

СВЯЗЬ ПОЛИМОРФИЗМА ГЕНОВ *ABCB9* И *COL22A1* С ПРЕДРАСПОЛОЖЕННОСТЬЮ ЧЕЛОВЕКА К ТЯЖЕЛЫМ ФОРМАМ КЛЕЩЕВОГО ЭНЦЕФАЛИТА

© 2019 г. А. В. Бархаш¹, *, А. А. Юрченко¹, Н. С. Юдин^{1,2}, И. В. Козлова³, И. А. Борищук⁴,
М. В. Смольникова⁵, О. И. Зайцева⁵, Л. Л. Позднякова⁶, М. И. Воевода^{1,2}, А. Г. Ромащенко¹

¹Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения
Российской академии наук, Новосибирск, 630090 Россия

²Новосибирский национальный исследовательский государственный университет, Новосибирск, 630090 Россия

³Научный центр проблем здоровья семьи и репродукции человека, Иркутск, 664003 Россия

⁴Иркутская областная инфекционная клиническая больница, Иркутск, 664043 Россия

⁵Федеральный исследовательский центр «Красноярский научный центр Сибирского отделения Российской академии
наук», Научно-исследовательский институт медицинских проблем Севера, Красноярск, 660022 Россия

⁶Городская инфекционная клиническая больница № 1, Новосибирск, 630099 Россия

*e-mail: barkhash@bionet.nsc.ru

Поступила в редакцию 16.04.2018 г.

После доработки 07.07.2018 г.

Поступила в редакцию 07.08.2018 г.

Клещевой энцефалит (КЭ) вызывается нейротропным РНК-содержащим вирусом из рода *Flavivirus*. КЭ характеризуется значительной вариабельностью клинических проявлений — от непаралитических форм (лихорадочная, менингеальная) до тяжелых паралитических (очаговых) форм (менингоэнцефалитическая, полиомиелитическая, полиоэнцефаломиелитическая). Результат взаимодействия между вирусом и хозяином (и, следовательно, течение и исход вирусного заболевания) в значительной степени зависит от генетически детерминированной способности иммунной системы организма хозяина (в частности, человека) подавлять развитие вирусной инфекции. Тем не менее наследственная предрасположенность к КЭ еще мало изучена в популяциях человека. В данной работе приведены результаты полноэкзомного секвенирования образцов ДНК 22 русских неиммунизированных больных с тяжелыми формами КЭ и 17 контрольных индивидов из той же популяции. Идентифицированы 16 однонуклеотидных полиморфизмов (single nucleotide polymorphism, SNP), ассоциированных с предрасположенностью к тяжелым формам КЭ. Частоты генотипов и аллелей трех из этих SNPs, локализованных в генах *ABCB9* (rs4148866, G/A, интрон), *COL22A1* (rs4909444, G/T, Ala938Asp) и *ITGAL* (rs1557672, G/A, интрон), были далее исследованы в выборках большего размера больных различными формами КЭ ($n = 177$) и в контрольной популяции ($n = 215$). В результате была впервые продемонстрирована ассоциация SNPs генов *ABCB9* и *COL22A1* с развитием тяжелых форм КЭ в популяции русских. Рассматривается гипотеза в отношении возможного механизма влияния интронного SNP гена *ABCB9* на процесс инфицирования вирусом КЭ у человека.

Ключевые слова: клещевой энцефалит, генетическая предрасположенность, полноэкзомное секвенирование, ген *ABCB9*, ген *COL22A1*.

DOI: 10.1134/S0016675819030032

Клещевой энцефалит (КЭ) — трансмиссивное природно-очаговое заболевание, вызываемое РНК-содержащим вирусом из рода *Flavivirus* семейства *Flaviviridae*. Вирус КЭ, включающий европейский, сибирский и дальневосточный подтипы, распространен на территории России и в ряде европейских стран. Заражение человека происходит преимущественно через укус инфицированного клеща (в основном *Ixodes ricinus* и *Ixodes persulcatus*). Ежегодно регистрируется около 6000–14000 случаев заболевания КЭ (в том

числе 3000–11000 случаев в России) [1–3]. Нейротропный вирус КЭ может преодолевать гематоэнцефалический барьер и вызывать серьезные повреждения центральной нервной системы (ЦНС), но примерно в 70–95% случаев заболевание протекает бессимптомно [2]. КЭ характеризуется значительной вариабельностью клинических проявлений — от непаралитических форм (лихорадочная, менингеальная) до тяжелых паралитических (очаговых) форм с поражением ЦНС (менин-

гоэнцефалитическая и др.), при этом тяжелые формы встречаются значительно реже [1–4].

Известно, что течение и исход инфекционных заболеваний в значительной степени зависят от особенностей защитных реакций организма хозяина. Индивидуальные различия в способности иммунной системы хозяина (в частности, человека) подавлять развитие инфекции предопределены генетически. Наследственная предрасположенность к некоторым инфекционным заболеваниям (СПИД, туберкулез, малярия и др.) интенсивно изучается, и для них уже идентифицированы некоторые генетические факторы человека, которые специфически предопределяют эффективность защитных реакций против этих инфекций [5, 6]. Однако генетическая предрасположенность человека к КЭ, в отличие от некоторых других вирусов семейства *Flaviviridae* (лихорадка Денге, гепатит С), сравнительно мало изучена [7].

