
ГЕНЕТИКА И МОЛЕКУЛЯРНАЯ БИОЛОГИЯ

GENETICS AND MOLECULAR BIOLOGY

УДК 575 + 519.234.6

ИНТЕРПРЕТАЦИЯ ДИНАМИКИ ЭНТРОПИИ ПРИ ПРОВЕДЕНИИ ГЕНЕТИЧЕСКИХ ИССЛЕДОВАНИЙ МЕТОДОМ СНИЖЕНИЯ МУЛЬТИФАКТОРНОЙ РАЗМЕРНОСТИ

В. А. ИВАНЮКОВИЧ¹⁾, С. Б. МЕЛЬНОВ²⁾, МА МИН²⁾

¹⁾Международный государственный экологический институт им. А. Д. Сахарова БГУ,
ул. Долгобродская, 23/1, 220070, г. Минск, Беларусь

²⁾Белорусский государственный университет физической культуры,
пр. Победителей, 105, 220020, г. Минск, Беларусь

Аннотация. Метод снижения мультифакторной размерности (*multifactor dimensionality reduction*, MDR) позволяет с высокой точностью подбирать комбинации предикторов (генов, аллелей), соответствующих тому или иному фенотипу (исходу). Однако граф изменения энтропии и дендрограмма взаимодействия, которые строятся в реализующем этот метод программном приложении MDR (версия 3.0.2), не характеризуют ген-генные взаимо-

Образец цитирования:

Иванюкович ВА, Мельнов СБ, Ма Мин. Интерпретация динамики энтропии при проведении генетических исследований методом снижения мультифакторной размерности. *Экспериментальная биология и биотехнология*. 2025;1:40–46. EDN: PKZDXT

For citation:

Ivaniukovich UA, Melnov SB, Ma Min. Entropy dynamics interpretation in genetic studies with the multifactor dimensionality reduction method. *Experimental Biology and Biotechnology*. 2025;1:40–46. Russian. EDN: PKZDXT

Авторы:

Владимир Александрович Иванюкович – кандидат физико-математических наук, доцент; доцент кафедры информационных технологий в экологии и медицине факультета мониторинга окружающей среды.

Сергей Борисович Мельнов – доктор биологических наук, профессор; профессор кафедры анатомии факультета оздоровительной физической культуры.

Ма Мин – аспирантка кафедры анатомии факультета оздоровительной физической культуры. Научный руководитель – С. Б. Мельнов.

Authors:

Uladzimir A. Ivaniukovich, PhD (physics and mathematics), docent; associate professor at the department of information technologies in ecology and medicine, faculty of environmental monitoring.

u.ivaniukovich@gmail.com

<https://orcid.org/0000-0002-7273-6645>

Sergei B. Melnov, doctor of science (biology), full professor; professor at the department of anatomy, faculty of health-improving physical education.

sbmelnov@gmail.com

<https://orcid.org/0000-0002-9820-4188>

Ma Min, postgraduate student at the department of anatomy, faculty of health-improving physical education.

<https://orcid.org/0009-0005-6003-7766>

действия. На примерах классификации спортсменов по данным их генотипов показано, что приложение строит граф изменения статистической энтропии при понижении размерности, который не соответствует силе и направленности реальных ген-генных взаимодействий, а описывает эффективность свертывания предикторов при объединении двух или более факторов в целях достижения максимальной точности классификации исходов.

Ключевые слова: генетика; моделирование; снижение мультифакторной размерности; ген-генные взаимодействия; генотип; фенотип; энтропия; статистический анализ; классификация.

