создание сложных моделей – до сих пор остается искусством, доступным немногим исследователям;
- итерационная разработка моделей – хотя такой подход подразумевается в цикле исследований, предлагаемом системной биологией, на самом деле он используется редко. По-видимому, основной причиной такой ситуации является недостаточное развитие соответствующей технологии и инструментария;
- объединение и повторное использование ранее созданных моделей – как правило, использование ранее созданных моделей в качестве составных

блоков при построении композиционных моделей вызывает существенные трудности, поскольку исходные модели не были предназначены для этого;

– тестирование и отладка моделей – данные процессы не формализованы и не автоматизированы;

– персонализация параметров моделей – настройка параметров модели под конкретный биологический объект (например, настройка параметров модели для определенного пациента на основе имеющихся индивидуальных клинических данных);

– использование численных данных, полученных с помощью высокопроизводительных технологий – системно-биологический подход предполагает применение этих данных для верификации и возможного последующего уточнения модели. Однако, только в небольшом количестве работ такие данные и результаты их анализа были использованы для построения моделей, например для создания полногеномной модели контроля генной экспрессии в фибробластах мыши (Schwanhausser B. et al., 2011), автоматической генерации полной модели клетки Mycoplasma genitalium на основе специально созданной базы данных (Karr et al., 2012, 2014; Maritan et al., 2022), для крупномасштабной механистической модели E. coli (Macklin et al., 2020). Инструмент для моделирования должен быть интегрирован с инструментом для анализа омиксных данных.

Таким образом, для эффективного создания и использования сложных модульных моделей биологических систем требуется развитие особого подхода и создание соответствующего инструментария (специализированные программные комплексы и базы данных).

Цель и задачи работы

Целью данной работы является разработка технологии и инструментария (программный комплекс и базы данных) для эффективного построения и использования сложных модульных моделей биологических систем в интеграции с анализом геномных/омиксных данных. Для достижения этой цели были поставлены и решены следующие задачи:

1. Разработать технологию для эффективного создания и использования сложных модульных моделей биологических систем.
2. Разработать программный комплекс (ПК), реализующий весь

инструментарий, необходимый для успешного использования предложенной технологии.

3. Включить в состав разрабатываемого комплекса программы и методы для анализа и графического представления генетических данных.

4. Разработать базу данных для интеграции широкого круга экспериментальных NGS (Next Generation Sequencing) данных по регуляции транскрипции для построения моделей регуляции транскрипции.

5. Продемонстрировать на практически важных задачах возможности применения разработанной технологии и программного комплекса для создания сложных моделей биологических систем.

6. Разработать технологию построения цифрового двойника пациента и продемонстрировать ее применимость на примере оптимизации выбора лекарственной терапии для лечения артериальной гипертензии.

Положения, выносимые на защиту

1. Разработана эффективная технология моделирования для итерационного создания модульных моделей сложных биологических систем, имеющих иерархическую структуру.

2. Разработан программный комплекс BioUML – Biological Universal Modeling Language (ПК BioUML, https:/biouml.org), реализующий инструментарий, необходимый для успешного использования предложенной технологии. В состав разработанного ПК также включены алгоритмы, программы, сценарии для анализа и графического представления биомедицинских данных.

3. Для построения моделей регуляции транскрипции разработана база данных GTRD (Gene Transcription Regulation Database, https:/gtrd.biouml.org), которая обеспечивает единообразную аннотацию, контроль качества, обработку и интеграцию большого объема экспериментальных NGS данных по регуляции транскрипции. На основе этих данных, используя методы машинного обучения разработаны модели для предсказания уровня транскрипции для наиболее экспериментально исследованных клеточных линий.

4. Используя разработанную технологию и ПК BioUML построены новые модульные модели сложных биологических систем:

– наиболее полная модульная модель апоптоза;

– иерархическая модель регуляция генной экспрессии в скелетных мышцах при физической нагрузке;

– модели эпидемиологического процесса COVID-19;

– модульная модель регуляции артериального давления у человека;

– модель терапии артериальной гипертензии основными классами лекарственных препаратов.

5. Разработана технология построения цифрового двойника пациента. Показана ее применимость на примере оптимизации выбора лекарственной терапии для лечения артериальной гипертензии.

Научная новизна

1. Разработана оригинальная технология моделирования для итерационного создания, тестирования и использования сложных модульных моделей биологических систем. Новизной этой технологии являются:

– универсальная мета-модель для представления широкого круга моделей биологических систем;

– расширение языка Antimony для текстового представления биологических моделей в формате SBML (Systems Biology Markup Language) и их графического представления в расширенной графической нотации SBGN (Systems Biology Graphical Notation).

2. Разработан специализированный ПК BioUML реализующий весь инструментарий, необходимый для успешного использования предложенной технологии для итерационного создания модульных моделей биологических систем. Уникальными особенностями ПК BioUML являются:

– интеграция методов анализа данных высокопроизводительных технологий с возможностями построения моделей сложных биологических систем (визуальное моделирование, автоматическая генерация моделей);

– поддержка широкого круга подходов к моделированию биологических систем: обычные дифференциальные уравнения с задержкой переменных по времени и мгновенными событиями, стохастическое моделирование, генерация моделей на основе правил (rule based modeling), одномерные модели кровотока на основе уравнений в частных производных, анализ метаболических потоков (flux balance analysis) и метод линейного

программирования, мультиагентное моделирование (agent based modelling). Отдельно стоит подчеркнуть, что в рамках разработанной нами технологии все эти методы можно одновременно использовать при построении модульных моделей, в этом случае для их численного моделирование их динамики используется мультиагентное моделирование.

Стоит особо отметить, что по результатам независимых сравнений (Maggioli et al., 2019), ПК BioUML признан единственным в мире симулятором биологических моделей, который проходит все тесты на правильность численного моделирования биологических систем SBML Test Suite Core v3.3.0, а также является самым быстрым симулятором.

3. Для графического представления геномных данных разработан новый геномный браузер, который по функционалу сопоставим с ведущими геномными браузерами (Ensembl genome browser, UCSC genome browser, JBrowse, Integrative Genomics Viewer, Integrated Genome Browser, NCBI Genome Data Viewer), а также имеет ряд уникальных возможностей: интерактивное объединение и пересечение треков с возможностью комбинации их отображения (объединение, только общие области, различия), отображение мастер-треков для базы данных GTRD.

4. Создана новая база данных GTRD, которая в настоящее время является крупнейшей в России базой данных по генетической информации (общий объем, включая исходные NGS данные, составляет 600+ ТБ) и одной из крупнейших в мире по регуляции транскрипции (по количеству ChIP-seq и ChIP-exo экспериментов). При создании базы данных GTRD были предложены следующие новые идеи/подходы для интеграции данных ChIP-seq экспериментов и достоверного выявления соответствующих районов связывания транскрипционных факторов (РСТФ):

– единообразная аннотация и обработка NGS данных при помощи сценариев;

– мета анализ результатов 4-х разных алгоритмов для выявления РСТФ: MACS/MACS2 (Zhang et al., 2008), SISSRs (Narlikar and Jothi, 2012), GEM (Guo et al., 2012) и PICS (Zhang et al., 2011) для наиболее достоверного выявления РСТФ;

– кластеры – объединение по разным экспериментальным условиям (клеточные типы, воздействие и т.п.) выявленных районов связывания для

заданного ТФ и алгоритма (см. выше);

– мета-кластеры – объединение таких кластеров по используемым алгоритмам, что позволяет получить уникальный набор районов связывания для заданного ТФ по всему геному;

– ранжирование РСТФ (rank aggregation) для выявления наиболее достоверных районов;

– мастер-трек – новая концепция для интеграции всех результатов анализа для заданного ТФ.

На основе полученных мета-кластеров РСТФ для 3-х клеточных линий построена оригинальная модель регуляции транскрипции, которая позволяет предсказать уровень экспрессии соответствующих стартов инициации транскрипции из базы данных FANTOM5.

5. Используя разработанную технологию и ПК BioUML построены новые модульные модели сложных биологических систем:

– наиболее полная модульная модель апоптоза, которая включает: 13 модулей, 286 белков и их комплексов, 684 реакции и 719 параметров;

– модульная агентная модель регуляции артериального давления у человека (2 агента (сердечно-сосудистая система и почка), 20 модулей, 25 обыкновенных дифференциальных уравнений (ОДУ), 160 алгебраических выражений, 160 переменных, 132 параметра, 10 дискретных событий). Она была калибрована и валидирована на большом количестве экспериментальных данных, соответствующих физиологическим параметрам человека в норме и патологии. Данная модель стала первой модульной моделью в базе данных Biomodels (ID – MODEL2202160001);

– с использованием агентной модели регуляции артериального давления у человека была разработана модель терапии артериальной гипертензии основными классами лекарственных препаратов, как по отдельности так и их комбинациями. Она была валидирована по наибольшему количеству клинических исследований, по сравнению со сходными моделями;

– модульная модель энергетического метаболизма в клетках скелетной мышцы;

– математическая модель, связывающая Ca2+-зависимый сигнальный путь с регуляцией экспрессии генов в клетках скелетной мышцы человека. При построении этой модели использовались экспериментально

подтвержденные сайты связывания ТФ CREB, FOS и JUN семейств в промоторных областях генов NR4A2, NR4A3 и PPARGC1A из базы данных GTRD;

– представленные выше две модели работы скелетных мышц были объединены в единую модульную модель, что впервые позволило построить математическую модель, описывающую изменение экспрессии отдельных генов в ответ на физическую нагрузку;

– модель эпидемиологического процесса COVID-19. Впервые для описания процесса заражения использовалось дифференциальное уравнение с задержкой, параметры которого определялись путем подгонки кривой инфицирования к реальным данным независимо от других параметров модели. Это позволяет с одной стороны более точно описать динамику инфицирования, а с другой – уменьшить количество параметров, определяемых в ходе решения обратной задачи, что делает эти параметры более идентифицируемыми.

6. Впервые была разработана технология построения цифрового двойника пациента и показана ее применимость на примере оптимизации выбора лекарственной терапии для лечения артериальной гипертензии. При использовании созданной модели регуляции артериального давления невозможно однозначно идентифицировать большое количество параметров этой модели по ограниченному данному из истории болезни пациента, данных клинических и лабораторных наблюдений. Для решения этой проблемы предложен принципиально новый подход, когда для каждого пациента строится популяция виртуальных пациентов (200-500), которые совпадают с заданным пациентом по наблюдаемым клиническим параметрам, но имеют различные значения внутренних скрытых параметров модели. Выбор оптимальной лекарственной терапии производится путем моделирования лечения различными антигипертензивными препаратами всей такой популяции и анализом однородности отклика отдельных пациентов.

Теоретическая значимость диссертационной работы

1. В диссертационной работе разработана новая технология для итерационного создания, тестирования и использования модульных моделей биологических систем.

2. При создании базы данных GTRD разработана новая технология объединения данных из различных NGS экспериментов в виде кластеров, мета-кластеров и мастер-треков.

3. Разработан ряд новых модульных моделей сложных биологических систем. 4. Разработана технология построения цифрового двойника пациента.

Практическая значимость диссертационной работы

1. ПК BioUML можно рассматривать как полнофункциональную отечественную биоинформатическую платформу, что особенно важно для обеспечения технологического суверенитета Российской Федерации в этой области. В этом качестве ПК BioUML используется в ряде отечественных проектов:

– В ходе реализации работ по контракту № 43-03/21/44/152 от 13.12.2021 "Исследование, обоснование и выбор программных решений для визуализации генетических данных и обеспечения инструментов для работы с генетической информацией в «Национальной базе генетической информации" на основе ПК BioUML созданы 2 подсистемы: "Геномный браузер" и "Рабочие окружения".

– Для создания биоинформатической платформы в рамках договора № 33.2022 на выполнение научно-исследовательских работ по теме: "Разработка и испытания биоинформатической платформы для хранения, анализа и графического представления данных, полученных при одномолекулярном секвенировании ДНК".

2. Sirius-web (https://sirius-web.org/) – это информационная платформа на основе ПК BioUML для проектов, связанных с анализом данных и моделированием для образовательных и научных проектов и активно используемая в Научно-технологическом университете «Сириус» для проведения исследований в области наук о жизни.

3. На основе ПК BioUML созданы коммерческие программы для визуализации и анализа геномных/омиксных данных:

– платформа Genome Enhancer (https://genexplain.com/genexplain-platform) – это специализированный вариант вариант ПК BioUML для анализа омиксных данных в автоматическом режиме (Kel et al., 2011), изначально разработанный ООО "Биософт.Ру" при поддержке гранта Фонда "Сколково";

– платформа geneXplain (https://genexplain.com/genexplain-platform) – это коммерческий вариант ПК BioUML, содержащий множество готовых сценариев для анализа омиксных данных, с использованием информации из баз данных TRANSFAC и TRANSPATH. Она разрабатывается компаниями geneXplain GmbH (Вольфенбютель, Германия) и ООО "Биософт.Ру" для анализа омиксных данных (Kel et al., 2011).

4. ПК BioUML используется как основной инструмент в курсах по системной биологии, проводимых в Новосибирском государственном университете и научно-технологическом университете "Сириус".

5. Разработанная технология создания модульных моделей биологических систем и реализующий ее ПК BioUML позволили перейти на более высокий уровень сложности создания таких моделей (сотни и тысячи переменных и уравнений), а также упростили и ускорили, с точки зрения создателей моделей, этот процесс.

6. Созданная база данных GTRD является высоко востребованной, широко используемой и цитируемой: три публикации в специализированных выпусках Nucleic Acids Research 2017, 2019 и 2021 года набрали в совокупности 400+ цитирований по версии Semantic Scholar (https://www.semanticscholar.org/), включая цитирования в журналах Nature и Science. На основе GTRD коллективом под руководством проф., чл.корр. РАН Макеева В.Ю. и д.б.н. Кулаковского И.В. были созданы ресурсы:

– HOCOMOCO – HOmo sapiens COmprehensive MOdel COllection – коллекция мотивов для сайтов связывания транскрипционных факторов (ССТФ) человека и мыши (https://hocomoco11.autosome.ru/);

– ADASTRA – Allelic Dosage-corrected Allele-Specific human Transcription factor binding sites (https://adastra.autosome.ru/) – коллекция данных по аллель-специфичному связыванию факторов транскрипции в геноме человека;

– ANANASTRA – ANnotation and enrichment ANAlysis of SNP sets with allele-specific TRAnscription factor binding (https://ananastra.autosome.ru/) – коллекция данных по влиянию SNP на аллель-специфичное связывание

факторов транскрипции в геноме человека;

– версия цистрома для человека и мыши (https://doi.org/10.1186/s13104-018-3856-x).

