GERMLINE MUTATIONS IN PATIENTS WITH ORAL MUCOSAL LEUKOPLAKIA AND SQUAMOUS CELL CARCINOMA: A PROSPECTIVE OBSERVATIONAL STUDY

Natalia A. Karpuk[1], Sergey P. Rubnikovich[2], Oksana Ch. Mazur[3], Ivan V. Zhyltsov[1], Ivan Yu. Karpuk[1], Alena P. Mikhalenka[3]

[1]Vitebsk State Order of Peoples' Friendship Medical University, Frunze Ave., 27, Vitebsk, 210009, Republic of Belarus
[2]Belarusian State Medical University, Dzerzhinski Ave., 83, Minsk, 220116, Republic of Belarus
[3]Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, Akademicheskaya str., 27, Minsk, 220072, Republic of Belarus

ABSTRACT

Background. The number of studies devoted to the molecular genetics of oral mucosal leukoplakia and squamous cell carcinoma is small, while the obtained results are usually preliminary in nature. We can assume the existence of region-specific pathogenic genetic variants involved in the development of oral mucosal leukoplakia and squamous cell carcinoma. With the knowledge of such variants, it would become possible to develop PCR (polymerase chain reaction) and NGS (next-generation sequencing) test systems for the detection of clinically significant germline mutations. **Objectives** — to identify pathogenic germline genetic variants in patients with oral mucosal leukoplakia accompanied by grade 1 epithelial dysplasia, as well as oral mucosal squamous cell carcinoma, using new-generation sequencing. **Methods.** Study design: prospective, observational, cross-sectional, without a control group. The sample included patients (48 persons) of either sex (18 years of age or older) with the following proven and morphologically confirmed diagnoses: oral mucosal leukoplakia accompanied by grade 1 squamous intraepithelial neoplasia of epithelium (24 people) and oral mucosal squamous cell carcinoma (24 people), who sought medical care at the Vitebsk Regional Clinical Dental Center and Vitebsk Regional Clinical Oncological Center in 2019–2020. The identified pathogenic and presumably pathogenic genetic variants involved in the development of these diseases were quantitatively assessed. The study was conducted at the Shareable Core Facilities GENOME of the Institute of Genetics and Cytology of the National Academy of Sciences of Belarus. In order to isolate deoxyribonucleic acid (DNA) from blood samples, a QIAamp DNA FFPE Tissue Kit (Qiagen, Germany) was used. The preparation of DNA libraries and sequencing were carried out by means of an Illumina NextSeq 550 sequencing system (Illumina, Inc., USA) using an Illumina Nextera DNA Exome kit (USA). Bioinformatic analysis was conducted using Illumina BaseSpace specialized software (USA) and Galaxy Project (Galaxy Community, an international non-profit project) in accordance with current guidelines. The obtained data were statistically processed employing specialized software packages Statistica 12 (StatSoft, Inc., USA) and MedCalc 18.9.1 (MedCalc Software, Ltd, Belgium). **Results.** Next-generation whole-exome sequencing of deoxyribonucleic acid samples isolated from the blood of patients with oral mucosal leukoplakia and squamous cell carcinoma has been conducted in the Republic of Belarus for the first time. The total number of unique germline genetic variants in the exome of both groups of patients was shown to be very high, yet most of them were not pathogenic. In the examined patients, the majority of germline mutations were found to be localized only in 19 exome genes: *MAP2K3, DNAH5, HSPG2, OBSCN, SYNE1, HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-A, HLA-B, PKD1L2, TTN, AHNAK2, PDE4DIP, MUC3A, MUC4, MUC12, MUC16,* and *MUC17*. In both clinical groups, the greatest number of genetic variants (> 40% of the total number) was detected in *MUC3A, MUC4, MUC12,* and *MUC16,* responsible for the synthesis of the glycoprotein mucin family. **Conclusion.** Oral mucosal leukoplakia and squamous cell carcinoma can arise from the pathogenic variants of *MUC3A, MUC4, MUC12*, and *MUC16*.

Keywords: genes, germline mutations, leukoplakia, cancer, oral mucosa

Conflict of interest: One of the authors (Prof. Rubnikovich, Dr. Sci. (Med.)) is a member of the Editorial Board of the Kuban Scientific Medical Bulletin. The authors are unaware of any other potential conflict of interest associated with this manuscript.

Compliance with ethical standards:
The study was approved by the Clinical Trial Ethics Committee (Minutes No. 2 of January 18, 2020) of the Vitebsk State Order of Peoples' Friendship Medical University (Frunze Ave., 27, Vitebsk, Republic of Belarus).

Author contributions: Karpuk N.A., Rubnikovich S.P., Mazur O.Ch., Zhyltsov I.V., Karpuk I.Yu., Mikhalenka A.P. — concept formulation and study design; Karpuk N.A., Mazur O.Ch., Karpuk I.Yu., Mikhalenka A.P. — data collection; Karpuk N.A., Rubnikovich S.P., Zhyltsov I.V. — analysis and interpretation of the obtained results; Karpuk N.A., Zhyltsov I.V., Karpuk I.Yu., Mikhalenka A.P. — literature review and statistical analysis; Karpuk N.A., Mazur O.Ch. — drafting of the manuscript and preparation of its final version; S.P. Rubnikovich, Zhyltsov I.V., Karpuk I.Yu., Mikhalenka A.P. — critical revision of the manuscript for valuable intellectual content.

ГЕРМИНАЛЬНЫЕ МУТАЦИИ У ПАЦИЕНТОВ ПРИ ЛЕЙКОПЛАКИИ И РАКЕ СЛИЗИСТОЙ ОБОЛОЧКИ РОТОВОЙ ПОЛОСТИ: ПРОСПЕКТИВНОЕ НАБЛЮДАТЕЛЬНОЕ ИССЛЕДОВАНИЕ

Н.А. Карпук[1], С.П. Рубникович[2], О.Ч. Мазур[3], И.В. Жильцов[1], И.Ю. Карпук[1], Е.П. Михаленко[3]

[1] Учреждение образования «Витебский государственный ордена Дружбы народов медицинский университет», пр-т Фрунзе, д. 27, 210009, г. Витебск, Республика Беларусь

[2] Учреждение образования «Белорусский государственный медицинский университет», пр. Дзержинского, д. 83, 220116, г. Минск, Республика Беларусь

[3] Государственное научное учреждение «Институт генетики и цитологии Национальной академии наук Беларуси», ул. Академическая, д. 27, 220072, г. Минск, Республика Беларусь

