

(19)



**Евразийское
патентное
ведомство**

(21) **202292332** (13) **A1**

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

(43) Дата публикации заявки
2022.11.17

(51) Int. Cl. *C12Q 1/686* (2018.01)

(22) Дата подачи заявки
2021.02.12

(54) **СПОСОБЫ И МАТЕРИАЛЫ ДЛЯ ОЦЕНКИ НУКЛЕИНОВЫХ КИСЛОТ**

(31) 62/977,066

(72) Изобретатель:

(32) 2020.02.14

Пападопулос Николас, Кинзлер

(33) US

Кеннет В., Вогельштейн Берт, Коэн

(86) PCT/US2021/017937

Джошуа Дэвид (US)

(87) WO 2021/163546 2021.08.19

(74) Представитель:

(71) Заявитель:

Медведев В.Н. (RU)

**ДЗЕ ДЖОНС ХОПКИНС
ЮНИВЕРСИТИ (US)**

(57) В изобретении предложены системы, наборы, композиции и способы для получения библиотек для секвенирования, а также рабочий процесс секвенирования (например, для идентификации мутаций). В некоторых вариантах осуществления в настоящем документе предложены системы и способы идентичного баркодирования обеих цепей матриц, а также основанного на методе ПЦР обогащения каждой цепи, которое не требует гибридизационного захвата.

202292332

A1

A1

202292332

ОПИСАНИЕ ИЗОБРЕТЕНИЯ

2420-575308EA/081

СПОСОБЫ И МАТЕРИАЛЫ ДЛЯ ОЦЕНКИ НУКЛЕИНОВЫХ КИСЛОТ

Настоящая заявка испрашивает приоритет по предварительной заявке на патент США 62/977,066, поданной 14 февраля 2020 года, которая включена в настоящий документ посредством отсылки во всей своей полноте.

ЗАЯВЛЕНИЕ О ФИНАНСИРОВАНИИ ИЗ ФЕДЕРАЛЬНОГО БЮДЖЕТА

Настоящее изобретение было сделано при правительственной поддержке по грантам CA062924, CA152753 и CA230691, выданным Национальными Институтами Здравоохранения США. Правительство обладает некоторыми правами на изобретение.

ОБЛАСТЬ ТЕХНИКИ, К КОТОРОЙ ОТНОСИТСЯ ИЗОБРЕТЕНИЕ

Настоящее изобретение относится к области секвенирования нуклеиновых кислот. В частности, изобретение относится к получению библиотек для секвенирования и рабочему процессу секвенирования для идентификации мутаций.

УРОВЕНЬ ТЕХНИКИ

Идентификация редких мутаций полезна в некоторых аспектах фундаментальной биологии, а также для улучшения клинического ведения пациентов. Области применения включают инфекционные заболевания, исследование иммунного репертуара, палентогенетику, судебную экспертизу, старение, неинвазивное пренатальное тестирование и рак. Технологии секвенирования следующего поколения (NGS) теоретически подходят для этого применения, и существует множество подходов NGS для обнаружения редких мутаций. Однако в случае традиционных подходов NGS частота ошибок самого секвенирования слишком высока, чтобы можно было надежно обнаруживать мутации, особенно такие мутации, которые присутствуют с низкой частотой в исходном образце.

Применение молекулярных штрих-кодов для мечения исходных молекул-матриц было разработано для преодоления различных препятствий при обнаружении редких мутаций. С помощью молекулярного баркодирования выполняют избыточное секвенирование ПЦР-генерируемого потомства каждой меченой молекулы, при этом ошибки секвенирования можно легко выявлять (Kinde et al., Proc Natl Acad Sci U S A 108:9530-9535 (2011)). Например, если заданный порог потомства баркодированной матричной молекулы содержит одну и ту же мутацию, то мутация считается подлинной ("супермутант"). Если меньше заданного порога потомства содержат представляющую интерес мутацию, то мутация считается артефактом. Описаны два типа молекулярных штрих-кодов: экзогенные и эндогенные. Экзогенные штрих-коды (также именуемые в настоящем документе экзогенными UID) содержат заданные или случайные нуклеотиды и добавляются при создании библиотеки или во время ПЦР. Эндогенные штрих-коды (также именуемые в настоящем документе эндогенными UID) формируются из последовательностей на 5'- и 3'-концах фрагментов (например, фрагментов, получаемых в результате случайной фрагментации ДНК, или фрагментов, присутствующих в

бесклеточном образце биологической жидкости) ДНК-матрицы, подлежащей анализу. Было доказано, что такие штрих-коды могут применяться для отслеживания ампликонов до первичной исходной матрицы, что позволяет проводить молекулярный подсчет и улучшать идентификацию истинных мутаций в клинически значимых образцах.

Разветвленные адаптеры для секвенирования спаренных концов были разработаны для проведения "дуплексного секвенирования", при котором каждую из двух цепей (Уотсона и Крика) исходного ДНК-дуплекса можно различать по направленности 5'→3', выявленной при секвенировании. Дуплексное секвенирование снижает количество ошибок секвенирования, потому что обе цепи ДНК с крайне малой вероятностью будут содержать идентичную мутацию, если эта мутация была ошибочно введена при получении библиотеки или секвенировании.

Однако существует множество проблем, которые ограничивают возможности применения и клиническую применимость молекулярного баркодирования. Например, сложно превратить большую часть исходных молекул-матриц в дуплексные молекулы с одинаковым штрих-кодом на каждой цепи (Schmitt et al., Proc Natl Acad Sci U S A 109:14508-14513 (2012); Schmitt et al., Nat Methods 12:423-425 (2015); и Newman et al., Nat Biotechnol 34:547-555 (2016)). Эта проблема представляет особую сложность, когда количество исходной ДНК ограничено (например, <33 нг), например, как в случае ДНК из бесклеточной плазмы, обычно используемой для жидкостной биопсии.

Получение библиотек для направленного секвенирования обычно включает присоединение адаптера к матрицам секвенирования, амплификацию библиотеки и гибридизационный захват для обогащения библиотеки представляющими интерес мишенями. Хотя гибридизационный захват эффективен для обогащения больших областей, представляющих интерес, он не подходит для небольших областей-мишеней (Springer et al., Elife 7:doi:10.7554/eLife.32143 (2018)) и демонстрирует плохое выделение дуплекса (Wang et al., Proc Natl Acad Sci U S A 112:9704-9709 (2015) и Wang et al., Elife 5:doi:10.7554/eLife.15175 (2016)). Последовательные раунды захвата могут частично преодолеть эти ограничения, но даже с такими улучшениями процент выделения дуплекса обычно составляет ~1%. CRISPR-DS может обеспечить выделение до 15%, но данный метод не удастся применять к внеклеточной ДНК. Если область-мишень очень мала (например, одно или несколько положений в геноме, представляющих особый интерес, например, требуемые для мониторинга заболевания в плазме крови) или количество доступной ДНК ограничено (например, <33 нг, что часто обнаруживается в плазме), подходы на основе захвата неэффективны.

Таким образом, существует потребность в улучшении получения библиотек секвенирования и рабочего процесса секвенирования, чтобы можно было обеспечить точную идентификацию мутаций, например редких мутаций, в клинически значимых образцах, таких как образцы жидкостной биопсии.

СУЩНОСТЬ ИЗОБРЕТЕНИЯ

В настоящем документе предложены способы и материалы, которые направлены

на решение указанных проблем путем предоставления подхода к идентичному баркодированию обеих цепей матриц и предоставления способа обогащения каждой цепи на основе метода ПЦР, который не требует гибридизационного захвата.

Данный документ относится к способам и материалам, которые могут применяться для обнаружения присутствия одной или более мутаций, присутствующих на обеих цепях двухцепочечной нуклеиновой кислоты (например, ДНК). В некоторых случаях способы обнаружения одной или более мутаций, присутствующих на обеих цепях двухцепочечной нуклеиновой кислоты, могут включать создание библиотеки дуплексов для секвенирования, имеющей дуплексный молекулярный штрих-код на каждом конце (например, 5'-конце и 3'-конце) каждой нуклеиновой кислоты в библиотеке, с получением библиотеки одноцепочечных последовательностей, полученных из цепей Уотсона, и библиотеки одноцепочечных последовательностей, полученных из цепей Крика, из библиотеки дуплексов для секвенирования, и обнаружение присутствия одной или более мутаций, присутствующих на обеих цепях двухцепочечной нуклеиновой кислоты в каждой одноцепочечной библиотеке.

Как продемонстрировано в настоящем документе, библиотека одноцепочечных ДНК, соответствующая цепям Уотсона матрицы двухцепочечной нуклеиновой кислоты, и библиотека одноцепочечных ДНК, соответствующая цепям Крика матрицы двухцепочечной нуклеиновой кислоты, могут быть созданы из библиотек секвенирования, содержащих дуплексные молекулярные штрих-коды, при этом каждая библиотека одноцепочечных ДНК может быть обогащена по отношению к области-мишени с использованием метода цепь-специфической якорной ПЦР, и область-мишени может быть секвенирована для обнаружения присутствия или отсутствия одной или более мутаций в области-мишени нуклеиновой кислоты. Например, способы и материалы, описанные в настоящем документе, которые могут применяться для обнаружения присутствия одной или более мутаций, присутствующих на обеих цепях двухцепочечной нуклеиновой кислоты, могут именоваться Системой секвенирования с точным определением последовательности без ошибок (от англ. - Sequence Ascertainment Free of Errors Sequencing System, SaferSeqS) и могут включать, например, конструирование библиотеки с созданием *in situ* двухцепочечных молекулярных штрих-кодов (см., например, Фиг. 22а), обогащение мишени с применением якорной ПЦР (см., например, Фиг. 22b) и реконструирование *in silico* молекул-матриц (см., например, Фиг. 22с). *Vona fide* мутации, присутствующие в исходных первичных матрицах, могут быть идентифицированы путем обнаружения изменений на обеих цепях одной и той же исходной молекулы нуклеиновой кислоты.

Возможность обнаружения одной или более мутаций, присутствующих в обеих цепях двухцепочечной нуклеиновой кислоты (например, достоверных соматических мутаций), предоставляет уникальную и нереализованную возможность точной и эффективной оценки множественных мутаций одновременно и по доступной цене. Применение способов и материалов, описанных в настоящем документе, для обнаружения

присутствия одной или более мутаций, присутствующих в обеих цепях двухцепочечной нуклеиновой кислоты (например, способов SaferSeqS), может позволить проводить идентификацию редких мутаций с высокой степенью достоверности при минимизации степени повреждения ДНК, объема проводимых реакций ПЦР и/или количества артефактов с повреждениями ДНК. Следует отметить, что термины "цепь Уотсона" и "цепь Крика" используются просто, чтобы различать две цепи двухцепочечной исходной последовательности нуклеиновой кислоты. Любую из цепей можно обозначить как "Уотсон" или "Крик", а затем другую цепь обозначить другим именем.

В некоторых вариантах осуществления в настоящем документе предложены способы, включающие: а) формирование реакционной смеси, включающей: i) множество фрагментов двухцепочечной ДНК, которые дефосфорилированы и имеют тупые концы, где каждый из фрагментов двухцепочечной ДНК включает цепи Уотсона и Крика; ii) множество адаптеров, где каждый из адаптеров содержит в направлении 5'→3': А) штрих-код и В) универсальную 3'-адаптерную последовательность; и iii) лигазу; и б) инкубирование реакционной смеси таким образом, чтобы: i) адаптеры были лигированы с 3'-концами цепей Уотсона и Крика, и ii) адаптеры не были лигированы с 5'-концами цепей Уотсона или Крика, с получением в результате двухцепочечных продуктов лигирования.

В некоторых вариантах осуществления каждый из множества адаптеров содержит уникальный штрих-код. В дополнительных вариантах осуществления каждый из двухцепочечных продуктов лигирования содержит цепь Уотсона только с одним штрих-кодом и цепь Крика только с одним штрих-кодом, который отличается от штрих-кода на цепи Уотсона. В дополнительных вариантах осуществления способы дополнительно включают этап: с) секвенирования по меньшей мере части двухцепочечных продуктов лигирования.

В конкретных вариантах осуществления в настоящем документе предложены способы, включающие: а) присоединение частично двухцепочечных 3'-адаптеров (3'PDSA) к 3'-концам цепей Уотсона и Крика популяции фрагментов двухцепочечной ДНК в анализируемом образце ДНК, где первая цепь 3'PDSA-адаптеров включает в направлении 5'→3': (i) первый сегмент, (ii) экзогенную UID последовательность, (iii) сайт отжига для 5'-адаптера и (iv) универсальную 3'-адаптерную последовательность, включающую сайт праймера для секвенирования R2, и где вторая цепь 3'PDSA включает в направлении 5'→3': (i) сегмент, комплементарный первому сегменту, и, необязательно, (ii) 3'-блокирующую группу, б) отжиг 5'-адаптеров на сайте отжига, где 5'-адаптеры включают в направлении 5'→3' (i) универсальную 5'-адаптерную последовательность, которая не комплементарна универсальной 3'-адаптерной последовательности и которая включает сайт праймера для секвенирования R1, и (ii) последовательность, комплементарную сайту отжига для 5'-адаптера; с) удлинение 5'-адаптеров через экзогенную UID последовательность и первый сегмент, с получением в результате комплемента экзогенной последовательности UID и комплемента первого сегмента, и d) ковалентное связывание 3'-конца комплемента первого сегмента с 5'-концами цепей

Уотсона и Крика двухцепочечных ДНК-фрагментов, с образование в результате множества лигированных с адаптером фрагментов двухцепочечной ДНК.

В некоторых вариантах осуществления в настоящем документе предложены способы, включающие: а) присоединение частично двухцепочечных 3'-адаптеров (3'PDSA) к 3'-концам цепей Уотсона и Крика популяции фрагментов двухцепочечной ДНК в анализируемом образце ДНК, где первая цепь 3'PDSA включает в направлении 5'→3': (i) первый сегмент, (ii) экзогенную UID последовательность, (iii) сайт отжига для 5'-адаптера и (iv) универсальную 3'-адаптерную последовательность, включающую сайт праймера для секвенирования R2, и где вторая цепь 3'PDSA включает в направлении 5'→3': (i) сегмент, комплементарный первому сегменту, и необязательно, (ii) 3'-блокирующую группу, б) отжиг 5'-адаптеров на сайте отжига, где 5'-адаптеры включают в направлении 5'→3': (i) универсальную последовательность 5'-адаптера, которая не комплементарна универсальной последовательности 3'-адаптера, которая включает сайт праймера для секвенирования R1, и (ii) последовательность, комплементарную сайту отжига для 5'-адаптера; в) удлинение 5'-адаптеров через экзогенную UID последовательность, с получением в результате комплемента экзогенной последовательности UID, и d) ковалентное связывание 3'-конца комплемента экзогенной последовательности UID с 5'-концами сегмента, комплементарного первому сегменту на каждой из цепей Уотсона и Крика фрагментов двухцепочечной ДНК, с образованием в результате множества лигированных с адаптером фрагментов двухцепочечной ДНК.

В некоторых вариантах осуществления в настоящем документе предложены способы, включающие: а) присоединение частично двухцепочечных 3'-адаптеров (3'PDSA) к 3'-концам цепей Уотсона и Крика популяции фрагментов двухцепочечной ДНК в анализируемом образце ДНК, где первая цепь 3'PDSA включает в направлении 5'→3': (i) первый сегмент, (ii) экзогенную UID последовательность, (iii) сайт отжига для 5'-адаптера и (iv) универсальную 3'-адаптерную последовательность, включающую сайт праймера для секвенирования R2, и где вторая цепь 3'PDSA включает в направлении 5'→3': (i) сегмент, комплементарный первому сегменту, и необязательно, (ii) 3'-блокирующую группу, б) отжиг 5'-адаптеров на сайте отжига, где 5'-адаптеры включают в направлении 5'→3': (i) универсальную 5'-адаптерную последовательность, которая не комплементарна универсальной 3'-адаптерной последовательности, и которая включает сайт праймера для секвенирования R1, и (ii) последовательность, комплементарную сайту отжига для 5'-адаптера; в) удлинение 5'-адаптеров через экзогенную UID последовательность и первый сегмент 3'PDSA, с получением в результате комплемента экзогенной последовательности UID и комплемента первого сегмента 3'PDSA, и d) ковалентное связывание 3'-конца комплемента первого сегмента 3'PDSA с 5'-концами цепей Уотсона и Крика фрагментов двухцепочечной ДНК, в результате чего образуется множество лигированных с адаптером фрагментов двухцепочечной ДНК.

В некоторых вариантах осуществления в настоящем документе предложены системы, наборы и композиции, включающие: а) популяцию частично двухцепочечных 3'-

адаптеров (3'PDSA), которые обеспечивают возможность лигирования с 3'-концами цепей Уотсона и Крика популяции двухцепочечных ДНК-фрагментов, где первая цепь 3'PDSA включает в направлении 5'→3': (i) первый сегмент, (ii) экзогенную UID последовательность, (iii) сайт отжига для 5'-адаптера и (iv) универсальную 3'-адаптерную последовательность, включающую сайт праймера для секвенирования R2, и где вторая цепь 3'PDSA включает в направлении 5'→3': (i) сегмент, комплементарный первому сегменту, и (ii) 3'-блокирующую группу; и б) популяцию 5'-адаптеров, которые обеспечивают возможность отжига с сайтом отжига, где 5'-адаптеры включают в направлении 5'→3': (i) универсальную последовательность 5'-адаптера, которая не комплементарна универсальной 3'-адаптерной последовательности, и которая включает сайт праймера для секвенирования R1, и (ii) последовательность, комплементарную сайту отжига для 3'-адаптера.

В других вариантах осуществления системы, наборы и композиции дополнительно включают: с) популяцию фрагментов двухцепочечной ДНК из биологического образца и/или с) реагенты для расщепления второй цепи 3'PDSA с образованием одноцепочечных 3'-адаптеров (3'SSA); и/или с) первый праймер, комплементарный универсальной 3'-адаптерной последовательности, и второй праймер, комплементарный универсальной 5'-адаптерной последовательности; и/или с) систему секвенирования; и/или с) якорный праймер для амплификации цепи Уотсона, комплементарный универсальной 3'-адаптерной последовательности, и d) якорный праймер для амплификации цепи Крика, комплементарный комплементу универсальной 5'-адаптерной последовательности; и/или с) первый набор пар мишень-селективных праймеров Уотсона, включающий: (i) один или больше первых мишень-селективных праймеров Уотсона, включающих последовательность, комплементарную части универсальной 3'-адаптерной последовательности, и (ii) один или больше вторых мишень-селективных праймеров Уотсона, причем каждый из одного или больше вторых мишень-селективных праймеров Уотсона включает мишень-селективную последовательность; и/или с) первый набор пар мишень-селективных праймеров Крика, включающий: (i) один или больше мишень-селективных праймеров Крика, включающих последовательность, комплементарную части универсальной 5'-адаптерной последовательности, и (ii) один или больше вторых мишень-селективных праймеров Крика, причем каждый из одного или больше вторых мишень-селективных праймеров Крика включает такую же селективную последовательность, как и последовательность второго мишень-селективного праймера Уотсона.

В некоторых вариантах осуществления способы дополнительно включают: амплификацию множества лигированных с адаптером фрагментов двухцепочечной ДНК при использовании первого праймера, комплементарного универсальной 3'-адаптерной последовательности, и второго праймера, комплементарного универсальной 5'-адаптерной последовательности, с получением в результате ампликонов, где ампликоны включают множество двухцепочечных матриц Уотсона и множество двухцепочечных матриц Крика.

В некоторых вариантах осуществления способы дополнительно включают: селективную амплификацию указанных двухцепочечных матриц Уотсона при использовании первого набора пар мишень-селективных праймеров Уотсона, причем первый набор пар мишень-селективных праймеров Уотсона включает: (i) первый мишень-селективный праймер Уотсона, включающий последовательность, комплементарную части универсальной 3'-адаптерной последовательности, и (ii) второй мишень-селективный праймер Уотсона, включающий мишень-селективную последовательность, с получением в результате продуктов амплификации мишени Уотсона. В дополнительных вариантах осуществления способы дополнительно включают: селективную амплификацию указанных двухцепочечных матриц Крика при использовании первого набора пар мишень-селективных праймеров Крика, причем первый набор пар праймеров Крика, селективных в отношении мишени, включает: (i) первый мишень-селективный праймер для амплификации цепи Крика, включающий последовательность, комплементарную комплементу части универсальной 5'-адаптерной последовательности, и (ii) второй мишень-селективный праймер для амплификации цепи Крика, включающий такую же мишень-селективную последовательность, что и последовательность второго мишень-селективного праймера Уотсона, с получением в результате продуктов амплификации мишени Крика. В некоторых вариантах осуществления популяцию фрагментов двухцепочечной ДНК инкубируют со смесью урацил-ДНК-гликозилазы и ДНК-гликозилазы-лиазы эндонуклеазы VIII перед лигированием каких-либо адаптеров.

В некоторых вариантах осуществления используемая полимеразы (например, для удлинения 5'-адаптерных последовательностей) обладает 5'→3'-экзонуклеазной активностью (например, может расщеплять вторую цепь 3'PDSA). В других вариантах осуществления используемая полимеразы (например, для удлинения 5'-адаптерных последовательностей) не обладает 5'→3'-экзонуклеазной активностью.

В других вариантах осуществления способы дополнительно включают: удаление второй цепи 3'PDSA с получением одноцепочечных 3'-адаптеров (3'SSA). В других вариантах осуществления удаление второй цепи происходит после этапа b) или перед этапом b), или во время этапа b). В некоторых вариантах осуществления удаление второй цепи 3'-PDSA включает контакт 3'-дуплексного адаптера с урацил-ДНК-гликозилазой (УДГ) для расщепления второй цепи. В дополнительных вариантах осуществления удаление второй цепи производят с помощью полимеразы с экзонуклеазной активностью, где полимеразы удлиняет 5'-адаптеры через экзогенную UID последовательность и первый сегмент.

В других вариантах осуществления способы дополнительно включают: определение чтений последовательности одного или более ампликонов. В других вариантах осуществления способы дополнительно включают: отнесение считанных последовательностей к семействам UID, где каждый член семейства UID включает одну и ту же экзогенную UID последовательность. В конкретных вариантах осуществления способы дополнительно включают: отнесение чтений последовательностей каждого

семейства UID к подсемейству Уотсона и подсемейству Крика на основе пространственного отношения экзогенной последовательности UID к последовательностям чтений R1 и R2. В других вариантах осуществления способы дополнительно включают: идентификацию нуклеотидной последовательности в качестве точно представленной цепи Уотсона анализируемого фрагмента ДНК, если по меньшей мере 50% (например, 50...75...95%) подсемейства Уотсона содержит данную последовательность. В других вариантах осуществления способы дополнительно включают: идентификацию нуклеотидной последовательности в качестве точно представленной цепи Крика анализируемого фрагмента ДНК, если по меньшей мере 50% (например, 50...75...90%) подсемейства Крика содержит данную последовательность.

В некоторых вариантах осуществления способы дополнительно включают: идентификацию мутации в нуклеотидной последовательности в виде точно представленной цепи Уотсона, где последовательность, точно представляющая цепь Уотсона, отличается от референсной последовательности, в которой такая мутация отсутствует. В дополнительных вариантах осуществления способы дополнительно включают: идентификацию мутации в нуклеотидной последовательности в виде точно представленной цепи Крика, где последовательность, точно представляющая цепь Крика, отличается от референсной последовательности, в которой такая мутация отсутствует. В других вариантах осуществления способы дополнительно включают: идентификацию мутации в анализируемом фрагменте ДНК, где мутация в нуклеотидной последовательности, точно представляющей цепь Уотсона, и мутация в нуклеотидной последовательности, точно представляющей цепь Крика, представляют собой одну и ту же мутацию. В некоторых вариантах осуществления каждый член семейства UID дополнительно включает одну и ту же эндогенную UID последовательность, где эндогенная UID последовательность включает конец фрагмента двухцепочечной ДНК из популяции. В других вариантах осуществления популяция фрагментов двухцепочечной ДНК имеет тупые концы.

В настоящем документе предложен способ обнаружения присутствия или отсутствия мутации в области-мишени двухцепочечной ДНК-матрицы, полученной из образца млекопитающего, и определения, присутствует ли мутация в обеих цепях двухцепочечной ДНК-матрицы, где способ включает: А) создание фрагментов двухцепочечной ДНК, каждый из которых имеет дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК; В) амплификацию фрагмента двухцепочечной ДНК, включающего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, с получением библиотеки амплифицированных дуплексов для секвенирования, где амплификация включает контакт фрагмента двухцепочечной ДНК, включающего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК с парой универсальных праймеров в условиях полногеномной ПЦР; С) необязательно, создание библиотеки одноцепочечных ДНК цепей Уотсона из библиотеки амплифицированных дуплексов для секвенирования; D)

необязательно, создание библиотеки одноцепочечных ДНК цепей Крика из библиотеки амплифицированных дуплексов для секвенирования; E) амплификацию области-мишени из библиотеки ДНК (например, библиотеки одноцепочечных ДНК) цепей Уотсона при использовании пары праймеров, состоящей из первого праймера, способного к гибридизации с областью-мишенью, и второго праймера, способного к гибридизации с 3'-дуплексным адаптером; F) амплификацию области-мишени из библиотеки ДНК (например, библиотеки одноцепочечных ДНК) цепей Крика при использовании пары праймеров, состоящей из первого праймера, способного к гибридизации с областью-мишенью, и второго праймера, способного к гибридизации с 5'-адаптером ; G) секвенирование области-мишени, амплифицированной из библиотеки ДНК (например, библиотеки одноцепочечных ДНК) (например, библиотеки ДНК (например, библиотеки одноцепочечных ДНК)) цепей Уотсона, с получением чтения секвенирования и обнаружением присутствия или отсутствия мутации в цепи Уотсона области-мишени; H) секвенирование области-мишени, амплифицированной из библиотеки ДНК (например, библиотеки одноцепочечных ДНК) (например, библиотеки ДНК (например, библиотеки одноцепочечных ДНК)) цепей Крика с получением чтения секвенирования и определением присутствия или отсутствия мутации в цепи Крика области-мишени; и I) группировку чтений секвенирования по молекулярному штрих-коду, присутствующему в каждом чтении секвенирования, для определения, присутствует ли мутация в обеих цепях двухцепочечной ДНК-матрицы. В некоторых вариантах осуществления получение фрагментов двухцепочечной ДНК, каждый из которых имеет дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, включает: i) лигирование 3'-дуплексного адаптера с каждым 3'-концом фрагмента двухцепочечной ДНК, полученного из фрагмента двухцепочечной ДНК. ДНК-матрица, где 3'-дуплексный адаптер включает: а) первый олигонуклеотид, включающий 5'-фосфат, первый молекулярный штрих-код и 3'-олигонуклеотид, гибридный с б) вторым олигонуклеотидом, включающим расщепляемую 3'-блокирующую группу, где 3'-олигонуклеотид олигонуклеотид и вторая олигонуклеотидная последовательность являются комплементарными; ii) расщепление расщепляемой 3'-блокирующей группы; iii) лигирование 5'-адаптера с каждым дефосфорилированным 5'-концом фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, где 5'-дуплексный адаптер включает олигонуклеотид, включающий второй молекулярный штрих-код, где второй молекулярный штрих-код отличается от первого молекулярного штрих-кода, где 5'-адаптер лигирован с двухцепочечным фрагментом ДНК перед первым молекулярным штрих-кодом, при этом остается гэп с одноцепочечной нуклеиновой кислотой между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером; и iv) заполнение гэпа с одноцепочечной нуклеиновой кислотой между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером с образованием фрагмента двухцепочечной ДНК, включающего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК. В некоторых вариантах осуществления создание библиотеки ДНК (например, библиотеки

одноцепочечных ДНК) (например, библиотеки ДНК (например, библиотеки одноцепочечных ДНК)) цепей Уотсона из библиотеки амплифицированных дуплексов для секвенирования включает: i) амплификацию первой аликвоты библиотеки амплифицированных дуплексов для секвенирования при использовании пары праймеров, состоящей из первого праймера и второго праймера, где первый праймер способен к гибридизации с цепью Уотсона, и где первый праймер включает метку, с получением двухцепочечных продуктов амплификации, содержащих меченые цепи Уотсона; ii) денатурацию двухцепочечных продуктов амплификации, содержащих меченые цепи Уотсона, с получением одноцепочечных меченых цепей Уотсона и одноцепочечных цепей Крика; и iii) выделение одноцепочечных меченых цепей Уотсона с получением библиотеки ДНК (например, библиотеки одноцепочечных ДНК) цепей Уотсона из библиотеки амплифицированных дуплексов для секвенирования.

В некоторых вариантах осуществления двухцепочечную ДНК-матрицу получают из образца млекопитающего, создание библиотеки ДНК (например, библиотеки одноцепочечных ДНК) цепей Крика из библиотеки амплифицированных дуплексов для секвенирования включает: i) амплификацию второй аликвоты библиотеки амплифицированных дуплексов для секвенирования при использовании пары праймеров, состоящей из первого праймера и второго праймера, где первый праймер способен к гибридизации с цепью Крика, и где первый праймер включает метку, с получением двухцепочечных продуктов амплификации, содержащих меченые цепи Крика; ii) денатурацию двухцепочечных продуктов амплификации, содержащих меченые цепи Крика, с получением одноцепочечных меченых цепей Крика и одноцепочечных цепей Уотсона; и iii) выделение одноцепочечных меченых цепей Крика с созданием библиотеки ДНК (например, библиотеки одноцепочечных ДНК) цепей Крика из библиотеки амплифицированных дуплексов для секвенирования. В некоторых вариантах осуществления млекопитающим является человек.

В некоторых вариантах осуществления способ дополнительно включает, перед созданием фрагмента двухцепочечной ДНК, содержащего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК: фрагментацию двухцепочечной ДНК с получением фрагментов двухцепочечной ДНК; дефосфорилирование 5'-концов фрагментов двухцепочечной ДНК; и затупление концов фрагментов двухцепочечной ДНК. В некоторых вариантах осуществления лигирование 3'-дуплексного адаптера с каждым 3'-концом фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, включает контакт 3'-дуплексного адаптера и фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, в присутствии лигазы. В некоторых вариантах осуществления лигаза является ДНК-лигазой T4.

В некоторых вариантах осуществления расщепление расщепляемой 3'-блокирующей группы включает контакт 3'-дуплексного адаптера с урацил-ДНК-гликозилазой (УДГ). В некоторых вариантах осуществления лигирование 5'-адаптера с каждым дефосфорилированным 5'-концом фрагмента двухцепочечной ДНК, полученного

с двухцепочечной ДНК-матрицы, включает контакт 5'-адаптера и фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, в присутствии лигазы. В некоторых вариантах осуществления лигаза является лигазой *Escherichia coli*.

В некоторых вариантах осуществления заполнение гэпа с одноцепочечной нуклеиновой кислотой между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером включает контакт 5'-конца фрагмента двухцепочечной ДНК и 5'-адаптера в присутствии полимеразы и дНТФ. В некоторых вариантах осуществления полимеразы является Taq-полимеразой.

В некоторых вариантах осуществления лигирование 5'-адаптера с каждым 5'-концом фрагмента двухцепочечной ДНК и заполнение гэпа между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером производят одновременно. В некоторых вариантах осуществления амплификация фрагмента двухцепочечной ДНК, включающего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, с получением библиотеки амплифицированных дуплексов для секвенирования включает контакт фрагмента двухцепочечной ДНК, включающего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, с парой универсальных праймеров в условиях ПЦР. В некоторых вариантах осуществления амплификация включает полногеномную ПЦР. В некоторых вариантах осуществления меченые праймеры являются биотинилированными праймерами, и где биотинилированные праймеры могут давать биотинилированные одноцепочечные цепи Уотсона и биотинилированные одноцепочечные цепи Крика. В некоторых вариантах осуществления этапы денатурации включают денатурацию NaOH, термическую денатурацию или их комбинацию.

В некоторых вариантах осуществления этапы выделения включают контакт меченых цепей Уотсона со стрептавидин-функционализированными сферами и контакт меченых цепей Крика со стрептавидин-функционализированными сферами. В некоторых вариантах осуществления этапы выделения дополнительно включают денатурацию немеченных цепей Уотсона и денатурацию немеченных цепей Уотсона. В некоторых вариантах осуществления этапы выделения дополнительно включают высвобождение биотинилированных одноцепочечных цепей Уотсона со стрептавидин-функционализированных сфер и высвобождение биотинилированных одноцепочечных цепей Крика со стрептавидин-функционализированных сфер. В некоторых вариантах осуществления меченые праймеры представляют собой фосфорилированные праймеры, при этом фосфорилированные праймеры могут давать фосфорилированные одноцепочечные цепи Уотсона и фосфорилированные одноцепочечные цепи Крика. В некоторых вариантах осуществления этапы денатурации включают расщепление экзонуклеазой фага лямбда.

В некоторых вариантах осуществления амплификация области-мишени из библиотеки ДНК (например, библиотеки одноцепочечных ДНК) цепей Уотсона дополнительно включает вторую амплификацию с использованием второй пары

праймеров, состоящей из первого праймера, способного к гибридизации с областью-мишенью, и второго праймера, способного к гибридизации с 3'-дуплексным адаптером; и где амплификация области-мишени из библиотеки ДНК (например, библиотеки одноцепочечных ДНК) цепей Крика дополнительно включает вторую амплификацию с использованием второй пары праймеров, состоящей из первого праймера, способного к гибридизации с областью-мишенью, и второго праймера, способного к гибридизации с 5'-адаптером. В некоторых вариантах осуществления этапы секвенирования включают секвенирование спаренных концов.

В настоящем документе также предложены способы обнаружения присутствия или отсутствия мутации в области-мишени двухцепочечной ДНК-матрицы, полученной из образца млекопитающего, и определения, присутствует ли мутация в обеих цепях двухцепочечной ДНК-матрицы, где способ включает: А) получение фрагментов двухцепочечной ДНК, каждый из которых имеет дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК; В) создание библиотеки ДНК (например, библиотеки одноцепочечных ДНК) цепей Уотсона и библиотеки ДНК (например, библиотеки одноцепочечных ДНК) цепей Крика из библиотеки амплифицированных дуплексов для секвенирования из фрагмента двухцепочечной ДНК, имеющего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК; С) амплификацию области-мишени с одноцепочечных цепей Уотсона при использовании пары праймеров, состоящей из первого праймера, способного к гибридизации с областью-мишенью, и второго праймера, способного к гибридизации с 3'-дуплексным адаптером; D) амплификацию области-мишени с одноцепочечных цепей Крика при использовании пары праймеров, состоящей из первого праймера, способного к гибридизации с областью-мишенью, и второго праймера, способного к гибридизации с 5'-адаптером; Е) секвенирование области-мишени, амплифицированной с библиотеки ДНК (например, библиотеки одноцепочечных ДНК) цепей Уотсона, с получением чтения последовательности и обнаружением присутствия или отсутствия мутации в цепи Уотсона области-мишени; F) секвенирование области-мишени, амплифицированной с библиотеки ДНК (например, библиотеки одноцепочечных ДНК) цепей Крика, с получением чтения последовательности и определением присутствия или отсутствия мутации в цепи Крика области-мишени; и G) группировку чтений секвенирования по молекулярному штрих-коду, присутствующему в каждом чтении секвенирования, для определения, присутствует ли мутация в обеих цепях двухцепочечной ДНК-матрицы.

В некоторых вариантах осуществления двухцепочечная ДНК-матрица представляет собой образец геномной ДНК, и создание фрагментов двухцепочечной ДНК, каждый из которых имеет дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, включает: i) лигирование 3'-дуплексного адаптера с каждым 3'-концом фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, где 3'-дуплексный адаптер включает: а) первый олигонуклеотид, включающий 5'-фосфат, первый молекулярный штрих-код и 3'-олигонуклеотид, гибридизированный с б) вторым

олигонуклеотидом, включающим расщепляемую 3'-блокирующую группу, где 3'-олигонуклеотид и вторая олигонуклеотидная последовательность являются комплементарными; ii) расщепление расщепляемой 3'-блокирующей группы; iii) лигирование 5'-адаптера с каждым дефосфорилированным 5'-концом фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, где 5'-дуплексный адаптер включает олигонуклеотид, включающий второй молекулярный штрих-код, где второй молекулярный штрих-код отличается от первого молекулярного штрих-кода, где 5'-адаптер лигирован с двухцепочечным фрагментом ДНК перед первым молекулярным штрих-кодом и оставляет гэп с одноцепочечной нуклеиновой кислотой между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером; и iv) заполнение гэпа с одноцепочечной нуклеиновой кислотой между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером с получением фрагмента двухцепочечной ДНК, включающего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК.

В некоторых вариантах осуществления двухцепочечная ДНК-матрица представляет собой образец внеклеточной ДНК, и создание библиотеки ДНК (например, библиотеки одноцепочечных ДНК) цепей Уотсона и библиотеки ДНК (например, библиотеки одноцепочечных ДНК) цепей Крика из библиотеки амплифицированных дуплексов для секвенирования из фрагмента двухцепочечной ДНК, имеющего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, включает: i) амплификацию фрагмента двухцепочечной ДНК, имеющего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, при использовании пары универсальных праймеров, состоящей из первого праймера и второго праймера, где амплификация включает контакт фрагмента двухцепочечной ДНК, включающего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, с парой праймеров в условиях полногеномной ПЦР, где первый праймер способен к гибридизации с цепью Уотсона, и где первый праймер биотинилирован, с получением двухцепочечных продуктов амплификации, имеющих биотинилированные цепи Уотсона; ii) контакт двухцепочечных продуктов амплификации, имеющих биотинилированные цепи Уотсона, со стрептавидин-функционализованными сферами в условиях, когда биотинилированные цепи Уотсона связываются со стрептавидин-функционализованными сферами; iii) денатурацию двухцепочечных продуктов амплификации, имеющих биотинилированные цепи Уотсона, при этом одноцепочечные биотинилированные цепи Уотсона остаются связанными со стрептавидин-функционализованными сферами, а одноцепочечные цепи Крика высвобождаются; iv) сбор одноцепочечных цепей Крика; v) высвобождение одноцепочечных биотинилированных цепей Уотсона со стрептавидин-функционализованных сфер; и vi) сбор одноцепочечных биотинилированных цепей Уотсона.

В некоторых вариантах осуществления двухцепочечная ДНК-матрица получена из образца млекопитающего. В некоторых вариантах осуществления млекопитающим является человек.

В некоторых вариантах осуществления способ дополнительно включает, перед созданием фрагмента двухцепочечной ДНК, имеющего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК: фрагментацию двухцепочечной ДНК с получением фрагментов двухцепочечной ДНК; дефосфорилирование 5'-концов фрагментов двухцепочечной ДНК; и затупление концов фрагментов двухцепочечной ДНК.

В некоторых вариантах осуществления лигирование 3'-дуплексного адаптера с каждым 3'-концом фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, включает контакт 3'-дуплексного адаптера и фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, в присутствии лигазы. В некоторых вариантах осуществления лигаза является ДНК-лигазой T4. В некоторых вариантах осуществления расщепляемая 3'-блокирующая группа включает контакт 3'-дуплексного адаптера с урацил-ДНК-гликозилазой (УДГ). В некоторых вариантах осуществления лигирование 5'-адаптера с каждым дефосфорилированным 5'-концом фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, включает контакт 5'-адаптера и фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, в присутствии лигазы. В некоторых вариантах осуществления лигаза является лигазой *Escherichia coli*.

В некоторых вариантах осуществления заполнение гэпа с одноцепочечной нуклеиновой кислотой между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером включает контакт 5'-конца фрагмента двухцепочечной ДНК и 5'-адаптера в присутствии полимеразы и дНТФ. В некоторых вариантах осуществления полимеразы является полимеразой Taq-В. В некоторых вариантах осуществления лигирование 5'-адаптера с каждым 5'-концом фрагмента двухцепочечной ДНК и заполнение гэпа между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером производят одновременно.

В некоторых вариантах осуществления амплификация фрагмента двухцепочечной ДНК, имеющего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, включает контакт фрагмента двухцепочечной ДНК, содержащего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, с парой праймеров в условиях ПЦР. В некоторых вариантах осуществления амплификация включает полногеномную ПЦР. В некоторых вариантах осуществления амплификация области-мишени с библиотеки ДНК (например, библиотеки одноцепочечных ДНК) цепей Уотсона дополнительно включает вторую амплификацию при использовании второй пары праймеров, состоящей из первого праймера, способного к гибридизации с областью-мишенью, и второго праймера, способного к гибридизации с 3'-дуплексным адаптером; и при этом амплификация области-мишени с библиотеки ДНК (например, библиотеки одноцепочечных ДНК) цепей Крика дополнительно включает вторую амплификацию при использовании второй пары праймеров, состоящей из первого праймера, способного к гибридизации с областью-мишенью, и второго праймера, способного к гибридизации с 5'-адаптером. В некоторых вариантах осуществления этапы секвенирования включают

секвенирование спаренных концов или секвенирование неспаренных концов.

В настоящем документе также предложены способы обнаружения присутствия или отсутствия мутации в области-мишени двухцепочечной ДНК-матрицы, полученной из образца млекопитающего, и определения, присутствует ли мутация в обеих цепях двухцепочечной ДНК-матрицы, где способ включает: А) получение фрагментов двухцепочечной ДНК, каждый из которых имеет дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК; В) амплификацию фрагментов двухцепочечной ДНК, каждый из которых имеет дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, при использовании пары универсальных праймеров, где амплификация включает контакт фрагмента двухцепочечной ДНК, включающего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, с парой праймеров в условиях полногеномной ПЦР; С) амплификацию области-мишени с цепи Уотсона амплифицированных фрагментов двухцепочечной ДНК, каждый из которых имеет дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, при использовании пары праймеров, состоящей из первого праймера, способного к гибридизации с областью-мишенью, и второго праймера, способного к гибридизации с 3'-дуплексным адаптером; D) амплификацию области-мишени с цепи Крика амплифицированных фрагментов двухцепочечной ДНК, каждый из которых имеет дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, при использовании пары праймеров, состоящей из первого праймера, способного к гибридизации с областью-мишенью, и второго праймера, способного к гибридизации с 5'-адаптером; Е) секвенирование области-мишени, амплифицированной с цепи Уотсона, с получением чтения последовательности и обнаружением присутствия или отсутствия мутации в цепи Уотсона области-мишени; F) секвенирование области-мишени, амплифицированной с цепи Крика, с получением чтения последовательности и определением присутствия или отсутствия мутации в цепи Крика области-мишени; G) группировку чтений секвенирования по молекулярному штрих-коду, присутствующему в каждом чтении секвенирования, для определения, присутствует ли мутация в обеих цепях двухцепочечной ДНК-матрицы.

В некоторых вариантах осуществления двухцепочечная ДНК-матрица представляет собой образец геномной ДНК, и создание фрагментов двухцепочечной ДНК, каждый из которых имеет дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, включает: i) лигирование 3'-дуплексного адаптера с каждым 3'-концом фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, где 3'-дуплексный адаптер включает: а) первый олигонуклеотид, включающий 5'-фосфат, первый молекулярный штрих-код и 3'-олигонуклеотид, гибридизированный с б) вторым олигонуклеотидом, включающий расщепляемую 3'-блокирующую группу, где 3'-олигонуклеотид и вторая олигонуклеотидная последовательность являются комплементарными; ii) расщепление расщепляемой 3'-блокирующей группы; iii) лигирование 5'-адаптера с каждым дефосфорилированным 5'-концом фрагмента

двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, где 5'-дуплексный адаптер включает олигонуклеотид, включающий второй молекулярный штрих-код, где второй молекулярный штрих-код отличается от первого молекулярного штрих-кода, где 5'-адаптер лигирован с двухцепочечным фрагментом ДНК перед первым молекулярным штрих-кодом, при этом остается гэп с одноцепочечной нуклеиновой кислотой между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером; и iv) заполнение гэпа с одноцепочечной нуклеиновой кислотой между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером с получением фрагмента двухцепочечной ДНК, включающего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК. В некоторых вариантах осуществления двухцепочечная ДНК-матрица представляет собой образец внеклеточной ДНК. В некоторых вариантах осуществления двухцепочечная ДНК-матрица представляет собой образец геномной ДНК. В некоторых вариантах осуществления млекопитающим является человек.

В некоторых вариантах осуществления способ дополнительно включает, перед созданием фрагмента двухцепочечной ДНК, имеющего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК: фрагментацию двухцепочечной ДНК с получением фрагментов двухцепочечной ДНК; дефосфорилирование 5'-концов фрагментов двухцепочечной ДНК; и затупление концов фрагментов двухцепочечной ДНК.

В некоторых вариантах осуществления лигирование 3'-дуплексного адаптера с каждым 3'-концом фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, включает контакт 3'-дуплексного адаптера и фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, в присутствии лигазы. В некоторых вариантах осуществления лигаза является ДНК-лигазой Т4. В некоторых вариантах осуществления расщепление расщепляемой 3'-блокирующей группы включает контакт 3'-дуплексного адаптера с урацил-ДНК-гликозилазой (УДГ). В некоторых вариантах осуществления лигирование 5'-адаптера с каждым дефосфорилированным 5'-концом фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, включает контакт 5'-адаптера и фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, в присутствии лигазы. В некоторых вариантах осуществления лигаза является лигазой *Escherichia coli*.

В некоторых вариантах осуществления заполнения гэпа с одноцепочечной нуклеиновой кислотой между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером включает контакт 5'-конца фрагмента двухцепочечной ДНК и 5'-адаптера в присутствии ДНК-полимеразы и дНТФ. В некоторых вариантах осуществления ДНК-полимераза является полимеразой Taq-В.

В некоторых вариантах осуществления лигирование 5'-адаптера с каждым 5'-концом фрагмента двухцепочечной ДНК и заполнение гэпа между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером производят одновременно. В некоторых вариантах осуществления амплификация фрагмента двухцепочечной ДНК, имеющего дуплексный

молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, включает контакт фрагмента двухцепочечной ДНК, содержащего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, с парой праймеров в условиях ПЦР. В некоторых вариантах осуществления амплификация включает полногеномную ПЦР. В некоторых вариантах осуществления амплификация области-мишени с библиотеки ДНК (например, библиотеки одноцепочечных ДНК) цепей Уотсона дополнительно включает вторую амплификацию при использовании второй пары праймеров, состоящей из первого праймера, способного к гибридизации с областью-мишенью, и второго праймера, способного к гибридизации с 3'-дуплексным адаптером; и где амплификация области-мишени с библиотеки ДНК (например, библиотеки одноцепочечных ДНК) цепей Крика дополнительно включает вторую амплификацию при использовании второй пары праймеров, состоящей из первого праймера, способного к гибридизации с областью-мишенью, и второго праймера, способного к гибридизации с 5'-адаптером. В некоторых вариантах осуществления этапы секвенирования включают секвенирование спаренных концов.

Также в настоящем документе предложены способы, включающие:

- a. присоединение частично двухцепочечных 3'-адаптеров к 3'-концам цепей Уотсона и Крика популяции фрагментов двухцепочечной ДНК в анализируемом образце ДНК, где первая цепь частично двухцепочечного 3'-адаптера включает, в направлении 5'→3': (i) первый сегмент, (ii) экзогенную UID последовательность, (iii) сайт отжига для 5'-адаптера и (iv) универсальную 3'-адаптерную последовательность, включающая сайт праймера для секвенирования R2, и где вторая цепь частично двухцепочечного 3'-адаптера включает, в направлении 5'→3': (i) сегмент, комплементарный первому сегменту, и (ii) 3'-блокирующую группу, где вторая цепь необязательно является расщепляемой;
- b. отжиг 5'-адаптеров с 3'-адаптерами посредством сайта отжига, где 5'-адаптеры включают, в направлении 5'→3': (i) универсальную 5'-адаптерную последовательность, которая не комплементарна универсальной 3'-адаптерной последовательности, и которая включает сайт праймера для секвенирования R1, и (ii) последовательность, комплементарную сайту отжига для 5'-адаптера;
- c. выполнение реакции, подобной ник-трансляции, с удлинением 5'-адаптеров через экзогенную UID последовательность 3'-адаптеров (например, с использованием ДНК-полимеразы) и ковалентным связыванием удлиненного 5'-адаптера с 5'-концами цепей Уотсона и Крика фрагментов двухцепочечной ДНК (например, с использованием лигазы);
- d. проведение первичной амплификации с целью амплификации лигированных с адаптером фрагментов двухцепочечной ДНК с получением ампликонов;
- e. определение чтений последовательности одного или более ампликонов одного или более лигированных с адаптером фрагментов двухцепочечной ДНК;
- f. отнесение чтений последовательностей к семействам UID, где каждый член семейства UID включает одну и ту же экзогенную UID последовательность;
- g. отнесение чтений последовательностей каждого семейства UID к подсемейству

Уотсона и подсемейству Крика на основе пространственного отношения экзогенной последовательности UID к последовательностям чтений R1 и R2;

h. идентификацию нуклеотидной последовательности в качестве точно представленной цепи Уотсона анализируемого фрагмента ДНК, когда пороговый процент членов подсемейства Уотсона содержит такую последовательность;

i. идентификацию нуклеотидной последовательности в качестве точно представленной цепи Крика анализируемого фрагмента ДНК, когда пороговый процент членов подсемейства Крика содержит такую последовательность;

j. идентификацию мутации в нуклеотидной последовательности, точно представляющей цепь Уотсона, когда последовательность, точно представляющая цепь Уотсона, отличается от референсной последовательности, в которой отсутствует такая мутация;

k. идентификацию мутации в нуклеотидной последовательности, точно представляющей цепь Крика, когда последовательность, точно представляющая цепь Крика, отличается от референсной последовательности, в которой отсутствует такая мутация; и

l. идентификацию мутации в анализируемом фрагменте ДНК, когда мутация в нуклеотидной последовательности, точно представляющей цепь Уотсона, и мутация в нуклеотидной последовательности, точно представляющей цепь Крика, являются одной и той же мутацией.

В некоторых вариантах осуществления каждый член семейства UID дополнительно включает одну и ту же эндогенную UID последовательность, где эндогенная UID последовательность включает конец фрагмента двухцепочечной ДНК из популяции. В некоторых вариантах осуществления эндогенная UID последовательность, включающая конец фрагмента двухцепочечной ДНК, включает по меньшей мере 8, 10 или 15 оснований. В некоторых вариантах осуществления экзогенная UID последовательность является уникальной для каждого фрагмента двухцепочечной ДНК. В некоторых вариантах осуществления экзогенная UID последовательность не является уникальной для каждого фрагмента двухцепочечной ДНК. В некоторых вариантах осуществления каждый член семейства UID включает одну и ту же эндогенную UID последовательность и одну и ту же экзогенную UID последовательность. В некоторых вариантах осуществления стадия (d) включает не больше 11 циклов ПЦР-амплификации. В некоторых вариантах осуществления стадия (d) включает не больше 7 циклов ПЦР-амплификации. В некоторых вариантах осуществления стадия (d) включает не больше 5 циклов ПЦР-амплификации. В некоторых вариантах осуществления стадия (d) включает по меньшей мере 1 цикл ПЦР-амплификации.

В некоторых вариантах осуществления перед определением чтений последовательности ампликоны обогащают одним или больше целевыми полинуклеотидами. В некоторых вариантах осуществления обогащение включает:

a. селективную амплификацию ампликонов цепей Уотсона, включающих

полинуклеотидную последовательность-мишень, с использованием первого набора пар мишень-селективных праймеров Уотсона, причем первый набор пар мишень-селективных праймеров Уотсона включает: (i) первый мишень-селективный праймер Уотсона, включающий последовательность, комплементарную части универсальной 3'-адаптерной последовательности, где, необязательно, часть универсальной 3'-адаптерной последовательности представляет собой сайт праймера для секвенирования R2 универсальной 3'-адаптерной последовательности, и (ii) второй мишень-селективный праймер Уотсона, включающий мишень-селективную последовательность, с получением в результате целевых продуктов амплификации Уотсона; и

в. селективную амплификацию ампликонов цепей Крика, включающих такую же полинуклеотидную последовательность-мишень, с использованием первого набора пар мишень-селективных праймеров Крика, причем первый набор пар мишень-селективных праймеров Крика включает: (i) первый мишень-селективный праймер Крика, включающий последовательность, комплементарную части универсальной 5'-адаптерной последовательности, где, необязательно, часть универсальной 5'-адаптерной последовательности представляет собой сайт праймера для секвенирования R1 универсальной 5'-адаптерной последовательности, и (ii) второй мишень-селективный праймер Крика, включающий такую же мишень-селективную последовательность в качестве последовательности второго мишень-селективного праймера Уотсона, с получением в результате целевых продуктов амплификации Крика.

В некоторых вариантах осуществления способ дополнительно включает очистку целевых продуктов амплификации Уотсона и целевых продуктов амплификации Крика от нецелевых полинуклеотидов. В некоторых вариантах осуществления способ дополнительно включает очистку, которая включает прикрепление целевых продуктов амплификации Уотсона и целевых продуктов амплификации Крика к твердой подложке. В некоторых вариантах осуществления первый мишень-селективный праймер Уотсона и первый мишень-селективный праймер Крика включают первый член пары аффинного связывания, и где твердая подложка включает второй член пары аффинного связывания. В некоторых вариантах осуществления первый член представляет собой биотин, а второй член представляет собой стрептавидин. В некоторых вариантах осуществления твердая подложка включает сферу, лунку, мембрану, пробирку, колонку, планшет, сефарозу, магнитную сферу или чип. В некоторых вариантах осуществления способ дополнительно включает удаление полинуклеотидов, которые не прикрепилась к твердой подложке.

В некоторых вариантах осуществления способ дополнительно включает:

а. дополнительную амплификацию целевых продуктов амплификации Уотсона с использованием второго набора мишень-селективных праймеров Уотсона, причем второй набор мишень-селективных праймеров Уотсона включает: (i) третий мишень-селективный праймер Уотсона, включающий последовательность, комплементарную части универсальной 3'-адаптерной последовательности, где, необязательно, часть универсальной 3'-адаптерной последовательности представляет собой сайт праймера для

секвенирования R2 универсальной 3'-адаптерной последовательности, и (ii) четвертый мишень-селективный праймер Уотсона, включающий, в направлении 5'→3', сайт праймера для секвенирования R1 и мишень-селективную последовательность, селективную в отношении того же полинуклеотида-мишени, с получением в результате целевых членов библиотеки Уотсона;

в. дополнительную амплификацию целевых продуктов Крика с использованием второго набора мишень-селективных праймеров Крика, причем второй набор мишень-селективных праймеров Крика включает: (i) третий мишень-селективный праймер Крика, включающий последовательность, комплементарную части универсальной 5'-адаптерной последовательности, где, необязательно, часть универсальной 5'-адаптерной последовательности представляет собой сайт праймера для секвенирования R1 универсальной 5'-адаптерной последовательности, и (ii) четвертый мишень-селективный праймер Крика, включающий, в направлении 5'→3', сайт праймера для секвенирования R2 и мишень-селективную последовательность, селективную в отношении того же полинуклеотида-мишени четвертого мишень-селективного праймера Уотсона, с получением в результате целевых членов библиотеки Крика.

В некоторых вариантах осуществления третьи мишень-селективные праймеры Уотсона и Крика дополнительно включают последовательность штрих-кода образца. В некоторых вариантах осуществления третий мишень-селективный праймер Уотсона дополнительно включает первую последовательность для связывания, которая обеспечивает гибридизацию с первым праймером для связывания на секвенаторе, и где третий мишень-селективный праймер Крика дополнительно включает вторую последовательность для связывания, которая обеспечивает гибридизацию со вторым праймером для связывания на секвенаторе. В некоторых вариантах осуществления четвертый мишень-селективный праймер Уотсона дополнительно включает вторую последовательность для связывания, и где четвертый мишень-селективный праймер Крика дополнительно включает первую последовательность для связывания. В некоторых вариантах осуществления первая последовательность для связывания представляет собой последовательность P7, а вторая последовательность для связывания представляет собой последовательность P5. В некоторых вариантах осуществления целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляют по меньшей мере 50% целевых полинуклеотидов в популяции фрагментов двухцепочечной ДНК. В некоторых вариантах осуществления целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляют по меньшей мере 70% целевых полинуклеотидов в популяции фрагментов двухцепочечной ДНК. В некоторых вариантах осуществления целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляют по меньшей мере 80% целевых полинуклеотидов в популяции фрагментов двухцепочечной ДНК. В некоторых вариантах осуществления целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляют по меньшей мере 90% целевых полинуклеотидов в популяции фрагментов двухцепочечной ДНК. В некоторых вариантах

осуществления целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляют по меньшей мере 50% популяции всех фрагментов ДНК. В некоторых вариантах осуществления целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляют по меньшей мере 70% популяции всех фрагментов ДНК. В некоторых вариантах осуществления целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляют по меньшей мере 80% популяции всех фрагментов ДНК. В некоторых вариантах осуществления целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляют по меньшей мере 90% популяции всех фрагментов ДНК.

Также в настоящем документе предложены способы, включающие:

а. присоединение адаптеров к популяции фрагментов двухцепочечной ДНК в анализируемом образце ДНК, где адаптеры включают двухцепочечную часть, включающую экзогенный UID, и разветвленную часть, включающую: (i) одноцепочечную 3'-адаптерную последовательность, включающую сайт праймера для секвенирования R2 и (ii) одноцепочечную 5'-адаптерную последовательность, включающую сайт праймера для секвенирования R1;

б. проведение первичной амплификации для амплификации лигированных с адаптером фрагментов двухцепочечной ДНК с получением ампликонов;

с. селективную амплификацию ампликонов цепей Уотсона, включающих полинуклеотидную последовательность-мишень, с использованием первого набора пар мишень-селективных праймеров Уотсона, причем первый набор пар мишень-селективных праймеров Уотсона включает: (i) первый мишень-селективный праймер Уотсона, включающий последовательность, комплементарную части универсальной 3'-адаптерной последовательности, где, необязательно, часть универсальной 3'-адаптерной последовательности представляет собой сайт праймера для секвенирования R2 универсальной 3'-адаптерной последовательности, и (ii) второй мишень-селективный праймер Уотсона, включающий мишень-селективную последовательность, с получением в результате целевых продуктов амплификации Уотсона;

д. селективную амплификацию ампликонов цепей Крика, включающих такую же полинуклеотидную последовательность-мишень, с использованием первого набора пар мишень-селективных праймеров Крика, причем первый набор пар мишень-селективных праймеров Крика включает: первый мишень-селективный праймер Крика, включающий последовательность, комплементарную части универсальной 5'-адаптерной последовательности, где, необязательно, часть универсальной 5'-адаптерной последовательности представляет собой сайт праймера для секвенирования R1 универсальной 5'-адаптерной последовательности, и (ii) второй мишень-селективный праймер Крика, включающий такую же мишень-селективную последовательность в качестве последовательности второго мишень-селективного праймера Крика, с получением в результате целевых продуктов амплификации Крика.

е. определение чтений последовательности целевых продуктов амплификации

Уотсона и целевых продуктов амплификации Крика;

f. отнесение чтений последовательностей к семействам UID, где каждый член семейства UID включает одну и ту же экзогенную UID последовательность;

g. отнесение чтений последовательностей каждого семейства UID к подсемейству Уотсона и подсемейству Крика на основе пространственного отношения экзогенной UID последовательности к последовательности чтений R1 и R2;

h. идентификацию нуклеотидной последовательности в качестве точно представленной цепи Уотсона анализируемого фрагмента ДНК, если пороговый процент членов семейства Уотсона содержит данную последовательность;

i. идентификацию нуклеотидной последовательности в качестве точно представленной цепи Крика анализируемого фрагмента ДНК, если пороговый процент членов семейства Крика содержит данную последовательность; и

j. идентификацию мутации в анализируемом фрагменте ДНК, когда нуклеотидная последовательность, точно представляющая цепь Уотсона, и нуклеотидная последовательность, точно представляющая цепь Крика, содержат одну и ту же мутацию.

В некоторых вариантах осуществления способ дополнительно включает очистку целевых продуктов амплификации Уотсона и целевых продуктов амплификации Крика от нецелевых полинуклеотидов. В некоторых вариантах осуществления способ дополнительно включает прикрепление целевых продуктов амплификации Уотсона и целевых продуктов амплификации Крика к твердой подложке. В некоторых вариантах осуществления первый селективный праймер для амплификации цепи Уотсона и первый селективный праймер для амплификации цепи Крика включают первый член пары аффинного связывания, и где твердая подложка включает второй член пары аффинного связывания. В некоторых вариантах осуществления первый член представляет собой биотин, а второй член представляет собой стрептавидин. В некоторых вариантах осуществления твердая подложка включает сферу, лунку, мембрану, пробирку, колонку, планшет, сефарозу, магнитную сферу или чип. В некоторых вариантах осуществления способ дополнительно включает удаление полинуклеотидов, не прикрепившихся к твердой подложке.

В некоторых вариантах осуществления способ дополнительно включает:

a. дополнительную амплификацию целевых продуктов амплификации Уотсона с использованием второго набора мишень-селективных праймеров Уотсона, причем второй набор мишень-селективных праймеров Уотсона включает: (i) третий мишень-селективный праймер Уотсона, включающий последовательность, комплементарную сайту праймера для секвенирования R2 универсальной 3'-адаптерной последовательности, и (ii) четвертый мишень-селективный праймер Уотсона, включающий, в направлении 5'→3', сайт праймера для секвенирования R1 и мишень-селективную последовательность, селективную в отношении такого же полинуклеотида-мишени, с получением в результате целевых членов библиотеки Уотсона;

b. дополнительную амплификацию целевых продуктов Крика с использованием

второго набора мишень-селективных праймеров Крика, причем второй набор мишень-селективных праймеров Крика включает: (i) третий мишень-селективный праймер для амплификации цепи Крика, включающий последовательность, комплементарную сайту праймера для секвенирования R1 универсальной 3'-адаптерной последовательности, и (ii) четвертый мишень-селективный праймер Крика, включающий, в направлении 5'→3', сайт праймера для секвенирования R2 и мишень-селективную последовательность, селективную в отношении такого же полинуклеотида-мишени четвертого мишень-селективного праймера Уотсона, с получением в результате целевых членов библиотеки Крика.

В некоторых вариантах осуществления третьи мишень-селективные праймеры Уотсона и Крика дополнительно включают последовательность штрих-кода образца. В некоторых вариантах осуществления третий мишень-селективный праймер Уотсона дополнительно включает первую последовательность для связывания, которая обеспечивает гибридизацию с первым праймером для связывания на секвенаторе, и где третий мишень-селективный праймер Крика дополнительно включает вторую последовательность для связывания, которая обеспечивает гибридизацию со вторым праймером для связывания на секвенаторе. В некоторых вариантах осуществления четвертый мишень-селективный праймер Уотсона дополнительно включает вторую последовательность для связывания, а четвертый мишень-селективный праймер Крика дополнительно включает первую последовательность для связывания. В некоторых вариантах осуществления первая последовательность для связывания представляет собой последовательность P7, а вторая последовательность для связывания представляет собой последовательность P5. В некоторых вариантах осуществления прикрепление включает прикрепление адаптеров с А-хвостом к популяции фрагментов двухцепочечной ДНК. В некоторых вариантах осуществления прикрепление включает прикрепление адаптеров с А-хвостом к обоим концам фрагментов ДНК в популяции.

В некоторых вариантах осуществления прикрепление включает:

а. прикрепление частично двухцепочечных 3'-адаптеров к 3'-концам цепей Уотсона и Крика популяции фрагментов двухцепочечной ДНК, где первая цепь частично двухцепочечного 3'-адаптера включает, в направлении 5'→3': (i) первый сегмент, (ii) необязательно, экзогенную UID последовательность, (iii) сайт отжига для 5'-адаптера и (iv) универсальную 3'-адаптерную последовательность, включающую сайт праймера для секвенирования R2, и где вторая цепь частично двухцепочечного 3'-адаптера включает, в направлении 5'→3': (i) сегмент, комплементарный первому сегменту, и (ii) 3'-блокирующую группу, где, необязательно, вторая цепь является расщепляемой; и

б. отжиг 5'-адаптеров с 3'-адаптерами посредством сайта отжига, где 5'-адаптеры включают, в направлении 5'→3': (i) универсальную 5'-адаптерную последовательность, которая не комплементарна универсальной 3'-адаптерной последовательности и которая включает сайт праймера для секвенирования R1, и (ii) последовательность, комплементарную сайту отжига для 5'-адаптера; и

с. проведение реакции, подобной ник-трансляции, для удлинения 5'-адаптеров через 3'-адаптеры (например, с использованием ДНК-полимеразы) и ковалентного связывания удлиненного 5'-адаптера с 5'-концами цепей Уотсона и Крика фрагментов двухцепочечной ДНК (например, с использованием лигазы).

В некоторых вариантах осуществления UID последовательность содержит эндогенную UID последовательность, включающую конец фрагмента двухцепочечной ДНК из популяции. В некоторых вариантах осуществления эндогенная UID последовательность, включающая конец фрагмента двухцепочечной ДНК, включает по меньшей мере 8, 10 или 15 оснований. В некоторых вариантах осуществления экзогенная UID последовательность является уникальной для каждого фрагмента двухцепочечной ДНК. В некоторых вариантах осуществления экзогенная UID последовательность не является уникальной для каждого фрагмента двухцепочечной ДНК. В некоторых вариантах осуществления каждый член семейства UID включает одну и ту же эндогенную UID последовательность и одну и ту же экзогенную UID последовательность.

В некоторых вариантах осуществления амплификация лигированных с адаптером фрагментов двухцепочечной ДНК с получением ампликонов включает не больше 11 циклов ПЦР-амплификации. В некоторых вариантах осуществления амплификация лигированных с адаптером фрагментов двухцепочечной ДНК с получением ампликонов включает не больше 7 циклов ПЦР-амплификации. В некоторых вариантах осуществления амплификация лигированных с адаптером фрагментов двухцепочечной ДНК с получением ампликонов включает не больше 5 циклов ПЦР-амплификации. В некоторых вариантах осуществления амплификация лигированных с адаптером фрагментов двухцепочечной ДНК с получением ампликонов включает по меньшей мере 1 цикл ПЦР-амплификации. В некоторых вариантах осуществления целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляют по меньшей мере 50% целевых полинуклеотидов в популяции фрагментов двухцепочечной ДНК. В некоторых вариантах осуществления целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляют по меньшей мере 70% целевых полинуклеотидов в популяции фрагментов двухцепочечной ДНК. В некоторых вариантах осуществления целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляют по меньшей мере 80% целевых полинуклеотидов в популяции фрагментов двухцепочечной ДНК. В некоторых вариантах осуществления целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляют по меньшей мере 90% целевых полинуклеотидов в популяции фрагментов двухцепочечной ДНК. В некоторых вариантах осуществления целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляют по меньшей мере 50% популяции всех фрагментов ДНК. В некоторых вариантах осуществления целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляют по меньшей мере 70% популяции всех фрагментов ДНК. В некоторых вариантах осуществления целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляют по меньшей мере 80% популяции всех фрагментов

ДНК. В некоторых вариантах осуществления целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляют по меньшей мере 90% популяции всех фрагментов ДНК.

В некоторых вариантах осуществления определение чтений последовательности позволяет определить последовательность обоих концов молекул-матриц. В некоторых вариантах осуществления определение обоих концов молекул-матриц включает секвенирование спаренных концов. В некоторых вариантах осуществления определение чтений последовательности включает однократное секвенирование по всей длине матриц с получением чтений последовательности. В некоторых вариантах осуществления определение чтений последовательностей включает секвенирование с помощью массивно-параллельного секвенатора. В некоторых вариантах осуществления массивно-параллельный секвенатор обеспечивает возможность определения чтений последовательности с обоих концов матричных полинуклеотидов. В некоторых вариантах осуществления популяция фрагментов двухцепочечной ДНК включает один или больше фрагментов длиной приблизительно 50-600 нуклеотидов. В некоторых вариантах осуществления популяция фрагментов двухцепочечной ДНК включает один или больше фрагментов длиной меньше 2000, меньше 1000, меньше 500, меньше 400, меньше 300 или меньше 250 нуклеотидов.

В некоторых вариантах осуществления способы, предложенные в настоящем документе, дополнительно включают, после первичной амплификации и перед селективной амплификацией, получение библиотек одноцепочечных (оц) ДНК, соответствующих смысловым и антисмысловым цепям ампликонов. В некоторых вариантах осуществления получение библиотеки одноцепочечных ДНК включает:

a. проведение реакции амплификации с использованием двух праймеров, где только один из двух праймеров включает первый член пары аффинного связывания, с получением в результате продуктов амплификации, включающих цепь, включающую первый член пары аффинного связывания, и цепь, не включающую первый член пары аффинного связывания;

b. контакт продуктов амплификации с твердой подложкой, где твердая подложка включает второй член пары аффинного связывания;

c. денатурацию продуктов амплификации с отделением цепи, включающей первый член пары аффинного связывания, от цепи, не включающей первый член пары аффинного связывания; и

d. очистку разделенных цепей, включающих первый член пары аффинного связывания, и разделенных цепей, не включающих первый член пары аффинного связывания.