ENTROPY DYNAMICS INTERPRETATION IN GENETIC STUDIES WITH THE MULTIFACTOR DIMENSIONALITY REDUCTION METHOD

U. A. IVANIUKOVICH^a, S. B. MELNOV^b, MA MIN^b

^aInternational Sakharov Environmental Institute, Belarusian State University,
23/1 Dawgabrodskaja Street, Minsk 220070, Belarus

^bBelarusian State University of Physical Culture, 105 Peramozhcaw Avenue,
Minsk 220020, Belarus

Corresponding author: U. A. Ivaniukovich (u.ivaniukovich@gmail.com)

Abstract. The multifactor dimensionality reduction (MDR) method allows to select the predictor (gene, allele) combinations corresponding to a particular phenotype (outcome) with high precision. However, the entropy change graph and interaction dendrogram constructed in the *MDR* (version 3.0.2) software implementing this method do not characterise gene – gene interactions. Using examples of athlete classification based on their genotypes, it is shown that the software constructs a statistical entropy change graph when reducing dimensionality, which does not correspond to the strength and direction of real gene – gene interactions, but describes the efficiency of predictor convolution when combining two or more factors in order to achieve maximum accuracy in classifying outcomes.

Keywords: genetics; simulation; multifactor dimensionality reduction; gene – gene interactions; genotype; phenotype; entropy; statistical analysis; classification.

Введение

При изучении фенотипов (исходов), связываемых с ген-генными или генно-средовыми взаимодействиями, часто предпринимаются попытки обнаружить и доказать причинно-следственную связь между данными взаимодействиями. Как правило, при этом рассматривается большое количество факторов (генов, локусов, условий окружающей среды и др.), которые могут оказывать воздействие на конечный результат, что делает анализ взаимодействий весьма сложным, но перспективным. Одним из популярных методов, обеспечивающих получение статистически обоснованных выводов о влиянии ген-генных и генно-средовых взаимодействий на исследуемые фенотипы, является метод снижения мультифакторной размерности (*multifactor dimensionality reduction*, MDR). Актуальная версия реализующего этот метод программного приложения *MDR* (версия 3.0.2) с открытым исходным кодом находится в свободном доступе на сайте *SourceForge* (<https://sourceforge.net/projects/mdr/>). Метод MDR – это непараметрический безмодельный метод обнаружения и описания нелинейных взаимодействий разной природы. По своей сути он является альтернативой логистической регрессии и может быть применен для анализа различных многофакторных процессов [1–3]. Данный метод позволяет уменьшить количество исследуемых параметров, статистически оценивая уровни повышения или понижения рисков формирования фенотипов и преобразовывая два или более фактора (признака) в один.

Большинство публикаций, связанных с MDR, посвящены анализу ген-генных и генно-средовых взаимодействий, ассоциированных с заболеваниями. Эффективность этого метода в изучении генной природы заболеваний подтверждена многочисленными прикладными и модельными исследованиями [4–9]. Также он применяется для классификации данных, не связанных с заболеваниями, например при оценке физического потенциала человека на основе его генотипа [10] и других объектов [11]. Подбирая параметры моделей, можно получить высокую точность классификации. Программа находит наилучшую модель, рассчитывает статистические данные, характеризующие ее точность, а также предоставляет графическую визуализацию результатов. Среди них исследователи чаще всего обращают внимание на графические объекты, построенные в разделе «Энтропия». Большинство авторов статей, ссылающихся на графические модели энтропии, на наш взгляд, некорректно их интерпретируют, связывая изменения энтропии с исследуемыми эффектами ген-генных взаимодействий [4; 6; 7; 9; 11].

Целью настоящей работы является демонстрация того, что предлагаемые программным приложением *MDR* (версия 3.0.2) графические модели динамики энтропии не описывают ген-генные и генно-средовые взаимодействия, а отображают изменения статистической энтропии, происходящие при объединении двух или более факторов в целях понижения размерности изучаемого процесса. Для подтверждения такого вывода в статье демонстрируются изменения рассчитанных методом *MDR* энтропий в зависимости от условий классификации и количества измерений.

