

(19)



**Евразийское
патентное
ведомство**

(11) **035092**(13) **B1**

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ

(45) Дата публикации и выдачи патента
2020.04.27

(51) Int. Cl. **C12N 15/10** (2006.01)
C12Q 1/68 (2006.01)

(21) Номер заявки
201692157

(22) Дата подачи заявки
2015.05.15

(54) СИНТЕЗ ДВУХЦЕПОЧЕЧНЫХ НУКЛЕИНОВЫХ КИСЛОТ

(31) **14168313.6**

(32) **2014.05.14**

(33) **EP**

(43) **2017.05.31**

(86) **PCT/EP2015/060777**

(87) **WO 2015/173402 2015.11.19**

(71)(73) Заявитель и патентовладелец:
**БУРВИНКЕЛЬ БАРБАРА;
ТУРЧИНОВИЧ АНДРЕЙ (DE)**

(72) Изобретатель:
**Турчинович Андрей, Суровы
Харальд, Бурвинкель Барбара (DE)**

(74) Представитель:
Нилова М.И. (RU)

(56) **US-A1-2008182239**

HOSHINO TATSUHIKO ET AL.: "A comparative study of microbial diversity and community structure in marine sediments using poly(A) tailing and reverse transcription-FOR", June 2013 (2013-06), FRONTIERS IN MICROBIOLOGY, VOL. 4, PAGE(S) ARTICLE NO.: 160, XP002742132, ISSN: 1664-302X(print) page 2, column 1, last paragraph - page 2, column 2, paragraph 2; figure 1

US-A1-2008009420

WO-A1-2013123442

SHI RUI ET AL.: "Poly(T) adaptor RT-PCR.", 2012, METHODS IN MOLECULAR BIOLOGY (CLIFTON, N.J.) 2012, VOL. 822, PAGE(S) 53-66, XP002742133, ISSN: 1940-6029, the whole document

PICELLI SIMONE ET AL.: "Full-length RNA-seq from single cells using Smart-seq2.", January 2014 (2014-01), NATURE PROTOCOLS JAN 2014, VOL. 9, NR. 1, PAGE(S) 171-181, XP002742134, ISSN: 1750-2799, the whole document
TURCHINOVICH ANDREY ET AL.: "Capture and Amplification by Tailing and Switching (CATS)", July 2014 (2014-07), RNA BIOLOGY, VOL. 11, NR. 7, PAGE(S) 817-828, XP002742135, ISSN: 1547-6286(print), the whole document

WO-A1-2015057319

US-A1-2014242581

(57) Настоящее изобретение относится к способу синтеза двухцепочечных нуклеиновых кислот из широкого спектра образцов и включает применение данных нуклеиновых кислот для глубокого анализа их последовательностей. Также настоящее изобретение относится к определенным реагентам, применяемым в способе согласно настоящему изобретению. Кроме того, настоящее изобретение относится к наборам, содержащим реагенты для применения в способе согласно настоящему изобретению, и к применению указанных наборов.

B1**035092****035092****B1**

Настоящее изобретение относится к способу синтеза двухцепочечных нуклеиновых кислот из широкого спектра образцов и включает применение данных нуклеиновых кислот для глубокого анализа последовательности. Также настоящее изобретение относится к определенным реагентам, применяемым в способе согласно настоящему изобретению. Кроме того, настоящее изобретение относится к наборам, содержащим реагенты для применения в способе согласно настоящему изобретению, и к применению указанных наборов.

