

Приоритизация генов, ассоциированных с патогенезом лейкоза у крупного рогатого скота

Н.С. Юдин^{1, 2}, Н.Л. Подколотный^{1, 3}, Т.А. Агаркова⁴, Е.В. Игнатьева^{1, 2}

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

² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

³ Институт вычислительной математики и математической геофизики Сибирского отделения Российской академии наук, Новосибирск, Россия

⁴ Сибирский федеральный научный центр агробиотехнологий Российской академии наук, р.п. Краснообск, Новосибирская область, Россия

Перспективным подходом для искоренения инфекционных болезней сельскохозяйственных животных, особенно при отсутствии эффективных методов лечения и профилактики, считается направленная селекция по генетическим маркерам (однонуклеотидным полиморфизмам, минисателлитам). Вирус лейкоза крупного рогатого скота (ВЛКРС) распространен по всему миру и представляет одну из основных проблем для развития животноводства и продовольственной безопасности в России. Однако, как показали недавние полногеномные исследования, чувствительность/резистентность к инфекции ВЛКРС носит полигенный характер. Целью исследования было создание каталога генов крупного рогатого скота и других видов млекопитающих, вовлеченных в процесс патогенеза инфекции ВЛКРС, и приоритизация генов каталога с помощью методов биоинформатики. Путем ручного сбора информации из открытых источников нами создан каталог генов крупного рогатого скота и других видов млекопитающих, вовлеченных в процесс патогенеза инфекции ВЛКРС, включающий 446 генов. Приоритизацию генов каталога осуществляли на основе следующих критериев: 1) ген ассоциирован с лейкозом по данным полногеномного анализа ассоциаций; 2) ген ассоциирован с лейкозом по данным анализа ассоциаций «случай–контроль»; 3) роль гена в развитии лейкоза изучена в экспериментах по нокауту у мышей; 4) белок, кодируемый геном, участвует в белок-белковых взаимодействиях с вирусной частицей либо отдельным вирусным белком; 5) ген имеет аннотацию терминами Gene Ontology, которые являются перепредставленными для данного списка генов; 6) ген участвует в биологических путях из баз KEGG или REACTOME, которые являются перепредставленными для данного списка генов; 7) белок, кодируемый геном, имеет неслучайно высокое количество белок-белковых взаимодействий с белками, кодируемыми другими генами из каталога. На основе каждого критерия гену присваивали ранг. Затем ранги суммировали и определяли общий ранг. Приоритизация списка, включающего 446 генов, позволила выделить пять наиболее вероятных генов-кандидатов (*TNF*, *LTB*, *BOLA-DQA1*, *BOLA-DRB3*, *ATF2*), которые могут влиять на чувствительность/устойчивость крупного рогатого скота к заболеванию лейкозом.

Ключевые слова: вирус лейкоза крупного рогатого скота; полногеномный анализ ассоциаций; ген; однонуклеотидный полиморфизм; ассоциация; приоритизация; GO анализ; белок-белковое взаимодействие.

Prioritization of genes associated with the pathogenesis of leukosis in cattle

N.S. Yudin^{1, 2}, N.L. Podkolodny^{1, 3},
T.A. Agarkova⁴, E.V. Ignatieva^{1, 2}

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

³ Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia

⁴ Siberian Federal Research Center of Agro-BioTechnologies, RAS, Krasnoobsk, Novosibirsk region, Russia

Selection by means of genetic markers is a promising approach to the eradication of infectious diseases in farm animals, especially in the absence of effective methods of treatment and prevention. Bovine leukemia virus (BLV) is spread throughout the world and represents one of the biggest problems for the livestock production and food security in Russia. However, recent genome-wide association studies have shown that sensitivity/resistance to BLV is polygenic. The aim of this study was to create a catalog of cattle genes and genes of other mammalian species involved in the pathogenesis of BLV-induced infection and to perform gene prioritization using bioinformatics methods. Based on manually collected information from a range of open sources, a total of 446 genes were included in the catalog of cattle genes and genes of other mammals involved in the pathogenesis of BLV-induced infection. The following criteria were used to prioritize 446 genes from the catalog: (1) the gene is associated with leukemia according to a genome-wide association study; (2) the gene is associated with leukemia according to a case-control study; (3) the role of the gene in leukemia development has been studied using knockout mice; (4) protein-protein interactions exist between the gene-encoded protein and either viral particles or individual viral proteins; (5) the gene is annotated with Gene Ontology terms that are over-represented for a given list of genes; (6) the gene participates in biological pathways from the KEGG or REACTOME databases, which are over-represented for a given list of genes; (7) the protein encoded by the gene has a high number of protein-protein interactions with proteins encoded by other genes from the catalog. Based on each criterion, a rank was assigned to each gene. Then the ranks were summarized and an overall rank was determined. Prioritization of 446 candidate genes allowed us to identify 5 genes of interest

(*TNF, LTB, BOLA-DQA1, BOLA-DRB3, ATF2*), which can affect the sensitivity/resistance of cattle to leukemia.

