

МУТАГЕНЕЗ И КАНЦЕРОГЕНЕЗ

С.К. Абилев, А.В. Тарасов,
В.А. Тарасов

Институт общей генетики
им. Н.И. Вавилова РАН,
Москва

✿ Предложен метод оценки прогностической эффективности батареи краткосрочных тестов при оценке потенциальной канцерогенной активности химических соединений для грызунов. Определены уровни доказанности канцерогенной активности 83-х химических соединений, достигаемые при их тестировании с помощью батареи из трех тестов — теста Эймса, теста по учету СХО в культуре клеток и теста по учету микроядер в костном мозге у грызунов. Исследуемая выборка использована в качестве обучающей для прогнозирования потенциальной канцерогенной активности новых химических соединений методом кросс-проверки.

✿ **Ключевые слова:** канцерогенная активность; прогностическая эффективность тест-систем; кросс-проверка; уровень доказанности.

ПРОГНОСТИЧЕСКАЯ ЭФФЕКТИВНОСТЬ КРАТКОСРОЧНЫХ ТЕСТ-СИСТЕМ ПРИ ОЦЕНКЕ КАНЦЕРОГЕННОЙ АКТИВНОСТИ ХИМИЧЕСКИХ СОЕДИНЕНИЙ

ВВЕДЕНИЕ

Применение краткосрочных тест-систем на генотоксичность для выявления потенциальных канцерогенов поставило проблему оценки их прогностической эффективности. В начале 70-х годов прошлого столетия были введены понятия чувствительности, специфичности и прогностической эффективности для оценки результатов теста Эймса [1]. Позже эти понятия были распространены на все тест-системы в качестве универсальных критериев оценки результатов тестирования. Чувствительность теста представляет собой долю правильно определенных канцерогенов, а специфичность — долю правильно определенных неканцерогенов, которые обычно вычисляются при ретроспективном анализе результатов применения той или иной тест-системы.

Показатели чувствительности и специфичности тестов использовались при разработке подходов для формирования батарей тестов. Одним из таких подходов является метод *CPBS (Carcinogen prediction and battery selection)*, основанный на применении теоремы гипотез Байеса для предсказания канцерогенной активности не тестированных соединений [2, 3]. Как метод *CPBS*, так и другие методы формирования батарей основаны на использовании показателей чувствительности и специфичности тестов, то есть вероятностных характеристик тест-систем [4, 5, 6, 7]. Однако вероятностные характеристики не позволяют определить какую информацию о канцерогенных свойствах мы получаем в ходе тестирования химического соединения при использовании батареи тестов. Ранее нами был разработан новый подход к оценке прогностической эффективности тест-систем и их батарей, основанный на расчете информации получаемой в ходе тестирования, то есть на замене вероятностных характеристик на информационные [8–10]. С использованием этого метода нами была проведена оценка эффективности тестирования и достигаемого при этом уровня доказанности мутагенной активности для половых клеток млекопитающих при использовании различных тест-систем и их сочетаний [11, 12].

Данная работа посвящена определению прогностической эффективности трех широко используемых тест-систем на генотоксичность и уровня (веса) доказанности канцерогенной активности для химических соединений, достигаемых при использовании этих тестов.

МАТЕРИАЛЫ И МЕТОДЫ

Характеристика выборки химических соединений. Источником информации при создании выборки химических соединений, канцерогенная активность которых была исследована на грызунах, служи-

ла компьютерная база данных CPDB-Carcinogenic Potential Date Base (<http://potency.berkeley.edu/cpdb.html>). Для ретроспективного анализа были выбраны химические соединения, исследованные на мышах, и были сформированы две группы: канцерогены и неканцерогены. Затем из базы данных GENE-TOX (<http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?GENETOX>) были выбраны данные по тестированию этих соединений в 5 тест-системах: тест Эймса (ТЭ), тесты по учету хромосомных aberrаций (ХА), генных мутаций (ГМ) и сестринских хроматидных обменов (СХО) в культурах клеток *in vitro* и микроядер в клетках костного мозга мышей *in vivo* (МЯ). В результате была сформирована вы-

борка, состоящая из 772-х химических соединений (в том числе 349 канцерогенов и 423 неканцерогенов). При формировании выборки для ретроспективного анализа руководствовались двумя критериями. Во-первых, химические соединения должны быть испытаны в 3-х из перечисленных тест-систем. Во-вторых, для каждой тест-системы должны содержаться результаты тестирования более 50% соединений. Всего было отобрано 83 соединения, из которых 8 были испытаны в 5 тест-системах, 6 — в 4-х тест-системах. Из 83-х соединений 81 протестировано в ТЭ, 73 — на СХО и 74 — на МЯ. Из дальнейшего рассмотрения выпали тесты на культурах клеток на индукцию ХА и ГМ, на которых были протестированы только 25 и

Таблица 1

Список канцерогенов и неканцерогенов, включенных в выборку для ретроспективного анализа

