

## ОТЗЫВ

официального оппонента на диссертационную работу

**Шмакова Сергея Анатольевича**

«Разработка биоинформатического подхода для поиска новых CRISPR-Cas систем»,  
предоставленную на соискание степени кандидата биологических наук по специальности  
03.01.09 – «Математическая биология, биоинформатика»

CRISPR-cas «иммунные системы» бактерий и архей уникальны по тому вниманию, которое уделяется им научным сообществом, прогрессу достигнутому в понимании основных механизмов их функционирования и скорости внедрения полученных результатов в прикладные технологии. Этапным в исследовании CRISPR-cas систем стало обнаружение белка Cas-9, который способен обнаруживать специфическую мишень и разрезать её. Это быстро нашло широчайшее применение в биотехнологии и даже в практической медицине, а скрининг новых микробных изолятов на предмет поиска новых CRISPR-Cas систем с новыми свойствами стал одним из наиболее **актуальных направлений** геномных исследований. Шесть новых систем второго класса обнаружено в рамках рецензируемой работы. Её фундаментальная **научная значимость** обусловлена также масштабностью предпринятого поиска, результаты которого весомо свидетельствуют о том, что основные системы CRISPR-Cas адаптивного ответа клеток, по-видимому, можно считать уже обнаруженными и, следовательно, существующее разнообразие можно считать репрезентативным для понимания не только функциональности и эволюционного разнообразия, но и для оценки защитного потенциала этих систем для бактерий и архей.

Диссертация изложена на 104 страницах машинописного текста и содержит разделы Введение, Обзор литературы, Материалы и Методы, Результаты и Обсуждение, Заключение и отдельный раздел «Результаты исследования», который содержит список выносимых на защиту положений. Диссертация иллюстрирована 16 ёмкими рисунками и двумя таблицами. Список цитируемой литературы включает 211 ссылок, а в Приложении даны ссылки на доступные в сети Интернет материалы опубликованных статей.

В очень объёмном **Введении** даётся базовое представление о механизмах формирования CRISPR кассет и классификация этих систем. В отдельных главах сформулирована цель и задачи исследования, приводится информация о научной новизне и практической значимости работы, отражён личный вклад соискателя, указаны положения, выносимые на защиту и степень их опубликованности.

В **Обзоре литературы** более подробно описаны процессы адаптации, т.е. механизма встраивания нового спэйсера в CRISPR кассету, биогенеза, т.е. экспрессии и созревания crRNA и интерференции, т.е. разрезания чужеродной ДНК или РНК. Очень хорошо описана хронология накопления знаний в этой области, информация о критериях, используемых для классификации CRISPR-Cas систем и представления об их эволюционном происхождении. В отдельной главе изложены пути практического применения этих систем. Обзор информативен, но только частично соответствует кругу задач, решаемых в диссертационной работе. Отражая современное представление о структурно-функциональной организации CRISPR-Cas систем, в нём, тем не менее, отсутствует раздел, посвящённый анализу биоинформатических методов, используемых для поиска CRISPR-Cas систем, перспективам дальнейшего поиска и о закономерно возникающих при этом проблемах. Это странно, т.к. рецензируемая работа представлена Сергеем Анатольевичем на соискание степени кандидата биологических наук по специальности «Математическая биология, биоинформатика» и именно биоинформатический поиск и анализ являются главными её составляющими.



