

И. А. ЕФРЕМОВ, Н. Н. ЛЕБЕДЕВА, Е. Ю. ЗЕМСКОВА, П. Л. ИВАНОВ

ИССЛЕДОВАНИЕ СВОЙСТВ МОЛЕКУЛЯРНО-ГЕНЕТИЧЕСКОЙ ИНДИВИДУАЛИЗИРУЮЩЕЙ СИСТЕМЫ НА ОСНОВЕ ХРОМОСОМНОГО ЛОКУСА D1S111 В АСПЕКТЕ СУДЕБНО-ЭКСПЕРТНОГО ТИПИРОВАНИЯ ДНК

ФГУ "Российский центр судебно-медицинской экспертизы" Росздрава; Институт молекулярной биологии им. В. А. Энгельгардта РАН; ИЦ РФ ГосНИИ генетика, Москва

Уровень полиморфизма локуса D1S111 хромосомной ДНК человека до настоящего времени охарактеризован недостаточно полно и на относительно малочисленных выборках. Кроме того, в данных литературы имеются определенные неясности и разночтения, касающиеся структурной организации аллелей указанного локуса.

Целью исследования явились сравнительный анализ референтных нуклеотидных последовательностей для локуса D1S111 и определение основных параметров его полиморфизма в репрезентативной выборке российского населения с целью практического использования в качестве индивидуализирующей молекулярно-генетической системы в судебно-экспертных исследованиях.

Изучение этих вопросов представляется актуальным с точки зрения повышения эффективности судебно-медицинских молекулярно-генетических идентификационных исследований с использованием указанного молекулярно-генетического маркера.

Ключевые слова: полиморфизм генов, молекулярно-генетические индивидуализирующие системы, локус D1S111, популяционные обследования населения России

PROPERTIES OF THE MOLECULAR-GENETIC INDIVIDUAL IDENTIFICATION SYSTEM ON THE BASIS OF D1S111 LOCUS IN TERMS OF FORENSIC-EXPERT TYPING OF DNA

I.A. Efremov, N.N. Lebedeva, E.Yu. Zemskova, P.L. Ivanov

The article presents a comparative analysis of reference nucleotide sequences for locus D1S111, estimation of basic parameters of this locus polymorphism in the representative sample of Russian population for use as an individual identification molecular-genetic system in forensic expert examinations.

Key words: gene polymorphism, molecular-genetic individual identification systems, D1S111 locus, population studies in Russia

Полиморфный мини-сателлит D1S111 впервые был выявлен с использованием мультилокусного зонда 33.6 в прицентромержной области длинного плеча хромосомы 1 [5—8]. Этот маркер расположен на 3'-конце гена CD3Z, кодирующего белок-предшественник зета-цепи T3-антигена — поверхностного гликопротеина T-лимфоцитов (TiT3-комплекс). Мини-сателлит D1S111 локализован в области Iq23, а содержащая его референтная последовательность AL359962 общей протяженностью более 67 000 п. н. — в области Iq23.3-q25.1.

В этом локусе выявлено не менее 14 аллелей, содержащих от 9 до 22 повторов, с гетерозиготностью от 53 до 72%.

В 1996 г. отечественными авторами была предложена праймерная система для амплификации аллелей этого маркера, определена нуклеотидная по-

следовательность одного из аллелей, предложена номенклатура аллелей согласно числу тандемных повторов, а также изучено частотное распределение аллелей в выборке из русской популяции Москвы [1]. Это обусловило последующее использование мини-сателлита D1S111 в качестве молекулярно-генетической индивидуализирующей системы в судебно-медицинских приложениях по установлению родства и идентификации личности на территории России и ближнего зарубежья.

