РАЗРАБОТКА ПОДХОДОВ К СОЗДАНИЮ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ДЛЯ ПРЕДСКАЗАНИЯ СКОРОСТИ ПРОЯВЛЕНИЯ ПЕРЕВИТОЙ ОПУХОЛИ В МЫШИНОЙ МОДЕЛИ РМЖ ЧЕЛОВЕКА

Аронов Д.А.^{1,2}, Жуков В.В.³, Моисеева Е.В.¹, Семушина С.Г.¹, Шишкин А.М.⁴,
Боженко В.К.⁴

¹ФГБУН Институт биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук, Москва, e-mail: evmoise@gmail.com;

²ФГБУ «Национальный медицинский исследовательский центр акушерства, гинекологии и перинатологии имени академика В.И. Кулакова» Министерства здравоохранения РФ, Москва;

³Российский университет дружбы народов, Москва;

⁴ФГБУ «Российский научный центр рентгенорадиологии» Минздрава России, Москва

Прогноз темпов роста опухоли необходим для выбора терапевтической стратегии для пациентов с РМЖ. Известно, что фоновые патологии внутренних органов пациента могут определять исход болезни. Мы предположили, что такие патологии могут оказывать влияние и на динамику роста опухоли, а гематологические и биохимические показатели крови и сыворотки, отражающие наличие анемии, нарушений системы гемостаза, заболеваний печени и сердечно-сосудистой системы, могут служить информативными маркерами для прогноза скорости проявления РМЖ. Для проверки этой гипотезы использовали перевиваемую мышиную модель РМЖ человека. У интактных самцов оригинальной линии мышей BLRB-Rb(8.17)11em индивидуально определяли количество эритроцитов, гемоглобина, гематокрит, средний объем эритроцита, среднее содержание гемоглобина в эритроците, среднюю концентрацию гемоглобина в эритроците, ширину распределения эритроцитов, количество тромбоцитов, тромбокрит, средний объем тромбоцитов, ширину распределения тромбоцитов, количество лейкоцитов, долю лимфоцитов, моноцитов, эозинофилов, палочкоядерных нейтрофилов, сегментоядерных нейтрофилов, уровень холестерина, триглицеридов, общего белка, альбумина, мочевины, креатинина, аспартатаминотрансферазы, аланинаминотрансферазы, билирубина, альфа-амилазы, лактатдегидрогеназы, щелочной фосфатазы и креатинкиназы в периферической крови. Через 2 недели самцам перевили спонтанную опухоль сингенной самки. Спустя 2 недели после перевивки самцов разделили на группы с ранним и поздним проявлением опухоли (диаметр больше или меньше 3,5 мм соответственно). Математическая модель, основанная на применении логистической регрессии, позволила получить удовлетворительный уровень классификации мышей на группы по скорости проявления опухоли на основании выявленного набора информативных лабораторных показателей. Полученные результаты демонстрируют, что такие параметры можно использовать для прогноза скорости роста РМЖ, что косвенным образом доказывает влияние фоновых патологий в момент возникновения опухоли на темпы роста РМЖ.

Ключевые слова: рак молочной железы, математическая модель, гематологические показатели крови, биохимические показатели крови.

APPROACHES TO MATHEMATICAL MODEL DEVELOPMENT FOR PREDICTING THE TRANSPLANTED TUMOR MANIFESTATION SPEED IN A MOUSE MODEL OF HUMAN BREAST CANCER

Aronov D.A.^{1,2}, Zhukov V.V.³, Moiseeva E.V.¹, Semushina S.G.¹, Shishkin A.M.⁴,
Bozhenko V.K.⁴

¹Shemyakin-Ovchinnikov Institute of bioorganic chemistry, Russian Academy of Sciences, Moscow, e-mail: evmoise@gmail.com;

²FSBI «National medical research center for obstetrics, gynecology and perinatology» Ministry of healthcare of the Russian Federation, Moscow;

³Peoples' Friendship University of Russia (RUDN University), Moscow;

⁴FSBI «Russian Scientific Center of Roentgenoradiology» Ministry of healthcare of the Russian Federation, Moscow