В некоторых вариантах осуществления первый член пары аффинного связывания представляет собой биотин, а второй член пары аффинного связывания представляет собой стрептавидин. В некоторых вариантах осуществления получение библиотеки одноцепочечных ДНК включает:

а. разделение ампликонов на две реакции амплификации, при этом в каждой реакции амплификации используется прямой праймер и обратный праймер, где фосфорилирован только один из двух праймеров, в результате чего образуются продукты амплификации, включающие фосфорилированную цепь и нефосфорилированную цепь;

б. контакт продуктов амплификации с экзонуклеазой, которая селективно расщепляет цепи с 5'-фосфатом.

В некоторых вариантах осуществления:

а. в первой реакции амплификации прямой праймер фосфорилирован, а обратный праймер не фосфорилирован;

б. во второй реакции амплификации обратный праймер фосфорилирован, а прямой праймер не фосфорилирован.

В некоторых вариантах осуществления экзонуклеаза представляет собой экзонуклеазу фага лямбда. В некоторых вариантах осуществления фосфорилирование идет по 5'-сайту.

В некоторых вариантах осуществления первичная амплификация включает:

а. амплификацию с использованием пары праймеров, в которой только один из двух праймеров включает первый член пары аффинного связывания, с получением в результате продуктов амплификации, включающих цепь, включающую первый член пары аффинного связывания, и цепь, не включающую первый член пары аффинного связывания;

б. контакт продуктов амплификации с твердой подложкой, где твердая подложка включает второй член пары аффинного связывания;

с. денатурацию продуктов амплификации с отделением цепи, включающей первый член пары аффинного связывания, от цепи, не включающей первый член пары аффинного связывания; и

д. очистку разделенных цепей, включающих первый член пары аффинного связывания, и разделенных цепей, не включающих первый член пары аффинного связывания.

В некоторых вариантах осуществления первый член пары аффинного связывания представляет собой биотин, а второй член пары аффинного связывания представляет собой стрептавидин. В некоторых вариантах осуществления чтения последовательностей семейства UID относят к подсемейству Уотсона, если экзогенная UID последовательность расположена после последовательности R2 и перед последовательностью R1. В некоторых вариантах осуществления чтения последовательностей семейства UID относят к подсемейству Крика, если экзогенная UID последовательность расположена после последовательности R1 и перед последовательностью R2. В некоторых вариантах осуществления чтения последовательностей семейства UID относят к подсемейству Уотсона, если экзогенная UID последовательность расположена ближе к последовательности R2 и дальше от последовательности R1. В некоторых вариантах осуществления чтения последовательностей семейства UID относят к подсемейству

Крика, если экзогенная UID последовательность расположена ближе к последовательности R1 и дальше от последовательности R2. В некоторых вариантах осуществления чтения последовательности семейства UID относят к подсемейству Уотсона, если экзогенная UID последовательность находится непосредственно после или в пределах 1-300, 1-70, 1-60, 1-50, 1-40, 1-30, 1-20, 1-10 или 1-5 нуклеотидов от последовательности R2. В некоторых вариантах осуществления чтения последовательности семейства UID относят к подсемейству Крика, если экзогенная UID последовательность находится непосредственно после или в пределах 1-300, 1-70, 1-60, 1-50, 1-40, 1-30, 1-20, 1-10 или 1-5 нуклеотидов от последовательности R1.

В некоторых вариантах осуществления популяция фрагментов двухцепочечной ДНК происходит из биологического образца. В некоторых вариантах осуществления биологический образец получен от субъекта. В некоторых вариантах осуществления субъектом является человек. В некоторых вариантах осуществления биологический образец представляет собой образец жидкости. В некоторых вариантах осуществления образец жидкости выбран из цельной крови, плазмы крови, сыворотки, мокроты, мочи, пота, слез, асцитной жидкости, спермы и бронхоальвеолярного лаважа. В некоторых вариантах осуществления образец жидкости представляет собой бесклеточный или по существу бесклеточный образец. В некоторых вариантах осуществления биологический образец представляет собой твердый биологический образец. В некоторых вариантах осуществления твердый биологический образец представляет собой образец опухоли.

В некоторых вариантах осуществления идентифицированная мутация присутствует в популяции фрагментов двухцепочечной ДНК с частотой 0,1% или меньше. В некоторых вариантах осуществления идентифицированная мутация присутствует в популяции фрагментов двухцепочечной ДНК с частотой от 0,1% до 0,00001%. В некоторых вариантах осуществления идентифицированная мутация присутствует в популяции фрагментов двухцепочечной ДНК с частотой от 0,1% до 0,01%. В некоторых вариантах осуществления определение чтений последовательностей включает определение чтений последовательностей с цепей Уотсона и Крика по меньшей мере 50% фрагментов двухцепочечной ДНК, включающих полинуклеотид-мишень в анализируемом образце ДНК. В некоторых вариантах осуществления определение чтений последовательностей включает определение чтений последовательностей как с цепей Уотсона, так и с цепей Крика по меньшей мере 70% фрагментов двухцепочечной ДНК, включающих полинуклеотид-мишень, в анализируемом образце ДНК. В некоторых вариантах осуществления определение чтений последовательностей включает определение чтений последовательностей с цепей Уотсона и Крика по меньшей мере 80% фрагментов двухцепочечной ДНК, включающих полинуклеотид-мишень в анализируемом образце ДНК. В некоторых вариантах осуществления определение чтений последовательностей включает определение чтений последовательностей как с цепей Уотсона, так и с цепей Крика по меньшей мере 90% фрагментов двухцепочечной ДНК, включающих полинуклеотид-мишень, в анализируемом образце ДНК. В некоторых вариантах

осуществления определение чтений последовательностей включает определение чтений последовательностей как с цепей Уотсона, так и с цепей Крика по меньшей мере 50% фрагментов двухцепочечной ДНК в анализируемом образце ДНК. В некоторых вариантах осуществления определение чтений последовательностей включает определение чтений последовательностей как с цепей Уотсона, так и с цепей Крика по меньшей мере 70% фрагментов двухцепочечной ДНК в анализируемом образце ДНК. В некоторых вариантах осуществления определение чтений последовательности включает определение чтений последовательностей как с цепей Уотсона, так и с цепей Крика по меньшей мере 80% фрагментов двухцепочечной ДНК в анализируемом образце ДНК. В некоторых вариантах осуществления определение чтений последовательностей включает определение чтений последовательностей как с цепей Уотсона, так и с цепей Крика по меньшей мере 90% фрагментов двухцепочечной ДНК в анализируемом образце ДНК.

В некоторых вариантах осуществления частота ошибок, связанных с идентификацией одной или более мутаций в анализируемых фрагментах ДНК в соответствии со способом согласно любому из предыдущих пунктов, снижена по меньшей мере в 2 раза, 4 раза, 5 раз, 10 раз, 20 раз, 30 раз, 40 раз, 50 раз, 60 раз, 70 раз, 80 раз, 90 раз или 100 раз по сравнению с альтернативным способом идентификации мутаций, который не требует одновременного обнаружения мутации в цепях Уотсона и в цепях Крика анализируемого фрагмента ДНК. В некоторых вариантах осуществления альтернативный способ включает стандартное молекулярное баркодирование или стандартное молекулярное баркодирование на основе метода ПЦР. В некоторых вариантах осуществления альтернативный способ включает:

- a. присоединение адаптеров к популяции фрагментов двухцепочечной ДНК в анализируемом образце ДНК, где адаптеры включают уникальный экзогенный UID;
- b. проведение первичной амплификации для амплификации лигированных с адаптером фрагментов двухцепочечной ДНК с получением ампликонов;
- c. определение чтений последовательностей одного или больше ампликонов одного или больше лигированных с адаптером фрагментов двухцепочечной ДНК;
- d. отнесение чтений последовательностей к семействам UID, где каждый член семейства UID включает одну и ту же экзогенную UID последовательность;
- e. идентификацию нуклеотидной последовательности в виде точно представленного анализируемого фрагмента ДНК, если пороговый процент членов семейства UID содержит такую последовательность; и
- f. идентификацию мутации во анализируемом фрагменте ДНК, если последовательность, идентифицированная как точно представляющая анализируемый фрагмент ДНК, отличается от референсной последовательности, в которой такая мутация отсутствует.

В некоторых вариантах осуществления частота ошибок, связанная с идентификацией одной или больше мутаций в анализируемых фрагментах ДНК в соответствии со способом согласно любому из предыдущих пунктов, составляет не

больше 1×10^{-2} , не больше 1×10^{-3} , не больше 1×10^{-4} , не больше 1×10^{-5} , не больше 1×10^{-6} , не больше 5×10^{-6} или не больше 1×10^{-7} .

В настоящем документе также предложены машиночитаемые носители, содержащие машиноисполняемые инструкции для анализа данных чтений последовательностей из образца нуклеиновой кислоты, где данные генерируются способом согласно любому из предыдущих пунктов. В некоторых вариантах осуществления машиночитаемый носитель содержит исполняемые инструкции для:

a. отнесения чтений последовательностей к семействам UID, где каждый член семейства UID включает одну и ту же экзогенную UID последовательность;

b. отнесения чтений последовательностей каждого семейства UID к подсемействам Уотсона и Крика на основе пространственного отношения экзогенной UID последовательности к последовательностям чтений R1 и R2;

c. идентификацию нуклеотидной последовательности в качестве точно представленной цепи Уотсона анализируемого фрагмента ДНК, если пороговый процент членов подсемейства Уотсона содержит такую последовательность;

d. идентификацию нуклеотидной последовательности в качестве точно представленной цепи Крика анализируемого фрагмента ДНК, если пороговый процент членов подсемейства Крика содержит такую последовательность;

e. идентификацию мутации в нуклеотидной последовательности, точно представляющей цепь Уотсона, если последовательность, точно представляющая цепь Уотсона, отличается от референсной последовательности, в которой такая мутация отсутствует;

f. идентификацию мутации в нуклеотидной последовательности, точно представляющей цепь Крика, если последовательность, точно представляющая цепь Крика, отличается от референсной последовательности, в которой такая мутация отсутствует;

g. идентификацию мутации в анализируемом фрагменте ДНК, если мутация в нуклеотидной последовательности, точно представляющей цепь Уотсона, и мутация в нуклеотидной последовательности, точно представляющей цепь Крика, являются одной и той же мутацией.

В некоторых вариантах осуществления машиночитаемый носитель содержит исполняемые инструкции для отнесения членов семейства UID к подсемейству Уотсона, если экзогенная UID последовательность расположена непосредственно после или в пределах 1-300 нуклеотидов от сайта связывания праймера для секвенирования R2. В некоторых вариантах осуществления машиночитаемый носитель содержит исполняемые инструкции для отнесения членов семейства UID к подсемейству Крика, если экзогенная UID последовательность расположена непосредственно после или в пределах 1-300 нуклеотидов от сайта связывания праймера для секвенирования R1. В некоторых вариантах осуществления машиночитаемый носитель содержит исполняемые инструкции для картирования чтений последовательностей в референсном геноме. В некоторых

вариантах осуществления референсный геном представляет собой референсный геном человека.

В некоторых вариантах осуществления машиночитаемый носитель дополнительно содержит машиноисполняемые инструкции для формирования отчета о возможных методах лечения на основе присутствия, отсутствия или количества мутации в образце. В некоторых вариантах осуществления машиночитаемый носитель дополнительно содержит машиноисполняемый код, который обеспечивает передачу данных по сети.

Также в настоящем документе предложены компьютерные системы, включающие:

a. блок памяти, выполненный с возможностью приема данных о последовательности из образца нуклеиновой кислоты, где данные генерируются способом согласно любому из предыдущих пунктов;

b. процессор, соединенный с возможностью связи с блоком памяти, где процессор содержит машиночитаемый носитель согласно любому из предыдущих пунктов.

В некоторых вариантах осуществления компьютерная система дополнительно включает систему секвенирования, выполненную с возможностью передачи данных в блок памяти. В некоторых вариантах осуществления компьютерная система дополнительно включает пользовательский интерфейс, выполненный с возможностью передачи или отображения отчета пользователю. В некоторых вариантах осуществления компьютерная система дополнительно включает цифровой процессор, выполненный с возможностью передачи результатов анализа данных по сети.

Также в настоящем документе предложены системы, включающие:

a. популяцию фрагментов двухцепочечной ДНК из биологического образца;

b. набор 3'-адаптеров, как описано в любом из предыдущих пунктов;

c. множество 5'-адаптеров, как описано в любом из предыдущих пунктов;

d. реагенты для проведения реакции, подобной ник-трансляции (например, с ДНК-полимеразой, специфичной к липким концам лигазой и урацил-ДНК-гликозилазой);

e. реагенты для обогащения ампликонов одним или более полинуклеотидами-мишенями; а также

f. систему секвенирования.

В некоторых вариантах осуществления система дополнительно включает компьютерную систему согласно любому из предыдущих пунктов.

Также в настоящем документе предложены наборы, включающие:

a. первый набор пар мишень-селективных праймеров Уотсона, включающий: (i) один или больше первых мишень-селективных праймеров Уотсона, включающих последовательность, комплементарную части универсальной 3'-адаптерной последовательности, где, необязательно, часть универсальной 3'-адаптерной последовательности представляет собой сайт праймера для секвенирования R2 универсальной 3'-адаптерной последовательности, и (ii) один или больше вторых мишень-селективных праймеров Уотсона, причем каждый из одного или больше вторых мишень-селективных праймеров Уотсона включает мишень-селективную последовательность;

b. первый набор пар мишень-селективных праймеров Крика, включающий: (i) один или больше мишень-селективных праймеров Крика, включающих последовательность, комплементарную части универсальной 5'-адаптерной последовательности, где, необязательно, часть универсального 5'-адаптера представляет собой сайт праймера для секвенирования R1 универсальной 5'-адаптерной последовательности, и (ii) один или больше вторых, мишень-селективных праймеров Крика, причем каждый из одного или больше вторых мишень-селективных праймеров Крика включает такую же мишень-селективную последовательность, как последовательность второго мишень-селективного праймера Уотсона;

c. второй набор пар мишень-селективных праймеров Уотсона, включающий: (i) один или больше третьих мишень-селективных праймеров Уотсона, включающих последовательность, комплементарную сайту праймера для секвенирования R2 универсальной 3'-адаптерной последовательности, и (ii) один или больше четвертых мишень-селективных праймеров Уотсона, причем каждый из одного или больше четвертых мишень-селективных праймеров Уотсона включает, в направлении 5'→3', сайт праймера для секвенирования R1 и мишень-селективную последовательность, селективную в отношении одного и того же полинуклеотида-мишени; и

d. второй набор мишень-селективных праймеров Крика, включающий: (i) один или больше третьих мишень-селективных праймеров Крика, включающих последовательность, комплементарную сайту праймера для секвенирования R1 универсальной 3'-адаптерной последовательности, и (ii) один или больше четвертых мишень-селективных праймеров Крика, причем каждый из одного или больше четвертых мишень-селективных праймеров Крика включает, в направлении 5'→3', сайт праймера для секвенирования R2 и мишень-селективную последовательность, селективную в отношении одного и того же полинуклеотида-мишени.

Если не указано иное, все используемые в настоящем документе технические и научные термины имеют то же значение, которое обычно известно среднему специалисту в области, к которой относится настоящее изобретение. Хотя способы и материалы, аналогичные или эквивалентные описанным в настоящем документе, могут использоваться для осуществления изобретения, подходящие способы и материалы описаны ниже. Все публикации, заявки на патенты, патенты и другие источники, указанные в настоящем документе, полностью включены посредством отсылки. В случае противоречия настоящее описание, включая определения, будет иметь преимущественную силу. Кроме того, материалы, способы и примеры носят исключительно иллюстративный характер и не предназначены для ограничения.

Подробности одного или больше вариантов осуществления изобретения представлены на прилагаемых чертежах и в описании ниже. Другие признаки, цели и преимущества изобретения будут очевидны из описания и чертежей, а также из формулы изобретения.

ОПИСАНИЕ ЧЕРТЕЖЕЙ

Фигура 1 содержит схему иллюстративного метода дуплексной якорной ПЦР. Дуплексные адаптеры с молекулярными штрих-кодами лигируют на концах фрагментов нуклеиновых кислот, имеющих тупые концы, с получением библиотеки дуплексов для секвенирования, и библиотеку дуплексов для секвенирования подвергают ПЦР с получением амплифицированной библиотеки дуплексов для секвенирования. Продукты амплификации в амплифицированной библиотеке дуплексов для секвенирования делят на две аликвоты, и каждую аликвоту подвергают ПЦР, где цепь Уотсона амплифицируют с первой аликвоты, а цепь Крика амплифицируют со второй аликвоты.

Фигура 2 содержит схему примерного второго раунда амплификации библиотеки, где цепи Уотсона, амплифицированные с первой аликвоты на Фигуре 1, подвергают ПЦР с использованием пары праймеров, где первый праймер биотинилирован, а второй праймер не биотинилирован, с получением библиотеки одноцепочечных ДНК, которую можно использовать для амплификации и оценки цепей Уотсона.

Фигура 3 содержит схему примерного второго раунда амплификации библиотеки, где цепи Крика, амплифицированные с первой аликвоты на Фигуре 1, подвергают ПЦР с использованием пары праймеров, где первый праймер не биотинилирован, а второй праймер биотинилирован, с получением библиотеки одноцепочечных ДНК, которую можно использовать для амплификации и оценки цепей Крика.

Фигура 4 содержит схему иллюстративной амплификации Уотсона.

Фигура 5 содержит схему иллюстративной амплификации Крика.

Фигура 6 содержит схему иллюстративной амплифицированной цепи Уотсона и иллюстративной амплифицированной цепи Крика.

Фигура 7 содержит схему иллюстративной гнездовой амплификации Уотсона.

Фигура 8 содержит схему примерного гнездовой амплификации Крика.

Фигура 9 содержит схему иллюстративного удаления 5'-фосфата.

Фигура 10 содержит схему примерного заполнения на 3'-конце фрагмента амплификации, имеющего выступающий 5'-конец, с получением продукта амплификации с тупым концом.

Фигура 11 содержит схему примерного 3'-дуплексного адаптера, включающего спейсер 3SpC3, экзогенную UID последовательность, содержащую молекулярный штрих-код, и 3'-олигонуклеотид (dT), гибридный с 3'-блокирующей группой, который может расщепляться урацил-ДНК-гликозилазой (УДГ).

Фигура 12 содержит схему примерного лигирования 3'-адаптера с использованием дуплексного 3'-адаптера. 5'-фосфат 3'-дуплексного адаптера лигируют с 3'-концом матрицы нуклеиновой кислоты.

Фигура 13 содержит схему примерного лигирования 5'-адаптера. В одной реакции блокирующая группа 3'-дуплексного адаптера расщепляется, и содержащийся 5'-адаптер лигируется с 5'-концом матрицы нуклеиновой кислоты посредством реакции ник-трансляции.

Фигура 14 содержит схему иллюстративной ПЦР-амплификации библиотеки.

Фигура 15 содержит схему иллюстративной амплификации Уотсона.

Фигура 16 содержит схему иллюстративной гнездовой амплификации Уотсона.

Фигура 17 содержит схему иллюстративной амплификации Крика.

Фигура 18 содержит схему иллюстративной гнездовой амплификации Крика.

Фигура 19 содержит схему конечных продуктов амплификации, полученных в иллюстративной дуплексной якорной ПЦР.

Фигура 20 содержит схему, на которой показано, каким образом секвенирование спаренных концов можно применять, чтобы отличить цепь Уотсона от цепи Крика вводимой нуклеиновой кислоты при использовании конечных продуктов амплификации, полученных в иллюстративной дуплексной якорной ПЦР.

Фигура 21 представлена схема, на которой показано, каким образом секвенирование спаренных концов можно применять, чтобы отличить цепь Уотсона от цепи Крика вводимой нуклеиновой кислоты при использовании конечных продуктов амплификации, полученных в иллюстративной дуплексной якорной ПЦР.

Фигура 22 содержит схему обзора иллюстративного метода SaferSeqS. (a) Подготовка библиотеки начинается с затупления концов (этап 1), при котором молекулы матрицы ДНК дефосфорилируют и затупляют. Затем 3'-адаптер, содержащий последовательность уникального идентификатора (UID) (узкая или широкая диагональная штриховка), лигируют с концами 3'-фрагмента (этап 2). Последовательности UID преобразуют в двухцепочечные штрих-коды после удлинения и лигирования 5'-адаптера (этап 3). Наконец, во время амплификации библиотеки (этап 4) создают избыточные ПЦР-копии каждой исходной молекулы-матрицы. (b) Обогащение мишени достигается с помощью цепь-специфических полугнездовых ПЦР. Амплифицированную библиотеку разделяют на Уотсон- и Крик-специфические реакции (этап 5), в которых селективно амплифицируют продукты, полученные с одной из цепей ДНК (этап 6). Дополнительная целевая специфичность и включение штрих-кодов образца обеспечивают с помощью второй гнездовой ПЦР (этап 7). Конечные продукты ПЦР (этап 8) подвергают секвенированию спаренных концов (этап 9). Эндогенный штрих-код представляет собой конец фрагмента матрицы перед конструированием библиотеки. (c) После секвенирования определяют, что чтения получены с цепи Уотсона или Крика. Поскольку каждая цепь исходных молекул-матриц помечена одним и тем же экзогенным штрих-кодом и имеет один и тот же эндогенный штрих-код, чтения, полученные с каждой из двух цепей одного и того же дуплекса исходной ДНК, могут быть сгруппированы вместе в семейство дуплексов. Разная штриховка и пунктир на правых концах цепей соответствуют разным штрих-кодам. В показанных примерах каждое семейство дуплексов содержит по восемь членов, четыре из которых представляют цепь Уотсона, и четыре представляют цепь Крика. В реальных экспериментах, описанных в данном документе, каждое семейство содержит по меньшей мере два члена с цепи Уотсона и два члена с цепи Крика, причем фактическое количество зависит от глубины секвенирования. *Vona fide* мутации, обозначенные звездочками внутри семейства *Vona fide* мутаций, присутствуют в обеих

исходных цепях ДНК дуплекса и, следовательно, встречаются как в семействе Уотсона, так и в семействе Крика. Напротив, ошибки ПЦР или секвенирования, обозначенные звездочками в семействе ошибок ПЦР или секвенирования, ограничены подмножеством чтений одной из двух цепей. Артефакты, специфические для цепей Уотсона (звездочки в семействе поврежденных цепей Уотсона) и специфические для цепей Крика (звездочки в семействе поврежденных цепей Крика), встречаются во всех копиях семейства Уотсона или Крика, но не в обоих семействах.

Фигура 23 содержит график, на котором показана аналитическая эффективность метода SaferSeqS. Частоты мутантных аллелей (MAF), определенные с помощью SaferSeqS, в сравнении с ожидаемыми частотами, когда ДНК из злокачественной опухоли, содержащей известную мутацию, смешивали с ДНК лейкоцитов здорового донора в соотношениях, варьирующих от 10% до 0,001%. Также был проанализирован 0% контрольный образец для определения специфичности представляющей интерес мутации. Сплошная линия представляет собой приближение модели линейной регрессии, в которой точка пересечения оси y была зафиксирована на нуле (наклон=0,776, $R^2 > 0,999$, $P = 3,95 \times 10^{-15}$).

Фигура 24. Высокое выделение дуплекса и эффективное обогащение мишеней с применением SaferSeqS. Тридцать три нг смешанных образцов вкДНК исследовали на присутствие одной из трех разных мутаций в TP53 (p.L264fs, p.P190L или p.R342X). Для каждого образца вкДНК получали три библиотеки, каждая из которых содержала ~11 нг вкДНК. (a) Медианное количество семейств дуплексов (т.е. цепей Уотсона и Крика, содержащих одни и те же эндогенные и экзогенные штрих-коды) составляло 89% (диапазон: от 65% до 102%) от количества исходных молекул-матриц. (b) Медианный процент целевых чтений составил 80% (диапазон: от 72% до 91%). Нижняя и верхняя петли соответствуют 25-му и 75-му процентилю, усы продолжают до 1,5-кратного межквартильного диапазона. Отдельные точки накладываются случайным образом для удобства визуализации.

Фигура 25 содержит графики, на которых показано обнаружение примерных мутаций в образцах жидкостной биопсии. Анализ 33 нг внеклеточной ДНК здоровых лиц, смешанной с внеклеточной ДНК плазмы крови онкобольного. Смеси создавали для получения высокой частоты (~0,5-1%) мутации, низкой частоты (~0,01-0,1%) мутации или без мутации. Смешанный образец TP53 p.R342X анализировали с помощью: (a) SafeSeqs или (b) SaferSeqS. Аналогичным образом, смешанный образец TP53 p.L264fs анализировали с помощью: (c) SafeSeqs и (d) SaferSeqS, и смешанный образец TP53 p.P190L анализировали с помощью: (e) SafeSeqs и (f) SaferSeqS. Номера мутаций обозначают каждую из 153 различных мутаций, наблюдаемых с помощью SafeSeqS (определено в таблице 8). Единственный "суперкалимутант", обнаруженный с помощью SafeSeqS (Таблица 9), находился за пределами геномной области, проанализированной с использованием SafeSeqS, и поэтому не показан.

Фигура 26. Ошибки SaferSeqS по сравнению с ошибками независимых от цепи,

основанных на лигировании методов молекулярного баркодирования. Анализ 33 нг внеклеточной ДНК здоровых лиц, смешанной с внеклеточной ДНК из плазмы крови онкобольного. Смеси приготавливали с получением высокой частоты (~0,5-1%) мутации, низкой частоты (~0,01-0,1%) мутации или без мутации. Смешанный образец TP53 p.R342X исследовали с помощью SaferSeqS, но при этом: (a) информацию о цепи в анализе игнорировали для имитации независимые от цепи методом молекулярного баркодирования на основе лигирования, или (b) информацию о цепи учитывали во время обнаружения мутации. Аналогичным образом смешанный образец TP53 p.L264fs исследовали: (c) без учета информации о цепи и (d) с применением метода SaferSeqS. Смешанный образец TP53 p.P190L исследовали аналогичным образом: (e) без учета информации о цепи и (f) с применением SaferSeqS. Номера мутаций определены в дополнительной Таблице 3. Звездочками обозначены смешанные мутации. Единственный неожиданный суперкалимутант, обнаруженный с помощью SaferSeqS, показан на (e).

Фигура 27. Исследование образцов плазмы крови онкобольных. Образцы внеклеточной ДНК плазмы крови пяти онкобольных, несущих восемь известных мутаций с частотой от 0,01% до 0,1%, исследовали с помощью ранее описанного метода молекулярного баркодирования на основе ПЦР ("SafeSeqS", а не "SaferSeqS") и с помощью SaferSeqS. Номера мутаций указаны в Таблице 11. Звездочками обозначены ожидаемые мутации. Единственный неожиданный суперкалимутант, обнаруженный с помощью SaferSeqS (Таблица 11), находился за пределами геномной области, проанализированной с помощью SafeSeqS, и поэтому не показан.

Фигура 28. Влияние эффективности ПЦР и числа циклов при выделении дуплекса. Вероятность выделения обеих цепей исходных дуплексов ДНК (ось y) отложена на кривых в зависимости от числа циклов амплификации библиотеки (ось x). Каждая панель на фигуре представляет предполагаемую эффективность ПЦР, указанную в верхней части панели. Показана доля продукта амплификации библиотеки, используемого в цепь-специфических ПЦР. Количество циклов амплификации библиотеки варьировало от одного до 11. Эффективность ПЦР варьировала от 100% до 50% с шагом 10%. Доля продукта амплификации библиотеки, используемого в каждой цепь-специфической ПЦР, варьировала от 50% до 1,4%. Вероятностное моделирование проводили, как описано в Примере 2.

На Фигуре 29 представлены графики, на которых показана мультиплексная панель для обнаружения иллюстративных драйверных мутаций генов, приводящих к возникновению рака. Выделение и охват 36 ампликонов, которые были успешно амплифицированы в мультиплексной панели. По горизонтальной оси показано положение после 3'-конца второго генспецифического праймера (GSP2). Постепенное снижение охвата с увеличением расстояния от 3'-конца праймера является следствием вводного паттерна фрагментации ДНК. Детали, касающиеся теоретического выхода чтений с определенной длиной ампликона, обсуждаются в Примере 2.

Фигура 30. Эффективность 48 пар праймеров, используемых в мультиплексной

панели для анализа областей драйверных генов, обычно соержащих мутацию при раке. Процент целевых чтений (т.е. доля от общего числа чтений, картируемых на заданной мишени) для каждой из 48 пар праймеров SaferSeqS, используемых в цепь-специфических ПЦР. Праймеры использовали в эквимоллярных концентрациях в каждой генспецифической ПЦР.

Фигура 31. Эффективность 62 пар праймеров. Процент целевых чтений (т.е. доля от общего числа чтений, соответствующих заданной мишени) для каждой из 62 пар праймеров SaferSeqS, протестированных в данный момент. Пятьдесят из 62 пар (81%) демонстрируют показатель целевого охвата больше 50%. Представленные результаты отражают одну попытку подбора праймеров.

На Фигуре 32 представлена иллюстративная компьютерная система, предназначенная для предоставления пользователю возможности проводить анализ образца нуклеиновой кислоты в соответствии со способом, описанным в настоящем документе.

ПОДРОБНОЕ ОПИСАНИЕ

Следует отметить, что при использовании в описании и прилагаемой формуле изобретения формы единственного числа "a", "an" и "the" включают ссылки в множественном числе, если из контекста прямо не следует иное.

"Нуклеотиды" и "нт" используются в настоящем документе попеременно для общего обозначения биологических молекул, которые включают нуклеиновые кислоты. Нуклеотиды могут содержать остатки, которые содержат известные пуриновые и пиримидиновые основания. Нуклеотиды могут содержать другие гетероциклические основания, которые были модифицированы. Такие модификации включают, например, метилированные пурины или пиримидины, ацилированные пурины или пиримидины, алкилированные рибозы или другие гетероциклы. Термины "полинуклеотиды", "нуклеиновая кислота" и "олигонуклеотиды" могут использоваться взаимозаменяемо. Они могут относиться к полимерной форме нуклеотидов любой длины, дезоксирибонуклеотидов или рибонуклеотидов, либо их аналогам. Полинуклеотиды могут иметь любую трехмерную структуру и выполнять любую известную или неизвестную функцию. Ниже приведены неограничивающие примеры полинуклеотидов: кодирующие или некодирующие области гена или фрагмента гена, локусы (локус), определенные в результате анализа сцепления, экзоны, интроны, матричная РНК (мРНК), транспортная РНК, рибосомная РНК, рибозимы, кДНК, рекомбинантные полинуклеотиды, разветвленные полинуклеотиды, плазмиды, векторы, выделенная ДНК любой последовательности, выделенная РНК любой последовательности, зонды на основе нуклеиновых кислот и праймеры. Полинуклеотид может включать не встречающиеся в природе последовательности. Полинуклеотид может включать модифицированные нуклеотиды, такие как метилированные нуклеотиды и аналоги нуклеотидов. Модификации нуклеотидной структуры, в случае присутствия, могут быть введены до или после сборки полимера. Последовательность нуклеотидов может быть прервана

ненуклеотидными компонентами. Полинуклеотид может быть дополнительно модифицирован после полимеризации, например, путем конъюгирования с метящим компонентом.

"Праймер" обычно представляет собой полинуклеотидную молекулу, включающую нуклеотидную последовательность (например, олигонуклеотид), обычно со свободной 3'-ОН группой, которая гибридизуется с матричной последовательностью (такой как полинуклеотид-мишень или продукт удлинения праймера) и способен вызывать полимеризацию полинуклеотида, комплементарного матрице.

Термин "млекопитающее" при использовании в настоящем документе включает как человека, так и не человека, и включает, без ограничения этим, человека, приматов за исключением человека, псовых, кошачьих, мышиных, бычьих, лошадиных и свинообразных.

Обзор

Данный документ относится к способам и материалам, которые могут применяться для точной идентификации мутации, присутствующей в образце нуклеиновой кислоты. В некоторых аспектах способ включает идентификацию мутации, если она одновременно присутствует в цепях Уотсона и в цепях Крика двухцепочечной матрицы нуклеиновой кислоты. Такие способы особенно полезны, если нужно отличить истинные мутации от артефактов, возникающих, например, из-за повреждения ДНК, при ПЦР, а также других артефактов секвенирования, что позволяет идентифицировать мутации с высокой степенью достоверности.

В некоторых случаях способы и материалы, описанные в настоящем документе, позволяют обнаруживать одну или более мутаций с низкой частотой ошибок. Например, способы и материалы, описанные в настоящем документе, могут применяться для обнаружения присутствия или отсутствия мутации нуклеиновой кислоты в матрице нуклеиновой кислоты с частотой ошибки меньше чем приблизительно 1% (например, меньше чем приблизительно 0,1%, меньше чем приблизительно 0,05% или меньше чем приблизительно 0,01%). В некоторых случаях способы и материалы, описанные в настоящем документе, могут применяться для обнаружения присутствия или отсутствия мутации нуклеиновой кислоты в матрице нуклеиновой кислоты с частотой ошибки от приблизительно 0,001% до приблизительно 0,01%. В некоторых случаях частота ошибки, связанная с идентификацией одной или более мутаций в анализируемых фрагментах ДНК согласно способу, описанному в настоящем документе, составляет не больше чем 1×10^{-2} , не больше чем 1×10^{-3} , не больше чем 1×10^{-4} , не больше чем 1×10^{-5} , не больше чем 1×10^{-6} , не больше чем 5×10^{-6} или не больше чем 1×10^{-7} . В некоторых случаях частота ошибок, связанных с идентификацией одной или более мутаций в анализируемых фрагментах ДНК согласно способу, описанному в настоящем документе, снижена по меньшей мере в 2 раза, 4 раза, 5 раз, 10 раз, 20 раз, 30 раз, 40 раз, 50 раз, 60 раз, 70 раз, 80 раз, 90 раз или 100 раз по сравнению с альтернативным способом идентификации мутаций, который не требует одновременного обнаружения мутации в цепях Уотсона и в цепях Крика

анализируемого фрагмента ДНК.

В некоторых вариантах осуществления альтернативный способ включает стандартное молекулярное баркодирование или стандартное молекулярное баркодирование на основе ПЦР с последующим секвенированием. В конкретных вариантах осуществления альтернативный способ включает: (a) присоединение адаптеров к популяции фрагментов двухцепочечной ДНК в анализируемом образце ДНК, где адаптеры включают уникальный экзогенный UID; (b) проведение первичной амплификации для амплификации лигированных с адаптером двухцепочечных фрагментов ДНК с получением ампликонов; (c) определение чтений последовательности одного или больше ампликонов одного или больше лигированных с адаптером фрагментов двухцепочечной ДНК; (d) отнесение чтений последовательностей к семействам UID, где каждый член семейства UID включает одну и ту же экзогенную UID последовательность; (e) идентификацию нуклеотидной последовательности в виде точно представленного анализируемого фрагмента ДНК, если пороговый процент членов семейства UID содержит такую последовательность; и (f) идентификацию мутации во фрагменте ДНК анализируемого вещества, если последовательность, идентифицированная в виде точно представленного анализируемого фрагмента ДНК, отличается от референсной последовательности, в которой такая мутация отсутствует.

В некоторых случаях способы и материалы, описанные в настоящем документе, могут применяться для достижения эффективного выделения дуплекса. Например, способы, описанные в настоящем документе, могут применяться для выделения продуктов ПЦР-амплификации, полученных как с цепи Уотсона, так и с цепи Крика матричной двухцепочечной нуклеиновой кислоты. В некоторых случаях способы, описанные в настоящем документе, могут применяться для достижения по меньшей мере 50% (например, приблизительно 50%, приблизительно 60%, приблизительно 70%, приблизительно 75%, приблизительно 80%, приблизительно 82%, приблизительно 85%, приблизительно 88%, приблизительно 90%, приблизительно 93%, приблизительно 95%, приблизительно 97%, приблизительно 99% или 100%) выделения дуплекса.

В некоторых случаях способы и материалы, описанные в настоящем документе, могут применяться для обнаружения мутации с низкой частотой мутантного аллеля. Например, способы, описанные в настоящем документе, могут применяться для обнаружения мутации, имеющей низкую частоту мутантного аллеля меньше чем приблизительно 1% (например, меньше чем приблизительно 0,1%, меньше чем приблизительно 0,05% или меньше чем приблизительно 0,01%). В некоторых случаях способы, описанные в данном документе, могут применяться для обнаружения мутации, имеющей низкую частоту аллеля приблизительно 0,001%.

В некоторых случаях описанные в настоящем документе способы могут применяться для обнаружения мутации, которая присутствует в образце анализируемой нуклеиновой кислоты с частотой 0,1% или меньше. В некоторых вариантах осуществления способы, описанные в настоящем документе, могут применяться для

обнаружения мутации, которая присутствует в образце анализируемой нуклеиновой кислоты с частотой от 0,1% до 0,00001%. В некоторых вариантах осуществления способы, описанные в настоящем документе, могут применяться для обнаружения мутации, которая присутствует в образце анализируемой нуклеиновой кислоты с частотой от 0,1% до 0,01%.

В некоторых случаях способы и материалы, описанные в настоящем документе, могут применяться для обнаружения мутации с минимальными фоновыми артефактными мутациями (или без них). В некоторых случаях способы, описанные в настоящем документе, могут применяться для обнаружения мутации с менее чем 0,01% фоновых артефактных мутаций. В некоторых случаях способы, описанные в настоящем документе, могут применяться для обнаружения мутации без фоновых артефактных мутаций.

В некоторых случаях способы обнаружения одной или более мутаций, присутствующих на обеих цепях двухцепочечной нуклеиновой кислоты, могут включать создание библиотеки дуплексов для секвенирования, где каждая нуклеиновая кислота в библиотеке имеет дуплексный молекулярный штрих-код на каждом конце (например, 5'-конце и 3'-конце), создание библиотеки полученных с цепи Уотсона одноцепочечных последовательностей и библиотеки полученных с цепи Крика одноцепочечных последовательностей из библиотеки дуплексов для секвенирования, и обнаружение присутствия одной или больше мутаций, присутствующих на обеих цепях двухцепочечной нуклеиновой кислоты в каждой библиотеке одноцепочечных последовательностей. Присутствие первого молекулярного штрих-кода в 3'-дуплексном адаптере и второго молекулярного штрих-кода в 5'-адаптере можно использовать, чтобы отличить продукты амплификации, полученные с цепи Уотсона, от продуктов амплификации, полученных с цепи Крика.

В некоторых случаях способы идентификации мутации включают: (a) присоединение частично двухцепочечных 3'-адаптеров к 3'-концам цепей Уотсона и Крика популяции фрагментов двухцепочечной ДНК в анализируемом образце ДНК, где первая цепь частично двухцепочечного 3'-адаптера включает, в направлении 5'→3': (i) первый сегмент, (ii) экзогенную UID последовательность, (iii) сайт отжига для 5'-адаптера и (iv) универсальную 3'-адаптерную последовательность, включающую сайт праймера для секвенирования R2, и где вторая цепь частично двухцепочечного 3'-адаптера включает, в направлении 5'→3': (i) сегмент, комплементарный первому сегменту, и (ii) 3'-блокирующую группу, где, необязательно, вторая цепь является расщепляемой; (b) отжиг 5'-адаптеров с 3'-адаптерами посредством сайта отжига, где 5'-адаптеры включают, в направлении 5'→3': (i) универсальную 5'-адаптерную последовательность, которая не комплементарна универсальной 3'-адаптерной последовательности, и которая включает сайт праймера для секвенирования R1, и (ii) последовательность, комплементарную сайту отжига для 5'-адаптера; (c) проведение реакции, подобной ник-трансляции, для удлинения 5'-адаптеров через экзогенную UID последовательность 3'-адаптеров (например, с использованием ДНК-полимеразы) и ковалентного связывания удлиненного 5'-адаптера с

5'-концами цепей Уотсона и Крика двухцепочечных ДНК-фрагментов (например, с использованием лигазы); (d) проведение первичной амплификации для амплификации лигированных с адаптером двухцепочечных фрагментов ДНК с получением ампликонов; (e) определение чтений последовательностей одного или больше ампликонов одного или больше лигированных с адаптером фрагментов двухцепочечной ДНК; (f) отнесение чтений последовательностей к семействам UID, где каждый член семейства UID включает одну и ту же экзогенную UID последовательность; (g) отнесение чтений последовательностей каждого семейства UID к подсемейству Уотсона и подсемейству Крика на основе пространственного отношения экзогенной последовательности UID к последовательностям чтения R1 и R2; (h) идентификацию нуклеотидной последовательности в качестве точно представленной цепи Уотсона анализируемого фрагмента ДНК, если пороговый процент членов подсемейства Уотсона содержит такую последовательность; (i) идентификацию нуклеотидной последовательности в качестве точно представленной цепи Крика анализируемого фрагмента ДНК, если пороговый процент членов подсемейства Крика содержит такую последовательность; (j) идентификацию мутации в нуклеотидной последовательности, точно представляющей цепь Уотсона, если последовательность, точно представляющая цепь Уотсона, отличается от референсной последовательности, в которой такая мутация отсутствует; (k) идентификацию мутации в нуклеотидной последовательности, точно представляющей цепь Крика, если последовательность, точно представляющая цепь Крика, отличается от референсной последовательности, в которой такая мутация отсутствует; и (l) идентификацию мутации в анализируемом фрагменте ДНК, если мутация в нуклеотидной последовательности, точно представляющей цепь Уотсона, и мутация в нуклеотидной последовательности, точно представляющей цепь Крика, представляют собой одну и ту же мутацию.

В некоторых случаях способы идентификации мутации включают: (a) присоединение адаптеров к популяции двухцепочечных фрагментов ДНК, где адаптеры включают двухцепочечную часть, включающую экзогенный UID, и разветвленную часть, включающую: (i) одноцепочечную 3'-адаптерную последовательность, включающую сайт праймера для секвенирования R2, и (ii) одноцепочечную 5'-адаптерную последовательность, включающую сайт праймера для секвенирования R1; (b) проведение первичной амплификации для амплификации лигированных с адаптером двухцепочечных фрагментов ДНК с получением ампликонов; (c) селективную амплификацию ампликонов цепей Уотсона, включающих полинуклеотидную последовательность-мишень, с использованием первого набора пар мишень-селективных праймеров Уотсона, причем первый набор пар мишень-селективных праймеров Уотсона включает: (i) первый мишень-селективный праймер Уотсона, включающий последовательность, комплементарную сайту праймера для секвенирования R2 универсальной 3'-адаптерной последовательности, и (ii) второй мишень-селективный праймер Уотсона, включающий мишень-селективную последовательность, с получением в результате целевых продуктов амплификации

Уотсона; (d) селективную амплификацию ампликонов цепей Крика, включающих такую же полинуклеотидную последовательность-мишень, с использованием первого набора мишень-селективных праймеров Крика, причем первый набор пар мишень-селективных праймеров Крика включает: первый мишень-селективный праймер Крика, включающий последовательность, комплементарную сайту праймера для секвенирования R1 универсальной 5'-адаптерной последовательности, и (ii) второй мишень-селективный праймер Крика, включающий такую же мишень-селективную последовательность, как и вторая мишень-селективная последовательность праймера Крика, с получением в результате целевых продуктов амплификации Крика; (e) определение чтений последовательностей целевых продуктов амплификации Уотсона и целевых продуктов амплификации Крика; (f) отнесение чтений последовательностей к семействам UID, где каждый член семейства UID включает одну и ту же экзогенную UID последовательность; (g) отнесение чтений последовательностей каждого семейства UID к подсемейству Уотсона и подсемейству Крика на основе пространственного отношения экзогенной UID последовательности к последовательностям чтения R1 и R2; (h) идентификацию нуклеотидной последовательности в качестве точно представленной цепи Уотсона анализируемого фрагмента ДНК, если пороговый процент членов семейства Уотсона содержит такую последовательность; (i) идентификацию нуклеотидной последовательности в качестве точно представленной цепи Крика анализируемого фрагмента ДНК, если пороговый процент членов семейства Крика содержит такую последовательность; и (j) идентификацию мутации в анализируемом фрагменте ДНК, если нуклеотидная последовательность, точно представляющая цепь Уотсона, и нуклеотидная последовательность, точно представляющая цепь Крика, содержат одну и ту же мутацию.

В некоторых случаях способы и материалы, описанные в настоящем документе, могут применяться для независимой оценки каждой цепи двухцепочечной нуклеиновой кислоты. Например, если мутацию нуклеиновой кислоты идентифицируют в независимо оцениваемых цепях двухцепочечной нуклеиновой кислоты, как описано в настоящем документе, материалы и способы, описанные в настоящем документе, могут применяться для определения того, из какой цепи двухцепочечной нуклеиновой кислоты происходит мутация нуклеиновой кислоты.

Любой подходящий способ может использоваться для создания библиотеки дуплексов для секвенирования. При использовании в настоящем документе библиотека дуплексов для секвенирования представляет собой множество фрагментов нуклеиновых кислот, включающих дуплексный молекулярный штрих-код на одном конце (например, на 5'-конце и/или 3'-конце) каждого фрагмента нуклеиновой кислоты в библиотеке, и может обеспечивать возможность секвенирования обеих цепей двухцепочечной нуклеиновой кислоты. В некоторых случаях образец нуклеиновой кислоты может быть фрагментирован с получением фрагментов нуклеиновой кислоты, при этом полученные фрагменты нуклеиновой кислоты можно использовать для создания библиотеки дуплексов для секвенирования. Фрагменты нуклеиновой кислоты, используемые для создания

библиотеки дуплексов для секвенирования, также могут именоваться в настоящем документе вводящей нуклеиновой кислотой. Например, когда фрагменты нуклеиновой кислоты, используемые для создания библиотеки дуплексов для секвенирования, представляют собой фрагменты ДНК, эти фрагменты ДНК также могут именоваться в настоящем документе вводящей ДНК. Библиотека дуплексов для секвенирования может включать любое подходящее количество фрагментов нуклеиновых кислот. В некоторых случаях создание библиотеки дуплексов для секвенирования может включать фрагментацию матрицы нуклеиновой кислоты и лигирование адаптеров с каждым концом каждого фрагмента нуклеиновой кислоты в библиотеке.

Анализируемый образец нуклеиновой кислоты

Матрицы нуклеиновых кислот в образце анализируемой нуклеиновой кислоты могут включать любой тип нуклеиновой кислоты (например, ДНК, РНК и гибриды ДНК/РНК). В некоторых случаях матрица нуклеиновой кислоты может быть матрицей двухцепочечной ДНК. Примеры нуклеиновой кислоты, которые могут использоваться в качестве матрицы в способах, описанных в настоящем документе, включают, без ограничения, геномную ДНК, свободную циркулирующую ДНК (сцДНК; например, циркулирующую опухолевую ДНК (цоДНК) и внеклеточную фетальную ДНК (вкфДНК)).

В некоторых вариантах осуществления матрицы нуклеиновой кислоты в образце нуклеиновой кислоты представляют собой фрагменты нуклеиновых кислот, например, фрагменты ДНК. В некоторых вариантах осуществления концы фрагмента ДНК представляют собой уникальные последовательности, которые могут использоваться в качестве эндогенного уникального идентификатора такого фрагмента. В некоторых вариантах осуществления фрагменты получают вручную. В некоторых вариантах осуществления фрагменты получают с помощью фрагментации, например, ферментативной фрагментации, фрагментации с использованием химических средств, акустической фрагментации, распыления, центрифужной фрагментации, фрагментации при дросселировании в тонкой трубке, игольной фрагментации, при обработке ультразвуком, рестрикционными эндонуклеазами, неспецифическими нуклеазами (например, ДНКазой I) и т.п. В некоторых вариантах осуществления фрагменты не получают вручную. В некоторых вариантах осуществления фрагменты получены из образца вкДНК.

В некоторых вариантах осуществления фрагмент нуклеиновой кислоты в образце нуклеиновой кислоты имеет некую длину. Длина может составлять приблизительно 4-1000 нуклеотидов. Длина может составлять приблизительно 60-300 нуклеотидов. Длина может составлять приблизительно 60-200 нуклеотидов. Длина может составлять приблизительно 140-170 нуклеотидов. Длина может быть меньше 500, меньше 400, меньше 300, меньше 250 или меньше 200 нуклеотидов.

В некоторых вариантах осуществления в качестве эндогенных UID используются концы матричных нуклеиновых кислот. Квалифицированный специалист может определить длину эндогенного UID, необходимую для уникальной идентификации

матрицы нуклеиновой кислоты, при использовании таких факторов, как, например, общая длина матрицы, сложность матриц нуклеиновой кислоты в разделенном или исходном образце нуклеиновой кислоты и т.п. В некоторых вариантах осуществления в качестве эндогенных UID используются 10-500 нуклеотидов на концах матриц нуклеиновых кислот. В некоторых вариантах осуществления в качестве эндогенных UID используются 15-100 нуклеотидов на концах матриц нуклеиновых кислот. В некоторых вариантах осуществления в качестве эндогенных UID используются 15-40 нуклеотидов на концах матриц нуклеиновых кислот. В некоторых вариантах осуществления по меньшей мере 10 нуклеотидов на концах матриц нуклеиновых кислот используются в качестве эндогенных UID. В некоторых вариантах осуществления в качестве эндогенных UID используются по меньшей мере 15 нуклеотидов на концах матриц нуклеиновых кислот. В некоторых вариантах осуществления в качестве эндогенного UID используется только один конец матрицы нуклеиновой кислоты.

В некоторых вариантах осуществления матрицы нуклеиновых кислот включают один или более полинуклеотидов-мишеней. Термины "полинуклеотид-мишень", "область-мишень", "представляющая интерес матрица нуклеиновой кислоты", "требуемый локус", "требуемая матрица" или "мишень" попеременно используются в настоящем документе для обозначения исследуемого полинуклеотида, представляющего интерес. В некоторых вариантах осуществления полинуклеотид-мишень содержит одну или больше последовательностей, которые представляют интерес и являются предметом исследования. Полинуклеотид-мишень может включать, например, геномную последовательность. Полинуклеотид-мишень может включать последовательность-мишень, присутствие, количество и/или нуклеотидная последовательность которой, или ее изменение, требуется определить.

Полинуклеотид-мишень может быть областью гена, ассоциированного с заболеванием. В некоторых вариантах осуществления ген представляет собой вероятную мишень лекарственного средства. Термин "вероятная мишень лекарственного средства" при использовании в настоящем документе обычно относится к гену или клеточному пути, который поддается модуляции при терапии заболевания. Заболеванием может быть рак. Таким образом, ген может быть известным геном, ассоциированным с раком.

В некоторых вариантах осуществления вводная нуклеиновая кислота, также указанная в настоящем документе образцом нуклеиновой кислоты, была получена из биологического образца. Биологический образец может быть получен от субъекта. В некоторых вариантах осуществления субъект является млекопитающим. Примеры млекопитающих, от которых может быть получена нуклеиновая кислота, используемая в качестве матрицы нуклеиновой кислоты в способах, описанных в настоящем документе, включают, без ограничения, человека, приматов кроме человека (например, обезьян), собак, кошек, овец, кроликов, мышей, хомяков и крыс. В некоторых вариантах осуществления субъектом является человек. В некоторых вариантах осуществления субъектом является растение.

Биологические образцы включают, без ограничения перечисленным, плазму крови, сыворотку, кровь, ткань, образец опухоли, стул, мокроту, слюну, мочу, пот, слезную жидкость, асцитную жидкость, бронхоальвеолярный лаваж, сперму, археологические образцы и образцы судебно-медицинской экспертизы. В конкретных вариантах осуществления биологический образец представляет собой твердый биологический образец, например, образец опухоли. В некоторых вариантах осуществления твердый биологический образец подвергают обработке. Твердый биологический образец может быть обработан путем фиксации в растворе формалина с последующей заливкой в парафин (например, представляет собой образец FFPE). В альтернативе обработка может включать замораживание образца перед проведением анализа на основе зонда. В некоторых вариантах осуществления образец не фиксируют и не замораживают. Нефиксированный, незамороженный образец может, только в качестве примера, храниться в растворе для хранения, предназначенном для хранения нуклеиновой кислоты.

В некоторых вариантах осуществления биологический образец представляет собой жидкий биологический образец. Жидкие биологические образцы включают, без ограничения перечисленным, плазму крови, сыворотку, кровь, мокроту, слюну, мочу, пот, слезы, асцитную жидкость, бронхоальвеолярный лаваж и сперму. В некоторых вариантах осуществления жидкий биологический образец является бесклеточным или по существу не содержит клеток. В конкретных вариантах осуществления биологический образец представляет собой образец плазмы или сыворотки крови. В некоторых вариантах осуществления жидкий биологический образец представляет собой образец цельной крови. В некоторых вариантах осуществления жидкий биологический образец включает мононуклеарные клетки периферической крови.

В некоторых вариантах осуществления образец нуклеиновой кислоты был выделен и очищен из биологического образца. Нуклеиновая кислота может быть выделена и очищена из биологического образца при использовании любых средств, известных в уровне техники. Например, биологический образец может быть обработан для высвобождения нуклеиновой кислоты из клеток или для отделения нуклеиновых кислот от нежелательных компонентов биологического образца (например, белков, клеточных стенок, других нежелательных примесей). Например, нуклеиновая кислота может быть выделена из биологического образца при использовании методов жидкостной экстракции (например, Trizol, DNazol). Нуклеиновую кислоту также можно выделять с использованием доступных в продаже наборов (например, набора Qiagen DNeasy, набора QIAamp, набора Qiagen Midi, набора QIAprep spin).

В некоторых вариантах осуществления биологический образец включает малые количества нуклеиновой кислоты. В некоторых вариантах осуществления биологический образец включает меньше чем приблизительно 500 нанограммов (нг) нуклеиновой кислоты. Например, биологический образец включает от приблизительно 30 нг до приблизительно 40 нг нуклеиновой кислоты.

Нуклеиновую кислоту можно концентрировать с помощью известных способов,

включающих, только в качестве примера, центрифугирование. Нуклеиновая кислота может быть связана на селективной мембране (например, силикагелевой) в целях очистки. Нуклеиновая кислота также может быть обогащена фрагментами требуемой длины, например, фрагментами, которые имеют длину меньше 1000, 500, 400, 300, 200 или 100 пар оснований. Такое обогащение в зависимости от размера могут проводить с использованием, например, вызванного ПЭГ осаждения, электрофоретического геля или хроматографического материала (Huber et al. (1993) *Nucleic Acids Res.* 21:1061-6), гельфильтрационной хроматографии, TSKgel (Kato et al., (1984) *J.Biochem.* 95:83-86), указанные публикации включены в настоящий документ посредством отсылки.

Полинуклеотиды, выделенные из биологического образца, можно селективно осаждают или концентрировать с использованием любых методов, известных в уровне техники.

В некоторых вариантах осуществления образец нуклеиновой кислоты включает меньше чем приблизительно 35 нг нуклеиновой кислоты. Например, образец нуклеиновой кислоты включает, может включать от приблизительно 1 нг до приблизительно 35 нг нуклеиновой кислоты (например, от приблизительно 1 нг до приблизительно 30 нг, от приблизительно 1 нг до приблизительно 25 нг, от приблизительно 1 нг до приблизительно 20 нг, от приблизительно 1 нг до приблизительно 15 нг, от приблизительно 1 нг до приблизительно 10 нг, от приблизительно 1 нг до приблизительно 5 нг, от приблизительно 5 нг до приблизительно 35 нг, от приблизительно 10 нг до приблизительно 35 нг, от приблизительно 15 нг до приблизительно 35 нг, от приблизительно 20 нг до приблизительно 35 нг, от приблизительно 25 нг до приблизительно 35 нг, от приблизительно 30 нг до приблизительно 35 нг, от приблизительно 5 нг до приблизительно 30 нг, от приблизительно 10 нг до приблизительно 25 нг, от приблизительно 15 нг до приблизительно 20 нг, от приблизительно 5 нг до приблизительно 10 нг, от приблизительно 10 нг до приблизительно 15 нг, от приблизительно 15 нг до приблизительно 20 нг, от приблизительно 20 нг до приблизительно 25 нг или от приблизительно 25 нг до приблизительно 30 нг нуклеиновой кислоты). В некоторых случаях образец нуклеиновой кислоты может включать нуклеиновую кислоту из генома, который включает больше чем приблизительно несколько сотен нуклеотидов нуклеиновой кислоты.

В некоторых случаях образец нуклеиновой кислоты может по существу не содержать примесей. Например, когда образец нуклеиновой кислоты представляет собой вкДНК-матрицу, вкДНК может по существу не содержать примесей геномной ДНК. В некоторых случаях образец вкДНК, по существу не содержащий примесей геномной ДНК, может содержать минимальное количество (или не содержать) высокомолекулярной (например, >1000 п.н.) ДНК. В некоторых случаях способы, описанные в настоящем документе, могут включать определение, является ли образец нуклеиновой кислоты по существу не содержащим примеси. Может использоваться любой подходящий способ определения, является ли образец нуклеиновой кислоты по существу не содержащим

примеси. Примеры способов, которые могут использоваться для определения, является ли образец нуклеиновой кислоты по существу не содержащим примеси, включают, например, систему TapeStation и Bioanalyzer. Например, при использовании системы TapeStation и/или Bioanalyzer для определения, является ли образец вкДНК по существу не содержащим примеси геномной ДНК, может использоваться заметный пик на ~180 п.н. (например, соответствующий моонуклеосомной ДНК), который указывает, что образец нуклеиновой кислоты по существу не содержит примесей геномной ДНК.

В некоторых случаях фрагменты нуклеиновых кислот, которые могут использоваться для создания библиотеки дуплексов для секвенирования (например, перед присоединением 3'-дуплексного адаптера к 3'-концам фрагментов нуклеиновых кислот), могут быть подвергнуты репарации концов. Для репарации концов матрицы нуклеиновой кислоты может использоваться любой подходящий метод. Например, реакции затупления (например, лигирование тупых концов) и/или реакции дефосфорилирования могут использоваться для репарации концов матрицы нуклеиновой кислоты. В некоторых случаях затупление может включать заполнение одноцепочечной области. В некоторых случаях затупление может включать отщепление одноцепочечной области. В некоторых случаях для репарации концов матрицы нуклеиновой кислоты могут использоваться реакции затупления и дефосфорилирования, как показано на Фигуре 9 и/или Фигуре 10.

Адаптеры

В некоторых вариантах осуществления способ включает присоединение адаптеров к популяции двухцепочечных фрагментов ДНК с получением популяции связанных с адаптером фрагментов двухцепочечной ДНК.

В некоторых вариантах осуществления адаптеры включают двухцепочечную часть, включающую экзогенный UID, и разветвленную часть, включающую: (i) одноцепочечную 3'-адаптерную последовательность и (ii) одноцепочечную 5'-адаптерную последовательность. В некоторых вариантах осуществления одноцепочечная 3'-адаптерная последовательность не комплементарна одноцепочечной 5'-адаптерной последовательности. В некоторых вариантах осуществления 3'-адаптерная последовательность включает второй (например, R2) сайт праймера для секвенирования, а 5'-адаптерная последовательность включает первый (например, R1) сайт праймера для секвенирования. Следует понимать, что сайты праймеров секвенирования "R1" и "R2" используются в системах секвенирования, которые дают чтения со спаренными концами, например, чтения с противоположных концов секвенируемого фрагмента ДНК. В некоторых вариантах осуществления праймер для секвенирования R1 используется для получения первой популяции чтений с первых концов фрагментов ДНК, а праймер для секвенирования R2 используется для получения второй популяции чтений с противоположных концов фрагментов ДНК. Первая популяция именуется в настоящем документе как "R1" или "чтение 1". Вторая популяция именуется в настоящем документе как "R2" или "Чтение 2". Чтения R1 и R2 могут быть выровнены как "парные чтения" или "спаренные пары", соответствующие каждой цепи анализируемого фрагмента

двухцепочечной ДНК.

В некоторых системах секвенирования, например, Illumina, используются так называемые праймеры "R1" и "R2", и чтения "R1" и "R2". Следует отметить, что термины "R1" и "R2", а также "Чтение 1" и "Чтение 2" в рамках настоящей заявки не ограничиваются тем значением, на которое ссылаются в отношении конкретной платформы секвенирования. Например, в случае использования секвенатора Illumina, праймер "R2" и соответствующее чтение R2, раскрытые в настоящем документе, могут относиться к праймеру и чтению "R2" Illumina, или могут относиться к праймеру и чтению "R1" Illumina, при условии, что праймер "R1" и соответствующее чтение R1, раскрытые в настоящем документе, относятся к другому праймеру и чтению Illumina. Следует пояснить, что в некоторых вариантах осуществления, где праймер "R2", представленный в настоящем документе, является праймером "R1" Illumina, дающим чтения "R1", соответствующий праймер "R1", представленный в настоящем документе, является праймером "R2" Illumina, дающим чтения "R2". Следует пояснить, что в некоторых вариантах осуществления, где праймер "R2", представленный в настоящем документе, является праймером "R2" Illumina, дающим чтения "R2", праймер "R1", представленный в настоящем документе, является праймером "R1" Illumina, дающим чтения R1.

В некоторых вариантах осуществления экзогенный UID уникален для каждого фрагмента двухцепочечной ДНК в образце нуклеиновой кислоты. В некоторых вариантах осуществления экзогенный UID не уникален для каждого фрагмента двухцепочечной ДНК.

В некоторых вариантах осуществления экзогенный UID имеет некую длину. Длина может быть приблизительно 2-4000 нт. Длина может быть приблизительно 6-100 нт. Длина может быть приблизительно 8-50 нт. Длина может быть приблизительно 10-20 нт. Длина может быть приблизительно 12-14 нт. В некоторых вариантах осуществления длина экзогенного UID достаточна для однозначного баркодирования молекул, при этом длина/последовательность экзогенного UID не препятствует последующим этапам амплификации.

В некоторых вариантах осуществления экзогенная UID последовательность не присутствует в матрице нуклеиновой кислоты. В некоторых вариантах осуществления экзогенная UID последовательность не присутствует в требуемой матрице, несущей требуемый локус. Такие уникальные последовательности могут быть созданы путем случайного подбора, например, с помощью машиночитаемого носителя, и отобраны с использованием BLAST в известных базах данных нуклеотидов, таких как, например, EMBL, GenBank или DDBJ. В некоторых вариантах осуществления экзогенная UID последовательность присутствует в матрице нуклеиновой кислоты. В таких случаях положение экзогенной последовательности UID в чтении последовательности используется, чтобы отличить экзогенную UID последовательность от последовательности в матрице нуклеиновой кислоты.

В некоторых вариантах осуществления экзогенная UID последовательность является случайной. В некоторых вариантах осуществления экзогенная UID последовательность представляет собой случайный N-мер. Например, если экзогенная UID последовательность имеет длину шесть нуклеотидов, то она может быть случайным гексамером. Если экзогенная UID последовательность имеет длину 12 нуклеотидов, то она может быть случайным 12-мером.

Экзогенные UID могут быть получены при использовании случайного присоединения нуклеотидов с образованием последовательности, имеющей длину, которая будет использоваться в качестве идентификатора. В каждом положении добавления может использоваться выбор из одного из четырех дезоксирибонуклеотидов. В альтернативе можно использовать выбор из одного из трех, двух или одного дезоксирибонуклеотида. Таким образом, UID может быть полностью случайным, в некоторой степени случайным или неслучайным в некоторых положениях.

В некоторых вариантах осуществления экзогенные UID не являются случайными N-мерами, а выбраны из заданного набора экзогенных UID последовательностей.

Примерные экзогенные UID, подходящие для применения в способах, раскрытых в настоящем документе, описаны в заявке PCT/US2012/033207, которая полностью включена в настоящий документ посредством отсылки.

Разветвленные адаптеры, описанные в настоящем документе, могут быть присоединены к фрагментам двухцепочечной ДНК любыми способами, известными в уровне техники.

В некоторых вариантах осуществления разветвленные адаптеры присоединяют к фрагментам двухцепочечной ДНК путем: (а) присоединения частично двухцепочечных 3'-адаптеров к 3'-концам цепей Уотсона и Крика популяции двухцепочечных фрагментов ДНК, где первая цепь частично двухцепочечного 3'-адаптера включает, в направлении 5'→3': (i) первый сегмент, (ii) экзогенную UID последовательность, (iii) сайт отжига для 5'-адаптера и (iv) универсальную 3'-адаптерную последовательность, включающую сайт праймера для секвенирования R2, и где вторая цепь частично двухцепочечного 3'-адаптера включает, в направлении 5'→3': (i) сегмент, комплементарный первому сегменту, и (ii) 3'-блокирующую группу, где, необязательно, вторая цепь является расщепляемой; (b) отжиг 5'-адаптеров с 3'-адаптерами посредством сайта отжига, где 5'-адаптеры включают, в направлении 5'→3': (i) универсальную 5'-адаптерную последовательность, которая не комплементарна универсальной 3'-адаптерной последовательности и которая включает сайт праймера для секвенирования R1, и (ii) последовательность, комплементарную сайту отжига для 5'-адаптера; и (c) проведение реакции, подобной ник-трансляции, с целью удлинения 5'-адаптеров через экзогенную UID-последовательность 3'-адаптеров (например, с использованием ДНК-полимеразы) и ковалентного связывания удлиненного 5'-адаптера с 5'-концами цепей Уотсона и Крика двухцепочечных фрагментов ДНК (например, с использованием лигазы).

В некоторых вариантах осуществления разветвленные адаптеры присоединяют к

фрагментам двухцепочечной ДНК путем: (а) присоединения 3'-дуплексного адаптера к 3'-концам цепей Уотсона и Крика популяции фрагментов двухцепочечной ДНК. 3'-дуплексный адаптер, также называемый в настоящем документе частично двухцепочечным 3'-адаптером, как описано в настоящем документе, представляет собой олигонуклеотидный комплекс, включающий молекулярный штрих-код, который может иметь первый олигонуклеотид (также называемый в настоящем документе "первой цепью") отождествленный (гибридизованный) со вторым олигонуклеотидом (также называемым в настоящем документе "второй цепью"), в результате чего часть (например, первая часть) 3'-дуплексного адаптера является двухцепочечной, а часть (например, вторая часть) 3'-дуплексного адаптера является одноцепочечной. В некоторых случаях первый олигонуклеотид 3'-дуплексного адаптера, описанный в настоящем документе, содержит первый сегмент, содержащий нуклеотиды, комплементарные нуклеотидам, присутствующим во втором олигонуклеотиде 3'-дуплексного адаптера (например, такой, что первый олигонуклеотид 3'-дуплексного адаптера и второй олигонуклеотид 3'-дуплексного адаптера могут отжигаться в комплементарной области). Примерная структура 3'-дуплексного адаптера может быть такой, как показано на рисунке 11.

Первый олигонуклеотид 3'-дуплексного адаптера, описанного в настоящем документе, может быть олигонуклеотидом, который включает 5'-фосфат и молекулярный штрих-код. Первый олигонуклеотид 3'-дуплексного адаптера, описанного в настоящем документе, может включать любое подходящее количество нуклеотидов. Любой подходящий молекулярный штрих-код может быть включен в первый олигонуклеотид 3'-дуплексного адаптера, описанного в настоящем документе. В некоторых случаях молекулярный штрих-код может включать случайную последовательность. В некоторых случаях молекулярный штрих-код может включать фиксированную последовательность. Примеры молекулярных штрих-кодов, которые могут быть включены в первый олигонуклеотид 3'-дуплексного адаптера, описанного в настоящем документе, включают, помимо прочего, IDT 8, IDT 10, ILMN 8, ILMN 10, доступные от Integrated DNA Technologies. Может использоваться любой подходящий тип молекулярного штрих-кода. В некоторых случаях молекулярный штрих-код содержит экзогенную UID последовательность. Экзогенные UID описаны в настоящем документе. Примеры олигонуклеотидов, которые включают 5'-фосфат и молекулярный штрих-код и которые могут быть включены в первый олигонуклеотид 3'-дуплексного адаптера, описанного в настоящем документе, включают, без ограничений, АТААААСGACGGCNNNNNNNNNNNNNNNNNAGATCGGAAGCACACGTCTGAACTCCAG *T*C (где звездочки обозначают фосфоротиоатные связи; SEQ ID NO:1), где NNNNNNNNNNNNNNNN (SEQ ID NO:2) представляет собой молекулярный штрих-код, и где количество нуклеотидов в молекулярном штрих-коде может составлять от 0 до приблизительно 25.

В некоторых вариантах осуществления первый олигонуклеотид 3'-дуплексного адаптера включает сайт отжига для 5'-адаптера.

В некоторых вариантах осуществления первый олигонуклеотид 3'-дуплексного адаптера включает универсальную 3'-адаптерную последовательность. В некоторых вариантах осуществления универсальная 3'-адаптерная последовательность включает сайт праймера для секвенирования R2.