Раздел **Материалы и Методы** включает всего 6 глав, в первой из которых указаны базы данных нуклеотидных последовательностей, использованные для поиска новых CRISPR-Cas систем. Кроме общедоступного депозитария NCBI, была создана представительная, актуальная на март 2016 г. индивидуальная коллекция архейных и бактериальных геномов. Перед включением в эту базу данных, частично аннотированные геномы были заново аннотированы с использованием MetaGeneMark\_v1.mod. Созданный ресурс включал нуклеотидные последовательности ~50,000 полностью или частично секвенированных геномов, кодирующих 182,301,555 белков, т.е. был достаточно обширным для поиска новых генов. В следующих 5 главах методической части перечислены стандартные алгоритмы и программные пакеты, использованные для решения поставленных задач. Для поиска гомологов генов *cas1* был использован алгоритм BLAST, а CRISPR кассеты искали программами CRISPRfinder и PILER-CR. Открытые рамки считывания были аннотированы MetaGeneMark\_v1.mod и RPS-BLAST. Для поиска доменов в белках была использована программа HHpred и CD-search, а для множественного выравнивания - MUSCLE и MAFFT. Для кластеризации последовательностей найденных гомологов использовали NCBI BLASTCLUST и UCLUST, а для филогенетического анализа – FastTree. Для сравнительного анализа кластеров были использованы программы HHSEARCH, UPGMA, HHALIGN и пакет программ PAML. Поиск протоспейсеров осуществляли с помощью алгоритма MEGABLAST.

Раздел **Результаты и обсуждение** состоит из четырех частей, в первой из них описан поиск новых CRISPR-Cas систем второго типа. Согласно очень краткому тексту, он осуществлялся в 2 этапа, вначале были собраны сайты, содержащие потенциальные ортологи генов *cas1* и участки, обладающие структурными свойствами CRISPR кассет. Затем было исследовано их генетическое окружение на предмет поиска крупных белков с доменами, типичными для Cas-белков. При этом участки, кодирующие белки с доменами, не свойственными известным функциям CRISPR-Cas систем были отброшены. Т.е. ограничениями использованного поискового подхода были: ко-локализация с ортологами *cas1* или CRISPR кассетами, крупный размер гена и потенциальная принадлежность белков к нуклеазам или другим белкам процессинга нуклеиновых кислот. Результаты суммированы в виде 4 рисунков, отражающих схему поиска новых CRISPR-Cas систем, их классификацию, доменную структуру новых эффекторных белков и 2 филогенетических дерева для эффекторов типов V-U и VI-B. Отмечается, что отличительной особенностью систем II и V типов является наличие RuvC-подобных нуклеазных доменов, а классификационными признаками для систем подтипа V-U стало отсутствие гена *cas1*, меньший размер эффекторных белков и большая гомология с белками TnpB. Отличительной особенностью систем VI типа является наличие эффекторного белка с двумя доменами HEPN.

Вторая часть результативного раздела посвящена оценке разнообразия CRISPR-Cas систем второго класса и построению их новой классификации. Эти данные суммированы в Таблице 2, которая отражает явное доминирование систем II типа и практически полное отсутствие систем второго класса у архей. В результате классификационного анализа предложено три типа и как минимум 10 подтипов CRISPR-Cas систем 2 класса, что, безусловно, является весомым результатом работы. Делается вывод о том, что «выполнена исчерпывающая оценка разнообразия систем второго класса показывающая распространение и дающая количественную оценку этих систем среди бактерий и архей». В этой связи важно вспомнить, что гены, кодирующие мембранные белки, были отсеяны на одном из первых этапов поиска новых CRISPR-Cas систем. Однако, в системах VI-B1 были обнаружены белки с трансмембранными доменами. Не исключено, что эти системы могут быть ассоциированными с мембран, что не типично для CRISPR-Cas систем и может вскрывать их неизвестное пока свойство, формируя поле для нового поиска.



Вопросы эволюционного возникновения CRISPR-Cas систем обсуждаются в третьей части результативного раздела. Предложенные гипотетические схемы базируются на предположении о независимом происхождении эффекторов разных типов и подтипов систем второго класса и представляются вполне разумными.

В последней части обсуждаются потенциальные способы применения новых CRISPR-Cas систем. К ним, безусловно, относятся возможность использования эффекторов VI типа для таргетирования РНК, принципиальная доступность более гибкой стратегии в выборе типа разрезания и систем, зависящих или независящих *tracrP*НК. Если введение Cas9 в лабораторную практику стало этапным для прикладного использования CRISPR-Cas систем, а открытие эндонуклеазы Spf1 предоставило более простой и более специфичный способ редактирования геномов, то новые системы на основе Cas13a открывают перспективы прямого воздействия на транскриптом или, что может быть даже более перспективным, специфического подавления роста целевых клеток.

