

(19)



**Евразийское  
патентное  
ведомство**

(21) **202392318** (13) **A1**

(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ**

(43) Дата публикации заявки  
**2023.10.31**

(51) Int. Cl. *C12Q 1/68* (2018.01)

(22) Дата подачи заявки  
**2020.05.11**

---

(54) **ПРАЙМЕРЫ И СПОСОБЫ АНАЛИЗА ДЛЯ СОЕДИНЕНИЯ ОБЛАСТЕЙ С  
ИСПОЛЬЗОВАНИЕМ ПОЛИМЕРАЗ**

---

(31) **62/846,149**

(32) **2019.05.10**

(33) **US**

(62) **202193081; 2020.05.11**

(71) Заявитель:  
**ТЕ ЧАЙНИЗ ЮНИВЕРСИТИ ОВ  
ГОНКОНГ (CN)**

(72) Изобретатель:

**Чань Квань Чэ, Гай Ванься, Ло Юйк-  
Мин Деннис (CN)**

(74) Представитель:

**Фелицына С.Б. (RU)**

---

(57) Конкретные прямой и обратный праймеры можно применять для связывания отдаленных областей одной и той же крупной молекулы ДНК с образованием молекулы ДНК меньшего размера. Обратный праймер R1 может содержать первый фрагмент, комплементарный концевой последовательности области А, и может содержать второй фрагмент, содержащий перекрывающуюся последовательность. Прямой праймер F2 может содержать первый фрагмент, комплементарный последовательности начала области В, причем указанный прямой праймер содержит комплементарную перекрывающуюся последовательность (например, такой же первый фрагмент или второй фрагмент), которая комплементарна указанной перекрывающейся последовательности. Первый фрагмент F2 может представлять собой весь праймер. Молекулы ДНК меньшего размера можно использовать для определения гаплотипов областей. Кроме того, описаны наборы, содержащие указанные конкретные прямые и обратные праймеры.

---

**A1**

**202392318**

**202392318**

**A1**

## ПРАЙМЕРЫ И СПОСОБЫ АНАЛИЗА ДЛЯ СОЕДИНЕНИЯ ОБЛАСТЕЙ С ИСПОЛЬЗОВАНИЕМ ПОЛИМЕРАЗ

### Уровень техники

У человека есть две копии каждой хромосомы, по одной от каждого из родителей. Представление о комбинации вариантов на одной и той же родительской хромосоме, т.е. гаплотипе, может быть важным с клинической точки зрения. В частности, информация о гаплотипе полезна для неинвазивного пренатального анализа на предмет моногенных заболеваний и расшифровки геномной наследственности плода (Hui et al. Clin Chem. 2017;63:513-524; Lo et al. Sci Transl Med. 2010;2:61ra91). Однако современные методы определения гаплотипов конкретного индивида могут быть дорогостоящими, характеризоваться низкой точностью и низкой пропускной способностью.

### Краткое описание изобретения

ПЦР-анализы с использованием полимераз с высокой пропускной способностью можно применять для соединения удаленных областей (например, разделенных последовательностями размером от 100 п.о. до 100 т.п.о.) одной и той же большой молекулы ДНК в молекулу ДНК меньшего размера, например, упрощая определения гаплотипа двух указанных областей. Полимеразы с высокой пропускной способностью могут добавлять дополнительный нуклеотид только к одному концу каждой цепи двуцепочечной ДНК (например, А на 3'-конце). Это может препятствовать образованию соединенных молекул ДНК меньшего размера (также называемых удлинненными ампликонами) или снижать их выход.

Для решения этих проблем можно использовать конкретные прямые и обратные праймеры. Например, обратный праймер R1 может содержать первый фрагмент, комплементарный концевой последовательности области А, и может содержать второй фрагмент, содержащий перекрывающуюся последовательность. Прямой праймер F2 может содержать первый фрагмент, комплементарный последовательности начала области В, причем указанный прямой праймер содержит комплементарную перекрывающуюся последовательность (например, такой же первый фрагмент или второй фрагмент), которая комплементарна указанной перекрывающейся последовательности. Первый фрагмент F2 может представлять собой весь праймер.

Лучшее понимание природы и преимуществ вариантов реализации данного изобретения можно получить, сверяясь со следующим подробным описанием и прилагаемыми чертежами.

## Термины

В настоящем документе термин «*фрагмент*» (например, фрагмент ДНК) может относиться к фрагменту полинуклеотидной или полипептидной последовательности, содержащему по меньшей мере 3 последовательных нуклеотида. Нуклеотидный фрагмент может сохранять биологическую активность и/или некоторые характеристики исходного полипептида. Нуклеотидный фрагмент может быть двуцепочечным или одноцепочечным, метилированным или неметилированным, интактным или содержать одноцепочечный разрыв, находиться в комплексе или не в комплексе с другими макромолекулами, например, липидными частицами, белками. Фрагмент может быть получен из определенного типа ткани, например, плода, опухоли, трансплантированного органа и т.д.

Термин «*анализ*» обычно относится к методике определения свойства нуклеиновой кислоты. Анализ (например, первый анализ или второй анализ) обычно относится к методике определения количества нуклеиновых кислот в образце, геномной принадлежности нуклеиновых кислот в образце, изменению количества копий нуклеиновых кислот в образце, статусу метилирования нуклеиновых кислот в образце, распределению размеров фрагментов нуклеиновых кислот в образце, мутационному статусу нуклеиновых кислот в образце или характеру фрагментации нуклеиновых кислот в образце. Любой анализ, известный специалисту в данной области техники, можно применять для обнаружения любых свойств нуклеиновых кислот, упомянутых в настоящем документе. Анализ может также относиться к методике объединения продуктов амплификации различных областей молекулы ДНК с образованием одной или более молекул ДНК. Свойства нуклеиновых кислот включают: последовательность, количество, геномную принадлежность, количество копий, состояние метилирования по одному или более нуклеотидным положениям, размер нуклеиновой кислоты, мутацию нуклеиновой кислоты по одному или более нуклеотидным положениям и характер фрагментации нуклеиновой кислоты (например, положение нуклеотида(ов), по которым фрагментирована нуклеиновая кислота). Термин «анализ» можно использовать взаимозаменяемо с термином «способ». Анализ или способ могут характеризоваться конкретной чувствительностью и/или специфичностью, и их относительную полезность в качестве диагностического инструмента можно измерить с использованием статистического показателя ROC-ППК.

Термин «*прочтение последовательности*» относится к цепочке нуклеотидов, секвенированной из любого фрагмента или всей молекулы нуклеиновой кислоты. Например, прочтение последовательности может представлять собой весь фрагмент нуклеиновой кислоты, присутствующий в биологическом образце. Например, прочтение

последовательности может представлять собой короткую цепочку нуклеотидов (например, 20-150 оснований), секвенированную из фрагмента нуклеиновой кислоты, короткую цепочку нуклеотидов на одном или обоих концах фрагмента нуклеиновой кислоты, или результат секвенирования всего фрагмента нуклеиновой кислоты, присутствующего в биологическом образце. Прочтения спаренной последовательности можно выровнять на эталонный геном, что может позволить получить длину фрагмента. Прочтение последовательности можно получить различными способами, например, с использованием методик секвенирования или с использованием зондов, например, в матрицах гибридизации или зондах захвата, или методик амплификации, например, полимеразной цепной реакции (ПЦР), или линейной амплификации с использованием единственного праймера, или изотермической амплификации, или на основе биофизических измерений, например, масс-спектрометрии. Прочтение последовательности можно получить путем секвенирования единичной молекулы. «Секвенирование единичной молекулы» относится к секвенированию единственной молекулы матричной ДНК для получения прочтения последовательности без необходимости интерпретировать информацию о последовательности оснований из клональных копий молекулы матричной ДНК. Секвенирование единичной молекулы может секвенировать всю молекулу или лишь фрагмент молекулы ДНК. Можно секвенировать большую часть молекулы ДНК, например, более 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% или 99%.

«Степень разделения» (или относительная встречаемость) соответствует разности или отношению, включающему два значения, например, два количества прочтений, содержащих два различных аллеля. Степень разделения может представлять собой простую разность или отношение. Например, прямое отношение  $x/y$ , а также  $x/(x + y)$  представляет собой степень разделения. Степень разделения может включать другие коэффициенты, например, множители. В качестве других примеров можно использовать разность или отношение функций значений, например, разность или отношение натуральных логарифмов ( $\ln$ ) двух значений. Степень разделения может включать разность и/или отношение.

В настоящем документе термин «классификация» относится к любому числу(числам) или другому символу(символам), которые связаны с конкретным свойством образца. Например, символ «+» (или слово «положительный») может означать, что образец классифицируется как содержащий делеции или амплификации. Классификация может быть двоичной (например, положительной или отрицательной) или характеризоваться несколькими уровнями классификации (например, шкала от 1 до 10,

или от 0 до 1).

Термины «*пороговое значение*» и «*порог*» относятся к заранее определенным числам, используемым в процессе. Например, размер порогового значения может относиться к размеру, при превышении которого фрагменты исключают. Пороговое значение может являться значением, выше или ниже которого применяют конкретную классификацию, например, классификацию состояния, например, наличие или отсутствие у субъекта состояния, или тяжесть состояния. Пороговое значение или порог может представлять собой «эталонное значение» или производное от эталонного значения, которое входит в конкретную классификацию или проводит границу между двумя или более классификациями, например, позволяет различать аллели, составляющие гаплотип. Такое эталонное значение можно определить различными способами, например, выбрать после получения и на основании получения выходного значения данных анализа, как должен принимать во внимание специалист в данной области техники. Например, можно определить параметры для двух разных групп субъектов с разными известными классификациями, а эталонное значение можно выбрать как типичное (например, среднее) значение для одной классификации или как значение, находящееся между двумя кластерами параметров. Соответственно, эталонных субъектов с известной классификацией гаплотипов можно использовать для определения эталонных уровней для различения различных гаплотипов. В качестве еще одного примера, эталонное значение можно определить на основе статистического моделирования образцов. Любой из указанных терминов можно использовать в любом указанном контексте. Специалист в данной области техники должен принимать во внимание, что можно выбрать порог для достижения желательной чувствительности и специфичности.

Краткое описание чертежей

На фиг. 1A-1D показана система ПЦР с четырьмя праймерами, используемая для амплификации двух различных областей-мишеней А и В.

На фиг. 2A-2C показана система ПЦР с тремя праймерами, используемая для амплификации двух различных областей-мишеней А и В.

На фиг. 3A-3F проиллюстрирован способ ПЦР с перекрывающимися праймерами с использованием соответствующих геномных положений с Т и А в областях А и В согласно вариантам реализации настоящего изобретения.

На фиг. 4A-4C проиллюстрирован способ ПЦР с перекрывающимися праймерами с использованием обратного праймера с добавленным А согласно вариантам реализации настоящего изобретения.

На фиг. 5A-5D проиллюстрирован способ ПЦР с перекрывающимися праймерами с

использованием обратного и прямого праймеров с добавленным А согласно вариантам реализации настоящего изобретения.

На фиг. 6 проиллюстрирован принцип способа ПЦР с перекрывающимися праймерами и лигированием адаптеров, использующего праймеры для стимуляции амплификации гибридных продуктов ПЦР согласно вариантам реализации настоящего изобретения.

На фиг. 7А-7В показан пример применения праймеров для стимуляции амплификации гибридных продуктов ПЦР согласно вариантам реализации настоящего изобретения.

На фиг. 8А-8В проиллюстрирован способ ПЦР с перекрывающимися праймерами для связывания более двух областей, обеспечивающий связывание друг с другом лишь определенных областей, согласно вариантам реализации настоящего изобретения.

На фиг. 9А-9В проиллюстрирован способ ПЦР с перекрывающимися праймерами для связывания более двух областей, обеспечивающий возможность связывания любых областей друг с другом, согласно вариантам реализации настоящего изобретения.

На фиг. 10А-10В показано, как можно применять праймеры для амплификации гибридных продуктов ПЦР с множественными областями в соответствии с вариантами реализации настоящего изобретения.

На фиг. 11 показаны шесть гаплотипов, содержащих различные комбинации генотипов в экзонах 1 и 3 *NUDT15*.

На фиг. 12А и 12В показаны кривые плавления, позволяющие различать образцы с различными генотипами *NUDT15* в областях экзона 1 (А) и экзона 3 (В).

На фиг. 13А-13F показано применение способа ПЦР с перекрывающимися праймерами к экзонам 1 и 3 гена *NUDT15* согласно вариантам реализации настоящего изобретения.

На фиг. 14А-14С показаны четыре области, подлежащие гаплотипированию, разделенные на несколько групп по две области в соответствии с вариантами реализации настоящего изобретения.

На фиг. 15А-15С показано образование гибридных продуктов из трех областей, по одной области из каждой из трех групп, согласно вариантам реализации настоящего изобретения.

На фиг. 16А-16В показано образование гибридных продуктов без добавления аденозина в праймеры согласно вариантам реализации настоящего изобретения.

На фиг. 17А-17В показано (А) расстояние между различными ОНП в молекуле ДНК и (В) конфигурации праймеров для ОНП согласно вариантам реализации настоящего

изобретения.

Фиг. 18 представляет собой таблицу, в которой показаны последовательности праймеров для амплификации каждой области согласно вариантам реализации настоящего изобретения.

Фиг. 19 представляет собой таблицу, в которой показаны фазы, определенные посредством цифровой ПЦР с перекрывающимися праймерами согласно вариантам реализации настоящего изобретения.

Фиг. 20 представляет собой таблицу, в которой показаны последовательности праймеров для ОНП в *гене HBV* согласно вариантам реализации настоящего изобретения.

На фиг. 21А-21В проиллюстрирован анализ фазовой сборки согласно вариантам реализации настоящего изобретения.

Фиг. 22 представляет собой таблицу, в которой показаны гаплотипы, определенные посредством ПЦР с перекрывающимися областями для семейных трио согласно вариантам реализации настоящего изобретения.

Фиг. 23 представляет собой блок-схему, на которой проиллюстрирован способ соединения разделенных областей молекул ДНК в биологическом образце субъекта согласно вариантам реализации настоящего изобретения.

На фиг. 24 проиллюстрирована система измерения согласно варианту реализации настоящего изобретения.

На фиг. 25 продемонстрирована блок-схема примера компьютерной системы, которую можно применять с системами и способами согласно вариантам реализации данного изобретения.

#### Подробное описание изобретения

Секвенирование множественных отдаленных областей молекулы ДНК может быть полезным. Интерес могут представлять лишь определенные области (например, однонуклеотидный полиморфизм [ОНП]) молекулы ДНК. Некоторые методики секвенирования могут быть ограничены по длине фрагмента, который можно секвенировать за один раз. Секвенирование промежуточных областей между областями, представляющими интерес, может напрасно занимать оборудование и время оператора. Обычные способы, связывающие две отдельные области для секвенирования друг с другом, сопряжены с рядом проблем. Контроль чистоты молекул ДНК, подлежащих амплификации, может являться сложной задачей. Такие обычные способы могут обладать низкой точностью. Методики, описанные в настоящем документе, позволяют использовать точные и эффективные методы соединения двух или более разделенных областей молекулы ДНК друг с другом. Затем полученные гибридные молекулы ДНК

можно секвенировать и/или гаплотипировать.

Некоторые варианты реализации могут обеспечить повышенный выход и соответствующую эффективность за счет использования полимераз с высокой пропускной способностью, характеризующихся склонностью к добавлению определенного нуклеотида (например, А) в положение в липком конце. В одном примере такая методика может включать конструирование обратного праймера, который гибридизуется в области с комплементарным нуклеотидом (например, Т) на конце, за счет чего возможно правильное связывание двух областей. Другая связываемая область может содержать указанный определенный нуклеотид после прямого праймера, за счет чего создается совпадающая перекрывающаяся область для связывания указанных областей. В еще одном примере сами праймеры могут включать перекрывающуюся область с указанным определенными нуклеотидом (и комплементарным нуклеотидом для некоторых праймеров в некоторых вариантах реализации), вставленную между двумя фрагментами. Соответственно, для связывания разделенных областей можно применять прямые праймеры с фрагментами, комплементарными фрагментам обратных праймеров.

В качестве других примеров с использованием или без использования таких высокопроизводительных полимераз, приводящих к смещению, в группы друг с другом можно связывать различные комбинации областей одной и той же молекулы. Эта комбинация областей могла бы обеспечить точное и дешевое гаплотипирование крупных геномных областей, охватывающие несколько исследуемых областей. Прямой и обратный праймеры можно применять для амплификации гибридного продукта. Друг с другом можно связать три или более областей. Одновременное объединение, фазирование и гаплотипирование большего числа областей могло бы быть выгодным, если одна или более из указанных областей являются гомозиготными, поскольку гомозиготные области не позволяют получать информацию о гаплотипе. Кроме того, можно представить наборы, включающие такие специально разработанные праймеры. Этих особенности подробно описаны ниже.

#### Методики определения гаплотипа

Доступен ряд способов для определения гаплотипов индивида. Эти способы в целом делятся на две категории: вычислительный вывод и экспериментальный анализ. В то же время существующие способы определения гаплотипов индивида имеют свои ограничения, включая относительно низкую точность, высокую стоимость, низкую пропускную способность и невозможность адресного анализа конкретной области. В настоящей заявке описаны способы, позволяющие точно определять гаплотипы индивида при относительно низких затратах и с высокой пропускной способностью.



### *А. Вычислительные подходы*

Вычислительные подходы часто используются для статистического анализа гаплотипов человека на основе информации о генотипах соответствующей популяции (Browning et al., *Nature Reviews Genetics*, 2011;12:703-714). Посредством анализа генотипа у большого количества индивидов в популяции можно определить общие гаплотипы популяции. Генотип тестируемого индивида определяют экспериментально и сравнивают с известными гаплотипами популяции, делая вывод о наиболее вероятных гаплотипах индивида.

Однако точность таких вычислительных способов зависит от этнической принадлежности и происхождения тестируемого индивида. В регионах с широким этническим составом способы вычислительного гаплотипирования могут не обеспечивать достаточную точность для клинических целей. Кроме того, точность этого подхода снижается при анализе гаплотипов на большом расстоянии и в областях с низким неравновесием по сцеплению. Например, в областях генома, кодирующих лейкоцитарный антиген человека (HLA), часто происходит рекомбинация. Следовательно, вычислительный анализ гаплотипа в этих областях с высокой рекомбинацией характеризуется низкой точностью.

### *В. Экспериментальный анализ*

В качестве альтернативы, гаплотипы индивида можно определять экспериментально. Принцип этих экспериментальных способов гаплотипирования основан на генотипировании различных областей одной длинной молекулы ДНК. Есть три общих подхода для достижения этой цели, включая (а) перекрестное связывание структурно проксимальных областей, (b) компарментализацию и (с) секвенирование с длинными прочтениями.

#### *Перекрестное связывание*

При перекрестном связывании используют тот факт, что хромосома укладывается в трехмерную (3-D) структуру, и различные части одной и той же хромосомы характеризуются значительно повышенной вероятностью сближения. Посредством перекрестного связывания ДНК, непосредственно сближенной в трехмерном пространстве, можно соединять участки ДНК, происходящие из одной и той же хромосомы, но разделенные несколькими тысячами пар оснований. Последовательности этих гибридных молекул можно использовать для фазирования аллелей однонуклеотидных полиморфизмов (ОНП). Примеры этого подхода включают методику Hi-C (Selvaraj et al. *Nat. Biotechnol* 2013;31:1111-8) и модифицированную методику Hi-C «Чикаго» (Putnam et al. *Genome Res.* 2016;26:342-50).

### Компартментализация

Подход, основанный на компартментализации, может использовать преимущества автоматизированных платформ для компартментализации. Принцип этого подхода заключается в разделении отдельных длинных молекул ДНК, полученных от тестируемого субъекта, в различных компартментах. Затем генотипируют ДНК из одного компартмента. Поскольку информация о генотипе получена из одной молекулы ДНК, а не из пары родительских хромосом, полученная информация об аллелях представляет гаплотип. Разделение отдельных длинных молекул ДНК в компартментах можно выполнять вручную (Peters et al. *Nature* 2012;478:190-5) или с использованием микрожидкостных систем (Zheng et al. *Nat. Biotechnol* 2016;34:303-11).

Для повышения рентабельности этого подхода длинную молекулу ДНК внутри одного компартмента фрагментируют на более мелкие фрагменты. Каждый короткий фрагмент, происходящий из одинаковых длинных молекул ДНК, кодируют одним и тем же индексом. Затем короткие фрагменты ДНК из разных компартментов можно объединить и секвенировать с использованием массового параллельного секвенирования. Фрагменты с одинаковым индексом можно использовать для построения гаплотипа (Amini et al. *Nat. Genet* 2014;46:1343-9). Примеры автоматизированных платформ, позволяющих индексировать фрагменты, включают систему Chromium производства 10x genomics и решения для фазированного секвенирования производства Illumina, но не ограничиваются ими.

Однако существующие способы гаплотипирования, основанные на этих подходах, в основном предназначены для гаплотипирования относительно больших областей генома. Использование этих способов для гаплотипирования относительно короткой области генома требует обогащения соответствующей области, например, путем гибридационного захвата, что делает его относительно дорогим и трудоемким.

### Секвенирование с длинными прочтениями

В еще одном подходе целую длинную молекулу ДНК можно непосредственно секвенировать для определения ее последовательности. Это можно выполнить с помощью платформ секвенирования нового поколения, например, с использованием технологии секвенирования единственной молекулы в реальном времени (от Pacific Biosystems) и технологии нанопорового секвенирования (от Oxford Nanopore Technologies), но не ограничиваясь ими. Однако эти системы также требуют предварительного обогащения рассматриваемых областей для секвенирования.

### Компартментализация с помощью ПЦР

Молекулы ДНК можно подвергать компартментализации в каплях эмульсии.

Каждая капля эмульсии может включать молекулы ДНК, содержащие две области, которые можно связать друг с другом посредством амплификации. Связывание двух областей друг с другом может повысить эффективность секвенирования и гаплотипирования. Для амплификации используют системы с четырьмя праймерами и тремя праймерами.

### *С. Система с четырьмя праймерами*

Для гаплотипирования конкретной области описан ряд способов на основе ПЦР. Например, Wetmur с соавт. разработали способ связывающей эмульсионной ПЦР (Wetmur et al. *Nucleic acid research* 2005;33:2615-9). В этом способе разбавленные матрицы ДНК распределяют в капли эмульсии, которые действуют как компартменты.

На фиг. 1А-1D показана система ПЦР с четырьмя праймерами, используемая для амплификации двух различных областей-мишеней А и В. На фиг. 1А показаны области А и В как две мишени, требующие фазирования. Прямой праймер F1 и обратный праймер R1 предназначены для амплификации области А. Прямой праймер F2 и обратный праймер R2 предназначены для амплификации области В. Точечная линия 105 представляет собой промежуточную область генома между областью А и областью В. Две области могут быть разделены несколькими тысячами пар оснований, эта линия изображена не в масштабе.

На фиг. 1В прямой праймер F2 и обратный праймер R1 частично комплементарны. В этом примере между F2 и R1 находятся 32 комплементарных нуклеотида, причем 10 нуклеотидов для каждого праймера специфичны по отношению к соответствующей области. Комплементарная область обозначена как перекрывающаяся область (OLR) 110. Молекулу биотина 115 присоединили к 5'-концу F2 и R1, например для обеспечения возможности удаления несвязанных цепей, как указано ниже.

На фиг. 1С 3'-концы продуктов ПЦР-амплификации области А и области В содержат OLR 110, так что они комплементарны друг другу. Следовательно, 3'-конец продуктов амплификации области А можно использовать в качестве праймера для продуктов амплификации области В и наоборот. Это приведет к соединению двух амплифицированных областей. Затем соединенные цепи ДНК можно удлинить вправо в верхней цепи и влево в нижней цепи, тем самым создавая относительно короткую молекулу, включающую отдаленные области А и В.

Соответственно, как показано на фиг. 1В, между обратным праймером для ампликона А и прямым праймером для ампликона В находятся 32 перекрывающихся нуклеотида. Это перекрывание позволяет связать два ампликона. При низкой концентрации матрицы ДНК и наличии лишь одной матрицы ДНК в каждой капле эмульсии связанные продукты ПЦР представляют собой гаплотип индивида.

Однако в этой системе связывание двух ампликонов является неполным (т.е. не все они связаны), и присутствуют несвязанные молекулы каждой области ДНК. Например, одноцепочечные молекулы ДНК, полученные в результате удлинения R1 или F2, охватывают только область А или область В, соответственно. Эти несвязанные молекулы ДНК могут влиять на последующий анализ гаплотипирования. Следовательно, эти несвязанные молекулы ДНК необходимо удалить путем связывания биотинированных праймеров для ПЦР и блокирования неудлиненной одноцепочечной ДНК.

На фиг. 1D несвязанные продукты ПЦР содержат связанные с биотином праймеры R1 и F2. Их можно удалить обработкой стрептавидиновыми гранулами.

#### *D. Система с тремя праймерами*

В еще одном способе, описанном Tyson с соавт., объединили систему ПЦР-амплификации с тремя праймерами путем разделения в каплях эмульсии для гаплотипирования структурно сложной области (Tyson et al. BMC Genomics 2012;13:693). Аналогично способу, описанному Wetmur с соавт., систему ПЦР настроили для использования в эмульсиях с очень низкой концентрацией ДНК-мишени, так что каждая капля эмульсии содержала одну длинную молекулу ДНК-мишени. В отличие от способа, описанного Wetmur с соавт., в способе, разработанном Tyson с соавт., использовали всего 3 праймера.

На фиг. 2A-2C показана система ПЦР с тремя праймерами, используемая для амплификации двух различных областей-мишеней А и В.

На фиг. 2A прямой праймер F3 и обратный праймер R3 используются для амплификации области А.

На фиг. 2B обратный праймер R3 состоит из двух частей. 3'-конец 205 содержит нуклеотиды, комплементарные последовательностям области А, так что он может служить праймером для амплификации области А. 5'-конец 210 R3 комплементарен В'-области области В.

На фиг. 2C после ПЦР-амплификации области А с использованием F3 и R3 продукты амплификации содержат 3'-конец 215, комплементарный В'-области. Этот амплифицированный продукт может служить праймером для амплификации области В вместе с обратным праймером R4. В результате области А и В можно связать друг с другом. Каждый связанный продукт ПЦР состоит из области А и области В одной молекулы ДНК и представляет собой гаплотип.

На основе способа, описанном Tyson с соавт., в качестве второго раунда амплификации для отбора связанных молекул, несущих конкретный аллель в области А, использовали аллель-специфическую ПЦР. Отбор связанных молекул можно выполнить

путем амплификации с использованием прямого праймера, специфичного по отношению к определенному аллелю в области А, и обратного праймера в конце области В. Таким образом, амплифицируется только один из гаплотипов, причем амплифицированный гаплотип соответствует конкретному аллелю в области А. Отобранные молекулы секвенировали с использованием секвенирования по Сэнгеру для определения последовательности в области В, чтобы обеспечить возможность фазирования аллелей в области А и области В.

#### *Е. Проблемы*

В этих способах на основе ПЦР используют ручное получение капельной эмульсии для распределения длинных молекул ДНК-матрицы в очень разбавленной концентрации по отдельным каплям. Эмульсию обычно образуют путем встряхивания смеси масла и водных реагентов, используемой для реакции ПЦР. Однако распределение молекул ДНК по отдельным каплям с использованием этого ручного способа сопряжено с рядом проблем.

Во-первых, нельзя точно контролировать объем капель. В результате существует относительно высокая вероятность того, что две или более молекулы ДНК будут размещены в одной крупной капле. Такая ситуация приведет к неправильному определению фазы аллелей в разных локусах. Более того, эффективность реакции ПЦР в различных каплях будет неодинакова. В этих предыдущих способах с использованием анализа на основе эмульсионной ПЦР связанные продукты затем амплифицировали с использованием аллель-специфической ПЦР для амплификации молекул ДНК, несущих один из двух аллелей в первом локусе (например, в области А). Затем выполняли секвенирование по Сэнгеру для определения аллеля во втором локусе (например, в области В), связанном с конкретным аллелем в первом локусе.

#### Способ на основе пцр с перекрывающимися праймерами

В настоящей заявке на патент авторы описывают надежные способы гаплотипирования, которые могут использовать микрожидкостные системы для связывания нескольких мишеней на одной молекуле ДНК, причем связанные молекулы секвенируют путем массового параллельного секвенирования. Эти способы могут быть особенно полезны для гаплотипирования областей, включающих варианты, которые нельзя генотипировать простыми методиками дискриминации аллелей, например, аллель-специфической ПЦР или с помощью аллель-специфических флуоресцентных зондов. В некоторых вариантах реализации можно использовать микрожидкостную цифровую ПЦР-систему для создания мелкодисперсных капель, например, большинство из которых содержат не более одной длинной молекулы ДНК, охватывающей рассматриваемую

область. ПЦР-амплификацию можно выполнять на двух или более рассматриваемых областях одной длинной молекулы ДНК с использованием высокоэффективной полимеразы, например, полимеразы Taq и двух наборов праймеров, но не ограничиваясь ей. Использование таких полимераз может значительно увеличить производительность за счет сокращения времени получения продуктов амплификации.

Существуют доступные для приобретения устройства, которые могут автоматически генерировать тысячи капель размером порядка нанолитра за короткое время, например, BioRad QX200 Droplet Generator, Elveflow Droplet Generator Pack и Micronit Microfluidic Droplet Generator, но не ограничиваясь ими. Кроме того, можно использовать другие способы создания отдельных компартментов для реакции ПЦР, например, микрофлюидные системы, но не ограничиваясь ими. Примеры таких систем включают систему Fluidigm BioMark и системы, предоставляемые Microfluidic Chipshop.

Хотя использование такой полимеразы увеличивает эффективность и может обеспечить однородность капель, такие полимеразы могут создавать проблемы из-за добавления дополнительного [A] к 3'-концу удлиненной последовательности. Такое добавление может привести к вероятному отсутствию комплементарности OLR 110 и конца 210. В вариантах реализации приведено возможное решение такой проблемы с сохранением возможности применения высокоэффективных полимераз.

*F. Добавление A в обратный праймер: последовательности в областях A и B - специфические положения генома, содержащие A и T*

Можно выбрать области для связывания, для которых наличие дополнительного A, добавляемого полимеразой на 3'-концах, является преимуществом. Можно выбрать область с T на 5'-конце для комплементарности дополнительному A, добавляемому полимеразой.

На фиг. 3A-3F проиллюстрирован способ ПЦР с перекрывающимися праймерами с использованием соответствующих геномных положений с T и A согласно вариантам реализации настоящего изобретения.

На фиг. 3A показана только цепь Уотсона длинной молекулы ДНК. Прямой праймер F5 и обратный праймер R5 используют для амплификации области A. Прямой праймер F6 и обратный праймер R6 используют для амплификации области B. Область A сконструирована так, что на 3'-конце 307 находится тимин. Область B специально сконструирована так, что нуклеотид, расположенный после (в 3'-направлении от) 3' конца праймера F6, представляет собой аденозин 308. Точечная линия между областями A и B представляет собой фрагмент длинной ДНК, например, размером от 100 до 100 т.п.о., представляющий собой промежуточную область 305 между областями A и B.

