








 pismavavilov.ru

doi 10.18699/letvjgb-2026-12-02

Оригинальное исследование

## Прогнозирование индивидуального риска развития рака легкого у больных хронической обструктивной болезнью легких с применением модели машинного обучения «случайный лес»

М.Л. Баканова <sup>1,2</sup> ✉, В.И. Минина <sup>1,3</sup>, А.Н. Глушков <sup>1</sup>, Я.А. Захарова <sup>1,3</sup>,  
А.В. Минин <sup>1,3</sup>, М.В. Катанахова <sup>1,3</sup>, М.Е. Черноусов<sup>3</sup>, Т.Г. Колесникова <sup>3</sup>

**Аннотация.** Рак представляет собой сложное генетическое и эпигенетическое заболевание, которое, несмотря на интенсивные исследования, все еще находится на лидирующих позициях по заболеваемости и смертности. Распространенной онкопатологией, являющейся ведущей причиной смертности от злокачественных новообразований, считается рак легкого (РЛ). Большинство существующих методов диагностики РЛ направлено на обнаружение уже имеющегося опухолевого процесса в легких. Мы предлагаем оценку риска развития РЛ до проявления патологии. Оценка потенциального риска заболевания проводится с использованием классификационной модели машинного обучения «случайный лес», обеспечивающей высокую эффективность и точность. В ходе исследования были проанализированы генетические данные SNP 33 генов на выборке из 300 здоровых мужчин и 214 пациентов с РЛ. Также были собраны анамнестические данные пациентов: возраст, вредные привычки в различных градациях (курение). Выявлены ассоциации риска развития РЛ и полиморфных локусов *XPD* rs13181, *TGFb1* rs1800471, *CHEK2* rs555607708, *EGFR* rs2227984, *EPHX1* rs1051740, *APEX1* rs1130409, *TNFA* rs1800629, *IL1b* rs16944. Разработано программное обеспечение, позволяющее на основе данных конкретного пациента определять риск развития у него рака легкого.

**Ключевые слова:** рак легкого; генетический полиморфизм; машинное обучение; информационные технологии; анализ данных; шахтеры

**Для цитирования:** Баканова М.Л., Минина В.И., Глушков А.Н., Захарова Я.А., Минин А.В., Катанахова М.В., Черноусов М.Е., Колесникова Т.Г. Прогнозирование индивидуального риска развития рака легкого у больных хронической обструктивной болезнью легких с применением модели машинного обучения «случайный лес». *Письма в Вавиловский журнал генетики и селекции*. 2026;12(1):5-14. doi 10.18699/letvjgb-2026-12-02

**Финансирование:** Работа выполнена в рамках государственного задания (№ гос. регистрации 1023032300035-4-3.1.3).

Original article

## Integrated program for prediction of the risk of lung cancer development in working coal mines of Kuzbass using the random forest machine learning model


M.L. Bakanova <sup>1,2</sup> ✉, V.I. Minina <sup>1,3</sup>, A.N. Glushkov <sup>1</sup>, Ya.A. Zakharova <sup>1,3</sup>,  
A.V. Minin <sup>1,3</sup>, M.V. Katanakhova <sup>1,3</sup>, M.E. Chernousov<sup>3</sup>, T.G. Kolesnikova <sup>3</sup>

**Abstract.** Cancer is a complex genetic and epigenetic disease, which, despite intensive research, still occupies a leading position in morbidity and mortality. A common oncopathology, which is the main cause of death from malignant neoplasms, is lung cancer (LC). The bulk of these lung cancer diagnostic methods are aimed at the appearance of an already luminous tumor process in the lungs. We offer an assessment of the risk of developing LC before the manifestation of nature. The assessment is carried out using the classification model of machine learning “random forest”, which allows with high efficiency and to determine the potential risk of developing LC. During the study, genetic data of SNP 33 genes were analyzed for a sample of 300 healthy men and 214 patients with LC. Anamnestic data of patients were also collected: age, bad habits in various gradations (smoking). As a result of the analysis of genetic risk factors using the SNPStats statistical analysis software package, the most significant inheritance models were identified for the polymorphic

<sup>1</sup> Федеральный исследовательский центр угля и углехимии Сибирского отделения Российской академии наук, Кемерово, Россия  
The Federal Research Center of Coal and Coal Chemistry of the Siberian Branch of the Russian Academy of Sciences, Kemerovo, Russia

<sup>2</sup> Кемеровский государственный медицинский университет, Кемерово, Россия  
Kemerovo State Medical University, Kemerovo, Russia

<sup>3</sup> Кемеровский государственный университет, Кемерово, Россия  
Kemerovo State University, Kemerovo, Russia

 mari-bakano@yandex.ru

© Баканова М.Л., Минина В.И., Глушков А.Н., Захарова Я.А., Минин А.В., Катанахова М.В., Черноусов М.Е., Колесникова Т.Г., 2026

loci *XPB* rs13181, *TGF $\beta$ 1* rs1800471, *CHEK2* rs555607708, *EGFR* rs2227984, *EPHX1* rs1051740, *APEX1* rs1130409, *TNF $\alpha$*  rs1800629, *IL1 $\beta$*  rs16944. To implement the claimed output, software was developed based on the data, which allows you to determine the risk of developing its LC in advance.

**Key words:** lung cancer; genetic polymorphism; machine learning; information technology; data analysis; coal mines

**For citation:** Bakanova M.L., Minina V.I., Glushkov A.N., Zakharova Ya.A., Minin A.V., Katanakhova M.V., Chernousov M.E., Kolesnikova T.G. Integrated program for prediction of the risk of lung cancer development in working coal mines of Kuzbass using the random forest machine learning model. *Pisma v Vavilovskii Zhurnal Genetiki i Seleksii = Lett Vavilov J Genet Breed.* 2026;12(1):5-14. doi 10.18699/letvjgb-2026-12-02

**Funding:** This study was funded by the state task 1023032300035-4-3.1.3.

## Введение

Изучение факторов, влияющих на риск развития рака, – одно из приоритетных направлений современной науки. Известно, что риск развития онкологического заболевания существенно возрастает при воздействии канцерогенных веществ, поступающих в организм человека из окружающей среды (Sung et al., 2021). Особую группу риска составляют работники угольной промышленности, подвергающиеся комбинированному воздействию производственных канцерогенов, среди которых лидирующую позицию занимают полициклические ароматические углеводы (ПАУ), тяжелые металлы, угольная пыль и продукты неполного сгорания органического топлива. Совокупность данных факторов создает мощный промотор канцерогенеза, а наличие у пациента хронической обструктивной болезни легких (ХОБЛ) дополнительно увеличивает этот риск.

Многостадийность онкогенеза подтверждается многочисленными исследованиями. Инициация злокачественного роста тесно связана с индивидуальными особенностями функционирования систем защиты соматических клеток. Наследственные вариации в генах, контролирующих репарацию ДНК, метаболизм ксенобиотиков, регуляцию клеточного цикла и апоптоз, рассматриваются как важные факторы, определяющие специфическую чувствительность организма к канцерогенным воздействиям.

