

(19)



Евразийское  
патентное  
ведомство

(21) 202091673 (13) A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки  
2021.02.03

(22) Дата подачи заявки  
2019.01.28

(51) Int. Cl. C12Q 1/68 (2018.01)  
C12Q 1/6806 (2018.01)  
C12Q 1/6827 (2018.01)  
C12Q 1/6844 (2018.01)  
C12Q 1/6855 (2018.01)  
C12Q 1/6858 (2018.01)  
C12Q 1/686 (2018.01)  
C12Q 1/6869 (2018.01)  
C12P 19/34 (2006.01)

(54) СПОСОБ АМПЛИФИКАЦИИ НУКЛЕИНОВЫХ КИСЛОТ

(31) 62/623,471

(32) 2018.01.29

(33) US

(86) PCT/US2019/015452

(87) WO 2019/148119 2019.08.01

(71) Заявитель:

СТ. ДЖУД ЧИЛДРЕН'С РИСЕРЧ  
ХОСПИТАЛ, ИНК. (US)

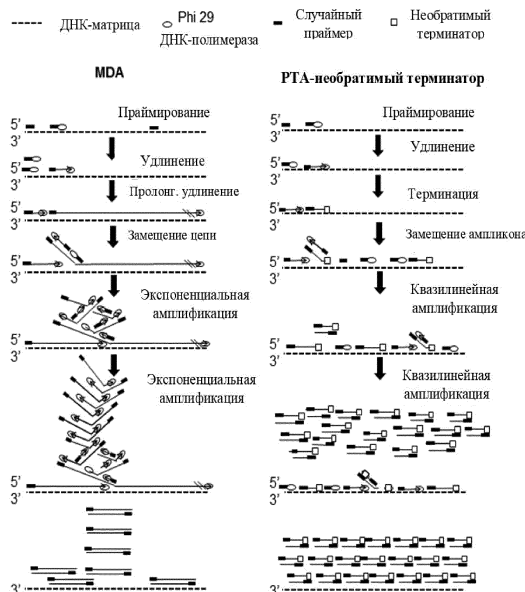
(72) Изобретатель:

Гавад Чарльз, Истон Джон, Гонзалез-  
Пена Вероника (US)

(74) Представитель:

Строкова О.В., Глухарёва А.О., Лыгу  
Т.Н., Угрюмов В.М., Христофоров  
А.А., Гизатуллина Е.М., Гизатуллин  
Ш.Ф., Костюшенкова М.Ю., Лебедев  
В.В., Парамонова К.В., Джермакян  
Р.В. (RU)

(57) В настоящем документе предусмотрены композиции и способы точной и масштабируемой направленной первичной матрицей амплификации (РТА) нуклеиновых кислот, способы секвенирования и их применения для исследований, диагностики и лечения.



A1

202091673

202091673

A1

# **СПОСОБ АМПЛИФИКАЦИИ НУКЛЕИНОВЫХ КИСЛОТ**

## **ОПИСАНИЕ**

### **Ссылка на родственную заявку**

**[0001]** Согласно настоящей заявке испрашивается приоритет в соответствии с предварительной заявкой на выдачу патента США № 62/623471, поданной 29 января 2018 г., полное содержание которой включено в настоящий документ посредством ссылки.

### **Предшествующий уровень техники настоящего изобретения**

**[0002]** Способы исследования, в которых используют амплификацию нуклеиновых кислот, например, секвенирование нового поколения, предоставляют большие объемы информации о сложных образцах, геномах и других источниках нуклеиновых кислот. Однако существует потребность в высокоточных, масштабируемых и эффективных способах амплификации и секвенирования нуклеиновых кислот для исследований, диагностики и лечения с использованием небольших образцов.

### **Краткое раскрытие настоящего изобретения**

**[0003]** В настоящем документе предусмотрены композиции, содержащие следующее: по меньшей мере одна целевая молекула нуклеиновой кислоты и библиотека ампликонов, где библиотека ампликонов содержит множество полинуклеотидов, полученных в результате амплификации по меньшей мере одной целевой молекулы нуклеиновой кислоты, где по меньшей мере некоторые из полинуклеотидов содержат терминаторный нуклеотид, где по меньшей мере 5% полинуклеотидов представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты. Кроме того, в настоящем документе предусмотрены композиции, в которых по меньшей мере 10% полинуклеотидов представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты. Кроме того, в настоящем документе предусмотрены композиции, в которых по меньшей мере 15% полинуклеотидов представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты. Кроме того, в настоящем документе предусмотрены композиции, в которых по меньшей мере 25% полинуклеотидов представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты. Кроме того, в настоящем документе предусмотрены композиции, в которых по меньшей мере 50% полинуклеотидов представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты. Кроме

того, в настоящем документе предусмотрены композиции, в которых 5-50% полинуклеотидов представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты. Кроме того, в настоящем документе предусмотрены композиции, в которых 5-25% полинуклеотидов представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты. Кроме того, в настоящем документе предусмотрены композиции, в которых не более 50% кумулятивной доли полинуклеотидов содержат последовательности, составляющие по меньшей мере 80% кумулятивной доли последовательностей по меньшей мере одной целевой молекулы нуклеиновой кислоты. Кроме того, в настоящем документе предусмотрены композиции, в которых не более 50% кумулятивной доли множества полинуклеотидов содержат последовательности, составляющие по меньшей мере 85% кумулятивной доли целевых последовательностей нуклеиновых кислот. Кроме того, в настоящем документе предусмотрены композиции, в которых не более 50% кумулятивной доли множества полинуклеотидов содержат последовательности, составляющие по меньшей мере 90% кумулятивной доли целевых последовательностей нуклеиновых кислот. Кроме того, в настоящем документе предусмотрены композиции, в которых библиотека ампликонов характеризуется коэффициентом Джинни, составляющим не более 0,5. Кроме того, в настоящем документе предусмотрены композиции, в которых библиотека ампликонов характеризуется коэффициентом Джинни, составляющим не более 0,4. Кроме того, в настоящем документе предусмотрены композиции, в которых множество полинуклеотидов имеют длину между около 50 и около 2000 нуклеотидов. Кроме того, в настоящем документе предусмотрены композиции, в которых полинуклеотиды имеют длину между около 400 и около 600 нуклеотидов. Кроме того, в настоящем документе предусмотрены композиции, в которых количество полинуклеотидов составляет 100-5000. Кроме того, в настоящем документе предусмотрены композиции, в которых количество полинуклеотидов составляет 250-1250. Кроме того, в настоящем документе предусмотрены композиции, в которых количество полинуклеотидов составляет по меньшей мере 100. Кроме того, в настоящем документе предусмотрены композиции, в которых количество полинуклеотидов составляет по меньшей мере 500. Кроме того, в настоящем документе предусмотрены композиции, в которых количество полинуклеотидов составляет по меньшей мере 1000. Кроме того, в настоящем документе предусмотрены композиции, в которых по меньшей мере некоторые из полинуклеотидов содержат штрих-код. Кроме того, в настоящем документе предусмотрены композиции, в которых штрих-код содержит штрих-код клетки. Кроме того, в настоящем документе предусмотрены композиции, в которых штрих-код содержит штрих-код образца. Кроме того, в настоящем документе предусмотрены

композиции, в которых по меньшей мере некоторые из полинуклеотидов содержат уникальный молекулярный идентификатор. Кроме того, в настоящем документе предусмотрены композиции, в которых множество полинуклеотидов содержит последовательности, по меньшей мере частично репрезентативные для генома. Кроме того, в настоящем документе предусмотрены композиции, в которых множество полинуклеотидов содержит последовательности, по меньшей мере частично репрезентативные по меньшей мере для двух геномов. Кроме того, в настоящем документе предусмотрены композиции, в которых множество полинуклеотидов содержит последовательности из кДНК. Кроме того, в настоящем документе предусмотрены композиции, в которых по меньшей мере 90% полинуклеотидов содержат терминаторный нуклеотид. Кроме того, в настоящем документе предусмотрены композиции, в которых по меньшей мере 98% полинуклеотидов содержат терминаторный нуклеотид. Кроме того, в настоящем документе предусмотрены композиции, в которых терминаторный нуклеотид прикреплен к 3' концу по меньшей мере некоторых полинуклеотидов. Кроме того, в настоящем документе предусмотрены композиции, в которых терминаторный нуклеотид выбран из группы, состоящей из следующего: нуклеотиды с модификацией альфа-группы, C3 спейсерные нуклеотиды, запертые нуклеиновые кислоты (LNA), инвертированные нуклеиновые кислоты, 2' фторнуклеотиды, 3' фосфорилированные нуклеотиды, 2'-О-метил-модифицированные нуклеотиды и транснуклеиновые кислоты. Кроме того, в настоящем документе предусмотрены композиции, в которых нуклеотиды с модификацией альфа-группы представляют собой альфа-тио-дидезоксинуклеотиды. Кроме того, в настоящем документе предусмотрены композиции, в которых терминаторный нуклеотид содержит модификации группы γ 3' углерода дезоксирибозы. Кроме того, в настоящем документе предусмотрены композиции, в которых терминаторный нуклеотид выбран из группы, состоящей из следующего: 3' блокированный обратимый терминатор, содержащий нуклеотиды, 3' неблокированный обратимый терминатор, содержащий нуклеотиды, терминаторы, содержащие 2' модификации дезоксинуклеотидов, терминаторы, содержащие модификации азотистого основания дезоксинуклеотидов, и их комбинации. Кроме того, в настоящем документе предусмотрены композиции, в которых терминаторные нуклеотиды выбраны из группы, состоящей из следующего: дидезоксинуклеотиды, инвертированные дидезоксинуклеотиды, 3' биотинилированные нуклеотиды, 3' аминокислотные нуклеотиды, 3'-фосфорилированные нуклеотиды, 3'-О-метилнуклеотиды, 3' углеродные спейсерные нуклеотиды, включая в себя 3' C3 спейсерные нуклеотиды, 3' C18 нуклеотиды, 3' гександиоловые спейсерные нуклеотиды, ациклонуклеотиды и их

комбинации. Кроме того, в настоящем документе предусмотрены композиции, причем композиция содержится в капле.

**[0004]** В настоящем документе предусмотрены композиции содержащий: по меньшей мере одна целевая молекула нуклеиновой кислоты и библиотека ампликонов, где библиотека ампликонов содержит множество полинуклеотидов, полученных в результате амплификации по меньшей мере одной целевой молекулы нуклеиновой кислоты, где по меньшей мере некоторые из полинуклеотидов содержат терминаторный нуклеотид, где не более 50% кумулятивной доли полинуклеотидов содержат последовательности, составляющие по меньшей мере 80% кумулятивной доли последовательностей по меньшей мере одной целевой молекулы нуклеиновой кислоты. Кроме того, в настоящем документе предусмотрены композиции, в которых не более 50% кумулятивной доли множества полинуклеотидов содержат последовательности, составляющие по меньшей мере 85% кумулятивной доли целевых последовательностей нуклеиновых кислот. Кроме того, в настоящем документе предусмотрены композиции, в которых не более 50% кумулятивной доли множества полинуклеотидов содержат последовательности по меньшей мере 90% кумулятивной доли целевых последовательностей нуклеиновых кислот. Кроме того, в настоящем документе предусмотрены композиции, в которых множество полинуклеотидов имеет длину между около 50 и около 2000 нуклеотидов. Кроме того, в настоящем документе предусмотрены композиции, в которых полинуклеотиды имеют длину между около 400 и около 600 нуклеотидов. Кроме того, в настоящем документе предусмотрены композиции, в которых количество полинуклеотидов составляет 100-5000. Кроме того, в настоящем документе предусмотрены композиции, в которых количество полинуклеотидов составляет 250-1250. Кроме того, в настоящем документе предусмотрены композиции, в которых количество полинуклеотидов составляет по меньшей мере 100. Кроме того, в настоящем документе предусмотрены композиции, в которых количество полинуклеотидов составляет по меньшей мере 500. Кроме того, в настоящем документе предусмотрены композиции, в которых количество полинуклеотидов составляет по меньшей мере 1000. Кроме того, в настоящем документе предусмотрены композиции, в которых по меньшей мере некоторые из полинуклеотидов содержат штрих-код. Кроме того, в настоящем документе предусмотрены композиции, в которых штрих-код содержит штрих-код клетки. Кроме того, в настоящем документе предусмотрены композиции, в которых штрих-код содержит штрих-код образца. Кроме того, в настоящем документе предусмотрены композиции, в которых по меньшей мере некоторые из полинуклеотидов содержат уникальный молекулярный идентификатор. Кроме того, в настоящем документе предусмотрены композиции, в которых множество полинуклеотидов содержит последовательности, по

меньшей мере частично репрезентативные для генома. Кроме того, в настоящем документе предусмотрены композиции, в которых множество полинуклеотидов содержит последовательности, по меньшей мере частично репрезентативные по меньшей мере для двух геномов. Кроме того, в настоящем документе предусмотрены композиции, в которых множество полинуклеотидов содержит последовательности из кДНК. Кроме того, в настоящем документе предусмотрены композиции, в которых по меньшей мере 90% полинуклеотидов содержат терминаторный нуклеотид. Кроме того, в настоящем документе предусмотрены композиции, в которых по меньшей мере 98% полинуклеотидов содержат терминаторный нуклеотид. Кроме того, в настоящем документе предусмотрены композиции, в которых терминаторный нуклеотид прикреплен к 3' концу по меньшей мере некоторых полинуклеотидов. Кроме того, в настоящем документе предусмотрены композиции, в которых терминаторный нуклеотид выбран из группы, состоящей из следующего: нуклеотиды с модификацией альфа-группы, C3 спейсерные нуклеотиды, запертые нуклеиновые кислоты (LNA), инвертированные нуклеиновые кислоты, 2' фторнуклеотиды, 3' фосфорилированные нуклеотиды, 2'-О-метил-модифицированные нуклеотиды и транснуклеиновые кислоты. Кроме того, в настоящем документе предусмотрены композиции, в которых нуклеотиды с модификацией альфа-группы представляют собой альфа-тио-дидезоксинуклеотиды. Кроме того, в настоящем документе предусмотрены композиции, в которых терминаторный нуклеотид содержит модификации группы г 3' углерода дезоксирибозы. Кроме того, в настоящем документе предусмотрены композиции, в которых терминаторный нуклеотид выбран из группы, состоящей из следующего: 3' блокированный обратимый терминатор, содержащий нуклеотиды, 3' неблокированный обратимый терминатор, содержащий нуклеотиды, терминаторы, содержащие 2' модификации дезоксинуклеотидов, терминаторы, содержащие модификации азотистого основания дезоксинуклеотидов, и их комбинации. Кроме того, в настоящем документе предусмотрены композиции, в которых терминаторные нуклеотиды выбраны из группы, состоящей из следующего: дидезоксинуклеотиды, инвертированные дидезоксинуклеотиды, 3' биотинилированные нуклеотиды, 3' аминонуклеотиды, 3'-фосфорилированные нуклеотиды, 3'-О-метилнуклеотиды, 3' углеродные спейсерные нуклеотиды, включая в себя 3' C3 спейсерные нуклеотиды, 3' C18 нуклеотиды, 3' гександиоловые спейсерные нуклеотиды, ациклонуклеотиды и их комбинации. Кроме того, в настоящем документе предусмотрены композиции, причем композиция содержится в капле.

**[0005]** В настоящем документе предусмотрены способы амплификации целевой молекулы нуклеиновой кислоты, причем способ включает следующее: приведение

в контакт образца, содержащего целевую молекулу нуклеиновой кислоты, по меньшей мере одного праймера для амплификации, по меньшей мере одной полимеразы нуклеиновых кислот и смеси нуклеотидов, где смесь нуклеотидов содержит по меньшей мере один терминаторный нуклеотид, который прекращает репликацию нуклеиновой кислоты полимеразой, и амплификация целевой молекулы нуклеиновой кислоты для создания множества терминированных продуктов амплификации, где репликация происходит путем репликации с замещением цепи. Кроме того, в настоящем документе предусмотрены способы, при которых амплификацию проводят при по существу изотермических условиях. Кроме того, в настоящем документе предусмотрены способы, при которых амплификацию проводят при условиях, в которых температура варьируется не более чем на 10° С. Кроме того, в настоящем документе предусмотрены способы, при которых амплификацию проводят при условиях, в которых температура варьируется не более чем на 5° С. Кроме того, в настоящем документе предусмотрены способы, при которых полимеразы нуклеиновых кислот представляет собой ДНК-полимеразу. Кроме того, в настоящем документе предусмотрены способы, при которых ДНК-полимераза представляет собой замещающую цепи ДНК-полимеразу. Кроме того, в настоящем документе предусмотрены способы, при которых полимеразы нуклеиновых кислот представляет собой следующее: полимеразы бактериофага phi29 (Ф29), генетически модифицированная ДНК-полимераза phi29 (Ф29), фрагмент Кленова ДНК-полимеразы I, ДНК-полимераза фага M2, ДНК-полимераза фага phiPRD1, ДНК-полимераза Bst, ДНК-полимераза большого фрагмента Bst, полимеразы экзо(-) Bst, ДНК-полимераза экзо(-) Bca, ДНК-полимераза Bsu, ДНК-полимераза Vent<sub>R</sub>, ДНК-полимераза Vent<sub>R</sub> (экзо-), ДНК-полимераза Deep Vent, ДНК-полимераза Deep Vent (экзо-), ДНК-полимераза IsoPol, ДНК-полимераза I, ДНК-полимераза Terminator, ДНК-полимераза T5, секвенаса, ДНК-полимераза T7, T7-секвенаса или ДНК-полимераза T4. Кроме того, в настоящем документе предусмотрены способы, при которых полимеразы нуклеиновых кислот характеризуется 3'→5' экзонуклеазной активностью, и по меньшей мере один терминаторный нуклеотид ингибирует 3'→5' экзонуклеазную активность. Кроме того, в настоящем документе предусмотрены способы, при которых полимеразы нуклеиновых кислот не характеризуется 3'→5' экзонуклеазной активностью. Кроме того, в настоящем документе предусмотрены способы, при которых полимеразы представляет собой следующее: ДНК-полимераза Bst, полимеразы экзо(-) Bst, ДНК-полимераза экзо(-) Bca, ДНК-полимераза Bsu, ДНК-полимераза Vent<sub>R</sub> (экзо-), ДНК-полимераза Deep Vent (экзо-), ДНК-полимераза фрагмента Кленова (экзо-) или ДНК-полимераза Terminator. Кроме того, в настоящем документе предусмотрены способы, при которых по меньшей мере один терминаторный нуклеотид содержит модификации группы

г 3' углерода дезоксирибозы. Кроме того, в настоящем документе предусмотрены способы, при которых по меньшей мере один терминаторный нуклеотид выбран из группы, состоящей из следующего: 3' блокированный обратимый терминатор, содержащий нуклеотиды, 3' неблокированный обратимый терминатор, содержащий нуклеотиды, терминаторы, содержащие 2' модификации дезоксинуклеотидов, терминаторы, содержащие модификации азотистого основания дезоксинуклеотидов, и их комбинации. Кроме того, в настоящем документе предусмотрены способы, при которых по меньшей мере один терминаторный нуклеотид выбран из группы, состоящей из следующего: дидезоксинуклеотиды, инвертированные дидезоксинуклеотиды, 3' биотинилированные нуклеотиды, 3' аминонуклеотиды, 3'-фосфорилированные нуклеотиды, 3'-О-метилнуклеотиды, 3' углеродные спейсерные нуклеотиды, включая в себя 3' C3 спейсерные нуклеотиды, 3' C18 нуклеотиды, 3' гександиоловые спейсерные нуклеотиды, ациклонуклеотиды и их комбинации. Кроме того, в настоящем документе предусмотрены способы, при которых по меньшей мере один терминаторный нуклеотид выбран из группы, состоящей из следующего: нуклеотиды с модификацией альфа-группы, C3 спейсерные нуклеотиды, запертые нуклеиновые кислоты (LNA), инвертированные нуклеиновые кислоты, 2' фторнуклеотиды, 3' фосфорилированные нуклеотиды, 2'-О-метилмодифицированные нуклеотиды и транснуклеиновые кислоты. Кроме того, в настоящем документе предусмотрены способы, при которых нуклеотиды с модификацией альфа-группы представляют собой альфа-тио-дидезоксинуклеотиды. Кроме того, в настоящем документе предусмотрены способы, при которых праймеры для амплификации имеют длину 4 - 70 нуклеотидов. Кроме того, в настоящем документе предусмотрены способы, при которых по меньшей мере один праймер для амплификации имеет длину 4 - 20 нуклеотидов. Кроме того, в настоящем документе предусмотрены способы, причем способ дополнительно включает дополнительную стадию амплификации с использованием ПЦР. Кроме того, в настоящем документе предусмотрены способы, при которых по меньшей мере один праймер для амплификации содержит рандомизированную область. Кроме того, в настоящем документе предусмотрены способы, при которых рандомизированная область имеет длину 4 - 20 нуклеотидов. Кроме того, в настоящем документе предусмотрены способы, при которых рандомизированная область имеет длину 8 - 15 нуклеотидов. Кроме того, в настоящем документе предусмотрены способы, в которых продукты амплификации имеют длину между около 50 и около 2000 нуклеотидов. Кроме того, в настоящем документе предусмотрены способы, в которых продукты амплификации имеют длину между около 200 и около 1000 нуклеотидов. Кроме того, в настоящем документе предусмотрены способы идентификации низкочастотных вариантов последовательности.



Кроме того, в настоящем документе предусмотрены способы, в которых низкочастотные варианты последовательности составляют  $\geq 0,01\%$  от общего числа последовательностей.

Кроме того, в настоящем документе предусмотрены способы, в которых низкочастотные варианты последовательности составляют  $\geq 0,05\%$  от общего числа последовательностей.

Кроме того, в настоящем документе предусмотрены способы, в которых низкочастотные варианты последовательности составляют  $\geq 0,10\%$  от общего числа последовательностей.

**[0006]** В настоящем документе предусмотрены способы секвенирования целевой молекулы нуклеиновой кислоты, причем способ включает следующее: приведение в контакт образца, содержащего целевую молекулу нуклеиновой кислоты, по меньшей мере одного праймера для амплификации, по меньшей мере одной полимеразы нуклеиновых кислот и смеси нуклеотидов, где смесь нуклеотидов содержит по меньшей мере один терминаторный нуклеотид, который прекращает репликацию нуклеиновой кислоты полимеразой, и амплификация целевой молекулы нуклеиновой кислоты для создания множества терминированных продуктов амплификации, где репликация происходит путем репликации с замещением цепи; удаление по меньшей мере одного терминаторного нуклеотида из терминированных продуктов амплификации; лигирование молекул, полученных в способе, с адапторами, тем самым создавая библиотеку продуктов амплификации; и секвенирование библиотеки продуктов амплификации. Кроме того, в настоящем документе предусмотрены способы, дополнительно включающие репарацию концов и добавление А-хвостов к 3'-концу ДНК с тупыми концами. Кроме того, в настоящем документе предусмотрены способы, при которых целевая нуклеиновая кислота представляет собой ДНК. Кроме того, в настоящем документе предусмотрены способы, при которых ДНК представляет собой кДНК. Кроме того, в настоящем документе предусмотрены способы, при которых ДНК представляет собой геномную ДНК. Кроме того, в настоящем документе предусмотрены способы, при которых по меньшей мере один праймер для амплификации содержит два или больше праймеров. Кроме того, в настоящем документе предусмотрены способы, при которых по меньшей мере один праймер для амплификации представляет собой случайный праймер. Кроме того, в настоящем документе предусмотрены способы, при которых по меньшей мере один праймер для амплификации содержит штрих-код. Кроме того, в настоящем документе предусмотрены способы, при которых штрих-код содержит штрих-код клетки. Кроме того, в настоящем документе предусмотрены способы, при которых штрих-код содержит штрих-код образца. Кроме того, в настоящем документе предусмотрены способы, при которых праймеры для амплификации содержат уникальный молекулярный идентификатор (UMI). Кроме того, в настоящем документе предусмотрены способы, причем способ дополнительно включает

денатурирование целевой нуклеиновой кислоты или геномной ДНК перед начальным отжигом праймера. Кроме того, в настоящем документе предусмотрены способы, при которых денатурацию проводят в щелочных условиях с последующей нейтрализацией. Кроме того, в настоящем документе предусмотрены способы, при которых образец, праймеры для амплификации, полимеразы нуклеиновых кислот и смесь нуклеотидов содержатся в микрожидкостном устройстве. Кроме того, в настоящем документе предусмотрены способы, при которых образец, праймеры для амплификации, полимеразы нуклеиновых кислот и смесь нуклеотидов содержатся в капле. Кроме того, в настоящем документе предусмотрены способы, при которых образец выбран из следующего: образцы ткани(ей), клетки, образцы биологической жидкости, образцы костного мозга, образцы спермы, биопсийные образцы, образцы рака, образцы опухоли, образцы клеточного лизата, криминалистические образцы, археологические образцы, палеонтологические образцы, образцы инфекции, производственные образцы, целые растения, части растения, образцы микробиоты, вирусные препараты, образцы почвы, образцы морской воды, образцы пресной воды, бытовые или промышленные образцы и их комбинации и изоляты. Кроме того, в настоящем документе предусмотрены способы, при которых биологические жидкости выбраны из следующего: кровь, моча, слюна, лимфатическая жидкость, спинномозговая жидкость (CSF), амниотическая жидкость, плевральная жидкость, перикардальная жидкость, асцит и водянистая влага. Кроме того, в настоящем документе предусмотрены способы, причем способ дополнительно включает дополнительную стадию амплификации с использованием ПЦР.

**[0007]** В настоящем документе предусмотрены способы секвенирования одиночной клетки, причем способ включает следующее: получение клеточного лизата из одиночной клетки; приведение в контакт клеточного лизата по меньшей мере с одним праймером для амплификации, по меньшей мере одной полимеразой нуклеиновых кислот и смесью нуклеотидов, где смесь нуклеотидов содержит по меньшей мере один терминаторный нуклеотид, который прекращает репликацию нуклеиновой кислоты полимеразой, и амплификация целевой молекулы нуклеиновой кислоты для создания множества терминированных продуктов амплификации, где репликация происходит путем репликации с замещением цепи; удаление по меньшей мере одного терминаторного нуклеотида из терминированных продуктов амплификации; лигирование молекул, полученных в способе, с адапторами, тем самым создавая библиотеку продуктов амплификации; и секвенирование библиотеки продуктов амплификации. Кроме того, в настоящем документе предусмотрены способы, при которых клеточный лизис сопровождается протеолизом. Кроме того, в настоящем документе предусмотрены

способы, при которых образец, по меньшей мере один праймер для амплификации, полимеразы нуклеиновых кислот и смесь нуклеотидов содержатся в микрожидкостном устройстве. Кроме того, в настоящем документе предусмотрены способы, при которых образец, по меньшей мере один праймер для амплификации, полимеразы нуклеиновых кислот и смесь нуклеотидов содержатся в капле. Кроме того, в настоящем документе предусмотрены способы, при которых по меньшей мере один праймер для амплификации прикреплен к твердой подложке. Кроме того, в настоящем документе предусмотрены способы, при которых твердая подложка представляет собой шарик. Кроме того, в настоящем документе предусмотрены способы, при которых по меньшей мере один праймер для амплификации прикреплен к твердой подложке посредством расщепляемого линкера. Кроме того, в настоящем документе предусмотрены способы, при которых по меньшей мере один праймер для амплификации содержит штрих-код. Кроме того, в настоящем документе предусмотрены способы, причем способ дополнительно включает расщепление расщепляемого линкера перед амплификацией. Кроме того, в настоящем документе предусмотрены способы, при которых клетку выбирают из следующего: клетка животного, растительная клетка, грибковая клетка, бактериальная клетка и клетка простейшего. Кроме того, в настоящем документе предусмотрены способы, при которых клетка животного представляет собой клетку человека. Кроме того, в настоящем документе предусмотрены способы, при которых клетку выбирают из следующего: клетка предимплантационного эмбриона, стволовая клетка, фетальная клетка, опухолевая клетка, предполагаемая раковая клетка, раковая клетка, клетка, подвергнутая процедуре редактирования генов, клетка из патогенного организма, клетка, полученная из криминалистического образца, клетка, полученная из археологического образца, и клетка, полученная из палеонтологического образца. Кроме того, в настоящем документе предусмотрены способы, при которых клетка предимплантационного эмбриона представляет собой бластомер. Кроме того, в настоящем документе предусмотрены способы, при которых бластомер получают из эмбриона восьмиклеточной стадии, полученного с помощью *in vitro* оплодотворения. Кроме того, в настоящем документе предусмотрены способы, дополнительно включающие определение наличия предрасполагающих к заболеванию вариантов зародышевой линии или соматических вариантов в эмбриональной клетке. Кроме того, в настоящем документе предусмотрены способы, при которых патогенный организм представляет собой бактерию, гриб или простейшее. Кроме того, в настоящем документе предусмотрены способы, при которых клетку, полученную из патогенного организма, получают из жидкости, взятой у пациента, образца микробиоты или постоянного изделия медицинского назначения. Кроме того, в

настоящем документе предусмотрены способы, дополнительно включающие стадию определения идентичности патогенного организма. Кроме того, в настоящем документе предусмотрены способы, дополнительно включающие определение наличия генетических вариантов, ответственных за устойчивость патогенного организма к лечению. Кроме того, в настоящем документе предусмотрены способы, при которых клетка представляет собой опухолевую клетку, предполагаемую раковую клетку или раковую клетку. Кроме того, в настоящем документе предусмотрены способы, дополнительно включающие определение наличия одной или более диагностических или прогностических мутаций. Кроме того, в настоящем документе предусмотрены способы, дополнительно включающие определение наличия вариантов зародышевой линии или соматических вариантов, ответственных за устойчивость к лечению. Кроме того, в настоящем документе предусмотрены способы, при которых клетка представляет собой клетку, подвергнутую процедуре редактирования генов. Кроме того, в настоящем документе предусмотрены способы, дополнительно включающие определение наличия незапланированных мутаций, вызванных процессом редактирования генов. Кроме того, в настоящем документе предусмотрены способы, дополнительно включающие определение истории клеточной родословной. Кроме того, в настоящем документе предусмотрены способы идентификации низкочастотных вариантов последовательности. Кроме того, в настоящем документе предусмотрены способы, при которых низкочастотные варианты последовательности составляют  $\geq 0,01\%$  от общего числа последовательностей. Кроме того, в настоящем документе предусмотрены способы, при которых низкочастотные варианты последовательности составляют  $\geq 0,05\%$  от общего числа последовательностей. Кроме того, в настоящем документе предусмотрены способы, при которых низкочастотные варианты последовательности составляют  $\geq 0,10\%$  от общего числа последовательностей. Кроме того, в настоящем документе предусмотрены способы, причем способ дополнительно включает дополнительную стадию амплификации с использованием ПЦР.

**[0008]** В настоящем документе предусмотрены способы определения мутагенности условия окружающей среды, причем способ включает следующее: воздействие на клетки условия окружающей среды; выделение одиночных клеток из популяции; получение клеточного лизата из одиночной клетки; приведение в контакт клеточного лизата по меньшей мере с одним праймером для амплификации, по меньшей мере одной полимеразой нуклеиновых кислот и смесью нуклеотидов, где смесь нуклеотидов содержит по меньшей мере один терминаторный нуклеотид, который прекращает репликацию нуклеиновой кислоты полимеразой, и амплификация целевой молекулы нуклеиновой кислоты для создания множества терминированных продуктов

амплификации, где репликация происходит путем репликации с замещением цепи; удаление по меньшей мере одного терминаторного нуклеотида из терминированных продуктов амплификации; лигирование молекул, полученных в способе, с адапторами, тем самым создавая библиотеку продуктов амплификации; и секвенирование библиотеки продуктов амплификации, и сравнение последовательностей продуктов амплификации по меньшей мере с одной эталонной последовательностью для идентификации мутаций. Кроме того, в настоящем документе предусмотрены способы, при которых одиночная клетка представляет собой клетку человека. Кроме того, в настоящем документе предусмотрены способы, при которых условие окружающей среды включает в себя химическое вещество. Кроме того, в настоящем документе предусмотрены способы, при которых условие окружающей среды включает в себя облучение. Кроме того, в настоящем документе предусмотрены способы, при которых условие окружающей среды включает в себя ультрафиолетовое излучение. Кроме того, в настоящем документе предусмотрены способы, при которых одиночные клетки происходят из печени, кожи, почки, крови или легкого. Кроме того, в настоящем документе предусмотрены способы, при которых по меньшей мере некоторые из продуктов амплификации содержат штрих-код. Кроме того, в настоящем документе предусмотрены способы, при которых штрих-код содержит штрих-код клетки. Кроме того, в настоящем документе предусмотрены способы, при которых штрих-код содержит штрих-код образца. Кроме того, в настоящем документе предусмотрены способы, при которых по меньшей мере некоторые из праймеров для амплификации содержат уникальный молекулярный идентификатор (UMI). Кроме того, в настоящем документе предусмотрены способы, причем способ дополнительно включает дополнительную стадию амплификации с использованием ПЦР.

#### **Включение посредством ссылки**

**[0009]** Все публикации, патенты и заявки на выдачу патентов, упомянутые в настоящем описании, включены в настоящий документ посредством ссылки в той же степени, как если бы было указано, что каждая отдельная публикация, патент или заявка на выдачу патента были специально и индивидуально включены посредством ссылки.

#### **Краткое описание графических материалов**

**[0010]** Новые признаки настоящего изобретения подробно изложены в прилагаемой формуле изобретения. Лучшее понимание признаков и преимуществ настоящего изобретения будет получено при обращении к нижеследующему подробному описанию, в котором представлены иллюстративные варианты осуществления, в которых

используются принципы настоящего изобретения, и сопровождающие графические материалы, на которых:

**[0011]** На фигуре 1А показано сравнение предшествующего способа амплификации со множественным замещением цепи (MDA) с одним из вариантов осуществления способа амплификации, направленной первичной матрицей (РТА), а именно способом РТА с необратимым терминатором.

**[0012]** На фигуре 1В показано сравнение способа РТА с необратимым терминатором с другим вариантом осуществления, а именно способом РТА с обратимым терминатором.

**[0013]** На фигуре 1С показано сравнение MDA и способа РТА с необратимым терминатором в отношении распространения мутаций.

**[0014]** На фигуре 1D показаны стадии способа, выполняемые после амплификации, которые включают удаление терминатора, репарацию концов и выполнение добавление А-хвостов к 3'-концу ДНК с тупыми концами перед лигированием адаптора. Затем перед секвенированием библиотеку объединенных клеток могут подвергать опосредованному гибридизацией обогащению в отношении всех экзонов или других конкретных областей, представляющих интерес. Клетку происхождения каждого прочтения идентифицируют штрих-кодом клетки (показан в виде зеленой и синей последовательностей).

**[0015]** На фигуре 2А показано распределение ампликонов по размерам после прохождения РТА с добавлением возрастающих концентраций терминаторов (верхний гель). Нижний гель показывает распределение ампликонов по размерам после прохождения РТА с добавлением возрастающих концентраций обратимого терминатора или добавлением возрастающих концентраций необратимого терминатора.

**[0016]** На фигуре 2В (GC) показано сравнение содержания GC секвенированных оснований для MDA и РТА.

**[0017]** На фигуре 2С показаны показатели оценки качества картирования (e) (mapQ), сопоставленные с геномом человека (p\_mapped) после того, как одиночные клетки подвергли РТА или MDA.

**[0018]** На фигуре 2D показан процент прочтений, сопоставленных с геномом человека (p\_mapped) после того, как одиночные клетки подвергли РТА или MDA.

**[0019]** На фигуре 2Е (ПЦР) показано сравнение процента прочтений, которые являются дубликатами ПЦР, для 20 миллионов подвергнутых субдискретизации прочтений после того, как одиночные клетки подвергли MDA или РТА.

**[0020]** На фигуре 3А показаны показатели оценки качества картирования (с) (mapQ2), сопоставленные с геномом человека (p\_mapped2) после того, как одиночные клетки подвергли РТА с обратимыми или необратимыми терминаторами.

**[0021]** На фигуре 3В показан процент прочтений, сопоставленные с геномом человека (p\_mapped2) после того, как одиночные клетки подвергли РТА с обратимыми или необратимыми терминаторами.

**[0022]** На фигуре 3С показана серия блочных диаграмм, на которых представлены выровненные прочтения для среднего процента прочтений, перекрывающихся с элементами Alu, с использованием различных способов. РТА характеризуется наибольшим количеством прочтений, выровненных с геномом.

**[0023]** На фигуре 3D показана серия блочных диаграмм, на которых представлены дубликации ПЦР для среднего процента прочтений, перекрывающихся с элементами Alu, с использованием различных способов.

**[0024]** На фигуре 3Е показана серия блочных диаграмм, на которых представлено содержание GC прочтений для среднего процента прочтений, перекрывающихся с элементами Alu, с использованием различных способов.

**[0025]** На фигуре 3F показана серия блочных диаграмм, на которых представлено качество картирования среднего процента прочтений, перекрывающихся с элементами Alu, с использованием различных способов. РТА характеризуется самым высоким качеством картирования исследованных способов.

**[0026]** На фигуре 3G показано сравнение широты охвата митохондриального генома SC с помощью различных способов WGA при фиксированной глубине секвенирования, составляющей 7,5X.

**[0027]** На фигуре 4А показана средняя глубина охвата 10 т.п.н. окон по хромосоме 1 после выбора клетки высококачественной MDA (представляющей ~ 50% клеток) по сравнению с клеткой, амплифицированной РТА со случайным праймером после субдискретизации каждой клетки до 40 миллионов парных прочтений. На фигуре показано, что MDA имеет меньшую однородность с гораздо большим количеством окон, которые имеют больше (прямоугольник А) или меньше (прямоугольник С), чем удвоенная средняя глубина охвата. Отсутствует охват как MDA, так и РТА на центромере из-за высокого содержания GC и низкого качества картирования повторяющихся областей (прямоугольник В).

**[0028]** На фигуре 4В показаны графики зависимости охвата секвенированием от положения в геноме для способов MDA и РТА (вверху). Нижние

блочные диаграммы показывают частоты аллелей для способов MDA и РТА по сравнению с совокупным образцом.

**[0029]** На фигуре 4С показан график зависимости среднего охвата от окна генома для однородности охвата экспериментов с MDA и РТА. РТА привела к значительно более равномерному охвату генома, чем MDA.

**[0030]** На фигуре 5А показан график доли охваченного генома в зависимости от числа прочтений генома для оценки охвата при увеличении глубины секвенирования для различных способов. Способ РТА приближается к двум совокупным образцам на каждой глубине, что является улучшением по сравнению с другими исследованными способами.

**[0031]** На фигуре 5В показан график зависимости коэффициента вариации охвата генома от количества прочтений для оценки однородности охвата. Было обнаружено, что способ РТА характеризуется наивысшей однородностью среди исследованных способов.

**[0032]** На фигуре 5С показан график Лоренца кумулятивной доли общего числа прочтений по сравнению с кумулятивной долей генома. Было обнаружено, что способ РТА характеризуется наивысшей однородностью среди исследованных способов.

**[0033]** На фигуре 5D показана серия блочных диаграмм рассчитанных коэффициентов Джини для каждого из исследованных способов, чтобы оценить разницу каждой реакции амплификации от идеальной однородности. Было обнаружено, что способ РТА воспроизводимо более однороден, чем другие исследованные способы.

**[0034]** На фигуре 5Е показан график доли обнаруженных совокупных вариантов в зависимости от количества прочтений. Частоту обнаружения вариантов для каждого из способов сравнивали с соответствующим совокупным образцом при увеличении глубины секвенирования. Для оценки чувствительности рассчитывали процент вариантов, обнаруженных в соответствующих совокупных образцах, которые были подвергнуты субдискретизации до 650 млн. прочтений, обнаруженных в каждой клетке на каждой глубине секвенирования (фиг. 3А). Улучшенный охват и однородность РТА привели к обнаружению на 30% больше вариантов по сравнению со способом Q-MDA, который являлся следующим по чувствительности способом.

**[0035]** На фигуре 5F показана серия блочных диаграмм среднего процента прочтений, перекрывающихся с элементами Alu. Способ РТА значительно уменьшил аллельный перекоп в этих гетерозиготных сайтах. Способ РТА более равномерно амплифицирует два аллеля в одной и той же клетке по сравнению с другими исследованными способами.



**[0036]** На фигуре 5G показан график зависимости специфичности обнаружений вариантов от числа прочтений для оценки специфичности обнаружений мутаций. Варианты, обнаруженные с помощью различных способов, которые не были обнаружены в совокупных образцах, считались ложноположительными. Способ РТА привел к наименьшему количеству ложноположительных обнаружений (наивысшая специфичность) исследованных способов.

**[0037]** На фигуре 5H показана доля ложноположительных замен оснований для каждого типа замены оснований с помощью различных способов. Без ограничения какой-либо теорией такие паттерны могут зависеть от полимеразы.

**[0038]** На фигуре 5I показана серия блочных диаграмм среднего процента прочтений, перекрывающихся с элементами Alu для ложноположительных обнаружений вариантов. Способ РТА привел к самым низким частотам аллелей для ложноположительных обнаружений вариантов.

**[0039]** На фигуре 6A представлено схематическое описание каталога чувствительности клонотипа к лекарственным средствам согласно настоящему раскрытию. Путем определения лекарственной чувствительности отдельных клонотипов можно создать каталог, из которого онкологи могут преобразовать клонотипы, идентифицированные в опухоли пациента, в список лекарственных средств, которые будут лучше всего нацелены на устойчивые популяции.

**[0040]** На фигуре 6B показано изменение количества лейкозных клонов с увеличением количества лейкозных клеток на клон после 100 имитаций. Используя частоту мутаций на клетку, имитация предсказывает огромное разнообразие меньших клонов, создаваемых по мере того, как одна клетка размножается до 10-100 миллиардов клеток (прямоугольник A). С помощью современных способов секвенирования выявляются только 1-5 клонов с самой высокой частотой (прямоугольник C). Согласно одному варианту осуществления настоящего изобретения представлены способы определения лекарственной устойчивости сотен клонов, которые находятся чуть ниже уровня обнаружения существующего в настоящее время способа (прямоугольник B).

**[0041]** На фигуре 7 показан иллюстративный вариант осуществления настоящего раскрытия. По сравнению с диагностическим образцом в нижнем ряду, культивирование без химиотерапии выбирало клон (красный прямоугольник, нижний правый угол), несущий активирующую мутацию KRAS. Напротив, этот клон был убит преднизолоном или даунорубицином (зеленый прямоугольник, верхний правый угол), в то время как клоны с более низкой частотой прошли положительный отбор (пунктирный прямоугольник).

**[0042]** Фигура 8 представляет собой обзор одного варианта осуществления настоящего изобретения, а именно экспериментальный дизайн для количественной оценки относительной чувствительности клонов с конкретными генотипами к конкретным лекарственным средствам.

**[0043]** На фигуре 9 (часть А) показаны шарики с олигонуклеотидами, прикрепленными с помощью расщепляемого линкера, уникальным клеточным штрих-кодом и случайным праймером. В части В показана одиночная клетка и шарик, инкапсулированные в одной и той же капле, с последующим лизисом клетки и расщеплением праймера. Затем капля может быть слита с другой каплей, содержащей смесь для амплификации РТА. В части С показано, что после амплификации капли разбивают, а ампликоны из всех клеток объединяют. Затем протокол в соответствии с настоящим раскрытием используют для удаления терминатора, репарации концов и добавление А-хвостов к 3'-концу ДНК с тупыми концами перед лигированием адаптора. Затем библиотека объединенных клеток подвергается опосредованному гибридизацией обогащению представляющими интерес экзонами перед секвенированием. Затем клетку происхождения каждого прочтения идентифицируют с помощью штрих-кода клетки.