Уровень техники

Для массового параллельного секвенирования (МПС) нуклеиновых кислот требуется получение амплифицированных библиотек, в которых участок ДНК, который необходимо отсеквенировать, расположен между известными 5'- и 3'-концевыми последовательностями. В современных способах конструирования библиотек для МПС применяют лигирование либо РНК-, либо ДНК-адаптера с 5'- и 3'-концами молекул РНК или ДНК из образца. Лигирование адаптеров представляет собой не только требующий больших затрат времени, но также низкоэффективный процесс, для которого необходимо исходное количество образцов нуклеиновых кислот на уровне микрограммов. Кроме того, полученные в результате этого библиотеки кДНК оказываются контаминированы побочными продуктами перекрестного лигирования и самолигирования адаптеров, и требуются дополнительные этапы очистки как до, так и после предварительной амплификации. Более десяти лет назад в Clontech Laboratories описали способ, в котором используется способность обратной транскриптазы вируса лейкоза мышей Молони (ОТ MMLV) переключаться на другую матрицу для присоединения выбранных адаптеров к 5'-концу кДНК, полученной из содержащих поли(А)-хвост молекул мРНК. В то же время последовательность 3'-адаптера включали в поли(dT)-праймер для обратной транскрипции. Данный принцип, получивший название SMART, на сегодняшний день используется в наборе для секвенирования Illumina Ultra Low RNA (Clontech) для получения полноразмерных кДНК-копий молекул мРНК из отдельной клетки. Тем не менее, в данном способе после синтеза матрицы все еще требуется: (1) фрагментация амплифицированной кДНК, (2) лигирование 5'-/3'-концевых адаптеров, специфичность которых определяется выбранной платформой, и (3) предварительная амплификация лигированных с адаптерами фрагментов ДНК. Хотя способ SMART позволяет получить кДНК для секвенирования, используя РНК в количестве, содержащемся в одной клетке, он требует больших затрат времени и средств и ограничен секвенированием мРНК. До настоящего времени подход, в котором использована способность ОТ MMLV к переключению на другую матрицу, еще не применялся для секвенирования (1) молекул РНК, отличных от длинных молекул РНК, и (2) любых молекул ДНК. В настоящем изобретении описан способ получения подготовленных для секвенирования двух- или одноцепочечных ДНК, предпочтительно библиотек ДНК, из количеств порядка пикограммов (пг) молекул либо РНК, либо ДНК за период времени, составляющий несколько часов. Малые молекулы (<150 п.о.) РНК или ДНК (например, миРНК (микроРНК), рiРНК (рiwiРНК), расщепленную или обработанную бисульфитом ДНК) можно непосредственно использовать в качестве исходного материала. Тем не менее, длинные РНК или ДНК необходимо сначала фрагментировать с помощью соответствующего подхода (например, обработки ультразвуком для ДНК или инкубации с Mg^{2+} для РНК). Способ согласно настоящему изобретению обеспечивает несколько преимуществ, которые включают существенное уменьшение количества времени, необходимого для получения подготовленной для секвенирования ДНК, которая может быть получена из ДНК или РНК, указанный способ значительно дешевле, чем любой из способов, известных из уровня техники. Цены современных доступных для приобретения наборов, позволяющих получить библиотеки кДНК для секвенирования РНК и ДНК нового поколения, находятся в диапазоне от 200 до 500 \$ за образец в зависимости от конкретного способа, типа набора и марки поставщика. Грубые оценки затрат, необходимых для получения одной библиотеки ДНК с применением способа согласно настоящему изобретению, по меньшей мере в 20 раз ниже, и способ согласно настоящему изобретению позволит секвенировать нуклеиновые кислоты из источников, из которых ранее секвенирование было невозможным вследствие минимальных количеств ДНК и/или РНК, которые можно было получить из образца. Примеры последних включают ДНК и РНК из небольших (диагностических) количеств жидкостей и твердых биоплатов, отдельных компартментов клеток (например, микроядер, эндоплазматического ретикулума), ископаемых остатков, останков вымерших организмов и криминалистических образцов, содержащих мелкие и высокофрагментированные молекулы ДНК. Настоящее изобретение отчасти основано на данных, что ДНК также может служить в качестве субстрата для обратной транскриптазы.

Краткое описание изобретения

В первом аспекте настоящего изобретения предложен способ синтеза двухцепочечной нуклеиновой кислоты с определенной 3'- и 5'-концевой последовательностью нуклеотидов с использованием образца, содержащего одноцепочечную нуклеиновую кислоту, включающий следующие этапы:

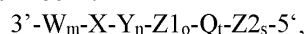
- а) обеспечение образца, содержащего одноцепочечную или двухцепочечную нуклеиновую кислоту, возможно денатурирование указанной двухцепочечной нуклеиновой кислоты;
- б) добавление по меньшей мере 5, предпочтительно от 10 до 50 последовательных нуклеотидов, к 3'-концу указанной одноцепочечной или двухцепочечной нуклеиновой кислоты;
- с) гибридизация иницирующего олигонуклеотида, комплементарного добавленной последователь-

ности нуклеотидов, и синтез кДНК или кРНК с помощью зависимой от матрицы ДНК- или РНК-полимеразы с получением двухцепочечной нуклеиновой кислоты;

д) гибридизация олигонуклеотида, переключающего матрицу (ОПМ), с указанной двухцепочечной нуклеиновой кислотой;

е) удлинение 3'-конца цепи кДНК или кРНК для синтеза двухцепочечной нуклеиновой кислоты, при этом одна цепь указанной нуклеиновой кислоты содержит иницирующий олигонуклеотид и кДНК или кРНК, которая комплементарна указанной одноцепочечной нуклеиновой кислоте и олигонуклеотиду, переключающему матрицу.