Также информация из базы данных GTRD была использована при создании широко известных международных баз данных:

– BaMM motif – Bayesian Markov Models (BaMMs) – для распознавания ССТФ (https://bammmotif.soedinglab.org/);

– mSigDB – Molecular Signatures DataBase, C3: regulatory target gene sets (https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp#C3).

Результаты диссертационной работы внедрены в научные исследования и образовательный процесс ряда организаций, включая: Институт аналитического приборостроения РАН, Институт белка РАН, Институт медико-биологических проблем РАН, Федеральный исследовательский центр информационных и вычислительных технологий, Новосибирский государственный университет, Казанский федеральный университет, Научно-технологический университет «Сириус» (имеются акты внедрения).

Апробация работы

Результаты работы были представлены и обсуждены на следующих российских и международных конференциях: “International Conference on Systems Biology” – ICSB-2002 (Стокгольм, Швеция), ICSB-2004 (Гайдельберг, Германия), ICSB-2006 (Йокохама, Япония), ICSB-2007 (Лонг Бич, США), ICSB-2008 (Гетеборг, Швеция), ICSB-2010 (Эдинбург, Великобритания), ICSB-2011 (Гейдельберг, Германия); "Биоинформатика регуляции и структуры генома (BGRS)", г. Новосибирск: BGRS-1998, BGRS-2000, BGRS-2002, BGRS-2004, BGRS-2006, BGRS-2008, BGRS/SB-2010, BGRS/SB-2012, BGRS/SB-2014, BGRS/SB-2016, BGRS/SB-2018, BGRS/SB-2020, BGRS/SB-2022); 7th Noordwijkerhout Symposium on Pharmakokinetics, Pharmacodynamics and Systems Pharmacology (Нордвейкерхаут, Нидерланды, 2013); International conference “High-throughput sequencing in genomes - 2013”, г. Новосибирск; German Conference on Bioinformatics, GCB’2003, Нойхерберг, Германия, 2003; Московская международная конференция “Биотехнологии и медицина”, Москва, 2006; «Молекулярная онкология», Всероссийская конференция с международным участием, Новосибирск, 2008;

Международная Московская конференция по компьютерной молекулярной биологии (MCCMB), г. Москва, в 2005 и 2011 гг.; международной конференции по современным проблемам математики, информатики и биоинформатики, посвященной 100-летию со дня рождения член-корреспондента АН СССР Алексея Андреевича Ляпунова (Новосибирск, 2011); международный семинаре «From virtual cell to virtual human and virtual patient» (Новосибирск, 2012); международной конференции по биомедицинской инженерии и компьютерных технологиях SIBIRCON (Новосибирск, 2015); XVI Российская конференция «Распределенные информационно-вычислительные ресурсы. Наука – цифровой экономике», DICR-2017 (Новосибирск, 2017); NETTAB - 17th International Workshop on Network Tools and Applications in Biology: Methods, Tools and Platforms for Personalized Medicine in the Big Data Era. - (Палермо, Италия, 2017); Workshop on BIOINFORMATICS OF GENE REGULATION. 30 Years TRANSFAC (Геттинген, Германия, 2018); Международный семинар "LIFE OF GENOMES 2018“ (Казань, 2018); XVII Международная конференция «Распределенные информационно-вычислительные ресурсы: Цифровые двойники и большие данные», DICR-2019 (Новосибирск, 2019); Всероссийская мультиконференция с международным участием "Биотехнология – медицине будущего" (Новосибирск, 2019); VII Съезд Вавиловского общества генетиков и селекционеров, посвященный 100-летию кафедры генетики СПбГУ, и ассоциированные симпозиумы (Санкт-Петербург, 2019); Международный конгресс "Биотехнология: состояние и перспективы развития" (Москва, 2019-02-27); X Всероссийская конференция «Новые подходы к изучению проблем физиологии экстремальных состояний» (Москва, 2021); Международный семинар «Сочинский диалог» «Диалоги о геномике: лучшие практики лабораторий РФ и Европы» (Сочи, 2021); IX Международный форум технологического развития «Технопром-2022» (Новосибирск, 2022); VI Сеченовский Международный Биомедицинский Саммит (SIBS-2022, Москва); Школа "Цифровая кардиология" (Москва, 2022); Международная конференция «Прикладная математика иммунологии и вирусологии» (Сочи, 2022); Sensors and their medical applications (Huawei, он-лайн конференция, 2022);VII Всероссийской научно-практической конференция по Искусственному интеллекту в здравоохранении и системам поддержки

принятия врачебных решений ITM-AI (он-лайн конференция, 2023).

Публикации. Материалы диссертационной работы отражены в 75 научных публикациях, включая: 34 публикации в журналах Q1 и Q2 Web of Science/Scopus, 12 публикаций в журналах Q3 и Q4, 3 главы в монографиях. Издано 1 учебное пособие. Получено 7 свидетельств на регистрацию программ и баз данных для ЭВМ.

Личный вклад автора. Автором предложена новая технология для итерационного создания, тестирования и использования сложных модульных моделей биологических систем, лично автором сформулированы соответствующие идеи и подходы**.**

В 2002 г. лично автором была создана первая версия ПК BioUML. В дальнейшем развитии этой платформы под постоянным руководством автора приняло участие более 30 человек. Также автором постоянно осуществлялись разработка и развитие архитектуры ПК BioUML, начиная от настольного приложения и заканчивая облачным веб-приложением. Автором лично написан начальный код ключевых Java библиотек и Java интерфейсов: для доступа к данным (ru.biosoft.access), для визуализации данных (ru.biosoft.graphics), для объектно-ориентированного представления данных (biouml.standard, ru.biosoft.bsa), мета-модель для представления биологических моделей (biouml.model), для расположения вершин графов на плоскости (ru.biosoft.graph) и ряд других.

Под руководством автора была создана база данных GTRD, при этом автором лично сформулированы ключевые идеи и подходы.

Под руководством автора, используя ПК BioUML, были разработаны приведенные в диссертационной работе модульные модели биологических систем**.** Автор также отбирал существующие математические модели, которые стали отправной точкой для построения соответствующих модульных моделей. Автором была разработана технология построения цифрового двойника пациента.

На основе архитектуры и кода ПК BioUML автором предложена архитектура универсальной научной платформы u-science – для графического представления и анализа данных из других предметных областей. Также автором разработана архитектура и ключевые идеи платформы Sirius-web.

Структура и объем диссертации. Диссертационная работа состоит из

введения, 9 глав, заключения и списка литературы содержащего 438 ссылок. Работа изложена на 395 страницах (включая 291 страницы основного текста и 24 приложения), содержит 120 рисунков, 20 таблиц.

СОДЕРЖАНИЕ РАБОТЫ

Глава 1. Обзор литературы

В разделе 1.1 рассмотрено понятие системной биологии и соответствующие экспериментальные методы – высокопроизводительные автоматизированные потоковые технологии для сбора большого количества численных данных (high throughput technologies, omics технологии). Отдельное внимание уделено воспроизводимости моделей биологических систем и используемым для этого международным стандартам (SBML, SBGN, SED-ML).

В разделе 1.2 рассмотрены основные методы и подходы моделирования биологических систем (математические методы, визуальное моделирование, модульное моделирование, агентное моделирование), а также рассмотрена концепция цифровых двойников и наиболее крупные, модели (или проекты, в рамках которых они были разработаны).

В разделе 1.3 приведено описание основных программных комплексов для моделирования биологических систем, с акцентом на те, которые поддерживают модульное моделирование (ProMoT, JigCell, iBioSim, RoadRunner, Systems Biology Simulation Core Library, TinkerCell, M2SL).

Раздел 1.4 посвящен методам и сценариям анализа генетических данных. Рассматриваются реестры методов и сценариев анализа генетических данных (Biocontaners, Dockstore, Biotools); языки для задания сценариев анализа генетических данных (CWL, WDL, Nexflow, Galaxy) и платформы для анализа генетических данных (R/Bioconductor, Jupyter, SoS, Galaxy, Taverna, Unigene, Nexflow).

В разделе 1.5. рассматривается графическое представление генетических данных при помощи геномных браузеров, визуализация метаболических путей и путей передачи сигнала в клетке, множественных выравниваний и филогенетических деревьев, а также 3-х мерных структур макромолекул.

В разделе 1.6. рассмотрены компьютерные программы, разработанные автором в 1994–2000 г.г. во время работы в лаборатории теоретической

генетики Института цитологии и генетики СО РАН под руководством академика РАН, профессора, д.б.н. Колчанова Н.А. и послужившие заделом для разработки ПК BioUML. Работа была начата в 1994 г. с разработки программы MGL для формального описания и графического представления геномных регуляторных последовательностей (Колпаков и Бабенко, 1996). В 1996-1997 г. основные классы были переведены с языка C++ на язык Java и были созданы MGL сервер и клиент (апплет TRRD Viewer) для визуализации регуляторных районов генов из базы данных TRRD (Heinemeyer et al., 1997, Kolchanov et al., 1999). В 1997 г. были разработаны новые принципы формального описания и графического представления генных сетей, реализованные в системе GeneNet (Kolpakov et al., 1998, Kolpakov and Ananko, 1999).

Глава 2. ПК BioUML

ПК BioUML разрабатывается с 2001 года по настоящее время под руководством автора в ООО “Биософт.Ру» и Федеральном исследовательском центре информационных и вычислительных технологий. Ниже приведены данные, характеризующие сложность ПК BioUML:

- период разработки 2001 – 2023 годы;
- затрачено на разработку 120+ человеко-лет;
- объем кода (не включая сторонние библиотеки):
- Java – 6 199 файлов, объем 36.8 Мб, ~ 1.2 млн. строк кода
- JavaScript – 74 файла, 1.5 Мб, ~ 50 000 строк кода.

ПК BioUML доступен для пользователя в 2-х вариантах:

1. BioUML workbench – Java приложение, которое устанавливается на компьютер пользователя;
2. BioUML сервер – в этом случае ПК BioUML устанавливается на сервере или в "облаке", а пользователь может использовать веб-браузер (рис. 2.1).

Архитектура ПК BioUML представлена на рис. 2.2. Более подробно она описана в работе Kolpakov et al., 2022.

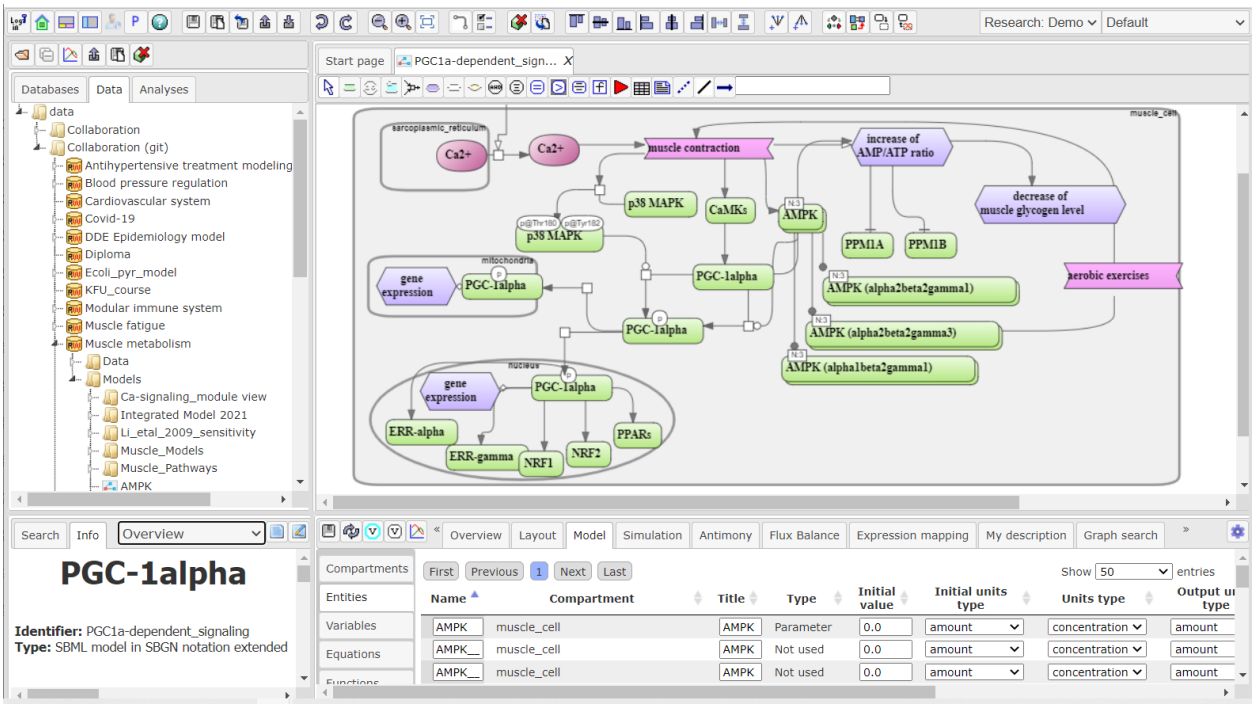

Рисунок 2.1. Интерфейс ПК BioUML: верхняя левая панель – репозиторий проектов; верхняя правая панель – документ для редактирования, здесь фрагмент диаграммы регуляции экспрессии гена PGC-1alpha в графической нотации SBGN; левая нижняя панель – инспектор свойств для просмотра и редактирования.

