

# Экспертиза диссертации

**Автор:** Кулемин Николай Александрович

**Название:** Биоинформатический анализ данных высокопроизводительного генотипирования в применении к поиску маркеров спортивной успешности.

**Объект исследования:** Результаты генотипирования профессиональных спортсменов, полученных с использованием высокоплотной чип-гибридизации и ионного полупроводникового секвенирования.

**Цель исследования:** Разработка комплексного биоинформатического протокола, предназначенного для первичной обработки и анализа данных высокопроизводительного генотипирования, проведения ассоциативных генетических исследований и аннотации выявленных генетических вариаций в группах профессиональных спортсменов.

## **Задачи работы:**

1) Разработать оптимизированный алгоритм обработки геномных данных, получаемых с использованием технологий ионного полупроводникового секвенирования (ThermoFisher) и высокоплотной чип-гибридизации (Illumina).

2) Используя полученные данные высокопроизводительного генотипирования групп профессиональных спортсменов, осуществить анализ генетических маркеров, ассоциированных с показателями выносливости, оксидативного стресса, эритропоеза и структуры мышечных волокон.

3) Разработать программный инструмент для комплексной аннотации медицински значимых генетических вариаций, включающий анализ известных нуклеотидных полиморфизмов и мутаций, размещенных в доступных базах данных, анализ потенциальной патогенности редких вариантов и впервые выявленных мутаций.

**Методы исследования:** Оптимизация протоколов биоинформатической обработки первичных данных секвенирования и чип-гибридизации путем применения альтернативных программных алгоритмов и настройки параметров их работы. Поиск значимых генетических ассоциаций с использованием методов

корреляционного анализа. Разработка программного продукта и базы данных при помощи интеллектуального анализа текстов и приемов программирования.

**Обзор литературы:** Обзор литературы достаточно полный и содержит 118 ссылок на литературные источники, представленные как классическими работами, так и современными. Обзор покрывает все основные вопросы, связанные с темой работы.

**Основные результаты:**

- 1) Разработан альтернативный протокол обработки данных ионного полупроводникового секвенирования, точность которого выше штатного на 0,5%, что позволяет при полноэкзомном формате исследования дополнительно корректно идентифицировать около 270 тысяч нуклеотидных позиций.
- 2) Разработан алгоритм обработки данных высокоплотной чип-гибридизации, который позволяет исключать ошибки, связанные с неправильным позиционированием генетических вариаций и предоставляет возможность пользовательской настройки выходных параметров качества идентификации вариантов.
- 3) В результате ассоциативного исследования впервые было показано, что полиморфизм rs1572312 гена NFIA-AS2 для представителей европейского этноса достоверно ассоциирован с уровнем гемоглобина в крови и, как следствие, с аэробными возможностями спортсменов.
- 4) Впервые показано, что полиморфизм rs11091046 гена AGTR2 для представителей европейского этноса достоверно ассоциирован со структурой мышечных волокон и показателями потребления кислорода.
- 5) Подтверждено, что полиморфизм rs4880 гена SOD2 для представителей европейского этноса влияет на устойчивость организма спортсмена к оксидативному стрессу, а также ассоциирован с изменением уровня креатинкиназы и креатинина.
- 6) На основе всех имеющихся геномных данных ведущих российских спортсменов (более 1200 испытуемых) показано, что на настоящее время не

выявлено единого для большинства мировых популяций генетического профиля элитного спортсмена-стайера.

7) Разработан программный продукт, который позволяет осуществлять индивидуальную аннотацию медицински информативных генетических вариантов, полученных различными методами высокопроизводительного генотипирования.

8) Предложен прототип модели для индивидуальной комплексной оценки структуры мышечных волокон, точность которого составила 81% в группах профессиональных спортсменов.

**Публикации результатов:** Результаты работы опубликованы в 5 статьях в международных журналах. В совместных публикациях диссертантом выполнена работа по анализу данных генотипирования всех исследованных профессиональных спортсменов контрольной группы с использованием разработанных методов и проведён поиск значимых ассоциаций геномных состояний и исследуемых фенотипических проявлений.

**Заключение:** Исследование выполнено на высоком научном уровне, самостоятельно, достаточно полно опубликовано. Тема исследования полностью соответствует специальности «математическая биология, биоинформатика» и соответствует профилю Совета.

Председатель комиссии:



д.б.н. М.С.Гельфанд

Члены комиссии:



д.б.н. А.А.Миронов



д.ф.-м.н. В.Ю.Макеев

Москва, 06 октября 2016 года.