Ранее мы впервые провели систематические исследования по поиску генов предрасположенности человека к КЭ в российской популяции. Используя подход генов-кандидатов, мы показали, что десять однонуклеотидных полиморфизмов (single nucleotide polymorphism, SNP) в шести генах (*OAS2*, *OAS3*, *CD209*, *TLR3*, *IL28B* и *IL10*), продукты которых являются важнейшими компонентами неспецифического иммунного ответа, ассоциированы с предрасположенностью к КЭ в популяции русских г. Новосибирска [8–11]. Подход с использованием генов-кандидатов позволяет ограничить число анализируемых вариантов генов, но нельзя исключать, что часть вариантов может быть упущена. Поэтому ранее мы впервые применили метод полноэкзомного секвенирования для поиска генов предрасположенности человека к КЭ. Результаты секвенирования экзомов у шести больных с тяжелыми формами КЭ и семи контрольных образцов позволили идентифицировать SNP rs17576 (A/G, Gln279Arg) гена *MMP9* (кодирующего матриксную металлопротеиназу 9), а последующее подтверждение этих данных методом ассоциативного анализа на выборках большего размера позволили рассматривать его в качестве генетического маркера предрасположенности к КЭ в популяции русских [12]. Учитывая, что степень восприимчивости/устойчивости к инфекционному заболеванию — сложный признак, необходим поиск дополнительных генетических локусов, предопределяющих реакцию организма на вирус КЭ, а также течение и исход заболевания. Целью настоящей работы было выявление новых генов, ассоциированных с предрасположенностью к тяжелым формам КЭ в популяции русских. В настоящей работе представлены результаты секвенирования экзомов 22 образцов ДНК больных тяжелыми формами КЭ и 17 образцов контрольной группы с последующей верифика-

цией полученных ассоциаций наиболее перспективных SNPs в расширенной выборке больных с различными формами КЭ и в контрольной популяции.

МАТЕРИАЛЫ И МЕТОДЫ

Пациенты

Были использованы 177 образцов ДНК, выделенных из крови неродственных больных КЭ, проходивших лечение в стационарах Новосибирска в 2002–2017 гг. ($n = 143$), Иркутска в 2014–2015 гг. ($n = 27$) и Красноярска в 2017 г. ($n = 7$). Все больные имели подтвержденный диагноз КЭ согласно общепринятым критериям (клинические симптомы, сезонность, факт укуса клещом, анализ спинномозговой жидкости, серологическая диагностика) [1]. Проведение данного исследования одобрено Комиссией по биоэтике ИЦиГ СО РАН; все пациенты дали письменное информированное согласие на участие в обследовании. Только пациенты, которые указали, что они не подвергались иммунизации до начала заболевания (профилактической вакцинации и/или введению специфического иммуноглобулина после укуса клеща), были включены в данное исследование. Все пациенты были русскими жителями вышеперечисленных городов. Общая группа больных КЭ была разделена на подгруппы больных, перенесших различные клинические формы КЭ: 1) тяжелые паралитические формы с поражением ЦНС (менингоэнцефалитическая, полиомиелитическая, полиоэнцефаломиелитическая) ($n = 65$); 2) более легкие непаралитические формы без поражения ЦНС ($n = 112$), включая лихорадочную ($n = 51$) и менингеальную ($n = 61$) формы.

Контроль

Образцы ДНК контрольной группы были собраны в Новосибирске и Иркутске. Контрольная группа г. Новосибирска ($n = 132$) состояла из двух когорт, собранных в 2003–2005 гг. в рамках программы НАPIEE (Health, Alcohol and Psychological factors In Eastern Europe) ($n = 17$) и в 1993–1994 гг. в рамках программы ВОЗ “MONICA” (MONItoring trends and determinants in Cardiovascular disease) ($n = 115$). Обе когорты включают русских, отобранных случайным образом из избирательных списков одного из районов Новосибирска. Контрольная группа г. Иркутска включала 100 случайно отобранных доноров крови (русские). Учитывая небольшую выборку больных КЭ из Красноярска, контрольная выборка из этого города изучена не была. Информация о реакции лиц, включенных в контрольную группу, на заражение вирусом КЭ отсутствовала.

Секвенирование

Для настоящей работы у 16 больных тяжелыми формами КЭ экзомы были обогащены с помощью набора SureSelect Human All exon V5 Kit (Agilent Technologies, США) и было проведено их секвенирование на платформе HiSeq Illumina 4000 согласно протоколу производителя с использованием парных прочтений длиной 150 пн (компания BGI, Гонконг, КНР). Также использовали ранее полученные данные по экзомам шести больных тяжелыми формами КЭ и 17 контрольных образцов [12, 13].

Выравнивание коротких прочтений на референсный геном и поиск генетических вариантов

После проверки прочтений (ридов) с помощью программы FastQC [14] остатки адаптеров, прочтения с неопределенными нуклеотидами и нуклеотиды низкого качества были удалены с помощью программы Trimmomatic [15] со следующими настройками: LEADING:24, TRAILING:24, SLIDINGWINDOW:4:23, MINLEN:50, параметры сходства для удаления адаптеров: 2 : 30 : 10.

Прочтения, отфильтрованные по качеству, были выравнены на референсный геном Hg19 с помощью программы BWA mem [16], и полученные SAM файлы были конвертированы в BAM формат, а затем отсортированы с помощью программы SAMtools [17]. Дальнейшая обработка файлов и поиск вариантов осуществлялись на основе руководства GATK Best Practices (<https://software.broadinstitute.org/gatk/best-practices/>) с помощью программы GATK 3.8 [18] и включали следующие последовательные шаги: RealignerTargetCreator, IndelRealigner, BaseRecalibrator на основе dbSNP138 [19], HaplotypeCaller, GenotypeGVCFs. На финальном этапе производилась рекалибровка качества полученных вариантов (VariantRecalibrator) на основе баз данных известных вариантов HarMap 3.3, Omni 2.5, dbSNP138 и 1000 genomes [20, 21]. В случае вариантов с инсерциями/делециями использовали базы данных Mills [22] и dbSNP138. Удаление не прошедших фильтр вариантов (ApplyRecalibration) осуществляли на основе 99%-ного порога чувствительности ($-ts_filter_level\ 99.0$) для однонуклеотидных вариантов и 99.5%-ного для таковых с инсерциями/делециями ($-ts_filter_level\ 99.5$).

