

(19)



**Евразийское
патентное
ведомство**

(21) **201991262** (13) **A1**

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2020.04.07

(22) Дата подачи заявки
2017.12.15

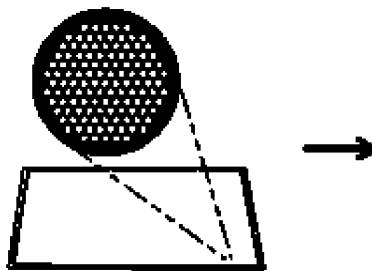
(51) Int. Cl. *A61K 35/12* (2015.01)
C07K 14/705 (2006.01)
C07K 14/725 (2006.01)
C12N 15/09 (2006.01)
C12N 15/10 (2006.01)
C12N 15/12 (2006.01)

(54) БИБЛИОТЕКИ ВАРИАНТОВ ИММУНОЛОГИЧЕСКОГО СИНАПСА И ИХ СИНТЕЗ

(31) 62/435,650
(32) 2016.12.16
(33) US
(86) PCT/US2017/066847
(87) WO 2018/112426 2018.06.21
(71) Заявитель:
ТВИСТ БАЙОСАЙЕНС
КОРПОРЕЙШН (US)

(72) Изобретатель:
Кокс Энтони, Чэнь Сыюань (US)
(74) Представитель:
Строкова О.В., Глухарёва А.О., Лыгу
Т.Н., Угрюмов В.М., Христофоров
А.А., Гизатуллин Е.М., Гизатуллин
Ш.Ф., Костюшенкова М.Ю., Лебедев
В.В., Парамонова К.В. (RU)

(57) В настоящем документе раскрыты способы создания высокоточных библиотек нуклеиновых кислот, кодирующих предварительно определенные варианты последовательности нуклеиновой кислоты. Последовательность нуклеиновой кислоты может кодировать весь или часть TCR или TCR-связывающего антигена. Степень вариации может быть полной, что дает в результате насыщенную библиотеку вариантов, или меньше, чем полной, что дает в результате ненасыщенную библиотеку вариантов. Описанные в настоящем документе библиотеки вариантных нуклеиновых кислот можно сконструировать для дальнейшей обработки путем транскрипции или трансляции. Библиотеки вариантных нуклеиновых кислот, описанные в настоящем документе, можно сконструировать для создания вариантных популяций РНК, ДНК и/или белка. Кроме того, в настоящем документе предусмотрен способ идентификации видов вариантов с повышенными или пониженными активностями, с применениями в регулировании биологических функций и разработке терапевтических средств для лечения или снижения такого заболевания, как рак.



A1

201991262

201991262

A1

БИБЛИОТЕКИ ВАРИАНТОВ ИММУНОЛОГИЧЕСКОГО СИНАПСА И ИХ

СИНТЕЗ

ОПИСАНИЕ

Ссылка на родственную заявку

[0001] Согласно настоящей заявке испрашивается преимущество в соответствии с предварительной заявкой на выдачу патента США № 62/435650, поданной 16 декабря 2016 г., которая полностью включена в настоящий документ посредством ссылки.

Предшествующий уровень техники настоящего изобретения

[0002] Краеугольным камнем синтетической биологии является процесс конструирования, сборки и испытания – итеративный процесс, который требует того, чтобы ДНК была доступной для быстрого и доступного по цене создания и оптимизации этих пользовательских путей и организмов. В фазе конструирования нуклеотиды А, С, Т и G, которые составляют ДНК, составляют в различные генные последовательности, которые будут включать в себя локус или путь, представляющий интерес, причем каждый вариант последовательности представляет конкретную гипотезу, которая будет проверена. Эти варианты последовательностей генов представляют собой подмножества пространства последовательностей, концепции, которая возникла в эволюционной биологии и относится ко всей совокупности последовательностей, которые составляют гены, геномы, транскриптом и протеом.

[0003] Множество различных вариантов, как правило, разрабатывают для каждого цикла конструирования - сборки - испытания, чтобы обеспечить адекватную выборку пространства последовательностей и максимизировать вероятность оптимизированной конструкции. Несмотря на простоту концепции, ограничивающие факторы процесса, связанные со скоростью, пропускной способностью и качеством традиционных способов синтеза, снижают скорость, с которой этот цикл продвигается, увеличивая время разработки. Неспособность в достаточной степени исследовать пространство последовательностей из-за высокой стоимости очень точного секвенирования ДНК и ограниченной пропускной способности современных технологий синтеза остается ограничивающей скоростью стадий.

[0004] Начиная с фазы сборки, два процесса заслуживают внимания: синтез нуклеиновой кислоты и синтез гена. Исторически синтез различных вариантов генов

осуществляли путем молекулярного клонирования. Несмотря на надежность, этот подход не масштабируется. Ранние попытки химического синтеза генов были сосредоточены на получении большого количества нуклеиновых кислот с гомологией перекрывающихся последовательностей. Затем их объединяли и подвергали множественным циклам полимеразной цепной реакции (ПЦР), что позволяло объединять перекрывающиеся нуклеиновые кислоты в полноразмерный двухцепочечный ген. Проведению этого способа препятствует ряд факторов, включая в себя кропотливое и трудоемкое конструирование, потребность в больших объемах фосфорамидитов, дорогостоящее сырье и производство наномолярных количеств конечного продукта, что значительно меньше, чем требуется для последующих стадий, а также большое количество отдельных нуклеиновых кислот, которое необходимо для начала синтеза одного гена на одном 96-луночном планшете.

[0005] Синтезирование нуклеиновых кислот на микроматрицах обеспечило значительное увеличение пропускной способности синтеза генов. Большое количество нуклеиновых кислот можно синтезировать на поверхности микроматрицы, затем отщепить от нее и объединить вместе. Каждая нуклеиновая кислота, предназначенная для определенного гена, содержит уникальную штрих-кодovou последовательность, которая позволила эту конкретную субпопуляцию полинуклеотидов выделить из пула и собрать в представляющий интерес ген. В этой фазе способа каждый субпул переносят в одну лунку в 96-луночном планшете, увеличивая пропускную способность до 96 генов. Хотя пропускная способность на два порядка выше, чем у классического способа, этот способ все еще не обеспечивает адекватной поддержки циклов конструирования, сборки, испытания, для которых требуются тысячи последовательностей одновременно, из-за недостаточной экономической эффективности и длительных сроков выполнения. Таким образом, существует потребность в более эффективном создании библиотек вариантов последовательностей.

[0006] Иммунная система обладает способностью искать и уничтожать вредные клетки. Т-клетки играют важную роль в таком процессе. В контексте надзора за раковыми клетками иммунные ответы Т-клеток, индуцированные эндогенным репертуаром Т-клеток, могут быть недостаточными для удаления таких раковых клеток. Иммунологический синапс включает в себя репертуар Т-клеточных белков, антигенпрезентирующих белков и антигенов. Генетически модифицированные компоненты этой системы обеспечивают средства для усиления опосредованных Т-клетками иммунных ответов. Однако внутренние клеточные факторы, а также иммуносупрессорные факторы в микроокружении опухоли могут ограничивать функцию таких генетически модифицированных Т-клеток. Таким образом, существует потребность

в улучшенных композициях и способах противораковой и противовирусной терапии, в которых используют опосредованную Т-клетками передачу сигналов.

Сущность изобретения

[0007] В настоящем документе предусмотрена библиотека нуклеиновых кислот, причем библиотека нуклеиновых кислот содержит по меньшей мере 3000 вариантных нуклеиновых кислот, причем каждая вариантная нуклеиновая кислота кодирует вариантный ген или фрагмент гена белка рецептора Т-клеток (TCR), и причем область, выбранная для вариации, охватывает вплоть до 1000 оснований в длину. Дополнительно в настоящем документе предусмотрена библиотека нуклеиновых кислот, в которой вариантный ген или фрагмент гена происходит из вариабельного домена. Дополнительно в настоящем документе предусмотрена библиотека нуклеиновых кислот, в которой вариабельный домен представляет собой вариабельный домен TCR-альфа, TCR-бета, TCR-гамма или TCR-дельта. Дополнительно в настоящем документе предусмотрена библиотека нуклеиновых кислот, в которой вариабельный домен является специфическим в отношении ракового антигена. Дополнительно в настоящем документе предусмотрена библиотека нуклеиновых кислот, в которой раковый антиген представляет собой MAGE A3, MAGE A12, MAGE A2, MAGE A6, NY-ESO-1 или CEA. Дополнительно в настоящем документе предусмотрена библиотека нуклеиновых кислот, в которой вариантный ген или фрагмент гена происходит из константного домена. Дополнительно в настоящем документе предусмотрена библиотека нуклеиновых кислот, в которой вариантный ген или фрагмент гена позволяет создавать белок TCR, характеризующийся повышенной специфичностью, авидностью, аффинностью, стабильностью или экспрессией.

[0008] В настоящем документе предусмотрена библиотека полинуклеотидов, причем библиотека полинуклеотидов содержит по меньшей мере 3000 вариантных нуклеиновых кислот, причем каждый полинуклеотид составляет по меньшей мере 15 оснований в длину, причем каждый полинуклеотид кодирует вариант в пределах вариабельного домена белка Т-клеточного рецептора (TCR) или его фрагмента, причем вариабельный домен содержит вплоть до 1000 оснований, и причем каждая вариантная нуклеиновая кислота содержит по меньшей мере одну вариацию в предварительно выбранном кодоне для аминокислотного остатка в области контакта с антигеном. Дополнительно в настоящем документе предусмотрена библиотека полинуклеотидов, в которой вариабельный домен представляет собой вариабельный домен TCR-альфа, TCR-бета, TCR-гамма или TCR-дельта. Дополнительно в настоящем документе предусмотрена библиотека полинуклеотидов, в которой антиген представляет собой раковый антиген.

Дополнительно в настоящем документе предусмотрена библиотека полинуклеотидов, в которой раковый антиген представляет собой MAGE A3, MAGE A12, MAGE A2, MAGE A6, NY-ESO-1 или CEA. Дополнительно в настоящем документе предусмотрена библиотека полинуклеотидов, в которой каждая вариантная нуклеиновая кислота содержит множество вариаций в предварительно выбранных кодонах для аминокислотных остатков в области контакта с антигеном. Дополнительно в настоящем документе предусмотрена библиотека полинуклеотидов, в которой множество вариаций в предварительно выбранных кодонах для аминокислотных остатков содержит вплоть до 100 остатков. Дополнительно в настоящем документе предусмотрена библиотека полинуклеотидов, в которой множество вариаций в предварительно выбранных кодонах для аминокислотных остатков содержит вплоть до 30 остатков. Дополнительно в настоящем документе предусмотрена библиотека полинуклеотидов, в которой множество вариаций в предварительно выбранных кодонах для аминокислотных остатков содержит вплоть до 5 остатков. Дополнительно в настоящем документе предусмотрена библиотека полинуклеотидов, в которой вариант позволяет создавать белок TCR, характеризующийся повышенной специфичностью, авидностью, аффинностью, стабильностью или экспрессией. Дополнительно в настоящем документе предусмотрена библиотека полинуклеотидов, в которой каждая вариантная нуклеиновая кислота дополнительно содержит по меньшей мере одну вариацию в предварительно выбранном кодоне для аминокислотного остатка в пределах константного домена белка Т-клеточного рецептора (TCR) или его фрагмента.

[0009] В настоящем документе предусмотрена библиотека полинуклеотидов, причем библиотека полинуклеотидов содержит по меньшей мере 3000 вариантных полинуклеотидов, причем каждый полинуклеотид составляет длину по меньшей мере 15 оснований в длину, причем каждый полинуклеотид кодирует вариант в пределах константного домена белка Т-клеточного рецептора (TCR) или его фрагмента, причем константный домен содержит вплоть до 1000 оснований, и причем вариант содержит по меньшей мере одну вариацию в предварительно выбранном кодоне для аминокислотного остатка в области контакта с антигеном. Дополнительно в настоящем документе предусмотрена библиотека полинуклеотидов, в которой константный домен представляет собой константный домен TCR-альфа, TCR-бета, TCR-гамма или TCR-дельта. Дополнительно в настоящем документе предусмотрена библиотека полинуклеотидов, в которой антиген представляет собой раковый антиген. Дополнительно в настоящем документе предусмотрена библиотека полинуклеотидов, в которой раковый антиген представляет собой MAGE A3, MAGE A12, MAGE A2, MAGE A6, NY-ESO-1 или CEA. Дополнительно в настоящем документе предусмотрена библиотека полинуклеотидов, в

которой каждый полинуклеотид содержит множество вариаций в предварительно выбранных кодонах для аминокислотных остатков. Дополнительно в настоящем документе предусмотрена библиотека полинуклеотидов, в которой множество вариаций в предварительно выбранных кодонах для аминокислотных остатков содержит вплоть до 100 остатков. Дополнительно в настоящем документе предусмотрена библиотека полинуклеотидов, в которой множество вариаций в предварительно выбранных кодонах для аминокислотных остатков содержит вплоть до 30 остатков. Дополнительно в настоящем документе предусмотрена библиотека полинуклеотидов, в которой множество вариаций в предварительно выбранных кодонах для аминокислотных остатков содержит вплоть до 5 остатков.

[0010] В настоящем документе предусмотрена библиотека нуклеиновых кислот, причем библиотека нуклеиновых кислот содержит по меньшей мере 10000 вариантных нуклеиновых кислот, причем каждая вариантная нуклеиновая кислота составляет от около 500 до около 1000 оснований в длину, причем каждая вариантная нуклеиновая кислота кодирует вариант эталонной последовательности, которая кодирует последовательность экзона для белка Т-клеточного рецептора или его фрагмента, и причем эталонная последовательность происходит от субъекта, который характеризуется наличием рака. Дополнительно в настоящем документе предусмотрена библиотека нуклеиновых кислот, причем библиотека содержит по меньшей мере около 1000000 вариантных нуклеиновых кислот, причем каждая вариантная нуклеиновая кислота составляет от около 500 до около 1000 оснований в длину, причем каждая вариантная нуклеиновая кислота кодирует вариант эталонной последовательности, которая кодирует последовательность экзона для белка Т-клеточного рецептора или его фрагмента, причем эталонную последовательность белка Т-клеточного рецептора или его фрагмента выбирают на основе сравнения последовательности гена Т-клеточного рецептора от множества субъектов в отношении последовательности нуклеиновой кислоты, которая является общей среди субъектов, и причем по меньшей мере у части субъектов диагностирован рак. Дополнительно в настоящем документе предусмотрена библиотека нуклеиновых кислот, причем рак представляет собой солидный рак или рак кроветворной системы. Дополнительно в настоящем документе предусмотрена библиотека нуклеиновых кислот, причем библиотека содержит около 10000000 вариантных нуклеиновых кислот. Дополнительно в настоящем документе предусмотрена библиотека нуклеиновых кислот, в которой длина каждой вариантной нуклеиновой кислоты составляет от около 600 до около 900 оснований. Дополнительно в настоящем документе предусмотрена библиотека нуклеиновых кислот, в которой каждая вариантная нуклеиновая кислота находится в

последовательности вектора. Дополнительно в настоящем документе предусмотрена библиотека нуклеиновых кислот, в которой последовательность вектора представляет собой последовательность вирусного вектора.

[0011] В настоящем документе предусмотрена библиотека белков, содержащая белки, кодируемые библиотекой нуклеиновых кислот, описанной в настоящем документе. В настоящем документе предусмотрена клеточная библиотека, содержащая библиотеку нуклеиновых кислот, описанную в настоящем документе.

[0012] В настоящем документе предусмотрен способ синтеза библиотеки нуклеиновых кислот, включающий: (а) предоставление первого набора предварительно выбранных полинуклеотидных последовательностей, кодирующих по меньшей мере 3000 вариантных последовательностей гена TCR или фрагмента гена TCR, причем каждая вариантная последовательность содержит по меньшей мере одну вариацию в предварительно выбранном кодоне для аминокислотного остатка в области контакта с антигеном; (b) синтез первого набора предварительно выбранных полинуклеотидных последовательностей; и (с) скрининг белков, кодируемых первым набором полинуклеотидных последовательностей, в отношении активности связывания. Дополнительно в настоящем документе предусмотрен способ синтеза библиотеки нуклеиновых кислот, при котором по меньшей мере одна вариация находится в кодирующей вариабельный домен области гена TCR или фрагмента гена TCR. Дополнительно в настоящем документе предусмотрен способ синтеза библиотеки нуклеиновых кислот, при котором вариабельный домен представляет собой вариабельный домен TCR-альфа, TCR-бета, TCR-гамма или TCR-дельта. Дополнительно в настоящем документе предусмотрен способ синтеза библиотеки нуклеиновых кислот, при котором по меньшей мере одна вариация находится в кодирующей константный домен области гена TCR или фрагмента гена TCR. Дополнительно в настоящем документе предусмотрен способ синтеза библиотеки нуклеиновых кислот, при котором константный домен представляет собой константный домен TCR-альфа, TCR-бета, TCR-гамма или TCR-дельта. Дополнительно в настоящем документе предусмотрен способ синтеза библиотеки нуклеиновых кислот, при котором антиген представляет собой раковый антиген. Дополнительно в настоящем документе предусмотрен способ синтеза библиотеки нуклеиновых кислот, при котором раковый антиген представляет собой MAGE A3, MAGE A12, MAGE A2, MAGE A6, NY-ESO-1 или CEA. Дополнительно в настоящем документе предусмотрен способ синтеза библиотеки нуклеиновых кислот, при котором каждая вариантная последовательность содержит вплоть до 100 вариаций в предварительно выбранных кодонах для аминокислотных остатков в области контакта с антигеном.

Дополнительно в настоящем документе предусмотрен способ синтеза библиотеки нуклеиновых кислот, при котором каждая последовательность вариантов содержит вплоть до 30 вариаций в предварительно выбранных кодонах для аминокислотных остатков в области контакта с антигеном. Дополнительно в настоящем документе предусмотрен способ синтеза библиотеки нуклеиновых кислот, дополнительно включающий: (а) предоставление второго набора предварительно выбранных полинуклеотидных последовательностей, кодирующих по меньшей мере одну вариантную последовательности гена TCR или фрагмента гена TCR, где каждая вариантная последовательность содержит по меньшей мере одну вариацию в предварительно выбранном кодоне для аминокислотного остатка в гене TCR или фрагменте гена TCR в кодирующей константный домен области; (b) синтез второго набора предварительно выбранных полинуклеотидных последовательностей; и (с) скрининг белков, кодируемых вторым набором полинуклеотидных последовательностей, в отношении второй активности. Дополнительно в настоящем документе предусмотрен способ синтеза библиотеки нуклеиновых кислот, при котором вторая активность представляет собой уничтожение раковых клеток, экспрессию белка или стабильность белка.

Включение посредством ссылки

[0013] Все публикации, патенты и заявки на выдачу патента, упомянутые в настоящем описании, включены в настоящий документ посредством ссылки в той же степени, как если бы было указано, что каждая отдельная публикация, патент или заявка на выдачу патента специально и отдельно включена в качестве ссылки.

Краткое описание графических материалов

[0014] На фигурах **1A-1D** изображен технологический процесс синтеза вариантных биологических молекул, включающий в себя стадию ПЦР-мутагенеза.

[0015] На фигурах **2A-2D** изображен технологический процесс создания нуклеиновой кислоты, содержащей последовательность нуклеиновой кислоты, которая отличается от эталонной последовательности нуклеиновой кислоты в одном предварительно определенном сайте кодона.

[0016] На фигурах **3A-3F** изображен альтернативный технологический процесс создания набора вариантов нуклеиновой кислоты из матричной нуклеиновой кислоты, причем каждый вариант содержит различную последовательность нуклеиновой кислоты в одном положении кодона. Каждый вариант нуклеиновой кислоты кодирует

различную аминокислоту в своем одном положении кодона, различные кодоны представлены X, Y и Z.

[0017] На фигурах **4A-4E** изображена эталонная аминокислотная последовательность (фиг. **4A**), содержащая ряд аминокислот, причем каждый остаток обозначен одним кружком, и варианты аминокислотные последовательности (фиг. **4B**, **4C**, **4D** и **4E**), полученные с использованием способов, описанных в настоящем документе. Эталонная аминокислотная последовательность и варианты последовательности кодируются нуклеиновыми кислотами и их вариантами, созданными с помощью описанных в настоящем документе способов.

[0018] На фигурах **5A-5B** изображена эталонная аминокислотная последовательность (фиг. **5A**, SEQ ID NO: 24) и библиотека вариантных аминокислотных последовательностей (фиг. **5B**, SEQ ID NO 25-31, соответственно, в порядке встречаемости), каждый вариант содержит один вариант остатка (обозначен "X"). Эталонная аминокислотная последовательность и варианты последовательности кодируются нуклеиновыми кислотами и их вариантами, созданными с помощью описанных в настоящем документе способов.

[0019] На фигурах **6A-6B** изображена эталонная аминокислотная последовательность (фиг. **6A**) и библиотека вариантных аминокислотных последовательностей (фиг. **6B**), причем каждый вариант содержит два сайта вариантов с одним положением. Каждый вариант обозначен кружками с разной штриховкой. Эталонная аминокислотная последовательность и варианты последовательности кодируются нуклеиновыми кислотами и их вариантами, созданными с помощью описанных в настоящем документе способов.

[0020] На фигурах **7A-7B** изображены эталонная аминокислотная последовательность (фиг. **7A**) и библиотека вариантных аминокислотных последовательностей (фиг. **7B**), причем каждый вариант содержит участок аминокислот (обозначен рамкой вокруг кружков), причем каждый участок содержит три сайта вариантов положения (кодирующих гистидин), отличающихся по последовательности от эталонной аминокислотной последовательности. Эталонная аминокислотная последовательность и варианты последовательности кодируются нуклеиновыми кислотами и их вариантами, созданными с помощью описанных в настоящем документе способов.

[0021] На фигурах **8A-8B** изображена эталонная аминокислотная последовательность (фиг. **8A**) и библиотека вариантных аминокислотных последовательностей (фиг. **8B**), причем каждый вариант содержит два участка аминокислотных последовательностей (обозначены рамкой вокруг кружков), причем

каждый участок содержит один сайт вариантов с одним положением (проиллюстрировано кружками со штриховкой), отличающийся по последовательности от эталонной аминокислотной последовательности. Эталонная аминокислотная последовательность и вариантные последовательности кодируются нуклеиновыми кислотами и их вариантами, созданными с помощью описанных в настоящем документе способов.

[0022] На фигурах **9A-9B** изображена эталонная аминокислотная последовательность (фиг. **9A**) и библиотека вариантных аминокислотных последовательностей (фиг. **9B**), причем каждый вариант содержит участок аминокислот (обозначен кружками со штриховкой), причем каждый участок содержит один сайт вариантов с несколькими положениями, отличающихся по последовательности от эталонной аминокислотной последовательности. На этой иллюстрации 5 положений подвергают вариации, где первое положение характеризуется отношением 50/50 K/R; второе положение характеризуется отношением 50/25/25 V/L/S, третье положение характеризуется отношением Y/R/D 50/25/25, четвертое положение характеризуется одинаковым отношением для всех аминокислот, и пятое положение характеризуется отношением 75/25 для G/P. Эталонная аминокислотная последовательность и вариантные последовательности кодируются нуклеиновыми кислотами и их вариантами, созданными с помощью описанных в настоящем документе способов.

[0023] На фигурах **10A-10C** показана структура и функция рецептора Т-клетки (TCR) (фиг. **10A**) и TCR, экспрессируемого на Т-клетке, взаимодействующей с антигенным пептидом на клетке-мишени, когда последовательность пептида-мишени презентируется подходящим главным комплексом гистосовместимости (MHC-1 для цитотоксических Т-клеток) (фиг. **10B**). TCR содержит две цепи: альфа (α) и бета (β) цепи. Обе цепи содержат константную область (c) и переменную область (v), шарнирную область (h), трансмембранную область (tm) и область цитоплазматического хвоста (ct). TCR ассоциирован с комплексом CD3, который содержит три трансмембранные сигнальные молекулы (CD3 ζ , CD3 $\delta\epsilon$ и CD3 $\gamma\epsilon$). Эффективная активация Т-клеток также требует одновременного связывания корецептора Т-клеток (CD8 для цитотоксических Т-клеток). “ss” относится к дисульфидному мостику. На фиг. **10C** проиллюстрирован иллюстративный технологический процесс для конструирования Т-клеток субъекта и введения их обратно субъекту.

[0024] На фигуре **11** изображено иллюстративное количество вариантов, полученных путем чередования секций двух экспрессионных кассет (например, промоторы, открытые рамки считывания и терминаторы) для создания библиотеки вариантов экспрессионных кассет.

[0025] На фигуре 12 представлена схема стадий, демонстрирующая иллюстративный технологический процесс синтеза гена, как раскрыто в настоящем документе.

[0026] На фигуре 13 проиллюстрирован пример вычислительной системы.

[0027] На фигуре 14 представлена блок-схема, иллюстрирующая архитектуру вычислительной системы.

[0028] На фигуре 15 показана схема, на которой продемонстрирована сеть, выполненная с возможностью включать в себя множество вычислительных систем, множество мобильных телефонов и карманных персональных компьютеров, а также сетевое устройство хранения данных (NAS).

[0029] На фигуре 16 показана блок-диаграмма многопроцессорной вычислительной системы с использованием совместного пространства памяти виртуальных адресов.

[0030] На фигуре 17 представлен график BioAnalyzer для продуктов реакции ПЦР, разделенных с помощью гель-электрофореза.

[0031] На фигуре 18 изображена электрофореграмма, показывающая 96 наборов продуктов ПЦР, причем каждый набор продуктов ПЦР отличается по последовательности от матричной нуклеиновой кислоты дикого типа в одном положении кодона, где одно положение кодона каждого набора находится в другом сайте в матричной последовательности нуклеиновой кислоты дикого типа. Каждый набор продуктов ПЦР содержит 19 вариантных нуклеиновых кислот, причем каждый вариант кодирует различную аминокислоту в своем одном положении кодона.

Подробное описание изобретения

[0032] В настоящем раскрытии используют, если не указано иное, общепринятые техники молекулярной биологии, которые находятся в пределах квалификации специалиста в настоящей области техники. Если не указано иное, все технические и научные термины, используемые в настоящем документе, имеют то же значение, которое обычно понимают специалисты в настоящей области техники.

[0033] Определения

[0034] Во всем настоящем раскрытии числовые признаки представлены в формате диапазона. Следует понимать, что описание в формате диапазона дано просто для удобства и краткости и не должно рассматриваться как негибкое ограничение объема каких-либо вариантов осуществления. Соответственно, следует считать, что описание диапазона конкретно раскрывает все возможные поддиапазоны, а также отдельные числовые значения

в пределах этого диапазона до десятых нижнего предела, если контекст явно не предписывает иное. Например, описание диапазона, такого как от 1 до 6, следует рассматривать как специально раскрывающее поддиапазоны, такие как от 1 до 3, от 1 до 4, от 1 до 5, от 2 до 4, от 2 до 6, от 3 до 6 и т. д., а также отдельные значения в этом диапазоне, например, 1,1, 2, 2,3, 5 и 5,9. Это применимо независимо от широты диапазона. Верхний и нижний пределы этих промежуточных диапазонов могут независимо включаться в меньшие диапазоны и также охватываются настоящим изобретением, с учетом любого специально исключенного предела в указанном диапазоне. Если указанный диапазон включает в себя один или оба из пределов, диапазоны, исключающие один или оба из этих включенных пределов, также включены в настоящее изобретение, если контекст явно не предписывает иное.

[0035] Используемая в настоящем документе терминология предназначена только для описания конкретных вариантов осуществления и не предназначена для ограничения какого-либо варианта осуществления. Подразумевается, что используемые в настоящем документе формы единственного числа также включают в себя формы множественного числа, если контекст явно не указывает на иное. Кроме того, следует понимать, что термины “содержит” и/или “содержащий” при использовании в настоящем описании определяют наличие заявленных признаков, целых чисел, стадий, операций, элементов и/или компонентов, но не исключают присутствия или добавление одного или нескольких других признаков, целых чисел, стадий, операций, элементов, компонентов и/или их групп. Используемый в настоящем документе термин “и/или” включает в себя любую и все комбинации одного или нескольких связанных перечисленных элементов.

[0036] Если специально не указано или не очевидно из контекста, используемый в настоящем документе термин “приблизительно” в отношении числа или диапазона чисел понимают как означающее указанное число и числа $\pm 10\%$ от него, или на 10% ниже указанного нижнего предела и на 10% выше указанного верхнего предела для значений, указанных для диапазона.

[0037] Используемые в настоящем документе термины “предварительно выбранная последовательность”, “предварительно установленная последовательность” или “предварительно определенная последовательность” используют взаимозаменяемо. Термины означают, что последовательность полимера известна и выбрана перед синтезом или сборкой полимера. В частности, различные аспекты настоящего изобретения описаны в настоящем документе прежде всего в отношении получения молекул нуклеиновых кислот, причем последовательность олигонуклеотида или полинуклеотида известна и выбрана до синтеза или сборки молекул нуклеиновой кислоты.

[0038] В настоящем документе предусмотрены способы и композиции для получения синтетических (т.е. *de novo* синтезированных или химически синтезированных) полинуклеотидов. Термины “олигонуклеиновая кислота”, “олигонуклеотид”, “олиго” и “полинуклеотид” определены как синонимы во всем настоящем описании. Библиотеки синтезированных полинуклеотидов, описанные в настоящем документе, могут содержать множество полинуклеотидов, совместно кодирующих один или несколько генов или фрагментов генов. В некоторых случаях библиотека полинуклеотидов содержит кодирующие или не кодирующие последовательности. В некоторых случаях библиотека полинуклеотидов кодирует множество последовательностей кДНК. Последовательности эталонных генов, на которых основаны последовательности кДНК, могут содержать интроны, тогда как последовательности кДНК исключают интроны. Полинуклеотиды, описанные в настоящем документе, могут кодировать гены или фрагменты генов организма. Иллюстративные организмы включают в себя без ограничения организмы-прокариоты (например, бактерии) и организмы-эукариоты (например, мыши, кролики, люди и приматы, отличные от человека). В некоторых случаях библиотека полинуклеотидов содержит один или несколько полинуклеотидов, каждый из которых представляет собой один или несколько полинуклеотидов, кодирующих последовательности для нескольких экзонов. Каждый полинуклеотид в библиотеке, описанной в настоящем документе, может кодировать разную последовательность, т.е. неидентичную последовательность. В некоторых случаях каждый полинуклеотид в библиотеке, описанной в настоящем документе, содержит по меньшей мере одну часть, которая является комплементарной последовательности другого полинуклеотида в пределах библиотеки. Полинуклеотидные последовательности, описанные в настоящем документе, могут, если не указано иное, включать ДНК или РНК.

[0039] В настоящем документе предусмотрены способы и композиции для получения синтетических (т.е. *de novo* синтезированных) генов. Библиотеки, содержащие синтетические гены, можно сконструировать различными способами, описанными более подробно в другом месте настоящего документа, такими как РСА (полимеразная цепная сборка), не относящиеся к РСА способы сборки генов или иерархическая сборка генов, объединение (“сшивание”) двух или больше двухцепочечных полинуклеотидов для получения более крупных звеньев ДНК (т.е., шасси). Библиотеки больших конструкторов могут включать в себя полинуклеотиды, которые составляют по меньшей мере 1, 1,5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400, 500 т.п.н. в длину или больше. Большие конструкторы могут быть ограничены независимо выбранным верхним пределом, составляющим около 5000, 10000, 20000 или 50000 пар

оснований. Синтез любого числа нуклеотидных последовательностей, кодирующих полипептидный сегмент, включая в себя последовательности, кодирующие нерибосомные пептиды (NRP), последовательности, кодирующие модули нерибосомной пептидсинтетазы (NRPS) и синтетические варианты, полипептидные сегменты других модульных белков, таких как антитела, полипептидные сегменты из других семейств белков, включая в себя не кодирующую ДНК или РНК, такие как регуляторные последовательности, например, промоторы, факторы транскрипции, энхансеры, киРНК, кшРНК, РНКи, микроРНК, малая ядрышковая РНК, происходящая из микроРНК, или любое функциональное или структурное звено ДНК или РНК, представляющее интерес. Ниже приведены неограничивающие примеры полинуклеотидов: кодирующие или не кодирующие области гена или фрагмента гена, межгенная ДНК, локусы (локус), определенные из анализа групп сцепления, экзоны, интроны, матричная РНК (мРНК), транспортная РНК, рибосомальная РНК короткая интерферирующая РНК (киРНК), короткая шпилечная РНК (кшРНК), микроРНК, малая ядрышковая РНК, рибозимы, комплементарная ДНК (кДНК), которая представляет собой ДНК-репрезентацию мРНК, обычно получаемую путем обратной транскрипции матричной РНК (мРНК) или путем амплификации; молекулы ДНК, полученные синтетическим путем или путем амплификации, геномная ДНК, рекомбинантные полинуклеотиды, разветвленные полинуклеотиды, плазмиды, векторы, выделенная ДНК любой последовательности, выделенная РНК любой последовательности, нуклеиновокислотные зонды и праймеры. кДНК, кодирующая ген или фрагмент гена, на которые ссылаются в настоящем документе, может содержать по меньшей мере одну область, кодирующую экзонную(ые) последовательность(и) без промежуточной интронной последовательности, встречающейся в соответствующей геномной последовательности. Альтернативно, соответствующая геномная последовательность для кДНК в первую очередь может не содержать интронную последовательность.

[0040] Конструирование вариаций в иммунологическом синапсе

[0041] В настоящем документе предусмотрены способы синтеза библиотеки вариантных нуклеиновых кислот, в которой каждая вариантная нуклеиновая кислота кодирует последовательность, которую подвергают вариации по сравнению с эталонной последовательностью белка Т-клеточного рецептора (TCR). Подвергнутая вариации последовательность может представлять собой последовательность нуклеиновой кислоты, которая кодирует фрагмент или полную последовательность белка TCR. В некоторых случаях подвергнутая вариации последовательность представляет собой пептид, служащий антигеном в иммунологическом синапсе. Полученная библиотека нуклеиновых кислот может представлять собой библиотеку олигонуклеотидов, библиотеку фрагментов генов

или библиотеку генов. В некоторых случаях библиотека нуклеиновой кислоты экспрессируется в клетках для создания библиотеки вариантов белка.

[0042] Ссылаясь на фиг. **10A-10C** белок TCR представляет собой гетеродимерный комплекс, содержащий α и β цепь, каждая из которых содержит внеклеточную, трансмембранную и внутриклеточную область. Обе цепи связаны дисульфидной связью, причем каждый рецептор обеспечивает сайт связывания антигена. Антиген, как правило, презентирован молекулой главного комплекса гистосовместимости (МНС). Вспомогательные молекулы, такие как CD4 для МНС класса II и CD8 для МНС класса I, также экспрессируются Т-клетками и участвуют в комплексе. Кроме того, комплекс CD3 также участвует и обеспечивает роль передачи сигнала в Т-клетке.

[0043] В настоящем документе предусмотрены способы увеличения уничтожения опухолевых клеток, опосредованного Т-клетками, путем конструирования Т-клеток. Например, образцы тканей можно взять у субъектов для получения информации о последовательности TCR, присутствующих в Т-клетках субъекта. Образцы ткани можно получить путем сбора крови, мононуклеарных клеток периферической крови или сбора выделенных инфильтрирующих опухоль лимфоцитов из иссеченной опухолевой ткани. Субъект может характеризоваться наличием рака или не характеризоваться наличием рака.

[0044] При получении образцов от множества субъектов их подвергают скринингу в отношении наличия профиля общего донора HLA. Например, HLA может представлять собой аллель HLA-A, HLA-B или HLA-C. Иллюстративные аллели приведены в таблице 1 ниже.

