

© Филатова Е.Н., Сахарнов Н.А., Князев Д.И., Преснякова Н.Б., Уткин О.В., 2020

УДК 616.9:616-037:57.087.1

Использование программы MiDA для поиска молекулярно-генетических маркеров риска развития осложнений инфекционных заболеваний

Е.Н. Филатова, Н.А. Сахарнов, Д.И. Князев, Н.Б. Преснякова, О.В. Уткин

ФБУН «Нижегородский научно-исследовательский институт эпидемиологии и микробиологии им. академика И.Н. Блохиной» Роспотребнадзора,
ул. Малая Ямская, д. 71, г. Нижний Новгород, 603950, Российская Федерация

Резюме: *Введение.* Поиск специфических молекулярно-генетических маркеров риска развития осложнений инфекционных заболеваний является актуальным направлением современной медико-биологической науки. *Материалы и методы.* Для решения этой задачи нами разработана программа для ЭВМ MiDA. В программе реализован комплексный подход, позволяющий проводить отбор потенциальных маркеров на основании показателей изменения уровня экспрессии ряда генов в группах сравнения и значимости признака для классификации – отнесения образцов к анализируемым группам. *Результаты исследования.* С помощью программы MiDA проведен поиск молекулярно-генетических маркеров риска развития тяжелой формы течения лихорадки денге и хронической формы бруцеллеза. В результате исследования в качестве маркера риска осложнения лихорадки денге предложен ген *HSPA6*, экспрессия которого в периферической крови пациентов с тяжелой формой течения болезни снижалась. К маркерам хронической формы течения бруцеллеза отнесли снижение экспрессии микроРНК *hsa-miR-198* и *hsa-miR-501-3p*, а также повышение экспрессии микроРНК *hsa-miR-618* в $CD4^+$ Т-лимфоцитах. *Выводы.* Продемонстрирована возможность применения программы MiDA для анализа больших данных, полученных с помощью современных технологий (секвенирование, биочипы и др.). Возможно расширение сферы применения программы для анализа экспрессии генов, транскриптов и белков при заболеваниях различного генеза, определения молекулярных механизмов реализации патологического процесса, поиска диагностических и прогностических маркеров заболевания, а также потенциальных мишеней для разработки средств таргетной терапии.

Ключевые слова: анализ данных, биочип, большие данные, бруцеллез, лихорадка денге, машинное обучение, программа для ЭВМ, риск осложнений, R.

Для цитирования: Филатова Е.Н., Сахарнов Н.А., Князев Д.И., Преснякова Н.Б., Уткин О.В. Использование программы MiDA для поиска молекулярно-генетических маркеров риска развития осложнений инфекционных заболеваний // Здоровье населения и среда обитания. 2020. № 1 (322). С. 51–56. DOI: <http://doi.org/10.35627/2219-5238/2020-322-1-51-56>

The Use of the MiDA Software for Establishing Molecular Genetic Markers of the Risk of Developing Infectious Disease Complications

Е.Н. Filatova, N.A. Sakharov, D.I. Knyazev, N.B. Presnyakova, O.V. Utkin

Academician I.N. Blokhina Nizhny Novgorod Scientific Research Institute of Epidemiology and Microbiology,
71 Malaya Yamskaya Street, Nizhny Novgorod, 603950, Russian Federation

Abstract: *Introduction.* The search for specific molecular and genetic markers of the risk of developing infectious disease complications is a current area of research in modern medical and biological science. *Materials and methods.* In order to solve this issue, we developed a MiDA software that implements an integrated approach allowing for selection of potential markers on the basis of indicators of expression fold change of a number of genes in the comparison groups and the feature importance for classification, i.e. the assignment of samples to the analyzed groups. *Results.* Using the MiDA software, we searched for molecular and genetic markers of the risk of developing severe dengue fever and chronic brucellosis. As a result of the study, the *HSPA6* gene was proposed as a risk marker for the dengue complication. *HSPA6* expression was reduced in the peripheral blood samples of severe dengue cases. Markers of chronic brucellosis included a decrease in the expression of miRNA *hsa-miR-198* and *hsa-miR-501-3p*, as well as an increase in the expression of miRNA *hsa-miR-618* in $CD4^+$ T-lymphocytes. *Conclusion.* We demonstrated the possibility of applying the MiDA software to the analysis of big data obtained using modern techniques (sequencing, biochips, etc.). It is possible to expand the scope of the software application in order to analyze the expression of genes, transcripts and proteins in diseases of various origins, to determine molecular mechanisms of the pathological process, to search for diagnostic and prognostic markers of the disease, as well as potential targets for the development of specific therapies.

Key words: data analysis, biochip, big data, brucellosis, dengue fever, machine learning, software, risk of complications, R.