**[0044]** На фигуре 10А показано включение клеточных штрих-кодов и/или уникальных молекулярных идентификаторов в реакции РТА с использованием праймеров, содержащих клеточные штрих-коды и/или уникальные молекулярные идентификаторы.

**[0045]** На фигуре 10В показано включение клеточных штрих-кодов и/или уникальных молекулярных идентификаторов в реакции РТА с использованием шпилечных праймеров, содержащих клеточные штрих-коды и/или уникальные молекулярные идентификаторы.

**[0046]** На фигуре 11А (РТА\_UMI) показано, что включение уникальных молекулярных идентификаторов (UMI) позволяет создавать консенсусные прочтения, снижая частоту ложноположительных результатов, вызванных секвенированием и другими ошибками, ведущими к повышенной чувствительности при выполнении обнаружения варианта зародышевой линии или соматического варианта.

**[0047]** На фигуре 11В показано, что сворачивание прочтений с одним и тем же UMI позволяет скорректировать смещения амплификации и другие смещения, которые могут привести к ложному обнаружению или ограниченной чувствительности при обнаружении вариантов количества копий.

**[0048]** На фигуре 12А показан график количества мутаций по сравнению с группами обработки для прямого измерения в эксперименте по мутагенности окружающей среды. Одиночные клетки человека подвергали воздействию носителя (VHC), маннозы

(MAN) или прямого мутагена N-этил-N-нитрозомочевины (ENU) при различных уровнях обработки и измеряли количество мутаций.

**[0049]** На фигуре 12B показана серия графиков количества мутаций по сравнению с различными группами обработки и уровнями обработки, дополнительно разделенными по типу мутаций оснований.

**[0050]** На фигуре 12C показано представление паттерна мутаций в тринуклеотидном контексте. Основания по оси y находятся в положении  $n-1$ , а основания по оси x находятся в положении  $n + 1$ . Более темные области указывают на более низкую частоту мутаций, а более светлые области указывают на более высокую частоту мутаций. Сплошные черные прямоугольники в верхнем ряду (мутации цитозина) указывают на то, что мутагенез цитозина происходит реже, когда за цитозином следует гуанин. Пунктирные черные прямоугольники в нижнем ряду (мутации тимина) указывают на то, что большинство мутаций тимина происходит в положениях, где аденин непосредственно предшествует тимину.

**[0051]** На фигуре 12D показан график сравнения местоположений известных сайтов гиперчувствительности к ДНКазе I в клетках CD34+ с соответствующими местоположениями из клеток, обработанных N-этил-N-нитрозомочевинной. Не наблюдалось значительного обогащения цитозиновыми вариантами.

**[0052]** На фигуре 12E показана доля мутаций, вызванных ENU, в сайтах гиперчувствительности (DH) к ДНКазе I. Сайты DH в CD34+ клетках, ранее каталогизированные Roadmap Epigenomics Project, использовали для исследования того, являются ли мутации ENU более распространенными в сайтах DH, которые представляют сайты открытого хроматина. Не было обнаружено значительного обогащения вариантами в сайтах DH и не наблюдалось обогащения вариантами, ограниченными цитозинами, в сайтах DH.

**[0053]** На фигуре 12F показана серия блочных диаграмм доли мутаций, индуцированных ENU, в геномных местоположениях с конкретными аннотациями. Никакого особого обогащения не наблюдалось в конкретных аннотациях в отношении вариантов (левые прямоугольники) в каждой клетке относительно доли генома (правые прямоугольники), которую содержит каждая аннотация.

### **Подробное раскрытие настоящего изобретения**

**[0054]** Существует потребность в разработке новых масштабируемых, точных и эффективных способов амплификации нуклеиновых кислот (включая в себя амплификацию генома одиночных клеток и множества клеток) и секвенирования, которые

преодолели бы ограничения в существующих способах за счет увеличения представленности последовательностей, однородности и точности воспроизводимым образом. В настоящем документе предусмотрены композиции и способы для обеспечения точной и масштабируемой амплификации, направленной первичной матрицей (РТА), и секвенирования. Кроме того, в настоящем документе предусмотрены способы определения однонуклеотидного варианта, изменения числа копий, клонотипирования и измерения мутагенности окружающей среды. Такие способы и композиции способствуют высокоточной амплификации целевых нуклеиновых кислот (или «матриц»), что увеличивает точность и чувствительность целевых приложений, таких как секвенирование нового поколения.

**[0055] Определения**

**[0056]** Если не указано иное, все технические и научные термины, используемые в настоящем документе, имеют то же значение, которое обычно понимается специалистом в настоящей области техники, к которой принадлежат настоящие изобретения.

**[0057]** В этом раскрытии числовые признаки представлены в формате диапазона. Следует понимать, что описание в формате диапазона дано просто для удобства и краткости и не должно толковаться как жесткое ограничение объема любых вариантов осуществления. Соответственно, следует считать, что описание диапазона конкретно раскрывает все возможные поддиапазоны, а также отдельные числовые значения в пределах этого диапазона с точностью до десятых долей единицы нижнего предела, если контекст явно не диктует иное. Например, описание диапазона, такого как от 1 до 6, следует рассматривать как конкретно раскрывающие поддиапазоны, такие как от 1 до 3, от 1 до 4, от 1 до 5, от 2 до 4, от 2 до 6, от 3 до 6 и т. д., а также отдельные значения в этом диапазоне, например, 1,1, 2, 2,3, 5 и 5,9. Это применимо независимо от широты диапазона. Верхний и нижний пределы этих промежуточных диапазонов могут независимо включаться в меньшие диапазоны, а также включены в настоящее изобретение, с учетом любого специально исключенного предела в указанном диапазоне. Если указанный диапазон включает в себя один или оба предела, диапазоны, исключаящие один или оба из этих включенных пределов, также включены в настоящее изобретение, если контекст явно не диктует иное.

**[0058]** Терминология, используемая в настоящем документе, предназначена только для описания конкретных вариантов осуществления и не предназначена для ограничения какого-либо варианта осуществления. Подразумевается, что используемые в настоящем документе формы единственного числа включают в себя также формы

множественного числа, если контекст явно не указывает иное. Кроме того, следует понимать, что термины «содержит» и/или «содержащий» при использовании в настоящем описании определяют наличие заявленных признаков, целых чисел, стадий, операций, элементов и/или компонентов, но не исключают наличие или добавление одного или более других признаков, целых чисел, стадий, операций, элементов, компонентов и/или их групп. Используемый в настоящем документе термин «и/или» включает в себя любые и все комбинации одного или более связанных перечисленных элементов.

**[0059]** Если специально не указано или не очевидно из контекста, используемый в настоящем документе термин «приблизительно» в отношении числа или диапазона чисел понимают как означающий указанное число и числа +/- 10% от них, или 10% ниже нижнего указанного предела и 10% выше указанного верхнего предела для значений, перечисленных для диапазона.

**[0060]** Используемые в настоящем документе термины «субъект», «пациент» или «индивидуум» относятся к животным, включая в себя млекопитающих, таких как, например, люди, ветеринарные животные (например, кошки, собаки, коровы, лошади, овцы, свиньи и т.д.) и используемые в качестве моделей заболеваний экспериментальные животные (например, мыши, крысы). Согласно настоящему изобретению можно использовать общепринятые техники молекулярной биологии, микробиологии и рекомбинантной ДНК в пределах квалификации в настоящей области техники. Такие техники полностью описаны в литературе. См., например, Sambrook, Fritsch & Maniatis, *Molecular Cloning: A Laboratory Manual*, 2-е изд. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (в настоящем документе "Sambrook *et al.*, 1989"); *DNA Cloning: A practical Approach*, Volumes I and II (D.N. Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed. 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. (1985); *Transcription and Translation* (B.D. Hames & S.J. Higgins, eds. (1984); *Animal Cell Culture* (R.I. Freshney, ed. (1986); *Immobilized Cells and Enzymes* (IRL Press, (1986); B. Perbal, *A practical Guide To Molecular Cloning* (1984); F.M. Ausubel *et al.* (eds.), *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. (1994); среди прочих.

**[0061]** Термин «нуклеиновая кислота» охватывает как многоцепочечные, так и одноцепочечные молекулы. В двухцепочечных или трехцепочечных нуклеиновых кислотах цепи нуклеиновой кислоты не обязательно должны быть одинаковой протяженности (т.е. двухцепочечная нуклеиновая кислота не обязательно должна быть двухцепочечной по всей длине обеих цепей). Описанные в настоящем документе матрицы нуклеиновых кислот могут характеризоваться любым размером в зависимости от образца (от небольших фрагментов свободно-циркулирующей ДНК до полных геномов), включая в

себя без ограничения 50-300 оснований, 100-2000 оснований, 100-750 оснований, 170-500 оснований, 100-5000 оснований, 50-10000 оснований или 50-2000 оснований в длину. В некоторых случаях матрицы имеют длину по меньшей мере 50, 100, 200, 500, 1000, 2000, 5000, 10000, 20000 50000, 100000, 200000, 500000, 1000000 или больше чем 1000000 оснований. Описанные в настоящем документе способы обеспечивают амплификацию нуклеиновых кислот, таких как матрицы нуклеиновых кислот. Описанные в настоящем документе способы дополнительно обеспечивают получение выделенных и по меньшей мере частично очищенных нуклеиновых кислот и библиотек нуклеиновых кислот. Нуклеиновые кислоты включают в себя без ограничения те, которые включают в себя следующее: ДНК, РНК, кольцевая РНК, сцДНК (свободно-циркулирующая ДНК), сцРНК (свободно-циркулирующая РНК), киРНК (короткая интерферирующая РНК), сцфДНК (свободно-циркулирующая фетальная ДНК), мРНК, тРНК, рРНК, миРНК (микроРНК), синтетические полинуклеотиды, аналоги полинуклеотидов, любая другая нуклеиновая кислота, соответствующая описанию, или любые их комбинации. Длина полинуклеотидов, если она указана, описывается как количество оснований и сокращается, например, н. (нуклеотиды), п.н. (пары нуклеотидов), т.п.н. (тысячи пар нуклеотидов) или млрд.п.н. (миллиарды пар нуклеотидов).

**[0062]** Используемый в настоящем документе термин «капля» относится к объему жидкости на капельном приводе. В некоторых случаях, например, капли могут являться водными или неводными или могут являться смесями или эмульсиями, включая в себя водные и неводные компоненты. В отношении неограничивающих примеров капельных жидкостей, которые могут быть подвергнуты операциям с каплями, см., например, международную патентную публикацию № WO2007/120241. Любую подходящую систему для формирования капель и управления ими можно использовать в вариантах осуществления представленных в настоящем документе. Например, в некоторых случаях используют капельный привод. В отношении неограничивающих примеров капельных приводов, которые можно использовать см., например, патент США № 6911132, 6977033, 6773566, 6565727, 7163612, 7052244, 7328979, 7547380, 7641779, публикации заявок на выдачу патента США №№ US20060194331, US20030205632, US20060164490, US20070023292, US20060039823, US20080124252, US20090283407, US20090192044, US20050179746, US20090321262, US20100096266, US20110048951, международную патентную публикацию № WO2007/120241. В некоторых случаях шарики предусмотрены в капле, в зазоре для операций с каплями или на поверхности для операций с каплями. В некоторых случаях шарики предусмотрены в резервуаре, который является внешним по отношению к зазору для операций с каплями или расположен отдельно от поверхности для

операций с каплями, и резервуар может быть связан с проточным каналом, который позволяет капле, включая в себя шарики, попадать в зазор для операций с каплями или в контакт с поверхностью для операций с каплями. Неограничивающие примеры техник капельных приводов для иммобилизации магниточувствительных шариков и/или немагнитно-чувствительных шариков и/или проведения протоколов операций с каплями с использованием шариков описаны в публикации на выдачу патента США № US20080053205, международных патентных публикациях WO2008/098236, WO2008/134153, WO2008/116221, WO2007/120241. Характеристики шариков можно использовать в вариантах осуществления мультиплексирования способов, описанных в настоящем документе. Примеры шариков, обладающих характеристиками, подходящими для мультиплексирования, а также способы обнаружения и анализа сигналов, излучаемых такими шариками, можно найти в публикации на выдачу патента США № US20080305481, US20080151240, US20070207513, US20070064990, US20060159962, US20050277197, US20050118574.

**[0063]** Используемый в настоящем документе термин «уникальный молекулярный идентификатор (UMI)» относится к уникальной последовательности нуклеиновой кислоты, которая прикреплена к каждой из множества молекул нуклеиновой кислоты. При включении в молекулу нуклеиновой кислоты UMI в некоторых случаях используется для корректировки смещения последующей амплификации путем прямого подсчета UMI, которые секвенируются после амплификации. Дизайн, встраивание и применение UMI описаны, например, в международной патентной публикации WO 2012/142213, Islam et al. Nat. Methods (2014) 11:163-166 и Kivioja, T. et al. Nat. Methods (2012) 9: 72-74.

**[0064]** Используемый в настоящем документе термин «штрих-код» относится к метке нуклеиновой кислоты, которую можно использовать для идентификации образца или источника материала нуклеиновой кислоты. Таким образом, если образцы нуклеиновой кислоты происходят из нескольких источников, нуклеиновые кислоты в каждом образце нуклеиновой кислоты в некоторых случаях метят разными метками нуклеиновой кислоты, чтобы можно было идентифицировать источник образца. Штрих-коды, также обычно называемые индексами, метками и т.п., хорошо известны специалистам в настоящей области техники. Можно использовать любой подходящий штрих-код или набор штрих-кодов. См., например, неограничивающие примеры, приведенные в патенте США № 8053192 и международной патентной публикации № WO2005/068656. Штрихкодирование одиночных клеток можно проводить, как описано, например, в публикации заявки на выдачу патента США № 2013/0274117.

**[0065]** Термины «твердая поверхность», «твердая подложка» и другие грамматические эквиваленты в настоящем документе относятся к любому материалу, который подходит или может быть модифицирован, чтобы быть подходящим для прикрепления праймеров, штрих-кодов и последовательностей, описанных в настоящем документе. Иллюстративные субстраты включают в себя без ограничения следующее: стекло и модифицированное или функционализированное стекло, пластмассы (включая в себя акрилы, полистирол и сополимеры стирола и других материалов, полипропилен, полиэтилен, полибутилен, полиуретаны, Teflon™ и т.д.), полисахариды, нейлон, нитроцеллюлоза, керамика, смолы, диоксид кремния, материалы на основе диоксида кремния (например, кремний или модифицированный кремний), углерод, металлы, неорганические стекла, пластмассы, пучки оптических волокон и множество других полимеров. Согласно некоторым вариантам осуществления твердая подложка содержит узорчатую поверхность, подходящую для иммобилизации праймеров, штрих-кодов и последовательностей в упорядоченной структуре.

**[0066]** Используемый в настоящем документе термин «биологический образец» включает в себя без ограничения ткани, клетки, биологические жидкости и их изоляты. Клетки или другие образцы, используемые в описанных в настоящем документе способах, в некоторых случаях выделяют из организма пациентов-людей, животных, растений, почвы или других образцов, содержащих микроорганизмы, такие как бактерии, грибы, простейшие и т.д. В некоторых случаях биологический образец получен от человека. В некоторых случаях биологический образец получен не от человека. В некоторых случаях клетки подвергают способам РТА, описанным в настоящем документе, и секвенированию. Варианты, обнаруженные по всему геному или в определенных местах, можно сравнить со всеми другими клетками, выделенными от этого субъекта, чтобы проследить историю клеточной родословной в исследовательских или диагностических целях.

**[0067]            Направленная первичной матрицей амплификация**

**[0068]** В настоящем документе описаны способы амплификации нуклеиновых кислот, такие как «направленная первичной матрицей амплификация (РТА)». Например, описанные в настоящем документе способы РТА схематически представлены на фиг. 1A-1D. С помощью способа РТА ампликоны предпочтительно создаются из первичной матрицы («прямые копии») с использованием полимеразы (например, полимеразы, замещающей цепь). Следовательно, ошибки передаются от дочерних ампликонов во время последующих амплификаций с меньшей скоростью по сравнению с MDA. Результатом является легко выполняемый способ, который, в отличие от существующих протоколов WGA, может точно и воспроизводимо усиливать низкий входной сигнал ДНК, включая в



себя геномы одиночных клеток, с высокой широтой охвата и однородностью. Более того, терминированные продукты амплификации можно подвергать направленному лигированию после удаления терминаторов, что позволяет прикрепить клеточный штрих-код к праймерам для амплификации, так что продукты из всех клеток можно объединить после прохождения параллельных реакций амплификации (фигура 1B).

**[0069]** В настоящем документе описаны способы, в которых для амплификации используют полимеразы нуклеиновых кислот с активностью замещения цепи. В некоторых случаях такие полимеразы обладают активностью замещения цепи и низкой частотой появления ошибок. В некоторых случаях такие полимеразы обладают активностью замещения цепи и корректирующей экзонуклеазной активностью, такой как 3'→5' корректирующая активность. В некоторых случаях полимеразы нуклеиновых кислот используют в сочетании с другими компонентами, такими как обратимые или необратимые терминаторы или дополнительные факторы замещения цепи. В некоторых случаях полимеразы обладают активностью замещения цепи, но не обладают корректирующей экзонуклеазной активностью. Например, в некоторых случаях такие полимеразы включают в себя полимеразу бактериофага phi29 (Ф29), которая также характеризуется очень низкой частотой появления ошибок в результате 3'→5' корректирующей экзонуклеазной активности (см., например, патенты США №№ 5198543 и 5001050). В некоторых случаях неограничивающие примеры замещающих цепи полимераз нуклеиновых кислот включают в себя, например, следующее: генетически модифицированная ДНК-полимераза phi29 (Ф29), фрагмент Кленова ДНК-полимеразы I (Jacobsen et al., *Eur. J. Biochem.* 45:623-627 (1974)), ДНК-полимераза фага M2 (Matsumoto et al., *Gene* 84:247 (1989)), ДНК-полимераза фага phiPRD1 (Jung et al., *Proc. Natl. Acad. Sci. USA* 84:8287 (1987); Zhu and Ito, *Biochim. Biophys. Acta.* 1219:267-276 (1994)), ДНК-полимераза Bst (например, ДНК-полимераза большого фрагмента Bst (Exo(-) Bst; Aliotta et al., *Genet. Anal. (Netherlands)* 12:185-195 (1996)), ДНК-полимераза экзо(-) Bca (Walker and Linn, *Clinical Chemistry* 42:1604-1608 (1996)), ДНК-полимераза Bsu, ДНК-полимераза Vent<sub>R</sub>, включая в себя ДНК-полимеразу Vent<sub>R</sub> (экзо-) (Kong et al., *J. Biol. Chem.* 268:1965-1975 (1993)), ДНК-полимераза Deep Vent, включая в себя ДНК-полимеразу Deep Vent (экзо-), ДНК-полимераза IsoPol, ДНК-полимераза I, ДНК-полимераза Terminator, ДНК-полимераза T5 (Chatterjee et al., *Gene* 97:13-19 (1991)), секвенза (U.S. Biochemicals), ДНК-полимераза T7, T7-секвенза, ДНК-полимераза T7 gp5, PRD1 ДНК-полимераза, ДНК-полимераза T4 (Kaboord and Benkovic, *Curr. Biol.* 5:149-157 (1995)). Дополнительные замещающие цепи полимеразы нуклеиновых кислот также совместимы со способами, описанными в настоящем документе. Способность данной полимеразы осуществлять репликацию с замещением цепи можно определить,

например, с использованием полимеразы в анализе репликации с замещением цепи (например, как описано в патенте США № 6977148). Такие анализы в некоторых случаях проводят при температуре, подходящей для оптимальной активности используемого фермента, например, 32°C для ДНК-полимеразы phi29, от 46°C до 64°C для ДНК-полимеразы экзо(-) Vst или от около 60°C до 70°C для фермента из гипертермофильного организма. Другой применимый анализ для выбора полимеразы представляет собой анализ с блокированием праймера, описанный в Kong et al., J. Biol. Chem. 268:1965-1975 (1993). Анализ состоит из анализа удлинения праймера с использованием матрицы оцДНК M13 в присутствии или при отсутствии олигонуклеотида, который гибридизуется выше удлиняющего праймера, чтобы заблокировать его продвижение. Другие ферменты, способные замещать блокирующий праймер в этом анализе, в некоторых случаях можно использовать для раскрытого способа. В некоторых случаях полимеразы включают dNTP и терминаторы приблизительно с одинаковой скоростью. В некоторых случаях отношение скоростей включения для dNTP и терминаторов для полимеразы, описанной в настоящем документе, составляют около 1:1, около 1,5:1, около 2:1, около 3:1, около 4:1, около 5:1, около 10:1, около 20:1, около 50:1, около 100:1, около 200:1, около 500:1 или около 1000:1. В некоторых случаях отношение скоростей включения для dNTP и терминаторов для полимеразы, описанной в настоящем документе, составляют 1:1 - 1000:1, 2:1 - 500:1, 5:1 - 100:1, 10:1 - 1000:1, 100:1 - 1000:1, 500:1 - 2000:1, 50:1 - 1500:1 или 25:1 - 1000:1.

**[0070]** В настоящем документе описаны способы амплификации, при которых замещение цепей можно облегчить с помощью фактора замещения цепи, такого как, например, хеликаза. Такие факторы в некоторых случаях используют в сочетании с дополнительными компонентами амплификации, такими как полимеразы, терминаторы или другой компонент. В некоторых случаях фактор замещения цепи используют с полимеразой, не обладающей активностью замещения цепи. В некоторых случаях фактор замещения цепи используют с полимеразой, обладающей активностью замещения цепи. Без ограничения какой-либо теорией факторы замещения цепи могут увеличивать скорость повторного праймирования более мелких двухцепочечных ампликонов. В некоторых случаях любая ДНК-полимераза, которая может выполнять репликацию с замещением цепи в присутствии фактора замещения цепи, подходит для использования в способе РТА, даже если ДНК-полимераза не выполняет репликацию с замещением цепи при отсутствии такого фактора. Факторы замещения цепи, применимый в репликации с замещением цепи, в некоторых случаях включают в себя (без ограничения) следующее: дополнительная субъединица полимеразы BMRF1 (Tsurumi et al., J. Virology 67(12):7648-7653 (1993)), белок, связывающий ДНК аденовируса (Zijderveld and van der Vliet, J. Virology 68(2): 1158-1164

(1994)), вирусный белок простого герпеса ICP8 (Boehmer and Lehman, J. Virology 67(2):711-715 (1993); Skaliter and Lehman, Proc. Natl. Acad. Sci. USA 91(22):10665-10669 (1994)); белки, связывающие одноцепочечную ДНК (SSB; Rigler and Romano, J. Biol. Chem. 270:8910-8919 (1995)); белок гена 32 фага T4 (Villemain and Giedroc, Biochemistry 35:14395-14404 (1996)); хеликаза-праймаза T7; белок SSB gp2.5 T7; Tte-UvrD (от *Thermoanaerobacter tengcongensis*), хеликаза тимуса телят (Siegel et al., J. Biol. Chem. 267:13629-13635 (1992)); бактериальный SSB (например, SSB *E. coli*), репликационный белок A (RPA) у эукариот, митохондриальный SSB человека (mtSSB) и рекомбиназы (например, белки семейства рекомбиназа A (RecA), T4 UvsX, Sak4 фага HK620, Rad51, Dmc1 или Radb). Комбинации факторов, которые облегчают замещение цепей и праймирование, также являются сопоставимыми со способами, описанными в настоящем документе. Например, хеликазу используют вместе с полимеразой. В некоторых случаях способ РТА включает применение белка, связывающего одноцепочечную ДНК (SSB, gp32 T4 или другого связывающего одноцепочечную ДНК белка), хеликазы и полимеразы (например, SauДНК-полимераза, Bsu-полимераза, Bst2.0, GspM, GspM2.0, GspSSD или другая подходящая полимеразы). В некоторых случаях обратные транскриптазы используют в сочетании с факторами замещения цепи, описанными в настоящем документе.

**[0071]** В настоящем документе описаны способы амплификации, включающие применение терминаторных нуклеотидов, полимераз и дополнительных факторов или условий. Например, такие факторы используют в некоторых случаях для фрагментации матрицы(матриц) нуклеиновой кислоты или ампликонов во время амплификации. В некоторых случаях такие факторы включают в себя эндонуклеазы. В некоторых случаях факторы включают в себя транспозазы. В некоторых случаях для фрагментации нуклеиновых кислот во время амплификации используют механический сдвиг. В некоторых случаях во время амплификации добавляют нуклеотиды, которые можно фрагментировать путем добавления дополнительных белков или условий. Например, в ампликоны включают урацил; обработка урацил-D-гликозилазой фрагментирует нуклеиновые кислоты в урацил-содержащих положениях. В некоторых случаях также используют дополнительные системы для селективной фрагментации нуклеиновых кислот, например, сконструированную ДНК-гликозилазу, которая расщепляет модифицированные пары оснований цитозин-пирен. (Kwon, et al. Chem Biol. 2003, 10(4), 351).

**[0072]** В настоящем документе описаны способы амплификации, включающие применение терминаторных нуклеотидов, которые прекращают репликацию нуклеиновой кислоты, таким образом, уменьшая размер продуктов амплификации. Такие

терминаторы в некоторых случаях используют в сочетании с полимеразам, факторами замещения цепи или другими компонентами амплификации, описанными в настоящем документе. В некоторых случаях терминаторные нуклеотиды снижают или уменьшают эффективность репликации нуклеиновой кислоты. Такие терминаторы в некоторых случаях снижают скорости удлинения по меньшей мере на 99,9%, 99%, 98%, 95%, 90%, 85%, 80%, 75%, 70% или по меньшей мере на 65%. Такие терминаторы в некоторых случаях снижают скорости удлинения на 50%-90%, 60%-80%, 65%-90%, 70%-85%, 60%-90%, 70%-99%, 80%-99% или 50%-80%. В некоторых случаях терминаторы снижают среднюю длину продукта ампликона по меньшей мере на 99,9%, 99%, 98%, 95%, 90%, 85%, 80%, 75%, 70% или по меньшей мере на 65%. Терминаторы в некоторых случаях снижают среднюю длину ампликона на 50%-90%, 60%-80%, 65%-90%, 70%-85%, 60%-90%, 70%-99%, 80%-99% или 50%-80%. В некоторых случаях ампликоны, содержащие терминаторные нуклеотиды, образуют петли или шпильки, которые снижают способность полимеразы использовать такие ампликоны в качестве матриц. Использование терминаторов в некоторых случаях снижает скорость амплификации в начальных сайтах амплификации за счет включения терминаторных нуклеотидов (например, дидезоксинуклеотидов, которые были модифицированы, чтобы сделать их устойчивыми к экзонуклеазам для прекращения удлинения ДНК), что приводит к меньшим по размеру продуктам амплификации. Путем получения продуктов амплификации меньшего размера, чем в используемых в настоящее время способах (например, средняя длина, составляющая 50-2000 нуклеотидов для способов РТА по сравнению со средней длиной продукта, составляющей > 10 000 нуклеотидов для способов МДА), продукты амплификации РТА в некоторых случаях подвергаются прямому лигированию адапторов без необходимости фрагментации, что позволяет эффективно встраивать штрих-коды клеток и уникальные молекулярные идентификаторы (UMI) (см. фигуры 1D, 2B-3E, 9, 10A и 10B).

**[0073]** Терминаторные нуклеотиды присутствуют в различных концентрациях в зависимости от таких факторов, как полимеразы, матрица или другие факторы. Например, количество терминаторных нуклеотидов в некоторых случаях выражают как отношение нетерминаторных нуклеотидов к терминаторным нуклеотидам в способе, описанном в настоящем документе. Такие концентрации в некоторых случаях позволяют контролировать длины ампликонов. В некоторых случаях отношение нетерминаторных нуклеотидов к терминаторным нуклеотидам составляет около 2:1, 5:1, 7:1, 10:1, 20:1, 50:1, 100:1, 200:1, 500:1, 1000:1, 2000:1 или 5000:1. В некоторых случаях отношение нетерминаторных нуклеотидов к терминаторным нуклеотидам составляет 2:1-10:1, 5:1-20:1, 10:1-100:1, 20:1-200:1, 50:1-1000:1, 50:1-500:1, 75:1-150:1 или 100:1-500:1. В

некоторых случаях по меньшей мере один из нуклеотидов, присутствующий в течение амплификации с использованием способа, описанного в настоящем документе, представляет собой терминаторный нуклеотид. Необязательно, чтобы каждый терминатор присутствовал в приблизительно одинаковой концентрации; в некоторых случаях отношения каждого терминатора, присутствующего в способе, описанном в настоящем документе, оптимизируют для конкретного набора условий реакции, типа образца или полимеразы. Без ограничения какой-либо теорией каждый терминатор может обладать разной эффективностью для включения в растущую полинуклеотидную цепь ампликона в ответ на спаривание с соответствующим нуклеотидом на матричной цепи. Например, в некоторых случаях терминатор, спаривающийся с цитозином, присутствует в концентрации на около 3%, 5%, 10%, 15%, 20%, 25% или 50% выше чем средняя концентрация терминатора. В некоторых случаях терминатор, спаривающийся с тиминном, присутствует в концентрации на около 3%, 5%, 10%, 15%, 20%, 25%, или 50% выше чем средняя концентрация терминатора. В некоторых случаях терминатор, спаривающийся с гуанином, присутствует в концентрации на около 3%, 5%, 10%, 15%, 20%, 25% или 50% выше чем средняя концентрация терминатора. В некоторых случаях терминатор, спаривающийся с аденином, присутствует в концентрации на около 3%, 5%, 10%, 15%, 20%, 25% или 50% выше чем средняя концентрация терминатора. В некоторых случаях терминатор, спаривающийся с урацилом, присутствует в концентрации на около 3%, 5%, 10%, 15%, 20%, 25% или 50% выше чем средняя концентрация терминатора. Любой нуклеотид, способный прекращать удлинение нуклеиновой кислоты полимеразой нуклеиновой кислоты, в некоторых случаях используют в качестве терминаторного нуклеотида в способах, описанных в настоящем документе. В некоторых случаях для прекращения репликации нуклеиновой кислоты используют обратимый терминатор. В некоторых случаях необратимый терминатор используют для прекращения репликации нуклеиновой кислоты. В некоторых случаях неограниченные примеры терминаторов включают в себя следующее: обратимые и необратимые нуклеиновые кислоты и аналоги нуклеиновых кислот, такие как, например, 3' блокированный обратимый терминатор, содержащий нуклеотиды, 3' неблокированный обратимый терминатор, содержащий нуклеотиды, терминаторы, содержащие 2' модификации дезоксинуклеотидов, терминаторы, содержащие модификации азотистого основания дезоксинуклеотидов, или любая их комбинация. Согласно одному варианту осуществления терминаторные нуклеотиды представляют собой дидезоксинуклеотиды. Другие нуклеотид модификации которые прекращают репликацию нуклеиновой кислоты и могут являться подходящими для осуществления на практике настоящего изобретения, включают в себя без ограничения

следующее: любые модификации группы г 3' углерода дезоксирибозы, такие как инвертированные дидезоксинуклеотиды, 3' биотинилированные нуклеотиды, 3' аминонуклеотиды, 3'-фосфорилированные нуклеотиды, 3'-О-метилнуклеотиды, 3' углеродные спейсерные нуклеотиды, включая в себя 3' C3 спейсерные нуклеотиды, 3' C18 нуклеотиды, 3' гександиоловые спейсерные нуклеотиды, ациклонуклеотиды и их комбинации. В некоторых случаях терминаторы представляют собой полинуклеотиды, содержащие 1, 2, 3, 4 или более оснований в длину. В некоторых случаях терминаторы представляют собой полинуклеотиды, содержащие 1, 2, 3, 4 или более оснований в длину. В некоторых случаях терминаторы не содержат обнаруживаемый фрагмент или метку (например, массовую метку, флуоресцентную метку, краситель, радиоактивный атом или другой обнаруживаемый фрагмент). В некоторых случаях терминаторы не содержат химический фрагмент, обеспечивающий возможность прикрепления обнаруживаемого фрагмента или метки (например, «клик» азид/алкин, партнер для присоединения конъюгата или другой химический элемент для прикрепления метки). В некоторых случаях все терминаторные нуклеотиды содержат одну и ту же модификацию, которая снижает амплификацию до области (например, сахарный фрагмент, основной фрагмент или фосфатный фрагмент) нуклеотида. В некоторых случаях по меньшей мере один терминатор характеризуется другой модификацией, которая снижает амплификацию. В некоторых случаях все терминаторы характеризуются по существу одинаковыми длинами волн флуоресцентного возбуждения или излучения. В некоторых случаях терминаторы без модификации фосфатной группы используют с полимеразми, которые не обладают корректирующей экзонуклеазной активностью. Терминаторы, при использовании с полимеразми, которые обладают экзонуклеазной активностью 3'→5' корректирующей экзонуклеазной активностью (такая как, например, phi29), которая может удалить терминаторный нуклеотид, в некоторых случаях дополнительно модифицируют, чтобы сделать их устойчивыми к экзонуклеазам. Например, дидезоксинуклеотиды модифицируют альфа-тиогруппой, которая создает фосфоротиоатную связь, которая делает эти нуклеотиды устойчивыми к 3'→5' корректирующей экзонуклеазной активности 3'→5' полимераз нуклеиновых кислот. Такие модификации в некоторых случаях снижают корректирующую экзонуклеазную активность по меньшей мере на 99,5%, 99%, 98%, 95%, 90% или по меньшей мере на 85%. Неограничивающие примеры других модификаций терминаторных нуклеотидов, обеспечивающих устойчивость к 3'→5' экзонуклеазной активности, включают в себя в некоторых случаях: нуклеотиды с модификацией альфа-группы, такие как альфа-тио-дидезоксинуклеотиды, создающие фосфоротиоатную связь, C3 спейсерные нуклеотиды, запертые нуклеиновые кислоты (LNA), инвертированные

нуклеиновые кислоты, 2' фторооснования, 3' фосфорилирование, 2'-О-метильные модификации (или другие 2'-О-алкильные модификации), модифицированные пропином основания (например, дезоксицитозин, дезоксиуридин), нуклеотиды L-ДНК, нуклеотиды L-РНК, нуклеотиды с инвертированными связями (например, 5'-5' или 3'-3'), 5' инвертированные основания (например, 5' инвертированные 2',3'-дидезокси dТ), метилфосфонатные скелеты и транснуклеиновые кислоты. В некоторых случаях нуклеотиды с модификацией включают в себя нуклеиновые кислоты с модифицированными основаниями, содержащие свободные 3' ОН-группы (например, 2-нитробензилалкилированные трифосфаты НОMedU, основания, содержащие модификацию с большими химическими группами, такими как твердые подложки или другой большой фрагмент). В некоторых случаях полимеразу с активностью замещения цепи, но без 3'->5' корректирующей экзонуклеазной активности используют с терминаторными нуклеотидами с модификациями или без них, чтобы сделать их устойчивыми к экзонуклеазам. Такие полимеразы нуклеиновых кислот включают в себя без ограничения следующее: ДНК-полимераза Bst, ДНК-полимераза Bsu, ДНК-полимераза Deep Vent (экзо-), ДНК-полимераза фрагмента Кленова (экзо-), ДНК-полимераза Terminator и Vent<sub>R</sub> (экзо-).

**[0074] Праймеры и библиотеки ампликонов**

**[0075]** В настоящем документе описаны библиотеки ампликонов, полученные в результате амплификации по меньшей мере одной целевой молекулы нуклеиновой кислоты. Такие библиотеки в некоторых случаях создают с использованием способов, описанных в настоящем документе, таких как способы с использованием терминаторов. Такие способы включают применение полимераз или факторов замещения цепи, терминаторных нуклеотидов (обратимых или необратимых) или других признаков и вариантов осуществления, описанных в настоящем документе. В некоторых случаях библиотеки ампликонов, созданные путем применения терминаторов, описанных в настоящем документе, дополнительно амплифицируют в последующей реакции амплификации (например, ПЦР). В некоторых случаях последующие реакции амплификации не содержат терминаторы. В некоторых случаях библиотеки ампликонов содержат полинуклеотиды, где по меньшей мере 50%, 60%, 70%, 80%, 90%, 95% или по меньшей мере 98% полинуклеотидов содержат по меньшей мере один терминаторный нуклеотид. В некоторых случаях библиотека ампликонов содержит целевую молекулу нуклеиновой кислоты, из которой была получена библиотека ампликонов. Библиотека ампликонов содержит множество полинуклеотидов, где по меньшей мере некоторые из полинуклеотидов представляют собой прямые копии (например, реплицируются

непосредственно из целевой молекулы нуклеиновой кислоты, такой как геномная ДНК, РНК или другая целевая нуклеиновая кислота). Например, по меньшей мере 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% или больше чем 95% полинуклеотидов ампликона представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты. В некоторых случаях по меньшей мере 5% полинуклеотидов ампликона представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты. В некоторых случаях по меньшей мере 10% полинуклеотидов ампликона представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты. В некоторых случаях по меньшей мере 15% полинуклеотидов ампликона представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты. В некоторых случаях по меньшей мере 20% полинуклеотидов ампликона представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты. В некоторых случаях по меньшей мере 50% полинуклеотидов ампликона представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты. В некоторых случаях 3%-5%, 3-10%, 5%-10%, 10%-20%, 20%-30%, 30%-40%, 5%-30%, 10%-50% или 15%-75% полинуклеотидов ампликона представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты. В некоторых случаях по меньшей мере некоторые из полинуклеотидов представляют собой прямые копии целевой молекулы нуклеиновой кислоты или дочернего (первая копия целевой нуклеиновой кислоты) потомства. Например, по меньшей мере 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% или больше чем 95% полинуклеотидов ампликона представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты или дочернего потомства. В некоторых случаях по меньшей мере 5% полинуклеотидов ампликона представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты или дочернего потомства. В некоторых случаях по меньшей мере 10% полинуклеотидов ампликона представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты или дочернего потомства. В некоторых случаях по меньшей мере 20% полинуклеотидов ампликона представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты или дочернего потомства. В некоторых случаях по меньшей мере 30% полинуклеотидов ампликона представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты или дочернего потомства. В некоторых случаях 3%-5%, 3%-10%, 5%-10%, 10%-20%, 20%-30%, 30%-40%, 5%-30%, 10%-50% или 15%-75% полинуклеотидов ампликона представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты или дочернего потомства. В некоторых случаях прямые копии



целевой нуклеиновой кислоты имеют длину 50-2500, 75-2000, 50-2000, 25-1000, 50-1000, 500-2000 или 50-2000 оснований. В некоторых случаях дочернее потомство имеет длину 1000-5000, 2000-5000, 1000-10000, 2000-5000, 1500-5000, 3000-7000 или 2000-7000 оснований. В некоторых случаях средняя длина продуктов амплификации ПГА составляет 25-3000 нуклеотидов в длину, 50-2500, 75-2000, 50-2000, 25-1000, 50-1000, 500-2000 или 50-2000 оснований в длину. В некоторых случаях ампликоны, созданные с помощью ПГА, составляют не более 5000, 4000, 3000, 2000, 1700, 1500, 1200, 1000, 700, 500 или не более 300 оснований в длину. В некоторых случаях ампликоны, созданные с помощью ПГА, составляют 1000-5000, 1000-3000, 200-2000, 200-4000, 500-2000, 750-2500 или 1000-2000 оснований в длину. Библиотеки ампликонов, созданные с использованием способов, описанных в настоящем документе, в некоторых случаях содержат по меньшей мере 1000, 2000, 5000, 10000, 100000, 200000, 500000 или больше чем 500000 ампликонов, содержащих уникальные последовательности. В некоторых случаях библиотека содержит по меньшей мере 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 2000, 2500, 3000 или по меньшей мере 3500 ампликонов. В некоторых случаях по меньшей мере 5%, 10%, 15%, 20%, 25%, 30% или больше чем 30% полинуклеотидов ампликона с длиной, составляющей меньше чем 1000 оснований, представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты. В некоторых случаях по меньшей мере 5%, 10%, 15%, 20%, 25%, 30% или больше чем 30% полинуклеотидов ампликона с длиной, составляющей не более 2000 оснований, представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты. В некоторых случаях по меньшей мере 5%, 10%, 15%, 20%, 25%, 30% или больше чем 30% полинуклеотидов ампликона с длиной, составляющей 3000-5000 оснований, представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты. В некоторых случаях отношение ампликонов - прямых копий к целевым молекулам нуклеиновой кислоты составляет по меньшей мере 10:1, 100:1, 1000:1, 10000:1, 100000:1, 1000000:1, 10000000:1 или больше чем 10000000:1. В некоторых случаях отношение ампликонов - прямых копий к целевым молекулам нуклеиновой кислоты составляет по меньшей мере 10:1, 100:1, 1000:1, 10000:1, 100000:1, 1000000:1, 10000000:1 или больше чем 10000000:1, где ампликоны - прямые копии составляют не более 700-1200 оснований в длину. В некоторых случаях отношение ампликонов - прямых копий и дочерних ампликонов к целевым молекулам нуклеиновой кислоты составляет по меньшей мере 10:1, 100:1, 1000:1, 10000:1, 100000:1, 1000000:1, 10000000:1 или больше чем 10000000:1. В некоторых случаях отношение ампликонов - прямых копий и дочерних ампликонов к целевым молекулам нуклеиновой кислоты составляет по меньшей мере 10:1, 100:1, 1000:1, 10000:1, 100000:1,

1000000:1, 10000000:1 или больше чем 10000000:1, где ампликоны - прямые копии составляют 700-1200 оснований в длину, и дочерние ампликоны составляют 2500-6000 оснований в длину. В некоторых случаях библиотека содержит около 50-10000, около 50-5000, около 50-2500, около 50-1000, около 150-2000, около 250-3000, около 50-2000, около 500-2000 или около 500-1500 ампликонов, которые представляют собой прямые копии целевой молекулы нуклеиновой кислоты. В некоторых случаях библиотека содержит около 50-10000, около 50-5000, около 50-2500, около 50-1000, около 150-2000, около 250-3000, около 50-2000, около 500-2000 или около 500-1500 ампликонов, которые представляют собой прямые копии целевой молекулы нуклеиновой кислоты или дочерних ампликонов. Библиотеки ампликонов, созданные с использованием способов, описанных в настоящем документе, в некоторых случаях подвергают дополнительным стадиям, таким как лигирование адаптора и дальнейшая ПЦР-амплификация. В некоторых случаях такие дополнительные стадии предшествуют стадии секвенирования.

