

Полногеномное секвенирование спортсменов: влияние нонсенс-мутаций на функциональные показатели

Булыгина Е.А.¹, Борисов О.В.^{2,3}, Валеева Е.В.^{1,4}, Семенова Е.А.^{1,2}, Ларин А.К.²,
Набиуллина Р.М.⁴, Мавлиев Ф.А.⁵, Ахатов А.М.⁵, Генерозов Э.В.², Ахметов И.И.^{2,4,6,7}

- 1 — Казанский (Приволжский) федеральный университет
420008, г. Казань, ул. Кремлевская, 18
- 2 — Федеральный научно-клинический центр физико-химической медицины Федерального медико-биологического агентства
119435, г. Москва, ул. Малая Пироговская, д. 1а
- 3 — Институт геномной статистики и биоинформатики, Университетская клиника Бонна
Venusberg-Campus 1, Building 11, 2nd Floor, 53127 Bonn, Germany
- 4 — Казанский государственный медицинский университет
420012, г. Казань, ул. Бутлерова, 49
- 5 — Поволжская государственная академия физической культуры, спорта и туризма
420010, г. Казань, Деревня Универсиады, д. 35
- 6 — Российский экономический университета им. Г.В. Плеханова
117997, г. Москва, Стремянный пер., 36
- 7 — Научно-исследовательский институт спорта, Ливерпульский университет им. Джона Мурса
Tom Reilly Building, Byrom Street, Liverpool L3 3AF, United Kingdom

Применение технологии полногеномного секвенирования предоставляет обширные возможности для обнаружения новых генетических маркеров, определяющих результативность спортсмена. Целью исследования является определение особенностей геномного профиля спортсменов высокого класса и выявление генетических вариантов, ассоциированных с некоторыми функциональными показателями, влияющими на спортивную успешность. Впервые проведено секвенирование геномов 20 профессиональных борцов татарской национальности. Найденные генетические варианты валидированы с использованием ДНК-микрочипов и/или импутации. Из всего спектра полиморфных вариантов отобраны мутации в гомозиготном состоянии, имеющие аннотацию «stop gain» (нонсенс-мутации) и показавшие ассоциацию со следующими фенотипами: мышечная масса, пиковая мощность нижних конечностей, время реакции и спринтерские способности. Частоты встречаемости мутантных аллелей таких вариантов сравнивались с частотами в других популяциях Волго-Уральского региона. В геномах борцов найдено около 11 млн полиморфных локусов, в среднем 3,62 млн однонуклеотидных замен и 617 тыс инделов на геном. Из них 347 вариантов потенциально вызывают преждевременную терминацию трансляции белка. Корреляция со спортивными фенотипами была выявлена для 6 нонсенс-мутаций (гены *ANKDD1B*, *SLC6A18*, *CCHCR1*, *VOPP1*, *ADAMTS12* и *ZACN*), в гомозиготном состоянии приводящих к достоверному изменению функциональных показателей спортсменов. У носителей замены *p.Tyr319** (rs7447815) в гене *SLC6A18* обнаружено снижение относительной пиковой мощности нижних конечностей. В то же время у борцов-татар была зарегистрирована меньшая частота мутантного аллеля в этом локусе по сравнению с башкирами и русскими. Также, согласно исследованиям GWAS, замена rs7447815 ассоциирована с риском развития остеоартроза и пониженной физической активностью. Количество полиморфных вариантов, найденных в геномах спортсменов, и доля нонсенс-мутаций в них соотносятся с результатами, описанными для других популяций. Нонсенс-мутация *p.Tyr319** (rs7447815) в гене *SLC6A18* является потенциальным маркером, ухудшающим значение относительной пиковой мощности нижних конечностей.

Ключевые слова: полногеномное секвенирование, нонсенс-мутация, борьба на поясах, мышечная масса, пиковая мощность, время реакции, скоростная способность.

Для цитирования: Булыгина Е.А., Борисов О.В., Валеева Е.В., Семенова Е.А., Ларин А.К., Набиуллина Р.М., Мавлиев Ф.А., Ахатов А.М., Генерозов Э.В., Ахметов И.И. Полногеномное секвенирование спортсменов: влияние нонсенс-мутаций на функциональные показатели. *Медицинская генетика* 2021; 20(4): 19-29.

DOI: 10.25557/2073-7998.2021.04.19-29

Автор для корреспонденции: Булыгина Е.А.; e-mail: boulygina@gmail.com

Финансирование. Работа выполнена за счет средств субсидии, выделенной Казанскому федеральному университету для выполнения государственного задания в сфере научной деятельности (№ 0671-2020-0058).

Конфликт интересов. Авторы заявляют об отсутствии конфликтов интересов.

Поступила: 23.03.2021.

Whole-genome sequencing of athletes: the effect of nonsense mutations on functional variables

Boulygina E.A.¹, Borisov O.V.^{2,3}, Valeeva E.V.^{1,4}, Semenova E.A.^{1,2}, Larin A.K.², Nabiullina R.M.⁴, Mavliev F.A.⁵, Akhatov A.M.⁵, Generozov E.V.², Ahmetov I.I.^{2,4,6,7}

- 1 — Kazan (Volga region) Federal University
18 Kremlyovskaya str., Kazan, 420008 Russian Federation
- 2 — Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency
Malaya Pirogovskaya str., 1a, Moscow, 119435, Russia
- 3 — Institute for Genomic Statistics and Bioinformatics, University Hospital Bonn
Venusberg-Campus 1, Building 11, 2nd Floor, 53127 Bonn, Germany
- 4 — Kazan State Medical University
49 Butlerova Str., Kazan, 420012, Russia
- 5 — Volga Region State Academy of Physical Culture, Sport and Tourism
35, Privolzhsky district, the Universiade Village, Kazan, 420010, Russia
- 6 — Plekhanov Russian University of Economics
Stremyanny lane, 36, Moscow, 117997, Russia
- 7 — Research Institute for Sport and Exercise Sciences, Liverpool John Moores University
Tom Reilly Building, Byrom Street, Liverpool L3 3AF, United Kingdom