For citation: Filatova EN, Sakharov NA, Knyazev DI, Presnyakova NB, Utkin OV. The Use of the MiDA Software for Establishing Molecular Genetic Markers of the Risk of Developing Infectious Disease Complications. *Zdorov'e naseleniya i sreda obitaniya*. 2020; 1(322): 51–56. (In Russian) DOI: <http://doi.org/10.35627/2219-5238/2020-322-1-51-56>

Information about the authors: Filatova E.N., <http://orcid.org/0000-0002-6683-7191>; Sakharov N.A., <http://orcid.org/0000-0003-3965-2033>; Knyazev D.I., <http://orcid.org/0000-0002-6880-7105>; Presnyakova N.B., <http://orcid.org/0000-0001-5257-6210>; Utkin O.V., <http://orcid.org/0000-0002-7571-525X>.

Введение. Борьба с инфекционными заболеваниями уже давно перешла из сферы общественного здравоохранения в область национальной безопасности. Рост инфекционной заболеваемости среди населения неизменно приводит к демографическим, социальным и экономическим потерям. При этом социально-экономический ущерб во многом обусловлен тяжестью течения инфекционного процесса, возможностью его хронизации и развития осложнений.

На сегодняшний день тяжесть течения инфекции, а также факт ее хронизации определяют по степени выраженности клинических симптомов, длительности их сохранения, возникновению осложнений и сопутствующих заболеваний. Специфические тесты, позволяющие не только констатировать неблагоприятное течение болезни, но и превентивно оценивать риски его возникновения, отсутствуют. Известно, что характер течения инфекционного процесса определяется как особенностями возбудителя

(например, вирулентностью штамма), так и индивидуальными особенностями организма-хозяина, в частности, состоянием его иммунной системы [1]. Таким образом, для оценки риска возникновения тяжелых форм течения инфекционных заболеваний возможно использование специфических молекулярно-генетических маркеров, характеризующих функциональное состояние иммунной системы пациента.

Поисковые работы, направленные на выявление молекулярно-генетических маркеров, подразумевают анализ геномов и/или транскриптомов. Данные об экспрессии тысяч генов и транскриптов получают с помощью высокопроизводительных технологий, таких как полногеномное секвенирование, биочипы и др. На втором этапе исследования осуществляют анализ полученных данных, для чего привлекают сложные математические и статистические инструменты. В ходе этого этапа из всей массы исследованных генов и/или транскриптов отбирают маркерные признаки. Наиболее часто применяют методы традиционной статистики (Т-тест и его модификации) с определением уровня изменения экспрессии генов/транскриптов [2, 3]. Другим подходом является построение модели классификации с применением методов машинного обучения. В этом случае отбор признаков осуществляют на основании значимости гена/транскрипта для классификации [4, 5]. Алгоритмы отбора признаков, комбинирующие оба подхода, позволяют сочетать универсальность, гибкость и высокую обобщающую способность методов машинного обучения с удовлетворяющими требованиям доказательной медицины критериями традиционной статистики [6], что обеспечивает высокую имплементацию полученных результатов в медицинскую практику.

Для реализации комбинированного подхода при поиске молекулярно-генетических маркеров среди множества генов/транскриптов нами разработана программа для ЭВМ MiDA (Microarray Data Analysis). Программа выполнена в виде пакета для языка программирования R и размещена в официальном репозитории The Comprehensive R Archive Network (<https://CRAN.R-project.org/package=MiDA>). В программе реализован алгоритм отбора признаков, основанный на применении Т-теста с поправкой на ожидаемую долю ложных отклонений в комбинации с классификацией методом градиентного бустинга над решающими деревьями. Отбор маркерных генов/транскриптов осуществляется на основании показателей уровня изменения их экспрессии (при условии статистической значимости $q < 0,05$) и значимости для классификации.

Цель исследования — оценка возможности применения программы MiDA для поиска молекулярно-генетических маркеров риска развития осложнений инфекционных заболеваний.

Материалы и методы. *Данные.* Для тестирования возможностей программы MiDA в поиске молекулярно-генетических маркеров риска развития осложнений мы использовали данные о состоянии транскриптома при двух инфекционных заболеваниях, бруцеллезе и лихорадке денге, полученные из репозитория GEO на портале NCBI.

Для поиска маркеров тяжелой формы течения лихорадки денге анализировали массив данных, полученных из образцов цельной периферической крови пациентов с нетяжелой формой лихорадки денге (ЛД, $n = 18$), тяжелой формой лихорадки денге (ТЛД, $n = 10$) и здоровых доноров (НОРМ-Д, $n = 9$) (серия образцов GSE51808, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse51808>). Транскриптомные данные были получены при помощи микрочипа платформы Affymetrix HT HG-U133+ PM Array Plate (Thermo Fisher Scientific, США), содержащего 54613 зондов к мРНК различных генов человека.