**[0076]** Библиотеки ампликонов полинуклеотидов, созданные с помощью способов РТА и композиций (терминаторы, полимеразы и т.д.), описанных в настоящем документе, в некоторых случаях характеризуются увеличенной однородностью. Однородность в некоторых случаях описывают с использованием кривой Лоренца (например, фигура 5С) или другого такого способа. Такие увеличения в некоторых случаях приводят к пониженному количеству прочтений секвенирования, необходимых для требуемого охвата целевой молекулы нуклеиновой кислоты (например, геномной ДНК, РНК или другой целевой молекулы нуклеиновой кислоты). Например, не более 50% кумулятивной доли полинуклеотидов содержат последовательности, составляющие по меньшей мере 80% кумулятивной доли последовательностей целевой молекулы нуклеиновой кислоты. В некоторых случаях не более 50% кумулятивной доли полинуклеотидов содержат последовательности, составляющие по меньшей мере 60% кумулятивной доли последовательностей целевой молекулы нуклеиновой кислоты. В некоторых случаях не более 50% кумулятивной доли полинуклеотидов содержат последовательности, составляющие по меньшей мере 70% кумулятивной доли последовательностей целевой молекулы нуклеиновой кислоты. В некоторых случаях не более 50% кумулятивной доли полинуклеотидов содержат последовательности, составляющие по меньшей мере 90% кумулятивной доли последовательностей целевой молекулы нуклеиновой кислоты. В некоторых случаях однородность описывают с использованием коэффициента Джинни (где коэффициент, составляющий 0, представляет идеальную равномерность библиотеки, а коэффициент, составляющий 1, представляет идеальную неравномерность). В некоторых случаях библиотеки ампликонов, описанные в

настоящем документе, характеризуются коэффициентом Джинни, составляющим не более 0,55, 0,50, 0,45, 0,40 или 0,30. В некоторых случаях библиотеки ампликонов, описанные в настоящем документе, характеризуются коэффициентом Джинни, составляющим не более 0,50. В некоторых случаях библиотеки ампликонов, описанные в настоящем документе, характеризуются коэффициентом Джинни, составляющим не более 0,40. Такие количественные показатели однородности в некоторых случаях зависят от количества полученных прочтений. Например, получают не более 100 млн., 200 млн., 300 млн., 400 млн. или не более 500 млн. прочтений. В некоторых случаях длина прочтения составляет около 50,75, 100, 125, 150, 175, 200, 225 или около 250 оснований в длину. В некоторых случаях количественные показатели однородности зависят от глубины охвата целевой нуклеиновой кислоты. Например, средняя глубина охвата составляет около 10X, 15X, 20X, 25X или около 30X. В некоторых случаях средняя глубина охвата составляет 10-30X, 20-50X, 5-40X, 20-60X, 5-20X или 10-20X. В некоторых случаях библиотеки ампликонов, описанные в настоящем документе, характеризуются коэффициентом Джинни, составляющим не более 0,55, где получили около 300 млн. прочтений. В некоторых случаях библиотеки ампликонов, описанные в настоящем документе, характеризуются коэффициентом Джинни, составляющим не более 0,50, где получили около 300 млн. прочтений. В некоторых случаях библиотеки ампликонов, описанные в настоящем документе, характеризуются коэффициентом Джинни, составляющим не более 0,45, где получили около 300 млн. прочтений. В некоторых случаях библиотеки ампликонов, описанные в настоящем документе, характеризуются коэффициентом Джинни, составляющим не более 0,55, где получили не более 300 млн. прочтений. В некоторых случаях библиотеки ампликонов, описанные в настоящем документе, характеризуются коэффициентом Джинни, составляющим не более 0,50, где получили не более 300 млн. прочтений. В некоторых случаях библиотеки ампликонов, описанные в настоящем документе, характеризуются коэффициентом Джинни, составляющим не более 0,45, где получили не более 300 млн. прочтений. В некоторых случаях библиотеки ампликонов, описанные в настоящем документе, характеризуются коэффициентом Джинни, составляющим не более 0,55, где средняя глубина охвата секвенирования составляет около 15X. В некоторых случаях библиотеки ампликонов, описанные в настоящем документе, характеризуются коэффициентом Джинни, составляющим не более 0,50, где средняя глубина охвата секвенирования составляет около 15X. В некоторых случаях библиотеки ампликонов, описанные в настоящем документе, характеризуются коэффициентом Джинни, составляющим не более 0,45, где средняя глубина охвата секвенирования составляет около 15X. В некоторых случаях библиотеки ампликонов, описанные в

настоящем документе, характеризуются коэффициентом Джинни, составляющим не более 0,55, где средняя глубина охвата секвенирования составляет по меньшей мере 15X. В некоторых случаях библиотеки ампликонов, описанные в настоящем документе, характеризуются коэффициентом Джинни, составляющим не более 0,50, где средняя глубина охвата секвенирования составляет по меньшей мере 15X. В некоторых случаях библиотеки ампликонов, описанные в настоящем документе, характеризуются коэффициентом Джинни, составляющим не более 0,45, где средняя глубина охвата секвенирования составляет по меньшей мере 15X. В некоторых случаях библиотеки ампликонов, описанные в настоящем документе, характеризуются коэффициентом Джинни, составляющим не более 0,55, где средняя глубина охвата секвенирования составляет не более 15X. В некоторых случаях библиотеки ампликонов, описанные в настоящем документе, характеризуются коэффициентом Джинни, составляющим не более 0,50, где средняя глубина охвата секвенирования составляет не более 15X. В некоторых случаях библиотеки ампликонов, описанные в настоящем документе, характеризуются коэффициентом Джинни, составляющим не более 0,45, где средняя глубина охвата секвенирования составляет не более 15X. Однородные библиотеки ампликонов, созданные с использованием способов, описанных в настоящем документе, в некоторых случаях подвергают дополнительным стадиям, таким как лигирование адаптора и дальнейшая ПЦР-амплификация. В некоторых случаях такие дополнительные стадии предшествуют стадии секвенирования.

**[0077]** Праймеры содержат нуклеиновые кислоты, используемые для праймирования описанных в настоящем документе реакций амплификации. Такие праймеры в некоторых случаях включают в себя без ограничения случайные дезоксинуклеотиды любой длины с модификациями или без них, чтобы сделать их устойчивыми к экзонуклеазам, случайные рибонуклеотиды любой длины с модификациями или без них, чтобы сделать их устойчивыми к экзонуклеазам, модифицированные нуклеиновые кислоты, такие как запертые нуклеиновые кислоты, ДНК- или РНК-праймеры, которые нацелены на конкретную область генома, и реакции, которые праймированы ферментами, такими как праймаза. В случае полногеномной РТА предпочтительно использовать набор праймеров, характеризующихся случайными или частично случайными нуклеотидными последовательностями. В образце нуклеиновой кислоты значительной сложности не обязательно знать конкретные последовательности нуклеиновой кислоты, присутствующие в образце, и нет необходимости конструировать праймеры так, чтобы они являлись комплементарными какой-либо конкретной последовательности. Скорее, сложность образца нуклеиновой кислоты приводит к

большому количеству различных целевых последовательностей гибридизации в образце, которые будут комплементарны различным праймерам случайной или частично случайной последовательности. Комплементарная часть праймеров для использования в РТА в некоторых случаях полностью рандомизирована, содержит только часть, которая является рандомизированной, или иным образом выборочно рандомизирована. Количество случайных положений оснований в комплементарной части праймеров в некоторых случаях, например, составляет от 20% до 100% от общего количества нуклеотидов в комплементарной части праймеров. В некоторых случаях количество случайных положений оснований в комплементарной части праймеров составляет от 10% до 90%, 15-95%, 20% -100%, 30% -100%, 50% -100%, 75-100% или 90-95% от общего количества нуклеотидов в комплементарной части праймеров. В некоторых случаях количество случайных положений оснований в комплементарной части праймеров составляет по меньшей мере 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% или не менее 90% от общего количества нуклеотидов в комплементарной части праймеров. Наборы праймеров, характеризующихся случайными или частично случайными последовательностями, в некоторых случаях синтезируют с использованием стандартных методик, позволяющих рандомизировать добавление любого нуклеотида в каждом положении. В некоторых случаях наборы праймеров состоят из праймеров одинаковой длины и/или гибридизационных характеристик. В некоторых случаях термин «случайный праймер» относится к праймеру, который может проявлять четырехкратную вырожденность в каждом положении. В некоторых случаях термин «случайный праймер» относится к праймеру, который может проявлять трехкратную вырожденность в каждом положении. Случайные праймеры, используемые в описанных в настоящем документе способах, в некоторых случаях содержат случайную последовательность, которая составляет 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 или более оснований в длину. В некоторых случаях праймеры содержат случайные последовательности длиной 3-20, 5-15, 5-20, 6-12 или 4-10 оснований. Праймеры могут также содержать не удлиняемые элементы, которые ограничивают последующую амплификацию созданных ими ампликонов. Например, праймеры с не удлиняемыми элементами в некоторых случаях содержат терминаторы. В некоторых случаях праймеры содержат терминаторные нуклеотиды, например, 1, 2, 3, 4, 5, 10 или более 10 терминаторных нуклеотидов. Праймеры не должны ограничиваться компонентами, которые добавляются извне в реакцию амплификации. В некоторых случаях праймеры создаются *in situ* путем добавления нуклеотидов и белков, которые способствуют праймированию. Например, ферменты, подобные праймазе, в комбинации с нуклеотидами в некоторых случаях используют для создания случайных праймеров для описанных в

настоящем документе способов. Ферменты, подобные праймазе, в некоторых случаях являются представителями суперсемейства ферментов DnaG или AEP. В некоторых случаях фермент, подобный праймазе, представляет собой TthPrimPol. В некоторых случаях праймазоподобным ферментом является хеликаза-праймаза T7 gp4. Такие праймазы в некоторых случаях используют с полимеразами или факторами замещения цепи, описанными в настоящем документе. В некоторых случаях праймазы инициируют праймирование дезоксирибонуклеотидами. В некоторых случаях праймазы инициируют праймирование рибонуклеотидами.

**[0078]** За амплификацией РТА может следовать отбор в отношении определенного подмножества ампликонов. Такой отбор в некоторых случаях зависит от размера, аффинности, активности, гибридизации с зондами или другого известного фактора отбора в настоящей области техники. В некоторых случаях отбор предшествует или следует за дополнительными стадиями, описанными в настоящем документе, такими как лигирование адаптора и/или амплификация библиотеки. В некоторых случаях отбор основан на размере (длине) ампликонов. В некоторых случаях выбирают меньшие ампликоны, которые с меньшей вероятностью подверглись экспоненциальной амплификации, которая обогащает продуктами, полученными из первичной матрицы, при дальнейшем преобразовании процесса амплификации из экспоненциальной в квазилинейный процесс амплификации (фиг. 1А). В некоторых случаях выбирают ампликоны, содержащие 50-2000, 25-5000, 40-3000, 50-1000, 200-1000, 300-1000, 400-1000, 400-600, 600-2000 или 800-1000 оснований в длину. Выбор размера в некоторых случаях происходит с использованием протоколов, например, с использованием твердофазной обратимой иммобилизации (SPRI) на карбоксилированных парамагнитных шариках для обогащения фрагментами нуклеиновых кислот определенных размеров или другого протокола, известного специалистам в настоящей области техники. Необязательно или в комбинации, отбор происходит посредством предпочтительной амплификации более мелких фрагментов во время ПЦР при подготовке библиотек для секвенирования, а также в результате предпочтительного образования кластеров из более мелких фрагментов библиотеки для секвенирования во время секвенирования Illumina. Другие стратегии отбора более мелких фрагментов также согласуются с описанными в настоящем документе способами и включают в себя без ограничения выделение фрагментов нуклеиновой кислоты определенного размера после гель-электрофореза, использование колонок с диоксидом кремния, которые связывают фрагменты нуклеиновой кислоты определенного размера, и использование других стратегий ПЦР, которые сильнее обогащают в отношении более мелких фрагментов.

**[0079]** Некомплементарная часть праймера, используемого в РГА, может включать в себя последовательности, которые можно использовать для дальнейших манипуляций и/или анализа амплифицированных последовательностей. Примером такой последовательности является «детекторная метка». Детекторные метки содержат последовательности, комплементарные детекторным зондам, и выявляются с помощью родственных детекторных зондов. На праймере может присутствовать одна, две, три, четыре или более четырех детекторных меток. Не существует фундаментальных ограничений на количество детекторных меток, которые могут присутствовать на праймере, за исключением размера праймера. В некоторых случаях на праймере присутствует одна детекторная метка. В некоторых случаях на праймере присутствуют две детекторные метки. Когда присутствуют несколько детекторных меток, они могут характеризоваться одинаковой последовательностью или могут характеризоваться разными последовательностями, причем каждая отличающаяся последовательность комплементарна другому детекторному зонду. В некоторых случаях несколько детекторных меток характеризуются одинаковой последовательностью. В некоторых случаях несколько детекторных меток характеризуются разной последовательностью.

**[0080]** Другим примером последовательности, которая может быть включена в некомплементарную часть праймера, является «адресная метка». Адресная метка характеризуется последовательностью, комплементарной адресному зонду. Адресные метки включаются на концах амплифицированных цепей. Если они присутствуют, то на праймере может находиться одна или более адресных меток. Нет никаких фундаментальных ограничений на количество адресных меток, которые могут присутствовать в праймере, за исключением размера праймера. Когда присутствуют несколько адресных меток, они могут характеризоваться одинаковой последовательностью или могут характеризоваться разными последовательностями, причем каждая отличающаяся последовательность комплементарна другому адресному зонду. Часть адресной метки может характеризоваться любой длиной, которая поддерживает конкретную и стабильную гибридизацию между адресной меткой и адресным зондом. В некоторых случаях нуклеиновые кислоты из более чем одного источника могут включать в себя переменную последовательность метки. Эта последовательность метки может характеризоваться длиной до 100 нуклеотидов, предпочтительно от 1 до 10 нуклеотидов в длину, наиболее предпочтительно от 4, 5 или 6 нуклеотидов в длину и содержит комбинации нуклеотидов. В некоторых случаях последовательность метки составляет 1-20, 2-15, 3-13, 4-12, 5-12 или 1-10 нуклеотидов в длину. Например, если для формирования метки выбирают шесть пар оснований и используют перестановку четырех разных

нуклеотидов, то можно создать в общей сложности 4096 якорей нуклеиновых кислот (например, шпильки), каждый с уникальной меткой из 6-ти оснований.

**[0081]** Описанные в настоящем документе праймеры могут присутствовать в растворе или иммобилизованы на твердой подложке. В некоторых случаях праймеры, несущие штрих-коды образца и/или последовательности УМІ, могут быть иммобилизованы на твердой подложке. Твердая подложка может представлять собой, например, один или более шариков. В некоторых случаях отдельные клетки контактируют с одним или более шариками, содержащими уникальный набор штрих-кодов образцов и/или последовательностей УМІ, чтобы идентифицировать индивидуальную клетку. В некоторых случаях лизаты из отдельных клеток контактируют с одним или более шариками, содержащими уникальный набор штрих-кодов образца и/или последовательностей УМІ, чтобы идентифицировать отдельные лизаты клеток. В некоторых случаях очищенная нуклеиновая кислота из отдельных клеток контактирует с одним или более шариками, содержащими уникальный набор штрих-кодов образцов и/или последовательностей УМІ, чтобы идентифицировать очищенную нуклеиновую кислоту из отдельной клетки. С шариками можно проводить манипуляции любым подходящим способом, известным в настоящей области техники, например, с использованием капельных приводов, как описано в настоящем документе. Шариками могут быть любого подходящего размера, включая в себя, например, микрошарики, микрочастицы, наночастицы и наночастицы. Согласно некоторым вариантам осуществления шариками обладают магнитной чувствительностью; согласно другим вариантам осуществления шариками не обладают значительной магнитной чувствительностью. Неограничивающие примеры подходящих шариков включают в себя следующее: микрошарики для проточной цитометрии, микрочастицы и наночастицы полистирола, функционализированные микрочастицы и наночастицы полистирола, микрочастицы и наночастицы полистирола с покрытием, микрошарики кремнезема, флуоресцентные микросферы и наносферы, функционализированные флуоресцентные микросферы и наносферы, флуоресцентные микросферы и наносферы с покрытием, окрашенные микрочастицы и наночастицы, магнитные микрочастицы и наночастицы, суперпарамагнитные микрочастицы и наночастицы (например, DYNABEADS®, поставляемые Invitrogen Group, Карлсбад, Калифорния), флуоресцентные микрочастицы и наночастицы, магнитные микрочастицы и наночастицы с покрытием, ферромагнитные микрочастицы и наночастицы, ферромагнитные микрочастицы и наночастицы с покрытием и описанные в патентных публикациях США №№ US20050260686, US20030132538, US20050118574, 20050277197, 20060159962. Шариками могут быть предварительно связаны с антителом, белком или антигеном, ДНК/РНК-зондом или любой другой молекулой,



обладающей аффинностью к требуемой мишени. Согласно некоторым вариантам осуществления праймеры, несущие штрих-коды образцов и/или последовательности UMI, могут находиться в растворе. Согласно определенным вариантам осуществления может быть представлено множество капель, при этом каждая капля во множестве несет штрих-код образца, который является уникальным для капли, и UMI, который является уникальным для молекулы, так что UMI повторяются много раз в пределах набора капель. Согласно некоторым вариантам осуществления отдельные клетки контактируют с каплей с уникальным набором штрих-кодов образца и/или последовательностей UMI, чтобы идентифицировать индивидуальную клетку. Согласно некоторым вариантам осуществления лизаты отдельных клеток контактируют с каплей с уникальным набором штрих-кодов образца и/или последовательностей UMI, чтобы идентифицировать лизаты отдельных клеток. Согласно некоторым вариантам осуществления очищенная нуклеиновая кислота из отдельных клеток контактирует с каплей с уникальным набором штрих-кодов образца и/или последовательностей UMI, чтобы идентифицировать очищенную нуклеиновую кислоту из отдельной клетки.

**[0082]** Праймеры РТА могут содержать специфический для последовательности или случайный праймер, штрих-код клетки и/или уникальный молекулярный идентификатор (UMI) (см., например, фигуры 10А (линейный праймер) и 10В (шпилечный праймер)). В некоторых случаях праймер включает в себя праймер, специфический для последовательности. В некоторых случаях праймер включает в себя случайный праймер. В некоторых случаях праймер содержит штрих-код клетки. В некоторых случаях праймер содержит штрих-код образца. В некоторых случаях праймер содержит уникальный молекулярный идентификатор. В некоторых случаях праймеры содержат два или более клеточных штрих-кода. Такие штрих-коды в некоторых случаях идентифицируют уникальный источник образца или уникальный рабочий процесс. Такие штрих-коды или UMI в некоторых случаях составляют 5, 6, 7, 8, 9, 10, 11, 12, 15, 20, 25, 30 или более 30 оснований в длину. Праймеры в некоторых случаях содержат по меньшей мере 1000, 10000, 50000, 100000, 250000, 500000,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$  или по меньшей мере  $10^{10}$  уникальных штрих-кодов или UMI. В некоторых случаях праймеры содержат по меньшей мере 8, 16, 96 или 384 уникальных штрих-кода или UMI. В некоторых случаях стандартный адаптор затем лигируют с продуктами амплификации перед секвенированием; после секвенирования прочтения сначала присваивают конкретной клетке на основе штрих-кода клетки. Подходящие адапторы, которые можно использовать вместе со способом РТА, включают в себя, например, адапторы xGen® Dual Index UMI, доступные от Integrated DNA Technologies (IDT). Затем прочтения из каждой клетки группируются с использованием

UMI, и прочтения с одним и тем же UMI могут быть свернуты в «консенсусное прочтение». Использование штрих-кода клетки позволяет объединить все клетки перед подготовкой библиотеки, так как позже они могут быть идентифицированы по штрих-коду клетки. Использование UMI для формирования консенсусного прочтения в некоторых случаях корректирует смещение ПЦР, улучшая обнаружение вариации числа копий (CNV) (фигуры 11A и 11B). Кроме того, ошибки секвенирования можно скорректировать, потребовав, чтобы фиксированный процент прочтений одной и той же молекулы содержал одну и ту же замену основания, обнаруженное в каждом положении. Этот подход был использован для улучшения обнаружения CNV и исправления ошибок секвенирования в совокупных образцах. В некоторых случаях UMI используют с описанными в настоящем документе способами, например, в патенте США 8835358 раскрыт принцип цифрового подсчета после прикрепления случайного амплифицируемого штрих-кода. Schmitt с соавт. и Fan с соавт. раскрыли аналогичные способы исправления ошибок секвенирования.

**[0083]** Описанные в настоящем документе способы могут дополнительно включать дополнительные стадии, включая в себя стадии, выполняемые с образцом или матрицей. Такие образцы или матрицы в некоторых случаях подвергают одной или более стадиям перед РТА. В некоторых случаях образцы, содержащие клетки, подвергают стадии предварительной обработки. Например, клетки подвергают лизису и протеолизу для увеличения доступности хроматина с использованием комбинации замораживания-оттаивания, Triton X-100, Tween 20 и протеиназы К. Другие стратегии лизиса также подходят для применения на практике описанных в настоящем документе способов. Такие стратегии включают в себя без ограничения лизис с использованием других комбинаций обработки детергентом, и/или лизоцимом, и/или протеазой, и/или физическое разрушение клеток, такое как обработка ультразвуком и/или щелочной лизис и/или гипотонический лизис. В некоторых случаях первичную матрицу или целевую молекулу подвергают стадии предварительной обработки. В некоторых случаях первичную матрицу (или целевую молекулу) денатурируют с использованием гидроксида натрия с последующей нейтрализацией раствора. Другие стратегии денатурирования также могут быть подходящими для применения на практике описанных в настоящем документе способов. Такие стратегии могут включать в себя без ограничения комбинации щелочного лизиса с другими основными растворами, повышение температуры образца и/или изменение концентрации соли в образце, добавление добавок, таких как растворители или масла, другую модификацию или любую их комбинацию. В некоторых случаях дополнительные стадии включают сортировку, фильтрацию или выделение образцов, матриц или ампликонов по размеру. Например, после амплификации описанными в настоящем

документе способами библиотеки ампликонов обогащают ампликонами с требуемой длиной. В некоторых случаях библиотеки ампликонов обогащают ампликонами с длиной, составляющей 50-2000, 25-1000, 50-1000, 75-2000, 100-3000, 150-500, 75-250, 170-500, 100-500 или 75-2000 оснований. В некоторых случаях библиотеки ампликонов обогащают ампликонами с длиной не более 75, 100, 150, 200, 500, 750, 1000, 2000, 5000 или не более 10000 оснований. В некоторых случаях библиотеки ампликонов обогащают ампликонами с длиной, составляющей по меньшей мере 25, 50, 75, 100, 150, 200, 500, 750, 1000 или по меньшей мере 2000 оснований.

**[0084]** Описанные в настоящем документе способы и композиции могут включать в себя буферы или другие составы. Такие буферы в некоторых случаях содержат поверхностно-активные вещества/детергенты или денатурирующие агенты (Твин-20, DMSO, DMF, пегилированные полимеры, содержащие гидрофобную группу, или другое поверхностно-активное вещество), соли (фосфат калия или натрия (одноосновный или двухосновный), хлорид натрия, хлорид калия, TrisHCl, хлорид или сульфат магния, соли аммония, такие как фосфат, нитрат или сульфат, EDTA), восстановители (DTT, TNP, DTE, бета-меркаптоэтанол, TCEP или другой восстановитель) или другие компоненты (глицерин, гидрофильные полимеры, такие как PEG). В некоторых случаях буферы используют в сочетании с такими компонентами, как полимеразы, факторы замещения цепи, терминаторы или другой компонент реакции, описанный в настоящем документе.

**[0085]** Молекулы нуклеиновой кислоты, амплифицированные описанными в настоящем документе способами, можно секвенировать и проанализировать с использованием способов, известных специалистам в настоящей области техники. Неограничивающие примеры способов секвенирования, которые в некоторых случаях используют, включают в себя, например, секвенирование путем гибридизации (SBH), секвенирование путем лигирования (SBL) (Shendure et al. (2005) Science 309:1728), секвенирование с количественным последовательным добавлением флуоресцентных нуклеотидов (QIFNAS), пошаговое лигирование и расщепление, резонансный перенос энергии флуоресценции (FRET), молекулярные маяки, расщепление репортерного зонда TaqMan, пиросеквенирование, флуоресцентное секвенирование *in situ* (FISSEQ), шарики FISSEQ (патент США № 7425431), wobble-секвенирование (публикация международной патентной заявки № WO2006/073504), мультиплексное секвенирование (публикация заявки на выдачу патента США № US2008/0269068; Porreca et al., 2007, Nat. Methods 4:931), секвенирование полимеризованных колоний (POLONY) (патенты США №№ 6432360, 6485944 и 6511803, и публикация международной патентной заявки № WO2005/082098), секвенирование по методу катящегося кольца с использованием наносетки (ROLONY)

(патент США № 9624538), аллель-специфические анализы лигирования олигонуклеотидов (например, анализ лигирования олигонуклеотидов (OLA), OLA одиночной матричной молекулы с использованием лигированного линейного зонда и считывание результатов амплификации по методу катящегося кольца (RCA), лигированные замыкающих кольцо зондов и/или OLA одиночной матричной молекулы с использованием лигированного циклического замыкающего кольцо зонда и считывания амплификации по методу катящегося кольца (RCA)), высокопроизводительные способы секвенирования, такие как, например, способы с использованием платформ Roche 454, Illumina Solexa, AB-SOLiD, Helicos, Polonator и т.п., а также технологии секвенирования на основе использования света (Landegren et al. (1998) *Genome Res.* 8:769-76; Kwok (2000) *Pharmacogenomics* 1:95-100; и Shi (2001) *Clin. Chem.* 0,47:164-172). В некоторых случаях амплифицированные молекулы нуклеиновых кислот секвенируют методом «дробовика».

**[0086] Способы и применения**

**[0087]** В настоящем документе описаны способы выявления мутаций в клетках с помощью способов РТА. Применение способа РТА в некоторых случаях приводит к улучшениям по сравнению с известными способами, например MDA. РТА в некоторых случаях характеризуется более низкой частотой обнаружения ложноположительных и ложноотрицательных вариантов, чем способ MDA. Геномы, такие как платиновые геномы NA12878, в некоторых случаях используют для определения того, приведет ли больший охват генома и однородность РТА к снижению частоты обнаружения ложноположительных вариантов. Без ограничения какой-либо теорией можно определить, что отсутствие распространения ошибок в РТА снижает частоту обнаружения ложноположительных вариантов. Баланс амплификации между аллелями с помощью этих двух способов в некоторых случаях оценивают путем сравнения частот аллелей обнаружений гетерозиготных мутаций в известных положительных локусах. В некоторых случаях библиотеки ампликонов, созданные с использованием РТА, дополнительно амплифицируют с помощью ПЦР.

**[0088]** Клетки, проанализированные с использованием описанных в настоящем документе способов, в некоторых случаях включают в себя опухолевые клетки. Например, циркулирующие опухолевые клетки можно выделить из жидкости, взятой у пациентов, такой как без ограничения кровь, костный мозг, моча, слюна, спинномозговая жидкость, плевральная жидкость, перикардальная жидкость, асцит или водянистая влага. Затем клетки подвергают описанным в настоящем документе способам (например, РТА) и секвенированию для определения бремени мутаций и комбинации мутаций в каждой клетке. Эти данные в некоторых случаях используют для диагностики конкретного

заболевания или в качестве инструментов для прогнозирования ответа на лечение. Аналогичным образом, в некоторых случаях клетки с неизвестным злокачественным потенциалом в некоторых случаях выделяют из жидкости, взятой у пациентов, такой как без ограничения кровь, костный мозг, моча, слюна, спинномозговая жидкость, плевральная жидкость, перикардальная жидкость, асцит или водянистая влага. После использования описанных в настоящем документе способов и секвенирования такие способы в дальнейшем используют для определения бремени мутаций и комбинации мутаций в каждой клетке. Эти данные в некоторых случаях используют для диагностики конкретного заболевания или в качестве инструментов для прогнозирования прогрессирования предракового состояния в явное злокачественное новообразование. В некоторых случаях клетки можно выделить из образцов первичной опухоли. Затем клетки могут подвергать РГА и секвенированию для определения бремени мутаций и комбинации мутаций в каждой клетке. Эти данные можно использовать для диагностики конкретного заболевания или в качестве инструментов для прогнозирования вероятности того, что злокачественное новообразование пациента является устойчивым к доступным противораковым лекарственным средствам. При воздействии на образцы различных химиотерапевтических средств обнаружили, что основные и минорные клоны обладают разной чувствительностью к конкретным лекарственным средствам, которая не обязательно коррелирует с наличием известной «мутации-драйвера», что позволяет предположить, что комбинации мутаций в клональной популяции определяют ее чувствительность к конкретным химиотерапевтическим лекарственным средствам. Без ограничения какой-либо теорией эти результаты позволяют предположить, что злокачественное новообразование можно легче устранить, если обнаружены предраковые поражения, которые еще не расширились и превратились в клоны, повышенное количество модификаций генома которых может сделать их более устойчивыми к лечению. См. Ma et al., 2018, «Pan-cancer genome and transcriptome analyses of 1,699 pediatric leukemias and solid tumors». Протокол геномики одиночных клеток в некоторых случаях используют для обнаружения комбинаций соматических генетических вариантов в одиночной раковой клетке или клонотипе в смеси нормальных и злокачественных клеток, выделенных из образцов пациента. Эту технологию в некоторых случаях в дальнейшем используют для идентификации клонотипов, которые подвергаются положительному отбору после воздействия лекарственных средств как *in vitro*, так и в организме пациентов. Как показано на фиг. 6А, путем сравнения выживших клонов, подвергшихся химиотерапии, с клонами, идентифицированными при постановке диагноза, можно создать каталог клонотипов рака, который документирует их устойчивость к конкретным лекарственным средствам. В некоторых случаях способы РГА

обнаруживают чувствительность определенных клонов в образце, состоящем из нескольких клонотипов, к существующим или новым лекарственным средствам, а также к их комбинациям, где способ может определять чувствительность конкретных клонов к лекарственному средству. Этот подход в некоторых случаях показывает эффективность лекарственного средства для конкретного клона, который может не быть обнаружен с помощью существующих измерений чувствительности к лекарственным средствам, которые учитывают чувствительность всех раковых клонов вместе в одном измерении. Когда РТА, описанный в настоящем документе, применяют к образцам пациентов, собранным во время постановки диагноза, с целью обнаружения раковых клонотипов в раке данного пациента, каталог чувствительности к лекарственным средствам можно использовать для поиска этих клонов и, таким образом, информирования онкологов о том, какое лекарственное средство или комбинация лекарственных средств не будут работать, и какое лекарственное средство или комбинация лекарственных средств, скорее всего, будут эффективны против этого рака пациента.

**[0089]** В настоящем документе описаны способы измерения мутагенности фактора окружающей среды. Например, клетки (одиночные или популяция) подвергаются потенциальному воздействию окружающей среды. Например, в некоторых случаях в способе используют клетки, происходящие из органов (печень, поджелудочная железа, легкие, толстая кишка, щитовидная железа или другой орган), тканей (кожа или другие ткани), крови или другого биологического источника. В некоторых случаях условия окружающей среды включают в себя тепло, свет (например, ультрафиолет), облучение, химическое вещество или любую их комбинацию. После определенного воздействия условий окружающей среды, в некоторых случаях минут, часов, дней или дольше, одиночные клетки выделяют и подвергают способу РТА. В некоторых случаях для маркировки образца используют молекулярные штрих-коды и уникальные молекулярные идентификаторы. Образец секвенируют и затем анализируют для идентификации мутаций, вызванных воздействием условий окружающей среды. В некоторых случаях такие мутации сравнивают с контрольным условием окружающей среды, таким как известное немутагенное вещество, носитель/растворитель или отсутствие условия окружающей среды. Такой анализ в некоторых случаях дает не только общее количество мутаций, вызванных условием окружающей среды, но также местонахождение и характер таких мутаций. В некоторых случаях из данных идентифицируют паттерны и их могут использовать для диагностики заболеваний или состояний. В некоторых случаях паттерны используют для прогнозирования будущих заболеваний или состояний. В некоторых случаях описанные в настоящем документе способы измеряют бремя мутаций,

местоположения и паттерны в клетке после воздействия агента окружающей среды, такого как, например, потенциальный мутаген или тератоген. Этот подход в некоторых случаях используют для оценки безопасности данного агента, в том числе его способности вызывать мутации, которые могут способствовать развитию заболевания. Например, способ можно использовать для прогнозирования канцерогенности или тератогенности агента в отношении определенных типов клеток после воздействия определенной концентрации определенного агента.

**[0090]** В настоящем документе описаны способы идентификации мутаций в клетках животных, растений или микроорганизмов, которые подверглись редактированию генома (например, с использованием технологий CRISPR). Такие клетки в некоторых случаях можно выделить и подвергнуть РГА и секвенированию для определения бремени мутаций и комбинации мутаций в каждой клетке. Частоту мутаций на клетку и локализацию мутаций, которые возникают в результате протокола редактирования генома, в некоторых случаях используют для оценки безопасности данного способа редактирования генома.

**[0091]** В настоящем документе описаны способы определения мутаций в клетках, которые используют для клеточной терапии, такой как без ограничения трансплантация индуцированных плюрипотентных стволовых клеток, трансплантация гемопоэтических или других клеток, которые не подвергались манипуляциям, или трансплантация гемопоэтических или других клеток, подвергшихся редактированию генома. Затем клетки могут подвергать РГА и секвенированию для определения бремени мутаций и комбинации мутаций в каждой клетке. Частоту мутаций на клетку и локализацию мутаций в продукте для клеточной терапии можно использовать для оценки безопасности и потенциальной эффективности продукта.

**[0092]** Согласно дополнительному варианту осуществления клетки можно выделить из бластомеров, созданных с помощью оплодотворения *in vitro*. Затем клетки можно подвергать РГА и секвенированию, чтобы определить бремя и комбинацию потенциально предрасполагающих к заболеванию генетических вариантов в каждой клетке. Профиль мутации клетки затем можно использовать для экстраполяции генетической предрасположенности бластомера к конкретным заболеваниям перед имплантацией.

**[0093]** Согласно другому варианту осуществления клетки микроорганизмов (например, бактерии, грибы, простейшие) можно выделить из растений или животных (например, из образцов микробиоты [например, микробиоты ЖКТ, микробиоты кожи и т.д.] или из жидкостей организма, таких как, например, кровь, костный мозг, моча, слюна, спинномозговая жидкость, плевральная жидкость, перикардальная жидкость, асцит или водянистая влага). Кроме того, клетки микроорганизмов можно выделить из постоянных

изделий медицинского назначения, таких как без ограничения внутривенные катетеры, уретральные катетеры, цереброспинальные шунты, искусственные клапаны, искусственные суставы или эндотрахеальные трубки. Затем клетки можно подвергать РГА и секвенированию для определения идентичности конкретного микроорганизма, а также для обнаружения наличия генетических вариантов микроорганизмов, которые предсказывают ответ (или устойчивость) к конкретным антимикробным агентам. Эти данные можно использовать для диагностики конкретного инфекционного заболевания и/или в качестве инструментов для прогнозирования ответа на лечение.

**[0094]** В настоящем документе описаны способы создания библиотек ампликонов из образцов, содержащих короткую нуклеиновую кислоту, с использованием описанных в настоящем документе способов РГА. В некоторых случаях РГА приводит к повышению точности и однородности амплификации более коротких нуклеиновых кислот. В некоторых случаях длина нуклеиновых кислот не превышает 2000 оснований. В некоторых случаях длина нуклеиновых кислот не превышает 1000 оснований. В некоторых случаях длина нуклеиновых кислот не превышает 500 оснований. В некоторых случаях нуклеиновые кислоты характеризуются длиной, составляющей не более 200, 400, 750, 1000, 2000 или 5000 оснований. В некоторых случаях образцы, содержащие короткие фрагменты нуклеиновых кислот, включают в себя без ограничения древнюю ДНК (возрастом сотни, тысячи, миллионы или даже миллиарды лет), образцы FFPE (фиксированные формалином, залитые парафином), внеклеточную ДНК или другой образец, содержащий короткие нуклеиновые кислоты.

**[0095]            Варианты осуществления**

**[0096]** В настоящем документе описаны способы амплификации целевой молекулы нуклеиновой кислоты, причем способ включает следующее: а) приведение в контакт образца, содержащего целевую молекулу нуклеиновой кислоты, одного или более праймеров для амплификации, полимеразы нуклеиновых кислот и смеси нуклеотидов, которая содержит один или более терминаторных нуклеотидов, которые прекращают репликацию нуклеиновой кислоты полимеразой, и б) инкубация образца при условиях, которые стимулируют репликацию целевой молекулы нуклеиновой кислоты для получения множества терминированных продуктов амплификации, где репликация происходит путем репликации с замещением цепи. Согласно одному варианту осуществления любого из приведенных выше способов способ дополнительно включает выделение из множества терминированных продуктов амплификации продуктов, которые составляют от около 50 до около 2000 нуклеотидов в длину. Согласно одному варианту осуществления любого из приведенных выше способов способ дополнительно включает выделение из множества



терминированных продуктов амплификации продуктов, которые составляют от около 400 до около 600 нуклеотидов в длину. Согласно одному варианту осуществления любого из приведенных выше способов способ дополнительно включает следующее: с) удаление концевых терминаторных нуклеотидов из терминированных продуктов амплификации; d) репарация концов и добавление А-хвостов к 3'-концу ДНК с тупыми концами, и е) лигирование молекул, полученных на стадии (d), с адапторами, и тем самым создавая библиотеку продуктов амплификации. Согласно одному варианту осуществления любого из приведенных выше способов способ дополнительно включает секвенирование продуктов амплификации. Согласно одному варианту осуществления любого из приведенных выше способов амплификацию проводят при по существу изотермических условиях. Согласно одному варианту осуществления любого из приведенных выше способов полимеразы нуклеиновых кислот представляет собой ДНК-полимеразу.

**[0097]** Согласно одному варианту осуществления любого из приведенных выше способов ДНК-полимераза представляет собой замещающую цепи ДНК-полимеразу. Согласно одному варианту осуществления любого из приведенных выше способов полимеразы нуклеиновых кислот выбрана из следующего: полимеразы бактериофага phi29 (Ф29), генетически модифицированная ДНК-полимераза phi29 (Ф29), фрагмент Кленова ДНК-полимеразы I, ДНК-полимераза фага M2, ДНК-полимераза фага phiPRD1, ДНК-полимераза Bst, ДНК-полимераза большого фрагмента Bst, полимеразы экзо(-) Bst, ДНК-полимераза экзо(-) Vsa, ДНК-полимераза Bsu, ДНК-полимераза Vent<sub>R</sub>, ДНК-полимераза Vent<sub>R</sub> (экзо-), ДНК-полимераза Deep Vent, ДНК-полимераза Deep Vent (экзо-), ДНК-полимераза IsoPol, ДНК-полимераза I, ДНК-полимераза Terminator, ДНК-полимераза T5, секвенза, ДНК-полимераза T7, T7-секвенза и ДНК-полимераза T4. Согласно одному варианту осуществления любого из приведенных выше способов полимеразы нуклеиновых кислот характеризуется 3'→5' экзонуклеазной активностью, и терминаторные нуклеотиды ингибируют такую 3'→5' экзонуклеазную активность. Согласно одному конкретному варианту осуществления терминаторные нуклеотиды выбраны из следующего: нуклеотиды с модификацией альфа-группы (например, альфа-тио-дидезоксинуклеотиды, создающие фосфоротиоатную связь), C3 спейсерные нуклеотиды, запертые нуклеиновые кислоты (LNA), инвертированные нуклеиновые кислоты, 2' фторнуклеотиды, 3' фосфорилированные нуклеотиды, 2'-О-метил-модифицированные нуклеотиды и транснуклеиновые кислоты. Согласно одному варианту осуществления любого из приведенных выше способов полимеразы нуклеиновых кислот не характеризуется 3'→5' экзонуклеазной активностью. Согласно одному конкретному варианту осуществления полимеразы выбрана из следующего: ДНК-полимераза Bst, полимеразы экзо(-) Bst, ДНК-

полимераза экзо(-) Bca, ДНК-полимераза Bsu, ДНК-полимераза Vent<sub>R</sub> (экзо-), ДНК-полимераза Deep Vent (экзо-), ДНК-полимераза фрагмента Кленова (экзо-) и ДНК-полимераза Terminator. Согласно одному конкретному варианту осуществления терминаторные нуклеотиды содержат модификации группы г 3' углерода дезоксирибозы. Согласно одному конкретному варианту осуществления терминаторные нуклеотиды выбраны из следующего: 3' блокированный обратимый терминатор, содержащий нуклеотиды, 3' неблокированный обратимый терминатор, содержащий нуклеотиды, терминаторы, содержащие 2' модификации дезоксинуклеотидов, терминаторы, содержащие модификации азотистого основания дезоксинуклеотидов, и их комбинации. Согласно одному конкретному варианту осуществления терминаторные нуклеотиды выбраны из следующего: дидезоксинуклеотиды, инвертированные дидезоксинуклеотиды, 3' биотинилированные нуклеотиды, 3' аминонуклеотиды, 3'-фосфорилированные нуклеотиды, 3'-О-метилнуклеотиды, 3' углеродные спейсерные нуклеотиды, включая в себя 3' C3 спейсерные нуклеотиды, 3' C18 нуклеотиды, 3' гександиоловые спейсерные нуклеотиды, ациклонуклеотиды и их комбинации. Согласно одному варианту осуществления любого из приведенных выше способов праймеры для амплификации составляют от 4 до 70 нуклеотидов в длину. Согласно одному варианту осуществления любого из приведенных выше способов продукты амплификации составляют от около 50 до около 2000 нуклеотидов в длину. Согласно одному варианту осуществления любого из приведенных выше способов целевая нуклеиновая кислота представляет собой ДНК (например, кДНК или геномная ДНК). Согласно одному варианту осуществления любого из приведенных выше способов праймеры для амплификации представляют собой случайные праймеры. Согласно одному варианту осуществления любого из приведенных выше способов праймеры для амплификации содержат штрих-код. Согласно одному конкретному варианту осуществления штрих-код содержит штрих-код клетки. Согласно одному конкретному варианту осуществления штрих-код содержит штрих-код образца. Согласно одному варианту осуществления любого из приведенных выше способов праймеры для амплификации содержат уникальный молекулярный идентификатор (UMI). Согласно одному варианту осуществления любого из приведенных выше способов способ включает денатурирование целевой нуклеиновой кислоты или геномной ДНК перед начальным отжигом праймера. Согласно одному конкретному варианту осуществления денатурацию проводят в щелочных условиях с последующей нейтрализацией. Согласно одному варианту осуществления любого из приведенных выше способов образец, праймеры для амплификации, полимеразы нуклеиновых кислот и смесь нуклеотидов содержатся в микрожидкостном устройстве. Согласно одному варианту осуществления

любого из приведенных выше способов образец, праймеры для амплификации, полимеразы нуклеиновых кислот и смесь нуклеотидов содержатся в капле. Согласно одному варианту осуществления любого из приведенных выше способов образец выбран из следующего: образцы ткани(ей), клетки, образцы биологической жидкости (например, кровь, моча, слюна, лимфатическая жидкость, спинномозговая жидкость (CSF), амниотическая жидкость, плевральная жидкость, перикардальная жидкость, асцит, водянистая влага), образцы костного мозга, образцы спермы, биопсийные образцы, образцы рака, образцы опухоли, образцы клеточного лизата, криминалистические образцы, археологические образцы, палеонтологические образцы, образцы инфекции, производственные образцы, целые растения, части растения, образцы микробиоты, вирусные препараты, образцы почвы, образцы морской воды, образцы пресной воды, бытовые или промышленные образцы и их комбинации и изоляты. Согласно одному варианту осуществления любого из приведенных выше способов образец представляет собой клетку (например, клетка животного [например, клетка человека], растительная клетка, грибковая клетка, бактериальная клетка и клетка простейшего). Согласно одному конкретному варианту осуществления клетку лизируют перед репликацией. Согласно одному конкретному варианту осуществления клеточный лизис сопровождается протеолизом. Согласно одному конкретному варианту осуществления клетка выбрана из следующего: клетка предимплантационного эмбриона, стволовая клетка, фетальная клетка, опухолевая клетка, предполагаемая раковая клетка, раковая клетка, клетка, подвергнутая процедуре редактирования генов, клетка из патогенного организма, клетка, полученная из криминалистического образца, клетка, полученная из археологического образца, и клетка, полученная из палеонтологического образца. Согласно одному варианту осуществления любого из приведенных выше способов образец представляет собой клетку из предимплантационного эмбриона (например, бластомер [например, бластомер, полученный из эмбриона восьмиклеточной стадии, полученного с помощью *in vitro* оплодотворения]). Согласно одному конкретному варианту осуществления способ дополнительно включает определение наличия предрасполагающих к заболеванию вариантов зародышевой линии или соматических вариантов в эмбриональной клетке. Согласно одному варианту осуществления любого из приведенных выше способов образец представляет собой клетку из патогенного организма (например, бактерии, гриба, простейшего). Согласно одному конкретному варианту осуществления клетку патогенного организма получают из жидкости, взятой у пациента, образца микробиоты (например, образец микробиоты ЖКТ, образец вагинальной микробиоты, образец микробиоты кожи и т.д.) или постоянного изделия медицинского назначения (например, внутривенный катетер,

уретральный катетер, цереброспинальный шунт, искусственный клапан, искусственный сустав, эндотрахеальная трубка и т.д.). Согласно одному конкретному варианту осуществления способ дополнительно включает стадию определения идентичности патогенного организма. Согласно одному конкретному варианту осуществления способ дополнительно включает определение наличия генетических вариантов, ответственных за устойчивость патогенного организма к лечению. Согласно одному варианту осуществления любого из приведенных выше способов образец представляет собой опухолевую клетку, предполагаемую раковую клетку или раковую клетку. Согласно одному конкретному варианту осуществления способ дополнительно включает определение наличия одной или более диагностических или прогностических мутаций. Согласно одному конкретному варианту осуществления способ дополнительно включает определение наличия вариантов зародышевой линии или соматических вариантов, ответственных за устойчивость к лечению. Согласно одному варианту осуществления любого из приведенных выше способов образец представляет собой клетку, подвергнутую процедуре редактирования генов. Согласно одному конкретному варианту осуществления способ дополнительно включает определение наличия незапланированных мутаций, вызванных процессом редактирования генов. Согласно одному варианту осуществления любого из приведенных выше способов способ дополнительно включает определение истории клеточной родословной. Согласно родственному аспекту настоящее изобретение относится к применению любого из приведенных выше способов для идентификации низкочастотных вариантов последовательности (например, вариантов, которые составляют  $\geq 0,01\%$  от общего числа последовательностей).