В некоторых случаях первый олигонуклеотид 3'-дуплексного адаптера, описанный в настоящем документе, также может включать один или больше признаков для предотвращения или уменьшения достройки в ходе ПЦР. Признак, который может предотвращать или уменьшать достройку в ходе ПЦР, может быть любым типом признака (например, химической модификацией). Примеры признака, который может предотвращать или уменьшать достройку в ходе ПЦР и может быть включен в первый олигонуклеотид 3'-дуплексного адаптера, описанного в настоящем документе, включают, без ограничения, 3SpC3 и 3Phos. Признак, который может предотвращать или уменьшать достройку в ходе ПЦР, может быть включен в первый олигонуклеотид 3'-дуплексного адаптера, описанного в настоящем документе, в любом подходящем положении внутри олигонуклеотида. В некоторых случаях молекула, которая может предотвращать или уменьшать достройку в ходе ПЦР, может быть включена в олигонуклеотид. В некоторых случаях молекула для предотвращения или уменьшения достройки в ходе ПЦР может быть включена на конце (например, 5'-конце) олигонуклеотида.

В конкретных вариантах осуществления первый олигонуклеотид 3'-дуплексного адаптера включает 5'-фосфат, первый сегмент, включающий нуклеотиды, которые комплементарны нуклеотидам, присутствующим во втором олигонуклеотиде 3'-дуплексного адаптера, экзогенную UID последовательность, сайт отжига для 5'-адаптера и универсальную 3'-адаптерную последовательность.

Второй олигонуклеотид 3'-дуплексного адаптера, описанный в настоящем документе, может быть олигонуклеотидом, который включает заблокированную 3'-группу (например, для уменьшения или исключения димеризации двух адаптеров). Вторым олигонуклеотидом 3'-дуплексного адаптера, описанного в настоящем документе, может включать любое подходящее количество нуклеотидов. В некоторых вариантах осуществления второй олигонуклеотид 3'-дуплексного адаптера комплементарен первому сегменту первого олигонуклеотида 3'-дуплексного адаптера. Иллюстративный олигонуклеотид, который включает заблокированную 3'-группу и может быть включен во второй олигонуклеотид 3'-дуплексного адаптера, описанного в настоящем документе, включает, без ограничения, GCCGUCGUUUAdT (SEQ ID NO:3).

Второй олигонуклеотид 3'-дуплексного адаптера, описанный в настоящем документе, может быть расщепляемым. Для расщепления второго олигонуклеотида 3'-дуплексного адаптера, описанного в настоящем документе, может использоваться любой подходящий метод. Например, УДГ может использоваться для расщепления второго олигонуклеотида 3'-дуплексного адаптера, описанного в настоящем документе.

В некоторых случаях 3'-дуплексный адаптер, описанный в настоящем документе, может включать первый олигонуклеотид, включающий последовательность

ATAAAACGACGGCNNNNNNNNNNNNNNNNNAGATCGGAAGAGCACACGTCTGAACTCC
AG*T*C/3SpC3 (SEQ ID NO:1), отожденную со вторым олигонуклеотидом, включающим
последовательность GCCGUCGUUUUAdT (SEQ ID NO:3).

В некоторых случаях 3'-дуплексный адаптер, описанный в настоящем документе, может включать коммерчески доступный адаптер. Примеры доступных в продаже адаптеров, которые могут использоваться в качестве (или могут использоваться для создания) 3'-дуплексного адаптера, описанного в настоящем документе, включают, помимо прочего, адаптеры в наборе Accel-NGS 2S DNA Library (Swift Biosciences, номер по кат. 21024). В некоторых случаях 3'-дуплексный адаптер, описанный в настоящем документе, может быть таким, как описано в Примере 1.

3'-адаптеры могут быть присоединены (например, ковалентно) к 3'-концам двухцепочечных ДНК-фрагментов при использовании любого подходящего способа. В некоторых вариантах осуществления 3'-адаптеры присоединяют путем лигирования. В некоторых вариантах осуществления лигирование включает использование лигазы. Примеры лигаз, которые могут использоваться для присоединения 3'-адаптера к 3'-концам каждого фрагмента нуклеиновой кислоты, включают, без ограничения, ДНК-лигазы T4, лигазы *E. coli* (например, фермент Y3), CircLigase I, CircLigase II, Taq-лигазу, лигазу T3, лигазу T7 и лигазу 9N.

После присоединения 3'-дуплексного адаптера (например, ковалентного присоединения) к 3'-концам каждого фрагмента нуклеиновой кислоты второй олигонуклеотид 3'-дуплексного адаптера, описанный в настоящем документе, может быть расщеплен, и 5'-адаптер может быть присоединен (например, ковалентно присоединен) к 5'-концам каждого фрагмента нуклеиновой кислоты. В некоторых вариантах осуществления последовательность 5'-адаптера не комплементарна первому олигонуклеотиду 3'-адаптера. В некоторых вариантах осуществления последовательность 5'-адаптера включает, в направлении 5'→3', сайт праймера для секвенирования R1 и последовательность, комплементарную сайту отжига 3'-адаптера.

В некоторых вариантах осуществления присоединение 5'-адаптера включает отжиг 5'-адаптера с 3'-адаптером посредством сайта отжига.

5'-адаптер может отжигаться с фрагментом нуклеиновой кислоты перед молекулярным штрих-кодом на 3'-дуплексном адаптере, при этом гэп (например, одноцепочечный фрагмент нуклеиновой кислоты), содержащий часть (например, молекулярный штрих-код) 3'-дуплексного адаптера, присутствует на фрагменте нуклеиновой кислоты. Гэп, содержащий часть 3'-дуплексного адаптера, может быть заполнен (например, с получением двухцепочечного фрагмента нуклеиновой кислоты). Для заполнения одноцепочечного гэпа может использоваться любой подходящий способ. Примеры способов, которые могут использоваться для заполнения одноцепочечного гэпа во фрагменте нуклеиновой кислоты, включают, без ограничения, полимеразы, такие как ДНК-полимеразы (например, Taq-полимеразы, такие как Taq-В-полимераза) и реакции ник-трансляции (например, одновременно включающие лигазу, такую как лигаза *E. coli*, и

полимеразу, такую как ДНК-полимераза). В случаях, когда заполнение одноцепочечного гэта во фрагменте нуклеиновой кислоты включает применение полимеразы, способ также может включать применение дезоксирибонуклеотидтрифосфатов (дНТФ, например, дАТФ, дГТФ, дЦТФ и дТТФ). В некоторых случаях присоединение 5'-адаптера к 5'-концу каждого фрагмента нуклеиновой кислоты и заполнение одноцепочечного гэта можно производить одновременно (например, в одной реакционной пробирке).

В некоторых случаях для присоединения адаптеров к матрицам могут использоваться альтернативные методы. Например, фрагменты нуклеиновых кислот могут быть обработаны нуклеазами, специфичными к одноцепочечным нуклеиновым кислотам (например, для расщепления оверхэнгов) с последующим лигированием, которые можно использовать для получения библиотеки дуплексов для секвенирования. Например, один нуклеотид может быть добавлен на 3'-концы каждого фрагмента нуклеиновой кислоты, а адаптеры (например, содержащие молекулярный штрих-код), содержащие комплементарное основание на 5'-конце, можно лигировать с каждым фрагментом нуклеиновой кислоты с получением библиотеки дуплексов для секвенирования, включающей матрицы, присоединенные к адаптерам.

Первичная амплификация матриц, присоединенных к адаптерам

После присоединения адаптера, матрицы с присоединенными к ним адаптерами могут быть амплифицированы (например, амплифицированы с помощью ПЦР) в первичной реакции амплификации. Для амплификации матриц, присоединенных к адаптерам, может использоваться любой подходящий способ. Иллюстративный способ, который может использоваться для амплификации матриц, присоединенных к адаптерам, включает, без ограничения, полногеномную ПЦР.

Любая подходящая пара праймеров может использоваться для амплификации матриц, присоединенных к адаптерам. В некоторых случаях может использоваться пару универсальных праймеров. Праймер может включать, без ограничения, от приблизительно 12 нуклеотидов до приблизительно 30 нуклеотидов. Примеры пар праймеров, которые могут использоваться для амплификации матриц, присоединенных к адаптерам, как описано в настоящем документе, включают, без ограничения, пары праймеров, описанные в Примере 1 и/или Примере 2.

При первичной амплификации могут использоваться любые подходящие условия ПЦР. ПЦР-амплификация может включать фазу денатурации, фазу отжига и фазу элонгации. Каждая фаза цикла амплификации может включать любые подходящие условия. В некоторых случаях фаза денатурации может включать температуру от приблизительно 90°C до приблизительно 105°C (например, от приблизительно 94°C до приблизительно 98°C) и время от приблизительно 1 секунды до приблизительно 5 минут (например, приблизительно от 10 секунд до 1 минуты). Например, фаза денатурации может включать температуру приблизительно 98°C в течение приблизительно 10 секунд. В некоторых случаях фаза отжига может включать температуру от приблизительно 50°C до приблизительно 72°C и время от приблизительно 30 секунд до приблизительно 90

секунд. В некоторых случаях фаза элонгации может включать температуру от приблизительно 55°C до приблизительно 80°C и время от приблизительно 15 секунд на тн синтезируемого ампликона до приблизительно 30 секунд на тн синтезируемого ампликона. В некоторых случаях фазы отжига и элонгации могут выполняться в одном цикле. Например, фаза отжига и фаза элонгации может включать температуру приблизительно 65°C в течение приблизительно 75 секунд.

Условия ПЦР, используемые при первичной амплификации, могут включать любое подходящее количество циклов ПЦР-амплификации. В некоторых случаях ПЦР-амплификация может включать от приблизительно 1 до приблизительно 50 циклов. В некоторых вариантах осуществления ПЦР-амплификация включает не больше 11 циклов. В некоторых вариантах осуществления ПЦР-амплификация включает не больше 7 циклов. В некоторых вариантах осуществления ПЦР-амплификация включает не больше 5 циклов.

В некоторых случаях, когда условия ПЦР включают термоактивируемую полимеразу, ПЦР-амплификация также может включать стадию инициализации. Например, ПЦР-амплификация может включать стадию инициализации перед выполнением циклов ПЦР-амплификации. В некоторых случаях стадия инициализации может включать температуру от приблизительно 94°C до приблизительно 98°C и время от приблизительно 15 секунд до приблизительно 1 минуты. Например, стадия инициализации может включать температуру приблизительно 98°C в течение приблизительно 30 секунд.

В некоторых случаях ПЦР-амплификация также может включать стадию выдерживания. Например, ПЦР-амплификация может включать стадию выдерживания после выполнения циклов ПЦР-амплификации, необязательно после выполнения любой заключительной стадии элонгации. В некоторых случаях стадия выдерживания может включать температуру от приблизительно 4°C до приблизительно 15°C в течение неопределенного периода времени.

В некоторых случаях библиотека дуплексов для секвенирования, созданная, как описано в настоящем документе (например, библиотека амплифицированных дуплексов для секвенирования), может быть очищена. Для очистки библиотеки дуплексов для секвенирования может использоваться любой подходящий способ. Иллюстративный способ, который может использоваться для очистки библиотеки дуплексов для секвенирования, включает, без ограничения, магнитные частицы (например, магнитные частицы с обратимой твердофазной иммобилизацией (SPRI)).

Необязательное получение библиотеки оцДНК

В некоторых случаях библиотека дуплексов для секвенирования может использоваться для создания библиотеки одноцепочечных последовательностей, полученных с цепи Уотсона, и библиотеки одноцепочечных последовательностей, полученных с цепи Крика. Создание библиотеки одноцепочечных последовательностей, полученных с цепи Уотсона, и библиотеки одноцепочечных последовательностей, полученных с цепи Крика, может свести к минимуму неспецифическую амплификацию

(например, с праймера, комплементарного лигированной последовательности, такой как 3'-дуплексный адаптер или 5'-адаптер). Любой подходящий способ может использоваться для создания библиотеки одноцепочечных последовательностей, полученных с цепи Уотсона, и библиотеки одноцепочечных последовательностей, полученных с цепи Крика (например, из библиотеки дуплексов для секвенирования, созданной, как описано в настоящем документе). В некоторых случаях библиотека одноцепочечных последовательностей, полученных с цепи Уотсона, и библиотека одноцепочечных последовательностей, полученных с цепи Крика, могут быть получены из библиотеки амплифицированных дуплексов для секвенирования при разделении продуктов амплификации по меньшей мере на две аликвоты и использования каждой аликвоты для ПЦР-амплификации, при которой цепь Уотсона амплифицируется с первой аликвоты, а цепь Крика амплифицируется со второй аликвоты. Например, первую аликвоту продуктов амплификации из библиотеки амплифицированных дуплексов для секвенирования можно подвергнуть ПЦР-амплификации с использованием пары праймеров, где первый праймер биотинилирован, а второй праймер не биотинилирован, с получением одноцепочечной библиотеки цепей Уотсона, а вторую аликвоту продуктов амплификации из библиотеки амплифицированных дуплексов для секвенирования можно подвергнуть ПЦР-амплификации с использованием пары праймеров, где первый праймер не биотинилирован, а второй праймер биотинилирован, с получением одноцепочечной библиотеки цепей Крика. В некоторых случаях можно создать библиотеку одноцепочечных последовательностей, полученных с цепи Уотсона, и библиотеку одноцепочечных последовательностей, полученных с цепи Крика, как показано на Фигуре 2 и Фигуре 3.

Любой подходящий способ может использоваться для создания библиотеки одноцепочечных последовательностей, полученных с цепи Уотсона, и библиотеки одноцепочечных последовательностей, полученных с цепи Крика, из библиотеки амплифицированных дуплексов для секвенирования. Например, продукты амплификации из библиотеки амплифицированных дуплексов для секвенирования можно разделить на первую ПЦР-амплификацию и вторую ПЦР-амплификацию, в которых только один из двух праймеров в паре ПЦР-праймеров содержит метку. Например, в первой ПЦР-амплификации может использоваться пара праймеров, которая включает праймер (например, первый праймер), который помечен, и праймер (например, второй праймер), который не помечен, а во второй ПЦР-амплификации может использоваться пара праймеров, которая включает праймер (например, первый праймер), который не помечен, и праймер (например, второй праймер), который помечен. Метка праймера может быть любой меткой, которая позволяет выделять продукт ПЦР-амплификации, полученный с меченого праймера. В некоторых случаях меченый праймер может быть биотинилированным праймером, а продукт ПЦР-амплификации, полученный с биотинилированного праймера, может быть выделен с использованием стрептавидина. Например, библиотека одноцепочечных последовательностей, полученных с цепи

Уотсона, и библиотека одноцепочечных последовательностей, полученных с цепи Крика, могут быть созданы при ПЦР-амплификации с использованием пары праймеров, включающей биотинилированный праймер и небитинилированный праймер. В некоторых случаях меченый праймер может быть фосфорилированным праймером, а продукт ПЦР-амплификации, полученный с фосфорилированного праймера, может быть получен при использовании нуклеазы фага лямбда. Например, библиотека одноцепочечных последовательностей, полученных с цепи Уотсона, и библиотека одноцепочечных последовательностей, полученных с цепи Крика, могут быть созданы при ПЦР-амплификации с использованием пары праймеров, включающей фосфорилированный праймер и нефосфорилированный праймер.

Любая подходящая пара праймеров может использоваться для создания библиотеки одноцепочечных последовательностей, полученных с цепи Уотсона, и библиотеки одноцепочечных последовательностей, полученных с цепи Крика (например, из библиотеки дуплексов для секвенирования, созданной, как описано в настоящем документе). Праймер может включать, без ограничения, от приблизительно 12 нуклеотидов до приблизительно 30 нуклеотидов. В некоторых случаях пара праймеров может включать по меньшей мере один праймер, который может направленно взаимодействовать (например, направленно взаимодействовать и связываться) с адаптерной последовательностью (например, адаптерной последовательностью, содержащей молекулярный штрих-код), присутствующей в продукте амплификации, полученном, как описано в настоящем документе (например, при лигировании 3'-дуплексного адаптера, включающего первый молекулярный штрих-код, и 5'-адаптера, включающего второй молекулярный штрих-код, с фрагментом нуклеиновой кислоты в библиотеке дуплексов для секвенирования до амплификации). Примеры пар праймеров, которые могут использоваться для создания библиотеки одноцепочечных последовательностей, полученных с цепи Уотсона, и библиотеки одноцепочечных последовательностей, полученных с цепи Крика, как описано в настоящем документе, включают, без ограничения, праймер P5 и праймер P7.

Любые подходящие условия ПЦР могут быть использованы для создания библиотеки одноцепочечных последовательностей, происходящих с цепи Уотсона, и библиотеки одноцепочечных последовательностей, происходящих с цепи Крика (например, из библиотеки дуплексов для секвенирования, созданной, как описано в настоящем документе). Амплификация ПЦР может включать фазу денатурации, фазу отжига и фазу удлинения. Каждая фаза цикла амплификации может включать любые подходящие условия. В некоторых случаях фаза денатурации может включать температуру от приблизительно 90°C до приблизительно 105°C и время от приблизительно 1 секунды до приблизительно 5 минут. Например, фаза денатурации может включать температуру приблизительно 98°C в течение приблизительно 10 секунд. В некоторых случаях фаза отжига может включать температуру от приблизительно 50°C до приблизительно 72°C и время от приблизительно 30 секунд до приблизительно 90

секунд. В некоторых случаях фаза элонгации может включать температуру от приблизительно 55°C до приблизительно 80°C и время от приблизительно 15 секунд на тн синтезируемого ампликона до приблизительно 30 секунд на тн синтезируемого ампликона. В некоторых случаях фаза элонгации отражает процессивность используемой полимеразы. В некоторых случаях фазы отжига и элонгации могут выполняться в одном цикле. Например, фаза отжига и фаза элонгации могут включать температуру приблизительно 65°C в течение приблизительно 75 секунд.

Условия ПЦР, используемые для создания библиотеки одноцепочечных последовательностей, полученных с цепи Уотсона, и библиотеки одноцепочечных последовательностей, полученных с цепи Крика (например, из библиотеки дуплексов для секвенирования, созданной, как описано в настоящем документе), могут включать любое подходящее количество циклов ПЦР-амплификации. В некоторых случаях ПЦР-амплификация может включать, без ограничения, от приблизительно 1 до приблизительно 50 циклов. Например, ПЦР-амплификация может включать приблизительно 4 цикла амплификации.

В некоторых случаях, когда условия ПЦР включают термоактивируемую полимеразу, ПЦР-амплификация также может включать стадию инициализации. Например, ПЦР-амплификация может включать стадию инициализации перед выполнением циклов ПЦР-амплификации. В некоторых случаях стадия инициализации может включать температуру от приблизительно 94°C до приблизительно 98°C и время от приблизительно 15 секунд до приблизительно 1 минуты. Например, стадия инициализации может включать температуру приблизительно 98°C в течение приблизительно 30 секунд.

В некоторых случаях ПЦР-амплификация также может включать стадию выдерживания. Например, ПЦР-амплификация может включать стадию выдерживания после выполнения циклов ПЦР-амплификации, необязательно после выполнения любого заключительной стадии элонгации. В некоторых случаях стадия выдерживания может включать температуру от приблизительно 4°C до приблизительно 15°C в течение неопределенного периода времени.

Любой подходящий способ может использоваться для разделения двухцепочечных продуктов амплификации на одноцепочечные продукты амплификации. В некоторых случаях двухцепочечные продукты амплификации можно денатурировать с разделением двухцепочечных продуктов амплификации на два одноцепочечных продукта амплификации. Примеры способов, которые могут использоваться для разделения двухцепочечного продукта амплификации на одноцепочечные продукты амплификации, включают, без ограничения, термическую денатурацию, химическую (например, NaOH) денатурацию и солевую денатурацию.

После ПЦР-амплификации меченые продукты ПЦР-амплификации могут быть выделены. Любой подходящий способ может использоваться для выделения меченых продуктов ПЦР-амплификации, полученных при использовании меченого праймера. В

случаях, когда меченый праймер является биотинилированным праймером, биотинилированные продукты амплификации (например, полученные с биотинилированного праймера) можно выделять с использованием стрептавидина (например, стрептавидин-функционализованных сфер). Например, когда библиотеку амплифицированных дуплексов для секвенирования дополнительно амплифицируют в первой ПЦР-амплификации при использовании пары праймеров, которая включает первый биотинилированный праймер и второй небитинилированный праймер, и во второй ПЦР-амплификации при использовании пары праймеров, которая включает первый небитинилированный праймер и второй биотинилированный праймер, биотинилированные продукты амплификации, полученные в результате первой ПЦР-амплификации, могут быть связаны со стрептавидин-функционализованными сферами (например, первым набором стрептавидин-функционализованных сфер), а биотинилированные продукты амплификации, полученные в результате второй ПЦР-амплификации, могут быть связаны со стрептавидин-функционализованными сферами (например, вторым набором стрептавидин-функционализованных сфер), при этом двухцепочечные продукты амплификации могут быть разделены (например, денатурированы) на одноцепочечные продукты амплификации. В некоторых случаях выделение биотинилированных продуктов ПЦР-амплификации также может включать высвобождение биотинилированных продуктов ПЦР-амплификации со стрептавидина (например, стрептавидин-функционализованных сфер). Разделение двухцепочечных продуктов амплификации, полученных в результате первой ПЦР-амплификации при использовании пары праймеров, включающей первый биотинилированный праймер и второй небитинилированный праймер, и второй ПЦР-амплификации при использовании пары праймеров, включающей первый небитинилированный праймер и второй биотинилированный праймер, могут позволять одноцепочечным продуктам амплификации, полученным с биотинилированных праймеров, оставаться связанными со стрептавидин-функционализованными сферами, тогда как одноцепочечные продукты амплификации, полученные с небитинилированных праймеров, могут быть денатурированы (например, денатурированы и расщеплены) со стрептавидин-функционализованных сфер, с получением в результате библиотеки одноцепочечных последовательностей, полученных с цепи Уотсона, и библиотеки одноцепочечных последовательностей, полученных с цепи Крика, из библиотеки дуплексов для секвенирования.

В случаях, когда меченый праймер является фосфорилированным праймером, фосфорилированные продукты амплификации (например, полученные с фосфорилированного праймера) могут быть выделены с использованием экзонуклеазы (например, экзонуклеазы фага лямбда). Например, когда библиотеку амплифицированных дуплексов для секвенирования дополнительно амплифицируют в первой ПЦР-амплификации при использовании пары праймеров, которая включает первый фосфорилированный праймер и второй нефосфорилированный праймер, и во второй ПЦР-

амплификации при использовании пары праймеров, которая включает первый нефосфорилированный праймер и второй фосфорилированный праймер, двухцепочечные продукты амплификации можно разделить на одноцепочечные продукты амплификации. Разделение двухцепочечных продуктов амплификации, полученных в результате первой ПЦР-амплификации при использовании пары праймеров, включающей первый фосфорилированный праймер и второй нефосфорилированный праймер, и второй ПЦР-амплификации при использовании пары праймеров, включающей первый нефосфорилированный праймер и второй фосфорилированный праймер, может позволить выделить одноцепочечные продукты амплификации, полученные с нефосфорилированных праймеров, тогда как одноцепочечные продукты амплификации, полученные с фосфорилированных праймеров, могут быть расщеплены экзонуклеазой фага лямбда, с получением в результате библиотеки одноцепочечных последовательностей, полученных с цепи Уотсона, и библиотеки одноцепочечных последовательностей, полученных с цепи Крика, из библиотеки дуплексов для секвенирования.

Обогащение мишенями

В некоторых вариантах осуществления любого из способов в настоящем документе, ампликоны, полученные в результате первичной амплификации, обогащают одним или больше полинуклеотидами-мишенями. В некоторых вариантах осуществления, перед обогащением мишенями, библиотеки одноцепочечных ДНК получают с ампликонов, полученных в результате первичной амплификации. Иллюстративные способы получения библиотек одноцепочечных ДНК описаны в настоящем документе.

Любой подходящий способ может использоваться для амплификации области-мишени из библиотеки продуктов амплификации (например, библиотеки дуплексов для секвенирования, библиотеки одноцепочечных последовательностей, полученных с цепи Уотсона, или библиотеки одноцепочечных последовательностей, полученных с цепи Крика, созданных, как описано в настоящем документе). В некоторых случаях область-мишень может быть амплифицирована с библиотеки продуктов амплификации при ПЦР-амплификации библиотеки продуктов амплификации с использованием пары праймеров, где некий праймер (например, первый праймер), который может направленно взаимодействовать (например, направленно взаимодействовать и связываться) с последовательностью адаптера (например, последовательностью адаптера, содержащей молекулярный штрих-код), присутствующей в продукте амплификации, полученном, как описано в настоящем документе (например, при лигировании 3'-дуплексного адаптера, включающего первый молекулярный штрих-код, и 5'-адаптера, включающего второй молекулярный штрих-код, с фрагментом нуклеиновой кислоты в библиотеке дуплексов для секвенирования перед амплификацией) и некий праймер (например, второй праймер), который может направленно взаимодействовать (например, направленно взаимодействовать и связываться) с областью-мишенью (например, областью, представляющей интерес). В некоторых случаях библиотека одноцепочечных

последовательностей, полученных с цепи Уотсона, и библиотека одноцепочечных последовательностей, полученных с цепи Крика, могут быть созданы, как показано на Фигуре 4 и Фигуре 5. В некоторых случаях библиотека одноцепочечных последовательностей, полученных с цепи Уотсона, и библиотека одноцепочечных последовательностей, полученных с цепи Крика, могут быть созданы, как описано в Примере 2.

В некоторых случаях область-мишень может быть амплифицирована с библиотеки продуктов амплификации (например, библиотеки дуплексов для секвенирования, библиотеки одноцепочечных последовательностей, полученных с цепи Уотсона, или библиотеки одноцепочечных последовательностей, полученных с цепи Крика, созданных, как описано в настоящем документе) в одной ПЦР-амплификации. Например, область-мишень может быть амплифицирована с библиотеки продуктов амплификации в ходе одной ПЦР-амплификации при использовании пары праймеров, включающей первый праймер, который может направленно взаимодействовать с адаптерной последовательностью (например, адаптерной последовательностью, содержащей молекулярный штрих-код), присутствующей в продукте амплификации, полученном, как описано в настоящем документе (например, при лигировании 3'-дуплексного адаптера, включающего первый молекулярный штрих-код, и 5'-адаптера, включающего второй молекулярный штрих-код, с фрагментом нуклеиновой кислоты в библиотеке дуплексов для секвенирования перед амплификацией), и второй праймер, который может направленно взаимодействовать с областью-мишенью. Например, область-мишень может быть амплифицирована с библиотеки продуктов амплификации в ходе одной ПЦР-амплификации, как показано на Фигуре 4, Фигуре 5, Фигуре 15 и Фигуре 17.

В некоторых случаях область-мишень может быть амплифицирована с библиотеки продуктов амплификации (например, библиотеки дуплексов для секвенирования, библиотеки одноцепочечных последовательностей, полученных с цепи Уотсона, или библиотеки одноцепочечных последовательностей, полученных с цепи Крика, созданных, как описано в настоящем документе) во множестве ПЦР-амплификаций. Множество ПЦР-амплификаций (например, первую ПЦР-амплификацию и последующую гнездовую ПЦР-амплификацию) можно использовать для повышения специфичности амплификации области-мишени. Например, область-мишень может быть амплифицирована с библиотеки продуктов амплификации в серии ПЦР-амплификаций, где в первой ПЦР-амплификации используется пара праймеров, включающая первый праймер, который может направленно взаимодействовать с адаптерной последовательностью (например, адаптерной последовательностью, содержащей молекулярный штрих-код), присутствующей в продукте амплификации, полученном, как описано в настоящем документе (например, при лигировании 3'-дуплексного адаптера, включающего первый молекулярный штрих-код, и 5'-адаптера, включающего второй молекулярный штрих-код, с фрагментом нуклеиновой кислоты в библиотеке дуплексов для секвенирования перед амплификацией), и второй праймер, который может направленно взаимодействовать с областью-мишенью,

и использовании продуктов амплификации, полученных в первой ПЦР-амплификации, в последующей гнездовой ПЦР-амплификации, в которой используется пара праймеров, включающая первый праймер, который может направленно взаимодействовать с последовательностью адаптера (например, последовательностью адаптера, содержащей молекулярный штрих-код), присутствующей в продукте амплификации, полученном, как описано в настоящем документе (например, при лигировании 3'-дуплексного адаптера включающего первый молекулярный штрих-код, и 5'-адаптера, включающего второй молекулярный штрих-код, с фрагментом нуклеиновой кислоты в библиотеке дуплексов для секвенирования перед амплификацией), и второй праймер, который может направленно взаимодействовать с последовательностью нуклеиновой кислоты из области-мишени, которая присутствует в продуктах амплификации, полученных в ходе первой ПЦР-амплификации. Например, область-мишень может быть амплифицирована с библиотеки продуктов амплификации в серии ПЦР-амплификаций, как показано на Фигуре 7, Фигуре 8, Фигуре 16 и Фигуре 18.

Любая подходящая пара праймеров может использоваться для амплификации области-мишени с библиотеки продуктов амплификации (например, библиотеки дуплексов для секвенирования, библиотеки одноцепочечных последовательностей, полученных с цепи Уотсона, или библиотеки одноцепочечных последовательностей, полученных с цепи Крика, созданных как описано в настоящем документе). Праймер может включать, без ограничения, от приблизительно 12 нуклеотидов до приблизительно 30 нуклеотидов. В некоторых случаях пара праймеров может включать праймер (например, первый праймер), который может направленно взаимодействовать (например, направленно взаимодействовать и связываться с) последовательностью адаптера (например, последовательностью адаптера, содержащей молекулярный штрих-код), присутствующей в продукте амплификации, полученном, как описано в настоящем документе (например, при лигировании 3'-дуплексного адаптера, включающего первый молекулярный штрих-код, и 5'-адаптера, включающего второй молекулярный штрих-код, с фрагментом нуклеиновой кислоты в библиотеке дуплексов для секвенирования перед амплификацией), и праймер (например, второй праймер), который может направленно взаимодействовать (например, направленно взаимодействовать и связываться) с областью-мишенью (например, областью, представляющей интерес). Примеры праймеров, которые могут направленно взаимодействовать с последовательностью адаптера, содержащей молекулярный штрих-код, присутствующей в продукте амплификации, полученном, как описано в настоящем документе (например, при лигировании 3'-дуплексного адаптера, включающего первый молекулярный штрих-код, и 5'-адаптера, включающего второй молекулярный штрих-код, с фрагментом нуклеиновой кислоты в библиотеке дуплексов для секвенирования перед амплификацией), включают, без ограничения, праймер с индексом i5 и праймер с индексом i7. Праймеры, которые могут направленно взаимодействовать с областью-мишенью, могут включать последовательность, комплементарную области-мишени. В случаях, когда область-

мишень представляет собой нуклеиновую кислоту, кодирующую TP53, примеры праймеров, которые могут направленно взаимодействовать с нуклеиновой кислотой, кодирующей TP53, включают, без ограничения, TP53_342_GSP1 и TP53_GSP2. В некоторых случаях, когда область-мишень является нуклеиновой кислотой, кодирующей TP53, праймер, направленный на нуклеиновую кислоту, кодирующую TP53, может быть таким, как описано в Примере 2.

В некоторых случаях один или оба праймера в паре праймеров, используемых для амплификации области-мишени с библиотеки продуктов амплификации (например, библиотеки дуплексов для секвенирования, библиотеки одноцепочечных последовательностей, полученных с цепи Уотсона, или библиотеки одноцепочечных последовательностей, полученных с цепи Крика, созданных, как описано в настоящем документе), могут включать один или больше молекулярных штрих-кодов.

В некоторых случаях один или оба праймера в паре праймеров, используемых для амплификации области-мишени с библиотеки продуктов амплификации (например, библиотеки дуплексов для секвенирования, библиотеки одноцепочечных последовательностей, полученных с цепи Уотсона, или библиотеки одноцепочечных последовательностей, полученных с цепи Крика, созданных, как описано в настоящем документе), могут включать одну или больше последовательностей присоединения (например, последовательностей присоединения для секвенирования следующего поколения).

В одном аспекте обогащение мишенями включает: (a) селективную амплификацию ампликонов цепей Уотсона, включающих полинуклеотидную последовательность-мишень, при использовании первого набора пар мишень-селективных праймеров Уотсона, причем первый набор пар мишень-селективных праймеров Уотсона включает: (i) первый мишень-селективный праймер Уотсона, включающий последовательность, комплементарную сайту праймера для секвенирования R2 универсальной 3'-адаптерной последовательности, и (ii) второй мишень-селективный праймер Уотсона, включающий мишень-селективную последовательность, с получением в результате целевых продуктов амплификации Уотсона; и (b) селективную амплификацию ампликонов цепей Крика, включающих такую же полинуклеотидную последовательность-мишень, при использовании первого набора пар мишень-селективных праймеров Крика, причем первый набор пар мишень-селективных праймеров Крика включает: (i) первый мишень-селективный праймер Крика, включающий последовательность, комплементарную сайту праймера для секвенирования R1 универсальной 5'-адаптерной последовательности, и (ii) второй мишень-селективный праймер Крика, включающий такую же последовательность мишень-селективного праймера, как и последовательность второго мишень-селективного праймера Уотсона, с получением в результате целевых продуктов амплификации Крика.

В некоторых вариантах осуществления способ дополнительно включает очистку целевых продуктов амплификации Уотсона и целевых продуктов амплификации Крика от нецелевых полинуклеотидов. В некоторых вариантах осуществления очистка включает

прикрепление целевых продуктов амплификации Уотсона и целевых продуктов амплификации Крика к твердой подложке. В некоторых вариантах осуществления первый мишень-селективный праймер Уотсона и первый мишень-селективный праймер Крика включают первый член пары аффинного связывания, и где твердая подложка включает второй член пары аффинного связывания. В некоторых вариантах осуществления первый член представляет собой биотин, а второй член представляет собой стрептавидин. В некоторых вариантах осуществления твердая подложка включает сферу, лунку, мембрану, пробирку, колонку, планшет, сефарозу, магнитную сферу или чип. В некоторых вариантах осуществления способ включает удаление полинуклеотидов, не прикрепившихся к твердой подложке.

В некоторых вариантах осуществления способ дополнительно включает: (a) дополнительную амплификацию целевых продуктов амплификации Уотсона с использованием второго набора мишень-селективных праймеров Уотсона, причем второй набор мишень-селективных праймеров Уотсона включает: (i) третий мишень-селективный праймер Уотсона, включающий последовательность, комплементарную сайту праймера для секвенирования R2 универсальной 3'-адаптерной последовательности, и (ii) четвертый мишень-селективный праймер Уотсона, включающий, в направлении 5'→3': сайт праймера для секвенирования R1 и мишень-селективную последовательность, селективную в отношении такого же полинуклеотида-мишени, с получением в результате целевых членов библиотеки Уотсона; (b) дополнительную амплификацию целевых продуктов амплификации Крика с использованием второго набора мишень-селективных праймеров Крика, причем второй набор мишень-селективных праймеров Крика включает: (i) третий мишень-селективный праймер Крика, включающий последовательность, комплементарную сайту праймера для секвенирования R1 универсальной 3'-адаптерной последовательности, и (ii) четвертый мишень-селективный праймер Крика, включающий, в направлении 5'→3': сайт праймера для секвенирования R2 и мишень-селективную последовательность, селективную в отношении такого же полинуклеотида-мишени четвертого мишень-селективного праймера Уотсона, с получением в результате целевых членов библиотеки Крика.

В некоторых вариантах осуществления третьи мишень-селективные праймеры Уотсона и Крика дополнительно включают последовательность штрих-кода образца. В некоторых вариантах осуществления третий мишень-селективный праймер Уотсона дополнительно включает первую последовательность присоединения, которая обеспечивает гибридизацию с первым праймером присоединения на секвенаторе, и где третий мишень-селективный праймер Крика дополнительно включает вторую последовательность присоединения, которая обеспечивает гибридизацию со вторым праймером присоединения на секвенаторе. В некоторых вариантах осуществления четвертый мишень-селективный праймер Уотсона дополнительно включает вторую последовательность присоединения, и четвертый мишень-селективный праймер Крика дополнительно включает первую последовательность присоединения. В некоторых

вариантах осуществления первая последовательность присоединения представляет собой последовательность P7, а вторая последовательность присоединения представляет собой последовательность P5.

Любые подходящие условия ПЦР могут использоваться для получения амплифицированной области-мишени, как описано в настоящем документе (например, с библиотеки продуктов амплификации, такой как библиотека дуплексов для секвенирования, библиотека одноцепочечных последовательностей, полученных с цепи Уотсона, или библиотека одноцепочечных последовательностей, полученных с цепи Крика). Типичные условия ПЦР описаны в настоящем документе. Условия ПЦР, используемые для получения амплифицированной области-мишени, как описано в настоящем документе (например, с библиотеки продуктов амплификации, такой как библиотека дуплексов для секвенирования, библиотека одноцепочечных последовательностей, полученных с цепи Уотсона, или библиотека одноцепочечных последовательностей, полученных с цепи Крика), могут включать любое подходящее количество циклов ПЦР-амплификации. В некоторых случаях ПЦР-амплификация может включать, без ограничения, от приблизительно 1 до приблизительно 50 циклов. Например, когда ПЦР-амплификация амплифицированной области-мишени включает одну ПЦР-амплификацию, ПЦР-амплификация может включать приблизительно 18 циклов амплификации. Например, когда ПЦР-амплификация амплифицированной области-мишени включает первую ПЦР-амплификацию и последующую гнездовую ПЦР-амплификацию, первая ПЦР-амплификация может включать приблизительно 18 циклов амплификации, а последующая гнездовая ПЦР-амплификация может включать приблизительно 10 циклов амплификации.

Примеры мишеней

Любая подходящая область-мишень (например, область, представляющая интерес) может быть амплифицирована с библиотеки продуктов амплификации (например, библиотеки дуплексов для секвенирования, библиотеки одноцепочечных последовательностей, полученных с цепи Уотсона, или библиотеки одноцепочечных последовательностей, полученных с цепи Крика, созданных, как описано в настоящем документе), и оценена на присутствие или отсутствие одной или более мутаций. В некоторых случаях область-мишень может представлять собой область нуклеиновой кислоты, в которой одна или более мутаций ассоциированы с заболеванием или нарушением. Примеры областей-мишеней, которые могут быть амплифицированы и оценены на предмет присутствия или отсутствия одной или более мутаций, включают, помимо прочего, нуклеиновую кислоту, кодирующую опухолевый белок p53 (TP53), нуклеиновую кислоту, кодирующую белок рака молочной железы 1 (BRCA1), нуклеиновую кислоту, кодирующую BRCA2, нуклеиновую кислоту, кодирующую полипептид гомолога фосфатазы и тензина (PTEN), нуклеиновую кислоту, кодирующую полипептид АКТ1, нуклеиновую кислоту, кодирующую полипептид APC, нуклеиновую кислоту, кодирующую полипептид CDKN2A, нуклеиновую кислоту, кодирующую

полипептид EGFR, нуклеиновую кислоту, кодирующую полипептид FBXW7, нуклеиновую кислоту, кодирующую полипептид GNAS, нуклеиновую кислоту, кодирующую полипептид KRAS, нуклеиновую кислоту, кодирующую полипептид NRAS, нуклеиновую кислоту, кодирующую полипептид PIK3CA, нуклеиновую кислоту, кодирующую полипептид BRAF, нуклеиновую кислоту, кодирующую полипептид CTNNA1, нуклеиновую кислоту, кодирующую полипептид FGFR2, нуклеиновую кислоту, кодирующую полипептид HRAS, и нуклеиновую кислоту, кодирующую полипептид PPP2R1A. В некоторых случаях область-мишень, которую могут амплифицировать и оценивать на предмет присутствия или отсутствия одной или более мутаций, может представлять собой нуклеиновую кислоту, кодирующую TP53. Например, нуклеиновая кислота, кодирующая TP53, может быть амплифицирована и оценена, как описано в Примере 2.

Любой подходящий способ может использоваться для оценки области-мишени (например, амплифицированной области-мишени) на предмет присутствия или отсутствия одной или более мутаций. В некоторых случаях один или больше методов секвенирования может использоваться для оценки амплифицированной области-мишени на предмет присутствия или отсутствия одной или более мутаций.

Определение последовательности

В некоторых случаях один или больше методов секвенирования может использоваться для оценки амплифицированной области-мишени с целью определения, присутствует ли мутация(и) одновременно в цепи Уотсона и в цепи Крика. В некоторых случаях чтения секвенирования могут использоваться для оценки амплифицированной области-мишени на присутствие или отсутствие одной или более мутаций и могут использоваться для определения, присутствует ли мутация(и) одновременно в цепи Уотсона и в цепи Крика. Примеры методов секвенирования, которые могут использоваться для оценки амплифицированной области-мишени на предмет присутствия или отсутствия одной или более мутаций, как описано в настоящем документе, включают, без ограничения, секвенирование с одним чтением, секвенирование спаренных концов, NGS и глубокое секвенирование. В некоторых вариантах осуществления секвенирование с одним чтением включает секвенирование по всей длине матриц с получением чтений последовательности. В некоторых вариантах осуществления секвенирование включает секвенирование спаренных концов. В некоторых вариантах осуществления секвенирование проводят с использованием массово-параллельного секвенатора. В некоторых вариантах осуществления массивно-параллельный секвенатор обладает возможностью определения чтений последовательности с обоих концов полинуклеотидов-матриц.

Анализ чтений последовательности

В некоторых вариантах осуществления чтения последовательности картируют на референсном геноме.

В некоторых вариантах осуществления чтения последовательности относят к

семействам UID. Семейство UID может включать чтения последовательностей с ампликонов, происходящих от исходной матрицы, например, от исходного фрагмента двухцепочечной ДНК из образца нуклеиновой кислоты.

В некоторых вариантах осуществления каждый член семейства UID включает одну и ту же экзогенную UID последовательность. В некоторых вариантах осуществления каждый член семейства UID дополнительно включает одну и ту же эндогенную UID последовательность. Эндогенные UID описаны в настоящем документе.

В некоторых вариантах осуществления каждый член семейства UID дополнительно включает одну и ту же экзогенную UID последовательность и одну и ту же эндогенную UID последовательность. В некоторых вариантах осуществления комбинация экзогенной UID последовательности и эндогенной UID последовательности является уникальной для семейства UID. В некоторых вариантах осуществления комбинация экзогенной UID последовательности и эндогенной UID последовательности не существует в другом семействе UID, представленном в образце нуклеиновой кислоты.

Количество членов семейства UID может зависеть от глубины секвенирования. В некоторых вариантах осуществления семейство UID включает по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500 или 1000 членов. В некоторых вариантах осуществления семейство UID включает приблизительно 2-1000 членов, приблизительно 2-500 членов, приблизительно 2-100 членов, приблизительно 2-50 членов или приблизительно 2-20 членов.

В некоторых вариантах осуществления чтения последовательности отдельного семейства UID относят к подсемейству Уотсона и подсемейству Крика. В некоторых вариантах осуществления чтения последовательности отдельного семейства UID относят к подсемействам Уотсона и Крика на основе ориентации вставки относительно адаптерных последовательностей. В некоторых вариантах осуществления ориентация вставки относительно адаптерных последовательностей определяется тем, как чтения последовательностей были выровнены как "парные чтения" или "спаренные чтения".

В некоторых вариантах осуществления отнесение чтений последовательностей к подсемействам Уотсона и Крика основано на пространственном отношении экзогенной UID последовательности к последовательностям чтений R1 и R2. В некоторых вариантах осуществления члены подсемейства Уотсона характеризуются экзогенной UID последовательностью, расположенной после последовательности R2 и перед последовательностью R1. В некоторых вариантах осуществления члены подсемейства Крика характеризуются экзогенной UID последовательностью, расположенной после последовательности R1 и перед последовательностью R2. В некоторых вариантах осуществления члены подсемейства Уотсона характеризуются экзогенной последовательностью UID, которая расположена в большей близости от последовательности R2 и в меньшей близости от последовательности R1. В некоторых

вариантах осуществления члены подсемейства Крика характеризуются экзогенной последовательностью UID, которая расположена в большей близости от последовательности R1 и в меньшей близости от последовательности R2. В некоторых вариантах осуществления члены подсемейства Уотсона характеризуются экзогенной последовательностью UID, расположенной непосредственно после или в пределах 1-70, 1-60, 1-50, 1-40, 1-30, 1-20, 1-10 или 1-5 нуклеотидов от последовательности R2. В некоторых вариантах осуществления члены подсемейства Крика характеризуются экзогенной последовательностью UID, расположенной непосредственно после или в пределах 1-70, 1-60, 1-50, 1-40, 1-30, 1-20, 1-10 или 1-5 нуклеотидов от последовательности R1.

В некоторых вариантах осуществления подсемейство UID (например, подсемейство Уотсона и/или подсемейство Крика) включает по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490 или 500 членов. В некоторых вариантах осуществления подсемейство UID (например, подсемейство Уотсона и/или подсемейство Крика) включает приблизительно 2-500 членов, приблизительно 2-100 членов, приблизительно 2-50 членов, приблизительно 2-20 членов или приблизительно 2-10 членов.

В некоторых вариантах осуществления определяют, что нуклеотидная последовательность точно представляет цепь Уотсона анализируемого фрагмента ДНК, например, фрагмента двухцепочечной ДНК из образца нуклеиновой кислоты, если пороговый процент (или процент, превышающий пороговое значение) членов подсемейства Уотсона содержит такую последовательность. В некоторых вариантах осуществления определяют, что нуклеотидная последовательность точно представляет цепь Крика анализируемого фрагмента ДНК, например, фрагмента двухцепочечной ДНК из образца нуклеиновой кислоты, если пороговый процент (или процент, превышающий порог) членов подсемейства Крика содержит такую последовательность.

Пороговые значения могут быть определены квалифицированным специалистом на основе, например, количества членов подсемейства, конкретной задачи эксперимента по секвенированию и конкретных параметров эксперимента по секвенированию. В некоторых вариантах осуществления пороговое значение устанавливают на уровне 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100%. В конкретных вариантах осуществления пороговое значение устанавливают на уровне 50%. Исключительно в качестве примера, в варианте осуществления, в котором пороговое значение установлено на 50%, определяют, что нуклеотидная последовательность точно представляет цепь Уотсона или Крика анализируемого фрагмента ДНК, например, фрагмента двухцепочечной ДНК из образца нуклеиновой кислоты, если по меньшей мере 50% членов подсемейства содержат такую последовательность. Только в качестве другого примера, в варианте осуществления, в

котором пороговое значение установлено на уровне 50%, определяют, что нуклеотидная последовательность точно представляет цепь Уотсона или Крика анализируемого фрагмента ДНК, например, фрагмента двухцепочечной ДНК из образца нуклеиновой кислоты, если больше 50% членов подсемейства содержат такую последовательность.

В некоторых вариантах осуществления определяют, что последовательность, точно представляющая цепь Уотсона анализируемого фрагмента ДНК, имеет мутацию. В некоторых вариантах осуществления определяют, что последовательность, точно представляющая цепь Уотсона анализируемого фрагмента ДНК, имеет мутацию, если последовательность отличается от референсной последовательности, в которой такая мутация отсутствует.

В некоторых вариантах осуществления определяют, что последовательность, точно представляющая цепь Крика анализируемого фрагмента ДНК, имеет мутацию. В некоторых вариантах осуществления определяют, что последовательность, точно представляющая цепь Крика анализируемого фрагмента ДНК, имеет мутацию, если последовательность отличается от референсной последовательности, в которой такая мутация отсутствует.

В некоторых вариантах осуществления определяют, что анализируемый фрагмент ДНК имеет мутацию, если последовательность, точно представляющая цепь Уотсона, и последовательность, точно представляющая цепь Крика, содержит одинаковую мутацию.

В некоторых случаях расположение молекулярного штрих-кода в чтениях секвенирования спаренных концов амплифицированной области-мишени можно использовать для определения того, с какой цепи матрицы двухцепочечной нуклеиновой кислоты была получена амплифицированная область-мишень. Например, если первое чтение секвенирования спаренных концов амплифицированной области-мишени указывает, что молекулярный штрих-код прочитывается последним, можно определить, что амплифицированная область-мишень получена со смысловой цепи матрицы нуклеиновой кислоты, а если первое чтение секвенирования спаренных концов амплифицированной области-мишени указывает, что молекулярный штрих-код считывается первым, можно определить, что амплифицированная область-мишень получена с антисмысловой цепи матрицы нуклеиновой кислоты. Например, если второе чтение секвенирования спаренных концов амплифицированной области-мишени указывает, что молекулярный штрих-код считывается первым, можно определить, что амплифицированная область-мишень получена с антисмысловой цепи матрицы нуклеиновой кислоты, а если второе чтение секвенирования спаренных концов амплифицированной области-мишени указывает, что молекулярный штрих-код считывается последним, можно определить, что амплифицированная область-мишень получена со смысловой цепи матрицы нуклеиновой кислоты. В некоторых случаях секвенирование спаренных концов может использоваться, чтобы отличать продукты амплификации, полученные с цепи Уотсона, от продуктов амплификации, полученных с цепи Крика, как показано на Фигуре 20 и Фигуре 21.

После секвенирования областей-мишеней (например, областей-мишеней, амплифицированных, как описано в настоящем документе), чтения секвенирования можно выравнивать с референсным геномом и группировать по молекулярному штрих-коду, присутствующему в каждом чтении секвенирования. В некоторых случаях можно определить, что чтения секвенирования, которые включают один и тот же молекулярный штрих-код и картированы как на цепи Уотсона, так и на цепи Крика матричной двухцепочечной нуклеиновой кислоты (например, и цепи Уотсона, и цепи Крика области-мишени), имеют дуплексное подтверждение. Например, если чтения секвенирования указывают на присутствие одной или больше мутаций в области-мишени, включают один и тот же молекулярный штрих-код и картированы как на цепи Уотсона, так и на цепи Крика области-мишени, можно определить, что мутация(и) имеет дуплексное подтверждение.

Наборы

В настоящем документе также предложены наборы. Наборы могут включать наборы пар праймеров для амплификации одного или более полинуклеотидов-мишеней.

В некоторых вариантах осуществления набор включает: (a) первый набор пар мишень-селективных праймеров Уотсона, включающий: (i) один или больше первых мишень-селективных праймеров Уотсона, включающих последовательность, комплементарную сайту праймера для секвенирования R2 универсальной 3'-адаптерной последовательности, и (ii) один или больше вторых мишень-селективных праймеров Уотсона, причем каждый из одного или больше вторых мишень-селективных праймеров Уотсона включает мишень-селективную последовательность; (b) первый набор пар мишень-селективных праймеров Крика, включающий: (i) один или больше мишень-селективных праймеров Крика, включающих последовательность, комплементарную сайту праймера секвенирования R1 универсальной 5'-адаптерной последовательности, и (ii) один или больше вторых мишень-селективных праймеров Крика, причем каждый из одного или больше вторых мишень-селективных праймеров Крика включает такую же мишень-селективную последовательность, как и последовательность второго мишень-селективного праймера Уотсона; (c) второй набор пар мишень-селективных праймеров Уотсона, включающий: (i) один или больше третьих мишень-селективных праймеров Уотсона, включающих последовательность, комплементарную сайту праймера для секвенирования R2 универсальной 3'-адаптерной последовательности, и (ii) один или больше четвертых мишень-селективных праймеров Уотсона, причем каждый из одного или больше четвертых мишень-селективных праймеров Уотсона включает, в направлении 5'→3': сайт праймера для секвенирования R1 и мишень-селективную последовательность, селективную в отношении такого же полинуклеотида-мишени; и (d) второй набор мишень-селективных праймеров Крика, включающий: (i) один или больше третьих мишень-селективных праймеров Крика, включающих последовательность, комплементарную сайту праймера секвенирования R1 универсальной 3'-адаптерной последовательности, и (ii) один или больше четвертых мишень-селективных праймеров

Крика, причем каждый из одного или больше четвертых мишень-селективных праймеров Крика включает, в направлении 5'→3': сайт праймера для секвенирования R2 и мишень-селективную последовательность, селективную в отношении такого же полинуклеотида-мишени.

Набор может включать наборы пар праймеров для мультиплексной амплификации множества полинуклеотидов-мишеней.

Машиночитаемый носитель

Также в настоящем документе предложен машиночитаемый носитель, содержащий машиноисполняемые инструкции, которые обеспечивают реализацию любого из способов, описанных в настоящем документе. Машиночитаемый носитель может содержать машиноисполняемые инструкции для анализа данных последовательности из образца нуклеиновой кислоты, где данные генерируются способом согласно любому из предыдущих пунктов.

Машиночитаемый носитель может обеспечивать реализацию способа полуавтоматизированного или автоматизированного анализа данных последовательности.

В некоторых вариантах осуществления машиночитаемый носитель содержит машиноисполняемые инструкции для: (a) отнесения чтений последовательностей к семействам UID, где каждый член семейства UID включает одну и ту же экзогенную UID последовательность; (b) отнесения чтений последовательностей каждого семейства UID к подсемействам Уотсона и Крика; (c) идентификацию нуклеотидной последовательности в виде точно представленной цепи Уотсона анализируемого фрагмента ДНК, если пороговый процент членов подсемейства Уотсона содержит такую последовательность; (d) идентификацию нуклеотидной последовательности в виде точно представленной цепи Крика анализируемого фрагмента ДНК, если пороговый процент членов подсемейства Крика содержит такую последовательность; (e) идентификацию мутации в нуклеотидной последовательности, точно представляющей цепь Уотсона, если последовательность, точно представляющая цепь Уотсона, отличается от референсной последовательности, в которой такая мутация отсутствует; (f) идентификацию мутации в нуклеотидной последовательности, точно представляющей цепь Крика, если последовательность, точно представляющая цепь Крика, отличается от референсной последовательности, в которой такая мутация отсутствует; и (g) идентификацию мутации в анализируемом фрагменте ДНК, если мутация в нуклеотидной последовательности, точно представляющей цепь Уотсона, и мутация в нуклеотидной последовательности, точно представляющей цепь Крика, представляют собой одну и ту же мутацию.

В некоторых вариантах осуществления машиночитаемый носитель содержит исполняемый код для отнесения членов семейства UID к подсемейству Уотсона или подсемейству Крика на основе пространственного отношения экзогенной UID последовательности к последовательности чтений R1 и R2. В некоторых вариантах осуществления машиноисполняемый код относит членов семейства UID к подсемейству Уотсона, если экзогенная UID последовательность расположена после

последовательности R2 и перед последовательностью R1. В некоторых вариантах осуществления машиноисполняемый код относит членов семейства UID к подсемейству Крика, если экзогенная UID последовательность расположена после последовательности R1 и перед последовательностью R2. В некоторых вариантах осуществления машиноисполняемый код назначает членов семейства UID подсемейству Уотсона, если экзогенная UID последовательность расположена в большей близости от последовательности R2 и в меньшей близости от последовательности R1. В некоторых вариантах осуществления машиноисполняемый код относит членов семейства UID к подсемейству Крика, если экзогенная UID последовательность находится в большей близости от последовательности R1 и в меньшей близости от последовательности R2. В некоторых вариантах осуществления машиноисполняемый код относит членов семейства UID к подсемейству Уотсона, если экзогенная UID последовательность расположена непосредственно после или в пределах 1-70, 1-60, 1-50, 1-40, 1-30, 1-20, 1-10 или 1-5 нуклеотидов от последовательности R2. В некоторых вариантах осуществления машиноисполняемый код относит членов семейства UID к подсемейству Крика, если экзогенная UID последовательность расположена непосредственно после по течению или в пределах 1-70, 1-60, 1-50, 1-40, 1-30, 1-20, 1-10 или 1-5 нуклеотидов от последовательности R1.

В некоторых вариантах осуществления машиночитаемый носитель содержит исполняемый код для картирования чтений последовательности на референсном геноме. В некоторых вариантах осуществления референсный геном представляет собой референсный геном человека.

В некоторых вариантах осуществления машиночитаемый носитель содержит исполняемый код для создания отчета об активности заболевания, прогнозе или терагностике на основе присутствия, отсутствия или количества мутации в образце. В некоторых вариантах осуществления заболеванием является рак.

В некоторых вариантах осуществления машиночитаемый носитель содержит исполняемый код для создания отчета о вариантах лечения на основе присутствия, отсутствия или количества мутации в образце.

В некоторых вариантах осуществления машиночитаемый носитель содержит исполняемый код для передачи данных по сети.

Компьютерная система

Также в настоящем документе предложены компьютерные системы. В некоторых вариантах осуществления компьютерная система содержит блок памяти, выполненный с возможностью приема и хранения данных о последовательности из образца нуклеиновой кислоты, данных, сгенерированных способом, описанным в настоящем документе; и процессор, соединенный с возможностью связи с запоминающим устройством, где процессор содержит машиночитаемый носитель, раскрытый в настоящем документе.

На **ФИГ. 32** показан пример компьютерной системы **900**, предназначенной для предоставления пользователю возможности исследования образца нуклеиновой кислоты в

соответствии с любым из способов, описанных в настоящем документе. Система **900** включает в себя центральный компьютерный сервер **901**, который запрограммирован для реализации иллюстративных способов, описанных в настоящем документе. Сервер **901** включает в себя центральный процессор (ЦП, также "процессор") **905**, который может быть одноядерным процессором, многоядерным процессором или множеством процессоров для параллельной обработки данных. Сервер **901** также включает в себя запоминающее устройство **910** (например, оперативное запоминающее устройство, постоянное запоминающее устройство, флэш-память); электронный накопитель **915** (например, жесткий диск); коммуникационный интерфейс **920** (например, сетевой адаптер) для связи с одной или больше другими системами, например, с системой секвенирования; и внешние устройства **925**, которые могут включать в себя кэш-память, другую память, хранилище данных и/или электронные графические адаптеры. Запоминающее устройство **910**, накопитель **915**, интерфейс **920** и внешние устройства **925** связаны с процессором **905** через коммуникационную шину (сплошные линии), такую как материнская плата. Накопитель **915** может быть накопителем данных для хранения данных. Сервер **901** оперативно связан с компьютерной сетью ("сетью") **930** посредством коммуникационного интерфейса **920**. Сеть **930** может быть Интернетом, интранетом и/или экстранетом, интранетом и/или экстранетом, которые связаны с Интернетом, телекоммуникационной сетью или сетью передачи данных. Сеть **930** в некоторых случаях при посредстве сервера **901** может реализовывать одноранговую сеть, которая может позволять устройствам, подключенным к серверу **901**, выполнять функцию клиента или сервера.

Накопитель **915** может хранить файлы, такие как данные о последовательностях, данные о последовательности штрих-кодов или любой аспект данных, связанных с изобретением. Накопитель данных **915** может быть связан с данными, относящимися к местоположениям ячеек в виртуальной сетке.

Сервер может связываться с одной или больше удаленными компьютерными системами по сети **930**. Одна или больше удаленных компьютерных систем могут быть, например, персональными компьютерами, ноутбуками, планшетами, телефонами, смартфонами или карманными персональными компьютерами.

В некоторых ситуациях система **900** включает в себя один сервер **901**. В других ситуациях система включает в себя множество серверов, сообщающихся друг с другом через интранет, экстранет и/или Интернет.

Сервер **901** может быть предназначен для хранения данных о последовательности, данных об образце нуклеиновой кислоты, данных о биологическом образце, данных о субъекте и/или другой потенциально значимой информации. Такая информация может храниться на накопителе **915** или сервере **901**, причем такие данные могут передаваться по сети.

Способы, описанные в настоящем документе, могут быть реализованы посредством машиночитаемого (например, компьютерным процессором) носителя (или

программного обеспечения), сохраненного на электронном накопителе сервера компьютерами, например, на запоминающем устройстве **910** или электронном накопителе **915**. Во время использования код может выполняться процессором **905**.

В некоторых случаях код может извлекаться из накопителя **915** и сохраняться на запоминающем устройстве **910** для быстрого доступа к нему процессора **905**. В некоторых ситуациях электронный накопитель **915** может быть исключен, а машиноисполняемые инструкции сохранены на запоминающем устройстве **910**. В альтернативе код может выполняться на второй компьютерной системе **940**.

Аспекты систем и способов, представленных в данном документе, таких как сервер **901**, могут быть реализованы при программировании. Различные аспекты технологии могут рассматриваться как "продукты" или "изделия", обычно в форме исполняемого машиной (или процессором) кода и/или связанных данных, которые перенесены на или реализованы в типе машиночитаемого носителя (например, компьютерного носителя). Машиноисполняемый код может храниться на электронном накопителе, таком как запоминающее устройство (например, постоянное запоминающее устройство, оперативное запоминающее устройство и флэш-память) или жесткий диск. Носители типа "накопители" могут включать любую возможную физическую память компьютеров, процессоров и т.п. или связанных с ними модулей, таких как различные полупроводниковые запоминающие устройства, ленточные накопители, дисковые накопители и т.п., которые могут обеспечивать долговременное хранение в любое время для составления программного обеспечения. Все программное обеспечение или его части могут периодически передаваться через Интернет или другие различные телекоммуникационные сети. Такие каналы связи, например, могут обеспечивать загрузку программного обеспечения с одного компьютера или процессора на другой, например, с сервера управления или центрального компьютера на компьютерную платформу сервера приложений. Таким образом, другой тип носителя, который может нести программные элементы, включает в себя оптические, электрические и электромагнитные волны, например, используемые через физические интерфейсы между локальными устройствами, через проводные и оптические стационарные сети и по различным радиоканалам. Физические элементы, которые передают такие волны, такие как проводные или беспроводные устройства, оптические линии связи и т.п., также могут рассматриваться в качестве носителей, несущих программное обеспечение. При использовании в настоящем документе термины, такие как компьютерный или машиночитаемый носитель, если они не ограничиваются постоянными, физическими "носителями информации", относятся к любому носителю, который участвует в передаче инструкций процессору для исполнения.

Следовательно, машиночитаемый носитель, такой как машиноисполняемый код, может принимать множество форм, включающих, без ограничения, физический носитель данных, носитель несущей волны или физическую передающую среду. Энергонезависимые носители данных могут включать в себя, например, оптические или магнитные диски, такие как любые запоминающие устройства на любом компьютере(ах) и

т.п., которые могут использоваться для реализации системы. Физические передающие среды могут включать: коаксиальные кабели, медные провода и оптоволокно (включая провода, составляющие шину в компьютерной системе). Передающая среда несущей волны может иметь форму электрических или электромагнитных сигналов, акустических или световых волн, таких как генерируемые во время радиочастотной (РЧ) и инфракрасной (ИК) передачи данных. Таким образом, распространенные формы машиночитаемых носителей включают, например: дискету, гибкий диск, жесткий диск, магнитную ленту, любой другой магнитный носитель, CD-ROM, DVD, DVD-ROM, любой другой оптический носитель, перфокарты, бумажную ленту, любой другой физический носитель с группой отверстий, RAM, ROM, PROM и EPROM, FLASH-EPROM, любое другое однокристалльное запоминающее устройство или картридж, несущую волну, передающую данные или инструкции, кабели или линии, передающие такую несущую волну, или любой другой носитель, с которого компьютер может считывать программный код и/или данные. Многие из этих форм машиночитаемых носителей могут использоваться для переноса одной или нескольких последовательностей одной или больше инструкций в процессор для исполнения.

Результаты анализа могут быть представлены пользователю с помощью пользовательского интерфейса, такого как графический пользовательский интерфейс.

Далее изобретение будет описано в следующих примерах, которые не ограничивают объем изобретения, описанный в формуле изобретения.

ПРИМЕРЫ

Пример 1: Дуплексная якорная ПЦР

Материалы и методы

Получение библиотеки дуплексной якорной ПЦР

Этот протокол позволяет получать дуплексные библиотеки при использовании набора Swift Accel-NGS 2S PCR-Free Library Kit (номера по кат. 20024 и 20096) и специфических усеченных адаптеров и праймеров. В некоторых случаях полноразмерные последовательности для присоединения P5 и P7 могут быть добавлены в библиотеку с помощью отдельной ПЦР для секвенирования на приборе Illumina.

Данный протокол предназначен для ПЦР пробирок, но он может быть масштабирован для ПЦР планшетов.

Материалы:

1. Набор Swift Accel-NGS 2S PCR-Free Library Kit (номера по кат. 20024 и 20096)

2. Адаптер 3' Swift N14 Adapter 1 v3A

а.

/5Phos/ATAAAACGACGGCNNNNNNNNNNNNNNNAGATCGGAAGAGCACACGTCTGAACTCCAG*T*C/3SpC3/(SEQ ID NO:1)

б. Заказ в IDT очистки с помощью ПААГЭ, масштаб синтеза 1 мкмоль, лиофилизированный

с. /3SpC3/ можно заменить /3Phos/, фосфоротиоатные связи не требуются, и

олигонуклеотид может быть очищен с помощью ВЭЖХ

3. 3' Swift Адаптер 2 v3'dT

a. GCCGUCGUUUUA/33dT/(SEQ ID NO:3)

b. /33dT/ является модификацией, не входящей в каталог IDT, для 3'-дезоксит

c. Заказ в IDT очистки с помощью ПААГЭ, масштаб синтеза 1 мкмоль, лиофилизированный

4. 5' Swift Адаптер

a. /5SpC3/A*C*ACTCTTCCCTACACGACGCTCTCCGATCT (SEQ ID NO:4)

b. Заказ в IDT очистки с помощью ПААГЭ, масштаб синтеза 1 мкмоль, лиофилизированный

c. /5SpC3/ и фосфоротиоатные связи не требуются, и олигонуклеотид должен быть очищен с помощью ВЭЖХ

d. Можно заменить реагентом truB2 из набора 2S Dual Indexing Kit (номер по кат. 28096)

5. Мастер-микс NEB Ultra II Q5 (номер по кат. M0544L)

6. Усеченный праймер P5

a. ACACTCTTCCCTACACGACGCTCTCCGATCT (SEQ ID NO:4)

b. Модификации не требуются, обессоленный в IDT, 100 мкМ в IDTE

7. Усеченный праймер P7

a. GACTGGAGTTCAGACGTGTGCTCTCCGATCT (SEQ ID NO:5)

b. Модификации не требуются, обессоленный в IDT, 100 мкМ в IDTE

8. Частицы SPRIselect (Beckman Coulter, номер по кат. B23317/B23318/B23319)

9. 80% EtOH (~2 мл на образец)

10. ПЦР-пробирки в стрипах (например, номер по кат GeneMate VWR 490003-710)

11. Магнитный штатив (например, Permagen MSRLV08)

12. Фермент USER (номер по кат. NEB M5505L), который представляет собой смесь урацил-ДНК-гликозилазы и ДНК гликозилазы-лиазы Эндонуклеазы VIII.

Подготовка специализированных адаптеров (могут быть изготовлены на заказ однократно крупной партией):

1. Если не используется реагент truB2 Swift, 5' Swift Адаптер ресуспендируют в концентрации 42 мкМ в TE с низкой концентрацией ЭДТА (входит в набор Swift 2S)

2. Ресуспендируют 3' Swift Адаптер N14 1 v3A до 100 мкМ в TE с низкой концентрацией ЭДТА (входит в набор Swift 2S). Хранение при -20°C для последующего использования.

3. Ресуспендируют 3' Swift Адаптер 2 v3'dT до 100 мкМ в TE с низкой концентрацией ЭДТА (входит в набор Swift 2S). Хранение при -20°C для последующего использования.

4. Отжиг 3' Swift Адаптера N14 1 v3A с 3' Swift Адаптером 2 v3'dT путем смешивания 100 мкл каждого олигонуклеотида при комнатной температуре. Пробирку обозначают 3' Swift N14 v3'dT Duplex Adapter, 50 мкМ. Конечная концентрация 3' дуплекс

адаптеров составляет 50 мкМ. Инкубируют в течение по меньшей мере 5 минут при комнатной температуре перед использованием. Хранение при -20°C для последующего использования.

Технические примечания:

Пробирки с ферментами могут быть извлечены с хранения на -20°C и помещены в лед приблизительно на 10 минут, чтобы ферменты могли достичь температуры 4°C перед пипетированием. Пипетирование ферментов при -20°C может привести к нехватке ферментных реагентов.

После размораживания реагентов до 4°C, реагенты (кроме ферментов) можно быстро встряхивать на вортексе, чтобы хорошо их перемешать. Все пробирки следует центрифугировать в микроцентрифуге, чтобы собрать содержимое перед открыванием.

Все мастер-миксы реагентов собирают на льду и измеряют объемы как соответствующие, используя 5% избыточный объем, чтобы скомпенсировать потери при пипетировании.

Реагенты следует добавлять к мастер-миксу в установленном порядке, как указано в протоколе.

Реагенты могут быть подготовлены заранее (например, чтобы гарантировать, что магнитные частицы не высохнут во время этапа подбора размера).

Этап 1: Репарация матрицы

1. Перенести 11 нг образца вкДНК в 0,2 мл пробирку для ПЦР и при необходимости довести объем образца до конечного объема 37 мкл при использовании ТЕ с низкой концентрацией ЭДТА.

2. Добавить в каждый образец по 3 мкл фермента USER.

3. Перемешать на вортексе и осторожно центрифугировать, чтобы собрать всю жидкость на дне пробирки.

4. Поместить образцы в амплификатор, установленный на 37°C, на 15 минут при отключенном нагреве крышки.

Этап 2: Репарация концов I

1. Аккуратно центрифугировать образцы, чтобы собрать любой образовавшийся конденсат.

2. Добавить 20 мкл предварительно смешанного Мастер-микс Репарации I (см. Таблицу 1) к каждому образцу, содержащему 40 мкл образца ДНК.

