

И. А. Ефремов, Ю. А. Серегин

РАСЧЕТЫ ИНДЕКСА И ВЕРОЯТНОСТИ ОТЦОВСТВА

В СУДЕБНО-МЕДИЦИНСКИХ ЭКСПЕРТИЗАХ

СЛУЧАЕВ СПОРНОГО РОДСТВА

Лаборатория молекулярной диагностики и геномной дактилоскопии
Государственного научного центра Российской Федерации «ГосНИИгенетика»

При установлении спорного отцовства (материнства) результаты экспертизы в большинстве случаев не вызывают сомнений лишь в случаях исключения родственной связи. В случаях же «подтверждения» (отцовство не исключается) ценность выполненного исследования определяется численной доказательностью экспертного вывода. Молекулярно-генетический метод (исследование ДНК) позволяет на заключительной стадии проводить численную оценку полученных результатов средствами теории вероятностей и математической статистики. Основой метода является сравнительный анализ аллелей в нескольких полиморфных локусах геномов ребенка и заявленных родителей. Поскольку выявляемые генотипы не являются уникальными для каждого человека, любой случай «подтверждения», независимо от количества исследованных локусов, не является логически неопровержимым: всегда существует вероятность случайного совпадения аллелей в сравниваемых образцах ДНК. Принципиальным вопросом является именно адекватность расчетов, поскольку определенное высокое значение («пороговое значение») уровня доказательности можно

считать необходимым и достаточным.

К настоящему времени во многих странах существуют подробные стандарты проведения таких экспертиз и последующей вероятностной оценки полученных результатов. Например, в США минимальное значение вероятности отцовства (Probability of Paternity, PP), приемлемое для рассмотрения ДНК-доказательства в суде, составляет не менее 95 %, а для многих штатов – 99 % и больше [1, 2, 8]. Разработаны и строго регламентированы используемые в исследовании методики, популяционные базы данных, специализированное программное обеспечение (например, с описанием к программе "Pater" можно ознакомиться на сайте *CBFMP*: <http://www.dna-view.com/downloads/documents/paterpdf.zip>) [6].

В России стандарты высокого уровня пока отсутствуют, и правильность интерпретации результатов и последующих расчетов очень часто зависит от компетентности эксперта. С другой стороны, результаты ДНК-исследования нередко принимаются судом безоговорочно, как самое веское доказательство среди прочих материалов дела, в первую очередь, в причине недостаточного понимания узких мест такого

рода экспертиз [12].

В то же время любой вероятностный расчет при исследовании ДНК основан на ряде неочевидных на первый взгляд, но существенных допущений. В отдельных случаях высокие значения вероятности отцовства могут быть некорректными и вводить в заблуждение, если в основу расчетов положены неверные исходные посылки.

Стандартно вероятность отцовства оценивается на основании формулы Байеса для условной вероятности. Для большинства тривиальных случаев априорно (до получения результатов ДНК-исследования) полагается, что никто из близких кровных родственников обследуемого мужчины (его отец, братья, другие дети) не входит в число «предполагаемых отцов». Также исходно обычно предполагается равное соотношение вероятностей двух взаимоисключающих гипотез:

$$p(\text{неОтец}): p(\text{Отец}) = 1:1,$$

где гипотеза *неОтец* – обследуемый мужчина не является отцом (исключение отцовства); гипотеза *Отец* – обследуемый мужчина является истинным отцом обследуемого ребенка.

То есть априорная вероятность (prior probability, pp) полагается равной 0,5. Эта величина является достаточно условной, да и в целом определение численного значения априорной вероятности находится, пожалуй, вне компетенции судебно-медицинского эксперта (скорее, в компетенции суда). Поэтому к настоящему времени в зарубежной экспертной практике при расчете вероятности отцовства требуется прямое указание как соответствующего значения априорной вероятности, так и всех априорных допущений (в том числе описание референтной популяционной группы). На многих официальных сайтах компаний, проводящих такого рода исследования, представлены образцы заключений, содержащие краткое описание интерпретационной части. Например, на сайте компании "Genetica DNA Laboratories, Inc." (США) приведен образец заключения со следующей формулировкой:

The alleged father, XXXXX, cannot be excluded as the biological father of the child named XXXXX. Based on testing results obtained from analyses of 5 different DNA probes, the probability of paternity is 99.9999%. This probability of paternity is calculated by comparing to an untested, unrelated man of the North American Caucasian population (assumes prior probability equals 0.50).

