

ISSN 2311-4495
 ISSN 2410-5155 (Online)
 УДК 575.1 + 616.12-007
<https://doi.org/10.18705/2311-4495-2026-13-1-16-30>

Выявление значимых генетических композиций при врожденных пороках сердца с использованием методов машинного обучения

А. В. Шабалдин^{1,2}, Р. Ю. Замараев², А. В. Сеницкая¹,
 М. В. Кобзев¹, И. К. Халивопуло¹, Я. В. Шкитин³,
 Е. В. Шабалдина³, Л. Н. Игишева¹

Контактная информация:

Шабалдин Андрей Владимирович,
 ФГБНУ «НИИ комплексных проблем
 сердечно-сосудистых заболеваний»,
 бульвар академика Л. С. Барбараша, д. 6,
 Кемерово, Россия, 650002.
 E-mail: weit2007@yandex.ru

¹ Федеральное государственное бюджетное научное учреждение «Научно-исследовательский институт комплексных проблем сердечно-сосудистых заболеваний» Министерства науки и высшего образования Российской Федерации, Кемерово, Россия

² Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр информационных и вычислительных технологий», Новосибирск, Россия

³ Федеральное государственное бюджетное образовательное учреждение высшего образования «Кемеровский государственный медицинский университет» Министерства здравоохранения Российской Федерации, Кемерово, Россия

Резюме

Генетические исследования показали ассоциативные связи между полиморфными участками генов, которые детерминируют процессы метаболизма ксенобиотиков, факторы транскрипции, сигнальные пути, мембранные рецепторы и цитокины с плейотропным эффектом, а также антиген-презентирующие молекулы, и врожденными пороками сердца (ВПС). Существенной проблемой при оценке влияния на формирование ВПС широкого набора полиморфных вариантов генов, не связанных между собой сигнальными путями, является их сложное, комбинированное взаимодействие, что требует применения развитой схемы анализа данных бинарной системы «больные — здоровые». **Цель** исследования: оценка значимости широкого набора генов для формирования врожденных пороков сердца. **Материалы и методы.** Проведено типирование полиморфных участков генов иммунного ответа, метаболизма ксенобиотиков, генов сигнальных путей и сравнение генетических профилей у детей со спорадическими (без семейной истории) септальными ВПС с отсутствием хромосомной патологии с контрольной условно-здоровой группой детей с использованием различных классификаторов. Выполнен анализ данных по широкому спектру полиморфных вариантов генов по развитой схеме: оценка структуры данных; выбор значимых полиморфных вариантов генов, ассоциированных с ВПС; обучение и тестирование нелинейных классификаторов; повторный выбор ассоциативных вариантов полиморфных участков генов с оптимизацией нелинейных классификаторов; построение линейного классификатора. **Результаты.** Выявлены наиболее значимые полиморфные варианты генов, детерминирующих формирование ВПС: *GSTP rs1871042*, *GSTP rs1695*, *CYP1A2 rs7622551*, *GATA6 rs10454095*, *TREM1 rs4711668*, *TLR1 rs5743551*, *IL6R rs2228145*, *IL6R rs2229238*, *IL8 rs4073*, *HLADRBI**. Получены модели типа случайного леса, опорных векторов с полиномиальным ядром и гауссовских процессов для классификации как с хорошей точностью по отдельным классам, так и со средней точностью.

Опробован способ повышения точности классификации за счет комбинирования моделей в простую систему голосования, что может иметь прикладной интерес для пилотных исследований широкого спектра полиморфных вариантов генов, не связанных между собой сигнальными путями.

Ключевые слова: дефект межжелудочковой перегородки, дефект межпредсердной перегородки, классификаторы, полиморфные варианты генов, генетический профиль, врожденные пороки сердца

Для цитирования: Шабалдин А.В., Замараев Р.Ю., Сеницкая А.В. и др. Выявление значимых генетических композиций при врожденных пороках сердца с использованием методов машинного обучения. *Трансляционная медицина*. 2026;13(1):16–30. <https://doi.org/10.18705/2311-4495-2026-13-1-16-30>; <https://elibrary.ru/DSXZYJ>

Identification of significant genetic compositions in congenital heart defects using machine learning methods

Andrey V. Shabaldin^{1,2}, Roman Yu. Zamaraev²,
Anna V. Sinitskaya¹, Maksim V. Kobzev¹, Ivan K. Khalivopulo¹,
Yan V. Shkitin³, Elena V. Shabdina³, Lyudmila N. Igisheva¹

Corresponding author:

Andrey V. Shabaldin,
Research Institute for Complex Issues of
Cardiovascular Diseases,
6 Boulevard Academician L.S. Barbarash,
Kemerovo, Russia, 650002.
E-mail: weit2007@yandex.ru

¹ Federal State Budgetary Institution “Research Institute for Complex Issues of Cardiovascular Diseases”, Kemerovo, Russia

² Federal Research Center for Information and Computational Technologies, Novosibirsk, Russia

³ Kemerovo State Medical University, Kemerovo, Russia

Abstract

Genetic studies have shown associations between polymorphic regions of genes that determine the metabolic processes of xenobiotics, transcription factors, signaling pathways, membrane receptors and cytokines with a pleiotropic effect, as well as antigen-presenting molecules, with congenital heart defects (CHD). A significant problem in assessing the impact on the formation of congenital heart disease of a wide set of polymorphic variants of genes unrelated to each other by signaling pathways is their complex, combined interaction, which requires the use of a developed scheme for analyzing data from the binary system “sick — healthy”. **Aim** of the study: to assess the significance of a wide set of genes for the formation of congenital heart defects. **Materials and methods.** Genotyping of polymorphic loci of immune response genes, xenobiotic metabolism genes, and signaling pathway genes was performed and comparison of genetic profiles in children with sporadic (without family history) septal congenital heart disease with the absence of chromosomal pathologies with a control conditionally healthy group of children using various classifiers. Data analysis was performed on a wide range of polymorphic gene variants according to a developed scheme: assessment of the data structure; selection of significant polymorphic variants of genes associated with congenital heart disease; training and testing of nonlinear classifiers; repeated selection of associative variants of polymorphic gene regions with optimization of nonlinear classifiers; construction of a linear classifier. **Results.** The most significant polymorphic variants of genes that determine the formation of congenital heart defects were identified: *GSTP rs1871042*, *GSTP rs1695*, *CYP1A2 rs7622551*, *GATA6 rs10454095*, *TREM1 rs4711668*, *TLR1 rs5743551*, *IL6R rs2228145*, *IL6R rs2229238*, *IL8 rs4073*, *HLADRBI**. Models such as random forest (RF), support vector machines with a polynomial kernel (PSVM) and Gaussian processes for classification (GPC) were obtained with good accuracy both for individual classes and average accuracy. A method has been tested to increase classification accuracy by combining models into a simple voting system, which may be of (applied) practical interest for pilot studies of a wide range of polymorphic variants of genes with unrelated signaling pathways.

Keywords: ventricular septal defect, atrial septal defect classifiers, polymorphic gene variants, genetic profile, congenital heart defects

For citation: Shabaldin AV, Zamaraev RYu, Sinitskaya AV, et al. Identification of significant genetic compositions in congenital heart defects using machine learning methods. *Translational Medicine*. 2026;13(1):16–30. (In Russ.) <https://doi.org/10.18705/2311-4495-2026-13-1-16-30>; <https://elibrary.ru/DSXZYJ>

Список сокращений: ВПС — врожденные пороки сердца, ДМЖП — дефект межжелудочковой перегородки, ДМПП — дефект межпредсердной перегородки, GPC — гауссовский процесс для классификации, PSVM — полиномиальное ядро, RF — случайный лес.

Введение

Многочисленные эпидемиологические исследования показали, что врожденные пороки сердца (ВПС) являются самой распространенной врожденной патологией, которая в отдельных популяциях мира поражает до 1 % живорожденных детей. ВПС вносят существенный вклад в показатели перинатальной, неонатальной и младенческой смертности, особенно это актуально для критических пороков сердца [1–3]. По результатам официальных статистических данных, ежегодно в России регистрируется более 20 тыс. случаев врожденных пороков сердца разной степени сложности у детей, из них 75 % нуждаются в оперативной коррекции порока [4]. Современные исследования показали, что порок сердца является доминирующим проявлением заболевания за счет нарушения гемодинамики, но, вполне вероятно, что у этих детей могут иметь место нарушения по другим системам (иммунная, эндокринная, метаболическая), которые могут быть детерминированы через наследуемые полиморфные варианты генов, ответственных за поддержание иммунометаболического гомеостаза. С этих позиций ВПС можно считать заболеванием всего организма.

В большинстве случаев (более 60 %) ВПС являются несиндромальной спорадической патологией, которую и по своему патогенезу можно отнести к мультифакториальным заболеваниям, не связанным с определенными хромосомными и генными мутациями [5]. Развитие сердечно-сосудистой системы эмбриона/плода в целом регулируется генами, унаследованными от матери и отца и координирующими клеточные процессы (пролиферацию, дифференцировку и апоптоз), что закладывает основу как для нормального развития сердца, так и для патологий [6].

Многочисленные генетические исследования показали ассоциативные связи между полиморфными участками генов, детерминирующих: процессы метаболизма эндо- и ксенобиотиков (*GST*, *CYP*), факторы

транскрипции (*GATA*, *CRELD*), сигнальные пути (*NOTCH*), мембранные рецепторы с плейотропным эффектом (*TREM*, *TLR*) и антиген-презентирующие молекулы (*HLA*), а также цитокины с плейотропным действием, и врожденными пороками сердца [7–18].