Key words: bovine leukemia virus; genome-wide association study; gene; single-nucleotide polymorphism; association; prioritization; GO analysis; protein-protein interactions.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Юдин Н.С., Подколотный Н.Л., Агаркова Т.А., Игнатиева Е.В. Приоритизация генов, ассоциированных с патогенезом лейкоза у крупного рогатого скота. Вавиловский журнал генетики и селекции. 2018;22(8):1063-1069. DOI 10.18699/VJ18.451

HOW TO CITE THIS ARTICLE:

Yudin N.S., Podkolodny N.L., Agarkova T.A., Ignatieva E.V. Prioritization of genes associated with the pathogenesis of leukosis in cattle. Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding. 2018;22(8):1063-1069. DOI 10.18699/VJ18.451 (in Russian)

Вирус лейкоза крупного рогатого скота (ВЛКРС) – это онкогенный дельта-ретровирус, инфицирующий крупный рогатый скот (КРС), овец и другие виды млекопитающих. ВЛКРС родственен Т-лимфотропному вирусу человека типа 1 (HTLV-1) и вирусу иммунодефицита человека типа 1 (HIV-1) (Coffin et al., 2002). Инфекция ВЛКРС протекает у большинства животных бессимптомно, тогда как у 30–40 % развивается персистентный лимфоцитоз, который характеризуется размножением инфицированных провирусом В-лимфоцитов и повышением числа лейкоцитов в периферической крови КРС. Менее чем у 10 % инфицированных животных развивается тяжелая форма заболевания, которая приводит к появлению летальных лимфом или лимфосарком в различных органах (Barez et al., 2015). ВЛКРС распространен по всему миру и переносится между животными через кровь, молоко или другие биологические жидкости, которые содержат ДНК провирусов (Gutiérrez et al., 2015). В России лейкоз КРС представляет одну из основных проблем для развития животноводства и продовольственной безопасности. В ряде хозяйств поголовье инфицированных животных составляет до 40 %. По данным за 2010–2014 гг., экономические потери от лейкоза КРС в России достигают 2.5 млрд руб. в год (Stepanova, 2016).

Традиционные меры борьбы с лейкозом включают изоляцию или ликвидацию инфицированного скота по результатам серологической и гематологической диагностики инфекции ВЛКРС, но эти меры экономически неприемлемы в хозяйствах с высокой долей инфицированных животных (Rodríguez et al., 2011). Перспективным подходом для ликвидации инфекционных болезней сельскохозяйственных животных, особенно при отсутствии эффективных методов лечения и профилактики, считается направленная селекция по генетическим маркерам, ассоциированным с устойчивостью к инфекции организма-хозяина (Bishop, 2014). Например, при геномной оценке молочного скота уже сейчас учитывается устойчивость к маститу (Koeck et al., 2012).

Известно, что некоторые породы КРС обладают врожденной устойчивостью к инфицированию ВЛКРС (FAO, 2007). Путем анализа фенотипов и родословных (Abdalla et al., 2013), а также с использованием матрицы аддитивной генетической дисперсии–ковариации (Abdalla et al., 2016) было показано, что коэффициент наследуемости чувствительности к ВЛКРС у животных голштинской и джерсейской пород составляет около 0.08. Проведенные ранее многочисленные исследования показали, что ДНК-

маркеры в генах-кандидатах (например, определенные аллели в гене *BOLA-DRB3*, кодирующем белок DRB3, относящийся к главному комплексу гистосовместимости МНС II) вносят существенный вклад в вариабельность к инфицированию ВЛКРС между отдельными животными (Xu et al., 1993; Juliarena et al., 2008; Miyasaka et al., 2013). При этом одни аллели в гене *BOLA-DRB3* ассоциированы с чувствительностью/резистентностью к ВЛКРС у животных голштинской породы и не ассоциированы у животных джерсейской породы. Однако, как показал недавний полногеномный анализ ассоциации (GWAS) чувствительности/резистентности к инфекции, чувствительность к ВЛКРС носит полигенный характер (Abdalla et al., 2016; Brym et al., 2016; Carignano et al., 2018).

Целью исследования было создание каталога генов КРС и других видов млекопитающих, вовлеченных в процесс патогенеза инфекции ВЛКРС, а также приоритизация генов каталога с помощью методов биоинформатики.

Материалы и методы

Создание каталога генов. Набор ключевых слов для поиска публикаций в PubMed был сформирован на основе методики, использованной нами ранее для поиска информации о генах человека, участвующих в ответе на вирус клещевого энцефалита (Ignatieva et al., 2017). В ходе этой работы был проведен экспертный анализ абстрактных и полных текстов научных статей, полученных из PubMed по названию вируса. Выявлены следующие типы экспериментов, указывающих на причастность того или иного гена к ответу организма-хозяина на вирус клещевого энцефалита: 1) идентификация аллельных вариантов генов, ассоциированных с тяжестью заболевания; 2) обнаружение влияния нокаута определенного гена на клинические показатели, характеризующие ход заболевания; 3) анализ дифференциальной экспрессии генов *in vivo* у больных людей и инфицированных животных либо *in vitro* в клетках, зараженных вирусом; 4) выявление прямых физических взаимодействий между белками организма-хозяина и вирусными белками (либо РНК) *in vitro*. Поэтому для поиска публикаций в PubMed, содержащих данные о генах, вовлеченных в патогенез лейкоза у крупного рогатого скота, нами были использованы следующие комбинации ключевых слов: 1) (Bovine leukosis OR Bovine leukemia virus) AND association; 2) (Bovine leukosis OR Bovine leukemia virus) AND Knockout; 3) (Bovine leukosis OR Bovine leukemia virus) AND Expression; 4) (Bovine leukosis OR Bovine leukemia virus) AND (PPI OR Physical interactions).