В качестве альтернативы $pp=0,5$ вероятность отцовства может рассчитываться и при других значениях априорной вероятности. Например, в программе "Pater" предусмотрена следующая форма:

Effect of prior probability on the probability of paternity:

prior	posterior
10 %	99.97 %
25 %	99.99 %
50 %	99.99 %+
75 %	99.99 %+
90 %	99.99 %+

Однако возможность адекватного (непредвзятого) выбора априорной вероятности отдельными авторами (тем же Ч. Бреннером) ставится под сомнение, либо вообще отвергается [6].

Существует и другой (менее спорный) стандартный показатель, определяющий значимость выполненной экспертизы: Комбинированный Индекс Отцовства (Combined Paternity Index, CPI, или просто – Индекс Отцовства). Например, в том же образце заключения "Genetica DNA Laboratories, Inc." приводятся обе величины:

Interpretation: Combined Parentage Index 1,703,289; Probability of Parentage 99.9999 %.

Терминологически обычно различают CPI (для нескольких локусов в целом) и PI (Paternity Index, для каждого отдельного локуса). Другой вариант, которого мы и будем придерживаться, – для отдельных локусов оперируют понятием Отношение Правдоподобия (Likelihood Ratio, LR), и тогда для панели локусов в целом используют термин PI.

Отношение правдоподобия LR стандартно рассчитывается для выявленных комбинаций генотипов ребенка и заявленных родителей отдельно для каждого локуса (пример расчетов будет описан ниже). Эта величина показывает, во сколько раз более вероятно, что наблюдаемое совпадение аллелей у заявленного отца и ребенка обусловлено тем, что именно этот мужчина является биологическим отцом, нежели истинным отцом является некто другой (случайный мужчина из той же референтной популяции), не подвергнутый ДНК-исследованию.

Если для исследованных локусов, исходя из их хромосомной локализации, предполагается независимое наследование аллелей (отсутствие сцепления между локусами), то только в этом случае значе-

ния LR могут быть перемножены и рассчитан индекс отцовства (PI):

$$PI = LR(\text{локус 1}) * LR(\text{локус 2}) * \dots * LR(\text{локус n})$$

Отметим, что все ограничения (помимо отсутствия сцепления между локусами), накладываемые на возможность перемножения значений LR или частот аллелей, описываются так называемым «правилом произведения» ("product rule", "multiplication rule") и подробно обсуждаются во многих работах [3, 5, 8].

Индекс отцовства можно интерпретировать следующим образом: исходя из полученных результатов исследования ДНК, в PI раз более вероятно, что именно обследуемый мужчина и передал ребенку «отцовский набор аллелей», чем такие результаты можно объяснить случайным совпадением генотипов.

В случае принятия априорной вероятности $pp=0,5$, вероятность отцовства PP рассчитывается из индекса отцовства PI по простой формуле:

$$PP = [PI/(1+PI)]$$

В России к настоящему времени в области молекулярно-генетических экспертиз существуют только краткие методические указания [13]. В них описан несколько иной алгоритм, чем в [11]. Кратко вместо отношений правдоподобия используются так называемые «локальные статистические частоты» r для каждого из исследованных локусов. Эта величина, по сути, является частотой встречаемости в популяции случайного мужчины, имеющего в генотипе аллель, идентичный аллелю отцовского происхождения в генотипе ребёнка [10]. Значения r рассчитываются по формуле:

$$r = 2p - p^2,$$

где p – частота аллеля отцовского происхождения в генотипе ребенка.

Затем определяется частота встречаемости в популяции человека (R), обладающего «потенциальным отцовским набором аллелей» («правило перемножения»):

$$R = r \text{ (локус 1)} * r \text{ (локус 2)} * \dots * r \text{ (локус n)}$$

Индекс отцовства определяется как величина, обратная R: $PI = (1/R)$, и далее вероятность отцовства рассчитывается как $PP = [PI / (1 + PI)]$ (априорная вероятность $pp = 0,5$).

