



КОНСТРУИРОВАНИЕ ПРАЙМЕРОВ ДЛЯ ПЦР В ПРОГРАММЕ PRIMER-BLAST

Козырева А. А., Злотина А. М., Головкин А. С., Калинина О. В.,
Костарева А. А.

Федеральное государственное бюджетное учреждение
«Национальный медицинский исследовательский центр имени
В. А. Алмазова» Министерства здравоохранения Российской
Федерации, Санкт-Петербург, Россия

Контактная информация

Козырева Александра Анатольевна,
ФГБУ «НМИЦ им. В. А. Алмазова»
Минздрава России,
ул. Акkuratова, д. 2, Санкт-Петербург,
Россия, 197341.
Email: fairb@yandex.ru

Статья поступила в редакцию 02.05.2021
и принята к печати 01.06.2021.



Резюме

Метод полимеразной цепной реакции (ПЦР) широко применяется для решения разнообразных задач. Так, в клинико-лабораторной диагностике ПЦР используют для детекции патогенов бактериальной и вирусной природы, при диагностике наследственных заболеваний и для установления отцовства. Существует три модификации метода, которые позволяют проводить качественную, полуколичественную и количественную оценку присутствия молекул нуклеиновых кислот. В основе метода лежит многократное удвоение фрагмента ДНК или РНК, в результате чего синтезируется значительное число копий интересующего участка нуклеиновой кислоты. Необходимыми компонентами реакции являются молекулы ДНК или РНК, которые служат матрицей для синтеза новых молекул; полимеразы — фермент, который синтезирует новые цепи ДНК; и короткие олигонуклеотиды — праймеры, которые необходимы для работы фермента и являются специфичными к фрагменту интереса. На данный момент имеется не так много инструментов для конструирования праймеров, которые были бы одновременно простыми в использовании и общедоступными. Одним из таких инструментов является онлайн-ресурс Primer-BLAST, интегрированный в базу данных NCBI. Его преимуществами являются информативность, наглядность и свободный доступ к актуальной информации баз данных, таких как GenBank, RefSeq, данных о геномах и транскриптах. Но и для этого ресурса отсутствуют подробные инструкции по конструированию праймеров. Данная статья является обновлением для выпущенного ранее учебного пособия и представляет пошаговое руководство по поиску целевых участков ДНК, подбору праймеров, и оценке качества имеющихся праймеров. В тексте отражен личный опыт авторов, призванный помочь исследователям, работающим с ПЦР.

Ключевые слова: базы данных, детекция, полимеразная цепная реакция, праймер, ПЦР, NCBI, primer-BLAST.

Для цитирования: Козырева А. А., Злотина А. М., Головкин А. С. и др. Конструирование праймеров для ПЦР в программе Primer-BLAST. Трансляционная медицина. 2021;8(3): 37-52. DOI: 10.18705/2311-4495-2021-8-3-37-52

PRIMER DESIGNING IN PRIMER-BLAST

Kozyreva A. A., Zlotina A. M., Golovkin A. S., Kalinina O. V., Kostareva A. A.

Almazov National Medical Research Centre, Saint Petersburg, Russia

Corresponding author:

Kozyreva Alexandra A.,
Almazov National Medical Research Centre,
Akkuratova street, 2, St. Petersburg, Russia,
197341.

E-mail: fairb@yandex.ru

Received 02 May 2021; accepted 01 June 2021.

Abstract

The polymerase chain reaction (PCR) is widely used in different areas. For example, in laboratory diagnostics, PCR is used to detect bacterial and viral pathogens, in the diagnosis of hereditary diseases, and to identify paternity and many other. There are three types of PCR — qualitative, semi-quantitative, and quantitative. The method is based on the ability of DNA polymerase to double the existing DNA strand, thus resulting in multiplication the number of copies of the region of interest. Necessary components of the reaction are DNA or RNA molecules, serving as a template for the of new molecules; polymerase — an enzyme synthesizes new DNA strands; short oligonucleotides — primers that are required for the enzyme work and are specific to the fragment of interest. Currently, there are not many primer designing tools that are easy-to-use and free. One of these tools is the Primer-BLAST online resource, which is integrated with the NCBI database. It is user-friendly and easy-to-use effective tool that is integrated with GenBank and RefSeq, and always is up-to-dated. However, even for this tool, there are no detailed instructions for the primers designing. This work is an update for a previously published tutorial and provides a step-by-step guide to find target DNA regions, primers designing, and performing primer quality control. The paper is largely based on the personal experience of the authors and is intended for researchers working with PCR.

Key words: databases, detection, PCR, polymerase chain reaction, primer, NCBI, primer-BLAST.

For citation: Kozyreva AA, Zlotina AM, Golovkin AS, et al. Primer designing in Primer-BLAST. Translational Medicine. 2021;8(3): 37-52. (In Russ.) DOI: 10.18705/2311-4495-2021-8-3-37-52

Введение

Метод полимеразной цепной реакции (ПЦР) широко применяется для решения разнообразных задач. Так, в медицине ПЦР используют для поиска патогенов бактериальной и вирусной природы, при диагностике наследственных заболеваний и для установления отцовства.

Разработка метода полимеразной цепной реакции стала одним из наиболее значимых открытий 20 века, которое легло в основу современных методов, применяемых в биомедицинских исследованиях в области молекулярной биологии, геномной инженерии, клинической лабораторной диагностики и других. В свою очередь, праймеры, ключевые компоненты полимеразной цепной реакции, явля-

ются важным параметром специфичности и эффективности результатов ПЦР. Нуклеотидный состав праймеров влияет на качество и температурный режим проведения ПЦР. В настоящий момент имеется достаточно много разнообразных программ для подбора праймеров, большинство из которых интегрированы в платные программные пакеты.

Международная база данных NCBI (Национальный центр биотехнологической информации) предоставляет открытый доступ к перекрестно-интегрированным генетическим базам данных и онлайн-ресурсам, в том числе к базе данных Nucleotide, которая включает в себя последовательности из постоянно обновляющихся источников, таких как GenBank, RefSeq, данные о геномах и транскриптомах, что

в совокупности обеспечивает уникальную основу для эффективного конструирования праймеров с использованием онлайн-блока Primer-BLAST [1].

Данная статья является обновлением для выпущенного ранее учебного пособия и представляет пошаговое руководство по поиску целевых участков ДНК, подбору праймеров и оценке качества имеющихся праймеров [2].

Общая характеристика праймеров

Праймерами называют короткие одноцепочечные синтетические фрагменты ДНК (олигонуклеотиды), имеющие длину от 15 до 50 оснований, комплементарные специфичным участкам матричной ДНК. Праймеры — один из ключевых компонентов ПЦР, от которых зависит качество продукта ПЦР. Праймеры выполняют функцию «затравки» для ДНК-полимеразы, которая присоединяет нуклеотиды к дочерней цепи по принципу комплементарности в направлении от 5' к 3' концу матричной ДНК.

Для получения целевого фрагмента в ПЦР используют два праймера, нуклеотидная последовательность каждого из которых комплементарна противоположным концам определенной цепи целевого фрагмента ДНК: прямой (forward) праймер соответствует началу амплифицируемого участка ДНК, обратный (reverse) праймер — концу амплифицируемого участка ДНК. Таким образом, прямой и обратный праймеры комплементарны противоположным цепям ДНК, что позволяет многократно амплифицировать именно целевой фрагмент ДНК, ограниченный этими праймерами.

При выборе праймеров рекомендуется соблюдать следующие правила:

- Нуклеотидная последовательность праймеров должна быть высокоспецифична матрице ДНК, чтобы обеспечить амплификацию чистого монопродукта в результате ПЦР без амплификации неспецифичных фрагментов ДНК.

- Нуклеотидная последовательность праймеров должна быть максимально (в идеале полностью) комплементарна последовательности ДНК, которые в геноме встречаются не чаще одного раза.

- Последние три нуклеотида 3'-конца праймера должны быть обязательно полностью комплементарны матрице ДНК, так как ДНК-полимераза присоединяет новые нуклеотиды к гидроксильным группам 3'-конца растущей цепи, и для успешного присоединения полимеразе требуется полное конформационное совпадение концевых нуклеотидов на 3'-конце.

- Нуклеотидная последовательность праймеров не должна содержать повторяющиеся элемен-

ты и палиндромы, поскольку это может привести к неизбирательному отжигу и, как следствие, к получению неспецифичного продукта.

- Область отжига праймеров должна происходить на консервативном участке матрицы ДНК, т. е. располагаться вне зон точечных мутаций, делеций или инсерций. При попадании праймера на зону с мутациями возможно отсутствие ПЦР-продукта из-за несоответствия нуклеотидной последовательности праймера к области отжига на матрице, синтез неспецифического продукта из-за некорректного отжига праймера в другой области.

- Оптимальная длина праймера от 16 до 25 нуклеотидов, так как более короткие праймеры могут слабо связываться с матрицей из-за малого количества образующихся водородных связей, быть менее специфичными, поскольку чем короче последовательность праймера, тем менее она уникальна и тем чаще она может встречаться в геноме.

- Соотношение АТ и GC нуклеотидов в последовательности праймера должно быть близко к 1:1. Следует учитывать, что большое количество АТ-оснований приводит к снижению температуры плавления праймеров, а большое количество GC-оснований — соответственно, к ее увеличению.

