

карт (управление слоями, отображение, пространственный анализ и топология и т.д.), в том числе отвечающим нуждам радиозкологии.

Разумеется, использование ГИС в радиационных исследованиях не ограничивается приведенными выше примерами. Арсенал их применения гораздо шире. Тем не менее в общих чертах ГИС предназначены для сбора, хранения, анализа, просмотра графической визуализации пространственных объектов.

Таким образом, эффективность обследования радиационно-экологического состояния региона может быть существенно повышена при использовании ГИС-технологий. Главным достоинством ГИС является не только то, что они позволяют накапливать и отображать получаемые при обследовании данные, но и, главным образом, предоставляют возможность сопоставления и взаимного анализа различных информационных слоев.

#### ЛИТЕРАТУРА

1. *Кащенко Н. А.* Геоинформационные системы / Н. А. Кащенко. – Нижний Новгород: Нижегород. гос. Архитектур.-строит. ун-т, 2012. – 130 с.
2. MapServer, Web-сайт фирмы MapServer [Электронный ресурс]. – Режим доступа: <https://www.mapserver.org/>. – Дата доступа : 25.02.2023.
3. QGIS., Web-сайт фирмы QGIS. [Электронный ресурс]. – Режим доступа: <https://qgis.org/ru/site/forusers/download.html/>. – Дата доступа : 25.02.2023.
4. MapInfo, Web-сайт фирмы MapInfo. [Электронный ресурс]. – Режим доступа: <http://mapinfo.ru/product/mapinfo-professional/>. – Дата доступа : 25.02.2023.
5. *Чунихин Л. А., Чеховский А. Л., Дроздов Д. Н.* Карта радоновой опасности территории Республики Беларусь // Радиационная гигиена. – 2016. – Т. 9, № 4. – С. 43–46.

## ВЛИЯНИЕ КЛАССИФИКАЦИИ ИСХОДОВ НА РЕЗУЛЬТАТ МОДЕЛИРОВАНИЯ И ПРОГНОЗИРОВАНИЯ МЕТОДОМ MDR INFLUENCE OF CLASSIFICATION OF OUTCOMES ON THE RESULT OF SIMULATION AND FORECASTING BY THE MDR METHOD

**В. А. Иванюкович<sup>1</sup>, Е. А. Николаенко<sup>1</sup>, С. Б. Мельнов<sup>2</sup>, Н. В. Жур<sup>3</sup>, Т. Л. Лебедь<sup>3</sup>  
U. Ivaniukovich<sup>1</sup>, K. Nikalayenka<sup>1</sup>, S. Melnov<sup>2</sup>, N. Zhyr<sup>3</sup>, T. Lebedz<sup>3</sup>**

<sup>1</sup>Учреждение образования «Международный государственный экологический институт имени А. Д. Сахарова» Белорусского государственного университета, МГЭИ им. А. Д. Сахарова БГУ  
г. Минск, Республика Беларусь  
[iva@iseu.by](mailto:iva@iseu.by)

<sup>2</sup>Белорусский государственный университет физической культуры,  
г. Минск, Республика Беларусь

<sup>3</sup>Полесский государственный университет  
г. Пинск, Республика Беларусь

<sup>1</sup>*International Sakharov Environmental Institute of Belarusian State University, ISEI BSU,  
Minsk, Republic of Belarus*

<sup>2</sup>*Belarusian State University of Physical Culture, Minsk, the Republic of Belarus*

<sup>3</sup>*Polesky State University, Pinsk, the Republic of Belarus*

Показано влияние правил классификации фенотипов на результат моделирования методом MDR. Проведено моделирование с целью установления связи генотипов спортсменов, занимающихся циклическими видами спорта, с их достижениями. Исследования проводились на данных, содержащих 7 генов 94 испытуемых и 11 генов 74 испытуемых. Небольшое изменение правил классификации приводит к существенному изменению моделью. При этом достигается высокая достоверность прогнозирования. Следовательно, ассоциировать построенные методом MDR модели с природой ген-генных взаимодействий не представляется возможным.

