

О.А.Образцова, К.А.Алейникова, А.А.Кубанов, Д.Г.Дерябин

ВАРИАБЕЛЬНОСТЬ НУКЛЕОТИДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ГЕНА *arp* У РОССИЙСКИХ ИЗОЛЯТОВ *TREPONEMA PALLIDUM*

Государственный научный центр дерматовенерологии и косметологии, Москва

Цель. Анализ варибельности нуклеотидных последовательностей внутреннего фрагмента гена *arp* у современных российских штаммов *T. pallidum subsp. pallidum*. *Материалы и методы.* В работе использованы 57 клинических изолятов, полученных в 2016 — 2017 гг. из специализированных медицинских учреждений дерматовенерологического профиля Центрального (Калужская область), Северо-Кавказского (Ставропольский край) и Сибирского (Республика Тыва) федеральных округов. Для исследования нуклеотидных последовательностей гена *arp* использована технология капиллярного секвенирования. *Результаты.* Предложена двухраундовая амплификация гена *arp*, обеспечивающая корректное прочтение его внутреннего участка. В составе данных участков описаны 4 варианта 60-нуклеотидных повторов, отличающихся по составу 6, 8 и 15 — 17 триплетов. Охарактеризованы различные комбинации подобных повторов, соответствующие референсному штамму Nichols, глобально распространенной геногруппе Street 14, а также и впервые обнаруженному региональному варианту Stavropol. *Заключение.* Продемонстрирована целесообразность секвенирования гена *arp* как инструмента повышения эффективности молекулярного типирования *T. pallidum*.

Журн. микробиол., 2018, № 3, С. 45—52

Ключевые слова: *Treponema pallidum subsp. pallidum*, молекулярное типирование, ген *arp*, секвенирование

О.А.Образцова, К.А.Алейникова, А.А.Кубанов, Д.Г.Дерябин

Arp GENE NUCLEOTID SEQUENCES VARIABILITY IN RUSSIAN *TREPONEMA PALLIDUM* ISOLATES

State Research Center of Dermatovenerology and Cosmetology, Moscow, Russia

Aim. Analysis of the *arp* gene internal fragment nucleotide sequences variability in modern russian *T. pallidum subsp. pallidum* strains. *Materials and methods.* 57 *T. pallidum* isolates obtained from specialized dermatovenerologic clinics of the Central (Kaluga), the North Caucasus (Stavropol) and the Siberian (Tyva) regions in 2016 — 2017 were used in the study. The sequencing of the *arp* gene was performed using capillary electrophoresis technology. *Results.* A two-round amplification of the *arp* gene have been proposed, which ensures a correct reading of its internal region. Four variants of 60-nucleotide repeats in the internal *arp* fragment are described, which differing in 6, 8 and 15 — 17 codons compositions. Various combinations of these repeats, corresponding to the reference Nichols strain, globally distributed Street 14 genogroup, and the firstly described Stavropol regional variant are shown. *Conclusion.* The prospect of *arp* gene sequencing as a way to increase *T. pallidum* molecular typing efficiency is postulated.

Zh. Mikrobiol. (Moscow), 2018, No. 3, P. 45—52

Key words: *Treponema pallidum subsp. pallidum*, molecular typing, *arp* gene, sequencing

ВВЕДЕНИЕ

Молекулярная эпидемиология, возникшая как результат интеграции молекулярно-генетических методов в традиционные эпидемиологические исследования, с конца XX века стала одним из основных инструментов кон-

троля над распространением и изменчивостью возбудителей инфекционных заболеваний [4]. В наибольшей степени этот подход оказался востребованным при анализе эпидемиологии некультивируемых патогенов, ярким примером которых является возбудитель сифилиса — *Treponema pallidum* subsp. *pallidum* [1].

Отправной точкой для разработки метода генотипирования бледной трепонемы стала расшифровка ее полного генома [6] и последующее сравнение с геномами других представителей вида *T. pallidum* [3, 12], позволившее идентифицировать на бактериальной хромосоме наиболее вариабельные локусы. На основании двух из них — генов *agr* и *trpII* — в 1998 году был предложен первый метод молекулярного типирования *T. pallidum* subsp. *pallidum* [13], в дальнейшем поддержанный CDC (США) и получивший широкое распространение при анализе эпидемиологии сифилиса во всем мире как CDC-метод [2, 10].

