

На правах рукописи

Грищенко Ирина Владимировна

**Исследование нестабильности экзогенных повторов (CGG)_n в клеточных
моделях**

1.5.3 – Молекулярная биология

Автореферат
диссертации на соискание ученой степени
кандидата биологических наук

Кольцово – 2022

Работа выполнена в Федеральном бюджетном учреждении науки «Государственный научный центр вирусологии и биотехнологии «Вектор» Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека.

Научный руководитель:	Юдкин Дмитрий Владимирович кандидат биологических наук, заведующий отделом геномных исследований ФБУН ГНЦ ВБ «Вектор» Роспотребнадзора
Официальные оппоненты:	Максимов Владимир Николаевич доктор медицинских наук, профессор, главный научный сотрудник с в. о. заведующего лабораторией молекулярно-генетических исследований терапевтических заболеваний Научно-исследовательского института терапии и профилактической медицины - филиала ФГБНУ Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук Егорова Татьяна Владимировна кандидат биологических наук, старший научный сотрудник лаборатории механизмов контроля трансляции ФГБНУ Институт молекулярной биологии им. В.А. Энгельгардта Российской академии наук
Ведущая организация:	ФГБНУ Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук

Защита состоится «27» мая 2022 г. в 11-00 на заседании диссертационного совета 64.1.001.01 при Федеральном бюджетном учреждении науки «Государственный научный центр вирусологии и биотехнологии «Вектор» Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека по адресу: 630559, Новосибирская обл., р.п. Кольцово, тел. 8 (383) 363-47-10.

С диссертацией можно ознакомиться в библиотеке и на сайте ФБУН ГНЦ ВБ «Вектор» Роспотребнадзора <http://www.vector.nsc.ru>.

Автореферат разослан «___» _____ 2022 г.

Ученый секретарь
диссертационного совета,
кандидат биологических наук

Т.С. Непомнящих

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования

Одной из особенностей геномов эукариот является обогащенность повторенными последовательностями ДНК. У человека не менее 2/3 генома составляют повторенные последовательности (de Koning A.P.J. et al., 2011). Эта часть содержит более миллиона tandemных микросателлитов – последовательностей ДНК, состоящих из коротких повторенных мономеров размером до 9 п. н. (Gymrek M., 2017; Willems T. et al., 2014), однако функции и биологические особенности этих последовательностей до сих пор остаются недостаточно изученными. Такие последовательности по своей природе склонны к изменению количества мономеров, но зачастую изменение в размере повтора не влечет за собой серьезных последствий для жизненно важных функций клетки. Однако в некоторых случаях нестабильность повторов может приводить к серьезным патологиям. Одной из ключевых форм изменения длины tandemных микросателлитов считается их экспансия – мутация, при которой происходит значительное увеличение числа мономеров в повторенной последовательности. Экспансия является причиной ряда нейродегенеративных заболеваний человека, таких, как фраксопатии (заболевания, связанные с увеличением CGG-повтора в 5'UTR гена *FMRI*), болезнь Хантингтона, боковой амиотрофический склероз, прогрессирующая миоклоническая эпилепсия и др. Известно, что увеличение длины различных повторов в геноме человека обуславливает развитие 36 неврологических и нервно-мышечных неизлечимых заболеваний (Grishchenko I.V., Purvinsh Y.V., Yudkin D.V., 2020).

Экспансия CGG-повтора в 5'UTR гена *FMRI* приводит к нарушению экспрессии этого гена и развитию фраксопатий, а именно синдрома ломкой X-хромосомы и ассоциированных с ломкой X-хромосомой заболеваний. Фраксопатии являются важной социальной проблемой в силу высокой распространенности в популяции, отсутствия таргетной терапии этих патологий, а также значительных социально-экономических затрат. Ежегодно на одного пациента с подтвержденной фраксопатией тратится до 10000 \$ в США (Sacco P. et al., 2013) и до 58862 € в странах Европы (Chevreul K. et al., 2016). При этом детальный механизм развития мутации до сих пор не установлен, что усложняет разработку терапевтических подходов, а также прогнозирование течения заболевания у пациентов. В связи с этим обнаружение причин экспансии является наиболее актуальным вопросом в данной области исследований.

В качестве возможного источника возникновения экспансии считаются альтернативные вторичные структуры ДНК (шпильки, G-квадруплексы и др.), способные нарушать процессы метаболизма ДНК: репликацию, репарацию и транскрипцию (Kononenko A.V., Ebersole T., Mirkin S.M., 2020; Krasilnikova M.M. et al., 2007; Polyzos A.A., McMurray C.T., 2017; Richard G.F. et al., 2000; Schmidt M.H.M., Pearson C.E., 2016; Voineagu I., Freudenreich C.H., Mirkin S.M., 2009). Обнаружение экспансии в тканях с низкой пролиферативной активностью, а также активная транскрипция генов, в которых обнаруживается нестабильность, подтверждают версию о зависимости экспансии от транскрипции и ассоциированной с ней репарации. (Lokanga R.A. et al., 2013), при этом экспериментальных доказательств о вовлеченности именно этих путей в формирование нестабильности недостаточно. Для детализации вклада путей метаболизма ДНК в нестабильность CGG-повторов необходимо иметь такую молекулярно-генетическую модель, в которой можно активировать эти каскады, чтобы обнаружить вклад каждого из них в наблюдаемую нестабильность. Наличие подобных модельных систем упростит понимание причинно-следственных связей в развитии нестабильности CGG-повтора и поможет смоделировать ее появление на различных этапах жизни клетки и онтогенеза, что крайне важно для понимания времени начала патогенеза синдрома ломкой X-хромосомы и ассоциированных заболеваний.

Целью работы является поиск особенностей вторичной структуры и размера CGG-повторов в геноме, а также анализ нестабильности этого типа повтора в модельных клеточных линиях при активации транскрипции. В связи с этим были поставлены следующие **задачи**:

1. Визуализировать формирование G-квадруплексов в районе локализации гена *FMR1* в клеточных линиях с различным размером CGG-повтора в промоторной области этого гена;
2. Измерить уровень обогащенности белками CSB и ERCC2, участвующими в репарации, ассоциированной с транскрипцией, в области повтора в гене *FMR1* в клеточных линиях с различными аллелями этого гена;
3. Исследовать нестабильность экзогенного CGG-повтора различной длины в эукариотической системе как без интеграции, так и с интеграцией в геном в зависимости от наличия транскрипции;
4. Определить уровень нестабильности CGG-повтора в различных геномных локусах человека в постоянных клеточных линиях, различных по аллелю гена *FMR1*.