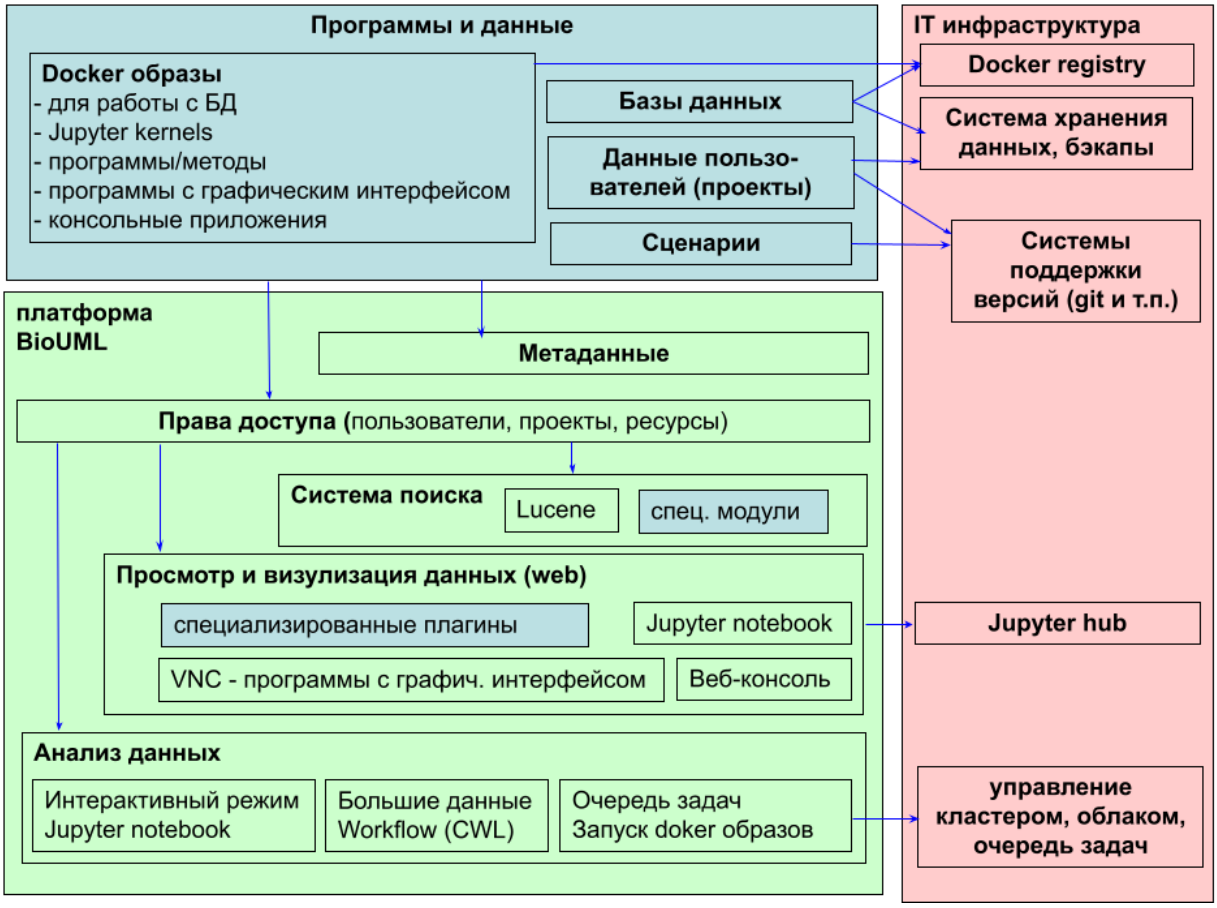

Рисунок 2.2. Архитектура ПК BioUML.

Глава 3. Технология создания и использования модульных моделей биологических систем

Мета-модель («модель модели» биологической системы) обеспечивает формализм для комплексного описания, графического представления и численного моделирования широкого круга биологических систем. Она включает три взаимосвязанных уровня описания (рис. 3.1):

1) граф – структура системы описывается как компартментализованный граф;

2) семантика – на этом уровне описывается биологическая (или другая) семантика каждого объекта. Для этого используется объектно-ориентированное представление биологических данных и приводятся ссылки на соответствующие записи в базах данных;

3) математическая модель – переменные, формулы, уравнения, события, состояния и переходы. Например, вершины графа могут быть переменными или состояниями системы, а ребра графа могут быть уравнениями, описывающими изменения этих переменных или условными переходами между состояниями.

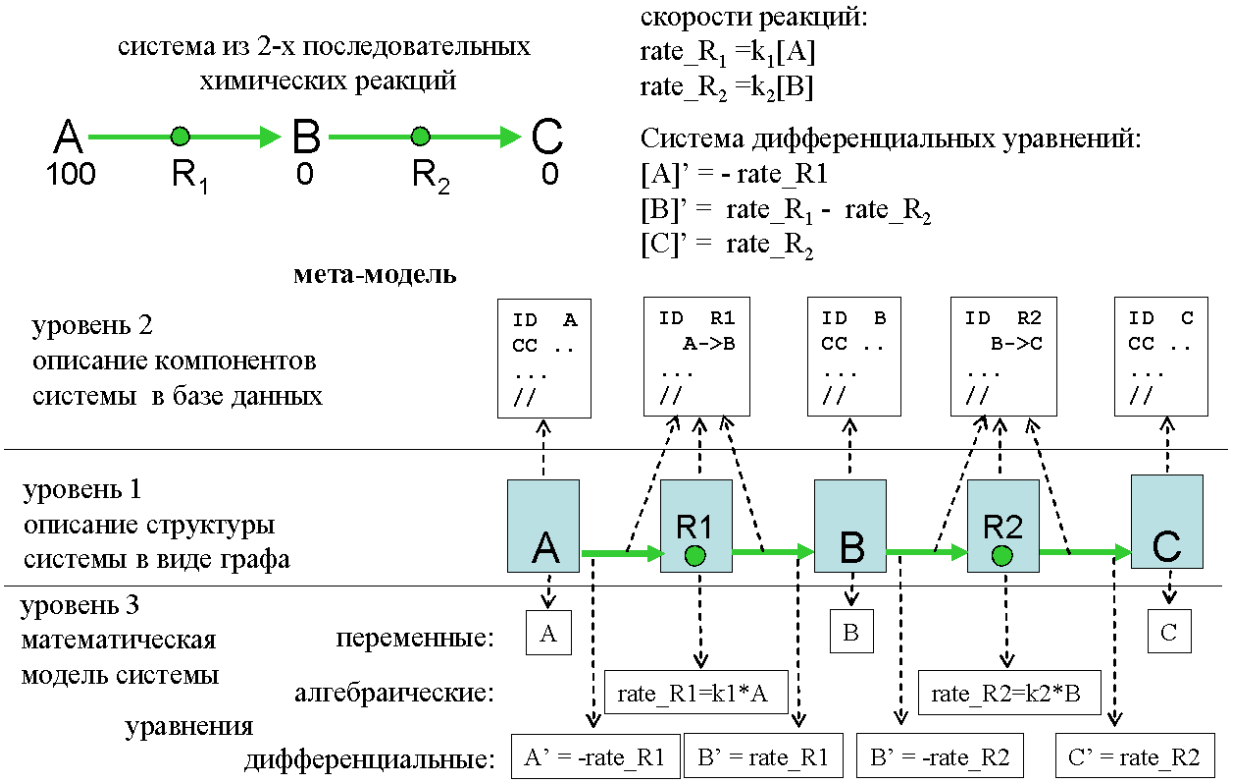

Рисунок 3.1. Пример отображения простой системы, состоящей из двух последовательных биохимических реакций и ее представление по уровням мета-модели.

Визуальное моделирование – упрощает создание сложных биологических моделей:

- пользователь создаёт и редактирует модель биологической системы в виде диаграммы;
- на основе созданной диаграммы ПК BioUML генерирует Java код;
- сгенерированный код компилируется и используется соответствующими решателями для численного моделирования динамики исходной системы.

Модульное моделирование – модель всей системы строится как комбинация моделей подсистем (модулей). Взаимодействие блоков моделируется связями между переменными модулей, которые указывают на пути передачи сигналов между ними. Модули представляют собой математические модели различных подсистем, также они могут содержать вложенные модульные модели, составляя иерархическую структуру.

Ключевой аспект при создании модульных моделей - четкая спецификация интерфейса модуля для взаимодействия с другими частями модели. В ПК BioUML для этого используется понятие портов (вход-выход и контакт-контакт для разделяемых переменных).

Агентное моделирование – является дальнейшим развитием идеи модульного моделирования в ПК BioUML и использует ту же графическую нотацию. Основное отличие от модульной модели – каждый модуль может иметь собственный решатель (simulation engine) и шаг по времени. Таким образом, в рамках агентной модели можно объединить как разные подходы к моделированию динамики биологических систем (например, системы дифференциальных уравнений и стохастическое моделирование), так и модули с существенно различным шагом по времени.

Текстовое представление модели, его синхронизация с графическим представлением – одна и та же модель представляется в текстовом и графическом виде. В качестве декларативного текстового представления был выбран язык Antimony (https://tellurium.readthedocs.io/en/latest/antimony.html), который позволяет представить в удобном человеко-читаемом виде SBML модели. Далее в этот язык был добавлен механизм аннотаций, при помощи которых для каждого элемента модели можно задавать дополнительные свойства, необходимые для его графического представления в расширенной графической нотации SBGN Process Diagrams.

Автоматическая генерация кода – происходит на основе формального описания биологической системы в виде разработанной мета-модели. Генерируемый код зависит от численного метода и типа диаграммы, используемой для дальнейших численных расчетов.

Для генерации каждого Java кода для каждого типа модели используются соответствующие шаблоны на языке Velocity (https://velocity.apache.org/). Стоит особо подчеркнуть, что для очень больших SBML моделей (10 000+ компонентов и реакций) разработан специальный шаблон. Это связано с тем, что Java ограничивает размер (64 кБ) выполняемой функции и объявлений переменных.

Численное моделирование – в ПК BioUML реализовано 6 основных типов решателей для численного моделирования динамики биологических систем:

1) алгебро-дифференциальные уравнения с дискретными событиями и задержкой по времени;
2) стохастические дифференциальные уравнения;
3) стохастические модели;
4) анализ стационарных потоков (Flux Balance Analysis, или FBA);
5) моделирования кровотока в основных артериях и артериолах;
6) агентное моделирование.

Определение параметров моделей на основе исходных экспериментальных данных осуществляется путем решения обратной задачи. Для этого реализовано 6 различных алгоритмов (Кутумова, 2012). В ПК BioUML реализована возможность одновременно оптимизировать параметры для нескольких экспериментов. Для этого вводятся типы параметров: *глобальный* – значение параметра одно, и это значение оптимизируется по всем экспериментам одновременно и *локальный* – значение параметра оптимизируется индивидуально для каждого эксперимента. Также в ПК BioUML реализован численный анализ идентифицируемости параметров (Raue et al., 2010).

Использование практик из разработки программного обеспечения, а именно методология экстремального программирования (XP, eXtreme Programming) была адаптирована для разработки моделей сложных биологических систем (Колпаков и соавт., 2022).

Автоматическое тестирование моделей – в соответствии с предложенной

технологией для каждой модели должен создаваться набор тестов, которые осуществляют проверку значений переменных при симулировании на соответствие заданным интервалам и экспериментальным данным.

Создание математических моделей сложных биологических систем эволюционным путем – является наиболее эффективным подходом. Например, самая современная модульная модель сердечно-сосудистой и почечной систем человека для регуляции артериального давления (Kutumova, 2021) все еще содержит часть исходной модели профессора Гайтона для общей регуляции кровообращения (Guyton, 1972), которой уже более 50 лет.

Глава 4. Анализ и графическое представление геномных данных

ПК BioUML содержит большой набор функций для графического представления и анализа омиксных данных, в первую очередь геномных: работа с биологическими базами данных (Ensembl, GeneOntology и другие), геномный браузер, методы, программы и сценарии анализа данных.

Разработана библиотека Java классов для объектно-ориентированного представления геномных данных. Информация из различных баз данных (Ensembl, GTRD и другие) и файлов с геномными данными в форматах bam, bed/bigBed, wig/bigWig, vcf, gff и др. может быть отображена в виде набора объектов соответствующих классов.

Для графического представления геномных данных разработан новый геномный браузер (рис. 4.1, Valeev et al., 2013), который по функционалу сопоставим с ведущими геномными браузерами (Ensembl genome browser, UCSC genome browser, JBrowse, Integrative Genomics Viewer, Integrated Genome Browser, NCBI Genome Data Viewer), но имеет ряд уникальных возможностей: интерактивное объединение и пересечение треков с возможностью комбинации их отображения.

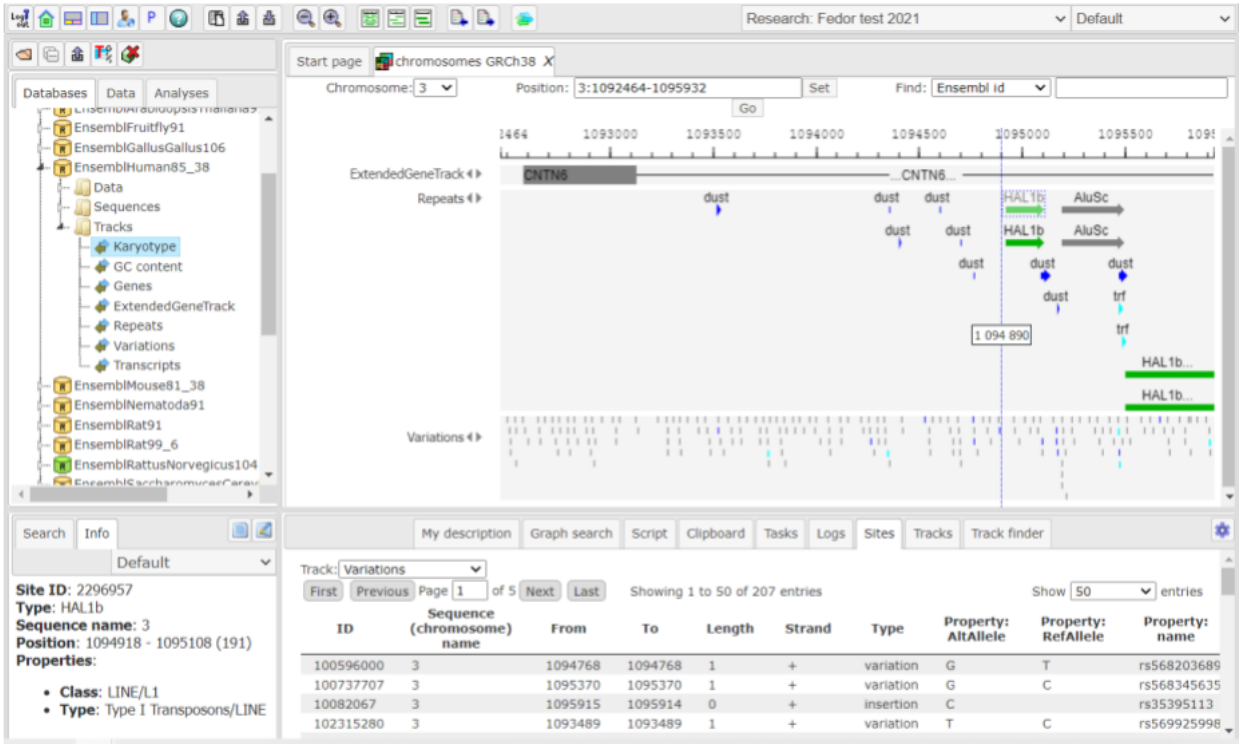

Рисунок 4.1. Представление треков из базы данных Ensembl в геномном браузере.

