



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(52) СПК

C12N 15/1065 (2019.08); C12Q 1/6834 (2019.08); C12Q 1/6874 (2019.08); C12Q 2525/191 (2019.08); C12Q 2563/179 (2019.08); C12Q 2565/514 (2019.08)

(21)(22) Заявка: 2017116989, 16.10.2015

(24) Дата начала отсчета срока действия патента:
16.10.2015

Дата регистрации:
19.12.2019

Приоритет(ы):

(30) Конвенционный приоритет:
17.10.2014 US 62/065,544;
05.05.2015 US 62/157,396;
16.10.2015 US 62/242,880

(43) Дата публикации заявки: 19.11.2018 Бюл. № 32

(45) Опубликовано: 19.12.2019 Бюл. № 35

(85) Дата начала рассмотрения заявки РСТ на
национальной фазе: 17.05.2017

(86) Заявка РСТ:
US 2015/056040 (16.10.2015)

(87) Публикация заявки РСТ:
WO 2016/061517 (21.04.2016)

Адрес для переписки:
129090, Москва, ул. Большая Спасская, д. 25,
строение 3, ООО "Юридическая фирма
Городисский и Партнеры"

(72) Автор(ы):

БЕТЛИ Джейсон Ричард (GB),
ГУНДЕРСОН Кевин Л. (US),
ЧЖАН Фань (US),
МЕЛЕМАН Ваутер (US),
ГОРМЛИ Нил Энтони (GB),
ИОАННОУ Августа (GB),
УИР Жаклин (GB),
ДЖЕКСОН Розамонд (GB),
ДЖЕНКИНС Гарет (GB),
МОРРЕЛЛ Натали (GB),
ПОХОЛОК Дмитрий К. (US),
СТИМЕРС Фрэнк Дж. (US),
НОРБЕРГ Стивен Дж. (US),
ХИ Молли (US),
КИА Амирали (US),
ГОРЫШИН Игорь (US),
ПАНТОЯ Риго (US)

(73) Патентообладатель(и):

ИЛЛУМИНА КЕМБРИДЖ ЛИМИТЕД
(US)

(56) Список документов, цитированных в отчете
о поиске: WO 2012061832, 10.05.2012. WO
2014108810, 17.07.2014. WO 2012106546,
09.08.2012. WO 2010048605, 29.04.2010. EP
2712931, 02.04.2014.

(54) ТРАНСПОЗИЦИЯ С СОХРАНЕНИЕМ СЦЕПЛЕНИЯ ГЕНОВ

(57) Реферат:

Изобретение относится к области биотехнологии, конкретно к способу получения иммобилизованной библиотеки фрагментов ДНК нуклеиновой кислоты-мишени со штрихкодами с использованием иммобилизованных на носителе транспозомных комплексов для захвата

специфических фрагментов ДНК. Полученные иммобилизованные на твердом носителе библиотеки могут быть использованы для идентификации геномных вариантов и получения информации о сцеплении генов. 39 з.п. ф-лы, 75 ил., 6 табл., 21 пр.



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(51) Int. Cl.

C12N 15/10 (2006.01)*C12Q 1/6806* (2018.01)*C12Q 1/6874* (2018.01)*C12Q 1/6876* (2018.01)**(12) ABSTRACT OF INVENTION**

(52) CPC

C12N 15/1065 (2019.08); *C12Q 1/6834* (2019.08); *C12Q 1/6874* (2019.08); *C12Q 2525/191* (2019.08); *C12Q 2563/179* (2019.08); *C12Q 2565/514* (2019.08)

(21)(22) Application: **2017116989, 16.10.2015**

(24) Effective date for property rights:
16.10.2015

Registration date:
19.12.2019

Priority:

(30) Convention priority:
17.10.2014 US 62/065,544;
05.05.2015 US 62/157,396;
16.10.2015 US 62/242,880

(43) Application published: **19.11.2018 Bull. № 32**(45) Date of publication: **19.12.2019 Bull. № 35**(85) Commencement of national phase: **17.05.2017**

(86) PCT application:
US 2015/056040 (16.10.2015)

(87) PCT publication:
WO 2016/061517 (21.04.2016)

Mail address:
129090, Moskva, ul. Bolshaya Spasskaya, d. 25,
stroenie 3, OOO "Yuridicheskaya firma
Gorodisskij i Partnery"

(72) Inventor(s):

BETLI Dzhejson Richard (GB),
GUNDERSON Kevin L. (US),
CHZHAN Fan (US),
MELEMAN Vauter (US),
GORMLI Nil Entoni (GB),
IOANNOU Avgusta (GB),
UIR Zhaklin (GB),
DZHEKSON Rozamond (GB),
DZHENKINS Gareth (GB),
MORRELL Natali (GB),
POKHOLOK Dmitrij K. (US),
STIMERS Frenk Dzh. (US),
NORBERG Stiven Dzh. (US),
KHI Molli (US),
KIA Amirali (US),
GORYSHIN Igor (US),
PANTOYA Rigo (US)

(73) Proprietor(s):

ILLUMINA KEMBRIDZH LIMITED (US)

(54) TRANSPOSITION WITH PRESERVATION OF GENE ADHESION

(57) Abstract:

FIELD: biotechnology.

SUBSTANCE: invention relates to biotechnology, specifically to a method of producing an immobilized library of DNA fragments of a target nucleic acid with bar codes using immobilized on a carrier transposable

complexes for capturing specific DNA fragments.

EFFECT: prepared immobilized libraries on solid carrier can be used to identify genomic variants and obtain information on the adhesion of genes.

40 cl, 75 dwg, 6 tbl, 21 ex

Родственные заявки

В настоящей заявке испрашивается приоритет предварительной заявки на патент США No: 62/065544, поданной 17 октября, 2014, и предварительной заявки на патент США No: 62/157396, поданной 5 мая, 2015, которые во всей своей полноте вводятся в

5 настоящее описание посредством ссылки.

Область, к которой относится изобретение

Варианты настоящего изобретения относятся к секвенированию нуклеиновых кислот. В частности, описанные здесь варианты способов и композиций относятся к получению матриц на основе нуклеиновых кислот и к получению данных для этих

10 последовательностей.

Предшествующий уровень техники

Детектирование специфических последовательностей нуклеиновых кислот, присутствующих в биологическом образце, применяется, например, как способ идентификации и классификации микроорганизмов, диагностики инфекционных

15 заболеваний, детектирования и характеристики генетических аномалий, идентификации генетических модификаций, ассоциированных с развитием рака, исследования генетической восприимчивости к развитию заболевания и оценки ответа на различные типы лечения. Общим методом детектирования специфических последовательностей нуклеиновых кислот в биологическом образце является секвенирование нуклеиновых

20 кислот.

Методика секвенирования нуклеиновых кислот была разработана, главным образом, на основе методов химического разложения, применяемых Мэксэмом и Гилбертом, и метода удлинения цепи, используемого Сэнгером. В настоящее время применяется несколько методов секвенирования, которые позволяют осуществлять параллельный

25 процессинг всех нуклеиновых кислот в одном раунде секвенирования. При этом, информация, получаемая после проведения одного раунда секвенирования, может быть слишком объемной.

Описание сущности изобретения

В одном из аспектов настоящего изобретения описаны способы получения библиотеки

30 фрагментов ДНК нуклеиновой кислоты-мишени со штрихкодами. Эти способы включают контактирование нуклеиновой кислоты-мишени с множеством транспозомных комплексов, каждый из которых включает транспозоны и транспозазы, где транспозоны содержат перенесенные цепи и неперенесенные цепи. По меньшей мере один из транспозонов транспозомного комплекса включает последовательность адаптера,

35 способную гибридизоваться с комплементарной последовательностью для захвата.

Нуклеиновую кислоту-мишень фрагментируют с образованием множества фрагментов, а затем множество перенесенных цепей встраивают в 5'-конец по меньшей мере одной цепи фрагментов с сохранением сцепления нуклеиновой кислоты-мишени. Множество

40 фрагментов нуклеиновой кислоты-мишени подвергают контактированию с множеством твердых носителей, каждый из которых содержит множество иммобилизованных олигонуклеотидов, где каждый из этих олигонуклеотидов включает комплементарную последовательность для захвата и первую последовательность со штрихкодом, где первая последовательность со штрихкодом, присутствующая на каждом твердом носителе из множества твердых носителей, отличается от первой последовательности

45 со штрихкодом, присутствующей на других твердых носителях из множества твердых носителей. Информацию о последовательности со штрихкодом переносят во фрагменты нуклеиновой кислоты-мишени, в результате чего получают иммобилизованную библиотеку двухцепочечных фрагментов, где по меньшей мере одна цепь помечена у

5'-конца первым штрихкодом так, чтобы по меньшей мере два фрагмента одной и той же нуклеиновой кислоты-мишени получали идентичную информацию по штрихкоду.

В одном из аспектов изобретения описаны способы получения информации о сцеплении последовательности нуклеиновой кислоты-мишени. Эти способы включают

- 5 контактирование нуклеиновой кислоты-мишени с множеством транспозомных комплексов, каждый из которых включает транспозоны и транспозазы, где транспозоны содержат перенесенные цепи и неперенесенные цепи, в которых по меньшей мере один из транспозонов транспозомного комплекса включает последовательность адаптера, способную гибридизоваться с комплементарной последовательностью для захвата.
- 10 Нуклеиновую кислоту-мишень фрагментируют с образованием множества фрагментов, а затем множество перенесенных цепей встраивают во множество фрагментов с сохранением сцепления нуклеиновой кислоты-мишени. Множество фрагментов нуклеиновой кислоты-мишени подвергают контактированию с множеством твердых носителей. Каждый из этих твердых носителей содержит множество иммобилизованных
- 15 олигонуклеотидов, которые включают комплементарную последовательность для захвата и первую последовательность со штрихкодом, где первая последовательность со штрихкодом, присутствующая на каждом твердом носителе из множества твердых носителей, отличается от первой последовательности со штрихкодом, присутствующей на других твердых носителях из множества твердых носителей. Информацию о
- 20 последовательности со штрихкодом переносят во фрагменты нуклеиновой кислоты-мишени, так, чтобы по меньшей мере два фрагмента одной и той же нуклеиновой кислоты-мишени получали идентичную информацию по штрихкоду. Затем определяют последовательность фрагментов нуклеиновой кислоты-мишени и последовательностей со штрихкодом. Информацию о сцеплении нуклеиновой кислоты-мишени определяют
- 25 путем идентификации последовательностей со штрихкодом. В некоторых вариантах осуществления изобретения, транспозазы транспозомных комплексов удаляют после транспозиции и последующей гибридизации последовательностей адаптеров транспозона с комплементарной последовательностью для захвата. В некоторых вариантах осуществления изобретения, транспозазы удаляют путем ДСН-обработки. В некоторых
- 30 вариантах осуществления изобретения, транспозазы удаляют путем обработки протеиназой.

В одном из аспектов изобретения описаны способы получения одновременной информации о фазах и статусе метилирования последовательности нуклеиновой кислоты-мишени. Эти способы включают контактирование нуклеиновой кислоты-мишени с

- 35 множеством транспозомных комплексов, каждый из которых включает транспозоны и транспозазы, где транспозоны содержат перенесенные цепи и неперенесенные цепи, в которых по меньшей мере один из транспозонов транспозомного комплекса включает последовательность адаптера, способную гибридизоваться с комплементарной последовательностью для захвата. Нуклеиновую кислоту-мишень фрагментируют с
- 40 образованием множества фрагментов, а затем множество перенесенных цепей встраивают во фрагменты нуклеиновой кислоты-мишени с сохранением сцепления нуклеиновой кислоты-мишени. Множество фрагментов нуклеиновой кислоты-мишени подвергают контактированию с множеством твердых носителей, каждый из которых содержит множество иммобилизованных олигонуклеотидов, где каждый из этих
- 45 олигонуклеотидов включает комплементарную последовательность для захвата и первую последовательность со штрихкодом, и где первая последовательность со штрихкодом, присутствующая на каждом твердом носителе из множества твердых носителей, отличается от первой последовательности со штрихкодом, присутствующей

на других твердых носителях из множества твердых носителей. Информацию о последовательности со штрихкодом переносят во фрагменты нуклеиновой кислоты-мишени, так, чтобы по меньшей мере два фрагмента одной и той же нуклеиновой кислоты-мишени получали идентичную информацию по штрихкоду. Затем фрагменты нуклеиновой кислоты-мишени, содержащие штрихкоды, подвергают обработке бисульфитом, в результате чего получают обработанные бисульфитом фрагменты нуклеиновой кислоты-мишени, содержащие штрихкоды. После этого определяют последовательность обработанных бисульфитом фрагментов нуклеиновой кислоты-мишени и последовательности со штрихкодом. Информацию о сцеплении нуклеиновой кислоты-мишени определяют путем идентификации последовательностей со штрихкодом.

В одном из аспектов изобретения описаны способы получения иммобилизованной библиотеки меченных фрагментов ДНК. Эти способы включают получение множества твердых носителей, имеющих транспозомные комплексы, иммобилизованные на этих носителях, где указанные транспозомные комплексы являются мультимерными, а транспозомные мономерные единицы одного и того же транспозомного комплекса являются сцепленными друг с другом, и где указанные транспозомные мономерные единицы содержат транспозазу, связанную с первым полинуклеотидом, где указанный первый полинуклеотид включает (i) 3'-часть, содержащую концевую последовательность транспозона, и (ii) первый адаптер, содержащий первый штрихкод. ДНК-мишень наносят на множество твердых носителей в условиях, при которых ДНК-мишень фрагментируется транспозомными комплексами, а 3'-концевую последовательность транспозона первого полинуклеотида переносят в 5'-конец по меньшей мере одной цепи фрагментов, в результате чего получают иммобилизованную библиотеку двухцепочечных фрагментов, в которых по меньшей мере одна цепь помечена у 5'-конца первым штрихкодом.

В одном из аспектов изобретения описаны способы получения секвенирующей библиотеки для определения статуса метилирования нуклеиновой кислоты-мишени. Эти способы включают фрагментирование нуклеиновой кислоты-мишени с получением двух или более фрагментов. Первую общую последовательность адаптера вводят в 5'-конец фрагментов нуклеиновой кислоты-мишени, где указанная последовательность адаптера содержит первую последовательность, связывающуюся с праймером, и аффинную группу, где указанная аффинная группа является одним из членов связывающейся пары. Фрагмент нуклеиновой кислоты-мишени подвергают денатурации. Затем, фрагмент нуклеиновой кислоты-мишени иммобилизуют на твердом носителе, где указанный твердый носитель включает другой член связывающейся пары, и где иммобилизация нуклеиновой кислоты-мишени представляет собой связывание связывающейся пары. Иммобилизованные фрагменты нуклеиновой кислоты-мишени подвергают обработке дисульфитом. Вторую общую последовательность адаптера включают в иммобилизованные фрагменты нуклеиновой кислоты-мишени, обработанные бисульфитом, где второй общий адаптер включает второй сайт связывания с праймером. Обработанные бисульфитом фрагменты нуклеиновой кислоты-мишени, иммобилизованные на твердом носителе, подвергают амплификации и получают секвенирующую библиотеку для определения статуса метилирования нуклеиновой кислоты-мишени.

В одном из аспектов изобретения описаны способы получения секвенирующей библиотеки для определения статуса метилирования нуклеиновой кислоты-мишени. Эти способы включают получение множества твердых носителей, содержащих

транспозомные комплексы, иммобилизованные на этих носителях. Транспозомные комплексы включают транспозоны и транспозазы, где транспозоны содержат перенесенные цепи и неперенесенные цепи. Перенесенная цепь включает (i) первую часть, расположенную у 3'-конца и содержащую последовательность распознавания транспозазы, и (ii) вторую часть, расположенную у 5'-конца по отношению к первой части, содержащей первую последовательность адаптера и первый член связывающейся пары. Первый член связывающейся пары связывается со вторым членом связывающейся пары на твердом носителе, что приводит к иммобилизации транспозона на твердом носителе. Первый адаптер также включает первую последовательность, связывающуюся с праймером. Неперенесенная цепь включает (i) первую часть, расположенную у 5'-конца и содержащую последовательность распознавания транспозазы, и (ii) вторую часть, расположенную у 3'-конца по отношению к первой части, содержащей вторую последовательность адаптера, где концевой нуклеотид у 3'-конца является заблокированным. Второй адаптер также включает вторую последовательность, связывающуюся с праймером. Нуклеиновую кислоту-мишень подвергают контактированию со множеством твердых носителей, содержащих иммобилизованные транспозомные комплексы. Нуклеиновую кислоту-мишень фрагментируют с образованием множества фрагментов, и множество перенесенных цепей встраивают в 5'-конец по меньшей мере одной цепи фрагментов, что приводит к иммобилизации фрагментов нуклеиновой кислоты-мишени на твердом носителе. 3'-конец фрагментированной нуклеиновой кислоты-мишени удлиняют с использованием ДНК-полимеразы. Неперенесенную цепь лигируют с 3'-концом фрагментированной нуклеиновой кислоты-мишени. Затем иммобилизованные фрагменты нуклеиновой кислоты-мишени подвергают обработке бисульфитом. 3'-конец иммобилизованных фрагментов нуклеиновой кислоты-мишени, разрушенных в процессе обработки бисульфитом, удлиняют с использованием ДНК-полимеразы так, чтобы 3'-конец иммобилизованных фрагментов нуклеиновой кислоты-мишени содержал гомополимерный «хвост». Вторую последовательность адаптера вводят в 3'-конец иммобилизованных фрагментов нуклеиновой кислоты-мишени, разрушенных в процессе обработки бисульфитом. Обработанные бисульфитом фрагменты нуклеиновой кислоты-мишени, иммобилизованные на твердом носителе, подвергают амплификации с использованием первого и второго праймера, в результате чего получают секвенирующую библиотеку для определения статуса метилирования нуклеиновой кислоты-мишени.

В одном из аспектов изобретения описаны способы получения секвенирующей библиотеки для определения статуса метилирования нуклеиновой кислоты-мишени. Эти способы включают контактирование нуклеиновой кислоты-мишени с транспозомными комплексами, содержащими транспозоны и транспозазы. Транспозоны содержат перенесенные цепи и неперенесенные цепи. Перенесенная цепь включает (i) первую часть, расположенную у 3'-конца и содержащую последовательность распознавания транспозазы, и (ii) вторую часть, расположенную у 5'-конца по отношению к первой части, содержащей первую последовательность адаптера и первый член связывающейся пары, где первый член связывающейся пары связывается со вторым членом связывающейся пары. Неперенесенная цепь включает (i) первую часть, расположенную у 5'-конца и содержащую последовательность распознавания транспозазы, и (ii) вторую часть, расположенную у 3'-конца по отношению к первой части, содержащей вторую последовательность адаптера, где концевой нуклеотид у 3'-конца является заблокированным, и где второй адаптер включает вторую

последовательность, связывающуюся с праймером. Нуклеиновую кислоту-мишень фрагментируют с образованием множества фрагментов, и множество перенесенных цепей встраивают в 5'-конец по меньшей мере одной цепи фрагментов, что приводит к иммобилизации фрагментов нуклеиновой кислоты-мишени на твердом носителе.

- 5 Фрагменты нуклеиновой кислоты-мишени, содержащие транспозонный конец, подвергают контактированию со множеством твердых носителей, содержащих второй член связывающейся пары, где связывание первого члена связывающейся пары со вторым членом связывающейся пары приводит к иммобилизации нуклеиновой кислоты-мишени на твердом носителе. 3'-конец фрагментированной нуклеиновой кислоты-мишени удлиняют с использованием ДНК-полимеразы. Неперенесенную цепь лигируют с 3'-концом фрагментированной нуклеиновой кислоты-мишени. Иммобилизованные фрагменты нуклеиновой кислоты-мишени подвергают обработке бисульфитом. 3'-конец иммобилизованных фрагментов нуклеиновой кислоты-мишени, разрушенных в процессе обработки бисульфитом, удлиняют с использованием ДНК-полимеразы так, чтобы 3'-конец иммобилизованных фрагментов нуклеиновой кислоты-мишени содержал гомополимерный «хвост». Вторую последовательность адаптера вводят в 3'-конец иммобилизованных фрагментов нуклеиновой кислоты-мишени, разрушенных в процессе обработки бисульфитом. Обработанные бисульфитом фрагменты нуклеиновой кислоты-мишени, иммобилизованные на твердом носителе, подвергают амплификации с использованием первого и второго праймера, в результате чего получают секвенирующую библиотеку для определения статуса метилирования нуклеиновой кислоты-мишени.

В некоторых вариантах осуществления изобретения, концевой нуклеотид у 3'-конца второго адаптера блокируют членом, выбранным из группы, состоящей из дидезокси-нуклеотида, фосфатной группы, тиофосфатной группы и азидогруппы.

В некоторых вариантах осуществления изобретения, аффинные молекулы могут быть членами связывающейся пары. В некоторых случаях, модифицированные нуклеиновые кислоты могут содержать первый член связывающейся пары, а зонд для захвата может содержать второй член связывающейся пары. В некоторых случаях, зонды для захвата могут быть иммобилизованы на твердой поверхности, где модифицированная нуклеиновая кислота может содержать первый член связывающейся пары, а зонд для захвата может содержать второй член связывающейся пары. В таких случаях, связывание первого и второго членов связывающейся пары приводит к иммобилизации модифицированной нуклеиновой кислоты на твердой поверхности.

35 Примерами связывающихся пар являются, но не ограничиваются ими, биотин-авидин, биотин-стрептавидин, биотин-нейтравидин, лиганд-рецептор, гормон-рецептор, лектин-гликопротеин, олигонуклеотид-комплементарный олигонуклеотид и антиген-антитело.

В некоторых вариантах осуществления изобретения, первую общую последовательность адаптера встраивают в 5'-конец фрагментов нуклеиновой кислоты-мишени посредством односторонней транспозиции. В некоторых вариантах осуществления изобретения, первую общую последовательность адаптера встраивают в 5'-конец фрагментов нуклеиновой кислоты-мишени посредством лигирования. В некоторых вариантах осуществления изобретения, введение второй общей последовательности адаптера в обработанные бисульфитом иммобилизованные фрагменты нуклеиновой кислоты-мишени включает (i) удлинение 3'-конца иммобилизованных фрагментов нуклеиновой кислоты-мишени с использованием концевой трансферазы так, чтобы они содержали гомополимерный «хвост»; (ii) гибридизацию олигонуклеотида, содержащего одноцепочечную гомополимерную часть

и двухцепочечную часть, включающую вторую общую последовательность адаптера, где одноцепочечная гомополимерная часть комплементарна гомополимерному «хвосту»; и (iii) лигирование второй общей последовательности адаптера с иммобилизованными фрагментами нуклеиновой кислоты-мишени для включения второй общей последовательности адаптера в обработанные бисульфитом иммобилизованные фрагменты нуклеиновой кислоты-мишени.

В некоторых вариантах осуществления изобретения, нуклеиновую кислоту-мишень выделяют из одной клетки. В некоторых вариантах осуществления изобретения, нуклеиновую кислоту-мишень выделяют из одной органеллы. В некоторых вариантах осуществления изобретения, нуклеиновой кислотой-мишенью является геномная ДНК. В некоторых вариантах осуществления изобретения, нуклеиновую кислоту-мишень подвергают перекрестному связыванию с другими нуклеиновыми кислотами. В некоторых вариантах осуществления изобретения, нуклеиновую кислоту-мишень выделяют из залитого в парафин образца, фиксированного формалином (FFPE). В некоторых вариантах осуществления изобретения, нуклеиновую кислоту-мишень подвергают перекрестному связыванию с белками. В некоторых вариантах осуществления изобретения, нуклеиновую кислоту-мишень подвергают перекрестному связыванию с ДНК. В некоторых вариантах осуществления изобретения, нуклеиновой кислотой-мишенью является ДНК, защищенная гистоном. В некоторых вариантах осуществления изобретения, гистоны удаляют из нуклеиновой кислоты-мишени. В некоторых вариантах осуществления изобретения, нуклеиновой кислотой-мишенью является неклеточная опухолевая ДНК. В некоторых вариантах осуществления изобретения, неклеточную опухолевую ДНК выделяют из плацентарной жидкости. В некоторых вариантах осуществления изобретения, неклеточную опухолевую ДНК выделяют из плазмы. В некоторых вариантах осуществления изобретения, плазму выделяют из цельной крови с использованием мембранного сепаратора, имеющего зону сбора плазмы. В некоторых вариантах осуществления изобретения, зона для сбора плазмы включает транспозомные комплексы, иммобилизованные на твердом носителе. В некоторых вариантах осуществления изобретения, нуклеиновой кислотой-мишенью является кДНК. В некоторых вариантах осуществления изобретения, твердым носителем является сфера. В некоторых вариантах осуществления изобретения, множество твердых носителей представляет собой множество сфер, имеющих различные размеры.

В некоторых вариантах осуществления изобретения, одна последовательность со штрихкодом присутствует во множестве олигонуклеотидов, иммобилизованных на каждом отдельном твердом носителе. В некоторых вариантах осуществления изобретения, различные последовательности со штрихкодом присутствуют во множестве олигонуклеотидов, иммобилизованных на каждом отдельном твердом носителе. В некоторых вариантах осуществления изобретения, перенос информации о последовательности со штрихкодом во фрагменты нуклеиновой кислоты-мишени осуществляют путем лигирования. В некоторых вариантах осуществления изобретения, перенос информации о последовательности со штрихкодом во фрагменты нуклеиновой кислоты-мишени осуществляют путем удлинения с использованием полимеразы. В некоторых вариантах осуществления изобретения, перенос информации о последовательности со штрихкодом во фрагменты нуклеиновой кислоты-мишени осуществляют путем лигирования и удлинения с использованием полимеразы. В некоторых вариантах осуществления изобретения, удлинение с использованием полимеразы осуществляют посредством удлинения 3'-конца нелигированной транспозонной цепи посредством ДНК-полимеразы с использованием лигированного

иммобилизованного олигонуклеотида в качестве матрицы. В некоторых вариантах осуществления изобретения, по меньшей мере часть последовательностей адаптера также содержит вторую последовательность со штрихкодом.

В некоторых вариантах осуществления изобретения, транспозомные комплексы являются мультимерными, где последовательности адаптера транспозонов каждой мономерной единицы отличаются от последовательностей другой мономерной единицы, присутствующей в том же самом транспозомном комплексе. В некоторых вариантах осуществления изобретения, последовательность адаптера также включает первую последовательность, связывающуюся с праймером. В некоторых вариантах осуществления изобретения, первый сайт связывания с праймером имеет последовательность, которая не является гомологичной последовательности для захвата или ее комплементу. В некоторых вариантах осуществления изобретения, олигонуклеотиды, иммобилизованные на твердом носителе, также содержат вторую последовательность, связывающуюся с праймером.

В некоторых вариантах осуществления изобретения, транспозомные комплексы являются мультимерными, где транспозонные мономерные единицы связаны друг с другом в одном и том же транспозомном комплексе. В некоторых вариантах осуществления изобретения, транспозаза транспозомной мономерной единицы связана с транспозазой другой транспозомной мономерной единицы одного и того же транспозомного комплекса. В некоторых вариантах осуществления изобретения, транспозоны транспозомной мономерной единицы связаны с транспозонами другой транспозомной мономерной единицы одного и того же транспозомного комплекса. В некоторых вариантах осуществления изобретения, транспозаза транспозомной мономерной единицы связана с транспозазой другой транспозомной мономерной единицы одного и того же транспозомного комплекса посредством ковалентной связи. В некоторых вариантах осуществления изобретения, транспозазы одной мономерной единицы связаны с транспозазой другой транспозомной мономерной единицы одного и того же транспозомного комплекса посредством дисульфидной связи. В некоторых вариантах осуществления изобретения, транспозоны транспозомной мономерной единицы связаны с транспозонами другой транспозомной мономерной единицы одного и того же транспозомного комплекса посредством ковалентной связи.

В некоторых вариантах осуществления изобретения, информация о сцеплении последовательности нуклеиновой кислоты-мишени представляет собой информацию о гаплотипе. В некоторых вариантах осуществления изобретения, информация о сцеплении последовательности нуклеиновой кислоты-мишени представляет информацию о геномных вариантах. В некоторых вариантах осуществления изобретения, геномные варианты выбраны из группы, состоящей из делеций, транслокаций, межхромосомных сцеплений генов, дупликаций и паралогов. В некоторых вариантах осуществления изобретения, олигонуклеотиды, иммобилизованные на твердом носителе, содержат частично двухцепочечную область и частично одноцепочечную область. В некоторых вариантах осуществления изобретения, частично одноцепочечная область олигонуклеотида содержит вторую последовательность со штрихкодом и вторую последовательность, связывающуюся с праймером. В некоторых вариантах осуществления изобретения, фрагменты нуклеиновой кислоты-мишени, имеющие штрихкоды, амплифицируют, а затем определяют последовательность фрагментов нуклеиновой кислоты-мишени. В некоторых вариантах осуществления изобретения, последующую амплификацию проводят в одном реакционном сосуде до определения последовательности фрагментов нуклеиновой кислоты-мишени. В некоторых вариантах

осуществления изобретения, третью последовательность со штрихкодом вводят во фрагменты нуклеиновой кислоты-мишени во время амплификации.

В некоторых вариантах осуществления изобретения, способы могут также включать объединение фрагментов нуклеиновой кислоты-мишени, содержащих штрихкоды из первой серии множества реакционных сосудов, в пул фрагментов нуклеиновой кислоты-мишени, содержащих штрихкоды; перераспределение пула фрагментов нуклеиновой кислоты-мишени, содержащих штрихкоды, с получением множества второй серии множества реакционных сосудов; и введение третьего штрихкода во фрагменты нуклеиновой кислоты-мишени путем амплификации фрагментов нуклеиновой кислоты-мишени во вторую серию множества реакционных сосудов до секвенирования.

В некоторых вариантах осуществления изобретения, способы могут также включать предварительное фрагментирование нуклеиновой кислоты-мишени до контактирования нуклеиновой кислоты-мишени с транспозомными комплексами. В некоторых вариантах осуществления изобретения, предварительное фрагментирование нуклеиновой кислоты-мишени представляет собой способ, выбранный из группы, состоящей из обработки ультразвуком и расщепления рестриктивными ферментами.

Краткое описание чертежей

На фигуре 1 представлена блок-схема, на которой проиллюстрирован метод связывания транспозом с поверхностью сфер.

На фигуре 2 проиллюстрированы стадии способа, описанного на фигуре 1.

На фигуре 3 схематически представлена диаграмма, иллюстрирующая способ тагментации на поверхности сфер.

На фигуре 4 представлена таблица данных о выходах ДНК для различных кластеров в способе тагментации на сферах, как показано на фигуре 3.

На фигуре 5 представлена таблица данных, полученных для другого примера воспроизводимости способа тагментации на сферах одинакового размера как показано на фигуре 3.

На фигурах 6А и 6В представлен график размера вставок для пула 1 и график размера вставок для пула 2 индексированных образцов на фигуре 5, соответственно.

На фигуре 7 представлена гистограмма воспроизводимости общего числа ридов и процента ридов, выровненных для проведения эксперимента, как показано на фигуре 5.

На фигурах 8А, 8В и 8С представлен график данных о размерах вставки в контрольной библиотеке, график данных о размерах вставки в тагментированной библиотеке на сферах, а также представлена таблица, в которой систематизированы данные, полученные в анализе на обогащение экзосом, соответственно.

На фигурах 9А, 9В и 9С представлены гистограмма дубликатов PF, гистограмма для фракции отобранных оснований и гистограмма для оснований, используемых в РСТ на мишени, соответственно, в анализе на обогащение экзосом.

На фигуре 10 представлена блок-схема, иллюстрирующая репрезентативный метод продуцирования транспозомных комплексов на поверхности сфер.

На фигурах 11, 12 и 13 проиллюстрированы стадии метода, описанного на фигуре 10.

На фигуре 14 схематически представлена диаграмма способа тагментации, проводимого с использованием сферы, покрытой транспозомой как показано на фигуре 13.

На фигуре 15 представлена репрезентативная схема образования транспозом на твердом носителе.

На фигуре 16 представлена репрезентативная схема получения библиотек сцепленных

нуклеиновых кислот с уникальными индексами.

На фигуре 17 представлена репрезентативная схема получения библиотек сцепленных нуклеиновых кислот с уникальными индексами.

На фигурах 18 и 19 проиллюстрированы захват одной СРТ-ДНК на одной клональной индексированной сфере, где СРТ-ДНК обертывает эту сферу.

На фигуре 20 схематически проиллюстрировано связывание Y-адаптера, иммобилизованного на твердой поверхности, с тагментированной ДНК посредством лигирования и заполнения брешей.

На фигуре 21 схематически проиллюстрировано получение таких Y-адаптеров в процессе лигирования СРТ-DNA с олигонуклеотидами, иммобилизованными на твердом носителе.

На фигуре 22 проиллюстрирован электрофорез в агарозном геле для удаления свободной транспозомы из библиотек сцепленных нуклеиновых кислот с помощью эксклюзионной хроматографии.

На фигуре 23 представлена репрезентативная схема получения библиотеки последовательностей специфических фрагментов ДНК методом «дробовика».

На фигуре 24 представлена репрезентативная схема сбора информации о последовательности клональной индексированной секвенирующей библиотеки.

На фигуре 25 представлены результаты оптимизации плотности зондов для захвата на сферах.

На фигуре 26 представлены результаты теста на возможность получения индексированных секвенирующих библиотек СРТ-ДНК на сферах путем внутримолекулярной гибридизации.

На фигуре 27 представлены результаты теста на возможность клонального индексирования.

На фигуре 28 представлен график, на котором проиллюстрирована частота секвенирующих ридов для конкретных расстояний в островках (внутри островков), а также между соседними выровненными островками ридов для матричной нуклеиновой кислоты после тагментации.

На фигурах 29А и 29В представлены репрезентативные методы получения информации о сцеплении генов на твердом носителе.

На фигурах 30 и 31 схематически представлены транспозиция индексированных клонированных сфер в одном реакционном сосуде (в одном сосуде) и результаты транспозиции.

На фигуре 32 схематически проиллюстрировано создание клональных транспозом на сферах с использованием 5'- или 3'-биотинилированных олигонуклеотидов.

На фигуре 33 показаны размеры библиотеки для транспозом на сферах.

На фигуре 34 показано влияние поверхностной плотности транспозом на размер вставки.

На фигуре 35 показано влияние исходной ДНК на размер распределения.

На фигуре 36 показаны размеры островков и распределение, достигаемое посредством реакции тагментации на сферах и в растворе.

На фигуре 37 проиллюстрировано клональное индексирование нескольких отдельных молекул ДНК, каждая из которых имеет уникальные индексы.

На фигуре 38 представлена диаграмма для устройства, применяемого в целях выделения плазмы из цельной крови.

На фигурах 39 и 40 представлена диаграмма для устройства, применяемого в целях выделения плазмы и последующего использования выделенной плазмы.

На **фигуре 41** представлена репрезентативная схема целевого фазирования посредством обогащения специфических областей генома.

На **фигуре 42** представлена репрезентативная схема фазирования экзона с использованием SNP, расположенных между экзонами.

5 На **фигуре 43** представлена репрезентативная схема одновременного фазирования и детектирования метилирования.

На **фигуре 44** представлена альтернативная репрезентативная схема одновременного фазирования и детектирования метилирования.

10 На **фигуре 45** представлена репрезентативная схема получения библиотек различных размеров с использованием клонально индексированных сфер различных размеров в одном анализе.

На **фиг. 46** представлена репрезентативная схема определения генетических вариантов библиотек, имеющих длины различных масштабов.

15 На **фиг. 47 А и В** представлены результаты детектирования гомозиготной 60 т.п.о.- делеции в хромосоме 1.

На **фиг. 48** представлены результаты детектирования сцепления генов способами согласно изобретению.

На **фиг. 49** представлены результаты детектирования генетических делеций способами согласно изобретению.

20 На **фиг. 50** представлены МЕ-последовательности до и после превращения под действием бисульфита.

На **фиг. 51** представлены результаты оптимизации эффективности превращения под действием бисульфита.

25 На **фиг. 52** представлены результаты превращения под действием бисульфита на графике IVC (на графике зависимости интенсивности от циклов на одно основание).

На **фиг. 53** представлены изображения, полученные с помощью электрофореза в агарозном геле для сцепленных с индексами библиотек после проведения ПЦР, осуществляемой после превращения под действием бисульфита (BSC).

30 На **фиг. 54** проиллюстрировано мечение сцепленных с индексами библиотек CPT-seq с помощью биоанализатора после обогащения этих библиотек без отбора по размеру.

На **фиг. 55** проиллюстрирован анализ библиотек в агарозном геле после обогащения.

На **фиг. 56** представлены результаты применения целевого гаплотипирования в области HLA в хромосоме.

На **фиг. 57** проиллюстрированы некоторые возможные механизмы замены МЕ.

35 На **фиг. 58** проиллюстрированы некоторые возможные механизмы замены МЕ.

На **фиг. 59** представлена часть транспозазы Tn5, в которой репрезентативные аминокислотные остатки Asp468, Tyr407, Asp461, Lys459, Ser458, Gly462, Ala466, Met470 могут быть заменены Cys.

40 На **фиг. 60** представлена часть транспозазы Tn5 с аминокислотными заменами S458C, K459C и A466C, введенными так, чтобы цистеиновые остатки могли образовывать дисульфидную связь между двумя мономерными единицами.

На **фиг. 61** представлена репрезентативная схема получения и использования биоконъюгата «димерная транспозаза (dTnp)-наночастица (NP)» (dTnp-NP) посредством наночастиц, покрытых амином.

45 На **фиг. 62** представлена репрезентативная схема конъюгирования транспозомного димера с твердым носителем, покрытым амином.

На **фиг. 63** представлен транспозомный комплекс Mu, в котором транспозонные концы являются сцепленными.

На фиг. 64 представлена диаграмма индексированных сцепленных ридов для сборки/фазирования псевдогенов и показано преимущество идентификации вариантов в псевдогене с использованием коротких фрагментов.

На фиг. 65 проиллюстрирован график замен индексов для 4 отдельных экспериментов, где такая замена представлена как % замененных индексов.

На фиг. 66 проиллюстрирован анализ размеров фрагментов, проведенный посредством титрования Ts-Tn5 на биоанализаторе Agilent BioAnalyzer.

На фиг. 67 представлена репрезентативная схема повышения выхода ДНК в соответствии с протоколом Epi-CPTSeq с применением ферментативных методов восстановления разрушенных элементов библиотеки после обработки бисульфитом.

На фиг. 68 А-С представлено несколько репрезентативных схем повышения выхода ДНК в соответствии с протоколом Epi-CPTSeq с применением ферментативных методов восстановления разрушенных элементов библиотеки после обработки бисульфитом.

На фиг. 69 представлена репрезентативная схема «спасения» матрицы методом рандомизированного удлинения праймера.

На фиг. 70 проиллюстрирована фрагментация библиотеки ДНК в процессе реакции превращения с использованием бисульфата натрия. На левой панели проиллюстрирована фрагментация в процессе превращения части ДНК, тагментированной на магнитных сферах, посредством бисульфата. На правой панели показаны следовые количества библиотек CPTSeq и Epi-CPTSeq (Me-CPTSeq), оцененные с помощью биоанализатора BioAnalyzer.

На фиг. 71 представлена репрезентативная схема и результаты TdT-опосредуемой реакции лигирования оцДНК.

На фиг. 72 представлены схемы и результаты TdT-опосредуемого восстановления связанной со сферой библиотеки после превращения посредством бисульфата натрия. На левой панели проиллюстрирована технологическая схема «спасения» библиотеки ДНК посредством бисульфитного превращения с использованием TdT-опосредуемой реакции лигирования. Результаты эксперимента по «спасению» библиотеки ДНК представлены на правой панели.

На фиг. 73 представлены результаты анализа на метилирование CPTSeq.

На фиг. 74 представлена репрезентативная схема превращения ДНК, присутствующей на сферах, посредством бисульфита.

На фиг. 75 А-В представлены результаты оптимизации эффективности после превращения посредством бисульфита.

Подробное описание изобретения

Варианты настоящего изобретения относятся к секвенированию нуклеиновых кислот. В частности, описанные здесь варианты способов и композиций относятся к получению матриц на основе нуклеиновых кислот и к получению данных для этих последовательностей.

В одном из своих аспектов, настоящее изобретение относится к способам тагментации (фрагментирования и мечения) нуклеиновой кислоты-мишени на твердом носителе для конструирования тагментированной библиотеки нуклеиновой кислоты-мишени. В одном из вариантов осуществления изобретения, твердым носителем являются сферы. В одном из вариантов осуществления изобретения, нуклеиновой кислоты-мишенью является ДНК.

В одном из своих аспектов, настоящее изобретение относится к способам и композициям твердого носителя, а также к разработанным на основе транспозазы способам, которые позволяют получить информацию о сцеплении нуклеиновой кислоты-

мишени. В некоторых вариантах осуществления изобретения, указанные композиции и способы позволяют получить информацию о сборке/фазировании.

В одном из своих аспектов, настоящее изобретение относится к способам и композициям, применяемым для получения информации о сцеплении генов посредством захвата сцепленных и перенесенных нуклеиновых кислот-мишеней на твердом носителе.

В одном из аспектов изобретения, описанные здесь композиции и способы применяются для анализа геномных вариантов. Репрезентативными геномными вариантами являются, но не ограничиваются ими, делеции, межхромосомные транслокации, дубликации, паралоги, межхромосомные сцепления генов. В некоторых вариантах осуществления изобретения, описанные здесь композиции и способы применяются для получения информации о фазировании геномных вариантов.

В одном из аспектов изобретения, описанные здесь композиции и способы применяются для фазирования специфических областей нуклеиновой кислоты-мишени. В одном из вариантов осуществления изобретения, нуклеиновой кислотой-мишенью является ДНК. В одном из вариантов осуществления изобретения, нуклеиновой кислотой-мишенью является геномная ДНК. В некоторых вариантах осуществления изобретения, нуклеиновой кислотой-мишенью является РНК. В некоторых вариантах осуществления изобретения, РНК представляет собой мРНК. В некоторых вариантах осуществления изобретения, нуклеиновой кислотой-мишенью является комплементарная ДНК (кДНК). В некоторых вариантах осуществления изобретения, нуклеиновая кислота-мишени выделена из одной клетки. В некоторых вариантах осуществления изобретения, нуклеиновая кислота-мишени выделена из опухолевых клеток кровотока. В некоторых вариантах осуществления изобретения, нуклеиновой кислотой-мишенью является неклеточная ДНК. В некоторых вариантах осуществления изобретения, нуклеиновой кислотой-мишенью является неклеточная опухолевая ДНК. В некоторых вариантах осуществления изобретения, нуклеиновая кислота-мишень выделена из залитых из залитых в парафин образцов ткани, фиксированных формалином. В некоторых вариантах осуществления изобретения, нуклеиновой кислотой-мишенью является перекрестно связанная нуклеиновая кислота-мишень. В некоторых вариантах осуществления изобретения, нуклеиновая кислота-мишень перекрестно связана с белками. В некоторых вариантах осуществления изобретения, нуклеиновая кислота-мишень перекрестно связана с нуклеиновой кислотой. В некоторых вариантах осуществления изобретения, нуклеиновой кислотой-мишенью является ДНК, защищенная гистоном. В некоторых вариантах осуществления изобретения, защищенную гистоном ДНК осаждают из клеточного лизата с использованием антител против гистонов, а затем гистоны удаляют.

В некоторых аспектах изобретения, индексированные библиотеки создают из нуклеиновой кислоты-мишени с использованием клонально индексированных сфер. В некоторых вариантах осуществления изобретения, тагментированная нуклеиновая кислота-мишень, в случае когда ДНК-мишень еще связана с транспозазой, может быть захвачена с использованием клонально индексированных сфер. В некоторых вариантах осуществления изобретения, специфические зонды для захвата используют в целях захвата специфической представляющей интерес области в нуклеиновой кислоте-мишени. Захваченные области нуклеиновой кислотой-мишени могут быть промыты в различных условиях жесткости, а затем амплифицированы, но необязательно, и секвенированы. В некоторых вариантах осуществления изобретения, зонд для захвата может быть биотинилированным. Комплекс биотинилированных зондов для захвата, гибридизованных со специфическими областями индексированных нуклеиновых кислот-

мишеней, может быть разделен с использованием стрептавидиновых сфер.

Репрезентативная схема целевого фазирования представлена на фиг. 41.

В некоторых аспектах изобретения, описанные здесь композиции и способы могут быть применены для фазирования экзонов. В некоторых вариантах осуществления изобретения, экзоны и промоторы могут быть обогащены. Маркеры, например, гетерозиготные SNP, находящиеся между экзонными областями, могут облегчать фазирование экзонов, а в частности, если между экзонами имеется большое расстояние. Репрезентативное фазирование экзонов проиллюстрировано на фиг. 42. В некоторых вариантах осуществления изобретения, индексированные сцепленные риды не могут одновременно охватывать (покрывать) гетерозиготные SNP соседних экзонов. Таким образом, это затрудняет фазирование двух или более экзонов. Описанные здесь композиции и способы также применяются для обогащения гетерозиготных SNP, расположенных между экзонами, например, для фазирования экзона 1 на SNP1, и SNP2 на экзон 2. Таким образом, с использованием SNP 1 может быть проведено фазирование экзона 1 и экзона 2, как показано на фиг. 42.

В одном из аспектов изобретения, описанные здесь композиции и способы могут быть применены для одновременного фазирования и детектирования метилирования. Детектирование метилирования посредством превращения в результате обработки бисульфитом (BSC) является проблематичным, поскольку реакция BSC является слишком жесткой для ДНК, то есть, приводит к фрагментации ДНК, а поэтому она удаляет информацию о сцеплении/фазировании. Способы, описанные в настоящей заявке, имеют дополнительное преимущество, заключающееся в том, что при их проведении не требуется дополнительной стадии очистки как в традиционных методах BSC, и это способствует увеличению выхода.

В одном из аспектов изобретения, описанные здесь композиции и способы могут быть применены для получения библиотек различных размеров в одном анализе. В некоторых вариантах осуществления изобретения, клонально индексированные сферы различных размеров могут быть использованы для получения библиотек различных размеров. На фигуре 1 проиллюстрирована блок-схема метода 100, применяемого для связывания транспозом с поверхностью сфер. Транспозомы могут быть присоединены к поверхности сфер любым химическим методом, который может быть осуществлен на олигонуклеотиде транспозона, транспозазе и твердой фазе. В одном из примеров, транспозомы связаны с поверхностью сфер посредством комплекса «биотин-стрептавидин». Метод 100 включает, но не ограничивается ими, нижеследующие стадии.

В одном из вариантов осуществления изобретения, транспозоны могут содержать секвенирующие сайты связывания с праймером. Репрезентативными последовательностями сайтов связывания с последовательностью являются, но не ограничиваются ими, AATGATACGGCGACCACCGAGATCTACAC (последовательность P5) и CAAGCAGAAGACGGCATACGAGAT (последовательность P7). В некоторых вариантах осуществления изобретения, транспозоны могут быть биотинилированными.