Научная новизна

Поскольку проблема возникновения экспансии повторов является актуальной и перспективной темой исследований многие научные коллективы по всему миру работают над ее решением. Однако еще не получено достаточного количества экспериментальных данных, на основе которых была бы построена единая модель, объясняющая механизм экспансии. Настоящее исследование посвящено определению особенностей нестабильности CGG-повторов.

В этой работе впервые с помощью антител визуализировано формирование альтернативных вторичных структур – G-квадруплексов – в районе промотора гена *FMR1*.

Также нами впервые была создана модель на основе клеток эукариот, позволяющая оценить вклад транскрипции в изменение размера CGG-повтора. С ее помощью были получены уникальные данные, проливающие свет на природу нестабильности CGG-повтора. Подтвержденное участие вклада транскрипции объясняет происхождение соматического мозаицизма, который наблюдается у пациентов с синдромом ломкой X-хромосомы, и ассоциированными с ним заболеваниями, в клетках, не способных к делению.

Также впервые была проанализирована нестабильность повтора (CGG)_n в других геномных локусах в клетках с различным генотипом по гену *FMR1* и определен уровень его нестабильности в этих районах.

Теоретическая и научно-практическая значимость работы

Результаты исследования являются ключевыми для понимания природы экспансии повторов. Нами получены фундаментальные знания о вкладе репликации и транскрипции в нестабильность CGG-повтора с помощью созданной генно-инженерной модельной системы. Подтвержденное участие вклада транскрипции объясняет природу соматического мозаицизма у пациентов с фраксопатиями, которая не имела однозначных гипотез происхождения. Установленные особенности нестабильности CGG-повтора лягут в основу однозначной модели экспансии. Полученные знания позволят разрабатывать таргетную терапию, которая в перспективе может предотвратить процесс экспансии, а, следовательно, и вызванные ей патологии. Кроме того, созданные модельные клеточные линии могут быть использованы в других исследованиях, посвященных фраксопатиям.

Значимость проведенных исследований подтверждается наличием публикаций, посвященных нестабильности повторов, в журналах, индексируемых основными базами данных (WoS, Scopus, РИНЦ).

Положения, выносимые на защиту

1. Склонность к формированию G-квадруплексов в терминальной области q-плеча X-хромосомы в районе гена *FMR1* выше в постоянных культурах В-лимфоцитов человека с премутацией и полной мутацией по этому гену.
2. Обогащенность белками CSB и ERCC2 в локусе гена *FMR1* в постоянных культурах В-лимфоцитов человека прямо коррелирует с уровнем транскрипции этого гена, а также обогащенность белком CSB вероятно связана с сокращением количества копий CGG-повтора премутантной длины.
3. Транскрипция последовательности, содержащей экзогенный CGG-повтор премутантной длины, приводит к сокращению количества мономеров этого повтора. Экзогенный CGG-повтор нормальной длины стабилен при транскрипции последовательности, содержащей этот повтор.
4. Изменение количества мономеров повтора (CGG)_n в 5'UTR гена *FMR1* не связано с размером этого типа повтора в других геномных локусах.

Степень достоверности и апробации результатов

Достоверность результатов работы определяется использованием методов, адекватных целям и задачам исследования, а также методов статистической обработки полученных результатов. Материалы диссертационной работы были представлены на 8 международных научных конференциях. По материалам диссертации опубликовано четыре статьи в журналах, рекомендованных ВАК.

Структура и объем диссертации

Работа состоит из введения, обзора литературы, описания материалов и методов, результатов и обсуждения, заключения, выводов, списка цитируемой литературы, содержащего 243 ссылки. Диссертация изложена на 169 страницах машинописного текста, содержит 9 таблиц, 19 рисунков и 4 приложения.

Публикации

По результатам работы опубликовано 4 статьи в российских и зарубежных журналах из перечня ВАК, а также 8 тезисов в материалах отечественных и зарубежных конференций.

Вклад автора

Основные экспериментальные результаты получены автором самостоятельно, либо при его непосредственном участии. Подготовка публикаций проводилась автором совместно с научным руководителем зав. отделом, канд. биол. наук Д. В. Юдкиным.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

ОБЗОР ЛИТЕРАТУРЫ

Обзор литературы включает 3 главы и заключение. В первом разделе рассматривается феномен нестабильности повторов в геноме человека. В этой

главе приведена информация обо всех заболеваниях, вызванных экспансией повторов, известных на сегодняшний день, а также механизм их патогенеза.

Следующая глава посвящена механизмам появления и развития нестабильности повторенных последовательностей и моделям, созданным для изучения этих процессов. Здесь описаны различные альтернативные вторичные структуры ДНК, являющиеся возможными источниками нестабильности, а также пути метаболизма ДНК (репликация, транскрипция и ассоциированная с ней репарация), нарушаемые данными вторичными структурами.

Последняя глава посвящена особенностям развития фраксопатий и причинам появления нестабильности CGG-повтора.

МАТЕРИАЛЫ И МЕТОДЫ

Эксперименты были осуществлены при помощи стандартных методик молекулярной биологии. В работе использованы следующие методы: культивирование различных культур эукариотических клеток (HEK293A(T), постоянные культуры В-лимфоцитов человека) поддержание культур клеток прокариот и их трансформация (различные штаммы клеток *E. coli*), выделение нуклеиновых кислот, трансфекция клеток эукариот иммуноокрашивание, флуоресцентная *in situ* гибридизация, иммунопреципитация хроматина и др. Для создания клеточных моделей были собраны плазмидные конструкции методами генетической инженерии.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Формирование специфических вторичных структур ДНК – G-квадруплексов в 5'UTR гена FMR1

Для протяженного CGG-повтора характерно формирование G-квадруплексов. Они могут являться вероятным источником развития нестабильности, нарушая различные процессы метаболизма ДНК в клетках. Для поиска локализации этих вторичных структур в локусе FRAXA использованы препараты метафазных хромосом и ядер фиксированных клеток трех постоянных линий В-лимфоцитов человека, обладающих нормальным размером CGG-повтора (GM06895), премутантным аллелем гена *FMR1* (GM06891) и полной мутацией в гене *FMR1* (GM04025). Мы ожидали, что эти клеточные линии будут демонстрировать разный уровень представленности G-квадруплексов в ломком сайте FRAXA в силу разной длины тракта CGG-повторов и различной эффективности образования им вторичных структур. Для обнаружения G-квадруплексов в клетке проводили иммуноокрашивание с помощью коммерчески доступных антител BG4 (антитело закодировано в плазмиде pSANG10-3F-BG4, №55756, Addgene). Для обнаружения локуса FRAXA на препараты фиксированных

клеток гибридизовали зонд, представляющий собой ВАС-клон RP11-489K19 (Empire Genomics) с полной последовательностью гена *FMRI*, методом FISH. Для каждой линии клеток было проанализировано 100 метафазных пластинок. На Рисунке 1 показана колокализация сигналов после иммуноокрашивания и FISH.