На фиг. 3В на 5'-конце обратного праймера R5 размещают дополнительную последовательность 310, комплементарную F6. Эта область праймера называется перекрывающейся областью (OLR). 3'-конец R5 используют для выполнения ПЦР-амплификации области А. Аденозин 309 отделяет OLR 310 от последовательности праймера 312 для области А. Аденозин 309 комплементарен тимину 307 в области А.

На фиг. 3С после ПЦР-амплификации к 3'-концу каждой из четырех удлиненных цепей (т.е. цепей W, X, Y и Z) полимеразы добавляет дополнительные остатки аденозина [А] (например, 319 и 329). Цепи X и Y - это две цепи ПЦР-ампликонов области А, а цепи W и Z - две цепи ПЦР-ампликона области В. Дополнительный остаток аденозина 319 на 3'-конце цепи X может повысить эффективность и специфичность последующего объединения ПЦР-ампликонов области А и области В. Более того, большинство существующих доступных для приобретения микрожидкостных систем используют полимеразу, добавляющую дополнительный [А] к 3'-концу удлиняемой последовательности.

С другой стороны, поскольку в каждой амплифицированной цепи области А находится дополнительный остаток аденозина, праймер F6 должен быть специально сконструирован так, чтобы нуклеотид, расположенный после (в 3'-направлении от) 3' конца праймера F6, являлся аденозином 308 на фиг. 3А и 3С. Аналогичным образом, сайт связывания R5 для отжига в области А должен представлять собой тимин [Т] (307, как проиллюстрировано на фиг. 3А).

На фиг. 3D цепь X и цепь Z отжигают друг с другом в области OLR. Тимин [Т] 307 на 5'-конце сайта отжига R5 в области А связывается с аденозином [А] 329 на 3'-конце цепи Z. Тимин [Т] 331, созданный аденозином 308, прилегающим к 3'-концу F6, связывается с аденозином [А] 319 на 3'-конце цепи X. Эти два сайта связывания А-Т улучшают эффективность и специфичность связывания цепи X и цепи Z.

На фиг. 3Е после использования праймеров цепи X и Z можно связать друг с другом посредством удлинения полимеразой. Т.е. можно использовать полимеразу для удлинения верхней цепи вправо (в 3'-направлении), а нижнюю цепь можно удлинить влево (в 5'-направлении).

На фиг. 3F внутри реакционных компартментов концентрации R5 и F6 могут быть ниже, чем, например, концентрации F5 и R6, так что после связывания происходит амплификация удлиненного ампликона. Примеры соотношения между этими группами праймеров (R5 и F6 по сравнению с F5 и R6) включают 1:100000, 1:50000, 1:10000, 1:5000, 1:1000, 1:500, 1:200, 1:150, 1:100, 1:80, 1:60, 1:40, 1:30, 1:20, 1:10 и 1:5, но не ограничиваются ими. Причина наличия более низких концентраций R5 и F6 заключается в

необходимости гарантировать их исчерпание после формирования достаточного количества цепей X и Z во избежание образования чрезмерного количества цепей Y и W. Отдельные цепи Y и W не несут информацию о гаплотипах в отношении областей A и B. Кроме того, цепи Y и W могут взаимодействовать или образовывать гибридные продукты с продуктами амплификации области A и области B из другого компартмента в последующем процессе. Это может влиять на точность способа гаплотипирования. Более высокие концентрации F5 и R6 позволяют этим двум праймерам осуществить окончательную ПЦР-амплификацию связанных последовательностей. В других вариантах реализации для амплификации удлиненных ампликонов можно применять отдельную процедуру амплификации или вообще не амплифицировать их, например, при использовании секвенирования единственной молекулы для анализа удлиненных ампликонов.

После получения удлиненных ампликонов их можно анализировать, например, с помощью аллель-специфической ПЦР или секвенирования.

На фиг. 3A область A сконструирована так, что на 3'-конце 307 находится тимин, а фрагмент комплементарного праймера R5 включает A 309 на конце. Такое специфичное расположение аденозина [A] 309 может быть полезно для объединения области A и области B. Для ПЦР-амплификации области B с использованием F6 и R6 к 3'-концу цепи Z добавили дополнительный аденозин [A] 329 (фиг. 3C). Последовательность OLR 310 совпадает с последовательностью OLR на цепи Z (фиг. 3D). Для создания гибридных продуктов, содержащих область A и область B, необходима полная комплементарность последовательностей на 3'-концах цепи X и цепи Z (рис. 3D). Внедрение аденозина [A] 309 в R5 необходимо для обеспечения комплементарности дополнительного аденозина [A] 329, добавленного к 3'-концу цепи Z, тимину [T] 307 в цепи X. Аналогичным образом, полимеразы добавляет дополнительный аденозин [A] 319 к 3'-концу цепи X. Аденозин [A] 308 в цепи W предназначен для обеспечения комплементарности аденозина [A] 319 тимину [T] 331 в цепи Z (фиг. 3D). Такое расположение обеспечивает более полную амплификацию с получением удлиненных ампликонов, охватывающих обе области, с большей эффективностью.

*G. Добавление A в обратный праймер: неспецифические положения в геноме*

Как описано выше на фиг. 3A, область A сконструирована так, что на 3'-конце 307 находится тимин, а конец фрагмента комплементарного праймера R5 содержит A 309. Однако в некоторых вариантах реализации комплементарный фрагмент праймера может заканчиваться непосредственно перед A, и A может являться дополнительным основанием (3 фрагмент) между комплементарным фрагментом и OLR-фрагментом. Таким образом,



на конце области А не обязательно должен находиться Т. Вместо этого дополнительный А может вызвать добавление Т при амплификации.

На фиг. 4А-4С проиллюстрирован способ ПЦР с перекрывающимися праймерами с использованием обратного праймера с добавленным А согласно вариантам реализации настоящего изобретения. На фиг. 4А показана молекула ДНК с областью А и областью В. На 3'-конце области А может не находиться тимин [Т], в отличие от области А на фиг. 3А. Соответственно, в этой альтернативной конструкции на фиг. 4А сайт отжига на 5'-конце R5 не обязательно должен представлять собой тимин [Т].

На фиг. 4В показан обратный праймер R5. Дополнительный аденозин [А] 409 можно поместить между последовательностью праймера 403 для области А и OLR 410.

На фиг. 4С показаны ПЦР-ампликоны, образованные праймерами. Дополнительный [А] 409 может вызывать появление [Т] 413 в цепи X для связывания с [А] 415 на 3' конце цепи Z. [А] 415 представляет собой дополнительный аденозин, добавленный полимеразой к 3'-концу ампликона.

Дополнительный аденозин [А] в обратном праймере может позволить конструировать области, на 3'-конце находится любой нуклеотид. Область можно сконструировать на основе факторов, на конце которых находится не тимин [Т]. Например, можно сконструировать области определенной длины или области, охватывающие определенные ОНП без учета наличия тимина на конце области.

#### *Н. Добавление А в обратный праймер и прямой праймер*

В сценарии, изображенном на фиг. 4А, область В по-прежнему должна содержать А 404 после области OLR. В некоторых вариантах реализации можно добавить дополнительный А как к обратному праймеру R5, так и к прямому праймеру F6.

На фиг. 5А-5D проиллюстрирован способ ПЦР с перекрывающимися праймерами с использованием обратного и прямого праймеров с добавленным А согласно вариантам реализации настоящего изобретения. В этой другой альтернативной конструкции 5'-концы как R5, так и F6 могут формировать искусственную последовательность.

На фиг. 5В две указанные искусственные последовательности комплементарны друг другу. Дополнительный аденозин [А] 509 можно поместить между последовательностями праймера 503 и OLR 510, причем А по-прежнему входит в состав искусственных последовательностей, но не в состав OLR.

На фиг. 5С аденозин [А] 509 на R5 приводит к появлению [Т] 513 на цепи X между OLR и ПЦР-ампликоном. Этот [Т] 513 связывается с [А] 523 на 3'-конце цепи Z. Аденозин 559 в F6 приводит к появлению [Т] 563 между OLR и ПЦР-ампликоном в цепи Z. Этот [Т] связывается с [А] на 3'-конце цепи X.

В этом примере OLR обратного праймера комплементарна OLR прямого праймера, поэтому требования к последовательности областей отсутствуют, и работа с OLR идет полностью в праймерах.

На фиг. 5D показано связывание цепи X и цепи Z в OLR. Каждую цепь можно удлинить с образованием двуцепочечной молекулы, содержащей обе цепи области A и обе цепи области B.

#### *1. Праймеры для амплификации гибридных продуктов ПЦР*

Специфичные праймеры можно применять для стимуляции амплификации гибридных продуктов ПЦР, содержащих две областями, а не продуктов амплификации, включающих только одну область. Специфичные праймеры могут амплифицировать гибридные продукты ПЦР без амплификации областей, которые сами еще не были амплифицированы. Эти праймеры можно использовать в высоких концентрациях по сравнению с другими праймерами для образования большого количества гибридных продуктов ПЦР.

#### *Прямой праймер без отдельной перекрывающейся области*

На фиг. 6 проиллюстрирован принцип способа ПЦР с перекрывающимися праймерами и лигированием адаптеров, использующего праймеры для стимуляции амплификации гибридных продуктов ПЦР. В реакционный компартмент добавляют матрицу ДНК и шесть праймеров: F1, R1, F2, R2, F3 и R3. Праймеры F1, R1, F2 и R2 могут быть аналогичны праймерам F5, R5, F6 и R6, соответственно, на фиг. 3A-3F. Область 1 и область 2 представляют собой две мишени, которые являются мишенью анализа фазирования гаплотипов. Область 1 может быть аналогична области A на фиг. 3, а область 2 может быть аналогична области B на фиг. 3.

Точечная линия представляет собой геномную область между областью 1 и областью 2. Эти две области могут быть разделены несколькими тысячами пар оснований, длина точечной линии изображена не в масштабе. F1 и R1 представляют собой праймеры для амплификации области 1. F2 и R2 представляют собой праймеры, применяемые для амплификации области 2. 5'-конец R1 содержит перекрывающуюся область (OLR), комплементарную F2. Это перекрывание позволяет связывать область 1 и область 2 (как показано на фиг. 3E и 3F).

Одно ключевое отличие на фиг. 6 по сравнению с фиг. 3A-3F и 4A-4C заключается в добавлении дополнительной пары праймеров F3 и R3. Эта пара праймеров используется для амплификации гибридных продуктов ПЦР, содержащих как область 1, так и область 2. Включение этой пары праймеров (F3 и R3) особенно полезно для объединения более чем двух областей. Использование для объединения более чем двух областей подробно

описано ниже. Поскольку 5'-конец 604 R1 комплементарен F2, продукты ПЦР области 1 с использованием F1 и R1 можно объединить с продуктами ПЦР области 2. Последовательность 5'-концов F1 и R2 (608 и 612) аналогична последовательности 3'-концов F3 и R3 (616 и 620), соответственно. Область перекрывания F1 и F3 обозначена как область адаптера секвенирования 1 (SAR1), а область перекрывания R2 и R3 обозначена как область адаптера секвенирования 2 (SAR2). В результате включения SAR1 и SAR2 F3 и R3 можно применять для амплификации гибридного продукта, содержащего область 1 и область 2.

Удлиненный ампликон, образованный F1, R1, F2 и R2, может быть аналогичен цепям на фиг. 3F. Однако поскольку F1 включает SAR1, а R2 включает SAR2, две цепи ДНК, расположенные на одном конце удлиненного ампликона, содержат SAR1 и последовательность, комплементарную SAR1. Две цепи ДНК, расположенные на другом конце, содержат SAR2 и последовательность, комплементарную SAR2. Затем праймеры F3 и R3 могут амплифицировать этот удлиненный ампликон. После этого амплифицированные удлиненные ампликоны содержат секции 624 F3 и 628 R3, предназначенные для секвенирования, и комплементарные им области, например, адаптеры, присоединенные к проточной ячейке. Таким образом, амплифицированные удлиненные ампликоны можно дополнительно амплифицировать с помощью праймеров F3 и R3, которые могут быть полностью комплементарны начальной или конечной последовательностям амплифицированных удлиненных ампликонов.

Для снижения влияния негибридных продуктов ПЦР области 1 и области 2, концентрации F3 и R3 можно задать выше, чем концентрации других праймеров F1, F2, R1 и R2. Примеры отношения концентраций F3 и/или R3 к любому одному или более из F1, F2, R1 и R2 включают отношения, большие или равные 5:1, 10:1, 20:1, 50:1, 75:1, 100:1, 200:1, 500:1, 1000:1, 2000:1, 5000:1, 10000:1, 20000:1, 50000:1 или 100000:1, но не ограничиваются ими. Более низкие концентрации F1, R1, F2 и R2 увеличивают вероятность их исчерпания после образования гибридных продуктов и отсутствия значительного количества негибридных продуктов. 5'-концы F3 и R3 предназначены для облегчения секвенирования гибридных продуктов. Негибридные продукты могут соединяться с негибридными продуктами из другого компартмента в последующих процессах анализа и влиять на точность анализа гаплотипирования.

В вариантах реализации последовательность адаптеров для платформ массового параллельного секвенирования, например, последовательности адаптеров секвенирования Illumina, можно разместить на 5'-конце 624 F3 и 628 R3, обеспечивая секвенирование гибридных продуктов. В некоторых вариантах реализации можно включить

последовательность, идентифицирующую образец, т.е. индекс образца, так что образцы, полученные из различных образцов или экспериментов, можно объединять для секвенирования или другого последующего анализа. После секвенирования секвенированные прочтения можно сопоставить с соответствующими исходными образцами на основе последовательностей индексов образцов. Эти последовательности адаптеров или индексы образцов являются примерами фрагментов F3 и R3, помеченных как «предназначенные для секвенирования» на фиг. 6.

Прямой праймер с отдельной перекрывающейся областью

На фиг. 7А показан еще один пример использования праймеров для стимуляции амплификации гибридных продуктов ПЦР. 5'-концы 704 и 708 каждого из R1 и F2 можно сконструировать как искусственную последовательность. Эти две искусственные последовательности должны быть комплементарны друг другу, аналогично R5 и F6 на фиг. 5. Как и на фиг. 6, можно добавить праймеры F3 и R3 для стимуляции амплификации гибридных продуктов ПЦР. Последовательность 5'-концов 712 и 716 F1 и R2 аналогична последовательности 3'-концов 720 и 724 F3 и R3, соответственно. Область перекрывания F1 и F3 обозначена как область адаптера секвенирования 1 (SAR1), а область перекрывания R2 и R3 обозначена как область адаптера секвенирования 2 (SAR2). F3 и R3 можно применять для амплификации гибридного продукта, содержащего область 1 и область 2. Для снижения влияния негибридных продуктов ПЦР области 1 и области 2 вносят более высокие концентрации F3 и R3 по сравнению с концентрациями других праймеров F1, F2, R1 и R2. Примеры отношения концентраций F3 и R3 к F1, F2, R1 и R2 включают отношения, большие или равные 5:1, 10:1, 20:1, 50:1, 75:1, 100:1, 200:1, 500:1, 1000:1, 2000:1, 5000:1, 10000:1, 20000:1, 50000:1 или 100000:1, но не ограничиваются ими. Более низкие концентрации F1, R1, F2 и R2 увеличивают вероятность их исчерпания после образования гибридных продуктов и отсутствия значительного количества негибридных продуктов.

На фиг. 7В показан гибридный продукт ПЦР. Гибридные продукты ПЦР содержат последовательность OLR между двумя областями-мишенями.

Связывание более чем двух областей

В некоторых вариантах реализации в удлиненном ампликоне можно связать друг с другом более одной области. Связывание более чем одной области может обеспечить возможность секвенирования большего количества областей, чем если бы эти области не были связаны друг с другом. 5'-конец обратного праймера для первой области можно сконструировать так, чтобы он был комплементарен фрагменту прямого праймера для области, расположенной далее (в направлении 3'). Шаблон для обратного и прямого

праймеров можно продолжать для вовлечения дополнительных областей после первой области.

*Ж. Некоторые области связаны друг с другом*

На фиг. 8А-8В проиллюстрирован способ ПЦР с перекрывающимися праймерами для связывания более двух областей, обеспечивающий связывание друг с другом лишь определенных областей, согласно вариантам реализации настоящего изобретения. На фиг. 8А и 8В OLR обратного праймера одной области специфично комплементарна концу еще одной области, аналогично фиг. 4А-4С, где OLR комплементарна последовательности следующего прямого праймера. Соответственно, при использовании различных вышеописанных стратегий можно сконструировать OLR для связывания более чем двух областей друг с другом.

На фиг. 8А OLR на 5'-конце обратных праймеров комплементарны прямым праймерам следующей области. Например, добавленная OLR 804 в R1 комплементарна F2, добавленная OLR 808 в R2 комплементарна F3, и т.д. Полученный удлиненный ампликон включает области 1-3 в указанном порядке или, возможно, удлиненные ампликоны только областей 1 и 2 и только областей 2 и 3. Использование повышенных концентраций праймеров F1 и R3 позволяет получить полностью удлиненные ампликоны. Для крайних праймеров можно использовать такие же концентрации, как и для получения только двух областей. С помощью этой конструкции можно амплифицировать несколько областей одного фрагмента длинной ДНК и связать их друг с другом. Фазу аллелей в различных областях можно определить с помощью последующего анализа, например, массового параллельного секвенирования, но не ограничиваясь им.

На фиг. 8В как обратный, так и прямой праймеры включают дополнительную OLR в дополнение к последовательности праймера. Последовательность OLR представляет собой специфичное специализированное связывание областей, например, последовательность OLR между R1 и F2 (812 и 816) отличается от последовательности OLR между R2 и F3. Полученный удлиненный ампликон включает области 1-3 в указанном порядке, аналогичном показанному на фиг. 8А, однако при этом используют способ, показанный на фиг. 5А-5D. Соответственно, в этой альтернативной конструкции OLR конструируют на обратных праймерах и прямых праймерах. OLR обратного праймера предыдущей области комплементарна OLR прямого праймера следующей области.

*К. Любые области могут соединяться друг с другом*

В некоторых вариантах реализации области OLR могут быть одинаковыми для всех областей. Таким образом, связывание не обязательно должно быть

последовательным, поскольку возможна любая попарная комбинация областей. Спаренные удлиненные ампликоны можно анализировать (например, секвенировать) для получения информации о гаплотипах. Например, информация о спаривании областей 1 и 2 вместе с информацией о спаривании 2 и 3 (или 1 и 3) может позволить получить информацию о гаплотипах для 1, 2 и 3. Таким образом, образования пар может быть достаточно. По сути, этот подход позволяет произвольно связывать любые две пары областей.

На фиг.9А-9В проиллюстрирован способ ПЦР с перекрывающимися праймерами для связывания более двух областей, обеспечивающий возможность связывания любых областей друг с другом, согласно вариантам реализации настоящего изобретения. В этом другом варианте OLR, расположенные на обратном и прямом праймерах, могут быть стандартными. На фиг. 9А показана молекула с тремя областями (областями 1, 2, 3) и связанными с ней прямым и обратным праймерами.

На фиг. 9В показана подробная информация о праймерах. Например, OLR 904, расположенная на 5'-конце обратного праймера R1 для области 1, комплементарна прямому праймеру, расположенному на 5'-конце прямых праймеров для областей 2 и 3 (а также прямому праймеру для областей 4, 5, 6 и т.д., при их наличии). Таким образом, все праймеры содержат OLR. Т.е. любые F- и R-праймеры комплементарны друг другу. При использовании такой конструкции аллели в области 1 можно фазировать с любыми другими областями (областями 2, 3, 4 и т.д.). Затем можно определить фазу всех областей.

В качестве примера предположим, что существуют три области с ОНП А/Т. Если известно, что А в области 1 связан с А в области 2, то эти два А находятся в одном и том же гаплотипе. А если А в области 1 связан с А в области 3, то можно определить, что гаплотип всех областей представляет собой ААА.

*L. Праймеры для амплификации гибридных продуктов ПЦР с несколькими областями*

Прямые и обратные праймеры, предназначенные для амплификации гибридных продуктов, можно использовать с несколькими областями. Прямые и обратные праймеры можно конструировать аналогично праймерам для связывания двух областей на фиг. 6, 7А и 7В.

На фиг. 10А показано, как использовать праймеры для амплификации гибридных продуктов ПЦР с несколькими областями. При использовании вышеописанных стратегий можно сконструировать OLR для связывания 3 или более областей друг с другом. OLR на 5'-конце обратных праймеров комплементарны прямым праймерам следующей области. С помощью этой конструкции можно амплифицировать несколько областей одного

фрагмента длиной ДНК и связать их друг с другом. Этот способ аналогичен способу, описанному на фиг. 8А-8В, за исключением добавления пары праймеров (F4 и R4), 3'-концы которых содержат ту же последовательность, что и 5'-конец F1 и R3, соответственно. Эту пару праймеров добавляют для амплификации полностью гибридных продуктов, включающих все области, подлежащие объединению. Предпочтительной амплификации полностью гибридного продукта можно достичь за счет использования более высоких концентраций F4 и R4 по сравнению с другими праймерами, включая F1, F2, F3, R1, R2 и R3. С помощью F4 и R4 можно амплифицировать только полностью гибридные продукты, связывающие все три области. Частично гибридные продукты, связывающие только две области, или негибридные продукты ПЦР любой из трех областей нельзя амплифицировать с помощью F4 и R4. Из-за более высоких концентраций F4 и R4 количество полностью гибридных продуктов может быть намного выше, чем количество негибридных продуктов и частично гибридных продуктов ПЦР. Примеры отношения концентраций F4 и R4 к F1, F2, R1 и R2 включают отношения, большие или равные 5:1, 10:1, 20:1, 50:1, 75:1, 100:1, 200:1, 500:1, 1000:1, 2000:1, 5000:1, 10000:1, 20000:1, 50000:1 или 100000:1, но не ограничиваются ими.

На фиг. 10В проиллюстрирована амплификация гибридных продуктов ПЦР, содержащих три области с прямыми праймерами с OLR в качестве фрагментов, отдельных от фрагментов, комплементарных указанным областям, в отличие от прямых праймеров, содержащих OLR, равнозначную фрагменту, комплементарную указанной области. Объединение продуктов ПЦР осуществляется аналогично схеме, показанной на фиг. 7А. На фиг. 10В использование F4 и R4 в более высоких концентрациях по сравнению с F1, F2, F3, R1, R2 и R3 может обогатить полностью гибридные продукты, аналогично использованию F4 и R4 на фиг. 10А.

Пример гаплотипирования с использованием гена *nudt15*

Авторы изобретения использовали гаплотипирование вариантов гена *NUDT15* в качестве примера для иллюстрирования этого способа гаплотипирования.

#### *M. NUDT15*

Ген *NUDT15* кодирует фермент нудикс-гидролазу 15, метаболизирующий 6-меркаптопурин (6-МР), тиопуриновый препарат, который является основным компонентом поддерживающего лечения при детском остром лимфобластном лейкозе (ОЛЛ). Ряд вариантов гена *NUDT15* человека являются критическими детерминантами непереносимости тиопурина (Moriyama et al., Nature Genetics, 2016;48:367-373). Варианты, расположенные в экзоне 1 и экзоне 3 гена *NUDT15*, могут приводить к потере функции нудикс-гидролазы 15 и к тяжелой иммуносупрессии при назначении этим пациентам

нормальной дозы 6-MP. Комбинация генотипов в этих вариантах дает ряд гаплотипов.

На фиг. 11 показаны шесть гаплотипов, содержащие различные комбинации генотипов в экзонах 1 и 3 *NUDT15*.

Для пациентов с наличием генетических вариантов в обоих экзонах 1 и 3 важно определить фазу вариантов в этих двух локусах, поскольку это может влиять на скорость метаболизма иммунодепрессантов и, следовательно, на оптимальную дозу, которую следует давать пациенту. Например, при обнаружении гетерозиготного варианта p.Val18\_Val19insGlyVal в экзоне 1 и гетерозиготного варианта p.Arg139Cys в экзоне 3 эти два варианта могут находиться на одной или разных хромосомах. Если эти два варианта находятся на одной хромосоме, у пациента имеют место гаплотипы \*1 и \*2. Поскольку у пациента есть нормальная копия гена *NUDT15*, он может продуцировать функциональную нудикс-гидролазу 15. Дозировку иммунодепрессанта следует скорректировать до умеренной дозы, как в случае наличия только одного варианта.

С другой стороны, если указанные два варианта находятся на двух разных хромосомах, у пациента имеют место гаплотипы \*3 и \*6. В этой ситуации пациент не может продуцировать функциональную нудикс-гидролазу 15, и дозу иммунодепрессанта потребуется заметно снизить. В мета-анализе, выполненном Yin с соавт. (Yin et al., *Oncotarget*, 2017;8:13575-85), пациентам с одной или двумя копиями аллелей риска *NUDT15* требовалась на 23% и 65% более низкая средняя суточная доза тиопурина по сравнению с пациентами с *NUDT15* дикого типа, соответственно.

Определение гаплотипов вариантов *NUDT15* является сложной задачей по двум причинам. Во-первых, расстояние между сайтами с высокой частотой мутаций в экзоне 1 и экзоне 3 составляет приблизительно 8 т.п.о. Подходы на основе ПЦР дальнего действия ограничены, поскольку эффективность и надежность ПЦР снижаются по мере увеличения размера ампликона, что приводит к образованию искусственных рекомбинантов. Другая сложность заключается в том, что V18V19insGV представляет собой вариант тандемного повтора из 6 пар оснований. Это приводит к невозможности использования способов фазирования гаплотипов, основанных на дискриминации аллелей праймерами или зондами, например, множественной дуплексной ПЦР (Regan et al., *PLoS One*, 2015;10:e0118270). Технологии, позволяющие получать информацию о локусах, разнесенных на большое расстояние, с помощью секвенирования с короткими прочтениями, т.е. 10X Genomics, или прямого секвенирования отдельной молекулы на большом протяжении, т.е. PacBio, относительно дороги и характеризуются низкой производительностью. Следовательно, данный локус является хорошим кандидатом для демонстрации возможностей этого нового способа гаплотипирования.



Шестьдесят три ребенка с ОЛЛ были набраны из отделения химической патологии и педиатрии больницы принца Уэльского (PWH), Гонконг. Эти пациенты были первоначально направлены на генотипирование *TPMT* после возникновения миелосупрессии во время лечения тиопурином. Генотипирование *NUDT15* выполняли для всех пациентов с помощью секвенирования по Сэнгеру. Родителей или бабушек/дедушек 4 пациентов, гетерозиготных по вариантам как V18V19insGV, так и R139C, также генотипировали для определения истинных гаплотипов соответствующих пациентов. Фазу двух указанных вариантов для этих пациентов определяли с помощью семейного анализа. Кроме того, 506 здоровых субъектов также проанализировали на предмет потенциальных вариантов в двух локусах *NUDT15*.

*N. Генотипирование гена NUDT15 посредством высокоточного анализа кинетики плавления продуктов амплификации*

Перед началом анализа гаплотипирования авторы изобретения определили субъектов, которые должны быть гетерозиготными по обоим рассматриваемым локусам, поскольку для выбора подходящей дозы лечения этих субъектов тиопурином должен быть полезен анализ гаплотипирования. Разработали два варианта высокоточного анализа кинетики плавления продуктов амплификации (HRM). Один вариант анализа был специфичен по отношению к сайту с высокой частотой мутаций в экзоне 1 гена *NUDT15*, содержащего варианты V18V19insGV и V18I. В еще одном варианте анализа амплифицировали фрагмент, охватывающий варианты R139C и R139H в экзоне 3. Последовательности праймеров перечислены в таблице 1.

Таблица 1. Праймеры для высокоточного анализа кинетики плавления продуктов амплификации гена *NUDT15*

Праймеры	Последовательность (5'-3')
Экзон 1, прямой	TATGACGGCCAGCGCAC
Экзон 1, обратный	ACGCAACGCGGATGCT
Экзон 3, прямой	CCTCCCCTGGACCAGCTT
Экзон 3, обратный	CCACCAGATGGTTCAGATCTTCT

Для каждого образца выполняли ПЦР в двух повторностях. Получали двадцать микролитров реакционной смеси, содержащей 10 мкл мастер-микса для высокоточного анализа кинетики плавления продуктов амплификации 2 × LightCycler® 480, прямой и обратный праймеры в конечной концентрации 300 нмоль/л (анализ экзона 1) или 250 нмоль/л (анализ экзона 3), MgCl<sub>2</sub> в конечной концентрации 3,5 ммоль/л (анализ экзона 1) или 2,5 ммоль/л (анализ экзона 3) и 20 нг ДНК. Для анализа экзона 1 в реакционную смесь для ПЦР добавляли 10% диметилсульфоксида для улучшения амплификации.

ПЦР выполняли с использованием прибора LightCycler® 96. Профиль ПЦР состоял из предварительного инкубирования при 95°C в течение 10 минут с последующей

программой амплификации. Для анализа экзона 1 программа амплификации состояла из 50 циклов, включавших 95°C в течение 15 секунд, снижение со скоростью 1,5°C/с до 59°C в течение 20 секунд и 72°C в течение 20 секунд. Для анализа экзона 3 программа амплификации состояла из 45 циклов, включавших 95°C в течение 10 секунд, 59°C в течение 15 секунд со скоростью снижения 2,2°C/с и 72°C в течение 15 секунд. После ПЦР-амплификации выполняли цикл плавления по следующему протоколу: 95°C в течение 1 минуты, 40°C в течение 1 минуты, а затем непрерывная регистрация флуоресценции при нагревании от 60°C до 90°C со скоростью 0,04°C/с. Данные высокоточного анализа кинетики плавления продуктов амплификации анализировали с помощью программного приложения LightCycler® 96. Образец дикого типа обрабатывали на каждой планшете в качестве отрицательного контроля.

Для 73 субъектов, имеющих отношение к ОЛЛ (т.е. 63 пациентов с ОЛЛ и 10 членов семей) выявляли мутационный статус в экзоне 1 и экзоне 3 гена *NUDT15* секвенированием по Сэнгеру. Шестнадцать из этих субъектов несли мутации в экзоне 1, включая 15 субъектов с гетерозиготным вариантом V18V19insGV и 1 субъекта с вариантом V18I. В отношении экзона 3 48 субъектов были гомозиготными по последовательности дикого типа, 21 субъект были гетерозиготными по R139C и дикому типу, а 4 субъекта были гомозиготными по R139C. Авторы изобретения выполнили высокоточный анализ кинетики плавления продуктов амплификации для экзона 1 и экзона 3 гена *NUDT15* для всех субъектов.