Для поиска маркеров хронической формы бруцеллеза использовали массив данных о профиле экспрессии некодирующих микроРНК в образцах CD4⁺ Т-лимфоцитов периферической крови пациентов с различными формами бруцеллеза (серия образцов GSE107554, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE107554>). Образцы получены от пациентов с острым бруцеллезом (ОБ, $n = 7$), хроническим бруцеллезом (ХБ, $n = 8$), а также здоровых доноров (НОРМ-Б, $n = 8$). Экспрессия микроРНК изучалась с помощью микрочипа платформы Agilent-046064 Unrestricted Human miRNA V19.0 Microarray (Agilent Technologies, США), содержащего 4440 зондов к различным микроРНК человека.

Обработка данных. Поскольку предварительная обработка результатов была осуществлена авторами депонированных данных, процедуры нормализации и вычитания фона нами не проводились. Для каждой исследуемой инфекции нами сформированы парные группы сравнения: ЛД — НОРМ-Д, ТЛД — НОРМ-Д, ТЛД — ЛД и ОБ — НОРМ-Б, ХБ — НОРМ-Б, ХБ — ОБ. С применением программы MiDA для каждой пары групп проводился анализ изменения уровня экспрессии признаков (генов/транскриптов) (функция *MiStatCount*), который рассчитывали как [(среднее группы 1 · 100 / среднее группы 2) – 100] (%). В дальнейшем осуществляли оценку значимости признаков для классификации с применением кросс-валидации (функция *MiBiClassGBODT*), модель классификации считали удовлетворительной при среднем показателе площади под кривой ошибок (AUC) более 0,75. Подбор оптимальных параметров классификации производили с использованием служебных функций пакета (*MiNTreesAdjust*, *MiShrinkAdjust*, *MiIntDepthAdjust*), позволяющих выбрать оптимальное значение коэффициентов для получения максимальной акkuratности классификации тестовых образцов. Расчет усредненного значения значимости признаков производили с применением функции *MiInflCount*.

Для каждой анализируемой пары групп сравнения определяли кандидатные маркеры — признаки, характеризующиеся наивысшими показателями уровня изменения экспрессии и значимости для классификации среди всех анализируемых генов/транскриптов. Использовали функцию *MiSelectSignif*, при этом пороговый уровень статистической значимости устанавливали равным 0,05, а пороговые уровни для показателей уровня изменения и значимости определяли эмпирически, исходя из

распределения указанных параметров. В случае невозможности построить удовлетворительную модель классификации для групп сравнения показатель значимости не учитывался.

На следующем этапе из полученных для каждой пары групп сравнения перечней кандидатных маркеров выбирали гены/транскрипты, позволяющие дифференцировать различные формы течения заболевания, – специфические маркеры. В качестве специфических маркеров тяжелой формы лихорадки денге выделяли гены/транскрипты, входящие в перечень кандидатных маркеров групп сравнения ТЛД – НОРМ-Д и ТЛД – ЛД, но отсутствующие в перечне кандидатных маркеров групп сравнения ЛД – НОРМ-Д. Аналогично в качестве специфических маркеров хронической формы бруцеллеза отбирали гены/транскрипты, входящие в перечень кандидатных маркеров групп сравнения ХБ – НОРМ-Б и ХБ – ОБ, но отсутствующие в перечне кандидатных маркеров групп сравнения ОБ – НОРМ-Б.

Результаты исследования. При поиске молекулярно-генетических маркеров риска развития тяжелой формы лихорадки денге были построены модели классификации для групп сравнения ЛД – НОРМ-Д ($AUC = 0,90 \pm 0,14$) и ТЛД – НОРМ-Д ($AUC = 0,95 \pm 0,11$). Для групп сравнения ТЛД – ЛД удовлетворитель-

ного классификатора построить не удалось ($AUC = 0,57 \pm 0,14$).

В результате анализа данных нами сформированы 3 списка кандидатных генов/транскриптов, экспрессия которых значимо различалась в группах сравнения. Кандидатными маркерами групп сравнения ЛД – НОРМ-Д стали мРНК генов: *BIRC5*, *BUB1*, *CDC6*, *CDCA2*, *CDYL2*, *CEP55*, *E2F7*, *KCTD14*, *MAPK1*, *MCM10*, *NCAPG*, *PBK*, *RFC2*, *SPC25*, *TTK*. Кандидатными маркерами групп сравнения ТЛД – НОРМ-Д явились мРНК генов: *ALDH1L2*, *ALG10*, *ANLN*, *BIRC5*, *BUB1*, *CAV1*, *CDC6*, *CDCA2*, *CEP55*, *CYTOR*, *DDR1*, *E2F7*, *HSPA6*, *MCM10*, *NCAPG*, *PBK*, *SPC25*, *ST7L*. Из вышеперечисленных генов в перечень кандидатных маркеров для групп сравнения ТЛД – ЛД вошли только мРНК генов *DDR1*, *CYTOR* и *HSPA6*. Обобщенные сведения о характере экспрессии мРНК выделенных генов в образцах периферической крови пациентов с лихорадкой денге представлены в табл. 1.