Таблица 1. Мастер-микс Репарации концов I

Реагент	Объем на образец
ТЕ с низкой концентрацией ЭДТА	13 мкл
Буфер W1	6 мкл
Фермент W2	1 мкл
Суммарный объем	20 мкл

3. Смешать на вортексе, осторожно отцентрифугировать, поместить в амплификатор и запустить на амплификаторе программу Репарации I в порядке,

описанном ниже.

а. 37°C, 5 минут, с включенным подогревом крышки (подогрев крышки установлен на 75°C)

б. 65°C, 2 минуты, с включенным подогревом крышки (подогрев крышки установлен на 75°C)

с. 37°C, 5 минут, с включенным подогревом крышки (подогрев крышки установлен на 75°C)

4. После окончания программы амплификатора провести легкое центрифугирование для сбора конденсата.

5. Очистить реакцию Репарации I путем добавления 120 мкл (2,0×) частиц SPRIselect. Перемешать на вортексе. Осторожно центрифугировать для сбора частиц и инкубировать при комнатной температуре в течение 5 минут.

6. Собрать частицы путем помещения образца в магнитный штатив на 5 минут.

7. Удалить и отбросить супернатант, не потревожив осадок.

8. Добавить 180 мкл свежеприготовленного 80% этанольного раствора образца, пока он находится в магнитном штативе. Стараться не потревожить осадок. Инкубировать в течение 30 секунд и затем тщательно удалить этанольный раствор пипеткой P20.

9. Повторить стадию выше для второй промывки 80% этанольным раствором.

10. Удалить остатки этанольного раствора пипеткой P20, позволить частицам высохнуть в течение ~30 секунд. Нужно стараться не пересушить частицы, и сразу перейти к стадии 1 Репарации концов 2.

Этап 3: Репарация концов 2

1. Добавить 50 мкл предварительно смешанного мастер-микса Репарации II (см. Таблицу 2) к частицам для каждого образца и перемешать на вортексе до гомогенного состояния.

Таблица 2. Мастер-микс Репарации концов II

Реагент	Объем на образец
TE с низкой концентрацией ЭДТА	30 мкл
Буфер G1	5 мкл
Реагент G2	13 мкл
Фермент G3	1 мкл
Фермент G4	1 мкл
Суммарный объем	50 мкл

2. Поместить образцы в амплификатор, установленный на 20°C в течение 20 минут при отключенном нагреве крышки.

3. После завершения программы амплификатора осторожно отцентрифугировать пробирки для сбора конденсата.

4. Очистить реакцию Репарации 2 путем добавления 90 мкл (1,8×) раствора ПЭГ/NaCl. Перемешать на вортексе. Осторожно отцентрифугировать для сбора частиц и инкубировать при комнатной температуре в течение 5 минут.

5. Собрать частицы путем помещения образца в магнитный штатив на 5 минут.
6. Удалить и отбросить супернатант, не потревожив осадок.
7. Добавить 180 мкл свежеприготовленного 80% этанольного раствора образца, пока он находится в магнитном штативе. Стараться не потревожить осадок. Инкубировать в течение 30 секунд и затем тщательно удалить этанольный раствор пипеткой Р20.
8. Повторить стадию выше для второй промывки 80% этанольным раствором.
9. Удалить остатки этанольного раствора пипеткой Р20, позволить частицам высохнуть в течение ~30 секунд. Нужно стараться не пересушить частицы, и сразу перейти к стадии 1 Лигирования 1.

Этап 4: Лигирование I

1. Добавить 30 мкл предварительно смешанного мастер-микса Лигирования I (см. Таблицу 3) к частицам для каждого образца и смешать на вортексе до гомогенного состояния.

Таблица 3. Мастер-микс Лигирования I

Реагент	Объем на образец
TE с низкой концентрацией ЭДТА	20 мкл
3' Swift N14 v3'dT Duplex Adapter, 50 мкМ	5 мкл
Буфер Y1	3 мкл
Фермент Y3	2 мкл
Суммарный объем	30 мкл

2. Поместить образцы в амплификатор, установленный на 25°C в течение 15 минут при отключенном нагреве крышки.
3. После завершения программы амплификатора осторожно отцентрифугировать пробирки для сбора конденсата.
4. Очистить реакцию Лигирования 1 путем добавления 36 мкл (1,2×) раствора ПЭГ/NaCl. Перемешать на вортексе. Осторожно отцентрифугировать для сбора частиц и инкубировать при комнатной температуре в течение 5 минут.
5. Собрать частицы путем помещения образца в магнитный штатив на 5 минут.
6. Удалить и отбросить супернатант, не потревожив осадок.
7. Добавить 180 мкл свежеприготовленного 80% этанольного раствора образца, пока он находится в магнитном штативе. Стараться не потревожить осадок. Инкубировать в течение 30 секунд и затем тщательно удалить этанольный раствор пипеткой Р20.
8. Повторить стадию выше для второй промывки 80% этанольным раствором.
9. Удалить остатки этанольного раствора пипеткой Р20, позволить частицам высохнуть в течение ~30 секунд. Нужно стараться не пересушить частицы, и сразу перейти к стадии 1 Лигирования 2.

Этап 5: Лигирование 2

1. Добавить 50 мкл предварительно смешанного мастер-микса Лигирования II (см. Таблицу 4) к частицам для каждого образца и перемешать на вортексе до гомогенного состояния.

Таблица 4. Мастер-микс Лигирования II

Реагент	Объем на образец
TE с низкой концентрацией ЭДТА	30 мкл
Реагент В1	5 мкл
3' Swift Адаптер, 50 мкМ или Реагент truB2	2 мкл
Реагент В3	9 мкл
Фермент В4	1 мкл
Фермент В5	2 мкл
Фермент В6	1 мкл
Суммарный объем	50 мкл

2. Поместить образцы в амплификатор, установленный на 40°C в течение 10 минут при отключенном нагреве крышки.

3. После завершения программы амплификатора осторожно отцентрифугировать пробирки для сбора конденсата.

4. Очистить реакцию Лигирования 1 путем добавления 52,5 мкл (1,05×) раствора ПЭГ/NaCl. Перемешать на вортексе. Осторожно отцентрифугировать для сбора частиц и инкубировать при комнатной температуре в течение 5 минут.

5. Собрать частицы путем помещения образца в магнитный штатив на 5 минут.

6. Удалить и отбросить супернатант, не потревожив осадок.

7. Добавить 180 мкл свежеприготовленного 80% этанольного раствора образца, пока он находится в магнитном штативе. Стараться не потревожить осадок. Инкубировать в течение 30 секунд и затем тщательно удалить этанольный раствор пипеткой Р20.

8. Повторить стадию выше для второй промывки 80% этанольным раствором.

9. Удалить остатки этанольного раствора пипеткой Р20, позволить частицам высохнуть в течение ~30 секунд. Нужно стараться не пересушить частицы, и сразу ресуспендировать частицы в 24 мкл TE с низкой концентрацией ЭДТА. Перемешать на вортексе и инкубировать в течение 2 минут.

10. Осторожно отцентрифугировать для сбора частиц и собрать частицы на магнитном штативе в течение 2 минут.

Этап 6: Амплификация ПЦР-библиотеки

1. Добавить 26 мкл предварительно смешанного мастер-микса для Амплификации ПЦР-библиотеки (см. Таблицу 5) в чистую пробирку для каждого образца.

Таблица 5. Мастер-микс Амплификации ПЦР-библиотеки

Реагент	Объем на образец
Мастер-микс NEB Ultra II Q5 (2×)	25 мкл
Усеченный праймер P5, 100 мкМ	0,5 мкл
Усеченный праймер P7, 100 мкМ	0,5 мкл
Суммарный объем	26 мкл

2. Тщательно перенести супернатант, содержащий готовую библиотеку после

лигирования в мастер-микс для Амплификации ПЦР-библиотеки.

3. Перенести остатки готовой библиотеки после лигирования с использованием пипетки P20. Нужно стараться перенести максимально возможное количество супернатанта.

4. Перемешать на вортексе, осторожно отцентрифугировать, поместить в амплификатор и запустить на амплификаторе программу Амплификации ПЦР-библиотеки в порядке, описанном ниже.

Таблица 6. Иллюстративная программа амплификатора для Амплификации ПЦР-библиотеки

Стадия	Температура	Время
Начальная денатурация	98°C	30 секунд
5 циклов ПЦР (4 стадии GOTO)	98°C	10 секунд
	65°C	75 секунд
Конечная достройка	65°C	5 минут
Выдерживание	4°C	Выдерживание

5. Очистить реакцию Амплификации ПЦР-библиотеки путем добавления 90 мкл (1,8×) частиц SPRIselect. Перемешать на вортексе. Осторожно отцентрифугировать для сбора частиц и инкубировать при комнатной температуре в течение 5 минут.

6. Собрать частицы путем помещения образца в магнитный штатив на 5 минут.

7. Удалить и отбросить супернатант, не потревожив осадок.

8. Добавить 180 мкл свежеприготовленного 80% этанольного раствора образца, пока он находится в магнитном штативе. Стараться не потревожить осадок. Инкубировать в течение 30 секунд и затем тщательно удалить этанольный раствор пипеткой P20.

9. Повторить стадию выше для второй промывки 80% этанольным раствором.

10. Удалить остатки этанольного раствора пипеткой P20, позволить частицам высохнуть в течение ~30 секунд. Нужно стараться не пересушить частицы, и сразу ресуспендировать частицы в 47 мкл ТЕ с низкой концентрацией ЭДТА. Перемешать на вортексе и инкубировать в течение 2 минут.

11. Осторожно отцентрифугировать для сбора частиц и собрать частицы на магнитном штативе в течение 2 минут.

12. Тщательно перенести супернатант, содержащий готовую PCR-амплифицированную библиотеку в чистую пробирку, не захватывая частицы.

13. Провести анализ 1 мкл амплифицированных библиотек с помощью TapeStation. Интенсивный пик должен присутствовать на ~300 пн, соответствующий лигированной с адаптером мононуклеосомной ДНК (180 пн+60 пн+59 пн).

14. Хранить библиотеки при -20°C.

Точное и эффективное обнаружение редких мутаций с помощью дуплексной якорной ПЦР

Библиотеки для секвенирования, включающие дуплексные молекулярные штрих-коды, были созданы путем последовательного лигирования двух адаптерных молекул с

вводной двухцепочечной ДНК. Сначала вводную ДНК подвергали репарации концов посредством реакций затупления и дефосфорилирования (Фиг. 9 и 10). После репарации концов 3'-адаптер, содержащий 5'-фосфат (3'-олиго #1), отоженный с коротким олигонуклеотидом с заблокированной 3'-группой (3'-олиго #2), лигировали с каждым 3'-концом вводной ДНК (Фиг. 12). Поскольку один из олигонуклеотидов содержит 3'-блокирующую группу, только олигонуклеотид, содержащий 5'-фосфат (3'-олиго № 1), ковалентно связывался к вводной ДНК на 3'-концах. Присоединенный 3'-олигонуклеотид также содержит молекулярный штрих-код, который однозначно метит каждую цепь (Фиг. 11). Затем 3'-олигонуклеотид, содержащий 3'-блокирующую группу, расщепляли и 5'-адаптерный олигонуклеотид лигировали с каждым 5'-концом посредством реакции, подобной ник-трансляции. В частности, 5'-адаптерный олигонуклеотид отжигается непосредственно перед молекулярным штрих-кодом на 3'-адаптерном олиго #1, оставляя гэп. Этот гэп заполнялся и закрывался в ходе реакции, подобной ник-трансляции, с образованием в результате дуплексного молекулярного штрих-кода *in situ* на каждом конце фрагментов ДНК (Фиг. 13). Полученные лигированные продукты очищали и амплифицировали посредством первичной полногеномной ПЦР (Фиг. 14).

Продукт после первичной полногеномной ПЦР можно дополнительно очищать с получением библиотек одноцепочечных (оц) НК, соответствующих смысловым и антисмысловым цепям (Фиг. 2 и 3).

Библиотеки амплифицированных ДНК обогащали требуемыми мишенями с использованием метода цепь-специфической якорной ПЦР. В таком ПЦР обогащении использовали один праймер, направленный на требуемую область, представляющую интерес, и второй праймер, направленный на лигированную адаптерную последовательность (Фиг. 4, 5, 15, 17). Для повышения специфичности обогащения мишенями может быть проведена вторая гнездовая ПЦР с использованием одного праймера, направленного на требуемую область, представляющую интерес, и второго праймера, направленного на лигированную адаптерную последовательность (Фиг. 7, 8, 16, 18). Для повышения специфичности обогащения мишенями может использоваться вторая гнездовая ПЦР для включения штрих-кодов образцов, а также необходимых последовательностей для присоединения, требуемых для секвенирования следующего поколения. Полученные библиотеки затем количественно оценивают, нормализуют и секвенируют.

После секвенирования чтения выравнивали с геномом и группировали по их молекулярному штрих-коду. Фрагменты, содержащие чтения с одним и тем же молекулярным штрих-кодом, которые одновременно картируются на смысловой и на антисмысловой цепи мишени, были обозначены как имеющие "дуплексное подтверждение". Мутации учитывались только в том случае, если они присутствовали в обеих цепях (фиг. 20 и 21).

Пример 2: Направленное ДНК секвенирование цепей ДНК Уотсона и Крика

Идентификация и количественная оценка редких последовательностей

нуклеиновых кислот важна для многих областей биологии и клинической медицины. В данном примере описан способ (названный SaferSeqS), который позволяет решить эту проблему путем: (i) эффективного введения идентичных молекулярных штрих-кодов в цепи Уотсона и Крика молекул-матриц и (ii) обогащения представляющих интерес областей генома с помощью нового цепь-специфического ПЦР-анализа. Этот способ можно применять для оценки мутаций в одном ампликоне или одновременно во множестве ампликонов, при этом он позволяет оценивать ограниченные количества ДНК, например, присутствующие в плазме крови, и снижает частоту ошибок существующих подходов молекулярного баркодирования на основе ПЦР, по меньшей мере на два порядка.

Результаты

Чтобы устранить неэффективность и допускаемые ошибки, обычно связанные с созданием библиотеки, была разработана стратегия, которая включает последовательное лигирование адаптерных последовательностей с 3'- и 5'-концами фрагментов ДНК и создание двухцепочечных молекулярных штрих-кодов *in situ* (Фиг. 22a). Создание молекулярных штрих-кодов *in situ* является ключевой инновацией нового способа получения библиотек. Ферменты, используемые для создания двухцепочечных молекулярных штрих-кодов *in situ*, обеспечивают уникальное баркодирование каждого фрагмента ДНК и исключают необходимость ферментной подготовки дуплексных адаптеров (Фиг. 22a, этапы 2 и 3). Адаптер содержал участок из 14 случайных нуклеотидов в виде экзогенного молекулярного штрих-кода (последовательность уникального идентификатора [UID]). Лигированные с адаптером фрагменты подвергали ограниченному числу циклов ПЦР для создания избыточных копий двух исходных цепей ДНК (Фиг. 22a, этап 4). Для ясности в данном иллюстративном варианте осуществления референсная последовательность UCSC (доступная на сайте genome.ucsc.edu/) была произвольно определена как цепь "Уотсона", а ее обратный комплемент - как цепь "Крика".

Еще одним нововведением в этом протоколе является применение для обогащения метода на основе полугнездовой ПЦР. Хотя полугнездовая ПЦР ранее уже использовалась для обогащения мишеней (см., например, Zheng et al., 2014, Nat Med 20:1479-1484), для ее применения к дуплексному секвенированию потребовались серьезные изменения. В частности, были проведены две отдельные ПЦР - одна для цепи Уотсона и одна для цепи Крика. В обеих ПЦР использовался один и тот же ген-специфический праймер, но при этом в каждой использовались разные якорные праймеры. ПЦР дубликаты, полученные с каждой цепи, можно различать по ориентации вставки относительно экзогенного UID (Фиг. 22b).

После секвенирования, чтения, соответствующие каждой цепи исходных ДНК дуплексов, группировали в семейства Уотсона и Крика. Каждый член семейства имел идентичный эндогенный штрих-код, представляющий последовательность, на одном конце исходного фрагмента матрицы, и идентичный экзогенный UID, введенный *in situ* во

время создания библиотеки. Мутации, присутствующие в >80% семейства цепей Уотсона, были названы "супермутантами Уотсона". Мутации, присутствующие в > 80% семейства цепей Крика, были названы "супермутантами Крика". А мутации, которые присутствовали в >80% семейств Уотсона и Крика с одинаковым UID ("дуплексное семейство"), были названы "суперкалифраджилистикэсприалидоушес мутантами" (от англ. *supercalifragilisticexpialidocious* - не имеющее значения слово и название одноименной песни из фильма "Мери Поппинс" 1964 года), называемыми в настоящем документе "суперкалимутантами" (Фиг. 22с).

В качестве первоначальной демонстрации SaferSeqS был проведен эксперимент со смешиванием, в котором ДНК с известной мутацией добавляли к ДНК лейкоцитов здорового человека в соотношении от 10% до 0%. Было сделано предположение, что эти смеси приведут к 15 400, 150, 15, 15, 8 или 0 суперкалимутантов на анализ. Доля целевых чтений (т.е. чтений, состоящих из предполагаемого ампликона) составила 88%, что намного выше, чем можно достичь с применением методов, основанных на гибридном захвате (см., например, Samorodnitsky et al., 2015 *Hum Mutat* 36:903-914). Более того, была продемонстрирована сильная корреляция между ожидаемой и наблюдаемой частотами аллелей по пяти порядкам величины (Фиг. 23, r -коэффициент Пирсона $>0,999$, $p=2,02 \times 10^{-12}$). Ни один мутант, соответствующий предварительно определенному, добавленному в смесь варианту, не наблюдали в ДНК здорового человека, что указывает на очень высокую специфичность представляющей интерес мутации. Специфичность также определяли для *любого* основания в ампликоне, а не только для запрашиваемого основания. В общей сложности из 37747670 оснований, запрошенных среди всех образцов ДНК, наблюдали только шесть суперкалимутантов, что соответствует частоте мутаций $1,59 \times 10^{-7}$ суперкалимутантов/пн (Таблица 7).

Таблица 7. Мутации, идентифицированные в экспериментах по проверке аналитической чувствительности и специфичности.

* Координаты относятся к выпуску референсного генома человека hg19 (Консорциум референсного генома GRCh37, февраль 2009 года).

Образец	Хром	Положение*	Референсное основание	Вариантное основание	Ген	Изменение кДНК	Изменение белка	Число суперкаллимутантов	Глубина дуплекса	MAF
Введение 10%	Хр.17	7579507	T	G	TP53	c.180A>C	p.P60P	3067	37686	8.14E+00
Введение 1%	Хр.17	7579507	T	G	TP53	c.180A>C	p.P60P	280	34997	8.00E-01
Введение 1%	Хр.17	7579560	A	G	TP53	c.127T>C	p.L43L	1	1481	6.75E-02
Введение 0,1%	Хр.17	7579507	T	G	TP53	c.180A>C	p.P60P	28	36764	7.62E-02
Введение 0,01%	Хр.17	7579507	T	G	TP53	c.180A>C	p.P60P	3	37872	7.92E-03
Введение 0,001%	Хр.17	7579507	T	G	TP53	c.180A>C	p.P60P	2	206682	9.68E-04
Введение 0,001%	Хр.17	7579518	C	G	TP53	c.169G>C	p.D57H	1	193317	5.17E-04
Введение 0,001%	Хр.17	7579547	G	A	TP53	c.140C>T	p.P47L	1	153733	6.50E-04
Введение 0,001%	Хр.17	7579575	G	-	TP53	c.112del	p.Q38del	1	37951	2.63E-03
Введение 0,001%	Хр.17	7579716	G	A	TP53	c.80C>T	p.P27L	1	14207	7.04E-03
Введение 0%	Хр.17	7579508	G	C	TP53	c.179C>G	p.P60R	1	200754	4.98E-04

Затем была предпринята попытка определить, можно ли применять SaferSeqS к клиническим образцам, в которых количество ДНК ограничено. Например, всего 33 нг ДНК часто присутствует в десяти мл образцов внеклеточной ДНК плазмы крови, используемых для жидкостной биопсии. Подавляющее большинство молекул ДНК-матриц в этих образцах относится к дикому типу, при этом всего одна или две мутантные матрицы из 10000 матриц дикого типа присутствуют в образцах, взятых у пациентов с низкой опухолевой нагрузкой. Чтобы точно обнаружить такое чрезвычайно малое количество мутантных матриц, анализ должен обеспечивать эффективное выделение последовательности исходных молекул.

Чтобы оценить SaferSeqS в такой сложной ситуации, внеклеточную ДНК плазмы онкобольных смешали с внеклеточной ДНК плазмы здоровых лиц, чтобы воспроизвести частоты мутаций, которые обычно наблюдаются в клинических образцах. В этих экспериментах 33 нг каждого образца исследовали на присутствие одной из трех различных мутаций в TP53. Медианный процент целевых чтений в 27 экспериментальных условиях (3 ампликона TP53 \times 3 образца \times 3 аликвоты/образец) составил 80% (диапазон: от 72% до 91%) (Фиг. 24а). Медианное количество дуплексных семейств (т.е. цепей Уотсона и Крика, содержащих одинаковые эндогенные и экзогенные штрих-коды), составляло 89% (диапазон: от 65% до 102%) от числа исходных молекул-матриц (Фиг. 24b). Более того, во всех шести образцах с добавкой примеси, представляющий интерес суперкалимутант был идентифицирован с ожидаемой частотой (Фиг. 25b, d, e, Таблица 9). Мутации с этой ожидаемой частотой также были идентифицированы в тех же образцах с использованием ранее описанного метода молекулярного баркодирования ("SafeSeqS", а не "SaferSeqS") (Фиг. 25а, b, c, Таблица 8). Преимуществом SaferSeqS была его специфичность. В общей сложности было обнаружено 1406 супермутантов, представляющих 153 различных мутации, наблюдаемых при использовании ранее описанного метода, что отражает среднюю частоту ошибок $9,39 \times 10^{-6}$ супермутантов/пн (Фиг. 25а, b, c, Таблица 8). Подавляющее большинство этих мутаций, предположительно, представляли собой ошибки полимеразы, которые возникали в ходе ранних циклов баркодирования только в одной из двух цепей. Аналогичным образом, если рассматривать только супермутанты Уотсона или супермутанты Крика (т.е. наблюдаемые только в одной из двух цепей, Фиг. 22с), а не суперкалимутанты, наблюдали частоту ошибок $6,56 \times 10^{-6}$ супермутантов/пн (Фиг. 26, Таблица 9). Напротив, только один суперкалимутант был обнаружен из 4947725 оснований, запрошенных с использованием SaferSeqS, что соответствует общей частоте мутаций $2,02 \times 10^{-7}$ (Таблица 9). Эти различия в специфичности между SaferSeqS и ранее описанными методами молекулярного баркодирования (т.е. теми, в которых используется прямая ПЦР или лигирование с адаптерами для включения молекулярных штрих-кодов перед секвенированием) были высокозначимыми ($P < 3,5 \times 10^{-10}$, двусторонний Z-критерий для пропорций, сравнивающих SaferSeqS с каждым из других методов).

Таблица 8. Сравнение мутаций, идентифицированных с помощью SafeSeqS и SaferSeqS.

(См. Приложение А),

Таблица 9. Сравнение мутаций, идентифицированных с помощью неспецифических к цепи молекулярных штрих-кодов и SaferSeqS.

(См. Приложение В),

В качестве дополнительной демонстрации клинической применимости SaferSeqS были обследованы пять онкобольных с минимальной опухолевой нагрузкой. В каждом случае мутации в первичных опухолях (а не в плазме крови) идентифицировали, как описано в другом источнике (Tie et al., *Sci Transl Med* 8:346ra392 (2016)). Плазму крови этих пациентов делили на две равные аликвоты и оценивали одну аликвоту с помощью метода молекулярного баркодирования, описанного в другом источнике (Kinde et al., *Proc Natl Acad Sci U S A* 108: 9530-9535 (2011)), а другую - с помощью SaferSeqS. В обоих случаях были подобраны праймеры, которые давали бы короткие ампликоны, направленные на интересующие мутации. Оценка с помощью ранее описанного метода баркодирования показала, что образцы плазмы крови несли в общем совокупности восемь мутаций, которые были первоначально идентифицированы в первичных опухолях. Частота этих мутаций в плазме варьировала от 0,01% до 0,1% (**Фиг. 27, Таблица 10**). В дополнение к восьми известным мутациям ранее описанный метод обнаружил 334 различных мутации, присутствующих с частотой до 0,013%, причем ни одна из них не была обнаружена в первичных опухолях этих пациентов. Эти 334 мутации включали 10347 супермутантов, что отражает среднюю частоту ошибок $1,23 \times 10^{-5}$ супермутантов/пн (**Фиг. 27, Таблица 10**). С помощью SaferSeqS восемь мутаций, обнаруженных в первичных опухолях, были обнаружены у всех пяти пациентов с частотами, аналогичными частотам, обнаруженным с помощью ранее описанного метода (**Фиг. 27, Таблица 10**). Однако из 8707755 запрошенных оснований только один дополнительный суперкалимутант (а не 334 мутации) был идентифицирован с помощью SaferSeqS, что соответствует средней частоте ошибок $1,15 \times 10^{-7}$ (**Таблица 10**). Это более чем 100-кратное улучшение специфичности по сравнению с ранее описанным методом молекулярного баркодирования было очень значимым ($P < 2,2 \times 10^{-16}$, двусторонний Z-критерий для пропорций).

Таблица 10. Мутации, идентифицированные с помощью SaferSeqS и SaferSeqS в образцах плазмы крови, полученных от онкобольных.

(См. Приложение С)

Затем исследовали, может ли SaferSeqS одновременно анализировать несколько мишеней, что может быть полезным для различных применений в области секвенирования. SaferSeqS допускает два типа мультиплексирования: один, при котором несколько мишеней анализируют в отдельных реакциях ПЦР, и другой, при котором несколько мишеней анализируют в одной и той же реакции ПЦР. Поскольку во время амплификации библиотеки создаются избыточные копии, полученные с цепей Уотсона и Крика, библиотеку можно разделить на несколько реакций ПЦР без нежелательного влияния на выделение образца. Например, если предположить, что эффективность ПЦР

составляет 70%, можно провести отдельный анализ до 22 мишеней с потерями при выделении <10%, если библиотека ДНК амплифицируется в 11 циклах ПЦР (Фиг. 28). На практике проводили анализ либо 100%, либо 4,4% библиотеки. Процент целевых чтений был одинаковым при использовании 100% или 4,4% библиотеки, причем 82% и 92% чтений были правильно картированы с предполагаемой областью. Количество выделенных дуплексных семейств также было одинаковым: выделено 7825 и 6769 в 100% и 4,4% разделениях библиотеки.

Хотя описанный выше подход мультиплексирования можно применять для одновременного анализа ограниченного числа мишеней, приложения, которые оценивают многие области генома, могут включать мультиплексирование в малом количестве реакций ПЦР. Чтобы оценить возможности мультиплексирования SaferSeqS в данной ситуации, было подобрано 48 праймеров для поиска в областях драйверных генов, которые обычно содержат мутации при раке (**Таблица 11**). Эти праймеры были объединены в две реакции: одна была направлена на 25, а другая на 23 области. Каждая из 48 пар праймеров специфично амплифицирует свои заданные мишени (Фигура 30), и 36 были признаны успешными, поскольку число дуплексных семейств составляло по меньшей мере 50% от числа, идентифицированного в одиночных реакциях. Из этих 36, медианная частота целевых чтений, полученных с цепи Уотсона, составила 95% (диапазон: от 39% до 97%), и медианная частота целевых чтений, полученных с цепи Крика, составила 95% (диапазон: от 39% до 98%). Что наиболее важно, мишени продемонстрировали относительно равномерное восстановление вводимых молекул с коэффициентом вариации всего лишь 17% (Фиг. 29). Длины секвенированных ампликонов (медиана 77 п.н., межквартильный диапазон: 71-83 п.н.) также были одинаковыми для всех ампликонов и соответствовали исходному размеру внеклеточной ДНК плазмы, составляющей ~167 п.н. $\pm 10,4$ п.н. (Фиг. 29).

Таблица 11. Состав и последовательности праймеров GSP мультиплексной панели.

*Координаты относятся к выпуску референсного генома человека hg19 (Консорциум референсного генома GRCh37, февраль 2009 года).

Звездочки в последовательностях праймеров обозначают фосфоротиоатные связи.

Ампликон	Мульти-плексный набор #	Последовательность GSP1	SEQ ID NO:	Последовательность GSP2	SEQ ID NO:	Хром. GSP2	Цепь GSP2	Начальное положение GSP2*	Конечное положение GSP2*
AKT1_17	1	TGCGTGGCTCTCACC*A*C	6	CACCCGCACGTCTGT*A*G	7	Xp.14	-	105,246,554	105,246,570
APC_1310	1	ACAGGATCTTCAGCTGAC*C*T	8	CTTCAGCTGACCTAGTCC*A*A	9	Xp.5	-	112,175,223	112,175,243
CDKN2A_51	1	CCTGGCTCTGACCATTCTG*T*T	10	CCATTCTGTTCTCTCTGG*C*A	11	Xp.9	-	21,971,209	21,971,228
EGFR_856	1	AACACCGCAGCATGTC*A*A	12	GCAGCATGTCAAGATCACA*G*A	13	Xp.7	+	55,259,486	55,259,506
FBXW7_369	1	TTGTGCAGAGTTCAGTTACSTT*A*G	14	GAGTTCAGTTACSTTAGGAGATTG*A*G	15	Xp.4	+	153,251,872	153,251,898
FBXW7_481	1	GGCCTGTCTCAATATCCCA*A*A	16	GTCTCAATATCCCAAACCCTAA*G*A	17	Xp.4	+	153,247,333	153,247,356
GNAS_202	1	GGAACCTGGTCTCAAAGATTC*C*A	18	GGTCTCAAAGATTCAGAAGTC*A*G	19	Xp.20	-	57,484,426	57,484,449
KRAS_14	1	TCTGAATTAGCTGIATCGTCAA*G*G	20	TATCGTCAAGGCACTCTTG*C*C	21	Xp.12	+	25,398,256	25,398,276
KRAS_144	1	CTGTATTTATTTTCAGTGTACTTACCTGT*C*T	22	AGTGTACTTACCTGTCTTGTC*T*T	23	Xp.12	+	25,378,536	25,378,559
NRAS_12	1	GACTGAGTACAAACTGGTG*G*T	24	AAACTGGTGGTGGTTGG*A*G	25	Xp.1	-	115,258,751	115,258,769
PIK3CA_80	1	TTTCGTAAGTGTACTCAAGAA*G*C	26	AGTGTACTCAAGAAGCAGAAA*G*G	27	Xp.3	+	178,916,827	178,916,850
PIK3CA_348	1	GCATCAGCATTTGACTTTACSTTAT*C*A	28	CATTTGACTTTACSTTATCAATGTCT*C*G	29	Xp.3	-	178,921,563	178,921,590
PIK3CA_1053	1	GCATGCTGTTTAATTGTGTG*G*A	30	TGTGTGGAAGATCCAATCCAT*T*T	31	Xp.3	-	178,952,106	178,952,128
PTEN_145	1	CCATAGAAATCTAGGGCCTC*T*T	32	TAGGGCCTCTGTGCCTT*T*A	33	Xp.10	-	89,692,953	89,692,972
TP53_25	1	GGCCTGCCCTTCCAA*T*G	34	TTCCAAATGGATCCACTC*A*C	35	Xp.17	+	7,579,820	7,579,838
TP53_60	1	GCAGCCTCTGGCATT*C*T	36	GGCATTCTGGGAGCTTC*A*T	37	Xp.17	+	7,579,487	7,579,505

TP53_125	1	GCCAGGCATTGAAGTCTC*A*T	38	ATTGAAGTCTCATGGAAGC*C*A	39	Xp.17	+	7,579,272	7,579,292
TP53_161	1	CCGTCATGTGCTGTGA*C*T	40	GTGCTGTGACTGCTTGTA*G*A	41	Xp.17	+	7,578,426	7,578,445
TP53_196	1	CCAAATACTCCACACGCAA*A*T	42	CACACGCAAATTCCTTC*C*A	43	Xp.17	+	7,578,240	7,578,259
TP53_245	1	GGAGTCTTCCAGTGTGATGA*T*G	44	GTGATGATGGTGAGGATG*G*G	45	Xp.17	+	7,577,514	7,577,533
TP53_264	1	CACAAACACGCACCTC*A*A	46	ACGCACCTCAAAGCTGT*T*C	47	Xp.17	+	7,577,119	7,577,137
TP53_305	1	AACTGCACCCTTGGTC*T*C	48	CCTCCACCGCTTCTTG*T*C	49	Xp.17	+	7,576,989	7,577,006
TP53_330	1	GAAACTTTCACCTTGATAAGAGG*T* C	50	TGATAAGAGGTCCCAAGACT*T* A	51	Xp.17	+	7,576,827	7,576,848
TP53_365	1	AGTAGGGCCAGGAAG*G*G	52	GGGCTGAGGTCACCTC*A*C	53	Xp.17	+	7,573,910	7,573,926
TP53_390	1	TGGGAGGCTGTCAGT*G*G	54	GGGAACAAGAAGTGAGAA*T*G	55	Xp.17	+	7,572,906	7,572,926
APC_1458	2	TGCTTAGGTCCACTCTCT*C*T	56	GTCCACTCTCTCTTTTCAG*C *A	57	Xp.5	-	112,175,668	112,175,690
BRAF_602	2	ACAACGTTCAAACTGATGG*G*A	58	ATGGGACCCACTCCAT*C*G	59	Xp.7	+	140,453,111	140,453,128
CTNNB1_32	2	TTAGTCACTGGCAGCAAC*A*G	60	GGCAGCAACAGTCTTAC*C*T	61	Xp.3	+	41,266,077	41,266,095
FBXW7_472	2	TAGAGGAAGAAGTCCCAAC*C*A	62	GAAGTCCCAACCATGACAAG*A* T	63	Xp.4	+	153,249,328	153,249,349
FBXW7_506	2	CTAACAAACCCTCCTGCCA*T*C	64	CCCTCCTGCCATCATATTGA*A* C	65	Xp.4	+	153,247,262	153,247,283
FGFR2_250	2	ACTCCTCCTTTCTTCCCT*C*T	66	TTTCTTCCCTCTCTCCACC*A*G	67	Xp.10	-	123,279,684	123,279,704
HRAS_10	2	GGAGCGATGACGGAATATAA*G*C	68	GGAATATAAGCTGGTGGTGG*T* G	69	Xp.11	-	534,296	534,317
KRAS_62	2	CTCATGTACTGGTCCCTCAT*T*G	70	GTCCCTCATTGCACTGTAC*T*C	71	Xp.12	+	25,380,251	25,380,271
NRAS_59	2	GTCAAACCTGTTTGTGGAC*A*T	72	CCTGTTGTTGGACATACTGGAT *A*C	73	Xp.1	-	115,256,538	115,256,562
PIK3CA_111	2	CACGGTTGCCTACTGG*T*T	74	GTTGCCTACTGGTCAATTACT* T*T	75	Xp.3	-	178,916,911	178,916,934
PIK3CA_542	2	GAACAGCTCAAAGCAATTTCTAC*A* C	76	AATTTCTACAGAGATCCTCTC* T*C	77	Xp.3	+	178,936,057	178,936,080
PPP2R1A_176	2	CGACTCCCAGGTACTTC*C*G	78	TACTTCCGGAACCTGTGC*T*C	79	Xp.19	+	52,715,940	52,715,959
PTEN_130	2	CATGTTGCAGCAATTCACT*G*T	80	CAATTCACTGTAAAGCTGGAAA* G*G	81	Xp.10	+	89,692,878	89,692,901
TP53_36	2	GGGACAGCATCAAATCATC*C*A	82	ATCAAATCATCCATTGCTTG*G* G	83	Xp.17	+	7,579,556	7,579,577

TP53_113	2	GTCACAGACTTGGCTGT*C*C	84	TTGGCTGTCCCAGAAT*G*C	85	Xp.17	+	7,579,328	7,579,345
TP53_138	2	ATCAACCCACAGCTGC*A*C	86	TGCACAGGGCAGGTC*T*T	87	Xp.17	+	7,578,499	7,578,515
TP53_180	2	CCAGCTGCTCACCAT*C*G	88	TCACCATCGCTATCTGAG*C*A	89	Xp.17	+	7,578,367	7,578,386
TP53_220	2	TGACAACCACCCTTAAC*C*C	90	CAGTTGCAAACCAGACCTC*A*G	91	Xp.17	+	7,578,161	7,578,181
TP53_260	2	GTGCAGGGTGGCAAG*T*G	92	GGCAAGTGGCTCCTG*A*C	93	Xp.17	+	7,577,482	7,577,498
TP53_282	2	CCTTTCTTGGGAGATTCTC*T*T	94	CGGAGATTCTCTTCCTCTG*T*G	95	Xp.17	+	7,577,069	7,577,089
TP53_315	2	AAATATTCTCCATCCAGTGGTT*T*C	96	CCAGTGGTTTCTTCTTGG*C*T	97	Xp.17	+	7,576,876	7,576,896
TP53_342	2	GGGCATCCTTGAGITCC*A*A	98	CTTGAGITCCAAGGCTC*A*T	99	Xp.17	+	7,573,974	7,573,993
TP53_380	2	TCAGGCCCTTCTGTCTTG*A*A	100	CCTTCTGTCTTGAACATGAGT*T *T	101	Xp.17	+	7,572,943	7,572,965

Для оценки множественных ампликонов можно использовать два иллюстративных способа. Первый включает параллельные ампликон-специфические ПЦР в разных лунках. В случае жидкостных биопсий для мониторинга рецидива заболевания, когда обычно наблюдают только несколько мутаций драйверных генов, эту стратегию можно легко применять, не беспокоясь о перекрестной гибридизации между праймерами или других проблемах, обычно встречающихся в случае мультиплексных реакций ПЦР. В случае других применений жидкостной биопсии, таких как скрининг, когда представляющая интерес мутация неизвестна, может применяться оценка гораздо большего количества ампликонов; например, комбинация нескольких пар праймеров в каждой лунке ПЦР. Данный пример показал, что по меньшей мере 18 ампликонов можно эффективно анализировать в одной лунке с применением SaferSeqS, при этом стратегии полугнездовой ПЦР, не включающие дуплексное секвенирование, показали, что можно совместно амплифицировать до 313 ампликонов.

Благодаря обеспечению эффективного обнаружения и количественной оценки редких генетических изменений, SaferSeqS позволяет разрабатывать высокочувствительные и специфичные методы молекулярной диагностики на основе ДНК, а также помогает ответить на ряд важных фундаментальных научных вопросов.

Методы

Образцы ДНК плазмы и периферической крови

ДНК очищали из 10 мл плазмы крови при использовании набора для выделения ДНК cfPure MAX Cell-Free DNA Extraction Kit (BioChain, номер по кат. K5011625MA), как указано производителями. ДНК из лейкоцитов периферической крови очищали при использовании набора QIAAsymphony DSP DNA Midi Kit (Qiagen, номер по кат. 937255), как указано производителем. Очищенную ДНК из всех образцов определяли количественно, как описано в другом источнике (см., например, Douville et al., 2019 bioRxiv, 660258).

Получение библиотеки

Был разработан специальный рабочий процесс получения библиотеки, который позволял эффективно выделять последовательности вводимых фрагментов ДНК и одновременно включать двухцепочечные молекулярные штрих-коды. Вкратце, библиотеки дуплексов для секвенирования были получены с внеклеточной ДНК или ДНК лейкоцитов при использовании набора Accel-NGS 2S DNA Library Kit (Swift Biosciences, номер по кат. 21024) со следующими модификациями: 1) ДНК предварительно обрабатывали тремя единицами фермента USER (New England BioLabs, номер по кат. M5505L) в течение 15 минут при 37°C для удаления урацилов; 2) отношения частиц SPRI/ПЭГ NaCl, используемых после каждой реакции, составляли 2,0×, 1,8×, 1,2× и 1,05× для Репарации концов 1, Репарации концов 2, Лигирования 1 и Лигирования 2 соответственно; 3) 50 мкМ специального 3'-адаптера (**Таблица 12**) заменяли Реагентом Y2; 4) Реагент B2 был заменен 42 мкМ специального 5'-адаптера (**Таблица 12**). Затем библиотеки подвергали ПЦР-амплификации в реакциях объемом 50 мкл при

использовании праймеров, направленных на лигированные адаптеры (**Таблица 12**). Условия реакции были следующими: 1× мастер-микс NEBNext Ultra II Q5 (New England BioLabs, номер по кат. M0544L), 2 мкМ универсального прямого праймера и 2 мкМ универсального обратного праймера (**Таблица 12**). Библиотеки амплифицировали при использовании 5, 7 или 11 циклов ПЦР, в зависимости от того, сколько экспериментов было запланировано, в соответствии со следующим протоколом: 98°C - 30 секунд, циклы 98°C - 10 секунд, 65°C - 75 секунд и выдерживание при 4°C. Если использовали 5 или 7 циклов, библиотеки амплифицировали в одиночных реакциях по 50 мкл. Если использовали 11 циклов, библиотеки делили на восемь аликвот и амплифицировали в восьми реакциях по 50 мкл, в каждую из которых добавляли дополнительные 0,5 единицы ДНК-полимеразы Q5® Hot Start высокой точности (New England BioLabs, номер по кат. M0493L), 1 мкл 10 мМ дНТФ (New England BioLabs, номер по кат. N0447L) и 0,4 мкл 25 мМ раствора MgCl₂ (New England BioLabs, номер по кат. B9021S). Продукты очищали с использованием 1,8× SPRI частиц (Beckman Coulter, номер по кат. B23317) и элюировали в буфере EB (Qiagen).

Таблица 12. Олигонуклеотиды для конструирования библиотек, ПЦР-анализов и секвенирования.

Звездочки в последовательностях праймеров обозначают фосфоротиоатные связи, другие пользовательские модификации указаны в колонке "Примечания".

Название олига	Последовательность олига	SEQ ID NO	Назначение/применение	Примечания
3'-N14 Адаптер Олиг #1	/5Phos/ATAAACGACGGCNNNNNNNNNNNNNAGATCGGAAGAGCACACGTCTGAACTCCAG*Т*С/3SpC3/	102	1/2 3'-адаптера, содержащего UID последовательность	/5Phos/ – модификация 5'-фосфата; /3SpC3/ – модификация спейсера 3' 3-углерода
3'-Адаптер Олиг #2	GCCGUCGUUUUA/33dT/	103	1/2 3'-адаптера, содержащего 3'-блок и расщепляемые остатки	/33dT/ – модифицированное основание 3'-дезокситимидин, U означает дезоксиуридин
5'-Адаптер	АСАСТСТТТСССТАСАСГАСГСТСТТССГАТСТ	104	Лигирование 5'-адаптера	
Праймер для амплификации библиотеки #1	АСАСТСТТТСССТАСАСГАСГСТСТТССГАТСТ	105	Амплификация библиотеки	
Праймер для амплификации библиотеки #2	GACTGGAGTTCAGACGTGTGCTCTTCCGATCT	106	Амплификация библиотеки	
GSP1 праймер	См. Таблицу 11	107	Первый праймер для гнездовой генспецифической ПЦР	Такой же праймер используется в обеих первых гнездовых ПЦР Уотсона и Крика
Якорный праймер Уотсона GSP1	GACTGGAGTTCAGACGT*G*Т	108	Праймер для Уотсон-специфической якорной ПЦР	
Якорный праймер Крика GSP1	АСАСТСТТТСССТАСАСГ*А*С	109	Праймер для Крик-специфической якорной ПЦР	
Праймер Уотсона GSP2	AATGATACGGCGACCACCGAGATCTACACXXXXXXXXXCGACGTA AAACGACGGCCAGT [GSP2]	110	Второй праймер для гнездовой Уотсон генспецифической ПЦР	[GSP2] означает последовательность генспецифической части (см. Таблицу 11), Xs означает индексную последовательность
Праймер Крика GSP2	CAAGCAGAAGACGGCATACGAGATXXXXXXXXXCACACAGGAAAC AGCTATGACCATG [GSP2]	111	Второй праймер для гнездовой Крик генспецифической ПЦР	[GSP2] означает последовательность генспецифической части (см. Таблицу 11), Xs означает индексную последовательность

Праймер Уотсона GSP2 якорный/индексированный	CAAGCAGAAGACGGCATAACGAGATXXXXXXXXGTGACTGGAGTT CAGACGTGTGCTCTTCCGAT*C*T	112	Второй праймер для Уотсон гнездовой якорной ПЦР и индексирования	Xs означает индексную последовательность
Праймер Крика GSP2 якорный/индексированный	AATGATACGGCGACCACCGAGATCTACACXXXXXXXXXACACTCT TTCCCTACACGACGCTCTTCCGAT*C*T	113	Второй праймер для Крик гнездовой якорной ПЦР и индексирования	Xs означает индексную последовательность
Собственный праймер для секвенирования Чтения 1	GAGATCTACACNNNNNNNCGACGTAACGACGGCCA*G*T	114	Секвенирование спаренных концов Illumina	Введение примеси в стандартный праймер для секвенирования чтения 1 Illumina
Собственный индексированный праймер для секвенирования	CATGGTCATAGCTGTTTCCCTGTG*T*G	115	Секвенирование спаренных концов Illumina	Введение примеси в стандартный индексный праймер для секвенирования Illumina
Собственный праймер для секвенирования Чтения 2	GGCATAACGAGATNNNNNNNCCACACAGGAAACAGCTATGACCA* T*G	116	Секвенирование спаренных концов Illumina	Введение примеси в стандартный индексный праймер для секвенирования чтения 2 Illumina

Конструирование библиотеки

Для решения проблемы неэффективности, связанной с созданием библиотеки, была разработана стратегия, которая включает последовательное лигирование адаптерных последовательностей с 3'- и 5'-концами фрагмента ДНК и создание двухцепочечных молекулярных штрих-кодов *in situ* (**Фиг. 22а**). После дефосфорилирования и репарации концов ДНК (**Фиг. 22а**, стадия 1) к 3'-концу фрагментов ДНК присоединяли адаптер (**Фиг. 22а**, стадия 2). Адаптер представлял собой частично двухцепочечный фрагмент ДНК с концевыми модификациями, которые были селективно лигированы с 3'-концами ДНК и предотвращали образование димера адаптера. В частности, этот адаптер состоял из одного олигонуклеотида, содержащего модификацию концевого 5'-фосфата (**Таблица 12**, 3' N14 адаптер олиг #1), который был гибридизован с другим олигонуклеотидом, содержащим 3'-блокирующую группу, и дезоксиуридины, замещенные дезокситимидинами (**Таблица 12**, 3' N14 адаптер олиг #2). Эта конструкция позволяла использовать адаптеры в высокой концентрации в реакции лигирования, что способствовало эффективному присоединению к 3'-концам без риска образования значительного димера или конкатемера. Кроме того, адаптер содержал участок из 14 случайных нуклеотидов в одном из двух олигонуклеотидов, который компрометировал одну цепь дуплексного UID. После лигирования 3'-адаптера второй адаптер (**Таблица 12**, 5'-Адаптер) лигировали с 5'-концами фрагмента ДНК посредством реакции, подобной ник-трансляции, состоящей из ДНК-полимеразы, специфичной к липким концам лигазы и урацил-ДНК-гликозилазы (**Фиг. 22а**, стадия 3). Совместное воздействие этих ферментов приводило к синтезу комплементарной цепи UID, расщеплению блокирующей части 3'-адаптера и лигированию удлиненного адаптера с 5'-концом фрагмента ДНК. Создание двухцепочечных молекулярных штрих-кодов *in situ* обеспечивает уникальное баркодирование каждого фрагмента ДНК и устраняет необходимость ферментной подготовки дуплексных адаптеров, что, как было отмечено, отрицательно влияет на выделение последовательности вводной ДНК. Наконец, лигированные с адаптером фрагменты подвергали ограниченному количеству циклов ПЦР для создания избыточных копий ("семейств" UID) двух исходных цепей ДНК (**Фиг. 22а**, стадия 4).

Влияние количества циклов амплификации библиотеки и эффективность

Количество циклов ПЦР и эффективность дубликации во время амплификации библиотеки можно регулировать с целью оптимизации параметров SaferSeqS. Поскольку SaferSeqS может включать разделение избыточных копий, получаемых с цепей Уотсона и Крика, в цепь-специфических ПЦР с целью обогащения мишенями, в предпочтительных вариантах осуществления должно быть создано необходимое количество копий, чтобы гарантировать высокую вероятность выделения дуплекса. Например, исходя из эффективности 100%, после одного цикла ПЦР каждый дуплекс ДНК-матрицы превращается в две двухцепочечные копии (одна из которых представляет каждую цепь), при этом существует лишь 25% вероятность правильного распределения этих двух копий, в результате чего одна полученная с цепи Уотсона копия разделяется на Уотсон-

специфическую ПЦР, а одна полученная с цепи Крика копия разделяется на Крик-специфическую ПЦР. Увеличение количества циклов ПЦР или повышение эффективности амплификации приводит к образованию большего количества избыточных копий, что, в свою очередь, увеличивает вероятность выделения последовательности исходного дуплекса ДНК.

Была разработана вероятностная модель для оценки количества циклов ПЦР и эффективности амплификации, требуемых для эффективного выделения дуплекса. Эта модель состояла из трех этапов: 1) моделирование количества потомства ПЦР, создаваемого во время амплификации библиотеки; 2) случайное разделение этих ПЦР-копий на реакции, специфичные к цепям Уотсона и цепям Крика; и 3) определение выделения дуплекса, то есть долю исходных дуплексов ДНК, в которых по меньшей мере одна копия, полученная с цепи Уотсона, распределена в специфичную к цепи Уотсона реакцию(и), и по меньшей мере одна копия, полученная с цепи Крика, распределена в специфичную к цепи Крика реакцию(и).

Количество ПЦР-копий исходных цепей матриц, созданных во время каждого цикла амплификации библиотеки, подчиняется биномиальному распределению. Для первого цикла ПЦР количество цепь-специфических копий было установлено равным одной. Следует отметить, что счет был установлен на одну (вместо двух), потому что первый цикл амплификации библиотеки служит просто для денатурации двух исходных цепей матрицы и их превращения в физически отдельные двухцепочечные формы. В течение последующих i -х циклов ПЦР каждая из n_i копий ПЦР может реплицироваться с вероятностью p (т.е. эффективностью амплификации), с получением в общей сложности n_{i+1} ПЦР копий, равных n_i . Бином(n_i, p). Этот процесс многократно повторяли, чтобы смоделировать количество потомства, получаемого после i циклов ПЦР. Формально общее количество образующихся копий ПЦР можно выразить следующим образом:

$$n_i = \sum_{j=1}^{i-1} \text{Бином}(n_j, p); n_1 = 1$$

После амплификации библиотеки каждый исходный дуплекс ДНК амплифицировали с образованием $n_{i, w}$ копий цепи Уотсона и $n_{i, c}$ копий цепи Крика, как описано выше. Каждая из копий $n_{i, w}$ и $n_{i, c}$ случайным образом распределяли в ПЦР-реакции, специфичные к цепи Уотсона и цепи Крика, с вероятностью q , равной доле библиотеки, используемой для каждой реакции. Когда библиотека разделена на специфичную к одной цепи Уотсона и специфичную к одной цепи Крика ПЦР, q равно 50%. Если библиотека разделена на две ПЦР, специфичных к цепи Уотсона и Крика, q равно 25%. Количество копий ПЦР, которые разделены на соответствующие цепь-специфические ПЦР ($N_{k, w}$ или $N_{k, c}$ для k -й Уотсон или Крик специфической ПЦР, соответственно), получают из биномиального распределения с n_i, W или n_i, C "исследований" и вероятности q "успеха" для копий Уотсона и Крика соответственно. Таким образом, вероятность разделения по меньшей мере одной копии ПЦР, полученной с

цепи Уотсона, на k -ю реакцию ПЦР, специфичную к цепи Уотсона, составляет:

$$P(N_{k,W} > 0) = 1 - (1 - q)^{n_{i,w}}$$

Аналогичным образом, вероятность разделения по меньшей мере одной копии ПЦР, полученной с цепи Крика, на k -ю Крик-специфическую реакцию ПЦР составляет:

$$P(N_{k,C} > 0) = 1 - (1 - q)^{n_{i,c}}$$

Обе цепи исходного дуплекса ДНК могут быть выделены только в том случае, если $N_{k,W}$ и $N_{k,C}$ больше нуля. Поскольку разделение потомства ПЦР является независимым, вероятностное выделение дуплекса прогнозируется следующим:

$$P(N_{k,W} > 0, N_{k,C} > 0) = [1 - (1 - q)^{n_{i,w}}][1 - (1 - q)^{n_{i,c}}]$$

Авторы варьировали эффективность ПЦР от 100% до 50%, количество циклов амплификации библиотеки от 1 до 11 и долю библиотеки, используемой для каждой реакции, от 50% до 1,4%. Для каждого условия провели 10000 симуляций вышеописанного процесса и представили среднее выделение дуплекса на **Фиг. 28**.

Размер и выделение фрагмента с помощью якорной полугнездовой ПЦР

Якорная полугнездовая ПЦР теоретически демонстрирует более высокое выделение последовательности молекул-матриц, чем традиционная ПЦР с ампликоном. В традиционной ПЦР с ампликоном молекула-матрица должна содержать сайты связывания как прямых, так и обратных праймеров, а также промежуточную последовательность, которая определяет ампликон. Напротив, в якорной полугнездовой ПЦР молекулы-матрицы должны одновременно нести два сайта связывания ген-специфических праймеров, чтобы можно было выделить их последовательности. Общая длина участков связывания ген-специфических праймеров, используемых в SaferSeqS, составляет примерно 30 п.н., тогда как длина ампликонов, используемых в SafeSeqS для профилирования вкДНК, обычно составляет 70-80 п.н. Формально, с учетом равномерно случайных координат начала/конца фрагмента, вероятность выделения последовательности молекулы-матрицы длины L равна $\frac{L-r}{L}$, где r - длина ампликона в случае традиционной ПЦР или длина комбинированного участка связывания ген-специфических праймеров в случае якорной полугнездовой ПЦР. Таким образом, для бесклеточных фрагментов ДНК размером ~167 п.н. якорная полугнездовая ПЦР теоретически может выделять на ~25% больше исходных фрагментов матриц, чем традиционная ПЦР с ампликоном. Кроме того, в отличие от традиционной ПЦР с ампликонами, которая дает заданные размеры продукта, обусловленные положениями прямого и обратного праймеров, якорная полугнездовая ПЦР дает фрагменты различной длины, причем только один из концов фрагмента определен положениями ген-специфических праймеров. Исходя из того, что молекулы-матрицы длиной L имеют

равномерно случайные координаты начала/конца, наблюдаемая длина фрагмента после якорной полугнездовой ПЦР будет равна $\frac{L-r}{2}$, где r представляет собой длину комбинированного участка связывания ген-специфических праймеров.

Пример осуществления биоинформационного конвейера SaferSeqS

В примере осуществления биоинформационного конвейера SaferSeqS, чтения Уотсона и Крика для каждого образца были сведены в один BAM файл и отсортированы по имени чтения с использованием SAMtools, чтобы можно было легко извлечь пары спаривания. Пользовательские сценарии в Python использовали для последующей реконструкции семейств дуплексов и идентификации супермутантов Уотсона, супермутантов Крика и суперкалимутантов.

Сначала чтения группировали в семейства UID, отмечая при этом, какие чтения были получены с цепи Уотсона и Крика, изучая значение их побитового флага (т.е. поля FLAG). Чтения, содержащие побитовые пофлаговые значения 99 и 147, получены с цепи Уотсона, а чтения, которые содержат побитовые флаги 83 и 163, получены с цепи Крика. Чтения с любыми другими значениями побитового флага исключались из последующего анализа. Побитовые флаги - это числовые значения, которые присваиваются парам чтений во время картирования. Их значение указывает на то, как спаренные чтения выравниваются в геноме относительно друг друга. Например, если чтение картировано на референсной цепи, а его комплемент картирован на обратной (комплементарной) цепи, то такая пара считывания была получена с цепи Уотсона. Аналогичным образом, если считывание картировано на обратной (комплементарной) цепи, а его комплемент картирован на референсной цепи, то такая пара чтений была получена с цепи Крика.

Во-вторых, при группировании семейств UID вводили два дополнительных критерия контроля качества с целью улучшения определения эндогенного молекулярного штрих-кода (т.е. координаты конца фрагмента): 1) чтения с мягким усечением на 5'- или 3'-концах фрагмента были исключены, 2) считывания должны были содержать ожидаемую последовательность постоянной метки (GCCGTCGTTTAT, SEQ ID NO: 117) сразу после экзогенного UID не более чем с одним некомплементарным основанием.

В-третьих, поскольку в этом примере количество возможных экзогенных UID последовательностей значительно превышает количество исходных молекул-матриц, "коллизии штрих-кодов", в которых две молекулы имеют одну и ту же экзогенную UID последовательность, но разные эндогенные UID, должны быть чрезвычайно редкими. В частности, ожидаемое количество коллизий штрих-кодов может быть вычислено из классического "парадокса дней рождения" и составляет:

$$E[X] = n \left\{ 1 - \left(1 - \frac{1}{N} \right)^{n-1} \right\}$$

где n равно количеству молекул матрицы, а N равно количеству возможных штрих-кодов. Для экзогенной UID последовательности длиной 14 п.н. (содержащей в общей сложности 268435456 возможных последовательностей) и 10000 эквивалентов генома

ожидаемое количество коллизий составляет 0,37 или 0,0037% от вводных данных. Поэтому для данного примера требовалось, чтобы каждая экзогенная UID последовательность могла быть связана только с одним эндогенным UID. В случаях, когда экзогенный UID был связан больше чем с одним эндогенным UID, сохраняли самое большое семейство, а все остальные отбрасывали.

Следует отметить, что в других параметрах схемы эксперимента могут использоваться не уникальные экзогенные UID, причем не уникальные экзогенные UID могут использоваться в комбинации с эндогенными UID для отнесения к семействам UID.

Наконец, поскольку сами по себе экзогенные штрих-коды подвержены ошибкам ПЦР и секвенирования, авторы исправили ошибки в последовательностях UID и перегруппировали семейства UID, используя метод сетевой смежности UMI-tools.

После сборки чтениц в семейства UID, супермутанты Уотсона, супермутанты Крик и суперкалимутанты были обозначены, как описано в других частях настоящего документа. Чтобы исключить распространенные полиморфизмы, были исключены все мутации в базе данных с агрегированной информацией по геномам (gnomeAD), присутствующие с частотой аллеля больше 0,1%. Чтения, включающие суперкалимутанты, подвергали окончательной проверке вручную для исключения возможных артефактов выравнивания.

Оценки частот неклональных соматических мутаций

ДНК, используемая в данном исследовании, была получена от группы лиц средним возрастом 30 лет. В результате ожидаемая частота неклональных соматических замен одиночных оснований в этих образцах составила 426 на диплоидный геном или примерно 7×10^{-8} мутаций на п.н. В данном исследовании авторы с помощью SaferSeqS оценивали в общей сложности 42695395 оснований из ДНК, полученной от здоровых контрольных субъектов. Среди этих 42695395 оснований было обнаружено 5 суперкалимутантов с заменой одного основания, что соответствует частоте мутаций 12×10^{-8} . Для определения, соответствует ли частота наблюдаемых суперкалимутантов предыдущим оценкам частот неклональных соматических мутаций в нормальных клетках крови, было вычислено следующее точное одностороннее биномиальное р-значение:

$$P(X \geq 5) = 1 - \sum_{k=0}^4 \binom{42,695,395}{k} (7 * 10^{-8})^k (1 - 7 * 10^{-8})^{42,695,395 - k} = 0,19$$

Таким образом, статистически значимое различие между количеством наблюдаемых суперкалимутантов и предсказанным количеством ассоциированных с возрастом неклональных соматических мутаций, происходящих из здоровых гемопоэтических стволовых клеток.

Якорная полугнездовая ПЦР

Обогащение мишенями представляющих интерес областей было выполнено с помощью критических модификаций якорной полугнездовой ПЦР, необходимых для дуплексного секвенирования. Во время разработки этого специализированного цепь-

специфического теста были оптимизированы различные условия реакции, включая количество циклов, концентрации праймеров и состав полимеразы. Окончательный оптимизированный протокол был следующим: первый раунд ПЦР проводили в реакционной смеси объемом 50 мкл при следующих условиях: 1× мастер-микс NEBNext Ultra II Q5 (New England BioLabs, номер по кат. M0544L), 2 мкМ праймера GSP1 и 2 мкМ короткого якорного праймера P7 для амплификации цепи Уотсона. Праймер GSP1 был специфичен к каждому ампликону, а короткий якорный праймер P7 использовали в качестве якорного праймера для цепи Уотсона всех ампликонов (**Таблицы 11 и 12**). Цепь Крика амплифицировали таким же путем в отдельной лунке, за исключением того, что короткий якорный праймер праймер P5 использовали вместо короткого праймера P7. Следует отметить, что праймер GSP1, используемый для амплификации цепи Уотсона, был идентичен праймеру GSP1, используемому для цепи Крика; единственное различие между ПЦР с цепями Уотсона и Крика заключалось в якорных праймерах. Обе реакции (с цепями Уотсона и Крика) амплифицировали с использованием 19 циклов в соответствии с протоколом амплификации, описанным выше.

Для цепи Уотсона был сформирован второй цикл ПЦР в реакциях объемом 50 мкл при использовании идентичных условий реакции, используемых для первого цикла ПЦР. Различия заключались в следующем: (i) Матрица: в качестве матрицы использовали 1% продукта первой якорной ПЦР с цепью Уотсона (вместо библиотеки, используемой в качестве матрицы для первой ПЦР), и (ii) Праймеры: ген-специфические праймеры GSP2 использовали вместо ген-специфических праймеров GSP1, и якорный индексирующий праймер P5 использовали вместо короткого якорного праймера P7. Второй раунд ПЦР для цепи Крика проводили идентично, за исключением следующего: (i) Матрица: первую ПЦР с цепью Крика использовали в качестве матрицы и (ii) Праймеры: индексирующий якорный праймер P7 использовали вместо индексирующего якорного праймера P5. Обе реакции (с цепями Уотсона и Крика) амплифицировали с использованием 17 циклов в соответствии с протоколом амплификации, описанным ранее. Последовательности праймеров, используемых для второго раунда ПЦР, приведены в **Таблице 12**. Продукты второго раунда ПЦР объединяли и очищали с использованием 1,8× частиц SPRI перед секвенированием.

Для экспериментов, в которых одновременно амплифицировали множество мишеней в одной реакции, условия ПЦР были идентичны описанным выше, за исключением следующего: (i) каждый ген-специфический праймер был включен в конечной концентрации 0,25 мкМ, и (ii) концентрация якорного праймера был включена в конечной концентрации 0,25 мкМ на мишень (например, в конечной концентрации 6,25 мкМ, если совместно амплифицировали 25 мишеней).

Секвенирование

Концентрации библиотек определяли при использовании набора KAPA Library Quantification Kit (KAPA Biosystems, номер по кат. KK4824), как описано производителем. Секвенирование проводили с использованием 2×75 чтений спаренных концов с двойной

индексацией по восьми основаниям на приборе Illumina MiSeq. Контрольную библиотеку PhiX с двойным индексом (номер по кат. SeqMatic TM-502-ND) добавляли к 25% от общего количества матриц, чтобы обеспечить разнообразие оснований во всех циклах. Пользовательские праймеры для секвенирования чтения 1, индекса и чтения 2 (**Таблица 12**) комбинировали со стандартными праймерами для секвенирования Illumina в конечной концентрации 1 мкМ.

Опрос мутаций и процесс анализа SaferSeqS

Анализ данных SafeSeqS проводили, как описано в другом источнике (см., например, Kinde et al., 2011 Proc Natl Acad Sci U S A 108:9530-9535), при использовании пользовательских скриптов на Python. Чтения секвенирования подвергали первоначальной обработке путем извлечения первых 14 нуклеотидов в качестве UID последовательности и маскирования адаптерных последовательностей при использовании инструмента IlluminaBasecallsToSam в пакете Picard (broadinstitute.github.io/picard). Затем чтения сопоставляли с референсным геномом hg19 при использовании BWA-MEM (версии 0.7.17) и сортировали по UID последовательности с помощью SAMtools. Семейства UID оценивали, если они состояли из 2 или более чтений и если >90% чтений были картированы в референсном геноме с ожидаемыми последовательностями праймеров. "Супермутанты" были идентифицированы как мутации, которые присутствовали в >95% картированных чтений и имели средний балл Phred больше 25.

Для анализа SaferSeqS был разработан специальный порядок анализа. Вкратце, чтения демультимплексовали, и цепь, с которой были получены чтения, идентифицировали при использовании индексных последовательностей. Для ясности и краткости, чтения, полученные с цепи Уотсона, называются "чтениями Уотсона", а чтения, полученные с цепи Крика, называются "чтениями Крика". Для чтений Уотсона первые 14 оснований чтения 1 были выделены в качестве UID последовательности. Поскольку ориентация вставки для цепи Крика является обратной, первые 14 оснований чтения 2 были выделены в качестве UID последовательности для чтения Крика. Адаптерные последовательности маскировали с помощью инструмента IlluminaBasecallsToSam в пакете Picard (broadinstitute.github.io/picard), а полученные в результате матрица-специфичные части чтений картировали на референсном геноме hg19 при использовании BWA-MEM (версии 0.7.17). После выравнивания картированные чтения Уотсона и Крика объединяли и сортировали с помощью SAMtools.

Скрипты на Python использовали для последующей реконструкции семейств дуплексов и идентификации супермутантов Уотсона, супермутантов Крика и суперкалимутантов. После коррекции ошибок ПЦР и секвенирования в последовательностях молекулярных штрих-кодов, как описано в другом источнике (см., например, Smith et al., 2017 Genome Res 27:491-499), чтения Уотсона и Крика, относящиеся к одному и тому же семейству дуплексов, группировали вместе для реконструкции последовательность исходной молекулы-матрицы. Чтобы исключить артефакты, возникающие на этапе конечной репарации при конструировании библиотеки,

основания на расстоянии меньше 10 оснований от 3'-адаптерной последовательности не учитывались для анализа мутаций. Супермутанты Уотсона и Крика были определены как мутации, присутствующие в >80% чтений Уотсона или Крика семейства дуплексов, соответственно. Суперкалимутанты были определены как мутации, присутствующие больше чем в 80% семейств Уотсона и Крика с одинаковым UID.

Статистический анализ

Непрерывные переменные были представлены как медианные значения и диапазон, тогда как категориальные переменные были представлены в виде целых чисел и процентов. Все статистические тесты проводили при использовании пакета R stats (версии 3.5.1).

Эти результаты демонстрируют, что SaferSeqS способен обнаруживать редкие мутации с чрезвычайно высокой специфичностью. Данный метод хорошо масштабируется, экономически эффективен и поддается автоматизации с высокой производительностью. SaferSeqS достигал 5-75-кратного улучшения выделения вводных данных по сравнению с существующими методами дуплексного секвенирования и может применяться к ограниченным количествам исходного материала, а также приводил к более чем 50-кратному улучшению коррекции ошибок по сравнению со стандартными методами на основе ПЦР с использованием молекулярных штрих-кодов (Фиг. 23, табл. 8). Он также обеспечивает >50-кратное улучшение коррекции ошибок по сравнению с оптимальными методами на основе лигирования, в которых используются только супермутанты Уотсона или Крика, но не суперкалимутанты (Фиг. 26, таблица 9). Оба сокращения полезны для обнаружения мутаций, присутствующих в одной копии или очень малом количестве копий, например, при скрининге рака и в условиях минимального остаточного заболевания. Наконец, поскольку он включает дуплексное секвенирование, SaferSeqS значительно более чувствителен, чем цифровая капельная ПЦР, при анализе одиночных ампликонов и, в отличие от цифровой капельной ПЦР, может быть в высокой степени мультиплексирован.

ДРУГИЕ ВАРИАНТЫ ОСУЩЕСТВЛЕНИЯ

Следует понимать, что хотя изобретение было описано в сочетании с его подробным описанием, предшествующее описание предназначено для иллюстрации, а не для ограничения объема изобретения, который определяется объемом прилагаемой формулы изобретения. Другие аспекты, преимущества и модификации входят в объем следующей формулы изобретения.

ВКЛЮЧЕНИЕ ПОСРЕДСТВОМ ОТСЫЛКИ

Все ссылки, выданные патенты и заявки на патенты, цитируемые в тексте настоящего описания, настоящим включены посредством отсылки во всей своей полноте во всех отношениях.

Таблица 8. Сравнение мутаций, обнаруженных с помощью SafeSeqS и SaferSeqS.

*Координаты относятся к выпуску референсного генома человека hg19 (Консорциум референсного генома GRCh37, февраль 2009 года).

**Номер мутации соответствует указанному на Фиг. 25.

NA, положение, не подвергнутое оценке с помощью SafeSeqS или SaferSeqS.