Использование для этих целей гена *agr* (acidic repeat protein) определялось особенностями его структуры: наличием консервативных терминальных участков (в том числе мембрана-связывающего домена и сайта для сигнальной пептидазы I), а также находящегося между ними вариабельного фрагмента из 60-нуклеотидных тандемных повторов, занимающего 51% от общей длины гена [9]. При этом предложенный подход к выявлению подобного полиморфизма основан на амплификации соответствующего участка гена *agr* и последующего определения количества 60-нуклеотидных повторов на основе характеристик электрофоретической подвижности ампликона в агарозном геле [13]. В зависимости от числа подобных повторов (от 2 до 22) идентифицируемому изоляту присваивается соответствующий цифровой индекс, в дальнейшем дополняемый буквенным обозначением, соответствующим определенному профилю полиморфизма длин фрагментов рестрикции продуктов амплификации генов *trpII* (16 вариантов).

Многолетний опыт использования CDC-метода заставил констатировать ряд ограничений его дискриминирующей способности, в том числе связанных с выраженным доминированием вариантов *T. pallidum* subsp. *pallidum* с четырнадцатью повторами в гене *agr* [2, 10]. В этой связи, для повышения разрешающей способности метода было предложено дополнительное секвенирование генов *trp0136*, *trp0319*, *trp0548* и 23S рРНК [5], а также вариабельного участка гена *trp0548* (позиции 131 — 215) [11]. При этом последний подход, позволивший разделить подтип 14a на 2, а 14d на 4 отдельные подгруппы, получил наибольшее распространение и в настоящее время интегрирован в систему молекулярного типирования при сифилисе.

Вместе с тем, анализ гена *agr* (слияние *trp0433* — 0434) свидетельствует о неполном раскрытии его дифференцирующего потенциала. Основанием для этого утверждения являются данные о неидентичном характере нуклеотидных последовательностей тандемных повторов, на основании которых могут быть выделены их I, II, III и II/III подтипы [9]. Другой особенностью может являться различный порядок расположения этих подтипов в составе внутреннего фрагмента гена *agr* [7]. Соответственно секвенирование гена *agr* вместо рутинной детекции ПЦР-продуктов его амплификации представляется перспективным направлением совершенствования CDC-метода, приводящим молекулярное типирование *T. pallidum* subsp. *pallidum* к стандартам мультилокусного секвенирования бактериальных геномов.

В этой связи, целью настоящего исследования явился анализ вариабельности нуклеотидных последовательностей внутреннего фрагмента гена *agr* у

современных российских штаммов *T. pallidum* subsp. *pallidum*, выполненный с использованием технологии капиллярного секвенирования (по Сенгеру).

МАТЕРИАЛЫ И МЕТОДЫ

В исследование включены 57 клинических изолятов, полученных в 2016 — 2017 гг. из специализированных медицинских учреждений дерматовенерологического профиля Центрального (Калужская область), Северо-Кавказского (Ставропольский край) и Сибирского (Республика Тыва) федеральных округов. В качестве референса использован *T. pallidum* subsp. *pallidum* штамм Nichols, культивируемый на тестикулярной модели у кроликов.

Выделение ДНК из исследуемого биоматериала проводилось с использованием набора реагентов «Проба-НК» (ДНК-технология, Россия). Присутствие генетического материала *T. pallidum* subsp. *pallidum* в полученных образцах подтверждалось методом ПЦР с праймерами к гену *po1A* (регион 1156 — 1531 пар оснований), кодирующему специфическую ДНК-полимеразу I данного микроорганизма [8].

Рутинное типирование подтвержденных клинических изолятов проводили с использованием CDC-метода [13] путем амплификации гена *arp*, а также амплификации генов подсемейства *trpII* с последующей рестрикцией эндонуклеазой Tru9 I (НПО СибЭнзим, Россия) и оценкой электрофоретической подвижности полученных продуктов в агарозном геле относительно маркеров молекулярных масс GeneRuler 100 bp Plus DNA Ladder (ThermoFisherScientific, США). В соответствии с рекомендацией Marra S.M. et al. [11] дополнительно проводили секвенирование вариабельного участка гена *trp0548*.