Согласно полученным данным, развитие лихорадки денге сопровождалось выраженными изменениями экспрессии многих генов вне зависимости от формы течения заболевания. В качестве специфического маркера риска возникновения тяжелой формы течения лихорадки денге можно выделить мРНК гена *HSPA6*. Экспрессия *HSPA6* в периферической крови

Таблица 1. Изменение экспрессии мРНК некоторых генов в периферической крови пациентов с лихорадкой денге и практически здоровых доноров

Table 1. Changes in the expression of mRNA of some genes in the peripheral blood of dengue patients and practically healthy donors

Ген/ Gene	Изменение уровня экспрессии (%), уровень статистической значимости q (значимость для классификации – при наличии) / Expression change (%), statistical significance q (significance for classification, if applicable)		
	Группа сравнения / Comparison group		
	ЛД – НОРМ-Д / Dengue cases vs Healthy donors	ТЛД – НОРМ-Д / Severe dengue cases vs Healthy donors	ТЛД – ЛД / Severe dengue cases vs Dengue cases
<i>ALG10</i>	38,11, q < 0,001 (–)	–47,16, q < 0,001 (5,57)	6,57, q = 0,148 (–)
<i>ALDH1L2</i>	76,96, q < 0,001 (–)	108,63, q < 0,001 (0,00)	17,86, q = 0,136 (–)
<i>ANLN</i>	41,06, q < 0,001 (–)	41,16, q < 0,001 (0,00)	–1,01, q = 0,800 (–)
<i>BIRC5</i>	110,86, q < 0,001 (0,00)	117,93, q < 0,001 (0,00)	2,00, q = 0,748 (–)
<i>BUB1</i>	92,67, q < 0,001 (0,00)	97,92, q < 0,001 (0,00)	2,52, q = 0,692 (–)
<i>CAV1</i>	70,00, q < 0,001 (–)	106,30, q < 0,001 (0,00)	19,07, q = 0,058 (–)
<i>CDC6</i>	112,54, q < 0,001 (0,00)	114,36, q < 0,001 (0,00)	0,90, q = 0,905 (–)
<i>CDCA2</i>	106,28, q < 0,001 (0,00)	112,71, q < 0,001 (0,00)	3,10, q = 0,687 (–)
<i>CDYL2</i>	11,48, q = 0,002 (9,03)	14,08, q < 0,001 (–)	–1,47, q = 0,600 (–)
<i>CEP55</i>	111,26, q < 0,001 (0,00)	109,26, q < 0,001 (0,00)	–0,97, q = 0,874 (–)
<i>CYTOR</i>	8,45, q < 0,001 (–)	16,25, q < 0,001 (0,00)	7,30, q = 0,008 (–)
<i>DDR1</i>	–13,99, q < 0,001 (–)	–19,69, q < 0,001 (54,43)	–6,57, q = 0,047 (–)
<i>E2F7</i>	104,96, q < 0,001 (0,00)	111,49, q < 0,001 (0,00)	3,19, q = 0,711 (–)
<i>HSPA6</i>	–0,49, q = 0,145 (–)	–14,33, q = 0,009 (40,00)	–6,65, q = 0,035 (–)
<i>KCTD14</i>	110,87, q < 0,001 (0,00)	94,52, q < 0,001 (–)	–7,76, q = 0,116 (–)
<i>MAPK1</i>	–7,57, q < 0,001 (72,09)	–19,31, q < 0,001 (–)	–3,04, q = 0,074 (–)
<i>MCM10</i>	128,13, q < 0,001 (0,00)	127,71, q < 0,001 (0,00)	–0,28, q = 0,972 (–)
<i>NCAPG</i>	97,41, q < 0,001 (0,00)	100,57, q < 0,001 (0,00)	1,59, q = 0,801 (–)
<i>PBK</i>	107,48, q < 0,001 (0,00)	114,59, q < 0,001 (0,00)	3,20, q = 0,662 (–)
<i>RFC2</i>	12,11, q < 0,001 (18,68)	11,77, q < 0,001 (–)	–2,45, q = 0,336 (–)
<i>SPC25</i>	98,98, q < 0,001 (0,00)	103,04, q < 0,001 (0,00)	2,16, q = 0,790 (–)
<i>ST7L</i>	26,26, q < 0,001 (–)	24,39, q < 0,001 (0,00)	0,84, q = 0,660 (–)
<i>TTK</i>	92,53, q < 0,001 (0,00)	96,47, q < 0,001 (–)	2,19, q = 0,768 (–)

Примечание. (–) – классификацию не проводили (группы сравнения ТЛД – ЛД) или параметр не вошел в перечень кандидатных маркеров (группы сравнения ЛД – НОРМ-Д и ТЛД – НОРМ-Д).