В частности, ранее проведенные исследования выявили ассоциации ВПС с полиморфизмами в иммунокомпетентных генах (*TREM*, *TLR*, *IL*, *HLADR*), которые указывают на роль детерминирования процессов нарушения кардиогенеза в эмбриональном периоде онтогенеза медиаторами и регуляторными молекулами иммунной системы [10, 16–19].

При исследовании генетики септальных ВПС были выявлены ассоциации с генами сигнальных путей и факторами транскрипции (*GATA*, *CRELD*, *NOTCH*), через мутации в которых могут быть детерминированы процессы патологического кардиогенеза [12–15].

Широко обсуждается роль экзогенных и эндогенных факторов в качестве тератогенов для формирующейся сердечно-сосудистой системы эмбриона/плода. Так показано, что мутации в генах, детерминирующих процессы детоксикации аэрополлютантов, лекарственных препаратов и витаминов (цитохромы, трансферазы: *GST*, *CYP*), ассоциированы с ВПС [7–9, 11].

В то же время существенной проблемой при оценке влияния на формирование ВПС широкого набора полиморфных вариантов генов, не связанных между собой сигнальными путями, и выделении из этой группы особо значимых мутаций является их сложное, комбинированное взаимодействие, что требует применения развитой схемы анализа данных бинарной системы «больные — здоровые».

Исходя из вышесказанного, была поставлена цель исследования — поиск в широком наборе генов значимых их композиций, определяющих формирование врожденных пороков сердца, с использованием различных приемов оценки и проверки значимости.

Материалы и методы

В соответствии с поставленной целью было проведено типирование полиморфных участков генов, описанных ниже, и сравнение генетических профилей у детей со спорадическими (без семейной истории) септальными ВПС с отсутствием хромосомной патологии с контрольной условно-здоровой группой

детей с использованием различных методов и приемов оценки значимости генов-признаков.

Характеристика обследованных групп.

Исследование выполнено на базе ФГБНУ «НИИ комплексных проблем сердечно-сосудистых заболеваний» Минобрнауки России и ГБУЗ Кемеровской области «Кемеровский областной клинический кардиологический диспансер им. ак. Л. С. Барбараша». Использовался регистр ВПС и ДНК детей с ВПС, находящихся в банке биологического материала [20]. Для исследования было отобрано 207 ДНК детей (**основная группа**) с септальными ВПС (70 детей с дефектом межжелудочковой перегородки — ДМЖП, 106 детей с дефектом межпредсердной перегородки — ДМПП, и 31 ребенок с сочетанием ДМПП и ДПЖП). Средний возраст детей на момент выделения ДНК и внесения информации в регистр составил $1,26 \pm 0,51$ года (от 0,1 до 8 лет). Критериями включения, применяемыми для отбора ДНК из регистра ВПС, в основную группу исследования для детей были:

- наличие установленного диагноза ВПС;
- средний возраст от 1 месяца до 18 лет;
- отсутствие сопутствующих патологий, включая генетические и хромосомные заболевания.

Критериями исключения из основной группы исследования для детей были:

- отсутствие установленного диагноза;
- наличие ВПС в сочетании с другими патологиями, включая генетические и хромосомные заболевания;
- возраст старше 18 лет.

Контрольная группа была сформирована из условно-здоровых детей ($n=105$). В нее вошли дети, проходившие плановое обследование в детских поликлиниках, являющихся клиническими базами Кемеровского государственного медицинского университета. Средний возраст пациентов составил $2,54 \pm 0,83$ года (от 1 до 8 лет).

Критериями для выбора ДНК из банка биологического материала для включения в контрольную группу были:

- отсутствие установленного диагноза;
- средний возраст от 1 месяца до 18 лет;
- отсутствие сопутствующих патологий, включая генетические и хромосомные заболевания.

Соответственно, критериями исключения из этой группы исследования для детей были:

- наличие ВПС в сочетании с другими патологиями, включая генетические;
- возраст более 18 лет.

Всего в исследование вошли ДНК 312 индивидуумов.

Молекулярно-генетические методы исследования.

Исходя из литературных данных было выбрано 39 полиморфных участков генов, с учетом полученных сведений об их ассоциациях с ВПС и/или распространенности минорного аллеля полиморфного сайта в популяции по данным НарМар не более 5%. Характеристики полиморфных участков генов, анализируемых в настоящем исследовании, представлены в таблице 1.

Представленные гены детерминировали молекулы, участвующие в метаболизме эндо- и ксенобиотиков, эмбриогенезе сердца, межклеточных взаимодействиях и врожденном иммунитете (*GSTP1*, *CYP1A1*, *CYP1A2*, *GATA6*, *CRELD1*, *NOTCH*, *TREMI*, *TLR1*, *TLR2*, *TLR4*, *TLR6*, *IL6*, *IL6R*, *IL8*, *IL10*, *TNFA*). Методом ПЦР проводили генотипирование исследуемых полиморфных вариантов генов с использованием TaqMan зондов (Thermo Fisher Scientific, США) на амплификаторе ViiA 7 RealTime PCR System (LifeTechnologies, США).

Кроме того, проведено типирование гена *HLADRBI** (аллели: 01, 04, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 18) (ген иммунного ответа — Ir-ген) с помощью коммерческого набора ООО «ДНК-технология» (Москва, Россия) согласно протоколу производителя.

Таблица 1. Характеристика исследуемых полиморфных вариантов генов

Table 1. Characteristics of the studied polymorphic gene variants

№	Ген (Gene)	Полиморфизм (SNV ID)	Хромосомная позиция в сборке GRCh38p12 (Chromosomal position)	Аллели по ThermoFisher [VIC/FAM]	Частота минорного аллеля в «1000 Genomes» (MAF)	Тип мутации (Variant type)	Запись аминокислотной замены (Amino acid substitution)
1	<i>TLR1</i>	rs5743611	chr4:38798593	C/G	G=0,022564	Missense Variant	Arg80Thr
2	<i>TLR1</i>	rs5743551	chr4:38806033	C/T	T=0,425719	5 Prime UTR Variant	

3	<i>TLR2</i>	rs3804099	chr4:153703504	C/T	C=0,414736	Synonymous Variant	Asn199Asn
4	<i>TLR2</i>	rs5743708	chr4:153705165	A/G	A=0,006789	Missense Variant	Arg753Gln
5	<i>TLR4</i>	rs4986790	chr9:117713024	A/G	G=0,059904	Missense Variant	Asp299Gly/Val
6	<i>TLR4</i>	rs4986791	chr9:117713324	C/T	T=0,040735	Missense Variant	Thr399Ile
7	<i>TLR6</i>	rs3775073	chr4:38828211	T/C	C=0,457468	Missense Variant	Lys421Asn
8	<i>TLR6</i>	rs5743810	chr4:38828729	A/G	A=0,115216	Missense Variant	Ser249Ala/Pro/Thr
9	<i>TREM1</i>	rs1817537	chr6:41276829	C/G	G=0,373003	Intron Variant	
10	<i>TREM1</i>	rs3804277	chr6:41277434	C/T	T=0,366613	Intron Variant	
11	<i>TREM1</i>	rs6910730	chr6:41278895	A/G	G=0,271166	Intron Variant	
12	<i>TREM1</i>	rs7768162	chr6:41287773	A/G	A=0,286741	Upstream Transcript Variant	
13	<i>TREM1</i>	rs2234246	chr6:41276002	C/T	T=0,366613	Non Coding Transcript Variant	
14	<i>TREM1</i>	rs4711668	chr6:41278735	T/C	T=0,296925	Intron Variant	
15	<i>TREM1</i>	rs9471535	chr6:41287752	C/T	C=0,16234	Upstream Transcript Variant	
16	<i>TREM1</i>	rs2234237	chr6:41282728	A/T	A=0,163938	Missense Variant	Thr25Ser
17	<i>IL6R</i>	rs2228145	chr1:154454494	C/A	C=0,293131	Missense Variant	Asp358Val/Ala
18	<i>IL6R</i>	rs2229238	chr1:154465420	G/T	T=0,203075	3 Prime UTR Variant	
19	<i>IL6</i>	rs1800796	chr7:22726627	C/G	C=0,313898	Intron Variant	
20	<i>IL6</i>	rs2069827	chr7:22725837	G/T	T=0,036542	Upstream Transcript Variant	
21	<i>IL8 (CXCL8)</i>	rs4073	chr4:73740307	A/T	T=0,477835	Upstream Transcript Variant	
22	<i>IL10</i>	rs1800871	chr1:206773289	A/G	A=0,434704	Upstream Transcript Variant	
23	<i>IL10</i>	rs1800872	chr1:206773062	T/G	T=0,434904	Upstream Transcript Variant	
24	<i>IL10</i>	rs1800896	chr1:206773552	T/C	C=0,272165	Upstream Transcript Variant	
25	<i>TNF</i>	rs1799964	chr6:31574531	C/T	C=0,21905	Upstream Transcript Variant	
26	<i>TNF</i>	rs361525	chr6:31575324	A/G	A=0,060903	Upstream Transcript Variant	
27	<i>TNF</i>	rs1800629	chr6:31575254	A/G	A=0,090256	Upstream Transcript Variant	
28	<i>NOTCH1</i>	rs13290979	chr9:136531182	A/G	G=0,485823	Intron Variant	
29	<i>GSTP1</i>	rs6591256	chr11:67582428	A/G	G=0,309505	Upstream Transcript Variant	
30	<i>GSTP1</i>	rs1695	chr11:67585218	A/G	G=0,352636	Missense Variant	Ile105Val/Phe
31	<i>GSTP1</i>	rs1871042	chr11:67586373	T/C	T=0,248203	Intron Variant	
32	<i>GSTP1</i>	rs1793068	chr11:106553564	NA	G=0,199281	Not announced	