Активация канцерогенов в организме осуществляется ферментами семейства цитохромов. В качестве факторов генетической предрасположенности к раку интенсивно исследуются полиморфизмы генов, ответственных за метаболизм ксенобиотиков – *CYP1A1* (цитохром P450 1A1, rs4646903) (Mohammadi et al., 2022), *CYP1A2* (цитохром P450 rs762551) (Iqbal et al., 2024), и за детоксикацию метаболитов – *GSTT1* (глутатион-S-трансфераза тета-1), *GSTM1* (глутатион-S-трансфераза мю-1) (Zhang et al., 2021). Для нейтрализации свободных радикалов в клетке работает целый спектр ферментов антиоксидантной системы, в том числе супероксиддисмутазы (SuperOxide Dismutase, Sod), каталазы (catalase, CAT) и глутатионпероксидазы (glutathione peroxidase, hGPX1) (Bender et al., 2008). Установлено, что значимую роль в формировании наследственной предрасположенности к геномной нестабильности и раку способны играть гены, контролирующие метаболизм фолатов (*MTHFR*, *MTR*, *MTRR*) (Gohari et al., 2021; Zhong et al., 2022). Активируется и мощная система репарации возникших повреждений. По мнению многих авторов, варианты генов *hOGG1* (8-оксогуанин-ДНК-гликозилаза, rs1052133) (Xu et al., 2022), *APEX1* (апуриновая/пиримидиновая эндонуклеаза 1, rs1130409) (Saad et al., 2021), *PAP1* (поли(АДФ-рибоза)-

полимераза 1, rs1136410) (Jin et al., 2020), *XPB* (ксеродерма пигментная, группа D, rs13181) (Zeng et al., 2024) ассоциированы с риском развития злокачественных новообразований. В большинстве цитируемых работ данные рисковые аллели обозначены как приводящие к снижению функциональной активности (loss-of-function) кодируемых ими белков. Рассматривают и роль факторов иммунного ответа в формировании онкологического риска. Значительный интерес в современных исследованиях рака вызывает фактор некроза опухоли-альфа (TNF- $\alpha$ ) как потенциального супрессора опухолей. Полиморфные варианты в гене TNF- $\alpha$  были идентифицированы как факторы риска для нескольких видов рака (Li et al., 2018; Parsaeian et al., 2021; HaghghiKian et al., 2024).

Однако результаты исследований, посвященных ассоциациям отдельных генетических вариантов с риском РЛ, часто противоречивы и зависят от популяционной составляющей когорты пациентов и контроля, что затрудняет их прямое клиническое применение для индивидуального прогнозирования. Отсюда возникает необходимость в комплексном анализе взаимодействия множества факторов, как генетических, так и средовых.

Широкие возможности в прогнозировании рисков формирования различных заболеваний предоставляют методы машинного обучения. Исследователи активно разрабатывают новые диагностические и терапевтические инструменты на основе машинного обучения, ориентированные на конкретные задачи в медицине. Например, крупные «языковые модели» – продвинутые алгоритмы обработки текста – могут прогнозировать развитие заболевания на основе результатов анализов (Акинин, 2024). Важную роль играет кластеризация – метод, который позволяет разделить пациентов на группы, обладающие общими характеристиками (Васяева, Белякова, 2017; Касюк и др., 2021). Алгоритмы кластеризации выявляют скрытые закономерности в медицинских данных (симптомы, анамнез, результаты анализов и т.д.), позволяя врачам более точно классифицировать пациентов и выбирать оптимальную стратегию лечения (Бога-тырь, 2023).

В период пандемии COVID-19 мощное развитие получили методы распознавания и классификации изображений с помощью нейронных сетей. Разработка алгоритмов для автоматической диагностики COVID-19 на основе рентгеновских КТ-снимков грудной клетки стала одним из приоритетных направлений. Такие системы позволяли быстро и точно определять наличие инфекции, снижая нагрузку на медицинский персонал и ускоряя процесс постановки диагноза (Лиманова, Морозов, 2021).

Один из лучших традиционных подходов к машинному обучению – алгоритм случайного леса (random forest), представляющий собой ансамбль деревьев решений, которые обучаются с использованием бэггинга (ансамблевый мета-алгоритм, предназначенный для улучшения стабильности и точности алгоритмов машинного обучения, применяемых в задачах классификации и регрессии), а разделение узлов ограничено случайным подмножеством исходного набора признаков (Becker et al., 2023). Этот алгоритм был выбран в качестве основного инструмента для построения прогностической модели в силу ряда преимуществ, актуальных для нашего исследования: 1) способность обрабатывать данные со смешанными типами признаков (категориальные и числовые); 2) устойчивость к переобучению и выбросам в данных за счет построения ансамбля деревьев; 3) осуществимость оценки важности каждого анализируемого признака (гена или фактора) для конечного прогнозирования; 4) простое применение, без сложной предобработки данных. Существует ряд исследований, в которых доказана эффективность метода в прогнозировании риска сердечно-сосудистых заболеваний и стратификации онкобольных (Weng et al., 2017).

Цель работы – оценить риск развития рака легкого у рабочих угольных шахт Кузбасса, больных ХОБЛ, с помощью комплексной программы прогнозирования «случайный лес» и разработать ПО с дружелюбным интерфейсом.

## Материалы и методы

Были обследованы 514 мужчин, рабочих угольных шахт Кузбасса. Материалом изучения послужила цельная периферическая кровь, забиравшаяся из локтевой вены в асептических условиях с использованием разовых вакуумных систем (пробирки «Вакутейнер» производства Weston Dickinson (BD), США). Образцы крови транспортировали в лабораторию цитогенетики Института экологии человека ФИЦ УУХ СО РАН в термоконтейнере (+4 °C).

Эксперимент соответствовал этическим стандартам биоэтического комитета (одобрен Этическим комитетом ИЭЧ ФИЦ УУХ СО РАН г. Кемерово, протокол № 1 от 01.07.2025), разработанным в соответствии с Хельсинкской декларацией Всемирной ассоциации «Этические принципы проведения научных медицинских исследований с участием человека» с поправками 2000 г. и «Правилами клинической практики в Российской Федерации», утвержденными Приказом Минздрава РФ от 19.06.2003 № 266.

После детального обследования врачами и специалистами Кемеровского областного клинического онкологического диспансера у 214 человек был диагностирован рак легкого. Критериями включения в основную группу служили: подтвержденный патоморфологический диагноз РЛ, принадлежность к этнической группе европеоидов, подписание пациентом добровольного информированного согласия на участие в исследовании, работа на угледобывающем предприятии (шахте) и наличие хронической обструктивной болезни легких (ХОБЛ), подтвержденной данными индивидуальными медицинскими карт. Критериями исключения являлись: установленный диагноз других злокачественных или доброкачественных заболеваний, наличие в анамнезе

**Таблица 1.** Характеристика обследованных групп  
**Table 1.** Characteristics of the studied groups

Характеристика группы	Больные раком легкого, N	Группа сравнения, N
Всего обследовано	214	300
Возраст, лет (min–max)	59.8 (35–78)	56.7 (30–73)
Курящие	179	147
Некурящие	35	153

терапии до начала исследования, а также диагностических процедур в предшествующие три месяца, отказ пациента от участия в исследовании. Все доноры заполняли подробную анкету.

Триста человек не имели признаков онкопатологии в анамнезе и не имели близких родственников (I, II степень родства) с онкозаболеваниями и составили группу сравнения. Отбор индивидов в контрольную группу проводился после их предварительного анкетирования с помощью специально разработанного опросника. Для группы контроля была исключена ХОБЛ на основании данных анкетирования и отсутствия соответствующего диагноза в медицинской документации. Критериями включения в контрольную группу служили: соответствие группе больных по возрасту, полу, этнической принадлежности и территориальному проживанию. Критериями исключения стали: употребление сильнодействующих препаратов и рентгенография в течение трех месяцев до начала исследования. Все доноры подписывали информированное добровольное согласие. Характеристика пациентов и индивидов из контрольной группы представлена в табл. 1.