В стадии 110, показанной на фигуре 1, были получены биотинилированные транспозоны P5 и P7. Эти транспозоны могут также включать одну или более индексных последовательностей (уникальный идентификатор). Репрезентативными индексными последовательностями являются, но не ограничиваются ими, TAGATCGC, CTCTCTAT, TATCCTCT, AGAGTAGA, GTAAGGAG, ACTGCATA, AAGGAGTA, СТАAGCCT. В другом примере, биотинилированными являются только транспозоны P5 или только транспозоны P7. В другом примере, транспозоны включают только мозаичные концевые (ME) последовательности или последовательности ME плюс дополнительные

последовательности, которые не являются последовательностями Р5 и Р7. В этом примере, последовательности Р5 и Р7 добавляют в последующей стадии ПЦР-амплификации.

В стадии 115, показанной на фигуре 1, была осуществлена сборка транспозом. Собранные транспозомы представляют собой смесь транспозом Р5 и Р7. Смесь транспозом Р5 и Р7 более подробно описаны со ссылкой на фигуры 11 и 12.

В стадии 120, показанной на фигуре 1, смеси транспозом Р5/Р7 связаны с поверхностью сфер. В этом примере, сферы представляют собой сферы, покрытые стрептавидином, а транспозомы связаны с последовательностью сфер посредством комплекса «биотин-стрептавидин». Сферы имеют различные размеры. В одном из примеров, размер сфер может составлять 2,8 мкм. В другом примере, размер сфер может составлять 1 мкм. Суспензия (например, 1 мкл) сфер размером 1 мкм имеет большую площадь поверхности на объем связывающихся транспозом. Благодаря доступности площади поверхности для связывания транспозом, число продуктов тагментирования на реакцию увеличивается.

На фигуре 2 проиллюстрированы стадии 110, 115 и 120 метода, показанного на фигуре 1. В этом примере, транспозомы представлены в виде дуплексов. В другом примере (не показаны) может быть использована другая структура, такая как шпилька, то есть, один олигонуклеотид с аутокомплементарными областями, способными образовывать деплекс.

В стадии 110 метода 100 было получено множество биотинилированных транспозомов Р5 210a и множество транспозомов Р7 210b. Транспозомы Р5 210a и транспозомы Р7 210b были биотинилированными.

В стадии 115 метода 100, транспозомы Р5 210a и транспозомы Р7 210b смешивают с транспозазой Тn5 215 с получением множества собранных транспозом 220.

В стадии 120 метода 100, транспозомы 220 связаны со сферой 225. Сферой 225 является сфера, покрытая стрептавидином. Транспозомы 220 связаны со сферой 225 посредством комплекса «биотин-стрептавидин».

В одном из вариантов осуществления изобретения, смесь транспозом может быть получена на твердом носителе, таком как поверхность сферы как показано на фигурах 10, 11, 12 и 13. В этом примере, олигонуклеотиды Р5 и Р7 сначала связывают с поверхностью сферы, а затем осуществляют сборку транспозомных комплексов.

На фигуре 3 схематически проиллюстрирована диаграмма репрезентативного способа тагментации 300 на поверхности сферы. На этой фигуре проиллюстрирован способ 300 для сферы 225 со связанными с ней транспозомами 220, как показано на фигуре 2.

Раствор ДНК 310 добавляют к суспензии сфер 225. Поскольку ДНК 310 контактирует с транспозомами 220, то ДНК является тагментированной (фрагментированной и меченной) и связанной со сферами 225 посредством транспозом 220. Связанная и тагментированная ДНК 310 может быть подвергнута ПЦР-амплификации с получением пула ампликонов 315 в растворе (без сфер). Ампликоны 315 могут быть перенесены на поверхность проточной кюветы 320. Протокол получения кластера (например, протокол мостиковой амплификации или протокол любой другой амплификации, который может быть проведен для создания кластера) может быть проведен для получения множества кластеров 325 на поверхности проточной кюветы 320. Кластеры 325 представляют собой продукты клональной амплификации тагментированной ДНК 310. Кластеры 325 уже готовы для проведения следующей стадии протокола секвенирования.

В другом варианте осуществления изобретения, транспозомы могут быть связаны с любой твердой поверхностью, такой как стенки микроцентрифужной пробирки.

В другом варианте получения смеси транспозомных комплексов на поверхности сфер, олигонуклеотиды сначала связывают с поверхностью сферы, а затем осуществляют сборку транспозомы. На фигуре 10 проиллюстрирована блок-схема репрезентативного метода 1000, применяемого для создания транспозомных комплексов на поверхности

сферы. Метод 1000 включает, но не ограничивается ими, нижеследующие стадии.

В стадии 1010, олигонуклеотиды P5 и P7 связывают с поверхностью сферы. В одном из примеров, олигонуклеотиды P5 и P7 являются биотинилированными, а сферой является сфера, покрытая стрептавидином. Эта стадия также схематически проиллюстрирована на диаграмме 1100 фигуры 11. Как показано на фигуре 11, олигонуклеотид P5 1110 и олигонуклеотид P7 1115 связаны с поверхностью сферы 1120. В этом примере, один олигонуклеотид P5 1110 и один олигонуклеотид P7 1115 связаны с поверхностью сферы 1120, однако, с поверхностью множества сфер 1120 могут быть связаны олигонуклеотиды P5 1110 и/или олигонуклеотиды P7 1115 в любом количестве. В одном из примеров, олигонуклеотид P5 1110 содержит последовательность праймера P5, индексную последовательность (уникальный идентификатор), секвенирующую последовательность праймера рида 1 и мозаичную концевую последовательность (МЕ). В этом примере, олигонуклеотид P7 1115 содержит последовательность праймера P7, индексную последовательность (уникальный идентификатор), секвенирующую последовательность праймера рида 2 и последовательность МЕ. В другом примере (не

показано), индексная последовательность присутствует только в олигонуклеотиде P5 1110. В другом примере (не показано), индексная последовательность присутствует только в олигонуклеотиде P7 1115. В еще одном примере (не показано), индексная последовательность отсутствует в олигонуклеотиде P5 1110 и в олигонуклеотиде P7 1115.

В стадии 1015, комплементарные мозаичные концевые (МЕ') олигонуклеотиды гибридизуют с олигонуклеотидами P5 и P7, связанными со сферами. Эта стадия также схематически проиллюстрирована на диаграмме 1200 на фигуре 12. Как показано на фигуре 12, комплементарные МЕ-последовательности (МЕ') 1125 гибридизованы с олигонуклеотидом P5 1110 и с олигонуклеотидом P7 1115. Комплементарные МЕ-последовательности (МЕ') 1125 (например, комплементарные МЕ-последовательности (МЕ') 1125a и комплементарные МЕ-последовательности (МЕ') 1125b) гибридизованы с МЕ-последовательностями олигонуклеотида P5 1110 и олигонуклеотида P7 1115, соответственно. Комплементарные МЕ-последовательности (МЕ') 1125 обычно имеют длину приблизительно 15 оснований и фосфорилированы у 5'-конца.

В стадии 1020, фермент транспозазу добавляют к олигонуклеотидам, связанным со сферами, с образованием смеси связанных со сферами транспозомных комплексов. Эта стадия также схематически проиллюстрирована на диаграмме 1300 на фигуре 13. Как показано на фигуре 13, фермент транспозазу добавляют с образованием множества транспозомных комплексов 1310. В этом примере, транспозомным комплексом 1310 является дуплексная структура, содержащая фермент транспозазу и две связанных с поверхностью олигонуклеотидных последовательности, а также гибридизованные с ними комплементарные МЕ-последовательности (МЕ') 1125. Так, например, транспозомный комплекс 1310a содержит олигонуклеотид P5 1110, гибридизованный с комплементарной МЕ-последовательностью (МЕ') 1125 и олигонуклеотид P7 1115, гибридизованный с комплементарной МЕ-последовательностью (МЕ') 1125 (то есть, P5:P7); транспозомный комплекс 1310b содержит два олигонуклеотида P5 1110, гибридизованных с комплементарными МЕ-последовательностями (МЕ') 1125 (то есть, P5:P5), а транспозомный комплекс 1310c содержит два олигонуклеотида P7 1117,

гибридизованных с комплементарными МЕ-последовательностями (МЕ') 1125 (то есть, Р7:Р7). Отношения транспозомных комплексов Р5:Р5, Р7:Р7 и Р5:Р7 могут составлять, например, 25:25:50, соответственно.

На фигуре 14 схематически проиллюстрирована диаграмма репрезентативного способа тагментации 1400, проводимого на покрытых транспозомой сферах 1120 как показано на фигуре 13. В этом примере, тагментацию осуществляют путем добавления сферы 1120, содержащей транспозомные комплексы 1310, к раствору ДНК 1410 в буфере для тагментации, а затем ДНК связывают с поверхностью сферы 1120 посредством транспозом 1310. Последовательная тагментация ДНК 1410 приводит к образованию множества мостиковых молекул 1415, расположенных между транспозомами 1310. Длина мостиковых молекул 1415 может зависеть от плотности транспозомных комплексов 1310 на поверхности сферы 1120. В одном из примеров, плотность транспозомных комплексов 1310 на поверхности сферы 1120 может быть скорректирована путем изменения количества олигонуклеотидов Р5 и Р7, связанных с поверхностью сферы 1120 в стадии 1010 метода 100 как показано на фигуре 10. В другом примере, плотность транспозомных комплексов 1310 на поверхности сферы 1120 может быть скорректирована путем изменения количества комплементарной МЕ-последовательности (МЕ'), гибридной с олигонуклеотидами Р5 и Р7 в стадии 1015 метода 1000 как показано на фигуре 10. В другом примере, плотность транспозомных комплексов 1310 на поверхности сферы 1120 может быть скорректирована путем изменения количества фермента транспозазы, добавленного в стадии 1020 метода 1000 как показано на фигуре 1.

Длина мостиковых молекул 1415 не зависит от количества сфер 1120, содержащих связанные с ними транспозомные комплексы 1310 и используемых в реакции тагментации. Аналогичным образом, добавление большего или меньшего количества ДНК 1410 в реакционную смесь для тагментации не приводит к изменению размера конечного тагментированного продукта, но может влиять на выход продукта реакции.

В одном из примеров, сферой 1120 является парамагнитная сфера. В этом примере, очистка продукта реакции тагментации может быть легко достигнута путем иммобилизации сфер 1120 на магните и их последующей промывки. Поэтому, тагментация и последующая ПЦР-амплификация могут быть осуществлены в одном реакционном сосуде («в одном сосуде»).

В одном из своих аспектов, настоящее изобретение относится к способам и композициям, разработанным на основе транспозазы, которые позволяют получить информацию о сцеплении нуклеиновой кислоты-мишени на твердом носителе. В некоторых вариантах осуществления изобретения, указанные композиции и способы позволяют получить информацию о сборке/фазировании. В одном из вариантов осуществления изобретения, твердым носителем является сфера. В одном из вариантов осуществления изобретения, нуклеиновой кислотой-мишенью является ДНК. В одном из вариантов осуществления изобретения, нуклеиновой кислотой-мишенью является геномная ДНК. В некоторых вариантах осуществления изобретения, нуклеиновой кислотой-мишенью является РНК. В некоторых вариантах осуществления изобретения, РНК представляет собой мРНК. В некоторых вариантах осуществления изобретения, нуклеиновой кислотой-мишенью является комплементарная ДНК (кДНК).

В некоторых вариантах осуществления изобретения, транспозоны могут быть иммобилизованы в виде димеров на твердом носителе, таком как сферы, с последующим связыванием транспозазы с транспозонами и с образованием транспозом.

В некоторых вариантах осуществления изобретения, которые, в частности, относятся

к образованию транспозом на твердых фазах посредством транспозонов, иммобилизованных на твердой фазе, и добавлению транспозазы, два транспозона могут быть иммобилизованы на твердом носителе в непосредственной близости друг от друга (предпочтительно, на фиксированном расстоянии). Этот метод имеет несколько преимуществ. Во-первых, два транспозона всегда иммобилизуются одновременно, при этом, линкер имеет оптимальную длину, а два этих транспозона имеют оптимальную ориентацию, что способствует эффективному образованию транспозом. Во-вторых, эффективность образования транспозомы не зависит от плотности транспозонов. Два транспозона всегда находятся в правильной ориентации, а расстояние между ними является достаточным для образования транспозом. В-третьих, при рандомизированной иммобилизации транспозонов на поверхностях, транспозоны располагаются на различных расстояниях друг от друга, а поэтому, только одна часть имеет оптимальную ориентацию и оптимальное расстояние, достаточное для эффективного образования транспозом. Следовательно, не все транспозоны превращаются в транспозомы, при этом, могут присутствовать твердофазные иммобилизованные транспозоны, не образующие комплексов. Эти транспозоны могут быть мишенью для транспозиции, поскольку МЕ-часть представляет собой двухцепочечную ДНК. Это может приводить к снижению эффективности транспозиции или к образованию нежелательных побочных продуктов. Таким образом, транспозомы могут быть получены на твердом носителе, а поэтому они могут быть использованы для получения информации о сцеплении генов посредством тагментации и секвенирования. Репрезентативная схема проиллюстрирована на фигуре 15. В некоторых вариантах осуществления изобретения, транспозоны могут быть иммобилизованы на твердом носителе другими способами, не относящимися к способам химического связывания. Репрезентативными методами иммобилизации транспозонов на твердом носителе могут быть, но не ограничиваются ими, аффинное связывание, такое как связывание «стрептавидин-биотин», «мальтоза - белок, связывающийся с мальтозой», «антиген-антитело», или гибридизация ДНК-ДНК или ДНК-РНК.

В некоторых вариантах осуществления изобретения, транспозомы могут быть подвергнуты предварительной сборке, а затем иммобилизованы на твердом носителе. В некоторых вариантах осуществления изобретения, транспозоны содержат уникальные индексы, штрихкоды и сайты связывания праймеров для амплификации. Транспозаза может быть добавлена в раствор, содержащий транспозоны, с образованием транспозомных димеров, которые могут быть иммобилизованы на твердом носителе. В одном из вариантов осуществления изобретения может быть получено множество наборов сфер, где каждый из этих наборов имеет одинаковый индекс для иммобилизованных транспозонов, в результате чего могут быть получены индексные сферы. К каждой серии индексированных сфер может быть добавлена нуклеиновая кислота-мишень, как показано на фигуре 29А.

В некоторых вариантах осуществления изобретения, к каждой серии индексированных сфер может быть добавлена нуклеиновая кислота-мишень, а затем могут быть проведены отдельные реакции тагментации и ПЦР-амплификации.

В некоторых вариантах осуществления изобретения, нуклеиновая кислота-мишень, индексированные сферы и транспозомы могут быть объединены в виде капелек так, чтобы ряд капелек составляли одну сферу с одной или более молекулами ДНК и адекватными транспозомами.

В некоторых вариантах осуществления изобретения, индексированные сферы могут быть объединены, а затем, в этот пул может быть добавлена нуклеиновая кислота-

мишень, после чего может быть проведена реакция тагментации и ПЦР-амплификации в одном реакционном сосуде («в одном сосуде»).

В одном из своих аспектов, настоящее изобретение относится к способам и композициям, применяемым для получения информации о сцеплении генов посредством захвата сцепленных и перенесенных нуклеиновых кислот-мишеней на твердом носителе. В некоторых вариантах осуществления изобретения, транспозицию, сохраняющую сцепление генов (СРТ), осуществляют на ДНК, но эта ДНК должна оставаться интактной (СРТ-ДНК), что позволяет получить библиотеки сцепленных нуклеиновых кислот. Информация о сцеплении может быть сохранена с использованием транспозазы для сохранения ассоциации фрагментов матричных нуклеиновых кислот, которые являются смежными в нуклеиновой кислоте-мишени. СРТ-ДНК может быть захвачена посредством гибридизации комплементарных олигонуклеотидов, имеющих уникальные индексы или штрихкоды и иммобилизованных на твердом носителе, например, на сферах (фигура 29В). В некоторых вариантах осуществления изобретения, олигонуклеотид, иммобилизованный на твердом носителе, может также включать, помимо штрихкодов, сайты связывания с праймером и уникальные молекулярные индексы (UMI).

Преимуществом такого использования транспозом для сохранения физической близости фрагментированных нуклеиновых кислот является повышение вероятности того, что фрагментированные нуклеиновые кислоты, происходящие от одной и той же исходной молекулы, например, хромосомы, будут давать одну и ту же информацию по уникальным штрихкодам и индексам олигонуклеотидов, иммобилизованных на твердом носителе. Это даст возможность получить библиотеку сцепленных секвенирующих нуклеиновых кислот с уникальными штрихкодами. Эта библиотека сцепленных секвенирующих нуклеиновых кислот может быть секвенирована для получения информации о смежных последовательностях.

На фигурах 16 и 17 схематически представлен репрезентативный вариант вышеуказанного аспекта изобретения, относящегося к получению библиотек сцепленных секвенирующих нуклеиновых кислот, имеющих уникальные штрихкоды или индексы. Этот репрезентативный метод основан на лигировании СРТ-ДНК с иммобилизованными олигонуклеотидами на твердом носителе, содержащими уникальные индексы и штрихкоды, и на проведении ПЦР с заменой цепи для получения секвенирующей библиотеки. В одном из вариантов осуществления изобретения, клональные индексированные сферы могут быть получены с использованием иммобилизованных последовательностей ДНК, таких как неспецифический или специфический праймер и индекс. Библиотеки сцепленных нуклеиновых кислот могут быть иммобилизованы на клональных индексированных сферах посредством гибридизации с иммобилизованными олигонуклеотидами с последующим их лигированием. Поскольку внутримолекулярная реакция гибридизации посредством захвата происходит гораздо быстрее, чем межмолекулярная гибридизация, то библиотеки, перенесенные с сохранением сцепления, будут «обертывать» сферы. На фигурах 18 и 19 проиллюстрированы захват СРТ-ДНК на клональных индексированных сферах и сохранение информации о сцеплении. ПЦР с заменой цепи позволяет осуществлять перенос информации о клональном индексе сферы на индивидуальную молекулу. Таким образом, каждая библиотека сцепленных нуклеиновых кислот может иметь уникальные индексы.

В некоторых вариантах осуществления изобретения, олигонуклеотид, иммобилизованный на твердом носителе, может содержать частично двухцепочечную структуру, где одна цепь иммобилизована на твердом носителе, а другая цепь частично комплементарна иммобилизованной цепи и образует Y-адаптер. В некоторых вариантах

осуществления изобретения, Y-адаптер, иммобилизованный на твердой поверхности, связан с тагментированной сцепленной ДНК посредством лигирования и заполнения брешей, как показано на фигуре 20.

В некоторых вариантах осуществления изобретения, Y-адаптер получают посредством захвата СРТ-ДНК с зондом/индексом на твердом носителе, таком как сферы. На фигуре 21 представлена репрезентативная схема получения таких Y-адаптеров. Использование этих Y-адаптеров дает уверенность, что каждый фрагмент может стать секвенирующей библиотекой. Это позволяет охватить большую область секвенирования.

В некоторых вариантах осуществления изобретения, свободные транспозомы могут быть отделены от СРТ-ДНК. В некоторых вариантах осуществления изобретения, такое отделение свободных транспозом осуществляют посредством эксклюзионной хроматографии. В одном из вариантов осуществления изобретения, такое отделение может быть достигнуто на колонках MicroSpin S-400 HR (GE Healthcare Life Sciences, Pittsburgh, PA). На фигуре 22 проиллюстрирован электрофорез в агарозном геле для СРТ-ДНК, отделенных от свободных транспозом.

Захват сцепленных и перенесенных нуклеиновых кислот-мишеней на твердом носителе посредством гибридизации имеет ряд уникальных преимуществ. Во-первых, этот метод основан на гибридизации, но не на транспозиции. Степень внутримолекулярной гибридизации выше степени межмолекулярной гибридизации. Таким образом, шансы получения библиотек со сцепленными и перенесенными нуклеиновыми кислотами на одной молекуле ДНК-мишени для обертывания уникальной индексированной сферы гораздо выше, чем шансы получения библиотек на двух или более других отдельных молекулах ДНК-мишени для обертывания уникальной индексированной сферы. Во-вторых, транспозиция ДНК и штриховое кодирование перенесенной ДНК могут быть осуществлены в две отдельных стадии. В-третьих, можно решить проблемы, связанные со сборкой активных транспозом на сферах и с оптимизацией плотности транспозомов на твердых поверхностях. В-четвертых, продукты ауто транспозиции могут быть удалены путем колоночной очистки. В-пятых, поскольку перенесенная и сцепленная ДНК содержит бреши, то такая ДНК является более гибкой, что будет снижать нагрузку на плотность транспозиции (размер вставок) по сравнению с плотностью, достигаемой с применением методов иммобилизации транспозомы на сферах. В-шестых, в этом методе могут быть использованы схемы комбинаторного штрихкодирования. В-седьмых, может быть облегчена процедура ковалентного связывания индексированных олигонуклеотидов со сферами. Таким образом, уменьшается вероятность замены индекса. В-восьмых, реакция тагментации и последующая реакция ПЦР-амплификации могут быть мультиплексными и могут быть проведены в одном реакционном сосуде («в одном сосуде»), что освобождает от необходимости проведения отдельных реакций для каждой индексной последовательности.

В некоторых вариантах осуществления изобретения, множество уникальных штрихкодов для всех нуклеиновых кислот-мишеней может быть встроено во время транспозиции. В некоторых вариантах осуществления изобретения, каждый штрихкод включает первую последовательность со штрихкодом и вторую последовательность со штрихкодом, имеющие сайт фрагментации, расположенные между ними. Первая последовательность со штрихкодом и вторая последовательность со штрихкодом могут быть идентифицированы или сконструированы так, чтобы они спаривались друг с другом. Спаривание может носить информативный характер, а поэтому первый штрихкод должен быть ассоциирован со вторым штрихкодом. Спаренные последовательности со штрихкодом могут быть преимущественно использованы для

сбора данных по секвенированию библиотеки матричных нуклеиновых кислот. Так, например, идентификация первой матричной нуклеиновой кислоты, содержащей первую последовательность со штрихкодом, и второй матричной нуклеиновой кислоты, содержащей вторую последовательность со штрихкодом, где указанная

5 последовательность спарена с первой последовательностью, показала, что первая и вторая матричные нуклеиновые кислоты представляют собой последовательности, смежные друг с другом в нуклеиновой кислоте-мишени. Такие методы могут быть применены для сборки репрезентативных последовательностей нуклеиновой кислоты-мишени *de novo*, где такая сборки не требует эталонного генома.

10 В одном из своих аспектов, настоящее изобретение относится к способам и композициям, применяемым для получения библиотеки последовательностей специфического фрагмента ДНК с использованием «дробовика».

В одном из вариантов осуществления изобретения, клональные индексированные сферы получают с использованием иммобилизованных олигонуклеотидных

15 последовательностей: неспецифических или специфических праймеров и уникальных индексов. Нуклеиновую кислоту-мишень добавляют к клональным индексированным сферам. В некоторых вариантах осуществления изобретения, нуклеиновой кислотой-мишенью является ДНК. В одном из вариантов осуществления изобретения, ДНК-мишень является денатурированной. ДНК-мишень гибридизуется сначала с праймерами,

20 содержащими уникальные индексы, иммобилизованные на твердой поверхности (например, на сферах), а затем с другими праймерами, имеющими тот же самый индекс. Праймеры на сферах амплифицируют ДНК. При этом, могут быть проведены один или более дополнительных раундов амплификации. В одном из вариантов осуществления изобретения, амплификация может быть осуществлена для целого генома с

25 использованием иммобилизованных на сфере праймеров, имеющих 3'-рандомизированную n-мерную последовательность. В предпочтительном варианте осуществления изобретения, рандомизированный n-мер содержит псевдокомплементарные основания (2-тиотимин, 2-амино-dA, N4-этилцитозин и т.п.), что предотвращает взаимодействия праймера с праймером в процессе амплификации

30 (Hoshika, S; Chen, F; Leal, NA; Benner, SA, Angew. Chem. Int. Ed. 49(32) 5554-5557 (2010). На фигуре 23 представлена репрезентативная схема получения библиотеки последовательностей специфического фрагмента ДНК с использованием «дробовика». Может быть генерирована клональная индексированная секвенирующая библиотека, которая может представлять собой библиотеку амплифицированных продуктов. В

35 одном из вариантов осуществления изобретения, такая библиотека может быть получена путем транспозиции. Информация о последовательности клональной индексированной библиотеки может быть использована для сбора информации о смежных последовательностях по данным об индексах, служащих в качестве ориентира. На фигуре 24 представлена репрезентативная схема сборки информации о

40 последовательности клональной индексированной секвенирующей библиотеки.

Способы согласно вышеуказанным вариантам изобретения имеют несколько преимуществ. Внутримолекулярная амплификация на сферах происходит гораздо быстрее, чем межмолекулярная амплификация. Таким образом, продукты на сферах имеют один и тот же индекс. Библиотека специфического фрагмента ДНК может быть

45 получена методом «дробовика». Рандомизированные праймеры амплифицируют матрицу в произвольных положениях, а поэтому библиотека, полученная методом «дробовика» и имеющая тот же самый индекс, может быть генерирована из специфической молекулы, а информация о последовательности может быть собрана с

использованием индексированной последовательности. Значительное преимущество способов согласно вышеуказанным вариантам изобретения заключается в том, что реакции могут быть мультиплексными и проводиться за одну реакцию (в одном реакционном сосуде), и для их проведения не требуется множества отдельных лунок.

- 5 Многие клональные индексированные сферы могут быть получены так, чтобы множество различных фрагментов могло иметь уникальные метки, что позволило бы дифференцировать родительские аллели для одних и тех же геномных областей. При высоком числе индексов, вероятность того, что копия ДНК отца и копия ДНК матери будут иметь один и тот же индекс для одной и той же геномной области, довольно мала.
- 10 Этот способ имеет то преимущество, что внутренние реакции проходят гораздо быстрее, чем внешние, а поэтому сферы будут, в основном, давать фактическое распределение в более широком физическом пространстве.

- В некоторых вариантах всех вышеуказанных аспектов изобретения, этот способ может быть применен для внеклеточных ДНК (свободной ДНК или свДНК) в анализах
- 15 на свДНК. В некоторых вариантах осуществления изобретения, свДНК получают из плазмы и плацентарной жидкости.

- В одном из вариантов осуществления изобретения, плазма может быть выделена из неразведенной цельной крови с использованием мембранного сепаратора плазмы, работающего на основе седиментации (Liu et al. Anal Chem. 2013 Nov. 5;85(21):10463-70).
- 20 В одном из вариантов осуществления изобретения, зона сбора плазмы сепаратора может содержать твердый носитель, включающий транспозомы. Твердый носитель, включающий транспозомы, может захватывать свДНК из выделенной плазмы по мере ее отделения от цельной крови и может концентрировать свДНК и/или тагментировать ДНК. В некоторых вариантах осуществления изобретения, тагментация также позволяет
- 25 вводить уникальные штрихкоды для последующего разуплотнения после секвенирования пула библиотек.

- В некоторых вариантах осуществления изобретения, зона сбора сепаратора может содержать маточную ПЦР-смесь (праймеры, нуклеотиды, буферы, металлы) и полимеразу. В одном из вариантов осуществления изобретения, маточная смесь может
- 30 использоваться в сухой форме так, чтобы ее можно было развести по мере выхода плазмы из сепаратора. В некоторых вариантах осуществления изобретения, праймерами являются рандомизированные праймеры. В некоторых вариантах осуществления изобретения, праймеры могут быть специфичными к конкретному гену. ПЦР-амплификация свДНК может приводить к получению библиотеки непосредственно из
- 35 выделенной плазмы.

- В некоторых вариантах осуществления изобретения, зона сбора сепаратора может содержать маточную ОТ-ПЦР-смесь (праймеры, нуклеотиды, буферы, металлы), обратную транскриптазу и полимеразу. В некоторых вариантах осуществления изобретения, праймерами являются рандомизированные праймеры или олиго-dT-
- 40 праймеры. В некоторых вариантах осуществления изобретения, праймеры могут быть специфичными к конкретному гену. Полученная кДНК может быть использована для секвенирования. Альтернативно, кДНК может быть обработана транспозомами, иммобилизованными на твердом носителе, для получения библиотеки последовательностей.

- 45 В некоторых вариантах осуществления изобретения, сепаратор плазмы может включать штрихкоды (штрихкоды 1D или 2D). В некоторых вариантах осуществления изобретения, аппарат для сепарации может включать устройство для забора крови. Это позволяет осуществлять прямую доставку крови в сепаратор плазмы и в устройство

для получения библиотеки. В некоторых вариантах осуществления изобретения, устройство может включать нижерасположенный анализатор последовательности. В некоторых вариантах осуществления изобретения, анализатором последовательности является одноразовый секвенатор. В некоторых вариантах осуществления изобретения, секвенатор позволяет установить очередность взятия образцов перед их серийным секвенированием. Альтернативно, секвенатор может обеспечивать рандомизированный доступ образцов в зону их секвенирования.

В некоторых вариантах осуществления изобретения, зона сбора плазмы может содержать субстраты на основе двуокиси кремния для концентрирования неклеточной ДНК.

Одновременное фазирование и детектирование метилирования

5-метилцитозин (5-Me-C) и 5-гидроксиметилцитозин (5-гидрокси-C) также известны как эпигенетические модификации, играющие важную роль в метаболизме и дифференцировке клеток, а также в развитии рака. Авторами настоящего изобретения было неожиданно обнаружено, что фазирование и одновременное детектирование метилирования может быть осуществлено с применением способов и композиций согласно изобретению. Способы согласно изобретению позволяют одновременно осуществлять CPT-seq на сферах (с индексированными сцепленными библиотеками) и детектирование метилирования ДНК. Так, например, отдельные библиотеки, полученные на сферах, могут быть обработаны бисульфитом с последующим превращением неметилированных, но не метилированных, цитозинов (C) в U, что позволяет детектировать 5-Me-C. С помощью дополнительного анализа на фазирование с использованием гетерозиготных SNP могут быть получены эпигенетические модификации-фазировочные блоки, имеющие размер в пределах мультимегаоснований.

В некоторых вариантах осуществления изобретения, размер анализируемой ДНК может составлять приблизительно от 100 оснований до мультимегаоснований. В некоторых вариантах осуществления изобретения, размер анализируемой ДНК может составлять приблизительно 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1300, 1500, 2000, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, 10000, 10500, 11000, 11500, 12000, 12500, 13000, 14000, 14500, 15000, 15500, 16000, 16500, 17000, 17500, 18000, 18500, 19000, 19500, 20000, 20500, 21000, 21500, 22000, 22500, 23000, 23500, 24000, 24500, 25000, 25500, 26000, 26500, 27000, 27500, 28000, 28500, 29500, 30000, 30500, 31000, 31500, 32000, 33000, 34000, 35000, 36000, 37000, 38000, 39000, 40000, 42000, 45000, 50000, 55000, 60000, 65000, 70000, 75000, 80000, 85000, 90000, 95000, 100000, 110000, 120000, 130000, 140000, 150000, 160000, 170000, 180000, 200000, 225000, 250000, 300000, 350000, 400000, 450000, 500000, 550000, 600000, 650000, 700000, 750000, 800000, 850000, 900000, 1000000, 1250000, 1500000, 2000000, 2500000, 3000000, 4000000, 5000000, 6000000, 7000000, 8000000, 9000000, 10000000, 15000000, 20000000, 30000000, 40000000, 50000000, 75000000, 100000000 или более оснований.

Другие эпигенетические модификации, такие как 5-гидрокси-C, продукты окисления ДНК, продукты алкилирования ДНК, футпринтинг на гистоне и т.д., также могут быть проанализированы на фазирование с применением описанных способов и композиций согласно изобретению.

В некоторых вариантах осуществления изобретения, ДНК сначала превращают в сцепленные с индексом библиотеки на твердом носителе. Отдельные индексированные библиотеки, размер которых гораздо меньше, чем размер исходной ДНК, менее предрасположены к фрагментации, поскольку такие отдельные библиотеки имеют меньший размер. Даже при потере небольшой фракции индексированных библиотек,

информация о фазировании еще сохраняется для всего диапазона индексируемой молекулы ДНК. Так, например, если 100 т.п.о.-молекула, полученная путем традиционного превращения под действием бисульфита (BSC), была фрагментирована наполовину, то сцепление ограничено 50 т.п.о. В описанных здесь способах, сначала индексируют 100 т.п.о.-библиотеку и даже в случае потери части отдельных библиотек, сцепление еще будет составлять ~100 т.п.о. (за исключением очень маловероятного события, когда все потерянные библиотеки находятся у одного конца молекулы ДНК). Кроме того, способы, описанные в настоящей заявке, имеют дополнительные преимущества, заключающиеся в том, что при проведении этих способов не требуются дополнительные стадии очистки по сравнению с традиционными методами превращения под действием бисульфита, что также способствует повышению выхода. В способах согласно изобретению, сферы, после их превращения под действием бисульфита, просто промывают. Кроме того, поскольку ДНК связана с твердой фазой, то может быть легко осуществлен буферный обмен с минимальной потерей ДНК (индексируемых библиотек) и уменьшением времени на обработку.

Репрезентативная схема одновременного фазирования и детектирования метилирования представлены на фиг. 43. Технологическая схема такого процесса состоит из фрагментации ДНК на сферах; лигирования областей-повторов размером 9 п.о. с заполнением брешей; удаления Tn5 посредством ДСН; и превращения отдельных библиотек на сферах под действием бисульфита. Превращение под действием бисульфита осуществляют в условиях денатурации для гарантии того, что соседние комплементарные библиотеки не будут подвергаться повторному отжигу, который снижает эффективность превращения под действием бисульфита. BCS превращает неметилированные С в U, а метилированные С не подвергаются такому превращению.

На фиг. 44 представлена альтернативная репрезентативная схема одновременного фазирования и детектирования метилирования. После получения секвенирующих библиотек посредством транспозиции, часть библиотек, лигированных с заполнением брешей, разлагается с образованием одноцепочечных матриц. Для превращения под действием бисульфита, одноцепочечные матрицы требуют более мягких условий, поскольку эти матрицы уже являются одноцепочечными, что снижает потери библиотек или повышает эффективность превращения под действием бисульфита. В одном из вариантов осуществления изобретения, смесь 3'-тио-защищенных транспозонов (Ехо-резистентных) и незащищенных транспозонов используют на одной и той же сфере. Ферменты, например, Ехо I, могут быть применены для расщепления библиотек, не защищенных тио-группой, что способствует превращению этих библиотек в одноцепочечные библиотеки. С использованием смеси тио-защищенных транспозонов/незащищенных транспозонов (50:50), 50% библиотек превращаются в одноцепочечные библиотеки (50% библиотек имеют один защищенный и один незащищенный транспозон, то есть, комплементарную цепь), 25% вообще не подвергались превращению (то есть, оба транспозона были тио-защищенными), а 25% подвергались превращению с удалением всей библиотеки (оба транспозона были не защищены).

Одним из недостатков метода бисульфитного превращения ДНК, связанной с твердой фазой, такой как стрептавидиновые магнитные сферы, является длительная обработка связанной со сферами ДНК бисульфитом натрия при высоких температурах, которая приводит к разрушению как ДНК, так и сфер. Для снижения степени повреждения ДНК, в реакционную смесь, до ее обработки бисульфитом, добавляют носитель ДНК (то есть, лямбда ДНК). Было установлено, что даже в присутствии носителя ДНК теряется приблизительно 80% исходной ДНК. В результате этого, блоки сцепления

CPTSeq имеют меньшее число членов, чем это наблюдается в протоколе проведения традиционного CPTSeq.

Поэтому, для увеличения выхода ДНК, в предложенный здесь протокол проведения Epi-CPTSeq добавляют несколько стратегий. Первая стратегия основана на снижении размера вставок в библиотеку путем введения более плотной совокупности транспозомных комплексов, связанных со стрептавидиновыми сферами. При снижении размера библиотеки, меньшее число элементов библиотеки будет подвергаться разрушению после обработки бисульфитом.

Второй стратегией увеличения выхода ДНК в соответствии с протоколом Epi-CPTSeq является ферментативное восстановление разрушенных элементов библиотеки. Целью такой стратегии восстановления является повторное присоединение общей 3'-последовательности, необходимой для амплификации библиотеки, к элементам библиотеки, связанным со сферами, поскольку эти элементы подвергались расщеплению и теряли свою 3'-часть в процессе обработки бисульфитом. После присоединения общей 3'-последовательности, эти элементы снова могут быть подвергнуты ПЦР-амплификации и секвенированию. На фигурах 67 и 68 представлена репрезентативная схема этой стратегии. Двухцепочечные элементы библиотеки CPTSeq были денатурированы и подвергнуты бисульфитному превращению (верхняя панель). В процессе бисульфитного превращения, одна из цепей ДНК была повреждена (средняя панель), что приводило к потере общей ПЦР-последовательности на 3'-конце. Стратегии «спасения» матрицы способствуют восстановлению общей 3'-последовательности (показанной зеленым), необходимой для ПЦР-амплификации (нижняя панель). В одном из примеров используются концевая трансфераза в присутствии 3'-фосфорилированного олиго-аттенюатора; последовательность, содержащая секвенирующий адаптер, а затем, олиго-dT-фрагмент (фигура 68A). Вкратце, TdT присоединяет фрагмент из 10-15 dA к 3'-концу разрушенного элемента библиотеки, что приводит к гибридизации с олиго-dT-частью олиго-аттенюатора. Образование этой гибридной ДНК приводит к прекращению реакции TdT и к образованию матрицы для последующего удлинения 3'-конца разрушенного элемента библиотеки под действием ДНК-полимеразы.

В альтернативной технологической схеме (фигура 68B), реакцию присоединения TdT-хвоста осуществляют в присутствии частично двухцепочечного олиго-аттенюатора, содержащего одноцепочечную олиго-dT-часть и 5'-фосфорилированную двухцепочечную секвенирующую часть адаптера. После прекращения реакции TdT, разрыв между последним добавлением dA и 5'-фосфорилированным олиго-аттенюатором репарируют с помощью ДНК-лигазы.

Обе описанные технологические схемы основаны на недавно разработанной регулируемой реакции присоединения TdT-хвоста, описанной в публикации заявки на патент США 20150087027. Общий секвенирующий адаптер может быть также присоединен к 3'-концу разрушенных элементов библиотеки путем только что встроенной оцДНК-матрицы, переключающей активность MMLV-RT. Короче говоря, MMLV RT и матрицу, переключающие олигонуклеотид (TS_oligo), добавляют к разрушенной ДНК (фигура 68C). В первой стадии этой реакции, обратная транскриптаза добавляет несколько дополнительных нуклеотидов к 3'-концам одноцепочечного фрагмента ДНК, и эти пары оснований с олиго (N)-последовательностью присутствуют у 3'-конца одного из TS_oligo. Затем, матрица на основе обратной транскриптазы, переключающая активность, добавляет последовательности общих гибридизованных праймеров к 3'-концу BSC-разрушенного элемента библиотеки, что приводит к восстановлению способности к ПЦР-амплификации в присутствии общих

секвенирующих праймеров.

При проведении третьей стратегии, для «спасения» элементов библиотеки, потерявших свои общие последовательности у 3'-конца в процессе бисульфитного превращения, может быть применен набор Epicentre's EpiGenome для осуществления метода
 5 конструирования библиотеки «после бисульфитного превращения». Как показано на фигуре 69, этот метод «спасения» библиотеки проводят с применением 3'-фосфорилированных олигонуклеотидов с общими последовательностями, за которыми расположен короткий фрагмент рандомизированной последовательности. Эти короткие рандомизированные последовательности гибридизуются с одноцепочечной ДНК,
 10 обработанной бисульфитом, а общие последовательности затем копируют на разрушенную цепь библиотеки с помощью ДНК-полимеразы.

На фигуре 74 проиллюстрирована четвертая стратегия усовершенствования методов секвенирования на сферах путем обработки бисульфитом. Первая общая последовательность, содержащая метку для захвата, ковалентно присоединена к 5'-
 15 концам ДНК. Первая общая последовательность может быть присоединена к ДНК различными методами, включая одностороннюю транспозицию (как показано на фигуре), лигирование адаптера или лигирование адаптера под действием концевой трансферазы (TdT) как описано в публикации заявки на патент США 20150087027.

Затем, ДНК денатурируют (например, путем инкубирования при высокой
 20 температуре) и присоединяют к твердому носителю. Если в качестве метки для захвата на CS1 служит биотин, то, например, ДНК может быть связана с применением стрептавидиновых магнитных сфер (как показано на фигуре). После присоединения к твердому носителю может быть легко осуществлен буферный обмен.

В следующей стадии осуществляют бисульфитное превращение оцДНК. ДНК, если
 25 она присутствует в одноцепочечной форме, должна быть легко доступна для бисульфитного превращения, и эффективность ее превращения до 95% может быть достигнута с помощью модифицированного варианта набора Promega's Methyl Edge BSC (фигура 75).

После бисульфитного превращения, вторую общую последовательность ковалентно
 30 присоединяют к 3'-концу оцДНК, связанной с твердым носителем. Для ковалентного присоединения олигонуклеотидов к оцДНК были применены несколько методов, описанных выше. С применением метода лигирования аттенюатора/адаптера TdT, эффективность лигирования может составлять >95%. В результате, конечные выходы библиотеки, полученные с применением предложенной технологической схемы
 35 MethylSeq, должны превышать выходы, полученные с применением уже существующих методов.

В конечной стадии осуществляют ПЦР для амплификации библиотеки и ее удаления из твердого носителя. ПЦР-праймеры могут быть сконструированы так, чтобы они присоединяли дополнительные общие последовательности, такие как секвенирующие
 40 адаптеры, к концам библиотеки MethylSeq.

Получение библиотек различных размеров в одном анализе

Точность сборки геномов зависит от применения технологий масштабирования различных длин. Так, например, все методы «дробовика» (100 п.о.), то есть, методы спаривания (~3 т.п.о.) с -Hi-C (в масштабе мегаоснований) представляют собой методы,
 45 которые применяют в целях последовательного улучшения сборки и увеличения длин контигов. Недостаток этого метода заключается в том, что для его осуществления необходимо проводить множество анализов, а поэтому такой многостадийный метод является очень трудоемким и дорогостоящим. Описанные здесь композиции и способы

могут быть применены к различным шкалам длин в одном анализе.

В некоторых вариантах осуществления изобретения, получение библиотеки может быть осуществлено в одном анализе с использованием твердого носителя различных размеров, например, сфер. Размер каждой сферы позволяет определить размер или интервал размеров специфической библиотеки и физический размер сферы, определяющей размер библиотеки. Все сферы различных размеров имеют уникальные клональные индексы, которые переносят в библиотеку. Так, например, библиотеки различных размеров получают с использованием каждой библиотеки с различными длинами, имеющими уникальный индекс. Библиотеки с различными длинами получают одновременно в одном и том же физическом пространстве, что позволяет снизить стоимость технологического процесса и усовершенствовать весь этот процесс. В некоторых вариантах осуществления изобретения, специфический твердый носитель каждого конкретного размера, например, сфера конкретного размера имеет уникальный индекс. В некоторых других вариантах осуществления изобретения, множество различных индексов твердых носителей одного и того же размера, например, сфер, также получают так, чтобы множество молекул ДНК могли иметь индексы, распределенные по интервалу данных значений. На фиг. 45 представлена репрезентативная схема создания библиотеки различных размеров с применением клональных индексированных сфер различных размеров в одном анализе.

В некоторых вариантах осуществления изобретения, размер генерируемых библиотек составляет приблизительно 50, 75, 100, 150, 200, 250, 300, 350, 400, 500, 600, 700, 800, 900, 1000, 1200, 1300, 1500, 2000, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, 10000, 10500, 11000, 11500, 12000, 12500, 13000, 14000, 14500, 15000, 15500, 16000, 16500, 17000, 17500, 18000, 18500, 19000, 19500, 20000, 20500, 21000, 21500, 22000, 22500, 23000, 23500, 24000, 24500, 25000, 25500, 26000, 26500, 27000, 27500, 28000, 28500, 29500, 30000, 30500, 31000, 31500, 32000, 33000, 34000, 35000, 36000, 37000, 38000, 39000, 40000, 42000, 45000, 50000, 55000, 60000, 65000, 70000, 75000, 80000, 85000, 90000, 95000, 100000, 110000, 120000, 130000, 140000, 150000, 160000, 170000, 180000, 200000, 225000, 250000, 300000, 350000, 400000, 450000, 500000, 550000, 600000, 650000, 700000, 750000, 800000, 850000, 900000, 1000000, 1250000, 1500000, 2000000, 2500000, 3000000, 4000000, 5000000, 6000000, 7000000, 8000000, 9000000, 10000000, 15000000, 20000000, 30000000, 40000000, 50000000, 75000000, 100000000 или более оснований.

В некоторых вариантах осуществления изобретения, обсуждаемые выше библиотеки, имеющие длины различных масштабов, могут быть использованы для сборки псевдогенов, паралогов и т.п. вместо библиотеки, имеющей длину одного большого масштаба. В некоторых вариантах осуществления изобретения, библиотеки, имеющие длины различных масштабов, получают одновременно в одном анализе. Преимущество этого метода состоит в том, что по меньшей мере одна шкала длин имеет уникальную область, которая охватывает только псевдоген и/или ген, но не то и другое. Так, например, вариантам, детектируемым по этой шкале длин, может быть присвоен уникальный вариант, по которому можно определить, является ли он геном или псевдогеном. То же самое правило справедливо и для вариантов различных копий, паралогов и т.п. Точность сборки зависит от применения различных шкал длин. С применением описанных здесь методов могут быть получены сцепленные с индексами библиотеки различных длин в одном анализе вместо отдельных различных библиотек с различными длинами. На фиг. 46 представлена репрезентативная схема определения генетических вариантов с применением библиотек с различными шкалами длин.

Анализ геномных вариантов

Описанные здесь композиции и способы относятся к анализам геномных вариантов. Репрезентативными геномными вариантами являются, но не ограничиваются ими, делеции, межхромосомные транслокации, дубликации, паралоги, межхромосомные сцепления генов. В некоторых вариантах осуществления изобретения, описанные здесь композиции и способы применяются для получения информации о фазировании геномных вариантов. В представленной ниже таблице проиллюстрированы репрезентативные межхромосомные сцепления генов.