Во втором аспекте настоящего изобретения предложен иницирующий олигонуклеотид, содержащий следующие элементы последовательности:



где

W в каждом случае независимо выбран из dA, dG, dC, dT и dU;

X выбран из dA, dG, dC, dT, dU, rA, rG, rC, rT и rU;

Y представляет собой полинуклеотид длиной по меньшей мере 10 нуклеотидов, последовательность которого на 80% или более состоит из идентичных нуклеотидов или динуклеотидов, выбранных из dA, dG, dC, dT, dU, rA, rG, rC, rT, rU, AC, AG, AT, AU, CA, CG, CT, CU, GA, GC, GT, GU, TA, TC, TG, TU, AA, CC, GG, TT, UU, UA, UC, UG и UT, и на другие не более чем 20% или менее состоит из нуклеотидов или динуклеотидов, которые отличны от основного нуклеотида или динуклеотида и также выбраны из dA, dG, dC, dT, dU, rA, rG, rC, rT, rU, AC, AG, AT, AU, CA, CG, CT, CU, GA, GC, GT, GU, TA, TC, TG, TU, AA, CC, GG, TT, UU, UA, UC, UG и/или UT, при условии, что X отличен от нуклеотида или динуклеотида, который составляет большую часть Y;

Q представляет собой последовательность идущих друг за другом вырожденных (неоднозначных) оснований ДНК, предпочтительно выбранных из N, V, H, D, B и J, где N представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT, dC и dG; B представляет собой продукт включения нуклеотида из эквимольной смеси dT, dC и dG; D представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dG; H представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dC; V представляет собой продукт включения нуклеотида из эквимольной смеси dA, dC и dG; J представляет собой продукт включения нуклеотида из смеси (0-100% dA) с (0-100% dG), с (0-100% dC), с (0-100% dT), с (0-100% dU), с (0-100% rA), с (0-100% rG), с (0-100% rC), с (0-100% rT), с (0-100% rU);

Z1 представляет собой полинуклеотид длиной по меньшей мере 5 нуклеотидов с определенной последовательностью, при этом указанная последовательность отличается от W_m-X-Y_n , предпочтительно указанная последовательность также отличается от Q_t-Z2_s ;

Z2 представляет собой полинуклеотид длиной по меньшей мере 5 нуклеотидов с определенной последовательностью, при этом указанная последовательность отличается от $W_m-X-Y_n-Z1_o-Q_t$;

m представляет собой целое число от 0 до 6, т.е. 0, 1, 2, 3, 4, 5 или 6;

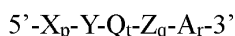
n представляет собой целое число от 10 до 100, если Y выбран из dA, dG, dC, dT, dU, rA, rG, rC, rT, и rU, целое число от 5 до 50, если Y выбран из AC, AG, AT, AU, CA, CG, CT, CU, GA, GC, GT, GU, TA, TC, TG, TU, AA, CC, GG, TT, UU, UA, UC, UG и UT;

o представляет собой 0 или 1;

s представляет собой 0 или 1; и

t представляет собой целое число от 0 до 6, т.е. 0, 1, 2, 3, 4, 5 или 6.

В третьем аспекте настоящего изобретения предложен олигонуклеотид, переключающий матрицу, содержащий следующие элементы последовательности



где

X представляет собой химическую группу, выбранную из группы, состоящей из остатка амина, биотина, глицерина, холестерина, диоксигенина, фтора или производных нуклеотида, включая нуклеотиды без азотистых оснований, дидезоксирибонуклеотиды, 3'-дезоксинуклеотиды, 2'-дезоксиинозин, 2'-дезоксиуридин;

Y представляет собой известную олигонуклеотидную последовательность;

Q представляет собой последовательность идущих друг за другом вырожденных (неоднозначных) оснований ДНК, предпочтительно выбранных из N, V, H, D, B и J, где N представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT, dC и dG; B представляет собой продукт включения нуклеотида из эквимольной смеси dT, dC и dG; D представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dG; H представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dC; V представляет собой продукт включения нуклеотида из эквимольной смеси dA, dC и dG; J представляет собой продукт включения нуклеотида из смеси (0-100% dA) с (0-100% dG), с (0-100% dC), с (0-100% dT), с (0-100% dU), с (0-100% rA), с (0-100% rG), с (0-100% rC), с (0-100% rT), с (0-100% rU);

Z представляет собой рибонуклеотид, выбранный из группы, состоящей из AMP, CMP, GMP, TMP и UMP;

A представляет собой химическую группу, выбранную из группы, состоящей из остатка амина, биотина, глицерина, холестерина, диоксигенина, фосфата, фтора или производных нуклеотида, включая нуклеотиды без азотистых оснований, дидезоксирибонуклеотиды, 3'-дезоксинуклеотиды, 2'-дезоксининозин, 2'-дезоксинуридин;

t представляет собой целое число от 0 до 6, т.е. 0, 1, 2, 3, 4, 5 или 6;

r представляет собой целое число от 0 до 10, т.е. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10;

q представляет собой целое число, равное по меньшей мере 1; и

г представляет собой целое число от 0 до 10, т.е. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10.