Tumor growth rate prediction is necessary for developing a therapeutic strategy for breast cancer (BC) patients. Concomitant pathologies of internal organs were shown to determine the patient outcome. We suggested that such pathologies could also affect the dynamics of tumor growth, thus, hematological and biochemical parameters of blood and serum, which reflect the presence of anemia, disorders of the hemostatic system, diseases of the liver and cardiovascular system, could serve as informative markers for predicting the growth of

the tumor. To test this hypothesis, a transplantable mouse model of human BC was used. The number of erythrocytes, hemoglobin, hematocrit, average erythrocyte volume, average erythrocyte hemoglobin content, average erythrocyte hemoglobin concentration, erythrocyte volume distribution, platelet count, plateletcrit, the width of the distribution of platelets, the number of leukocytes, the proportion of lymphocytes, monocytes, eosinophils, band neutrophils, segmented neutrophils, cholesterol, triglycerides, total protein, albumin, urea, creatinine, aspartate aminotransferase, alanine aminotransferase, bilirubin, alpha-amylase, LDH, alkaline phosphatase and creatine kinase were measured in peripheral blood of intact males of the original BLRB-Rb(8.17)1Iem mouse strain. After 2 weeks, the naturally occurred tumor of the syngeneic female was transplanted to males of BLRB mouse strain. Two weeks after tumor inoculation, the males were divided into groups with early and late manifestation of the tumor (diameter greater or less than 3.5 mm). A mathematical model based on the logistic regression allowed us to obtain satisfactory results of classification mice into groups according to the rate of tumor manifestation and based on the set of informative laboratory parameters. These results demonstrate that such parameters could be used to predict the growth rate of breast cancer, which indirectly proves the influence of background pathologies at the time of the onset of the tumor on the growth of breast cancer.

Keywords: breast cancer, mathematical model, hematological blood parameters, biochemical blood parameters

Рак молочной железы (РМЖ) является наиболее распространенным онкологическим заболеванием у женщин в России. Морфологическое разнообразие типов РМЖ определяет различия в скорости роста опухоли у пациентов. Как клинический исход пациента [1], так и исход терапии (в частности, химиотерапии [2]) может зависеть от динамики опухолевого роста, поэтому заблаговременный прогноз роста РМЖ необходим для выбора наиболее оптимального терапевтического подхода для пациента. Наличие фоновых заболеваний при РМЖ, как правило, отягощает течение болезни, хотя до сих пор не выявлено связи между патологией внутренних органов и степенью агрессивности РМЖ [3]. При этом давно известно, что ряд гематологических и биохимических показателей крови и сыворотки, являющихся клиническими маркерами заболеваний органов и систем, можно применять для прогноза исхода пациентов с РМЖ [4, 5]. Также известны изменения таких показателей у женщин с диагностированным РМЖ по сравнению со здоровыми донорами [6, 7] в зависимости от стадии заболевания [8, 9] и проводимой терапии [10, 11]. Мы предположили, что некоторые патологические процессы внутренних органов, о которых можно косвенно судить по гематологическим и биохимическим показателям крови и сыворотки, могут оказывать влияние на скорость проявления первичной опухоли. Для проверки этого предположения в эксперименте *in vivo* и выявления информативных биомаркеров для прогноза скорости проявления опухоли необходимо применять адекватные мышинные модели РМЖ человека. Существенное влияние на рост опухоли оказывают качественный и количественный состав микроокружения на местном уровне, а также возраст, половые гормоны и иммунный статус реципиента на системном уровне. Чтобы нивелировать значение этих факторов, необходимо унифицировать реципиентов в исследовании. Для этого наиболее подходящей моделью можно считать перевиваемую мышиную модель РМЖ с использованием стареющих самцов высокоракковой линии BLRB-Rb(8.17)1Iem (далее – BLRB) одного возраста, содержащихся в конвенциональных (не-SPF) условиях. Таким

образом, перевивка естественно возникшей опухоли от сингенной самки унифицирует показатели опухолевого микроокружения; пол и возраст реципиентов уравнивают их гормональный и иммунный статус, а конвенциональные условия содержания создадут возможность для развития фоновых возраст-ассоциированных патологий внутренних органов, схожих с таковыми у пациентов с РМЖ.

