

© КОЛЛЕКТИВ АВТОРОВ, 2020

УДК 616.12-008.331.1:616.133:575-07

**В. В. Бенберин¹, Т. А. Вощенкова¹, Р. Ж. Карабаева¹, А. С. Сибгатов¹, Д. Б. Бабенко²,
А. А. Турмухамбетова²**

ДИАГНОСТИЧЕСКИЕ ПОКАЗАТЕЛИ ГЕНЕТИЧЕСКИХ ФАКТОРОВ В СПОСОБНОСТИ КЛАССИФИЦИРОВАТЬ ПАЦИЕНТОВ С РЕМОДЕЛИРОВАНИЕМ СОННЫХ АРТЕРИЙ ОТ УСЛОВНО ЗДОРОВЫХ НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ

¹РГП «Больница Медицинского центра Управления делами Президента РК» (Нур-Султан, Казахстан);

²НАО «Медицинский университет Караганды» (Караганда, Казахстан)

Получена модель для дифференцирования пациентов с ремоделированием сонных артерий и условно здоровых людей на основе машинного обучения с использованием генетических данных казахской популяции.

За период 2018-2020 гг. рекрутированы 561 участник исследования, из которых 357 были с ремоделированием сонных артерий, 204 – условно здоровые. Генотипирование осуществлено с помощью технологии QuantStudio TM 12K Flex Real-Time PCR (Applied Biosystems) с использованием панели из 118 полиморфизмов. После процедуры контроля качества 43 полиморфизма были исключены из дальнейшего анализа.

Результаты генотипирования 75 полиморфизмов были использованы для построения бинарной классификационной модели в виде дерева решений на основе Fast-and-frugal trees (FFTs) алгоритма. Наилучшая модель из 5 полиморфизмов обладала 52% чувствительностью, 64% специфичностью и 56% точностью в способности проводить разделение между пациентами с РСА и условно-здоровыми людьми на основе генетических данных.

Ключевые слова: ремоделирование сонных артерий, машинное обучение, классификационная модель, дерево решений

Сердечно-сосудистые заболевания (ССЗ) являются одной из главных проблем здравоохранения в Республике Казахстан. В 2018 г. доля смертности от ССЗ составляла 23,4% (167 случаев на 100 тыс. при общей смертности 713,75 на 100 тыс.), обеспечивая лидирующую позицию в структуре смертности [3]. Следует отметить, что гипертония является одним из ведущих заболеваний в структуре ССЗ с тенденцией к увеличению (855,6 случаев в 2008 г. и 1 296,5 случаев в 2018 г.). В 2015 г. порядка 27% казахстанцев страдало повышенным артериальным давлением [2]. Артериальная гипертензия (АГ) приводит к поражению органов-мишеней, включая сосуды и миокард, и как следствие может развиваться ремоделирование сонных артерий (РСА) и ремоделирование миокарда (РМ). В то же время РСА и РМ являются факторами повышенного кардиоваскулярного риска [5].

Для оценки изменений состояния сосудистых структур наиболее безопасными и доступными являются ультразвуковые методы [1]. Эффективно для визуализации сосудов и определения их структурных поражений триплексное цветовое сканирование [4]. Данные подходы активно используются в диагностике РСА.

В последние годы активно используются генетические исследования для понимания роли наследственного компонента в развитии заболеваний. Наиболее популярными являются полногеномные исследования (GWAS), и

существуют работы, в которых удалось определить ассоциацию генетических факторов с РСА на основе размера стенки сосудов. В исследовании N. Franceschini et al. (Nat Commun, 2018) определены 6 генетических полиморфизмов (rs13225723, rs2912063, rs11196033, rs148147734, rs200482500, rs139302128), которые связаны с утолщением интимы сонной артерии и 7 полиморфизмов (rs201648240, rs224904, rs6907215, rs11785239, rs844396, rs7412, rs515135), ассоциированных с уменьшением толщины интимы сонной артерии [6].

Цель работы – изучить диагностический вклад генетических факторов в способности определить ремоделирование сонных артерий на основе машинного обучения.

МАТЕРИАЛЫ И МЕТОДЫ

Отбор пациентов осуществлялся на базе Больницы Медицинского центра Управления делами Президента РК. Исследование проводилось в соответствии с этическими нормами (разрешение Комитета по биоэтике КГМУ №305 от 19 мая 2017 г.). Все участники исследования согласились участвовать в исследовании добровольно и подписали соответствующее информированное согласие.

