

ОТЗЫВ

официального оппонента на диссертационную работу

Колпакова Федора Анатольевича

«Компьютерное моделирование биологических систем и анализ биомедицинских данных», представленную на соискание ученой степени доктора биологических наук по специальности 1.5.8 – «Математическая биология, биоинформатика».

Диссертация Ф.А. Колпакова суммирует и обобщает его результаты, полученные почти за 30 лет в области разработки технологии и программного обеспечения для компьютерного моделирования биологических систем, а также анализа и визуализации биомедицинских данных, в первую очередь связанных с регуляцией транскрипции.

Целью диссертационной работы являлась разработка технологии и инструментария (программный комплекс и базы данных) для эффективного построения и использования моделей сложных биологических систем в интеграции с анализом геномных/омиксных данных.

Актуальность. Создание моделей сложных биологических систем до сих пор остается искусством, доступным немногим исследователям. Цикл разработки таких моделей, как правило, составляет несколько лет. Задача моделирования может быть существенно упрощена и ускорена путем разработки подходящих именно для моделирования биологических систем технологии и инструментария - программный комплекс. При этом такой программный комплекс должен быть интегрирован с инструментом для анализа и визуализации биомедицинских данных.

Также актуальной и перспективной является задача построения цифрового двойника пациента, что открывает новые возможности для персонализированной медицины.

Для обеспечения технологического суверенитета Российской Федерации важной является задача разработки полнофункциональной отечественной биоинформатической платформы.

Диссертационная работа имеет комплексный характер. В ней представлены **теоретические, технологические и прикладные результаты.**

Теоретическая значимость. В диссертационной работе разработана новая технология (методология) для итерационного создания, тестирования

и использования модульных моделей биологических систем. Основной новизной этого подхода являются, предложенные автором:

- мета-модель для формализованного представления широкого круга моделей биологических систем;
- возможность синхронного представления и модификации моделей биологических систем как в графическом, так и текстовом виде;
- адаптация практик разработки программного обеспечения к построению сложных моделей биологических систем;
- создание среды, в которой модели могут эволюционировать подобно живым существам.

Колпаковым Ф.А. предложена технология построения цифрового двойника пациента. Разработан новый подход для решения проблемы неидентифицируемости части параметров математической модели на основе ограниченных данных реального пациента. Для этого для одного реального пациента находится множество возможных решений, которые представляются в виде популяции виртуальных пациентов.

Для анализа и интеграции NGS данных по регуляции транскрипции автором предложен новый подход объединения данных из различных экспериментов в виде кластеров, мета-кластеров и мастер-треков.

Основными **практическими результатами** диссертационной работы являются:

- 1) программный комплекс BioUML, который, с одной стороны, содержит весь инструментарий, необходимый для успешного использования предложенной технологии построения сложных моделей биологических систем, а с другой стороны - содержит множество методов, программ, сценариев для анализа и графического представления биомедицинских данных, что позволяет рассматривать программный комплекс BioUML как полнофункциональную биоинформатическую платформу;
- 2) компьютерные модели биологических систем. Используя разработанную технологию и программный комплекс BioUML под руководством автора был разработан ряд сложных модульных 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.

- 3) База данных GTRD - которая содержит единообразно аннотированные, проанализированные и интегрированные данные 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*.

База данных GTRD востребована научным сообществом, широко используется и цитируется: три публикации в специализированных выпусках *Nucleic Acids Research* 2017, 2019 и 2021 года набрали в совокупности 400+ цитирований, включая цитирования в журналах *Nature* и *Science*.

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», что подтверждено соответствующими актами внедрения.

Структура и объем диссертации. Диссертационная работа состоит из введения, 8 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.

Результаты комплексной диссертационной работы изложены на хорошем уровне, понятным языком с достаточной степенью детализации.

Во введении обосновывается актуальность выбора темы исследований, приведены цели, задачи и положения, выносимые на защиту, а также приведены научная новизна, теоретическая и практическая значимость, апробация работы и личный вклад автора.