**[0098]** Согласно родственному аспекту настоящее изобретение относится к набору, содержащему полимеразу нуклеиновых кислот, один или более праймеров для амплификации, смесь нуклеотидов, содержащую один или более терминаторных нуклеотидов, и необязательно инструкции по применению. Согласно одному варианту осуществления наборов согласно настоящему изобретению полимеразы нуклеиновых кислот представляет собой замещающую цепи ДНК-полимеразу. Согласно одному варианту осуществления наборов согласно настоящему изобретению полимеразы нуклеиновых кислот выбрана из следующего: полимеразы бактериофага phi29 (Ф29), генетически модифицированная ДНК-полимераза phi29 (Ф29), фрагмент Кленова ДНК-полимеразы I, ДНК-полимераза фага M2, ДНК-полимераза фага phiPRD1, ДНК-полимераза Bst, ДНК-полимераза большого фрагмента Bst, полимеразы экзо(-) Bst, ДНК-полимераза экзо(-) Bca, ДНК-полимераза Bsu, ДНК-полимераза Vent<sub>r</sub>, ДНК-полимераза Vent<sub>r</sub> (экзо-), ДНК-полимераза Deep Vent, ДНК-полимераза Deep Vent (экзо-), ДНК-полимераза IsoPol,

ДНК-полимераза I, ДНК-полимераза Terminator, ДНК-полимераза T5, секвенаса, ДНК-полимераза T7, T7-секвенаса и ДНК-полимераза T4. Согласно одному варианту осуществления наборов согласно настоящему изобретению полимераз нуклеиновых кислот характеризуется 3'→5' экзонуклеазной активностью, и терминаторные нуклеотиды ингибируют такую 3'→5' экзонуклеазную активность (например, нуклеотиды с модификацией альфа-группы [например, альфа-тио-дидезоксинуклеотиды], C3 спейсерные нуклеотиды, запертые нуклеиновые кислоты (LNA), инвертированные нуклеиновые кислоты, 2' фторнуклеотиды, 3' фосфорилированные нуклеотиды, 2'-О-метилмодифицированные нуклеотиды, транснуклеиновые кислоты). Согласно одному варианту осуществления наборов согласно настоящему изобретению полимераз нуклеиновых кислот не характеризуется 3'→5' экзонуклеазной активностью (например, ДНК-полимераза Bst, полимеразы экзо(-) Bst, ДНК-полимераза экзо(-) Bca, ДНК-полимераза Bsu, ДНК-полимераза Vent<sub>R</sub> (экзо-), ДНК-полимераза Deep Vent (экзо-), ДНК-полимераза фрагмента Кленова (экзо-), ДНК-полимераза Terminator). Согласно одному конкретному варианту осуществления терминаторные нуклеотиды содержат модификации группы γ 3' углерода дезоксирибозы. Согласно одному конкретному варианту осуществления терминаторные нуклеотиды выбраны из следующего: 3' блокированный обратимый терминатор, содержащий нуклеотиды, 3' неблокированный обратимый терминатор, содержащий нуклеотиды, терминаторы, содержащие 2' модификации дезоксинуклеотидов, терминаторы, содержащие модификации азотистого основания дезоксинуклеотидов, и их комбинации. Согласно одному конкретному варианту осуществления терминаторные нуклеотиды выбраны из следующего: дидезоксинуклеотиды, инвертированные дидезоксинуклеотиды, 3' биотинилированные нуклеотиды, 3' аминонуклеотиды, 3'-фосфорилированные нуклеотиды, 3'-О-метилнуклеотиды, 3' углеродные спейсерные нуклеотиды, включая в себя 3' C3 спейсерные нуклеотиды, 3' C18 нуклеотиды, 3' гександиоловые спейсерные нуклеотиды, ациклонуклеотиды и их комбинации.

**[0099]** В настоящем документе описаны способы амплификации генома, причем способ включает следующее: а) приведение в контакт образца, содержащего геном, множества праймеров для амплификации (например, два или больше праймеров), полимеразы нуклеиновых кислот и смеси нуклеотидов, которая содержит один или более терминаторных нуклеотидов, которые прекращают репликацию нуклеиновой кислоты полимеразой, и б) инкубация образца при условиях, которые стимулируют репликацию генома, для получения множества терминированных продуктов амплификации, где репликация происходит путем репликации с замещением цепи. Согласно одному варианту осуществления любого из приведенных выше способов способ дополнительно включает

выделение из множества терминированных продуктов амплификации продуктов, которые составляют от около 50 до около 2000 нуклеотидов в длину. Согласно одному варианту осуществления любого из приведенных выше способов способ дополнительно включает выделение из множества терминированных продуктов амплификации продуктов, которые составляют от около 400 до около 600 нуклеотидов в длину. Согласно одному варианту осуществления любого из приведенных выше способов способ дополнительно включает следующее: с) удаление концевых терминаторных нуклеотидов из терминированных продуктов амплификации; d) репарация концов и добавление А-хвостов к 3'-концу ДНК с тупыми концами, и е) лигирование молекул, полученных на стадии (d), с адапторами, и тем самым создавая библиотеку продуктов амплификации. Согласно одному варианту осуществления любого из приведенных выше способов способ дополнительно включает секвенирование продуктов амплификации. Согласно одному варианту осуществления любого из приведенных выше способов амплификацию проводят при по существу изотермических условиях. Согласно одному варианту осуществления любого из приведенных выше способов полимеразы нуклеиновых кислот представляет собой ДНК-полимеразу.

**[00100]** Согласно одному варианту осуществления любого из приведенных выше способов ДНК-полимеразы представляет собой замещающую цепи ДНК-полимеразу. Согласно одному варианту осуществления любого из приведенных выше способов полимеразы нуклеиновых кислот выбрана из следующего: полимеразы бактериофага phi29 (Ф29), генетически модифицированная ДНК-полимераза phi29 (Ф29), фрагмент Кленова ДНК-полимеразы I, ДНК-полимеразы фага M2, ДНК-полимеразы фага phiPRD1, ДНК-полимеразы Bst, ДНК-полимеразы большого фрагмента Bst, полимеразы экзо(-) Bst, ДНК-полимеразы экзо(-) Vsa, ДНК-полимеразы Bsu, ДНК-полимеразы Vent<sub>R</sub>, ДНК-полимеразы Vent<sub>R</sub> (экзо-), ДНК-полимеразы Deep Vent, ДНК-полимеразы Deep Vent (экзо-), ДНК-полимеразы IsoPol, ДНК-полимеразы I, ДНК-полимеразы Terminator, ДНК-полимеразы T5, секвеназа, ДНК-полимеразы T7, T7-секвеназа и ДНК-полимеразы T4. Согласно одному варианту осуществления любого из приведенных выше способов полимеразы нуклеиновых кислот характеризуется 3'→5' экзонуклеазной активностью, и терминаторные нуклеотиды ингибируют такую 3'→5' экзонуклеазную активность. Согласно одному конкретному варианту осуществления терминаторные нуклеотиды выбраны из следующего: нуклеотиды с модификацией альфа-группы (например, альфа-тио-дидезоксинуклеотиды, создающие фосфоротиоатную связь), C3 спейсерные нуклеотиды, запертые нуклеиновые кислоты (LNA), инвертированные нуклеиновые кислоты, 2' фторнуклеотиды, 3' фосфорилированные нуклеотиды, 2'-О-метил-модифицированные нуклеотиды и

транснуклеиновые кислоты. Согласно одному варианту осуществления любого из приведенных выше способов полимеразы нуклеиновых кислот не характеризуется 3'→5' экзонуклеазной активностью. Согласно одному конкретному варианту осуществления полимеразы выбрана из следующего: ДНК-полимеразы Bst, полимеразы экзо(-) Bst, ДНК-полимеразы экзо(-) Bca, ДНК-полимеразы Bsu, ДНК-полимеразы Vent<sub>R</sub> (экзо-), ДНК-полимеразы Deep Vent (экзо-), ДНК-полимеразы фрагмента Кленова (экзо-) и ДНК-полимеразы Terminator. Согласно одному конкретному варианту осуществления терминаторные нуклеотиды содержат модификации группы γ 3' углерода дезоксирибозы. Согласно одному конкретному варианту осуществления терминаторные нуклеотиды выбраны из следующего: 3' блокированный обратимый терминатор, содержащий нуклеотиды, 3' неблокированный обратимый терминатор, содержащий нуклеотиды, терминаторы, содержащие 2' модификации дезоксинуклеотидов, терминаторы, содержащие модификации азотистого основания дезоксинуклеотидов, и их комбинации. Согласно одному конкретному варианту осуществления терминаторные нуклеотиды выбраны из следующего: дидезоксинуклеотиды, инвертированные дидезоксинуклеотиды, 3' биотинилированные нуклеотиды, 3' аминонуклеотиды, 3'-фосфорилированные нуклеотиды, 3'-O-метилнуклеотиды, 3' углеродные спейсерные нуклеотиды, включая в себя 3' C3 спейсерные нуклеотиды, 3' C18 нуклеотиды, 3' гександиоловые спейсерные нуклеотиды, ациклонуклеотиды и их комбинации. Согласно одному варианту осуществления любого из приведенных выше способов праймеры для амплификации составляют от 4 до 70 нуклеотидов в длину. Согласно одному варианту осуществления любого из приведенных выше способов продукты амплификации составляют от около 50 до около 2000 нуклеотидов в длину. Согласно одному варианту осуществления любого из приведенных выше способов целевая нуклеиновая кислота представляет собой ДНК (например, кДНК или геномную ДНК). Согласно одному варианту осуществления любого из приведенных выше способов праймеры для амплификации представляют собой случайные праймеры. Согласно одному варианту осуществления любого из приведенных выше способов праймеры для амплификации содержат штрих-код. Согласно одному конкретному варианту осуществления штрих-код содержит штрих-код клетки. Согласно одному конкретному варианту осуществления штрих-код содержит штрих-код образца. Согласно одному варианту осуществления любого из приведенных выше способов праймеры для амплификации содержат уникальный молекулярный идентификатор (UMI). Согласно одному варианту осуществления любого из приведенных выше способов способ включает денатурирование целевой нуклеиновой кислоты или геномной ДНК перед начальным отжигом праймера. Согласно одному конкретному варианту осуществления

денатурацию проводят в щелочных условиях с последующей нейтрализацией. Согласно одному варианту осуществления любого из приведенных выше способов образец, праймеры для амплификации, полимеразы нуклеиновых кислот и смесь нуклеотидов содержатся в микрожидкостном устройстве. Согласно одному варианту осуществления любого из приведенных выше способов образец, праймеры для амплификации, полимеразы нуклеиновых кислот и смесь нуклеотидов содержатся в капле. Согласно одному варианту осуществления любого из приведенных выше способов образец выбран из следующего: образцы ткани(ей), клетки, образцы биологической жидкости (например, кровь, моча, слюна, лимфатическая жидкость, спинномозговая жидкость (CSF), амниотическая жидкость, плевральная жидкость, перикардальная жидкость, асцит, водянистая влага), образцы костного мозга, образцы спермы, биопсийные образцы, образцы рака, образцы опухоли, образцы клеточного лизата, криминалистические образцы, археологические образцы, палеонтологические образцы, образцы инфекции, производственные образцы, целые растения, части растения, образцы микробиоты, вирусные препараты, образцы почвы, образцы морской воды, образцы пресной воды, бытовые или промышленные образцы и их комбинации и изоляты. Согласно одному варианту осуществления любого из приведенных выше способов образец представляет собой следующее: клетка (например, клетка животного [например, клетка человека], растительная клетка, грибковая клетка, бактериальная клетка и клетка простейшего). Согласно одному конкретному варианту осуществления клетку лизируют перед репликацией. Согласно одному конкретному варианту осуществления клеточный лизис сопровождается протеолизом. Согласно одному конкретному варианту осуществления клетка выбрана из следующего: клетка из предимплантационного эмбриона, стволовая клетка, фетальная клетка, опухолевая клетка, предполагаемая раковая клетка, раковая клетка, клетка, подвергнутая процедуре редактирования генов, клетка из патогенного организма, клетка, полученная из криминалистического образца, клетка, полученная из археологического образца, и клетка, полученная из палеонтологического образца. Согласно одному варианту осуществления любого из приведенных выше способов образец представляет собой клетку из предимплантационного эмбриона (например, бластомер [например, бластомер, полученный из эмбриона восьмиклеточной стадии, полученного с помощью *in vitro* оплодотворения]). Согласно одному конкретному варианту осуществления способ дополнительно включает определение наличия предрасполагающих к заболеванию вариантов зародышевой линии или соматических вариантов в эмбриональной клетке. Согласно одному варианту осуществления любого из приведенных выше способов образец представляет собой клетку из патогенного организма (например, бактерии, гриба,



простейшего). Согласно одному конкретному варианту осуществления клетку патогенного организма получают из жидкости, взятой у пациента, образца микробиоты (например, образец микробиоты ЖКТ, образец вагинальной микробиоты, образец микробиоты кожи и т.д.) или постоянного изделия медицинского назначения (например, внутривенный катетер, уретральный катетер, цереброспинальный шунт, искусственный клапан, искусственный сустав, эндотрахеальная трубка и т.д.). Согласно одному конкретному варианту осуществления способ дополнительно включает стадию определения идентичности патогенного организма. Согласно одному конкретному варианту осуществления способ дополнительно включает определение наличия генетических вариантов, ответственных за устойчивость патогенного организма к лечению. Согласно одному варианту осуществления любого из приведенных выше способов образец представляет собой опухолевую клетку, предполагаемую раковую клетку или раковую клетку. Согласно одному конкретному варианту осуществления способ дополнительно включает определение наличия одной или более диагностических или прогностических мутаций. Согласно одному конкретному варианту осуществления способ дополнительно включает определение наличия вариантов зародышевой линии или соматических вариантов, ответственных за устойчивость к лечению. Согласно одному варианту осуществления любого из приведенных выше способов образец представляет собой клетку, подвергнутую процедуре редактирования генов. Согласно одному конкретному варианту осуществления способ дополнительно включает определение наличия незапланированных мутаций, вызванных процессом редактирования генов. Согласно одному варианту осуществления любого из приведенных выше способов способ дополнительно включает определение истории клеточной родословной. Согласно родственному аспекту настоящее изобретение относится к применению любого из приведенных выше способов для идентификации низкочастотных вариантов последовательности (например, вариантов, которые составляют  $\geq 0,01\%$  от общего числа последовательностей).

**[00101]** Согласно родственному аспекту настоящее изобретение относится к набору, содержащему полимеразу нуклеиновых кислот, один или более праймеров для амплификации, смесь нуклеотидов, содержащую один или более терминаторных нуклеотидов, и необязательно инструкции по применению. Согласно одному варианту осуществления наборов согласно настоящему изобретению полимеразы нуклеиновых кислот представляет собой замещающую цепи ДНК-полимеразу. Согласно одному варианту осуществления наборов согласно настоящему изобретению полимеразы нуклеиновых кислот выбрана из следующего: полимеразы бактериофага phi29 (Ф29), генетически модифицированная ДНК-полимераза phi29 (Ф29), фрагмент Кленова ДНК-

полимеразы I, ДНК-полимераза фага M2, ДНК-полимераза фага phiPRD1, ДНК-полимераза Bst, ДНК-полимераза большого фрагмента Bst, полимеразы экзо(-) Bst, ДНК-полимераза экзо(-) Bca, ДНК-полимераза Bsu, ДНК-полимераза Vent<sub>R</sub>, ДНК-полимераза Vent<sub>R</sub> (экзо-), ДНК-полимераза Deep Vent, ДНК-полимераза Deep Vent (экзо-), ДНК-полимераза IsoPol, ДНК-полимераза I, ДНК-полимераза Terminator, ДНК-полимераза T5, секвенза, ДНК-полимераза T7, T7-секвенза и ДНК-полимераза T4. Согласно одному варианту осуществления наборов согласно настоящему изобретению полимеразы нуклеиновых кислот характеризуется 3'→5' экзонуклеазной активностью, и терминаторные нуклеотиды ингибируют такую 3'→5' экзонуклеазную активность (например, нуклеотиды с модификацией альфа-группы [например, альфа-тио-дидезоксинуклеотиды], C3 спейсерные нуклеотиды, запертые нуклеиновые кислоты (LNA), инвертированные нуклеиновые кислоты, 2' фторнуклеотиды, 3' фосфорилированные нуклеотиды, 2'-О-метилмодифицированные нуклеотиды, транснуклеиновые кислоты). Согласно одному варианту осуществления наборов согласно настоящему изобретению полимеразы нуклеиновых кислот не характеризуется 3'→5' экзонуклеазной активностью (например, ДНК-полимераза Bst, полимеразы экзо(-) Bst, ДНК-полимераза экзо(-) Bca, ДНК-полимераза Bsu, ДНК-полимераза Vent<sub>R</sub> (экзо-), ДНК-полимераза Deep Vent (экзо-), ДНК-полимераза фрагмента Кленова (экзо-), ДНК-полимераза Terminator). Согласно одному конкретному варианту осуществления терминаторные нуклеотиды содержат модификации группы γ 3' углерода дезоксирибозы. Согласно одному конкретному варианту осуществления терминаторные нуклеотиды выбраны из следующего: 3' блокированный обратимый терминатор, содержащий нуклеотиды, 3' неблокированный обратимый терминатор, содержащий нуклеотиды, терминаторы, содержащие 2' модификации дезоксинуклеотидов, терминаторы, содержащие модификации азотистого основания дезоксинуклеотидов, и их комбинации. Согласно одному конкретному варианту осуществления терминаторные нуклеотиды выбраны из следующего: дидезоксинуклеотиды, инвертированные дидезоксинуклеотиды, 3' биотинилированные нуклеотиды, 3' аминонуклеотиды, 3'-фосфорилированные нуклеотиды, 3'-О-метилнуклеотиды, 3' углеродные спейсерные нуклеотиды, включая в себя 3' C3 спейсерные нуклеотиды, 3' C18 нуклеотиды, 3' гександиоловые спейсерные нуклеотиды, ациклонуклеотиды и их комбинации.

### **ПРИМЕРЫ**

**[00102]** Следующие ниже примеры представлены для более ясной иллюстрации принципа и осуществления на практике вариантов осуществления, раскрытых в настоящем документе, специалистам в настоящей области техники, и их не следует

истолковывать как ограничение объема любых заявленных вариантов осуществления. Если не указано иное, все части и проценты являются массовыми.

**[00103] ПРИМЕР 1: Направленная первичной матрицей амплификация (РТА)**

**[00104]** Хотя РТА можно использовать для амплификации любой нуклеиновой кислоты, она является особенно применимой для амплификации целого генома, поскольку позволяет охватывать больший процент клеточного генома более однородным и воспроизводимым образом и с более низкими показателями ошибок, чем используемые в настоящее время способы, такие как, например, амплификация с множественным замещением цепи (MDA), позволяя избежать таких недостатков используемых в настоящее время способов, как экспоненциальная амплификация в тех местах, где полимераза сначала удлиняет случайные праймеры, что приводит к случайному избыточному представлению локусов и аллелей и распространению мутации (см. фигуры 1А-1С).

**[00105] Клеточная культура**

**[00106]** Клетки NA12878 человека (Coriell Institute) поддерживали в среде RPMI с добавлением 15% FBS и 2 мМ L-глутамина и 100 единиц/мл пенициллина, 100 мкг/мл стрептомицина и 0,25 мкг/мл амфотерицина В (Gibco, Life Technologies). Клетки высевали с плотностью  $3,5 \times 10^5$  клеток/мл. Культуры разделяли каждые 3 дня и поддерживали во влажной камере при 37°C с 5% CO<sub>2</sub>.

**[00107] Выделение одиночных клеток и WGA**

**[00108]** После культивирования клеток NA12878 в течение минимум трех дней после посева при плотности  $3,5 \times 10^5$  клеток/мл 3 мл клеточной суспензии осаждали при 300×g в течение 10 минут. Затем среду отбрасывали и клетки промывали три раза 1 мл буфера для промывки клеток (1X PBS, содержащий 2% FBS без Mg<sup>2</sup> или Ca<sup>2</sup>), центрифугированных при 300×g, 200×g и, наконец, 100×g в течение 5 минут. Затем клетки ресуспендировали в 500 мкл буфера для промывания клеток. За этим следовало окрашивание с помощью 100 нМ кальцеина АМ (Molecular Probes) и 100 нг/мл йодида пропидия (PI; Sigma-Aldrich) для различения популяции живых клеток. Клетки загружали в проточный цитометр BD FACScan (FACSAria II) (BD Biosciences), который был тщательно очищен с помощью ELIMINase (Decon Labs) и откалиброван с использованием флуоресцентных шариков Accudrop (BD Biosciences) для сортировки клеток. Одиночную клетку из кальцеин АМ-положительной, PI-отрицательной фракции отсортировали в каждую лунку 96-луночного планшета, содержащего 3 мкл PBS (Qiagen, набор REPLI-g SC Kit) с 0,2% Tween 20 в клетках, которые должны были подвергнуться РТА (Sigma-Aldrich).

Несколько лунок намеренно оставили пустыми, чтобы использовать их в качестве контроля без матрицы (NTC). Сразу после сортировки планшеты кратковременно центрифугировали и помещали на лед. Затем клетки замораживали как минимум в течение ночи при  $-20^{\circ}\text{C}$ . На следующий день WGA-реакции собирали на установке для предварительной ПЦР, которая обеспечивает постоянное положительное давление воздуха, отфильтрованного HEPA, и которую обеззараживают УФ-светом в течение 30 минут перед каждым экспериментом.

**[00109]** MDA проводили с использованием набора REPLI-g Single Cell Kit (Qiagen) с модификациями, которые, как было показано ранее, улучшают однородность амплификации. В частности, устойчивые к экзонуклеазам случайные праймеры (ThermoFisher) добавляли в буфер D2 (REPLI-g Single Cell Kit, Qiagen) до конечной концентрации 125 мкМ в буфере D2. 4 мкл полученной смеси для лизиса/денатурирования добавляли в пробирки, содержащие одиночные клетки, перемешивали вихревым способом, кратковременно центрифугировали и инкубировали на льду в течение 10 минут. Лизаты клеток нейтрализовали добавлением 3 мкл стоп-раствора (REPLI-g Single Cell Kit, Qiagen), перемешивали встряхиванием, кратковременно центрифугировали и помещали при комнатной температуре. За этим следовало добавление 40 мкл смеси для амплификации перед инкубацией при  $30^{\circ}\text{C}$  в течение 8 часов, после чего амплификацию прекращали нагреванием до  $65^{\circ}\text{C}$  в течение 3 минут.

**[00110]** РТА проводили путем сначала дополнительного лизирования клеток после замораживания-оттаивания путем добавления 2 мкл предварительно охлажденного раствора смеси 1:1 5% Triton X-100 (Sigma-Aldrich) и 20 мг/мл протеиназы К (Promega). Затем клетки перемешивали вихревым способом и кратковременно центрифугировали перед тем, как поместить при  $40^{\circ}$  в течение 10 минут. Затем к лизированным клеткам добавляли 4 мкл буфера D2 (REPLI-g Single Cell Kit, Qiagen) и 1 мкл 500 мкМ случайного праймера, устойчивого к экзонуклеазам, для денатурирования ДНК перед перемешивания вихревым способом, центрифугирования и помещения при  $65^{\circ}$  в течение 15 минут. Затем добавляли 4 мкл стоп-раствора при комнатной температуре (REPLI-g Single Cell Kit, Qiagen) и образцы перемешивали вихревым способом и центрифугировали. 56 мкл смеси для амплификации (REPLI-g Single Cell Kit, Qiagen), которая содержала альфа-тио-ddNTP (альфа-тио-дидезоксинуклеотиды) в равных соотношениях при концентрации 1200 мкМ в конечной реакции амплификации. Затем образцы помещали при  $30^{\circ}\text{C}$  на 8 часов, после чего амплификацию прекращали нагреванием до  $65^{\circ}\text{C}$  в течение 3 минут.

**[00111]** После стадии амплификации ДНК из обеих реакций, MDA и РТА, очищали с использованием магнитных шариков AMPure XP (Beckman Coulter) при соотношении шариков к образцу 2:1, и выход измеряли с использованием набора Qubit

dsDNA HS Assay Kit с помощью флуориметра Qubit 3.0 согласно инструкции производителя (Life Technologies).

**[00112]**        Получение библиотеки

**[00113]**        Реакции с MDA привели к продукции 40 мкг амплифицированной ДНК. 1 мкг продукта фрагментировали в течение 30 минут согласно протоколу КАРА HyperPlus после добавления кондиционирующего раствора (КАРА Biosystems). Затем образцы прошли подготовку стандартной библиотеки с помощью 15 мкМ адапторов с двойным индексом (Integrated DNA Technologies) и 4 цикла ПЦР. Каждая реакция РТА генерировала от 40 до 60 нг материала, который использовали для получения библиотеки секвенирования ДНК целиком с использованием набора КАРА HyperPlus без фрагментации. При лигировании использовали 2,5 мкМ адапторы с UMI и двойными индексами (Integrated DNA Technologies), а в заключительной амплификации использовали 15 циклов ПЦР. Затем библиотеки очищали с использованием двустороннего SPRI с коэффициентами 0,65X и 0,55X для правосторонней и левосторонней селекции соответственно. Конечные библиотеки количественно определяли с использованием набора Qubit dsDNA BR Assay Kit и 2100 Bioanalyzer (Agilent Technologies) перед секвенированием на платформе Illumina NextSeq. Все платформы секвенирования Illumina, включая в себя NovaSeq, также совместимы с этим протоколом.

**[00114]**        Анализ данных

**[00115]**        Прочтения секвенирования демультиплексировали на основе штрих-кода клетки с использованием Bcl2fastq. Затем прочтения укорачивали с помощью trimmomatic, после чего проводили выравнивание по hg19 с помощью BWA. Прочтения подвергли параллельной маркировке с помощью Picard с последующим локальным выравниванием и повторной калибровкой базы с использованием GATK 4.0. Все файлы, используемые для расчета показателей качества, подвергали субдискретизации до двадцати миллионов прочтений с помощью Picard DownSampleSam. Количественные показатели качества получали из окончательного файла BAM с использованием qualimap, а также Picard AlignmentSummaryMetrics и CollectWgsMetrics. Общий охват генома также оценивали с помощью Preseq.

**[00116]**        Обнаружение вариантов

**[00117]**        Однонуклеотидные варианты и индели обнаруживали с использованием GATK UnifiedGenotyper из GATK 4.0. Стандартные критерии фильтрации с использованием практических рекомендаций GATK использовали для всех этапов процесса (<https://software.broadinstitute.org/gatk/best-practices/>). Варианты числа копий обнаруживали с использованием Control-FREEC (Boeva et al., Bioinformatics, 2012,

28(3):423-5). Структурные варианты также обнаруживали с помощью CREST (Wang et al., Nat Methods, 2011, 8(8):652-4).

**[00118]**            Результаты

Как показано на фиг. 3А и 3В, показатели картирования и показатели качества картирования амплификации только с дидезоксинуклеотидами («обратимые») составляют 15,0 +/- 2,2 и 0,8 +/- 0,08, соответственно, в то время как включение устойчивых к экзонуклеазам альфа-тио-дидезоксинуклеотидных терминаторов («необратимые») приводит к показателям картирования и показателям качества картирования 97,9 +/- 0,62 и 46,3 +/- 3,18 соответственно. Эксперименты также проводили с использованием обратимого ddNTP и различных концентраций терминаторов. (Фигура 2А, внизу)

**[00119]**            На фигурах 2В-2Е показаны сравнительные данные, полученные на одиночных клетках человека NA12878, которые подверглись MDA (в соответствии со способом Dong, X. et al., Nat Methods. 2017, 14(5):491-493) или РТА. В то время как оба протокола давали сравнимо низкие показатели дубликации ПЦР (MDA 1,26% +/- 0,52 по сравнению с РТА 1,84% +/- 0,99) и GC% (MDA 42,0 +/- 1,47 по сравнению с РТА 40,33 +/- 0,45), РТА продуцировала ампликоны меньшего размера. Процент прочтений, которые картировали и у которых показатели качества картирования были также значительно выше для РТА по сравнению с MDA (РТА 97,9 +/- 0,62 по сравнению с MDA 82,13 +/- 0,62 и РТА 46,3 +/- 3,18 по сравнению с MDA 43,2 +/- 4,21, соответственно. В целом, РТА производит более удобные, картируемые данные по сравнению с MDA. На фигуре 4А показано, что по сравнению с MDA, РТА значительно улучшила однородность амплификации с большей шириной охвата и меньшим количеством областей, где охват падает почти до 0. Использование РТА позволяет идентифицировать низкочастотные варианты последовательностей в популяции нуклеиновых кислот, включая в себя варианты, которые составляют  $\geq 0,01\%$  от общего числа последовательностей. РТА можно успешно использовать для амплификации генома одиночной клетки.

**[00120]**            **ПРИМЕР 2: Сравнительный анализ РТА**

**[00121]**            *Сопоставительный анализ РТА и поддержания клеток и выделения SCMDA*

**[00122]**            Лимфобластоидные клетки из проекта 1000 Genome Project NA12878 (Coriell Institute, Камден, Нью-Джерси, США) поддерживали в среде RPMI с добавлением 15% FBS, 2 mM L-глутамина, 100 единиц/мл пенициллина, 100 мкг/мл стрептомицина и 0,25 мкг/мл амфотерицина В). Клетки высевали с плотностью  $3,5 \times 10^5$  клеток/мл и разделяли каждые 3 дня. Их поддерживали во влажной камере при 37°C с 5% CO<sub>2</sub>. Перед выделением одиночных клеток 3 мл суспензии клеток, которые размножились за предыдущие 3 дня,

центрифугировали при  $300\times g$  в течение 10 минут. Осажденные клетки промывали три раза 1 мл буфера для промывки клеток (1X PBS, содержащий 2% FBS без  $Mg^{2+}$  или  $Ca^{2+}$ ), где их последовательно центрифугировали при  $300\times g$ ,  $200\times g$  и, наконец,  $100\times g$  в течение 5 минут для удаления мертвых клеток. Затем клетки ресуспендировали в 500 мкл буфера для промывания клеток с последующим окрашиванием 100 нМ кальцеина АМ и 100 нг/мл йодида пропидия (PI) для различения популяции живых клеток. Клетки загружали в проточный цитометр BD FACScan (FACSAria II), который был тщательно очищен с помощью ELIMINase и откалиброван с использованием флуоресцентных шариков Accudrop. Одиночную клетку из кальцеин-АМ-положительной, PI-отрицательной фракции сортировали в каждой лунке 96-луночного планшета, содержащего 3 мкл PBS с 0,2% твина 20. Несколько лунок намеренно оставляли пустыми, чтобы использовать их в качестве контроля без матрицы. Сразу после сортировки планшеты кратковременно центрифугировали и помещали на лед. Затем клетки замораживали как минимум в течение ночи при  $-80^{\circ}C$ .

**[00123]**            *Эксперименты PTA и SCMDA*

**[00124]**            Реакции WGA собирали на установке для предварительной ПЦР, которая обеспечивает постоянное положительное давление с помощью воздуха, отфильтрованного HEPA, и которую обеззараживают УФ-светом в течение 30 минут перед каждым экспериментом. MDA проводили в соответствии с методологией SCMDA с использованием набора REPLI-g Single Cell Kit в соответствии с опубликованным протоколом (Dong et al. Nat. Meth. 2017, 14, 491-493). В частности, устойчивые к экзонуклеазам случайные праймеры добавляли в конечной концентрации 12,5 мкМ в буфер для лизиса. 4 мкл полученной смеси для лизиса добавляли в пробирки, содержащие одиночные клетки, пипетировали три раза для перемешивания, быстро центрифугировали и инкубировали на льду в течение 10 минут. Лизаты клеток нейтрализовали добавлением 3 мкл буфера для гашения, перемешивали пипетированием 3 раза, кратковременно центрифугировали и помещали на лед. После этого добавляли 40 мкл смеси для амплификации перед инкубацией при  $30^{\circ}C$  в течение 8 часов, после чего амплификацию прекращали нагреванием до  $65^{\circ}C$  в течение 3 минут. PTA проводили путем сначала дополнительного лизиса клеток после замораживания-оттаивания путем добавления 2 мкл предварительно охлажденного раствора смеси 1:1 5% Triton X-100 и 20 мг/мл протеиназы К. Затем клетки перемешивали вихревым способом и кратковременно центрифугировали перед тем, как поместить при  $40^{\circ}C$  в течение 10 минут. Затем к лизированным клеткам добавляли 4 мкл денатурирующего буфера и 1 мкл 500 мкМ устойчивого к экзонуклеазам случайного праймера для денатурирования ДНК перед перемешивания

вихревым способом, центрифугированием и помещением при 65°C в течение 15 минут. Затем добавляли 4 мкл гасящего раствора при комнатной температуре, и образцы перемешивали вихревым способом и центрифугировали. 56 мкл смеси для амплификации, которая содержала альфа-тио-ddNTP в равных соотношениях при концентрации 1200 мкМ в конечной реакции амплификации. Затем образцы помещали при 30°C на 8 часов, после чего амплификацию прекращали нагреванием до 65°C в течение 3 минут. После амплификации SCMDA или РТА ДНК очищали с использованием магнитных шариков AMPure XP при соотношении шариков к образцу 2:1, и выход измеряли с помощью набора Qubit dsDNA HS Assay Kit с помощью флуорометра Qubit 3.0 в соответствии с инструкциями производителя. Эксперименты с РТА также проводили с использованием обратимых ddNTP и различных концентраций терминаторов. (Фигура 2А, сверху)

**[00125]**            *Получение библиотеки*

**[00126]**            1 мкг продукта SCMDA фрагментировали в течение 30 минут в соответствии с протоколом HyperPlus после добавления кондиционирующего раствора. Затем образцы подвергали подготовке стандартной библиотеки с использованием 15 мкМ уникальных адапторов с двойным индексом и 4 цикла ПЦР. Весь продукт каждой реакции РТА использовали для получения библиотеки секвенирования ДНК с использованием набора KAPA HyperPlus без фрагментации. При лигировании использовали 2,5 мкМ уникального адаптора с двойным индексом, а в окончательной амплификации использовали 15 циклов ПЦР. Затем библиотеки из SCMDA и РТА визуализировали на 1% Agarose E-Gel. Фрагменты размером 400-700 п.н. вырезали из геля и выделяли с помощью набора для восстановления ДНК геля. Конечные библиотеки количественно определяли с использованием набора Qubit dsDNA BR Assay Kit и Agilent 2100 Bioanalyzer перед секвенированием на NovaSeq 6000.

**[00127]**            *Анализ данных*

**[00128]**            Данные усекали с использованием trimmomatic, за которым следовало выравнивание по hg19 с использованием BWA. Прочтения подвергли параллельной маркировке с помощью Picard, после чего производили локальную корректировку и повторную калибровку базы с использованием практических рекомендаций GATK 3.5. Все файлы подвергали субдискретизации до указанного числа прочтений с помощью Picard DownSampleSam. Количественные показатели качества получали из окончательного bam-файла с использованием qualimap, а также Picard AlignmentMetricsSummary и CollectWgsMetrics. Строили кривые Лоренца и коэффициенты Джини рассчитывали с использованием htSeqTools. Обнаружение SNV выполняли с помощью UnifiedGenotyper, которые затем фильтровали с использованием стандартных



рекомендуемых критериев ( $QD < 2,0 \parallel FS > 60,0 \parallel MQ < 40,0 \parallel SOR > 4,0 \parallel MQRankSum < -12,5 \parallel ReadPosRankSum < -8,0$ ). Никакие области не исключали из анализа, и не выполняли никакую другую нормализацию данных или манипуляции. Количественные показатели секвенирования для исследованных способов приведены в таблице 1.

**[00129] Таблица 1: Сравнение количественных показателей секвенирования между исследованными способами.**

	PTA	Набор для MDA 2	PicoPlex	MALBAC	LIANTI	Набор для MDA 1	DOP PCR
% картирования генома	97	88	55	79	92	65	52
% извлечения генома (300M прочтений)	95	75	43	60	82	73	23
% CV охвата (300M прочтений)	0,8	1,8	3	2,5	1,1	2	3,5
Чувствительность SNV, % (300M прочтений)	76	50	15	34	49	46	5
Специфичность SNV, % (300M прочтений)	93	91	56	47	88	90	35

CV = коэффициент вариации; SNV = однонуклеотидная вариация; значения относятся к 15-кратному охвату.

**[00130]** *Ширина и однородность охвата генома*

**[00131]** Выполнили всесторонние сравнения PTA со всеми распространенными способами WGA одиночных клеток. Для этого PTA и улучшенную версию MDA, называемую MDA одиночных клеток ((Dong et al. Nat. Meth. 2017, 14, 491-493) (SCMDA), проводили на 10 клетках NA12878 каждую. Кроме того, эти результаты для клеток, которые подверглись амплификации с помощью DOP-PCR ((Zhang et al. PNAS 1992, 89, 5847-5851), MDA, набор 1 (Dean et al. PNAS 2002, 99, 5261-5266), MDA, набор 2, MALBAC (Zong et al. Science 2012, 338, 1622-1626), LIANTI (Chen et al., Science 2017, 356, 189-194) , или PicoPlex (Langmore, Pharmacogenomics 3, 557-560 (2002)) сравнивали с использованием данных, полученных в рамках исследования LIANTI.

**[00132]** Для нормализации по образцам необработанные данные из всех образцов выравнивали и подвергали предварительной обработке для обнаружения вариантов с использованием одного и того же конвейера. Затем bam-файлы подвергали субдискретизации до 300 миллионов прочтений каждый перед выполнением сравнений.

Важно отметить, что продукты РТА и SCMDA не подвергали скринингу перед выполнением дальнейших анализов, в то время как все другие способы подвергали скринингу на охват генома и однородность перед отбором клеток самого высокого качества, которые использовали в последующих анализах. Следует отметить, что SCMDA и РТА сравнивали с объемными диплоидными образцами NA12878, в то время как все другие способы сравнивали с объемными диплоидными фибробластами BJ1, которые использовали в исследовании LIANTI. Как видно на фигурах 3C-3F, у РТА был самый высокий процент прочтений, выровненных с геномом, а также самое высокое качество картирования. РТА, LIANTI и SCMDA имели сходное содержание GC, все из которых были ниже, чем у других способов. Показатели дубликации ПЦР были сходными для всех способов. Кроме того, способ РТА позволил более мелким матрицам, таким как митохондриальный геном, обеспечить более высокие показатели охвата (аналогичные более крупным каноническим хромосомам) по сравнению с другими исследованными способами (фиг. 3G).

**[00133]** Затем сравнивали широту и однородность охвата всех способов. Примеры графиков охвата хромосомы 1 показаны для SCMDA и РТА, где показано, что РТА имеет значительно улучшенную однородность охвата (фиг. 4B и 4C). Затем рассчитывали коэффициенты охвата для всех способов с использованием возрастающего числа прочтений. РТА приближается к двум совокупным образцам на каждой глубине, что является значительным улучшением по сравнению со всеми другими способами (фиг. 5A). Затем авторы настоящего изобретения использовали две стратегии для измерения однородности охвата. Первый подход заключался в вычислении коэффициента вариации охвата при увеличении глубины секвенирования, когда было обнаружено, что РТА более однородна, чем все другие способы (фиг. 5B). Вторая стратегия заключалась в вычислении кривых Лоренца для каждого субдискретизированного bam-файла, где снова было обнаружено, что РТА характеризуется наибольшей однородностью (фиг. 5C). Чтобы измерить воспроизводимость однородности амплификации, рассчитывались коэффициенты Джини для оценки отличия каждой реакции амплификации от идеальной однородности (de Bourcy et al., PloS one 9, e105585 (2014)). Опять же было показано, что РТА является воспроизводимо более однородной, чем другие способы (фиг. 5D).

**[00134]** *Чувствительность SNV*

**[00135]** Чтобы определить влияние этих различий на производительность способов амплификации на обнаружение SNV, показатели частоты обнаружения вариантов для каждого сравнивали с соответствующим совокупным образцом при увеличении глубины секвенирования. Для оценки чувствительности сравнивали процент вариантов,

обнаруженных в соответствующих совокупных образцах, которые подвергли субдискретизации до 650 миллионов прочтений, которые были обнаружены в каждой клетке на каждой глубине секвенирования (фиг. 5E). Улучшенные охват и однородность РТА привели к обнаружению на 45,6% больше вариантов по сравнению с набором для MDA 2, который являлся следующим по чувствительности способом. Исследование сайтов, названных гетерозиготными в совокупном образце, показало, что РТА значительно уменьшила аллельный перекося в этих гетерозиготных сайтах (фиг. 5F). Это открытие подтверждает утверждение, что РТА не только характеризуется более равномерной амплификацией по геному, но также более равномерно амплифицирует два аллеля в одной и той же клетке.