Диагноз ремоделирования сонных артерий ставился на основании результатов цветного дуплексного сканирования, проводимого на сердечно-сосудистом ультразвуковом сканере Vivid E9 от GE Healthcare's (США). Значения толщины комплекса интимы-медиа сон-

ных артерий более 0,9 мм были приняты за PCA.

Выделение ДНК и генотипирование. Выделение ДНК из биообразцов крови проводилось с использованием набора реактивов PurLink Genomic DNA Mini Kit (Invitrogen, USA) в соответствии с протоколом производителя. Генотипирование образцов проводилось с использованием чипа на 118 однонуклеотидных полиморфизмах и технологии QuantStudio TM 12K Flex Real-Time PCR (Applied Biosystems).

Анализ первичных данных, полученных с прибора QuantStudio 12K, осуществлялся с помощью облачного сервиса ThermoFisher (<https://apps.thermofisher.com>) и с использованием R statistics.

Статистический анализ. Анализ результатов генотипирования проводился в среде R с использованием пакетов CompareGroups, FFTrees с использованием параметров по умолчанию. Оценка на нормальность распределения проводилась как графически, так и с помощью метода Шапиро – Уилкса (рис. 1).

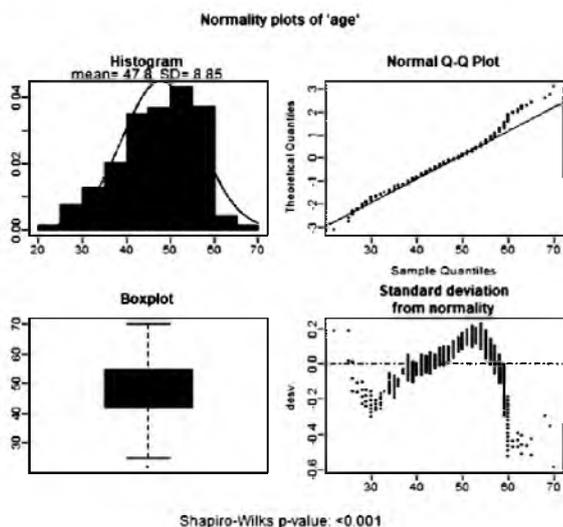


Рисунок 1 – Распределение параметра возраст

Для оценки и подготовки базы данных с результатами генотипирования использовался *dataPreparationRpackage*. Пропущенные данные

Таблица 1 – Половозрастная структура участников

Показатель	Условно здоровые (n=176)	PCA (n=337)	p
Возраст (n=513)	44.0 [38.0;50.0]	51.0 [44.0;56.0]	<0.001
Пол (n=474)			0.563
Женский	76 (45.8%)	151 (49.0%)	
Мужской	90 (54.2%)	157 (51.0%)	

возраст оценивался методом Манна – Уитни, так выборка не имела нормального распределение; указаны медиана и квантили 1 и 3 в квадратных скобках

восполнялись на основе метода *RandomForest* с учетом заполненных данных для конкретного образца и данных всей базы.

Оценка диагностической ценности результатов генотипирования проводилась на основании параметров чувствительности, специфичности и точности.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

За период с апреля 2018 г. по апрель 2020 г. было рекрутировано 357 пациентов с PCA и 204 условно здоровых, исключаящих ремоделирование.

Проанализирована половозрастная структура участников исследования (табл. 1). Так, возраст в среднем был выше в группе пациентов с ремоделированием сонных артерий ($p < 0,001$), а распределение по полу имело схожее соотношение в обеих группах ($p = 0,543$).

Из дальнейшего анализа 34 из 118 полиморфизмов были исключены, так как не преодолели порога в 98% (доля полученных результатов составила ниже 98%). Для оставшихся 84 SNP пропущенные значения были заполнены на основе алгоритма *RandomForest* с параметрами $pm.m.k=3$, $maxiter=15$, $num.trees=10000$. На следующем этапе были исключены полиморфизмы, которые не имели варибельности в выборке (доля одного генотипа составляла 100%). В результате была получена база из 75 полиморфизмов.

На основе данных полиморфизмов была построена бинарная классификационная модель в виде дерева решений на основе Fast-and-frugal trees (FFTs) (рис. 2). Интерпретация включает в себя 6 шагов: 1) если $rs1454157 \neq \{CC, TT\}$, значит здоровые; 2) если $rs11646213 = \{AT, TT\}$, значит PCA; 3) если $rs4868731 = \{AG\}$, значит PCA; 4) если $rs2081015 \neq \{CT, CC\}$, значит здоровые; 5) если $rs113296370 \neq \{AC\}$, значит здоровые; 6) иначе PCA.