Данные о генах/белках экстрагировали из отобранных публикаций вручную. Названия генов корректировали в соответствии с их официальными названиями из базы данных Entrez Gene. Сведения о каждом гене/белке помещали в определенную категорию (таблицу) в соответствии с типом доказательства его участия в патогенезе инфекционного процесса. Наименование категорий (таблиц) и их описание приведены в табл. 1.

Функциональная аннотация генов осуществлялась с помощью системы DAVID (Database for Annotation, Visualization and Integrated Discovery) (Huang et al., 2009). В качестве референсного списка генов использовали список генов из полного генома *Bos taurus*. Термины из словаря базы Gene Ontology и биологические пути из баз KEGG и REACTOME считали перепредставленными для данного списка генов при $FDR < 0.05$ и $Fold\ Enrichment > 3$.

Обучающее множество генов содержало 21 ген из созданного нами каталога. Каждый из них удовлетворял хотя бы одному критерию (табл. S14)¹. Первый критерий – ген содержался также в диаграмме bta05166 базы KEGG Pathway, описывающей процесс инфицирования клеток организма-хозяина вирусом HTLV-1 (Human T-cell leukemia virus 1). Таких генов было найдено 19. Вирус HTLV-1 является наиболее близкородственным представителем того же семейства (*Deltaretrovirus*), что и ВЛКРС (Coffin et al., 2002), на основе чего можно ожидать, что процесс инфицирования вирусами HTLV-1 и ВЛКРС активирует сходные процессы в клетках организма-хозяина. Второй критерий – ген имел следующие экспериментальные свидетельства (см. табл. 1) его вовлеченности в ответ на инфекцию вирусом ВЛКРС: аллельные варианты по данным экспериментов «случай–контроль»; нокаут; наличие физических (белок-белковых) взаимодействий с белками ВЛКРС (табл. S2, S3, S6). Таких генов было выявлено шесть.

Реконструкция сетей белок-белковых взаимодействий. Данные о белок-белковых взаимодействиях экстрагировали из базы данных IID (Integrated Interactions Database, версия от 2018-05) (Kotlyar et al., 2016).

Поскольку протеом вида *Bos taurus* изучен еще не в достаточной степени, поиск белок-белковых взаимодействий производили для списка генов человека, ортологичных генам из каталога. Поиск соответствия между символами генов вида *Bos taurus* и символами ортологичных генов вида *Homo sapiens* осуществляли в программе Ortholog Conversions (<https://biobdnet-abcc.ncifcrf.gov/db/dbOrtho.php>) ресурса bioDBnet. В отдельных случаях поиск ортологов выполняли с помощью геномного браузера UCSC genome browser (<https://genome.ucsc.edu/>), который визуализирует выравнивание последовательности ДНК искомого гена с последовательностями ортологичных генов других видов.

На основе данных из базы IID нами были реконструированы две сети: сеть всех белок-белковых взаимодействий между генами/белками человека (H_PPI) и сеть белок-белковых взаимодействий для генов/белков, ассоциированных с лейкозом (L_PPI). Сеть H_PPI содержала $N = 18238$ вершин (белков), 1562726 ребер (взаимодей-

ствий) и имела среднюю степень вершины 171. Сеть L_PPI содержала $n = 428$ вершин, 2352 ребра и имела среднюю степень вершины, равную 11.

Расчет значимости генов на основе сети белок-белковых взаимодействий. Для расчета индекса значимости или веса w_i некоторого гена g_i , кодирующего белок p_i , была использована следующая процедура.

1. Формирование для белка p_i таблицы 2×2 , описывающей соотношение числа белков, взаимодействующих с белком p_i в сети L_PPI и в сети H_PPI.

Сеть*	Число белков		Размер сети за вычетом белка p_i
	взаимодействующих с белком p_i	не взаимодействующих с белком p_i	
Сеть L_PPI	$n_{11} = d_i$	$n_{12} = n - 1 - d_i$	$n_{1.} = n - 1$
Сеть H_PPI \ L_PPI	$n_{21} = D_i - d_i$	$n_{22} = N - D_i - n + d_i$	$n_{2.} = N - n$
Сеть H_PPI	$n_{.1} = D_i$	$n_{.2} = N - 1 - D_i$	$n_{..} = N - 1$

*Здесь H_PPI \ L_PPI означает сеть PPI человека, из которой удалены белки, входящие в сеть L_PPI. Точка в индексе означает суммирование по этому индексу; d_i и D_i – число взаимодействий белка p_i в сети L_PPI и H_PPI соответственно.

2. Расчет отношения шансов (odds ratio – OR) для каждого белка p_i :

$$OR = \frac{d_i / (n - 1 - d_i)}{(D_i - d_i) / (N - D_i - n + d_i)} = \frac{d_i (N - D_i - n + d_i)}{(D_i - d_i) (n - 1 - d_i)} \quad (1)$$

Далее мы будем использовать логарифмическую форму статистики $\ln(OR)$, которая асимптотически сходится к нормальному распределению $N(0, \sigma)$ (Vsevolozhskaya, Zaykin, 2018), где

$$\sigma = \sigma(\ln(OR)) = \sqrt{\frac{1}{d_i} + \frac{1}{n - 1 - d_i} + \frac{1}{D_i - d_i} + \frac{1}{N - D_i - n + d_i}} \quad (2)$$

Нулевая гипотеза H_0 : $\ln(OR) = 0$ означает, что отношение числа белков, взаимодействующих с белком p_i , к числу не взаимодействующих с ним белков в сети L_PPI равно этому показателю в сети с белками, не входящими в сеть L_PPI, т.е. белок p_i взаимодействует с белками из сети L_PPI так же, как и с другими белками.