Глава 1 - Обзор литературы - начинается с описания понятия системной биологии, соответствующих экспериментальных методов, а также используемых стандартов и форматов для обеспечения воспроизводимости моделей биологических систем. В следующих двух

разделах рассмотрены основные методы и программные комплексы для моделирования биологических систем. Далее описаны методы и сценарии анализа генетических данных и их графического представления. В отдельном разделе приведен обзор предшествующих работ автора, выполненных в 1994–2000 г.г. в лаборатории теоретической генетики Института Цитологии и Генетики СО РАН под руководством академика РАН, профессора, д.б.н. Колчанова Н.А., которые послужили заделом для разработки BioUML.

Глава 2 - ПК BioUML - посвящена описанию архитектуры, интерфейса пользователя и технологий, использованных при создании программного комплекса BioUML. Часть информации вынесена в отдельные приложения (этапы развития, основные модули, Java библиотека для доступа и поиска информации в базах данных, Java библиотека графических объектов).

Глава 3 посвящена технологии создания и использования модульных моделей биологических систем. В основе этой технологии лежит разработанная автором мета-модель для комплексного описания, графического представления и численного моделирования широкого круга биологических систем. Далее описаны использование визуального, модульного, агентного и численного моделирования, текстовое представление модели и его синхронизация с графическим представлением, автоматическая генерация кода, определение параметров моделей и создание математических моделей сложных биологических систем эволюционным путем. Часть информации также вынесена в отдельные приложения (Java интерфейсы для задания графической нотации, исходная графическая нотация ПК BioUML на основе GeneNet, расширение и синхронизация SBGN Process Diagram и Antimony для графического и текстового представления моделей).

Глава 4 - Анализ и графическое представление биомедицинских данных - содержит описание возможностей программного комплекса BioUML для анализа и графического представления геномных данных. Для графического представления геномных данных разработан новый геномный браузер, который по функционалу сопоставим с ведущими геномными браузерами, а также имеет ряд уникальных возможностей.

Глава 5 посвящена описанию базы данных GTRD - Gene Transcription Regulation Database, созданной под руководством автора. Данная глава начинается с постановки задачи, далее подробно описаны использованные

методы и подходы для решения этих задач. В следующем разделе описано использование базы данных GTRD для построения цистрома для человека и мыши. На основе БД GTRD коллективом чл.-

корр. РАН, д.б.н., проф. Макеева В.Ю. и д.б.н. Кулаковского И.В. были созданы ресурсы: HOCOMOCO – коллекция мотивов для сайтов связывания транскрипционных факторов человека и мыши; ADASTR – коллекция данных по аллель-специфичному связыванию факторов транскрипции в геноме человека; ANANASTRA – веб-сервер для аннотации SNP на аллель-специфичное связывание факторов транскрипции в геноме человека. Также информация из GTRD была использована при создании известных международных ресурсов BaMM motif и mSigDB. Последний раздел посвящен обсуждению GTRD: сравнению с другими базами данных по ChIP-seq экспериментам, полнота покрытия транскрипционных факторов и их сайтов связывания.

Глава 6 - Моделирование сложных биологических систем. В первом разделе детально рассмотрены основные этапы предложенной технологии на примере создания модульной модели апоптоза. Следующие разделы посвящены моделям регуляции генной экспрессии при физической нагрузке в скелетных мышцах, эпидемиологии COVID-19, регуляции артериального давления у человека.

Глава 7 - Цифровой двойник пациента. Приведено описание разработанной автором технологии построения цифрового двойника пациента. Описаны генерация виртуальной популяции для одного реального пациента. Следующие разделы посвящены моделированию лекарственной терапии артериальной гипертензии и ее валидации на основе клинических данных. Далее описаны настройка параметров модели под конкретного пациента (персонализация) и моделирование лечения пациента. На примере реального пациента показана применимость предложенного подхода для оптимизации выбора лекарственной терапии для лечения артериальной гипертензии.

Глава 8 - описаны другие программные продукты, созданные на основе технологии BioUML, в том числе: компоненты прототипа «Национальной базы генетической информации»; биоинформатическая платформа для хранения, анализа и графического представления данных, полученных при одномолекулярном секвенировании ДНК; платформа Sirius-web для проектов, связанных с анализом данных и моделированием

для образовательных и научных проектов; платформы geneXplain и Genome Enhancer для анализа омиксных данных.