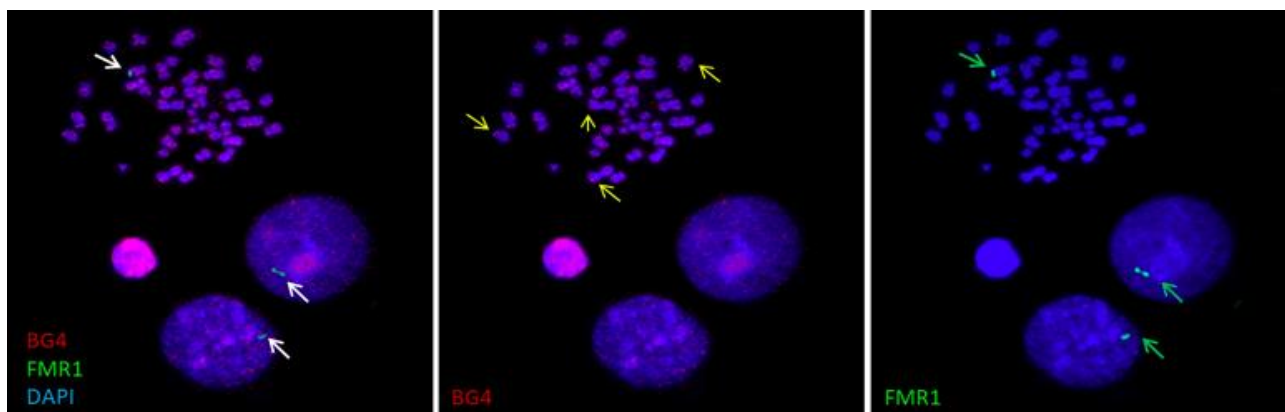


Рисунок 1. Пример локализации антитела BG4 (красный) и ВАС RP11-489K19, несущего последовательность *FMRI* (зеленый) на метафазных хромосомах и ядрах клеток (окрашены DAPI) линии GM04025. Колокализации сигналов иммуноокрашивания и FISH отмечены белой стрелкой. Сигналы зонда для FISH на основе ВАС RP11-489K19 обозначены зелеными стрелками. G-квадруплексы в районе теломер отмечены желтыми стрелками.

В результате последовательного иммуноокрашивания и FISH гибридизации мы обнаружили колокализацию сигналов антитела BG4 и зонда с последовательностью гена *FMRI* на X-хромосоме в постоянных линиях клеток В-лимфоцитов GM06891 и GM04025 в 86% и 78% случаев, соответственно. В клеточной линии GM06895 обогащенности G-квадруплексами локуса FRAXA обнаружено не было (колокализация встречалась в 12% случаев). Внутренним контролем специфичности полученных антител к G-квадруплексам служат сигналы локализации в районе теломер, поскольку для олигонуклеотидов, содержащих теломерный повтор, показано взаимодействие с BG4, а также доказано формирование G-квадруплексов теломерными последовательностями в клетках (Biffi G. et al., 2013; Henderson A. et al., 2014). Таким образом, было подтверждено наличие G-квадруплексов в районе ломкого сайта FRAXA в постоянных клеточных культурах В-лимфоцитов человека, несущих премутацию и полную мутацию гена *FMRI* при низкой представленности этих структур в клеточной линии с нормальным генотипом. Таким образом, можно предполагать, что преодоление порогового размера между нормой и премутацией приводит к увеличению вероятности возникновения таких структур в ДНК. Также наши наблюдения косвенно свидетельствует о связи G-

квадруплексов с нестабильностью CGG-повтора в гене *FMRI*, поскольку нестабильность повтора выражена в премутантном аллеле этого гена.

Локализация белков TC-NER в районе CGG-повтора гена *FMRI* в В-лимфоцитах человека

Один из возможных механизмов формирования нестабильности CGG-повтора заключается в возникновении нарушений, вызванных альтернативными вторичными структурами, при транскрипции и ассоциированной с ней репарации (TC-NER), однако этот путь развития нестабильности остается плохо изученным. На предыдущем этапе работы мы подтвердили формирование вторичных структур G-квадруплексов в районе гена *FMRI*, поэтому было решено исследовать влияет ли их присутствие на распределение белков каскада TC-NER.

В нашем исследовании мы выбрали два белка каскада TC-NER, которые потенциально могут участвовать в нестабильности CGG-повтора. Первым белком является геликаза CSB, реагирующая на остановку транскрипционного комплекса. Она определяет повреждение на молекуле ДНК при активации TC-NER во время транскрипции различных геномных локусов (van Gool A.J. et al., 1997). Кроме того на мышинных моделях с нокаутом этого гена показано, что отсутствие CSB обеспечивает уменьшение экспансии, но не предотвращает сокращение количества мономеров в тракте (CGG)_n в гене *FMRI* (Zhao X.N., Usdin K., 2015). Однако в этом исследовании не было установлено точного пути участия белка CSB в нестабильности CGG-повтора, а определена лишь частота появления экспансии или сокращения в ряду поколений у мышей при его отсутствии. Другим важным компонентом каскада репарации является геликаза ERCC2, представляющая собой кроме того важный транскрипционный фактор. Для этого белка показана способность взаимодействовать с G-квадруплексами в различных локусах генома, установленная методами ChIP-Seq (Gray L.T. et al., 2014). Однако локализация ERCC2 в локусе FRAXA не исследована.