**[00136]**            *Специфичность SNV*

**[00137]**            Чтобы оценить специфичность обнаружений мутаций, варианты, обнаруженные в каждой одиночной клетке, не обнаруженные в соответствующем совокупном образце, считались ложноположительными. Низкотемпературный лизис SCMDA значительно снизил количество ложноположительных обнаружений вариантов (фиг. 5G). Способы с использованием термостабильных полимераз (MALBAC, PicoPlex и DOP-PCR) показали дальнейшее снижение специфичности обнаружения SNV с увеличением глубины секвенирования. Без ограничения какой-либо теорией это, вероятно, является результатом значительного увеличения частоты ошибок этих полимераз по сравнению с ДНК-полимеразой phi29. Кроме того, паттерны замены оснований, наблюдаемые в ложноположительных обнаружениях, также оказываются зависимыми от полимеразы (фиг. 5H). Как видно на фиг. 5G, модель подавления распространения ошибок в РТА поддерживается более низкой частотой ложноположительных обнаружений SNV в РТА по сравнению со стандартными протоколами MDA. Кроме того, РТА характеризуется самыми низкими частотами аллелей ложноположительных обнаружений вариантов, что снова согласуется с моделью подавления распространения ошибки с РТА (фиг. 5I).

**[00138]**            **ПРИМЕР 3: Прямое измерение мутагенности окружающей среды (DMEM)**

**[00139]**            РТА использовали для проведения нового анализа мутагенности, который обеспечивает основу для проведения исследований токсикогеномики человека с высоким разрешением на уровне всего генома. Предыдущие исследования, такие как тест Эймса, основаны на бактериальной генетике для проведения измерений, которые, как предполагается, являются репрезентативными для клеток человека, но дают лишь ограниченную информацию о количестве и паттернах мутаций, индуцированных в каждой подвергнутой воздействию клетке. Чтобы преодолеть эти ограничения, разработали

систему мутагенеза человека «прямое измерение мутагенности окружающей среды (DMEM)», в которой отдельные клетки человека подвергали воздействию соединения окружающей среды, выделяли в виде одиночных клеток и подвергали секвенированию одиночных клеток для выявления новых мутаций, индуцируемых в каждой клетке.

**[00140]** Клетки пуповинной крови, которые экспрессируют стволовой маркер/маркер предшественника CD34, подвергали воздействию возрастающих концентраций прямого мутагена N-этил-N-нитрозомочевина (ENU). Известно, что ENU характеризуется относительно низкой константой субстрата Суэйна-Скотта и, следовательно, было показано, что она преимущественно действует через двухступенчатый механизм SN1, который приводит к предпочтительному алкилированию O4-тимина, O2-тимина и O2-цитозина. Посредством ограниченного секвенирования целевых генов было показано, что ENU также отдает предпочтение изменениям Т на А (А на Т), Т на С (А на G) и С на Т (G на А) у мышей, что значительно отличается от паттерна, наблюдаемого у *E. coli*.

**[00141]** *Выделение и размножение клеток пуповинной крови для экспериментов по мутагенности*

**[00142]** ENU (CAS 759-73-9) и D-маннит (CAS 69-65-8) помещали в раствор при их максимальной растворимости. Свежую обработанную антикоагулянтом пуповинную кровь (CB) получали из банка пуповинной крови Сент-Луиса. CB разводили 1:2 PBS и мононуклеарные клетки (MNC) выделяли центрифугированием в градиенте плотности на Ficoll-Paque Plus в соответствии с инструкциями производителя. Затем MNC CB, экспрессирующие CD34, отбирали иммуномагнитно с использованием набора микрошариков CD34 человека и системы магнитной сортировки клеток (MACS) в соответствии с инструкциями производителя. Количество и жизнеспособность клеток оценивали с помощью счетчика клеток Luna FL. CD34 +клетки CB высевали с плотностью  $2,5 \times 10^4$  клеток/мл в StemSpan SFEM с добавлением 1X добавки для размножения CD34+, 100 единиц/мл пенициллина и 100 мкг/мл стрептомицина, где они размножались в течение 96 часов, прежде чем перейти к воздействию мутагена.

**[00143]** *Прямое измерение мутагенности окружающей среды (DMEM)*

**[00144]** Размноженные CD34+ клетки пуповинной крови культивировали в StemSpan SFEM с добавлением 1X добавки для размножения CD34+, 100 единиц/мл пенициллина и 100 мкг/мл стрептомицина. Клетки подвергали воздействию ENU в концентрациях 8,54, 85,4 и 854 мкМ, D-маннита при 1152,8 и 11528 мкМ, или 0,9% хлорида натрия (контроль носителем) в течение 40 часов. Суспензии одиночных клеток из клеток, обработанных лекарственным средством, и образцов контроля носителем собирали и окрашивали для определения жизнеспособности, как описано выше. Сортировку

одиночных клеток проводили, как описано выше. Выполняли РТА и готовили библиотеки с использованием упрощенного и улучшенного протокола в соответствии с общими методами описанных в настоящем документе способов и примером 2.

**[00145]**            *Анализ данных DMEM*

**[00146]**            Данные, полученные от клеток в экспериментах DMEM, отсекали с использованием Trimmomatic, выровнены по GRCh38 с использованием BWA и далее обработаны с использованием практических рекомендаций GATK 4.0.1 без отклонения от рекомендуемых параметров. Генотипирование выполняли с помощью HaplotypeCaller, где совместные генотипы снова фильтровали с использованием стандартных параметров. Вариант считался результатом мутагена только в том случае, если он имел показатель качества Phred не менее 100 и был обнаружен только в одной клетке, но не был обнаружен в совокупном образце. Тринуклеотидный контекст каждого SNV определяли путем извлечения окружающих оснований из эталонного генома с помощью программных средств. Количество мутаций и контекст были визуализированы с использованием ggplot2 и heatmap2 в R.

**[00147]**            Чтобы определить, были ли мутации обогащены сайтами гиперчувствительности к ДНКазе I (DHS) в клетках CD34 +, рассчитывали долю SNV в каждом образце, которые перекрываются с сайтами DHS из 10 наборов данных первичных клеток CD34 +, полученных Roadmap Epigenomics Project. Сайты DHS увеличивали на 2 нуклеосомы или на 340 оснований в любом направлении. Каждый набор данных DHS спаривали с образцом одиночной клетки, где авторы настоящего изобретения определили долю генома человека с как минимум 10-кратным охватом в той клетке, которая перекрыта с DHS, который сравнивали с долей SNV, которые были обнаружены на перекрытых сайтах DHS.

**[00148]**            *Результаты*

**[00149]**            В соответствии с этими исследованиями наблюдали дозозависимое увеличение количества мутаций каждой клетки, где аналогичное количество мутаций было обнаружено в самой низкой дозе ENU по сравнению с контролем носителем или токсичными дозами маннита (фиг. 12A). Также в соответствии с предыдущей работой на мышах с использованием ENU, наиболее распространенными мутациями являются Т на А (А на Т), Т на С (А на G) и С на Т (G на А). Также наблюдали другие три типа изменений оснований, хотя трансверсия С на G (G на С), по-видимому, встречается редко (фиг. 12B). Изучение тринуклеотидного контекста SNV показывает два различных паттерна (фиг. 12C). Первый паттерн заключается в том, что мутагенез цитозина, по-видимому, случается редко, когда за цитозином следует гуанин. Цитозин, за которым следует гуанин, обычно

метилован по пятому углеродному сайту в геномах человека, что является маркером гетерохроматина. Без ограничения какой-либо теорией, было выдвинуто предположение, что 5-метилцитозин не подвергается алкилированию с помощью ENU из-за недоступности в гетерохроматине или в результате неблагоприятных условий реакции с 5-метилцитозином по сравнению с цитозином. Чтобы проверить первую гипотезу, местоположения сайтов мутаций сравнивали с известными сайтами гиперчувствительности к ДНКазе I в клетках CD34 +, которые были каталогизированы Roadmap Epigenomics Project. Как видно на фиг. 12D, не наблюдалось обогащения цитозиновых вариантов в сайтах гиперчувствительности к ДНКазе I. Кроме того, в сайтах ДН не наблюдалось обогащения вариантами, ограниченными цитозинами (фиг. 12E). Кроме того, большинство тиминовых вариантов встречаются там, где аденин присутствует перед тиминном. Аннотации геномных признаков для вариантов существенно не отличались от аннотаций этих признаков в геноме (рис. 12F).

**[00150] ПРИМЕР 4: Массовое параллельное секвенирование ДНК одиночных клеток**

**[00151]** Используя РТА, устанавливают протокол массового параллельного секвенирования ДНК. Сначала к случайному праймеру добавляют штрих-код клетки. Используют две стратегии для минимизации любого смещения при амплификации, вызванного клеточным штрих-кодом: 1) удлинение размера случайного праймера и/или 2) создание праймера, который закольцован на себе, чтобы предотвратить связывание клеточного штрих-кода с матрицей (фиг. 10B). После того, как оптимальная стратегия праймера определена, вплоть до 384 отсортированных клеток масштабируют с помощью, например, устройства для манипуляций с жидкостями Mosquito HTS, которое может с высокой точностью дозировать даже вязкие жидкости объемом вплоть до 25 нл. Этот жидкостный манипулятор также снижает затраты на реагенты приблизительно в 50 раз за счет использования 1 мкл реакции РТА вместо стандартного реакционного объема, составляющего 50 мкл.

**[00152]** Протокол амплификации переводят в капли путем доставки праймера со штрих-кодом клетки в каплю. Необязательно используют твердые подложки, такие как шарики, созданные с использованием стратегии смешения-разделения. Подходящие шарики доступны, например, от ChemGenes. Олигонуклеотид в некоторых случаях содержит случайный праймер, клеточный штрих-код, уникальный молекулярный идентификатор и расщепляемую последовательность или спейсер для высвобождения олигонуклеотида после того, как шарик и клетка инкапсулируются в одной и той же капле. Во время этого процесса оптимизируют концентрации матрицы, праймера, dNTP, альфа-тио-ddNTP и полимеразы для объема в нижнем нанолитровом диапазоне в каплях.

Оптимизация в некоторых случаях включает в себя применение более крупных капель для увеличения реакционного объема. Как видно на фиг. 9, этот процесс требует двух последовательных реакций для лизиса клеток, за которыми следует WGA. Первую каплю, которая содержит лизированные клетки и шарик, объединяют со второй каплей с амплификационной смесью. В качестве альтернативы или в комбинации клетку инкапсулируют в шарик гидрогеля перед лизисом, а затем оба шарика можно добавить к капле масла. См. Lan, F. et al., *Nature Biotechnol.*, 2017, 35:640-646).

**[00153]** Дополнительные способы включают использование микролунок, которые в некоторых случаях захватывают 140000 одиночных клеток в реакционных камерах емкостью 20 пиколитров на устройстве размером с предметное стекло микроскопа 3 дюйма × 2 дюйма. Подобно способам на основе капель, эти лунки объединяют клетку с шариком, который содержит штрих-код клетки, что позволяет проводить массовую параллельную обработку. См. Gole et al., *Nature Biotechnol.*, 2013, 31:1126-1132).

**[00154] ПРИМЕР 5: Применение РТА при остром лимфобластном лейкозе у детей (ALL)**

**[00155]** Выполнили секвенирование экзома одиночных клеток отдельных лейкозных клеток, несущих транслокацию ETV6-RUNX1, при этом измерили приблизительно 200 кодирующих мутаций на клетку, только 25 из которых присутствовали в достаточном количестве клеток, чтобы их можно было обнаружить с помощью стандартного массового секвенирования у этого пациента. Затем нагрузку мутаций на клетку объединяли с другими известными характеристиками этого типа лейкоза, такими как частота связанных с репликацией мутаций (1 кодирующая мутация/300 делений клеток), время от начала до постановки диагноза (4,2 года) и размер популяции на момент постановки диагноза (100 миллиардов клеток) для создания *in silico* имитации развития заболевания. Неожиданно обнаружили, что даже при том, что считалось генетически простым раком, таким как ALL у детей, на момент постановки диагноза у этого пациента имеется около 330 миллионов клонов с различными профилями кодирующих мутаций. Интересно, что, как видно на фиг. 6B, только от одного до пяти наиболее распространенных клонов (прямоугольник C) выявляются с помощью стандартного массового секвенирования; существуют десятки миллионов клонов, которые состоят из небольшого количества клеток и, следовательно, с меньшей вероятностью будут иметь клиническое значение (прямоугольник A). Соответственно, предусмотрены способы усиления чувствительности обнаружения, чтобы можно было обнаруживать клоны, составляющие по меньшей мере 0,01% (1:10000) клеток (прямоугольник B), поскольку это страта, в которой находится наиболее устойчивое заболевание, которое вызывает рецидив.

**[00156]** Учитывая такое огромное генетическое разнообразие популяции, была выдвинута гипотеза, что существуют клоны, которые более устойчивы к лечению у данного пациента. Чтобы проверить эту гипотезу, образец помещают в культуру и лейкозные клетки подвергают воздействию увеличивающихся концентраций стандартных химиотерапевтических препаратов от ALL. Как видно на фигуре 7, в контрольных образцах и в образцах, получавших самую низкую дозу аспарагиназы, клон, несущий активирующую мутацию KRAS, продолжал размножаться. Однако этот клон оказался более чувствительным к преднизолону и даунорубину, тогда как другие ранее не обнаруживаемые клоны могли быть более четко обнаружены после лечения этими лекарственными средствами (фигура 7, пунктирная линия). Этот подход также использовал массовое секвенирование образцов, получивших лечение. Использование секвенирования ДНК одиночных клеток в некоторых случаях позволяет определять разнообразие и клонотипы растущих популяций.

**[00157]** *Создание каталога чувствительности к лекарственным средствам клонотипов ALL*

**[00158]** Как показано на фигуре 8, чтобы составить каталог лекарственной чувствительности клонотипов ALL, берут аликвоту диагностического образца и проводят секвенирование одиночных клеток 10000 клеток для определения численности каждого клонотипа. Параллельно диагностические лейкозные клетки подвергают воздействию стандартных лекарственных средств от ALL (винкристин, даунорубин, меркаптопурин, преднизолон и аспарагиназа), а также группы таргетных лекарственных средств (ибрутиниб, дасатаниб и руксолитиниб) *in vitro*. Отбирают живые клетки и выполняют секвенирование ДНК одиночной клетки по меньшей мере на 2500 клетках на каждое воздействие лекарственного средства. Наконец, образцы костного мозга от тех же пациентов после того, как они завершили 6 недель лечения, сортируют на наличие живых клеток остаточного прелейкоза и лейкоза с использованием установленных протоколов для исследований массового секвенирования. Затем РТА используют для выполнения секвенирования ДНК одиночной клетки десятков тысяч клеток масштабируемым, эффективным и экономичным способом, что позволяет достичь следующих целей.

**[00159]** *От клонотипов к каталогу лекарственной чувствительности*

После получения данных секвенирования устанавливают клонотипы каждой клетки. Для этого обнаруживают варианты и определяют клонотипы. Использование РТА ограничивает выпадение аллелей и смещение охвата, вносимую во время используемых в настоящее время способов WGA. Провели систематическое сравнение инструментов для обнаружения вариантов из одиночных клеток, подвергшихся MDA, и обнаружили, что



недавно разработанный инструмент Monovar характеризуется самой высокой чувствительностью и специфичностью (Zafar et al., Nature Methods, 2016, 13:505-507). После того, как были обнаружены варианты, определяют, имеют ли две клетки один и тот же клонотип, несмотря на то, что некоторые обнаружения вариантов отсутствуют из-за выпадения аллелей. Для этого можно использовать смешанную модель многомерных распределений Бернулли (Gawad et al., Proc. Natl. Acad. Sci. USA, 2014, 111(50):17947-52). После установления того, что клетки имеют один и тот же клонотип, определяют, какие варианты включить в каталог. Включены гены, которые соответствуют любому из следующих критериев: 1) они являются несинонимичными вариантами, обнаруженными в любой из горячих точек мутагенеза или вариантов с потерей функции (сдвиг рамки считывания, нонсенс, сплайсинг), которые встречаются в известном гене-супрессоре опухоли, идентифицированном в крупных проектах по секвенированию генома рака у детей; 2) это варианты, которые регулярно обнаруживаются в образцах рецидива рака; и 3) они являются рецидивирующими вариантами, которые подвергаются положительному отбору в текущих исследованиях массового секвенирования остаточной болезни, поскольку пациенты с ALL проходят 6 недель лечения. Если у клонов нет хотя бы двух вариантов, отвечающих этим критериям, они не включаются в каталог. По мере выявления большего числа генов, связанных с устойчивостью к лечению или рецидивом заболевания, клоны могут быть «спасены» и включены в каталог. Чтобы определить, подвергся ли клонотип положительному или отрицательному отбору между контролем и лечением лекарственным средством, используют точный тест Фишера для идентификации клонов, которые значительно отличаются от контроля. Клоны будут добавлены в каталог только в том случае, если будет показано, что по крайней мере две согласованные комбинации мутаций имеют одинаковую корреляцию с воздействием определенного лекарственного средства. Известные активирующие мутации в онкогенах или мутации с потерей функции в супрессорах опухолей в одном и том же гене будут считаться эквивалентными для клонов. Если клонотипы не совсем совпадают, общие мутации будут внесены в каталог. Например, если клонотип 1 представляет собой A+B+C, а клонотип 2 представляет собой B+C+D, клонотип B+C будет внесен в каталог. Если идентифицированы гены, которые регулярно мутируют в устойчивых клетках с ограниченным количеством сопутствующих мутаций, эти клоны могут быть свернуты в функционально эквивалентные клонотипы.

**[00160] ПРИМЕР 6: Способ РТА**

**[00161]** Способ РТА выполняют с использованием общих способов согласно примеру 1 с модификациями. Согласно одному варианту осуществления терминаторы заменены стандартными dNTP, а добавки используют для замедления удлинения во время

амплификации. Согласно другому варианту осуществления терминаторы заменяют стандартными dNTP, а полимеразу замещения цепи модифицируют, чтобы замедлить скорость ее удлинения. Согласно другому варианту осуществления терминаторы заменяют на dNTP, которые во время удлинения включаются медленнее, чем стандартные нуклеотиды, или которые приводят к более медленной реакции удлинения после включения, чем из матрицы, содержащей стандартные нуклеотиды. Такие dNTP с медленным включением в некоторых случаях являются устойчивыми к нуклеазам.

**[00162] ПРИМЕР 7: Способ шпильки или петли с терминаторами**

**[00163]** Образец необязательно лизируют (например, одиночную клетку), а матричную ДНК («целевая молекула нуклеиновой кислоты») образца подвергают квазислучайному праймированию и линейной амплификации. На стадии квазислучайного праймирования используют смесь терминаторов и dNTP. Праймеры предназначены для создания шпилечных или петлевых структур, которые являются менее эффективными матрицами для дальнейшей амплификации, чем исходная матричная ДНК образца. Это приводит к увеличению доли ампликонов, происходящих из исходной матрицы образца. Затем библиотеку ампликонов дополнительно амплифицируют со стадией экспоненциальной амплификации для создания библиотеки для секвенирования. В некоторых случаях лизис, линейная амплификация и экспоненциальная амплификация происходят в одном контейнере. Альтернативно или в комбинации терминаторы используют на стадии экспоненциальной амплификации. В некоторых случаях стандартные dNTP используют во время линейной амплификации, а терминаторы используют во время стадии экспоненциальной амплификации. Использование терминаторов приводит к уменьшению амплификации не исходной матрицы по сравнению с не терминаторными нуклеотидами.

**[00164] ПРИМЕР 8: Рекомбиназная полимеразная амплификация (RPA) с терминаторами**

**[00165]** Образец необязательно лизируют (например, одиночную клетку), и матричную ДНК образца подвергают действию реакционной смеси RPA (в отношении иллюстративной процедуры см. Daher et al., Clin. Chem. 2016, 62 (7), 947 -958), содержащей рекомбиназу, связывающий одноцепочечную ДНК белок, праймеры, полимеразу и смесь терминаторов и dNTP. Например, рекомбиназа представляет собой RecA, а белок, связывающий одноцепочечную ДНК, представляет собой SSB. В некоторых случаях рекомбиназа представляет собой T4 UvsX, а белок, связывающий одноцепочечную ДНК, представляет собой T4 gp32. Различные полимеразы включают в себя без ограничения полимеразу Sau или полимеразу Bsu. В некоторых случаях к реакционной смеси добавляют

дополнительные агенты, такие как полиэтиленгликоль или Carbowax20M. В некоторых случаях для амплификации матрицы образцов РНК добавляют обратную транскриптазу. В некоторых случаях используют полностью или частично рандомизированные праймеры. Ампликоны, генерируемые RPA, необязательно подвергают дополнительным стадиям, таким как лигирование с адапторами, экспоненциальная амплификация, секвенирование или любая их комбинация. Использование терминаторов приводит к уменьшению амплификации не исходной матрицы по сравнению с не терминаторными нуклеотидами.

**[00166] ПРИМЕР 9: Зависимая от хеликазы амплификация (HDA) с терминаторами**

**[00167]** Образец необязательно лизируют (например, одиночную клетку), и матричную ДНК образца подвергают воздействию реакционной смеси HDA (в отношении иллюстративной процедуры см. Yang et al., *Chembiochem* 2015, 16 (9), 1365-1370), содержащей хеликазу, полимеразу и смесь терминаторов и dNTP. Например, полимеразы представляют собой Bst2.0, GspM, GspM2.0, GspSSD или другую полимеразу, а хеликаза представляет собой термофильную хеликазу, Tte-UvrD или другую хеликазу. В некоторых случаях добавляют дополнительный белок, связывающий одноцепочечную ДНК. В некоторых случаях для амплификации матрицы образцов РНК добавляют обратную транскриптазу. В некоторых случаях используют полностью или частично рандомизированные праймеры. Ампликоны, генерируемые HDA, необязательно подвергают дополнительным стадиям, таким как лигирование с адапторами, экспоненциальная амплификация, секвенирование или любая их комбинация. Использование терминаторов приводит к уменьшению амплификации не исходной матрицы по сравнению с не терминаторными нуклеотидами.

**[00168]** Хотя в настоящем документе были показаны и описаны предпочтительные варианты осуществления настоящего изобретения, для специалистов в настоящей области техники будет очевидно, что такие варианты осуществления представлены только в качестве примера. Многочисленные вариации, изменения и замены теперь будут очевидны специалистам в настоящей области техники, не отклоняясь от настоящего изобретения. Следует понимать, что различные альтернативы описанным в настоящем документе вариантам осуществления изобретения можно использовать при практическом применении настоящего изобретения. Предполагается, что нижеследующая формула изобретения определяет объем настоящего изобретения и что способы и структуры в рамках этой формулы изобретения и их эквиваленты охвачены ею.

## **ФОРМУЛА ИЗОБРЕТЕНИЯ**

1. Композиция, содержащая:  
по меньшей мере одну целевую молекулу нуклеиновой кислоты и библиотеку ампликонов,  
где библиотека ампликонов содержит множество полинуклеотидов, полученных в результате амплификации по меньшей мере одной целевой молекулы нуклеиновой кислоты,  
где по меньшей мере некоторые из полинуклеотидов содержат терминаторный нуклеотид,  
где по меньшей мере 5% полинуклеотидов представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты.
2. Композиция по п. 1, в которой по меньшей мере 10% полинуклеотидов представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты.
3. Композиция по п. 1 или 2, в которой по меньшей мере 15% полинуклеотидов представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты.
4. Композиция по любому из пп. 1-3, в которой по меньшей мере 25% полинуклеотидов представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты.
5. Композиция по любому из пп. 1-4, в которой по меньшей мере 50% полинуклеотидов представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты.
6. Композиция по п. 1, в которой 5-50% полинуклеотидов представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты.
7. Композиция по п. 1, в которой 5-20% полинуклеотидов представляют собой прямые копии по меньшей мере одной целевой молекулы нуклеиновой кислоты.

8. Композиция по любому из пп. 1-7, в которой не более 50% кумулятивной доли полинуклеотидов содержат последовательности, составляющие по меньшей мере 80% кумулятивной доли последовательностей по меньшей мере одной целевой молекулы нуклеиновой кислоты.

9. Композиция по любому из пп. 1-8, в которой не более 50% кумулятивной доли множества полинуклеотидов содержат последовательности, составляющие по меньшей мере 85% кумулятивной доли целевых последовательностей нуклеиновых кислот.

10. Композиция по любому из пп. 1-9, в которой не более 50% кумулятивной доли множества полинуклеотидов содержат последовательности, составляющие по меньшей мере 90% кумулятивной доли целевых последовательностей нуклеиновых кислот.

11. Композиция по любому из пп. 1-9, в которой библиотека ампликонов характеризуется коэффициентом Джинни, составляющим не более 0,5.

12. Композиция по любому из пп. 1-9, в которой библиотека ампликонов характеризуется коэффициентом Джинни, составляющим не более 0,4.

13. Композиция по любому из пп. 1-12, в которой множество полинуклеотидов имеет длину между около 50 и около 2000 нуклеотидов.

14. Композиция по любому из пп. 1-13, в которой полинуклеотиды имеют длину между около 400 и около 600 нуклеотидов.

15. Композиция по любому из пп. 1-14, в которой количество полинуклеотидов составляет 100-5000.

16. Композиция по любому из пп. 1-14, в которой количество полинуклеотидов составляет 250-1250.

17. Композиция по любому из пп. 1-14, в которой количество полинуклеотидов составляет по меньшей мере 100.

18. Композиция по любому из пп. 1-14, в которой количество полинуклеотидов составляет по меньшей мере 500.

19. Композиция по любому из пп. 1-14, в которой количество полинуклеотидов составляет по меньшей мере 1000.

20. Композиция по любому из пп. 1-19, в которой по меньшей мере некоторые из полинуклеотидов содержат штрих-код.

21. Композиция по п. 16, в которой штрих-код содержит штрих-код клетки.

22. Композиция по п. 16 или 21, в которой штрих-код содержит штрих-код образца.

23. Композиция по любому из пп. 1-22, в которой по меньшей мере некоторые из полинуклеотидов содержат уникальный молекулярный идентификатор.

24. Композиция по любому из пп. 1-23, в которой множество полинуклеотидов содержит последовательности, по меньшей мере частично репрезентативные для генома.

25. Композиция по любому из пп. 1-23, в которой множество полинуклеотидов содержит последовательности, по меньшей мере частично репрезентативные по меньшей мере для двух геномов.

26. Композиция по любому из пп. 1-23, в которой множество полинуклеотидов содержит последовательности из кДНК.

27. Композиция по любому из пп. 1-26, в которой по меньшей мере 90% полинуклеотидов содержат терминаторный нуклеотид.

28. Композиция по любому из пп. 1-27, в которой по меньшей мере 98% полинуклеотидов содержат терминаторный нуклеотид.

29. Композиция по любому из пп. 1-28, в которой терминаторный нуклеотид прикреплен к 3' концу по меньшей мере некоторых полинуклеотидов.

30. Композиция по любому из пп. 1-29, в которой терминаторный нуклеотид выбран из группы, состоящей из следующего: нуклеотиды с модификацией альфа-группы, С3 спейсерные нуклеотиды, запертые нуклеиновые кислоты (LNA), инвертированные нуклеиновые кислоты, 2' фторнуклеотиды, 3' фосфорилированные нуклеотиды, 2'-О-метил-модифицированные нуклеотиды и транснуклеиновые кислоты.

31. Композиция по любому из пп. 1-30, в которой нуклеотиды с модификацией альфа-группы представляют собой альфа-тио-дидезоксинуклеотиды.

32. Композиция по любому из пп. 1-31, в которой терминаторный нуклеотид содержит модификации группы г 3' углерода дезоксирибозы.

33. Композиция по любому из пп. 1-32, в которой терминаторный нуклеотид выбран из группы, состоящей из следующего: 3' блокированный обратимый терминатор, содержащий нуклеотиды, 3' неблокированный обратимый терминатор, содержащий нуклеотиды, терминаторы, содержащие 2' модификации дезоксинуклеотидов, терминаторы, содержащие модификации азотистого основания дезоксинуклеотидов, и их комбинации.

34. Композиция по любому из пп. 1-32, в которой терминаторные нуклеотиды выбраны из группы, состоящей из следующего: дидезоксинуклеотиды, инвертированные дидезоксинуклеотиды, 3' биотинилированные нуклеотиды, 3' аминонуклеотиды, 3'-фосфорилированные нуклеотиды, 3'-О-метилнуклеотиды, 3' углеродные спейсерные нуклеотиды, включая в себя 3' С3 спейсерные нуклеотиды, 3' С18 нуклеотиды, 3' гександиоловые спейсерные нуклеотиды, ациклонуклеотиды и их комбинации.

35. Композиция по любому из пп. 1-34, причем композиция содержится в капле.

36. Композиция, содержащая:

по меньшей мере одну целевую молекулу нуклеиновой кислоты и библиотеку ампликонов,

где библиотека ампликонов содержит множество полинуклеотидов, полученных в результате амплификации по меньшей мере одной целевой молекулы нуклеиновой кислоты,

где по меньшей мере некоторые из полинуклеотидов содержат терминаторный нуклеотид,

где не более 50% кумулятивной доли полинуклеотидов содержат последовательности, составляющие по меньшей мере 80% кумулятивной доли последовательностей по меньшей мере одной целевой молекулы нуклеиновой кислоты.

37. Композиция по п. 36, в которой не более 50% кумулятивной доли множества полинуклеотидов содержат последовательности, составляющие по меньшей мере 85% кумулятивной доли целевых последовательностей нуклеиновых кислот.

38. Композиция по п. 36 или 37, в которой не более 50% кумулятивной доли множества полинуклеотидов содержат последовательности, составляющие по меньшей мере 90% кумулятивной доли целевых последовательностей нуклеиновых кислот.

39. Композиция по любому из пп. 36-38, в которой множество полинуклеотидов имеют длину между около 50 и около 2000 нуклеотидов.

40. Композиция по любому из пп. 36-39, в которой полинуклеотиды имеют длину между около 400 и около 600 нуклеотидов.

41. Композиция по любому из пп. 36-40, в которой по меньшей мере некоторые из полинуклеотидов содержат штрих-код.

42. Композиция по п. 41, в которой штрих-код содержит штрих-код клетки.

43. Композиция по п. 41 или 42, в которой штрих-код содержит штрих-код образца.

44. Композиция по любому из пп. 36-43, в которой по меньшей мере некоторые из полинуклеотидов содержат уникальный молекулярный идентификатор.

45. Композиция по любому из пп. 36-44, в которой множество полинуклеотидов содержит последовательности, по меньшей мере частично репрезентативные для генома.



46. Композиция по любому из пп. 36-44, в которой множество полинуклеотидов содержит последовательности, по меньшей мере частично репрезентативные по меньшей мере для двух геномов.

47. Композиция по любому из пп. 36-44, в которой множество полинуклеотидов содержит последовательности из кДНК.

48. Композиция по любому из пп. 36-47, в которой по меньшей мере 90% полинуклеотидов содержат терминаторный нуклеотид.

49. Композиция по любому из пп. 36-48, в которой по меньшей мере 98% полинуклеотидов содержат терминаторный нуклеотид.

50. Композиция по любому из пп. 36-49, в которой терминаторный нуклеотид прикреплен к 3' концу по меньшей мере некоторых полинуклеотидов.

51. Композиция по любому из пп. 36-50, в которой терминаторный нуклеотид выбран из группы, состоящей из следующего: нуклеотиды с модификацией альфа-группы, С3 спейсерные нуклеотиды, запертые нуклеиновые кислоты (LNA), инвертированные нуклеиновые кислоты, 2' фторнуклеотиды, 3' фосфорилированные нуклеотиды, 2'-О-метил-модифицированные нуклеотиды и транснуклеиновые кислоты.

52. Композиция по любому из пп. 36-51, в которой нуклеотиды с модификацией альфа-группы представляют собой альфа-тио-дидезоксинуклеотиды.

53. Композиция по любому из пп. 36-52, в которой терминаторный нуклеотид содержит модификации группы г 3' углерода дезоксирибозы.

54. Композиция по любому из пп. 36-53, в которой терминаторный нуклеотид выбран из группы, состоящей из следующего: 3' блокированный обратимый терминатор, содержащий нуклеотиды, 3' неблокированный обратимый терминатор, содержащий нуклеотиды, терминаторы, содержащие 2' модификации дезоксинуклеотидов, терминаторы, содержащие модификации азотистого основания дезоксинуклеотидов, и их комбинации.

55. Композиция по любому из пп. 36-53, в которой терминаторные нуклеотиды выбраны из группы, состоящей из следующего: дидезоксинуклеотиды, инвертированные дидезоксинуклеотиды, 3' биотинилированные нуклеотиды, 3' аминонуклеотиды, 3'-фосфорилированные нуклеотиды, 3'-О-метилнуклеотиды, 3' углеродные спейсерные нуклеотиды, включая в себя 3' С3 спейсерные нуклеотиды, 3' С18 нуклеотиды, 3' гександиоловые спейсерные нуклеотиды, ациклонуклеотиды и их комбинации.

56. Композиция по любому из пп. 36-54, причем композиция содержится в капле.

57. Способ амплификации целевой молекулы нуклеиновой кислоты, включающий:

а. приведение в контакт образца, содержащего целевую молекулу нуклеиновой кислоты, по меньшей мере одного праймера для амплификации, по меньшей мере одной полимеразы нуклеиновых кислот и смеси нуклеотидов, где смесь нуклеотидов содержит по меньшей мере один терминаторный нуклеотид, который прекращает репликацию нуклеиновой кислоты полимеразой, и

б. амплификация целевой молекулы нуклеиновой кислоты для создания множества терминированных продуктов амплификации, где репликация происходит путем репликации с замещением цепи.

58. Способ по п. 37, в котором амплификацию проводят при по существу изотермических условиях.

59. Способ по п. 37 или 58, в котором амплификацию проводят при условиях, в которых температура варьируется не более чем на 10° С.

60. Способ по любому из пп. 37-59, в котором амплификацию проводят при условиях, в которых температура варьируется не более чем на 5° С.

61. Способ по любому из пп. 37-60, в котором полимеразы нуклеиновых кислот представляет собой ДНК-полимеразу.

62. Способ по п. 61, в котором ДНК-полимераза представляет собой замещающую цепи ДНК-полимеразу.

63. Способ по любому из пп. 37-62, в котором полимеразы нуклеиновых кислот представляет собой следующее: полимеразы бактериофага  $\phi$ 29 ( $\Phi$ 29), генетически модифицированная ДНК-полимераза  $\phi$ 29 ( $\Phi$ 29), фрагмент Кленова ДНК-полимеразы I, ДНК-полимераза фага M2, ДНК-полимераза фага  $\phi$ PRD1, ДНК-полимераза Bst, ДНК-полимераза большого фрагмента Bst, полимеразы экзо(-) Bst, ДНК-полимераза экзо(-) Bca, ДНК-полимераза Bsu, ДНК-полимераза Vent<sub>R</sub>, ДНК-полимераза Vent<sub>R</sub> (экзо-), ДНК-полимераза Deep Vent, ДНК-полимераза Deep Vent (экзо-), ДНК-полимераза IsoPol, ДНК-полимераза I, ДНК-полимераза Terminator, ДНК-полимераза T5, секвенза, ДНК-полимераза T7, T7-секвенза или ДНК-полимераза T4.

64. Способ по любому из пп. 37-63, в котором полимеразы нуклеиновых кислот характеризуется 3'→5' экзонуклеазной активностью, и по меньшей мере один терминаторный нуклеотид ингибирует 3'→5' экзонуклеазную активность.

65. Способ по любому из пп. 37-63, в котором полимеразы нуклеиновых кислот не характеризуется 3'→5' экзонуклеазной активностью.

66. Способ по любому из пп. 37-63, в котором полимеразы представляет собой следующее: ДНК-полимераза Bst, полимеразы экзо(-) Bst, ДНК-полимераза экзо(-) Bca, ДНК-полимераза Bsu, ДНК-полимераза Vent<sub>R</sub> (экзо-), ДНК-полимераза Deep Vent (экзо-), ДНК-полимераза фрагмента Кленова (экзо-) или ДНК-полимераза Terminator.

67. Способ по любому из пп. 37-66, в котором по меньшей мере один терминаторный нуклеотид содержит модификации группы г 3' углерода дезоксирибозы.

68. Способ по любому из пп. 37-67, в котором по меньшей мере один терминаторный нуклеотид выбирают из группы, состоящей из следующего: 3' блокированный обратимый терминатор, содержащий нуклеотиды, 3' неблокированный обратимый терминатор, содержащий нуклеотиды, терминаторы, содержащие 2' модификации дезоксинуклеотидов, терминаторы, содержащие модификации азотистого основания дезоксинуклеотидов, и их комбинации.

69. Способ по любому из пп. 37-68, в котором по меньшей мере один терминаторный нуклеотид выбирают из группы, состоящей из следующего: дидезоксинуклеотиды, инвертированные дидезоксинуклеотиды, 3' биотинилированные нуклеотиды, 3'

аминонуклеотиды, 3'-фосфорилированные нуклеотиды, 3'-О-метилнуклеотиды, 3' углеродные спейсерные нуклеотиды, включая в себя 3' C3 спейсерные нуклеотиды, 3' C18 нуклеотиды, 3' гександиоловые спейсерные нуклеотиды, ациклонуклеотиды и их комбинации.

70. Способ по любому из пп. 37-69, в котором по меньшей мере один терминаторный нуклеотид выбирают из группы, состоящей из следующего: нуклеотиды с модификацией альфа-группы, C3 спейсерные нуклеотиды, запертые нуклеиновые кислоты (LNA), инвертированные нуклеиновые кислоты, 2' фторнуклеотиды, 3' фосфорилированные нуклеотиды, 2'-О-метил-модифицированные нуклеотиды и транснуклеиновые кислоты.

71. Способ по любому из пп. 37-70, в котором нуклеотиды с модификацией альфа-группы представляют собой альфа-тио-дидезоксинуклеотиды.

72. Способ по любому из пп. 37-71, в котором праймеры для амплификации имеют длину 4 - 70 нуклеотидов.

73. Способ по любому из пп. 37-72, в котором по меньшей мере один праймер для амплификации имеет длину 4 - 20 нуклеотидов.

74. Способ по любому из пп. 37-73, в котором по меньшей мере один праймер для амплификации содержит рандомизированную область.

75. Способ по п. 74, в котором рандомизированная область имеет длину 4 - 20 нуклеотидов.

76. Способ по п. 74 или 75, в котором рандомизированная область имеет длину 8 - 15 нуклеотидов.

77. Способ по любому из пп. 37-76, в котором продукты амплификации имеют длину между около 50 и около 2000 нуклеотидов.

78. Способ по любому из пп. 37-77, в котором продукты амплификации имеют длину между около 200 и около 1000 нуклеотидов.

79. Способ по любому из пп. 37-78, в способ дополнительно включает дополнительную стадию амплификации с использованием ПЦР.

80. Способ секвенирования целевой молекулы нуклеиновой кислоты, включающий:

а. приведение в контакт образца, содержащего целевую молекулу нуклеиновой кислоты, по меньшей мере одного праймера для амплификации, по меньшей мере одной полимеразы нуклеиновых кислот и смеси нуклеотидов, где смесь нуклеотидов содержит по меньшей мере один терминаторный нуклеотид, который прекращает репликацию нуклеиновой кислоты полимеразой, и

б. амплификация целевой молекулы нуклеиновой кислоты для создания множества терминированных продуктов амплификации, где репликация происходит путем репликации с замещением цепи;

с. удаление по меньшей мере одного терминаторного нуклеотида из терминированных продуктов амплификации;

д. лигирование молекул, полученных на стадии (с), с адапторами, тем самым создавая библиотеку продуктов амплификации; и

е. секвенирование библиотеки продуктов амплификации.

81. Способ по п. 80, дополнительно включающий репарацию концов и добавление А-хвостов к 3'-концу ДНК с тупыми концами.

82. Способ по любому из пп. 80 или 81, в котором целевая нуклеиновая кислота представляет собой ДНК.

83. Способ по п. 82, в котором ДНК представляет собой кДНК.

84. Способ по п. 82, в котором ДНК представляет собой геномную ДНК.

85. Способ по любому из пп. 80-84, в котором по меньшей мере один праймер для амплификации содержит два или больше праймеров.

86. Способ по любому из пп. 80-85, в котором по меньшей мере один праймер для амплификации представляет собой случайный праймер.

87. Способ по любому из пп. 80-86, в котором по меньшей мере один праймер для амплификации содержит штрих-код.

88. Способ по п. 87, в котором штрих-код содержит штрих-код клетки.

89. Способ по п. 87 или 88, в котором штрих-код содержит штрих-код образца.

90. Способ по любому из пп. 80-89, причем способ дополнительно включает дополнительную стадию амплификации с использованием ПЦР.

91. Способ по любому из пп. 80-90, в котором праймеры для амплификации содержат уникальный молекулярный идентификатор (UMI).

92. Способ по любому из пп. 80-91, причем способ дополнительно включает денатурирование целевой нуклеиновой кислоты или геномной ДНК перед начальным отжигом праймера.

93. Способ по п. 92, в котором денатурацию проводят в щелочных условиях с последующей нейтрализацией.

94. Способ по любому из пп. 80-93, в котором образец, праймеры для амплификации, полимеразы нуклеиновых кислот и смесь нуклеотидов содержатся в микрожидкостном устройстве.

95. Способ по любому из пп. 80-94, в котором образец, праймеры для амплификации, полимеразы нуклеиновых кислот и смесь нуклеотидов содержатся в капле.

96. Способ по любому из пп. 80-95, в котором образец выбран из следующего: образцы ткани(ей), клетки, образцы биологической жидкости, образцы костного мозга, образцы спермы, биопсийные образцы, образцы рака, образцы опухоли, образцы клеточного лизата, криминалистические образцы, археологические образцы, палеонтологические образцы, образцы инфекции, производственные образцы, целые растения, части растения, образцы микробиоты, вирусные препараты, образцы почвы, образцы морской воды, образцы пресной воды, бытовые или промышленные образцы и их комбинации и изоляты.

97. Способ по п. 96, в котором биологические жидкости выбраны из следующего: кровь, моча, слюна, лимфатическая жидкость, спинномозговая жидкость (CSF), амниотическая жидкость, плевральная жидкость, перикардальная жидкость, асцит и водянистая влага.

98. Способ секвенирования одиночной клетки, включающий:

- a. получение клеточного лизата из одиночной клетки;
- b. приведение в контакт клеточного лизата по меньшей мере с одним праймером для амплификации, по меньшей мере одной полимеразой нуклеиновых кислот и смесью нуклеотидов, где смесь нуклеотидов содержит по меньшей мере один терминаторный нуклеотид, который прекращает репликацию нуклеиновой кислоты полимеразой, и
- c. амплификация целевой молекулы нуклеиновой кислоты для создания множества терминированных продуктов амплификации, где репликация происходит путем репликации с замещением цепи;
- d. удаление по меньшей мере одного терминаторного нуклеотида из терминированных продуктов амплификации;
- e. лигирование молекул, полученных на стадии (d), с адапторами, тем самым создавая библиотеку продуктов амплификации; и
- f. секвенирование библиотеки продуктов амплификации.

