

(19)



Евразийское
патентное
ведомство

(21) 201991209 (13) A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2019.12.30

(51) Int. Cl. C40B 50/14 (2006.01)
C12Q 1/68 (2018.01)
C12N 15/10 (2006.01)

(22) Дата подачи заявки
2017.11.17

(54) БИБЛИОТЕКИ ПОЛИНУКЛЕОТИДОВ, ХАРАКТЕРИЗУЮЩИЕСЯ
КОНТРОЛИРУЕМОЙ СТЕХИОМЕТРИЕЙ, И ИХ СИНТЕЗ

(31) 62/424,302; 62/548,307; 62/558,666

(32) 2016.11.18; 2017.08.21; 2017.09.14

(33) US

(86) PCT/US2017/062391

(87) WO 2018/094263 2018.05.24

(71) Заявитель:

ТВИСТ БАЙОСАЙЕНС
КОРПОРЕЙШН (US)

(72) Изобретатель:

Зейтоун Рамси Ибрагим, Чэнь
Сьюань (US)

(74) Представитель:

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

(57) В настоящем документе предусмотрены композиции, способы и системы, относящиеся к библиотекам полинуклеотидов, характеризующимся предварительно выбранной стехиометрией в отношении видов полинуклеотидов так, чтобы библиотеки обеспечивали предварительно определенные результаты применения, например контролируемое представление после амплификации и однородное обогащение после связывания с целевыми последовательностями. Кроме того, в настоящем документе предусмотрены полинуклеотидные зонды и их применения для однородного и точного секвенирования нового поколения.



A1

201991209

201991209

A1

БИБЛИОТЕКИ ПОЛИНУКЛЕОТИДОВ, ХАРАКТЕРИЗУЮЩИЕСЯ КОНТРОЛИРУЕМОЙ СТЕХИОМЕТРИЕЙ, И ИХ СИНТЕЗ

ОПИСАНИЕ

Ссылка на родственные заявки

[0001] Согласно настоящей заявке испрашивается преимущество в соответствии с предварительной заявкой на выдачу патента США № 62/424302, поданной 18 ноября 2016 г., предварительной заявкой на выдачу патента США № 62/448307, поданной 21 августа 2017 г., и предварительной заявкой на выдачу патента США № 62/558666, поданной на 14 сентября 2017 г., каждая из которых полностью включена в настоящий документ посредством ссылки.

Перечень последовательностей

[0002] Настоящая заявка содержит перечень последовательностей, который был представлен в электронном виде в формате ASCII и настоящим полностью включен посредством ссылки. Указанная копия ASCII, созданная 13 ноября 2017 года, называется 44854-730_601_SL.txt и имеет размер 5304 байта.

Предшествующий уровень техники настоящего изобретения

[0003] Высокоэффективный химический генный синтез с высокой точностью и низкой стоимостью играет центральную роль в биотехнологии и медицине, а также в фундаментальных биомедицинских исследованиях. Генный синтез de novo является мощным инструментом для фундаментальных биологических исследований и биотехнологических применений. Хотя известны различные способы синтеза относительно коротких фрагментов в небольшом масштабе, эти техники часто характеризуются недостатками в отношении масштабируемости, автоматизации, скорости, точности и стоимости.

Включение посредством ссылки

[0004] 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 раскрытие настоящего изобретения

[0005] В настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов содержит по меньшей мере 5000 полинуклеотидов, причем каждый из по меньшей мере 5000 полинуклеотидов присутствует в таком количестве, что после гибридизации с геномными фрагментами и секвенирования гибридизованных геномных фрагментов библиотека полинуклеотидов обеспечивает по меньшей мере 30-кратную глубину прочтения по меньшей мере 90 процентов оснований геномных фрагментов при условиях для вплоть до 55-кратной теоретической глубины прочтения для оснований геномных фрагментов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов обеспечивает по меньшей мере 30-кратную глубину прочтения по меньшей мере 95 процентов оснований геномных фрагментов при условиях для вплоть до 55-кратной теоретической глубины прочтения для оснований геномных фрагментов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов обеспечивает по меньшей мере 30-кратную глубину прочтения по меньшей мере 98 процентов оснований геномных фрагментов при условиях для вплоть до 55-кратной теоретической глубины прочтения для оснований геномных фрагментов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов обеспечивает по меньшей мере 90 процентов уникальных прочтений для оснований геномных фрагментов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов обеспечивает по меньшей мере 95 процентов уникальных прочтений для оснований геномных фрагментов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов обеспечивает по меньшей мере 90 процентов оснований геномных фрагментов с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов обеспечивает по меньшей мере 95 процентов оснований геномных фрагментов с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов обеспечивает по меньшей мере 90 процентов геномных фрагментов с процентным отношением GC, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов, с глубиной прочтения в пределах приблизительно 1,5-кратной

по отношению к средней глубине прочтения. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов обеспечивает по меньшей мере приблизительно 80 процентов геномных фрагментов с процентным отношением повторяющейся последовательности или последовательности вторичной структуры, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов, с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем каждый из геномных фрагментов составляет приблизительно 100 оснований - приблизительно 500 оснований в длину. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, в которых по меньшей мере приблизительно 80 процентов по меньшей мере из 5000 полинуклеотидов представлены в количестве в пределах по меньшей мере приблизительно 1,5-кратном по отношению к среднему представлению для библиотеки полинуклеотидов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, в которых по меньшей мере 30 процентов по меньшей мере из 5000 полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, в которых по меньшей мере приблизительно 15 процентов по меньшей мере из 5000 полинуклеотидов содержат полинуклеотиды с процентным отношением повторяющейся последовательности или последовательности вторичной структуры, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, в которых по меньшей мере 5000 полинуклеотидов кодируют по меньшей мере 1000 генов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов содержит по меньшей мере 100000 полинуклеотидов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов содержит по меньшей мере 700000 полинуклеотидов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, в которых по меньшей мере 5000 полинуклеотидов содержат по меньшей мере одну последовательность экзона. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, в которых по меньшей мере 700000 полинуклеотидов содержат по меньшей мере один набор полинуклеотидов, в совокупности содержащих одну последовательность экзона. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, в

которых по меньшей мере 700000 полинуклеотидов содержат по меньшей мере 150000 наборов.

[0006] В настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов содержит по меньшей мере 5000 полинуклеотидов, причем каждый из полинуклеотидов составляет приблизительно 20 - 200 оснований в длину, причем множество полинуклеотидов кодируют последовательности из каждого экзона для по меньшей мере 1000 предварительно выбранных генов, причем каждый полинуклеотид содержит молекулярный маркер, причем каждый из по меньшей мере 5000 полинуклеотидов присутствует в таком количестве, что после гибридизации с геномными фрагментами и секвенирования гибридизованных геномных фрагментов библиотека полинуклеотидов обеспечивает по меньшей мере 30-кратную глубину прочтения по меньшей мере 90 процентов оснований геномных фрагментов при условиях для вплоть до 55-кратной теоретической глубины прочтения для оснований геномных фрагментов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов обеспечивает по меньшей мере 30-кратную глубину прочтения по меньшей мере 95 процентов оснований геномных фрагментов при условиях для вплоть до 55-кратной теоретической глубины прочтения для оснований геномных фрагментов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов обеспечивает по меньшей мере 30-кратную глубину прочтения по меньшей мере 98 процентов оснований геномных фрагментов при условиях для вплоть до 55-кратной теоретической глубины прочтения для оснований геномных фрагментов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов обеспечивает по меньшей мере 90 процентов уникальных прочтений для оснований геномных фрагментов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов обеспечивает по меньшей мере 95 процентов уникальных прочтений для оснований геномных фрагментов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов обеспечивает по меньшей мере 90 процентов оснований геномных фрагментов с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов обеспечивает по меньшей мере 95 процентов оснований геномных фрагментов с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов,

причем библиотека полинуклеотидов обеспечивает больше чем 90 процентов геномных фрагментов с процентным отношением GC, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов, с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов обеспечивает больше чем приблизительно 80 процентов геномных фрагментов с процентным отношением повторяющейся последовательности или последовательности вторичной структуры, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов, с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем каждый из геномных фрагментов составляет приблизительно 100 оснований - приблизительно 500 оснований в длину. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем больше чем приблизительно 80 процентов по меньшей мере из 5000 полинуклеотидов представлены в количестве в пределах по меньшей мере приблизительно 1,5-кратного по отношению к среднему представлению для библиотеки полинуклеотидов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем больше чем 30 процентов по меньшей мере из 5000 полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем больше чем приблизительно 15 процентов по меньшей мере из 5000 полинуклеотидов содержат полинуклеотиды с процентным отношением повторяющейся последовательности или последовательности вторичной структуры, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов содержит по меньшей мере 100000 полинуклеотидов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, причем библиотека полинуклеотидов содержит по меньшей мере 700000 полинуклеотидов. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, в которых по меньшей мере 700000 полинуклеотидов содержат по меньшей мере один набор полинуклеотидов, в совокупности содержащих одну последовательность экзона. Дополнительно в настоящем документе предусмотрены библиотеки полинуклеотидов, в которых по меньшей мере 700000 полинуклеотидов содержат по меньшей мере 150000 наборов.

[0007] В настоящем документе предусмотрены способы создания библиотеки полинуклеотидов, причем способ предусматривает следующее: получение предварительно определенных последовательностей, кодирующих по меньшей мере 5000 полинуклеотидов; синтезирование по меньшей мере 5000 полинуклеотидов; и амплификация по меньшей мере 5000 полинуклеотидов с помощью полимеразы с образованием библиотеки полинуклеотидов, причем больше чем приблизительно 80 процентов по меньшей мере из 5000 полинуклеотидов представлены в количестве в пределах по меньшей мере приблизительно 2-кратного по отношению к среднему представлению для библиотеки полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых больше чем приблизительно 80 процентов по меньшей мере из 5000 полинуклеотидов представлены в количестве в пределах по меньшей мере приблизительно 1,5-кратного по отношению к среднему представлению для библиотеки полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых больше чем 30 процентов по меньшей мере из 5000 полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов. Дополнительно в настоящем документе предусмотрены способы, при которых больше чем приблизительно 15 процентов по меньшей мере из 5000 полинуклеотидов содержат полинуклеотиды с процентным отношением повторяющейся последовательности или последовательности вторичной структуры, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека полинуклеотидов характеризуется суммарной частотой ошибок, составляющей меньше чем 1 на 800 оснований по сравнению с предварительно определенными последовательностями без исправления ошибок. Дополнительно в настоящем документе предусмотрены способы, при которых предварительно определенные последовательности кодируют по меньшей мере 700000 полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых синтез по меньшей мере 5000 полинуклеотидов происходит на структуре, характеризующейся поверхностью, причем поверхность содержит множество кластеров, причем каждый кластер содержит множество локусов; и причем каждый из по меньшей мере 5000 полинуклеотидов удлиняется из различного локуса множества локусов. Дополнительно в настоящем документе предусмотрены способы, при которых множество локусов содержит вплоть до 1000 локусов на кластер. Дополнительно в настоящем документе предусмотрены способы, при которых множество локусов содержит вплоть до 200 локусов на кластер.

[0008] В настоящем документе предусмотрены способы амплификации библиотеки полинуклеотидов, причем способ предусматривает следующее: получение распределения амплификации по меньшей мере для 5000 полинуклеотидов; кластеризация по меньшей мере 5000 полинуклеотидов из распределения амплификации на два или больше бинов на основании по меньшей мере одного признака последовательности, причем признак последовательности представляет собой процентное содержание GC, процентное содержание повторяющейся последовательности или процентное содержание вторичной структуры; коррекция относительной частоты полинуклеотидов по меньшей мере в одном бине для создания библиотеки полинуклеотидов, характеризующейся предварительно выбранным представлением; синтезирование библиотеки полинуклеотидов, характеризующейся предварительно выбранным представлением; и амплификация библиотеки полинуклеотидов, характеризующейся предварительно выбранным представлением. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере один признак последовательности представляет собой процентное содержание GC. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере один признак последовательности представляет собой процентное содержание вторичной структуры. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере один признак последовательности представляет собой процентное содержание повторяющейся последовательности. Дополнительно в настоящем документе предусмотрены способы, при которых содержание повторяющейся последовательности содержит последовательности с 3 или больше аденинами. Дополнительно в настоящем документе предусмотрены способы, при которых содержание повторяющейся последовательности содержит повторяющиеся последовательности по меньшей мере на одном конце полинуклеотида. Дополнительно в настоящем документе предусмотрены способы, при которых указанные полинуклеотиды подвергнуты кластеризации в бины на основании аффинности одной или нескольких полинуклеотидных последовательностей в отношении связывания с целевой последовательностью. Дополнительно в настоящем документе предусмотрены способы, при которых число последовательностей в нижних 30 процентах бинов характеризуется по меньшей мере на 50 процентов большим представлением в последующем применении после коррекции по сравнению с числом последовательностей в нижних 30 процентах бинов до коррекции. Дополнительно в настоящем документе предусмотрены способы, при которых число последовательностей в верхних 30 процентах бинов характеризуется по меньшей мере на 50 процентов большим представлением в последующем применении

после коррекции по сравнению с числом последовательностей в верхних 30 процентах бинов до коррекции.

[0009] В настоящем документе предусмотрены способы секвенирования геномной ДНК, предусматривающие следующее: приведение в контакт любой из библиотек полинуклеотидов, описанных в настоящем документе, с множеством геномных фрагментов; обогащение по меньшей мере одного геномного фрагмента, который связывается с библиотекой, для создания по меньшей мере одного обогащенного целевого полинуклеотида; и секвенирование по меньшей мере одного обогащенного целевого полинуклеотида. Дополнительно в настоящем документе предусмотрены способы, при которых множество обогащенных целевых полинуклеотидов содержит библиотеку кДНК. Дополнительно в настоящем документе предусмотрены способы, при которых длина по меньшей мере 5000 полинуклеотидов составляет приблизительно 80 - приблизительно 200 оснований. Дополнительно в настоящем документе предусмотрены способы, при которых каждый из геномных фрагментов составляет приблизительно 100 оснований - приблизительно 500 оснований в длину. Дополнительно в настоящем документе предусмотрены способы, при которых приведение в контакт происходит в растворе. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере 5000 полинуклеотидов являются по меньшей мере частично комплементарными геномным фрагментам. Дополнительно в настоящем документе предусмотрены способы, при которых выделение предусматривает следующее: (i) захват гибридизационных пар полинуклеотид/геномный фрагмент на твердой подложке; и (ii) высвобождение множества геномных фрагментов для создания обогащенных целевых полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых секвенирование дает в результате по меньшей мере 30-кратную глубину прочтения по меньшей мере 95 процентов оснований геномных фрагментов при условиях для вплоть до 55-кратной теоретической глубины прочтения для оснований геномных фрагментов. Дополнительно в настоящем документе предусмотрены способы, при которых секвенирование дает в результате по меньшей мере 30-кратную глубину прочтения по меньшей мере 98 процентов оснований геномных фрагментов при условиях для вплоть до 55-кратной теоретической глубины прочтения для оснований геномных фрагментов. Дополнительно в настоящем документе предусмотрены способы, при которых секвенирование дает в результате по меньшей мере 90 процентов уникальных прочтений для оснований геномных фрагментов. Дополнительно в настоящем документе предусмотрены способы, при которых секвенирование дает в результате по меньшей мере 95 процентов уникальных прочтений для оснований геномных фрагментов. Дополнительно

в настоящем документе предусмотрены способы, при которых секвенирование дает в результате по меньшей мере 90 процентов оснований геномных фрагментов с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения. Дополнительно в настоящем документе предусмотрены способы, при которых секвенирование дает в результате по меньшей мере 95 процентов оснований геномных фрагментов с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения. Дополнительно в настоящем документе предусмотрены способы, при которых секвенирование дает в результате по меньшей мере 90 процентов геномных фрагментов с процентным отношением GC, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов, с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения. Дополнительно в настоящем документе предусмотрены способы, при которых секвенирование дает в результате по меньшей мере приблизительно 80 процентов геномных фрагментов с процентным отношением повторяющейся последовательности или последовательности вторичной структуры, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов, с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере приблизительно 80 процентов по меньшей мере из 5000 полинуклеотидов представлены в количестве в пределах по меньшей мере приблизительно 1,5-кратного по отношению к среднему представлению для библиотеки полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере 30 процентов по меньшей мере из 5000 полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере 15 процентов по меньшей мере из 5000 полинуклеотидов содержат полинуклеотиды с процентным отношением повторяющейся последовательности или последовательности вторичной структуры, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере 5000 полинуклеотидов кодируют по меньшей мере 1000 генов. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека полинуклеотидов содержит по меньшей мере 100000 полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека полинуклеотидов содержит по меньшей мере 700000 полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере 5000 полинуклеотидов

содержат по меньшей мере одну последовательность экзона. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере 700000 полинуклеотидов содержат по меньшей мере один набор полинуклеотидов, в совокупности содержащих одну последовательность экзона. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере 700000 полинуклеотидов содержат по меньшей мере 150000 наборов.

Краткое описание графических материалов

[0010] На фигуре **1A** изображен схематический рабочий процесс, включая в себя применение первой библиотеки полинуклеотидов, измерение смещения из результата применения, конструирование и синтезирование второй библиотеки полинуклеотидов с контролируемой стехиометрией и применение второй библиотеки полинуклеотидов для получения требуемого результата представления.

[0011] На фигуре **1B** изображена схема обогащения целевых полинуклеотидов с помощью библиотеки связывающихся с мишенью полинуклеотидов.

[0012] На фигуре **1C** изображен иллюстративный рабочий процесс для обогащения и секвенирования образца нуклеиновой кислоты.

[0013] На фигуре **2** изображена схема создания библиотек полинуклеотидов из кластерной амплификации.

[0014] На фигуре **3A** изображена пара полинуклеотидов для нацеленного воздействия и обогащения. Полинуклеотиды содержат комплементарные связывающиеся с мишенью (вставочные) последовательности, а также сайты связывания праймера.

[0015] На фигуре **3B** изображена пара полинуклеотидов для нацеленного воздействия и обогащения. Полинуклеотиды содержат комплементарные связывающиеся с целевой последовательностью (вставочные) последовательности, сайты связывания праймера и нецелевые последовательности.

[0016] На фигуре **4A** изображена конфигурация связывания полинуклеотида с целевой последовательностью большего полинуклеотида. Целевая последовательность короче, чем область связывания полинуклеотида, и область связывания полинуклеотида (или вставочная последовательность) смещена относительно целевой последовательности, и также связывается с частью смежной последовательности.

[0017] На фигуре **4B** изображена конфигурация связывания полинуклеотида с целевой последовательностью большего полинуклеотида. Длина целевой последовательности меньше или равна области связывания полинуклеотида, и область

связывания полинуклеотида находится в центре целевой последовательности, и также связывается с частью смежной последовательности.

[0018] На фигуре **4C** изображена конфигурация связывания полинуклеотида с целевой последовательностью большего полинуклеотида. Целевая последовательность немного длиннее, чем область связывания полинуклеотида, и область связывания полинуклеотида находится в центре целевой последовательности с буферной областью с каждой стороны.

[0019] На фигуре **4D** изображена конфигурация связывания полинуклеотида с целевой последовательностью большего полинуклеотида. Целевая последовательность длиннее, чем область связывания полинуклеотида, и области связывания двух полинуклеотидов перекрываются, чтобы охватить целевую последовательность.

[0020] На фигуре **4E** изображена конфигурация связывания полинуклеотида с целевой последовательностью большего полинуклеотида. Целевая последовательность длиннее, чем область связывания полинуклеотида, и области связывания двух полинуклеотидов перекрываются, чтобы охватить целевую последовательность.

[0021] На фигуре **4F** изображена конфигурация связывания полинуклеотида с целевой последовательностью большего полинуклеотида. Целевая последовательность длиннее, чем область связывания полинуклеотида, и области связывания двух полинуклеотидов не перекрываются, чтобы охватить целевую последовательность, оставляя пропуск **405**.

[0022] На фигуре **4G** изображена конфигурация связывания полинуклеотида с целевой последовательностью большего полинуклеотида. Целевая последовательность длиннее, чем область связывания полинуклеотида, и области связывания трех полинуклеотидов перекрываются, чтобы охватить целевую последовательность.

[0023] На фигуре **5** представлена диаграмма стадий, на которой продемонстрирован иллюстративный рабочий процесс синтеза генов согласно раскрытию, представленному в настоящем документе.

[0024] На фигуре **6** показана вычислительная система.

[0025] На фигуре **7** показана блок-диаграмма, на которой проиллюстрирована архитектура вычислительной системы.

[0026] На фигуре **8** показана диаграмма, 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).

[0027] На фигуре 9 показана блок-диаграмма многопроцессорной вычислительной системы с использованием совместного пространства памяти виртуальных адресов.

[0028] На фигуре 10 показано изображение планшета с 256 кластерами, причем каждый кластер содержит 121 локус с полинуклеотидами, удлиняющимися из них.

[0029] На фигуре 11А показан график представления полинуклеотидов (частота полинуклеотида в зависимости от представленности, виде измеренного поглощения) по всему планшету из синтеза 29040 уникальных полинуклеотидов из 240 кластеров, причем каждый кластер содержит 121 полинуклеотид.

[0030] На фигуре 11В показан график измерения частоты полинуклеотида в зависимости от поглощения представленности (в виде измеренного поглощения) по каждому отдельному кластеру, с контрольными кластерами, указанными в рамке.

[0031] На фигуре 12 показан график измерений частоты полинуклеотида в зависимости от представленности (в виде измеренного поглощения) по четырем отдельным кластерам.

[0032] На фигуре 13А показан график частоты в зависимости от частоты ошибок по всему планшету из синтеза 29040 уникальных полинуклеотидов из 240 кластеров, причем каждый кластер содержит 121 полинуклеотид.

[0033] На фигуре 13В показан график измерения частоты ошибок полинуклеотида в зависимости от частоты по каждому отдельному кластеру, с контрольными кластерами, указанными в рамке.

[0034] На фигуре 14 показан график измерений частоты полинуклеотида в зависимости от частоты ошибок по четырем кластерам.

[0035] На фигуре 15 показан график содержания GC как меры количества полинуклеотидов в зависимости от процентного отношения на каждый полинуклеотид.

[0036] На фигуре 16 показаны графики с результатами ПЦР с двумя различными полимеразми. На каждой диаграмме изображено число полинуклеотидов (0 - 2000) в зависимости от наблюдаемой частоты ("0 - 35", измеренной в импульсах на 100000).

[0037] На фигуре 17 представлена диаграмма с количественным определением однородности популяции полинуклеотидов после амплификации, которая была зарегистрирована.

[0038] На фигуре 18 показан график, на котором показано влияние избыточной амплификации на выпадения последовательностей.

[0039] На фигуре **19** изображены графики процентного содержания GC на частоту полинуклеотида (на 100000 прочтений) в объединенных неамплифицированных и амплифицированных популяциях полинуклеотидов.

[0040] На фигуре **20** показан график процентного содержания GC на частоту полинуклеотида (на 100000 прочтений) для двух отдельных прогонов после амплификации кластеров.

[0041] На фигуре **21A** показан график процентного содержания GC на частоту полинуклеотида для сбалансированной по содержанию GC библиотеки полинуклеотидов.

[0042] На фигуре **21B** показан график процентного содержания GC на частоту полинуклеотида для библиотеки полинуклеотидов с сильным смещением в сторону высокого и низкого содержания GC.

[0043] На фигуре **21C** показан график процентного содержания GC на частоту полинуклеотида для библиотеки полинуклеотидов с умеренным смещением в сторону высокого и низкого содержания GC.

[0044] На фигуре **21D** показан график процентного содержания GC на частоту полинуклеотида для библиотеки полинуклеотидов со смещением в сторону низкого содержания GC.

[0045] На фигуре **21E** показан график процентного содержания GC на частоту полинуклеотида для библиотеки полинуклеотидов со смещением в сторону высокого содержания GC.

[0046] На фигуре **22** показан график процентного содержания GC на частоту полинуклеотида для теоретической библиотеки из 13000 полинуклеотидов с последовательностями, характеризующимися 15% - 85% содержанием GC.

[0047] На фигуре **23** показан график числа полинуклеотидов в зависимости от частоты полинуклеотидов (на 100000 прочтений) для сбалансированной по содержанию GC библиотеки полинуклеотидов.

[0048] На фигуре **24A** изображен график, на котором показано количество сэмпинга, необходимого для получения 80% покрытия секвенирования для сбалансированной по содержанию GC библиотеки полинуклеотидов, по сравнению с теоретическим максимумом монодисперсной библиотеки.

[0049] На фигуре **24B** изображен график, на котором показано количество сэмпинга, необходимого для получения 90% покрытия секвенирования для сбалансированной по содержанию GC библиотеки полинуклеотидов по сравнению с теоретическим максимумом монодисперсной библиотеки.

[0050] На фигуре **25** показан график количества полинуклеотидов в зависимости от частоты полинуклеотидов (импульсы на 1000000 прочтений) для библиотеки, содержащей полинуклеотиды длиной 80 нуклеотидов.

[0051] На фигуре **26** показан график количества полинуклеотидов в зависимости от частоты полинуклеотидов (импульсы на 1000000 прочтений) для библиотеки, содержащей полинуклеотиды длиной 120 нуклеотидов.

[0052] На фигуре **27** изображены графики, на которых показана средняя частота полинуклеотидов (на 1000000 прочтений) для сбалансированных по содержанию GC библиотек, содержащих полинуклеотиды длиной как 80, так и 120 нуклеотидов.

[0053] На фигуре **28** изображены графики, на которых показан эффект числа циклов ПЦР-амплификации, содержания GC и выбора ДНК-полимеразы на представление полинуклеотидных последовательностей.

[0054] На фигуре **29** показан график выпадений последовательностей как функции от циклов амплификации для двух различных высокоточных полимераз.

[0055] На фигуре **30** изображены графики, на которых показан эффект различных ДНК-полимераз на представление последовательностей. Одну и ту же библиотеку полинуклеотидов амплифицировали в течение 15 циклов либо с помощью ДНК-полимеразы 1, либо с помощью ДНК-полимеразы 2.

[0056] На фигуре **31A** изображено количество избыточного секвенирования, необходимого для достижения заданной глубины прочтения целевой последовательности с использованием библиотеки экзомных зондов без контролируемой стехиометрии.

[0057] На фигуре **31B** изображено снижение избыточного секвенирования, необходимого для достижения заданной глубины прочтения для целевой последовательности с использованием библиотеки экзомных зондов с контролируемой стехиометрией по сравнению с библиотекой экзомных зондов без контролируемой стехиометрии.

[0058] На фигуре **32A** показан график процентного отношения оснований, обладающих 1x, 20x или 30x глубиной прочтения секвенирования (X покрытия) как для набора сравнения А экзомных зондов, так и библиотеки зондов 1 с контролируемой стехиометрией.

[0059] На фигуре **32B** показан график процентного отношения оснований, обладающих 1X или 10X глубиной прочтения секвенирования (покрытие X), нормализованной при 4,5 млрд.п.н. секвенирования для панели наборов сравнения экзомных зондов и библиотеки зондов 1 с контролируемой стехиометрией.

[0060] На фигуре **33** изображен синтез библиотеки полинуклеотидных зондов различных масштабов как функции от числа полинуклеотидов в библиотеке.

[0061] На фигуре **34** изображено сравнение между покрытием (число оснований) как функцией от глубины прочтения библиотеки сравнения зондов на основе матрицы по сравнению с библиотекой зондов 2 с контролируемой стехиометрией.

[0062] На фигуре **35A** изображено сравнение между покрытием как функцией от глубины прочтения библиотеки сравнения зондов на основе матрицы по сравнению с библиотекой зондов 2 с контролируемой стехиометрией для мишеней с содержанием GC, составляющим 10-30% и 30-50%.

[0063] На фигуре **35B** изображено сравнение между покрытием как функцией от глубины прочтения библиотеки сравнения зондов на основе матрицы по сравнению с библиотекой зондов 2 с контролируемой стехиометрией для мишеней с содержанием GC, составляющим больше чем 50-70% и больше чем 70-90%.

[0064] На фигуре **36A** изображено сравнение между процентным отношением (%) частоты мишеней библиотеки сравнения зондов на основе матрицы в зависимости от 0,1-, 1- и 3-кратных концентраций библиотеки зондов 3 с контролируемой стехиометрией.

[0065] На фигуре **36B** изображено сравнение между глубиной прочтения библиотеки сравнения зондов на основе матрицы в зависимости от 0,1-, 1- и 3-кратных концентраций библиотеки зондов 3 с контролируемой стехиометрией.

[0066] На фигуре **37** изображено сравнение между процентными отношениями уникальных прочтений и целевых оснований при 1X, 20X и 30X глубине прочтения набора сравнения экзона в сравнении с библиотекой зондов 4 с контролируемой стехиометрией.

[0067] На фигуре **38** изображено сравнение между процентными отношениями уникальных прочтений и целевых оснований при 1X, 20X и 30X глубине прочтения набора сравнения экзона в сравнении с библиотекой зондов 4 с контролируемой стехиометрией.

Подробное раскрытие настоящего изобретения

[0068] 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.

[0069] В настоящем документе предусмотрены способы конструирования, синтеза и контроля стехиометрии больших библиотек полинуклеотидов. Когда первая популяция полинуклеотидов подвергается предварительной стадии применения, например, для амплификации, в качестве зондов захвата для обогащения и синтеза генов, последующие реакции амплификации популяции полинуклеотидов могут приводить к смещенному репрезентативному выходу из-за дисперсии в полинуклеотидной последовательности, в результате чего определенные полинуклеотиды представлены более широко, чем другие.

Фиг. 1А. Полученное в результате смещение, наблюдаемое из выхода этого предварительного применения, измеряют и используют для контроля первой популяции полинуклеотидов с предварительно выбранной стехиометрией, например, относительной частотой полинуклеотидов в популяции с учетом любого числа признаков последовательности, таких как содержание GC, повторяющиеся последовательности, хвостовые аденины, вторичная структура, аффинность по отношению к связыванию целевой последовательности или модифицированные нуклеотиды. После модификации стехиометрии полинуклеотидов вторую популяцию полинуклеотидов разрабатывают и синтезируют с предварительно выбранной стехиометрией, чтобы исправить нежелательные эффекты смещения, связанные со стадией применения. В некоторых случаях подвергание второй популяции контролируемой стехиометрии полинуклеотидов стадии применения, такой как ПЦР-амплификация, приводит к сбалансированному выходу, такому как популяция амплифицированных полинуклеотидов с очень однородным представлением, или неоднородное представление с предварительно выбранным сдвигом в представлении. См. **фиг. 1А**, нижние графики. В некоторых случаях способы, описанные в настоящем документе, предусматривают контроль над представлением последовательности полинуклеотидных зондов так, что полинуклеотидная популяция обеспечивает очень однородную частоту захвата целевой последовательности (**фиг. 1В**). Например, образец полинуклеотидов **100** содержит целевые полинуклеотиды **101**. Контакт образца **100** со связывающимися с мишенью полинуклеотидами **103** при соответствующих условиях **102** приводит к образованию гибридизационных пар **104**, которые отделены от нецелевых полинуклеотидов в образце **100**. Денатурация и разделение пар **104** высвобождает обогащенные целевые полинуклеотиды **107** для последующих применений, таких как секвенирование. Кроме того, в настоящем документе предусмотрены синтезированные *de novo* полинуклеотиды для применения в гибридизации с геномной ДНК, например, в контексте процесса секвенирования. На первой стадии иллюстративного рабочего процесса секвенирования (**фиг. 1С**) образец нуклеиновой кислоты **108**, содержащий целевые полинуклеотиды, фрагментируют путем механического или ферментативного сдвига с

образованием библиотеки фрагментов **109**. Адаптеры **115**, необязательно содержащие последовательности праймеров и/или штрих-коды, лигируют для образования библиотеки, маркированной адаптером **110**. Затем эту библиотеку необязательно амплифицируют и проводят гибридизацию со связывающимися с мишенью полинуклеотидами **117**, которые гибридизуются с целевыми полинуклеотидами, вместе с блокирующими полинуклеотидами **116**, которые предотвращают гибридизацию между связывающимися с мишенью полинуклеотидами **117** и адаптерами **115**. Захват гибридизационных пар, представляющих собой целевой полинуклеотид-связывающийся с мишенью полинуклеотид **112**, и связывающихся с мишенью полинуклеотидов **117** позволяет выделять/обогащать целевые полинуклеотиды **113**, которые затем необязательно амплифицируют и секвенируют **114**.

[0070] Определения

[0071] 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, исключаящие один или оба из этих включенных пределов, также включены в настоящее изобретение, если контекст явно не предписывает иное.

[0072] 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.

[0073] 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.

[0074] 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.

[0075] Термин “нуклеиновая кислота” охватывает двух- или трехцепочечные нуклеиновые кислоты, а также одноцепочечные молекулы. В двухцепочечных или трехцепочечных нуклеиновых кислотах цепи нуклеиновых кислот не обязательно должны быть параллельными (т.е. двухцепочечная нуклеиновая кислота не обязательно должна быть двухцепочечной по всей длине обеих цепей). Последовательности нуклеиновых кислот, если они приведены, перечислены в направлении от 5' до 3', если не указано иное. Способы, описанные в настоящем документе, обеспечивают получение выделенных нуклеиновых кислот. Способы, описанные в настоящем документе, дополнительно обеспечивают получение выделенных и очищенных нуклеиновых кислот. Длина полинуклеотидов, если она указана, описана как число оснований и сокращенно, например, н. (nt, нуклеотидов), п.н. (bp, пар нуклеотидов), т.п.н. (kb, тысяч пар нуклеотидов) или млрд.п.н. (Gb, миллиардов пар нуклеотидов).

[0076] 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. Описанная в настоящем документе библиотека полинуклеотидов может содержать по меньшей мере 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10000, 20000, 30000, 50000, 100000, 200000, 500000, 1000000 или больше чем 1000000 полинуклеотидов. Описанная в настоящем документе библиотека полинуклеотидов может содержать не больше чем 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10000, 20000, 30000, 50000, 100000, 200000, 500000 или не больше чем 1000000 полинуклеотидов. Описанная в настоящем документе библиотека полинуклеотидов может содержать 10 - 500, 20 - 1000, 50 - 2000, 100 - 5000, 500 - 10000, 1000 - 50000, 10000 - 50000, 100000 - 500000 или 50000 - 1000000 полинуклеотидов. Описанная в настоящем документе библиотека полинуклеотидов может содержать приблизительно 370000; 400000; 500000 или больше различных полинуклеотидов.

[0077] 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 конструкции могут быть ограничены независимо выбранным верхним пределом, составляющим приблизительно 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 может в первую очередь не содержать интронную последовательности.

[0078] De novo синтез малых популяций полинуклеотидов для реакций амплификации

[0079] В настоящем документе описаны способы синтеза полинуклеотидов с поверхности, например, планшета. В некоторых случаях полинуклеотиды синтезируют на кластере локусов для удлинения полинуклеотида, высвобождают и затем впоследствии

подвергают реакции амплификации, например, ПЦР. Иллюстративный рабочий процесс синтеза полинуклеотидов из кластера изображен на фиг. 2. Кремниевый планшет **201** включает в себя множественные кластеры **203**. В пределах каждого кластера находятся множественные локусы **221**. Полинуклеотиды синтезируют **207 de novo** на планшете **201** из кластера **203**. Полинуклеотиды отщепляют **211** и удаляют **213** из планшета с образованием популяции высвобожденных полинуклеотидов **215**. Популяцию высвобожденных полинуклеотидов **215** затем амплифицируют **217** с образованием библиотеки амплифицированных полинуклеотидов **219**.

[0080] В настоящем документе предусмотрены способы, при которых амплификация полинуклеотидов, синтезированных на кластере, обеспечивает усиленный контроль над представлением полинуклеотидов по сравнению с амплификацией полинуклеотидов по всей поверхности структуры без такого кластерного расположения. В некоторых случаях амплификация полинуклеотидов, синтезированных с поверхности, характеризующейся кластерным расположением локусов для удлинения полинуклеотидов, обеспечивает преодоление отрицательных эффектов на представление за счет повторного синтеза больших популяций полинуклеотидов. Иллюстративные отрицательные эффекты на представление за счет повторного синтеза больших популяций полинуклеотидов включают в себя без ограничения смещение за счет амплификации, являющееся результатом высокого/низкого содержания GC, повторяющихся последовательностей, хвостовых аденинов, вторичной структуры, аффинности по отношению к связыванию целевой последовательности или модифицированных нуклеотидов в полинуклеотидной последовательности.

[0081] Кластерная амплификация в противоположность амплификации полинуклеотидов по всему планшету без кластерного расположения может приводить к более плотному распределению вокруг среднего. Например, если 100000 прочтений выбирают в качестве образца случайным образом, среднее от 8 прочтений на последовательность будет давать на выходе библиотеку с распределением, приблизительно в 1,5 раза (1,5X) выше среднего. В некоторых случаях одна кластерная амплификация дает в результате не более чем приблизительно 1,5X, 1,6X, 1,7X, 1,8X, 1,9X или 2,0X от среднего. В некоторых случаях одна кластерная амплификация дает в результате по меньшей мере приблизительно 1,0X, 1,2X, 1,3X, 1,5X, 1,6X, 1,7X, 1,8X, 1,9X, или 2,0X от среднего.

[0082] Описанные в настоящем документе способы кластерной амплификации по сравнению с амплификацией по всему планшету могут давать в результате библиотеку полинуклеотидов, которая требует меньшего секвенирования для эквивалентного представления последовательности. В некоторых случаях требуется по меньшей мере на

10%, по меньшей мере на 20%, по меньшей мере на 30%, по меньшей мере на 40%, по меньшей мере на 50%, по меньшей мере на 60%, по меньшей мере на 70%, по меньшей мере на 80%, по меньшей мере на 90% или по меньшей мере на 95% меньше секвенирования. В некоторых случаях требуется вплоть до на 10%, вплоть до на 20%, вплоть до на 30%, вплоть до на 40%, вплоть до на 50%, вплоть до на 60%, вплоть до на 70%, вплоть до на 80%, вплоть до на 90% или вплоть до на 95% меньше секвенирования. Иногда требуется на 30% меньше секвенирования после кластерной амплификации по сравнению с амплификацией по всему планшету. Секвенирование полинуклеотидов в некоторых случаях подтверждают с помощью высокопроизводительного секвенирования, такого как секвенирование нового поколения. Секвенирование библиотеки секвенирования можно проводить с помощью любой соответствующей технологии секвенирования, включая в себя без ограничения одномолекулярное секвенирование в режиме реального времени (SMRT), полонисеквенирование, секвенирование путем лигирования, секвенирование с обратимым терминатором, секвенирование с обнаружением протонов, ионное полупроводниковое секвенирование, секвенирование через нанопоры, электронное секвенирование, пиросеквенирование, секвенирование по Максаму-Гилберту, секвенирование методом «обрыва цепи» (например, секвенирование Сэнгера), +S секвенирование или секвенирование путем синтеза. Число раз, сколько идентифицируется один нуклеотид или полинуклеотид, или “прочтение”, определяют как глубину секвенирования или глубину прочтения. В некоторых случаях глубина прочтения называется кратностью покрытия, например, 55-кратное (или 55X) покрытие, необязательно описывающее процентное отношение оснований.

[0083] В некоторых случаях амплификация из кластерного расположения по сравнению с амплификацией по всему планшету приводит к меньшему количеству выпадений, или последовательностей, которые не обнаруживаются после секвенирования продукта амплификации. Выпадения могут представлять собой АТ и/или GC. В некоторых случаях число выпадений составляет не более чем приблизительно 1%, 2%, 3%, 4% или 5% популяции полинуклеотидов. В некоторых случаях число выпадений равно нулю.

[0084] Описанный в настоящем документе кластер содержит набор дискретных, неперекрывающихся локусов для синтеза полинуклеотидов. Кластер может содержать приблизительно 50-1000, 75-900, 100-800, 125-700, 150-600, 200-500 или 300-400 локусов. В некоторых случаях каждый кластер включает в себя 121 локус. В некоторых случаях каждый кластер включает в себя приблизительно 50-500, 50-200, 100-150 локусов. В некоторых случаях каждый кластер включает в себя по меньшей мере приблизительно 50, 100, 150, 200, 500, 1000 или больше локусов. В некоторых случаях один планшет включает

в себя 100, 500, 10000, 20000, 30000, 50000, 100000, 500000, 700000, 1000000 или больше локусов. Локус может представлять собой пятно, лунку, микролунку, канал или стержень. В некоторых случаях каждый кластер характеризуется по меньшей мере 1-, 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-кратной или большей избыточностью отдельных признаков, поддерживающих удлинение полинуклеотидов, характеризующихся идентичной последовательностью.

[0085] Конструкция библиотек полинуклеотидов, характеризующихся контролируемой стехиометрией

[0086] В настоящем документе предусмотрены способы конструирования и синтеза библиотеки полинуклеотидов, при которых количество (или стехиометрия) каждого вида полинуклеотидов (т.е. характеризующегося последовательностью, отличающейся от другого полинуклеотида в библиотеке) доводят до предварительно определенного количества так, чтобы требуемый результат контролировался в последующем применении. В связи с этим, в настоящем документе предусмотрены способы контролируемой и предварительно определенной модификации стехиометрии видов полинуклеотидов. Например, распределение видов полинуклеотидов после реакции амплификации можно контролировать с использованием описанных в настоящем документе способов. Распределение видов полинуклеотидов предварительно выбирают для обеспечения очень равномерного захвата целевых последовательностей, например, с использованием панели полинуклеотидов для анализов на основе гибридизации, например, для анализа секвенирования. Более того, описанные в настоящем документе способы обеспечивают конструирование полинуклеотидной библиотеки последовательностей с одним или несколькими признаками последовательности, которые, как правило, будут приводить к неоднородным продуктам амплификации или продуктам захвата за счет структурных признаков определенных “проблемных” полинуклеотидных последовательностей, причем “проблемные” полинуклеотидные последовательности содержат одно или несколько свойств, ассоциированных с созданием смещения при применении библиотеки полинуклеотидов. Иллюстративные свойства “проблемных” полинуклеотидных последовательностей для контроля стехиометрии с использованием способов, описанных в настоящем документе, включают в себя без ограничения следующее: высокое или низкое содержание GC или AT, повторяющиеся последовательности, хвостовые аденины, вторичная структура, аффинность по отношению к связыванию целевой последовательности (для амплификации, обогащения или обнаружения), стабильность, температура плавления, биологическая активность, способность к сборке в фрагменты большего размера, последовательности, содержащие модифицированные нуклеотиды или аналоги нуклеотидов или любое другое свойство полинуклеотидов для создания второй

полинуклеотидной библиотеки последовательностей на основании спрогнозированных или эмпирических данных. В некоторых случаях библиотеку последовательностей получают для контроля стехиометрии, и организуют или осуществляют ее кластеризацию (бинирование) в две или больше предварительно установленные группы (бины) на основании одного или нескольких признаков последовательности. В некоторых случаях два или больше бинов представляют индивидуальные уникальные последовательности. В некоторых случаях бины представляют диапазоны значений на основании одного или нескольких установленных признаков последовательности, каждое из которых содержит по меньшей мере 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% или по меньшей мере 99% общих последовательностей. В некоторых случаях бины представляют диапазоны значений, каждое из которых содержит не более чем 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% или меньше чем 100% общих последовательностей. В одном примере бины можно установить по процентному отношению содержания GC, при этом множественные бины представляют диапазон 25-75% с инкрементом, составляющим 5% (например, 25-29%, 30-34%, 35-39% и т.д.), один бин представляет меньше чем 25%, и один бин представляет больше чем 75% содержание GC. Для каждого бина присваивается значение представленности, представляющее стехиометрию молекул для всех последовательностей в каждом бине. В некоторых случаях значение представленности изначально устанавливают на 100, что приводит к равному представлению последовательности на один бин. В некоторых случаях контроль стехиометрии осуществляют с использованием полученных данных смещения за счет применения для увеличения, уменьшения или поддержания значения представленности для каждого бина. Также используют другие способы коррекции представленности последовательности в соответствии с настоящим описанием изобретения. В некоторых случаях ранее полученное распределение используют для определения начальных значений представленности.

[0087] В некоторых случаях данные смещения за счет применения получают с помощью алгоритмов прогнозирования. Данные смещения за счет применения можно получить эмпирически или можно получить из библиотеки с неконтролируемой или ранее контролируемой стехиометрией. Например, данные смещения за счет применения получают из амплификации библиотеки полинуклеотидов; частоту полинуклеотидов на бин после амплификации наносят на график в зависимости от бинов процентного отношения GC для установления смещения за счет амплификации как функции от процентного отношения содержания GC. В другом примере данные смещения за счет применения получают из данных секвенирования нового поколения (NGS) после обогащения целевых последовательностей с помощью библиотеки полинуклеотидных

зондов; число прочтений на целевой ген используют для сортировки последовательностей зондов в бины; число прочтений на целевой ген наносят на график в зависимости от бинов прочтений NGS для установления смещения секвенирования NGS как функции от последовательности полинуклеотидного зонда. В другом примере данные смещения за счет применения можно получить из результата клеточного анализа, такого как флуоресценция, после обработки клеток вектором, содержащим библиотеку полинуклеотидов; число прочтений на последовательность, идентифицированное в флуоресцентных клетках, используют для сортировки последовательностей зонда в бины; число прочтений на последовательность наносят на график в зависимости от числа бинов прочтений для установления смещения как функции от последовательности полинуклеотидного зонда.

[0088] После контроля стехиометрии модифицированную библиотеку последовательностей синтезируют для создания библиотеки полинуклеотидов с контролируемой стехиометрией. В некоторых случаях библиотеку с контролируемой стехиометрией используют для последующего применения. В некоторых случаях данные из последующего применения с помощью библиотеки полинуклеотидов с контролируемой стехиометрией используют для проведения дополнительных раундов стехиометрической модификации библиотеки.

[0089] Создание библиотек полинуклеотидов с контролируемой стехиометрией содержания GC

[0090] В настоящем документе предусмотрены способы синтеза библиотек полинуклеотидов с установленным свойством, таким как содержание GC, повторяющиеся последовательности, хвостовые аденины, вторичная структура, аффинность по отношению к связыванию целевой последовательности или модифицированные нуклеотиды, для создания второй популяции полинуклеотидов на основании спрогнозированных или эмпирических данных. Например, когда библиотека полинуклеотидов выбрана для синтеза, чтобы получить в результате установленное содержание GC после амплификации, коррекция представления видов для полинуклеотидов в библиотеке на стадии синтеза в зависимости от содержания GC дает в результате улучшенное представление полинуклеотидов после амплификации. Содержание GC в библиотеке полинуклеотидов может составлять по меньшей мере 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% или больше чем 95%. В некоторых случаях содержание GC в библиотеке полинуклеотидов составляет не больше чем 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% или меньше чем 100%. В некоторых случаях содержание GC находится в диапазоне, составляющем приблизительно 5-95%, 10-90%, 30-80%, 40-75% или 50-70%.

[0091] Описанные в настоящем документе библиотеки полинуклеотидов можно скорректировать по их содержанию GC. В некоторых случаях библиотеки полинуклеотидов предпочтительно характеризуются высоким содержанием GC. Например, сконструирована библиотека, в которой увеличенная частота полинуклеотидов характеризуется содержанием GC в диапазоне, составляющем приблизительно 40% - приблизительно 90%. В некоторых случаях библиотеки полинуклеотидов содержат низкое содержание GC. Например, сконструирована библиотека, в которой увеличенная частота полинуклеотидов характеризуется содержанием GC, находящимся в диапазоне, составляющем приблизительно 10% - приблизительно 60%. Можно сконструировать библиотеку, предпочтительно характеризующуюся высоким и низким содержанием GC. Например, можно сконструировать библиотеку, в которой увеличенная частота полинуклеотидов характеризуется содержанием GC, главным образом, в диапазоне, составляющем приблизительно 10% - приблизительно 30%, и в диапазоне, составляющем приблизительно 70% - приблизительно 90%. В некоторых случаях библиотека предпочтительно характеризуется однородным содержанием GC. Например, частота полинуклеотидов является однородной с содержанием GC в диапазоне, составляющем приблизительно 10% - приблизительно 90%. В некоторых случаях библиотека содержит полинуклеотиды с процентным отношением GC, составляющим приблизительно 10% - приблизительно 95%. В некоторых случаях описанная в настоящем документе библиотека содержит полинуклеотиды, характеризующиеся больше чем 30% различных полинуклеотидов с процентным отношением GC, составляющим от 10% до 30% или от 70 до 90%. В некоторых случаях описанная в настоящем документе библиотека содержит полинуклеотиды, характеризующиеся меньше чем приблизительно 15% полинуклеотидов с процентным отношением GC, составляющим от 10% до 30% или от 60 до 90%.

[0092] Создание библиотек полинуклеотидов с заданным содержанием GC в некоторых случаях происходит путем комбинации по меньшей мере 2 библиотек полинуклеотидов с различным содержанием GC. В некоторых случаях по меньшей мере 2, 3, 4, 5, 6, 7, 10 или больше чем 10 библиотек полинуклеотидов комбинируют для создания популяции полинуклеотидов с заданным содержанием GC. В некоторых случаях не больше чем 2, 3, 4, 5, 6, 7 или 10 библиотек полинуклеотидов комбинируют для создания популяции неидентичных полинуклеотидов с заданным содержанием GC.

[0093] В некоторых случаях содержание GC корректируют путем синтеза меньшего или большего количества полинуклеотидов на кластер. Например, по меньшей мере 25, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 или больше чем 1000 неидентичных полинуклеотидов синтезируют на одном кластере. В некоторых случаях не больше чем

приблизительно 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 неидентичных полинуклеотидов синтезируют на одном кластере. В некоторых случаях 50 - 500 неидентичных полинуклеотидов синтезируют на одном кластере. В некоторых случаях 100 - 200 неидентичных полинуклеотидов синтезируют на одном кластере. В некоторых случаях приблизительно 100, приблизительно 120, приблизительно 125, приблизительно 130, приблизительно 150, приблизительно 175 или приблизительно 200 неидентичных полинуклеотидов синтезируют на одном кластере.