Note. (–) – not classified (comparison groups: Severe dengue cases vs Dengue fever cases) or the parameter was not included in the list of candidate markers (comparison groups: Dengue fever cases vs Healthy donors and Severe dengue cases vs Healthy donors).

при тяжелой форме лихорадки значительно снижалась, а при неосложненной форме течения заболевания оставалась в пределах значений здоровых доноров.

При поиске молекулярно-генетических маркеров риска хронизации бруцеллезной инфекции не удалось построить удовлетворительные модели классификации для групп сравнения ОБ – НОРМ-Б ($AUC = 0,68 \pm 0,24$) и ХБ – НОРМ-Б ($AUC = 0,69 \pm 0,13$). Для групп сравнения ХБ – ОБ классификация была проведена с удовлетворительным результатом ($AUC = 0,78 \pm 0,19$).

Кандидатными маркерами групп сравнения ОБ – НОРМ-Б стали микроРНК: hsa-miR-18b-3p, hsa-miR-92b-5p, hsa-miR-422a, hsa-miR-587, hsa-miR-887, hsa-miR-941, hsa-miR-3127-3p, hsa-miR-3611, hsa-miR-4449, hsa-miR-4520a-3p, hsa-miR-4522. Кандидатными маркерами групп сравнения ХБ – НОРМ-Б явились микроРНК: hsa-miR-138-2-3p, hsa-miR-198, hsa-miR-302d-5p, hsa-miR-501-3p, hsa-miR-587, hsa-miR-599, hsa-miR-618, hsa-miR-661, hsa-miR-887, hsa-miR-3124-5p, hsa-miR-4449, hsa-miR-5011-5p. Из вышеперечисленных микроРНК в перечень кандидатных маркеров для групп сравнения ХБ – ОБ вошли hsa-miR-18b-3p, hsa-miR-198, hsa-miR-501-3p, hsa-miR-618, hsa-miR-941 и hsa-miR-4522.

Обобщенные сведения о характере экспрессии выделенных микроРНК представлены в табл. 2.

Согласно полученным результатам, развитие бруцеллеза вне зависимости от формы

течения заболевания сопровождалось усилением экспрессии микроРНК hsa-miR-587, hsa-miR-887, hsa-miR-4449 и hsa-miR-5011-5p и снижением экспрессии hsa-miR-3124-5p в CD4⁺ Т-лимфоцитах периферической крови. При этом острая форма бруцеллеза дополнительно характеризовалась снижением экспрессии hsa-miR-941 и hsa-miR-4522, а также усилением экспрессии hsa-miR-18b-3p по сравнению со здоровыми донорами. В качестве специфических маркеров риска возникновения хронической формы течения бруцеллеза предложены микроРНК hsa-miR-198, hsa-miR-501-3p и hsa-miR-618. Экспрессия hsa-miR-198 и hsa-miR-501-3p при хронической форме заболевания снижалась, а экспрессия hsa-miR-618 повышалась. Важно отметить, что при отсутствии признаков хронизации инфекции данные показатели оставались в пределах значений здоровых доноров.

Применение современных высокопроизводительных технологий при поисковых работах, направленных на выявление маркерных молекул того или иного патологического процесса, приводит к необходимости анализа экспрессии большого количества генов/транскриптов. При увеличении числа анализируемых признаков неизменно возрастает количество молекул, потенциально способных выступить в качестве маркеров. Таким образом, на современном этапе поисковые работы заключаются не столько в обнаружении дифференциально экспрессирующихся генов/транскриптов в ходе развития

Таблица 2. Изменение экспрессии некоторых микроРНК в CD4⁺ Т-лимфоцитах периферической крови пациентов с бруцеллезом и здоровых доноров

Table 2. Changes in the expression of some miRNAs in CD4⁺ T lymphocytes of peripheral blood of patients with brucellosis and healthy donors