Поиск генетических ассоциаций

Полученные варианты были аннотированы в программе ANNOVAR [23] для получения информации об аминокислотных заменах, о патогенности вариантов, их частотах в базах данных генетического разнообразия человека, известных ассоциациях с различными заболеваниями.

Поиск генетических ассоциаций проводили с помощью программы Plink 1.9 [24] *assoc fisher-midp mperm = 1000000* с использованием точного теста Фишера и с помощью одного миллиона пермутаций для каждого варианта. Дополнительно для каждого варианта мы подсчитывали соответствие равновесию Харди–Вайнберга отдельно в выборке пациентов и контрольных образцах ($-hardy\ midp$), проводили пермутационный тест на различия в числе негенотипированных индивидов между выборками ($-test-missing\ perm\ midp$). Качество вариантов, демонстрирующих ассоциации, дополнительно проверяли визуально в нескольких индивидуальных BAM файлах с помощью консольного браузера *samtools tview* на отсутствие систематических артефактов выравнивания и покрытия. Потенциально важные варианты отбирались по итогам ручной фильтрации на отсутствие ошибок генотипирования, отклонения от равновесия Харди–Вайнберга, различий в числе негенотипированных образцов между контрольными и больными индивидами с минимальным *P*-значением по Фишеру равным 0.0001.

Генотипирование SNPs генов *ABCB9*, *COL22A1* и *ITGAL*

Генотипирование SNPs rs4148866 гена *ABCB9*, rs4909444 гена *COL22A1* и rs1557672 гена *ITGAL* проводили методом ПЦР с последующим рестрикционным анализом. Для генотипирования по каждому из SNPs были подобраны оригинальные праймеры, при этом в ПЦР-продукт был введен новый сайт рестрикции путем замены нуклеотида в одном из двух праймеров согласно [25]. Основные параметры генотипирования по изученным SNPs приведены в табл. 1. Размеры продуктов ПЦР и рестрикционных фрагментов определяли с помощью электрофореза в 5–6%-ном полиакриламидном геле (ПААГ) и окрашивания раствором бромистого этидия.

Статистические методы

Соответствие частот генотипов равновесию Харди–Вайнберга оценивали с использованием критерия χ^2 и компьютерной программы SHHW [26]. Сравнение частот генотипов и аллелей между изученными группами проводили по точному тесту Фишера с помощью программы SPSS (версия 11.0). Различия между группами считались статистически достоверными при уровне значимости $P < 0.05$.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В результате секвенирования экзотов у 16 пациентов с тяжелыми формами КЭ было получено от 18.6 до 29.8 млн “сырых” прочтений на обра-

Таблица 1. Параметры генотипирования изученных однонуклеотидных полиморфизмов (SNP) генов *ABCB9*, *COL22A1* и *ITGAL*

Ген, SNP	Последовательность праймеров	Температура отжига праймеров, °С	Длина ПЦР-продукта, пн	Эндонуклеаза рестрикции	Длина рестрикционных фрагментов, пн
<i>ABCB9</i> rs4148866	5'-TCCCAGGAGTCGGGTACC-3' 5'-GGTCGATGAACTCGAACACC-3'	60	124	<i>HpaII</i>	A/A: 124 A/G: 124, 106, 18 G/G: 106, 18
<i>COL22A1</i> rs4909444	5'-TCTTCCTTGCCACAGGGCG-3' 5'-AGCTGCTCCTTCAGTGCCTG-3'	63	139	<i>HspAI</i>	T/T: 139 T/G: 139, 121, 18 G/G: 121, 18
<i>ITGAL</i> rs1557672	5'-GCAGCTGCCCTGGGAGCC-3' 5'-TGAGATCCCAGTGCATGAAGTGG-3'	66	147	<i>HpaII</i>	A/A: 147 A/G: 147, 129, 18 G/G: 129, 18

Примечание. Нуклеотиды, обозначенные полужирным шрифтом, были заменены в структуре праймера для образования нового сайта узнавания эндонуклеазой рестрикции [25].

зец. После удаления прочтений с низким качеством секвенирования и наличием длинных технических последовательностей осталось от 84.1 до 85.9% от этого количества. Медиана покрытия для разных образцов составляла от 41× до 63×, средняя покрытия – от 47.7× до 71.3×, что позволяет достаточно надежно выявлять генетические варианты. Процент картированных чтений составлял 99.7–99.9%.

По итогам поиска генетических вариантов было получено 128739 высококачественных однонуклеотидных вариантов и 12562 инсерций/делеций для всей выборки из 39 образцов, включая больных КЭ и контрольную группу. Анализ генетических ассоциаций и ручная фильтрация вариантов по качеству и возможной связи с заболеванием позволили выявить 16 SNPs, потенциально связанных с предрасположенностью к тяжелым формам КЭ, для тестирования их на расширенных выборках (табл. 2). Обращает на себя внимание, что только три из 16 SNPs локализованы в экзонах, а остальные – в интронах и межгенных промежутках; этот результат не является артефактом, поскольку значительное число SNPs за пределами экзонов уже было достоверно выявлено с помощью полноэкзомного секвенирования [27].