На фиг. 12A и 12B показаны кривые плавления, позволяющие различать образцы с различными генотипами *NUDT15* в областях экзона 1 (A) и экзона 3 (B). На фиг. 12A анализ HRM, специфичный по отношению к экзону 1 гена *NUDT15*, позволил дифференцировать гетерозиготный вариант V18V19insGV и гетерозиготный вариант V18I от последовательностей дикого типа на основании формы нормированных пиков плавления. На фиг. 12B анализ HRM позволял дифференцировать два генотипа в экзоне 3 гена *NUDT15* (т.е. гетерозиготный по R139C и гомозиготный по R139C) от последовательностей дикого типа.

*О. Анализ гаплотипов гена NUDT15 на основе цифровой капельной ПЦР с перекрывающимися праймерами*

Для определения фазы вариантов в двух локусах *NUDT15* авторы изобретения разработали способ гаплотипирования с использованием капельной цифровой ПЦР с перекрывающимися праймерами. ДНК с высокой молекулярной массой, выделенную из лейкоцитарного слоя, распределяли по каплям нанолитрового размера для выполнения связывающей ПЦР с использованием системы капельной цифровой ПЦР BioRad Droplet

Digital PCR. Концентрацию матрицы ДНК контролировали так, чтобы заполнить только приблизительно 10% капель, так что > 99% капель должны содержать только одну матрицу ДНК, охватывающую две рассматриваемые области, или не содержать матрицы ДНК. ПЦР с перекрывающимися праймерами выполняли так, чтобы обеспечить амплификацию двух областей, несущих указанные варианты, с одной длинной молекулы ДНК и связывание их друг с другом.

На фиг. 13А-13F показано применение способа ПЦР с перекрывающимися праймерами к экзонам 1 и 3 гена *NUDT15* согласно вариантам реализации настоящего изобретения. Фиг. 13А-13F зеркально отображают фиг. 3А-3F, хотя можно использовать и методики, описанные в других способах, например, на фиг. 4А-4С и 5А-5D.

На фиг. 13А анализ на основе ПЦР с перекрывающимися праймерами, используемый для гаплотипирования гена *NUDT15*, включает две пары праймеров, одну пару (F1 и R1) для амплификации вариантной области в экзоне 1, а другую пару (F2 и R2) для амплификации вариантной области в экзоне 3.

На фиг. 13В обратный праймер для экзона 1 (R1) содержит линкерную последовательность на 5'-конце, комплементарную последовательности прямого праймера для экзона 3 (F2).

На фиг. 13С в начальных раундах ПЦР две области амплифицируются по отдельности. Две ПЦР выполняли одновременно в одной капле. Как обсуждалось ранее, полимеразы Taq добавляет дополнительные основания аденозина [A] на 3'-конце удлиненных цепей ДНК.

На фиг. 13D OLR, сконструированная для 5'-конца R1, делает 3'-конец цепи X комплементарным 3'-концу цепи Z.

На фиг. 13Е две указанные цепи можно связать друг с другом посредством удлинения с помощью полимеразы Taq или других полимераз.

На фиг. 13F выполняют дальнейшую ПЦР-амплификацию связанного продукта с использованием праймеров F1 и R2, поскольку концентрации F1 и R2 превышают концентрации R1 и F2. Размер гибридного продукта ПЦР должен составлять 190 п.о. в случае гаплотипа, несущего мутацию p.V18V19insGV, или 184 п.о. в противном случае.

Таблица 2. Праймеры для гаплотипирования гена *NUDT15* посредством ПЦР с перекрывающимися праймерами

Праймеры	Последовательность (5'-3')
Прямой праймер (F1)	TATGACGGCCAGCGCAC
Обратный праймер (R1)	CCAGGGGAGGTAGTTCTTCCCACGCAACGCGGATGCT
Прямой праймер (F2)	GGGAAGAACTACCTCCCCTGG
Обратный праймер (R2)	CCACCAGATGGTTCAGATCTTCT

Реакционные смеси для ПЦР-получали в четырех повторностях для каждого образца в объеме 20 мкл; они состояли из 2,5 нг матричной ДНК, 10 мкл смеси для зондов 2 × ddPCR Supermix for Probes (не содержащей дУТФ) (Bio-Rad), F1 и R2 в конечной концентрации 1 мкмоль/л и R1 и F2 в конечной концентрации 30 нмоль/л. Затем реакционную смесь для ПЦР разделяли с использованием генератора капель для цифровой ПЦР QX100/QX200 (Bio-Rad). Температурный профиль анализа представлял собой: инициацию при 95°C в течение 10 минут, затем 45 циклов, включавших 94°C в течение 30 секунд и 61°C в течение 1 минуты, и заключительное инкубирование при 98°C в течение 10 минут.

После ПЦР объединяли капли из повторных лунок для одного образца и выделяли продукты ПЦР в соответствии с протоколом производителя с небольшой поправкой в виде использования 40 мкл буфера TE для каждой лунки вместо 20 мкл. Эта поправка привела к лучшему дроблению капель.

Для обогащения желательных гибридных продуктов ПЦР длиной 184 или 190 п.о. выделенные продукты ПЦР дополнительно очищали с использованием гранул AMPure XP Beads в соответствии с протоколом производителя для двойного отбора по размеру на основе гранул. Отношение гранулы/ДНК (т.е. отношение совокупного объема добавленных гранул к объему исходного образца) для первого и второго отбора с помощью гранул составляло 0,8 и 1,6, соответственно. Затем анализировали профиль размера очищенных гибридных продуктов ПЦР с использованием D1000 SreenTape и реагентов на приборе 4200 TapeStation (Agilent Technologies).

#### *Р. Секвенирование гибридных продуктов ПЦР гена NUDT15*

Гибридные продукты ПЦР анализировали посредством секвенирования следующего поколения, хотя можно использовать и другие методы измерения, если они подходят для последовательностей связанных областей. Индексированную библиотеку конструировали для гибридных продуктов ПЦР из каждого образца с использованием набора для подготовки библиотеки КАРА (Kapa Biosystems) в соответствии с рекомендациями производителя. До 48 библиотек с индексами мультиплексировали и секвенировали с использованием платформы секвенирования Illumina MiSeq (76 × 2 цикла). Шесть возможных последовательностей гибридных продуктов ПЦР, соответствующих шести гаплотипам, использовали в качестве эталонных последовательностей для картирования данных секвенирования (прочтений последовательностей). Прочтения со спаренными концами выравнивали на 6 эталонных последовательностей с использованием программы Bowtie 2 (<http://bowtie-bio.sourceforge.net/bowtie2/>). Для дальнейшего анализа использовали прочтения со

спаренными концами, оба конца которых соответственно выравнивали на эталонную последовательность без несовпадений.

Прочтения, выровненные на каждую последовательность гаплотипов в образце, подсчитывали с помощью сценария Perl. Затем рассчитывали процент прочтений, картированных с каждым гаплотипом (обозначенный как  $N_{ap}\%$ ) в образце, путем деления количества прочтений, картированных с каждым гаплотипом, на общее количество прочтений, картированных со всеми гаплотипами. Авторы изобретения задали  $N_{ap}\% < 10\%$ ,  $40\% - 60\%$  и  $> 90\%$  в качестве пороговых значений для присутствия 0, 1 и 2 копий гаплотипов в образце, соответственно. В других вариантах реализации в качестве порогового значения для присутствия 0 гаплотипов в образце можно использовать другие процентные значения, например, 1%, 2%, 5%, 15%, 20%, но не ограничиваясь ими. Аналогичным образом, примеры 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70% можно использовать в качестве пороговых значений для присутствия 1 гаплотипа. Аналогичным образом, примеры 75%, 80%, 85%, 92%, 94%, 98%, 99% можно использовать в качестве пороговых значений для присутствия 2 гаплотипов.

Комбинации гаплотипов (т.е. диплотипов) можно было определить по результатам генотипирования с помощью анализа HRM для большинства из 73 субъектов, за исключением 9 субъектов, являющихся двойными гетерозиготами по мутациям V18V19insGV и R139C. Посредством семейного анализа выявили, что восемь из этих 9 субъектов представляли собой \*1\*2. Для непосредственного определения диплотипов авторы изобретения выполнили цифровую капельную ПЦР с перекрывающимися праймерами с последующим секвенированием следующего поколения для всех субъектов. Следуя критериям, описанным выше, подсчитали количество копий каждого из 6 потенциальных гаплотипов для всех субъектов. Сводные результаты представлены в таблице 3. Диплотипы гена *NUDT15* были правильно выявлены у всех субъектов.

Таблица 3. Гаплотипы гена *NUDT15*, определенные с помощью анализа на основе цифровой капельной ПЦР с перекрывающимися праймерами у субъектов, имеющих отношение к ОЛЛ.

Диплотип	Кол-во образцов	Копии гаплоидного гена <i>NUDT15</i>					
		*1	*2	*3	*4	*5	*6
*1*1	44	2	0	0	0	0	0
*1*2	9	1	1	0	0	0	0
*1*3	11	1	0	1	0	0	0
*1*6	4	1	0	0	0	0	1
*2*3	2	0	1	1	0	0	0
*3*3	2	0	0	2	0	0	0
*3*5	1	0	0	1	0	1	0

*Q. Генотипирование и гаплотипирование гена NUDT15 у 506 здоровых субъектов*

Для оценки эффективности и применимости способа HRM и способа гаплотипирования на основе цифровой ПЦР с перекрывающимися праймерами для анализа гена NUDT15 авторы изобретения дополнительно проанализировали 506 здоровых субъектов.

Вначале выполнили скрининг 506 субъектов на варианты *NUDT15* с использованием двух анализов на основе HRM. Результаты представлены в таблице 4. Непосредственно по результатам генотипирования оказалось можно определить шесть различных диплотипов, включая \*1\*1, \*1\*3, \*1\*5, \*1\*6, \*2\*3 и \*3\*3. Кроме того, 22 образца содержали двойные гетерозиготные варианты для V18V19insGV и R139C, а у одного субъекта генотипы в экзоне 1 были неизвестны. Для определения диплотипов этих 23 образцов использовали цифровую ПЦР с перекрывающимися праймерами с последующим секвенированием.

У всех 22 субъектов-сложных гетерозигот по результатам секвенирования определили диплотип \*1\*2. Субъект с неизвестным генотипом в экзоне 1 был идентифицирован как несущий мутации V18V19insGV и V18I. Таким образом, диплотип этого субъекта представлял собой \*5\*6, что является ранее неизвестным диплотипом. Частоты диплотипов *NUDT15* в данной группе также перечислены в таблице 4. Эти результаты сопоставимы с результатами, полученными на основе фазированных данных проекта 1000 Genomes для Восточной Азии (Moriyama et al. *Nat Genet.* 2016;48:367-373).

Таблица 4. Диплотипы гена *NUDT15* у контрольных субъектов, определенные с помощью HRM и цифровой ПЦР с перекрывающимися праймерами.

Генотип экзона 1 по HRM	Генотип экзона 3 по HRM	Диплотип, определенный с помощью HRM	Диплотип, определенный с помощью цифровой ПЦР с перекрывающимися праймерами	Кол-во образцов	Частота диплотипа
ДТ/ДТ	ДТ/ДТ	*1*1	-	380	75,1%
ДТ/ДТ	R139C/ДТ	*1*3	-	62	12,3%
V18I/ДТ	ДТ/ДТ	*1*5	-	12	2,4%
V18V19insGV/ДТ	ДТ/ДТ	*1*6	-	22	4,3%
V18V19insGV/ДТ	R139C/R139C	*2*3	-	5	1,0%
ДТ/ДТ	R139C/R139C	*3*3	-	2	0,4%
V18V19insGV/ДТ	R139C/ДТ	-	*1*2	22	4,3%
Неизвестно	ДТ/ДТ	-	*5*6	1	0,2%

Гаплотипирование с анализом фазовой сборки

Гаплотипирование может включать определение аллелей более чем в двух гетерозиготных локусах и, следовательно, более чем в двух регионах с ОНП. В таких

случаях эти области могут быть собраны в гибридные продукты, содержащие лишь две или три области. Затем можно определить гаплотип путем анализа полученных продуктов из двух или трех областей. Продукты, составленные из нескольких различных областей, можно связать друг с другом посредством общей области. Этот анализ гаплотипирования с использованием общих областей для связывания различных гибридных продуктов друг с другом подробно описан ниже.

#### *R. Группировка областей*

На фиг. 14А, 14В и 14С показаны четыре области, разделенные на несколько групп по две области. На фиг. 14А в этом типичном примере показаны четыре области. Однако количество областей не ограничено и может быть равно любому количеству областей, подлежащих фазированию. У каждой из четырех областей есть соответствующие прямой и обратный праймеры, возможно, содержащие фрагменты праймера и перекрывающиеся фрагменты.

На фиг. 14В показаны две группы областей и конфигурация праймеров. Для наглядности четыре области разделены на две группы (группу А и группу В). Количество групп зависит от количества областей, подлежащих фазированию или гаплотипированию; кроме того, длину ДНК можно определить с помощью последующего анализа, например, путем массового параллельного секвенирования. Например, с помощью платформы секвенирования Illumina можно выполнить секвенирование спаренных концов по 150 п.о. на каждом конце. При такой схеме для любых гибридных молекул ДНК выполняется секвенирование фрагмента общей длиной 300 п.о. Исходя из того, что размер амплифицированного продукта каждой области составляет 75 п.о., а длина связывающих областей - 25 п.о., секвенирование 300 п.о. может выявить последовательность до 3 областей в гибридной молекуле. При такой схеме для группировки областей можно использовать до трех групп. Если необходимо гаплотипировать 30 областей, то каждая группа может состоять из 10 областей. Каждая группа может быть содержать одинаковое или неодинаковое количество областей. Если последующий анализ может выявить последовательность более длинных молекул ДНК, можно использовать большее количество групп.

На фиг. 14В также показана конфигурация праймеров на основе двух групп. Праймеры в каждой группе содержат одинаковые перекрывающиеся области, комплементарные перекрывающимся областям другой группы. В данном случае перекрывающаяся область 1 (OLR1), расположенная на 5'-концах обратных праймеров группы А, комплементарна перекрывающейся области 2 (OLR2), расположенной на 5'-концах прямых праймеров группы В. Эта конструкция праймеров аналогична способу,

описанному на фиг. 7А выше. F5 и R5 представляют собой пару праймеров, которые могут специфически амплифицировать гибридные продукты, включающие область из группы А и область из группы В. Концентрации F5 и R5 могут быть выше концентраций других праймеров (от F1 до F4 и от R1 до R4). При этих повышенных концентрациях F5 и R5 после ПЦР с перекрывающимися праймерами концентрации гибридных продуктов могут быть намного выше, чем концентрации негибридных продуктов ПЦР.

На фиг. 14С показаны гибридные продукты, включающие области из группы А и группы В. После ПЦР с перекрывающимися праймерами любой член группы А может быть объединен с членом группы В. В данном примере есть два члена в группе А (область 1 и область 3) и два члена в группе В (область 2 и область 4). Таким образом, есть четыре типа гибридных продуктов: область 1, объединенная с областью 2, область 1, объединенная с областью 4, область 3, объединенная с областью 2, и область 3, объединенная с областью 4. Используя эту конфигурацию, можно контролировать размер гибридных продуктов. В данном примере все гибридные продукты содержат только две области. Ниже приведен пример с тремя областями.

Преимущество объединения меньших групп областей по сравнению с методом последовательного связывания, описанным на фиг. 10А или фиг. 10В, заключается в возможности фазирования большего количества областей. Например, платформа для секвенирования Illumina является одной из часто используемых платформ для определения последовательности отдельных молекул. Количество областей, которые можно фазировать с использованием способа последовательного объединения, описанного на фиг. 10А или фиг. 10В, ограничено длиной регистрируемых последовательных прочтений. Например, используя секвенирование спаренных концов по 300 нуклеотидов на каждом конце, можно секвенировать молекулы длиной до 600 нуклеотидов. Если длина каждой области, подлежащей фазированию, составляет 50 нуклеотидов, то с помощью способа последовательного объединения можно фазировать максимум 12 областей.

Однако при использовании способа группировки, описанного в настоящем документе, теоретически не существует максимального ограничения количества областей, подлежащих фазированию. В этой схеме многолокусную ПЦР, позволяющую одновременно амплифицировать целевые области (области, подлежащие гаплотипированию), можно выполнять в одном компартменте. Затем амплифицированные продукты этих областей объединяют с помощью способа, описанного в настоящей заявке. Описаны способы увеличения количества множественных реакций в одной реакции (Wei et al. *Journal of Virological Methods* 2008;151:132-139; Mo et al. *Clin Chem Lab Med*; 50:649-



654). Например, при использовании четырех групп, каждая из которых включает 20 областей, можно фазировать в общей сложности 80 областей. Каждый гибридный продукт может содержать четыре области. Если каждая область состоит из 50 нуклеотидов, то гибридный продукт должен содержать 200 нуклеотидов, что находится в пределах возможностей платформы секвенирования Illumina.

Гаплотипирование может также включать биоинформатический анализ, который определяет принадлежность разных аллелей к одному и тому же гаплотипу на основе характера спаривания. Соответственно, количество областей, подлежащих гаплотипированию, может не ограничиваться длиной матричной ДНК, используемой для ПЦР с перекрывающимися праймерами. Общую область между двумя наборами гибридных продуктов ПЦР можно использовать для вместе гаплотипов других областей друг с другом в составе гибридных продуктов. Например, можно определить гаплотипы трех областей (например, ОНП 1, ОНП 2 и ОНП 3), если определены фаза ОНП 1 и ОНП 2 и фаза ОНП 2 и ОНП 3. Этот анализ можно повторить для гаплотипирования нескольких областей, как описано ниже.

На фиг. 15А, 15В и 15С показано формирование гибридных продуктов, содержащих три области, по одной области из каждой из трех групп. На фиг. 15А проиллюстрировано случайное разделение областей на три группы в качестве примера. Вместе с тем, для этого способа можно использовать более 3 групп (например, 4, 5, 6, 7 или более). В иллюстративных целях области последовательно приписаны к трем группам. Последовательность приписывания областей к группам может не влиять на результаты. В некоторых вариантах реализации приписывание областей к группам может быть скорее случайным, чем последовательным.

На фиг. 15В показана конфигурация праймеров на основе групп. OLR, расположенная на 5'-конце каждого обратного праймера в одной группе, комплементарна OLR, расположенной на 5'-конце каждого прямого праймера в следующей группе. В этом типичном примере OLR обратных праймеров группы А комплементарны OLR прямых праймеров группы В, причем эти OLR показаны точечными линиями. OLR обратных праймеров группы В комплементарны OLR прямых праймеров группы С, причем эти OLR показаны пунктирными линиями.

Пара внешних праймеров Fz и Rz используется для амплификации полностью гибридных продуктов. Fz содержит OLR, совпадающую с 5'-концом F1 и F4, причем эта OLR изображена двойными сплошными линиями. Rz содержит OLR, совпадающую с 5'-концом R3 и R6, причем эта OLR изображена двойными пунктирными линиями. Концентрации Fz и Rz могут превышать концентрации всех других праймеров,

используемых для амплификации конкретных областей, для стимуляции амплификации гибридных продуктов трех областей. Примеры отношения концентраций Fz и Rz к другим праймерам включают отношения, большие или равные 5:1, 10:1, 20:1, 50:1, 75:1, 100:1, 200:1, 500:1, 1000:1, 2000:1, 5000:1, 10000:1, 20000:1, 50000:1 или 100000:1, но не ограничиваются ими.

На фиг. 15С показан гибридный продукт ПЦР. При такой схеме каждый полностью гибридный продукт содержит три области: одну из группы А, одну из группы В и одну из группы С.

*S. ПЦР с перекрывающимися праймерами с использованием полимеразы без добавления аденозина на 3'-конце*

В вариантах реализации можно выполнять ПЦР с перекрывающимися праймерами с использованием полимеразы, не добавляющей аденозин на 3'-конце удлиняемой цепи ДНК. На фиг. 16А показана конструкция праймеров. Проиллюстрированная конструкция аналогична конструкции, описанной на фиг. 7А. Однако праймер R1 на фиг. 16А модифицирован путем удаления фланкирующего аденозина между последовательностью праймера для области 1 и OLR. Аналогичным образом, праймер F2 модифицирован путем удаления фланкирующего аденозина между последовательностью праймера для области 2 и OLR. Праймер F3 модифицирован путем удаления фланкирующего тимина между SAR1 и последовательностью, расположенной на 5'-конце, которая предназначена для секвенирования. Праймер R3 модифицирован путем удаления фланкирующего тимина между SAR2 и последовательностью, расположенной на 5'-конце, которая предназначена для секвенирования. На фиг. 16В показан гибридный продукт ПЦР, содержащий две области. Способы, описанные на фиг. 14А-14С и фиг. 15А-15С? можно применять с этой модификацией конструкции праймеров и использованием полимеразы без добавления дополнительного аденозина на 3'-конце удлиняемой цепи ДНК.

*T. Влияние длины между областями*

Области, расположенные на различном расстоянии друг от друга, связали друг с другом. Затем гибридные области проанализировали с целью определить, как длина влияет на точность анализа гаплотипов.

Пример конфигурации 10 ОНП

На фиг. 17А показаны значения длины между различными ОНП в молекуле ДНК. Эти разные значения длины использовали для определения влияния длины на способ ПЦР с перекрывающимися праймерами. Целью каждой ПЦР с перекрывающимися праймерами было объединение двух областей, одна из которых является ОНП1, а другая может являться одним из ОНП2, ОНП3, ОНП4, ОНП5, ОНП6, ОНП7, ОНП8, ОНП9 или ОНП10.

Расстояние между двумя областями, подлежащими объединению, составляет от 2,5 до 100 т.п.о.

На фиг. 17В показаны конфигурации праймеров для ОНП. Две области объединяли с использованием способа, описанного на фиг. 7А и фиг. 7В. 5'-конец обратного праймера для амплификации ОНП1 (праймер R1) включает перекрывающуюся область 1 (OLR1). 5'-концы прямых праймеров для амплификации ОНП2- ОНП10 включают перекрывающуюся область 2 (OLR2). OLR1 и OLR2 комплементарны друг другу. При такой схеме продукты ПЦР ОНП1 можно объединить с каждым из продуктов ПЦР ОНП2 - ОНП10. Однако возможность правильного фазирования ОНП1 и одного из ОНП2 - ОНП10 может зависеть от наличия интактной длинной молекулы ДНК, охватывающей указанные два ОНП, подлежащие фазированию, в одном компартменте, который, например, может быть каплей при цифровой капельной ПЦР.

### Результаты

Авторы изобретения отобрали беременную женщину в отделении акушерства и гинекологии больницы принца Уэльского (РWH), Гонконг. Получили образцы периферической крови и собрали лейкоцитарный слой матери. Кроме того, получили образец плода, взяв образцы плаценты после родов. Лейкоцитарный слой матери и ткань плаценты плода генотипировали с использованием платформы Illumina для анализа микрочипов (HumanOmni2.5). Отобрали 10 ОНП, расположенных в генах *гемоглобина гамма 2* и *гемоглобина эпсилон 1* хромосомы 11. Расстояние до второго, третьего, четвертого, пятого, шестого, седьмого, восьмого, девятого и десятого ОНП от первого ОНП составляет примерно 2,5, 5, 10, 20, 30, 40, 50, 75 и 100 т.п.о., соответственно. ОНП гомозиготны у плода и гетерозиготны у матери. В этом случае плод унаследовал два идентичных гаплотипа, по одному от каждого родителя. Затем на основе информации о гаплотипе плода и информации о генотипе матери можно определить два гаплотипа матери. Эта информация о гаплотипе матери может служить золотым стандартом для определения точности способа на основе цифровой ПЦР с перекрывающимися праймерами.

На фиг. 18 показаны последовательности праймеров для амплификации каждой области. В первом столбце показаны идентификаторы 10 ОНП. Во втором столбце указаны координаты ОНП на хромосоме 11. В третьем столбце указано расстояние от первого ОНП в тысячах пар оснований. В четвертом столбце указан генотип ОНП. В пятом столбце приведена последовательность прямого праймера. В шестом столбце приведена последовательность обратного праймера. Обратный праймер для ОНП1 содержит OLR1 (выделена желтым и подчеркнута). OLR1 комплементарна каждой OLR2

(выделенной зеленым и подчеркнутой) прямых праймеров для ОНП2 – ОНП10. Основание «А», выделенное красным полужирным шрифтом, представляет собой дополнительное основание между OLR и последовательностью праймера.

ДНК с высокой молекулярной массой выделяли из лейкоцитарного слоя матери с помощью набора MagAttract HMW DNA Kit (Qiagen). Для каждого анализа получали реакционную смесь для ПЦР в объеме 20 мкл, состоящую из 3,2 нг матрицы ДНК, 10 мкл смеси 2 × ddPCR Supermix for Probes (не содержащей дУТФ) (Bio-Rad), каждый из пары праймеров с внутренним адаптером в конечной концентрации 1 мкмоль/л, обратный праймер для ОНП 1 и прямой праймер для другого ОНП в конечной концентрации 40 нмоль/л и прямой праймер для ОНП 1 и обратный праймер для другого ОНП в конечной концентрации 50 нмоль/л. Затем реакционную смесь для ПЦР разделяли с использованием генератора капель для цифровой ПЦР QX100/QX200 (Bio-Rad). Использовали следующий температурный профиль анализа: инициация при 95°C в течение 10 минут; затем 50 циклов, состоящих из 94°C в течение 45 секунд, 59°C в течение 1 минуты и 72°C в течение 3 минут; и заключительное инкубирование при 98°C в течение 10 минут.

После ПЦР капли диспергировали и выделяли продукты ПЦР в соответствии с протоколом производителя с небольшой поправкой, заключающейся в использовании 60 мкл буфера TE для каждой лунки вместо 20 мкл. Эта поправка привела к лучшему дроблению капель.

Продукты ПЦР дополнительно очищали с использованием гранул AMPure XP Beads в соответствии с протоколом производителя для удаления остаточных праймеров и других нежелательных коротких фрагментов (размером < 100 п.о.). Отношение гранулы/ДНК (т.е. отношение объема добавленных гранул к объему образца) составляло 0,9.

Для получения библиотеки для секвенирования лигировали последовательности внешнего адаптера с гибридными продуктами ПЦР с помощью ПЦР. Праймеры с внешним адаптером включали прямой праймер с последовательностью: 5'-AATGATACGGCGACCACCGAGATCTACACATACGAGATCCGT-3' и обратный праймер с индексом с последовательностью: 5'-CAAGCAGAAGACGGCATACGAGAT-индекс-GTGACTGGAGTTC-3'. Для каждого образца получали реакционную смесь для ПЦР в объеме 50 мкл, состоящую из гибридных продуктов ПЦР, 25 мкл смеси КАРА HiFi HotStart ReadyMix (Roche) и обоих праймеров с внешними адаптерами в конечной концентрации 0,5 мкмоль/л. Температурный профиль анализа представлял собой: инициацию при 98°C в течение 45 секунд; затем 10 циклов, состоящих из 98°C в течение 15 секунд, 60°C в течение 30 секунд и 72°C в течение 30 секунд; и заключительное

инкубирование при 72°C в течение 1 минуты.

Для обогащения желательных последовательностей библиотек длиной от 200 до 300 п.о. выделенные продукты ПЦР дополнительно очищали с использованием гранул AMPure XP Beads в соответствии с протоколом производителя для двойного отбора по размеру на основе гранул. Отношение гранулы/ДНК (т.е. отношение совокупного объема добавленных гранул к объему исходного образца) для первого и второго отбора с помощью гранул составляло 0,6 и 1,4, соответственно. Затем анализировали профиль размера очищенных библиотек с использованием D1000 ScreenTape и реагентов на приборе 4200 TapeStation (Agilent Technologies).

Библиотеки индексов мультиплексировали и секвенировали с использованием платформы секвенирования Illumina MiSeq (76 × 2 цикла). Библиотеку PhiX Control v2 (Illumina) добавляли для контроля качества. Добавление PhiX также могло улучшить качество секвенирования, поскольку оно увеличивало сложность библиотеки, особенно для библиотек с низким разнообразием и с высокой повторяемостью, например, библиотеки гибридной ДНК. Примеры процентного содержания добавленной PhiX включают 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70% и 80%, но ограничиваются ими. В одном варианте реализации добавили 33% PhiX. Авторы получили пользовательские праймеры для секвенирования, используемые для прочтения 1 для цикла анализа на системе Illumina. 2/3 пользовательских праймеров для прочтения 1 смешали с 1/3 праймеров Illumina для прочтения 1. Пользовательские праймеры для прочтения 1 позволяли секвенировать гибридные продукты ПЦР, в то время как праймеры для прочтения 1 Illumina позволяли секвенировать библиотеки PhiX. Последовательность пользовательского праймера для прочтения 1 представляет собой 5'-CATACGAGATCCGTAATCGGGAAGCTGAAG-3'.

Прочтения при секвенировании укорачивали с помощью Cutadapt (Martin et al. *EMBnet. Journal.* 2011;17:10-12) для удаления последовательностей адаптеров и последовательности OLR. Последовательность хромосомы 11 строили для использования в качестве эталона. Для выравнивания прочтений секвенирования на эталон использовали программное обеспечение BWA (Li et al. *Bioinformatics.* 2009;25:1754-1760). Прочтения секвенирования правильно картировали с гибридными последовательностями, т.е. один фрагмент картировали с областью ОНП 1, а другой фрагмент - с одной из девяти других областей ОНП.

Затем определили два гаплотипа матери, как описано выше. Гаплотипы, несущие аллель С и аллель Т в ОНП 1, обозначили как Нар I и Нар II, соответственно. Для Нар I использовали сценарий Python для подсчета фрагментов с различными комбинациями

двух аллелей из двух областей, обозначенных как контиги аллелей. Для каждой пары областей (ОНП) собрали два первых подсчитанных контига аллелей и вычислили их относительные частоты, которые в сумме составляют 100 процентов. Пороговое значение частоты первого контига аллеля для идентификации его как гаплотипа установили как  $> 60\%$ . Если частота первого контига аллеля составляла  $< 60\%$ , гаплотип не определялся. В других вариантах реализации можно использовать другие пороговые процентные значения для идентификации первого контига аллеля в качестве гаплотипа в образце, например,  $> 65\%$ ,  $70\%$ ,  $80\%$ ,  $90\%$ , но не ограничиваясь ими. Аналогичным образом проанализировали прочтения секвенирования для Нар II.

На фиг. 19 показаны фазы, определенные методом цифровой ПЦР с перекрывающимися праймерами. В первом столбце перечислены пары ОНП 1 с другим ОНП. Аллели 10 ОНП гаплотипов перечислены по порядку. Второй и третий столбцы предназначены для Нар I в ОНП 1. Во втором столбце указана частота комбинации аллелей для ОНП 1 и другого ОНП. В третьем столбце перечислены определенные аллели для Нар I. Для каждого сайта аллеля он может содержать значение основания (аллеля), прочерк для сайта, не охваченного спаренными прочтениями, или крест для неопределенного сайта. Четвертый и пятый столбцы предназначены для Нар II в ОНП 1. В четвертом столбце указана частота комбинации аллелей для ОНП 1 и другого ОНП. В пятом столбце перечислены определенные аллели для Нар II.