33	<i>CYP1A1</i>	rs1048943	chr15:74720644	C/T	C=0,133387	Missense Variant	Ile462Phe/Val/Leu
34	<i>CYP1A2</i>	rs762551	chr15:74749576	C/A	C=0,370208	Intron Variant	
35	<i>CYP1A2 (Deletion)</i>	rs35694136	chr15:74747272	T/-	*=0,384984	Upstream Variant	
36	<i>CRELD1</i>	rs73118372	chr3: 9943989	T/C	C=0,009	Missense Variant	Met383Arg
37	<i>CRELD1</i>	rs9878047	chr3:9943773	T/C	C=0,266	Intron variant	
38	<i>CRELD1</i>	rs3774207	chr3:9943974	C/T	T=0,261	Synonyms variant	Amino acid codon H {CAC} >H [CAT]
39	<i>GATA6</i>	rs10454095	chr:22197478	C/T	T=0,351	Intron variant	

Подготовка и анализ данных. В качестве зависимой переменной введен признак наличия ВПС (больной, метка 1) или отсутствия ВПС (здоровый, метка 0). Независимыми переменными являются аллели исследованных генов. Для каждого аллеля был выставлен свой балл (0 — отсутствие аллеля, 1 — аллель присутствовал в гетерозиготе, 2 — аллель был в гомозиготе). Исследуемые аллели были пронумерованы: всего в анализ вошло 93 аллеля (табл. 2).

Для полиморфного варианта гена *CYP1A2* (Deletion) rs3569413, где присутствию Т-аллеля противопоставляется его делеция, запись проводилась как Т (наличие аллеля Т) и del (наличие делеции).

Современный уровень развития методов анализа и вычислительных технологий позволяет использовать для построения моделей данных все имеющиеся в настоящем исследовании показатели (аллели). Основная цель такого исследования — ранжирование

Таблица 2. Перечень и нумерация исследуемых аллелей полиморфных вариантов генов

Table 2. List and numbering of the studied alleles of polymorphic gene variants

1. <i>GSTP1RS1793068G</i>	17. <i>CYP1A2RS35694136T</i>	33. <i>TREM1RS7768162A</i>	49. <i>TLR4RS4986791C</i>	65. <i>IL8RS4073A</i>	81. <i>HLADRB1*4</i>
2. <i>GSTP1RS1793068T</i>	18. <i>CYP1A2RS35694136del</i>	34. <i>TREM1RS7768162G</i>	50. <i>TLR4RS4986791T</i>	66. <i>IL8RS4073T</i>	82. <i>HLADRB1*7</i>
3. <i>GSTPRS6591256A</i>	19. <i>CYP1A2RS762551C</i>	35. <i>TREM1RS4711668T</i>	51. <i>TLR4RS4986790A</i>	67. <i>IL10RS1800871A</i>	83. <i>HLADRB1*8</i>
4. <i>GSTPRS6591256G</i>	20. <i>CYP1A2RS762551A</i>	36. <i>TREM1RS4711668C</i>	52. <i>TLR4RS4986790G</i>	68. <i>IL10RS1800871G</i>	84. <i>HLADRB1*9</i>
5. <i>GSTPRS1871042C</i>	21. <i>GATA6RS10454095C</i>	37. <i>TREM1RS2234246C</i>	53. <i>TLR6RS3775073T</i>	69. <i>IL10RS1800896T</i>	85. <i>HLADRB1*10</i>
6. <i>GSTPRS1871042T</i>	22. <i>GATA6RS10454095T</i>	38. <i>TREM1RS2234246T</i>	54. <i>TLR6RS3775073C</i>	70. <i>IL10RS1800896C</i>	86. <i>HLADRB1*11</i>
7. <i>GSTPRS1695A</i>	23. <i>NOTCHRS13290979A</i>	39. <i>TREM1RS9471535C</i>	55. <i>TLR6RS5743810A</i>	71. <i>IL10RS1800872T</i>	87. <i>HLADRB1*12</i>
8. <i>GSTPRS1695G</i>	24. <i>NOTCHRS13290979G</i>	40. <i>TREM1RS9471535T</i>	56. <i>TLR6RS5743810G</i>	72. <i>IL10RS1800872G</i>	88. <i>HLADRB1*13</i>
9. <i>CRELD1RS73118372C</i>	25. <i>TREM1RS6910730A</i>	41. <i>TLR1RS5743611C</i>	57. <i>IL6RS1800796C</i>	73. <i>TNFRS1800629A</i>	89. <i>HLADRB1*14</i>
10. <i>CRELD1RS73118372T</i>	26. <i>TREM1RS6910730G</i>	42. <i>TLR1RS5743611G</i>	58. <i>IL6RS1800796G</i>	74. <i>TNFRS1800629G</i>	90. <i>HLADRB1*15</i>
11. <i>CRELD1RS9878047C</i>	27. <i>TREM1RS3804277C</i>	43. <i>TLR1RS5743551C</i>	59. <i>IL6RS2069827G</i>	75. <i>TNFRS361525A</i>	91. <i>HLADRB1*16</i>
12. <i>CRELD1RS9878047T</i>	28. <i>TREM1RS3804277T</i>	44. <i>TLR1RS5743551T</i>	60. <i>IL6RS2069827T</i>	76. <i>TNFRS361525G</i>	92. <i>HLADRB1*17</i>
13. <i>CRELD1RS3774207C</i>	29. <i>TREM1RS1817537C</i>	45. <i>TLR2RS5743708A</i>	61. <i>IL6RRS2228145C</i>	77. <i>TNFRS1799964C</i>	93. <i>HLADRB1*18</i>
14. <i>CRELD1RS3774207T</i>	30. <i>TREM1RS1817537G</i>	46. <i>TLR2RS5743708G</i>	62. <i>IL6RRS2228145A</i>	78. <i>TNFRS1799964T</i>	
15. <i>CYP1A1RS1048943C</i>	31. <i>TREM1RS2234237A</i>	47. <i>TLR2RS3804099C</i>	63. <i>IL6RRS2229238C</i>	79. <i>HLADRB1*1</i>	
16. <i>CYP1A1RS1048943T</i>	32. <i>TREM1RS2234237T</i>	48. <i>TLR2RS3804099T</i>	64. <i>IL6RRS2229238T</i>	80. <i>HLADRB1*3</i>	

Примечание: Заливкой отмечены аллели, определенные как информативные на первом этапе исследования по ранговой корреляции.

Note: The highlighted alleles are those identified as informative at the first stage of the study based on rank correlation.

показателей по значимости, что позволяет открыть внутренние взаимосвязи и соотношения взаимодействующих аллелей. В методическом плане это крайне важно для понимания принципов формирования ВПС.

Также ранжирование показателей по значимости позволяет понизить размерность данных и повысить устойчивость классификаторов для решения прикладных задач, прежде всего, диагностики вновь поступающих пациентов.

Поэтому для достижения поставленной цели была реализована следующая развернутая схема анализа данных:

1. Оценка структуры данных и первичный отбор по ранговой корреляции значимых полиморфных вариантов генов, ассоциированных с ВПС;
2. Выбор лучших классификаторов через многократное обучение и тестирование;
3. Повторный выбор ассоциативных вариантов полиморфных участков генов по значимости для точности лучших классификаторов;
4. Оценка «направленности» генов относительно классов и пробное построение линейного классификатора.

Результаты и обсуждение

На первом этапе значимость генов исследовалась методами ранговой корреляции с вектором принадлежности пациентов к классу больных и здоровых. Используются две меры — коэффициент корреляции Кендалла (τ) и коэффициент корреляции Спирмена (ρ)

[21], ввиду того, что показатели имеют размерность в виде баллов.

На рисунке 1 представлены значения указанных коэффициентов корреляции аллелей с классом пациентов. Многие гены составляют разнонаправленные пары аллелей с (почти) равными по модулю коэффициентами корреляции. Максимальное значение корреляции по модулю составляет $\rho=0,16805$ и $\tau=0,16170$ для полиморфного участка гена *IL6R* rs2228145 и его аллелей С № 61 и А № 62. Оба коэффициента между собой хорошо согласованы, их собственный коэффициент корреляции Пирсона равен 0,999. Причем аллель С имел отрицательный знак, а аллель А — положительный, что указывает на возможную предикторную роль минорного аллеля А полиморфного участка гена *IL6R*, обуславливающей миссенс-мутацию с заменой аминокислоты (Asp358Val/Ala).

Кроме того, этот этап показал, что для многих исследуемых полиморфных участков генов утверждается связь между аллелями. Также обнаруживается очень низкая средняя корреляция набора генов с классом индивидуумов, то есть многие из них совершенно «слепые» относительно классов больных и здоровых. Поэтому выбор наиболее значимых генов определенно является актуальной задачей.