99. Способ по п. 98, в котором клеточный лизис сопровождается протеолизом.

100. Способ по п. 98 или 99, в котором образец, по меньшей мере один праймер для амплификации, полимеразы нуклеиновых кислот и смесь нуклеотидов содержатся в микрожидкостном устройстве.

101. Способ по любому из пп. 98-100, в котором образец, по меньшей мере один праймер для амплификации, полимеразы нуклеиновых кислот и смесь нуклеотидов содержатся в капле.

102. Способ по любому из пп. 98-101, в котором по меньшей мере один праймер для амплификации прикреплен к твердой подложке.

103. Способ по п. 102, в котором твердая подложка представляет собой шарик.

104. Способ по пп. 102 или 103, в котором по меньшей мере один праймер для амплификации прикреплен к твердой подложке посредством расщепляемого линкера.

105. Способ по любому из пп. 98-104, в котором по меньшей мере один праймер для амплификации содержит штрих-код.

106. Способ по п. 104, причем способ дополнительно включает расщепление расщепляемого линкера перед амплификацией.

107. Способ по любому из пп. 98-106, причем способ дополнительно включает дополнительную стадию амплификации с использованием ПЦР.

108. Способ по любому из пп. 98-107, в котором клетку выбирают из следующего: клетка животного, растительная клетка, грибковая клетка, бактериальная клетка и клетка простейшего.

109. Способ по п. 108, в котором клетка животного представляет собой клетку человека.

110. Способ по любому из пп. 98-109, в котором клетку выбирают из следующего: клетка из предимплантационного эмбриона, стволовая клетка, фетальная клетка, опухолевая клетка, предполагаемая раковая клетка, раковая клетка, клетка, подвергнутая процедуре редактирования генов, клетка из патогенного организма, клетка, полученная из криминалистического образца, клетка, полученная из археологического образца, и клетка, полученная из палеонтологического образца.

111. Способ по п. 110, в котором клетка предимплантационного эмбриона представляет собой бластомер.

112. Способ по п. 111, в котором бластомер получают из эмбриона восьмиклеточной стадии, полученного с помощью *in vitro* оплодотворения.



113. Способ по любому из пп. 110-112, дополнительно включающий определение наличия предрасполагающих к заболеванию вариантов зародышевой линии или соматических вариантов в эмбриональной клетке.

114. Способ по п. 110, в котором патогенный организм представляет собой бактерию, гриб или простейшее.

115. Способ по п. 114, в котором клетку, полученную из патогенного организма, получают из жидкости, взятой у пациента, образца микробиоты или постоянного изделия медицинского назначения.

116. Способ по п. 114 или 115, дополнительно включающий стадию определения идентичности патогенного организма.

117. Способ по любому из пп. 110-116, дополнительно включающий определение наличия генетических вариантов, ответственных за устойчивость патогенного организма к лечению.

118. Способ по любому из пп. 98-110, в котором клетка представляет собой опухолевую клетку, предполагаемую раковую клетку или раковую клетку.

119. Способ по любому из пп. 98-110 или 118, дополнительно включающий определение наличия одной или более диагностических или прогностических мутаций.

120. Способ по любому из пп. 98-110, 118, или 119, дополнительно включающий определение наличия вариантов зародышевой линии или соматических, ответственных за устойчивость к лечению.

121. Способ по любому из пп. 98-110, в котором клетка представляет собой клетку, подвергнутую процедуре редактирования генов.

122. Способ по п. 121, дополнительно включающий определение наличия незапланированных мутаций, вызванных процессом редактирования генов.

123. Способ по любому из пп. 98-122, дополнительно включающий определение истории клеточной родословной.

124. Применение способа по любому из пп. 37-123 для идентификации низкочастотных вариантов последовательности.

125. Применение по п. 124, в котором низкочастотные варианты последовательности составляют  $\geq 0,01\%$  от общего числа последовательностей.

126. Применение по п. 124, в котором низкочастотные варианты последовательности составляют  $\geq 0,05\%$  от общего числа последовательностей.

127. Применение по п. 124, в котором низкочастотные варианты последовательности составляют  $\geq 0,10\%$  от общего числа последовательностей.

128. Способ определения мутагенности условия окружающей среды, включающий:

- a. воздействие на клетки условия окружающей среды;
- b. выделение одиночных клеток из популяции;
- c. получение клеточного лизата из одиночной клетки;
- d. приведение в контакт клеточного лизата по меньшей мере с одним праймером для амплификации, по меньшей мере одной полимеразой нуклеиновых кислот и смесью нуклеотидов, где смесь нуклеотидов содержит по меньшей мере один терминаторный нуклеотид, который прекращает репликацию нуклеиновой кислоты полимеразой, и
- e. амплификация целевой молекулы нуклеиновой кислоты для создания множества терминированных продуктов амплификации, где репликация происходит путем репликации с замещением цепи;
- f. удаление по меньшей мере одного терминаторного нуклеотида из терминированных продуктов амплификации;
- g. лигирование молекул, полученных на стадии (f), с адапторами, тем самым создавая библиотеку продуктов амплификации; и
- h. секвенирование библиотеки продуктов амплификации, и
- i. сравнение последовательностей продуктов амплификации по меньшей мере с одной эталонной последовательностью для идентификации мутаций.

129. Способ по п. 128, в котором одиночная клетка представляет собой клетку человека.

130. Способ по п. 128 или 129, в котором условие окружающей среды включает в себя химическое вещество.

131. Способ по п. 128 или 129, в котором условие окружающей среды включает в себя облучение.

132. Способ по п. 128 или 129, в котором условие окружающей среды включает в себя ультрафиолетовое излучение.

133. Способ по любому из пп. 128-132, в котором одиночные клетки происходят из печени, кожи, почки, крови или легкого.

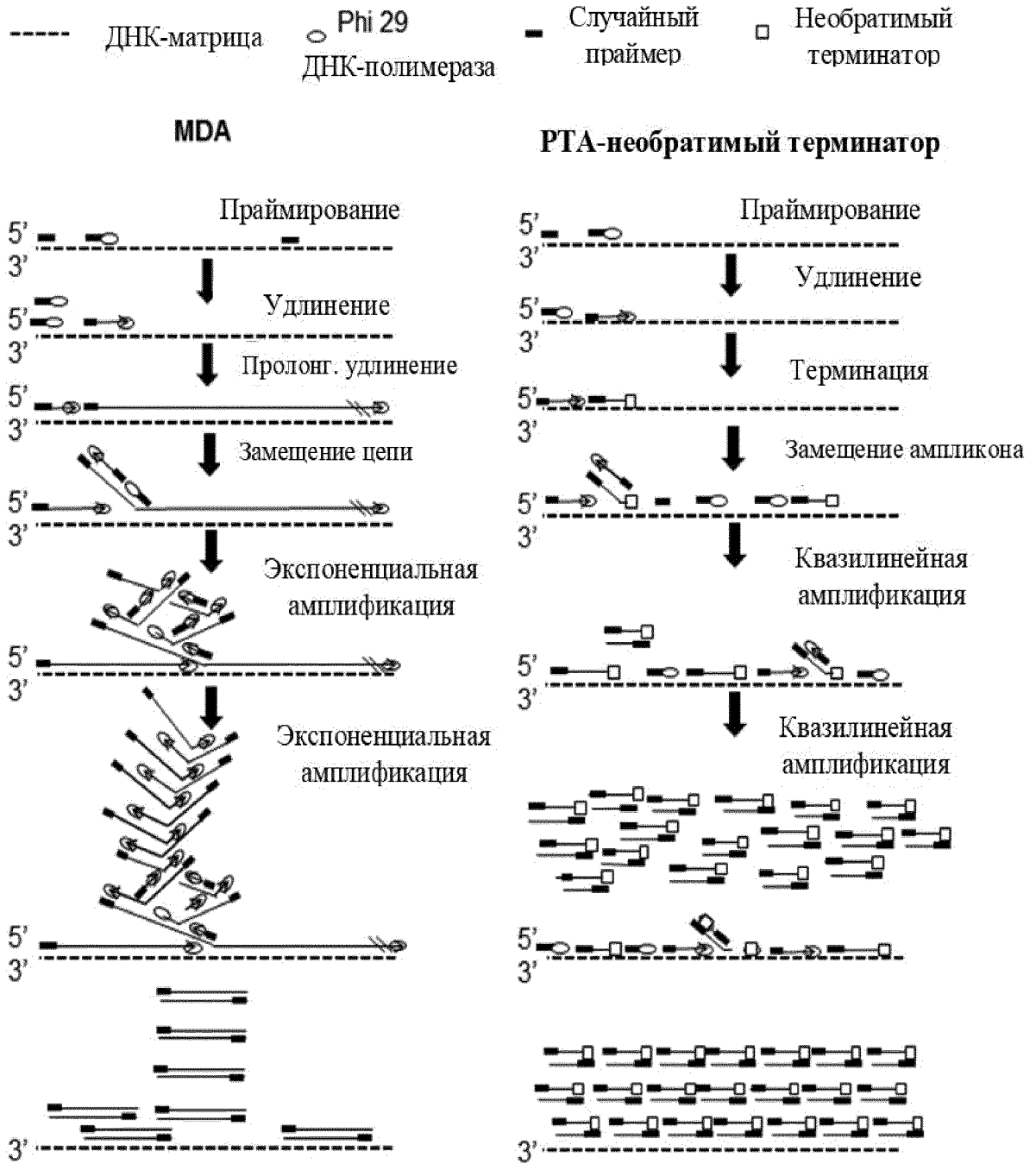
134. Способ по любому из пп. 128-133, в котором по меньшей мере некоторые из продуктов амплификации содержат штрих-код.

135. Способ по п. 134, в котором штрих-код содержит штрих-код клетки.

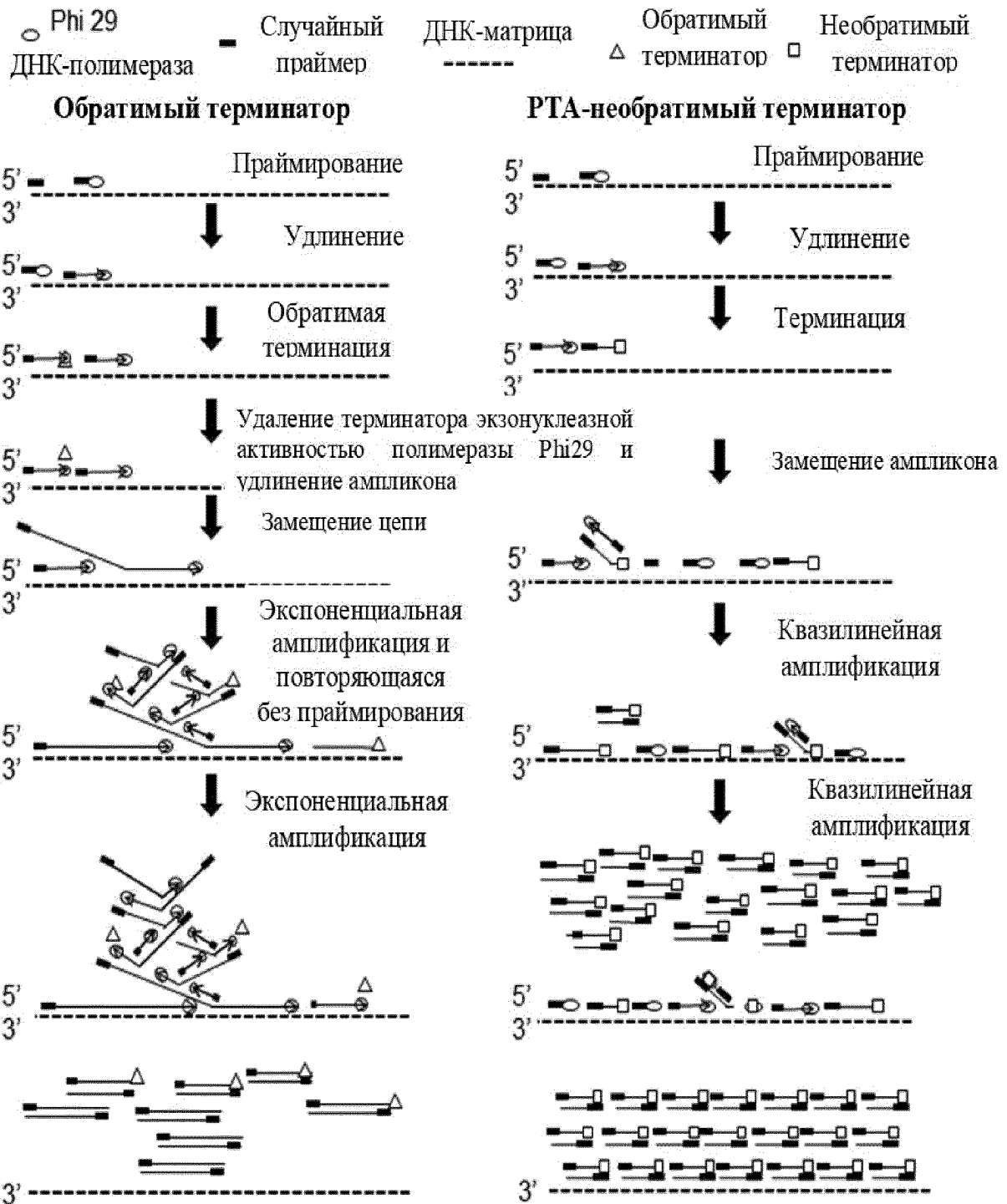
136. Способ по п. 134 или 135, в котором штрих-код содержит штрих-код образца.

137. Способ по любому из пп. 128-136, в котором по меньшей мере некоторые из праймеров для амплификации содержат уникальный молекулярный идентификатор (UMI).

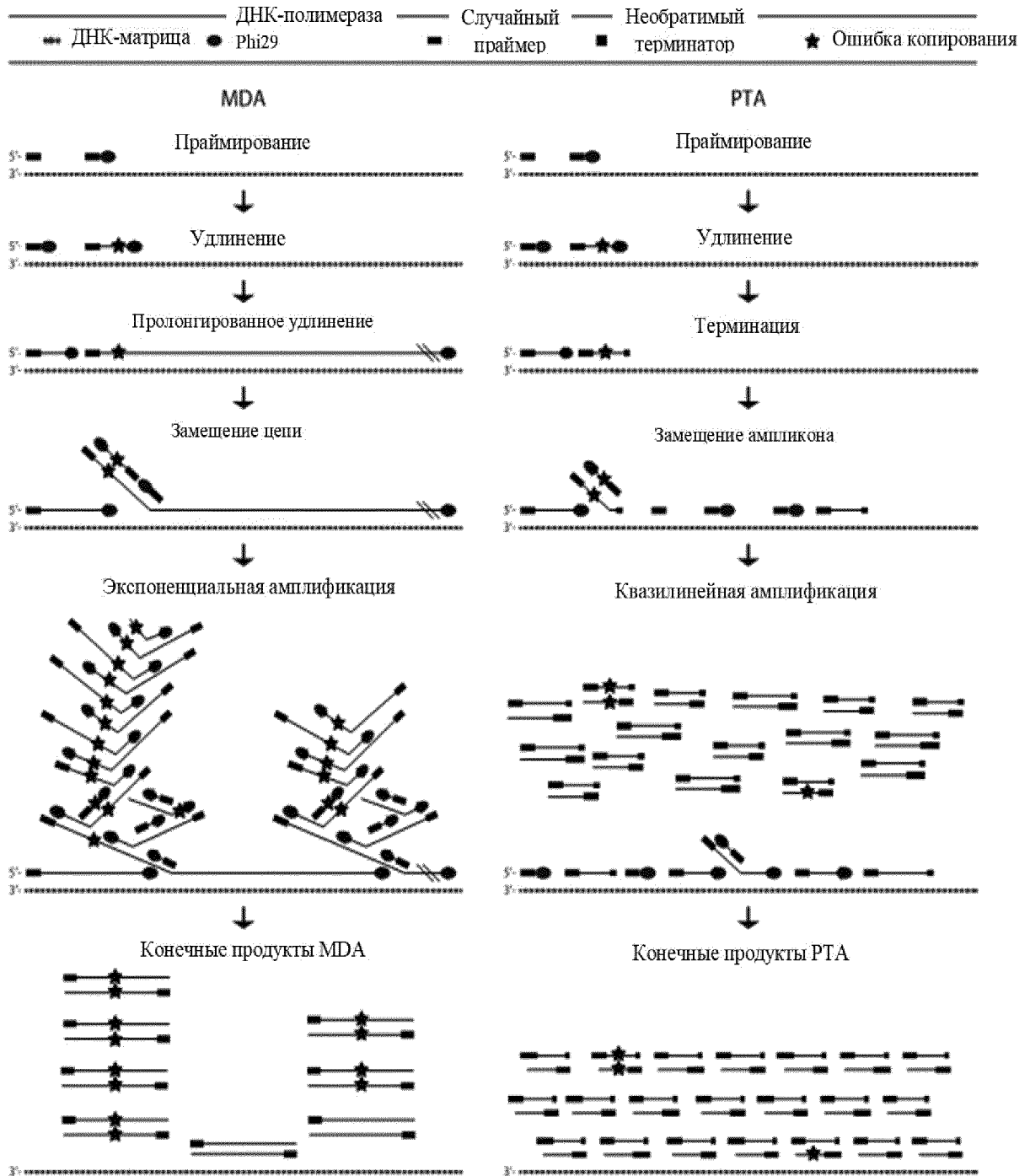
138. Способ по любому из пп. 128-137, причем способ дополнительно включает дополнительную стадию амплификации с использованием ПЦР.



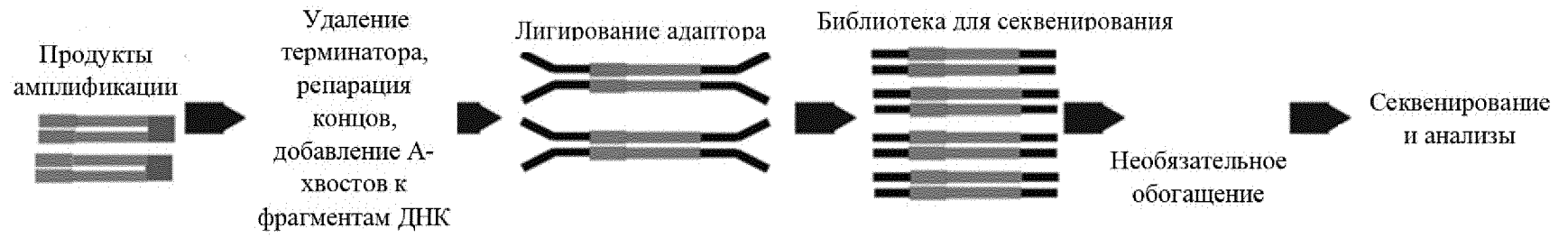
ФИГУРА 1А

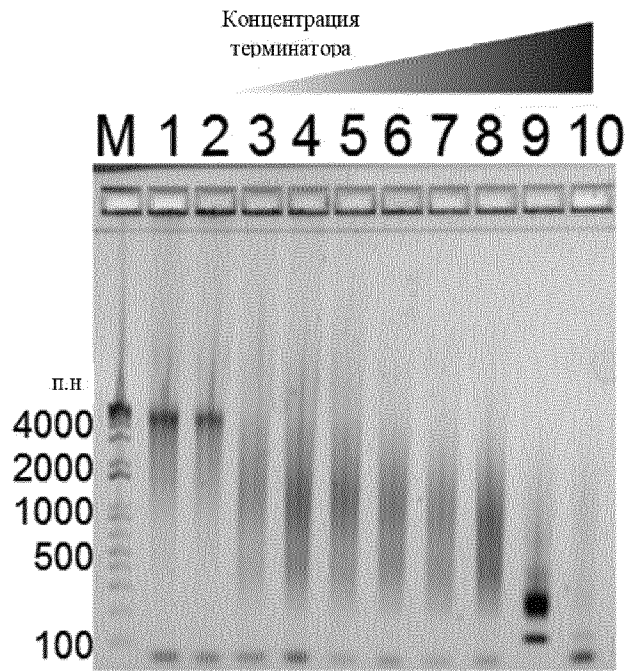


ФИГУРА 1В

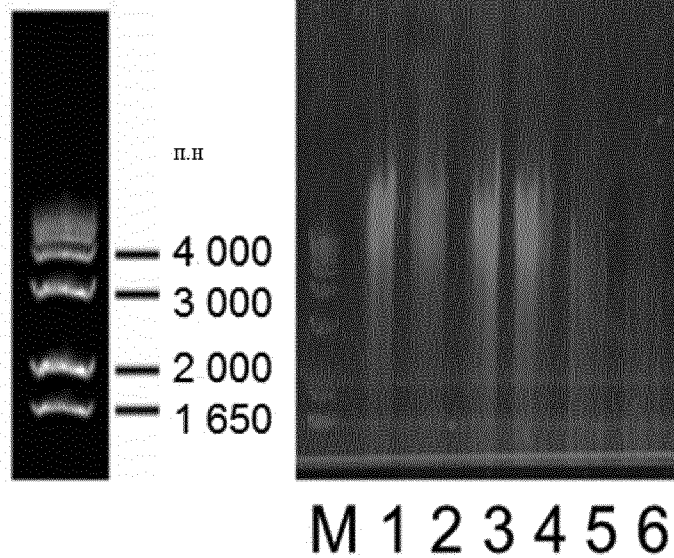


ФИГУРА 1С



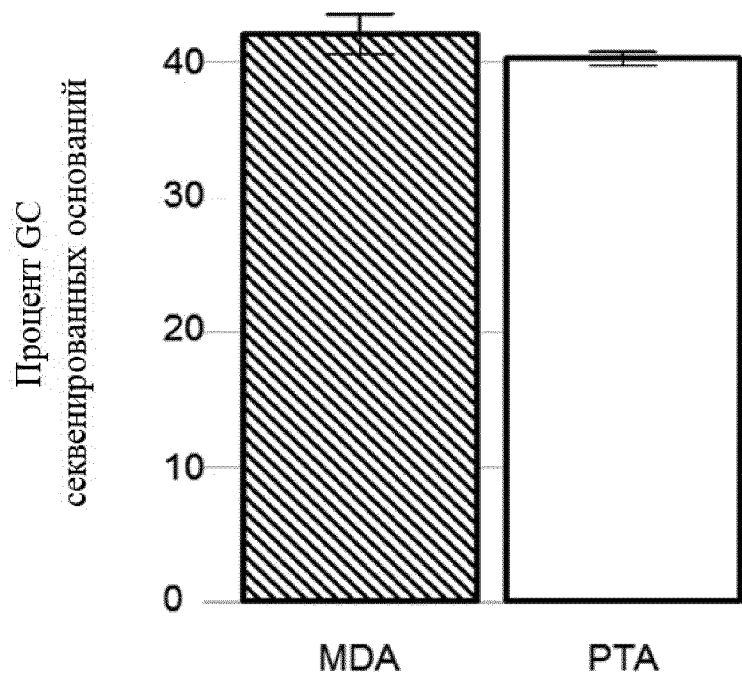


- М-лэддер
1. без ddNTP (полож. контроль)
  2. без ddNTP (полож. контроль)
  3. обратимый ddNTP – конц. 250 мкМ
  4. обратимый ddNTP – конц. 500 мкМ
  5. необратимый ddNTP – конц. 250 мкМ
  6. необратимый ddNTP – конц. 500 мкМ

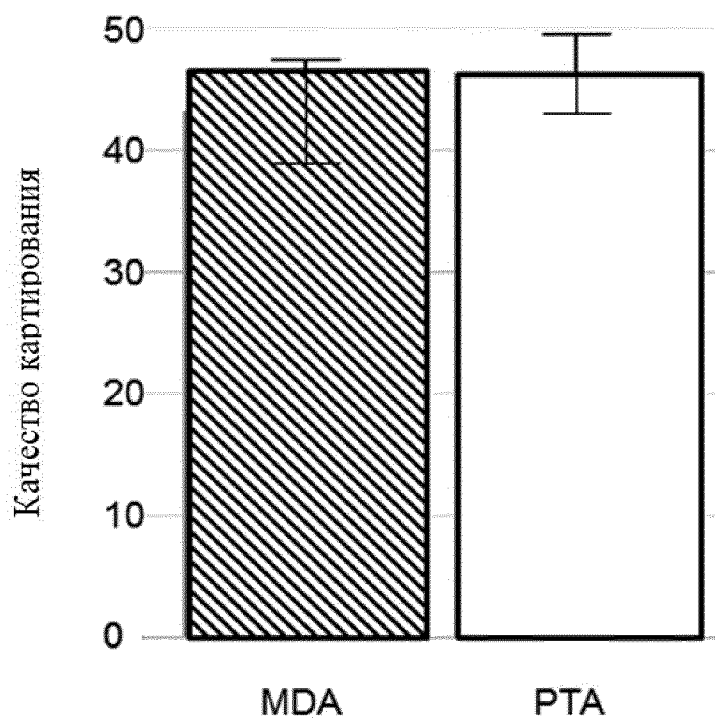


ФИГУРА 2А

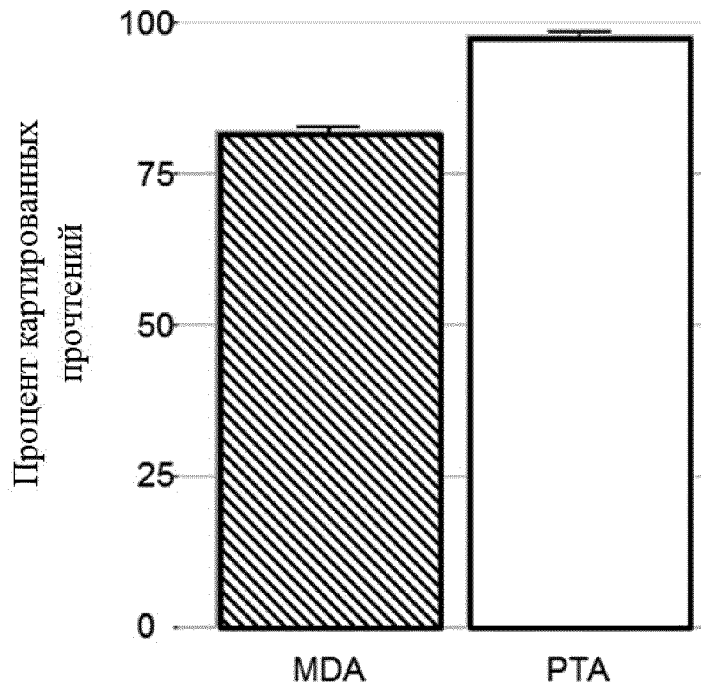




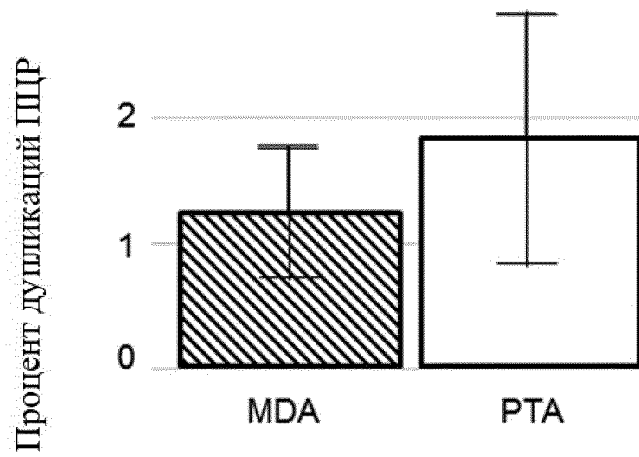
ФИГУРА 2В



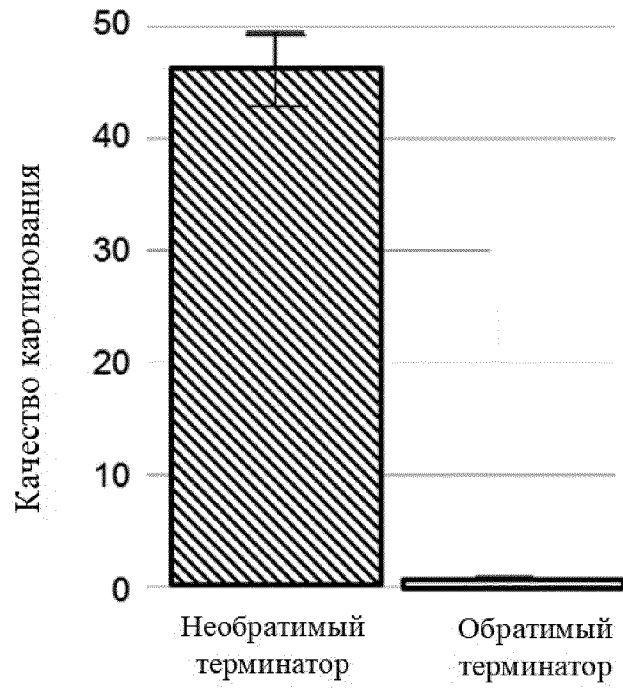
ФИГУРА 2С



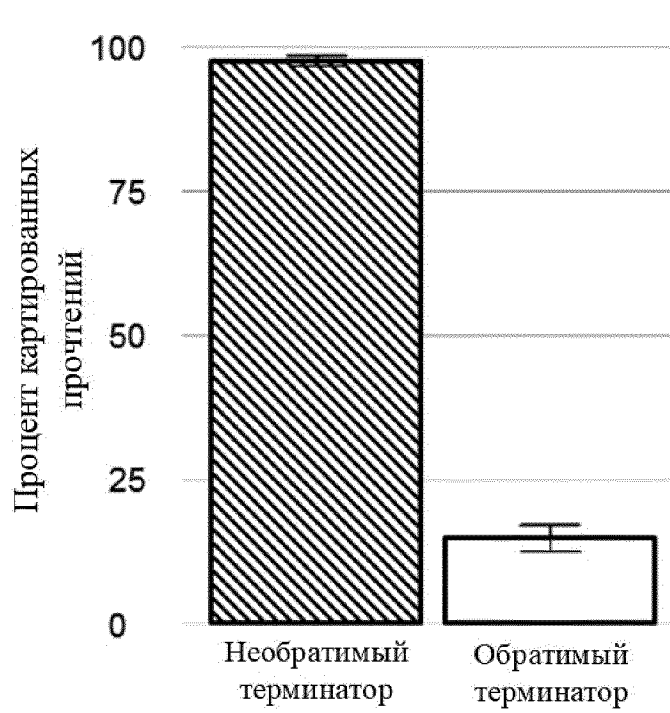
ФИГУРА 2D



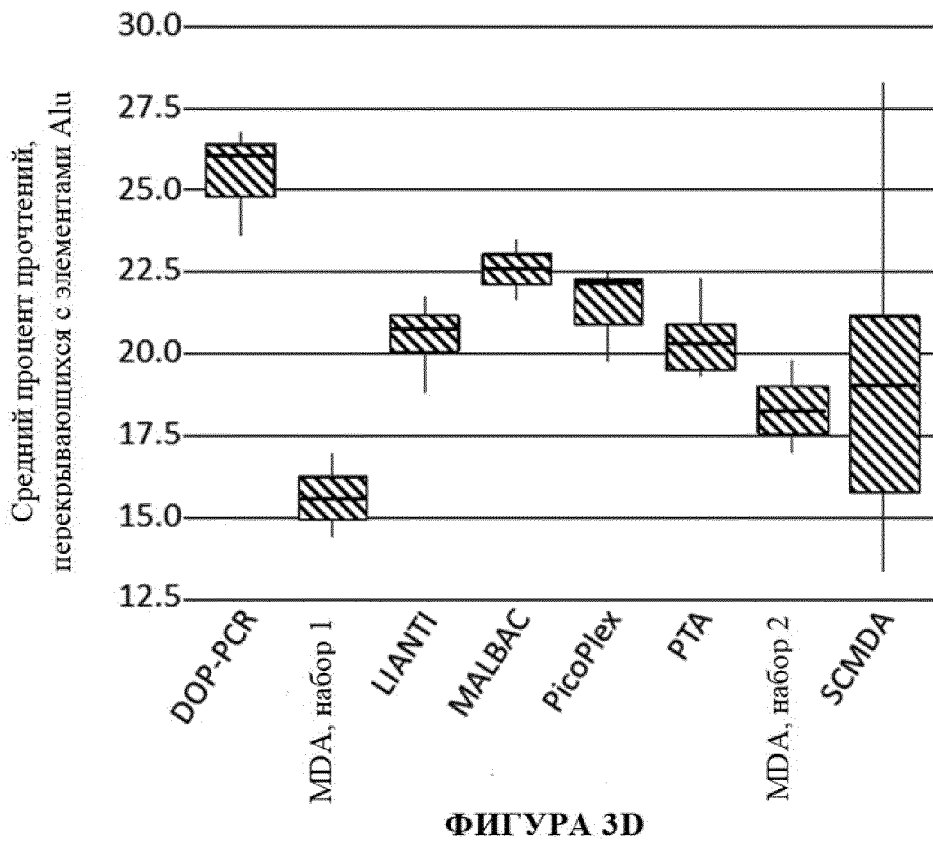
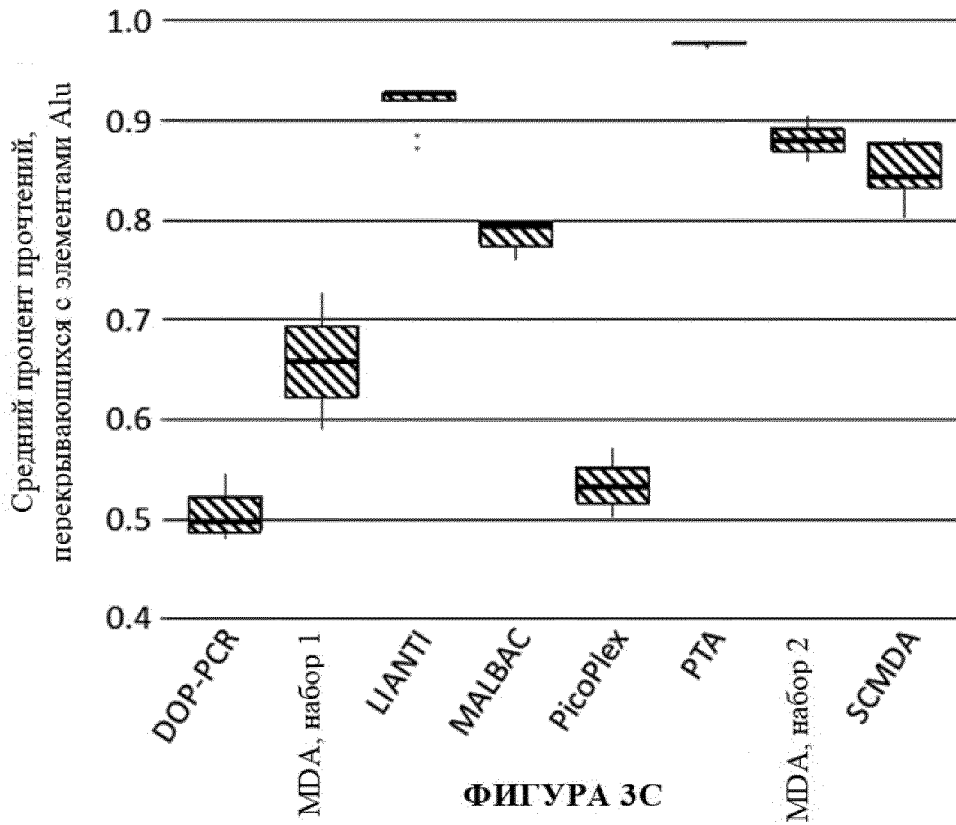
ФИГУРА 2E

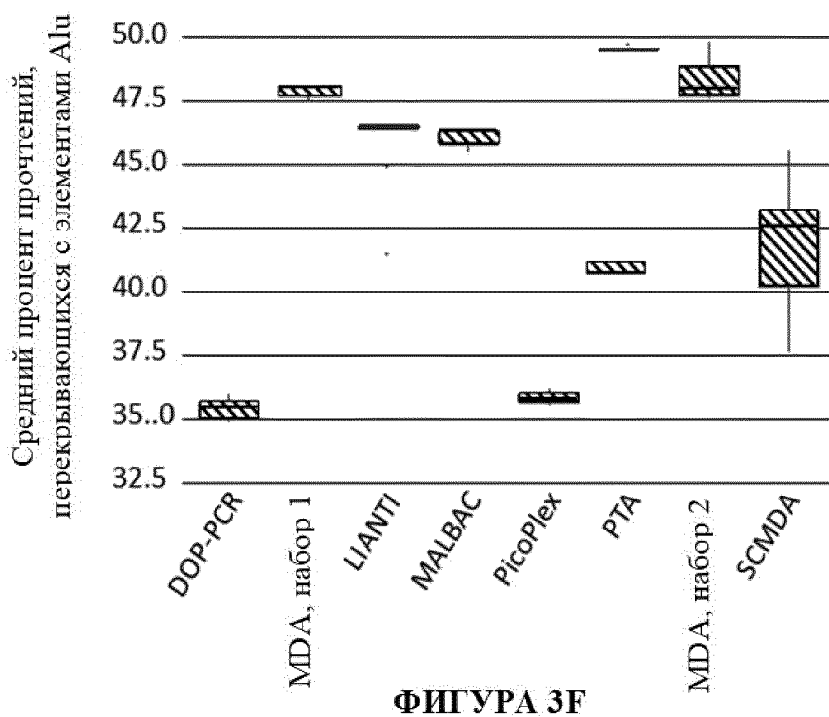
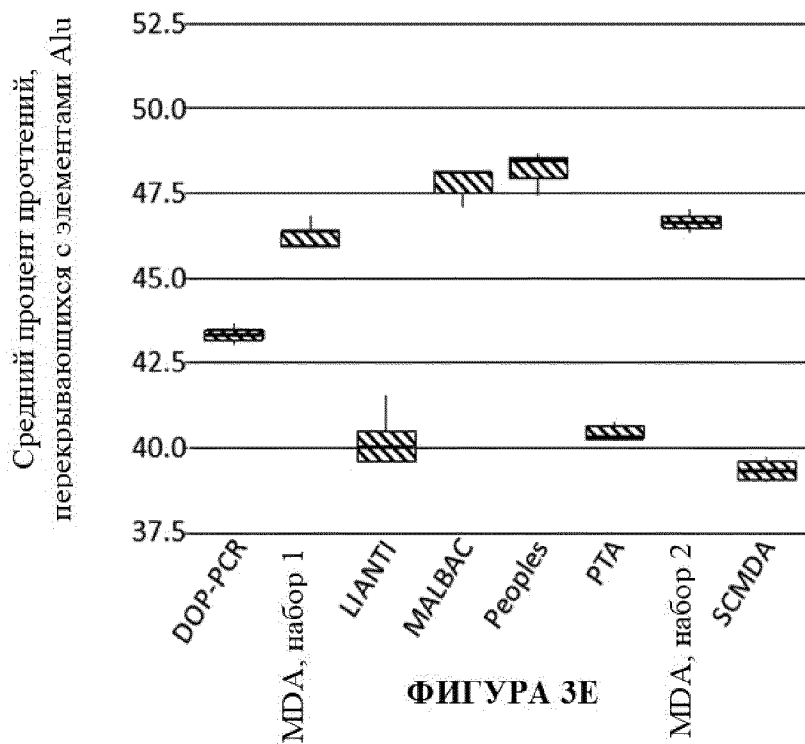


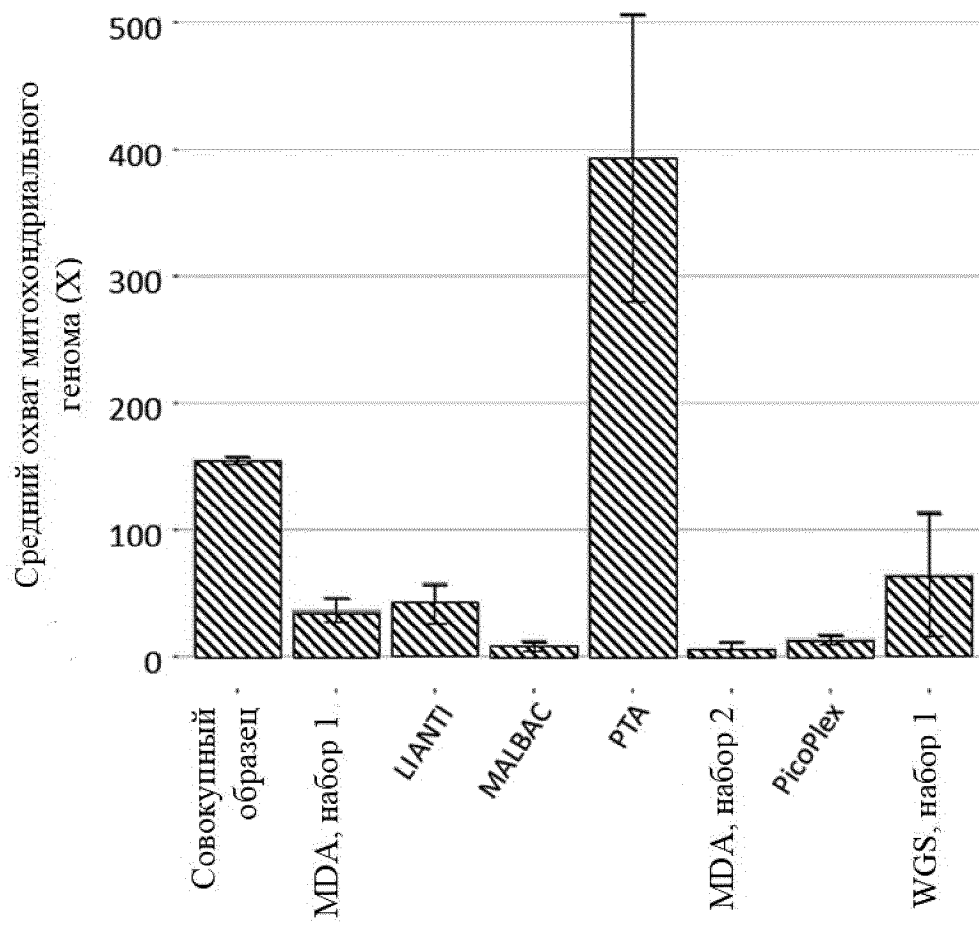
ФИГУРА 3А



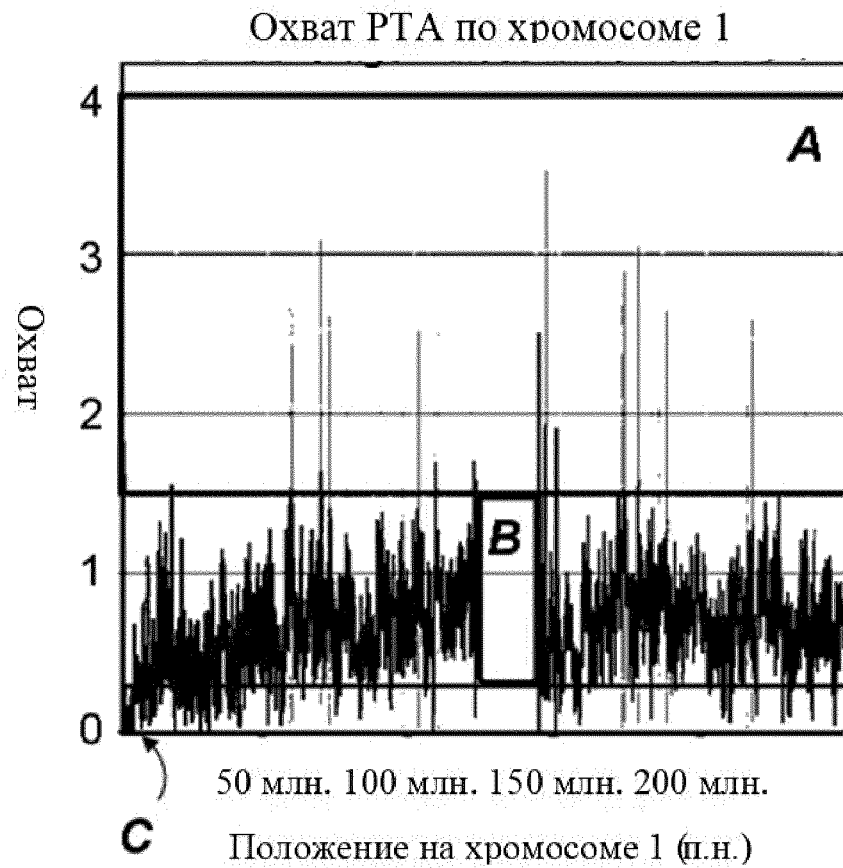
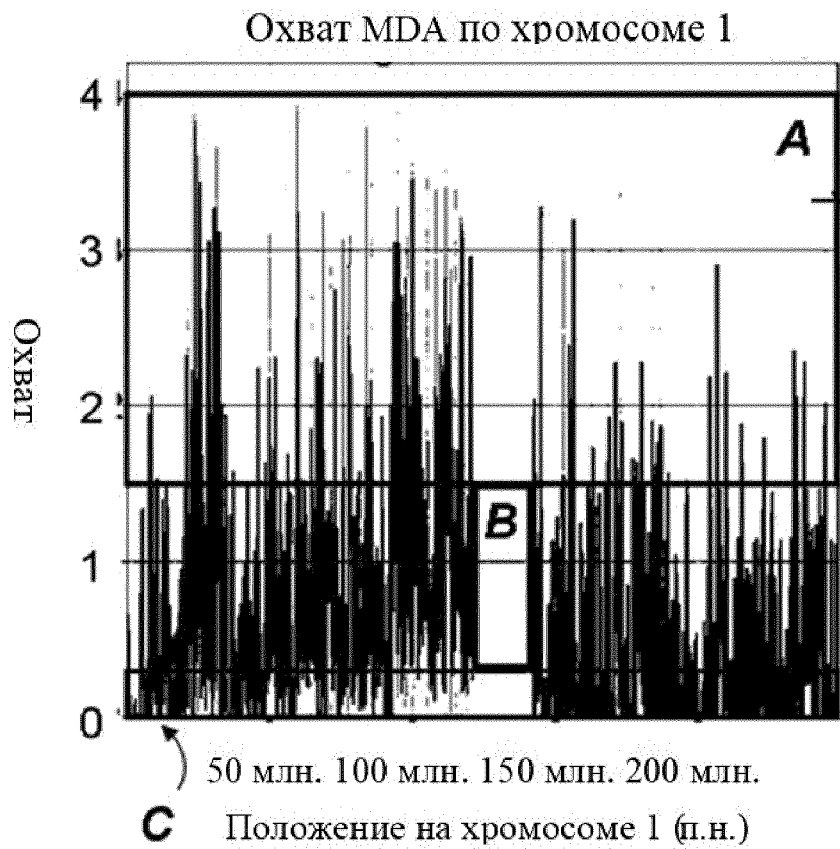
ФИГУРА 3В