Как показано на фиг. 19, способ на основе ПЦР с перекрывающимися праймерами физически может правильно фазировать две области, разделенные промежутком до 50 т.п.о. (расстояние от ОНП 8 до ОНП 1) на основе пороговой частоты  $60\%$ . При определении только Нар I фазирование было успешным до 75 т.п.о. (ОНП 1). Точность этого способа определяется качеством и целостностью ДНК. Способ, позволяющий выделять ДНК высокого качества и большего размера, может повысить точность и увеличить расстояние, подлежащее фазированию с помощью этого способа на основе ПЦР с перекрывающимися праймерами.

#### *U. Пример гаплотипирования с использованием гена бета-субъединицы гемоглобина*

Авторы изобретения использовали гаплотипирование вариантов гена *бета-субъединицы гемоглобина (HBB)* в качестве примера для иллюстрирования этого способа гаплотипирования для фазирования нескольких областей. Ген *HBB* кодирует  $\beta$ -субъединицу гемоглобина. Мутации в гене *HBB* могут вызывать  $\beta$ -талассемию - аутосомно-рецессивное заболевание, характеризующееся тяжелой анемией. Для унаследования данного заболевания у индивида должны быть две мутации от обоих

родителей. Lo с соавт. разработали подход к неинвазивному выявлению наследования плодом двух мутаций  $\beta$ -талассемии от обоих родителей (Lo et al. *Sci Transl Med.* 2010;2:61ra91). Авторы получили генотип отца и гаплотип матери. По данным секвенирования ДНК из плазмы матери можно обнаружить мутацию отца, унаследованную плодом. Для оценки наследования мутаций матери плодом авторы использовали анализ относительной дозы гаплотипа (RHDO) с целью выявления наследования плодом гаплотипа матери, содержащего мутацию. Для RHDO-анализа требуется информация о гаплотипе матери. Lo с соавт. получили гаплотип матери из образца ворсинок хориона в этом контрольно-проверочном исследовании. Позже они использовали платформу 10X genomics для определения фазы гаплотипа матери (Hui et al. *Clin Chem.* 2017;63:513-524). Эти способы гаплотипирования могут быть сложными или дорогостоящими. В настоящем документе авторы изобретения продемонстрировали потенциальные варианты применения способа на основе цифровой ПЦР с перекрывающимися праймерами для неинвазивного пренатального выявления моногенных заболеваний путем фазирования гаплотипов в гене *HBB*.

Для проверки точности способа на основе цифровой ПЦР с перекрывающимися праймерами использовали геномную ДНК семейных трио, включающих отца, мать и плод. Авторы изобретения отобрали четыре трио в отделении химической патологии и педиатрии больницы принца Уэльского (PWH), Гонконг. Затем получали лейкоцитарный слой отца, лейкоцитарный слой матери и ткани плаценты плода. Используя способ, проиллюстрированный на фиг. 15А, 15В и 15С, разработали варианты анализа для фазирования 20 ОНП. 20 ОНП разделили на две группы: группа А включала ОНП 1, 4, 6, 8, 10, 11, 13, 15, 17 и 20; а группа В включала ОНП 2, 3, 5, 7, 9, 12, 14, 16, 18 и 19. В других вариантах реализации можно применять другие схемы группировки, например, ОНП 1–10 в качестве группы А и ОНП 11–20 в качестве группы В, и ОНП 1 в качестве группы А и ОНП 2–20 в качестве группы В, но не ограничиваясь ими. Последовательности внутренних праймеров аналогичны праймерам, проиллюстрированным на фиг. 15В. 5'-конец обратных праймеров для области в группе А комплементарен 5'-концу прямых праймеров в группе В.

На фиг. 20 показаны последовательности праймеров для ОНП в гене *HBB*. В первом столбце показан номер области для 20 ОНП в гене *HBB*. Во втором столбце показан идентификатор ОНП. В третьем столбце показана последовательность прямого праймера в направлении от 5' к 3'. В четвертом столбце показана последовательность обратного праймера в направлении от 5' к 3'. Обратный праймер для каждого ОНП в группе А содержит OLR1 (выделена желтым и подчеркнута). OLR1 комплементарна

OLR2 (выделена зеленым и подчеркнута) прямых праймеров для каждого ОНП в группе В. Основание «А» выделено красным полужирным шрифтом и представляет собой дополнительное основание между OLR и последовательностью праймера.

ДНК, выделенную из лейкоцитарного слоя отца, лейкоцитарного слоя матери или плаценты плода, распределяли по каплям для выполнения связывающей ПЦР с использованием системы капельной цифровой ПЦР BioRad Droplet Digital PCR.

Для каждого образца получали реакционную смесь для ПЦР в объеме 20 мкл, состоящую из 3,2 нг матрицы ДНК, 10 мкл смеси  $2 \times$  ddPCR Supermix for Probes (не содержащей дУТФ) (Bio-Rad), каждый из пары праймеров с внутренним адаптером в конечной концентрации 1 мкмоль/л, каждый обратный праймер в группе А и прямой праймер в группе В в конечной концентрации 40 нмоль/л и каждый прямой праймер в группе А и обратный праймер в группе В в конечной концентрации 50 нмоль/л. Затем реакционную смесь для ПЦР разделяли с использованием генератора капель для цифровой ПЦР QX100/QX200 (Bio-Rad). Использовали следующий температурный профиль анализа: инициация при 95°C в течение 10 минут; затем 50 циклов, состоящих из 94°C в течение 45 секунд, 59°C в течение 1 минуты и 72°C в течение 3 минут; и заключительное инкубирование при 98°C в течение 10 минут. Гибридные продукты ПЦР выделяли из капель и лигировали с внешними адаптерами для секвенирования с использованием протокола, описанного в настоящем документе. Проиндексированные библиотеки мультиплексировали и секвенировали на платформе секвенирования Illumina MiSeq. Прочтения секвенирования укорачивали и выравнивали на хромосому 11. Для сборки гаплотипов 20 ОНП использовали сценарий Python.

На фиг. 21А и 21В проиллюстрирован анализ фазовой сборки. В этом примере принцип анализа фазовой сборки продемонстрирован путем фазирования четырех ОНП. На фиг. 21А показаны четыре ОНП для фазирования, разделенные на две группы: группа А содержит ОНП 1 и 3 (серые треугольники), а группа В - ОНП 2 и 4 (белые треугольники).

На фиг. 21В показано определение гаплотипов из гибридных продуктов ПЦР двух областей. На основе данных секвенирования можно связать любой аллель из группы А (серые прямоугольники) с любым аллелем из группы В (белые прямоугольники). Два фазированных аллеля в гибридном продукте позволяют выявить часть гаплотипов. Затем можно определить два гаплотипа с использованием всех фазированных аллелей.

Например, в гибридном продукте ПЦР, содержащем ОНП 1 и ОНП 4, можно определить, что область X802 содержит аллель А ОНП 1, а область X804 содержит аллель Т ОНП 4. В гибридном продукте ПЦР, содержащем ОНП 1 и ОНП 2, можно определить,



что область X806 содержит аллель А ОНП 1, а область X808 содержит аллель G ОНП 2. ОНП 1 одного и того же аллеля А затем используют для связывания ОНП 2 (аллель G) и ОНП 4 (аллель Т). Аналогичный процесс можно использовать для определения ОНП 3.

Гибридный продукт для одного гаплотипа можно отличить от гибридного продукта для другого гаплотипа по несовпадающим аллелям. Например, область X806 содержит аллель А ОНП 1, а область X810 - аллель Т ОНП 1. Кроме того, область X808 содержит аллель G ОНП 2, а область X812 - аллель А ОНП 2. В результате определяют, что гибридный продукт области X806 и области X808 является гаплотипом, отличным от гибридного продукта области X810 и области X812.

На фиг. 22 показана таблица гаплотипов, полученных способом на основе ПЦР с перекрывающимися праймерами для семейных трио. В первом столбце таблицы показаны индивиды, гаплотипы которых определили с помощью ПЦР с перекрывающимися праймерами. В таблице показаны четыре семьи из отца, матери и плода. Во втором столбце показана метка гаплотипа. В остальных столбцах показаны аллели 20 ОНП.

Гаплотипы каждого индивида определяли способом на основе ПЦР с перекрывающимися праймерами. Гаплотипы, определенные этим способом, полностью совпадали с гаплотипами, определенными на основании генотипов семейного трио, состоящего из отца, матери и ребенка. Сопоставление гаплотипов, определенных с помощью анализа генотипов, с определением с помощью ПЦР с перекрывающимися праймерами, подтверждает точность способа на основе ПЦР с перекрывающимися праймерами. Для каждой семьи гаплотип, выделенный синим цветом, представляет гаплотип, переданный ребенку от отца, а гаплотип, выделенный красным, представляет гаплотип, переданный ребенку от матери. Родительский гаплотип, переданный потомку, также обозначен звездочкой во втором столбце.

#### Способ

Фиг. 23 представляет собой блок-схему, иллюстрирующую способ соединения разделенных областей молекул ДНК в биологическом образце субъекта согласно вариантам реализации настоящего изобретения.

В блоке 2310 молекулы ДНК из биологического образца разделяют на множество компартментов (например, капель). Разделение может быть таким, что компартменты обычно включают не более одной молекулы ДНК. В некоторых компартментах может быть менее одной молекулы ДНК. Например, компартменты могут представлять собой капли. В качестве примеров можно привести существующие устройства, доступные для приобретения, которые могут автоматически генерировать тысячи капель нанолитрового размера за короткое время, например, BioRad QX200 Droplet Generator, Elveflow Droplet

Generator Pack и Micronit Microfluidic Droplet Generator, но не ограничиваясь ими. В компартменты можно добавлять различные реагенты (например, праймеры). Другие компартменты включают пробирку Eppendorf, лунку планшета для ПЦР или любой другой контейнер.

В наборе компартментов, содержащих молекулу ДНК, каждый компартмент может содержать определенные реагенты. Например, первый компартмент из множества компартментов может содержать: (1) первую молекулу ДНК, содержащую область А и область В, разделенные промежуточной областью, (2) полимеразу, характеризующуюся склонностью к добавлению определенного нуклеотида в положение в липком конце, (3) обратный праймер R1, содержащий первый фрагмент, комплементарный концевой последовательности области А, обратный праймер, содержащий второй фрагмент, содержащий перекрывающуюся последовательность, и (4) прямой праймер F2, содержащий первый фрагмент, комплементарный последовательности начала области В, прямой праймер, содержащий комплементарную перекрывающуюся последовательность, комплементарную указанной перекрывающейся последовательности. Компартмент также может содержать прямой праймер F1 в начале области А и обратный праймер R2 в конце области В.

Праймеры гибридизуются с 3'-концом одной цепи для ее удлинения до 3'-конца другой цепи. Прямой праймер гибридизуется с одной из цепей Крика или Уотсона на 3'-конце для ее удлинения до 3'-конца другой цепи. Если прямой праймер гибридизуется с цепью Крика, обратный праймер гибридизуется с 3'-концом цепи Уотсона для ее удлинения до 3'-конца цепи Крика. Последовательность начала области находится на 5'-конце цепи Уотсона. Концевая последовательность области находится на 3'-конце цепи Уотсона. Аналогичным образом, если прямой праймер гибридизуется с цепью Уотсона, обратный праймер гибридизуется с 3'-концом цепи Крика для ее удлинения до 3'-конца цепи Уотсона. Последовательность начала области находится на 5'-конце цепи Крика. Концевая последовательность области находится на 3'-конце цепи Крика.

В качестве примеров, длина промежуточной области может быть меньше или равна 100 т.п.о., включая 2,5 т.п.о., 5 т.п.о., 10 т.п.о., 20 т.п.о., 30 т.п.о., 40 т.п.о., 50 т.п.о., 75 т.п.о. или любому диапазону между этими величинами.

Первая молекула ДНК может содержать области, показанные на фиг. 3А или других аналогичных фигурах. На фиг. 13А показана область А, представляющая собой экзон 1 гена NUDT15, и область В, представляющая собой экзон 3 гена NUDT15.

Как объяснялось выше, полимеразы могут характеризоваться склонностью к добавлению определенного нуклеотида (например, А) в положение в липком конце.

Липкий конец - это конец двуцепочечной ДНК, содержащей по меньшей мере один крайний нуклеотид, не гибридизованный с другой цепью. Липкие концы также называют свесами или выступами. Положение в липком конце - это положение, которое не гибридизуется с другой цепью и может соответствовать первому из таких положений. А 319 или 329 на фиг. 3С является примером. Другие полимеразы могут характеризоваться склонностью к добавлению других нуклеотидов, кроме А, и в вариантах реализации также можно использовать такие полимеразы. В некоторых вариантах реализации, включая описанные ниже, полимеразы не обязательно должны характеризоваться склонностью к добавлению определенного нуклеотида.

Например, обратный праймер R1 может соответствовать праймерам, обозначенным как R5 на фиг. 3А, 4А и 5А, или обозначенным как R1 на фиг. 8А, 9А и 13А. Например, первый фрагмент может соответствовать последовательности праймера, например, 403 на фиг. 4В. Например, второй фрагмент, содержащий перекрывающуюся последовательность, может соответствовать OLR 310, 410, 510 и другим аналогичным фрагментам, описанным в настоящем документе. Обратный праймер R1 может содержать определенный нуклеотид (например, А 409 или 509) между первым фрагментом и вторым фрагментом. В некоторых вариантах реализации обратный праймер R1 может не содержать указанный определенный нуклеотид, если в способе используют полимеразу, не добавляющую дополнительный аденозин к 3'-концу цепи ДНК, которую она удлиняет.

Например, прямой праймер F2 может соответствовать праймеру, обозначенному как F6 на фиг. 3А, 4А и 5А, или обозначенному как F2 на фиг. 8А, 9А и 13А. Например, первый фрагмент может соответствовать последовательности праймера, например, 553 на фиг. 5В. Например, комплементарная перекрывающаяся последовательность может представлять собой первый фрагмент (например, соответствующий началу области В), например, F6 на фиг. 3А и 4А или F2 на фиг. 8А. Соответственно, первый фрагмент прямого праймера может содержать комплементарную перекрывающуюся последовательность.

В качестве еще одного примера, комплементарная перекрывающаяся последовательность может соответствовать второму фрагменту прямого праймера F2, например, OLR 550 на фиг. 5В и другим аналогичным фрагментам, описанным в настоящем документе. Соответственно, прямой праймер F2 может содержать второй фрагмент, содержащий комплементарную перекрывающуюся последовательность. Прямой праймер F2 может содержать определенный нуклеотид (например, А) между первым фрагментом и вторым фрагментом, например, как изображено на фиг. 5В. В некоторых вариантах реализации прямой праймер F2 может не содержать указанный

определенный нуклеотид, если в способе используют полимеразу, не добавляющую дополнительный аденозин к цепи ДНК, которую она удлиняет.

В блоке 2320 полимеразу и обратный праймер R1 используют для амплификации области А с целью получения первого набора ампликонов, содержащих определенный нуклеотид (например, А 319 в цепи X) на первом 3'-конце и содержащих комплементарный нуклеотид (например, Т 513 в цепи X) между участками последовательностей, соответствующими первому фрагменту и второму фрагменту обратного праймера R1. Указанный комплементарный нуклеотид комплементарен указанному определенному нуклеотиду. Один из участков последовательности может являться фрагментом комплементарной перекрывающейся последовательности, например, OLR на фиг. 5С. В некоторых вариантах реализации, если полимеразу не добавляет дополнительное основание (например, аденозин) в цепь ДНК, которую она удлиняет, дополнительные нуклеотиды, расположенные на 3'- и 5'-концах OLR, отсутствуют или не требуются.

Комплементарный нуклеотид может входить или не входить в состав первой молекулы ДНК в молекуле ДНК, например, как изображено на фиг. 3А-3F (входит) и фиг. 4А-4С (не входит). Соответственно, комплементарный нуклеотид в первом наборе ампликонов может находиться в первой молекуле ДНК (например, в концевом положении липкого конца).

В блоке 2330 полимеразу и прямой праймер F2 используют для амплификации области В с целью получения второго набора ампликонов, содержащих определенный нуклеотид (например, А 329 в цепи Z на фиг. 3С) на втором 3'-конце и содержащих комплементарный нуклеотид (например, Т 563 в цепи Z) на противоположном конце перекрывающейся последовательности. В некоторых вариантах реализации второй набор ампликонов может не содержать указанный определенный нуклеотид и комплементарный нуклеотид.

В блоке 2340 полимеразу используют для удлинения первого набора ампликонов с целью включения области В и второго набора ампликонов с целью включения области А. Таким образом, можно получать удлиненные ампликоны, содержащие область А и область В и не содержащие промежуточной области. Прямой праймер F1 и обратный праймер R2 можно использовать для амплификации удлиненных ампликонов.

Удлинение можно выполнять в том же компартменте или в рамках отдельного анализа. При выполнении в том же компартменте, как описано выше, концентрации F1 и R2 могут превышать концентрации F2 и R1, так что в конечном итоге большинство продуктов должны являться удлиненными ампликонами.

В блоке 2350 удлиненные ампликоны анализируют для определения гаплотипов. Для анализа удлиненных ампликонов можно выполнять любое количество подходящих анализов. Например, можно выполнить секвенирование, например, с использованием систем секвенирования Illumina (например, HiSeq2500, NextSeq 550 и MiSeq), секвенатора Pacific Biosystems и системы секвенирования Ion Proton, но не ограничиваясь этим. В качестве еще одного примера можно выполнить аллель-специфическую ПЦР для обнаружения по меньшей мере одного аллеля в области А и по меньшей мере еще одного аллеля в области В. Независимо от конкретной методики, можно обнаружить первый аллель в области А, а второй аллель в области В, что означает, что указанные первый аллель и второй аллель относятся к одному и тому же гаплотипу.

Аллель (например, первый аллель в области А или второй аллель в области В) можно обнаружить путем измерения количества аллеля в соответствующей области. Количество может представлять собой концентрацию, процент, долю или счет. Указанное количество можно сравнить с пороговым значением. Если указанное количество больше порогового значения, аллель можно определить как подлежащий обнаружению. Пороговое значение может представлять собой 50%, 60%, 70%, 80%, 90%, 95% или эквивалентные доли.

В некоторых вариантах реализации удлиненные ампликоны можно анализировать для определения наличия амплификации или делеции. Удлиненные ампликоны можно секвенировать для получения прочтений секвенирования. Прочтения секвенирования можно выровнять на эталонный геном. Амплификации и делеции можно выявлять по несовпадению прочтения секвенирования и эталонного генома.

В некоторых вариантах реализации к прямому праймеру F2 и обратному праймеру R1 можно добавить молекулу, что позволяет захватить и удалить несвязанные (неудлиненные) ампликоны.

#### *V. Праймеры для амплификации гибридных продуктов*

Способ может включать применение прямого праймера Fz и обратного праймера Rz для амплификации гибридных продуктов. Примеры прямого праймера Fz и обратного праймера Rz включают F3 и R3 на фиг. 6, R4 и F4 на фиг. 10А, или Fz или Rz на фиг. 15В). Прямой праймер F1 может содержать второй фрагмент, который не является первым фрагментом. Первый компартмент может дополнительно содержать прямой праймер Fz, содержащий первый фрагмент, идентичный второму фрагменту прямого праймера F1. Прямой праймер Fz также может содержать второй фрагмент. Прямой праймер Fz может содержать комплементарный нуклеотид. Обратный праймер R2 может содержать второй фрагмент, который не является первым фрагментом. Первый компартмент также может

дополнительно содержать обратный праймер R<sub>z</sub>, содержащий первый фрагмент, идентичный второму фрагменту обратного праймера R<sub>2</sub>. Обратный праймер R<sub>z</sub> может содержать второй фрагмент. Обратный праймер R<sub>z</sub> может содержать комплементарный нуклеотид.

Гибридные продукты, образованные с использованием прямого праймера F<sub>1</sub> и обратного праймера R<sub>2</sub>, можно амплифицировать с использованием прямого праймера F<sub>z</sub> и обратного праймера R<sub>z</sub>. Удлиненные ампликоны, образованные с использованием прямого праймера F<sub>1</sub> и обратного праймера R<sub>2</sub>, могут являться первыми удлиненными ампликонами. Первые удлиненные ампликоны могут содержать второй фрагмент прямого праймера F<sub>1</sub> и второй фрагмент обратного праймера R<sub>2</sub>. Способ может включать амплификацию первых удлиненных ампликонов с использованием полимеразы прямого праймера F<sub>z</sub> и обратного праймера R<sub>z</sub> для получения вторых удлиненных ампликонов. Каждый второй удлиненный ампликон может содержать второй фрагмент прямого праймера F<sub>z</sub> и второй фрагмент обратного праймера R<sub>z</sub>.

В вариантах реализации с использованием прямого праймера F<sub>z</sub> и обратного праймера R<sub>z</sub> для амплификации гибридных продуктов полимеразы может не характеризоваться склонностью добавлять определенный нуклеотид. Соответственно, указанные прямой и обратный праймеры (F<sub>z</sub> и R<sub>z</sub>) могут не содержать нуклеотид, комплементарный указанному определенному нуклеотиду, в конце перекрывающейся последовательности.

#### *W. Объединение трех областей*

В некоторых вариантах реализации можно использовать способы объединения трех областей первой молекулы ДНК. Например, три области можно объединить, как показано на фиг. 8А, 8В, 10А или 10В. На этих фигурах области обозначены как область 1, область 2 и область 3, однако описаны варианты реализации, которые относятся к областям А, области В и области С. Обозначение областей числом или буквой не влияет на указанные способы. Области могут представлять собой область А (например, обозначенную как область 1 на некоторых фигурах), область В (например, обозначенную как область 2 на некоторых фигурах) и область С (например, обозначенную как область 3 на некоторых фигурах). Промежуточная область между областью А и областью В может являться первой промежуточной областью. Область С может быть отделена от области В второй промежуточной областью. Перекрывающаяся последовательность второго фрагмента обратного праймера R<sub>1</sub> может являться первой перекрывающейся последовательностью. Комплементарная перекрывающаяся последовательность прямого праймера F<sub>2</sub> может являться первой комплементарной перекрывающейся последовательностью.

Первый компартмент также может содержать обратный праймер R2, содержащий первый фрагмент, комплементарный концевой последовательности области В. Обратный праймер R2 может содержать второй фрагмент, содержащий вторую перекрывающуюся последовательность. Первый компартмент также может содержать прямой праймер F3, содержащий первый фрагмент, комплементарный последовательности начала области С. Прямой праймер F3 может содержать вторую комплементарную перекрывающуюся последовательность, комплементарную указанной второй перекрывающейся последовательности. Вторая комплементарная перекрывающаяся последовательность может находиться в первом фрагменте прямого праймера F3 или во втором фрагменте прямого праймера F3.

Способ может включать амплификацию области С с использованием полимеразы и прямого праймера F3 для получения третьего набора ампликонов, содержащих определенный нуклеотид на третьем 3'-конце и содержащих комплементарный нуклеотид на противоположном конце указанной второй перекрывающейся последовательности. Амплификация области В для получения второго набора ампликонов может включать использование обратного праймера R2.

Способ может дополнительно включать удлинение трех наборов ампликонов с использованием полимеразы. Первые удлиненные ампликоны (содержащие область А и область В) можно удлинить с целью включения области С. Третий набор ампликонов можно удлинить с целью включения области А и области В. В результате удлинения можно получить вторые удлиненные ампликоны, содержащие область А, область В и область С и не содержащие первую промежуточную область и вторую промежуточную область.

Удлинение первых удлиненных ампликонов с целью включения области С может включать гибридизацию со второй перекрывающейся последовательностью по направлению к четвертому 3'-концу (в концевой последовательности области С). Удлинение третьего набора ампликонов с целью включения области А и области В может включать гибридизацию со второй комплементарной перекрывающейся последовательностью по направлению ко второму 3'-концу (в последовательности начала области В).

В вариантах реализации, где происходит объединение трех или более областей, полимеразы может не характеризоваться склонностью к добавлению определенного нуклеотида. Аналогичным образом, обратные праймеры могут содержать указанный определенный нуклеотид в конце указанной перекрывающейся последовательности, а прямые праймеры могут не содержать нуклеотид, комплементарный указанному

определенному нуклеотиду, в конце комплементарной перекрывающейся последовательности.

#### *X. Объединение областей из разных групп*

Способ также может включать образование гибридных продуктов ПЦР, не содержащих все области в одном гибридном продукте. Например, можно объединить три области, однако указанные области можно объединить в два гибридных продукта, каждый из которых содержит две области (например, на фиг. 14). Например, первая молекула ДНК может содержать область D (например, область 4 на фиг. 14). Область D может быть отделена от области A (например, области 1 на фиг. 14) второй промежуточной областью. Первый компартмент может дополнительно содержать прямой праймер F4, содержащий первый фрагмент, комплементарный последовательности начала области D. Прямой праймер F4 может содержать второй фрагмент, содержащий комплементарную перекрывающуюся последовательность, которая также находится в прямом праймере F2. Первый компартмент также может содержать обратный праймер R4, содержащий первый фрагмент, комплементарный концевой последовательности области D.

Способ может включать амплификацию области D с использованием полимеразы и прямого праймера F4 с целью получения третьего набора ампликонов. Третий набор ампликонов может содержать определенный нуклеотид на третьем 3'-конце и комплементарный нуклеотид на противоположном конце перекрывающейся последовательности. С использованием полимеразы можно удлинить первый набор ампликонов с целью включения области D. С использованием полимеразы можно удлинить третий набор ампликонов с целью включения области A. Можно получить вторые удлиненные ампликоны, содержащие область A и область D. Вторые удлиненные ампликоны могут не содержать вторую промежуточную область. Первые удлиненные ампликоны представляют собой удлиненные ампликоны, содержащие область A и область B (например, область 2 на фиг. 14).

В некоторых вариантах реализации способы могут включать связывание областей из трех различных групп, например, на фиг. 15A. Области обозначены буквами, но могут иметь. В примере связывания 4 областей область A является областью из первой группы (например, область 1 на фиг. 15A). Область B (например, область 2 на иллюстрации) и область D (например, область 5 на иллюстрации) являются областями из второй группы. Область C (например, область 3 на иллюстрации) является областью из третьей группы. Область A (область 1) можно соединить с областью B (областью 2) или с областью D (областью 5). Область B (область 2) или область D (область 5) можно соединить с областью C (областью 3). Комплементарная последовательность в прямом праймере F2



может являться первой комплементарной перекрывающейся последовательностью. Перекрывающаяся последовательность в обратном праймере R1 может являться первой перекрывающейся последовательностью. Первый компартмент может дополнительно содержать обратный праймер R2, содержащий первый фрагмент, комплементарный концевой последовательности области В (области 2). Обратный праймер R2 может содержать второй фрагмент, содержащий вторую перекрывающуюся последовательность. Первый компартмент также может содержать обратный праймер R4, содержащий первый фрагмент, комплементарный концевой последовательности области D (области 5). Обратный праймер R4 может содержать второй фрагмент, содержащий вторую перекрывающуюся последовательность. Первый компартмент может дополнительно содержать прямой праймер F3, содержащий первый фрагмент, комплементарный последовательности начала области С (области 3) первой молекулы ДНК. Область С (область 3) может быть отделена от области В (область 2) третьей промежуточной областью. Первая промежуточная область или вторая промежуточная область может содержать область С (область 3). Прямой праймер F3 может содержать второй фрагмент, содержащий вторую комплементарную перекрывающуюся последовательность.

Аmplификация области В (область 2) может включать применение обратного праймера R2. Амплификация области D (область 5) может включать применение обратного праймера R4. Способ может дополнительно включать амплификацию области С (области 3) с использованием полимеразы и прямого праймера F3 с целью получения четвертого набора ампликонов. Способ может дополнительно включать удлинение вторых удлиненных ампликонов с использованием полимеразы с целью включения области С (области 3) с образованием третьих удлиненных ампликонов, содержащих область А (область 1), область D (область 5) и область С (область 3). Способ может также включать удлинение первых удлиненных ампликонов с использованием полимеразы с целью включения области С (области 3) с образованием четвертых удлиненных ампликонов, содержащих область А (область 1), область В (область 2) и область С (область 3).

В некоторых вариантах реализации можно соединить две области, выбирая первую область из первой группы и вторую область из второй группы, например, на фиг. 14А. Например, первая группа может содержать область А (т.е. область 1 на иллюстрации) и область С (т.е. область 3 на иллюстрации). Вторая группа может содержать область В (т.е. область 2 на иллюстрации) и область D (т.е. область 4 на иллюстрации). Как разъяснялось выше, можно объединить область А (область 1) и область В (область 2), и можно объединить область А (область 1) и область D (область 4). Кроме того, можно объединить область С (область 3) и область В (область 2), и можно объединить область С (область 3) и

область D (область 4). Первый компартмент может содержать обратный праймер R3, содержащий первый фрагмент, комплементарный концевой последовательности области C (области 3) первой молекулы ДНК. Область C (область 3) может быть отделена от области B (область 2) третьей промежуточной областью. Область C (область 3) может быть отделена от области D (область 4) четвертой промежуточной областью. Обратный праймер R3 может содержать второй фрагмент, содержащий перекрывающуюся последовательность, которая является такой же перекрывающейся последовательностью в обратном праймере R1.

Способ может включать амплификацию области C (области 3) с использованием полимеразы и обратного праймера R3 с целью получения четвертого набора ампликонов, содержащих определенный нуклеотид на четвертом 3'-конце и комплементарный нуклеотид на противоположном конце перекрывающейся последовательности. Способ также может включать удлинение четвертого набора ампликонов с использованием полимеразы с целью включения области B (область 2), а также второго набора ампликонов с целью включения области C (области 3), тем самым получая третьи удлиненные ампликоны, содержащие область C (область 3) и область B (область 2) и не содержащие третью промежуточную область. Способ может дополнительно включать удлинение четвертого набора ампликонов с использованием полимеразы с целью включения области D (область 4), а также третьего набора ампликонов с целью включения области C (области 3), тем самым получая третьи удлиненные ампликоны, содержащие область C (область 3) и область D (область 4) и не содержащие четвертую промежуточную область.

В вариантах реализации, где используют различные комбинации по меньшей мере двух областей в одном компартменте, полимеразы могут не характеризоваться склонностью к добавлению определенного нуклеотида. Соответственно, обратные праймеры могут содержать указанный определенный нуклеотид в конце указанной перекрывающейся последовательности, а прямые праймеры могут не содержать нуклеотид, комплементарный указанному определенному нуклеотиду, в конце комплементарной перекрывающейся последовательности. Прямой и обратный праймеры Fz и Rz могут не содержать указанный определенный нуклеотид в конце перекрывающейся последовательности.

Анализ фазовой сборки, например, описанный на фиг. 21B, можно выполнять на первых удлиненных ампликонах и вторых удлиненных ампликонах. Первый аллель можно обнаружить в области A (например, области 1 на фиг. 21B) в первых удлиненных ампликонах. Второй аллель можно обнаружить в области B (например, области 2 на фиг. 21B) в первых удлиненных ампликонах. Первый аллель также можно обнаружить в

области А во вторых удлинённых ампликонах. Третий аллель можно обнаружить в области D (например, в области 4 на фиг. 21В) во вторых удлинённых ампликонах. В результате обнаружения первого аллеля в области А обоих удлинённых ампликонов можно обнаружить первый аллель в области А, второй аллель в области В и третий аллель в области D в одном и том же гаплотипе.