В четвертом аспекте настоящего изобретения предложена нуклеиновая кислота, содержащая иницирующий олигонуклеотид согласно второму аспекту настоящего изобретения.

В пятом аспекте настоящего изобретения предложен набор, содержащий

а) реагент, способный добавлять нуклеотиды к 3'-концу одноцепочечной нуклеиновой кислоты, предпочтительно фермент, более предпочтительно поли(А)-полимеразу или концевую трансферазу (КТ), и возможно блокирующий нуклеотид, предпочтительно 3d-NTP, 3-Me-NTP и ddNTP,

б) фермент обратную транскриптазу,

с) иницирующий олигонуклеотид согласно второму аспекту и

д) олигонуклеотид, переключающий матрицу, согласно третьему аспекту.

В шестом аспекте настоящего изобретения предложен чип, содержащий по меньшей мере одну нуклеиновую кислоту, содержащую иницирующий олигонуклеотид согласно четвертому аспекту настоящего изобретения.

В седьмом аспекте настоящего изобретения предложено применение указанного набора и применение указанной синтезированной двухцепочечной нуклеиновой кислоты в персонализированной медицине; для контроля (мониторинга) терапии; предсказания, прогнозирования, раннего обнаружения заболевания человека или животного или для криминалистического анализа последовательностей нуклеиновых кислот вирусов, бактерий, животных или растений или клеток, полученных из них.

Перечень фигур

В следующем разделе описано содержание фигур, включенных в настоящее описание. Для этого, пожалуйста, также обращайтесь к подробному описанию изобретения выше и/или ниже.

Фиг. 1. Схематическое представление способов получения кДНК с применением комбинации наращивания хвоста поли(А/дА) и способности ОТ MMLV переключать матрицу. Вкратце, короткую одноцепочечную РНК или фрагменты ДНК полиаденилируют или полидезоксиаденилируют с помощью либо поли(А)-полимеразы либо концевой дезоксинуклеотидилтрансферазы. Затем осуществляют синтез комплементарной цепи ДНК в присутствии заякоренного поли(дТ)-олигонуклеотида, содержащего специально разработанную последовательность 3'-адаптера. Возможно указанный олигонуклеотид содержит три различных нуклеотида на 3'-конце, т.е. С, G или А (=V в схематическом представлении на фиг. 1). Когда обратная транскриптаза достигает 5'-конца матрицы РНК (или ДНК), активность данного фермента как терминальной трансферазы приводит к добавлению дополнительных нуклеотидов (преимущественно dC), которые не кодируются матрицей. На следующем этапе в реакцию обратной транскрипции добавляют олигонуклеотид, переключающий матрицу, содержащий три концевых нуклеотида gG и специально разработанные 5'-адаптерные последовательности, который служит второй матрицей для обратной транскриптазы. Считают, что комплементарное взаимодействие трех последовательных нуклеотидов gG на 3'-конце ОПМ и dC-богатой удлиненной последовательности кДНК способствует переключению матрицы. Вторую цепь кДНК получают в ходе первого цикла стандартной полимеразной цепной реакции (ПНР) с прямым праймером, который либо полностью, либо частично комплементарен 3'-концу первой цепи кДНК. Более того, обратный праймер, используемый для ПЦР-амплификации кДНК (вместе с прямым праймером), либо полностью, либо частично комплементарен 3'-концу второй цепи кДНК.