- Разница в температурах отжига прямого и обратного праймеров должна быть в пределах 1-2 градусов, при увеличении разницы в температуре отжига между праймерами на 3 и более градусов оптимальную температуру отжига подбирают экспериментальным путем, при этом повышается вероятность амплификации неспецифических продуктов. Повышение температуры отжига праймеров улучшает качество ПЦР-продукта за счет уменьшения неспецифичного связывания праймера с матрицей, но при этом количество ПЦР-продукта может быть невелико. Понижение температуры отжига праймеров исключает преждевременную диссоциацию праймеров с матрицей на этапе элонгации ДНК, что дает увеличение количество ПЦР-продукта, но при этом есть риск появления неспецифичных ПЦР-продуктов.

- Для улучшения качества отжига праймеров рекомендуется подбирать нуклеотидные последовательности праймеров так, чтобы последние 2-4 нуклеотида на 3'-конце праймера содержали GC-основания. Для праймера с высоким содержанием GC-оснований данное условие становится менее строгим и позволяет допустить наличие 1-2 АТ-оснований на 3'-конце праймера.

- На 3'-конце праймера не должно быть моно- (например, CCCC), ди- (например, GCGCGC) или тринуклеотидных повторов, поскольку такие по-

вторые могут существенно уменьшать специфичность отжига праймеров, так как достаточно часто встречаются в геноме.

- Нуклеотидные последовательности праймеров не должны образовывать стабильные вторичные структуры, быть само- или взаимокomплементарными. Наличие вторичных структур на праймере, в том числе шпильки на 3'-конце праймера, приводит к уменьшению эффективности связывания праймера с матрицей, затруднению присоединения ДНК-полимеразы и, соответственно, отсутствию элонгации. Взаимная комплементарность праймеров между собой (как прямых, так и обратных) и друг с другом особенно на 3'-концах приводит к образованию и амплификации димеров праймеров в ущерб целевому продукту, при этом чем длиннее участки взаимной комплементарности, тем эффективнее проходит амплификация димеров праймеров.

Размер ПЦР-продукта зависит от поставленной задачи. Для клонирования допустим ПЦР-продукт размером до 1500 п.о., для секвенирования по Сэнгеру он может быть до 1000 п.о. При секвенировании больших фрагментов может потребоваться подбор внутреннего праймера.

Общая информация о базе данных NCBI

Сайт Национального центра биотехнологической информации США (NCBI: <https://www.softwarcbi.nlm.nih.gov>) представляет открытый доступ к различным программным ресурсам и международным генетическим базам данных [3]. Одной из таких платформ является BLAST (Basic Local Alignment Search Tool, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) [4], которая находит области сходства между нуклеотидными или аминокислотными последовательностями, сравнивает их с базами данных последовательностей и вычисляет статистическую значимость. Онлайн-версия BLAST общедоступна, интегрирована с базами данных NCBI, графически отображает результаты поиска в реальном времени, позволяя синхронно анализировать полученные результаты в различных базах данных NCBI. Имеется возможность загрузить офлайн-версию программы для Windows или Linux, но в этом случае от пользователя требуется умение работать с командной строкой.

Для конструирования праймеров используется онлайн-блок Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome), в котором пользователь самостоятельно задает конкретные параметры праймеров, выбирает целевую базу данных, далее программа подключается к базам данных и серверам NCBI,

обрабатывает запрос и отправляет результаты обратно в браузер пользователя в графическом формате. На экране браузера результат выводится в виде списка со всеми найденными совпадениями, включая графическое изображение, что позволяет быстро оценить приемлемость расположения праймеров относительно целевого участка ДНК, который необходимо амплифицировать.

Поиск целевой последовательности ДНК для подбора праймеров с использованием онлайн-ресурса Gene

Онлайн-ресурс Gene (<https://www.ncbi.nlm.nih.gov/gene/?term=>) содержит различную информацию о широком круге видов, включая номенклатуру, референсные (эталонные) последовательности (RefSeqs), генетические вариации, фенотипы и locus-специфичные ресурсы [5].

Для поиска целевой последовательности ДНК в строке поиска необходимо указать название гена (например, *menin 1* или *MEN1*) и нажать Enter. В выпадающем списке результатов, в котором будут даны все варианты данного гена, имеющиеся в базе данных NCBI, необходимо выбрать ген, соответствующий анализируемому виду, например *Homo sapiens*. В браузере откроется страница (<https://www.ncbi.nlm.nih.gov/gene/4221>), содержащая подробную информацию о данном гене, в том числе информацию о структуре и хромосомной локализации гена, в случае аннотированных генов — информацию о геномных регионах, транскриптах и продуктах, уровне экспрессии, ссылки на дополнительные и библиографические ресурсы.

В разделе с графическими изображениями геномных регионов и транскриптов (Genomic regions, transcripts, and products), который является составной частью геномного браузера NCBI Graphical Sequence Viewer (<https://www.ncbi.nlm.nih.gov/tools/sviewer/>), в поле Genomic Sequence необходимо выбрать референсную последовательность для хромосомной аннотации или референсный ген. В данном разделе графически отображается не полный ген, а транскрипты мРНК, у которых количество кодирующих участков может различаться в зависимости от изоформы. Каждый вариант сплайсинга мРНК имеет свой уникальный номер.

В случае поиска хромосомной позиции в поле Genomic Sequence нужно выбрать номер релиза аннотации генома («сборки»), например, для человека это NC_000011.9 Chromosome 11 Reference GRCh37.p13 Primary Assemble или NC_000011.10 Chromosome 11 Reference GRCh38.p13 Primary Assemble в зависимости от того, для какой версии аннотации была задана искомая позиция в геноме

(рис. 1). В поле поиска геномного браузера в окне Find указывается хромосомная позиция, полиморфизм (например, rs386134260) или нуклеотидная последовательность, желательно не короче 15 нуклеотидов для повышения специфичности поиска (рис. 1). Для поиска полиморфизмов или нуклеотидных последовательностей номер сборки генома значения не имеет.

В случае поиска определенного экзона, а также если искомая позиция была задана от начала гена, то в качестве референса необходимо выбрать в поле "Genomic Sequence» RefSeqGene (референсный образец гена). Референс не привязан к позиции в геноме, его нуклеотидные последовательности учитываются не от начала хромосомы, а от начала самого гена. Он есть, как правило, у генов, по которым накоплено достаточно информации. На рисунке 2 показан тот же участок гена, что и выше. Видны не только номера экзонов, но и эталонная матричная РНК и последовательности белка, соответствующие кодирующим частям мРНК.

В случае, если нужно найти позицию, заданную по мРНК, необходимо знать точный вариант транскрипта, так как из-за сплайсинга варианты транскриптов отличаются по длине, числу экзонов и длине кодирующих последовательностей.

Подробные объяснения по каждому отображаемому элементу геномного браузера: гены, РНК, единичная нуклеотидная замена (SNP) и т. д., можно получить, если нажать знак вопроса в верхнем правом углу интерфейса и выбрать пункт Legends (<https://www.ncbi.nlm.nih.gov/tools/sviewer/legends/>). Чтобы передвинуть влево или вправо видимую часть гена, необходимо навести курсор на графическое изображение транскрипта, нажать и держать левую кнопку мыши в области окна с графикой и смещать картинку в нужную сторону.

Для выделения фрагмента генома нужно подвести курсор к «линейке» в верхней части графического изображения, нажать левую кнопку мыши и смещать до нужного места, чтобы выделить одновременно несколько участков генома, надо удерживать кнопку Ctrl на клавиатуре. Если подвести мышку к выделенной зоне в области «линейки» и подождать, появится всплывающее меню (рис. 3), верхняя строчка которого показывает положение выделенного участка в выбранной нами геномной последовательности (NC или NG) и его размер. В меню Zoom On Range увеличивает масштаб выделенного фрагмента на все окно браузера; Zoom To Sequence дает максимальное увеличение масштаба до нуклеотидной последовательности; Set New Marker On Range позволяет поставить метки, Primer-BLAST (Selection) открывает в новом окне

онлайн-ресурс BLAST® blastn suite с фиксированными координатами выделенного фрагмента ДНК для поиска праймеров (см. Query subrange); Download FASTA (Selection) позволяет скачать и сохранить нуклеотидную последовательность выделенного фрагмента в формате FASTA.

У поставленных пользователем маркеров есть свое меню, оно появляется, если навести курсор на его заголовок на линейке и немного подождать. Для конструирования праймеров маркеры используются по большей части для того, чтобы отметить какие-то элементы на некоторый срок. Больше информации о маркерах можно найти здесь: <https://www.ncbi.nlm.nih.gov/tools/sviewer/markers/>.

Маркеры могут быть как точечными (проставляются автоматически, при поиске конкретной хромосомной позиции или полиморфизма), а могут захватывать некоторый диапазон — протяженный участок ДНК (в этом случае маркеры ставятся вручную, над выделенным диапазоном).

Маркеры обычно используют, чтобы зафиксировать в геномном браузере точное местоположение цели своего поиска. Особенно это удобно, если цель находится внутри экзона протяженного размера (более 1000 п.о.), в интроне, или, к примеру, захватывает гибридный участок экзон + часть интрона. Если видимый диапазон последовательности ДНК в геномном браузере смещался в ту или иную сторону, то маркер позволяет сразу вернуться к нужной цели.

Если целевых участков в исследуемой ДНК несколько, то можно отметить их все, затем оценить их взаиморасположение. Иногда получается подобрать одну пару праймеров к 2–3 целям, расположенным относительно близко друг к другу.