Выделение геномной ДНК проводили стандартным фенол-хлороформным методом (Великов, 2013) из 300 мкл цельной крови. Образец считали чистым при соотношении поглощения  $A_{260}/A_{280}$  в диапазоне 1.8–2.0 и  $A_{260}/A_{230} > 2.0$ .

Идентификация генетического компонента и множественных генетических взаимодействий была выполнена с помощью пакета ANDSytem путем реконструкции молекулярно-генетической сети, состоящей из генов, белков и их взаимодействий, из которой для анализа были взяты полиморфные локусы *hOGG1* (rs1052133), *PARP1* (rs1136410), *APEX1* (rs1130409), *XPD* (rs13181), *XRCC1* (rs25489, rs25487, rs1799782), *XPG* (rs17655), *XPC* (rs2228001), *ATM* (rs1801516), *NBS1* (rs1805794), *XRCC4* (rs2075685), *Ligase IV* (rs1805389), *CAT* (rs1001179), *TGFβ* (rs1800469), *EPHX1* (rs1051740), *CYP1A1* (rs4646903), *CYP1A2* (rs762551), *GSTM1* (del), *GSTT1* (del), *TP53* (rs1042522), *TERT* (rs2736100), *CYP2D6* (rs35742686), *CYP2E1* (rs2031920), *GSTP1* (rs1138272, rs1695), *XRCC2* (rs3218536), *XRCC3* (rs861539), *MTHFR* (rs1801133), *MTR* (rs1805087), *SOD2* (rs4880), *GPx1* (rs1050450), *TNFA* (rs1800629), *IL1b* (rs16944) с последующим генотипированием. Подбор генов-кандидатов осуществлялся с учетом функционального характера полиморфизма, наличия сопряженности с изменением активности и/или количества соответствующего фермента, а также ассоциаций с различными биологическими эффектами и болезнями.

Варианты генов *hOGG1* (rs1052133), *PARP1* (rs1136410), *APEX1* (rs1130409), *XPD* (rs13181), *XRCC1* (rs25489, rs25487, rs1799782), *XPG* (rs17655), *XPC* (rs2228001), *ATM* (rs1801516), *NBS1* (rs1805794), *XRCC4* (rs2075685), *Ligase IV* (rs1805389), *CAT* (rs1001179), *TGFβ* (rs1800469), *EPHX1* (rs1051740) изучали методом аллель-специфической ПЦР (НПФ «Литех», Москва), а варианты генов *CYP1A1* (rs4646903), *CYP1A2* (rs762551), *GSTM1* (del), *GSTT1* (del), *TP53* (rs1042522), *TERT* (rs2736100), *CYP2D6* (rs35742686), *CYP2E1* (rs2031920), *GSTP1* (rs1138272, rs1695), *XRCC2* (rs3218536), *XRCC3* (rs861539), *MTHFR* (rs1801133), *MTR* (rs1805087), *SOD2* (rs4880), *Gpx1* (rs1050450), *TNFα* (rs1800629), *IL1b* (rs16944) – ПЦР в реальном времени с использованием конкурирующих TaqMan-зондов (ООО «СибДНК», Новосибирск). Результаты генотипирования перепроверялись путем повторного генотипирования 20 % проб.

Математическую обработку полученных данных проводили с использованием программ Statistica 10.0 (StatSoft Inc., США) и онлайн-инструмента SNPStats (<http://bioinfo.iconcologia.net/SNPstats>), предназначенного для анализа ассоциаций генетических полиморфизмов.

Частоту аллелей ( $p_i$ ) и генотипов ( $P_i$ ) изучаемых генов рассчитывали по стандартной формуле

$$P_i = N_i / N,$$

где  $N_i$  – число  $i$ -х генотипов (аллелей);  $N$  – объем выборки.

Сравнение частот аллелей и генотипов генов проводили с помощью четырехпольной таблицы сопряженности с поправкой Йетса на непрерывность вариации ( $\chi^2$ ). Отвергали нулевую гипотезу при достигнутом уровне значимости ( $p$ )  $\leq 0.05$ .

Взаимосвязи полиморфных локусов в аддитивной, доминантной, сверхдоминантной, рецессивной, лог-аддитивной моделях выявляли с помощью SNPStats (<http://bioinfo.iconcologia.net/SNPstats>) и логистической регрессии. Для выбора модели использовали информационный критерий Акаике (AIC).

Для логистической модели экспоненту отдельного коэффициента регрессии ( $\beta$ ) расценивали как отношение шансов (OR). Дополнительно рассчитывали 95 % доверительный интервал. Существенность построенной модели проверялась тестом отношения правдоподобия и его значимости  $p_{adj}$ .  $OR_{adj}$  равный 1 свидетельствовал об отсутствии ассоциации, при  $OR_{adj}$  меньше 1 ассоциацию считали отрицательной, при  $OR_{adj}$  больше 1 – положительной.

Для программного обеспечения, позволяющего прогнозировать риск появления РЛ на основе вариантов генов человека, была использована классификационная модель машинного обучения «случайный лес» (sklearn.ensemble.RandomForestClassifier <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble>).

## Результаты

По итогу анализа генетических факторов риска с помощью SNPStats были выявлены наиболее значимые модели наследования для полиморфных локусов *XPD* rs13181, *TGFb1* rs1800471, *CHEK2* rs555607708, *EGFR* rs2227984, *EPHX1* rs1051740, *APEX1* rs1130409, *TNFα* rs1800629, *IL1b* rs16944, которые были включены в дальнейшую работу.

Распределение частот генотипов перечисленных полиморфных локусов в группах больных раком легкого и здоровых рабочих угольных шахт Кузбасса представлено в табл. 2.

Результаты OR для полиморфных локусов *XPD* rs11813, *TGFb1* rs1800471, *CHEK2* rs555607708, *EGFR* rs2227984, *EPHX1* rs1051740, *APEX1* rs1130409, *TNFα* rs1800629, *IL1b* rs16944 и статуса курения приведены в табл. 3.

Корреляционный анализ исследуемого набора данных и его признаков показал (рис. 1), что наиболее сильные линейные связи имеются между РЛ и девятью показателями (информация о статусе курения пациента, результаты генотипирования значимых полиморфных локусов *XPD* rs13181, *TGFb1* rs1800471, *CHEK2* rs555607708, *EGFR* rs2227984, *EPHX1* rs1051740, *APEX1* rs1130409, *TNFα* rs1800629, *IL1b* rs16944), которые были отобраны для построения классификационных моделей.

Метод «случайный лес» подразумевает создание совокупности кросс-валидированных решающих деревьев. Каждое из таких деревьев проходит обучение на подвыборке данных, включающей информацию лишь по части биомаркеров и наблюдений, и валидируется на подвыборке, не использованной для его построения (бэггинг). Решающее дерево как метод классификации строит свою иерархию на основе расчета определенной метрики. Такой метрикой здесь выступает метрика под названием «неопределенность Джини». Признак, у которого эта метрика минимальна, становится вершиной дерева (точкой разделения выборки на две подвыборки). Выбирается минимальное значение, так как чем ближе оно к нулю, тем меньше неопределенности и больше уверенности в классе. В модели использовалось 11 деревьев решений с максимальной глубиной до 5 вершин. Пример одного из 11 деревьев, построенных методом «случайный лес», приведен на рис. 2. Данный метод конструирует несколько решающих деревьев на различных частях обучающей выборки и, объединяя результаты отдельных деревьев, создает модель таким образом, чтобы снизить дисперсию и повысить общую устойчивость и точность прогноза.