Альтернативная гипотеза H_1 : $\ln(OR) > 0$ означает, что белок p_i чаще взаимодействует с белками из сети L_PPI, чем с другими белками.

Для гена g_i в качестве индекса значимости I_i будем использовать нижнюю границу 95 % доверительного интервала для статистики $\ln(OR)$:

$$I_i = \ln(OR) - 1.96 \sigma \quad (3)$$

Значение $I_i > 0$ будет указывать на то, что гипотеза H_0 отвергается. В этом случае мы будем использовать значение I_i как значимость или вес гена g_i , полученный на основе сети белок-белковых взаимодействий:

$$\text{если } I_i > 0, \text{ то } w_i = I_i, \text{ иначе } w_i = 0. \quad (4)$$

Приоритизацию генов осуществляли на основе нескольких критериев, после чего по каждому критерию гену присваивали ранг. Затем ранги суммировали и определяли общий ранг. В работе использовали следующие критерии.

¹ Табл. S1–S15 см. в Приложении по адресу: <http://www.bionet.nsc.ru/vogis/download/pict-2018-22/appx18.pdf>

Таблица 1. Категории генов/белков КРС и других видов млекопитающих, которые были включены в каталог генов, вовлеченных в патогенез инфекции ВЛКРС

Категория (тип доказательства)	Описание	Число		Номер таблицы в Приложении
		генов	публикаций	
Полногеномный анализ ассоциаций	Гены, внутри или рядом с которыми находятся SNP, ассоциированные с патогенезом инфекции ВЛКРС по данным полногеномного анализа ассоциаций	53	3	S1
Аллельные варианты	Аллельные варианты SNP в этих генах были ассоциированы с чувствительностью/устойчивостью к инфекции ВЛКРС в исследованиях типа «случай–контроль»*	3	6	S2
Нокаут	Нокаут этого гена у мышей делает их более чувствительными к введению плазмиды, содержащей нуклеотидную последовательность ВЛКРС	1	1	S3
Экспрессия**	Гены, кодирующие мРНК (или белки), экспрессия которых повышается/снижается при инфицировании ВЛКРС	25	29	S4
ДНК-микрочип	Гены, кодирующие мРНК, экспрессия которых повышается/снижается у животных с персистентным лимфоцитозом, по данным гибридизации на ДНК-микрочипе	368	1	S5
Физическое взаимодействие	Гены, кодирующие белки, которые вступают в прямые физические взаимодействия с белками ВЛКРС	3	3	S6
Полный каталог	Каталог генов КРС и других видов млекопитающих, вовлеченных в патогенез инфекции ВЛКРС	446	43	S7

* Клиническая тяжесть заболевания или иммунологические параметры у инфицированных животных были ассоциированы с одним из аллелей SNP.

** Если имелось доказательство, что уровень активной формы белка изменялся в ответ на инфицирование ВЛКРС, то ген, кодирующий этот белок, также был включен в эту категорию.

Критерий 1. Ген ассоциирован с лейкозом по данным экспериментов GWAS. Ассоциированным генам присваивается ранг 1, остальным генам присваивается ранг 0.

Критерий 2. Ген ассоциирован с лейкозом по данным анализа «случай–контроль» для отдельного гена. Ассоциированным генам присваивается ранг 1, остальным генам – ранг 0.

Критерий 3. Роль гена в развитии лейкоза изучена в экспериментах по нокауту у мышей. Таким генам присваивается ранг 1, остальным генам – 0.

Критерий 4. Белок, кодируемый геном, участвует в белок-белковых взаимодействиях с вирусной частицей либо отдельным вирусным белком. Такие гены получают ранг 1, а остальным генам присваивается ранг 0.

Критерий 5. Ген имеет аннотацию терминами Gene Ontology, которые, по данным нашего анализа, являются перепредставленными для списка генов каталога ($FDR < 0.05$, $Fold\ Enrichment > 3$). Гены, аннотированные такими терминами, получают ранг 1, остальным генам присваивается ранг 0.

Критерий 6. Ген участвует в биологических путях из баз KEGG или REACTOME, которые, по данным нашего анализа, являются перепредставленными для списка генов каталога ($FDR < 0.05$, $Fold\ Enrichment > 3$). Такие гены получают ранг 1, а остальным генам присваивается ранг 0.

Критерий 7. Ранг каждого гена приравнивался индексу значимости (или веса) w_i , рассчитанному на основе описанной выше процедуры анализа сетей белок-белковых взаимодействий H_PPI и L_PPI .

Примененные в данной работе критерии выполняют две задачи. Первая задача состояла в том, чтобы отобрать гены, имеющие наиболее убедительные экспериментальные свидетельства их вовлеченности в ответ на

инфекцию ВЛКРС (критерии 1–4). Вторая задача – придать вес генам/белкам, принадлежащим к функционально связанному подмножеству, контролирующему патогенез заболевания, т. е. приоритизировать их.

Результаты

Создание каталога генов

Поиск по PubMed выявил 43 экспериментальные статьи, описывающие вовлечение генов/белков КРС и других видов млекопитающих в процесс инфицирования ВЛКРС. В результате ручной аннотации данного набора статей нами был создан каталог, включающий 446 генов КРС и других видов млекопитающих, вовлеченных в процесс патогенеза инфекции ВЛКРС (табл. 1).