The influence of phenotype classification rules on the result of modeling by the MDR method is shown. Simulations were carried out in order to establish the connection between the genotypes of athletes involved in cyclic sports and their achievements. The studies were conducted on data containing 7 genes from 94 test subjects and 11 genes from 74 test subjects. A small change in the classification rules leads to a significant change in the model. At the same time, high reliability of forecasting is achieved. Therefore, it is not possible to associate the models constructed by the MDR method with the nature of gene-gene interactions.

*Ключевые слова:* Моделирование, метод снижения мультифакторной размерности, программное обеспечение, ген-генные взаимодействия, генотип, фенотип, спортсмены, статистический анализ.

*Keywords:* Modeling, Multifactor Dimensionality Reduction Method, software, gene-gene interactions, genotype, phenotype, athletes, statistical analysis.

<https://doi.org/10.46646/SAKH-2023-2-301-305>

При изучении исходов, связанных с генетическими особенностями организма (заболеваний, неординарных возможностей и других), как правило, оцениваются межгенные взаимодействия и предпринимаются попытки обнаружить и доказать причинно-следственную связь между ними и исследуемой проблемой. При этом рассматривается большое количество генов или локусов, которые могут оказывать влияющее воздействие. Одним из популярных методов, который позволяет получать статистически обоснованные выводы о влиянии ген-генных взаимодействий на исследуемые фенотипы (исходы) является метод снижения мультифакторной размерности MDR (Multifactor Dimensionality Reduction). Версия MDR 3.0.2 программного обеспечения с открытым исходным кодом, реализующего этот метод, доступна на сайте <https://sourceforge.net/projects/mdr/>. Это непараметрический интеллектуальный метод обнаружения и описания нелинейных взаимодействий различной природы. По своей сути он является альтернативой логистической регрессии и может быть применен для анализа различных многофакторных процессов [1]. MDR позволяет уменьшить количество исследуемых параметров, оценивая уровни повышенного или пониженного рисков формирования фенотипов и преобразовывая два или более признака в один.

Большое количество публикаций посвящено исследованию ген-генных и генно-средовых взаимодействий, ассоциированных с заболеваниями. Эффективность данного метода в изучении генной природы заболеваний подтверждена многочисленными прикладными и модельными исследованиями [2]. Целью данной работы является изучение влияния классификации исходов на формирование модели, пригодной для прогнозирования развития изучаемого фенотипа. Были исследованы локусы генных карт спортсменов, занимающихся циклическими видами спорта и достигших высоких результатов (кандидаты в мастера спорта, мастера спорта, мастера спорта международного класса), и контрольной группы. Анализировались результаты измерений, содержащие 7 и 11 генов, формирующих генотип спортсмена. Предполагалось, что генотип ассоциируется с результатом, достигнутым спортсменом. Были предприняты попытки связать комбинации аллелей с перспективностью спортсмена, занимающегося конкретным видом спорта и, при необходимости, рекомендовать ему занятие другими видами спорта, в которых его результаты могут быть более высокими.

Предполагалось, что достигнутые спортсменом результаты ассоциированы с его генотипом, который определяет физические возможности человека. Классификация спортсменов на два класса основывалась на их достижениях, при этом группы формировались по-разному:

- a. кандидаты в мастера спорта – 0, мастера спорта и мастера спорта международного класса – 1;
- b. кандидаты в мастера спорта и мастера спорта – 0, мастера спорта международного класса – 1;
- c. контрольная группа – 0, спортсмены – 1.

Результаты моделирования оценивались по данным статистического анализа, реализованного в программе MDR. Для визуализации моделей использовались дендограммы и графы, показывающие характер и силу ген-генных или генно-средовых взаимодействий. Красные линии соответствуют выраженному синергизму, оранжевые – умеренному синергизму, коричневые – аддитивному взаимодействию, зеленые – умеренному антагонизму и синие – выраженному антагонизму взаимодействия. Сила взаимодействия описывается долей вклада в энтропию и направлением (знак + или –). Качество моделей оценивалось по статистическим показателям полного набора данных и по значениям согласованности перекрестных проверок (Cross Validation Consistency – CVC) по 10 в каждой модели.