Цель исследования: продемонстрировать возможность создания математической модели для прогноза скорости проявления перевитой опухоли на основании гематологических и биохимических показателей крови и сыворотки, измеренных у интактных мышей до перевивки им опухоли.

Материалы и методы исследования. Использовали перевиваемую мышиную модель РМЖ. У самцов мыши линии BLRB (n=64), характеризующейся высокой частотой естественно возникающего РМЖ у стареющих самок, за две недели до перевивки опухоли взяли образцы крови объемом 200 ± 50 мкл; с помощью гематологического анализатора «MICROS 60 OT-18» (ABX, Франция) измеряли следующие показатели: количество эритроцитов (RBC), гемоглобина (HGB), гематокрит (HCT), средний объем эритроцита (MCV), среднее содержание гемоглобина в эритроците (MCH), среднюю концентрацию гемоглобина в эритроците (MCHC), ширину распределения эритроцитов (RDW), количество тромбоцитов (PLT), тромбоцит (PCT), средний объем тромбоцитов (MPV), ширину распределения тромбоцитов (PDW), количество лейкоцитов (WBC), долю лимфоцитов (LY), моноцитов (MO), эозинофилов (EO). Лейкоцитарную формулу определяли путем микроскопического анализа мазков крови, окрашенных по Романовскому–Гимзе. С помощью биохимического анализатора «COBAS FARA» (Roche, Швейцария) в сыворотке крови измерили следующие маркеры: уровень холестерина (CHOL), триглицеридов (TRIG), общего белка (TP), альбумина (ALB), мочевины (UREA), креатинина (CREA), аспартатаминотрансферазы (AST), аланинаминотрансферазы (ALT), билирубина (BIL), альфа-амилазы (AMYL), лактатдегидрогеназы (LDH), щелочной фосфотазы (ALP) и креатинкиназы (CK). Спустя две недели гомогенат, полученный из естественно возникшей опухоли сингенной с реципиентами самки, инъецировали в жировую подушечку в область правых подмышечных лимфатических узлов в дозе 10^7 клеток/мышь. Ежедневно измеряли средний диаметр опухоли, как описано в [12]. На день 13 после перевивки опухоли самцов классифицировали на: группу 1 (быстро проявляющиеся опухоли, n=18, диаметр $>3,5$ мм на день 13) и группу 2 (медленно проявляющиеся опухоли, n=46, диаметр $<3,5$ мм на день 13). Далее проводили разведочный анализ данных для определения показателей, оптимальных для построения модели прогноза. В качестве математической модели индивидуального

прогноза скорости проявления опухоли использовали логистическую регрессию. Реализация модели выполнена на языке Python 3.8 с использованием модуля scikit-learn 0.22.

Результаты исследования и их обсуждение

В качестве критерия быстрого ($n=18$) или медленного ($n=46$) проявления РМЖ использовали среднее значение диаметра опухолевого узла в 3,5 мм на день 13 после перевивки. Опухоль диаметром менее 3,5 мм возможно диагностировать только путем пальпации тканей молочной железы (период до клинического проявления – субклинический период), в то время как новообразования, размер которых превышает это значение, можно диагностировать визуально. День 13 был выбран в качестве медианного значения окончания субклинического периода роста опухоли в популяции самок с естественно возникшим РМЖ [13]. Целесообразность такого подхода была подтверждена математически путем аппроксимации индивидуальной динамики роста опухоли для каждого реципиента. Общая закономерность роста перевитой опухоли в данной модели заключается в наличии трех этапов: 1) размер опухолевого узла начинает нарастать экспоненциально (субклинический период); 2) скорость роста снижается, и размер опухолевого узла начинает увеличиваться линейно; 3) скорость роста замедляется еще сильнее, а размер опухолевого узла остается практически неизменным. Такое поведение динамики роста с математической точки зрения можно описать сигмоидой – гладкой монотонной нелинейной функцией, имеющей форму буквы S, в частности логистической функцией (уравнение Ферхюльста, часто используемое в экологии для описания динамики роста некоторой популяции в условиях ограниченности ресурсов, рис. 1).