*TGFb1* был выбран отправной точкой, потому что имеет наименьший показатель неопределенности Джини среди остальных признаков на первой итерации.

На основании предсказаний каждого из построенных деревьев решений пациент причисляется к одной из групп (здоровые доноры или пациенты с раком легкого), финальное предсказание классификатора определяется большинством голосов построенных деревьев.

Оценка параметров моделей (обучение) производилась на обучающей выборке и была направлена на минимизацию предсказательных ошибок алгоритма. Валидация обученных моделей заключалась в оценке их предсказательной способности на тестовой выборке. Результаты работы модели представлены на рис. 3. График отображает частоту встречаемости генотипа в зависимости от того, больной (status 1) или здоровый (status 0) человек. По вертикальной оси отложено количество записей (в %), которые содержат тот или иной генотип по подвыборкам здоровых/больных людей. В результате на стадии обучения коэффициент точности достиг 80 %, а на тестовой выборке – 90 %.

**Таблица 2.** Распределение частот генотипов изученных полиморфных локусов в группах больных раком легкого и здоровых рабочих угольных шахт Кузбасса

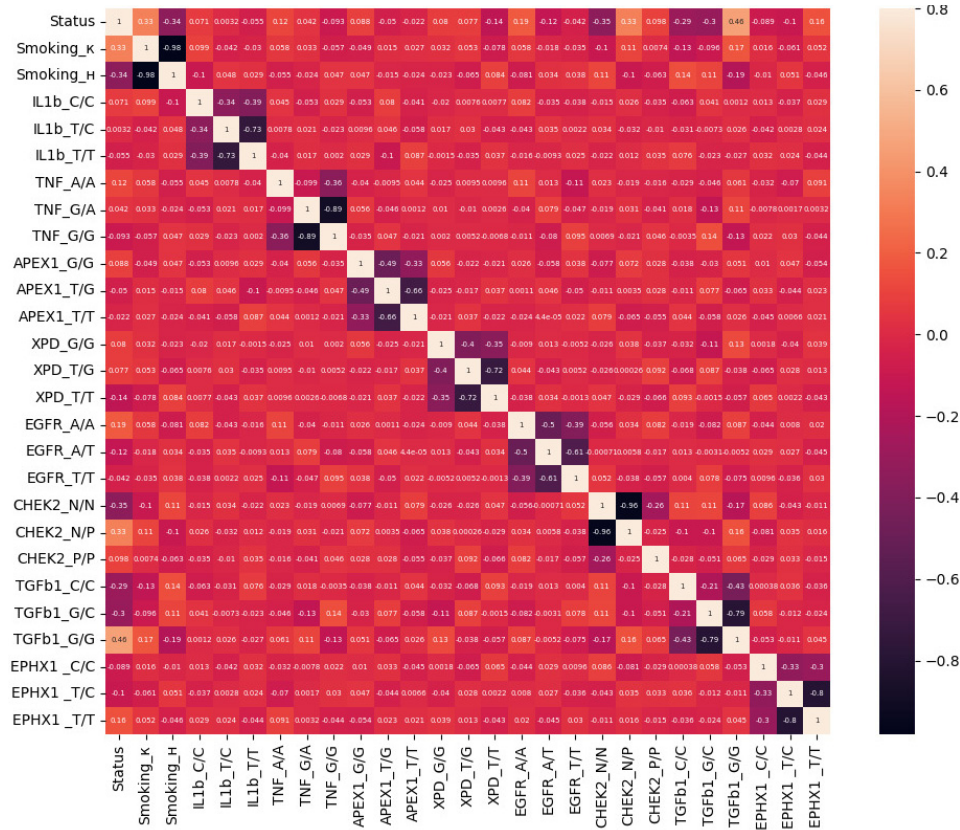
**Table 2.** Distribution of genotype frequencies of the studied polymorphic loci in groups of patients with lung cancer and healthy workers in Kuzbass coal mines

Локус	Генотипы	Контроль, N (%)	Больные раком легкого, N (%)	$\chi^2$	p
<i>TGFb1</i> rs1800471	GG	125 (41.7)	192 (89.7)	119.9	0.001
	GC	125 (41.7)	22 (10.3)		
	CC	50 (16.7)	0		
<i>CHEK2</i> rs555607708	GG	298 (99.3)	174 (81.3)	16.3	0.001
	G/delC	2 (0.7)	37 (17.3)		
	delC/delC	0	3 (1.4)		
<i>EGFR</i> rs2227984	TT	97 (32.3)	60 (28)	22.7	0.001
	AT	157 (52.3)	82 (38.3)		
	AA	46 (15.3)	72 (33.6)		
<i>XPB (ERCC2)</i> rs13181	TT	142 (47.3)	62 (29)	16.8	0.00001
	TG	122 (40.7)	110 (51.4)		
	GG	36 (12)	42 (19.6)		
<i>EPHX1</i> rs1051740	TT	102 (34)	110 (51.4)	14.9	0.001
	TC	161 (53.7)	90 (42.1)		
	CC	37 (12.3)	14 (6.5)		
<i>APEX1</i> rs1130409	TT	94 (31.3)	68 (31.8)	4.12	0.043
	TG	161 (53.7)	99 (46.3)		
	GG	45 (15)	47 (22)		
<i>TNFa</i> rs1800629	GG	246 (82)	161 (75.2)	4.896	0.027
	GA	49 (16.3)	41 (19.2)		
	AA	5 (1.7)	12 (5.6)		
<i>IL1b</i> rs16944	T/T	149 (49.7)	87 (40.7)	4.536	0.034
	T/C	114 (38)	86 (40.2)		
	C/C	37 (12.3)	41 (19.2)		

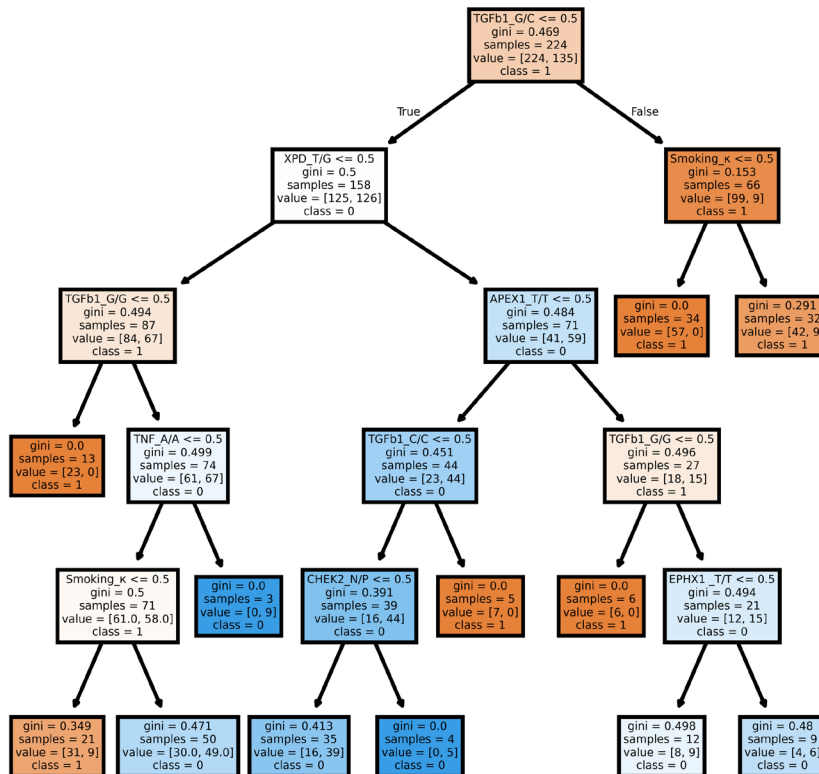
**Таблица 3.** Ассоциации полиморфных локусов *XPB* rs13181, *TGFb1* rs1800471, *CHEK2* rs555607708, *EGFR* rs2227984, *EPHX1* rs1051740, *APEX1* rs1130409, *TNFa* rs1800629, *IL1b* rs16944 и статуса курения с риском развития рака легкого у рабочих угольных шахт Кузбасса, больных ХОБЛ

