

# Прогнозирование распространения опухоли рака головы и шеи на окружающие лимфатические узлы на основе генетического тестирования

Угаров И.В.<sup>1</sup>, Черных В.Б.<sup>1,2</sup>, Иванов Н.В.<sup>1</sup>, Масленников В.В.<sup>1</sup>, Остапенко Д.К.<sup>1</sup>

1 — ООО «эксДжен Сайбернетикс»

129344, г. Москва, ул. Верхоянская, д.18, корп. 2.

2 — ФГБНУ «Медико-генетический научный центр имени академика Н.П. Бочкова»

115522, г. Москва, ул. Москворечье, д. 1

Рак головы и шеи (HNSC) – полигенное заболевание с тяжелыми инвалидизирующими последствиями и высокой смертностью, занимает 6 место среди онкологических заболеваний по распространенности в мире. Рост заболеваемости у людей до 40 лет, частое диагностирование в продвинутой стадии требуют разработки мер профилактики данного заболевания. Одним из способов решения такой задачи является выявление и использование в практике биомаркеров, выявляющих изменения на ранней доклинической стадии, и разработка подхода для оценки суммарного риска многофакторной патологии. Разработан подход и выявлены рисковые паттерны, характерные для пациентов с раком головы и шеи, математическая модель учитывает полиморфизмы в 17 генах и показывает чувствительность более 16% и специфичность более 93%. Валидация на тестовой выборке показала чувствительность 10% и специфичность 82%, что позволяет считать, что модель выявляет пациентов с высоким риском при наличии ключевых маркеров.

**Ключевые слова:** прогнозирование риска, рак головы и шеи, полиморфизм, математическая модель

**Для цитирования:** Угаров И.В., Черных В.Б., Иванов Н.В., Масленников В.В., Остапенко Д.К. Прогнозирование распространения опухоли рака головы и шеи на окружающие лимфатические узлы на основе генетического тестирования. *Медицинская генетика* 2020; 19(6): 103-105

**DOI:** 10.25557/2073-7998.2020.06.103-105

**Автор для корреспонденции:** Угаров И.В.; e-mail: iugarov@yandex.ru

**Финансирование.** Исследование не имело спонсорской поддержки.

**Конфликт интересов.** Авторы декларируют отсутствие конфликта интересов.

**Поступила:** 20.05.2020

## Predicting the spread of head and neck cancer tumors to surrounding lymph nodes based on genetic testing

Ugarov I.V.<sup>1</sup>, Chernykh V.B.<sup>1,2</sup>, Ivanov N.V.<sup>1</sup>, Maslennikov V.V.<sup>1</sup>, Ostapenko D.K.<sup>1</sup>

1 — xGen Cybernetics

Verkhoyanskaya str., 18, building 2, Moscow, 129344, Russia

2 — Research Centre for Medical Genetics

Moskvorechye str., 1, Moscow, 115522, Russia Head and neck cancer (HNSC) is a polygenic disease with severe disabling consequences and high mortality. It is the 6th most common cancer in the world. The increase in the incidence in people under 40 years of age, frequent diagnosis at an advanced stage requires the development of prevention measures for this disease. One of the ways to solve this problem is to identify and use in practice biomarkers that detect changes at an early preclinical stage and an approach for assessing the total risk of multifactorial pathology. An approach was developed and risk patterns typical for patients with head and neck cancer were identified. The mathematical model takes into account polymorphisms in 17 genes and shows a sensitivity of more than 16% and a specificity of more than 93%. Validation on a test sample showed 10% sensitivity and 82% specificity, which suggests that the prognosis identifies patients with high risk in the presence of key markers.

**Keywords:** risk prediction, head and neck cancer, polymorphism, mathematical model

**For citation:** Ugarov I.V., Chernykh V.B., Ivanov N.V., Maslennikov V.V., Ostapenko D.K. Predicting the spread of head and neck cancer tumors to surrounding lymph nodes based on genetic testing. *Medical genetics*. 2020; 19(6): 103-105 (In Rus)

**DOI:** 10.25557/2073-7998.2020.06.103-105

**Corresponding author:** Ugarov I.V.; e-mail: iugarov@yandex.ru

**Funding.** The study had no sponsorship.

**Conflict of interest.** The authors declare no conflict of interest.

**Accepted:** 20.05.2020

**Р**ак головы и шеи (HNSC) — полигенное заболевание с тяжелыми инвалидизирующими последствиями и высокой смертностью, занимает 6 место среди онкологических заболеваний по распространенности в мире. Известно, что HNSC развивается в течение многих лет и в этот период существует несколько неопластических участков преобразования. В литературе представлено более 200 генов, полиморфизмы в которых ассоциированы с HNSC [1,2], однако отсутствуют рекомендации по выявлению рискованных маркеров. Рост заболеваемости у людей до 40 лет, частое диагностирование в продвинутой стадии требует разработки мер профилактики данного заболевания. Одним из способов решения такой задачи является выявление и использование в практике биомаркеров, выявляющих изменения на ранней доклинической стадии и разработка подхода для анализа результатов NGS, лишённого ранее описанных недостатков и обеспечивающего выявление вариантов, характерных для многофакторной патологии.