АННОТАЦИЯ

Введение. Количество исследований, посвященных молекулярной генетике лейкоплакий и плоскоклеточного рака слизистой оболочки ротовой полости, невелико, а результаты их, как правило, носят предварительный характер. Можно предполагать, что существуют региональные особенности патогенных генетических вариантов, ассоциированных с развитием лейкоплакий слизистой оболочки ротовой полости и плоскоклеточного рака слизистой оболочки ротовой полости. Знание подобных вариантов позволило бы разработать ПЦР (полимеразной цепной реакции) — и NGS-тест-системы (Next-generation sequencing) для выявления клинически значимых герминальных мутаций. **Цель исследования —** используя метод высокопроизводительного секвенирования, идентифицировать патогенные герминальные генетические варианты у пациентов с лейкоплакиями слизистой оболочки полости рта и дисплазией эпителия 1-й степени, а также с плоскоклеточным раком слизистой оболочки полости рта. **Методы.** Дизайн исследования: проспективное, основанное на наблюдении, поперечное (cross-sectional), без контрольной группы. В выборку включены пациенты (48 человек), любого пола 18 лет и старше с установленными и морфологически подтвержденными диагнозами: лейкоплакия слизистой оболочки ротовой полости с плоскоклеточной интраэпителиальной неоплазией эпителия 1-й степени (24 человека) и плоскоклеточный рак слизистой оболочки ротовой полости (24 человека), обратившиеся за медицинской помощью в учреждения здравоохранения «Витебский областной клинический стоматологический центр» и «Витебский областной клинический онкологический диспансер» в 2019–2020 гг. Проведена количественная оценка найденных патогенных и вероятно патогенных генетических вариантов, ассоциированных с развитием данных заболеваний. Исследование выполнялось на базе Центра коллективного пользования «Геном» Института генетики и цитологии Национальной академии наук Беларуси. Для выделения дезоксирибонуклеиновой кислоты (ДНК) из образцов крови использовали диагностический набор QIAamp DNA FFPE Tissue Kit (Qiagen, Германия). Подготовку ДНК-библиотек и секвенирование выполняли при помощи секвенатора Illumina Next Seq 550 (Illumina, Inc, США) с использованием набора реагентов для полноэкзомного секвенирования Illumina Nextera DNA Exome (США). Биоинформационный анализ был выполнен с использованием специализированного программного обеспечения Illumina Base Space (США) и Galaxy Project (The Galaxy Community, некоммерческий международный проект) в соответствии с актуальными рекомендациями. Статистическая обработка данных выполнялась при помощи специализированных программных пакетов Statistica (версия 12) (StatSoft, Inc., США) и MedCalc (версия 18.9.1) (Med Calc Software Ltd, Бельгия). **Результаты.** Впервые в Республике Беларусь выполнено высокопроизводительное полноэкзомное секвенирование образцов дезоксирибонуклеиновой кислоты, выделенных из крови пациентов с лейкоплакиями и плоскоклеточным раком слизистой оболочки полости рта. Показано, что общее количество уникальных герминальных генетических вариантов в экзоме пациентов с обоими заболеваниями весьма велико, но бóльшая их часть не является патогенными. Установлено, что большинство герминальных мутаций у изученных пациентов локализовано всего в 19 генах экзома: *MAP2K3, DNAH5, HSPG2, OBSCN, SYNE1, HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-A, HLA-B, PKD1L2, TTN, AHNAK2, PDE4DIP, MUC3A, MUC4, MUC12, MUC16, MUC17*. Наибольшее количество генетических вариантов (>40% от их общего количества) у пациентов из обеих изученных клинических групп было выявлено в генах *MUC3A, MUC4, MUC12* и *MUC16*, ответственных за синтез семейства гликопротеидов-муцинов. **Заключение.** Развитие лейкоплакии слизистой оболочки ротовой полости и плоскоклеточного рака слизистой оболочки ротовой полости может быть ассоциировано с патогенными вариантами генов *MUC3A, MUC4, MUC12* и *MUC16*.

Ключевые слова: гены, герминальные мутации, лейкоплакия, рак, слизистая оболочка ротовой полости

INTRODUCTION

Polymerase chain reaction (PCR) is widely used as a routine molecular genetic method for diagnosing neoplasms, offering several advantages, specifically, high diagnostic sensitivity and specificity, as well as a relatively low cost of a single study. However, PCR diagnostics can detect only known mutations in a limited set of genes, which restricts the diagnostic and prognostic capabilities of this method [1].

The specified disadvantages can be avoided when using new-generation sequencing (NGS), which can identify the primary genetic sequence for any genes of interest. A significant disadvantage of this method consists in the high cost per sequencer run; however, this drawback is overcome by the use of the so-called targeted sequencing, i.e., sequencing of a pre-selected set of genes. This type of sequencing is significantly cheaper than sequencing the entire genome or a large part of it, as well as a PCR detection of a single genetic abnormality [2, 3].

According to world statistics, oncological diseases are currently the second most frequent cause of death in developed countries and the third cause of death in developing countries. Oral mucosal and lip cancer account for 3% of all cancers. It is known that 11% of all oral mucosal diseases have a high risk of malignant transformation [4]. At the initial stages, the process is reversible, while timely and rational treatment of oral mucosal diseases accompanied by epithelial dysplasia prevents the development of malignant neoplasms in the oral mucosa[1] [5]. Noteworthy is that the five-year survival rate of patients with malignant neoplasms of oral mucosa does not exceed 50% [6].

The number of studies devoted to the molecular genetics of oral mucosal leukoplakia (OML) and squamous cell carcinoma (OMSCC) is small, while the obtained results are usually preliminary in nature. It has been shown that oral mucosal squamous cell carcinoma (OMSCC) can be associated with mutations in the *NOTCH* family [7]; *Mcm2* (accompanied by the increased expression of this gene) [8]; *TP53* (a pathogenic mutation of this gene *TP53Arg72Pro* is described) [9]; *FBXL5*, *UGT2B15*, *UGT2B28*, *KANSL1*, *GSTT1* and *DUSP22* [10]; the *RAS* family (*Ha-ras*, *Ki-ras*, and *N-ras*) [11], *FAT1* and *COL9A1* (genetic variants *rs28647489* and *rs550675*, respectively) [12], etc.

In the Republic of Belarus, such studies have not been conducted before. We can assume the existence of region-specific pathogenic genetic variants involved in the development of OML and OMSCC. With the knowledge of such variants, it would become possible to develop PCR and NGS test systems for the detection of clinically significant germline mutations, which, in turn, would expand and supplement existing protocols for the care of patients with oral mucosal diseases, specifically in terms of ensuring early diagnosis of dysplastic processes in the oral mucosal epithelium and predicting their course and outcome specifics. In theory, an analysis examining the individual profile of pathogenic mutations involved in dysplastic and neoplastic processes helps to personalize treatment regimens for patients, thus achieving maximum treatment efficacy.