Разработана библиотека Java классов для объектно-ориентированного представления методов анализа данных и их параметров в виде объектов JavaBeans (рис. 4.2, сверху, справа). Информация из файлов с методами и сценариями анализа данных в форматах CWL, WDL и Nexftlow может быть отображена в виде набора объектов соответствующих классов. Используя технологию BeanExplorer автоматически генерируются формы для задания значений параметров для соответствующих методов анализа данных.

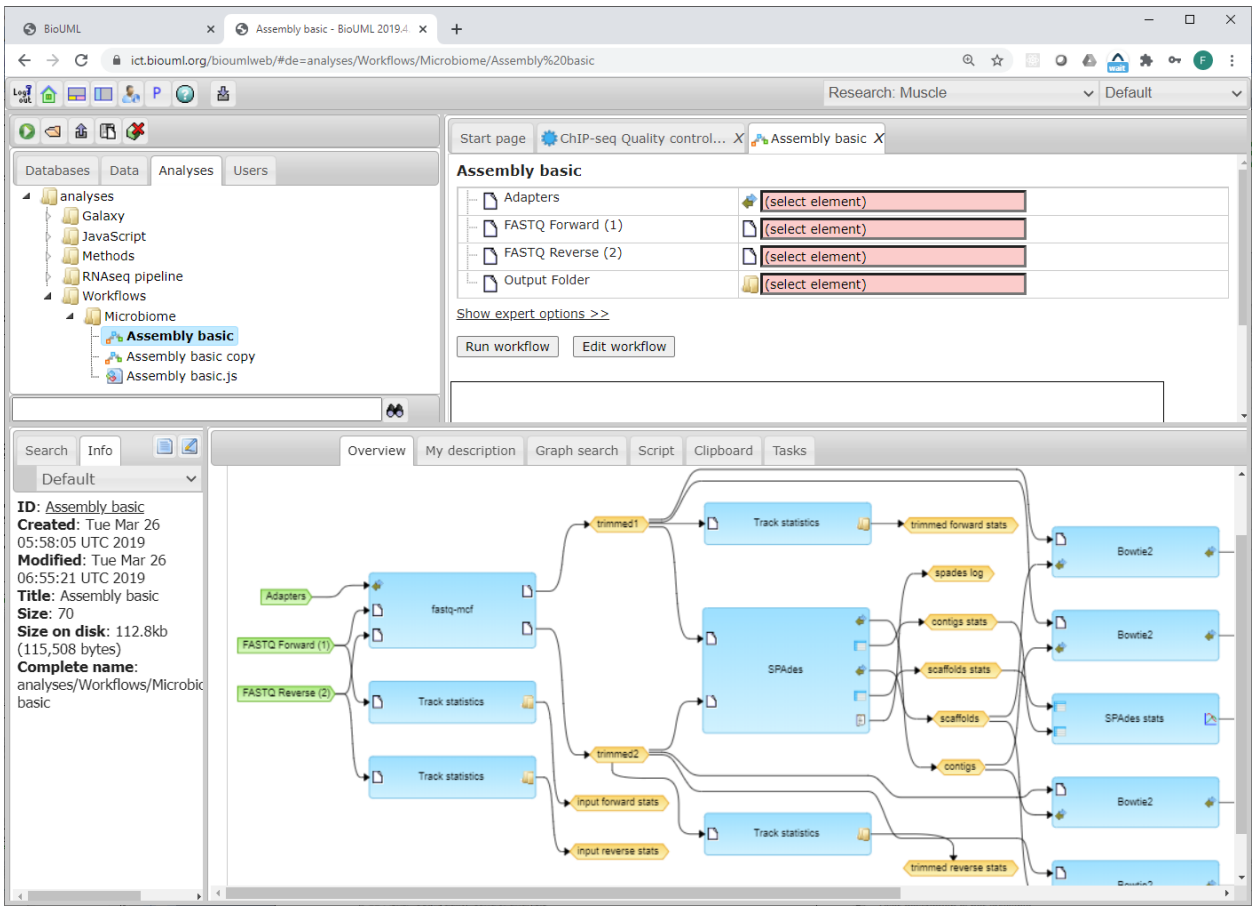

Рисунок 4.2. Графическое представление сценария анализа данных (снизу, справа) и форма для задания его параметров для запуска (сверху, справа).

Глава 5. База данных GTRD

В настоящее время в базе данных GTRD накапливаются данные для 10 видов организмов: *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Gallus gallus*, *Danio rerio*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* и *Arabidopsis thaliana*.

Текущий рабочий процесс для единообразной аннотации и анализа NGS данных по регуляции транскрипции в базе данных GTRD представлен на рис. 5.1. Статистика по объему аннотированных NGS данных для версии 23.07 приведена в таблице 5.1.

Используя сопоставленные клеточные типы GTRD и баз данных Ensembl, ENCODE, FANTOM 5, и GTEx, были сопоставлены различные типы экспериментов из этих баз данных, но проведенные на одних и тех же клеточных линиях. Детальное описание работ по этому пункту приведено в публикациях Kulyashov et al., 2020a, b.

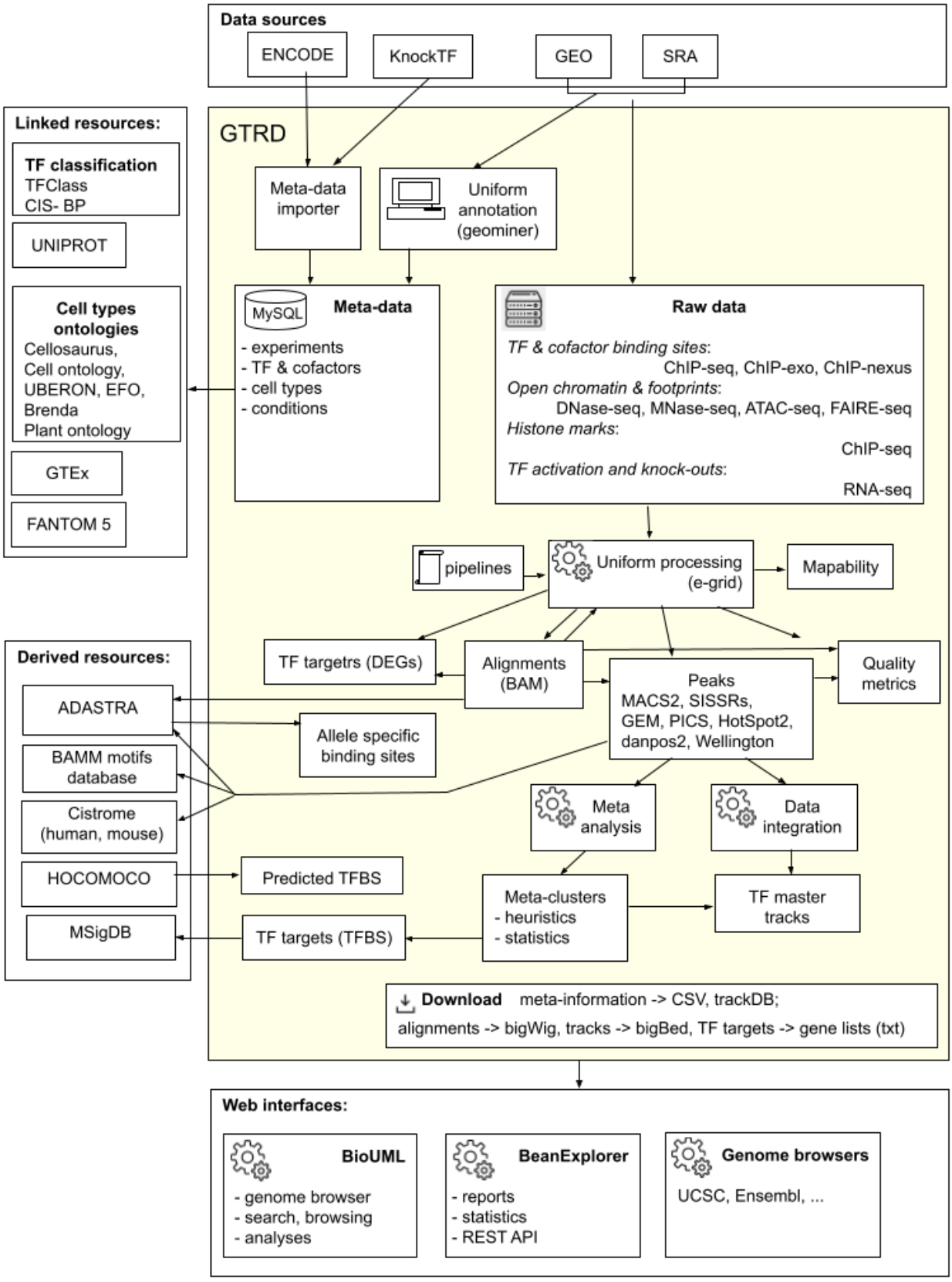

Рисунок 5.1. Текущий рабочий процесс для единообразной аннотации и анализа NGS данных в базе данных GTRD (Kolmykov et al., 2021).

Таблица 5.1. Статистика базы данных GTRD (версия 23.07) по объему аннотированных NGS данных по регуляции транскрипции

	Транскрипционные факторы и кофакторы		Гистоновые модификации ChIP-seq	Открытый хроматин				Метили-рование	Экспрессия при модификации транскрипционных факторов	
	ChIP-seq	ChIP-exo		DNase-seq	ATAC-seq	MNase-seq	FAIRE-seq	WGBS	RNA-seq	CAGE-seq
Mus musculus	21 456	134	16 284	928	15 368	607	144	85	5 326	1 018
Homo sapiens	29 797	2 021	22 987	1 898	14 190	247	346	117	3 664	1 816
Rattus norvegicus	469	2	680	12	165	4	0	0	20	55
Gallus gallus	107	0	259	15	116	0	1	0	0	119
Drosophila melanogaster	3 680	51	2 056	47	767	331	90	0	108	0
Caenorhabditis elegans	2 058	0	868	94	127	46	0	0	0	0
Arabidopsis thaliana	1 106	0	2 162	14	285	143	22	1	12	0
Saccharomyces cerevisiae		1 120	1 862	9	192	1 358	102	0	8	0
Danio rerio	122	0	597	0	433	7	4	0	17	0
Schizosaccharomyces pombe	670	13	919	0	6	174	0		4	0
Всего (версия 23.07)	**62 947**	**3 341**	**48 674**	**3 017**	**31 649**	**2 944**	**709**	**203**	**9 159**	**3 008**

Начиная с первой публично-доступной версии GTRD (Yevshin et al., 2017) мы разрабатываем и совершенствуем методы, чтобы интегрировать данные о пиках из разных ChIP-seq экспериментов для заданного ТФ, обработанных разными алгоритмами поиска РСТФ. В последующем, для более достоверного выявления РСТФ на основе мета-анализа результатов ChIP-seq анализа, был разработан новый метод METARA – METa Analysis of ChIP-seq datasets through the RA approach (рис. 5.5.1, Kolmykov et al., 2019, 2020, 2021). Этот метод представляет собой трех-этапное применение подхода коллективного выбора (Rank Aggregation Approach, RA-approach).

На основе полученных мета-кластеров РСТФ для 3 клеточных линий (HepG2, K562 и HEK293 – эти клеточные линии наиболее полно исследованы с помощью ChIP-Seq экспериментов) методом пошаговой линейной регрессии построена модель регуляции транскрипции, которая позволяет предсказать уровень экспрессии соответствующих стартов инициации транскрипции из базы данных FANTOM5 (Sharipov et al, 2020).

Существенная часть работы, описанной в этой главе, была выполнена под руководством автора в рамках гранта РНФ № 19-14-00295 "Интеграция и анализ омиксных данных по регуляции транскрипции, оценка влияния SNV".

Глава 6. Моделирование сложных биологических систем

В главе 6 продемонстрированы возможности разработанной технологии и ПК BioUML для создания и использования сложных моделей биологических систем.

Модульная модель апоптоза была первой моделью, которая нами разработана, используя описанную выше технологию. Она включает: 13 модулей, 286 белков и их комплексов, 684 реакции и 719 параметров (рис. 6.1.). При построении данной модели были разработаны новые подмодели для сигнальных путей CD95 и NF-κB. Более подробно данная модель описана в работах Kutumova et al., 2012; Kutumova et al., 2013; Кутумова и соавт., 2012.

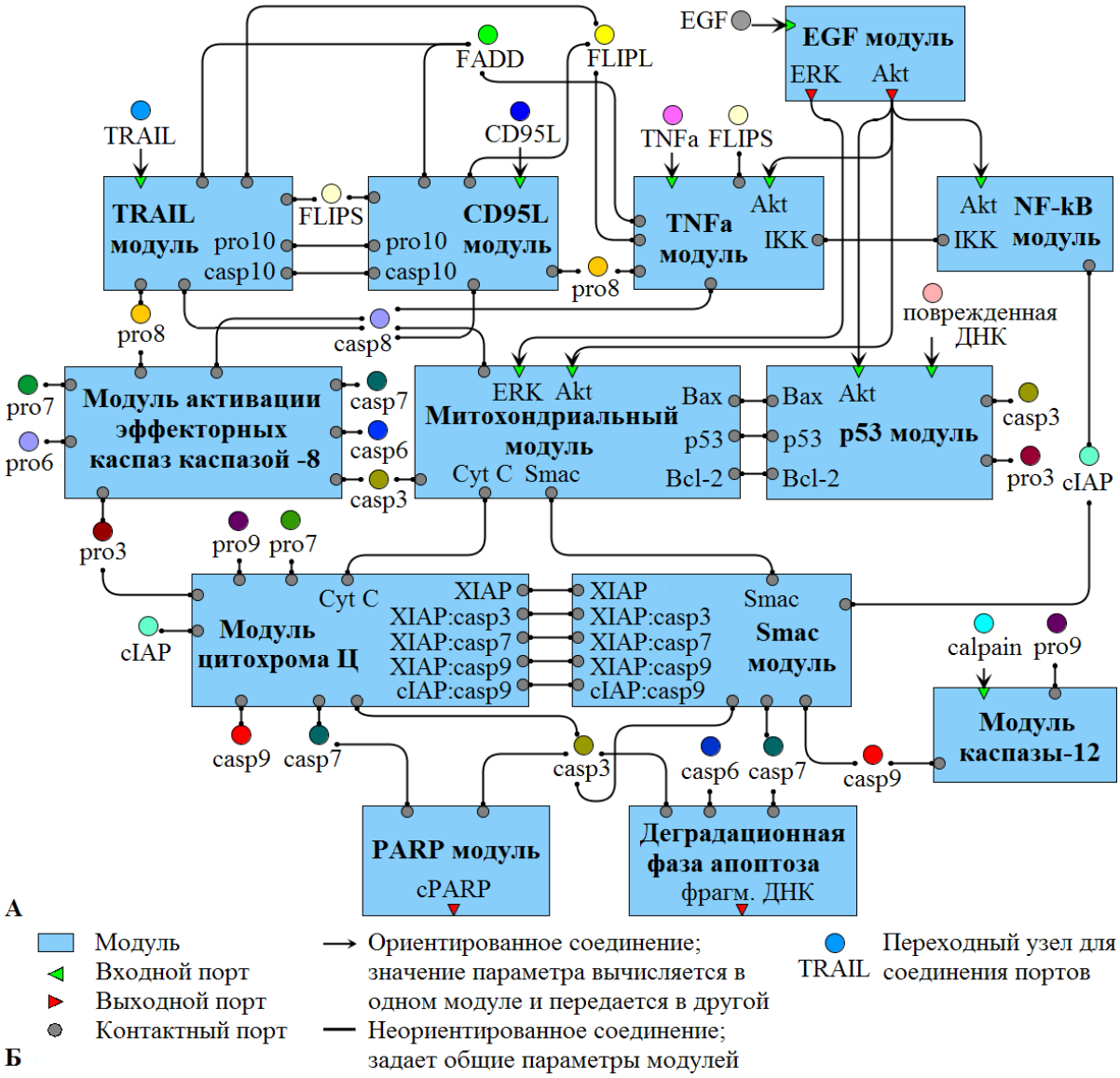

Рисунок 6.1. Модульная модель апоптоза (Kutumova et al., 2012). А. Модель, состоящая из 13 функциональных модулей. Б. Графическая нотация представления модульной модели в ПК BioUML.

