

Молекулярно-генетический механизм ПЦР-диагностики заболевания COVID-19, вызываемого коронавирусом SARS-CoV-2

Г.Г. ГОНЧАРЕНКО

В ходе исследования были охарактеризованы молекулярно-генетические механизмы ПЦР-диагностики коронавируса SARS-CoV-2 и приведены примеры надежных ДНК-праймеров для идентификации целевых генов SARS-CoV-2, кодирующих вирусный спайковый S и нуклеокапсидный N белки. Рассмотрен механизм мультиплексного ПЦР-тестирования при использовании трех пар праймеров для S, N и RdRP генов SARS-CoV-2.

Ключевые слова: ПЦР-диагностика, коронавирус SARS-CoV-2, ДНК-праймеры.

The study characterized the molecular genetic mechanisms of PCR diagnostics of the SARS-CoV-2 coronavirus and provided examples of reliable DNA primers for identifying the target SARS-CoV-2 genes encoding the viral spike S and nucleocapsular N proteins. The mechanism of multiplex PCR testing using three pairs of primers for S, N and RdRp SARS-CoV-2 genes is considered.

Keywords: PCR-diagnostics, SARS-CoV-2 coronavirus, DNA-primers.

Введение. Особо актуальным примером успешного применения технологии молекулярно-генетической идентификации в медицине является заболевание COVID-19, вызываемое коронавирусом SARS-CoV-2. Точная диагностика данного коронавирусного заболевания традиционными методами требует существенных временных и материальных затрат. При этом диагностика заболевания происходит, когда пациент явно болен и инфекция уже существенно проникла в дыхательную систему организма.

В то же время, вследствие большой способности коронавируса SARS-CoV-2 передаваться от человека к человеку при отсутствии явных симптомов, **необходимо было разработать метод** обнаружения коронавируса, для того чтобы быстро и точно диагностировать его у людей, не проявляющих внешних признаков болезни.

Фармацевтические компании и научные лаборатории министерств здравоохранения развитых стран в кратчайшие сроки разработали высокотехнологичные молекулярно-генетические методы ПЦР-диагностики коронавируса SARS-CoV-2 в биологическом материале человека. Однако во многих доступных в продаже наборах для ПЦР-диагностики SARS-CoV-2 нет открытой информации **о точной последовательности ДНК-праймеров**, которые в них содержатся, что делает невозможным проверить и подтвердить чувствительность и надежность каждого комплекта.

В связи с этим целью нашего исследования было охарактеризовать молекулярно-генетический механизм ПЦР-диагностики коронавируса SARS-CoV-2 и привести примеры надежных ДНК-праймеров для идентификации целевых генов SARS-CoV-2, включая ген, кодирующий вирусный спайковый S белок.

Характеристика генома коронавируса SARS-CoV-2. Коронавирусы MERS-CoV, SARS-CoV и его модификации (штаммы), в том числе SARS-CoV-2, относятся к роду *Betacoronavirus* из семейства *Coronaviridae*. Все представители этого семейства являются одноцепочечными РНК-содержащими, причем в их геномах РНК всегда в форме РНК (+), т. е. 5'-3' и следовательно, эти вирусы относятся по Балтимору к IV группе.

Вирионная геномная РНК SARS-CoV-2 представляет собой одноцепочечную молекулу 5'-РНК-3' (РНК+) размером около 30 кб (30 000 нуклеотидов), которая на 5'-конце имеет метилированный гуанин (кэп), а на 3'-конце полиА-хвост, подобно эукариотическим мРНК. Это «дезинформирует» инфицируемые вирусом эукариотические клетки человека, и они начинают синтезировать на своих рибосомах, но по матрице коронавирусной РНК+ чужеродные для себя белки SARS-CoV-2.

На рисунке 1 представлено схематическое изображение структуры коронавируса SARS-CoV-2, показывающее основные вирусные белки, их названия и локализацию. На по-

верхности коронавируса хорошо видны так называемые спайковые белки S (шипы), формирующие как бы корону SARS-CoV-2 (рисунок 1). Именно с помощью этих S шипов коронавирус прикрепляется к мембранам эпителиальных клеток человека и проникает в их цитоплазму. Хорошо видно, что спайковые S-белки прикреплены к вирусной мембране, в состав которой входят мембранные белки M и конвертные белки E.

