

ФАНО РОССИИ



Федеральное государственное бюджетное учреждение науки Институт общей генетики им. Н.И. Вавилова Российской академии наук (ИОГен РАН)

ул. Губкина, д. 3, г. Москва, ГСП-1, 119991
Тел.: (499) 135-62-13, (499) 135-20-41
Факс: (499) 132-89-62

E-mail: iogen@vigg.ru
http: www.vigg.ru

22.11.2016 №112 02-14/4

На № _____

“УТВЕРЖДАЮ”

Зам. директора по научной работе Института общей
генетики им. Н. И. Вавилова Российской академии
наук д.б.н., проф. С.К.Абилев

“21” ноября 2016 г.

ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

о научно-практической ценности диссертации Кулемина Николая Александровича на тему «Биоинформатический анализ данных высокопроизводительного генотипирования в применении к поиску маркеров спортивной успешности», представленной на соискание ученой степени кандидата биологических наук по специальности 03.01.09 – математическая биология, биоинформатика.

В настоящее время в результате развития технологий высокопроизводительного секвенирования и других методов масштабного генотипирования изучение генетической детерминации фенотипических особенностей основываются на обработке больших массивов данных, требующих биоинформатического анализа. Диссертация на соискание ученой степени кандидата биологических наук Н.А.Кулемина посвящена разработке комплексного протокола, предназначенного для первичной биоинформатической обработки и анализа данных генотипирования (ионного полупроводникового секвенирования и чип-гибридизации),

проведения ассоциативных исследований и аннотации выявленных генетических вариантов с фенотипическими признаками.

Для достижения поставленной цели автором работы решены три задачи: оптимизирован алгоритм обработки геномных данных, получаемых с использованием технологий ионного полупроводникового секвенирования и высокоплотной чип-гибридизации; проведен анализ генетических маркеров, ассоциированных с показателями выносливости, оксидативного стресса, эритропоза и структуры мышечных волокон у спортсменов на основе полученных генетических данных; разработан программный инструмент для комплексной аннотации медицински значимых генетических вариаций, включающий анализ известных однонуклеотидных полиморфизмов и мутаций, размещенных в доступных базах данных, анализ потенциальной патогенности редких вариантов и впервые выявленных мутаций.

Цели и задачи диссертации четко сформулированы и адекватно отражают тему работы.

Диссертационная работа построена по традиционному плану, содержит введение, три главы (обзор литературы, материалы и методы, и результаты и обсуждение), заключение и выводы. Общий объем рукописи диссертации составляет 130 страниц. Работа иллюстрирована 14 рисунками и 19 таблицами, содержит библиографический список из 118 наименований.

Обзор литературы рассматривает применение биоинформатических методов в современных генетических исследованиях и демонстрирует знание автором литературы по теме исследования. Раздел завершается обоснованием поставленных в диссертационной работе задач.

Раздел «Материалы и методы» включает описание исследованных групп, методов генетического анализа и биомедицинских исследований, описание использованного программного обеспечения и аппаратных средств.

Раздел «Результаты и обсуждение» включает подразделы, соответствующие поставленным задачам исследования.

Основные результаты диссертации изложены в 5 публикациях, соответствующие перечню ВАК рецензируемых научных журналов, и доложены на различных российских и международных конференциях. Основные положения, выносимые на защиту, сформулированы четко.

В диссертационной работе проанализирован значительный массив экспериментальных данных. Для проведения исследования были использованы генетические данные 17668 человек, из которых 3543 являются профессиональными спортсменами различных стран мира. Для 80 ведущих атлетов сборной России было проведено измерение объема максимального потребления кислорода. В анализе мышечной структуры приняли участие 40 атлетов, относящихся к группе выносливости. Для 61 спортсмена-стайера был проведен классический тест состава венозной крови (гемоглобин, моноциты, эритроциты и прочие показатели). У 1444 спортсменов сборной России были протестированы уровни различных показателей метаболизма.