Лучшая модель показала относительно невысокие параметры чувствительности (52%) и специфичности (64%) при общей точности в 56% (рис. 2). Данная модель включает в себя 5 полиморфизмов и в зависимости от резуль-

Таблица 2 – Диагностические параметры полиморфизмов в классификации между здоровыми и пациентами с PCA

SNPs	Генотипы	Чувствительность	Специфичность	Точность
rs1454157	CC, TT	0.638655	0.470588	0.554622
rs11646213	AT, TT	0.627451	0.47549	0.551471
rs4868731	AG	0.336134	0.759804	0.547969
rs2081015	CT, CC	0.714286	0.372549	0.543417
rs113296370	AC	0.296919	0.789216	0.543067

татов генотипирования для конкретного полиморфизма делает вывод.

Проанализированы результаты чувствительности, специфичности и точности, рассчитанные отдельно для каждого полиморфизма, входящие в модель классификации (табл. 2).

Модель на основе 5 полиморфизмов дает небольшое преимущество в точности классификации в сравнении с отдельными полиморфизмами (56% против 55%), а чувствительность и специфичность модели отражает усредненные показатели полиморфизмов, входящие в модель. В результате модель позволяет получить специфичность в 64% при 52% чувствительности, какими показателями не обладает полиморфизм по отдельности.

Несмотря на то, что генетические данные имеют относительно невысокие диагностические показатели в способности классифицировать больных с PCA от здоровых, тем не менее, генотипирование для оценки шансов развития PCA может играть важную роль. Работа R. Zh. Karabayeva et al., проведенная на казахской популяции, указывает на наличие связи между rs17016480, rs923109 и ремоделированием при гипертензии.

Таким образом, классификационная модель на основе машинного обучения продемонстрировала относительно невысокие показатели чувствительности, специфичности и точности в способности проводить разделение между пациентами с PCA и условно-здоровыми людьми на основе генетических данных.

ЛИТЕРАТУРА

1 Белова Л. А. Гипертоническая энцефалопатия: клинико-патогенетические подтипы, классификация, диагностика /Л. А. Белова, В. В. Машин. – Ульяновск: УлГУ, 2010. – 210 с.

2 Информация здравоохранения для Европейского региона / <https://gateway.euro.who.int/ru/hfa-explorer/2020>.

3 Каргабаева Б. А. Здоровье населения Республики Казахстан и деятельность организаций здравоохранения /Б. А. Каргабаева, Ж. К. Алдажарова, А. А. Кенесова и др. http://www.rcrz.kz/index.php/ru/?option=com_content&view=article&id=9736

http://www.rcrz.kz/index.php/ru/?option=com_content&view=article&id=9736

4 Шумилина М. В. Комплексная ультразвуковая диагностика патологии периферических сосудов. Учеб.-метод. рук. – М.: НЦССХ им. А.Н. Бакулева РАМН, 2012. – 384 с.

5 GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes /N. Franceschini, C. Giambartolomei, P. S. de Vries et al. //Nat. Commun. – 2018. – №9. – 5141P.

6 Williams B. Guidelines for the Management of Arterial Hypertension: The Task Force for the Management of Arterial Hypertension of the European Society of Cardiology (ESC) and the European Society of Hypertension (ESH) /B. Williams, G. Mancia //Eur. Heart J. – 2018. – №39. – P. 3021-3104.

REFERENCES

1 Belova L. A. Gipertonicheskaja jencefalopatija: kliniko-patogeneticheskie podtipy, klassifikacija, diagnostika /L. A. Belova, V. V. Mashin. – Ul'janovsk: UIGU, 2010. – 210 s.

2 Informacija zdravoohranenija dlja Evropejskogo regiona / <https://gateway.euro.who.int/ru/hfa-explorer/2020>.

3 Kargabaeva B. A. Zdorov'e naselenija Respubliki Kazahstan i dejatel'nost' organizacij zdravoohranenija /B. A. Kargabaeva, Zh. K. Aldazharova, A. A. Kenesova i dr. http://www.rcrz.kz/index.php/ru/?option=com_content&view=article&id=9736

4 Shumilina M. V. Kompleksnaja ul'trazvukovaja diagnostika patologii perifericheskikh sosudov. Ucheb.-metod. ruk. – М.: NССSH им. А.Н. Бакулева РАМН, 2012. – 384 с.