Цель работы: создание математической модели для прогнозирования распространения HNSC на близлежащие лимфатические узлы на основе генетического тестирования.

### Материалы и методы

Выполнен анализ реферируемых российских и зарубежных научных статей по следующим ключевым словам: плоскоклеточный рак головы и шеи, предрасположенность, полиморфизм. Для оценки работы модели использовали данные с TumorPortal (<http://www.tumorportal.org/>). Для отладки и обучения математической модели прогноза развития HNSC на основе данных сформированы две независимые группы: пациенты с плоскоклеточным HNSC (n=384) и пациенты с другими нозологиями (n=6362). Из группы пациентов с HNSC для отладки и обучения математической модели распространения опухоли HNSC на близлежащие лимфатические узлы сформированы две группы: пациенты с плоскоклеточным HNSC со стадией N0 (без распространения опухоли в близлежащие лимфатические узлы) (n=143) и пациенты с плоскоклеточным HNSC с распространением опухоли в близлежащие лимфатические узлы (n=162). Для валидации прогноза развития HNSC сформированы две независимые группы: пациенты с плоскоклеточным HNSC (n=238) и пациенты с другими нозологиями (n=1570). Из 238 пациентов с HNSC сформированы две независимые группы для валидации прогноза распространения опухоли HNSC на близлежащие лимфатические узлы: пациенты с плоскоклеточным

HNSC со стадией N0 (без распространения опухоли в близлежащие лимфатические узлы) (n=85) и пациенты плоскоклеточным HNSC с распространением опухоли в близлежащие лимфатические узлы (n=153). Для формирования математической модели и выявления сочетаний генотипов (паттернов), применен набор алгоритмов интеллектуального анализа данных [3], что позволило выявить генетические маркеры, недоступные для анализа стандартными статистическими методами. Система работает по принципу «черного ящика» автоматически выбирая из ряда встроенных алгоритмов, оптимальный для оценки чувствительности и специфичности.

### Результаты работы

Разработан метод, который позволяет выявить скрининговые паттерны многофакторных состояний с учетом полигенной природы заболеваний и неопределенности патогенного эффекта отдельных генетических вариантов.

Паттерны и математическая модель, созданная на основе метода, учитывают генетические варианты в более чем 500 генов с чувствительностью 42% и специфичностью более 90%. В качестве примера сравним гены выявленных паттернов и базы данных Tumorportal. Рисковые паттерны охватывают все высокосignимые мутированные гены по данным Tumorportal, из них 6 генов, являются основными в паттернах (*TP53*, *CASP8*, *CDKN2A*, *FAT1*, *NOTCH1*, *PIK3CA*), остальные входят в группу генов окружения. Валидация на тестовой выборке показала чувствительность 39% и специфичность 92%. Можно предположить, что так как на данный момент отсутствует скрининговый метод для оценки генетической предрасположенности к данному заболеванию, то применение созданного авторами подхода позволяет существенно повысить точность выявления обследуемых, имеющих характерные генетические варианты. и, следовательно, проводить доклиническую профилактику заболеваний.

Также, разработана математическая модель, позволяющая выявить рисковые паттерны распространения опухоли плоскоклеточного HNSC в близлежащие лимфатические узлы. Математическая модель учитывает полиморфизмы в 17 генах и показывает чувствительность более 16% и специфичность более 93%. Интересно, что основным геном в паттернах является *NSD1*, мутации в котором описаны при синдромальном гигантизме и детской миелоидной лейкемии, остальные гены оказывают модифицирующее влияние. К самым частым в паттернах генам окружения относятся *ALK*,

*BCOR, FUBP1, MED23, NUP210L, PIK3R1, RAB3GAP1, SETD2, SMAD4, ZNF217 и ZNF750.* Валидация на тестовой выборке показала чувствительность 10% и специфичность 82%, что позволяет говорить о том, что прогноз выявляет пациентов с высоким риском при наличии ключевых маркеров. Однако выборка недостаточна для достоверной оценки.

Использование данной модели совместно с оценкой предрасположенности к плоскоклеточному HNSC позволит существенно повысить эффективность выявления лиц с повышенным риском разви-

тия и распространения опухоли и профилактики заболевания.

### Литература/ References

1. Kang H., Kiess A., Chung C.H. Emerging biomarkers in head and neck cancer in the era of genomics. *Nat Rev Clin Oncol* 2015; 12: 11–26.
2. Lawrence M. S. et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 2014; 505: 495–501.
3. Witten I.H., Frank E., Hall M.A. *Data Mining: Practical Machine Learning Tools and Techniques*. — 3rd Edition. — Morgan Kaufmann, 2011. — p. 664.