Химические соединения	Номер по CAS	Тест-системы			Канцерогенная активность	
		ТЭ	СХО	МЯ	На мышах	На крысах
Канцерогены						
Озон	10028-15-6	1	1	0,54*	1	0
Гидразин сульфат	10034-93-2	1	1	0	1	1
Бензил хлорид	100-44-7	1	1	0	1	0
1,2-Дибромэтан	106-93-4	1	1	0	1	1
Катехол	120-80-9	0	1	0	1	1
Гидрохинон	123-31-9	0	1	1	1	1
Трис(2,3-дибромпропил)фосфат	126-72-7	1	1	1	1	1
Каптан	133-06-2	1	1	1	1	1
Мелфалан	148-82-3	1	1	0,54	1	1
Этиленимин	151-56-4	1	0,63	0,54	1	0
Аурамин-О	2465-27-2	1	1	0	1	1
Краситель синий НС №1	2784-94-3	1	1	0	1	1
Хлорамбуцил	305-03-3	1	1	1	1	1
АФ-2 (консервант)	3688-53-7	1	1	0	1	1
Метранидазол	443-48-1	1	1	0	1	1
Формальдегид	50-00-0	1	1	0	1	1
Циклофосфамид	50-18-0	1	1	1	1	1
ДДТ	50-29-3	0	0,63	0	1	1
Бенза(а)пирен	50-32-8	1	1	1	1	1
Триметилфосфат	512-56-1	1	0,63	1	1	0
Уретан	51-79-6	1	1	1	1	1
ТиоТЕФ	52-24-4	1	1	1	1	1
2-Нитро-р-фенилендиамин	5307-14-2	1	1	0	1	0
N-гидрокси-2-ацетиламинофлуорен	53-95-2	1	1	0,54	1	1
2-Ацетиламинофлуорен	53-96-3	1	1	1	1	1
Диэтилстильбэстрол	56-53-1	0	1	1	1	1
1,1-Диметилгидразин	57-14-7	1		0	1	0
β-Пропиолактон	57-57-8	1	1	0	1	1
7,12-Диметилбенза(а)антрацен	57-97-6	1	1	1	1	1
5-Нитро-2-фуральдегид семикарбазон	59-87-0	1	0,63	0	1	1
3-Аминотриазол	61-82-5	0	1	0	1	1
Дихлорфос	62-73-7	1	1	0	1	1
N-нитрозодиметиламин	62-75-9	1	1	1	1	1

Продолжение таблицы 1

Метилметансульфонат	66-27-3	1	1	1	1	0
Хлороформ	67-66-3	0	1	0	1	1
Бензол	71-43-2	0	1	1	1	1
p,p-DDE	72-55-9	0	0,63	0,54	1	0
Винил хлорид	75-01-4	1	1	1	1	1
Метилен хлорид	75-09-2	1	1	0	1	1
Этиленоксид	75-21-8	1	1	1	1	1
IQ	76180-96-6	1	1	0	1	1
Трихлорэтилен	79-01-6	0	0	1	1	1
Диметилкарбамил хлорид	79-44-7	1	1	1	1	0
2-Нафтиламин	91-59-8	1	1	1	1	1
Нитрозодибутиламин	924-16-3	1	1	0	1	1
Сафрол	94-59-7	0	1	0	1	1
Стирен оксид	96-09-3	1	1	0	1	1
Этилентиомочевина	96-45-7	1	0	0	1	1
Стирол	100-42-5	1	1	1	0	1
Эпихлоргидрин	106-89-8	1	1	0	0	1
Толуол	108-88-3	0	0	1	0	1
Афлатоксин В1	1162-65-8	1	1	1	0	1
Кверцетин	117-39-5	1	1	1	0	1
Сахарин, натриевая соль	128-44-9	0	1	0	0	1
Ацетат свинца	301-04-2	0	0,63	1	0	1
Карбарил	63-25-2	1	1	0	0	1
Этиловый спирт	64-17-5	0	1	0	0	1
Нитрит натрия	7632-00-0	1	1	1	0	1
Неканцерогены						
Оксид азота	10024-97-2	0,76	0	0,54	0	0
Резорцинол	108-46-3	0	0	0	0	0
2,4-Динитротолуол	121-14-2	1	0,63	0	0	0
Малатион	121-75-5	0	1	1	0	0
Малениновый гидразид	123-33-1	0	1	0	0	0
Смесь ксилолов (m- и o-ксилолы)	1330-20-7	0	0	0	0	0
Ерфлуран	13838-16-9	0	0	0,54	0	0
Метилпаратрион	298-00-0	1	1	1	0	0
Аспирин	50-78-2	0	0,63	0	0	0
Пронеталол гидрохлорид	51-02-5	0	0,63	0	0	0
Сахароза	57-50-1	0	0	0	0	0
Кофенин	58-08-2	0	1	1	0	0
Меготрексат	59-05-2	0	1	1	0	0
Диметоат	60-51-5	1	1	1	0	0
1,1,1-Трихлорэтан технический	71-55-6	1	1	0	0	0
Хлорид натрия	7647-14-5	0	0	0	0	0
Фтор, натриевая соль	7681-49-4	0	1	1	0	0
Гипохлорид натрия	7681-52-9	0	1	0,54	0	0
Мышьяк, натриевая соль	7784-46-5	0,76	1	1	0	0
Сахарин	81-07-2	0	0	0,54	0	0
Хлорпропамид	94-20-2	0	1	1	0	0
2,4-Дихлорфеноксиуксусная кислота	94-75-7	0	1	0	0	0
γ-Бутиролактон	96-48-0	0	1	0	0	0
Эугенол	97-53-0	0	0	1	0	0
4-Нитро-о-фенилендиамин	99-56-9	1	1	0	0	0

Примечание: 1 — наличие активности; 0 — отсутствие активности; * — рассчитанные значения вероятности получения позитивного результата при тестировании.

38 веществ соответственно. Кроме того, в базу данных для анализируемой выборки добавили результаты их испытаний на канцерогенную активность на крысах, и выборку разделили на канцерогены и неканцерогены. При этом к неканцерогенам отнесли соединения, которые не проявили активности на обоих видах грызунов (табл. 1).