**Table 3.** Associations of polymorphic loci *XPB* rs13181, *TGFb1* rs1800471, *CHEK2* rs555607708, *EGFR* rs2227984, *EPHX1* rs1051740, *APEX1* rs1130409, *TNFa* rs1800629, *IL1b* rs16944 and smoking status with the risk of developing lung cancer in Kuzbass coal mine workers with COPD

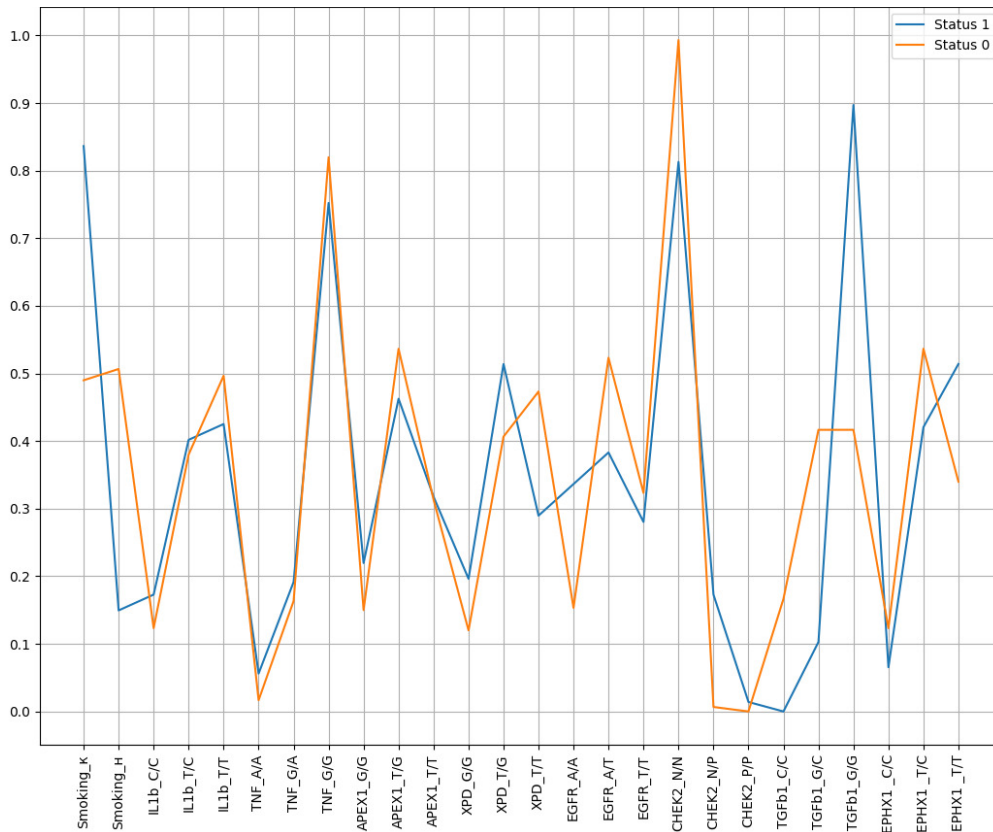
Факторы	OR	CI	p
<i>XPB</i> rs13181	2.20	1.52–3.20	0.000001
Курение (статус)	5.32	3.47–8.16	0.001
<i>TGFb1</i> rs1800471	0.10	0.06–0.17	0.000001
<i>CHEK2</i> rs555607708	34.25	8.18–143.46	0.000001
<i>EGFR</i> rs2227984	2.80	1.83–4.27	0.000001
<i>EPHX1</i> rs1051740	0.56	0.42–0.74	0.000001
<i>TNFa</i> rs1800629	3.50	1.22–10.10	0.01
<i>IL1b</i> rs16944	1.45	1.03–2.05	0.033
<i>APEX1</i> rs1130409	1.59	1.01–2.51	0.04



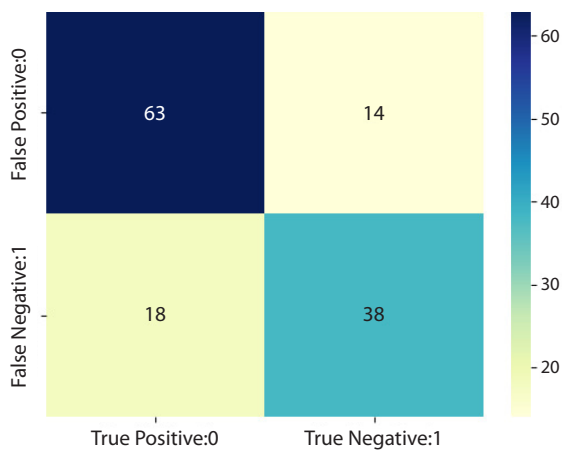
**Рис. 1.** Матрица коэффициентов связей.  
**Fig. 1.** Relationship coefficient matrix.



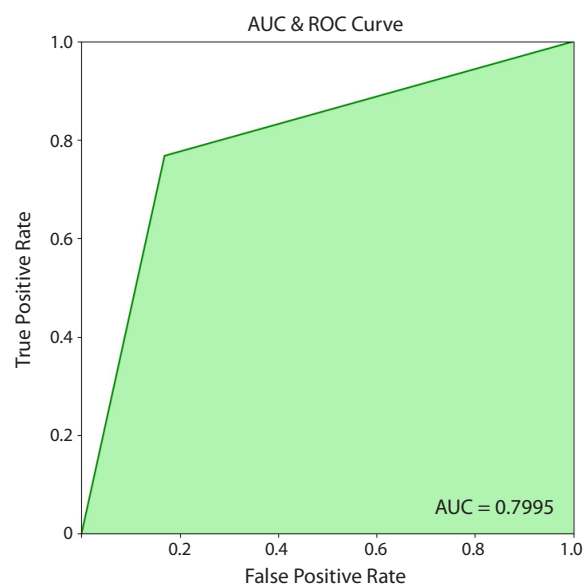
**Рис. 2.** Пример дерева решений, полученных в результате обучения многофакторного классификационного алгоритма «случайный лес» на экспериментальных данных по девяти маркерам.  
**Fig. 2.** An example of a decision tree obtained as a result of training a multifactor random forest classification algorithm on experimental data for 9 markers.



**Рис. 3.** Статистические характеристики выборки.  
**Fig. 3.** Statistical characteristics of the sample.



**Рис. 4.** Матрица погрешностей.  
**Fig. 4.** Error matrix.



**Рис. 5.** График ROC-кривой для оценки предсказательной способности классификационных алгоритмов.  
**Fig. 5.** ROC curve plot for assessing the predictive ability of classification algorithms.