[0094] В некоторых случаях содержание GC корректируют путем синтеза неидентичных полинуклеотидов различной длины. Например, длина каждого из синтезированных неидентичных полинуклеотидов может составлять по меньшей мере или приблизительно по меньшей мере 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 300, 400, 500, 2000 нуклеотидов или больше. Длина синтезированных неидентичных полинуклеотидов может составлять не более чем или приблизительно не более чем 2000, 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 синтезированных неидентичных полинуклеотидов может попадать в диапазон 10-2000, 10-500, 9-400, 11-300, 12-200, 13-150, 14-100, 15-50, 16-45, 17-40, 18-35 и 19-25.

[0095] Создание библиотек полинуклеотидов с контролируемой стехиометрией содержания повторяющейся последовательности

[0096] Описанную в настоящем документе библиотеку полинуклеотидов можно синтезировать с заданным распределением повторяющейся последовательности. В некоторых случаях коррекция библиотеки полинуклеотидов в отношении содержания повторяющейся последовательности дает в результате улучшенное представление полинуклеотидов.

[0097] Повторяющаяся последовательность может представлять собой повтор одного нуклеотида или повтор блока из двух или больше нуклеотидов. В некоторых случаях повторяющаяся последовательность составляет по меньшей мере 2, 3, 4, 5, 6, 7, 8, 9 или по меньшей мере 10 нуклеотидов. В некоторых случаях повторяющаяся последовательность составляет не более чем 2, 3, 4, 5, 6, 7, 8, 9 или не более чем 10 нуклеотидов. В некоторых случаях блок нуклеотидов содержит по меньшей мере 2, 3, 4, 5, 10, 15, 25, 50, 100, 200, 500 или по меньшей мере 1000 нуклеотидов. В некоторых случаях блок нуклеотидов содержит не более чем 2, 3, 4, 5, 10, 15, 25, 50, 100, 200, 500 или не более чем 1000 нуклеотидов. Повторяющаяся последовательность может быть расположена во внутреннем или концевом расположении синтезированного полинуклеотида большего размера. Концевое расположение может находиться вблизи 5', 3' или как 5', так и 3' концов полинуклеотида.

В некоторых случаях повторяющаяся последовательность расположена в пределах по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9 или по меньшей мере 10 нуклеотидов конца. В некоторых случаях повторяющаяся последовательность расположена в пределах не более чем 1, 2, 3, 4, 5, 6, 7, 8, 9 или не более чем 10 нуклеотидов конца. В некоторых случаях повторяющийся нуклеотид представляет собой аденин. В некоторых случаях повторяющаяся последовательность расположена на конце полинуклеотида, например, полиадениновый хвост.

[0098] Содержание повторяющейся последовательности в библиотеке полинуклеотидов может составлять по меньшей мере 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% или больше чем 95%. В некоторых случаях содержание повторяющейся последовательности в библиотеке полинуклеотидов составляет не более чем 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% или не больше чем 100%. В некоторых случаях содержание повторяющейся последовательности находится в диапазоне, составляющем приблизительно 5-95%, 10-90%, 30-80%, 40-75% или 50-70%.

[0099] Библиотеки полинуклеотидов можно скорректировать по их содержанию повторяющейся последовательности. В некоторых случаях библиотеки полинуклеотидов предпочтительно характеризуются высоким содержанием повторяющейся последовательности. Например, сконструирована библиотека, в которой увеличенная частота полинуклеотидов характеризуется содержанием повторяющейся последовательности в диапазоне, составляющем приблизительно 40% - приблизительно 90%. В некоторых случаях библиотеки полинуклеотидов содержат низкое содержание повторяющейся последовательности. Например, сконструирована библиотека, в которой увеличенная частота полинуклеотидов характеризуется содержанием повторяющейся последовательности в диапазоне, составляющем приблизительно 10% - приблизительно 60%. Можно сконструировать библиотеку, предпочтительно характеризующуюся высоким и низким содержанием повторяющейся последовательности. Например, можно сконструировать библиотеку, в которой увеличенная частота полинуклеотидов характеризуется содержанием повторяющейся последовательности преимущественно в диапазоне, составляющем приблизительно 10% - приблизительно 30% и в диапазоне, составляющем приблизительно 70% - приблизительно 90%. В некоторых случаях библиотека предпочтительно характеризуется однородным содержанием повторяющейся последовательности. Например, частота полинуклеотидов является однородной с содержанием повторяющейся последовательности в диапазоне, составляющем приблизительно 10% - приблизительно 90%. В некоторых случаях библиотека содержит полинуклеотиды с процентным отношением повторяющейся последовательности,

составляющим приблизительно 10% - приблизительно 95%. В некоторых случаях описанная в настоящем документе библиотека содержит полинуклеотиды, характеризующиеся больше чем 30% различных полинуклеотидов, характеризующихся процентным отношением повторяющейся последовательности от 10% до 30% или от 70 до 90%. В некоторых случаях описанная в настоящем документе библиотека содержит полинуклеотиды, характеризующиеся меньше чем приблизительно 15% полинуклеотидов, характеризующихся процентным отношением повторяющейся последовательности от 10% до 30% или от 60 до 90%.

[0100] Создание библиотек полинуклеотидов с заданным содержанием повторяющейся последовательности в некоторых случаях происходит путем комбинации по меньшей мере 2 библиотек полинуклеотидов с различным содержанием повторяющейся последовательности. В некоторых случаях по меньшей мере 2, 3, 4, 5, 6, 7, 10 или больше чем 10 библиотек полинуклеотидов комбинируют для создания популяции полинуклеотидов с заданным содержанием повторяющейся последовательности. В некоторых случаях не больше чем 2, 3, 4, 5, 6, 7 или 10 библиотек полинуклеотидов комбинируют для создания популяции неидентичных полинуклеотидов с заданным содержанием повторяющейся последовательности.

[0101] В некоторых случаях содержание повторяющейся последовательности корректируют путем синтеза меньшего или большего количества полинуклеотидов на кластер. Например, по меньшей мере 25, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 или больше чем 1000 неидентичных полинуклеотидов синтезируют на одном кластере. В некоторых случаях не больше чем приблизительно 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 неидентичных полинуклеотидов синтезируют на одном кластере. В некоторых случаях 50 - 500 неидентичных полинуклеотидов синтезируют на одном кластере. В некоторых случаях 100 - 200 неидентичных полинуклеотидов синтезируют на одном кластере. В некоторых случаях приблизительно 100, приблизительно 120, приблизительно 125, приблизительно 130, приблизительно 150, приблизительно 175 или приблизительно 200 неидентичных полинуклеотидов синтезируют на одном кластере.

[0102] В некоторых случаях содержание повторяющейся последовательности корректируют путем синтеза неидентичных полинуклеотидов различной длины. Например, длина каждого из синтезированных неидентичных полинуклеотидов может составлять по меньшей мере или приблизительно по меньшей мере 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 300, 400, 500, 2000 нуклеотидов или больше. Длина синтезированных неидентичных полинуклеотидов может составлять не более чем или приблизительно не более чем 2000, 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 синтезированных неидентичных полинуклеотидов может попадать в следующий диапазон: 10-2000, 10-500, 9-400, 11-300, 12-200, 13-150, 14-100, 15-50, 16-45, 17-40, 18-35 и 19-25.

[0103] Создание библиотек полинуклеотидов с контролируемой стехиометрией содержания вторичной структуры

[0104] Описанную в настоящем документе библиотеку полинуклеотидов можно синтезировать с заданным содержанием вторичной структуры. В некоторых случаях коррекция библиотеки полинуклеотидов в отношении содержания вторичной структуры дает в результате улучшенное представление полинуклеотидов.

[0105] Вторичная структура может содержать три или больше нуклеотидов в одной или нескольких полинуклеотидных цепях, которые образуют структуру, такую как спираль (например, альфа-спираль), бета-лист, «петля-на-стебле», псевдоузел, гомодимер или гетеродимер. «Петля-на-стебле» может представлять собой шпильную петлю, внутреннюю петлю, выпетливание или мультипетлю. Тип вторичной структуры и их потенциал к образованию можно спрогнозировать из данных в отношении последовательностей. Укладка или гибридизация неразветвленных последовательностей во вторичные структуры могут происходить в то время, когда полинуклеотиды прикреплены к твердой подложке или после расщепления в растворе.

[0106] Содержание вторичной структуры в библиотеке полинуклеотидов может составлять по меньшей мере 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% или больше чем 95%. В некоторых случаях содержание вторичной структуры в библиотеке полинуклеотидов составляет не более чем 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% или не больше чем 100%. В некоторых случаях содержание вторичной структуры находится в диапазоне, составляющем приблизительно 5-95%, 10-90%, 30-80%, 40-75% или 50-70%.

[0107] Библиотеки полинуклеотидов можно скорректировать по их содержанию вторичной структуры. В некоторых случаях библиотеки полинуклеотидов предпочтительно характеризуются высоким содержанием вторичной структуры. Например, сконструирована библиотека, в которой увеличенная частота полинуклеотидов характеризуется содержанием вторичной структуры в диапазоне, составляющем приблизительно 40% - приблизительно 90%. В некоторых случаях библиотеки полинуклеотидов содержат низкое содержание вторичной структуры. Например, сконструирована библиотека, в которой увеличенная частота полинуклеотидов характеризуется содержанием вторичной структуры, которое находится в диапазоне, составляющем приблизительно 10% - приблизительно 60%. Можно сконструировать

библиотеку, предпочтительно характеризующуюся высоким и низким содержанием вторичной структуры. Например, можно сконструировать библиотеку, в которой увеличенная частота полинуклеотидов характеризуется содержанием вторичной структуры преимущественно в диапазоне, составляющем приблизительно 10% - приблизительно 30%, и в диапазоне, составляющем приблизительно 70% - приблизительно 90%. В некоторых случаях библиотека предпочтительно характеризуется однородным содержанием вторичной структуры. Например, частота полинуклеотидов является однородной с содержанием вторичной структуры в диапазоне, составляющем приблизительно 10% - приблизительно 90%. В некоторых случаях библиотека содержит полинуклеотиды с процентным отношением вторичной структуры, составляющим приблизительно 10% - приблизительно 95%. В некоторых случаях описанная в настоящем документе библиотека содержит полинуклеотиды, характеризующиеся больше чем 30% различных полинуклеотидов, характеризующихся процентным отношением вторичной структуры, составляющим 10% - 30% или 70 - 90%. В некоторых случаях описанная в настоящем документе библиотека содержит полинуклеотиды, характеризующиеся меньше чем приблизительно 15% полинуклеотидов, характеризующихся процентным отношением вторичной структуры, составляющим 10% - 30% или 60 - 90%.

[0108] Создание библиотек полинуклеотидов с заданным содержанием вторичной структуры в некоторых случаях происходит путем комбинации по меньшей мере 2 библиотек полинуклеотидов с различным содержанием повторяющейся последовательности. В некоторых случаях по меньшей мере 2, 3, 4, 5, 6, 7, 10 или больше чем 10 библиотек полинуклеотидов комбинируют для создания популяции полинуклеотидов с заданным содержанием вторичной структуры. В некоторых случаях не больше чем 2, 3, 4, 5, 6, 7 или 10 библиотек полинуклеотидов комбинируют для создания популяции неидентичных полинуклеотидов с заданным содержанием вторичной структуры.

[0109] В некоторых случаях содержание вторичной структуры корректируют путем синтеза меньшего или большего количества полинуклеотидов на один кластер. Например, по меньшей мере 25, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 или больше чем 1000 неидентичных полинуклеотидов синтезируют на одном кластере. В некоторых случаях не больше чем приблизительно 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 неидентичных полинуклеотидов синтезируют на одном кластере. В некоторых случаях 50 - 500 неидентичных полинуклеотидов синтезируют на одном кластере. В некоторых случаях 100 - 200 неидентичных полинуклеотидов синтезируют на одном кластере. В некоторых случаях приблизительно 100, приблизительно 120, приблизительно

125, приблизительно 130, приблизительно 150, приблизительно 175 или приблизительно 200 неидентичных полинуклеотидов синтезируют на одном кластере.

[0110] В некоторых случаях содержание вторичной структуры корректируют путем синтеза неидентичных полинуклеотидов различной длины. Например, длина каждого из синтезированных неидентичных полинуклеотидов может составлять по меньшей мере или приблизительно по меньшей мере 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 300, 400, 500, 2000 нуклеотидов или больше. Длина синтезированных неидентичных полинуклеотидов может составлять не более чем или приблизительно не более чем 2000, 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 синтезированных неидентичных полинуклеотидов может попадать в следующий диапазон: 10-2000, 10-500, 9-400, 11-300, 12-200, 13-150, 14-100, 15-50, 16-45, 17-40, 18-35 и 19-25.

[0111] Создание библиотек полинуклеотидов с контролируемой стехиометрией содержания последовательностей

[0112] В некоторых случаях библиотеку полинуклеотидов синтезируют с заданным распределением требуемых полинуклеотидных последовательностей. В некоторых случаях коррекция библиотеки полинуклеотидов для обогащения конкретных требуемых последовательностей приводит к улучшенным результатам последующего применения.

[0113] Одну или несколько конкретных последовательностей можно выбрать на основании их оценки в последующем применении. В некоторых случаях оценка представляет собой аффинность связывания с целевыми последовательностями для амплификации, обогащения или обнаружения, стабильность, температуру плавления, биологическую активность, способность к сборке в фрагменты большего размера или другое свойство полинуклеотидов. В некоторых случаях оценка является эмпирической или спрогнозированной на основании предшествующих экспериментов и/или компьютерных алгоритмов.

[0114] Выбранные последовательности в библиотеке полинуклеотидов могут составлять по меньшей мере 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% или больше чем 95% последовательностей. В некоторых случаях выбранные последовательности в библиотеке полинуклеотидов составляют не более чем 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% или не более чем 100% последовательностей. В некоторых случаях выбранные последовательности находятся в диапазоне, составляющем приблизительно 5-95%, 10-90%, 30-80%, 40-75% или 50-70% последовательностей.

[0115] Библиотеки полинуклеотидов можно скорректировать по частоте каждой выбранной последовательности. В некоторых случаях библиотеки полинуклеотидов

предпочтительно содержат повышенное количество выбранных последовательностей. Например, сконструирована библиотека, в которой увеличенная частота полинуклеотидов выбранных последовательностей находится в диапазоне, составляющем приблизительно 40% - приблизительно 90%. В некоторых случаях библиотеки полинуклеотидов содержат низкое количество выбранных последовательностей. Например, сконструирована библиотека, в которой увеличенная частота полинуклеотидов выбранных последовательностей находится в диапазоне, составляющем приблизительно 10% - приблизительно 60%. Можно сконструировать библиотеку, предпочтительно характеризующуюся повышенной и пониженной частотой выбранных последовательностей. В некоторых случаях библиотека предпочтительно содержит однородное представление последовательностей. Например, частота полинуклеотидов является однородной по отношению к частоте выбранных последовательностей, находясь в диапазоне, составляющем приблизительно 10% - приблизительно 90%. В некоторых случаях библиотека содержит полинуклеотиды с частотой выбранных последовательностей, составляющей приблизительно 10% - приблизительно 95% последовательностей.

[0116] Создание библиотек полинуклеотидов с заданной частотой выбранных последовательностей в некоторых случаях происходит путем комбинации по меньшей мере 2 библиотек полинуклеотидов с различной частотой содержания выбранных последовательностей. В некоторых случаях по меньшей мере 2, 3, 4, 5, 6, 7, 10 или больше чем 10 библиотек полинуклеотидов комбинируют для создания популяции полинуклеотидов с заданной частотой выбранных последовательностей. В некоторых случаях не больше чем 2, 3, 4, 5, 6, 7 или 10 библиотек полинуклеотидов комбинируют для создания популяции неидентичных полинуклеотидов с заданной частотой выбранных последовательностей.

[0117] В некоторых случаях частоту выбранных последовательностей корректируют путем синтеза меньшего или большего количества полинуклеотидов на один кластер. Например, по меньшей мере 25, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 или больше чем 1000 неидентичных полинуклеотидов синтезируют на одном кластере. В некоторых случаях не больше чем приблизительно 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 неидентичных полинуклеотидов синтезируют на одном кластере. В некоторых случаях 50 - 500 неидентичных полинуклеотидов синтезируют на одном кластере. В некоторых случаях 100 - 200 неидентичных полинуклеотидов синтезируют на одном кластере. В некоторых случаях приблизительно 100, приблизительно 120, приблизительно

125, приблизительно 130, приблизительно 150, приблизительно 175 или приблизительно 200 неидентичных полинуклеотидов синтезируют на одном кластере.

[0118] В некоторых случаях частоту выбранных последовательностей корректируют путем синтеза неидентичных полинуклеотидов различной длины. Например, длина каждого из синтезированных неидентичных полинуклеотидов может составлять по меньшей мере или приблизительно по меньшей мере 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 300, 400, 500, 2000 нуклеотидов или больше. Длина синтезированных неидентичных полинуклеотидов может составлять не более чем или приблизительно не более чем 2000, 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 синтезированных неидентичных полинуклеотидов может попадать в следующий диапазон: 10-2000, 10-500, 9-400, 11-300, 12-200, 13-150, 14-100, 15-50, 16-45, 17-40, 18-35 и 19-25.

[0119] Структуры полинуклеотидных зондов

[0120] Библиотеки полинуклеотидных зондов можно использовать для обогащения конкретных целевых последовательностей в большей популяции полинуклеотидов образца. В некоторых случаях каждый из полинуклеотидных зондов содержит связывающуюся с мишенью последовательность, комплементарную одной или нескольким целевым последовательностям, одну или несколько не связывающихся с мишенью последовательностей и один или несколько сайтов связывания праймера, таких как универсальные сайты связывания праймера. Связывающиеся с мишенью последовательности, которые являются комплементарными или по меньшей мере частично комплементарными, в некоторых случаях связываются (гибридизуются) с целевыми последовательностями. Сайты связывания праймера, такие как универсальные сайты связывания праймера, облегчают одновременную амплификацию всех представителей библиотеки зондов или субпопуляции представителей. В некоторых случаях зонды дополнительно содержат последовательность штрих-кода или индекса. Штрих-коды представляют собой последовательности нуклеиновой кислоты, которые позволяют идентифицировать некоторый признак полинуклеотида, с которым ассоциирован штрих-код. После секвенирования область штрих-кода образует собой индикатор для идентификации характеристики, ассоциированной с кодирующей областью или источником образца. Штрих-коды можно сконструировать с подходящей длиной, чтобы обеспечить достаточную степень идентификации, например, по меньшей мере приблизительно 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, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 или больше оснований в длину. Множественные штрих-коды, такие как

приблизительно 2, 3, 4, 5, 6, 7, 8, 9, 10 или больше штрих-кодов, можно использовать на одной и той же молекуле, необязательно разделенные с помощью не штрих-кодовых последовательностей. Согласно некоторым вариантам осуществления каждый штрих-код в множестве штрих-кодов отличается от другого штрих-кода в множестве по меньшей мере тремя положениями оснований, например, по меньшей мере приблизительно 3, 4, 5, 6, 7, 8, 9, 10 положениями или больше. В некоторых случаях полинуклеотиды лигированы с одним или несколькими молекулярными (или аффинными) маркерами, такими как малая молекула, пептид, антиген, металл или белок, с образованием зонда для последующего захвата представляющих интерес целевых последовательностей. В некоторых случаях два зонда, которые обладают комплементарными связывающимися с мишенью последовательностями, которые способны к гибридизации, образуют двухцепочечную пару зондов.

[0121] Описанные в настоящем документе зонды могут являться комплементарными целевым последовательностям, которые представляют собой последовательности в геноме. Описанные в настоящем документе зонды могут являться комплементарными целевым последовательностям, которые представляют собой экзонные последовательности в геноме. Описанные в настоящем документе зонды могут являться комплементарными целевым последовательностям, которые представляют собой интронные последовательности в геноме. В некоторых случаях зонды содержат связывающуюся с мишенью последовательность, комплементарную целевой последовательности, и по меньшей мере одну не связывающуюся с мишенью последовательность, которая не является комплементарной мишени. В некоторых случаях связывающаяся с мишенью последовательность зонда составляет приблизительно 120 нуклеотидов в длину или по меньшей мере 10, 15, 20, 25, 50, 75, 100, 110, 120, 125, 140, 150, 160, 175, 200, 300, 400, 500 или больше чем 500 нуклеотидов в длину. Связывающаяся с мишенью последовательность составляет в некоторых случаях не больше чем 10, 15, 20, 25, 50, 75, 100, 125, 150, 175, 200 или не больше чем 500 нуклеотидов в длину. Связывающаяся с мишенью последовательность зонда составляет в некоторых случаях приблизительно 120 нуклеотидов в длину или приблизительно 10, 15, 20, 25, 40, 50, 60, 70, 80, 85, 87, 90, 95, 97, 100, 105, 110, 115, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 135, 140, 145, 150, 155, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 175, 180, 190, 200, 210, 220, 230, 240, 250, 300, 400 или приблизительно 500 нуклеотидов в длину. Связывающаяся с мишенью последовательность составляет в некоторых случаях приблизительно 20 - приблизительно 400 нуклеотидов в длину, или приблизительно 30 - приблизительно 175, приблизительно 40 - приблизительно 160, приблизительно 50 -

приблизительно 150, приблизительно 75 - приблизительно 130, приблизительно 90 - приблизительно 120, или приблизительно 100 - приблизительно 140 нуклеотидов в длину. Не связывающаяся(иеся) с мишенью последовательность(и) зонда составляет(ют) в некоторых случаях по меньшей мере приблизительно 20 нуклеотидов в длину или по меньшей мере приблизительно 1, 5, 10, 15, 17, 20, 23, 25, 50, 75, 100, 110, 120, 125, 140, 150, 160, 175 или больше чем приблизительно 175 нуклеотидов в длину. Не связывающаяся с мишенью последовательность часто составляет не больше чем приблизительно 5, 10, 15, 20, 25, 50, 75, 100, 125, 150, 175 или не больше чем приблизительно 200 нуклеотидов в длину. Не связывающаяся с мишенью последовательность зонда часто составляет приблизительно 20 нуклеотидов в длину или приблизительно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150 или приблизительно 200 нуклеотидов в длину. Не связывающаяся с мишенью последовательность в некоторых случаях составляет приблизительно 1 - приблизительно 250 нуклеотидов в длину, или приблизительно 20 - приблизительно 200, приблизительно 10 - приблизительно 100, приблизительно 10 - приблизительно 50, приблизительно 30 - приблизительно 100, приблизительно 5 - приблизительно 40 или приблизительно 15 - приблизительно 35 нуклеотидов в длину. Не связывающаяся с мишенью последовательность часто содержит последовательности, которые не являются комплементарными целевой последовательности, и/или содержат последовательности, которые не используют для связывания с праймерами. В некоторых случаях не связывающаяся с мишенью последовательность содержит повтор одного нуклеотида, например, полиаденин или политимидин. Зонд часто не содержит ни одной или содержит по меньшей мере одну не связывающуюся с мишенью последовательность. В некоторых случаях зонд содержит одну или две не связывающиеся с мишенью последовательности. Не связывающаяся с мишенью последовательность может являться смежной с одной или несколькими связывающимися с мишенью последовательностями в зонде. Например, не связывающаяся с мишенью последовательность расположена на 5' или 3' конце зонда. В некоторых случаях не связывающаяся с мишенью последовательность прикреплена к молекулярному маркеру или спейсеру.

[0122] В некоторых случаях не связывающаяся(иеся) с мишенью последовательность(и) может(могут) представлять собой сайт связывания праймера. Каждый из сайтов связывания праймера часто составляет по меньшей мере приблизительно 20 нуклеотидов в длину или по меньшей мере приблизительно 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 или по меньшей мере приблизительно 40 нуклеотидов в длину. Каждый сайт связывания праймера в некоторых случаях составляет не больше чем

приблизительно 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 или не больше чем приблизительно 40 нуклеотидов в длину. Каждый сайт связывания праймера в некоторых случаях составляет приблизительно 10 - приблизительно 50 нуклеотидов в длину, или приблизительно 15 - приблизительно 40, приблизительно 20 - приблизительно 30, приблизительно 10 - приблизительно 40, приблизительно 10 - приблизительно 30, приблизительно 30 - приблизительно 50, или приблизительно 20 - приблизительно 60 нуклеотидов в длину. В некоторых случаях полинуклеотидные зонды содержат по меньшей мере два сайта связывания праймера. В некоторых случаях сайты связывания праймера могут представлять собой универсальные сайты связывания праймера, причем все зонды содержат идентичные последовательности связывания праймера на этих сайтах. В некоторых случаях пара полинуклеотидных зондов, нацеленных на конкретную последовательность и обратнoкомплементарную ей последовательность (например, область геномной ДНК), представлены как **300** на фиг. **3А**, они содержат первую связывающуюся с мишенью последовательность **301**, вторую связывающуюся с мишенью последовательность **302**, первую не связывающуюся с мишенью последовательность **303** и вторую не связывающуюся с мишенью последовательность **304**. Например, пара полинуклеотидных зондов, комплементарных конкретной последовательности (например, область геномной ДНК).

[0123] В некоторых случаях первая связывающаяся с мишенью последовательность **301** является обратнoкомплементарной второй связывающейся с мишенью последовательности **302**. В некоторых случаях обе связывающиеся с мишенью последовательности химически синтезируют перед амплификацией. В альтернативном расположении пара полинуклеотидных зондов, нацеленных на конкретную последовательность и обратнoкомплементарную ей последовательность (например, область геномной ДНК), представлены как **305** на фиг. **3В**, они содержат первую связывающуюся с мишенью последовательность **301**, вторую связывающуюся с мишенью последовательность **302**, первую не связывающуюся с мишенью последовательность **303**, вторую не связывающуюся с мишенью последовательность **304**, третью не связывающуюся с мишенью последовательность **306** и четвертую не связывающуюся с мишенью последовательность **307**. В некоторых случаях первая связывающаяся с мишенью последовательность **301** является обратнoкомплементарной второй связывающейся с мишенью последовательности **302**. В некоторых случаях одна или несколько не связывающихся с мишенью последовательностей содержат полиаденин или политимидин.

[0124] В некоторых случаях оба зонда в паре метят по меньшей мере с помощью одного молекулярного маркера. В некоторых случаях ПЦР используют для введения

молекулярных маркеров (посредством праймеров, содержащих молекулярный маркер) на зонды во время амплификации. В некоторых случаях молекулярный маркер содержит один или несколько из следующего: биотин, фолат, полигистидин, FLAG-маркер, глутатион или другой молекулярный маркер в соответствии с настоящим описанием. В некоторых случаях зонды метят на 5' конце. В некоторых случаях зонды метят на 3' конце. В некоторых случаях как 5', так и 3' концы метят с помощью молекулярного маркера. В некоторых случаях 5' конец первого зонда в паре метят с помощью по меньшей мере одного молекулярного маркера и 3' конец второго зонда в паре метят с помощью по меньшей мере одного молекулярного маркера. В некоторых случаях спейсер присутствует между одним или несколькими молекулярными маркерами и нуклеиновыми кислотами зонда. В некоторых случаях спейсер может содержать алкил, многоатомный спирт или полиаминоцепь, пептид или полинуклеотид. Твердая подложка, используемая для захвата комплексов зонд-целевая нуклеиновая кислота, в некоторых случаях представляет собой микрошарик или поверхность. Твердая подложка в некоторых случаях содержит стекло, пластик или другой материал, способный содержать фрагмент, обеспечивающий захват, который будет связывать молекулярный маркер. В некоторых случаях микрошарик представляет собой магнитный микрошарик. Например, зонды, меченные биотином, захватывают с помощью магнитного микрошарика, содержащего стрептавидин. Зонды приводят в контакт с библиотекой нуклеиновых кислот для обеспечения связывания зондов с целевыми последовательностями. В некоторых случаях блокирующие полинуклеиновые кислоты добавляют для предотвращения связывания зондов с одной или несколькими адаптерными последовательностями, прикрепленными к целевым нуклеиновым кислотам. В некоторых случаях блокирующие полинуклеиновые кислоты содержат один или несколько аналогов нуклеиновых кислот. В некоторых случаях в блокирующих полинуклеиновых кислотах урацил замещен тиминном на одном или нескольких положениях.

[0125] Описанные в настоящем документе зонды могут содержать комплементарные связывающиеся с мишенью последовательности, которые связываются с одной или несколькими целевыми последовательностями нуклеиновой кислоты. В некоторых случаях целевые последовательности представляют собой любую последовательность нуклеиновой кислоты ДНК или РНК. В некоторых случаях целевые последовательности могут быть длиннее, чем вставка-зонд. В некоторых случаях целевые последовательности могут быть короче, чем вставка-зонд. В некоторых случаях целевые последовательности могут быть той же длины, что и вставка-зонд. Например, длина целевой последовательности может составлять по меньшей мере или приблизительно по меньшей мере 2, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 300, 400, 500, 1000, 2000,

5000, 12000, 20000 нуклеотидов или больше. Длина целевой последовательности может составлять не более чем или приблизительно не более чем 20000, 12000, 5000, 2000, 1000, 500, 400, 300, 200, 150, 100, 50, 45, 35, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 2 нуклеотидов или меньше. Длина целевой последовательности может попадать в диапазон, составляющий 2-20000, 3-12000, 5-5, 5000, 10-2000, 10-1000, 10-500, 9-400, 11-300, 12-200, 13-150, 14-100, 15-50, 16-45, 17-40, 18-35 и 19-25. Последовательности зонда могут представлять собой целевые последовательности, ассоциированные с конкретными генами, заболеваниями, регуляторными путями или другими биологическими функциями в соответствии с настоящим описанием.

[0126] В некоторых случаях одна вставка-зонд **403** является комплементарной одной или нескольким целевым последовательностям **402** (фигуры **4A-4G**) в полинуклеиновой кислоте большего размера **400**. Иллюстративная целевая последовательность представляет собой экзон. В некоторых случаях один или несколько зондов нацелены на одну целевую последовательность (фигуры **4A-4G**). В некоторых случаях один зонд может быть нацелен больше чем на одну целевую последовательность. В некоторых случаях связывающаяся с мишенью последовательность зонда нацелена как на целевую последовательность **402**, так и на смежную последовательность **401** (фиг. **4A** и **4B**). В некоторых случаях первый зонд нацелен на первую область и вторую область целевой последовательности, а второй зонд нацелен на вторую область и третью область целевой последовательности (фиг. **4D** и фиг. **4E**). В некоторых случаях множество зондов нацелено на одну целевую последовательность, причем связывающиеся с мишенью последовательности множества зондов содержат одну или несколько последовательностей, которые перекрываются в отношении комплементарности с областью целевой последовательности (фиг. **4G**). В некоторых случаях вставки-зонды не перекрываются в отношении комплементарности с областью целевой последовательности. В некоторых случаях по меньшей мере 2, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 300, 400, 500, 1000, 2000, 5000, 12000, 20000 или больше чем 20000 зондов нацелены на одну целевую последовательность. В некоторых случаях не больше чем 4 зонда направлены на одно перекрытие целевой последовательности, или не больше чем 3, 2, 1, или ни один из зондов не нацелен на одно перекрытие целевой последовательности. В некоторых случаях один или несколько зондов не нацелены на все основания в целевой последовательности, оставляя один или несколько пропусков (фиг. **4C** и фиг. **4F**). В некоторых случаях пропуски находятся вблизи середины целевой последовательности **405** (фиг. **4F**). В некоторых случаях пропуски **404** находятся на 5' или 3' концах целевой последовательности (фиг. **4C**). В некоторых случаях пропуски составляют 6 нуклеотидов в длину. В некоторых случаях пропуски составляют не больше

чем 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40 или не больше чем 50 нуклеотидов в длину. В некоторых случаях пропуски составляют по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40 или по меньшей мере 50 нуклеотидов в длину. В некоторых случаях длина пропусков находится в диапазонах, составляющих 1-50, 1-40, 1-30, 1-20, 1-10, 2-30, 2-20, 2-10, 3-50, 3-25, 3-10 или 3-8 нуклеотидов в длину. В некоторых случаях набор зондов, нацеленных на последовательность, не содержит перекрывающиеся области среди зондов в наборе, если гибридизуется с комплементарной последовательностью. В некоторых случаях набор зондов, нацеленных на последовательность, не содержит каких-либо пропусков среди зондов в наборе, если гибридизуется с комплементарной последовательностью. Зонды можно сконструировать, чтобы сделать максимальным однородное связывание с целевыми последовательностями. В некоторых случаях зонды конструируют для минимизации связывающихся с мишенью последовательностей с высоким или низким содержанием GC, вторичной структурой, повторяющимися/палиндромными последовательностями или другим признаком последовательности, который может препятствовать связыванию зонда с мишенью. В некоторых случаях один зонд может быть нацелен на множество целевых последовательностей.

[0127] Описанная в настоящем документе библиотека зондов может содержать по меньшей мере 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10000, 20000, 50000, 100000, 200000, 500000, 1000000 или больше чем 1000000 зондов. Библиотека зондов может содержать не больше чем 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10000, 20000, 50000, 100000, 200000, 500000 или не больше чем 1000000 зондов. Библиотека зондов может содержать 10 - 500, 20 - 1000, 50 - 2000, 100 - 5000, 500 - 10000, 1000 - 5000, 10000 - 50000, 100000 - 500000 или 50000 - 1000000 зондов. Библиотека зондов может содержать приблизительно 370000; 400000; 500000 или больше различных зондов.

[0128] Применения секвенирования нового поколения

[0129] Последующие применения библиотек полинуклеотидов могут включать в себя секвенирование нового поколения. Например, обогащение целевых последовательностей с помощью библиотеки полинуклеотидных зондов с контролируемой стехиометрией дает в результате более эффективное секвенирование. Производительность библиотеки полинуклеотидов для захвата или гибридизации с мишенями можно установить по ряду различных измерений, описывающих эффективность, достоверность и точность. Например, показатели (метрики) Пикара содержат такие переменные, как размер библиотеки HS (количество уникальных молекул в библиотеке, которые соответствуют целевым областям, рассчитанные из пар прочтений), среднее целевое покрытие (процентное отношение оснований, достигающих конкретного уровня покрытия), глубина

покрытия (число прочтений, включающих в себя заданный нуклеотид), кратность обогащения (прочтения последовательностей, однозначно соответствующие мишени образом/прочтения, соответствующие общему образцу, умноженное на общую длину образца/длину мишени), процентное отношение оснований, не связанных с зондами (процентное отношение оснований, не соответствующих основаниям зондов), используемые основания на мишени, коэффициент отсева по AT или GC, штраф на кратность избыточного покрытия для 80% оснований (fold 80 base penalty) (кратность необходимого избыточного покрытия, чтобы поднять 80 процентов ненулевых мишеней до среднего уровня покрытия), процентное отношение мишеней с нулевым покрытием, прочтения PF (число прочтений, прошедших фильтр качества), процентное отношение выбранных оснований (сумма оснований на зонде и оснований вблизи зонда, деленная на общее количество выровненных оснований), процентное отношение дублирования или другая переменная в соответствии с настоящим описанием изобретения.

[0130] Глубина прочтения (глубина секвенирования, или сэмплинг) представляет общее количество раз, когда секвенированный фрагмент нуклеиновой кислоты (“прочтение”) получают для последовательности. Теоретическую глубину прочтения определяют как ожидаемое число раз прочтения одного и того же нуклеотида при условии, что прочтения идеально распределены по идеализированному геному. Глубину прочтения выражают как функцию от % покрытия (или ширина покрытия). Например, 10 миллионов прочтений генома из 1 миллиона оснований, идеально распределенных, теоретически дают в результате 10X глубину прочтения 100% последовательностей. Экспериментально большее число прочтений (повышенная теоретическая глубина прочтения, или избыточный сэмплинг) может понадобиться для получения требуемой глубины прочтения для процентного отношения целевых последовательностей. Обогащение целевых последовательностей библиотекой зондов с контролируемой стехиометрией увеличивает эффективность последующего секвенирования, поскольку меньшее количество общих прочтений понадобится для получения экспериментального результата с приемлемым числом прочтений для требуемого процентного отношения целевых последовательностей. Например, в некоторых случаях 55X теоретическая глубина прочтения целевых последовательностей дает в результате по меньшей мере 30X покрытие по меньшей мере 90% последовательностей. В некоторых случаях не больше чем 55X теоретическая глубина прочтения целевых последовательностей дает в результате по меньшей мере 30X глубину прочтения по меньшей мере 80% последовательностей. В некоторых случаях не больше чем 55X теоретическая глубина прочтения целевых последовательностей дает в результате по меньшей мере 30X глубину прочтения по меньшей мере 95% последовательностей. В

некоторых случаях не больше чем 55X теоретическая глубина прочтения целевых последовательностей дает в результате по меньшей мере 10X глубину прочтения по меньшей мере 98% последовательностей. В некоторых случаях 55X теоретическая глубина прочтения целевых последовательностей дает в результате по меньшей мере 20X глубину прочтения по меньшей мере 98% последовательностей. В некоторых случаях не больше чем 55X теоретическая глубина прочтения целевых последовательностей дает в результате по меньшей мере 5X глубину прочтения по меньшей мере 98% последовательностей. Увеличение концентрации зондов во время гибридизации с мишенями может приводить к увеличению глубины прочтения. В некоторых случаях концентрацию зондов увеличивают по меньшей мере в 1,5, 2,0, 2,5, 3, 3,5, 4, 5 раз или больше чем в 5 раз. В некоторых случаях увеличение концентрации зонда дает в результате по меньшей мере 1000% увеличение, или 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300%, 500%, 750%, 1000%, или больше чем 1000% увеличение глубины прочтения. В некоторых случаях увеличение концентрации зонда в 3 раза дает в результате 1000% увеличение глубины прочтения.

[0131] Частота встречаемости мишени представляет процентное отношение прочтений секвенирования, которые соответствуют требуемым целевым последовательностям. В некоторых случаях библиотека полинуклеотидных зондов с контролируемой стехиометрией дает в результате частоту встречаемости мишени, составляющую по меньшей мере 30%, или по меньшей мере 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, или по меньшей мере 90%. Увеличение концентрации полинуклеотидных зондов во время контакта с целевыми нуклеиновыми кислотами приводит к увеличению частоты встречаемости мишени. В некоторых случаях концентрация зондов увеличивается по меньшей мере в 1,5, 2,0, 2,5, 3, 3,5, 4, 5 или больше чем в 5 раз. В некоторых случаях увеличение концентрации зонда приводит по меньшей мере к 20% увеличению, или 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 200%, 300% или по меньшей мере 500% увеличению целевого связывания. В некоторых случаях увеличение концентрации зонда в 3 раза дает в результате 20% увеличение частоты встречаемости мишени.

[0132] Однородность покрытия в некоторых случаях рассчитывают как глубину прочтения как функцию идентичности целевой последовательности. Повышенная однородность покрытия дает в результате пониженное число прочтений секвенирования, необходимых для получения требуемой глубины прочтения. Например, свойство целевой последовательности может влиять на глубину прочтения, например, высокое или низкое содержание GC или AT, повторяющиеся последовательности, хвостовые аденины, вторичная структура, аффинность к связыванию целевой последовательности (для

амплификации, обогащения или обнаружения), стабильность, температура плавления, биологическая активность, способность к сборке в фрагменты большего размера, последовательности, содержащие модифицированные нуклеотидов или аналоги нуклеотидов или любое другое свойство полинуклеотидов. Обогащение целевых последовательностей с помощью библиотек полинуклеотидных зондов с контролируемой стехиометрией дает в результате повышенную однородность покрытия после секвенирования. В некоторых случаях 95% последовательностей характеризуются глубиной прочтения, которая находится в пределах 1-кратной от средней глубины прочтения библиотеки, или приблизительно 0,05-, 0,1-, 0,2-, 0,5-, 0,7-, 1-, 1,2-, 1,5-, 1,7- или приблизительно в пределах 2-кратной от средней глубины прочтения библиотеки. В некоторых случаях 80%, 85%, 90%, 95%, 97% или 99% последовательностей характеризуются глубиной прочтения, которая находится в пределах 1-кратной от среднего значения.

[0133] Обогащение целевых нуклеиновых кислот с помощью библиотеки полинуклеотидных зондов

[0134] Описанную в настоящем документе библиотеку зондов можно использовать для обогащения целевых полинуклеотидов, присутствующих в популяции полинуклеотидов образца, для различных последующих применений. В некоторых случаях образец получают из одного или нескольких источников и популяцию полинуклеотидов образца выделяют с использованием общепринятых техник, известных в настоящей области техники. Образцы получают (в качестве неограничивающего примера) из биологических источников, таких как слюна, кровь, ткань, кожа, или полностью синтетических образцов. Множество полинуклеотидов, полученных из образца, фрагментируют, проводят репарацию концов и аденилируют с образованием двухцепочечного фрагмента нуклеиновой кислоты образца. В некоторых случаях репарацию концов осуществляют путем обработки с помощью одного или нескольких ферментов, таких как ДНК-полимераза Т4, фермент Кленова и полинуклеотидкиназа Т4 в соответствующем буфере. Добавляют нуклеотид-«липкий конец» для облегчения лигирования к адаптерам, в некоторых случаях с 3' - 5' экзонуклеазой без фрагмента Кленова и dATP.

[0135] Адаптеры можно лигировать по обоим концам фрагментов полинуклеотидов образца с помощью лигазы, такой как лигаза Т4, с получением библиотеки маркированных адаптерами полинуклеотидных цепей, и библиотеку маркированных адаптерами полинуклеотидов амплифицируют с помощью праймеров, таких как универсальные праймеры. В некоторых случаях адаптеры представляют собой адаптеры Y-образной

формы, содержащие один или несколько сайтов связывания праймера, одну или несколько прививающих областей и одну или несколько индексных областей. В некоторых случаях одна или несколько индексных областей присутствуют на каждой цепи адаптера. В некоторых случаях прививающие области комплементарны поверхности проточной кюветы и облегчают секвенирование нового поколения библиотек образца. В некоторых случаях адаптеры Y-образной формы содержат частично комплементарные последовательности. В некоторых случаях Y-образные адаптеры содержат один тимидиновый «липкий конец», которые гибридизуются с адениновым «липким концом» маркированных двухцепочечным адаптером полинуклеотидных цепей. Y-образные адаптеры могут содержать модифицированные нуклеиновые кислоты, которые являются устойчивыми к расщеплению. Например, тиофосфатный остов используют для прикрепления тимидинового «липкого конца» к 3' концу адаптеров. Библиотеку двухцепочечных фрагментов нуклеиновой кислоты образца затем денатурируют в присутствии блокаторов адаптеров. Блокаторы адаптеров минимизируют нецелевую гибридизацию зондов с адаптерными последовательностями (вместо целевых последовательностей), присутствующими на маркированных адаптерах полинуклеотидных цепях. Денатурацию проводят в некоторых случаях при 96°C, или при приблизительно 85, 87, 90, 92, 95, 97, 98 или приблизительно 99°C. Нацеленную на полинуклеотиды библиотеку (библиотеку зондов) денатурируют в гибридизационном растворе, в некоторых случаях при 96°C, приблизительно при 85, 87, 90, 92, 95, 97, 98 или 99°C. Денатурированную библиотеку маркированных адаптерами полинуклеотидов и гибридизационный раствор инкубируют в течение подходящего периода времени и при подходящей температуре для обеспечения гибридизации зондов с их комплементарными целевыми последовательностями. В некоторых случаях подходящая температура гибридизации составляет приблизительно 45 - 80°C или по меньшей мере 45, 50, 55, 60, 65, 70, 75, 80, 85 или 90°C. В некоторых случаях температура гибридизации составляет 70°C. В некоторых случаях подходящее время гибридизации составляет 16 часов, или по меньшей мере 4, 6, 8, 10, 12, 14, 16, 18, 20, 22 или больше чем 22 часа, или приблизительно 12 - 20 часов. Буфер для связывания затем добавляют к зондам, гибридизованным с маркированными адаптерами полинуклеотидами, и твердую подложку, содержащую фрагмент для захвата используют для селективного связывания зондов, гибридизованных с полинуклеотидами, маркированными адаптерами. Твердую подложку отмывают с помощью буфера для удаления несвязанных полинуклеотидов до добавления элюирующего буфера для высвобождения обогащенных, маркированных полинуклеотидных фрагментов с твердой подложки. В некоторых случаях твердую

подложку отмывают 2 раза или 1, 2, 3, 4, 5 или 6 раз. Обогащенную библиотеку маркированных адаптерами полинуклеотидных фрагментов амплифицируют и обогащенную библиотеку секвенируют.

[0136] Множество нуклеиновых кислот (т.е. геномную последовательность) можно получить из образца и фрагментировать, необязательно осуществить репарацию концов и аденилировать. Адаптеры лигируют по обоим концам полинуклеотидных фрагментов с получением библиотеки маркированных адаптерами полинуклеотидных цепей и библиотеку маркированных адаптерами полинуклеотидов амплифицируют. Библиотеку маркированных адаптерами полинуклеотидов затем денатурируют при высокой температуре, предпочтительно при 96°C, в присутствии блокаторов адаптеров. Нацеленную на полинуклеотиды библиотеку (библиотеку зондов) денатурируют в гибридизационном растворе при высокой температуре, предпочтительно приблизительно 90 - 99°C, и объединяют с денатурированной библиотекой маркированных полинуклеотидов в гибридизационном растворе в течение приблизительно 10 - 24 часов при приблизительно 45 - 80°C. Буфер для связывания затем добавляют к зондам, гибридизованным с маркированными полинуклеотидами, и твердую подложку, содержащую фрагмент для захвата, используют для селективного связывания с зондами, гибридизованными с полинуклеотидами, маркированными адаптерами. Твердую подложку отмывают один или несколько раз с помощью буфера, предпочтительно приблизительно 2 - 5 раз для удаления несвязанных полинуклеотидов до добавления элюирующего буфера для высвобождения обогащенных, маркированных адаптерами полинуклеотидных фрагментов с твердой подложки. Обогащенную библиотеку маркированных адаптерами полинуклеотидных фрагментов амплифицируют и затем библиотеку секвенируют. Альтернативные переменные эксперимента, такие как время инкубации, температуры, объемы реакции/концентрации реакции, количество отмывок или другие переменные в соответствии с настоящим описанием изобретения, также используют в настоящем способе.

[0137] Популяцию полинуклеотидов можно обогатить перед лигированием адаптера. В одном примере множество полинуклеотидов получают из образца, фрагментируют, необязательно осуществляют репарацию концов и денатурируют при высокой температуре, предпочтительно 90-99°C. Нацеленную на полинуклеотиды библиотеку (библиотеку зондов) денатурируют в гибридизационном растворе при высокой температуре, предпочтительно приблизительно 90 - 99°C, и объединяют с денатурированной библиотекой маркированных полинуклеотидов в гибридизационном растворе в течение приблизительно 10 - 24 часов при приблизительно 45 - 80°C. Буфер для

связывания затем добавляют к зондам, гибридизованным с маркированными полинуклеотидами, и твердую подложку, содержащую фрагмент для захвата, используют для селективного связывания с зондами, гибридизованными с полинуклеотидами, маркированными адаптерами. Твердую подложку отмывают один или несколько раз с помощью буфера, предпочтительно приблизительно 2 - 5 раз для удаления несвязанных полинуклеотидов перед добавлением элюирующего буфера для высвобождения обогащенных, маркированных адаптерами полинуклеотидных фрагментов с твердой подложки. Обогащенные полинуклеотидные фрагменты затем полиаденилируют, адаптеры лигируют по обоим концам полинуклеотидных фрагментов с получением библиотеки маркированных адаптерами полинуклеотидных цепей, и библиотеку маркированных адаптерами полинуклеотидов амплифицируют. Библиотеку маркированных адаптерами полинуклеотидов затем секвенируют.

[0138] Нацеленную на полинуклеотиды библиотеку также можно использовать для фильтрации ненужных последовательностей из множества полинуклеотидов путем гибридизации ненужных фрагментов. Например, множество полинуклеотидов получают из образца и фрагментируют, необязательно осуществляют репарацию концов и аденилируют. Адаптеры лигируют по обоим концам полинуклеотидных фрагментов с получением библиотеки маркированных адаптерами полинуклеотидных цепей и библиотеку маркированных адаптерами полинуклеотидов амплифицируют. Альтернативно стадии аденилирования и лигирования адаптеров вместо этого выполняют после обогащения полинуклеотидов образца. Библиотеку маркированных адаптерами полинуклеотидов затем денатурируют при высокой температуре, предпочтительно 90-99°C, в присутствии блокаторов адаптеров. Фильтрующую полинуклеотиды библиотеку (библиотеку зондов), разработанную для удаления ненужных, нецелевых последовательностей, денатурируют в гибридизационном растворе при высокой температуре, предпочтительно приблизительно 90 - 99°C, и объединяют с денатурированной библиотекой маркированных полинуклеотидов в гибридизационном растворе в течение приблизительно 10 - 24 часов при приблизительно 45 - 80°C. Буфер для связывания затем добавляют к зондам, гибридизованным с маркированными полинуклеотидами и твердую подложку, содержащую фрагмент для захвата, используют для селективного связывания с зондами, гибридизованными с полинуклеотидами, маркированными адаптерами. Твердую подложку отмывают один или несколько раз с помощью буфера, предпочтительно приблизительно 1 - 5 раз для элюирования несвязанных маркированных адаптерами полинуклеотидных фрагментов. Обогащенную библиотеку несвязанных маркированных адаптерами

полинуклеотидных фрагментов амплифицируют и затем амплифицированную библиотеку секвенируют.

[0139] Высокопараллельный de novo синтез нуклеиновых кислот

[0140] В настоящем документе описан подход на основе платформы, в котором используют миниатюризацию, распараллеливание и вертикальную интеграцию сквозного процесса от синтеза полинуклеотидов 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 100 - 1000 раз по сравнению с традиционными способами синтеза с получением вплоть до приблизительно 1000000 полинуклеотидов за один высокопараллельный прогон. В некоторых случаях один описанный в настоящем документе кремниевый планшет обеспечивает синтез приблизительно 6100 неидентичных полинуклеотидов. В некоторых случаях каждый из неидентичных полинуклеотидов расположен в пределах кластера. Кластер может содержать 50 - 500 неидентичных полинуклеотидов.