The use of whole-genome sequencing technology provides extensive opportunities for the detection of new genetic markers that determine athletic performance. *The aim* of the study was to characterize the genome profile of high-class athletes and to identify genetic variants associated with some functional variables which can affect sports success. *Materials and methods.* For the first time, whole-genome sequencing of 20 professional belt wrestlers of Tatar nationality was carried out. Genetic variants were validated using microarray technology and/or imputation procedure. From the entire spectrum of variants those that were annotated as nonsense mutations, were found at the homozygous state and showed association with muscle mass, lower-extremity peak power, reaction time, and sprinting abilities were selected. The mutant allele frequencies of these variants were compared with the frequencies in other populations of the Volga-Ural region. *Results.* About 11 million polymorphic loci were found in the genomes of wrestlers, an average of 3.62 million single nucleotide polymorphisms and 617 thousand indels per genome. Of these, 347 variants potentially cause premature termination of protein translation. Correlation with sports phenotypes was found for 6 nonsense mutations (*ANKDD1B*, *SLC6A18*, *CCHCR1*, *VOPP1*, *ADAMTS12*, and *ZACN* genes), which in a homozygous state lead to significant changes in functional indicators of athletes. Wrestlers with *p.Tyr319** (rs7447815) substitution in the *SLC6A18* gene had decreased relative lower-extremity peak power. At the same time, a lower mutant allele frequency at this locus was registered in Tatar wrestlers compared to the Bashkirs and Russians. Also, according to the GWAS, the rs7447815 substitution is associated with the risk of osteoarthritis and decreased physical activity. *Conclusions.* The number of polymorphic variants found in the genomes of athletes and the proportion of nonsense mutations in them correlate with the results described for other populations. The nonsense mutation *p.Tyr319** (rs7447815) in the *SLC6A18* gene is a potential marker that reduces the relative peak power of the lower extremities.

Keywords: whole-genome sequencing, nonsense mutation, belt wrestling, muscle mass, peak power, reaction time, speed performance.

For citation: Boulygina E.A., Borisov O.V., Valeeva E.V., Semenova E.A., Larin A.K., Nabiullina R.M., Mavliev F.A., Akhatov A.M., Generozov E.V., Ahmetov I.I. Whole-genome sequencing of athletes: the effect of nonsense mutations on functional variables. *Meditinskaya genetika [Medical genetics]* 2021; 20(4): 19-29. (In Russ.).

DOI: 10.25557/2073-7998.2021.04.19-29

Corresponding author: E.A. Boulygina; **e-mail:** boulygina@gmail.com

Funding. This work was funded by the subsidy allocated to Kazan Federal University for the state assignment in the sphere of scientific activities (№ 0671-2020-0058).

Conflict of interest. Authors declare no conflicts of interest.

Accepted: 23.03.2021.

Высокопроизводительное секвенирование становится рутинным методом как в диагностике генетических заболеваний и оценке рисков их развития, так и в определении предрасположенностей к различным видам деятельности, в том числе к спорту. В настоящее время найдено около 200 генетических вариантов — маркеров быстроты, силы, выносливости и психологических особенностей [1-3]. Наиболее рас-

пространенными методами поиска новых ассоциаций «генотип-фенотип» является анализ генов-кандидатов и секвенирование на ДНК-микрочипах. Однако даже микрочипы с максимальной плотностью (более 4,5 млн однонуклеотидных полиморфизмов) детектируют состояние только известных маркеров, оставляя неизученной большую часть генома. На помощь может прийти полногеномное секвенирование, позволяющее

при достаточной глубине покрытия описать полный генетический профиль спортсмена. Данные микрочипов в этом случае могут служить основой для валидации обнаруженных вариантов и для импутации отдельных непокрытых локусов.

В данной работе мы представляем результаты полногеномного секвенирования спортсменов-борцов. Борьба – ациклический вид спорта с переменной интенсивностью, связанный со значительными мышечными напряжениями. Исход поединка зависит от специальной скоростно-силовой выносливости спортсмена, его координации, скорости реагирования и принятия решений, гибкости и других качеств, в большой степени обусловленных генотипом [4]. Целью нашего исследования является выявление особенностей геномного профиля спортсменов, а также поиск делегирующих мутаций и оценка их влияния на некоторые функциональные показатели, лежащие в основе спортивной успешности.

Материалы и методы

В исследовании приняли участие 20 профессиональных спортсменов-мужчин, специализирующихся в татарской национальной борьбе на поясах «Корэш» и имеющих спортивную квалификацию (первый взрослый разряд или кандидат в мастера спорта (КМС)). Средний возраст участников составлял $20 \pm 4,4$ лет (mean \pm SD), средний индекс массы тела – $24,6 \pm 2,7$ кг/м². Все спортсмены относят себя к казанским татарам и проживают в Республике Татарстан.

От всех участников исследования было получено информированное согласие на использование биологического материала в данной работе. Протокол исследования был одобрен этическим комитетом Федерального научно-клинического центра физико-химической медицины Федерального медико-биологического агентства России (протокол заседания №2017/04 от 04.07.2017). Исследование было проведено в соответствии с Хельсинкской декларацией Всемирной медицинской ассоциации.