Таблица 1. Иллюстративные аллели

аллель HLA-A:	аллель HLA-C:	аллель HLA-B:
010101g	0102	0702
02010101g	0202/0210	0705g
0205	0302	0714
0206	0303	0801g
0217	030401/030403	1301
03010101g	030402	1302
0302	0332 (V)	1401
0325 (V)c	04010101g	1402
1101	0403	15010101g
2301g	0501g	1503
24020101g	0602	1506

240301g	070101g	1510
2501	0702	1517
2601	070401g	1518
29010101g	0713	1534
2902	0732N (V)	180101g
3001	0801	1803
3002	0802	2702
3004	0803	270502g
3101	1202	2707
3201	1203	350101g
3208	1402	3502
3301	1502	3503
3303	1505	3508
3402	1601	3512
3601	1602	3701
6601	1604	3801
680101	1701g	3901
680102g		3906
6802		400101g
7401g		400201g

[0045] В настоящем документе описаны различные способы идентификации областей нуклеиновой кислоты, представляющих интерес для отбора для создания библиотеки вариантов, характеризующейся предварительно выбранным разнообразием в пределах этой области. Например, Т-клетки можно получить от множества субъектов. По меньшей мере часть субъектов может характеризоваться наличием рака, или ни у одного из субъектов может быть не диагностирован рак. Затем Т-клетки секвенируют, и среди субъектов идентифицируют общую последовательность в последовательности гена, кодирующей белок TCR. В некоторых случаях 1, 2, 3, 4, 5 или более отдельных участков общих последовательностей идентифицируют в гене, кодирующем TCR, среди субъектов. Общая(ие) последовательность(и), как правило, представляет(ют) собой последовательность экзона в последовательности гена TCR. Интронную последовательность также могут проанализировать среди субъектов. Альтернативно, Т-клетки получают от одного субъекта до и после диагностирования рака. Генетический профиль генов TCR в Т-клетках сравнивают аналогичным образом, чтобы

идентифицировать общие последовательности для вариации. Альтернативно, Т-клетки получают от одного субъекта, у которого диагностирован рак или отсутствие рака, и генетический профиль генов TCR в Т-клетках сравнивают с базой данных последовательностей. База данных последовательностей может содержать генетические профили TCR от субъектов с аналогично картированным HLA.

[0046] Затем общую последовательность подвергают вариации с использованием способов, описанных в настоящем документе, включая в себя (i) синтез de novo библиотеки вариантов нуклеиновых кислот праймеров с последующим ПЦР-мутагенезом, (ii) синтез de novo полной новой вариантной версии общей последовательности или (iii) синтез de novo нескольких фрагментов вариантной версии общей последовательности для отжига и сборки, включая в себя полимеразную цепную сборку.

[0047] Общая(ие) последовательность(и), описанная(ые) в настоящем документе, может(могут) составлять около 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 или больше оснований в длину. В некоторых случаях общая(ие) последовательность(и) составляет(ют) по меньшей мере около 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 или больше оснований в длину. В некоторых случаях общая последовательность составляет около 100 - около 1000, около 100 - около 2000 или около 400 - около 900 оснований в длину. В некоторых случаях идентифицируют несколько общих последовательностей. В некоторых случаях для вариации выбирают около 100, 250, 500, 1000 или больше общих кодирующих TCR генных последовательностей среди субъектов. Вариация включает в себя конструирование по меньшей мере около 500, 1000, 2000, 5000, 10000, 20000, 50000, 100000 или больше вариантов для каждой из выбранных общих последовательностей TCR. Всего библиотека вариантов может давать в результате 10^6 , 10^8 , 10^{10} , 10^{11} , 10^{12} , 10^{13} или больше различных нуклеиновых кислот.

[0048] Описанные в настоящем документе библиотеки вариантных последовательностей TCR могут содержать общую последовательность из любой области нуклеотидной последовательности TCR, кодирующей ген или фрагмент гена. Иллюстративные области нуклеотидной последовательности TCR для вариации гена или фрагмента гена включают в себя без ограничения: переменная область α цепи, переменная область β цепи, константная область α цепи, константная область β цепи, шарнирная область, трансмембранная область или область цитоплазматического хвоста. В некоторых случаях библиотека вариантных последовательностей TCR содержит общую последовательность для вариации, причем презентируется молекулой главного комплекса гистосовместимости нуклеиновые кислоты, кодирующие остатки за пределами

определяющей комплементарность области переменных доменов. В некоторых случаях библиотека вариантных последовательностей TCR содержит общую последовательность для вариации, причем для вариации выбирают нуклеиновые кислоты, кодирующие остатки в пределах определяющей комплементарности области переменных доменов.

[0049] Каждая из последовательностей, кодирующих α и бета-цепи TCR, содержит отдельные переменные (V), дополнительные (D), соединительные (J) генные сегменты и константные (C) гены. Локус α TCR (хромосома 14) состоит из 70-80 генных сегментов $V\alpha$, каждому из которых предшествует экзон, кодирующий лидерную последовательность (L). Кластер из 61 генного сегмента $J\alpha$ расположен на значительном расстоянии от генных сегментов $V\alpha$. За генными сегментами $J\alpha$ следует один ген C, который содержит отдельные экзоны для константного и шарнирного доменов и один экзон, кодирующий трансмембранную и цитоплазматическую области (не показано). Локус TCR β (хромосома 7) характеризуется другой организацией: кластер из 52 функциональных генных сегментов $V\beta$ расположен на расстоянии от двух отдельных кластеров, каждый из которых содержит один генный сегмент D, вместе с шестью или семью генными сегментами J и одним геном C. Каждый ген C TCR β содержит отдельные экзоны, кодирующие константный домен, шарнир, трансмембранную область и цитоплазматическую область (не показано). Локус TCR α прерывается между генными сегментами J и V другим локусом Т-клеточного рецептора - локусом TCR δ . Согласно некоторым вариантам осуществления общая последовательность, выбранная для вариации, включает в себя один или несколько из следующего: генный сегмент $V\alpha$, генный сегмент $J\alpha$, генный сегмент $C\alpha$, генный сегмент $V\beta$, генный сегмент $J\beta$ или генный сегмент $C\beta$.

[0050] Иллюстративные переменные гены α -цепи TCR включают в себя без ограничения TRAV1-1 (TRAV11, TCRAV1S1, TCRAV7S1), TRAV1-2, TRAV2, TRAV3, TRAV4 (TCRAV20S1, TCRAV4S1), TRAV5, TRAV6 (TCRAV5S1), TRAV7, TRAV8-1, TRAV8-2, TRAV8-3, TRAV8-4, TRAV8-5 (TRAV85, TCRAV8S5), TRAV8-6, TRAV8-7, TRAV9-1, TRAV9-2, TRAV10 (TCRAV10S1, TCRAV24S1), TRAV11, TRAV12-1, TRAV12-2, TRAV12-3 (TRAV123, TCRAV2S2, TCRAV12S3), TRAV13-1, TRAV13-2, TRAV14DV4 (TRAV14/DV4, TCRAV6S1-hDV104S1, hADV14S1), TRAV15 (TCRAV15S1), TRAV16 (TCRAV16S1, TCRAV9S1), TRAV17 (TCRAV17S1, TCRAV3S1), TRAV18 (TCRAV18S1), TRAV19 (TCRAV12S1, TCRAV19S1), TRAV20, TRAV21, TRAV22 (TCRAV13S1, TCRAV22S1), TRAV23DV6 (TRAV23/DV6, TCRAV17S1), TRAV24, TRAV25 (TCRAV25S1, TCRAV32S1), TRAV26-1, TRAV26-2 (TRAV262, TCRAV4S1, TCRAV26S2), TRAV27, TRAV28, TRAV29DV5 (TRAV29/DV5), TRAV30, TRAV31 (TCRAV31S1), TRAV32, TRAV33 (C14orf12), TRAV34 (TCRAV26S1, TCRAV34S1), TRAV35

(TCRAV25S1, TCRAV35S1), TRAV36DV7 (TRAV36/DV7), TRAV37, TRAV38-1 (TRAV381, TCRAV14S2, TCRAV38S1), TRAV38-2DV8 (TRAV38-2/DV8, TRAV382DV8, TCRAV14S1, hADV38S2), TRAV39 (TCRAV27S1, TCRAV39S1), TRAV40 (TCRAV31S1, TCRAV40S1) и TRAV41 (TCRAV19S1, TCRAV41S1). Иллюстративные генные последовательности для вариабельных генов α цепи TCR перечислены в таблице 2.

[0051] Иллюстративные вариабельные гены бета-цепи TCR включают в себя без ограничения TRBV1 (TCRBV27S1P, TCRBV1S1P) TRBV2 (TCRBV22S1A2N1T, TCRBV2S1), TRBV3-1 (TRBV31, TCRBV3S1, TCRBV9S1A1T), TRBV3-2 (TRBV32, TCRBV3S2, TCRBV9S2A2PT), TRBV4-1 (TRBV41, TCRBV4S1, TCRBV7S1A1N2T, BV07S1J2.7), TRBV4-2 (TRBV42, TCRBV4S2, TCRBV7S3A2T), TRBV4-3 (TRBV43, TCRBV4S3, TCRBV7S2A1N4T), TRBV5-1 (TRBV51, TCRBV5S1, TCRBV5S1A1T), TRBV5-2 (TRBV52, TCRBV31S1, TCRBV5S2P), TRBV5-3 (TRBV53, TCRBV5S3, TCRBV5S5P), TRBV5-4 (TRBV54, TCRBV5S4, TCRBV5S6A3N2T), TRBV5-5 (TRBV55, TCRBV5S3A2T, TCRBV5S5), TRBV5-6 (TRBV56, TCRBV5S2, TCRBV5S6), TRBV5-7 (TRBV57, TCRBV5S7, TCRBV5S7P), TRBV5-8 (TRBV58, TCRBV5S4A2T, TCRBV5S8), TRBV6-1 (TRBV61, TCRBV13S3, TCRBV6S1), TRBV6-2 (TRBV62, TCRBV13S2A1T, TCRBV6S2), TRBV6-3 (TRBV63, TCRBV13S9/13S2A1T, TCRBV6S3), TRBV6-4 (TRBV64, TCRBV13S5, TCRBV6S4), TRBV6-5 (TRBV65, TCRBV13S1, TCRBV6S5), TRBV6-6 (TRBV66, TCRBV13S6A2T, TCRBV6S6), TRBV6-7 (TRBV67, TCRBV13S8P, TCRBV6S7), TRBV6-8 (TRBV68, TCRBV13S7P, TCRBV6S8), TRBV6-9 (TRBV69, TCRBV13S4, TCRBV6S9), TRBV7-1 (TRBV71, TCRBV6S7P, TCRBV7S1), TRBV7-2 (TRBV72, TCRBV6S5A1N1, TCRBV7S2), TRBV7-3 (TRBV73, TCRBV6S1A1N1, TCRBV7S3), TRBV7-4 (TRBV74, TCRBV6S8A2T, TCRBV7S4), TRBV7-5 (TRBV75, TCRBV6S9P, TCRBV7S5), TRBV7-6 (TRBV76, TCRBV6S3A1N1T, TCRBV7S6), TRBV7-7 (TRBV77, TCRBV6S6A2T, TCRBV7S7), TRBV7-8 (TRBV78, TCRBV6S2A1N1T, TCRBV7S8), TRBV7-9 (TRBV79, TCRBV6S4A1, TCRBV7S9), TRBV8-1 (TRBV81, TCRBV30S1P, TCRBV8S1P), TRBV8-2 (TRBV82, TCRBV32S1P), TRBV9 (TCRBV1S1A1N1, TCRBV9S1), TRBV10-1 (TRBV101, TCRBV10S1, TCRBV12S2A1T, TCRBV12S2), TRBV10-2 (TRBV102, TCRBV10S2, TCRBV12S3), TRBV10-3 (TRBV103, TCRBV10S3, TCRBV12S1A1N2), TRBV11-1 (TRBV111, TCRBV11S1, TCRBV21S1), TRBV11-2 (TRBV112, TCRBV11S2, TCRBV21S3A2N2T), TRBV11-3 (TRBV113, TCRBV11S3, TCRBV21S2A2), TRBV12-1 (TRBV121, TCRBV12S1, TCRBV8S4P), TRBV12-2 (TRBV122, TCRBV12S2, TCRBV8S5P), TRBV12-3 (TRBV123, TCRBV12S3, TCRBV8S1), TRBV12-4 (TRBV124, TCRBV12S4, TCRBV8S2A1T), TRBV12-5 (TRBV125, TCRBV12S5, TCRBV8S3), TRBV13 (TCRBV13S1, TCRBV23S1A2T), TRBV14 (TCRBV14S1, TCRBV16S1A1N1), TRBV15 (TCRBV15S1,

TCRBV24S1A3T), TRBV16 (TCRBV16S1, TCRBV25S1A2PT), TRBV17 (TCRBV17S1, TCRBV26S1P), TRBV18 (TCRBV18S1), TRBV19 (TCRBV17S1A1T, TCRBV19S1), TRBV20-1 (TRBV201, TCRBV20S1, TCRBV2S1), TRBV21-1 (TRBV211, TCRBV10S1P, TCRBV21S1), TRBV22-1 (TCRBV22S1, TCRBV29S1P), TRBV23-1 (TRBV231, TCRBV19S1P, TCRBV23S1), TRBV24-1 (TRBV241, TCRBV15S1, TCRBV24S1), TRBV25-1 (TRBV251, TCRBV11S1A1T, TCRBV25S1), TRBV26 (TCRBV26S1, TCRBV28S1P), TRBV27 (TCRBV14S1, TCRBV27S1), TRBV28 (TCRBV28S1, TCRBV3S1), TRBV29-1 (TRBV291, TCRBV29S1, TCRBV4S1A1T) и TRBV30 (TCRBV20S1A1N2, TCRBV30S1).

Иллюстративные генные последовательности для переменных генов бета-цепи TCR приведены в таблице 2.

Таблица 2. Варибельные последовательности альфа-цепи Т-клеточного рецептора и варибельные последовательности бета-цепи Т-клеточного рецептора

SEQ ID NO	Название гена	Регистр. номер	Последовательность
32	TRAV1-1	NC_000014.9	ATGTGGGGAGCTTTCCTTCTCTATGTTTCCATGAAGATG GGAGGTGAGTCTCAATCTAATAGTAAATGCTGCTAGGA ATTTTCAAAACAATTTCTTTTCAGCTAAATTATTGCAA TTTTGACATTTGTAATGAGAGTATTTCTGAATATGCAT TTTCCTAACGTGGTGCTAATTGTCCTCCTGTTACTATTGC TGCTGCTGTTACTGCAACCATTTATTTTCAGTCTAAGAAA TTCTCCCATCAATGGCAGTTCTTTTGTGACCACATGGAA GCATCATTTAAAAAATTATTCCAATAGTTTTTGGAGGAA ACATCATTTTTTAATAATGATGGGGCTTCTGGGGGTGCTG CCCTAGTAACAATCATGTATCTTGTCTATAGGCACTGCAG GACAAAGCCTTGAGCAGCCCTCTGAAGTGACAGCTGTG GAAGGAGCCATTGTCCAGATAAACTGCACGTACCAGAC ATCTGGGTTTTATGGGCTGTCCTGGTACCAGCAACATGA TGGCGGAGCACCCACATTTCTTTCTTACAATGCTCTGGA TGGTTTGGAGGAGACAGGTCGTTTTTCTTCATTCTTAG TCGCTCTGATAGTTATGGTTACCTCCTTCTACAGGAGCT CCAGATGAAAGACTCTGCCTCTTACTTCTGCGCTGTGAG AGA
33	TRAV4	NG_001332.3	ATGAGGCAAGTGGCGAGAGTGATCGTGTTCTGACCCT GAGTGAGTTATTTTGGGATGAAGAGGAATGGGATCTGG GCCTGATGATGCTGGAAAGGAATCTGGAACCTTGCCTGC TAGCAGTTGCTCTTTATCCAAGATGTAGAGGGATAGCTT CAGGGTTCCATTTTCTCCAGGCAGCTCCTGAGCATTTA TGTGAGATGTGTCTCAGGGCAGCAAAGATTTTCGGGGTT TCTAATCCCTAGGGTCTATCCATGGGAGGGAGATAGTA AACCTGACACTTCTGTGTATGCCTGGGATAAATATTTTT GTACTGAAACATAAATGAGAAGTAAATTGTATGTACAT GTATATGGCAGGGAAGTAGAGCGTGGGGATGGATTTGA CTGTGAGAGGGAACCGTGGGGCTACTGAGTGGGACCGT CTAACTTACCTTGCTTTCAGGATAGGGGCTAGGATTGTG

			TTTACTCCCATAGGTACTTTGAGCCTTGCTAAGACCAC CCAGCCCATCTCCATGGACTCATATGAAGGACAAGAAG TGAACATAACCTGTAGCCACAACAACATTGCTACAAAT GATTATATCACGTGGTACCAACAGTTTCCCAGCCAAGGA CCACGATTTATTATTCAAGGATACAAGACAAAAGTTAC AAACGAAGTGGCCTCCCTGTTTATCCCTGCCGACAGAAA GTCCAGCACTCTGAGCCTGCCCCGGGTTTCCCTGAGCGA CACTGCTGTGTACTACTGCCTCGTGGGTGACA
34	TRAV8-1	NG_0013 32.3	ATGCTCCTGTTGCTCATACCAGTGCTGGGGATGATTTTT GCCCTGAGTGAGTAACATTCTATTATGGTCTCTAGTTCC ACAGAAGTAACTGTTTTCTGATTCAAATCTTAGTAGAAA TACTTTTCATAGACAAGTCTGCACTATTTTCACTGATAC AACATTGATTTTTTTCAGGAGATGCCAGAGCCCAGTCTGT GAGCCAGCATAACCACCACGTAATTCTCTCTGAAGCAG CCTCACTGGAGTTGGGATGCAACTATTCCTATGGTGGAA CTGTTAATCTCTTCTGGTATGTCCAGTACCCTGGTCAAC ACCTTCAGCTTCTCCTCAAGTACTTTTCAGGGGATCCAC TGGTTAAAGGCATCAAGGGCTTTGAGGCTGAATTTATAA AGAGTAAATTCTCCTTTAATCTGAGGAAACCCTCTGTGC AGTGGAGTGACACAGCTGAGTACTTCTGTGCCGTGAAT GC
35	TRBV2	NG_0013 33.2	ATGGATACCTGGCTCGTATGCTGGGCAATTTTTAGTCTC TTGAAAGCAGGTGCGATGCTTAGACTCTAGGAAATTCTTG CTTTGAACTTACCTAAGACAATTCTAAACCATTCTCTTA ATCTTCTTCTTTTCTCACAGGACTCACAGAACCTGAAGT CACCCAGACTCCCAGCCATCAGGTCACACAGATGGGAC AGGAAGTGATCTTGCGCTGTGTCCCCATCTCTAATCACT TATACTTCTATTGGTACAGACAAATCTTGGGGCAGAAAG TCGAGTTTCTGGTTTCCTTTTATAATAATGAAATCTCAG AGAAGTCTGAAATATTCGATGATCAATTCTCAGTTGAAA GGCCTGATGGATCAAATTTCACTCTGAAGATCCGGTCCA CAAAGCTGGAGGACTCAGCCATGTACTTCTGTGCCAGC AGTGAAGC
36	TRBV3-1	NG_0013 33.2	ATGGGCTGCAGGCTCCTCTGCTGTGTGGTCTTCTGCCTC CTCCAAGCAGGTGAGTCCCGGGCCCAGGTGACATGATC CTATTGGAGTCCCTAAGCCTTTTACCATGACAACAACA GCAGGCCGTCTCCTAGGATTTGCCTGAATTCTGCTTCTTT CCTTTGCAGGTCCCTTGGACACAGCTGTTTCCCAGACTC CAAAATACCTGGTCACACAGATGGGAAACGACAAGTCC ATTAAATGTGAACAAAATCTGGGCCATGATACTATGTAT TGGTATAAACAGGACTCTAAGAAATTTCTGAAGATAAT GTTTAGCTACAATAATAAGGAGCTCATTATAAATGAAA CAGTTCCAAATCGCTTCTCACCTAAATCTCCAGACAAAG CTCACTTAAATCTTCACATCAATTCCCTGGAGCTTGGTG ACTCTGCTGTGTATTTCTGTGCCAGCAGCCAAGA
37	TRBV9	NG_0013 33.2	ATGGGCTTCAGGCTCCTCTGCTGTGTGGCCTTTTGTCTCC TGGGAGCAGGTGAGTCCCTGGGCACAACCTTGAAAGTCTC CGATCTTCATTTCTTGTCCCTGAAATGCATGTGGGCCAA CGATGGCTTCAGCAGGAGGCTTTCTTCTGTGCCTTATGG TTAACTTTTGTCTTCTGACACACAGGCCCAGTGGATTCT GGAGTCACACAAACCCCAAAGCACCTGATCACAGCAAC TGGACAGCGAGTGACGCTGAGATGCTCCCCTAGGTCTG

			GAGACCTCTCTGTGTACTGGTACCAACAGAGCCTGGACC AGGGCCTCCAGTTCCTCATTCACTATTATAATGGAGAAG AGAGAGCAAAAGGAAACATTCTTGAACGATTCTCCGCA CAACAGTTCCTGACTTGCCTCTGAACTAAACCTGAGC TCTCTGGAGCTGGGGGACTCAGCTTTGTATTTCTGTGCC AGCAGCGTAG
38	TRBV15	NG_0013 33.2	ATGGGTCCTGGGCTTCTCCACTGGATGGCCCTTTGTCTC CTTGGAACAGGTGAGTACTGGGCAGAAAGGAAATCTTT GAGCAAAGCTATCTTGTCTCAGTCTGCACCTTTCATTC ACAGCAGTAACACTGTTCTCCTTAACTCTGACTCCAAAT TTGTCTTCTTTCTCTACAGGTCATGGGGATGCCATGGTC ATCCAGAACCCAAGATACCAGGTTACCCAGTTTGAAA GCCAGTGACCCTGAGTTGTTCTCAGACTTTGAACCATAA CGTCATGTACTGGTACCAGCAGAAGTCAAGTCAGGCCC CAAAGCTGCTGTTCCACTACTATGACAAAGATTTTAACA ATGAAGCAGACACCCCTGATAACTTCCAATCCAGGAGG CCGAACACTTCTTTCTGCTTTCTTGACATCCGCTCACCAG GCCTGGGGGACACAGCCATGTACCTGTGTGCCACCAGC AGAGA

[0052] Общая последовательность, выбранная для вариации в библиотеке, описанной в настоящем документе, может включать в себя один или несколько из генного сегмента $J\alpha$, генного сегмента $C\alpha$, генного сегмента $J\beta$ или генного сегмента $C\beta$. Иллюстративные гены $J\alpha$ включают в себя без ограничения TRAJ1, TRAJ2, TRAJ3, TRAJ4, TRAJ5, TRAJ6, TRAJ7, TRAJ8, TRAJ9, TRAJ10, TRAJ11, TRAJ12, TRAJ13, TRAJ14, TRAJ15, TRAJ16, TRAJ17, TRAJ18, TRAJ19, TRAJ20, TRAJ21, TRAJ22, TRAJ23, TRAJ24, TRAJ25, TRAJ26, TRAJ27, TRAJ28, TRAJ29, TRAJ30, TRAJ31, TRAJ32, TRAJ33, TRAJ34, TRAJ35, TRAJ36, TRAJ37, TRAJ38, TRAJ39, TRAJ40, TRAJ41, TRAJ42, TRAJ43, TRAJ44, TRAJ45, TRAJ46, TRAJ47, TRAJ48, TRAJ49, TRAJ50, TRAJ51, TRAJ52, TRAJ53, TRAJ54, TRAJ55, TRAJ56, TRAJ57, TRAJ58, TRAJ59, TRAJ60 и TRAJ61. В некоторых случаях ген $C\alpha$ представляет собой TRAC. Иллюстративные гены $J\beta$ включают в себя без ограничения TRBJ1-1 (TRBJ11, TCRBJ1S1), TRBJ1-2 (TRBJ12, TCRBJ1S2), TRBJ1-3 (TRBJ13, TCRBJ1S3), TRBJ1-4 (TRBJ14, TCRBJ1S4), TRBJ1-5 (TRBJ15, TCRBJ1S5), TRBJ1-6 (TRBJ16, TCRBJ1S6), TRBJ2-1 (TRBJ21, TCRBJ2S1), TRBJ2-2 (TRBJ22, TCRBJ2S2), TRBJ2-2P (TRBJ22P), TRBJ2-3 (TRBJ23, TCRBJ2S3), TRBJ2-4 (TRBJ24, TCRBJ2S4), TRBJ2-5 (TRBJ25, TCRBJ2S5), TRBJ2-6 (TRBJ26, TCRBJ2S6) и TRBJ2-7 (TRBJ27, TCRBJ2S7). В некоторых случаях ген $C\beta$ представляет собой TRBC1 (TCRBC1, BV05S1J2.2) или TRBC2.

[0053] Описанная в настоящем документе библиотека вариантов TCR может содержать варианты для последовательностей локуса TCR-дельта (TCR δ). В некоторых случаях последовательность, кодирующая цепь TCR δ , содержит вариабельные (V),

дополнительные (D), соединительные (J) генные сегменты или константный (C) ген. Иллюстративные гены включают в себя без ограничения TRD (TCRDV1), TRDV1 (hDV101S1), TRDV2 (hDV102S1, MGC117421), TRDV3 (hDV103S1), TRDD1, TRDD2, TRDD3, TRDJ1, TRDJ2, TRDJ3, TRDJ4 и TRDC.

[0054] Описанная в настоящем документе библиотека вариантов TCR может содержать варианты для последовательностей локуса TCR-гамма (TCR γ). В некоторых случаях последовательность, кодирующая цепь TCR γ , содержит переменные (V), дополнительные (D), соединительные (J) генные сегменты или константный (C) ген. Иллюстративные гены включают в себя без ограничения TRG, TRGV1 (V1S1P), TRGV2 (VIS2), TRGV3 (V1S3), TRGV4 (V1S4), TRGV5 (V1S5), TRGV5P (V1S5P), TRGV6 (V1S5P, TCRGV5P), TRGV7 (V1S7P), TRGV8 (V1S8), TRGV9 (V2), TRGV10 (V3P), TRGV11 (V4P), TRGVA (V5P), TRGVB (V6P), TRGJ1 (J1), TRGJ2 (J2), TRGJP (JP), TRGJP1 (JP1), TRGJP2 (JP2), TRGC1 (C1), TRGC2 (TRGC2(2X), TRGC2(3X))

[0055] В настоящем документе предусмотрены библиотеки TCR, содержащие нуклеиновые кислоты, кодирующие вариантный TCR, причем каждая нуклеиновая кислота кодирует предварительно определенный вариант по меньшей мере одной предварительно определенной эталонной последовательности нуклеиновой кислоты. В некоторых случаях предварительно определенная эталонная последовательность представляет собой последовательность нуклеиновой кислоты, кодирующую белок, и библиотека вариантов содержит последовательности, кодирующие вариацию по меньшей мере одного кодона так, чтобы множество различных вариантов одного остатка в последующем белке, кодируемом синтезированной нуклеиновой кислотой, создавались с помощью стандартного процесса трансляции. В некоторых случаях библиотека вариантов TCR содержит подвергнутые вариации нуклеиновые кислоты, совместно кодирующие вариации в нескольких положениях. В некоторых случаях библиотека вариантов содержит последовательности, кодирующие вариацию по меньшей мере одного кодона цепи TCR α , TCR β , TCR δ или TCR γ . В некоторых случаях библиотека вариантов содержит последовательности, кодирующие вариацию по меньшей мере одного кодона переменного домена цепи TCR β . В некоторых случаях библиотека вариантов содержит последовательности, кодирующие вариацию по меньшей мере одного кодона переменного домена цепи TCR β . В некоторых случаях библиотека вариантов содержит последовательности, кодирующие вариацию по меньшей мере одного кодона переменного домена цепей TCR α и TCR β . В некоторых случаях библиотека вариантов содержит последовательности, кодирующие вариацию нескольких кодонов переменного домена цепи TCR α , цепи TCR β или их комбинаций. Иллюстративное количество кодонов для вариации включает в себя без ограничения по

меньшей мере или около 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 225, 250, 275, 300 или больше чем 300 кодонов.

[0056] В некоторых случаях библиотеки вариантов TCR содержат вариацию в одном или нескольких экзонах варибельного гена альфа-цепи TCR. В некоторых случаях библиотеки вариантов TCR содержат вариацию в одном или нескольких экзонах варибельного гена бета-цепи TCR. Например, по меньшей мере или около 10, 15, 20, 25, 30, 40, 50, 75, 100, 150, 200 или больше чем 200 нуклеиновых кислот одного или нескольких экзонов варибельного гена α -цепи TCR подвергают вариации. В некоторых случаях по меньшей мере или около 10, 15, 20, 25, 30, 40, 50, 75, 100, 150, 200 или больше чем 200 нуклеиновых кислот одного или нескольких экзонов варибельного гена бета-цепи TCR подвергают вариации. Иллюстративные экзонные последовательности варибельных генов альфа-цепи TCR и бета-цепи TCR показаны в таблице 3.

Таблица 3. Экзонные последовательности варибельных генов альфа-цепи TCR и бета-цепи TCR

SEQ ID NO	Название гена	Регистр. номер	Последовательность
39	TRAV1-1	X04939.1	ATGTGGGGAGCTTTCCTTCTCTATGTTTCCATGAAGATG GGAGGCACTGCAGGACAAAGCCTTGAGCAGCCCTCTGA AGTGACAGCTGTGGAAGGAGCCATTGTCCAGATAAACT GCACGTACCAGACATCTGGGTTTTATGGGCTGTCCTGGT ACCAGCAACATGATGGCGGAGCACCCACATTTCTTTCTT ACAATGGTCTGGATGGTTTGGAGGAGACAGGTCGTTTTT CTTCATTCCTTAGTCGCTCTGATAGTTATGGTTACCTCCT TCTACAGGAGCTCCAGATGAAAGACTCTGCCTCTTACTT CTGCGCTGTTTTTAACCAGGCAGGAAGTCTCTGATCTT TGGGAAGGGAACCACCTTATCAGTGAGTTCCAATATCC AGAACCCT
40	TRAV8-1	X04949.1	ATGCTCCTGTTGCTCATACCAGTGCTGGGGATGATTTTT GCCCTGAGAGATGCCAGAGCCCAGTCTGTGAGCCAGCA TAACCACCACGTAATTCTCTCTGAAGCAGCCTCACTGGA GTTGGGATGCAACTATTCCTATGGTGGAAGTGTAAATCT CTTCTGGTATGTCCAGTACCCTGGTCAACACCTTCAGCT TCTCCTCAAGTACTTTTCAGGGGATCCACTGGTTAAAGG CATCAAGGGCTTTGAGGCTGAATTTATAAAGAGTAAATT CTCCTTTAATCTGAGGAAACCCTCTGTGCAGTGGAGTGA CACAGCTGAGTACTTCTGTGCCGTGAATGAATACGACTA CAAGCTCAGCTTTGGAGCCGGAACCACAGTAACTGTAA GAGCAAATATCCAGAACCC

41	TRAV13-1	X04944.1	ATGACATCCATTTCGAGCTGTATTTATATTCCTGTGGCTG CAGCTGGACTTGGTGAATGGAGAGAATGTGGAGCAGCA TCCTTCAACCCTGAGTGTCCAGGAGGGAGACAGCGCTG TTATCAAGTGTACTTATTCAGACAGTGCCTCAAACCTACT TCCCTTGGTATAAAGCAAGAACTTGGAAAAAGACCTCAG CTTATTATAGACATTCGTTCAAATGTGGGCGAAAAGAA AGACCAACGAATTGCTGTTACATTGAACAAGACAGCCA AACATTTCTCCCTGCACATCACAGAGACCCAACCTGAAG ACTCGGCTGTCTACTTCTGTGCAGCAAAGCGCAAGGCCT CTAGCAACACAGGCAAACCTAATCTTTGGGCAAGGGACA ACTTTACAAGTAAAACCAGATATCCAGAACCCT
42	TRBV2	M62379.1	GAACCTGAAGTCACCCAGACTCCCAGCCATCAGGTCAC ACAGATGGGACAGGAAGTGATCTTGCCTGTGTCCCCA TCTCTAATCACTTATACTTCTATTGGTACAGACAAATCTT GGGGCAGAAAGTCGAGTTTCTGGTTTCTTTTATAATAA TGAAATCTCAGAGAAGTCTGAAATATTCGATGATCAATT CTCAGTTGAAAGGCCTGATGGATCAAATTTCACTCTGAA GATCCGGTCCACAAAGCTGGAGGACTCAGCCATGTACT TCTGTGCCAGCAGT
43	TRBV3-1	L06889	ATGGGCTTCAGGCTCCTCTGCTGCGTCGCCTTCTGCCTC CTCCAAGCAGGTCCCTTGGACACAGCTGTTTCCCAGACT CCAAAATACCTGGTCACACAGATGGGAAACGACAAGTC CATTAATGTGAACAAAATCTGGGCCATGATACTATGTA TTGGTATAAACAGGACTCTAAGAAATTTCTGAAGATAAT GTTTAGCTACAATAACAAGGAGATCATTATAAATGAAA CAGTTCCAAATCGATTCTCACCTAAATCTCCAGACAAAG CTAAATTAAATCTTCACATCAATTCCTGGAGCTTGGTG ACTCTGCTGTGTATTTCTGTGCCAGC
44	TRBV15	M62376	ACCCCGACCTCTCAACCCAGGAATCAGAGCCTGAGACA GACAGATGCTTCATTCCTGTATGGGGTGGTATTCCTGCC ATGGGTCCTGGGCTTCTCCACTGGATGGCCCTTTGTCTC CTTGGAACAGGTCATGGGGATGCCATGGTCATCCAGAA CCCAAGATAACGGGTTACCCAGTTTGGAAAGCCAGTGA CCCTGAGTTGTTCTCAGACTTTGAACCATAACGTCATGT ACTGGTACCAGCAGAAGTCAAGTCAGGCCCCAAAGCTG CTGTTCCACTACTATAACAAAGATTTTAACAATGAAGCA GACACCCCTGATAACTTCCAATCCAGGAGGCCGAACAC TTCTTTCTGCTTTCTAGACATCCGCTCACCAGGCCTGGG GGACGCAGCCATGTACCAGTGTGCCACCAGC

[0057] Библиотеки, содержащие нуклеиновые кислоты, кодирующие варианты TCR, как описано в настоящем документе, содержат различные длины аминокислот при трансляции. В некоторых случаях длина каждого из аминокислотных фрагментов или средняя длина синтезированной аминокислоты может составлять по меньшей мере или около 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150 или больше чем 150 аминокислот. В некоторых случаях длина аминокислоты находится в диапазоне, составляющем около 15 - 150, 20 - 145, 25 - 140, 30 - 135, 35 - 130, 40 - 125, 45 - 120, 50 - 115, 55 - 110, 60 - 110, 65 - 105, 70 - 100 или 75 - 95

аминокислот. В некоторых случаях длина аминокислоты находится в диапазоне, составляющем около 22 - около 75 аминокислот.

[0058] Ряд вариантных последовательностей по меньшей мере для одной области цепи TCR для вариации *de novo* синтезируют с использованием способов, описанных в настоящем документе. В некоторых случаях ряд вариантных последовательностей *de novo* синтезируют для вариабельной области альфа-цепи TCR или бета-цепи TCR. Число вариантных последовательностей может составлять по меньшей мере или около 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500 или больше чем 500 последовательностей. В некоторых случаях число вариантных последовательностей находится в диапазоне, составляющем около 10 - 500, 25 - 475, 50 - 450, 75 - 425, 100 - 400, 125 - 375, 150 - 350, 175 - 325, 200 - 300, 225 - 375, 250 - 350 или 275 - 325 последовательностей.

[0059] В настоящем документе предусмотрены вариантные нуклеиновые кислоты, кодирующие вариантные TCR, причем вариантные TCR являются специфическими в отношении антигена. В некоторых случаях антиген участвует или ассоциирован с заболеванием, нарушением или состоянием. Например, антиген ассоциирован с пролиферативным заболеванием, опухолевым заболеванием, воспалительным заболеванием, иммунологическим нарушением, аутоиммунным заболеванием, инфекционным заболеванием, вирусным заболеванием, аллергической реакцией, паразитарной реакцией, реакцией «трансплантат-против-хозяина» или реакцией «хозяин-против-трансплантата». В некоторых случаях антигена представляет собой антиген, экспрессируемый на опухолевой клетке. В некоторых случаях антиген ассоциирован с таким патогеном, как вирус или бактерия.

[0060] В некоторых случаях вариантные TCR распознают антигены, которые являются рестриктированными в отношении конкретной ткани. Например, вариантные TCR рестриктированы в отношении нежизнеспособных клеточных линий или тканей. В некоторых случаях вариантные TCR распознают антигены из продуктов мутированных генов.

[0061] В настоящем документе предусмотрены библиотеки вариантных TCR, причем вариантные TCR кодируют варианты в области контакта с антигеном. В некоторых случаях остатки для вариации представляют собой предварительно выбранные или спрогнозированные остатки, которые контактируют с антигеном. В некоторых случаях остатки для вариации представляют собой предварительно выбранные или спрогнозированные остатки, расположенные в связывающем кармане.