Известная мутация	Тип образца	Номер мутации*	Хром.	Положение*	Референсное основание	Вариантное основание	Ген	Замена в кДНК	Замена в Белке	#Супермутантов (SafeSeqS)	Всего UID (SaferSeqS)	SafeSeqS MAF	#Суперкаллимутантов (SaferSeqS)	Глубина дуплекса (SaferSeqS)	SaferSeqS MAF
TP53 p.L264fs	вДНК, высокое подмешивание	1	хр.17	7,577,132	C	A	TP53	c.808G>T	p.S268I	1	601,826	1.66E-06	NA	NA	NA
TP53 p.L264fs	вДНК, контроль дилого типа	1	хр.17	7,577,132	C	A	TP53	c.808G>T	p.S268I	1	1,53E-06	NA	NA	NA	NA
TP53 p.L264fs	вДНК, высокое подмешивание	2	хр.17	7,577,132	C	T	TP53	c.808G>A	p.S268N	1	601,826	1.66E-06	NA	NA	NA
TP53 p.L264fs	вДНК, низкое подмешивание	2	хр.17	7,577,132	C	T	TP53	c.808G>A	p.S268N	1	539,498	1.85E-06	NA	NA	NA
TP53 p.L264fs	вДНК, контроль дилого типа	2	хр.17	7,577,132	C	T	TP53	c.808G>A	p.S268N	1	655,959	1.53E-06	NA	NA	NA
TP53 p.L264fs	вДНК, контроль дилого типа	3	хр.17	7,577,133	T	C	TP53	c.805A>G	p.S269G	1	655,959	1.53E-06	NA	NA	NA
TP53 p.L264fs	вДНК, контроль дилого типа	4	хр.17	7,577,133	T	G	TP53	c.805A>C	p.S268R	1	655,959	1.53E-06	NA	NA	NA
TP53 p.L264fs	вДНК, высокое подмешивание	5	хр.17	7,577,134	G	A	TP53	c.804C>T	p.N268N	1	601,826	1.66E-06	NA	NA	NA
TP53 p.L264fs	вДНК, высокое подмешивание	6	хр.17	7,577,134	G	C	TP53	c.804C>G	p.N268K	1	601,826	1.66E-06	NA	NA	NA
TP53 p.L264fs	вДНК, высокое подмешивание	7	хр.17	7,577,134	G	T	TP53	c.804C>A	p.N268K	4	601,826	6.65E-06	NA	NA	NA
TP53 p.L264fs	вДНК, контроль дилого типа	7	хр.17	7,577,134	G	T	TP53	c.804C>A	p.N268K	1	655,959	1.53E-06	NA	NA	NA
TP53 p.L264fs	вДНК, высокое подмешивание	8	хр.17	7,577,135	T	C	TP53	c.803A>G	p.N268S	1	601,826	1.66E-06	NA	NA	NA
TP53 p.L264fs	вДНК, низкое подмешивание	9	хр.17	7,577,136	T	C	TP53	c.802A>G	p.N268D	2	539,498	3.71E-06	NA	NA	NA
TP53 p.L264fs	вДНК, контроль дилого типа	9	хр.17	7,577,136	T	C	TP53	c.802A>G	p.N268D	3	655,959	4.58E-06	NA	NA	NA
TP53 p.L264fs	вДНК, высокое подмешивание	10	хр.17	7,577,137	C	A	TP53	c.801G>T	p.R267R	1	601,826	1.66E-06	NA	NA	NA
TP53 p.L264fs	вДНК, высокое подмешивание	11	хр.17	7,577,137	C	T	TP53	c.801G>A	p.R267R	3	601,826	4.98E-06	NA	NA	NA
TP53 p.L264fs	вДНК, контроль дилого типа	11	хр.17	7,577,137	C	T	TP53	c.801G>A	p.R267R	4	655,959	6.10E-06	NA	NA	NA
TP53 p.L264fs	вДНК, высокое подмешивание	12	хр.17	7,577,138	C	A	TP53	c.800G>T	p.R267L	1	601,826	1.66E-06	0	7,081	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	12	хр.17	7,577,138	C	A	TP53	c.800G>T	p.R267L	1	655,959	1.53E-06	0	7,305	0.00E+00
TP53 p.L264fs	вДНК, высокое подмешивание	13	хр.17	7,577,138	C	T	TP53	c.800G>A	p.R267Q	12	601,826	1.96E-05	0	7,081	0.00E+00
TP53 p.L264fs	вДНК, низкое подмешивание	13	хр.17	7,577,138	C	T	TP53	c.800G>A	p.R267Q	12	539,498	2.22E-06	0	7,519	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	13	хр.17	7,577,138	C	T	TP53	c.800G>A	p.R267Q	26	655,959	3.97E-06	0	7,305	0.00E+00
TP53 p.L264fs	вДНК, высокое подмешивание	14	хр.17	7,577,139	G	A	TP53	c.799C>T	p.R267W	53	601,826	8.81E-05	0	7,005	0.00E+00
TP53 p.L264fs	вДНК, низкое подмешивание	14	хр.17	7,577,139	G	A	TP53	c.799C>T	p.R267W	2	539,498	3.71E-06	0	7,500	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	14	хр.17	7,577,139	G	A	TP53	c.799C>T	p.R267W	53	655,959	8.08E-05	0	7,282	0.00E+00
TP53 p.L264fs	вДНК, низкое подмешивание	15	хр.17	7,577,139	G	T	TP53	c.799C>A	p.R267R	1	539,498	1.85E-06	0	7,500	0.00E+00
TP53 p.L264fs	вДНК, высокое подмешивание	16	хр.17	7,577,140	T	C	TP53	c.798A>G	p.G266G	1	601,826	1.66E-06	0	7,025	0.00E+00
TP53 p.L264fs	вДНК, низкое подмешивание	17	хр.17	7,577,141	C	A	TP53	c.797G>T	p.G266V	1	539,498	1.85E-06	0	7,426	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	17	хр.17	7,577,141	C	A	TP53	c.797G>T	p.G266V	1	655,959	1.53E-06	0	7,187	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	18	хр.17	7,577,141	C	G	TP53	c.797G>C	p.G266A	1	655,959	1.53E-06	0	7,187	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	19	хр.17	7,577,141	C	T	TP53	c.797G>A	p.G266E	1	655,959	1.53E-06	0	7,187	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	20	хр.17	7,577,142	C	A	TP53	c.796G>T	p.G269*	1	655,959	1.53E-06	0	7,159	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	21	хр.17	7,577,142	C	A	TP53	c.796G>T	p.G269*	1	655,959	1.53E-06	0	7,159	0.00E+00
TP53 p.L264fs	вДНК, высокое подмешивание	22	хр.17	7,577,143	C	A	TP53	c.795G>T	p.L266L	1	601,826	1.66E-06	0	6,923	0.00E+00
TP53 p.L264fs	вДНК, низкое подмешивание	23	хр.17	7,577,143	C	T	TP53	c.795G>A	p.L265L	3	539,498	5.58E-06	0	7,373	0.00E+00
TP53 p.L264fs	вДНК, низкое подмешивание	24	хр.17	7,577,144	A	G	TP53	c.794T>C	p.L265P	1	539,498	1.85E-06	0	7,352	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	24	хр.17	7,577,144	A	G	TP53	c.794T>C	p.L265P	2	655,959	3.05E-06	0	7,106	0.00E+00
TP53 p.L264fs	вДНК, высокое подмешивание	25	хр.17	7,577,144	A	T	TP53	c.794T>A	p.L265Q	1	601,826	1.66E-06	0	6,849	0.00E+00
TP53 p.L264fs	вДНК, высокое подмешивание	26	хр.17	7,577,144	AGT	-	TP53	c.792_794del	p.L264fs	4,307	601,826	6.28E-03	43	6,849	6.28E-03
TP53 p.L264fs	вДНК, низкое подмешивание	26	хр.17	7,577,144	AGT	-	TP53	c.792_794del	p.L264fs	398	539,498	7.38E-04	1	7,352	1.39E-04
TP53 p.L264fs	вДНК, высокое подмешивание	27	хр.17	7,577,145	G	A	TP53	c.793C>T	p.L265L	1	601,826	1.66E-06	0	6,775	0.00E+00
TP53 p.L264fs	вДНК, низкое подмешивание	27	хр.17	7,577,145	G	A	TP53	c.793C>T	p.L265L	2	539,498	3.71E-06	0	7,270	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	27	хр.17	7,577,145	G	A	TP53	c.793C>T	p.L265L	1	655,959	1.53E-06	0	7,071	0.00E+00
TP53 p.L264fs	вДНК, низкое подмешивание	28	хр.17	7,577,147	A	G	TP53	c.791T>C	p.L264W	1	539,498	1.85E-06	0	7,205	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	29	хр.17	7,577,147	AG	-	TP53	c.790_791del	p.L243fs	12	655,959	1.85E-06	0	6,985	0.00E+00
TP53 p.L264fs	вДНК, высокое подмешивание	30	хр.17	7,577,148	G	T	TP53	c.790C>A	p.L264I	1	601,826	1.66E-06	0	6,740	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	30	хр.17	7,577,148	G	T	TP53	c.790C>A	p.L264I	1	655,959	1.53E-06	0	6,923	0.00E+00
TP53 p.L264fs	вДНК, низкое подмешивание	31	хр.17	7,577,151	T	C	TP53	c.787A>G	p.N263D	1	539,498	1.85E-06	0	7,110	0.00E+00
TP53 p.L264fs	вДНК, высокое подмешивание	32	хр.17	7,577,152	A	G	TP53	c.788T>C	p.G262G	1	601,826	1.66E-06	0	6,823	0.00E+00
TP53 p.L264fs	вДНК, низкое подмешивание	32	хр.17	7,577,152	A	G	TP53	c.788T>C	p.G262G	1	539,498	1.85E-06	0	7,042	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	32	хр.17	7,577,152	A	G	TP53	c.788T>C	p.G262G	6	655,959	8.16E-06	0	6,796	0.00E+00
TP53 p.L264fs	вДНК, низкое подмешивание	33	хр.17	7,577,153	C	A	TP53	c.786G>T	p.G262V	1	539,498	1.85E-06	0	6,985	0.00E+00
TP53 p.L264fs	вДНК, высокое подмешивание	34	хр.17	7,577,153	C	T	TP53	c.785G>A	p.G262D	1	601,826	1.66E-06	0	6,562	0.00E+00
TP53 p.L264fs	вДНК, низкое подмешивание	34	хр.17	7,577,153	C	T	TP53	c.785G>A	p.G262D	1	539,498	1.85E-06	0	6,985	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	34	хр.17	7,577,153	C	T	TP53	c.785G>A	p.G262D	1	655,959	1.53E-06	0	6,780	0.00E+00
TP53 p.L264fs	вДНК, высокое подмешивание	35	хр.17	7,577,154	C	A	TP53	c.784G>T	p.G262C	2	601,826	1.66E-06	0	6,539	0.00E+00
TP53 p.L264fs	вДНК, низкое подмешивание	35	хр.17	7,577,154	C	A	TP53	c.784G>T	p.G262C	13	539,498	2.41E-05	0	6,984	0.00E+00
TP53 p.L264fs	вДНК, высокое подмешивание	36	хр.17	7,577,154	C	T	TP53	c.784G>A	p.G262S	1	601,826	1.66E-06	0	6,539	0.00E+00
TP53 p.L264fs	вДНК, низкое подмешивание	36	хр.17	7,577,154	C	T	TP53	c.784G>A	p.G262S	2	539,498	3.71E-06	0	6,964	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	37	хр.17	7,577,155	A	G	TP53	c.783T>C	p.S261S	2	655,959	3.05E-06	0	6,680	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	38	хр.17	7,577,156	C	T	TP53	В интроне	В интроне	1	655,959	1.53E-06	0	6,854	0.00E+00
TP53 p.L264fs	вДНК, высокое подмешивание	38	хр.17	7,577,156	A	G	TP53	В интроне	В интроне	1	601,826	1.66E-06	0	6,310	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	40	хр.17	7,577,159	C	A	TP53	В интроне	В интроне	1	655,959	1.53E-06	0	6,547	0.00E+00
TP53 p.L264fs	вДНК, высокое подмешивание	41	хр.17	7,577,169	C	T	TP53	В интроне	В интроне	1	601,826	1.66E-06	0	6,281	0.00E+00
TP53 p.L264fs	вДНК, низкое подмешивание	42	хр.17	7,577,161	C	G	TP53	В интроне	В интроне	5	539,498	8.27E-06	0	6,205	0.00E+00
TP53 p.L264fs	вДНК, высокое подмешивание	45	хр.17	7,577,162	A	G	TP53	В интроне	В интроне	3	601,826	4.98E-06	0	6,092	0.00E+00
TP53 p.L264fs	вДНК, низкое подмешивание	43	хр.17	7,577,162	A	G	TP53	В интроне	В интроне	1	539,498	1.85E-06	0	6,505	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	44	хр.17	7,577,162	A	G	TP53	В интроне	В интроне	1	655,959	1.53E-06	0	6,334	0.00E+00
TP53 p.L264fs	вДНК, низкое подмешивание	46	хр.17	7,577,163	G	A	TP53	В интроне	В интроне	1	539,498	1.85E-06	0	6,491	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	45	хр.17	7,577,163	G	A	TP53	В интроне	В интроне	1	655,959	1.53E-06	0	6,324	0.00E+00
TP53 p.L264fs	вДНК, низкое подмешивание	47	хр.17	7,577,163	G	C	TP53	В интроне	В интроне	1	539,498	1.85E-06	0	6,491	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	46	хр.17	7,577,164	G	A	TP53	В интроне	В интроне	1	655,959	1.53E-06	0	6,317	0.00E+00
TP53 p.L264fs	вДНК, контроль дилого типа	48	хр.17	7,577,164	G	C	TP53	В интроне	В интроне	1	655,959	1.53E-06	0	6,317	0.00E+00
TP53 p.L264fs	вДНК, высокое подмешивание	48													

Известная мутация	Тип образца	Номер мутации**	Хром.	Положение*	Референсное основание	Вариантное основание	Ген	Замена в кДНК	Замена в белке	#Супермутантов (SafeSeq3)	Всего UID (SafeSeq3)	SafeSeq3 MAF	#Суперкаллимутантов (SaferSeq3)	Глубина дуплекса (SaferSeq3)	SaferSeq3 MAF
TRP3 p.R190L	вДНК, низкое подмешивание	NA	xp.17	7,577,220	A	-	TP53	В интроне	В интроне	NA	NA	NA	1	1,526	6.55E-04
TRP3 p.R190L	вДНК, контроль дикого типа	1	xp.17	7,578,264	A	A	TP53	c.584_S85fsT	p.M95fs	1	497,273	2.01 E-06	0	7,182	0.00E+00
TRP3 p.R190L	вДНК, низкое подмешивание	2	xp.17	7,578,264	G	A	TP53	c.585C>T	p.R1195I	4	436,466	9.16E-06	0	8,712	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	2	xp.17	7,578,264	G	A	TP53	c.585C>T	p.R1195I	24	497,273	4.83E-05	0	7,182	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	3	xp.17	7,578,265	A	G	TP53	c.584T>C	p.R1195T	1	497,273	2.01 E-06	0	7,185	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	4	xp.17	7,578,266	T	C	TP53	c.583A>G	p.R1196V	1	497,273	2.01 E-06	0	7,088	0.00E+00
TRP3 p.R190L	вДНК, низкое подмешивание	5	xp.17	7,578,267	A	-	TP53	c.582del	p.L194fs	2	436,466	4.58E-06	0	8,478	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	6	xp.17	7,578,267	A	G	TP53	c.582T>C	p.L194L	1	502,574	1.98E-06	0	8,470	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	6	xp.17	7,578,267	A	G	TP53	c.582T>C	p.L194L	3	497,273	6.03E-06	0	7,052	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	7	xp.17	7,578,267	A	T	TP53	c.582T>A	p.L194L	1	502,574	1.98E-06	0	8,470	0.00E+00
TRP3 p.R190L	вДНК, низкое подмешивание	8	xp.17	7,578,268	A	G	TP53	c.581T>C	p.L194P	2	436,466	4.58E-06	0	8,447	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	8	xp.17	7,578,268	A	G	TP53	c.581T>C	p.L194P	1	497,273	2.01 E-06	0	7,031	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	9	xp.17	7,578,268	A	T	TP53	c.581T>A	p.L194H	1	502,574	1.98E-06	0	8,440	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	10	xp.17	7,578,269	G	A	TP53	c.580C>T	p.L194F	1	502,574	1.98E-06	0	8,238	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	11	xp.17	7,578,269	G	T	TP53	c.580C>A	p.L194I	10	502,574	1.98E-06	0	8,239	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	12	xp.17	7,578,270	A	G	TP53	c.578T>C	p.H193H	1	497,273	2.01 E-06	0	8,815	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	13	xp.17	7,578,272	G	A	TP53	c.577C>T	p.H193Y	1	502,574	1.98E-06	0	8,134	0.00E+00
TRP3 p.R190L	вДНК, низкое подмешивание	13	xp.17	7,578,272	G	A	TP53	c.577C>T	p.H193Y	2	436,466	4.58E-06	0	8,175	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	14	xp.17	7,578,272	G	T	TP53	c.577C>A	p.H193N	3	502,574	5.97E-06	0	8,134	0.00E+00
TRP3 p.R190L	вДНК, низкое подмешивание	14	xp.17	7,578,272	G	T	TP53	c.577C>A	p.H193N	3	436,466	6.87E-06	0	8,175	0.00E+00
TRP3 p.R190L	вДНК, низкое подмешивание	15	xp.17	7,578,273	C	A	TP53	c.579G>T	p.Q192H	1	436,466	2.29E-06	0	8,152	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	16	xp.17	7,578,273	C	T	TP53	c.579G>A	p.Q192Q	1	502,574	1.98E-06	0	8,100	0.00E+00
TRP3 p.R190L	вДНК, низкое подмешивание	16	xp.17	7,578,273	C	T	TP53	c.579G>A	p.Q192Q	17	436,466	3.89E-05	0	8,152	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	16	xp.17	7,578,273	C	T	TP53	c.579G>A	p.Q192Q	4	497,273	8.04E-06	0	8,733	0.00E+00
TRP3 p.R190L	вДНК, низкое подмешивание	17	xp.17	7,578,274	T	C	TP53	c.575A>G	p.Q192R	1	436,466	2.29E-06	0	8,125	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	17	xp.17	7,578,274	T	C	TP53	c.575A>G	p.Q192R	1	497,273	2.01 E-06	0	8,702	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	18	xp.17	7,578,274	T	G	TP53	c.575A>C	p.Q192P	1	502,574	1.98E-06	0	8,063	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	18	xp.17	7,578,275	G	T	TP53	c.575G>A	p.Q192K	20	502,574	7.97E-05	0	8,030	0.00E+00
TRP3 p.R190L	вДНК, низкое подмешивание	19	xp.17	7,578,275	G	T	TP53	c.574C>A	p.Q192K	1	436,466	2.29E-06	0	8,000	0.00E+00
TRP3 p.R190L	вДНК, низкое подмешивание	20	xp.17	7,578,275	GAG	-	TP53	c.572_574del	p.P181fs	32	436,466	7.33E-05	0	8,026	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	21	xp.17	7,578,276	A	C	TP53	c.573T>G	p.P191P	7	497,273	1.41E-05	0	8,408	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	22	xp.17	7,578,276	A	G	TP53	c.573T>C	p.P191P	1	502,574	1.98E-06	0	7,811	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	23	xp.17	7,578,277	G	T	TP53	c.572C>A	p.P191H	1	497,273	2.01 E-06	0	8,375	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	24	xp.17	7,578,278	G	A	TP53	c.571C>T	p.P191S	7	502,574	1.39E-05	0	7,738	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	24	xp.17	7,578,278	G	A	TP53	c.571C>T	p.P191S	1	497,273	2.01 E-06	0	8,347	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	25	xp.17	7,578,278	G	T	TP53	c.571C>A	p.P191T	1	502,574	1.98E-06	0	7,738	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	26	xp.17	7,578,280	G	A	TP53	c.569C>T	p.P190L	3,780	502,574	7.52E-03	82	7,704	8.08E-03
TRP3 p.R190L	вДНК, низкое подмешивание	26	xp.17	7,578,280	G	A	TP53	c.569C>T	p.P190L	269	436,466	6.62E-04	7	7,715	8.07E-04
TRP3 p.R190L	вДНК, низкое подмешивание	27	xp.17	7,578,280	G	C	TP53	c.569C>A	p.P190R	1	436,466	2.29E-06	0	8,000	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	27	xp.17	7,578,280	G	C	TP53	c.569C>G	p.P190R	1	497,273	2.01 E-06	0	8,518	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	28	xp.17	7,578,280	G	T	TP53	c.569C>A	p.P190H	1	502,574	1.98E-06	0	7,704	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	29	xp.17	7,578,281	G	T	TP53	c.568C>A	p.P190T	1	502,574	1.98E-06	0	7,851	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	30	xp.17	7,578,282	G	A	TP53	c.567C>T	p.A188A	1	497,273	2.01 E-06	0	6,236	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	31	xp.17	7,578,282	G	T	TP53	c.567C>A	p.A188A	1	502,574	1.98E-06	0	7,638	0.00E+00
TRP3 p.R190L	вДНК, низкое подмешивание	31	xp.17	7,578,282	G	T	TP53	c.567C>A	p.A188A	18	436,466	3.67E-05	0	7,831	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	32	xp.17	7,578,283	G	A	TP53	c.566C>T	p.A188V	31	502,574	6.17E-05	0	7,566	0.00E+00
TRP3 p.R190L	вДНК, низкое подмешивание	32	xp.17	7,578,283	G	A	TP53	c.566C>T	p.A188V	12	436,466	2.75E-05	0	7,590	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	32	xp.17	7,578,283	G	A	TP53	c.566C>T	p.A188V	20	497,273	4.02E-06	0	8,196	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	33	xp.17	7,578,283	G	T	TP53	c.566C>A	p.A188D	1	502,574	2.01 E-06	0	6,195	0.00E+00
TRP3 p.R190L	вДНК, низкое подмешивание	34	xp.17	7,578,284	C	G	TP53	c.565G>C	p.A188P	1	436,466	2.29E-06	0	7,532	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	34	xp.17	7,578,284	C	G	TP53	c.565G>C	p.A188P	2	497,273	4.02E-06	0	6,145	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	35	xp.17	7,578,284	C	T	TP53	c.565G>A	p.A188T	3	502,574	5.97E-06	0	7,552	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	36	xp.17	7,578,285	C	G	TP53	c.564G>C	p.L188L	28	497,273	5.23E-05	0	5,981	0.00E+00
TRP3 p.R190L	вДНК, низкое подмешивание	37	xp.17	7,578,285	C	T	TP53	c.564G>A	p.L188L	1	436,466	2.29E-06	0	7,388	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	38	xp.17	7,578,286	A	G	TP53	c.563T>C	p.L188P	1	497,273	2.01 E-06	0	5,873	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	38	xp.17	7,578,287	G	C	TP53	c.562DG	p.L188V	1	497,273	2.01 E-06	0	6,948	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	40	xp.17	7,578,287	G	T	TP53	c.562DA	p.L188M	9	502,574	1.79E-05	0	7,252	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	41	xp.17	7,578,288	A	C	TP53	c.561T>G	p.G187G	1	497,273	2.01 E-06	0	5,797	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	42	xp.17	7,578,288	A	G	TP53	c.561T>C	p.G187G	2	502,574	3.98E-06	0	7,196	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	43	xp.17	7,578,288	A	T	TP53	c.561T>A	p.G187G	1	502,574	3.98E-06	0	7,198	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	44	xp.17	7,578,288	C	A	TP53	c.560G>T	p.G187V	1	497,273	2.01 E-06	0	6,794	0.00E+00
TRP3 p.R190L	вДНК, низкое подмешивание	45	xp.17	7,578,288	C	G	TP53	c.560G>C	p.G187A	1	436,466	2.29E-06	0	7,165	0.00E+00
TRP3 p.R190L	вДНК, низкое подмешивание	46	xp.17	7,578,289	C	T	TP53	c.560G>A	p.G187D	1	436,466	2.29E-06	0	7,165	0.00E+00
TRP3 p.R190L	вДНК, высокое подмешивание	47	xp.17	7,578,290	C	A	TP53	В интроне	В интроне	1	502,574	1.98E-06	0	7,165	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	47	xp.17	7,578,290	C	A	TP53	В интроне	В интроне	6	497,273	1.21E-05	0	6,770	0.00E+00
TRP3 p.R190L	вДНК, контроль дикого типа	48	xp.17	7,578,290	C	A	TP53	В интроне	В интроне	6	497,273	2.01 E-06	0	5,770	0.00E+00
TRP3 p.R342X	вДНК, контроль дикого типа	1	xp.17	7,573,986	A	G	TP53	c.1031T>C	p.L344F	7	505,165	1.36E-05	0	7,480	0.00E+00
TRP3 p.R342X	вДНК, низкое подмешивание	2	xp.17	7,573,987	G	A	TP53	c.1030C>T	p.L344L	13	505,165	2.57E-05	0	9,041	0.00E+00
TRP3 p.R342X	вДНК, контроль дикого типа	2	xp.17	7,573,987	G	A	TP53	c.1030C>T	p.L344L	3	505,165	5.94E-06	0	7,463	0.00E+00
TRP3 p.R342X	вДНК, контроль дикого типа	3	xp.17	7,573,987	G	T	TP53	c.1030G>A	p.L344M	1	505,165	1.98E-06	0	7,463	0.00E+00
TRP3 p.R342X	вДНК, высокое подмешивание	4	xp.17	7,573,988	C	T	TP53	c.1028G>A	p.E343E	3	447,331	6.71 E-06	0	8,778	0.00E+00
TRP3 p.R342X	вДНК, низкое подмешивание	4	xp.17	7,573,988	C	T	TP53	c.1028G>A	p.E343E	20	505,165	3.96E-05	0	8,897	0.00E+00
TRP3 p.R342X	вДНК, низкое подмешивание	5	xp.17	7,573,989	T	A	TP53	c.1028A>T	p.E343V	1	505,165	1.98E-06	0	8,867	0.00E+00
TRP3 p.R342X	вДНК, низкое подмешивание	6	xp.17	7,573,989	T	C	TP53	c.1028A>G	p.E343G	2	505,165	3.96E-0			

Известная мутация	Тип образца	Номер мутации**	Хром.	Положение*	Референсное основание	Вариантное основание	Ген	Замена в кДНК	Замена в белке	#Супермутантов (SafeSeqS)	Всего UID (SafeSeqS)	SafeSeqS MAF	#Суперкаллимутантов (SafSeqS)	Глубина дуплекса (SafSeqS)	SafeSeqS MAF
TR53 p.R342X	вДНК, контроль дигено типа	11	xp.17	7,574,003	G	A	TP53	c.1024>T	p.R342*	73	505,185	1.45E-04	0	7,215	0.00E+00
TR53 p.R342X	сroNA, низкое подмешивание	12	xp.17	7,574,003	G	C	TP53	c.1024C>G	p.R342G	1	505,185	1.09E-06	0	8,746	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	13	xp.17	7,574,004	G	A	TP53	c.1023C>T	p.F341F	3	447,331	6.71E-06	0	8,406	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	13	xp.17	7,574,004	G	A	TP53	c.1023C>T	p.F341F	2	505,515	3.96E-06	0	8,508	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	13	xp.17	7,574,004	G	A	TP53	c.1023C>T	p.F341F	3	505,185	5.94E-06	0	7,128	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	14	xp.17	7,574,004	G	C	TP53	c.1023C>G	p.F341L	1	447,331	2.24E-06	0	8,406	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	15	xp.17	7,574,004	G	T	TP53	c.1023C>A	p.F341L	1	505,185	1.09E-06	0	7,128	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	16	xp.17	7,574,005	A	G	TP53	c.1022T>C	p.F341S	7	447,331	1.59E-05	0	8,380	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	16	xp.17	7,574,005	A	G	TP53	c.1022T>C	p.F341S	6	505,515	1.49E-05	0	8,486	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	16	xp.17	7,574,005	A	G	TP53	c.1022T>C	p.F341S	4	505,185	7.82E-06	0	7,104	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	17	xp.17	7,574,005	A	T	TP53	c.1022T>A	p.F341Y	1	505,515	1.95E-06	0	8,486	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	18	xp.17	7,574,006	A	G	TP53	c.1021T>C	p.F341L	2	447,331	4.47E-06	0	8,356	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	19	xp.17	7,574,007	C	A	TP53	c.1020G>T	p.M340I	1	447,331	2.24E-06	0	8,274	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	19	xp.17	7,574,007	C	A	TP53	c.1020G>T	p.M340I	1	505,185	1.09E-06	0	7,038	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	20	xp.17	7,574,007	C	T	TP53	c.1020G>A	p.M340I	2	505,515	3.96E-06	0	8,398	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	21	xp.17	7,574,009	T	C	TP53	c.1018A>G	p.M340V	1	505,515	1.09E-06	0	8,285	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	22	xp.17	7,574,010	C	A	TP53	c.1017G>T	p.E336D	1	447,331	2.24E-06	0	8,156	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	22	xp.17	7,574,010	C	A	TP53	c.1017G>T	p.E336D	1	505,515	1.09E-06	0	8,271	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	22	xp.17	7,574,010	C	A	TP53	c.1017G>T	p.E336D	1	505,185	1.09E-06	0	8,917	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	23	xp.17	7,574,010	T	TP53	c.1017G>A	p.E336E	1	505,185	1.09E-06	0	8,917	0.00E+00	
TR53 p.R342X	вДНК, низкое подмешивание	24	xp.17	7,574,011	T	A	TP53	c.1016A>G	p.E336V	1	505,515	1.09E-06	0	8,246	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	25	xp.17	7,574,011	T	C	TP53	c.1016A>G	p.E336G	1	447,331	2.24E-06	0	8,137	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	25	xp.17	7,574,011	T	C	TP53	c.1016A>G	p.E336G	2	505,515	3.96E-06	0	8,249	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	25	xp.17	7,574,011	T	C	TP53	c.1016A>G	p.E336G	2	505,185	3.96E-06	0	8,893	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	26	xp.17	7,574,012	C	A	TP53	c.1015G>T	p.E336*	2	505,515	3.96E-06	0	8,207	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	27	xp.17	7,574,012	C	T	TP53	c.1015G>A	p.E336K	16	447,331	3.35E-05	0	8,089	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	27	xp.17	7,574,012	C	T	TP53	c.1015G>A	p.E336K	2	505,515	3.96E-06	0	8,207	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	27	xp.17	7,574,012	C	T	TP53	c.1015G>A	p.E336K	31	505,185	6.14E-05	0	8,848	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	28	xp.17	7,574,013	G	A	TP53	c.1014C>T	p.F338F	4	447,331	8.84E-06	0	8,302	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	28	xp.17	7,574,013	G	A	TP53	c.1014C>T	p.F338F	1	505,515	1.09E-06	0	8,098	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	28	xp.17	7,574,013	G	A	TP53	c.1014C>T	p.F338F	14	505,185	2.77E-05	0	8,604	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	29	xp.17	7,574,013	G	T	TP53	c.1014C>A	p.F338L	1	505,185	1.09E-06	0	8,304	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	30	xp.17	7,574,014	A	G	TP53	c.1013T>C	p.F338S	3	447,331	6.71E-06	0	7,352	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	30	xp.17	7,574,014	A	G	TP53	c.1013T>C	p.F338S	1	505,515	1.09E-06	0	8,091	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	30	xp.17	7,574,014	A	G	TP53	c.1013T>C	p.F338S	1	505,185	1.09E-06	0	8,793	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	31	xp.17	7,574,015	A	G	TP53	c.1012T>G	p.F338L	2	447,331	4.47E-06	0	7,933	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	32	xp.17	7,574,016	G	A	TP53	c.1011C>T	p.R337R	1	447,331	2.24E-06	0	7,922	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	32	xp.17	7,574,016	G	A	TP53	c.1011C>T	p.R337R	6	505,185	1.19E-05	0	8,786	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	33	xp.17	7,574,017	C	A	TP53	c.1010G>T	p.R337L	1	447,331	2.24E-06	0	7,841	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	33	xp.17	7,574,017	C	A	TP53	c.1010G>T	p.R337L	1	505,515	1.09E-06	0	7,952	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	34	xp.17	7,574,017	C	T	TP53	c.1010G>A	p.R337H	4	447,331	8.84E-06	0	7,841	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	34	xp.17	7,574,017	C	T	TP53	c.1010G>A	p.R337H	69	505,515	1.36E-04	0	7,952	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	34	xp.17	7,574,017	C	T	TP53	c.1010G>A	p.R337H	33	505,185	6.53E-05	0	6,969	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	35	xp.17	7,574,018	G	A	TP53	c.1009T>C	p.R337C	20	447,331	4.47E-05	0	7,305	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	35	xp.17	7,574,018	G	A	TP53	c.1009T>C	p.R337C	26	505,515	5.14E-05	0	7,915	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	35	xp.17	7,574,018	G	A	TP53	c.1009T>C	p.R337C	50	505,185	0.93E-05	0	8,842	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	37	xp.17	7,574,018	G	T	TP53	c.1008C>A	p.R337S	1	505,185	1.09E-06	0	8,842	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	38	xp.17	7,574,019	C	T	TP53	c.1008G>T	p.E336D	1	447,331	2.24E-06	0	7,866	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	38	xp.17	7,574,019	C	T	TP53	c.1008G>A	p.E336E	1	447,331	2.24E-06	0	7,866	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	38	xp.17	7,574,019	C	T	TP53	c.1008G>A	p.E336E	1	505,185	1.09E-06	0	6,608	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	39	xp.17	7,574,020	T	A	TP53	c.1007A>T	p.E336V	1	505,185	1.09E-06	0	8,571	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	40	xp.17	7,574,020	T	C	TP53	c.1007A>G	p.E336C	2	447,331	4.47E-06	0	7,960	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	40	xp.17	7,574,020	T	C	TP53	c.1007A>G	p.E336C	3	505,185	5.94E-06	0	8,571	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	41	xp.17	7,574,021	C	A	TP53	c.1006G>T	p.E336*	2	447,331	4.47E-06	0	7,438	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	42	xp.17	7,574,022	A	G	TP53	c.1005T>C	p.R336R	1	447,331	2.24E-06	0	7,401	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	42	xp.17	7,574,022	A	G	TP53	c.1005T>C	p.R336R	1	505,515	1.09E-06	0	7,549	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	43	xp.17	7,574,023	C	A	TP53	c.1004G>T	p.R336L	3	447,331	6.71E-06	0	7,352	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	44	xp.17	7,574,023	C	T	TP53	c.1004G>A	p.R335H	2	447,331	4.47E-06	0	7,352	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	44	xp.17	7,574,023	C	T	TP53	c.1004G>A	p.R335H	9	505,515	1.78E-05	0	7,492	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	44	xp.17	7,574,023	C	T	TP53	c.1004G>A	p.R335H	24	505,185	4.75E-05	0	8,440	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	45	xp.17	7,574,024	G	A	TP53	c.1003C>T	p.R335C	19	447,331	4.25E-05	0	7,337	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	45	xp.17	7,574,024	G	A	TP53	c.1003C>T	p.R335C	1	505,515	6.13E-05	0	7,462	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	45	xp.17	7,574,024	G	A	TP53	c.1003C>T	p.R335C	1	505,185	3.96E-06	0	6,411	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	46	xp.17	7,574,024	G	T	TP53	c.1003C>A	p.R335S	15	505,515	2.67E-05	0	7,462	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	47	xp.17	7,574,025	C	A	TP53	c.1002G>T	p.G334G	1	505,185	3.96E-06	0	6,341	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	47	xp.17	7,574,025	C	G	TP53	c.1002G>C	p.G334G	1	505,515	1.09E-06	0	7,373	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	48	xp.17	7,574,025	C	G	TP53	c.1002G>C	p.G334G	1	505,185	1.09E-06	0	6,341	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	48	xp.17	7,574,025	C	T	TP53	c.1002G>A	p.G334G	1	447,331	2.24E-06	0	7,549	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	49	xp.17	7,574,025	C	T	TP53	c.1002G>A	p.G334G	1	505,515	1.78E-05	0	7,373	0.00E+00
TR53 p.R342X	вДНК, контроль дигено типа	49	xp.17	7,574,025	C	T	TP53	c.1002G>A	p.G334G	1	505,185	1.09E-06	0	6,341	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	50	xp.17	7,574,026	C	A	TP53	c.1001G>T	p.G334V	1	447,331	2.24E-06	0	7,234	0.00E+00
TR53 p.R342X	вДНК, высокое подмешивание	51	xp.17	7,574,027	C	A	TP53	c.1000G>T	p.G334W	1	447,331	4.47E-06	0	7,156	0.00E+00
TR53 p.R342X	вДНК, низкое подмешивание	51	xp.17	7,574,027	C	A	TP53	c.1000G>T	p.G334W	1	505,515	1.09E-06	0	7,286	0.00E+00
TR53 p.R342X	вД														

Таблица 9. Сравнение мутаций, идентифицированных с помощью независимых от цепи молекулярных штрих-кодов и SaferSeq5. Координаты относятся к выпуску референсного генома человека hg19 (Консорциум референсного генома GRCh37, февраль 2009 года).

**Номера мутаций соответствуют указанным на Фиг. 26.

***Общая глубина определена как сумма глубин Уотсона и Крика. Следует обратить внимание, что эта цифра соответствует общему количеству оцененных цепей, а не дуплексов. Общее количество супермутантов аналогично определено как сумма супермутантов Уотсона и Крика.

****Средняя частота мутантных аллелей Уотсона/Крика (MAF) представляет собой среднюю MAF, наблюдаемую для двух цепей, взвешенную по их соответствующим глубинам.

Известная мутация	Тип образца	Номер мутации**	Хром.	Положение*	Референсное основание	Вариантное основание	Ген	Изменение в днк	Изменение в белке	#Супермутантов Уотсона	Глубина Уотсона	MAF Уотсона	#Супермутантов Крика	Глубина Крика	MAF Крика	Общее # супермутантов***	Общая глубина	Средняя MAF Уотсона/Крика***	#Суперкалпм утангов (SaferSeq5)	Глубина дуплекса (SaferSeq5)	MAF SaferSeq5
TP53 p.L264fs	вднк, низкое подмешивание	1	xp.17	7,577,138	C	T	TP53	c.900G>A	p.R287Q	1	10,167	9.84E-05	0	8,576	0.00E+00	18,746	5.33E-05	0	7,519	0.00E+00	
TP53 p.L264fs	вднк, высокое подмешивание	2	xp.17	7,577,139	G	A	TP53	c.799C>T	p.R287W	1	9,894	1.03E-04	0	8,109	0.00E+00	17,803	5.62E-05	0	7,067	0.00E+00	
TP53 p.L264fs	вднк, высокое подмешивание	3	xp.17	7,577,138	G	T	TP53	c.799C>A	p.R267R	1	9,894	1.03E-04	0	8,109	0.00E+00	17,803	5.62E-05	0	7,067	0.00E+00	
TP53 p.L264fs	вднк, высокое подмешивание	4	xp.17	7,577,140	C	T	TP53	c.798A>G	p.G269G	1	9,945	1.04E-04	0	8,062	0.00E+00	17,707	5.65E-05	0	7,025	0.00E+00	
TP53 p.L264fs	вднк, высокое подмешивание	5	xp.17	7,577,142	C	T	TP53	c.798A>G	p.G269R	1	9,945	1.05E-04	0	7,959	0.00E+00	17,562	5.69E-05	0	6,963	0.00E+00	
TP53 p.L264fs	вднк, высокое подмешивание	19	xp.17	7,577,144	AGT	-	TP53	c.792_794del	p.L264fs	62	9,418	9.59E-05	43	7,870	5.46E-03	105	17,298	9.07E-03	43	6,949	9.29E-03
TP53 p.L264fs	вднк, низкое подмешивание	6	xp.17	7,577,144	AGT	-	TP53	c.792_794del	p.L264fs	2	9,958	2.01E-04	1	8,394	1.19E-04	3	18,352	1.63E-04	1	7,352	1.39E-04
TP53 p.L264fs	вднк, контроль дикого типа	6	xp.17	7,577,144	AGT	-	TP53	c.792_794del	p.L264fs	1	10,422	9.69E-05	1	8,597	1.16E-04	2	18,019	1.05E-04	0	7,106	0.00E+00
TP53 p.L264fs	вднк, низкое подмешивание	7	xp.17	7,577,158	C	T	TP53	В интроне	В интроне	1	8,060	1.10E-04	0	7,898	0.00E+00	16,788	5.96E-05	0	6,725	0.00E+00	
TP53 p.L264fs	вднк, высокое подмешивание	8	xp.17	7,577,162	A	G	TP53	В интроне	В интроне	1	8,391	1.19E-04	0	7,027	0.00E+00	15,418	6.49E-05	0	6,062	0.00E+00	
TP53 p.L264fs	вднк, высокое подмешивание	9	xp.17	7,577,166	A	T	TP53	В интроне	В интроне	0	8,488	0.00E+00	1	7,191	1.39E-04	1	15,679	6.38E-05	0	6,282	0.00E+00
TP53 p.L264fs	вднк, высокое подмешивание	10	xp.17	7,577,177	A	T	TP53	В интроне	В интроне	1	7,732	1.29E-04	0	6,491	0.00E+00	14,223	7.93E-05	0	5,631	0.00E+00	
TP53 p.L264fs	вднк, высокое подмешивание	11	xp.17	7,577,218	-	-	TP53	В интроне	В интроне	0	1,995	0.00E+00	1	1,911	6.21E-04	3,905	2.77E-04	0	1,418	0.00E+00	
TP53 p.L264fs	вднк, контроль дикого типа	11	xp.17	7,577,218	-	-	TP53	В интроне	В интроне	0	2,142	0.00E+00	1	1,688	5.92E-04	3,830	2.81E-04	0	1,430	0.00E+00	
TP53 p.L264fs	вднк, низкое подмешивание	12	xp.17	7,577,220	A	T	TP53	В интроне	В интроне	1	2,071	4.83E-04	1	1,701	5.68E-04	3,772	5.30E-04	1	1,526	5.55E-04	
TP53 p.L264fs	вднк, низкое подмешивание	13	xp.17	7,577,303	-	-	TP53	В интроне	В интроне	0	592	0.00E+00	1	496	2.02E-03	1,089	9.19E-04	0	464	0.00E+00	
TP53 p.L264fs	вднк, высокое подмешивание	14	xp.17	7,577,368	G	A	TP53	В интроне	В интроне	0	470	0.00E+00	1	407	2.46E-03	877	1.14E-03	0	365	0.00E+00	
TP53 p.L264fs	вднк, высокое подмешивание	15	xp.17	7,577,739	G	T	TP53	В интроне	В интроне	1	1	1.00E+00	0	6	0.00E+00	7	1.43E-01	0	1	0.00E+00	
TP53 p.L264fs	вднк, высокое подмешивание	16	xp.17	7,577,742	T	C	TP53	В интроне	В интроне	1	8	1.00E+00	0	8	0.00E+00	9	1.11E-01	0	1	0.00E+00	
TP53 p.P190L	вднк, низкое подмешивание	1	xp.17	7,578,263	G	A	TP53	c.589G>A	p.V197M	1	11,812	9.57E-05	0	8,992	1.47E-05	21,204	4.99E-05	0	17,204	8.65E-05	
TP53 p.P190L	вднк, контроль дикого типа	1	xp.17	7,578,260	C	T	TP53	c.589G>A	p.V197M	1	10,578	9.46E-05	0	8,587	0.00E+00	19,143	5.22E-05	0	7,287	0.00E+00	
TP53 p.P190L	вднк, низкое подмешивание	2	xp.17	7,578,282	C	T	TP53	c.587G>A	p.R199Q	1	11,691	8.55E-05	0	8,793	0.00E+00	21,484	4.65E-05	0	8,764	0.00E+00	
TP53 p.P190L	вднк, контроль дикого типа	2	xp.17	7,578,282	C	T	TP53	c.587G>A	p.R199Q	1	10,481	9.54E-05	0	8,488	0.00E+00	19,969	5.27E-05	0	7,219	0.00E+00	
TP53 p.P190L	вднк, высокое подмешивание	3	xp.17	7,578,263	G	A	TP53	c.586G>T	p.R199*	0	11,880	0.00E+00	1	9,969	1.01E-04	21,749	4.60E-05	0	8,715	0.00E+00	
TP53 p.P190L	вднк, низкое подмешивание	4	xp.17	7,578,263	G	T	TP53	c.586G>A	p.R199R	1	11,672	8.57E-05	0	9,779	0.00E+00	21,451	4.66E-05	0	8,750	0.00E+00	
TP53 p.P190L	вднк, высокое подмешивание	5	xp.17	7,578,277	G	T	TP53	c.572G>A	p.P191H	1	10,564	9.47E-05	0	8,794	0.00E+00	19,358	5.17E-05	0	7,770	0.00E+00	
TP53 p.P190L	вднк, высокое подмешивание	6	xp.17	7,578,278	G	A	TP53	c.571G>A	p.P191H	1	10,564	9.50E-05	0	8,790	0.00E+00	19,358	5.17E-05	0	7,738	0.00E+00	
TP53 p.P190L	вднк, высокое подмешивание	7	xp.17	7,578,278	G	A	TP53	c.570del	p.P190fs	1	10,456	9.93E-05	0	8,733	0.00E+00	19,189	5.20E-05	0	7,715	0.00E+00	
TP53 p.P190L	вднк, низкое подмешивание	7	xp.17	7,578,278	-	-	TP53	c.570del	p.P190fs	2	10,310	1.94E-04	0	8,865	0.00E+00	19,875	1.05E-04	0	7,721	0.00E+00	
TP53 p.P190L	вднк, высокое подмешивание	8	xp.17	7,578,280	G	A	TP53	c.569G>T	p.P190L	66	10,480	9.16E-03	63	8,721	7.22E-03	159	18,201	8.28E-03	62	7,704	8.09E-03
TP53 p.P190L	вднк, низкое подмешивание	8	xp.17	7,578,280	G	A	TP53	c.569G>T	p.P190L	10	10,299	9.71E-04	7	8,658	8.06E-04	17	19,957	8.97E-04	7	7,715	9.07E-04
TP53 p.P190L	вднк, контроль дикого типа	9	xp.17	7,578,284	C	A	TP53	c.565G>T	p.A189G	0	8,929	1.12E-04	0	7,228	0.00E+00	16,157	5.19E-05	0	6,145	0.00E+00	
TP53 p.P190L	вднк, контроль дикого типа	9	xp.17	7,578,285	C	A	TP53	c.564G>T	p.L189L	1	8,710	1.15E-04	0	7,038	0.00E+00	15,748	6.39E-05	0	5,981	0.00E+00	
TP53 p.P190L	вднк, контроль дикого типа	11	xp.17	7,578,269	C	G	TP53	c.560G>C	p.G187A	1	8,425	1.19E-04	0	6,909	0.00E+00	15,234	6.56E-05	0	5,794	0.00E+00	
TP53 p.P190L	вднк, низкое подмешивание	12	xp.17	7,578,269	G	A	TP53	В интроне	В интроне	0	9,375	0.04E-04	0	8,045	0.00E+00	17,620	5.85E-05	0	7,154	0.00E+00	
TP53 p.P190L	вднк, высокое подмешивание	13	xp.17	7,578,314	G	T	TP53	В интроне	В интроне	0	8,197	1.00E-04	0	6,851	0.00E+00	15,048	6.65E-05	0	6,041	0.00E+00	
TP53 p.P190L	вднк, низкое подмешивание	14	xp.17	7,578,320	-	-	TP53	В интроне	В интроне	1	2,029	4.93E-04	0	1,654	0.00E+00	3,683	2.72E-04	0	1,502	0.00E+00	
TP53 p.P190L	вднк, высокое подмешивание	15	xp.17	7,578,328	-	-	TP53	В интроне	В интроне	1	1,828	5.47E-04	0	1,467	0.00E+00	3,295	3.03E-04	0	1,315	0.00E+00	
TP53 p.P190L	сiDNA контроль дикого типа	16	xp.17	7,578,363	C	A	TP53	В интроне	В интроне	0	1,604	0.00E+00	1	1,255	7.97E-04	2,859	3.00E-04	0	1,064	0.00E+00	
TP53 p.P190L	вднк, контроль дикого типа	17	xp.17	7,578,385	C	T	TP53	c.545G>A	p.C182Y	1	1,351	7.43E-04	0	1,075	6.00E+00	2,428	4.12E-04	0	914	0.00E+00	
TP53 p.P190L	вднк, контроль дикого типа	18	xp.17	7,578,397	G	A	TP53	c.533del	p.H178fs	1	597	1.68E-03	0	493	0.00E+00	1,080	9.26E-04	0	420	0.00E+00	
TP53 p.P190L	вднк, низкое подмешивание	19	xp.17	7,578,421	G	A	TP53	c.509G>T	p.L178I	0	792	0.00E+00	0	697	0.00E+00	1,489	1.52E-03	0	898	0.00E+00	
TP53 p.P190L	вднк, низкое подмешивание	20	xp.17	7,578,475	G	A	TP53	c.445G>T	p.P152L	0	477	0.00E+00	1	401	2.49E-03	878	1.14E-03	0	365	0.00E+00	
TP53 p.P190L	вднк, низкое подмешивание	21	xp.17	7,578,712	TTTT	-	TP53	В интроне	В интроне	0	7	0.00E+00	1	9	1.11E-01	16	6.25E-02	0	1	0.00E+00	
TP53 p.P190L	вднк, высокое подмешивание	22	xp.17	7,578,714	T	C	TP53	В интроне	В интроне	0	9	0.00E+00	1	7	1.43E-01	16	6.25E-02	0	2	0.00E+00	
TP53 p.P190L	вднк, контроль дикого типа	23	xp.17	7,578,716	C	T	TP53	В интроне	В интроне	0	1	0.00E+00	1	17	5.88E-02	18	5.68E-02	0	1	0.00E+00	
TP53 p.P190L	вднк, высокое подмешивание	24	xp.17	7,578,744	CT	-	TP53	В интроне	В интроне	1	8	0.00E+00	0	33	0.00E+00	41	2.44E-02	0	5	0.00E+00	
TP53 p.R342X	вднк, высокое подмешивание	1	xp.17	7,575,999	A	C	TP53	c.1031T>G	p.L344R	1	12,030	8.31E-05	0	10,379	0.00E+00	22,409	4.46E-05	0	8,947	0.00E+00	
TP53 p.R342X	вднк, высокое подмешивание	2	xp.17	7,574,003	G	A	TP53	c.1024C>T	p.R342*	81	11,840	5.24E-03	45	10,025	4.48E-03	106	21,865	4.98E-03	43	8,940	4.98E-03
TP53 p.R342X	вднк, низкое подмешивание	2	xp.17	7,574,003	G	A	TP53	c.1024C>T	p.R342*	9	11,889	7.70E-04	6	10,092	5.95E-04	15	21,781	8.89E-04	8	8,749	8.89E-04
TP53 p.R342X	вднк, контроль дикого типа	2	xp.17	7,574,003	G	A	TP53	c.1024C>T	p.R342*	1	11,223	8.91E-05	0	9,482	0.00E+00	20,705	4.83E-05	0	7,215	0.00E+00	
TP53 p.R342X	вднк, контроль дикого типа	3	xp.17	7,574,003	G	A	TP53	c.1024C>G	p.R342G	0	11,223	8.91E-05	0	9,482	0.00E+00	20,705	4.83E-05	0	7,215	0.00E+00	
TP53 p.R342X	вднк, низкое подмешивание	4	xp.17	7,574,005	A	C	TP53	c.1022T>C	p.F341S	2	11,343	1.76E-04	0	9,794	0.00E+00	21,137	9.46E-05	0	8,486	0.00E+00	
TP53 p.R342X	вднк, контроль дикого типа	5	xp.17	7,574,009	T	C	TP53	c.1018A>G	p.M340V	1	10,778	9.28E-05	0	9,108	0.00E+00	19,886	5.03E-05	0	6,930	0.00E+00	
TP53 p.R342X	вднк, высокое подмешивание	6	xp.17	7,574,012	C	T	TP53	c.1015G>A	p.E339K	1	10,620	9.19E-05	0	9,403	0.00E+00	20,323	4.92E-05	0	8,909	0.00E+00	

Таблица 10. Мутации, идентифицированные с помощью SafeSeqS и SaferSeqS в образцах плазмы крови, полученных от онкобольных.

* Координаты относятся к выпуску референсного генома человека hg19 (Консорциум референсного генома GRCh37, февраль 2009 года).

**Номер мутации соответствует указанному на Фиг. 27.

NA, положение, не подвергнувшееся оценке с помощью SafeSeqS или SaferSeqS.

Идентификатор павелта	Номер Мутации**	Хром.	Положение*	Референсное основание	Вариантное основание	Ген	Изменение в ДНК	Изменение в белке	Мутация ранее была выявлена?	#Супермутантов (SaferSeqS)	Всего UID (SafeSeqS)	SafeSeqS MAF	#Суперкалмутантов (SaferSeqS)	Глубина дуплекса (SaferSeqS)	SaferSeqS MAF
LCRA1797	1	Xp.12	25.398.277	T	C	KRAS	c.42A>G	p.V14V	Нет	1	1,059,932	9.43E-07	0	3,980	0.00E+00
LCRA1797	2	Xp.12	25.398.279	C	A	KRAS	c.40G>T	p.V14L	Нет	2	1,059,932	1.89E-06	0	3,950	0.00E+00
LCRA1797	3	Xp.12	25.398.279	C	T	KRAS	c.40G>A	p.V14I	Да	634	1,059,932	5.98E-04	3	3,950	7.59E-04
LCRA1797	4	Xp.12	25.398.280	G	A	KRAS	c.39C>T	p.G13G	Нет	2	1,059,932	1.89E-06	0	3,936	0.00E+00
LCRA1797	5	Xp.12	25.398.280	G	T	KRAS	c.39C>A	p.G13G	Нет	2	1,059,932	1.89E-06	0	3,936	0.00E+00
LCRA1797	6	Xp.12	25.398.281	C	A	KRAS	c.38G>T	p.G13V	Нет	3	1,059,932	2.83E-06	0	3,915	0.00E+00
LCRA1797	7	Xp.12	25.398.281	C	T	KRAS	c.38G>A	p.G13D	Нет	1	1,059,932	9.43E-07	0	3,915	0.00E+00
LCRA1797	8	Xp.12	25.398.282	C	A	KRAS	c.37G>T	p.G13C	Нет	1	1,059,932	9.43E-07	0	3,829	0.00E+00
LCRA1797	9	Xp.12	25.398.283	A	G	KRAS	c.36T>C	p.G12G	Нет	2	1,059,932	1.89E-06	0	3,808	0.00E+00
LCRA1797	10	Xp.12	25.398.284	C	A	KRAS	c.35G>T	p.G12V	Нет	58	1,059,932	5.28E-05	0	3,794	0.00E+00
LCRA1797	11	Xp.12	25.398.284	C	G	KRAS	c.35G>C	p.G12A	Нет	16	1,059,932	1.51E-05	0	3,794	0.00E+00
LCRA1797	12	Xp.12	25.398.285	C	A	KRAS	c.34G>T	p.G12C	Нет	5	1,059,932	4.72E-06	0	3,774	0.00E+00
LCRA1797	13	Xp.12	25.398.285	C	T	KRAS	c.34G>A	p.G12S	Нет	2	1,059,932	1.89E-06	0	3,774	0.00E+00
LCRA1797	14	Xp.12	25.398.286	A	C	KRAS	c.33T>G	p.A11A	Нет	1	1,059,932	9.43E-07	0	3,746	0.00E+00
LCRA1797	15	Xp.12	25.398.287	G	A	KRAS	c.32C>T	p.A11V	Нет	3	1,059,932	2.83E-06	0	3,735	0.00E+00
LCRA1797	16	Xp.12	25.398.287	G	T	KRAS	c.32C>A	p.A11D	Нет	1	1,059,932	9.43E-07	0	3,735	0.00E+00
LCRA1797	17	Xp.12	25.398.288	C	A	KRAS	c.31G>T	p.A11S	Нет	3	1,059,932	2.83E-06	0	3,724	0.00E+00
LCRA1797	18	Xp.12	25.398.288	C	T	KRAS	c.31G>A	p.A11T	Нет	4	1,059,932	3.77E-06	0	3,724	0.00E+00
LCRA1797	19	Xp.12	25.398.289	T	C	KRAS	c.30A>G	p.G10G	Нет	26	1,059,932	2.45E-05	0	3,631	0.00E+00
LCRA1797	20	Xp.12	25.398.289	T	G	KRAS	c.30A>C	p.G10G	Нет	1	1,059,932	9.43E-07	0	3,631	0.00E+00
LCRA1797	21	Xp.12	25.398.291	C	A	KRAS	c.28G>T	p.G10*	Нет	1	1,059,932	9.43E-07	0	3,590	0.00E+00
LCRA1797	22	Xp.12	25.398.293	A	G	KRAS	c.26T>C	p.V9A	Нет	4	1,059,932	3.77E-06	0	3,481	0.00E+00
LCRA1797	23	Xp.12	25.398.294	C	A	KRAS	c.25G>T	p.V9F	Нет	75	1,059,932	7.08E-05	0	3,458	0.00E+00
LCRA1797	24	Xp.12	25.398.296	A	G	KRAS	c.23T>C	p.V8A	Нет	2	1,059,932	1.89E-06	0	3,419	0.00E+00
LCRA1797	25	Xp.12	25.398.297	C	A	KRAS	c.22G>T	p.V8L	Нет	3	1,059,932	2.83E-06	0	3,409	0.00E+00
LCRA1797	26	Xp.12	25.398.297	C	T	KRAS	c.22G>A	p.V8I	Нет	1	1,059,932	9.43E-07	0	3,409	0.00E+00
LCRA1797	27	Xp.12	25.398.298	C	T	KRAS	c.21G>A	p.V7V	Нет	1	1,059,932	9.43E-07	0	3,346	0.00E+00
LCRA1877	1	Xp.7	140.453.129	A	C	BRAF	c.1806T>G	p.S602S	Нет	1	1,210,434	8.26E-07	0	12,796	0.00E+00
LCRA1877	2	Xp.7	140.453.132	T	C	BRAF	c.1803A>G	p.K601K	Нет	1	1,210,434	8.26E-07	0	12,236	0.00E+00
LCRA1877	3	Xp.7	140.453.134	T	C	BRAF	c.1801A>G	p.K601E	Нет	4	1,210,434	3.30E-06	0	11,975	0.00E+00
LCRA1877	4	Xp.7	140.453.136	A	G	BRAF	c.1799T>C	p.V600A	Нет	5	1,210,434	4.13E-06	0	11,808	0.00E+00
LCRA1877	5	Xp.7	140.453.136	A	T	BRAF	c.1799T>A	p.V600E	Да	931	1,210,434	7.69E-04	16	11,808	1.36E-03
LCRA1877	6	Xp.7	140.453.137	C	A	BRAF	c.1789G>T	p.V600L	Нет	1	1,210,434	2.48E-06	0	11,742	0.00E+00
LCRA1877	7	Xp.7	140.453.137	C	G	BRAF	c.1789G>C	p.V600L	Нет	3	1,210,434	8.26E-07	0	11,742	0.00E+00
LCRA1877	8	Xp.7	140.453.137	C	T	BRAF	c.1789G>A	p.V600M	Нет	7	1,210,434	5.78E-06	0	11,742	0.00E+00
LCRA1877	9	Xp.7	140.453.138	T	C	BRAF	c.1797A>G	p.T599T	Нет	5	1,210,434	4.13E-06	0	11,633	0.00E+00
LCRA1877	10	Xp.7	140.453.138	T	G	BRAF	c.1797A>C	p.T599T	Нет	1	1,210,434	8.26E-07	0	11,633	0.00E+00
LCRA1877	11	Xp.7	140.453.139	G	A	BRAF	c.1796C>T	p.T599I	Нет	13	1,210,434	1.07E-05	0	11,594	0.00E+00
LCRA1877	12	Xp.7	140.453.142	G	T	BRAF	c.1793C>A	p.A598D	Нет	4	1,210,434	3.30E-06	0	11,365	0.00E+00
LCRA1877	13	Xp.7	140.453.146	G	T	BRAF	c.1789C>A	p.L597I	Нет	5	1,210,434	4.13E-06	0	11,143	0.00E+00
LCRA1877	14	Xp.7	140.453.147	A	G	BRAF	c.1788T>C	p.G596G	Нет	1	1,210,434	8.26E-07	0	11,109	0.00E+00
LCRA1877	15	Xp.7	140.453.148	C	T	BRAF	c.1787G>A	p.G596D	Нет	3	1,210,434	2.48E-06	0	10,993	0.00E+00
LCRA1877	16	Xp.7	140.453.149	A	G	BRAF	c.1786G>T	p.G596C	Нет	1	1,210,434	8.26E-07	0	10,242	0.00E+00
LCRA1877	17	Xp.7	140.453.149	C	T	BRAF	c.1786G>A	p.G596G	Нет	2	1,210,434	1.65E-06	0	10,242	0.00E+00
LCRA1877	18	Xp.7	140.453.150	A	C	BRAF	c.1785T>G	p.F595L	Нет	2	1,210,434	1.65E-06	0	10,201	0.00E+00
LCRA1877	19	Xp.7	140.453.150	A	G	BRAF	c.1785T>C	p.F595F	Нет	9	1,210,434	7.44E-06	0	10,201	0.00E+00
LCRA1877	20	Xp.7	140.453.151	A	C	BRAF	c.1784T>G	p.F595C	Нет	1	1,210,434	8.26E-07	0	10,140	0.00E+00
LCRA1877	21	Xp.7	140.453.153	A	T	BRAF	c.1782T>A	p.D594E	Нет	1	1,210,434	8.26E-07	0	10,096	0.00E+00
LCRA1877	22	Xp.7	140.453.154	T	C	BRAF	c.1781A>G	p.D594G	Нет	2	1,210,434	1.65E-06	0	10,043	0.00E+00
LCRA1877	23	Xp.7	140.453.155	C	A	BRAF	c.1780G>T	p.D594Y	Нет	1	1,210,434	8.26E-07	0	9,966	0.00E+00
LCRA1877	24	Xp.7	140.453.156	A	G	BRAF	c.1779T>C	p.G593G	Нет	1	1,210,434	8.26E-07	0	9,936	0.00E+00
LCRA1877	25	Xp.7	140.453.158	C	A	BRAF	c.1777G>T	p.G593C	Нет	5	1,210,434	4.13E-06	0	9,832	0.00E+00
LCRA1877	26	Xp.7	140.453.159	T	C	BRAF	c.1776A>G	p.I592M	Нет	21	1,210,434	1.73E-05	0	9,796	0.00E+00
LCRA1877	27	Xp.7	140.453.160	-	T	BRAF	c.1774_1775insA	p.I592I	Нет	1	1,210,434	8.26E-07	0	9,723	0.00E+00
LCRA1877	28	Xp.7	140.453.161	-	-	BRAF	c.1774del	p.I592I	Нет	25	1,210,434	2.07E-05	0	9,625	0.00E+00
LCRA1896	1	Xp.7	140.453.130	G	T	BRAF	c.1805C>A	p.S602Y	Нет	2	1,212,620	1.65E-06	0	26,949	0.00E+00
LCRA1896	2	Xp.7	140.453.132	T	C	BRAF	c.1803A>G	p.K601K	Нет	3	1,212,620	2.47E-06	0	25,980	0.00E+00
LCRA1896	3	Xp.7	140.453.134	T	C	BRAF	c.1801A>G	p.K601E	Нет	7	1,212,620	5.77E-06	0	25,373	0.00E+00
LCRA1896	4	Xp.7	140.453.135	C	G	BRAF	c.1800G>C	p.V600V	Нет	1	1,212,620	8.25E-07	0	25,166	0.00E+00
LCRA1896	5	Xp.7	140.453.136	A	G	BRAF	c.1799T>C	p.V600A	Нет	3	1,212,620	2.47E-06	0	24,997	0.00E+00
LCRA1896	6	Xp.7	140.453.136	A	T	BRAF	c.1799T>A	p.V600E	Да	492	1,212,620	4.06E-04	13	24,997	5.20E-04
LCRA1896	7	Xp.7	140.453.137	C	G	BRAF	c.1798G>C	p.V600L	Нет	2	1,212,620	1.65E-06	0	24,826	0.00E+00
LCRA1896	8	Xp.7	140.453.137	C	T	BRAF	c.1798G>A	p.V600M	Нет	11	1,212,620	9.07E-06	0	24,826	0.00E+00
LCRA1896	9	Xp.7	140.453.138	T	A	BRAF	c.1797A>T	p.T599T	Нет	1	1,212,620	8.25E-07	0	24,543	0.00E+00
LCRA1896	10	Xp.7	140.453.138	T	C	BRAF	c.1797A>G	p.T599T	Нет	1	1,212,620	8.25E-07	0	24,543	0.00E+00
LCRA1896	11	Xp.7	140.453.138	C	G	BRAF	c.1797A>C	p.T599T	Нет	1	1,212,620	8.25E-07	0	24,543	0.00E+00
LCRA1896	12	Xp.7	140.453.139	G	A	BRAF	c.1796C>T	p.T599I	Нет	21	1,212,620	1.73E-05	0	24,464	0.00E+00
LCRA1896	13	Xp.7	140.453.140	T	A	BRAF	c.1795A>T	p.T599S	Нет	1	1,212,620	8.25E-07	0	24,134	0.00E+00
LCRA1896	14	Xp.7	140.453.142	G	A	BRAF	c.1793C>T	p.A598V	Нет	5	1,212,620	4.12E-06	0	24,013	0.00E+00
LCRA1896	15	Xp.7	140.453.142	G	T	BRAF	c.1793C>A	p.A598D	Нет	2	1,212,620	1.65E-06	0	24,013	0.00E+00
LCRA1896	16	Xp.7	140.453.143	C	T	BRAF	c.1792G>A	p.A598T	Нет	1	1,212,620	8.25E-07	0	23,963	0.00E+00
LCRA1896	17	Xp.7	140.453.145	A	G	BRAF	c.1790T>C	p.L597P	Нет	3	1,212,620	2.47E-06	0	23,763	0.00E+00
LCRA1896	18	Xp.7	140.453.146	G	A	BRAF	c.1789C>T	p.L597L	Нет	2	1,212,620	1.65E-06	0	23,525	0.00E+00
LCRA1896	19	Xp.7	140.453.146	G	T	BRAF	c.1789C>A	p.L597I	Нет	6	1,212,620	4.95E-06	0	23,525	0.00E+00
LCRA1896	20	Xp.7	140.453.147	A	T	BRAF	c.1788T>A	p.G596G	Нет	1	1,212,620	8.25E-07	0	23,451	0.00E+00
LCRA1896	21	Xp.7	140.453.148	C	T	BRAF	c.1787G>A	p.G596D	Нет	1	1,212,620	4.12E-06	0	23,085	0.00E+00
LCRA1896	22	Xp.7	140.453.149	C	T	BRAF	c.1786G>A	p.G596S							