Для локализации белков CSB и ERCC2 в локусе FRAXA мы использовали хроматин, выделенный из постоянных клеточных культур В-лимфоцитов, использованных на предыдущем этапе исследования: GM06895 (норма), GM06891 (премутация), GM04025 (полная мутация). После проведения иммунопреципитации, локализацию белков оценивали в трех локусах (CGG1, CGG2, CGG3), расположенных на участках в районе тракта CGG-повтора в гене *FMRI* с помощью ПЦР в реальном времени. Кроме сайтов предполагаемой локализации CSB и ERCC2 в гене *FMRI*, мы использовали промоторную

область гена домашнего хозяйства *GAPDH* в качестве положительного контроля иммунопреципитации.

При исследовании представленности белков TC-NER было показано, что клеточная линия GM06891, несущая CGG-повтор премутантной длины, содержит в 1,53 раза меньше белка CSB ($p = 0,075$, $\alpha = 0,1$) и в 1,9 раз меньше белка ERCC2 ($p = 0,061$, $\alpha = 0,1$) в локусах CGG1, CGG2 и CGG3, относительно клеточной линии GM06895 с CGG-повтором нормальной длины. Также было показано отсутствие обогащенности белками CSB и ERCC2 в районе CGG-повтора в линии GM04025, где отсутствует экспрессия мРНК гена *FMR1*, а, следовательно, и комплексы транскрипции и белки репарации, ассоциированной с транскрипцией (Рисунок 2).

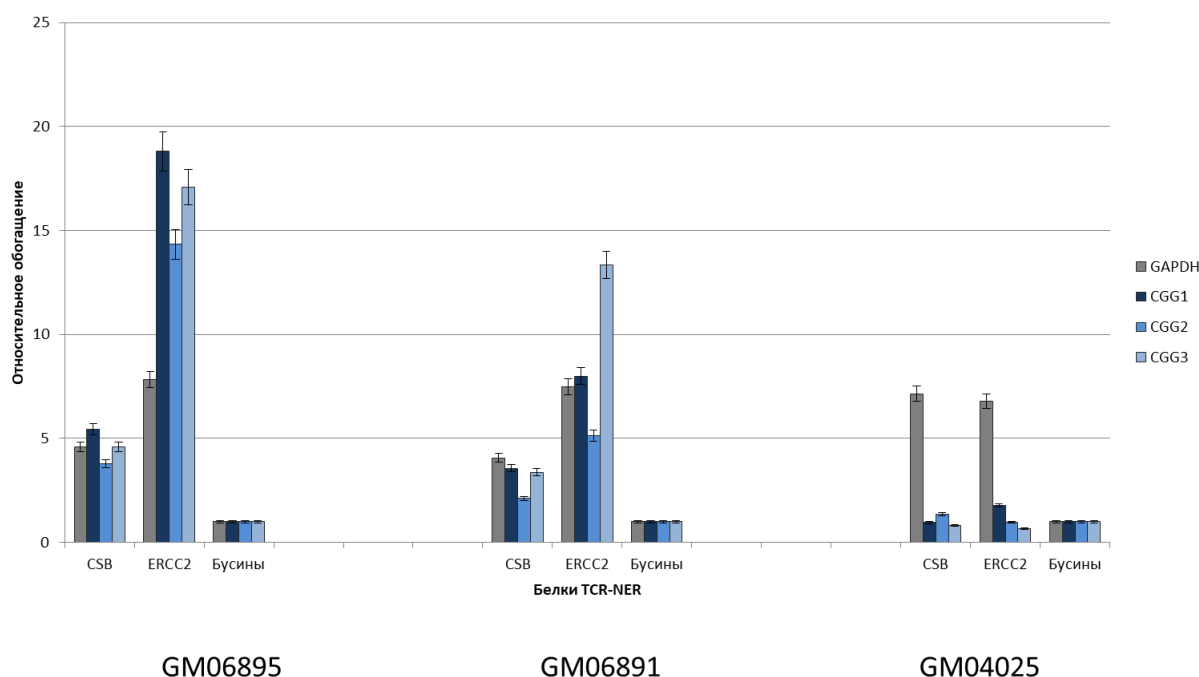


Рисунок 2. Локализация белков TC-NER в районе CGG-повтора в гене *FMR1*. Данные приведены для трех клеточных линий: GM06895, GM06891, GM04025. В этих культурах клеток была установлена локализация белков CSB и ERCC2 в 4 локусах. Локусы CGG1, CGG2 и CGG3 расположены в промоторной области гена *FMR1*, четвертый локус является внутренним положительным контролем и расположен в промоторной области гена домашнего хозяйства *GAPDH*. Отрицательным контролем служила реакция ПЦР в реальном времени с образцом ДНК после инкубации с агарозными бусинами без добавления специфичных антител.

Ранее нами было показано, что в клеточной линии GM06891 уровень транскрипции снижается относительно клеточной линии GM06895, которая обладает нормальным генотипом по гену *FMR1* (Dolskiy A.A. et al., 2021), что является нетипичным для клеток с премутацией. Однако эти наблюдения

согласуются с результатами о количестве белков TC-NER в районе CGG-повтора в линии GM06891. Кроме того для этой клеточной линии показано сокращение CGG-повтора в течение 120 дней культивирования, при этом мы обнаружили пониженное количество белка CSB относительно клеточной линии с нормальным генотипом в локусах CGG1, CGG2 и CGG3. Это наблюдение подтверждает данные, полученные на модельных мышах, о том, что снижение CSB уменьшает вероятность экспансии, но не влияет на сокращение CGG-повтора (Zhao X.N., Usdin K., 2015). Отсутствие как экспрессии гена *FMR1*, так и белков, ассоциированных с транскрипцией, в локусе FRAXA в клеточной линии GM04025 говорит об отсутствии вклада транскрипции в нестабильность полной мутации. Таким образом, показано, что транскрипционные факторы и белки TC-NER присутствуют в непосредственной близости от CGG-повтора в линии GM06891, также их представленность связана с уровнем транскрипции гена *FMR1* и с вероятностью развития нестабильности CGG-повтора. При этом мы не обнаружили однозначной связи между наличием G-квадруплексов и количеством белков, участвующих в транскрипции и TC-NER.