[0062] Описанные в настоящем документе библиотеки вариантных TCR содержат одну или несколько мутаций в библиотеке. В некоторых случаях библиотеки вариантных TCR представляют собой библиотеки одного варианта, содержащие варианты на одном сайте во всей библиотеке. В некоторых случаях библиотеки вариантных TCR представляют собой библиотеки множественных вариантов, содержащие варианты на нескольких сайтах. В некоторых случаях количество сайтов составляет 2, 3, 4, 5, 6, 7, 8, 9, 10 или больше чем 10 сайтов.

[0063] В настоящем документе предусмотрены библиотеки, в которых один или несколько предварительно выбранных кодонов в гене TCR или фрагменте гена TCR кодируют вариантные аминокислотные остатки для создания вариации в полученных библиотеках вариантных белков TCR. В некоторых случаях вплоть до 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 аминокислотных остатков подвергают вариации. В некоторых случаях вплоть до 30 аминокислотных остатков подвергают вариации. В некоторых случаях вплоть до 5 аминокислотных остатков подвергают вариации. В некоторых случаях вплоть до 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 или больше чем 1000 аминокислотных остатков подвергают вариации. В некоторых случаях по меньшей мере или около 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 аминокислотных остатков подвергают вариации. В некоторых случаях по меньшей мере или около 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 или больше чем 1000 аминокислотных остатков подвергают вариации. В некоторых случаях все аминокислотные остатки в предварительно выбранной области подвергают вариации. В некоторых случаях библиотеки вариантных TCR характеризуются высокой степенью разнообразия. В некоторых случаях библиотеки содержат по меньшей мере или около 10^6 , 10^8 , 10^9 , 10^{10} , 10^{11} , 10^{12} или 10^{13} вариантов. В некоторых случаях библиотеки содержат по меньшей мере или около 10^9 вариантов.

[0064] В настоящем документе предусмотрены вариантные нуклеиновые кислоты, кодирующие вариантный TCR для создания растворимых вариантных TCR. В некоторых случаях растворимые вариантные TCR являются специфическими в отношении антигена. В некоторых случаях растворимые вариантные TCR содержат одну или несколько цепей из цепи TCR α , цепи TCR β , цепи TCR δ , цепи TCR γ или их комбинаций. В некоторых случаях растворимые вариантные TCR содержат одну или несколько цепей цепи TCR α . В некоторых случаях растворимые вариантные TCR содержат одну или несколько цепей цепи TCR β . В некоторых случаях растворимые вариантные TCR содержат одну или несколько цепей цепи TCR α и одну или несколько цепей цепи TCR β . В некоторых случаях количество цепей TCR составляет 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 или больше чем 10 цепей TCR, причем цепь TCR представляет собой цепь TCR α , цепь TCR β , цепь TCR δ , цепь TCR γ или их комбинации.

[0065] Растворимый вариантный TCR, описанный в настоящем документе, может содержать варианты в цепи TCR α , TCR β , TCR δ , TCR γ или их комбинациях. В некоторых случаях растворимые вариантные TCR содержат варианты в цепи TCR α . В некоторых случаях растворимые вариантные TCR содержат варианты в цепи TCR β . В некоторых случаях растворимые вариантные TCR содержат варианты в цепи TCR α и цепи TCR β .

[0066] Описанные в настоящем документе растворимые вариантные TCR могут содержать линкер между цепями TCR. Например, линкер представляет собой пептидную последовательность между цепями TCR. В некоторых случаях растворимые вариантные TCR содержат одну или несколько линкерных последовательностей. В некоторых случаях количество линкерных последовательностей составляет 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 или больше чем 10 линкерных последовательностей. В некоторых случаях количество линкерных последовательностей находится в диапазоне, составляющем 1-10, 2-8, 3-7 или 4-6 линкерных последовательностей.

[0067] Различные способы получения растворимых TCR предусмотрены в настоящем документе. В некоторых случаях библиотеки вариантных TCR сначала синтезируют *de novo* и затем переносят в экспрессионные векторы (т.е. в бактериальные клетки или клетки млекопитающих). В некоторых случаях очищенные TCR получают для создания растворимых TCR. Растворимые TCR могут не содержать доменов или остатков, необходимые для вставки или стабилизации в клеточной мембране. Растворимые TCR можно дополнительно модифицировать, чтобы они находились в форме слитого белка. Иллюстративные дополнительные компоненты, подлежащие включению в слитый белок, включают в себя без ограничения домены иммуноглобулина (Ig) IgG, IgA и IgD, IgM или IgE. В частности, слитый белок конструируют так, чтобы он включал в себя растворимый TCR и константный домен (Fc) или его фрагмент домена Ig. Иллюстративные подтипы IgG включают в себя без ограничения IgG1, IgG2, IgG3 и IgG4. Например, такой слитый белок может состоять из растворимого одноцепочечного Т-клеточного рецептора, генетически связанного с Fc-областью тяжелой цепи IgG1 человека. Fc-области, выбранные для слитых белков, можно выбрать на основании связывания с определенными Fc-рецепторами, например, Fc γ RI, Fc γ RIIa, Fc γ RIIb/c, Fc γ RIIIa, Fc γ RIIIb и FcRn.

[0068] Описанные в настоящем документе библиотеки с областями общей последовательности, выбранной для вариации, могут характеризоваться предварительно выбранным профилем вариации. Например, по меньшей мере около 1, 5, 10, 15, 20, 30, 40, 50 или больше нуклеиновых кислот в общей последовательности могут быть выбраны для инвариантности или могут быть выбраны, чтобы характеризоваться ограниченным профилем вариации. В некоторых случаях вариантные нуклеиновые кислоты, каждая из

которых кодирует вариантный TCR с профилем вариантных кодонов, встроены в экспрессионные векторы и ими трансфицируют (например, путем электропорации, трансфекции или трансдукции) клетки млекопитающих для скрининга.

[0069] Описанные в настоящем документе библиотеки вариантных нуклеиновых кислот можно использовать для создания библиотек вариантных белков. Такие библиотеки белков можно подвергнуть скринингу в отношении изменений в активности TCR, например, аффинности связывания или авидности, стабильности, экспрессии белка, или уничтожения клеток-мишеней (например, уничтожение раковых клеток), или активности ADCC. Скрининг может включать в себя анализы связывания с антигеном(ами), ассоциированным(и) с конкретной формой рака, представляющей интерес. Рак может представлять собой солидный рак или гемобластоз. В некоторых случаях рак представляет собой рак мочевого пузыря, рак легкого, рак головного мозга, меланому, рак молочной железы, неходжкинскую лимфому, рак шейки матки, рак яичников, рак толстой и прямой кишки, рак поджелудочной железы, рак пищевода, рак предстательной железы, рак почки, рак кожи, лейкоз, рак щитовидной железы, рак печени или рак матки. В некоторых случаях антиген представляет собой MART1, CEA, gp100, NY-ESO-1, WT1, MUC1, HER-2/neu, MAGE-A3, p53, PSMA, GD2, протеиназу 3 (PR1), тирозиназу, сурвивин, PSA, hTERT и EphA2. Иллюстративные антигены для применения в анализах связывания включают в себя без ограничения те, которые представлены в таблице 4.

Таблица 4. Иллюстративные антигены для применения в анализах связывания

Название пептида антигена	Последовательность белка	Мишень-заболевание
MAGE A3	KVAELVHFL	Меланома
MAGE A12	KMAELVHFL	Меланома
MAGE A2	KMVELVHFL	Меланома
MAGE A6	KVAKLVHFL	Меланома
NY-ESO-1	SLLMWITQV	Многие формы рака, включая в себя меланому и саркому
пептид CEA	IMIGVLVGV	Метастатический рак толстой и прямой кишки

[0070] Скрининг может включать в себя *in vitro*, *in vivo* или *ex vivo* анализы. Клетки для скрининга включают в себя первичные клетки, полученные из организма живых субъектов или из клеточных линий. Клетки можно получить от прокариот (например, бактерий и грибов) или эукариот (например, животных и растений). Иллюстративные клетки животных включают в себя без ограничения клетки мыши, кролика, примата и насекомого. В некоторых случаях клетки для скрининга включают в себя клеточную

линию, включая в себя без ограничения клеточную линию яичника китайского хомячка (СНО), клеточную линию эмбриональной почки человека (НЕК) или клеточную линию почки новорожденного хомяка (ВНК). В некоторых случаях описанные в настоящем документе библиотеки нуклеиновых кислот также могут быть доставлены в многоклеточный организм. Иллюстративные многоклеточные организмы включают в себя без ограничения растение, мышь, кролика, примата и насекомое.

[0071] Библиотеки вариантных TCR, созданные с использованием способов, описанных в настоящем документе, можно подвергнуть скринингу для выбора модифицированной последовательности гена TCR, обеспечивая белковый комплекс с улучшенной аффинностью (мера силы взаимодействия между эпитопом и антигенсвязывающим сайтом антитела) для опухолевого антигена. Кроме того, оценивают дополнительные функциональные соображения, такие как экспрессия вариантного гена, авидность (мера общей силы комплекса антитело-антиген), стабильность и специфичность клеток-мишеней (т.е. раковых клеток). В некоторых случаях повышенная специфичность комплекса TCR обеспечивает пониженную перекрестную реактивность к антигенам, не связанным с раком, по сравнению с эталонным комплексом невариантного TCR. В некоторых случаях библиотеки вариантных TCR подвергают скринингу в отношении локализации в клетке. В некоторых случаях библиотеки вариантных TCR подвергают скринингу для идентификации правильно локализованных вариантных TCR.

[0072] В настоящем документе предусмотрены библиотеки вариантных TCR с улучшенной аффинностью или специфичностью к антигену, которые экспрессируются в клетках. В некоторых случаях библиотеки вариантных TCR экспрессируются в клетках для создания клеток с измененной специфичностью к антигену. В некоторых случаях клетки представляют собой Т-клетки. В некоторых случаях библиотека вариантов содержит последовательности, кодирующие вариацию по меньшей мере одного кодона переменного домена альфа-цепи TCR или бета-цепи TCR. В некоторых случаях библиотека вариантов содержит последовательности, кодирующие вариацию множества кодонов переменного домена альфа-цепи TCR или бета-цепи TCR. Например, библиотека вариантов содержит вариацию последовательности в таблице 2 или таблице 3. В некоторых случаях библиотеки вариантных TCR содержат вариацию в одном или нескольких экзонах переменного гена альфа-цепи TCR. Иллюстративные последовательности экзонов для вариации показаны в таблице 2 или таблице 3. В некоторых случаях библиотеки вариантных TCR содержат вариацию в одном или нескольких экзонах переменного гена бета-цепи TCR.

[0073] В некоторых случаях библиотеки вариантных TCR, которые экспрессируются в клетках, используют для идентификации вариантных TCR с улучшенной экспрессией вариантных генов, авидностью, стабильностью, аффинностью или специфичностью. Например, создают первую библиотеку вариантных TCR, которая содержит улучшенную специфичность к опухолевому антигену. В некоторых случаях после идентификации вариантных TCR с улучшенной экспрессией гена, авидностью, стабильностью, аффинностью или специфичностью эти вариантные TCR дополнительно подвергают вариации для получения второй библиотеки вариантных TCR со вторым улучшением. Например, вариантные TCR с улучшенной специфичностью дополнительно подвергают вариации для идентификации вариантов, дополнительно содержащих улучшенную стабильность. В некоторых случаях вторая библиотека содержит варианты в той же области, что и первая библиотека. В некоторых случаях вторая библиотека содержит варианты в другой области по сравнению с первой библиотекой. Например, первая библиотека содержит варианты в варибельной области TCR, а вторая библиотека содержит варианты в константной области TCR. В некоторых случаях создают несколько библиотек вариантных TCR. В некоторых случаях количество библиотек вариантных TCR составляет по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 или более 12 библиотек вариантов. В некоторых случаях каждая из библиотек вариантов характеризуется улучшениями в экспрессии генов, авидности, стабильности, аффинности или специфичности.

[0074] В некоторых случаях клетки, сконструированные для экспрессии вариантных TCR, вводят субъекту. В некоторых случаях клетки являются аутологичными, т.е. происходят из собственных клеток субъекта. Альтернативно, клетки, экспрессирующие вариантные TCR, являются аллогенными, т.е. происходят от другого субъекта со сходным типом ткани. В некоторых случаях клетки адаптированы для субъекта. В некоторых случаях клетки совместимы с локальной тканью.

[0075] Библиотеки вариантных TCR, созданные с использованием способов, описанных в настоящем документе, можно подвергнуть скринингу для выбора по меньшей мере одной модифицированной последовательности, кодирующей ген TCR, обеспечивающей усиленный профиль экспрессии сконструированного белкового комплекса TCR в клетках. Стратегии повышения относительной экспрессии генов включают в себя оптимизацию кодонов, введение одной или нескольких дисульфидных связей между цепями TCR, введение последовательности остатка мыши в домен константной области, вариацию некомплементарной кодирующей области варибельного(ых) домена(ов), использование технологии TALEN или CRISPR для селективного нокдауна или остановки экспрессии эндогенного белка TCR в

модифицированных Т-клетках, доставку иммуностимулирующих цитокинов к модифицированным Т-клеткам или комбинацию любой или всех этих стратегий. Вставка дисульфидной связи (например, в шарнирную область) подвергнутой вариации последовательности гена, кодирующей α или β субъединицы TCR, обеспечивает эффективную сборку сконструированных вариантов при экспрессии в клетках. Количество дисульфидных связей, добавленных к последовательности гена TCR, описанной в настоящем документе, может составлять, по меньшей мере, 1, 2, 3, 4, 5 или более.

[0076] В некоторых случаях описанные в настоящем документе способы синтеза *de novo* обеспечивают создание библиотеки вариантных антигенов для идентификации антигена, которые являются высокоэффективными как при связывании с комплексом вариантного TCR, полученным способами, описанными в настоящем документе, а также с картированной молекулой МНС I субъекта. Таким образом, как комплекс TCR, так и антигенсвязывающую часть можно оптимизировать для усиленного взаимодействия в иммунологическом синапсе. Для целей скрининга нуклеиновые кислоты, кодирующие библиотеку вариантных антигенов, сначала синтезируются *de novo*, а затем переносят в экспрессионные векторы для экспрессии (т.е. в клетках бактерий или млекопитающих) с целью получения очищенного пептида для анализов связывания с комплексами TCR и комплексами МНС I.

[0077] Синтез библиотеки вариантов

[0078] Описанные в настоящем документе способы обеспечивают синтез библиотеки нуклеиновых кислот, каждая из которых кодирует предварительно определенный вариант по меньшей мере одной предварительно определенной эталонной последовательности нуклеиновой кислоты. В некоторых случаях предварительно определенная эталонная последовательность представляет собой последовательность нуклеиновой кислоты, кодирующую белок, и библиотека вариантов содержит последовательности, кодирующие вариацию по меньшей мере одного кодона так, чтобы множество различных вариантов одного остатка в последующем белке, кодируемом синтезированной нуклеиновой кислотой, создавались с помощью стандартного процесса трансляции. Синтезированные конкретные изменения в последовательности нуклеиновой кислоты можно вводить с помощью встраивания нуклеотидных изменений в перекрывающиеся или содержащие «тупые концы» нуклеиновокислотные праймеры. Альтернативно, популяция нуклеиновых кислот может совместно кодировать длинную нуклеиновую кислоту (например, ген) и ее варианты. В этой конфигурации популяцию нуклеиновых кислот можно гибридизовать и подвергать стандартным техникам молекулярной биологии с образованием длинной нуклеиновой кислоты (например, гена) и

ее вариантов. Когда длинная нуклеиновая кислота (например, ген) и ее варианты экспрессируются в клетках, создается библиотека вариантов белка. Аналогично в настоящем документе предусмотрены способы синтеза библиотеки вариантов, кодирующей последовательности РНК (например, микроРНК, кшРНК и мРНК) или последовательности ДНК (например, области энхансера, промотора, UTR и терминатора). Кроме того, в настоящем документе предусмотрены последующие применения для вариантов, выбранных из библиотек, синтезированных с использованием способов, описанных в настоящем документе. Последующие применения включают в себя идентификацию вариантных нуклеиновокислотных или белковых последовательностей с усиленными биологически важными функциями, например, биохимической аффинностью, ферментативной активностью, изменениями в клеточной активности, и для лечения или профилактики болезненного состояния

[0079] *Синтез с последующим ПЦР-мутагенезом*

[0080] Первый способ синтеза библиотеки вариантов нуклеиновых кислот относится к способам ПЦР-мутагенеза (насыщающего или ненасыщающего). В этом технологическом процессе синтезируют множество нуклеиновых кислот, причем каждая нуклеиновая кислота кодирует предварительно определенную последовательность, которая представляет собой предварительно определенный вариант эталонной последовательности нуклеиновой кислоты. Ссылаясь на фигуры, иллюстративный технологический процесс изображен на фиг. **1A-1D**, где нуклеиновые кислоты образуются на поверхности. На фиг. **1A** показан развернутый вид одного кластера поверхности со 121 локусом. Каждая нуклеиновая кислота, изображенная на фиг. **1B**, представляет собой праймер, который можно использовать для амплификации из эталонной последовательности нуклеиновой кислоты для получения библиотеки вариантов длинных нуклеиновых кислот, фиг. **1C**. Затем библиотеку вариантов длинных нуклеиновых кислот необязательно подвергают транскрипции и/или трансляции для создания библиотеки вариантов РНК или белка, фиг. **1D**. На этом иллюстративном изображении показано устройство, характеризующееся по существу плоскую поверхность, которое используют для синтеза нуклеиновых кислот *de novo*, фиг. **1A**. В некоторых случаях устройство содержит кластер локусов, где каждый локус является сайтом для удлинения нуклеиновой кислоты. В некоторых случаях один кластер содержит все варианты нуклеиновых кислот, необходимые для создания требуемой библиотеки вариантных последовательностей. В альтернативной конфигурации планшет содержит поле локусов, которые не разделены на кластеры.

[0081] В некоторых случаях нуклеиновые кислоты, синтезированные в пределах кластера (например, как видно на фиг. **1A**), амплифицируют с помощью ПЦР.

Такая конфигурация может обеспечить улучшенное представление нуклеиновых кислот по сравнению с амплификацией неидентичных нуклеиновых кислот по всей планшете без кластерного расположения. В некоторых случаях амплификация нуклеиновых кислот, синтезируемых на поверхностях локусов в пределах кластера, преодолевает негативные воздействия на репрезентацию из-за повторного синтеза больших популяций нуклеиновых кислот, характеризующихся нуклеиновыми кислотами с высоким содержанием GC. В некоторых случаях кластер, описанный в настоящем документе, содержит около 50-1000, 75-900, 100-800, 125-700, 150-600, 200-500, 50-500 или 300-400 отдельных локусов. В некоторых случаях локус представляет собой пятно, лунку, микролунку, канал или стержень. В некоторых случаях каждый кластер характеризуется по крайней мере 1X, 2X, 3X, 4X, 5X, 6X, 7X, 8X, 9X, 10X или более избыточностью отдельных признаков, поддерживающих удлинение нуклеиновых кислот с идентичной последовательностью.

[0082] Описанная в настоящем документе синтезированная *de novo* библиотека нуклеиновых кислот может содержать множество нуклеиновых кислот, каждая из которых характеризуется по меньшей мере одной вариантной последовательностью в первом положении, в положении “х”, и каждую вариантную нуклеиновую кислоту используют в качестве праймера в первом раунде ПЦР для создания первого продукта удлинения. В этом примере положение “х” в первой нуклеиновой кислоте **220** кодирует вариантную последовательность кодона, т.е. один из 19 возможных вариантов из эталонной последовательности. См. фиг. **2А**. Вторую нуклеиновую кислоту **225**, содержащую последовательность, перекрывающую последовательность первой нуклеиновой кислоты, также используют в качестве праймера в отдельном раунде ПЦР для получения второго продукта удлинения. Кроме того, внешние праймеры **215**, **230** можно использовать для амплификации фрагмента из длинной последовательности нуклеиновой кислоты. Полученные продукты амплификации представляют собой фрагменты длинной последовательности нуклеиновой кислоты **235**, **240**. См. фиг. **2В**. Затем фрагменты длинной последовательности нуклеиновой кислоты **235**, **240** гибридизуют и подвергают реакции удлинения с образованием варианта длинной нуклеиновой кислоты **245**. См. фиг. **2С**. Перекрывающиеся концы первого и второго продуктов удлинения могут служить праймером второго раунда ПЦР, тем самым создавая третий продукт удлинения (фиг. **2D**), который содержит вариант. Для увеличения выхода вариант длинной нуклеиновой кислоты амплифицируют в реакции, включающей в себя ДНК-полимеразу, реагенты для амплификации и внешние праймеры **215**, **230**. В некоторых случаях вторая нуклеиновая кислота содержит последовательность, смежную, но не включающую в себя сайт варианта. В альтернативной конфигурации создают первую нуклеиновую кислоту, которая содержит

область, которая перекрывается со второй нуклеиновой кислотой. В этом сценарии первую нуклеиновую кислоту синтезируют с вариацией в одном кодоне для вплоть до 19 вариантов. Вторая нуклеиновая кислота не содержит вариантную последовательность. Необязательно первая популяция содержит первые варианты нуклеиновых кислот и дополнительные нуклеиновые кислоты, кодирующие варианты в другом сайте кодона. Альтернативно, первую нуклеиновую кислоту и вторую нуклеиновую кислоту можно сконструировать для лигирования по тупым концам.

[0083] Альтернативный способ ПЦР-мутагенеза показан на фиг. **3A-3F**. В таком способе матричную молекулу нуклеиновой кислоты **300**, содержащую первую и вторую цепи **305**, **310**, амплифицируют в реакции ПЦР, содержащей первый праймер **315** и второй праймер **320** (фиг. **3A**). Реакция амплификации включает в себя урацил в качестве нуклеотидного реагента. Создают меченый урацилом продукт удлинения **325** (фиг. **3B**), необязательно очищают, и он служит матрицей для последующей реакции ПЦР с использованием первой нуклеиновой кислоты **335** и множества второй нуклеиновой кислоты **330** для создания первых продуктов удлинения **340** и **345**. (Фиг. **3C-3D**). В этом способе множество второй нуклеиновой кислоты **330** содержит нуклеиновые кислоты, кодирующие вариантные последовательности (обозначенные как X, Y и Z на фиг. **3C**). Меченую урацилом матричную нуклеиновую кислоту расщепляют специфическим для урацила эксцизионным реагентом, например, USER digest, доступным коммерчески от New England Biolabs. Вариант **335** и различные кодоны **330** с вариантами X, Y и Z добавляют и проводят ограниченную стадию ПЦР для создания фиг. **3D**. После расщепления урацилсодержащей матрицы перекрывающиеся концы продуктов удлинения служат для инициирования реакции ПЦР с первыми продуктами удлинения **340** и **345**, действующими в качестве праймеров в комбинации с первым внешним праймером **350** и вторым внешним праймером **355**, в результате чего образуется библиотека молекул нуклеиновых кислот **360**, содержащая множество вариантов X, Y и Z на варианте сайте. Фиг. **3F**.

[0084] *De novo синтез популяции с вариантными и невариантными частями длинной нуклеиновой кислоты*

[0085] Во втором способе синтеза библиотеки вариантов поверхность используют для de novo синтеза множественных фрагментов длинной нуклеиновой кислоты, причем по меньшей мере один из фрагментов синтезируют в нескольких версиях, причем каждая версия происходит из различной вариантной последовательности. В такой конфигурации все фрагменты, необходимые для сборки библиотеки вариантных нуклеиновых кислот широкого диапазона, синтезируют de novo. Синтезированные фрагменты могут содержать перекрывающуюся последовательность, так чтобы после

синтеза библиотека фрагментов подвергалась гибридизации. После гибридизации можно провести реакцию удлинения для заполнения любых комплементарных пропусков.

[0086] Альтернативно, синтезированные фрагменты можно амплифицировать с помощью праймеров и затем подвергнуть либо лигированию по «тупым концам», либо гибридизации с перекрытием. В некоторых случаях устройство содержит кластер локусов, где каждый локус представляет собой сайт для удлинения нуклеиновой кислоты. В некоторых случаях один кластер содержит все варианты нуклеиновых кислот и другие последовательности фрагментов предварительно определенной длинной нуклеиновой кислоты для создания требуемой библиотеки вариантных последовательностей нуклеиновых кислот. Кластер может содержать около 50 - 500 локусов. В некоторых конфигурациях кластер содержит более 500 локусов.

[0087] Каждую отдельную нуклеиновую кислоту в первой популяции нуклеиновых кислот можно создать на отдельном, индивидуально адресуемом локусе кластера. Один вариант нуклеиновой кислоты может быть представлен множеством индивидуально адресуемых локусов. Каждый вариант в первой популяции нуклеиновых кислот может быть представлен 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 или более раз. В некоторых случаях каждый вариант в первой популяции нуклеиновых кислот представлен на 3 или менее локусах. В некоторых случаях каждый вариант в первой популяции нуклеиновых кислот представлен на двух локусах. В некоторых случаях каждый вариант в первой популяции нуклеиновых кислот представлен только на одном локусе.

[0088] В настоящем документе предусмотрены способы получения библиотек нуклеиновых кислот с пониженной избыточностью. В некоторых случаях вариантные полинуклеотиды можно создать без необходимости синтезировать вариантный полинуклеотид более 1 раза для получения требуемого вариантного полинуклеотида. В некоторых случаях согласно настоящему раскрытию предусмотрены способы создания вариантных полинуклеотидов без необходимости синтезировать вариантный полинуклеотид более чем 1, 2, 3, 4, 5 раз, 6, 7, 8, 9, 10 или более раз для создания требуемого вариантного полинуклеотида.

[0089] Вариантные полинуклеотиды можно создать без необходимости синтезировать вариантный полинуклеотид на более чем 1 отдельном сайте для получения требуемого вариантного полинуклеотида. Согласно настоящему раскрытию предусмотрены способы создания вариантных полинуклеотидов без необходимости синтеза вариантного полинуклеотида на более чем 1 сайте, 2 сайтах, 3 сайтах, 4 сайтах, 5 сайтах, 6 сайтах, 7 сайтах, 8 сайтах, 9 сайтах или 10 сайтах для создания требуемого вариантного полинуклеотида. В некоторых случаях полинуклеотид синтезируют не более

чем на 6, 5, 4, 3, 2 или 1 отдельном сайте. Один и тот же полинуклеотид можно синтезировать на 1, 2 или 3 отдельных локусах на поверхности.

[0090] В некоторых случаях количество локусов, представляющих один вариантный полинуклеотид, является функцией количества материала нуклеиновой кислоты, необходимого для последующей обработки, например, реакции амплификации или клеточного анализа. В некоторых случаях количество локусов, представляющих один вариантный полинуклеотид, является функцией доступных локусов в одном кластере.

[0091] В настоящем документе предусмотрены способы создания библиотеки нуклеиновых кислот, содержащей варианты нуклеиновых кислот, различающиеся по множеству сайтов в эталонной нуклеиновой кислоте. В таких случаях каждую библиотеку вариантов создают на индивидуально адресуемом локусе в пределах кластера локусов. Следует понимать, что число сайтов вариантов, представленных библиотекой нуклеиновых кислот, будет определяться количеством индивидуально адресуемых локусов в кластере и количеством требуемых вариантов на каждом сайте. В некоторых случаях каждый кластер содержит около 50 - 500 локусов. В некоторых случаях каждый кластер содержит 100 - 150 локусов.

[0092] В иллюстративной конфигурации 19 вариантов представлены на сайте вариантов, соответствующем кодонам, кодирующим каждую из 19 возможных вариантов аминокислот. В другом иллюстративном случае 61 вариант представлен на варианте сайта, соответствующем триплетам, кодирующим каждую из 19 возможных вариантов аминокислот. В неограничивающем примере кластер содержит 121 индивидуально адресуемый locus. В этом примере популяция нуклеиновых кислот содержит 6 копий каждого варианта с одним сайтом ($6 \text{ копий} \times 1 \text{ вариантный сайт} \times 19 \text{ вариантов} = 114 \text{ локусов}$), 3 копии каждого варианта с двумя сайтами ($3 \text{ копии} \times 2 \text{ вариантных сайта} \times 19 \text{ вариантов} = 114 \text{ локусов}$) или 2 копии каждого из вариантов с тремя сайтами ($2 \text{ репликации} \times 3 \text{ вариантных сайта} \times 19 \text{ вариантов} = 114 \text{ локусов}$). В некоторых случаях популяция нуклеиновых кислот содержит варианты на четырех, пяти, шести или более шести вариантных сайтах.

[0093] *Вариация кодонов*

[0094] Описанные в настоящем документе библиотеки вариантов нуклеиновых кислот могут содержать множество нуклеиновых кислот, причем каждая нуклеиновая кислота кодирует вариантную последовательность кодона по сравнению с эталонной последовательностью нуклеиновой кислоты. В некоторых случаях каждая нуклеиновая кислота первой популяции нуклеиновых кислот содержит вариант на одном вариантном сайте. В некоторых случаях первая популяция нуклеиновых кислот содержит множество

вариантов на одном вариантном сайте так, что первая популяция нуклеиновых кислот содержит больше чем один вариант на одном и том же вариантном сайте. Первая популяция нуклеиновых кислот может содержать нуклеиновые кислоты, совместно кодирующие множественные варианты кодонов на одном и том же вариантном сайте. Первая популяция нуклеиновых кислот может содержать нуклеиновые кислоты, совместно кодирующие вплоть до 19 или больше кодонов в одном и том же положении. Первая популяция нуклеиновых кислот может содержать нуклеиновые кислоты, совместно кодирующие вплоть до 60 триплетов варианта в одном и том же положении, или первая популяция нуклеиновых кислот может содержать нуклеиновые кислоты, совместно кодирующие вплоть до 61 различных триплетов кодонов в одном и том же положении. Каждый вариант может кодировать кодон, который дает в результате различную аминокислоту во время трансляции. В таблице 5 представлен перечень каждого кодона, возможного (и репрезентативной аминокислоты) для вариантного сайта.

Таблица 5. Перечень кодонов и аминокислот

Аминокислоты	Однобукв. код	Трехбукв. код	Кодоны			
Аланин	A	Ala	GCA	GCC	GCG	GCT
Цистеин	C	Cys	TGC	TGT		
Аспарагиновая кислота	D	Asp	GAC	GAT		
Глутаминовая кислота	E	Glu	GAA	GAG		
Фенилаланин	F	Phe	TTC	TTT		
Глицин	G	Gly	GGA	GGC	GGG	GGT
Гистидин	H	His	CAC	CAT		
Изолейцин	I	Iso	ATA	ATC	ATT	
Лизин	K	Lys	AAA	AAG		
Лейцин	L	Leu	TTA	TTG	CTA	CTC
Метионин	M	Met	ATG			
Аспарагин	N	Asn	AAC	AAT		
Пролин	P	Pro	CCA	CCC	CCG	CCT
Глутамин	Q	Gln	CAA	CAG		
Аргинин	R	Arg	AGA	AGG	CGA	CGC
Серин	S	Ser	AGC	AGT	TCA	TCC
Треонин	T	Thr	ACA	ACC	ACG	ACT
Валин	V	Val	GTA	GTC	GTG	GTT
Триптофан	W	Trp	TGG			
Тирозин	Y	Tyr	TAC	TAT		

[0095] Популяция нуклеиновых кислот может содержать подвергнутые вариации нуклеиновые кислоты, совместно кодирующие вплоть до 20 вариаций кодонов в нескольких положениях. В таких случаях каждая нуклеиновая кислота в популяции содержит вариацию для кодонов в больше чем одном положении в одной и той же нуклеиновой кислоте. В некоторых случаях каждая нуклеиновая кислота в популяции содержит вариацию для кодонов в 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 или более кодонах в одной нуклеиновой кислоте. В некоторых случаях каждая вариантная длинная нуклеиновая кислота содержит вариацию для кодонов в 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 или более кодонах в одной длинной нуклеиновой кислоте. В некоторых случаях популяция вариантных нуклеиновых кислот содержит вариацию для кодонов в 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 или более кодонах одной нуклеиновой кислоте. В некоторых случаях популяция вариантных нуклеиновых кислот содержит вариацию для кодонов по меньшей мере в около 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 или более кодонах в одной длинной нуклеиновой кислоте.

[0096] В настоящем документе предусмотрены способы, при которых вторую популяцию нуклеиновых кислот создают на втором кластере, содержащем множество индивидуально адресуемых локусов. Вторая популяция нуклеиновых кислот может содержать множество вторых нуклеиновых кислот, которые являются константными для каждого положения кодона (т.е. кодируют одну и ту же аминокислоту в каждом положении). Вторая нуклеиновая кислота может перекрываться по меньшей мере с частью первых нуклеиновых кислот. В некоторых случаях вторые нуклеиновые кислоты не содержат вариантный сайт, представленный на первых нуклеиновых кислотах. Альтернативно, вторая популяция нуклеиновых кислот может содержать множество вторых нуклеиновых кислот, в которых изменяется одно или несколько положений кодонов.

[0097] В настоящем документе предусмотрены способы синтеза библиотеки нуклеиновых кислот, при которых создают одну популяцию нуклеиновых кислот, содержащую варианты в нескольких положениях кодонов. Первую популяцию нуклеиновых кислот можно создать на первом кластере, содержащем множество индивидуально адресуемых локусов. В таких случаях первая популяция нуклеиновых кислот содержит варианты в различных положениях кодонов. В некоторых случаях различные сайты являются последовательными (т.е. кодирующими последовательные аминокислоты). Первая популяция нуклеиновых кислот может содержать подвергнутые вариации нуклеиновые кислоты, совместно кодирующие вплоть до 19 вариантов кодонов

на одном и том же или дополнительном вариантном сайте. Первая популяция нуклеиновых кислот может включать в себя множество первых нуклеиновых кислот, которое содержит вплоть до 19 вариантов в положении x , вплоть до 19 вариантов в положении y и вплоть до 19 вариантов в положении z . В такой конфигурации каждый вариант кодирует различную аминокислоту так, что вплоть до 19 вариантов аминокислот кодируются на каждом из различных вариантных сайтов. В дополнительном случае вторую популяцию нуклеиновых кислот создают на втором кластере, содержащем множество индивидуально адресуемых локусов. Вторая популяция нуклеиновых кислот может содержать множество вторых нуклеиновых кислот, которые являются константными для каждого положения кодона (т.е. кодируют одну и ту же аминокислоту в каждом положении). Вторые нуклеиновые кислоты могут перекрываться по меньшей мере с частью первых нуклеиновых кислот. Вторые нуклеиновые кислоты могут не содержать вариантный сайт, представленный на первых нуклеиновых кислотах.

[0098] Библиотеки вариантных нуклеиновых кислот, созданные с помощью способов, описанных в настоящем документе, обеспечивают создание библиотек вариантных белков. В первой иллюстративной конфигурации матричная нуклеиновая кислота кодирует последовательность, которая при транскрипции и трансляции дает в результате эталонную аминокислотную последовательность (фиг. 4A) с рядом положений кодонов, указанных одним кружком. Варианты нуклеиновой кислоты матрицы можно создать с использованием способов, описанных в настоящем документе. В некоторых случаях один вариант присутствует в нуклеиновой кислоте, давая в результате одну аминокислотную последовательность (фиг. 4B). В некоторых случаях больше чем один вариант присутствует в нуклеиновой кислоте, причем варианты разделены одним или несколькими кодонами, давая в результате белок с интервалом между вариантными остатками (фиг. 4C). В некоторых случаях больше чем один вариант присутствует в нуклеиновой кислоте, причем варианты являются последовательными и соседними или следуют друг за другом, давая в результате разделенные интервалами участки вариантных остатков (фиг. 4D). В некоторых случаях два участка вариантов присутствуют в нуклеиновой кислоте, причем каждый участок вариантов содержит последовательные и соседние или следующие друг за другом варианты (фиг. 4E).