В некоторых вариантах реализации первый аллель может не обнаруживаться в области А во вторых удлинённых ампликонах. Вместо этого в области А во вторых удлинённых ампликонах можно обнаружить четвёртый аллель. В этих случаях можно определить, что третий аллель в области D не принадлежит к тому же гаплотипу, что и первый аллель в области А.

Вышеупомянутый процесс можно повторить для каждого из множества компартментов с целью определения гаплотипа в каждом из компартментов, содержащих молекулу ДНК.

#### Наборы

Варианты реализации могут включать наборы для соединения разделённых областей молекул ДНК в биологическом образце субъекта. Набор может содержать полимеразу, характеризующуюся склонностью к добавлению определённого нуклеотида в положение в липком конце. В некоторых вариантах реализации полимеразы может не характеризоваться склонностью к добавлению определённого нуклеотида в положение в липком конце. Варианты реализации включают наборы для применения прямого праймера Fz и обратного праймера Rz для амплификации гибридных продуктов, наборы для объединения трёх или более областей и наборы для объединения различных комбинаций по меньшей мере двух областей в одном и том же компартменте, аналогично способам, описанным выше.

Набор также может включать обратный праймер R1, содержащий первый фрагмент, комплементарный концевой последовательности области А первой молекулы ДНК. Обратный праймер R1 может содержать второй фрагмент, содержащий перекрывающуюся последовательность. Указанный определённый нуклеотид, к добавлению которого полимеразы характеризуется склонностью, может находиться между первым фрагментом и вторым фрагментом.

Набор может дополнительно содержать прямой праймер F2, содержащий первый фрагмент, комплементарный последовательности начала области В первой молекулы ДНК. Область В может быть отделена от области А промежуточной областью. Прямой праймер F2 может содержать комплементарную перекрывающуюся последовательность, комплементарную указанной перекрывающейся последовательности обратного праймера

R1. Комплементарная перекрывающаяся последовательность может находиться во втором фрагменте прямого праймера F2, причем указанный второй фрагмент не является первым фрагментом. В некоторых вариантах реализации первый фрагмент прямого праймера F2 может содержать комплементарную перекрывающуюся последовательность.

Набор может также содержать прямой праймер F1, содержащий первый фрагмент, комплементарный последовательности начала области А. Прямой праймер F1 может содержать второй фрагмент, не являющийся первым фрагментом.

Набор может дополнительно содержать обратный праймер R2, содержащий первый фрагмент, комплементарный концевой последовательности области В. Обратный праймер R может содержать второй фрагмент, не являющийся первым фрагментом.

Набор может содержать прямой праймер Fz и обратный праймер Rz, которые можно использовать для амплификации гибридных продуктов. Прямой праймер Fz может содержать первый фрагмент, идентичный второму фрагменту прямого праймера F1. Обратный праймер Rz может содержать первый фрагмент, идентичный второму фрагменту обратного праймера R2. Прямой праймер Fz может содержать нуклеотид, комплементарный указанному определенному нуклеотиду. Обратный праймер Rz может содержать нуклеотид, комплементарный указанному определенному нуклеотиду.

Отношение прямого праймера Fz к прямому праймеру F1 или любому праймеру, кроме обратного праймера Rz, может составлять по меньшей мере 5:1. В некоторых вариантах реализации указанное соотношение может составлять по меньшей мере 10:1, 20:1, 50:1, 75:1, 100:1, 200:1, 500:1, 1000:1, 2000:1, 5000:1, 10000:1, 20000:1, 50000:1 или 100000:1. Отношение обратного праймера Rz к обратному праймеру R2 или любому праймеру, кроме прямого праймера Fz, может составлять по меньшей мере 5:1, 10:1, 20:1, 50:1, 75:1, 100:1, 200:1, 500:1, 1000:1, 2000:1, 5000:1, 10000:1, 20000:1, 50000:1 или 100000:1.

Каждый из прямого праймера Fz и обратного праймера Rz может содержать второй фрагмент. Второй фрагмент может быть предназначен для секвенирования. Например, второй фрагмент может являться индексом образца или последовательностью адаптера для конкретной платформы секвенирования.

В некоторых вариантах реализации набор может дополнительно содержать обратный праймер R2, содержащий первый фрагмент, комплементарный концевой последовательности области В. Обратный праймер R2 может содержать второй фрагмент, содержащий вторую перекрывающуюся последовательность. Набор также может содержать прямой праймер F3, содержащий первый фрагмент, комплементарный последовательности начала области С первой молекулы ДНК. Область С может быть

отделена от области В второй промежуточной областью. Прямой праймер F3 может содержать вторую комплементарную перекрывающуюся последовательность, комплементарную указанной второй перекрывающейся последовательности обратного праймера R2. Вторая перекрывающаяся последовательность может не являться перекрывающейся последовательностью в обратном праймере R1. В некоторых вариантах реализации вторая перекрывающаяся последовательность может являться перекрывающейся последовательностью в обратном праймере R1.

В некоторых вариантах реализации набор может содержать обратный праймер R3. Обратный праймер R3 может содержать первый фрагмент, комплементарный концевой последовательности области С. Обратный праймер R3 может содержать второй фрагмент, не являющийся первым фрагментом. В области С обратный праймер R3 может содержать первый фрагмент, идентичный второму фрагменту обратного праймера R3.

Перекрывающиеся последовательности в праймерах могут состоять по меньшей мере из 12 оснований, включая 12-15, 15-20, 20-25 или более 25 оснований. Фрагменты праймеров, комплементарные последовательности области, могут состоять по меньшей мере из 12 оснований, включая 12-15, 15-20, 20-25 или более 25 оснований.

Фрагменты праймеров, комплементарные последовательности области, могут находиться на 3'-конце праймера. Фрагменты праймеров, которые являются перекрывающимися последовательностями или комплементарны перекрывающимся последовательностям, могут находиться на 5'-конце праймера. Прямой праймер Fz и обратный праймер Rz могут содержать фрагменты, идентичные прямому праймеру F1, а обратный праймер R2 может находиться на 3'-конце прямого праймера Fz и обратного праймера Rz.

Набор может содержать первую молекулу ДНК. Набор может содержать реагенты для ПЦР, микрожидкостной картридж для создания отдельных компартментов для ПЦР и праймеры для реакций объединения.

#### Примеры систем

На фиг. 24 проиллюстрирована система измерения 2400 согласно варианту реализации настоящего изобретения. Показанная система включает образец 2405, например, молекулы ДНК без клеток, в держателе образца 2410, где образец 2405 может контактировать с системой анализа 2408 для получения сигнала физической характеристики 2415. Примером держателя образца может быть проточная кювета, которая содержит зонды и/или праймеры для анализа, или пробирка, через которую движется капля (с каплей, содержащей систему анализа). Физическую характеристику 2415 (например, интенсивность флуоресценции, напряжение или ток) образца определяют

с помощью детектора 2420. Детектор 2420 может выполнять измерения с интервалами (например, периодическими интервалами), получая значения данных, которые составляют сигнал данных. В одном варианте реализации аналого-цифровой преобразователь несколько раз преобразует аналоговый сигнал от детектора в цифровую форму. Держатель образца 2410 и детектор 2420 могут формировать устройство анализа, например, устройство секвенирования, которое выполняет секвенирование согласно вариантам реализации, описанным в настоящем документе. Сигнал данных 2425 отправляется из детектора 2420 к логической системе 2430. Сигнал данных 2425 может храниться в локальной памяти 2435, внешней памяти 2440 или запоминающем устройстве 2445.

Логическая система 2430 может представлять собой или содержать компьютерную систему, ASIC, микропроцессор и т.д. Она также может содержать или быть соединена с дисплеем (например, монитором, светодиодным дисплеем и т.д.) и устройством ввода пользователя (например, мышью, клавиатурой, кнопками и т.д.). Логическая система 2430 и другие компоненты могут входить в состав автономной или подключенной к сети компьютерной системы, или быть непосредственно присоединены к или встроены в устройство (например, устройство секвенирования), которое включает детектор 2420 и/или держатель образца 2410. Логическая система 2430 также может содержать программное обеспечение, которое выполняется в процессоре 2450. Логическая система 2430 может содержать машиночитаемый носитель, хранящий инструкции для управления системой 2400 в целях реализации любого из способов, описанных в настоящем документе. Например, логическая система 2430 может подавать команды в систему, которая включает держатель образца 2410, так что выполняется секвенирование или другие физические операции. Такие физические операции могут выполняться в определенном порядке, например, с добавлением и удалением реагентов в определенном порядке. Такие физические операции может выполнять робототехническая система, например, включая роботизированную руку, которую можно использовать для получения образца и выполнения анализа.

Любая из упомянутых в настоящем документе компьютерных систем может использовать любое подходящее количество подсистем. Примеры таких подсистем показаны на фиг. 25 в компьютерной системе 2510. В некоторых вариантах реализации компьютерная система включает одно компьютерное устройство, причем подсистемы могут быть компонентами компьютерного устройства. В других вариантах реализации компьютерная система может содержать множество компьютерных устройств, каждое из которых является подсистемой и содержит внутренние компоненты. Компьютерная система может включать настольные и портативные компьютеры, планшеты, мобильные

телефоны и другие мобильные устройства.

Подсистемы, показанные на фиг. 25, соединены системной шиной 2575. Показаны дополнительные подсистемы, например, принтер 2574, клавиатура 2578, запоминающее устройство(а) 2579, монитор 2576 (например, экран дисплея, например, светодиодный дисплей), соединенный с адаптером дисплея 2582, и т.д. Периферийные устройства и устройства ввода/вывода (ВВОД/ВЫВОД), соединенные с контроллером 71 ввода/вывода, можно подключать к компьютерной системе с помощью любого количества средств, известных в данной области техники, например, порта ввода/вывода (ВВОД/ВЫВОД) 2577 (например, USB, FireWire®). Например, порт ввода/вывода 2577 или внешний интерфейс 2581 (например, Ethernet, Wi-Fi и т.д.) можно использовать для подключения компьютерной системы 2510 к глобальной сети, например, Интернет, устройству ввода - мыши, или сканеру. Межкомпонентное соединение через системную шину 2575 позволяет центральному процессору 2573 обмениваться данными с каждой подсистемой и контролировать выполнение множества инструкций из системной памяти 2572 или запоминающего устройства(ств) 2579 (например, несъемного диска, например, жесткого диска, или оптического диска), а также обмен информацией между подсистемами. Системная память 2572 и/или запоминающее устройство(а) 2579 может представлять собой машиночитаемый носитель. Еще одной подсистемой является устройство сбора данных 2585, например, камера, микрофон, акселерометр и т.п. Любые из упомянутых в настоящем документе данных могут выводиться из одного компонента в другой компонент и могут выводиться пользователю.

Компьютерная система может содержать множество одинаковых компонентов или подсистем, например, связанных друг с другом с помощью внешнего интерфейса 2581, внутреннего интерфейса или через съемные запоминающие устройства, которые можно подключать и переносить от одного компонента к другому компоненту. В некоторых вариантах реализации компьютерные системы, подсистемы или устройства могут обмениваться данными по сети. В таких случаях один компьютер может считаться клиентом, а другой компьютер - сервером, причем каждый из них может быть частью одной и той же компьютерной системы. Каждый клиент и сервер могут содержать множество систем, подсистем или компонентов.

Аспекты вариантов реализации можно реализовать в форме логики управления с использованием аппаратных схем (например, специализированной интегральной схемы или программируемой матрицы логических элементов) и/или с использованием компьютерного программного обеспечения с, в целом, программируемым процессором в модульном или интегрированном виде. В контексте настоящего изобретения процессор

может включать одноядерный процессор, многоядерный процессор на одном интегрированном кристалле, или множество процессорных блоков на одной печатной плате или в сети, а также специализированное оборудование. На основе изобретения и идей, представленных в настоящем документе, специалист в данной области техники должен узнать о других путях и/или способах реализации вариантов реализации данного изобретения с использованием аппаратных средств и комбинации аппаратных средств и программного обеспечения и принимать их во внимание.

Любой из программных компонентов или функций, описанных в данной заявке, можно реализовать в виде программного кода, который должен выполняться процессором с использованием любого подходящего языка программирования, например, Java, C, C++, C#, Objective-C, Swift или скриптового языка программирования, например, Perl или Python, с использованием, например, стандартных или объектно-ориентированных способов. Программный код может храниться в виде серии инструкций или команд на машиночитаемом носителе для хранения и/или передачи. Подходящий энергонезависимый машиночитаемый носитель может включать оперативную память (ОЗУ), постоянное запоминающее устройство (ПЗУ), магнитный носитель, например, жесткий диск или гибкий диск, или оптический носитель, например, компакт-диск (CD) или DVD (цифровой универсальный диск), или диск Blu-ray, флэш-память и т.п. Машиночитаемый носитель может быть любой комбинацией таких устройств хранения или передачи.

Такие программы также можно кодировать и передавать с использованием несущих сигналов, адаптированных для передачи через проводные, оптические и/или беспроводные сети, соответствующих множеству протоколов, включая Интернет. По существу, машиночитаемый носитель можно создать с использованием сигнала данных, закодированного с помощью таких программ. Машиночитаемый носитель, закодированный программным кодом, можно упаковать с совместимым устройством или представить отдельно от других устройств (например, путем загрузки через Интернет). Любой такой машиночитаемый носитель может поставляться с или в составе компьютерного продукта (например, жесткого диска, компакт-диска или всей компьютерной системой), а также с или в составе различных компьютерных продуктов в пределах системы или сети. Компьютерная система может содержать монитор, принтер или другой подходящий дисплей для предоставления пользователю любых результатов, упомянутых в настоящем документе.

Любой из описанных в настоящем документе способов может быть полностью или частично выполнен компьютерной системой, содержащей один или более процессоров,



которые можно сконфигурировать для выполнения поэтапных действий. Таким образом, в вариантах реализации могут быть предложены компьютерные системы, сконфигурированные для выполнения этапов любого из способов, описанных в настоящем документе, возможно, с различными компонентами, выполняющими соответствующий этап или соответствующую группу этапов. Хотя этапы представлены в пронумерованном виде, этапы описанных в настоящем документе способов можно выполнять в одно и то же время, в разное время, или в другом порядке. Кроме того, элементы этих этапов можно использовать с элементами других этапов из других способов. Кроме того, этап может быть полностью или частично необязательным. Кроме того, любой из этапов любого из способов можно выполнять с помощью модулей, блоков, схем или других средств системы для выполнения этих этапов.

Конкретные детали конкретных вариантов реализации можно комбинировать любым подходящим способом без отступления от сущности и объема вариантов реализации настоящего изобретения. В то же время другие варианты реализации изобретения могут указывать на конкретные варианты реализации, относящиеся к каждому отдельному аспекту, или к конкретным комбинациям этих отдельных аспектов.

Приведенное выше описание типичных вариантов реализации настоящего изобретения представлено в целях иллюстрации и описания. Оно не является исчерпывающим и не предусматривает ограничения настоящего изобретения описанной точной формой, и в свете изложенного выше возможны его многочисленные модификации и варианты.

Указание существительных в единственном и множественном числе служит для обозначения «одного или большего количества», если противоположное не указано явным образом. При использовании «или» имеется в виду «включающее или», а не «исключающее или», если противоположное не указано явным образом. Ссылка на «первый» компонент не обязательно требует предоставления второго компонента. Более того, ссылка на «первый» или «второй» компонент не ограничивает упомянутый компонент конкретным местоположением, если это не указано явным образом. Под термином «основанный на» имеется в виду «основанный по меньшей мере частично на».

Все патенты, заявки на патенты, публикации и описания, упомянутые в настоящем документе, полностью включены в настоящий документ посредством ссылки для всех целей. Ни один из них не рассматривается в качестве предшествующего уровня техники.

## ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ соединения разделенных областей молекул ДНК в биологическом образце субъекта, включающий:

распределение множества молекул ДНК из биологического образца во множество компартиментов, причем первый компартимент из множества компартиментов содержит:

первую молекулу ДНК, содержащую область А и область В, разделенные промежуточной областью,

полимеразу, характеризующуюся склонностью к добавлению определенного нуклеотида в положение в липком конце,

обратный праймер R1, содержащий первый фрагмент, комплементарный концевой последовательности области А, обратный праймер R1, содержащий второй фрагмент, содержащий перекрывающуюся последовательность,

прямой праймер F2, содержащий первый фрагмент, комплементарный последовательности начала области В, причем указанный прямой праймер F2 содержит комплементарную перекрывающуюся последовательность, комплементарную указанной перекрывающейся последовательности;

прямой праймер F1, содержащий первый фрагмент, комплементарный последовательности начала области А, и содержащий второй фрагмент,

обратный праймер R2, содержащий первый фрагмент, комплементарный концевой последовательности области В, и содержащий второй фрагмент,

прямой праймер Fz, содержащий первый фрагмент, идентичный второму фрагменту прямого праймера F1, и содержащий второй фрагмент, и

обратный праймер Rz, содержащий первый фрагмент, идентичный второму фрагменту обратного праймера R2, и содержащий второй фрагмент;

амплификацию области А с использованием полимеразы и обратного праймера R1 с целью получения первого набора ампликонов, содержащих указанный определенный нуклеотид на первом 3'-конце и содержащих комплементарный нуклеотид между участками последовательностей, соответствующими первому фрагменту и второму фрагменту обратного праймера R1, причем указанный комплементарный нуклеотид комплементарен указанному определенному нуклеотиду;

амплификацию области В с использованием полимеразы и прямого праймера F2 с целью получения второго набора ампликонов, содержащих указанный определенный нуклеотид на втором 3'-конце и содержащих комплементарный нуклеотид на противоположном конце перекрывающейся последовательности;

удлинение с использованием полимеразы первого набора ампликонов с целью

включения области В, а также второго набора ампликонов с целью включения области А, что позволяет получить первые удлиненные ампликоны, содержащие область А и область В и не содержащие промежуточную область, причем каждый первый удлиненный ампликон содержит второй фрагмент прямого праймера F1 и второй фрагмент обратного праймера R2; и

амплификацию первых удлиненных ампликонов с использованием полимеразы, прямого праймера Fz и обратного праймера Rz с целью получения вторых удлиненных ампликонов, причем каждый второй удлиненный ампликон содержит второй фрагмент прямого праймера Fz и второй фрагмент обратного праймера Rz.

2. Способ по п.1, в котором прямой праймер F2 включает второй фрагмент с комплементарной перекрывающейся последовательностью.

3. Способ по п.1, в котором первый фрагмент прямого праймера F2 включает комплементарную перекрывающуюся последовательность.

4. Способ соединения разделенных областей молекул ДНК в биологическом образце субъекта, включающий:

распределение множества молекул ДНК из биологического образца во множество компартментов, причем первый компартмент из множества компартментов содержит:

первую молекулу ДНК, содержащую область А и область В, разделенные первой промежуточной областью,

полимеразу, характеризующуюся склонностью к добавлению определенного нуклеотида в положение в липком конце,

обратный праймер R1, содержащий первый фрагмент, комплементарный концевой последовательности области А, обратный праймер R1, содержащий второй фрагмент, содержащий первую перекрывающуюся последовательность,

прямой праймер F2, содержащий первый фрагмент, комплементарный последовательности начала области В, прямой праймер F2, содержащий первую комплементарную перекрывающуюся последовательность, комплементарную указанной первой перекрывающейся последовательности,

обратный праймер R2, содержащий первый фрагмент, комплементарный концевой последовательности области В, обратный праймер R2, содержащий второй фрагмент, содержащий вторую перекрывающуюся последовательность, и

прямой праймер F3, содержащий первый фрагмент, комплементарный последовательности начала области С первой молекулы ДНК, область С, отделенную от области В второй промежуточной областью, прямой праймер F3, содержащий вторую комплементарную перекрывающуюся последовательность, комплементарную указанной

второй перекрывающейся последовательности;

амплификацию области А с использованием полимеразы и обратного праймера R1 с целью получения первого набора ампликонов, содержащих указанный определенный нуклеотид на первом 3'-конце и содержащих комплементарный нуклеотид между участками последовательностей, соответствующими первому фрагменту и второму фрагменту обратного праймера R1, причем указанный комплементарный нуклеотид комплементарен указанному определенному нуклеотиду;

амплификацию области В с целью получения второго набора ампликонов, содержащих указанный определенный нуклеотид на втором 3'-конце и содержащих комплементарный нуклеотид на противоположном конце перекрывающейся последовательности, с использованием полимеразы, прямого праймера F2 и обратного праймера R2;

амплификацию области С с использованием полимеразы и прямого праймера F3 с целью получения третьего набора ампликонов;

удлинение с помощью полимеразы:

первого набора ампликонов, содержащих область В и область С,

третьего набора ампликонов с целью включения области А и области В, что позволяет получить удлиненные ампликоны, содержащие область А, область В и область С и не содержащие первую промежуточную область и вторую промежуточную область.

5. Способ соединения разделенных областей молекул ДНК в биологическом образце субъекта, включающий:

распределение множества молекул ДНК из биологического образца во множество компартментов, причем первый компартмент из множества компартментов содержит:

первую молекулу ДНК, содержащую область А и область В, разделенные первой промежуточной областью,

полимеразу, характеризующуюся склонностью к добавлению определенного нуклеотида в положение в липком конце,

обратный праймер R1, содержащий первый фрагмент, комплементарный концевой последовательности области А, обратный праймер R1, содержащий второй фрагмент, содержащий перекрывающуюся последовательность,

прямой праймер F2, содержащий первый фрагмент, комплементарный последовательности начала области В, прямой праймер F2, содержащий второй фрагмент, содержащий комплементарную перекрывающуюся последовательность, комплементарную указанной перекрывающейся последовательности, и

прямой праймер F4, содержащий первый фрагмент, комплементарный

последовательности начала области D первой молекулы ДНК, область D, отделенную от области A второй промежуточной областью, прямой праймер R4, содержащий второй фрагмент, содержащий комплементарную перекрывающуюся последовательность,

амплификацию области A с использованием полимеразы и обратного праймера R1 с целью получения первого набора ампликонов, содержащих указанный определенный нуклеотид на первом 3'-конце и содержащих комплементарный нуклеотид между участками последовательностей, соответствующими первому фрагменту и второму фрагменту обратного праймера R1, причем указанный комплементарный нуклеотид комплементарен указанному определенному нуклеотиду;

амплификацию области B с использованием полимеразы и прямого праймера F2 с целью получения второго набора ампликонов, содержащих указанный определенный нуклеотид на втором 3'-конце и содержащих комплементарный нуклеотид на противоположном конце перекрывающейся последовательности;

амплификацию области D с использованием полимеразы и прямого праймера F4 с целью получения третьего набора ампликонов;

удлинение с использованием полимеразы первого набора ампликонов с целью включения области B, а также второго набора ампликонов с целью включения области A, что позволяет получить первые удлиненные ампликоны, содержащие область A и область B и не содержащие первую промежуточную область; и

удлинение с использованием полимеразы первого набора ампликонов с целью включения области D, и третьего набора ампликонов с целью включения области A, что позволяет получить вторые удлиненные ампликоны, содержащие область A и область D и не содержащие вторую промежуточную область.

6. Способ по п.5, дополнительно включающий:

обнаружение первого аллеля в области A в первых удлиненных ампликонах,  
обнаружение второго аллеля в области B в первых удлиненных ампликонах,  
обнаружение первого аллеля в области A во вторых удлиненных ампликонах, и  
обнаружение третьего аллеля в области D во вторых удлиненных ампликонах,  
тем самым обнаруживая, что первый аллель в области A, второй аллель в области B и третий аллель в области D находятся в одном и том же гаплотипе.

7. Способ по п.5, в котором:

комплементарная перекрывающаяся последовательность представляет собой первую комплементарную перекрывающуюся последовательность,

перекрывающаяся последовательность представляет собой первую перекрывающуюся последовательность,

первый компартмент дополнительно включает в себя:

обратный праймер R2, имеющий первый фрагмент, комплементарный последовательности конца области В, причем обратный праймер R2 включает второй фрагмент со второй перекрывающейся последовательностью,

обратный праймер R4, имеющий первый фрагмент, комплементарный последовательности конца области D, причем обратный праймер R4 включает второй фрагмент со второй перекрывающейся последовательностью, и

прямой праймер F3, имеющий первый фрагмент, комплементарный последовательности начала области С первой молекулы ДНК, причем область С отделена от области В третьей промежуточной областью, прямой праймер F3 включает второй фрагмент со второй комплементарной перекрывающейся последовательностью,

причем амплификация области В включает использование обратного праймера R2, а амплификация области D включает использование обратного праймера R4,

причем способ дополнительно включает:

амплификацию с использованием полимеразы и прямого праймера F3 области С для получения четвертого набора ампликонов,

удлинение с использованием полимеразы вторых удлиненных ампликонов для включения области С, тем самым получая третьи удлиненные ампликоны, которые включают область А, область D и область С, и

удлинение с использованием полимеразы первых удлиненных ампликонов для включения области С, тем самым получая четвертые удлиненные ампликоны, которые включают область А, область В и область С.

8. Набор для соединения разделенных областей молекул ДНК в биологическом образце субъекта, содержащий:

полимеразу, характеризующуюся склонностью к добавлению определенного нуклеотида в положение в липком конце;

обратный праймер R1, содержащий первый фрагмент, комплементарный последовательности конца области А первой молекулы ДНК, обратный праймер R1, содержащий второй фрагмент, содержащий перекрывающуюся последовательность;

прямой праймер F2, содержащий первый фрагмент, комплементарный последовательности начала области В первой молекулы ДНК, область В, отделенную от области А промежуточной областью, прямой праймер F2, содержащий комплементарную перекрывающуюся последовательность, комплементарную указанной перекрывающейся последовательности;

прямой праймер F1, содержащий первый фрагмент, комплементарный

последовательности начала области А, и содержащий второй фрагмент,

обратный праймер R2, содержащий первый фрагмент, комплементарный концевой последовательности области В, и содержащий второй фрагмент,

прямой праймер Fz, содержащий первый фрагмент, идентичный второму фрагменту прямого праймера F1, и содержащий второй фрагмент, и

обратный праймер Rz, содержащий первый фрагмент, идентичный второму фрагменту обратного праймера R2, и содержащий второй фрагмент.

9. Набор по п.8, в котором прямой праймер F2 включает второй фрагмент с комплементарной перекрывающейся последовательностью.

10. Набор по п.8, в котором первый фрагмент прямого праймера F2 включает комплементарную перекрывающуюся последовательность.

11. Набор для соединения разделенных областей молекул ДНК в биологическом образце субъекта, содержащий:

полимеразу, характеризующуюся склонностью к добавлению определенного нуклеотида в положение в липком конце;

обратный праймер R1, содержащий первый фрагмент, комплементарный последовательности конца области А первой молекулы ДНК, обратный праймер R1, содержащий второй фрагмент, содержащий первую перекрывающуюся последовательность;

прямой праймер F2, содержащий первый фрагмент, комплементарный последовательности начала области В первой молекулы ДНК, область В, отделенную от области А промежуточной областью, прямой праймер F2, содержащий первую комплементарную перекрывающуюся последовательность, комплементарную указанной первой перекрывающейся последовательности;

обратный праймер R2, содержащий первый фрагмент, комплементарный последовательности конца области В, обратный праймер R2, содержащий второй фрагмент, содержащий вторую перекрывающуюся последовательность, и

прямой праймер F3, содержащий первый фрагмент, комплементарный последовательности начала области С первой молекулы ДНК, область С, отделенную от области В второй промежуточной областью, прямой праймер F3, содержащий вторую комплементарную перекрывающуюся последовательность, комплементарную указанной второй перекрывающейся последовательности.

12. Набор для соединения разделенных областей молекул ДНК в биологическом образце субъекта, содержащий:

полимеразу, характеризующуюся склонностью к добавлению определенного

нуклеотида в положение в липком конце;

обратный праймер R1, содержащий первый фрагмент, комплементарный последовательности конца области А первой молекулы ДНК, обратный праймер R1, содержащий второй фрагмент, содержащий перекрывающуюся последовательность;

прямой праймер F2, содержащий первый фрагмент, комплементарный последовательности начала области В первой молекулы ДНК, область В, отделенную от области А промежуточной областью, прямой праймер F2, содержащий комплементарную перекрывающуюся последовательность, комплементарную указанной перекрывающейся последовательности; и

прямой праймер F4, содержащий первый фрагмент, комплементарный последовательности начала области D первой молекулы ДНК, область D, отделенную от области А второй промежуточной областью, прямой праймер F4, содержащий комплементарную перекрывающуюся последовательность.

13. Набор по п.12, в котором:

комплементарная перекрывающаяся последовательность представляет собой первую комплементарную перекрывающуюся последовательность, а

перекрывающаяся последовательность представляет собой первую перекрывающуюся последовательность,

причем набор дополнительно содержит:

обратный праймер R2 с первым фрагментом, комплементарным последовательности конца области В, причем обратный праймер R2 включает второй фрагмент со второй перекрывающейся последовательностью,

обратный праймер R4 с первым фрагментом, комплементарным последовательности конца области D, причем обратный праймер R4 включает второй фрагмент со второй перекрывающейся последовательностью, и

прямой праймер F3 с первым фрагментом, комплементарным последовательности начала области С первой молекулы ДНК, причем область С отделена от области В третьей промежуточной областью, прямой праймер F3 включает второй фрагмент со второй комплементарной перекрывающейся последовательностью.