Таблица 1: Межхромосомные сцепления генов

Межхромосомный									СРТ-детектируемый
BP1 Хр	BP1-старт	BP1-конец	длина	BP2	BP2-старт	BP2-конец	длина	Допол.	
chr2	3930448	3935237	4789	chr12	124494444	124,500,974	6,530	100	Да
chr2	11948918	11954246	5328	chr21	18143482	18,149,316	5,834	104	Да
chr3	73158440	73162742	4302	chr17	41379324	41384937	5,613	80	Да
chr3	73158440	73162742	4302	chr17	41397704	41,403,894	6,190	132	Да
chr3	75985574	75999286	13712	chr20	26198677	26,213,907	15,230	271	Да
chr3	97545352	97548214	2862	chr20	51820873	51826137	5,264	54	Нет
chr3	111271573	111275046	3472	chr8	128529232	128,538,779	9,547	138	Да
chr5	45603558	45620710	17152	chr22	39709007	39715736	6,729	64	Нет
chr7	81789161	81794473	5312	chr10	60900343	60904922	4,579	66	Да
chr7	100936524	100940377	3853	chr11	114417947	114425186	7,239	45	Да
chr7	111050826	111056166	5340	chr12	108199590	108203183	3,593	55	Да
chr8	28431852	28436443	4591	chr17	79758758	79762966	4,208	51	Нет
chr9	12624450	12627475	3025	chr22	29065133	29067995	2,862	11	Да
chr9	96381359	96386424	5065	chr13	19645775	19650874	5,099	29	Да
chr9	101473886	101479887	6001	chr22	32575696	32579748	4,052	67	Да
chr9	109772276	109777707	5431	chr20	29873930	29878940	5,010	70	Да
chr9	121298728	121305690	6962	chr11	11980608	12002075	21,467	178	Да
chr12	56987510	56991801	4291	chr15	39995972	39999958	3,986	25	Да
chr13	55877158	55882393	5235	chr19	12896441	12904258	7,817	72	Да
chr13	63618669	63624153	5484	chr17	21666588	21670266	3,678	115	Да

В таблице 2 представлены репрезентативные делеции, присутствующие в хромосоме

Таблица 2: Репрезентативные делеции в хромосоме 1

		Делеции							
Хр	BP1-старт	BP1-конец	BP2-старт	BP2-конец	Размер (п.о.)		CPT-сигнал		
5	chr1	1219982	1223434	3452	1225577	1229105	3528	2143	Да
	chr1	9591985	9595341	3356	9597227	9600089	2862	1896	Да
	chr1	14432636	14436290	3654	14437570	14441375	3805	1280	Да
	chr1	25154228	25158676	4448	25161514	25165093	3579	2838	Да
	chr1	26456898	26459838	2940	26463789	26468542	4753	3951	Да
10	chr1	34985513	34991181	5668	34992539	34996862	4323	1358	Да
	chr1	56827433	56831099	3666	56834909	56838758	3849	3810	Нет
	chr1	63702441	63705334	2893	63708226	63712091	3865	2892	??
	chr1	71234176	71237339	3163	71239506	71243649	4143	2167	Да
	chr1	73450444	73454045	3601	73454999	73458655	3656	954	??
15	chr1	81400468	81404472	4004	81408850	81413863	5013	4378	Да
	chr1	83465559	83469370	3811	83477194	83482174	4980	7824	Да
	chr1	84513313	84517907	4594	84524680	84528485	3805	6773	Да
	chr1	84708940	84711988	3046	84715625	84719559	3934	3639	Да
	chr1	97657769	97660911	3142	97669287	97672164	2877	8376	Нет
20	chr1	1.06E+08	1.06E+08	4804	1.06E+08	1.06E+08	3723	6370	Да
	chr1	1.07E+08	1.07E+08	5095	1.07E+08	1.07E+08	6141	887	??
	chr1	1.09E+08	1.09E+08	2999	1.09E+08	1.09E+08	4018	3899	Да
	chr1	1.1E+08	1.1E+08	5069	1.1E+08	1.1E+08	4470	2801	??
	chr1	1.13E+08	1.13E+08	5237	1.13E+08	1.13E+08	4826	3241	Да
25	chr1	1.16E+08	1.16E+08	4594	1.16E+08	1.16E+08	4531	3084	??
	chr1	1.45E+08	1.45E+08	5596	1.45E+08	1.45E+08	4721	3724	Да
	chr1	1.59E+08	1.59E+08	4541	1.59E+08	1.59E+08	3981	1856	Да (двойная делеция)
	chr1	1.79E+08	1.79E+08	4565	1.79E+08	1.79E+08	5191	2238	Да
	chr1	1.79E+08	1.79E+08	4480	1.79E+08	1.79E+08	3917	2479	Да
30	chr1	1.85E+08	1.85E+08	4475	1.85E+08	1.85E+08	3316	6243	Да
	chr1	1.97E+08	1.98E+08	4323	1.98E+08	1.98E+08	3694	2522	Нет
	chr1	2.08E+08	2.08E+08	4995	2.08E+08	2.08E+08	5216	3458	Да
	chr1	2.08E+08	2.08E+08	3589	2.08E+08	2.08E+08	4964	2954	Да
	chr1	2.1E+08	2.1E+08	2973	2.1E+08	2.1E+08	2943	7803	Да (двойная делеция)
	chr1	2.3E+08	2.3E+08	3828	2.3E+08	2.3E+08	3079	8440	Да
	chr1	2.32E+08	2.32E+08	3531	2.32E+08	2.32E+08	3758	1699	??
	chr1	2.38E+08	2.38E+08	3848	2.38E+08	2.38E+08	3920	2112	Да
	chr1	2.48E+08	2.48E+08	3407	2.48E+08	2.48E+08	3873	6003	Да

В некоторых вариантах осуществления изобретения, нуклеиновая кислота-мишень может быть фрагментирована до ее обработки транспозомами. Репрезентативными методами фрагментации являются, но не ограничиваются ими, обработка ультразвуком, механический сдвиг и расщепление рестриктирующими ферментами. Фрагментирование нуклеиновой кислоты-мишени до тагментации (фрагментирования и мечения) применяется преимущественно для сборки/фазирования псевдогенов (например, CYP2D6). Длинные «островки» (>30 т.п.о.) индексированных сцепленных ридов охватывают псевдогены А и А' как показано на фигуре 64. Из-за высокой гомологии последовательностей трудно определить, какой из вариантов принадлежит гену А, а какой гену А'. Более короткие варианты будут связывать один вариант псевдогенов с уникальными соседними последовательностями. Такие более короткие островки могут быть получены путем фрагментирования нуклеиновой кислоты-мишени до тагментации.

Сцепленные транспозомы

В некоторых вариантах осуществления изобретения, транспозазы являются мультимерными в транспозомном комплексе, например, они образуют димеры,

тетрамеры и т.п. в транспозомном комплексе. Авторами настоящего изобретения было неожиданно обнаружено, что сцепление мономерных транспозаз в мультимерном транспозомном комплексе или сцепление транспозонных концов транспозомного мономера в мультимерном транспозомном комплексе имеет несколько преимуществ.

5 Во-первых, сцепление транспозаз или транспозонов приводит к образованию комплексов, которые являются более стабильными и представляют крупную фракцию в активном состоянии. Во-вторых, более низкие концентрации транспозом могут быть использованы для фрагментирования посредством реакции транспозиции. В-третьих, сцепление приводит к уменьшению числа замен мозаичных концов (МЕ) транспозомных
10 комплексов, и тем самым к уменьшению вероятности смешивания штрихкодов или молекул адаптера. Такая замена МЕ-концов возможна в том случае, когда комплексы расположены на определенном расстоянии друг от друга и преобразованы, или в том случае, когда транспозомы иммобилизованы на твердом носителе посредством стрептавидина/биотина, если взаимодействие стрептавидина/биотина может нарушаться
15 и восстанавливаться, или если наблюдается контаминация. Авторы настоящего изобретения отмечают, что значительная модификация или замена МЕ-концов наблюдается в различных реакционных условиях. В некоторых вариантах осуществления изобретения, замена может составлять до 15%. Процент замен увеличивается при высокой концентрации солевого буфера и снижается в глутаматном буфере. На фигурах
20 57 и 58 представлены некоторые возможные механизмы замены МЕ.

В некоторых вариантах осуществления изобретения, субъединицы транспозазы в транспозомном комплексе могут быть связаны друг с другом ковалентными и нековалентными связями. В некоторых вариантах осуществления изобретения, мономеры транспозазы могут быть связаны до получения транспозомного комплекса
25 (перед добавлением транспозонов). В некоторых вариантах осуществления изобретения, мономеры транспозазы могут связываться после образования транспозомы.

В некоторых вариантах осуществления изобретения, нативные аминокислотные остатки могут быть заменены аминокислотой цистеином (Cys) в мультимерной пограничной области для стимуляции образования дисульфидной связи. Так, например,
30 в транспозазе Tn5, Asp468, Tyr407, Asp461, Lys459, Ser458, Gly462, Ala466, Met470 могут быть заменены Cys для стимуляции образования дисульфидной связи между мономерными субъединицами, как показано на фигурах 59 и 60. Для транспозазы Mos-1, репрезентативными аминокислотам, которые могут быть заменены цистеином, являются, но не ограничиваются ими, Leu21, Leu32, Ala35, His20, Phe17, Phe36, Ile16,
35 Thr13, Arg12, Gln10, Glu9 как показано на фиг. 61. В некоторых вариантах осуществления изобретения, модифицированные транспозазы, в которых аминокислотные остатки заменены цистеином, могут быть химически связаны друг с другом перекрестными связями под действием химического перекрестно-сшивающего линкера посредством малеимидных или пиридилдитиоловых реакционноспособных групп. Репрезентативные
40 химические перекрестно-сшивающие линкеры являются коммерчески доступными и поставляются компаниями Pierce Protein Biology/ThermoFisher Scientific (Grand Island, NY, USA).

В некоторых вариантах осуществления изобретения, транспозомные мультимерные комплексы могут быть ковалентно связаны с твердым носителем. Репрезентативными
45 твердыми носителями являются, но не ограничиваются ими, наночастицы, сферы, поверхности проточной кюветы, колоночные матрицы. В некоторых вариантах осуществления изобретения, твердые поверхности могут быть покрыты аминокислотными группами. Модифицированная транспозаза, в которой аминокислотные остатки были заменены

цистеином, может быть подвергнута химической реакции перекрестного связывания с такими аминокислотными группами посредством перекрестно-сшивающего линкера, связывающего амин с сульфгидрилом (то есть, сукцинимидил-4-(N-малеимидометил)циклогексан-1-карбоксилатом (SMCC)). Репрезентативная схема представлена на фигуре 62. В некоторых вариантах осуществления изобретения, перекрестно-сшивающий линкер, такой как малеимид-ПЭГ-биотин, может быть использован для присоединения dTnp к твердой поверхности, покрытой стрептавидином.

В некоторых вариантах осуществления изобретения, ген транспозазы может быть модифицирован для экспрессии мультимерного белка в одном полипептиде. Так, например, гены Tn5 или Mos-1 могут быть модифицированы для экспрессии двух белков Tn5 или Mos-1 в одном полипептиде. Аналогичным образом, ген транспозазы Mu может быть модифицирован так, чтобы он кодировал четыре единицы транспозазы mu в одном полипептиде.

В некоторых вариантах осуществления изобретения, транспозонные концы транспозомной мономерной единицы могут быть сцеплены с образованием сцепленного транспозомного мультимерного комплекса. Сцепление транспозонных концов позволяет встраивать сайты праймеров, секвенирующие праймеры, праймеры для амплификации или определить, какую роль может играть ДНК в образовании гДНК без фрагментирования ДНК-мишени. Встраивание таких функциональных групп имеет определенные преимущества при проведении анализов на гаплотипы или анализов на мечение области стыка, где необходимо получить информацию об интактных молекулах, или где важное значение имеет взятие части образцов из пробы. В некоторых вариантах осуществления изобретения, транспозонные концы транспозом Mu могут быть сцеплены с транспозазой Mu/транспозоном, имеющими конфигурацию типа «петли». Поскольку Mu является тетрамером, то возможны различные конфигурации, включая, но не ограничиваясь ими, сцепление R2UJ и/или R1UJ с R2J и/или R1J. При таких конфигурациях, R2UJ и R1UJ могут быть сцеплены, а могут быть и не сцеплены с R2J и R1J, соответственно. На фигуре 63 представлен транспозомный комплекс Mu, в котором транспозонные концы являются сцепленными. В некоторых вариантах осуществления изобретения, транспозонные концы Tn5 или транспозонные концы транспозом Mos1 могут быть сцепленными.

Используемый здесь термин «транспозон» означает двухцепочечную ДНК, которая имеет только нуклеотидные последовательности (последовательности «транспозонных концов»), необходимые для образования комплекса с ферментом транспозазой или интегразой, которые являются функциональными в реакции транспозиции *in vitro*. Транспозон образует «комплекс» или «синаптический комплекс» или «транспозомный комплекс» или «транспозомную композицию», содержащую транспозазу или интегразу, которая распознает транспозон и связывается с транспозоном, где указанный комплекс обладает способностью встраивать или переносить транспозон в ДНК-мишень, с которой его инкубируют в реакции транспозиции *in vitro*. Транспозон имеет две комплементарные последовательности, состоящие из «перенесенной последовательности транспозона» или «перенесенной цепи» и «неперенесенной последовательности транспозона» или «неперенесенной цепи». Так, например, один транспозон, образующий комплекс с гиперактивной транспозазой Tn5 (например, EZ-Tn5™ Transposase, EPICENTRE Biotechnologies, Madison, Wis., USA), которая является активной в реакции транспозиции *in vitro*, содержит перенесенную цепь, имеющую нижеследующую «перенесенную транспозонную последовательность»:

5' AGATGTGTATAAGAGACAG 3'

и неперенесенную цепь, имеющую нижеследующую «неперенесенную транспозонную последовательность»:

5' STGTCTCTTATACACATCT 3'.

3'-конец перенесенной цепи присоединяют к ДНК-мишени или переносят в ДНК-мишень посредством реакции транспозиции *in vitro*. Неперенесенную цепь, имеющую транспозонную последовательность, комплементарную перенесенной последовательности транспозонного конца, не присоединяют к ДНК-мишени или не переносят в ДНК-мишень посредством реакции транспозиции *in vitro*. В некоторых вариантах осуществления изобретения, транспозонные последовательности могут содержать одну или более из таких последовательностей, как штрихкод, последовательность адаптера, последовательность метки, последовательность, связывающуюся с праймером, последовательность для захвата и уникальную последовательность молекулярного идентификатора (UMI).

Используемый здесь термин «адаптер» означает последовательность нуклеиновой кислоты, которая может содержать штрихкод; последовательность, связывающуюся с праймером; последовательность для захвата; последовательность, комплементарную последовательности для захвата; уникальную последовательность молекулярного идентификатора (UMI); аффинную группу и рестрикционный сайт.

Используемый здесь термин «информация о сцеплении» означает пространственную взаимосвязь между двумя или более фрагментами ДНК, выявленную исходя из общей информации. Общий аспект такой информации может относиться к взаимодействиям смежных фрагментов, отдельных фрагментов и фрагментов, находящихся на пространственном расстоянии друг от друга. Информация, относящаяся к этим взаимодействиям, в свою очередь, облегчает иерархическую сборку и мечения ридов последовательности, происходящих от фрагментов ДНК. Такая информация о сцеплении повышает эффективность и точность такой сборки или картирования, поскольку традиционные методы сборки или картирования, применяемые в комбинации со стандартным секвенированием методом «дробовика», не позволяют точно определить родственные геномные ориджины или координаты ридов отдельных последовательностей, так как они находятся в пространственной взаимосвязи между двумя или более фрагментами ДНК, от которых происходят риды отдельных последовательностей. Поэтому, в соответствии с описанными здесь вариантами осуществления изобретения, способы сбора информации о сцеплении генов могут быть осуществлены методами сцепления генов в близких положениях для определения пространственной взаимосвязи смежных генов; методами сцепления генов в положениях средней дальности друг от друга для определения пространственной взаимосвязи отдельных генов; методами сцепления генов, расположенных далеко друг от друга для определения пространственной взаимосвязи генов на далеких расстояниях. Эти методы облегчают оценку точности и качества сборки или картирования последовательностей ДНК и могут быть применены в комбинации с любым методом секвенирования, описанным выше.

Получение информации о сцеплении генов включает сбор информации о родственных геномных ориджаинах или координатах ридов отдельных последовательностей, так как они находятся в пространственной взаимосвязи между двумя или более фрагментами ДНК, от которых происходят риды отдельных последовательностей. В некоторых вариантах осуществления изобретения, получение информации о сцеплении генов включает сбор информации о последовательностях неперекрывающихся ридов.

В некоторых вариантах осуществления изобретения, информация о сцеплении

последовательностей нуклеиновой кислоты-мишени представляет собой информацию о гаплотипе. В некоторых вариантах осуществления изобретения, информация о сцеплении последовательности нуклеиновой кислоты-мишени представляет собой информацию о геномных вариантах.

5 Используемый здесь термин «сохранение сцепления нуклеиновой кислоты-мишени» при фрагментировании нуклеиновой кислоты означает сохранение порядка расположения фрагментов нуклеиновой кислоты в одной и той же нуклеиновой кислоте-мишени.

10 Используемый здесь термин «по меньшей мере часть» и/или его грамматические эквиваленты может означать любую часть от всего количества. Так, например, «по меньшей мере часть» может означать по меньшей мере приблизительно 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99%, 99,9% или 100% от всего количества.

Используемый здесь термин «приблизительно» означает $\pm 10\%$.

15 Используемый здесь термин «секвенирующий рид» и/или его грамматические эквиваленты может относиться к повторяющимся физическим или химическим стадиям, проводимым для получения сигналов, указывающих на порядок расположения мономеров в полимере. Сигналы могут указывать на порядок расположения мономеров при разрешении одного мономера или при более низком разрешении. В конкретных
20 вариантах осуществления изобретения, стадии могут быть инициированы на нуклеиновой кислоте-мишени и осуществлены для получения сигналов, указывающих на порядок расположения оснований в нуклеиновой кислоте-мишени. Такой способ может быть осуществлен до его типичного завершения, обычно определяемого точкой, при которой сигналы больше не могут распознавать основания мишени с достаточной степенью
25 достоверности. При необходимости, завершение процедуры может быть осуществлено раньше, например, после получения нужного количества информации о последовательности. Секвенирование рида может быть осуществлено на одной молекуле нуклеиновой кислоты-мишени или одновременно на группе молекул нуклеиновой кислоты-мишени, имеющих одну и ту же последовательность, или одновременно на
30 группе нуклеиновых кислот-мишеней, имеющих различные последовательности. В некоторых вариантах осуществления изобретения, секвенирование рида прекращают после того, когда уже не могут быть получены сигналы от одной или более молекул нуклеиновой кислоты-мишени, от которых исходит этот сигнал. Так, например, секвенирование рида может быть инициировано для одной или более молекул
35 нуклеиновой кислоты-мишени, присутствующих на твердофазном субстрате, и завершено после удаления одной или более молекул нуклеиновой кислоты-мишени из субстрата. Секвенирование может быть завершено посредством какого-либо иного способа прекращения детектирования нуклеиновых кислот-мишеней, присутствующих на субстрате в начале раунда секвенирования. Репрезентативные методы секвенирования
40 описаны в патенте США No. 9029103, который во всей своей полноте вводятся в настоящее описание посредством ссылки.

Используемый здесь термин «представление секвенирования» и/или его грамматические эквиваленты могут означать информацию о порядке и типе мономерных единиц в полимере. Так, например, такая информация может указывать на порядок и
45 тип нуклеотидов в нуклеиновой кислоте. Информация может быть представлена в любом формате, включая, но не ограничиваясь ими, рисунок, изображение, электронный носитель, ряд символов, ряд чисел, ряд букв, цветовую гамму и т.п. Информация может быть получена при разрешении одного мономера или при более низком разрешении.

Репрезентативным полимером является нуклеиновая кислота, такая как ДНК или РНК, имеющая нуклеотидные единицы. Ряд букв «А», «Т», «G» и «С» означает хорошо известное представление последовательности ДНК, которое может быть скорректировано при разрешении одного нуклеотида в фактической последовательности молекулы ДНК. Другими репрезентативными полимерами являются белки, имеющие аминокислотные звенья, и полисахариды, имеющие сахаридные звенья.

Твердый носитель

Во всем описании настоящей заявки, термины «твердый носитель» и «твердая поверхность» являются синонимами. В некоторых вариантах осуществления изобретения, твердый носитель или его поверхность не являются плоскими, то есть, представляют собой внутреннюю или внешнюю поверхность пробирки или сосуда. В некоторых вариантах осуществления изобретения, твердый носитель включает микросферы или сферы. Используемый здесь термин «микросферы» или «сферы» или «частицы» или его грамматические эквиваленты означают небольшие дискретные частицы. Подходящими материалами для сфер являются, но не ограничиваются ими, пластик, керамические изделия, стекло, полистирол, метилстирол, акриловые полимеры, парамагнитные материалы, соль тория, графит, диоксид титана, латекс или перекрестно-связанные декстраны, такие как сефароза, целлюлоза, нейлон, перекрестно-связанные мицеллы и тефлон, а также любые другие материалы, которые могут быть применены для получения твердых носителей. Ценным руководством является «Руководство по детектированию микросфер» («Microsphere Detection Guide», Bangs Laboratories, Fishers Ind.). В некоторых вариантах осуществления изобретения, микросферы представляют собой магнитные микросферы или сферы. В некоторых вариантах осуществления изобретения, сферы могут иметь цветовые коды. Так, например, могут быть использованы микросферы MicroPlex®, поставляемые Luminex, Austin, TX.

Сферы необязательно должны иметь сферическую форму, например, могут быть использованы частицы неправильной формы. Альтернативно или дополнительно, сферы могут быть пористыми. Сферы имеют размеры от нанометров, то есть, приблизительно 10 нм, до миллиметров в диаметре, то есть, 1 мм, причем, предпочтительными являются сферы размером приблизительно от 0,2 микрона до 200 микрон, а особенно предпочтительно, приблизительно от 0,5 до 6 микрон, хотя в некоторых вариантах осуществления изобретения могут быть использованы сферы меньшего или большего размера. В некоторых вариантах осуществления изобретения, диаметр сфер может составлять приблизительно 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,5, 2, 2,5, 2,8, 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5, 10, 10,5, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150 или 200 мкм.

Транспозомы

«Транспозома» включает интегрирующий фермент, такой как интеграза или транспозаза и нуклеиновую кислоту, содержащую сайт распознавания интеграции, такой как сайт распознавания транспозазы. В описанных здесь вариантах, транспозаза может образовывать функциональный комплекс с сайтом распознавания транспозазы, способным катализировать реакции транспозиции. Транспозаза может связываться с сайтом распознавания транспозазы и встраивать сайт распознавания транспозазы в нуклеиновую кислоту-мишень по механизму, иногда называемому «тагментацией». В некоторых таких процедурах встраивания, одна цепь сайта распознавания транспозазы может быть перенесена в нуклеиновую кислоту-мишень. В одном из примеров, транспозома содержит димерную транспозазу, включающую две субъединицы и две несмежных транспозонных последовательности. В другом примере, транспозома

содержит транспозазу, включающую димерную транспозазу, содержащую две субъединицы и смежную транспозонную последовательность.

Некоторые варианты изобретения могут включать применение гиперактивной транспозазы Tn5 и сайта распознавания транспозазы Tn5-типа (Goryshin and Reznikoff, J. Biol. Chem., 273:7367 (1998)), или транспозазы MuA и сайта распознавания транспозазы Mu, содержащего концевые последовательности R1 и R2 (Mizuuchi, K., Cell, 35: 785, 1983; Savilahti, H, et al., EMBO J., 14: 4893, 1995). Репрезентативный сайт распознавания транспозазы, который образует комплекс с гиперактивной транспозазой Tn5 (например, EZ-Tn5™ Transposase, Epicentre Biotechnologies, Madison, Wisconsin), включает
 10 нижеследующие перенесенные цепи из 19 оснований (иногда обозначаемые «М» или «МЕ») и неперенесенные цепи: 5' AGATGTGTATAAGAGACAG 3' и 5' CTGTCTCTTATACACATCT 3', соответственно. Последовательности МЕ могут быть также использованы после их оптимизации специалистом.

Другими примерами систем транспозиции, которые могут быть применены в
 15 некоторых вариантах описанных здесь композиций и способов, являются Tn552 *Staphylococcus aureus* (Colegio et al., J. Bacteriol., 183: 2384-8, 2001; Kirby C et al., Mol. Microbiol., 43: 173-86, 2002), Ty1 (Devine & Boeke, Nucleic Acids Res., 22: 3765-72, 1994 и публикация Международной заявки WO 95/23875), транспозон Tn7 (Craig, N L, Science. 271: 1512, 1996; Craig, N L, Review in: Curr Top Microbiol Immunol., 204:27-48, 1996), Tn/O
 20 и IS10 (Kleckner N, et al., Curr Top Microbiol Immunol., 204:49-82, 1996), транспозаза Mariner (Lampe D J, et al., EMBO J., 15: 5470-9, 1996), Tc1 (Plasterk R H, Curr. Topics Microbiol. Immunol., 204: 125-43, 1996), элемент P (Gloor, G. B, Methods Mol. Biol., 260: 97-114, 2004), Tn3 (Ichikawa & Ohtsubo, J. Biol. Chem. 265:18829-32, 1990), бактериальные встраиваемые последовательности (Ohtsubo & Sekine, Curr. Top. Microbiol. Immunol. 204: 1-26, 1996),
 25 ретровирусы (Brown, et al., Proc. Natl. Acad. Sci. USA, 86:2525-9, 1989) и ретротранспозон дрожжей (Boeke & Corces, Annu Rev Microbiol. 43:403-34, 1989). Другими примерами являются IS5, Tn10, Tn903, IS911, Sleeping Beauty, SPIN, hAT, PiggyBac, Hermes, TcBuster, AeBuster1, Tol2 и сконструированные варианты ферментов семейства транспозаз (Zhang et al., (2009) PLoS Genet. 5:e1000689. Epub 2009 Oct 16; Wilson C. et al (2007) J. Microbiol.
 30 Methods 71:332-5).

Другими примерами интеграз, которые могут быть применены в описанных здесь способах и композициях, являются ретровирусные интегразы и последовательности распознавания таких ретровирусных интеграз, например, интеграз, происходящих от ВИЧ-1, ВИЧ-2, SIV, PFV-1, RSV.

35 Штрихкоды

Обычно, штрихкод может включать одну или более нуклеотидных последовательностей, которые могут быть использованы для идентификации одной или более конкретных нуклеиновых кислот. Штрихкод может представлять собой
 40 искусственную последовательность, либо он может представлять собой природную последовательность, образованную в процессе транспозиции, такую как идентичные фланкирующие геномные последовательности ДНК (г-коды) на конце ранее образованных фрагментов ДНК, находящихся в юкта-положении. В некоторых вариантах осуществления изобретения, штрихкод представляют собой искусственные последовательности, которые отсутствуют в последовательности нуклеиновой кислоты-
 45 мишени и могут быть использованы для идентификации одной или более последовательностей нуклеиновой кислоты-мишени.

Штрихкод может содержать по меньшей мере приблизительно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 или более смежных нуклеотидов. В некоторых

вариантах осуществления изобретения, штрихкод содержит по меньшей мере приблизительно 10, 20, 30, 40, 50, 60, 70 80, 90, 100 или более смежных нуклеотидов. В некоторых вариантах осуществления изобретения, по меньшей мере часть штрихкодов в группе нуклеиновых кислот, имеющих эти штрихкоды, отличаются друг от друга. В
 5 некоторых вариантах осуществления изобретения, по меньшей мере приблизительно 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99% штрихкодов являются различными. В других таких вариантах, все штрихкоды являются различными. Разнообразие различных штрихкодов в группе нуклеиновых кислот, имеющих эти штрихкоды, может быть создано рандомизированным или нерандомизированным
 10 методом.

В некоторых вариантах осуществления изобретения, транспозонная последовательность включает по меньшей мере один штрихкод. В некоторых вариантах осуществления изобретения, транспозомы включают две несцепленных транспозонных последовательности, где первая транспозонная последовательность содержит первый
 15 штрихкод, а вторая транспозонная последовательность содержит второй штрихкод. В некоторых вариантах осуществления изобретения, транспозонная последовательность включает штрихкод, содержащий первую последовательность штрихкода и вторую последовательность штрихкода. В некоторых вышеуказанных вариантах, первая последовательность штрихкода может быть идентифицирована или сконструирована
 20 так, чтобы она спаривалась со второй последовательностью штрихкода. Так, например, известная первая последовательность штрихкода может быть присоединена ко второй известной последовательности штрихкода с использованием эталонной таблицы, включающей множество первых и вторых последовательностей штрихкодов, которые, как известно, будут спариваться друг с другом.

В другом примере, первая последовательность штрихкода может включать такую же последовательность штрихкода, как и вторая последовательность штрихкода. В другом примере, первая последовательность штрихкода может включать обратный комплемент второй последовательности штрихкода. В некоторых вариантах осуществления изобретения, первая последовательность штрихкода и вторая
 30 последовательность штрихкода являются различными. Первая и вторая последовательности штрихкода могут содержать двоичный код.

В некоторых вариантах описанных здесь композиций и способов, штрихкоды используют для получения матричных нуклеиновых кислот. Совершенно очевидно, что очень большое число доступных штрихкодов позволяет присвоить каждой
 35 матричной молекуле нуклеиновой кислоты уникальный идентификационный номер. Уникальная идентификация каждой молекулы в смеси матричных нуклеиновых кислот может быть использована в различных целях. Так, например, уникально идентифицированные молекулы могут быть применены для идентификация отдельных молекул нуклеиновой кислоты; образцов, имеющих множество хромосом; геномов;
 40 клеток; типов клеток; клеточных патологий, а в частности, например, для секвенирования гаплотипов, для дифференциации родительских аллелей, для секвенирования метагенома и для секвенирования генома в образце.

Репрезентативными последовательностями штрихкодов являются, но не ограничиваются ими, TATAGCCT, ATAGAGGC, CCTATCCT, GGCTCTGA, AGGCGAAG,
 45 TAATCTTA, CAGGACGT и GTACTGAC.

Сайты праймеров.

В некоторых вариантах осуществления изобретения, транспозонная последовательность может включать «секвенирующий адаптер» или «сайт

секвенирующего адаптера», то есть, можно сказать, что он содержит область, включающую один или более сайтов, которые могут гибридизоваться с праймером. В некоторых вариантах осуществления изобретения, транспозонная последовательность может включать по меньшей мере первый сайт праймера, используемый для амплификации, секвенирования и т.п. Репрезентативными последовательностями сайтов связывания с последовательностью являются, но не ограничиваются ими, ААТГАТАСГГСГАССАССГАТСТАСАС (последовательность Р5) и СААГСАГААГАСГГСАТАСГАТ (последовательность Р7).

Нуклеиновые кислоты-мишени

Нуклеиновой кислотой-мишенью может быть любая представляющая интерес нуклеиновая кислота. Нуклеиновыми кислотами-мишенями могут быть ДНК, РНК, пептид-содержащая нуклеиновая кислота, морфолино-нуклеиновая кислота, блокированная нуклеиновая кислота, гликолевая нуклеиновая кислота, нуклеиновая кислота, содержащая треозу, смешанные образцы нуклеиновых кислот, полиплоидная ДНК (то есть, ДНК растения), а также их смеси и гибриды. В предпочтительном варианте осуществления изобретения, в качестве нуклеиновой кислоты-мишени используются гемномная ДНК или ее амплифицированные копии. В другом предпочтительном варианте осуществления изобретения используются кДНК, митохондриальная ДНК или ДНК хлоропластов. В некоторых вариантах осуществления изобретения, нуклеиновой кислотой-мишенью является мРНК.

В некоторых вариантах осуществления изобретения, нуклеиновая кислота-мишень происходит от одной клетки или из фракций одной клетки. В некоторых вариантах осуществления изобретения, нуклеиновая кислота-мишень происходит от одной органеллы. Репрезентативными моноорганеллами являются, но не ограничиваются ими, одно ядро, одна митохондрия и одна рибосома. В некоторых вариантах осуществления изобретения, нуклеиновую кислоту-мишень выделяют из залитого в парафин образца, фиксированного формалином (FFPE). В некоторых вариантах осуществления изобретения, нуклеиновой кислотой-мишенью является перекрестно-связанная нуклеиновая кислота. В некоторых вариантах осуществления изобретения, нуклеиновая кислота-мишень перекрестно связана с белком. В некоторых вариантах осуществления изобретения, нуклеиновой кислотой-мишенью является перекрестно-связанная ДНК. В некоторых вариантах осуществления изобретения, нуклеиновой кислотой-мишенью является ДНК, защищенная гистоном. В некоторых вариантах осуществления изобретения, гистоны удаляют из нуклеиновой кислоты-мишени. В некоторых вариантах осуществления изобретения, нуклеиновую кислоту-мишень выделяют из нуклеосом. В некоторых вариантах осуществления изобретения, нуклеиновую кислоту-мишень выделяют из нуклеосом, из которых были удалены ядерные белки.

Нуклеиновая кислота-мишень может содержать любую нуклеотидную последовательность. В некоторых вариантах осуществления изобретения, нуклеиновая кислота-мишень содержит гомополимерные последовательности. Нуклеиновая кислота-мишень может включать повторяющиеся последовательности. Повторяющиеся последовательности могут иметь любую длину, включая, например, 2, 5, 10, 20, 30, 40, 50, 100, 250, 500 или 1000 нуклеотидов или более. Повторяющиеся последовательности могут повторяться либо непрерывно, либо с перерывами любое количество раз, например, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 или 20 раз или более.

В некоторых описанных здесь вариантах осуществления изобретения может быть использована одна нуклеиновая кислота-мишень. В других вариантах осуществления

изобретения может быть использовано множество нуклеиновых кислот-мишеней. В таких вариантах осуществления изобретения, множество нуклеиновых кислот-мишеней может включать множество одних и тех же нуклеиновых кислот-мишеней, множество различных нуклеиновых кислот-мишеней, где некоторые нуклеиновые кислоты-мишени являются одинаковыми, или множество нуклеиновых кислот-мишеней, где все нуклеиновые кислоты-мишени являются различными. Варианты, в которых используется множество нуклеиновых кислот-мишеней, могут быть осуществлены в мультиплексном формате, так, чтобы реагенты могли доставляться одновременно к нуклеиновым кислотам-мишеням, например, в одну или более камер или на поверхность массива. В некоторых вариантах осуществления изобретения, множество нуклеиновых кислот-мишеней может включать, по существу, весь геном конкретного организма. Множество нуклеиновых кислот-мишеней может включать по меньшей мере часть генома конкретного организма, например, по меньшей мере приблизительно 1%, 5%, 10%, 25%, 50%, 75%, 80%, 85%, 90%, 95% или 99% генома. В конкретных вариантах осуществления изобретения, эта часть может иметь верхний предел, составляющий максимум приблизительно 1%, 5%, 10%, 25%, 50%, 75%, 80%, 85%, 90%, 95%, или 99% генома.

Нуклеиновые кислоты-мишени могут быть получены из любого источника. Так, например, нуклеиновые кислоты-мишени могут быть получены из молекул нуклеиновой кислоты, выделенных из одного организма, или из групп молекул нуклеиновой кислоты, выделенных из природных источников, которые включают один или более организмов. Источниками молекул нуклеиновой кислоты являются, но не ограничиваются ими, органеллы, клетки, ткани, органы или организмы. Клетками, которые могут служить в качестве источников молекул нуклеиновой кислоты-мишени, могут быть прокариотические клетки (бактериальные клетки, например, клетки бактерий рода *Escherichia*, *Bacillus*, *Serratia*, *Salmonella*, *Staphylococcus*, *Streptococcus*, *Clostridium*, *Chlamydia*, *Neisseria*, *Treponema*, *Mycoplasma*, *Borrelia*, *Legionella*, *Pseudomonas*, *Mycobacterium*, *Helicobacter*, *Erwinia*, *Agrobacterium*, *Rhizobium* и *Streptomyces*); архециты, такие как кренархециты, наноархециты или эуриархециты; или эукариотические клетки, такие как клетки грибов (например, дрожжей), растений, простейших и других паразитов, а также клетки животных (включая насекомых (например, *Drosophila* spp.), нематод (например, *Caenorhabditis elegans*) и млекопитающих (например, крыс, мышей, обезьян, приматов, не являющихся человеком, и человека). Нуклеиновые кислоты-мишени и матричные нуклеиновые кислоты могут быть обогащены некоторым представляющими интерес последовательностями с применением различных методов, хорошо известных специалистам. Примеры таких методов описаны в публикации Международной заявки No. WO/2012/108864, которая во всей своей полноте вводится в настоящее описание посредством ссылки. В некоторых вариантах осуществления изобретения, нуклеиновые кислоты могут быть дополнительно обогащены в процессе получения матричных библиотек. Так, например, нуклеиновые кислоты могут быть обогащены некоторыми последовательностями до встраивания транспозом, после встраивания транспозом и/или после амплификации нуклеиновых кислот.

Кроме того, в некоторых вариантах осуществления изобретения, нуклеиновые кислоты-мишени и/или матричные нуклеиновые кислоты могут быть подвергнуты высокой степени очистки, например, нуклеиновые кислоты могут по меньшей мере приблизительно на 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% или 100% не содержать примесей, имеющих до проведения описанных здесь способов. В некоторых вариантах осуществления изобретения, предпочтительно, применять способы, которые, как

известно специалистам, обеспечивают сохранение качества и размера нуклеиновой кислоты-мишени, например, выделение и/или прямая транспозиция ДНК-мишени могут быть осуществлены в агарозных слоях. Транспозиция может быть также осуществлена непосредственно в клетках, в популяции клеток, в лизатах и в неочищенной ДНК.

5 В некоторых вариантах осуществления изобретения, нуклеиновая кислота-мишень может быть получена из биологического образца или образца, взятого у пациента. Используемый здесь термин «биологический образец» или «образец, взятый у пациента» включает образцы, такие как ткани и физиологические жидкости. Термин «физиологические жидкости» может включать, но не ограничивается ими, кровь, сыворотку, плазму, слюну, цереброспинальную жидкость, плевральные выпоты, слезы, грудное молоко, лимфу, мокроту, мочу, амниотическую жидкость и сперму. Образец может включать физиологическую жидкость, которая является «неклеточной». Термин «неклеточная физиологическая жидкость» включает менее, чем приблизительно 1% (масс./масс.) от всего клеточного материала. Примерами неклеточных физиологических жидкостей являются плазма или сыворотка. Образцом может быть образец природного или синтетического происхождения (то есть, клеточный образец, превращенный во

10

15

внеклеточный).

В некоторых вариантах вышеописанных способов, нуклеиновая кислота-мишень может быть фрагментирована (например, путем обработки ультразвуком, гидролиза рестриктирующими ферментами или другими механическими способами) до ее обработки транспозомами.

20

Используемый здесь термин «плазма» означает неклеточную жидкость, присутствующую в крови. «Плазма» может быть выделена из крови путем удаления всех клеточных элементов крови методами, известными специалистам (например, путем центрифугирования, фильтрации и т.п.).

25

Используемые во всем описании артикли «а» или «an» означают «один или более», если это не оговорено особо.

Используемые здесь словосочетания «например», «так, например», «такой как», «включает», «включая» или их варианты не должны рассматриваться как ограничение и должны интерпретироваться как «не ограничивающие» или «без ограничений».

30

В нижеследующих примерах описаны иллюстративные варианты осуществления изобретения, которые не должны рассматриваться как ограничение объема описанного здесь изобретения.

Примеры

35 Пример 1. Выход ДНК-кластера, достигаемый с применением способа тагментации на сферах

Выход ДНК-кластера оценивали методом тагментации на сферах как показано на фигуре 3, и результаты такой оценки представлены в таблице на фигуре 4. В этом примере, 50, 250 и 1000 нг человеческой ДНК NA12878 тагментировали с использованием одной и той же партии сфер для тагментации (2,8 мкм-сферы). Вторую 50 нг-аликвоту ДНК NA12878 тагментировали с использованием второй партии сфер для тагментации (полный повтор; 2,8 мкм-сферы). Тагментированные образцы ДНК, связанные со сферами, подвергали ПЦР-амплификации и очищали. Аликвоту (5,4 мкл) каждого очищенного ПЦР-продукта (количественно не оцененного) подвергали 270-кратному разведению и получали маточные растворы образцов в концентрации приблизительно 50 пМ. Для каждого образца, 50 пМ маточного раствора разводили до 15, 19, 21 и 24 пМ. Разведенные образцы загружали на проточную кювету для получения и секвенирования кластера. Полученные данные показали, что при одном и том же

40

45

разведении (~50 пМ), число кластеров составляло 100-114% для трех различных исходных уровней (то есть, 50, 250 и 1000 нг) при использовании одной и той же серии сфер. Число кластеров для 50 нг полного повтора (на другой партии сфер) составляло 81%. Другие разведения (15, 19, 21 и 24 пМ) давали одно и то же число кластеров в концентрации приблизительно 10%. Полученные данные показали, что сферы играют значительную роль в регуляции выхода, и такой выход является воспроизводимым для различных исходных ДНК и для различных повторов.

Пример 2. Воспроизводимость способа тагментации на сферах

Воспроизводимость способа тагментации на сферах, проиллюстрированного на фигуре 3, показана на фигуре 5. В этом примере, для получения тагментированной ДНК с использованием 50 и 500 нг исходной ДНК NA12878 брали шесть различных препаратов индексированных сфер (помеченных цифрами от 1 до 6; 2,8 мкм-сферы), полученных при «одной и той же» плотности транспозом. Тагментированную ДНК подвергали ПЦР-амплификации и очищали. 12 очищенных ПЦР-продуктов объединяли с получением двух смесей (пула 1 и пула 2) из шести препаратов на двух дорожках HiSeq. Каждый пул включал 3-50 нг и 3-500 нг образцов на дорожку. Данные в таблице 500 представляют собой медианный размер вставки и средний размер вставки для каждого индексированного образца.

Пример 3. Размер вставки пула 1 и размер вставки пула 2

Размер вставки пула 1 и размер вставки пула 2 представлены на фигуре 6А (график 600) и представлены на фигуре 6В (график 650), соответственно, для индексированных образцов, проиллюстрированных на фигуре 5. Эти данные также показали, что размер вставок равномерно распределен по шести различным препаратам индексированных сфер. Тагментация на сферах представляет собой механизм регуляции размера вставок и выхода ДНК.

Пример 4. Воспроизводимость общего числа ридов.

Воспроизводимость общего числа ридов и процента выровненных ридов в эксперименте, описанном на фигуре 5, проиллюстрирована на фигуре 7 (гистограмма 700). При обеих исходных концентрациях (50 нг и 500 нг), общее число ридов является аналогичным для одного и того же препарата, полученного на индексированных сферах. Четыре из шести препаратов, полученных на индексированных сферах (индексы 1, 2, 3 и 6), имеют почти одинаковые выходы; при этом препараты на индексированных сферах 4 и 5 имеют некоторую степень вариабельности, которая может быть обусловлена присутствием индексной последовательности.

В одном из применений, способ тагментации на сферах может быть применен в анализе на обогащение экзома, где указанный анализ включает стадию тагментации, например, осуществляемую в соответствии с протоколом Illumina's Nextera® Rapid Capture Enrichment. В этом анализе на обогащение экзома (то есть, проводимом по протоколу Illumina's Nextera® Rapid Capture Enrichment), для фрагментирования геномной ДНК была проведена тагментация в растворе (Nextera). Затем для получения представляющих специфических генных фрагментов использовали геноспецифические праймеры. После этого осуществляли два цикла обогащения, а затем полученные фрагменты обогащали с помощью ПЦР и секвенировали.

Для оценки возможности применения способа тагментации на сферах в анализе на обогащение экзома, человеческую ДНК NA12878 тагментировали с использованием 25, 50, 100, 150, 200 и 500 нг исходной ДНК. Контрольную библиотеку (NA00536) получали из 50 нг исходной ДНК в соответствии со стандартным протоколом. Каждая исходная ДНК имеет свой индекс (уникальный идентификатор). В соответствии со

стандартными методами и для гарантии получения достаточного количества фрагментов проводили десять циклов ПЦР с использованием активированной маточной смеси полимераз (ЕРМ). В соответствии с протоколом, амплификацию проводили в течение 3 минут при 72°C, 30 секунд при 98°C, а затем 10 циклов по 10 секунд при 98°C, 30 секунд при 65°C и 1 минуту при 72°C. Затем образцы хранили при 10°C. После этого, образцы обрабатывали методом обогащения экзона и секвенировали.

Пример 5. Размер вставок в контрольной библиотеке и в тагментированной библиотеке на сферах в анализе на обогащение экзона

На фигурах 8А, 8В и 8С представлен график 800, на котором указан размер вставки в контрольной библиотеке, график 820, на котором указан размер вставки в тагментированной библиотеке на сферах, и данные, систематизированные в таблице 840, соответственно, для анализа на обогащение экзона. Эти данные показали, что тагментированные библиотеки на сферах имели более широкий диапазон размеров вставки по сравнению с контрольной библиотекой, но при этом, размер вставки почти не зависит от размера исходной ДНК в образцах.

Пример 6. Оценка качества последовательностей ридов

На фигурах 9А, 9В и 9С представлены: гистограмма 900, на которой представлены процент дубликаций, проходящих через фильтры (дубликаций РF); гистограмма 920 для РСТ-выбранных оснований и гистограмма 940 для оснований, используемых в РСТ на мишени, соответственно, в анализе на обогащение экзона, описанном на фигурах 8А, 8В и 8С. Как показано на фигуре 9А, процент дубликаций РF является показателем числа дублицированных ридов на проточной кювете. В идеальном случае, это число будет низким (как в данном случае) и будет гарантировать, что все кластеры дадут данные, которые могут быть использованы для получения нужных результатов.

На фигуре 9В представлены РСТ-отобранные основания, которые определяют число ридов, последовательности которых расположены в представляющем интерес сайте или поблизости от этого сайта, и которые должны быть обогащены при проведении процедуры обогащения. В идеальном случае, это число будет близким к единице, что будет указывать на успешное осуществление процедуры обогащения и на то, что риды, которые не должны быть обогащены, не пройдут такую процедуру.

На фигуре 9С представлены основания, используемые для РСТ на мишени, которые определяют число ридов, имеющих фактическую последовательность по всем представляющим интерес конкретным основаниям в обогащенной области. В идеальном случае, все обогащенные риды должны иметь последовательность по всем представляющим интерес конкретным основаниям в обогащенном риде, но из-за рандомизированного характера тагментации и вариабельности длин вставок, могут быть обогащены риды, которые не были полностью секвенированы по всей представляющей интерес области.

Для оптимизации распределения размеров вставок были применены два метода. В одном из примеров, для удаления фрагментов, которые являются слишком мелкими или слишком крупными, может быть применен метод SPRI-очистки. SPRI-очистка представляет собой способ удаления фрагментов, размер которых выше или ниже нужного размера, посредством селективной преципитации ДНК исходя из размера и удерживания осажденной или неосажденной ДНК, если это необходимо (то есть, первую стадию проводят для осаждения только той ДНК, размер которой больше нужного размера, и удерживания более мелких растворимых фрагментов). Затем, более мелкие фрагменты осаждают, и в это время нежелательные очень мелкие фрагменты (которые еще присутствуют в растворе) удаляют и осажденную ДНК удерживают, промывают

и повторно солибилизируют с получением ДНК нужных размеров. В другом примере, пространство активных транспозом на поверхности сфер может быть использовано для регуляции распределения размеров вставок. Так, например, бреши на поверхности сфер могут быть заполнены неактивными транспозомами (например, транспозомами, содержащими неактивные транспозоны).