Рис. 1. А – логистическая функция, разделенная на три части, соответствующие трем фазам роста: а – начальная (экспоненциальная) фаза роста, б – фаза линейного роста, в – фаза плато; Б – приближение логистической функцией средних размеров образований (с указанием 95%-ного ДИ), $R^2 = 0,996$

Для определения момента перехода между фазами экспоненциального и линейного роста динамика роста опухолевого образования для каждой мыши аппроксимировалась логистическим уравнением (1). Подбор неизвестных параметров K, r, t_0, s_0 осуществлялся методом наименьших квадратов.

$$S(t) = \frac{K}{1 + e^{-r(t-t_0)}} + s_0 \quad 1)$$

Момент перехода между первым и вторым этапами характеризовался как пик увеличения скорости роста опухоли, после которого наступал обратный процесс – сначала темпы увеличения скорости роста начинали постепенно снижаться, а затем снижалась и сама скорость. Математически это соответствует максимуму второй производной логистической функции (рис. 2).

Рис. 2. А – приближение динамики роста отдельной мыши логистической функцией;
Б – график второй производной заданной функции; вертикальная черта соответствует
максимальному значению второй производной функции – моменту перехода между первым и
вторым этапами роста образования

Расчет дня перехода к линейному периоду роста опухоли показал, что в среднем это происходит на 16-й день после перевивки (медиана 15,3; 95%-ный ДИ: 10,9–20,4) при среднем диаметре опухолевого узла 3,55 мм. Полученные результаты подтверждают биологическое обоснование определения субклинического периода. День 13 был подтвержден в качестве оптимального критерия для деления популяции самцов на группу 1 (быстрое проявление опухоли: диаметр >3,5 мм на день 13) и группу 2 (медленное проявление опухоли: диаметр <3,5 мм на день 13) как наиболее близкий к теоретическому среднему дню окончания первой фазы роста.

Далее на этапе разведочного анализа данных было проведено исследование основных характеристик параметров и взаимосвязей между ними. Оценка нормальности

распределений параметров проводилась с использованием критерия Шапиро–Уилка. Большая часть параметров не имела нормального распределения из-за наличия выбросов, которые можно объяснить наличием у части реципиентов патологий внутренних органов. В таких случаях распределение параметра приводили к нормальному путем степенного преобразования.

Далее были исследованы корреляции между параметрами роста РМЖ и показателями крови и сыворотки, проведено сравнение средних значений (t -критерия Стьюдента, U -критерия Манна–Уитни) и дисперсий (критерия Фишера, критерия Ансари–Бредли) параметров в группах с быстрым и медленным проявлением РМЖ. Средние значения показателей по группам не отличались статистически значимо, хотя стоит отметить наличие некоторых тенденций к различиям по уровням ТР ($p=0,07$), НСТ ($p=0,09$), RBC ($p=0,10$), MCHC ($p=0,19$), TRIG ($p=0,19$). Анализ дисперсий показал различия между группами с быстрым и медленным проявлением РМЖ для параметров HGB ($p=0,03$), ALB ($p=0,02$), ALT ($p=0,03$). В заключительной части разведочного анализа данных были рассмотрены корреляции и совместные распределения показателей крови и сыворотки. Некоторые параметры образуют сильно коррелирующие группы, которые можно объяснить с биологической точки зрения (например, RBC, HGB и НСТ отражают основные характеристики эритроцитов, тогда как PLT, PCT, MPV – тромбоцитов). Такие группы в свою очередь могут иметь значимые обратные корреляции с параметрами других групп (например, группа RBC, HGB, НСТ имеет умеренную обратную корреляцию с группой MCH, MCHC). Такие взаимосвязи накладывают ограничение на прямое использование параметров в качестве предикторов в математических моделях, так как большое количество методов (например, линейный дискриминантный анализ) предполагает некоррелированность факторов. Исследование совместных распределений не показало явной возможности разделения групп с быстрым и медленным проявлением РМЖ линейной или нелинейной функцией. В результате было выдвинуто предположение о наличии возможности разделения групп в пространстве высокой размерности, то есть учитывающем многие анализируемые параметры.