Функциональная аннотация генов с помощью системы DAVID

С помощью системы DAVID были выявлены 200 перепредставленных терминов из словаря биологических процессов базы Gene Ontology, а также группы генов, аннотированные этими терминами (см. табл. 2 и табл. S8). Большая часть выявленных GO терминов характеризует процессы, связанные с ответом организма на вирус и иммунными реакциями. Объединенный список генов, связанных со всеми 200 перепредставленными GO терминами, включал 162 гена (табл. S9).

Кроме того, было выявлено, что гены из рассматриваемого списка неслучайно часто встречаются в семнадцати биологических путях из баз KEGG и REACTOME (15 и 2 соответственно) (табл. S10). Большинство выявленных биологических путей связано с развитием заболеваний, таких как воспалительное заболевание кишечника, болезнь

Шагаса, малярия, африканский трипаносомоз, туберкулез, корь, грипп А, ревматоидный артрит, лейшманиоз, стафилококковая инфекция, коклюш, герпес. Объединенный список генов, связанных со всеми семнадцатью перепредставленными биологическими путями, включал 81 ген (табл. S11).

Значимость генов в сети белок-белковых взаимодействий

На основе анализа сетей H_PPI и L_PPI нами выявлено, что 167 генов из каталога имеют положительные значения индекса значимости w_i (табл. S12).

Наибольшие значения w_i (от 1.20 до 1.77) имели гены *SLC7A7*, *TYROBP*, *CD79A*, *MS4A1*, *CTSS*, *CD3E*, *CIQA*, *TLR8*, *IL2RA*, *CD2*, *FGL2*, *IL6R*, *CD63*, *TNFRSF1B*. Белковые продукты большинства из них участвуют в регуляции иммунных процессов (табл. S13). Кроме того, корреляционный анализ значения характеристики гена w_i и индекса принадлежности к обучающей выборке показал наличие высокого уровня корреляции $R = 0.187$ с уровнем значимости $p\text{-value} < 0.001$. Эти наблюдения указывают на то, что предложенный нами метод расчета значимости генов на основе сети белок-белковых взаимодействий адекватно оценивает значимость гена в контексте целевого признака и может быть использован в качестве дополнительного критерия при решении задачи приоритизации генов.

Ранжирование генов

На основе описанной процедуры ранжирования (см. Материалы и методы) были определены ранги всех генов из сформированного нами каталога (446 генов). Наивысшие итоговые ранги (5.77 и 5.05) выявлены у генов *TNF* и *HLA-DPA1/BOLA-DQA1*. Еще три гена (*HLA-DRB1/BOLA-DRB3*, *ATF2*, *LTB*) набрали более 4 баллов (табл. 3). У двадцати генов (*CTSS*, *CD3E*, *LTA*, *CIQA*, *IL2RA*, *CD2*, *IL6R*, *TNFRSF1B*, *CTLA4*, *CSNK2B*, *IL10*, *LCP2*, *SELPLG*, *IL6*, *CIQC*, *CD19*, *IL1B*, *IL18*, *SYK*, *HGF*) суммарный балл принимал значение от 3 до 3.36. Полные сведения по ранжированию генов представлены в табл. S15.

Обсуждение

Для выявления приоритетных генов-кандидатов, связанных с развитием лейкоза, нами собран каталог, включающий 446 генов, выявленных на основе различных экспериментальных подходов (GWAS, исследование белок-белковых взаимодействий *in vitro* и т. д.). Оказалось,

Таблица 2. Некоторые перепредставленные (FDR < 0.05, Fold Enrichment > 10) термины* из словаря биологических процессов базы Gene Ontology, выявленные при анализе 446 генов из каталога системой DAVID

Термин GO	Число генов	Fold Enrichment	FDR
Interferon-alpha biosynthetic process	5	37.4	4.7E-03
Regulation of interferon-alpha biosynthetic process	5	37.4	4.7E-03
Interferon-gamma biosynthetic process	9	24.0	1.2E-06
Regulation of interferon-gamma biosynthetic process	7	21.8	5.3E-04
Regulation of tolerance induction	6	18.7	1.7E-02
Tolerance induction	6	16.0	4.1E-02
Positive regulation of activated T cell proliferation	8	13.6	2.1E-03
Activated T cell proliferation	10	12.1	1.4E-04
Regulation of activated T cell proliferation	9	11.6	1.2E-03
Positive regulation of interferon-gamma production	16	11.5	6.5E-09
Positive regulation of alpha-beta T cell activation	12	11.5	6.9E-06
Regulation of interferon-gamma production	22	11.0	4.2E-13
Interferon-gamma production	22	10.4	1.9E-12
Cytokine biosynthetic process	22	10.3	2.3E-12
Regulation of alpha-beta T cell activation	14	10.3	1.0E-06
B cell receptor signaling pathway	9	10.2	3.5E-03
Negative regulation of mononuclear cell proliferation	11	10.0	1.7E-04
Negative regulation of lymphocyte proliferation	11	10.0	1.7E-04

* Полный список перепредставленных терминов, выявленных при FDR < 0.05 и Fold Enrichment > 3, приведен в табл. S8.