МикроРНК / MicroRNA	Изменение уровня экспрессии (%), уровень статистической значимости q (значимость для классификации – при наличии) / Expression change (%), statistical significance q (significance for classification, if applicable)		
	Группа сравнения / Comparison group		
	ОБ – НОРМ-Б / Acute brucellosis cases vs Healthy donors	ХБ – НОРМ-Б / Chronic brucellosis cases vs Healthy donors	ХБ – ОБ / Chronic vs Acute brucellosis cases
hsa-miR-18b-3p	15,88, q = 0,049 (–)	4,84, q = 0,064 (–)	–5,55, q = 0,049 (0,00)
hsa-miR-92b-5p	–29,09, q = 0,047 (–)	–21,70, q = 0,096 (–)	10,47, q = 0,679 (–)
hsa-miR-138-2-3p	4,94, q = 0,146 (–)	–14,06, q = 0,049 (–)	–0,94, q = 0,779 (–)
hsa-miR-198	–4,12, q = 0,167 (–)	–6,82, q = 0,037 (–)	–8,34, q = 0,027 (6,40)
hsa-miR-302d-5p	32,63, q = 0,093 (–)	–42,55, q = 0,039 (–)	–7,13, q = 0,138 (–)
hsa-miR-422a	55,98, q = 0,047 (–)	1,24, q = 0,917 (–)	–5,62, q = 0,117 (–)
hsa-miR-501-3p	–0,35, q = 0,930 (–)	–13,52, q = 0,043 (–)	–15,22, q = 0,049 (0,00)
hsa-miR-587	13,05, q = 0,037 (–)	10,81, q = 0,045 (–)	18,05, q = 0,385 (–)
hsa-miR-599	12,89, q = 0,082 (–)	13,64, q = 0,049, (–)	0,74, q = 0,659 (–)
hsa-miR-618	–13,87, q = 0,185 (–)	15,00, q = 0,032 (–)	9,27, q = 0,049 (8,98)
hsa-miR-661	6,41, q = 0,248 (–)	21,04, q = 0,047 (–)	–3,94, q = 0,393 (–)
hsa-miR-887	15,80, q = 0,037 (–)	14,29, q = 0,032 (–)	–6,45, q = 0,051 (–)
hsa-miR-941	–43,86, q = 0,039 (–)	–1,17, q = 0,872 (–)	43,21, q = 0,049 (0,00)
hsa-miR-3124-5p	–9,47, q = 0,037 (–)	–14,50, q = 0,038 (–)	–5,52, q = 0,362 (–)
hsa-miR-3127-3p	13,02, q = 0,048 (–)	–4,89, q = 0,274 (–)	–2,94, q = 0,368 (–)
hsa-miR-3611	–15,36, q = 0,043 (–)	–11,89, q = 0,238 (–)	4,09, q = 0,764 (–)
hsa-miR-4449	14,22, q = 0,037 (–)	12,07, q = 0,039 (–)	–2,26, q = 0,447 (–)
hsa-miR-4520a-3p	–26,37, q = 0,046 (–)	–13,06, q = 0,184 (–)	18,05, q = 0,385 (–)
hsa-miR-4522	–14,74, q = 0,037 (–)	–3,26, q = 0,333 (–)	15,22, q = 0,049 (0,00)
hsa-miR-5011-5p	6,39, q = 0,043 (–)	28,40, q = 0,049 (–)	1,46, q = 0,576 (–)

Примечание. (–) – классификацию не проводили (группы сравнения ОБ – НОРМ-Б и ХБ – НОРМ-Б) или параметр не вошел в перечень кандидатных маркеров (группы сравнения ХБ – ОБ).

Note. (–) – not classified (comparison groups: Acute brucellosis cases vs Healthy donors and Chronic brucellosis cases vs Healthy donors) or the parameter was not included in the list of candidate markers (comparison groups: Chronic vs Acute brucellosis cases).

болезни, сколько в эффективном отборе наиболее подходящих кандидатных молекул из доступного множества. С точки зрения математики и статистики, процедура отбора признаков сводится к установлению критериев, которым выделяемый признак должен соответствовать.

В использованных нами алгоритмах такими критериями являются пороговое значение показателя изменения уровня экспрессии гена/транскрипта в сравниваемых группах (при обязательном условии статистической значимости изменения) и пороговое значение показателя значимости гена/транскрипта для классификации. В научной литературе используют различные значения порога изменения уровня экспрессии (от 1,3- до 4,0-кратного), равно как и различные значения порога уровня статистической значимости (от 0,2 до 0,01 после внесения поправки на множественные сравнения) [3, 7, 8]. Слишком завышенное пороговое значение изменения уровня экспрессии может привести к отсеиванию биологически значимых маркеров, а слишком заниженный — к включению незначимых молекул. В конечном счете наиболее целесообразным представляется определение порога изменения уровня экспрессии исходя из распределения анализируемых значений, поскольку на численное соотношение показателей экспрессии генов/транскриптов в двух сравниваемых группах оказывает влияние множество факторов. Некоторые факторы являются объективными, например, изначальный уровень экспрессии генов в ткани. Другие субъективны и зависят от используемой для детекции экспрессии генов/транскриптов технологии, а также методов обработки и анализа, например, способа нормализации данных.