В заключении отмечается, что разработанная технология и поддерживающий ее программный комплекс ПК BioUML существенно ускоряет и упрощает процесс создания, 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, в которой модели могут эволюционировать подобно живым существам.

Выводы работы отражают полученные результаты и соответствуют поставленным задачам.

Текст диссертации позволяет заключить, что эта работа является завершённой, содержит ряд новых научных результатов, как теоретических, так и практических.

Научные положения, выносимые на защиту, выводы, сформулированные в диссертации полностью **обоснованы, их достоверность и новизна** не вызывает сомнений.

Это подтверждается тщательным тестированием разработанного программного комплекса BioUML, а также валидацией и тестированием созданных моделей сложных биологических систем. По результатам независимых сравнений программный комплекс 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.

Независимыми рецензентами статьи Kutumova et al., 2021 отмечено, что представленная в диссертационной работе математическая модель регуляции артериального давления у человека является одной из лучших в мире: "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".

Публикации. Результаты диссертационной работы представлены в 75 научных публикациях, из них 46 – публикации в журналах Web of Science и Scopus, причем 34 публикации относятся к Q1 и Q2 Web of Science

или Scopus. Все работы опубликованы в соавторстве. Издано 1 учебное пособие. Получено 7 свидетельств на регистрацию программ и баз данных для ЭВМ.

Также результаты исследования были представлены на большом количестве научных конференций в России и за рубежом.

Автореферат. Структура и текст автореферата соответствуют структуре и тексту диссертации. Автореферата изложен на 47 страницах, содержит 3 таблицы и 10 рисунков.

Замечания и рекомендации.

В работе большинство разработанных моделей создается путем объединения известных из литературы моделей подсистем. Позволяет ли такой подход моделировать эмергентные свойства систем и если да, то какие.

Большинство разработанных моделей предлагается использовать в описательной моде через подбор параметров, дающий наилучшее совпадение с экспериментальными данными. Такой подход часто используется в моделировании, но для доказательства правильности модели он недостаточен. Хотелось бы получить ответ, какие предсказания можно сделать из разработанных в работе моделей и какие простейшие экспериментальные проверки этих предсказаний можно было бы сделать.

Отметим также, что визуальная оценка совпадения результатов численного моделирования с данными клинических исследований (рис. 7.5.1.) не должна подменять статистическую оценку, которая, к сожалению, отсутствует.

Эти замечания, однако, не влияют на общую высокую оценку работы.

Заключение. Диссертационная работа Колпакова Федора Анатольевича «Компьютерное моделирование биологических систем и анализ биомедицинских данных», представленная на соискание ученой степени доктора биологических наук по специальности 1.5.8 – математическая биология, биоинформатика, является законченной научно-квалификационной работой, полностью соответствующей требованиям, предъявляемым к диссертациям на соискание ученой степени доктора наук согласно п. 9-14 «Положение о присуждении ученых степеней», утвержденного Постановлением Правительства РФ от 24.09.2013 № 842, а

её автор, Колпаков Федор Анатольевич, безусловно достоин присуждения
искомой ученой степени доктора биологических наук по специальности
1.5.8. — математическая биология, биоинформатика.

Дата: 16 января 2024 г.

Официальный оппонент

Самсонова Мария Георгиевна

доктор биологических наук, профессор Высшей школы прикладной
математики и вычислительной физики, заведующая научно-
исследовательской лабораторией «Математическая биология и
биоинформатика», ФГАОУ ВО «Санкт-Петербургский политехнический
университет Петра Великого»
г. Санкт-Петербург

Федеральное государственное автономное образовательное учреждение
высшего образования «Санкт-Петербургский политехнический
университет Петра Великого»

Адрес: 29, Политехническая ул., Санкт-Петербург, 195251 Россия

Телефон: +7 812 290-9645

E-mail: m.samsonova@spbstu.ru

Подпись Самсонова М.Г.
УДОСТОВЕРЯЮ
Ведущий специалист
по кадрам
«18» 01

