

БИОЛОГИЯ**BIOLOGY**

УДК 575.113.1:575.21:577.24

<https://doi.org/10.29235/1561-8323-2021-65-5-582-591>

Поступило в редакцию 17.06.2021

Received 17.06.2021

**В. А. Лемеш¹, В. Н. Кипень¹, М. В. Богданова¹, А. А. Буракова¹,
член-корреспондент А. Г. Булгак², А. В. Байда³, О. В. Зотова², О. И. Добыш¹**

¹*Институт генетики и цитологии Национальной академии наук Беларуси, Минск, Республика Беларусь*

²*Республиканский научно-практический центр «Кардиология», Минск, Республика Беларусь*

³*Белорусская медицинская академия последипломного образования, Минск, Республика Беларусь*

**ОПРЕДЕЛЕНИЕ ВОЗРАСТА ЧЕЛОВЕКА ПО ОБРАЗЦАМ КРОВИ
НА ОСНОВАНИИ АНАЛИЗА МЕТИЛИРОВАНИЯ CpG-ДИНУКЛЕОТИДОВ**

Аннотация. На основании биоинформатического и статистического анализа GEO-проектов по определению полногеномного профиля метилирования ДНК человека сформирован перечень из 27 CpG-динуклеотидов с высоким прогностическим потенциалом для создания моделей предсказания возраста человека по образцам крови. Определен уровень метилирования для 245 образцов индивидов из Республики Беларусь, рассчитаны коэффициенты корреляции R и построены математические модели определения возраста человека. Среднее значение точности предсказания возраста по образцам крови с использованием 12 CpG-динуклеотидов составило 3,4 года (для мужчин – 3,3, для женщин – 3,5). Полученные результаты будут положены в основу разработки калькуляторов предсказания возраста индивида по биологическим следам для экспертов-криминалистов.

Ключевые слова: криминалистика, определение возраста, кровь, метилирование ДНК, Human Methylation 450k BeadChip

Для цитирования. Определение возраста человека по образцам крови на основании анализа метилирования CpG-динуклеотидов / В. А. Лемеш [и др.] // Докл. Нац. акад. наук Беларуси. – 2021. – Т. 65, № 5. – С. 582–591. <https://doi.org/10.29235/1561-8323-2021-65-5-582-591>

**Valentina A. Lemesh¹, Viachaslau N. Kipen¹, Marina V. Bahdanava¹, Arina A. Burakova¹,
Corresponding Member Alexander G. Bulgak², Alexander V. Bayda³, Volga V. Zotova², Volga I. Dobysh¹**

¹*Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus*

²*Scientific and Practical Centre «Cardiology», Minsk, Republic of Belarus*

³*Belarusian Medical Academy of Postgraduate Education, Minsk, Republic of Belarus*

**DETERMINING THE HUMAN CHRONOLOGICAL AGE FROM THE BLOOD SAMPLES
ON THE BASIS OF THE ANALYSIS OF CpG-DINUCLEOTIDES METHYLATION**

Abstract. Based on the bioinformatic and statistical analysis of the GEO-projects to determine the genome-wide profile of human DNA methylation, a list of 27 CpG dinucleotides with a high predictive potential was formed to create models for prediction of the human age from blood samples. The methylation level was determined for 245 samples of individuals from the Republic of Belarus. The correlation coefficients R were calculated, and the mathematical models for determining the age of an individual were constructed. The average accuracy value of the age prediction from blood samples using 12 CpG-dinucleotides was 3.4 years (for men – 3.3, for women – 3.5). The results obtained will be used as a basis for development of calculators for predicting the age of an individual based on the biological traces for forensic experts.

Keywords: forensic, age determination, blood, DNA methylation, Human Methylation 450k BeadChip

For citation. Lemesh V. A., Kipen V. N., Bahdanava M. V., Burakova A. A., Bulgak A. G., Bayda A. V., Zotova V. V., Dobysh V. I. Determining the human chronological age from the blood samples on the basis of the analysis of CpG-dinucleotides methylation. *Doklady Natsional'noi akademii nauk Belarusi = Doklady of the National Academy of Sciences of Belarus*, 2021, vol. 65, no. 5, pp. 582–591 (in Russian). <https://doi.org/10.29235/1561-8323-2021-65-5-582-591>

Введение. Метилирование ДНК является процессом прижизненной химической модификации молекулы ДНК путем добавления метильной группы к цитозину в составе CpG-динуклеотида с помощью ферментов ДНК-метилтрансфераз, причем последовательность нуклеотидов не изменяется. В геноме представлены два типа распределения CpG-динуклеотидов: рассеянные CpG в виде одиночных динуклеотидов (около 80 %) и целые области CpG, называемые CpG-островками [1]. Одиночные CpG обнаруживаются чаще всего в межгенных и реже – в транскрибируемых последовательностях. CpG-островки располагаются вблизи структурных генов, преимущественно в 5'-регионах, которые содержат регуляторные последовательности, характерные для промоторов [2].

Изучение распределения CpG-сайтов по геному, степени их метилирования является актуальным для понимания эпигенетических процессов в геноме организмов, поскольку было показано, что профиль метилирования ДНК меняется с возрастом [3]. Впервые изменения уровня метилирования ДНК в онтогенезе были идентифицированы при анализе различных органов и жизненных стадий североатлантических рыб [4]. Позже было показано, что возрастное глобальное ДНК-гипометилирование наблюдается у разных видов, включая крыс, мышей и человека [5].

Развитие геномных биотехнологий и расширение их применения в эпигеномных исследованиях позволило установить специфические возрастные изменения профилей метилирования определенных генов или геномных областей. Были выявлены маркеры, уровни метилирования которых могут быть использованы для определения возраста индивида [6]. S. Horvath сформулировал концепцию «эпигенетических часов» [7]. При помощи регрессионной модели автор выделил 353 CpG-локуса, которые являются основой «эпигенетических часов», поскольку анализ уровней метилирования данных локусов позволяет определить возраст биологического образца и является стабильным при анализе различных типов клеток и тканей.

Одной из областей, в которой важно изучение паттернов метилирования ДНК, является криминалистика, поскольку возможно определять возраст подозреваемого по оставленным на месте преступления образцам и фрагментам биологических тканей, что может помочь при идентификации останков и сузить круг подозреваемых.