Также с помощью маркеров можно оценить, как праймеры расположены по отношению к целевому участку ДНК, и оценить характеристики районов ДНК, в которых они находятся. Для этого отмечают целевой участок ДНК, затем вводят в поле поиска геномного браузера последовательности прямого и обратного праймеров, фиксируя каждый раз установкой маркера найденный и выделенный участок ДНК, затем можно уменьшить масштаб (при поиске коротких последовательностей масштаб геномного браузера выставляется по умолчанию на максимальное увеличение) и оценить результат визуально.

Маркеры сохраняются в геномном браузере при изменении конфигурации его содержимого (через меню Tracks, см. ниже), но сбрасываются при обновлении страницы интернет-браузера или смене текущей аннотации генома.

Меню для участков без выделенных диапазонов появляется при наведении курсора в области «ли-

нейки» и нажатием правой кнопки мыши (рис. 4). В этом режиме Select a Range позволяет выбрать нужный диапазон, Set New Marker At Position ставить маркеры, Primer-BLAST (Visible Range) переходить к дизайну праймеров в онлайн-ресурс BLAST® blastn suite, в данном режиме праймеры будут подбираться ко всей области генома, видимой в окне геномного браузера.

Если в процессе поиска области для праймера нужно быстро получить дополнительную информацию о каком-либо элементе, достаточно привести курсор мышки в виде руки на него, и появится справочное окно (рис. 5), компоновка которого меняется в зависимости от типа просматриваемого элемента. Сначала располагается информация о названии, типе, длине, расположении и т. д. просматриваемого элемента, далее о позиции, на которую указывает курсор. При наличии нуклеотидной последовательности данного элемента в формате FASTA будет дана ссылка на выгрузку соответствующего файла. Дополнительно в меню имеются ссылки для быстрого перехода на различные базы NCBI и внешние ресурсы, связанные с этим объектом.

После того, как интересующий участок найден, для удобства поиска праймеров необходимо изменить отображаемые данные в геномном браузере, для этого в правой части окна геномного браузера выберите Tracks, далее активируйте Configure Tracks (рис. 6).

В появившемся окне в левой части нужно активировать закладку Search Tracks и отметить следующие пункты (рис. 7), наиболее критичные для оптимального дизайна праймеров:

- G+C content (50bp): позволяет оптимально выбрать места для гибридизации праймеров и корректно соблюсти соотношение AT/GC нуклеотидов.

- CpG islands (участки ДНК, богатые метилированными GC-парами). Если такие участки входят в состав праймера, то при проведении ПЦР рекомендуется добавлять DMSO, а также немного увеличивать количество олигонуклеотидов для улучшения выхода ПЦР продукта, или же использовать реагенты от различных производителей, предназначенные для амплификации GC-богатых участков генома.

- Repeats identified by WindowMasker, Repeats identified by RepeatMasker (повторяющиеся элементы в геноме): расположение праймеров внутри этих повторов крайне нежелательно.

Если какой-то параметр не удастся быстро найти, можно воспользоваться полем поиска в верхней части окна (рис. 7).

В результате геномный браузер поменяет свой вид: будут отмечены области CpG островков (чер-

ные полоски), повторы и мобильные элементы генома (черные и синие линии) (Рис. 8). Волнистая полоса будет отражать содержание GC-пар в геноме, при этом тонкая линия внутри этого элемента соответствует 50 %-му количеству GC-пар.

Для зарегистрированных на сайте NCBI пользователей имеется возможность сохранить наборы треков в меню Tracks → My NCBI Track Collections. Следует обратить внимание на то, что все коллекции будут привязаны к сборкам и референсам, которые были выбраны при поиске нужного фрагмента генома. Например, если набор треков сохранен для версии геномной аннотации человека GRCh37, а в активном графическом интерфейсе выбрана версия геномной аннотации человека GRCh38, то сохраненный ранее набор треков в списке коллекций будет не виден до тех пор, пока не переключить версию геномной аннотации человека на GRCh37. Подробная информация о коллекциях треков расположена на сайте <https://www.ncbi.nlm.nih.gov/tools/sviewer/faq/>.

Далее необходимо выбрать потенциальные сайты гибридизации праймеров и геномной ДНК так, чтобы нежелательные элементы генома (описаны выше), в том числе делеции и инсерции, не попали в область потенциальных сайтов связывания. С целью визуального анализа таких сайтов видимый диапазон геномного браузера может быть сдвинут, как описано выше, изменен масштаб (для визуализации делеций и инсерций) пока не будет достигнут желаемый результат. В тех местах, где хотелось бы найти сайты связывания праймеров, содержание GC-пар должно быть от 50 до 65 % максимум. На участках ДНК, где содержание GC-пар приближается к 80–100 %, онлайн-ресурс Primer-BLAST подбирает праймеры только при условии, что указаны специальные требования.

На рисунке 9 показан типичный случай выбора потенциальных мест для связывания с праймерами. Видимый в геномном браузере потенциальный участок поиска праймеров не идеальный, присутствуют нежелательные элементы. В первую очередь исключаются все районы с повторами. Если есть зоны, избыточные инсерциями/делециями, то стоит избегать этих мест также, как повторов. Далее по возможности исключаются участки с высоким содержанием GC-пар. Иногда приходится оставлять какие-то не совсем подходящие варианты, если в целом результат выбора недостаточно хорош.

На рисунке 10 представлен окончательный результат подбора потенциальных сайтов для поиска праймеров. Выделены три области, к которым можно подобрать пары праймеров (1 и 2, 1 и 3). Первая потенциальная пара (области 1 и 2) опти-

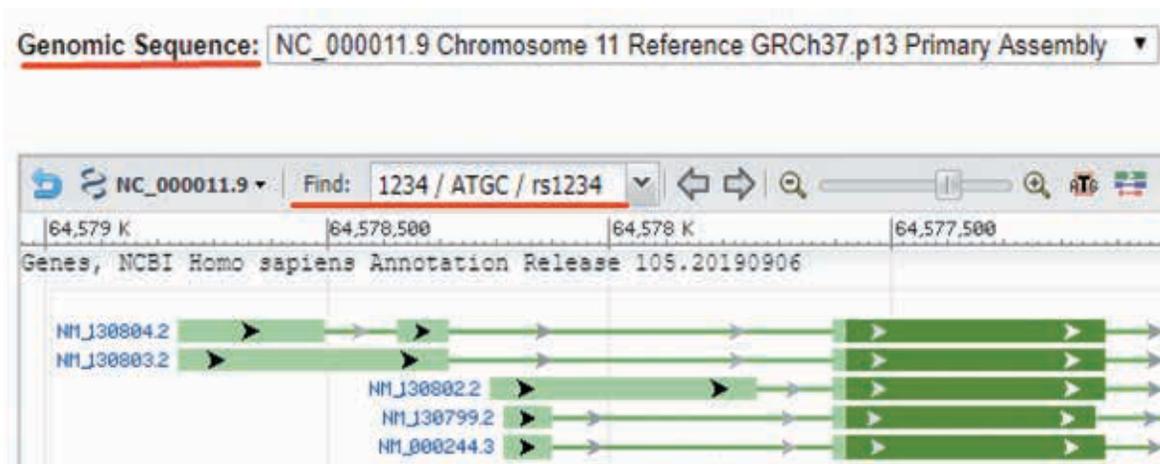


Рис. 1. Вид геномного браузера в режиме хромосомных аннотаций (сборки генома). Хромосомная аннотация имеет префикс NC_, транскрипты мРНК — префикс NM_. Темно-зеленым цветом отмечены кодирующие участки гена, светло-зеленым — некодирующие

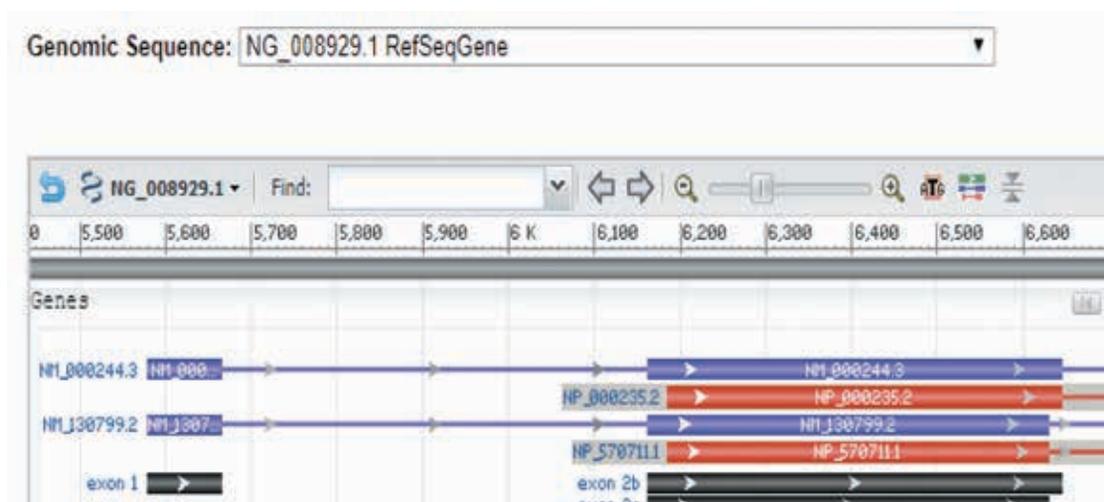


Рис. 2. Вид геномного браузера в режиме референсного образца гена. Референсный ген имеет префикс NG_, белки — NP_