При проведении расчетов значений информации и уровня доказанности использовались специально созданные компьютерные программы с помощью VBA для программ Microsoft Excel и Microsoft Access.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Математические модели. Для перехода от вероятностного описания процесса биотестирования к информационным характеристикам процесса тестирования нами были использованы основные понятия теории информации [13]. В рамках развиваемых представлений полная информация о канцерогенной активности химических соединений зависит от априорной информации об их канцерогенности и от информации, полученной в ходе тестирования. В том случае, когда источник получения априорной информации не связан с биотестированием, полная информация равняется:

$$I_{\pi} = I_0 + I, \quad (1)$$

где I_{π} — полная информация; I_0 — априорная информация; I — полученная информация.

Информация, или точнее селективная информация, как ее определили Винер [14] и Шеннон [15], является мерой изменения неопределенности, в данном случае разности между априорной неопределенностью и апостериорной неопределенностью:

$$I(A_i) = H_0 - H(A_i), \quad (2)$$

где $I(A_i)$ — полученная в ходе тестирования информация при A_i -ом результате; H_0 — априорная неопределенность; $H(A_i)$ — апостериорная неопределенность при A_i -ом результате.

Полученная информация в зависимости от результата может приобретаться и теряться (т.е. быть как положительной, так и отрицательной). Средняя информация, в отличие от полученной для каждого варианта испытаний, никогда не бывает отрицательной:

$$I_{cp} = \sum p(A_i)I(A_i), \quad (3)$$

где $p(A_i)$ — вероятность получения A_i -ого результата.

Величины априорной H_0 и апостериорной $H(A_i)$ неопределенностей определяются значениями соот-

ветственно априорной вероятности и апостериорной вероятности A_i -ого результата испытаний:

$$H_0 = p_0 \log_2 1/p_0 + (1 - p_0) \log_2 1/(1 - p_0), \quad (4)$$

$$H(A_i) = p(k/A_i) \log_2 1/p(k/A_i) + [1 - p(k/A_i)] \times \log_2 1/[1 - p(k/A_i)], \quad (5)$$

где p_0 — априорная вероятность наличия у вещества канцерогенных свойств; $p(k/A_i)$ — апостериорная вероятность наличия у вещества канцерогенных свойств в том случае, если в ходе испытаний получен A_i -й результат; k — канцерогенная активность.

Для оценки уровня доказанности канцерогенной активности у испытанных соединений нами введена специальная функция, названная функцией веса (уровня) доказанности наличия канцерогенной активности (*weight-of-evidence*):

$$W_{\kappa}(A_i) = \frac{1 + I(A_i) \text{Sign}(p(A_i) - 0,5)}{2}, \quad (6)$$

где $W_{\kappa}(A_i)$ — уровень доказанности канцерогенности соединения (κ), достигаемый при A_i -ом результате тестирования; $I(A_i)$ — полная апостериорная информация, полученная при A_i -ом результате в тесте; $p(A_i)$ — апостериорная вероятность; Sign — функция равная +1, если $p(A_i) > 0,5$, равная -1, если $p(A_i) < 0,5$, равная 0, если $p(A_i) = 0,5$.

Значения этой функции веса доказанности меняются от 0 до 1, или от 0 до 100%.

Для оценки значений апостериорной вероятности нами использован развитый Фишером дискриминантный анализ. При оценке апостериорной вероятности наличие у химических соединений канцерогенной активности выступает в качестве зависимой переменной, в то время как результаты тестирования — в качестве независимых переменных. Однако при проведении дискриминантного анализа необходимо формализовать отсутствующие значения результатов тестирования. Это можно сделать, если результаты испытаний представить в форме вероятности получения позитивного результата. Тогда при позитивном и негативном исходе испытаний эта вероятность оказывается равной 0 или 1 соответственно, а при отсутствии испытаний эта вероятность может быть представлена ее правдоподобной оценкой [11, 12].

При проведении анализа конкретной выборки возможна прямая оценка доли химических соединений, показавших позитивный результат для данной тест-системы. В общем случае для выборок, характеризующихся различным соотношением между канце-

Таблица 2

Чувствительность и специфичность тест-систем, рассчитанные по данным тестирования химических соединений

Тест-системы	Протестировано всего	Канцерогенов — 58				Неканцерогенов — 25			
		из них:	результаты тестирования		чувствительность теста	из них:	результаты тестирования		специфичность теста
			+	—			+	—	
Тест Эймса	81	58	44	14	0,76	23	5	18	0,78
Тест на СХО <i>in vitro</i>	73	51	48	3	0,94	22	14	8	0,36
Микроядерный тест <i>in vivo</i>	74	53	26	27	0,49	21	9	12	0,57

рогенами и неканцерогенами, вероятность получения позитивного результата оказывается равной:

$$p(+)=p_0\alpha+(1-p_0)\beta, \quad (7)$$

где α и β — значения чувствительности и специфичности данного теста.

Первый член этого равенства представляет собой долю правильных позитивных результатов, второй член — долю ложных позитивных результатов. Используя значения чувствительности и специфичности, представленные в табл. 2, были рассчитаны значения $p(+)$ при $p_0 = 0,5$, и полученные таким образом значения $p(+)$ использовались в качестве правдоподобных оценок отсутствующих данных. В табл. 1 представлены трансформированные таким образом данные для $p_0 = 0,5$, которые служили основой для проведения дискриминантного анализа.

ОПРЕДЕЛЕНИЕ ПРОГНОСТИЧЕСКОЙ ЭФФЕКТИВНОСТИ ТЕСТ-СИСТЕМ

В табл. 2 приведены данные тестирования 83 отобранных соединений в 3-х тест-системах и результаты расчета чувствительности и специфичности этих тест-систем. Видно, что тест на МЯ обладает низкой чувствительностью (0,49) и столь же низкой специфичностью. Тест на СХО имеет очень высокий пока-

затель чувствительности (0,94) и очень низкий показатель специфичности (0,36). Чувствительность теста Эймса составила 0,76 и была ниже этого показателя для теста на СХО, в то же время специфичность его была самой высокой среди всех рассматриваемых тест-систем.