[0099] В настоящем документе предусмотрены способы создания библиотеки вариантов нуклеиновых кислот, в которых каждый вариант содержит вариант кодона с одним положением. В одном случае матричная нуклеиновая кислота содержит ряд положений кодонов, причем иллюстративные аминокислотные остатки показаны кружками с соответствующим им белковым кодоном в однобуквенном коде, фиг. 5A. На фиг. 5B

показана библиотека вариантов аминокислот, кодируемых библиотекой вариантных нуклеиновых кислот, где каждый вариант содержит вариант с одним положением, показанный как “X”, расположенный на одном отличающемся сайте. Первый вариант положения содержит любой кодон для замены аланина, второй вариант с любым кодоном, кодируемым библиотекой вариантных нуклеиновых кислот для замены триптофана, третий вариант с любым кодоном для замены изолейцина, четвертый вариант с любым кодоном для замены лизина, пятый вариант с любым кодоном для замены аргинина, шестой вариант с любым кодоном для замены глутаминовой кислоты, и седьмой вариант с любым кодоном для замены глутамина. Если все или меньше чем все варианты кодонов кодируются библиотекой вариантных нуклеиновых кислот, в результате создается соответствующая популяция вариантов аминокислотных последовательностей после экспрессии белка (т.е. стандартных клеточных событий транскрипции ДНК с последующими событиями трансляции и процессинга).

[00100] В некоторых конфигурациях библиотеку создают с несколькими сайтами вариантов с одним положением. Как показано на фиг. **6А**, представлена матрица дикого типа. На фиг. **6В** показана полученная аминокислотная последовательность с двумя сайтами вариантов кодона с одним положением, причем каждый вариант кодона, кодирующий различную аминокислоту, показан кружками с различной штриховкой.

[00101] В настоящем документе предусмотрены способы создания библиотеки, характеризующейся участком с несколькими сайтами вариантов с одним положением. Каждый участок нуклеиновой кислоты может характеризоваться 1, 2, 3, 4, 5 или больше вариантами. Каждый участок нуклеиновой кислоты может характеризоваться по меньшей мере 1 вариантом. Каждый участок нуклеиновой кислоты может характеризоваться по меньшей мере 2 вариантами. Каждый участок нуклеиновой кислоты может характеризоваться по меньшей мере 3 вариантами. Например, участок из 5 нуклеиновых кислот может характеризоваться 1 вариантом. Участок из 5 нуклеиновых кислот может характеризоваться 2 вариантами. Участок из 5 нуклеиновых кислот может характеризоваться 3 вариантами. Участок из 5 нуклеиновых кислот может характеризоваться 4 вариантами. Например, участок из 4 нуклеиновых кислот может характеризоваться 1 вариантом. Участок из 4 нуклеиновых кислот может характеризоваться 2 вариантами. Участок из 4 нуклеиновых кислот может характеризоваться 3 вариантами. Участок из 4 нуклеиновых кислот может характеризоваться 4 вариантами.

[00102] В некоторых случаях все варианты с одним положением могут кодировать одну и ту же аминокислоту, например, гистидин. Как показано на фиг. **7А**, представлена эталонная аминокислотная последовательность. В этой конфигурации

участок нуклеиновой кислоты кодирует несколько сайтов вариантов с одним положением и при экспрессии дает в результате аминокислотную последовательность, характеризующуюся всеми вариантами с одним положением, кодирующими гистидин, фиг. **7B**. Согласно некоторым вариантам осуществления библиотека вариантов, синтезированная с помощью способов, описанных в настоящем документе, не кодирует больше чем 4 остатка гистидина в полученной аминокислотной последовательности.

[00103] В некоторых случаях библиотека вариантов нуклеиновых кислот, созданная с помощью способов, описанных в настоящем документе, обеспечивает экспрессию аминокислотных последовательностей, которые содержат отдельные участки вариации. Матричная аминокислотная последовательность показана на фиг. **8A**. Участок нуклеиновых кислот может характеризоваться только 1 вариантным кодоном в двух участках и при экспрессии может давать в результате аминокислотную последовательность, показанную на фиг. **8B**. Варианты показаны на фиг. **8B** кружками с различной штриховкой, чтобы показать вариацию в аминокислотах в различном положении в одном участке.

[00104] В настоящем документе предусмотрены способы и устройства для синтеза библиотек нуклеиновых кислот с 1, 2, 3 или больше вариантами кодонов, причем вариант для каждого сайта является селективно контролируемым. Отношение двух аминокислот для варианта с одним сайтом может составлять около 1:100, 1:50, 1:10, 1:5, 1:3, 1:2, 1:1. Отношение трех аминокислот для варианта с одним сайтом может составлять около 1:1:100, 1:1:50, 1:1:20, 1:1:10, 1:1:5, 1:1:3, 1:1:2, 1:1:1, 1:10:10, 1:5:5, 1:3:3, или 1:2:2. На фиг. **9A** показана эталонная аминокислотная последовательность дикого типа, кодируемая последовательностью нуклеиновой кислоты дикого типа. На фиг. **9B** показана библиотека вариантов аминокислот, причем каждый вариант содержит участок последовательности (показанный кружками со штриховкой), причем каждое положение может характеризоваться определенным отношением аминокислот в полученной библиотеке вариантных белков. Полученная библиотека вариантных белков кодируется библиотекой вариантных нуклеиновых кислот, созданной с помощью способов, описанных в настоящем документе. На этой иллюстрации 5 положений подвергают вариации: первое положение **900** характеризуется отношением K/R, составляющим 50/50, второе положение **910** характеризуется отношением V/L/S, составляющим 50/25/25, третье положение **920** характеризуется отношением Y/R/D, составляющим 50/25/25, четвертое положение **930** характеризуется равным отношением для всех 20 аминокислот, и пятое положение **940** характеризуется отношением для G/P, составляющим 75/25. Описанные в настоящем документе отношения являются только иллюстративными.

[00105] *Вариация в экспрессионных кассетах*

[00106] В некоторых случаях создают синтезированную библиотеку вариантов, которая кодирует часть экспрессионного конструкта. Иллюстративные части экспрессионного конструкта включают в себя промотор, открытую рамку считывания и область терминации. В некоторых случаях экспрессионный конструкт кодирует одну, две, три или больше экспрессионных кассет. Можно создать библиотеку полинуклеотидов, кодирующую вариацию кодонов на одном сайте или нескольких сайтах, в отдельных областях, которые составляют части экспрессионной кассеты конструкта, как показано на фиг. 11. Для создания двух экспрессионных кассет конструкта синтезировали варианты нуклеиновых кислоты, кодирующие по меньшей мере часть вариантной последовательности первого промотора **1110**, первой открытой рамки считывания **1120**, первого терминатора **1130**, второго промотора **1140**, второй открытой рамки считывания **1150** или последовательности второго терминатора **1160**. После раундов амплификации, как описано в предыдущих примерах, создают библиотеку из 1024 экспрессионных конструктов. На фиг. 11 показан только один пример конфигурации. В некоторых случаях дополнительные регуляторные последовательности, такие как нетранслируемая регуляторная область (UTR) или область энхансера, также включены в экспрессионную кассету, описанную в настоящем документе. Экспрессионная кассета может содержать 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 или больше компонентов, для которых варианты последовательности создают с помощью способов, описанных в настоящем документе. В некоторых случаях экспрессионный конструкт содержит больше чем один ген в мультицистронном векторе. В одном примере синтезированные нуклеиновые кислоты ДНК вводят в вирусные векторы (например, лентивирусные) и затем упаковывают для трансдукции в клетки, или невирусные векторы для переноса в клетки с последующим скринингом и анализом.

[00107] Экспрессионные векторы для вставки нуклеиновых кислот, раскрытых в настоящем документе, содержат клетки млекопитающих, например, человека, не являющегося человеком примата, свиньи, кролика и мыши. Иллюстративные экспрессионные векторы включают в себя без ограничения экспрессионные векторы млекопитающих: pSF-CMV-NEO-NH2-PPT-3XFLAG, pSF-CMV-NEO-COOH-3XFLAG, pSF-CMV-PURO-NH2-GST-TEV, pSF-OXB20-COOH-TEV-FLAG(R)-6His, pCEP4 pDEST27, pSF-CMV-Ub-KrYFP, pSF-CMV-FMDV-daGFP, вектор pEF1a-mCherry-N1, вектор pEF1a-tdTomato, pSF-CMV-FMDV-Hygro, pSF-CMV-PGK-Puro, pMCP-tag(m) и pSF-CMV-PURO-NH2-CMYC. Нуклеиновыми кислотами, синтезированными с помощью способов, описанных в настоящем документе, можно трансфицировать клетки с помощью различных способов, известных в настоящей области техники, включая в себя без ограничения трансфекцию, трансдукцию и электропорацию. Иллюстративные

исследуемые клеточные функции включают в себя без ограничения изменения в клеточной пролиферации, смерти, миграции/адгезии, метаболической и сигнальной активности.

[00108] Высокопараллельный синтез нуклеиновых кислот

[00109] В настоящем документе предусмотрен подход на основе платформы, в котором используют миниатюризацию, распараллеливание и вертикальную интеграцию сквозного процесса от синтеза нуклеиновых кислот до сборки генов в нанолунках на кремнии для создания революционной платформы синтеза. Описанные в настоящем документе устройства обеспечивают, с той же площадью основания, что и у 96-луночного планшета, кремниевую платформу синтеза, способную увеличивать производительность в 1000 раз или больше по сравнению с традиционными способами синтеза с получением вплоть до приблизительно 1000000 или больше нуклеиновых кислот или 10000 или больше генов за один высокопараллельный прогон.

[00110] С появлением секвенирования нового поколения геномные данные высокого разрешения стали важным фактором для исследований, которые углубляются в биологические роли различных генов как в нормальной биологии, так и в патогенезе заболеваний. В основе этого исследования лежит центральная догма молекулярной биологии и концепция “переноса информации о последовательности остаток за остатком”. Геномная информация, закодированная в ДНК, транскрибируется в сообщение, которое затем транслируется в белок, который представляет собой активный продукт в пределах данного биологического пути.

[00111] Другой интересной областью исследований является открытие, разработка и производство терапевтических молекул, сфокусированных на высокоспецифичной клеточной мишени. Библиотеки последовательностей ДНК с большим разнообразием лежат в основе портфеля разрабатываемых препаратов для целенаправленной терапии. Генные мутанты используют для экспрессии белков в цикле конструирования, сборки и испытания белковой инженерии, который в идеале завершается оптимизированным геном для высокой экспрессии белка с высокой аффинностью к его терапевтической мишени. В качестве примера рассмотрим связывающий карман рецептора. Возможность одновременного испытания всех перестановок последовательностей всех остатков в связывающем кармане позволит провести тщательное исследование, увеличивая шансы на успех. Насыщающий мутагенез, при котором исследователь пытается создать все возможные мутации в определенном сайте в пределах рецептора, представляет собой один из подходов к этой проблеме разработки. Несмотря на то, что это дорого, требует много времени и является трудоемким, это позволяет вводить каждый вариант в каждое положение. Напротив, комбинаторный мутагенез, где несколько выбранных положений

или короткий участок ДНК можно модифицировать в значительной степени, создает неполный репертуар вариантов со смещенным представлением.

[00112] Для ускорения разработки портфеля лекарственных средств библиотека с желаемыми вариантами, доступными с требуемой частотой в правильном положении, доступная для испытания, иными словами, прецизионная библиотека, позволяет снизить затраты, а также длительность цикла для скрининга. В настоящем документе предусмотрены способы синтеза синтетических библиотек вариантных нуклеиновых кислот, которые обеспечивают точное введение каждого предполагаемого варианта с желаемой частотой. Для конечного пользователя это означает возможность не только тщательно анализировать пространство последовательностей, но и эффективно запрашивать эти гипотезы, снижая стоимость и время скрининга. Общегеномное редактирование может объяснить важные пути, библиотеки, в которой каждый вариант и перестановку последовательностей можно исследовать в отношении оптимальной функциональности, и тысячи генов можно использовать для реконструкции целых путей и геномов, чтобы переконструировать биологические системы для обнаружения лекарственных средств.

[00113] В первом примере само лекарственное средство можно оптимизировать с использованием способов, описанных в настоящем документе. Например, для улучшения конкретной функции антитела конструируют и синтезируют библиотеку вариантных нуклеиновых кислот, кодирующую часть антитела. Библиотеку вариантных нуклеиновых кислот для антитела можно затем создавать с помощью способов, описанных в настоящем документе (например, с помощью ПЦР-мутагенеза с последующей вставкой в вектор). Затем антитело экспрессируется в клеточной линии-продуценте и его подвергают скринингу в отношении усиленной активности. Примеры скрининга включают в себя изучение модуляции аффинности связывания с антигеном, стабильности или эффекторной функции (например, ADCC, функции комплемента или апоптоза). Иллюстративные области для оптимизации антитела включают в себя без ограничения область Fc, область Fab, переменную область области Fab, константную область области Fab, переменный домен тяжелой цепи или легкой цепи (VH или VL) и определяющие специфическую комплементарность области (CDR) VH или VL.

[00114] Библиотеки нуклеиновых кислот, синтезированные с помощью описанных в настоящем документе способов, могут экспрессироваться в различных клетках, ассоциированных с болезненным состоянием. Клетки, ассоциированные с болезненным состоянием, включают в себя клеточные линии, образцы тканей, первичные клетки от субъекта, культивируемые клетки, выделенные от субъекта, или клетки в

модельной системе. Иллюстративные модельные системы включают в себя без ограничения модели болезненного состояния с использованием растений и животных.

[00115] Для идентификации вариантной молекулы, связанной с профилактикой, уменьшением или лечением болезненного состояния, описанная в настоящем документе библиотека вариантных нуклеиновых кислот экспрессируется в клетке, ассоциированной с болезненным состоянием, или в клетке, в которой может быть вызвано болезненное состояние. В некоторых случаях средство используют, чтобы вызвать болезненное состояние в клетках. Иллюстративные инструменты для индукции болезненного состояния включают в себя без ограничения систему рекомбинации Cre/Lox, индукцию воспаления LPS и стрептозотоцин для индукции гипогликемии. Клетки, ассоциированные с болезненным состоянием, могут представлять собой клетки из модельной системы или культивируемые клетки, а также клетки от субъекта с конкретным болезненным состоянием. Иллюстративные болезненные состояния включают в себя бактериальное, грибковое, вирусное, аутоиммунное или пролиферативное нарушение (например, рак). В некоторых случаях библиотека вариантных нуклеиновых кислот экспрессируется в модельной системе, клеточной линии или первичных клетках, полученных от субъекта, и ее подвергают скринингу в отношении изменений по меньшей мере одной клеточной активности. Иллюстративные клеточные активности включают в себя без ограничения пролиферацию, прогрессию клеточного цикла, клеточную смерть, адгезию, миграцию, размножение, клеточную передачу сигналов, выработку энергии, использование кислорода, метаболическую активность и старение, ответ на повреждение свободными радикалами или любую их комбинацию.

[00116] Субстраты

[00117] В настоящем документе предусмотрены субстраты, содержащие множество кластеров, причем каждый кластер содержит множество локусов, которые поддерживают прикрепление и синтез нуклеиновых кислот. Используемый в настоящем документе термин “локус” относится к отдельной области на структуре, которая обеспечивает подложку для полинуклеотидов, кодирующих одну предварительно определенную последовательность, удлиняющуюся от поверхности. В некоторых случаях локус находится на двухмерной поверхности, например, по существу плоской поверхности. В некоторых случаях локус относится к отдельному поднятому или опущенному участку на поверхности, например, лунке, микролунке, каналу или стержню. В некоторых случаях поверхность локуса содержит материал, который активно функционализирован для прикрепления по меньшей мере к одному нуклеотиду для синтеза нуклеиновых кислот, или предпочтительно, к популяции идентичных нуклеотидов для синтеза популяции

нуклеиновых кислот. В некоторых случаях нуклеиновая кислота относится к популяции нуклеиновых кислот, кодирующих одну и ту же последовательность нуклеиновой кислоты. В некоторых случаях поверхность устройства включает в себя одну или множество поверхностей субстрата.

[00118] Средние показатели частоты ошибок для нуклеиновых кислот, синтезированных в пределах библиотеки с использованием предусмотренных систем и способов, могут составлять меньше чем 1 на 1000, меньше чем 1 на 1250, меньше чем 1 на 1500, меньше чем 1 на 2000, меньше чем 1 на 3000 или реже. В некоторых случаях средние показатели частоты ошибок для нуклеиновых кислот, синтезированных в пределах библиотеки с использованием предусмотренных систем и способов, составляют меньше чем 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1100, 1/1200, 1/1250, 1/1300, 1/1400, 1/1500, 1/1600, 1/1700, 1/1800, 1/1900, 1/2000, 1/3000 или меньше. В некоторых случаях средние показатели частоты ошибок для нуклеиновых кислот, синтезированных в пределах библиотеки с использованием предусмотренных систем и способов, составляют меньше чем 1/1000.

[00119] В некоторых случаях показатели суммарной частоты ошибок для нуклеиновых кислот, синтезированных в пределах библиотеки с использованием предусмотренных систем и способов, составляют меньше чем 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1100, 1/1200, 1/1250, 1/1300, 1/1400, 1/1500, 1/1600, 1/1700, 1/1800, 1/1900, 1/2000, 1/3000 или меньше по сравнению с предварительно определенными последовательностями. В некоторых случаях показатели суммарной частоты ошибок для нуклеиновых кислот, синтезированных в пределах библиотеки с использованием предусмотренных систем и способов, составляют меньше чем 1/500, 1/600, 1/700, 1/800, 1/900 или 1/1000. В некоторых случаях показатели суммарной частоты ошибок для нуклеиновых кислот, синтезированных в пределах библиотеки с использованием предусмотренных систем и способов, составляют меньше чем 1/1000.

[00120] В некоторых случаях фермент для исправления ошибок можно использовать для нуклеиновых кислот, синтезированных в пределах библиотеки с использованием предусмотренных систем и способов. В некоторых случаях показатели суммарной частоты ошибок для нуклеиновых кислот с исправлением ошибок могут составлять меньше чем 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1100, 1/1200, 1/1300, 1/1400, 1/1500, 1/1600, 1/1700, 1/1800, 1/1900, 1/2000, 1/3000 или меньше по сравнению с предварительно определенными последовательностями. В некоторых случаях показатели суммарной частоты ошибок с исправлением ошибок для нуклеиновых кислот, синтезированных в пределах библиотеки с использованием предусмотренных систем и

способов, могут составлять меньше чем 1/500, 1/600, 1/700, 1/800, 1/900 или 1/1000. В некоторых случаях показатели суммарной частоты ошибок с исправлением ошибок для нуклеиновых кислот, синтезированных в пределах библиотеки с использованием предусмотренных систем и способов, могут составлять меньше чем 1/1000.

[00121] Частота ошибок может ограничивать ценность синтеза генов для производства библиотек вариантов генов. С частотой ошибок, составляющей 1/300, около 0,7% клонов в гене из 1500 пар оснований будут правильными. Поскольку большинство ошибок синтеза полинуклеотидов приводят к мутациям со сдвигом рамки считывания, более 99% клонов в такой библиотеке не производят полноразмерный белок. Снижение частоты ошибок на 75% увеличило бы долю правильных клонов в 40 раз. Способы и композиции согласно настоящему раскрытию позволяют быстро синтезировать новые полинуклеотиды и генные библиотеки *de novo* с показателями частоты ошибок, которые ниже, чем обычно наблюдаются для способов синтеза генов, как из-за улучшенного качества синтеза, так и из-за применимости способов исправления ошибок, которые обеспечивают возможность осуществления в значительной степени параллельного и эффективного по времени способа. Соответственно можно синтезировать библиотеки со вставкой, делецией, заменой оснований или общими показателями частоты ошибок, которые составляют ниже 1/300, 1/400, 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1250, 1/1500, 1/2000, 1/2500, 1/3000, 1/4000, 1/5000, 1/6000, 1/7000, 1/8000, 1/9000, 1/10000, 1/12000, 1/15000, 1/20000, 1/25000, 1/30000, 1/40000, 1/50000, 1/60000, 1/70000, 1/80000, 1/90000, 1/100000, 1/125000, 1/150000, 1/200000, 1/300000, 1/400000, 1/500000, 1/600000, 1/700000, 1/800000, 1/900000, 1/1000000 или меньше, во всей библиотеке или в больше чем 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%, 99%, 99,5%, 99,8%, 99,9%, 99,95%, 99,98%, 99,99% библиотеки или больше. Способы и композиции согласно настоящему раскрытию дополнительно относятся к большим библиотекам синтетических полинуклеотидов и генов с низкими показателями частоты ошибок, ассоциированными по меньшей мере с 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%, 99%, 99,5%, 99,8%, 99,9%, 99,95%, 99,98%, 99,99% или больше полинуклеотидов или генов по меньшей мере в подмножестве библиотеки, при соотнесении с последовательностями без ошибок по сравнению с предварительно определенной/предварительно выбранной последовательностью. В некоторых случаях по меньшей мере 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%, 99%, 99,5%, 99,8%, 99,9%, 99,95%, 99,98%, 99,99% или больше полинуклеотидов или генов в выделенном объеме в пределах библиотека характеризуются одинаковой последовательностью. В некоторых случаях по меньшей мере 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%,

99%, 99,5%, 99,8%, 99,9%, 99,95%, 99,98%, 99,99% или больше из любых полинуклеотидов или генов, характеризующихся больше чем 95%, 96%, 97%, 98%, 99%, 99,5%, 99,6%, 99,7%, 99,8%, 99,9% или больше сходством или идентичностью, характеризуются одинаковой последовательностью. В некоторых случаях частоту ошибок, относящуюся к заданному локусу на полинуклеотиде или гене, оптимизируют. Таким образом, каждый из данного локуса или множества выбранных локусов одного или нескольких полинуклеотидов или генов как части большой библиотеки может характеризоваться частотой ошибок, которая составляет меньше чем 1/300, 1/400, 1/500, 1/600, 1/700, 1/800, 1/900, 1/1000, 1/1250, 1/1500, 1/2000, 1/2500, 1/3000, 1/4000, 1/5000, 1/6000, 1/7000, 1/8000, 1/9000, 1/10000, 1/12000, 1/15000, 1/20000, 1/25000, 1/30000, 1/40000, 1/50000, 1/60000, 1/70000, 1/80000, 1/90000, 1/100000, 1/125000, 1/150000, 1/200000, 1/300000, 1/400000, 1/500000, 1/600000, 1/700000, 1/800000, 1/900000, 1/1000000 или меньше. В разных случаях такие оптимизированные по ошибкам локусы могут содержать по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 30000, 50000, 75000, 100000, 500000, 1000000, 2000000, 3000000 или больше локусов. Оптимизированные по ошибкам локусы могут быть распределены по меньшей мере на 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 30000, 75000, 100000, 500000, 1000000, 2000000, 3000000 или больше полинуклеотидов или генов.

[00122] Показатели частоты ошибок могут быть достигнуты с исправлением ошибок или без нее. Показатели частоты ошибок могут быть достигнуты во всей библиотеке или в больше чем 80%, 85%, 90%, 93%, 95%, 96%, 97%, 98%, 99%, 99,5%, 99,8%, 99,9%, 99,95%, 99,98%, 99,99% библиотеки или больше.

[00123] В настоящем документе предусмотрены структуры, которые могут содержать поверхность, которая поддерживает синтез множества нуклеиновых кислот, характеризующихся различными предварительно определенными последовательностями на адресуемых положениях на общей подложке. В некоторых случаях устройство обеспечивает подложку для синтеза больше чем 2000; 5000; 10000; 20000; 30000; 50000; 75000; 100000; 200000; 300000; 400000; 500000; 600000; 700000; 800000; 900000; 1000000; 1200000; 1400000; 1600000; 1800000; 2000000; 2500000; 3000000; 3500000; 4000000; 4500000; 5000000; 10000000 или больше неидентичных нуклеиновых кислот. В некоторых случаях устройство обеспечивает подложку для синтеза больше чем 2000; 5000; 10000; 20000; 30000; 50000; 75000; 100000; 200000; 300000; 400000; 500000; 600000; 700000;

800000; 900000; 1000000; 1200000; 1400000; 1600000; 1800000; 2000000; 2500000; 3000000; 3500000; 4000000; 4500000; 5000000; 10000000 или больше нуклеиновых кислот, кодирующих отдельные последовательности. В некоторых случаях по меньшей мере часть нуклеиновых кислот содержит идентичные последовательность или может быть синтезирована с идентичной последовательностью.

[00124] В настоящем документе предусмотрены способы и устройства для получения и роста нуклеиновых кислот, составляющих около 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900 или 2000 оснований в длину. В некоторых случаях длина образованной нуклеиновой кислоты составляет около 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200 или 225 оснований в длину. Нуклеиновая кислота может составлять по меньшей мере 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 оснований в длину. Нуклеиновая кислота может составлять 10 - 225 оснований в длину, 12 - 100 оснований в длину, 20 - 150 оснований в длину, 20 - 130 оснований в длину или 30 - 100 оснований в длину.

[00125] В некоторых случаях нуклеиновые кислоты синтезируют на отдельных локусах субстрата, причем каждый локус поддерживает синтез популяции нуклеиновых кислот. В некоторых случаях каждый локус поддерживает синтез популяции нуклеиновых кислот с последовательностью, отличной от популяции нуклеиновых кислот, выращенных на другом локусе. В некоторых случаях локусы устройства расположены в пределах множества кластеров. В некоторых случаях устройство содержит по меньшей мере 10, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 11000, 12000, 13000, 14000, 15000, 20000, 30000, 40000, 50000 или больше кластеров. В некоторых случаях устройство содержит больше чем 2000; 5000; 10000; 100000; 200000; 300000; 400000; 500000; 600000; 700000; 800000; 900000; 1000000; 1100000; 1200000; 1300000; 1400000; 1500000; 1600000; 1700000; 1800000; 1900000; 2000000; 300000; 400000; 500000; 600000; 700000; 800000; 900000; 1000000; 1200000; 1400000; 1600000; 1800000; 2000000; 2500000; 3000000; 3500000; 4000000; 4500000; 5000000 или 10000000 или больше отдельных локусов. В некоторых случаях устройство содержит около 10000 отдельных локусов. Количество локусов в пределах одного кластера варьируется в разных случаях. В некоторых случаях каждый кластер включает в себя 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 130, 150, 200, 300, 400, 500 или больше локусов. В некоторых случаях каждый кластер включает в себя около 50-500 локусов. В некоторых случаях каждый кластер включает в себя около 100-200 локусов. В некоторых случаях каждый кластер включает в себя около 100-150 локусов. В некоторых случаях каждый кластер включает в себя около 109, 121, 130 или 137

локусов. В некоторых случаях каждый кластер включает в себя около 19, 20, 61, 64 или больше локусов.

[00126] Число отдельных нуклеиновых кислот, синтезированных на устройстве, может зависеть от числа отдельных локусов, доступных в субстрате. В некоторых случаях плотность локусов в пределах кластера устройства составляет по меньшей мере или около 1 locus на мм^2 , 10 локусов на мм^2 , 25 локусов на мм^2 , 50 локусов на мм^2 , 65 локусов на мм^2 , 75 локусов на мм^2 , 100 локусов на мм^2 , 130 локусов на мм^2 , 150 локусов на мм^2 , 175 локусов на мм^2 , 200 локусов на мм^2 , 300 локусов на мм^2 , 400 локусов на мм^2 , 500 локусов на мм^2 , 1000 локусов на мм^2 или больше. В некоторых случаях устройство содержит около 10 локусов на мм^2 - около 500 мм^2 , около 25 локусов на мм^2 - около 400 мм^2 , около 50 локусов на мм^2 - около 500 мм^2 , около 100 локусов на мм^2 - около 500 мм^2 , около 150 локусов на мм^2 - около 500 мм^2 , около 10 локусов на мм^2 - около 250 мм^2 , около 50 локусов на мм^2 - около 250 мм^2 , около 10 локусов на мм^2 - около 200 мм^2 или около 50 локусов на мм^2 - около 200 мм^2 . В некоторых случаях расстояние от центров двух смежных локусов в пределах кластера составляет около 10 мкм - около 500 мкм, около 10 мкм - около 200 мкм или около 10 мкм - около 100 мкм. В некоторых случаях расстояние от двух центров смежных локусов составляет больше чем около 10 мкм, 20 мкм, 30 мкм, 40 мкм, 50 мкм, 60 мкм, 70 мкм, 80 мкм, 90 мкм или 100 мкм. В некоторых случаях расстояние от центров двух смежных локусов составляет меньше чем около 200 мкм, 150 мкм, 100 мкм, 80 мкм, 70 мкм, 60 мкм, 50 мкм, 40 мкм, 30 мкм, 20 мкм или 10 мкм. В некоторых случаях каждый locus характеризуется шириной, составляющей около 0,5 мкм, 1 мкм, 2 мкм, 3 мкм, 4 мкм, 5 мкм, 6 мкм, 7 мкм, 8 мкм, 9 мкм, 10 мкм, 20 мкм, 30 мкм, 40 мкм, 50 мкм, 60 мкм, 70 мкм, 80 мкм, 90 мкм или 100 мкм. В некоторых случаях каждый locus характеризуется шириной, составляющей около 0,5 мкм - 100 мкм, около 0,5 мкм - 50 мкм, около 10 мкм - 75 мкм или около 0,5 мкм - 50 мкм.

[00127] В некоторых случаях плотность кластеров в пределах устройства составляет по меньшей мере или около 1 кластер на 100 мм^2 , 1 кластер на 10 мм^2 , 1 кластер на 5 мм^2 , 1 кластер на 4 мм^2 , 1 кластер на 3 мм^2 , 1 кластер на 2 мм^2 , 1 кластер на 1 мм^2 , 2 кластера на 1 мм^2 , 3 кластера на 1 мм^2 , 4 кластера на 1 мм^2 , 5 кластеров на 1 мм^2 , 10 кластеров на 1 мм^2 , 50 кластеров на 1 мм^2 или больше. В некоторых случаях устройство содержит около 1 кластер на 10 мм^2 - около 10 кластеров на 1 мм^2 . В некоторых случаях расстояние от центров двух смежных кластеров составляет меньше чем около 50 мкм, 100 мкм, 200 мкм, 500 мкм, 1000 мкм, или 2000 мкм, или 5000 мкм. В некоторых случаях расстояние от центров двух смежных кластеров составляет около 50 мкм - около 100 мкм, около 50 мкм - около 200 мкм, около 50 мкм - около 300 мкм, около 50 мкм - около 500 мкм

и около 100 мкм - около 2000 мкм. В некоторых случаях расстояние от центров двух смежных кластеров составляет около 0,05 мм - около 50 мм, около 0,05 мм - около 10 мм, около 0,05 мм - около 5 мм, около 0,05 мм - около 4 мм, около 0,05 мм - около 3 мм, около 0,05 мм - около 2 мм, около 0,1 мм - 10 мм, около 0,2 мм - 10 мм, около 0,3 мм - около 10 мм, около 0,4 мм - около 10 мм, около 0,5 мм - 10 мм, около 0,5 мм - около 5 мм или около 0,5 мм - около 2 мм. В некоторых случаях каждый кластер характеризуется диаметром или шириной вдоль одного измерения, составляющими около 0,5 - 2 мм, около 0,5 - 1 мм или около 1 - 2 мм. В некоторых случаях каждый кластер характеризуется диаметром или шириной вдоль одного измерения, составляющими около 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,1, 1,2, 1,3, 1,4, 1,5, 1,6, 1,7, 1,8, 1,9 или 2 мм. В некоторых случаях каждый кластер характеризуется внутренним диаметром или шириной вдоль одного измерения, составляющими около 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,1, 1,15, 1,2, 1,3, 1,4, 1,5, 1,6, 1,7, 1,8, 1,9 или 2 мм.

[00128] Устройство может характеризоваться около размером стандартного 96-луночного планшета, например, около 100 - 200 мм на около 50 - 150 мм. В некоторых случаях устройство характеризуется диаметром меньше чем или равным около 1000 мм, 500 мм, 450 мм, 400 мм, 300 мм, 250 мм, 200 мм, 150 мм, 100 мм или 50 мм. В некоторых случаях диаметр устройства составляет около 25 мм - 1000 мм, около 25 мм - около 800 мм, около 25 мм - около 600 мм, около 25 мм - около 500 мм, около 25 мм - около 400 мм, около 25 мм - около 300 мм или около 25 мм - около 200. Неограничивающие примеры размера устройства включают в себя около 300 мм, 200 мм, 150 мм, 130 мм, 100 мм, 76 мм, 51 мм и 25 мм. В некоторых случаях устройство характеризуется площадью плоской поверхности, составляющей по меньшей мере около 100 мм²; 200 мм²; 500 мм²; 1000 мм²; 2000 мм²; 5000 мм²; 10000 мм²; 12000 мм²; 15000 мм²; 20000 мм²; 30000 мм²; 40000 мм²; 50000 мм² или больше. В некоторых случаях толщина устройства составляет около 50 мм - около 2000 мм, около 50 мм - около 1000 мм, около 100 мм - около 1000 мм, около 200 мм - около 1000 мм или около 250 мм - около 1000 мм. Неограничивающие примеры толщины устройства включают в себя 275 мм, 375 мм, 525 мм, 625 мм, 675 мм, 725 мм, 775 мм и 925 мм. В некоторых случаях толщина устройства варьируется вместе с диаметром и зависит от состава субстрата. Например, устройство, содержащее материалы, отличные от кремния, характеризуется другой толщиной, чем кремниевое устройство того же диаметра. Толщину устройства можно определить по механической прочности используемого материала, и устройство должно быть достаточно толстым, чтобы выдерживать свой собственный вес без растрескивания во время обработки. В некоторых случаях структура содержит множество устройств, описанных в настоящем документе.

[00129] Материалы поверхности

[00130] Настоящее изобретение относится к устройству, содержащему поверхность, причем поверхность модифицирована для поддержания синтеза полинуклеотидов в предварительно определенных местах, и полученной в результате этого низкой частоте ошибок, низкой частоте выпадения, высокому выходу и высокому уровню представления олигонуклеотидов. Согласно некоторым вариантам осуществления поверхности устройства для синтеза полинуклеотидов, предусмотренного в настоящем документе, изготовлены из различных материалов, способных к модификации для поддержания реакции синтеза полинуклеотидов *de novo*. В некоторых случаях устройства являются достаточно проводящими, например, способны формировать однородные электрические поля на всем устройстве или на части устройства. Устройство, описанное в настоящем документе, может содержать гибкий материал. Иллюстративные гибкие материалы включают в себя без ограничения модифицированный нейлон, немодифицированный нейлон, нитроцеллюлозу и полипропилен. Описанное в настоящем документе устройство может содержать жесткий материал. Иллюстративные жесткие материалы включают в себя без ограничения стекло, плавкий кремнезем, кремний, диоксид кремния, нитрид кремния, пластмассы (например, политетрафторэтилен, полипропилен, полистирол, поликарбонат и их смеси, а также металлы (например, золото, платина). Раскрытое в настоящем документе устройство может быть изготовлено из материала, содержащего кремний, полистирол, агарозу, декстран, целлюлозные полимеры, полиакриламиды, полидиметилсилоксан (PDMS), стекло или любую их комбинацию. В некоторых случаях раскрытое в настоящем документе устройство изготовлено из комбинации перечисленных в настоящем документе материалов или любого другого подходящего материала, известного в настоящей области техники.

[00131] Перечень показателей прочности при растяжении для иллюстративных материалов, описанных в настоящем документе, представлен следующим образом: нейлон (70 МПа), нитроцеллюлоза (1,5 МПа), полипропилен (40 МПа), кремний (268 МПа), полистирол (40 МПа), агароза (1-10 МПа), полиакриламид (1-10 МПа), полидиметилсилоксан (PDMS) (3,9-10,8 МПа). Твердые подложки, описанные в настоящем документе, могут характеризоваться прочностью при растяжении от 1 до 300, от 1 до 40, от 1 до 10, от 1 до 5 или от 3 до 11 МПа. Твердые подложки, описанные в настоящем документе, могут характеризоваться прочностью при растяжении, составляющей около 1, 1,5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 20, 25, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 270 МПа или больше. В некоторых случаях устройство, описанное в настоящем документе, содержит твердую подложку для синтеза полинуклеотидов, которая находится в форме гибкого

материала, способного храниться в непрерывной петле или бобине, такой как лента или гибкий лист.

[00132] Модуль Юнга измеряет сопротивление материала упругой (восстанавливаемой) деформации под нагрузкой. Перечень модулей Юнга для жесткости иллюстративных материалов, описанных в настоящем документе, представлен следующим образом: нейлон (3 ГПа), нитроцеллюлоза (1,5 ГПа), полипропилен (2 ГПа), кремний (150 ГПа), полистирол (3 ГПа), агароза (1-10 ГПа), полиакриламид (1-10 ГПа), полидиметилсилоксан (PDMS) (1-10 ГПа). Твердые подложки, описанные в настоящем документе, могут характеризоваться модулями Юнга от 1 до 500, от 1 до 40, от 1 до 10, от 1 до 5 или от 3 до 11 ГПа. Твердые носители, описанные в настоящем документе, могут характеризоваться модулями Юнга, составляющими около 1, 1,5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 20, 25, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 400, 500 ГПа или более. Поскольку соотношение между гибкостью и жесткостью обратно пропорционально друг другу, гибкий материал обладает низким модулем Юнга и значительно меняет свою форму под нагрузкой.