Модульная модель регуляции генной экспрессии при физической нагрузке в скелетных мышцах (Киселев и соавт., 2019; Акбердин и соавт, 2020; Akberdin et al. 2021) состоит из модулей, упорядоченных по 2 иерархическим уровням:

1) <u>физиологический уровень</u> – на этом уровне моделируется кровоток и транспорт метаболитов в/из мышцы. Для более реалистичного описания механизмов функционирования мышечной ткани при разных типах нагрузок учтено функционирование мышечных волокон двух типов (быстрые и медленные), имеющих различные паттерны активации в ответ на физическую

нагрузку в зависимости от ее интенсивности.

2) клеточный уровень – на этом уровне моделируется работа одной мышечной клетки, при этом данный уровень разделен на подуровни: метаболизм, передача сигнала в клетке и регуляция генной экспрессии.

Модели эпидемиологического процесса COVID-19 были реализованы в рамках гранта № РФФИ 20-04-60355 "Разработка мультимасштабной иммуно-эпидемиологической математической модели COVID-19 с учетом воздействия на экономику региона и сценариев действия органов власти" под руководством автора по 3-м направлениям:

– эпидемиологическая модель на основе систем дифференциальных уравнений с задержкой (Kiselev et al., 2021) на основе модели (Westerhoff and Kolodkin, 2020;

– агентная модель эпидемиологического процесса на основе существовавшей ранее агентной модели распространения коронавируса REINA (Tuomisto et al., 2020);

– модель инфекционного процесса в организме человека (Afonyushkin et al., 2022).

Модели регуляции артериального давления у человека – в данной разделе кратко рассмотрены 5 моделей:

1. модель Гайтона (Guyton et al., 1972), которая была создана в модульном виде (18 модулей, 229 параметров, 133 операций присваивания (ОП), 39 ОДУ; Киселев, 2016), используя ПК BioUML;

2. Резервуарная модель была предложена Ю. В. Солодянниковым в 1994 году (Солодяников, 1994) и в дальнейшем была расширена (Proshin A.P., Solodyannikov Y.V., 2006), чтобы учесть большой и малый круг кровообращения. Состоит из 6 модулей, 58 переменных, 7 ОДУ, 14 ОП и 2 события для мгновенного перехода между состояниями – систолы и диастолы.

3. Долговременная модель почечной регуляции, предложенная Карааслaном и соавторами (Karaaslan et al., 2004), использующая идеи своих предшественников – моделей Гайтона, Уттамшинга и Колмана-Холла – детально описывает механизмы почечной регуляции кровотока. Включает 6 модулей, 85 переменных, 11 ОДУ, 61 ОП.

4. Гидродинамическая модель артериальной системы, полученная из

уравнений Навье-Стокса посредством интегрирования их по поперечному сечению сосуда (Lamponi, 2004). Данная модель описывает кровоток в артериальном дереве, состоящем из 55 наиболее крупных артерий человека и позволяет получить значения давления, потока крови и других параметров в каждой точке каждой артерии. Для поиска численных решений ГДМ предложен оригинальный алгоритм упаковки артериального дерева и использован метод ортогональной прогонки (Бибердорф и соавт., 2015).

5. Модульная агентная модель регуляции артериального давления человека (Kutumova et al., 2021), которая является объединением и развитием моделей 1, 3 и 4. По мнению рецензента данной статьи "Никто со времен Гайтона и его команды в Миссисипи в 1960-70-х годах не совершил такого качественного скачка в области моделирования сердечно-сосудистой регуляции кровяного давления на уровне всего организма"[1]. На рис. 6.2 представлена расширенная версия данной модели, в ней добавлен модуль «Фармакодинамика», отвечающий за терапевтический эффект антигипертензивных препаратов (см. главу 7). Зеленым отмечены модули, взятые из расширенной модели Гайтона – Карааслана, синим – из модели Солодянникова. Красными стрелками отмечены точки воздействия антигипертензивных препаратов.

Была выполнена калибровка модели, чтобы найти состояние равновесия в пределах нормальных диапазонов показателей жизнедеятельности для здорового взрослого человека. Мы проверили возможности модели воспроизводить равновесные состояния с аномальными физиологическими значениями, связанными с различными сочетаниями сердечно-сосудистых заболеваний (таких как системная гипертензия, хроническая сердечная недостаточность, легочная гипертензия и др.). Для создания и валидации модели были использованы данные из более 200 научных исследований, охватывающих известные модели сердечно-сосудистой и почечной функций человека, платформы биомоделирования и клинические измерения физиологических величин в норме и при патологии (Kutumova et al., 2021).

[1] "No-one since Guyton and his team in Mississippi in the 1960-70s has made such quantum leap in the field of modelling whole body cardiovascular regulation of blood pressure and its control."

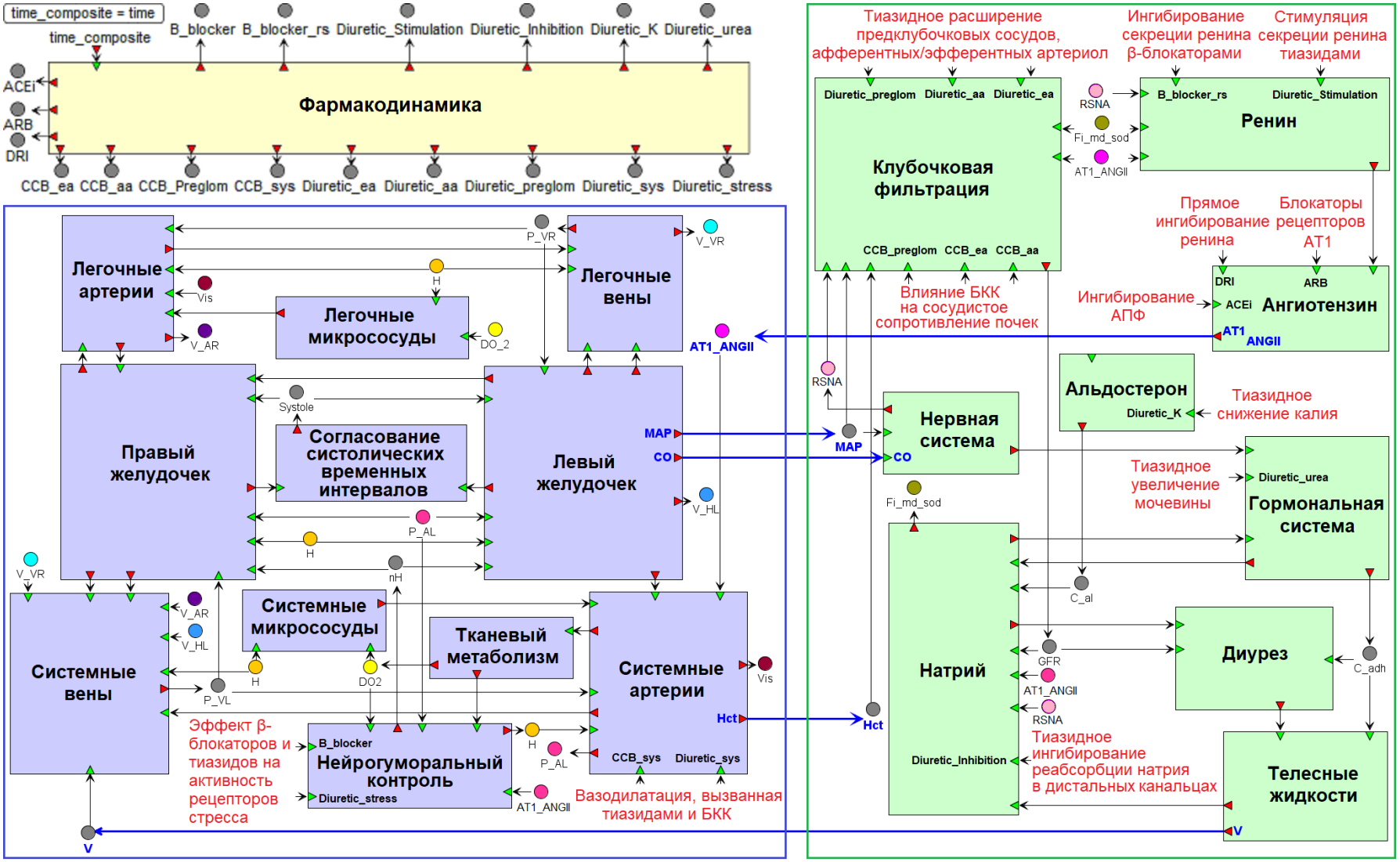

Рисунок 6.2. Комплексная модель регуляции артериального давления человека, реализованная в платформе BioUML. Зеленым отмечены модули, взятые из расширенной модели Гайтона – Карааслана, синим – из модели Солодянникова. Красными стрелками отмечены точки воздействия антигипертензивных препаратов.

Глава 7. Цифровой двойник пациента

Впервые концепция цифрового двойника была предложена профессором Мичиганского университета Майклом Гривсом (Michael Grieves) еще в 2002 г. Цифровой двойник содержит три основные части (Grieves, 2014). В применении к цифровому двойнику пациента это выглядит следующим образом:

1. физический продукт в реальном пространстве – реальный пациент;
2. виртуальный продукт в виртуальном пространстве – виртуальный пациент – математическая модель физиологических и биохимических процессов в организме пациента;
3. данные и информацию, которые объединяют виртуальный и физический продукт – персонализация математической модели на основе данных электронной медицинской карты, генетических данных, опросников, носимых мобильных устройств и т.п.

Таким образом, "виртуальный пациент" – это «образ» реального пациента, который представлен математической моделью физиологических и биохимических процессов в организме человека, причем параметры этой модели подобраны так, чтобы максимально соответствовать наблюдаемым клиническим данным для конкретного реального пациента (Lehrach 2016).

В зависимости от заболевания, патофизиологических механизмов и применяемой терапии виртуальные модели пациентов могут быть реализованы как на довольно высоком уровне абстракции, так и включая модели, представляющие релевантные клетки, типы клеток или органы, которые могут взаимодействовать не только друг с другом, но и с окружающей средой. Необходимо отметить, что пока невозможно построить единую математическую модель для всех заболеваний. Для каждой группы болезней необходима своя математическая модель виртуального пациента.

Нами предложена технология построения виртуального пациента и оптимизации выбора лекарственной терапии на примере лечения артериальной гипертензии. Это сложное мультифакторное заболевание, для лечения которого используются разные классы лекарственных препаратов, и часто врач пробует несколько схем лечения, чтобы найти оптимальный вариант для конкретного пациента.

Предложенная автором технология состоит из 4 шагов:

1) **построение модульной математической модели биохимии и физиологии** человека с достаточным уровнем детализации для заданной болезни. Для этого предложен следующий подход – создать набор основных блоков, а из них собирать модель под заданного пациента и болезнь (как из блоков конструктора "Лего"). Созданная для этого математическая модель регуляции артериального давления у человека была описана в предыдущей главе (рис. 6.2).

2) **моделирование фармакодинамики лекарственных препаратов** – для основных классов антигипертензивных препаратов (АГП, таблица 6.1): были определены их точки воздействия на построенную модель регуляции артериального давления (рис. 5.2) и были построены соответствующие модели фармакодинамики. Для валидации полученной модели были использованы данные клинических исследований, найденные в литературе. Для этого была сгенерирована популяция виртуальных пациентов с повышенным артериальным давлением и смоделировано их лечение АГП (рис. 6.1). Более подробно этот подход описан в работах Кутумовой и соавт. 2019 и Kutumova et al., 2022.

3) **персонализация модели** – задание параметров модели для заданного пациента – наиболее сложная задача. Часть параметров модели могут быть идентифицированы следующим образом:

- взяты из истории болезни пациента;
- путем расчета ряда физиологических параметров на основе известных зависимостей от роста, веса, пола и возраста пациента;
- решение обратной задачи, используя данные из истории болезни заданного пациента. Однако, при решении обратной задачи только меньшая часть параметров является идентифицируемыми.

Таким образом, большая часть параметров модели не может быть однозначно идентифицирована для заданного пациента. Чтобы преодолеть эту проблему был предложен следующий подход:

- отбор неидентифицируемых параметров, наиболее значимо влияющие на артериальное давление;
- генерация популяции виртуальных пациентов (200-500), так чтобы значения неидентифицируемых параметров могли существенно варьировать в заданных физиологических границах, и наблюдаемые параметры

соответствовали данным из истории болезни.

4) **“Лечение” созданной популяции виртуальных пациентов**, т.е. моделирование воздействия АГП. При этом каждый виртуальный пациент реагирует на “лечение” по-своему, не для всех оно будет эффективно, что соответствует реальной клинической практике. Это позволяет выделить группы виртуальных пациентов со схожей реакцией на лекарственный препарат и определить какие именно параметры определяют разделение на эти группы. В частности, такой анализ может позволить установить какие еще лабораторные исследования нужно провести для данного пациента, чтобы понять к какой группе он относится и какая лекарственная терапия будет для него наиболее эффективной.