Фиг. 2. Для конструирования библиотек ДНК, подходящих для платформ Illumina MiSeq или HiSeq, мы использовали адаптерные последовательности из набора для секвенирования малых РНК NEBnext (New England Biolabs). Последовательность, соответствующую 5'-адаптеру, включали в ОПМ, и последовательность 3'-адаптера использовали для дизайна концевого участка поли(дТ)-праймера (фиг. 2А). Использовали либо 1 нг, либо 5 пг РНК и ДНК длиной 22 н.т. в качестве начального материала (матрицы) для получения библиотеки ДНК (фиг. 2В). Эффективность синтеза кДНК была одинакова для ДНК и РНК. Когда использовали 1 нг нуклеиновых кислот, то обнаружили один продукт ПЦР после 17 циклов ПЦР-амплификации (разведение кДНК в реакционной смеси ПЦР 1/100). Когда использовали 5 пг нуклеиновых кислот в качестве затравки, то количество циклов ПЦР, необходимое для амплификации кДНК, увеличивалось до 26. Когда использовали разведение кДНК в реакционной смеси ПЦР 10/100, то количество циклов, необходимое для получения библиотеки ДНК, уменьшалось пропорционально (результаты не представлены). Единственным контаминирующим побочным продуктом в данной реакции был избыток праймеров для ПЦР, большинство которых можно удалить с помощью очистки на колонке. С помощью секвенирования по Сэнгеру дополнительно подтвердили, что кДНК, полученная из синтети-

ческой короткой ДНК, была чистой (результаты не представлены).

Фиг. 3. Критические параметры протокола I получения библиотеки ДНК. В первую очередь, критической для оптимального выхода кДНК является реакция присоединения поли(А)-хвоста. Слишком длинные поли(А)-хвосты в конечном счете уменьшают эффективную концентрацию поли(dT)-праймера, что не только уменьшит количество кДНК, но также приведет к образованию размытого пятна побочных продуктов большего размера на геле, поскольку поли(dT)-праймер будет гибридизоваться с различными сайтами на поли(А)-хвосте. На верхней панели фиг. 3А показана электрофореграмма, полученная в результате электрофореза в 3% агарозном геле, для 1 нг cel-miR-39, к которой присоединили поли(А)-хвост, используя различное время инкубации и различные концентрации АТР. На нижней панели показана электрофореграмма, полученная в результате электрофореза в 3% агарозном геле библиотек ДНК, полученных из 1 нг соответствующих содержащих поли(А)-хвост cel-miR-39 с применением 100 нМ ILPdTPo. На фиг. 3В показана электрофореграмма, полученная в результате электрофореза в 3% агарозном геле библиотек ДНК, полученных из 1 нг cel-miR-39 (к которым присоединяли поли(А)-хвост в течение 10 мин, применяя различные концентрации АТР), применяя либо 100 нМ заякоренного на одном основании поли(dT)-праймера (ILPdTPo), либо 100 нМ заякоренного на двух основаниях поли(dT)-праймера (ILPcTPt). На фиг. 3С показаны электрофореграммы, полученные в результате электрофореза в 3% агарозном геле библиотек ДНК, полученных из 1 нг cel-miR-39 (к которым присоединяли поли(А)-хвост в течение 10 мин, применяя 0,1 мМ АТР) с применением ОТ MMLV различных торговых марок и либо 1 мкМ, либо 0,1 мкМ ОПМ8. Верхняя фигура: ПЦР-амплификацию молекул кДНК осуществляли на протяжении 17 циклов. Нижняя фигура: ПЦР-амплификацию молекул кДНК осуществляли на протяжении 21 цикла. Присоединение поли(А)-хвоста в течение 10 мин и конечная концентрация АТР 0,1 мМ привели к хорошим результатам клонирования РНК длиной 22 н.т. Во-вторых, поставщик и торговая марка ОТ MMLV оказались критичны для чувствительности данного подхода. Таким образом, из 6 коммерчески доступных ОТ MMLV Superscribe II (Invitrogen), ОТ SMARTScribe (Clontech) и ОТ SMART (Clontech) позволяли наиболее эффективно получить детектируемые количества кДНК после амплификации по данному протоколу, тогда как для SuperScribe III (Invitrogen), ОТ Multiscribe (Applied Biosystems) и MMLV от NEB требовалось 4 дополнительных цикла амплификации, чтобы библиотеку ДНК было видно на агарозном геле (фиг. 3С). Данный феномен можно объяснить тем фактом, что у различных вариантов ОТ MMLV могут быть различные активности РНКазы Н и терминальной трансферазы (считают, что последняя из упомянутых способствует реакции переключения матрицы). Таким образом, выбор обратной транскриптазы с активностью РНКазы Н предпочтителен.

Фиг. 4. Критические параметры протокола II получения библиотеки кДНК.