LCRA1896	25	Xp.7	140,453,167	C	T	BRAF	c.1778G>A	P.G593D	Het	1	1212,620	8.25E-07	0	20,753	0.00E+00
LCRA1896	26	Xp.7	140,453,168	C	G	BRAF	c.1777G>C	P.G593R	Het	1	1,212,620	8.26E-07	0	20,841	0.00E+00
LCRA1896	27	Xp.7	140,453,159	C	T	BRAF	c.1777G>A	P.G593S	Het	2	1,212,620	1.65E-06	0	20,841	0.00E+00
LCRA1896	28	Xp.7	140,453,159	T	C	BRAF	c.1776A>G	P.I562M	Het	8	1,212,620	6.89E-06	0	20,568	0.00E+00
LCRA1938	1	Xp.12	25,398,277	T	C	KRAS	c.429G>C	P.V14V	Het	5	5,401,596	9.28E-07	0	21,700	0.00E+00
LCRA1938	2	Xp.12	25,398,278	A	G	KRAS	c.417T>C	P.V14A	Het	4	5,401,596	7.41E-07	0	21,843	0.00E+00
LCRA1938	3	Xp.12	25,398,279	C	A	KRAS	c.40G>T	p.V14L	Het	6	5,401,596	1.11E-06	0	21,402	0.00E+00
LCRA1938	4	Xp.12	25,398,279	G	T	KRAS	c.40G>A	p.V14I	Het	249	5,401,596	4.55E-05	0	21,402	0.00E+00
LCRA1938	5	Xp.12	25,398,280	C	A	KRAS	c.39C>T	P.G13G	Het	503	5,401,596	9.31E-05	0	21,286	0.00E+00
LCRA1938	6	Xp.12	25,398,280	G	T	KRAS	c.39C>A	P.G13G	Het	18	5,401,596	3.33E-06	0	21,286	0.00E+00
LCRA1938	7	Xp.12	25,398,281	C	A	KRAS	c.38G>T	P.G13V	Het	5	5,401,596	9.28E-07	0	21,202	0.00E+00
LCRA1938	8	Xp.12	25,398,281	C	T	KRAS	c.38G>A	P.G13D	Het	3	5,401,596	5.55E-07	0	21,202	0.00E+00
LCRA1938	9	Xp.12	25,398,282	C	A	KRAS	c.37G>T	P.G13C	Het	1	5,401,596	1.85E-07	0	20,798	0.00E+00
LCRA1938	10	Xp.12	25,398,282	C	T	KRAS	c.37G>A	P.G13S	Het	1	5,401,596	1.85E-07	0	20,798	0.00E+00
LCRA1938	11	Xp.12	25,398,283	A	G	KRAS	c.38T>C	P.G12G	Het	8	5,401,596	1.48E-06	0	20,873	0.00E+00
LCRA1938	12	Xp.12	25,398,284	C	A	KRAS	c.35G>T	P.G12V	Δa	744	5,401,596	1.38E-04	3	20,577	1.46E-04
LCRA1938	13	Xp.12	25,398,284	C	G	KRAS	c.35G>C	P.G12A	Het	175	5,401,596	3.24E-05	0	20,577	0.00E+00
LCRA1938	14	Xp.12	25,398,284	C	T	KRAS	c.35G>A	P.G12D	Het	3	5,401,596	5.55E-07	0	20,577	0.00E+00
LCRA1938	15	Xp.12	25,398,285	C	A	KRAS	c.34G>T	P.G12C	Het	5	5,401,596	9.28E-07	0	20,358	0.00E+00
LCRA1938	16	Xp.12	25,398,285	C	T	KRAS	c.34G>A	P.G12S	Het	4	5,401,596	7.41E-07	0	20,358	0.00E+00
LCRA1938	17	Xp.12	25,398,286	A	C	KRAS	c.33T>G	p.A11A	Het	4	5,401,596	7.41E-07	0	20,202	0.00E+00
LCRA1938	18	Xp.12	25,398,286	A	G	KRAS	c.33T>C	p.A11A	Het	13	5,401,596	2.41E-06	0	20,202	0.00E+00
LCRA1938	19	Xp.12	25,398,286	A	T	KRAS	c.33T>A	p.A11A	Het	1	5,401,596	1.85E-07	0	20,202	0.00E+00
LCRA1938	20	Xp.12	25,398,287	G	A	KRAS	c.32C>T	P.A11V	Het	15	5,401,596	2.78E-06	0	20,131	0.00E+00
LCRA1938	21	Xp.12	25,398,287	G	A	KRAS	c.32C>A	P.A11D	Het	8	5,401,596	1.48E-06	0	20,131	0.00E+00
LCRA1938	22	Xp.12	25,398,288	C	A	KRAS	c.31G>T	P.A11S	Het	130	5,401,596	2.41E-05	0	20,011	0.00E+00
LCRA1938	23	Xp.12	25,398,288	C	G	KRAS	c.31G>C	p.A11P	Het	1	5,401,596	1.85E-07	0	20,011	0.00E+00
LCRA1938	24	Xp.12	25,398,288	C	T	KRAS	c.31G>A	p.A11T	Het	13	5,401,596	2.41E-06	0	20,011	0.00E+00
LCRA1938	25	Xp.12	25,398,289	T	A	KRAS	c.30A>T	P.G10G	Het	18	5,401,596	3.33E-06	0	19,488	0.00E+00
LCRA1938	26	Xp.12	25,398,289	T	C	KRAS	c.30A>G	P.G10G	Het	140	5,401,596	2.59E-05	0	19,488	0.00E+00
LCRA1938	27	Xp.12	25,398,290	C	A	KRAS	c.29G>T	P.G10V	Het	2	5,401,596	3.70E-07	0	19,404	0.00E+00
LCRA1938	28	Xp.12	25,398,290	C	T	KRAS	c.29G>A	P.G10E	Het	1	5,401,596	1.85E-07	0	19,404	0.00E+00
LCRA1938	29	Xp.12	25,398,291	C	A	KRAS	c.28G>T	P.G10*	Het	2	5,401,596	3.70E-07	0	19,177	0.00E+00
LCRA1938	30	Xp.12	25,398,291	C	T	KRAS	c.28G>A	P.G10R	Het	1	5,401,596	1.85E-07	0	19,177	0.00E+00
LCRA1938	31	Xp.12	25,398,292	A	G	KRAS	c.27T>C	p.V9V	Het	6	5,401,596	1.11E-06	0	18,730	0.00E+00
LCRA1938	32	Xp.12	25,398,292	A	T	KRAS	c.27T>A	p.V9V	Het	1	5,401,596	1.85E-07	0	18,730	0.00E+00
LCRA1938	33	Xp.12	25,398,293	A	G	KRAS	c.26T>C	p.V9A	Het	2	5,401,596	5.70E-07	0	18,575	0.00E+00
LCRA1938	34	Xp.12	25,398,294	C	A	KRAS	c.25G>T	p.V9F	Het	8	5,401,596	1.48E-06	0	18,560	0.00E+00
LCRA1938	35	Xp.12	25,398,294	C	G	KRAS	c.25G>C	p.V8L	Het	3	5,401,596	5.55E-07	0	18,590	0.00E+00
LCRA1938	36	Xp.12	25,398,294	C	T	KRAS	c.25G>A	p.V9I	Het	2	5,401,596	3.70E-07	0	18,690	0.00E+00
LCRA1938	37	Xp.12	25,398,297	C	A	KRAS	c.22G>T	p.V8L	Het	10	5,401,596	1.85E-06	0	18,212	0.00E+00
LCRA1938	38	Xp.12	25,398,297	C	T	KRAS	c.22G>A	p.V8I	Het	1	5,401,596	1.85E-07	0	18,212	0.00E+00
LCRA1938	39	Xp.12	25,398,298	C	A	KRAS	c.21G>C	p.V7V	Het	3	5,401,596	5.55E-07	0	17,926	0.00E+00
LCRA1938	40	Xp.12	25,398,298	C	G	KRAS	c.21G>C	p.V7V	Het	1	5,401,596	1.85E-07	0	17,926	0.00E+00
LCRA1938	41	Xp.12	25,398,298	C	T	KRAS	c.21G>A	p.V7V	Het	12	5,401,596	5.22E-06	0	17,926	0.00E+00
LCRA1938	1	Xp.17	7,573,695	A	C	TP53	c.1031T>G	P.L344R	Het	2	8,164,879	2.45E-07	0	26,650	0.00E+00
LCRA1938	2	Xp.17	7,573,696	A	G	TP53	c.1031T>C	P.L344P	Het	124	8,164,879	1.52E-05	0	26,650	0.00E+00
LCRA1938	3	Xp.17	7,573,697	G	A	TP53	c.1030C>T	P.L344L	Het	24	8,164,879	2.94E-06	0	26,543	0.00E+00
LCRA1938	4	Xp.17	7,573,697	G	C	TP53	c.1030C>G	P.L344V	Het	1	8,164,879	1.22E-07	0	26,543	0.00E+00
LCRA1938	5	Xp.17	7,573,697	G	T	TP53	c.1030C>A	P.L344M	Het	16	8,164,879	1.96E-06	0	26,543	0.00E+00
LCRA1938	6	Xp.17	7,573,698	C	A	TP53	c.1028G>T	P.E343D	Het	16	8,164,879	1.96E-06	0	26,026	0.00E+00
LCRA1938	7	Xp.17	7,573,698	C	G	TP53	c.1028G>C	P.E343D	Het	1	8,164,879	1.22E-07	0	26,026	0.00E+00
LCRA1938	8	Xp.17	7,573,698	C	T	TP53	c.1028G>A	P.E343E	Het	12	8,164,879	1.47E-06	0	25,028	0.00E+00
LCRA1938	9	Xp.17	7,573,699	T	C	TP53	c.1028A>G	P.E343G	Het	5	8,164,879	6.12E-07	0	28,993	0.00E+00
LCRA1938	10	Xp.17	7,574,000	C	A	TP53	c.1027G>T	p.E343*	Het	36	8,164,879	4.41E-06	0	28,895	0.00E+00
LCRA1938	11	Xp.17	7,574,000	C	T	TP53	c.1027G>A	p.E343K	Het	6	8,164,879	7.35E-07	0	28,895	0.00E+00
LCRA1938	12	Xp.17	7,574,001	T	A	TP53	c.1026A>T	P.R342R	Het	1	8,164,879	1.22E-07	0	28,748	0.00E+00
LCRA1938	13	Xp.17	7,574,001	T	C	TP53	c.1026A>G	P.R342R	Het	8	8,164,879	7.35E-07	0	28,748	0.00E+00
LCRA1938	14	Xp.17	7,574,002	C	A	TP53	c.1026G>T	P.R342L	Het	16	8,164,879	2.33E-06	0	28,606	0.00E+00
LCRA1938	15	Xp.17	7,574,002	C	T	TP53	c.1026G>A	P.R342Q	Het	754	8,164,879	6.23E-05	0	28,606	0.00E+00
LCRA1938	16	Xp.17	7,574,003	G	A	TP53	c.1024C>T	P.R342*	Het	594	8,164,879	6.91E-05	0	28,489	0.00E+00
LCRA1938	17	Xp.17	7,574,003	G	T	TP53	c.1024C>A	P.R342R	Het	14	8,164,879	1.71E-06	0	28,489	0.00E+00
LCRA1938	18	Xp.17	7,574,004	G	A	TP53	c.1023C>T	P.F341F	Het	12	8,164,879	1.47E-06	0	28,054	0.00E+00
LCRA1938	19	Xp.17	7,574,004	G	T	TP53	c.1023C>A	P.F341L	Het	93	8,164,879	1.14E-05	0	28,054	0.00E+00
LCRA1938	20	Xp.17	7,574,005	A	G	TP53	c.1022T>C	P.F341S	Het	33	8,164,879	4.04E-06	0	27,845	0.00E+00
LCRA1938	21	Xp.17	7,574,006	A	C	TP53	c.1021T>G	P.F341V	Het	1	8,164,879	1.22E-07	0	27,849	0.00E+00
LCRA1938	22	Xp.17	7,574,006	A	G	TP53	c.1021T>C	P.F341L	Het	10	8,164,879	1.22E-06	0	27,849	0.00E+00
LCRA1938	23	Xp.17	7,574,007	C	A	TP53	c.1020G>T	P.M340I	Het	8	8,164,879	8.89E-07	0	27,576	0.00E+00
LCRA1938	24	Xp.17	7,574,007	C	T	TP53	c.1020C>A	P.M340I	Het	1	8,164,879	1.22E-07	0	27,576	0.00E+00
LCRA1938	25	Xp.17	7,574,008	A	G	TP53	c.1019T>C	P.M340T	Het	1	8,164,879	1.22E-07	0	27,125	0.00E+00
LCRA1938	26	Xp.17	7,574,009	T	C	TP53	c.1018A>G	P.M340V	Het	2	8,164,879	2.45E-07	0	27,118	0.00E+00
LCRA1938	27	Xp.17	7,574,010	C	A	TP53	c.1017C>T	P.E339D	Het	5	8,164,879	6.12E-07	0	27,020	0.00E+00
LCRA1938	28	Xp.17	7,574,010	C	T	TP53	c.1017C>A	p.E339E	Het	3	8,164,879	3.67E-07	0	27,020	0.00E+00
LCRA1938	29	Xp.17	7,574,011	T	A	TP53	c.1016A>G	P.E339V	Het	2	8,164,879	2.45E-07	0	26,928	0.00E+00
LCRA1938	30	Xp.17	7,574,011	T	C	TP53	c.1016A>C	P.E339G	Het	31	8,164,879	3.89E-06	0	26,928	0.00E+00
LCRA1938	31	Xp.17	7,574,011	T	G	TP53	c.1015T>A	p.E339A	Het	1	8,164,879	1.22E-07	0	26,928	0.00E+00
LCRA1938	32	Xp.17	7,574,012	C	A	TP53	c.1015G>T	p.E339*	Het	19	8,164,879	2.33E-06	0	26,794	0.00E+00
LCRA1938	33	Xp.17	7,574,012	C	T	TP53	c.1015G>A	p.E339K	Het	432	8,164,879	6.29E-05	0	26,794	0.00E+00
LCRA1938	34	Xp.17	7,574,013	G	A	TP53	c.1014C>T	P.F338F	Het	299	8,164,879	3.66E-05	0	26,595	0.00E+00
LCRA1938	35	Xp.17	7,574,013	G	T	TP5									

LCRA 1938	45	Xp.17	7,574,018	G	T	TP53	c.1008C>A	p.R337S	No	20	8,164,879	2,45E-06	0	25,837	0.00E+00
LCRA 1938	46	Xp.17	7,574,019	C	A	TP53	c.1008G>T	p.E336D	No	2	8,164,879	2,45E-07	0	25,789	0.00E+00
LCRA 1938	47	Xp.17	7,574,019	C	T	TP53	c.1008G>A	p.E339E	No	1	8,164,879	8,57E-07	0	25,789	0.00E+00
LCRA 1938	48	Xp.17	7,574,020	T	A	TP53	c.1007A>T	p.E336V	No	7	8,164,879	1,22E-07	0	25,576	0.00E+00
LCRA 1938	49	Xp.17	7,574,020	T	C	TP53	c.1007G>T	p.E336S	No	34	8,164,879	4,16E-06	0	25,576	0.00E+00
LCRA 1938	50	Xp.17	7,574,021	C	A	TP53	c.1008G>T	p.E336*	No	19	8,164,879	2,33E-06	0	25,340	0.00E+00
LCRA 1938	51	Xp.17	7,574,021	C	G	TP53	c.1008G>C	p.E336Q	No	1	8,164,879	1,22E-07	0	25,340	0.00E+00
LCRA 1938	52	Xp.17	7,574,022	A	G	TP53	c.1005T>C	p.R335R	No	5	8,164,879	6,12E-07	0	25,255	0.00E+00
LCRA 1938	53	Xp.17	7,574,023	C	A	TP53	c.1004G>T	p.R335L	No	18	8,164,879	2,20E-06	0	25,128	0.00E+00
LCRA 1938	54	Xp.17	7,574,023	C	G	TP53	c.1004G>C	p.R335P	No	1	8,164,879	1,22E-07	0	25,128	0.00E+00
LCRA 1938	55	Xp.17	7,574,023	C	T	TP53	c.1004G>A	p.R335H	No	222	8,164,879	2,72E-05	0	25,128	0.00E+00
LCRA 1938	56	Xp.17	7,574,024	G	A	TP53	c.1003G>T	p.R335C	No	310	8,164,879	3,80E-05	0	25,060	0.00E+00
LCRA 1938	57	Xp.17	7,574,024	G	C	TP53	c.1003G>G	p.R335G	No	11	8,164,879	1,35E-06	0	25,060	0.00E+00
LCRA 1938	58	Xp.17	7,574,024	G	T	TP53	c.1003G>A	p.R335S	No	37	8,164,879	4,53E-06	0	25,060	0.00E+00
LCRA 1938	59	Xp.17	7,574,025	C	A	TP53	c.1002G>T	p.G334G	No	13	8,164,879	1,59E-06	0	24,841	0.00E+00
LCRA 1938	60	Xp.17	7,574,025	C	G	TP53	c.1002G>C	p.G334G	No	2	8,164,879	2,45E-07	0	24,841	0.00E+00
LCRA 1938	61	Xp.17	7,574,025	C	T	TP53	c.1002G>A	p.G334G	No	7	8,164,879	2,45E-06	0	24,841	0.00E+00
LCRA 1938	62	Xp.17	7,574,026	C	A	TP53	c.1001G>T	p.G334V	No	20	8,164,879	8,57E-07	0	24,771	0.00E+00
LCRA 1938	63	Xp.17	7,574,026	C	T	TP53	c.1001G>A	p.G334E	No	6	8,164,879	7,35E-07	0	24,771	0.00E+00
LCRA 1938	64	Xp.17	7,574,027	C	A	TP53	c.1000G>T	p.G334W	No	22	8,164,879	2,69E-06	0	24,567	0.00E+00
LCRA 1938	65	Xp.17	7,574,027	C	T	TP53	c.1000G>A	p.G334R	No	9	8,164,879	1,10E-06	0	24,567	0.00E+00
LCRA 1938	66	Xp.17	7,574,028	A	G	TP53	c.999T>C	p.R333R	No	43	8,164,879	5,27E-06	0	24,061	0.00E+00
LCRA 1938	67	Xp.17	7,574,029	C	A	TP53	c.998G>T	p.R333L	No	13	8,164,879	1,59E-06	0	23,693	0.00E+00
LCRA 1938	68	Xp.17	7,574,029	C	T	TP53	c.998G>A	p.R333H	No	365	8,164,879	4,47E-05	0	23,693	0.00E+00
LCRA 1938	69	Xp.17	7,574,036	G	-	TP53	В интроне	В интроне	No	1	8,164,879	1,22E-07	0	22,874	0.00E+00
LCRA 1938	NA	Xp.17	7,574,123	G	A	TP53	В интроне	В интроне	No	NA	NA	NA	0	3,033	3.30E-04
LCRA 1947	1	Xp.5	112,175,203	A	G	APC	c.p112A>G	p.I1304M	No	19	4,868,437	3,89E-06	0	11,927	0.00E+00
LCRA 1947	2	Xp.5	112,175,204	G	A	APC	c.3913G>A	p.A1305T	No	16	4,868,437	3,27E-06	0	12,152	0.00E+00
LCRA 1947	3	Xp.5	112,175,204	G	C	APC	c.3913G>C	p.A1305S	No	1	4,868,437	2,05E-07	0	12,152	0.00E+00
LCRA 1947	4	Xp.5	112,175,204	G	T	APC	c.3913G>T	p.A1305S	No	17	4,868,437	3,48E-06	0	12,152	0.00E+00
LCRA 1947	5	Xp.5	112,175,205	C	A	APC	c.3914C>A	p.A1305E	No	15	4,868,437	3,07E-06	0	12,232	0.00E+00
LCRA 1947	6	Xp.5	112,175,205	C	T	APC	c.3914C>T	p.A1305V	No	29	4,868,437	5,93E-06	0	12,232	0.00E+00
LCRA 1947	7	Xp.5	112,175,206	A	C	APC	c.3915A>C	p.A1305A	No	14	4,868,437	2,87E-06	0	12,267	0.00E+00
LCRA 1947	8	Xp.5	112,175,206	A	G	APC	c.3915A>G	p.A1305A	No	57	4,868,437	1,17E-05	0	12,267	0.00E+00
LCRA 1947	9	Xp.5	112,175,206	A	T	APC	c.3915A>T	p.A1305A	No	5	4,868,437	1,02E-06	0	12,267	0.00E+00
LCRA 1947	10	Xp.5	112,175,207	G	A	APC	c.3916G>C	p.E1306Q	No	155	4,868,437	3,17E-05	0	12,467	0.00E+00
LCRA 1947	11	Xp.5	112,175,207	G	T	APC	c.3916G>T	p.E1306*	No	2	4,868,437	4,09E-07	0	12,467	0.00E+00
LCRA 1947	12	Xp.5	112,175,208	A	G	APC	c.3917A>G	p.E1306G	No	8	4,868,437	1,64E-06	0	12,516	0.00E+00
LCRA 1947	13	Xp.5	112,175,210	A	G	APC	c.3918A>G	p.I1307V	No	1	4,868,437	2,05E-07	0	12,553	0.00E+00
LCRA 1947	14	Xp.5	112,175,212	A	G	APC	c.3921A>G	p.I1307M	No	8	4,868,437	1,84E-06	0	12,654	0.00E+00
LCRA 1947	15	Xp.5	112,175,212	AAAAA	-	APC	c.3921_3925del	p.M3076	Yes	1,871	4,868,437	3,42E-04	8	12,654	6.32E-04
LCRA 1947	16	Xp.5	112,175,214	A	G	APC	c.3923A>G	p.K1308R	No	2	4,868,437	4,09E-07	0	12,825	0.00E+00
LCRA 1947	17	Xp.5	112,175,215	A	C	APC	c.3924A>C	p.K1308R	No	1	4,868,437	2,05E-07	0	12,874	0.00E+00
LCRA 1947	18	Xp.5	112,175,215	A	G	APC	c.3924A>G	p.K1308K	No	9	4,868,437	1,84E-06	0	12,874	0.00E+00
LCRA 1947	19	Xp.5	112,175,216	A	T	APC	c.3925G>T	p.E1309*	No	1	4,868,437	2,05E-07	0	13,012	0.00E+00
LCRA 1947	20	Xp.5	112,175,217	A	G	APC	c.3926A>G	p.E1309G	No	1	4,868,437	2,05E-07	0	13,043	0.00E+00
LCRA 1947	21	Xp.5	112,175,218	A	G	APC	c.3927A>G	p.E1309E	No	1	4,868,437	2,05E-07	0	13,086	0.00E+00
LCRA 1947	22	Xp.5	112,175,219	A	G	APC	c.3928A>G	p.K1310E	No	1	4,868,437	2,05E-07	0	13,123	0.00E+00
LCRA 1947	23	Xp.5	112,175,220	A	C	APC	c.3929A>C	p.K1310T	No	1	4,868,437	2,05E-07	0	13,141	0.00E+00
LCRA 1947	24	Xp.5	112,175,220	A	G	APC	c.3929A>G	p.K1310R	No	4	4,868,437	8,19E-07	0	13,141	0.00E+00
LCRA 1947	25	Xp.5	112,175,221	G	T	APC	c.3930G>T	p.K1310N	No	4	4,868,437	4,09E-07	0	13,216	0.00E+00
LCRA 1947	1	Xp.12	25,398,277	T	A	KRAS	c.42A>T	p.V14V	No	2	2,527,462	1,58E-06	0	10,238	0.00E+00
LCRA 1947	2	Xp.12	25,398,277	T	C	KRAS	c.42A>G	p.V14V	No	16	2,527,462	6,33E-06	0	10,238	0.00E+00
LCRA 1947	3	Xp.12	25,398,277	T	G	KRAS	c.42A>C	p.V14V	No	1	2,527,462	3,96E-07	0	10,238	0.00E+00
LCRA 1947	4	Xp.12	25,398,278	A	G	KRAS	c.41T>C	p.V14A	No	1	2,527,462	3,96E-07	0	10,221	0.00E+00
LCRA 1947	5	Xp.12	25,398,279	C	A	KRAS	c.40G>T	p.V14L	No	5	2,527,462	1,98E-06	0	10,126	0.00E+00
LCRA 1947	6	Xp.12	25,398,279	C	T	KRAS	c.40G>A	p.V14I	No	250	2,527,462	9,89E-05	0	10,126	0.00E+00
LCRA 1947	7	Xp.12	25,398,280	G	A	KRAS	c.39C>T	p.G13G	No	9	2,527,462	3,56E-06	0	10,093	0.00E+00
LCRA 1947	8	Xp.12	25,398,280	G	T	KRAS	c.39C>A	p.G13G	No	6	2,527,462	2,37E-06	0	10,093	0.00E+00
LCRA 1947	9	Xp.12	25,398,281	C	A	KRAS	c.39C>T	p.G13V	No	8	2,527,462	3,17E-06	0	10,050	0.00E+00
LCRA 1947	10	Xp.12	25,398,281	C	T	KRAS	c.39C>A	p.G13D	No	84	2,527,462	3,32E-05	0	10,050	0.00E+00
LCRA 1947	11	Xp.12	25,398,282	C	A	KRAS	c.37G>T	p.G13C	No	105	2,527,462	4,15E-05	0	9,879	0.00E+00
LCRA 1947	12	Xp.12	25,398,282	C	G	KRAS	c.37G>C	p.G13R	No	18	2,527,462	7,12E-06	0	9,879	0.00E+00
LCRA 1947	13	Xp.12	25,398,283	A	G	KRAS	c.36T>C	p.G12G	No	102	2,527,462	4,04E-05	0	9,839	0.00E+00
LCRA 1947	14	Xp.12	25,398,284	C	-	KRAS	c.35del	p.G12fs	No	117	2,527,462	4,63E-05	0	9,810	0.00E+00
LCRA 1947	15	Xp.12	25,398,284	C	A	KRAS	c.35G>T	p.G12V	Yes	1,171	2,527,462	4,63E-04	5	9,810	5.10E-04
LCRA 1947	16	Xp.12	25,398,284	C	G	KRAS	c.36C>C	p.G12A	No	212	2,527,462	8,39E-05	0	9,810	0.00E+00
LCRA 1947	17	Xp.12	25,398,284	C	T	KRAS	c.36G>T	p.G12D	No	117	2,527,462	4,32E-05	0	9,810	0.00E+00
LCRA 1947	18	Xp.12	25,398,285	C	A	KRAS	c.34G>T	p.G12C	No	2	2,527,462	7,91E-07	0	9,742	0.00E+00
LCRA 1947	19	Xp.12	25,398,285	C	T	KRAS	c.34G>A	p.G12S	No	2	2,527,462	7,91E-07	0	9,742	0.00E+00
LCRA 1947	20	Xp.12	25,398,286	A	C	KRAS	c.33T>G	p.A11A	No	1	2,527,462	3,96E-07	0	9,695	0.00E+00
LCRA 1947	21	Xp.12	25,398,286	A	G	KRAS	c.33T>C	p.A11A	No	22	2,527,462	8,70E-06	0	9,695	0.00E+00
LCRA 1947	22	Xp.12	25,398,287	G	A	KRAS	c.32C>T	p.A11V	No	7	2,527,462	2,77E-06	0	9,689	0.00E+00
LCRA 1947	23	Xp.12	25,398,287	G	T	KRAS	c.32C>A	p.A11D	No	3	2,527,462	1,19E-06	0	9,689	0.00E+00
LCRA 1947	24	Xp.12	25,398,288	C	A	KRAS	c.31C>T	p.A11S	No	3	2,527,462	1,19E-06	0	9,629	0.00E+00
LCRA 1947	25	Xp.12	25,398,288	C	T	KRAS	c.31C>G	p.A11D	No	8	2,527,462	7,91E-07	0	9,629	0.00E+00
LCRA 1947	26	Xp.12	25,398,288	C	T	KRAS	c.31G>A	p.A11T	No	1	2,527,462	3,96E-07	0	9,629	0.00E+00
LCRA 1947	27	Xp.12	25,398,289	T	A	KRAS	c.30A>T	p.G10G	No	18	2,527,462	7,12E-06	0	9,412	0.00E+00
LCRA 1947	28	Xp.12	25,398,289	T	C	KRAS	c.30A>G	p.G10G	No	114	2,527,462	4,51E-05	0	9,412	0.00E+00
LCRA 1947	29	Xp.12	25,398,289												

LCRA1947	39	Xp.12	25,388,296	A	G	KRAS	c.23T>C	p.V8A	Het	8	2,527,462	3.17E-06	0	8,910	0.00E+00
LCRA1947	40	Xp.12	25,388,297	C	-		c.22del	p.V8fs	Het	1	2,527,462	3.96E-07	0	8,882	0.00E+00
LCRA1947	41	Xp.12	25,388,297	C	A	KRAS	c.22G>T	p.V8L	Het	1	2,527,462	3.96E-07	0	8,882	0.00E+00
LCRA1947	42	Xp.12	25,388,297	C	T	KRAS	c.22G>A	p.V8I	Het	3	2,527,462	1.19E-06	0	8,882	0.00E+00
LCRA1947	43	Xp.12	25,388,298	C	A	KRAS	c.21G>T	p.V7V	Het	117	2,527,462	4.63E-05	0	8,757	0.00E+00
LCRA1947	44	Xp.12	25,388,298	C	T	KRAS	c.21G>A	p.V7V	Het	41	2,527,462	1.62E-05	0	8,757	0.00E+00
LCRA1947	1	Xp.17	7,578,398	G	-	TP53	c.532del	p.H178fs	Het	54	4,490,801	1.20E-05	0	12,277	0.00E+00
LCRA1947	2	Xp.17	7,578,398	GG	-	TP53	c.531_S52del	p.H178fs	Het	1	4,490,801	2.23E-07	0	12,277	0.00E+00
LCRA1947	3	Xp.17	7,578,398	G	A	TP53	c.531C>T	p.P177P	Het	32	4,490,801	7.13E-06	0	12,289	0.00E+00
LCRA1947	4	Xp.17	7,578,398	G	C	TP53	c.531C>G	p.P177P	Het	1	4,490,801	2.23E-07	0	12,289	0.00E+00
LCRA1947	5	Xp.17	7,578,398	G	T	TP53	c.531C>A	p.P177P	Het	1	4,490,801	2.23E-07	0	12,289	0.00E+00
LCRA1947	6	Xp.17	7,578,400	G	A	TP53	c.530C>T	p.P177L	Het	7	4,490,801	1.56E-06	0	12,207	0.00E+00
LCRA1947	7	Xp.17	7,578,400	G	C	TP53	c.530C>G	p.P177R	Het	1	4,490,801	2.23E-07	0	12,207	0.00E+00
LCRA1947	8	Xp.17	7,578,400	G	T	TP53	c.530C>A	p.P177H	Het	2	4,490,801	4.45E-07	0	12,207	0.00E+00
LCRA1947	9	Xp.17	7,578,401	G	A	TP53	c.529C>T	p.P177S	Het	7	4,490,801	1.56E-06	0	12,189	0.00E+00
LCRA1947	10	Xp.17	7,578,401	G	T	TP53	c.529C>A	p.P177T	Het	7	4,490,801	1.56E-06	0	12,189	0.00E+00
LCRA1947	11	Xp.17	7,578,402	G	A	TP53	c.5280T	p.C176C	Het	41	4,490,801	9.13E-06	0	12,142	0.00E+00
LCRA1947	12	Xp.17	7,578,402	G	T	TP53	c.5280A	p.C176-	Het	10	4,490,801	2.23E-06	0	12,142	0.00E+00
LCRA1947	13	Xp.17	7,578,403	C	A	TP53	c.527G>T	p.C176F	Het	9	4,490,801	2.00E-06	0	12,044	0.00E+00
LCRA1947	14	Xp.17	7,578,403	C	T	TP53	c.527G>A	p.C176Y	Het	35	4,490,801	7.79E-06	0	12,044	0.00E+00
LCRA1947	15	Xp.17	7,578,404	A	G	TP53	c.526T>C	p.C176R	Het	89	4,490,801	1.54E-05	0	11,985	0.00E+00
LCRA1947	16	Xp.17	7,578,405	G	A	TP53	c.5260T	p.R175R	Het	10	4,490,801	2.23E-06	0	11,811	0.00E+00
LCRA1947	17	Xp.17	7,578,405	G	T	TP53	c.5260A	p.R175R	Het	1	4,490,801	2.23E-07	0	11,811	0.00E+00
LCRA1947	18	Xp.17	7,578,406	C	A	TP53	c.524G>T	p.R175L	Het	7	4,490,801	1.56E-06	0	11,801	0.00E+00
LCRA1947	19	Xp.17	7,578,406	C	G	TP53	c.524G>C	p.R175P	Het	2	4,490,801	4.45E-07	0	11,801	0.00E+00
LCRA1947	20	Xp.17	7,578,406	C	T	TP53	c.524G>A	p.R175H	Дя	2,588	4,490,801	7.99E-04	8	11,801	6.78E-04
LCRA1947	21	Xp.17	7,578,407	G	A	TP53	c.523C>T	p.R175C	Het	78	4,490,801	1.74E-05	0	11,776	0.00E+00
LCRA1947	22	Xp.17	7,578,407	G	T	TP53	c.523C>A	p.R175S	Het	12	4,490,801	2.67E-06	0	11,776	0.00E+00
LCRA1947	23	Xp.17	7,578,408	C	A	TP53	c.522G>T	p.R174S	Het	5	4,490,801	1.11E-06	0	11,713	0.00E+00
LCRA1947	24	Xp.17	7,578,408	C	T	TP53	c.522G>A	p.R174R	Het	9	4,490,801	2.00E-06	0	11,713	0.00E+00
LCRA1947	25	Xp.17	7,578,409	C	A	TP53	c.521G>T	p.R174M	Het	96	4,490,801	2.14E-05	0	11,700	0.00E+00
LCRA1947	26	Xp.17	7,578,409	C	G	TP53	c.521G>C	p.R174T	Het	1	4,490,801	2.23E-07	0	11,700	0.00E+00
LCRA1947	27	Xp.17	7,578,409	C	T	TP53	c.521G>A	p.R174K	Het	18	4,490,801	4.01E-06	0	11,700	0.00E+00
LCRA1947	28	Xp.17	7,578,410	T	A	TP53	c.520A>T	p.R174W	Het	41	4,490,801	9.13E-06	0	11,665	0.00E+00
LCRA1947	29	Xp.17	7,578,410	T	C	TP53	c.520A>G	p.R174G	Het	579	4,490,801	1.28E-04	0	11,665	0.00E+00
LCRA1947	30	Xp.17	7,578,410	T	G	TP53	c.520A>C	p.R174R	Het	13	4,490,801	2.69E-06	0	11,665	0.00E+00
LCRA1947	31	Xp.17	7,578,411	C	A	TP53	c.519G>T	p.V173V	Het	2	4,490,801	4.45E-07	0	11,633	0.00E+00
LCRA1947	32	Xp.17	7,578,411	C	T	TP53	c.519G>A	p.V173V	Het	11	4,490,801	2.45E-06	0	11,633	0.00E+00
LCRA1947	33	Xp.17	7,578,412	A	G	TP53	c.518T>C	p.V173A	Het	100	4,490,801	2.23E-05	0	11,525	0.00E+00
LCRA1947	34	Xp.17	7,578,412	A	T	TP53	c.518T>A	p.V173E	Het	1	4,490,801	2.23E-07	0	11,525	0.00E+00
LCRA1947	35	Xp.17	7,578,413	C	A	TP53	c.517G>T	p.V173L	Het	6	4,490,801	1.34E-06	0	11,507	0.00E+00
LCRA1947	36	Xp.17	7,578,413	C	T	TP53	c.517G>A	p.V173M	Het	10	4,490,801	2.23E-06	0	11,507	0.00E+00
LCRA1947	37	Xp.17	7,578,414	A	G	TP53	c.516T>C	p.V172V	Het	47	4,490,801	11.37E-06	0	11,375	0.00E+00
LCRA1947	38	Xp.17	7,578,414	A	T	TP53	c.516T>A	p.V172A	Het	2	4,490,801	4.45E-07	0	11,375	0.00E+00
LCRA1947	39	Xp.17	7,578,415	A	G	TP53	c.515T>C	p.V172A	Het	4	4,490,801	8.91E-07	0	11,320	0.00E+00
LCRA1947	40	Xp.17	7,578,415	A	T	TP53	c.515T>A	p.V172D	Het	1	4,490,801	2.23E-07	0	11,320	0.00E+00
LCRA1947	41	Xp.17	7,578,416	C	A	TP53	c.514G>T	p.V172F	Het	7	4,490,801	1.56E-06	0	11,305	0.00E+00
LCRA1947	42	Xp.17	7,578,416	C	T	TP53	c.514G>A	p.V172I	Het	2	4,490,801	4.45E-07	0	11,305	0.00E+00
LCRA1947	43	Xp.17	7,578,417	C	A	TP53	c.513G>T	p.E171D	Het	4	4,490,801	8.91E-07	0	11,246	0.00E+00
LCRA1947	44	Xp.17	7,578,417	C	T	TP53	c.513G>A	p.E171E	Het	10	4,490,801	2.23E-06	0	11,246	0.00E+00
LCRA1947	45	Xp.17	7,578,418	T	C	TP53	c.512A>G	p.E171G	Het	16	4,490,801	3.56E-06	0	11,185	0.00E+00
LCRA1947	46	Xp.17	7,578,419	C	A	TP53	c.511G>T	p.E171*	Het	2	4,490,801	4.45E-07	0	11,147	0.00E+00
LCRA1947	47	Xp.17	7,578,419	C	T	TP53	c.511G>A	p.E171K	Het	5	4,490,801	1.11E-06	0	11,147	0.00E+00
LCRA1947	48	Xp.17	7,578,420	C	A	TP53	c.510G>T	p.T170T	Het	11	4,490,801	2.45E-06	0	11,058	0.00E+00
LCRA1947	49	Xp.17	7,578,420	C	T	TP53	c.510G>A	p.T170T	Het	166	4,490,801	3.70E-05	0	11,058	0.00E+00
LCRA1947	50	Xp.17	7,578,421	G	A	TP53	c.509C>T	p.T170M	Het	92	4,490,801	2.05E-05	0	11,038	0.00E+00
LCRA1947	51	Xp.17	7,578,421	G	T	TP53	c.509C>A	p.T170K	Het	9	4,490,801	2.00E-06	0	11,038	0.00E+00
LCRA1947	52	Xp.17	7,578,422	T	C	TP53	c.508A>G	p.T170A	Het	147	4,490,801	3.27E-05	0	10,975	0.00E+00
LCRA1947	53	Xp.17	7,578,423	C	A	TP53	c.507G>T	p.M169I	Het	5	4,490,801	1.11E-06	0	10,932	0.00E+00
LCRA1947	54	Xp.17	7,578,423	C	G	TP53	c.507C>C	p.M169H	Het	1	4,490,801	2.23E-07	0	10,932	0.00E+00
LCRA1947	55	Xp.17	7,578,423	C	T	TP53	c.507G>A	p.M169H	Het	6	4,490,801	1.34E-06	0	10,932	0.00E+00
LCRA1947	56	Xp.17	7,578,424	A	G	TP53	c.506T>C	p.M169T	Het	33	4,490,801	7.35E-06	0	10,916	0.00E+00
LCRA1947	57	Xp.17	7,578,425	T	C	TP53	c.505A>G	p.M169V	Het	11	4,490,801	2.45E-06	0	10,884	0.00E+00
LCRA1947	58	Xp.17	7,578,426	G	A	TP53	c.504C>T	p.H189H	Het	8	4,490,801	1.78E-06	0	10,813	0.00E+00
LCRA1947	59	Xp.17	7,578,426	G	T	TP53	c.504C>A	p.H189Q	Het	12	4,490,801	2.67E-06	0	10,813	0.00E+00
LCRA1947	60	Xp.17	7,578,427	T	C	TP53	c.503A>C	p.H189R	Het	1	4,490,801	3.13E-06	0	10,773	0.00E+00
LCRA1947	61	Xp.17	7,578,428	G	A	TP53	c.502C>T	p.H188Y	Het	25	4,490,801	5.57E-06	0	10,751	0.00E+00
LCRA1947	62	Xp.17	7,578,428	C	T	TP53	c.502C>A	p.H188H	Het	7	4,490,801	1.56E-06	0	10,751	0.00E+00
LCRA1947	63	Xp.17	7,578,429	C	A	TP53	c.501G>T	p.Q167H	Het	5	4,490,801	1.11E-06	0	10,648	0.00E+00
LCRA1947	64	Xp.17	7,578,429	C	G	TP53	c.501G>C	p.Q167H	Het	4	4,490,801	8.91E-07	0	10,648	0.00E+00
LCRA1947	65	Xp.17	7,578,429	C	T	TP53	c.501G>A	p.Q167Q	Het	160	4,490,801	3.56E-05	0	10,648	0.00E+00
LCRA1947	66	Xp.17	7,578,430	T	A	TP53	c.500A>T	p.Q167L	Het	15	4,490,801	3.34E-06	0	10,570	0.00E+00
LCRA1947	67	Xp.17	7,578,430	T	C	TP53	c.500A>G	p.Q167R	Het	487	4,490,801	1.08E-04	0	10,570	0.00E+00
LCRA1947	68	Xp.17	7,578,430	T	G	TP53	c.500A>C	p.Q167P	Het	13	4,490,801	2.89E-06	0	10,570	0.00E+00
LCRA1947	69	Xp.17	7,578,430	TG	-	TP53	c.499_500del	p.Q167fs	Het	1	4,490,801	2.23E-07	0	10,570	0.00E+00
LCRA1947	70	Xp.17	7,578,431	G	A	TP53	c.499C>T	p.Q167*	Het	33	4,490,801	7.35E-06	0	10,554	0.00E+00
LCRA1947	71	Xp.17	7,578,431	G	T	TP53	c.499C>A	p.Q167K	Het	16	4,490,801	3.56E-06	0	10,554	0.00E+00
LCRA1947	72	Xp.17	7,578,432	G	C	TP53	c.498A>G	p.S166S	Het	30	4,490,801	6.69E-06	0	10,404	0.00E+00
LCRA1947	73	Xp.17	7,578,433	G	A	TP53	c.497C>T	p.S166L	Het	8	4,490,801	1.78E-06	0	10,382	0.00E+00
LCRA1947	74	Xp.17													

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ, включающий:

а) присоединение частично двухцепочечных 3'-адаптеров (3'PDSA) к 3'-концам обеих цепей Уотсона и Крика популяции фрагментов двухцепочечной ДНК в анализируемом образце ДНК,

где первая цепь 3'PDSA-адаптеров включает, в направлении 5'→3': (i) первый сегмент, (ii) экзогенную UID последовательность, (iii) сайт отжига для 5'-адаптера, и (iv) универсальную 3'-адаптерную последовательность, включающую сайт праймера для секвенирования R2, и

где вторая цепь 3'PDSA-адаптеров включает, в направлении 5'→3': (i) сегмент, комплементарный первому сегменту, и (ii) 3'-блокирующую группу,

б) отжиг 5'-адаптеров на указанном сайте отжига, где 5'-адаптеры включают, в направлении 5'→3': (i) универсальную 5'-адаптерную последовательность, которая не комплементарна универсальной 3'-адаптерной последовательности, и которая включает сайт праймера для секвенирования R1, и (ii) последовательность, комплементарную сайту отжига для 5'-адаптера;

с) удлинение 5'-адаптеров через экзогенную UID последовательность и указанный первый сегмент, с получением в результате комплемента указанной экзогенной UID последовательности и комплемента указанного первого сегмента, и

д) ковалентное связывание 3'-конца указанного комплемента указанного первого сегмента с 5'-концом цепей Уотсона и Крика фрагментов двухцепочечной ДНК, с получением в результате множества лигированных с адаптером фрагментов двухцепочечной ДНК.

2. Способ по п.1, дополнительно включающий: амплификацию указанного множества лигированных с адаптером фрагментов двухцепочечной ДНК с использованием первого праймера, комплементарного указанной универсальной 3'-адаптерной последовательности, и второго праймера, комплементарного комплементу указанной универсальной 5'-адаптерной последовательности, с получением в результате ампликонов, где указанные ампликоны включают множество двухцепочечных матриц Уотсона и множество двухцепочечных матриц Крика.

3. Способ по п.2, дополнительно включающий: селективную амплификацию указанных двухцепочечных матриц Уотсона с использованием первого набора пар мишень-селективных праймеров Уотсона, причем первый набор пар мишень-селективных праймеров Уотсона включает: (i) первый мишень-селективный праймер Уотсона, включающий последовательность, комплементарную части универсальной 3'-адаптерной последовательности, и (ii) второй мишень-селективный праймер Уотсон, включающий мишень-селективную последовательность, с получением в результате целевых продуктов амплификации Уотсона.

4. Способ по п.3, дополнительно включающий: селективную амплификацию указанных двухцепочечных матриц Крика с использованием первого набора пар мишень-

селективных праймеров Крика, причем первый набор пар мишень-селективных праймеров Крика включает: (i) первый мишень-селективный праймер Крика, включающий последовательность, комплементарную комплементу части универсальной 5'-адаптерной последовательности, и (ii) второй мишень-селективный праймер Крика, включающий такую же мишень-селективную последовательность, как и последовательность второго мишень-селективного праймера Уотсона, с получением в результате целевых продуктов амплификации Крика.

5. Способ по п.1, дополнительно включающий: удаление указанной второй цепи указанных 3'PDSA-адаптеров с получением одноцепочечных 3'-адаптеров (3'SSA-адаптеров).

6. Способ по п.5, где указанное удаление указанной второй цепи происходит после этапа b) или перед этапом b), или во время этапа b).

7. Способ по п.5, где указанная вторая цепь включает один или больше дезоксиуридинов, и где указанное удаление указанной второй цепи указанных 3'PDSA-адаптеров включает контакт 3'-дуплексного адаптера с урацил-ДНК-гликозилазой (УДГ), с расщеплением указанной второй цепи.

8. Способ по п.5, где указанное удаление указанной второй цепи производят с помощью полимеразы с экзонуклеазной активностью, где указанная полимераза удлиняет 5'-адаптеры через экзогенную UID последовательность и указанный первый сегмент.

9. Способ по п.2, дополнительно включающий: определение чтений последовательности одного или больше указанных ампликонов.

10. Способ по п.9, дополнительно включающий: отнесение чтений последовательности к семействам UID, где каждый член семейства UID включает одну и ту же экзогенную UID последовательность.

11. Способ по п.10, дополнительно включающий: отнесение чтений последовательности каждого семейства UID к подсемейству Уотсона и подсемейству Крика на основе пространственных отношений экзогенной UID последовательности к последовательности чтений R1 и R2.

12. Способ по п.11, дополнительно включающий: идентификацию нуклеотидной последовательности как точно представляющей цепь Уотсона анализируемого фрагмента ДНК, если по меньшей мере 50% подсемейства Уотсона содержит такую последовательность.

13. Способ по п.12, дополнительно включающий: идентификацию нуклеотидной последовательности как точно представляющей цепь Крика анализируемого фрагмента ДНК, если по меньшей мере 50% подсемейства Крика содержит такую последовательность.

14. Способ по п.12, дополнительно включающий: идентификацию мутации в нуклеотидной последовательности как точно представляющей цепь Уотсона, если последовательность, точно представляющая цепь Уотсона, отличается от референсной последовательности, в которой такая мутация отсутствует.

15. Способ по п.14, дополнительно включающий: идентификацию мутации в нуклеотидной последовательности как точно представляющей цепь Крика, если последовательность, точно представляющая цепь Крика, отличается от референсной последовательности, в которой такая мутация отсутствует.

16. Способ по п.15, дополнительно включающий: идентификацию мутации в анализируемом фрагменте ДНК, если мутация в нуклеотидной последовательности, точно представляющей цепь Уотсона, и мутация в нуклеотидной последовательности, точно представляющей цепь Крика, является одной и той же мутацией.

17. Способ по п.10, где каждый член семейства UID дополнительно включает одну и ту же эндогенную UID последовательность, где эндогенная UID последовательность включает конец фрагмента двухцепочечной ДНК из популяции.

18. Способ по п.1, где указанная популяция фрагментов двухцепочечной ДНК имеет тупые концы.

19. Система, включающая:

а) популяцию частично двухцепочечных 3'-адаптеров (3'PDSA-адаптеров) с возможностью лигирования с 3'-концами обеих цепей Уотсона и Крика популяции фрагментов двухцепочечной ДНК,

где первая цепь 3'PDSA-адаптеров включает, в направлении 5'→3': (i) первый сегмент, (ii) экзогенную UID последовательность, (iii) сайт отжига для 5'-адаптера, и (iv) универсальную 3'-адаптерную последовательность, включающую сайт праймера для секвенирования R2, и

где вторая цепь 3'PDSA-адаптеров включает, в направлении 5'→3': (i) сегмент, комплементарный первому сегменту, и (ii) 3'-блокирующую группу; и

б) популяцию 5'-адаптеров с возможностью отжига на указанном сайте отжига, где 5'-адаптеры включают, в направлении 5'→3': (i) универсальную 5'-адаптерную последовательность, которая не комплементарна универсальной 3'-адаптерной последовательности, и которая включает сайт праймера для секвенирования R1, и (ii) последовательность, комплементарную сайту отжига для 3'-адаптера.

20. Система по п.19, дополнительно включающая: с) указанную популяцию фрагментов двухцепочечной ДНК из биологического образца.

21. Система по п.20, где указанная популяция фрагментов двухцепочечной ДНК имеет тупые концы.

22. Система по п.19, дополнительно включающая: с) реагенты для расщепления указанной второй цепи указанных 3'PDSA-адаптеров с получением одноцепочечных 3'-адаптеров (3'SSA-адаптеров).

23. Система по п.19, дополнительно включающая: с) первый праймер, комплементарный указанной универсальной 3'-адаптерной последовательности, и второй праймер, комплементарный комплементу указанной универсальной 5'-адаптерной последовательности.

24. Система по п.19, дополнительно включающая: с) якорный праймер Уотсона,

комплементарный указанной универсальной 3'-адаптерной последовательности, и d) якорный праймер Крика, комплементарный комплементу указанной универсальной 5'-адаптерной последовательности.

25. Система по п.19, дополнительно включающая:

c) первый набор пар мишень-селективных праймеров Уотсона, включающий: (i) один или больше первых мишень-селективных праймеров Уотсона, включающих последовательность, комплементарную части универсальной 3'-адаптерной последовательности, и (ii) один или больше вторых мишень-селективных праймеров Уотсона, где каждый из одного или больше вторых мишень-селективных праймеров Уотсона включает мишень-селективную последовательность, и

d) первый набор пар мишень-селективных праймеров Крика, включающий: (i) один или больше мишень-селективных праймеров Крика, включающих последовательность, комплементарную комплементу части универсальной 5'-адаптерной последовательности, и (ii) один или больше вторых мишень-селективных праймеров Крика, где каждый из одного или больше вторых мишень-селективных праймеров Крика включает такую же мишень-селективную последовательность, как и последовательность второго мишень-селективного праймера Уотсона.

26. Способ, включающий:

a) формирование реакционной смеси, включающей:

i) множество фрагментов двухцепочечной ДНК, которые дефосфорилированы и имеют тупые концы, где каждый из указанных фрагментов двухцепочечной ДНК включает цепи Уотсона и Крика;

ii) множество адаптеров, где каждый из указанных адаптеров включает, в направлении 5'→3': А) штрих-код и В) универсальную 3'-адаптерную последовательность; и

iii) лигазу; и

b) инкубирование указанной реакционной смеси таким образом, что: i) адаптеры лигируются с 3'-концами цепей Уотсона и Крика, и ii) адаптеры не лигируются с 5'-концами цепей Уотсона или цепей Крика, с получением в результате двухцепочечных продуктов лигирования.

27. Способ по п.26, где каждое указанное множество адаптеров включает уникальный штрих-код.

28. Способ по п.27, где каждый из указанных двухцепочечных продуктов лигирования включает цепь Уотсона только с одним штрих-кодом и цепь Крика только с одним штрих-кодом, отличающимся от указанного штрих-кода на указанной цепи Уотсона.

29. Способ обнаружения присутствия или отсутствия мутации в области-мишени двухцепочечной ДНК-матрицы, полученной из образца млекопитающего, и определения, присутствует ли мутация в обеих цепях двухцепочечной ДНК-матрицы, где способ включает:

А) получение фрагментов двухцепочечной ДНК, каждый из которых имеет дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК;

В) амплификацию фрагмента двухцепочечной ДНК, включающего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК с получением библиотеки амплифицированных дуплексов для секвенирования, где амплификация включает контакт фрагмента двухцепочечной ДНК, включающего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, с парой универсальных праймеров в условиях полногеномной ПЦР;

С) необязательно, создание библиотеки одноцепочечных ДНК цепей Уотсона из библиотеки амплифицированных дуплексов для секвенирования;

Д) необязательно, создание библиотеки одноцепочечных ДНК цепей Крика из библиотеки амплифицированных дуплексов для секвенирования;

Е) амплификацию области-мишени с библиотеки ДНК цепей Уотсона при использовании пары праймеров, включающей первый праймер, способный к гибридизации с областью-мишенью, и второй праймер, способный к гибридизации с 3'-дуплексным адаптером;

Ф) амплификацию области-мишени с библиотеки ДНК цепей Крика при использовании пары праймеров, включающей первый праймер, способный к гибридизации с областью-мишенью, и второй праймер, способный к гибридизации с 5'-адаптером;

Г) секвенирование области-мишени, амплифицированной с библиотеки ДНК цепей Уотсона с получением чтения секвенирования и обнаружением присутствия или отсутствия мутации в цепи Уотсона области-мишени;

Н) секвенирование области-мишени, амплифицированной с библиотеки ДНК цепей Крика с получением чтения секвенирования и обнаружением присутствия или отсутствия мутации в цепи Крика области-мишени;

Д) группировку чтений секвенирования по молекулярному штрих-коду, присутствующему в каждом чтении секвенирования, для определения, присутствует ли мутация на обеих цепях двухцепочечной ДНК-матрицы.

30. Способ по п.29, где получение фрагментов двухцепочечной ДНК, каждый из которых имеет дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, включает:

i) лигирование 3'-дуплексного адаптера с каждым 3'-концом фрагмента двухцепочечной ДНК, полученного с матрицы двухцепочечной ДНК, где 3'-дуплексный адаптер включает: а) первый олигонуклеотид, включающий 5'-фосфат, первый молекулярный штрих-код и 3'-олигонуклеотид, отождествленный со б) вторым олигонуклеотидом, включающим расщепляемые 3'-блокирующие группы, где 3'-олигонуклеотид и вторая олигонуклеотидная последовательность комплементарны;

ii) расщепление расщепляемых 3'-блокирующих групп;

iii) лигирование 5'-адаптеров с каждым дефосфорилированным 5'-концом

фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, где 5'-дуплексный адаптер включает олигонуклеотид, включающий второй молекулярный штрих-код, где второй молекулярный штрих-код отличается от первого молекулярного штрих-кода, где 5'-адаптер лигирован с фрагментом двухцепочечной ДНК перед первым молекулярным штрих-кодом, и сохранение гэта с одноцепочечной нуклеиновой кислотой между 5'-концами фрагмента двухцепочечной ДНК и 5'-адаптерами; и

iv) заполнение гэта одноцепочечной нуклеиновой кислоты между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером, с получением фрагмента двухцепочечной ДНК, включающим дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК.

31. Способ по п.29, где создание библиотеки ДНК цепей Уотсона из библиотеки амплифицированных дуплексов для секвенирования включает:

i) амплификацию первой аликвоты библиотеки амплифицированных дуплексов для секвенирования с использованием пары праймеров, состоящей из первого праймера и второго праймера, где первый праймер способен к гибридизации с цепью Уотсона, и где первый праймер включает метку, с получением двухцепочечных продуктов амплификации, имеющих меченые цепи Уотсона;

ii) денатурацию двухцепочечных продуктов амплификации, имеющих меченые цепи Уотсона, с получением одноцепочечных меченых цепей Уотсона и одноцепочечных цепей Крика; и

iii) выделение одноцепочечных меченых цепей Уотсона с получением библиотеки ДНК цепей Уотсона из библиотеки амплифицированных дуплексов для секвенирования.

32. Способ по любому из пп.29-31, где двухцепочечная ДНК-матрица получена из образца млекопитающего, при этом создание библиотеки ДНК цепей Крика из библиотеки амплифицированных дуплексов для секвенирования включает:

i) амплификацию второй аликвоты библиотеки амплифицированных дуплексов для секвенирования с использованием пары праймеров, включающей первый праймер и второй праймер, где первый праймер способен к гибридизации с цепью Крика, и где первый праймер включает метку, с получением двухцепочечных продуктов амплификации, имеющих меченые цепи Крика;

ii) денатурацию двухцепочечных продуктов амплификации, имеющих меченые цепи Крика, с получением одноцепочечных меченых цепей Крика и одноцепочечных цепей Уотсона; и

iii) выделение одноцепочечных меченых цепей Крика с получением библиотеки ДНК цепей Крика из библиотеки амплифицированных дуплексов для секвенирования.

33. Способ по любому из пп.29-32, где млекопитающим является человек.

34. Способ по любому из пп.29-33, где способ дополнительно включает, перед получением фрагмента двухцепочечной ДНК, имеющего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК:

фрагментацию двухцепочечной ДНК с получением фрагментов двухцепочечной

ДНК;

дефосфорилирование 5'-концов фрагментов двухцепочечной ДНК; и затупление концов фрагментов двухцепочечной ДНК.

35. Способ по любому из пп.29-34, где лигирование 3'-дуплексных адаптеров с каждым 3'-концом фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, включает контакт 3'-дуплексного адаптера и фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, в присутствии лигазы.

36. Способ по п.35, где лигаза является ДНК-лигазой T4.

37. Способ по любому из пп.29-36, где расщепление расщепляемых 3'-блокирующих групп включает контакт 3'-дуплексного адаптера с урацил-ДНК-гликозилазой (УДГ).

38. Способ по любому из пп.29-37, где лигирование 5'-адаптера с каждым дефосфорилированным 5'-концом фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, включает контакт 5'-адаптера и фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, в присутствии лигазы.

39. Способ по п.38, где лигаза является лигазой *Escherichia coli*.

40. Способ по любому из пп.29-39, где заполнение гэпа с одноцепочечной нуклеиновой кислотой между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером включает контакт 5'-конца фрагмента двухцепочечной ДНК и 5'-адаптера в присутствии полимеразы и дНТФ.

41. Способ по п.40, где полимеразы являются Taq-полимеразой.

42. Способ по любому из пп.29-31, где лигирование 5'-адаптера с каждым 5'-концом фрагмента двухцепочечной ДНК и заполнение гэпа между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером производят параллельно.

43. Способ по любому из пп.29-42, где амплификация фрагмента двухцепочечной ДНК, включающего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, с получением амплифицированной библиотеки дуплексов для секвенирования, включает контакт фрагмента двухцепочечной ДНК, включающего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, с парой универсальных праймеров в условиях ПЦР.

44. Способ по п.43, где амплификация включает полногеномную ПЦР.

45. Способ по любому из пп.29-44, где меченые праймеры являются биотинилированными праймерами, и где биотинилированные праймеры могут давать биотинилированные одноцепочечные цепи Уотсона и биотинилированные одноцепочечные цепи Крика.

46. Способ по п.45, где стадии денатурации включают денатурацию NaOH, термическую денатурацию или их комбинацию.

47. Способ по п.45 или по п.46, где стадии выделения включают контакт меченых цепей Уотсона со стрептавидин-функционализированными частицами и контакт меченых

цепей Крика со стрептавидин-функционализированными частицами.

48. Способ по п.47, где стадии выделения дополнительно включают денатурацию немеченых цепей Уотсона и денатурацию немеченых цепей Уотсона.

49. Способ по п.47 или по п.48, где стадии выделения дополнительно включают высвобождение биотинилированных одноцепочечных цепей Уотсона со стрептавидин-функционализированных частиц и высвобождение биотинилированных одноцепочечных цепей Крика со стрептавидин-функционализированных частиц.

50. Способ по любому из пп.29-44, где меченые праймеры являются фосфорилированными праймерами, и где фосфорилированные праймеры могут давать фосфорилированные одноцепочечные цепи Уотсона и фосфорилированные одноцепочечные цепи Крика.

51. Способ по п.50, где стадии денатурации включают расщепление экзонуклеазой фага лямбда.

52. Способ по любому из пп.29-51, где амплификация области-мишени с библиотеки ДНК цепей Уотсона дополнительно включает вторую амплификацию с использованием второй пары праймеров, включающей первый праймер, способный к гибридизации с областью-мишенью, и второй праймер, способный к гибридизации с 3'-дуплексным адаптером; и где амплификация области-мишени с библиотеки ДНК цепей Крика дополнительно включает вторую амплификацию с использованием второй пары праймеров, включающей первый праймер, способный к гибридизации с областью-мишенью, и второй праймер, способный к гибридизации с 5'-адаптером.

53. Способ по любому из пп.29-52, где этапы секвенирования включают секвенирование спаренных концов.

54. Способ обнаружения присутствия или отсутствия мутации в области-мишени двухцепочечной ДНК-матрицы, полученной из образца млекопитающего, и определения, присутствует ли мутация в обеих цепях двухцепочечной ДНК-матрицы, где способ включает:

А) получение фрагментов двухцепочечной ДНК, каждый из которых имеет дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК;

В) создание библиотеки ДНК цепей Уотсона и библиотеки ДНК цепей Крика из библиотеки амплифицированных дуплексов для секвенирования из фрагмента двухцепочечной ДНК, имеющего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК;

С) амплификацию области-мишени с одноцепочечных цепей Уотсона с использованием пары праймеров, состоящей из первого праймера, способного к гибридизации с областью-мишенью, и второго праймера, способного к гибридизации с 3'-дуплексным адаптером;

Д) амплификацию области-мишени с одноцепочечных цепей Крика с использованием пары праймеров, состоящей из первого праймера, способного к гибридизации с областью-мишенью, и второго праймера, способного к гибридизации с 5'-

адаптером;

Е) секвенирование области-мишени, амплифицированной с библиотеки ДНК цепей Уотсона, с получением чтения секвенирования и обнаружением присутствия или отсутствия мутации в цепи Уотсона области-мишени;

Ф) секвенирование области-мишени, амплифицированной с библиотеки ДНК цепей Крика, с получением чтения секвенирования и обнаружением присутствия или отсутствия мутации в цепи Крика области-мишени;

Г) группировку чтений секвенирования по молекулярному штрих-коду, присутствующему в каждом чтении секвенирования, с определением, присутствует ли мутация в обеих цепях двухцепочечной ДНК-матрицы.

55. Способ по п.54, где двухцепочечная ДНК-матрица представляет собой образец ДНК, и получение фрагментов двухцепочечной ДНК, каждый из которых имеет дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, включает:

i) лигирование 3'-дуплексного адаптера с каждым 3'-концом фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, где 3'-дуплексный адаптер включает: а) первый олигонуклеотид, включающий 5'-фосфат, первый молекулярный штрих-код и 3'-олигонуклеотид, отожженный с б) вторым олигонуклеотидом, включающим расщепляемую 3'-блокирующую группу, где 3'-олигонуклеотид и последовательность второго олигонуклеотида комплементарны;

ii) расщепление расщепляемой 3'-блокирующей группы;

iii) лигирование 5'-адаптера с каждым дефосфорилированным 5'-концом фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, где 5'-дуплексный адаптер включает олигонуклеотид, включающий второй молекулярный штрих-код, где второй молекулярный штрих-код отличается от первого молекулярного штрих-кода, где 5'-адаптер лигирован на фрагменте двухцепочечной ДНК перед первым молекулярным штрих-кодом, и сохранение гэта с одноцепочечной нуклеиновой кислотой между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером; и

iv) заполнение гэта с одноцепочечной нуклеиновой кислотой между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером с получением фрагмента двухцепочечной ДНК, включающего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК.

56. Способ по п.54, где двухцепочечная ДНК-матрица является образцом внеклеточной ДНК, и создание библиотеки ДНК цепей Уотсона и библиотеки ДНК цепей Крика из библиотеки амплифицированных дуплексов для секвенирования из фрагмента двухцепочечной ДНК, имеющего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, включает:

i) амплификацию фрагмента двухцепочечной ДНК, имеющего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, с использованием пары универсальных праймеров, состоящей из первого праймера и

второго праймера, где амплификация включает контакт фрагмента двухцепочечной ДНК, включающего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, с парой праймеров в условиях полногеномной ПЦР, где первый праймер способен к гибридизации с цепью Уотсона, и где первый праймер биотинилирован, с получением двухцепочечных продуктов амплификации, имеющих биотинилированные цепи Уотсона;

ii) контакт двухцепочечных продуктов амплификации, имеющих биотинилированные цепи Уотсона, со стрептавидин-функционализированными частицами при условиях, где биотинилированные цепи Уотсона связываются со стрептавидин-функционализированными частицами;

iii) денатурацию двухцепочечных продуктов амплификации, биотинилированных цепей Уотсона, с сохранением одноцепочечных биотинилированных цепей Уотсона связанными со стрептавидин-функционализированными частицами и высвобождение одноцепочечных цепей Крика;

iv) сбор одноцепочечных цепей Крика;

v) высвобождение одноцепочечных биотинилированных цепей Уотсона со стрептавидин-функционализированных частиц; и

vi) сбор одноцепочечных биотинилированных цепей Уотсона.

57. Способ по любому из пп.54-56, где двухцепочечная ДНК-матрица получена из образца млекопитающего.

58. Способ по любому из пп.54-57, где млекопитающим является человек.

59. Способ по любому из пп.54-58, где способ дополнительно включает, перед получением фрагмента двухцепочечной ДНК, имеющего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК:

фрагментацию двухцепочечной ДНК с получением фрагментов двухцепочечной ДНК;

дефосфорилирование 5'-концов фрагментов двухцепочечной ДНК; и

затупление концов фрагментов двухцепочечной ДНК.

60. Способ по любому из пп.54-59, где лигирование 3'-дуплексного адаптера с каждым 3'-концом фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, включает контакт 3'-дуплексного адаптера и фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, в присутствии лигазы.

61. Способ по п.60, где лигаза является ДНК-лигазой T4.

62. Способ по любому из пп.54-61, где расщепление расщепляемой 3'-блокирующей группы включает контакт 3'-дуплексного адаптера с урацил-ДНК-гликозилазой (УДГ).

63. Способ по любому из пп.54-62, где лигирование 5'-адаптера с каждым дефосфорилированным 5'-концом фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, включает контакт 5'-адаптера и фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы в присутствии

лигазы.

64. Способ по п.63, где лигаза является лигазой *Escherichia coli*.

65. Способ по любому из пп.54-64, где заполнение гэпа с одноцепочечной нуклеиновой кислотой между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером включает контакт 5'-конца фрагмента двухцепочечной ДНК и 5'-адаптера в присутствии полимеразы и дНТФ.

66. Способ по п.65, где полимеразы является Taq-В полимеразой.

67. Способ по любому из пп.54-66, где лигирование 5'-адаптера с каждым 5'-концом фрагмента двухцепочечной ДНК и заполнение гэпа между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером производят параллельно.

68. Способ по любому из пп.54-67, где амплификация фрагмента двухцепочечной ДНК, имеющего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, включает контакт фрагмента двухцепочечной ДНК, включающего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, с парой праймеров в условиях ПЦР.

69. Способ по п.68, где амплификация включает полногеномную ПЦР.

70. Способ по любому из пп.54-69, где амплификация области-мишени с библиотеки ДНК цепей Уотсона, дополнительно включает вторую амплификацию с использованием второй пары праймеров, включающей первый праймер, способный к гибридизации с областью-мишенью, и второй праймер, способный к гибридизации с 3'-дуплексным адаптером; и где амплификация области-мишени с библиотеки ДНК цепей Крика, дополнительно включает вторую амплификацию с использованием второй пары праймеров, включающей первый праймер, способный к гибридизации с областью-мишенью, и второй праймер, способный к гибридизации с 5'-адаптером.

71. Способ по любому из пп.54-70, где стадии секвенирования включают секвенирование спаренных концов или секвенирование неспаренных концов.

72. Способ обнаружения присутствия или отсутствия мутации в области-мишени двухцепочечной ДНК-матрицы, полученной из образца млекопитающего, и определения, присутствует ли мутация в обеих цепях двухцепочечной ДНК-матрицы, где способ включает:

А) получение фрагментов двухцепочечной ДНК, каждый из которых имеет дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК;

В) амплификацию фрагментов двухцепочечной ДНК, каждый из которых имеет дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, с использованием пары универсальных праймеров, где амплификация включает контакт фрагмента двухцепочечной ДНК, включающего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, с парой праймеров в условиях полногеномной ПЦР;

С) амплификацию области-мишени с цепи Уотсона амплифицированных фрагментов двухцепочечной ДНК, каждый из которых имеет дуплексный молекулярный

штрих-код на каждом конце фрагмента двухцепочечной ДНК, с использованием пары праймеров, состоящей из первого праймера, способного к гибридизации с областью-мишенью, и второго праймера, способного к гибридизации с 3'-дуплексным адаптером;

Д) амплификацию области-мишени с цепи Крика амплифицированных фрагментов двухцепочечной ДНК, каждый из которых имеет дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, с использованием пары праймеров, состоящей из первого праймера, способного к гибридизации с областью-мишенью, и второго праймера, способного к гибридизации с 5'-адаптером;

Е) секвенирование области-мишени, амплифицированной с цепи Уотсона, с получением чтения секвенирования и обнаружением присутствия или отсутствия мутации в цепи Уотсона области-мишени;

Ф) секвенирование области-мишени, амплифицированной с цепи Крика, с получением чтения секвенирования и обнаружением присутствия или отсутствия мутации в цепи Крика области-мишени;

Г) группировку чтений секвенирования по молекулярному штрих-коду, присутствующему в каждом чтении секвенирования, с определением, присутствует ли мутация в обоих цепях двухцепочечной ДНК-матрицы.

73. Способ по п.72, где двухцепочечная ДНК-матрица является образцом геномной ДНК, и получение фрагментов двухцепочечной ДНК, каждый из которых имеет дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, включает:

i) лигирование 3'-дуплексного адаптера с каждым 3'-концом фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, где 3'-дуплексный адаптер включает: а) первый олигонуклеотид, включающий 5'-фосфат, первый молекулярный штрих-код и 3'-олигонуклеотид, отожженный с б) вторым олигонуклеотидом, включающим расщепляемую 3'-блокирующую группу, где 3'-олигонуклеотид и последовательность второго олигонуклеотида комплементарны;

ii) расщепление расщепляемой 3'-блокирующей группы;

iii) лигирование 5'-адаптера с каждым дефосфорилированным 5'-концом фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, где 5'-дуплексный адаптер включает олигонуклеотид, включающий второй молекулярный штрих-код, где второй молекулярный штрих-код отличается от первого молекулярного штрих-кода, где 5'-адаптер лигирован на фрагменте двухцепочечной ДНК перед первым молекулярным штрих-кодом, и сохранение гэпа с одноцепочечной нуклеиновой кислотой между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером; и

iv) заполнение гэпа с одноцепочечной нуклеиновой кислотой между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером с получением фрагмента двухцепочечной ДНК, включающего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК.