The paper **aims** to identify pathogenic germline genetic variants in patients with OML accompanied by grade 1 epithelial dysplasia, as well as in OMSCC patients, using new-generation sequencing.

MATERIALS

Study design

A prospective, observational, cross-sectional, without a control group of 48 blood samples from patients with oncological diseases of the oral mucosa.

Study conditions

The study was conducted at the Shareable Core Facilities GENOME of the Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, engaging highly qualified specialists from the Laboratory of Molecular Genetics, Laboratory of Human Genetics, and the Bioinformatics Sector of this state research institution.

---

[1] Gubaidulina E.Ya., Tsegel'nik L.N., Luzina V.V., Chergeshtov Yu.I. *A Practical Guide to the Outpatient Care in Surgical Dentistry*, Moscow, MIA, 2007.

Eligibility criteria
Inclusion criteria
Patients of either sex (18 years of age or older) with morphologically confirmed OML accompanied by grade 1 squamous intraepithelial neoplasia of epithelium or with morphologically confirmed OMSCC who signed an informed consent to participate in the study.

Exclusion criteria
Patients under 18 years old, pregnant women, patients with a severe course or decompensation of concomitant somatic diseases, as well as patients who refused to sign an informed consent to participate in the study.

Removal criteria
A change in the diagnosis, pregnancy detection, or decompensation of one or several concomitant somatic diseases in patients from the sample, which occurred in the course of the study.

Description of the eligibility criteria (diagnostic criteria)
Diagnostic criteria for patients with OML accompanied by grade 1 squamous intraepithelial neoplasia of epithelium are as follows: the main morphological element observed on the seemingly unchanged oral mucosa is a pearly grayish-white patch having distinct edges, which constitutes an irregular epithelial opacification; such patches are not raised as compared to the surrounding oral mucosal areas and cannot be scraped off; in the affected areas, the mucous membrane forms folds; histologically, the following phenomena are observed: hyperkeratosis, weakly expressed acanthosis, an atypical transformation of squamous epithelium, affecting not more than one third of its thickness; luminescence study shows the affected area in a blue glow; optical coherence tomography (OCT) yields a layered image, differentiating between two horizontally arranged layers.

The diagnostic criteria for patients with OMSCC are as follows: presence of a neoplasm or a long-term nonhealing ulcer (possibly having necrotic plaque) in the oral cavity; long, slow course of the disease not amenable to conservative treatment; failure of conservative treatment; an increase in the nidus size despite adequate treatment; hardening around or at the base of the nidus; bleeding and bad breath; dense, enlarged, painless regional lymph nodes; histologically, malignant epithelial cells are detected, which may be arranged in bundles, cords, or irregularly shaped nests. These cells are similar to the stratified epithelium. The neoplasm destroys the basement membrane forcing its way into the underlying connective tissue.

Selection of group members
During the study, 24 blood samples were collected from patients with OML and 24 blood samples were collected from patients with OMSCC.

Target parameters in the study
Main parameter in the study
Number of detected pathogenic and presumably pathogenic genetic variants linked to the development of OML accompanied by grade 1 squamous intraepithelial neoplasia of epithelium and OMSCC.

Additional parameters in the study
Additional parameters are not provided in this study.

Methods for measuring the target parameters
Laboratory test methods
Peripheral venous blood was drawn from the cubital vein in the morning on an empty stomach into 10 ml vacuum tubes filled with K2 EDTA (dipotassium ethylenediaminetetraacetic acid), followed by centrifugation to obtain plasma, which was deep frozen for subsequent storage (t° = −80°).

DNA was isolated using a QIAamp DNA FFPE Tissue Kit (Qiagen, Germany). All operations involved in preparing DNA libraries for sequencing were performed step by step in strict accordance with the directions for use provided for the Illumina Nextera DNA Exome kit (Illumina, Inc., USA)[2,3] [13, 14].

The Nextera DNA Exome kit is designed for sequencing the so-called exome, a part of the human genome that includes exons, i.e., sequences that are directly translated into proteins. The human exome comprises approximately 180,000 exons, which corresponds to approximately 1% of the entire genome or 30 million nucleotide pairs [15]. Nevertheless, exome mutations account for up to 85% of all mutations causing various diseases, both congenital and acquired [16]. The kit requires ≈ 40 ng of DNA for sample preparation while providing an analytical sensitivity of > 96% and an analytical specificity of 99.9998% (when detecting single-nucleotide polymorphisms).

A whole-exome sequencing of 48 altered tissue samples (24 OML tissue samples and 24 OMSCC tissue samples) was conducted using an Illumina NextSeq 550 sequencing system (Illumina, Inc., USA). This study was conducted to detect congenital (germline) genetic variants (mutations) most likely involved in the development of OML and OMSCC. These genetic variants are present in 100% of body cells, which means that they can be detected in leukocyte DNA; in this case, with ≥ 85% of reads covered at 20× depth, the Nextera DNA Exome kit provides sufficient coverage to reliably detect and document them.

Bioinformatic analysis
A bioinformatic analysis examining the whole-exome DNA sequencing was conducted using Illumina BaseSpace specialized software (USA) and Galaxy Project (the Galaxy Community, an international non-profit project within the open science movement) in accordance with current guidelines [17, 18, 19].

Variables (predictors, confounders, and effect modifiers)
The only predictors of OML and OMSCC development examined in the present study are pathogenic and presumably pathogenic germline genetic variants detected in the exome of patients with the respective diseases. Pathogenic variants of regulatory genes, as well as epigenetic modifications of exome genes (e.g., their methylation) not considered in this study, can serve as effect modifiers. Another effect modifier could be

---

the simultaneous presence of several pathogenic exome gene variants in the patient. This study may be confounded by unaccounted external influences (e.g., smoking habits or human papillomavirus effect) that can cause dysplasia or a malignant transformation of oral mucosal epithelial cells without the formation of pathogenic (oncogenic) exome gene variants, in particular, by causing their epigenetic modifications or damaging regulatory genes.