Таким образом, в работе [13] вместо отношений правдоподобия для каждого из исследованных локусов рассчитываются «локальные статистические частоты». Последующая расчетная часть совпадает с описанной для LR, и с этой точки зрения LR и r являются обратными величинами: $LR \sim 1/r$. В зависимости от комбинаций выявленных гено-

типов и частот совпадающих аллелей LR может быть как меньше, так и больше $1/r$, что в итоге отражается на соответствующих величинах индекса и вероятности отцовства (в конце будет приведен пример сравнительных расчетов).

Важно, что в отличие от $1/r$, LR может принимать значения и меньше 1 (при совпадении по частым аллелям), и с этой точки зрения алгоритм, использующий отношения правдоподобия, является более консервативным и, соответственно, более предпочтительным для экспертной практики.

Когда отцовство в целом (исходя из сравнительного анализа генотипов) не исключается, значения $LR < 1$ для отдельных локусов свидетельствуют «против» отцовства, уменьшая значения CPI и PP. В этом плане актуально не общее количество исследован-

ных в экспертизе локусов, а их «качество»: малопригодны низкополиморфные локусы с доминированием 1-2 аллелей.

Для словесной интерпретации ценности выполненной экспертизы часто используется классификация, разработанная Конрадом Хуммелем [7]. Исходно в этой классификации различным «пороговым» значениям вероятности отцовства (рассчитанным при априорной вероятности 0,5) приписывались различные словесные формулировки [7, 14]. В терминах отношений правдоподобия и индекса отцовства данная классификация также может быть успешно использована [6]. Учитывая относительно широкое распространение (в том числе в России) «предикатов Хуммеля», уместно привести здесь классификацию полностью (таблица)

Таблица

Словесная классификация численных значений вероятности отцовства*

PP, %**	Словесные формулировки			PI***
	Немецкий	Английский	Русский	
(99,9)	Vaterschaft praktisch erwiesen	Paternity practically proven	Отцовство практически доказано	399
99,75	Vaterschaft hochst wahrscheinlich	Highly likely	Отцовство в высшей степени вероятно	
99	Vaterschaft sehr wahrscheinlich	Very likely	Отцовство весьма вероятно	99
90	Vaterschaft wahrscheinlich	Likely	Отцовство вероятно	19
(50)	(Ohne Predikat)	(No verbal predicate)	Зона неопределенности	9
10	Vaterschaft unwahrscheinlich	Unlikely	Отцовство маловероятно	1/9
5	Vaterschaft sehr unwahrscheinlich	Very unlikely	Отцовство весьма маловероятно	1/19
1	Vaterschaft hochst unwahrscheinlich	Highly unlikely	Отцовство в высшей степени маловероятно	1/99
0,2 (0,1)	Vaterschaft praktisch ausgeschlossen	Practically excluded	Отцовство практически исключается	1/399

Примечания: 1. *Классификация приводится по [6], и русский вариант несколько отличается от описанного в [14]. 2. **PP – вероятность отцовства, рассчитанная для $pp = 0,5$. 3. ***PI – эквивалентные численные значения индекса отцовства.

Согласно этой классификации, величина численных оценок в определенной степени отражается в «строгости» словесных формулировок. Так, при превышении «порогового» значения $PI=399$ ($PP=0,9975$) значимость экспертного вывода, независимо от реальных величин PI и PP , описывается словосочетанием «отцовство практически доказано». Ясно, что некорректная (или менее консервативная) процедура расчета может повлечь за собой «ужесточение» словесной формулировки: например, экспертный вывод «отцовство весьма вероятно» ошибочно подменяется более значимым «отцовство в высшей степени вероятно».

В определенной степени «предикаты Хуммеля» позволяют ответить на следующий вопрос: насколько убедительными можно считать численные результаты экспертизы по ДНК, особенно при отсутствии других (не биологических) доказательств?

Корректность расчетной части экспертизы в огромной степени определяется используемыми эталонными (референтными) аллельными частотами. Такие частоты определяются при генотипировании репрезентативной выборки (популяционная база данных; подробнее см. [8]). Практически для всех локусов, используемых в судебно-медицинских приложениях, продемонстрирована высокая межрасовая (и в меньшей степени – внутрирасовая) вариабельность уровня и характера полиморфизма. Для исследованных популяций характерны различия как в числе, так и в частотном распределении аллелей по отдельным локусам.