Подход к определению хронологического возраста, основанный на выявлении уровня метилирования ДНК в конкретных CpG-динуклеотидах, является наиболее сбалансированным с позиции точность/стоимость для задач криминалистики [8–10]. Изучение метилирования ДНК в связи с определением возраста представителей населения Союзного государства и, в частности, Республики Беларусь ранее не проводилось, а увеличение точности представленных в научной литературе данных при определении возраста конкретного индивида, а не в среднем по выборке, требует специальных исследований, так как было продемонстрировано, что уровень метилирования CpG-динуклеотидов может различаться в зависимости от этногеографического происхождения индивидов и факторов окружающей среды [11; 12].

Цель исследования – оценить точность предсказания возраста человека с использованием математических моделей на основании результатов анализа метилирования CpG-динуклеотидов в образцах крови индивидов из Республики Беларусь.

Материалы и методы исследования. Все процедуры, выполненные в исследовании с участием людей, соответствуют этическим стандартам институционального и/или национального комитета по исследовательской этике и Хельсинкской декларации 1964 г. и ее последующим изменениям или сопоставимым нормам этики. От каждого из включенных в исследование участника было получено информированное добровольное согласие.

Биологические образцы. После разъяснительной беседы и добровольного согласия в письменной форме у 275 индивидов (мужчины – 119, женщины – 156) осуществлен забор венозной крови в пробирки BD Vacutainer K2E. Средний возраст добровольцев составил $57,3 \pm 16,9$ лет (для мужчин – $58,4 \pm 16,8$ лет; для женщин – $56,4 \pm 17,0$ лет), диапазон – от 18 до 93 полных лет.

ДНК экстрагировали с использованием набора MagMAX™ DNA Multi-Sample Kit (ThermoFisher, США) в соответствии с инструкциями производителя. Выделенную ДНК количественно оценивали с использованием NanoPhotometer N50 (IMPLEN, США).

Отбор CpG-динуклеотидов для анализа. Нами проведен анализ существующих тест-систем, а также выполнен самостоятельный биоинформатический поиск CpG-динуклеотидов с высоким потенциалом для определения возраста по данным GEO-проектов (www.ncbi.nlm.nih.gov/gds). J. Naue и соавт. предложили использовать тест-систему для определения возраста по образцам крови, включающую 13 CpG-динуклеотидов [13]. Проведенный нами биоинформатический анализ данных полногеномного профиля метилирования в рамках восьми GEO-проектов подтвердил высокий прогностический потенциал десяти из 13 CpG-динуклеотидов (cg02337583, cg02872426, cg06784991, cg06874016, cg07553761, cg11807280, cg16054275, cg16867657, cg18473521 и cg25410668) для предсказания возраста по образцам крови для различных этнорегиональных групп [14]. Кроме того, дополнительно отобраны для молекулярно-генетических исследований белорусской популяции CpG-динуклеотиды, статистически значимо ассоциированные с хронологическим возрастом: cg00481951 (ген *SST*, NCBI Gene ID – 6750), cg05213896 (*IL4I1*, 259307), cg07547549 (*SLC12A5*, 57468), cg08128734 (*RASSF5*, 83593), cg08468401, cg08928145 (*TSSK6*, 83983), cg10501210 (*MIR29B2CHG*, 100128537), cg12757011 (*SLC4A10*, 57282), cg14361627 (*LOC105375508*, 105375508), cg18384097 (*PTPN7*, 5778), cg19283806 (*CCDC102B*, 79839), cg19671120 (*CNGA3*, 1261), cg22454769 (*FHL2*, 2274) и cg24079702 (*FHL2*, 2274). Таким образом, для определения профиля метилирования ДНК образцов крови индивидов из Республики Беларусь нами отобраны 27 CpG-динуклеотидов.

Бисульфитное секвенирование с использованием SNaPshot. Праймеры для анализа CpG-динуклеотидов cg02872426, cg06784991, cg06874016, cg07553761, cg11807280, cg16054275, cg16867657, cg18473521, cg25410668, Chr.3:g.51707136, Chr.4:g.16573797, Chr.7:g.130734357 и Chr.20:g.63980491 представлены в [13]. Праймеры для анализа CpG-динуклеотидов cg00481951, cg07547549, cg08928145, cg12757011, cg14361627, cg18384097 и cg19671120 представлены в [15]. Праймеры и SBE-олигонуклеотиды для анализа CpG-динуклеотидов cg05213896, cg08128734, cg08468401, cg10501210, cg19283806, cg22454769 и cg24079702 смоделированы нами с использованием программы BiSearch (<http://bisearch.enzim.hu/>) (табл. 1).

Бисульфитная конверсия ДНК проводилась для 200–500 нг с использованием MethylEdge® Bisulfite Conversion System Kit (Promega, США). ПЦР проводили в объеме 20 мкл, содержащем 10–15 нг конвертированной ДНК, 1 ед. ArtStart DNA полимеразы (ArtBioTech, Беларусь), 2 мкл 10x буфера для ПЦР (в состав которого входил хлорид магния в концентрации 3,0 мМ), 200 нМ каждого дНТФ и 0,4–1,0 мкМ каждого праймера. Протокол амплификации на термоциклере ProFlex PCR System (ThermoFisher, США) включал следующие этапы: 95 °С – 4 мин, далее 34 цикла (94 °С – 20 с, 56 °С – 30 с, 72 °С – 45 с), финальная элонгация при 72 °С в течение 7 мин. Далее 5 мкл амплификата очищали с использованием Exo-CIP™ Rapid PCR Cleanup Kit (NEB, США). Реакцию микросеквенирования (SBE, Single-base extension) проводили с использованием SNaPshot Kit (Applied Biosystems, США) согласно протоколу производителя, каждая реакция включала 3 мкл очищенного на предыдущей стадии амплификата и 0,2–0,4 мкМ соответствующего SBE-олигонуклеотида. Далее 10 мкл амплификата очищали с использованием 1 ед. FastAP Thermosensitive Alkaline Phosphatase (ThermoFisher, США).

Электрофоретическая детекция продуктов реакции микросеквенирования проводилась с использованием генетического анализатора ABI PRISM 3500 и программного обеспечения GeneMapper® 5.0 (Applied Biosystems, США).