Второй критерий — пороговое значение показателя значимости гена/транскрипта для классификации — позволяет оценить вклад каждой анализируемой молекулы в отнесение образца к одной из двух групп. Предполагается, что гены/транскрипты, обладающие наибольшим весом в модели классификации, также являются биологически значимыми маркерами. Однако зачастую для построения качественной модели классификации алгоритм машинного обучения использует лишь два-три показателя, которым присваивается максимальная значимость. Тысячи других молекул при этом получают нулевую значимость и незаслуженно отсеиваются [9]. Кроме того, высокая значимость гена/транскрипта для классификации не гарантирует наличие статистической значимости изменения его экспрессии при развитии патологии, что затрудняет имплементацию полученных результатов в медицинскую практику.

В разработанной нами программе MiDA реализуется комплексный подход, позволяющий производить отбор признаков с учетом обоих критериев. Так, введение в анализ критерия значимости гена/транскрипта для классификации позволяет включать в перечень маркеров молекулы со сравнительно низким изменением уровня экспрессии (например, маркер развития тяжелой формы лихорадки денге HSPA6). В то же время в качестве потенциальных маркеров включаются и гены/транскрипты, не обладающие значимостью для классифика-

ции, но характеризующиеся сравнительно высоким уровнем изменения экспрессии в группах сравнения (например, все маркеры хронического бруцеллеза). Полученные нами результаты демонстрируют высокую гибкость и универсальность использованного подхода.

Последним этапом анализа выявленных маркеров является определение целесообразности их применения, что обусловлено биологической функцией молекул в контексте анализируемой патологии. Так, тяжелая (геморрагическая) форма лихорадки денге характеризуется экссудацией плазмы крови вследствие повышения проницаемости сосудов. Выявленный нами маркер риска развития тяжелой формы заболевания — белок теплового стресса HSPA6 (Hsp70B) — участвует в фолдинге множества протеинов, обеспечивает выживание клеток при тепловом стрессе и воздействии ингибиторов протеасом, играет важную роль в активации макрофагов. На модели культуры клеток первичного эндотелия артерий легкого продемонстрирована способность маркерного белка понижать проницаемость эндотелиального монослоя [10]. Таким образом, снижение экспрессии HSPA6 в крови пациентов с лихорадкой денге может рассматриваться как потенциальный маркер риска возникновения осложнений заболевания в виде тяжелой формы.

Хроническая форма течения бруцеллеза характеризуется развитием реактивно-аллергических изменений и нарушением состава иммунокомпетентных клеток, в частности, снижением содержания CD4⁺ Т-лимфоцитов [11]. Известно, что выявленные нами маркеры ХБ — микроРНК hsa-miR-198 и hsa-miR-618 — регулируют дифференцировку и рост клеток за счет модуляции элементов PI3K/АКТ/mTOR-сигнального пути. На моделях опухолевых линий продемонстрировано, что снижение экспрессии hsa-miR-198 и усиление экспрессии hsa-miR-618 ингибирует пролиферацию и рост клеток [12, 13]. Другой маркер ХБ, hsa-miR-501-3p, подавляет экспрессию протоонкогена MYCN [14], что приводит к гиперэкспрессии хемокинов, а также повышению активации NKT-клеток, Т-лимфоцитов и моноцитов [12]. Подавление экспрессии hsa-miR-501-3p при ХБ индуцирует экспрессию MYCN, снижая напряженность иммунитета. Таким образом, выявленные нами микроРНК могут рассматриваться в качестве маркеров риска возникновения осложнений и хронизации бруцеллезной инфекции.

Выводы. С помощью программы MiDA проведен поиск молекулярно-генетических маркеров риска возникновения осложнений двух инфекционных заболеваний — лихорадки денге и бруцеллеза. Выявлено снижение экспрессии мРНК гена белка теплового стресса HSPA6 в периферической крови у пациентов с тяжелой формой течения лихорадки денге. Развитие хронической формы бруцеллеза сопровождалось изменением экспрессии регуляторов пролиферации и активации клеток: снижением экспрессии микроРНК hsa-miR-198 и hsa-miR-501-3p и повышением экспрессии микроРНК hsa-miR-618 в CD4⁺ Т-лимфоцитах. Выявленные нами молекулы могут рассматриваться в качестве маркеров риска развития осложненных

форм течения данных модельных заболеваний. Реализованный в программе MiDA подход, основанный на комбинации методов отбора признаков, дал возможность продемонстрировать ее пригодность для решения поставленных задач. Сфера применения программы может быть расширена с позиций анализа экспрессии генов/транскриптов и белков при заболеваниях различного генеза, определения молекулярных механизмов патоморфоза, поиска диагностических и прогностических маркеров заболевания, а также мишеней для разработки средств таргетной терапии.