Statistical procedures
Principles behind sample size determination

The required sample size was determined by factoring in Type I error probability level of $(\alpha) \leq 0.05$, which is standard for biomedical research. The desirable power of the study was 90%. We also proceeded from the assumption that pathogenic genetic variants involved in the formation of neoplasms occur extremely rarely in the population (not more often than in 0.001% of cases), while in the cells of pathologically altered tissues, these variants should occur in at least 30% of samples. Due to the initial assumption that pathogenic (oncogenic) genetic variants occur much more frequently in tissue affected by OML and OMSCC than in the reference human genome that was used for the read alignment, we employed the *SampleSizeCalculation* function (TwoProportions, Z-test) of Statistica 12 (StatSoft, Inc., USA) having a 1-tailed hypothesis option in order to determine the required sample size. Thus, the estimated sample size amounted to 23 people; however, it was increased to 24 people as a single Nextera DNA Exome kit allows 12 DNA samples to be processed.

Statistical methods

The obtained data were statistically processed using specialized software packages Statistica 12 (StatSoft, Inc., USA) and MedCalc 18.9.1 (MedCalc Software, Ltd, Belgium).

The central tendency and the spread of analyzed quantitative indicators were described in the form of median-quartile characteristics: median, 25th and 75th quartiles. Categorical variables were compared using the χ^2 test and Fisher's exact test; the statistical significance of differences in quantitative characteristics was determined using the Mann–Whitney U test. Genetic variants statistically linked to the development of oral mucosal squamous cell carcinoma were detected using Spearman's Rho, as well as the logistic regression analysis. The regression analysis included parameters having a significance level of $p \leq 0.05$. In order to assess the influence of individual genetic variants on the potential development of the pathology under study, odds ratios (OR) and hazard ratios (HR), as well as their 95% confidence intervals (CI), were determined. In all cases, the identified correlations were considered statistically significant at a significance level of $p < 0.05$, with $p \leq 5 \times 10^{-8}$ as the optimal significance level, generally recognized among bioinformaticians, which unambiguously indicates the presence of a relationship between a genetic variant and a phenotype [20].

RESULTS
Sampling

The sample included patients with proven and confirmed diagnoses of OML and OMSCC who had sought medical care

at the Vitebsk Regional Clinical Dental Center (OML) and the Vitebsk Regional Clinical Oncological Center (OMSCC). Patients were included in both samples in the order in which they had sought medical care, provided they met the inclusion criteria and did not fit the exclusion criteria (without any additional conditions). The block diagram of the study design is presented in Fig. 1.

Characteristics of the study sample (groups)

A total of 24 patients with the morphologically confirmed diagnosis of OML accompanied by grade 1 squamous intraepithelial neoplasia of epithelium (15 men and 9 women) were included in the study. The average age of the patients was 59 years (min — 42 years old, max — 72 years old, 95% CI: 57–65 years old). In all cases, squamous cell leukoplakia was observed, which is the most common form of leukoplakia in the population.

The study also includes 24 patients diagnosed with oral mucosal cancer (13 men and 11 women). The average age of the patients was 60.5 years (min — 38 years old, max — 75 years old, 95% CI: 55–65 years old). All considered cases revealed a primary tumor; in addition, all patients were diagnosed with squamous cell carcinoma.

Main study result
Quality assessment of whole-exome sequencing

The proportion of GC bases in the read genetic sequences ranges from 49 to 51%, averaging 50%; the average length of reads amounts to approximately 148 bases, which best corresponds to the capabilities of the Illumina NextSeq 550; average proportions of nucleotide sequences are as follows: read at least once — 97.8%, ≥ 5 times — 91.8%, ≥ 10 times — 83.7%, ≥ 30 times — 52.9%; median of the read coverage — 35.4X, average coverage — 42.5X, and proportion of successfully aligned nucleotide sequences in the total exome — 99.9%.

Considering that in this case, DNA sequencing was performed to identify germline variants, excessive read coverage is not required, while the most important indicator in this situation (of those presented in the table) is the average proportion of nucleotide sequences read ten or more times, which should be $\geq 60\%$ (in this case, 83.7%, which indicates that the quality of our sequencing results is high enough).

Results of bioinformatic analysis examining data from whole-exome sequencing of patients' blood samples

The bioinformatic analysis of sequencing results revealed certain genetic variants in almost each of the 24,918 human exome genes that can be sequenced using the Illumina Nextera DNA Exome kit; the total number of detected unique genetic variants amounted to 124,475 in the OML patient group and 152,767 in the OMSCC patient group (an average of 11,688 mutant genes and ≈33,600 genetic variants in each examined patient). However, the overwhelming majority of detected nucleotide sequence variants proved to be of no clinical significance, having no substantial effect on either protein synthesis or protein function.

Nevertheless, a significant number of variants were found in a relatively small number of genes. In patients with OML, these were (in ascending numerical order of variants) *MAP2K3* (72 variants detected on average), *DNAH5* (315 variants),
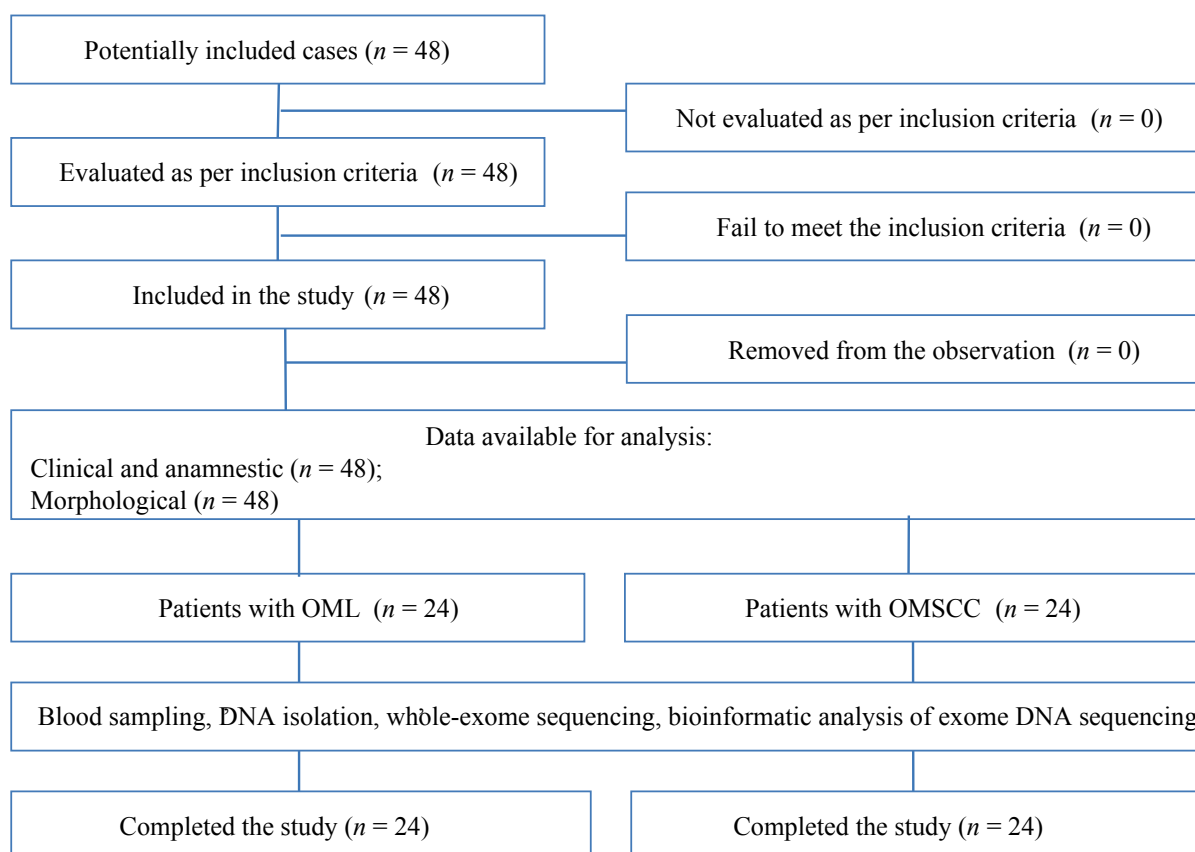
```
┌─────────────────────────────────────────┐
│ Potentially included cases (n = 48)      │
└─────────────────────────────────────────┘
                   │              ┌──────────────────────────────────────────────┐
                   │──────────────│ Not evaluated as per inclusion criteria (n = 0)│
                   │              └──────────────────────────────────────────────┘
┌─────────────────────────────────────────┐
│ Evaluated as per inclusion criteria (n = 48)│
└─────────────────────────────────────────┘
                   │              ┌──────────────────────────────────────────────┐
                   │──────────────│ Fail to meet the inclusion criteria (n = 0)  │
                   │              └──────────────────────────────────────────────┘
┌─────────────────────────────────────────┐
│ Included in the study (n = 48)           │
└─────────────────────────────────────────┘
                   │              ┌──────────────────────────────────────────────┐
                   │──────────────│ Removed from the observation (n = 0)         │
                   │              └──────────────────────────────────────────────┘
```