Сбор образцов, секвенирование и поиск генетических вариантов

Забор венозной крови общим объемом 9 мл производился утром, натощак, в вакуумные пробирки с антикоагулянтом К2 ЭДТА. Образцы крови до экстракции ДНК хранили при температуре -20°C . Геномная ДНК была выделена с использованием набора Wizard Genomic DNA Purification Kit (Promega, США). ДНК-библиотеки были отсекуены на платформе Illumina HiSeq 2500 с использованием на-

бора HiSeq SBS Kit v4 (Illumina, США) в режиме парных прочтений длиной 125 п. н. со средней глубиной покрытия $9,9x \pm 5,9x$. Полученные риды были картированы на референсную сборку генома человека hg19 с использованием алгоритма BWA [5]. Поиск однонуклеотидных полиморфизмов (ОНП) в геномах был произведен с использованием Strelka2 [6]. Валидация вариантов проводилась путем генотипирования на ДНК-микрочипах плотностью более 900 тыс. ОНП с использованием HumanOmniExpress Bead-Chips (Illumina, США), а также импутации полученных генотипов с помощью Eagle, версия 2.4 [7], и Minimac4 [8]. Варианты были проаннотированы при помощи инструмента Annovar [9] по базам данных RefSeq, ClinVar, dbSNP 138 и gnomAD [10-13]. Функциональный анализ генов, несущих нонсенс-мутации (аннотация «stop gain»), был проведен с помощью PANTHER [14].

Функциональные тесты

Измерение мышечной массы спортсменов производилось методом биоимпедансного анализа на электронных весах Tanita MC 980 MA (Япония) утром натощак. Пиковая мощность нижних конечностей оценивалась в 30-секундном анаэробном тесте Вингейта на ножном эргометре. Относительные показатели мышечной массы и пиковой мощности получали путем нормализации абсолютных значений на вес борцов.

Время реакции определялось с помощью компьютерного теста «Светофор»: на экране компьютера демонстрируются световые сигналы разных цветов, и при появлении зеленого сигнала испытуемый должен как можно быстрее нажать на кнопку мыши. Интервалы между сигналами составляют от 0,5 до 5 с, первые пять попыток являются пробными и не регистрируются. Из следующих пяти попыток выбираются три лучшие, по ним вычисляется среднее время реакции.

Также спортсменам предлагалось самостоятельно оценить свои скоростные способности и ответить на вопрос, насколько хорошо они пробегают короткие дистанции. Варианты ответов «отлично», «хорошо», «удовлетворительно» записывались как 3, 2 и 1, соответственно.

Статистический анализ

Среди мутаций, приводящих к появлению преждевременного стоп-кодона, были отобраны варианты, по которым спортсмены имеют мутантный гомозиготный генотип. Был произведен анализ корреляции этих ОНП с функциональными показателями по критерию Спирмена по аддитивной и рецессивной генетическим моделям. Различия в фенотипических характеристиках между спортсменами-носителями мутант-

ных аллелей в гомозиготном состоянии и остальными борцами оценивались с использованием двустороннего *t*-критерия Стьюдента или непараметрического критерия Манна-Уитни для независимых выборок (только для результатов опроса). Частоты аллелей ОНП, влияющих на значение показателя в гомозиготном состоянии, проверялись на соответствие равновесию Харди-Вайнберга, а также на значимость различий с группами контроля по критерию хи-квадрат. Данными для сравнения послужили результаты полноэкзомного секвенирования, а также секвенирования на микрочипах представителей других народов Волго-Уральского региона (ВУР) (татары-неспорстмены, мишари, чуваша, башкиры, русские) [15, 16].

Нонсенс-мутации, показавшие ассоциацию в корреляционных тестах, были изучены на предмет наличия статистически достоверной связи со спортивными фенотипами в базе результатов полногеномного анализа ассоциаций UK Biobank и GWAS Catalog [17].

За критический уровень значимости во всех статистических тестах было принято значение $p < 0,05$. Все вычисления проводились в среде R, версия 3.4.3.

Результаты

По результатам полногеномного секвенирования и поиска вариантов в геномах борцов было найдено около 11 млн полиморфных локусов. Для образцов с глубиной покрытия от 12х среднее количество однонуклеотидных замен составляет около 3,62 млн на геном, инсерций и делеций – около 617 тыс. Если принимать во внимание глубину покрытия, эти показатели согласуются с количеством найденных генетических вариантов для других популяций Евразии [18-20]. В экзонах локализовано чуть более 0,5% мутаций, примерно столько же аннотировано как «патогенные» или «вероятно патогенные» [21]. Около 1,45 млн (13,2%) всех вариантов не представлено в базе данных dbSNP.

Общая схема отбора вариантов для дальнейшего анализа представлена на **рис. 1**.

В геномах спортсменов обнаружено 347 мутаций с аннотацией «stop gain» (в среднем 121 мутация на геном), потенциально приводящих к потере функции 298 белков. Данные белки входят в группы транспортеров веществ и молекул внутри клетки и за ее пределами, обонятельных и вкусовых рецепторов, кератинов, а также участвуют в иммунных реакциях, регуляции клеточного цикла, сперматогенезе и прочих процессах (**рис. 2**).

Из 347 нонсенс-вариантов были отфильтрованы локусы, генотипированные не для всех спортсменов, и отображены варианты, гомозиготные по мутантно-

му аллелю хотя бы у одного участника исследования. Из 86 таких вариантов корреляция со спортивными показателями была выявлена для 13 локусов по аддитивной модели и для 11 – по рецессивной; итого было обнаружено 19 уникальных вариантов, показавших значимую ассоциацию с физическими качествами (**табл. 1**). Из них отличия в средних значениях показателя для борцов с гетерозиготными и немутантными гомозиготными генотипами и для носителей мутант-