Процент метилирования (0–100 %) для каждого CpG-динуклеотида рассчитывали согласно формуле: интенсивность флуоресцентного сигнала для G (метилированные, т. е. неконвертированные цитозины; согласно правилу комплементарности C-G) разделить на сумму флуоресцентных сигналов G и A (неметилированные, т. е. конвертированные в T цитозины; согласно правилу комплементарности T-A).

Статистический анализ данных. В ходе биоинформатического и статистического анализа были рассчитаны: коэффициенты корреляции R по Спирману (Spearman rank correlation coefficient); коэффициенты детерминации (R^2), а также скорректированные R^2 , равные доле дисперсии зависимой переменной «Возраст», обусловленной влиянием независимых переменных, представляющих собой уровень метилирования CpG-динуклеотидов; значения среднего абсолютного

Т а б л и ц а 1. **Информация об CpG-динуклеотидах, отобранных для исследования на основании биоинформатического анализа GEO-проектов**

Table 1. **Information about CpG dinucleotides selected for research based on bioinformatics analysis of GEO projects**

CpG-динуклеотид CpG dinucleotide	Ген Gene	Хромосомная позиция* Chromosomal position*
cg05213896	<i>IL4I1</i>	Chr.19:g.49890396
1	F-праймер (5'>3'): AACTTAAAAACACCTTAATA R-праймер (5'>3'): AATTTGAAGGTGTTGAAGGT SBE-олигонуклеотид (5'>3'): (T)n-ACAACAACAACAAAAAATAATAC	
cg08128734	<i>RASSF5</i>	Chr.1:g.206512090
2	F-праймер (5'>3'): TTTATTTGGGTATTTGGAAGGT R-праймер (5'>3'): CCATTTCTATACTTCATTTTCT SBE-олигонуклеотид (5'>3'): (T)n-CTTAAACTCTAAAACCAAACAC	
cg08468401	–	Chr.3:g.14261631
3	F-праймер (5'>3'): TTTTAGTTGTTTGTTGGTGT R-праймер (5'>3'): TTTCTTTAAACTTAATCTCATTC SBE-олигонуклеотид (5'>3'): (T)n-ATAACATAACTCCTAACACCTC	
cg10501210	<i>MIR29B2CHG</i>	Chr.1:g.207823675
4	F-праймер (5'>3'): GAAAAGAAATTTAAAGTGGGAA R-праймер (5'>3'): ACCAAATCTAAAACATTCRAAAC SBE-олигонуклеотид (5'>3'): (T)n-CCAAAATTTAAATCTACRCAAAC	
cg19283806	<i>CCDC102B</i>	Chr.18:g.68722183
5	F-праймер (5'>3'): GGGGAATGTTTGATTTATTT R-праймер (5'>3'): TAAAAAAATTACCCTTCTCT SBE-олигонуклеотид (5'>3'): (T)n-ACAAAAATAACAACCAAAAAAATAATAC	
cg22454769	<i>FHL2</i>	Chr.2:g.105399310
6	F-праймер (5'>3'): TGTTAGGGTTTTTTTTTYGTGT R-праймер (5'>3'): TAAAACCTTCTCCAATCTCCC SBE-олигонуклеотид (5'>3'): (T)n-CAAACCTCRTCCRAAACTC	
cg24079702	<i>FHL2</i>	Chr.2:g.105399314
7	F-праймер (5'>3'): TGTTAGGGTTTTTTTTTYGTGT R-праймер (5'>3'): TAAAACCTTCTCCAATCTCCC SBE-олигонуклеотид (5'>3'): (T)n-CAAACCTCRTCCRAAACTC	

Примечание: * – GRCh38.p13 (GCF_000001405.39)
Note: * – GRCh38.p13 (GCF_000001405.39)

отклонения (Mean Absolute Deviation, MAD) от истинного возраста и среднеквадратическая ошибка модели (Root Mean Square Error, RMSE). Данные уровня метилирования CpG-динуклеотидов нормализованы с использованием алгоритма Batch normalization ($n = 30$). Статистический анализ проводили с использованием программного обеспечения SPSS v.20.0.

Результаты и их обсуждение. Коэффициенты корреляции R . Коэффициенты R для 27 CpG-динуклеотидов (с указанием половой принадлежности), рассчитанные по результатам молекулярно-генетического анализа 275 образцов крови, представлены в табл. 2. Наибольшие абсолютные значения R (по модулю) для всей выборки показаны для CpG-динуклеотидов cg19283806 (–0,811), cg14361627 (0,767), cg08128734 (–0,707), cg16867657 (0,675), cg22454769 (0,674), cg07553761 (0,652), cg08468401 (–0,637), cg11807280 (–0,618).

Для 16 CpG-динуклеотидов корреляция с возрастом была положительной, для 11 – отрицательной. Для CpG-динуклеотида cg08928145 статистически значимая ассоциация с возрастом отсутствовала как для мужчин, так и для женщин, несмотря на показанную ассоциацию в рамках биоинформатического анализа GEO-проектов. Соответственно, этот CpG-динуклеотид в дальнейшем моделировании не использовался.

Регрессионная модель предсказания возраста. При использовании множественной линейной регрессии (26 CpG-динуклеотидов) в качестве модели предсказания возраста по образцам крови в общей группе индивидов (без разделения по полу) значение MAD составило 3,2 года, RMSE –

Т а б л и ц а 2. Коэффициенты корреляции (R) с возрастом для 27 CpG-динуклеотидов
 Table 2. Correlation coefficients (R) with age for 27 CpG dinucleotides