Data available for analysis:
Clinical and anamnestic (n = 48);
Morphological (n = 48)

Patients with OML (n = 24)   Patients with OMSCC (n = 24)

Blood sampling, DNA isolation, whole-exome sequencing, bioinformatic analysis of exome DNA sequencing

Completed the study (n = 24)   Completed the study (n = 24)

Fig. 1. Block diagram of the study design.
Note: the block diagram was created by the authors (as per STROBE recommendations); Abbreviations: OML — oral mucosal leukoplakia; OMSCC — oral mucosal squamous cell carcinoma; DNA — deoxyribonucleic acid.
Рис. 1. Блок-схема дизайна проведенного исследования.
Примечание: блок-схема выполнена авторами (согласно рекомендациям STROBE). Сокращения: ЛСОРП — лейкоплакия слизистой оболочки ротовой полости; ПРСОРП — плоскоклеточный рак слизистой оболочки ротовой полости; ДНК — дезоксирибонуклеиновая кислота.

HSPG2 (383 variants), *OBSCN* (731 variants), *SYNE1* (733 variants), *HLA-DRB1* (766 variants), *HLA-DQB1* (842 variants), *TTN* (911 variants), *AHNAK2* (1029 variants), *HLA-A* (1038 variants), *PDE4DIP* (1210 variants), *MUC12* (1293 variants), *MUC3A* (1561 variants), *MUC4* (1680 variants), and *MUC16* (1955 variants) — a total of 14,519 variants in 15 genes (Fig. 2).

A similar situation was observed in patients with OMSCC: in this group, variants were most frequently found in *MAP2K3* (176 variants), *MUC17* (206 variants), *SYNE1* (596 variants), *PKD1L2* (625 variants), *AHNAK2* (838 variants), *HLA-DRB1* (856 variants), *HLA-B* (888 variants), *HLA-DQB1* (1033 variants), *HLA-A* (1057 variants), *PDE4DIP* (1214 variants), *HLA-DQA1* (1280 variants), *MUC12* (1299 variants), *MUC4* (1433 variants), *MUC3A* (1539 variants), and *MUC16* (1864 variants) — a total of 14,904 variants in 15 genes (Fig. 3).

Clearly, the greatest number of genetic variants was detected in the same genes for both OLM and OMSCC patients, which suggests that there are no significant differences in the localization of germline mutations in these patient groups. Table 1 provides data on the statistical significance of differences in

the occurrence frequency of the identified nucleotide sequence variants in identical genes between the groups of patients with OML and OMSCC.

It follows from Table 1 that the specified genetic variants are generally significantly more frequent in patients with OML than in OMSCC patients; however, the differences in the absolute and relative occurrence frequency of germline mutations between these groups are small, while the statistical significance of these differences results from the large total number of identified genetic variants, which almost certainly makes this difference clinically insignificant.