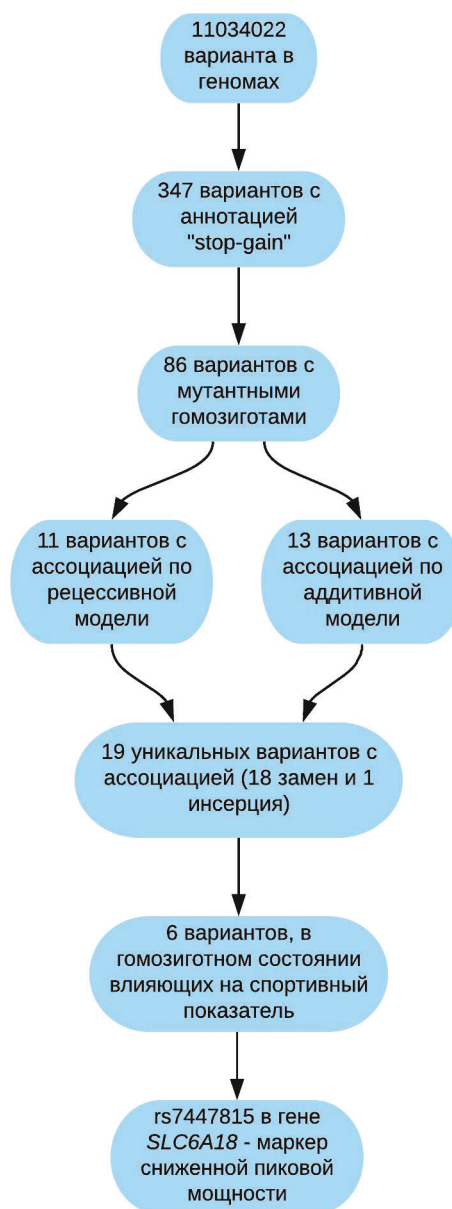


Рис. 1. Схема отбора вариантов для ассоциативного анализа.

ных гомозигот были выявлены для 6 полиморфных вариантов, локализованных в генах *ANKDD1B*, *SLC6A18*, *CCHCR1*, *VOPPI*, *ADAMTS12* и *ZACN* (табл. 2). Фенотипы GWAS из категорий опорно-двигательной, дыхательной, нервной систем, метаболизма жиров, углеводов и антропометрических показателей, ассоциированные с этими вариантами, представлены в табл. 3.

Обсуждение

В ходе изучения данных полногеномного секвенирования спортсменов-борцов было обнаружено, что количество мутаций, приводящих к появлению преждевременного стоп-кодона, а также паттерн их распределения в геноме согласуется с приведенными ранее результатами [22, 23]. Показано, что нонсенс-мутация *p.Trp480** в гене *ANKDD1B* ассоциирована с пиковой мощностью: образование стоп-кодона приводит к ухудшению этого показателя у борцов. Ген *ANKDD1B* кодирует малоизученный белок, имеющий в своем составе анкириновые повторы и «домен смерти». Известно, что мутации в подобных белках ассоциированы с онкологическими и неврологическими заболеваниями [24]. Частота мутантного аллеля

у спортсменов составляет 0,53 и значимо не отличается от таковой в контрольных выборках. Согласно базе данных UK Biobank, вариант rs34358 способствует увеличению безжировой массы тела, что не согласуется с обнаруженной нами ассоциацией, так как пиковая мощность напрямую зависит от мышечной массы, и мутантный аллель должен снижать оба эти показателя. К тому же эта мутация не показала ассоциацию с пиковой мощностью, нормализованной на вес борцов, что также может свидетельствовать о ложноположительном результате.

С относительной пиковой мощностью ассоциированы варианты в трех генах: *SLC6A18*, *CCHCR1* и *VOPP*. Полиморфизм *p.Tyr319** (rs7447815) в гене *SLC6A18*, кодирующем белок из семейства транспортеров Na^+/Cl^- -зависимых нейромедиаторов, снижает мощность нижних конечностей. Было показано, что при некоторых заболеваниях носительство минорного аллеля G в данном локусе может иметь протекторный эффект [25], однако в нашем случае полиморфизм ухудшает способности спортсменов, и частота мутантного аллеля у борцов значимо ниже, чем в выборках башкир и русских – 0,2 против 0,5 и 0,6, соответственно ($p < 0,01$). Также, согласно исследованиям GWAS, ва-

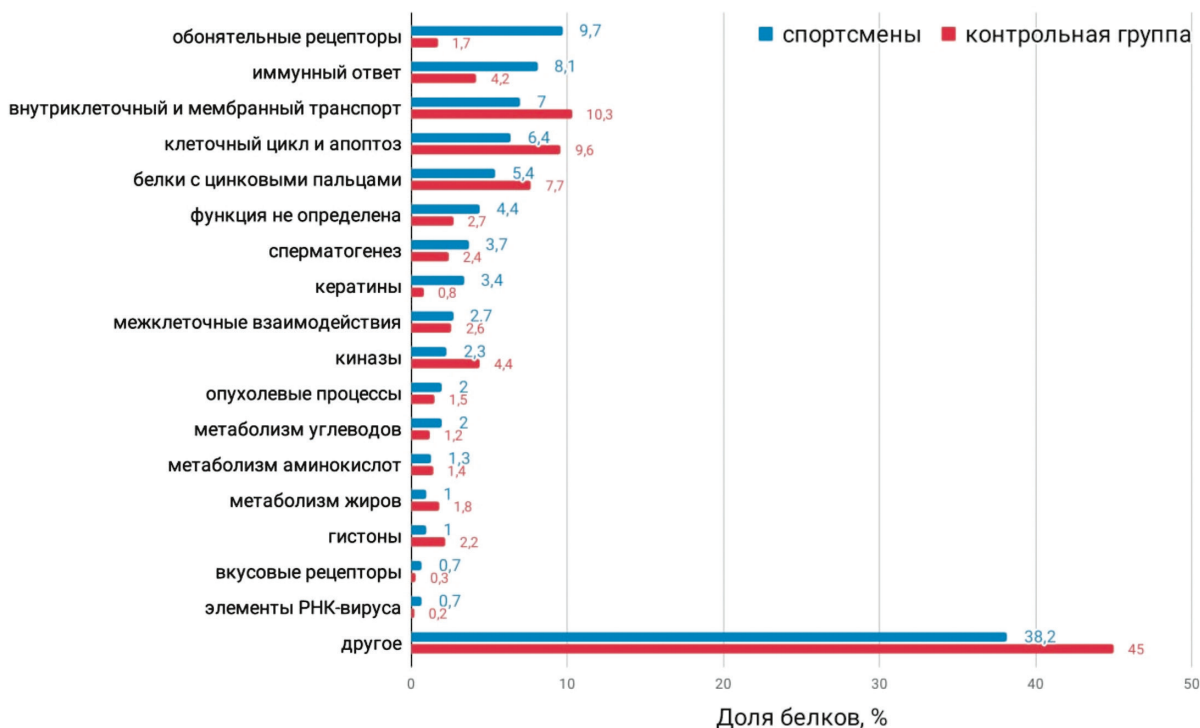


Рис. 2. Распределение функциональных групп белков, несущих нонсенс-мутации, в геномах спортсменов и в экзомах контрольной группы татар-неспортсменов.