Для дальнейшего анализа нами было выбрано три наиболее перспективных SNPs, включая два SNPs, частоты которых больше всего различались между контролем и больными КЭ при тестировании по Фишеру (rs4148866 в интроне гена *ABCB9*, несинонимичный SNP rs4909444 (G/T, Ala938Asp) в экзоне 41 гена *COL22A1*) (табл. 2), а также SNP rs1557672 в интроне гена *ITGAL*, который кодирует белок, вовлеченный в функционирование иммунной системы (см. ниже). Принимая во внимание небольшой размер выборки для полноэкзомного

секвенирования, была проведена верификация наблюдаемых ассоциаций: частоты генотипов и аллелей по этим трем SNPs были определены в расширенной выборке больных различными формами КЭ из Новосибирска, Иркутска и Красноярска (всего 177 человек), включая подгруппы больных очаговыми тяжелыми формами (ТФ) с поражением ЦНС и больных более легкими формами без поражения ЦНС (лихорадочная (ЛФ), менингеальная (МФ)), а также в контрольной популяционной выборке (табл. 3). Следует отметить, что различий по частотам генотипов и аллелей между подгруппами больных КЭ из Новосибирска и Иркутска, а также между контрольными выборками обнаружено не было, поэтому соответствующие группы больных и контроля из этих городов были объединены (данные не приведены). Распределение частот генотипов в контрольной популяционной выборке соответствовало равновесию Харди–Вайнберга.

Не было обнаружено достоверных различий по частотам генотипов и аллелей SNP rs1557672 (G/A) в интроне гена *ITGAL* между больными КЭ и контролем, а также между подгруппами больных ТФ и более легкими формами КЭ и/или контролем (табл. 3). Продукт гена *ITGAL* относится к интегринам, является трансмембранным белком и участвует в процессах взаимодействия лейкоцитов с эндотелиальными клетками и другими типами клеток и др. [28]. То есть, несмотря на то что ген *ITGAL* по данным полноэкзомного секвенирования потенциально может рассматриваться как ген-кандидат, это не подтверждается на расширенной выборке.

SNP rs4148866 (G/A) гена *ABCB9* транспортера расположен в интроне 6 на расстоянии 33 нуклеотидов от 3'-сайта сплайсинга (3'-ss)

Таблица 2. Генетические варианты, ассоциированные с предрасположенностью к тяжелым формам клещевого энцефалита, по результатам полноэкзомного секвенирования

Позиция на хромосоме	Ген, SNP, локализация	Частота аллеля у пациентов	Частота аллеля в контрольной группе	<i>P</i> (по Фишеру)	<i>P</i> (100 тыс. пермутаций)	<i>P</i> -отклонения от равновесия Харди–Вайнберга в контрольной группе
12:123425575	<i>ABCB9</i> rs4148866 (T/C) интрон	0.09091	0.5882	2.25E-06	0.000145	0.2301
8:139701209	<i>COL22A1</i> rs4909444 (T/G) экзон	0.1591	0.6471	9.14E-06	3.00E-05	0.218
12:123460962	<i>OGFOD2</i> rs884956 (T/C) интрон	0.6364	0.1471	1.58E-05	6.00E-05	0.6442
19:623064	<i>POLRMT</i> rs398048312 (T/TC) интрон	0	0.3235	1.84E-05	6.50E-05	0.8042
19:623670	<i>POLRMT</i> rs41554313 (A/G) интрон	0	0.3235	1.84E-05	6.50E-05	0.8042
23:144711814	<i>SPANXN1</i> ; <i>SLITRK2</i> rs1727458 (G/T) межгенный	0.25	0.7353	1.94E-05	0.00044	0.02006
23:144711815	<i>SPANXN1</i> ; <i>SLITRK2</i> rs1781157 (A/T) межгенный	0.25	0.7353	1.94E-05	0.00044	0.02006
12:123664514	<i>MPHOSPH9</i> rs1727301 (A/G) экзон	0.1364	0.5882	2.02E-05	0.000155	0.8122

Таблица 2. Окончание

Позиция на хромосоме	Ген, SNP, локализация	Частота аллеля у пациентов	Частота аллеля в контрольной группе	<i>P</i> (по Фишеру)	<i>P</i> (100 тыс. пермутаций)	<i>P</i> -отклонения от равновесия Харди–Вайнберга в контрольной группе
16:30510571	<i>ITGAL</i> rs1557672 (T/C) интрон	0.1591	0.6176	2.54E-05	0.00013	0.4635
12:123476586	<i>PITPNM2</i> rs7135296 (G/A) интрон	0.6364	0.1562	3.58E-05	0.000205	0.1613
12:123466111	<i>ARL6IP4</i> rs5742290 (C/T) интрон	0.6364	0.1765	4.96E-05	7.00E-05	0.7055
12:123810873	<i>SBNO1</i> rs61388686 (G/A) экзон	0.6364	0.1765	4.96E-05	7.00E-05	0.7055
4:120265443	<i>FABP2; LINC01061</i> rs10006630 (C/A) межгенный	0.4286	0.03333	6.39E-05	0.001165	0.5
9:137623555	<i>COL5A1</i> rs3109675 (C/T) интрон	0.5682	0.125	6.83E-05	0.00028	0.09677
22:30051705	<i>NF2</i> rs769771815 (C/CT) интрон	0	0.2941	6.87E-05	3.00E-05	0.1647
17:31266419	<i>TMEM98</i> rs2726134 (T/C) интрон	0.1136	0.5294	7.74E-05	0.00019	0.8221

Примечание. Полу жирным шрифтом отмечены SNPs, для которых проведено тестирование в расширенных выборках больных и контроля. *P* – уровень значимости.