[0141] Описанные в настоящем документе способы обеспечивают синтез библиотеки полинуклеотидов, каждый 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 специфические изменения в последовательности нуклеиновой кислоты можно ввести путем встраивания нуклеотидных изменений в праймеры перекрывающихся полинуклеотидов или полинуклеотидов с тупыми концами. Альтернативно популяция полинуклеотидов может совместно кодировать длинную нуклеиновую кислоту (например, ген) и ее варианты. В этой конфигурации популяцию полинуклеотидов 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.

[0142] Субстраты

[0143] В настоящем документе предусмотрены субстраты, содержащие множество кластеров, причем каждый кластер содержит множество локусов, которые поддерживают прикрепление и синтез полинуклеотидов. 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 полинуклеотидов, или предпочтительно, к популяции идентичных нуклеотидов для синтеза популяции полинуклеотидов. В некоторых случаях полинуклеотид относится к популяции полинуклеотидов, кодирующих одну и ту же последовательность нуклеиновой кислоты. В некоторых случаях поверхность устройства включает в себя одну или множество поверхностей субстрата.

[0144] В настоящем документе предусмотрены структуры, которые могут содержать поверхность, которая поддерживает синтез множества полинуклеотидов, 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; 2,500000; 3000000; 3500000; 4000000; 4500000; 5000000; 10000000 или больше неидентичных полинуклеотидов. 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 полинуклеотидов, кодирующих отдельные последовательности. В некоторых случаях по меньшей мере часть полинуклеотидов содержит идентичные последовательность или может быть синтезирована с идентичной последовательностью.

[0145] В настоящем документе предусмотрены способы и устройства для получения и роста полинуклеотидов, составляющих приблизительно 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 образованного полинуклеотида составляет приблизительно 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200 или 225 оснований в длину. Полинуклеотид может составлять по меньшей мере 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 или 100 оснований в длину. Полинуклеотид 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.

[0146] В некоторых случаях полинуклеотиды синтезируют на отдельных локусах субстрата, причем каждый локус поддерживает синтез популяции полинуклеотидов. В некоторых случаях каждый локус поддерживает синтез популяции полинуклеотидов с последовательностью, отличной от популяции полинуклеотидов, 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; 3000000; 4000000; 5000000; 6000000; 7000000; 8000000; 9000000; 10000000; 12000000; 14000000; 16000000; 18000000; 20000000; 2,5000000; 30000000; 35000000; 40000000; 45000000; 50000000 или 100000000 или больше отдельных локусов. В некоторых случаях устройство содержит приблизительно 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, 1000 или больше локусов. В некоторых случаях каждый кластер включает в себя приблизительно 50-500 локусов. В некоторых случаях каждый кластер включает в себя приблизительно 100-200 локусов. В некоторых случаях каждый кластер включает в себя приблизительно 100-150 локусов. В

некоторых случаях каждый кластер включает в себя приблизительно 109, 121, 130 или 137 локусов. В некоторых случаях каждый кластер включает в себя приблизительно 19, 20, 61, 64 или больше локусов.

[0147] Число отдельных полинуклеотидов, 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 приблизительно 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 приблизительно 10 локусов на мм^2 - приблизительно 500 мм^2 , приблизительно 25 локусов на мм^2 - приблизительно 400 мм^2 , приблизительно 50 локусов на мм^2 - приблизительно 500 мм^2 , приблизительно 100 локусов на мм^2 - приблизительно 500 мм^2 , приблизительно 150 локусов на мм^2 - приблизительно 500 мм^2 , приблизительно 10 локусов на мм^2 - приблизительно 250 мм^2 , приблизительно 50 локусов на мм^2 - приблизительно 250 мм^2 , приблизительно 10 локусов на мм^2 - приблизительно 200 мм^2 или приблизительно 50 локусов на мм^2 - приблизительно 200 мм^2 . В некоторых случаях расстояние от центров двух смежных локусов в пределах кластера составляет приблизительно 10 мкм - приблизительно 500 мкм, приблизительно 10 мкм - приблизительно 200 мкм или приблизительно 10 мкм - приблизительно 100 мкм. В некоторых случаях расстояние от двух центров смежных локусов составляет больше чем приблизительно 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 приблизительно 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 приблизительно 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 приблизительно 0,5 мкм - 100 мкм, приблизительно 0,5 мкм - 50 мкм, приблизительно 10 мкм - 75 мкм или приблизительно 0,5 мкм - 50 мкм.

[0148] В некоторых случаях плотность кластеров в пределах устройства составляет по меньшей мере или приблизительно 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 приблизительно 1 кластер на 10 мм^2 - приблизительно 10 кластеров на 1 мм^2 . В некоторых случаях расстояние от центров двух смежных кластеров составляет меньше чем приблизительно 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 приблизительно 50 мкм - приблизительно 100 мкм, приблизительно 50 мкм - приблизительно 200 мкм, приблизительно 50 мкм - приблизительно 300 мкм, приблизительно 50 мкм - приблизительно 500 мкм и приблизительно 100 мкм - приблизительно 2000 мкм. В некоторых случаях расстояние от центров двух смежных кластеров составляет приблизительно 0,05 мм - приблизительно 50 мм, приблизительно 0,05 мм - приблизительно 10 мм, приблизительно 0,05 мм - приблизительно 5 мм, приблизительно 0,05 мм - приблизительно 4 мм, приблизительно 0,05 мм - приблизительно 3 мм, приблизительно 0,05 мм - приблизительно 2 мм, приблизительно 0,1 мм - 10 мм, приблизительно 0,2 мм - 10 мм, приблизительно 0,3 мм - приблизительно 10 мм, приблизительно 0,4 мм - приблизительно 10 мм, приблизительно 0,5 мм - 10 мм, приблизительно 0,5 мм - приблизительно 5 мм или приблизительно 0,5 мм - приблизительно 2 мм. В некоторых случаях каждый кластер характеризуется диаметром или шириной вдоль одного измерения, составляющими приблизительно 0,5 - 2 мм, приблизительно 0,5 - 1 мм или приблизительно 1 - 2 мм. В некоторых случаях каждый кластер характеризуется диаметром или шириной вдоль одного измерения, составляющими приблизительно 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 приблизительно 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.

[0149] Устройство может характеризоваться приблизительно размером стандартного 96-луночного планшета, например, приблизительно 100 - 200 мм на приблизительно 50 - 150 мм. В некоторых случаях устройство характеризуется диаметром меньше чем или равным приблизительно 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 приблизительно 25 мм - 1000 мм, приблизительно 25 мм - приблизительно 800 мм, приблизительно 25 мм - приблизительно 600 мм, приблизительно 25 мм - приблизительно 500 мм, приблизительно 25 мм - приблизительно 400 мм, приблизительно 25 мм - приблизительно 300 мм или приблизительно 25 мм - приблизительно 200. Неограничивающие примеры размера устройства включают в себя приблизительно 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

приблизительно 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 приблизительно 50 мм - приблизительно 2000 мм, приблизительно 50 мм - приблизительно 1000 мм, приблизительно 100 мм - приблизительно 1000 мм, приблизительно 200 мм - приблизительно 1000 мм или приблизительно 250 мм - приблизительно 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.

[0150] Материалы поверхности

[0151] 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.

[0152] 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 приблизительно 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.

[0153] 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 приблизительно 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.

[0154] 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).

[0155] Структура поверхности

[0156] В настоящем документе предусмотрены устройства, содержащие приподнятые и/или опущенные элементы. Одним из преимуществ таких элементов является увеличение площади поверхности для поддержания синтеза полинуклеотидов. 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 полинуклеотидов. 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 полинуклеотидов синтезируют в множестве локусов кластера.

[0157] В некоторых случаях структура выполнена с возможностью обеспечения контролируемых потоков и путей массопереноса для синтеза полинуклеотидов на поверхности. В некоторых случаях конфигурация устройства обеспечивает контролируемое и равномерное распределение путей массопереноса, времени химического воздействия и/или эффективности промывки во время синтеза полинуклеотидов. В некоторых случаях конфигурация устройства позволяет повысить эффективность вытеснения, например, путем предоставления достаточного объема для выращивания полинуклеотида, так чтобы исключенный объем за счет растущего полинуклеотида 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 полинуклеотид. В некоторых случаях трехмерная структура обеспечивает управляемый поток жидкости для быстрого обмена химическими веществами.

[0158] 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 приблизительно 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% или 100% гена. Ген может варьироваться вплоть до приблизительно 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95% или 100%.

[0159] Неидентичные полинуклеотиды 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 полинуклеотид может кодировать последовательность, составляющую 50%, 60%, 70%, 80%, 85%, 90%, 95% или больше гена. В некоторых случаях полинуклеотид может кодировать последовательность, составляющую 80%, 85%, 90%, 95% или больше гена.

[0160] 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 полинуклеотидов. 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 полинуклеотидов реагентами соседних пятен. В некоторых случаях устройство, такое как синтезатор полинуклеотидов, используют для размещения реагентов в различных местах синтеза полинуклеотидов. Субстраты, характеризующиеся трехмерными признаками, сконфигурированы так, чтобы обеспечить возможность синтеза большого количества полинуклеотидов (например, больше чем приблизительно 10000) с низкой частотой ошибок (например, меньше чем приблизительно 1:500, 1:1000, 1:1500, 1:2000; 1:3000; 1:5000 или 1:10000). В некоторых случаях устройство содержит элементы с плотностью, составляющей приблизительно или больше чем приблизительно 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.

[0161] 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 приблизительно 0,05 мм - приблизительно 50 мм, приблизительно 0,05 мм - приблизительно 10 мм, приблизительно 0,05 мм - приблизительно 5 мм, приблизительно 0,05 мм - приблизительно 4 мм, приблизительно 0,05 мм - приблизительно 3 мм, приблизительно 0,05 мм - приблизительно 2 мм, приблизительно 0,05 мм - приблизительно 1 мм, приблизительно 0,05 мм - приблизительно 0,5 мм, приблизительно 0,05 мм - приблизительно 0,1 мм, приблизительно 0,1 мм - 10 мм, приблизительно 0,2 мм - 10 мм, приблизительно 0,3 мм - приблизительно 10 мм, приблизительно 0,4 мм - приблизительно 10 мм, приблизительно 0,5 мм - 10 мм, приблизительно 0,5 мм -

приблизительно 5 мм или приблизительно 0,5 мм - приблизительно 2 мм. В некоторых случаях ширина лунки, содержащей кластер, составляет приблизительно 0,05 мм - приблизительно 50 мм, приблизительно 0,05 мм - приблизительно 10 мм, приблизительно 0,05 мм - приблизительно 5 мм, приблизительно 0,05 мм - приблизительно 4 мм, приблизительно 0,05 мм - приблизительно 3 мм, приблизительно 0,05 мм - приблизительно 2 мм, приблизительно 0,05 мм - приблизительно 1 мм, приблизительно 0,05 мм - приблизительно 0,5 мм, приблизительно 0,05 мм - приблизительно 0,1 мм, приблизительно 0,1 мм - 10 мм, приблизительно 0,2 мм - 10 мм, приблизительно 0,3 мм - приблизительно 10 мм, приблизительно 0,4 мм - приблизительно 10 мм, приблизительно 0,5 мм - 10 мм, приблизительно 0,5 мм - приблизительно 5 мм или приблизительно 0,5 мм - приблизительно 2 мм. В некоторых случаях ширина кластера составляет меньше чем или приблизительно 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 приблизительно 1,0 - 1,3 мм. В некоторых случаях ширина кластера составляет приблизительно 1,150 мм. В некоторых случаях ширина лунки составляет меньше чем или приблизительно 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 приблизительно 1,0 - 1,3 мм. В некоторых случаях ширина лунки составляет приблизительно 1,150 мм. В некоторых случаях ширина кластера составляет приблизительно 0,08 мм. В некоторых случаях ширина лунки составляет приблизительно 0,08 мм. Ширина кластера может относиться к кластерам в пределах двумерного или трехмерного субстрата.

[0162] В некоторых случаях высота лунки составляет приблизительно 20 мкм - приблизительно 1000 мкм, приблизительно 50 мкм - приблизительно 1000 мкм, приблизительно 100 мкм - приблизительно 1000 мкм, приблизительно 200 мкм - приблизительно 1000 мкм, приблизительно 300 мкм - приблизительно 1000 мкм, приблизительно 400 мкм - приблизительно 1000 мкм или приблизительно 500 мкм - приблизительно 1000 мкм. В некоторых случаях высота лунки составляет меньше чем приблизительно 1000 мкм, меньше чем приблизительно 900 мкм, меньше чем приблизительно 800 мкм, меньше чем приблизительно 700 мкм или меньше чем приблизительно 600 мкм.

[0163] В некоторых случаях устройство содержит множество каналов, соответствующих множеству локусов в пределах кластера, причем высота или глубина канала составляет приблизительно 5 мкм - приблизительно 500 мкм, приблизительно 5 мкм - приблизительно 400 мкм, приблизительно 5 мкм - приблизительно 300 мкм, приблизительно 5 мкм - приблизительно 200 мкм, приблизительно 5 мкм - приблизительно

100 мкм, приблизительно 5 мкм - приблизительно 50 мкм или приблизительно 10 мкм - приблизительно 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.

[0164] 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 приблизительно 1 мкм - приблизительно 1000 мкм, приблизительно 1 мкм - приблизительно 500 мкм, приблизительно 1 мкм - приблизительно 200 мкм, приблизительно 1 мкм - приблизительно 100 мкм, приблизительно 5 мкм - приблизительно 100 мкм или приблизительно 10 мкм - приблизительно 100 мкм, например, приблизительно 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 приблизительно 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 приблизительно 1 мкм - приблизительно 500 мкм, приблизительно 1 мкм - приблизительно 200 мкм, приблизительно 1 мкм - приблизительно 100 мкм, приблизительно 5 мкм - приблизительно 200 мкм, приблизительно 5 мкм - приблизительно 100 мкм, приблизительно 5 мкм - приблизительно 50 мкм или приблизительно 5 мкм - приблизительно 30 мкм, например, приблизительно 20 мкм.

[0165] Модификации поверхности

[0166] 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.

[0167] 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, осажденного на поверхность, поддерживает локализацию капель на поверхности. В зависимости от выбранного расположения компоновки близость локусов и/или площадь контакта жидкости в локусах могут изменяться.

[0168] 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.

[0169] 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 полинуклеотида. Поверхностная энергия, или гидрофобность, поверхности является фактором, определяющим аффинность нуклеотида в отношении прикрепления к поверхности. В некоторых случаях способ функционализации устройства может предусматривать следующее: (а) 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.

[0170] В некоторых случаях органофункциональная молекула алкоксисилана включает в себя диметилхлороктодецилсилан, метилдихлороктодецилсилан, трихлороктодецилсилан, триметилорктодецилсилан, триэтилоорктодецилсилан 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.

[0171] 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.

[0172] 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.

[0173] В настоящем документе предусмотрены устройства, которые могут содержать компоновку средств, способных связываться с нуклеозидом. 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-триэтоксисилилпропил)-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.

[0174] 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.

[0175] 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.

[0176] Синтез полинуклеотидов

[0177] Способы настоящего раскрытия для синтеза полинуклеотидов могут предусматривать процессы, включая в себя фосфорамидитную химию. В некоторых случаях синтез полинуклеотидов включает в себя сочетание основания с фосфорамидитом. Синтез полинуклеотидов 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 полинуклеотидов может включать в себя кэпирование непрореагировавших сайтов. В некоторых случаях кэпирование является необязательным. Синтез полинуклеотидов может также включать в себя окисления, или стадию окисления,

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

[0178] Синтез полинуклеотидов с использованием фосфорамидитного способа может включать в себя последующее добавление строительного блока фосфорамидита (например, нуклеозидного фосфорамидита) к растущей полинуклеотидной цепи для образования фосфиттриэфирной связи. Фосфорамидитный синтез полинуклеотидов происходит в направлении 3' - 5'. Фосфорамидитный синтез полинуклеотидов 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 полинуклеотидов, используемый в настоящем документе, предусматривает 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).

[0179] После сочетания способы фосфорамидитного синтеза полинуклеотидов необязательно предусматривают стадию кэпирования. На стадии кэпирования растущий полинуклеотид обрабатывают кэпирующим средством. Стадия кэпирования является

применимой для блокирования непрореагировавших связанных с субстратом 5'-ОН-групп после сочетания из дальнейшего удлинения цепи, предотвращая образование полинуклеотидов 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 полинуклеотидов снижает частоту ошибок по сравнению с синтезом без кэпирования. В качестве примера стадия кэпирования предусматривает обработку связанного с субстратом полинуклеотида смесью уксусного ангидрида и 1-метилимидазола. После стадии кэпирования устройство дополнительно промывают.

[0180] 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 предусматривает, что фосфиттриэфир окисляется до тетракоординированного фосфаттриэфира, защищенного предшественника встречающейся в природе фосфатдиэфирной межнуклеозидной связи. В некоторых случаях окисление растущего полинуклеотида 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 растущий полинуклеотид 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).

[0181] Для того чтобы последующий цикл включения нуклеозидов происходил посредством сочетания, защищенный 5'-конец связанного с устройством растущего полинуклеотида 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 связанный с устройством полинуклеотид отмывают после деблокирования. В некоторых случаях эффективная промывка после деблокирования способствует получению синтезированных полинуклеотидов, характеризующихся низкой частотой ошибок.

[0182] Способы синтеза полинуклеотидов, как правило, предусматривают 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 сульфурование. В некоторых случаях одна или несколько стадий промывки предшествуют или следуют после одной или всех стадий.

[0183] Способы синтеза полинуклеотидов на основе фосфорамидита предусматривают серию химических стадий. В некоторых случаях одна или несколько стадий способа синтеза предусматривают циклическое использование реагентов, где одна или несколько стадий способа предусматривают нанесение на устройство реагента, примененного на стадии. Например, реагенты циклически проходят серию стадий осаждения жидкости и вакуумной сушки. Для субстратов, содержащих трехмерные элементы, такие как лунки, микроячейки, каналы и тому подобное, реагенты необязательно пропускают через одну или несколько участков устройства через лунки и/или каналы.

[0184] 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 приблизительно 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 полинуклеотидов, 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 приблизительно 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 приблизительно 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 приблизительно 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.

[0185] Предусмотренные в настоящем документе способы синтеза полинуклеотидов 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, уридиновые строительные блоки или их

аналоги/модифицированные версии. В некоторых случаях библиотеки полинуклеотидов синтезируют в параллельном режиме на субстрате. Например, устройство, содержащее приблизительно или по меньшей мере приблизительно 100; 1000; 10000; 30000; 75000; 100000; 1000000; 2000000; 3000000; 4000000 или 5000000 выделенных локусов, способно поддерживать синтез по меньшей мере такого же числа отдельных полинуклеотидов, причем полинуклеотид, кодирующий отдельную последовательность, синтезируют на выделенном локусе. В некоторых случаях библиотеку полинуклеотидов синтезируют на устройстве с низкими частотами ошибок, описанном в настоящем документе, меньше чем приблизительно 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 полинуклеотидов, синтезированной с низкой частотой ошибок с использованием субстратов и способов, описанных в настоящем документе, получают меньше чем приблизительно 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.

[0186] В некоторых случаях описанные в настоящем документе способы обеспечивают создание библиотеки полинуклеотидов, содержащей варианты полинуклеотиды, различающиеся по множеству сайтов кодонов. В некоторых случаях полинуклеотид 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.

[0187] 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.

[0188] В некоторых случаях полинуклеотид 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 полинуклеотид может содержать множественные сайты вариантных сайтов кодонов, причем ни один из вариантных сайтов кодонов не является смежным друг с другом. В некоторых случаях полинуклеотид 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.

[0189] Ссылаясь на фигуры, на фиг. 5 проиллюстрирован пример рабочего процесса синтеза нуклеиновых кислот (например, генов) из более коротких полинуклеотидов. Рабочий процесс, как правило, делят на следующие стадии: (1) синтез *de novo* библиотеки одноцепочечных полинуклеотидов, (2) объединение полинуклеотидов 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.

[0190] После того, как выбраны большие полинуклеотиды для создания, предварительно определенную библиотеку полинуклеотидов конструируют для синтеза *de novo*. Известны различные подходящие способы получения полинуклеотидных матриц высокой плотности. В примере рабочего процесса предусмотрен поверхностный слой **501** устройства. В этом примере химический состав поверхности изменяют с целью улучшения процесса синтеза полинуклеотидов. 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 микроячейки, 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.

[0191] Получение полинуклеотидных матриц *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 полинуклеотидов, предназначено для поэтапного высвобождения реагентов таким образом, что множественные полинуклеотиды удлиняются в параллельном режиме по одному остатку за раз, образуя олигомеры с предварительно определенной последовательностью нуклеиновой кислоты **502**. В некоторых случаях полинуклеотиды отщепляют от поверхности на этой стадии. Расщепление включает в себя расщепление газом, например, аммиаком или метиламином.

[0192] Созданные библиотеки полинуклеотидов 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 полинуклеотидов **503**. До или после герметизации **504** полинуклеотидов в реагент добавляют для высвобождения

полинуклеотидов из субстрата. В иллюстративном рабочем процессе полинуклеотиды высвобождаются после герметизации нанореактора **505**. После высвобождения фрагменты одноцепочечных полинуклеотидов гибридизуются, чтобы охватить всю длинную последовательность ДНК. Частичная гибридизация **505** возможна, потому что каждый синтезированный полинуклеотид сконструирован так, что содержит небольшую часть, перекрывающуюся по меньшей мере с одним другим полинуклеотидом в популяции.

[0193] После гибридизации начинают реакцию ПЦР. Во время полимеразных циклов полинуклеотиды отжигаются с комплементарными фрагментами и пропуски заполняются полимеразой. Каждый цикл увеличивает длину различных фрагментов случайным образом в зависимости от того, какие полинуклеотиды находят друг друга. Комплементарность между фрагментами позволяет сформировать полный большой диапазон двухцепочечной ДНК **506**.

[0194] После завершения ПЦР нанореактор отделяют от устройства **507** и располагают для взаимодействия с устройством, содержащим праймеры для ПЦР **508**. После герметизации нанореактор подвергают ПЦР **309** и амплифицируют более крупные нуклеиновые кислоты. После ПЦР **510** нанокамеру открывают **511**, добавляют реагенты для исправления ошибок **512**, камеру герметизируют **513**, и происходит реакция исправления ошибок для удаления несовпадающих пар оснований и/или цепей с плохой комплементарностью из двухцепочечной продуктов ПЦР-амплификации **514**. Нанореактор открывают и отделяют **515**. Продукт с исправленными ошибками затем подвергают дополнительным стадиям обработки, таким как ПЦР и молекулярное штрих-кодирование, а затем упаковывают **522** для отправки **523**.

[0195] 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 **516**, герметизацию пластины с камерой, содержащей продукт амплификации с исправленными ошибками **517**, и выполнение дополнительного цикла амплификации **518**. Нанореактор открывают **519** и продукты объединяют **520** и секвенируют **521**. После того как приемлемое определение контроля качества выполнено, упакованный продукт **522** получает одобрение для отправки **523**.

[0196] В некоторых случаях нуклеиновую кислоту, созданную с помощью такого рабочего процесса, как представлен на фиг. **5**, 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 полинуклеотидов, предназначено для поэтапного высвобождения реагентов таким образом, чтобы множественные полинуклеотиды удлинялись в параллельном режиме по одному остатку за раз, образуя олигомеры с предварительно определенной последовательностью нуклеиновой кислоты **502**.

[0197] Большие библиотеки полинуклеотидов, характеризующиеся низким показателями частоты ошибок

[0198] Средние показатели частоты ошибок для полинуклеотидов, 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 полинуклеотидов, 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 полинуклеотидов, синтезированных в пределах библиотеки с использованием предусмотренных систем и способов, составляют меньше чем 1/1000.

[0199] В некоторых случаях показатели суммарной частоты ошибок для полинуклеотидов, 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 полинуклеотидов, 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 полинуклеотидов, синтезированных в пределах библиотеки с использованием предусмотренных систем и способов, составляют меньше чем 1/1000.

[0200] В некоторых случаях фермент для исправления ошибок можно использовать для полинуклеотидов, синтезированных в пределах библиотеки с использованием предусмотренных систем и способов. В некоторых случаях показатели суммарной частоты ошибок для полинуклеотидов 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 полинуклеотидов, 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 полинуклеотидов, синтезированных в пределах библиотеки с использованием предусмотренных систем и способов, могут составлять меньше чем 1/1000.

[0201] Частота ошибок может ограничивать ценность синтеза генов для производства библиотек вариантов генов. С частотой ошибок, составляющей 1/300, приблизительно 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.

[0202] 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.

[0203] Вычислительные системы

[0204] 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.

[0205] Вычислительную систему **600**, показанную на фиг. **6** можно понимать как логическое устройство, которое может считывать инструкции с носителя **611** и/или сетевого порта **605**, который необязательно может быть подключен к серверу **609** с фиксированным носителем **612**. Такая система, как показана на фиг. **6**, может включать в себя ЦП (центральный процессор) **601**, дисководы **603**, дополнительные устройства ввода, такие как клавиатура **615** и/или мышь **616** и дополнительный монитор **607**. 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 **622**, как показано на фиг. **6**.

[0206] На фиг. **7** представлена блок-схема, иллюстрирующая первый пример архитектуры вычислительной системы **700**, которую можно использовать в связи с иллюстративными примерами согласно настоящему изобретению. Как показано на фиг. **7**, иллюстративная вычислительная система может включать в себя процессор **702** 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.

[0207] Как показано на фиг. **7**, высокоскоростной кэш **704** может быть подключен или встроен в процессор **702** для обеспечения высокоскоростной памяти для команд или данных, которые недавно или часто используются процессором **702**. Процессор **702** подключен к северному мосту **706** процессорной шиной **708**. Северный мост **706** соединен

с оперативной памятью (ОЗУ) **710** шиной памяти **712** и управляет доступом к ОЗУ **710** процессором **702**. Северный мост **706** также соединен с южным мост **714** с помощью шины чипсета **716**. Южный мост **714**, в свою очередь, соединен с периферийной шиной **718**. 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 **718**. В некоторых альтернативных архитектурах функциональность северного моста может быть встроена в процессор вместо использования отдельного чипа северного моста. В некоторых случаях система **700** может включать в себя карту акселератора **722**, присоединенную к периферийной шине **718**. 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.

[0208] Программное обеспечение и данные хранятся во внешнем хранилище **724** и могут быть загружены в ОЗУ **710** и/или в кэш **704** для использования процессором. Система **700** 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 **700** также включает в себя сетевые интерфейсные карты (NIC) **720** и **721**, 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.

[0209] Фиг. 8 представляет собой диаграмму, на которой показана сеть **800** с множеством вычислительных систем **802a** и **802b**, множеством сотовых телефонов и карманных персональных компьютеров **802c** и сетевым хранилищем (NAS) **804a** и **804b**. В иллюстративных примерах системы **802a**, **802b** и **802c** могут управлять хранением данных и оптимизировать доступ к данным для данных, хранящихся в сетевом хранилище (NAS) **804a** и **804b**. Математическую модель можно использовать для данных и оценивать с использованием распределенной параллельной обработки по вычислительным системам **802a** и **802b**, а также по системам сотовых телефонов и карманных персональных

компьютеров **802c**. Вычислительные системы **802a** и **802b** и системы сотовых телефонов и карманных персональных компьютеров **802c** также могут обеспечивать параллельную обработку для адаптивной реструктуризации данных, хранящихся в сетевом хранилище (NAS) **804a** и **804b**. На фиг. 8 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. «Лезвия» 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.

[0210] На фиг. 9 представлена блок-схема многопроцессорной вычислительной системы **900**, использующей совместное пространство памяти виртуальных адресов в соответствии с иллюстративным примером. Система включает в себя множество процессоров **902a-f**, которые могут обращаться к подсистеме совместной памяти **904**. Система включает в себя множество программируемых аппаратных процессоров алгоритмов памяти **906a-f** в подсистеме памяти **904**. Каждый MAP **906a-f** может содержать память **908a-f** и один или несколько программируемых пользователем вентильных матриц (FPGA) **910a-f**. MAP образует собой конфигурируемый функциональный блок, и конкретные алгоритмы или части алгоритмов могут быть предоставлены FPGA **910a-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 **908a-f**, что позволяет ей выполнять задачи независимо и асинхронно от соответствующего микропроцессора **902a-f**. В этой конфигурации MAP может передавать результаты непосредственно в другой MAP для конвейерной обработки и параллельного выполнения алгоритмов.

[0211] 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.

[0212] 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. 9, 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 722, показанная на фиг. 7.

[0213] Дополнительные способы и композиции

[0214] В настоящем документе предусмотрены способы создания библиотеки полинуклеотидов, предусматривающие следующее:

получение предварительно определенных последовательностей, кодирующих по меньшей мере приблизительно 5000 неидентичных полинуклеотидов; получение структуры, характеризующейся поверхностью, причем поверхность содержит множество кластеров; синтезирование по меньшей мере приблизительно 5000 неидентичных полинуклеотидов, причем каждый из по меньшей мере 5000 неидентичных полинуклеотидов удлиняется из различного локуса; и амплификация по меньшей мере 5000 неидентичных полинуклеотидов с образованием библиотеки полинуклеотидов, причем больше чем приблизительно 80% по меньшей мере из 5000 неидентичных полинуклеотидов представлены в количестве в пределах по меньшей мере приблизительно 2-кратного по отношению к среднему представлению для библиотеки полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых больше чем приблизительно 80% по меньшей мере из 5000 неидентичных полинуклеотидов представлены в количестве

в пределах по меньшей мере приблизительно 1,5-кратного по отношению к среднему представлению для библиотеки полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых больше чем приблизительно 90% по меньшей мере из 5000 неидентичных полинуклеотидов представлены в количестве в пределах по меньшей мере приблизительно 2-кратного по отношению к среднему представлению для библиотеки полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых больше чем приблизительно 90% по меньшей мере из 5000 неидентичных полинуклеотидов представлены в количестве в пределах по меньшей мере приблизительно 1,5-кратного по отношению к среднему представлению для библиотеки полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека полинуклеотидов содержит меньшее количество выпадений по сравнению с продуктом амплификации из способа с использованием структуры, характеризующейся поверхностью некластеризованных локусов. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим по меньшей мере приблизительно 10%. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим не более чем 95%. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим приблизительно 10% - приблизительно 95%. Дополнительно в настоящем документе предусмотрены способы, при которых больше чем 30% по меньшей мере из приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим 10% - 30% или 70 - 90%. Дополнительно в настоящем документе предусмотрены способы, при которых меньше чем приблизительно 15% по меньшей мере из приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим 10% - 30% или 60 - 90%. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением повторяющейся последовательности, составляющим по меньшей мере приблизительно 10%. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением повторяющейся последовательности, составляющим не более

чем 95%. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением повторяющейся последовательности, составляющим приблизительно 10% - приблизительно 95%. Дополнительно в настоящем документе предусмотрены способы, при которых больше чем 30% по меньшей мере из приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением повторяющейся последовательности составляющим, 10% - 30% или 70 - 90%. Дополнительно в настоящем документе предусмотрены способы, при которых меньше чем приблизительно 15% по меньшей мере из приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением повторяющейся последовательности составляющим, 10% - 30% или 60 - 90%. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением вторичной структуры, составляющим по меньшей мере приблизительно 10%. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением вторичной структуры, составляющим не более чем 95%. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением вторичной структуры, составляющим приблизительно 10% - приблизительно 95%. Дополнительно в настоящем документе предусмотрены способы, при которых больше чем 30% по меньшей мере из приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением вторичной структуры, составляющим 10% - 30% или 70 - 90%. Дополнительно в настоящем документе предусмотрены способы, при которых меньше чем приблизительно 15% по меньшей мере из приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением вторичной структуры, составляющим 10% - 30% или 60 - 90%. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека полинуклеотидов кодирует библиотеку вариантов. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере 5000 неидентичных полинуклеотидов библиотеки кодируют по меньшей мере один ген. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере 5000 неидентичных полинуклеотидов библиотеки кодируют по меньшей мере 50

генов. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека полинуклеотидов кодирует по меньшей мере один ген. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека полинуклеотидов кодирует по меньшей мере часть антитела, фермента или пептида. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека полинуклеотидов характеризуется суммарной частотой ошибок, составляющей меньше чем 1 на 500 оснований по сравнению с предварительно определенными последовательностями без исправления ошибок. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека полинуклеотидов характеризуется суммарной частотой ошибок, составляющей меньше чем 1 на 1000 оснований по сравнению с предварительно определенными последовательностями без исправления ошибок. Дополнительно в настоящем документе предусмотрены способы, при которых предварительно определенные последовательности кодируют по меньшей мере 700000 неидентичных полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых каждый кластер содержит 50 - приблизительно 500 локусов для синтеза полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых каждый кластер содержит вплоть до приблизительно 500 локусов для синтеза полинуклеотидов.

[0215] В настоящем документе предусмотрены способы создания библиотеки полинуклеотидов, предусматривающие следующее: получение предварительно определенных последовательностей, кодирующих по меньшей мере приблизительно 5000 неидентичных полинуклеотидов; получение структуры, характеризующейся поверхностью, причем поверхность содержит множество кластеров; синтезирование по меньшей мере приблизительно 5000 неидентичных полинуклеотидов, причем каждый из по меньшей мере 5000 неидентичных полинуклеотидов удлиняется из различного локуса; и амплификация по меньшей мере 5000 неидентичных полинуклеотидов с образованием библиотеки полинуклеотидов, причем библиотека полинуклеотидов характеризуется частотой правильных последовательностей, составляющей больше чем 75% после реакции амплификации. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека полинуклеотидов характеризуется частотой правильных последовательностей, составляющей больше чем 85% после реакции амплификации. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим по меньшей мере приблизительно 10%. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей

мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим не более чем 95%. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим приблизительно 10% - приблизительно 95%. Дополнительно в настоящем документе предусмотрены способы, при которых больше чем 30% по меньшей мере из приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим 10% - 30% или 70 - 90%. Дополнительно в настоящем документе предусмотрены способы, при которых меньше чем приблизительно 15% по меньшей мере из приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим 10% - 30% или 60 - 90%. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека полинуклеотидов содержит меньшее количество выпадений по сравнению с продуктом амплификации из способа с использованием структуры, характеризующейся поверхностью некластеризованных локусов. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека полинуклеотидов кодирует библиотеку вариантов. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека полинуклеотидов кодирует по меньшей мере один ген. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека полинуклеотидов кодирует по меньшей мере часть антитела, фермента или пептида. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека полинуклеотидов характеризуется суммарной частотой ошибок, составляющей меньше чем 1 на 500 оснований по сравнению с предварительно определенными последовательностями без исправления ошибок. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека полинуклеотидов характеризуется суммарной частотой ошибок, составляющей меньше чем 1 на 1000 оснований по сравнению с предварительно определенными последовательностями без исправления ошибок. Дополнительно в настоящем документе предусмотрены способы, при которых предварительно определенные последовательности кодируют по меньшей мере 700000 неидентичных полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере 5000 неидентичных полинуклеотидов кодируют по меньшей мере один ген. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере 5000 неидентичных полинуклеотидов кодируют по меньшей мере 50 генов. Дополнительно в настоящем документе предусмотрены способы, при которых каждый кластер содержит 50 -

приблизительно 500 локусов для синтеза полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых каждый кластер содержит вплоть до приблизительно 500 локусов для синтеза полинуклеотидов.

[0216] В настоящем документе предусмотрены библиотеки нуклеиновых кислот, содержащие по меньшей мере 5000 неидентичных полинуклеотидов, причем по меньшей мере 5000 неидентичных полинуклеотидов представляют собой продукты амплификации синтезированных полинуклеотидов, и причем больше чем приблизительно 80% по меньшей мере из 5000 неидентичных полинуклеотидов представлены в количестве в пределах по меньшей мере приблизительно 2-кратного по отношению к среднему представлению для библиотек нуклеиновых кислот. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых больше чем приблизительно 80% по меньшей мере из 5000 неидентичных полинуклеотидов представлены в количестве в пределах по меньшей мере приблизительно 1,5-кратной представленности для библиотек нуклеиновых кислот. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых больше чем приблизительно 90% по меньшей мере из 5000 неидентичных полинуклеотидов представлены в количестве в пределах по меньшей мере приблизительно 2-кратного по отношению к среднему представлению для библиотек нуклеиновых кислот. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых больше чем приблизительно 90% по меньшей мере из 5000 неидентичных полинуклеотидов представлены в количестве в пределах по меньшей мере приблизительно 1,5-кратной представленности для библиотек нуклеиновых кислот. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, причем библиотеки нуклеиновых кислот содержат меньшее количество выпадений по сравнению с продуктом амплификации из способа с использованием структуры, характеризующейся поверхностью некластеризованных локусов. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим по меньшей мере приблизительно 10%. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим не более чем 95%. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим приблизительно 10% - приблизительно 95%. Дополнительно в настоящем документе предусмотрены библиотеки

полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением вторичной структуры, составляющим приблизительно 10% - приблизительно 95%. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых больше чем 30% по меньшей мере из приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением вторичной структуры, составляющим 10% - 30% или 70 - 90%. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых меньше чем приблизительно 15% по меньшей мере из приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением вторичной структуры, составляющим 10% - 30% или 60 - 90%. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, причем библиотека полинуклеотидов кодирует библиотеку вариантов. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых по меньшей мере 5000 неидентичных полинуклеотидов кодируют по меньшей мере один ген. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых по меньшей мере 5000 неидентичных полинуклеотидов кодируют по меньшей мере 50 генов. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, причем библиотека полинуклеотидов кодирует по меньшей мере часть антитела, фермента или пептида. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых предварительно определенные последовательности кодируют по меньшей мере 700000 неидентичных полинуклеотидов.

[0217] В настоящем документе предусмотрены библиотеки нуклеиновых кислот, содержащие по меньшей мере 5000 неидентичных полинуклеотидов, где содержание GC является контролируемым, и причем библиотеки обеспечивают частоту правильных последовательностей, составляющую больше чем 75% после реакции амплификации. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, характеризующиеся частотой правильных последовательностей, составляющей больше чем 85% после реакции амплификации. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим по меньшей мере приблизительно 10%. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим не более чем 95%. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых по

меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим приблизительно 10% - приблизительно 95%. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых больше чем 30% по меньшей мере из приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим 10% - 30% или 70 - 90%. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых меньше чем приблизительно 15% по меньшей мере из приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим 10% - 30% или 60 - 90%.

[0218] В настоящем документе предусмотрены библиотеки нуклеиновых кислот, содержащие по меньшей мере 5000 неидентичных полинуклеотидов, где содержание повторяющейся последовательности является контролируемым, и причем библиотеки обеспечивают частоту правильных последовательностей, составляющую больше чем 75% после реакции амплификации. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, причем библиотека полинуклеотидов характеризуется частотой правильных последовательностей, составляющей больше чем 85% после реакции амплификации. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением повторяющейся последовательности, составляющим по меньшей мере приблизительно 10%. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением повторяющейся последовательности, составляющим не более чем 95%. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением повторяющейся последовательности, составляющим приблизительно 10% - приблизительно 95%. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых больше чем 30% по меньшей мере из приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением повторяющейся последовательности, составляющим, 10% - 30% или 70 - 90%. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых меньше чем приблизительно 15% по меньшей мере из приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся

процентным отношением повторяющейся последовательности составляющим, 10% - 30% или 60 - 90%.

[0219] В настоящем документе предусмотрены библиотеки нуклеиновых кислот, содержащие по меньшей мере 5000 неидентичных полинуклеотидов, где содержание вторичной структуры, кодируемое по меньшей мере 5000 неидентичных полинуклеотидов, является предварительно выбранным, и причем библиотеки обеспечивают частоту правильных последовательностей, составляющую больше чем 75% после реакции амплификации. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, причем библиотеки нуклеиновых кислот характеризуются частотой правильных последовательностей, составляющей больше чем 85% после реакции амплификации. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением вторичной структуры, составляющим по меньшей мере приблизительно 10%. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением вторичной структуры, составляющим не более чем 95%. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых по меньшей мере приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением вторичной структуры, составляющим приблизительно 10% - приблизительно 95%. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых больше чем 30% по меньшей мере из приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением вторичной структуры, составляющим 10% - 30% или 70 - 90%. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых меньше чем приблизительно 15% по меньшей мере из приблизительно 5000 неидентичных полинуклеотидов содержат полинуклеотиды, характеризующиеся процентным отношением повторяющейся последовательности составляющим, 10% - 30% или 60 - 90%. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, кодирующие библиотеку вариантов. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, кодирующие по меньшей мере один ген. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, кодирующие по меньшей мере часть антитела, фермента или пептида. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот,

характеризующиеся суммарной частотой ошибок, составляющей меньше чем 1 на 500 оснований по сравнению с предварительно определенными последовательностями без исправления ошибок. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, характеризующиеся суммарной частотой ошибок, составляющей меньше чем 1 на 1000 оснований по сравнению с предварительно определенными последовательностями без исправления ошибок. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых предварительно определенные последовательности кодируют по меньшей мере 700000 неидентичных полинуклеотидов. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых по меньшей мере 5000 неидентичных полинуклеотидов кодируют по меньшей мере один ген. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых по меньшей мере 5000 неидентичных полинуклеотидов кодируют по меньшей мере 50 генов.

[0220] В настоящем документе предусмотрены способы амплификации библиотеки полинуклеотидов, предусматривающие следующее: получение распределения амплификации по меньшей мере для 5000 неидентичных полинуклеотидов; кластеризация по меньшей мере 5000 неидентичных полинуклеотидов из распределения амплификации на два или больше бинов на основании по меньшей мере одного признака последовательности; и коррекция представления для синтеза каждого из неидентичных полинуклеотидов на основании частоты числа по меньшей мере 5000 неидентичных полинуклеотидов в каждом из двух или больше бинов для создания библиотеки полинуклеотидов, характеризующейся предварительно выбранным представлением; синтезирование библиотеки полинуклеотидов, характеризующейся предварительно выбранным представлением; и амплификация библиотеки полинуклеотидов, характеризующейся предварительно выбранным представлением. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере один признак последовательности представляет собой процентное содержание GC. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере один признак последовательности представляет собой процентное содержание повторяющейся последовательности. Дополнительно в настоящем документе предусмотрены способы, при которых по меньшей мере один признак последовательности представляет собой процентное содержание вторичной структуры. Дополнительно в настоящем документе предусмотрены способы, при которых повторяющиеся последовательности содержат 3 или больше аденинов. Дополнительно в настоящем документе предусмотрены способы, при которых повторяющиеся последовательности находятся на одном или обоих концах полинуклеотида.

Дополнительно в настоящем документе предусмотрены способы, при которых указанные полинуклеотиды подвергнуты кластеризации в бины на основании аффинности одной или нескольких полинуклеотидных последовательностей в отношении связывания с целевой последовательностью. Дополнительно в настоящем документе предусмотрены способы, при которых число последовательностей в нижних 30% бинов характеризуется по меньшей мере на 50% большим представлением в последующем применении после коррекции по сравнению с числом последовательностей в нижних 30% бинов до коррекции. Дополнительно в настоящем документе предусмотрены способы, при которых число последовательностей в верхних 30% бинов характеризуется по меньшей мере на 50% большим представлением в последующем применении после коррекции по сравнению с числом последовательностей в верхних 30% бинов до коррекции. Дополнительно в настоящем документе предусмотрены способы, при которых указанное распределение амплификации получают эмпирически. Дополнительно в настоящем документе предусмотрены способы, при которых указанное распределение амплификации получают посредством алгоритма прогнозирования. Коррекция в некоторых случаях включает в себя контроль стехиометрии полинуклеотидов в библиотеке.

[0221] В настоящем документе предусмотрены библиотеки нуклеиновых кислот, содержащие по меньшей мере 100000 неидентичных полинуклеотидов, причем каждый неидентичных полинуклеотид кодирует по меньшей мере одну различную экзонную последовательность, и причем каждый из по меньшей мере приблизительно 80% по меньшей мере из 100000 неидентичных полинуклеотидов присутствует в библиотеке полинуклеотидов в количестве в пределах 2-кратного от средней частоты для каждого из неидентичных полинуклеотидов в библиотеке. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, причем библиотеки нуклеиновых кислот представляют собой библиотеки ампликонов, и причем каждый из по меньшей мере приблизительно 80% множества неидентичных полинуклеотидов присутствует в библиотеках ампликонов в количестве в пределах 2-кратного от средней частоты для каждого из неидентичных полинуклеотидов в библиотеках. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых секвенирование библиотеки при вплоть до 55-кратной теоретической глубине прочтения дает в результате по меньшей мере 90% оснований, характеризующихся по меньшей мере 30-кратной глубиной прочтения. Дополнительно в настоящем документе предусмотрены библиотеки нуклеиновых кислот, в которых секвенирование библиотеки при вплоть до 55-кратной теоретической глубине прочтения дает в результате по меньшей мере 98% оснований, характеризующихся по меньшей мере 10-кратной глубиной прочтения.

[0222] В настоящем документе предусмотрены способы синтеза библиотеки полинуклеотидов, предусматривающие следующее: (a) получение предварительно определенных последовательностей для по меньшей мере 100000 неидентичных полинуклеотидов, причем каждый неидентичный полинуклеотид кодирует одну или нескольких частей геномной ДНК; (b) синтезирование по меньшей мере 100000 неидентичных полинуклеотидов; и (c) амплификация по меньшей мере 100000 неидентичных полинуклеотидов для создания библиотеки полинуклеотидов, причем по меньшей мере приблизительно 75% полинуклеотидов в библиотеке не содержат ошибок по сравнению с предварительно определенными последовательностями для по меньшей мере 100000 неидентичных полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека полинуклеотидов представляет собой библиотеку ампликонов, и причем каждый из по меньшей мере приблизительно 80% множества неидентичных полинуклеотидов присутствует в библиотеке ампликонов в количестве в пределах 2-кратного от средней частоты для каждого из неидентичных полинуклеотидов в библиотеке. Дополнительно в настоящем документе предусмотрены способы, при которых каждый неидентичный полинуклеотид кодирует один или несколько экзонов. Дополнительно в настоящем документе предусмотрены способы, при которых каждый из амплифицированных неидентичных полинуклеотидов содержит по меньшей мере один молекулярный маркер.

[0223] В настоящем документе предусмотрены способы синтеза библиотеки полинуклеотидов, предусматривающие следующие: (a) амплификация первой библиотеки по меньшей мере из 2000 неидентичных полинуклеотидов; (b) идентификация распределения последовательностей в первой библиотеке как функции одного или нескольких признаков последовательности; и (c) изменение относительного отношения последовательностей в первой библиотеке на основании распределения последовательности для создания второй библиотеки, так что не больше чем 2,5-кратный сэмплинг второй библиотеки дает в результате по меньшей мере 80% покрытия секвенирования. Дополнительно в настоящем документе предусмотрены способы, при которых один или несколько признаков последовательности содержат процентное содержание GC. Дополнительно в настоящем документе предусмотрены способы, при которых один или несколько признаков последовательности содержит процентное содержание повторяющейся последовательности. Дополнительно в настоящем документе предусмотрены способы, при которых один или несколько признаков последовательности содержит процентное содержание вторичной структуры. Дополнительно в настоящем документе предусмотрены способы, при которых один или несколько признаков

последовательности содержит покрытие секвенирования. Дополнительно в настоящем документе предусмотрены способы, при которых не больше чем 1,7-кратный сэмплинг дает в результате по меньшей мере 80% покрытия секвенирования. Дополнительно в настоящем документе предусмотрены способы, при которых не больше чем 2,5-кратный сэмплинг дает в результате по меньшей мере 90% покрытия секвенирования. Дополнительно в настоящем документе предусмотрены способы, при которых способ дополнительно предусматривает синтезирование второй библиотеки. Дополнительно в настоящем документе предусмотрены способы, при которых способ дополнительно предусматривает амплификацию второй библиотеки. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека содержит по меньшей мере 5000 полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека содержит по меньшей мере 10000 полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека содержит по меньшей мере 30000 полинуклеотидов.

[0224] В настоящем документе предусмотрены способы целевого обогащения, предусматривающие следующее: приведение в контакт библиотеки по меньшей мере из 2000 неидентичных, двухцепочечных полинуклеотидов с популяцией полинуклеотидов образца, содержащих целевые нуклеиновые кислоты, причем каждый из по меньшей мере 2000 неидентичных полинуклеотидов содержит следующее: (с 5' до 3'): первую нецелевую последовательность и вторую нецелевую последовательность; и вставочную последовательность, которая комплементарна одной или нескольким целевым последовательностям нуклеиновой кислоты; захват целевых последовательностей нуклеиновой кислоты, которые гибридизуются с одним или несколькими из по меньшей мере 2000 неидентичных полинуклеотидов на твердой подложке; и высвобождение захваченных целевых нуклеиновых кислот для создания обогащенной библиотеки целевых полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых каждый полинуклеотид дополнительно содержит по меньшей мере один молекулярный маркер. Дополнительно в настоящем документе предусмотрены способы, при которых каждая нецелевая последовательность дополнительно содержит сайт связывания праймера. Дополнительно в настоящем документе предусмотрены способы, при которых первая нецелевая последовательность расположена на 5' конце полинуклеотида, и вторая нецелевая последовательность расположена на 3' конце полинуклеотида. Дополнительно в настоящем документе предусмотрены способы, при которых один или несколько молекулярных маркеров прикреплены к 5' концу полинуклеотида. Дополнительно в настоящем документе предусмотрены способы, при

которых один или несколько молекулярных маркеров прикреплены к 3' концу полинуклеотида. Дополнительно в настоящем документе предусмотрены способы, при которых один или несколько молекулярных маркеров и полинуклеотид соединены с помощью спейсера. Дополнительно в настоящем документе предусмотрены способы, при которых вставочная последовательность комплементарна по меньшей мере одному экзону. Дополнительно в настоящем документе предусмотрены способы, при которых один или несколько молекулярных маркеров представляют собой биотин, фолат, полигистидин, FLAG-маркер или глутатион. Дополнительно в настоящем документе предусмотрены способы, при которых один или несколько молекулярных маркеров представляют собой две молекулы биотина. Дополнительно в настоящем документе предусмотрены способы, при которых твердая подложка представляет собой магнитный микрошарик. Дополнительно в настоящем документе предусмотрены способы, при которых первая нецелевая последовательность и вторая нецелевая последовательность составляют от 20 до 40 оснований в длину. Дополнительно в настоящем документе предусмотрены способы, при которых вставочная последовательность составляет от 90 до 200 оснований в длину. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека содержит по меньшей мере 5000 полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека содержит по меньшей мере 10000 полинуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых библиотека содержит по меньшей мере 30000 полинуклеотидов.