Высокая генетическая гетерогенность популяций является камнем преткновения при выборе эталонных аллельных частот, и в настоящее время в

этом вопросе отсутствует единство мнений. Как следствие, нет и единой практики использования частот, определенных для крупных популяционных выборок (например, для различных рас). Часто лаборатории разных стран в своей практике предпочитают использовать частоты, определенные для более узких популяционных групп (сформированных, например, по национальному или региональному признаку и обладающих меньшей внутренней гетерогенностью). В России аналогичные популяционно-генетические данные пока фрагментарны. В отсутствие эталонных частот для конкретной популяции можно использовать доступные данные для других популяций, применяя специальные коэффициенты пересчета или проводя «консервативную оценку» аллельных частот.

В отдельных работах продемонстрирована существенная вариация значений вероятности отцовства как функции этнической принадлежности. Поэтому в случаях, когда этническая (или расовая) принадлежность истинного отца неизвестна, корректным видится использование эталонных частот для нескольких крупных популяционных групп (или рас). Из получаемых значений индекса и вероятности отцовства в качестве «консервативной оценки» выбирается наименьшее.

Исходя из изложенных выше теоретических предпосылок, нами была разработана компьютерная программа для расчета индекса и вероятности отцовства. Программа написана в системе программирования Inprise Delphi 5.0. Минимальные системные требования: процессор 486MHz, RAM 16Mb, Microsoft Windows 95/98/Me, Microsoft Excel 95.

За основу интерпретации взяты следующие априорные условия:

- для гомозиготных генотипов не предполагается возможность «ложной гомозиготности»;

- материнство рассматривается как бесспорное;

- полагается полная невозможность мутаций;

- помимо заявленного отца, истинным отцом ребенка может быть любой случайный мужчина из референтной популяции, не подвергнутый ДНК-исследованию и не являющийся родственником обследуемого мужчины.

После запуска программы пользователь указывает наблюдаемые комбинации генотипов ребенка и заявленных родителей (или только одного из родителей) для каждого из исследованных локусов. Реализована возможность сохранения и последующего редактирования указанных данных.

Для каждого локуса отношения правдоподобия рассчитываются (команда «рассчитать») по стандартным алгоритмам [4, 9]. При этом программа адресуется к эталонным частотам аллелей исследуемых локусов, содержащимся в отдельных файлах формата Excel (в директории Data). Частоты аллелей для некоторых локусов, уже представленные в соответствующих файлах, являются консервативной оценкой для населения Центральной Европы (сравнительный анализ различных популяционных выборок, печатные работы). Пользователь может самостоятельно менять эталонные частоты для имеющихся локусов по своему усмотрению, а также создавать новые аналогичные файлы для любых других локусов.

Результаты расчетов (LR , PP) и генотипы обследуемых по всем локусам сохраняются в отдельном файле формата

Excel и могут быть прямо использованы при оформлении заключения эксперта. Программа содержит справочное меню; в перспективе предполагается доработка программы для гибкого выбора априорных условий (другие значения p , учет родственных связей).

В заключение рассмотрим пример расчета отношения правдоподобия. Пусть по отдельному локусу выявлена следующая комбинация генотипов:

мать	pp
ребенок	pq
предполагаемый отец	pq

По сути, LR является отношением вероятности того, что заявленный отец является отцом ребёнка (числитель) к вероятности того, что отцом этого ребёнка является другой мужчина (знаменатель).

Мать с генотипом pp с вероятностью 100% передает ребёнку аллель p , а предполагаемый отец может передать аллель q с вероятностью 50%. Соответственно числитель для LR равен $1 \cdot 1/2 = 1/2$, то есть вероятность рождения у таких родителей ребёнка с таким генотипом равна 50%.

Знаменатель рассчитывается следующим образом. От матери по-прежнему ребёнок с вероятностью 100% получает аллель p . Тогда аллель q он должен унаследовать от «случайного» отца, который может быть или гомозиготой pp , или гетерозиготой pq .