Материалы и методы исследования

В работе изучалась связь генотипов спортсменов, занимающихся циклическими видами спорта, с уровнем их достижений. Анализировались данные измерений, формирующие генотип спортсмена и содержащие 7 генов (*5HTT*, *ACE*, *AT2R1*, *GPX1*, *MTHFR*, *PPAPGCI*, *PPARA*). Они были собраны в ходе исследований с участием 94 испытуемых, проводившихся на протяжении последних 10 лет на базе Белорусского государственного университета физической культуры (кафедра анатомии), Полесского государственного университета (лаборатория лонгитудинальных исследований) и Международного государственного экологического института имени А. Д. Сахарова. Генотипирование осуществлялось методом ПЦР-анализа по стандартным методикам, описанным ранее¹ [12; 13]. Генотипы спортсменов оценивались в соответствии с достигнутыми результатами (кандидаты в мастера спорта, мастера спорта, мастера спорта международного класса). Предполагалось, что уровень достижений зависит от физиологических особенностей спортсмена, которые связаны с его генотипом. Классификация проводилась методом *MDR* с применением программного приложения *MDR* (версия 3.0.2) [1–3]. Построения моделей, в которых фенотип разделялся на три класса, давали низкую точность, что может быть объяснено недостаточным количеством измерений. Разделение спортсменов на два класса также основывалось на их достижениях, при этом группы формировались по разным правилам.

Правило 1: кандидаты в мастера спорта и мастера спорта – класс 0, мастера спорта международного класса – класс 1.

Правило 2: кандидаты в мастера спорта – класс 0, мастера спорта и мастера спорта международного класса – класс 1.

При моделировании с использованием данных, содержащих 7 генов, количество измерений в группах, сформированных по правилу 1, составляло 54 (класс 0) и 40 (класс 1), а в группах, сформированных по правилу 2, – 29 (класс 0) и 65 (класс 1).

Результаты и их обсуждение

Модели, построенные на основе исследуемых данных, обеспечили высокую точность классификации фенотипов, что подтверждено результатами статистического анализа, реализованного в программном приложении. Так, лучшая модель, построенная для групп, сформированных по правилу 1, характеризуется следующими статистическими показателями: точностью 0,94, чувствительностью 0,98, специфичностью 0,91, сбалансированной точностью (равна половине суммы чувствительности и специфичности) 0,94, отношением шансов 382,20 (доверительный интервал от 42,86 до 407,81), каппой Коэна 0,87, $\chi^2 = 71,86$ ($p < 0,0001$). Для лучшей модели, построенной для групп, сформированных по правилу 2, точность составляет 0,90, чувствительность – 0,86, специфичность – 1,00, сбалансированная точность – 0,93, отношение шансов стремится к бесконечности, каппа Коэна равна 0,79, $\chi^2 = 61,80$ ($p < 0,0001$). В обоих случаях значения согласованности перекрестных проверок (*cross validation consistency*, *CVC*) давали результат 10 из 10.

Приведенные статистические показатели убедительно свидетельствуют о том, что метод *MDR* и реализующее его программное приложение *MDR* (версия 3.0.2) могут быть использованы для прогнозирования достижений спортсменов в конкретных видах спорта по данным их генных профилей.

На рис. 1 и 2 представлены построенные программным приложением графы изменения энтропии для моделей, рассчитанных для групп, сформированных по правилам 1 и 2 соответственно. В интеллектуальном анализе данных важную роль играет изменение количества информации, которое оценивается в мерах энтропии (т. е. какая часть энтропии в статусе «случай – контроль» удаляется при рассмотрении фактора (в нашем случае генотипа)). Эта величина указана в вершинах графа. В методе *MDR* также оценивается количество информации о взаимодействии, полученной при объединении двух генотипов.

¹Лебедь Т. Л., Мельнов С. Б. Молекулярно-генетическое типирование полиморфизмов: генетический прогноз антропометрических характеристик спортсменов-гребцов : метод. рекомендации. Пинск : ПолесГУ, 2016. 25 с. ; Мельнов С. Б., Лебедь Т. Л., Кипень В. Н. Основы молекулярно-генетического анализа // Современные проблемы биохимии. Методы исследований : учеб. пособие / Е. В. Барковский [и др.] ; под ред. проф. А. А. Чиркина. Минск : Выш. шк., 2013. С. 404–438.

Веса ребер показывают силу (значение веса ребра в мерах энтропии) и направленность (знак в весе ребра) взаимодействий. Цвет линии указывает на характер взаимодействия между факторами: красный цвет соответствует выраженному синергизму, оранжевый – умеренному синергизму, коричневый – аддитивному взаимодействию, зеленый – умеренному антагонизму, синий – выраженному антагонизму.