Таблица 7.1. Мишени воздействия АГП на модель регуляции артериального давления

АГП (Класс АГП)	**Модуль**	**Точки воздействия**	**Коэффициент воздействия**	**Доза (мг/день)**
Алискирен (Ингибиторы ренина)	Ангиотензин	*PRA* – активность ренина	$\frac{C_c}{IC_{50}+C_c}$	150
				300
Лозартан (Блокаторы AT_1-рецепторов ангиотензина II)	Ангиотензин	c_{AT1} – скорость связывания AT_1-рецепторов с ангиотензином II	0.85	50
			0.94	100
Бисопролол (β-адреноблокаторы)	Нервногуморальный контроль	*rS* – рецептор стресса	0.98	10
Эналаприл (Ингибиторы АПФ)	Ангиотензин	c_{ACE} – скорость расщепления ангиотензина I до ангиотензина II	0.97	20
Амлодипин (Блокаторы кальциевых каналов)	Клубочковая фильтрация	R_{aa} – сопротивление афферентных артериол	0.56	5
		R_{ea} – сопротивление эфферентных артериол	0.26	
		$R_{preglom}$ – сопротивление предклубочковых артерий	0.56	
	Системные артерии	F_{ALVL} – поток крови из артерий в вены большого круга кровообращения	0.26	

С применением предложенного нами подхода было проведено моделирование лечения артериальной гипертензии для шести реальных пациентов. Предложенный алгоритм позволил получить достаточно точный

прогноз лечения выбранным АГП для конкретного пациента. В таблице 6.2 приведен прогноз лечения АГП для одного из реальных пациентов, у которого исходное давление было 152/100, а после лечения лозартаном в дозировке 50 мг снизилось до 132/94. Прогноз на виртуальной популяции дает снижение среднего давления до 124/94.

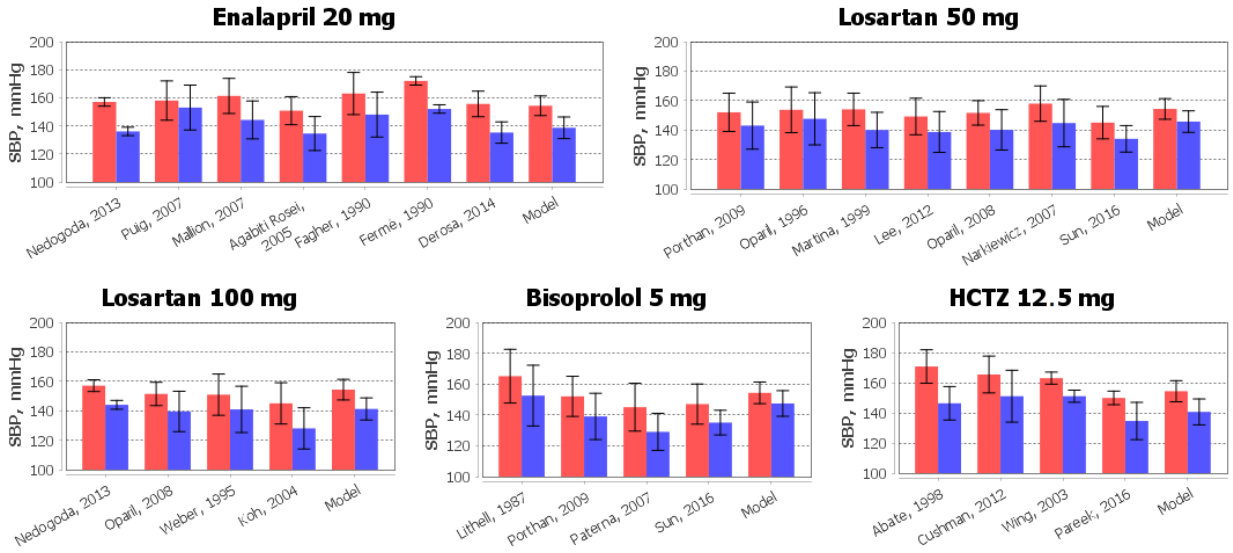

Рисунок 7.2. Сравнение данных моделирования с данными клинических исследований (Kutumova et al., 2022).

Таблица 7.2. Прогноз лечения АГП для одного из реальных пациентов (размер популяции виртуальных пациентов – 200)

Препарат	**Систолическое давление**		**Диастолическое давление**		**Слабый эффект (<10 мм рт. ст.)**
	До лечения	**После лечения**	**До лечения**	**После лечения**	
Алискирен, 150 мг	150.1 ± 2.9	125.5 ± 9.9	102.2 ± 2.9	89.7 ± 5.2	13%
Лозартан, 50мг		**123.5 ± 5.8**		**93.6 ± 3.1**	**18%**
Бисопролол, 10мг		129.4 ± 3.9		83.1 ± 2.3	27.1%
Эналаприл, 20мг		135.3±9.5		93.7±3.5	33.6%
Амлодипин, 5мг		139±4.68		87±3.52	38.6%
Гидрохлортиазид, 25 мг		139±6.8		90.8±6.7	72%

Практическим результатом проведенной нами работы является разработка с применением ПК BioUML нового метода, который реализован следующим образом: на вход подаются имеющиеся клинические данные пациента, после этого программа генерирует множество виртуальных пациентов, для которых она предсказывает наиболее вероятный эффект их лечения разными лекарственными препаратами и устанавливает какие неизвестные персональные параметры нужно определить, чтобы сделать более точный

выбор лекарственного препарата. Однако для внедрения этой программы в медицинскую практику нужно пройти еще большой путь – апробировать ее на большом количестве пациентов и пройти процесс сертификации.

Глава 8. Другие разработки на основе ПК BioUML

Модульная архитектура ПК BioUML позволила создать на его основе ряд различных программ для моделирования, анализа и графического представления данных, не только в области биологии и медицины, но и в других предметных областях. В результате на основе ПК BioUML сейчас разрабатывается универсальная научная платформа u-science.

Компоненты прототипа «Национальной базы генетической информации". На основе ПК BioUML созданы 2 подсистемы:

– Геномный браузер – обеспечивает представление генетической информации и связанных треков метаинформации (структурная аннотация, функциональная аннотация и др.) в графическом интерактивном режиме;

– Рабочие окружения – содержит сценарии и программы для решения следующих задач: сборка и аннотирования первичных последовательностей, поиск и выравнивание последовательностей, анализ ДНК и РНК, анализ транскриптомных данных, анализ экспрессии генов, функциональный анализ, филогенетический анализ, поиск и анализ вариаций геномов.

Биоинформатическая платформа для хранения, анализа и графического представления данных, полученных при одномолекулярном секвенировании ДНК как часть проекта «Разработка опытного образца одномолекулярного секвенатора ДНК, набора реагентов и расходных материалов для него».

Платформа Sirius-web (https://sirius-web.org/) – это информационная платформа для проектов, связанных с анализом данных и моделированием для образовательных и научных проектов. Однако, задачи, поставленные при ее разработке, делают ее полезной для использования в других научных и научно-образовательных организациях, связанных с информационными технологиями. Архитектура платформы Sirius-web представлена на рис. 8.1.

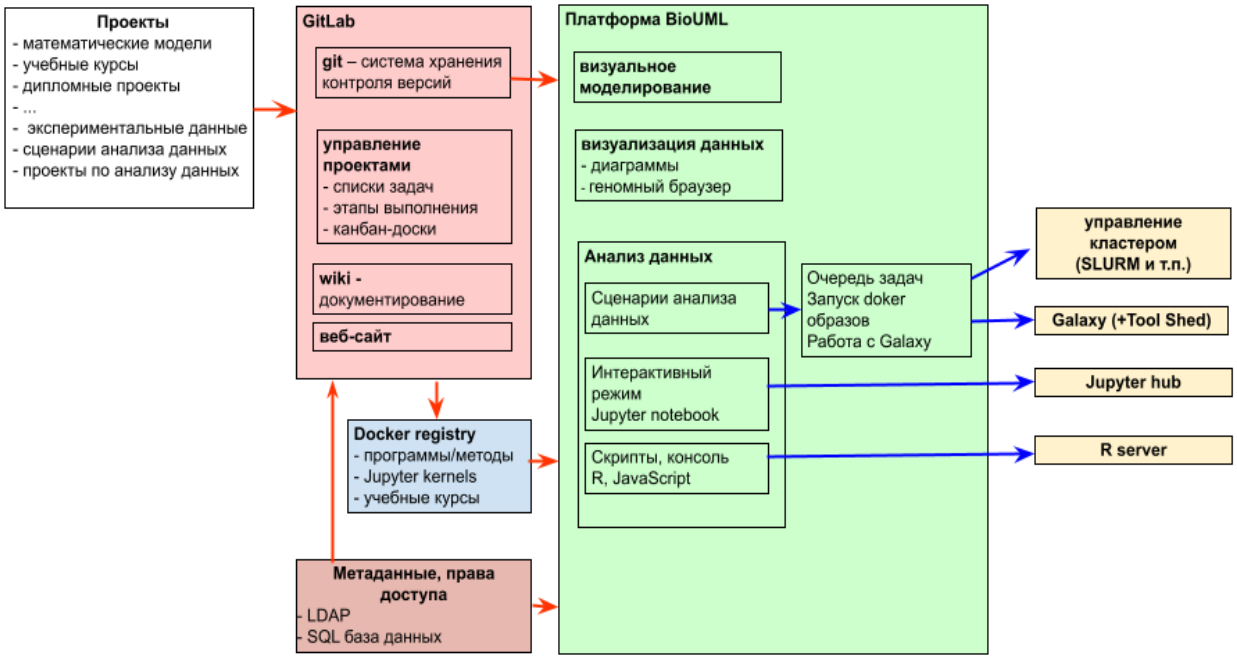

Рисунок 8.1. Архитектура платформы Sirius-web.

Сейчас эта платформа активно используется в Научно-технологическом университете "Сириус" для:

– ведения курсов "Введение в программирование", "Системная биология";

– выполнения студентами дипломных проектов;

– выполнения научных исследований, связанных с моделированием сложных биологических систем.

Платформа geneXplain (https://genexplain.com/genexplain-platform) – это коммерческий вариант ПК BioUML, разрабатываемый компаниями ООО "Биософт.Ру" и geneXplain GmbH (Вольфенбютель, Германия) для анализа омиксных данных (Kel et al., 2011). В ее основе лежат 3 основные идеи:

1) создать большой набор готовых сценариев для анализа различных типов омиксных данных, которые позволят пользователю, не являющемуся экспертом в области биоинформатики, проанализировать его данные. Другими словами, платформа geneXplain подобна "книге рецептов", где пользователь просто должен выбрать подходящий ему "рецепт", указать входные данные и на выходе получит готовый отчет и набор файлов с результатами анализа.

2) использование коммерческих баз данных для глубокого анализа омиксных данных. Для этого используются следующие базы данных:

- TRANSFAC® (https://genexplain.com/transfac) – база данных ТФ и их

сайтов связывания, также содержит большую коллекцию весовых матриц для определения ССТФ (Wingender 2008, Wingender et al., 2018).

- TRANSPATH® (https://genexplain.com/transpath) – база данных метаболических путей и путей передачи сигнала в клетке у млекопитающих (Krull et al., 2006).

3) собственные методы глубокого анализа омиксных данных с использованием баз данных TRANSFAC и TRANSPATH.

Платформа Genome Enhancer (https://genexplain.com/genexplain- platform) – специализированный вариант ПК BioUML для анализа омиксных данных в автоматическом режиме (Kel et al., 2011). Ее основные идеи, предложенные А.Э. Келем :

– от генома к мишени – по загруженным пользователем омиксным данным платформа автоматически определяет: гены, изменившие экспрессию; соответствующие им мастер-регуляторы; возможные лекарственные препараты или химические вещества, которые могли бы активировать или ингибировать мастер-регуляторы, чтобы вернуть экспрессию генов к нормальному состоянию;

– максимально упростить интерфейс пользователя для анализа данных;

– максимально автоматизировать анализ омиксных данных – в зависимости от загруженных пользователем данных платформа автоматически составляет нужный сценарий анализа этих данных;

– представить результаты анализа данных в виде отчета, сходного по содержанию и оформлению с готовой научной статьей.

ЗАКЛЮЧЕНИЕ

Нами создан ПК BioUML, реализующий весь инструментарий, необходимый для успешного использования предложенной технологии для создания и использования сложных модульных моделей биологических систем. Это существенно ускоряет и упрощает процесс построения, отладки и валидации сложных моделей биологических систем, содержащих сотни и тысячи переменных и уравнений.

Платформа Sirius-web, которая сочетает в себе возможности платформы BioUML для моделирования и анализа данных и платформы GitLab для хранения, управления и документирования проектов, создают такую среду,

на которой модели могут эволюционировать подобно живым существам:

– размножение – репозиторий GitLab, из которого можно легко скопировать математические модели. Подробная документация о структуре и параметрах моделей также представлена в репозитории и важна для дальнейшего развития моделей;

– мутации – редактор (как визуальный, так и текстовый) для редактирования модели;

– рекомбинация – поддержка модульного подхода для составления сложных моделей из готовых модулей;

– естественный отбор – методы подгонка параметров модели под экспериментальные данные и методы тестирования для проверки соответствия модели определенным условиям и экспериментальным данным.

Предложенная технология и разработанный ПК BioUML позволили создать и валидировать самые сложные в мире модульные модели процессов апоптоза, регуляции генной экспрессии при работе скелетных мышц и регуляции артериального давления.

Отдельно выделим модель регуляции генной экспрессии при работе скелетных мышц, поскольку при ее создании были использованы все основные разработки, представленные в данной диссертации:

– разработанная технология сложных модульных моделей;

– ПК BioUML, реализующий данную технологию;

– ПК BioUML для анализа геномных данных: 1) анализ CAGE-seq данных для выделения генов, чья экспрессия значимо меняется при физической нагрузке; 2) анализ ChIP-seq данных для выделения подтвержденные районов связывания ТФ CREB, FOS и JUN;

– база данных GTRD, в которой представлены результаты единообразного анализа данных ChIP-seq;

– платформа Sirius-web для хранения созданных моделей, воспроизводимых вычислительных экспериментов над созданными моделями в виде Jupyter ноутбуков.

Создание ПК BioUML играет важную роль с точки зрения обеспечения технологического суверенитета Российской Федерации в области моделирования сложных биологических систем и анализа биомедицинских данных. Данный комплекс является уникальной полнофункциональной отечественной биоинформатической платформой. В этом качестве ПК

BioUML используется в ряде отечественных проектов (см. раздел "Практическая значимость").