[0225] В настоящем документе предусмотрены библиотеки зондов, содержащие множество частично комплементарных двухцепочечных полинуклеотидов, причем каждый содержит: первый полинуклеотид, содержащий следующее: первая нецелевая и вторая нецелевая последовательность; и первая вставочная последовательность, которая является комплементарной одной или нескольким целевым последовательностям нуклеиновой кислоты; второй полинуклеотид, содержащий следующее: первая нецелевая последовательность и вторая нецелевая последовательность; и вторая вставочная последовательность, которая является комплементарной первой вставочной последовательности; причем первый полинуклеотид и второй полинуклеотид частично гибридизованы. Дополнительно в настоящем документе предусмотрены библиотеки, в которых каждая цепь двухцепочечных полинуклеотидов дополнительно содержит по меньшей мере два молекулярных маркера. Дополнительно в настоящем документе предусмотрены библиотеки, в которых первая нецелевая последовательность и вторая нецелевая последовательность не являются комплементарными. Дополнительно в настоящем документе предусмотрены библиотеки, в которых первая нецелевая

последовательность расположена на 5' конце полинуклеотида, и вторая нецелевая последовательность расположена на 3' конце полинуклеотида. Дополнительно в настоящем документе предусмотрены библиотеки, в которых один или несколько молекулярных маркеров прикреплены к 5' концу полинуклеотида. Дополнительно в настоящем документе предусмотрены библиотеки, в которых один или несколько молекулярных маркеров прикреплены к 3' концу полинуклеотида. Дополнительно в настоящем документе предусмотрены библиотеки, в которых один или несколько молекулярных маркеров и полинуклеотид соединены с помощью спейсера. Дополнительно в настоящем документе предусмотрены библиотеки, в которых вставочная последовательность является комплементарной по меньшей мере одному экзону. Дополнительно в настоящем документе предусмотрены библиотеки, в которых один или несколько молекулярных маркеров представляют собой биотин, фолат, полигистидин, FLAG-маркер или глутатион. Дополнительно в настоящем документе предусмотрены библиотеки, в которых один или несколько молекулярных маркеров представляют собой две молекулы биотина. Дополнительно в настоящем документе предусмотрены библиотеки, в которых твердая подложка представляет собой магнитный микрошарик. Дополнительно в настоящем документе предусмотрены библиотеки, в которых первая нецелевая последовательность и вторая нецелевая последовательность составляют от 20 до 40 оснований в длину. Дополнительно в настоящем документе предусмотрены библиотеки, в которых вставочная последовательность составляет от 90 до 200 оснований в длину.

[0226] В настоящем документе предусмотрены способы конструирования библиотеки зондов, предусматривающие следующее: получение библиотеки целевых последовательностей; и конструирование библиотеки вставочных последовательностей, комплементарных целевым последовательностям, причем конструирование предусматривает следующее: создание вставочных последовательностей, комплементарных целевым последовательностям, если целевая последовательность короче в длину, чем вставочная последовательность; создание вставочных последовательностей, по меньшей мере частично комплементарных целевым последовательностям, если целевая последовательность короче в длину, чем вставочная последовательность + X; или создание набора вставочных последовательностей, по меньшей мере частично комплементарных общей целевой последовательности, если целевая последовательность длиннее, чем вставочная последовательность + X, причем X представляет собой число последовательных оснований, на которые не нацелена вставочная последовательность; повторение стадии (b) для каждой целевой последовательности в библиотеке для создания библиотеки вставочных последовательностей. Дополнительно в настоящем документе предусмотрены

способы, при которых X составляет меньше чем 30 нуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых X составляет меньше чем 10 нуклеотидов. Дополнительно в настоящем документе предусмотрены способы, при которых X составляет приблизительно 6 нуклеотидов.

[0227] В настоящем документе предусмотрены способы секвенирования нового поколения, предусматривающие приведение в контакт описанной в настоящем документе библиотеки с образцом, содержащим множество целевых полинуклеотидов; обогащение по меньшей мере одного целевого полинуклеотида, который связывается с библиотекой; и секвенирование по меньшей мере одного обогащенного целевого полинуклеотида.

[0228] В настоящем документе предусмотрены способы секвенирования нового поколения, предусматривающие следующее: приведение в контакт описанной в настоящем документе библиотеки с образцом, содержащим множество полинуклеотидов; отделение по меньшей мере одного полинуклеотида в образце, который связывается с библиотекой, от по меньшей мере одного полинуклеотида, который не связывается с библиотекой; и секвенирование по меньшей мере одного полинуклеотида, который не связывается с библиотекой.

ПРИМЕРЫ

[0229] Следующие примеры приведены с целью иллюстрации различных вариантов осуществления настоящего изобретения и не предназначены для ограничения настоящего изобретения 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 изобретения. Изменения в настоящем документе и другие применения, которые охватываются сущностью настоящего изобретения, как определено объемом формулы изобретения, будут понятны специалистам в настоящей области техники.

[0230] Пример 1: Функционализация поверхности субстрата

[0231] Субстрат функционализировали для поддержания прикрепления и синтеза библиотеки полинуклеотидов. Поверхность субстрата сначала подвергали влажной очистке с использованием раствора «пираньи», содержащего 90% H_2SO_4 и 10% H_2O_2 , в течение 20 минут. Субстрат 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 субстрат очищали плазмой, подвергая поверхность субстрата воздействию O_2 . Прибор SAMCO PC-300 использовали для плазмохимического травления O_2 при 250 Вт в течение 1 мин в нисходящем режиме.

[0232] Очищенную поверхность подложки активно функционализировали раствором, содержащим N-(3-триэтоксисилилпропил)-4-гидроксипиридин, 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 субстрата покрывали резистом с использованием прибора для нанесения растворов центрифугированием Brewer Science 200X. Фоторезист SPRTM 3612 наносили методом центрифугирования на субстрат при 2500 об/мин в течение 40 с. Субстрат предварительно прокачивали в течение 30 мин при 90°C на горячей плите Брюера. Субстрат подвергали фотолитографии с использованием установки для совмещения и экспонирования Karl Suss MA6. Субстрат 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. Субстрат прокачивали в течение 30 мин при 100°C в печи с последующей визуальной проверкой дефектов литографии с использованием Nikon L200. Процесс удаления накали использовали для удаления остаточного резиста с использованием прибора SAMCO PC-300 для плазмохимического травления O_2 при 250 Вт в течение 1 мин.

[0233] Поверхность субстрата пассивно функционализировали с помощью 100 мкл раствора перфтороктилтрихлорсилана, смешанного с 10 мкл легкого минерального масла. Субстрат помещали в камеру, накачивали насосом в течение 10 мин, а затем клапан закрывали для насоса и оставляли стоять в течение 10 мин. Камеру вентилировалась. Субстрат 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 субстрат замачивали в течение 5 минут в 500 мл изопропанола при комнатной температуре с обработкой ультразвуком при максимальной мощности. Субстрат погружали в 300 мл этанола с крепостью 200 и продували N_2 . Функционализированную поверхность активировали для поддержания синтеза полинуклеотидов.

[0234] Пример 2: Синтез 50-мерной последовательности на устройстве для синтеза полинуклеотидов

[0235] 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.

[0236] Последовательность, составляющую 50-мер, являлась такой, как описано в SEQ ID NO.: 1. 5'AGACAATCAACCATTTGGGGTGGACAGCCTTGACCTCTAGACTTCGGCAT##TTTTT TTTTT3' (SEQ ID NO.: 1), где # обозначает тимидин-сукцинилгексамид-CED-фосфoramидит (CLP-2244 от ChemGenes), который представляет собой расщепляемый линкер, обеспечивающий высвобождение полинуклеотидов с поверхности во время снятия защитных групп.

[0237] Синтез выполняли с использованием стандартной химии синтеза ДНК (сочетание, кэпирование, окисление и деблокирование) согласно протоколу в таблице 1 и синтезатора ABI.

Таблица 1

Таблица 1		
Общее название процесса 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 фосфoramидита + активатора)	Промывка манифольда активатора	2
	Активатор в проточную кювету	6
	Активатор + фосфoramидит в проточную кювету	6
	Активатор в проточную кювету	0,5
	Активатор + фосфoramидит в проточную кювету	5
	Активатор в проточную кювету	0,5

Таблица 1		
Общее название процесса синтеза ДНК	Стадия способа	Время (с)
	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
	Промывка системы ацетонитрилом	4
Добавление оснований ДНК (поток фосфорамидита + активатора)	Промывка манифольда активатора	2
	Активатор в проточную кювету	5
	Активатор + фосфорамидит в проточную кювету	18
	Инкубировать в течение 25 с	25
Промывка (промывка потоком 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 потоком ацетонитрила)	Промывка системы ацетонитрилом	4

Таблица 1		
Общее название процесса синтеза ДНК	Стадия способа	Время (с)
	Ацетонитрил в проточную кювету	15
	Промывка системы ацетонитрилом	4
Окисление (поток окислителя)	Окислитель в проточную кювету	18
Промывка (промывка потоком 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
	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

Таблица 1		
Общее название процесса синтеза ДНК	Стадия способа	Время (с)
	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

[0238] 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.

[0239] 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 полинуклеотидов. Извлеченные полинуклеотиды затем анализировали на небольшом РНК-чипе BioAnalyzer (данные не показаны).

[0240] Пример 3: Синтез 100-мерной последовательности на устройстве для синтеза полинуклеотидов

[0241] 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.: 2) 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 полинуклеотиды, экстрагированные с поверхности, анализировали на приборе BioAnalyzer (данные не показаны).

[0242] Все десять образцов из двух чипов дополнительно амплифицировали с помощью ПЦР с использованием прямого (5'ATGCGGGGTTCTCATCATC3'; SEQ ID NO.: 3) и обратного (5'CGGGATCCTTATCGTCATCG3'; SEQ ID NO.: 4) 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 полинуклеотида, извлеченного с поверхности, и воды до 50 мкл) с использованием следующей программы термочиклирования:

98°C, 30 с

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

72°C, 2 мин

[0243] 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 2 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 2

Пятно	Частота ошибок	Эффективность цикла
1	1/763 п.н.	99,87%
2	1/824 п.н.	99,88%
3	1/780 п.н.	99,87%
4	1/429 п.н.	99,77%
5	1/1525 п.н.	99,93%

[0245] Наконец, в таблице 3 обобщенно представлены характеристики ошибок для последовательностей, полученных из образцов полинуклеотидов из пятен 1-10.

[illegible]

основания ROI										
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

[0246] Пример 4: Параллельная сборка 29040 уникальных полинуклеотидов

[0247] Структуру, содержащую 256 кластеров **1005**, каждый из которых содержит 121 locus на плоском кремниевом планшете **1001**, получали, как показано на фиг. **10**. Развернутый вид кластера показан в **1010** с 121 locusом. Locusы из 240 из 256 кластеров обеспечивали прикрепление и подложку для синтеза полинуклеотидов, характеризующихся отличающимися последовательностями. Синтез полинуклеотидов проводили с помощью фосфорамидитной химии с использованием общих способов из примера 3. Locusы из 16 кластеров из 256 кластеров представляли собой контрольные кластеры. Глобальное распределение 29040 синтезированных уникальных полинуклеотидов (240 x 121) показано на фиг. **11A**. Библиотеки полинуклеотидов синтезировали с высокой однородностью. 90% последовательностей присутствовали в сигналах в пределах 4-кратного распределения от среднего, обеспечивая 100% представление. Распределение измеряли для каждого кластера, как показано на фиг. **11B**. Распределение уникальных полинуклеотидов, синтезированных в 4 репрезентативных кластерах, показано на фиг. **12**. На глобальном уровне присутствовали все полинуклеотиды в прогоне и 99% полинуклеотидов характеризовались представленностью, которая находилась в пределах 2-кратной от среднего, указывая на однородность синтеза. Этому наблюдению не противоречили наблюдения на уровне кластера.

[0248] Частоту ошибок для каждого полинуклеотида определяли с использованием генного секвенатора Illumina MiSeq. Распределение частоты ошибок для 29040 уникальных полинуклеотидов показано на фиг. **13А** и в среднем оно составило приблизительно 1 на 500 оснований, при этом некоторые показатели частоты ошибок были ниже чем 1 на 800 оснований. Распределение измеряли для каждого кластера, как показано на фиг. **13В**. Распределение частоты ошибок для уникальных полинуклеотидов в четырех репрезентативных кластерах показано на фиг. **14**. Библиотеку из 29040 уникальных полинуклеотидов синтезировали меньше чем за 20 часов.

[0249] Анализ процентного отношения GC в зависимости от представления полинуклеотидов по всем из 29040 уникальных полинуклеотидов показал, что синтез являлся однородным несмотря на содержание GC, фиг. **15**.

[0250] Пример 5: ПЦР-амплификация библиотеки синтезированных полинуклеотидов

[0251] 9996 полинуклеотидов, каждый 100 оснований в длину, рандомизированных последовательностей с различным содержанием GC, от 20 до 80% GC конструировали и синтезировали на структуре с расположением, подобным описанному в примере 3. Для определения влияния ПЦР-амплификации на представление GC популяцию полинуклеотидов амплифицировали в течение 6 или 20 циклов с ДНК-полимеразой высокой точности (ДНК-полимераза 1). Альтернативно популяцию полинуклеотидов амплифицировали с использованием двух других высокоточных ферментов для ПЦР в течение 6, 8, 10 или 15 циклов, чтобы определить, влияет ли выбор полимеразы на общее представление последовательностей после амплификации. После ПЦР-амплификации образцы подготавливали для секвенирования нового поколения и секвенировали на платформе Illumina MiSeq. SE прочтения по 150 п.н. генерировали до приблизительно глубины прочтения, составляющей 100X. Необработанные файлы FASTQ анализировали. Представление полинуклеотидов с использованием любой полимеразы в течение 6, 10 или 15 циклов изображено на фиг. **16**. Однородность представления полинуклеотидов оценивали для различных условий и обобщенно представили в таблице 4.

Таблица 4

	Циклы	% в пределах 1,5x	% в пределах 2x
Полимераза 1	6	72,1%	92,6%
	8	76,1%	90,3%
	10	70,9%	86,6%

	15	64,1%	82,7%
Полимераза 2	6	91,9%	98,9%
	8	89,9%	98,1%
	10	90,1%	98,4%
	15	89,2%	97,9%

[0252] Количество выпадений для каждой амплифицированной популяции полинуклеотидов определяли количественно, как показано на фиг. 15, зависимость циклов амплификации от доли популяции ниже 10% от среднего порогового значения. Выпадения полимеразы 1 росли быстро, тогда как выпадения полимеразы 2 оставались относительно постоянными.

[0253] Оценивали влияние избыточной амплификации на распределение GC, фиг. 18. Как правило, полинуклеотиды с содержанием GC от 30% до 70% следовали линии тренда, $Y = X$, и увеличивались по частоте с увеличением числа циклов. Полинуклеотиды с содержанием GC более 70% встречались, как правило, несколько более часто после 20 циклов, тогда как полинуклеотиды с содержанием GC менее 30% встречались, как правило, несколько чаще после 6 циклов.

[0254] Пример 6. Сравнение представления полинуклеотидов из амплификации целого планшета с параллельной кластерной амплификацией полинуклеотидов

[0255] Полинуклеотиды синтезировали на структуре, содержащей 256 кластеров, каждый из которых содержал 121 локус на плоском кремниевом планшете, произведенном, как показано на фиг. 10. Синтез полинуклеотидов проводили с помощью фосфорамидитной химии с использованием общих способов из примера 3. Полинуклеотиды на структуре расщепляли и комбинировали.

[0256] Полинуклеотиды комбинировали по всему планшету и амплифицировали. После амплификации наблюдали заметное смещение и отклонение от среднего GC, как видно на линии, показанной на фиг. 19. В результате потребовалось больше секвенирования и наблюдали больше выпадений.

[0257] Распределение полинуклеотидов из амплификации кластеров показано на фиг. 20. В прогоне 1 и прогоне 2 распределение частот от среднего (линия) составляло приблизительно 8, и отклонение от среднего являлось приблизительно 1,7-кратным. Процентное отношение GC находилось в диапазоне от 17% до 94%. На фиг. 23 и фиг. 20

показано, что существует воспроизводимость, и популяция полинуклеотидов показывает сильное снижение смещения GC (фиг. 20). Кроме того, отсутствовали выпадения и потребовалось на 30% меньше секвенирования.

Пример 7. Синтезированные библиотеки полинуклеотидов с различным содержанием GC

[0258] Библиотеку из 13000 полинуклеотидных последовательностей, характеризующихся содержанием GC, составляющим приблизительно 15% - приблизительно 85%, предварительно выбирали для синтеза (фиг. 22). Первую библиотеку полинуклеотидов синтезировали на структуре и синтез проводили с помощью фосфорамидитной химии с использованием общих способов из примера 3. Полинуклеотиды на структуре расщепляли и комбинировали с последующей амплификацией для создания ПЦР-смещенной библиотеки полинуклеотидов. Полинуклеотидные последовательности в библиотеке распределяли по бинам в соответствии с содержанием GC и стехиометрию каждого бина корректировали с учетом наблюдаемого смещения GC, созданного за счет ПЦР-амплификации. Например, полинуклеотиды, содержащие повышенное или пониженное содержание GC, характеризовались повышенными начальными концентрациями, которые приводили к однородному стехиометрическому представлению после амплификации. Это эффективно снижает или устраняет вызванное ПЦР смещение GC из стадии амплификации. Вторую библиотеку полинуклеотидов синтезировали на структуре и синтез проводили с помощью фосфорамидитной химии с использованием общих способов из примера 3. Полинуклеотиды на структуре расщепляли и комбинировали с последующей амплификацией для создания в высокой степени однородной библиотеки полинуклеотидов (фиг. 23) с однородным представлением GC после амплификации (фиг. 21A). Одним из преимуществ сбалансированной по содержанию GC библиотеки является то, что она требует меньше сэмпинга для необходимого покрытия сэмпинга. Например, показатели частоты сэмпинга для 80% и 90% покрытия библиотеки достигали теоретического минимума для монодисперсной библиотеки (фиг. 24A и фиг. 24B). Кроме того, синтезировали библиотеки полинуклеотидов, которые предпочтительно характеризовались различными степенями как высокого, так и низкого содержания GC (фигуры 21B и 21C соответственно). Кроме того, синтезировали библиотеку полинуклеотидов, предпочтительно характеризующуюся низким содержанием GC (фиг. 21D) или высоким содержанием GC (фиг. 21E).

[0259] Пример 8. Сбалансированные по GC библиотеки полинуклеотидов, синтезированные с длинами полинуклеотидов, составляющими 80- и 120-мер

[0260] Библиотеку, содержащую приблизительно 20000 уникальных полинуклеотидов, каждый из которых составляет 80 нуклеиновых кислот в длину, конструировали и уравнивали по содержанию GC с использованием общего способа согласно примеру 7 и синтезировали на структуре; синтез проводили с помощью фосфорамидитной химии с использованием общих способов из примера 3. Кроме того, синтезировали сходную библиотеку, содержащую полинуклеотиды, каждый из которых составляет 120 нуклеиновых кислот в длину (фиг. 26). Обе библиотеки показали в высокой степени однородное распределение, с идентифицированными >99% уникальных последовательностей. Библиотеки также проявляли однородность в отношении отклонения в содержании GC, с высоким согласием между повторностями (фиг. 27), с низким числом полинуклеотидов на хвостах, подверженных шуму.

[0261] Пример 9. Оценка содержания GC после повторной амплификации библиотеки полинуклеотидов

[0262] Библиотеку полинуклеотидов, состоящую из 9,996 уникальных полинуклеотидов, характеризующихся содержанием GC, составляющим 20-80%, каждый из которых составляет 100 оснований в длину, синтезировали на структуре с помощью фосфорамидитной химии с использованием общих способов из примера 3. Библиотеку амплифицировали в течение или 8, или 15 циклов ПЦР с двумя различными высокоточными ДНК-полимеразами и частоту полинуклеотидов в популяции сравнивали между этими двумя состояниями (фиг. 28). Линия идентичности (черная пунктирная линия) указывает на полинуклеотиды с одинаковой частотой в популяции после или 8, или 15 циклов ПЦР. Последовательности выше линии идентичности являются избыточно представленными в популяции после 15 циклов, и последовательности ниже линии являются недостаточно представленными в популяции после 15 циклов, по отношению к 8 циклам амплификации. В этом случае полимераза 1 проявляет смещение GC с увеличением циклов ПЦР. Последовательности с высоким содержанием GC (GC выше 70%, серый средней насыщенности) согласно наблюдениям являлись избыточно представленными, и последовательности с низким содержанием GC (GC ниже 30%, показанные насыщенным серым) являлись недостаточно представленными после 15 циклов амплификации, по сравнению с 8 циклами. Кроме того, большая величина вариации в обогащении в пределах сходных процентных отношений GC указывает на то, что другие факторы, отличные от содержания GC, такие как образование шпилек или гомополимерные фрагменты, могут

влиять на смещение за счет амплификации. Полимераза 2 не проявляла такого же смещения представления последовательностей.

[0263] Пример 10. Оценка выпадений и представления после повторной амплификации библиотеки полинуклеотидов

[0264] Исследовали смещение, вводимое за счет амплификации с различными ферментами ДНК-полимеразами. Библиотеку полинуклеотидов согласно примеру 9 амплифицировали или с ДНК-полимеразой 1 (фиг. 29, темная линия), или ДНК-полимеразой 2 (фиг. 29, светлая линия) в течение 6, 8, 10 или 15 циклов. Увеличение циклов ПЦР коррелировало с увеличенной частотой выпадений полинуклеотидных последовательностей, где частоту выпадений определяют как последовательности с представленностью, составляющей меньше чем 10% от среднего. Степень этого эффекта зависела от ДНК-полимеразы, используемой для амплификации. Большая доля последовательностей выпадала после амплификации с ДНК-полимеразой 1 (приблизительно в 20 раз больше выпадений после 15 циклов) по сравнению с амплификацией с ДНК-полимеразой 2. Различные полимеразы могут являться оптимальными для амплификации различных последовательностей библиотеки в зависимости от содержания GC, длины и сложности последовательности.

[0265] Библиотеки, амплифицированные в течение 15 циклов ПЦР с каждой ДНК-полимеразой, исследовали более подробно для оценки представления полинуклеотидных последовательностей (фиг. 30). Амплификация библиотеки полинуклеотидов с помощью ДНК-полимеразы 1 давала в результате в 20 раз больше выпадений последовательностей после 15 циклов ПЦР (фиг. 29). Распределение полинуклеотидов, амплифицированных с помощью ДНК-полимеразы 1, было больше чем распределение полинуклеотидов, амплифицированных с помощью ДНК-полимеразы 2. Распределение полинуклеотидов библиотеки, амплифицированной с помощью ДНК-полимеразы 1, характеризовалось 64% последовательностей, находящихся в пределах 1,5-кратного значения от среднего. Когда эту же библиотеку амплифицировали с помощью ДНК-полимеразы 2, >89% последовательностей находились в пределах 1,5-кратного значения от среднего, указывая на то, что ДНК-полимераза 2 амплифицировала библиотеку с гораздо более низким смещением, чем ДНК-полимераза 1. Смещение, введенное в библиотеку, амплифицированную с помощью ДНК-полимеразы 1, увеличивает скрининг, необходимый для покрытия библиотеки полинуклеотидов.

[0266] Пример 11. Применение библиотеки полинуклеотидов с контролируемой стехиометрией для нацеленного воздействия на экзон с помощью секвенирования нового поколения (NGS)

[0267] Первую нацеленную на кДНК библиотеку полинуклеотидов (библиотеку зондов), содержащую вплоть до 370000 или больше неидентичных полинуклеотидов, которые перекрываются с одним или несколькими генными экзонами, конструируют и синтезируют на структуре с помощью фосфорамидитной химии с использованием общих способов из примера 3. Полинуклеотиды лигируют с молекулярным маркером, таким как биотин, с использованием ПЦР (или напрямую во время твердофазного синтеза) с образованием зонда для последующего захвата целевых экзонов, представляющих интерес. Зонды гибридизуются с последовательностями в библиотеке геномных нуклеиновых кислот и их отделяют от не связывающихся последовательностей. Несвязанных зонды отмывают, оставляя целевую библиотеку, обогащенную последовательностями кДНК. Обогащенную библиотеку затем секвенируют с использованием NGS и прочтения для каждого предполагаемого гена измеряют как функцию от зонда(ов) кДНК, используемого(ых) для нацеливания на ген.

[0268] В некоторых случаях на частоту прочтений целевой последовательности влияет представленность целевой последовательности, связывание с зондом, вторичная структура или другие факторы, которые снижают представление после секвенирования целевой последовательности, несмотря на обогащение. Стехиометрический контроль библиотеки полинуклеотидов проводят путем модификации стехиометрии первой нацеленной на кДНК библиотеки полинуклеотидов для получения второй нацеленной на кДНК полинуклеотидов библиотеки, с увеличенной стехиометрией в отношении последовательностей полинуклеотидных зондов, что приводит к меньшему числу прочтений. Эту вторую нацеленную на кДНК библиотеку конструируют и синтезируют на структуре с помощью фосфорамидитной химии с использованием общих способов из примера 3, и используют для обогащения экзонов последовательностей библиотеки целевой геномной ДНК, как описано ранее.

[0269] Пример 12. Множественные итерации стехиометрического контроля с помощью библиотеки экзомных зондов

[0270] Библиотеку экзомных зондов синтезировали и исследовали с использованием общего способа согласно примеру 11. Проводили множественные итерации стехиометрической модификации, приводящие в результате к библиотеке зондов с контролируемой стехиометрией, библиотеке 1. По сравнению с некоторыми наборами сравнения обогащения экзона она приводила к значительно меньшему числу прочтений

секвенирования для получения требуемого покрытия мишеней. Для точного секвенирования требуется 30X глубина прочтения по меньшей мере 90% оснований целевого экзона, и избыточное секвенирование (теоретическая глубина прочтения, больше чем 30X глубины прочтения) часто необходимо для компенсации проблем с однородностью. Библиотека экзомных зондов с контролируемой стехиометрией способна достигать 30X глубины прочтения 90% целевых оснований с 55X теоретической глубиной прочтения (фиг. 32A), что представляло собой значительно меньшее покрытие секвенирования, чем требовалось с помощью другого набора сравнения обогащения экзона, и более быструю скорость обработки последовательностей (число образцов за прогон, таблица 5). При нормировании к 4,5 млрд.п.н. секвенирования библиотека зондов с контролируемой стехиометрией обеспечивала 10X глубину прочтения >95% всех целевых оснований, и это было значительно выше, чем все другие наборы сравнения экзомных зондов, с которыми проводили сравнение (фиг. 32B).

Таблица 5

Необходимое среднее покрытие секвенирования	Образцов на прогон
Набор сравнения обогащения экзона А	4
Экзомные зонды с контролируемой стехиометрией (библиотека 1)	17

[0271] Пример 13: Получение гибридизационных панелей

[0272] Нацеленные на полинуклеотиды библиотеки получали с использованием общего способа согласно примеру 11, которые нацеленно воздействуют на конкретные гены, заболевания, комбинации панелей или пользовательские экзоны. Размеры реакции составляли 10^3 по своим масштабам, а размеры панели зондов находились в диапазоне, составляющем приблизительно 80 - приблизительно 900000 зондов (фиг. 33).

[0273] Пример 14: Получение панели из 70000 зондов

[0274] Нацеленную на полинуклеотиды библиотеку (библиотеку зондов), содержащую 70000 неидентичных полинуклеотидов, конструировали и синтезировали на структуре с помощью фосфорамидитной химии с использованием общих способов из примера 3, и контролировали содержание GC с использованием общего способа согласно примеру 11 для создания библиотеки 2. Распределение прочтений после секвенирования показано на фиг. 34, и покрытие мишеней, распределенных по бинам в отношении GC, показано на фиг. 35A и фиг. 35B.

[0275] Пример 15: Получение панели из 2544 зондов

[0276] Нацеленную на полинуклеотиды библиотеку (библиотеку зондов), содержащую 2544 неидентичных полинуклеотидов, конструировали и синтезировали на структуре с помощью фосфорамидитной химии с использованием общих способов из примера 3, и стехиометрию контролировали с использованием общего способа согласно примеру 11 для создания библиотеки 3. Частота встречаемости мишени показана на фиг. **36А**, и частота покрытия показана на фиг. **36В**. Целевое обогащение с помощью библиотеки 3 приводило как к повышенной частоте встречаемости мишени, так и повышенной частоте покрытия по сравнению с набором сравнения № 2 на основе матрицы.

[0277] Пример 16: Получение образцов и обогащение с помощью нацеленной на полинуклеотиды библиотеки

[0278] Геномную ДНК (гДНК) получают из образца и фрагментируют ферментативно в буфере для фрагментации, проводят репарацию концов и аденилируют на 3'-конце. Адаптеры лигируют по обоим концам фрагментов геномной ДНК с получением библиотеки маркированных адаптерами цепей гДНК, и маркированную адаптерами библиотеку ДНК амплифицируют с помощью высокоточной полимеразы. Библиотеку гДНК затем денатурируют до отдельных цепей при 96°C в присутствии блокаторов адаптеров. Нацеленную на полинуклеотиды библиотеку (библиотеку зондов) денатурируют в гибридизационном растворе при 96°C и комбинируют с денатурированной, маркированной библиотекой гДНК в гибридизационном растворе в течение 16 часов при 70°C. Буфер для связывания затем добавляют к зондам, гибридизованным с маркированной гДНК, и магнитные микрошарики, содержащие стрептавидин, используют для захвата биотинилированных зондов. Микрошарики отделяют от раствора с использованием магнита и микрошарики отмывают три раза с помощью буфера для удаления несвязанных адаптеров, гДНК и блокаторов адаптеров до добавления элюирующего буфера для высвобождения обогащенных, маркированных фрагментов гДНК из микрошариков. Обогащенную библиотеку маркированных фрагментов гДНК амплифицируют с высокоточной полимеразой для получения выходов, достаточных для создания кластера, и затем библиотеку секвенируют с использованием прибора NGS.

[0279] Пример 17: Общее получение образцов и обогащение с помощью нацеленной на полинуклеотиды библиотеки

[0280] Множество полинуклеотидов получают из образца и фрагментируют, необязательно проводят репарацию концов и аденилируют. Адаптеры лигируют по обоим концам полинуклеотидных фрагментов с получением библиотеки маркированных адаптерами полинуклеотидных цепей и библиотеку маркированных адаптерами

полинуклеотидов амплифицируют. Библиотеку маркированных адаптерами полинуклеотидов затем денатурируют при высокой температуре, предпочтительно при 96°C, в присутствии блокаторов адаптеров. Нацеленную на полинуклеотиды библиотеку (библиотеку зондов) денатурируют в гибридизационном растворе при высокой температуре, предпочтительно приблизительно 90 - 99°C, и комбинируют с денатурированной, маркированной библиотекой полинуклеотидов в гибридизационном растворе в течение приблизительно 10 - 24 часов при приблизительно 45 - 80°C. Буфер для связывания затем добавляют к зондам, гибридизованным с маркированными полинуклеотидами, и твердую подложку, содержащую фрагмент для захвата, используют для селективного связывания с зондами, гибридизованными с полинуклеотидами, маркированными адаптерами. Твердую подложку отмывают один или несколько раз буфером, предпочтительно приблизительно 2 - 5 раз для удаления несвязанных полинуклеотидов до добавления элюирующего буфера для высвобождения обогащенных, маркированных адаптерами полинуклеотидных фрагментов с твердой подложки. Обогащенную библиотеку маркированных адаптерами полинуклеотидных фрагментов амплифицируют и затем библиотеку секвенируют.

[0281] Пример 18: Общее обогащение перед маркированием с помощью нацеленной на полинуклеотиды библиотеки

[0282] Множество полинуклеотидов получают из образца, фрагментируют и необязательно проводят репарацию концов. Фрагментированный полинуклеотидный образец затем денатурируют при высокой температуре, предпочтительно при 96°C. Нацеленную на полинуклеотиды библиотеку (библиотеку зондов) денатурируют в гибридизационном растворе при высокой температуре, предпочтительно приблизительно 90 - 99°C, и комбинируют с библиотекой денатурированных полинуклеотидов в гибридизационном растворе в течение приблизительно 10 - 24 часов при приблизительно 45 - 80°C. Буфер для связывания затем добавляют к зондам, гибридизованным с полинуклеотидами, и твердую подложку, содержащую фрагмент для захвата, используют для селективного связывания с зондами, гибридизованными с фрагментированными полинуклеотидами. Твердую подложку отмывают один или несколько раз буфером, предпочтительно приблизительно 2 - 5 раз для удаления несвязанных полинуклеотидов до добавления элюирующего буфера для высвобождения обогащенных полинуклеотидных фрагментов с твердой подложки. Обогащенные полинуклеотиды аденилируют, адаптеры лигируют по обоим концам полинуклеотидов с получением обогащенной библиотеки маркированных адаптерами полинуклеотидных цепей и библиотеку маркированных

адаптерами полинуклеотидов амплифицируют. Обогащенную библиотеку маркированных адаптерами полинуклеотидных фрагментов затем секвенируют.

[0283] Пример 19: Общее получение образцов и фильтрование с помощью нацеленной на полинуклеотиды библиотеки

[0284] Множество полинуклеотидов получают из образца и фрагментируют, необязательно проводят репарацию концов и аденилируют. Адаптеры лигируют по обоим концам фрагментов полинуклеотидов с получением библиотеки маркированных адаптерами полинуклеотидных цепей и библиотеку маркированных адаптерами полинуклеотидов амплифицируют. Библиотеку маркированных адаптерами полинуклеотидов затем денатурируют при высокой температуре, предпочтительно при 96°C, в присутствии блокаторов адаптеров. Фильтрующую полинуклеотиды библиотеку (библиотеку зондов), сконструированную для удаления ненужных, нецелевых последовательностей, денатурируют в гибридизационном растворе при высокой температуре, предпочтительно приблизительно 90 - 99°C, и комбинируют с денатурированной библиотекой маркированных полинуклеотидов в гибридизационном растворе в течение приблизительно 10 - 24 часов при приблизительно 45 - 80°C. Буфер для связывания затем добавляют к зондам, гибридизованным с маркированными полинуклеотидами, и твердую подложку, содержащую фрагмент для захвата, используют для селективного связывания с зондами, гибридизованными с полинуклеотидами, маркированными адаптерами. Твердую подложку отмывают один или несколько раз буфером, предпочтительно приблизительно 1 - 5 раз для элюирования целевых маркированных адаптерами полинуклеотидных фрагментов. Обогащенную библиотеку целевых маркированных адаптерами полинуклеотидных фрагментов амплифицируют и затем библиотеку секвенируют.

[0285] Пример 20: Получение библиотеки 160-мерных зондов

[0286] Библиотеку, содержащую по меньшей мере 1000 зондов, синтезировали на структуре, и синтез проводили с помощью фосфорамидитной химии с использованием общих способов из примера 3. Каждый зонд является двухцепочечным, и каждая цепь зонда содержит состоящую из 120 нуклеотидов связывающуюся с мишенью последовательность, комплементарную мишени. Каждый зонд дополнительно содержит состоящий из 20 нуклеотидов сайт связывания прямого праймера и состоящий из 20 нуклеотидов сайт связывания обратного праймера. Каждую цепь зондов метят в положении 5' двумя молекулами биотина.

[0287] Пример 21: Получение библиотеки 210-мерных зондов, содержащей не связывающуюся с мишенью последовательность

[0288] Библиотеку, содержащую по меньшей мере 1000 зондов, синтезировали на структуре, и синтез проводили с помощью фосфорамидитной химии с использованием общих способов из примера 3. Каждый зонд является двухцепочечным, и каждая цепь зонда содержит состоящую из 120 нуклеотидов связывающуюся с мишенью последовательность, комплементарную мишени. Каждый зонд дополнительно содержит состоящий из 20 нуклеотидов сайт связывания прямого праймера, состоящий из 20 нуклеотидов сайт связывания обратного праймера, состоящую из 25 нуклеотидов, расположенную 5' не связывающуюся с мишенью последовательность, содержащую полиаденин, и состоящую из 25 нуклеотидов, расположенную 3' не связывающуюся с мишенью последовательность, содержащую полиаденин. Каждую цепь зондов метят в положении 5' двумя молекулами биотина.

[0289] Пример 22: 210-мерные зонды, нацеленные на экзон 1 HLA человека

[0290] Библиотеку, содержащую по меньшей мере 1000 зондов, синтезировали на структуре, и синтез проводили с помощью фосфорамидитной химии с использованием общих способов из примера 3. Каждый зонд является двухцепочечным, и каждая цепь зонда содержит состоящую из 120 нуклеотидов связывающуюся с мишенью последовательность, комплементарную области экзона 1 HLA человека. Каждый зонд дополнительно содержит состоящий из 20 нуклеотидов сайт связывания прямого праймера, состоящий из 20 нуклеотидов сайт связывания обратного праймера, состоящую из 25 нуклеотидов, расположенную 5' не связывающуюся с мишенью последовательность, содержащую полиаденин, и состоящую из 25 нуклеотидов, расположенную 3' не связывающуюся с мишенью последовательность, содержащую полиаденин. Каждую цепь зондов метят в положении 5' двумя молекулами биотина.

[0291] Пример 22: Способ конструирования библиотеки неперекрывающихся зондов

[0292] По меньшей мере 100 целевых последовательностей предоставляют и сортируют по отдельным категориям на основании длины по сравнению с требуемой длиной комплементарной связывающейся с мишенью последовательностью зонда. Например, категории включают в себя без ограничения следующее: (a) мишени короче, чем длина вставки, (b) мишени короче или равные длине вставки + X, и (c) мишени длиннее, чем длина вставки + X, где X представляет собой требуемую длину пропуска, на которую не нацелен зонд. На целевую последовательность в категории (a) нацелены связывающиеся с мишенью последовательности, которые либо расположены по центру, либо выровнены по

левой или правой стороне целевой последовательности, в зависимости от сложности нецелевой области (повторяющаяся последовательность, последовательность с высоким/низким содержанием GC, палиндромная последовательность и т. д.), к которым также будет комплементарна вставка. На целевые последовательности в категории (b) нацелено воздействуют таким же образом, как и в категории (a), где X представляет собой требуемую длину пропуска, на которую не нацелена связывающаяся с мишенью последовательность. Для мишеней в категории (c) общую длину мишени делят на длину связывающейся с мишенью последовательности и округляют до ближайшего целочисленного значения, которое представляет число связывающихся с мишенью последовательностей, необходимое для полного нацеливания на все целевые последовательности для создания набора вставок для целевой последовательности. Необязательно, количество связывающихся с мишенью последовательностей может быть уменьшено, причем после уменьшения пропуска между связывающимися с мишенью последовательностями являются меньше требуемой длины пропуска Y. Этот общий процесс затем повторяют для каждой целевой последовательности, образуя библиотеку вставок. Связывающиеся с мишенью последовательности в библиотеке затем модифицируют путем добавления одной или нескольких нецелевых последовательностей, содержащих одну или несколько последовательностей праймера, библиотеку синтезируют на структуре, синтез выполняют с помощью фосфорамидитной химии с использованием общих способов из примера 3, и зонды метят молекулярным(и) маркером(ами).

[0293] Пример 23: Способ конструирования библиотеки перекрывающихся зондов

[0294] По меньшей мере 100 целевых последовательностей предоставляют и сортируют по отдельным категориям на основании длины по сравнению с требуемой длиной комплементарной связывающейся с мишенью последовательностью зонда. Например, категории включают в себя без ограничения следующее: (a) мишени короче, чем длина вставки, (b) мишени короче или равны длине вставки + X, и (c) мишени длиннее, чем длина вставки + X, где X представляет собой требуемую длину пропуска, на которую не нацелен зонд. На целевую последовательность в категории (a) нацелены связывающиеся с мишенью последовательности, которые либо расположены по центру, либо выровнены по левой или правой стороне целевой последовательности, в зависимости от сложности нецелевой области (повторяющаяся последовательность, последовательность с высоким/низким содержанием GC, палиндромная последовательность и т. д.), к которым также будет комплементарна вставка. На целевые последовательности в категории (b) нацелено воздействуют таким же образом, как и в категории (a), где X представляет собой

требуемую длину пропуска, на которую не нацелена связывающаяся с мишенью последовательность. Для мишеней в категории (с) общую длину мишени делят на длину связывающейся с мишенью последовательности и округляют до ближайшего целочисленного значения, которое представляет число связывающихся с мишенью последовательностей, необходимое для полного нацеливания на все целевые последовательности. Затем комплементарные связывающиеся с мишенью последовательности разносят по целевой последовательности (необязательно равномерно), позволяя перекрываться для создания набор вставок для целевой последовательности. Этот общий процесс затем повторяют для каждой целевой последовательности, образуя библиотеку вставок. Связывающиеся с мишенью последовательности в библиотеке затем модифицируют путем добавления одной или нескольких нецелевых последовательностей, содержащих одну или несколько последовательностей праймера, библиотеку синтезируют на структуре, синтез выполняют с помощью фосфорамидитной химии с использованием общих способов из примера 3, и зонды метят молекулярным(и) маркером(ами).

[0295] Пример 24: Способ конструирования смешанной библиотеки зондов

[0296] Библиотеку зондов синтезируют в соответствии с общим способом согласно примеру 22 и 23 с модификацией. Набор, содержащий неперекрывающиеся вставки, перекрывающиеся вставки или смешанные (перекрывающиеся и неперекрывающиеся) вставки, создают для каждой целевой последовательности.

[0297] Пример 25: Полинуклеотидные зонды для нацеленного воздействия на экзон

[0298] Полинуклеотидные зонды могут нацеливаться на экзоны в геноме, и ген может содержать множество экзонов. Например, ген лейкоцитарного антигена человека (HLA) содержит семь экзонов, три из которых приведены в таблице 6.

Таблица 6

ИД последов.	Описание	Длина экзона (п.н.)	Геномная последовательность, содержащая экзон (подчеркнут)
5	экзон 1 гена HLA человека	63	CCCACCGGGACTCAGATTCTCCCCAGACGCCG AGGATGGTGCTCATGGCGCCCCGAACCCTCCT CCTGCTGCTCTCAGGGGCCCTGGCCCTGACCCA GACCTGGGCGCGTGAGTGCAGGGTCTGCAGGG AAATGGGCGCGTGAGTGCAGGGTCTGC
6	экзон 2 гена HLA человека	270	GCTCCCAGGTTCCCACTCCATGAGGTATTTCTA CACCACCATGTCCCGGCCCGGCCGCGGGGAGC CCCGCTTCATCTCCGTCGGCTACGTGGACGATA CGCAGTTCGTGCGGTTTCGACAGCGACGACGCG AGTCCGAGAGAGGAGCCGCGGGCGCCGTGGAT GGAGCGGGAGGGGGCCAAAGTATTGGGACCGG AACACACAGATCTGCAAGGCCAGGCACAGAC

			<u>TGAACGAGAGAACCTGCGGATCGCGCTCCGCT</u> <u>ACTACAACCAGAGCGAGGGCGGTGAGTTGACC</u> <u>CCGG</u>
7	экзон 5 гена HLA человека	117	<u>TAGCAGGGTTCAGGGTTCCCTCACCTTCCCCCT</u> <u>TTTCCCAGCCATCTTCCCAGCCCACCGTCCCCA</u> <u>TCGTGGGCATCGTTGCTGGCTTGGTTCTACTTG</u> <u>TAGCTGTGGTCACTGGAGCTGTGGTTCGCTGCTG</u> <u>TAATGTGGAGGAAGAAGAGCTCAGGTAAGGA</u> <u>AGGGGT</u>

Для данного экзона различные комбинации связывающихся с мишенью последовательностей и не связывающихся с мишенью последовательностей можно использовать для конструирования зондов различных конфигураций и длины. Неограничивающие конструкции зондов, нацеленных на экзон 1 HLA показаны в качестве примеров в таблице 7, и показана последовательность только одной цепи зонда. Размер не связывающейся(ихся) с мишенью последовательности(ей), связывающейся с мишенью последовательности и общие длины зондов перечислены в таблице 8.

Таблица 7

ID послед.	Мишень	5' не связывающаяся с мишенью последовательность	Связывающаяся с мишенью последовательность	3' не связывающаяся с мишенью последовательность
8	экзон 1 HLA человека	<u>GTTACCCAAGA</u> <u>ACGCAGCTGAT</u> <u>TCTCCCCAGAC</u> <u>GCCGAGGATGG</u> <u>TGCTC</u>	ATGGCGCCCCGAACCCTCCT CCTGCTGCTCTCAGGGGCCC TGGCCCTGACCCAGACCTGG GCG	<u>CGTGAGTGCAG</u> <u>GGTCTGCAGGG</u> <u>AAATGGTAGTG</u> <u>TCGGAGGTCGT</u> <u>TCCT</u>
9	экзон 2 HLA человека	<u>GTTACCCAAGA</u> <u>ACGCAGCTG</u>	GGTTCCTCACTCCATGAGGTA TTTCTACACCACCATGTCCC GGCCCGGCGCCGGGGAGCC CCGCTTCATCTCCGTCGGCT ACGTGGACGATACGCAGTTC GTGCGGTTTCGACAGCGACGA C	<u>TAGTGTCGGAG</u> <u>GTCGTTTCCT</u>
10	экзон 2 HLA человека	<u>GTTACCCAAGA</u> <u>ACGCAGCTG</u>	TGGATGGAGCGGGAGGGGC CAAAGTATTGGGACCGGAAC ACACAGATCTGCAAGGCCCA GGCACAGACTGAACGAGAG AACCTGCGGATCGCGCTCCG CTACTACAACCAGAGCGAGG GC	<u>TAGTGTCGGAG</u> <u>GTCGTTTCCT</u>
11	экзон 2 HLA человека	<u>GTTACCCAAGA</u> <u>ACGCAGCTG</u>	GGCTACGTGGACGATACGCA GTTTCGTGCGGTTTCGACAGCG ACGACGCGAGTCCGAGAGA GGAGCCGCGGGCGCCGTGG ATGGAGCGGGAGGGGCCAA AGTATTGGGACCGGAACACA CAG	<u>TAGTGTCGGAG</u> <u>GTCGTTTCCT</u>

12	экзон 1 HLA человека	<u>GTTACCCAAGA</u> <u>ACGCAGCTG</u>	ATGGCGCCCCGAACCCTCCT CCTGCTGCTCTCAGGGGCCC TGGCCCTGACCCAGACCTGG GCG	<u>TAGTGTCGGAG</u> <u>GTCGTTTCCT</u>
13	экзон 2 HLA человека	AAAAAAAAAAAA AAAAAAAAAAAA <u>AAAGTTACCCA</u> <u>AGAACGCAGCT</u> <u>G</u>	GGTTCCCCTCCATGAGGTA TTTCTACACCACCATGTCCC GGCCCGGCGCCGGGGAGCC CCGCTTCATCTCCGTCGGCT ACGTGGACGATACGCAGTTC GTGCGGTTCGACAGCGACGA C	<u>TAGTGTCGGAG</u> <u>GTCGTTTCCTAAA</u> AAAAAAAAAAAA AAAAAAAAAAAA
14	экзон 2 HLA человека	<u>GTTACCCAAGA</u> <u>ACGCAGCTGAA</u> AAAAAAAAAAAA AAAAAAAAAAAA A	GGTTCCCCTCCATGAGGTA TTTCTACACCACCATGTCCC GGCCCGGCGCCGGGGAGCC CCGCTTCATCTCCGTCGGCT ACGTGGACGATACGCAGTTC GTGCGGTTCGACAGCGACGA C	AAAAAAAAAAAA AAAAAAAAAAAA AAATAGTGTCG <u>GAGGTCGTTTCCT</u>
15	экзон 5 HLA человека	<u>GTTACCCAAGA</u> <u>ACGCAGCTG</u>	AGCCATCTTCCCAGCCCACC GTCCCCATCGTGGGCATCGT TGCTGGCTTGTTTCTACTTGT A	<u>TAGTGTCGGAG</u> <u>GTCGTTTCCT</u>
16	экзон 5 HLA человека	<u>GTTACCCAAGA</u> <u>ACGCAGCTG</u>	GCTGTGGTCACTGGAGCTGT GGTCGCTGCTGTAATGTGGA GGAAGAAGAGCTCAG	<u>TAGTGTCGGAG</u> <u>GTCGTTTCCT</u>

Последовательность праймера не связывающейся(ихся) с мишенью последовательности(ей) подчеркнута.

Таблица 8

ID послед.	Мишень	Длина (п.н.) 5' не связывающейся с мишенью последовательности	Длина (п.н.) связывающейся с мишенью последовательности	Длина (п.н.) 3' не связывающейся с мишенью последовательности	Общая длина зонда (п.н.)
8	экзон 1 HLA человека	49	63	48	160
9	экзон 2 HLA человека	20	120	20	160
10	экзон 2 HLA человека	20	120	20	160
11	экзон 2 HLA человека	20	120	20	160

12	экзон 1 HLA человека	20	63	20	103
13	экзон 2 HLA человека	45	120	45	210
14	экзон 2 HLA человека	45	120	45	210
15	экзон 5 HLA человека	20	62	20	102
16	экзон 5 HLA человека	20	55	20	95

Различные расположения множества зондов можно использовать для покрытия данного экзона, например, экзона 2 HLA человека. Зонды, содержащие SEQ ID: 9 или 10, содержат набор и нацелены на экзон 2 HLA человека, но вместе оставляют пропуск в целевом экзоне и не содержат перекрывающихся последовательностей (см. фиг. 4F). Зонды, содержащие SEQ ID: 11, нацелены на экзон 2 HLA, и содержат связывающуюся с мишенью последовательность, которая перекрывается с SEQ ID: 9 и 10 (см. фиг. 4G). Зонды, каждый из которых содержит SEQ ID 15 или 16, содержат набор, нацелены на экзон 5 HLA человека и не нацелены на какие-либо перекрывающиеся области целевых или нецелевых областей. Зонды, соответствующие SEQ ID 8-16, синтезируются на структуре, и синтез проводят с помощью фосфорамидитной химии с использованием общих способов из примера 3 и необязательно метят по меньшей мере одного молекулярного маркера, такого как биотин.