CpG-динуклеотиды CpG dinucleotides	Мужчины ($n = 119$) Men ($n = 119$)		Женщины ($n = 156$) Women ($n = 156$)	
	R	p -уровень p -level	R	p -уровень p -level
cg00481951	0,182	4,80E-02	0,161	4,51E-02
cg02872426	-0,423	1,42E-06	-0,485	1,12E-06
cg05213896	0,515	2,12E-09	0,572	6,29E-15
cg06784991	0,334	2,03E-04	0,454	2,60E-09
cg06874016	-0,383	1,74E-05	-0,375	1,40E-06
cg07547549	0,441	5,33E-07	0,508	1,35E-11
cg07553761	0,629	1,91E-14	0,675	4,28E-22
cg08128734	-0,682	1,27E-17	-0,730	2,90E-27
cg08468401	-0,530	5,51E-10	-0,713	1,48E-25
cg08928145	0,078	0,398	0,046	0,568
cg10501210	-0,430	1,09E-06	-0,496	4,64E-11
cg11807280	-0,605	3,15E-13	-0,619	7,51E-18
cg12757011	0,336	1,88E-04	0,363	3,25E-06
cg14361627	0,776	3,55E-25	0,751	1,56E-29
cg16054275	-0,417	2,37E-06	-0,370	2,05E-06
cg16867657	0,610	1,72E-13	0,716	7,70E-26
cg18384097	-0,308	6,47E-04	-0,300	1,42E-04
cg18473521	0,503	5,62E-09	0,450	3,88E-09
cg19283806	-0,789	1,65E-26	-0,820	3,23E-39
cg19671120	0,431	1,01E-06	0,243	2,21E-03
cg22454769	0,571	1,19E-11	0,734	1,29E-27
cg24079702	0,106	0,252	0,238	2,73E-03
cg25410668	0,484	2,40E-08	0,686	4,88E-23
Chr.3:g.51707136	0,234	1,04E-02	0,465	9,58E-10
Chr.4:g.16573797	-0,508	3,55E-09	-0,482	1,92E-10
Chr.7:g.130734357	0,470	6,81E-08	0,466	8,98E-10
Chr.20:g.63980491	-0,466	9,48E-08	-0,608	3,99E-17

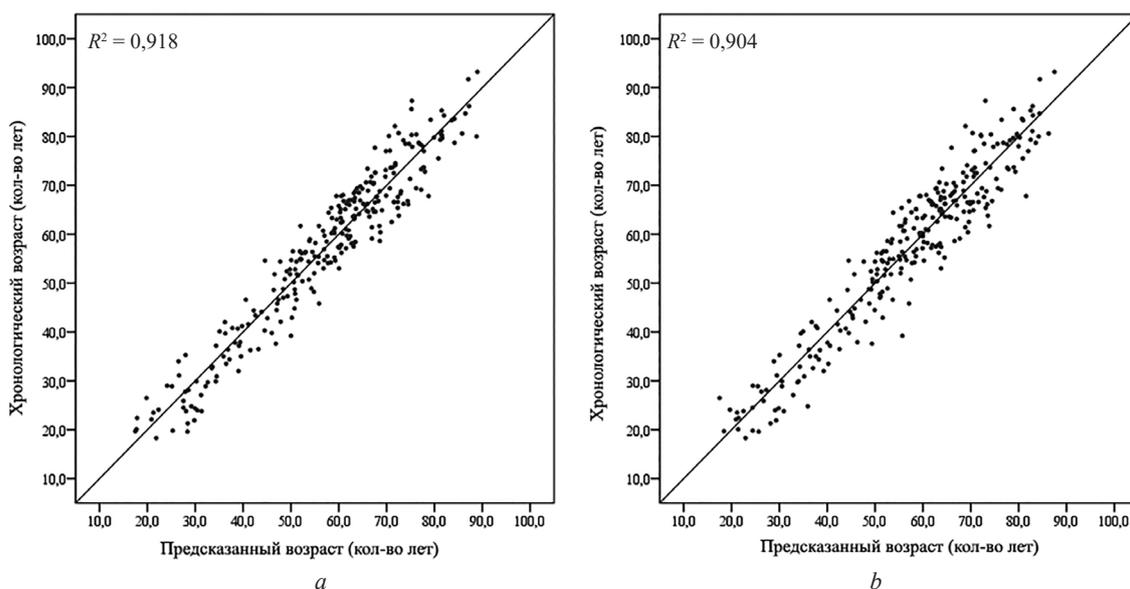


Рис. 1. Точность предсказания возраста индивида по образцам крови с использованием 26 (а) и 12 (б) CpG-динуклеотидов

Fig. 1. The accuracy of predicting the age of an individual from blood samples using 26 (a) and 12 (b) CpG dinucleotides

2,3 года (рис. 1, *a*). Полученные нами результаты сопоставимы с аналогичными исследованиями [13; 16–19].

Скорректированный процент объясненной дисперсии R^2 составил не менее 91,8 %. По влиянию на изменение коэффициента детерминации (R^2) CpG-динуклеотиды расположились в следующей последовательности: cg19283806 (67,9 %), cg14361627 (10,4 % к предыдущему значению), cg22454769 (4,5 %), cg11807280 (2,3 %), cg08468401 (1,4 %), cg05213896 (1,4 %) и cg16867657 (0,7 %). Совокупный вклад остальных маркеров составил не более 3,2 %. Возраст с точностью до 3 лет корректно предсказан для 42,9 % индивидов, с точностью до 5 лет – для 72,0 %, с точностью до 7 лет – для 86,9 %, для 13,1 % ошибка предсказания возраста составила более 7 лет.

При использовании в множественной линейной регрессии только 12 CpG-динуклеотидов с максимальным влиянием на R^2 (cg19283806, cg14361627, cg22454769, cg11807280, cg08468401, cg05213896, cg16867657, cg07547549, cg25410668, cg08128734, cg10501210, Chr.4:g.16573797) значение MAD составило 3,4 года, RMSE – 2,3 года (рис. 1, *b*). Скорректированный процент объясненной дисперсии R^2 составил не менее 90,4 %. Возраст с точностью до 3 лет корректно предсказан для 40,0 % индивидов, с точностью до 5 лет – для 64,0 %, с точностью до 7 лет – для 84,7 %, для 15,3 % ошибка предсказания возраста составила более 7 лет.

При создании модели предсказания возраста по 12 CpG-динуклеотидам для образцов крови в группе мужчин значение MAD составило 3,3 года, RMSE – 2,3 года (рис. 2, *a*). Скорректированный процент объясненной дисперсии R^2 составил не менее 90,2 %. Возраст с точностью до 3 лет корректно предсказан для 47,9 % индивидов, с точностью до 5 лет – для 67,2 %, с точностью до 7 лет – для 88,2 %, для 11,8 % ошибка предсказания возраста составила более 7 лет.