74. Способ по п.73, где двухцепочечная ДНК-матрица является образцом

внеклеточной ДНК.

75. Способ по любому из пп.72-74, где двухцепочечная ДНК-матрица является образцом геномной ДНК.

76. Способ по любому из пп.72-75, где млекопитающим является человек.

77. Способ по любому из пп.72-76, где способ дополнительно включает, перед получением фрагмента двухцепочечной ДНК, имеющего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК:

фрагментацию двухцепочечной ДНК с получением фрагментов двухцепочечной ДНК;

дефосфорилирование 5'-концов фрагментов двухцепочечной ДНК; и

затупление концов фрагментов двухцепочечной ДНК.

78. Способ по любому из пп.72-77, где лигирование 3'-дуплексного адаптера с каждым 3'-концом фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, включает контакт 3'-дуплексного адаптера и фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, в присутствии лигазы.

79. Способ по п.50, где лигаза является ДНК-лигазой T4.

80. Способ по любому из пп.72-79, где расщепление расщепляемой 3'-блокирующей группы включает контакт 3'-дуплексного адаптера с урацил-ДНК-гликозилазой (УДГ).

81. Способ по любому из пп.72-80, где лигирование 5'-адаптера с каждым дефосфорилированным 5'-концом фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, включает контакт 5'-адаптера и фрагмента двухцепочечной ДНК, полученного с двухцепочечной ДНК-матрицы, в присутствии лигазы.

82. Способ по п.81, где лигаза является лигазой *Escherichia coli*.

83. Способ по любому из пп.72-82, где заполнение гэпа с одноцепочечной нуклеиновой кислотой между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером включает контакт 5'-конца фрагмента двухцепочечной ДНК и 5'-адаптера в присутствии ДНК-полимеразы и дНТФ.

84. Способ по п.83, где ДНК-полимераза является Taq-В полимеразой.

85. Способ по любому из пп.72-84, где лигирование 5'-адаптера с каждым 5'-концом фрагмента двухцепочечной ДНК и заполнение гэпа между 5'-концом фрагмента двухцепочечной ДНК и 5'-адаптером производят параллельно.

86. Способ по любому из пп.72-85, где амплификация фрагмента двухцепочечной ДНК, имеющего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, включает контакт фрагмента двухцепочечной ДНК, включающего дуплексный молекулярный штрих-код на каждом конце фрагмента двухцепочечной ДНК, с парой праймеров в условиях ПЦР.

87. Способ по п.86, где амплификация включает полногеномную ПЦР.

88. Способ по любому из пп.72-87, где амплификация области-мишени с

библиотеки ДНК цепей Уотсона дополнительно включает вторую амплификацию с использованием второй пары праймеров, включающей первый праймер, способный к гибридизации с областью-мишенью, и второй праймер, способный к гибридизации с 3'-дуплексным адаптером; и где амплификация области-мишени с библиотеки ДНК цепей Крика дополнительно включает вторую амплификацию с использованием второй пары праймеров, включающей первый праймер, способный к гибридизации с областью-мишенью, и второй праймер, способный к гибридизации с 5'-адаптером.

89. Способ по любому из пп.72-88, где стадии секвенирования включают секвенирование спаренных концов.

90. Способ, включающий:

а. присоединение частично двухцепочечных 3'-адаптеров с 3'-концами и обеими цепями Уотсона и Крика популяции фрагментов двухцепочечной ДНК в анализируемом образце ДНК, где первая цепь частично двухцепочечного 3'-адаптера включает, в направлении 5'→3': (i) первый сегмент, (ii) экзогенную UID последовательность, (iii) сайт отжига для 5'-адаптера и (iv) универсальную 3'-адаптерную последовательность, включающую сайт праймера для секвенирования R2, и где вторая цепь частично двухцепочечного 3'-адаптера включает, в направлении 5'→3': (i) сегмент, комплементарный первому сегменту, и (ii) 3'-блокирующую группу, где, необязательно, вторая цепь является расщепляемой;

б. отжиг 5'-адаптеров с 3'-адаптерами посредством сайта отжига, где 5'-адаптеры включают, в направлении 5'→3': (i) универсальную 5'-адаптерную последовательность, которая не комплементарна универсальной 3'-адаптерной последовательностью, и которая включает сайт праймера для секвенирования R1, и (ii) последовательность, комплементарную сайту отжига для 5'-адаптера;

с. проведение реакции, подобной ник-трансляции, с удлинением 5'-адаптеров через экзогенную UID последовательность 3'-адаптеров и ковалентным связыванием удлинённых 5'-адаптеров с 5'-концами цепей Уотсона и Крика фрагментов двухцепочечной ДНК;

д. проведение первичной амплификации для амплификации лигированных с адаптером фрагментов двухцепочечной ДНК с получением ампликонов;

е. определение чтений последовательности одного или больше ампликонов одного или больше лигированных с адаптером фрагментов двухцепочечной ДНК;

ф. отнесение чтений последовательности к семействам UID, где каждый член семейства UID включает одну и ту же экзогенную UID последовательность;

г. отнесение чтений последовательности каждого семейства UID к подсемейству Уотсона и подсемейству Крика на основе пространственных отношений экзогенной последовательности UID с последовательностью чтений R1 и R2;

h. идентификацию нуклеотидной последовательности как точно представляющей цепь Уотсона анализируемого фрагмента ДНК, если пороговый процент членов подсемейства Уотсона содержат такую последовательность;

i. идентификацию нуклеотидной последовательности как точно представляющей цепь Крика анализируемого фрагмента ДНК, если пороговый процент членов подсемейства Крика содержат последовательность;

j. идентификацию мутации в нуклеотидной последовательности, точно представляющей цепь Уотсона, если последовательность, точно представляющая цепь Уотсона, отличается от референсной последовательности, в которой такая мутация отсутствует;

k. идентификацию мутации в нуклеотидной последовательности, точно представляющей цепь Крика, если последовательность, точно представляющая цепь Крика, отличается от референсной последовательности, в которой такая мутация отсутствует; и

l. идентификацию мутации в анализируемом фрагменте ДНК, если мутация в нуклеотидной последовательности, точно представляющей цепь Уотсона, и мутация в нуклеотидной последовательности, точно представляющей цепь Крика, является одной и той же мутацией.

91. Способ по п.90, где каждый член семейства UID дополнительно включает одну и ту же эндогенную UID последовательность, где эндогенная UID последовательность включает конец фрагмента двухцепочечной ДНК из популяции.

92. Способ по п.91, где эндогенная UID последовательность, включающая конец фрагмента двухцепочечной ДНК, включает по меньшей мере 8, 10 или 15 оснований.

93. Способ по любому из пп.90-92, где экзогенная UID последовательность является уникальной для каждого фрагмента двухцепочечной ДНК.

94. Способ по любому из пп.90-92, где экзогенная UID последовательность не является уникальной для каждого фрагмента двухцепочечной ДНК.

95. Способ по любому из пп.91-94, где каждый член семейства UID включает одну и ту же эндогенную UID последовательность и одну и ту же экзогенную UID последовательность.

96. Способ по любому из предыдущих пп., где стадия (d) включает не больше 11 циклов ПЦР-амплификации.

97. Способ по п.96, где стадия (d) включает не больше 7 циклов ПЦР-амплификации.

98. Способ по п.97, где стадия (d) включает не больше 5 циклов ПЦР-амплификации.

99. Способ по любому из предыдущих пп., где стадия (d) включает по меньшей мере 1 цикл ПЦР-амплификации.

100. Способ по любому из предыдущих пп., где, перед определением чтений последовательности, ампликоны обогащают одним или больше полинуклеотидами-мишенями.

101. Способ по п.100, где обогащение включает:

a. селективную амплификацию ампликонов цепей Уотсона, включающих

последовательность полинуклеотида-мишени, с использованием первого набора пар мишень-селективных праймеров Уотсона, причем первый набор пар мишень-селективных праймеров Уотсона включает: (i) первый мишень-селективный праймер Уотсона, включающий последовательность, комплементарную части универсальной 3'-адаптерной последовательности, где, необязательно, часть универсальной 3'-адаптерной последовательности является сайтом праймера для секвенирования R2 универсальной 3'-адаптерной последовательности, и (ii) второй мишень-селективный праймер Уотсона, включающий мишень-селективную последовательность, с получением в результате целевых продуктов амплификации Уотсона; и

в. селективную амплификацию ампликонов цепей Крика, включающих такую же последовательность полинуклеотида-мишени, с использованием первого набора пар мишень-селективных праймеров Крика, причем первый набор пар мишень-селективных праймеров Крика включает: (i) первый мишень-селективный праймер Крика, включающий последовательность, комплементарную части универсальной 5'-адаптерной последовательности, где, необязательно, часть универсальной 5'-адаптерной последовательности является сайтом праймера для секвенирования R1 универсальной 5'-адаптерной последовательности, и (ii) второй мишень-селективный праймер Крика, включающий такую же мишень-селективную последовательность, как и последовательность второго мишень-селективного праймера Уотсона, с получением в результате целевых продуктов амплификации Крика.

102. Способ по п.101, включающий очистку целевых продуктов амплификации Уотсона и целевых продуктов амплификации Крика от нецелевых полинуклеотидов.

103. Способ по п.102, где очистка включает прикрепление целевых продуктов амплификации Уотсона и целевых продуктов амплификации Крика к твердой подложке.

104. Способ по п.103, где первый мишень-селективный праймер Уотсона и первый мишень-селективный праймер Крика включает первый член пары аффинного связывания, и где твердая подложка, включает второй член пары аффинного связывания.

105. Способ по п.104, где первый член является биотином, а второй член является стрептавидином.

106. Способ по любому из пп.102-105, где твердая подложка включает частицу, лунку, мембрану, пробирку, колонку, планшет, сефарозу, магнитную частицу или чип.

107. Способ по любому из пп.102-106, включающий удаление полинуклеотидов, которые не прикрепились к твердой подложке.

108. Способ по любому из пп.101-107, включающий:

а. дополнительную амплификацию целевых продуктов амплификации Уотсона с использованием второго набора мишень-селективных праймеров Уотсона, причем второй набор мишень-селективных праймеров Уотсона включает: (i) третий мишень-селективный праймер Уотсона, включающий последовательность, комплементарную части универсальной 3'-адаптерной последовательности, где, необязательно, часть универсальной 3'-адаптерной последовательности является сайтом праймера для

секвенирования R2 универсальной 3'-адаптерной последовательности, и (ii) четвертый мишень-селективный праймер Уотсона, включающий, в направлении 5'→3', сайт праймера для секвенирования R1 и мишень-селективную последовательность, селективную в отношении того же полинуклеотида-мишени, с получением в результате целевых членов библиотеки Уотсона;

в. дополнительную амплификацию целевых продуктов амплификации Крика с использованием второго набора мишень-селективных праймеров Крика, причем второй набор мишень-селективных праймеров Крика включает: (i) третий мишень-селективный праймер Крика, включающий последовательность, комплементарную части универсальной 5'-адаптерной последовательности, где, необязательно, часть универсальной 5'-адаптерной последовательности является сайтом праймера для секвенирования R1 универсальной 5'-адаптерной последовательности, и (ii) четвертый мишень-селективный праймер Крика, включающий, в направлении 5'→3', сайт праймера для секвенирования R2 и мишень-селективную последовательность, селективную в отношении того же полинуклеотида-мишени четвертого мишень-селективного праймера Уотсона, с получением в результате целевых членов библиотеки Крика.

109. Способ по п.108, где третий мишень-селективные праймеры Уотсона и Крика дополнительно включают последовательность штрих-кода образца.

110. Способ по п.108 или 109, где третий мишень-селективный праймер Уотсона дополнительно включает первую последовательность присоединения, которая обеспечивает гибридизацию с первым праймером присоединения на секвенаторе, и где третий мишень-селективный праймер Крика дополнительно включает вторую последовательность присоединения, которая обеспечивает гибридизацию со вторым праймером присоединения на секвенаторе.

111. Способ по любому из пп.108-110, где четвертый мишень-селективный праймер Уотсона дополнительно включает вторую последовательность присоединения, и где четвертый мишень-селективный праймер Крика дополнительно включает первую последовательность присоединения.

112. Способ по п.110 или 111, где первая последовательность присоединения является последовательностью P7, и где вторая последовательность присоединения является последовательностью P5.

113. Способ по любому из пп.101-112, где целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляет по меньшей мере 50% целевых полинуклеотидов в популяции фрагментов двухцепочечной ДНК.

114. Способ по п.113, где целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляет по меньшей мере 70% целевых полинуклеотидов в популяции фрагментов двухцепочечной ДНК.

115. Способ по п.114, где целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляет по меньшей мере 80% целевых полинуклеотидов в популяции фрагментов двухцепочечной ДНК.

116. Способ по п.115, где целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляет по меньшей мере 90% целевых полинуклеотидов в популяции фрагментов двухцепочечной ДНК.

117. Способ по любому из пп.101-112, где целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляет по меньшей мере 50% популяции всех фрагментов ДНК.

118. Способ по п.117, где целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляет по меньшей мере 70% популяции всех фрагментов ДНК.

119. Способ по п.118, где целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляет по меньшей мере 80% популяции всех фрагментов ДНК.

120. Способ по п.119, где целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляет по меньшей мере 90% популяции всех фрагментов ДНК.

121. Способ, включающий:

a. присоединение адаптеров к популяции фрагментов двухцепочечной ДНК в анализируемом образце ДНК, где адаптеры включают двухцепочечную часть, включающую экзогенный UID, и разветвленную часть, включающую: (i) одноцепочечную 3'-адаптерную последовательность, включающую сайт праймера для секвенирования R2, и (ii) одноцепочечную 5'-адаптерную последовательность, включающую сайт праймера для секвенирования R1;

b. проведение первичной амплификации для амплификации лигированных с адаптером фрагментов двухцепочечной ДНК с получением ампликонов;

c. селективную амплификацию ампликонов цепей Уотсона, включающих последовательность полинуклеотида-мишени, с использованием первого набора пар мишень-селективных праймеров Уотсона, причем первый набор пар мишень-селективных праймеров Уотсона включает: (i) первый мишень-селективный праймер Уотсона, включающий последовательность, комплементарную части универсальной 3'-адаптерной последовательности, где, необязательно, часть универсальной 3'-адаптерной последовательности является сайтом праймера для секвенирования R2 универсальной 3'-адаптерной последовательности, и (ii) второй мишень-селективный праймер Уотсона, включающий мишень-селективную последовательность, с получением в результате целевых продуктов амплификации Уотсона;

d. селективную амплификацию ампликонов цепей Крика, включающих такую же последовательность полинуклеотида-мишени, с использованием первого набора пар мишень-селективных праймеров Крика, причем первый набор пар мишень-селективных праймеров Крика включает: первый мишень-селективный праймер Крика, включающий последовательность, комплементарную части универсальной 5'-адаптерной последовательности, где, необязательно, часть универсальной 5'-адаптерной последовательности является сайт праймера для секвенирования R1 универсальной 5'-адаптерной последовательности, и (ii) второй мишень-селективный праймер Крика, включающий такую же мишень-селективную последовательность, как и

последовательность второго мишень-селективного праймера Крика, с получением в результате целевых продуктов амплификации Крика.

e. определение чтений последовательности целевых продуктов амплификации Уотсона и целевых продуктов амплификации Крика;

f. отнесение чтений последовательности к семействам UID, где каждый член семейства UID включает одну и ту же экзогенную UID последовательность;

g. отнесение чтений последовательности каждого семейства UID к подсемейству Уотсона и подсемейству Крика на основе пространственных отношений экзогенной последовательности UID с последовательностью чтений R1 и R2;

h. идентификацию нуклеотидной последовательности как точно представляющей цепь Уотсона анализируемого фрагмента ДНК, если пороговый процент членов семейства Уотсона содержит такую последовательность;

i. идентификацию нуклеотидной последовательности как точно представляющей цепь Крика анализируемого фрагмента ДНК, если пороговый процент членов семейства Крика содержит такую последовательность; и

j. идентификацию мутации в анализируемом фрагменте ДНК, если нуклеотидная последовательность, точно представляющая цепь Уотсона, и нуклеотидная последовательность, точно представляющая цепь Крика, содержат одну и ту же мутацию.

122. Способ по п.121, включающий очистку целевых продуктов амплификации Уотсона и целевых продуктов амплификации Крика от нецелевых полинуклеотидов.

123. Способ по п.122, где очистка включает прикрепление целевых продуктов амплификации Уотсона и целевых продуктов амплификации Крика к твердой подложке.

124. Способ по п.123, где первый мишень-селективный праймер Уотсона и первый мишень-селективный праймер Крика включает первый член пары аффинного связывания, и где твердая подложка включает второй член пары аффинного связывания.

125. Способ по п.124, где первый член является биотином, а второй член является стрептавидином.

126. Способ по любому из пп.122-125, где твердая подложка включает частицу, лунку, мембрану, пробирку, колонку, планшет, сефарозу, магнитную частицу или чип.

127. Способ по любому из пп.122-126, включающий удаление полинуклеотидов, которые не прикрепились к твердой подложке.

128. Способ по любому из пп.121-127, включающий:

a. дополнительную амплификацию целевых продуктов амплификации Уотсона с использованием второго набора мишень-селективных праймеров Уотсона, причем второй набор мишень-селективных праймеров Уотсона включает: (i) третий мишень-селективный праймер Уотсона, включающий последовательность, комплементарную сайту праймера для секвенирования R2 универсальной 3'-адаптерной последовательности, и (ii) четвертый мишень-селективный праймер Уотсона, включающий, в направлении 5'→3', сайт праймера для секвенирования R1 и мишень-селективную последовательность, селективную в отношении такого же полинуклеотида-мишени, с получением в результате

целевых членов библиотеки Уотсона;

в. дополнительную амплификацию целевых продуктов амплификации Крика с использованием второго набора мишень-селективных праймеров Крика, причем второй набор мишень-селективных праймеров Крика включает: (i) третий мишень-селективный праймер Крика, включающий последовательность, комплементарную сайту праймера для секвенирования R1 универсальной 3'-адаптерной последовательности, и (ii) четвертый мишень-селективный праймер Крика, включающий, в направлении 5'→3', сайт праймера для секвенирования R2 и мишень-селективную последовательность, селективную в отношении такого же полинуклеотида-мишени четвертого мишень-селективного праймера Уотсона, с получением в результате целевых членов библиотеки Крика, с получением в результате целевых членов библиотеки Крика.

129. Способ по п.128, где третий мишень-селективные праймеры Уотсона и Крика дополнительно включают последовательность штрих-кода образца.

130. Способ по п.128 или 129, где третий мишень-селективный праймер Уотсона дополнительно включает первую последовательность присоединения, которая обеспечивает гибридизацию с первым праймером присоединения на секвенаторе, и где третий мишень-селективный праймер Крика дополнительно включает вторую последовательность присоединения, которая обеспечивает гибридизацию со вторым праймером присоединения на секвенаторе.

131. Способ по любому из пп.128-130, где четвертый мишень-селективный праймер Уотсона дополнительно включает вторую последовательность присоединения, и где четвертый мишень-селективный праймер Крика дополнительно включает первую последовательность присоединения.

132. Способ по п.130 или 131, где первая последовательность присоединения является последовательностью P7, и где вторая последовательность присоединения является последовательностью P5.

133. Способ по любому из пп.121-131, где присоединение включает присоединение адаптеров с А-хвостом к популяции фрагментов двухцепочечной ДНК.

134. Способ по п.133, где присоединение включает присоединение адаптеров с А-хвостом к обоим концам фрагментов ДНК в популяции.

135. Способ по любому из пп.121-131, где присоединение включает:

а. присоединение частично двухцепочечных 3'-адаптеров к 3'-концам обеих цепей Уотсона и Крика популяции фрагментов двухцепочечной ДНК, где первая цепь частично двухцепочечного 3'-адаптера включает, в направлении 5'→3': (i) первый сегмент, (ii) необязательно, экзогенную UID последовательность, (iii) сайт отжига для 5'-адаптера, и (iv) универсальную 3'-адаптерную последовательность, включающую сайт праймера для секвенирования R2, и где вторая цепь частично двухцепочечного 3'-адаптера включает, в направлении 5'→3': (i) сегмент, комплементарный первому сегменту, и (ii) 3'-блокирующую группу, где, необязательно, вторая цепь является расщепляемой; и

б. отжиг 5'-адаптеров с 3'-адаптерами посредством сайта отжига, где 5'-адаптер

включает, в направлении 5'→3': (i) универсальную 5'-адаптерную последовательность, которая не комплементарна универсальной 3'-адаптерной последовательности, и которая включает сайт праймера для секвенирования R1, и (ii) последовательность, комплементарную сайту отжига для 5'-адаптера; и

с. проведение реакции, подобной ник-трансляции, с удлинением 5'-адаптера через 3'-адаптер и ковалентным связыванием удлинённого 5'-адаптера с 5'-концом цепей Уотсона и Крика фрагментов двухцепочечной ДНК.

136. Способ по любому из пп.121-135, где UID последовательность включает эндогенную UID последовательность, включающую конец фрагмента двухцепочечной ДНК из популяции.

137. Способ по п.136, где эндогенная UID последовательность, включающая конец фрагмента двухцепочечной ДНК, включает по меньшей мере 8, 10 или 15 оснований.

138. Способ по любому из пп.121-136, где экзогенная UID последовательность является уникальной для каждого фрагмента двухцепочечной ДНК.

139. Способ по любому из пп.121-136, где экзогенная UID последовательность не является уникальной для каждого фрагмента двухцепочечной ДНК.

140. Способ по любому из пп.136-139, где каждый член семейства UID включает одну и ту же эндогенную UID последовательность и одну и ту же экзогенную UID последовательность.

141. Способ по любому из пп.121-140, где амплификация лигированного адаптером, фрагменты двухцепочечной ДНК, чтобы произвести ампликоны включает не больше, чем 11 циклов амплификации ПЦР.

142. Способ по п.141, где амплификация лигированных с адаптером фрагментов двухцепочечной ДНК с получением ампликонов включает не больше 7 циклов ПЦР-амплификации.

143. Способ по п.142, где амплификация лигированных с адаптером фрагментов двухцепочечной ДНК с получением ампликонов включает не больше 5 циклов ПЦР-амплификации.

144. Способ по любому из предыдущих пп., где амплификация лигированных с адаптером фрагментов двухцепочечной ДНК с получением ампликонов включает по меньшей мере 1 цикл ПЦР-амплификации.

145. Способ по любому из пп.121-143, где целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляет по меньшей мере 50% целевых полинуклеотидов в популяции фрагментов двухцепочечной ДНК.

146. Способ по п.145, где целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляет по меньшей мере 70% целевых полинуклеотидов в популяции фрагментов двухцепочечной ДНК.

147. Способ по п.146, где целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляет по меньшей мере 80% целевых полинуклеотидов в популяции фрагментов двухцепочечной ДНК.

148. Способ по п.147, где целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляет по меньшей мере 90% целевых полинуклеотидов в популяции фрагментов двухцепочечной ДНК.

149. Способ по любому из пп.121-143, где целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляет по меньшей мере 50% популяции всех фрагментов ДНК.

150. Способ по п.149, где целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляет по меньшей мере 70% популяции всех фрагментов ДНК.

151. Способ по п.150, где целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляет по меньшей мере 80% популяции всех фрагментов ДНК.

152. Способ по п.151, где целевые члены библиотеки Уотсона и целевые члены библиотеки Крика представляет по меньшей мере 90% популяции всех фрагментов ДНК.

153. Способ по любому из предыдущих пп., где определение чтений последовательности позволяет определять последовательность обоих концов молекул-матриц.

154. Способ по п.153, где определение обоих концов молекул-матриц включает секвенирование спаренных концов.

155. Способ по любому из предыдущих пп., где определение чтений последовательности включает секвенирование неспаренных концов по длине матриц с получением чтений последовательности.

156. Способ по любому из предыдущих пп., где определение чтений последовательности включает секвенирование с помощью массивно-параллельного секвенатора.

157. Способ по п.156, где массивно-параллельный секвенатор выполнен с возможностью определения чтений последовательности с обоих концов полинуклеотидов-матриц.

158. Способ по любому из предыдущих пп., где популяция фрагментов двухцепочечной ДНК включает один или больше фрагментов, которые имеют длину приблизительно 50-600 нт.

159. Способ по любому из предыдущих пп., где популяция фрагментов двухцепочечной ДНК включает один или больше фрагментов, которые имеют длину меньше 2000, меньше 1000, меньше 500, меньше 400, меньше 300 или меньше 250 нт.

160. Способ по любому из пп.101-159, дополнительно включающий, после первичной амплификации и перед селективной амплификацией, получение библиотек одноцепочечных (оц) ДНК, соответствующих смысловой и антисмысловой цепям ампликонов.

161. Способ по п.160, где получение библиотеки оцДНК включает:

а. проведение реакции амплификации с использованием двух праймеров, где только один из этих двух праймеров включает первый член пары аффинного связывания, с получением в результате продуктов амплификации, включающих цепь, включающую

первый член пары аффинного связывания, и цепь, не включающую первый член пары аффинного связывания;

b. контакт продуктов амплификации с твердой подложкой, где твердая подложка включает второй член пары аффинного связывания;

c. денатурацию продуктов амплификации с отделением цепи, включающей первый член пары аффинного связывания, от цепи, не включающей первый член пары аффинного связывания; и

d. очистку отделенных цепей, включающих первый член пары аффинного связывания, и отделенных цепей, не включающих первый член пары аффинного связывания.

162. Способ по п.161, где первым членом пары аффинного связывания является биотин, а вторым членом пары аффинного связывания является стрептавидин.

163. Способ по п.160, где получение библиотеки оцДНК включает:

a. разделение ампликонов на две реакции амплификации, причем в каждой реакции амплификации используется прямой праймер и обратный праймер, где только один из этих двух праймеров фосфорилирован, с получением в результате продуктов амплификации, включающих фосфорилированную цепь и нефосфорилированную цепь;

b. контакт продуктов амплификации с экзонуклеазой, которая селективно расщепляет цепи с 5'-фосфатами.

164. Способ по п.163, где:

a. в первой реакции амплификации прямой праймер фосфорилирован, а обратный праймер не фосфорилирован;

b. во второй реакции амплификации обратный праймер фосфорилирован, а прямой праймер не фосфорилирован.

165. Способ по п.163, где экзонуклеаза является экзонуклеазой фага лямбда.

166. Способ по любому из пп.163-165, где фосфорилирование производят по 5'-сайту.

167. Способ по любому из пп.90-153, где первичная амплификация включает:

a. амплификацию с использованием пары праймеров, где только один из этих двух праймеров в паре праймеров включает первый член пары аффинного связывания, с получением в результате продуктов амплификации, включающих цепь, включающую первый член пары аффинного связывания, и цепь, не включающую первый член пары аффинного связывания;

b. контакт продуктов амплификации с твердой подложкой, где твердая подложка включает второй член пары аффинного связывания;

c. денатурацию продуктов амплификации с отделением цепи, включающей первый член пары аффинного связывания, от цепи, не включающей первый член пары аффинного связывания; и

d. очистку отделенных цепей, включающих первый член пары аффинного связывания, и отделенных цепей, не включающих первый член пары аффинного

связывания.

168. Способ по п.167, где первым членом пары аффинного связывания является биотин, а вторым членом пары аффинного связывания является стрептавидин.

169. Способ по любому из предыдущих пп., где чтения последовательности в семействе UID относят к подсемейству Уотсона, если экзогенная UID последовательность расположена после последовательности R2 и перед последовательностью R1.

170. Способ по любому из предыдущих пп., где чтения последовательности в семействе UID относят к подсемейству Крика, если экзогенная UID последовательность расположена после последовательности R1 и перед последовательностью R2.

171. Способ по любому из предыдущих пп., где чтения последовательности в семействе UID относят к подсемейству Уотсона, если экзогенная UID последовательность расположена в большей близости от последовательности R2 и меньшей близости от последовательности R1.

172. Способ по любому из предыдущих пп., где чтения последовательности в семействе UID относят к подсемейству Крика, если экзогенная UID последовательность расположена в большей близости от последовательности R1 и меньшей близости от последовательности R2.

173. Способ по любому из предыдущих пп., где чтения последовательности в семействе UID относят к подсемейству Уотсона, если экзогенная UID последовательность расположена непосредственно после или в пределах 1-300, 1-70, 1-60, 1-50, 1-40, 1-30, 1-20, 1-10 или 1-5 нуклеотидов от последовательности R2.

174. Способ по любому из предыдущих пп., где чтения последовательности в семействе UID относят к подсемейству Крика, если экзогенная UID последовательность расположена непосредственно после или в пределах 1-300, 1-70, 1-60, 1-50, 1-40, 1-30, 1-20, 1-10 или 1-5 нуклеотидов от последовательности R1.

175. Способ по любому из предыдущих пп., где популяция фрагментов двухцепочечной ДНК происходит из биологического образца.

176. Способ по п.175, где биологический образец получен от субъекта.

177. Способ по п.176, где субъектом является субъект-человек.

178. Способ по любому из пп.175-177, где биологический образец является жидким образцом.

179. Способ по п.178, где жидкий образец выбран из цельной крови, плазмы крови, сыворотки, мокроты, мочи, пота, слезной жидкости, асцитной жидкости, спермы и бронхоальвеолярного лаважа.

180. Способ по п.178, где жидкий образец является бесклеточным или по существу бесклеточным образцом.

181. Способ по любому из пп.175-177, где биологический образец является твердым биологическим образцом.

182. Способ по п.181, где твердый биологический образец является образцом опухоли.

183. Способ по любому из предыдущих пп., где идентифицированная мутация присутствует в популяции фрагментов двухцепочечной ДНК с частотой 0,1% или меньше.

184. Способ по п.183, где идентифицированная мутация присутствует в популяции фрагментов двухцепочечной ДНК с частотой от 0,1% до 0,00001%.

185. Способ по п.183, где идентифицированная мутация присутствует в популяции фрагментов двухцепочечной ДНК с частотой от 0,1% до 0,01%.

186. Способ по любому из предыдущих пп., где определение чтений последовательности включает определение чтений последовательности с обеих цепей Уотсона и Крика по меньшей мере 50% фрагментов двухцепочечной ДНК, включающих целевой полинуклеотид, в анализируемом образце ДНК.

187. Способ по п.186, где определение чтений последовательности включает определение чтений последовательности с обеих цепей Уотсона и Крика по меньшей мере 70% фрагментов двухцепочечной ДНК, включающих целевой полинуклеотид, в анализируемом образце ДНК.

188. Способ по п.187, где определение чтений последовательности включает определение чтений последовательности с обеих цепей Уотсона и Крика по меньшей мере 80% фрагментов двухцепочечной ДНК, включающих целевой полинуклеотид, в анализируемом образце ДНК.

189. Способ по п.188, где определение чтений последовательности включает определение чтений последовательности с обеих цепей Уотсона и Крика по меньшей мере 90% фрагментов двухцепочечной ДНК, включающих целевой полинуклеотид, в анализируемом образце ДНК.

190. Способ по любому из предыдущих пп., где определение чтений последовательности включает определение чтений последовательности с обеих цепей Уотсона и Крика по меньшей мере 50% фрагментов двухцепочечной ДНК в анализируемом образце ДНК.

191. Способ по любому из предыдущих пп., где определение чтений последовательности включает определение чтений последовательности с обеих цепей Уотсона и Крика по меньшей мере 70% фрагментов двухцепочечной ДНК в анализируемом образце ДНК.

192. Способ по любому из предыдущих пп., где определение чтений последовательности включает определение чтений последовательности с обеих цепей Уотсона и Крика по меньшей мере 80% фрагментов двухцепочечной ДНК в анализируемом образце ДНК.

193. Способ по любому из предыдущих пп., где определение чтений последовательности включает определение чтений последовательности с обеих цепей Уотсона и Крика по меньшей мере 90% фрагментов двухцепочечной ДНК в анализируемом образце ДНК.

194. Способ по любому из предыдущих пп., где частота ошибок, связанных с идентификацией одной или больше мутаций в анализируемых фрагментах ДНК согласно

способу по любому из предыдущих пп., снижена по меньшей мере в 2 раза, 4 раза, 5 раз, 10 раз, 20 раз, 30 раз, 40 раз, 50 раз, 60 раз, 70 раз, 80 раз, 90 раз или 100 раз по сравнению с альтернативным способом идентификации мутаций, который не требует, чтобы мутация была обнаружена в обеих цепях Уотсона и Крика анализируемого фрагмента ДНК.

195. Способ по п.194, где альтернативный способ включает стандартное молекулярное баркодирование или стандартное молекулярное баркодирование на основе ПЦР.

196. Способ по п.195, где альтернативный способ включает:

a. присоединение адаптеров к популяции фрагментов двухцепочечной ДНК в анализируемом образце ДНК, где адаптеры включают уникальный экзогенный UID;

b. проведение первичной амплификации для амплификации лигированных с адаптером фрагментов двухцепочечной ДНК с получением ампликонов;

c. определение чтений последовательности одного или больше ампликонов одного или больше лигированных с адаптером фрагментов двухцепочечной ДНК;

d. отнесение чтений последовательности к семействам UID, где каждый член семейства UID включает одну и ту же экзогенную UID последовательность;

e. идентификацию нуклеотидной последовательности как точно представляющей анализируемый фрагмент ДНК, если пороговый процент членов семейства UID содержат такую последовательность; и

f. идентификацию мутации в анализируемом фрагменте ДНК, если последовательность, идентифицированная как точно представляющая анализируемый фрагмент ДНК, отличается от референсной последовательности, в которой такая мутация отсутствует.

197. Способ по любому из предыдущих пп., где частота ошибок, связанных с идентификацией одной или больше мутаций в анализируемых фрагментах ДНК согласно способу по любому из предыдущих пп., составляет не больше 1×10^{-2} , не больше 1×10^{-3} , не больше 1×10^{-4} , не больше 1×10^{-5} , не больше 1×10^{-6} , не больше 5×10^{-6} или не больше 1×10^{-7} .

198. Машиночитаемый носитель, содержащий машиноисполняемые инструкции для анализа данных чтений последовательностей из образца нуклеиновой кислоты, где данные генерируются способом по любому из предыдущих пп.

199. Машиночитаемый носитель по п.198, содержащий исполняемые инструкции для:

a. отнесения чтений последовательности к семействам UID, где каждый член семейства UID включает одну и ту же экзогенную UID последовательность;

b. отнесение чтений последовательности каждого семейства UID к подсемействам Уотсона и Крика на основе пространственных отношений экзогенной последовательности UID к последовательности чтений R1 и R2;

c. идентификацию нуклеотидной последовательности как точно представляющей цепь Уотсона анализируемого фрагмента ДНК, если пороговый процент членов подсемейства Уотсона содержат такую последовательность;

d. идентификацию нуклеотидной последовательности как точно представляющей цепь Крика анализируемого фрагмента ДНК, если пороговый процент членов подсемейства Крика содержат такую последовательность;

e. идентификацию мутации в нуклеотидной последовательности, точно представляющей цепь Уотсона, если последовательность, точно представляющая цепь Уотсона, отличается от референсной последовательности, в которой такая мутация отсутствует;

f. идентификацию мутации в нуклеотидной последовательности, точно представляющей цепь Крика, если последовательность, точно представляющая цепь Крика, отличается от референсной последовательности, в которой такая мутация отсутствует;

g. идентификацию мутации в анализируемом фрагменте ДНК, когда мутация в нуклеотидной последовательности, точно представляющей цепь Уотсона, и мутация в нуклеотидной последовательности, точно представляющей цепь Крика, являются одной и той же мутацией.

200. Машиночитаемый носитель по п.199, содержащий отнесение членов семейства UID к подсемейству Уотсона, если экзогенная UID последовательность расположена непосредственно после или в пределах 1-300 нуклеотидов от сайта связывания праймера для секвенирования R2.

201. Машиночитаемый носитель по любому из предыдущих пп., содержащий отнесение членов семейства UID к подсемейству Крика, если экзогенная UID последовательность расположена непосредственно после или в пределах 1-300 нуклеотидов от сайта связывания праймера для секвенирования R1.

202. Машиночитаемый носитель по любому из предыдущих пп., содержащий картирование чтений последовательности в референсном геноме.

203. Машиночитаемый носитель по п.202, где референсный геном является человеческим референсным геномом.

204. Машиночитаемый носитель по любому из предыдущих пп., дополнительно включающий машиноисполняемые инструкции с формированием отчета о вариантах лечения на основе присутствия, отсутствия или количества мутации в образце.

205. Машиночитаемый носитель по любому из предыдущих пп., дополнительно содержащий машиноисполняемый код, который обеспечивает передачу данных по сети.

206. Компьютерная система, включающая:

a. блок памяти, выполненный с возможностью получения данных по последовательности из образца нуклеиновой кислоты, где данные сгенерированы способом по любому из предыдущих пп.;

b. процессор, соединенный с возможностью связи с блоком памяти, где процессор включает машиночитаемый носитель по любому из предыдущих пп.

207. Компьютерная система по п.206, дополнительно включающая систему секвенирования, выполненную с возможностью передачи данных в блок памяти.

208. Компьютерная система по любому из предыдущих пп., дополнительно включающая пользовательский интерфейс, выполненный с возможностью передачи или отображения отчета пользователю.

209. Компьютерная система по любому из предыдущих пп., дополнительно включающая цифровой процессор, выполненный с возможностью передачи результатов анализа данных по сети.

210. Система, включающая:

- a. популяцию фрагментов двухцепочечной ДНК из биологического образца;
- b. популяцию 3'-адаптеров, как описано в любом из предыдущих пп.;
- c. популяцию 5'-адаптеров, как описано в любом из предыдущих пп.;
- d. реагенты для выполнения реакции, подобной ник-трансляции;
- e. реагенты для обогащения ампликонов одним или больше полинуклеотидами-мишенями; и
- f. систему секвенирования.

211. Система по п.210, дополнительно включающая компьютерную систему по любому из предыдущих пп.

212. Набор, включающий:

a. первый набор пар мишень-селективных праймеров Уотсона, включающий: (i) один или больше первых мишень-селективных праймеров Уотсон, включающих последовательность, комплементарную части универсальной 3'-адаптерной последовательности, где, необязательно, часть универсальной 3'-адаптерной последовательности является сайтом праймера для секвенирования R2 универсальной 3'-адаптерной последовательности, и (ii) один или больше вторых мишень-селективных праймеров Уотсон, где каждый из одного или больше вторых мишень-селективных праймеров Уотсон включает мишень-селективную последовательность;

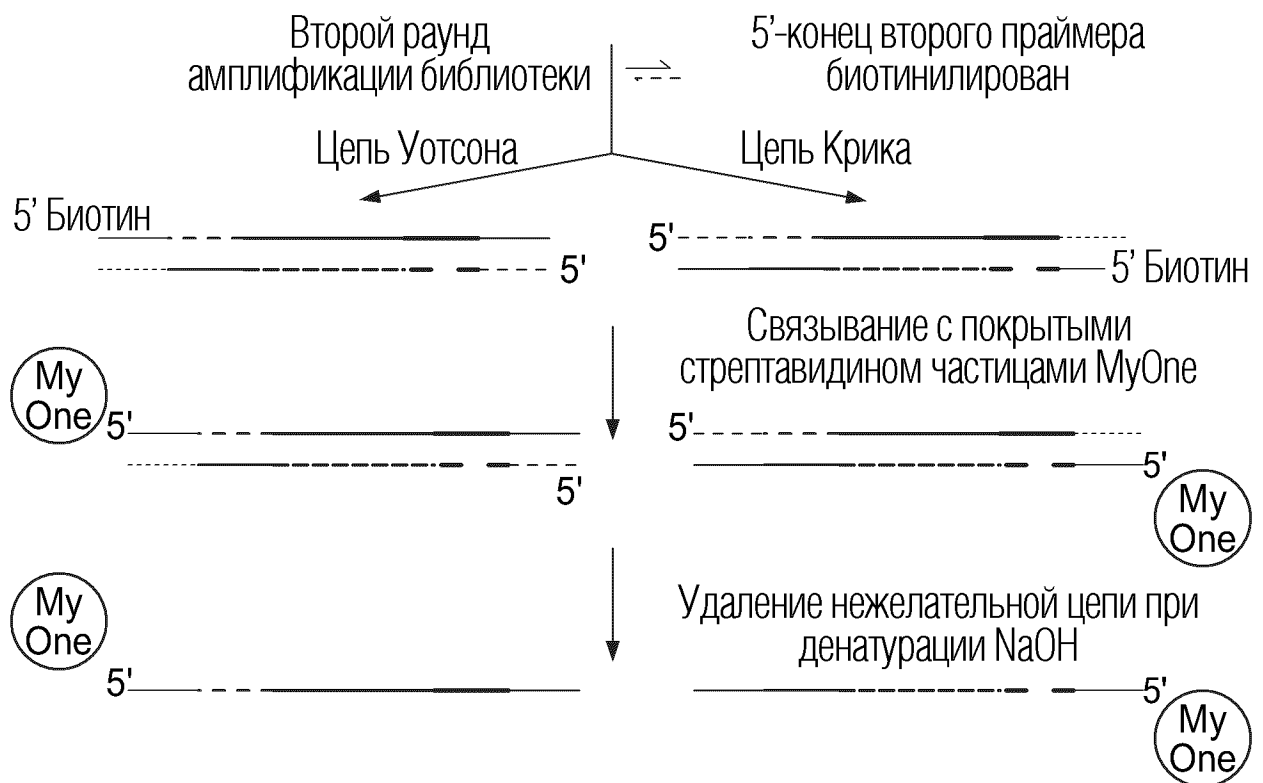
b. первый набор пар мишень-селективных праймеров Крика, включающий: (i) один или больше мишень-селективных праймеров Крика, включающих последовательность, комплементарную части универсальной 5'-адаптерной последовательности, где, необязательно, часть универсальной 5'-адаптерной последовательности является сайтом праймера для секвенирования R1 универсальной 5'-адаптерной последовательности, и (ii) один или больше вторых мишень-селективных праймеров Крика, где каждый из одного или больше вторых мишень-селективных праймеров Крика включает одну и ту же мишень-селективную последовательность, как и последовательность второго мишень-селективного праймера Уотсона;

c. второй набор пар мишень-селективных праймеров Уотсона, включающий: (i) один или больше третьих мишень-селективных праймеров Уотсона, включающих последовательность, комплементарную сайту праймера для секвенирования R2 универсальной 3'-адаптерной последовательности, и (ii) один или больше четвертых мишень-селективных праймеров Уотсона, где каждый из одного или больше четвертых мишень-селективных праймеров Уотсона включает, в направлении 5'→3', сайт праймера

для секвенирования R1 и мишень-селективную последовательность, селективную в отношении такого же целевого полинуклеотида; и

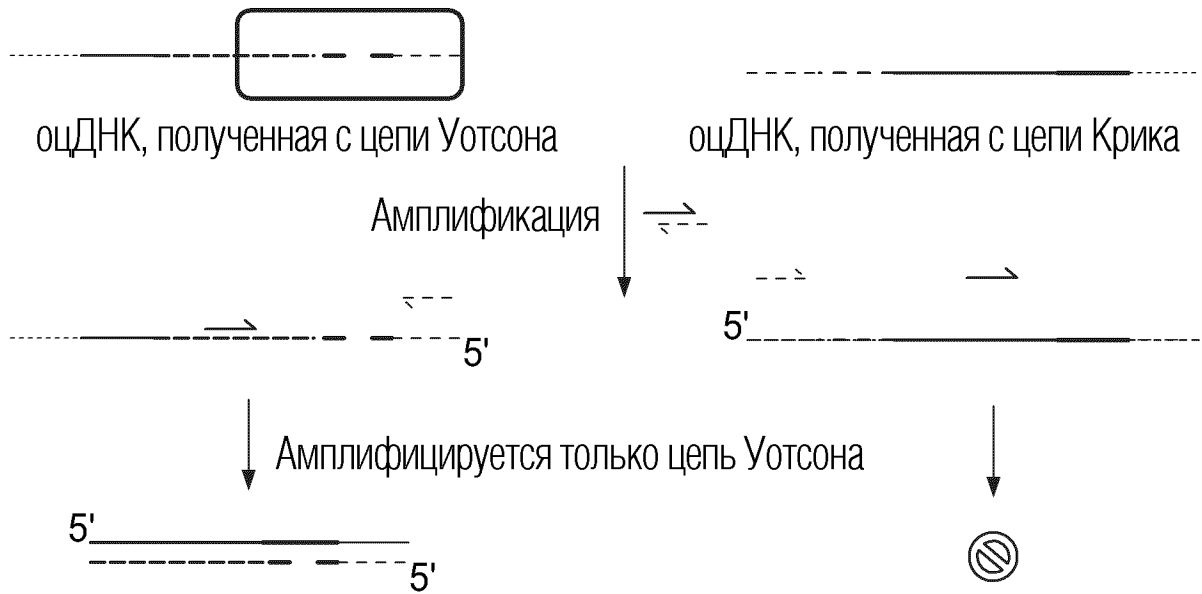
d. второй набор мишень-селективных праймеров Крика, включающий: (i) один или больше третьих мишень-селективных праймеров Крика, включающих последовательность, комплементарную сайту праймера для секвенирования R1 универсальной 3'-адаптерной последовательности, и (ii) один или больше четвертых мишень-селективных праймеров Крика, где каждый из одного или больше четвертых мишень-селективных праймеров Крика включает, в направлении 5'→3', сайт праймера для секвенирования R2 и мишень-селективную последовательность, селективную в отношении такого же полинуклеотида-мишени.

По доверенности

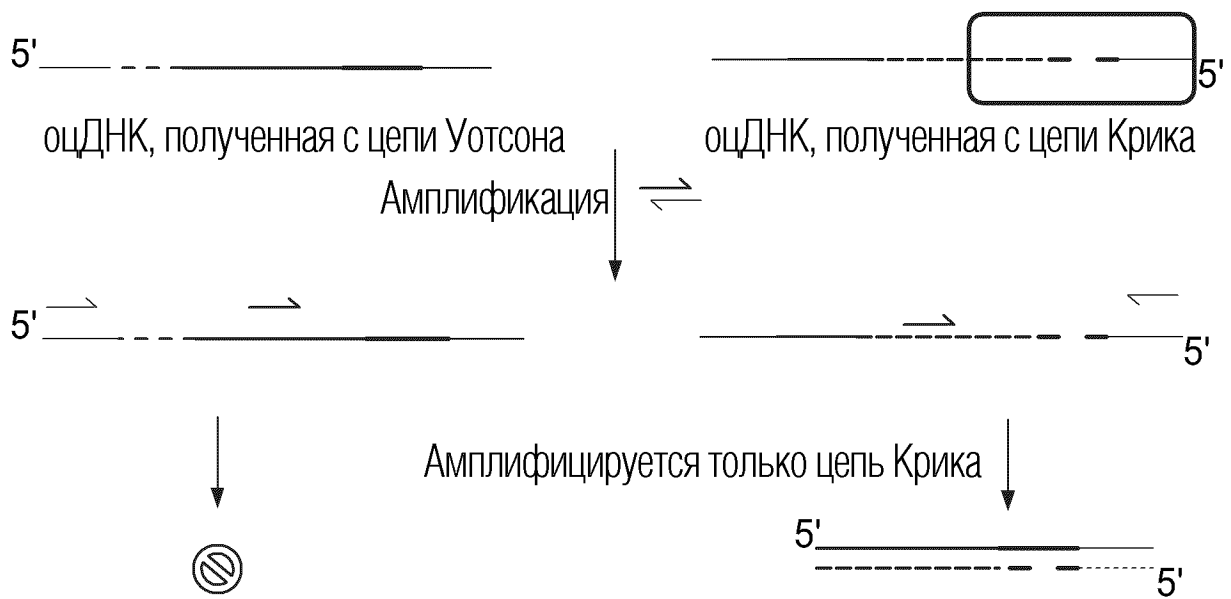


Готовая библиотека оцДНК для амплификации цепи Крика

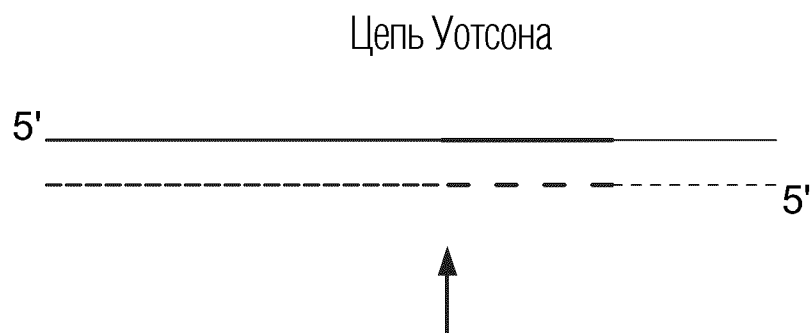
ФИГ. 3



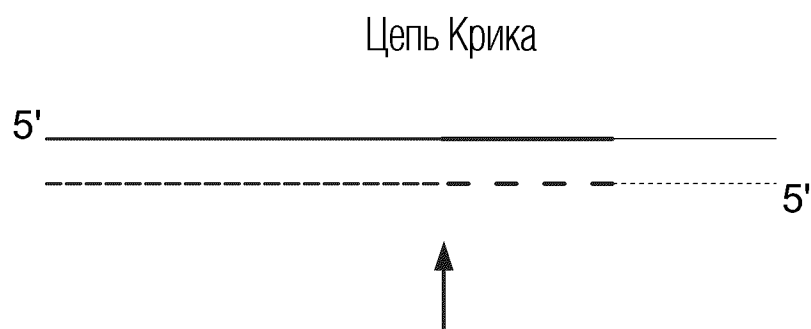
ФИГ. 4



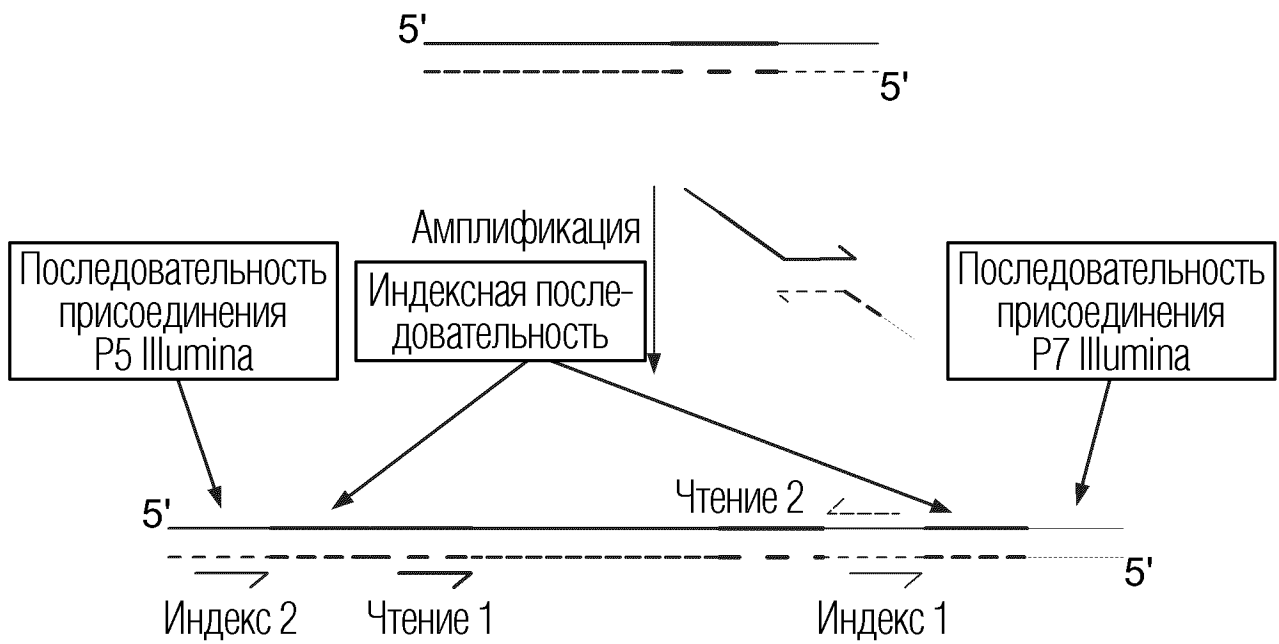
ФИГ. 5



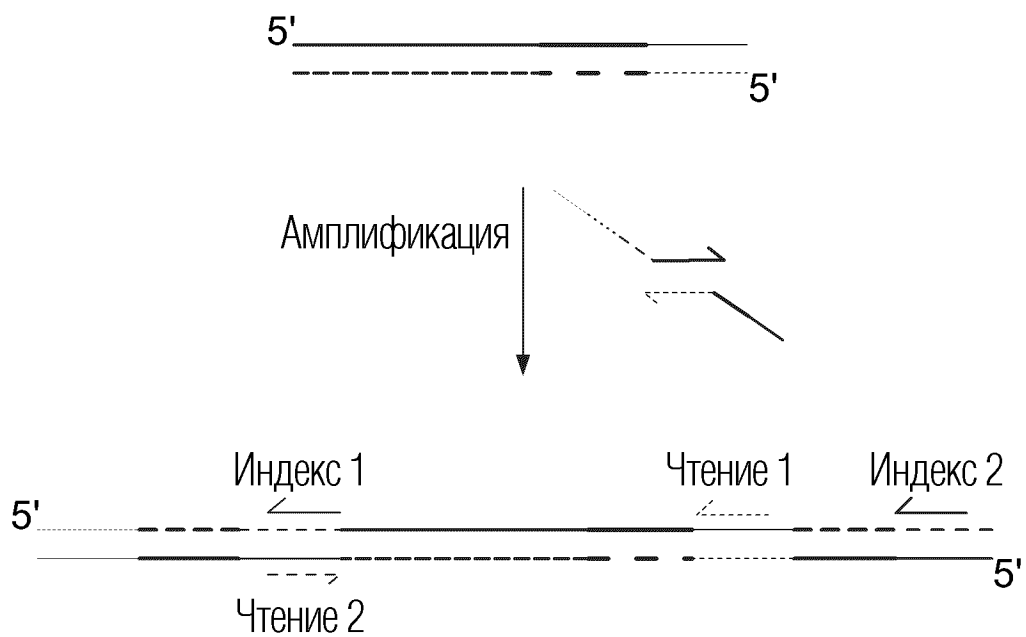
Один и тот же эндогенный UID
(конец молекулы) и экзогенный UID



ФИГ. 6

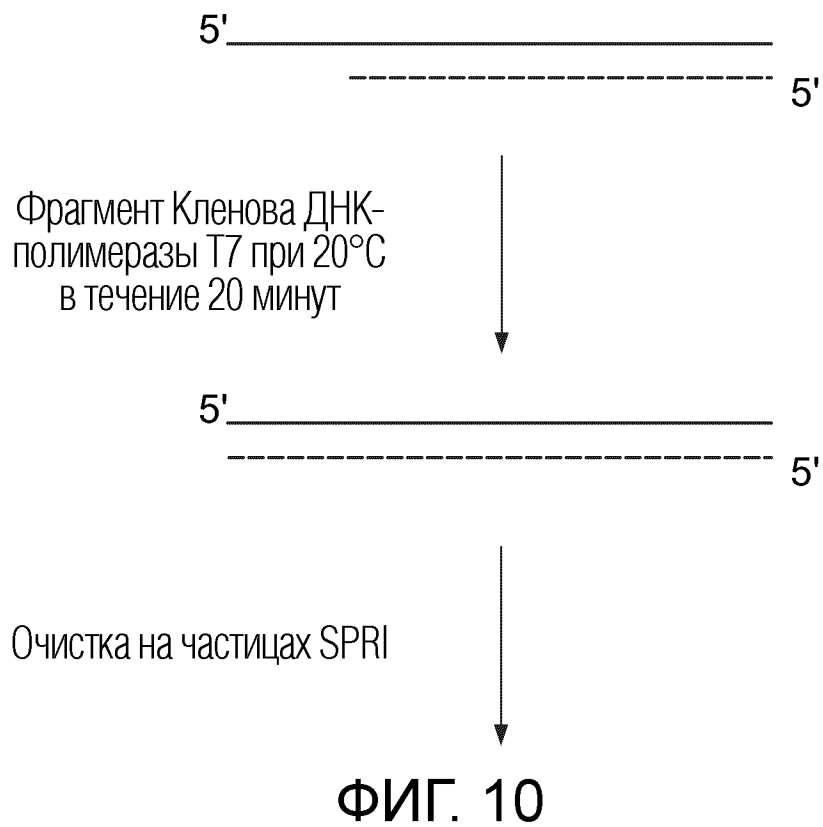
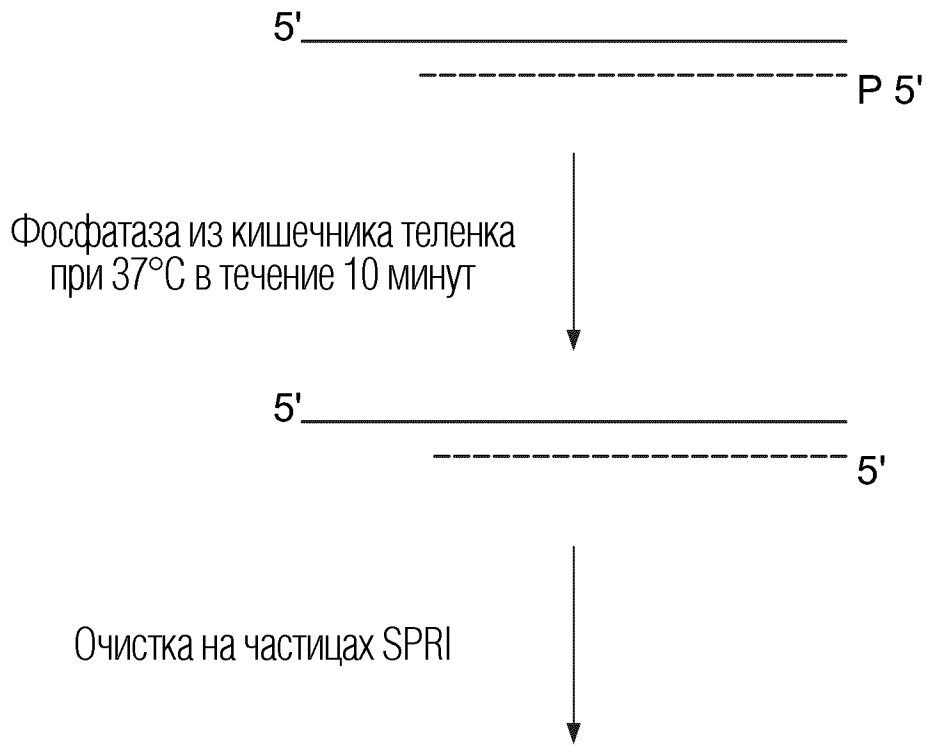