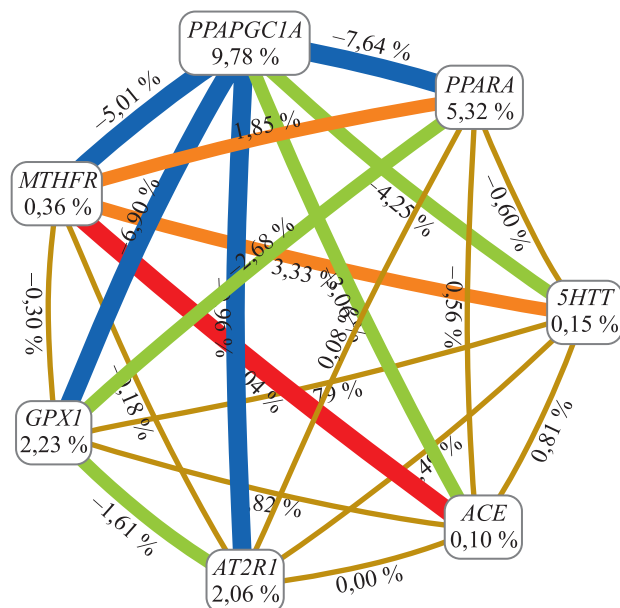


Рис. 1. Граф изменения энтропии для двух классов фенотипов:
класс 0 – кандидаты в мастера спорта и мастера спорта,
класс 1 – мастера спорта международного класса

Fig. 1. Graph of entropy changes for two classes of phenotypes:
class 0 – candidates for master of sports and masters of sports,
class 1 – masters of sports of international class

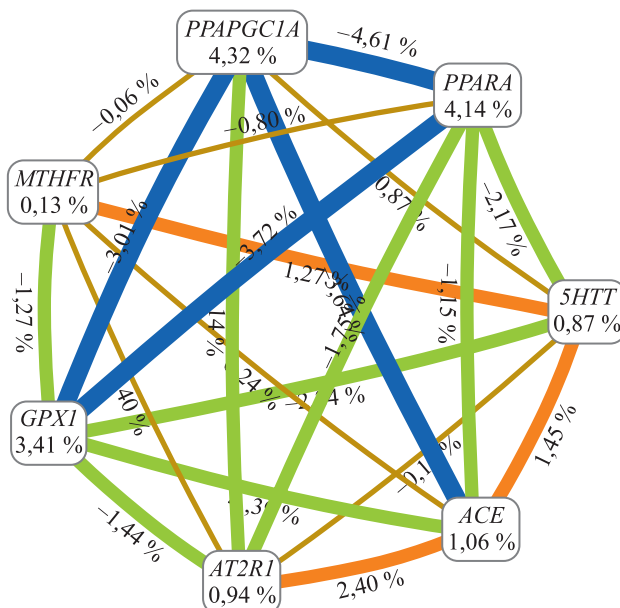


Рис. 2. Граф изменения энтропии для двух классов фенотипов:
класс 0 – кандидаты в мастера спорта,
класс 1 – мастера спорта и мастера спорта международного класса

Fig. 2. Graph of entropy changes for two classes of phenotypes:
class 0 – candidates for master of sports,
class 1 – masters of sports and masters of sports of international class

Представленные выше результаты указывают на существенные различия в характере взаимодействий между рассматриваемыми генами при несущественном изменении правил классификации. Очевидно, что в обеих моделях ген-генные взаимодействия должны быть одинаковыми (или по крайней мере очень близкими).

Наблюдаемые отличия могут быть объяснены только тем, что на графах приведены изменения статистической энтропии, характеризующей изменение количества информации, обусловленное учетом как самих генотипов, так и их взаимодействия. На графах показано, в каком направлении (знак в весе ребра) и с какой силой (значение веса ребра в мерах энтропии) изменится порядок в системе при замене двух взаимодействующих факторов одним. Эти данные не имеют ничего общего с природой ген-генных или генно-средовых взаимодействий, но позволяют с максимальной точностью оценить связь исхода с предикторами. В нашем случае это связь уровня достижений спортсмена (фенотипа) с комбинацией аллелей исследуемых генов (генотипом).

