

## Отзыв

официального оппонента на диссертационную работу

Кулемина Николая Александровича

на тему: «Биоинформационический анализ данных высокопроизводительного генотипирования в применении к поиску маркеров спортивной успешности»,

представленную на соискание ученой степени  
кандидата биологических наук по специальности

03.01.09 - математическая биология, биоинформатика.

Диссертационная работа Кулемина Николая Александровича посвящена поиску генетических маркеров, характеризующих наиболее успешных профессиональных спортсменов, на основе биоинформационического анализа данных высокопроизводительного генотипирования. Выбранная автором тематика представляет интерес как с практической точки зрения, так и с точки зрения фундаментальных научных исследований. Практическая значимость работы объясняется ценностью ее результатов для дальнейшего анализа генетически заложенных способностей в молодых спортсменах и возможностью подбора индивидуальных параметров тренировочного процесса, на основе полученных выводов работы. В связи с этим, разработка новых методов биоинформационического анализа данных высокопроизводительного генотипирования и успешное их применение для поиска маркеров, характеризующих наиболее успешных профессиональных спортсменов, имеет большое значение для персонализированного подхода в спорте.

Для поиска генетических маркеров автором оптимизирован алгоритм обработки данных, получаемых с использованием технологий ионного полупроводникового секвенирования и высокоплотной чип-гибридизации.

Это выгодно отличает работу Николая Александровича на фоне остальных подобных исследований, поскольку, во-первых, позволило исключить многие типичные ошибки при анализе данных высокопроизводительного генотипирования. К примеру, разработанный альтернативный протокол обработки данных ионного полупроводникового секвенирования позволил дополнительно корректно идентифицировать около 270 тысяч нуклеотидных позиций при полноэкзонном методе анализа генетической информации. Во-вторых, использование нового протокола позволяет рассчитывать на получение новых результатов при анализе других фенотипически различных групп. Важно отметить, что из интереса практического применения результатов своей работы автор включил в изложение описание собственного разработанного программного продукта для индивидуальной аннотации медицински информативных генетических вариантов.

Диссертация построена по стандартному плану и включает в себя введение, 3 главы, выводы и список литературы. Общий объем диссертации, включая 14 рисунков и 19 таблиц, составляет 130 страниц, в том числе список литературы включает 118 наименований.

Во введении дана общая характеристика работы, сформулированы цели и задачи исследования. Глава 1 представляет собой обзор литературы о современных методиках получения генетической информации (секвенирование и чип-гибридизация), а также о биоинформационических подходах к последующему анализу полученных данных.

Глава 2 включает описание материалов и методов, использованных в работе. Полученные результаты описаны в Главе 3, которая содержит четыре раздела. Первый раздел посвящен разработке и оптимизации алгоритма обработки результатов ионного полупроводникового секвенирования. Показано, что предложенный автором протокол анализа таких данных, состоящий из нескольких свободно распространяемых программных пакетов, позволяет в среднем на 0,5% точнее определять

геномные состояния. Второй раздел посвящен проблемам технического соответствия результатов чип-гибридизационного анализа и секвенирования. Автором предложен программный протокол решения поставленной задачи, что позволяет избегать технических ошибок при проведении массовых ассоциативных исследований. В третьем разделе приводятся примеры проведенных авторов ассоциативных исследований с применением разработанных методик. В процессе исследований были проанализированы геномы ведущих спортсменов России и были обнаружены новые генетические маркеры, которые значимо отделяют ведущих спортсменов-стайеров и спринтеров. Четвертый раздел полностью посвящен разработке программного обеспечения для автоматической медицинской аннотации геномных состояний с использованием открытых источников данных, а также созданной модели по неинвазивному определению состава мышечных волокон на основе генетической информации.

Результаты, описанные в диссертации, опубликованы в четырех научных статьях в международных рецензируемых журналах из списка ВАК, а также в одном российском журнале из списка ВАК. Основные результаты были доложены автором на 10 международных конференциях, 8 из которых были опубликованы в сборниках трудов конференций. Автореферат полностью отражает все основные выводы и результаты работы.

К достоинствам рецензируемой работы важно отнести успешность применения результатов современного высокопроизводительного секвенирования. Кроме того, разнообразие методик биоинформационического анализа данных генотипирования, говорит о высоком качестве исследований.

Однако работа не лишена и недостатков. Так автор заметно часто забывает цитировать литературные источники, на которые явным образом ссылается в тексте. К примеру, на странице 98, говорится что «В ходе

проведенных ранее исследований были сформированы списки кандидатных полиморфизмов, являющихся наиболее перспективными для поиска ассоциаций». Данный список автор использует в своей работе в качестве кандидатных полиморфизмов, однако ссылки на работу, где можно было бы уточнить детали формирования списка кандидатных генов, не приводятся.

Также, автор слишком углубляется в изложение «Обзора литературы». Так, к примеру, приводятся особенности построения fastq-файлов: какие в них есть строки, какие разделители. Все это говорит о, безусловно, высоком уровне теоретической подготовки автора и, безусловно, эти знания необходимы автору для успешной реализации исследования. Однако такое подробное описание кажется излишним с учетом значительного числа публикаций на данную тему в соответствующей литературе. Скорее всего, значительное внимание, уделенное базовым вопросам, привело к тому, что автор не успел уделить должного внимания разработанного им программного инструмента комплексной аннотации медицински значимых генетических вариаций. Данные разработанные автором инструменты по своей общности превосходят практический интерес маркеров, характеризующих наиболее успешных профессиональных спортсменов.

Однако вышеперечисленные недостатки не являются принципиальными и не снижают высокого качества представленной работы и полученных результатов. Несомнены также квалификация автора как исследователя в области биоинформатики.

Изучение текста диссертации Кулемина Н.А. на тему: «Биоинформационический анализ данных высокопроизводительного генотипирования в применении к поиску маркеров спортивной успешности» показало, что работа носит законченный характер, обладает актуальностью, новизной, а также теоретической и практической значимостью. Работа полностью соответствует требованиям,

предъявляемым к кандидатским диссертациям, согласно пункту 9 Положения «О порядке присуждения ученых степеней» (Постановление правительства Российской Федерации № 842 от 24 сентября 2013 г.). Её автор, Н.А.Кулемин заслуживает присуждения ученой степени кандидата биологических наук по специальности 03.01.09 – «Математическая биология, биоинформатика».

## Кандидат биологических наук,

Заведующий лабораторией постгеномных исследований  
Федерального государственного бюджетного учреждения науки  
Института молекулярной биологии им. В.А. Энгельгардта

Российской академии наук

(ФГБУН ИМБ РАН)

119991, г. Москва, ул. Вавилова, 32

[rhizamoeba@mail.ru](mailto:rhizamoeba@mail.ru)

+7-499-135-23-91

/ Кудрявцева Анна Викторовна /

Учрежденное сенатом Училище РАН  
Докторант А.Н.