С помощью дискриминантного анализа нами была рассчитана апостериорная вероятность наличия канцерогенной активности для каждого соединения, а затем, используя служебную компьютерную программу «Union 1001», были рассчитаны значения получаемой информации в результате тестирования для каждого соединения как в отдельной тест-системе, так и в батарее тест-систем, а также значения достигаемого при этом уровня доказанности канцерогенной активности. В табл. 3 приведены значения средней информации, получаемой при тестировании 83-х рассматриваемых нами соединений как в отдельных тест-системах, так и в батареях тест-систем.

Как видно из табл. 3, результаты тестирования в ТЭ на данной выборке имеют наибольшую прогностическую эффективность ($I_{cp} = 0,26$), и результаты тестирования в батареях тестов ТЭ+СХО ($I_{cp} = 0,29$) и ТЭ+МЯ ($I_{cp} = 0,26$) не приводят к значительному повышению этой эффективности. Батарея из 3-х тестов ТЭ+СХО+МЯ по эффективности ($I_{cp} = 0,29$) не отличалась от батареи из двух тестов ТЭ+СХО ($I_{cp} = 0,29$). Особое внимание обращают результаты

Таблица 3

Эффективность тест-систем и их батарей в прогнозировании канцерогенной активности химических соединений для грызунов

Тест-системы и их батареи	Средняя информация	% правильно классифицированных*	
		Канцерогенов	Неканцерогенов
Тест Эймса (ТЭ)	0,26	75,86	72,00
Тест на СХО <i>in vitro</i> (СХО)	0,11	75,86	72,00
Микроядерный тест <i>in vivo</i> (МЯ)	0	53,45	48,00
ТЭ + СХО	0,29	74,14	76,00
ТЭ + МЯ	0,26	75,86	72,00
СХО + МЯ	0,11	82,76	44,00
ТЭ + СХО + МЯ	0,29	74,14	76,00

Примечание. * — классификация соединений на основе значений апостериорной вероятности.

теста на МЯ — отсутствие полной информативности в силу низкой чувствительности и столь же низкой специфичности (табл. 1 и 2). Низкая эффективность теста на СХО, по-видимому, связана с его очень низкой специфичностью.

Дискриминантный анализ позволяет определить значения апостериорной вероятности принадлежности химических соединений к классу канцерогенов или к классу неканцерогенов. При этом принято считать, что соединения, для которых апостериорная вероятность $p > 0,5$ относятся к классу канцерогенов и, напротив, при апостериорной вероятности $p < 0,5$ соединения относятся к классу неканцерогенов. Используя классификационные функции или значения апостериорной вероятности, можно получить классификационную матрицу. В табл. 3 приведены результаты классификации 83-х химических соединений (канцерогенов и неканцерогенов) на основе значений апостериорной вероятности, которые выражены в долях правильно определенных (классифицированных) канцерогенов и неканцерогенов, которые в сущности аналогичны чувствительности и специфичности тест-систем и их батарей. Доля правильно классифицированных соединений по всей выборке представляет собой конкордантность.

Уровень доказанности канцерогенности, достигаемый при тестировании. Для каждого из 83-х химических соединений, входящих в выборку, были получены значения апостериорной вероятности и на

этой основе значения полученной информации (I_p) и уровня доказанности канцерогенности (w_k). При отсутствии экспериментальных данных пустые ячейки были заполнены значениями вероятности получения позитивного результата, рассчитанной по формуле 7. Уровень доказанности рассчитывали по формуле 6.

На рис. 1а приведено распределение канцерогенов и неканцерогенов по уровням доказанности канцерогенности по результатам тестирования в батарее тестов ТЭ+СХО+МЯ. Видно, что основная масса канцерогенов располагается в правой части графика, то есть в области значений уровня доказанности $> 0,5$, тогда как неканцерогены — в левой части графика в области значений $< 0,5$. Вместе с тем, часть канцерогенов оказывается в области низких значений уровня доказанности, а часть неканцерогенов в области высоких значений, то есть наблюдаются «ложно-негативные» и «ложно-позитивные» результаты. Кроме того, часть веществ расположилась в области 0,5, что указывает на недостаточность доказанности наличия или отсутствия канцерогенности. По-видимому, это связано с низкой специфичностью теста на СХО и низкой чувствительностью теста на МЯ. В этой связи было проведено сравнение распределения канцерогенов и неканцерогенов по уровням доказанности, полученных в результате тестирования в батарее ТЭ+СХО+МЯ, с распределением по уровням доказанности по итогам тестирования только в ТЭ. На