Такой вывод подтверждают и результаты моделирования для групп, сформированных по правилу 1, но при меньшем количестве измерений: случайным образом выбраны около 67 % измерений в классе 0 (использованы 34 измерения) и классе 1 (отобраны 29 измерений). Сформированный программным приложением граф изображен на рис. 3. Здесь различия с рис. 1 еще более значимы, несмотря на то что модели построены на основе данных из одной совокупности, классифицируемых по одинаковому правилу. Такие различия можно объяснить тем, что уменьшение количества обрабатываемых данных приводит к понижению на каждом шаге алгоритма точности выбора комбинаций генов для уменьшения размерности.

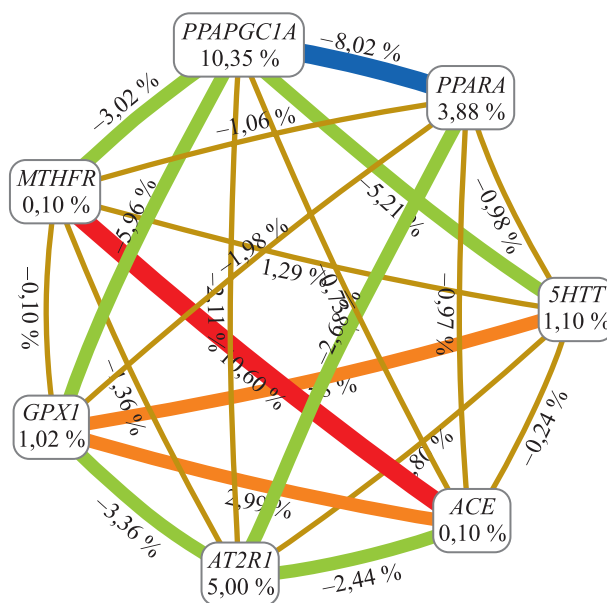


Рис. 3. Граф изменения энтропии, соответствующий условиям рис. 1, но построенный для меньшего количества измерений

Fig. 3. Entropy change graph corresponding to the conditions of fig. 1, but built for fewer dimensions

На основе созданных моделей изменения энтропии в рамках метода MDR формируются пути понижения размерности. Они наглядно представлены на дендрограммах взаимодействия, построенных методом иерархического кластерного анализа. На рис. 4–6 приведены дендрограммы взаимодействия, которые отвечают графам изменения энтропии на рис. 1–3 соответственно и показывают, в какой последовательности пары факторов заменяются одним фактором с максимально подобным эффектом. Чем короче линия, соединяющая два фактора, тем сильнее взаимодействие, проявляющееся в изменении энтропии. Видно, что пути понижения размерности для разных условий моделирования также сильно отличаются.

Подобные результаты получены и при моделировании на основе данных, включающих 11 генов (*5HTT*, *5HT2A*, *ACE*, *AGT_Thr174Met*, *AT2R1*, *BDKRB2*, *CYP1A2*, *GPX1*, *MTHFR*, *PPAPGC1A*, *PPARA*) и 74 испытуемых.

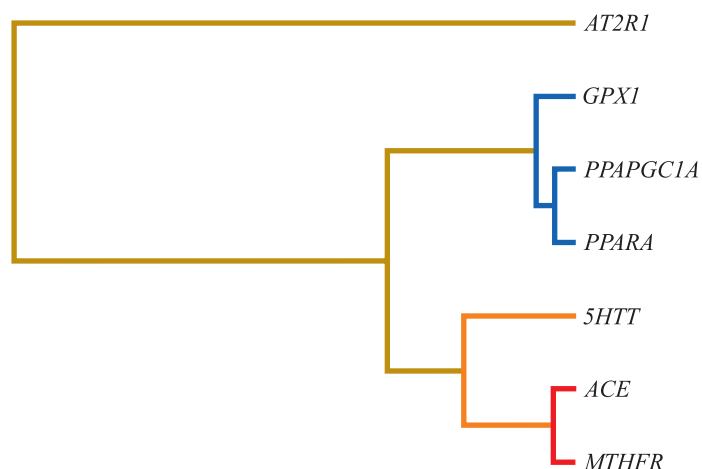


Рис. 4. Дендрограмма взаимодействия, иллюстрирующая последовательность понижения размерности на основе данных модели, изображенной на рис. 1