Для дополнительного контроля анализировались данные генотипирования 11521 образцов, размещенных в банке данных Illumina iControlDB. Для тестирования программного протокола использовались результаты проекта «Genome In A Bottle», который является общепринятым стандартом для сравнения различных методов секвенирования геномной ДНК человека.

Актуальность. Высокопроизводительное секвенирование и генотипирование на чипах порождает большие массивы данных, для биоинформатической обработки которых разрабатываются различные подходы. Результатом биоинформатической обработки является определение генетических вариантов отдельных индивидов. Следующим этапом исследования связи генотипа и фенотипа является определение связи между генетическими вариантами и фенотипическими особенностями. Выявленные генетические варианты, ассоциированные с изучаемыми особенностями, подлежат аннотации (описанию функций генов и возможной функциональной значимости выявленных в нем полиморфизмов). Для обработки информации, получаемой на разных этапах исследования, разрабатываются различные биоинформатические протоколы, реализованные в виде компьютерных программных с различными вариантами пользовательских настроек. Оптимизация таких протоколов, разработка комплексных программ, позволяющих обрабатывать информацию о результатах генотипирования, получаемую в разных форматах на разных этапах исследования, проводить ассоциативные исследования и аннотацию выявленных генетических вариантов, ассоциированных с фенотипическими признаками, представляется весьма актуальной. В диссертационной работе Н.А.Кулемина разработаны комплексные программные протоколы, позволяющие оптимизировать обработку данных на всех указанных этапах.

Применение разработанных протоколов проведено на примере выявления полногеномного исследования ассоциации генетических вариантов с фенотипическими признаками, важными для успешности в спорте. Эта область исследований активно развивается в последние годы, получаемые в ней результаты перспективные для практического применения и весьма актуальны.

Научная новизна и значимость полученных результатов. Разработаны биоинформатические алгоритмы, предназначенные для увеличения точности обработки данных, получаемых при высокопроизводительном ионном полупроводниковом секвенировании, а также модифицирован алгоритм обработки данных чип-гибридизации. Алгоритмы реализованы в виде программного продукта, позволяющего совместно использовать информацию, полученную с использованием разных технологий (секвенирования и генотипирования на чипах) в одном эксперименте. Разработанные алгоритмы позволяют при анализе полноэкзомных данных ионного

полупроводникового секвенирования дополнительно корректно идентифицировать около 270 тысяч полиморфизмов.

Разработан программный продукт, позволяющий проводить индивидуальную аннотацию медицински информативных генетических вариантов, полученных различными методами высокопроизводительного генотипирования, перспективный для применения в развитии персонализированной медицины.

С использованием разработанных оптимизированных алгоритмов проанализирован массив данных высокоплотной чип-гибридизации для более 1000 профессиональных спортсменов Российской Федерации и выявлены новые ассоциации различных фенотипических показателей с генетическими маркерами.

Впервые показано, что полиморфизм rs4880 в гене митохондриальной супероксиддисмутазы 2 *SOD2* ассоциирован с устойчивостью спортсменов к оксидативному стрессу и изменением уровней креатинкиназы, креатинина и иных маркеров мышечного повреждения.

Показана ассоциация полиморфизма rs11091046 в гене рецептора ангиотензина II типа 2 *AGTR2* со структурой мышц и показателями потребления кислорода, и полиморфизма rs1572312 гена ядерного фактора I A *NFIA-AS2* с уровнем гемоглобина в крови и, как следствие, с аэробными возможностями спортсменов.

Отдельную практическую ценность представляет разработанный алгоритм, позволяющий на основе комплексного анализа полученных генетических данных осуществлять оценку индивидуальных медицинских рисков. Использование алгоритмов такого рода приближает внедрение персонализированной медицины в реальную клиническую практику.