Геномная РНК+ SARS-CoV-2 расположена внутри коронавируса в виде одноцепочечной молекулы 5'–3' и скреплена с нуклеокапсидным белком N. В самом центре вирусной частицы показано наличие важного коронавирусного фермента РНК-зависимой РНК полимеразы (RdRP) (рисунок 1).

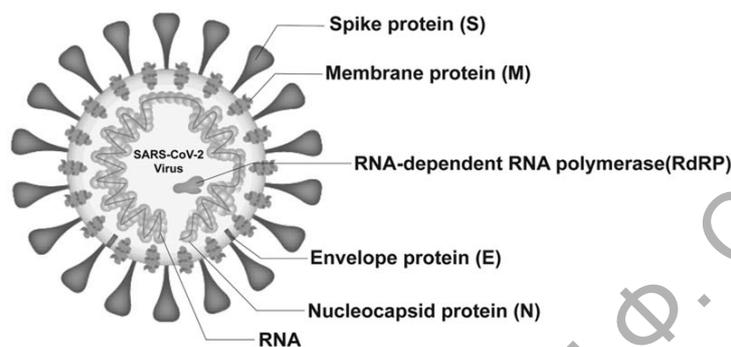


Рисунок 1 – Схематическое изображение структуры коронавируса SARS-CoV-2, показывающее основные вирусные белки, их названия и локализацию [1]

Генетическая карта генома SARS-CoV-2 с обозначением основных вирусных генов, их размеров и порядка расположения представлена на рисунке 2. Гены, кодирующие основные структурные белки коронавируса, расположены в правой части генома SARS-CoV-2.

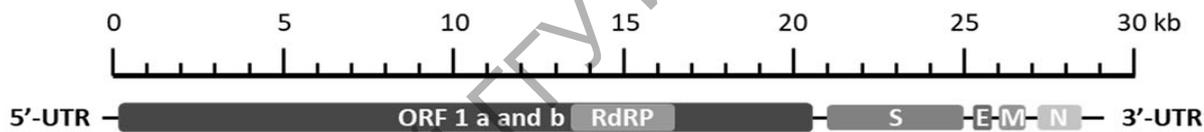


Рисунок 2 – Генетическая карта SARS-CoV-2 с обозначением основных вирусных генов, их размеров и порядка расположения [1]

Необходимо подчеркнуть, что геном коронавируса SARS-CoV-2 является относительно небольшим, нуклеотидная последовательность всех его генов давно установлена и в целом составляет 29 903 нуклеотида. Огромную роль в исследовании коронавирусов SARS-CoV, MERS и SARS-CoV-2 внесли молекулярные генетики под руководством профессора Ральфа Барика (Ralf Baric) из университета Северной Каролины и его ученицы (аспирантки) Ши Женли (Shi Zhengli), которая вернулась в Китай и возглавила молекулярно-генетическую лабораторию в институте вирусологии в Ухане [2]–[5]. Именно в китайском Ухане в конце 2019 г. появились первые пациенты заболевшие COVID-19, вызываемые новым коронавирусом, получившим название 2019-nCoV, а позднее SARS-CoV-2 [6]. В международной базе данных GISAID [7] уже к осени 2020 г. были представлены более 40 тысяч секвенированных геномов SARS-CoV-2 у людей с заболеванием. База сиквенсов SARS-CoV-2 постоянно пополняется, и это позволяет анализировать и отслеживать возникновение и скорость распространение возникающих *de novo* мутаций.

В работе китайских генетиков, посвященной анализу мутаций у вируса SARS-CoV, вызвавшего заболевание атипичной пневмонии в Китае в 2003 г., было показано, что скорость мутаций составила в среднем $\approx 1,5 \cdot 10^{-3}$ на нуклеотид в год [8]. По данным, полученным итальянскими исследователями, касающихся мутабельности коронавируса SARS-CoV-2, на первую половину 2020 г. она составила $\approx 6,5 \cdot 10^{-3}$ на нуклеотид в год, что несколько выше, чем у SARS-CoV [9].