Тем не менее уровень полиморфизма этого маркера до настоящего времени охарактеризован недостаточно полно и на относительно малочисленных выборках. Кроме того, в данных литературы имеются определенные неясности и разночтения, касающиеся структурной организации аллелей указанного локуса. Изучение этих вопросов пред-

Референтные нуклеотидные последовательности для мини-сателлита D1S111

GenBank, номер доступа	Дата публикации	Определение	Размер и тип последовательности	Размер ПЦР-фрагмента, п. н.
M30548	03.08.93	Human minisatellite region detected by myoglobin 33-repeat probe, clone lambda 33.6	548 п. н., линейная ДНК	518
—	1996	Нуклеотидная последовательность аллеля 14 из работы [5]	566 п. н., линейная ДНК	566
AK128376	19.02.04	Homo sapiens cDNA FLJ46519 fis, clone THYMU3033649, highly similar to T-cell surface glycoprotein CD3 zeta chain precursor	3008 п. н., линейная кДНК	604
AL359962	22.11.04	Human DNA sequence from clone RP11-104L21 on chromosome 1q23.3-25.1 Contains the 3' end of the POU2F1 gene for POU domain class 2 transcription factor 1, the 3' end of the CD3Z gene for CD3Z antigen zeta polypeptide (TiT3 complex) and two CpG islands, complete sequence	67 173 п. н., линейная ДНК	531

ставляется актуальным с точки зрения повышения эффективности судебно-медицинских молекулярно-генетических идентификационных исследований с использованием этого маркера.

Целью исследования явился сравнительный анализ референтных нуклеотидных последовательностей для локуса D1S111. Также определяли основные параметры полиморфизма указанного локуса в репрезентативной выборке российского населения с целью практического использования в качестве индивидуализирующей молекулярно-генетической системы на территории России.

На I этапе работы проводили сравнительный анализ четырех известных нуклеотидных последовательностей, ограниченных локализацией использованных праймеров [1] (табл. 1).

Отметим, что нуклеотидная последовательность мини-сателлита D1S111 впервые была представлена в электронной базе данных GenBank в 1993 г., согласно опубликованным ранее сведениям [5]. На конец 2004 г. в этой базе данных были представлены 3 референтные последовательности для мини-сателлита D1S111 (см. табл. 1). Четвертая известная нуклеотидная последовательность, соответствующая аллельному варианту 14, описана ранее в работе [1].

По результатам проведенного анализа можно сделать вывод о том, что тандемный повтор D1S111 имеет длину в среднем 37 п. н. и состоит из трех субповторов (табл. 2). В некоторых случаях длина каждого субповтора может отличаться на 1–2 п. н. от указанной величины из-за внутренних делеций или (реже) вставок.

Суммарно сравнительный анализ референтных нуклеотидных последовательностей позволяет предположить следующую "усредненную" структуру для этого мини-сателлита:

Консервативная последовательность → [полный повтор, 37 п. н.]_n → [субповторы 2 и 3] → [полный повтор с делецией 1 основания в субповторе 3, 36 п. н.] → Консервативная последовательность.

Таблица 2

Структура усредненного тандемного повтора мини-сателлита D1S111

Субповтор	Длина (в нуклеотидах)	Последовательность
1	12	tgg agg aag ggc
2	11	tgg agg agg gc
3	14	tcc gga gga agg gc

Так, например, исходя из этой схемы, "усредненный" аллель 13 мини-сателлита D1S111 должен состоять из 12 полных (или почти полных) повторов и еще 1 повтора, в котором отсутствует субповтор 1. Действительно, именно такая структура соответствует представленному на рис. 1 фрагменту длиной 531 п. н. референтной последовательности AL359962.

Соответственно усредненный размер наиболее короткого обнаруженного нами аллеля 9, амплифицируемого с использованием указанных праймеров, должен составлять 383 п. н. Рассчитанные подобным образом размеры для всех известных аллелей приведены в табл. 3. Указывая размеры с точностью до 1 п. н., мы учитываем, однако, возможность их различия в пределах нескольких (1–3) пар нуклеотидов внутри каждой аллельной группы из-за возможных нуклеотидных вставок и делеций. Указанные размеры аллелей отличаются от данных, приведенных в работе [1], на 2 п. н. и, на наш взгляд, являются более точными.