Создание модельных систем, содержащих экзогенный (CGG)*n* повтор

Дизайн и сборка генно-инженерных конструкций

В постоянных клеточных линиях, использованных в предыдущих экспериментах, процесс транскрипции через CGG-повтор не контролируем исследователем. Для установления однозначного вклада транскрипции в нестабильность транскрибируемых микросателлитов нужна модель, в которой была бы возможность активировать или останавливать этот процесс. Подобной моделью служит трансгенная клеточная культура с индуцируемой транскрипцией последовательности, содержащей экзогенный тракт CGG-повторов.

Мы разработали две плазмидные конструкции, которые в дальнейшем были использованы для липофекции эукариотических клеток для их экспрессии. Главными компонентами плазмид являются три последовательно расположенных функциональных элемента: эукариотический промотор, индуцируемый доксициклином, экзогенный фрагмент ДНК, содержащий тракт CGG-повторов, и ORF зеленого флуоресцентного белка (GFP). Благодаря такому взаимному расположению фрагментов в кассете можно установить, что транскрипция идет через участок со встроенным повтором за счет синтеза GFP, а также появляется возможность отслеживать изменения в размере повтора в ответ на активацию транскрипции.

Для сборки конструкций мы использовали фрагменты, содержащие тракт CGG-повторов нормальной и премутантной длины. Показано, что оба повтора транскрибируются в постоянных клеточных линиях человека, при этом G-квадруплексы формирует только премутантный CGG-повтор. Кроме того, показано, что премутантный аллель более склонен к изменению длины, чем другие типы аллелей (Lokanga R.A. et al., 2013). Эти характеристики могут привести к разному уровню нестабильности во время транскрипции в созданных модельных клеточных линиях, что позволит обнаружить особенности природы нестабильности. Повторы нормальной и премутантной длины получены методом ПЦР с использованием ДНК, выделенной из постоянных клеточных линий В-лимфоцитов человека с нормальным размером повтора и премутацией по гену *FMR1*. Схемы разработанных плазмидных конструкций представлены на рисунке 3.

А



Б

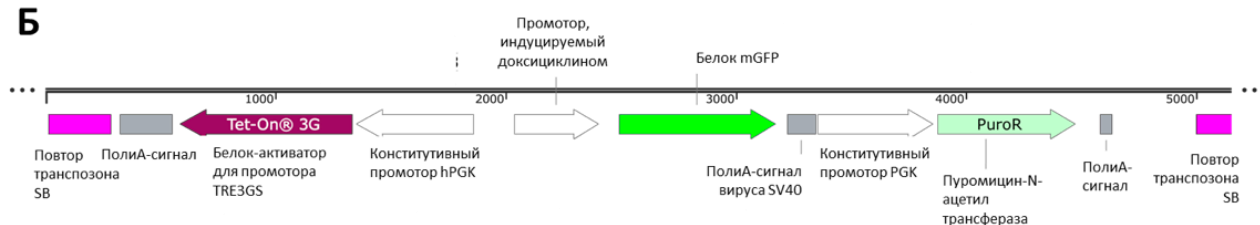


Рисунок 3. Схема плазмид pCDH (А) и pSBi (Б). EGFP – зеленый флуоресцентный белок с усиленной флуоресценцией; EF1 – конститутивный промотор гена *EF1-α*; rtTA – трансаактиватор, взаимодействующий с доксициклином; DsRedExpress – красный флуоресцентный белок; Tet-On 3G – последовательность, кодирующая трансаактиватор, который взаимодействует с доксициклином и затем активирует индуцируемый промотор.

Первая плазмидная конструкция основана на векторе pCDH (Рисунок 3, А). Особенностью данного вектора является наличие ориджина репликации вируса SV40. Благодаря ему плаزمида может поддерживаться в клетках HEK293T, экспрессирующих большой Т-антиген вируса SV40. В клеточных линиях, не синтезирующих этот белок, плазмида не реплицируется и элиминируется в течение 10 суток при интенсивном клеточном делении (Hacker D.L., Wurm F.M., 2011; Piechaczek C. et al., 1999). В клеточной линии HEK293T можно отслеживать влияние как транскрипции и TC-NER, так и репликации на

экзогенный CGG-повтор. На основе клеток HEK293T были получены трансгенные клеточные линии: HEK293T-pCDH-5, HEK293T-pCDH-89 (числа в конце названия обозначают размер клонированного в вектор CGG-повтора). В модельной клеточной линии HEK293A, содержащей данную конструкции, можно оценить вклад транскрипции и TC-NER в нестабильность CGG-повтора, полностью исключив влияние репликации. Эта клеточная линия не содержит большой Т-антиген вируса SV-40, поэтому плазмида pCDH в ней не реплицируется, а лишь транскрибируется. На основе клеточной культуры HEK293A нами были получены трансгенные линии клеток: HEK293A-pCDH-5, HEK293A-pCDH-59 (числа в конце названия обозначают размер экзогенного CGG-повтора).

Однако система на основе плазмиды pCDH имеет свои недостатки. При интенсивном клеточном делении плазмида, которая не реплицируется, будет элиминироваться из клеток. Для отслеживания вклада транскрипции в нестабильность повтора после 10 суток культивирования трансгенной клеточной линии, необходимо использовать модельные клетки, в которых плазмида с экзогенным CGG-повтором под индуцируемым промотором постоянно поддерживается в культуре, как в клеточной линии HEK293T, трансфицированной плазмидой pCDH. Однако при постоянной экспрессии большого Т-антигена происходит интенсивная репликация плазмиды, содержащей ORI вируса SV40 (Van Craenenbroeck K., Vanhoenacker P., Naegeman G., 2000), что приводит к гибели клеток и не позволяет использовать трансгенную клеточную культуру в течение продолжительного времени. Поэтому был разработан вектор на основе плазмиды pSBi (Рисунок 3, Б), которая является частью системы транспозон/транспозаза Sleeping Beauty. Трансген интегрируется в геном и не ведет к гибели клеток. С использованием плазмиды pSBi были получены стабильные во времени клеточные линии HEK293A-pSBi-25 и HEK293A-pSBi-165 (числа обозначают размер клонированного в вектор CGG-повтора).

Оптимизация культивирования прокариотических клеток, несущих плазмиду с CGG-повтором

Для наработки плазмид, содержащих CGG-повтор, не предложено универсального протокола, поэтому мы провели оптимизацию протокола трансформации и культивирования бактериальных клеток.