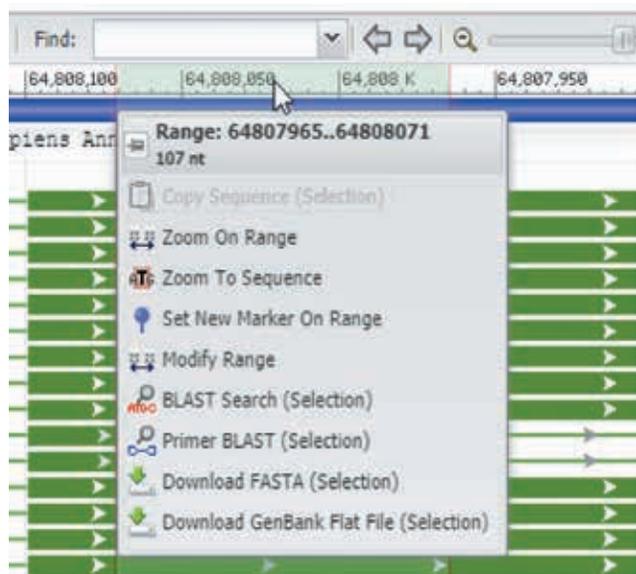


Рис. 3. Меню для выделенного участка ДНК

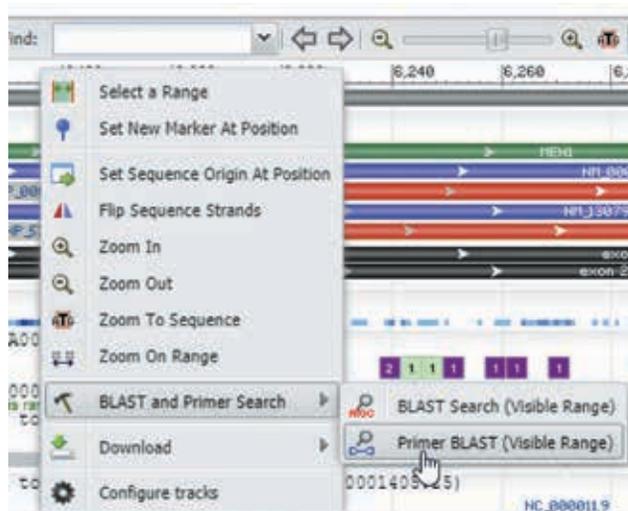


Рис. 4. Меню, открытое щелчком правой кнопки мыши. Показано подменю, связанное с программой BLAST

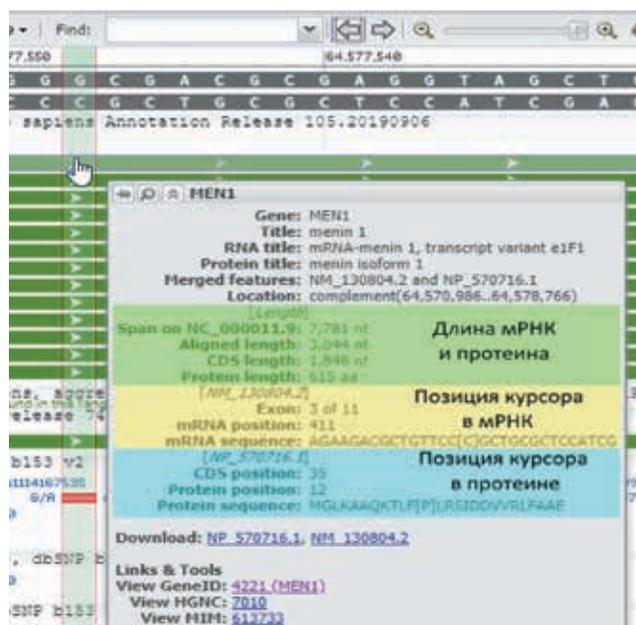


Рис. 5. Пример справочного окна с данными о просматриваемом элементе генома с комментариями авторов

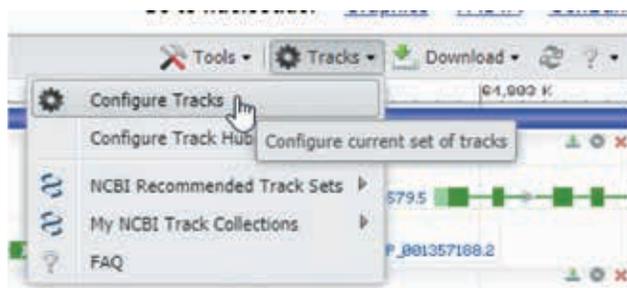


Рис. 6. Меню конфигурации отображаемых элементов геномного браузера.

Справа от меню Tracks есть меню загрузки — в нем можно скачать (целиком или избирательно) последовательность в форматах FASTA и GenBank, а также сохранить содержимое окна браузера в форматах PDF или SVG

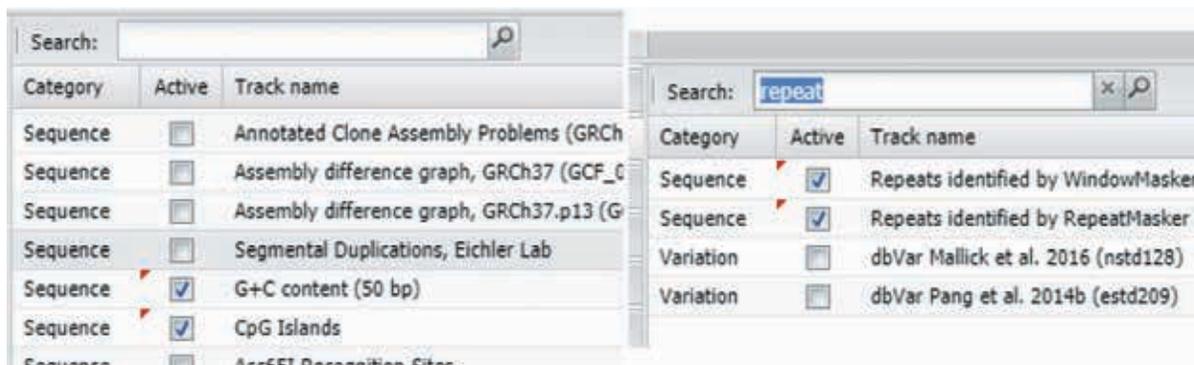


Рис. 7. Фрагменты всплывающего меню Configure Page

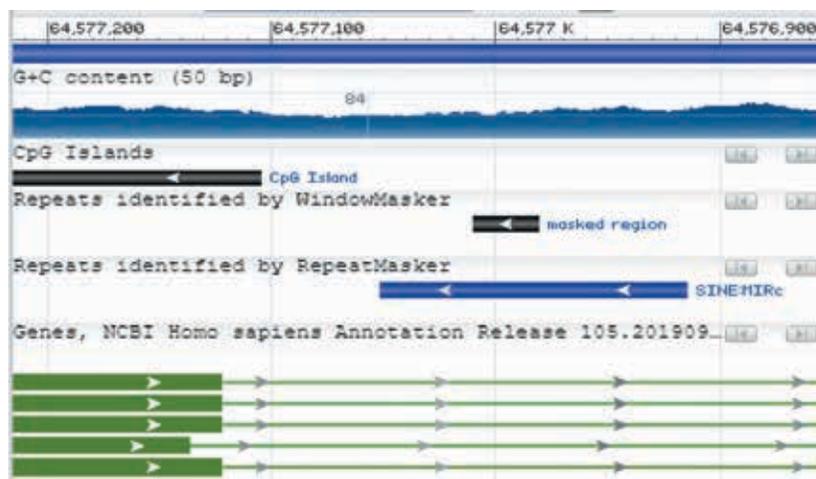


Рис. 8. Вид геномного браузера после активации описанных в тексте элементов (Tracks). Все элементы в геномном браузере подписаны, а при установке курсора над любым из них откроется информация

мальна по количеству GC-пар. Вторая (области 1 и 3) пара теоретически будет больше отличаться по температуре отжига между праймерами, так как область 3 более богата GC-основаниями по сравнению с 1 областью, соответственно у обратного (область 3) праймера температура отжига будет выше, чем у прямого (область 1) праймера.

Для дальнейшего поиска праймеров в видимом диапазоне геномного браузера используется онлайн-ресурс Primer-BLAST (Visible Range). В сложных случаях внутри подобранного фрагмента геномной ДНК можно выделить участки для поиска разных вариантов праймеров с использованием онлайн-ресурса Primer-BLAST (Selection). Например, для поиска праймеров в заданных конкретных участках ДНК (Рис. 10) выделяются диапазоны 1 и 2 или 1 и 3 (при нажатии Ctrl), и далее активируется запуск команды Primer-BLAST (Selection).

Поиск праймеров с использованием Primer-BLAST

Инструмент Primer-BLAST откроется в новом окне, при этом геномный браузер остается открытым. Выбор потенциальных участков матрицы ДНК для поиска праймеров проводился с использованием баз данных NCBI, конструирование праймеров будет проходить с использованием ресурса BLAST. В этом разделе будут рассмотрены только компоненты, которые нужны для дизайна праймеров. В программе BLAST для каждого компонента есть комментарии и объяснения, чтобы их получить, достаточно нажать круглый значок вопроса.