Процесс построения предиктивной вероятностной модели классификации мышей на группы с быстрым и медленным проявлением РМЖ был организован следующим образом: 64 мыши были разделены на обучающую ($n=46$) и тестовую ($n=18$) выборки. Разделение проходило с учетом классовой стратификации (так как распределение по классам не является равномерным). Далее на обучающей выборке была построена предиктивная модель. Гиперпараметры модели определялись с использованием перекрестной проверки. Эффективность модели измерялась на тестовой выборке.

В качестве базовых параметров для построения предиктивной модели на основе результатов разведочного анализа были отобраны следующие: масса тела мыши на момент перевивки опухоли, HGB, MCV, RDW, MCHC, PCT, MPV, TP, ALB, TRIG. Структура модели была организована следующим образом. Сначала проводили степенное преобразование со стандартизацией для данных с распределением, отличным от нормального. В результате многие параметры после данного шага принимали вид стандартного нормального распределения. Далее в модель были добавлены латентные переменные, представляющие собой попарные произведения некоторых параметров. Снижение размерности после добавления латентных переменных проходило при помощи сингулярного разложения матрицы данных и отборе k максимальных сингулярных значений (оптимальное значение для k составило 20). В качестве модели классификации использовалась логистическая регрессия с L_1 - и L_2 -регуляризациями (эластичная сеть).

Результаты работы модели на тестовой выборке приведены в таблице. Чувствительность модели для группы с быстрым проявлением РМЖ составила 60%, специфичность – 84,6%, ROC AUC=0,754.

Результаты работы модели классификации на тестовой выборке

		Наблюдаемые значения		Всего
		Быстрое проявление	Медленное проявление	
Предсказанные значения	Быстрое проявление	3	2	5
	Медленное проявление	2	11	13
Всего		5	13	18

Однако данные результаты нельзя считать реферативными для оценки обобщающей способности модели, так как объем выборки достаточно мал. Оценка обобщающей способности была проведена методом Монте Карло – многократным случайным делением выборки на обучающую и тестовую, с последующим усреднением результатов. Среднее значение чувствительности составило 61% (95%-ный ДИ: 55,3–66,7%), специфичности – 69% (95%-ный ДИ: 63,9–74,1%).

Представленные результаты демонстрируют возможность построения математических моделей для прогноза динамики роста РМЖ с основой на данных общедоступных клинических тестов. Гематологические показатели HGB, MCV, MCHC и

RDW являются ключевыми характеристиками состояния эритроцитов и применяются в клинике для дифференциальной диагностики анемий [14]. Уровни PCT и MPV позволяют диагностировать нарушения системы свертывания крови; лейкоцитоз и сдвиг лейкоцитарной формулы могут свидетельствовать о наличии инфекционных процессов, заболеваниях печени и почек. Следовательно, информативными оказались гематологические показатели крови, отражающие два наиболее часто наблюдаемых осложнения РМЖ: анемию и нарушения системы гемостаза [6, 15]. Общий белок сыворотки крови (TP) считается показателем баланса синтеза и расщепления или потери белков; белок альбумин (ALB) на 80% определяет коллоидно-осмотическое давление плазмы, его доля составляет половину от всех плазматических белков. Недостаток этого белка может говорить о патологических процессах в печени или сердечно-сосудистой системе. У пациентов с РМЖ часто наблюдается повышение уровня TP при снижении уровня ALB в сыворотке крови [7, 9]. Триглицериды (TRIG) являются информативным показателем липидного обмена. TRIG образуются в печени из глицерина и высших жирных кислот и попадают в кровоток в виде липопротеидов очень низкой плотности (ЛПОНП). Повышенный уровень триглицеридов часто приводит к заболеваниям сердечно-сосудистой системы и ассоциирован с высоким риском развития ишемической болезни сердца. Следовательно, биохимические показатели сыворотки крови, информативные для диагностики заболеваний печени и сердечно-сосудистой системы, оказались прогностическими маркерами для предсказания скорости проявления перевитой опухоли в мышинной модели РМЖ человека.

Таким образом, можно сделать вывод о том, что индивидуальная предрасположенность к более раннему или позднему проявлению перевитой опухоли может быть выявлена по ряду гематологических и биохимических показателей, измеренных заблаговременно у интактных реципиентов и отражающих основные патофизиологические изменения внутренних органов и систем.