ПЦР-диагностика генетического материала SARS-CoV-2. Тест на наличие вируса в биологическом материале пациента проводится на основе *метода полимеразной цепной реакции (ПЦР)*, разработанной еще в 90-е гг. XX ст. молекулярным генетиком Кэрри Мюллисом [10]. Суть метода состоит в том что, на основе пары сконструированных коротких фрагментов ДНК (*ДНК-праймеров*), комплементарных конкретному участку определенного гена, можно *амплифицировать (размножить)* данный участок гена и затем диагностировать его видовую принадлежность.

Поскольку генетическим материалом SARS-CoV-2 является РНК+, то после ее выделения из вируса нужно преобразовать РНК+ в комплементарную ей двух цепочечную молекулу ДНК (кДНК), необходимую для ПЦР-анализа. Эта процедура проводится с помощью двух специальных ферментов *обратная транскриптаза* и *ДНК полимеразы I*. Именно на фрагменты полученной двухцепочечной молекулы кДНК SARS-CoV-2 *конструируется пара праймеров* для ПЦР-амплификации конкретных участков ДНК целевых генов с последующей видовой идентификацией этих участков на принадлежность к коронавирусу SARS-CoV-2.

При конструировании праймеров важно учитывать их длину, комплементарность выделенному участку ДНК, температуру плавления, соотношение G/C пар. Одним из важных параметров для успешной амплификации является правильно подобранная *концентрация праймеров*, поскольку в ходе ПЦР-реакции часто происходит димеризация и гетеромиризация праймеров, а также их конкуренция за ДНК-полимеразы и дезоксинуклеотидтрифосфаты (dNTPs). Неоптимизированные комплекты праймеров могут произвольно давать ложноположительные результаты, что было показано в исследовании нескольких молекулярно-генетических лабораторий [1], [11]. Имеется сообщение о некорректных результатах в более 70 % случаев, как произошло в Чехии, при использовании закупленных зарубежных тестов [12]. Следовательно, компетентная проверка на валидность и тестирование коммерческих ПЦР-наборов, в особенности ДНК-праймеров, является крайне актуальной. В этой связи, закупка зарубежных ПЦР тест-систем для идентификации коронавируса SARS-CoV-2 в биологическом материале пациентов *должна осуществляться только при наличии информации о структуре используемых в наборе праймеров.*

Конструирование ДНК праймеров для ПЦР идентификации гена S белка SARS-CoV-2. Геном коронавируса SARS-CoV-2 содержит 10 генов и насчитывает 29 903 нуклеотида. Величина S гена, кодирующего спайковый белок, по результатам сиквенса в базе NCBI [13], составляет 3821 нуклеотид, и на карте генома SARS-CoV-2 он локализован в промежутке от 21563 нуклеотида до 25384 нуклеотида [13].

На рисунке 3 представлена нуклеотидная последовательность фрагмента ДНК гена S белка SARS-CoV-2 величиной 600 нуклеотидов от 23881 до 24480 нуклеотида [13] с указанием места прикрепления сконструированной пары праймеров (выделено жирным шрифтом). Сконструированная оптимальная пара праймеров позволяет амплифицировать фрагмент ДНК гена S белка с фланкирующими участками величиной 291 нуклеотидная пара.

Структуры прямого и обратного праймеров приведены ниже:

прямой: 5'- GAGATGCTGGCTTCATCAAA- 3'

обратный: 5'- GGTTGGCAATCAATTTTGG- 3'

```

23881 acaagacaaa aacaccacaag aagtttttgc acaagtcaaa caaatttaca aaacaccacc
23941 aattaaagat tttggtggtt ttaatttttc acaaatatta ccagatccat caaaaccaag
24001 caagaggtca tttattgaag atctactttt caacaaagtg acacttgcatg atgctggctt
24061 catcaaacaa tatggtgatt gccttgggta tattgctgct agagacctca tttgtgcaca
24121 aaagtttaac ggccttactg ttttgccacc tttgctcaca gatgaaatga ttgctcaata
24181 cacttctgca ctgtagcgg gtacaatcac ttctggttg accttgggtg caggtgctgc
24241 attacaata ccatttgcta tgcaaatggc ttataggtt aatggtattg gagttacaca
24301 gaatgttctc tatgagaacc aaaattgat tgccaaccaa ttaaatagtg ctattggcaa
24361 aattcaagac tcactttctt ccacagcaag tgcacttggg aaacttcaag atgtggcaaa
24421 ccaaaatgca caagctttaa acacgcttgt taaacaactt agctccaatt ttggtgcaat