Noteworthy is that in both patient groups, *MUC3A*, *MUC4*, *MUC12*, and *MUC16* feature the highest number of the identified variants (44.7% and 41.2% in the OML and OMSCC patient groups, respectively); these genes encode various mucin proteins that play an important role in the formation of protective mucosal barriers on epithelial surfaces, as well as participating in epithelial renewal and differentiation. Insufficient production or decreased functional activity of mucins can lead to chronic damage of oral mucosal epithelial cells, which, in its turn, can trigger both keratosis (process underly-
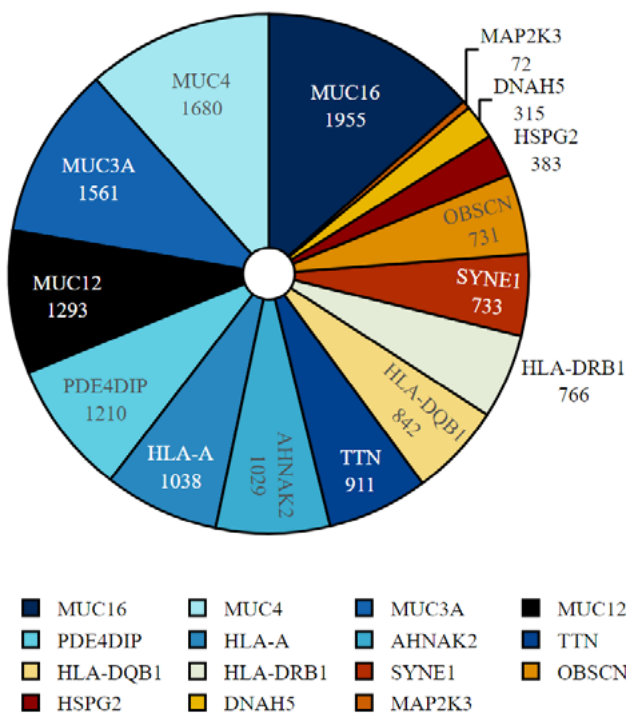
Fig. 2. Genes having the largest number of variants in patients with oral mucosal leukoplakia.
Note: the figure was created by the authors.
Рис. 2. Гены с наибольшим количеством вариантов у пациентов с лейкоплакией слизистой оболочки полости рта.
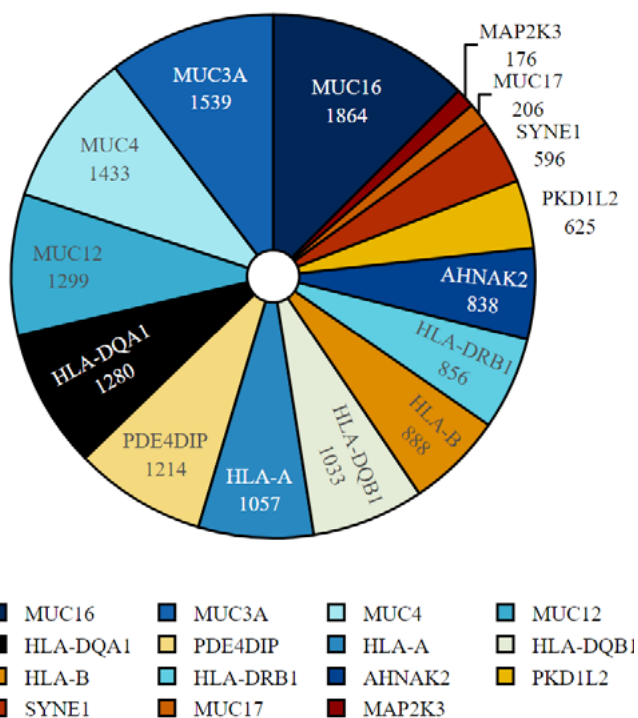*Примечание: рисунок выполнен авторами.*

Fig. 3. Genes having the largest number of variants in patients with oral mucosal squamous cell carcinoma.
Note: the figure was created by the authors.
Рис. 3. Гены с наибольшим количеством вариантов у пациентов с плоскоклеточным раком слизистой оболочки ротовой полости.
Примечание: рисунок выполнен авторами.

Table 1. Statistical significance of differences in the occurrence frequency of genetic variants in different exome genes between the groups of patients with oral mucosal leukoplakia and squamous cell carcinoma
Таблица 1. Статистическая значимость различий частоты встречаемости генетических вариантов в различных генах экзома в группах пациентов с лейкоплакией слизистой оболочки полости рта и плоскоклеточным раком слизистой оболочки ротовой полости

Gene	OML		OMSCC		Statistical significance of differences as per the chi-squared test, p
	Number of genetic variants	% of the total number of genetic variants ($n = 124{,}475$)	Number of genetic variants	% of the total number of genetic variants ($n = 152{,}767$)	
MAP2K3	72	0.058	176	0.12	<0.0001*
SYNE1	733	0.59	596	0.39	<0.0001*
HLA-DRB1	766	0.62	856	0.56	0.059
HLA-DQB1	842	0.68	1033	0.68	0.99
AHNAK2	1029	0.83	838	0.55	<0.0001*
HLA-A	1038	0.83	1057	0.69	<0.0001*
PDE4DIP	1210	0.97	1214	0.79	<0.0001*
MUC12	1293	1.04	1299	0.85	<0.0001*
MUC3A	1561	1.25	1539	1.01	<0.0001*
MUC4	1680	1.35	1433	0.94	<0.0001*
MUC16	1955	1.57	1864	1.22	<0.0001*

*Note: the table was compiled by the authors; * – the difference is statistically significant ($p \leq 0.05$). Abbreviations: OML — oral mucosal leukoplakia; OMSCC — oral mucosal squamous cell carcinoma.*
*Примечание: таблица составлена авторами; * — разница статистически значима ($p \leq 0,05$). Сокращения: ЛСОРП — лейкоплакия слизистой оболочки полости рта; ПРСОРП — плоскоклеточный рак слизистой оболочки ротовой полости.*

ing OML) and a malignant transformation of epithelial cells [21, 22, 23].

Additional study results

All results obtained in the course of the study are presented in the previous sections.

DISCUSSION

Main findings of the study

The total number of unique germline genetic variants was shown to be quite high in the exome of both patient groups — at least 124,475 in the group of OML patients and at least 152,767 in the group of OMSCC patients (an average of 11,688 altered genes and ≈33,600 genetic variants in each examined patient). In the patients from the studied sample (both OML and OMSCC), the majority of germline mutations were found to be localized only in 19 exome genes: *MAP2K3*, *DNAH5*, *HSPG2*, *OBSCN*, *SYNE1*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-A*, *HLA-B*, *PKD1L2*, *TTN*, *AHNAK2*, *PDE4DIP*, *MUC3A*, *MUC4*, *MUC12*, *MUC16*, and *MUC17*. In both clinical groups, the greatest number of genetic variants was detected in *MUC3A*, *MUC4*, *MUC12*, and *MUC16*, responsible for the synthesis of the glycoprotein mucin family (44.7% and 41.2% in the OML and OMSCC patient groups, respectively). Mucins play an important role in the formation of protective mucosal barriers on epithelial surfaces, as well as participating in epithelial renewal and differentiation [24]. Insufficient production or decreased functional activity of mucins can lead to chronic damage of oral mucosal epithelial cells, which may trigger both keratosis and a malignant transformation of epithelial cells [25].

Research limitations

Given that it is impossible to significantly increase the sample size of patients due to the high cost of the study, the identified relationships between the pathogenic genetic variants of exome genes and the development of OML and OMSCC may be ambiguous. The present study did not examine the pathogenic variants of regulatory genes and the epigenetic modifications of exome genes at all levels, which excludes an analysis examining the effect of the specified phenomena on the development of OML and OMSCC. The high cost of whole-exome sequencing, as well as the complexity involved in the bioinformatic analysis of obtained data, limits the use of this method

in clinical practice even when certain pathogenic exome gene variants are clearly linked to the development of the diseases under study.