Для выбора значимых показателей и некоторого повышения достоверности этого выбора использована обобщенная мера: $Overall=|\rho|+|\tau|$ (рис. 1). Значимым считался аллель, у которого сумма коэффициентов

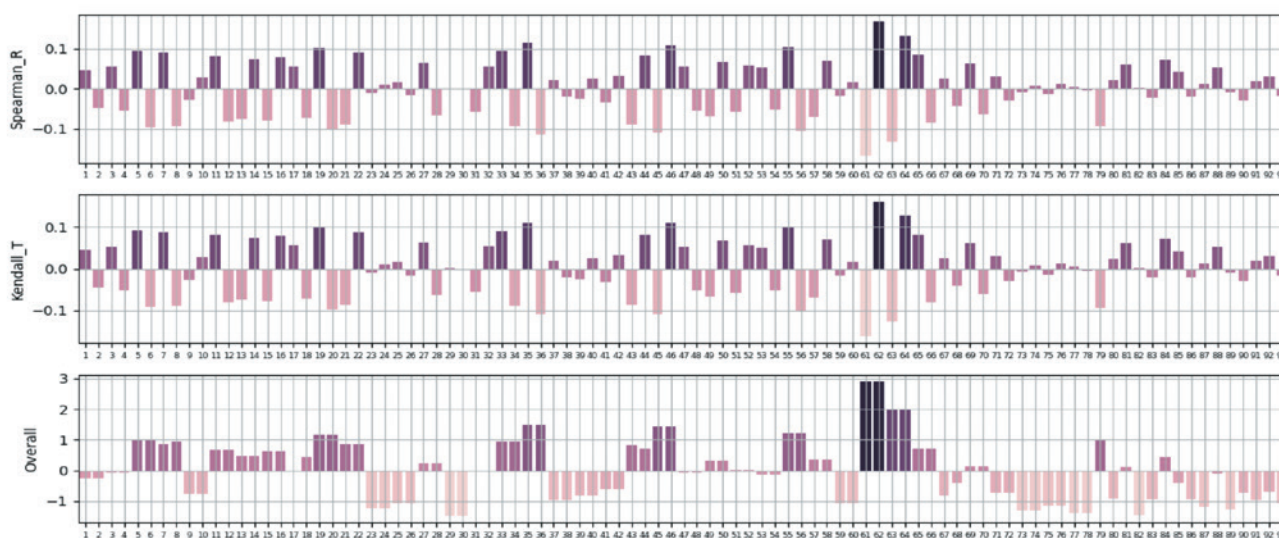


Рис. 1. Коэффициенты ранговой корреляции генов (баллов аллелей) с меткой класса индивидуумов (больные/здоровые) и комбинированная мера оценки их значимости

Fig. 1. Rank correlation coefficients of genes (allele scores) with the class label of individuals (patients/healthy) and a combined measure for assessing their significance

выше среднего по совокупности. Для наглядности эта мера представлена на рисунке 1 в стандартизованном виде, здесь условие отбора гена принимает вид: $Overall > 0$.

По данному условию прошло 44 аллеля, которые отмечены в таблице 2. Как видно из этой таблицы, 40 показателей — это аллельные варианты 20 генов, 3 аллеля — гена *HLADRBI** и одна делеция в полиморфном участке гена *CYP1A2* rs35694136. Для всех 20 генов показано разнонаправленное действие их аллелей (рис. 1: Spearman, Kendall) в отношении формирования ВПС. Для гена *HLADRBI** два аллеля 04 и 09 имели, предположительно, предикторное влияние на формирование ВПС.

На стадии разведочного анализа было произведено пробное обучение широкого набора классификаторов: метод ближайших соседей (K-Nearest Neighbors, KNN); метод опорных векторов (Support Vector Machines) с линейным ядром (LSVM) и полиномиальным ядром (PSVM); гауссовский процесс для классификации (Gaussian Process Classification, GPC); дерево решений (Classification and Regression Tree, CART); случайный лес (Random Forest, RF); нейронная сетевая модель в виде многослойного перцептрона (Multi Layer Perceptron, MLP); адаптивный бустинг (Adaptive Boosting, ABOOST); наивный байесовский классификатор (Naive Bayes Classifier, NBC); линейный бинарный классификатор (Linear Binary Classifier, LBC).

Модели обучались с настройками по умолчанию на общей для всех случайной обучающей и тестовой выборках с использованием всех аллелей исследуемых полиморфных вариантов генов, которые были определены как информативные на первом этапе исследования (табл. 2).

Ввиду сильного дисбаланса (~1:2) численности классов больных и здоровых, их представительство в обучающих выборках выравнивалось — выбирались по 82 случайных индивидуума. На рисунке 2 представлены матрицы ошибок моделей на полной выборке — по главной диагонали указаны доли правильных классификаций, по обратной — доли ошибочных, под псевдонимом моделей в заголовках указана их средняя точность. Как видно из рисунка 2, отдельные модели даже без оптимизации демонстрируют хорошую точность по обоим классам.

Далее, в ходе вычислительного эксперимента с многократной генерацией обучающих выборок по описанной выше схеме, был получен набор лучших моделей — результирующие матрицы ошибок на полном наборе данных представлены на рисунке 3.

На следующем этапе был проведен повторный отбор аллелей по значимости из 44 отобранных ранее. Здесь преследовалась цель дальнейшего понижения размерности задачи путем отбраковки аллелей с использованием полученных лучших классификаторов. Чтобы оценить направленность или влияние аллелей на детерминирование формирования ВПС,

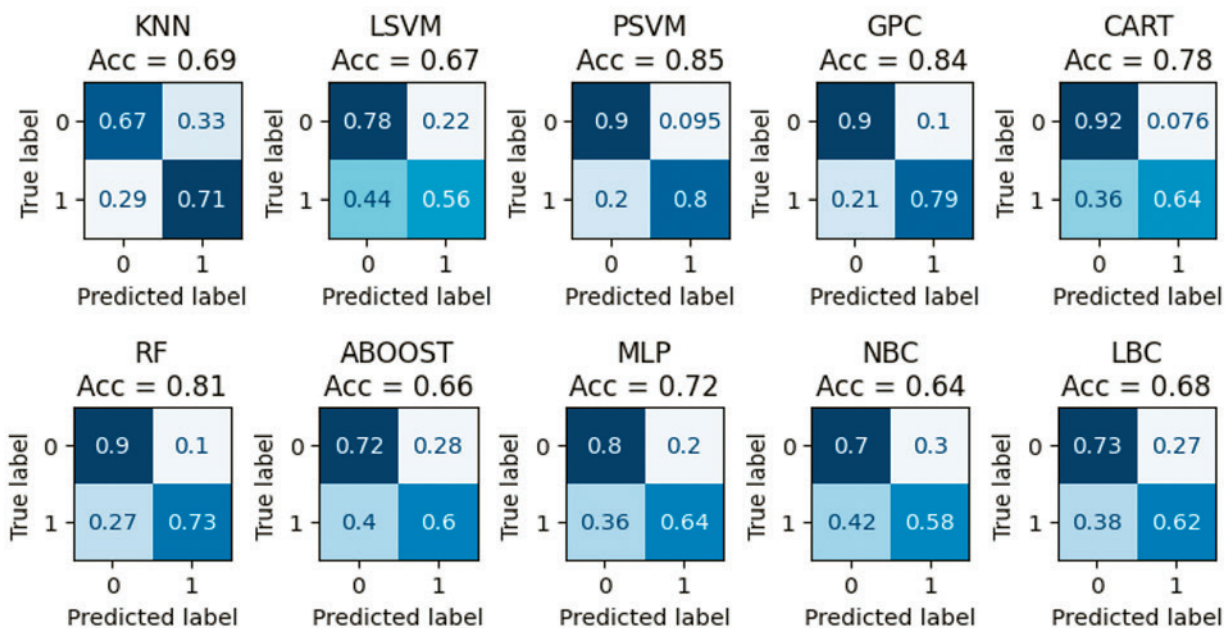


Рис. 2. Классификация полного набора пациентов по 44 генам: матрицы ошибок с нормировкой по истинному классу (True label)

Fig. 2. Results of training classifiers with default settings: confusion matrices normalized by true class (True label)

в нелинейных классификаторах использована известная мера — «важность при перестановке» (Permutation Importance) [22]. По упомянутому алгоритму для экземпляров моделей PSVM, GPC и RF были вычислены значимости аллелей. Как и ранее, для получения обобщенной оценки значимости *Overall*, частные оценки были просуммированы и приведены к стандартизированному виду. Полная картина представлена на рисунке 4.

В итоге в список значимых по условию *Overall*>0 вошли 17 аллелей: 5. *GSTPRS1871042C*; 6. *GSTPRS1871042T*; 7. *GSTPRS1695A*; 8. *GSTPRS1695G*; 19. *CYP1A2RS762551C*; 20. *CYP1A2RS762551A*; 21. *GATA6RS10454095C*; 22. *GATA6RS10454095T*; 35. *TREM1RS4711668T*; 36. *TREM1RS4711668C*; 43. *TLR1_RS5743551C*; 44. *TLR1_RS5743551T*; 61. *IL6RRS2228145C*; 62. *IL6RRS2228145A*; 63. *IL6RRS2229238C*; 64. *IL6RRS2229238T*; 65. *IL8RS4073A*.

Как видно из списка, в финальный отбор попали 9 пар аллелей, то есть 9 исследуемых генов, а также аллель гена *HLADRBI*01*. Для сокращенного по списку генов набора данных были обучены те же модели и построены матрицы ошибок, результат представлен на рисунке 5.