Подбор праймеров для секвенирования вариабельного фрагмента гена *arp* осуществляли с использованием банка аннотированных нуклеотидных последовательностей Национального центра биотехнологической информации США (GenBank NCBI USA) [<http://www.ncbi.nlm.nih.gov/genbank/>] и on-line сервиса nucleotide BLAST [<http://www.ncbi.nlm.nih.gov/>].

Амплификация гена *arp* проводилась с использованием термоциклера DNA Engine Dyad Peltier Thermal Cyler (Bio-Rad, США). Очистка ПЦР-продукта после первого этапа амплификации проведена с использованием набора «QIAquick PCR Purification Kit» (Qiagen, Германия). Очищенный ПЦР-продукт был использован для второго этапа амплификации с применением меченых терминирующих нуклеотидов из набора реагентов Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, США). Осажденные продукты ПЦР использовали для проведения секвенирования вариабельного фрагмента гена *arp* на приборе 3130 Genetic Analyzer (Applied Biosystems, США) с применением программного обеспечения 3130 Data Collection v.3.0.

Первичная расшифровка нуклеотидных последовательностей вариабельного фрагмента гена *arp* проведена в программе Sequencing Analysis 5.3.1. Для выравнивания гена *arp* на референсные сиквенсы *T. pallidum* subsp. *pallidum* использована программа Mega 5.

РЕЗУЛЬТАТЫ

Исследование 57 образцов генетического материала *T. pallidum* в соответствии с рутинной процедурой CDC относило их исключительно к 14 молекулярному субтипу по гену *arp*. Наличие гена *arp* с четырнадцатью 60-нуклеотидными тандемными повторами продемонстрировано и в референс-

ном геноме *T. pallidum* штамм Nichols. Последующее типирование по генам *trpII* и *trp0548* позволило частично разделить анализируемую выборку, выделив в ней доминантный (14 d/f — 89,5%) и минорные (14 b/g и 14 c/f по 3,5%; 14 b/f и 14 i/f по 1,7%) молекулярные субтипы.

Однако однотипный характер гена *agr* существенно снизил эффективность дискриминации данных клинических изолятов. Тем самым полученный результат подтвердил декларированную необходимость совершенствования CDC-метода, в частности, путем углубленного анализа гена *agr* с применением технологии капиллярного секвенирования.

Практическое решение этой задачи потребовало внесения изменений в протокол амплификации гена *agr*, поскольку использование предложенных Pillay A. et al. [13] прямого и обратного праймеров как на этапе фрагментации анализируемого участка ДНК, так и в реакции амплификации с применением меченых терминирующих нуклеотидов, не обеспечивало качественного прочтения концевых участков анализируемых нуклеотидных последовательностей. В этой связи, нами была предложена процедура двухраундовой амплификации гена *agr* с использованием двух пар прямого и обратного праймеров (табл. 1). При этом использование первой (внешней) пары праймеров ориентировано на получение первичного ампликона размером 1643 п.о., расположенного между 388 и 1541 нуклеотидами в гене *agr* и наряду с его центральным участком из 60-нуклеотидных тандемных повторов дополнительно включающего фрагменты 5'- и 3'-концевых консервативных участков. В свою очередь, на этапе секвенирования с использованием терминирующих нуклеотидов использовалась вторая (внутренняя) пара праймеров, амплифицирующая внутренний участок гена *agr*, полученного после первого раунда амплификации (рис. 1). Подобный подход обеспечил качественное прочтение участка гена *agr* между 414 и 1409 нуклеотидами (размером 1015 п.о.), результатом которого являлось как четкое определение количества 60-нуклеотидных повторов, так и их первичных нуклеотидных последовательностей.

Результаты проведенного секвенирования во всех случаях подтвердили присутствие во внутреннем участке гена *agr* четырнадцати тандемных повторов, одновременно зафиксировав вариации их нуклеотидного состава, а также

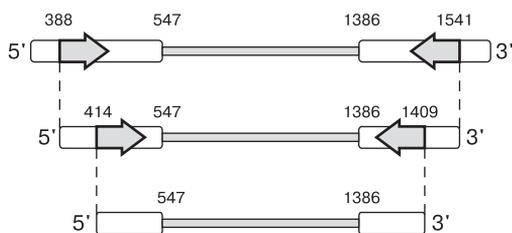


Рис. 1. Схема двухраундовой амплификации гена *agr* *T. pallidum* subsp. *pallidum* с использованием внешней и внутренней пар праймеров.