[00133] В некоторых случаях раскрытое в настоящем документе устройство содержит основу из диоксида кремния и поверхностный слой из оксида кремния. Альтернативно устройство может характеризоваться основой из оксида кремния. Поверхность устройства, представленного в настоящем документе, может быть текстурированной, что приводит к увеличению общей площади поверхности для синтеза полинуклеотидов. Устройство, раскрытое в настоящем документе, может содержать по меньшей мере 5%, 10%, 25%, 50%, 80%, 90%, 95% или 99% кремния. Раскрытое в настоящем документе устройство может быть изготовлено из пластины со структурой «кремнием на диэлектрике» (SOI).

[00134] Субстраты, устройства и реакторы, предусмотренные в настоящем документе, изготовлены из любого разнообразия материалов, подходящих для способов и композиций, описанных в настоящем документе. В определенных случаях материалы устройства изготовлены так, чтобы проявлять низкий уровень связывания нуклеотидов. В некоторых случаях материалы устройства модифицируют для создания отдельных поверхностей, которые проявляют высокий уровень связывания нуклеотидов. В некоторых случаях материалы устройства являются прозрачными для видимого и/или ультрафиолетового света. В некоторых случаях материалы устройства являются достаточно проводящими, например, способны формировать однородные электрические поля на всем субстрате или на части субстрата. В некоторых случаях проводящие материалы соединены с электрическим заземлением. В некоторых случаях устройство является теплопроводящим или изолированным. В некоторых случаях материалы являются химически стойкими и

жаростойкими, чтобы поддерживать химические или биохимические реакции, например, процессы реакции синтеза нуклеиновых кислот. В некоторых случаях устройство содержит гибкие материалы. Гибкие материалы включают в себя без ограничения модифицированный нейлон, немодифицированный нейлон, нитроцеллюлозу и полипропилен и тому подобное. В некоторых случаях устройство содержит жесткие материалы. Жесткие материалы включают в себя без ограничения стекло, плавкий кремнезем, кремний, диоксид кремния, нитрид кремния, пластмассы (например, политетрафторэтилен, полипропилен, полистирол, поликарбонат и их смеси, а также металлы (например, золото, платина). В некоторых случаях устройство изготовлено из материала, содержащего кремний, полистирол, агарозу, декстран, целлюлозные полимеры, полиакриламиды, полидиметилсилоксан (PDMS), стекло или любую их комбинацию. В некоторых случаях устройство изготовлено из комбинации перечисленных в настоящем документе материалов или любого другого подходящего материала, известного в настоящей области техники.

[00135] Структура поверхности

[00136] В настоящем документе предусмотрены устройства, содержащие приподнятые и/или опущенные элементы. Одним из преимуществ таких элементов является увеличение площади поверхности для поддержания синтеза нуклеиновых кислот. В некоторых случаях устройство, характеризующееся приподнятыми и/или опущенными элементами, называется трехмерным субстратом. В некоторых случаях трехмерное устройство содержит один или несколько каналов. В некоторых случаях один или несколько локусов содержат канал. В некоторых случаях каналы являются доступными для осаждения реагентов посредством устройства для осаждения, такого как синтезатор нуклеиновых кислот. В некоторых случаях реагенты и/или жидкости накапливаются в большей лунке в жидкостной связи с одним или несколькими каналами. Например, устройство содержит множество каналов, соответствующих множеству локусов с кластером, и множество каналов находятся в жидкостной связи с одной лункой кластера. В некоторых способах библиотеку нуклеиновых кислот синтезируют в множестве локусов кластера.

[00137] В некоторых случаях структура выполнена с возможностью обеспечения контролируемых потоков и путей массопереноса для синтеза нуклеиновых кислот на поверхности. В некоторых случаях конфигурация устройства обеспечивает контролируемое и равномерное распределение путей массопереноса, времени химического воздействия и/или эффективности промывки во время синтеза нуклеиновых кислот. В некоторых случаях конфигурация устройства позволяет повысить эффективность

вытеснения, например, путем предоставления достаточного объема для выращивания нуклеиновой кислоты, так чтобы исключенный объем за счет растущей нуклеиновой кислоты не занимал больше чем 50, 45, 40, 35, 30, 25, 20, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1% или меньше от первоначально доступного объема, который доступен или пригоден для выращивания нуклеиновой кислоты. В некоторых случаях трехмерная структура обеспечивает управляемый поток жидкости для быстрой замены химического воздействия.

[00138] В настоящем документе предусмотрены способы синтеза количества ДНК, составляющего 1 фМ, 5 фМ, 10 фМ, 25 фМ, 50 фМ, 75 фМ, 100 фМ, 200 фМ, 300 фМ, 400 фМ, 500 фМ, 600 фМ, 700 фМ, 800 фМ, 900 фМ, 1 пМ, 5 пМ, 10 пМ, 25 пМ, 50 пМ, 75 пМ, 100 пМ, 200 пМ, 300 пМ, 400 пМ, 500 пМ, 600 пМ, 700 пМ, 800 пМ, 900 пМ или больше. В некоторых случаях библиотека полинуклеотидов может занимать длину, составляющую около 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% или 100% гена. Ген может варьироваться вплоть до около 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95% или 100%.

[00139] Неидентичные нуклеиновые кислоты совместно могут кодировать последовательность, составляющую по меньшей мере 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95% или 100% гена. В некоторых случаях нуклеиновая кислота может кодировать последовательность, составляющую 50%, 60%, 70%, 80%, 85%, 90%, 95% или больше гена. В некоторых случаях нуклеиновая кислота может кодировать последовательность, составляющую 80%, 85%, 90%, 95% гена или больше.

[00140] В некоторых случаях сегрегация достигается за счет физической структуры. В некоторых случаях сегрегация достигается путем дифференциальной функционализации поверхности, создавая активные и пассивные области для синтеза нуклеиновых кислот. Дифференциальная функционализация также достигается путем изменения гидрофобности по всей поверхности устройства, создавая тем самым эффекты угла контакта с водой, которые вызывают образование микрошариков или смачивание нанесенных реагентов. Использование более крупных структур может уменьшить разбрызгивание и перекрестное загрязнение отдельных участков синтеза нуклеиновых кислот реагентами соседних пятен. В некоторых случаях устройство, такое как синтезатор нуклеиновых кислот, используют для размещения реагентов в различных местах синтеза нуклеиновых кислот. Субстраты, характеризующиеся трехмерными признаками, сконфигурированы так, чтобы обеспечить возможность синтеза большого количества нуклеиновых кислот (например, больше чем около 10000) с низкой частотой ошибок (например, меньше чем около 1:500, 1:1000, 1:1500, 1:2000; 1:3000; 1:5000 или 1:10000). В

некоторых случаях устройство содержит элементы с плотностью, составляющей около или больше чем около 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 300, 400 или 500 элементов на мм².

[00141] Лунка устройства может характеризоваться одинаковой или отличающейся шириной, высотой и/или объемом по сравнению с другой лункой субстрата. Канал устройства может характеризоваться одинаковой или отличающейся шириной, высотой и/или объемом по сравнению с другим каналом субстрата. В некоторых случаях ширина кластера составляет около 0,05 мм - около 50 мм, около 0,05 мм - около 10 мм, около 0,05 мм - около 5 мм, около 0,05 мм - около 4 мм, около 0,05 мм - около 3 мм, около 0,05 мм - около 2 мм, около 0,05 мм - около 1 мм, около 0,05 мм - около 0,5 мм, около 0,05 мм - около 0,1 мм, около 0,1 мм - 10 мм, около 0,2 мм - 10 мм, около 0,3 мм - около 10 мм, около 0,4 мм - около 10 мм, около 0,5 мм - 10 мм, около 0,5 мм - около 5 мм или около 0,5 мм - около 2 мм. В некоторых случаях ширина лунки, содержащей кластер, составляет около 0,05 мм - около 50 мм, около 0,05 мм - около 10 мм, около 0,05 мм - около 5 мм, около 0,05 мм - около 4 мм, около 0,05 мм - около 3 мм, около 0,05 мм - около 2 мм, около 0,05 мм - около 1 мм, около 0,05 мм - около 0,5 мм, около 0,05 мм - около 0,1 мм, около 0,1 мм - 10 мм, около 0,2 мм - 10 мм, около 0,3 мм - около 10 мм, около 0,4 мм - около 10 мм, около 0,5 мм - 10 мм, около 0,5 мм - около 5 мм или около 0,5 мм - около 2 мм. В некоторых случаях ширина кластера составляет меньше чем или около 5 мм, 4 мм, 3 мм, 2 мм, 1 мм, 0,5 мм, 0,1 мм, 0,09 мм, 0,08 мм, 0,07 мм, 0,06 мм или 0,05 мм. В некоторых случаях ширина кластера составляет около 1,0 - 1,3 мм. В некоторых случаях ширина кластера составляет около 1,150 мм. В некоторых случаях ширина лунки составляет меньше чем или около 5 мм, 4 мм, 3 мм, 2 мм, 1 мм, 0,5 мм, 0,1 мм, 0,09 мм, 0,08 мм, 0,07 мм, 0,06 мм или 0,05 мм. В некоторых случаях ширина лунки составляет около 1,0 - 1,3 мм. В некоторых случаях ширина лунки составляет около 1,150 мм. В некоторых случаях ширина кластера составляет около 0,08 мм. В некоторых случаях ширина лунки составляет около 0,08 мм. Ширина кластера может относиться к кластерам в пределах двумерного или трехмерного субстрата.

[00142] В некоторых случаях высота лунки составляет около 20 мкм - около 1000 мкм, около 50 мкм - около 1000 мкм, около 100 мкм - около 1000 мкм, около 200 мкм - около 1000 мкм, около 300 мкм - около 1000 мкм, около 400 мкм - около 1000 мкм или около 500 мкм - около 1000 мкм. В некоторых случаях высота лунки составляет меньше чем около 1000 мкм, меньше чем около 900 мкм, меньше чем около 800 мкм, меньше чем около 700 мкм или меньше чем около 600 мкм.

[00143] В некоторых случаях устройство содержит множество каналов, соответствующих множеству локусов в пределах кластера, причем высота или глубина

канала составляет около 5 мкм - около 500 мкм, около 5 мкм - около 400 мкм, около 5 мкм - около 300 мкм, около 5 мкм - около 200 мкм, около 5 мкм - около 100 мкм, около 5 мкм - около 50 мкм или около 10 мкм - около 50 мкм. В некоторых случаях высота канала составляет меньше чем 100 мкм, меньше чем 80 мкм, меньше чем 60 мкм, меньше чем 40 мкм или меньше чем 20 мкм.

[00144] В некоторых случаях диаметр канала, локуса (например, в по существу плоском субстрате) или как канала, так и локуса (например, в трехмерном устройстве, в котором локус соответствует каналу) составляет около 1 мкм - около 1000 мкм, около 1 мкм - около 500 мкм, около 1 мкм - около 200 мкм, около 1 мкм - около 100 мкм, около 5 мкм - около 100 мкм или около 10 мкм - около 100 мкм, например, около 90 мкм, 80 мкм, 70 мкм, 60 мкм, 50 мкм, 40 мкм, 30 мкм, 20 мкм или 10 мкм. В некоторых случаях диаметр канала, локуса или как канала, так и локуса составляет меньше чем около 100 мкм, 90 мкм, 80 мкм, 70 мкм, 60 мкм, 50 мкм, 40 мкм, 30 мкм, 20 мкм или 10 мкм. В некоторых случаях расстояние от центра двух смежных каналов, локусов или каналов и локусов составляет около 1 мкм - около 500 мкм, около 1 мкм - около 200 мкм, около 1 мкм - около 100 мкм, около 5 мкм - около 200 мкм, около 5 мкм - около 100 мкм, около 5 мкм - около 50 мкм или около 5 мкм - около 30 мкм, например, около 20 мкм.

[00145] Модификации поверхности

[00146] В различных случаях модификации поверхности используют для химического и/или физического изменения поверхности с помощью аддитивного или субтрактивного процесса для изменения одно или несколько химических и/или физических свойств поверхности устройства или выбранного участка или области поверхности устройства. Например, модификации поверхности включают в себя без ограничения следующее: (1) изменение смачивающих свойств поверхности, (2) функционализация поверхности, т.е. обеспечение наличия, модификация или замена поверхностных функциональных групп, (3) дефункционализация поверхности, т.е. удаление поверхностных функциональных групп, (4) иное изменение химического состава поверхности, например, путем травления, (5) увеличение или уменьшение шероховатости поверхности, (6) нанесение покрытия на поверхность, например, покрытия, которое проявляет смачивающие свойства, которые отличаются от смачивающих свойств поверхности и/или (7) осаждение частиц на поверхности.

[00147] В некоторых случаях добавление химического слоя поверх поверхности (называемого усилитель адгезии) облегчает структурированную компоновку локусов на поверхности субстрата. Иллюстративные поверхности для применения усиления адгезии включают в себя без ограничения стекло, кремний, диоксид кремния и

нитрид кремния. В некоторых случаях усилитель адгезии представляет собой химическое вещество с высокой поверхностной энергией. В некоторых случаях второй химический слой осаждают на поверхность субстрата. В некоторых случаях второй химический слой характеризуется низкой поверхностной энергией. В некоторых случаях поверхностная энергия химического слоя, осажженного на поверхность, поддерживает локализацию капель на поверхности. В зависимости от выбранного расположения компоновки близость локусов и/или площадь контакта жидкости в локусах могут изменяться.

[00148] В некоторых случаях поверхность устройства или выделенные локусы, на которые осаждают нуклеиновые кислоты или другие фрагменты, например, для синтеза полинуклеотидов, являются гладкими или по существу плоскими (например, двумерными) или имеют неровности, такие как приподнятые или опущенные элементы (например, трехмерные элементы). В некоторых случаях поверхность устройства модифицируют одним или несколькими различными слоями соединений. Такие модифицирующие слои, представляющие интерес, включают в себя без ограничения неорганические и органические слои, такие как металлы, оксиды металлов, полимеры, небольшие органические молекулы и тому подобное. Неограничивающие полимерные слои включают в себя пептиды, белки, нуклеиновые кислоты или их миметики (например, пептидные нуклеиновые кислоты и тому подобное), полисахариды, фосфолипиды, полиуретаны, сложные полиэфиры, поликарбонаты, полимочевины, полиамиды, полиэтиленамины, полиарилсульфиды, полисилоксаны, полиимиды, полиацетаты и любые другие подходящие соединения, описанные в настоящем документе или известные в настоящей области техники. В некоторых случаях полимеры являются гетерополимерными. В некоторых случаях полимеры являются гомополимерными. В некоторых случаях полимеры содержат функциональные фрагменты или являются конъюгированными.

[00149] В некоторых случаях выделенные локусы устройства функционализируют одним или несколькими фрагментами, которые увеличивают и/или уменьшают поверхностную энергию. В некоторых случаях фрагмент является химически инертным. В некоторых случаях фрагмент разработан для поддержания желаемой химической реакции, например, одного или нескольких процессов в реакции синтеза нуклеиновой кислоты. Поверхностная энергия, или гидрофобность, поверхности является фактором, определяющим аффинность нуклеотида в отношении прикрепления к поверхности. В некоторых случаях способ функционализации устройства может включать: (a) создание устройства, характеризующегося поверхностью, которая содержит диоксид кремния; и (b) силанизирование поверхности с использованием подходящего

силанизирующего средства, описанного в настоящем документе или известного в настоящей области техники, например, органофункциональной молекулы алкоксисилана.

[00150] В некоторых случаях органофункциональная молекула алкоксисилана включает в себя диметилхлороктодецилсилан, метилдихлороктодецилсилан, трихлороктодецилсилан, триметилхлороктодецилсилан, триэтилохлороктодецилсилан или любую их комбинацию. В некоторых случаях поверхность устройства содержит поверхность, функционализированную с помощью следующего: полиэтилен/полипропилен (функционализированный гамма-облучением или окислением хромовой кислоты и восстановлением до гидроксильной поверхности), в высокой степени сшитый полистирол-дивинилбензол (дериватизированный хлорметилированием и аминированный до бензиламиновой функциональной поверхности), нейлон (концевые аминогексильные группы являются непосредственно реакционноспособными или протравлены восстановленным политетрафторэтиленом). Другие способы и функционализирующие средства описаны в патенте США № 5474796, который полностью включен в настоящее описание посредством ссылки.

[00151] В некоторых случаях поверхность устройства функционализируют путем контакта с дериватирующей композицией, которая содержит смесь силанов, в условиях реакции, эффективных для связывания силанов с поверхностью устройства, как правило, посредством реакционноспособных гидрофильных фрагментов, присутствующих на поверхности устройства. Силанизация, как правило, покрывает поверхность путем самосборки с органофункциональными молекулами алкоксисилана.

[00152] Кроме того, можно использовать различные силоксановые функционализирующие реагенты, известные в настоящее время в настоящей области техники, например, для снижения или увеличения поверхностной энергии. Органофункциональные алкоксисиланы можно классифицировать в соответствии с их органическими функциями.

[00153] В настоящем документе предусмотрены устройства, которые могут содержать компоновку средств, способных связываться с нуклеотидом. В некоторых случаях устройство может быть покрыто активным средством. В некоторых случаях устройство может быть покрыто пассивным средством. Иллюстративные активные средства для включения в материалы для покрытия, описанные в настоящем документе, включают в себя без ограничения N-(3-триэтоксипропил)-4-гидоксибутирамид (HAPS), 11-ацетоксиундецилтриэтоксисилан, н-децилтриэтоксисилан, (3-аминопропил)триметоксисилан, (3-аминопропил)триметоксисилан, (3-аминопропил)триэтоксисилан, 3-глицидоксипропилтриметоксисилан (GOPS), 3-йод-

пропилтриметоксисилан, бутилальдегидтриметоксисилан, вторичные димерные аминоалкилсилоксаны, (3-аминопропил)-диэтоксиметилсилан, (3-аминопропил)-диметилэтоксисилан и (3-аминопропил)-триметоксисилан, (3-глицидоксипропил)-диметилэтоксисилан, глицидокситриметоксисилан, (3-меркаптопропил)-триметоксисилан, 3-4-эпоксициклогексил-этилтриметоксисилан и (3-меркаптопропил)-метилдиметоксисилан, аллилтрихлорхлорсилан, 7-окт-1-енилтрихлорхлорсилан или бис(3-триметоксисилилпропил)амин.

[00154] Иллюстративные пассивные средства для включения в материал покрытия, описанный в настоящем документе, включают в себя без ограничения перфтороктилтрихлорсилан; тридекафтор-1,1,2,2-тетрагидрооктил)трихлорсилан; 1Н,1Н,2Н,2Н-фтороктилтриэтоксисилан (FOS); трихлор(1Н,1Н,2Н,2Н-перфтороктил)силан; трет-бутил-[5-фтор-4-(4,4,5,5-тетраметил-1,3,2-диоксаборолан-2-ил)-индол-1-ил]диметилсилан; CYTOP™; Fluorinert™; перфтороктилтрихлорсилан (PFOTCS); перфтороктилдиметилхлорсилан (PFODCS); перфтордецилтриэтоксисилан (PFDTES); пентафторфенилдиметилпропилхлорсилан (PFPTES); перфтороктилтриэтоксисилан; перфтороктилтриметоксисилан; октилхлорсилан; диметилхлороктадецилсилан; метилдихлороктадецилсилан; трихлороктадецилсилан; триметилоктадецилсилан; триэтилоктадецилсилан или октадецилтрихлорсилан.

[00155] В некоторых случаях функционализирующее средство включает в себя углеводородный силан, такой как октадецилтрихлорсилан. В некоторых случаях функционализирующее средство включает в себя 11-ацетоксиундецилтриэтоксисилан, н-децилтриэтоксисилан, (3-аминопропил)триметоксисилан, (3-аминопропил)триэтоксисилан, глицидилоксипропил/триметоксисилан и N-(3-триэтоксисилилпропил)-4-гидроксибутирамид.

[00156] Синтез полинуклеотидов

[00157] Способы настоящего раскрытия для синтеза нуклеиновых кислот могут включать процессы, включая в себя фосфорамидитную химию. В некоторых случаях синтез нуклеиновых кислот включает в себя сочетание основания с фосфорамидитом. Синтез нуклеиновых кислот может включать в себя сочетание основания за счет осаждения фосфорамидитом в условиях сочетания, при этом одно и то же основание необязательно осаждается с помощью фосфорамидита больше одного раза, т.е. происходит двойное сочетание. Синтез нуклеиновых кислот может включать в себя кэпирование непрореагировавших сайтов. В некоторых случаях кэпирование является необязательным. Синтез нуклеиновых кислот может также включать в себя окисления, или стадию окисления, или стадии окисления. Синтез нуклеиновых кислот может включать в себя

деблокирование, детритилирование и сульфурирование. В некоторых случаях синтез нуклеиновых кислот включает в себя окисление или сульфурирование. В некоторых случаях между одной или каждой стадией во время реакции синтеза нуклеиновых кислот устройство промывают, например, с использованием тетразола или ацетонитрила. Временные рамки для любой одной стадии в способе фосфорамидитного синтеза могут составлять меньше чем около 2 мин, 1 мин, 50 с, 40 с, 30 с, 20 с и 10 с.

[00158] Синтез нуклеиновых кислот с использованием фосфорамидитного способа может включать в себя последующее добавление строительного блока фосфорамидита (например, нуклеозидного фосфорамидита) к растущей цепи нуклеиновой кислоты для образования фосфиттриэфирной связи. Фосфорамидитный синтез нуклеиновых кислот происходит в направлении 3' - 5'. Фосфорамидитный синтез нуклеиновых кислот обеспечивает контролируемое добавление одного нуклеотида к растущей цепи нуклеиновой кислоты за один цикл синтеза. В некоторых случаях каждый цикл синтеза включает в себя стадию сочетания. Фосфорамидитное сочетание включает в себя образование фосфиттриэфирной связи между активированным нуклеозидным фосфорамидитом и нуклеозидом, связанным с субстратом, например, через линкер. В некоторых случаях нуклеозидный фосфорамидит подают в устройство в активированной форме. В некоторых случаях нуклеозидный фосфорамидит подают в устройство с активатором. В некоторых случаях нуклеозидные фосфорамидиты подают в устройство в 1,5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100-кратном избытке или большем избытке по сравнению со связанными с субстратом нуклеозидами. В некоторых случаях добавление нуклеозидного фосфорамидита проводят в безводной среде, например, в безводном ацетонитриле. После добавления нуклеозидного фосфорамидита устройство необязательно промывают. В некоторых случаях стадию сочетания повторяют один или несколько дополнительных раз, необязательно со стадией промывки между добавками нуклеозидного фосфорамидита к субстрату. В некоторых случаях способ синтеза нуклеиновых кислот, используемый в настоящем документе, включает 1, 2, 3 или больше последовательных стадий сочетания. До сочетания во многих случаях с нуклеозида, связанного с устройством, снимают защитную группу путем удаления защитной группы, прием защитная группа функционирует для предотвращения полимеризации. Как правило, защитной группой является 4,4'-диметокситритил (DMT).

[00159] После сочетания способы фосфорамидитного синтеза нуклеиновых кислот необязательно включают стадию кэпирования. На стадии кэпирования растущую нуклеиновую кислоту обрабатывают кэпирующим средством. Стадия кэпирования является применимой для блокирования непрореагировавших связанных с субстратом 5'-

ОН-групп после сочетания из дальнейшего удлинения цепи, предотвращая образование нуклеиновых кислот с внутренними делециями оснований. Кроме того, фосфорамидиты, активированные 1Н-тетразолом, могут в небольшой степени реагировать с положением Об гуанозина. Не будучи связанными какой-либо теорией, при окислении I₂/водой этот побочный продукт, возможно, через миграцию Об-N7, может подвергнуться депуринизации. Не содержащие пурин сайты могут в конечном итоге отщепляться в ходе окончательного снятия защитных групп с полинуклеотида, тем самым снижая выход продукта полной длины. Модификации Об можно удалить обработкой кэпирующим реагентом перед окислением I₂/водой. В некоторых случаях включение стадии кэпирования во время синтеза нуклеиновых кислот снижает частоту ошибок по сравнению с синтезом без кэпирования. В качестве примера стадия кэпирования включает обработку связанной с субстратом нуклеиновой кислоты смесью уксусного ангидрида и 1-метилимидазола. После стадии кэпирования устройство дополнительно промывают.

[00160] В некоторых случаях после добавления нуклеозидного фосфорамидита и необязательно после кэпирования и одной или нескольких стадий отмывки, связанную с устройством растущую нуклеиновую кислоту окисляют. Стадия окисления включает, что фосфиттриэфир окисляется до тетракоординированного фосфаттриэфира, защищенного предшественника встречающейся в природе фосфатдиэфирной межнуклеозидной связи. В некоторых случаях окисление растущей нуклеиновой кислоты достигается обработкой йодом и водой, необязательно в присутствии слабого основания (например, пиридина, лутидина, коллидина). Окисление можно провести в безводных условиях с использованием, например, трет-бутилгидропероксида или (1S)-(+)-(10-камфорсульфонил)оксазиридина (CSO). В некоторых способах стадию кэпирования выполняют после окисления. Вторая стадия кэпирования позволяет высушить устройство, поскольку остаточная вода от окисления, которая может сохраняться, может ингибировать последующее сочетание. После окисления устройство и растущую нуклеиновую кислоту необязательно промывают. В некоторых случаях стадию окисления заменяют стадией сульфирования для получения полинуклеотидных тиофосфатов, причем после сульфирования можно выполнять любые стадии кэпирования. Многие реагенты способны эффективно переносить серу, включая в себя без ограничения 3-(диметиламинометилен)амино-3Н-1,2,4-дитиазол-3-тион, DDTT, 3Н-1,2-бензодитиол-3-он-1,1-диоксид, также известный как реагент Бокажа, и N,N,N',N'-тетраэтилтиурамдисульфид (TETD).

[00161] Для того чтобы последующий цикл включения нуклеозидов происходил посредством сочетания, защищенный 5'-конец связанной с устройством растущей

нуклеиновой кислоты удаляют, так чтобы первичная гидроксильная группа вступала в реакцию со следующим нуклеозидным фосфорамидитом. В некоторых случаях защитной группой является DMT, и деблокирование происходит с помощью трихлоруксусной кислоты в дихлорметане. Проведение детритилирования в течение продолжительного времени или с более сильными, чем рекомендуемые, растворами кислот может привести к усилению депуринизации полинуклеотида, связанного с твердой подложкой, и, таким образом, снижает выход требуемого полноразмерного продукта. Способы и композиции согласно раскрытию, описанные в настоящем изобретении, обеспечивают условия контролируемого деблокирования, ограничивающие нежелательные реакции депуринизации. В некоторых случаях связанную с устройством нуклеиновую кислоту отмывают после деблокирования. В некоторых случаях эффективная промывка после деблокирования способствует получению синтезированных нуклеиновых кислот, характеризующихся низкой частотой ошибок.

[00162] Способы синтеза нуклеиновых кислот, как правило, включают повторяющуюся последовательность следующих стадий: нанесение защищенного мономера на активно функционализированную поверхность (например, локус) для связывания либо с активированной поверхностью, линкером, либо с мономером с ранее снятой защитной группой; снятие защитной группы с нанесенного мономера так, чтобы он вступал в реакцию с последующим нанесенным защищенным мономером; и нанесение другого защищенного мономера для связывания. Одна или несколько промежуточных стадий включают в себя окисление или сульфурирование. В некоторых случаях одна или несколько стадий промывки предшествуют или следуют после одной или всех стадий.

[00163] Способы синтеза нуклеиновых кислот на основе фосфорамидита включают серию химических стадий. В некоторых случаях одна или несколько стадий способа синтеза включают циклическое использование реагентов, где одна или несколько стадий способа предусматривают нанесение на устройство реагента, применимого на стадии. Например, реагенты циклически проходят серию стадий осаждения жидкости и вакуумной сушки. Для субстратов, содержащих трехмерные элементы, такие как лунки, микролунки, каналы и тому подобное, реагенты необязательно пропускают через одну или несколько участков устройства через лунки и/или каналы.

[00164] Описанные в настоящем документе способы и системы относятся к синтезирующим полинуклеотиды устройствам для синтеза полинуклеотидов. Синтез может происходить в параллельном режиме. Например, по меньшей мере или около по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850,

900, 1000, 10000, 50000, 75000, 100000 или больше полинуклеотидов можно синтезировать в параллельном режиме. Общее число нуклеиновых кислот, которые можно синтезировать в параллельном режиме может составлять 2-100000, 3-50000, 4-10000, 5-1000, 6-900, 7-850, 8-800, 9-750, 10-700, 11-650, 12-600, 13-550, 14-500, 15-450, 16-400, 17-350, 18-300, 19-250, 20-200, 21-150, 22-100, 23-50, 24-45, 25-40, 30-35. Специалистам в настоящей области техники понятно, что общее число полинуклеотидов, синтезированных в параллельном режиме, может попадать в любой диапазон, ограниченный любым из этих значений, например 25-100. Общее число полинуклеотидов, синтезированных в параллельном режиме, может попадать в любой диапазон, определяемый любым из значений, служащих конечными точками диапазона. Общая молярная масса полинуклеотидов, синтезированных в пределах устройства, или молярная масса каждого из полинуклеотидов может составлять по меньшей мере или по меньшей мере около 10, 20, 30, 40, 50, 100, 250, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 25000, 50000, 75000, 100000 пикомоль или больше. Длина каждого из полинуклеотидов или средняя длина полинуклеотидов в пределах устройства может составлять по меньшей мере или около по меньшей мере 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 300, 400, 500 нуклеотидов или больше. Длина каждого из полинуклеотидов или средняя длина полинуклеотидов в пределах устройства может составлять не более чем или около не более чем 500, 400, 300, 200, 150, 100, 50, 45, 35, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10 нуклеотидов или меньше. Длина каждого из полинуклеотидов или средняя длина полинуклеотидов в пределах устройства может попадать в диапазон, составляющий 10-500, 9-400, 11-300, 12-200, 13-150, 14-100, 15-50, 16-45, 17-40, 18-35, 19-25. Специалистам в настоящей области техники понятно, что длина каждого из полинуклеотидов или средняя длина полинуклеотидов в пределах устройства может попадать в любой диапазон, ограниченный любым из указанных значений, например, 100-300. Длина каждого из полинуклеотидов или средняя длина полинуклеотидов в пределах устройства может попадать в любой диапазон, определяемый любым из значений, служащих конечными точками диапазона.

[00165] Предусмотренные в настоящем документе способы синтеза нуклеиновых кислот на поверхности позволяют проводить синтез с высокой скоростью. В качестве примера синтезируют по меньшей мере 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 125, 150, 175, 200 нуклеотидов в час или больше. Нуклеотиды включают в себя адениновые, гуаниновые, тиминные, цитозинные, уридинные строительные блоки или их аналоги/модифицированные версии. В некоторых случаях библиотеки нуклеиновых кислот синтезируют в параллельном режиме на субстрате. Например, устройство, содержащее

около или по меньшей мере около 100; 1000; 10000; 30000; 75000; 100000; 1000000; 2000000; 3000000; 4000000 или 5000000 выделенных локусов, способно поддержать синтез по меньшей мере такого же числа отдельных нуклеиновых кислот, причем нуклеиновую кислоту, кодирующую отдельную последовательность, синтезируют на выделенном локусе. В некоторых случаях библиотеку нуклеиновых кислот синтезируют на устройстве с низкими частотами ошибок, описанном в настоящем документе, меньше чем около за три месяца, две месяца, один месяц, три недели, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 дня, 24 часа или меньше. В некоторых случаях нуклеиновые кислоты большего размера, собранные из библиотеки нуклеиновых кислот, синтезированной с низкой частотой ошибок с использованием субстратов и способов, описанных в настоящем документе, получают меньше чем около за три месяца, две месяца, один месяц, три недели, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 дня, 24 часа или меньше.

[00166] В некоторых случаях описанные в настоящем документе способы обеспечивают создание библиотеки нуклеиновых кислот, содержащей варианты нуклеиновые кислоты, различающиеся по множеству сайтов кодонов. В некоторых случаях нуклеиновая кислота может содержать 1 сайт, 2 сайта, 3 сайта, 4 сайта, 5 сайтов, 6 сайтов, 7 сайтов, 8 сайтов, 9 сайтов, 10 сайтов, 11 сайтов, 12 сайтов, 13 сайтов, 14 сайтов, 15 сайтов, 16 сайтов, 17 сайтов, 18 сайтов, 19 сайтов, 20 сайтов, 30 сайтов, 40 сайтов, 50 сайтов или больше вариантных сайтов кодонов.

[00167] В некоторых случаях один или несколько сайтов вариантных сайтов кодонов могут являться смежными. В некоторых случаях один или несколько сайтов вариантных сайтов кодонов могут не являться смежными и разделены 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 или больше кодонами.

[00168] В некоторых случаях нуклеиновая кислота может содержать множественные сайты вариантных сайтов кодонов, причем все варианты сайты кодонов являются смежными друг с другом, образуя участок вариантных сайтов кодонов. В некоторых случаях нуклеиновая кислота может содержать множественные сайты вариантных сайтов кодонов, причем ни один из вариантных сайтов кодонов не является смежным друг с другом. В некоторых случаях нуклеиновая кислота может содержать множественные сайты вариантных сайтов кодонов, причем некоторые из вариантных сайтов кодонов являются смежными друг с другом, образуя участок вариантных сайтов кодонов, а некоторые из вариантных сайтов кодонов не являются смежными друг с другом.

[00169] Ссылаясь на фигуры, на фиг. 12 проиллюстрирован пример рабочего процесса синтеза нуклеиновых кислот (например, генов) из более коротких нуклеиновых кислот. Рабочий процесс, как правило, делят на следующие стадии: (1) синтез de novo

библиотеки одноцепочечных нуклеиновых кислот, (2) объединение нуклеиновых кислот с образованием более крупных фрагментов, (3) исправление ошибок, (4) контроль качества и (5) отправка. Перед синтезом *de novo* предварительно выбирают предполагаемую последовательность нуклеиновой кислоты или группу последовательностей нуклеиновой кислоты. Например, для создания предварительно выбирают группу генов.

[00170] После того, как выбраны большие нуклеиновые кислоты для создания, предварительно определенную библиотеку нуклеиновых кислот конструируют для синтеза *de novo*. Известны различные подходящие способы получения нуклеиновокислотных матриц высокой плотности. В примере рабочего процесса предусмотрен поверхностный слой **1201** устройства. В этом примере химический состав поверхности изменяют с целью улучшения процесса синтеза нуклеиновых кислот. Области с низкой поверхностной энергией создают для отталкивания жидкости, в то время как области с высокой поверхностной энергией создают для притяжения жидкостей. Сама поверхность может быть в форме плоской поверхности или содержать изменения в форме, такие как выступы или микролунки, которые увеличивают площадь поверхности. В примере рабочего процесса выбранные молекулы с высокой поверхностной энергией выполняют двойную функцию поддержания химии ДНК, как раскрыто в публикации международной патентной заявки WO/2015/021080, которая полностью включена в настоящее описание посредством ссылки.

[00171] Получение матриц нуклеиновых кислот *in situ* проводят на твердой подложке и при этом используют способ удлинения на один нуклеотид для удлинения нескольких олигомеров в параллельном режиме. Устройство для осаждения материала, такое как синтезатор нуклеиновых кислот, предназначено для поэтапного высвобождения реагентов таким образом, что множественные нуклеиновые кислоты удлиняются в параллельном режиме по одному остатку за раз, образуя олигомеры с предварительно определенной последовательностью нуклеиновой кислоты **1202**. В некоторых случаях нуклеиновые кислоты отщепляют от поверхности на этой стадии. Расщепление включает в себя расщепление газом, например, аммиаком или метиламином.

[00172] Созданные библиотеки нуклеиновых кислот помещают в реакционную камеру. В этом иллюстративном рабочем процессе реакционная камера (также называемая “нанореактором”) представляет собой лунку с кремниевым покрытием, содержащую реагенты для ПЦР и опущенную на библиотеку нуклеиновых кислот **1203**. До или после герметизации **1204** нуклеиновых кислот в реагент добавляют для высвобождения нуклеиновых кислот из субстрата. В иллюстративном рабочем процессе нуклеиновые кислоты высвобождаются после герметизации нанореактора **1205**. После высвобождения

фрагменты одноцепочечных нуклеиновых кислот гибридизуются, чтобы охватить всю длинную последовательность ДНК. Частичная гибридизация **1205** возможна, потому что каждая синтезированная нуклеиновая кислота сконструирована так, что содержит небольшую часть, перекрывающуюся по меньшей мере с одной другой нуклеиновой кислотой в пуле.