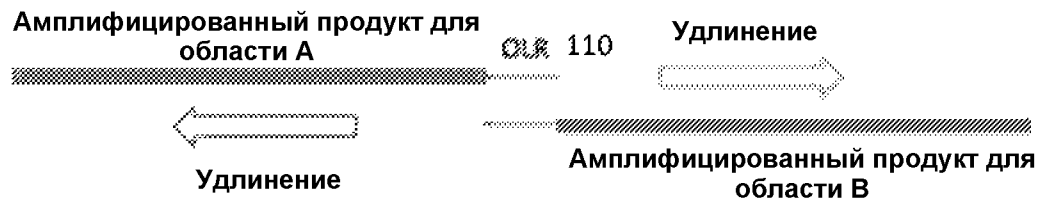




ФИГ. 1А



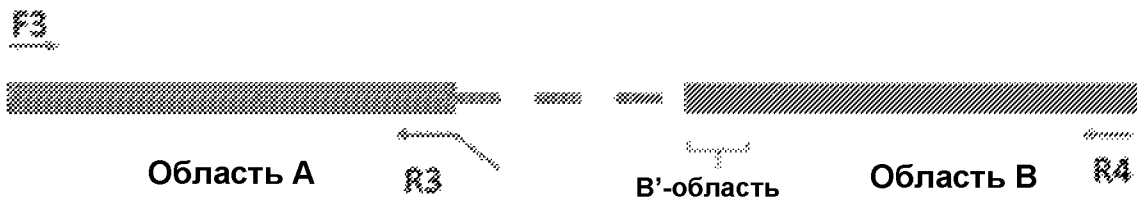
ФИГ. 1В



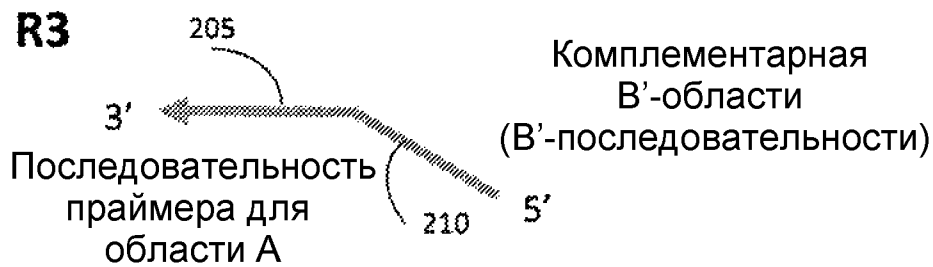
ФИГ. 1С



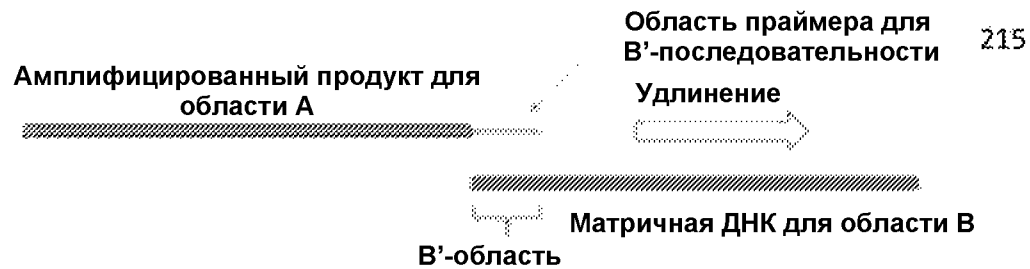
ФИГ. 1D



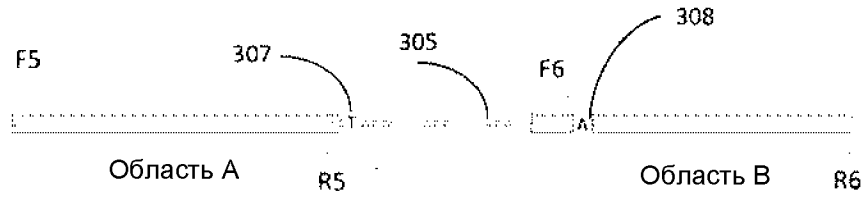
ФИГ. 2А



ФИГ. 2В



ФИГ. 2С



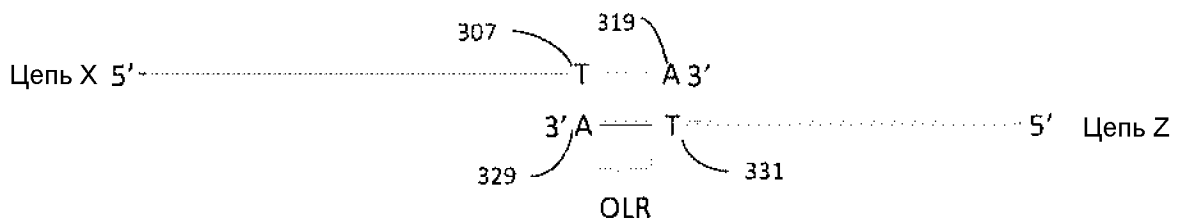
**ФИГ. 3А**



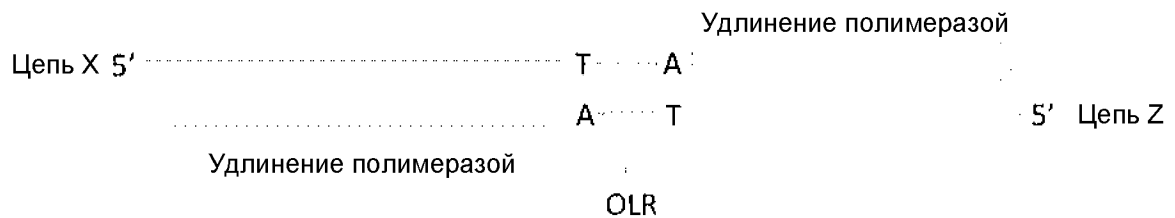
**ФИГ. 3В**



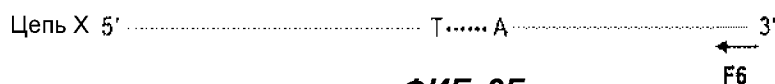
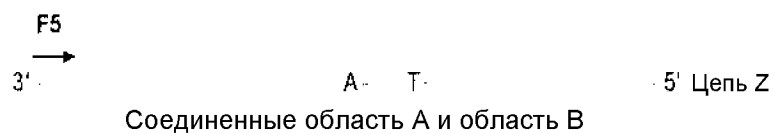
**ФИГ. 3С**



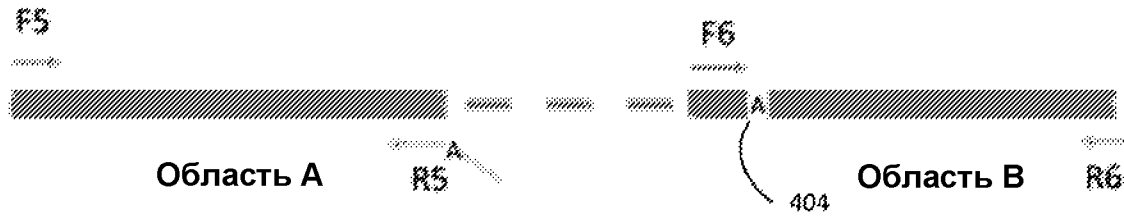
**ФИГ. 3D**



**ФИГ. 3Е**



**ФИГ. 3F**



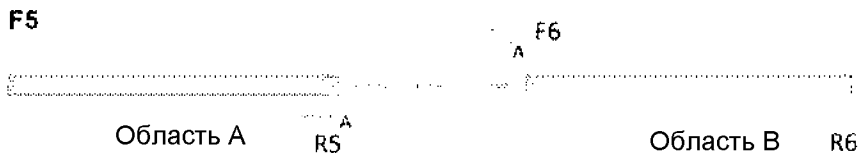
ФИГ. 4А



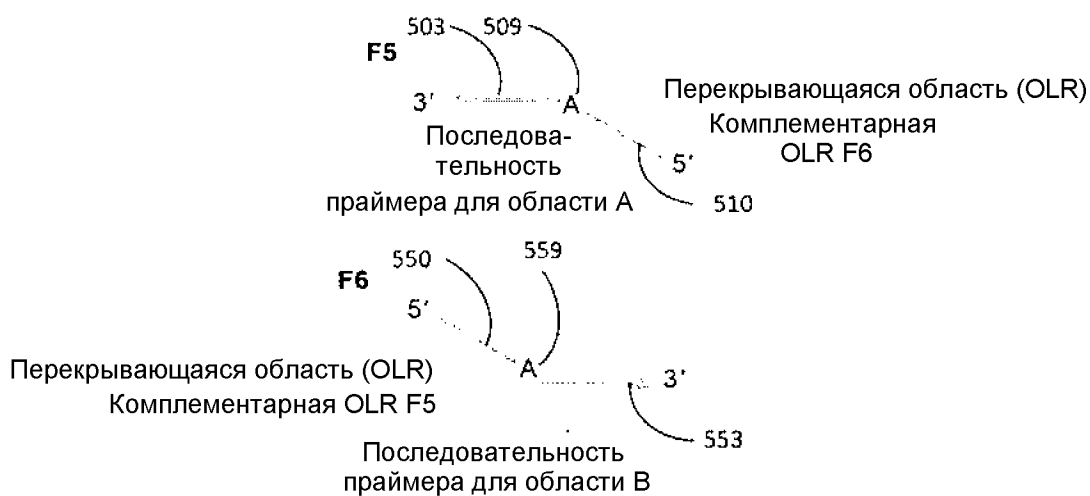
ФИГ. 4В



ФИГ. 4С



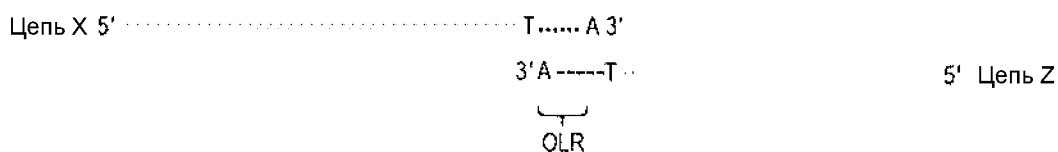
ФИГ. 5А



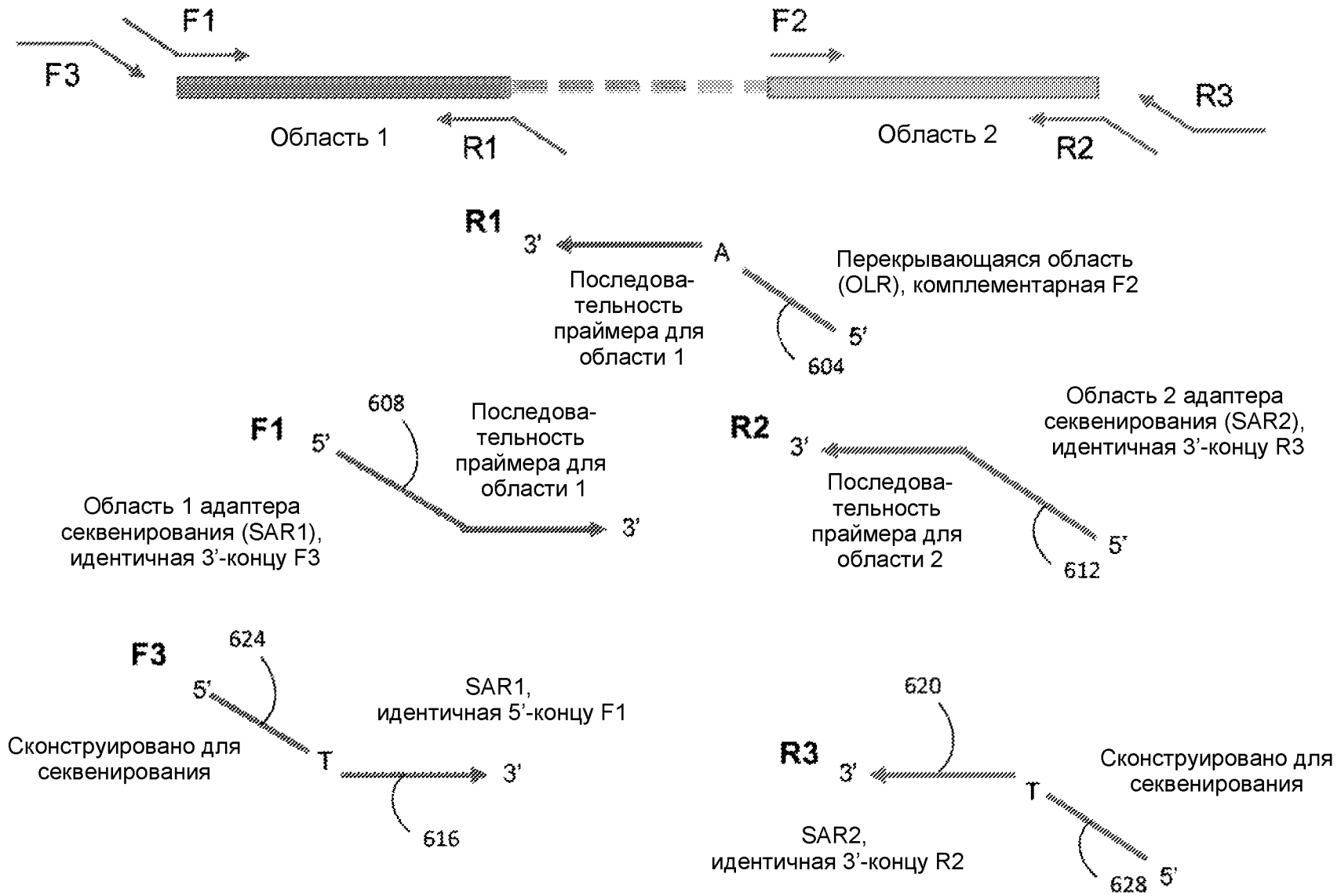
ФИГ. 5В



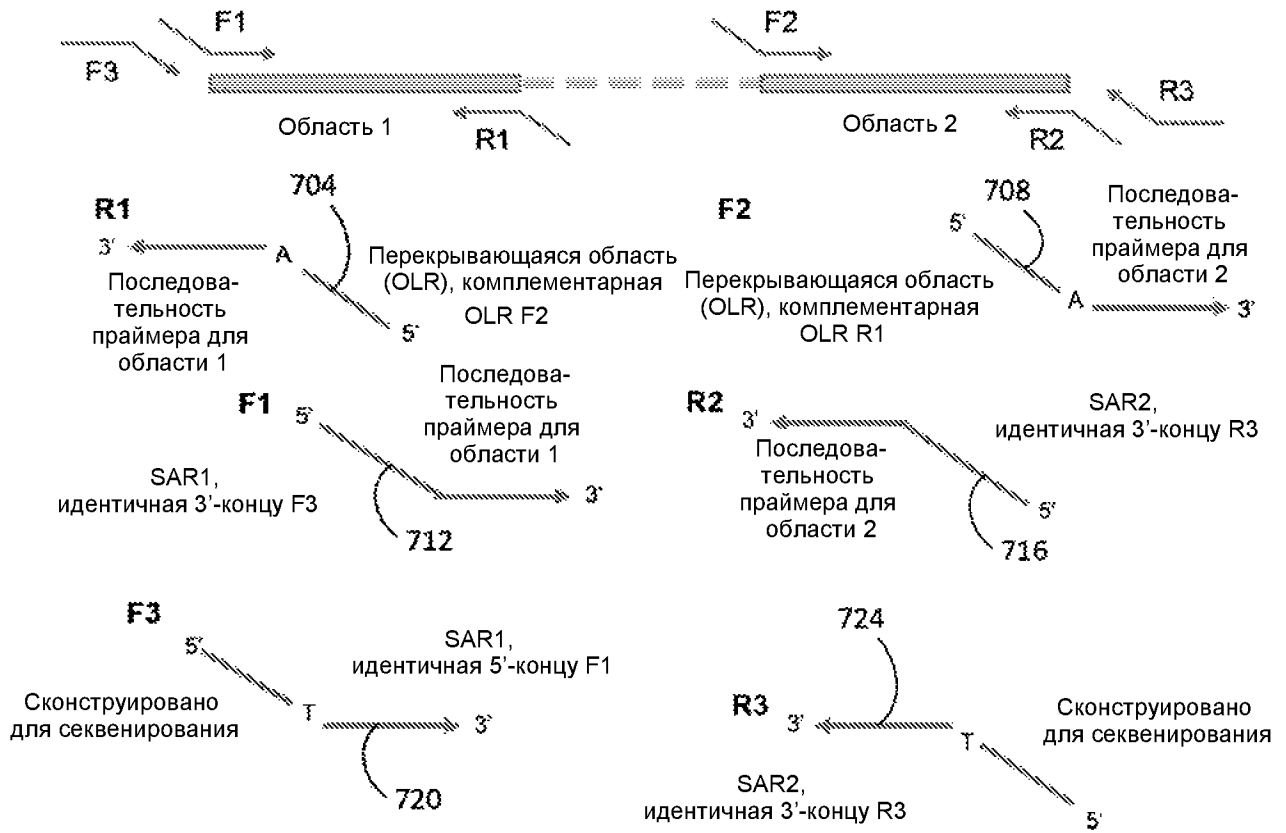
ФИГ. 5С



ФИГ. 5D



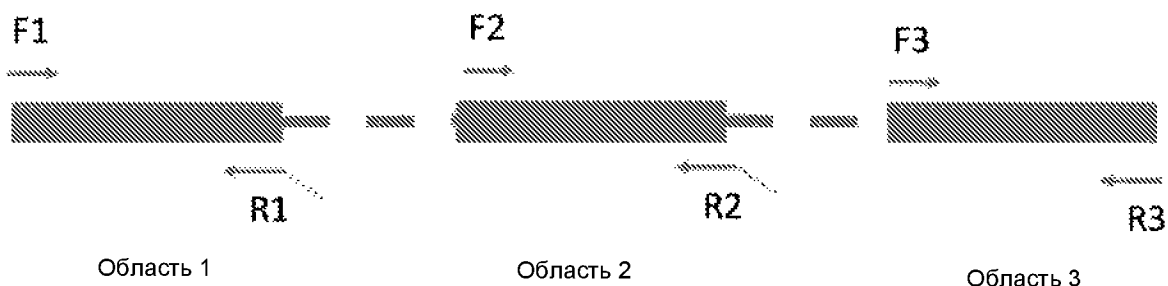
608  
ФИГ. 6



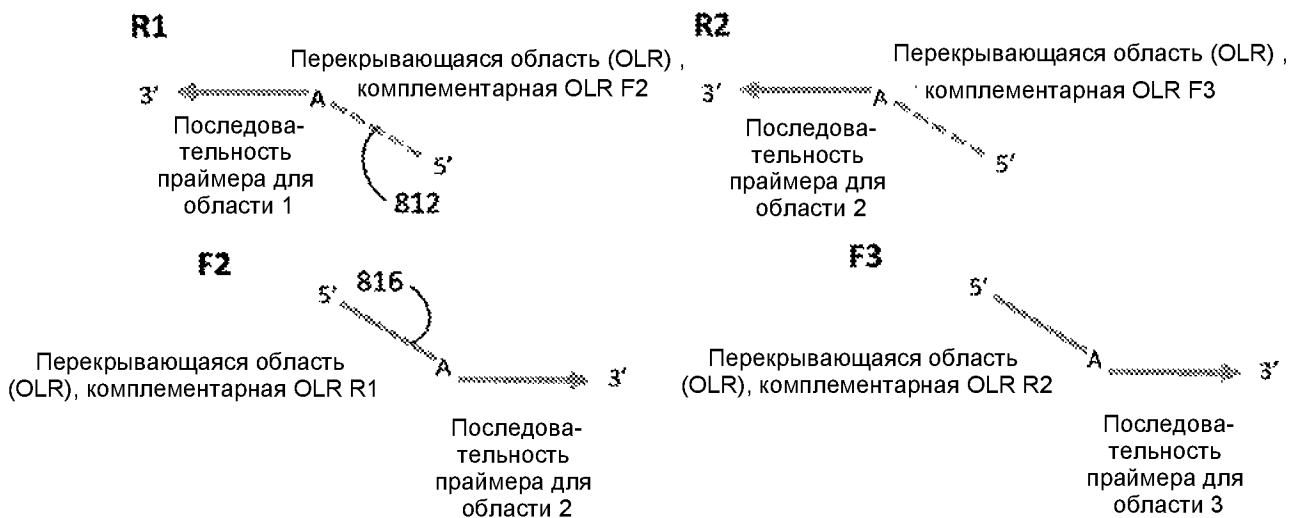
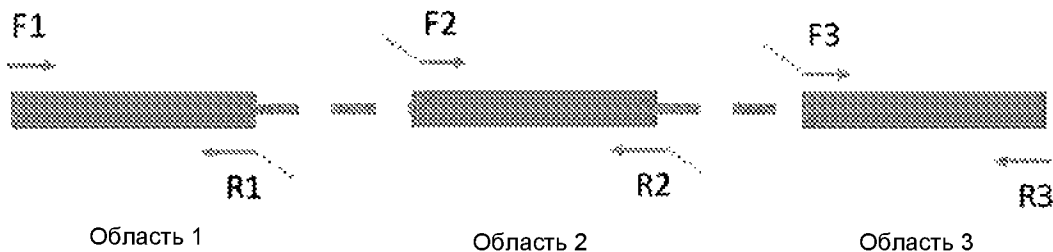
ФИГ. 7А