Кроме прогнозирования перспективности спортсмена, представлялось интересным изучить устойчивость MDR-моделей к незначительным изменениям в формировании групп или, другими словами, классификации исходов в исследовании. В работе сравнивались модели ген-генных взаимодействий, которые ассоциировались с физическими возможностями спортсмена в конкретном виде спорта. На рис. 1 показаны дендограмма и граф ген-генных взаимодействий в модели, построенной методом MDR по 7 генам из генных карт 94 спортсменов мужского и женского пола. В одну группу (класс 0) вошли кандидаты в мастера спорта и мастера спорта (54 спортсмена), во вторую (класс 1) – мастера спорта международного класса (40 спортсменов).

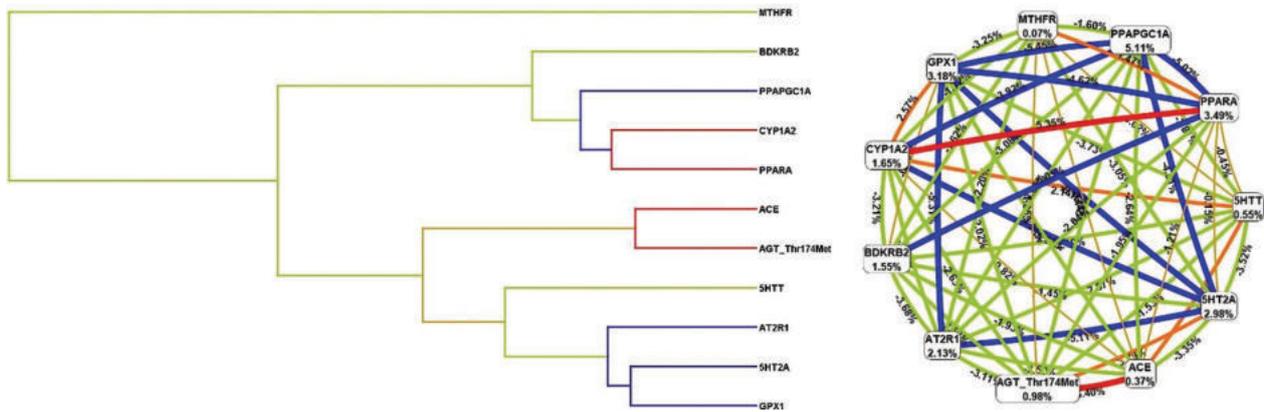


Рисунок 1 – Дендограмма и граф ген-генных взаимодействий в модели, построенной методом MDR при классификации 0 – кандидаты в мастера спорта и мастера спорта, 1 – мастера спорта международного класса (7 генов, 94 измерения). Характер ген-генного взаимодействия при формировании фенотипа, влияющего на физические возможности спортсмена, показан цветом линий: красный соответствует выраженному синергизму, оранжевый – умеренному синергизму, коричневый – аддитивному взаимодействию, зеленый – умеренному антагонизму и синий – выраженному антагонизму. Сила и направленность ген-генного взаимодействия показаны в долях энтропии

Определены наборы аллелей для моделей, построенных на различных количествах ген, которые дают достоверные прогнозы. Построенные на одном и двух генах модели достоверно не прогнозируют класс. Лучшая модель на трех атрибутах построена на генах SHTT, PPAPGC1A, PPARA и характеризуется отношением шансов  $OR=\infty$ ,  $\chi^2 = 10,9$  ( $p=0,001$ ), Карра = 0,84. Для предсказания класса могут быть использованы две комбинации аллелей: класс 0 – SS, GlySer, GC и класс 1 – LL, GlyGly, GG.