```

Рисунок 3 – Нуклеотидная последовательность фрагмента гена S белка SARS-CoV-2 от 23881 до 24480 нуклеотида

Электрофоретическое фракционирование на предмет видовой идентификации полученного фрагмента ДНК проводилось в агарозном геле стандартными методами [14], [15]. На электрофореграмме (рисунок 4), демонстрирующей спектр фрагмента ДНК S гена, полученного после ПЦР-амплификации с праймерами приведенными выше хорошо видно, что в образцах 1, 3 величина электрофоретических фракций составляет около 300 нуклеотидных пар. Иными словами, размер полученных фрагментов ДНК практически совпадает с ожидаемой величиной в 291 нуклеотидную пару и однозначно указывает на их принадлежность именно к гену S белка SARS-CoV-2. Если в анализируемом биологическом материале человека отсутствует короновиральная матрица кДНК, то сконструированные праймеры не присоединятся к матрице и соответственно никакой участок ДНК не будет амплифицирован, и в этом случае на электрофореграмме фракции ДНК будут также отсутствовать, как в образцах 2, 4 и 6 (рисунок 4).

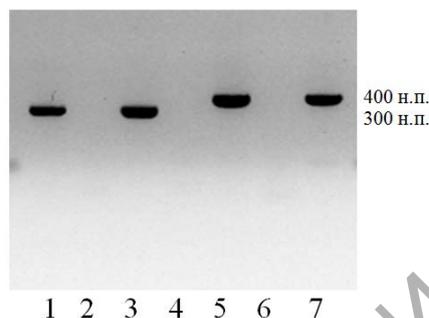


Рисунок 4 – Электрофореграмма демонстрирующая спектры фрагментов ДНК гена S белка и N белка SARS-CoV-2, полученных после ПЦР-амплификации с построенными праймерами: 1, 3 – фрагменты ДНК гена S белка величиной около 300 нуклеотидных пар; 5, 7 – фрагменты ДНК гена N белка величиной около 400 нуклеотидных пар

Конструирование ДНК праймеров для ПЦР идентификации гена N белка SARS-CoV-2. Величина другого гена SARS-CoV-2, который кодирует N белок по результатам сиквенса в базе NCBI [13], составляет 1259 нуклеотидов, и на карте генома SARS-CoV-2 он расположен в районе от 28274 нуклеотида до 29533 нуклеотида [13].

На рисунке 5 представлена нуклеотидная последовательность фрагмента ДНК SARS-CoV-2 величиной 600 нуклеотидов от 28981 до 29580 нуклеотида [13] с указанием места прикрепления сконструированной пары праймеров (выделено жирным шрифтом). Сконструированная пара позволяет амплифицировать фрагмент гена N белка с фланкирующими участками величиной 399 нуклеотидных пар.

Структуры прямого и обратного праймеров для ПЦР-амплификации фрагмента ДНК гена N белка SARS-CoV-2 имеют следующий состав:

прямой: 5'- AAGGAAATTTTGCCACSTTTGCT- 3'

обратный: 5'- GAGTCAGCACTGCTCATGGA- 3'

```

28981 taaaggccaa caacaacaag gccaaactgt cactaagaaa tctgctgctg aggcttctaa
29041 gaagcctcgg caaaaaacgta ctgccactaa agcatacaat gtaacacaag ctttcggcag
29101 acgtgggtcca gaacaaaccc aaggaaattt tggggaccag gaactaatca gacaaggaac
29161 tgattacaaa cattggccgc aaattgcaca atttgcccc agcgcttcag cgttcttcgg
29221 aatgtcgcgc attggcatgg aagtcaaccc ttcgggaacg tggttgacct acacaggtgc
29281 catcaaattg gatgacaaag atccaaattt caaagatcaa gtcattttgc tgaataagca
29341 tattgacgca tacaacaacat tcccaccaac agagcctaaa aaggacaaaa agaagaaggc
29401 tgatgaaact caagccttac cgcagagaca gaagaacacg caaactgtga ctcttcttcc
29461 tgctgcagat ttggatgatt tctccaaaca attgcaacaa tccatgagca gtgctgactc
29521 aactcaggcc taaactcatg cagaccacac aaggcagatg ggctatataa acgttttctg
  