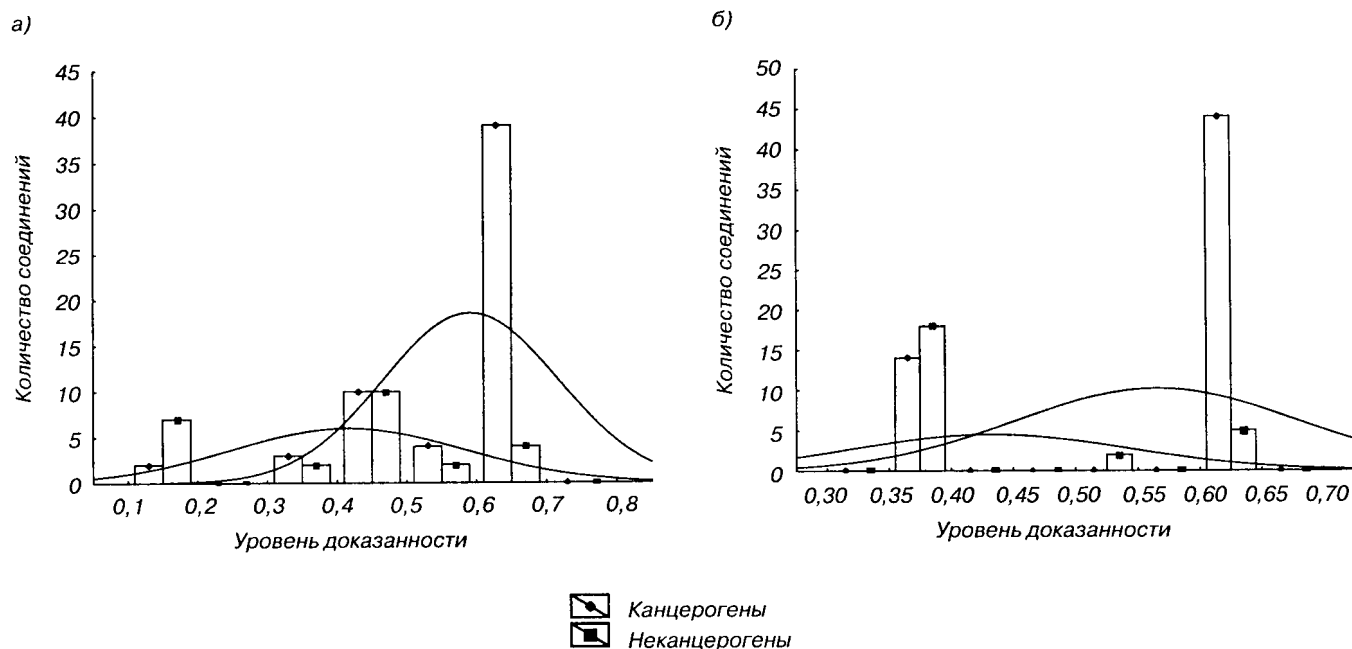


Рис. 1. Распределение соединений по уровням доказанности канцерогенности по результатам тестирования в батарее ТЭ+СХО+МЯ (А) и в тесте Эймса (Б)

рис. 1б приведено распределение канцерогенов и неканцерогенов по уровням доказанности по результатам испытания в тесте Эймса. Видно, что вещества более полярно разделяются на канцерогены и на неканцерогены, и очень мало веществ в области неопределенности, то есть в области 0,5. Кроме того, снижено количество неканцерогенов, определенных как канцерогены. По-видимому, это связано с более высокой специфичностью теста Эймса по сравнению с другими тестами в рамках данной выборки. Ясно, что чувствительность и специфичность зависят от долей правильно определенных канцерогенов и неканцерогенов на основании наличия или отсутствия у тестируемых соединений генетической активности в тест-системах или их батарей. Известно, что канцерогенная активность, определяемая на грызунах, видоспецифична, то есть соединения, проявляющие канцерогенную активность на мышах могут не проявлять ее на крысах, и наоборот (табл. 1).

Поэтому как прогностические показатели тест-систем, так и уровни доказанности, будут зависеть от состава выборки.

Каждое из соединений в анализируемой выборке характеризуется тем или другим исходом испытаний в тест-системах. Очевидно, что соединения, показавшие одинаковые результаты при проведении этих испытаний, характеризуются одними и теми же значениями полученной информации и уровня доказанности. В конечном счете, нам необходимо провести оценку прогностической значимости различных вариантов исхода испытаний. В табл. 4 сгруппированы данные по уровням доказанности, достигаемые при различных вариантах исхода тестирования, с которыми мы столкнулись при анализе использующейся в работе выборки химических соединений.

Как видно из табл. 4, почти все канцерогены, проявляющие генетическую активность в батарее тестов ТЭ+СХО+МЯ (во всех 3-х тестах или в ТЭ + СХО)

Таблица 4

Распределение химических соединений по уровням доказанности канцерогенной активности в зависимости от результатов тестирования

Химические соединения	Уровень доказанности, w	Результаты тестирования (ТЭ/СХО/М)
Трис(2,3-дибромпропил)фосфат, каптан, хлорамбуцил, циклофосфамид, бенза(а)пирен, уретан, тиоТЕФ, 2-ацетиламинофлуорен, 7,12-диметилбенза(а)антрацен, афлатоксин В1, N-нитрозодиметил-амин, стирол, метилметансульфонат, винилхлорид, этиленоксид, диметилкарбамил хлорид, 2-нафтиламин, кварцетин, нитрит натрия, <i>метилпаратион, диметоат*</i>	0,67	+/+/+
Озон, мелфалан, N-гидрокси-2-ацетиламинофлуорен	0,66	+/+/нт
Гидразин сульфат, бензилхлорид, 1,2-дибромэтан, аурамин-О, краситель синий НС №1, IQ, карбарил, АФ-2 (консервант), метранидазол, формальдегид, 2-нитро-р-фенилендиамин, β -пропиолактон, дихлорфос, метилхлорид, нитрозодибутиламин, стирен оксид, эпихлоргидрин, <i>1,1,1-трихлорэтан технический, 4-нитро-о-фенилендиамин</i>	0,66	+/+/-
Мышьяк натриевая соль	0,58	нт/+/+
Этиленимин	0,57	+/нт/+
Триметилфосфат	0,57	+/нт/нт
1,1-Диметилгидразин, 5-нитро-2-фуральдегид семикарбазон, 2,4-динитротолуол	0,56	+/нт/-
Этилентиомочевина	0,50	+/-/-
Оксид азота	0,45	нт/-/нт
Малатион, кофеин, метотрексат, хлорпропамид, фтор (натриевая соль), <i>гидрохинон, бензол, диэтил-стильбэстрол</i>	0,45	-/+/+
Гипохлорит натрия	0,45	-/+/нт
<i>Катехол, 3-аминотриазол, хлороформ, сафрол, сахарин (натриевая соль), этиловый спирт, малеиновый гидразид, 2,4-дихлорфеноксиуксусная кислота, γ-бутиролактон</i>	0,45	-/+/-
<i>ДДТ</i>	0,36	-/нт/+
<i>р,р-ДДЕ</i>	0,35	-/нт/нт
Аспирин, пронеталол гидрохлорид, <i>ацетат свинца</i>	0,35	-/нт/-
<i>Трихлорэтилен, толуол</i>	0,19	-/-/+
Смесь ксилолов (м- и о-ксилолы), ерфлуран	0,19	-/-/нт
Резорцинол	0,19	-/-/+
Сахарин, эугенол, сахароза, хлорид натрия	0,19	-/-/-