Предсказательная способность многофакторных классификационных моделей оценивалась при помощи ROC-анализа. Для определения эффективности работы классификатора построены матрица погрешностей (рис. 4) и график ROC-кривой (рис. 5) для тестовых прогнозов. Матрица погрешностей наглядно показывает распределение правильных и ошибочных предсказаний модели на тестовой выборке: количество истинно положительных (true positive, TP), истинно отрицательных (true negative, TN), ложнополо-

жительных (false positive, FP) и ложноотрицательных (false negative, FN) случаев. На основе этой матрицы рассчитываются такие метрики, как точность (accuracy), чувствительность (sensitivity) и специфичность (specificity).

На графике ROC-кривой ось *X* представляет долю ложных положительных классификаций, ось *Y* – долю верных положительных классификаций. Количественную интерпретацию ROC дает показатель AUC – площадь, ограниченная ROC-кривой и осью доли ложных положительных классификаций. Для построения модели значение AUC составило 0.7995 (см. рис. 5). Значение AUC > 0.5 свидетельствует, что модель работает лучше случайного угадывания, а значение, приближающееся к 0.8, указывает на ее существенную предсказательную силу и клинически потенциальную значимость. Данные выводы сделаны на основе общепринятой шкалы: значение AUC в диапазоне 0.7–0.8 считается приемлемым, 0.8–0.9 – отличным. Кривая модели стремится вверх и вправо, что также визуально подтверждает ее эффективность.

С помощью классификационной модели машинного обучения «случайный лес» была разработана компьютерная программа прогнозирования риска развития РЛ у рабочих угольных шахт Кузбасса.

Набор данных, использованный для обучения моделей, содержит информацию о 514 работниках угольных предприятий Кемеровской области. Для каждого человека указаны следующие данные:

- информация о генетических вариантах: *IL1b* rs16944, *TNFA* rs1800629, *APEX1* rs1130409, *XPB (ERCC2)* rs13181, *TGFB1* rs1800471, *EGFR* rs2227984, *EPHX1* rs1051740, *CHEK2* rs555607708, которые по результатам проведенных ранее исследований могут быть связаны с развитием рака легких у шахтеров;
- курит ли человек (да/нет);
- наличие заболевания раком легких: болен или здоров;
- если человек болен, то указан диагноз и стадия развития болезни.

Набор данных был разделен на обучающую и тестовую выборки. В тестовую выборку вошло 153 наблюдения, остальные использовались для обучения моделей.

Данные являются подмножеством набора данных, собранных в г. Кемерово.

Для того чтобы результат работы модели был высоким, необходимы: настройка модели; регулирование гиперпараметров, которые отвечают за внутреннюю структуру модели, количество деревьев, глубину деревьев; случайная мера; число признаков для расщепления; минимальное число объектов на листьях и использование для построения деревьев подвыборки с возвращением.

## Обсуждение

В настоящем исследовании осуществлен поиск комплекса факторов, модулирующих индивидуальную чувствительность организма человека к канцерогенным эффектам, который позволяет с высокой эффективностью и точностью определять потенциальный риск формирования рака легкого у рабочих угольных шахт.

В результате молекулярно-генетического анализа вовлеченными в канцерогенез легких были полиморфизмы ключевых генов воспалительных путей, таких как *TNFA* rs1800629, *IL1b* rs16944, *TGFB1* rs1800471.

TNF- $\alpha$  (фактор некроза опухоли альфа) – провоспалительный цитокин, который является неотъемлемой частью

различных биологических функций и, в частности, участвует в патогенезе воспалительных, аутоиммунных и злокачественных новообразований (Li et al., 2018; Parsaeian et al., 2021).

TGF- $\beta$  (трансформирующий фактор роста) входит в семейство многофункциональных цитокинов, регулирующих широкий спектр клеточных реакций. Накопленные данные свидетельствуют о критической роли TGF- $\beta$  как стимулятора опухолевой пролиферации, ангиогенеза и опухолевой инвазии при разных типах рака (Miyazono et al., 2018).

IL1b является важным цитокином, регулирующим экспрессию нескольких генов, вовлеченных в различные воспалительные процессы (Dinarello, 1996, 2002). Устойчивая индукция IL1b усиливает интенсивность воспалительного ответа и создает воспалительную микросреду, благоприятную для инициации и/или промоции опухоли (Dinarello, 2006). Было показано, что ген *IL1b* может быть индуцирован сигаретным дымом в клетках человека *in vitro* (Hellermann et al., 2002).

В нашем исследовании выявлена связь между курением и риском РЛ. Курение – главный фактор риска развития рака легких, вне зависимости от его типа. Канцерогенность табачного дыма на легкие была установлена в ходе эпидемиологических наблюдений, начатых еще в 1950-х годах, и впоследствии получила официальное подтверждение от органов здравоохранения (Malhotra et al., 2016).

При сравнении частот полиморфных вариантов ферментов репарации ДНК больших РЛ с группой здоровых было выявлено различие частот встречаемости аллелей и генотипов генов *XPB* rs13181 и *APEX1* rs1130409.

Ген *APEX1* (apurinic/aprimidinic endonuclease 1) кодирует апуриновую/апиридиновою эндонуклеазу 1, которая представляет собой важный фермент, играющий центральную роль в восстановлении поврежденных оснований и разрывов одноцепочечной ДНК, вызванных окислением или алкилированием, путем удаления AP-сайтов (Izumi et al., 2000).

Белок *XPB* имеет большое значение для распознавания и восстановления различных повреждений ДНК. В ряде исследований обнаружена связь между уровнем этих повреждений и полиморфным вариантом *XPB Gln751Gln* (Wu et al., 2014; Lawania et al., 2019).

Кроме того, получены статистически значимые ассоциации РЛ с полиморфным вариантом, кодирующим трансмембранный рецептор семейства рецепторных тирозинкиназ ErbB (*EGFR* rs2227984). Онкогенная роль *EGFR* при РЛ часто связана с активирующими мутациями в участке ДНК, кодирующем тирозинкиназный домен рецептора (Hynes, Lane, 2005). Полиморфный вариант гена *EGFR* rs2227984, возникающий в результате замены А→Т в позиции 2073, приводит к расщеплению фрагмента «дикого» типа размером 207 п.н. на фрагменты в 150 и 57 п.н. В результате образуется усеченная форма *EGFR*, формирующая укороченный транскрипт без трансмембранного домена, который не может связываться внутри мембраны и не может действовать как передатчик сигнала (Shintani et al., 1999).

Значимые взаимосвязи были получены между РЛ и геном-супрессором клеточного роста *CHEK2*. *CHEK2* кодирует белок checkpoint протеинкиназу 2 (Chk2) – один из ключе-

вых регуляторов клеточного цикла (Hirao et al., 2000; Cai et al., 2009). Есть данные, что мутация *1100delC*, следствием которой является делеция цитозина в 1100-м положении гена *CHEK2*, ассоциирована с повышенным риском опухолей (Meijers-Heijboer et al., 2002).

Проведенный анализ роли полиморфных вариантов ферментов биотрансформации ксенобиотиков показал значимую модель наследования гена *EPHX1*. *EPHX1* представляет собой фермент метаболизма ксенобиотиков – микросомальную эпоксидгидролазу 1. Эпоксидгидролаза-1 гидрирует монозамещенные, 1,1-дизамещенные и цис-1,2-дизамещенные эпоксиды и эпоксиды на циклических системах. Она инактивирует эпоксиды после ферментов первой фазы (Гуляева, Райс, 2005). Было показано, что полиморфизмы одиночных нуклеотидов (SNP) в *EPHX1* влияют на активность фермента, способствуя активации канцерогена (Tilak et al., 2011).