Был проведен анализ сцепления в способе тагментации на сферах. В таблице 3 указано число случаев, в которых 0, 1, 2 или 3 ряда присутствуют 1000 п.о.-окнах с общим индексом. Были получены сферы с 9 различными индексированными транспозомами, и эти сферы были использованы для тагментации небольшого количества человеческих ДНК. Полученные ряды выравнивали и анализировали на их число в 1000 п.о.- или 10 т.п.о.-окнах, имеющих один и тот же индекс. Некоторые ряды в небольшом окне с общим индексом могут продуцироваться случайно, и предсказание вероятности продуцирования таких рядов проиллюстрировано в таблице 3 и в таблице 4 в строке «случайное продуцирование». Числа в строке «Сферы» означают фактическое число рядов с общим индексом в 1000 п.о.-окне (таблица 3) или 10 т.п.о.-окне (таблица 4). Как показано в таблице 3 и таблице 4, фактическое число обнаруженных одинаковых индексов в 1000 п.о.-окне или в 10 т.п.о.-окне значительно выше, чем это ожидается в случае случайного продуцирования. «0» окон означает, число случаев, в которых конкретное 1000 п.о.-окно не имело индексированных рядов, картированных в этом окне. Это число является наибольшим, поскольку было секвенировано лишь очень небольшое количество человеческих геномов, а большинство окон не имело рядов, выровненных по этим окнам. «1» означает число случаев, в которых только один ряд был картирован по 1000 п.о. (или по 10 т.п.о.)-окну; «2» означает число случаев, в которых 2 ряда имеют общий индекс в 1000 п.о. (или по 10 т.п.о.)-окне; и т.п. Эти данные позволяют предположить, что во всех 1400 случаях, один и тот же фрагмент ДНК (свыше 10 т.п.о.) был тагментирован на одной и той же сфере по меньшей мере от двух до 5 раз из всех приблизительно 15000 событий тагментации. Поскольку эти фрагменты имеют общий индекс, то маловероятно, но возможно, что эти фрагменты происходят от одной и той же сферы. Если фрагменты имеют один и тот же индекс, то маловероятно, что они образовались случайно, а скорее всего, они происходят от одной и той же сферы.

Таблица 3. Число рядов в 1000 п.о.-окнах с общим индексом

	0	1	2	3
Сфера	25913666	15220	305	7
Случайное продуцирование	25913334	15855	9	0

В таблице 4 указано число рядов (до 5) в 10 т.п.о.-окнах с общим индексом.

Таблица 4. Число рядов в 10 т.п.о.-окнах с общим индексом

	0	1	2	3	4	5
Сфера	2578669	12683	1267	169	28	3
Случайное продуцирование	2577012	15742	64	1	0	0

Пример 7. Отделение свободных транспозом от СРТ-ДНК

После транспозиции, реакционную смесь, содержащую СРТ-ДНК и свободные транспозомы, подвергали колоночной хроматографии на сефакриле S-400 и эксклюзионной хроматографии на сефакриле S-200, и полученные данные были представлены на фигуре 22. СРТ-ДНК обозначали как NCP ДНК.

Пример 8. Оптимизация плотность зондов для захвата на сферах

Плотности зондов А7 и В7 для захвата были оптимизированы на 1 мкм-сферах, и результаты такой оптимизации представлены на фигуре 25. Дорожки 1 (А7) и 3 (В7) имели более высокие плотности зондов, а дорожки 2 (А7) и 4 (В7) имели плотности зондов 10000-100000 на 1 мкм-сферу. Продукт лигирования зонда для захвата с молекулой-мишенью анализировали в агарозном геле. Зонды с плотностью приблизительно 10000-100000 на сферу давали более высокую эффективность лигирования, чем зонды с более высокой плотностью.

Пример 9. Тесты на возможность получения индексированных секвенирующих библиотек СРТ-ДНК на сферах методом внутримолекулярной гибридизации.

Транспозомы были получены путем смешивания транспозонов, имеющих последовательности для захвата А7' и В7', которые являются комплементарными последовательностям для захвата А7 и В7 на сферах с гиперактивной транспозазой Tn5. Высокомолекулярную геномную ДНК смешивали с транспозомами, в результате чего получали СРТ-ДНК. Отдельно были получены сферы с иммобилизованными олигонуклеотидами: Р5-А7, Р7-В7 или Р5-А7+Р7-В7, где Р5 и Р7 представляют собой последовательности, связывающиеся с праймерами, а А7 и В7 представляют собой последовательности для захвата, комплементарные последовательностям А7' и В7', соответственно. Сферы, содержащие только Р5-А7, только Р7-В7, Р5-А7+Р7-В7 или смесь сфер Р5-А7 и Р7-В7, обрабатывали СРТ-ДНК, и к реакционной смеси добавляли лигазу для определения эффективности гибридизации иммобилизованных олигонуклеотидов с перенесенной ДНК. Результаты представлены на фигуре 26. Секвенирование библиотек осуществляли только в том случае, когда Р5-А7 и Р7-В7 были иммобилизованы вместе на одной сфере (дорожка 4), на что указывали полосы высокомолекулярного продукта на агарозном геле. Результаты указывали на высокую эффективность внутримолекулярной гибридизации и доказали целесообразность получения индексированных секвенирующих библиотек СРТ-ДНК на сферах методом внутримолекулярной гибридизации.

Пример 10. Тесты на целесообразность клонального индексирования.

Было получено несколько серий транспозом. В одной серии, гиперактивную транспозазу Tn5 смешивали с транспозонными последовательностями Tnp1, имеющими 5'-биотин, в результате чего получали транспозому 1. В другой серии, Tnp2, имеющие уникальный индекс 2, смешивали с 5'-биотином, в результате чего получали транспозому 2. В другой серии, гиперактивную транспозазу Tn5 смешивали с транспозонными последовательностями Tnp3, имеющими 5'-биотин, в результате чего получали транспозому 3. В другой серии, Tnp4, имеющие уникальный индекс 4, смешивали с 5'-биотином, в результате чего получали транспозому 4. Каждую из транспозом 1 и 2 и транспозом 3 и 4 отдельно смешивали со стрептавидиновыми сферами, в результате чего получали серию сфер 1 и серию сфер 2. Затем две серии сфер смешивали и инкубировали с геномной ДНК и буфером для тагментации, что приводило к стимуляции тагментации геномной ДНК. После этого осуществляли ПЦР-амплификацию тагментированных последовательностей. Амплифицированную ДНК секвенировали для анализа на встраивание индексных последовательностей. Если тагментация органичивается сферами, то большинство фрагментов будут иметь индексы Tnp1/Tnp2 и Tnp3/Tnp4. В случае внутримолекулярной гибридизации, фрагменты могут иметь индексы Tnp1/Tnp4, Tnp2/Tnp3, Tnp1/Tnp3 и Tnp2/Tnp4. Результаты секвенирования после 5 и 10 циклов ПЦР представлены на фигуре 27. Контроль имеет все четыре транспозона, смешанные друг с другом и иммобилизованные на сферах. Результаты показали, что большинство последовательностей имеют индексы Tnp1/Tnp2 или Tnp3/

Тпр4, что подтверждает целесообразность клонального индексирования. У контроля, каких-либо различий в индексах не наблюдалось.

Пример 11. Транспозиция индексированных клональных сфер за одну реакцию

Было получено 96 серий индексированных сфер с транспозомами. Отдельные
 5 индексированные транспозомы были получены путем смешивания транспозона, содержащего олигонуклеотид, включающий мозаичную концевую последовательность Tn5 (ME) у 5'-конца и индексную последовательность. Отдельные индексированные транспозомы иммобилизовали на сферах посредством взаимодействия «стрептавидин-биотин». Транспозомы, присутствующие на сферах, промывали, и все 96 отдельно
 10 индексированных транспозом на сферах объединяли. Олигонуклеотиды, комплементарные ME-последовательности и содержащие индексную последовательность, гибридизовали с иммобилизованным олигонуклеотидом, в результате чего получали транспозоны с уникальными индексами. 96 серий индексированных клональных сфер с транспозомами объединяли и инкубировали с
 15 высокомолекулярной (HMW) геномной ДНК в присутствии буфера для тагментации Nextera в одной пробирке.

Сферы промывали и транспозазу удаляли путем обработки реакционной смеси 0,1% ДСН. Тагментированную ДНК амплифицировали с индексированными праймерами и секвенировали в проточной кювете PE HiSeq v2 с использованием набора для получения
 20 кластера TrueSeq v3, после чего данные секвенирования анализировали.

Наблюдалось образование кластеров или островков ридов. График, на котором проиллюстрированы наименьшие расстояния между ридами для каждой последовательности, указывал, в основном, на присутствие главных пиков, то есть, одного пика в присутствии кластера (проксимального пика) и другого пика между
 25 кластерами (дистального пика). Этот метод и его результаты схематически представлены на фигурах 30 и 31. Размеры островков составляют приблизительно 3-10 т.п.о. Процент охватываемых оснований составляет приблизительно 5-10%. Размеры вставок геномной ДНК составляют приблизительно 200-300 оснований.

Пример 12. Размеры библиотек для транспозом, присутствующих на сферах
 30 Сначала транспозомы собирали в растворе путем смешивания первого олигонуклеотида, имеющего последовательность ME', второго олигонуклеотида, имеющего последовательность ME-штрихкод-P5/P7, и транспозазы Tn5. В первой стадии, первый олигонуклеотид, имеющий последовательность ME', биотинилировали у 3'-конца. Во второй стадии, олигонуклеотид, имеющий последовательность ME-
 35 штрихкод-P5/P7, биотинилировали у 5'-конца. К каждой из полученных транспозом в различных концентрациях (10 нМ, 50 нМ и 200 нМ) добавляли серии стрептавидиновых сфер так, чтобы транспозомы были иммобилизованы на этих стрептавидиновых сферах. Сферы промывали и добавляли высокомолекулярную геномную ДНК, а затем осуществляли тагментацию. В некоторых случаях, тагментированную ДНК
 40 обрабатывали 0,1% ДСН, а в других случаях, тагментированную ДНК не обрабатывали. Тагментированную ДНК подвергали ПЦР-амплификации за 5-8 циклов и секвенировали. Эта процедура схематически представлена на фигуре 32.

Как показано на фигуре 33, ДСН-обработка повышала эффективность амплификации и качество секвенирования. Олигонуклеотиды с 3'-биотином имели большие размеры
 45 библиотеки для транспозом.

На фигуре 34 проиллюстрировано влияние плотности поверхности транспозомы на размер вставки. Транспозомы с 5'-биотином имели меньшие размеры библиотеки и давали большее число побочных продуктов аутоинсерции.

Пример 13. Титрование исходной ДНК

Различные количества высокомолекулярной ДНК-мишени добавляли к клонально индексируемым сферам с плотностью 50 мМ Тп5:транспозон, и инкубировали в течение 15 или 60 минут при 37°C или в течение 60 минут при комнатной температуре.

5 Транспозомы состояли из олигонуклеотидов с 3'-биотином. Затем осуществляли тагментацию и реакционную смесь обрабатывали 0,1% ДСН и подвергали ПЦР-амплификации. Амплифицированную ДНК секвенировали. На фигуре 35 проиллюстрировано влияние исходной ДНК на распределение по размеру. Реакции с 10 пг исходной ДНК давали наименьший сигнал. Характер распределения по размеру
10 был аналогичным для исходных ДНК в концентрации 20, 40 и 200 пг.

Пример 14. Размер и распределение островков, оцениваемые методами, осуществляемыми в растворе и на сферах

Было проведено сравнение размера и распределения островков, оцениваемых методами, осуществляемыми в растворе и на сферах. В методе, осуществляемом в
15 растворе, 96 транспозом, каждая из которых имела транспозоны с уникальным индексом, подвергали сборке в 96-луночном планшете. Затем добавляли высокомолекулярную геномную ДНК и осуществляли реакцию тагментации. Реакционный продукт обрабатывали 0,1% ДСН и подвергали ПЦР-амплификации. Амплифицированные продукты секвенировали.

20 В методе, осуществляемом на сферах, 96 транспозом, каждая из которых имела транспозоны с уникальным индексом, подвергали сборке в 96-луночном планшете. Олигонуклеотиды состояли из 3'-концевого биотина. В каждый 96-луночный планшет добавляли стрептавидиновые сферы и смесь инкубировали так, чтобы транспозомы были иммобилизованы на стрептавидиновых сферах. Сферы отдельно промывали и
25 собирали, а затем добавляли высокомолекулярную геномную ДНК и осуществляли реакцию тагментации в одном реакционном сосуде (в одном сосуде). Реакционный продукт обрабатывали 0,1% ДСН и подвергали ПЦР-амплификации. Амплифицированные продукты секвенировали.

В случае негативного контроля, все 96 транспозонных последовательностей, каждая
30 из которых имела уникальный индекс, смешивали. Олигонуклеотиды состояли из 3'-концевого биотина. Транспозомы получали из отдельно смешанных индексируемых транспозонов. Затем к смеси добавляли стрептавидиновые сферы. После этого добавляли высокомолекулярную геномную ДНК и осуществляли реакцию тагментации. Реакционный продукт обрабатывали 0,1% ДСН и подвергали ПЦР-амплификации.
35 Амплифицированные продукты секвенировали.

Была построена кривая зависимости числа ридов внутри островков от размера островков. Результаты, представленные на фигуре 36, показали, что островки (проксимальные риды), наблюдаемые при использовании клональных индексируемых сфер в одном сосуде, были аналогичны островкам, наблюдаемым при проведении
40 метода в растворе. При смешивании индексируемых транспозонов до образования транспозомы, островки (проксимальные риды) не наблюдались. Смешивание транспозонов до образования транспозомы приводило к образованию сфер с различными индексами/транспозомами на сферу, то есть, неклональных сфер.

Пример 15. Структурный анализ варианта с помощью CPT-seq

45 Детектирование гетерозиготной делеции в 60 т.п.о.

Данные секвенирования были представлены в виде файлов fastq и подвергнуты разделению для создания отдельного файла fastq для каждого штрихкода. Файлы fastq, созданные после секвенирования CPT, подвергали разделению в соответствии с их

индексами и сопоставляли путем выравнивания с эталонным геномом с последующим удалением дубликатов. Хромосомы сканировали по окну 5 т.п.о./1 т.п.о., в котором регистрировали число индексов, указывающих на то, что любые риды находятся в окне сканирования. С точки зрения статистики, для гетерозиготной области делеции, по сравнению с ее соседними областями, для создания библиотеки доступна только половина от всего количества ДНК, а поэтому число индексов должно составлять приблизительно половину, как их соседняя половина. Гетерозиготная делеция 60 т.п.о. в хромосоме 1 NA12878 была обнаружена после сканирования по окну 5 т.п.о. для 9216 индексированных данных секвенирования СРТ и представлена на фиг. 47А и 47В.

Детектирование сцепления генов

Файлы fastq, созданные после секвенирования СРТ, подвергали разделению в соответствии с их индексами и сопоставляли путем выравнивания с эталонным геномом с последующим удалением дубликатов. Хромосомы сканировали по 2 т.п.о.-окну. Каждое 2 т.п.о.-окно представляет собой вектор 36864, в котором каждый элемент регистрирует число ридов, имеющих уникальный индекс и присутствующих в этом 2 т.п.о.-окне. Для каждой пары 2 т.п.о.-окон (X,Y) по всему геному вычисляли взвешенный индекс Джаккарда. Этот индекс *de facto* указывает на расстояние между парами (X,Y) в образце. Эти индексы проиллюстрированы в виде тепловой карты, представленной на фиг. 48, где каждая экспериментальная точка представляет пару 2 т.п.о.-окон сканирования; левый верхний квадрат относится к паре X,Y области 1, нижний правый квадрат относится к паре X,Y области 2, а верхний правый квадрат относится к паре X,Y области пересечения областей 1 и 2. Сигнал сцепления генов представлен горизонтальной линией в середине этого графика.

Детектирование делеций

Файлы fastq, созданные после секвенирования СРТ, подвергали разделению в соответствии с их индексами и сопоставляли путем выравнивания с эталонным геномом с последующим удалением дубликатов. Хромосомы сканировали по 2 т.п.о.-окну. Результаты детектирования генетических делеций представлены на фиг. 49.

Пример 16. Детектирование фазирования и метилирования

Оптимизация эффективности превращения под действием бисульфита

Превращение оценивали в области МЕ (области мозаичного элемента) и гДНК для сцепленных с индексом библиотек CPT-Seq на сферах. Систему превращения под действием бисульфита (Promega's MethylEdge Bisulfite Conversion system) оптимизировали для повышения эффективности.

Условие	ДНК	Сферы	BSC-обработка
1	10 нг	Нет	1 час @ 60°C/0,3М NaOH
2		Да	1 час @ 60°C/0,3М NaOH
3			1 час @ 60°C/1М NaOH
4			1 час @ 65°C/0,3М NaOH

МЕ-последовательности анализировали для определения эффективности обработки путем бисульфитного превращения, и результаты анализа представлены на фиг. 50. Бисульфитное превращение (BSC) для сцепленных с индексами библиотек, связанных со сферами, составляло 95%. Аналогичные ПЦР-выходы, наблюдаемые в условиях обработки дисульфитом > более жестких условиях обработки дисульфитом, не приводили к разложению библиотек, как показано на фиг. 51. При этом наблюдалось приблизительно 95%-ное BSC-превращение сцепленных с индексами библиотек на сферах. Переменными величинами, исследуемыми для оценки улучшения качества BSC (C®U), являются температура и концентрация NaOH (денатурация). Условия 60°C и 1М

NaOH или °C и 0,3 М NaOH давали хороший результат.

После секвенирования BSC-превращенных CPT-seq-библиотек на сферах наблюдалась ожидаемая структура секвенирующего ряда. Процентные соотношения оснований представлены на графике IVC на фиг. 52.

5 На фиг. 53 проиллюстрирован электрофорез в агарозном геле для сцепленных с индексами библиотек после проведения ПЦР и после бисульфитного превращения. При этом наблюдались библиотеки ожидаемого размера в пределах 200-500 п.о. Реакция, проводимая в отсутствие ДНК, не давала каких-либо сцепленных с индексами библиотек.

Пример 17. Целевое фазирование

10 Было проведено обогащение сцепленных с индексами библиотек CPT-seq во всем геноме. На фиг. 54 проиллюстрировано мечение с использованием биоанализатора для сцепленных с индексами библиотек CPT-seq во всем геноме перед обогащением без выбора по размеру. На фиг. 55 проиллюстрирован проводимый в агарозном геле анализ библиотек после обогащения.

15 Статистические методы обогащения для области HLA представлены ниже:

	Образец ID	C3
	Название образца:	HLA-зонды
	Размер слоя:	150
	Общая длина эталона мишени:	5062748
20	Общее число ридов, прошедших через фильтр (PF):	2516
	Процент Q30:	94,90%
	Общее число выровненных ридов:	2498
	Процент выровненных ридов:	99,40%
	Выровненные риды мишени:	840
	Обогащение ридов:	30,80%
25	Процент спаренных дублированных ридов:	12,70%
	Медианная длина фрагмента:	195

На фигуре 56 представлены результаты целевого гаплотипирования области HLA в хромосоме. Слева проиллюстрировано обогащение библиотеки сцепленных с индексами ридов во всем геноме. Каждая небольшая гистограмма представляет короткую индексированную библиотеку. Кластерами индексированных библиотек являются «островки», то есть, области, которые были клонально индексированы на одной сфере с одним и тем же индексом, и следовательно, эти риды являются проксимальными (типа «островков») на уровне генома. Обогащение (см. «Селективное обогащение нуклеиновых кислот» (Selective enrichment of nucleic acids) в заявке WO 2012108864 A1) библиотек в области мишени представлено справа. Риды обогащали для HLA-области. Кроме того, если риды были отобраны по индексам и выровнены с геномом, то их снова представляли в виде структуры «островков», что указывает на сохранение информации о целостности сцепленных с индексами ридов.

Пример 18. Замены индексов

40 Для оценки замен мозаичных концов (ME) транспозомных комплексов были получены сферы с различными индексами. После смешивания определяли замены индексов путем секвенирования библиотек, а затем регистрировали эти индексы для каждой библиотеки. Проценты «замененных» индексов были вычислены по формуле: $(D4+D5+E3+E5+f4)/$ (сумма из всех 96) и представлены на фигуре 65.

45 Пример 19. Снижение размера вставки библиотеки путем получения более плотных групп транспозомных комплексов, связанных со стрептавидиновыми сферами

Стрептавидиновые магнитные сферы нагружали 1х, 6х и 12х концентрациями транспозомного комплекса TsTn5. Протокол секвенирования Epi-CPT осуществляли

для сферы каждого типа. Конечный ПЦР-продукт загружали на биоанализатор Agilent BioAnalyzer для анализа, проиллюстрированного на фигуре. Фрагменты библиотек Epi-CPT seq были меньше и давали больший выход при загрузке большего количества TsTn5 на сферы.

Пример 20. Фрагментирование библиотеки ДНК в процессе превращения под действием бисульфата натрия

После превращения под действием бисульфита, ДНК разрушалась, что приводило к потере общих последовательностей (CS2), необходимых для ПЦР-амплификации. Библиотеки фрагментов ДНК CPTSeq и Epi-CPTSeq (Me-CPTSeq) анализировали на биоанализаторе. Из-за разрушения ДНК в процессе бисульфитного превращения, библиотека Epi-CPTSeq давала в 5 раз меньший выход и менее равномерное распределение по размеру по сравнению с библиотекой CPTSeq, как показано на фигуре 70.

Пример 21. TdT-опосредуемая реакция лигирования оцДНК

Была протестирована возможность восстановления концов ДНК посредством лигирования, опосредуемого концевой трансферазой (TdT). Вкратце, 5 пмоль матричной оцДНК инкубировали с TdT (10/50 ед.), с дуплексом «аттенуатор/адаптер» (0/15/25 пмоль) и с ДНК-лигазой (0/10 ед.) в течение 15 минут при 37°C. Продукты удлинения/лигирования ДНК анализировали на геле, содержащем TBE-мочевину, и полученные результаты были представлены на фигуре 71. Добавление всех реакционных компонентов приводило к почти полному лигированию молекулы адаптера (дорожки 5-8).

Возможность восстановления концов ДНК посредством лигирования, опосредуемого концевой трансферазой (TdT), была протестирована для связанной со сферой библиотеки, подвергнутой реакции превращения бисульфатом натрия, и результаты были представлены на фигуре 72. Вкратце, ДНК тагментировали на сферах (первые две дорожки), обрабатывали реагентами, входящими в набор для реакции превращения под действием бисульфата Promega's MethylEdge (дорожки 3 и 4) и подвергали процедуре «спасения» ДНК в соответствии с протоколом (дорожки 5 и 6). После реакции «спасения» наблюдалось явное увеличение выхода и размера библиотеки ДНК. При этом также наблюдалось увеличение количества аутоинсерционных транспозонов (SI), что указывало на эффективное лигирование молекулы адаптера.

Результаты анализа методом секвенирования метил-CPTSeq представлены на фигуре 73.

(57) Формула изобретения

1. Способ получения библиотеки фрагментов ДНК нуклеиновой кислоты-мишени со штрихкодами, где указанный способ включает:

а. контактирование нуклеиновой кислоты-мишени с множеством транспозомных комплексов, каждый из которых включает:

транспозоны и транспозазы, где транспозоны содержат перенесенные цепи и неперенесенные цепи и где по меньшей мере один из транспозонов транспозомного комплекса включает последовательность адаптера, способную гибридизоваться с комплементарной последовательностью для захвата;

б. фрагментирование нуклеиновой кислоты-мишени с образованием множества фрагментов и встраивание множества перенесенных цепей в 5'-конец по меньшей мере одной цепи фрагментов с сохранением сцепления нуклеиновой кислоты-мишени;

с. контактирование множества фрагментов нуклеиновой кислоты-мишени со

множеством твердых носителей, каждый из которых содержит множество иммобилизованных олигонуклеотидов, где каждый из этих олигонуклеотидов включает комплементарную последовательность для захвата и первую последовательность со штрихкодом и где первая последовательность со штрихкодом, происходящая от каждого

5 твердого носителя из множества твердых носителей, отличается от первой последовательности со штрихкодом, происходящей от других твердых носителей из множества твердых носителей;

d. перенос информации о последовательности со штрихкодом во фрагменты нуклеиновой кислоты-мишени с получением библиотеки двухцепочечных фрагментов,

10 где по меньшей мере одна цепь помечена у 5'-конца первым штрихкодом и где по меньшей мере два фрагмента одной и той же нуклеиновой кислоты-мишени получают идентичную информацию по штрихкоду.

2. Способ по п. 1, где одна последовательность со штрихкодом присутствует во множестве иммобилизованных олигонуклеотидов на каждом отдельном твердом

15 носителе.

3. Способ по п. 1, где различные последовательности со штрихкодом присутствуют во множестве иммобилизованных олигонуклеотидов на каждом отдельном твердом носителе.

4. Способ по любому из пп. 1-3, где перенос информации о последовательности со штрихкодом во фрагменты нуклеиновой кислоты-мишени осуществляют путем

20 лигирования.

5. Способ по любому из пп. 1-3, где перенос информации о последовательности со штрихкодом во фрагменты нуклеиновой кислоты-мишени осуществляют путем удлинения под действием полимеразы.

6. Способ по любому из пп. 1-3, где перенос информации о последовательности со штрихкодом во фрагменты нуклеиновой кислоты-мишени осуществляют путем лигирования и удлинения под действием полимеразы.

25

7. Способ по любому из пп. 5, 6, где удлинение под действием полимеразы осуществляют путем удлинения 3'-конца нелигированной транспозонной цепи под действием ДНК-полимеразы с использованием лигированного иммобилизованного олигонуклеотида в качестве матрицы.

30

8. Способ по любому из пп. 1-7, где по меньшей мере часть последовательностей адаптера также включает вторую последовательность со штрихкодом.

9. Способ по любому из пп. 1-8, где транспозомные комплексы являются

35 мультимерными и где последовательности адаптера транспозонов каждой мономерной единицы отличаются от другой мономерной единицы в одном и том же транспозомном комплексе.

10. Способ по любому из пп. 1-9, где последовательность адаптера дополнительно включает первую последовательность, связывающуюся с праймером.

11. Способ по п. 10, где первый сайт связывания с праймером не имеет последовательности, гомологичной последовательности для захвата или комплементу последовательности для захвата.

40

12. Способ по любому из пп. 1-11, где олигонуклеотиды, иммобилизованные на твердом носителе, также содержат вторые последовательности, связывающиеся с праймером.

45

13. Способ по пп. 1-12, где транспозомные комплексы являются мультимерными и где транспозомные мономерные единицы связаны друг с другом в одном и том же транспозомном комплексе.

14. Способ по п. 13, где транспозаза транспозомной мономерной единицы связана с другой транспозазой другой транспозомной мономерной единицы одного и того же транспозомного комплекса.

15. Способ по п. 13, где транспозоны транспозомной мономерной единицы связаны с транспозонами другой транспозомной мономерной единицы одного и того же транспозомного комплекса.

16. Способ по любому из пп. 1-15, где информация о сцеплении последовательности нуклеиновой кислоты-мишени представляет собой информацию о гаплотипе.

17. Способ по любому из пп. 1-15, где информация о сцеплении последовательности нуклеиновой кислоты-мишени представляет собой информацию о геномных вариантах.

18. Способ по п. 17, где геномные варианты выбраны из группы, состоящей из делеций, транслокаций, межхромосомных сцеплений генов, дупликаций и паралогов.

19. Способ по любому из пп. 1-18, где олигонуклеотиды, иммобилизованные на твердом носителе, содержат частично двухцепочечную область и частично одноцепочечную область.

20. Способ по п. 19, где частично одноцепочечная область олигонуклеотида содержит вторую последовательность со штрихкодом и вторую последовательность, связывающуюся с праймером.

21. Способ по любому из пп. 1-20, где фрагменты нуклеиновой кислоты-мишени, имеющие штрихкоды, амплифицируют, а затем определяют последовательность фрагментов нуклеиновой кислоты-мишени.

22. Способ по п. 21, где стадии (а)-(d) и последующую амплификацию проводят в одном реакционном сосуде до определения последовательности фрагментов нуклеиновой кислоты-мишени.

23. Способ по п. 21, где третью последовательность со штрихкодом вводят во фрагменты нуклеиновой кислоты-мишени во время амплификации.

24. Способ по любому из пп. 1-22, который также включает:
объединение фрагментов нуклеиновой кислоты-мишени, содержащих штрихкоды стадии (d) из множества первой серии реакционных сосудов в пул фрагментов нуклеиновой кислоты-мишени, содержащих штрихкоды;

перераспределение пула фрагментов нуклеиновой кислоты-мишени, содержащих штрихкоды, с получением множества второй серии реакционных сосудов; и
введение третьего штрихкода во фрагменты нуклеиновой кислоты-мишени путем амплификации фрагментов нуклеиновой кислоты-мишени во вторую серию реакционных сосудов до секвенирования.

25. Способ по любому из пп. 1-24, который также включает предварительное фрагментирование нуклеиновой кислоты-мишени до контактирования нуклеиновой кислоты-мишени с транспозомными комплексами.

26. Способ по п. 25, где предварительное фрагментирование нуклеиновой кислоты-мишени представляет собой способ, выбранный из группы, состоящей из обработки ультразвуком и расщеплением рестриктивными ферментами.

27. Способ по любому из пп. 1-26, где нуклеиновую кислоту-мишень выделяют из одной клетки.

28. Способ по любому из пп. 1-27, где нуклеиновую кислоту-мишень выделяют из одной органеллы.

29. Способ по любому из пп. 1-28, где нуклеиновой кислотой-мишенью является геномная ДНК.

30. Способ по любому из пп. 1-29, где нуклеиновую кислоту-мишень подвергают

перекрестному связыванию с другими нуклеиновыми кислотами.

31. Способ по любому из пп. 1-30, где нуклеиновой кислотой-мишенью является неклеточная опухолевая ДНК.

5 32. Способ по п. 31, где неклеточную опухолевую ДНК выделяют из плацентарной жидкости.

33. Способ по п. 31, где неклеточную опухолевую ДНК выделяют из плазмы.

34. Способ по п. 33, где плазму выделяют из цельной крови с использованием мембранного сепаратора, имеющего зону сбора плазмы.

10 35. Способ по п. 34, где зона для сбора плазмы включает транспозомные комплексы, иммобилизованные на твердом носителе.

36. Способ по любому из пп. 1-28, где нуклеиновой кислотой-мишенью является кДНК.

37. Способ по любому из пп. 1-28, где нуклеиновую кислоту-мишень выделяют из залитого в парафин образца ткани, фиксированного формалином.

15 38. Способ по любому из пп. 1-28, где нуклеиновой кислотой-мишенью является ДНК, защищенная гистоном.

39. Способ по любому из пп. 1-38, где твердым носителем является сфера.

40. Способ по пп. 1-26, где множество твердых носителей представляет собой множество сфер и где множество сфер имеют различные размеры.

20

25

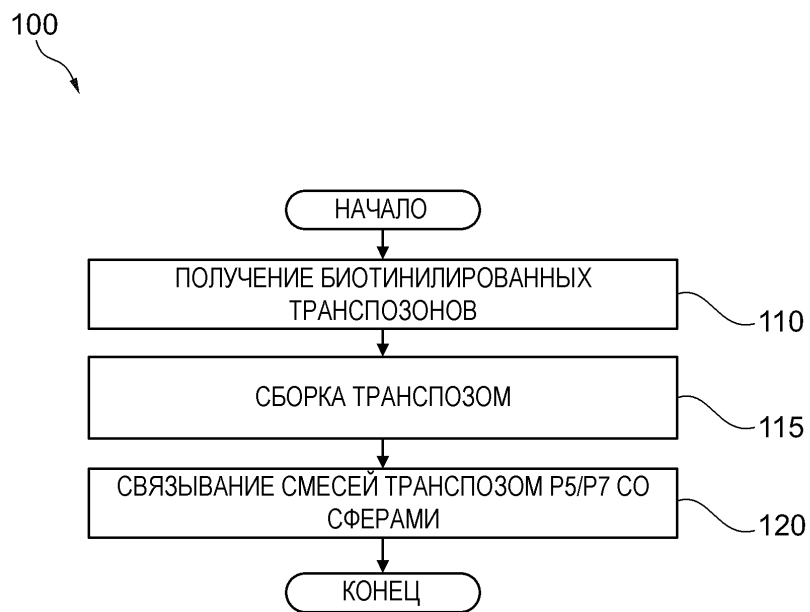
30

35

40

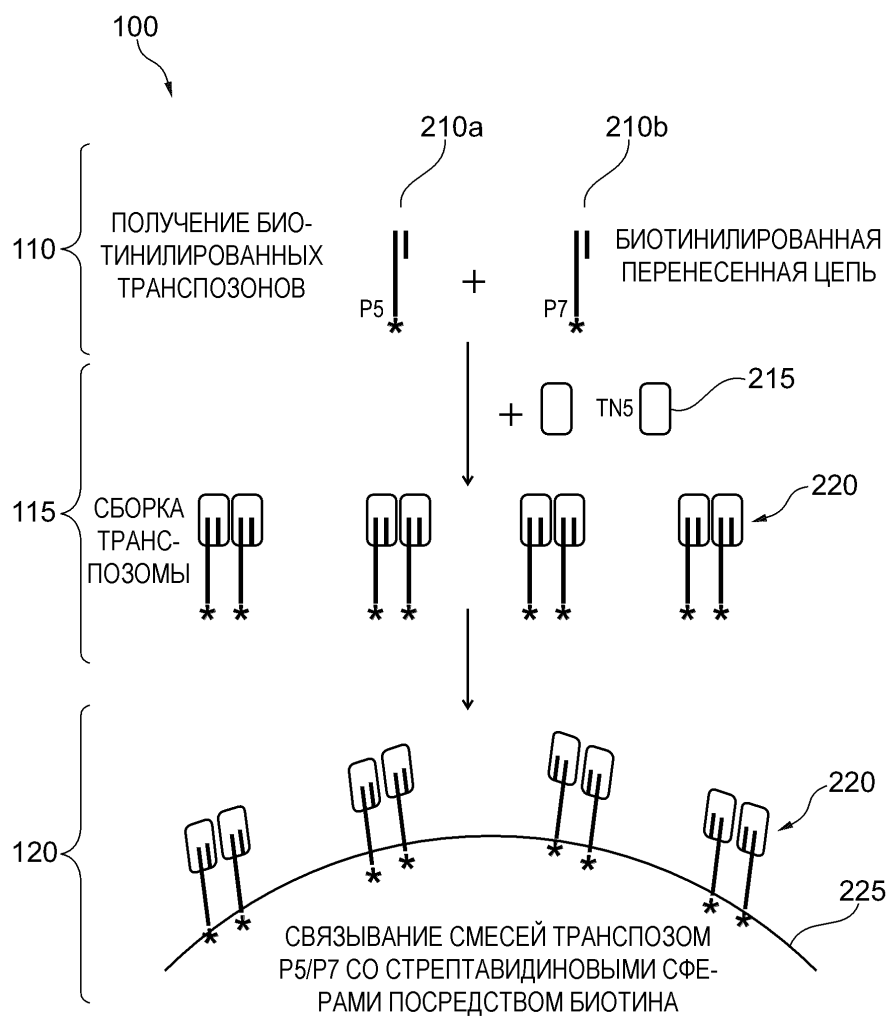
45

1/76



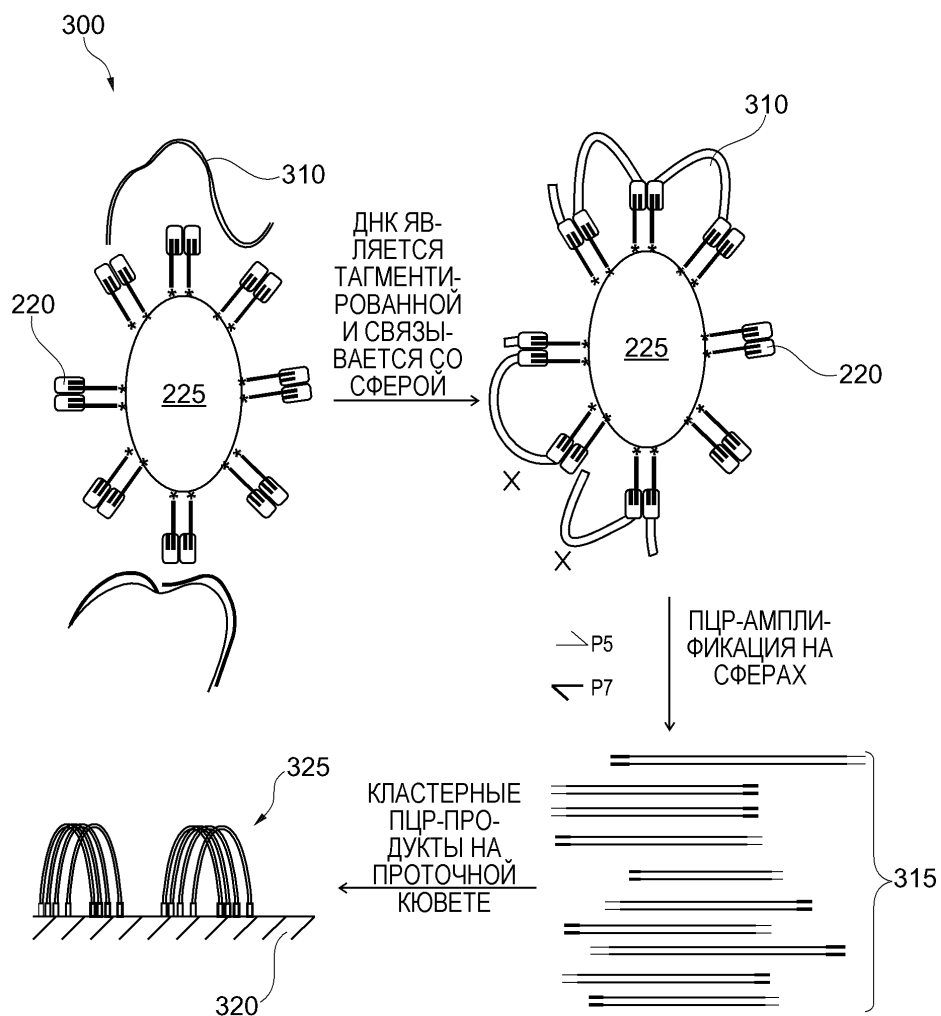
ФИГ. 1

2/76



ФИГ. 2

3/76



ФИГ. 3

4/76

400



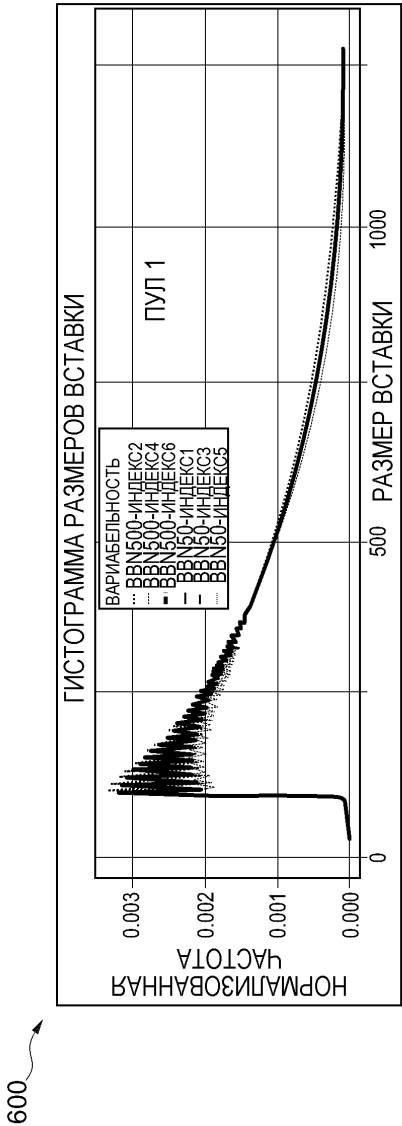
ОБРАЗЕЦ	ЧИСЛО КЛАСТЕРОВ С РАВНЫМ ОБЪЕМОМ ПЦР-ПРОДУКТА							
ОБРАЗЕЦ	15		19		21		24	
	К/мм ²	%	К/мм ²	%	К/мм ²	%	К/мм ²	%
50 нг	466	100	541	100	618	100	668	100
250 нг	504	108	598	110	703	114	740	111
1000 нг	512	109	602	111	670	108	716	107
*50 нг ПОЛНО-ГО ПОВТОРА	377	81	455	84	516	83	555	83

ФИГ. 4

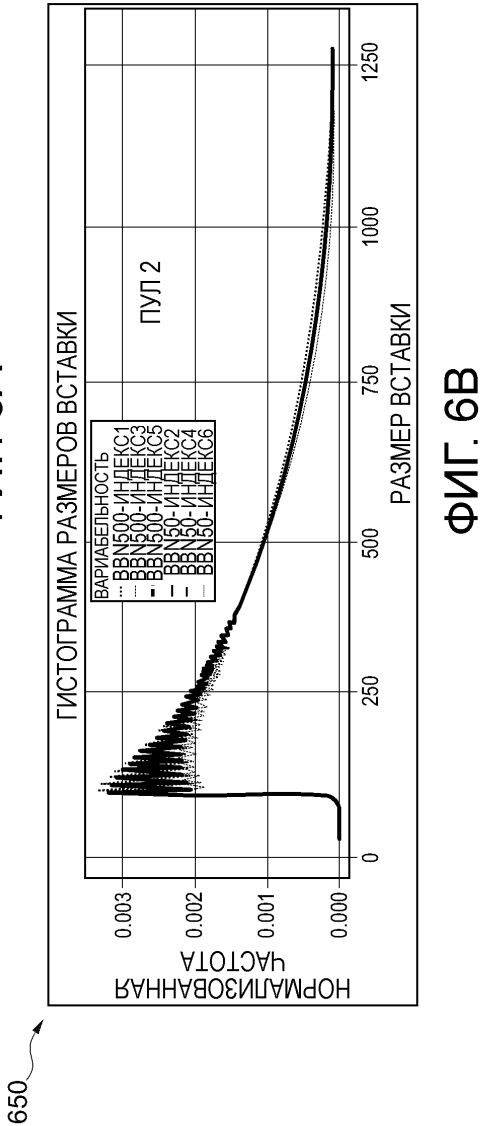
500 ↗

МЕДИАННЫЙ РАЗМЕР ВСТАВКИ		СРЕДНИЙ РАЗМЕР ВСТАВКИ
BBN50-ИНДЕКС1	341	409.78
BBN500-ИНДЕКС1	344	415.89
BBN50-ИНДЕКС2	337	405.97
BBN500-ИНДЕКС2	346	419.91
BBN50-ИНДЕКС3	330	400.35
BBN500-ИНДЕКС3	327	398.65
BBN50-ИНДЕКС4	321	388.71
BBN500-ИНДЕКС4	328	400.75
BBN50-ИНДЕКС5	327	397.62
BBN500-ИНДЕКС5	328	399.88
BBN50-ИНДЕКС6	324	389.35
BBN500-ИНДЕКС6	331	401.50

ФИГ. 5

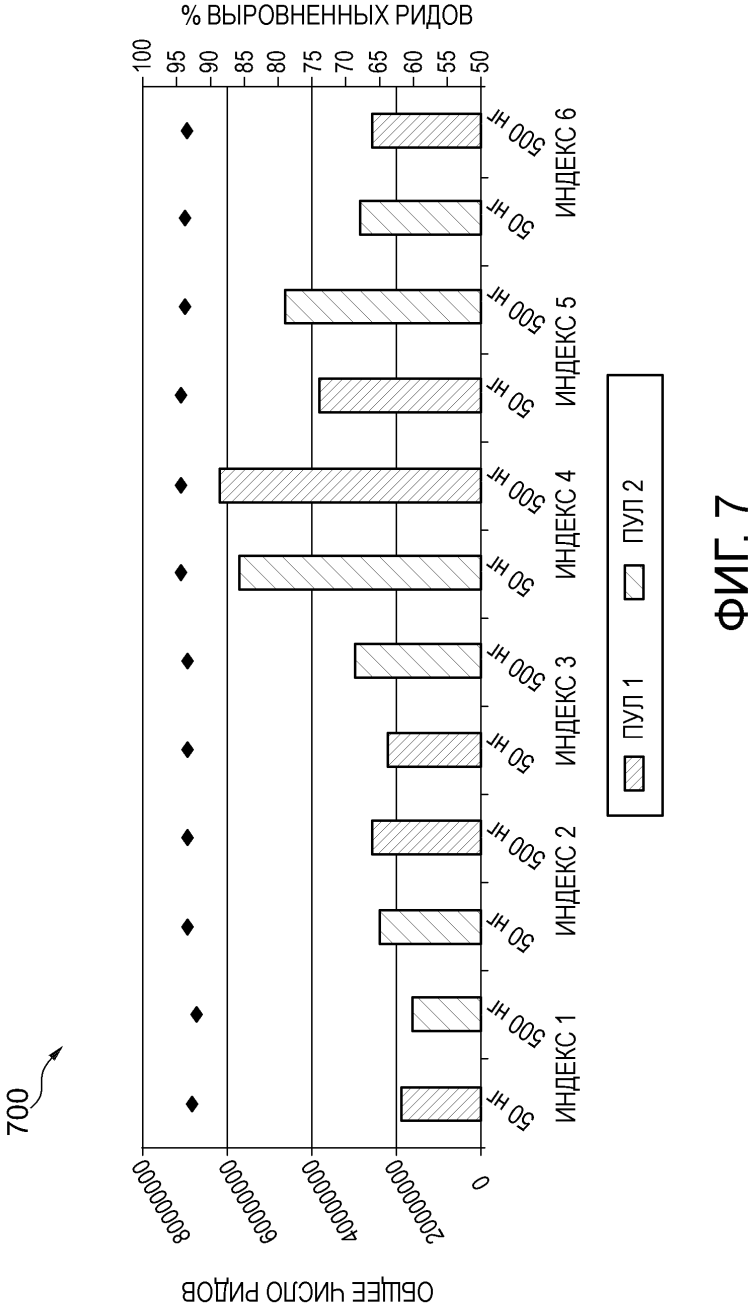


ФИГ. 6А

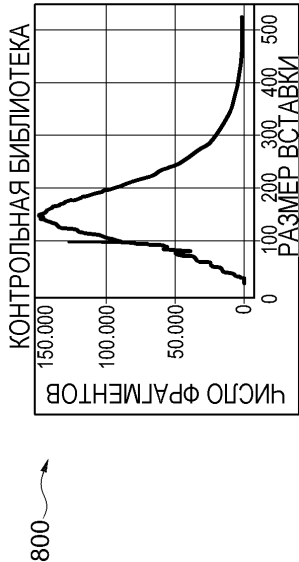


ФИГ. 6В

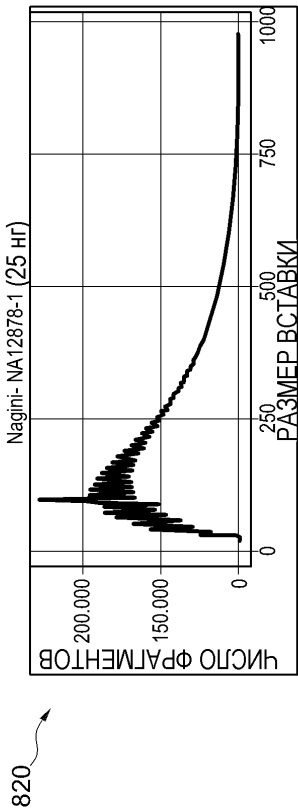
7/76



ФИГ. 7



ФИГ. 8А



ФИГ. 8В

840

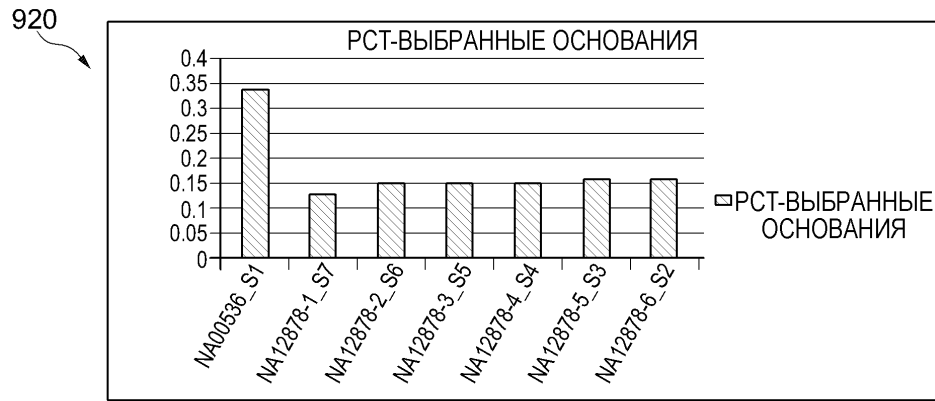
	КОНТРОЛЬ	25 нг	50 нг	100 нг	150 нг	200 нг	500 нг
СРЕДНИЙ	196.8	267.4	287.3	288.8	288.6	295.0	282.7
МЕДИАННЫЙ	136	206	222	224	223	230	220
МОДА РАСПРЕДЕЛЕНИЯ	149	98	98	98	98	98	98