ФИГ. 7В

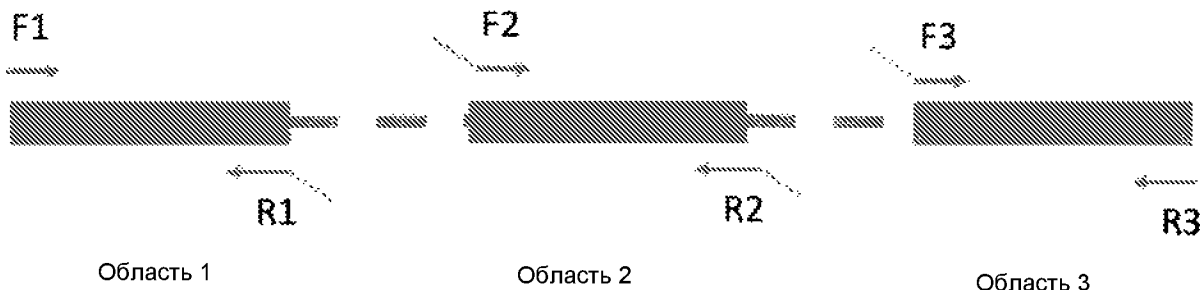


ФИГ. 8А

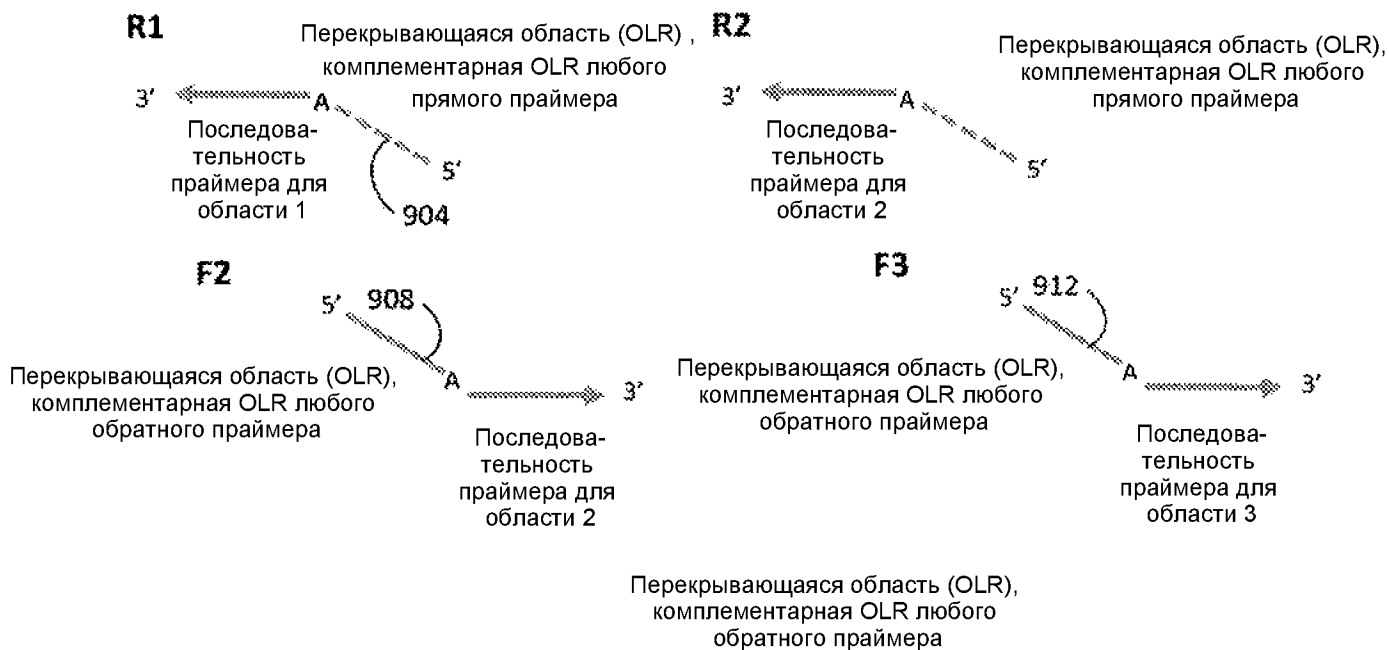


ФИГ. 8В

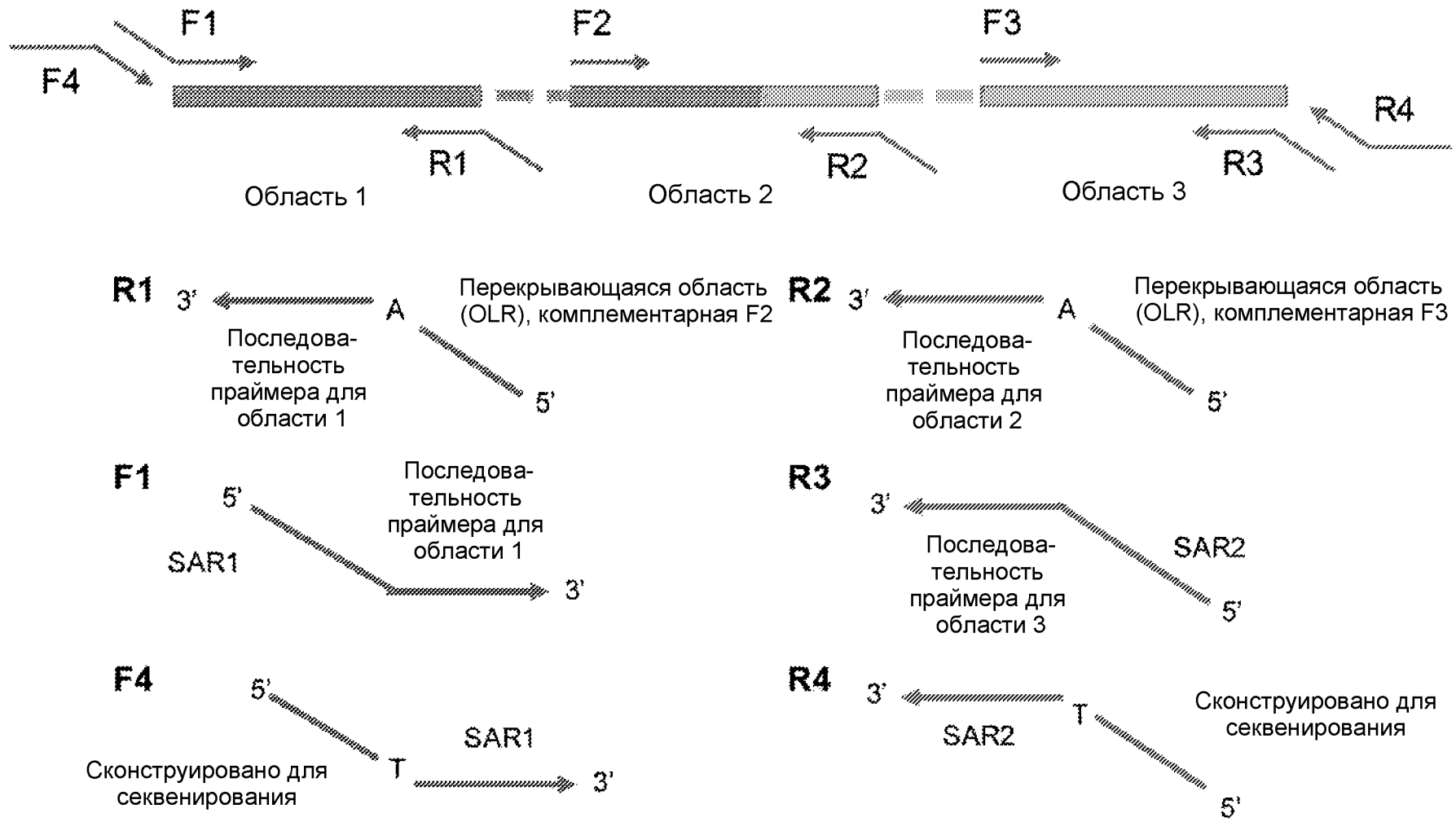




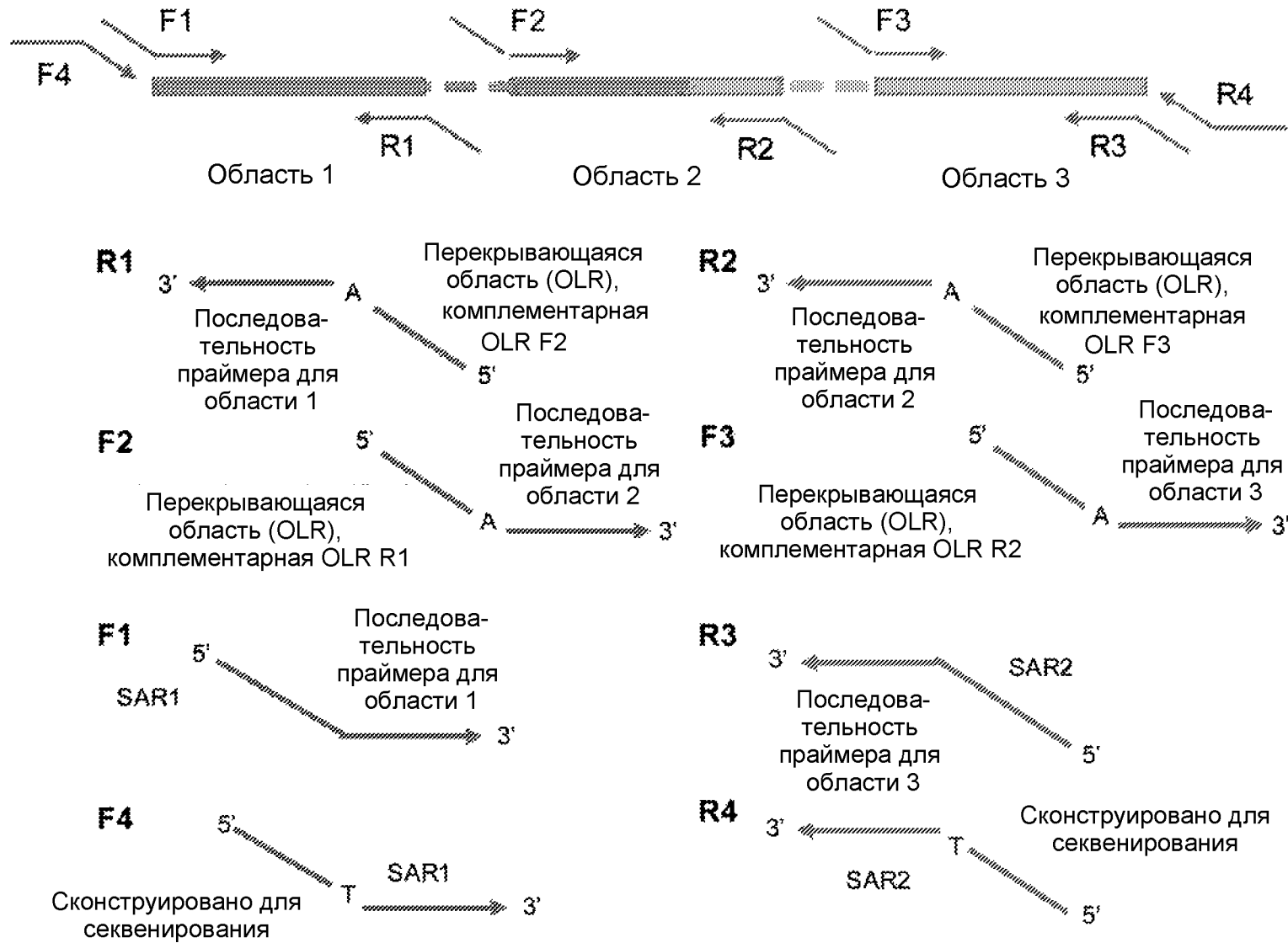
ФИГ. 9А



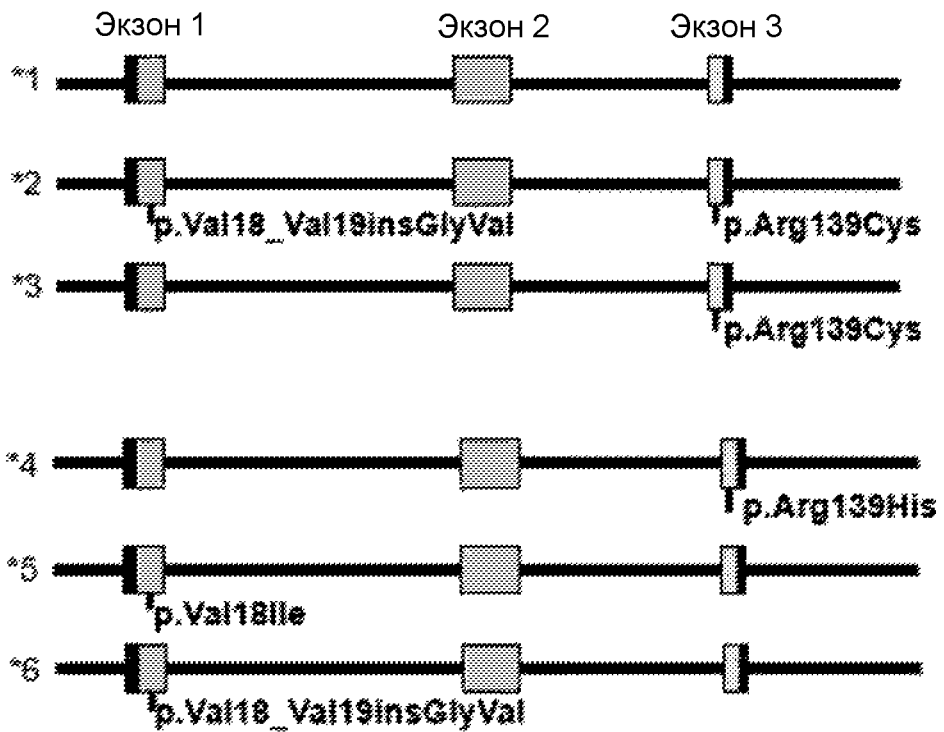
ФИГ. 9В



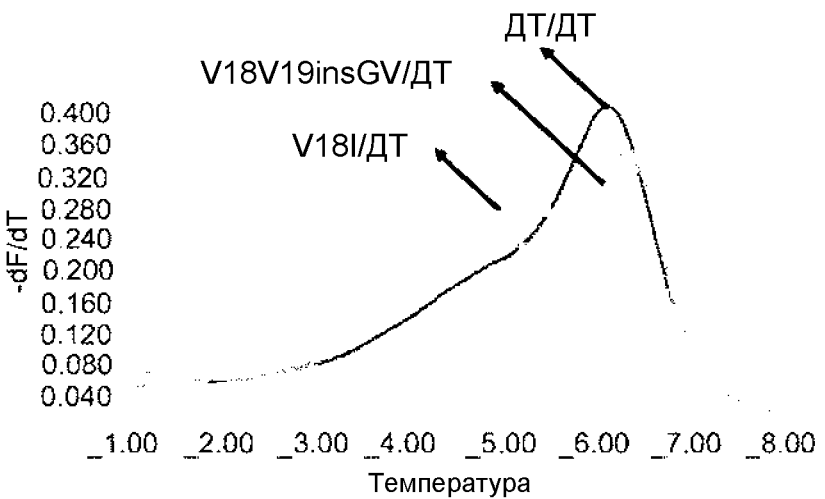
ФИГ. 10А



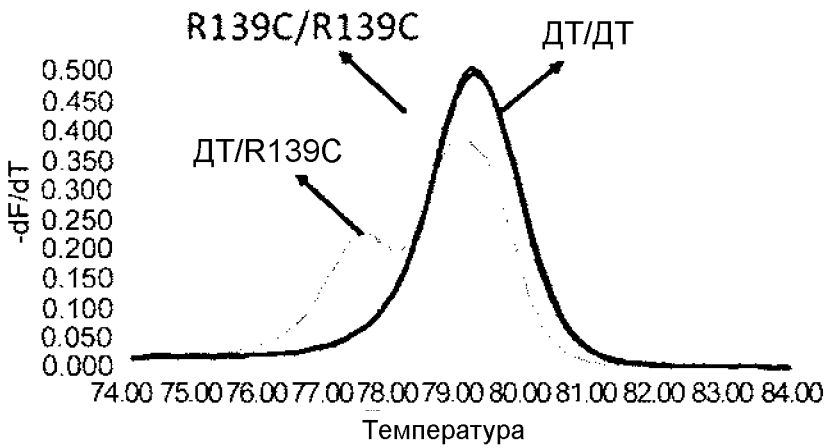
ФИГ. 10В



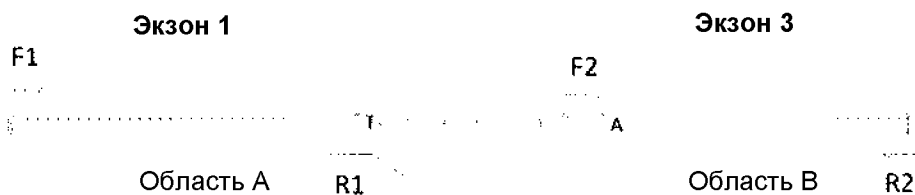
ФИГ. 11



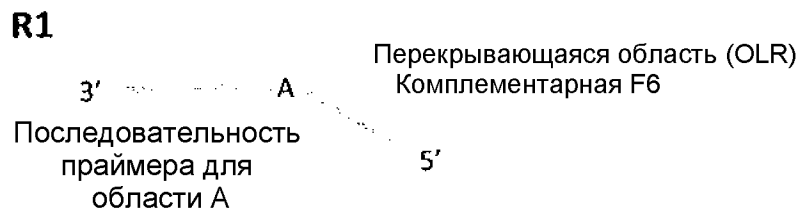
ФИГ. 12А



ФИГ. 12В

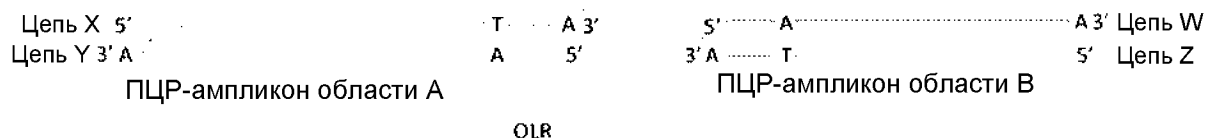


ФИГ. 13А

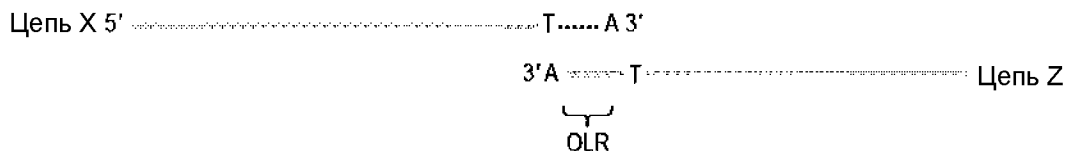


ФИГ. 13В

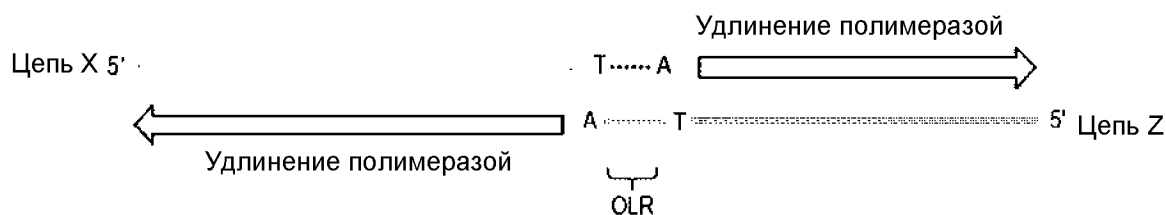
Дополнительные остатки аденозина, добавленные полимеразой к 3'-концу ПЦР-ампликона



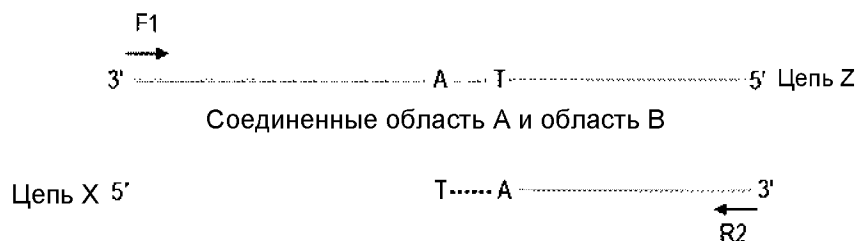
ФИГ. 13С



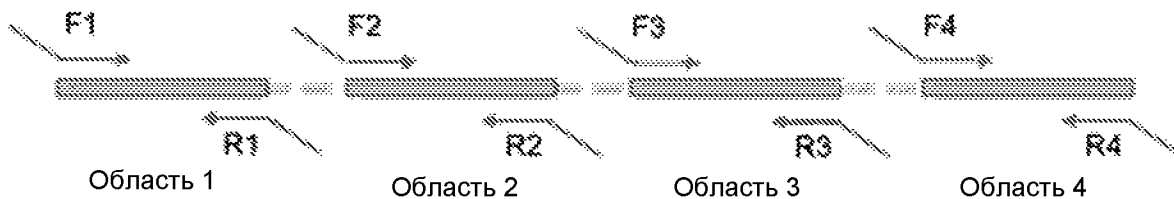
ФИГ. 13D



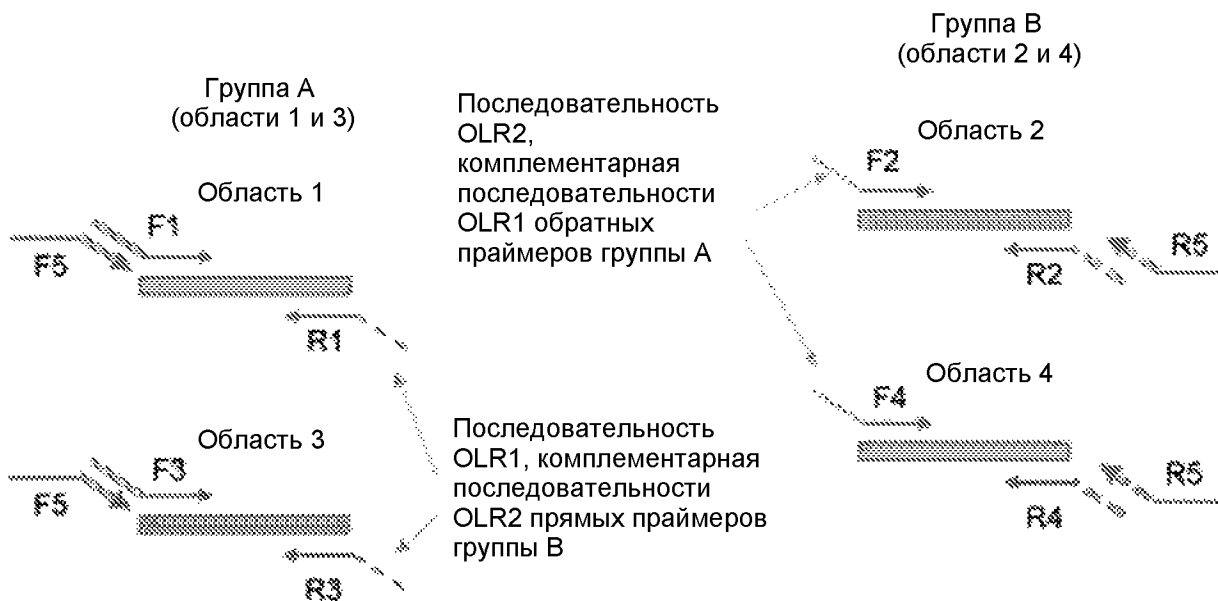
ФИГ. 13E



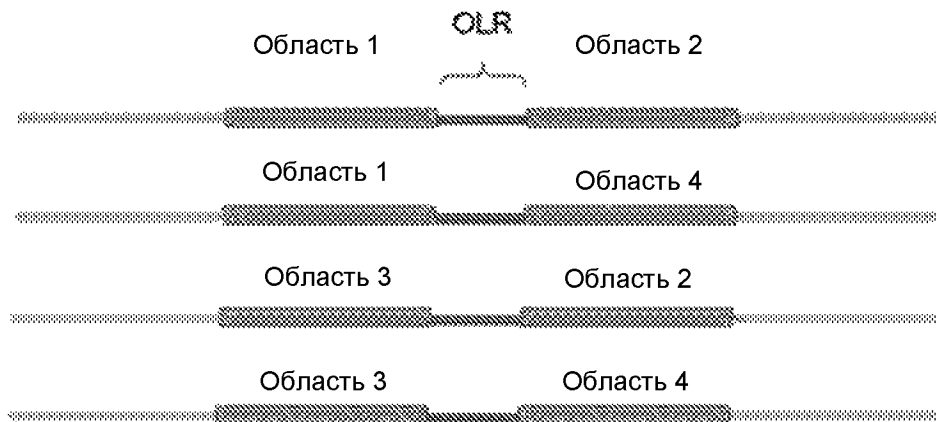
ФИГ. 13F



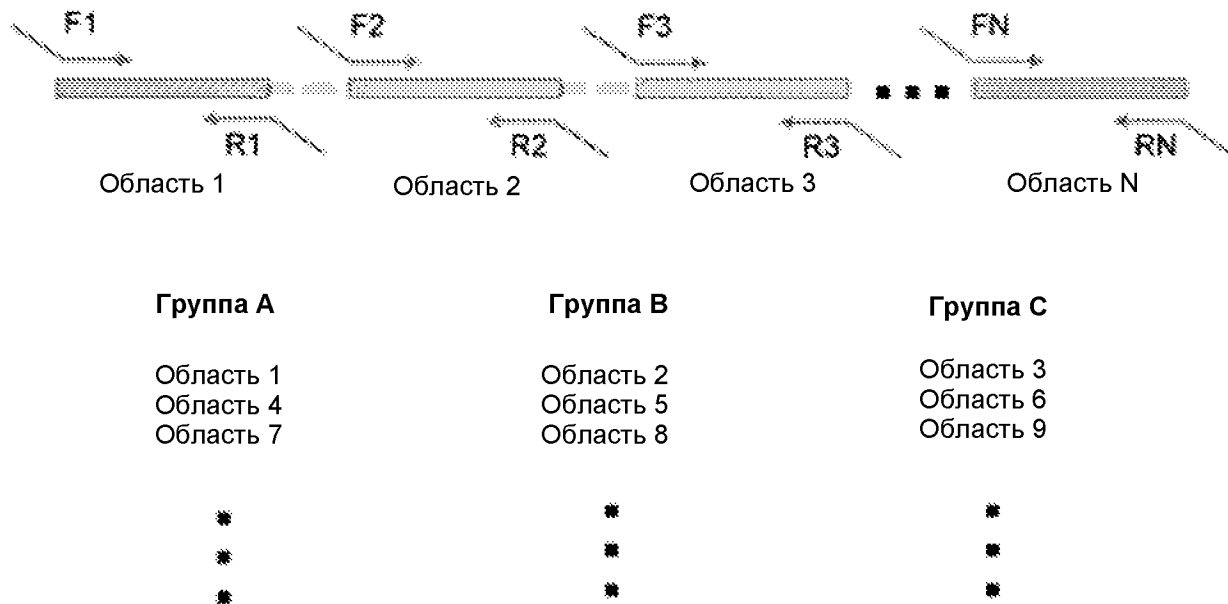
ФИГ. 14А



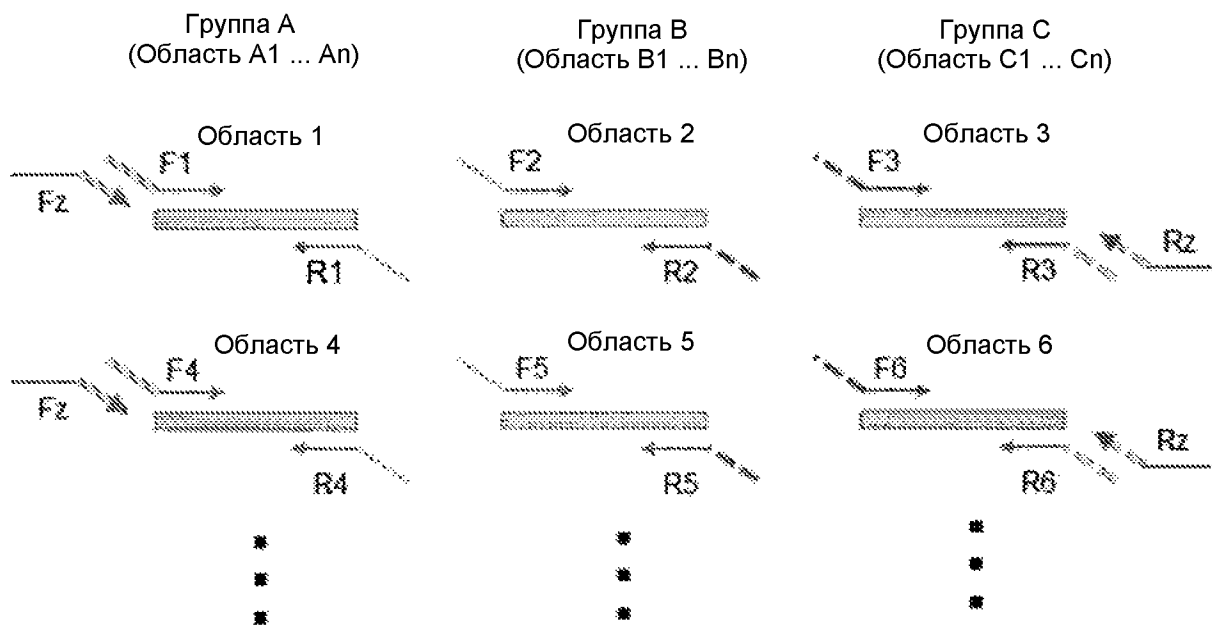
ФИГ. 14В



ФИГ. 14С



**ФИГ. 15А**

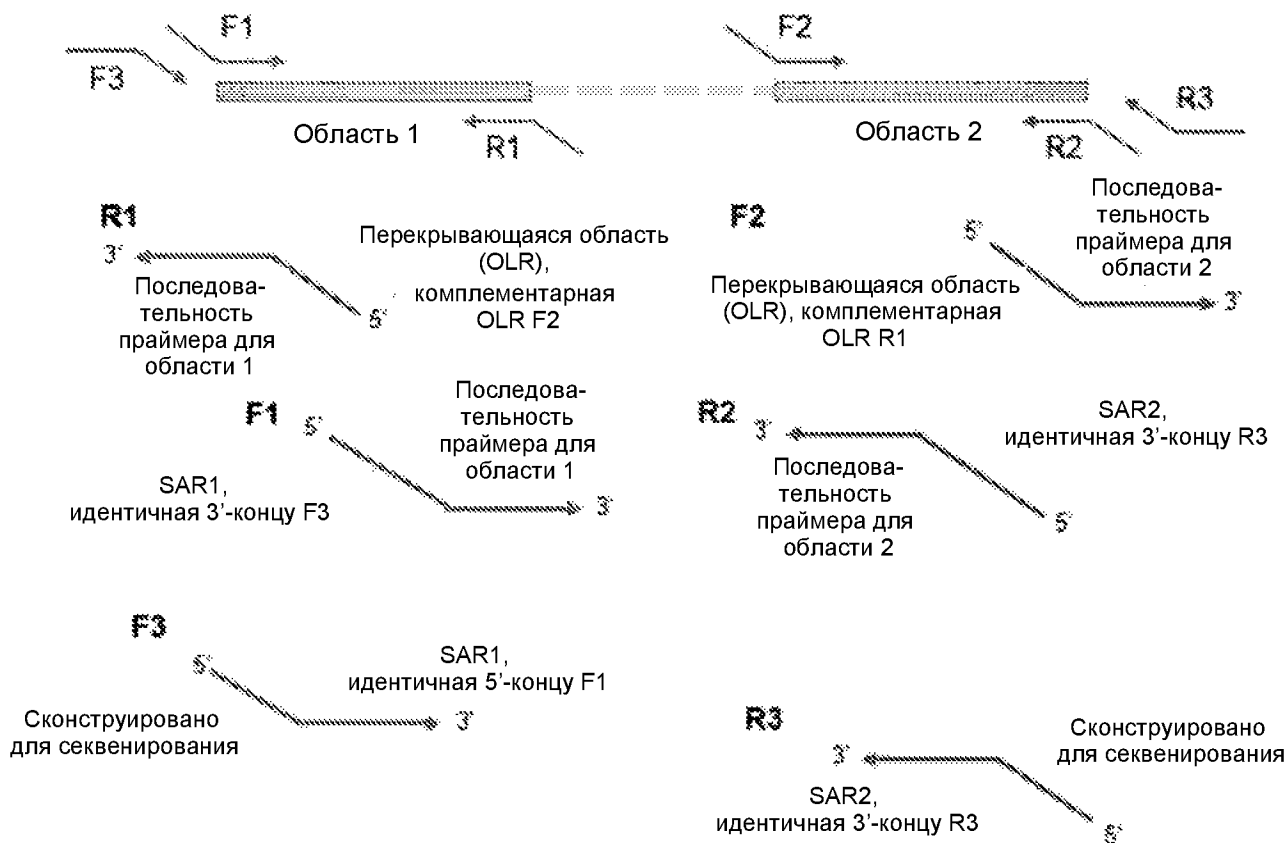


**ФИГ. 15В**

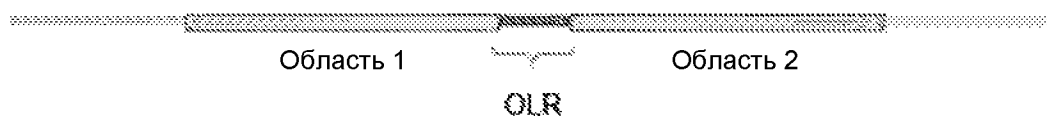


**ФИГ. 15С**

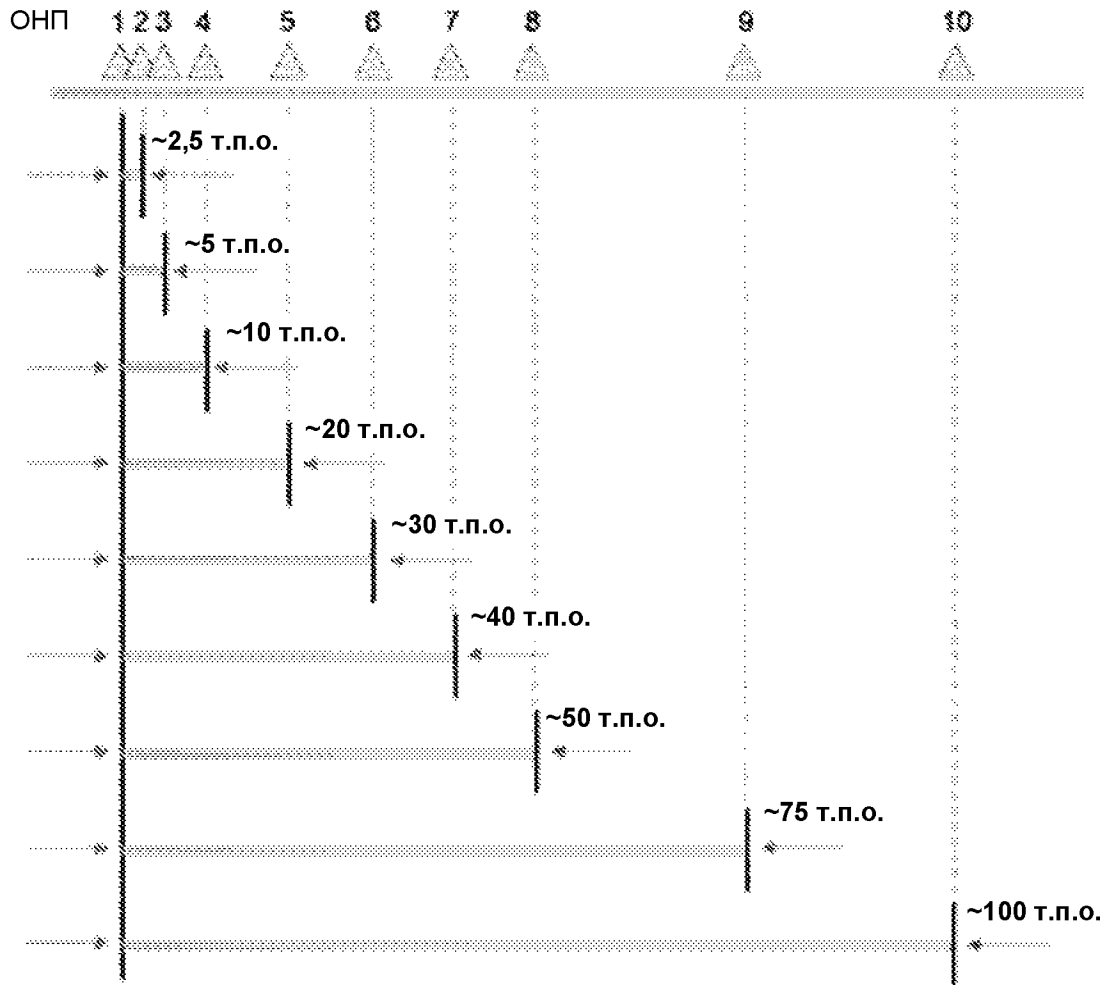




ФИГ. 16А

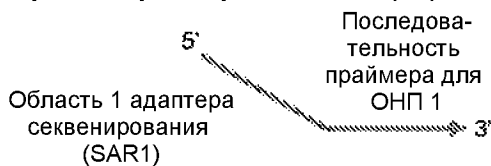


ФИГ. 16В

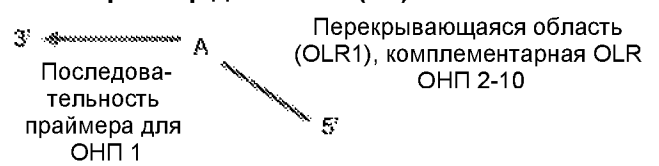


ФИГ. 17А

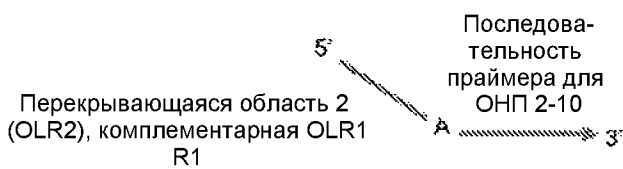
Прямой праймер для ОП 1 (F1)



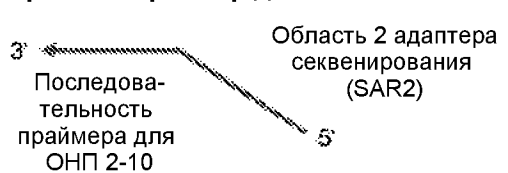
Обратный праймер для ОП 1 (R1)



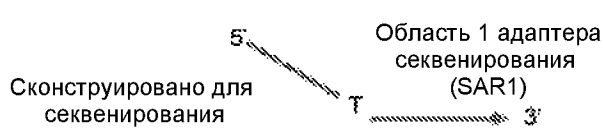
Прямой праймер для ОП 2-10



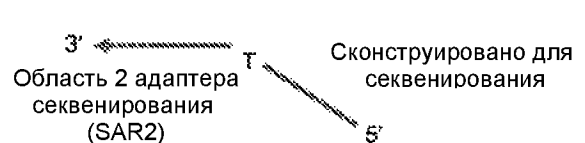
Обратный праймер для ОП 2-10



Внешний прямой праймер (Fz)



Внешний обратный праймер (Rz)



ФИГ. 17В

Идентификатор ОНП	Координата на хромосоме 11	Расстояние от первого ОНП (т.п.о.)	Генотип	Прямой праймер (5' - 3')	Обратный праймер (5' - 3')
rs11036868	5331270	-	T/C	AATCGGGAAGCT GAAGGCCAAGGA CTGTAGAGACAT AGATGA	<u>CAC</u> TACCGTC <u>GGATGACAT</u> A CTGGAGAGAT TTCATCCTTGG AG
rs78930059	5333450	2.18	G/T	<u>ATGTCATCCGACG</u> <u>GTAGTG</u> AGCAAT GACAGAGITTTGA CATATTGATT	GTGCTCTTCC GATCTGTTAG TGAATCAATGT GTAAATCATA AAGTAATT
rs10768807	5336582	5.31	T/A	<u>ATGTCATCCGACG</u> <u>GTAGTG</u> ACCTTCC TCCATAATCACAT CTGCTA	GTGCTCTTCC GATCTCTGTT GCACCAACCT AATATAAGGA A
rs4910551	5343046	11.78	G/A	<u>ATGTCATCCGACG</u> <u>GTAGTG</u> AGGAGT GACAATACAGAA AAAAATAGAGG	GTGCTCTTCC GATCTGAGGG GATTTGTATC CGTTGTTT
rs10837882	5351633	20.36	T/A	<u>ATGTCATCCGACG</u> <u>GTAGTG</u> ACCTACA TCTCCATACTTCT TGGCAAT	GTGCTCTTCC GATCTATGAT CATTCTTGAT GAGAAAGAG AAG
rs2736557	5361361	30.09	G/T	<u>ATGTCATCCGACG</u> <u>GTAGTG</u> AGGAGA GAGGAGTGGGCC GTA	GTGCTCTTCC GATCTAAAGA AAATTAGTTT CAGAAACTCC TCC
rs884222	5371742	40.47	A/T	<u>ATGTCATCCGACG</u> <u>GTAGTG</u> ATATCCA TCAATACAACACC TTCCATAG	GTGCTCTTCC GATCTACAGT AGATGTTAGA TGCCCTTACA CA
rs2736543	5381334	50.06	A/G	<u>ATGTCATCCGACG</u> <u>GTAGTG</u> ATACCTT AAAGGTTAAGGC CTGAATG	GTGCTCTTCC GATCTGGGGA AGGTGAAGTA TCAACACA
rs4910781	5406896	75.63	A/C	<u>ATGTCATCCGACG</u> <u>GTAGTG</u> AAGCAA ACTCAGAGGGTT GGAGTAC	GTGCTCTTCC GATCTTATTTT TGTAGCTCTG ACAGTTGTCT CC
rs10838096	5431296	100.03	A/G	<u>ATGTCATCCGACG</u> <u>GTAGTG</u> ACCCTGA CTCCATGCAGGA GA	GTGCTCTTCC GATCTGTTGA CTTTGATGGC TGCCCTG