Таблица 3. Частоты генотипов и аллелей по SNPs генов *ABCВ9*, *COL22A1* и *ITGAL* у больных различными формами клещевого энцефалита (КЭ) и в контрольной популяционной группе

Генотипы и аллели	Частота генотипа/аллеля: % (число ^а)							Р
	Контроль	всего	Больные КЭ				тяжелые формы с поражением ЦНС	
			формы без поражения ЦНС	ЛФ	МФ	МФ		
<i>ABCВ9</i> , rs4148866								
<i>G/G</i>	33.5 (67) ^b	44.1 (78)	36.6 (41) ^c	47.1 (24)	27.9 (17) ^e	56.9 (37) ^{b, c, e}	0.001 ^b , 0.012 ^c , 0.001 ^e	
<i>G/A</i>	50.0 (100)	41.8 (74)	44.6 (50)	43.1 (22)	45.9 (28)	36.9 (24)	>0.05	
<i>A/A</i>	16.5 (33) ^b	14.1 (25)	18.8 (21) ^c	9.8 (5)	26.2 (16) ^e	6.2 (4) ^{b, c, e}	0.039 ^b , 0.024 ^c , 0.003 ^e	
<i>G</i>	58.5 ^b	65.0	58.9 ^c	68.8	50.8 ^e	75.4 ^{b, c, e}	0.001 ^b , 0.002 ^c , <0.001 ^e	
<i>A</i>	41.5 ^b	35.0	41.1 ^c	31.4	49.2 ^e	24.6 ^{b, c, e}	0.001 ^b , 0.002 ^c , <0.001 ^e	
<i>n</i>	200	177	112	51	61	65		
χ^2 HW	0.18							
<i>COL22A1</i> , rs4909444								
<i>G/G</i>	47.0 (101)	47.3 (79)	42.9 (45)	42.6 (20)	43.1 (25)	54.8 (34)	>0.05	
<i>G/T</i>	40.0 (86)	39.5 (66)	40.0 (42)	31.9 (15)	46.6 (27)	38.7 (24)	>0.05	
<i>T/T</i>	13.0 (28)	13.2 (22)	17.1 (18) ^c	25.5 (12) ^d	10.3 (6)	6.5 (4) ^{c, d}	0.007 ^d	
<i>G</i>	67.0	67.1	62.9	58.5	66.4	74.2	>0.05	
<i>T</i>	33.0	32.9	37.1	41.5	33.6	25.8	>0.05	
<i>n</i>	215	167	105	47	58	62		
χ^2 HW	1.97							
<i>ITGAL</i> , rs1557672								
<i>G/G</i>	62.7 (126)	64.5 (93)	66.3 (65)	67.5 (29)	65.4 (36)	60.9 (28)	>0.05	
<i>G/A</i>	31.3 (63)	29.9 (43)	28.6 (28)	30.2 (13)	27.3 (15)	32.6 (15)	>0.05	
<i>A/A</i>	6.0 (12)	5.6 (8)	5.1 (5)	2.3 (1)	7.3 (4)	6.5 (3)	>0.05	
<i>G</i>	78.4	79.5	80.6	82.6	79.1	77.2	>0.05	
<i>A</i>	21.6	20.5	19.4	17.4	20.9	22.8	>0.05	
<i>n</i>	201	144	98	43	55	46		
χ^2 HW	1.16							

Примечание: КЭ – клещевой энцефалит; ЦНС – центральная нервная система; ЛФ – лихорадочная форма; МФ – менингеальная форма; *n* – размер выборки; *P* – уровни значимости; χ^2 HW – значения теста χ^2 на соответствие распределения частот генотипов равновесию Харди–Вайнберга.
^а Число человек с данным генотипом.
^{б-с} Уровни значимости при сравнении больных тяжелыми формами КЭ с поражением ЦНС и контролем и с больными другими формами КЭ: ^b с контролем, ^c с больными КЭ без поражения ЦНС (ЛФ + МФ), ^d с больными ЛФ КЭ, ^e с больными МФ КЭ.

(https://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=4148866). Нами обнаружено увеличение частоты гомозигот *G/G* у больных ТФ КЭ (56.9%) по сравнению с больными МФ (27.9%) ($P = 0.001$), объединенной группой больных с ЛФ и МФ (ЛФ + МФ) (36.6%) ($P = 0.012$) и контролем (33.5%) ($P = 0.001$). Частота аллеля *G* также повышена у больных ТФ КЭ (75.4%) по сравнению с больными МФ (50.8%) ($P < 0.001$), ЛФ + МФ (58.9%) ($P = 0.002$) и контрольной группой (58.5%) ($P = 0.001$). Для этого SNP показано уменьшение частоты генотипа *A/A* у больных ТФ КЭ (6.2%) по сравнению с больными МФ (26.2%) ($P = 0.003$), ЛФ + МФ (18.8%) ($P = 0.024$) и контролем (16.5%) ($P = 0.039$) (табл. 3). Для SNP rs4909444 (*G/T*, Ala938Asp) гена *COL22A1* было выявлено достоверное уменьшение частоты гомозигот *T/T* у больных ТФ КЭ (6.5%) по сравнению с больными ЛФ (25.5%) ($P = 0.007$). Частота данного генотипа понижена в 2 раза у больных ТФ по сравнению с контрольной группой (13.0%), однако различия не достигают порога достоверности (табл. 3).

В результате ассоциация двух из трех SNPs, выявленных методом полноэкзомного секвенирования, с предрасположенностью к тяжелым формам КЭ была подтверждена при анализе расширенной выборки образцов больных и контроля. Поиск литературы в базе данных PubMed по ключевым словам “ABCВ9, tick-borne encephalitis” и “COL22A1, tick-borne encephalitis” не выявляет публикаций, в которых было бы продемонстрировано участие этих белков в различных аспектах патогенеза КЭ. Таким образом, в настоящей работе впервые выявлено возможное влияние продуктов генов *ABCВ9* и *COL22A1* на результат инфекции человека вирусом КЭ. Ген *ABCВ9*, локализованный на хромосоме 12 (12q24.31), кодирует белок, относящийся к суперсемейству ABC (ATP-Binding Cassette) мембранных транспортеров. Эти белки представлены у всех живых организмов; 50 таких белков обнаружены у человека [29]. Они трансформируют химическую энергию гидролиза АТФ в механическую для переноса субстратов (пептидов) через биологические мембраны [30, 31]. ABC-семейство белков подразделяют на семь подсемейств (A–G), а также на два типа по характеру организации структуры белка. Полный тип содержит два АТФ-связывающих домена (nucleotide-binding domain, NBD) и два субстратных домена с трансмембранными спиралями (membrane-spanning domain, MSD); другой тип (половинный) имеет по одному NBD и MSD. Белок *ABCВ9* (второе название – TAPL, transporter associated with antigen processing like) относится к половинному типу (с 10 возможными трансмембранными спиралями в MSD) подсемейства В. Члены этого семейства участвуют в презентации антигенов и перемещении пептидов из цитозоля в просвет лизосомы [29, 32–34]. Известно не-