Изложенные данные использовали как опорные параметры для определения основных характеристик полиморфизма локуса D1S111 в репрезентативной выборке российского населения.

Обследовали более 1000 человек из разных регионов Российской Федерации, преимущественно русской национальности (по паспортным данным). По составу это была выборка русских с включен-

```

acaatgtgagtagaggagacc tccacatttgacctggaagt
tggaggaagggc tggaggagggc tccagaggaagggc
tggaggaagggc tggaggagggc tccagaggaagggc
tggaggaagggc tggaggagggc tccggaggaagggc
tggaggaagggc tggaggagggc tccggaggaagggc
tggaggaagggc tggaggagggc tccggaggaagggc
tggaggaagggc tggaggagggc tccggaggaagggc
tggaggaagggc tggaggagggc tccggaggaagggc
tggaggaagggc tggaggagggc tccggaggaagggc
tggaggaagggc tggaggagggc tccggaggaagggc
tggaggaagggc tggaggagggc tccggaggaagggc
tggaggaagggc tggaggagggc tccagaggaaggc
ggttgctcctcactctgtggt

```

Рис. 1. Фрагмент референтной нуклеотидной последовательности AL359962, ограниченной локализацией использованных праймеров (выделены рамками). Отдельные повторы, имеющие длину 37 нуклеотидов, вынесены построчно.

Таблица 3

Распределение аллелей локуса D1S111 в двух выборках ($M \pm m$)

Аллель	Размер, п.н.	Выборка 1		Выборка 2	
		число наблюдений	частота аллеля	число наблюдений	частота аллеля
9	383	27	0,023 ± 0,004	33	0,030 ± 0,005
10	420	18	0,015 ± 0,004	11	0,010 ± 0,003
11	457	5	0,004 ± 0,002	6	0,006 ± 0,002
12	494	131	0,110 ± 0,009	139	0,128 ± 0,010
13	531	10	0,008 ± 0,003	6	0,006 ± 0,002
14	568	11	0,009 ± 0,003	5	0,005 ± 0,002
15	605	402	0,338 ± 0,014	317	0,291 ± 0,014
16	642	12	0,010 ± 0,003	15	0,014 ± 0,004
17	679	22	0,019 ± 0,004	23	0,021 ± 0,004
18	716	447	0,376 ± 0,014	310	0,284 ± 0,014
19	753	43	0,036 ± 0,005	149	0,137 ± 0,010
20	790	10	0,008 ± 0,003	13	0,012 ± 0,003
21	827	21	0,018 ± 0,004	26	0,024 ± 0,005
22	864	20	0,017 ± 0,004	24	0,022 ± 0,004
23	901	3	0,003 ± 0,001	6	0,006 ± 0,002
24	938	3	0,003 ± 0,001	3	0,003 ± 0,002
25	975	2	0,002 ± 0,001	4	0,004 ± 0,002
26	1012	1	0,001 ± 0,001	0	0
Суммарно		1188	1,000	1090	1,003

ными мигрантами из других популяций, которую можно охарактеризовать как выборку из гетерогенной популяции. Общую выборку разделили на 2 группы: в 1-ю (выборка 1) вошли фигуранты судебно-медицинских экспертиз, назначенных в Бюро судебно-медицинской экспертизы Департамента здравоохранения Москвы в 1999—2003 гг. (594 неродственных лица); во 2-ю (выборка 2) — фигуранты судебно-медицинских экспертиз, назначенных в Российском центре судебно-медицинской экспертизы МЗ и СР РФ в 2001—2004 гг. (545 неродственных лиц).

В качестве биологического материала для анализа использовали образцы периферической крови обследованных.

Для амплификации аллелей локуса D1S111 использовали праймерную систему, описанную в работе [1]. Разделение продуктов полимеразной цепной реакции (ПЦР) осуществляли в 1,5—2% агарозном геле с последующей окраской бромистым этидием.

Аллельные варианты локуса D1S111 идентифицировали по внешним стандартам молекулярных масс двумя методами.