Примечание: * — курсивом выделены соединения, определенные неправильно (ложно-положительные результаты при $w > 0,5$ и ложно-отрицательные результаты при $w < 0,5$)

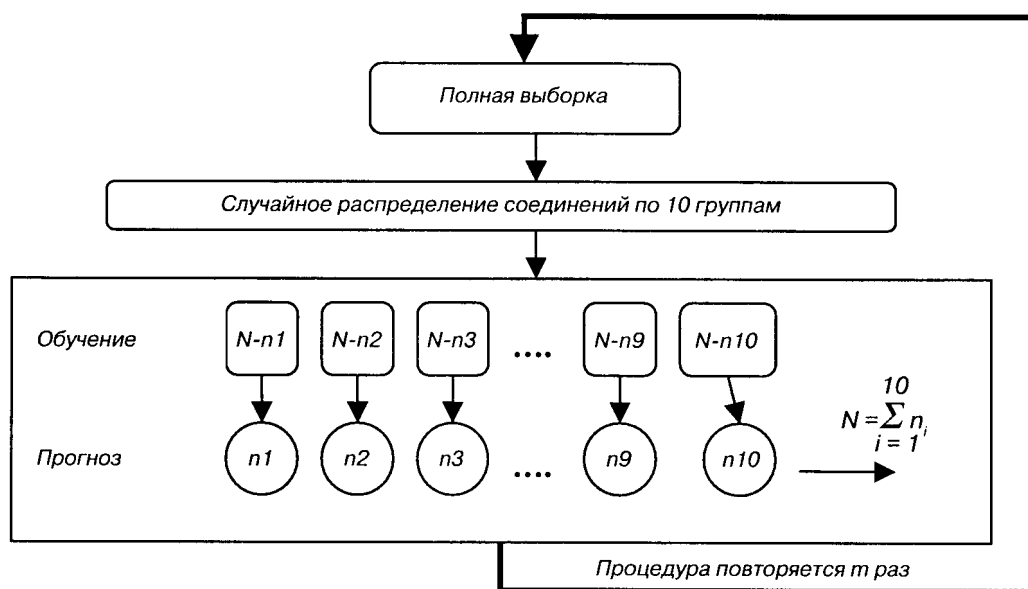


Рис. 2. Схема перекрестных проверок

находятся в группе с наивысшими уровнями доказанности $w = 0,66$ и $0,67$. В этой же группе находятся 4 соединения, не показавшие канцерогенной активности как на мышах, так и на крысах, (метилпаратин, диметоат, 1,1,1-трихлорэтан технический и 4-нитро-фенилендиамин). Ряд канцерогенов, которые не обладают способностью индуцировать мутации в тесте Эймса, оказались в группе соединений с низким уровнем доказанности ($w < 0,5$). К ним относятся гормональный препарат диэтилстильбэстрол, пестициды ДДТ и р,р-ДДЕ, этиловый спирт, толуол, бензол и другие.

Анализ связи между уровнями доказанности и результатами тестирования показал, что наибольшую прогностическую значимость в отношении канцерогенной активности на грызунах имеют результаты, полученные в тесте Эймса. В этом отношении вторыми по значимости являются результаты теста на индукцию СХО. Результаты теста на микродрова не имеют существенного значения. Эти данные противоречат полученным ранее результатам о низкой прогностической эффективности теста Эймса в отношении мутагенной активности химических соединений для половых клеток млекопитающих [11, 12].

Применение метода кросс-проверок для прогнозирования. Выборку, представленную в табл. 1, можно использовать в качестве обучающей выборки и применить различные статистические методы, которые позволят оценить с какой эффективностью будет осуществляться прогноз наличия либо отсутствия канцерогенной активности у химических соединений, не вошедших в обучающую выборку. Для этого разработан и используется метод кросс-про-

верок (*cross-validation*), когда общая выборка соединений разбивается на обучающую и контрольную.