Нонсенс-мутации в геномах борцов, ассоциированные со спортивными показателями

Ген	Вариант	Глубина покрытия позиции (mean ± SD)	Частота мутантного аллеля	Спортивный показатель	Коэффициент корреляции (p-value)	Модель наследования при анализе ассоциаций
<i>ANKRD36C</i>	rs74941794, A>C	40,3 ± 30,1	0,45 борцы 0,21 европейцы* 0,13 азиаты*	Время реакции	0,46 (0,04)	рецессивная
<i>C2orf83</i>	rs2176186, C>T	10,5 ± 6,2	0,45 борцы 0,21 татары [15]	Относительная мышечная масса	-0,45 (0,047)	рецессивная
<i>CCDC66</i>	rs150150392, T>TGGGGTAAGCA	5,5 ± 5,5	0,4 борцы 0,45 татары [15]	Скоростные способности	-0,45 (0,046)	аддитивная
				Относительная пиковая мощность	-0,45 (0,044)	рецессивная
<i>ADAMTS12</i>	rs1530507, A>T	10,4 ± 6,6	0,75 борцы 0,61 татары [15] 0,72 русские [15] 0,68 мишари, чуваши [15] 0,47 башкиры [15]	Скоростные способности	-0,67 (0,001)	рецессивная
<i>ANKDD1B</i>	rs34358, G>A	10,2 ± 5,5	0,53 борцы 0,54 татары [16] 0,56 русские [16]	Скоростные способности	-0,47 (0,035)	аддитивная
				Пиковая мощность	-0,45 (0,046)	рецессивная
				Время реакции	0,45 (0,046)	рецессивная
<i>SLC6A18</i>	rs7447815, C>G	8,3 ± 6,2	0,2 борцы 0,21 татары [15] 0,61 русские [15] 0,37 мишари [15] 0,32 чуваши [15] 0,5 башкиры [15]	Пиковая мощность	-0,47 (0,037)	аддитивная
				Относительная пиковая мощность	-0,48 (0,034)	аддитивная
<i>CCHCR1</i>	rs3130453, C>T	9,1 ± 6,4	0,55 борцы 0,47 татары [15] 0,4 русские [16] 0,55 мишари [15] 0,32 чуваши [15] 0,5 башкиры [15]	Относительная пиковая мощность	0,51 (0,021)	рецессивная
<i>VOPPI</i>	rs815957, G>A	7,4 ± 4,9	0,33 борцы 0,45 татары [15] 0,27 русские [15] 0,44 мишари [15] 0,32 чуваши [15] 0,33 башкиры [15]	Время реакции	-0,46 (0,041)	аддитивная
				Относительная пиковая мощность	0,46 (0,04)	рецессивная
<i>IDO2</i>	rs4503083, T>A	9,9 ± 6,6	0,15 борцы 0,26 татары [15]	Скоростные способности	0,48 (0,034)	аддитивная
<i>OR4C5</i>	rs77053100, G>A	19,8 ± 11,8	0,45 борцы 0,21 европейцы* 0,08 азиаты*	Индекс массы тела	-0,45 (0,046)	аддитивная
				Скоростные способности	-0,49 (0,027)	аддитивная

Ген	Вариант	Глубина покрытия позиции (mean ± SD)	Частота мутантного аллеля	Спортивный показатель	Коэффициент корреляции (p-value)	Модель наследования при анализе ассоциаций
<i>OR56A5</i>	rs61730422, G>A	10,4 ± 7,2	0,1 борцы 0,11 татары [15]	Относительная мышечная масса	-0,45 (0,049)	аддитивная
<i>EML3</i>	rs35156678, G>A	6,9 ± 4,7	0,25 борцы 0,34 татары [15]	Пиковая мощность	0,46 (0,042)	аддитивная
				Скоростные способности	0,61 (<0,01)	аддитивная
<i>PRB4</i>	rs12829245, G>A	8 ± 5,2	0,23 борцы 0,03 татары [15]	Время реакции	-0,46 (0,04)	аддитивная
<i>MAP2K3</i>	rs55796947, C>T	14,8 ± 9,5	0,5 борцы 0,5 татары [15]	Индекс массы тела	-0,47 (0,038)	аддитивная
				Пиковая мощность	-0,49 (0,027)	аддитивная
<i>ZACN</i>	rs1043149, C>T	8,9 ± 6	0,25 борцы 0,16 татары [15] 0,21 русские [16] 0,29 мишари [15] 0,18 чувашаи [15]	Относительная мышечная масса	0,46 (0,04)	рецессивная
<i>FUT2</i>	rs601338, G>A	6,4 ± 4,3	0,33 борцы 0,45 татары [15]	Относительная мышечная масса	-0,52 (0,019)	рецессивная
<i>KRTAP13-2</i>	rs877346, A>T	7,5 ± 5,5	0,48 борцы 0,45 татары [15]	Индекс массы тела	0,51 (0,022)	аддитивная
				Относительная мышечная масса	-0,60 (0,005)	аддитивная
<i>MAGEB16</i>	rs4829392, C>T	4,6 ± 3,9	0,4 борцы 0,63 татары [15]	Относительная пиковая мощность	-0,53 (0,016)	рецессивная (вариант на X хромосоме)
<i>H2BFM</i>	rs2301384, C>T	3,3 ± 2,6	0,3 борцы 0,37 татары [15]	Относительная мышечная масса	-0,48 (0,031)	рецессивная (вариант на X хромосоме)

*Для вариантов, по которым нет информации о частоте аллелей в контрольных популяциях ВУР [15, 16], приведена частота мутантного аллеля у европейцев и азиатов из базы данных dbSNP.