Для выборки 1 в качестве стандарта прямого соответствия использовали ис-

кусственно синтезированную "псевдоаллельную лестницу", содержащую 17 негомологичных фрагментов ДНК длиной от 381 до 973 п. н., с шагом в 37 п. н., которые по размерам соответствовали аллелям 9—25 мини-сателлита D1S111.

Аллельные варианты локуса D1S111 для выборки 2 определяли другим методом: не с непосредственным поиском прямого соответствия по размеру, а опосредованно, расчетным путем, с помощью аппаратно-программной системы Kodak EDAS 120 ("Кодак", США), с использованием в качестве маркера ДНК pBR322/AluI. Для вероятностно-статистического анализа данных использовали компьютерные программы RXC (G. Carmody, Канада), Arlequin [13] и PowerStats [15].

На рис. 2 (см. вклейку) представлены репрезентативные результаты типирования полиморфизма длины амплифицированных фрагментов (ПДАФ) ДНК локуса D1S111 двумя методами: с использованием "псевдоаллельной лестницы" для этого маркера и высокомолекулярного стандарта ДНК pBR322/AluI.

В табл. 4 представлено распределение генотипов в двух исследованных выборках. В каждой выборке выявлено примерно 40% генотипов от общего числа возможных. При этом для обеих выборок наиболее частыми оказались 3 генотипа: гетерозиготы 15—18, а также гомозиготы 18—18 и 15—15.

Всего выявили 18 аллелей, содержащих от 9 до 26 повторов (табл. 5). Наиболее распространенными оказались аллели 15 и 18, частота встречаемости которых превысила 0,33 в выборке 1 и 0,28 в выборке 2. Примечательно, что в ходе исследования обнаружили 4 новых, т. е. не описанных ранее, вы-

Таблица 4

Распределение генотипов локуса D1S111 в двух популяционных выборках

Аллели		10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
9	1\2			2			5		1	8	8		4	1			
10				3		1	2			3			2				
11							2	1		2					1		
12	3	4		9\9	1	1	43			40	16	5	6	2	1		1
13		1	1	2	1\0		1			2	2						
14	1			1		1\0	2	1									
15	8	4	1	44	2	2	79\52	1	10	88	38	4	6	7	1	2	1
16							2		1	1	5	1	1	2	1		
17	2	1		4			4			7	3				1		
18	7	4	3	43	2	4	144	8	10	95\59	23	2	7	8			1
19	2			7			17		1	13	0\24	1		4			1
20		1		1		1	1	1		2	1	1\0					
21	2	1					6	1		7	2						
22		2		3			7			6			1				
23							1			1				1		1	
24				1						1			1				
25							1			1							
26										1							

Примечание. Под выделенной диагональю приведено распределение генотипов для выборки 1, над диагональю — для выборки 2.

Основные параметры прикладной информативности локуса DIS111 в двух выборках

Параметр	Выборка 1	Выборка 2
Число обследованных	594	545
Число выявленных аллелей, в скобках — число повторов	18 (9—26)	17 (9—25)
Число наблюдавшихся генотипов из общего количества возможных	66 (39%) из 171	65 (42%) из 153
Наблюдаемая гетерозиготность	0,685	0,732
Ожидаемая гетерозиготность (несмещенная оценка)	0,729 ± 0,008	0,797 ± 0,006
Вероятность случайного совпадения генотипов (pM)	0,116	0,071
Дискриминирующий потенциал	0,884	0,929
Потенциал исключения отцовства	0,406	0,480
Усредненный индекс отцовства	1,59	1,87
Индекс полиморфизма	0,69	0,77

сокомолекулярных аллеля для этого маркера: с 23-го (901 п. н.) по 26-й (1012 п. н.). При этом аллель 26 встретился только 1 раз в гетерозиготном состоянии в выборке 1.

Основные популяционные характеристики локуса DIS111 для обеих выборок приведены в табл. 5. По представленным данным, информативность DIS111 для исследованной выборки превышает или сравнима с информативностью таких мини-сателлитов, как IL1RN-VNTR, RB1-VNTR, IgH-5'VNTR, D17S5, но уступает таким индивидуализирующим системам, как D1S80 и APOB-3'VNTR [1, 3, 4, 14, 16].