Таблица 1. Праймеры, использованные для амплификации фрагментов гена *agr* *T. pallidum* subsp. *pallidum*

Название	Последовательность 5'-3'	Длина ПЦР-продукта (п.о.)
Внешние праймеры (первый раунд амплификации)		
agr_F_1	CAA GTC AGG ACG GAC TGT CC	1643
agr_R_1	GGT ATC ACC TGG GGA TGC	
Внутренние праймеры (второй раунд амплификации)		
agr_F_2	ATC TTT GCC GTC CCG TGT GC	1015
agr_R_2	CCG AGT GGG ATG GCT GCT TC	

различный характер взаимного расположения подобных неидентичных повторов.

Анализируемые повторы характеризовались высоким содержанием GC-пар (67%), наличием двух палиндромных последовательностей (CCGGCC и GAGGGGAG), а также вариабельностью пяти триплетов в позициях 6, 8, 15, 16 и 17 (табл. 2). Наиболее частым вариантом являлся

60-нуклеотидный повтор с триплетом GTG в 6 и AAG в 8 позициях, а также триплетами GAG, CGT и GAG в 15 – 17 позициях. Ранее исключительное присутствие аналогичных 60-нуклеотидных повторов было показано у *T. pallidum*, выделенных от обезьян, *T. pallidum* subsp. *pertenuis* и *T. pallidum* subsp. *endemicum* [7], что позволяет предполагать их эволюционную первичность, составившую основу для последующей дивергенции гена *agr*. Вторым по частоте встречаемости являлся повтор, отличающийся от описанного выше вариантом единственным однонуклеотидным полиморфизмом — заменой пиримидинового основания во втором положении 6 триплетов, приводящим к изменению кодируемой аминокислоты с аланина на валин (GCG → GTG; A → V). Более существенными оказывались отличия третьего и четвертого вариантов, нуклеотидные замены в 15 – 17 триплетов которых произошли по типу транзиции, т.е. путем замещения пуринового основания на пуриновое, а пиримидиновое на пиримидиновое. При этом замены пуриновых оснований во втором положении 15 (GAG → GGG) и 16 (CGT → CAT) трипле-

Таблица 2. Варианты нуклеотидных и аминокислотных последовательностей tandemных повторов в гене/белке *agr* *T. pallidum*.

	№ триплетов (аминокислоты)																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Тип повтора																				
I						GT G										GA G				
II						V										R				
III						<u>G</u> C										CGT				
IV																R				

Примечание. Подчеркнуты нуклеотидные и аминокислотные замены относительно I варианта tandemного повтора.

тов повлекли за собой изменение кодируемых аминокислот с глутаминовой кислоты на глицин ($E \rightarrow G$) и с аргинина на гистидин ($R \rightarrow H$), а нуклеотидная замена в третьем положении 17 триплета ($GAG \rightarrow GAA$), напротив, не сопровождалась изменением кодируемой аминокислотной последовательности. Наконец, различия между третьим и четвертым вариантами 60-нуклеотидного повтора заключались в замене двух пуриновых оснований 8 триплета с соответствующим изменением кодируемой аминокислоты с лизина на глицин ($AAG \rightarrow GGG$; $K \rightarrow G$). В целом полученные данные достаточно хорошо согласуются с известными представлениями о вариантах нуклеотидных последовательностей тандемных повторов гена *agr* [7, 9], в то время как предложенная нами нумерация наиболее корректно отражает вероятные последовательные изменения их первичной структуры.

Последующий анализ комбинаций расположения различных вариантов 60-нуклеотидных тандемных повторов во внутреннем участке гена *agr* привел к заключению существованию нескольких молекулярных субтипов данного гена (рис. 2).