В качестве параметров для оптимизации были выбраны следующие характеристики: используемый бактериальный штамм, тип трансформации, температура инкубации и продолжительность культивирования. Для наработки

плазмид с экзогенным CGG-повтором мы использовали два коммерчески доступных штамма *E.coli*: DH5alpha (Thermo FS) и NEB Stable (New England Biolabs). Уровень нестабильности CGG повтора в этих штаммах не известен, однако характеристики их теоретически подходят для наработки плазмид с повторенными последовательностями. В генотипе клеток присутствуют две значимые мутации, которые повышают эффективность поддержания плазмид: мутация в белке *recA* (замена глицина на аспарагиновую кислоту), что приводит к уменьшению активности рекомбиназы и нарушению гомологичной рекомбинации. Другая мутация нарушает работу внутриклеточной эндонуклеазы I (*endA*), что способствует сохранению экзогенной ДНК внутри бактериальной клетки. *E. coli* штамм NEB Stable также несут мутантные гены *recA* и *endA*.

Для выбранных штаммов были проверены два протокола трансформации: электропорация и трансформация, опосредованная тепловым шоком.

После трансформации бактерии выращивали на агаризованной среде LB в чашках Петри при различных температурах: комнатной и оптимальной для используемого бактериального штамма (37°C для DH5alpha и 30°C для NEB Stable). Комнатная температура была выбрана, поскольку в таких условиях бактериальные клетки делятся менее интенсивно, а значит можно ожидать, что процесс репликации будет происходить точнее и приведет к меньшей нестабильности CGG-повтора (Kumar P., Libchaber A., 2013).

Инкубация в жидкой среде при выбранном температурном режиме проведена в пяти повторях (один повтор соответствует единичной колонии бактерии). Такое количество колоний было использовано для обнаружения более точных корреляций между стабильностью CGG-повтора и способом культивирования бактериальных клеток. Через сутки после начала инкубации, из части культуры была выделена плазмидная ДНК, оставшуюся часть клеток инкубировали в течение двух суток и на третьи сутки также выделяли плазмидную ДНК. При сравнении данных о размере CGG-повтора в плазмидах на первые и третьи сутки можно установить влияние репликации бактериальной культуры на стабильность повтора.

Для исследования нестабильности CGG повтора в бактериальных культурах были выбраны плазмиды pSBi-25 и pSBi-165. Нами были получены следующие данные о стабильности аллелей N и PM в различных условиях культивирования в бактериях (Таблица 1).

Полученные результаты подтверждают гипотезу о том, что CGG-повтор который соответствует нормальному аллелю не склонен к

сокращению/экспансии, поскольку он был стабилен во всех исследованных условиях. В предыдущих экспериментах мы показали также отсутствие альтернативных вторичных структур в клеточной линии с нормальным размером повтора. Все это говорит о высокой стабильности повтора нормальной длины. Для плазмид, несущих CGG-повтор премутантной длины мы показали, что оптимальным методом получения таких конструкций является трансформация методом теплового шока клеток штамма NEB Stable и дальнейшее культивирование при 30°C не более 24 часов.

Таблица 1. Результаты культивирования бактерий, содержащих плазмиду с экзогенным повтором (CGG)_n в разных условиях.

Размер повтора в плазмиде	Временн ая точка	DH5alpha				NEB Stable			
		Электропорация		Тепловой шок		Электропорация		Тепловой шок	
		Комнатная температура	37°C	Комнатная температура	37°C	Комнатная температура	30°C	Комнатная температура	30°C
25 CGG-повторов	1 сутки	25 повторов во всех колониях	25 повторов во всех колониях	25 повторов во всех колониях	25 повторов во всех колониях	25 повторов во всех колониях	25 повторов во всех колониях	25 повторов во всех колониях	25 повторов во всех колониях
	3 сутки	25 повторов во всех колониях	25 повторов во всех колониях	25 повторов во всех колониях	25 повторов во всех колониях	25 повторов во всех колониях	25 повторов во всех колониях	25 повторов во всех колониях	25 повторов во всех колониях
165 CGG-повторов	1 сутки	165	165	165/75	165/142	165	165	165	165
		165	165	25	165/100/30	165	165	165	165
		165	165	165/110/25	25	150/90	165	165	165
		90	165	165	25	165	165	165	165
		165	125	165/98	165	165	90	30	165
	3 сутки	165	165	165/75	165/142	165	165/25	165	30
		165	165	25	30	165	165/108	165	30
		165	165	165/110/25	25	90	165	165	30
		90	165/58	165	25	165	165	165	30
		165	125	165/98	165	165	90	30	30

Зеленым цветом обозначены ячейки, в которых указаны плазмиды, выделенные из бактериальных культур, в которых размер повтора не изменился. Желтым цветом отмечены ячейки, в которых произошло изменение в размере повтора в плазмидах при трансформации и/или культивировании. Кроме цветовой маркировки был использован красный жирный шрифт, которым выделены размеры повтора в плазмидах, отличающиеся от исходного варианта. Черный жирный шрифт отражает дальнейшую стабильность изменившихся CGG-повторов. Для удобства каждой бактериальной культуре присвоен номер от 1 до 5.

Определение уровня нестабильности экзогенного повтора (CGG)_n в модельных клеточных линиях эукариот

Модели CGG-повтора без интеграции генно-инженерной кассеты в геном и определение в них уровня нестабильности экзогенного повтора

В клеточных линиях HEK293A и HEK293T, несущих неинтегрированную в геном плазмиду pCDH с экзогенным трактом CGG-повторов, мы изучили влияние транскрипции и репликации на стабильность повторенной последовательности. Созданные клеточные модели позволяют полностью разделить влияние этих двух процессов на изменение CGG-повтора.

Нами были получены следующие линии на основе плазмиды pCDH, отличающиеся длиной экзогенного тракта CGG-повторов: HEK293A-pCDH-5, HEK293T-pCDH-5, HEK293A-pCDH-59, HEK293T-pCDH-89 (числа в конце названия обозначают размер экзогенного CGG-повтора).

В таблице 2 представлены данные о нестабильности экзогенного CGG-повтора во всех клеточных популяциях. В результате при культивировании в течение 10 суток показано, что склонность экзогенного CGG-повтора к изменению длины незначительна, но существует тенденция к изменению в присутствии транскрипции. Важно иметь в виду, что изменения размера микросателлитных повторов на несколько триплетов происходят в рамках природного полиморфизма, поэтому необходимы данные о нестабильности CGG-повтора в зависимости от транскрипции и репликации после 10 суток.