Распределение аллелей в целом оказалось сходным в обеих выборках, за исключением двух важных моментов. Во-первых, самый распространенный в выборке 1 аллель 18 в выборке 2 встречался существенно реже (0,376 против 0,284). Во-вторых, аллель 19 в выборке 2 наблюдался в 3,8 раза чаще, чем в выборке 1 (0,137 против 0,036). Кроме того, уровень полиморфизма для выборки 2 оказался существенно выше, чем для выборки 1.

Наиболее вероятное объяснение этому заключается в том, что при электрофоретическом определении размеров амплифицированных фрагментов по нелокусному внешнему высокомолекулярному стандарту ДНК pBR322/AluI имеет место систем-

Таблица 6

Результаты тестов на соответствие РХВ распределения генотипов в двух выборках ($M \pm m$)

Выборка	RXC		Arlequin, вероятность**
	значение критерия *	вероятность	
1-я	45,1498 ^{3*}	0,9999 ± 0,0000	0,0956 ± 0,0067
	57,3198 ^{4*}	0,9998 ± 0,0000	
2-я	69,6927 ^{3*}	0,8100 ± 0,0012	0,0028 ± 0,0010
	85,2789 ^{4*}	0,8295 ± 0,0012	

Примечание. * — алгоритм согласно [12], выполнено 100 000 перестановок. ** — тест, аналогичный точному тесту Фишера, использован модифицированный алгоритм марковских цепей [9]. Параметры теста: Epsilon Value = $1e - 12$; отброшено 10 000 шагов; выполнен 1 млн шагов [11]. ^{3*} — значение критерия χ^2 ; ^{4*} — G-статистики.

ная ошибка. При этом аллель 18 (716 п. н.) может ошибочно определяться как аллель 19 (753 п. н.).

Это объяснение выглядит достаточно убедительным в свете ранее полученных в работе [2] данных, касающихся изучения свойств гетерологичных маркерных систем применительно к фрагментному анализу ДНК, а именно что определение размеров амплифицированных фрагментов по нелокусным внешним стандартам ДНК как в агарозном, так и в полиакриламидном геле сложным образом зависит от экспериментальных условий и может быть некорректным в определенных диапазонах длин. Тем не менее для прояснения этого вопроса требуются отдельные исследования, которые нами сейчас проводятся.

Для проверки соответствия в обеих выборках распределения генотипов локуса DIS111 равновесию Харди-Вейнберга (РХВ) использовали 3 статистических теста. Следует заметить, что в настоящее время для такого рода проверок используют различные статистические тесты. Однако выявление возможных (особенно незначительных) отклонений весьма затруднительно при большом числе нулевых или уникальных классов наблюдений, что имело место и в нашем случае. В отношении мини- и микросателлитных локусов в последнее время обычно используют точные (вероятностные) тесты и в первую очередь критерий Фишера, который в данном случае является наиболее достоверным [17].

По результатам различных тестов (табл. 6), распределение генотипов только для выборки 1 не отклонялось от РХВ. Для выборки 2 при использовании точного теста Фишера выявлено достоверное отклонение от РХВ ($p = 0,003 \pm 0,001$). Сравнение численных значений различных тестов позволяет заключить, что программа RXC дает завышенные значения вероятностей и непригодна для выявления незначительных отклонений от равновесия. Отметим, что эту программу ранее широко использовали для подобных тестов на РХВ.

Кроме того, выявлено статистически значимое различие ($p = 0,0000 \pm 0,0000$) в распределении генотипов между выборками 1 и 2 (использовали точный тест на различия между выборками со следующими параметрами: марковские цепи длиной 1 млн шагов, отброшено 10 тыс. шагов, Epsilon Value = $1e - 12$, [11], программа Arlequin).