Комбинация внутреннего участка гена *agr* у референсного штамма Nichols при прочтении от 5'-конца заключалось в нерегулярном чередовании повторов I и II типов, завершающихся двумя повторами IV типа на 3'-конце фрагмента. Однако формула подобной комбинации 5'-II-II-I-II-I-I-II-II-I-I-I-IV-IV-3' не была зафиксирована ни у одного из современных российских клинических изолятов *T. pallidum*. На этом фоне наиболее распространенной в анализируемой выборке оказывалась комбинация 5'-II-II-I-II-I-I-II-II-IV-I-I-I-III-IV-3', установленная у 54 из 57 исследованных клинических изолятов и по своей последовательности соответствующая внутреннему фрагменту гена *agr* штамма Street 14 [12]. Тем самым соотнесение полученных данных с представлениями о существовании в глобальной популяции *T. pallidum subsp. pallidum* двух обособленных генетических кластеров, известными представителями которых являются штаммы Nichols (Dallas, Chicago B и др.) и Street 14 (Philadelphia-1, AZ 11 и др.), приводило к заключению о принадлежности российских изолятов к последней, также как это было показано для возбудителей сифилиса, циркулирующих в странах Восточной Европы [5]. На этом фоне у трех клинических изолятов, поступивших из Северо-Кавказского федерального округа (Ставропольский край), была обнаружена ранее неизвестная комбинация внутреннего участка гена *agr*, состоящая только из повторов I и II типов, описываемая формулой 5'-II-II-I-II-I-I-II-II-I-II-II-I-I-3' и обозначенная нами как молекулярный вариант Stavropol (рис. 2). Тем самым подобная находка развивает предположение о существовании в глобальной популяции *T. pallidum subsp. pallidum* более значительного разнообразия гена *agr*, что ранее было показано на при-

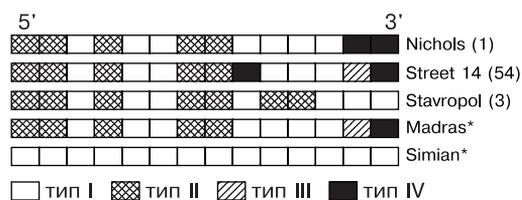


Рис. 2. Структура внутренних участков гена *agr*, состоящих из различных вариантов 60-нуклеотидных повторов.

Варианты нуклеотидных повторов I — IV аналогичны табл. 2. Справа указано наименование известной или впервые описанной комбинации повторов (в скобках — количество штаммов из анализируемой выборки; * комбинации приведены по данным литературы).

мере одного из регионально распространенных вариантов — выделенного в Индии штамма Madras [7].

ОБСУЖДЕНИЕ

Молекулярное типирование *T. pallidum* subsp. *pallidum*, основанное на исследовании вариабельных генов *agr*, *trpII* и *trp0548*, в настоящее время стало распространенным и востребованным инструментом слежения за глобальной и региональной эпидемиологией сифилиса [1, 2]. В то же время, длительный опыт использования данного метода заставил констатировать ряд ограничений его дискриминирующей способности, в том числе определяемых высокой частотой встречаемости гена *agr* с четырнадцатью 60-нуклеотидными тандемными повторами.

Предложенным в настоящем исследовании подходом к повышению эффективности генотипирования *T. pallidum* явился переход от электрофоретической детекции ампликонов гена *agr* к секвенированию его внутреннего вариабельного участка, реализованный с использованием оригинальной двухраундовой процедуры амплификации.

Проведенное секвенирование гена *agr* позволило получить три согласующихся результата: 1) отнести все исследуемые гены к вариантам с 14 нуклеотидными повторами; 2) подтвердить существование 4 вариантов 60-нуклеотидных повторов, отличающихся по составу 6, 8 и 15 — 17 триплетов [7, 9]; 3) описать различные комбинации подобных неидентичных повторов. При этом принципиально важным результатом стало выявление в российской популяции нескольких вариантов гена *agr*. Подавляющее большинство из них (54 из 57) относились к варианту, характерному для геногруппы Street 14, что впервые показывает её доминирование в обследованных субъектах Российской Федерации. С другой стороны, вариант последовательного расположения 60-нуклеотидных повторов, характерный для геногруппы Nichols, был зафиксирован только у соответствующего референсного штамма, что свидетельствует о произошедшем полном вытеснении данного генотипа. Наконец еще одной принципиально важной находкой стало обнаружение ранее неизвестного варианта гена *agr*, имеющего четко ограниченный ареал распространения.