Fig. 4. Interaction dendrogram illustrating the sequence of dimensionality reduction based on data from the model shown in fig. 1

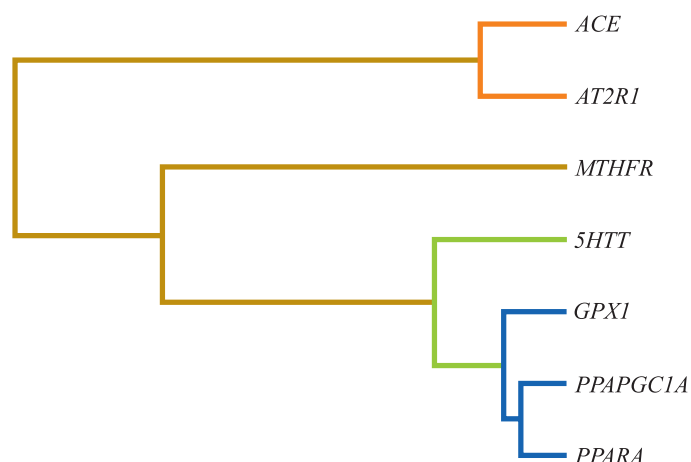


Рис. 5. Дендрограмма взаимодействия, иллюстрирующая последовательность понижения размерности на основе данных модели, изображенной на рис. 2

Fig. 5. Interaction dendrogram illustrating the sequence of dimensionality reduction based on data from the model shown in fig. 2

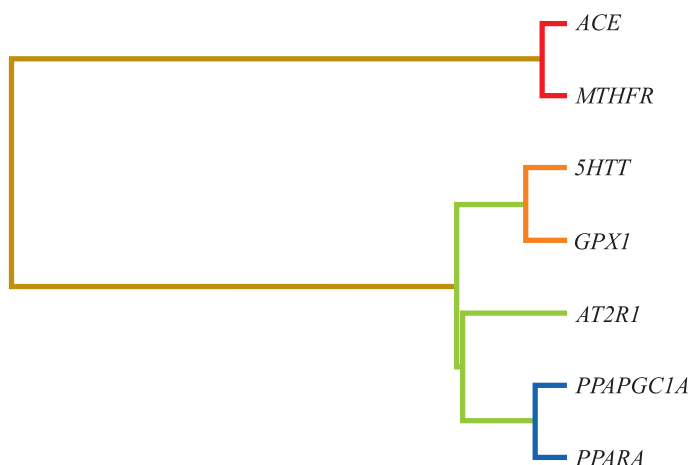


Рис. 6. Дендрограмма взаимодействия, иллюстрирующая последовательность понижения размерности на основе данных модели, изображенной на рис. 3

Fig. 6. Interaction dendrogram illustrating the sequence of dimensionality reduction based on data from the model shown in fig. 3

Заключение

Представленные результаты моделирования с помощью метода MDR демонстрируют, что графы изменения энтропии и дендрограммы взаимодействия, построенные программным приложением *MDR* (версия 3.0.2), не отражают взаимодействие генов, а являются иллюстрацией статистических инструментов, используемых для понижения размерности исследуемых систем в целях проведения классификации. Выводы о ген-генных и генно-средовых взаимодействиях могут быть сделаны на основе иной информации, предоставляемой приложением, например по данным графической модели или по результатам статистических оценок лучшей модели.