Возможно, выявленные различия между выборками 1 и 2 в распределении аллелей и генотипов, параметрах прикладной информативности, а также обнаруженное отклонение от равновесия РХВ для выборки 2 обусловлены тем, что использованному при обследовании последней методу идентификации амплифицированных аллелей присуща определенная системная ошибка. Соответственно может оказаться, что основные популяционные характеристики, полученные по выборке 1, более адекватно характеризуют полиморфизм мини-сателлита DIS111 в русской популяции. Эти отклонения в настоящее время являются предметом дополнительных исследований.

Уточненные молекулярно-генетические характеристики и установленные параметры аллельного полиморфизма мини-сателлитного маркера DIS111 могут быть использованы в качестве референтных величин для интерпретации результатов судебно-медицинских идентификационных экс-

пертиз и исследований в области создания баз молекулярно-генетических данных для населения России.

ЛИТЕРАТУРА

1. *Ефремов И. А., Чистяков Д. А., Носиков В. В.* // Молек. биол. — 1996. — Т. 30, № 2. — С. 307—318.
2. *Земскова Е. Ю., Иванов П. Л.* // 5-й Всероссийский съезд судебных медиков. Астрахань, 28—30 августа 2000. — С. 244—246.
3. *Чистяков Д. А., Гаврилов Д. К., Овчинников И. В., Носиков В. В.* // Молекул. биол. — 1993. — Т. 27, № 6. — С. 1304—1314.
4. *Chistyakov D., Savost'yanov K., Turakulov R. et al.* // Mol. Genet. Metab. — 2000. — Vol. 70, N 3. — P. 214—218.
5. *Jeffreys A. J., Wilson V., Thein S. L.* // Nature. — 1985. — Vol. 314. — P. 67—73.
6. *Jeffreys A. J., Wilson V., Thein S. L.* // Nature. — 1985. — Vol. 316. — P. 76—79.
7. *Jeffreys A. J., Wilson V., Neumann R., Keyte J.* // Nucl. Acids Res. — 1988. — Vol. 16. — P. 10953—10971.
8. *Jeffreys A. J., MacLeod A., Neumann R. et al.* // Genomisc. — 1990. — Vol. 70. — P. 449—452.
9. *Levene H.* // Ann. Mathemat. statist. — 1949. — Vol. 20. — P. 91—94.
10. *Maiste P. J., Weir B. S.* // Genetica. — 1995. — Vol. 96. — P. 125—138.
11. *Raymond M., Rousset F.* // Evolution. — 1995. — Vol. 49. — P. 1280—1283.
12. *Roff D. A., Bentzen P.* // Mol. Biol. Evol. — 1989. — Vol. 6. — P. 539—545.
13. *Schneider S., Roessli D., Excoffier L.* Arlequin ver. 2.000: A Software for Population Genetics Data Analysis. Genetic and Biometry Laboratory, University of Geneva, 2000.
14. *Smolyanitsky A. G., Smolyanitskaya A. I., Popov V. L. et al.* // Forens. Sci. Int. — 2003. — Vol. 137. — P. 100—103.
15. *Tereba A.* // Profiles in DNA. — 1999. — Vol. 2 (3). — P. 14.
16. *Verbenko D. A., Pogoda T. V., Spitsyn V. A. et al.* // Eur. J. Hum. Genet. — 2003. — Vol. 11. — P. 444—451.
17. *Weir B. S.* Genetic Data Analysis II. — Sinauer, 1996.

Рис. 2. Электрофореграмма амплифицированных фрагментов ДНК (2,5% агарозный гель, окрашивание этидиумбромидом). Результаты типирования ПДАФ ДНК локуса D1S11 для выборки из 3 неродственных человек (1—3), в которой представлены: *a* — 5 аллельных вариантов, в том числе достаточно редкий вариант 9 (указан стрелкой). *M* — аллельный маркер локуса D1S11, содержащий аллельные фрагменты с 9-го (381 п. н.) по 25-й (973 п. н.); *б* — 4 аллельных варианта. *M* — маркер молекулярных масс pBR322/AluI.