При проведении кросс-проверок оценивается вероятность канцерогенной активности у химических соединений контрольной выборки, то есть соединений, не входящих в обучающую выборку. Значение этой вероятности зависит от состава обучающей выборки. В результате возникает проблема разбиения общей выборки на обучающую и контрольную. Для того, чтобы устранить субъективизм при этом разбиении выборки на обучающую и контрольную и оценить величины погрешности оценки при разных способах разбиения, нами предложен метод, схема которого представлена на рис. 2., реализованный в программу,

На первом этапе выборка соединений случайно разбивается на десять групп таким образом, чтобы каждое соединение было представлено лишь в одной группе и один раз. Эти десять групп соединений представляют собой контрольные выборки. Каждой контрольной выборке соответствует своя обучающая выборка, представляющая собой соединения общей выборки, за исключением соответствующей контрольной. Обучающая выборка используется для количественного определения параметров, позволяющих дискриминировать соединения на группы канцерогенов и неканцерогенов. Подробности применения дискриминантного анализа для этих целей описаны нами ранее [10, 11, 12]. Для каждого соединения контрольной выборки проводилась оценка вероятности наличия канцерогенной активности и рассчитывались значения полученной информации и уровня доказанности. Эта процедура повторяется для каждой контрольной выборки. В результате мы име-

Таблица 5

Результаты прогнозирования канцерогенной активности методом кросс-проверок

Химические соединения	Уровень доказанности, w	Химические соединения	Уровень доказанности, w
Метилпаратион	0,73 ± 0,03	Нитрозодибутиламин	0,64 ± 0,03
Диметоат	0,72 ± 0,04	2,4-Динитротолуол	0,60 ± 0,01
4-Нитро-о-фенилендиамин	0,72 ± 0,03	Мышьяк натриевая соль	0,60 ± 0,02
1,1,1-Трихлорэтан технический	0,69 ± 0,04	5-Нитро-2-фуральдегид семикарбазон	0,57 ± 0,02
N-гидрокси-2-ацетиламинофлуорен	0,68 ± 0,03	Этиленимин	0,56 ± 0,02
Трис(2,3-дибромпропил) фосфат	0,68 ± 0,02	Триметилфосфат	0,56 ± 0,02
Дихлорфос	0,68 ± 0,05	1,1-Диметилгидразин	0,56 ± 0,02
Этилен оксид	0,68 ± 0,05	Меготрексат	0,47 ± 0,02
β-Пропиолактон	0,67 ± 0,02	2,4-Дихлорфеноксиуксусная кислота	0,47 ± 0,01
Озон	0,67 ± 0,04	Малатион	0,47 ± 0,02
Бенза(а)пирен	0,67 ± 0,03	Хлорпропамид	0,47 ± 0,01
Метилметансульфонат	0,67 ± 0,03	Фтор (натриевая соль)	0,47 ± 0,01
Краситель синий НС №1	0,67 ± 0,03	Этиленглимочевина	0,47 ± 0,04
Диметилкарбамил хлорид	0,67 ± 0,04	Малениновый гидразид	0,47 ± 0,02
2-Нафтиламин	0,67 ± 0,03	γ-Бутиролактон	0,46 ± 0,01
Мелфалан	0,66 ± 0,02	Оксид азота	0,46 ± 0,03
Аурамин-О	0,66 ± 0,05	Кофеин	0,45 ± 0,02
Циклофосфамид	0,66 ± 0,02	Гипохлорид натрия	0,44 ± 0,02
Метиленхлорид	0,66 ± 0,03	Бензол	0,43 ± 0,03
7,12-Диметилбенза(а)антрацен	0,66 ± 0,03	Гидрохинон	0,43 ± 0,03
Винилхлорид	0,66 ± 0,02	Этиловый спирт	0,43 ± 0,04
Хлорамбуцил	0,66 ± 0,04	3-Аминотриазол	0,43 ± 0,04
N-нитрозодиметиламин	0,66 ± 0,02	Диэтилстильбэстрол	0,43 ± 0,02
Каптан	0,66 ± 0,03	Катехол	0,42 ± 0,03
2-Нитро-р-фенилендиамин	0,66 ± 0,04	Сафрол	0,42 ± 0,04
Уретан	0,66 ± 0,03	Хлороформ	0,42 ± 0,02
ТиоГЕФ	0,66 ± 0,03	Сахарин (натриевая соль)	0,40 ± 0,04
Нитрит натрия	0,66 ± 0,01	Пронеталол гидрохлорид	0,36 ± 0,02
Метранидазол	0,65 ± 0,03	Аспирин	0,35 ± 0,04
АФ-2 (консервант)	0,65 ± 0,02	р,р-ДДЕ	0,32 ± 0,03
Афлатоксин В1	0,65 ± 0,03	Ацетат свинца	0,32 ± 0,02
2-Ацетиламинофлуорен	0,65 ± 0,02	ДДТ	0,31 ± 0,03
Бензилхлорид	0,65 ± 0,03	Резорцинол	0,21 ± 0,03
1,2-Дибромэтан	0,65 ± 0,02	Сахарин	0,21 ± 0,02
Карбарил	0,65 ± 0,03	Сахароза	0,21 ± 0,03
Гидразин сульфат	0,65 ± 0,03	Хлорид натрия	0,21 ± 0,03
Кверцетин	0,65 ± 0,03	Эугенол	0,20 ± 0,04
Стирен оксид	0,65 ± 0,04	Ерфлуран	0,20 ± 0,03
Формальдегид	0,65 ± 0,03	Смесь ксилолов	0,19 ± 0,04
Эпихлоргидрин	0,65 ± 0,02	Трихлорэтилен	0,14 ± 0,03
Стирол	0,65 ± 0,02	Толуол	0,13 ± 0,03
IQ	0,64 ± 0,03		

ем для каждого соединения прогностическую оценку — значения вероятности, информации и уровня доказанности.