Таблица 2. Изменение размера повтора при непродолжительном культивировании трансгенных клеточных линий.

Линия клеток	Репликация плазмиды	Активация транскрипции доксициклином	Временные точки, сутки									
			1	2	3	4	5	6	7	8	9	10
НЕКА-pCDH-5	Нет	Да	5	5	5	5	5	5	5	5	5	5
		Нет	5	5	5	5	5	5	5	5	5	5
НЕКА-pCDH-59	Нет	Да	59	59	59	59	59	59	62	59	57	64
		Нет	59	59	59	59	59	59	59	62	59	64
НЕКТ-pCDH-5	Да	Да	5	5	5	5	5	5	5	5	5	5
		Нет	5	5	5	5	5	5	5	5	5	5
НЕКТ-pCDH-89	Да	Да	89	89	89	89	89	89	90	89	87	89
		Нет	89	89	89	89	89	89	86	89	89	89

Зеленым цветом обозначены временные точки, в которых размер повтора не изменился. Желтым цветом отмечены временные точки, в которых произошло изменение в размере повтора при культивировании. Кроме цветовой маркировки был использован красный жирный шрифт, которым выделены размеры повтора, отличающиеся от исходного варианта.

Модели с интеграцией экзогенного CGG-повтора в геном и определение уровня его нестабильности

Для интеграции в геном и оценки влияния транскрипции на нестабильность CGG-повтора при длительном культивировании клеток были созданы плазмиды на основе системы транспозон/транспозаза Sleeping Beauty. Нами были получены три постоянные клеточные линии на основе клеток HEK293A и собранных плазмид pSBi-25 и pSBi-165: HEK293A-pSBi-N (25 CGG-повторов) и HEK293A-pSBi-PM (165 CGG-повторов). В культуре HEK293A-pSBi-N мы ожидали увидеть подтверждение стабильности CGG-повтора нормальной длины при длительном культивировании. В клетках HEK293A-pSBi-PM мы рассчитывали обнаружить больший вклад транскрипции и TC-NER в нестабильность CGG-повтора, соответствующего премутации, чем при культивировании в течение 10 суток. Для каждой исходной линии клеток, несущей CGG-повтор, были получены по 2 линии методом отбора единичных колоний, с интеграцией транспозона в различные локусы генома, которые были использованы для большей достоверности оценки транскрипции и TC-NER в нестабильность CGG-повтора: HEK293A-pSBi-N-①, HEK293A-pSBi-N-②, HEK293A-pSBi-PM-①, HEK293A-pSBi-PM-②. Также была использована контрольная линия HEK293A-pSBi-без CGG.

Полученные клеточные линии культивировали в течение 6 месяцев. Каждый месяц проводили сбор образцов для оценки размера повтора, уровня экспрессии GFP на уровне РНК и белка.

В таблице 3 приведена оценка размера повтора в начале эксперимента после отбора единичных клонов (t0), в первый месяц (t1) и шестой месяц культивирования (t2) в клеточных линиях с экзогенным CGG-повтором.

Таблица 3. Изменение размера повтора при культивировании трансгенных клеточных линий в течение 6 месяцев.

Линия клеток	Размер повтора после отбора клонов (t0)	Активация транскрипции с помощью доксициклина	Временная точка	
			1 месяц культивирования (t1)	6 месяцев культивирования (t2)
HEK293 A-pSBi-N-①	25	Да	25	25
		Нет	25	25
HEK293 A-pSBi-N-②	25	Да	25	25
		Нет	25	25
HEK293 A-pSBi-PM-①	165	Да	170/77	178/77/32
		Нет	165	172/71
HEK293 A-pSBi-PM-②	165	Да	167/80/63/62/61	38/32
		Нет	165/65/63/60	172/62/61/60/38

Зеленым цветом обозначены временные точки, в которых размер повтора не изменился. Желтым цветом отмечены временные точки, в которых произошло изменение в размере повтора при культивировании. Кроме цветовой маркировки был использован красный жирный шрифт, которым выделены размеры повтора, отличающиеся от исходного варианта.

В результате анализа показано, что размер повтора в норме оказался постоянным во всех временных точках. Наоборот, повтор соответствующий премутации оказался нестабильным. Уже спустя месяц культивирования мы детектировали мозаицизм по размеру повтора в клетках линии HEK293A-pSBi-PM-① и HEK293A-pSBi-PM-② при активации транскрипции. Интересно, что склонность к экспансии/сокращению к первому месяцу культивирования (t1) в разных клонах отличалась. При этом мы также наблюдали появление дополнительных укороченных аллелей в линии HEK293A-pSBi-PM-② без индукции транскрипции после 1 месяца культивирования. В конце эксперимента (спустя 6 месяцев культивирования) в линии HEK293A-pSBi-PM-① дох произошла экспансия с 170 до 178 CGG-триплетов, однако этот вариант аллеля не был основным. Также в этой линии клеток произошло сокращение повтора до нормы и меньшей премутации. Полная утрата крупного аллеля в клетках HEK293A-pSBi-PM-② дох к 6 месяцу культивирования и сокращение повтора вплоть до нормы свидетельствует об очевидном вкладе транскрипции в нестабильность повтора. В отсутствии активной транскрипции происходили экспансия и сокращение, однако, с сохранением исходного типа аллеля.

Таким образом, при культивировании клеточных линий HEK293A-pSBI-PM была показана нестабильность CGG-повтора как в сторону экспансии, так и в сторону сокращения при активации транскрипции. При этом с течением времени транскрипция вносит больший вклад в сокращение CGG-повтора, поскольку мы наблюдаем утрату крупных аллелей к 6 месяцу культивирования в клетках, где CGG-повтор транскрибируется, и сокращение размера повтора вплоть до нормы. Таким образом, мы получили однозначные свидетельства вклада транскрипции и TC-NER в нестабильность CGG-повтора. Склонность повтора к сокращению при активной транскрипции подтверждает участие этого процесса в качестве основного механизма развития соматического мозаицизма, наблюдающегося у пациентов, который характеризуется уменьшением крупных аллелей вплоть до нормы.