Частота встречаемости в популяции гомозиготного мужчины равна q^2 , и с вероятностью 100% он передаст ребёнку аллель q . Вероятность рождения ребёнка с заданным генотипом составит $1 \cdot 1 \cdot q^2 = q^2$.

Частота встречаемости в популяции гетерозиготного по аллелю q мужчины равна $2q(1-q)$, и с вероятностью 50% он передаст ребёнку аллель q . У такой пары родителей веро-

ятность рождения ребёнка с заданным генотипом составит $1 \cdot 1/2 \cdot 2q(1-q) = q(1-q)$.

Знаменатель будет равен сумме этих двух вероятностей: $q^2 + q(1-q) = q$.

Окончательно отношение правдоподобия для такой комбинации генотипов составит $LR = 1/(2q)$. Аналогичным образом расписываются все остальные возможные комбинации.

Используя алгоритм «локальных статистических частот», для этой же комбинации генотипов имеем следующее. Аллелем отцовского происхождения у ребёнка является q . Частота встречаемости в популяции «случайного» мужчины с этим аллелем (гомоили гетерозиготного) составляет: $r = (2q - q^2)$.

При сравнении двух методов расчета получаем $LR < 1/r$ (поскольку $2q > (2q - q^2)$), и при таких комбинациях генотипов оценка PI и PP через отношения правдоподобия более консервативна.

литература

1. *AABB*: American Association of Blood Banks. Standards for parentage testing. – 2d ed. – American Association of Blood Banks, Bethesda, 1994.
2. *Brenner C. H.* Calculation of paternity index // Inclusion Probabilities in Paternity Testing. – American Association of Blood Banks, Arlington, 1983. – P. 632-638.
3. *Brenner C. H.* Evaluation of the Product Rule // Advances in Forensic Haemogenetics. – 1994. – Vol. 5, Springer Verlag. – P. 453-455.
4. *Brenner C. H., Morris J. W.* Paternity Index Calculations in Single Locus Hypervariable DNA Probes: Validation and Other Studies // Proceedings for The International Symposium on Human Identification. – Promega Corporation, 1989. – P. 21-53.
5. *Buckleton J. S., Walsh S., Harbison S. A.* The fallacy of

independence testing and the use of the product rule // *Sci Justice*. – 2001. – Vol. 41 (2). – P. 81-84.

6. *CBFMP*: Charles Brenner's Forensic Mathematics Page: <http://www.dna-view.com>.

7. *Hummel K., Ihm P., Schmidt V.* Biostatistische Abstammungsbegutachtung mit blutgruppenbefunden. – Stuttgart, 1971

8. *NRC*: National Research Council. The Evaluation of Forensic DNA Evidence. – Washington: Natl. Acad. Press, 1996.

9. *Walsh K. A., Buckleton J. S.* A discussion of the law of mutual independence and its application to blood group frequency data // *J. Forensic Sci. Soc.* – 1988. – Vol. 28 (2). – P. 95-98.

10. *Walsh K. A., Buckleton J. S.* Calculating the frequency of occurrence of a blood type for a "random man" // *J. Forensic Sci. Soc.* – 1991. – Vol. 31(1). – P. 49-58.

11. *О возможных затруднениях молекулярно-генетической экспертизы при недостаточно высокой индивидуализирующей значимости результатов (на примере сложного случая оспариваемого материнства) / И. А. Ефремов, В. В. Носиков, Е. Ю. Скоблилов, А. Ф. Законова, П. Л. Иванов // Суд.-мед. экспертиза. – 2001 – №44(1); – С. 11-17.*

12. *Животовский Л. А.* ДНК-идентификация в судебно-медицинской экспертизе: должна ли она приниматься судом безоговорочно? // *Материалы конф. «Актуальные вопросы идентификации личности».* – СПб., 1999. – С. 75-76.

13. *Иванов П. Л.* Использование индивидуализирующих систем на основе полиморфизма длины амплифицированных фрагментов (ПДАФ) ДНК в судебно-медицинской экспертизе идентификации личности и установления родства: Метод. указания (утверждены Минздравом России 19.01.99) // *Суд.-мед. экспертиза.* – 1999. – № 42 (4). – С. 35-41.

14. *Туманов А. К.* Основы судебно-медицинской экспертизы вещественных доказательств. – М.: Медицина, 1975. – 232 с.