Список литературы

1. Friberg S., Mattson S. On the growth rates of human malignant tumors: Implications for medical decision making. *Journal of Surgical Oncology*. 1997. vol. 65. no. 4. P. 284–297.
2. Coates A.S., Colleoni M., Goldhirsch A. Is adjuvant chemotherapy useful for women with luminal a breast cancer? *J. Clin. Oncol.* 2012. vol. 30. no. 12. P. 1260–1263.
3. Sogaard M., Thomsen R.W., Bossen K.S., Sorensen H.T., Norgaard M. The impact of comorbidity on cancer survival: a review. *Clin. Epidemiol.* 2013. vol. 5(Suppl 1). P. 3–29.

4. Simon K. Studies of diagnosis and prognosis of breast carcinoma from serum findings. *Medizinische Monatsschrift*. 1957. vol. 11. no. 12. P. 794-800.
5. Lee Y.T. Biochemical and hematological tests in patients with breast carcinoma: correlations with extent of disease, sites of relapse, and prognosis. *Journal of surgical oncology*. 1985. vol. 29. no. 4. P. 242-248.
6. Olufemi A.E., Saliu O.A., Layiwola A.M., Olayide A.S. Haematological Parameters in Female Breast Cancer Patients in South Western Nigeria. *Int. J. Med. Health Sci*. 2013. vol. 2. no. 4. P. 398-401.
7. Al-Muhtaseb S.I. Serum and saliva protein levels in females with breast cancer. *Oncology Letters*. 2014. vol. 8. no. 6. P. 2752-2756.
8. Rana A.P.S., Kaur M., Zonunsanga B., Puri A., Kuka A.S. Preoperative Peripheral Blood Count in Breast Carcinoma: Predictor of Prognosis or a Routine Test. *Int. J. Breast Cancer*. 2015. vol. 2015. P. 964-392.
9. Paz M., Júnior Gomes A.L., Islam M.T., Tabrez S., Jabir N.R., Alam M.Z., Machado K.C., de Alencar M., Machado K.C., Ali E.S., Mishra S.K., Gomes L.F., Sobral A., e Sousa J., de Souza G.F., Melo-Cavalcante A., da Silva J. Assessment of chemotherapy on various biochemical markers in breast cancer patients. *Journal of cellular biochemistry*. 2017. vol. 119. no. 3. P. 2923–2928.
10. Shrivastava S., Singh N., Nigam A.K., Chandel S.S., Shrivastava R., Kumar S. Comparative study of hematological parameters along with effect of chemotherapy and radiotherapy in different stages of breast cancer. *Int. J. Res. Med. Sci*. 2016. vol. 5, no. 1. P. 311-315.
11. Chauhan P., Yadav R., Kaushal V., Beniwal P. Evaluation of serum biochemical profile of breast cancer patients. *Health Sciences*. 2016. vol. 5. no. 7. P. 1-7.
12. Semushina S.G., Aronov D.A., Moiseeva E.V. Local Interleukin-2 Immunotherapy of Breast Cancer: Benefit and Risk in a Spontaneous Mouse Model. *Pathol. Oncol. Res*. 2019. vol. 25. no. 3. P. 945• 951.
13. Moiseeva E.V. Anti-breast Cancer Drug Testing. Original Approaches. Novel Set of Mouse Models. Saarbrücken, Germany: Lambert academic publishing, 2009. p. 210.
14. Jameel T., Baig M., Ahmed I., Hussain M.B., Alkhamaly M.B.D. Differentiation of beta thalassemia trait from iron deficiency anemia by hematological indices. *Pak. J. Med. Sci*. 2017. vol. 33. no. 3. P. 665• 669.
15. Okuturlar Y., Gunaldi M., Tiken E.E., Oztosun B., Inan Y.O., Ercan T., Tuna S., Kaya A.O., Harmankaya O., Kumbasar A. Utility of peripheral blood parameters in predicting breast cancer risk. *Asian Pac. J. Cancer Prev*. 2015. vol. 16. no. 6. P. 2409-2412.