ВЫВОДЫ

1. Разработана эффективная технология для итерационного создания, тестирования и повторного использования модульных моделей сложных биологических систем, имеющих иерархическую структуру.
2. Разработан ПК BioUML реализующий весь инструментарий, необходимый для успешного использования предложенной технологии.
3. В составе разработанного ПК BioUML включены методы, программы, сценарии для анализа и графического представления биомедицинских данных.
4. Создана новая база данных GTRD, которая обеспечивает единообразную аннотацию, контроль качества, обработку и интеграцию большого объема экспериментальных NGS данных по регуляции транскрипции для 10 видов эукариот: *Homo sapiens, Mus musculus, Rattus norvegicus, Gallus gallus, Danio rerio, Caenorhabditis elegans, Drosophila melanogaster, Saccharomyces cerevisiae, Schizosaccharomyces pombe* и *Arabidopsis thaliana*. На основе методов машинного обучения разработаны модели для предсказания уровня транскрипции для наиболее экспериментально исследованных клеточных линий на основе собранных и интегрированных NGS данных.
5. Продемонстрированы возможности разработанной технологии и ПК BioUML для создания и использования сложных моделей биологических систем, включая: процессы, происходящие на клеточном уровне (метаболизм, транскрипция, отдельные пути передачи сигнала, апоптоз); регуляцию артериального давления у человека; лекарственную терапию артериальной гипертензии; эпидемиологию COVID-19.
6. Разработана технология построения цифрового двойника пациента. Показана ее применимость на примере оптимизации выбора лекарственной терапии для лечения артериальной гипертензии.

СВИДЕТЕЛЬСТВА НА РЕГИСТРАЦИЮ ПРОГРАММ ДЛЯ ЭВМ

1. **Колпаков Ф.А.**, Ананько Е.А., Колчанов Н.А., Колесов Г.Б., Подколодная О.А., Игнатьева Е.В. База данных GeneNet. РОСПАТЕНТ, св-во № 990006 от 15.02.1999 г.
2. **Колпаков Ф.А.** BioUML workbench (BioUML). Свидетельство о государственной регистрации программы для ЭВМ № 2008611606 от 27.03.2008 г.
3. **Колпаков Ф.А.** BioUML Enterprise Edition. Свидетельство о государственной регистрации программы для ЭВМ № 2008612975 от 19.06.2008 г.
4. **Колпаков Ф.А.** Модульная математическая модель регуляции липидного метаболизма на уровне организма человека. Свидетельство о государственной регистрации программы для ЭВМ № 2015618214 от 03.08.2015 г.
5. Евшин И.С., **Колпаков Ф.А.**, Шарипов Р.Н. GTRD – gene transcription regulation database. Свидетельство о государственной регистрации базы данных № 2017620872 от 10.08.2017 г.
6. Кель А.Э., **Колпаков Ф.А.**, Шарипов Р.Н., Киселев И.Н., Стельмашенко Д.Е., Мандрик Н.В., Евшин И.С., Пинтус С.С. Программа «От генома к мишени» // Свидетельство о государственной регистрации программы для ЭВМ № 2017616241 от 02.06.2017 г.
7. Кель А.Э., Киселев И.Н., **Колпаков Ф.А.**, Пинтус С.С. Модуль графической аннотации экспериментальных данных. Свидетельство о государственной регистрации программы для ЭВМ № 2017615715 от 15.06.2017.

СПИСОК ОСНОВНЫХ ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

Журналы из списка Web of Science и/или Scopus, Q1-Q2

1. **Kolpakov F**., Akberdin I.R., Kiselev I.N., Kolmykov S.K., Kondrakhin Y., Kulyashov M., Kutumova E.O., Pintus S.S., Ryabova A., Sharipov R.N., Yevshin I.S., Zhatchenko S., Kel A. BioUML – towards a universal research platform // Nucleic Acids Research – 2022. – V.50(W1). – P.W124-W131.
2. Kutumova E., Kiselev I., Sharipov R., Lifshits G., **Kolpakov F.** Mathematical modeling of antihypertensive therapy // Frontiers in Physiology – 2022. – V.13. – Art.1070115.
3. Boytsov A., Abramov S., Aiusheeva A., Kasianova A., Baulin E., Kuznetsov I., Aulchenko Y., Kolmykov S.K., Yevshin I.S., **Kolpakov F.,** Vorontsov I.E., Makeev V.J., Kulakovskiy I.V. ANANASTRA: annotation and enrichment analysis of allele-specific transcription factor binding at SNPs // Nucleic Acids Research – 2022. – V.50(W1). –P.W51-W56.
4. Kutumova E.O., Akberdin I.R., Kiselev I.N., Sharipov R.N., Egorova V.S., Syrocheva A.O., Parodi A., Zamyatnin A.A., **Kolpakov F.A.** Physiologically Based Pharmacokinetic Modeling of Nanoparticle Biodistribution: A Review of Existing Models, Simulation Software, and Data Analysis Tools // International Journal of Molecular Sciences – 2022. – V.23(20). – Art.12560.
5. Afonyushkin V., Akberdin I.R., Kozlova Y., Schukin I., Mironova T., Bobikova A., Cherepushkina V., Donchenko N., Poletaeva Y., **Kolpakov F.A**. Multicompartmental Mathematical Model of SARS-CoV-2 Distribution in Human Organs and Their Treatment // Mathematics – 2022. – V.10(11). – Art.1925.
6. Kutumova E.O., Akberdin I.R., Kiselev I.N., Sharipov R.N., **Kolpakov F.A**. Modular Representation of Physiologically Based Pharmacokinetic Models: Nanoparticle Delivery to Solid Tumors in Mice as an Example // Mathematics – 2022. – V.10(7). – Art.1176.
7. Makhnovskii P.A., Gusev O.A., Bokov R.O., Gazizova G.R., Vepkhvadze T., Lysenko E.A., Vinogradova O.L., **Kolpakov F.A.,** Popov D.V. Alternative transcription start sites contribute to acute-stress-induced transcriptome response in human skeletal muscle // Human Genomics – 2022. – V.16(1). –Art.24.

8. Abramov S., Boytsov A., Bykova D., Penzar D.D., Yevshin I., Kolmykov S.K., Fridman M.V., Favorov A.V., Vorontsov I.E., Baulin E., **Kolpakov F.A.,** Makeev V.J., Kulakovskiy I.V. Landscape of allele-specific transcription factor binding in the human genome // Nature Communications – 2021. – V.12. – Art.2751.
9. Akberdin I.R., Kiselev I.N., Pintus S.S., Sharipov R.N., Vertyshev A.Y., Vinogradova O.L., Popov D.V., **Kolpakov F.A.** A modular mathematical model of exercise-induced changes in metabolism, signaling, and gene expression in human skeletal muscle // International Journal of Molecular Sciences – 2021. – V.22(19). – Art.10353.
10. Kolmykov S., Yevshin I.S., Kulyashov M., Sharipov R.N., Kondrakhin Y., Makeev V.J., Kulakovskiy I.V., Kel A., **Kolpakov F.** GTRD: an integrated view of transcription regulation // Nucleic Acids Research – 2021. – V.49(D1). – P.D104-D111.
11. Kutumova E., Kiselev I., Sharipov R., Lifshits G., **Kolpakov F.** Thoroughly Calibrated Modular Agent-Based Model of the Human Cardiovascular and Renal Systems for Blood Pressure Regulation in Health and Disease // Frontiers in Physiology – 2021. – V.12. – Art.746300.
12. Makhnovskii P.A., Bokov R.O., **Kolpakov F.A.,** Popov D.V. Transcriptomic Signatures and Upstream Regulation in Human Skeletal Muscle Adapted to Disuse and Aerobic Exercise // International Journal of Molecular Sciences – 2021. – V.22(3). – Art.1208.
13. Pintus S.S., Akberdin I.R., Yevshin I.S., Makhnovskii P.A., Tyapkina O., Nigmetzyanov I., Nurullin L., Devyatiyarov R., Shagimardanova E., Popov D., **Kolpakov F.A.**, Gusev O., Gazizova G.R. Genome-Wide Atlas of Promoter Expression Reveals Contribution of Transcribed Regulatory Elements to Genetic Control of Disuse-Mediated Atrophy of Skeletal Muscle // Biology (Basel) – 2021. – V.10(6). – Art.557.
14. Makhnovskii P.A., Zgoda V.G., Bokov R.O., Shagimardanova E.I., Gazizova G.R., Gusev O.A., Lysenko E.A., **Kolpakov F.A**., Vinogradova O.L., Popov D.V. Regulation of Proteins in Human Skeletal Muscle: The Role of Transcription // Scientific Reports – 2020. – V.10(1). – Art.3514.
15. Sharipov R.N., Kondrakhin Y.V., Ryabova A.S., Yevshin I.S., **Kolpakov F.A.** Assessment of transcriptional importance of cell line-specific features based on GTRD and FANTOM5 data // PloS ONE – 2020. – V.15(12). – Art.e0243332.
16. Keating S.M., Waltemath D., et al. SBML Level 3: an extensible format for the exchange and reuse of biological models // Molecular Systems Biology – 2020. – V.16. – Art.e9110.
17. Kel A., Boyarskikh U., Stegmaier P., Leskov L.S., Sokolov A.V., Yevshin I.S., Mandrik N., Stelmashenko D., Koschmann J., Kel-Margoulis O., Krull M., Martínez-Cardús A., Moran S., Esteller M., **Kolpakov F.,** Filipenko M., Wingender E. Walking pathways with positive feedback loops reveal DNA methylation biomarkers of colorectal cancer // BMC Bioinformatics – 2019. – V.20. – Art.119.
18. Kolmykov S.K., Kondrakhin Y.V., Yevshin I.S., Sharipov R.N., Ryabova A.S., **Kolpakov F.A.** Population size estimation for quality control of ChIP-Seq datasets // PloS ONE – 2019. – V.14(8). – Art.e0221760.
19. **Kolpakov F.,** Akberdin I., Kashapov T., Kiselev I., Kolmykov S., Kondrakhin Y., Kutumova E., Mandrik N., Pintus S., Ryabova A., Sharipov R.N., Yevshin I., Kel A. BioUML: an integrated environment for systems biology and collaborative analysis of biomedical data // Nucleic Acids Research – 2019. – V.47(W1). – P.W225-W233.
20. Yevshin I., Sharipov R., Kolmykov S., Kondrakhin Y., **Kolpakov F.** GTRD: a database on gene transcription regulation-2019 update // Nucleic Acids Research – 2018. – V. 47(D1). – P. D100-D105.
21. Kel A., Boyarskikh U., Stegmaier P., Leskov L.S., Sokolov A.V., Yevshin I., Mandrik N., Stelmashenko D., Koschmann J., Kel-Margoulis O., Krull M., Martínez-Cardús A., Moran S., Esteller M., **Kolpakov F.,** Filipenko M., Wingender E. Walking pathways with positive feedback loops reveal DNA methylation biomarkers of colorectal cancer // BMC Bioinformatics

– 2019. –V. 20(Suppl 4). – Art.119.
22. Kulakovskiy, I.V., Vorontsov, I.E., Yevshin, I.S., Sharipov, R.N., Fedorova, A.D., Rumynskiy, E.I., Medvedeva, Y.A., Magana-Mora, A., Bajic, V.B., Papatsenko, D.A., **Kolpakov, F.A.,** Makeev, V.J. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis // Nucleic Acids Research – 2018. – V.46(D1). – P.D252-D259.
23. Boyarskikh U., Pintus S., Mandrik N., Stelmashenko D., Kiselev I., Evshin I., Sharipov R., Stegmaier P., **Kolpakov F**, Filipenko M., Kel A. Computational master-regulator search reveals mTOR and PI3K pathways responsible for low sensitivity of NCI-H292 and A427 lung cancer cell lines to cytotoxic action of p53 activator Nutlin-3 // BMC Med Genomics – 2018. – V.11(Suppl. 1). – Art.12.
24. Yevshin I., Sharipov R., Valeev T., Kel A., Kolpakov F. GTRD: a database of transcription factor binding sites identified by ChIP-seq experiments // Nucleic Acids Research – 2017. – V.45(D1). – P.D61-D67.
25. Kulakovskiy I.V., Makeev V.J., Vorontsov I.E., Kasianov A.S., Medvedeva Y.A., Yevshin I.S., **Kolpakov F.A.**, Soboleva A.V., Ashoor H., Ba-Alawi W., Bajic V.B. HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models // Nucleic Acids Research – 2016. – V.44(D1). – P.D116-D125.
26. Kutumova E., Sharipov R., **Kolpakov F.**, Zinovyev A. Model composition through model reduction: a combined model of CD95 and NF-κB signaling pathways // BMC Systems Biology – 2013. – V.7. – Art.13.
27. Kel A., **Kolpakov F.,** Poroikov V., and Selivanova G. geneXplain -identification of causal biomarkers and drug targets in personalized cancer pathways // Journal of Biomolecular Techniques – 2011. – V. 22(Supplement). – Art.S16.
28. Waltemath D., Adams R., Bergmann F.T., Hucka M., **Kolpakov F.,** Miller A.K., Nickerson D., Moraru I.I., Sahle S., Snoep J.L., Le Novère N. Reproducible computational biology experiments with SED-ML – the simulation experiment description markup language // BMC Systems Biology – 2011. – V.5. – Art.198.
29. Le Novère N., Hucka M., Mi H., Moodie S., Schreiber F., Sorokin A., Demir E., Wegner K., Aladjem M.I., Wimalaratne S.M., Bergman F.T., Gauges R., Ghazal P., Kawaji H., Li L., Matsuoka Y., Villéger A., Boyd S.E., Calzone L., Courtot M., Dogrusoz U., Freeman T.C., Funahashi A., Ghosh S., Jouraku A., Kim S., **Kolpakov F.,** Luna A., Sahle S., Schmidt E., Watterson S., Wu G., Goryanin I., Kell D.B., Sander C., Sauro H., Snoep J.L., Kohn K., Kitano H. The Systems Biology Graphical Notation // Nature Biotechnology – 2009. – V.27(8). – P.735-741.
30. **Kolpakov F,** Poroikov V, Sharipov R, Kondrakhin Y, Zakharov A, Lagunin A, Milanesi L, Kel A. CYCLONET – an integrated database on cell cycle regulation and carcinogenesis // Nucleic Acids Research – 2007. – V 35(Database issue). – D550-D556.
31. **Kolpakov F.A.**, Ananko E.A. Interactive data input into the GeneNet database // Bioinformatics – 1999. – V.15(7-8). – P.713-714.
32. Kolchanov N.A., Ananko E.A., Podkolodnaya O.A., Ignatieva E.V., Stepanenko I.L., Kel-Margoulis O.V., Kel A.E., Merkulova T.I., Goryachkovskaya T.N., Busygina T.V., **Kolpakov F.A.,** Podkolodny N.L., Naumochkin A.N., Romashchenko A.G. Transcription Regulatory Regions Database (TRRD):its status in 1999 // Nucleic Acids Research – 1999. – V.27(1). – P.303-306.
33. **Kolpakov F.A.,** Ananko E.A., Kolesov G.B., Kolchanov N.A. GeneNet: a gene network database and its automated visualization // Bioinformatics – 1998. – V.14(6). – P. 529-537.
34. Heinemeyer T., Wingender E., Reuter I., Hermjakob H., Kel A.E., Kel O.V., Ignatieva E.V., Ananko E.A., Podkolodnaya O.A., **Kolpakov F.A.,** Podkolodny N.L., Kolchanov N.A. Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL // Nucleic Acids Research – 1998. – V.26(1). – P.362-367.