[00173] После гибридизации начинают реакцию РСА. Во время полимеразных циклов нуклеиновые кислоты отжигаются с комплементарными фрагментами и пропуски заполняются полимеразой. Каждый цикл увеличивает длину различных фрагментов случайным образом в зависимости от того, какие нуклеиновые кислоты находят друг друга. Комплементарность между фрагментами позволяет сформировать полный большой диапазон двухцепочечной ДНК **1206**.

[00174] После завершения РСА нанореактор отделяют от устройства **1207** и располагают для взаимодействия с устройством, содержащим праймеры для ПЦР **1208**. После герметизации нанореактор подвергают ПЦР **1209** и амплифицируют более крупные нуклеиновые кислоты. После ПЦР **1210** наноканалу открывают **1211**, добавляют реагенты для исправления ошибок **1212**, канал герметизируют **1213**, и происходит реакция исправления ошибок для удаления несовпадающих пар оснований и/или цепей с плохой комплементарностью из двухцепочечной продуктов ПЦР-амплификации **1214**. Нанореактор открывают и отделяют **1215**. Продукт с исправленными ошибками затем подвергают дополнительным стадиям обработки, таким как ПЦР и молекулярное штрих-кодирование, а затем упаковывают **1222** для отправки **1223**.

[00175] В некоторых случаях принимают меры по контролю качества. После исправления ошибок стадии контроля качества включают в себя, например, взаимодействие с пластиной, содержащей праймеры секвенирования, для амплификации продукта с исправленными ошибками **1216**, герметизацию пластины с камерой, содержащей продукт амплификации с исправленными ошибками **1217**, и выполнение дополнительного цикла амплификации **1218**. Нанореактор открывают **1219** и продукты объединяют **1220** и секвенируют **1221**. После того как приемлемое определение контроля качества выполнено, упакованный продукт **1222** получает одобрение для отправки **1223**.

[00176] В некоторых случаях нуклеиновую кислоту, созданную с помощью такого рабочего процесса, как представлен на фиг. **12**, подвергают мутагенезу с использованием перекрывающихся праймеров, раскрытых в настоящем документе. В некоторых случаях библиотеку праймеров создают путем получения *in situ* на твердой подложке и используют процесс удлинения на один нуклеотид для удлинения нескольких олигомеров в параллельном режиме. Устройство для осаждения, такое как синтезатор

нуклеиновых кислот, предназначено для поэтапного высвобождения реагентов таким образом, чтобы множественные нуклеиновые кислоты удлинялись в параллельном режиме по одному остатку за раз, образуя олигомеры с предварительно определенной последовательностью нуклеиновой кислоты **1202**.

[00177] Вычислительные системы

[00178] Любая из систем, описанных в данном документе, может быть функционально связана с компьютером и может быть автоматизирована через компьютер локально или удаленно. В различных случаях способы и системы согласно настоящему раскрытию могут дополнительно содержать программы из системы программного обеспечения для вычислительных систем и их использование. Соответственно, компьютеризированное управление для синхронизации функций дозирования/вакуума/пополнения, таких как управление и синхронизация движения устройства осаждения материала, действия дозирования и приведение в действие вакуума, находятся в пределах объема настоящего раскрытия. Вычислительные системы могут быть запрограммированы на взаимодействие между заданной пользователем последовательностью оснований и положением устройства осаждения материала, чтобы доставлять правильные реагенты в указанные области субстрата.

[00179] Вычислительную систему **1300**, показанную на фиг. **13** можно понимать как логическое устройство, которое может считывать инструкции с носителя **1311** и/или сетевого порта **1305**, который необязательно может быть подключен к серверу **1309** с фиксированным носителем **1312**. Такая система, как показана на фиг. **13**, может включать в себя ЦП (центральный процессор) **1301**, дисководы **1303**, дополнительные устройства ввода, такие как клавиатура **1315** и/или мышь **1316** и дополнительный монитор **1307**. Передача данных может быть достигнута через указанную среду передачи данных на сервер в локальном или удаленном местоположении. Среда передачи данных может включать в себя любые средства передачи и/или приема данных. Например, среда передачи данных может являться сетевым соединением, беспроводным соединением или Интернет-соединением. Такое соединение может обеспечить связь через Интернет. Предполагается, что данные, относящиеся к настоящему раскрытию, могут передаваться по таким сетям или соединениям для приема и/или просмотра стороной **1322**, как показано на фиг. **13**.

[00180] На фиг. **14** представлена блок-схема, иллюстрирующая первый пример архитектуры вычислительной системы **1400**, которую можно использовать в связи с иллюстративными примерами согласно настоящему изобретению. Как показано на фиг. **14**, иллюстративная вычислительная система может включать в себя процессор **1402** для обработки команд. Неограничивающие примеры процессоров включают в себя: процессор

Intel Xeon™, процессор AMD Opteron™, 32-разрядный процессор RISC ARM 1176JZ(F)-S v1.0™ Samsung, процессор ARM Cortex-A8 Samsung S5PC100™, процессор ARM Cortex-A8 Apple A4™, процессор Marvell PXA 930™ или функционально эквивалентный процессор. Несколько потоков выполнения можно использовать для параллельной обработки. В некоторых случаях можно также использовать несколько процессоров или процессоры с несколькими ядрами, будь то в одной вычислительной системе, в кластере или распределенных по системам по сети, состоящей из множества устройств - компьютеров, мобильных телефонов и/или карманных персональных компьютеров.

[00181] Как показано на фиг. 14, высокоскоростной кэш 1404 может быть подключен или встроен в процессор 1402 для обеспечения высокоскоростной памяти для команд или данных, которые недавно или часто используются процессором 1402. Процессор 1402 подключен к северному мосту 1406 процессорной шиной 1408. Северный мост 1406 соединен с оперативной памятью (ОЗУ) 1410 шиной памяти 1412 и управляет доступом к ОЗУ 1410 процессором 1402. Северный мост 1406 также соединен с южным мостом 1414 с помощью шины чипсета 1416. Южный мост 1414, в свою очередь, соединен с периферийной шиной 1418. Периферийная шина может представлять собой, например, PCI, PCI-X, PCI Express или другую периферийную шину. Северный мост и южный мост часто называются чипсетом процессора, и они управляют передачей данных между процессором, ОЗУ и периферийными компонентами на периферийной шине 1418. В некоторых альтернативных архитектурах функциональность северного моста может быть встроена в процессор вместо использования отдельного чипа северного моста. В некоторых случаях система 1400 может включать в себя карту акселератора 1422, присоединенную к периферийной шине 1418. Акселератор может включать в себя программируемые пользователем вентильные матрицы (FPGA) или другое аппаратное обеспечение для ускорения определенной обработки. Например, акселератор можно использовать для адаптивной реструктуризации данных или для оценки алгебраических выражений, используемых при обработке расширенного набора.

[00182] Программное обеспечение и данные хранятся во внешнем хранилище 1424 и могут быть загружены в ОЗУ 1410 и/или в кэш 1404 для использования процессором. Система 1400 включает в себя операционную систему для управления системными ресурсами; неограничивающие примеры операционных систем включают в себя: Linux, Windows™, MACOS™, BlackBerry OS™, iOS™ и другие функционально эквивалентные операционные системы, а также прикладное программное обеспечение, работающее поверх операционной системы, для управления хранением и оптимизацией данных в соответствии с иллюстративными случаями согласно настоящему раскрытию. В этом примере система

1400 также включает в себя сетевые интерфейсные карты (NIC) **1420** и **1421**, подключенные к периферийной шине для обеспечения сетевых интерфейсов для внешнего хранилища, такого как сетевое хранилище (NAS) и другие вычислительные системы, которые можно использовать для распределенной параллельной обработки.

[00183] Фиг. **15** представляет собой диаграмму, на которой показана сеть **1500** с множеством вычислительных систем **1502a** и **1502b**, множеством сотовых телефонов и карманных персональных компьютеров **1502c** и сетевым хранилищем (NAS) **1504a** и **1504b**. В иллюстративных примерах системы **1502a**, **1502b** и **1502c** могут управлять хранением данных и оптимизировать доступ к данным для данных, хранящихся в сетевом хранилище (NAS) **1504a** и **1504b**. Математическую модель можно использовать для данных и оценивать с использованием распределенной параллельной обработки по вычислительным системам **1502a** и **1502b**, а также по системам сотовых телефонов и карманных персональных компьютеров **1502c**. Вычислительные системы **1502a** и **1502b** и системы сотовых телефонов и карманных персональных компьютеров **1502c** также могут обеспечивать параллельную обработку для адаптивной реструктуризации данных, хранящихся в сетевом хранилище (NAS) **1504a** и **1504b**. На фиг. **15** проиллюстрирован только пример, и большое разнообразие других компьютерных архитектур и систем можно использовать в сочетании с различными примерами согласно настоящему раскрытию. Например, сверхкомпактный сервер (blade-сервер) можно использовать для обеспечения параллельной обработки. Blade-секции процессора могут быть соединены через объединительную плату для обеспечения параллельной обработки. Хранилище также может быть подключено к объединительной плате или как сетевое хранилище (NAS) через отдельный сетевой интерфейс. В некоторых иллюстративных случаях процессоры могут поддерживать отдельные области памяти и передавать данные через сетевые интерфейсы, объединительную плату или другие соединители для параллельной обработки другими процессорами. В других случаях некоторые или все процессоры могут использовать совместное пространство памяти виртуальных адресов.

[00184] На фиг. **16** представлена блок-схема многопроцессорной вычислительной системы **1600**, использующей совместное пространство памяти виртуальных адресов в соответствии с иллюстративным примером. Система включает в себя множество процессоров **1602a-f**, которые могут обращаться к подсистеме совместной памяти **1604**. Система включает в себя множество программируемых аппаратных процессоров алгоритмов памяти **1606a-f** в подсистеме памяти **1604**. Каждый MAP **1606a-f** может содержать память **1608a-f** и один или несколько программируемых пользователем вентильных матриц (FPGA) **1610a-f**. MAP образует собой конфигурируемый

функциональный блок, и конкретные алгоритмы или части алгоритмов могут быть предоставлены FPGA **1610a-f** для обработки в тесной координации с соответствующим процессором. Например, MAP можно использовать для оценки алгебраических выражений, касающихся модели данных, и для выполнения адаптивной реструктуризации данных в иллюстративных примерах. В этом примере каждый MAP доступен всем процессорам для этих целей. В одной конфигурации каждый MAP может использовать прямой доступ к памяти (DMA) для доступа к ассоциированной памяти **1608a-f**, что позволяет ей выполнять задачи независимо и асинхронно от соответствующего микропроцессора **1602a-f**. В этой конфигурации MAP может передавать результаты непосредственно в другой MAP для конвейерной обработки и параллельного выполнения алгоритмов.

[00185] Вышеупомянутые компьютерные архитектуры и системы являются только примерами, и в связи с иллюстративными примерами можно использовать большое разнообразие других компьютерных архитектур и систем компьютеров, сотовых телефонов и карманных персональных компьютеров, включая в себя системы с использованием любой комбинации общих процессоров, копроцессоров, FPGA и других программируемых логических устройств, системы-на-кристалле (SOC), специализированных интегральных схем (ASIC) и других элементов обработки и логики. В некоторых случаях вся или часть вычислительной системы может быть реализована программно или аппаратно. Любое разнообразие носителей данных можно использовать в связи с иллюстративными примерами, включая в себя оперативное запоминающее устройство, жесткие диски, флэш-память, ленточные накопители, дисковые массивы, сетевое хранилище данных (NAS) и другие локальные или распределенные устройства и системы хранения данных.

[00186] В иллюстративных примерах вычислительная система может быть реализована с использованием программных модулей, выполняющихся на любой из вышеперечисленных или других компьютерных архитектур и систем. В других случаях функции системы могут быть реализованы частично или полностью в программно-аппаратных средствах, программируемых логических устройствах, таких как программируемые пользователем вентильные матрицы (FPGA), как показано на фиг. **13**, система-на-кристалле (SOC), специализированные интегральные схемы (ASIC) или другие элементы обработки и логики. Например, процессор задания и оптимизатор могут быть реализованы с аппаратным ускорением посредством использования карты аппаратного акселератора, такой как карта акселератора **1322**, показанная на фиг. **13**.

[00187] Приведенные ниже примеры представлены для более ясной иллюстрации принципа и практической реализации вариантов осуществления, раскрытых в настоящем документе, для специалистов в настоящей области техники и не должны

рассматриваться как ограничивающие объем любых заявленных вариантов осуществления. Если не указано иное, все части и проценты приведены в расчете на массу.

ПРИМЕРЫ

[00188] Следующие примеры приведены с целью иллюстрации различных вариантов осуществления настоящего раскрытия и не предназначены для ограничения настоящего раскрытия каким-либо образом. Настоящие примеры вместе со способами, описанными в настоящем документе, в настоящее время представляют предпочтительные варианты осуществления, являются иллюстративными и не предназначены для ограничения объема настоящего раскрытия. Изменения в настоящем документе и другие применения, которые охватываются сущностью настоящего раскрытия, как определено объемом формулы изобретения, будут понятны специалистам в настоящей области техники.

[00189] Пример 1: Функционализация поверхности устройства

[00190] Устройство функционализировали для поддержания прикрепления и синтеза библиотеки нуклеиновых кислот. Поверхность устройства сначала подвергали влажной очистке с использованием раствора «пираньи», содержащего 90% H_2SO_4 и 10% H_2O_2 , в течение 20 минут. Устройство промывали в нескольких стаканах деионизированной водой, выдерживали в течение 5 минут под водопроводным краном с деионизированной водой и высушивали N_2 . Затем субстрат замачивали в NH_4OH (1:100; 3 мл: 300 мл) в течение 5 минут, промывали деионизированной водой с использованием ручного дозатора-пистолета, пропитывали в трех последовательных стаканах деионизированной водой в течение 1 мин каждый и затем снова промывали деионизированной водой с использованием ручного дозатора-пистолета. Затем устройство очищали плазмой, подвергая поверхность устройства воздействию O_2 . Прибор SAMCO PC-300 использовали для плазмохимического травления O_2 при 250 Вт в течение 1 мин в нисходящем режиме.

[00191] Очищенную поверхность устройства активно функционализировали раствором, содержащим N-(3-триэтоксисилилпропил)-4-гидроксибутирамид, с использованием системы печи для осаждения из паровой фазы YES-1224P со следующими параметрами: от 0,5 до 1 торр, 60 мин, 70°C, испаритель 135°C. Поверхность устройства покрывали резистом с использованием прибора для нанесения растворов центрифугированием Brewer Science 200X. Фоторезист SPR™ 3612 наносили методом центрифугирования на устройство при 2500 об/мин в течение 40 с. Устройство предварительно прокаливали в течение 30 мин при 90°C на горячей плите Брюера. Устройство подвергали фотолитографии с использованием установки для совмещения и

экспонирования Karl Suss MA6. Устройство экспонировали в течение 2,2 с и проявляли в течение 1 мин в MSF 26A. Оставшийся проявитель промывали с использованием ручного дозатора-пистолета и субстрат замачивали в воде в течение 5 мин. Устройство прокаливали в течение 30 мин при 100°C в печи с последующей визуальной проверкой дефектов литографии с использованием Nikon L200. Процесс удаления накипи использовали для удаления остаточного резиста с использованием прибора SAMCO PC-300 для плазмохимического травления O₂ при 250 Вт в течение 1 мин.

[00192] Поверхность устройства пассивно функционализировали с помощью 100 мкл раствора перфтороктилтрихлорсилана, смешанного с 10 мкл легкого минерального масла. Устройство помещали в камеру, накачивали насосом в течение 10 мин, а затем клапан закрывали для насоса и оставляли стоять в течение 10 мин. Камера вентилировалась. Устройство очищали от резиста, выполняя два погружения в течение 5 минут в 500 мл NMP при 70°C с обработкой ультразвуком при максимальной мощности (9 на системе Crest). Затем устройство замачивали в течение 5 минут в 500 мл изопропанола при комнатной температуре с обработкой ультразвуком при максимальной мощности. Устройство погружали в 300 мл этанола с крепостью 200 и продували N₂. Функционализированную поверхность активировали, чтобы она служила в качестве подложки для синтеза нуклеиновых кислот.

[00193] Пример 2: Синтез 50-мерной последовательности на устройстве для синтеза полинуклеотидов

[00194] Двухмерное устройство для синтеза полинуклеотидов собирали в проточную кювету, которую соединяли с проточной кюветой (Applied Biosystems, синтезатор ДНК ABI394). Устройство для синтеза полинуклеотидов, однородно функционализированное с помощью N-(3-триэтоксисилилпропил)-4-гидроксибутирамида (Gelest), использовали для синтеза иллюстративного полинуклеотида из 50 п.н. ("50-мерного полинуклеотида") с использованием описанных в настоящем документе способов синтеза полинуклеотидов.

[00195] Последовательность, составляющую 50-мер, являлась такой, как описано в

SEQ	ID	NO.:	20.
5'AGACAATCAACCATTTGGGGTGGACAGCCTTGACCTCTAGACTTCGGCAT##TTTTT			
TTTTT3' (SEQ ID NO.: 20), где # обозначает тимидин-сукцинилгексамид-CED-			
фосфорамидит (CLP-2244 от ChemGenes), который представляет собой расщепляемый			
линкер, обеспечивающий высвобождение олигонуклеотидов с поверхности во время снятия			
защитных групп.			

[00196] Синтез выполняли с использованием стандартной химии синтеза ДНК (сочетание, кэпирование, окисление и деблокирование) согласно протоколу в таблице 6 и синтезатора ABI.

Таблица 6. Протокол синтеза

Таблица 6		
Название процесса общего синтеза ДНК	Стадия способа	Время (с)
Промывка (промывка потоком ацетонитрила)	Промывка системы ацетонитрилом	4
	Ацетонитрил в проточную кювету	23
	Промывка системы N ₂	4
	Промывка системы ацетонитрилом	4
Добавление оснований ДНК (поток фосфорамидита + активатора)	Промывка распределительного коллектора активатором	2
	Активатор в проточную кювету	6
	Активатор + фосфорамидит в проточную кювету	6
	Активатор в проточную кювету	0,5
	Активатор + фосфорамидит в проточную кювету	5
	Активатор в проточную кювету	0,5
	Активатор + фосфорамидит в проточную кювету	5
	Активатор в проточную кювету	0,5
	Активатор + фосфорамидит в проточную кювету	5
	Инкубировать в течение 25 с	25
Промывка (промывка потоком ацетонитрила)	Промывка системы ацетонитрилом	4
	Ацетонитрил в проточную кювету	15
	Промывка системы N ₂	4

Таблица 6		
Название процесса общего синтеза ДНК	Стадия способа	Время (с)
	Промывка системы ацетонитрилом	4
Добавление оснований ДНК (поток фосфорамидита + активатора)	Промывка распределительного коллектора активатором	2
	Активатор в проточную кювету	5
	Активатор + фосфорамидит в проточную кювету	18
	Инкубировать в течение 25 с	25
Промывка (промывка потокм ацетонитрила)	Промывка системы ацетонитрилом	4
	Ацетонитрил в проточную кювету	15
	Промывка системы N ₂	4
	Промывка системы ацетонитрилом	4
Кэпирование (кэп А+В, 1:1, поток)	Кэп А+В в проточную кювету	15
Промывка (промывка потокм ацетонитрила)	Промывка системы ацетонитрилом	4
	Ацетонитрил в проточную кювету	15
	Промывка системы ацетонитрилом	4
Окисление (поток окислителя)	Окислитель в проточную кювету	18
Промывка (промывка потокм ацетонитрила)	Промывка системы ацетонитрилом	4
	Промывка системы N ₂	4
	Промывка системы ацетонитрилом	4
	Ацетонитрил в проточную кювету	15
	Промывка системы ацетонитрилом	4
	Ацетонитрил в проточную кювету	15

Таблица 6		
Название процесса общего синтеза ДНК	Стадия способа	Время (с)
	Промывка системы N ₂	4
	Промывка системы ацетонитрилом	4
	Ацетонитрил в проточную кювету	23
	Промывка системы N ₂	4
	Промывка системы ацетонитрилом	4
Деблокирование (поток деблокирующего средства)	Деблокирующее средство в проточную кювету	36
Промывка (промывка потоком ацетонитрила)	Промывка системы ацетонитрилом	4
	Промывка системы N ₂	4
	Промывка системы ацетонитрилом	4
	Ацетонитрил в проточную кювету	18
	Промывка системы N ₂	4,13
	Промывка системы ацетонитрилом	4,13
	Ацетонитрил в проточную кювету	15

[00197] Комбинацию фосфорамидита/активатора доставляли аналогично доставке объемных реагентов через проточную кювету. Стадии сушки не проводили, так как среда все время оставалась "влажной" за счет реагента.

[00198] Ограничитель потока удаляли из синтезатора ABI 394 для обеспечения более быстрого потока. Без ограничителя потока скорости потока для амидитов (0,1 М в ACN), активатора (0,25 М бензоилтиотетразола ("БТТ"; 30-3070-xx от GlenResearch) в ACN) и окислителя (Ох) (0,02 М I₂ в 20% пиридина, 10 % воды и 70% THF) составляла примерно 100 мкл/с, для ацетонитрила ("ACN") и реагентов для кэпирования (смесь кэпа А и кэпа В 1: 1, где кэп А представляет собой уксусный ангидрид в смеси THF/пиридина, а кэп В составляет 16% 1-метилимидизола в THF) составляла примерно ~ 200 мкл/с, а для деблокирующего средства (3% дихлоруксусная кислота в толуоле) примерно ~ 300 мкл/с (по сравнению с ~ 50 мкл/с для всех реагентов с ограничителем потока). Наблюдали время для полного выталкивания окислителя, синхронизацию по времени для химических потоков скорректировали соответствующим образом и вводили дополнительную промывку

с помощью ACN между различными химическими веществами. После синтеза полинуклеотидов с чипа снимали защитные группы в газообразном аммиаке в течение ночи при 75 фунт/кв.дюйм. Пять капель воды наносили на поверхность для восстановления нуклеиновых кислот. Извлеченные нуклеиновые кислоты затем анализировали на небольшом РНК-чипе BioAnalyzer (данные не показаны).

[00199] Пример 3: Синтез 100-мерной последовательности на устройстве для синтеза полинуклеотидов

[00200] Такой же процесс, как описано в примере 2, для синтеза 50-мерной последовательности, использовали для синтеза 100-мерного полинуклеотида ("100-мер полинуклеотид"; 5'

CGGGATCCTTATCGTCATCGTCGTACAGATCCCGACCCATTTGCTGTCCACCAGTCA TGCTAGCCATACCATGATGATGATGATGATGAGAACCCCGCAT##TTTTTTTTTTT3', где # обозначает тимидин-сукцинилгексамид-CED-фосфорамидит (CLP-2244 от ChemGenes); SEQ ID NO.: 21) на двух разных кремниевых чипах, первый из которых равномерно функционализирован с помощью N-(3-триэтоксисилилпропил)-4-гидроксибутирамид, а второй функционализирован с помощью смеси 5/95 11-ацетоксиундецилтриэтоксисилана и н-децилтриэтоксисилана и нуклеиновые кислоты, экстрагированные с поверхности, анализировали на приборе BioAnalyzer (данные не показаны).

[00201] Все десять образцов из двух чипов дополнительно амплифицировали с помощью ПЦР с использованием прямого (5'ATGCGGGGTTCTCATCATC3'; SEQ ID NO.: 22) и обратного (5'CGGGATCCTTATCGTCATCG3'; SEQ ID NO.: 23) праймера в 50 мкл смеси для ПЦР (25 мкл мастер-микса NEB Q5, 2,5 мкл 10 мкМ прямого праймера, 2,5 мкл 10 мкМ обратного праймера, 1 мкл нуклеиновой кислоты, извлеченной с поверхности, и воды до 50 мкл) с использованием следующей программы термоциклирования:

98°C, 30 с

98°C, 10 с; 63°C, 10 с; 72°C, 10 с; повторить 12 циклов

72°C, 2 мин

[00202] Продукты ПЦР также анализировали на BioAnalyzer (данные не показаны), демонстрируя резкие пики в положении 100-мер. Затем амплифицированные с помощью ПЦР образцы клонировали и секвенировали. В таблице 7 обобщенно представлены результаты секвенирования Сэнгера для образцов, взятых из пятен 1-5 из чипа 1, и для образцов, взятых из пятен 6-10 из чипа 2.

Таблица 7. Результаты секвенирования

[00204] Наконец, в таблице 8 обобщенно представлены характеристики ошибок для последовательностей, полученных из образцов полинуклеотидов из пятен 1-10.

[illegible]

оснований ROI										
Небольшая вставка ROI	1	0	0	0	0	0	0	0	0	0
Делеция одного основания ROI	0	0	0	0	0	0	0	0	0	0
Количество больших делеций	0	0	1	0	0	1	1	0	0	0
Мутация: G>A	2	2	1	2	1	0	2	1	2	1
Мутация: T>C	0	0	0	1	0	0	0	0	0	0
Количество ошибок ROI	3	2	2	3	1	1	3	1	2	1
Частота ошибок ROI	Ош.: ~1 на 834	Ош.: ~1 на 1350	Ош.: ~1 на 1282	Ош.: ~1 на 708	Ош.: ~1 на 2500	Ош.: ~1 на 2667	Ош.: ~1 на 876	Ош.: ~1 на 2900	Ош.: ~1 на 1400	Ош.: ~1 на 2349
Частота ошибок ROI без праймера (MP)	Ош. MP: ~1 на 763	Ош. MP: ~1 на 824	Ош. MP: ~1 на 780	Ош. MP: ~1 на 429	Ош. MP: ~1 на 1525	Ош. MP: ~1 на 1615	Ош. MP: ~1 на 531	Ош. MP: ~1 на 1769	Ош. MP: ~1 на 854	Ош. MP: ~1 на 1451

[00205] Пример 4: Создание библиотеки нуклеиновых кислот с помощью мутагенеза в одном положении на одном сайте

[00206] Праймеры нуклеиновой кислоты *de novo* синтезировали для применения в серии реакций ПЦР для создания библиотеки вариантов нуклеиновых кислот матричной нуклеиновой кислоты, см. фигуры **2A-2D**. Четыре типа праймеров создавали на фиг. **2A**: внешний 5' праймер **215**, внешний 3' праймер **230**, внутренний 5' праймер **225** и внутренний 3' праймер **220**. Внутренний 5' праймер / первую нуклеиновую кислоту **220** и внутренний 3' праймер / вторую нуклеиновую кислоту **225** создавали с использованием способа синтеза нуклеиновых кислот, как в общем изложено в таблице **6**. Внутренний 5' праймер / первая нуклеиновая кислота **220** представляет набор из вплоть до 19 праймеров предварительно определенной последовательности, где каждый праймер в наборе отличается от другого на один кодон, на одном сайте последовательности.

[00207] Синтез нуклеиновых кислот проводили на устройстве, характеризующемся по меньшей мере двумя кластерами, причем каждый кластер характеризуется 121 индивидуально адресуемым локусом.

[00208] Внутренний 5' праймер **225** и внутренний 3' праймер **220** синтезировали в отдельных кластерах. Внутренний 5' праймер **225** подвергли репликации 121 раз, с удлинением на 121 локус в пределах одного кластера. Для внутреннего 3' праймера **220** каждый из 19 праймеров вариантных последовательностей удлинялся на 6 различных локусах, получая в результате удлинение 114 нуклеиновых кислот на 114 различных локусах.

[00209] Синтезированные нуклеиновые кислоты отщепляли от поверхности устройства и переносили в пластиковый флакон. Первую реакцию ПЦР проводили с использованием фрагментов длинной последовательности нуклеиновой кислоты **235**, **240** для амплификации матричной нуклеиновой кислоты, как проиллюстрировано на фиг. **2B**. Вторую реакцию ПЦР проводили с использованием комбинации праймеров и продуктов первой реакции ПЦР в качестве матрицы, как проиллюстрировано на фиг. **2C-2D**. Анализ продуктов второй ПЦР проводили на BioAnalyzer, как показано на графике на фиг. **17**.

[00210] Пример 5: Создание библиотеки нуклеиновых кислот, содержащей 96 различных наборов вариантов с одним положением

[00211] Четыре набора праймеров, как в общем показано на фиг. **2A** и представлено в примере 2, создавали с использованием синтеза нуклеиновой кислоты de novo. Для внутреннего 5' праймера **220** создавали 96 различных наборов праймеров, причем каждый набор праймеров нацелен на один различный кодон, расположенный на одном сайте матричной нуклеиновой кислоты. Для каждого набора праймеров создавали 19 различных вариантов, каждый из которых содержал кодон, кодирующий различную аминокислоту на одном сайте. Два раунда ПЦР проводили с использованием созданных праймеров, как в общем показано на фиг. **2A-2D** и описано в примере 2. 96 наборов продуктов амплификации визуализировали на электрофореграмме (фиг. **18**), которую использовали для расчета 100%-ного показателя успешности амплификации.

[00212] Пример 6: Создание библиотеки нуклеиновых кислот, содержащей 500 различных наборов вариантов с одним положением

[00213] Четыре набора праймеров, как в общем показано на фиг. **2A** и представлено в примере 2, создавали с использованием синтеза нуклеиновой кислоты de novo. Для внутреннего 5' праймера **220** создавали 500 различных наборов праймеров, причем каждый набор праймеров нацелен на один различный кодон, расположенный в пределах одного сайта матричной нуклеиновой кислоты. Для каждого набора праймеров

создавали 19 различных вариантов, каждый из которых содержал кодон, кодирующий различную аминокислоту на одном сайте. Два раунда ПЦР проводили с использованием созданных праймеров, как в общем показано на фиг. 2А и описано в примере 2. На электрофореграммах показан каждый из 500 наборов продуктов ПЦР, характеризующийся популяцией нуклеиновых кислот с 19 вариантами в одном различном сайте (данные не показаны). Всесторонний анализ секвенирования библиотеки показал более чем 99% показатель успешности среди предварительно выбранных мутаций кодонов (данные последовательности и данные анализа не показаны).

[00214] Пример 7: Праймеры односайтового мутагенеза для 1 положения

[00215] Пример конструкции вариации кодона представлен в таблице 9 для желтого флуоресцентного белка. В этом случае один кодон из 50-мера последовательности подвергают вариации 19 раз. Вариантная последовательность нуклеиновой кислоты показана жирным шрифтом. Последовательность праймера дикого типа представляет собой: ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT (SEQ ID NO.: 1). В этом случае кодон дикого типа кодирует валин, что показано подчеркиванием в SEQ ID NO.: 1. Следовательно, 19 вариантов ниже исключают кодон, кодирующий валин. В альтернативном примере, если нужно рассмотреть все триплеты, то будут созданы все 60 вариантов, включая в себя альтернативную последовательность для кодона дикого типа.

Таблица 9. Вариантные последовательности

SEQ ID NO.	Вариантная последовательность	Вариантный кодон
2	atg T TTAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	F
3	atg T TAAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	L
4	atg A TTAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	I
5	atg T CTAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	S
6	atg C CTAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	P
7	atg A CTAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	T
8	atg G CTAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	A
9	atg T ATAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	Y
10	atg C ATAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	H
11	atg C AAAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	Q
12	atg A ATAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	N
13	atg A AAAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	K
14	atg G ATAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	D
15	atg G AAAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	E
16	atg T GTAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	C

17	atgTGGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	W
18	atgCGTAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	R
19	atgGGTAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT	G

[00216] Пример 8: Варианты нуклеиновой кислоты с двумя положениями на одном сайте

[00217] De novo синтез нуклеиновых кислот проводили при условиях, сходных с условиями, описанными в примере 2. Создавали один кластер на устройстве, который содержал синтезированные предварительно определенные варианты нуклеиновой кислоты для 2 последовательных положений кодонов на одном сайте, причем каждое положение представляет собой кодон, кодирующий аминокислоту. В этой конфигурации 19 вариантов на каждое положение создавали для 2 положений с 3 копиями каждой нуклеиновой кислоты, давая в результате 114 синтезированных нуклеиновых кислот.

[00218] Пример 9: Варианты нуклеиновой кислоты с двумя положениями на нескольких сайтах

[00219] De novo синтез нуклеиновых кислот проводили при условиях, сходных с условиями, описанными в примере 2. Создавали один кластер на устройстве, который содержал синтезированные предварительно определенные варианты нуклеиновой кислоты для 2 непоследовательных положений кодонов, причем каждое положение представляет собой кодон, кодирующий аминокислоту. В этой конфигурации 19 вариантов на каждое положение создавали для 2 положений.

[00220] Пример 10: Варианты нуклеиновой кислоты с тремя положениями на одном сайте

[00221] De novo синтез нуклеиновых кислот проводили при условиях, сходных с условиями, описанными в примере 2. Создавали один кластер на устройстве, который содержал синтезированные предварительно определенные варианты эталонной нуклеиновой кислоты для 3 последовательных положений кодонов. В конфигурации с 3 последовательными положениями кодонов 19 вариантов на каждое положение создавали для 3 положений с 2 копиями каждой нуклеиновой кислоты, и это давало в результате 114 синтезированных нуклеиновых кислот.

[00222] Пример 11: Варианты нуклеиновой кислоты с тремя положениями на нескольких сайтах

[00223] De novo синтез нуклеиновых кислот проводили при условиях, сходных с условиями, описанными в примере 2. Создавали один кластер на устройстве, который содержал синтезированные предварительно определенные варианты эталонной

нуклеиновой кислоты для по меньшей мере 3 непоследовательных положений кодонов. В пределах предварительно определенной области расположение кодонов, кодирующих 3 остатка гистидина, подвергали вариации.

[00224] Пример 12: Варианты нуклеиновой кислоты с несколькими положениями на нескольких сайтах

[00225] De novo синтез нуклеиновых кислот проводили при условиях, сходных с условиями, описанными в примере 2. Создавали один кластер на устройстве, который содержал синтезированные предварительно определенные варианты эталонной нуклеиновой кислоты для 1 или нескольких положений кодонов в 1 или нескольких участках. Пять положений подвергали вариации в библиотеке. Первое положение, кодируемое кодонами для полученного отношения K/R 50/50 в экспрессированном белке; второе положение, кодируемое кодонами для полученного отношения V/L/S 50/25/25 в экспрессированном белке, третье положение, кодируемое кодонами для полученного отношения Y/R/D 50/25/25 в экспрессированном белке, четвертое положение, кодируемое кодонами для полученного равного отношения для всех аминокислот в экспрессированном белке, и пятое положение, кодируемое кодонами для полученного отношения G/P 75/25 в экспрессированном белке.

[00226] Пример 13: Компоненты модульной плазмиды для экспрессии разнообразных пептидов

[00227] Создают библиотеку полинуклеотидов, как в примерах 4-6 и 8-12, кодирующую вариацию кодонов на одном сайте или нескольких сайтах для каждой из отдельных областей, которые составляют части кассеты экспрессионного конструкта, как показано на фиг. 11. Для создания двух кассет экспрессионного конструкта синтезировали варианты нуклеиновых кислот, кодирующие по меньшей мере часть вариантной последовательности первого промотора 1110, первой открытой рамки считывания 1120, первого терминатора 1130, второго промотора 1140, второй открытой рамки считывания 1150 или последовательности второго терминатора 1160. После раундов амплификации, как описано в предыдущих примерах, создают библиотеку из 1024 экспрессионных конструктов.

[00228] Пример 14: Варианты с одним положением на нескольких сайтах

[00229] Создают библиотеку полинуклеотидов, как в примерах 4-6 и 8-12, кодирующую вариацию кодонов на одном сайте или нескольких сайтах в области, кодирующей по меньшей мере часть нуклеиновой кислоты. Создают библиотеку вариантов нуклеиновых кислот, причем библиотека состоит из вариантов с одним положением на нескольких сайтах. См., например, фиг. 6B.