Вкладка Primers for target on the template

Раздел PCR Template

В поле Range будут автоматически проставлены хромосомные позиции — границы диапазона, которые были выбраны ранее для конструирования праймеров. Эти значения являются обязательными, изменять их не следует.

Раздел Primer Parameters (рис. 11).

В поле PCR product size указывается минимальное и максимальное значение ПЦР-продукта.

- Максимальное значение продукта ПЦР ограничено видимым полем геномного браузера, либо крайними положениями выделенных диапазонов в геномном браузере (окно осталось открытым), при необходимости это значение можно уменьшать вручную.

- Обратите внимание, что при подборе праймеров для двух выделенных диапазонов значение минимума не проставляется, так как размер ПЦР-продукта уже принудительно ограничен этими районами.

- При подборе праймеров по всему видимому диапазону геномного браузера можно (и нужно) менять параметр минимального размера ПЦР-продукта, так как «по умолчанию» он равен 70 п.о. Если нужно подбирать праймеры к протяженному участку, то размер ПЦР продукта должен быть длиннее этого участка. Например, для секвенирования по Сэнгеру (в силу технических особенностей) к размеру цели прибавляется 70–100 п.о. с обеих сторон, таким образом ПЦР-продукт будет длиннее цели на 140–200 пар нуклеотидов. Это значение и нужно проставить в качестве минимального размера ПЦР-продукта. BLAST будет подбирать праймеры с длиной в пределах от заданного минимума до максимума, пользователю останется визуально проконтролировать и выбрать варианты, имеющие достаточный запас расстояния от праймеров до цели.

В поле #of primers to return указывается количество праймеров, выдаваемых программой, «по умолчанию» это 10 пар.

В поле Primer melting temperatures (T_m) указываются значения минимума (Min), максимума (Max) и максимальная разница в температурах плавления обоих праймеров (Max T_m difference). «По умолчанию» значение Max T_m difference равно 3 градусам, при дизайне праймеров желательнее уменьшить это значение до 1–1,5 градусов. Если в выбранных для праймеров участках генома содержание GC-пар значительно выше 50 %, то максимальную температуру плавления праймеров (Max) иногда рекомендуется увеличивать на 1–2 градуса.

Раздел Exon/intron selection предназначен для работы с матричной РНК, для подбора праймеров не имеет значения.

Раздел Primer Pair Specificity Checking Parameters

В поле Database указывается база данных для поиска праймеров (рис. 12). Если задать Genomes for selected organisms (primary reference assembly only) поиск праймеров будет идти только по сборкам генома человека (GRCh37.p13 или GRCh38.p13). Эти сборки содержат только геномную информацию и не включают альтернативные локусы, митохондриальную ДНК. Остальные базы данных либо не подходят по основному компоненту, либо содержат избыточные записи, которые будут мешать процессу поиска конкретных праймеров.

В поле Organism «по умолчанию» стоит Homo sapiens.

Остальные поля данного раздела отражают нуклеотидный состав 3'-конца праймера и степень

его специфичности, параметры которых, как правило, принимаются «по умолчанию».

Поле Get Primers запускает процесс поиска праймеров. Чтобы результаты поиска открылись в новом окне, необходимо поставить «галочку» в поле Show results in a new window (рис. 13), это позволит при необходимости вернуться к заданным для поиска параметрам праймеров и изменить их, после чего запустить повторный поиск.

Вкладка Advanced parameters (рис. 14)

В данном разделе можно отметить дополнительные параметры для повышения требований к будущим праймерам, особенно на «сложных» участках ДНК-матрицы.

Раздел Primer Pair Specificity Checking Parameters (название повторяется)

Все параметры данного раздела принимаются «по умолчанию», так как изменения могут негативно отразиться на специфичности праймеров или на успешности поиска.

Раздел Primer Parameters

Сначала рассмотрим параметры, которые имеют первостепенное значение для подбора оптимальной пары праймеров.

В поле Primer GC content (%) «по умолчанию» стоит большой разброс значений от 20 до 80 %. Обычно Primer-BLAST конструирует праймеры для участков ДНК с 50 %-м содержанием GC-пар, но иногда по тем или иным причинам дает результаты, заметно отличающиеся от 50 %-го уровня (в пределах значений, стоящих «по умолчанию»). Рекомендуется сразу выставлять значение 50 % в поле Max, чтобы повысить «строгость» выбора праймеров по содержанию GC-пар. Если после этого итоговый результат окажется неудовлетворительным, значения в этих полях (Min и Max) можно подвергнуть коррекции, в пределах от 40 до 65 % соответственно.

В поле Max Self Complementarity (For old secondary structure alignment model only) и в поле Max Pair Complementarity (For old secondary structure alignment model only) «по умолчанию» стоят значения комплементарности 8 и 3 для любого участка праймера и для 3'-концов соответственно. Чем ниже значения комплементарности, тем меньше вероятность того, что праймеры будут образовывать димеры. В связи с этим данные значения рекомендуется по возможности снизить до 6 для поля Any и 2 для поля «3'-конец».

Поле SNP handling является обязательным дополнительным параметром, который позволяет

избегать известные районы, содержащие полиморфизмы с высокой частотой встречаемости.

Остальные параметры раздела Primer Parameters используются тогда, когда получить приемлемый результат выбора праймеров с ранее используемыми параметрами не удается.

В поле Primer Size указывается длина праймеров (Min, Opt, Max). Если приходится подбирать праймеры на участках ДНК с однородным составом, например содержащих тандемные или мононуклеотидные повторы, то иногда приходится увеличить длину праймеров для повышения их уникальности (чем длиннее последовательность нуклеотидов, тем реже вероятность ее повтора в геноме). Также данное поле используется для подбора праймеров с заданной длиной.

В поле Max Poly-X указывается максимально возможное количество мононуклеотидных повторов в праймере («по умолчанию» стоит 5). Если в результате поиска праймеров предлагаются пары праймеров, все из которых имеют 5 мононуклеотидов в составе, это значение можно снизить, чтобы попытаться найти альтернативу предложенным вариантам пар праймеров.

В поле Max GC in primer 3' end указывается возможное количество GC-нуклеотидов подряд на 3'-конце («по умолчанию» стоит 5). Снижение этого значения увеличивает вероятность обогащения 3'-конца AT-парами.

В полях Repeat filter и Low complexity filter указаны параметры фильтрации повторяющихся последовательностей генома. Исключают эти параметры в последнюю очередь, желательно в сочетании с использованием поля Primer Size, когда необходимо подобрать праймеры на участках ДНК, содержащих тандемные или мононуклеотидные повторы. В этом случае в поле Repeat filter выставляется значение None, в поле Low complexity filter снимается «галочка».

Обратите внимание, что поле Low complexity filter также используется для проверки качества праймеров (см. ниже по тексту).

Значения остальных полей, в том числе весь раздел Internal hybridization oligo parameters, вкладки Advanced parameters используются без изменений (выставлены «по умолчанию»).

Поле Get Primers (повторяется после вкладки Advanced parameters) запускает процесс поиска праймеров с учетом дополнительных параметров. Чтобы результаты поиска открылись в новом окне, необходимо также, как и ранее, поставить «галочку» в поле Show results in a new window (рис. 13). Если праймеры подбираются по сборке генома, отличающейся от актуальной (на момент написания —

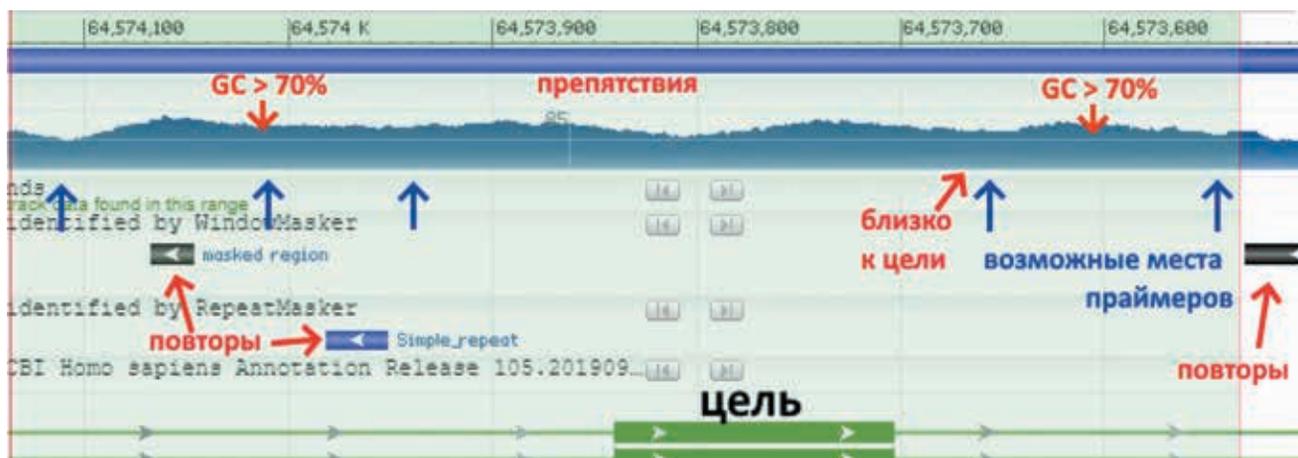


Рис. 9. Типовой случай визуального анализа в окне геномного браузера потенциальных сайтов связывания праймеров с матрицей ДНК