Линейные классификаторы показывали одни из худших результатов по точности на всех этапах, что на фоне хорошей точности лучших нелинейных моделей и при большом числе показателей подталкивает к мысли о наличии в анализируемых данных каких-либо структурных нелинейностей, например, типа XOR (The XOR problem) [25]. В основе такой сложной структуры данных, по всей видимости, лежит разнонаправленность исследуемых генов, детерминирующих метаболизм эндо- и ксенобиотиков, межклеточные взаимодействия, факторы транскрипции, врожденный и адаптивный иммунитет, с одной

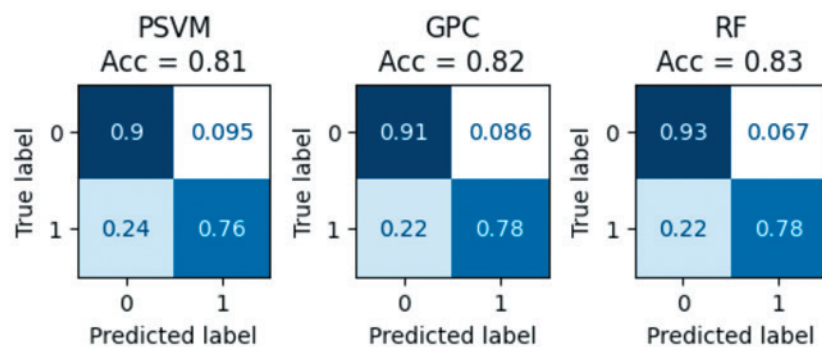


Рис. 3. Классификация полного набора пациентов по 44 генам: матрицы ошибок с нормировкой по истинному классу (True label)

Fig. 3. Classification of the full set of patients based on 44 genes: confusion matrices normalized by the true class (True label)

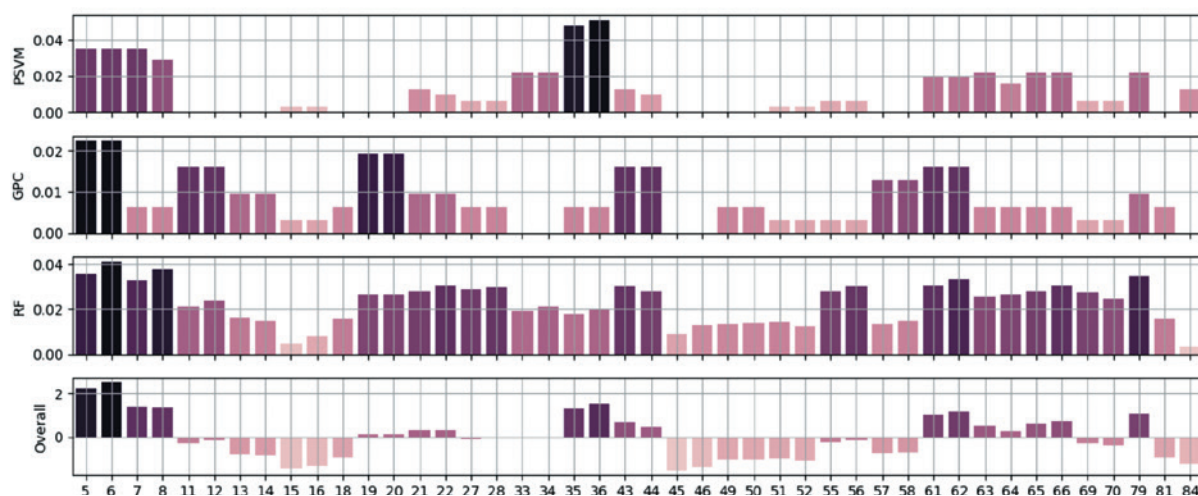


Рис. 4. Значимость аллелей в моделях по схеме Permutation Importance

Fig. 4. Importance of alleles in models according to the Permutation Importance scheme

стороны, и некоторый полиморфизм в рамках ВПС (дети с изолированными ДМПП, с изолированными ДМЖП, а также дети с сочетанием ДМПП и ДМЖП).

Единственный способ использовать линейные модели для существенно нелинейных данных — это генерация синтетических показателей, создающих новые измерения, в которых классы, возможно, окажутся линейно разделимыми. С этой идеей бы проведен численный эксперимент по обучению

логнормального классификатора по схеме 5-фолдовой кросс-валидации из уравновешенной выборки по 82 случайных представителя от обоих классов (больные и здоровые). Синтетические данные генерировались только из мультипликативных комбинаций по две, три и четыре аллели. На рисунке 6 представлены матрицы ошибок классификаторов, обученных на таких синтетических показателях при классификации полного набора данных.

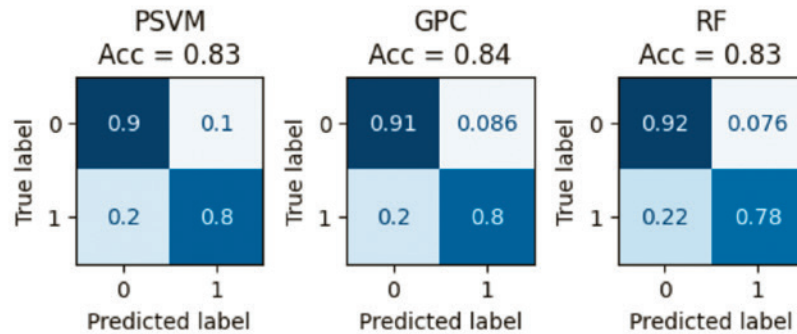


Рис. 5. Классификация полного набора пациентов по 19 генам: матрицы ошибок с нормировкой по истинному классу (True label)

Fig. 5. Classification of the full set of patients based on 19 genes: confusion matrices normalized by the true class (True label)

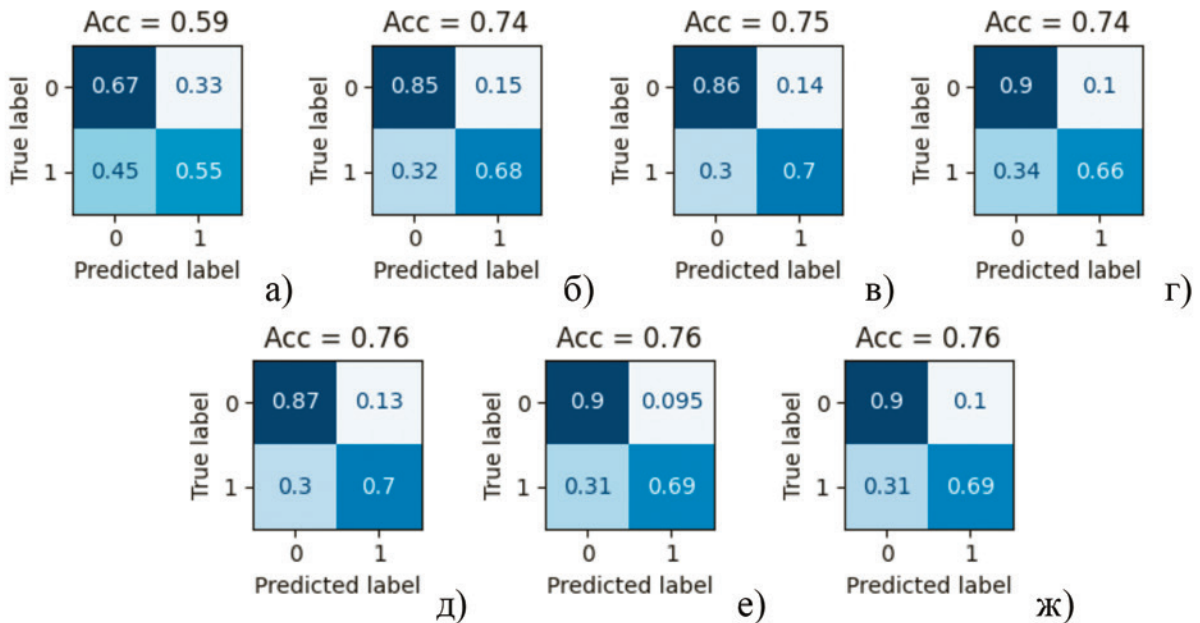


Рис. 6. Матрицы ошибок логнормального классификатора при разных составах мультипликативных синтетических показателей: а) один аллель от каждого гена — всего 10 показателей; б) комбинации по 2 (всего 45); в) комбинации по 3 (всего 120); г) комбинации по 4 (всего 210); д) комбинации по 2 и 3 (всего 165); е) комбинации по 3 и 4 (330); ж) комбинации по 2, 3 и 4 (всего 375)

Fig. 6. Confusion matrices of the log-normal classifier for different compositions of multiplicative synthetic indicators: а) one allele from each gene — 10 indicators in total; б) combinations of 2 (45 in total); в) combinations of 3 (120 in total); д) combinations of 4 (210 in total); е) combinations of 2 and 3 (165 in total); ф) combinations of 3 and 4 (330); г) combinations of 2, 3, and 4 (375 in total)

Для двух моделей — позитивные показатели (баллы аллелей) в логнормальном классификаторе и синтетические показатели — был выполнен ROC-анализ, представленный на рисунке 7. Обращает внимание следующее: результат обучения на всех 10 самых значимых генах (рис. 6, а, и 7, а) заметно хуже, чем результат обучения на полном наборе исходных данных (рис. 2, LBC), но этот результат, в свою очередь, заметно хуже любого, полученного на синтетических данных. Отсюда можно сделать вывод: либо в лучших по значимости генах остались «слепые» относительно целевого класса, либо действительно имеет место существенная нелинейность, поскольку даже сложные мультипликативные комбинации полностью ее не устранили.