ФИГ. 8С

9/76



ФИГ. 9А

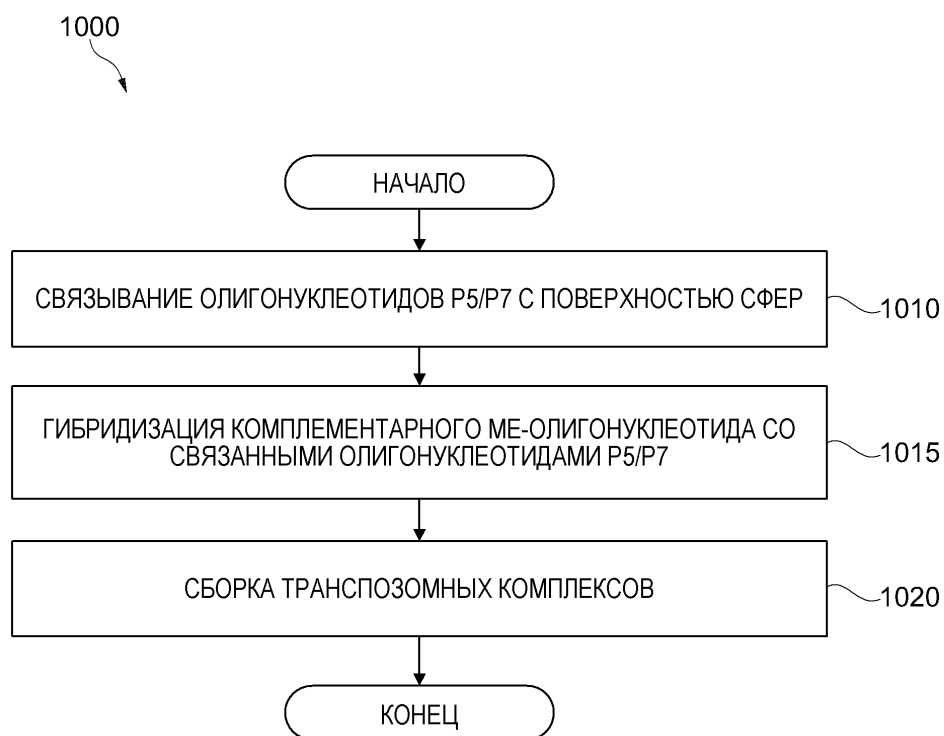


ФИГ. 9В



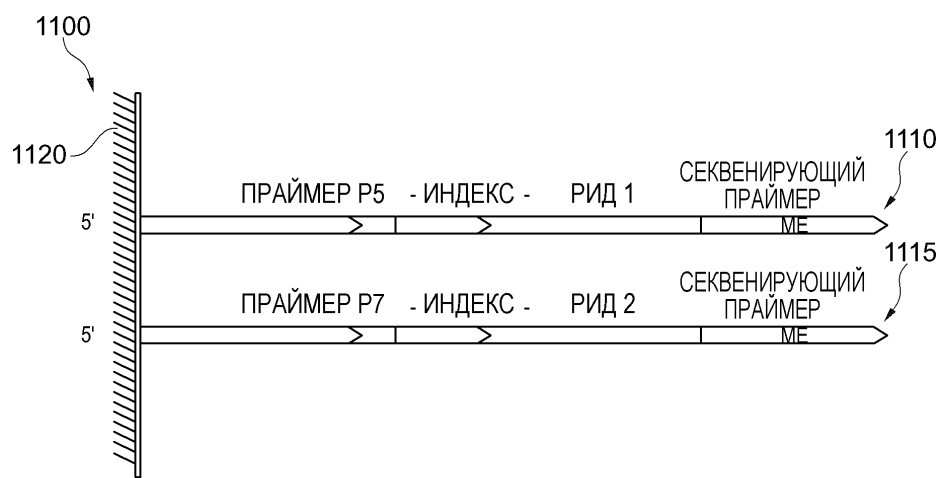
ФИГ. 9С

10/76



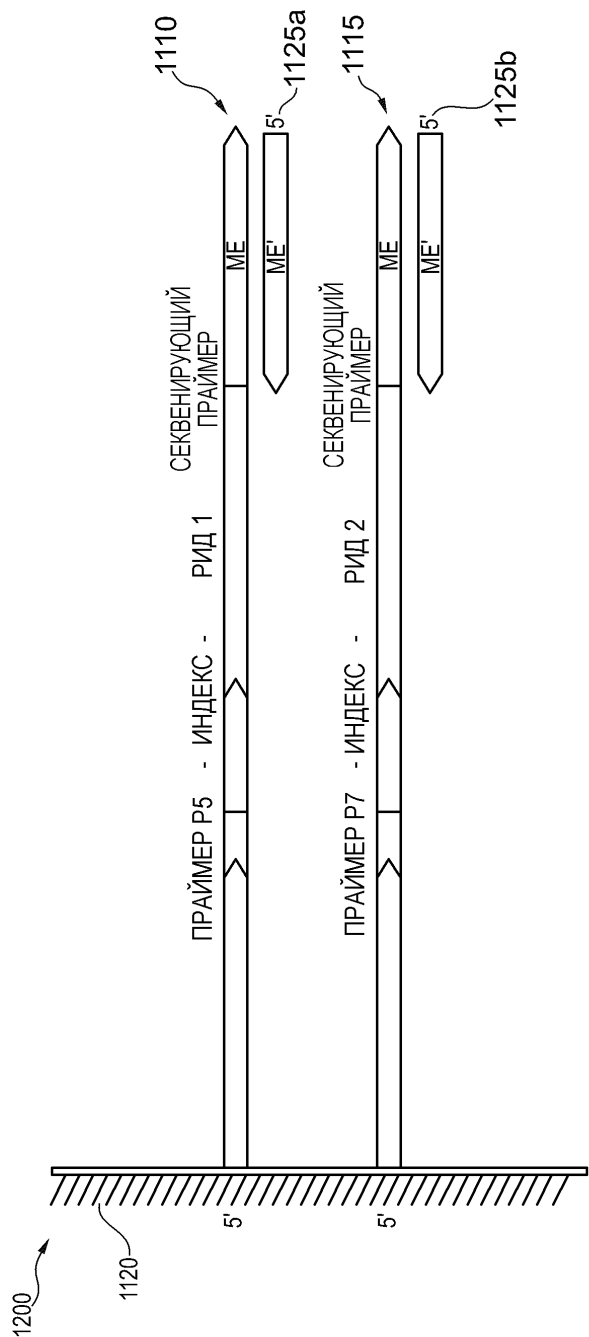
ФИГ. 10

11/76



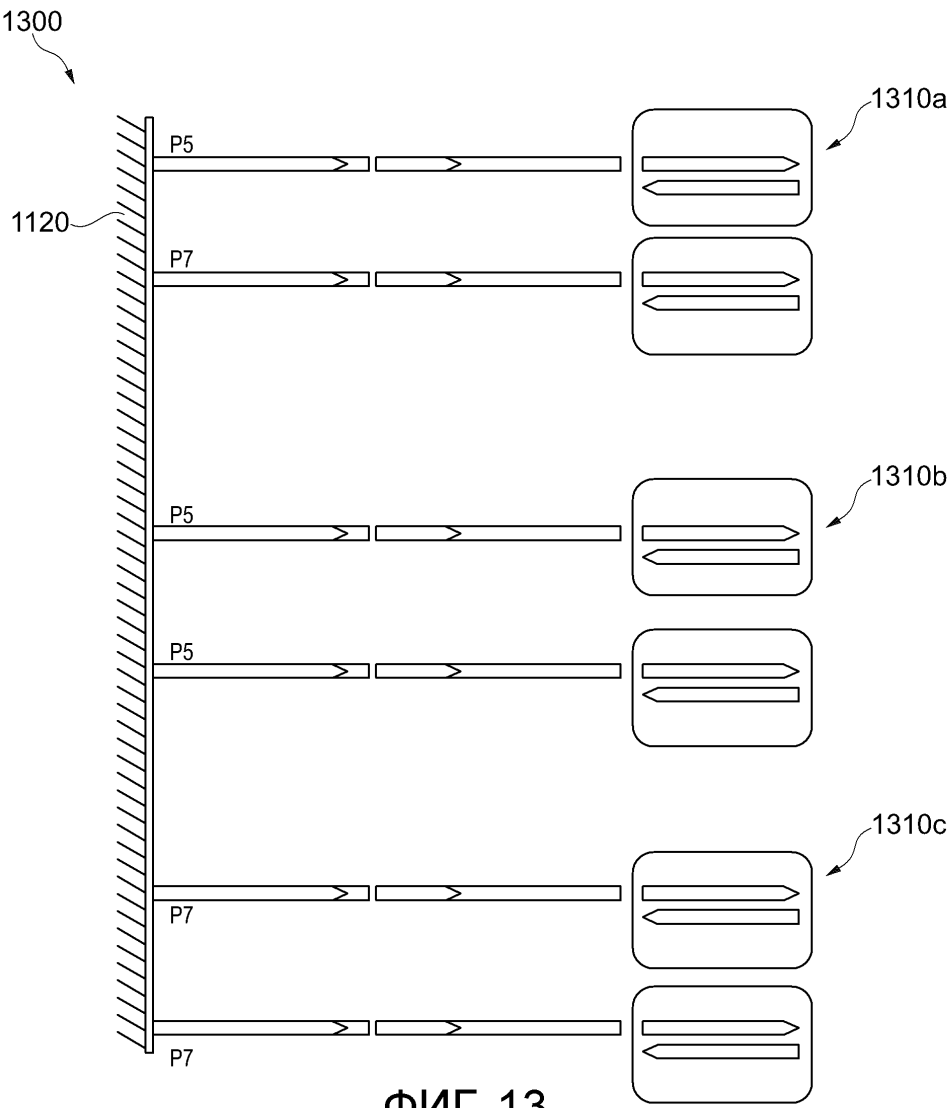
ФИГ. 11

12/76

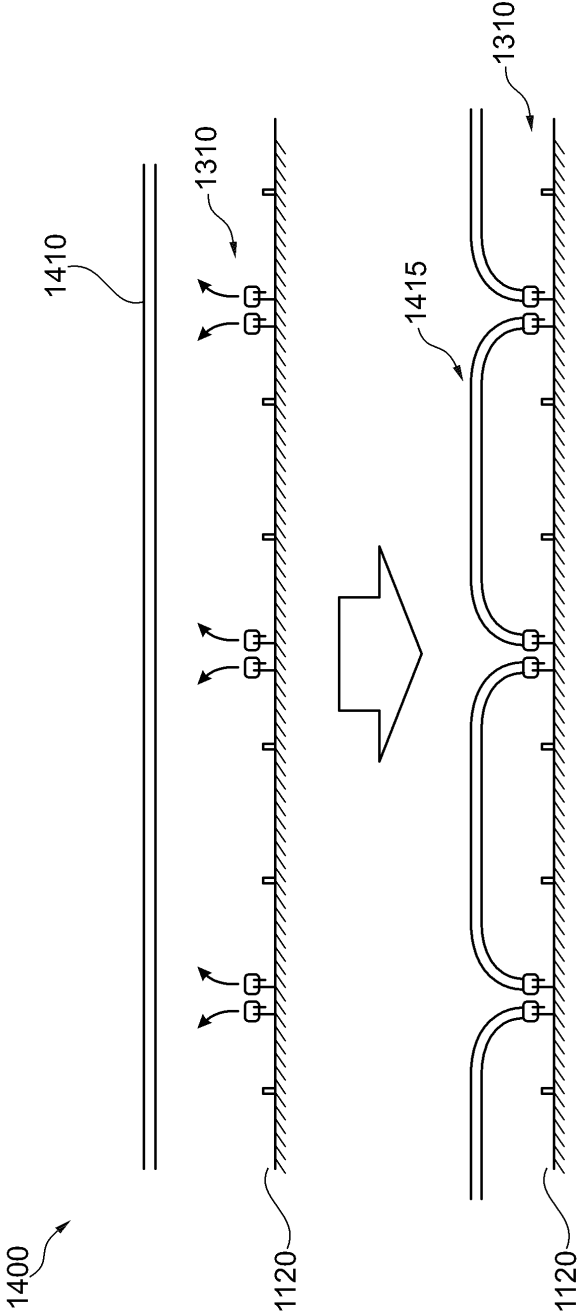


ФИГ. 12

13/76

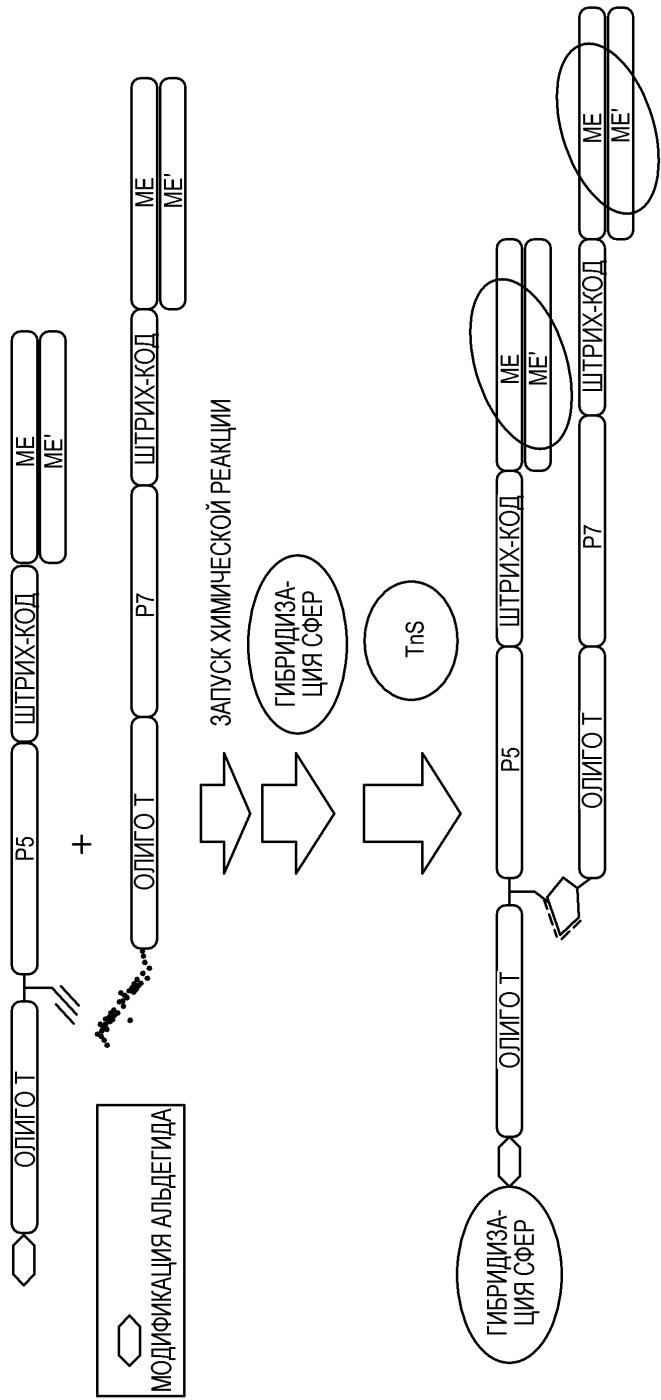


14/76

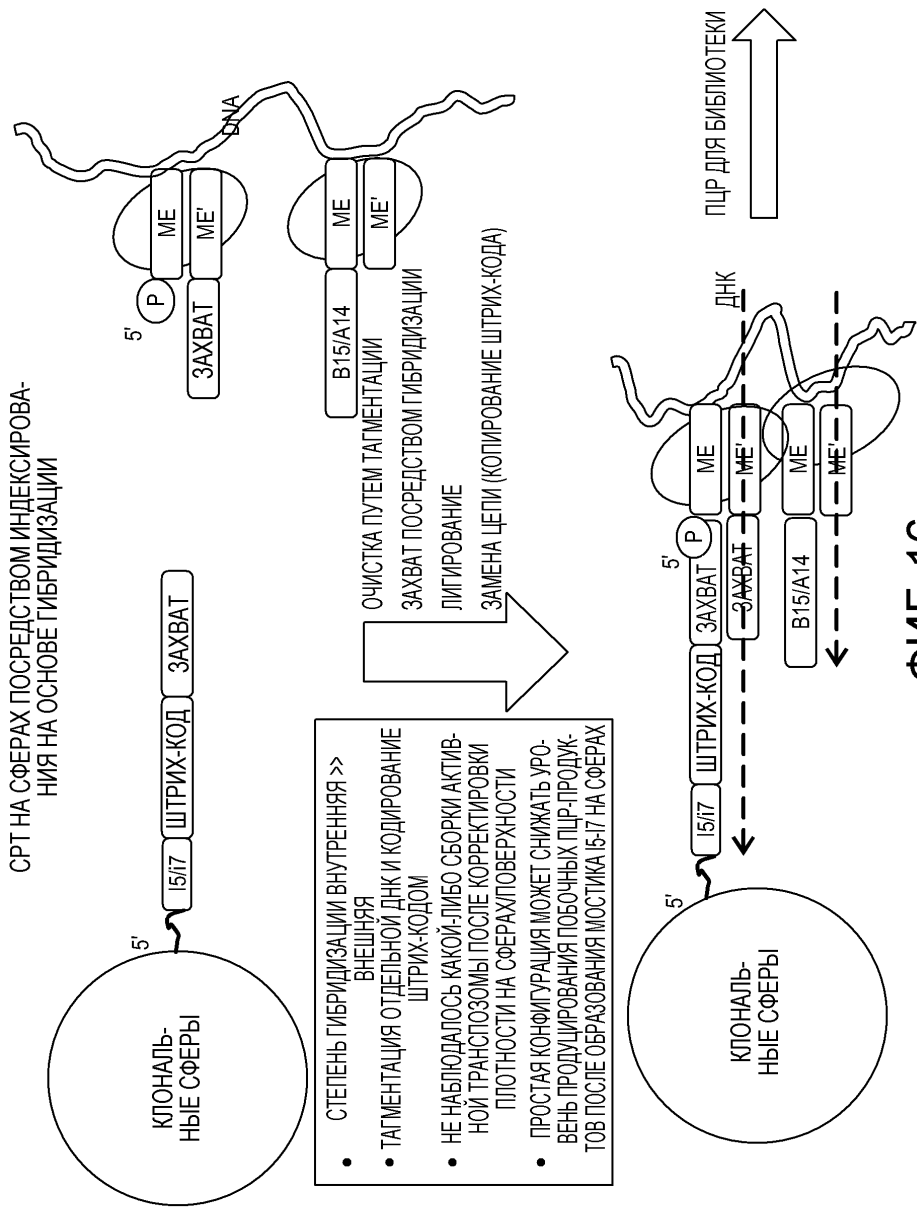


ФИГ. 14

15/76



ФИГ. 15

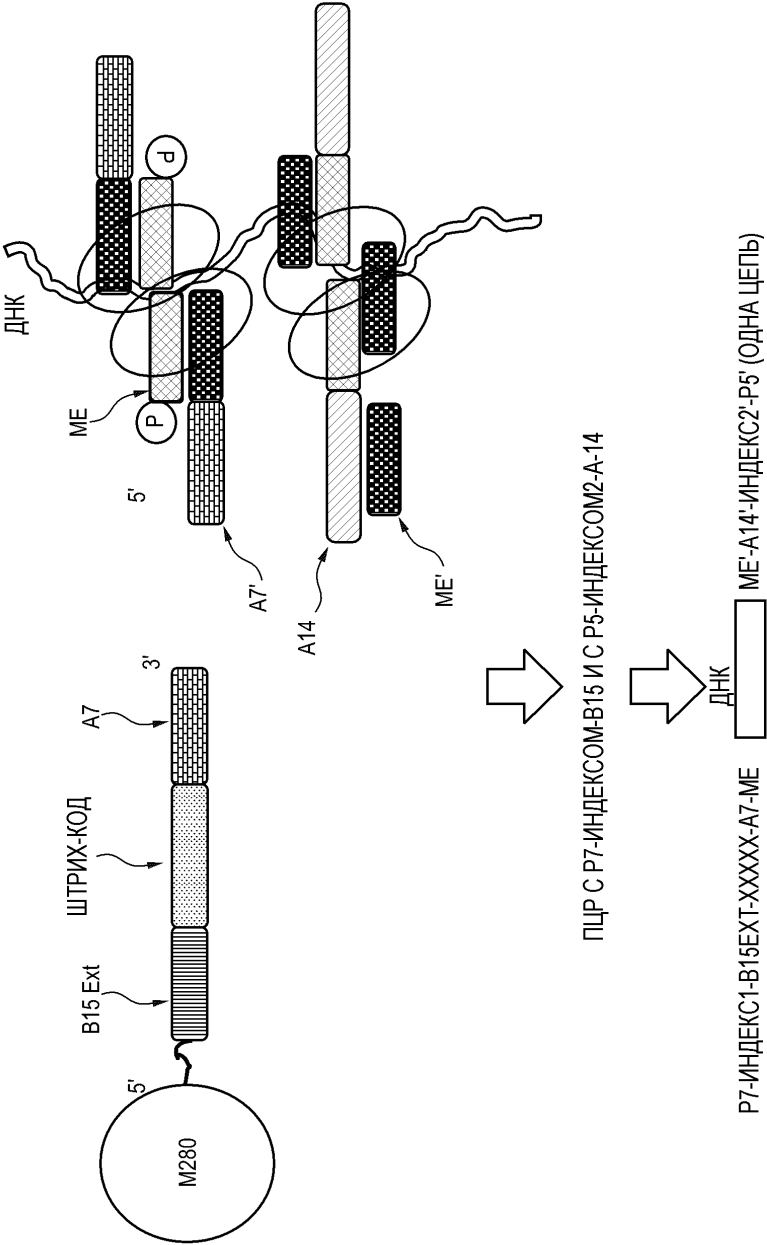


ФИГ. 16

- СТЕПЕНЬ ГИБРИДИЗАЦИИ ВНУТРЕННЯЯ >> ВНЕШНЯЯ
- ТАГМЕНТАЦИЯ ОТДЕЛЬНОЙ ДНК И КОДИРОВАНИЕ ШТРИХ-КОДОМ
- НЕ НАБЛЮДАЕТСЯ КАКОЙ-ЛИБО СБОРКИ АКТИВНОЙ ТРАНСПОЗОМЫ ПОСЛЕ КОРРЕКТИРОВКИ ПЛОТНОСТИ НА СФЕРАХ/ПОВЕРХНОСТИ
- ПРОСТАЯ КОНФИГУРАЦИЯ МОЖЕТ СНИЖАТЬ УРОВЕНЬ ПРОДУЦИРОВАНИЯ ПОБОЧНЫХ ПЦР-ПРОДУКТОВ ПОСЛЕ ОБРАЗОВАНИЯ МОСТИКА 15-17 НА СФЕРАХ

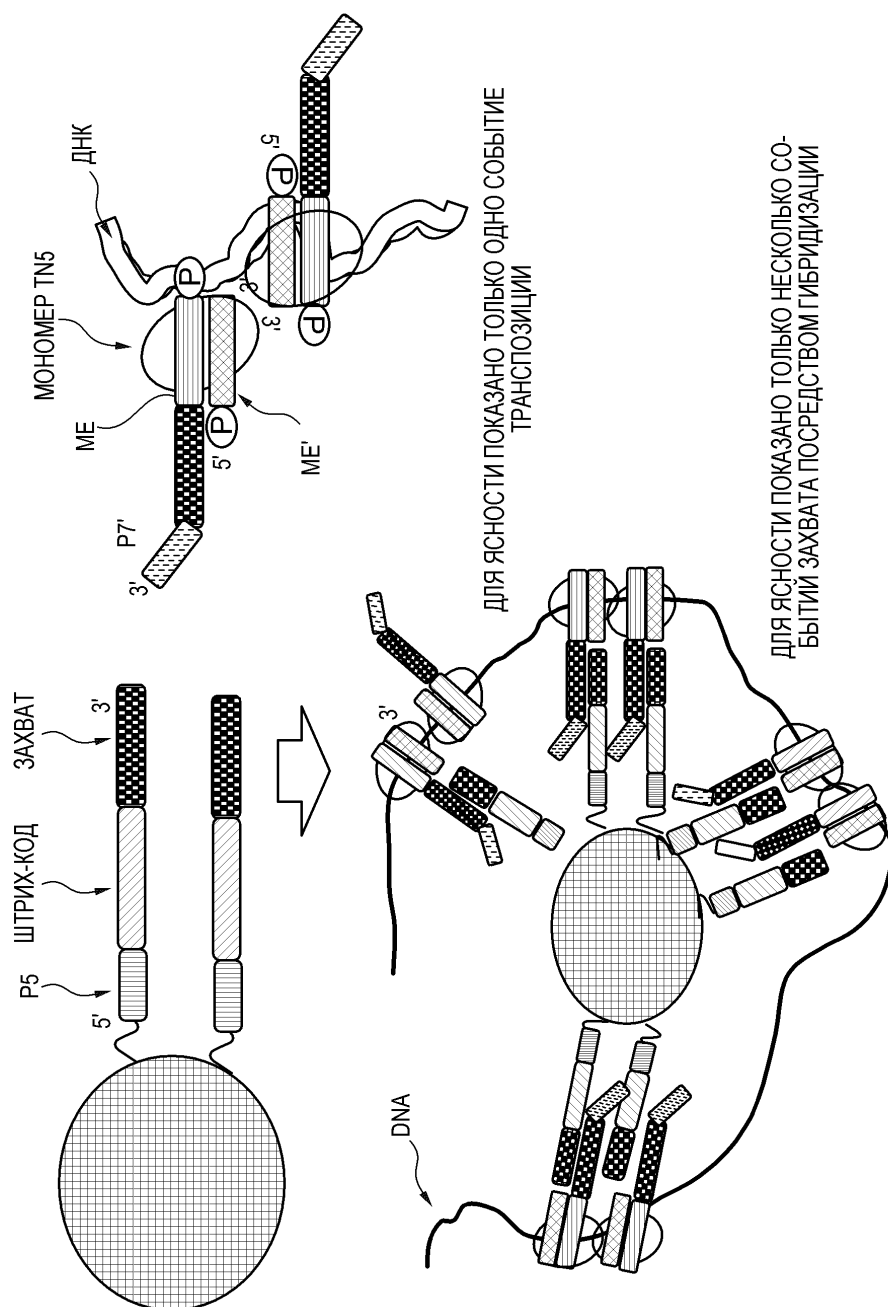
17/76

ТРАНСПОЗИЦИЯ С СОХРАНЕНИЕМ СЦЕПЛЕНИЯ ГЕНОВ (СРТ) НА СФЕРАХ



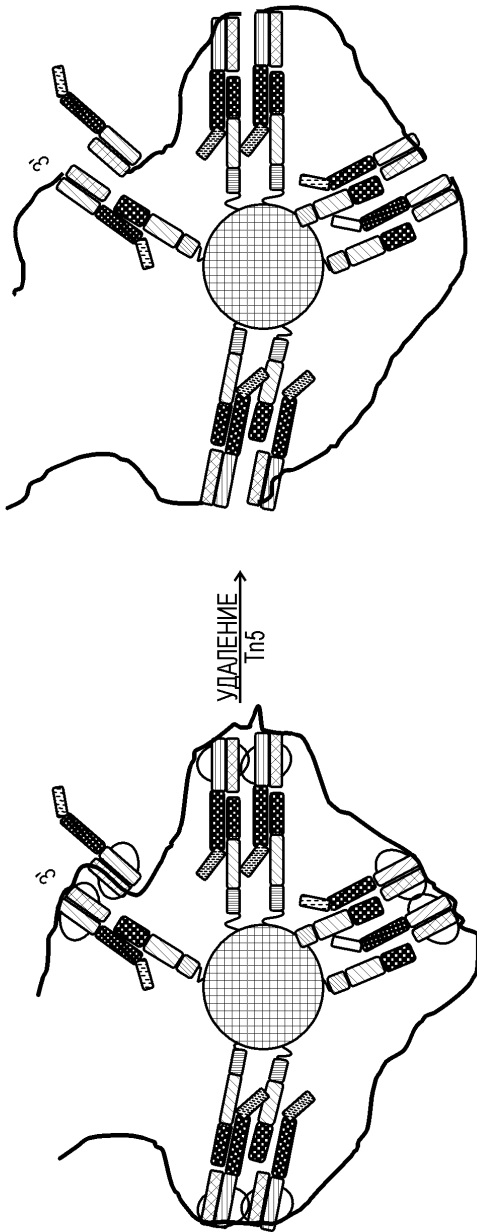
ФИГ. 17

18/76



ΦΙΓ. 18

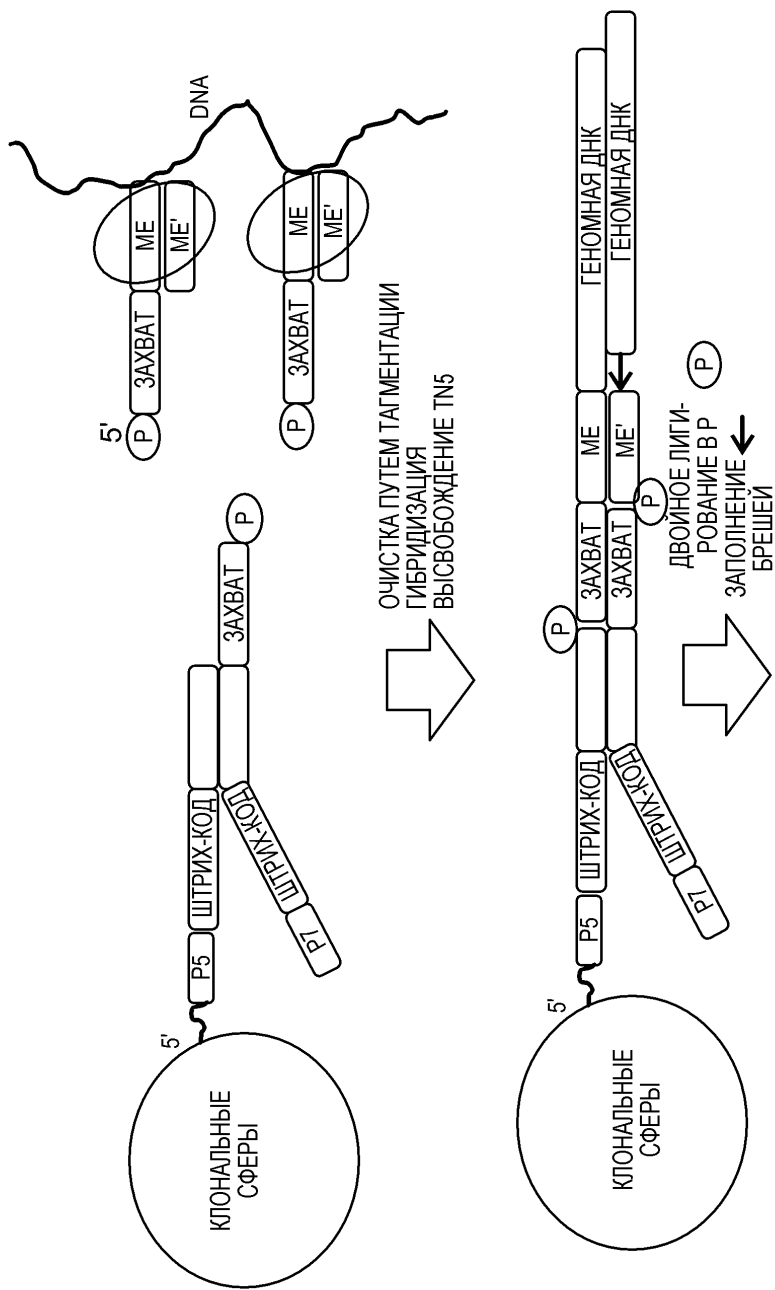
19/76



ПОСЛЕ ЗАХВАТА ПОСРЕДСТВОМ ГИБРИДИЗАЦИИ, СЦЕПЛЕНИЕ ГЕНОВ СОХРА-
НЯЮТСЯ ПОСЛЕ УДАЛЕНИЯ Tn5. ДНК ГИБРИДИЗОВАЛИ СО СФЕРОЙ С ИСПОЛЬ-
ЗОВАНИЕМ ИНДЕКСИРОВАННЫХ ПРАЙМЕРОВ, ЧТО СПОСОБСТВОВАЛО СО-
ХРАНЕНИЮ ВСЕХ ОТДЕЛЬНЫХ БИБЛИОТЕК НА СФЕРЕ

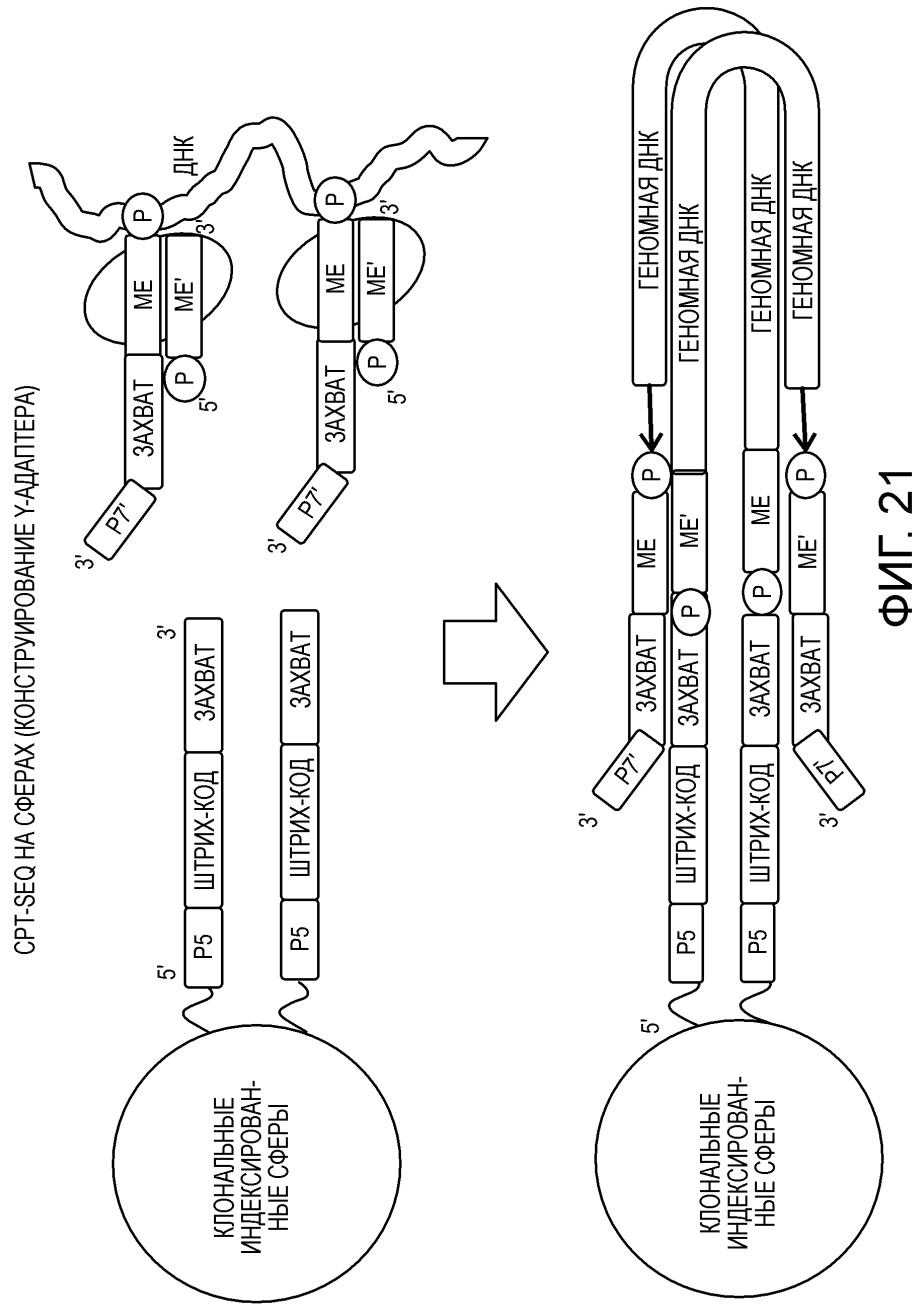
ФИГ. 19

20/76

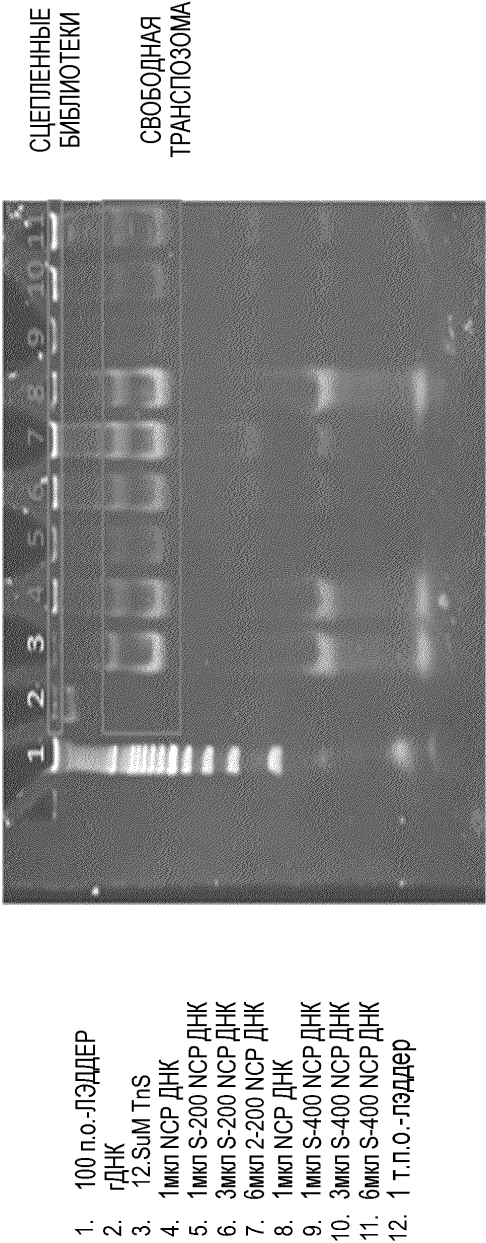


ФИГ. 20

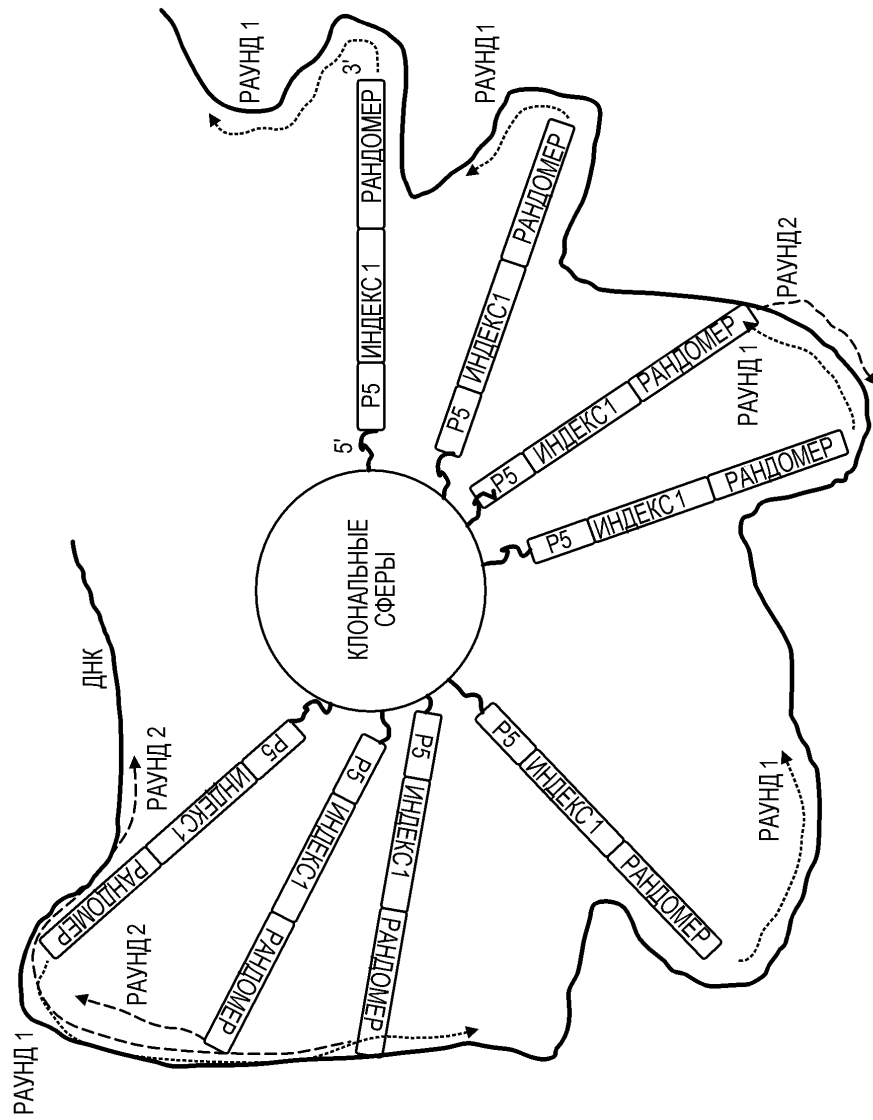
21/76



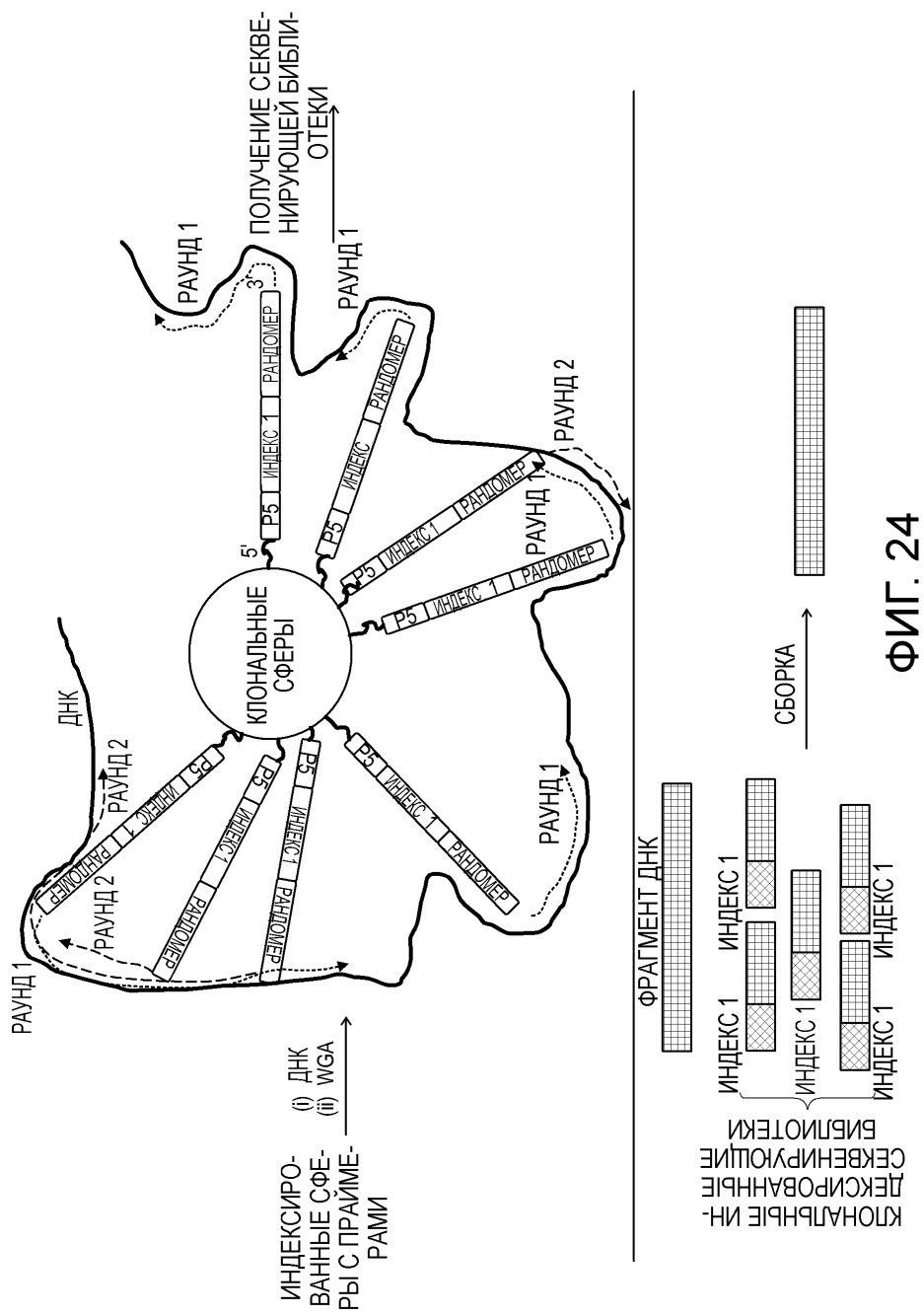
22/76



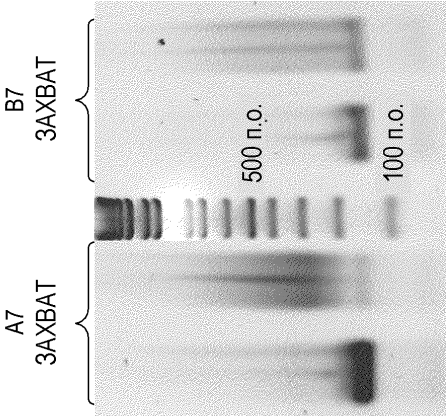
ФИГ. 22



ФИГ. 23



ОПТИМИЗАЦИЯ ПЛОТНОСТИ ЗАХВАТА НА СФЕРАХ

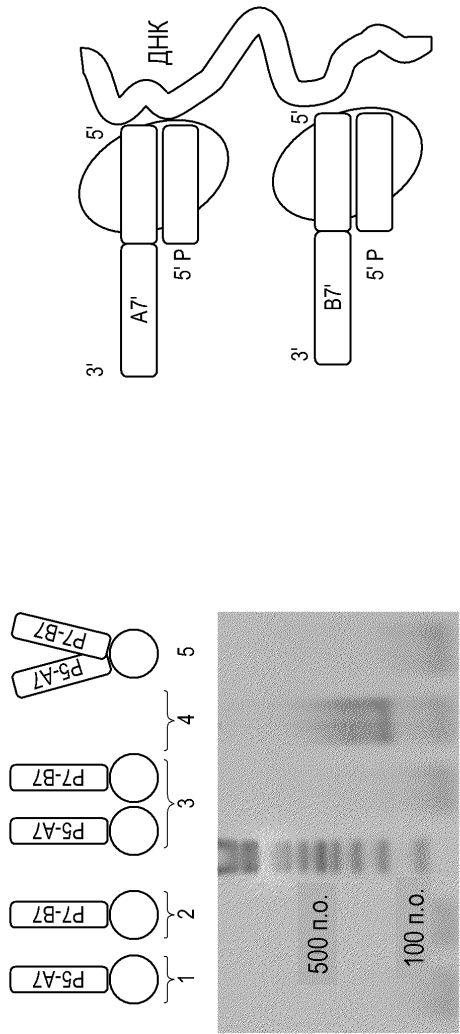


	ПЛОТНОСТЬ (РАЗВЕД- НИЕ)ЗОНДОВ	ЛИГИРОВАНИЕ
1	1x	НЕТ
2	10x	ДА
3	1x	НЕТ
4	10x	ДА

ФИГ. 25

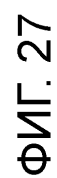
26/76

ВОЗМОЖНОСТЬ СРТ-СЕКВЕНИРОВАНИЯ НА СФЕРАХ: ВНУТРИМОЛЕКУЛЯРНЫЕ БИБЛИОТЕКИ

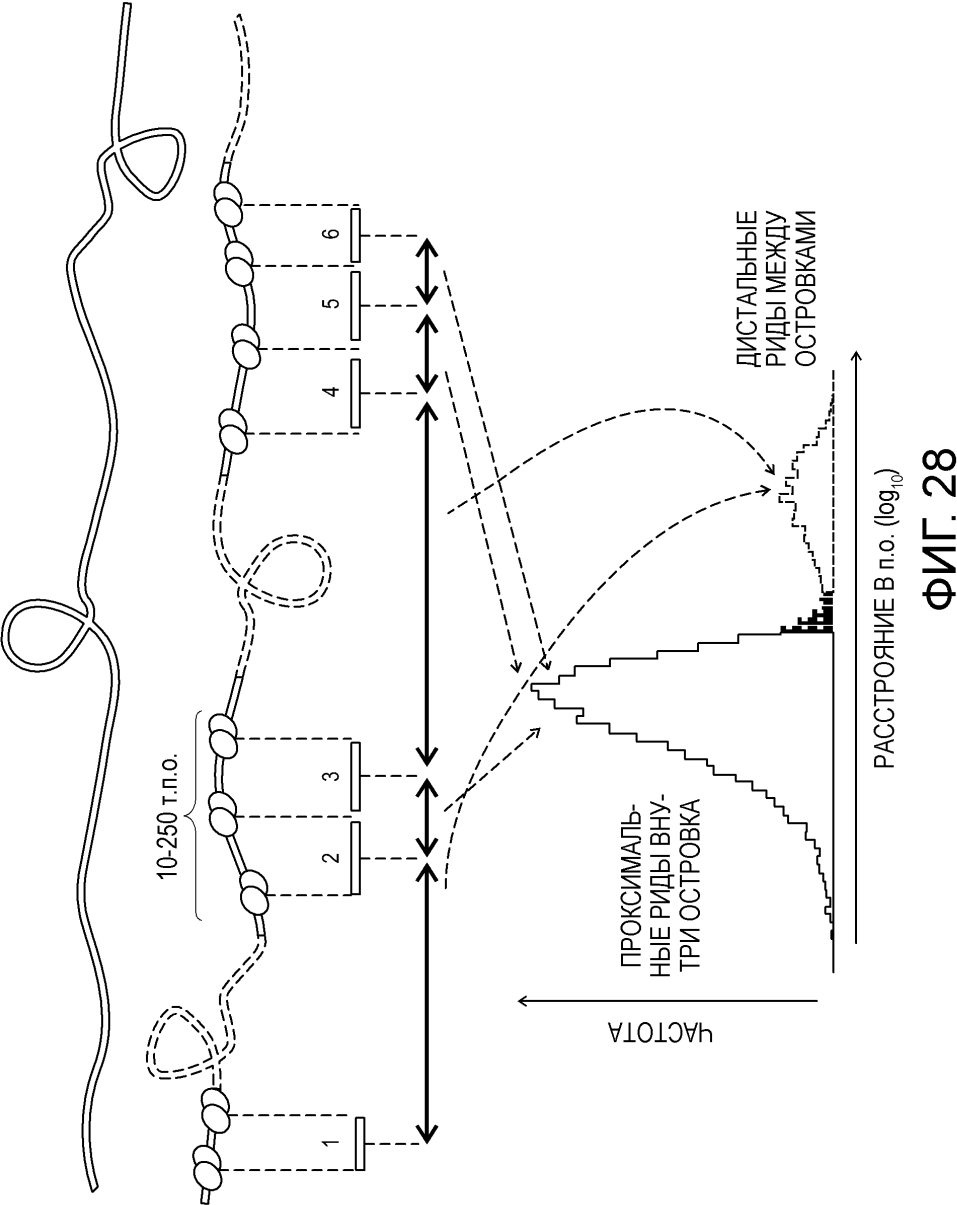


ФИГ. 26

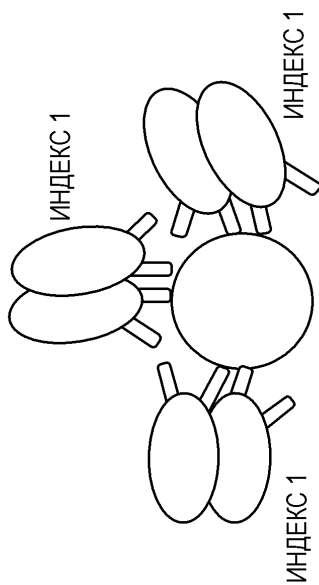
СФЕРЫ	ДОБАВЛЕНИЕ ЛИГАЗЫ
1 A7	ДА
2 B7	ДА
3 СФЕРЫ A7 & B7	ДА
4 A7+B7	ДА
5 A7+B7	НЕТ