Interpretation of the study results

A preliminary analysis of the obtained results suggests that the number of exome gene variants detected in OML and OMSCC patients as compared to the reference human genome is rather high, yet most of these variants are presumably benign and have no effect on the development of the considered diseases. Nevertheless, it seems that the large number of detected variants for *MUC3A*, *MUC4*, *MUC12*, and *MUC16* is not accidental: these genes are responsible for the formation of mucins that play an important role in the formation of protective mucosal barriers on epithelial surfaces, as well as participating in epithelial renewal and differentiation. Insufficient production or decreased functional activity of mucins can lead to the chronic damage of oral mucosal epithelial cells, which, in turn, may trigger keratosis and dysplasia, as well as a malignant transformation of epithelial cells. In order to assess the pathogenicity of the identified genetic variants and determine their involvement in the development of OML and OMSCC, further bioinformatics analysis of the obtained sequencing data is required.

CONCLUSION

Next-generation whole-exome sequencing of DNA samples isolated from the blood of patients with OML and OMSCC has been conducted in the Republic of Belarus for the first time. In the examined sample, the majority of germline mutations were found to be localized only in 19 exome genes; the greatest number of genetic variants in patients from both clinical groups were detected in *MUC3A*, *MUC4*, *MUC12*, and *MUC16* genes, i.e., 44.7% and 41.2% in the OML and OMSCC patient groups, respectively. Mucin genes encode epithelial proteins of the glycoprotein mucin family, playing an important role in the formation of protective mucosal barriers on epithelial surfaces, as well as participating in epithelial renewal and differentiation. The occurrence of pathological processes in oral mucosa may be attributed to the pathogenic variants of such genes, as the insufficient production or decreased functional activity of mucins may trigger both keratosis and a malignant transformation of epithelial cells.

REFERENCES / СПИСОК ЛИТЕРАТУРЫ

1. Hameed M. Molecular diagnosis of soft tissue neoplasia: clinical applications and recent advances. *Expert. Rev. Mol. Diagn.* 2014;14(8):961-977. DOI: 10.1586/14737159.2014.946909
2. Chen M., Zhao H. Next-generation sequencing in liquid biopsy: cancer screening and early detection. *Hum. Genomics.* 2019;13(1):34. DOI: 10.1186/s40246-019-0220-8
3. Acosta A.M., Al Rasheed M.R.H., Pins M.R., Borgen K.R., Panchal D., Rogozinska M., Wiley E.L., Behm F.G., Mohapatra G. The role of next-generation sequencing in the differential diagnosis of composite neoplasms. *Hum. Pathol.* 2018;81:78-88. DOI: 10.1016/j.humpath.2018.06.022
4. Kalavrezos N., Scully C. Mouth Cancer for Clinicians. Part 1: Cancer. *Dent. Update.* 2015;42(3):250-252, 255-256, 259-2560. DOI: 10.12968/denu.2015.42.3.250
5. Abati S., Bramati C., Bondi S., Lissoni A., Trimarchi M. Oral Cancer and Precancer: A Narrative Review on the Relevance of Early Diagno-

sis. *Int. J. Environ. Res. Public. Health.* 2020;17(24):9160. DOI: 10.3390/ijerph17249160
6. Ranganathan K., Kavitha L. Oral epithelial dysplasia: Classifications and clinical relevance in risk assessment of oral potentially malignant disorders. *J. Oral. Maxillofac. Pathol.* 2019;23(1):19-27. DOI: 10.4103/jomfp.JOMFP_13_19
7. Mao L. NOTCH mutations: multiple faces in human malignancies. *Cancer Prev. Res (Phila).* 2015;8(4):259-261. DOI: 10.1158/1940-6207.CAPR-15-0063
8. Gouvêa A.F., Santos Silva A.R., Speight P.M., Hunter K., Carlos R., Vargas P.A., de Almeida O.P., Lopes M.A. High incidence of DNA ploidy abnormalities and increased Mcm2 expression may predict malignant change in oral proliferative verrucous leukoplakia. *Histopathology.* 2013;62(4):551-562. DOI: 10.1111/his.12036
9. Zarate A.M., Don J., Secchi D., Carrica A., Galindez Costa F., Panico R., Brusa M., Barra J.L., Brunotto M. Study of the TP53 codon 72

polymorphism in oral cancer and oral potentially malignant disorders in Argentine patients. *Tumour. Biol.* 2017;39(5):1010428317699113. DOI: 10.1177/1010428317699113

10. Ribeiro I.P., Marques F., Barroso L., Rodrigues J., Caramelo F., Melo J.B., Carreira I.M. Genomic profile of oral squamous cell carcinomas with an adjacent leukoplakia or with an erythroleukoplakia that evolved after the treatment of primary tumor: A report of two cases. *Mol. Med. Rep.* 2017;16(5):6780-6786. DOI: 10.3892/mmr.2017.7428

11. Krishna A., Singh S., Singh V., Kumar V., Singh U.S., Sankhwar S.N. Does Harvey-Ras gene expression lead to oral squamous cell carcinoma? A clinicopathological aspect. *J. Oral. Maxillofac. Pathol.* 2018;22(1):65-72. DOI: 10.4103/jomfp.JOMFP_246_17

12. Chung C.M., Hung C.C., Lee C.H., Lee C.P., Lee K.W., Chen M.K., Yeh K.T., Ko Y.C. Variants in FAT1 and COL9A1 genes in male population with or without substance use to assess the risk factors for oral malignancy. *PLoS One.* 2019;14(1):e0210901. DOI: 10.1371/journal.pone.0210901

13. Huss W.J., Hu Q., Glenn S.T., Gangavarapu K.J., Wang J., Luce J.D., Quinn P.K., Brese E.A., Zhan F., Conroy J.M., Paragh G., Foster B.A., Morrison C.D., Liu S., Wei L. Comparison of SureSelect and NexteraExome Capture Performance in Single-Cell Sequencing. *Hum. Hered.* 2018;83(3):153-162. DOI: 10.1159/000490506

14. Díaz-de Usera A., Lorenzo-Salazar J.M., Rubio-Rodríguez L.A., Muñoz-Barrera A., Guillen-Guio B., Marcelino-Rodríguez I., García-Olivares V., Mendoza-Alvarez A., Corrales A., Íñigo-Campos A., González-Montelongo R., Flores C. Evaluation of Whole-Exome Enrichment Solutions: Lessons from the High-End of the Short-Read Sequencing Scale. *J. Clin. Med.* 2020;9(11):3656. DOI: 10.3390/jcm9113656

15. Corominas J., Smeekens S.P., Nelen M.R., Yntema H.G., Kamsteeg E.J., Pfundt R., Gilissen C. Clinical exome sequencing-Mistakes and caveats. *Hum. Mutat.* 2022;43(8):1041-1055. DOI: 10.1002/humu.24360