Вся эта процедура, включая разбиение общей выборки на контрольную и обучающую, повторяется заданное число раз. В данном случае это число рав-

нялось 10. В результате мы получаем 10 значений оценок вышеуказанных параметров эффективности анализа. Затем для каждого химического соединения рассчитывается среднее и стандартное отклонение. С использованием этого метода, нами была проведена оценка уровня доказанности канцерогенной

активности для каждого соединения по результатам тестирования в батарее ТЭ+СХО+МЯ. Полученные данные представлены в табл. 5. Значения w представляют собой прогнозируемые уровни доказанности со стандартными ошибками, полученные в результате 10-кратной кросс-проверки. Это означает, что канцерогенная активность для грызунов любого химического соединения, не входящего в данную выборку, может быть спрогнозирована этим методом на основании результатов, полученных при тестировании в батарее из 3-х тестов ТЭ+СХО+МЯ.

Таким образом, показано, что результаты, полученные в тесте Эймса, имеют наибольшую прогностическую значимость в отношении канцерогенной активности на грызунах. Вторыми по значимости являются результаты теста на индукцию СХО. Используемая в работе выборка может служить в качестве обучающей для прогнозирования канцерогенной активности химических соединений методом кросс проверок.

Предложенная в данной работе на примере 3-х тест-систем методология оценки прогностической значимости батарей тестов позволяет проводить оценку эффективности тестирования и достигаемых при этом значений уровня доказанности канцерогенной опасности для произвольных батарей, включающих в себя различное число и набор тестов. Работа выполнена при финансовой поддержке гранта РФФИ 902-04-49849 и Программы фундаментальных исследований РАН «Динамика генофондов растений, животных и человека» (проект № 24П-ИОГ-0302004).

Литература

1. *McCann J., Choi E., Yamasaki E., Ames B.N.* Detection of carcinogens as mutagens in the Salmonella/microsome test: Assay of 300 chemicals // *Proc. Natl. Acad. Sci. (USA)*. — 1975. — Vol. 72. — P. 5135–5139.
2. *Chankong V., Haimes Y.Y., Rosenkranz H.S., Pet-Edwards J.* The carcinogenicity prediction and battery selection (CPBS) method: a Bayesian approach // *Mutation Res.* — 1985. — Vol. 153. — P. 135–166.
3. *Pet-Edwards J., Chankong V., Haimes Y.Y., Rosenkranz H.S.* Application of the carcinogenicity prediction and battery selection (CPBS) method to the Gene-Tox data base // *Mutation Res.* — 1985. — Vol. 153. — P. 187–200.
4. *Heinze J.E., Poulson N.K.* The optimal design of batteries of short-term tests for detecting carcinogens // *Mutation Res.* — 1983. — Vol. 117. — P. 259–269.
5. *Lave L.D., Omenn G.S., Heffernan K.D., Dranoff G.* A model for selecting short-term tests of carcinogenicity // *J. Am. Coll. Toxicol.* — 1983. — Vol. 2. — P. 125–130.
6. *Lave L.D., Omenn G.S.* Cost-effectiveness of short-term tests for carcinogenicity // *Nature (London)*. — 1986. — Vol. 324. — P. 29–34.
7. *Benigni R., Giuliani A.* Which rules for assembling short-term test batteries to predict carcinogenicity? // *Molec. Toxicol.* — 1987. — Vol. 1. — P. 143–166.
8. *Тарасов В.А.* Принципы качественной и количественной оценки генетической опасности химических веществ / Докл. Междунар. симп. Мутагены и канцерогены окружающей среды и наследственность человека // М.: 1994. — С. 3–66.
9. *Тарасов В.А., Асланян М.М., Абилев С.К.* Принципы формализованной количественной оценки опасности химических соединений для человека // *Генетика*. — 1999. — Т. 35. — С. 1585–1599.
10. *Тарасов В.А., Тарасов А.В., Любимова И.К., Асланян М.М.* Проблема количественной оценки опасности химических соединений в генетической токсикологии // *Усп. соврем. биол.* — 2002. — Т. 122, № 2. — С. 136–147.
11. *Тарасов В.А., Велибеков Р.М., Любимова И.К., Асланян М.М.* Низкая эффективность краткосрочных тестов при оценке потенциальной мутагенной опасности химических соединений // *Генетика*. — 2001. — Т. 37, № 7. — С. 1008–1017.
12. *Тарасов В.А., Абилев С.К., Велибеков Р.М., Асланян М.М.* Эффективность батарей тестов при оценке потенциальной мутагенной опасности химических соединений // *Генетика*. — 2003. — Т. 39, № 10. — С. 1406–1417.
13. *Кастлер К.* Азбука теории информации / В кн.: Теория информации в биологии // М.: Иностран. литература. — 1960. — С. 9–53.
14. *Wiener N.* Cybernetics, or control and communication in the animal and the machine // N.Y.: J. Wiley and Sons. Inc. — 1948.
15. *Shannon C.E.* A mathematical theory of communication // *Bell Syst. Techn. J.* — 1948. — Vol. 27. — P. 379–423.

A prognostic efficiency of short-term tests — systems at an estimation of cancerogenic activity of chemical compounds

Abilev S.K., Tarasov A.V., Tarasov V.A.

Vavilov Institute of the General Genetics, Russian Academy of Sciences, Moscow

☼ **SUMMARY:** The method of an estimation prognostic efficiency of the battery of short-term tests at an estimation of potential cancerogenic activity of chemicals is offered. Levels of validity of cancerogenic activity 83 chemical compounds is carried out results at their testing in battery tests — test Ames, the test for SCE in cells *in vitro*, and the MNT — test *in vivo*. Researched date base is used as the trainee for forecasting potential cancerogenic activity of new chemical compounds by a method of cross-check.

☼ **KEY WORDS:** cancerogenic activity; prognostic efficiency of short-term tests; weight of evidence for carcinogenicity.