Таблица 3. Ранги пяти наиболее значимых генов из каталога генов, вовлеченных в патогенез инфекции ВЛКРС, согласно критериям 1–7, и их итоговая оценка

Символ гена		Критерии							Суммарный балл
человека	быка	1	2	3	4	5	6	7	
<i>TNF</i>	<i>TNF</i>	1	1	1	0	1	1	0.77	5.77
<i>HLA-DPA1</i>	<i>BOLA-DQA1</i>	1	1	0	0	1	1	1.05	5.05
<i>HLA-DRB1</i>	<i>BOLA-DRB3</i>	1	1	0	0	1	1	0.45	4.45
<i>ATF2</i>	<i>ATF2</i>	0	0	0	1	1	1	1.11	4.11
<i>LTB</i>	<i>LTB</i>	1	0	0	0	1	1	1.03	4.03

Примечание. Подчеркнуты символы генов, принадлежащих контрольной выборке.

что группы генов, выявленных каждым из методов, слабо перекрываются между собой. Так, ни один ген из списков дифференциально экспрессируемых генов не был подтвержден другими экспериментальными методиками. Мы связываем это обстоятельство с тем, что дифференциальная экспрессия изучалась в клетках определенного типа (дендритные клетки, Т-хелперы и др.) и в определенный временной период (табл. S4 и S5), а ассоциативные исследования (табл. S1 и S2) позволяют выявить гены, влияющие на интегральные характеристики всего организма на всех этапах патологического процесса. Отсутствие пересечения между группами генов, полученными в экспериментах по белок-белковым взаимодействиям либо нокауту, с другими группами генов, по-видимому, связано с очень малым количеством генов, исследованных в этих типах экспериментов.

Для приоритизации нами использованы две группы качественно различных критериев. Первая группа (критерии 1–4) отражает наличие экспериментальных данных, свидетельствующих об ассоциации гена с заболеванием либо об известном механизме его участия в биологическом процессе. Совокупное количество генов, положительно оцененных хотя бы по одному из четырех критериев, составляет 12 % от их общего числа (54 гена из 446). При этом ни один ген не имел положительных баллов по всем четырем критериям, и только ген *TNF* имел положительную оценку по трем критериям из четырех.

Получив такой результат, мы приняли решение разработать добавочные критерии. При этом основывались на предположении, что некая часть генов/белков из каталога представляет собой функционально связанное подмножество, контролирующее течение заболевания (лейкоз). Принадлежность гена/белка к этому подмножеству может указывать на то, что ген/белок задействован в ответе на инфекцию. Таким образом, были предложены критерии 5–7, характеризующие функциональную связь каждого из генов/белков со всем множеством других генов. Критерии этой группы положительно оценивали 162 (GO анализ), 81 (Pathway анализ) и 167 (ранг в сети белок-белковых взаимодействий) генов, что составляло 36, 18 и 37 % от общего количества генов соответственно. В целом положительную оценку хотя бы по одному из указанных критериев получили 235 генов (53 %). Положительную оценку по каждому из трех критериев (критерии 5–7) получили 57 генов (13 %). Анализ оценок всего множества генов из каталога, выявленных на основе каждого из трех критериев, и индексов принадлежности генов к обучающей выборке показал наличие значимых корреляций (во всех случаях p -value < 0.001).

В результате ранжирования наибольшее количество баллов (от 4 до 5.77) получили пять генов (*TNF*, *BOLA-DQA1*, *BOLA-DRB3*, *ATF2*, *LTB*) (см. табл. 3). Четыре из них (*TNF*, *LTB*, *BOLA-DQA1*, *BOLA-DRB3*) оказались вовлечены в иммунный ответ. Гены *TNF* и *LTB* кодируют белки из семейства белков фактора некроза опухолей, в которое входят около 48 белков (Kim et al., 2005). Оба гена были идентифицированы как потенциальные гены-кандидаты при полногеномном анализе ассоциаций с инфицированием ВЛКРС (Carignano et al., 2018). Аллель G в промоторе гена *TNF*, на расстоянии 824 п. н. от старта транскрипции,

ассоциирован с развитием лимфомы у вирусоносителей лейкоза КРС, повышенным числом провирусов в геноме и низкой транскрипционной активностью промотора *TNF* при люциферазном анализе *in vitro* (Konnai et al., 2006; Lendez et al., 2015). Нокаут этого гена у мышей делает их более чувствительными к введению плазмиды, содержащей нуклеотидную последовательность ВЛКРС (Müller et al., 2003). Белок LTB синтезируется активированными Т- и В-лимфоцитами, естественными киллерами, образует гетеротример с лимфотоксином-альфа и, таким образом, «заякоривает» лимфотоксин-альфа на клеточной мембране лимфоцита (Nakamura et al., 1995). Такой гетеротример функционирует как лиганд для рецептора TNFRSF3/LTBR и участвует в развитии иммунного ответа, обеспечивая межклеточную коммуникацию (Crowe et al., 1994). Считается, что основная функция LTB заключается в стимулировании развития лимфоидной ткани, в первую очередь лимфатических узлов (Onder et al., 2013).

Гены *BOLA-DQA1* и *BOLA-DRB3* относятся к генам главного комплекса совместимости, локализованного на хромосоме 23. Показано, что различные аллели гена *BOLA-DRB3* ассоциированы с чувствительностью/резистентностью к инфицированию ВЛКРС (Lewin, 1994). По-видимому, резистентность к вирусу связана с присутствием полярного мотива в позициях 70–71 полипептида в области вероятного пептид-связывающего домена (Xu et al., 1993). Аллель DRB3.2*0902 достоверно ассоциирован с генетической устойчивостью к развитию персистентного лимфоцитоза и пониженным числом провирусов (Juliarena et al., 2008). Аллели *BoLA-DQA1*0204* и *BoLA-DQA1*10012* были ассоциированы с низкой и высокой провирусной нагрузкой соответственно (Miyasaka et al., 2013).