сколько вариантов белка *ABCВ9* в результате альтернативного сплайсинга про-мРНК этого гена: три варианта белка, кодируемых экзонами 2–11 с различными фрагментами экзона 12, и белок с делецией, кодируемой экзоном 6 [35]. Этот экзон кодирует последовательность из 43 аминокислот в полноразмерном белке, включающую две последние трансмембранные спирали в MSD (ТМ9 и ТМ10), проксимальные к NBD [35–37]. Полагают, что делеция изменяет субстратную специфичность ABC-транспортера. Следует отметить, что изоформа белка с этой делецией обнаружена у человека и грызунов [37], что указывает на консервативность явления и высокую вероятность событий альтернативного сплайсинга; тканеспецифическая экспрессия данной изоформы наблюдается в яичниках и в мозгу [35]. Изучаемый нами SNP rs4148866 расположен в интроне 6 на расстоянии 21 нуклеотид от 3'-ss AG динуклеотида и полипиримидинового (Py) тракта (строго соответствующего в интроне 6 консенсусу [38]) и вблизи несовершенного мотива branch point (BP), с которыми связывается один из центральных регуляторов сплайсинга (белок U2AF, участвующий в формировании сплайсосомы) [39]. Можно предположить, что аллель *G* SNP rs4148866 благоприятствует формированию сплайсосомы на участке 3'-ss конца интрона 6 и 5'-ss впереди экзона 6 в про-мРНК *ABCВ9* транспортера. Изменение субстратной специфичности белка *ABCВ9*, обусловленное делецией, может ухудшить эффективность транслокации к лизосомам продуктов гидролиза протеазами, в том числе длинных пептидов с РНК-зависимой РНК-полимеразной активностью вируса [34]. Это обстоятельство может стимулировать репликацию антисенс-РНК, далее сенс-РНК и генерализацию процесса инфекции вирусом КЭ, т.е. приводить к тяжелым формам болезни [40].

Ген *COL22A1* локализован на хромосоме 8 (8q24.3) и кодирует коллаген типа XXII, относящийся к фибрилл-ассоциированным коллагенам (FACIT). На настоящий момент этот белок мало изучен; известно, что он вносит вклад в стабилизацию мышечно-сухожильных соединений [41, 42]. Механизмы, посредством которых продукт гена *COL22A1* может быть вовлечен в процессы патогенеза КЭ, не ясны и требуют дальнейшего изучения.

Таким образом, в настоящей работе с применением метода полноэкзомного секвенирования были идентифицированы два новых генетических маркера (SNPs rs4148866 гена *ABCВ9* и rs4909444 гена *COL22A1*), связанных, по нашим данным, с предрасположенностью к тяжелым формам КЭ с поражением ЦНС в популяции русских. Принимая во внимание, что полиморфизм этих генов впервые был исследован на возможную связь с восприимчивостью/устойчивостью

человека к вирусу КЭ, дальнейшее изучение ассоциации этих генов с использованием независимых выборок, выборок большего размера, а также в популяциях других этнических групп, живущих на эндемичных по вирусу КЭ территориях, представляется целесообразным. Идентификация генов, участвующих в формировании предрасположенности человека к КЭ, важна для улучшения понимания механизмов взаимодействия между вирусом и хозяином при инфекции вирусом КЭ и, соответственно, патогенеза заболевания, а также для выявления групп повышенного риска развития тяжелых форм КЭ.

Работа была поддержана грантом Российского научного фонда (№ 16-15-00127).