Для прогнозирования индивидуального риска развития РЛ у больных хронической обструктивной болезнью легких использовался мощный современный метод машинного обучения «случайный лес» (случайный лес деревьев решений). Мы применили итеративный способ отбора наиболее важных факторов. Выбор лучшей модели осуществлен по наименьшей ошибке на тестовых данных. Алгоритм метода оформлен на языке R – мощном и универсальном инструменте для статистического анализа и визуализации, обладающем огромным количеством специализированных модулей для биологии (геномика, экология, молекулярная биология и др.). Использование R позволяет существенно сократить время обработки данных, автоматизировать рутинные операции, получить более глубокие результаты и обеспечить воспроизводимость анализа, что критически важно для нашего исследования (Задорожный, 2023).

В настоящее время для оценки риска РЛ широко применяются клинично-демографические модели (например, PLCOm2012, LLP), которые преимущественно основаны на возрасте, стаже курения, индексе массы тела и наличии ХОБЛ. Также стоит упомянуть различные генетические шкалы с ограниченным количеством полиморфизмов генов метаболизма канцерогенов.

В отличие от упомянутых выше подходов, разработанная нами модель позволяет: 1) учесть специфическую когорту профессиональных пациентов (рабочих угольных шахт с ХОБЛ); 2) включить различные генетические маркеры, относящиеся к разным патогенетическим путям, и ключевой средовой фактор – курение; 3) эффективно выявлять нелинейные взаимодействия и оценивать вклад каждого признака. Именно такая комбинация составляет основную новизну работы. Дополнительно модель позволяет определить важность (весомость) того или иного признака (генотипа) в процессе принятия решения. Это следует из алгоритма построения дерева принятия решений, так как оно всегда выбирает самый влияющий признак из имеющихся (остающихся после предыдущих выборов). Таким образом, на первом уровне дерева расположен наиболее важный признак, затем менее важный и далее по убывающей. Собранный нами банк данных в дальнейшем позволит включить дополнительные клинические и анамнестические параметры в будущие итерации модели для ее совершенствования.

## Заключение

В рамках исследования, проведенного на шахтерах Кузбасса с ХОБЛ, предложен метод оценки индивидуального риска развития РЛ. Метод основан на молекулярно-генетическом анализе полиморфных локусов *XPB* rs13181, *TGFB1* rs1800471, *CHEK2* rs555607708, *EGFR* rs2227984, *EPHX1* rs1051740, *APEX1* rs1130409, *TNFA* rs1800629, *IL1b* rs16944 и определении статуса курения пациентов. Полученные результаты анализируются с помощью классификационной модели машинного обучения «случайный лес». Данная модель обучена для прогнозирования высокой или низкой вероятности развития рака легкого.

## Список литературы / References

- Акинин А.А. Применение больших языковых моделей для диагностики. *Научно-технический вестник Поволжья*. 2024;11:138-140 [Akinin A.A. Application of large language models for diagnostics. *Scientific and Technical Volga Region Bulletin*. 2024;11:138-140 (in Russian)]
- Богатырь И.Н. Алгоритмы кластеризации в анализе медицинских данных. *Научно-технический вестник Поволжья*. 2023;12:586-588 [Bogatyr I.N. Clustering algorithms in medical data analysis. *Scientific and Technical Volga Region Bulletin*. 2023;12:586-588 (in Russian)]
- Васяева Д.О., Белякова А.С. Обзор алгоритмов кластеризации с целью обработки медицинских данных. *Алгоритмы, методы и системы обработки данных*. 2017;2(36):17-22 [Vasyaeva D.O., Belyakova A.S. Review of clustering algorithms for the purpose of processing medical data. *Algoritmy, Metody i Sistemy Obrabotki Danykh*. 2017;2(36):17-22 (in Russian)]
- Великов В.А. Молекулярная биология. Саратов, 2013 [Velikov V.A. Molecular Biology. Saratov, 2013 (in Russian)]
- Гуляева Л.Ф., Райс Р.Х. Биологические эффекты токсических соединений. Новосибирск, 2005 [Gulyaeva L.F., Rice R.H. Biological Effects of Toxic Compounds. Novosibirsk, 2005 (in Russian)]
- Задорожный С.С. Статистическая обработка данных на языке R. М., 2023 [Zadorozhny S.S. Statistical Data Processing in the Language of R. Moscow, 2023 (in Russian)]
- Касюк С.Т., Диденко Г.А., Степанова О.А. Кластерный анализ данных медицинских исследований с использованием языка R. *Современная наука: актуальные проблемы теории и практики. Серия: Естественные и Технические Науки*. 2021;4/2:23-32. doi 10.37882/2223-2966.2021.04-2.11 [Kasyuk S.T., Didenko G.A., Stepanova O.A. Cluster analysis of medical research data with R. *Modern Science: Actual Problems of Theory and Practice*. 2021;4/2:23-32. doi 10.37882/2223-2966.2021.04-2.11 (in Russian)]
- Лиманова Н.И., Морозов Д.А. Обзор моделей компьютерного зрения для определения наличия заболевания COVID-19. *Евразийское Научное Объединение*. 2021;4/1:57-59 [Limanova N.I., Morozov D.A. Review of computer vision models for determining the presence of COVID-19 disease. *Yevraziyskoye Nauchnoye Ob'yedineniye*. 2021;4/1:57-59 (in Russian)]
- Becker T., Rousseau A.J., Geubbelmans M., Burzykowski T., Valkenburg D. Decision trees and random forests. *Am J Orthod Dentofacial Orthop*. 2023;164(6):894-897. doi 10.1016/j.ajodo.2023.09.011
- Bender G., Poyner R.R., Reed G.H. Identification of the substrate radical intermediate derived from ethanolamine during catalysis by ethanolamine ammonia-lyase. *Biochemistry*. 2008;47(43):11360-11366. doi 10.1021/bi801316v
- Cai Z., Chehab N.H., Pavletich N.P. Structure and activation mechanism of the CHK2 DNA damage checkpoint kinase. *Mol Cell*. 2009;35(6):818-829. doi 10.1016/j.molcel.2009.09.007
- Dinarello C.A. Biologic basis for interleukin-1 in disease. *Blood*. 1996;87(6):2095-2147. doi 10.1182/blood.V87.6.2095.bloodjournal.8762095
- Dinarello C.A. The IL-1 family and inflammatory diseases. *Clin Exp Rheumatol*. 2002;20(5 Suppl.27):S1-13