28/76



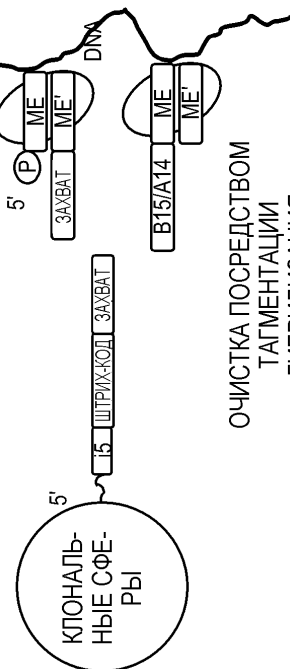
КЛОНАЛЬНЫЕ ИНДЕКСИ- РОВАННЫЕ ТРАНСПОЗИ- МЫ НА СФЕРАХ



ДНК (ОРАНЖЕВЫМ) ПЕРЕНОСИЛИ С ИСПОЛЬЗОВАНИЕМ ИНДЕКСИРОВАННЫХ ТРАНСПОЗОМ НА СФЕРАХ

ФИГ. 29А

КЛОНАЛЬНЫЕ ИНДЕКСИ-
РОВАННЫЕ ЗОНДЫ ДЛЯ
ЗАХВАТА НА СФЕРАХ И
СРТ-ДНК



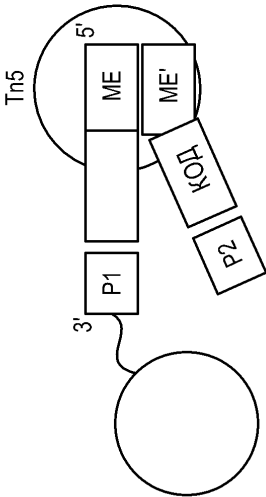
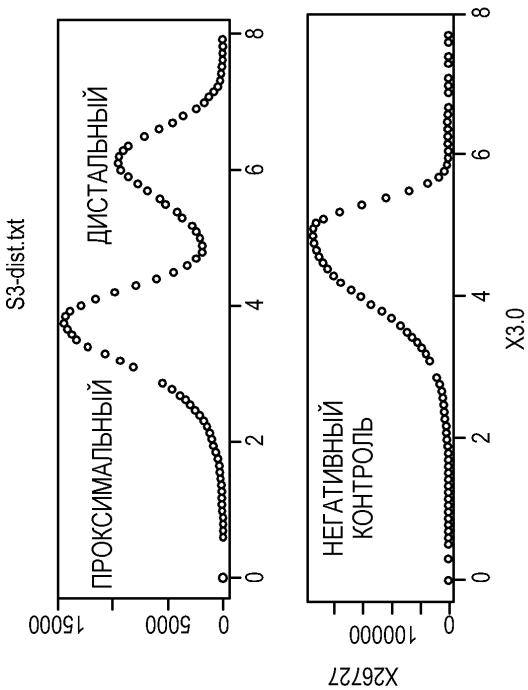
ОЧИСТКА ПОСРЕДСТВОМ
ТАГМЕНТАЦИИ
ГИБРИДИЗАЦИЯ

ПЕРЕНЕСЕННУЮ ДНК ЗАХВАТЫВАЛИ
НА КЛОНАЛЬНЫХ ИНДЕКСИРОВАННЫХ
СФЕРАХ: ИНДЕКС ПЕРЕНОСИЛИ НА
СРТ-ДНК

Фиг. 29В

30/76

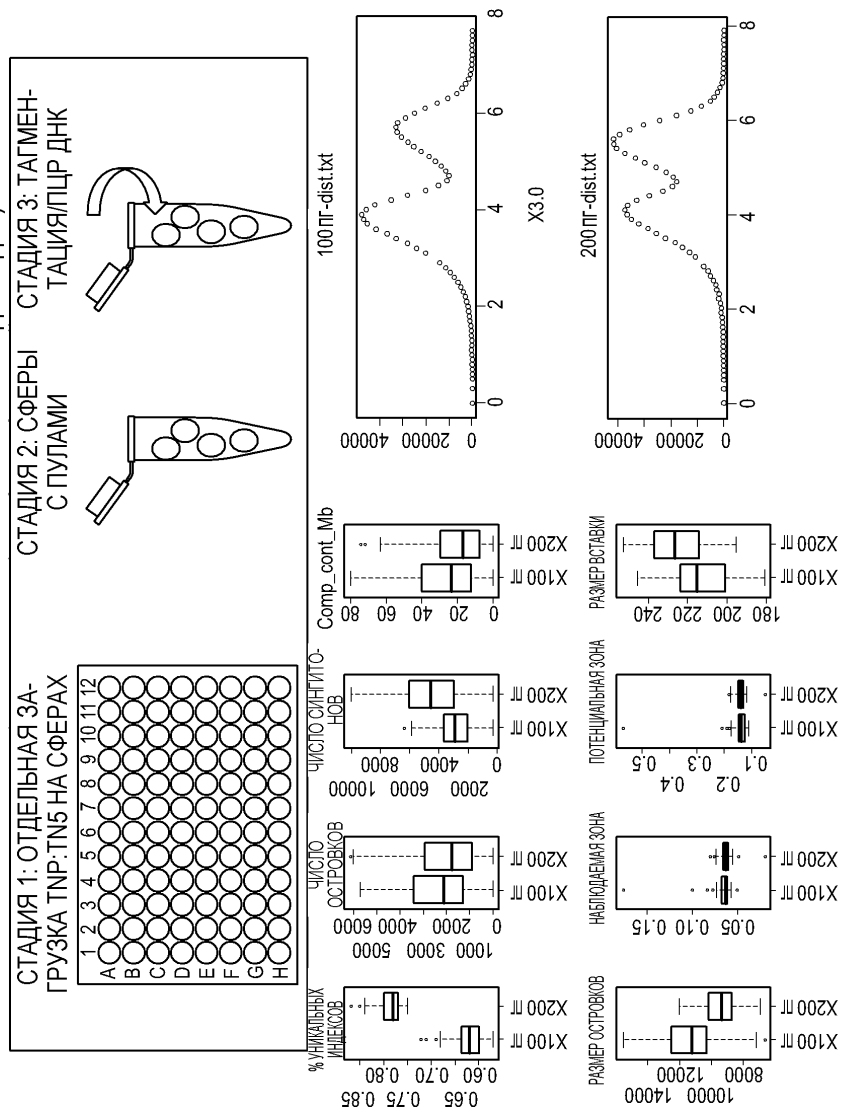
ТРАСПОЗИЦИЮ 96 ИНДЕКСИРОВАННЫХ КЛОНАЛЬ-
НЫХ СФЕР ОСУЩЕСТВЛЯЛИ В ОДНОМ СОСУДЕ



(ДЛЯ ЯСНОСТИ ПОКАЗАН ТОЛЬКО МОНОМЕР)

ФИГ. 30

МЕТОД, ПРОВДИМЫЙ НА СФЕРАХ В 96-ЛУЧНОМ ПЛАНШЕТЕ (ТРАНСПОЗОМЫ НА СФЕРАХ)(ПЛОТНОСТЬ ПОВЕРХНОСТИ 200 НМ/100 ПГ ИЛИ 200 ПГ ИСХОДНОЙ ДНК)

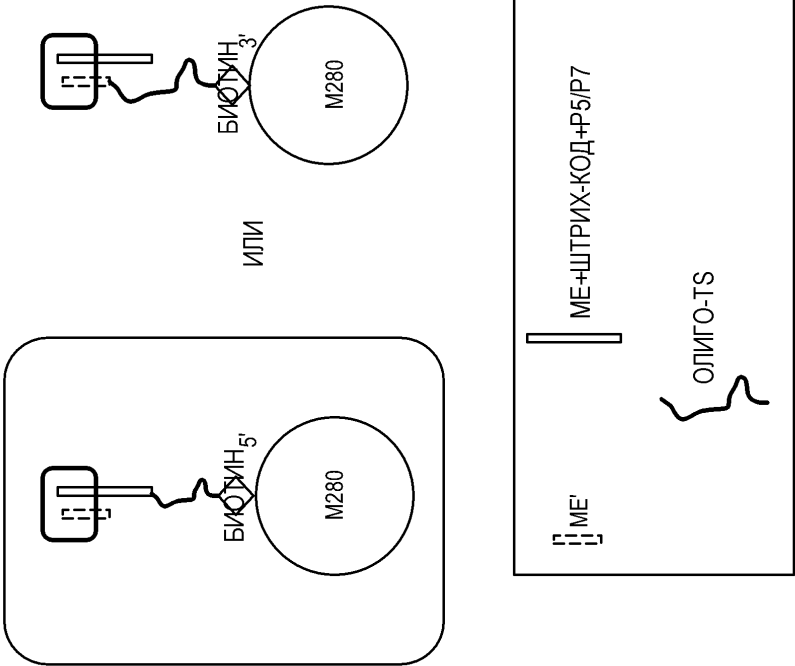


ΦΙΓ. 31

32/76

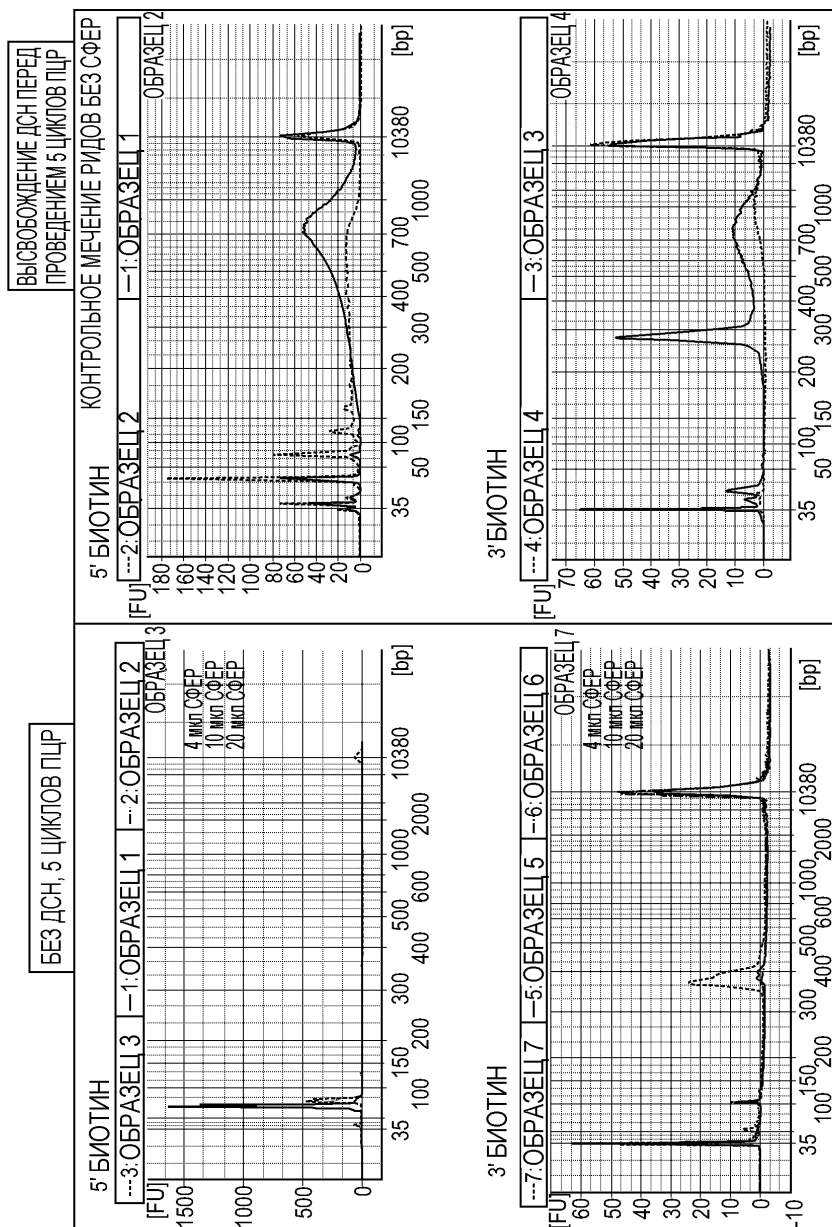
- СНАЧАЛА СБОРКА КОМПЛЕКСА ТНР:ТN5
- ЗАГРУЗКА НА СФЕРЫ
- ПРОМЫВКА СФЕР
- СМЕШИВАНИЕ СФЕР
- ДОБАВЛЕНИЕ ДНК И ТАГМЕНТАЦИЯ
- ПЦР

5'-БИОТИН-МЕ-ОЛИГОНУКЛЕОТИДЫ И 3'-БИОТИН-МЕ'-ОЛИГОНУКЛЕОТИДЫ НА СФЕРАХ



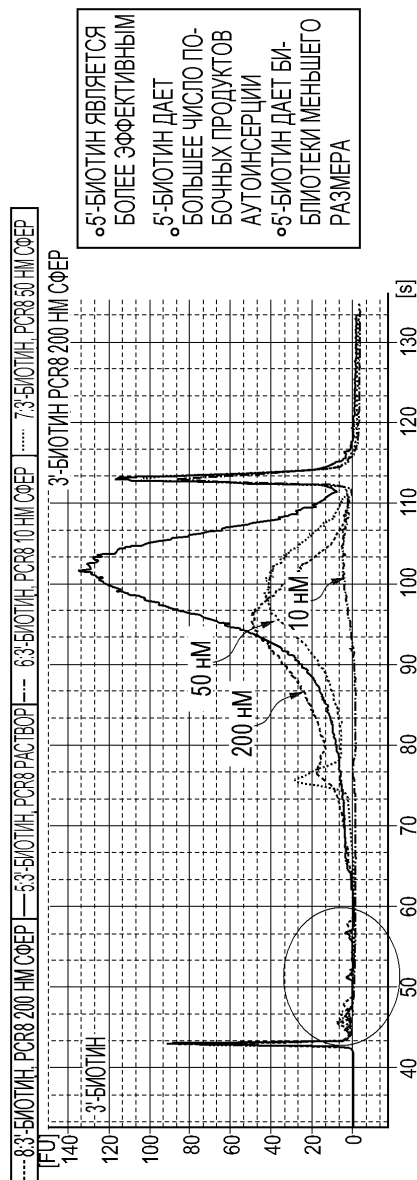
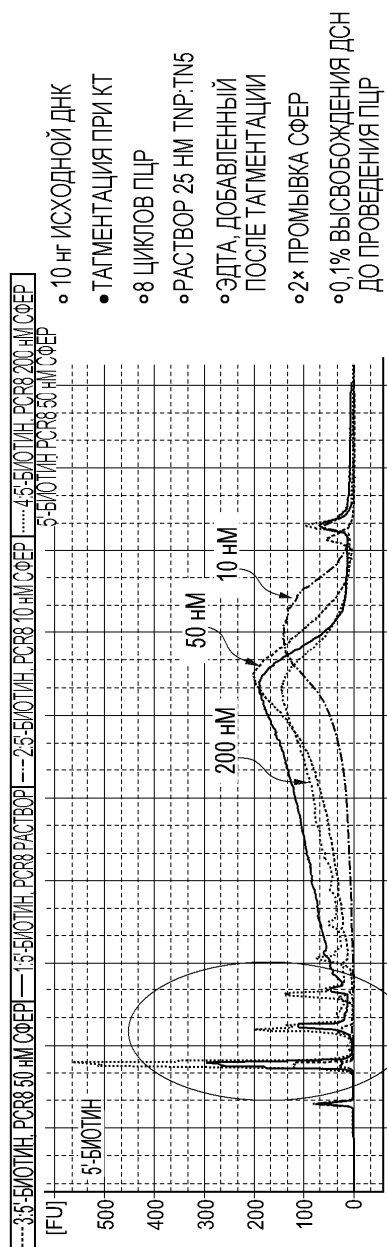
ФИГ. 32