Лучшая модель на четырех атрибутах построена на генах ACE, AT2R1, GPX1, PPARA и характеризуется отношением шансов  $OR=\infty$ ,  $\chi^2 = 7,54$  ( $p=0,006$ ), Карра = 0,81. Для предсказания класса могут быть использованы две комбинации аллелей: класс 0 – ID, AA, CT, GC и класс 1 – DD, AC, CC, GG.

Результаты, полученные на моделях с 5–7 генами, не позволили проводить прогнозирование.

На рис. 2 показаны дендограмма и граф ген-генных взаимодействий, построенные на той же группе испытуемых с таким же набором генов, но классифицированных по-другому. В одну группу (класс 0) вошли кандидаты в мастера спорта (29 спортсменов), во вторую (класс 1) – мастера спорта и мастера спорта международного класса (75 спортсменов).

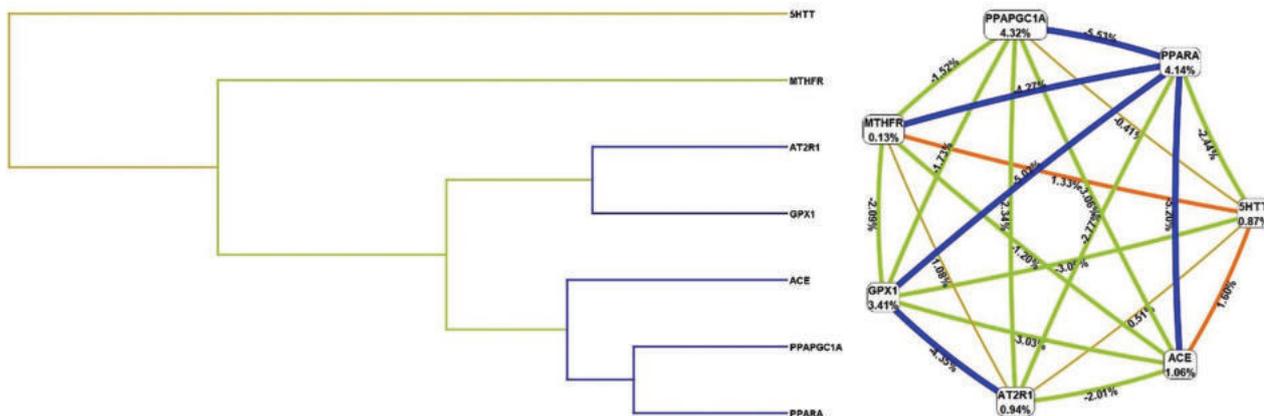


Рисунок 2 – Дендограмма и граф ген-генных взаимодействий в модели, построенной методом MDR при классификации 0 – кандидаты в мастера спорта, 1 – мастера спорта и мастера спорта международного класса (7 генов, 94 измерения). Цвет линий соответствует цветам на рис. 1

Лучшая модель на трех атрибутах построена на генах SHTT, ACE, PPARA и характеризуется отношением шансов  $OR=60,0$  (доверительный интервал 3,03, 1185),  $\chi^2 = 11,42$  ( $p=0,0007$ ), Карра = 0,74. Для предсказания класса могут быть использованы три комбинации аллелей: класс 0 – LL, II, GG и класс 1 – LS, DD, GG и LL, DD, GG.

Лучшая модель на четырех атрибутах построена на генах ACE, AT2R1, GPX1, PPARA и характеризуется отношением шансов  $OR=\infty$ ,  $\chi^2 = 11,73$  ( $p=0,006$ ), Карра = 0,84. Для предсказания класса так же могут быть использованы три комбинации аллелей: класс 0 – LL, II, AC, GG и класс 1 – LL, DD, AC, GG и LL, DD, AA, GG.

Результаты, полученные на остальных моделях, как и в предыдущем случае, не позволили проводить прогнозирование.