Показана электрофореграмма, полученная в результате электрофореза в 3% агарозном геле библиотек ДНК, полученных из 1 нг cel-miR-39 (к которой присоединяли поли(А)-хвост в течение 10 мин, применяя 0,1 мМ АТР) с применением различных олигонуклеотидов, переключающих матрицу (ОПМ), в конечной концентрации 1 мкМ. Верхняя фигура: ПЦР-амплификацию молекул кДНК осуществляли на протяжении 17 циклов. Нижняя фигура: ПЦР-амплификацию молекул кДНК осуществляли на протяжении 21 цикла (фиг. 4А). Структура ОПМ, похоже, была критична для чувствительности и производительности данного способа. ОПМ, состоящие как только из ДНК, так и только из РНК не позволили получить какое-либо адекватное количество целевой кДНК после 17 циклов предварительной амплификации ПЦР. Это можно объяснить тем фактом, что последовательность трех рибо-G обладает гораздо большей аффинностью для переключения матрицы, чем последовательность трех дезоксирибо-G, при этом чистый РНК олигонуклеотид склонен образовывать значительные вторичные структуры, которые уменьшают доступность 3'-конца. Более того, когда использовали ОПМ с четырьмя концевыми рибо-G нуклеотидами вместо трех, выход кДНК резко снизился (фиг. 4А), предположительно вследствие способности четырех последовательных G образовывать квадруплексные структуры. Также исследовали вариант блокирования концевой 3'-ОН группы ОПМ для предотвращения присоединения к ней поли(А)-хвоста, которое может произойти, если поли(А)-полимераза не полностью деактивирована. Несмотря на проведение термической деактивации поли(А)-полимеразы E. coli в течение 20 мин при 65°C перед реакцией обратной транскрипции, использование 3'-ОН-блокировки ОПМ будет обязательным в том случае, если (1) присоединение поли(А)-хвоста и обратную транскрипцию осуществляют одновременно или (2) присоединение поли(А)-хвоста к РНК невозможно инактивировать нагреванием. Неожиданно блокирование 3'-ОН-конца ОПМ с помощью либо монофосфата, либо биотина подавляло синтез кДНК при используемых условиях (фиг. 4А). Тем не менее, если 3'-ОН-группу ОПМ блокировали с помощью фосфата или дидезоксицитидина (ddC), аналогичные количества кДНК-продукта появлялись на четыре цикла ПЦР позже. На фиг. 4В показана электрофореграмма, полученная в результате электрофореза в 4% агарозном геле (слева), и анализ с помощью Agilent Bioanalyser (справа) библиотек ДНК, полученных из 1 нг cel-miR-39 (к которым присоединяли поли(А)-хвост в течение 10 мин, применяя 0,1 мМ АТР) с применением ОПМ3 с неблокированным 5'-концом или ОПМ8 с заблокированным биотином 5'-концом. Обратите внимание на небольшую фракцию более длинных (на уровне ~30 п.о.) библиотек ДНК, которые, вероятно, соответствуют продуктам эпизодов вторичного переключения матрицы (белая стрелка).

Фиг. 5. Критические параметры протокола III получения библиотек ДНК. Панель А.

Показаны электрофореграммы, полученные в результате электрофореза в 3% агарозном геле библиотек ДНК, полученных из 1 нг различных матриц - олигонуклеотидов cel-miR-39 (к РНК присоединяли поли(А)-хвост в течение 10 мин, применяя 0,1 мМ АТР; к ДНК присоединяли поли(dA)-хвост в течение 30 мин, применяя 0,1 мМ АТР). Эффективность синтеза библиотек ДНК по ДНК-матрицам, содержащим 5'-биотин, была значительно ниже, чем таковая для матриц, содержащих 5'-ОН или 5'-фосфат.

Панель В.

Показана электрофореграмма, полученная в результате электрофореза в 3% агарозном геле библиотек ДНК, полученных из 1 нг cel-miR-39 РНК (к которым присоединяли поли(А)-хвост в течение 10 мин, применяя 0,1 мМ АТР) с применением либо воды, либо 20% ДМСО (5% в конечном объеме реакционной смеси) в качестве среды для реакции обратной транскрипции. Добавление ДМСО не нарушает эффективность получения библиотек ДНК.

Фиг. 6. Получение библиотек ДНК из РНК и ДНК человека. Панель А.

Показана электрофореграмма, полученная в результате электрофореза в 4% агарозном геле (слева) библиотек ДНК, полученных из 1 нг контрольной cel-miR-39 (C39R) и 1 нг поли(А)-обогащенной РНК, выделенной из клеток U2OS, которую фрагментировали путем инкубации с ионами магния в течение 10 мин (R10). Кроме того, одну библиотеку ДНК получили из образца R10, который не был предварительно обработан полинуклеотидкиназой T4 (ПНК T4) перед присоединением поли(А)-хвоста (указан как -ПНК). Количество циклов ПЦР, использованных для предварительной амплификации библиотек кДНК, и концентрация обратного праймера с поли(dT) (ILPdTPo) указаны для каждого образца под электрофореграммой. Библиотеки ДНК, которые секвенировали на Illumina MiSeq, вырезали из агарозного геля, выделяли с помощью набора PureLink Gel Purification и анализировали с помощью высокочувствительных ДНК-чипов Agilent Bioanalyser (справа).