16. Chalmers Z.R., Connelly C.F., Fabrizio D., Gay L., Ali S.M., Ennis R., Schrock A., Campbell B., Shlien A., Chmielecki J., Huang F., He Y., Sun J., Tabori U., Kennedy M., Lieber D.S., Roels S., White J., Otto G.A., Ross J.S., Garraway L., Miller V.A., Stephens P.J., Frampton G.M. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med.* 2017;9(1):34. DOI: 10.1186/s13073-017-0424-2

17. Kanzi A.M., San J.E., Chimukangara B., Wilkinson E., Fish M., Ramsuran V., de Oliveira T. Next Generation Sequencing and Bioinformatics Analysis of Family Genetic Inheritance. *Front Genet.* 2020;11:544162. DOI: 10.3389/fgene.2020.544162

18. Feltes B.C., Poloni J.F., Nunes I.J.G., Faria S.S., Dorn M. Multi-Approach Bioinformatics Analysis of Curated Omics Data Provides a Gene Expression Panorama for Multiple Cancer Types. *Front. Genet.* 2020; 11: 586602. DOI: 10.3389/fgene.2020.586602

19. Fox A.J., Hiemenz M.C., Lieberman D.B., Sukhadia S., Li B., Grubb J., Candrea P., Ganapathy K., Zhao J., Roth D., Alley E., Loren A., Morrissette J.J. Next Generation Sequencing for the Detection of Actionable Mutations in Solid and Liquid Tumors. *J. Vis. Exp.* 2016;(115):52758. DOI: 10.3791/52758

20. Buzdugan L., Kalisch M., Navarro A., Schunk D., Fehr E., Bühlmann P. Assessing statistical significance in multivariable genome wide association analysis. *Bioinformatics.* 2016;32(13):1990-2000. DOI: 10.1093/bioinformatics/btw128

21. Kashyap B., Kullaa A.M. Regulation of mucin 1 expression and its relationship with oral diseases. *Arch. Oral. Biol.* 2020;117:104791. DOI: 10.1016/j.archoralbio.2020.104791

22. Kumar M.H., Sanjai K., Kumarswamy J., Keshavaiah R., Papaiah L., Divya S. Expression of MUC1 mucin in potentially malignant disorders, oral squamous cell carcinoma and normal oral mucosa: An immunohistochemical study. *J. Oral. Maxillofac. Pathol.* 2016;20(2):214-218. DOI: 10.4103/0973-029X.185916

23. Villa A., Celentano A., Glurich I., Borgnakke W.S., Jensen S.B., Peterson D.E., Delli K., Ojeda D., Vissink A., Farah C.S. World Workshop on Oral Medicine VII: Prognostic biomarkers in oral leukoplakia: A systematic review of longitudinal studies. *Oral. Dis.* 2019;25 Suppl 1(Suppl 1):64-78. DOI: 10.1111/odi.13087

24. Bhatia R., Gautam S.K., Cannon A., Thompson C., Hall B.R., Aithal A., Banerjee K., Jain M., Solheim J.C., Kumar S., Batra S.K. Cancer-associated mucins: role in immune modulation and metastasis. *Cancer Metastasis. Rev.* 2019;38(1-2):223-236. DOI: 10.1007/s10555-018-09775-0

25. Kasprzak A., Adamek A. Mucins: the Old, the New and the Promising Factors in Hepatobiliary Carcinogenesis. *Int. J.Mol. Sci.* 2019;20(6):1288. DOI: 10.3390/ijms20061288

INFORMATION ABOUT THE AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Natalia A. Karpuk — Cand. Sci. (Med.), Assoc. Prof.; General Dentistry and Prosthodontic Dentistry Department with a course at the Faculty of Advanced Training and Retraining, Vitebsk State Order of Peoples' Friendship Medical University.

Карпук Наталья Анатольевна — кандидат медицинских наук, доцент; доцент кафедры общей и ортопедической стоматологии с курсом факультета повышения квалификации и переподготовки кадров учреждения образования «Витебский государственный ордена Дружбы народов медицинский университет».

https://orcid.org/0000-0001-9991-7034

Sergey P. Rubnikovich — Dr. Sci. (Med.), corresponding member of the National Academy of Sciences of Belarus; Prof.; Rector of Belarusian State Medical University.

Рубникович Сергей Петрович — доктор медицинских наук, член-корреспондент Национальной академии наук Беларуси; профессор; ректор учреждения образования «Белорусский государственный медицинский университет».

https://orcid.org/0000-0002-7450-3757

Oksana Ch. Mazur — Researcher, Laboratory of Ecological Genetics and Biotechnology, Institute of Genetics and Cytology of the National Academy of Sciences of Belarus.

Мазур Оксана Чеславовна — научный сотрудник лаборатории экологической генетики и биотехнологии государственного научного учреждения «Институт генетики и цитологии Национальной академии наук Беларуси».

https://orcid.org/0000-0002-6093-4548

Ivan V. Zhyltsov — Dr. Sci. (Med.), Prof.; Head of the Department for Evidence-Based Medicine and Clinical Diagnostics at the Faculty of Advanced Training and Retraining, Vitebsk State Order of Peoples' Friendship Medical University.

Жильцов Иван Викторович — доктор медицинских наук, профессор; заведующий кафедрой доказательной медицины и клинической диагностики факультета повышения квалификации и переподготовки кадров учреждения образования «Витебский государственный ордена Дружбы народов медицинский университет».

https://orcid.org/0000-0002-4912-2880

Ivan Yu. Karpuk — Dr. Sci. (Med.), Assoc. Prof.; Dean of the Dental Faculty, Prof. at the Department for Evidence-Based Medicine and Clinical Diagnostics at the Faculty of Advanced Training and Retraining, Vitebsk State Order of Peoples' Friendship Medical University.

Карпук Иван Юрьевич — доктор медицинских наук, доцент; декан стоматологического факультета, профессор кафедры доказательной медицины и клинической диагностики факультета повышения квалификации и переподготовки кадров учреждения образования «Витебский государственный ордена Дружбы народов медицинский университет».

https://orcid.org/0000-0001-9991-7035

Alena P. Mikhalenka — Cand. Sci. (Biol.); Leading Researcher, Laboratory of Ecological Genetics and Biotechnology, Institute of Genetics and Cytology of the National Academy of Sciences of Belarus.

Михаленко Елена Петровна — кандидат биологических наук; ведущий научный сотрудник лаборатории экологической гене-тики и биотехнологии государственного научного учреждения «Институт генетики и цитологии Национальной академии наук Беларуси».
https://orcid.org/0000-0003-4543-2862