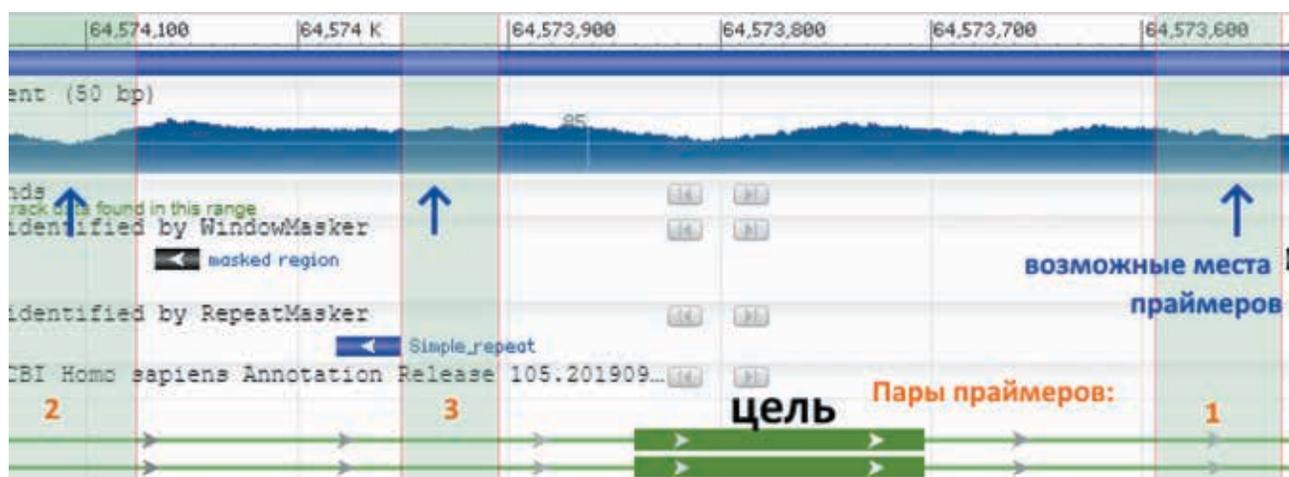


Рис. 10. Итоговый результат подбора потенциальных участков для поиска праймеров

Рис. 11. Пункты раздела Primer Parameters. Желтым цветом отмечены измененные параметры

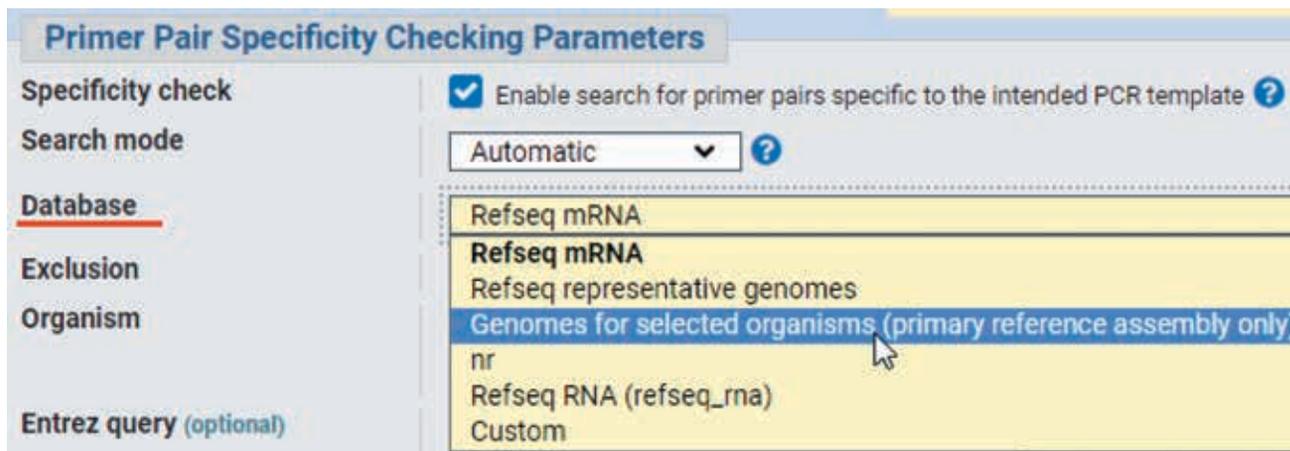


Рис. 12. Выбор базы данных для поиска праймеров



Рис. 13. Поле старта поиска праймеров

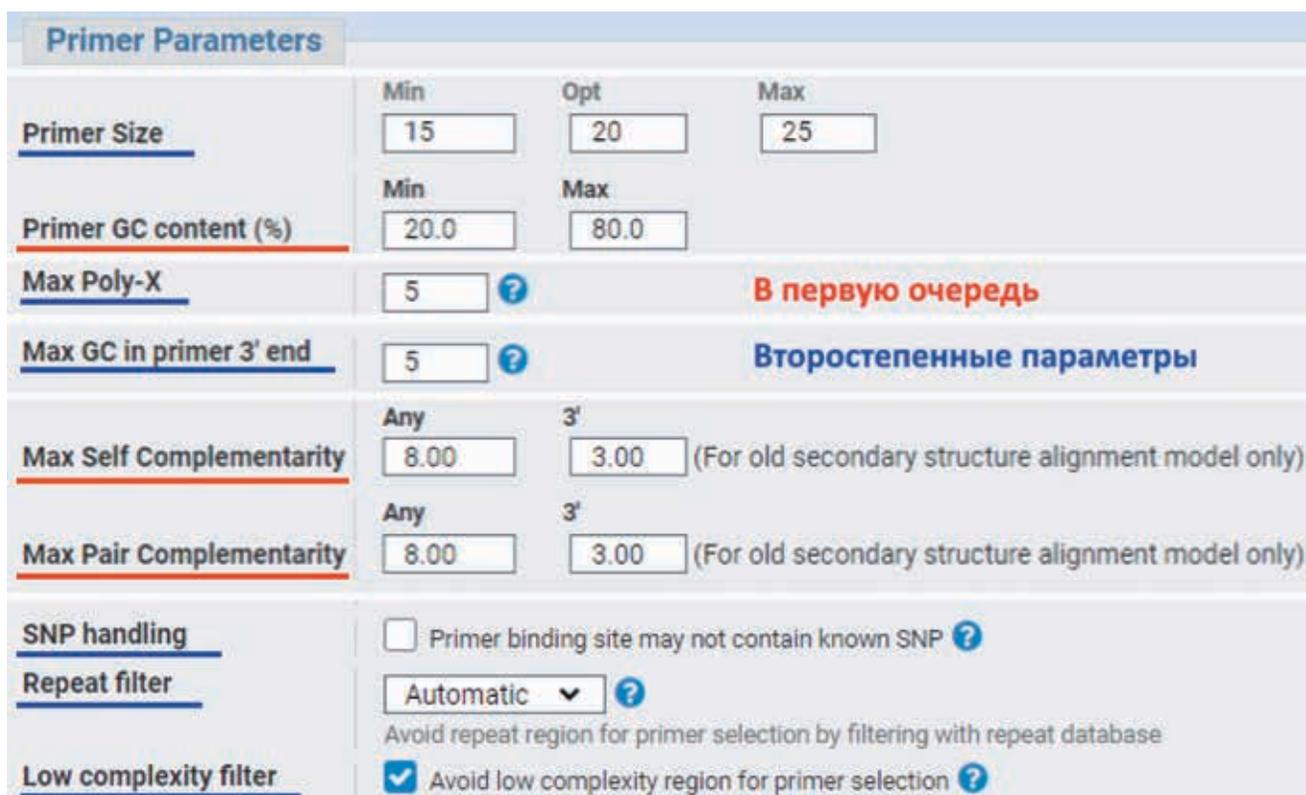


Рис. 14. Параметры раздела Primer Parameters во вкладке Advanced parameters. Показаны только те пункты, которые обычно используются для дизайна праймеров

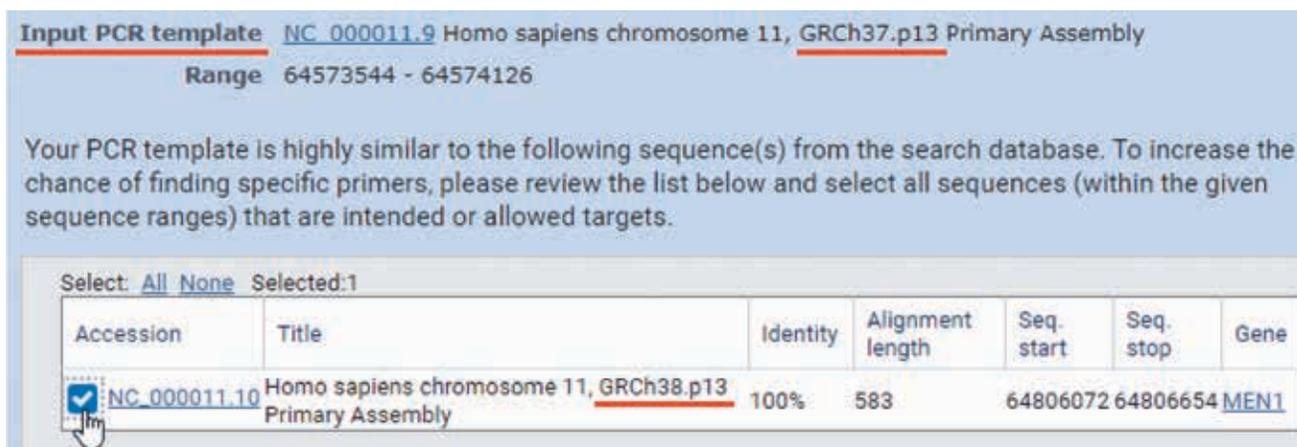


Рис. 15. Дополнительное окно после старта Get Primers, для вовлечения актуальной сборки генома человека (на момент написания — GRCh38.p13) в поиск праймеров