Список литературы (пп. 2–10, 12–15 см. References)

1. Железникова Г.Ф. Инфекция и иммунитет: стратегии обеих сторон. Медицинская Иммунология. 2006. Т. 8. № 5-6. С. 597-614.
11. Логвиненко О.В., Ракитина Е.Л., Пономаренко Д.Г. и др. Особенности иммунологических показателей крови у больных различными формами бруцеллеза. Инфекция и иммунитет. 2013. Т. 3. № 3. С. 275-278.

References

1. Zheleznikova GF Infection and immunity: strategies from both sides. *Med Immunol* 2006; 8(5-6):597-614. (In Russian). DOI:10.15789/1563-0625-2006-5-6-597-614
2. Cao Z, Wang Y, Sun Y, et al. Effective and stable feature selection method based on filter for gene signature identification in paired microarray data [abstract]. 2013 IEEE International Conference on Bioinformatics and Biomedicine. 2013. P. 189-192.
3. McCarthy DJ, Smyth GK. Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics*. 2009; 25(6):765-771. DOI:10.1093/bioinformatics/btp053
4. Adewale AJ, Dinu I, Yasui Y. Boosting for correlated binary classification. *J Comput Graph Stat*. 2010; 19(1):140-153.
5. Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *Ann Statist*. 2000; 28(2):337-407.
6. Mieth B, Kloft M, Rodríguez JA, et al. Combining multiple hypothesis testing with machine learning increases the statistical power of genome-wide association studies. *Sci Rep*. 2016; 6:36671. DOI:10.1038/srep36671
7. Huggins CE, Domenighetti AA, Ritchie ME, et al. Functional and metabolic remodelling in GLUT4-deficient hearts

confers hyper-responsiveness to substrate intervention. *J Mol Cell Cardiol*. 2008; 44(2):270-80. DOI:10.1016/j.yjmcc.2007.11.020

8. Jeffery IB, Higgins DG, Culhane AC. Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. *BMC Bioinformatics*. 2006; 7:359. DOI:10.1186/1471-2105-7-359
9. Pirooznia M, Yang JY, Yang MQ, et al. A comparative study of different machine learning methods on microarray gene expression data. *BMC Genomics*. 2008; 9(Suppl 1):S13. DOI:10.1186/1471-2164-9-S1-S13
10. Kondrikov D, Fulton D, Dong Z, et al. Heat shock protein 70 prevents hyperoxia-induced disruption of lung endothelial barrier via caspase-dependent and AIF-dependent pathways. *PLoS One*. 2015; 10(6):e0129343. DOI:10.1371/journal.pone.0129343
11. Logvinenko OV, Rakitina CL, Ponomarenko DG, et al. The features of blood immunological parameters in patients with various forms of brucellosis. *Infektsiya i immunitet*. 2013; 3(3):275-278. (In Russian). DOI:10.15789/2220-7619-2013-3-275-278
12. Wei D, Miao Y, Yu L, et al. Downregulation of microRNA-198 suppresses cell proliferation and invasion in retinoblastoma by directly targeting PTEN. *Mol Med Rep*. 2018; 18(1):595-602. DOI: 10.3892/mmr.2018.8979
13. Yi L, Yuan Y. MicroRNA-618 modulates cell growth via targeting PI3K/Akt pathway in human thyroid carcinomas. *Indian J Cancer*. 2015; 52(Suppl 3):E186-189. DOI: 10.4103/0019-509X.186577
14. Zhang CG, Yang F, Li YH, et al. miR-501-3p sensitizes glioma cells to cisplatin by targeting MYCN. *Mol Med Rep*. 2018; 18(5):4747-4752. DOI: 10.3892/mmr.2018.9458
15. Song L, Ara T, Wu HW, et al. Oncogene MYCN regulates localization of NKT cells to the site of disease in neuroblastoma. *J Clin Invest*. 2007; 117(9):2702-12. DOI:10.1172/JCI30751

Контактная информация:

Филатова Елена Николаевна, кандидат биологических наук, ведущий научный сотрудник лаборатории молекулярной биологии и биотехнологии, ФБУН ННИИЭМ им. академика И.Н. Блохиной Роспотребнадзора
e-mail: filatova@nniem.ru

Corresponding author:

Elena Filatova, Candidate of Biological Sciences, Leading Researcher, Laboratory of Molecular Biology and Biotechnology, Academician I.N. Blokhina Nizhny Novgorod Scientific Research Institute of Epidemiology and Microbiology
e-mail: filatova@nniem.ru

Статья получена: 08.07.2019

Принята в печать: 25.12.2019