Пятый из числа наиболее значимых генов – *ATF2*. Он кодирует транскрипционный фактор activating transcription factor 2, известный также как CREB2 (Cyclic AMP-responsive element-binding protein 2). Белок ATF2 способен индуцировать экспрессию длинного терминального повтора ВЛКРС *in vitro* (Willems et al., 1992).

Группа, включавшая двадцать следующих по значимости генов (от 3 до 3.36 балла), содержала существенное количество генов, связанных с функционированием иммунной системы (табл. S15). Среди таких генов был *LTA*, кодирующий лимфотоксин-альфа, относящийся к семейству белков фактора некроза опухолей (Kim et al., 2005). Уровень белка LTA в крови является хорошим предиктором скорости развития СПИД у ВИЧ-инфицированных пациентов (Medrano et al., 1998). Шесть из двадцати генов, имевших оценку от 3 до 3.36 балла (*IL2RA*, *IL6R*, *IL10*, *IL6*, *IL1B*, *IL18*), кодируют интерлейкины и/или их рецепторы, и три гена (*CD3E*, *CD2*, *CD19*) кодируют поверхностные антигены, экспрессирующиеся на лимфоидных клетках.

Таким образом, нами создан каталог генов КРС и других видов млекопитающих, вовлеченных в процесс патогенеза инфекции ВЛКРС, включающий 446 генов. Приоритизация полученного списка генов методами биоинформатики позволила выделить пять наиболее вероятных генов-кандидатов, которые могут влиять на чувствительность/устойчивость КРС к заболеванию лейкозом. Экспертный анализ данных литературы выявил потенциальную воз-

возможность участия каждого из пяти высокоприоритетных генов в патогенезе заболевания, подтвердив тем самым эффективность выбранного нами метода приоритизации. Дальнейшие экспериментальные исследования этих генов позволят получить информацию, значимую для понимания механизмов патогенеза лейкоза у КРС, что крайне важно для выработки практических рекомендаций для ликвидации этой вирусной инфекции, в том числе с помощью геномной и маркер-ассоциированной селекции.

Благодарности

Исследование поддержано Комплексной программой фундаментальных исследований СО РАН II.1 (проект № 0324-2018-0027) и выполнено с использованием оборудования ЦКП «Центр генетических ресурсов лабораторных животных» ФИЦ ИЦиГ СО РАН, поддержанного Минобрнауки России (уникальный идентификатор проекта RFMEFI62117X0015).

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Список литературы / References

Abdalla E.A., Peñagaricano F., Byrem T.M., Weigel K.A., Rosa G.J. Genome-wide association mapping and pathway analysis of leukosis incidence in a US Holstein cattle population. *Anim. Genet.* 2016; 47(4):395-407. DOI 10.1111/age.12438.

Abdalla E.A., Rosa G.J., Weigel K.A., Byrem T. Genetic analysis of leukosis incidence in United States Holstein and Jersey populations. *J. Dairy Sci.* 2013;96(9):6022-6029. DOI 10.3168/jds.2013-6732.

Barez P.Y., de Brogniez A., Carpentier A., Gazon H., Gillet N., Gutiérrez G., Hamaidia M., Jacques J.R., Perike S., Neelature Sriramareddy S., Renotte N., Staumont B., Reichert M., Trono K., Willems L. Recent Advances in BLV Research. *Viruses.* 2015;7:6080-6088. DOI 10.3390/v7112929.

Bishop S.C. Disease Genetics: Successes, Challenges and Lessons Learnt. Proc. 10th World Congr. Genet. Appl. to Livest. Prod., 2014.

Brym P., Bojarojc-Nosowicz B., Olenski K., Hering D.M., Rusc A., Kaczmarczyk E., Kaminski S. Genome-wide association study for host response to bovine leukemia virus in Holstein cows. *Vet. Immunol. Immunopathol.* 2016;175:24-35. DOI 10.1016/j.vetimm.2016.04.012.

Carignano H.A., Roldan D.L., Beribe M.J., Raschia M.A., Amadio A., Nani J.P., Gutierrez G., Alvarez I., Trono K., Poli M.A., Miretti M.M. Genome-wide scan for common SNPs affecting bovine leukemia virus infection level in dairy cattle. *BMC Genomics.* 2018;19:142. DOI 10.1186/s12864-018-4523-2.

Coffin J.M., Hughes S.H., Varmus H.E. (Eds.). *Retroviruses.* N. Y.: Cold Spring Harbor Laboratory Press, 2002.

Crowe P.D., VanArsdale T.L., Walter B.N., Ware C.F., Hession C., Ehrensfels B., Browning J.L., Din W.S., Goodwin R.G., Smith C.A. A lymphotoxin-beta-specific receptor. *Science.* 1994;264(5159):707-710.

FAO. The State of the World's Animal Genetic Resources for Food and Agriculture. Commission on genetic resources for food and agriculture organization of the United Nations. Rome, 2007. Available at <http://www.fao.org/docrep/010/a1250e/a1250e00.htm>

Gutiérrez G., Lomonaco M., Alvarez I., Fernandez F., Trono K. Characterization of colostrum from dams of BLV endemic dairy herds. *Vet. Microbiol.* 2015;177:366-369. DOI 10.1016/j.vetmic.2015.03.001.