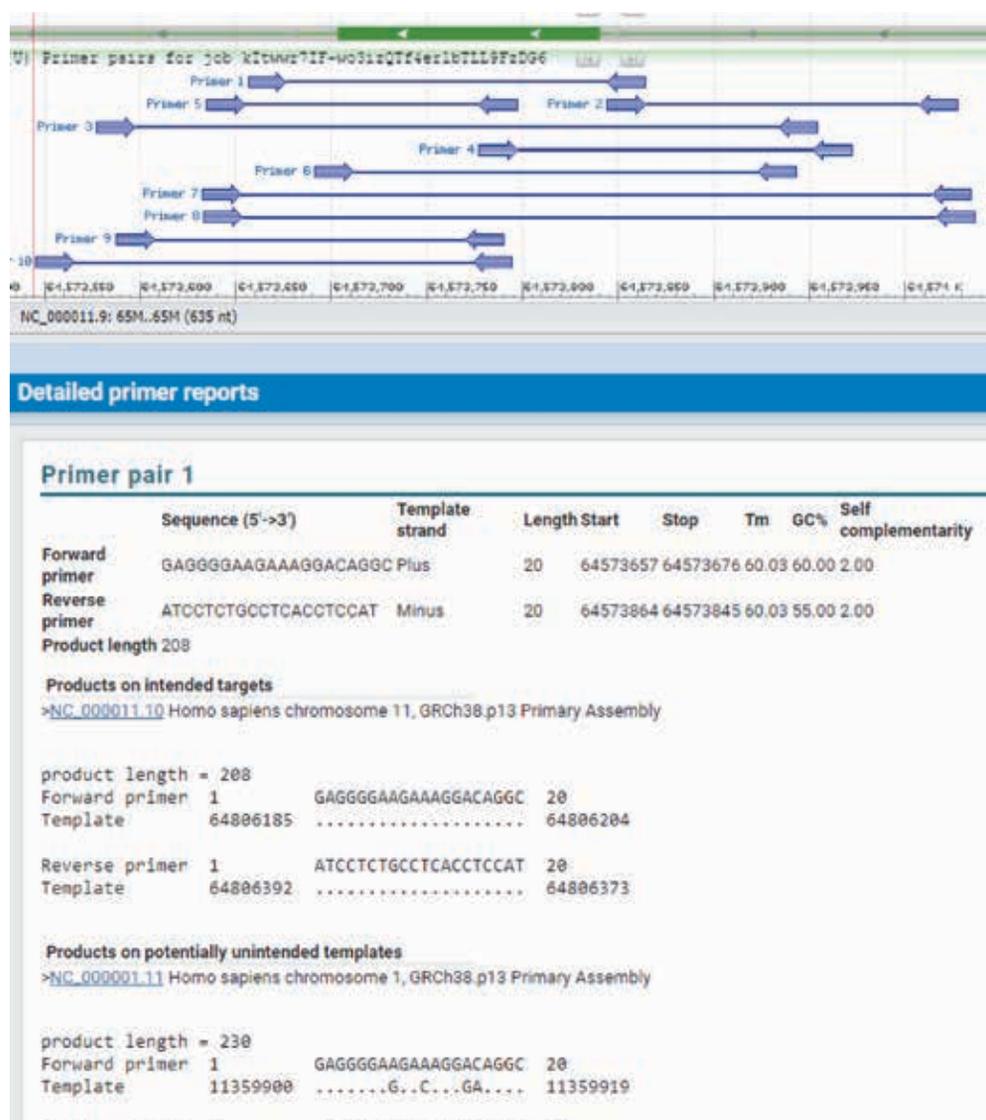


Рис. 16. Результат поиска праймеров, подобранных по видимому диапазону геномного браузера. Относительно расположения цели подходят только пары 1, 3, 6, 7 и 8. Остальные пары праймеров не захватывают цель полностью. В списке показана только первая пара праймеров; присутствует неспецифический продукт длиной 230 п.о.

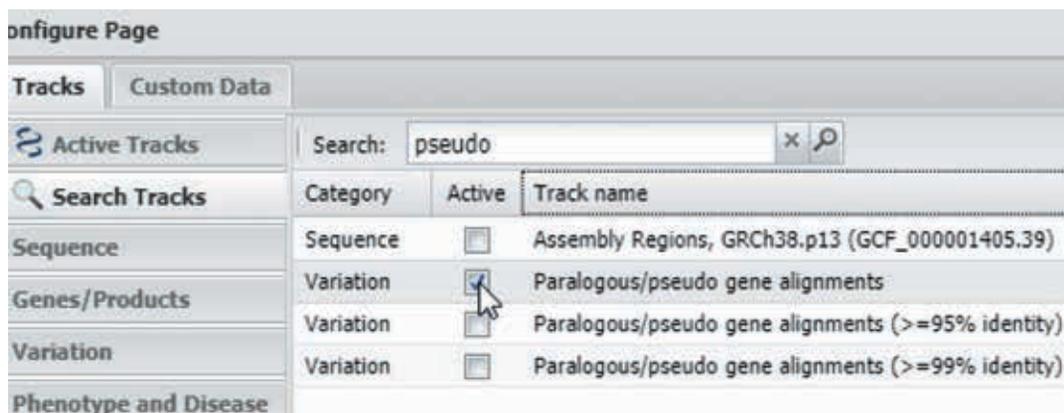


Рис. 17. Окно поиска псевдогенов. Отмечен трек, который отображает все псевдогены независимо от процента сходства с целевым участком ДНК

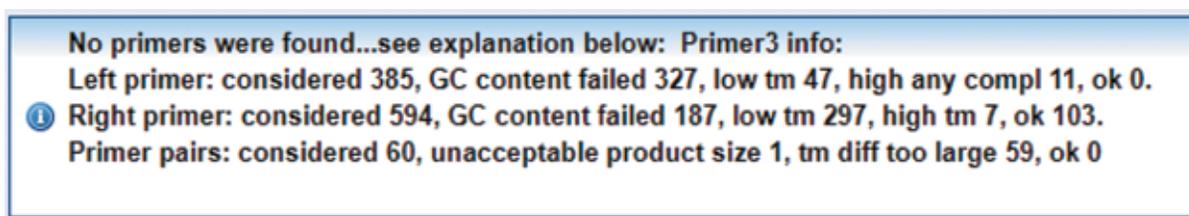


Рис. 18. Пример 1. Отрицательный результат поиска праймеров. Расшифровка сообщения:
 левый праймер: найдено 385 праймеров, содержание GC-пар не соблюдено у 327, слишком низкая Tm у 47, высокая степень димеризации (параметр Any) 11, хороших праймеров нет;
 правый праймер: найдено 594 праймера, содержание GC-пар не соблюдено у 187, слишком низкая Tm у 297, слишком высокая Tm у 7, хороших праймеров 103;
 пары праймеров: найдено 60 пар, недопустимый размер продукта у одной пары, слишком большая разница в температурах плавления праймеров у 59, приемлемых вариантов нет

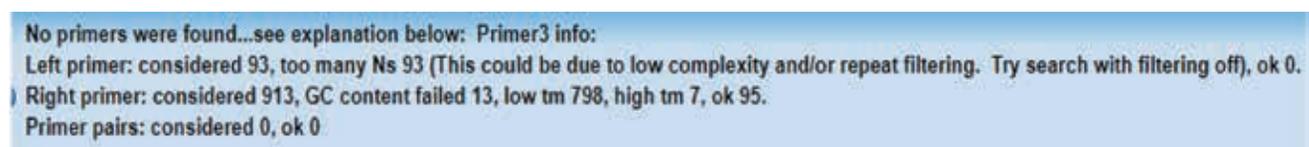


Рис. 19. Пример 2. Отрицательный результат поиска праймеров на участке ДНК, содержащем многочисленные тандемные или мононуклеотидные повторы

GRCh38.p13), например GRCh37.p13, после запуска старта в новом окне появится запрос на дополнительное использование похожих последовательностей из актуальной сборки на момент поиска, чтобы увеличить шансы выбора наиболее специфичных праймеров. Для включения текущей версии сборки в подбор праймеров необходимо поставить «галочку» в соответствующем окне (рис. 15) и нажать Submit, после чего в онлайн-режиме будет запущен процесс подбора праймеров. Обратите внимание, что если изначально в геномном браузере (рис. 1) выбрана актуальная версия сборки генома, то этого запроса после старта Get Primers не будет.

Анализ результатов поиска праймеров

В первую очередь оценивается расположение пары праймеров относительно целевого участка ДНК (рис. 16) и количество ПЦР-продуктов, которые потенциально могут быть амплифицированы с использованием данной пары праймеров (в идеале должна быть амплификация только одного участка матрицы ДНК, содержащего цель). Затем внимание уделяется нуклеотидному составу праймеров (особенно на 3'-концах) и их комплементарности друг другу, проверяется содержание GC-пар. Окончательный выбор делается на основе анализа разницы между температурами плавления прямого

и обратного праймеров, с учетом средней температуры плавления (T_m).