РАЗМЕРЫ БИБЛИОТЕК ДЛЯ ТРАНСПОЗОМ НА СФЕРАХ

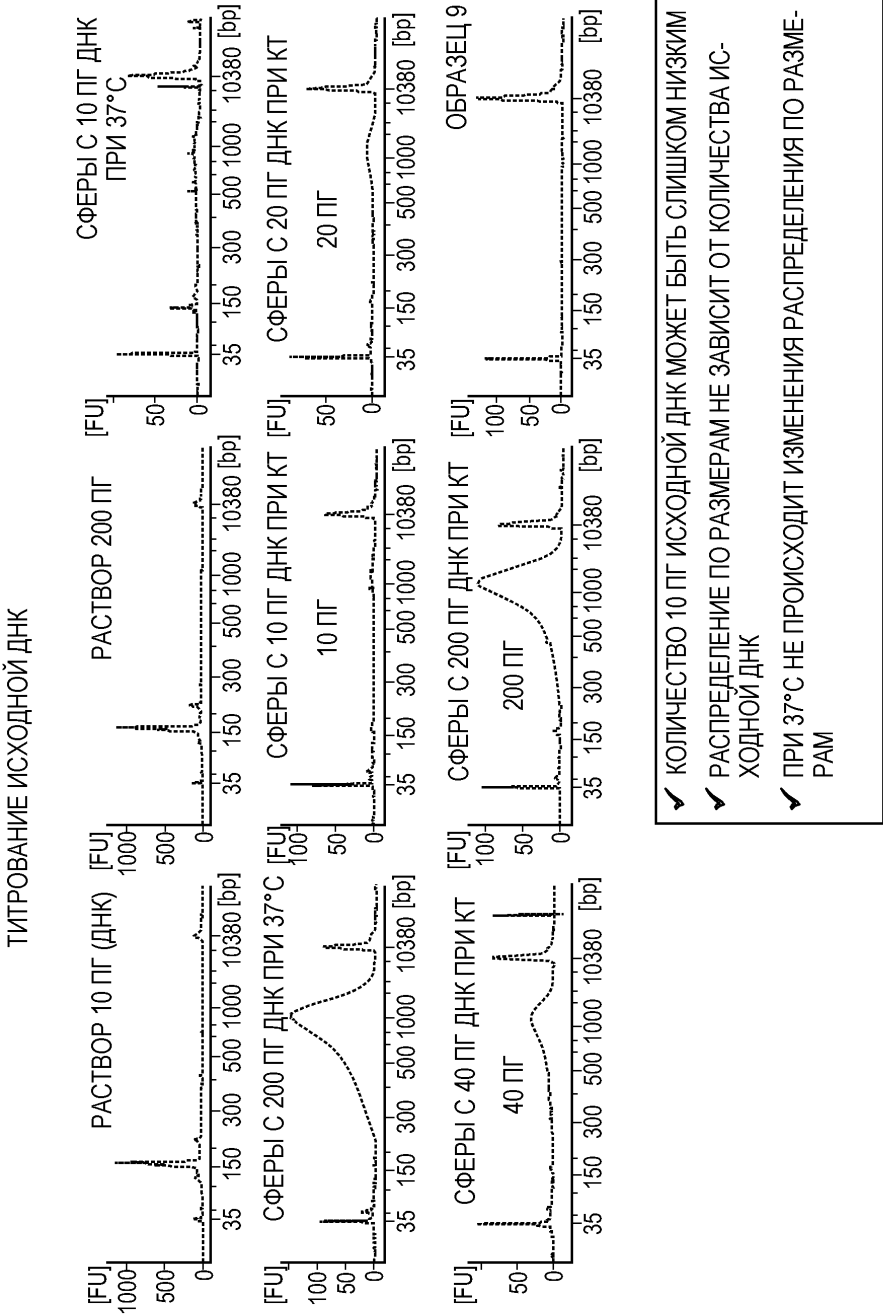


Φιλ. 33

ПЛОТНОСТЬ ПОВЕРХНОСТИ ТРАНСПОЗОМЫ В ЗАВИСИМОСТИ ОТ РАЗМЕРА ВСТАВКИ



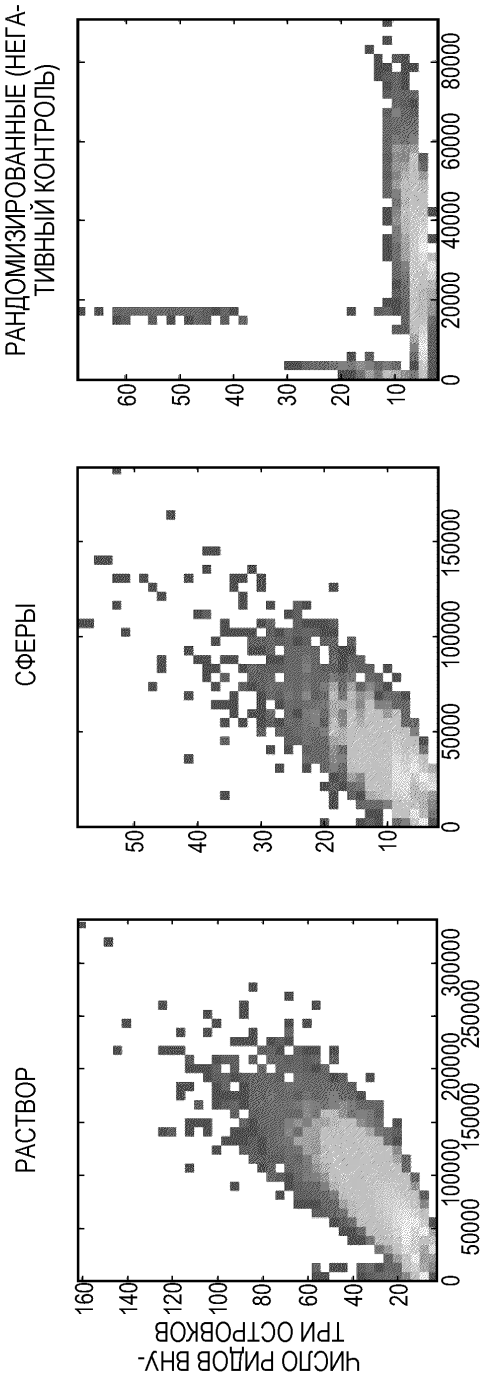
ФИГ. 34



ФИГ. 35

36/76

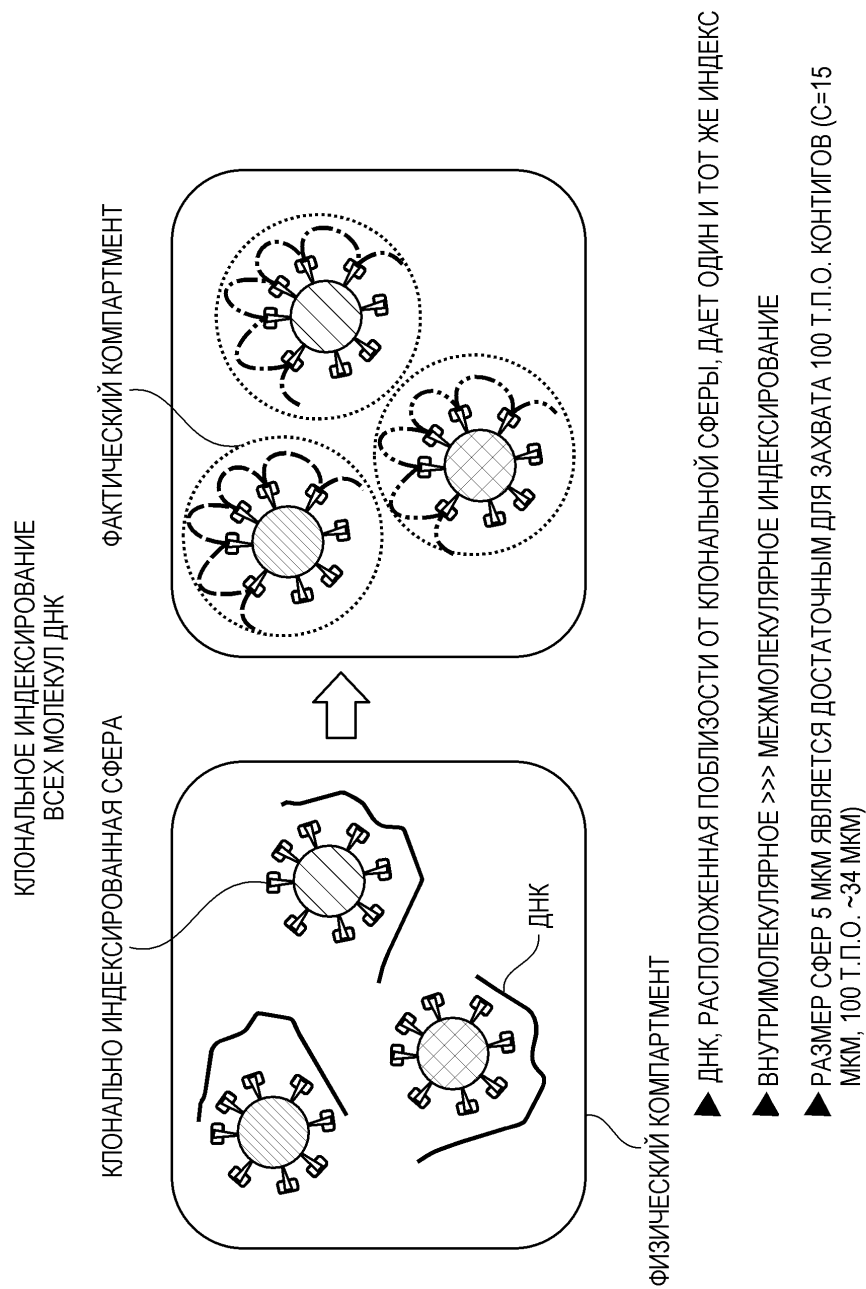
РАЗМЕР И РАСПРЕДЕЛЕНИЕ ОСТРОВКОВ ДЛЯ ИН-
ДЕКСИРОВАНИЯ СФЕР В ОДНОМ СОСУДЕ



РАЗМЕР ОСТРОВКОВ

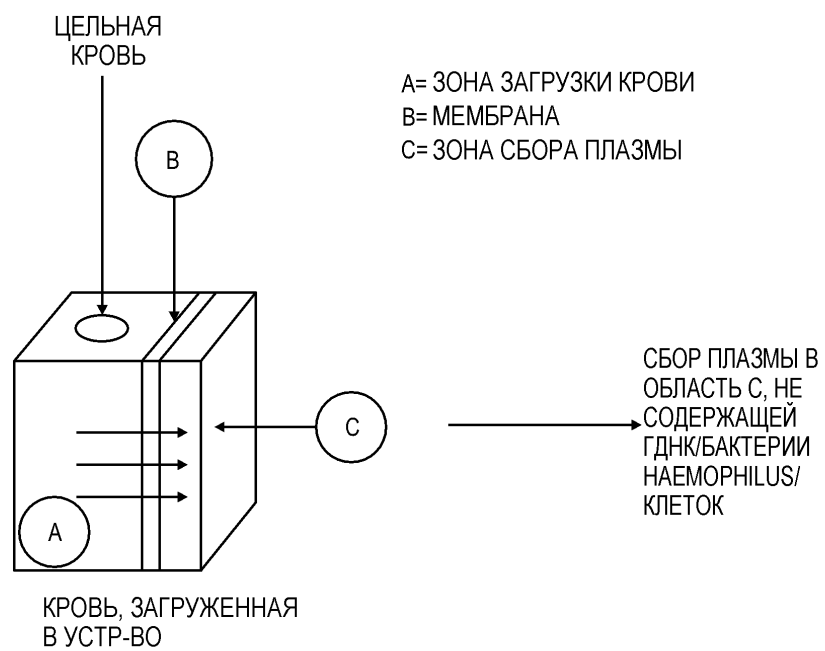
ФИГ. 36

37/76



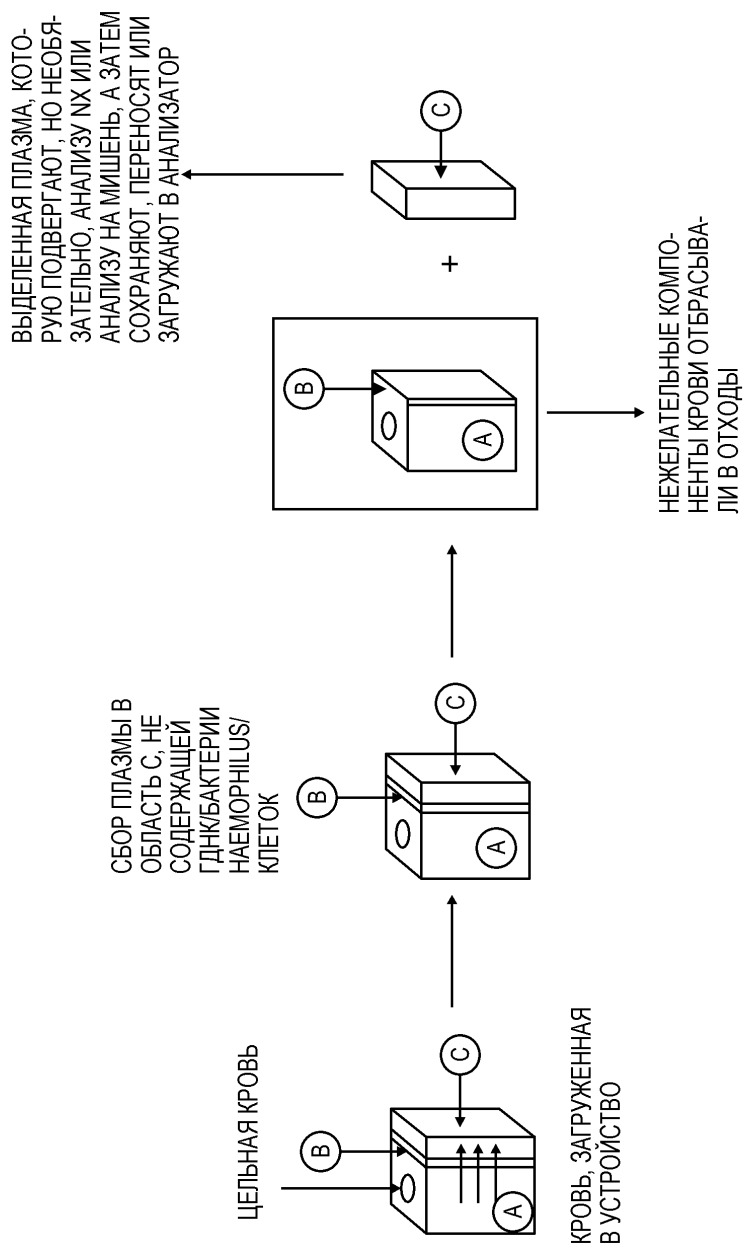
ФИГ. 37

38/76



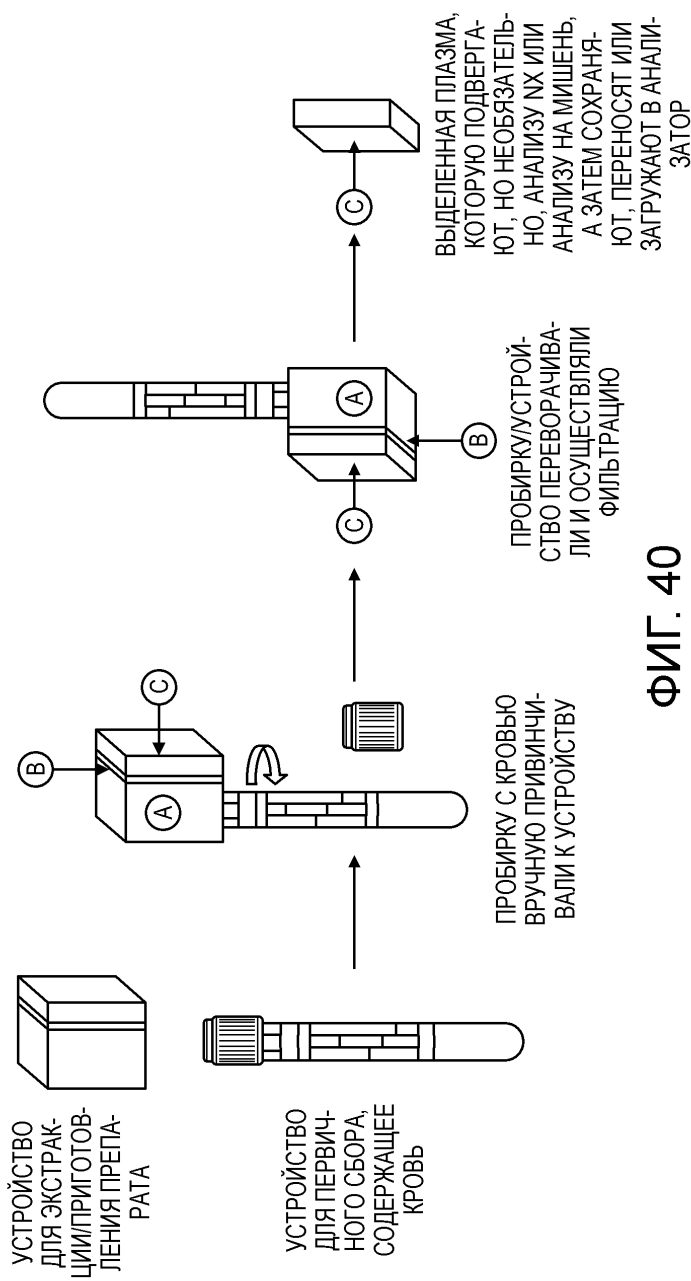
ФИГ. 38

39/76



ФИГ. 39

40/76

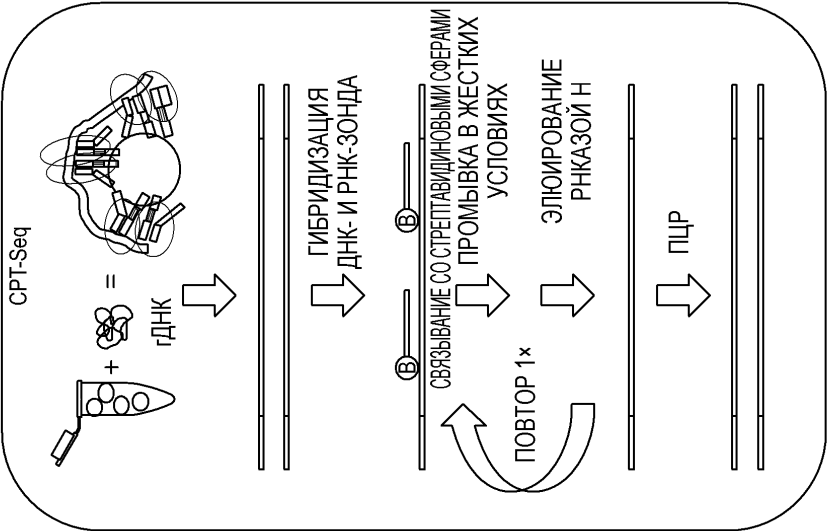


ФИГ. 40

41/76

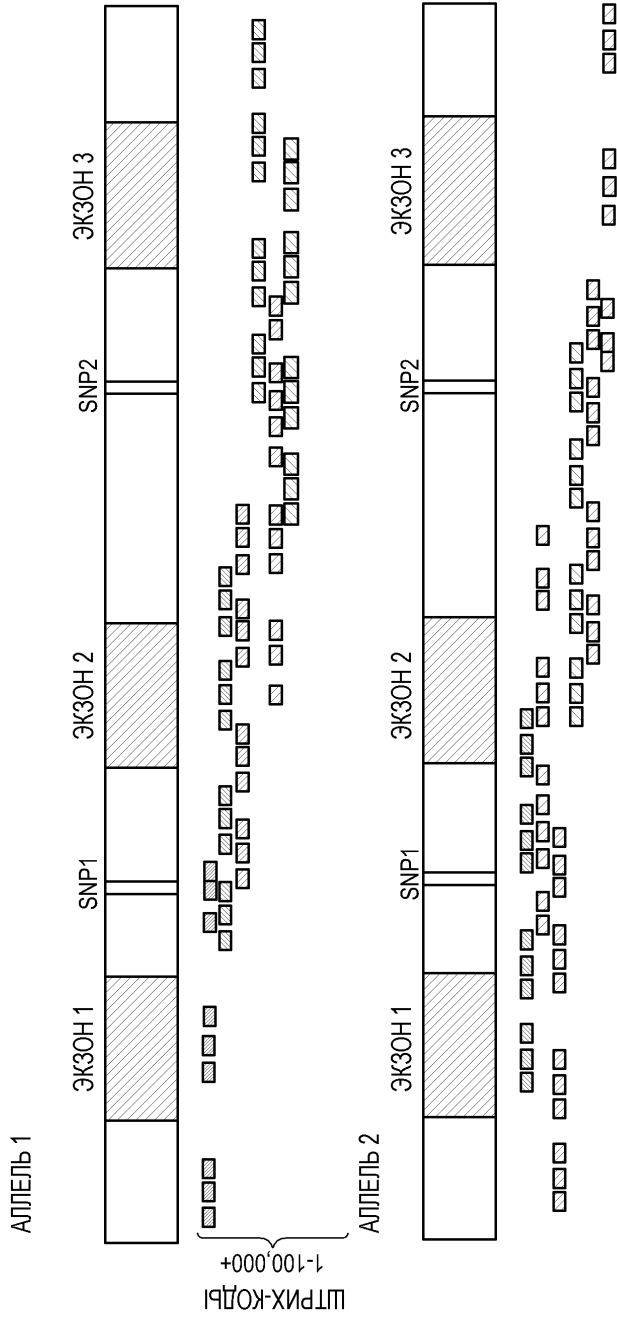
ОБОГАЩЕНИЕ CPT-SEQ

- ЦЕЛЕВОЕ ОБОГАЩЕНИЕ СЦЕПЛЕННЫХ С ИНДЕКСАМИ БИБЛИОТЕК ВСЕГО ГЕНОМА
- ИНДЕКСАЦИЯ ОБРАЗЦА+СО-ВМЕСТИМОЕ ОБОГАЩЕНИЕ

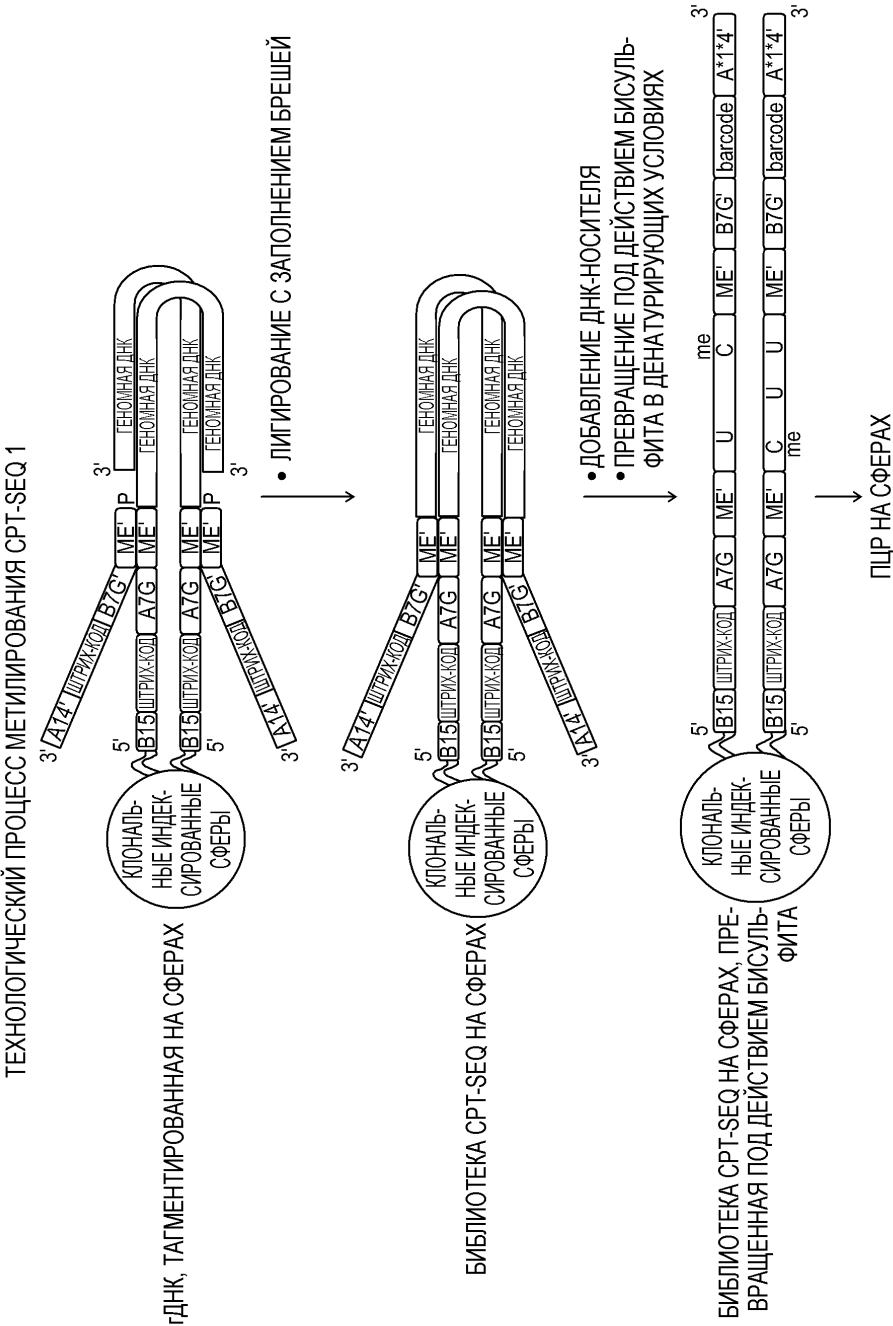


ФИГ. 41

ФАЗИРОВАНИЕ ЭКЗОМА: ОБОГАЩЕНИЕ ЭКЗОМА БИБЛИОТЕК
CPT-SEQ

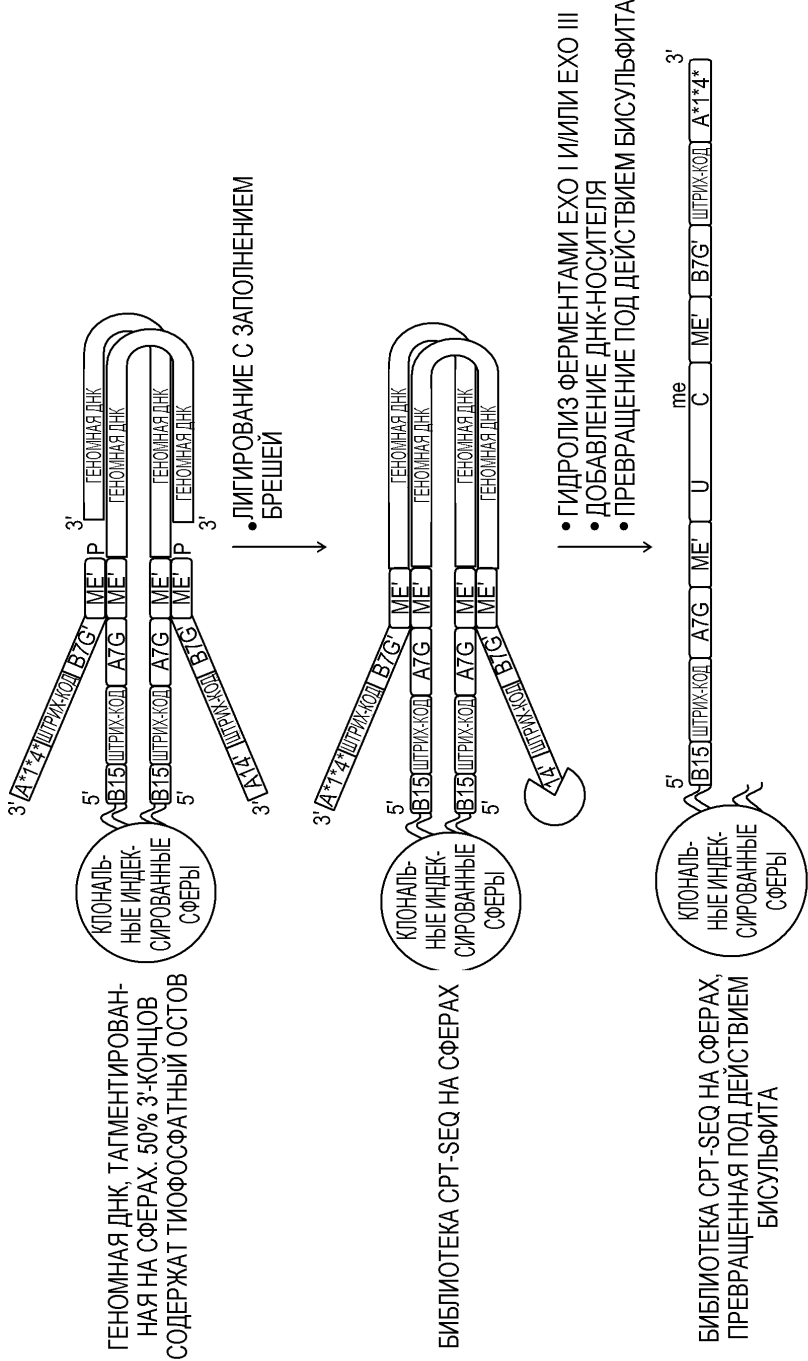


ФИГ. 42



ФИГ. 43

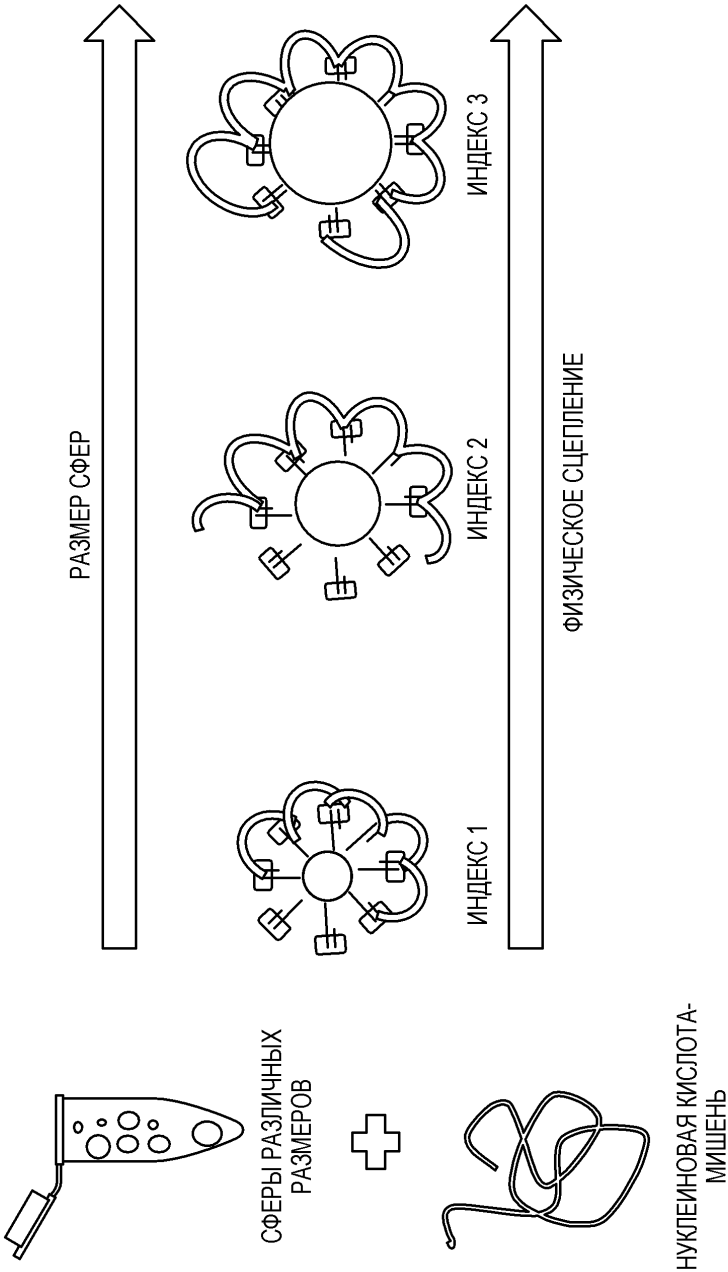
РЕПРЕЗЕНТАТИВНЫЙ ТЕХНОЛОГИЧЕСКИЙ ПРОЦЕСС МЕТИЛИРОВАНИЯ СРТ-SEQ 2



ФИГ. 44

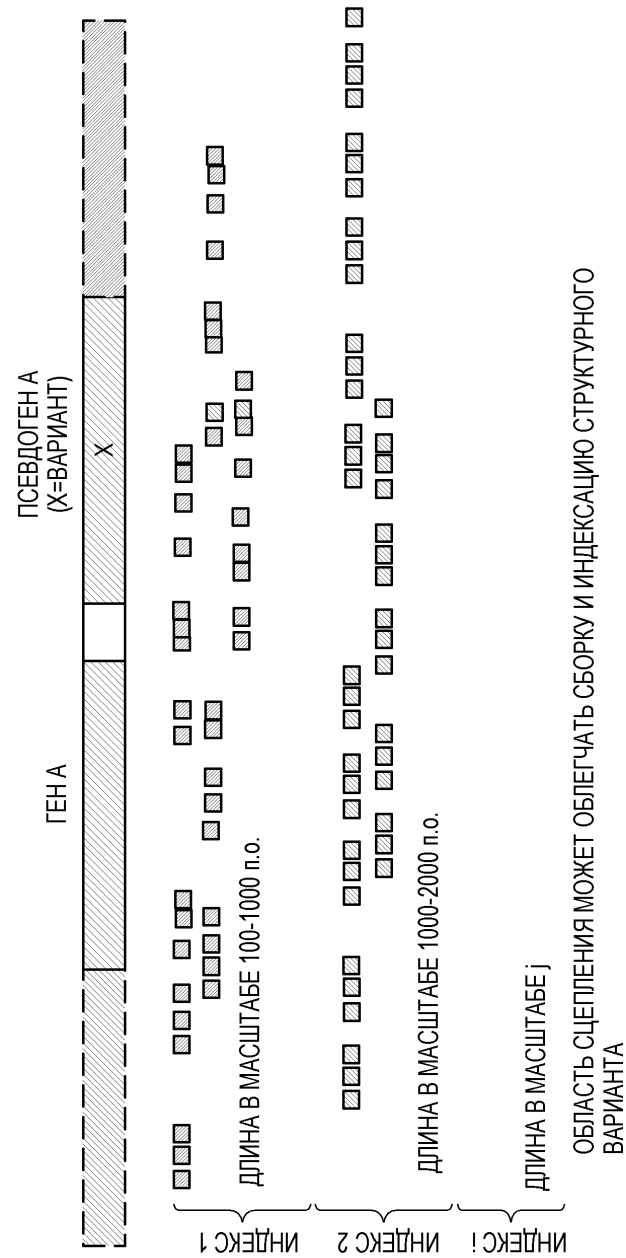
45/76

ПРЕПАРАТЫ БИБЛИОТЕК, ИМЕЮЩИЕ ДЛИНЫ В РАЗЛИЧНЫХ МАСШТАБАХ, В
ОДНОМ АНАЛИЗЕ



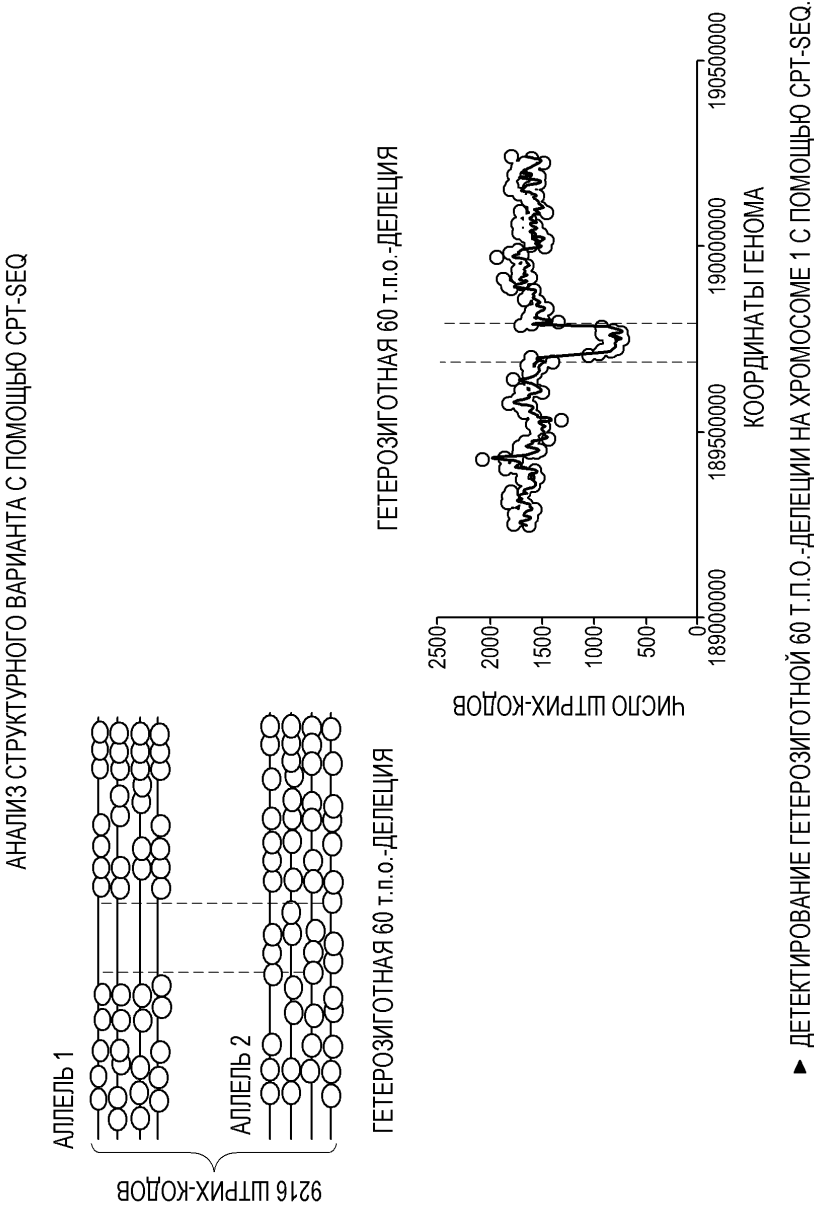
ФИГ. 45

ПРИМЕР СБОРКИ/ХАРАКТЕРИЗАЦИИ SV ДЛЯ БИБЛИОТЕК,
ИМЕЮЩИХ ДЛИНЫ В РАЗЛИЧНЫХ МАСШТАБАХ



ФИГ. 46

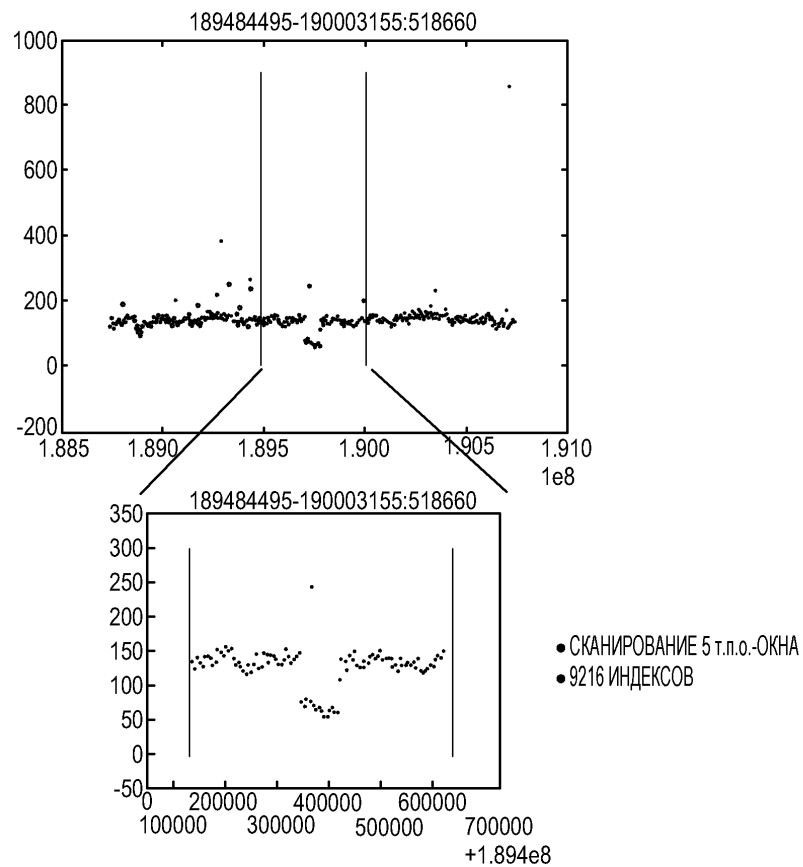
47/76



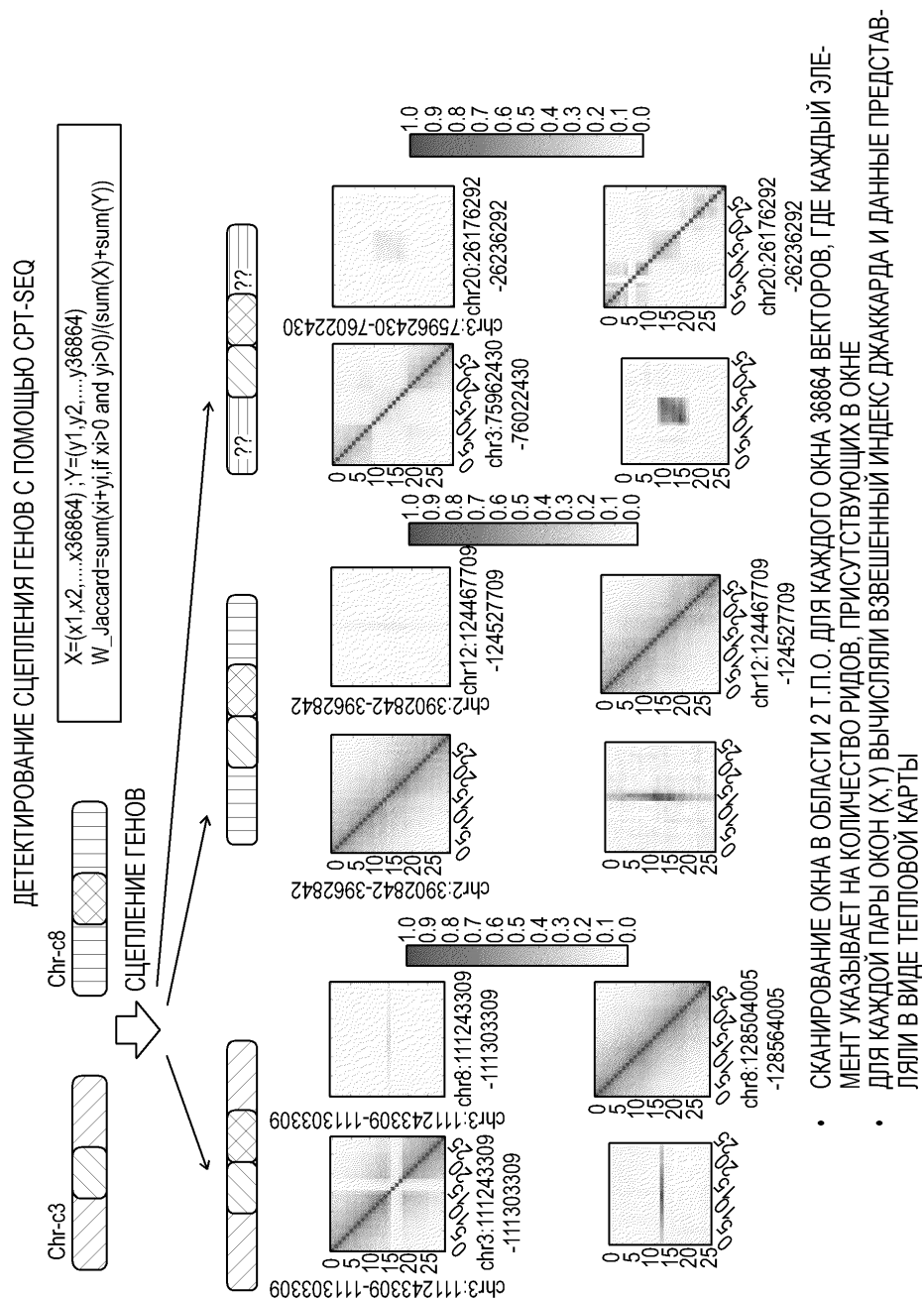
ФИГ. 47А

48/76

ГЕТЕРОЗИГОТНАЯ 60 Т.П.О.-ДЕЛЕ-
ЦИЯ НА ХРОМОСОМЕ 1 НА12878



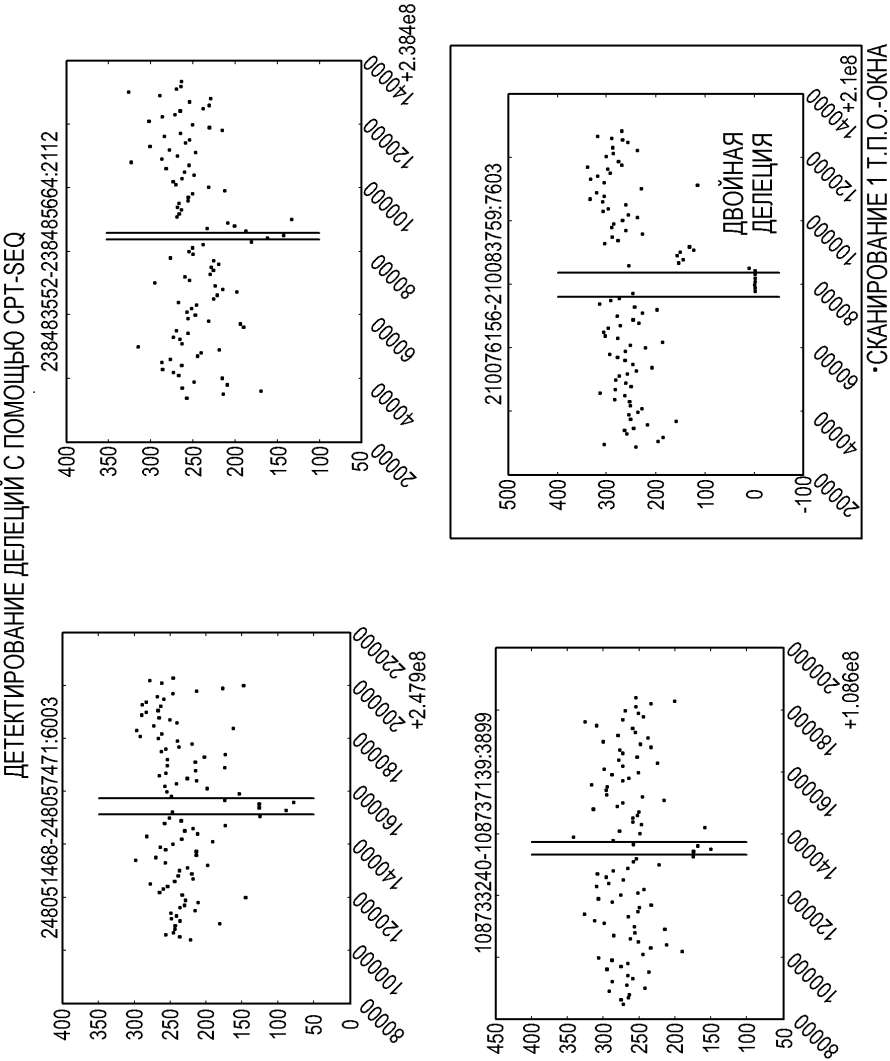
ФИГ. 47В



ΦΠΓ. 48

- сканирование окна в области 2 т.п.о. для каждого окна 36864 векторов, где каждый элемент указывает на количество рядов, присутствующих в окне
- для каждой пары окон (X,Y) вычисляли взвешенный индекс Джаккарда и данные представляли в виде тепловой карты

50/76



ФИГ. 49

ОПТИМИЗАЦИЯ ЭФФЕКТИВНОСТИ ОБРАБОТКИ ПОСРЕДСТВОМ БИСУЛЬФИТНОГО ПРЕВРАЩЕНИЯ

- ОПТИМИЗИРОВАННАЯ СИСТЕМА БИСУЛЬФИТНОГО ПРЕВРАЩЕНИЯ PROMEGA METHYLENBLUE
ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ПРЕВРАЩЕНИЯ

УСЛОВИЕ	ДНК	СФЕРЫ	BSC-ОБРАБОТКА
1	10 нг	НЕТ	1 ЧАС @ 60°C/0,3М NaOH
2			1 ЧАС @ 60°C/0,3М NaOH
3		ДА	1 ЧАС @ 60°C/1М NaOH
4			1 ЧАС @ 65°C/0,3М NaOH

- МЕ-ПОСЛЕДОВАТЕЛЬНОСТЬ АНАЛИЗИРОВАЛИ ДЛЯ ОПРЕДЕЛЕНИЯ ЭФФЕКТИВНОСТИ ОБРАБОТКИ ПОСРЕДСТВОМ БИСУЛЬФИТНОГО ПРЕВРАЩЕНИЯ

НЕПРЕВРАЩЕННАЯ МЕ	GATGTGTATAAGAGACAG
BSC МЕ	AATATATATAAAAAACAA

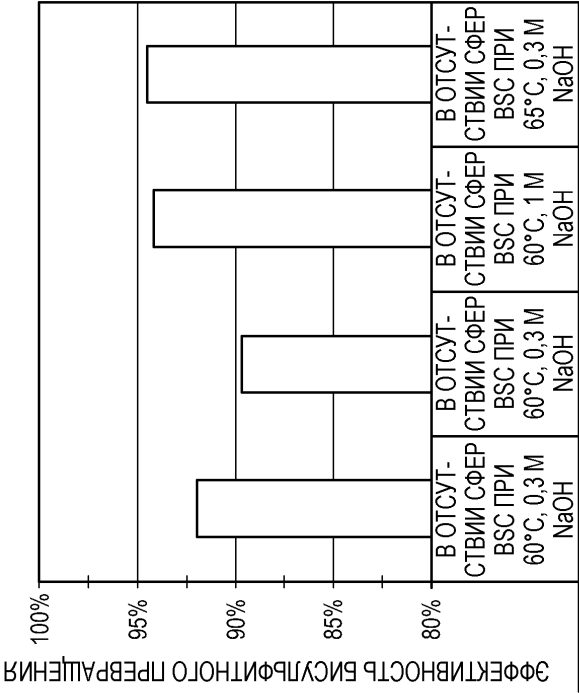


ФИГ. 50

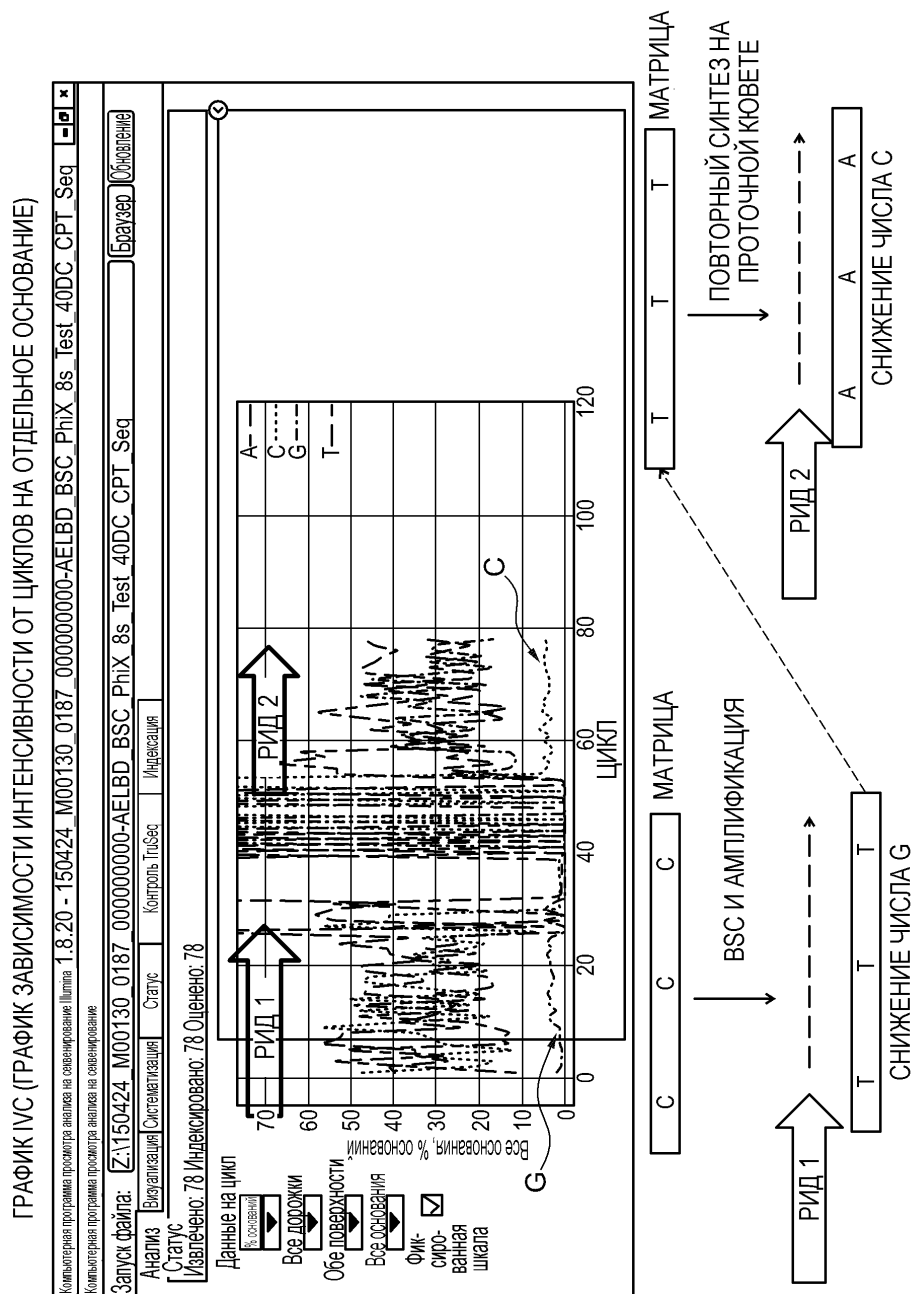
ОПТИМИЗАЦИЯ ЭФФЕКТИВНОСТИ ОБРАБОТКИ ПОСРЕДСТВОМ БИСУЛЬФИТНОГО ПРЕВРАЩЕНИЯ

РЕЗУЛЬТАТЫ

- 95% СЦЕПЛЕННЫХ С ИНДЕКСАМИ БИБЛИОТЕК, ПОДВЕРГНУТЫХ БИСУЛЬФИТНОМУ ПРЕВРАЩЕНИЮ (BSC), БЫЛИ ПРИСОЕДИНЕНЫ К СФЕРАМ
- АНАЛОГИЧНЫЕ ВЫХОДЫ ПЦР-ПРОДУКТОВ, НАБЛЮДАЕМЫЕ В УСЛОВИЯХ БИСУЛЬФИТНОГО ПРЕВРАЩЕНИЯ > В БОЛЕЕ ЖЕСТКИХ УСЛОВИЯХ БИСУЛЬФИТНОГО ПРЕВРАЩЕНИЯ, НЕ ПРИВОДИЛИ К РАЗЛОЖЕНИЮ БИБЛИОТЕК



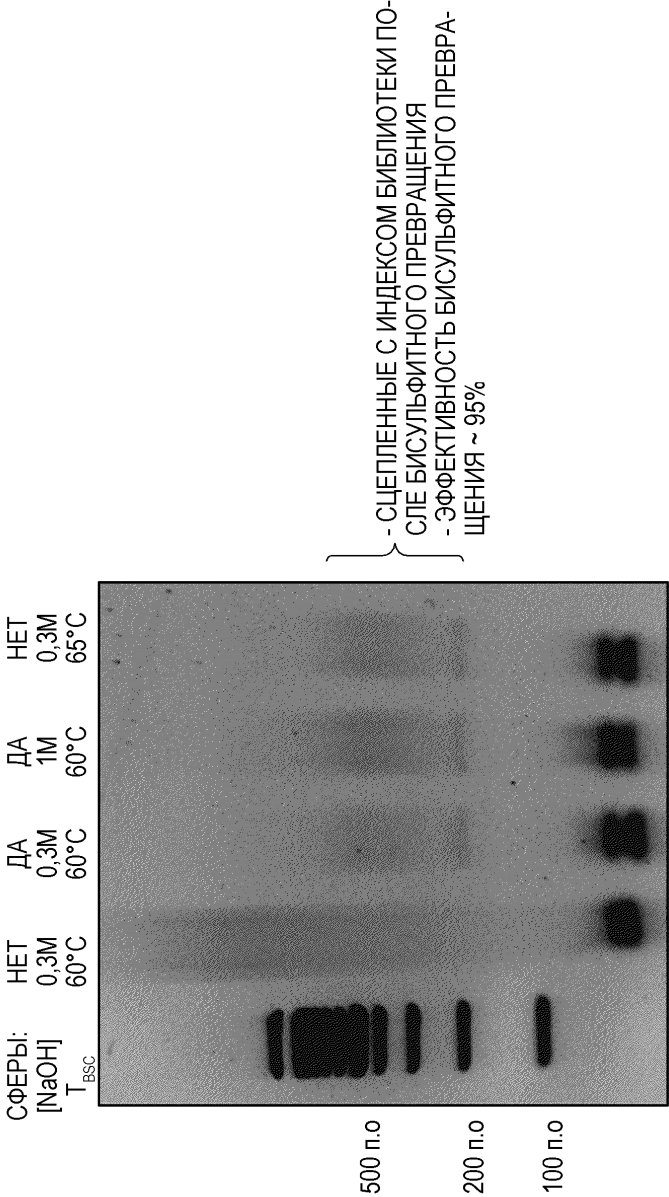
ФИГ. 51



Фиг. 52

54/76

ОПТИМИЗАЦИЯ ЭФФЕКТИВНОСТИ БИСУЛЬФИТНОГО ПРЕВРАЩЕНИЯ

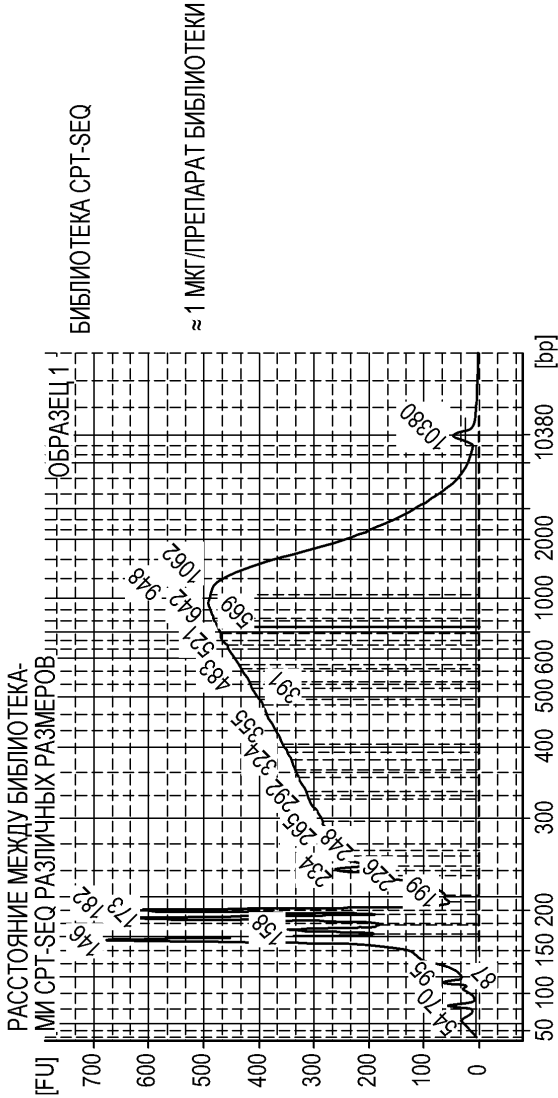


ФИГ. 53

55/76

ОБОГАЩЕНИЕ БИБЛИОТЕК CPT-SEQ

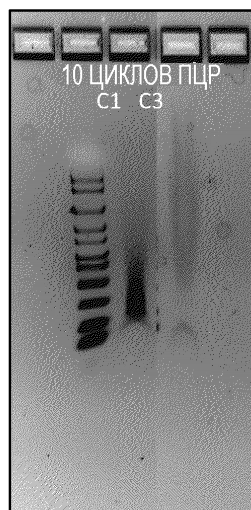
УСЛОВИЯ	БИБЛИОТЕКА	СЕРИЯ ЗОНДОВ	[Зонд] пМ/к.о.	ОБЪЕМ ГИБРИДИЗАЦИИ	ГРАДИЕНТ ГИБРИДИЗАЦИИ	ЕНВ	ЭЛЮИРОВАНИЕ	ПРОМЫВКИ	SMB
1	1,1 мкг CPT-Seq	ДНК по Данну А,В,К	100 пМ	25 мкл	95>68 (В ТЕЧ. НОЧИ)	ФОРМА, NXT BLKRS	ЕТ1	СТАНДАРТ	3х
2	1,1 мкг CPT-Seq	ЕЕХ	2,5 мкл	25 мкл	95>68 (В ТЕЧ. НОЧИ)	ФОРМА, NXT BLKRS	ЕТ1	СТАНДАРТ	3х
3	1,1 мкг CPT-Seq	РНК по Данну А,В,К	100 пМ	25 мкл	95>65 (В ТЕЧ. НОЧИ)	ФОРМА, NXT BLKRS	РНКАЗА Н	СТАНДАРТ	3х



ФИГ. 54

56/76

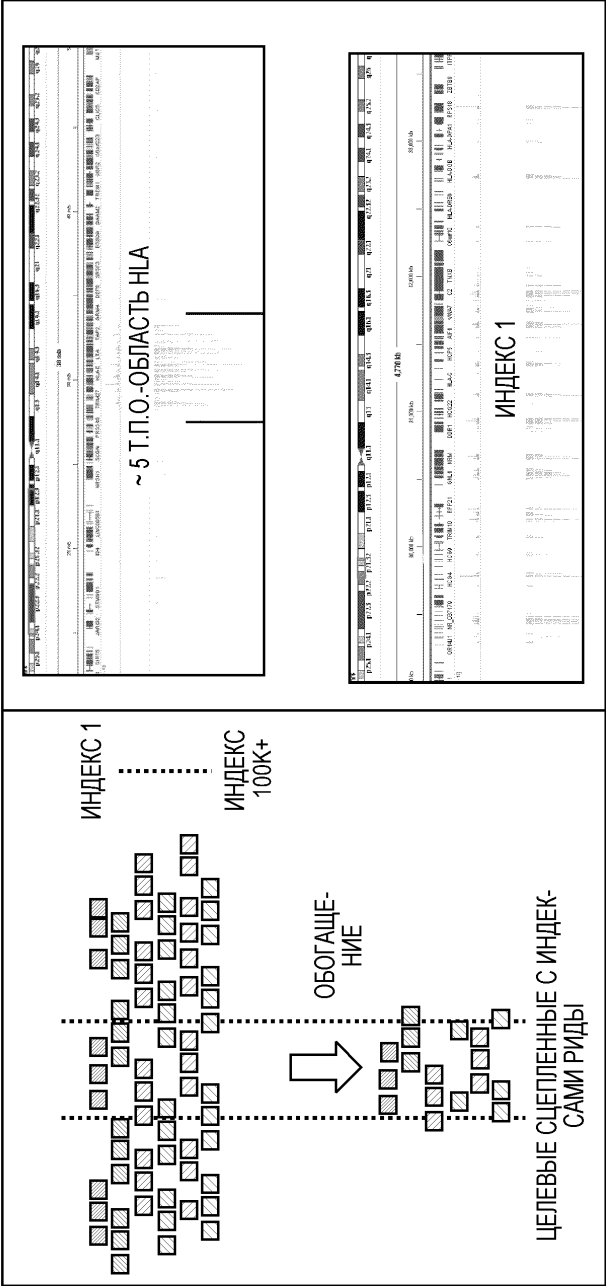
БИБЛИОТЕКИ СРТ-SEQ ПОСЛЕ ОБОГАЩЕНИЯ



ЦЕЛЕВЫЕ СЕКВЕНИРУЮЩИЕ БИБЛИОТЕКИ,
СЦЕПЛЕННЫЕ С ИНДЕКСАМИ

ФИГ. 55

ЦЕЛЕВОЕ ГАПЛОТИПИРОВАНИЕ, ПРИМЕНЯЕМОЕ
ДЛЯ НЛА



ФИГ. 56

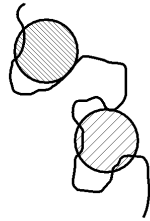
58/76

МЕХАНИЗМЫ ЗАМЕНЫ

1. ДНК СВЯЗЫВАЕТСЯ С ДВУМЯ РАЗЛИЧНЫМИ ИНДЕКСИРОВАННЫМИ СФЕРАМИ

- ВОЗМОЖНЫЕ РАСТВОРЫ:

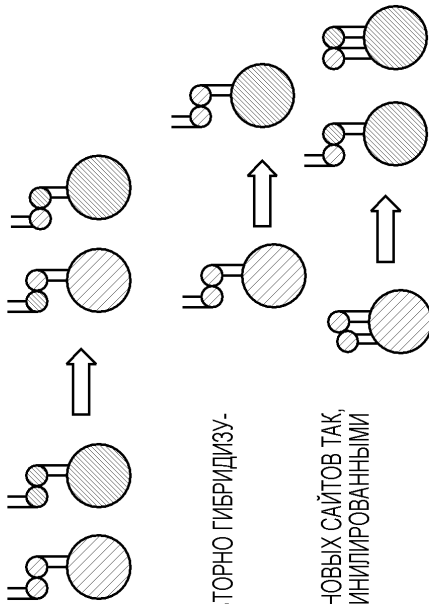
- БОЛЬШЕ ПРОСТРАНСТВА МЕЖДУ СФЕРАМИ (РАЗВЕДЕНИЯ БОЛЬШЕГО КОЛИЧЕСТВА ДНК ПЕРЕД ДОБАВЛЕНИЕМ)
- ДОБАВЛЕНИЕ СФЕР-ИМИТАТОРОВ ДЛЯ УДАЛЕНИЯ БОЛЬШЕГО КОЛИЧЕСТВА КОМПЛЕКСА, СВЯЗАННОГО СО СФЕРАМИ



2. РАЗДЕЛЕНИЕ КОМПЛЕКСОВ И ИХ ПОВТОРНОЕ ОБРАЗОВАНИЕ

- ВОЗМОЖНЫЕ РАСТВОРЫ:

- СТАДИИ С БОЛЕЕ ЛЕГКИМ ПОМЕШИВАНИЕМ
- УДАЛЕНИЕ МОНОМЕРОВ, НЕ СВЯЗАННЫХ С SA, ПУТЕМ ПРОМЫВКИ
- ПЕРЕКРЕСТНОЕ СВЯЗЫВАНИЕ?
- ОБРАЗОВАНИЕ НОВЫХ КОМПЛЕКСОВ?