СПИСОК ЛИТЕРАТУРЫ

1. *Gritsun T.S., Lashkevich V.A., Gould E.A.* Tick-borne encephalitis // *Antiviral Res.* 2003. V. 57. P. 129–146. doi 10.1016/S0166-3542(02)00206-1
2. *Ruzek D., Dobler G., Donoso Mantke O.* Tick-borne encephalitis: pathogenesis and clinical implications // *Travel Med. Infect. Dis.* 2010. V. 8. P. 223–232. doi 10.1016/j.tmaid.2010.06.004
3. *Suss J.* Tick-borne encephalitis 2010: epidemiology, risk areas, and virus strains in Europe and Asia—an overview // *Ticks Tick-Borne Dis.* 2011. V. 2. P. 2–15. doi 10.1016/j.ttbdis.2010.10.007
4. *Bogovic P., Strle F.* Tick-borne encephalitis: a review of epidemiology, clinical characteristics, and management // *World J. Clin. Cases.* 2015. V. 3. P. 430–441. doi 10.12998/wjcc.v3.i5.430
5. *Chapman S.J., Hill A.V.* Human genetic susceptibility to infectious disease // *Nat. Rev. Genet.* 2012. V. 13. P. 175–188. doi 10.1038/nrg3114
6. *Khor C.C., Hibberd M.L.* Host-pathogen interactions revealed by human genome-wide surveys // *Trends Genet.* 2012. V. 28. P. 233–243. doi 10.1016/j.tig.2012.02.001
7. *Юдин Н.С., Бархаш А.В., Максимов В.Н. и др.* Генетическая предрасположенность человека к заболеваниям, вызываемым вирусами семейства *Flaviviridae* // *Мол. биология.* 2018. Т. 52. № 2. С. 190–209.
8. *Barkhash A.V., Perelygin A.A., Babenko V.N. et al.* Variability in the 2'-5'-oligoadenylate synthetase gene cluster is associated with human predisposition to tick-borne encephalitis virus-induced disease // *J. Infect. Dis.* 2010. V. 202. № 12. P. 1813–1818. doi 10.1086/657418
9. *Barkhash A.V., Perelygin A.A., Babenko V.N. et al.* Single nucleotide polymorphism in the promoter region of the *CD209* gene is associated with human predisposition to severe forms of tick-borne encephalitis // *Antiviral Res.* 2012. V. 93. № 1. P. 64–68. doi 10.1016/j.antiviral.2011.10.017
10. *Barkhash A.V., Voevoda M.I., Romaschenko A.G.* Association of single nucleotide polymorphism rs3775291 in the coding region of the *TLR3* gene with predisposition to tick-borne encephalitis in a Russian population // *Antiviral Res.* 2013. V. 99. № 2. P. 136–138. doi 10.1016/j.antiviral.2013.05.008
11. *Barkhash A.V., Babenko V.N., Voevoda M.I., Romaschenko A.G.* Association of *IL28B* and *IL10* gene polymorphism with predisposition to tick-borne encephalitis in a Russian population // *Ticks Tick-Borne Dis.* 2016. V. 7. № 5. P. 808–812. doi 10.1016/j.ttbdis.2016.03.019
12. *Barkhash A.V., Yurchenko A.A., Yudin N.S. et al.* A matrix metalloproteinase 9 (*MMP9*) gene single nucleotide polymorphism is associated with predisposition to tick-borne encephalitis virus-induced severe central nervous system disease // *Ticks Tick-Borne Dis.* 2018. V. 9. № 4. P. 763–767. doi 10.1016/j.ttbdis.2018.02.010
13. *Ovsyannikova A.K., Rymar O.D., Shakhshneider E.V. et al.* ABCC8-related maturity-onset diabetes of the young (MODY12): clinical features and treatment perspective // *Diabetes Ther.* 2016. V. 7. № 3. P. 591–600. doi 10.1007/s13300-016-0192-9
14. *Andrews S.* FastQC: a quality control tool for high throughput sequence data. Version 0.11.2. Cambridge, UK: Babraham Institute, 2014. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
15. *Bolger A.M., Lohse M., Usadel B.* Trimmomatic: a flexible trimmer for Illumina sequence data // *Bioinformatics.* 2014. V. 30. № 15. P. 2114–2120. doi 10.1093/bioinformatics/btu170
16. *Li H., Durbin R.* Fast and accurate short read alignment with Burrows–Wheeler transform // *Bioinformatics.* 2009. V. 25. № 14. P. 1754–1760. doi 10.1093/bioinformatics/btp324
17. *Li H., Handsaker B., Wysoker A. et al.* The sequence alignment/map format and SAMtools // *Bioinformatics.* 2009. V. 25. № 16. P. 2078–2079. doi 10.1093/bioinformatics/btp352
18. *McKenna A., Hanna M., Banks E. et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data // *Genome Res.* 2010. V. 20. № 9. P. 1297–1303. doi 10.1101/gr.107524.110
19. *Sherry S.T., Ward M.H., Kholodov M. et al.* dbSNP: the NCBI database of genetic variation // *Nucl. Acids Res.* 2001. V. 29. № 1. P. 308–311.
20. *International HapMap Consortium et al.* A second generation human haplotype map of over 3.1 million SNPs // *Nature.* 2007. V. 449. № 7164. P. 851–861.
21. *1000 Genomes Project Consortium et al.* An integrated map of genetic variation from 1,092 human genomes // *Nature.* 2012. V. 491. № 7422. P. 56–65. doi 10.1038/nature11632
22. *Mills R.E., Pittard W.S., Mullaney J.M. et al.* Natural genetic variation caused by small insertions and deletions in the human genome // *Genome Res.* 2011. V. 21. № 6. P. 830–839. doi 10.1101/gr.115907.110
23. *Wang K., Li M., Hakonarson H.* ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data // *Nucl. Acids Res.* 2010. V. 38. № 16. e164. doi 10.1093/nar/gkq603
24. *Purcell S., Neale B., Todd-Brown K. et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses // *Am. J. Hum. Gen.* 2007. V. 81. № 3. P. 559–575.