Таким образом, исследование выявило 9 полиморфных участков генов, с вовлечением как мажорных, так и минорных аллелей, а также один аллель *HLADRB1*01*, которые могут принимать важное значение в формировании ВПС. В данном случае имеют место уникальные комбинации этих аллелей, своеобразные «генетические профили», которые четко разграничиваются нелинейными классификаторами по классу (дети с ВПС и условно-здоровые дети). Этими полиморфными участками генов оказались: *GSTP rs1871042*, *GSTP rs1695*, *CYP1A2 rs7622551*, *GATA6 rs10454095*, *TREM1 rs4711668*, *TLR1 rs5743551*, *IL6R rs2228145*, *IL6R rs2229238*, *IL8 rs4073*, *HLADRB1**.

Необходимо отметить, что в «генетических профилях», характерных для детей с ВПС, могут быть

любые аллели (мажорные и минорные) из этих од- нонуклеотидных точковых полиморфных участков генов, а аллель *HLADRB1*01* может как присут- ствовать, так и отсутствовать в них. Также крайне трудно описать характерные гаплотипы по этим полиморфным участкам генов для спорадических септальных ВПС без хромосомной патологии.

С учетом данных о характеристиках этих поли- морфных участков исследуемых генов, представлен- ных в таблице 1, можно сделать предварительное заключение о том, что в гаплотипы, ассоциированные с ВПС, входят мажорные аллели *GSTP rs1871042*, *GSTP rs1695*, *IL8 rs4073*, минорные аллели *CYP1A2 rs7622551*, *GATA6 rs10454095*, *TREM1 rs4711668*, *TLR1 rs5743551*, *IL6R rs2228145*, *IL6R rs2229238*, а аллель *HLADRB1*01* отсутствует. Причем только два полиморфных варианта этих генов кодируют *missens*-мутацию, приводящую к замене аминокисло- ты: *GSTP rs1695* и *IL6R rs2228145*. Соответственно, предположительный для ВПС гаплотип кодирует только мутацию (Asp358Val/Ala) в рецепторе к ин- терлейкину-6 (*IL6R rs2228145*). Тем самым, выявлены наиболее значимые полиморфные варианты генов, детерминирующих формирование ВПС. Эти данные указывают и на роль интерлейкина-6 и его рецептора в формировании септальных ВПС, а также в разви- тии иммунопатологических состояний у этих детей на этапах лечения и реабилитации порока сердца [26]. О значении мутаций в генах цитокинов при ВПС также неоднократно говорилось, в том числе и в ранее проведенных исследованиях [19, 26–28]. Известно,

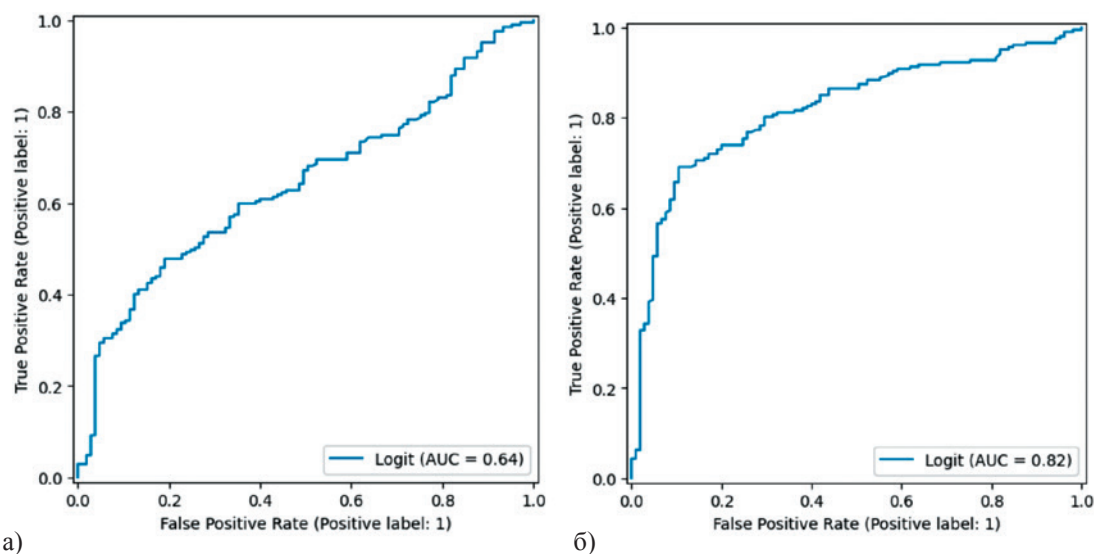


Рис. 7. ROC-кривая логнормального классификатора: а) обученного на 10 значимых генах; б) обученного на синтетических данных, составленных из мультипликативных комбинаций по 3 гена

Fig. 7. ROC curve of the log-normal classifier: а) trained on 10 significant genes; б) trained on synthetic data composed of multiplicative combinations of 3 genes

что рецепторы к цитокинам обладают плеiotропным эффектом, через активирование большого количества внутриклеточных сигнальных путей [19]. Этим же эффектом обладает и рецептор к интерлейкину-6, который экспрессируется на прогениторных клетках сердечно-сосудистой системы с 5–7 недели гестации [29]. Дефицит передачи сигнала в эти клетки может быть ключевым в патогенезе ВПС. С другой стороны, увеличение несвязанного IL-6 может способствовать нарушению в цитокиновой сети и, как следствие, активации пироптоза, в том числе прогениторных клеток формирующейся сердечно-сосудистой системы. Показаны высокие концентрации интерлейкина-6 в пуповинной крови детей, родившихся с ВПС [30]. Этот феномен требует дальнейшего изучения.

Выводы:

1. Выделенные значимые гены и их отношения проявились как таковые на всех стадиях отбора и не противоречат современным представлениям о механизмах формирования ВПС; ими оказались: *GSTP rs1871042*, *GSTP rs1695*, *CYP1A2 rs7622551*, *GATA6 rs10454095*, *TREM1 rs4711668*, *TLR1 rs5743551*, *IL6R rs2228145*, *IL6R rs2229238*, *IL8 rs4073*, *HLADRBI**.

2. Предположительный для ВПС гаплотип содержит: минорные аллели *CYP1A2 rs7622551*, *GATA6 rs10454095*, *TREM1 rs4711668*, *TLR1 rs5743551*, *IL6R rs2228145*, *IL6R rs2229238* и кодирует мутацию (Asp358Val/Ala) в рецепторе к интерлейкину-6 (*IL6R rs2228145*).

3. Обнаружена сильная нелинейность в данных (баллах аллелей) по отношению к классам больных и здоровых, ввиду разностороннего действия продуктов этих генов на ВПС, которая не устраняется полностью даже при введении в обучающие данные синтетических показателей; это ставит под сомнение использование линейных моделей в практических приложениях.

4. Получены классификаторы типа случайного леса (RF), опорных векторов с полиномиальным ядром (PSVM) и гауссовских процессов для классификации (GPC) с хорошей точностью для диагностирования ВПС у новых пациентов, что может представлять интерес для прикладных исследований по созданию систем диагностирования.

Конфликт интересов / Conflict of interest

Авторы заявили об отсутствии потенциального конфликта интересов. / The authors declare no conflict of interest.

Финансирование / Funding

Работа выполнена в рамках основной темы ФНИ ФГБНУ «НИИ комплексных проблем сердечно-сосудистых заболеваний» Минобрнауки России

№ 0419-2022-0001 «Молекулярные, клеточные и биомеханические механизмы патогенеза сердечно-сосудистых заболеваний в разработке новых методов лечения заболеваний сердечно-сосудистой системы на основе персонифицированной фармакотерапии, внедрения малоинвазивных медицинских изделий, биоматериалов и тканеинженерных имплантатов» (№ госрегистрации 122011900095-2 от 19.01.2022).

В части анализа данных работа выполнена в рамках государственного задания Минобрнауки России для Федерального исследовательского центра информационных и вычислительных технологий. Базовая тема: Информационные системы мониторинга и управления безопасностью социально-природно-техногенных систем, цифровые модели материалов и конструкций, автоматизации технологических процессов и мониторинга объектов техносферы (регистрационный номер ЕГИСУ НИОКТР — 125051306026-9).

Соответствие нормам этики / Compliance with ethical principles

Исследование одобрено локальным этическим комитетом. / The study was approved by the Local Ethics Committee.

Авторы заявляют об отсутствии использования генеративного искусственного интеллекта. / The authors declare no use of Generative AI in the preparation of this manuscript.

Список источников / References

1. Саперова Е. В., Вахлова И. В. Врожденные пороки сердца у детей: распространенность, факторы риска, смертность. *Вопросы современной педиатрии*. 2017;16(2):126–133. <https://doi.org/10.15690/vsp.v16i2.1713>
2. Saperova EV, Vahlova IV. Congenital heart diseases in children: incidence, risk factors, mortality. *Current Pediatrics*. 2017;16(2):126–133. (In Russ.) <https://doi.org/10.15690/vsp.v16i2.1713>
3. Xu J, Li Q, Deng L, et al. Global, regional, and national epidemiology of congenital heart disease in children from 1990 to 2021. *Front Cardiovasc Med*. 2025;12:1522644. <https://doi.org/10.3389/fcvm.2025.1522644>
4. Официальный веб-сайт Европейского союза [Интернет]. Доступно по ссылке: https://eu-rd-platform.jrc.ec.europa.eu/eurocat/eurocat-data/prevalence_en.
5. Бокерия Л. А., Шаталова К. В. Детская кардиохирургия: руководство для врачей. М.: НЦССХ им. А. Н. Бакулева; 2016. С. 864.
6. Bockeria LA, Shatalova KV. Pediatric cardiac surgery: a guide for physicians. Moscow: A. N. Bakulev Scientific Center for Cardiovascular Surgery; 2016. 864 p. (In Russ.)