ФИГ. 18

Спаривание ОНП	Нар I: аллель С в ОНП 1 (Эталон: 5'-СТААТТТААГ-3')		Нар II: аллель Т в ОНП 1 (Эталон: 5'-ТГТГАГАГСА-3')	
	Комбинация аллелей (частота)	Определенный Нар I	Комбинация аллелей (частота)	Определенный Нар II
ОНП 1 – ОНП 2	C—T (80.5%) C—G (19.5%)	CT-----	T—G (84.7%) T—T (15.3%)	TG-----
ОНП 1 – ОНП 3	C—A (89%) C—T (11%)	C—A-----	T—T (85.5%) T—A (14.5%)	T—T-----
ОНП 1 – ОНП 4	C—A (81.3%) C—G (18.7%)	C—A-----	T—G (83.1%) T—A (16.9%)	T—G-----
ОНП 1 – ОНП 5	C—T (67.8%) C—A (32.2%)	C---T-----	T—A (90.3%) T—T (9.7%)	T---A-----
ОНП 1 – ОНП 6	C—T (80.2%) C—G (19.8%)	C-----T---	T—G (86.4%) T—T (13.6%)	T-----G---
ОНП 1 – ОНП 7	C—T (67.2%) C—A (32.8%)	C-----T---	T—A (65.2%) T—T (34.8%)	T-----A---
ОНП 1 – ОНП 8	C—A (68.6%) C—G (31.4%)	C-----A--	T—G (61.7%) T—A (38.3%)	T-----G--
ОНП 1 – ОНП 9	C—A (65.8%) C—C (34.2%)	C-----A-	T—A (53%) T—C (47%)	T-----x-
ОНП 1 – ОНП 10	C—G (58.1%) C—A (41.9%)	C-----x	T—A (50.4%) T—G (49.6%)	T-----x

ФИГ. 19

№ области	Идентификатор ОНП	Координата на хромосоме 11	Прямой праймер (5' - 3')	Обратный праймер (5' - 3')
1	rs7930833	5222527	AATCGGGAAGCTGA AGCTTGATAGTTTCT ACTTTGGGTGATAA TT	<u>CACTACCGTCCGGATGAC</u> <u>AT</u> AATGGGAAGACATGT ATATAATCTTAGAGATA
2	rs1236487 2	5222914	<u>ATGTCATCCGACGG</u> <u>TAGTG</u> AGCAGAAAG AGATAATATTAAGG ATAATCAG	GTGCTCTCCGATCTGC AGAAAGAGATAATATT AAGGATAATCAG
3	rs1083762 8	5223174	<u>ATGTCATCCGACGG</u> <u>TAGTG</u> AGTCAAATG GATATTTGTATCTGA AAAATC	GTGCTCTCCGATCTAT GCATCTTGATGATTAGA ATTGCA
4	rs2187610	5224176	AATCGGGAAGCTGA AGTCATGGTTCACCT TTCATTTGTTT	<u>CACTACCGTCCGGATGAC</u> <u>AT</u> ACATCAAACATAAAA AATTTCCACACAA
5	rs1076868 2	5224277	<u>ATGTCATCCGACGG</u> <u>TAGTG</u> ACAAATTCA AAAAACGGCTGCA	GTGCTCTCCGATCTAA GAAATCAAAGAAGAA AATCTAATATTCA
6	rs1083763 0	5224812	AATCGGGAAGCTGA AGGTAATAAGACAG TAGTGAATATCAAG CTACAAA	<u>CACTACCGTCCGGATGAC</u> <u>AT</u> AGTTAGGACTGAGA AGAATTTGAAAGG
7	rs1083763 1	5225126	<u>ATGTCATCCGACGG</u> <u>TAGTG</u> AGTATCTCTA AGCAAGAGAACTGA GTGG	GTGCTCTCCGATCTATT TGCTTATCCATGCATCTCT CAG
8	rs7110263	5225282	AATCGGGAAGCTGA AGTTTTCTGAGGGAT GAATAAGGCAT	<u>CACTACCGTCCGGATGAC</u> <u>AT</u> AATGGCAACAGCCC CTGAT
9	rs1609812	5225911	<u>ATGTCATCCGACGG</u> <u>TAGTG</u> ATACAATTTA TATGCAGAAATATTT ATATGCAG	GTGCTCTCCGATCTATT TCTGGGTTAAGGCAATA GCAA
10	rs7480526	5226503	AATCGGGAAGCTGA AGTTCCCATTTCTAAA CTGTACCCTGTT	<u>CACTACCGTCCGGATGAC</u> <u>AT</u> ATGGTTAAGTTCAT GTCATAGGAAGG

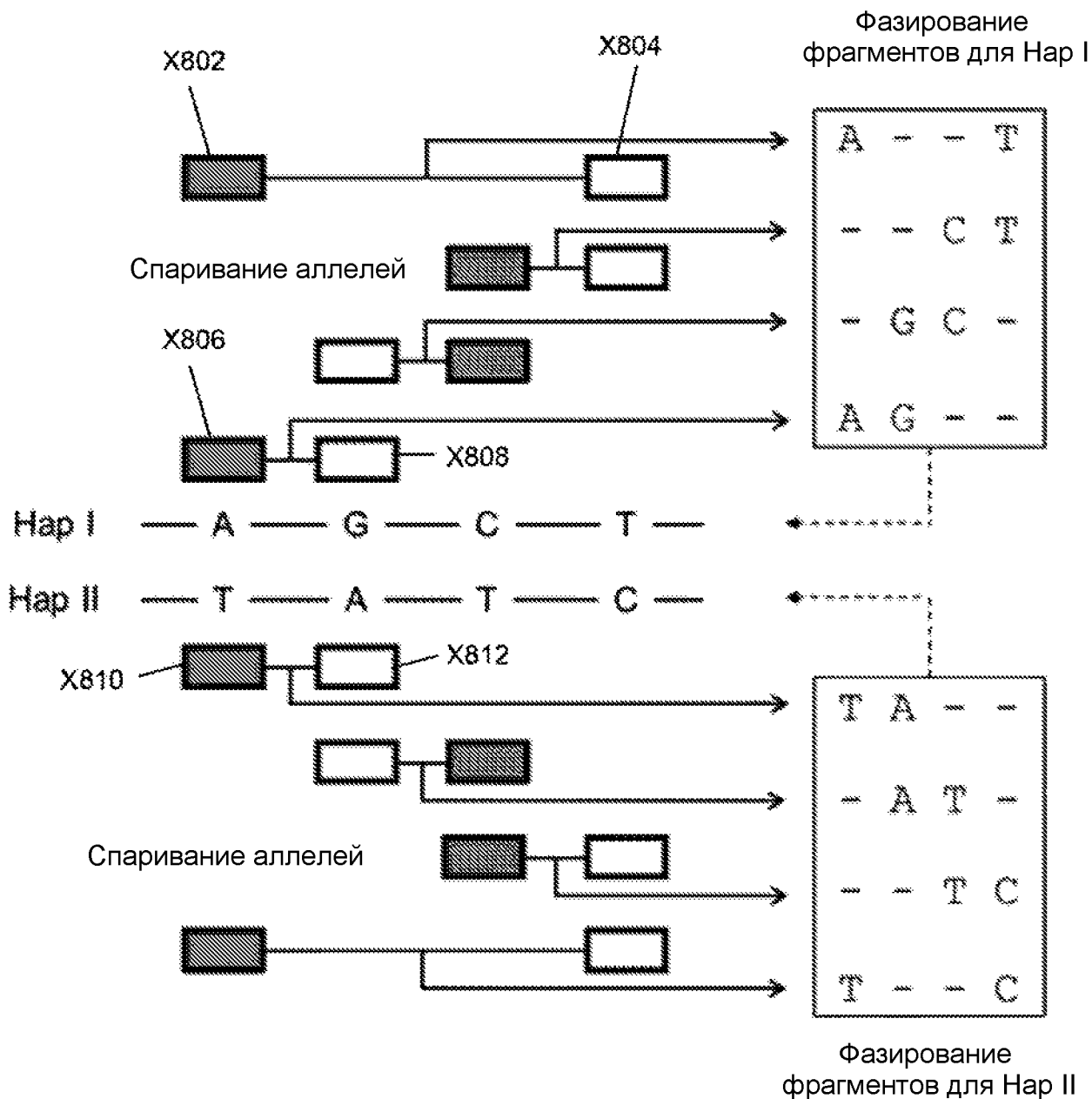
ФИГ. 20

№ области	Идентификатор ОНП	Координата на хромосоме 11	Прямой праймер (5' - 3')	Обратный праймер (5' - 3')
11	rs10768683	5226561	AATCGGGAAGCTGAAG CCTATGACATGAACTTA ACCATAGAAA	<u>CACTACCGTCGGATGAC</u> <u>ATATGAGAACTTCAGGG</u> TGAGTCTATG
12	rs10742583	5227411	<u>ATGTCATCCGACGGTA</u> <u>GTGATTCCTCCCAATTTTCT</u> TATTACACAAATAAGA	GTGCTCTTCCGATCTAA ACTCTTCCACTTTTAGTG CATCAA
13	rs11036364	5227774	AATCGGGAAGCTGAAG TTAGTTACTTATTAGGT TTTGGGAAACAAG	<u>CACTACCGTCGGATGAC</u> <u>ATATTTTTGACTGCAT</u> TAAGAGGTCTCTAGT
14	rs7936823	5228938	<u>ATGTCATCCGACGGTA</u> <u>GTGATGAAAAGACGTCT</u> TTAAAATATTTTAAATG TT	GTGCTCTTCCGATCTAA AATCTAACAGCCAAGTC AAATCTG
15	rs3813727	5234682	AATCGGGAAGCTGAAG CTGTCATGTGTGTCTTG ACTCAGAA	<u>CACTACCGTCGGATGAC</u> <u>ATATAGATGCGGTGGGG</u> AGATATG
16	rs10837643	5236808	<u>ATGTCATCCGACGGTA</u> <u>GTGATAAGGCCATGAGG</u> ACTGTTATTTG	GTGCTCTTCCGATCTAG CCAAGTCTTTGGAATTA ACAGAC
17	rs4283007	5237260	AATCGGGAAGCTGAAG CCGCATCCAGCCAGGA TA	<u>CACTACCGTCGGATGAC</u> <u>ATATAACAAGAGTAGATG</u> CTATTCCTTCCACTTT
18	rs4402323	5237362	<u>ATGTCATCCGACGGTA</u> <u>GTGATAGATAGGGACA</u> AATTGAAGCAGAA	GTGCTCTTCCGATCTCA TTCCCTACCTCTAATCA ACAGTT
19	rs4910736	5237759	<u>ATGTCATCCGACGGTA</u> <u>GTGATAGAAGTATGAAGT</u> AAGACATAGTCCCCAA	GTGCTCTTCCGATCTATT AAAATCGATACTGAGTT CTAAAATCATC
20	rs2105819	5238497	AATCGGGAAGCTGAAG GCTCTCCACAATTCTAA TCTCAGTTC	<u>CACTACCGTCGGATGAC</u> <u>ATATATTGATAAGATAAT</u> ATGAAAACACAGAATTC A

ФИГ. 20 (продолж.)



ФИГ. 21А



ФИГ. 21В

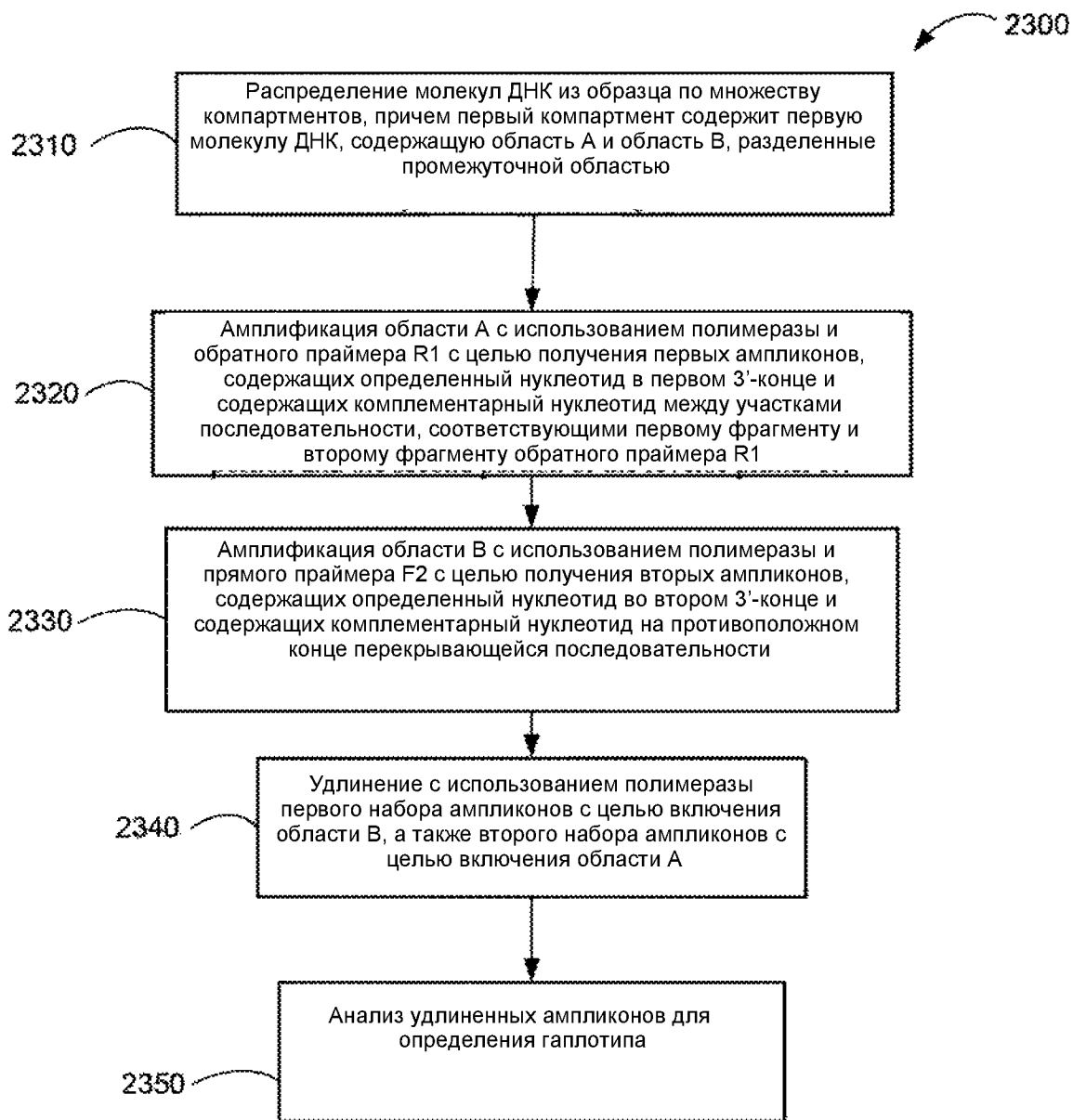
	№ ОНП	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Семья 1																					
Отец	Нар I*	C	A	C	A	C	T	G	A	C	A	G	G	A	T	A	C	G	A	G	C
	Нар II	C	G	G	A	C	A	G	A	A	A	G	A	G	A	T	A	C	A	G	T
Мать	Нар I	C	G	G	A	C	A	G	A	A	A	A	A	G	A	T	A	C	A	G	T
	Нар II*	G	A	G	G	G	T	T	G	A	G	A	G	A	T	A	C	G	A	C	C
Плод	Гаплотип от отца	C	A	C	A	C	T	G	A	C	A	G	G	A	T	A	C	G	A	G	C
	Гаплотип от матери	G	A	G	G	G	T	T	G	A	G	A	G	A	T	A	C	G	A	C	C
Семья 2																					
Отец	Нар I	G	A	G	G	G	T	T	G	A	G	A	A	G	A	T	A	C	A	C	T
	Нар II*	G	A	G	G	G	T	T	G	A	G	A	G	A	T	A	C	G	A	C	C
Мать	Нар I*	C	G	G	A	C	A	G	A	A	A	G	G	A	T	A	C	G	A	G	C
	Нар II	G	A	G	G	G	T	T	G	A	G	A	A	G	A	T	A	C	A	C	T
Плод	Гаплотип от отца	G	A	G	G	G	T	T	G	A	G	A	G	A	T	A	C	G	A	C	C
	Гаплотип от матери	C	G	G	A	C	A	G	A	A	A	G	G	A	T	A	C	G	A	G	C
Семья 3																					
Отец	Нар I	C	A	C	A	C	T	G	A	C	A	G	A	G	A	T	A	C	A	G	T
	Нар II*	C	A	G	A	C	T	G	A	C	A	G	G	A	T	A	C	G	A	G	C
Мать	Нар I	C	A	C	A	C	T	G	A	C	A	G	G	A	T	A	C	G	A	G	C
	Нар II*	G	A	G	G	G	T	T	G	A	G	G	G	A	T	A	C	G	A	C	C
Плод	Гаплотип от отца	C	A	G	A	C	T	G	A	C	A	G	G	A	T	A	C	G	A	G	C
	Гаплотип от матери	G	A	G	G	G	T	T	G	A	G	G	G	A	T	A	C	G	A	C	C

ФИГ. 22

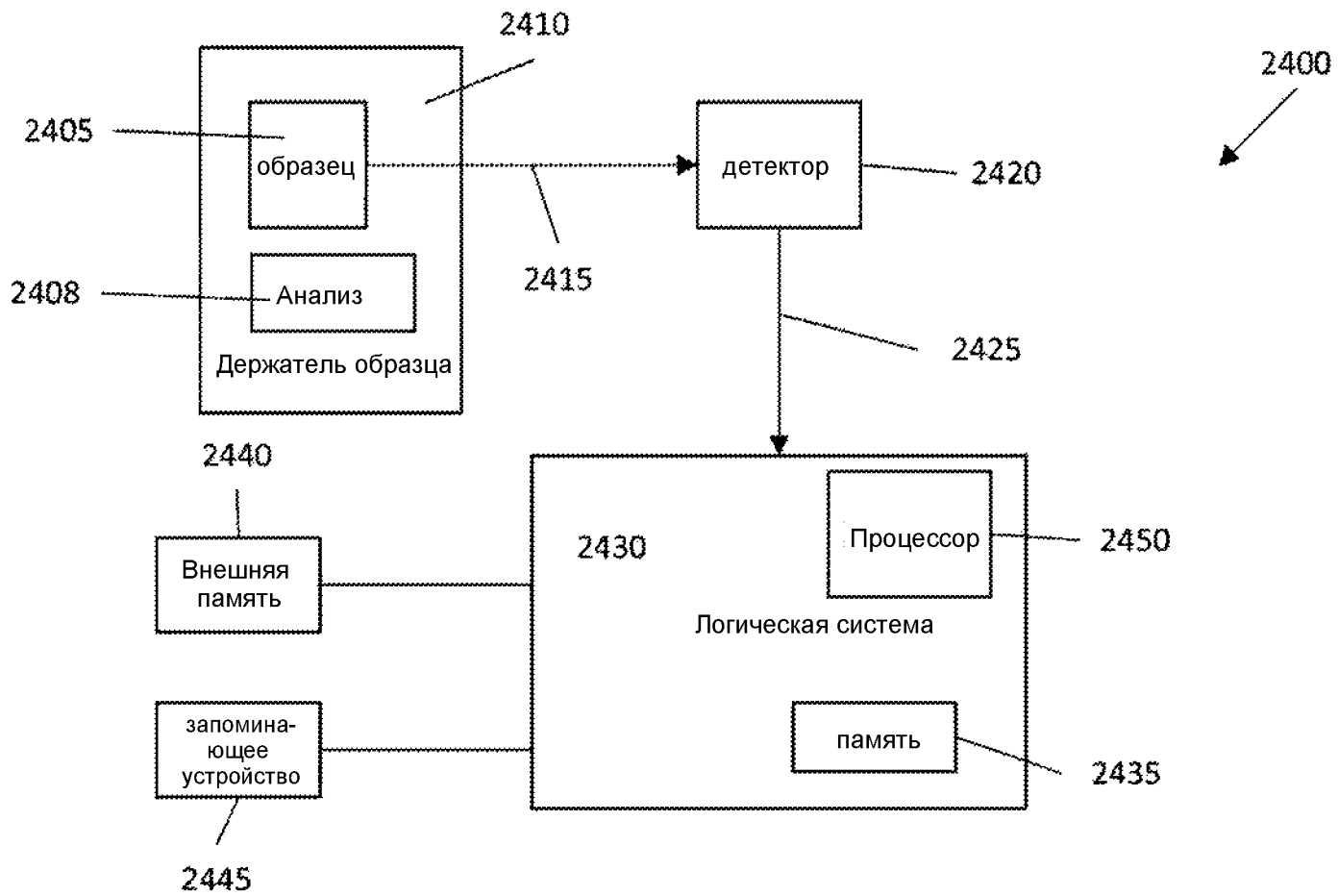


	№ ОНП	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
		Семья 4																			
Отец	Нар I*	G	A	G	G	G	T	T	G	A	G	A	A	G	A	T	A	C	A	C	T
	Нар II*	G	A	G	G	G	T	T	G	A	G	A	A	G	A	T	A	C	A	C	T
Мать	Нар I*	C	G	G	A	C	A	G	A	A	A	A	A	G	A	T	A	C	A	G	T
	Нар II	G	A	G	G	G	T	T	A	A	G	A	A	A	T	A	C	G	A	C	C
Плод	Гапло-тип от отца	G	A	G	G	G	T	T	G	A	G	A	A	G	A	T	A	C	A	C	T
	Гапло-тип от матери	C	G	G	A	C	A	G	A	A	A	A	A	G	A	T	A	C	A	G	T

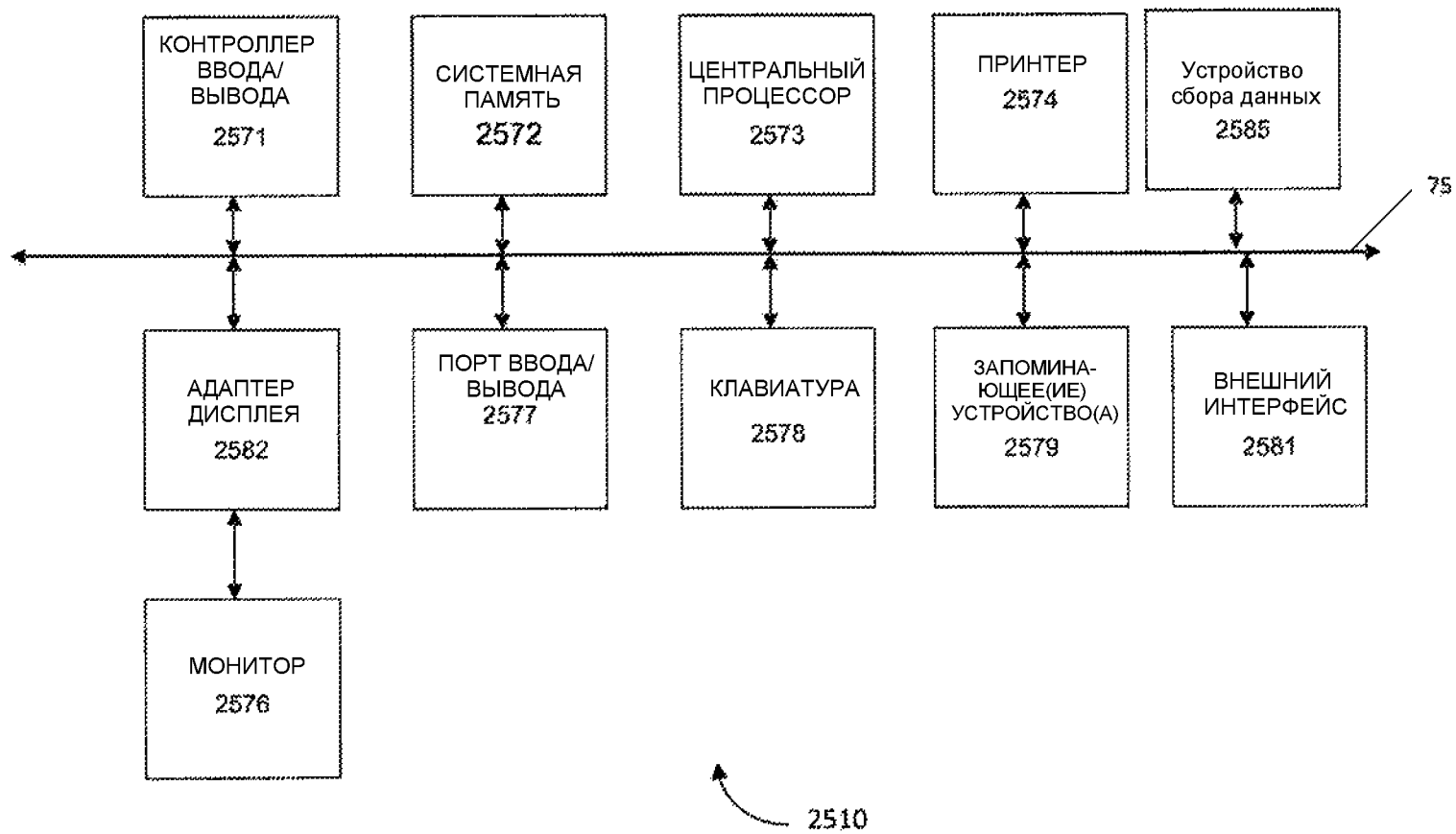
ФИГ. 22 (продолж.)



ФИГ. 23



ФИГ. 24



ФИГ. 25

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference <b>20C51387</b>	<b>FOR FURTHER ACTION</b>	see Form PCT/ISA/220 as well as, where applicable, item 5 below.
International application No. <b>PCT/CN2020/089560</b>	International filing date ( <i>day/month/year</i> ) <b>11 May 2020</b>	(Earliest) Priority Date ( <i>day/month/year</i> ) <b>10 May 2019</b>
Applicant <b>THE CHINESE UNIVERSITY OF HONG KONG</b>		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 4 sheets.

It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

a. With regard to the **language**, the international search was carried out on the basis of:

the international application in the language in which it was filed.

a translation of the international application into \_\_\_\_\_ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

b.  This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6bis(a)).

c.  With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2.  **Certain claims were found unsearchable** (see Box No. II).

3.  **Unity of invention is lacking** (see Box No. III).

4. With regard to the **title**,

the text is approved as submitted by the applicant.

the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

the text is approved as submitted by the applicant.

the text has been established, according to Rule 38.2, by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. With regard to the **drawings**,

a. the figure of the **drawings** to be published with the abstract is Figure No. \_\_\_\_\_

as suggested by the applicant.

as selected by this Authority, because the applicant failed to suggest a figure.

as selected by this Authority, because this figure better characterizes the invention.

b.  none of the figures is to be published with the abstract.

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/089560

<b>A. CLASSIFICATION OF SUBJECT MATTER</b>		
C12Q 1/68(2018.01)i		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols)		
C12Q		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
CNABS, SIPOABS, DWPI, VEN, USTXT, WOTXT, CNKI, NCBI, baidu, patents, incopat, wanfang, the Chinese University of HONG KONG, CHAN Kwan Chee, GAI Wanxia, LO Yuk-Ming Dennis, NUDT15, nudix hydrolase 15, HBB, hemoglobin subunit beta, two regions, pcr, overhang, overlapping, exon, snp, R1: CCAGGGGAGGTAGTCTTCCCACGCAACGCGGATGCT, F2: GGGAAGAACTACCTCCCCTGG, F1: TATGACGGCCAGCGCAC, R2: CCACCAGATGGTTCAGATCTTCT, A, B, C, D, Fz, Rz, F3, F4, TATGACGGCCAGCGCAC, ACGCAACGCGGATGCT, CCTCCCCTGGACCAGCTT, CCACCAGATGGTTCAGATCTTCT, AATGATACGGCCGACCACCGAGATCTACACATACGAGATCCGT, CAAGCAGAAGACGGCATAACGAGAT-index-GTGACTGGAGTTC, custom read 1 primer: CATACGAGATCCGTAATCGGGAAGCTGAAG, Illumina Adapter Sequences.		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KADKHODAEI, S.et al. "Multiple overlap extension PCR (MOE-PCR): an effective technical shortcut to high throughput synthetic biology." <i>RSC Adv.</i> , Vol. 6, 08 July 2016 (2016-07-08), pages 66682-66694	1-43
A	WANESKOG, M.et al. "Multi-fragment site-directed mutagenic overlap extension polymerase chain reaction as a competitive alternative to the enzymatic assembly method." <i>Analytical Biochemistry.</i> , Vol. 444, 29 September 2013 (2013-09-29), pages 32-37	1-43
A	US 2008/0194418 A1 (JOHNSON, E.A.et al.) 14 August 2008 (2008-08-14) the whole document	1-43
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
08 July 2020		10 August 2020
Name and mailing address of the ISA/CN		Authorized officer
National Intellectual Property Administration, PRC 6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088 China		ZHU,Ning
Facsimile No. (86-10)62019451		Telephone No. 86-(10)-53961956

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/089560

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2003/0190634 A1 (BARANY, F.et al.) 09 October 2003 (2003-10-09) the whole document	1-43
A	GEU-FLORES, F.et al. "USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products." <i>Nucleic Acids Research</i> , Vol. 35, No. 7, 27 March 2007 (2007-03-27), e55, pages 1-6	1-43
A	CANTSILIERIS, S.et al. "Targeted Capture and High-Throughput Sequencing Using Molecular Inversion Probes (MIPs)." <i>Methods Mol Biol</i> , Vol. 1492, 01 July 2017 (2017-07-01), pages 95-106, Author manuscript, pages 1-11	1-43
A	ILLUMINA, INC. "Illumina Adapter Sequences." <i>Illumina Document # 1000000002694 v01.</i> , 28 February 2016 (2016-02-28), pages 1, 35	1-43

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/CN2020/089560**

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
US	2008/0194418	A1	14 August 2008	AU	2006244042	B2	29 July 2010
				EP	1885882	A2	13 February 2008
				US	9365893	B2	14 June 2016
				NZ	564095	A	26 June 2009
				DE	602006019855	D1	10 March 2011
				AU	2006244042	A1	16 November 2006
				EP	1885882	B1	26 January 2011
				WO	2006122215	A2	16 November 2006
				WO	2006122215	A3	22 March 2007
				CA	2612145	A1	16 November 2006
				DK	1885882	T3	11 April 2011
				AT	497020	T	15 February 2011
US	2003/0190634	A1	09 October 2003	US	8367322	B2	05 February 2013