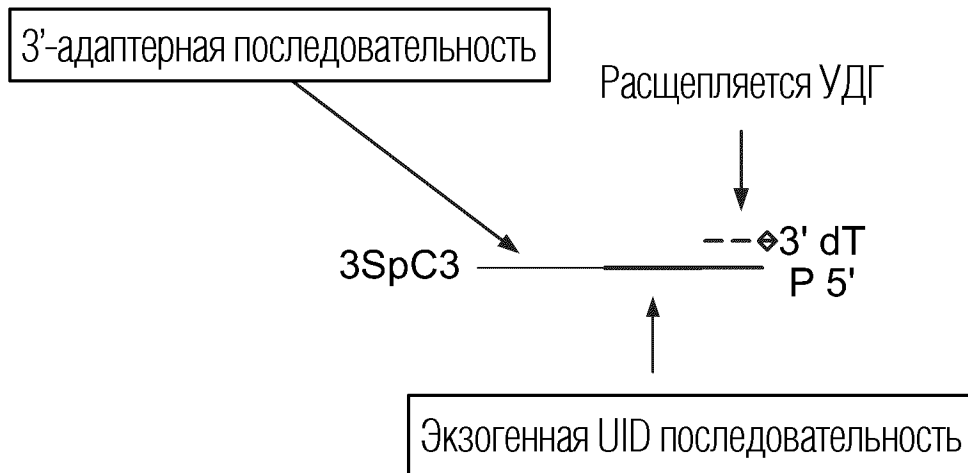
ФИГ. 7



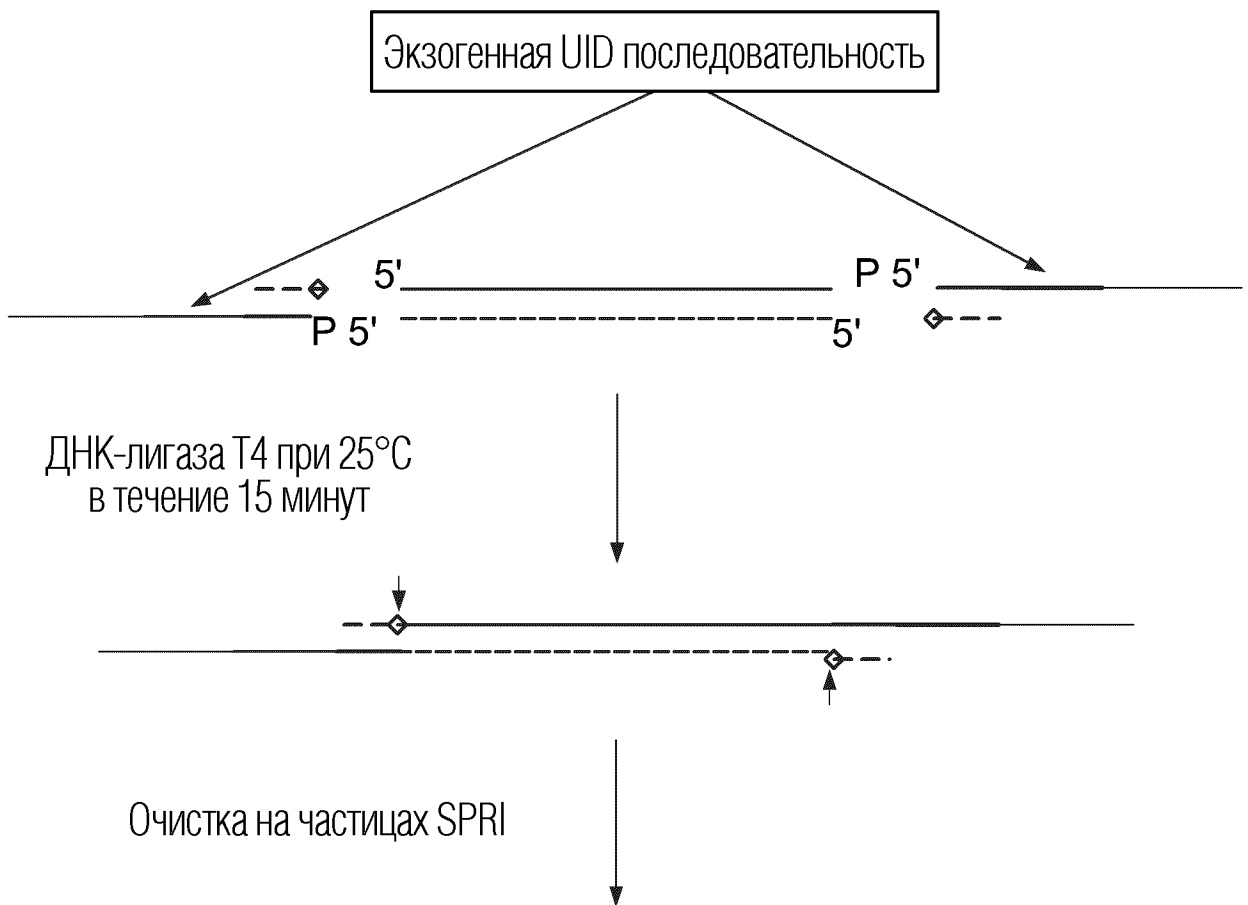
ФИГ. 8

6/27

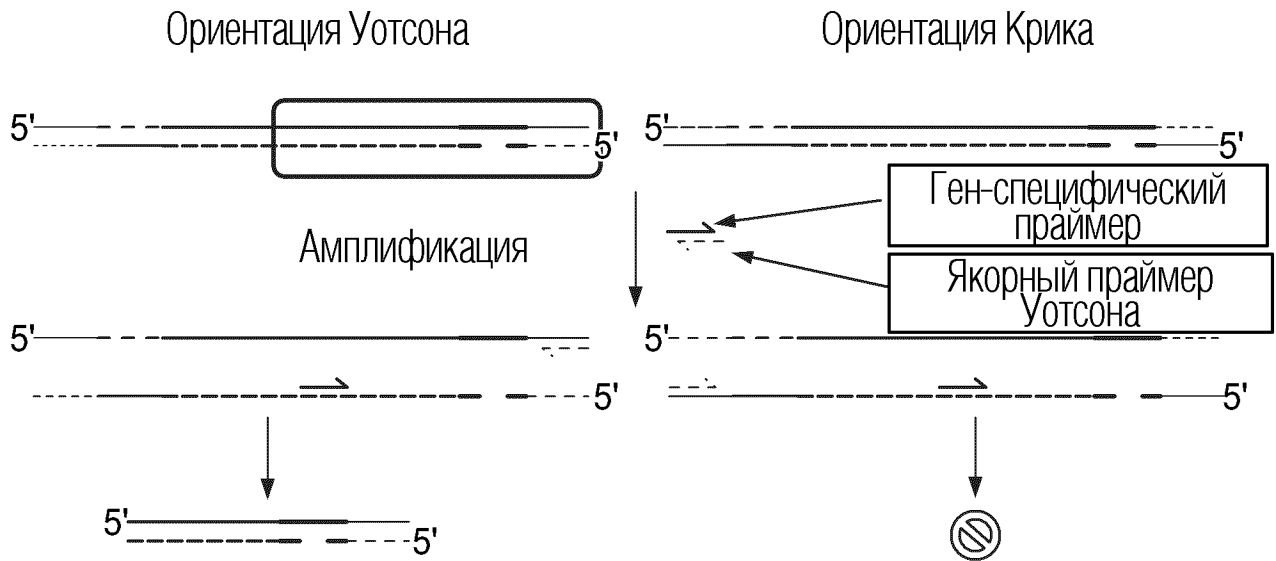




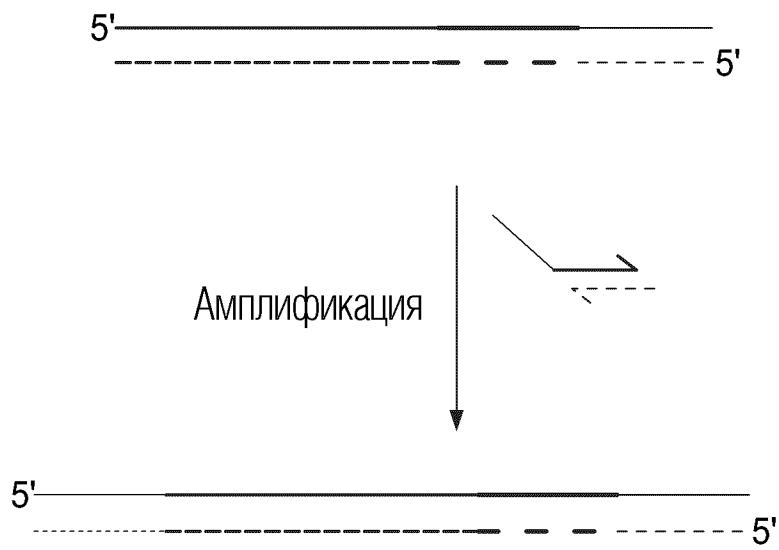
ФИГ. 11



ФИГ. 12

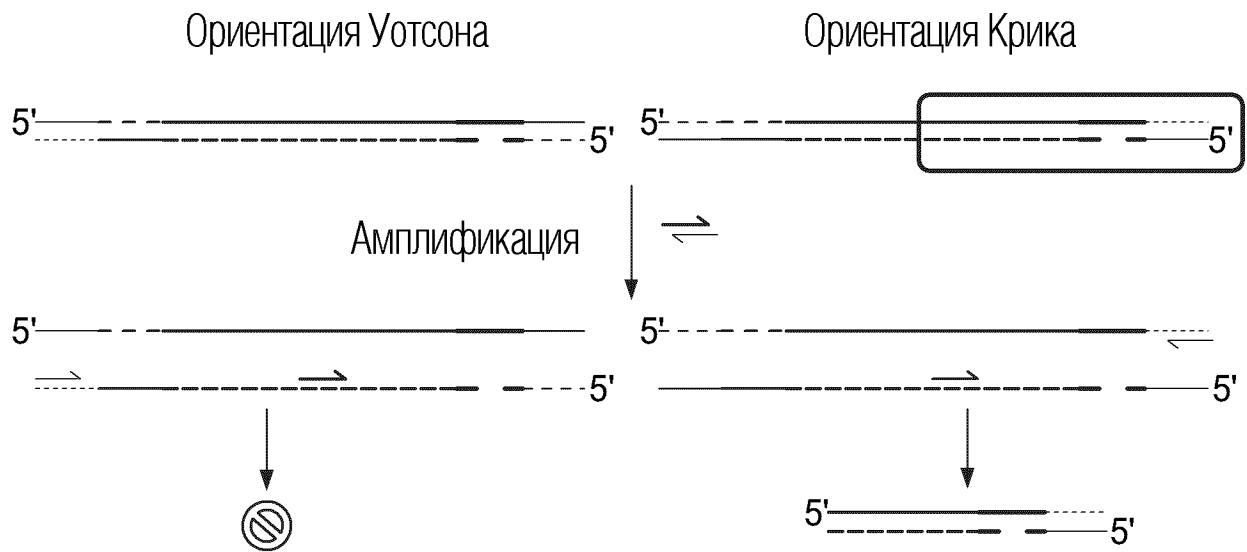


ФИГ. 15

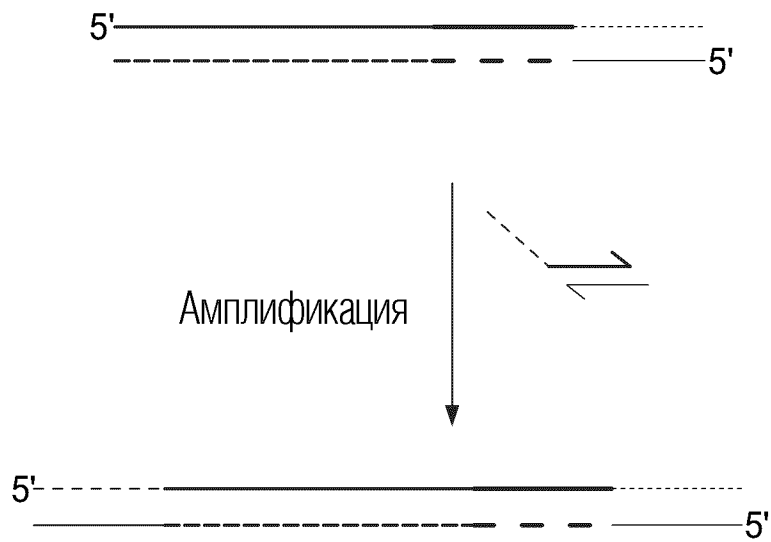


ФИГ. 16

10/27

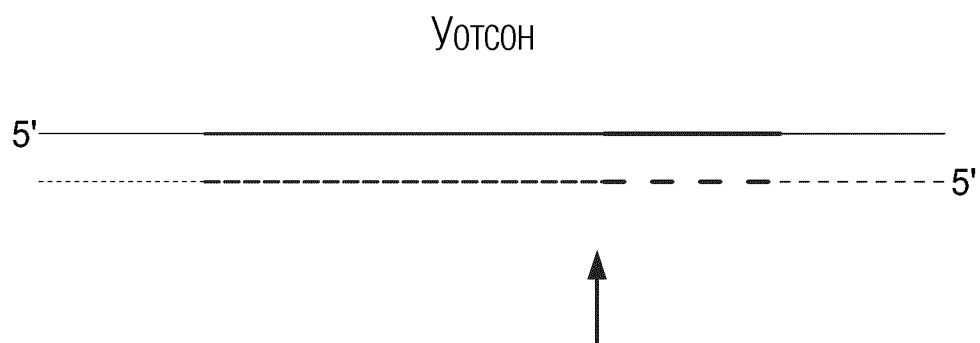


ФИГ. 17

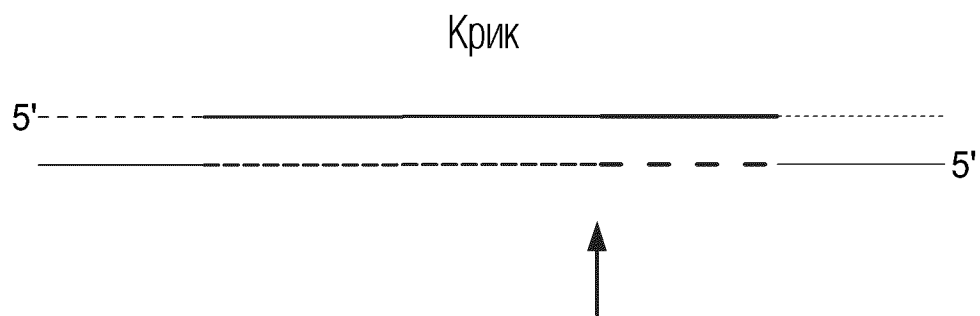


ФИГ. 18

11/27



Один и тот же эндогенный UID (конец молекулы) и экзогенный UID



ФИГ. 19

12/27

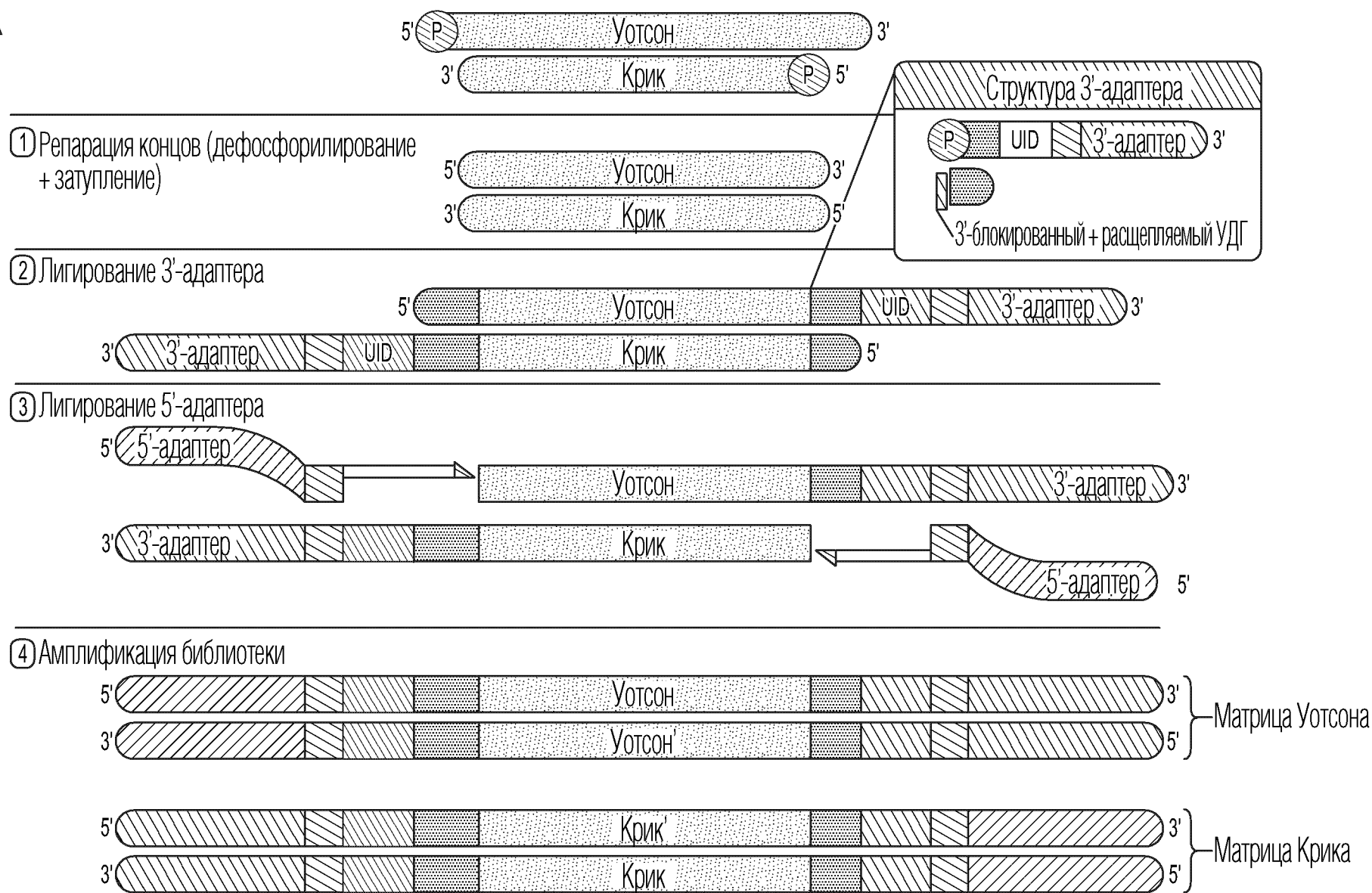


ФИГ. 20



ФИГ. 21

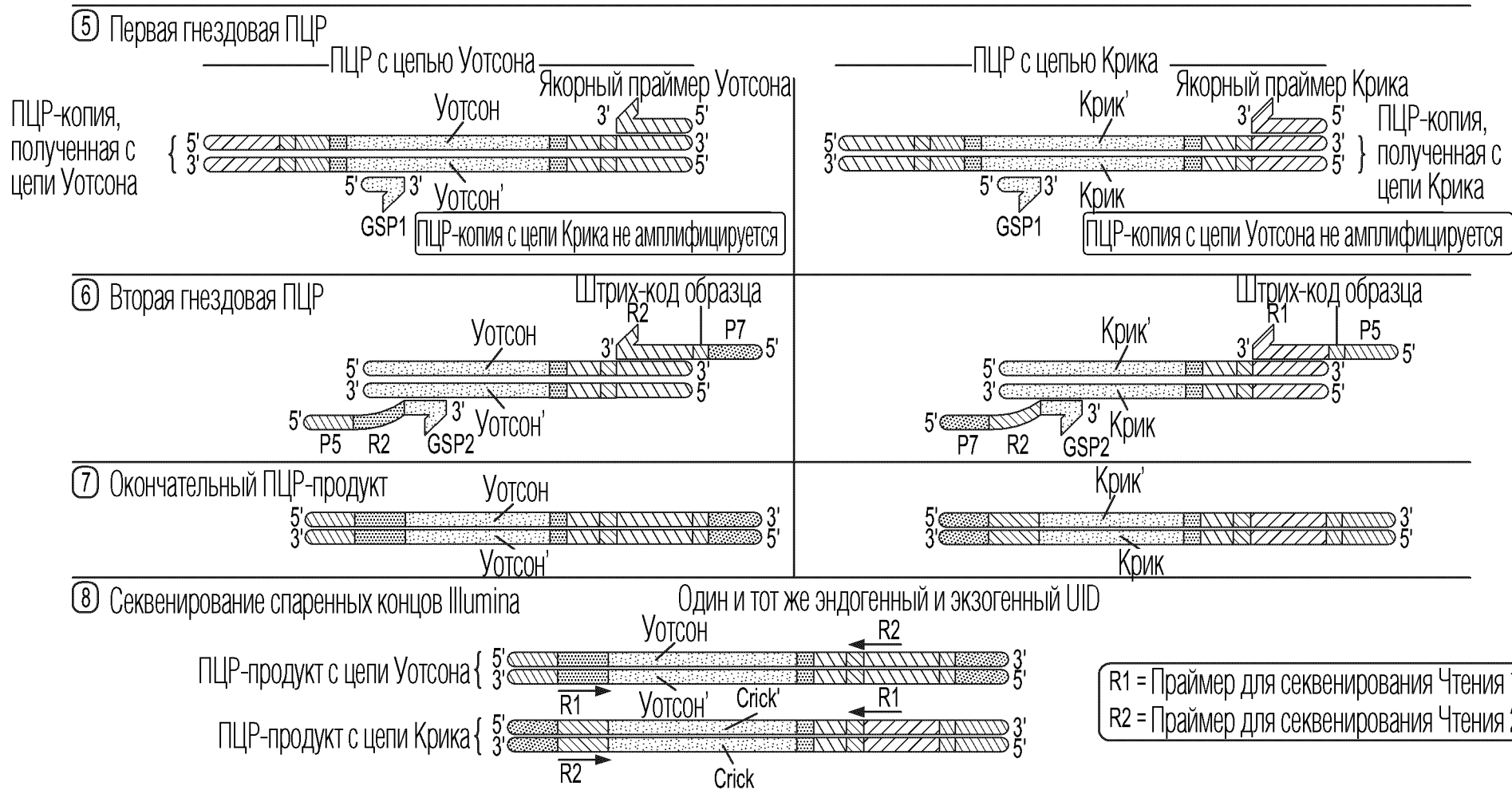
A



13/27

ФИГ. 22

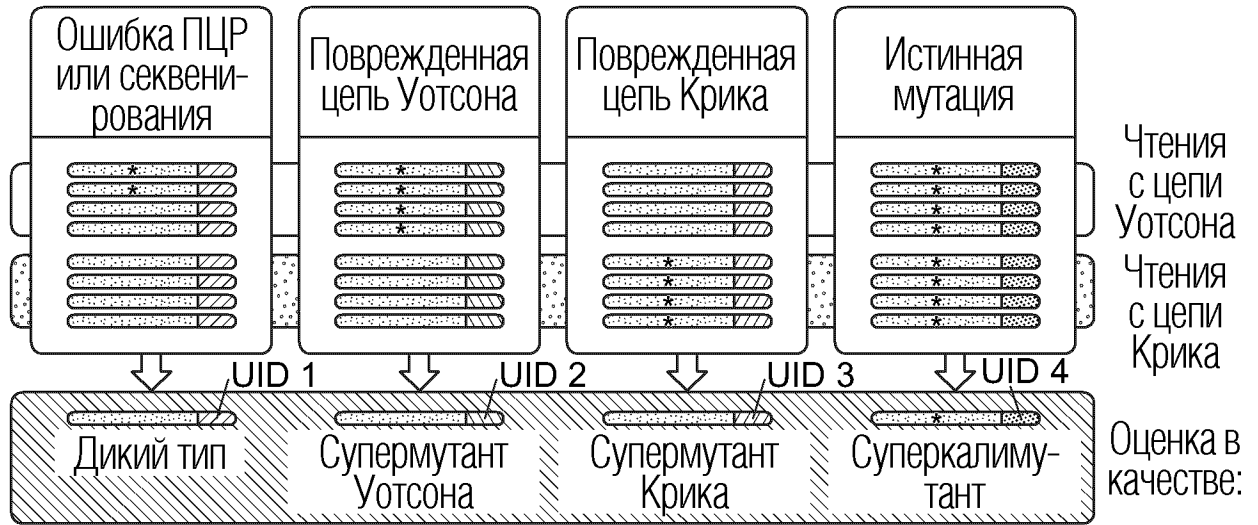
B



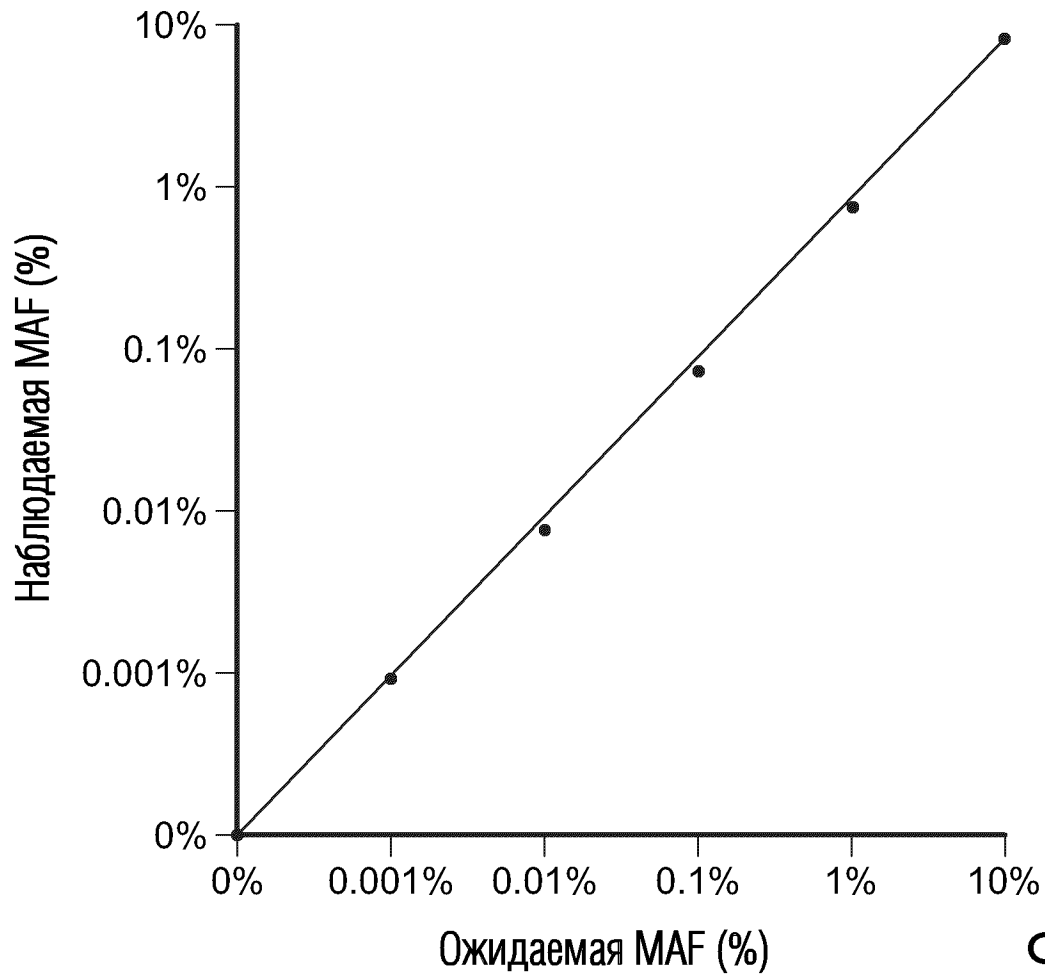
14/27

ФИГ. 22 (Продол.)

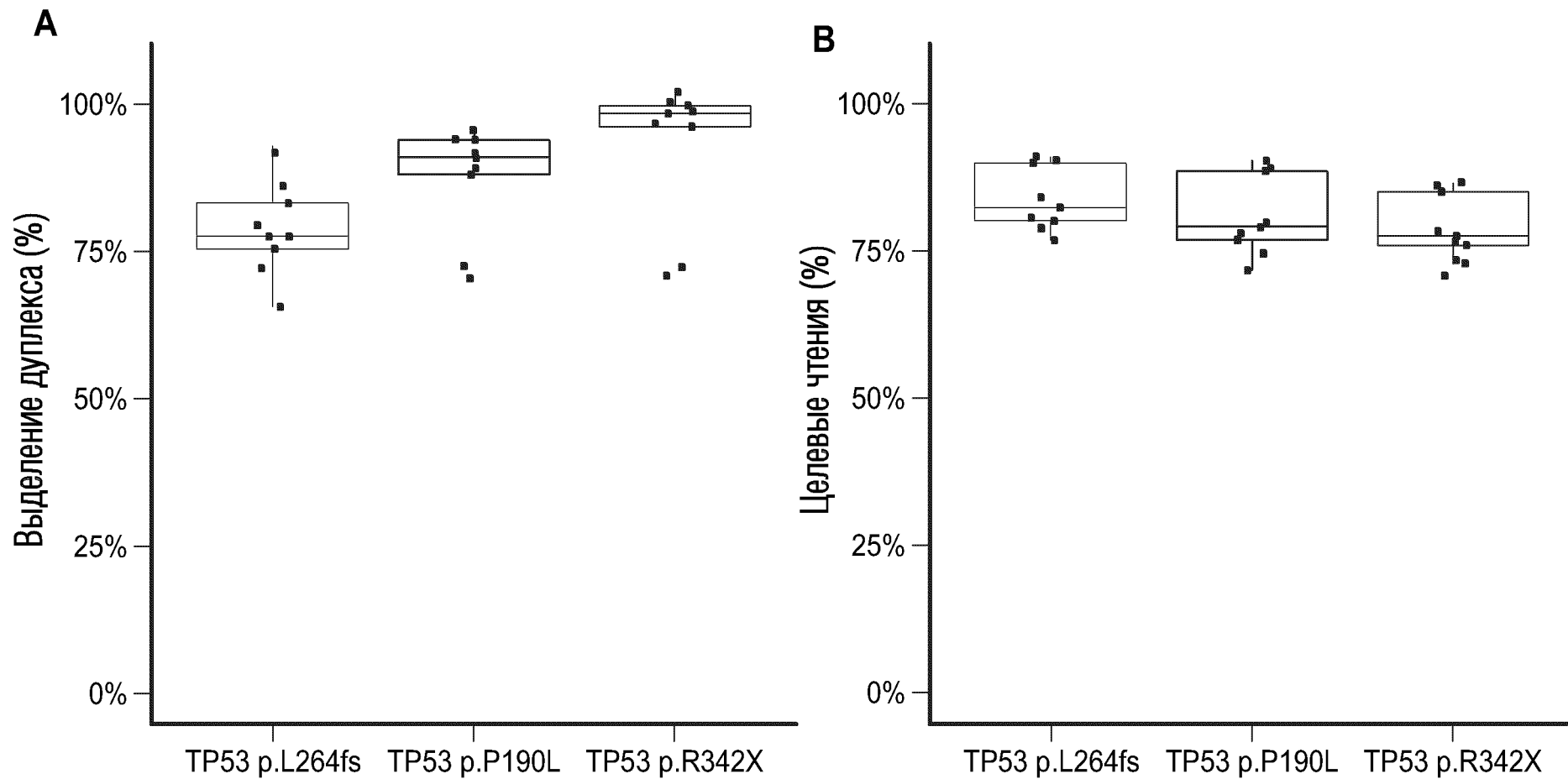
С



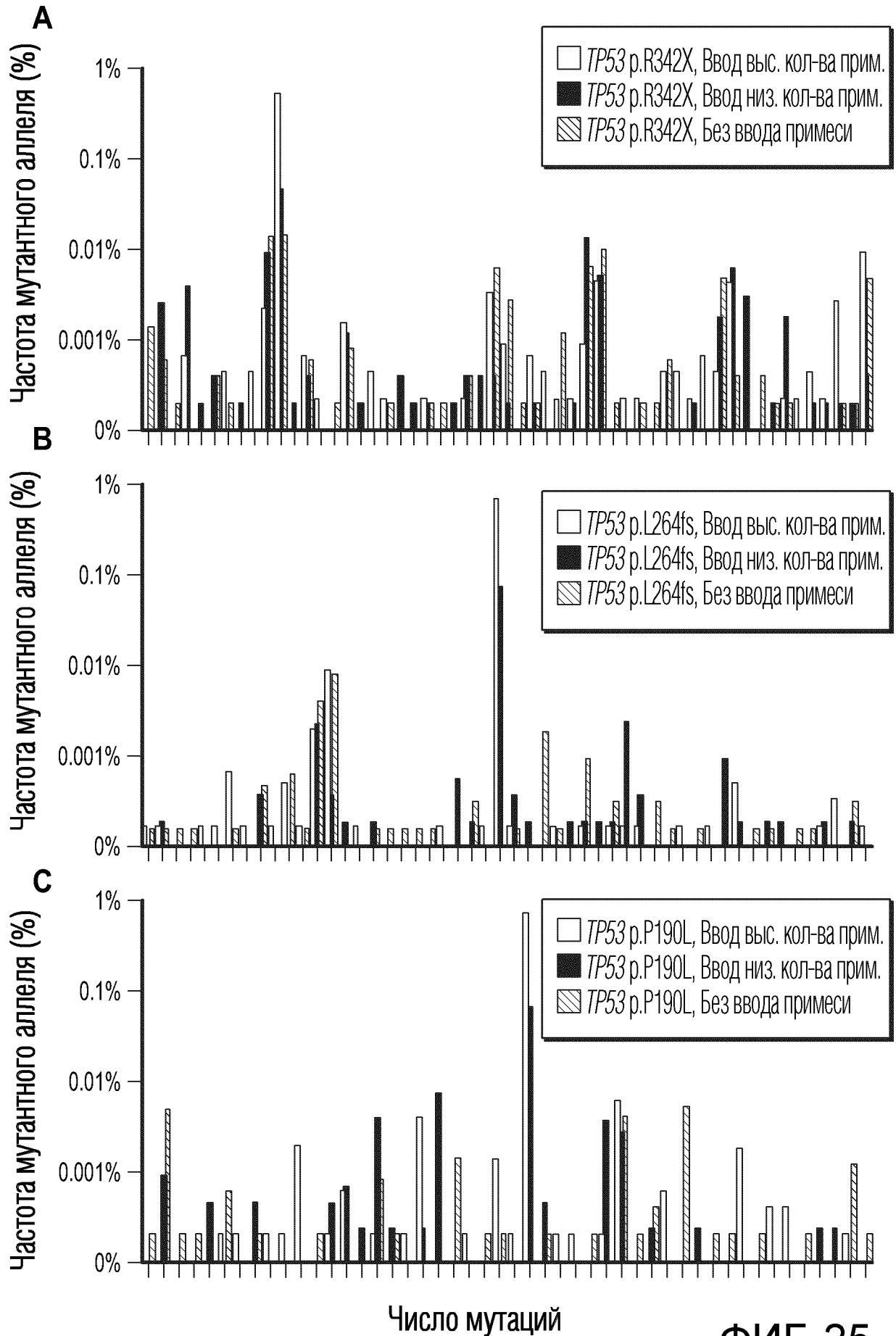
ФИГ. 22 (Продол.)

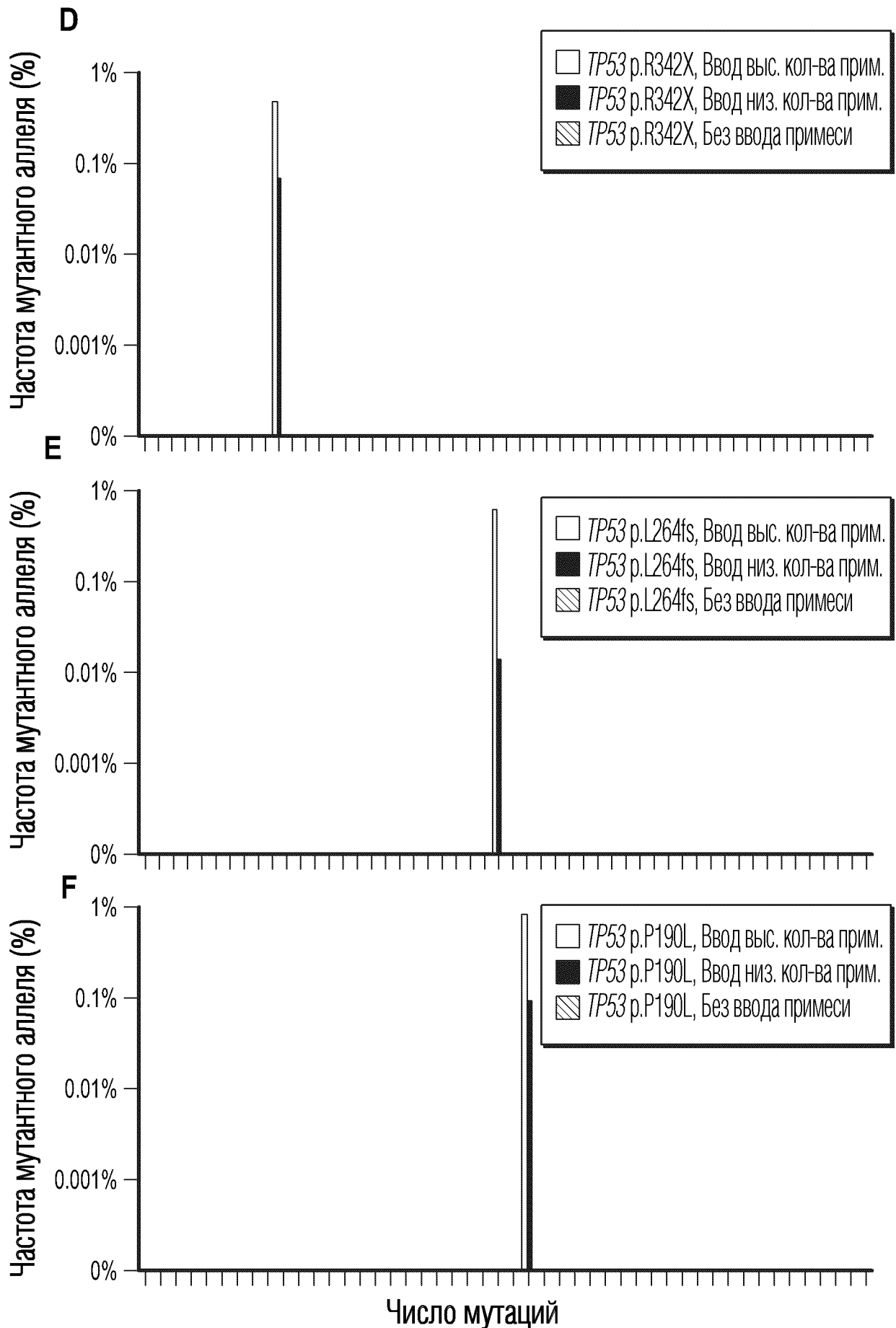


ФИГ. 23

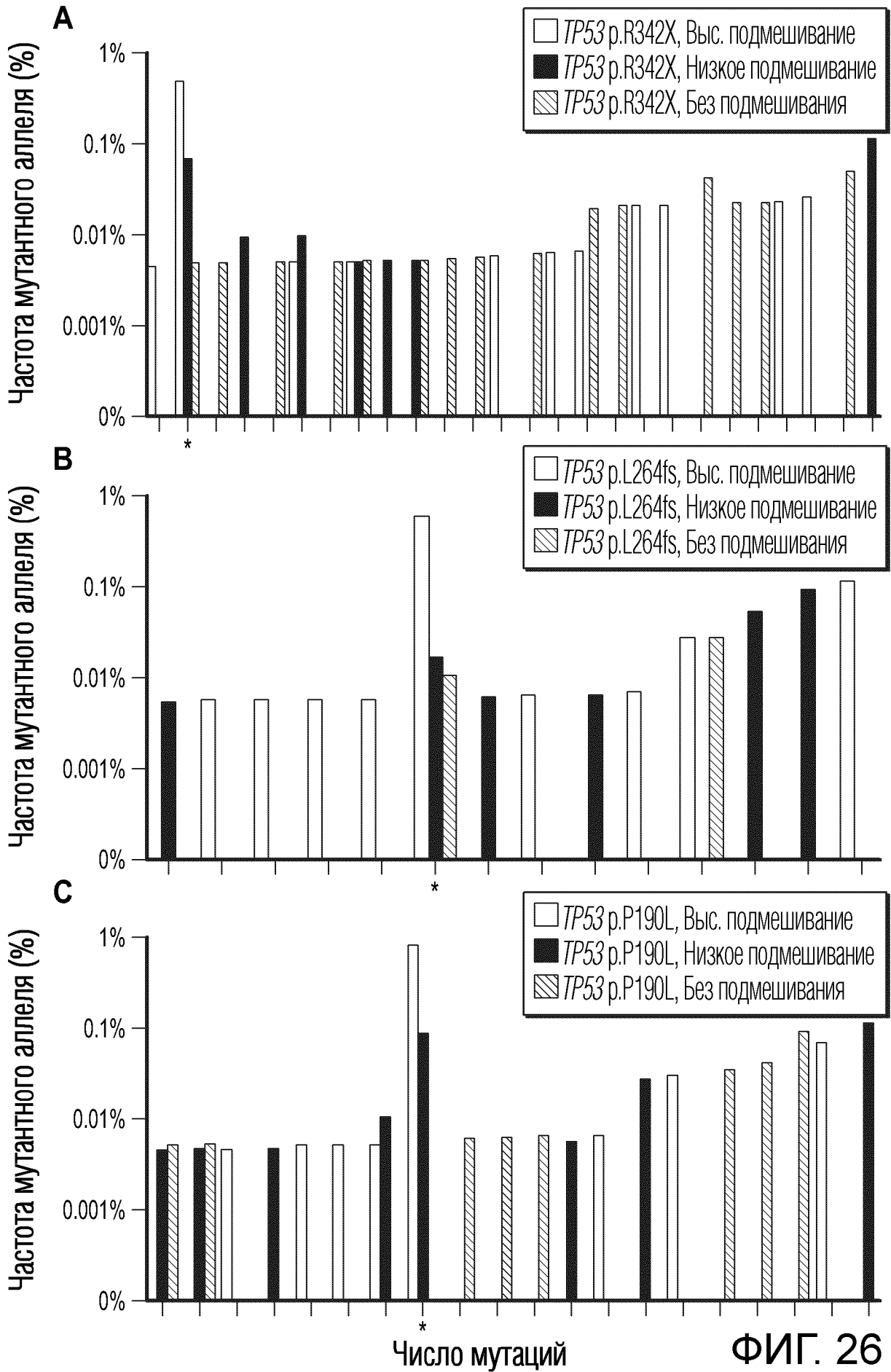


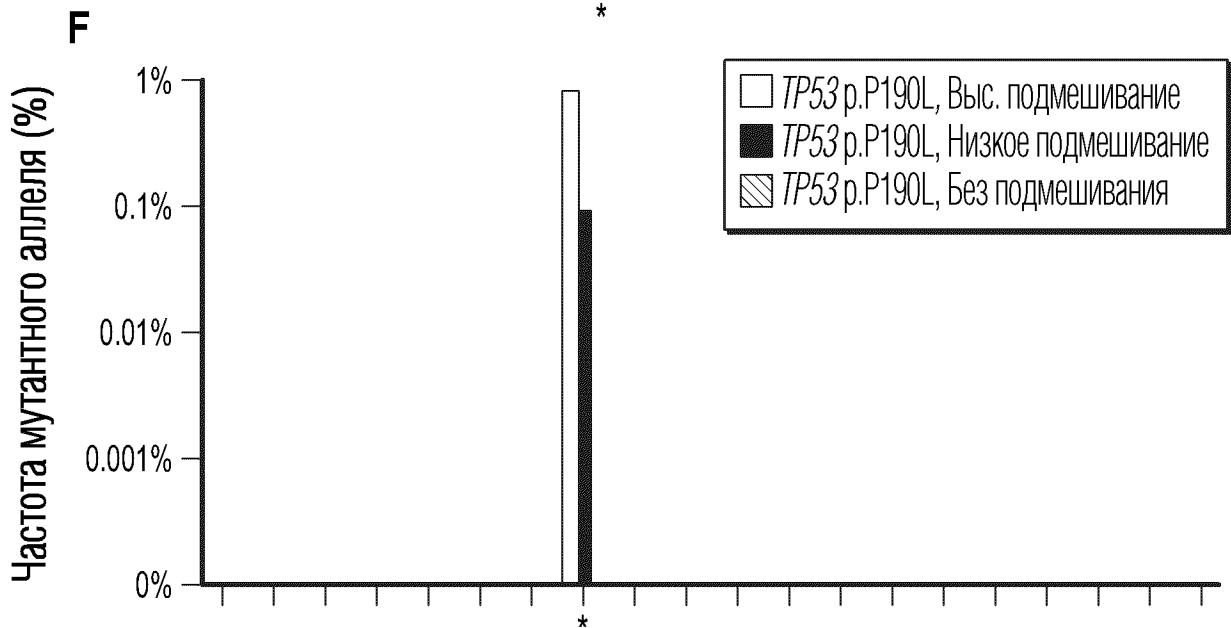
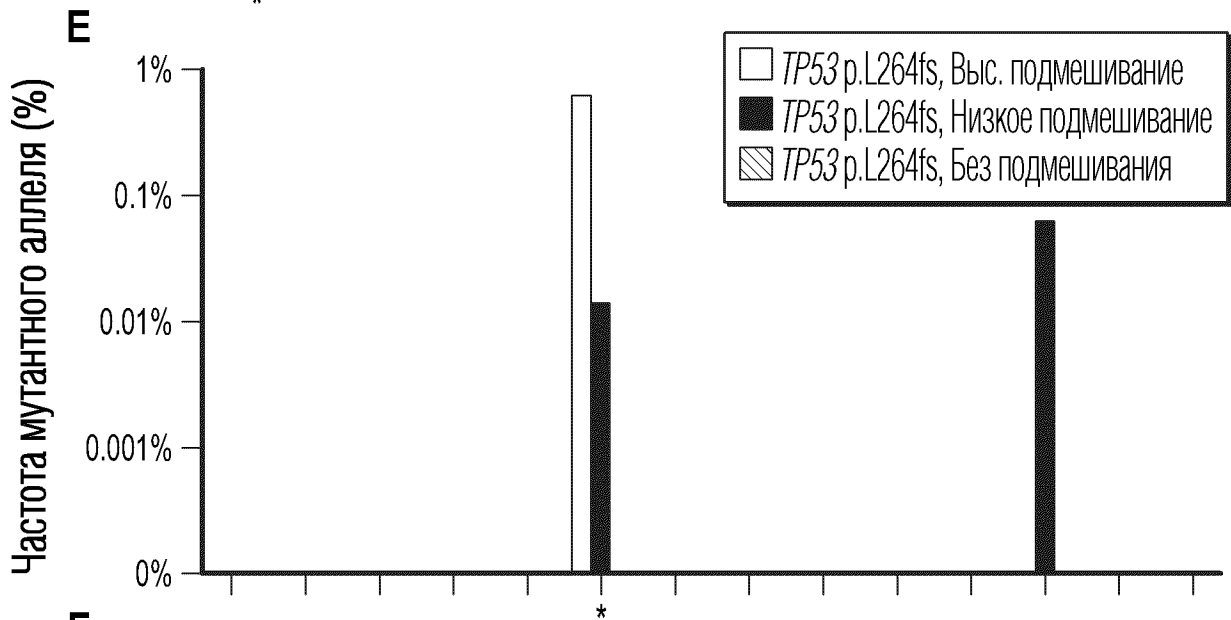
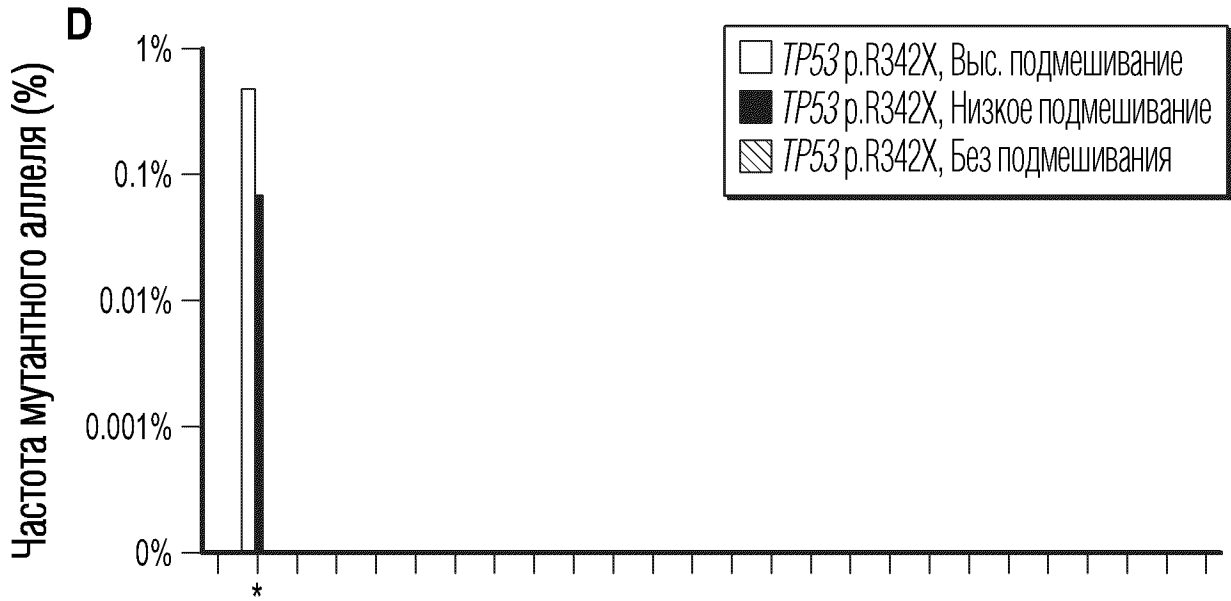
ФИГ. 24





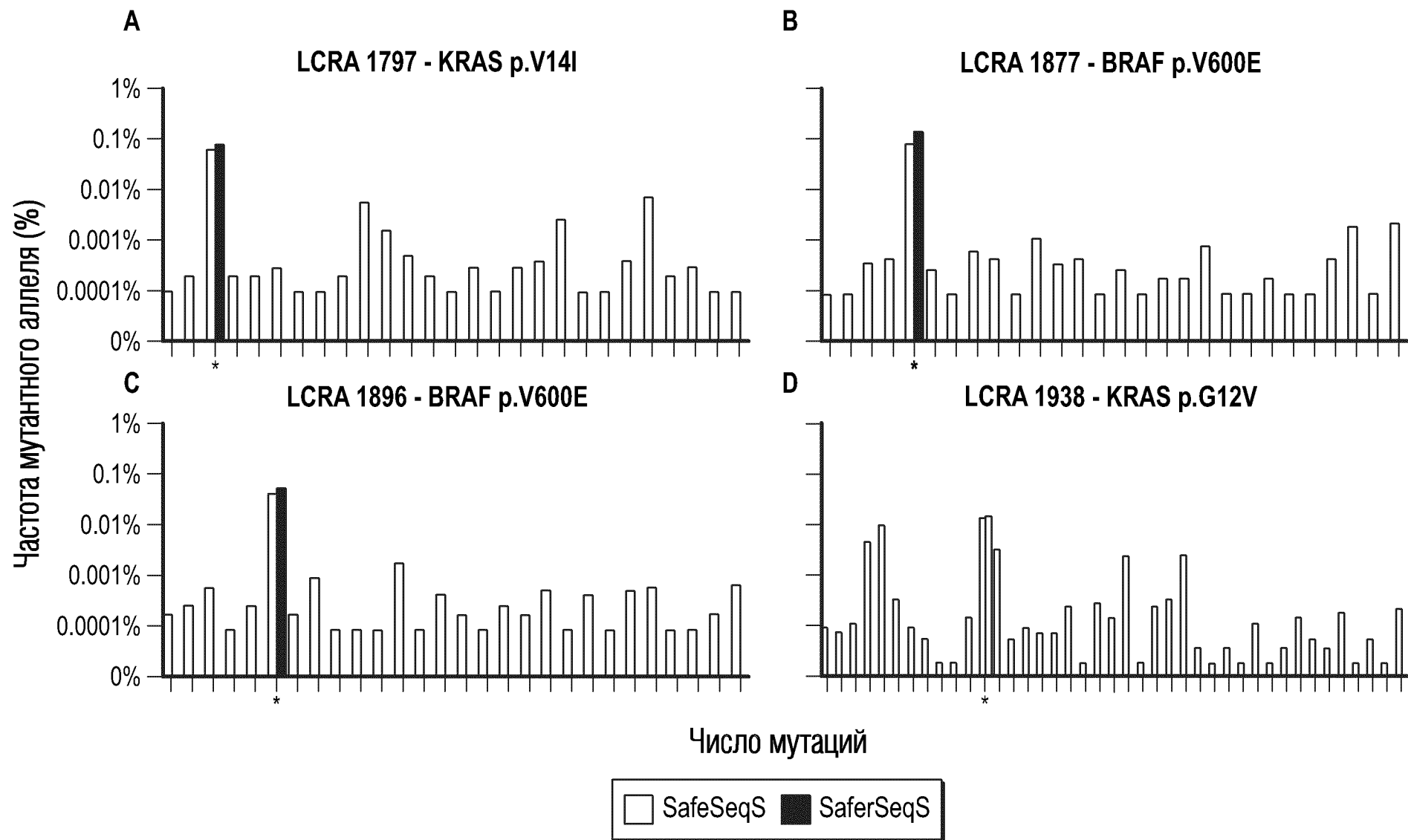
ФИГ. 25 (Продол.)





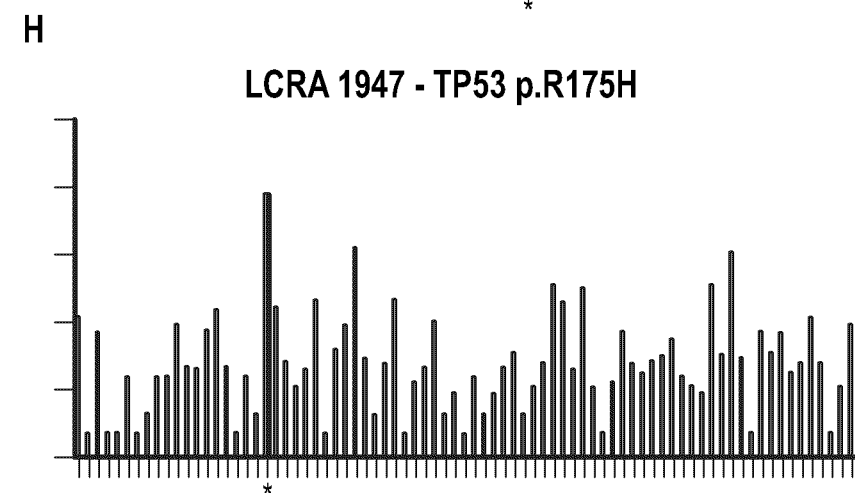
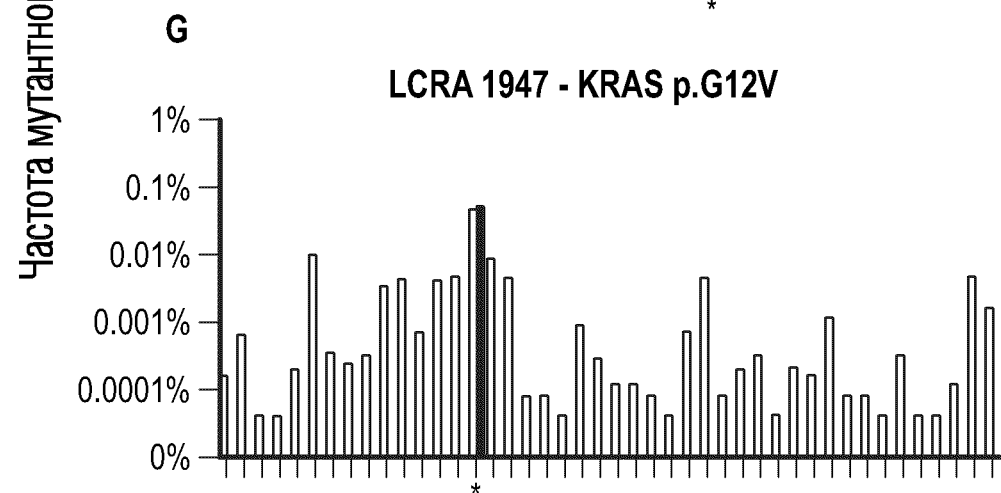
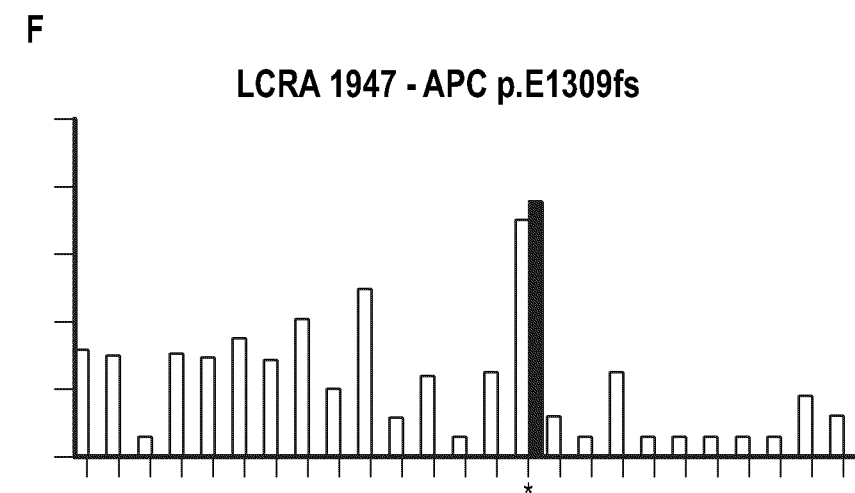
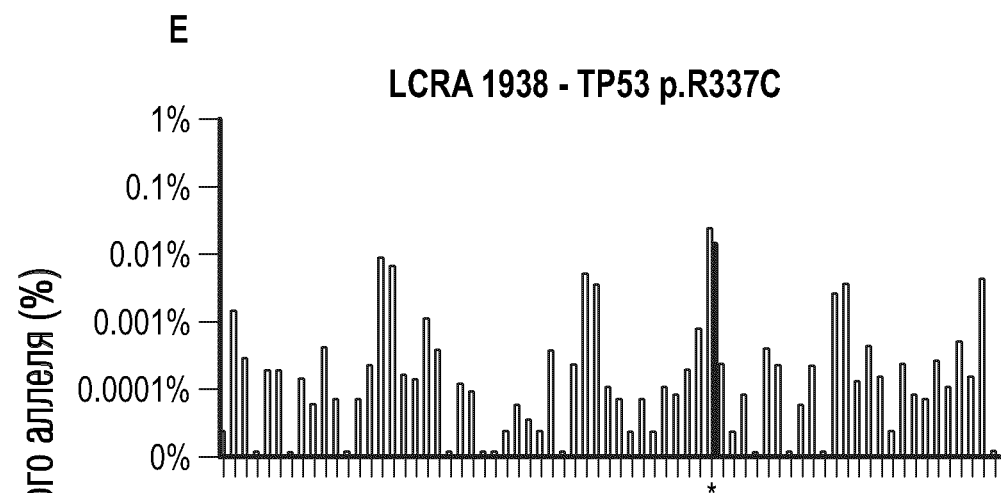
Число мутаций

ФИГ. 26 (Продол.)



21/27

ФИГ. 27



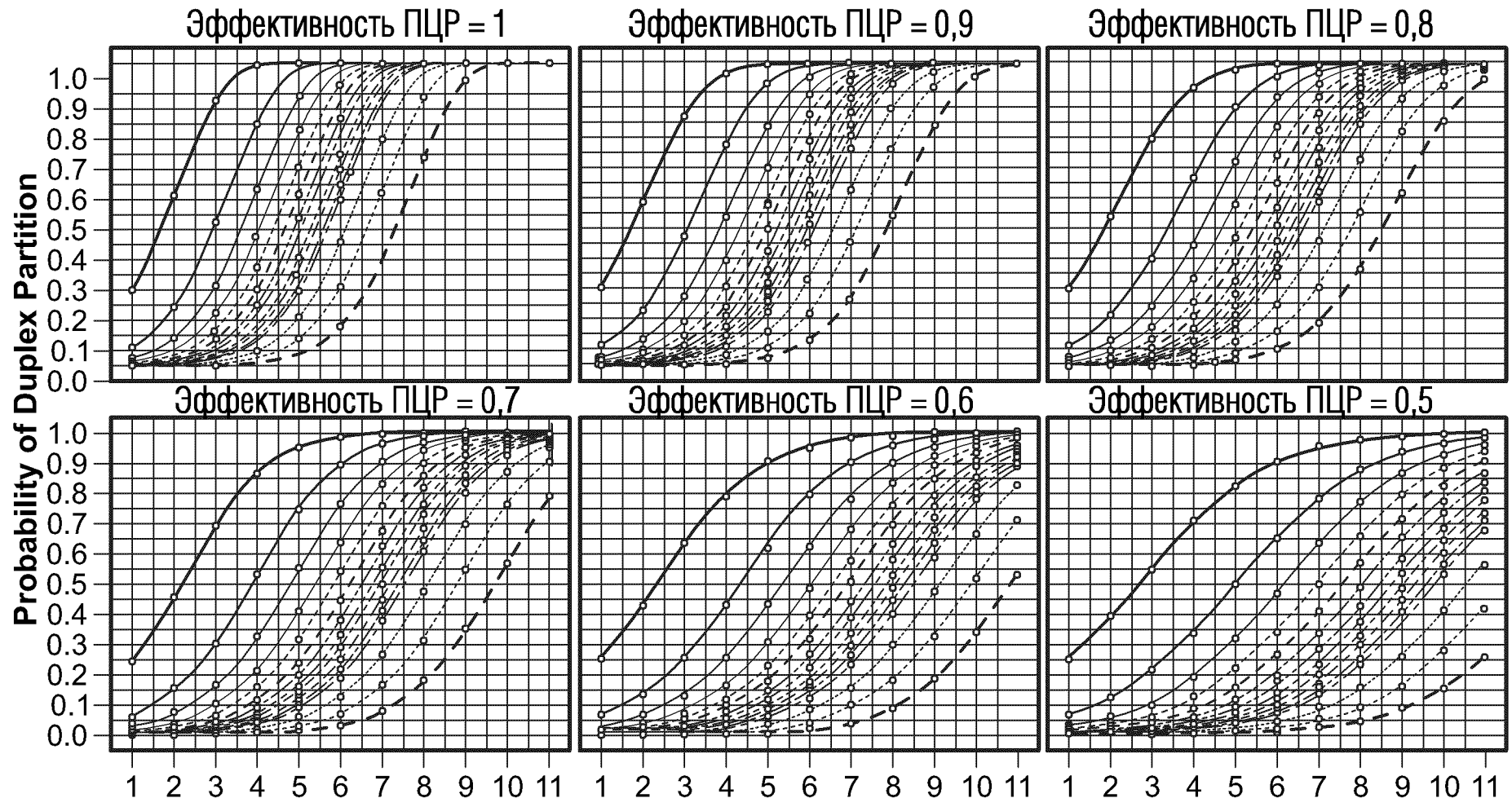
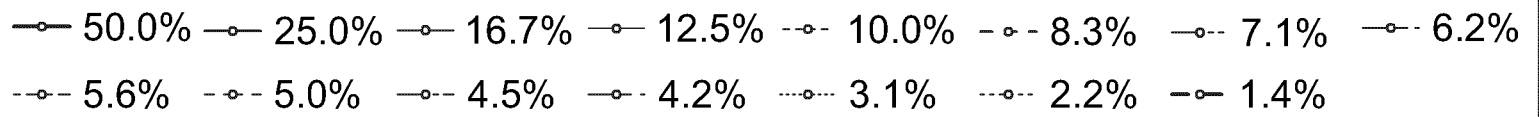
Число мутаций



22/27

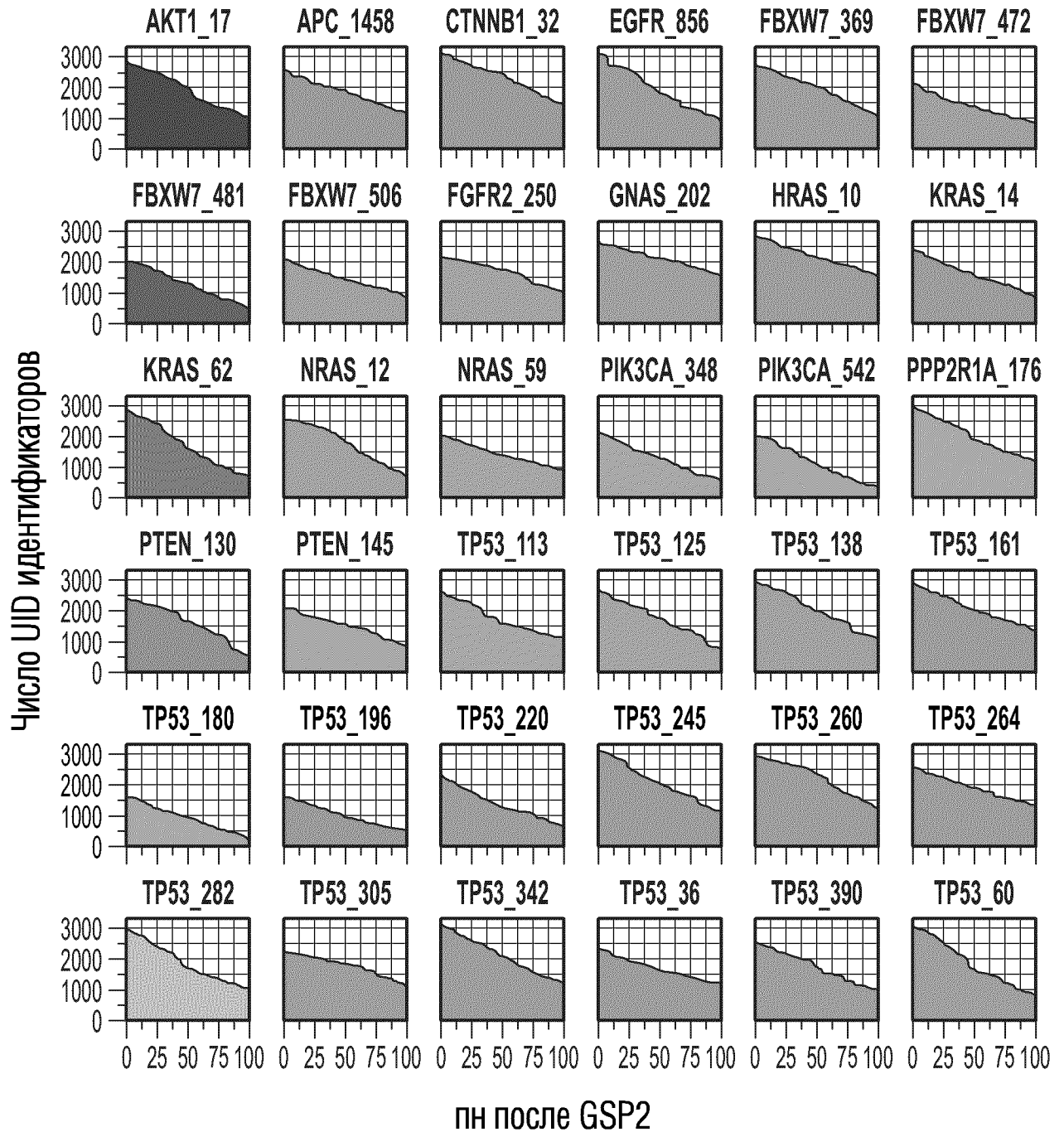
ФИГ. 27 (Продол.)

% библиотеки

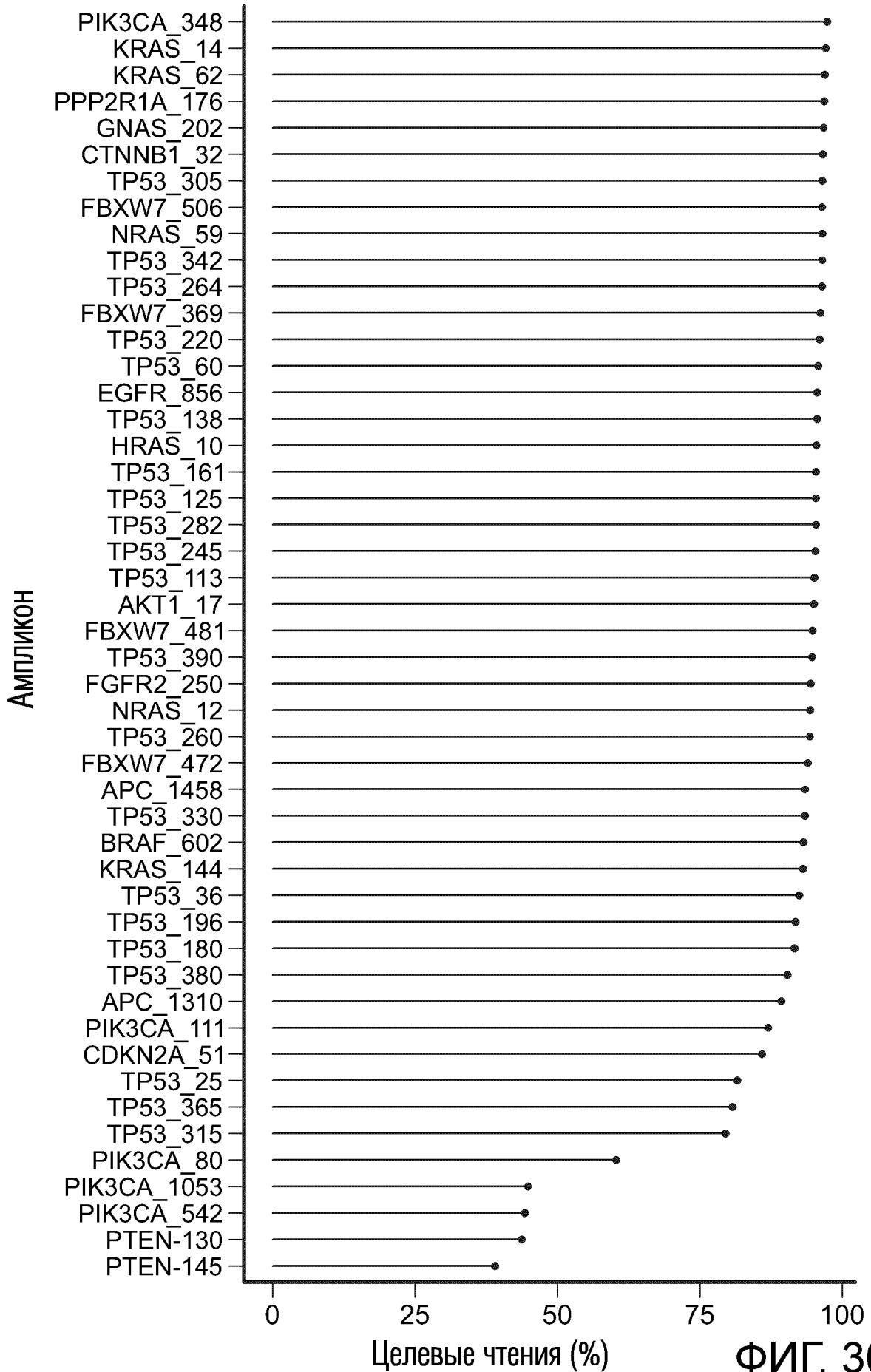


Циклы амплификации библиотеки

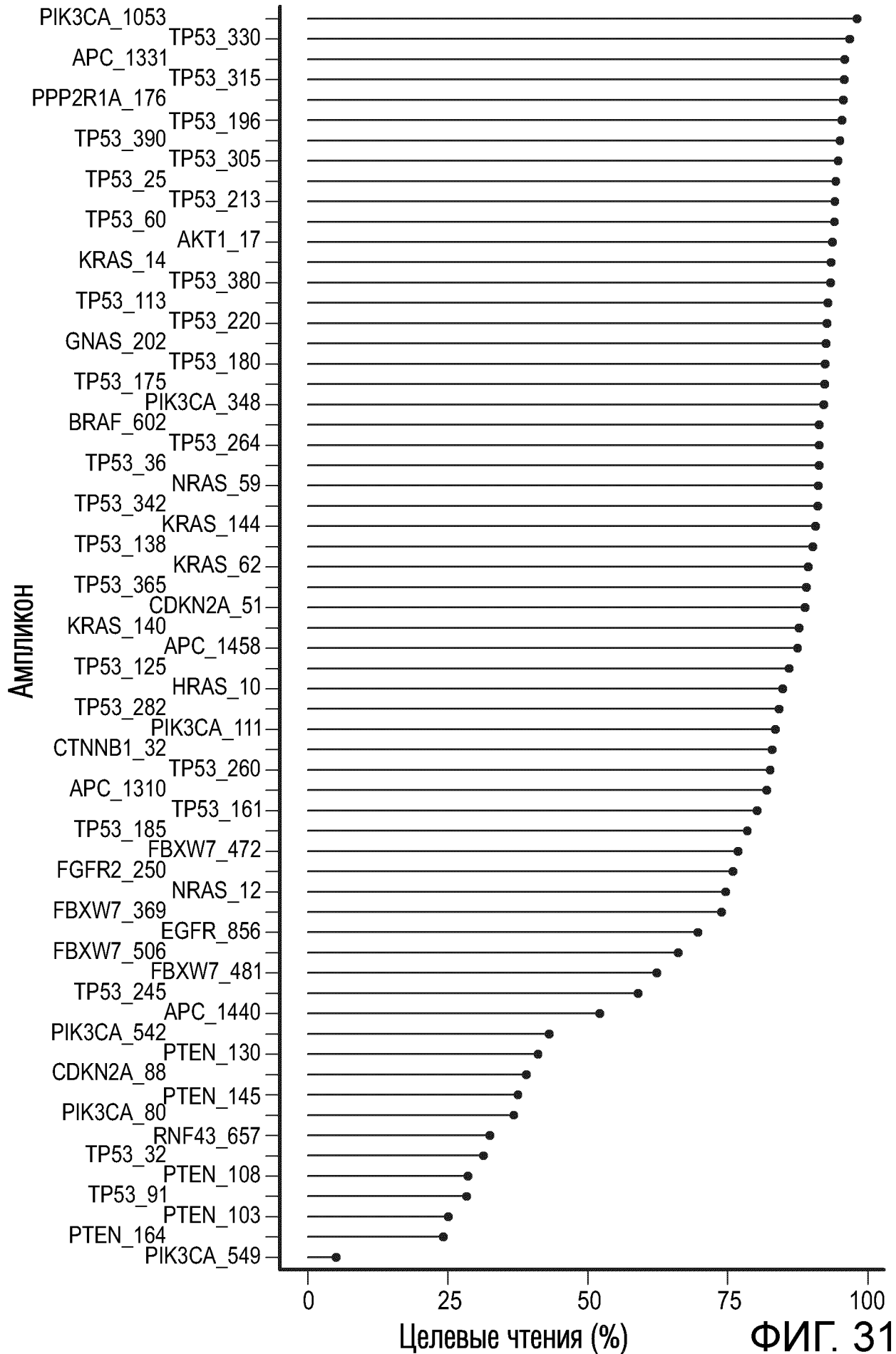
ФИГ. 28



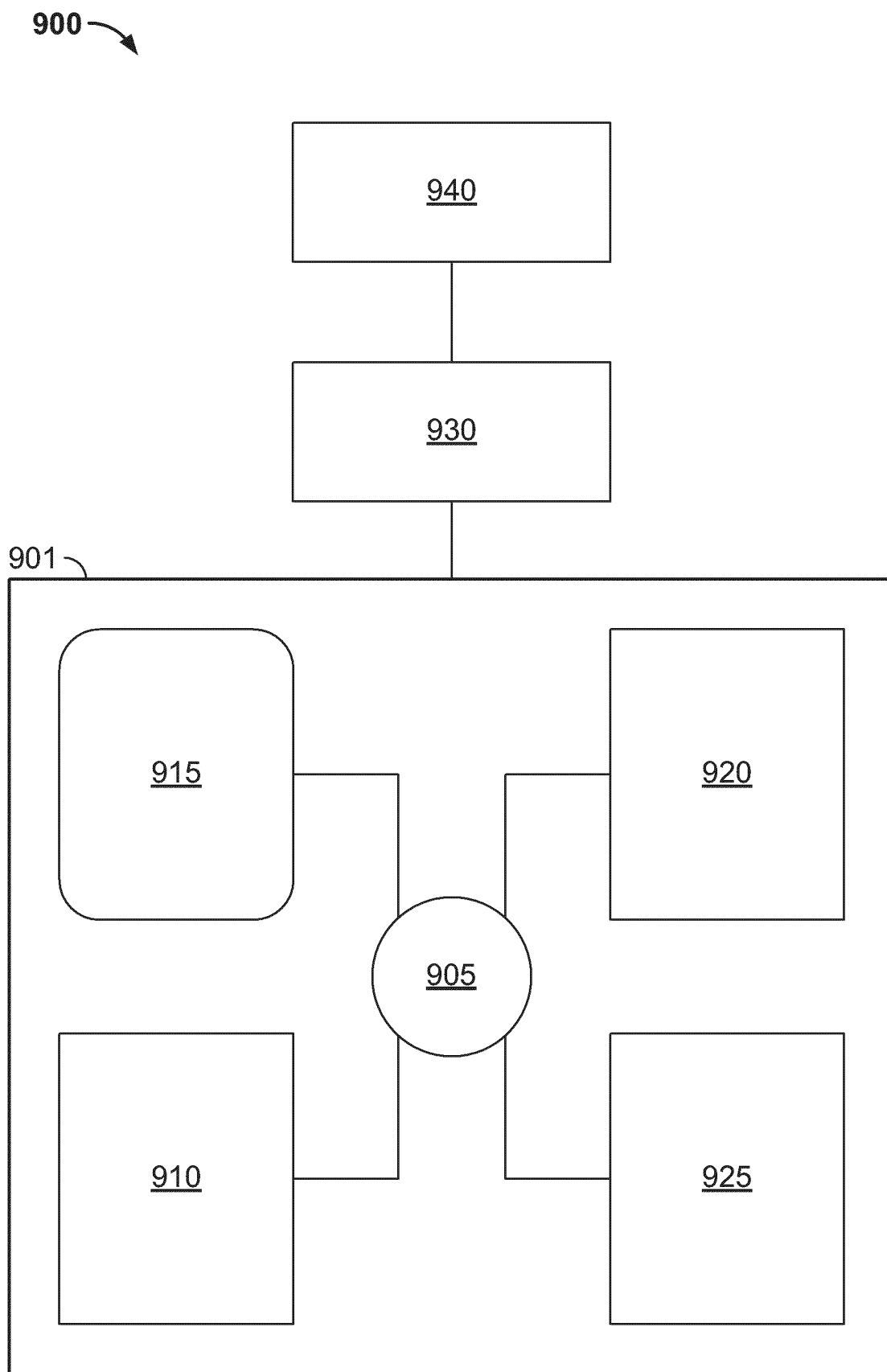
ФИГ. 29



ФИГ. 30



ФИГ. 31



ФИГ. 32