Если изначально по всем разделам (см. выше) были выставлены необходимые требования, то из всех параметров останется проанализировать только нуклеотидную последовательность праймеров. Это единственный параметр, который онлайн-ресурс не оценивает автоматически. Так, для улучшения качества гибридизации праймеры на 3'-конце должны содержать GC-основания, но Primer-BLAST не всегда соблюдает этот параметр. Далее задача пользователя — решить, какая пара праймеров будет наиболее эффективно работать для его целей. Например, пара праймеров с АТ-нуклеотидами на 3'-конце, но амплифицирующая только один целевой участок ДНК (результат — один ПЦР-продукт), или пара праймеров с GC нуклеотидами на 3'-конце, но имеющая сродство к нескольким участкам матрицы ДНК, и, соответственно, будут получены несколько ПЦР-продуктов.

Завершающим этапом проверки праймеров является проверка специфичности гибридизации, так как даже абсолютно уникальные праймеры для целевой ДНК-матрицы могут иногда гибридизоваться с неспецифичными последовательностями. С этой целью в новом окне браузера открывается программа Primer-BLAST; если в ней сохранились ранее установленные параметры, то все настройки сбрасываются с помощью кнопки *Reset page*, которая расположена справа от заголовка раздела *PCR Template*. В разделе *Primer parameters* в поле *Use my own forward primer* и поле *Use my own reverse primer* (см. рис. 11) вписываются нуклеотидные последовательности прямого и обратного праймеров соответственно. Выбирается база данных *Genomes for selected organisms (primary reference assembly only)*. Обратите внимание, обязательно нужно снять флажок с фильтра *Low complexity* в расширенных параметрах, после чего снова запустить процесс поиска праймеров. Ранее во время дизайна праймеров в онлайн-ресурсе Primer-BLAST был задан ряд ограничений, а сейчас ограничения сняты и использование районов с повторами разрешено. Результат поиска праймеров в данном случае покажет все возможные ПЦР-продукты для анализируемой пары праймеров, включая неспецифичные участки, если все-таки есть вероятность гибридизации данной пары на различных участках. Если пары праймеров подобраны идеально — программа выдаст только один целевой ПЦР-продукт, что дополнительно позволит пользователю удостовериться в правильности выбора пары праймеров.

Таким же образом можно проверить любые другие праймеры, например, полученные из какой-либо статьи или от другого человека. Primer-BLAST покажет пользователю температуру плавления обоих праймеров (прямого и обратного), процентное содержание GC-пар, вероятность димеризации, размер ПЦР-продукта и ген, комплементарный последовательностям праймеров.

Возможные проблемы при подборе праймеров и пути их решения

В данном разделе рассматриваются некоторые проблемы, которые могут встретиться при подборе праймеров:

- Праймеры не уникальны — после завершающей проверки с использованием онлайн-ресурса Primer-BLAST определяются несколько разных мест для гибридизации праймеров. Целесообразно вернуться к заданным параметрам поиска и увеличить строгость параметров поиска (см. вкладку *Advanced parameters*).

- Если после неоднократных изменений параметров поиска так и не удается получить пару праймеров, которая будет специфично амплифицировать только целевой ПЦР-продукт, то рекомендуется выбирать пару праймеров, которая будет амплифицировать целевой ПЦР-продукт, отличающийся от неспецифичных ПЦР-продуктов не менее чем на 100–200 нуклеотидов, чтобы их можно было легко разделить в агарозном гель-электрофорезе. Дополнительным критерием выбора пары праймеров будет высокая температура отжига.

- Наличие псевдогенов, дублирующих исследуемый участок ДНК. После запуска *Get Primers*, в случае наличия псевдогенов, появится запрос на использование подобных последовательностей (аналогично рис. 15). В поле *Identity* будет приведен процент сходства, как правило, меньше 100 %, а в поле *Gene* — название псевдогена. В этом случае очень велика вероятность того, что ПЦР-продуктов будет два, причем одинаковой длины. Если нет возможности избежать псевдогена, сдвинув район поиска в любую сторону, то решением проблемы будет поиск праймеров для участка ДНК гораздо большего размера, чем планировалось ранее, чтобы перекрыть этот псевдоген. Такой длинноразмерный ПЦР-продукт, в свою очередь, можно использовать как матрицу для последующей ПЦР (*nested PCR*) с внутренней парой праймеров, которая будет амплифицировать нужный фрагмент ДНК-матрицы, что позволит в результате получить специфический ПЦР-продукт. Для поиска псевдогенов и определения их границ используют закладку *Search Tracks* (см. рис. 6 и 7) в меню конфигу-

рации треков геномного браузера, в поле Search указывают pseudo (рис. 17). Браузер отобразит псевдогены, соответствующие последовательностям исследуемого гена, даже если они находятся на другой хромосоме.

- Результат поиска праймеров отрицательный, онлайн-ресурс Primer-BLAST выдает сообщение со списком проблем, препятствующих поиску праймеров. Необходимо проанализировать полученную информацию и отредактировать заданные критерии поиска праймеров (рис. 18 и 19).

В этом случае можно сначала повысить допустимое содержание GC-пар (температура плавления связана с нуклеотидным составом, при повышении количества GC-пар температура повысится автоматически) и провести поиск праймеров заново. При сохранении информации о большой разнице в температурах плавления праймеров, придется расширить допуск максимума до значения «по умолчанию». Если проблема с димеризацией продолжает возникать, то остается ослабить это требование (но не более чем до значения «по умолчанию») или изменить участок потенциального поиска праймера, так как при прочих решенных проблемах, вопрос димеризации праймеров оказывается связан с нуклеотидным составом зоны поиска.

В этом примере кроме проблем, описанных выше, один из праймеров попал на участок ДНК со слишком низкой сложностью нуклеотидного состава. Для получения результата необходимо отключить фильтрацию повторов. При этом целесообразно расширить, сузить или сдвинуть район поиска праймеров.

Заключение

Онлайн-ресурс Primer-BLAST является одним из лучших инструментов для конструирования и поиска праймеров. Рекомендуется помимо основной пары праймеров сохранять несколько дополнительных вариантов до тех пор, пока подобранные праймеры не будут проверены на практике. Не следует забывать, что любые биоинформатические ресурсы работают на депонированных на данный момент нуклеотидных последовательностях, а с учетом непрерывного мутационного процесса не удивительно, что иногда все же случается так, что теоретически идеальные праймеры на практике оказываются не слишком хороши, а не слишком хорошие — отлично работают.

Конфликт интересов / Conflict of interest

Авторы заявили об отсутствии потенциально-го конфликта интересов. / The authors declare no conflict of interest.

Список литературы / References

1. Ye J, Coulouris G, Zaretskaya I, et al. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. 2012;13:134. DOI: 10.1186/1471-2105-13-134.
2. Kozyreva AA, Kalinina OV, Zlotina AM, et al. Dizain praimerov s ispol'zovaniem baz dannykh NCBI: uchebnoe posobie dlya aspirantov i magistrantov k prakticheskim zanyatiyam. SPb.: Izd-vo SPBGEHTU «ЛЕНТИ», 2021. 31 p. [Козырева А.А., Калинина О.В., Злотина А.М. и др. Дизайн праймеров с использованием баз данных NCBI: учебное пособие для аспирантов и магистрантов к практическим занятиям. СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2021. 31 с.].
3. NCBI. National Center for Biotechnology Information Search database. <https://www.ncbi.nlm.nih.gov/> (01 May 2021).
4. Primer-BLAST. <https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi> (01 May 2021).
5. Gene <https://www.ncbi.nlm.nih.gov/gene/?term=> (01 May 2021).

Информация об авторах:

Козырева Александра Анатольевна, к.б.н., старший научный сотрудник Института молекулярной биологии и генетики, ФГБУ «НМИЦ им. В. А. Алмазова» Минздрава России;

Злотина Анна Михайловна, к.б.н., научный сотрудник Института молекулярной биологии и генетики, ФГБУ «НМИЦ им. В. А. Алмазова» Минздрава России;

Головкин Алексей Сергеевич, д.м.н., ведущий научный сотрудник Института молекулярной биологии и генетики, ФГБУ «НМИЦ им. В. А. Алмазова» Минздрава России;

Калинина Ольга Викторовна, д.б.н., декан факультета биомедицинских наук Института медицинского образования, ФГБУ «НМИЦ им. В. А. Алмазова» Минздрава России;

Костарева Анна Александровна, д.м.н., директор Института молекулярной биологии и генетики, ФГБУ «НМИЦ им. В. А. Алмазова» Минздрава России.

Author information:

Kozyreva Alexandra A., PhD, Senior Research Fellow, Institute of Molecular Biology and Genetics, Almazov National Medical Research Centre;

Zlotina Anna M., PhD, Research Fellow, Institute of Molecular Biology and Genetics, Almazov National Medical Research Centre;

Golovkin Alexey S., Dr. Sci., MD, Leading Researcher, Institute of Molecular Biology and Genetics, Almazov National Medical Research Centre;

Kalinina Olga V., Dr. Sci., Dean of the Faculty of Biomedical Sciences, Institute of Medical Education, Almazov National Medical Research Centre;

Kostareva Anna A., Dr. Sci., MD, Director of the Institute of Molecular Biology and Genetics, Almazov National Medical Research Centre.