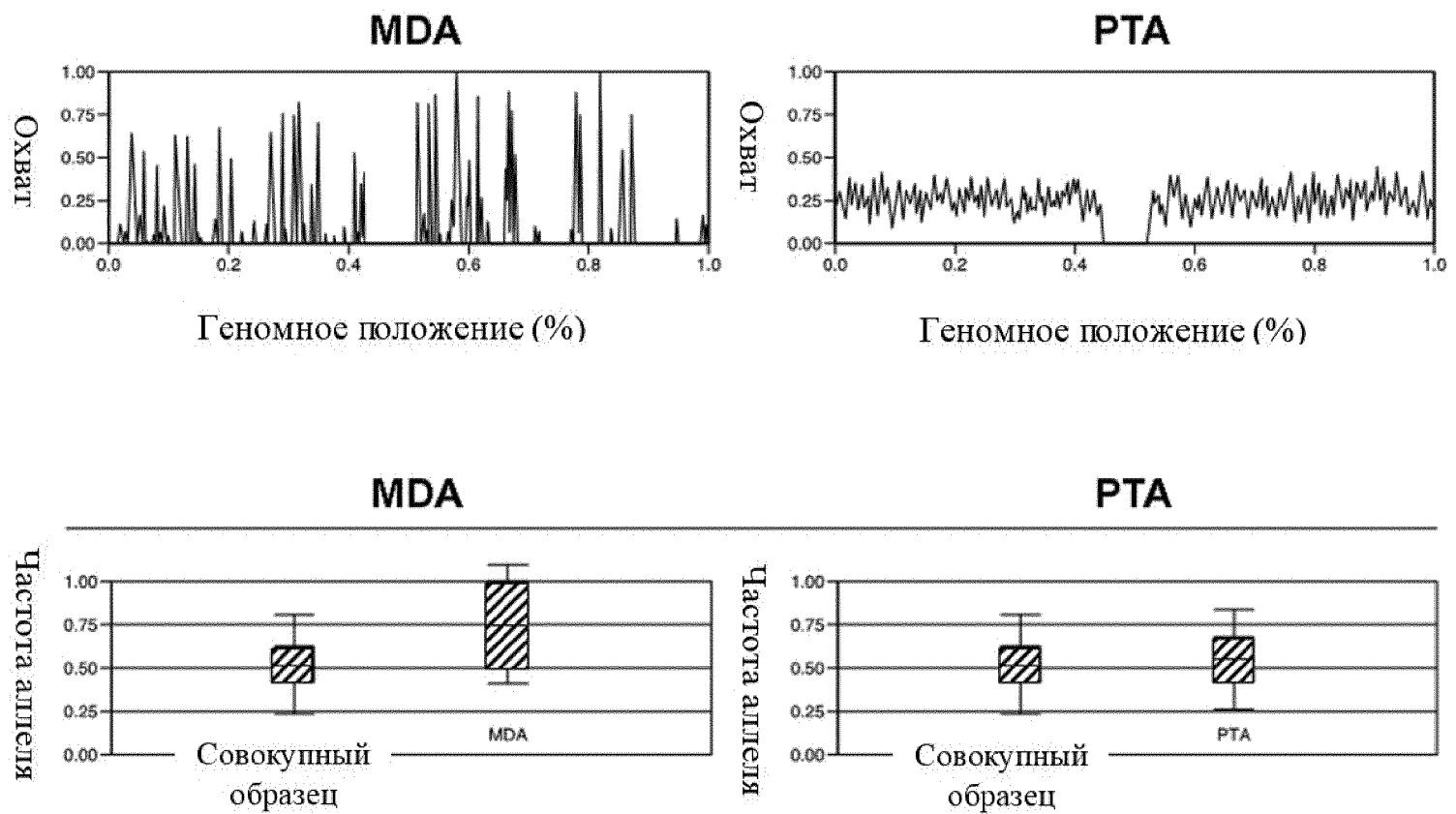




ФИГУРА 3G



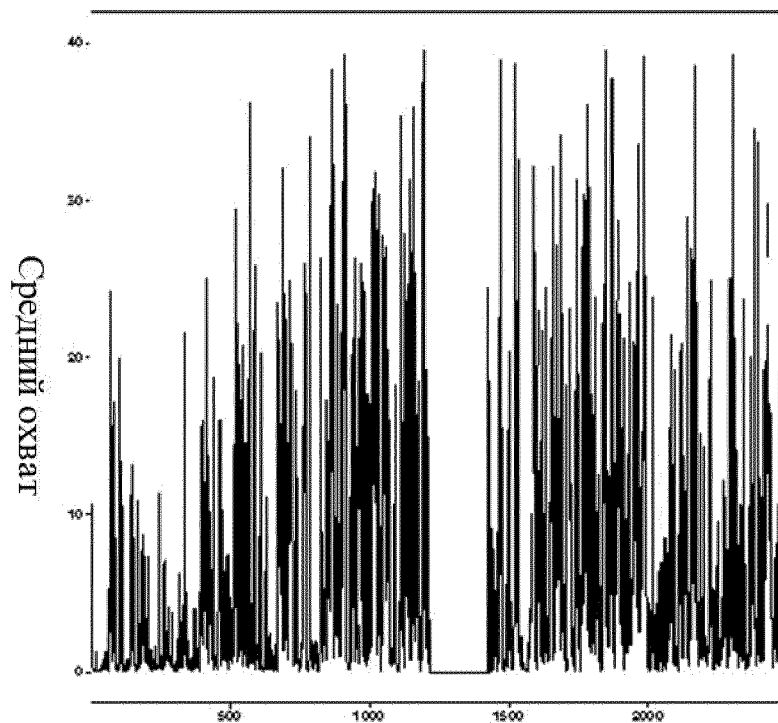
ФИГУРА 4В



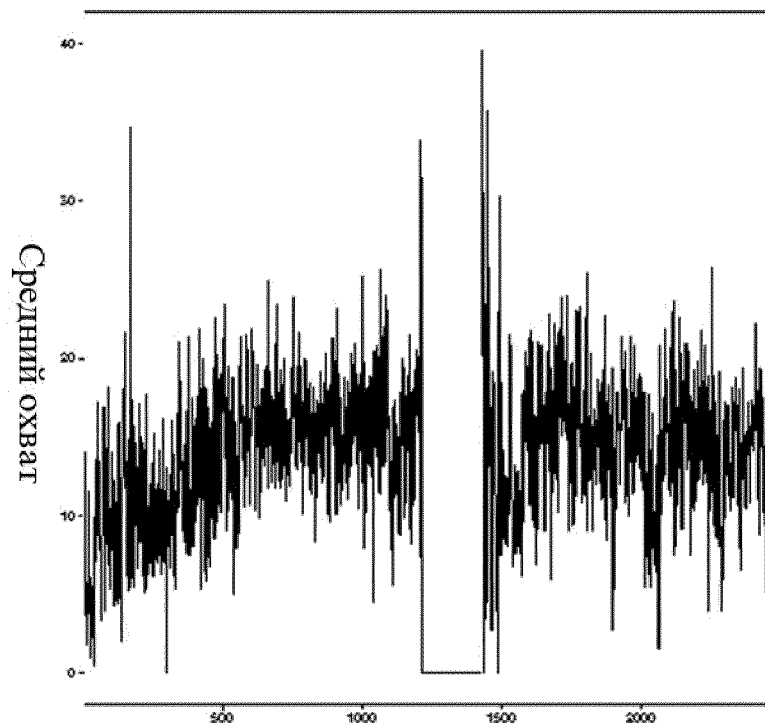


Средний охват в 1 т.п.н. окнах по хромосоме 1 MDA Средний охват в 1 т.п.н. окнах по хромосоме 1 РТА

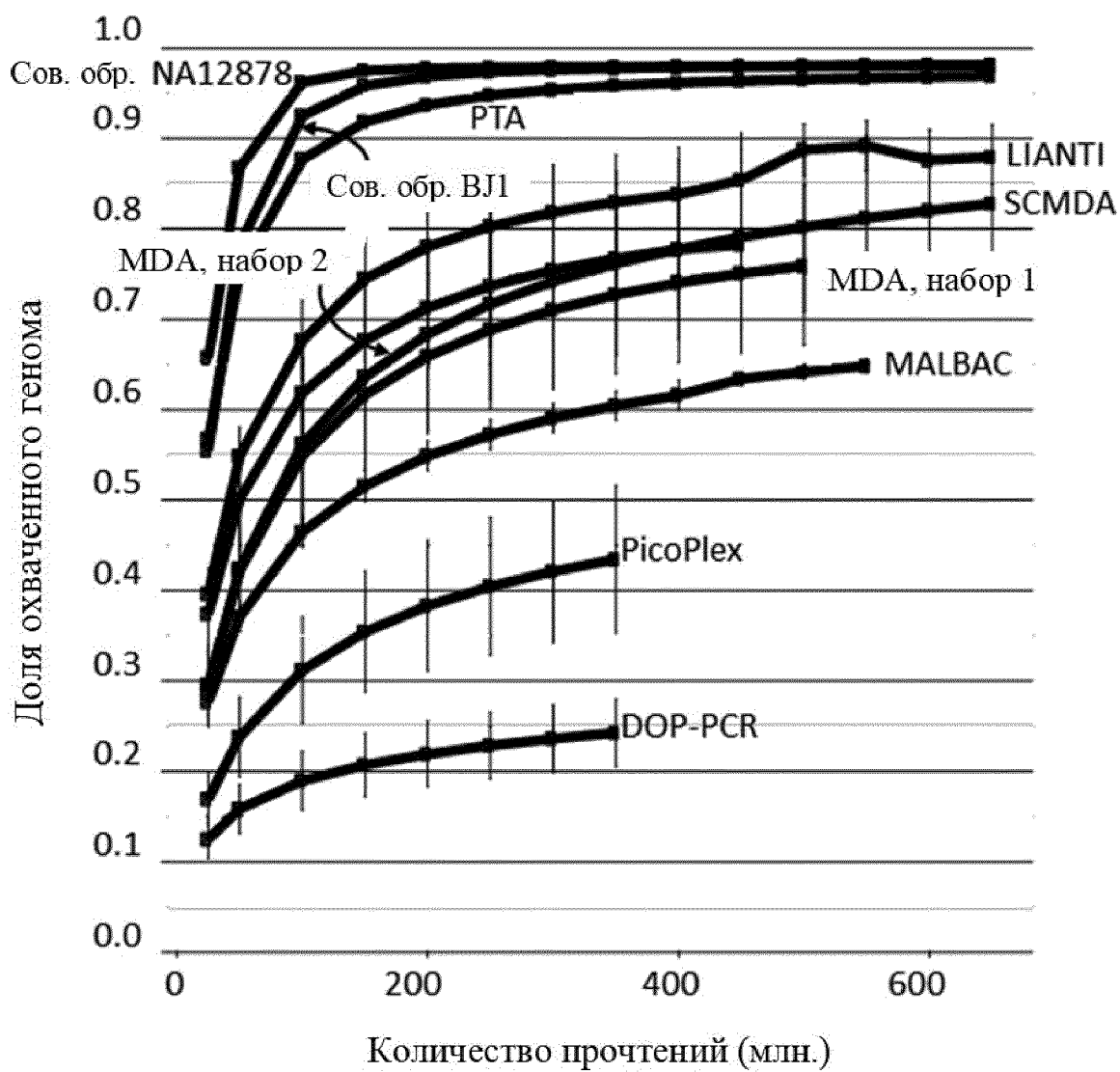
ФИГУРА 4С



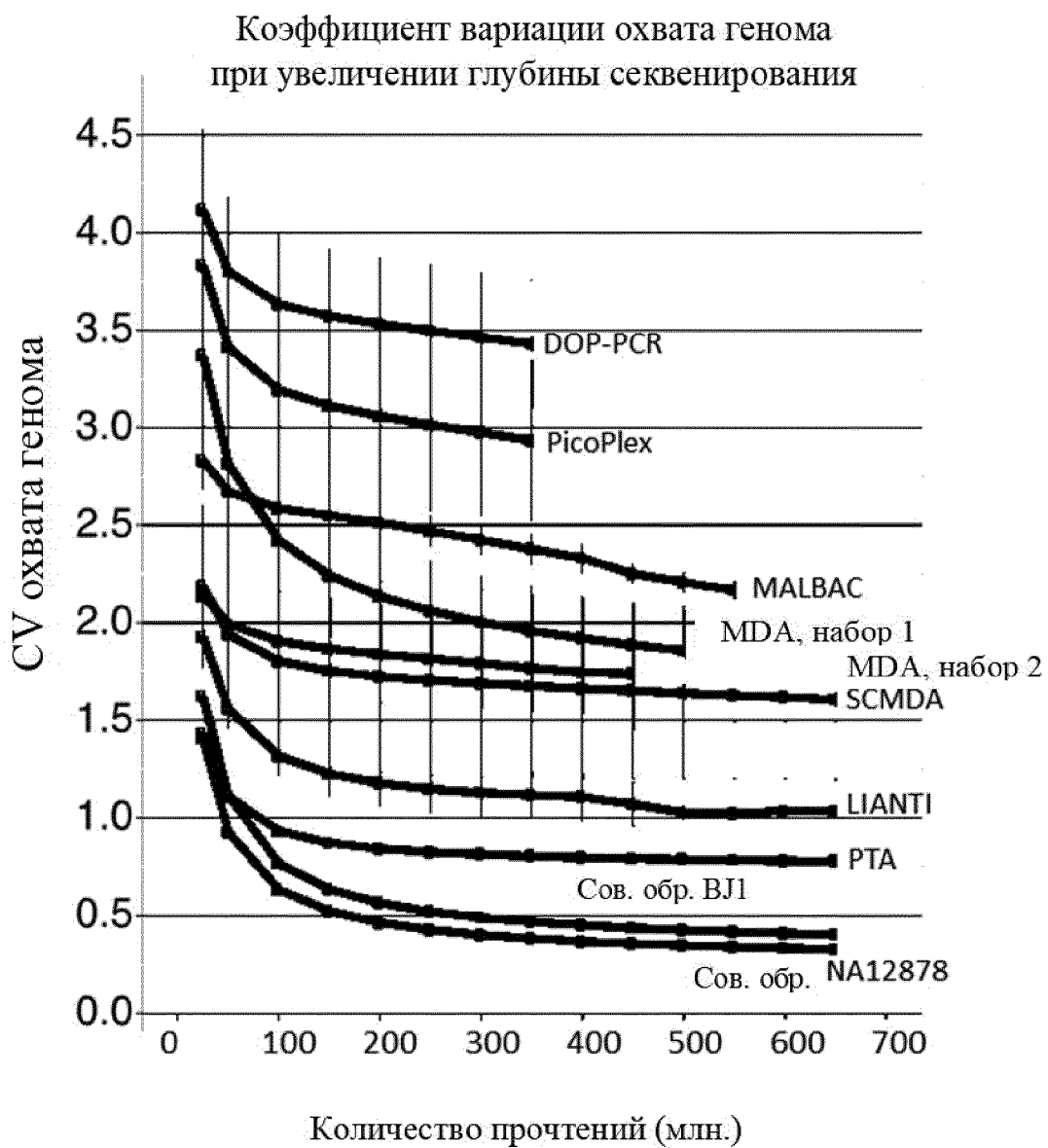
Геномное окно (1 т.п.н.)



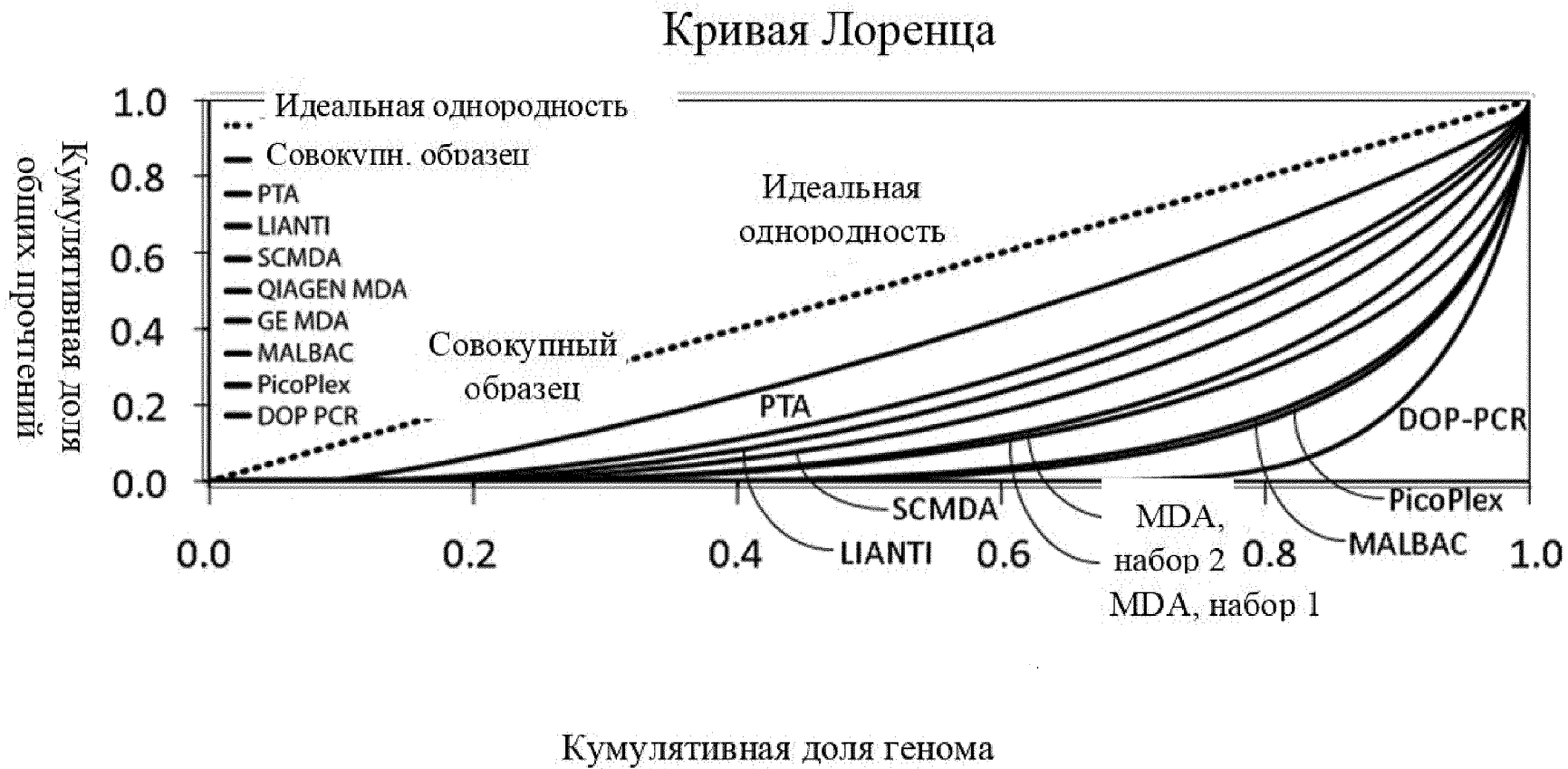
Геномное окно (1 т.п.н.)

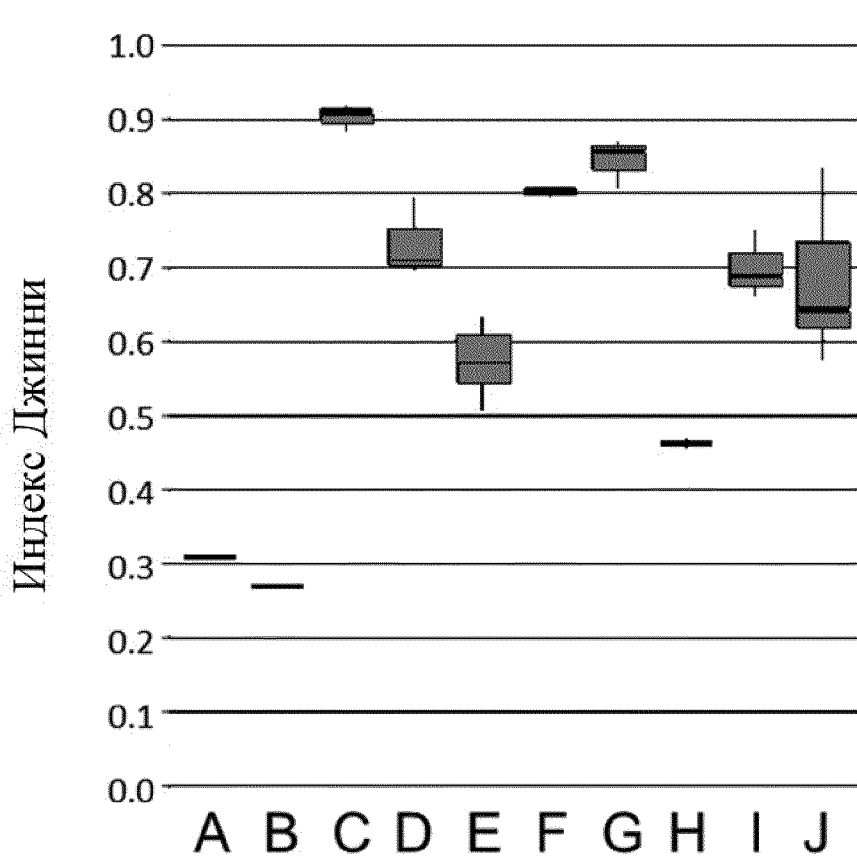


ФИГУРА 5А



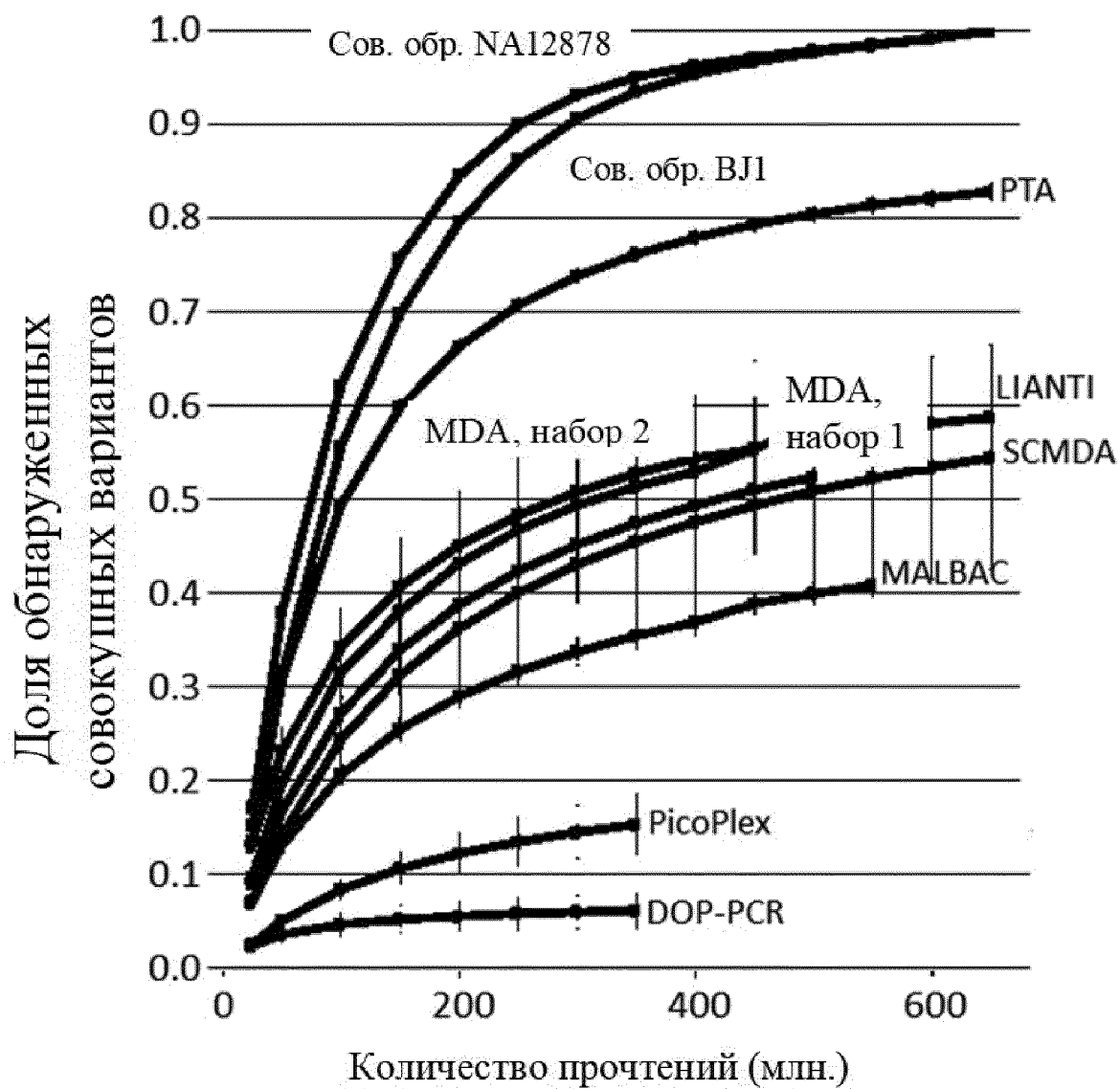
ФИГУРА 5В



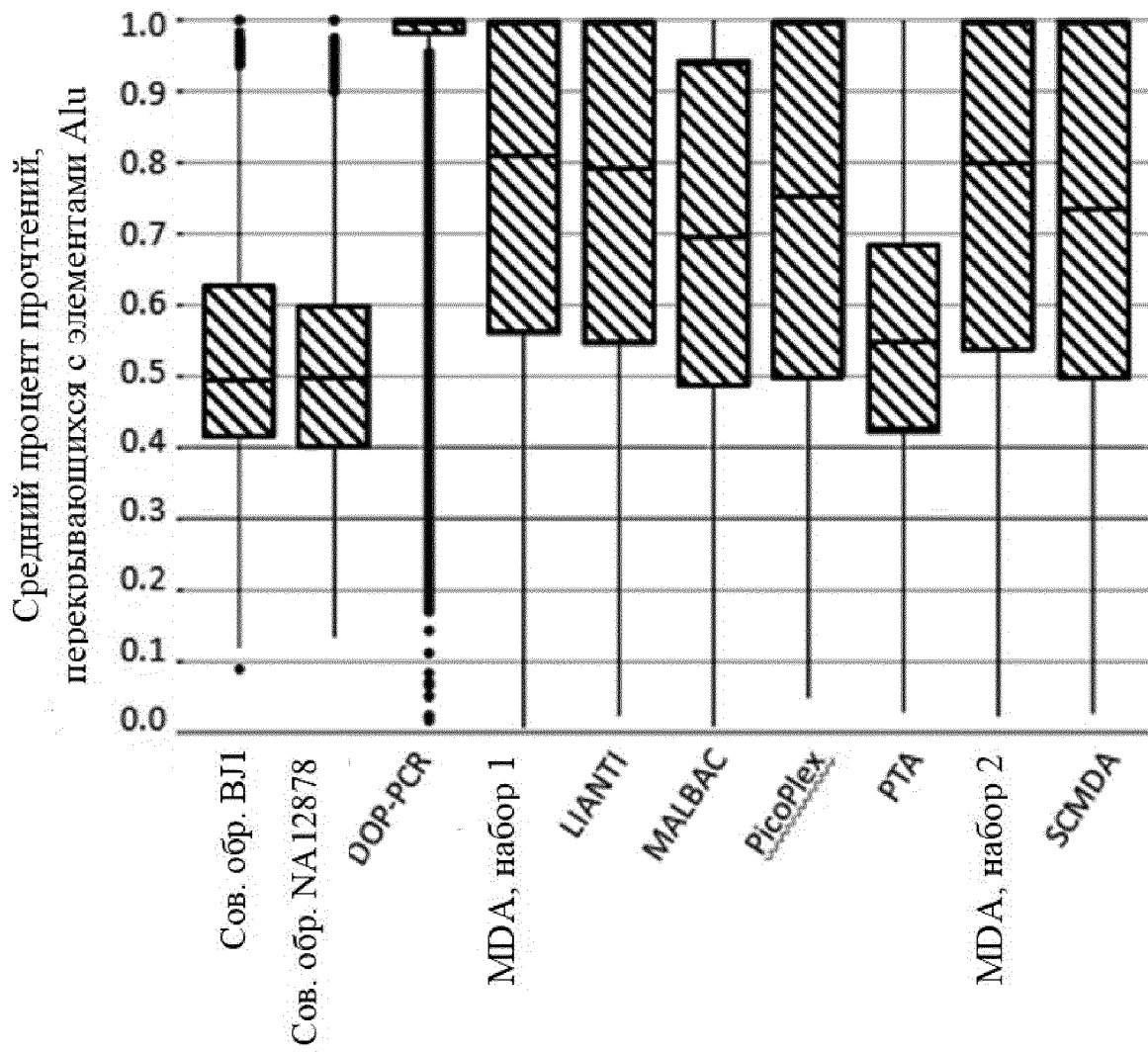


- A. Совокупн. обр. ВJ1
- B. Совокупн. обр. NA12878
- C. DOP-PCR
- D. MDA, набор 1
- E. LIANTI
- F. MALBEC
- G. PicoPLex
- H. PTA
- I. MDA, набор 2
- J. SCMDA

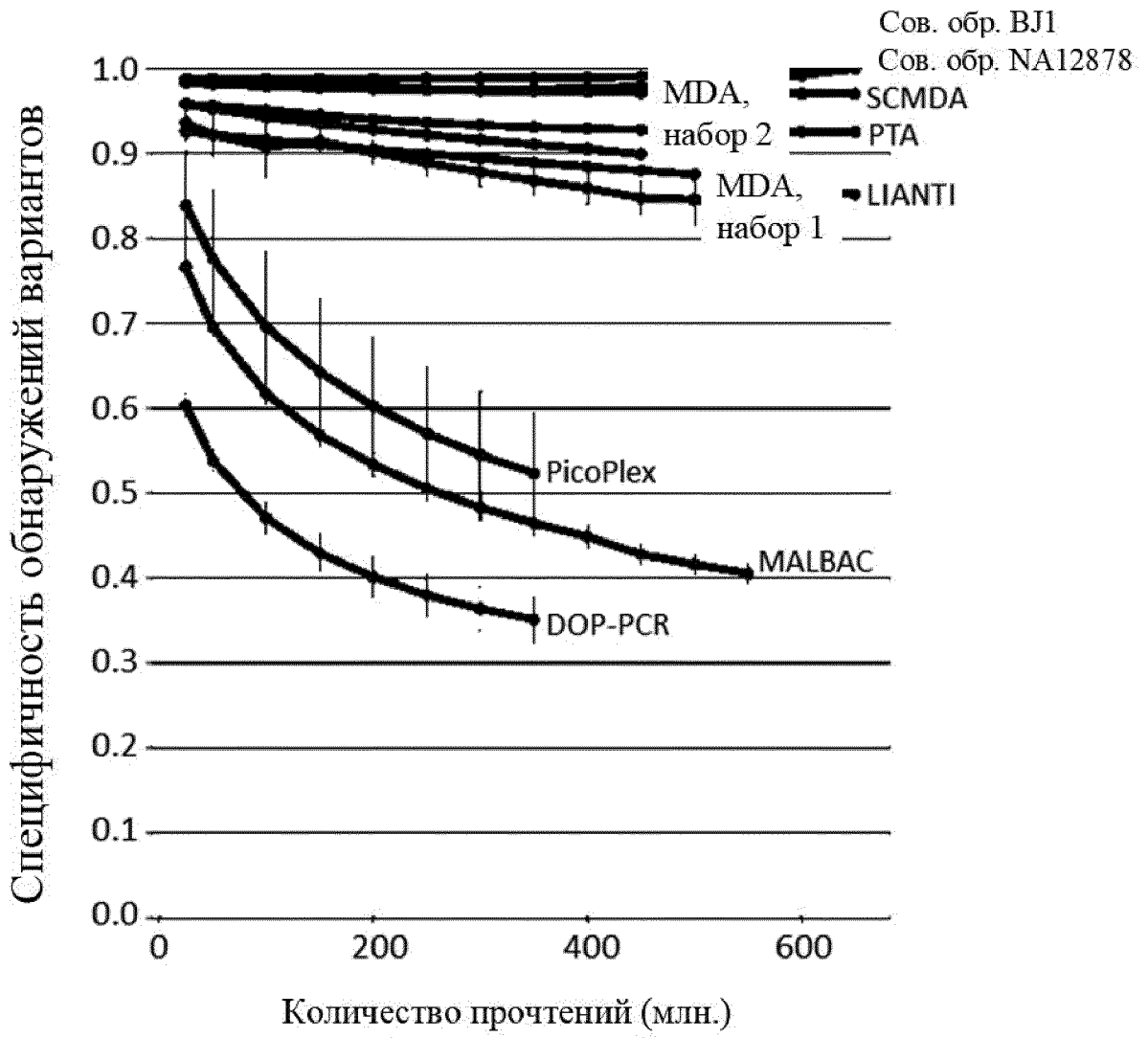
**ФИГУРА 5D**



ФИГУРА 5Е

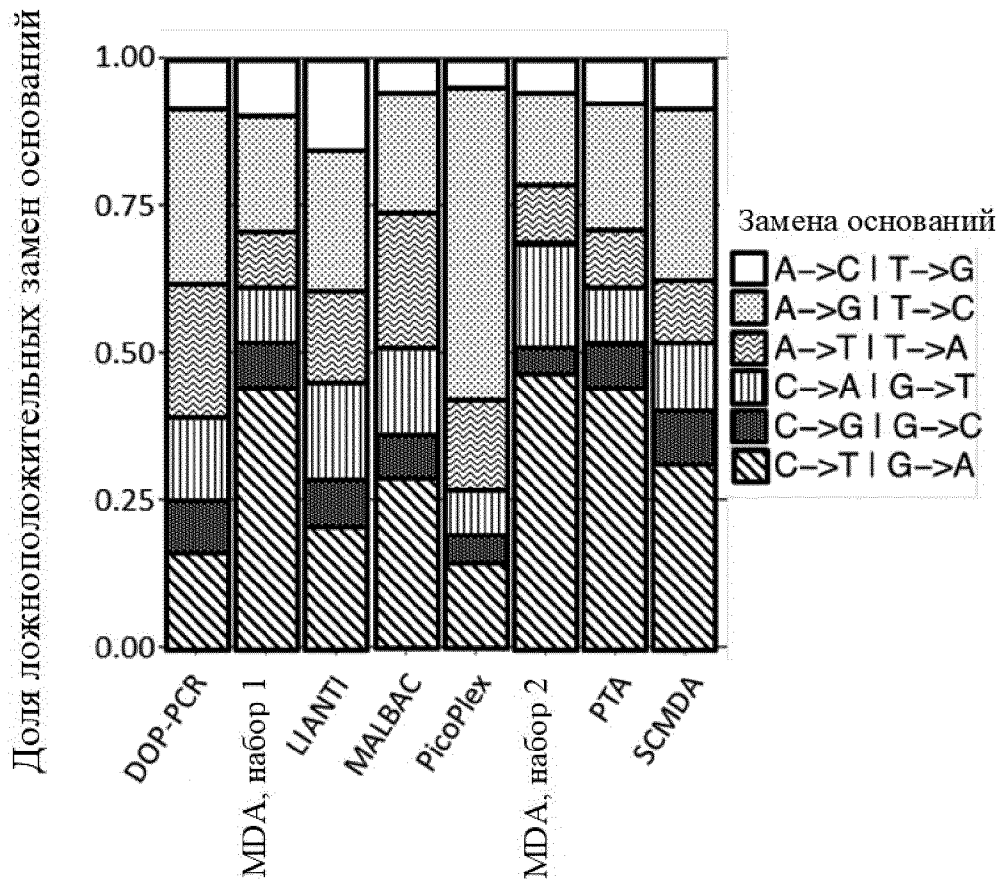


ФИГУРА 5F

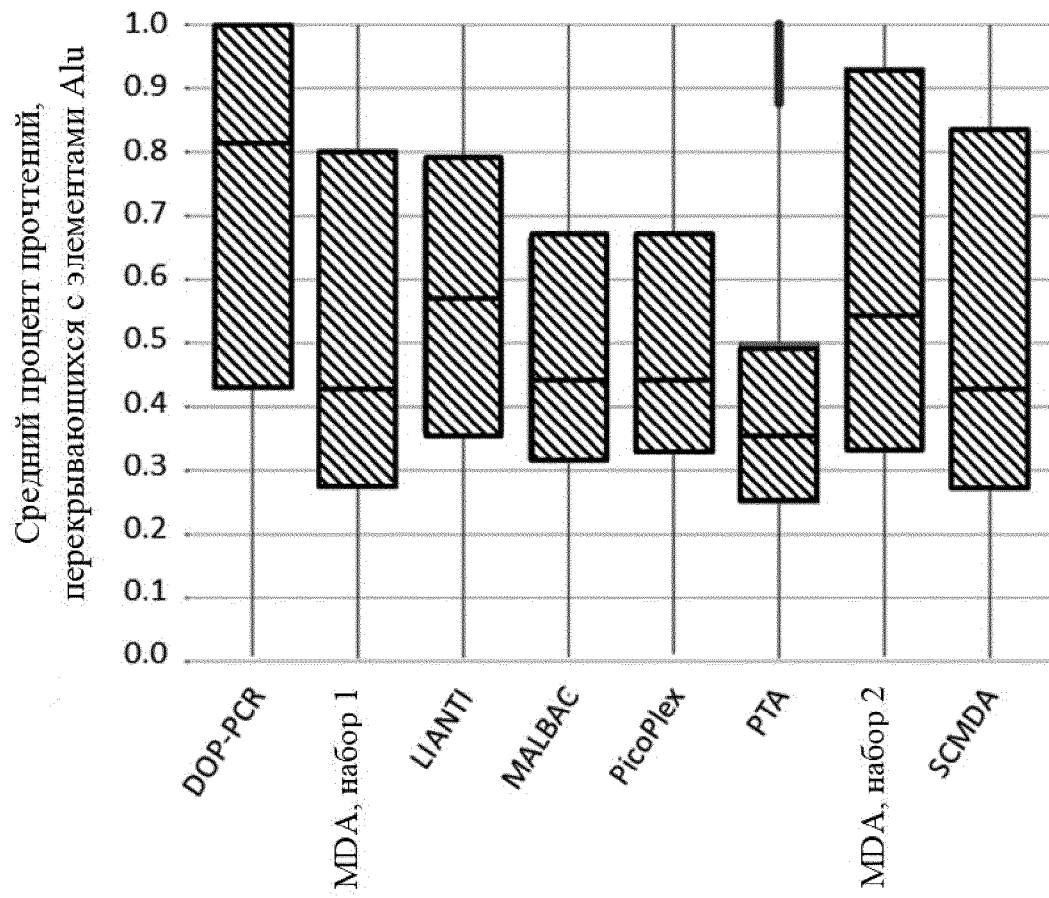


ФИГУРА 5G





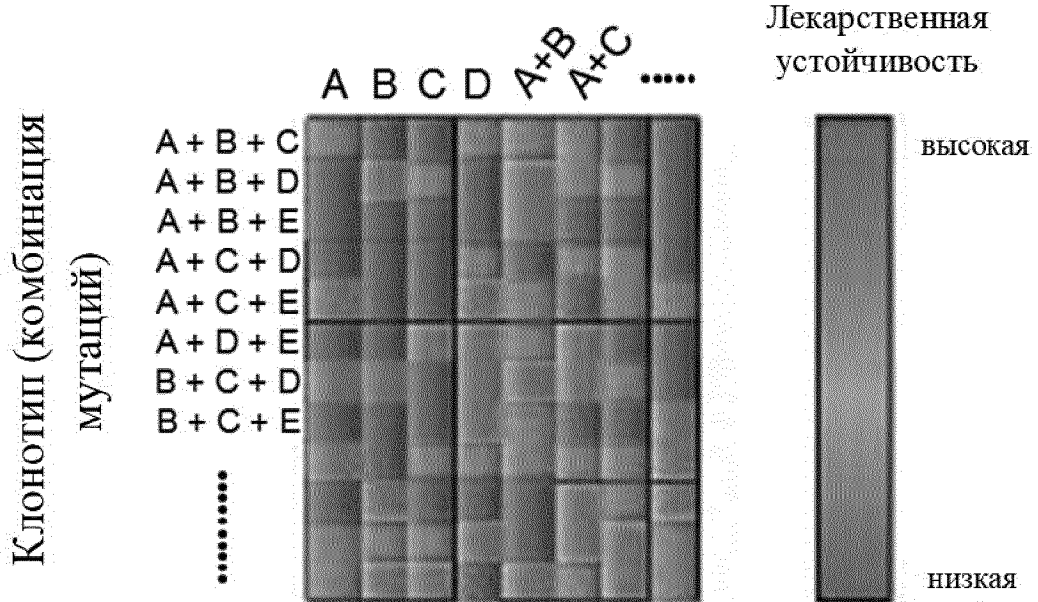
ФИГУРА 5Н



ФИГУРА 5I

ALL предшественников  
В-клеток

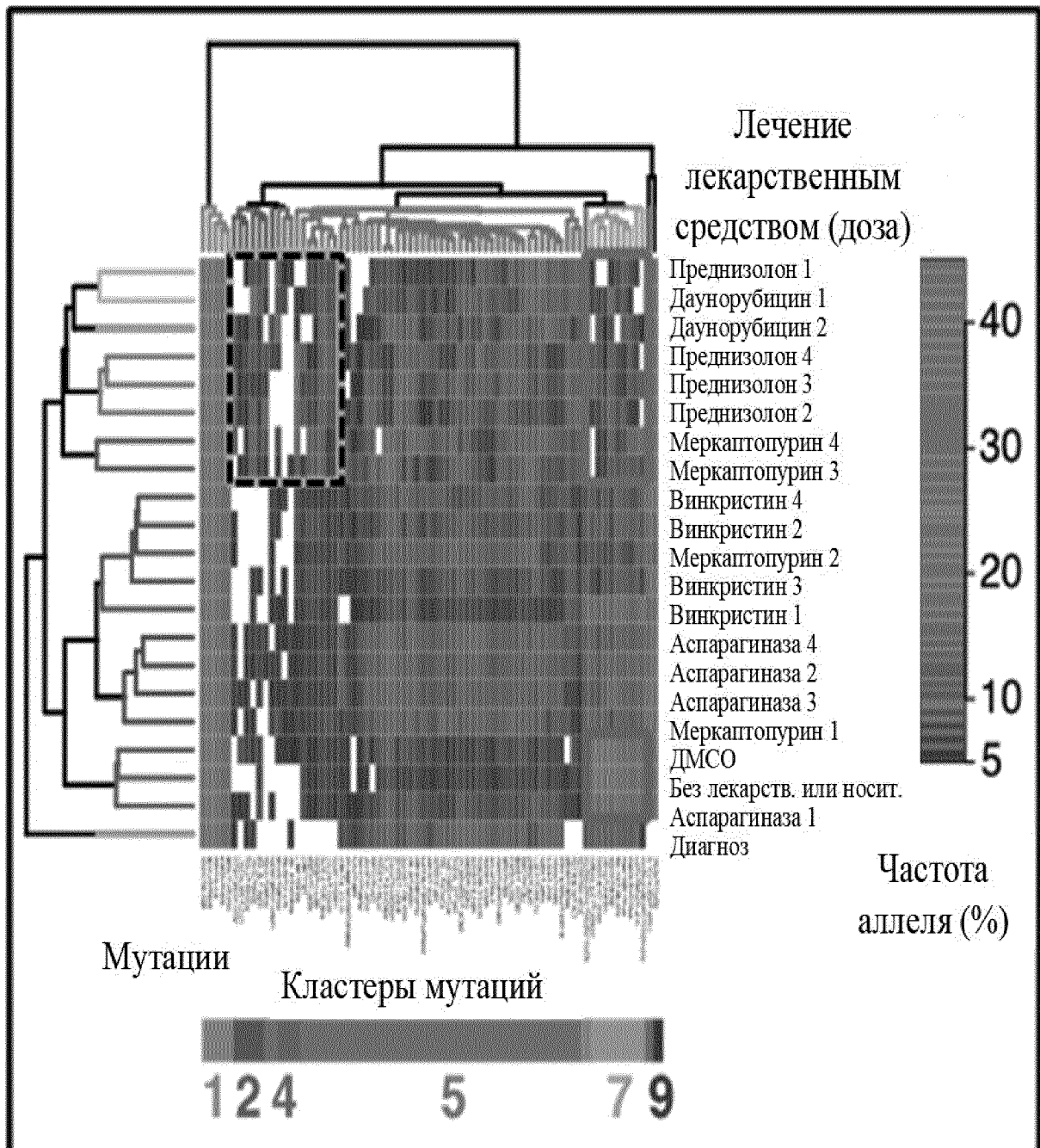
Лекарственное(ые)  
средство(а)