Полученные результаты суммированы в кратком **Заключении**, в котором, кроме констатации факта обнаружения шести новых подтипов во втором классе CRISPR-Cas систем, делается акцент на том, что новые системы этого класса попадают всего в две категории: эффекторы, использующие RuvC подобную нуклеазу для гидролиза ДНК и эффекторы, использующие HEPN РНКазный домен для атаки РНК. Важно, что выявленные в работе особенности CRISPR-Cas систем этого класса свидетельствуют об их эволюционном происхождении из мобильных элементов, изначально кодирующих разные нуклеазы. Важно также, что обнаруженные в работе системы имеют новые комбинации функциональных свойств, что перспективно для их технологического использования.

Суть проведённой работы отражена в последнем разделе **Результаты исследования**. Он содержит шесть утверждений, но не все сформулированы в виде выводов, традиционно ожидаемых в конце диссертационной работы. В п. 4, например, можно было бы указать диспропорцию распространения CRISPR-Cas систем у бактерий и архей, а в п. 6 указать преимущества новых систем для биотехнологического применения. **Недостатками работы** также являются: (1) отсутствие раздела, содержащего описание и анализ биоинформатических методов, использованных ранее для поиска CRISPR-Cas систем, (2) отсутствие какого-либо комментария по поводу обнаружения белков с потенциальным трансмембранным модулем при условии, что образцы, содержащие гены мембранных белков, были удалены из анализируемых наборов, а также (3) чрезвычайно большое число текстуальных погрешностей (в среднем больше одной на страницу), которое может быть прощательным только для человека с ограниченным знанием русского языка. Ярким примером является неоднократное использование словосочетания «соседский ген» (т.е. ген соседа) вместо «соседний ген». Примером необоснованной транслитерации является «билобная структура», а примером громоздкого и несогласованного предложения: «В данных научных работах была показана возможность запрограммировать Cas9 с помощью sgPНК (...) чтобы внести двуцепочечный разрыв в ДНК, что может быть использовано для внесения инделей используя механизм репарации ДНК не гомологичного соединения концов или добавления нового материала ДНК используют ДНК шаблоны и механизм репарации по гомологии». Текст автореферата отредактирован гораздо лучше, но и в нём присутствуют такие словосочетания как *распознавательная доля* (распознающая или распознаваемая?). Достоинством работы, наоборот, являются ссылки на иллюстративный материал опубликованных статей.

В целом, диссертационная работа Сергея Анатольевича является цельным, хорошо спланированным и законченным в рамках поставленных задач исследованием. Она



выполнены в двух ведущих лабораториях, внесших основной вклад в понимание структурной организации и механизмов функционирования CRISPR-Cas систем. Фундаментальная значимость и достоверность полученных результатов не вызывает никаких сомнений и подтверждена публикациями в высокорейтинговых журналах. Автореферат соответствует содержанию диссертации. По актуальности темы, объему и важности проведенных исследований и сделанных выводов работа, несомненно, соответствует требованиям «Положения о порядке присуждения учёных степеней», утвержденного Постановлением Правительства Российской Федерации от 24 сентября 2013 г. №842, с изменениями Постановления Правительства Российской Федерации от 21 апреля 2016 года №335, предъявляемым к диссертациям на соискание ученой степени доктора наук, а сам диссертант, Шмаков Сергей Анатольевич заслуживает присуждения учёной степени кандидата биологических наук по специальности 03.01.09 – "Математическая биология, биоинформатика".

Зав. Лабораторией функциональной геномики и клеточного стресса  
Института биофизики клетки РАН  
д.б.н., профессор



Озолин О.Н.

Адрес: 142290, г. Пущино, ул. Институтская, д. 3,  
Телефон: +7(496)773-91-40  
E-mail: [ozoline@rambler.ru](mailto:ozoline@rambler.ru)