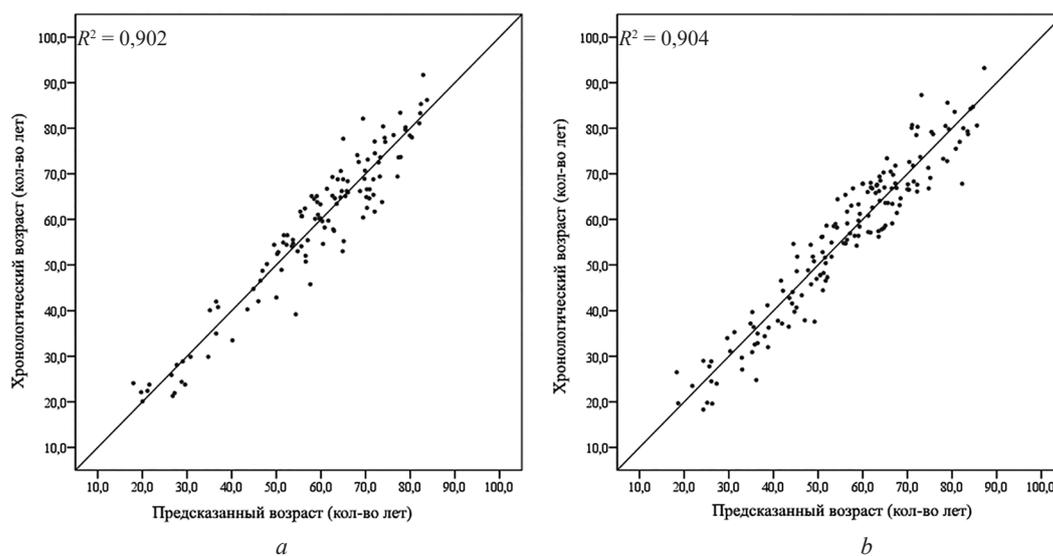


Рис. 2. Точность предсказания возраста для мужчин (*a*) и для женщин (*b*) по образцам крови с использованием 12 CpG-динуклеотидов

Fig. 2. The accuracy of age prediction for men (*a*) and for women (*b*) from blood samples using 12 CpG dinucleotides

При создании модели предсказания возраста по 12 CpG-динуклеотидам для образцов крови в группе женщин значение MAD составило 3,5 года, RMSE – 2,3 года (рис. 2, *b*). Скорректированный процент объясненной дисперсии R^2 составил не менее 90,4 %. Возраст с точностью до 3 лет корректно предсказан для 37,2 % индивидов, с точностью до 5 лет – для 66,7 %, с точностью до 7 лет – для 87,8 %, для 12,2 % ошибка предсказания возраста составила более 7 лет.

Таким образом, лучшая точность предсказания возраста индивида (без указания половой принадлежности) по образцам крови составила не более 3,2 года при анализе 26 CpG-динуклеотидов и 3,4 года при анализе 12 CpG-динуклеотидов.

Вероятность отнесения образца к возрастной группе. Известно, что показатели R^2 , MAD и RMSE отражают общую точность модели и позволяют сравнивать модели между собой. Для

характеристики точности предсказания зависимой переменной «Возраст» для конкретного образца в криминалистической практике целесообразно рассчитывать также вероятность отнесения образца к возрастной группе, например, « ≤ 40 лет», « >40 и ≤ 60 лет» и « >60 лет».

На рис. 3 представлена информация о количестве индивидов (без разделения по полу), выраженном в процентах в пределах трех возрастных групп, для которых с использованием регрессионной модели из 12 CpG-динуклеотидов были предсказаны значения возраста в пределах ошибки « ≤ 3 лет», « >3 и ≤ 5 лет», « >5 и ≤ 7 лет» и « >7 лет».

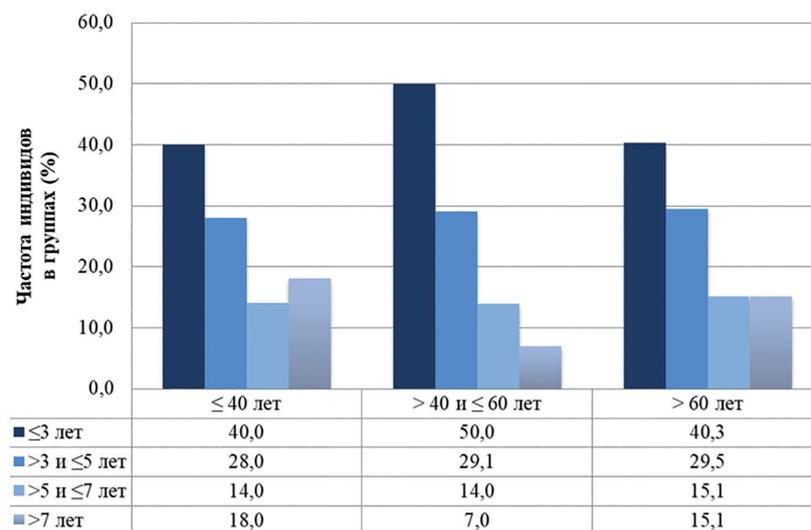


Рис. 3. Процент индивидов с корректно предсказанным возрастом в пределах заданных интервалов ошибки для трех возрастных групп

Fig. 3. Percentage of individuals with correctly predicted ages within specified error intervals for three age groups

Наибольший процент корректных расчетов приходится на возрастную группу « >40 и ≤ 60 лет» – для 79,1 % всех индивидов ошибка предсказания возраста была меньше 5 лет, для двух других групп этот показатель находился в диапазоне 68,0–69,8 %. Также из рис. 3 следует, что в группах « ≤ 40 лет» « >60 лет» ошибка предсказания возраста более 7 лет составляет 18,0 и 15,1 % соответственно. Такая тенденция может быть обусловлена более высокой дисперсией для уровня метилирования анализируемых CpG-динуклеотидов в крайних возрастных группах, что может быть связано с широким диапазоном нормы при физиологических процессах развития и старения организма.