Подобные результаты получены и при моделировании на основе 11 генов из генных карт 74 спортсменов, занимающихся тем же видом спорта. Данные, представленные на рис. 3, соответствуют классификации, когда в одну группу отобраны кандидаты в мастера спорта и мастера спорта (41 спортсмен, класс 0), во вторую (класс 1) – мастера спорта международного класса (33 спортсмена). При таком наборе данных метод успешно проводит классификацию на модели, включающей два гена при отношении шансов  $OR=\infty$ ,  $\chi^2 = 10,5$  ( $p=0,0012$ ), Карра = 0,86. Модель построена на генах ACE, AGT\_Thr174Met. Для предсказания класса могут быть использованы две комбинации аллелей: класс 0 – II, CC и класс 1 – II, CT.

Лучшая модель на трех атрибутах построена на генах 5HT2A, ACE, AGT\_Thr174Met и характеризуется отношением шансов  $OR=\infty$ ,  $\chi^2 = 15,28$  ( $p<0,0001$ ), Карра = 0,82. Для предсказания класса могут быть использованы четыре комбинации аллелей: класс 0 – TT, ID, CC и TT, II, CC, класс 1 – TC, ID, CC и TC, II, CT.

Лучшая модель на четырех атрибутах построена на генах AGT\_Thr174Met, BDKRB2, CYP1A2, PPARA и характеризуется отношением шансов  $OR=\infty$ ,  $\chi^2 = 14,19$  ( $p=0,0002$ ), Карра = 0,86. Для предсказания класса могут быть использованы четыре комбинации аллелей: класс 0 – CC, +9/-9, AC, GG и CC, +9/-9, AA, GC и CC, +9/+9, AA, GG, класс 1 – CC, +9/-9, CC, GG.

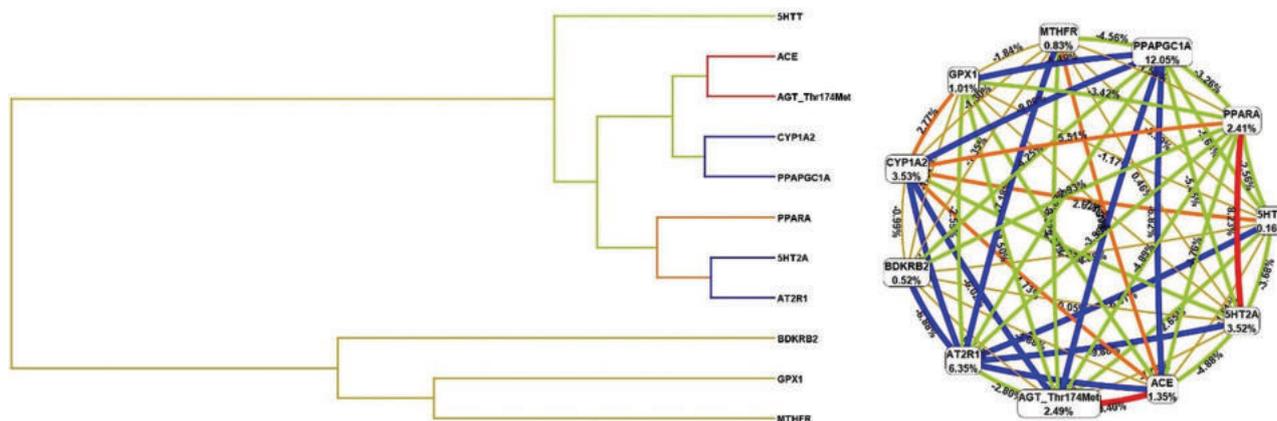


Рисунок 3 – Дендограмма и граф ген-генных взаимодействий в модели, построенной методом MDR при классификации 0 – кандидаты в мастера спорта и мастера спорта, 1 – мастера спорта международного класса (11 генов, 74 измерения). Цвет линий соответствует цветам на рис. 1

Лучшая модель на пяти атрибутах построена на генах 5HTT, 5HT2A, GPX1, MTHFR, PPARA и характеризуется отношением шансов  $OR=\infty$ ,  $\chi^2 = 10,0$  ( $p=0,0016$ ), Карра = 1,0. Для предсказания класса могут быть использованы две комбинации аллелей: класс 0 – SS, TT, CT, CT, GlySer, класс 1 – LL, TC, CT, CT, GlySer.