Таблица 2

Нонсенс-мутации в геномах борцов, в гомозиготном состоянии влияющие на спортивный показатель

Спортивный показатель	Ген	Мутация	Аллель, улучшающий показатель	p-value корреляционного теста
Пиковая мощность	<i>ANKDD1B</i>	rs34358, G>A	G	0,044
Относительная пиковая мощность	<i>SLC6A18</i>	rs7447815, C>G	C	0,025
	<i>CCHCR1</i>	rs3130453, C>T	T	0,022
	<i>VOPPI</i>	rs815957, G>A	A	<0,001
Время реакции	<i>VOPPI</i>	rs815957, G>A	A	0,032
Скоростные способности	<i>ADAMTS12</i>	rs1530507, A>T	A	0,004
Относительная мышечная масса	<i>ZACN</i>	rs1043149, C>T	T	0,003

Фенотипы GWAS, ассоциированные с нонсенс-мутациями

Ген	Мутация	Фенотип GWAS	Бета	p
<i>ANKK1B</i>	rs34358	Холестерин	-0,048	<0,001
		Безжировая масса левой ноги	0,033	<0,001
		Безжировая масса правой руки	0,012	<0,001
		Безжировая масса тела	0,165	<0,001
		Безжировая масса левой руки	0,012	<0,001
		Безжировая масса корпуса тела	0,075	<0,001
		Средний объем тромбоцита	0,013	<0,001
		Сахарный диабет, со слов участника проекта	0,058	<0,001
		Ожирение	0,055	<0,001
		Остеопороз, со слов участника проекта	-0,054	0,006
		Остеохондропатии	0,15	0,028
<i>SLC6A18</i>	rs7447815	Вес	0,831	0,013
		Физические нагрузки за последние 4 недели	-0,019	0,016
		Остеоартроз	0,022	0,017
		Остеоартроз, локализованный, первичный	0,033	0,033
		Сахарный диабет 1 типа с поражением почек	-0,331	0,045
<i>CCHCR1</i>	rs3130453	Объем форсированного выдоха за 1 секунду, лучший показатель	-0,012	<0,001
		Сахарный диабет 1 типа	0,165	<0,001
		Средний корпускулярный гемоглобин	-0,025	<0,001
		Сахарный диабет 2 типа	0,062	<0,001
		Физические нагрузки за последние 4 недели	-0,02	<0,001
		Остеопороз	0,039	0,033
<i>VOPPI</i>	rs815957	Врожденные сердечно-сосудистые аномалии	0,081	0,006
		Инсульт, диагностированный	0,056	0,007
		Инсульт, со слов участника проекта	0,057	0,011
		Безжировая масса левой ноги	-0,007	0,016
		Минеральная плотность пяточной кости	0,006	0,023
		Безжировая масса тела	-0,035	0,029
		Безжировая масса корпуса тела	-0,017	0,043
		Безжировая масса правой ноги	-0,006	0,043
<i>ADAMTS12</i>	rs1530507	Повышенное кровяное давление	0,016	0,010
		Общий холестерин	-0,04	0,011
		Холестерин липопротеинов низкой плотности	-0,04	0,011
		Эссенциальная гипертензия, со слов участника проекта	-0,1	0,015
		Фасциит	-0,065	0,017
		Сахарный диабет 2 типа	-0,03	0,024
		Ювенильный остеохондроз	0,248	0,032

Ген	Мутация	Фенотип GWAS	Бета	p
		Заболевания мышц, связок и фасций	-0,054	0,034
		Остеопороз	0,044	0,037
		Остеохондропатии	0,155	0,043
		Остановка сердца	0,109	0,047
ZACN	rs1043149	Вес	0,15	<0,001
		Жировая масса правой ноги	0,017	<0,001
		Жировая масса тела	0,09	0,002
		Остеoarтроз	0,027	0,028
		Сахарный диабет 2 типа с поражением глаз	0,107	0,044

риант rs7447815 повышает риск развития остеоартроза и ассоциируется с пониженной физической активностью.

Вариант *p.Trp78** (rs3130453) в гене *ССHCR1* повышает значение относительной пиковой мощности. Ген кодирует альфа-спиральный стержневой белок, участвующий в процессах стероидогенеза, регуляции цитоскелета и дифференциации мышечных волокон. Мутации в гене ассоциированы с псориазом [26] и диабетом 2 типа [27], а вариант rs3130453, согласно исследованиям GWAS, увеличивает риск остеопороза, снижает показатели дыхания и количество гемоглобина, что не может не сказываться негативно на спортивных успехах. Такими же взаимоисключающими выглядят результаты для варианта *p.Arg83** (rs815957) в гене *VOPP*. Согласно нашим данным, эта мутация увеличивает относительную пиковую мощность и снижает время реакции. Ген, в котором она расположена, является онкогеном, экспрессирующимся во многих видах опухолей [28]. Индивиды с мутантным аллелем А обладают меньшей безжировой массой, а частоты мутантных аллелей в обоих генах не отличались в выборке борцов и в контрольных популяциях, что также говорит о ложноположительном характере обнаруженной ассоциации.

Также обнаружено, что мутантный аллель Т в локусе rs1530507 гена *ADAMTS12* (замена *p.Leu230**) отрицательно коррелирует со спринтерскими способностями борцов. Ген кодирует белок эндопептидазу, которая, с одной стороны, обладает противоопухолевой активностью [29], а с другой — способностью расщеплять матриксный белок хряща, что приводит к артриту (данные GWAS). Частота минорного аллеля у борцов наиболее высокая — 0,75, значимые отличия по частоте выявлены с башкирами (0,47), однако распределение частот генотипов в выборке спортсменов не соответствует равновесию Харди-Вайнберга ($p < 0,05$).