3. СВЯЗЬ SA-БИОТИН РАЗРЫВАЕТСЯ И КОМПЛЕКС ПОВТОРНО ГИБРИДИЗУЕТСЯ С ДРУГОЙ СФЕРОЙ

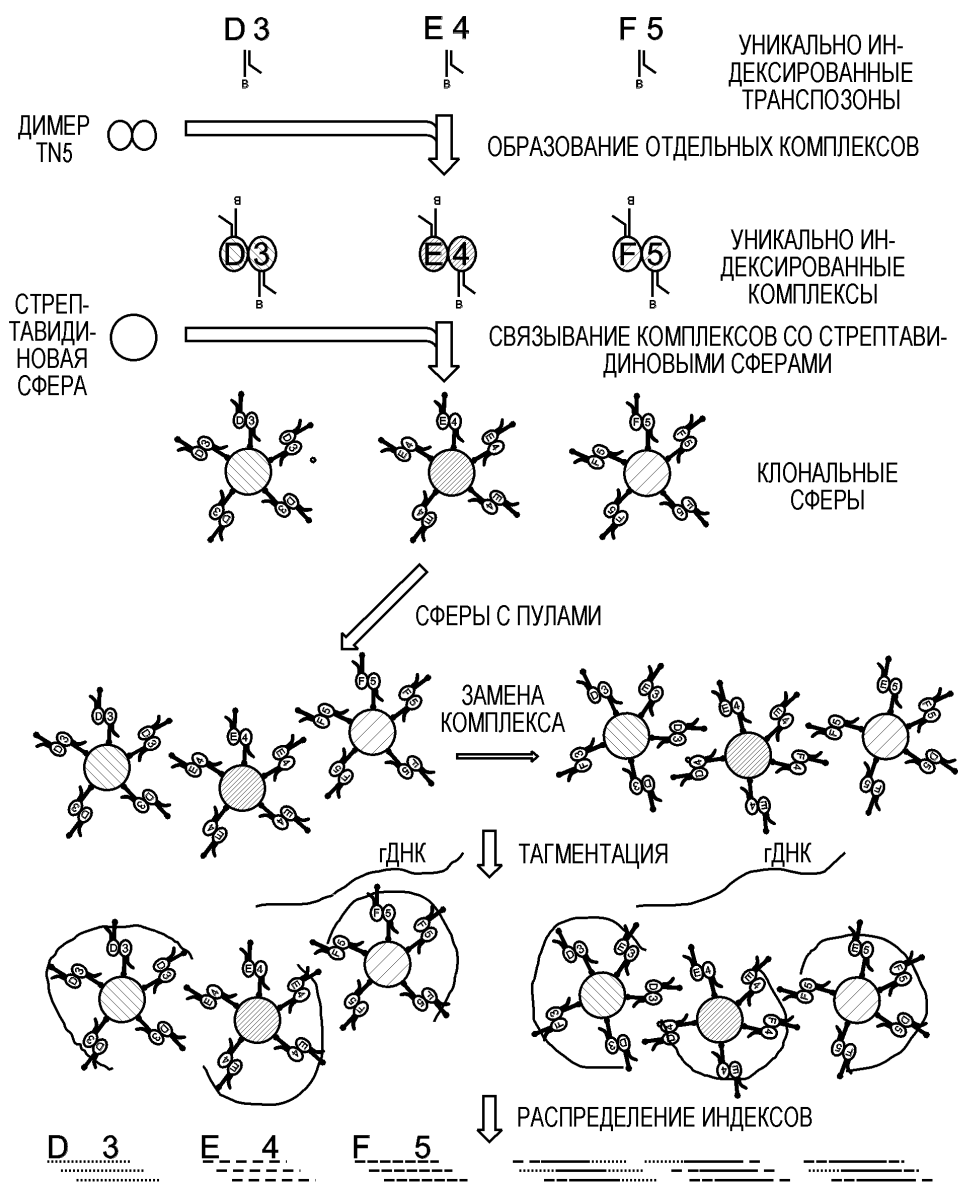
- ВОЗМОЖНЫЕ РАСТВОРЫ:

- СТРЕПТАВИДИНОВЫЕ СФЕРЫ
- БЛОКИРОВАНИЕ СВОБОДНЫХ СТРЕПТАВИДИНОВЫХ САЙТОВ ТАК, ЧТОБЫ ОНИ НЕ МОГЛИ СВЯЗЫВАТЬСЯ С БИОТИНИЛИРОВАННЫМИ ОЛИГОНУКЛЕОТИДАМИ

4. КОНТАМИНАЦИЯ

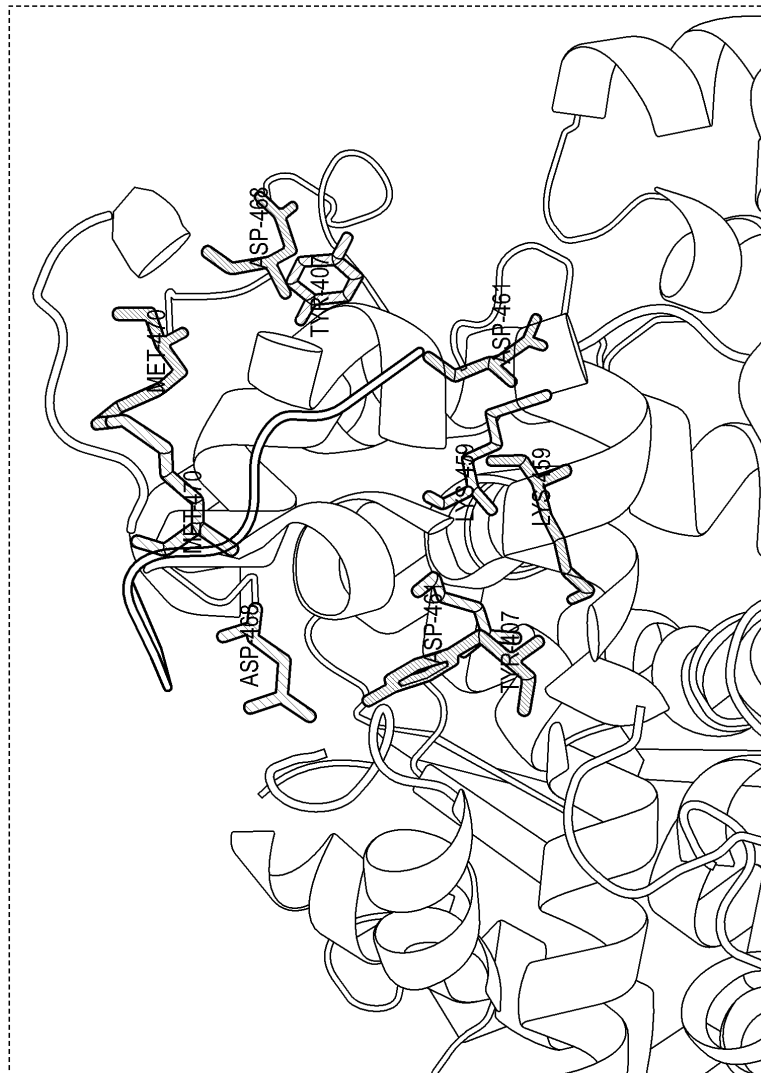
ФИГ. 57

59/76



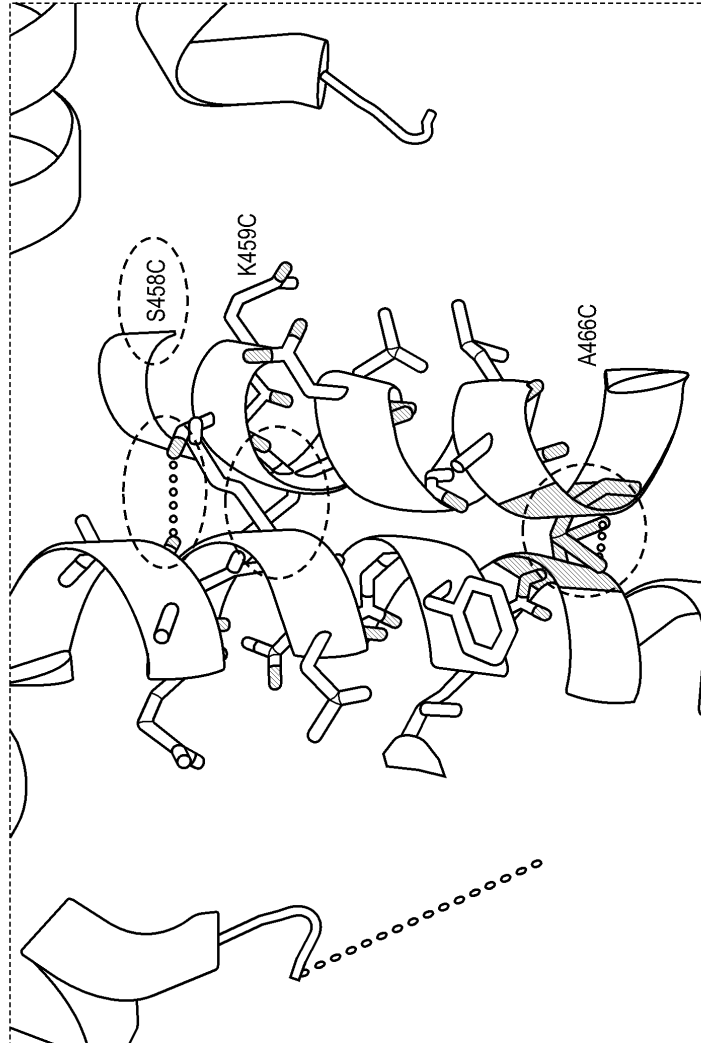
ФИГ. 58

60/76



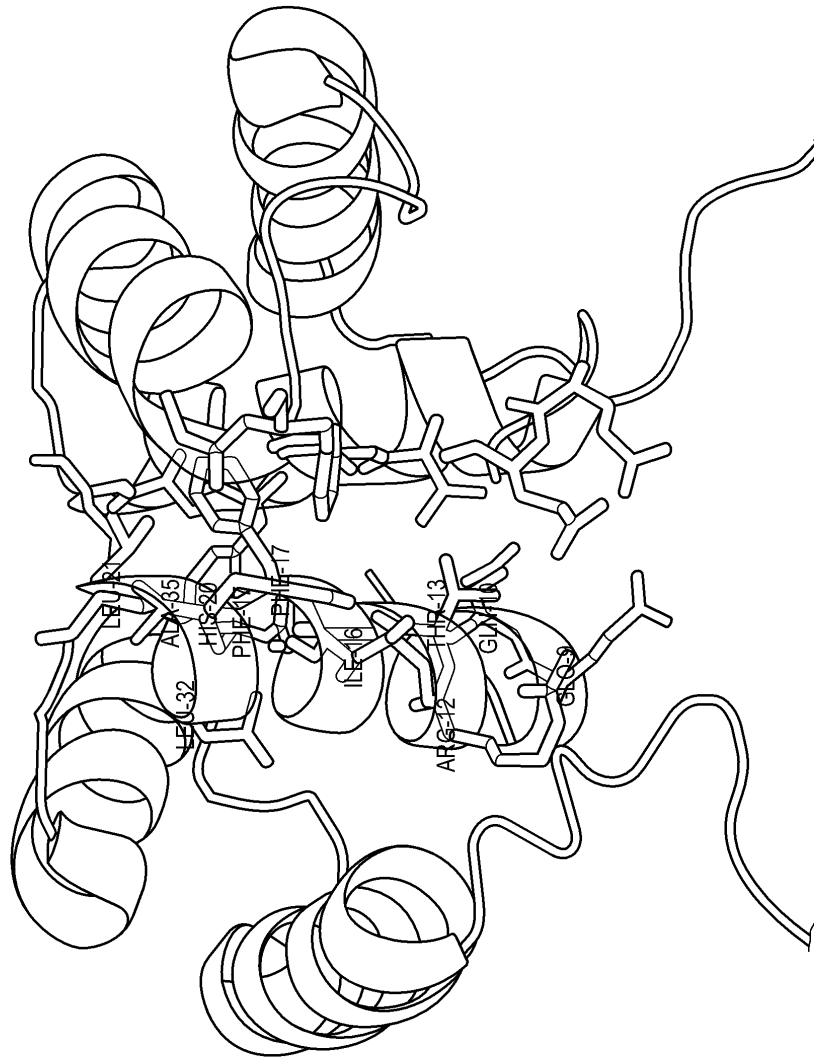
ФИГ. 59

61/76



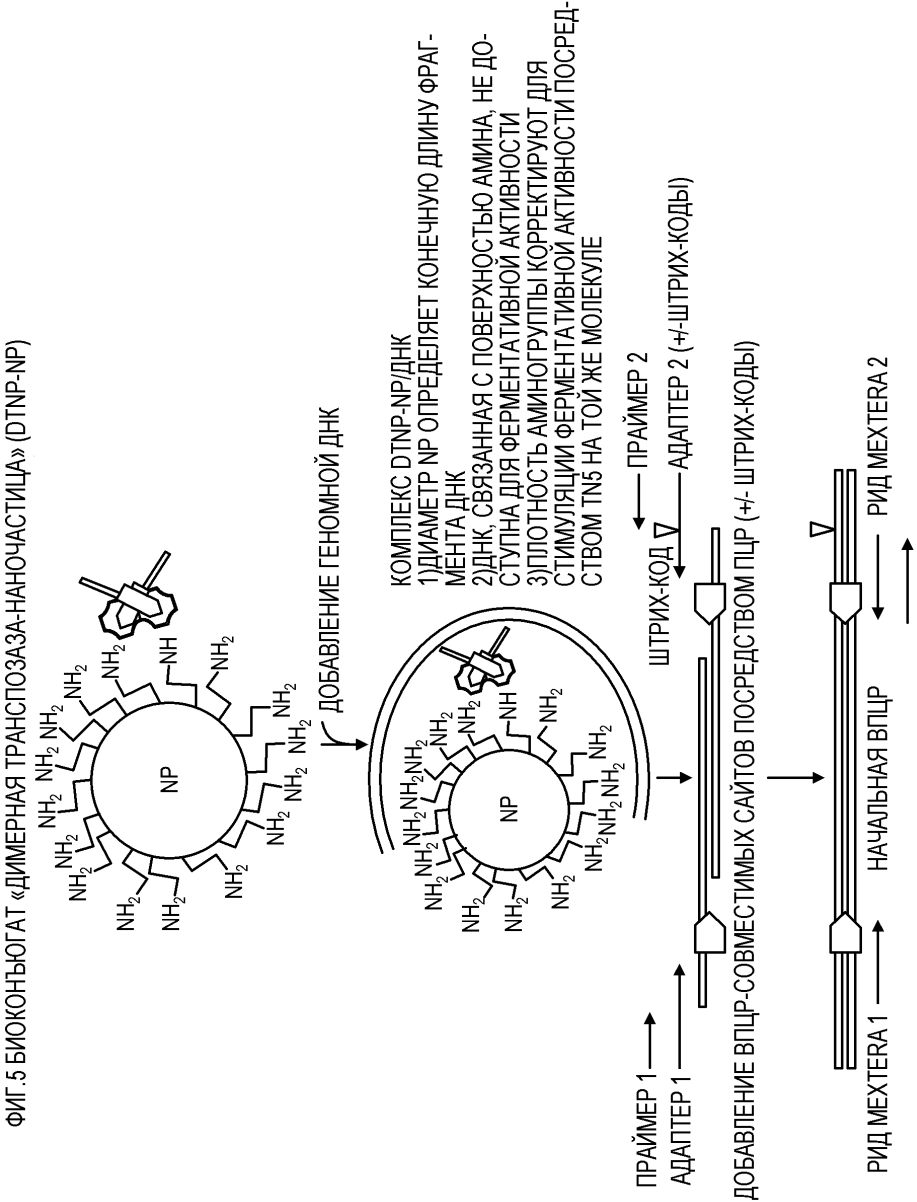
ФИГ. 60

62/76

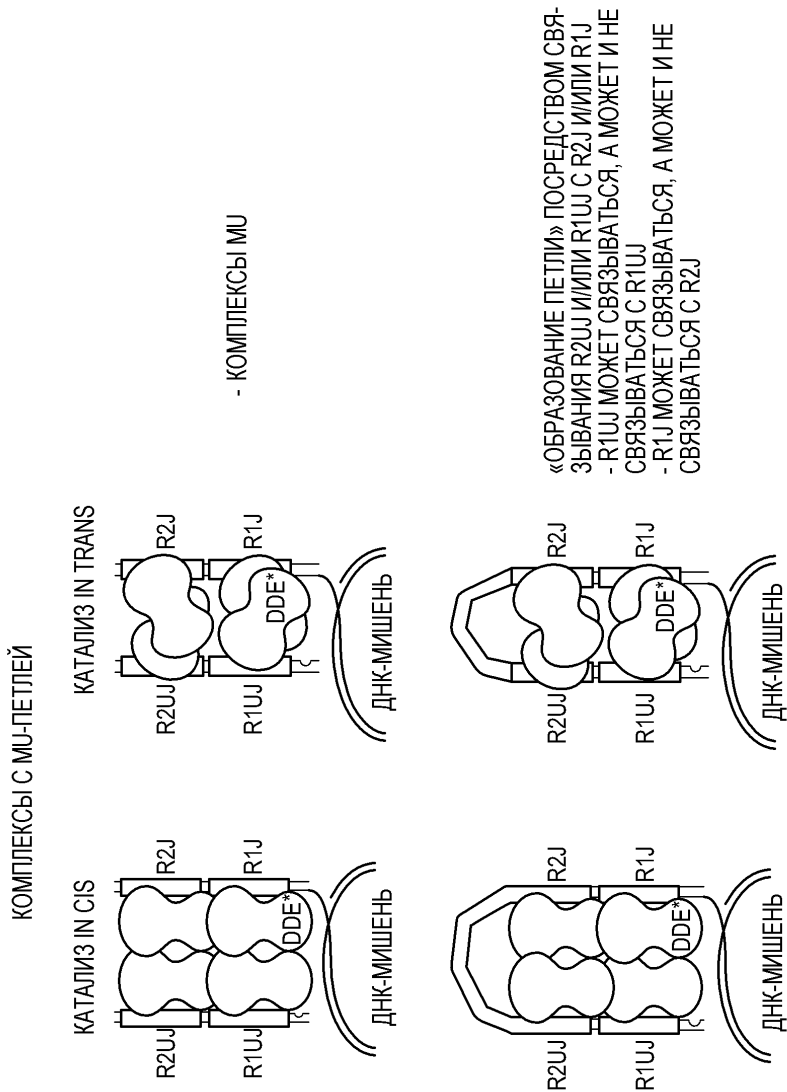


ФИГ. 61

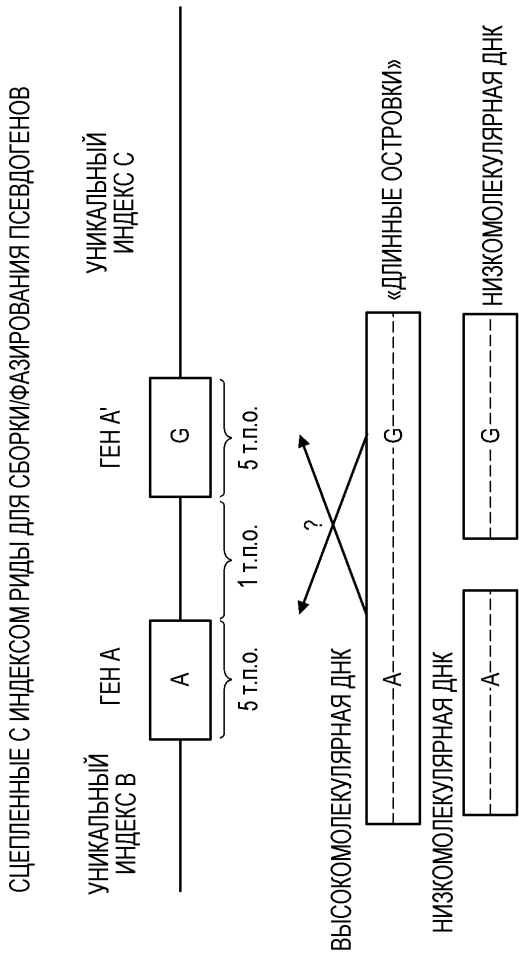
63/76



ФИГ. 62



ФИГ. 63



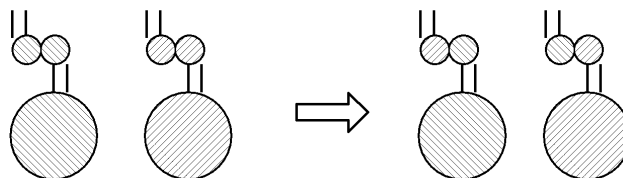
- ПРИМЕР СУР2D6
 - ДЛИННЫЕ ОСТРОВКИ (>30т.п.о.) СЦЕПЛЕННЫХ С ИНДЕКСАМИ РИДОВ ОХВАТЫВАЮТ ПСЕВДОГЕНЫ, ГЕН А И ГЕН А'. ИЗ-ЗА ДЛИННОЙ ПОСЛЕДОВАТЕЛЬНОСТИ НЕВОЗМОЖНО ОПРЕДЕЛИТЬ, КАКОЙ ВАРИАНТ ПРИНАДЛЕЖИТ ГЕНУ А, А КАКОМУ ГЕНУ А'.
 - БОЛЕЕ КОРОТКИЕ ОСТРОВКИ СВЯЗЫВАЮТСЯ ТОЛЬКО С ОДНИМ ВАРИАНТОМ В ОДНОМ ИЗ ПСЕВДОГЕНОВ С УНИКАЛЬНОЙ ОКРУЖАЮЩЕЙ ИХ ПОСЛЕДОВАТЕЛЬНОСТЬЮ
- ЭТО МОЖЕТ БЫТЬ ДОСТИГНУТО ПУТЕМ ФРАГМЕНТИРОВАНИЯ ДНК ПЕРЕД СРТ-SEQ ИЛИ С ИСПОЛЬЗОВАНИЕМ КОМБИНАЦИИ ФРАГМЕНТИРОВАННОЙ И НЕФРАГМЕНТИРОВАННОЙ ДНК

ФИГ. 64

66/76

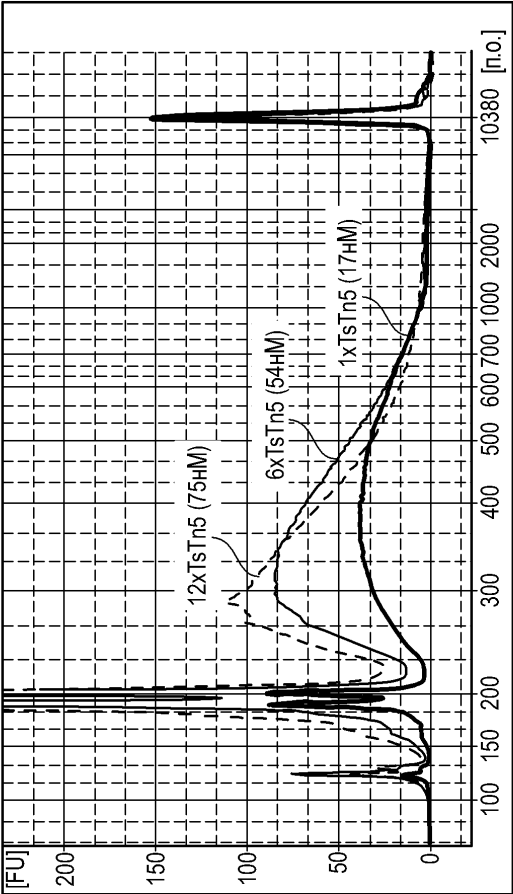


ФИГУРА 1: ЗАМЕНА ИНДЕКСОВ. БЫЛИ ПОЛУЧЕНЫ СФЕРЫ С РАЗЛИЧНЫМИ ИНДЕКСАМИ. ПОСЛЕ СМЕШИВАНИЯ, ЗАМЕНА ИНДЕКСОВ БЫЛА ОПРЕДЕЛЕНА ПУТЕМ СЕКВЕНИРОВАНИЯ И РЕГИСТРАЦИИ ИНДЕКСОВ ДЛЯ КАЖДОЙ БИБЛИОТЕКИ. % «ЗАМЕНЕННЫХ» ИНДЕКСОВ ВЫЧИСЛЯЛИ КАК $(D4+D5+E3+E5+F4)$ (СУММА ВСЕХ 96) И РЕГИСТРИРОВАЛИ НА ФИГУРЕ.

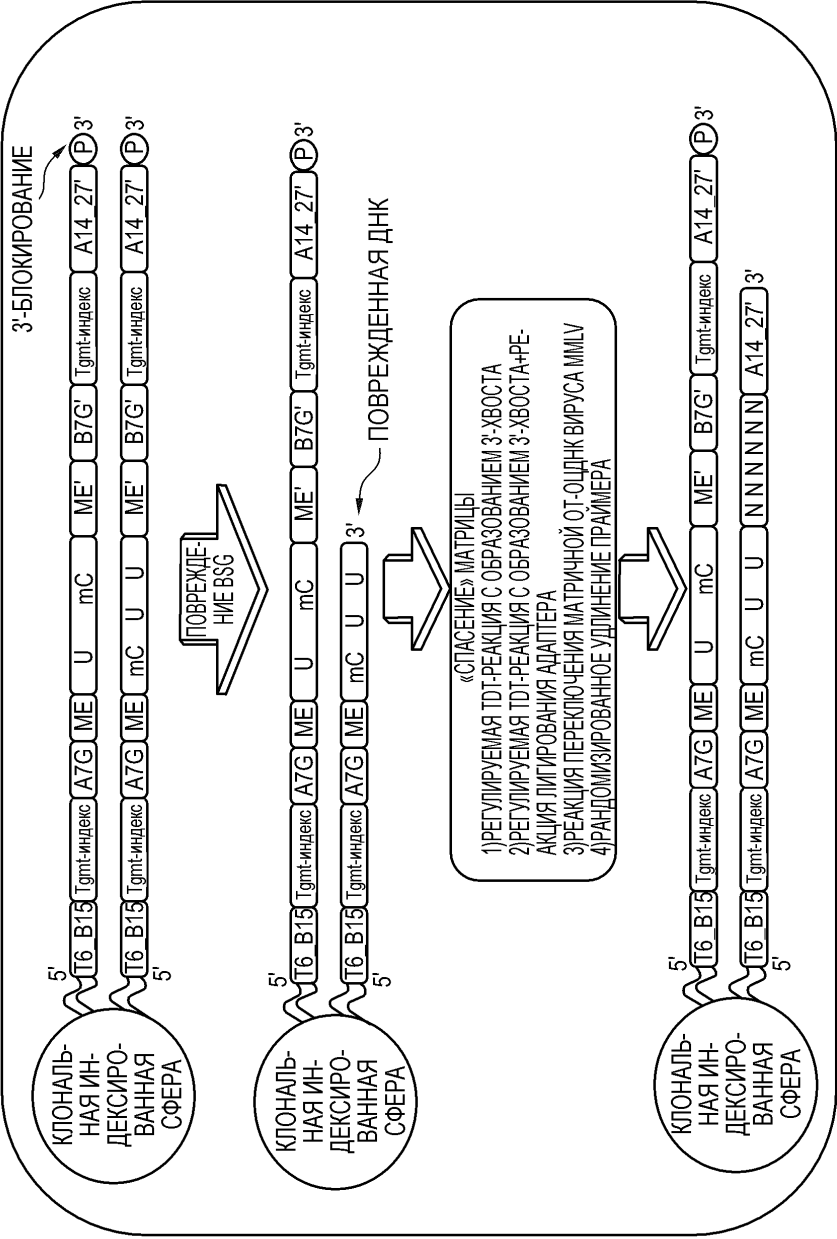


ФИГ. 65

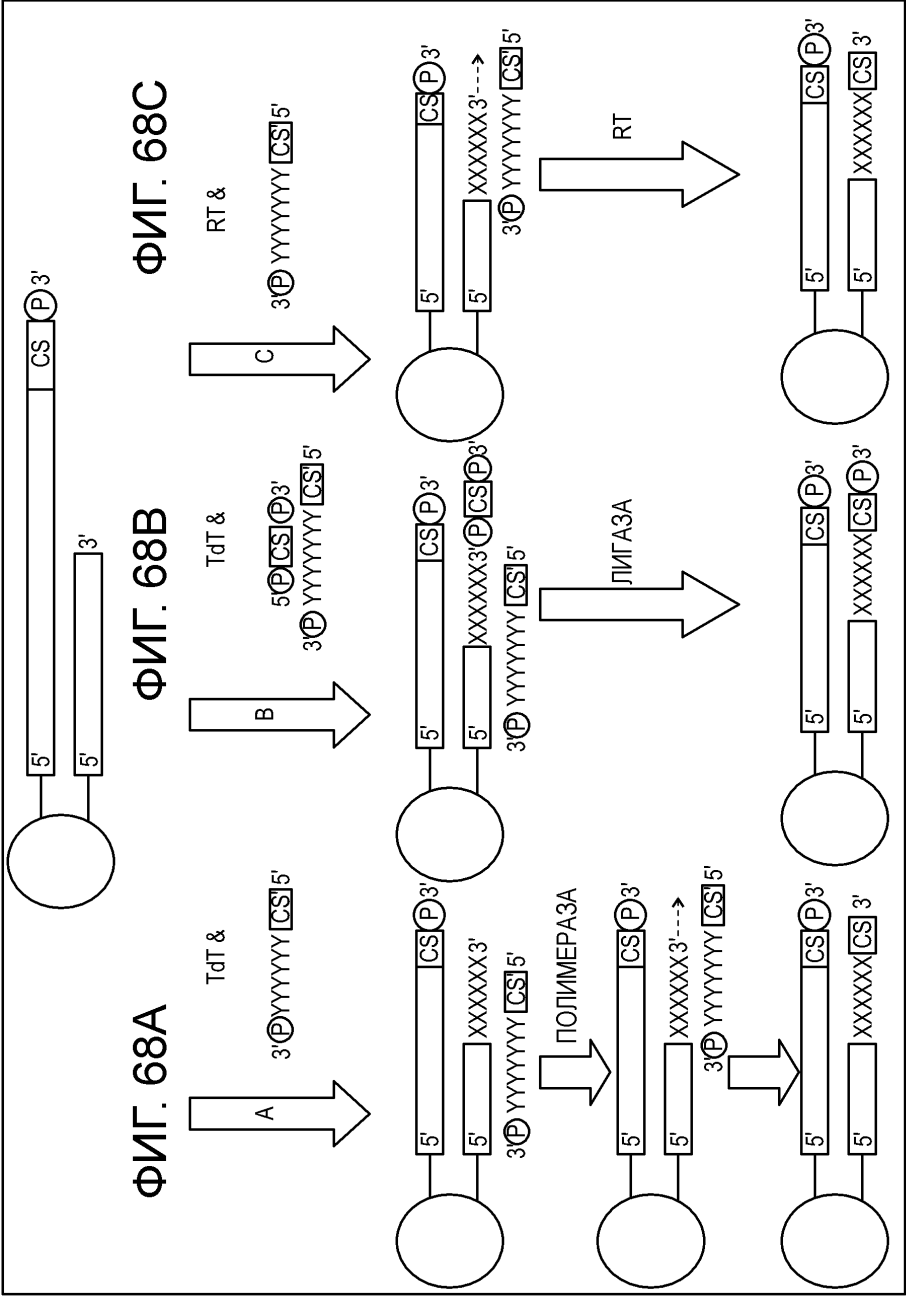
67/76



ФИГ. 66

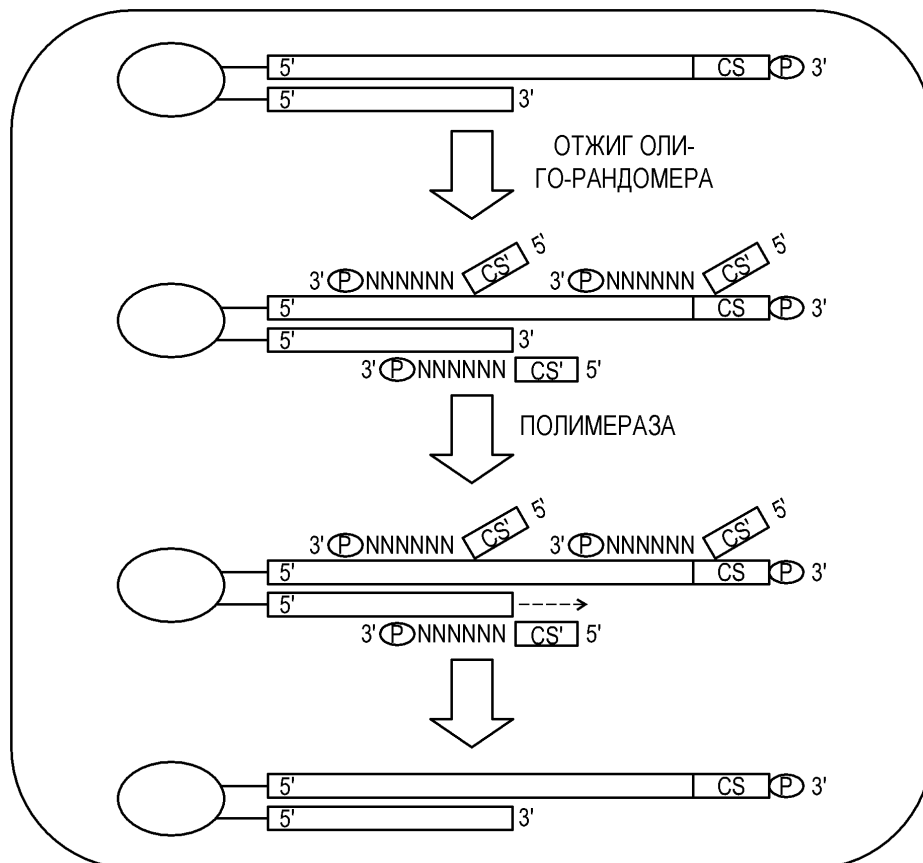


ФИГ. 67



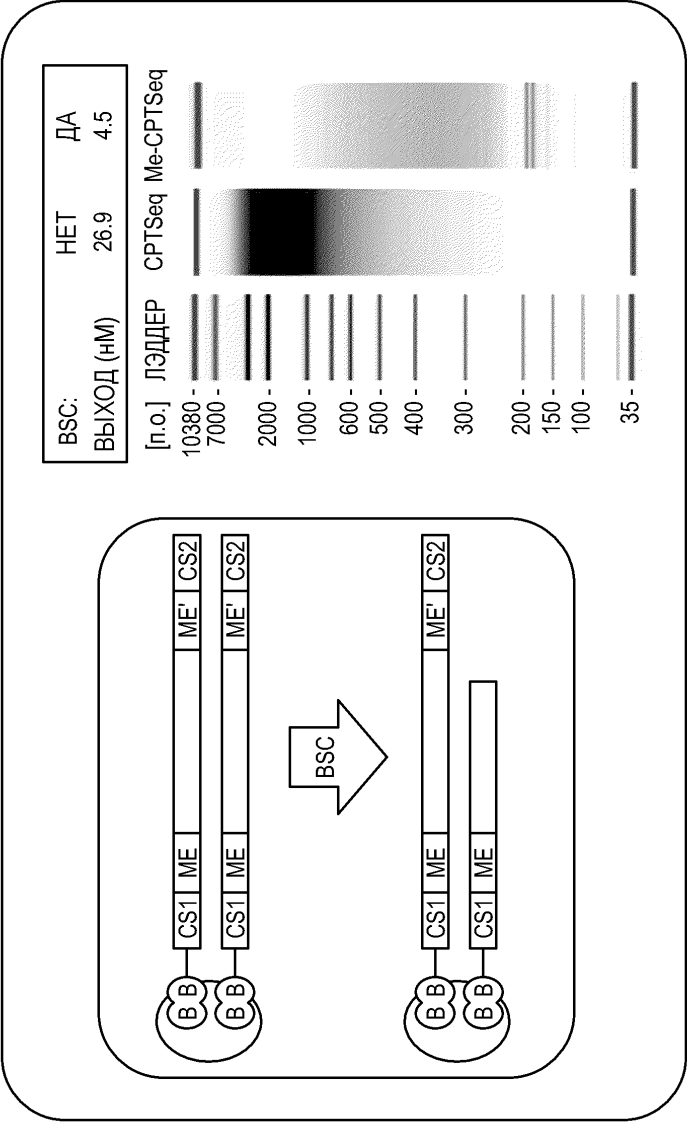
ФИГ. 68

70/76

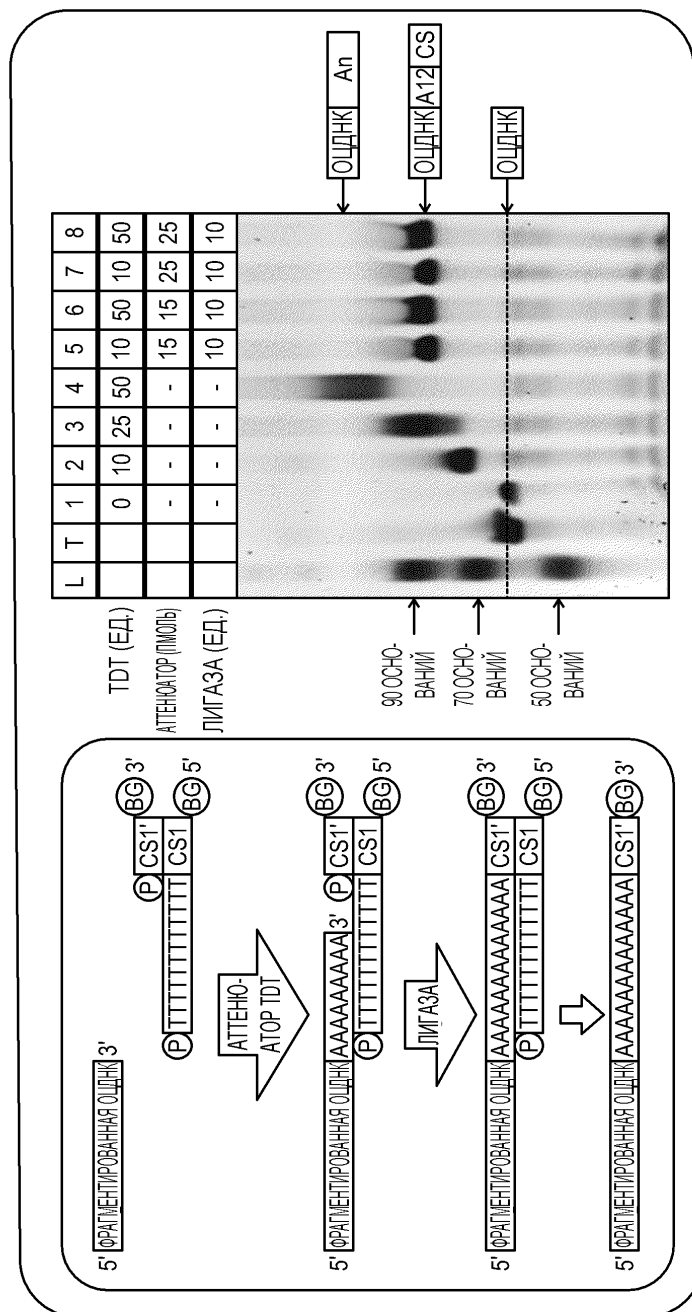


ФИГ. 69

71/76

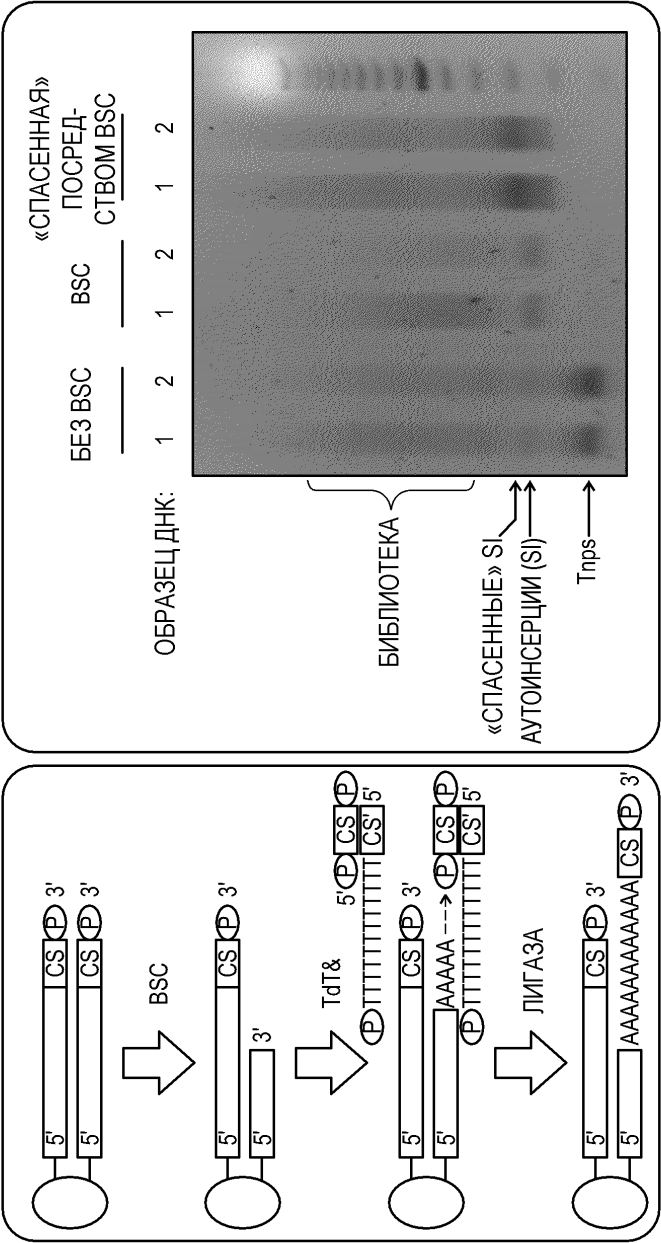


ФИГ. 70



Фиг. 71

73/76

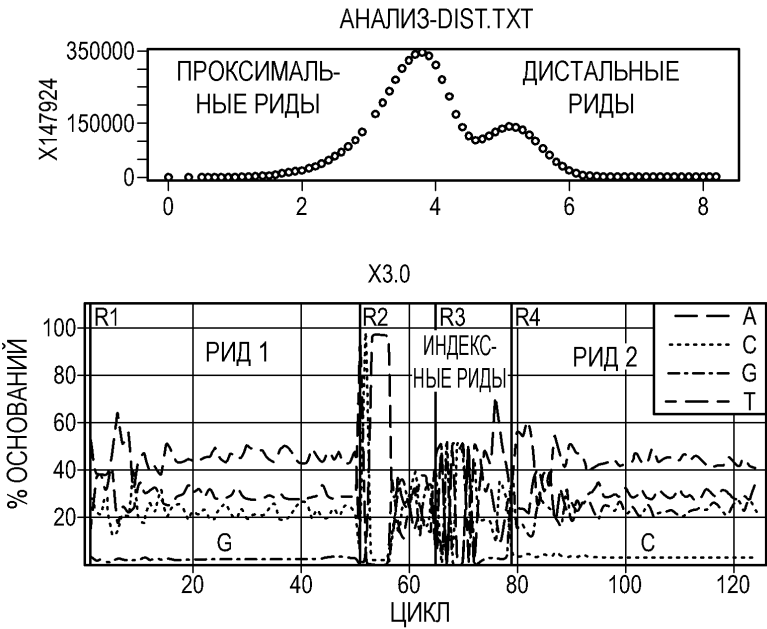


ФИГ. 72

74/76

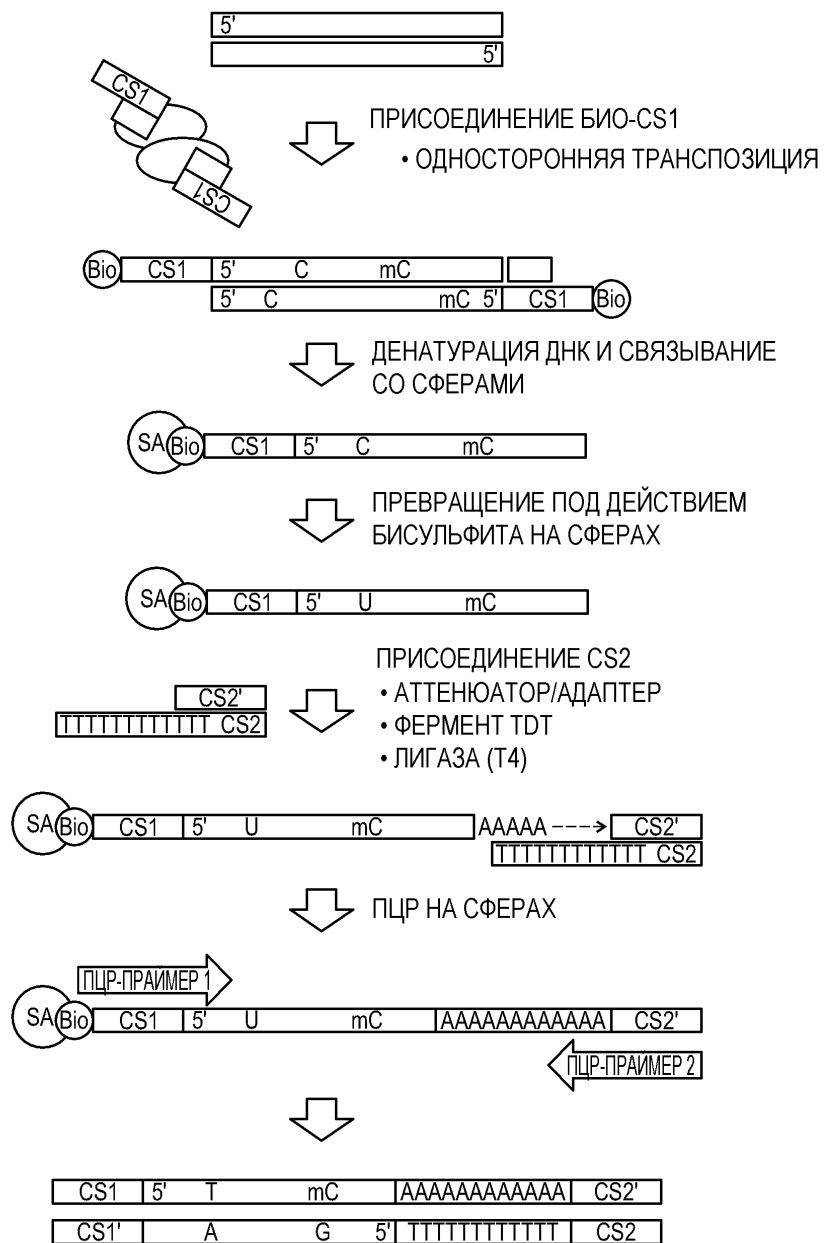
СТАТИСТИЧЕСКИЙ АНАЛИЗ МЕТИЛ-СРТ-SEQ

- НАБЛЮДАЕМЫЕ «ОСТРОВКИ» ФАЗИРОВАННЫХ РИДОВ
 - СРЕДНИЙ РАЗМЕР ОСТРОВКА 17 Т.П.О.
 - >50% УНИКАЛЬНЫХ РИДОВ
 - 88% С БЫЛИ ПРЕВРАЩЕНЫ В U



ФИГ. 73

75/76



ФИГ. 74

76/76

ОПТИМИЗАЦИЯ ЭФФЕКТИВНОСТИ ОБРАБОТКИ ПОСРЕДСТВОМ БИСУЛЬФИТНО-ГО ПРЕВРАЩЕНИЯ
ОПТИМИЗИРОВАННАЯ СИСТЕМА БИСУЛЬФИТНОГО ПРЕВРАЩЕНИЯ PROMEGA METHYLENBLUE ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ПРЕВРАЩЕНИЯ

УСЛОВИЕ	ДНК	СФЕРЫ	BSC-ОБРАБОТКА
1	10 нг	НЕТ	1 ЧАС @ 60°C/0,3M NaOH
2			1 ЧАС @ 60°C/0,3M NaOH
3		ДА	1 ЧАС @ 60°C/1M NaOH
4			1 ЧАС @ 65°C/0,3M NaOH

МЕ-ПОСЛЕДОВАТЕЛЬНОСТЬ АНАЛИЗИРОВАЛИ ДЛЯ ОПРЕДЕЛЕНИЯ ЭФФЕКТИВНОСТИ ОБРАБОТКИ ПОСРЕДСТВОМ БИСУЛЬФИТНОГО ПРЕВРАЩЕНИЯ

НЕПРЕВРАЩЕННАЯ МЕ	GATGTGTATAAGAGACAG
BSC ME	AATATATATAAAAAACAA

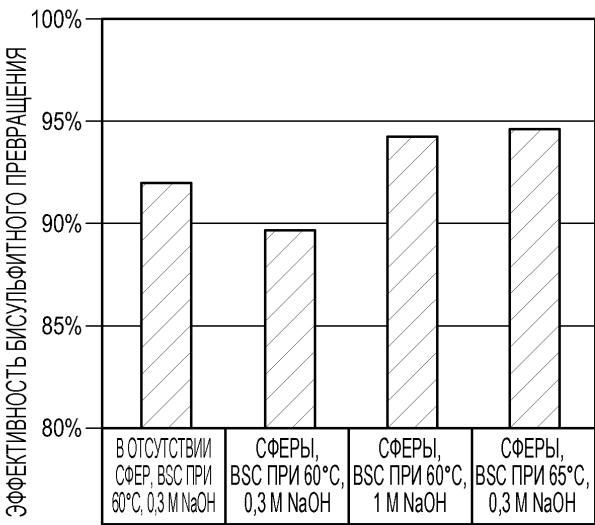
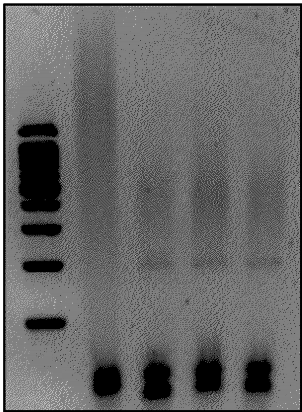
P7'	B15'	ШТРИХ-КОД	A7G'	МЕ'	ГЕНОМНАЯ ДНК	МЕ	B7G	ШТРИХ-КОД	A*1*4*	P5
P7	B15	ШТРИХ-КОД	A7G	МЕ	ГЕНОМНАЯ ДНК	МЕ'	B7G'	ШТРИХ-КОД	A14'	P5'

ФИГ. 75А

РЕЗУЛЬТАТЫ

- 95% БИСУЛЬФИТНОЕ ПРЕВРАЩЕНИЕ, ОСУЩЕСТВЛЯЕМОЕ С ИСПОЛЬЗОВАНИЕМ 1М NAOH ИЛИ ПУТЕМ ОБРАБОТКИ НАГРЕВАНИЕМ ПРИ 65°C
- АНАЛОГИЧНЫЕ ВЫХОДЫ ПЦР-ПРОДУКТОВ, НАБЛЮДАЕМЫЕ В УСЛОВИЯХ БИСУЛЬФИТНОГО ПРЕВРАЩЕНИЯ > В БОЛЕЕ ЖЕСТКИХ УСЛОВИЯХ БИСУЛЬФИТНОГО ПРЕВРАЩЕНИЯ, НЕ ПРИВОДИЛИ К РАЗЛОЖЕНИЮ БИБЛИОТЕК

СФЕРЫ: НЕТ ДА ДА ДА
[NaOH]: 0.3M 0.3M 1M 0.3M
T_{asc}: 60C 60C 60C 65C



ФИГ. 75В