```

Рисунок 5 – Нуклеотидная последовательность фрагмента гена N белка SARS-CoV-2 от 28981 до 29580 нуклеотида

На электрофореграмме в образцах 5 и 7 (рисунок 4), демонстрирующих спектр фрагмента ДНК гена N белка, полученного после ПЦР-амплификации с построенными праймерами, хорошо видно, что величина фракций составляет около 400 нуклеотидных пар. Следовательно, и в этом случае размер полученных фракций ДНК хорошо совпадает с ожидаемой величиной в 399 нуклеотидных пар и четко указывает на его принадлежность именно к гену N белка коронавируса SARS-CoV-2.

Мультиплексное тестирование набора ДНК праймеров для ПЦР идентификации SARS-CoV-2. Рассмотренные выше примеры конструирования и тестирования праймеров для ПЦР-амплификации фрагментов генов S и N, с последующей видовой идентификацией, относятся к *традиционному (простому) ПЦР анализу* для выявления SARS-CoV-2. На практике более эффективным и точным является *мультиплексный метод*, при котором в одной ПЦР реакции используется сразу 3 или 4 пары праймеров для разных генов SARS-CoV-2.

Во многих работах, кроме генов S и N, часто используются фрагменты ДНК гена RdRP, который кодирует у коронавируса фермент *РНК-зависимую РНК полимеразу*. На генетической карте SARS-CoV-2 ген RdRP располагается в центре генома (рисунок 1) в районе от 13442 до 16236 нуклеотидов [13]. Сконструированная оптимальная пара праймеров (прямой: 5'- GCTCGCAAACATACAACCTG- 3' и обратный: 5'- CATTAACATTTGGCCGTGACA- 3') позволяет амплифицировать фрагмент ДНК гена RdRP с фланкирующими участками величиной 202 нуклеотидные пары.

Использование мультиплексного ПЦР-анализа с набором трех пар праймеров для фрагментов S, N и RdRP генов позволяет тестировать более 25 % состава генома коронавируса SARS-CoV-2 в одной реакции. В этом случае на электрофореграмме спектр одного образца выявляется сразу тремя фракциями величиной 202, 291 и 399 нуклеотидных пар, что повышает эффективность и точность идентификации коронавирусной ДНК в биологическом материале пациента. Совокупность проведенных выше исследований убедительно демонстрирует *хорошую работу и надежность сконструированных праймеров* для выявления фрагментов S, N и RdRP генов коронавируса SARS-CoV-2 в биологическом материале.

ПЦР-анализ в реальном времени (PCR real time). Многие современные лаборатории оснащены более дорогими *амплификаторами нового поколения* и могут проводить *ПЦР-анализ в реальном времени (PCR real time)*. ПЦР в реальном времени позволяет не только визуализировать непосредственно ход реакции амплификации, но и дает возможность автоматизировать и ускорить весь процесс. В то же время при этом методе невозможно визуализировать образование ложных димеров ДНК-праймеров.

Традиционные методы ПЦР-анализа удобны, менее дорогостоящи и дают возможность непосредственно фиксировать образование димеров ДНК-праймеров, однако их невозможно автоматизировать из-за стадии проведения электрофореза.

Таким образом, в ходе нашего исследования был охарактеризован молекулярно-генетический механизм ПЦР-диагностики коронавируса SARS-CoV-2 и приведены примеры надежных ДНК-праймеров для идентификации целевых генов SARS-CoV-2, включая гены кодирующие вирусный спайковый S белок и нуклеокапсидный N белок. Рассмотрен механизм мультиплексного ПЦР-тестирования при использовании трех пар праймеров для S, N и RdRP генов SARS-CoV-2.

Автор выражает благодарность членам научного клуба молекулярной генетики и геномики, выпускникам биологического факультета ГГУ им. Ф. Скорины Андрейченко Никите и Нарскину Игорю за помощь и содействие в анализе структуры S и N генов SARS-CoV-2.