Журналы из списка Web of Science и/или Scopus, Q3-Q4

35. Акбердин И.Р., Вертышев А.Ю., Пинтус С.С., Попов Д.В., **Колпаков Ф.А.** Математическая модель, связывающая Ca2+-зависимый сигнальный путь с регуляцией экспрессии генов в клетках скелетной мышцы человека // Математическая биология и биоинформатика – 2020. – Т.15(1). – С.20-39.

36. Киселев И.Н., Акбердин И.Р., Вертышев А.Ю., Попов Д.В., **Колпаков Ф.А.** Модульная графическая модель энергетического метаболизма в клетках скелетной мышцы // Математическая биология и биоинформатика – 2019. – Т.14(2). – С. 373-392.

37. Киселев И.Н., Кутумова Е.О., Колпакова А.Ф., Лифшиц Г.И., **Колпаков Ф.А.** Математическое моделирование действия антигипертензивных препаратов // Математическая биология и биоинформатика – 2019. – Т.14(1). – С. 233-256.

38. Vorontsov I.E., Fedorova A.D., Yevshin I.S., Sharipov R.N., **Kolpakov F.A.**, Makeev V.J., Kulakovskiy I.V. Genome-wide map of human and mouse transcription factor binding sites aggregated from ChIP-Seq data 06 Biological Sciences 0604 Genetics // BMC Research Notes – 2018. – V.11(1). – Art.756.

39. Kondrakhin Y., Valeev T., Sharipov R., Yevshin I., **Kolpakov F**., Kel A. Prediction of protein-DNA interactions of transcription factors linking proteomics and transcriptomics data // EuPA Open Proteomics – 2016. – V.13. – P.14-23.

40. Киселев И. Н., Бибердорф Э. А., Баранов В. И., Комлягина Т. Г., Мельников В. Н., Суворова И. Ю., Кривощеков С. Г., **Колпаков Ф. А.** Персонализация параметров и валидация модели сердечно-сосудистой системы человека // Математическая биология и биоинформатика – 2015. – Т.10(2). – С.526-547.

41. Kutumova E. O., Kiselev I. N., Sharipov R. N., Lavrik I. N., **Kolpakov F. A.** A modular model of the apoptosis machinery // Advances in Experimental Medicine and Biology – 2012. – V.736. – P.235-245.

42. Кутумова Е.О., Зиновьев А.Ю., Шарипов Р.Н., **Колпаков Ф.А.** Применение методов редукции для построения комплексной модели апоптотических путей // Математическая биология и биоинформатика – 2012. – Т.7(2). – С.572-588.

43. Киселев И. Н, Семисалов Б. В, Бибердорф Э. А., Шарипов Р. Н., Блохин А. М., **Колпаков Ф.А.** Модульное моделирование сердечно-сосудистой системы человек // Математическая биология и биоинформатика – 2012. – Т.7(2). – С.703-736.

44. Колчанов Н.А., Ананько Е.А., **Колпаков Ф.А.,** Подколодная О.А., Игнатьева Е.В., Горячковская Т.Н., Степаненко И.Л. Генные сети // Молекулярная биология – 2000. – V.34(4). – P.533-544.

45. Ананько Е.А., **Колпаков Ф.А.**, Колесов Г.Б., Колчанов Н.А. Автоматическая генерация схем генных сетей на основе их формализованного описания в базе данных GeneNet // Биофизика – 1999. – Т.44(4). – С. 628-631.

46. **Колпаков Ф.А.**, Бабенко В.Н. Компьютерная система MGL – инструмент для составления выборок, графического представления и анализа регуляторных геномных последовательностей // Молекулярная биология – 1997. – Т.31(4). – С.647-655.

Журналы из списка Web of Science и/или Scopus, без Q

47. Akberdin I.R., Kiselev I.N., Pintus S.S., Vertyshev A., Makhnovskii P.A., Popov D.V., **Kolpakov F.A.** Integrated modular model linking metabolism, signaling transduction and gene expression regulation in human skeletal muscle // CEUR Workshop Proceedings – 2020. – V.2569. – P.1-5.

48. Kolmykov S.K., Evshin I.S., **Kolpakov F.A**. Analysis of NGS Data on the Transcriptional Regulation // CEUR Workshop Proceedings – 2020. – V.2569. – P.19-22.

49. Kulyashov M.A., Kolmykov S.K., Evshin I.S., **Kolpakov F.A.** Description, Characteristic And Algorithm For Creation Of A Dictionary Of Cell Types And Tissues In The Gtrd Database // CEUR Workshop Proceedings – 2020. – V.2569. – P.13-18.

Рецензируемые журналы, рекомендованные ВАК

50. Миронова Т.Е., Афонюшкин В.Н., Козлова Ю.Н., Бобикова А.С., Коптев В.Ю., Черепушкина В.С., Сигарева Н.А., **Колпаков Ф.А.** Изучение защитных эффектов вироцидных препаратов на модели коронавирусной пневмонии // Ветеринария и кормление – 2020. - № 7. – С.35-38.

51. Баранов В.И., Бибердорф Э.А., Киселев И.Н., **Колпаков Ф.А.,** Комлягина Т.Г., Кривощеков С.Г., Мельников В.Н., Суворова И.Ю. Персонализированная одномерная модель сердечно-сосудистой системы человека // Сибирский научный медицинский журнал – 2016. – Т.36(1). – С.70-79.

Прочие журналы

52. Стельмашенко Д.Е., Колпаков Ф.А., Евшин И.С., Мандрик Н.В., Рябова А.С., Пинтус С.С., Киселев И.Н., Шарипов Р.Н., Штегмайер Ф., Кошманн Д., Кель-Маргулис О.В., Крул М., Тандон Н., Вингендер Э., Кель А.Э. Персонализированная медицина в один клик: автоматическая идентификация терапевтической мишени по ДНК и РНК пациента // Лаборатория и производство. – 2019. – №3 (7). – С. 76-86.

53. Boyarskikh U.A., Pintus S.S., Mandrik N.V.,. Stelmashenko D.E, Kiselev I.N., Sharipov R.N., Kolpakova A.F., Stegmaier P., Filipenko M.L., Kolpakov F.A., Kel A.E.. Master-regulators driving resistance of non-small cell lung cancer cells to p53 reactivator Nutlin-3 // Virtual Biology – 2017. – V.4. – P.1–31.

54. Kondrakhin Yu., Valeev T., Sharipov R., Yevshin I., **Kolpakov F.** Toolkit for ChIP-seq based comparative analysis of the PWM methods for prediction of transcription factor binding sites // Virtual Biology – 2014. – V.2. – P.19-33.

55. Kiselev I., **Kolpakov F.** BioUML: plugin for population-based modeling // Virtual Biology – 2014. – V.2. – P.7-15.

56. Kiselev I. N., **Kolpakov F. A.** Modular Modeling of Biological Systems // Virtual Biology – 2013. – V.1. – P.30-44.

57. Valeev T., Yevshin I., **Kolpakov F.** BioUML genome browser // Virtual Biology – 2013. – V.1. – P.15-26.

58. Kutumova E., Ryabova A., Valeev T., **Kolpakov F**. BioUML plug-in for nonlinear parameter estimation using multiple experimental data // Virtual Biology – 2013. – V.1. – P.47-58.

Главы в книгах

59. Kiselev I.N., Semisalov B.V., Sharipov R.N., **Kolpakov F.A.** Modular approach to modeling complex biological systems. Circulatory System and Arterial Hypertension: Experimental Investigation, Mathematical and Computer Simulation / Eds. Ivanova L. N., Markel A. L., Blokhin A. M., Mishchenko E. V. Nova Science Publishers, Inc. 2012. – 258 pp. – ISBN-10: 1622572750.

60. Kiselev I.N., Semisalov B.V., Sharipov R.N., **Kolpakov F.A.** Modular modeling of the human cardiovascular system. Circulatory System and Arterial Hypertension: Experimental Investigation, Mathematical and Computer Simulation / Eds. L. N. Ivanova, A. L. Markel, A. M. Blokhin, E. V. Mishchenko. Nova Science Publishers, Inc. 2012. – 258 pp. – ISBN-10: 1622572750.

61. **Колпаков Ф.А.**, Шарипов Р.Н., Евшин И.С., Леонова Т.И., Семисалов Б.В. Компьютерное моделирование системы кровообращения. Система кровообращения и артериальная гипертония: биофизические и генетико-физиологические механизмы, математическое и компьютерное моделирование / под ред. Л.Н. Ивановой, А.М.Блохина, А.Л. Маркеля. Новосибирск, Россия. Издательство СО РАН 2008. – 252 с. ISBN: 978-5-7692-1021-1.

Учебные пособия

62. Бибердорф Э.А., Блохин Ф.М., Киселев И.Н. Семисалов Б.В. Колпаков Ф.А. Система кровообращения человека – математическое и компьютерное моделирование: учебное пособие. Новосибирск: РИЦ НГУ 2015. – 187 с.

Труды конференций (избранное)

63. Киселев И.Н., Колпакова А.Ф., **Колпаков Ф.А.** Математическое моделирование фармакокинетики и фармакодинамики алискирена // XVI Российская конференция «Распределенные информационно - вычислительные ресурсы. Наука – цифровой экономике» (DICR-2017): Труды XVI Всероссийской конференции (4-7 декабря 2017 г.). Новосибирск / Под ред. О.Л. Жижимова, А.М. Федотова. – Новосибирск: ИВТ СО РАН. – 2017. – С.218-223.

64. Киселев И.Н., Шарипов Р.Н., Мандрик Н.В., **Колпаков Ф.А.** Популяционное моделирование сердечно-сосудистой системы человека в платформе BioUML // XVI Российская конференция «Распределенные информационно - вычислительные ресурсы. Наука – цифровой экономике» (DICR-2017): Труды XVI Всероссийской конференции (4-7 декабря 2017 г.). Новосибирск / Под ред. О.Л. Жижимова, А.М. Федотова. – Новосибирск: ИВТ СО РАН. – 2017. – С.212-217.

65. Pintus S., Valeev T.F., Yevshin I.S., **Kolpakov F.A.** Biostore: a cloud-compatible hub for bioinformatics related tools and platforms // Tenth International Conference on Genome Regulation and Structure\Systems Biology (BGRS\SB-2016): Novosibirsk, 29 August – 2 September, 2016. Abstracts. – Novosibirsk. – 2016. – P.237.

66. Mandrik N., Kutumova E.O., Kolpakov F.A. Rule-based modeling in BioUML // Proceedings of The Second International Conference “Mathematical Modeling and High Performance Computing in Bioinformatics, Biomedicine and Biotechnology” (within BGRS\SB-2016): Novosibirsk, Russia. 29 August – 2 September 2016. – Novosibirsk. – 2016. – P.72.

67. Kiselev I.N., **Kolpakov F.A.**, Biberdorf E.A., Baranov V.I., Komlyagina T.G., Suvorova I.Y., Melnikov V.N., Krivoschekov S.G. Patient-specific 1d model of the human cardiovascular system // Proceedings - 2015 International Conference on Biomedical Engineering and Computational Technologies, SIBIRCON 2015.Institute of Electrical and Electronics Engineers Inc.– 2015. – P.76-81.

68. **Kolpakov F.A.,** Kiselev I.N., Valeev T.F., Bukharov A., Mandrik N., Kondrakhin Y.V., Yevshin I.S., Kolpakova A. BioUML Platform - From Systems Biology to Systems Pharmacology // 7th Noordwijkerhout Symposium on Pharmakokinetics, Pharmacodynamics and Systems Pharmacology. Systems Pharmacology in Drug Discovery & Development: Noordwijkerhout, The Netherlands, 23-25 April 2014. – Noordwijkerhout: Leiden Academic Centre for Drug Research. – 2014. – P.156-157.

69. Yevshin I.S., Valeev T.F., Sharipov R.N., **Kolpakov F.A.** BioUML – software platform for analysis of Next generation sequencing data using collaborative and reproducible research // International conference “High-throughput sequencing in genomes - 2013”: Novosibirsk, July 21-25, 2013. – Novosibirsk: Institute of Chemical Biology and Fundamental Medicine. – 2013. – P.66.

70. Semisalov B.V., Kiselev I.N., Sharipov R.N., **Kolpakov F.A.** BioUML: plugin for stochastic modeling of biological systems // The 8th International Conference on Bioinformatics of Genome Regulation and Structure/Systems Biology: Novosibirsk, June 25-29 2012. – Novosibirsk. – 2012. – P.283.

71. **Kolpakov F**., Puzanov M., Koshukov A. BioUML: visual modeling, automated code generation and simulation of biological systems // Proceedings of the 5th International Conference on Bioinformatics of Genome Regulation and Structure (BGRS’2006), July 16-22, 2006. – Novosibirsk. –2006. – P.281-284.

72. **Kolpakov F.A.** BioUML – open source extensible workbench for systems biology // Proceedings of the 4th International Conference on Bioinformatics of Genome Regulation and Structure (BGRS’2004), July 25-30, 2004. – Novosibirsk. –2004. – P.77-80.

73. **Kolpakov F.A.** Comprehensive model for formalized description, visualization and simulation of biological systems // Proceedings of German Conference on Bioinformatics. October 12-14, 2003. – Neuherberg/Garching, Germany. – 2003. – V.II. – P.155-157.
74. **Kolpakov F.A.** BioUML – framework for visual modeling and simulation of biological systems. // Proceedings of the 3rd International Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2002). – Novosibirsk. – 2002. – V.2. P.128-131.
75. Kel A.E., Kondrakhin Y.V., Kolpakov **Ph.A.**, Kel O.V., Romashenko A.G., Wingender E., Milanesi L., Kolchanov N.A. Computer tool FUNSITE for analysis of eukaryotic regulatory genomic sequences // Proceedings. International Conference on Intelligent Systems for Molecular Biology. – 1995. – V. 3. – P. 197-205.