Таким образом, результаты проведенного исследования принципиально подтвердили целесообразность секвенирования гена *agr* для повышения эффективности молекулярного типирования *T. pallidum*, позволив дифференцировать вариант данного гена с четырнадцатью 60-нуклеотидными тандемными повторами на несколько молекулярных субтипов. Планируемое увеличение количества исследуемых клинических изолятов с вовлечением субъектов Российской Федерации с традиционно высоким уровнем заболеваемости сифилисом потенциально должно еще более расширить дискриминирующую эффективность предложенного метода, позволив охарактеризовать пандемический и региональные молекулярные варианты *T. pallidum* subsp. *pallidum*. Одновременно продолжение исследований в обозначенном направлении явится шагом, приводящим молекулярное типирование возбудителя сифилиса к стандартам мультилокусного секвенирования бактериальных геномов.

Исследование выполнено в рамках Государственного задания ФГБУ «ГНЦДК» Минздрава России на плановый период 2018 — 2020 гг.

ЛИТЕРАТУРА

1. Белозоров А.П., Сокол О.А. Молекулярное типирование микроорганизма *Treponema pallidum* подвид *pallidum*. Дерматология та венерология. 2013, 4 (62): 5-11.
2. Кубанов А.А., Воробьев Д.В., Обухов А.П., Образцова О.А., Дерябин Д.Г. Молекулярная эпидемиология *Treponema pallidum* в приграничном регионе Российской Федерации (Республика Тыва). Молекулярная генетика, микробиология и вирусология. 2017, 1: 30-34.
3. Sejková D., Zobaníková M., Chen L. et al. Whole genome sequences of three *Treponema pallidum* ssp. *pertenue* strains: yaws and syphilis treponemes differ in less than 0.2% of the genome sequence. PLoS Negl. Trop. Dis. 2012, 6(1). doi: 10.1371.
4. Dekker M.C., van Duijn C.M. Prospects of genetic epidemiology in the 21st century. Eur. J. Epidemiol. 2003, 18 (7): 607-616.
5. Flasarová M., Pospíšilová P., Mikalová L. et al. Sequencing-based molecular typing of *Treponema pallidum* strains in the Czech Republic: all identified genotypes are related to the sequence of the SS14 strain. Acta Derm. Venereol. 2012, 92: 669-674.
6. Fraser C.M., Norris S.J., Weinstock G.M. et al. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. Science. 1998, 281 (5375): 375-388.
7. Harper K.N., Liu H., Ocampo P.S. et al. The sequence of the acidic repeat protein (arp) gene differentiates venereal from nonvenereal *Treponema pallidum* subspecies, and the gene has evolved under strong positive selection in the subspecies that causes syphilis. FEMS Immunol. Med. Microbiol. 2008, 53 (3): 322-332.
8. Liu H., Rodes B., Chen C.Y., Steiner B. New tests for syphilis: rational design of a PCR method for detection of *Treponema pallidum* in clinical specimens using unique regions of the DNA polymerase I gene. J. Clin. Microbiol. 2001, 39 (5): 1941-1946.
9. Liu H., Rodes B., George R., Steiner B. Molecular characterization and analysis of a gene encoding the acidic repeat protein (Arp) of *Treponema pallidum*. J. Med. Microbiol. 2007, 56 (6): 715-721.
10. Ma D.Y., Giacani L., Centurión-Lara A. The molecular epidemiology of *Treponema pallidum* subspecies *pallidum*. Sex Health. 2015, 12 (2): 141-147.
11. Marra C.M., Sahi S.K., Tantaló L.C. et al. Enhanced molecular typing of *Treponema pallidum*: geographical distribution of strain types and association with neurosyphilis. J. Infect. Dis. 2010, 202: 1380-1388.
12. Matejkova P., Strouhal M., Smajs D. et al. Complete genome sequence of *Treponema pallidum* ssp. *pallidum* strain SS14 determined with oligonucleotide arrays. BMC Microbiol, 2008, 8:76. doi:10.1186/1471-2180-8-76.
13. Pillay A., Liu H., Chen C.Y. et al. Molecular subtyping of *Treponema pallidum* subspecies *pallidum*. Sex. Transm. Dis. 1998, 25 (8): 408-414.

Поступила 25.12.17

Контактная информация: Дерябин Дмитрий Геннадьевич, д.м.н., проф.,
107076, Москва, ул. Короленко, 3, стр. 6, р.т. (499)785-20-40