Формирование стоп-кодона в локусе rs1043149 гена *ZACN* (замена *p.Gln281**), кодирующего белок ионного канала, повышает относительную мышечную массу спортсменов. Однако, согласно исследованиям GWAS, мутация повышает также жировую массу и общий вес, и среди фенотипов, ассоциированных с данным полиморфизмом, нет других признаков, прямо влияющих на успешность спортсмена. Значимых различий в частотах аллелей и генотипов между борцами и контрольными популяциями также не выявлено.

Таким образом, из всех нонсенс-мутаций, приводящих к прекращению синтеза белка, найденных в геномах борцов, только замена *p.Tyr319** (rs7447815) в гене *SLC6A18* показала однонаправленную ассоциацию с функциональным показателем на всех этапах исследования: мутация достоверно снижает значение относительной пиковой мощности нижних конечностей (разница в значениях мощности между борцами-носителями исходного генотипа СС и мутантного GG составляет более 2-х единиц), имеет меньшую частоту у борцов-татар по сравнению с башкирами и русскими, а также увеличивает риск развития остеоартроза.

Среди ограничений исследования можно назвать малый размер выборки и небольшую глубину прочтения геномов. Найденная ассоциация требует репликации на дополнительных независимых выборках спортсменов. Также планируется оценить эффект других видов мутаций, приводящих к потере функции белка, — в сплайс-сайтах и сдвига рамки считывания — на изученные функциональные показатели и на другие спортивные фенотипы.

Выводы

Данные полногеномного секвенирования борцов-татар в отношении количества обнаруженных однонуклеотидных замен и инделов, а также нонсенс-мута-

ций и их функциональной роли сопоставимы с геномными последовательностями, полученными в проектах по исследованию генетической вариабельности различных популяций.

Нонсенс-мутация *p.Tyr319** (rs7447815) в гене *SLC6A18* ассоциирована с меньшей пиковой мощностью нижних конечностей. Мутантный аллель G в этом локусе является фактором, негативно влияющим на силовые качества борцов, и носителям этого аллеля и, в особенности, генотипа GG следует обратить внимание на проработку соответствующих групп мышц при планировании тренировок.

Литература

- Maciejewska-Skrendo A., Sawczuk M., Cięszczyk P. et al. Genes and power athlete status. In *Sports, Exercise, and Nutritional Genomics: Current Status and Future Directions*. Academic Press, 2019:41-72.
- Semenova E.A., Fuku N., Ahmetov I.I. Genetic profile of elite endurance athletes. In *Sports, Exercise, and Nutritional Genomics: Current Status and Future Directions*. Academic Press, 2019:73-104.
- Valeeva E.V., Ahmetov I.I., Rees T. Psychogenetics and sport. In *Sports, Exercise, and Nutritional Genomics: Current Status and Future Directions*. Academic Press, 2019:147-165.
- Бондарева Э.А., Шиян В.В., Спицын В.А., Година Е.З. Ассоциации четырех полиморфных генетических систем (ACE, EPAS1, ACTN3 и NOS3) со спортивной успешностью в борьбе самбо. Вестник Московского университета. Серия 23. Антропология. 2010(1),36-45.
- Li H., Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589-595.
- Kim S., Scheffler K., Halpern A.L., et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods*. 2018;15(8):591-594.
- Loh P.R., Danecek P., Palamara P.F., et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genet*. 2016;48(11):1443.
- Das S., Forer L., Schönerr S., et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284-1287.
- Wang K., Li M., Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
- O'Leary N., Wright M., Brister J., et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016;44(D1):D733-45.
- Landrum M.J., Lee J.M., Benson M., et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46(D1):D1062-D1067.
- Sherry S.T., Ward M.H., Kholodov M., et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29(1):308-11.
- Karczewski K.J., Francioli L.C., Tiao G., et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443.
- Thomas P.D., Campbell M.J., Kejariwal A., et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res*. 2003;13:2129-2141.
- Boulygina E.A., Lukianova E., Grigoryeva T., et al. Lessons from the whole exome sequencing effort in populations of Russia and Tajikistan. *BioNanoScience*. 2016;6(4):540-542.
- Triska P., Chekanov N., Stepanov V., et al. Between Lake Baikal and the Baltic Sea: genetic history of the gateway to Europe. *BMC Genet*. 2017;18(1):5-20.
- Carvalho-Silva D., Pierleoni A., Pignatelli M., et al. Open Targets Platform: new developments and updates two years on. *Nucleic Acids Res*. 2019;47(D1):D1056-D1065.
- Zhernakova D.V., Brukhin V., Malov S., et al. Genome-wide sequence analyses of ethnic populations across Russia. *Genomics*. 2020;112(1):442-458.
- Seidually M., Blazyte A., Jeon S., et al. Decoding a highly mixed Kazakh genome. *Hum Genet*. 2020;139(5):557-568.
- Francioli L.C., Menelaou A., Pulit S.L., et al. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet*. 2014;46(8):818.
- Boulygina E.A., Borisov O.V., Valeeva E.V., et al. Whole genome sequencing of elite athletes. *Biol Sport*. 2020;37(1):295-304.
- MacArthur D.G., Balasubramanian S., Frankish A., et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science*. 2012;335(6070):823-828.
- Fujikura K. Premature termination codons in modern human genomes. *Sci Rep*. 2016;6:22468.
- Yang Y., Zhao H., Boomsma D.I., et al. Molecular genetic overlap between migraine and major depressive disorder. *Eur J Hum Genet*. 2018;26(8):1202-1216.
- Matsumoto K., Shimodaira M., Nakagawa T., et al. Association study: SLC6A18 gene and myocardial infarction. *Clin Biochem*. 2011;44(10-11):789-794.
- Tiala I., Suomela S., Huuhtanen J., et al. The CCHCR1 (HCR) gene is relevant for skin steroidogenesis and downregulated in cultured psoriatic keratinocytes. *J Mol Med*. 2007;85(6):589-601.
- Brenner L.N., Mercader J.M., Robertson C.C., et al. Analysis of glucocorticoid-related genes reveal CCHCR1 as a new candidate gene for type 2 diabetes. *J Endocr Soc*. 2020.
- Bonin F., Taouis K., Azorin P., et al. VOPPI promotes breast tumorigenesis by interacting with the tumor suppressor WWOX. *BMC Biol*. 2018;16(1):1-16.
- Cal S., Argüelles J.M., Fernández P.L., et al. Identification, characterization, and intracellular processing of ADAM-TS12, a novel human disintegrin with a complex structural organization involving multiple thrombospondin-1 repeats. *J Biol Chem*. 2001;276(21):17932-17940.