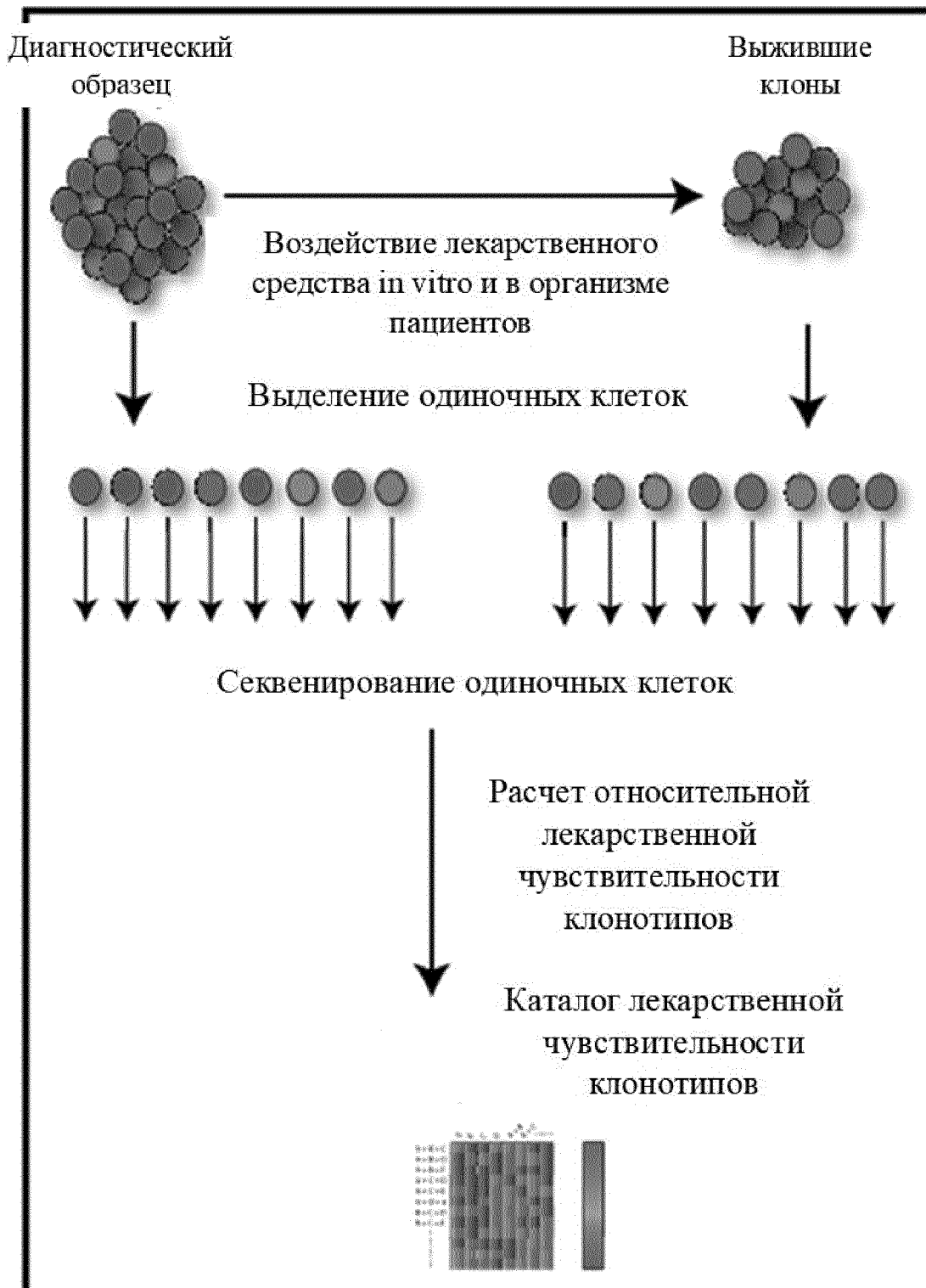
ФИГУРА 6А

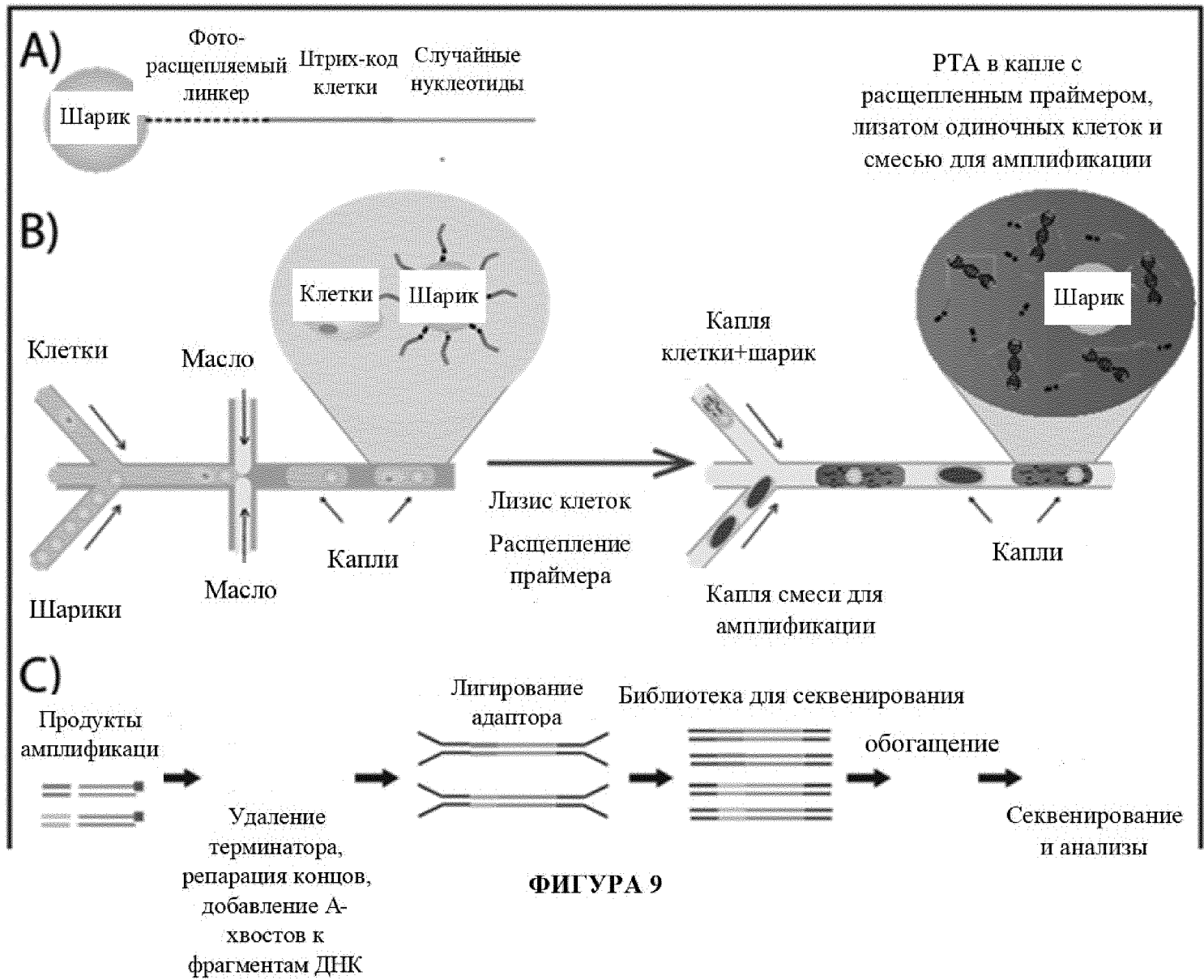


ФИГУРА 6В

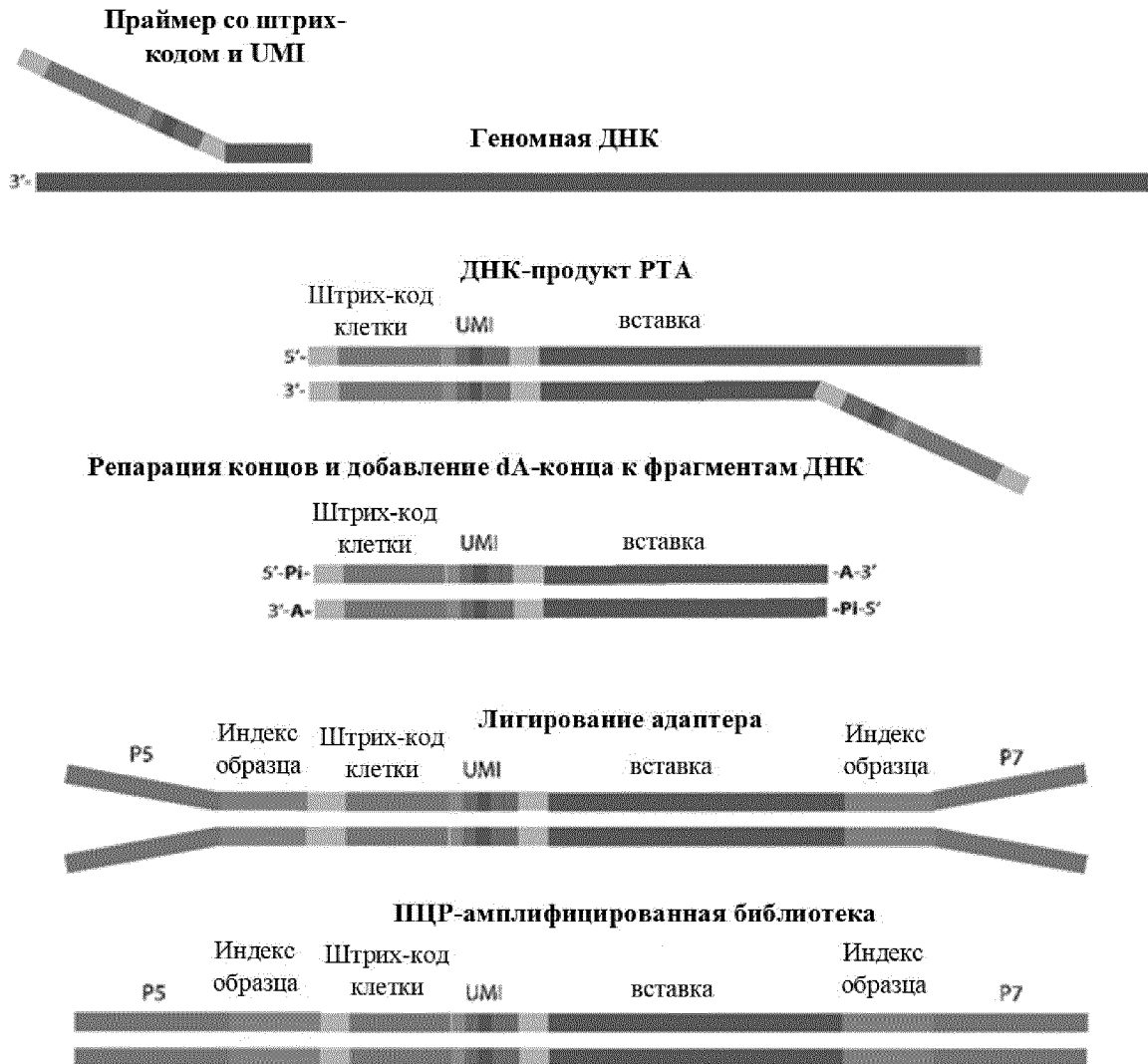


ФИГУРА 7

**ФИГУРА 8**



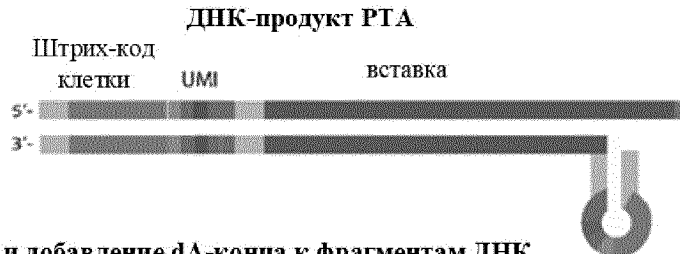
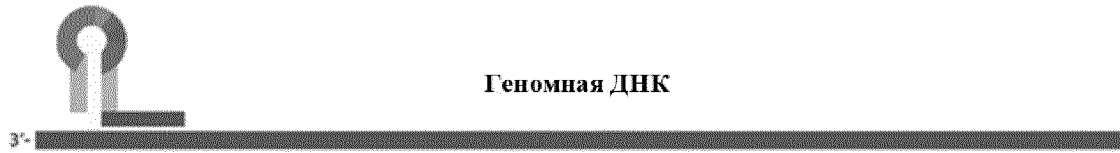
ФИГУРА 9



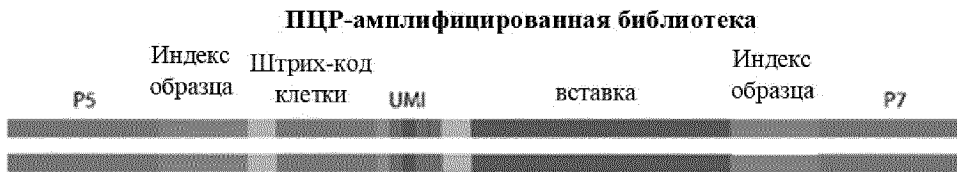
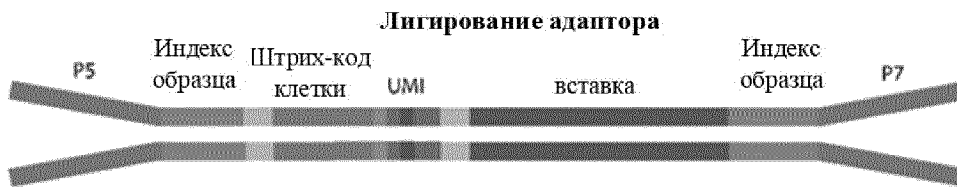
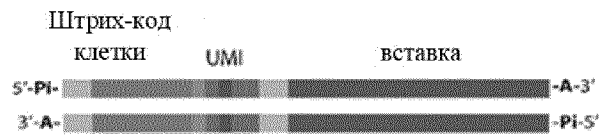
ФИГУРА 10А



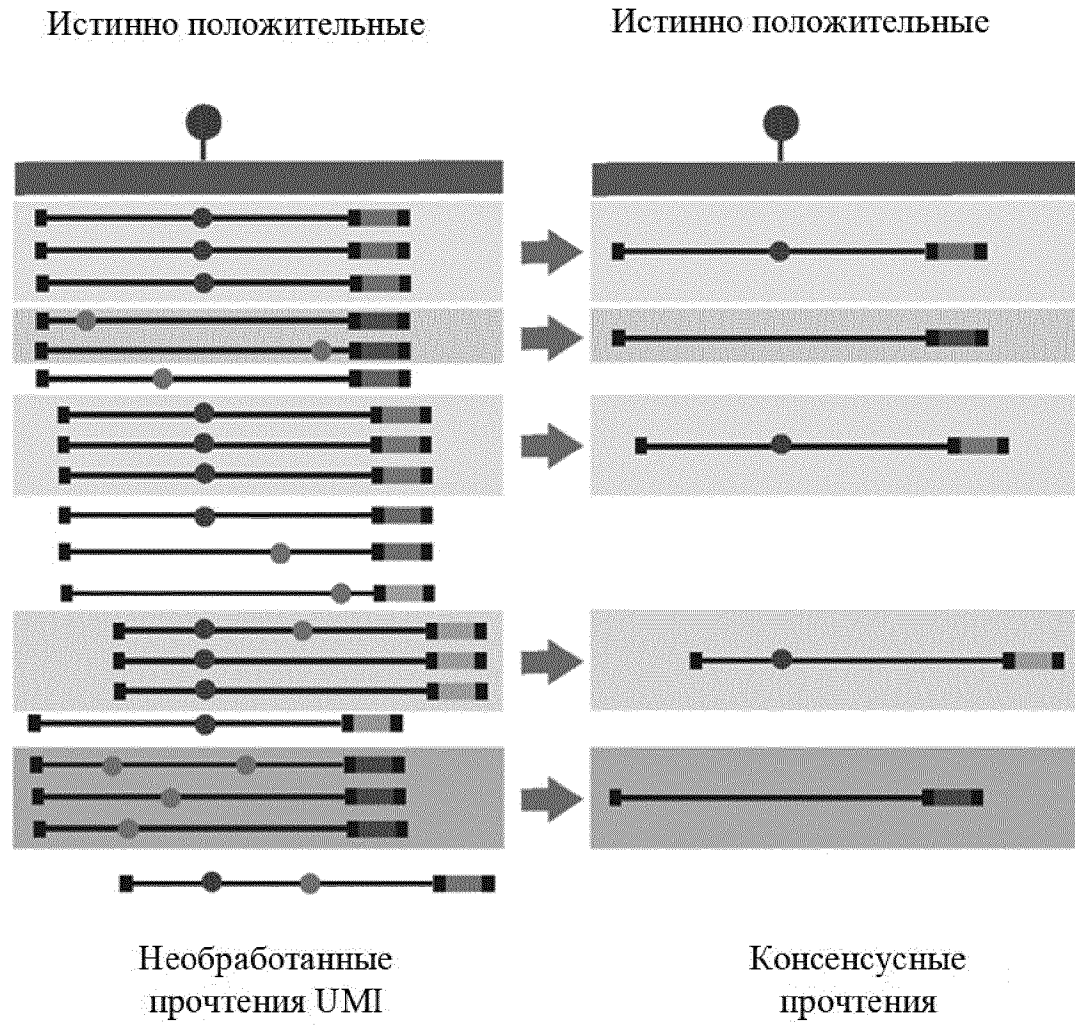
**Праймер со штрих-кодом и UMI**



**Репарация концов и добавление dA-конца к фрагментам ДНК**



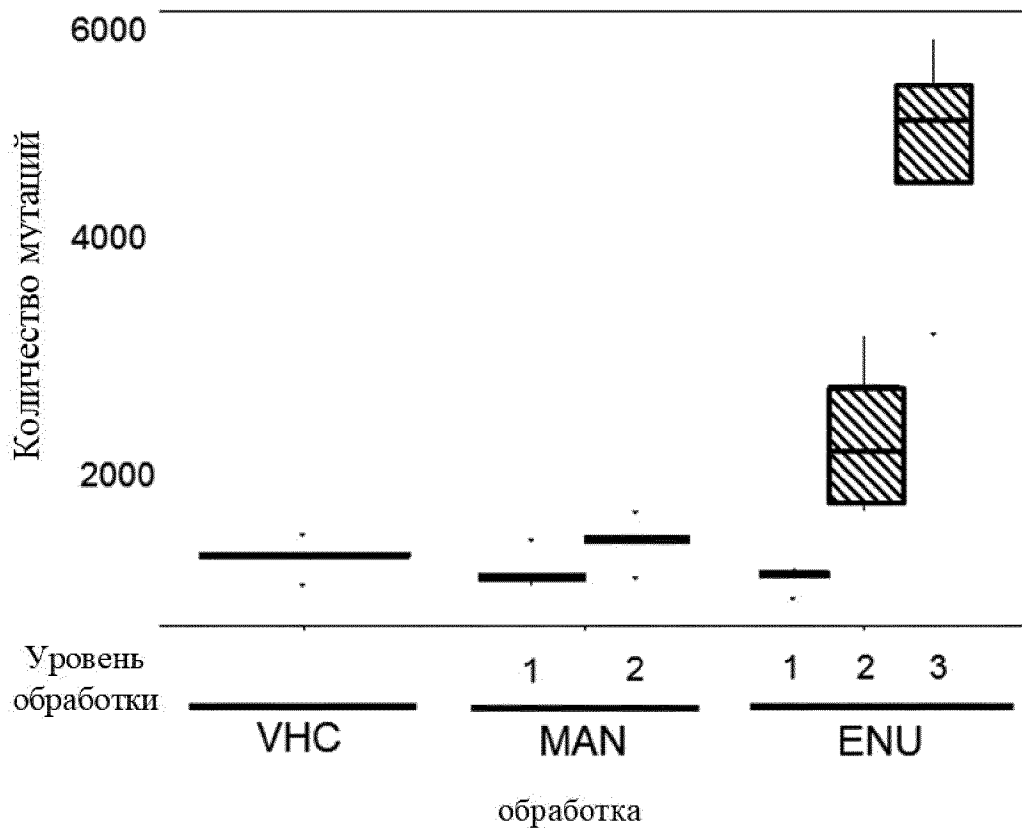
**ФИГУРА 10В**



ФИГУРА 11А

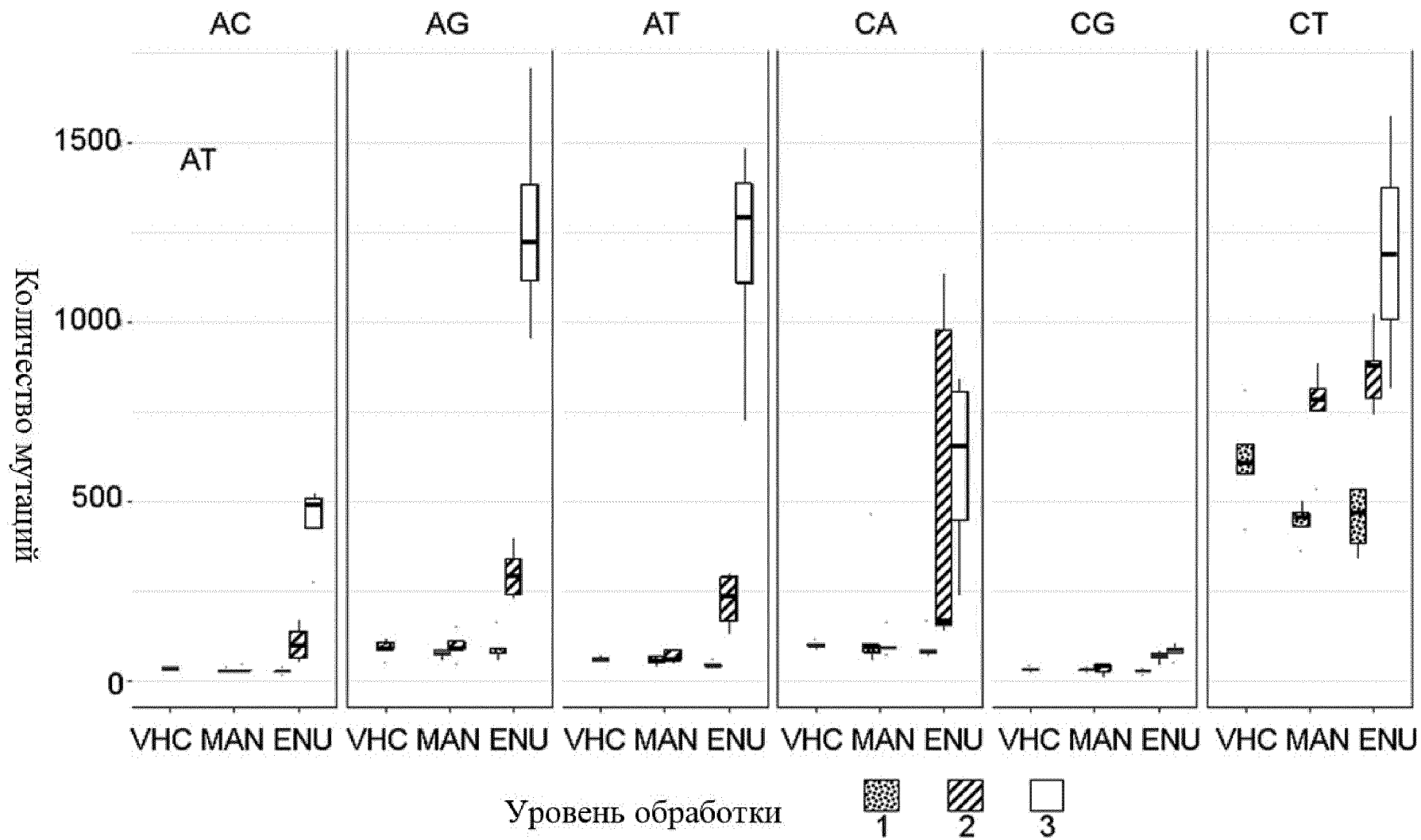
ФИГУРА 11В

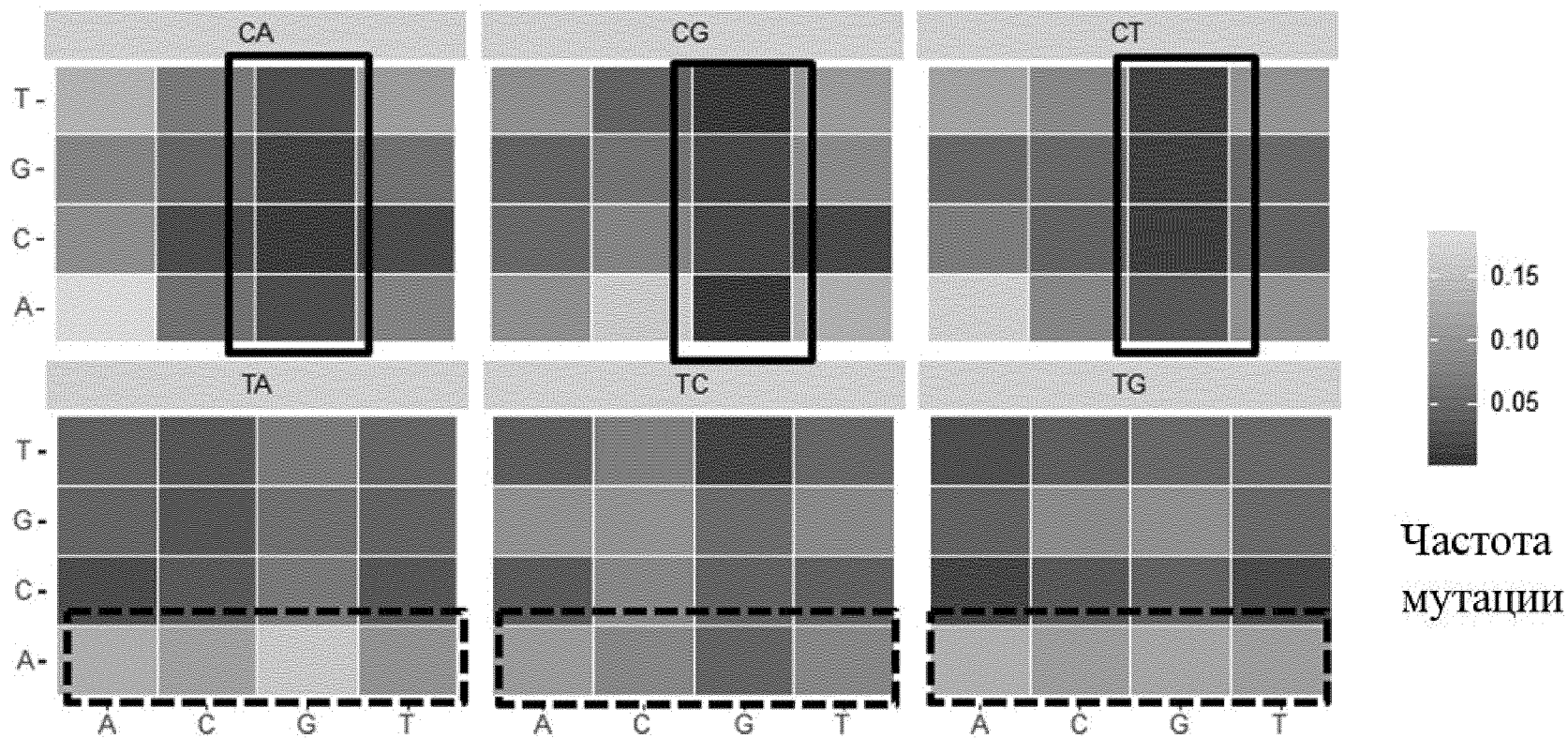




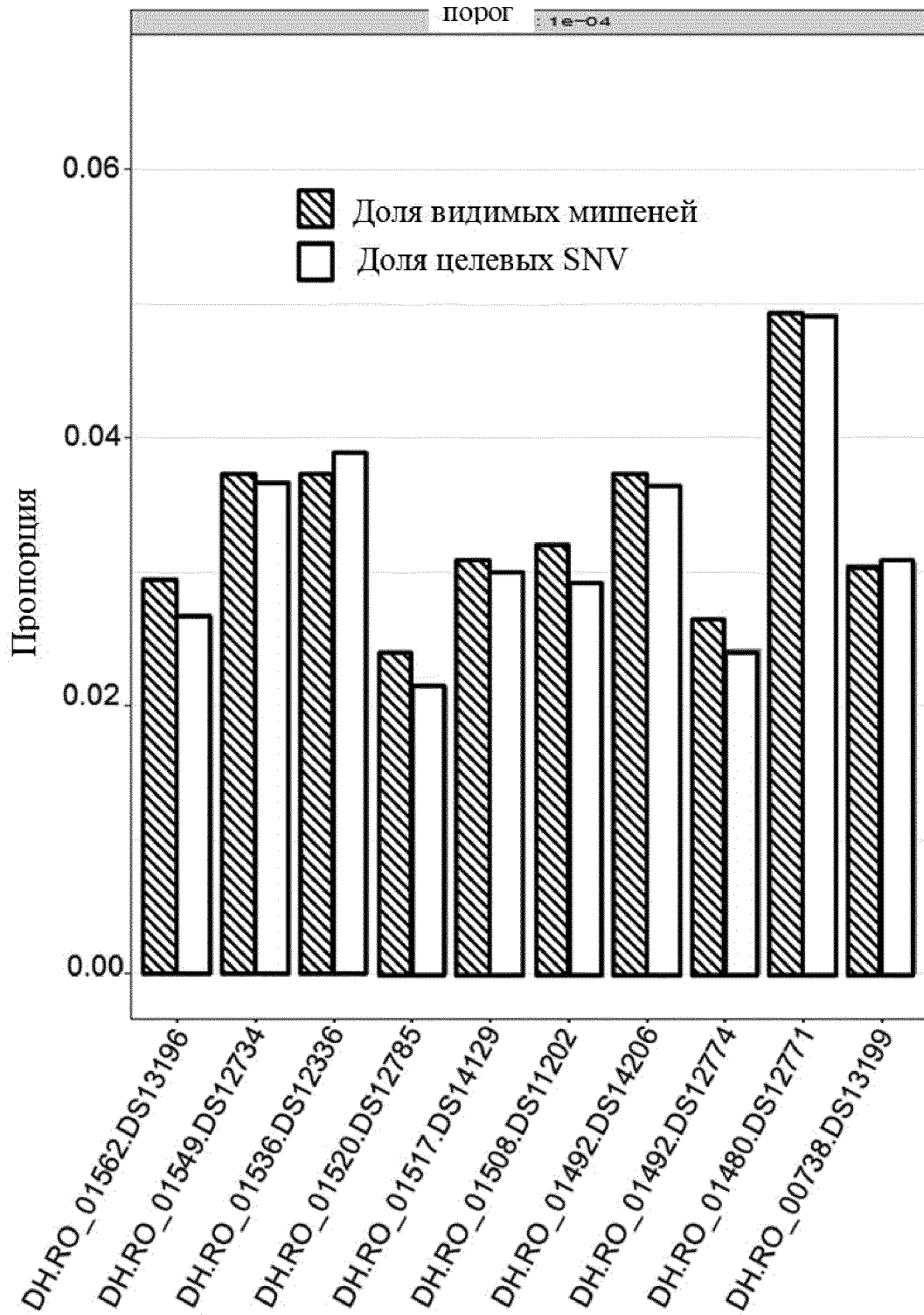
ФИГУРА 12А

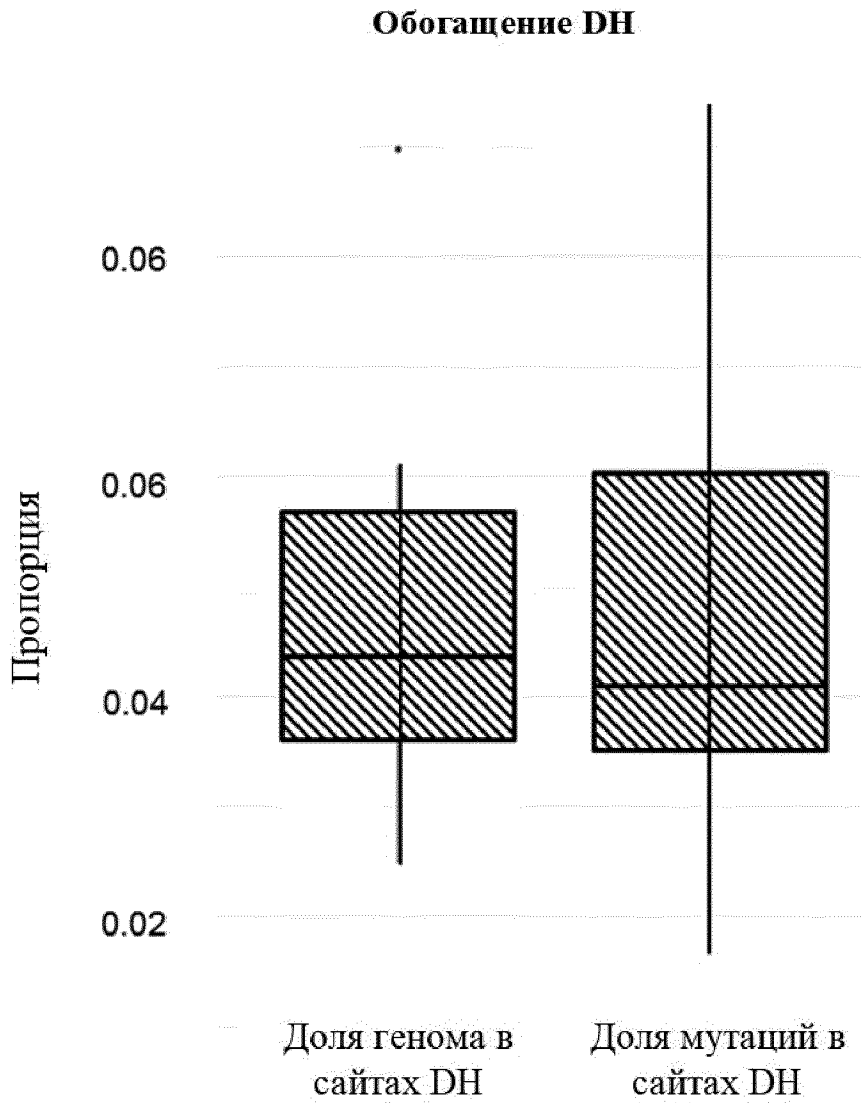
ФИГУРА 12В



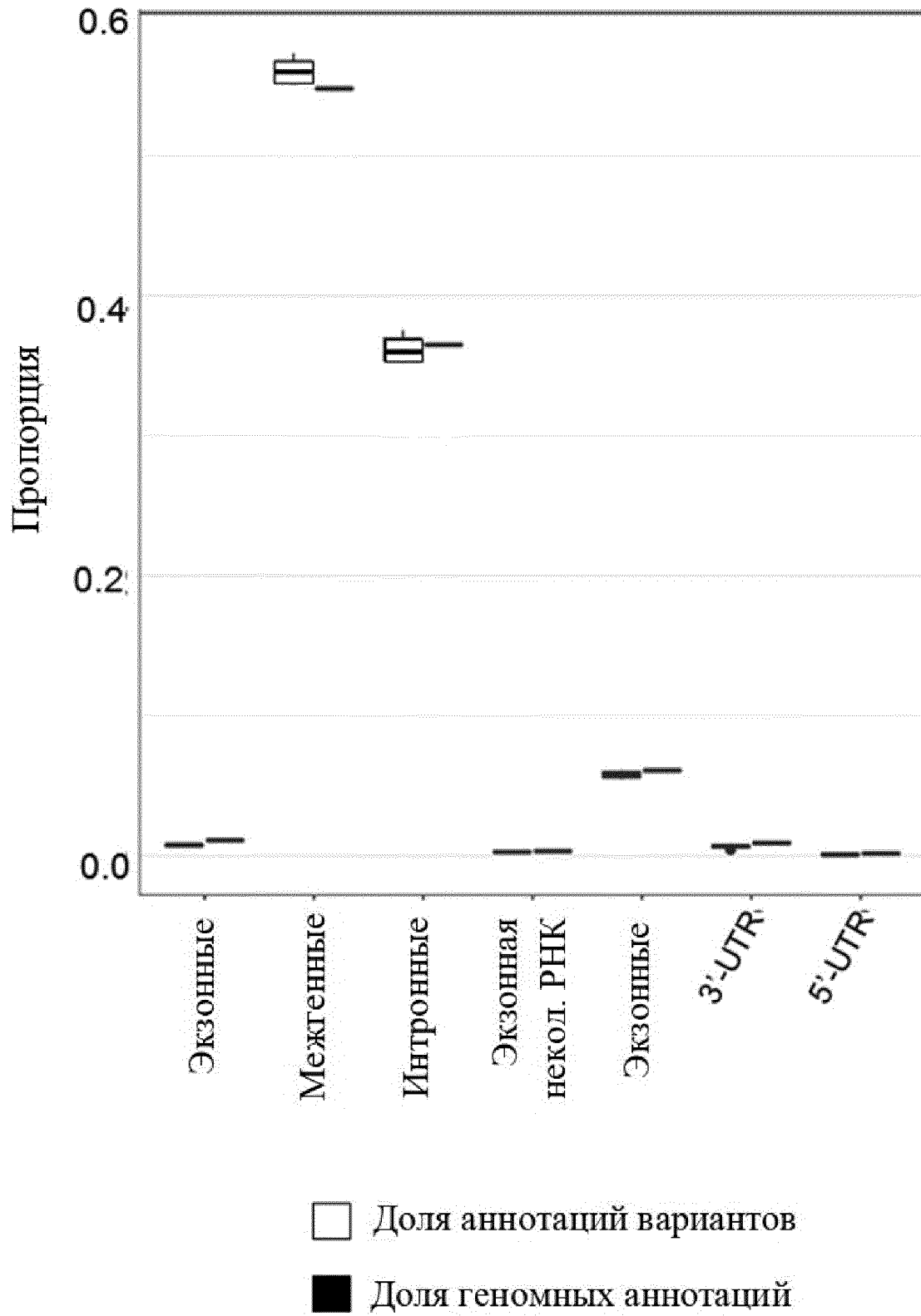


**ФИГУРА 12С**



**ФИГУРА 12Е**





ФИГУРА 12F