Лучшая модель на пяти атрибутах построена на генах 5HTT, 5HT2A, GPX1, MTHFR, PPARA и характеризуется отношением шансов  $OR=\infty$ ,  $\chi^2 = 10,0$  ( $p=0,0016$ ), Карра = 1,0. Для предсказания класса могут быть использованы две комбинации аллелей: класс 0 – SS, TT, CT, CT, GlySer, класс 1 – LL, TC, CT, CT, GlySer.

Лучшая модель на шести атрибутах построена на генах 5HT2A, BDKRB2, CYP1A2, MTHFR, PPARA и характеризуется отношением шансов  $OR=\infty$ ,  $\chi^2 = 6,67$  ( $p=0,0098$ ), Карра = 0,8. Для предсказания класса могут быть использованы две комбинации аллелей: класс 0 – TC, +9/-9, AC, CT, GlySer, GG, класс 1 – TC, +9/-9, AA, CT, GlySer, GG.

Результаты, полученные на моделях с большим количеством генов, не позволили проводить прогнозирование.

Результаты моделирования, проведенного на той же группе испытуемых с таким же набором генов, но классифицированных по-другому, представлены на рис. 4. В одну группу (класс 0) вошли кандидаты в мастера спорта (22 спортсмена), во вторую (класс 1) – мастера спорта и мастера спорта международного класса (52 спортсмена).

Лучшая модель на двух генах CYP1A2, PPARA характеризуется параметрами: отношение шансов  $OR=55,0$  (доверительный интервал 2,83, 1068,42),  $\chi^2 = 10,125$  ( $p=0,0015$ ), Карра = 0,75. Для предсказания класса могут быть использованы две комбинации аллелей: класс 0 – AA, GC и класс 1 – CC, GG.

Лучшая модель на трех атрибутах построена на генах 5HTT, 5HT2A, AGT\_Thr174Met и характеризуется отношением шансов  $OR=\infty$ ,  $\chi^2 = 13,84$  ( $p=0,0002$ ), Карра = 0,73. Для предсказания класса могут быть использованы три комбинации аллелей: класс 0 – LL, TT, CC, класс 1 – LL, TC, CC и LL, TT, CT.

Лучшая модель на четырех атрибутах построена на генах 5HTT, 5HT2A, AGT\_Thr174Met, AT2R1 и характеризуется отношением шансов  $OR=\infty$ ,  $\chi^2 = 16,62$  ( $p<0,0001$ ), Карра = 0,86. Для предсказания класса могут быть использованы пять комбинаций аллелей: класс 0 – LL, TT, CC, AC и четыре комбинации для класса 1 – (SS, TC, CC, AC), (SS, TT, CC, AA), (LL, TC, CC, AC) и (LL, TC, CC, AA).

Лучшая модель на пяти атрибутах построена на генах ACE, BDKRB2, CYP1A2, PPARA и характеризуется отношением шансов  $OR=\infty$ ,  $\chi^2 = 6,67$  ( $p=0,0098$ ), Карра = 0,8. Для предсказания класса

могут быть использованы две комбинации аллелей: класс 0 – DD, +9/-9, AC, Gly/Ser, GG, класс 1 – DD, +9/-9, AA, Gly/Ser, GG.

Лучшая модель на шести атрибутах построена на генах 5HTT, ACE, AGT\_Thr174Met, BDKRB2, PPAPGC1A, PPARA и характеризуется отношением шансов  $OR=\infty$ ,  $\chi^2 = 6,67$  ( $p=0,0098$ ), Кappa = 0,8. Для предсказания класса могут быть использованы две комбинации аллелей: класс 0 – SS, DD, CC, +9/-9, Gly/Ser, GG, класс 1 – LL, DD, CC, +9/-9, Gly/Ser, GG.