5 GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes /N. Franceschini, C. Giambartolomei, P. S. de Vries et al. //Nat. Commun. – 2018. – №9. – 5141P.

6 Williams B. Guidelines for the Management of Arterial Hypertension: The Task Force for the Management of Arterial Hypertension of the European Society of Cardiology (ESC) and the European Society of Hypertension (ESH) /B. Williams, G. Mancia //Eur. Heart J. – 2018. – №39. – P. 3021-3104.

Поступила 08.07.2020 г.

V. V. Benberin¹, T. A. Voshchenkova¹, R. Zh. Karabayeva¹, A. S. Sibagatova¹, D. B. Babenko²,
A. A. Turmukhambetova²

DIAGNOSTIC INDICATORS OF GENETIC FACTORS IN THE ABILITY TO CLASSIFY PATIENTS WITH CAROTID ARTERY
REMODELING FROM CONDITIONALLY HEALTHY ON THE BASIS OF MACHINE LEARNING

¹RSE «Hospital of the Medical Center of the Administrative Department of the President of the Republic of
Kazakhstan» (Nur-Sultan, Kazakhstan);

²NP JSC «Karaganda medical university» (Karaganda, Kazakhstan)

A model was obtained to differentiate between patients with carotid artery remodeling and conventionally healthy people based on machine learning using genetic data from the Kazakh population

For the period 2018-2020, 561 study participants were recruited, of which 357 had PCA and 204 were conditionally healthy. Genotyping was performed using QuantStudio TM 12K Flex Real-Time PCR technology (Applied Biosystems) using a panel of 118 polymorphisms. After the quality control procedure, 43 polymorphisms were excluded from further analysis.

The results of genotyping of 75 polymorphisms were used to build a binary classification model in the form of a «decision tree» based on the Fast-and-frugal trees (FFT) algorithm. The best model of 5 polymorphisms had 52% sensitivity, 64% specificity, and 56% accuracy in the ability to distinguish between RSA patients and conventionally healthy people based on genetic data.

Key words: carotid artery remodeling, machine learning, classification model, decision tree

В. В. Бенберин¹, Т. А. Вощенкова¹, Р. Ж. Қарабаева¹, А. С. Сибгатовва¹, Д. Б. Бабенко², А. А. Тұрмұхамбетова²
ГЕНЕТИКАЛЫҚ ФАКТОРЛАРДЫҢ ДИАГНОСТИКАЛЫҚ КӨРСЕТКІШТЕРІ, МАШИНАЛАРДЫ ОҚЫТУ НЕГІЗІНДЕ
ШАРТТЫ ДЕНІ САУЛЫҚТАН ЖАҢАРТЫЛҒАН, КАРОЦИДТІ АРТЕРИЯСЫ МЕНЕН НАУҚАСТАРДЫ КЛАССИЯЛАУ
МҮМКІНДІГІ

¹Қазақстан Республикасы Президентінің Әкімшілік бөлімінің медициналық орталығының ауруханасы
(Нұр-Сұлтан, Қазақстан);

²Қарағанды медициналық университеті (Қарағанды, Қазақстан)

Қазақ популяциясының генетикалық деректерін пайдалана отырып, Машиналық оқыту негізінде ұйқылы артерияларды (РСА) және шартты-сау адамдарды ремоделдейтін пациенттер арасында саралау жүргізу үшін үлгі алынды.

2018-2020 жылдар аралығында 561 зерттеу қатысушысы жұмысқа қабылданды, оның ішінде 357-і РСА және 204-і шартты-сау. Генотиптеу 118 полиморфизм панелін қолдана отырып, QuantStudio TM 12K Flex real-Time PCR (Applied Biosystems) технологиясының көмегімен жасалды. Котрол процедурасынан кейін 43 полиморфизм одан әрі талдаудан шығарылды.

75 полиморфизмнің генотиптеу нәтижелері алгоритмнің Fast-and-frugal trees (FFT) негізінде «шешімдер ағашы» түрінде екілік жіктеу моделін құру үшін қолданылды. 5 полиморфизмнен тұратын ең үздік модель 52% сезімталдыққа, 64% ерекшелікке және 56% дәлдікке ие болды.

Кілт сөздер: каротид артериясын қайта құру, машиналық оқыту, классификация моделі, шешім ағашы