- Dinarelli C.A. The paradox of pro-inflammatory cytokines in cancer. *Cancer Metastasis Rev.* 2006;25(3):307-313. doi 10.1007/s10555-006-9000-8
- Gohari M., Moghimi M., Aarafi H., Shajari A., Jafari-Nedooshan J., Lookzadeh M.H., Mirjalili S.R., Neamatzadeh H. Association of MTHFR 677C>T, 1298A>C and MTR 2756A>G polymorphisms with susceptibility to childhood retinoblastoma: a systematic review and meta-analysis. *Fetal Pediatr Pathol.* 2021;40(6):612-625. doi 10.1080/15513815.2020.1721738
- HaghighiKian S.M., Shirinzadeh-Dastgiri A., Ershadi R., Vakili-Ojarood M., Barahman M., Dastgheib S.A., Asadian F., Shiri A., Aghasipour M., Rahmani A., Aghili K., Neamatzadeh H. Correlation of TNF- $\alpha$  polymorphisms with susceptibility to lung cancer: evidence from a meta-analysis based on 29 studies. *BMC Cancer.* 2024;24(1):1113. doi 10.1186/s12885-024-12854-x
- Hellermann G.R., Nagy S.B., Kong X., Lockey R.F., Mohapatra S.S. Mechanism of cigarette smoke condensate-induced acute inflammatory response in human bronchial epithelial cells. *Respir Res.* 2002; 3(1):22. doi 10.1186/rr172
- Hirao A., Kong Y.Y., Matsuoka S., Wakeham A., Ruland J., Yoshida H., Liu D., Elledge S.J., Mak T.W. DNA damage-induced activation of p53 by the checkpoint kinase Chk2. *Science.* 2000;287(5459):1824-1827. doi 10.1126/science.287.5459.1824
- Hynes N.E., Lane H.A. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat Rev Cancer.* 2005;5(5):341-354. doi 10.1038/nrc1609. Erratum in: *Nat Rev Cancer.* 2005;5(7):580
- Iqbal M.S., Sardar N., Peng K., Almutairi L.A., Duan X., Tanvir F., Attia K.A., Zeng G., Gu D. Association between CYP1A2 gene variants -163 C/A (rs762551) and -3860 G/A (rs2069514) and bladder cancer susceptibility. *BMC Cancer.* 2024;24(1):880. doi 10.1186/s12885-024-12553-7
- Izumi T., Hazra T.K., Boldogh I., Tomkinson A.E., Park M.S., Ikeda S., Mitra S. Requirement for human AP endonuclease 1 for repair of 3'-blocking damage at DNA single-strand breaks induced by reactive oxygen species. *Carcinogenesis.* 2000;21(7):1329-1334
- Jin J., Robeson H., Fagan P., Orloff M.S. Association of PARP1-specific polymorphisms and haplotypes with non-small cell lung cancer subtypes. *PLoS One.* 2020;15(12):e0243509. doi 10.1371/journal.pone.0243509
- Lawania S., Singh A., Sharma S., Singh N., Behera D. The multi-faceted high order polymorphic synergistic interactions among nucleotide excision repair genes increase the risk of lung cancer in North Indians. *Mutat Res.* 2019;816-818:111673. doi 10.1016/j.mrfmmm.2019.111673
- Li L., Liu J., Liu C., Lu X. The correlation between TNF- $\alpha$ -308 gene polymorphism and susceptibility to cervical cancer. *Oncol Lett.* 2018; 15(5):7163-7167. doi 10.3892/ol.2018.8246
- Malhotra J., Malvezzi M., Negri E., La Vecchia C., Boffetta P. Risk factors for lung cancer worldwide. *Eur Respir J.* 2016;48(3):889-902. doi 10.1183/13993003.00359-2016
- Meijers-Heijboer H., van den Ouweland A., Klijn J., Wasielewski M., de Snoo A., Oldenburg R., Hollestelle A., ... Futreal P.A., Nathanson K.L., Weber B., Rahman N., Stratton M.R.; CHEK2-Breast Cancer Consortium. Low-penetrance susceptibility to breast cancer due to CHEK2\*1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet.* 2002;31(1):55-59. doi 10.1038/ng879
- Miyazono K., Katsuno Y., Koinuma D., Ehata S., Morikawa M. Intracellular and extracellular TGF- $\beta$  signaling in cancer: some recent topics. *Front Med.* 2018;12(4):387-411. doi 10.1007/s11684-018-0646-8
- Mohammadi H., Momeni Roochi M., Rezaei F., Garajei A., Heidar H., Ghaderi B., Sadeghi M. Association between the CYP1A1 MspI polymorphism and risk of head and neck cancer: a meta-analysis. *Sci Rep.* 2022;12(1):1527. doi 10.1038/s41598-022-05274-z
- Parsaeian S.F., Asadian F., Karimi-Zarchi M., Setayesh S., Javaheri A., Tabatabaie R.S., Dastgheib S.A., Golestanpour H., Neamatzadeh H. A meta-analysis for association of XRCC3 rs861539, MTHFR rs1801133, IL-6 rs1800795, IL-12B rs3212227, TNF- $\alpha$  rs1800629, and TLR9 rs352140 polymorphisms with susceptibility to cervical carcinoma. *Asian Pac J Cancer Prev.* 2021;22(11):3419-3431. doi 10.31557/APJCP.2021.22.11.3419
- Saad A.M., Abdel-Megied A.E.S., Elbaz R.A., Hassab El-Nabi S.E., Elshazli R.M. Genetic variants of APEX1 p.Asp148Glu and XRCC1 p.Gln399Arg with the susceptibility of hepatocellular carcinoma. *J Med Virol.* 2021;93(11):6278-6291. doi 10.1002/jmv.27217
- Shintani S., Matsuo K., Crohin C.C., McBride J., Tsuji T., Donoff R.B., Posner M., Todd R., Wong D.T. Intragenic mutation analysis of the human epidermal growth factor receptor (EGFR) gene in malignant human oral keratinocytes. *Cancer Res.* 1999;59(16):4142-4147
- Sung H., Ferlay J., Siegel R.L., Laversanne M., Soerjomataram I., Jemal A., Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209-249. doi 10.3322/caac.21660
- Tilak A.R., Kumar S., Jain M., Pant M.C., Das B.C., Guleria R., Mittal B., Mathur N., Kumar A. Association of functionally important polymorphism of microsomal epoxide hydrolase gene (EPHX1) with lung cancer susceptibility. *Cancer Invest.* 2011;29(6):411-418. doi 10.3109/07357907.2011.587070
- Weng S.F., Reys J., Garibaldi J.M., Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One.* 2017;12(4):e0174944. doi 10.1371/journal.pone.0174944
- Wu K.G., He X.F., Li Y.H., Xie W.B., Huang X. Association between the XPD/ERCC2 Lys751Gln polymorphism and risk of cancer: evidence from 224 case-control studies. *Tumour Biol.* 2014;35(11):11243-11259. doi 10.1007/s13277-014-2379-x
- Xu H., Zhang M., Hao Z., Liang C. hOGG1 rs1052133 polymorphism and prostate cancer risk: a Chinese case-control study and meta-analysis. *Med Sci Monit.* 2022;28:e938012. doi 10.12659/MSM.938012
- Zeng W., Xu W., Long W. The association between XPD rs13181 and rs1799793 polymorphism and oral cancer risk: evidence from a meta-analysis. *BMC Cancer.* 2024;24(1):738. doi 10.1186/s12885-024-12503-3
- Zhang W.P., Yang C., Xu L.J., Wang W., Song L., He X.F. Individual and combined effects of GSTM1, GSTT1, and GSTP1 polymorphisms on lung cancer risk: a meta-analysis and re-analysis of systematic meta-analyses. *Medicine (Baltimore).* 2021;100(26):e26104. doi 10.1097/MD.00000000000026104
- Zhong G., Luo X., Li J., Liao Y., Gui G., Sheng J. MTRR rs1532268 polymorphism and gastric cancer risk: evidence from a meta-analysis. *J Int Med Res.* 2022;50(5):3000605221097486. doi 10.1177/03000605221097486

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 19.09.2025. После доработки 04.12.2025. Принята к публикации 12.01.2026.