Результаты, полученные на моделях с большим количеством генов, не позволили проводить прогнозирование.

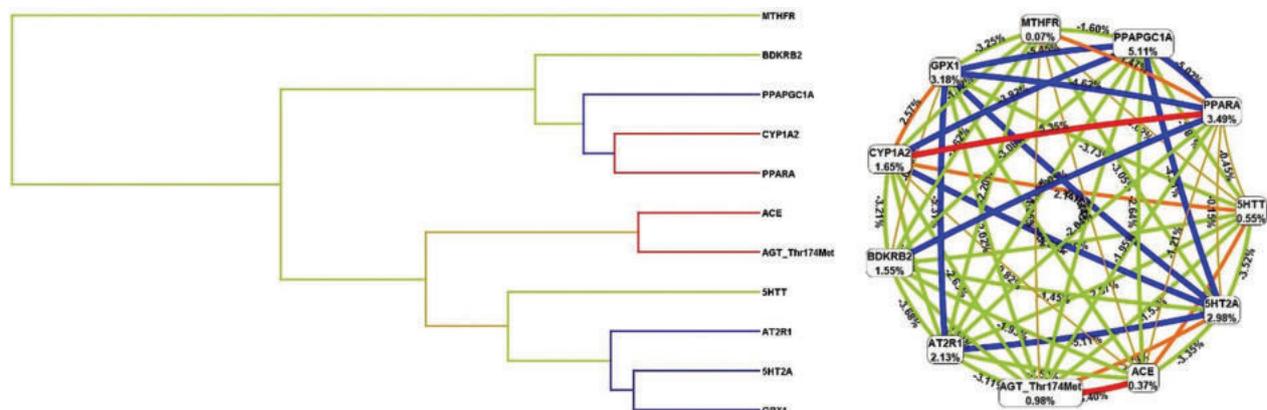


Рисунок 4 – Дендограмма и граф ген-генных взаимодействий в модели, построенной методом MDR при классификации 0 – кандидаты в мастера спорта, 1 – мастера спорта и мастера спорта международного класса (11 генов, 74 измерения). Цвет линий соответствует цветам на рис. 1

Полученные результаты наглядно демонстрируют, что незначительные изменения в классификации исходов (фенотипов) приводят к существенным изменениям статистической модели прогнозирования, создаваемой методом MDR. По-разному выглядят не только дендограмма и граф ген-генных взаимодействий, но и наборы генов в моделях, и комбинации аллелей, которые могут быть использованы для прогнозирования. При этом различные модели позволяют получать прогноз с высокой точностью, что подтверждается статистическими характеристиками. Однако ассоциировать построенные в нашем исследовании методом MDR модели с природой ген-генных взаимодействий не представляется корректным. Возможно, модели будут соответствовать природе ген-генных взаимодействий при условии выбора не только размера и корректного ранжирования групп, но и подбора комплекса генов, определяющих работу одной цепи биохимических реакций (например, РААС, нейромедиаторные, углеводный, липидный обмен и т.д.).

#### ЛИТЕРАТУРА

1. Multifactor Dimensionality Reduction [Электронный ресурс] 24.12.2014. – Режим доступа: <https://sourceforge.net/projects/mdr/>. – Дата доступа: 27.02.2023.
2. Пономарева И.В. Использование метода Multifactor Dimensionality Reduction (MDR) и его модификаций для анализа ген-генных и генно-средовых взаимодействий при генетико-эпидемиологических исследованиях (обзор). / И.В. Пономарева // Научные результаты биомедицинских исследований. 2019. Т. 5, № 1. С. 4-21. DOI: 10.18413/2313-8955-2019-5-1-0-1, URL: <http://rrmedicine.ru/journal/article/1605>.