5. Checchia PA, Laussen PC. The cardiac intensive care unit perspective on hemodynamic monitoring of oxygen transport balance. *Pediatr. Crit. Care Med.* 2011;12(4):69–71. <https://doi.org/10.1097/PCC.0b013e3182211d3d>
6. Suluba E, Shuwei L, Xia Q, Mwanga A. Congenital heart diseases: genetics, non-inherited risk factors, and signaling pathways. *Egyptian Journal of Medical Human Genetics.* 2020;21:1–12. <https://doi.org/10.1186/s43042-020-0050-1>
7. Брайко О. П., Лазарев К. Ю., Полоников А. В. и др. Молекулярно-генетический анализ полиморфизма Val432Leu гена CYP1B1 и G590A гена NAT2 у детей с изолированным дефектом межжелудочковой перегородки в Краснодарском крае. *Кубанский научный медицинский вестник.* 2013;5:52–55. <https://www.elibrary.ru/rbpfvp>
- Brayko OP, Lazarev KYu, Polonikov AV, et al. Molecular genetic analysis of the CYP1B1 gene Val432Leu and NAT2 gene G590A polymorphisms in children with isolated ventricular septal defect in the Krasnodar region. *Kuban Scientific Medical Bulletin.* 2013;5:52–55. (In Russ.) <https://www.elibrary.ru/rbpfvp>
8. Лазарев К. Ю., Брайко О. П., Голубцов В. И. и др. Молекулярно-генетический анализ полиморфизмов Val432Leu гена CYP1B1, G590A гена NAT2 и C3435T гена ABCB1 у детей с изолированным дефектом предсердной перегородки в Краснодарском крае. *Актуальные проблемы медицины.* 2013;23(18(161)):152–157. <https://www.elibrary.ru/rzvfqz>
- Lazarev KYu, Brayko OP, Golubtsov VI, et al. Molecular genetic analysis of the CYP1B1 gene Val432Leu, NAT2 gene G590A and ABCB1 gene C3435T polymorphisms in children with isolated atrial septal defect in the Krasnodar region. *Current Problems of Medicine.* 2013;23(18(161)):152–157. (In Russ.) <https://www.elibrary.ru/rzvfqz>
9. Швецов Я. Д., Лазарев К. Ю., Бушуева О. Ю. и др. Исследование ассоциации полиморфизма I462V гена CYP1A1 с развитием врожденного дефекта межжелудочковой перегородки сердца в Краснодарском крае. *Научные ведомости Белгородского государственного университета. Серия: Медицина. Фармация.* 2015;10(207):108–112. <https://elibrary.ru/uzcyxz>
- Shvetsov YaD, Lazarev KYu, Bushueva OYu, et al. Association study of the CYP1A1 gene I462V polymorphism with the development of congenital ventricular septal defect in the Krasnodar region. *Scientific Bulletin of Belgorod State University. Series: Medicine. Pharmacy.* 2015;10(207):108–112. (In Russ.) <https://elibrary.ru/uzcyxz>
10. Цепочкина А. В., Хуторная М. В., Шабалдин А. В., Понасенко А. В. Особенности распределения генотипов полиморфных вариантов rs2234246 и rs4711668 TREM-1 у детей с дуктус-зависимыми врожденными пороками сердца. *Трансляционная медицина.* 2019;6(4):5–12. <https://doi.org/10.18705/2311-4495-2019-6-4-5-12>
- Tsepokina AV, Khutorная MV, Shabaldin AV, Ponasenko AV. The role of gene TREM-1 at children who have operation congenital heart diseases. *Translational Medicine.* 2019;6(4):5–12. (In Russ.) <https://doi.org/10.18705/2311-4495-2019-6-4-5-12>
11. Li N, Kang H, Liu Z, et al. Association of maternal phthalates exposure and metabolic gene polymorphisms with congenital heart diseases: a multicenter case-control study. *BMC Pregnancy Childbirth.* 2024;24(1):167. <https://doi.org/10.1186/s12884-024-06343-z>
12. Kukshal P, Joshi RO, Kumar A, et al. Case-control association study of congenital heart disease from a tertiary pediatric cardiac centre from North India. *BMC Pediatr.* 2023;23(1):290. <https://doi.org/10.1186/s12887-023-04095-x>
13. Wu H, Wu H, He Y, et al. Functional characterization of GATA6 genetic variants associated with mild congenital heart defects. *Biochem Biophys Res Commun.* 2023;641:77–83. <https://doi.org/10.1016/j.bbrc.2022.12.004>
14. Ghosh P, Bhaumik P, Ghosh S, et al. Polymorphic haplotypes of CRELD1 differentially predispose Down syndrome and euploids individuals to atrioventricular septal defect. *American journal of medical genetics. Part A.* 2012;158A(11):2843–2848. <https://doi.org/10.1002/ajmg.a.35626>
15. Helle E, Córdova-Palomera A, Ojala T, et al. Loss of function, missense, and intronic variants in NOTCH1 confer different risks for left ventricular outflow tract obstructive heart defects in two European cohorts. *Genet Epidemiol.* 2019;43(2):215–226. <https://doi.org/10.1002/gepi.22176>
16. Шабалдин А. В., Синицкая А. В., Шмулевич С. А. Роль генов цитокинов и Toll-подобных рецепторов в патогенезе врожденных пороков сердца. *Медицинская иммунология.* 2022;24(3):605–616. <https://doi.org/10.15789/1563-0625-ROC-2488>
- Shabaldin AV, Sinitskaya AV, Shmulevich SA. Role of cytokine and Toll-like receptor genes in pathogenesis of inborn heart disease. *Medical Immunology.* 2022;24(3):605–616. (In Russ.) <https://doi.org/10.15789/1563-0625-ROC-2488>
17. Шабалдин А. В., Цепочкина А. В., Шмулевич С. А., Понасенко А. В. Генетические предикторы спорадических врожденных пороков сердца у детей. *Молекулярная медицина.* 2022;20(1):53–58. <https://doi.org/10.29296/24999490-2022-01-09>, <https://elibrary.ru/mtweuj>
- Shabaldin AV, Tsepokina AV, Shmulevich SA, Ponasenko AV. Genetic predictors of sporadic congenital heart defects in children. *Molecular Medicine.* 2022;20(1):53–58. (In Russ.) <https://doi.org/10.29296/24999490-2022-01-09>, <https://elibrary.ru/mtweuj>
18. Шабалдин А. В., Цепочкина А. В., Долгих О. В. и др. Комбинация аллелей HLA-DRB1 как фактор, вызывающий риски спорадических врожденных пороков сердца и врожденных пороков развития без хромосомных заболеваний. *Анализ риска для здоровья.* 2021;1:133–142. <https://doi.org/10.21668/health.risk/2021.1.14.eng>
- Shabaldin AV, Tsepokina AV, Dolgikh OV, et al. Combination of HLA-DRB1 alleles as a factor causing risks of sporadic congenital heart defects and congenital malformations without chromosome diseases. *Health Risk Analysis.* 2021;1:133–142. (In Russ.) <https://doi.org/10.21668/health.risk/2021.1.14.eng>