Панель В. Слева.

Показана электрофореграмма, полученная в результате электрофореза в 3% агарозном геле библиотек ДНК, полученных из приблизительно 3 нг обработанной бисульфитом ДНК из клеток U2OS (фиг. 6В). Кроме того, одну библиотеку ДНК получили из образца В, который был предварительно обработан ПНК T4 перед реакцией присоединения поли(dA) (указан как +ПНК). В библиотеке отрицательного контроля использовали 1 мкл воды (H₂O).

Справа.

Электрофореграмма Agilent Bioanalyser, на которой показаны очищенные в геле библиотеки ДНК, полученные из 1 нг обогащенной поли(А) РНК из клеток U2OS, которую фрагментировали путем инкубации с ионами магния в течение 5 мин (R5), обработанной бисульфитом ДНК (В) и 1 нг контрольной cel-miR-39 РНК (C39R).

Панель С.

Электрофореграмма, полученная в результате электрофореза в 4% агарозном геле (слева) и с помощью Agilent Bioanalyser (справа) библиотек ДНК, полученных из приблизительно 150 пг ДНК плазмы крови человека, выделенной из двух здоровых доноров (DI и DII). В контрольных экспериментах использовали либо воду (H₂O), либо 1 нг синтетической ДНК cel-miR-39 (C39D). Количество циклов ПЦР, используемых для предварительной амплификации библиотек кДНК, и концентрация обратного праймера с поли(dT) (ILPdTPo) указаны для каждого образца под электрофореграммой.

Панель D.

Электрофореграмма, полученная в результате электрофореза в 4% агарозном геле библиотек ДНК, полученных из приблизительно 200 пг РНК плазмы крови человека, выделенной из двух здоровых доноров (RI и RII). В контрольных экспериментах использовали либо воду (H₂O), либо 1 нг синтетической РНК cel-miR-39 (C39R). Кроме того, библиотеки ДНК получали из образцов циркулирующей РНК, которые предварительно не обрабатывали ПНК T4 перед присоединением поли(А)-хвоста (указан как -ПНК). Количество циклов ПЦР, используемых для предварительной амплификации библиотек кДНК, и концентрация обратного праймера с поли(dT) (ILPdTPo) указаны для каждого образца под электрофореграммой. Библиотеки ДНК, полученные из циркулирующей РНК плазмы обоих индивидов, очищали из агарозного геля, тем не менее, только библиотеку RI секвенировали на Illumina MiSeq.

Фиг. 7. Достоверность считывания индексной последовательности для образцов библиотек, объединенных на одной дорожке секвенирования, в зависимости от последовательности иницирующего олигонуклеотида.

Показаны соответствующие соотношения считываний индексной последовательности, которые были обнаружены без ошибок и с одной ошибкой в зависимости от состава последовательности иницирующего олигонуклеотида, используемого для получения двухцепочечной нуклеиновой кислоты, а точнее от части, которая комплементарна поли(А)-хвосту, созданному на предыдущем этапе данного способа. Восемь библиотек фрагментов ДНК получали параллельно из идентичных количеств (1 нг) исходного источника материала (геномной ДНК человека), применяя четыре различных иницирующих олигонуклеотида, как показано, и используя два повтора для каждого олигонуклеотида. Восемь полученных библиотек предварительно амплифицировали с праймерами, подходящими для мультиплексного секвенирования на современных системах секвенатора Illumina, при этом каждый обратный праймер содержал от-

личную последовательность индекса, которую использовали для определения библиотеки, к которой принадлежит данное считывание. Используемые индексные последовательности соответствуют индексным последовательностям Illumina 1-8, и каждая индексная последовательность отличается по меньшей мере по 3 положениям от всех остальных. Эквивалентные количества каждой библиотеки объединяли и секвенировали с одного конца на одной дорожке с 70 циклами для считывания ДНК последовательности в направлении 1 и 6 циклами для считывания индексной последовательности на системе Illumina MiSeq. Для каждой из восьми библиотек с различными индексными последовательностями регистрировали соответствующее количество считываний последовательности индекса, не содержащих ошибки или содержащих одну ошибку. Представлены средние значения частоты считываний индексных последовательностей с 0 или 1 ошибкой для двух библиотек, полученных с помощью каждого из четырех различных типов иницирующих олигонуклеотидов, соответственно где планки погрешностей обозначают отклонение от средних значений для отдельных библиотек. Очевидно, что применение иницирующего олигонуклеотида "20G" позволяет значительно повысить достоверность считывания последовательности индекса по сравнению с иницирующим олигонуклеотидом "30A", при этом сохраняется такая же эффективность получения библиотеки фрагментов ДНК.