Нестабильность (CGG)_n в различных геномных локусах в постоянных клеточных линиях человека

В геноме человека обнаружен 6101 локус, содержащий CGG-повтор. Одной из задач этого исследования было определить является ли нестабильность CGG-повтора в промоторной области гена *FMR1* уникальной особенностью самого локуса или же она вызвана нарушением работы какого-либо белка, участвующего в репликации, транскрипции и/или TC-NER, на тракте повторов (CGG)_n. Для этого мы проводили исследование нестабильности CGG-повтора в различных геномных локусах в нескольких отличающихся по аллелю гена *FMR1* клеточных линиях человека (GM06895, GM06865, CPG144, CPG10, GM06891, GM06897, CPG18, GM04025). Выбранные линии позволяют рассмотреть связь изменения CGG-повторов в геномных локусах со всеми аллельными вариантами локуса *FRAXA*. При возможном нарушении работы ферментативных каскадов во время прохождения через повторенную последовательность можно ожидать появление экспансии CGG-повтора во всех локусах, содержащих эту последовательность. В пользу этой гипотезы говорят проведенные нами ранее эксперименты с использованием модельных клеточных линий эукариот, в которых CGG-повтор был нестабилен, хотя располагался в окружении, отличном от нативного локуса. Если же изменение размера повтора будет наблюдаться только в локусе *FRAXA*, то это станет свидетельством появления экспансии в этом регионе, обусловленной его геномным или хроматиновым окружением.

В нашем исследовании мы изучили нестабильность CGG-повтора в 64 локусах, расположенных в промоторных областях генов человека. Подбор генов для анализа осуществлялся следующим образом. Координаты 5'UTR генов были взяты из аннотации генома человека GENECODE (версия 36),

последовательности 5'UTR были экстрагированы по координатам из геномной сборки (версия GRCh38.p13) с помощью инструмента bedtools. Для поиска последовательностей, содержащих CGG-повтор в 5'UTR был использован инструмент STR-FM (Fungtammasan A. et al., 2015), с настройками минимальной длины STR (15 п. н.). Определение размера повтора осуществлялось по методике амплификации GC-богатых матриц.

В результате мы получили данные о размере повтора во всех исследуемых локусах. В 42 из 64 локусов полиморфизм по размеру повтора не наблюдался. Размер ПЦР-продукта соответствовал ожидаемому размеру и был одинаков во всех клеточных линиях. В 22 из 64 локусов продукты амплификации отличались по размеру от предсказанных по длине. Был обнаружен полиморфизм, при котором изменения происходили на несколько триплетов (от 3 до 15) относительно ожидаемого размера. Однако обнаруженный полиморфизм не был связан с нестабильностью CGG-повтора в гене *FMR1*: локусы с увеличенным повтором встречались в клеточных линиях с нормальным вариантом гена *FMR1*, и наоборот – локусы, содержащие повтор нормального размера наблюдались в клеточных линиях с полной мутацией по гену *FMR1*.

Выявленные особенности позволяют предполагать, что увеличение размера повтора в одном участке генома не взаимосвязаны с его размером в другом. Полученные данные свидетельствуют о независимом от других генов развитии экспансии в локусе FRAXA.

ВЫВОДЫ

1. Показана повышенная склонность к формированию G-квадруплексов в районе локализации гена *FMR1* в терминальной области q-плеча X-хромосомы в клеточных линиях с увеличенным CGG-повтором с помощью иммуноокрашивания и флуоресцентной *in situ* гибридизации.
2. Обнаружена взаимосвязь представленности белков CSB и ERCC2, участвующих в ассоциированной с транскрипцией репарации, в районе CGG-повтора в гене *FMR1* с уровнем его транскрипционной активности, а также взаимосвязь снижения количества белка CSB и склонности CGG-повтора премутантной длины к сокращению количества копий мономеров.
3. Экзогенный CGG-повтор премутантной длины показывает тенденцию к нестабильности при активации транскрипции в трансгенных модельных клеточных линиях, культивируемых в течение 10 суток, при этом экзогенный CGG-повтор нормальной длины стабилен при репликации и активации транскрипции.

4. При интеграции в геном после 6 месяцев культивирования экзогенный CGG-повтор премутантной длины демонстрирует склонность к экспансии при репликации и сокращению в ответ на активацию транскрипции, в то же время экзогенный CGG-повтор нормальной длины не изменяется при аналогичных условиях.

5. Нестабильность CGG-повтора в гене *FMR1* в культуре клеток не связана с нестабильностью этого же типа повторов в других геномных локусах. В разных районах генома нестабильность CGG-повтора развивается независимо.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

1. **Грищенко, И. В.**; Тулупов, А. А.; Рымарева, Ю. М.; Петровский, Е. Д.; Савелов, А. А.; Коростышевская, А. М.; Максимова, Ю. В.; Шорина, А. Р.; Шитик, Е. М.; Юдкин, Д. В. Трансгенная клеточная линия с индуцируемой транскрипцией для исследования механизмов экспансии (CGG)_n повторов // Вавиловский журнал генетики и селекции. – 2021. – Т. 25. – № 1. – С. 117–124. doi: 10.18699/VJ21.014; (Scopus, WoS, РИНЦ)
2. **Grishchenko, I. V.**, Purvinsh, Y. V., Imatdinov, I. R., Kulemzin, S. V., Yudkin, D. V. Cell system with inducible transcription for the study of CGG repeat instability // Eur. J. Hum. Genet. – 2020. – V. 28. – № S1. – P. 359–359. doi: 10.1038/s41431-020-00739-z; (Scopus, WoS)
3. **Grishchenko I. V.**, Purvinsh Y. V., Yudkin D. V. Mystery of Expansion: DNA Metabolism and Unstable Repeats // Mechanisms of Genome Protection and Repair. Part of the Advances in Experimental Medicine and Biology book series / Ed by. D.O. Zharkov. Cham: Springer International Publishing, – 2020. P. 101–124. doi: 10.1007/978-3-030-41283-8_7; (Scopus, WoS, РИНЦ)
4. Yudkin, D. V., **Grishchenko, I. V.**, Tulupov, A. A., Rymareva, Y. M., Maksimova, Y. V., Shorina A.R. Somatic instability of CGG repeats in the FMR1 gene is a factor in symptom severity in Fragile X syndrome patients // Eur. J. Hum. Genet. – V. 27. – P. 1387–1388. doi: 10.1038/s41431-019-0494-2; (Scopus, WoS)

Разные этапы работы были поддержаны грантами РФФ (№15-15-10001, №18-15-00099) и РФФИ (№ 18-34-00336).