Huang D.W., Sherman B.T., Lempicki R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 2009;4:44-57. DOI 10.1038/nprot.2008.211.

Ignatieva E.V., Igoshin A.V., Yudin N.S. A database of human genes and a gene network involved in response to tick-borne encephalitis virus infection. *BMC Evol. Biol.* 2017;17(Suppl.2):259. DOI 10.1186/s12862-017-1107-8.

Juliarena M.A., Poli M., Sala L., Ceriani C., Gutierrez S., Dolcini G., Rodríguez E.M., Marino B., Rodríguez-Dubra C., Esteban E.N. Association of BLV infection profiles with alleles of the *BoLA-DRB3.2* gene. *Anim. Genet.* 2008;39:432-438. DOI 10.1111/j.1365-2052.2008.01750.x.

Kim J.Y., Moon S.M., Ryu H.J., Kim J.J., Kim H.T., Park C., Kimm K., Oh B., Lee J.K. Identification of regulatory polymorphisms in the TNF-TNF receptor superfamily. *Immunogenetics.* 2005;57(5):297-303.

Koeck A., Miglior F., Kelton D.F., Schenkel F.S. Health recording in Canadian Holsteins: data and genetic parameters. *J. Dairy Sci.* 2012; 95:4099-4108.

Konnai S., Usui T., Ikeda M., Kohara J., Hirata T., Okada K., Ohashi K., Onuma M. Tumor necrosis factor-alpha genetic polymorphism may contribute to progression of bovine leukemia virus-infection. *Microbes Infect.* 2006;8(8):2163-2171.

Kotlyar M., Pastrello C., Sheahan N., Jurisica I. Integrated interactions database: tissue-specific view of the human and model organism interactomes. *Nucleic Acids Res.* 2016;44(D1):D536-D541. DOI 10.1093/nar/gkv1115.

Lendez P.A., Passucci J.A., Poli M.A., Gutierrez S.E., Dolcini G.L., Ceriani M.C. Association of TNF- α gene promoter region polymorphisms in bovine leukemia virus (BLV)-infected cattle with different proviral loads. *Arch. Virol.* 2015;160(8):2001-2007. DOI 10.1007/s00705-015-2448-5.

Lewin H.A. Host genetic mechanism of resistance and susceptibility to a bovine retroviral infection. *Anim. Biotechnol.* 1994;5:183-191. DOI 10.1080/10495399409525820.

Medrano F.J., Leal M., Arienti D., Rey C., Zagliani A., Torres Y., Sanchez-Quijano A., Lissen E., Clerici M. Tumor necrosis factor beta and soluble APO-1/Fas independently predict progression to AIDS in HIV-seropositive patients. *AIDS Res. Hum. Retroviruses.* 1998;14:835-843.

Miyasaka T., Takeshima S.N., Jimba M., Matsumoto Y., Kobayashi N., Matsuhashi T., Sentsui H., Aida Y. Identification of bovine leukocyte antigen class II haplotypes associated with variations in bovine leukemia virus proviral load in Japanese Black cattle. *Tissue Antigens.* 2013;81:72-82. DOI 10.1111/tan.12041.

Müller C., Coffey T.J., Koss M., Teifke J.P., Langhans W., Werling D. Lack of TNF alpha supports persistence of a plasmid encoding the bovine leukaemia virus in TNF^{-/-} mice. *Vet. Immunol. Immunopathol.* 2003;92:15-22.

Nakamura T., Tashiro K., Nazarea M., Nakano T., Sasayama S., Honjo T. The murine lymphotoxin-beta receptor cDNA: isolation by the signal sequence trap and chromosomal mapping. *Genomics.* 1995; 30(2):312-319.

Onder L., Danuser R., Scandella E., Firmer S., Chai Q., Hehlhans T., Stein J.V., Ludewig B. Endothelial cell-specific lymphotoxin- β receptor signaling is critical for lymph node and high endothelial venule formation. *J. Exp. Med.* 2013;210(3):465-473. DOI 10.1084/jem.20121462.

Rodríguez S.M., Florins A., Gillet N., de Brogniez A., Sánchez-Alcaraz M.T., Boxus M., Boulanger F., Gutiérrez G., Trono K., Alvarez I., Vagnoni L., Willems L. Preventive and therapeutic strategies for bovine leukemia virus: lessons for HTLV. *Viruses.* 2011;3:1210-1248. DOI 10.3390/v3071210.

Stepanova T.V. Analysis of the economic damage caused by bovine leukemia from 2010 to 2014 in the Russian Federation. *Russ. J. Agric. Socio-Economic Sci.* 2016;8(56):49-56. DOI 10.18551/rjoas.2016-08.08.

Vsevolozhskaya O.A., Zaykin D.V. Put the odds on your side: a new measure for epidemiological associations. 2018. <https://arxiv.org/pdf/1806.04251.pdf>

Willems L., Kettmann R., Chen G., Portetelle D., Burny A., Derse D. A cyclic AMP-responsive DNA-binding protein (CREB2) is a cellular transactivator of the bovine leukemia virus long terminal repeat. *J. Virol.* 1992;66:766-772.

Xu A., van Eijk M.J., Park C., Lewin H.A. Polymorphism in *BoLA-DRB3* exon 2 correlates with resistance to persistent lymphocytosis caused by bovine leukemia virus. *J. Immunol.* 1993;151:6977-6985.