[0299] Пример 26: Захват геномной ДНК с помощью библиотеки полинуклеотидных зондов

[0300] Нацеленную на полинуклеотиды библиотеку, содержащую по меньшей мере 500000 неидентичных полинуклеотидов, нацеленных на экзон человека, конструировали и синтезировали на структуре с помощью фосфорамидитной химии с использованием общих способов из примера 3 и стехиометрию контролировали с использованием общего способа согласно примеру 11 для создания библиотеки 4. Полинуклеотиды затем метили с помощью биотина и затем растворяли с образованием раствора библиотеки экзонных зондов. Высушенный пул индексной библиотеки получали из образцов геномной ДНК (гДНК) с использованием общего способа согласно примеру 16.

[0301] Раствор библиотеки экзомных зондов, гибридизационный раствор, смесь блокаторов А и смесь блокаторов В смешивали путем импульсного встряхивания в течение 2 секунд. Гибридизационный раствор нагревали при 65°C в течение 10 минут или до полного растворения всего осадка, а затем доводили до комнатной температуры на рабочем столе в течение 5 дополнительных минут. 20 мкл гибридизационного раствора и 4 мкл раствора библиотеки экзомных зондов добавляли в тонкостенную пробирку в стрипе для ПЦР объемом 0,2 мл и осторожно перемешивали пипетированием. Объединенный гибридизационный раствор/раствор экзомных зондов нагревали до 95°C в течение 2 минут в термоциклере с крышкой при 105°C и сразу же охлаждали на льду в течение по меньшей мере 10 минут. Затем раствору давали остыть до комнатной температуры на рабочем столе в течение 5 минут. Пока гибридизационный раствор/раствор библиотеки экзомных зондов охлаждали, к каждому образцу геномной ДНК добавляли воду по 9 мкл, и 5 мкл смеси блокаторов А и 2 мкл смеси блокаторов В добавляли к высушенному пулу индексной библиотеки в тонкостенной пробирке для ПЦР объемом 0,2 мл. Затем раствор смешивали путем осторожного пипетирования. Пробирку с объединенной библиотекой/блокатором нагревали при 95°C в течение 5 минут в термоциклере с крышкой при 105°C, затем доводили до комнатной температуры на рабочем столе в течение не более 5 минут, прежде чем переходить к следующей стадии. Гибридизационную смесь/раствор зондов смешивали пипеткой и добавляли ко всем 24 мкл пробирки с объединенной библиотекой/блокатором. Всю лунку с реакцией захвата хорошо перемешивали путем осторожного пипетирования, чтобы избежать образования пузырьков. Пробирку с образцом подвергали импульсному вращению, чтобы убедиться, что пробирка плотно закрыта. Реакцию захвата/гибридизации нагревали при 70°C в течение 16 часов в термоциклере для ПЦР с температурой крышки, составляющей 85°C.

[0302] Буфер для связывания, промывочный буфер 1 и промывочный буфер 2 нагревали при 48°C до тех пор, пока весь осадок не растворился в растворе. 700 мкл промывочного буфера 2 разделяли на аликвоты на одну реакцию захвата и предварительно нагревали до 48°C. Связывающие стрептавидин микрошарики и микрошарики для очистки ДНК уравнивали при комнатной температуре в течение по меньшей мере 30 минут. Полимеразу, такую как КАРА HiFi HotStart ReadyMix, и праймеры для амплификации размораживали на льду. После размораживания реагентов их перемешивали импульсным встряхиванием в течение 2 секунд. Готовили 500 мкл 80-процентного этанола на одну реакцию захвата. Связывающие стрептавидин микрошарики предварительно уравнивали при комнатной температуре и встряхивали до гомогенизации. 100 мкл связывающих стрептавидин микрошариков добавляли в чистую 1,5 мл

микроцентрифужную пробирку на одну реакцию захвата. 200 мкл буфера для связывания добавляли в каждую пробирку и каждую пробирку перемешивали путем пипетирования до гомогенизации. Пробирку размещали на магнитном штативе. Связывающие стрептавидин микрошарики осаждали в течение 1 минуты. Пробирку снимали и выбрасывали прозрачный супернатант, следя за тем, чтобы не нарушить осадок микрошариков. Пробирку снимали с магнитного штатива и промывки повторяли еще два раза. После третьей промывки пробирку убирали, а прозрачный супернатант выбрасывали. Добавляли последние 200 мкл буфера для связывания и микрошарики ресуспендировали путем встряхивания до гомогенного состояния.

[0303] После завершения реакции гибридизации крышку термоциклера открывали и весь объем реакции захвата быстро переносили (36-40 мкл) в промытые связывающие стрептавидин микрошарики. Смесь перемешивали в течение 30 минут при комнатной температуре на шейкере, шюттель-аппарате или вращающем устройстве со скоростью, достаточной для поддержания гомогенизации реакции захвата/раствора для связывающихся со стрептавидином микрошариков. Реакцию захвата/раствор для связывающихся со стрептавидином микрошариков удаляли из смесителя и центрифугировали с помощью импульсной центрифуги, чтобы весь раствор находился на дне пробирки. Образец помещали на магнитный штатив, и микрошарики, связывающиеся со стрептавидином, осаждали, оставляя прозрачный супернатант в течение 1 минуты. Прозрачный супернатант удаляли и выбрасывали. Пробирку снимали с магнитного штатива и добавляли 200 мкл промывочного буфера при комнатной температуре с последующим перемешиванием путем пипетирования до гомогенизации. Пробирку центрифугировали с помощью импульсной центрифуги для обеспечения того, чтобы весь раствор находился на дне пробирки. Термоциклер был запрограммирован со следующим условием (таблица 9).

[0304] Температуру нагретой крышки выставляли на 105°C.

Таблица 9

Стадия	Температура	Время	Число циклов
1	98°C	45 секунд	1
2	98°C	15 секунд	9
	60°C	30 секунд	
	72°C	30 секунд	
3	72°C	1 минута	1
4	4°C	фиксация	

[0305] Праймеры для амплификации (2,5 мкл) и полимеразу, такую как КАРА HiFi HotStart ReadyMix (25 мкл) добавляли к пробирке, содержащей густую суспензию воды/связывающих стрептавидин микрошариков, и содержимое пробирки перемешивали пипетированием. Пробирку затем разделяли на две реакции. Пробирку центрифугировали на импульсной центрифуге и переносили в термоциклер и начинали программу циклов согласно таблице 9. Когда программа термоциклера завершалась, образцы удаляли из блока и сразу же подвергали очистке. Микрошарики для очистки ДНК, предварительно уравновешенные при комнатной температуре, встряхивали до гомогенизации. 90 мкл (1,8х) гомогенизированных микрошариков для очистки ДНК добавляли в пробирку и хорошо перемешивали путем встряхивания. Пробирку инкубировали в течение 5 минут при комнатной температуре и помещали на магнитный штатив. Микрошарики для очистки ДНК осаждали, оставляя прозрачный супернатант в течение 1 минуты. Прозрачный супернатант выбрасывали, а пробирку оставляли на магнитном штативе. Осадок микрошариков для очистки ДНК промывали 200 мкл свежеприготовленного 80-процентного этанола, инкубировали в течение 1 минуты, затем удаляли и этанол выбрасывали. Промывку повторяли один раз, всего две промывки, держа пробирку на магнитном штативе. Весь оставшийся этанол удаляли и выбрасывали пипеткой по 10 мкл, следя за тем, чтобы не нарушать осадок микрошариков для очистки ДНК. Осадок микрошариков для очистки ДНК сушили на воздухе на магнитном штативе в течение 5-10 минут или до тех пор, пока осадок не становился сухим. Пробирку вынимали из магнитного штатива и добавляли 32 мкл воды, перемешивали пипетированием до гомогенизации и инкубировали при комнатной температуре в течение 2 минут. Пробирку помещали на магнитный штатив на 3 минуты или до полного осаждения микрошариков. 30 мкл прозрачного супернатанта извлекали и переносили в чистую тонкостенную пробирку в стрипе для ПЦР объемом 0,2 мл, следя за тем, чтобы не нарушать осадок микрошариков для очистки ДНК. Средняя длина фрагмента составляла от приблизительно 375 п.н. - приблизительно 425 п.н. с использованием диапазона настройки от 150 до 1000 п.н. на приборе для анализа. В идеале конечные значения концентрации составляют по меньшей мере приблизительно 15 нг/мкл. Каждый захват определяли количественно и подтверждали с использованием секвенирования нового поколения (NGS).

[0306] Обобщенно показатели NGS представлены в таблице 10, таблице 11 и на фиг. 37 в сравнении с набором сравнения для захвата экзона (набор сравнения D). Библиотека 4 содержит зонды, которые соответствуют повышенному процентному отношению мишеней-экзонов, чем набор сравнения D. Это приводит к меньшему секвенированию для

получения сравнимого качества и покрытия целевых последовательностей с использованием библиотека 4.

Таблица 10

Показатель NGS	Набор сравнения D	Библиотека 4
Длина мишени	38,8 млн.п.н.	33,2 млн.п.н.
Длина зонда	50,8 млн.п.н.	36,7 млн.п.н.
Эффективность конструкции зонда	76,5%	90,3%
Одновременный захват	8 одновременно	8 одновременно
Прочтения PF	57,7 М	49,3 М
Нормализованное покрытие	150X	150X
Размер библиотеки HS	30,3 М	404,0 М
Процентное отношение дубликаций	32,5%	2,5%
Кратность обогащения	43,2	48,6
Штраф на кратность увеличения покрытия для 80% оснований (fold 80 base penalty)	1,84	1,40

Таблица 11

Показатель NGS	Набор сравнения D	Библиотека 4
Процентное отношение уникальных прочтений, прошедших через фильтр контроля качества (PCT_PF_UQ_READS)	67,6%	97,5%
Процентное отношение целевых оснований при 1X	99,8%	99,8%
Процентное отношение целевых оснований при 20X	90,3%	99,3%
Процентное отношение целевых оснований при 30X	72,4%	96,2%

[0307] Сравнение перекрывающихся целевых областей как для набора D, так и для библиотеки 4 (общее число прочтений, нормализованных к 96X покрытию) показано в таблице 12 и на фиг. 37. Библиотеку 4 обрабатывали в виде 8 образцов на гибридизацию, а набор D в виде 2 образцов на гибридизацию. Кроме того, для обеих библиотек сравнивали однонуклеотидный полиморфизм и сигналы делеции в рамке считывания из перекрывающихся областей с областями с высокой достоверностью,

идентифицированными из справочных данных NA12878 «Genome in a Bottle» (таблица 13). Библиотека 4 функционировала аналогично или лучше (с более высокой точностью инделей), чем набор D, при идентификации SNP и инделей.

Таблица 12

Показатель NGS	Набор сравнения D	Библиотека 4
Процентное отношение уникальных прочтений, прошедших через фильтр контроля качества (PCT_PF_UQ_READS)	94,60%	97,7%
Процентное отношение выбранных оснований	79%	80%
Процентное отношение целевых оснований при 1X	100%	100%
Процентное отношение целевых оснований при 20X	90%	96%
Процентное отношение целевых оснований при 30X	71%	77%
Кратность обогащения	44,9	49,9
Штраф на кратность увеличения покрытия для 80% оснований (fold 80 base penalty)	1,76	1,4
Размер библиотеки HS	122 М	267 М

Таблица 13

Варианты	Набор сравнения D		Библиотека 4	
	Точность	Чувствительность	Точность	Чувствительность
Однонуклеотидные полиморфизмы (SNP)	98,59%	99,23%	99,05%	99,27%
Делеции в рамке считывания (индел)	76,42%	94,12%	87,76%	94,85%
Всего	98,14%	99,15%	98,85%	99,20%

[0308] Точность представляет отношение истинно положительных сигналов к общему количеству (истинных и ложных) положительных сигналов. Чувствительность представляет собой отношение истинно положительных сигналов к общим истинным значениям (истинно положительным и ложно отрицательным).

[0309] Хотя предпочтительные варианты осуществления настоящего изобретения были показаны и описаны в настоящем документе, 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 изобретения. Следует понимать, что различные альтернативы вариантам осуществления настоящего изобретения, описанным в настоящем документе, можно использовать при практическом применении настоящего изобретения. Подразумевается, что приведенная ниже формула изобретения определяет объем настоящего изобретения и что таким образом будут охватываться способы и структуры в пределах объема этой формулы изобретения и их эквиваленты.

ПЕРЕЧЕНЬ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

<110> ТВИСТ БАЙОСАЙЕНС КОРПОРЕЙШН

<120> БИБЛИОТЕКИ ПОЛИНУКЛЕОТИДОВ, ХАРАКТЕРИЗУЮЩИЕСЯ КОНТРОЛИРУЕМОЙ
СТЕХИОМЕТРИЕЙ, И ИХ СИНТЕЗ

<130> 44854-730.601

<140>

<141>

<150> 62/558666

<151> 2017-09-14

<150> 62/548307

<151> 2017-08-21

<150> 62/424302

<151> 2016-11-18

<160> 16

<170> PatentIn версии 3.5

<210> 1

<211> 62

<212> DNA

<213> Искусственная последовательность

<220>

<223> Описание искусственной последовательности: синтетический
олигонуклеотид

<220>

<221> modified_base

<222> (51)..(52)

<223> Тимидин-сукцинилгексамид-CED-фосфорамидит

<400> 1

agacaatcaa ccatttgggg tggacagcct tgacctctag acttcggcat tttttttttt 60

tt 62

<210> 2

<211> 112

<212> DNA

<213> Искусственная последовательность

<220>

<223> Описание искусственной последовательности: синтетический
полинуклеотид

<220>

<221> modified_base

<222> (101)..(102)

<223> Тимидин-сукцинилгексамид-CED-фосфорамидит

<400> 2

cgggatcctt atcgatcatcg tcgtacagat cccgacccat ttgctgtcca ccagtcatgc 60

tagccataacc atgatgatga tgatgatgag aacccccgcat tttttttttt tt	112
<210> 3	
<211> 19	
<212> DNA	
<213> Искусственная последовательность	
<220>	
<223> Описание искусственной последовательности: синтетический праймер	
<400> 3	
atgcgggggtt ctcacatc	19
<210> 4	
<211> 20	
<212> DNA	
<213> Искусственная последовательность	
<220>	
<223> Описание искусственной последовательности: синтетический праймер	
<400> 4	
cgggatacctt atcgatcatc	20
<210> 5	
<211> 156	
<212> DNA	
<213> Homo sapiens	
<400> 5	
cccaccggga ctcagattct ccccagacgc cgaggatggt gctcatggcg ccccgaaacc	60
tcctcctgct gctctcaggg gccctggccc tgaccagac ctgggcgcgt gagtgcaggg	120
tctgcaggga aatgggcgcg tgagtgcagg gtctgc	156
<210> 6	
<211> 293	
<212> DNA	
<213> Homo sapiens	
<400> 6	
gctcccaggt tccactcca tgaggtatth ctacaccacc atgtcccggc ccggcgccgg	60
ggagccccgc ttcattctcg tcggctacgt ggacgatacg cagttcgtgc ggttcgacag	120
cgacgacgcg agtccgagag aggagccgcg ggcgcctgg atggagcggg agggggccaaa	180
gtattgggac cggaacacac agatctgcaa ggcccaggca cagactgaac gagagaacct	240
gcggatcgcg ctccgctact acaaccagag cgagggcggt gaggtagacc cgg	293
<210> 7	
<211> 169	
<212> DNA	
<213> Homo sapiens	

<400> 7
tagcagggtc aggggttcct caccttcccc ctttttccca gccatcttcc cagcccaccg 60
tccccatcgt gggcatcggt gctggcttgg ttctacttgt agctgtggtc actggagctg 120
tggtcgctgc tgtaatgtgg aggaagaaga gctcaggtaa ggaaggggt 169

<210> 8
<211> 160
<212> DNA
<213> Homo sapiens

<400> 8
gttacccaag aacgcagctg attctcccca gacgccgagg atgggtgctca tggcgccccg 60
aaccctcctc ctgctgctct cagggggcct ggccctgacc cagacctggg cgcgtgagtg 120
caggggtctgc agggaaatgg tagtgctcga ggtcgttcct 160

<210> 9
<211> 160
<212> DNA
<213> Homo sapiens

<400> 9
gttacccaag aacgcagctg ggttcccact ccatgaggta tttctacacc accatgtccc 60
ggcccggcgc cggggagccc cgcttcatct ccgtcggcta cgtggacgat acgcagttcg 120
tgcggttcga cagcgacgac tagtgctcga ggtcgttcct 160

<210> 10
<211> 160
<212> DNA
<213> Homo sapiens

<400> 10
gttacccaag aacgcagctg tggatggagc gggaggggcc aaagtattgg gaccggaaca 60
cacagatctg caaggccag gcacagactg aacgagagaa cctgcggatc gcgctccgct 120
actacaacca gagcgagggc tagtgctcga ggtcgttcct 160

<210> 11
<211> 160
<212> DNA
<213> Homo sapiens

<400> 11
gttacccaag aacgcagctg ggctacgtgg acgatacgca gttcgtgcgg ttcgacagcg 60
acgacgcgag tccgagagag gagccgcggg cgccgtggat ggagcgggag gggccaaagt 120
attgggaccg gaacacacag tagtgctcga ggtcgttcct 160

<210> 12
<211> 103
<212> DNA

<213> Homo sapiens

<400> 12
gttacccaag aacgcagctg atggcgcccc gaaccctcct cctgctgctc tcaggggccc 60
tggccctgac ccagacctgg gcgtagtgct ggaggtcgctt cct 103

<210> 13
<211> 210
<212> DNA
<213> Homo sapiens

<400> 13
aaaaaaaaaa aaaaaaaaaa aaaaagttac ccaagaacgc agctggggttc ccactccatg 60
aggtattttct acaccaccat gtcccggccc ggcgccgggg agccccgctt catctccgctc 120
ggctacgtgg acgatacgca gttcgtgcgg ttcgacagcg acgactagtg tcggaggtcg 180
ttcctaataaa aaaaaaaaaa aaaaaaaaaa 210

<210> 14
<211> 210
<212> DNA
<213> Homo sapiens

<400> 14
gttacccaag aacgcagctg aaaaaaaaaa aaaaaaaaaa aaaaagggttc ccactccatg 60
aggtattttct acaccaccat gtcccggccc ggcgccgggg agccccgctt catctccgctc 120
ggctacgtgg acgatacgca gttcgtgcgg ttcgacagcg acgacaaaaa aaaaaaaaaa 180
aaaaaaaaaa tagtgctcga ggtcgttcct 210

<210> 15
<211> 102
<212> DNA
<213> Homo sapiens

<400> 15
gttacccaag aacgcagctg agccatcttc ccagcccacc gtcccctatg tgggcacgtg 60
tgctggcttg gttctacttg tatagtgtcg gaggtcgctt ct 102

<210> 16
<211> 95
<212> DNA
<213> Homo sapiens

<400> 16
gttacccaag aacgcagctg gctgtgggtc ctggagctgt ggtcgctgct gtaatgtgga 60
ggaagaagag ctcagtagtg tcggaggtcg ttctt 95

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Библиотека полинуклеотидов, причем библиотека полинуклеотидов содержит по меньшей мере 5000 полинуклеотидов, причем каждый из по меньшей мере 5000 полинуклеотидов присутствует в таком количестве, что после гибридизации с геномными фрагментами и секвенирования гибридизованных геномных фрагментов библиотека полинуклеотидов обеспечивает по меньшей мере 30-кратную глубину прочтения по меньшей мере 90 процентов оснований геномных фрагментов при условиях для вплоть до 55-кратной теоретической глубины прочтения для оснований геномных фрагментов.

2. Библиотека полинуклеотидов по п. 1, причем библиотека полинуклеотидов обеспечивает по меньшей мере 30-кратную глубину прочтения по меньшей мере 95 процентов оснований геномных фрагментов при условиях для вплоть до 55-кратной теоретической глубины прочтения для оснований геномных фрагментов.

3. Библиотека полинуклеотидов по п. 1, причем библиотека полинуклеотидов обеспечивает по меньшей мере 30-кратную глубину прочтения по меньшей мере 98 процентов оснований геномных фрагментов при условиях для вплоть до 55-кратной теоретической глубины прочтения для оснований геномных фрагментов.

4. Библиотека полинуклеотидов по п. 1, причем библиотека полинуклеотидов обеспечивает по меньшей мере 90 процентов уникальных прочтений для оснований геномных фрагментов.

5. Библиотека полинуклеотидов по п. 1, причем библиотека полинуклеотидов обеспечивает по меньшей мере 95 процентов уникальных прочтений для оснований геномных фрагментов.

6. Библиотека полинуклеотидов по п. 1, причем библиотека полинуклеотидов обеспечивает по меньшей мере 90 процентов оснований геномных фрагментов с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения.

7. Библиотека полинуклеотидов по п. 1, причем библиотека полинуклеотидов обеспечивает по меньшей мере 95 процентов оснований геномных фрагментов с глубиной

прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения.

8. Библиотека полинуклеотидов по п. 0, причем библиотека полинуклеотидов обеспечивает по меньшей мере 90 процентов геномных фрагментов с процентным отношением GC, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов, с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения.

9. Библиотека полинуклеотидов по п. 0, причем библиотека полинуклеотидов обеспечивает по меньшей мере приблизительно 80 процентов геномных фрагментов с процентным отношением повторяющейся последовательности или последовательности вторичной структуры, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов, с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения.

10. Библиотека полинуклеотидов по п. 1, в которой каждый из геномных фрагментов составляет приблизительно 100 оснований - приблизительно 500 оснований в длину.

11. Библиотека полинуклеотидов по п. 1, в которой по меньшей мере приблизительно 80 процентов по меньшей мере из 5000 полинуклеотидов представлены в количестве в пределах по меньшей мере приблизительно 1,5-кратного по отношению к среднему представлению для библиотеки полинуклеотидов.

12. Библиотека полинуклеотидов по п. 1, в которой по меньшей мере 30 процентов по меньшей мере из 5000 полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов.

13. Библиотека полинуклеотидов по п. 1, в которой по меньшей мере приблизительно 15 процентов по меньшей мере из 5000 полинуклеотидов содержат полинуклеотиды с процентным отношением повторяющейся последовательности или последовательности вторичной структуры, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов.

14. Библиотека полинуклеотидов по п. 1, в которой по меньшей мере 5000 полинуклеотидов кодируют по меньшей мере 1000 генов.

15. Библиотека полинуклеотидов по п. 1, причем библиотека полинуклеотидов содержит по меньшей мере 100000 полинуклеотидов.

16. Библиотека полинуклеотидов по п. 1, причем библиотека полинуклеотидов содержит по меньшей мере 700000 полинуклеотидов.

17. Библиотека полинуклеотидов по п. 1, в которой по меньшей мере 5000 полинуклеотидов содержат по меньшей мере одну последовательность экзона.

18. Библиотека полинуклеотидов по п. 16, в которой по меньшей мере 700000 полинуклеотидов содержат по меньшей мере один набор полинуклеотидов, в совокупности содержащих одну последовательность экзона.

19. Библиотека полинуклеотидов по п. 18, в которой по меньшей мере 700000 полинуклеотидов содержат по меньшей мере 150000 наборов.

20. Библиотека полинуклеотидов, причем библиотека полинуклеотидов содержит по меньшей мере 5000 полинуклеотидов, причем каждый из полинуклеотидов составляет приблизительно 20 - 200 оснований в длину, причем множество полинуклеотидов кодируют последовательности из каждого экзона по меньшей мере для 1000 предварительно выбранных генов, причем каждый полинуклеотид содержит молекулярный маркер, причем каждый из по меньшей мере 5000 полинуклеотидов присутствуют в таком количестве, что после гибридизации с геномными фрагментами и секвенирования гибридизованных геномных фрагментов библиотека полинуклеотидов обеспечивает по меньшей мере 30-кратную глубину прочтения по меньшей мере 90 процентов оснований геномных фрагментов при условиях для вплоть до 55-кратной теоретической глубины прочтения для оснований геномных фрагментов.

21. Библиотека полинуклеотидов по п. 20, причем библиотека полинуклеотидов обеспечивает по меньшей мере 30-кратную глубину прочтения по меньшей мере 95 процентов оснований геномных фрагментов при условиях для вплоть до 55-кратной теоретической глубины прочтения для оснований геномных фрагментов.

22. Библиотека полинуклеотидов по п. 20, причем библиотека полинуклеотидов обеспечивает по меньшей мере 30-кратную глубину прочтения по меньшей мере 98 процентов оснований геномных фрагментов при условиях для вплоть до 55-кратной теоретической глубины прочтения для оснований геномных фрагментов.

23. Библиотека полинуклеотидов по п. 20, причем библиотека полинуклеотидов обеспечивает по меньшей мере 90 процентов уникальных прочтений для оснований геномных фрагментов.

24. Библиотека полинуклеотидов по п. 20, причем библиотека полинуклеотидов обеспечивает по меньшей мере 95 процентов уникальных прочтений для оснований геномных фрагментов.

25. Библиотека полинуклеотидов по п. 20, причем библиотека полинуклеотидов обеспечивает по меньшей мере 90 процентов оснований геномных фрагментов с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения.

26. Библиотека полинуклеотидов по п. 20, причем библиотека полинуклеотидов обеспечивает по меньшей мере 95 процентов оснований геномных фрагментов с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения.

27. Библиотека полинуклеотидов по п. 20, причем библиотека полинуклеотидов обеспечивает больше чем 90 процентов геномных фрагментов с процентным отношением GC, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов, с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения.

28. Библиотека полинуклеотидов по п. 20, причем библиотека полинуклеотидов обеспечивает больше чем приблизительно 80 процентов геномных фрагментов с процентным отношением повторяющейся последовательности или последовательности вторичной структуры, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов, с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения.

29. Библиотека полинуклеотидов по п. 20, в которой каждый из геномных фрагментов составляет приблизительно 100 оснований - приблизительно 500 оснований в длину.

30. Библиотека полинуклеотидов по п. 20, в которой больше чем приблизительно 80 процентов по меньшей мере из 5000 полинуклеотидов представлены в количестве в пределах по меньшей мере приблизительно 1,5-кратного по отношению к среднему представлению для библиотеки полинуклеотидов.

31. Библиотека полинуклеотидов по п. 20, в которой больше чем 30 процентов по меньшей мере из 5000 полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов.

32. Библиотека полинуклеотидов по п. 20, в которой больше чем приблизительно 15 процентов по меньшей мере из 5000 полинуклеотидов содержат полинуклеотиды с процентным отношением повторяющейся последовательности или последовательности вторичной структуры, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов.

33. Библиотека полинуклеотидов по п. 20, причем библиотека полинуклеотидов содержит по меньшей мере 100000 полинуклеотидов.

34. Библиотека полинуклеотидов по п. 20, причем библиотека полинуклеотидов содержит по меньшей мере 700000 полинуклеотидов.

35. Библиотека полинуклеотидов по п. 34, в которой по меньшей мере 700000 полинуклеотидов содержат по меньшей мере один набор полинуклеотидов, в совокупности содержащих одну последовательность экзона.

36. Библиотека полинуклеотидов по п. 35, в которой по меньшей мере 700000 полинуклеотидов содержат по меньшей мере 150000 наборов.

37. Способ создания библиотеки полинуклеотидов, причем способ предусматривает следующее:

а. получение предварительно определенных последовательностей, кодирующих по меньшей мере 5000 полинуклеотидов;

б. синтезирование по меньшей мере 5000 полинуклеотидов; и

с. амплификация по меньшей мере 5000 полинуклеотидов с помощью полимеразы с образованием библиотеки полинуклеотидов, в которой больше чем приблизительно 80 процентов по меньшей мере из 5000 полинуклеотидов представлены в количестве в пределах по меньшей мере приблизительно 2-кратного по отношению к среднему представлению для библиотеки полинуклеотидов.

38. Способ по п. 37, при котором больше чем приблизительно 80 процентов по меньшей мере из 5000 полинуклеотидов представлены в количестве в пределах по меньшей мере приблизительно 1,5-кратного по отношению к среднему представлению для библиотеки полинуклеотидов.

39. Способ по п. 37, при котором больше чем 30 процентов по меньшей мере из 5000 полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов.

40. Способ по п. 37, при котором больше чем приблизительно 15 процентов по меньшей мере из 5000 полинуклеотидов содержат полинуклеотиды с процентным отношением повторяющейся последовательности или последовательности вторичной структуры, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов.

41. Способ по п. 37, при котором библиотека полинуклеотидов характеризуется суммарной частотой ошибок, составляющей меньше чем 1 на 800 оснований по сравнению с предварительно определенными последовательностями без исправления ошибок.

42. Способ по п. 37, при котором предварительно определенные последовательности кодируют по меньшей мере 700000 полинуклеотидов.

43. Способ по п. 37, при котором синтез по меньшей мере 5000 полинуклеотидов происходит на структуре, характеризующейся поверхностью, причем поверхность содержит множество кластеров, причем каждый кластер содержит множество локусов; и причем каждый из по меньшей мере 5000 полинуклеотидов удлиняется из различного локуса множества локусов.

44. Способ по п. 43, при котором множество локусов содержит вплоть до 1000 локусов на кластер.

45. Способ по п. 43, при котором множество локусов содержит вплоть до 200 локусов на кластер.

46. Способ амплификации библиотеки полинуклеотидов, причем способ предусматривает следующее:

а. получение распределения амплификации по меньшей мере для 5000 полинуклеотидов;

б. кластеризация по меньшей мере 5000 полинуклеотидов из распределения амплификации на два или больше бинов на основании по меньшей мере одного признака последовательности, причем признак последовательности представляет собой процентное содержание GC, процентное содержание повторяющейся последовательности или процентное содержание вторичной структуры;

с. коррекция относительной частоты полинуклеотидов по меньшей мере в одном бине для создания библиотеки полинуклеотидов, характеризующейся предварительно выбранным представлением;

д. синтезирование библиотеки полинуклеотидов, характеризующейся предварительно выбранным представлением; и

е. амплификация библиотеки полинуклеотидов, характеризующейся предварительно выбранным представлением.

47. Способ по п. 46, при котором по меньшей мере один признак последовательности представляет собой процентное содержание GC.

48. Способ по п. 46, при котором по меньшей мере один признак последовательности представляет собой процентное содержание вторичной структуры.

49. Способ по п. 46, при котором по меньшей мере один признак последовательности представляет собой процентное содержание повторяющейся последовательности.

50. Способ по п. 49, при котором содержание повторяющейся последовательности содержит последовательности с 3 или больше аденинами.

51. Способ по п. 49, при котором содержание повторяющейся последовательности содержит повторяющиеся последовательности по меньшей мере на одном конце полинуклеотида.

52. Способ по п. 46, при котором указанные полинуклеотиды подвергают кластеризации в бины на основании аффинности одной или нескольких полинуклеотидных последовательностей в отношении связывания с целевой последовательностью.

53. Способ по п. 46, при котором число последовательностей в нижних 30 процентах бинов характеризуется по меньшей мере на 50 процентов большим представлением в последующем применении после коррекции по сравнению с числом последовательностей в нижних 30 процентах бинов до коррекции.

54. Способ по п. 46, при котором число последовательностей в верхних 30 процентах бинов характеризуется по меньшей мере на 50 процентов большим представлением в последующем применении после коррекции по сравнению с числом последовательностей в верхних 30 процентах бинов до коррекции.

55. Способ секвенирования геномной ДНК, предусматривающий следующее:

(а) приведение в контакт библиотеки по любому из пп. 1-36 с множеством геномных фрагментов;

(b) обогащение по меньшей мере одного геномного фрагмента, который связывается с библиотекой, для создания по меньшей мере одного обогащенного целевого полинуклеотида; и

(с) секвенирование по меньшей мере одного обогащенного целевого полинуклеотида.

56. Способ по п. 55, при котором множество обогащенных целевых полинуклеотидов содержит библиотеку кДНК.

57. Способ по п. 55, при котором длина по меньшей мере 5000 полинуклеотидов составляет приблизительно 80 - приблизительно 200 оснований.

58. Способ по п. 55, при котором каждый из геномных фрагментов составляет приблизительно 100 оснований - приблизительно 500 оснований в длину.

59. Способ по п. 55, при котором приведение в контакт происходит в растворе.

60. Способ по п. 55, при котором по меньшей мере 5000 полинуклеотидов являются по меньшей мере частично комплементарными геномным фрагментам.

61. Способ по п. 55, при котором выделение предусматривает (i) захват гибридизационных пар полинуклеотид/геномный фрагмент на твердой подложке; и (ii) высвобождение множества геномных фрагментов для создания обогащенных целевых полинуклеотидов.

62. Способ по п. 55, при котором секвенирование дает в результате по меньшей мере 30-кратную глубину прочтения по меньшей мере 95 процентов оснований геномных фрагментов при условиях для вплоть до 55-кратной теоретической глубины прочтения для оснований геномных фрагментов.

63. Способ по п. 55, при котором секвенирование дает в результате по меньшей мере 30-кратную глубину прочтения по меньшей мере 98 процентов оснований геномных фрагментов при условиях для вплоть до 55-кратной теоретической глубины прочтения для оснований геномных фрагментов.

64. Способ по п. 55, при котором секвенирование дает в результате по меньшей мере 90 процентов уникальных прочтений для оснований геномных фрагментов.

65. Способ по п. 55, при котором секвенирование дает в результате по меньшей мере 95 процентов уникальных прочтений для оснований геномных фрагментов.

66. Способ по п. 55, при котором секвенирование дает в результате по меньшей мере 90 процентов оснований геномных фрагментов с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения.

67. Способ по п. 55, при котором секвенирование дает в результате по меньшей мере 95 процентов оснований геномных фрагментов с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения.

68. Способ по п. 55, при котором секвенирование дает в результате по меньшей мере 90 процентов геномных фрагментов с процентным отношением GC, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов, с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения.

69. Способ по п. 55, при котором секвенирование дает в результате по меньшей мере приблизительно 80 процентов геномных фрагментов с процентным отношением повторяющейся последовательности или последовательности вторичной структуры, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов, с глубиной прочтения в пределах приблизительно 1,5-кратной по отношению к средней глубине прочтения.

70. Способ по п. 55, при котором по меньшей мере приблизительно 80 процентов по меньшей мере из 5000 полинуклеотидов представлены в количестве в пределах по меньшей мере приблизительно 1,5-кратного по отношению к среднему представлению для библиотеки полинуклеотидов.

71. Способ по п. 55, при котором по меньшей мере 30 процентов по меньшей мере из 5000 полинуклеотидов содержат полинуклеотиды с процентным отношением GC, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов.

72. Способ по п. 55, при котором по меньшей мере 15 процентов по меньшей мере из 5000 полинуклеотидов содержат полинуклеотиды с процентным отношением повторяющейся последовательности или последовательности вторичной структуры, составляющим от 10 процентов до 30 процентов или от 70 процентов до 90 процентов.

73. Способ по п. 55, при котором по меньшей мере 5000 полинуклеотидов кодируют по меньшей мере 1000 генов.

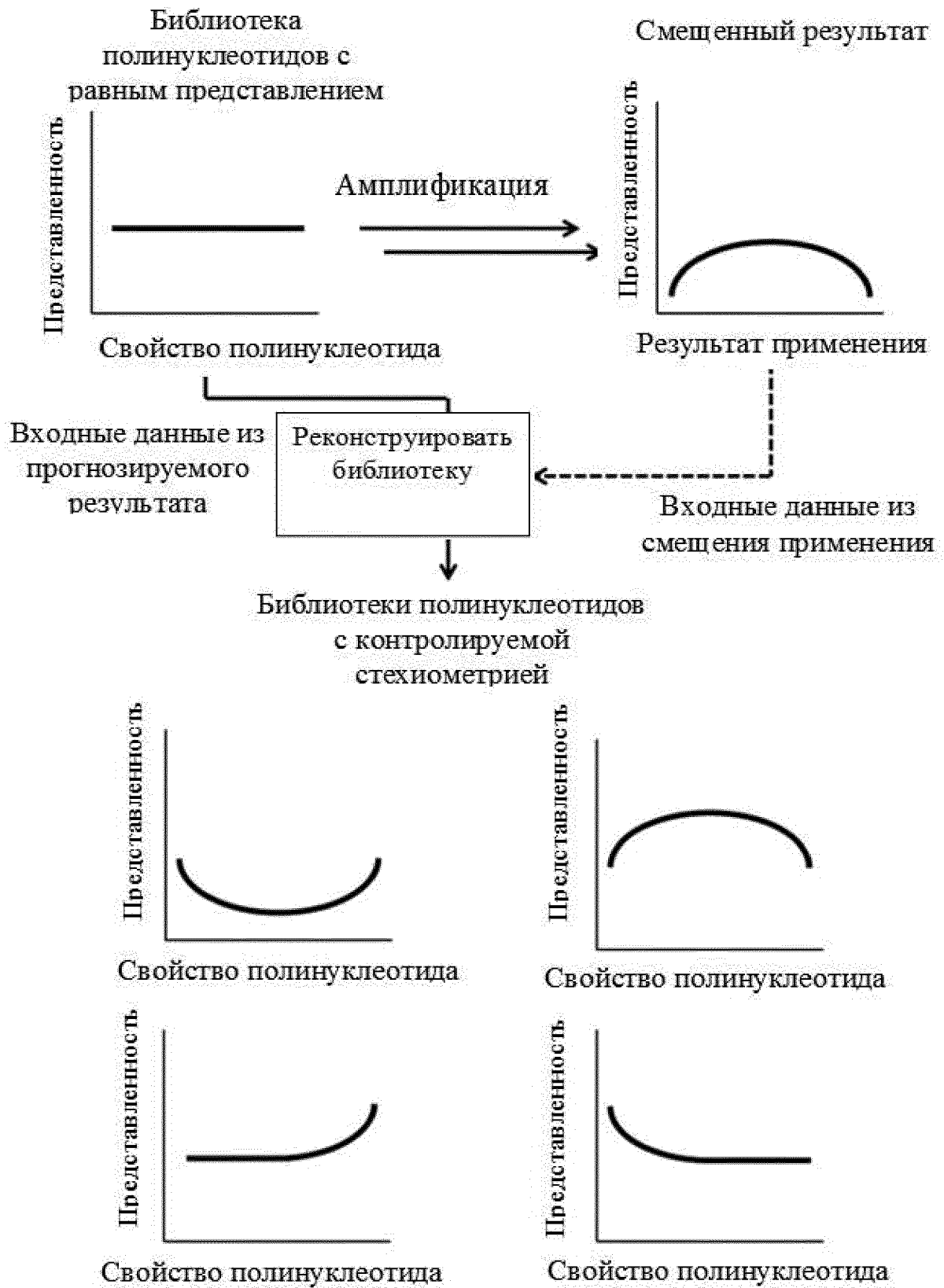
74. Способ по п. 55, при котором библиотека полинуклеотидов содержит по меньшей мере 100000 полинуклеотидов.

75. Способ по п. 55, при котором библиотека полинуклеотидов содержит по меньшей мере 700000 полинуклеотидов.

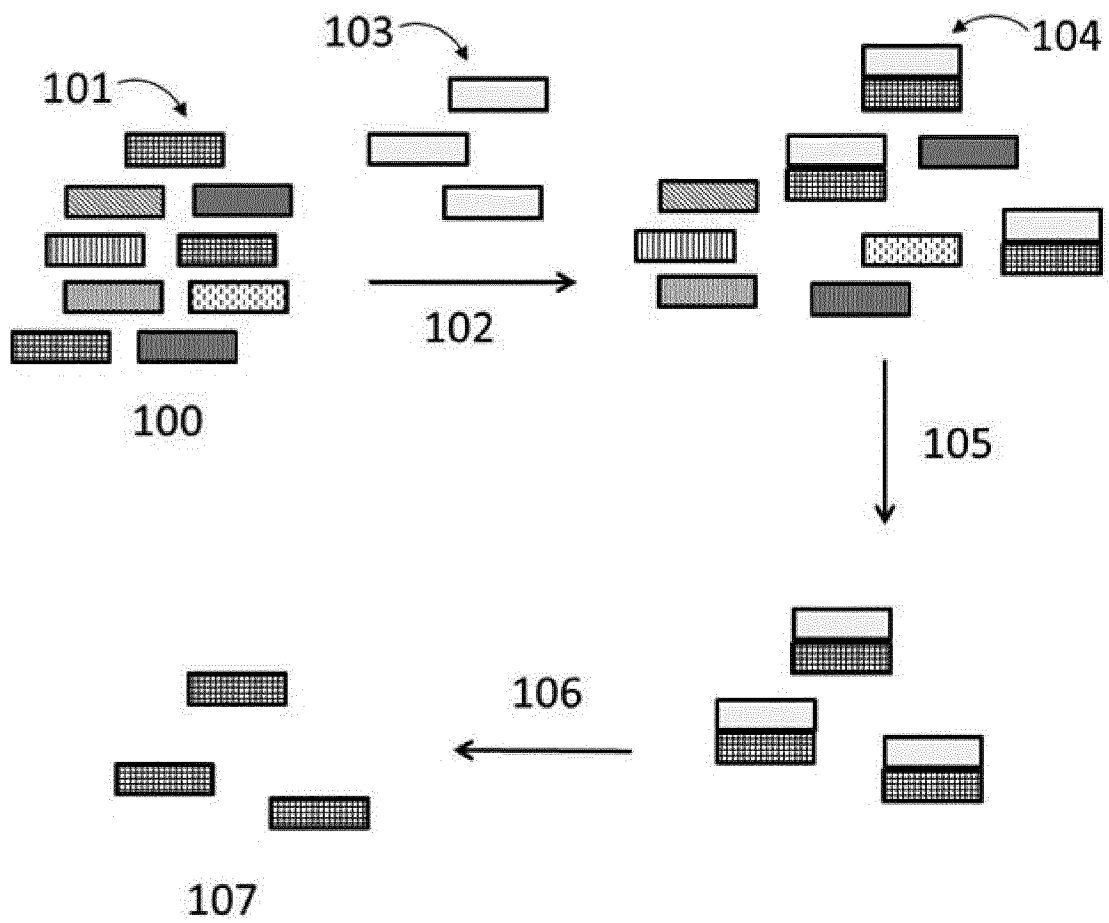
76. Способ по п. 55, при котором по меньшей мере 5000 полинуклеотидов содержат по меньшей мере одну последовательность экзона.

77. Способ по п. 75, при котором по меньшей мере 700000 полинуклеотидов содержат по меньшей мере один набор полинуклеотидов, в совокупности содержащих одну последовательность экзона.

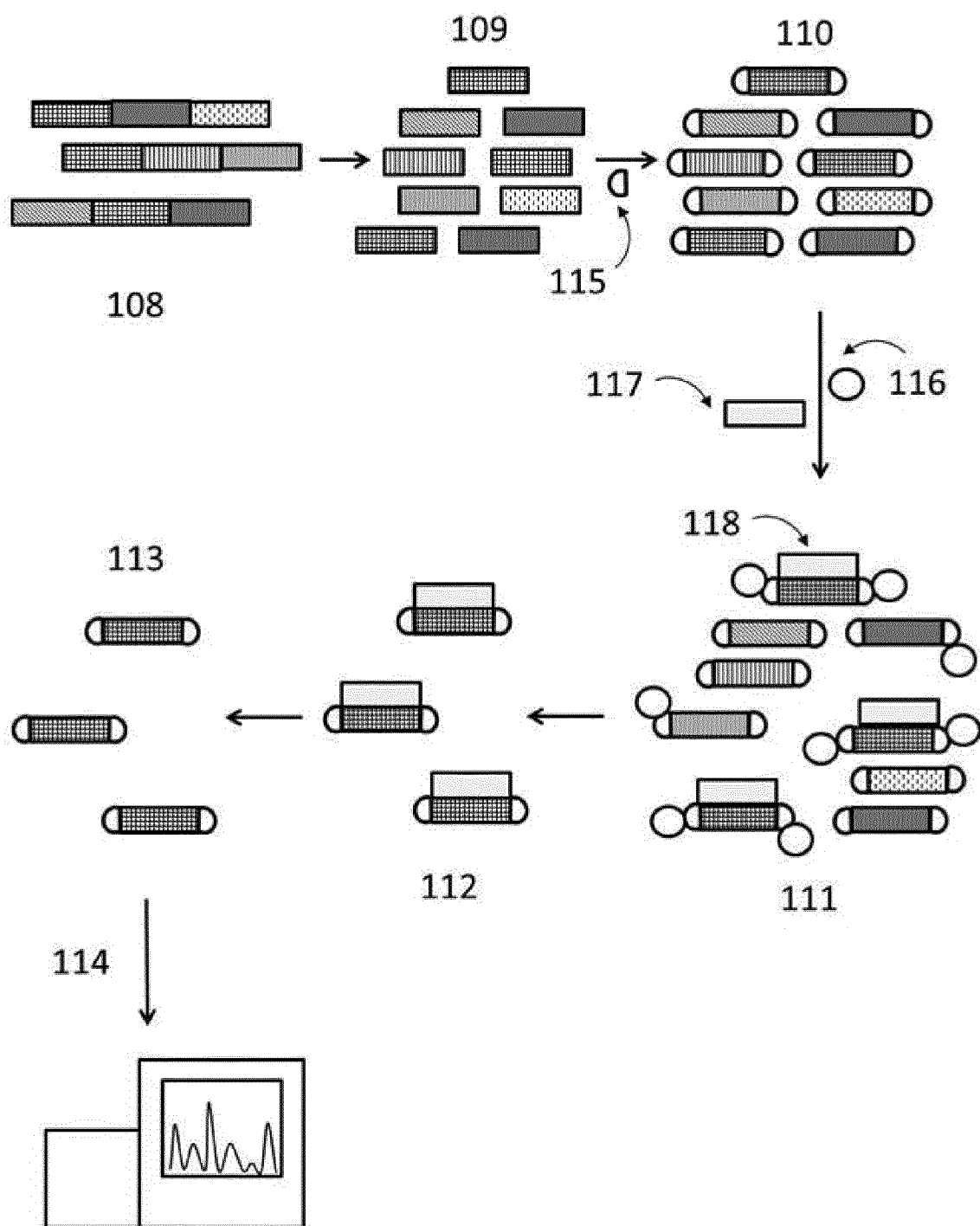
78. Способ по п. 77, при котором по меньшей мере 700000 полинуклеотидов содержат по меньшей мере 150000 наборов.