[00230] Пример 15: Синтез библиотеки вариантов

[00231] De novo синтез нуклеиновых кислот проводят при условиях, сходных с условиями, описанными в примере 2. По меньшей мере 30000 неидентичных нуклеиновых кислот синтезируют de novo, причем каждая из неидентичных нуклеиновых кислот кодирует различные варианты кодонов аминокислотной последовательности. По меньшей мере 30000 неидентичных синтезированных нуклеиновых кислот характеризуются суммарной частотой ошибок, составляющей менее 1 на 1000 оснований по сравнению с предварительно определенными последовательностями для каждой из по меньшей мере 30000 неидентичных нуклеиновых кислот. Библиотеку используют для ПЦР-мутагенеза длинной нуклеиновой кислоты, и образуется по меньшей мере 30000 неидентичных вариантных нуклеиновых кислот.

[00232] Пример 16: Синтез библиотеки вариантов на основе кластеров

[00233] De novo синтез нуклеиновых кислот проводили при условиях, сходных с условиями, описанными в примере 2. Создавали один кластер на устройстве, который содержал синтезированные предварительно определенные варианты эталонной нуклеиновой кислоты для 2 положений кодонов. В конфигурации с 2 последовательными положениями кодонов 19 вариантов на положение создают для 2 положений с 2 копиями каждой нуклеиновой кислоты, и это дает в результате 38 синтезированных нуклеиновых кислот. Каждая вариантная последовательность составляет 40 оснований в длину. В одном и том же кластере создают дополнительную невариантную последовательность нуклеиновых кислот, причем дополнительные невариантные нуклеиновые кислоты и вариантные нуклеиновые кислоты совместно кодируют 38 вариантов кодирующей последовательности гена. Каждая из нуклеиновых кислот содержит по меньшей мере одну область, обратную комплементарную другой нуклеиновой кислоте. Нуклеиновые кислоты в кластере высвобождают с помощью расщепления газообразным аммиаком. Штифт, содержащий воду, контактирует с кластером, собирает нуклеиновые кислоты и перемещает нуклеиновые кислоты в небольшой флакон. Флакон также содержит реагенты ДНК-полимеразы для реакции полимеразной циклической сборки (РСА). Нуклеиновые кислоты отжигаются, пропуски заполняются с помощью реакции удлинения, и в результате образуются двухцепочечные молекулы ДНК, образуя библиотеку вариантных нуклеиновых кислот. Библиотеку вариантных нуклеиновых кислот необязательно подвергают действию рестрикционного фермента, затем лигируют в экспрессионные векторы.

[00234] Пример 17: Создание библиотеки вариантных нуклеиновых кислот TCR

[00235] Для создания библиотеки TCR с разнообразием, составляющим 10^8 , предпринимают следующие стадии. Мононуклеарные клетки периферической крови получают от 10 людей - доноров и выделяют Т-клетки. Все 10 субъектов характеризуются HLA, который, согласно картированию, характеризуется соответствующим аллелем, например, HLA-A*2:01. Из 10 людей - доноров 7 субъектов являются здоровыми донорами и 3 являются донорами с формой рака. Гены, вовлеченные в комплекс TCR, секвенируют из Т-клеток каждого донора, и идентифицируют 1000 наиболее распространенных последовательностей TCR. Каждая из общих последовательностей TCR будет охватывать от 600 до 900 пар оснований.

[00236] Конструируют 10000 предварительно определенных вариантных последовательностей нуклеиновых кислот для каждой из 1000 наиболее распространенных последовательностей TCR, получая в результате библиотеку ДНК с общим разнообразием, составляющим по меньшей мере 10^8 . Каждую вариантную последовательность вводят в экспрессионный вектор млекопитающего для экспрессии гена.

[00237] Пример 18: Экспрессия и скрининг библиотеки комплексов вариантных белков TCR

[00238] Библиотеку вариантных генов TCR переносят в клетки млекопитающих для создания библиотеки популяций клеток, причем каждая клеточная популяция экспрессирует различный вариантный белок TCR. Библиотеку белков подвергают скринингу на наличие комплексов TCR с улучшенной аффинностью (мера силы взаимодействия между эпитопом и антигенсвязывающим сайтом антитела) в отношении опухолевого антигена, такого как MAGE A3 или NY-ESO-1. Также оценивают дополнительные функциональные критерии, такие как экспрессия вариантного гена, авидность (мера общей силы комплекса антитело-антиген), стабильность и специфичность мишени.

[00239] Пример 19: Получение и доставка сконструированных Т-клеток

[00240] Т-клетки собирают от субъекта, у которого диагностировали рак и генетически конструируют с новым Т-клеточным рецептором (TCR), выбранным после проведения анализа в примере 18. После короткого периода экспансии *in vitro* и прохождения специфических для продукта критериев выпуска продукт Т-клеток вводят тому же самому субъекту. См. фиг. 10С.

[00241] Пример 20: Библиотеки вариантных TCR

[00242] Синтез нуклеиновых кислот *de novo* проводят в условиях, аналогичных описанным в примере 2, для создания нуклеиновых кислот, кодирующих варианты TCR. Варианты TCR содержат варианты в области контакта с антигеном между

TCR и опухолевым антигеном, таким как MAGE A3 или NY-ESO-1. Варианты создают на одном сайте во всей библиотеке вариантных TCR для вплоть до 30 остатков, в результате чего получают первую библиотеку с общим разнообразием, составляющим около 10^9 . Остатки для вариации находятся в переменных доменах альфа-цепи TCR и бета-цепи TCR.

[00243] Первая библиотека, содержащая вариантные TCR, затем экспрессируется в Т-клетках. Вариантные TCR подвергают скринингу *in vitro* против опухолевых антигенов в отношении специфичности вариантных TCR к опухолевому антигену. Вариантные TCR, которые являются высокоспецифичными в отношении опухолевого антигена, затем подвергают дальнейшему варьированию для создания второй библиотеки. Вторая библиотека содержит вариацию остатков, расположенных в константных доменах альфа-цепи TCR и бета-цепи TCR. Вторая библиотека экспрессируется в Т-клетках и ее подвергают скринингу *in vitro* в отношении второго улучшения, например, авидности, стабильности, аффинности или экспрессии.

[00244] Выбранные варианты гена TCR или генных фрагментов TCR, характеризующиеся желательными признаками, после скрининга продуктов из первой и/или второй библиотек вариантов отбирают для разработки потенциального терапевтического средства, необязательно в форме растворимого белка TCR, в котором отсутствует область для закрепления белка к Т-клетке или в форме гена для экспрессии в сконструированной Т-клетке как часть иммунотерапии, включающей в себя перепрограммирование Т-клеток (либо от субъекта, либо от совпадающего по HLA донора) для нацеливания на раковые клетки у субъекта, страдающего от рака.

[00245] Несмотря на то, что в настоящем документе были показаны и описаны предпочтительные варианты осуществления настоящего раскрытия, специалистам в настоящей области техники будет очевидно, что такие варианты осуществления представлены только в качестве примера. Специалистам в настоящей области техники станут очевидными многочисленные варианты, изменения и замены без отклонения от настоящего раскрытия. Следует понимать, что различные альтернативы вариантам осуществления настоящего раскрытия, описанным в настоящем документе, можно использовать при практическом применении настоящего раскрытия. Подразумевается, что приведенная ниже формула изобретения определяет объем настоящего раскрытия и что таким образом будут охватываться способы и структуры в пределах объема этой формулы изобретения и их эквиваленты.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Библиотека нуклеиновых кислот, причем библиотека нуклеиновых кислот содержит по меньшей мере 3000 вариантных нуклеиновых кислот, причем каждая вариантная нуклеиновая кислота кодирует вариантный ген или фрагмент гена для белка Т-клеточного рецептора (TCR), и причем область, выбранная для вариации, охватывает вплоть до 1000 оснований в длину.

2. Библиотека нуклеиновых кислот по п. 1, в которой вариантный ген или фрагмент гена происходит из варибельного домена.

3. Библиотека нуклеиновых кислот по п. 2, в которой варибельный домен представляет собой варибельный домен TCR-альфа, TCR-бета, TCR-гамма или TCR-дельта.

4. Библиотека нуклеиновых кислот по п. 2, в которой варибельный домен является специфическим в отношении ракового антигена.

5. Библиотека нуклеиновых кислот по п. 4, причем раковый антиген представляет собой MAGE A3, MAGE A12, MAGE A2, MAGE A6, NY-ESO-1 или CEA.

6. Библиотека нуклеиновых кислот по п. 1, в которой вариантный ген или фрагмент гена происходит из константного домена.

7. Библиотека нуклеиновых кислот по п. 1, в которой вариантный ген или фрагмент гена позволяет создать белок TCR, характеризующийся увеличенной специфичностью, авидностью, аффинностью, стабильностью или экспрессией.

8. Библиотека полинуклеотидов, причем библиотека полинуклеотидов содержит по меньшей мере 3000 вариантных нуклеиновых кислот, причем каждый полинуклеотид составляет по меньшей мере 15 оснований в длину, причем каждый полинуклеотид кодирует вариант в пределах варибельного домена белка Т-клеточного рецептора (TCR) или его фрагмента, причем варибельный домен содержит вплоть до 1000 оснований, и причем каждая вариантная нуклеиновая кислота содержит по меньшей мере одну вариацию на предварительно выбранном кодоне для аминокислотного остатка в области контакта с антигеном.

9. Библиотека полинуклеотидов по п. 8, в которой вариабельный домен представляет собой вариабельный домен TCR-альфа, TCR-бета, TCR-гамма или TCR-дельта.

10. Библиотека полинуклеотидов по п. 8, причем антиген представляет собой раковый антиген.

11. Библиотека полинуклеотидов по п. 10, причем раковый антиген представляет собой MAGE A3, MAGE A12, MAGE A2, MAGE A6, NY-ESO-1 или CEA.

12. Библиотека полинуклеотидов по п. 8, в которой каждая вариантная нуклеиновая кислота содержит множество вариаций на предварительно выбранных кодонах для аминокислотных остатков в области контакта с антигеном.

13. Библиотека полинуклеотидов по п. 12, в которой множество вариаций на предварительно выбранных кодонах для аминокислотных остатков содержит вплоть до 100 остатков.

14. Библиотека полинуклеотидов по п. 12, в которой множество вариаций на предварительно выбранных кодонах для аминокислотных остатков содержит вплоть до 30 остатков.

15. Библиотека полинуклеотидов по п. 12, в которой множество вариаций на предварительно выбранных кодонах для аминокислотных остатков содержит вплоть до 5 остатков.

16. Библиотека полинуклеотидов по п. 8, в которой вариант позволяет создать белок TCR, характеризующийся увеличенной специфичностью, авидностью, аффинностью, стабильностью или экспрессией.

17. Библиотека полинуклеотидов по п. 8, в которой каждая вариантная нуклеиновая кислота дополнительно содержит по меньшей мере одну вариацию на предварительно выбранном кодоне для аминокислотного остатка в пределах константного домена белка Т-клеточного рецептора (TCR) или его фрагмента.

18. Библиотека полинуклеотидов, причем библиотека полинуклеотидов содержит по меньшей мере 3000 вариантных полинуклеотидов, причем каждый полинуклеотид составляет по меньшей мере 15 оснований в длину, причем каждый полинуклеотид кодирует вариант в пределах константного домена белка Т-клеточного рецептора (TCR) или его фрагмента, причем константный домен содержит вплоть до 1000 оснований, и причем вариант содержит по меньшей мере одну вариацию на предварительно выбранном кодоне для аминокислотного остатка в области контакта с антигеном.

19. Библиотека полинуклеотидов по п. 18, в которой константный домен представляет собой константный домен TCR-альфа, TCR-бета, TCR-гамма или TCR-дельта.

20. Библиотека полинуклеотидов по п. 18, причем антиген представляет собой раковый антиген.

21. Библиотека полинуклеотидов по п. 20, причем раковый антиген представляет собой MAGE A3, MAGE A12, MAGE A2, MAGE A6, NY-ESO-1 или CEA.

22. Библиотека полинуклеотидов по п. 18, в которой каждый полинуклеотид содержит множество вариаций на предварительно выбранных кодонах для аминокислотных остатков.

23. Библиотека полинуклеотидов по п. 22, в которой множество вариаций на предварительно выбранных кодонах для аминокислотных остатков содержит вплоть до 100 остатков.

24. Библиотека полинуклеотидов по п. 22, в которой множество вариаций на предварительно выбранных кодонах для аминокислотных остатков содержит вплоть до 30 остатков.

25. Библиотека полинуклеотидов по п. 22, в которой множество вариаций на предварительно выбранных кодонах для аминокислотных остатков содержит вплоть до 5 остатков.

26. Библиотека нуклеиновых кислот, причем библиотека нуклеиновых кислот содержит по меньшей мере 10000 вариантных нуклеиновых кислот, причем каждая вариантная нуклеиновая кислота составляет около 500 - около 1000 оснований в длину,

причем каждая вариантная нуклеиновая кислота кодирует вариант эталонной последовательности, которая кодирует последовательность экзона для белка Т-клеточного рецептора или его фрагмента, и причем эталонная последовательность получена от субъекта, который характеризуется наличием рака.

27. Библиотека нуклеиновых кислот, причем библиотека содержит по меньшей мере около 1000000 вариантных нуклеиновых кислот, причем каждая вариантная нуклеиновая кислота составляет около 500 - около 1000 оснований в длину, причем каждая вариантная нуклеиновая кислота кодирует вариант эталонной последовательности, которая кодирует последовательность экзона для белка Т-клеточного рецептора или его фрагмента, причем эталонная последовательность белка Т-клеточного рецептора или его фрагмента выбрана на основании сравнения последовательности гена Т-клеточного рецептора от множества субъектов в отношении последовательности нуклеиновой кислоты, которая является общей среди субъектов, и причем по меньшей мере у части субъектов диагностирован рак.

28. Библиотека нуклеиновых кислот по любому из пп. 26 или 27, причем рак представляет собой солидный рак или рак кроветворной системы.

29. Библиотека нуклеиновых кислот по любому из пп. 26 или 27, причем библиотека содержит около 10000000 вариантных нуклеиновых кислот.

30. Библиотека нуклеиновых кислот по любому из пп. 26 или 27, в которой каждая вариантная нуклеиновая кислота составляет около 600 - около 900 оснований в длину.

31. Библиотека нуклеиновых кислот по любому из пп. 26 или 27, причем каждая вариантная нуклеиновая кислота находится в последовательности вектора.

32. Библиотека нуклеиновых кислот по п. 31, причем последовательность вектора представляет собой последовательность вирусного вектора.

33. Библиотека белков, содержащая белки, кодируемые библиотекой нуклеиновых кислот по любому из пп. 1-7 или 26-32.

34. Клеточная библиотека, содержащая библиотеку нуклеиновых кислот по любому из пп. 0-7 или 26-32.

35. Способ синтеза библиотеки нуклеиновых кислот, включающий:

а. предоставление первого набора предварительно выбранных полинуклеотидных последовательностей, кодирующих по меньшей мере 3000 вариантных последовательностей гена TCR или фрагмента гена TCR, причем каждая вариантная последовательность содержит по меньшей мере одну вариацию на предварительно выбранном кодоне для аминокислотного остатка в области контакта с антигеном;

б. синтез первого набора предварительно выбранных полинуклеотидных последовательностей; и

с. скрининг белков, кодируемых первым набором полинуклеотидных последовательностей, в отношении активности связывания.

36. Способ по п. 35, в котором по меньшей мере одна вариация находится в кодирующей вариабельный домен области гена TCR или фрагмента гена TCR.

37. Способ по п. 36, в котором вариабельный домен представляет собой вариабельный домен TCR-альфа, TCR-бета, TCR-гамма или TCR-дельта.

38. Способ по п. 35, в котором по меньшей мере одна вариация находится в кодирующей константный домен области гена TCR или фрагмента гена TCR.

39. Способ по п. 38, в котором константный домен представляет собой константный домен TCR-альфа, TCR-бета, TCR-гамма или TCR-дельта.

40. Способ по п. 35, в котором антиген представляет собой раковый антиген.

41. Способ по п. 40, в котором раковый антиген представляет собой MAGE A3, MAGE A12, MAGE A2, MAGE A6, NY-ESO-1 или CEA.

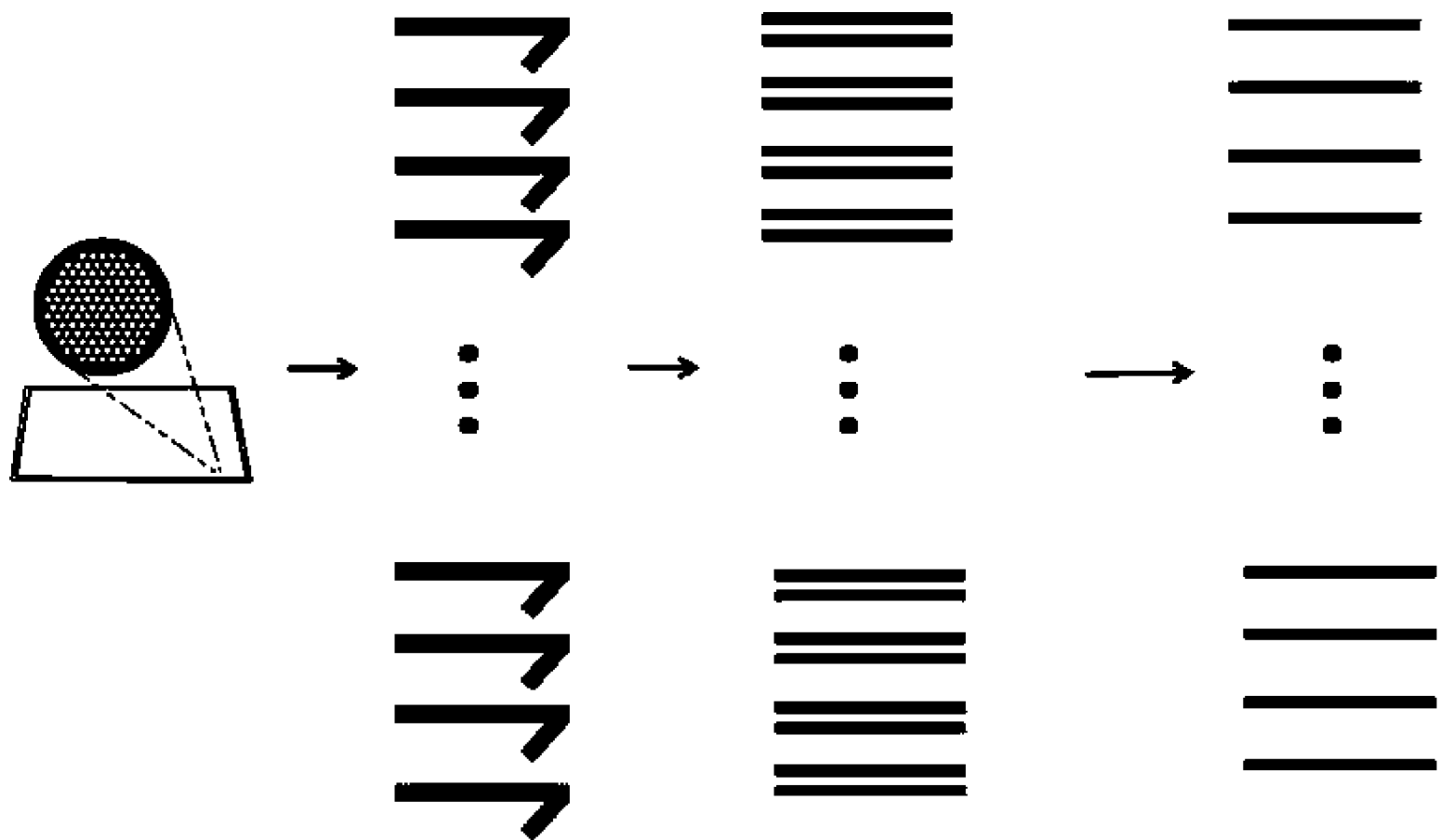
42. Способ по п. 38, в котором каждая вариантная последовательность содержит вплоть до 100 вариаций на предварительно выбранных кодонах для аминокислотных остатков в области контакта с антигеном.

43. Способ по п. 38, в котором каждая вариантная последовательность содержит вплоть до 30 вариаций на предварительно выбранных кодонах для аминокислотных остатков в области контакта с антигеном.

44. Способ по п. 38, дополнительно включающий:

- а. предоставление второго набора предварительно выбранных полинуклеотидных последовательностей, кодирующих по меньшей мере одну вариантную последовательность гена TCR или фрагмента гена TCR, где каждая вариантная последовательность содержит по меньшей мере одну вариацию на предварительно выбранном кодоне для аминокислотного остатка в гене TCR или фрагменте гена TCR в области, кодирующей константный домен;
- б. синтез второго набора предварительно выбранных полинуклеотидных последовательностей; и
- с. скрининг белков, кодируемых вторым набором полинуклеотидных последовательностей, в отношении второй активности.

45. Способ по п. 44, в котором вторая активность представляет собой уничтожение раковых клеток, экспрессия белка или стабильность белка.



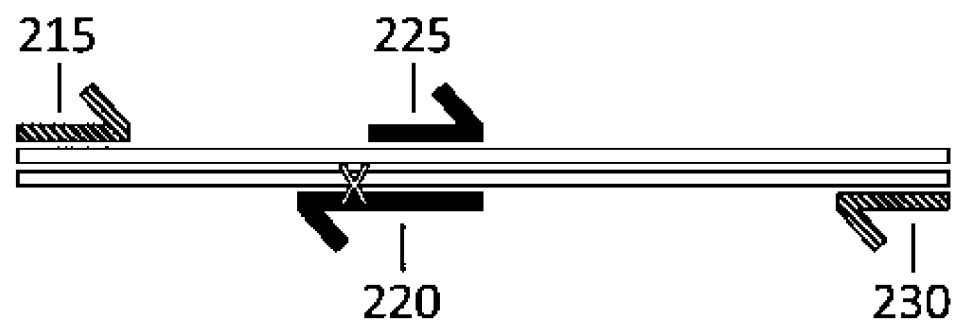
ФИГУРА 1А

ФИГУРА 1В

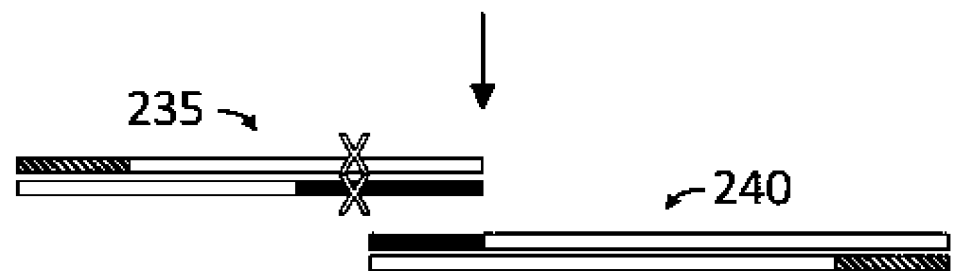
ФИГУРА 1С

ФИГУРА 1D

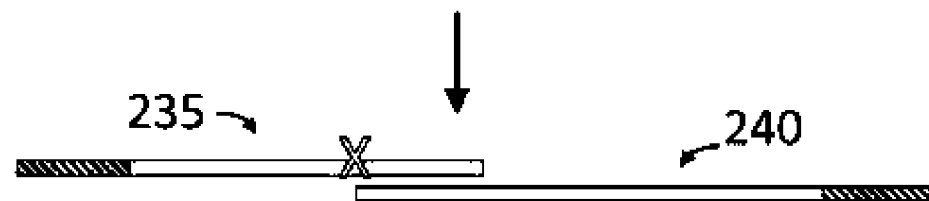
ФИГУРА 2А



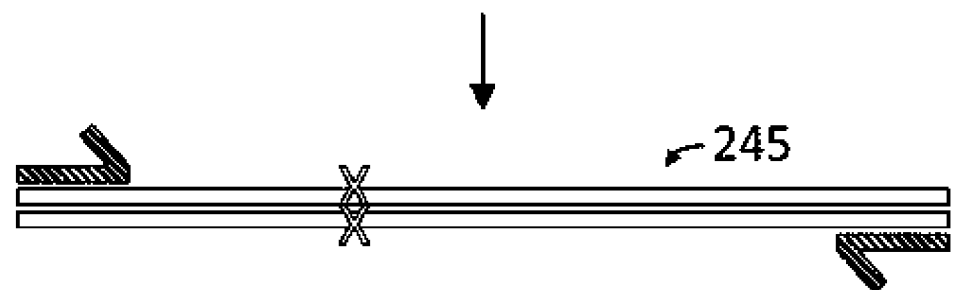
ФИГУРА 2В



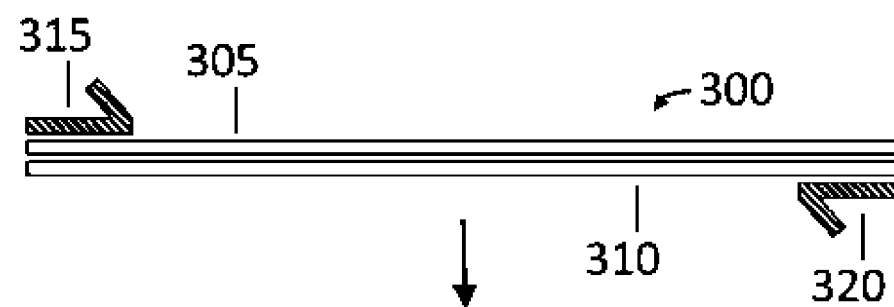
ФИГУРА 2С



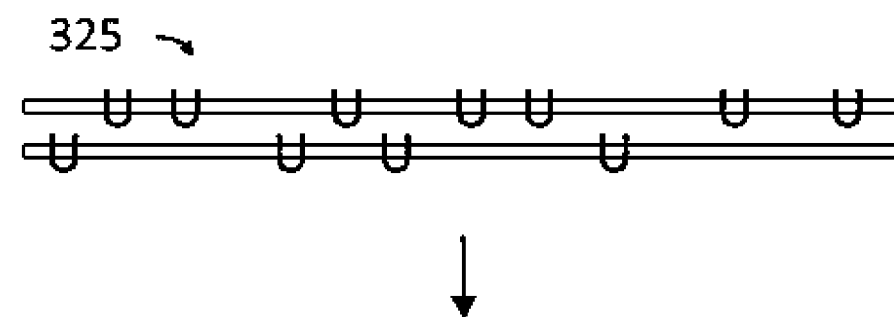
ФИГУРА 2D



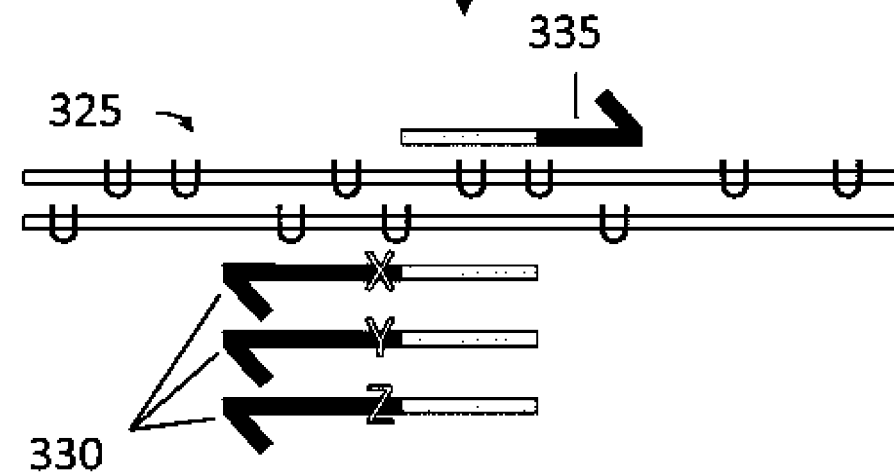
ФИГУРА 3А



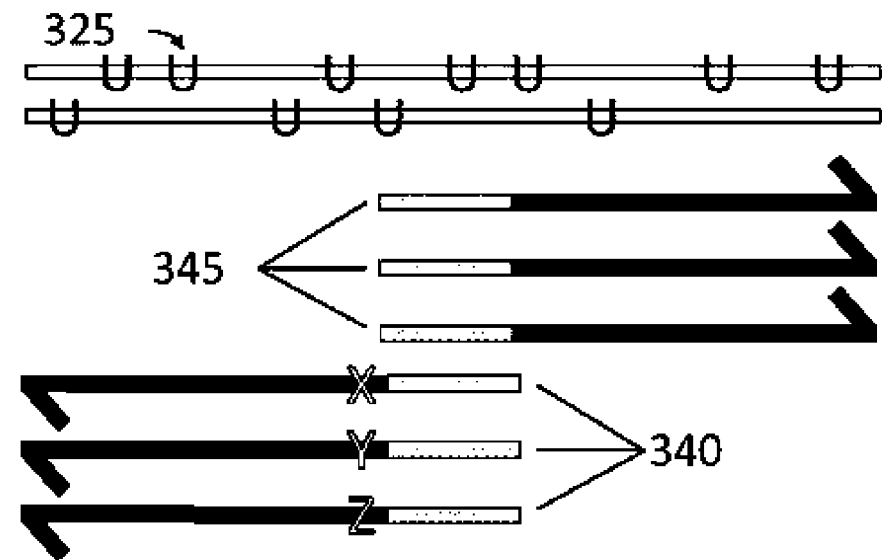
ФИГУРА 3В



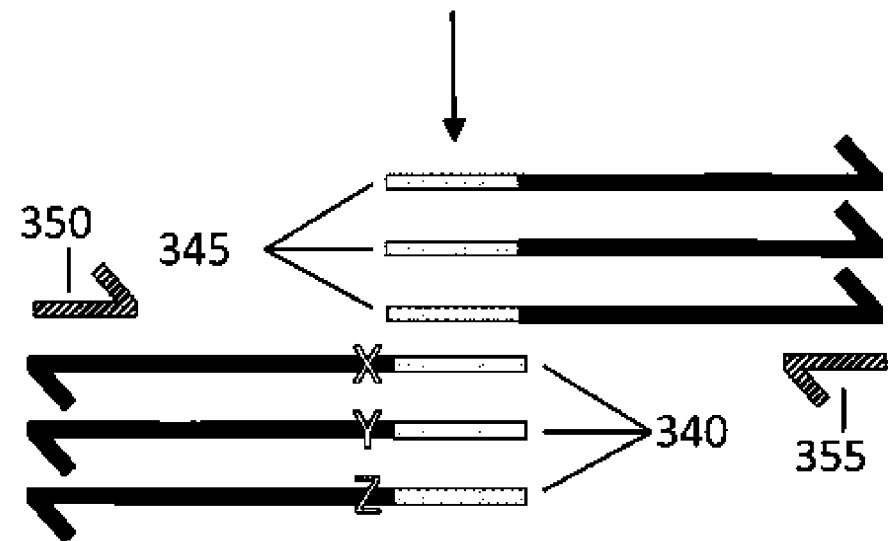
ФИГУРА 3С



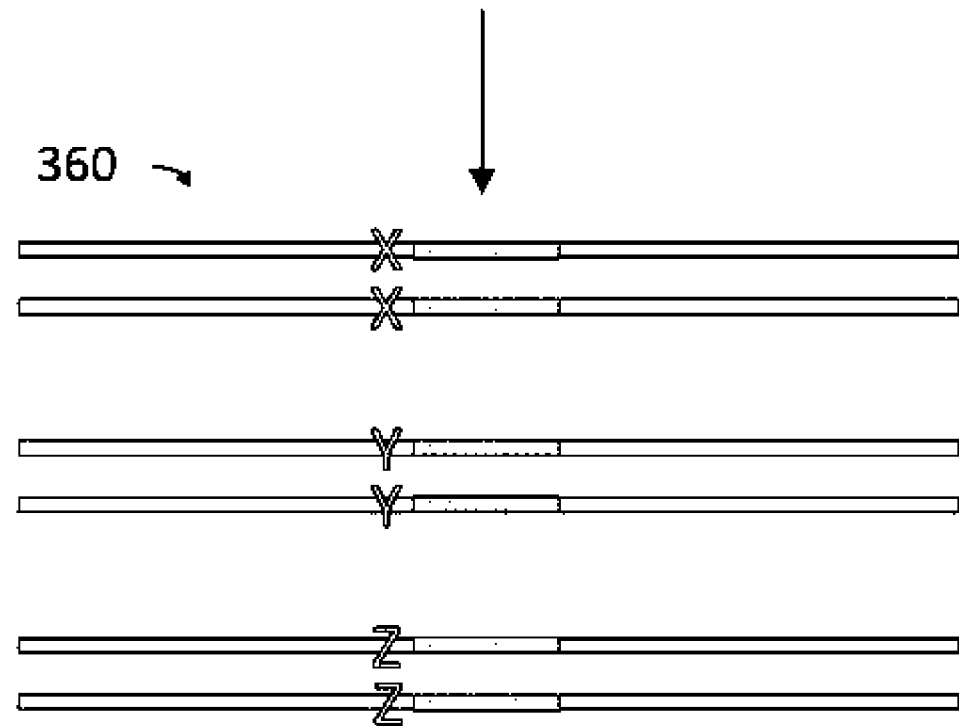
ФИГУРА 3D



ФИГУРА 3E



ФИГУРА 3F



ФИГУРА 4А



ФИГУРА 4В



ФИГУРА 4С



ФИГУРА 4D



ФИГУРА 4Е



ФИГУРА 5А

A W I K R E Q



ФИГУРА 5В

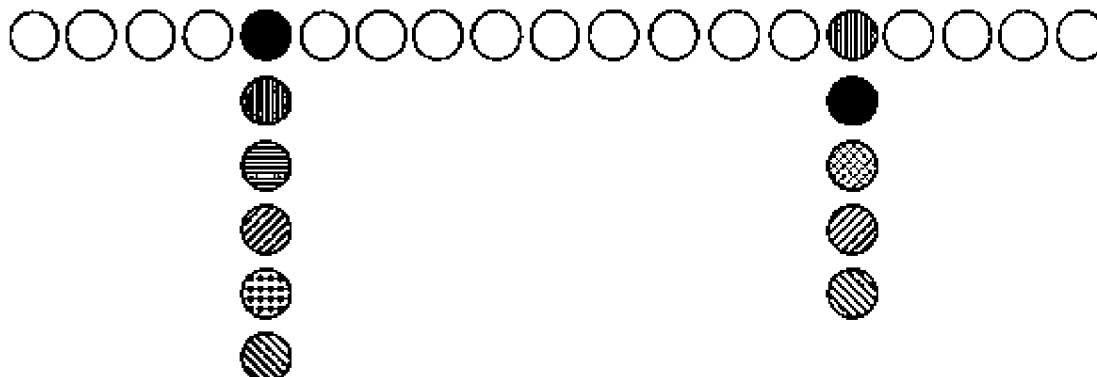
X W I K R E Q
 A X I K R E Q
 A W X K R E Q
 A W I X R E Q
 A W I K X E Q
 A W I K R X Q
 A W I K R E X