19. Шабалдин А. В., Шмулевич С. А., Сеницкая А. В., Замараев Р. Ю. Роль межгенных взаимодействий между генами цитокинов и Toll-подобных рецепторов в этиологии врожденных пороков сердца. *Кардиоваскулярная терапия и профилактика*. 2023;22(2):3368. <https://doi.org/10.15829/1728-8800-2023-3368>
- Shabaldin AV, Shmulevich SA, Sinitskaya AV, Zamaraev RYu. Role of intergenic interactions between cytokine and Toll-like receptor genes in the etiology of congenital heart defects. *Cardiovascular Therapy and Prevention*. 2023;22(2):3368. (In Russ.) <https://doi.org/10.15829/1728-8800-2023-3368>
20. Электронный регистр пациентов с врожденным пороком сердца. Свидетельство о регистрации программы для ЭВМ RU 2015660991. Заявка № 2015617738 от 24.08.2015: опубликовано 14.10.2015 / Игишева Л. Н., Шабалдин А. В., Шмулевич С. А., Джун И. Е.
21. Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников. М.: Физматлит; 2006. 816 с.
- Kobzar AI. Applied mathematical statistics. For Engineers and Scientists. M.: Fizmatlit; 2006. 816 p. (In Russ.)
22. Breiman L. Random forests. *Machine Learning*. 2001;45:5–32. <https://doi.org/10.1023/A:1010933404324>.
23. Vinge R, Mckelvey T. Understanding support vector machines with polynomial kernels // 27th European Signal Processing Conference (EUSIPCO), A Coruna, Spain, 02–06 September 2019. IEEE. 2023;230–231. <https://doi.org/10.23919/EUSIPCO.2019.8903042>
24. Yoshikawa Yu, Tomoharu I. Gaussian process regression with interpretable sample-wise feature weights. *IEEE transactions on neural networks and learning systems*. 2021;1–15. <https://doi.org/10.1109/TNNLS.2021.3131234>
25. Гольдберг Й. Нейросетевые методы в обработке естественного языка. М.: ДМК Пресс; 2019. 282 с.
- Goldberg Y. Neural network methods in natural language processing. M.: DMK Press; 2019. 282 p. (In Russ.)
26. Игишева Л. Н., Аникеенко А. А., Шмулевич С. А., Сизова И. Н. К вопросу комплексной оценки здоровья детей, оперированных по поводу врожденных пороков сердца. *Комплексные проблемы сердечно-сосудистых заболеваний*. 2019;8(4S):42–50. <https://doi.org/10.17802/2306-1278-2019-8-4S-42-50>
- Igisheva LN, Anikeenko AA, Shmulevich SA, Sizova IN. To the issue of the health comprehensive assessment of children who underwent surgical correction of congenital heart diseases. *Complex Issues of Cardiovascular Diseases*. 2019;8(4S):42–50. (In Russ.) <https://doi.org/10.17802/2306-1278-2019-8-4S-42-50>
27. Atasoy Karakas L, Tugrul D, Sahin Uysal N, et al. Associations between IL-1 α , IL-1 β , TNF α , and IL-6 variations, and susceptibility to transposition of the great arteries. *BMC Cardiovasc Disord*. 2022;22(1):229. <https://doi.org/10.1186/s12872-022-02670-1>
28. Noor Islam D, Auliahadi S. Polymorphisms in congenital heart disease and extracardiac disorders. *Clin Chim Acta*. 2026;579:120693. <https://doi.org/10.1016/j.cca.2025.120693>
29. Feng Y, Ye D, Wang Z, et al. The role of interleukin-6 family members in cardiovascular diseases. *Front Cardiovasc Med*. 2022;9:818890. <https://doi.org/10.3389/fcvm.2022.818890>
30. Clausen H, Friberg E, Lantering K, et al. Newborn screening for high-risk congenital heart disease by dried blood spot biomarker analysis. *JAMA Netw Open*. 2024;7(6):e2418097. <https://doi.org/10.1001/jamanetworkopen.2024.18097>

Информация об авторах:

Шабалдин Андрей Владимирович — доктор медицинских наук, профессор, ведущий научный сотрудник лаборатории пороков сердца ФГБНУ «НИИ комплексных проблем сердечно-сосудистых заболеваний» Минобрнауки России, Кемерово, Россия; старший научный сотрудник лаборатории моделирования геоэкологических систем (совместно с ИВЭП СО РАН) ФГБНУ «Федеральный исследовательский центр информационных и вычислительных технологий», Новосибирск, Россия, weit2007@yandex.ru, <https://orcid.org/0000-0002-8785-7896>;

Замараев Роман Юрьевич — кандидат технических наук, старший научный сотрудник лаборатории геоинформационного моделирования ФГБНУ «Федеральный исследовательский центр информационных и вычислительных технологий», Новосибирск, Россия, zamaraev@ict.sbras.ru, <https://orcid.org/0000-0003-4822-4794>;

Сеницкая Анна Викторовна — кандидат биологических наук, старший научный сотрудник лаборатории геномной медицины ФГБНУ «НИИ комплексных проблем сердечно-сосудистых заболеваний» Минобрнауки России, Кемерово, Россия, annacerokina@mail.ru, <https://orcid.org/0000-0002-4467-8732>;

Халивопуло Иван Константинович — кандидат медицинских наук, заведующий отделением кардиохирургии № 2 ФГБНУ «НИИ комплексных проблем сердечно-сосудистых заболеваний» Минобрнауки России, Кемерово, Россия, halivopulo@mail.ru;

Кобзев Максим Валерьевич — клинический ординатор по детской кардиологии ФГБНУ «НИИ комплексных проблем сердечно-сосудистых заболеваний» Минобрнауки России, Кемерово, Россия, kobzev.maksim@gmail.com;

Шкитин Ян Владимирович — клинический ординатор ФГБОУ ВО «Кемеровский государственный медицинский университет» Минздрава России, Кемерово, Россия, yanshkitin1998@mail.ru;

Шабалдина Елена Викторовна — доктор медицинских наук, доцент, заведующая кафедрой оториноларингологии ФГБОУ ВО «Кемеровский государственный медицинский университет» Минздрава России, Кемерово, Россия, weit@ya.ru, <https://orcid.org/0000-0002-3002-2863>;

Игишева Людмила Николаевна — доктор медицинских наук, доцент, ведущий научный сотрудник лаборатории пороков сердца ФГБНУ «НИИ комплексных проблем сердечно-сосудистых заболеваний» Минобрнауки России, Кемерово, Россия, igisheva@kemcardio.ru, <https://orcid.org/0000-0002-7102-3571>.

Вклад авторов:

Замараев Р. Ю. — концепция и дизайн работы, написание статьи, статистическая обработка, исправление статьи, утверждение окончательного варианта статьи; Шабалдин А. В. — концепция и дизайн работы, написание статьи, статистическая обработка, утверждение окончательного варианта статьи; Игишева Л. Н. — концепция и дизайн работы, исправление статьи, утверждение окончательного варианта статьи; Сеницкая А. В. — сбор и анализ данных, иммуногенетические исследования, утверждение окончательного варианта статьи; Халивопуло И. К. — сбор и анализ данных, иммуногенетические исследования, утверждение окончательного варианта статьи; Кобзев М. В. — сбор и анализ данных, иммуногенетические исследования, утверждение окончательного варианта статьи; Шкитин Я. В. — сбор и анализ данных, иммуногенетические исследования, утверждение окончательного варианта статьи; Шабалдина Е. В. — сбор и анализ данных, иммуногенетические исследования, утверждение окончательного варианта статьи.

Authors information:

Andrey V. Shabaldin, MD, DSc, Professor, Leading researcher at the Laboratory of Heart Diseases of the Federal State Budgetary Scientific Institution Research Institute of Complex Problems of Cardiovascular Diseases of the Ministry of Science and Higher Education of the Russian Federation, Kemerovo, Russia; senior researcher at the Laboratory for Modeling Geo-Ecological Systems (with the IVEP SB RAS) Federal State Budgetary Institution Federal Research Center for Information and Computing Technologies, Novosibirsk, Russia, weit2007@yandex.ru, <https://orcid.org/0000-0002-8785-7896>;

Roman Yu. Zamaraev, PhD in Technical Sciences, Senior Researcher at the laboratory of geoinformation modeling of the Federal State Budgetary Institution Federal Research Center for Information and Computing Technologies, Novosibirsk, Russia, zamaraev@ict.sbras.ru, <https://orcid.org/0000-0003-4822-4794>;

Anna V. Sinitskaya, PhD in Biological Sciences, Researcher at the laboratory of genomic medicine of the Federal State Budgetary Scientific Institution Research Institute of Complex Problems of Cardiovascular Diseases of the Ministry of Science and Higher Education of the Russian Federation, Kemerovo, Russia, annacepokina@mail.ru, <https://orcid.org/0000-0002-4467-8732>;

Ivan K. Khalivopulo, MD, PhD, Head of the department of cardiac surgery No. 2 of the Federal State Budgetary Institution Scientific Research Institute of Complex Problems of Cardiovascular Diseases of the Ministry of Science and Higher Education of the Russian Federation, Kemerovo, Russia, halivopulo@mail.ru;

Maksim V. Kobzev, Resident in pediatric cardiology the Federal State Budgetary Institution Scientific Research Institute of Complex Problems of Cardiovascular Diseases of the Ministry of Science and Higher Education of the Russian Federation, Kemerovo, Russia, kobzev.maksimus@gmail.com;

Yan V. Shkitin, Clinical resident of the Federal State Budgetary Educational Institution of Higher Education “Kemerovo State Medical University” of the Ministry of Health of the Russian Federation, Kemerovo, Russia, yanshkitin1998@mail.ru;

Elena V. Shabaldina, MD, DSc, Associate Professor, Head of the department of otorhinolaryngology, Kemerovo State Medical University, Ministry of Health of the Russian Federation, Kemerovo, Russia, weit@ya.ru, <https://orcid.org/0000-0002-3002-2863>;

Lyudmila N. Igisheva, MD, DSc, Associate Professor, Leading researcher at the Laboratory of Heart Defects, Federal State Budgetary Scientific Institution Research Institute for Complex Issues of Cardiovascular Diseases, Ministry of Science and Higher Education of the Russian Federation, Kemerovo, Russia, igisheva@kemcardio.ru, <https://orcid.org/0000-0002-7102-3571>.

Contribution of the authors:

Zamaraev R. Yu. — the concept and work design, article writing, statistical processing, editing, approval of the final version; Shabaldin A. V. — the concept and work design, statistical processing, approval of the final version; Igisheva L. N. — the concept and work design, editing, approval of the final version; Sinitskaya A. V. — data collection and analysis, immunogenetic studies, approval of the final version; Halivopulo I. K. — data collection and analysis, immunogenetic studies, approval of the final version; Kobzev M. V. — data collection and analysis, immunogenetic studies, approval of the final version; Shkitin Ya. V. — data collection and analysis, immunogenetic studies, approval of the final version; Shabaldina E. V. — data collection and analysis, immunogenetic studies, approval of the final version.

Поступила в редакцию / Received: 17.09.2024

Принята к публикации / Accepted: 20.03.2026