References

- Maciejewska-Skrendo A., Sawczuk M., Cięszczyk P. et al. Genes and power athlete status. In *Sports, Exercise, and Nutritional Genomics: Current Status and Future Directions*. Academic Press, 2019:41-72.
- Semenova E.A., Fuku N., Ahmetov I.I. Genetic profile of elite endurance athletes. In *Sports, Exercise, and Nutritional Genomics: Current Status and Future Directions*. Academic Press, 2019:73-104.
- Valeeva E.V., Ahmetov I.I., Rees T. Psychogenetics and sport. In *Sports, Exercise, and Nutritional Genomics: Current Status and Future Directions*. Academic Press, 2019:147-165.
- Bondareva E., Shiyon V., Spitsyn V., Godina E. Assotsiatsii chetyrekh polimorfnykh genicheskikh sistem (ASE, EPAS1, ACTN3 i NOS3) so sportivnoy uspešnost'yu v bor'be sambo [Associations of four polymorphisms (ACE, EPAS1, ACTN3 and NOS3) with high achievements in sambo wrestling]. *Vestnik Moskovskogo Universiteta. Seria XXIII. Antropologia [Moscow University Anthropology Bulletin]*, 2010; 1: 36-45 (In Russ.)
- Li H., Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589-595.
- Kim S., Scheffler K., Halpern A.L., et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods*. 2018;15(8):591-594.
- Loh P.R., Danecek P., Palamara P.F., et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genet*. 2016;48(11):1443.

8. Das S., Forer L., Schönherr S., et al. Next-generation genotype imputation service and methods. *Nat Genet.* 2016;48(10):1284-1287.
9. Wang K., Li M., Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.
10. O'Leary N., Wright M., Brister J., et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 2016;44(D1):D733-45.
11. Landrum M.J., Lee J.M., Benson M., et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 2018;46(D1):D1062-D1067.
12. Sherry S.T., Ward M.H., Kholodov M., et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001;29(1):308-11.
13. Karczewski K.J., Francioli L.C., Tiao G., et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581(7809):434-443.
14. Thomas P.D., Campbell M.J., Kejariwal A., et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 2003;13:2129-2141.
15. Boulygina E.A., Lukianova E., Grigoryeva T., et al. Lessons from the whole exome sequencing effort in populations of Russia and Tajikistan. *BioNanoScience.* 2016;6(4):540-542.
16. Triska P., Chekanov N., Stepanov V., et al. Between Lake Baikal and the Baltic Sea: genomic history of the gateway to Europe. *BMC Genet.* 2017;18(1):5-20.
17. Carvalho-Silva D., Pierleoni A., Pignatelli M., et al. Open Targets Platform: new developments and updates two years on. *Nucleic Acids Res.* 2019;47(D1):D1056-D1065.
18. Zhernakova D.V., Brukhin V., Malov S., et al. Genome-wide sequence analyses of ethnic populations across Russia. *Genomics.* 2020;112(1):442-458.
19. Seidually M., Blazyte A., Jeon S., et al. Decoding a highly mixed Kazakh genome. *Hum Genet.* 2020;139(5):557-568.
20. Francioli L.C., Menelaou A., Pulit S.L., et al. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet.* 2014;46(8):818.
21. Boulygina E.A., Borisov O.V., Valeeva E.V., et al. Whole genome sequencing of elite athletes. *Biol Sport.* 2020;37(1):295-304.
22. MacArthur D.G., Balasubramanian S., Frankish A., et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science.* 2012;335(6070):823-828.
23. Fujikura K. Premature termination codons in modern human genomes. *Sci Rep.* 2016;6:22468.
24. Yang Y., Zhao H., Boomsma D.I., et al. Molecular genetic overlap between migraine and major depressive disorder. *Eur J Hum Genet.* 2018;26(8):1202-1216.
25. Matsumoto K., Shimodaira M., Nakagawa T., et al. Association study: SLC6A18 gene and myocardial infarction. *Clin Biochem.* 2011;44(10-11):789-794.
26. Tiala I., Suomela S., Huuhtanen J., et al. The CCHCR1 (HCR) gene is relevant for skin steroidogenesis and downregulated in cultured psoriatic keratinocytes. *J Mol Med.* 2007;85(6):589-601.
27. Brenner L.N., Mercader J.M., Robertson C.C., et al. Analysis of glucocorticoid-related genes reveal CCHCR1 as a new candidate gene for type 2 diabetes. *J Endocr Soc.* 2020.
28. Bonin F., Taouis K., Azorin P., et al. VOPP1 promotes breast tumorigenesis by interacting with the tumor suppressor WWOX. *BMC Biol.* 2018;16(1):1-16.
29. Cal S., Argüelles J.M., Fernández P.L., et al. Identification, characterization, and intracellular processing of ADAM-TS12, a novel human disintegrin with a complex structural organization involving multiple thrombospondin-1 repeats. *J Biol Chem.* 2001;276(21):17932-17940.