X = любой кодон

ФИГУРА 6А



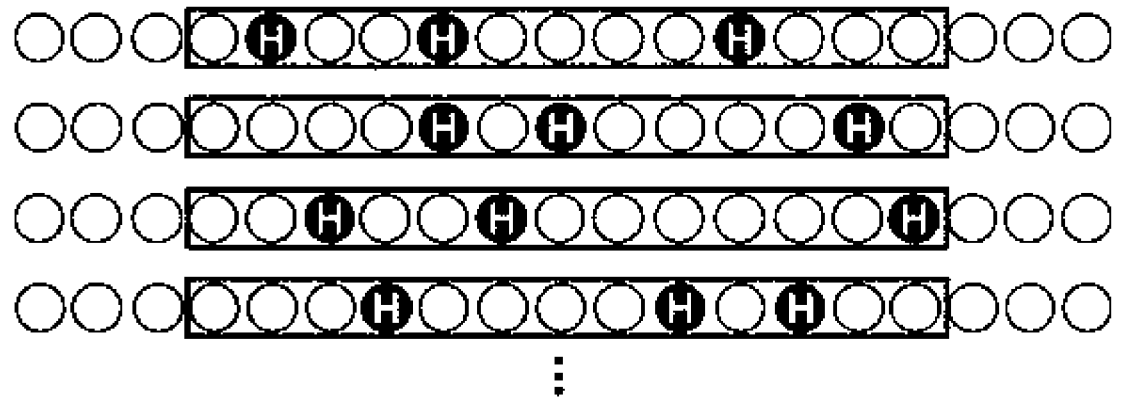
ФИГУРА 6В



ФИГУРА 7А



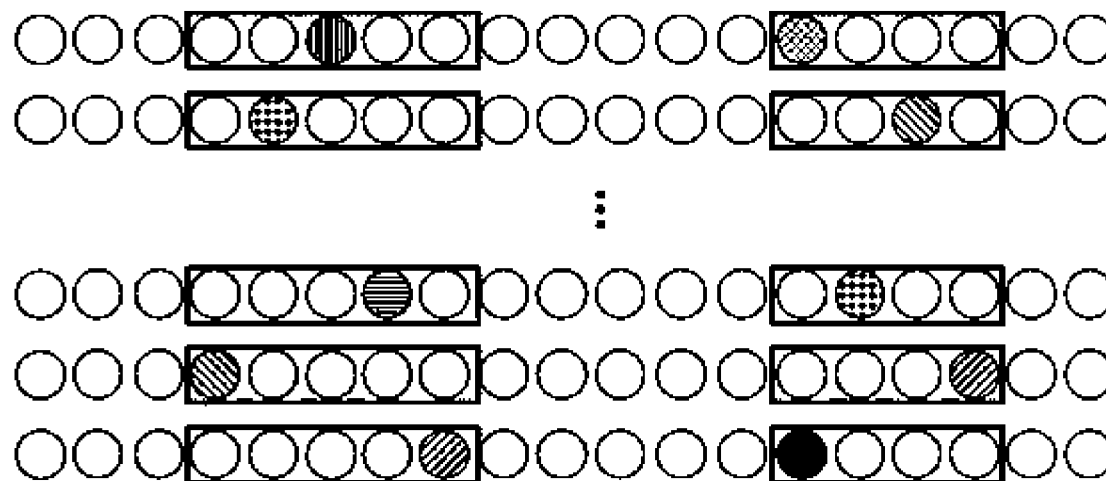
ФИГУРА 7В



ФИГУРА 8А



ФИГУРА 8В



ФИГУРА 9А



ФИГУРА 9В



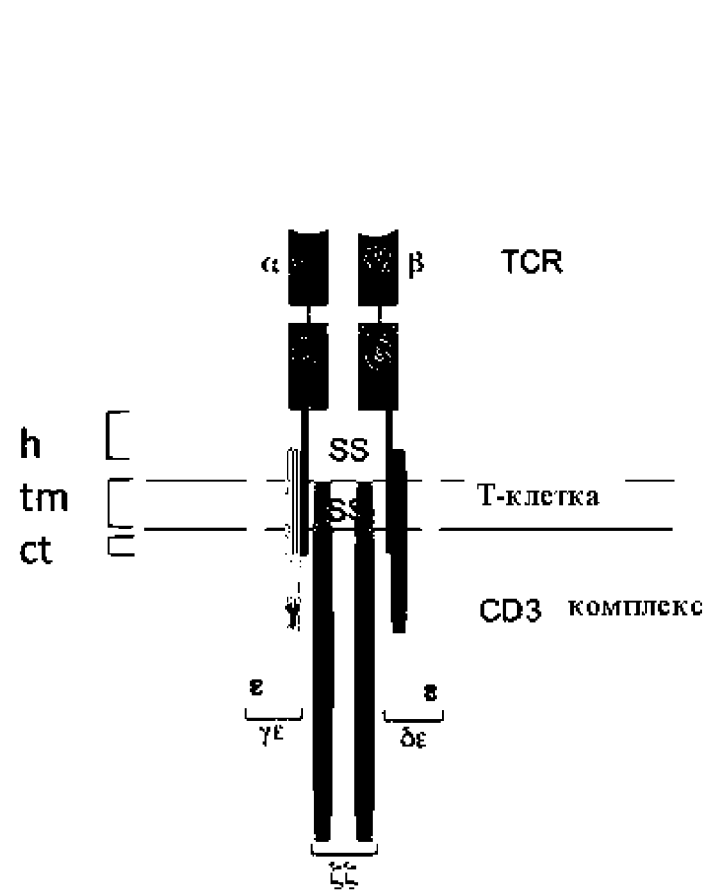
900

910

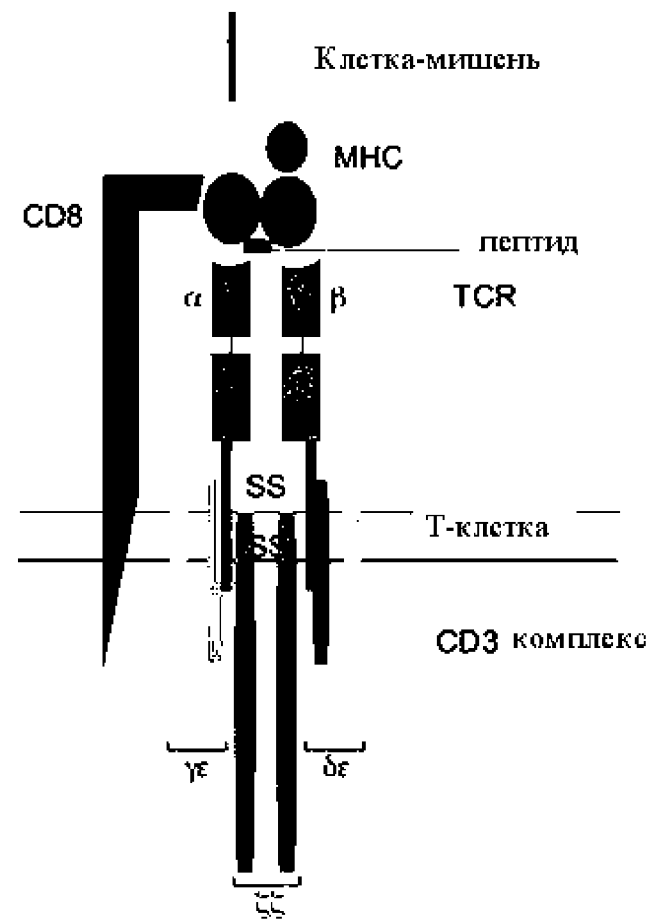
920

930

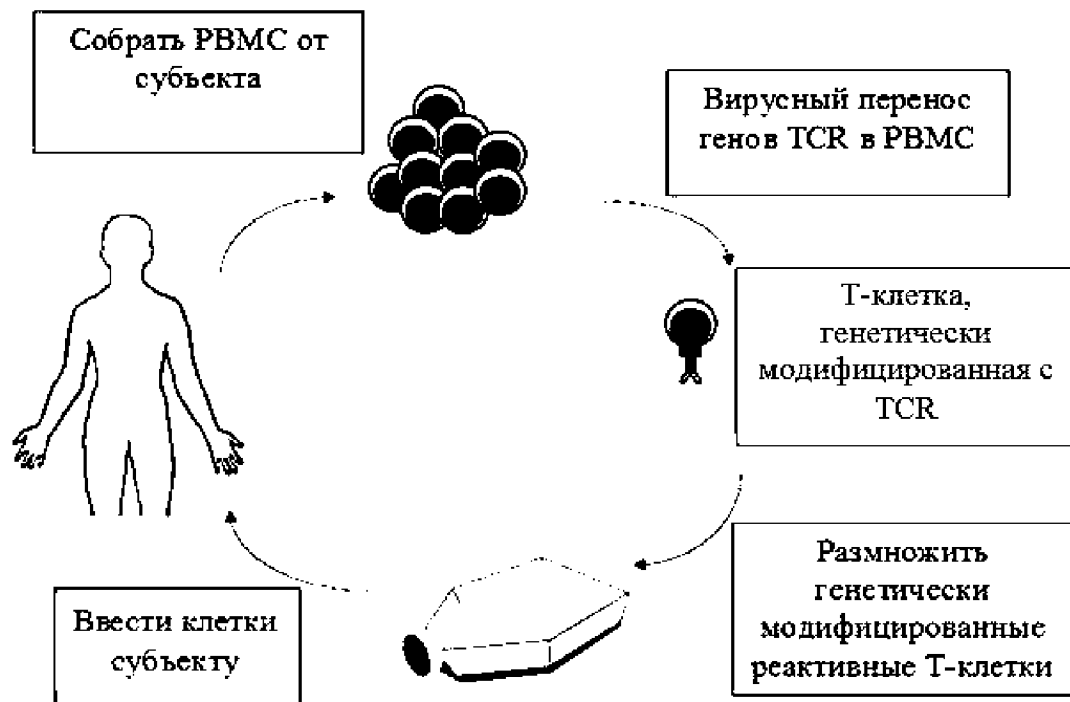
940

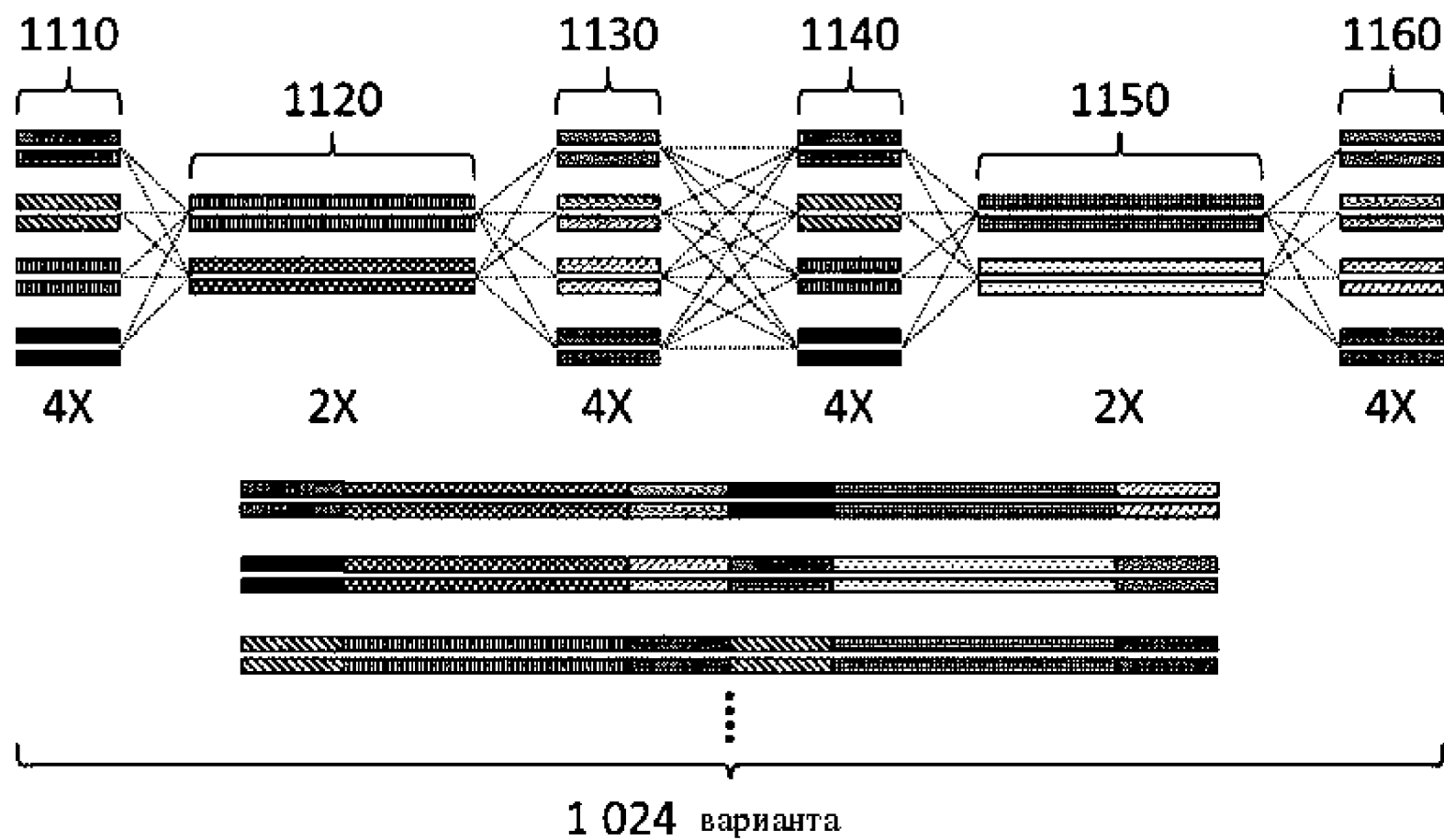


ФИГУРА 10А

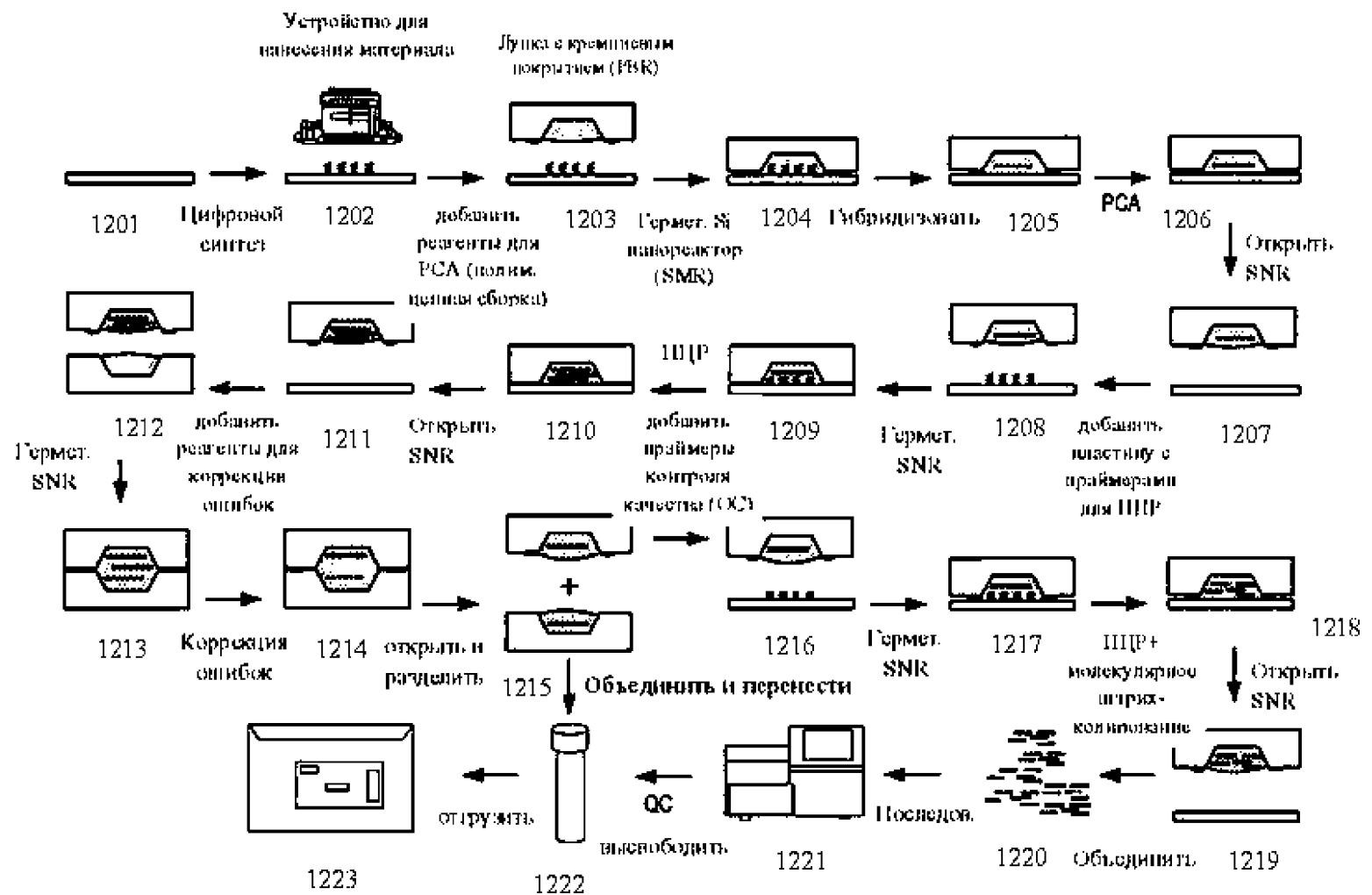


ФИГУРА 10В

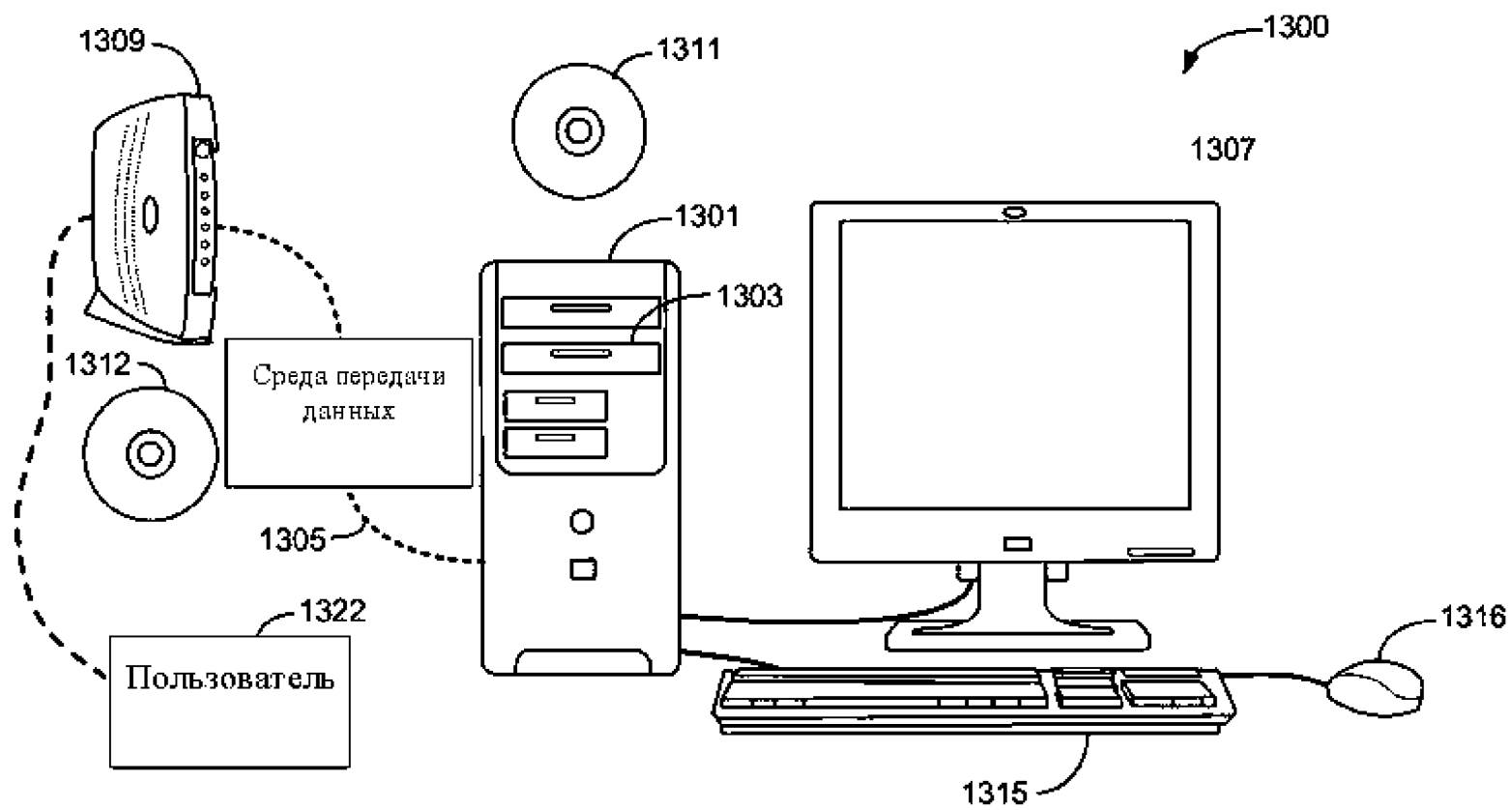
**ФИГУРА 10С**



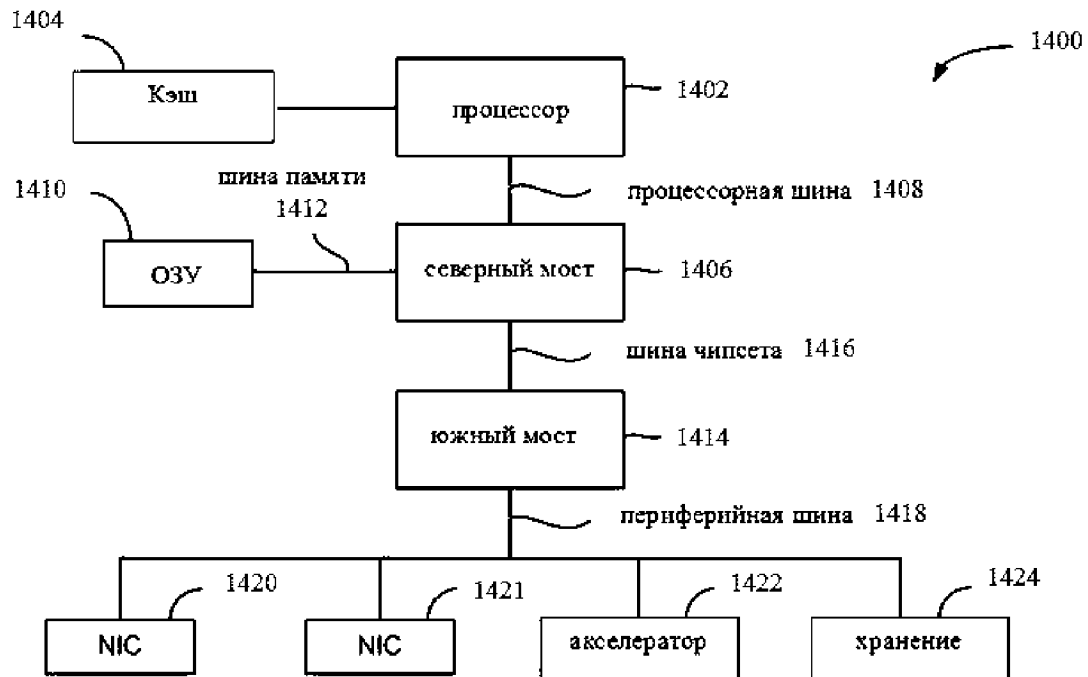
ФИГУРА 11



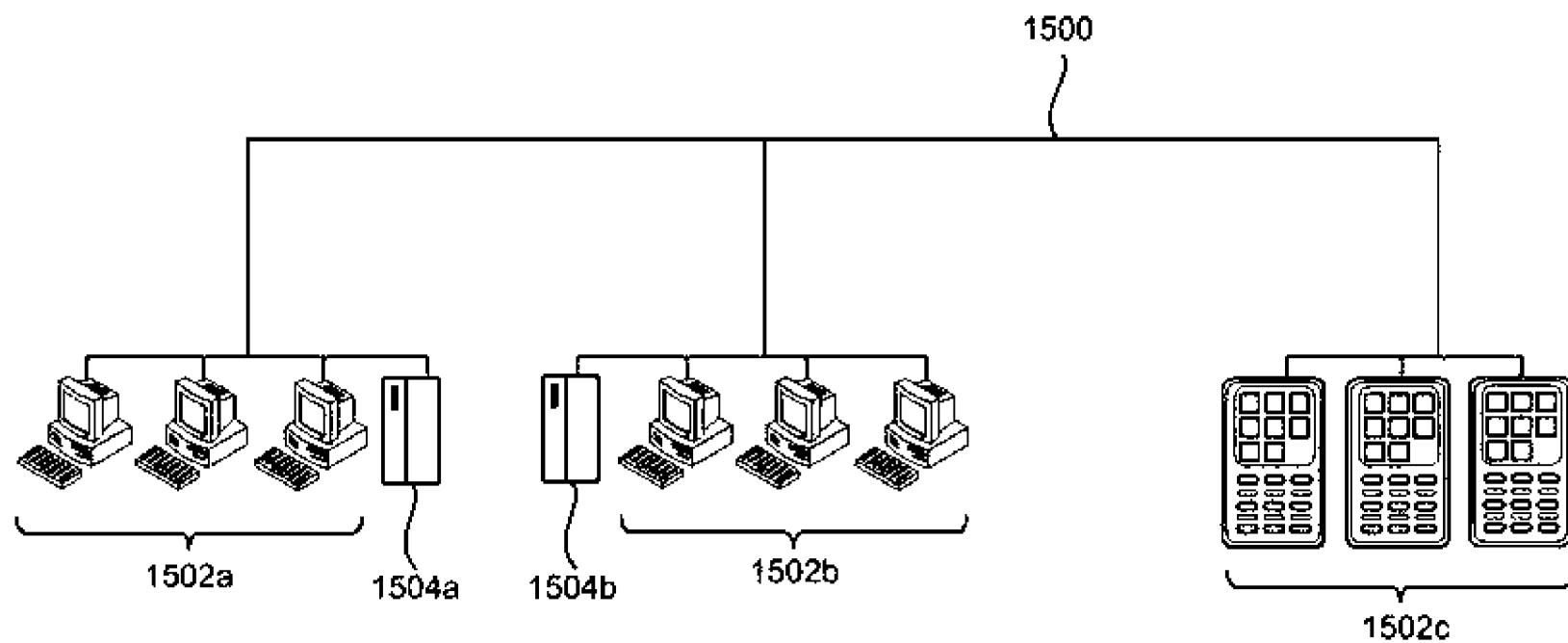
ФИГУРА 12



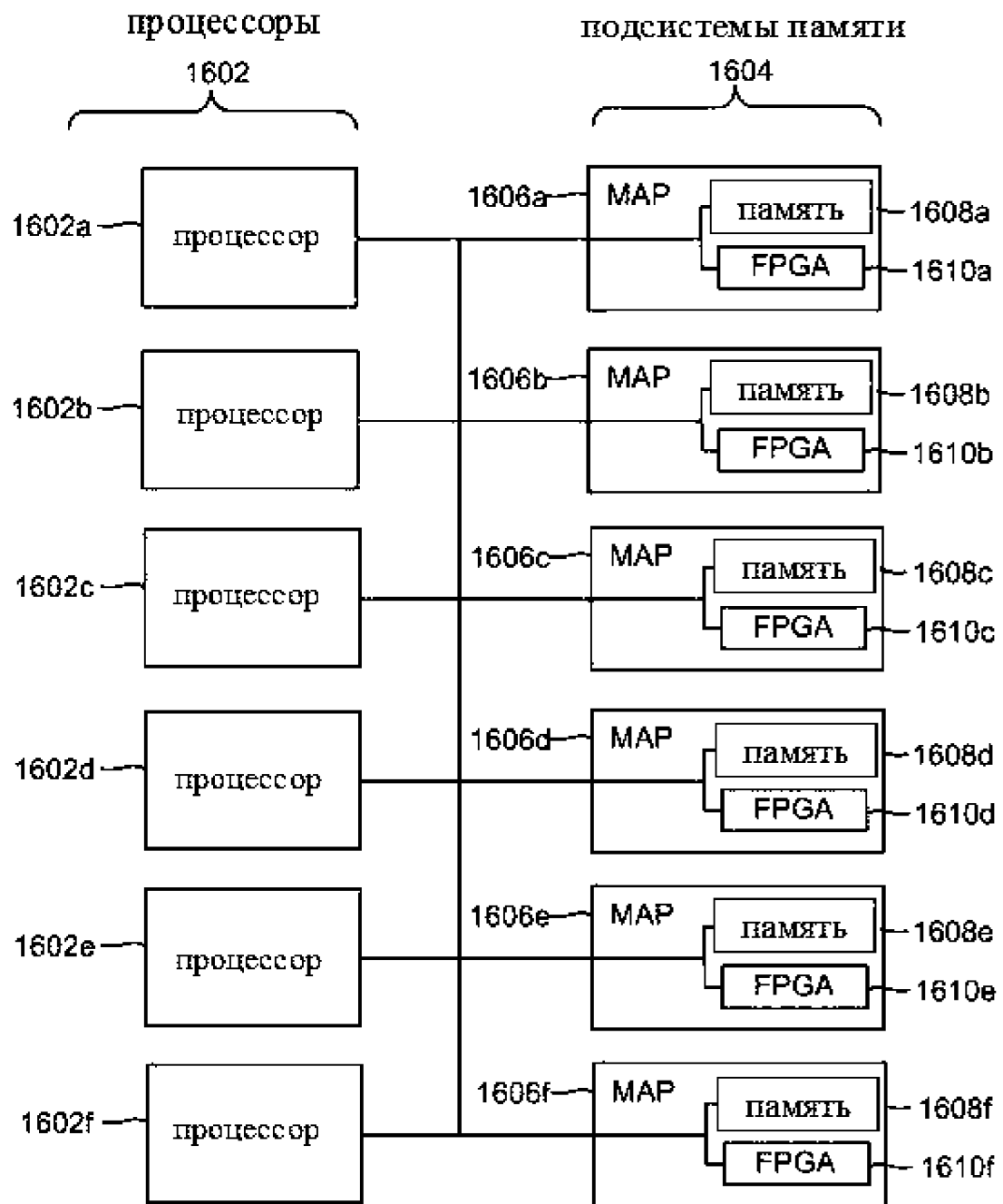
ФИГУРА 13

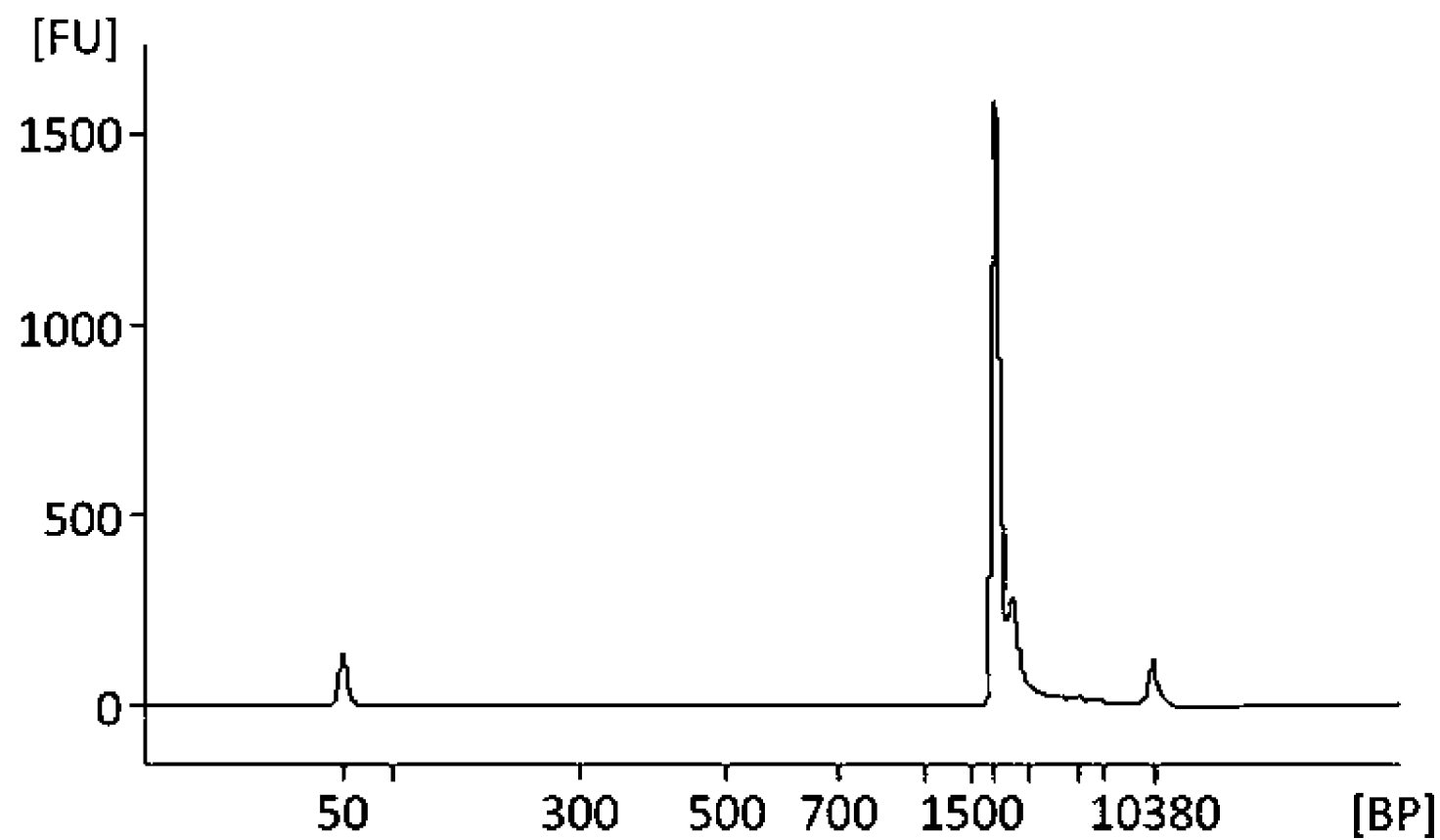


ФИГУРА 14

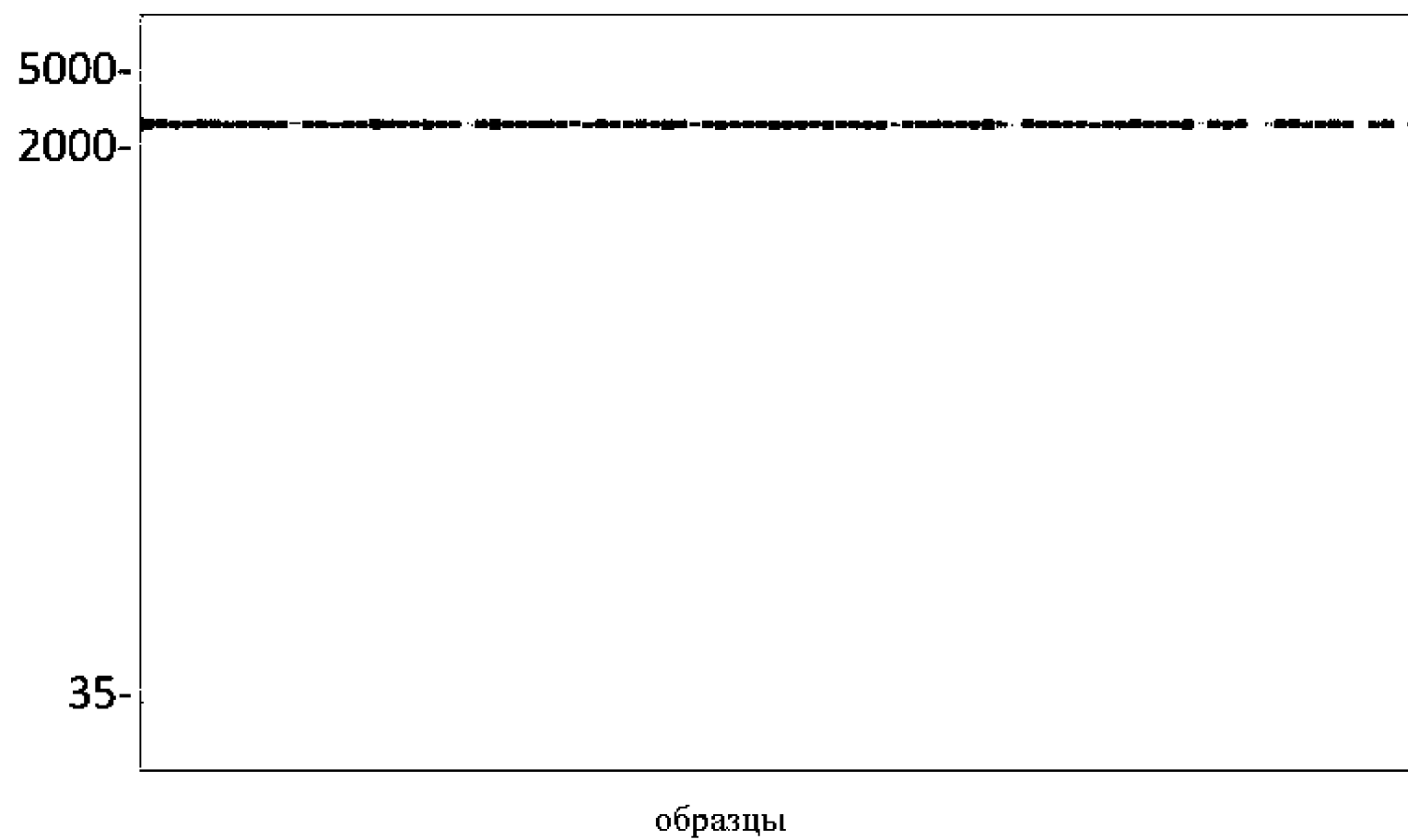


ФИГУРА 15

**ФИГУРА 16**



ФИГУРА 17



ФИГУРА 18