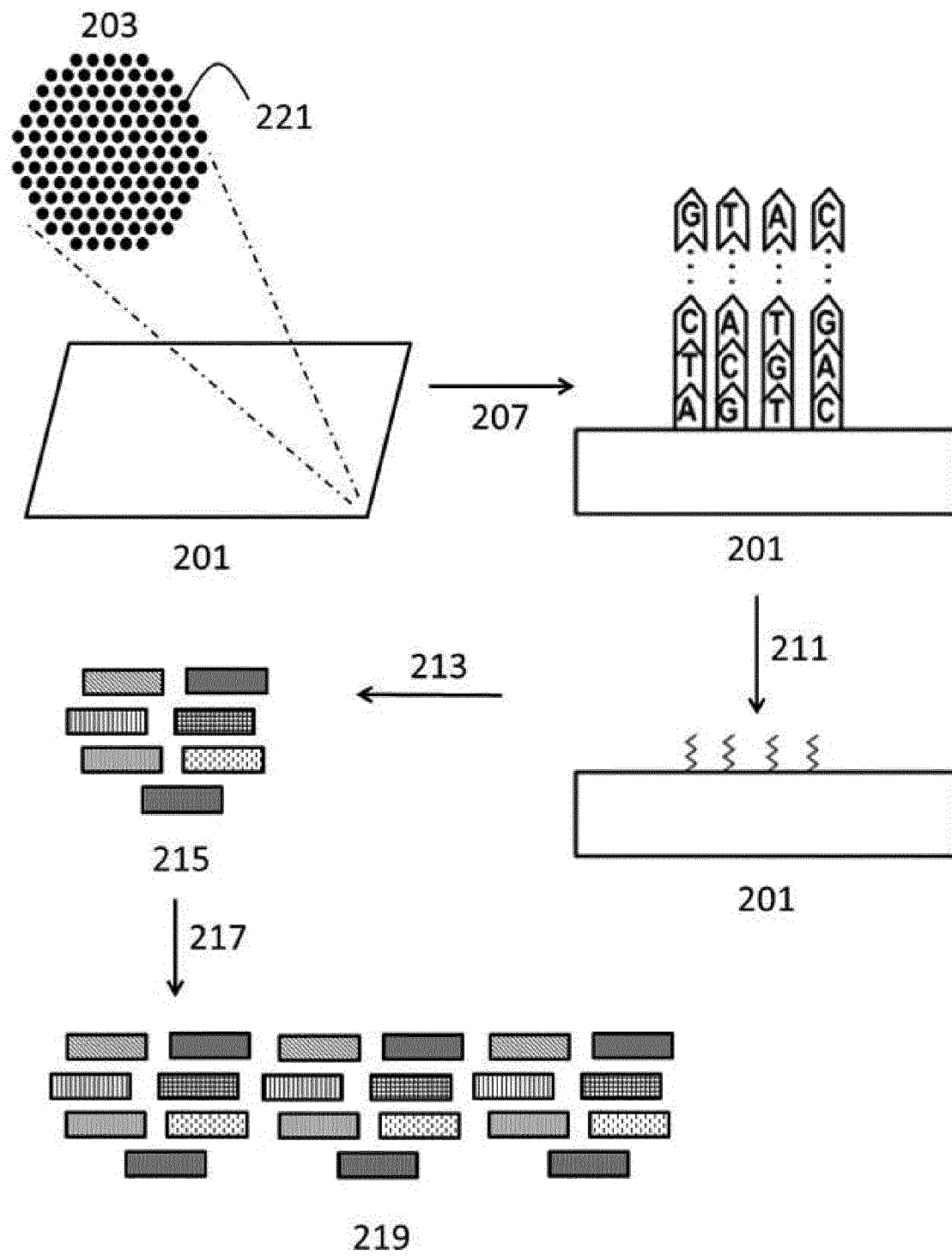
ФИГУРА 1А



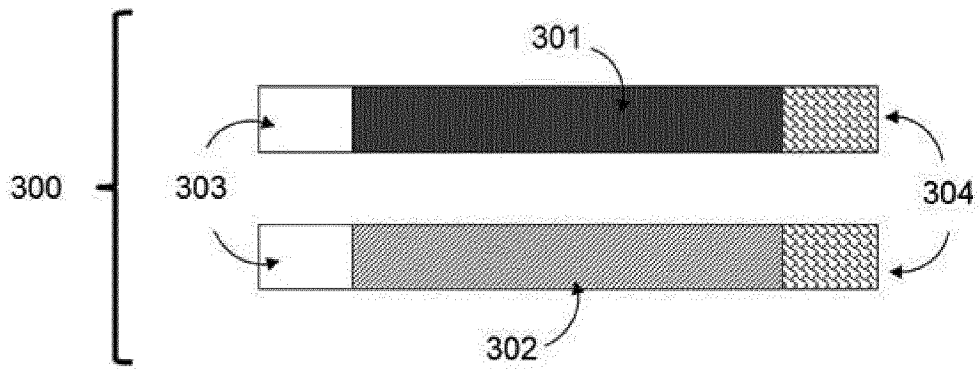
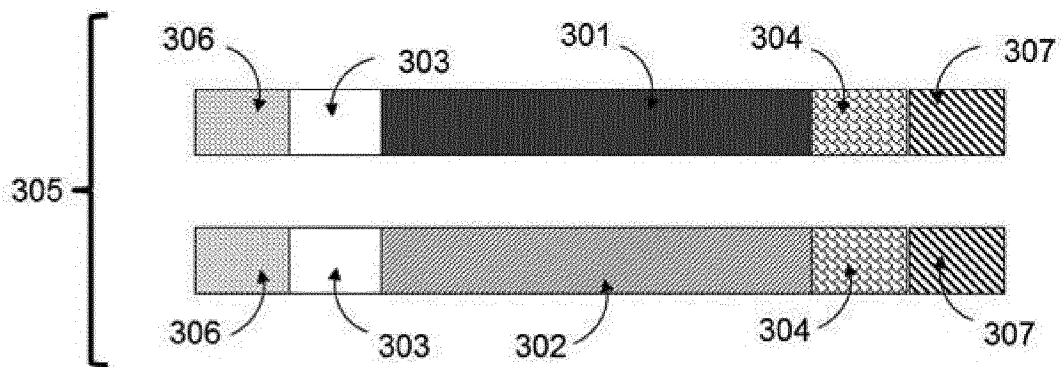
ФИГУРА 1В