Для криминалистов в судебной практике при выяснении предположительного возраста неизвестного индивида может быть важна информация не только о конкретном возрасте, но и об отнесении данного субъекта к определенной возрастной группе, например: «моложе 40 лет» или «старше 40 лет» и т. п. Точность отнесения неизвестного индивида к конкретной группе по результатам анализа метилирования CpG-динуклеотидов будет выше, чем при ответе на вопрос о приближенном к истинному значению хронологического возраста, так как количество групп в анализе с использованием, например, логистической регрессии, будет составлять в большинстве случаев не более двух. Например, при разбиении всего массива образцов на две категории (без разделения по полу) и использовании логистической регрессии на основании 12 CpG-динуклеотидов в среднем по модели с вероятностью 92,0 % имеется возможность сделать заключение, что возраст неизвестного индивида составляет меньше 40 лет, с вероятностью 98,2 % – больше 40 лет, с вероятностью 89,7 % – меньше 50 лет, с вероятностью 99,0 % – больше 50 лет; с вероятностью 88,2 % – меньше 60 лет, с вероятностью 89,9 % – больше 60 лет.

Для уточнения предсказанного возраста также может быть использована двухэтапная схема: отнесение неизвестного образца к определенной возрастной группе; и предсказание значения

возраста в годах (с уровнем точности в пределах предсказательной модели) уже в пределах возрастной группы.

Заключение. На основании данных, представленных в открытом доступе на платформе GEO NCBI для проектов по определению полногеномного профиля метилирования ДНК, нами определены CpG-динуклеотиды с высокими значениями линейной зависимости между уровнем метилирования и хронологическим возрастом. Для молекулярно-генетического анализа были отобраны 27 CpG-динуклеотидов с высоким прогностическим потенциалом для создания модели предсказания возраста человека по образцам крови.

Для данных CpG-динуклеотидов определен уровень метилирования в образцах крови для 275 индивидов из Республики Беларусь, рассчитаны коэффициенты корреляции R с возрастом. С использованием множественной линейной регрессии создана модель для предсказания возраста, основанная на анализе 12 CpG-динуклеотидов cg05213896, cg07547549, cg08128734, cg08468401, cg10501210, cg11807280, cg14361627, cg16867657, cg19283806, cg22454769, cg25410668 и Chr.4:g.16573797. Средняя ошибка предсказания возраста составила 3,4 года (3,3 – для мужчин, 3,5 – для женщин).

При использовании логистической регрессии на основании 12 CpG-динуклеотидов в среднем по модели с вероятностью 92,8 % имеется возможность сделать заключение, что возраст неизвестного индивида составляет больше или меньше определенного возраста.

Предложенная нами модель предсказания возраста на основании профиля метилирования CpG-динуклеотидов по образцам крови отличается относительной простотой, так как в анализе используется небольшое количество маркеров, и, таким образом, она может быть задействована в криминалистических лабораториях молекулярно-генетической направленности. Точность предложенной нами модели соответствует мировому уровню [13; 16–19].

Разработка методики определения вероятного возраста индивида по характеристике его ДНК и доведение ее до широкого практического применения в криминалистике позволит повысить эффективность проведения оперативно-розыскных мероприятий и установления личности подозреваемых, что обеспечит повышение раскрываемости преступлений, скорости реагирования правоохранительных структур на совершенные преступления.

Благодарности. Работа выполнена в рамках мероприятия 2 «Разработка методики определения вероятного возраста индивида по характеристике его ДНК» научно-технической программы Союзного государства «Разработка инновационных геногеографических и геномных технологий идентификации личности и индивидуальных особенностей человека на основе изучения генофондов регионов Союзного государства» (ДНК-идентификация).

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Acknowledgments. The research was carried out within the scientific study 2 «Development of a method for determining the probable age of an individual by the characteristics of his DNA» of the scientific and technical program of the Union State «Development of innovative gеногеографическая and genomic technologies for identification of personality and individual characteristics of a person based on the study of gene pools of regions of the Union State (DNA identification)».

Conflict of interests. The authors declare no conflicts of interest.

Список использованных источников

1. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond / P. A. Jones // *Nature Reviews Genetics*. – 2012. – Vol. 13, N 7. – P. 484–492. <https://doi.org/10.1038/nrg3230>
2. Antequera, F. High levels of De Novo methylation and altered chromatin structure at CpG islands in cell lines / F. Antequera, J. Boyes, A. Bird // *Cell*. – 1990. – Vol. 62, N 3. – P. 503–514. [https://doi.org/10.1016/0092-8674\(90\)90015-7](https://doi.org/10.1016/0092-8674(90)90015-7)
3. Park, S. K. Lessons learned from gene expression profile studies of aging and caloric restriction / S. K. Park, T. A. Prolla // *Ageing Research Reviews*. – 2005. – Vol. 4, N 1. – P. 55–65. <https://doi.org/10.1016/j.arr.2004.09.003>
4. Алтухов, Ю. П. Внутривидовое генетическое разнообразие: мониторинг и принципы сохранения / Ю. П. Алтухов // *Генетика*. – 1995. – Т. 31, № 10. – С. 1333–1357.
5. CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function / H. S. Jang [et al.] // *Genes*. – 2017. – Vol. 8, N 6. – Art. 148. <https://doi.org/10.3390/genes8060148>
6. A methylome-wide study of aging using massively parallel sequencing of the methyl-CpG-enriched genomic fraction from blood in over 700 subjects / J. L. McClay [et al.] // *Human Molecular Genetics*. – 2014. – Vol. 23, N 5. – P. 1175–1185. <https://doi.org/10.1093/hmg/ddt511>
7. Horvath, S. DNA methylation age of human tissues and cell types / S. Horvath // *Genome Biology*. – 2013. – Vol. 14, N 10. – P. 1–19. <https://doi.org/10.1186/gb-2013-14-10-r115>
8. Генетика – судебной экспертизе Беларуси / А. Кильчевский [и др.] // *Наука и Инновации*. – 2020. – № 10 (212). – С. 22–28.