Библиографические ссылки

1. Hahn LW, Ritchie MD, Moore JH. Multifactor dimensionality reduction software for detecting gene – gene and gene – environment interactions. *Bioinformatics*. 2003;19(3):376–382. DOI: 10.1093/bioinformatics/btf869.
2. Motsinger AA, Ritchie MD. Multifactor dimensionality reduction: an analysis strategy for modelling and detecting gene – gene interactions in human genetics and pharmacogenomics studies. *Human Genomics*. 2006;2:318. DOI: 10.1186/1479-7364-2-5-318.
3. Yang C-H, Lin Y-D, Yang C-S, Chuang L-Y. An efficiency analysis of high-order combinations of gene – gene interactions using multifactor dimensionality reduction. *BMC Genomics*. 2015;16:489. DOI: 10.1186/s12864-015-1717-8.
4. Пономаренко ИВ. Использование метода *multifactor dimensionality reduction* (MDR) и его модификаций для анализа ген-генных и генно-средовых взаимодействий при генетико-эпидемиологических исследованиях (обзор). *Научные результаты биомедицинских исследований*. 2019;5(1):4–21. DOI: 10.18413/2313-8955-2019-5-1-0-1.
5. Singh SA, Ghosh SK. Polymorphisms of XRCC1 and XRCC2 DNA repair genes and interaction with environmental factors influence the risk of nasopharyngeal carcinoma in Northeast India. *Asian Pacific Journal of Cancer Prevention*. 2016;17(6):2811–2819. DOI: 10.14456/apjcp.2016.13.
6. Liao ZY, Yang S, Hu S, Liu J, Mao YJ, Sun SQ. Association between gene polymorphisms and SNP-SNP interactions of the matrix metalloproteinase 2 signaling pathway and the risk of vascular senescence. *Biomedical and Environmental Sciences*. 2024;37(2):146–156. DOI: 10.3967/bes2024.016.
7. Бочарова АВ, Марусин АВ, Макеева ОА, Жукова ИА, Жукова НГ, Федоренко ОЮ и др. Межгенные взаимодействия при заболеваниях с нарушениями когнитивных функций человека. *Медицинская генетика*. 2022;21(7):24–29. DOI: 10.25557/2073-7998.2022.07.24-29.
8. Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, et al. Multifactor dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *The American Journal of Human Genetics*. 2001;69(1):138–147.
9. Сташкевич ДС, Беляева СВ, Филиппова ЮЮ, Бурмистрова АЛ. Межгенные взаимодействия генов цитокинов *IL1*, *IL1RA*, *IL6*, *IL10* и *TNFA* у больных синдромом раздраженного кишечника и неспецифическим язвенным колитом русской популяции Челябинской области. *Медицинская иммунология*. 2018;20(6):921–926. DOI: 10.15789/1563-0625-2018-6-921-926.
10. Иванюкович ВА, Николаенко ЕА, Мельнов СБ, Жур НВ, Лебедь ТЛ. Влияние классификации исходов на результат моделирования и прогнозирования методом MDR. В: Родькин ОИ, Герменчук МГ, редакторы. *Сахаровские чтения 2023 года: экологические проблемы XXI века. Материалы 23-й Международной научной конференции; 18–19 мая 2023 г.; Минск, Беларусь. Часть 2*. Минск: ИВЦ Минфина; 2023. с. 301–305.
11. Кипень ВН, Михайлова МЕ, Снытков ЕВ, Романишко ЕЛ, Иванова ЕВ, Шейко РИ. Биоинформатический анализ геномов коммерческих пород домашних свиней для идентификации породоспецифичных SNP. *Известия Национальной академии наук Беларуси. Серия аграрных наук*. 2021;59(4):464–476. DOI: 10.29235/1817-7204-2021-59-4-464-476.
12. Козлова АС, Лебедь ТЛ, Малиновская ЮВ, Мельнов СБ. Генетические маркеры сердечно-сосудистой патологии спортсменов спорта высших достижений. *Экологический вестник*. 2014;2:42–49. EDN: QTMWGY.
13. Шепелевич НВ, Лебедь ТЛ, Мельнов СБ. Особенности генетического профиля выносливости у спортсменов-гребцов. *Экологический вестник*. 2013;4:20–24. EDN: HEJUNO.

Получена 29.06.2024 / исправлена 31.10.2024 / принята 01.11.2024.
Received 29.06.2024 / revised 31.10.2024 / accepted 01.11.2024.