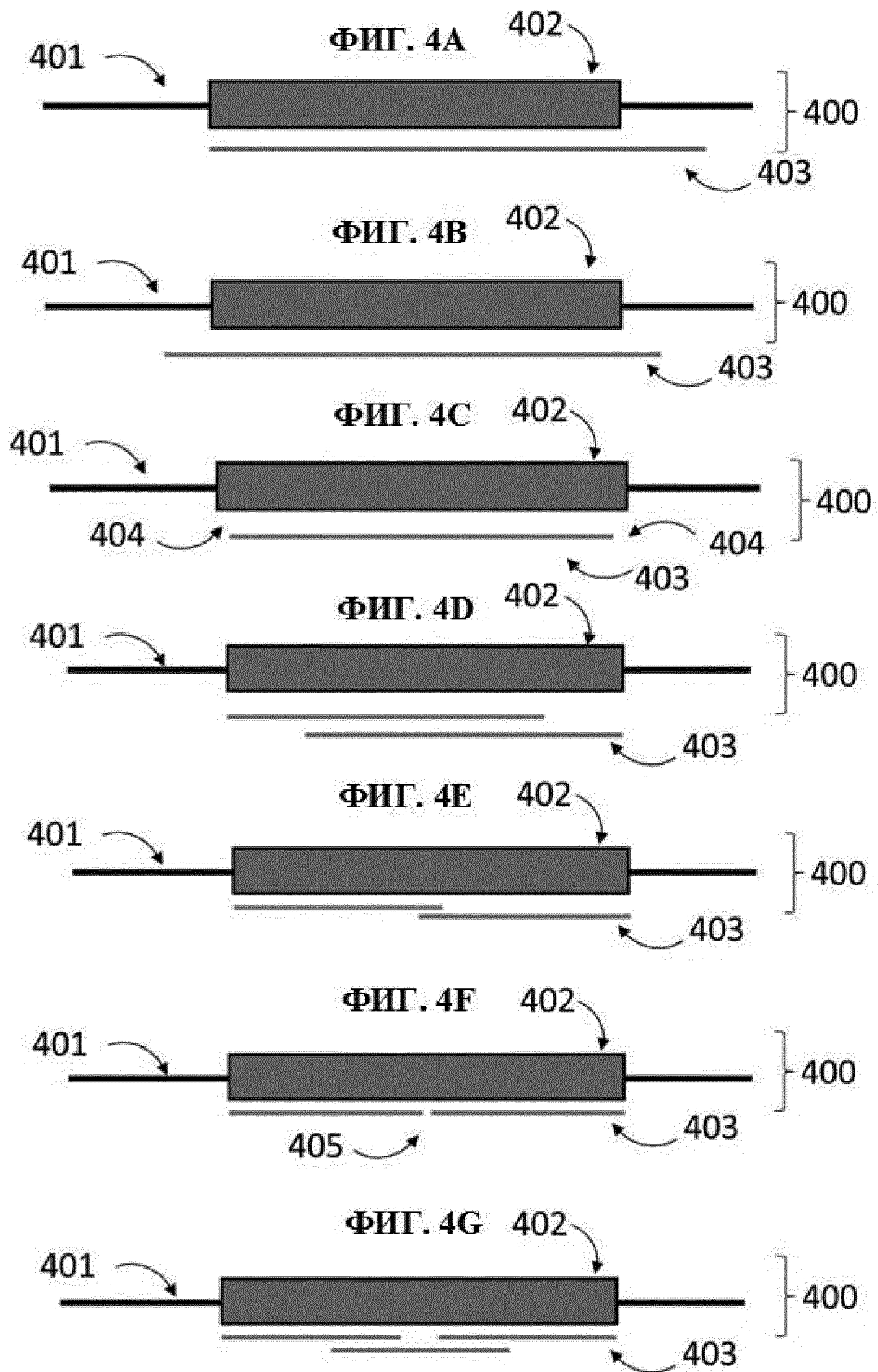


ФИГУРА 1С

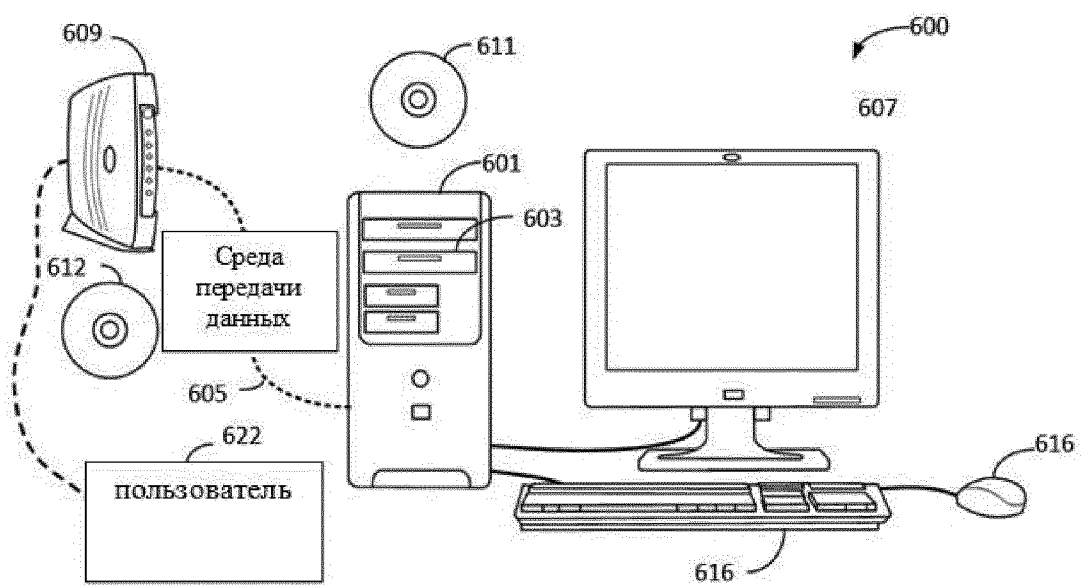


ФИГУРА 2

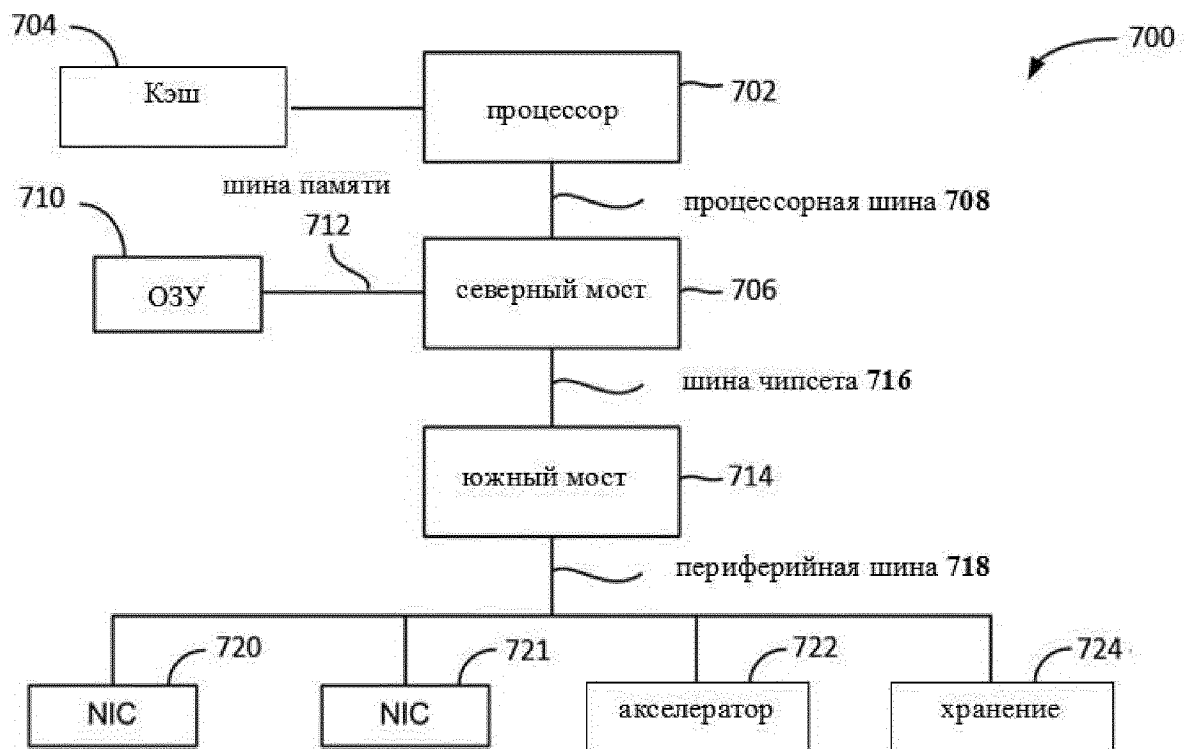
**ФИГУРА 3А****ФИГУРА 3В**

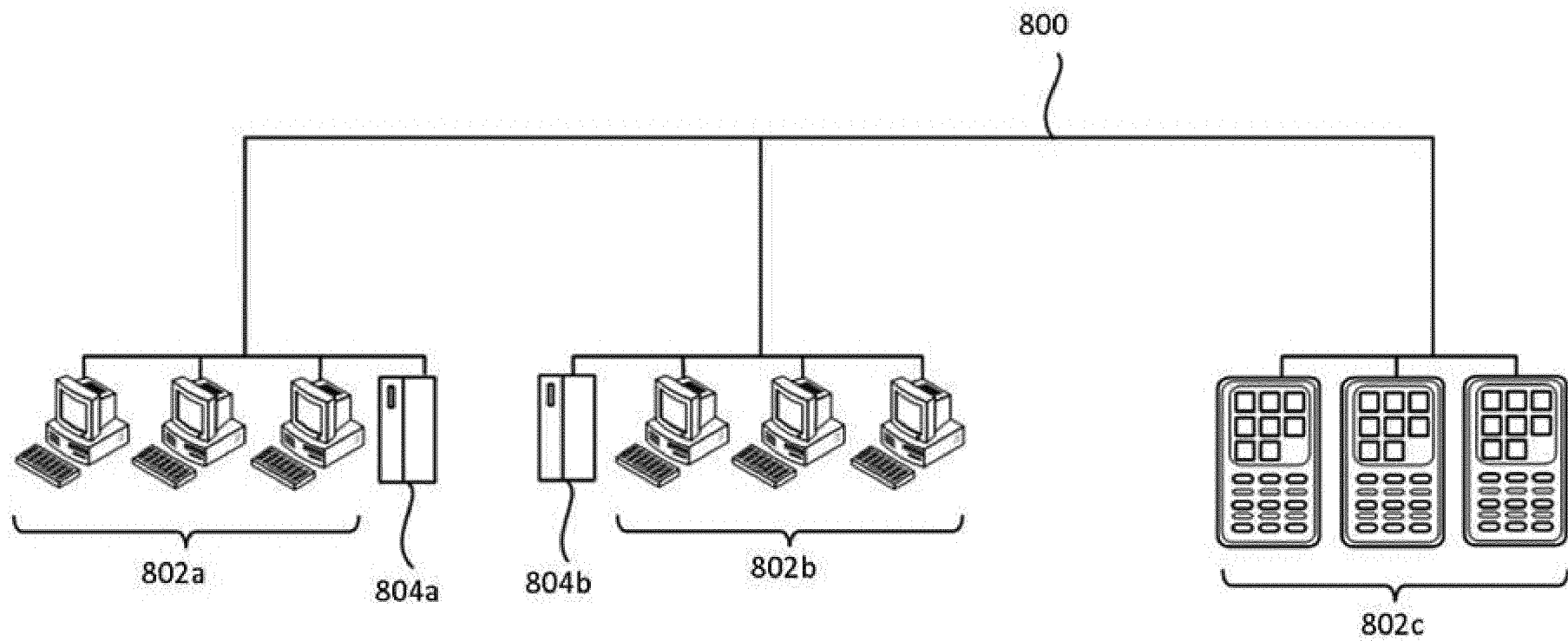


ФИГУРЫ 4А-4Г

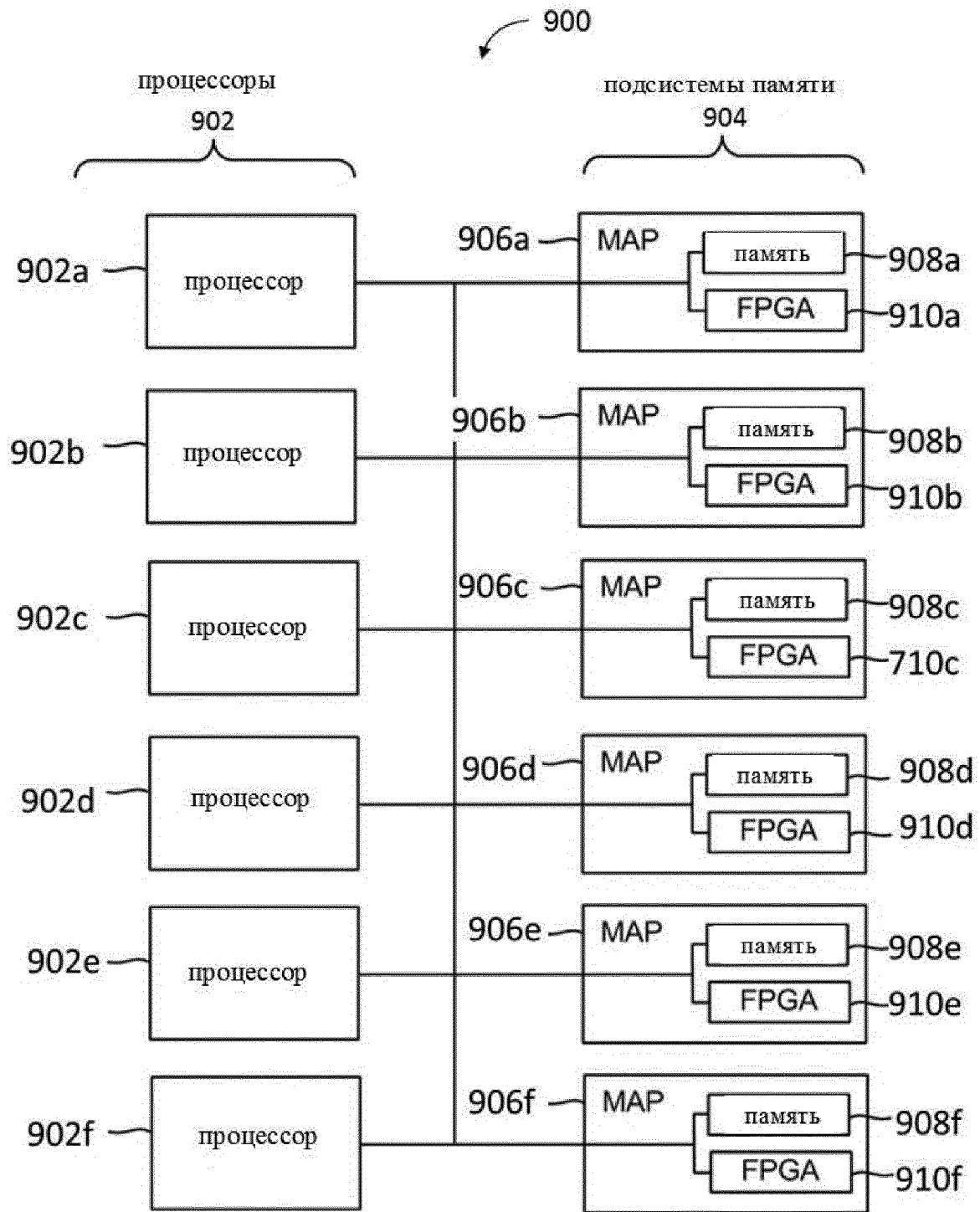


ФИГУРА 6

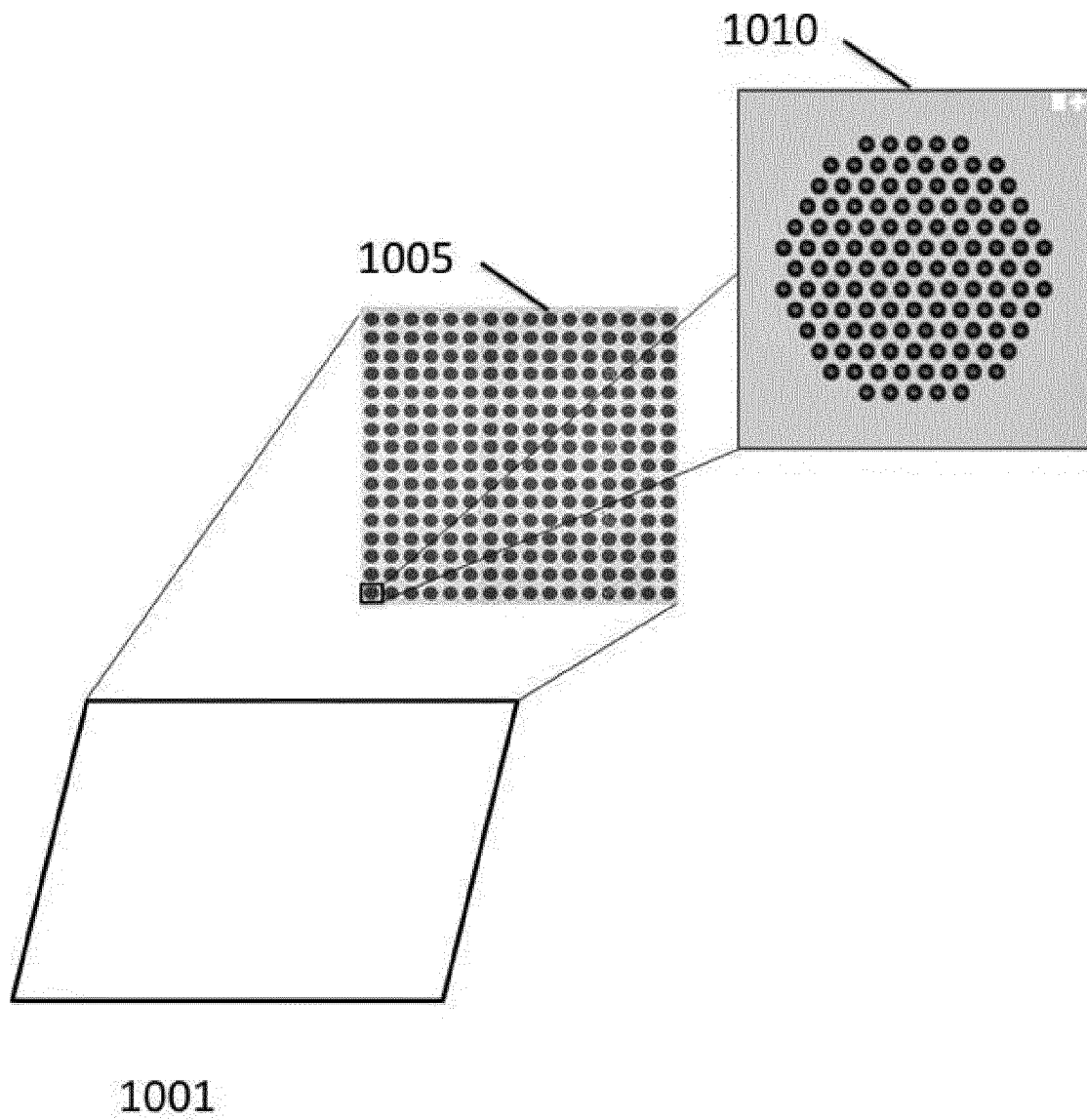
**ФИГУРА 7**



ФИГУРА 8



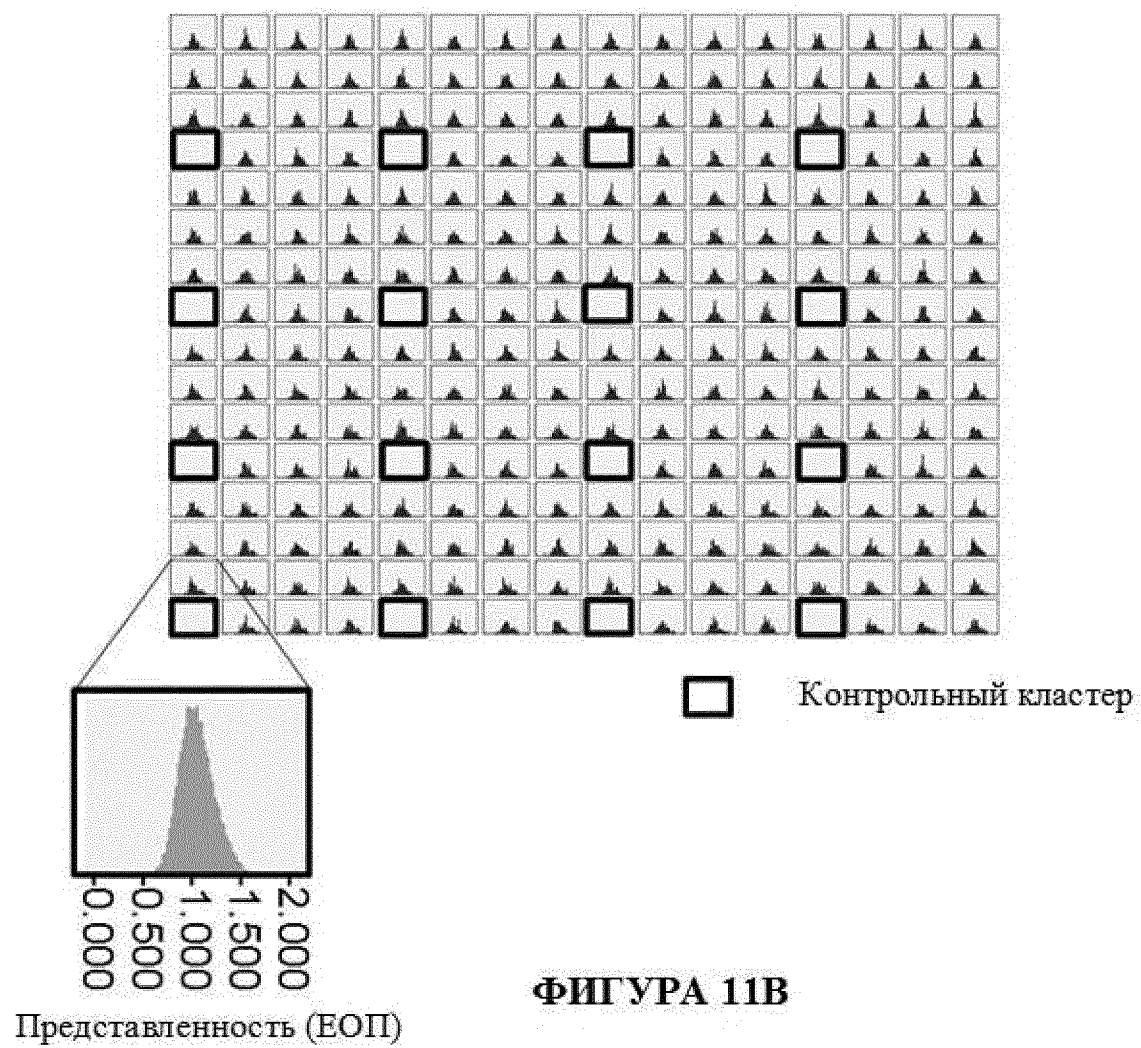
ФИГУРА 9



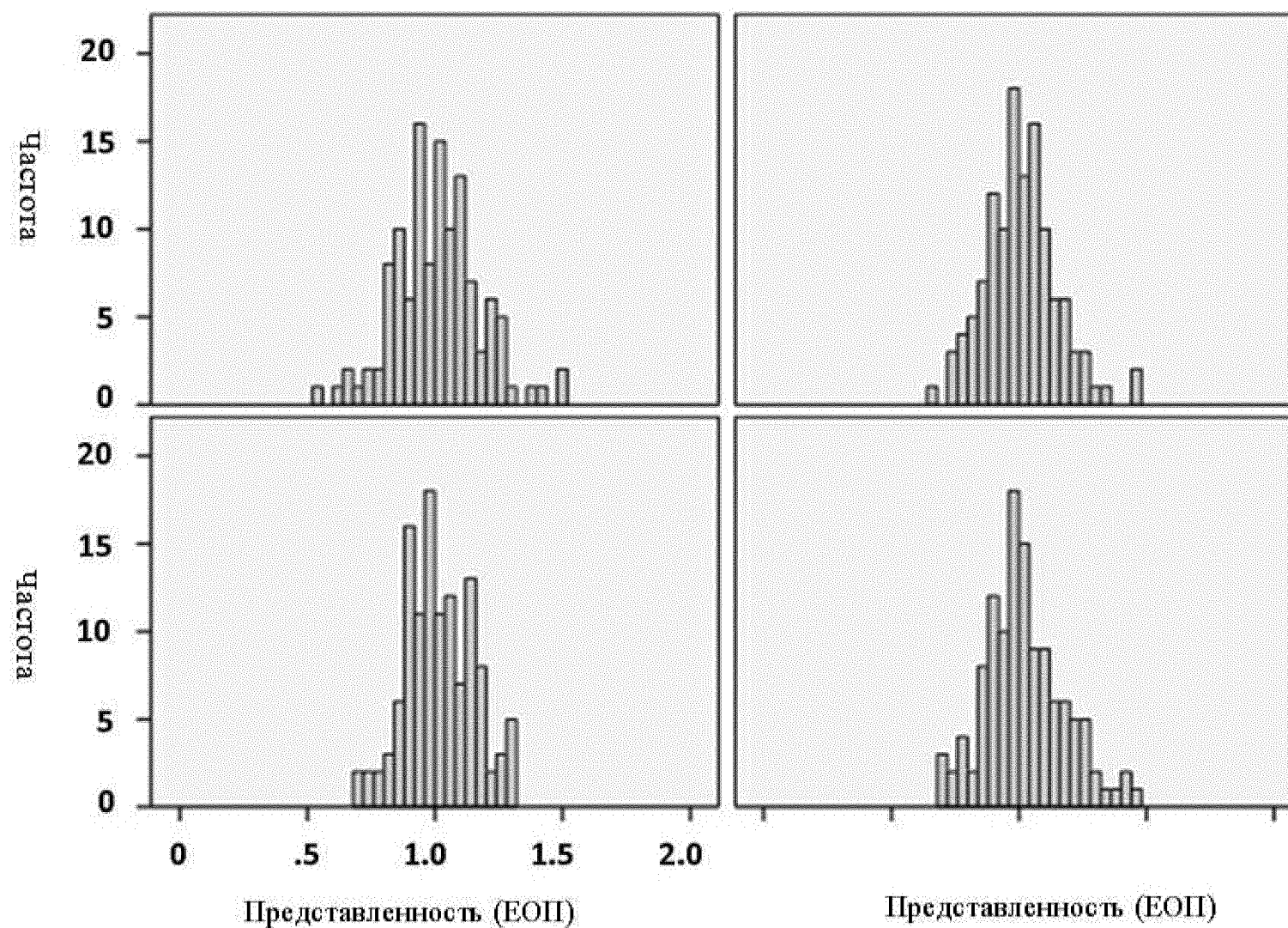
ФИГУРА 10



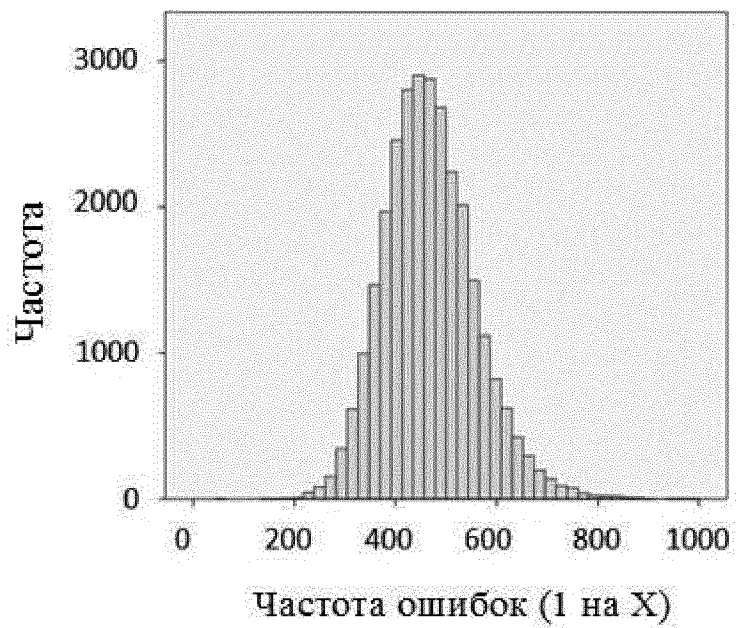
ФИГУРА 11А



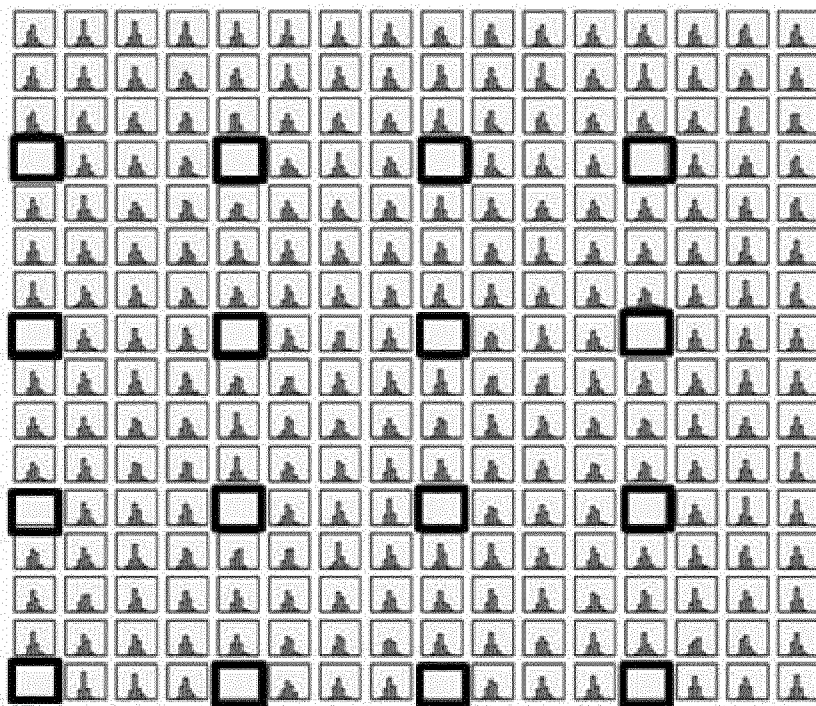
ФИГУРА 11В



ФИГУРА 12

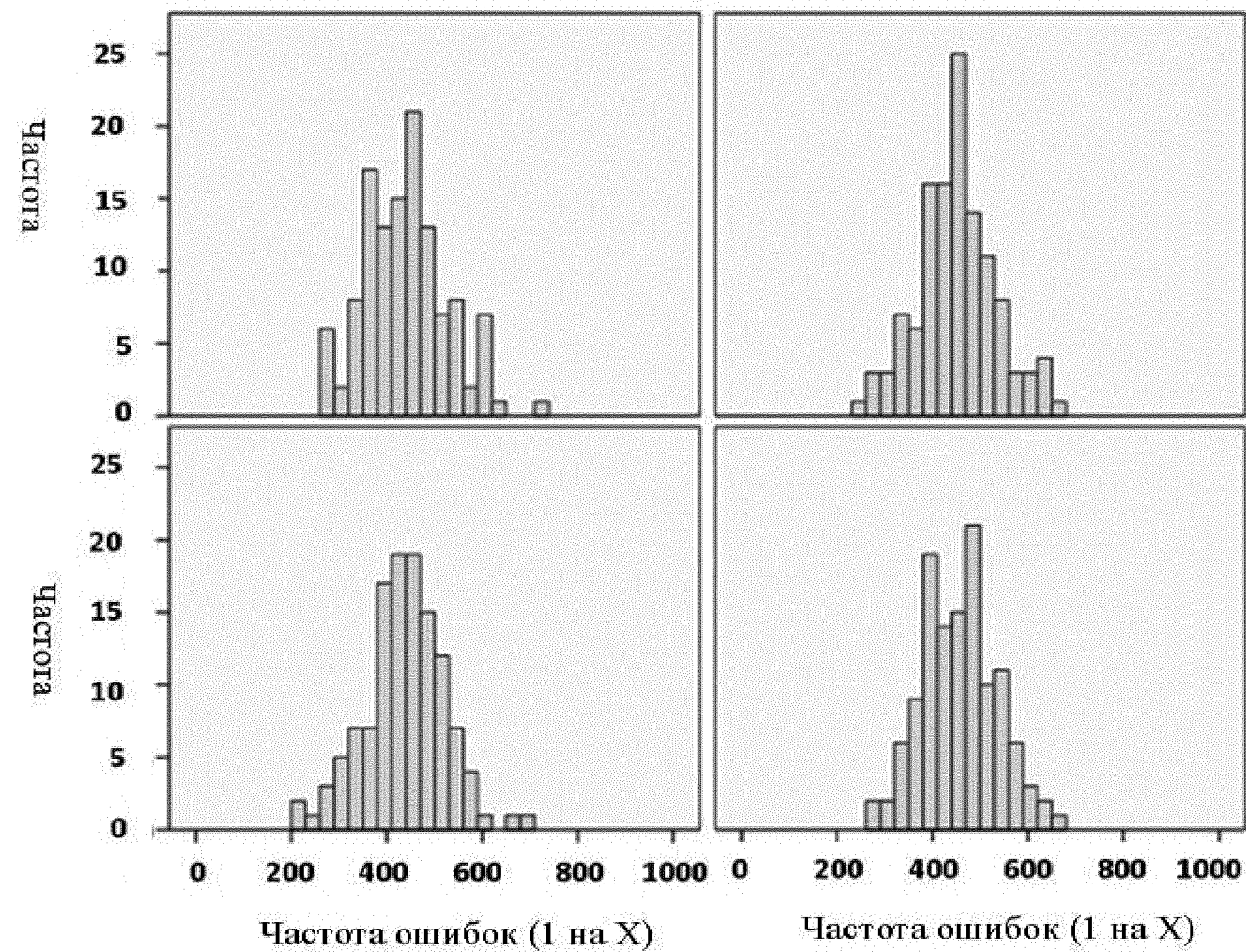


ФИГУРА 13А

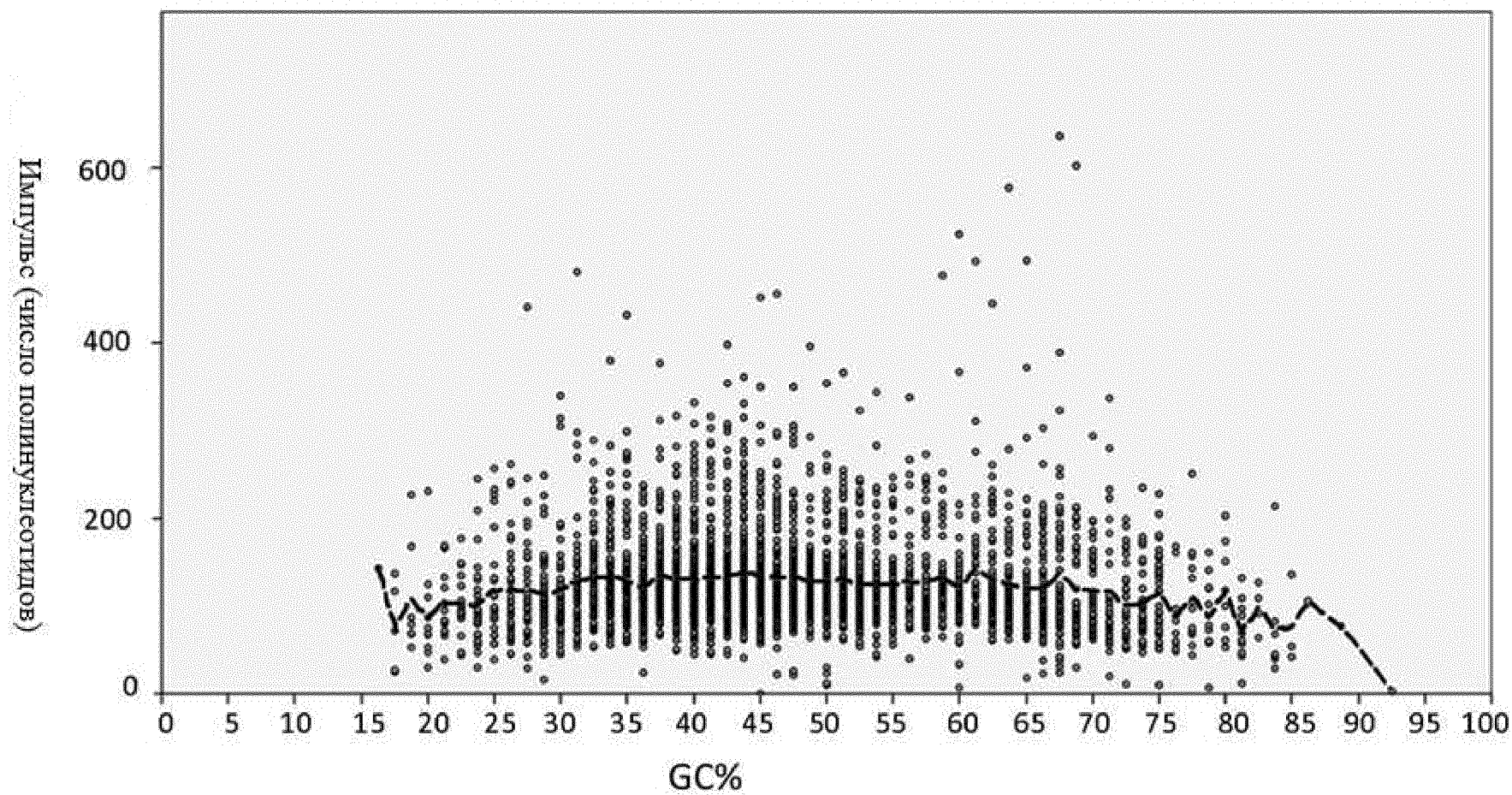


 Контрольный кластер

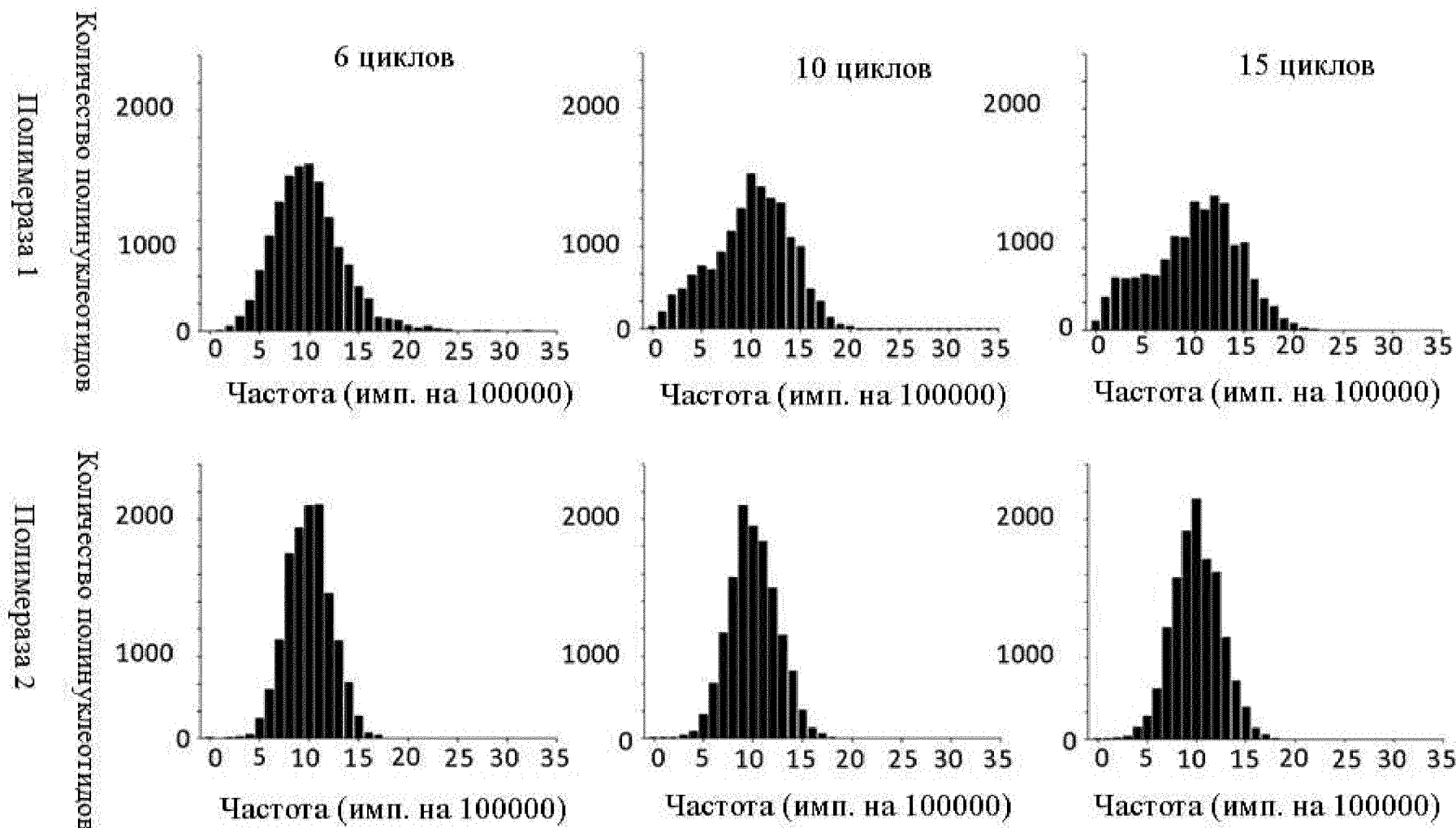
ФИГУРА 13В



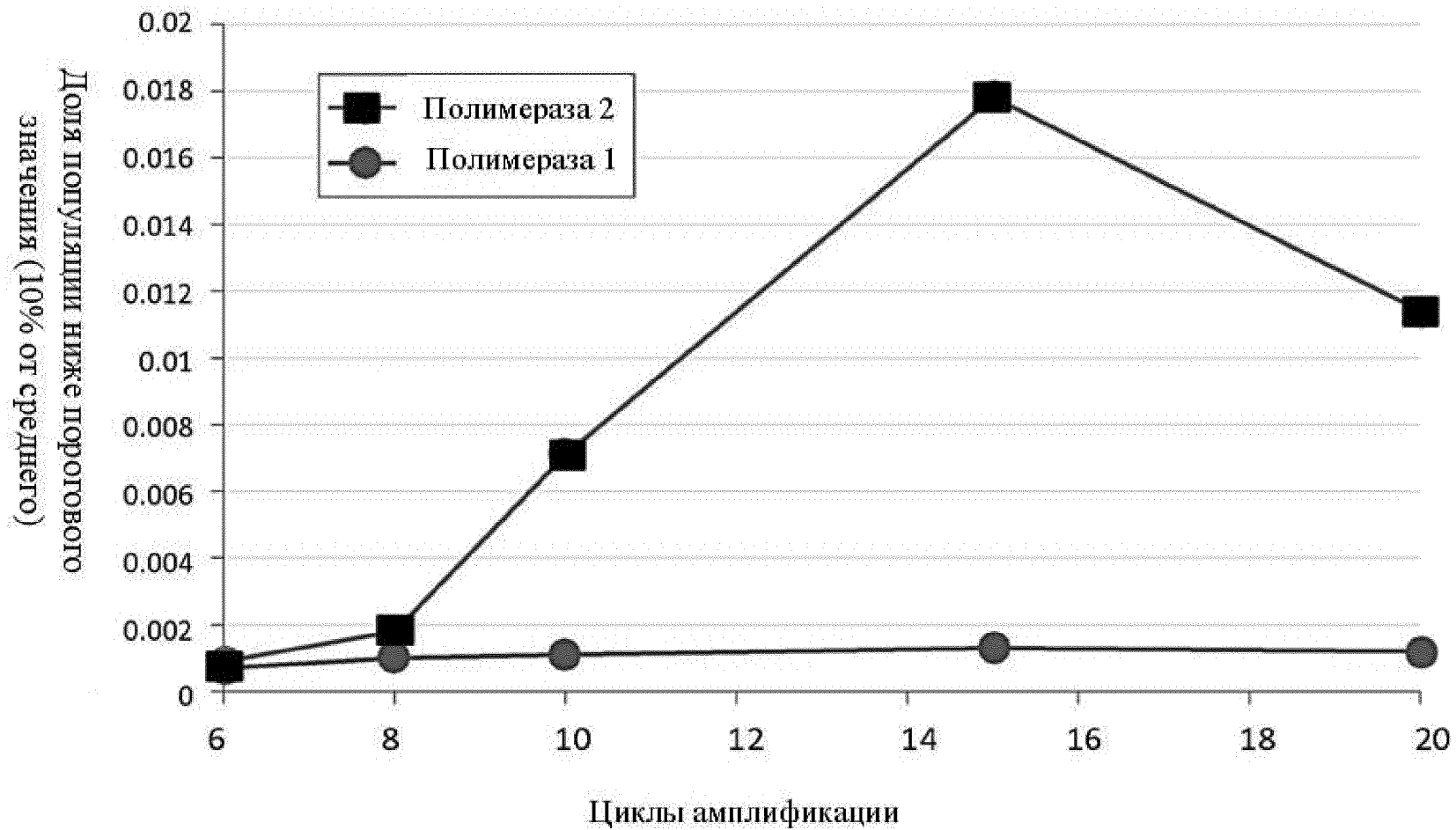
ФИГУРА 14



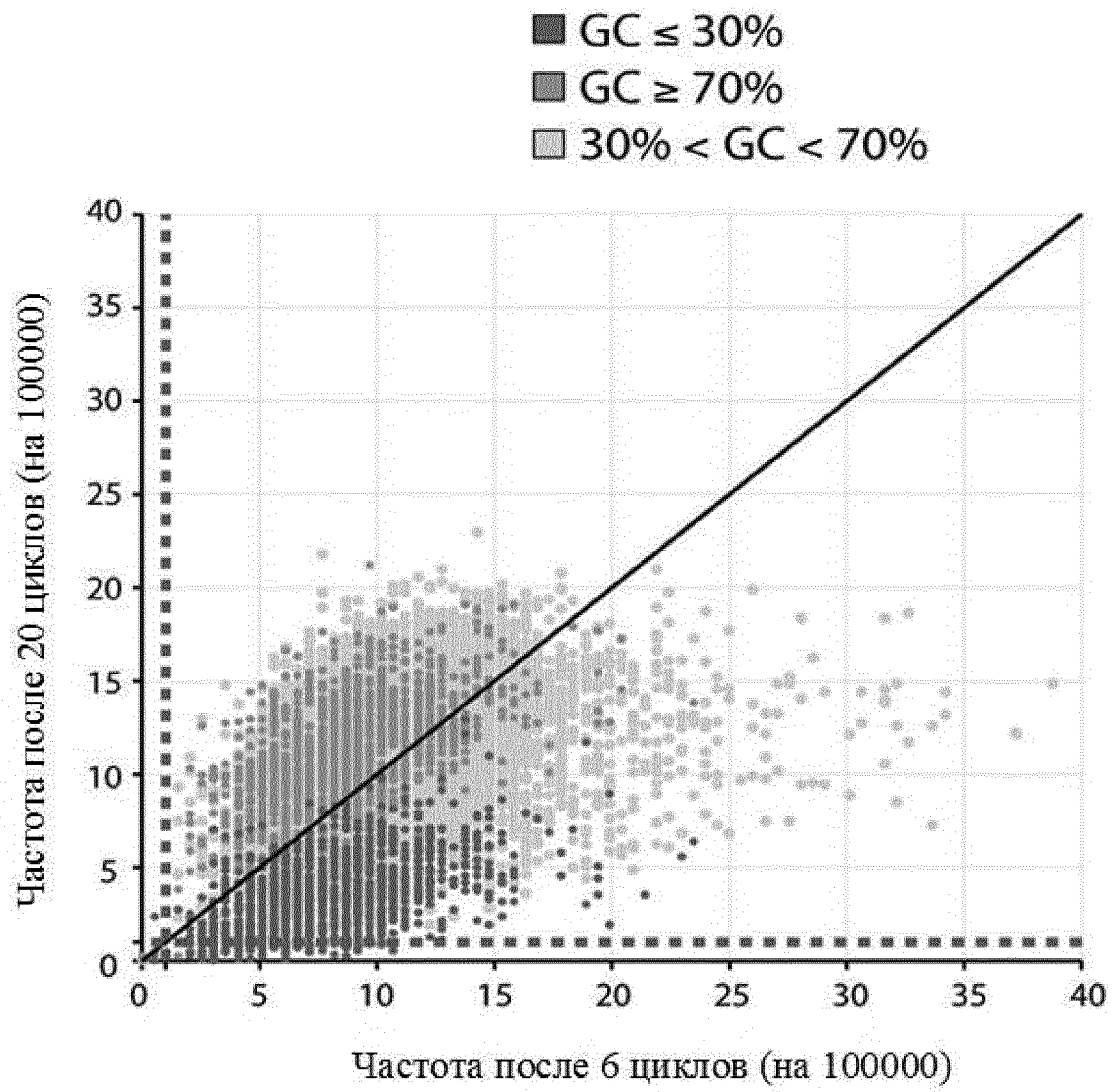
ФИГУРА 15



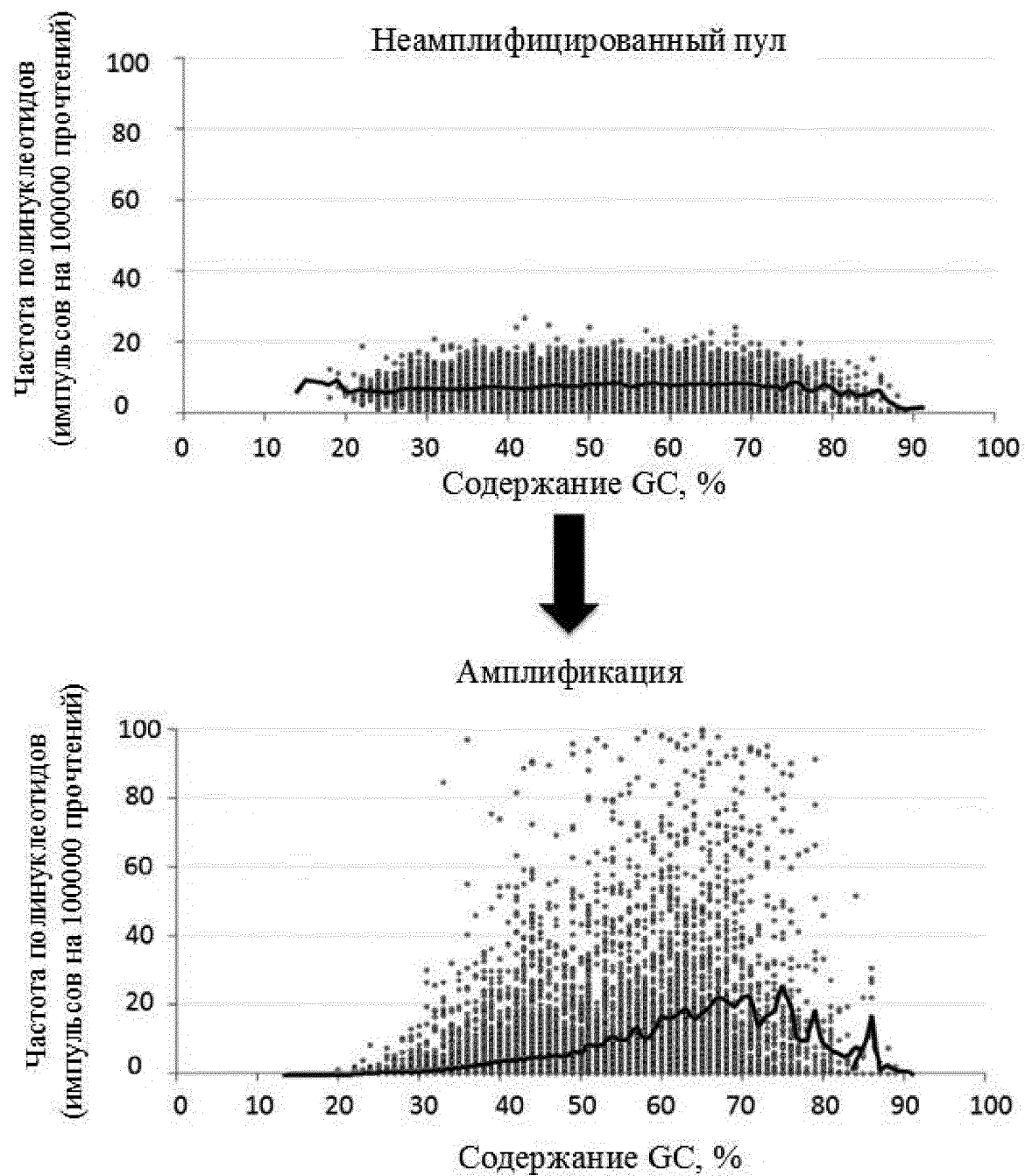
ФИГУРА 16

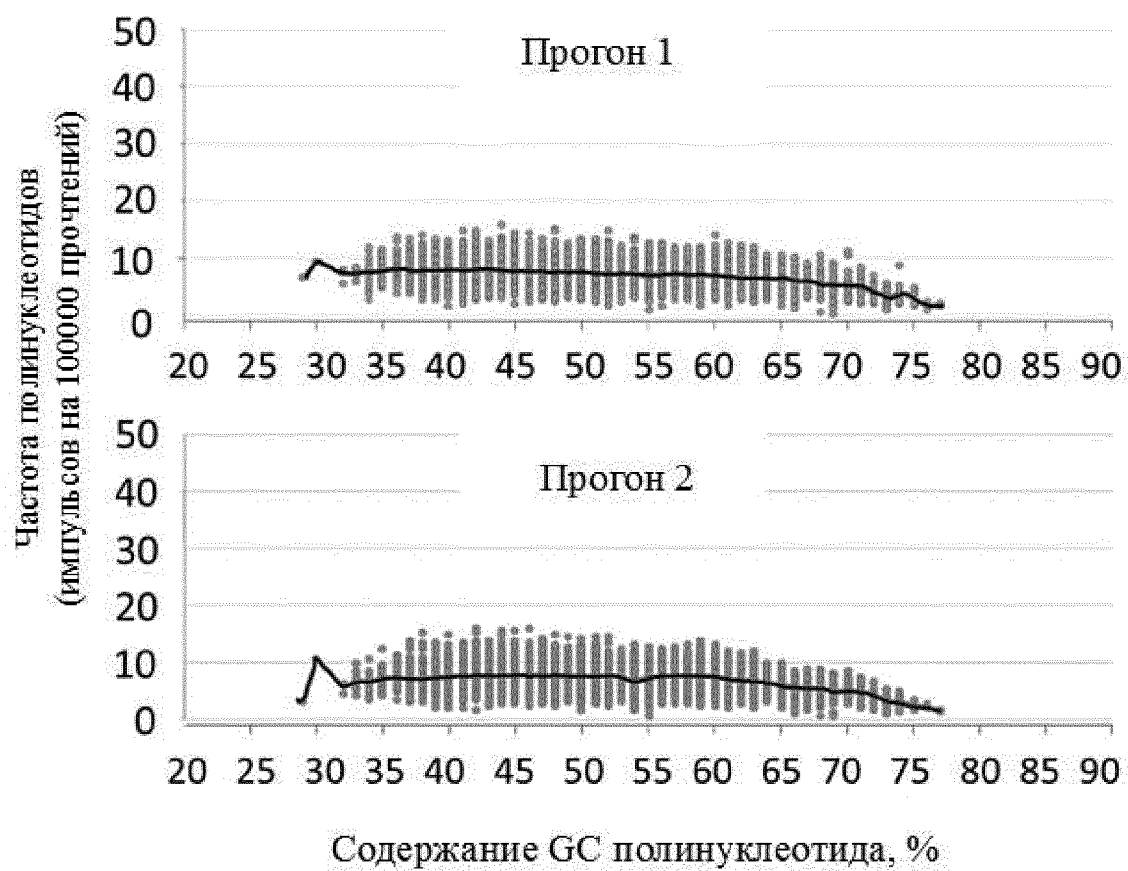


ФИГУРА 17

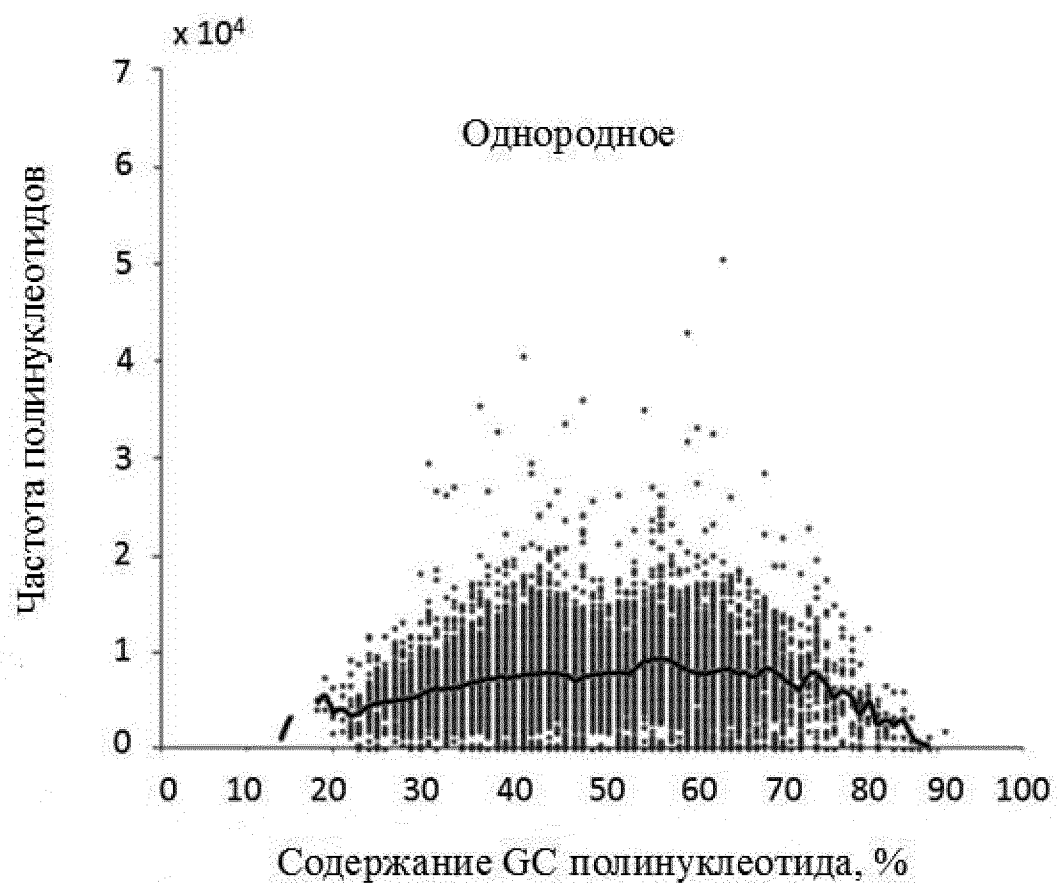


ФИГУРА 18

**ФИГУРА 19**

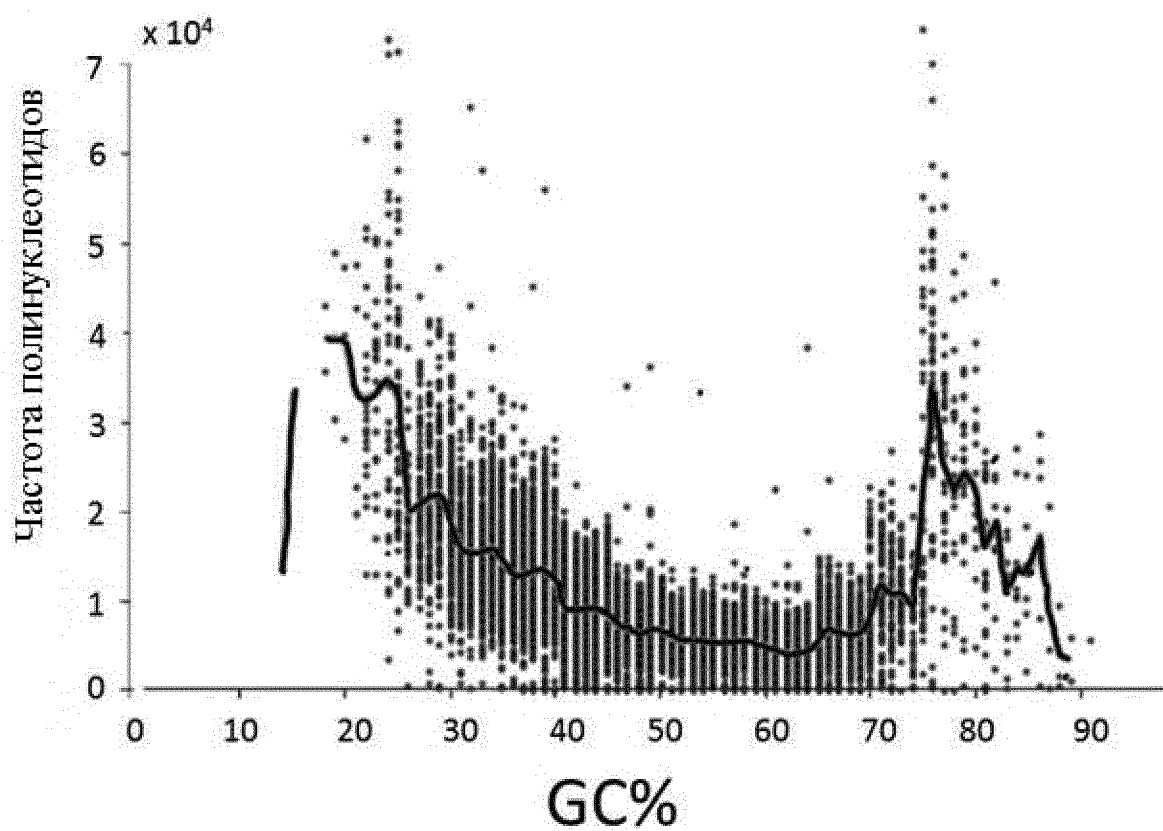


ФИГУРА 20

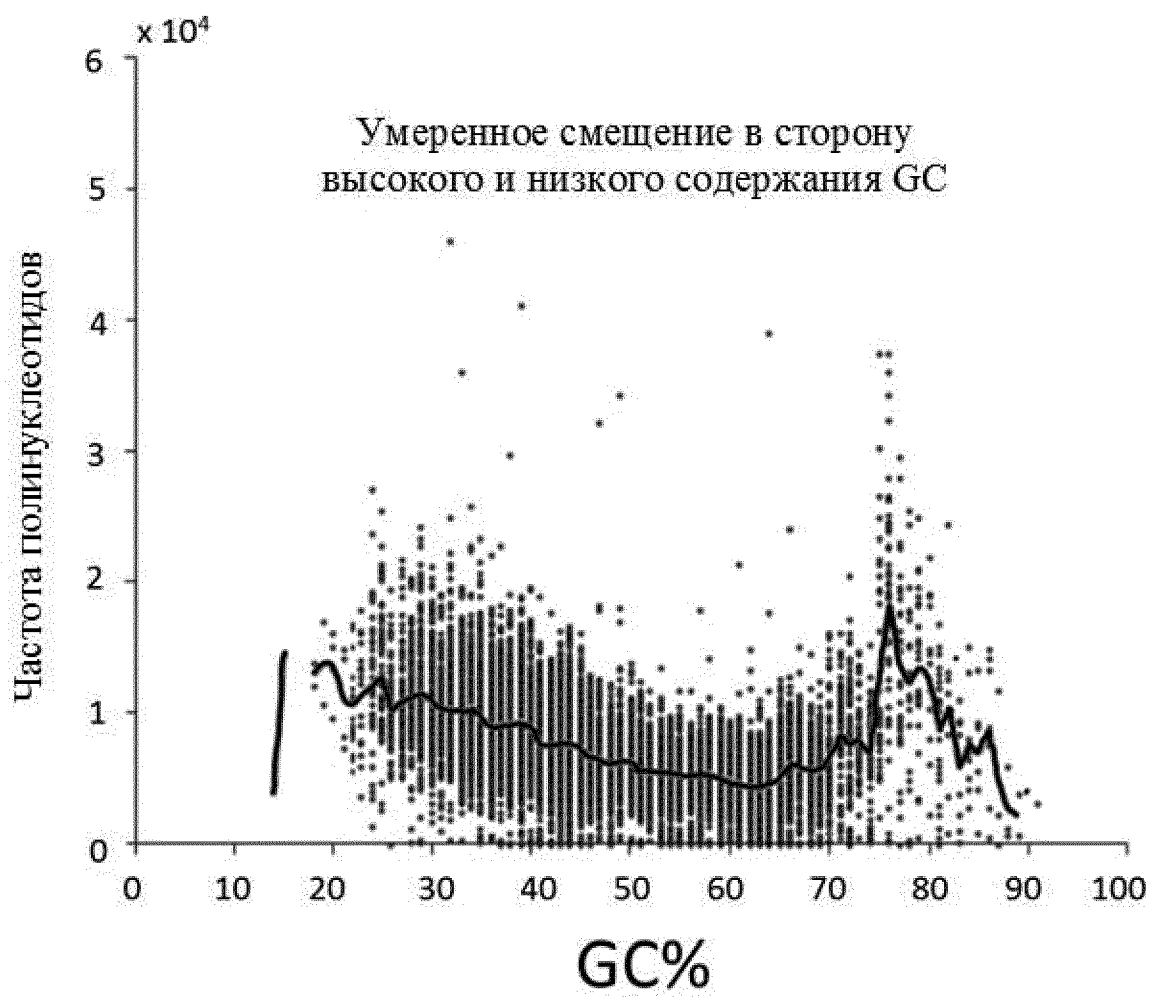


ФИГУРА 21А

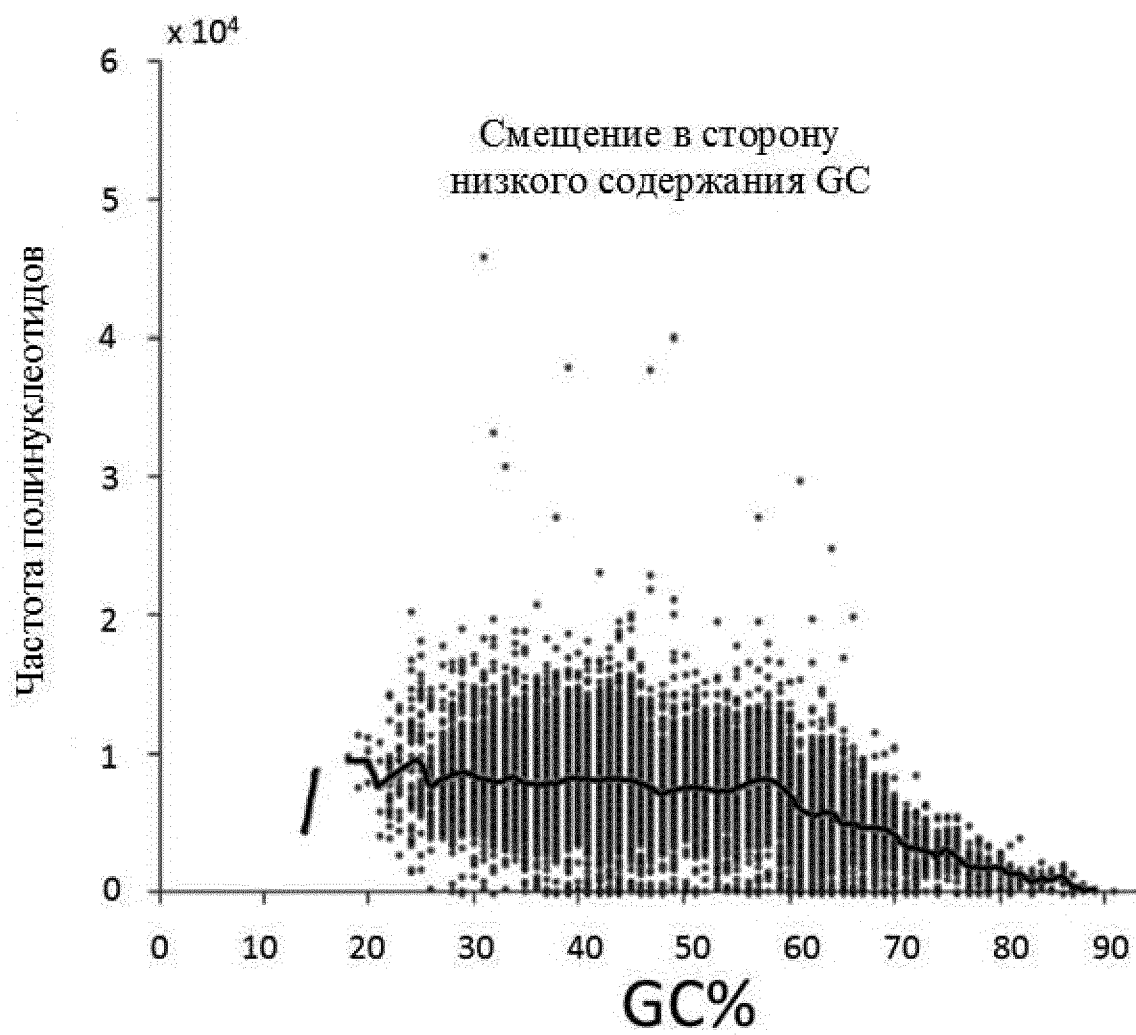
Сильное смещение в сторону
высокого и низкого содержания GC



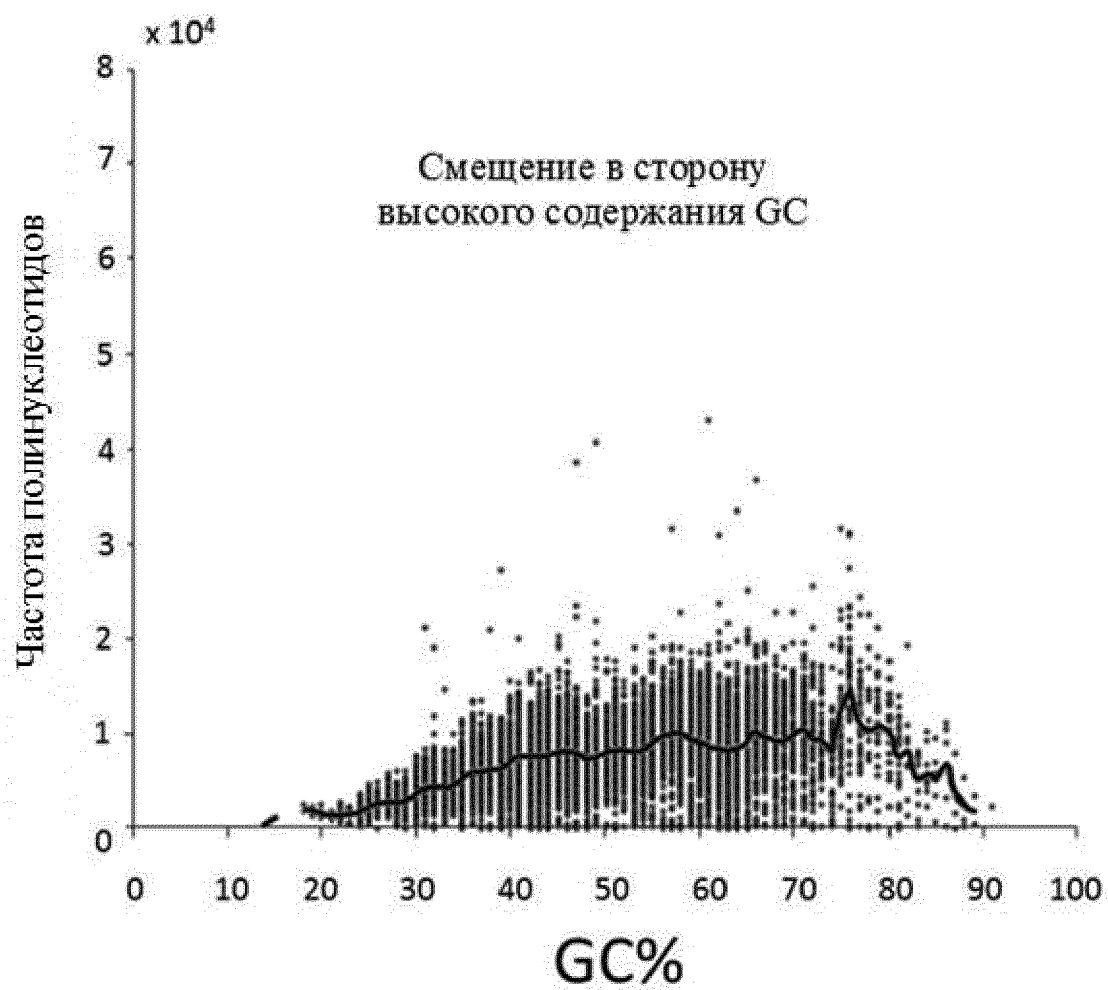
ФИГУРА 21В



ФИГУРА 21С

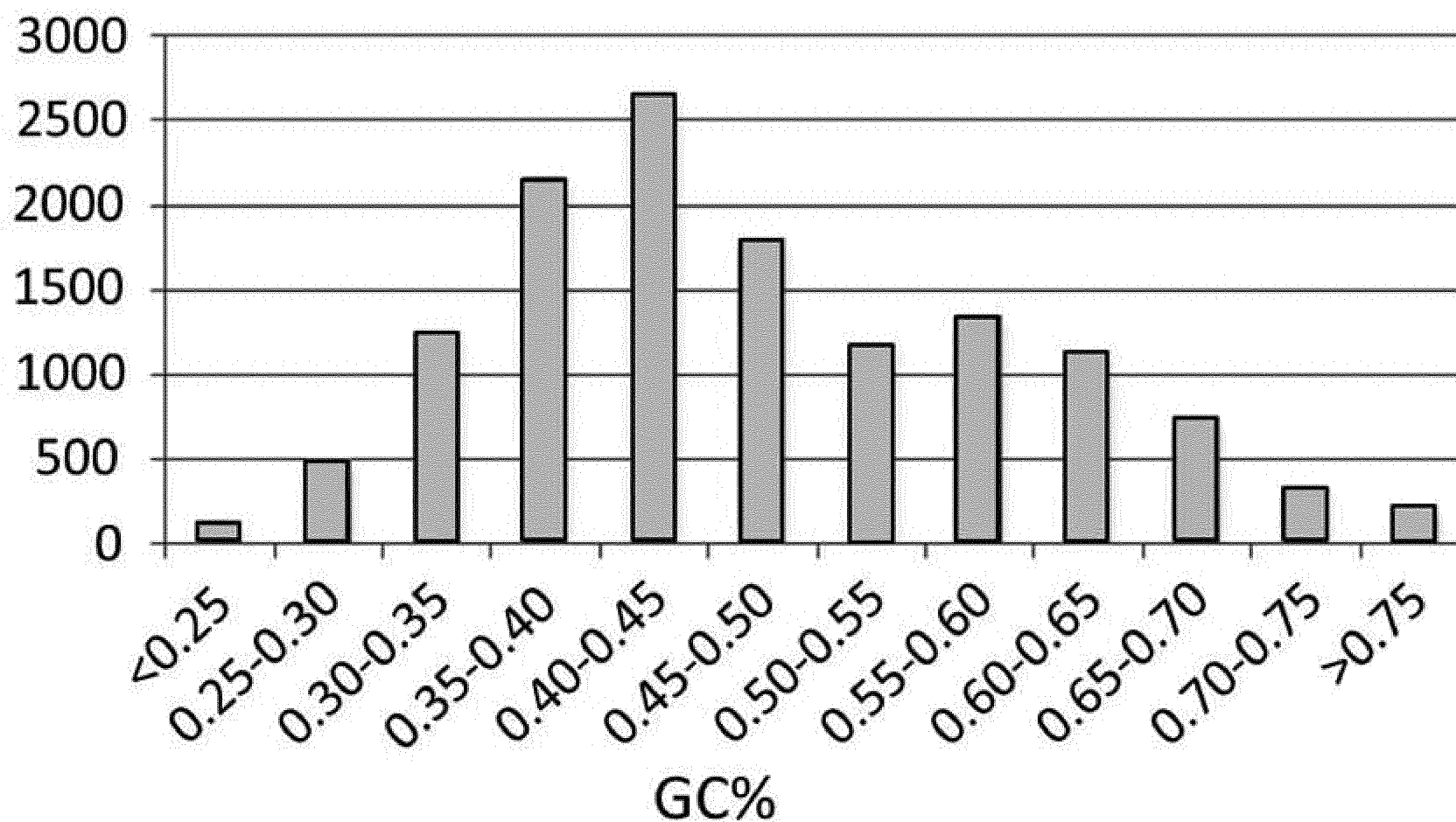


ФИГУРА 21D

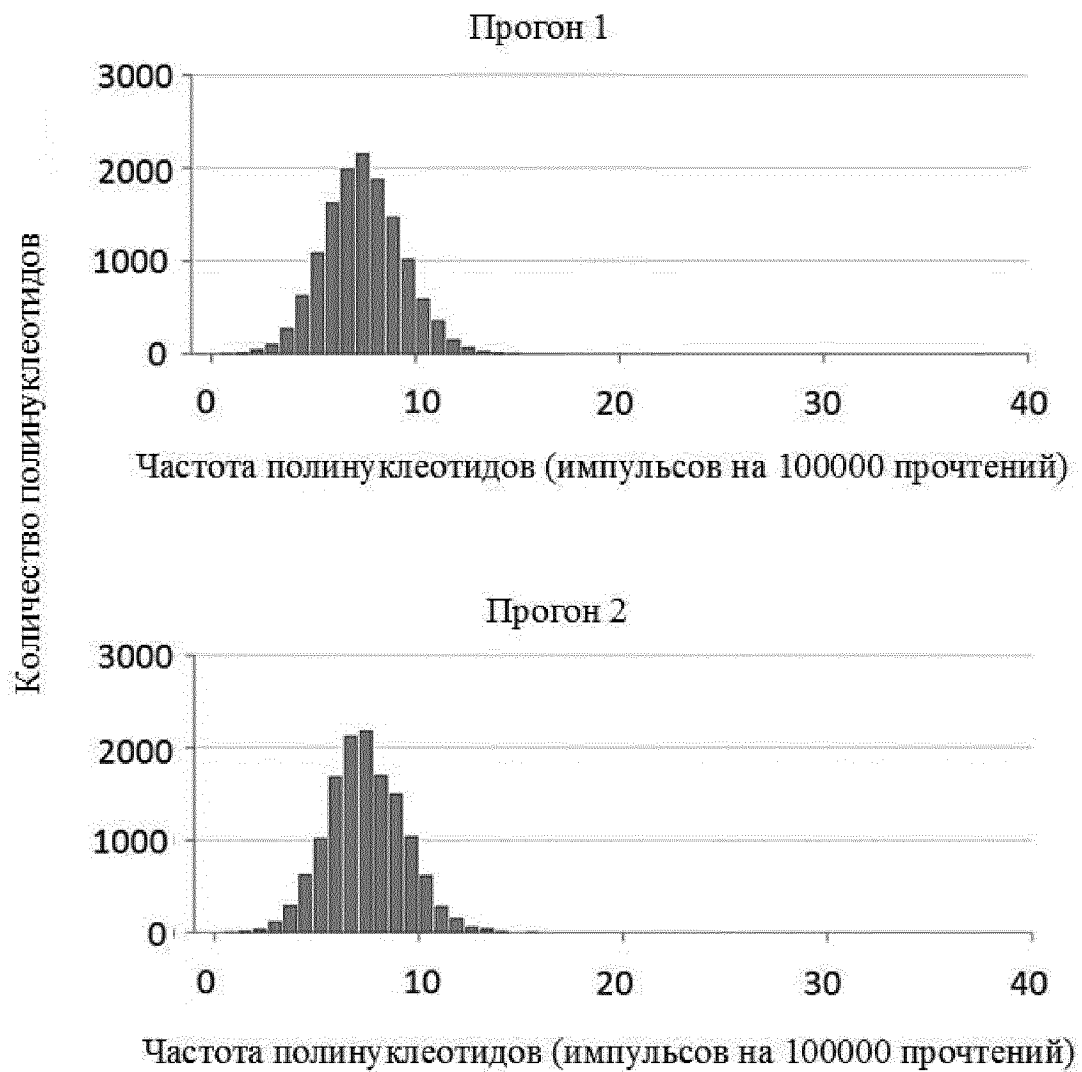


ФИГУРА 21Е

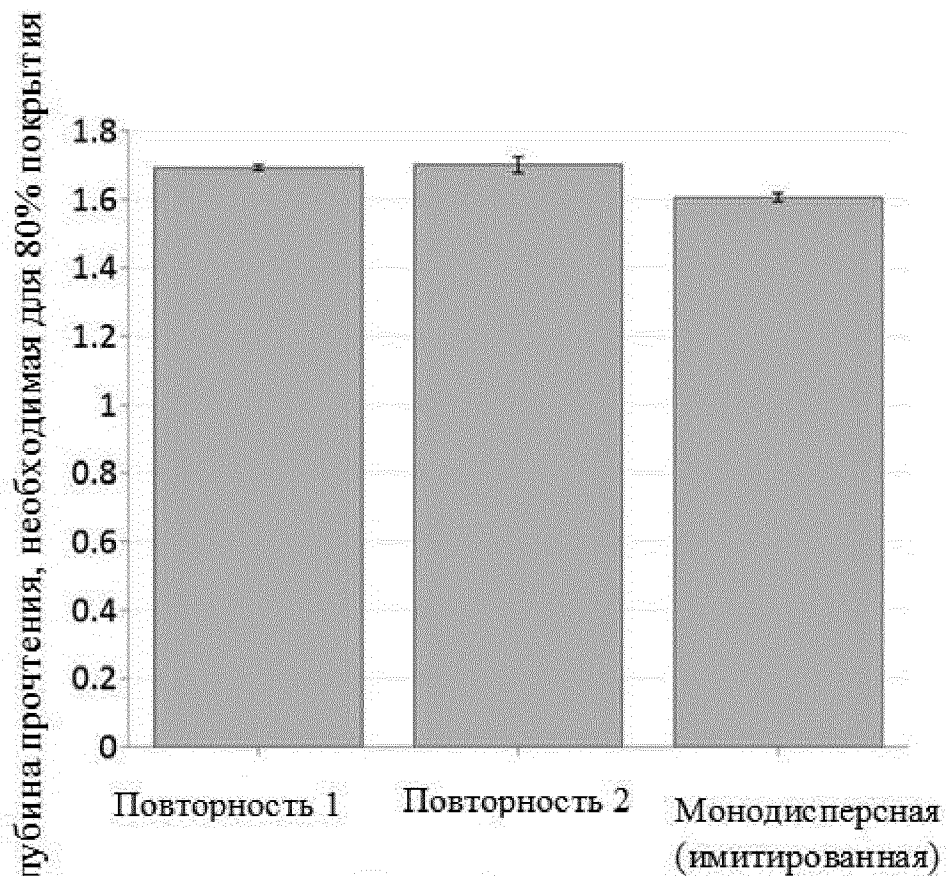
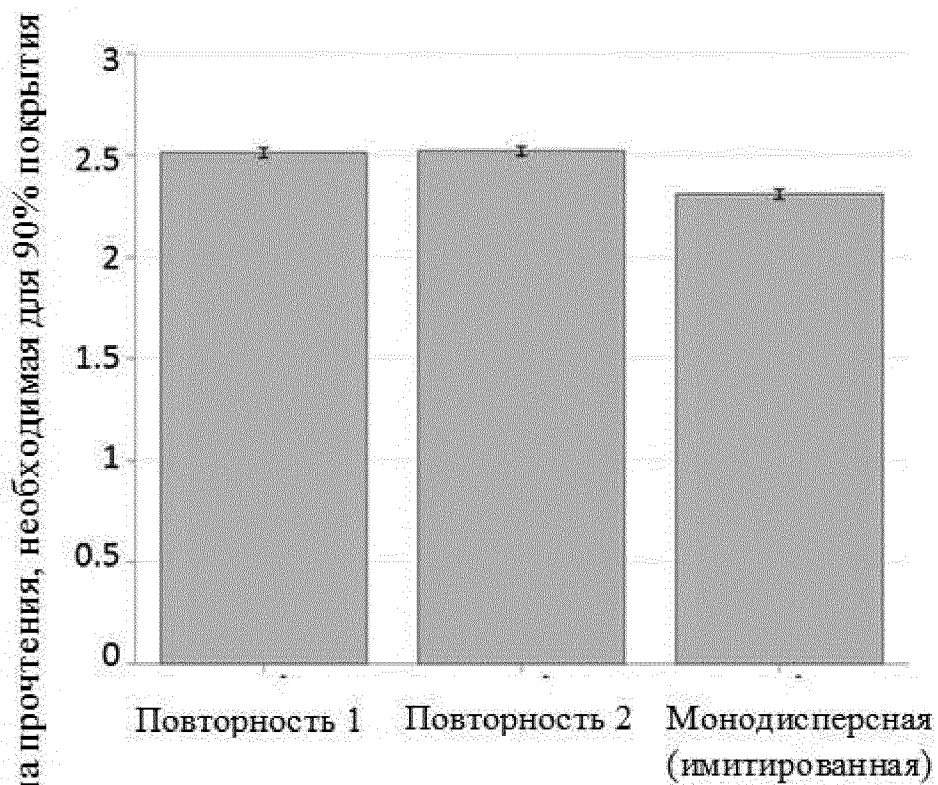
Колічність поліну/клієтидів