9. Genome-wide methylation profiles reveal quantitative views of human aging rates / G. Hannum [et al.] // *Mol. Cell.* – 2013. – Vol. 49, N 2. – P. 359–367. <https://doi.org/10.1016/j.molcel.2012.10.016>
10. DNA Methylation Analysis: Choosing the Right Method / S. Kurdyukov, M. Bullock // *Biology (Basel)*. – 2016. – Vol. 5, N 1. – P. e1–e21. <https://doi.org/10.3390/biology5010003>
11. Impact of genetic ancestry on chronological age prediction using DNA methylation analysis / J. Fleckhaus [et al.] // *Forensic Science International: Genetics Supplement Series*. – 2017. – Vol. 6. – P. e399–e400. <https://doi.org/10.1016/j.fsigs.2017.09.162>
12. Donkin, I. Sperm epigenetics and influence of environmental factors / I. Donkin, R. Barres // *Mol. Metab.* – 2018. – Vol. 14. – P. 1–11. <https://doi.org/10.1016/j.molmet.2018.02.006>
13. Chronological age prediction based on DNA methylation: Massive parallel sequencing and random forest regression / J. Naue [et al.] // *Forensic Sci. Int. Genet.* – 2017. – Vol. 31. – P. 19–28. <https://doi.org/10.1016/j.fsigen.2017.07.015>
14. Обоснование минимального объема выборки для предсказания хронологического возраста человека / В. Н. Кипень [и др.] // *Молекулярная и прикладная генетика*. – 2021. – Т. 30. – С. 39–48.
15. DNA methylation-based age prediction from saliva: High age predictability by combination of 7 CpG markers / S. R. Hong [et al.] // *Forensic Sci. Int. Genet.* – 2017. – Vol. 29. – P. 118–125. <https://doi.org/10.1016/j.fsigen.2017.04.006>
16. Tracking age-correlated DNA methylation markers in the young / A. Freire-Aradas [et al.] // *Forensic Sci. Int. Genet.* – 2018. – Vol. 36. – P. 50–59. <https://doi.org/10.1016/j.fsigen.2018.06.011>
17. Identification and evaluation of age-correlated DNA methylation markers for forensic use / J. L. Park [et al.] // *Forensic Sci. Int. Genet.* – 2016. – Vol. 23. – P. 64–70. <https://doi.org/10.1016/j.fsigen.2016.03.005>
18. DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing / A. Vidaki [et al.] // *Forensic Sci. Int. Genet.* – 2017. – Vol. 28. – P. 225–236. <https://doi.org/10.1016/j.fsigen.2017.02.009>
19. Development of a forensically useful age prediction method based on DNA methylation analysis / R. Zbiac-Piekarska [et al.] // *Forensic Sci. Int. Genet.* – 2015. – Vol. 17. – P. 173–179. <https://doi.org/10.1016/j.fsigen.2015.05.001>

References

1. Jones P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, 2012, vol. 13, no. 7, pp. 484–492. <https://doi.org/10.1038/nrg3230>
2. Antequera F., Boyes J., Bird A. High levels of De Novo methylation and altered chromatin structure at CpG islands in cell lines. *Cell*, 1990, vol. 62, no. 3, pp. 503–514. [https://doi.org/10.1016/0092-8674\(90\)90015-7](https://doi.org/10.1016/0092-8674(90)90015-7)
3. Park S. K., Prolla T. A. Lessons learned from gene expression profile studies of aging and caloric restriction. *Ageing Research Reviews*, 2005, vol. 4, no. 1, pp. 55–65. <https://doi.org/10.1016/j.arr.2004.09.003>
4. Altuhov Y. P. Intraspecific genetic diversity: monitoring and conservation principles. *Genetika* [Russian Journal of Genetics], 1995, vol. 31, no. 10, pp. 1333–1357 (in Russian).
5. Jang H. S., Shin W. J., Lee J. E., Do J. T. CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. *Genes*, 2017, vol. 8, no. 6, art. 148. <https://doi.org/10.3390/genes8060148>
6. McClay J. L., Aberg K. A., Clark S. L., Nerella S., Kumar G., Xie L. Y., Hudson A. D., Harada A., Hultman C. M., Magnusson P. K. E., Sullivan P. F., Van Den Oord E. A methylome-wide study of aging using massively parallel sequencing of the methyl-CpG-enriched genomic fraction from blood in over 700 subjects. *Human Molecular Genetics*, 2014, vol. 23, no. 5, pp. 1175–1185. <https://doi.org/10.1093/hmg/ddt511>
7. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biology*, 2013, vol. 14, no. 10, pp. 1–19. <https://doi.org/10.1186/gb-2013-14-10-r115>
8. Kilchevskij A., Mosse I., Shapturenko M., Burakova A. Genetics – forensic examination of Belarus. *Nauka i Innovacii = The Science and Innovations*, 2020, no. 10 (212), pp. 22–28 (in Russian).
9. Hannum G., Guinney J., Zhao L., Zhang L., Hughes G., Sada S., Klotzle B., Bibikova M., Fan J., Gao Y., Deconde R., Chen M., Rajapakse I., Friend S., Ideker T., Zhang K. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell*, 2013, vol. 49, no. 2, pp. 359–367. <https://doi.org/10.1016/j.molcel.2012.10.016>
10. Kurdyukov S., Bullock M. DNA Methylation Analysis: Choosing the Right Method. *Biology*, 2016, vol. 5, no. 1, pp. e1–e21. <https://doi.org/10.3390/biology5010003>
11. Fleckhaus J., Freire-Aradas A., Rothschild M. A., Schneider P. M. Impact of genetic ancestry on chronological age prediction using DNA methylation analysis. *Forensic Science International: Genetics Supplement Series*, 2017, vol. 6, pp. e399–e400. <https://doi.org/10.1016/j.fsigs.2017.09.162>
12. Donkin I., Barres R. Sperm epigenetics and influence of environmental factors. *Molecular Metabolism*, 2018, vol. 14, pp. 1–11. <https://doi.org/10.1016/j.molmet.2018.02.006>
13. Naue J., Hoefsloot H. C. J., Mook O. R. F., Rijlaarsdam-Hoekstra L., van der Zwalm M. C. H., Henneman P., Kloosterman A. D., Verschure P. J. Chronological age prediction based on DNA methylation: Massive parallel sequencing and random forest regression. *Forensic Science International: Genetics*, 2017, vol. 31, pp. 19–28. <https://doi.org/10.1016/j.fsigen.2017.07.015>
14. Kipen V. N., Bahdanava M. V., Burakova A. A., Dobysh O. I., Zotova O. V., Koroleva T. S., Bulgak A. G., Baida A. V., Bruskin S. A., Lemesh V. A. Determination of minimum sample size for the prediction of human chronological age. *Molekulyarnaya i prikladnaya genetika* [Molecular and Applied Genetics], 2021, vol. 30, pp. 39–48 (in Russian).
15. Hong S. R., Jung S., Lee E. H., Shin K., Yang W. I. Y., Lee H. Y. DNA methylation-based age prediction from saliva: High age predictability by combination of 7 CpG markers. *Forensic Science International: Genetics*, 2017, vol. 29, pp. 118–125. <https://doi.org/10.1016/j.fsigen.2017.04.006>