25. *Neff M.M., Turk E., Kalishman M.* Web-based primer design for single nucleotide polymorphism analysis // *Trends Genet.* 2002. V. 18. № 12. P. 613–615. doi 10.1016/S0168-9525(02)02820-2
26. *Zaykin D.V., Pudovkin A.I.* Two programs to estimate significance of χ^2 values using pseudo-probability tests // *J. Hered.* 1993. V. 84. P. 152.
27. *Guo Y., Long J., He J. et al.* Exome sequencing generates high quality data in non-target regions // *BMC Genomics.* 2012. V. 13. P. 194. doi 10.1186/1471-2164-13-194
28. *Wang Y., Shu Y., Xiao Y. et al.* Hypomethylation and overexpression of ITGAL (CD11a) in CD4(+) T cells in systemic sclerosis // *Clin. Epigenetics.* 2014. V. 6. P. 25. doi 10.1186/1868-7083-6-25
29. *Gizaw M., Anandakumar P.* A Review on ATP binding cassette (ABC) transporters // *Int. J. Pharma Res. Health Sci.* 2017. V. 5. № 2. P. 1607–1615. doi 10.21276/ijprhs.2017.02.01
30. *Holland I.B., Blight M.A.* ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans // *J. Mol. Biol.* 1999. V. 293. № 2. P. 381–399. doi 10.1006/jmbi.1999.2993
31. *Jones P.M., George A.M.* The ABC transporter structure and mechanism: perspectives on recent research // *Cell. Mol. Life Sci.* 2004. V. 61. № 6. P. 682–699. doi 10.1007/s00018-003-3336-9
32. *Ohara T., Ohashi-Kobayashi A., Maeda M.* Biochemical characterization of transporter associated with antigen processing (TAP)-like (ABCB9) expressed in insect cells // *Biol. Pharm. Bull.* 2008. V. 31. № 1. P. 1–5. doi 10.1248/bpb.31.1
33. *Bangert I., Tumulka F., Abele R.* The lysosomal polypeptide transporter TAPL: more than a housekeeping factor? // *Biol. Chem.* 2011. V. 392. № 1–2. doi 10.1515/BC.2011.007
34. *Zollmann T., Bock C., Graab P., Abele R.* Team work at its best – TAPL and its two domains // *Biol. Chem.* 2015. V. 396. № 9–10. doi 10.1515/hsz-2014-0319
35. *Kobayashi A., Hori S., Suita N., Maeda M.* Gene organization of human transporter associated with antigen processing-like (TAPL, ABCB9): analysis of alternative splicing variants and promoter activity // *Biochem. Biophys. Res. Commun.* 2003. V. 309. № 4. P. 815–822. doi 10.1016/j.bbrc.2003.08.081
36. *Zhao C., Tampé R., Abele R.* TAP and TAP-like – brothers in arms? // *Naunyn Schmiedebergs Arch. Pharmacol.* 2006. V. 372. № 6. P. 444–450. doi 10.1007/s00210-005-0028-z
37. *Zhang F., Zhang W., Liu L. et al.* Characterization of ABCB9, an ATP binding cassette protein associated with lysosomes // *J. Biol. Chem.* 2000. V. 275. № 30. P. 23287–23294. doi 10.1074/jbc.M001819200
38. *Burset M., Seledtsov I.A., Solovvey V.V.* SpliceDB: database of canonical and non-canonical mammalian splice sites // *Nucleic Acids Res.* 2001. V. 29. № 1. P. 255–259.
39. *Förch P., Merendino L., Martínez C., Valcárcel J.* U2 small nuclear ribonucleoprotein particle (snRNP) auxiliary factor of 65 kDa, U2AF65, can promote U1 snRNP recruitment to 5' splice sites // *Biochem J.* 2003. V. 372. P. 235–240. doi 10.1042/BJ20021202
40. *Lindenbach B.D., Thiel H.-J., Rice C.M.* Flaviviridae: the viruses and their replication // *Fields Virology*. 5th ed. Philadelphia: Lippincott-Raven Publ., 2007. P. 1101–1152.
41. *Koch M., Schulze J., Hansen U. et al.* A novel marker of tissue junctions, collagen XXII // *J. Biol. Chem.* 2004. V. 279. № 21. P. 22514–22521. doi 10.1074/jbc.M400536200
42. *Charvet B., Guiraud A., Malbouyres M. et al.* Knockdown of col22a1 gene in zebrafish induces a muscular dystrophy by disruption of the myotendinous junction // *Development.* 2013. V. 140. № 22. P. 4602–4613. doi 10.1242/dev.096024

Association of *ABCB9* and *COL22A1* Gene Polymorphism with Human Predisposition to Severe Forms of Tick-Borne Encephalitis

A. V. Barkhash^{a,*}, A. A. Yurchenko^a, N. S. Yudin^{a,b}, I. V. Kozlova^c, I. A. Borishchuk^d, M. V. Smolnikova^e, O. I. Zaitseva^e, L. L. Pozdnyakova^f, M. I. Voevoda^{a,b}, and A. G. Romaschenko^g

^aFederal Research Center Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, 630090 Russia

^bNovosibirsk State University, Novosibirsk, 630090 Russia

^cScientific Centre for Family Health and Human Reproduction Problems, Irkutsk, 664003 Russia

^dIrkutsk Regional Infectious Clinical Hospital, Irkutsk, 664043 Russia

^eScientific Research Institute of Medical Problems of the North, Federal Research Center “Krasnoyarsk Science Center of Siberian Branch of Russian Academy of Sciences”, Krasnoyarsk, 660022 Russia

^fCity Infectious Clinical Hospital no. 1, Novosibirsk, 630099 Russia

*e-mail: barkhash@bionet.nsc.ru

Tick-borne encephalitis (TBE) is caused by a neurotropic RNA virus from the *Flavivirus* genus. TBE is characterized by a significant variability of clinical manifestations from nonparalytic forms (fever, meningitis) to severe paralytic (focal) forms (meningoencephalitis, poliomyelitis, polioencephalomyelitis). The result of interaction between a virus and a host (and, consequently, the viral disease course and outcome) largely de-

pends on genetically determined ability of the host (particularly, human) organism immune system to suppress the development of viral infection. However, hereditary predisposition to TBE has been rather poorly studied in human populations to date. In this study, the results of whole-exome sequencing of DNA samples from 22 Russian non-immunized TBE patients with severe disease and 17 control individuals from the same population are presented. As a result, sixteen single nucleotide polymorphisms (SNPs) associated with predisposition to severe TBE were identified. The genotype and allele frequencies for three of these SNPs localized in the *ABCB9* (rs4148866, G/A, intron), *COL22A1* (rs4909444, G/T, Ala938Asp), and *ITGAL* (rs1557672, G/A, intron) genes were then studied in larger samples of patients with different clinical forms of TBE ($n = 177$) and in the control population ($n = 215$). As a result, the association of the *ABCB9* and *COL22A1* gene SNPs with the development of severe forms of TBE was for the first time demonstrated in the Russian population. The hypothesis regarding a mechanism of possible effect of the *ABCB9* gene intronic SNP on the process of human infection with TBE virus is considered.

Keywords: tick-borne encephalitis, genetic predisposition, whole-exome sequencing, *ABCB9* gene, *COL22A1* gene.