Впервые была разработана и апробирована модель для неинвазивной оценки структуры мышечных волокон на основании генетического анализа. Точность разработанного прототипа модели индивидуальной комплексной оценки структуры мышечных волокон составила 81% в исследованных группах профессиональных спортсменов.

Рекомендации по использованию результатов диссертации. Полученные результаты могут быть использованы при чтении спецкурсов на факультетах биологии, в медицинских ВУЗах, на курсах повышения квалификации медицинских работников, биологов и в спортивных учреждениях. С другой стороны, разработанные алгоритмы могут послужить базисом для проведения последующих молекулярно-генетических и биоинформатических исследований по поиску наследственных факторов, влияющих на важные для спорта физиологические характеристики и иные фенотипические признаки. Исследования такого типа активно проводятся как в Институте общей генетики РАН, так и в других ведущих профильных НИИ: Институте

молекулярной генетики РАН, Институте молекулярной биологии РАН, Институте цитологии и генетики СО РАН, Институте биоорганической химии РАН, Институте проблем передачи информации РАН, факультете биоинженерии и биоинформатики МГУ.

Общие замечания. В качестве замечания можно отметить использование автором в ряде случаев некорректной терминологии. Так, на стр. 8, 16, 24 автореферата и 120 рукописи диссертации используется термин «европеоидный этнос» или «европейский этнос». В использованных словосочетаниях объединены антропологический (европеоидный) и географический (европейский) термины с культурным (этнос). Этнос (или этническая группа) характеризуются рядом культурных особенностей, таких как язык или самосознание, но не антропологических или географических. В данном контексте следовало бы использовать термин «популяции».

К стилистическим огрехам рукописи следует отнести использование термина «аллель» то в мужском, то в женском роде. Исторически сложилось так, что московская генетическая школа считает «аллель» словом мужского рода, а ленинградская (ныне Санкт-Петербургская) - женского. Оба подхода правомерны, но в одной рукописи следовало бы придерживаться единообразия.

В генетической литературе принято приводить названия генов, о которых идет речь, при их первом упоминании в тексте. Однако автором рукописи эти названия не приведены. Исключением является ген *SOD2*, но и его название приведено в фразе, стилистически некорректной: «Одним из кандидатных генов, вариабельность которого может быть ассоциирована с повреждением мышц, является супероксид дисмутаза 2 типа (*SOD2*)». Супероксид дисмутаза является ферментом, а не геном. Хотя по смыслу понятно, что речь идет о гене, но перед названием фермента слово «ген» следовало бы вставить. Сделанные замечания являются исключительно стилистическими и никак не снижают научной ценности работы.

ЗАКЛЮЧЕНИЕ

Диссертационная работа Кулемина Николая Александровича на тему «Биоинформатический анализ данных высокопроизводительного генотипирования в применении к поиску маркеров спортивной успешности» является законченной научно-исследовательской работой, выполненной на высоком научном уровне, сделанные выводы обоснованы, базируются на большом экспериментальном материале, который был корректно интерпретирован. Автореферат отражает основные положения диссертации. По актуальности темы, научному уровню, теоретической и значимости результатов диссертация отвечает требованиям п. 9 «Положения о порядке присуждения ученых степеней», утвержденного постановлением

Правительства Российской Федерации № 842 от 24 сентября 2013 г., предъявляемым к кандидатским диссертациям, а ее автор заслуживает присуждения ему искомой степени кандидата биологических наук по специальности 03.01.09 – математическая биология, биоинформатика.

Отзыв обсужден на заседании научного семинара отдела геномики и генетики человека ИОГен РАН (Протокол № 16-11/11/16 от “16” ноября 2016 г.).

Отзыв подготовила:

Зав. лаборатории анализа генома ИОГен РАН

д.б.н.

С. А. Боринская

21 ноября 2016 г.

Подпись заведующего лабораторией анализа генома ИОГен РАН, доктора биологических наук Светланы Александровны Боринской заверяю

Ученый секретарь ИОГен РАН

Д.б.н. О. А. Огаркова