16. Freire-Aradas A., Phillips C., Giron-Santamaria L., Ana Mosquera-Miguel A., Gomez-Tato A., Angeles Casares de Cal M., Alvarez-Dios J., Lareu M. V. Tracking age-correlated DNA methylation markers in the young. *Forensic Science International: Genetics*, 2018, vol. 36, pp. 50–59. <https://doi.org/10.1016/j.fsigen.2018.06.011>

17. Park J.-L., Kim J. H., Seo E., Bae D. H., Kim S.-Y., Lee H.-C., Woo K.-M., Kim Y. S. Identification and evaluation of age-correlated DNA methylation markers for forensic use. *Forensic Science International: Genetics*, 2016, vol. 23, pp. 64–70. <https://doi.org/10.1016/j.fsigen.2016.03.005>

18. Vidaki A., Ballard D., Aliferi A., Miller T. H., Barron L. P., Court S. D. DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing. *Forensic Science International: Genetics*, 2017, vol. 28, pp. 225–236. <https://doi.org/10.1016/j.fsigen.2017.02.009>

19. Zbiac-Piekarska R., Spolnicka M., Kupiec T., Parys-Proszek A., Makowska Z., Pateczka A., Kucharczyk K., Płoski R., Branicki W. Development of a forensically useful age prediction method based on DNA methylation analysis. *Forensic Science International: Genetics*, 2015, vol. 17, pp. 173–179. <https://doi.org/10.1016/j.fsigen.2015.05.001>

Информация об авторах

Лемеш Валентина Александровна – канд. биол. наук, доцент, заведующий лабораторией. Институт генетики и цитологии НАН Беларуси (ул. Академическая, 27, 220072, Минск, Республика Беларусь). E-mail: v.lemesh@igc.by.

Кипень Вячеслав Николаевич – канд. биол. наук, вед. науч. сотрудник. Институт генетики и цитологии НАН Беларуси (ул. Академическая, 27, 220072, Минск, Республика Беларусь). E-mail: v.kipen@igc.by.

Богданова Марина Владимировна – канд. биол. наук, вед. науч. сотрудник. Институт генетики и цитологии НАН Беларуси (ул. Академическая, 27, 220072, Минск, Республика Беларусь). E-mail: M.Bogdanova@igc.by.

Буракова Арина Александровна – мл. науч. сотрудник. Институт генетики и цитологии НАН Беларуси (ул. Академическая, 27, 220072, Минск, Республика Беларусь). E-mail: arina.burakova@mail.ru.

Булгак Александр Григорьевич – член-корреспондент НАН Беларуси, д-р мед. наук, профессор, гл. науч. сотрудник. РНПЦ «Кардиология» (ул. Розы Люксембург, 110Б, 220036, Минск, Республика Беларусь). E-mail: bulag@mail.ru.

Байда Александр Васильевич – канд. мед. наук, заведующий кафедрой. Белорусская медицинская академия последипломного образования (ул. Петруся Бровки, 3, 220013, Минск, Республика Беларусь). E-mail: alvab@tut.by.

Зотова Ольга Владимировна – канд. мед. наук, врач-кардиолог, научный сотрудник. РНПЦ «Кардиология» (ул. Розы Люксембург, 110Б, 220036, Минск, Республика Беларусь). E-mail: doxtar@tut.by.

Добыш Ольга Игоревна – мл. науч. сотрудник. Институт генетики и цитологии НАН Беларуси (ул. Академическая, 27, 220072, Минск, Республика Беларусь). E-mail: olyadobysh@gmail.com.

Information about the authors

Lemesh Valentina A. – Ph. D. (Biology), Assistant professor, Head of the Laboratory. Institute of Genetics and Cytology of the National Academy of Sciences of Belarus (27, Akademicheskaya Str., 220072, Minsk, Republic of Belarus). E-mail: v.lemesh@igc.by.

Kipen Viachaslau N. – Ph. D. (Biology), Leading Researcher. Institute of Genetics and Cytology of the National Academy of Sciences of Belarus (27, Akademicheskaya Str., 220072, Minsk, Republic of Belarus). E-mail: v.kipen@igc.by.

Bahdanava Marina V. – Ph. D. (Biology), Leading researcher. Institute of Genetics and Cytology of the National Academy of Sciences of Belarus (27, Akademicheskaya Str., 220072, Minsk, Republic of Belarus). E-mail: M.Bogdanova@igc.by.

Burakova Arina A. – Junior researcher. Institute of Genetics and Cytology of the National Academy of Sciences of Belarus (27, Akademicheskaya Str., 220072, Minsk, Republic of Belarus). E-mail: arina.burakova@mail.ru.

Bulgak Alexander G. – Corresponding Member, D. Sc. (Medicine), Professor, Chief researcher. Scientific and Practical Centre «Cardiology» (110B, Roza Lyuksemburg Str., 220036, Minsk, Republic of Belarus). E-mail: bulag@mail.ru.

Bayda Alexander V. – Ph. D. (Medicine), Head of the Laboratory. Belarusian Medical Academy of Postgraduate Education (3, P. Brovka Str., 220013, Minsk, Republic of Belarus). E-mail: alvab@tut.by.

Zotova Volga V. – Ph. D. (Medicine), Cardiologist, Researcher. Scientific and Practical Centre «Cardiology» (110B, Roza Lyuksemburg Str., 220036, Minsk, Republic of Belarus). E-mail: doxtar@tut.by.

Dobysh Volga I. – Junior researcher. Institute of Genetics and Cytology of the National Academy of Sciences of Belarus (27, Akademicheskaya Str., 220072, Minsk, Republic of Belarus). E-mail: olyadobysh@gmail.com.