Литература

1. Park, M. Optimization of primer sets and detection protocols for SARS-CoV-2 of coronavirus disease 2019 (COVID-19) using CR and real-time PCR / M. Park., J. Won, C.B. Yoon., C. J. Lee. // *Experimental & Molecular Medicine*. – 2020. – V. 52. – P. 963–977. – DOI: 10.1038/s 12276-020-0452-7.
2. Yount, B. Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus / B. Yount, K. M. Curtis, E. A. Fritz, L. E. Hensley [et. al] // *PNAS USA*. – 2003. – Vol. 100. – P. 12995-13000. – DOI: 10.1073/pnas.1735582100.

3. Becker, M. Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. / M. M. Becker, R. L. Graham, E. F. Donaldson [et. al] // PNAS USA. – 2008. – Vol. 105. – P. 19944–3053. – DOI: 10.1073/pnas.0808116105.
4. Menachery, V. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence / V. D. Menachery, B. L. Yount, K. Debbink [et. al] // Nature medicine. – 2015. – V. 21. – P. 1508-1513. – DOI: 10.1038/nm.3985.
5. Menachery, V. SARS-like WIV1-CoV poised for human emergence / V. D. Menachery, B. L. Yount, A. C. Sims, K. Debbinka [et. al] // PNAS USA. – 2016. – Vol. 113. – P. 3048–3053. – DOI: 10.1073/pnas.1517719113.
6. Wu, F. A new coronavirus associated with human respiratory disease in China / F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu [et. al] // Nature. – 2020. – Vol. 579. – P. 263–269. – DOI: 10.1038/s41586-020-2008-3.
7. GISAID [Electronic resource]. – Mode of access : <https://www.gisaid.org>. – Date of access : 10.07.2020.
8. Zhao, Z. Moderate mutation rate in the SARS coronavirus genome and its implications / Z. Zhao, H. Li, X. Wu [et. al] // BMC Evol Biol. – 2004. – V. 4. – P. 21. – DOI: 10.1186/1471-2148-4-21.
9. Benvenuto, D. The global spread of 2019-nCoV : a molecular evolutionary analysis / D. Benvenuto, M. Giovanetti, M. Salemi, M. Prosperi [et. al] // Pathogens and Global Health. – 2020. – V. 114 (2). – P. 64–67. – DOI:10.1080/20477724.2020.1725339.
10. Mullis, K. B. Specific synthesis of DNA in vitro a Polymerase Catalysed Chain Reaction / K. B. Mullis, F. A. Falloona // Method in Enzymol. – 1987. – Vol. 155. – P. 335–349. – DOI: 10.1016/0076-6879(87)55023-6.
11. Ruis-Vilalba, A. Amplification of nonspecific products in quantitative polymerase chain reactions (qPCR) / A. Ruis-Vilalba, E. van Pelt-Verkuit [et. al] // Biomol. Detect. Quantif. – 2017. – Vol. 14. – P. 7–18. – DOI: 10.1016/j.bdq.2017.10.001.
12. 80 % of Rapid COVID-19 Tests the Czech Republic Bought From China are Wrong [Electronic resource] // Prague Morning. – Mode of access : <https://praguemorning.cz/80-of-rapid-covid-19-tests-the-czech-republic-bought-from-china-are-wrong/>. – Date of access : 10.07.2020.
13. NCBI [Electronic resource]: National Center for Biotechnology Information Search database. – Mode of access : <https://www.ncbi.nlm.nih.gov/nuccore/1798174254>. – Date of access : 14.07.2021.
14. Knoblauch, A. The binding of Mss4 to α -integrin subunits regulates matrix metalloproteinase activation and fibronectin remodeling / A. Knoblauch, G. Goncharenko, S. Ludwig, V. Wixler // FASEB Journal. – 2007. – № 2. – P. 497–510. – DOI: 10.1096/fj.06-7022.
15. Walter, B. The Mss4 protein is to regulate stress response and apoptosis / B. Walter, C. Nordhoft, G. Goncharenko, S. Ludwig, V. Wixler // Cell Death and Disease. – 2012. – № 3, e297. – DOI: 10.1038.