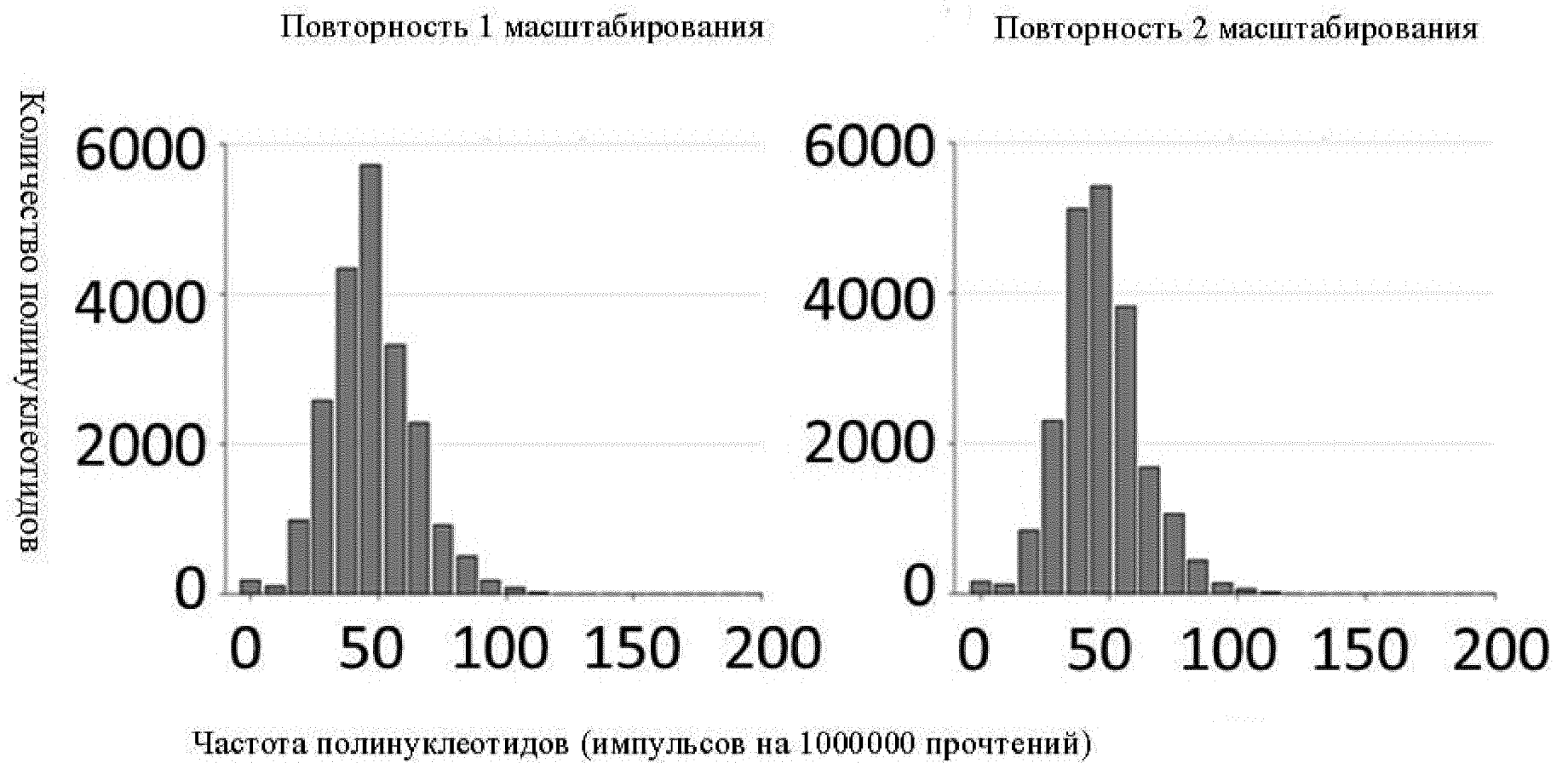


ФИГУРА 22

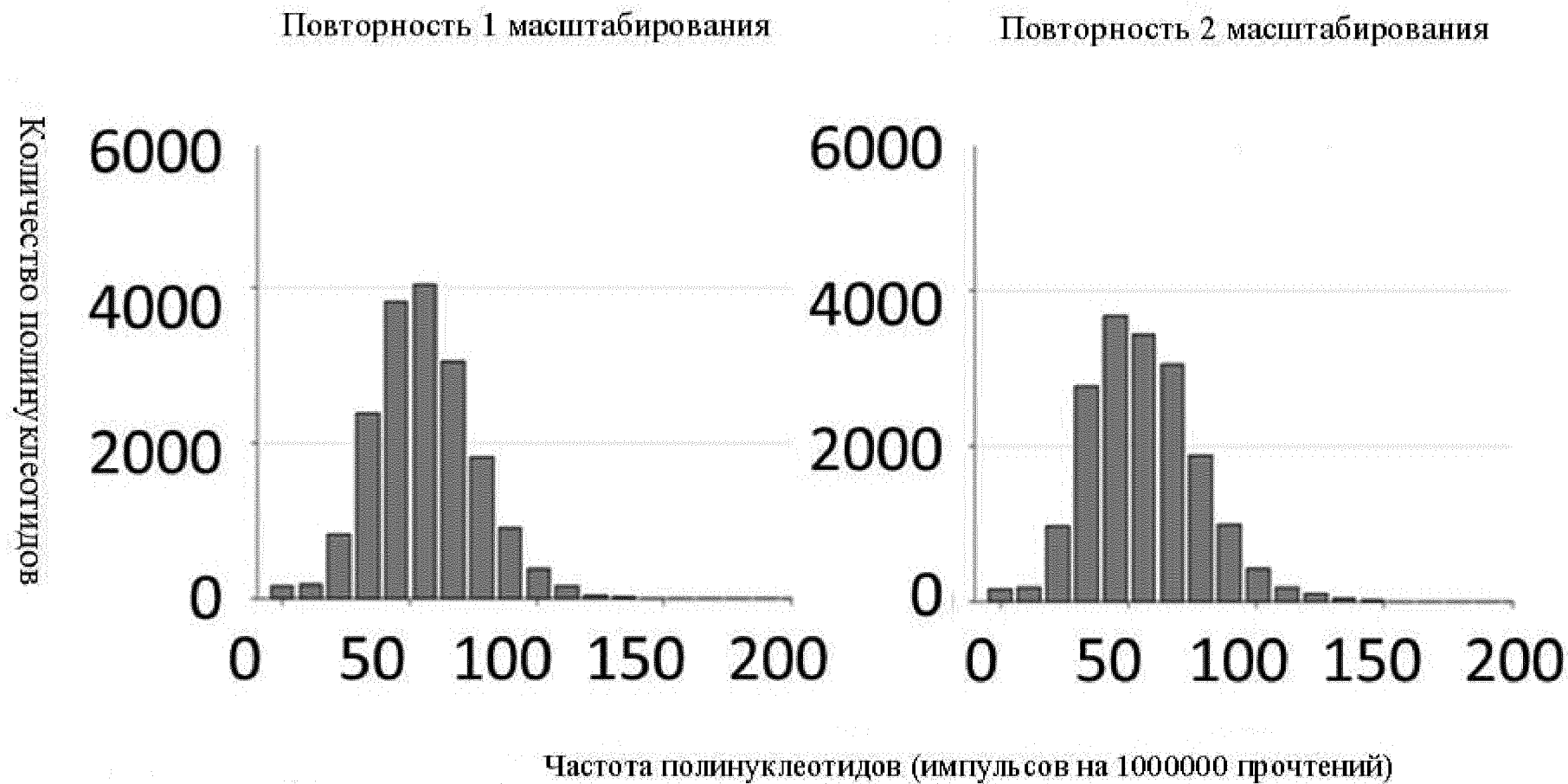


ФИГУРА 23

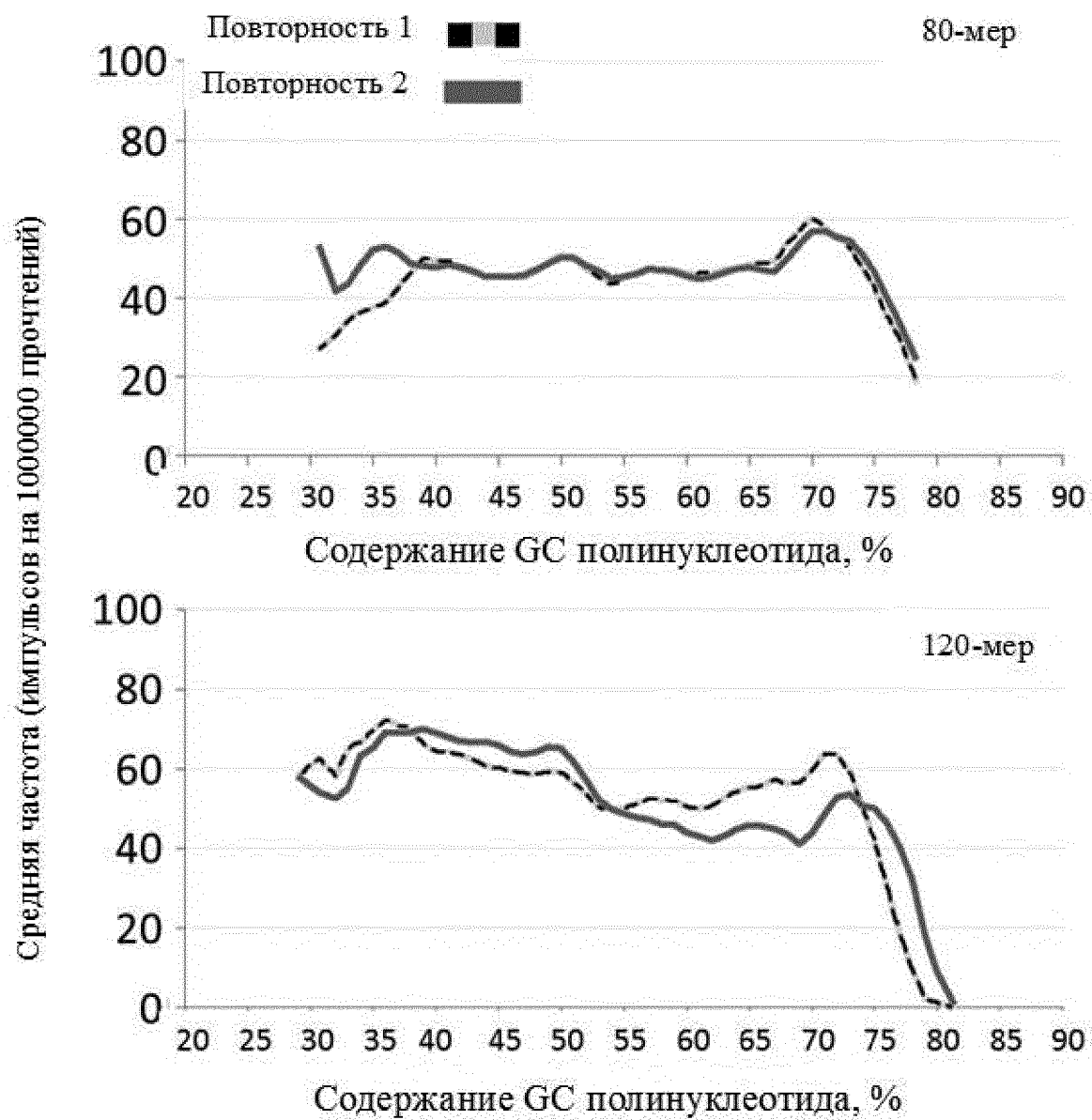
**ФИГУРА 24А****ФИГУРА 24В**



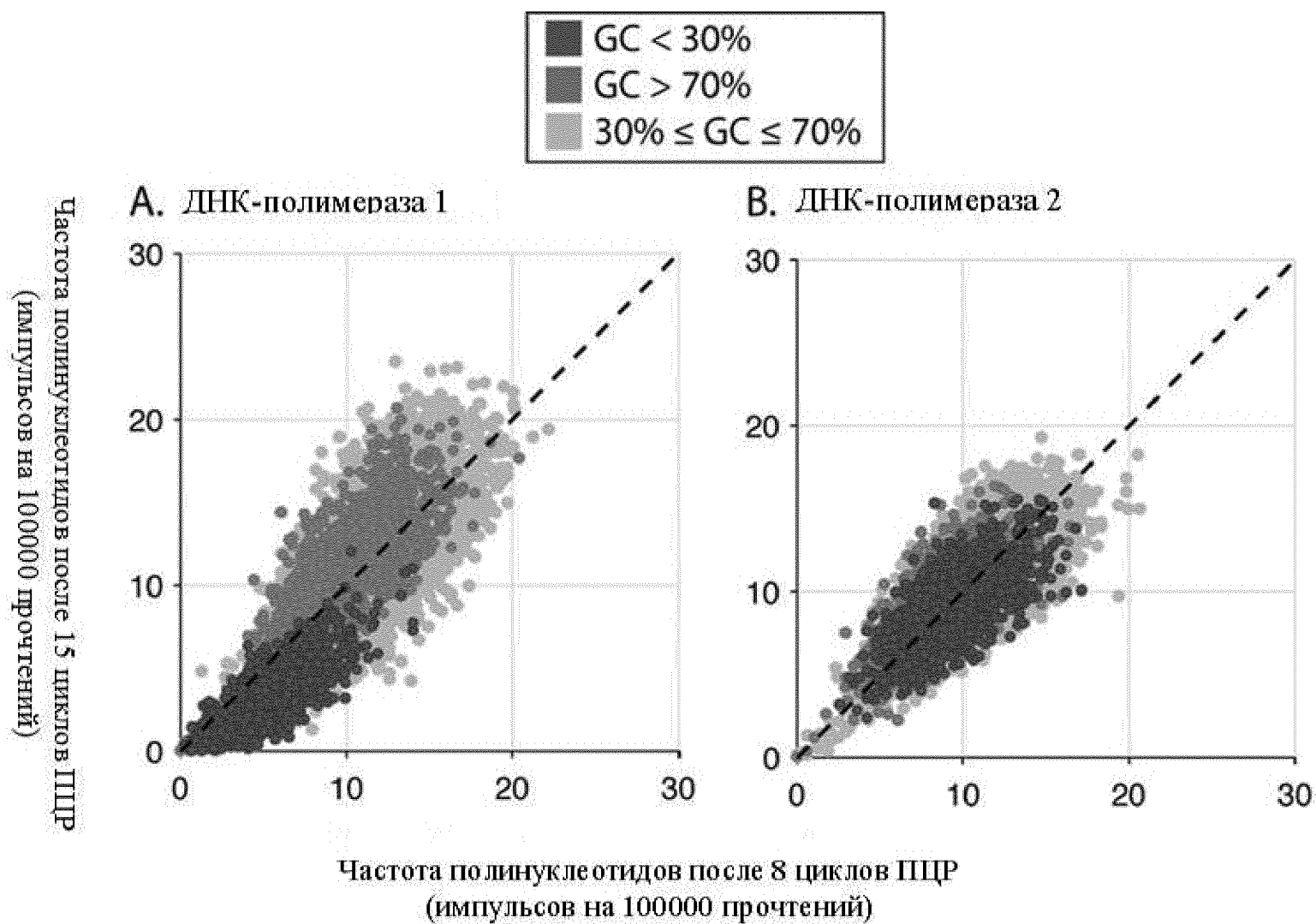
ФИГУРА 25



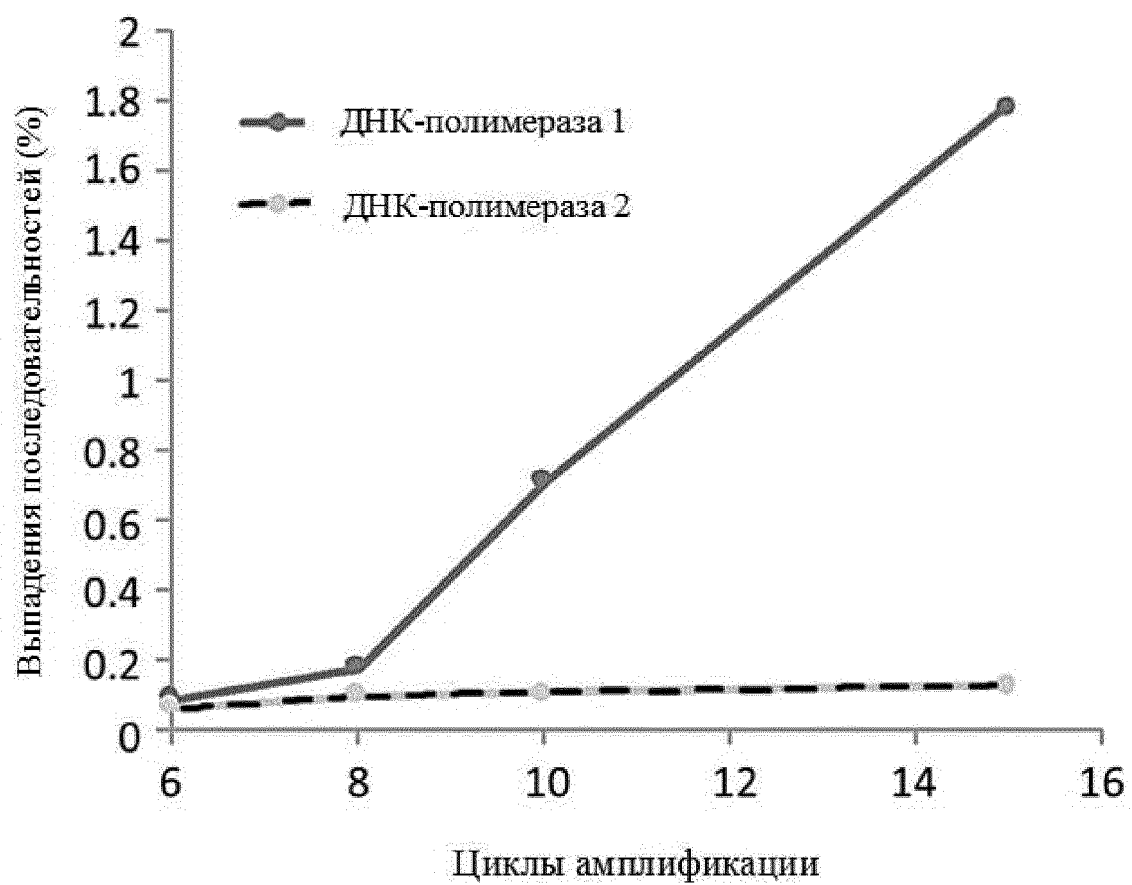
ФИГУРА 26



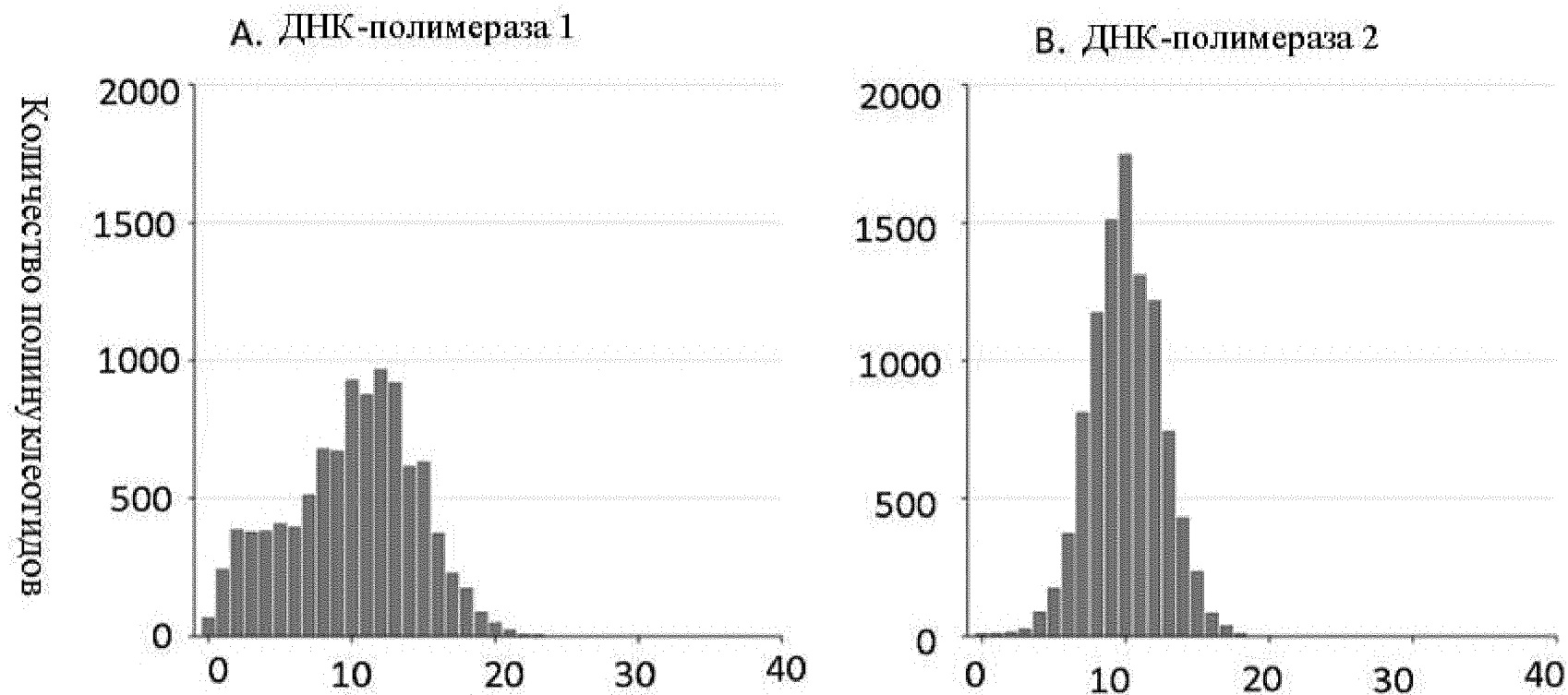
ФИГУРА 27



ФИГУРА 28

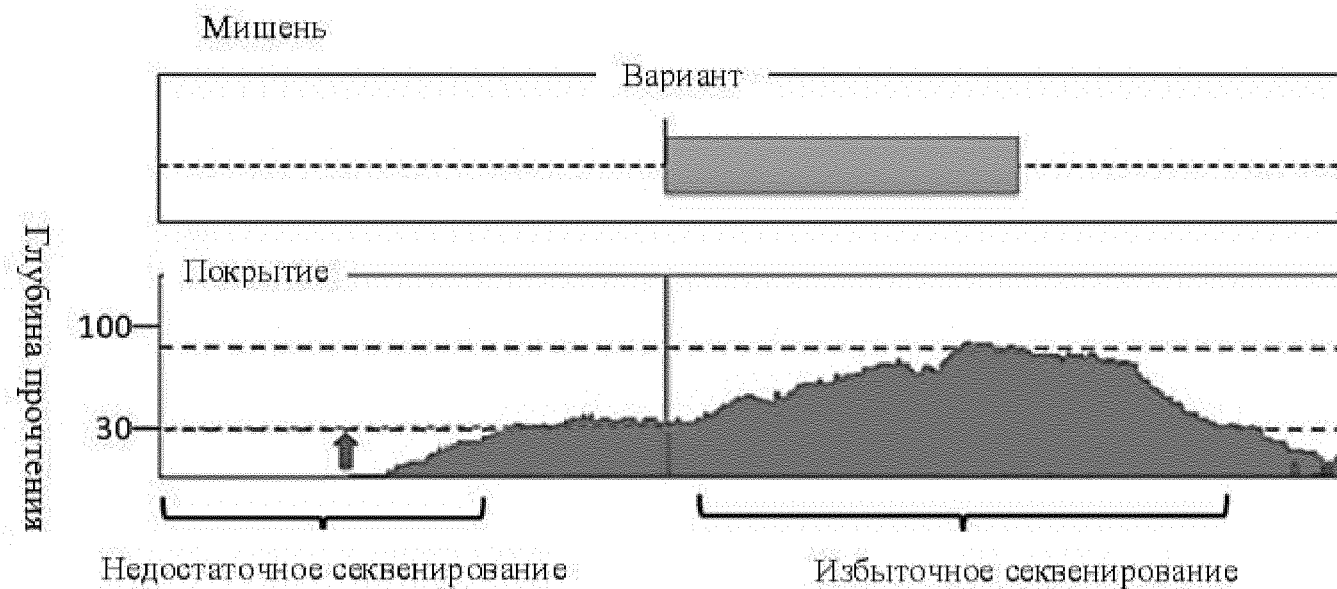


ФИГУРА 29

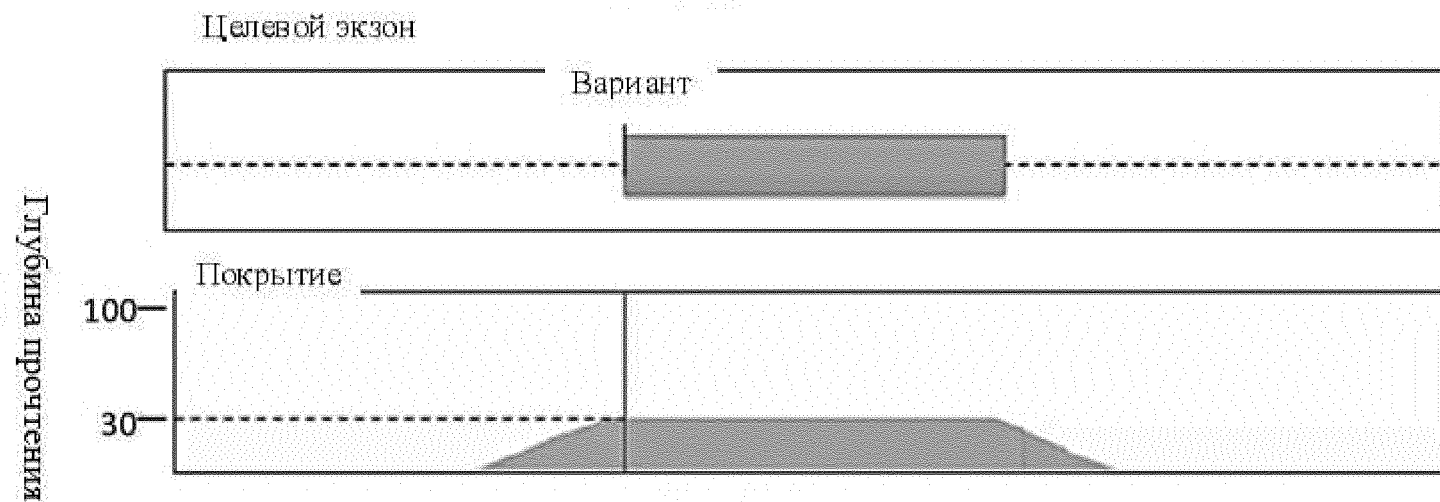


(импульсы полинуклеотидов на 100000 прочтений)

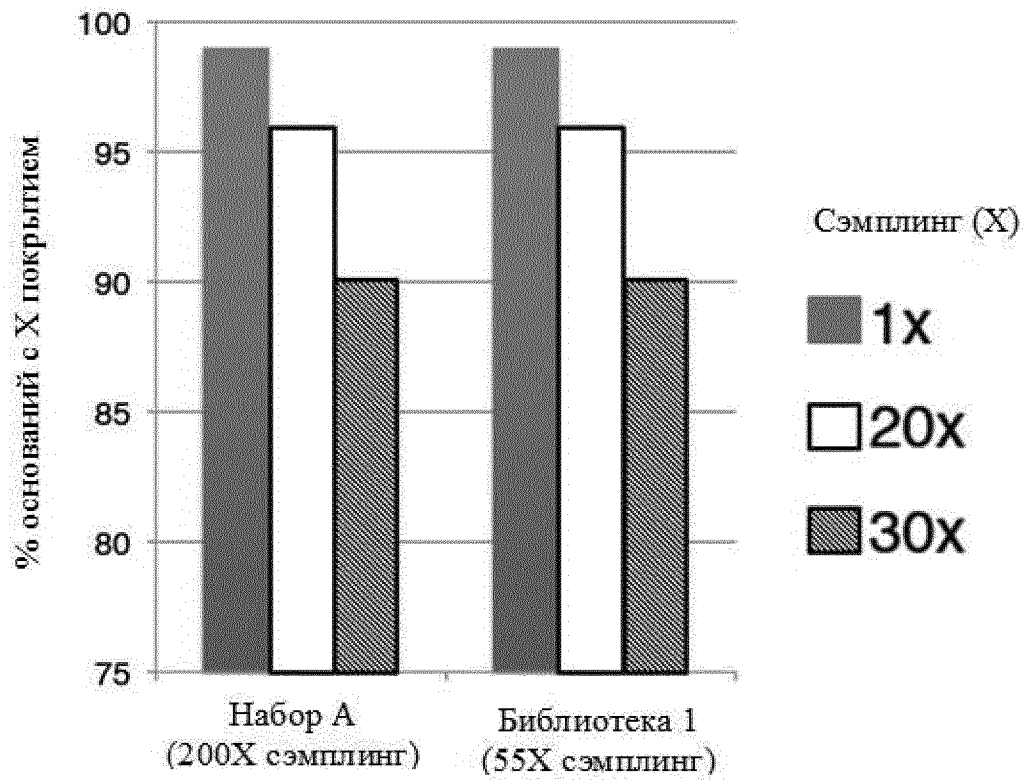
ФИГУРА 30



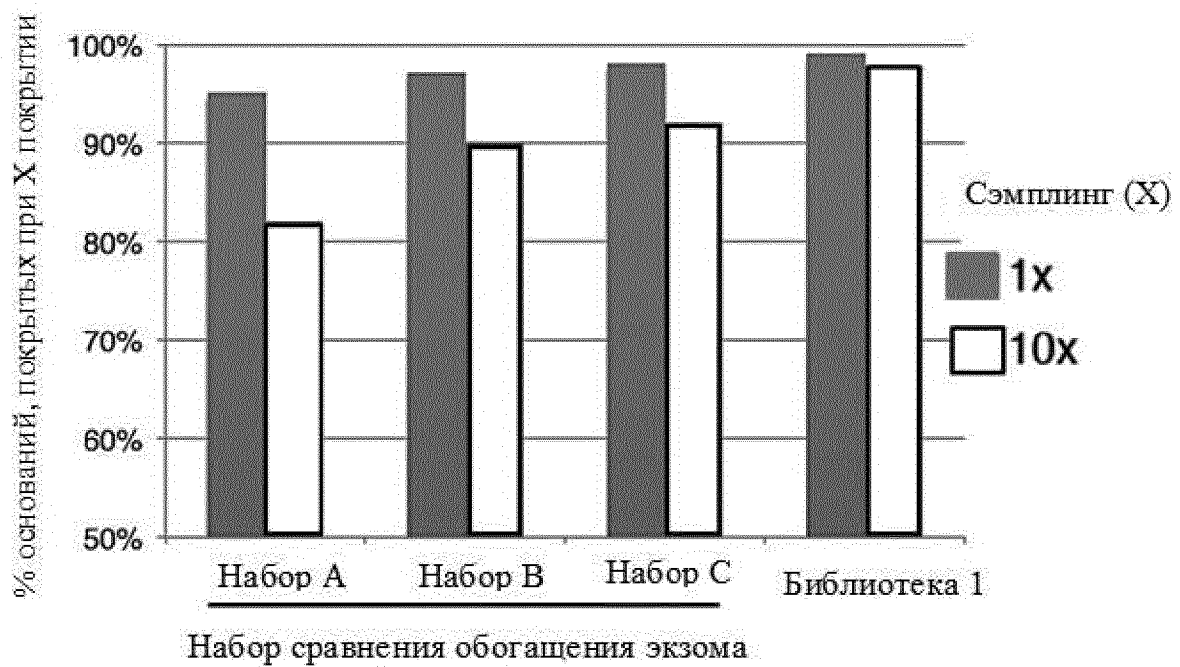
ФИГУРА 31А



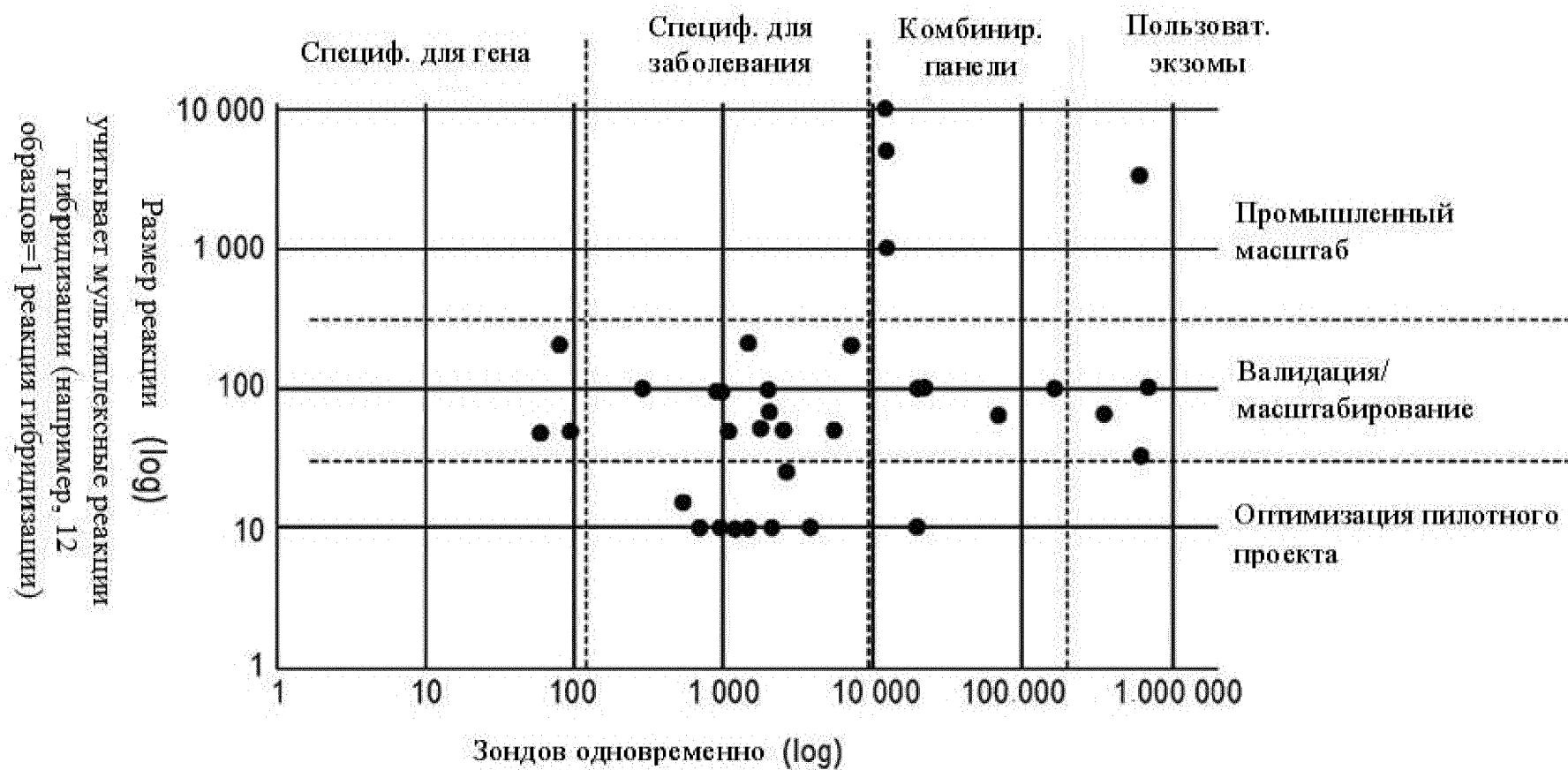
ФИГУРА 31В



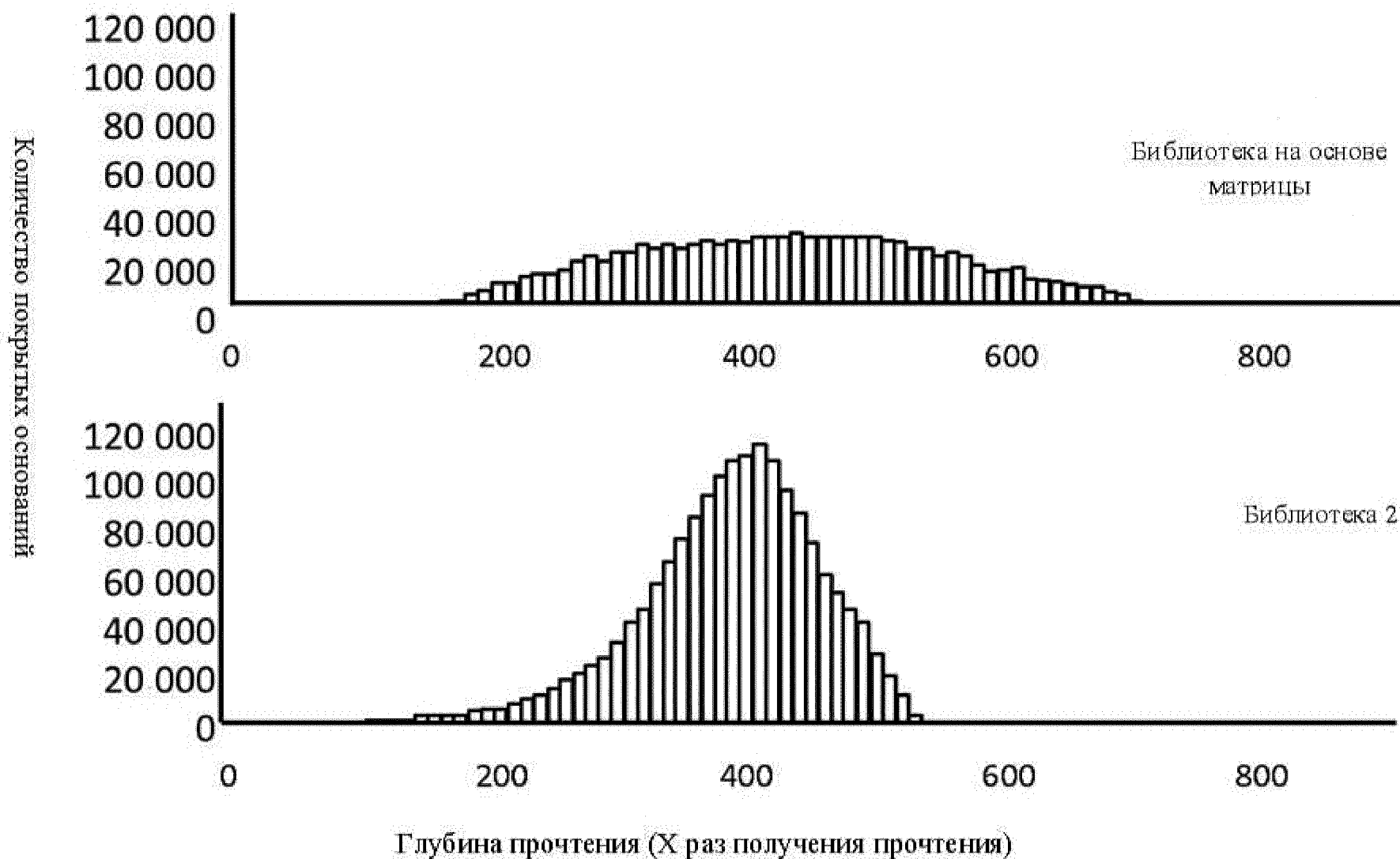
ФИГУРА 32А



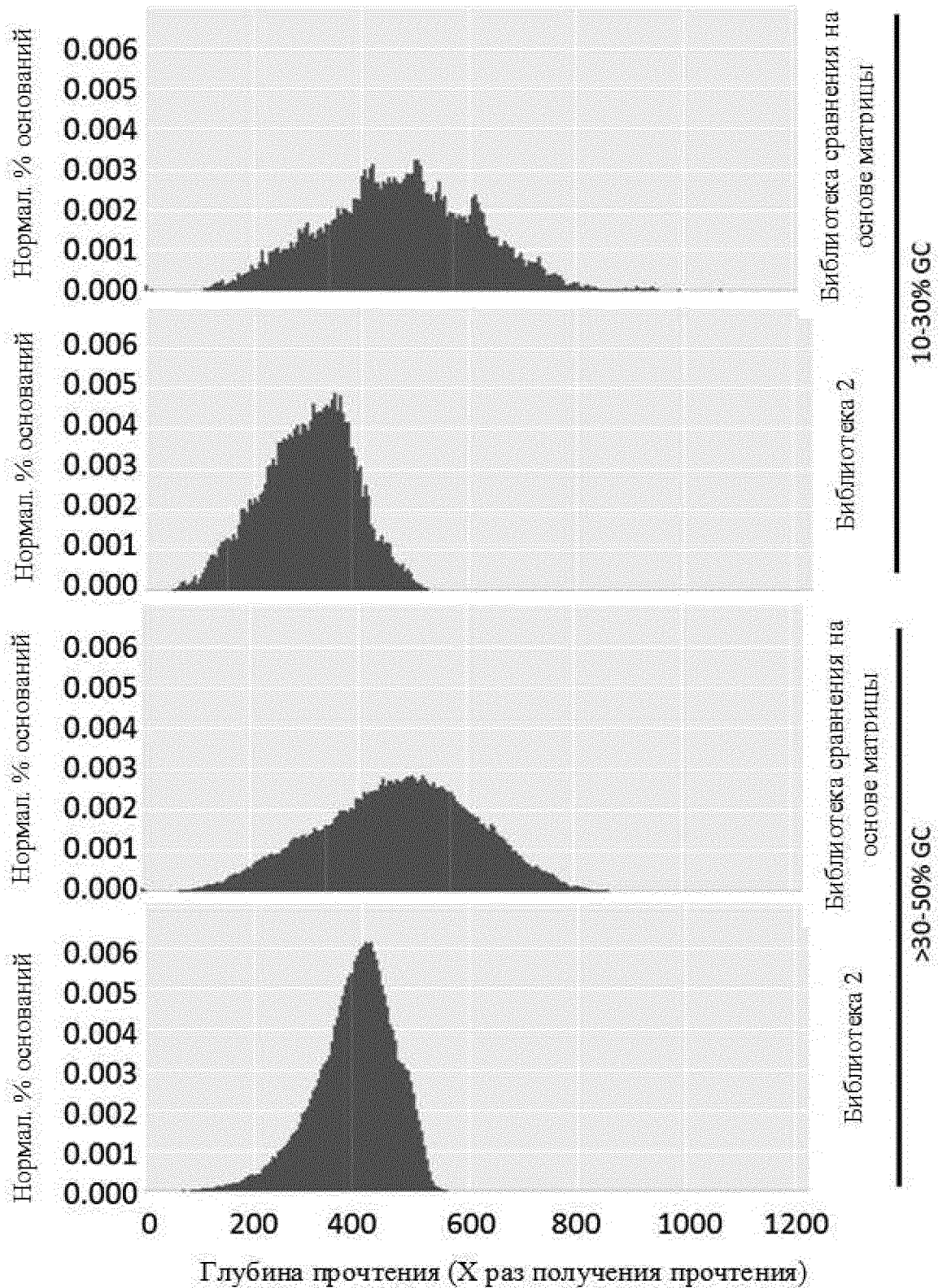
ФИГУРА 32В



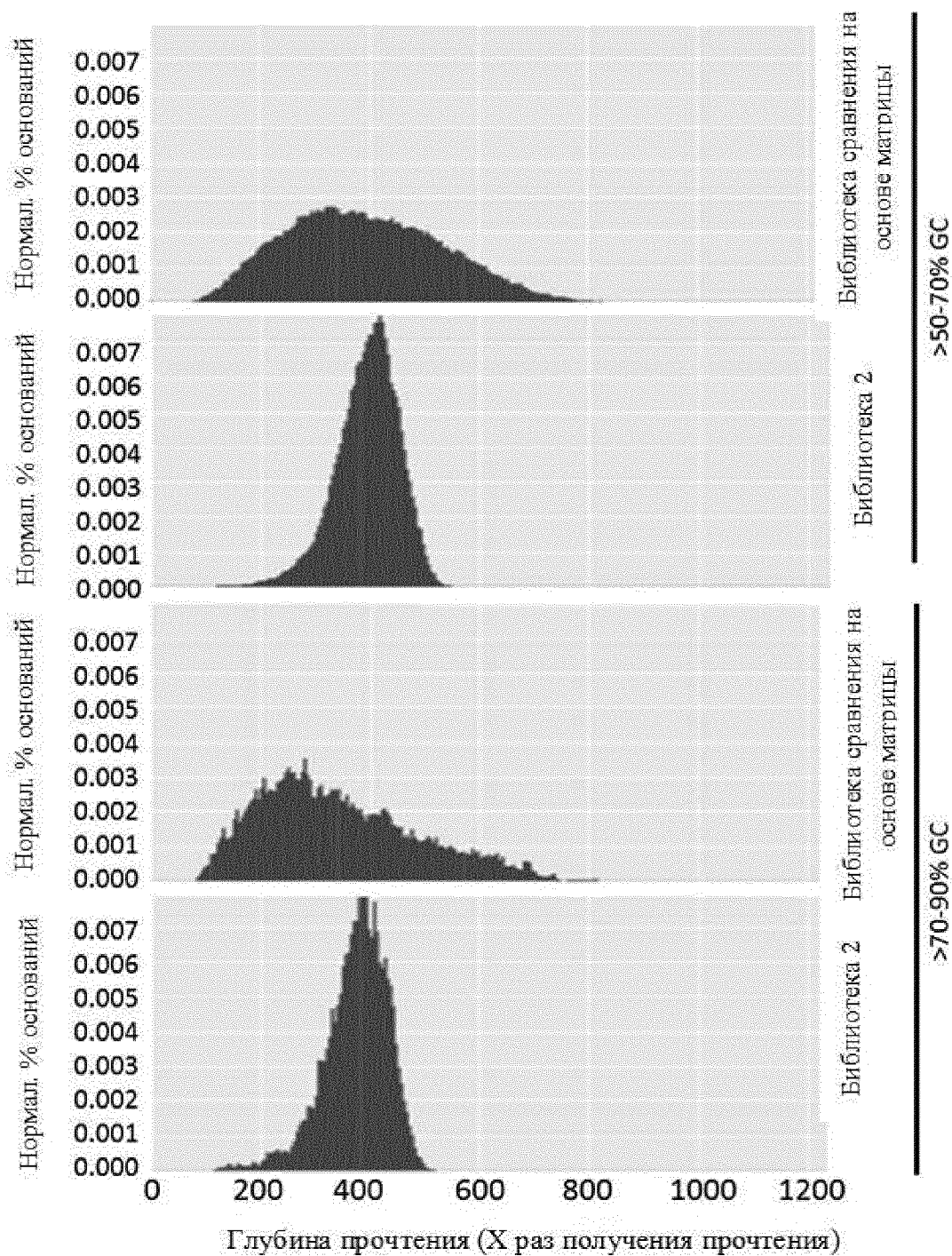
ФИГУРА 33



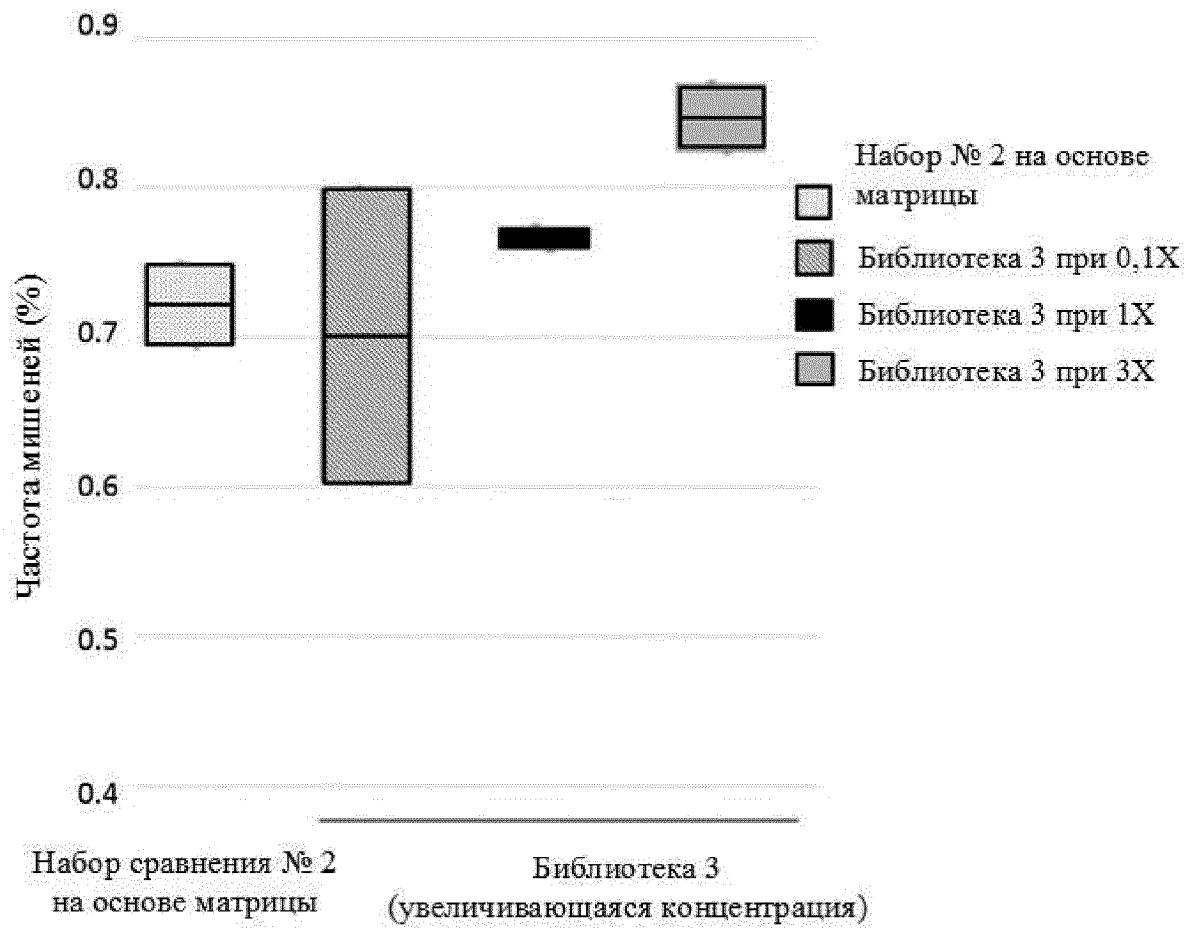
ФИГУРА 34



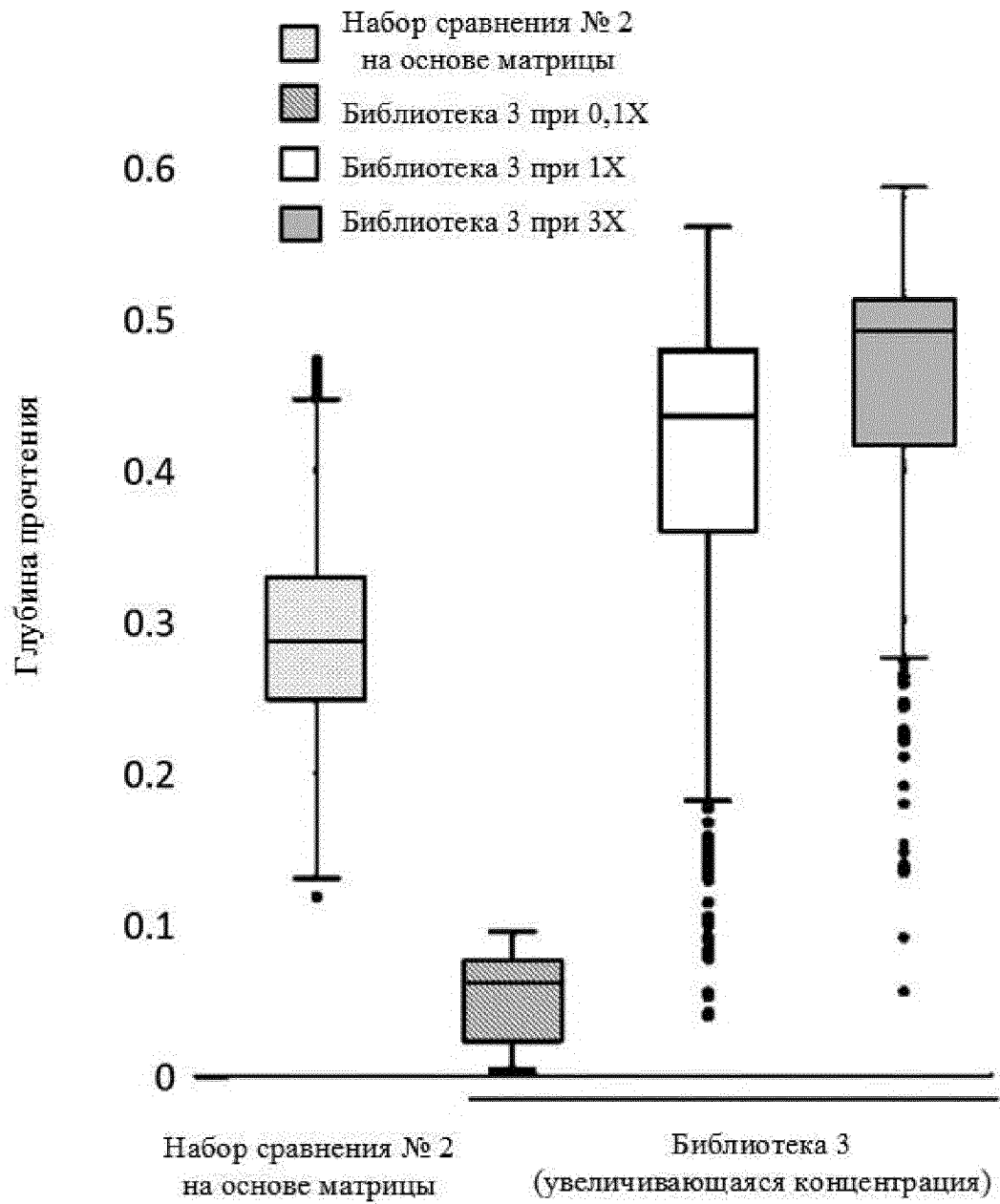
ФИГУРА 35А



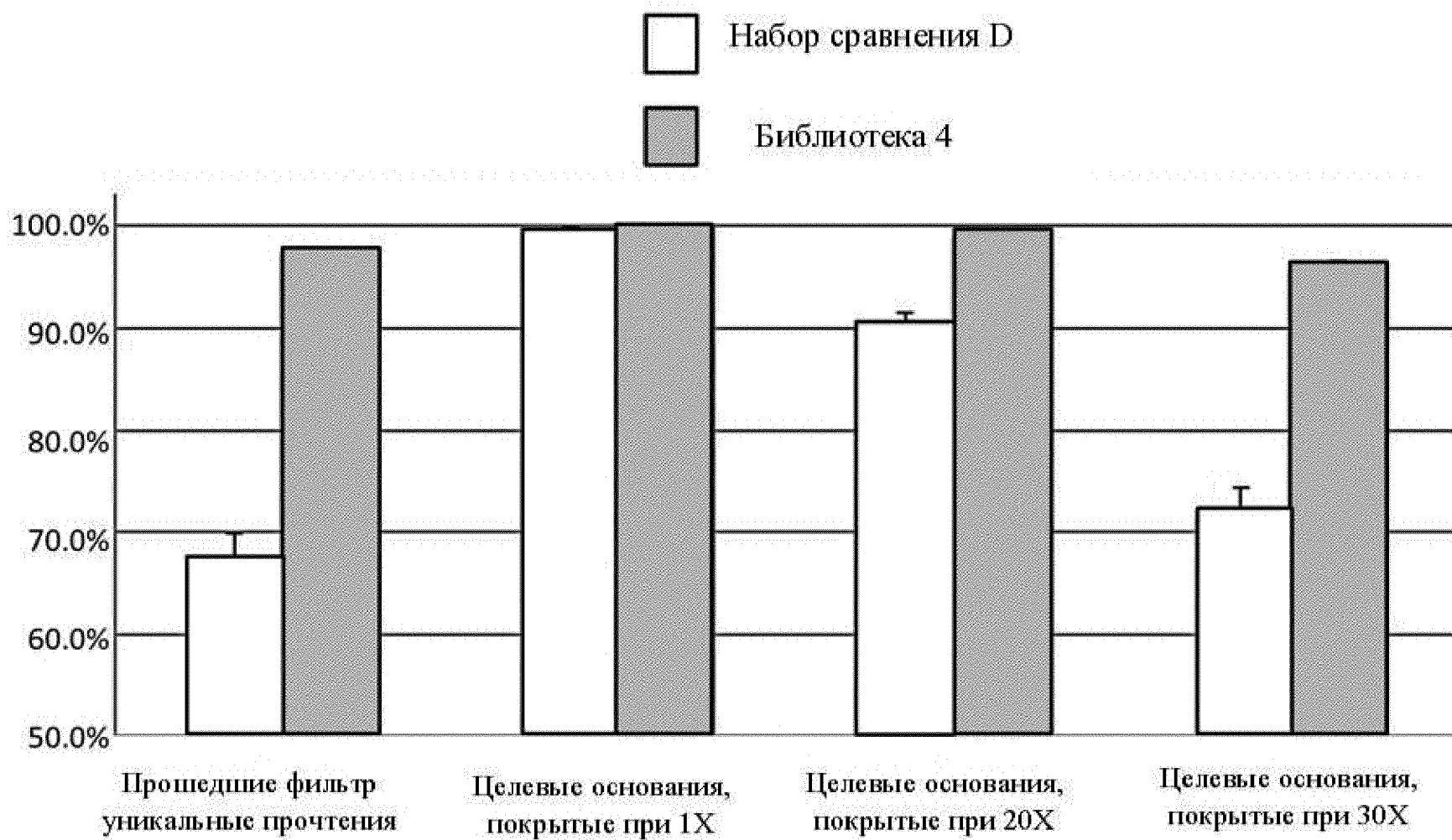
ФИГУРА 35В



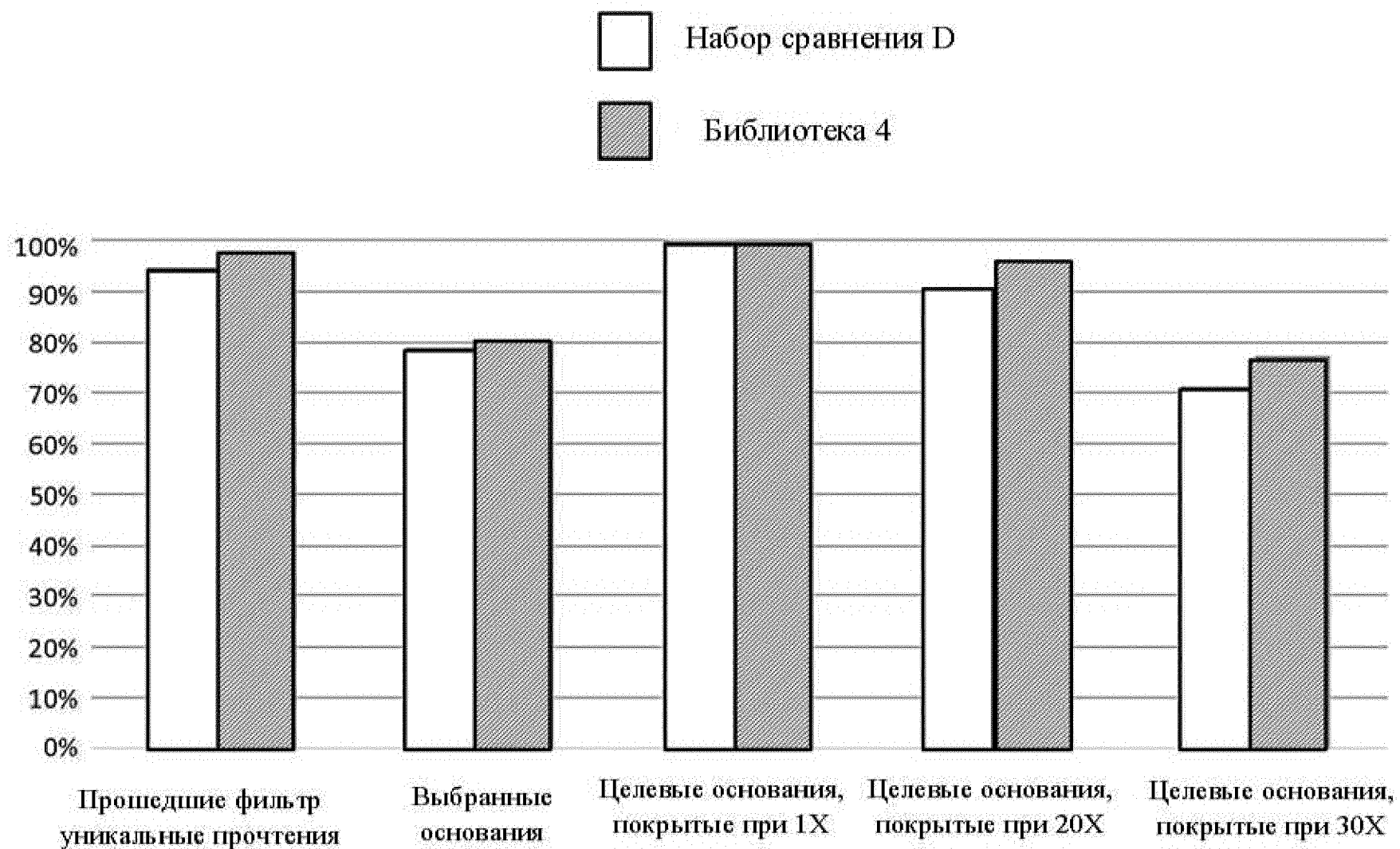
ФИГУРА 36А



ФИГУРА 36В



ФИГУРА 37



ФИГУРА 38