Фиг. 8. Преимущество контролируемого по сравнению с неконтролируемым добавлением полинуклеотидного хвоста к матрицам ДНК и РНК.

В данном примере продемонстрировано полезное влияние контролируемого добавления поли(А)- и поли(dA)-хвоста на выработку кДНК, полученной из синтетической ДНК cel-miR-39 (слева) и РНК cel-miR-39 (справа). Контролируемое добавление поли(А)- и поли(dA)-хвоста обеспечивает возможность более эффективного получения библиотек с применением такой же концентрации праймера для обратной транскрипции, и/или когда концентрация АТФ в растворе находится на субоптимальном уровне. Если отношение АТФ (или dАТФ) к матрице РНК (или ДНК) выше, чем оптимальное, то это приведет к образованию длинных (>300 н.т.) хвостов. Длинные полинуклеотидные хвосты снижают эффективную концентрацию поли(dT)-праймера, что уменьшает выход библиотеки и вызывает образование размытого пятна побочных продуктов большего размера на геле, поскольку избыток поли(dT)-праймера гибридизуется с сайтом в большом поли(А)-хвосте.

А. Электрофореграмма, полученная с помощью электрофореза в 3% агарозном геле библиотек ДНК, полученных после присоединения поли(dA)-хвоста к 1 нг ДНК-матрицы cel-miR-39 и с применением 10 нМ обратного праймера с поли(dT) (ILPdTPo) либо в присутствии (С), либо в отсутствие (NC) блокирующего нуклеотида ddАТФ (отношение dАТФ/ddАТФ составляет 1/50). Стоит отметить, что значительно большего выхода библиотеки после контролируемого присоединения поли(dA)-хвоста добились при такой же концентрации обратного праймера.

В. Электрофореграмма, полученная в результате электрофореза в 3% агарозном геле библиотек ДНК, полученных после присоединения поли(А)-хвоста к 1 нг РНК-матрицы cel-miR-39 либо в присутствии (С), либо в отсутствие (NC) блокирующего нуклеотида 3d-АТФ (отношение АТФ/3d-АТФ составляет 1/30). Стоит отметить, что отношение АТФ к РНК-матрице (1 мМ АТФ к 1 нг 22 н.т. матрицы) было субоптимальным. Также стоит отметить, что значительно большего выхода библиотеки в отсутствие размытого пятна побочных продуктов большего размера добились с помощью контролируемого присоединения хвоста.

Подробное описание изобретения

Перед ознакомлением с подробным описанием настоящего изобретения, представленным ниже, следует понять, что настоящее изобретение не ограничено конкретной методикой, протоколами и реагентами, описанными в данной заявке, так как они могут изменяться. Также должно быть очевидно, что терминология используется в данном тексте исключительно с целью описания конкретных вариантов реализации, и не предполагается, что она ограничивает объем настоящего изобретения, который будет ограничен только прилагаемой формулой изобретения. Если не указано иначе, все технические и научные термины, используемые в данном тексте, имеют те же значения, которые обычно понимает средний специалист в данной области.

В тексте настоящего описания цитируется несколько документов. Каждый из документов, цитируемых в данном тексте (включая все патенты, заявки на патент, научные публикации, спецификации, инструкции производителя и т.д.) выше или ниже, настоящим полностью включен посредством ссылки. Ни один материал, включенный в описание, не должен рассматриваться как признание того, что настоящее изобретение не дает права датировать такую публикацию более ранним числом в силу предшествующего изобретения. Некоторые из документов, цитируемых в данном тексте, описывают как "включенные посредством ссылки". В случае противоречия между определениями или идеями таких включенных ссылок и определениями или идеями, изложенными в настоящем описании, текст настоящего описания имеет преимущественную силу.

В следующем описании будут описаны элементы настоящего изобретения. Данные элементы перечислены в конкретных вариантах реализации, тем не менее, должно быть очевидно, что их можно объединить любым способом и в любом количестве, чтобы получить дополнительные варианты реализации. Различным образом описанные примеры и предпочтительные варианты реализации не должны истолко-

ываться как ограничивающие настоящее изобретение только явно описанными вариантами реализации. Следует понимать, что в настоящем описании обоснованы и включены варианты реализации, в которых объединены явно описанные варианты реализации с любым количеством описанных и/или предпочтительных элементов. Более того, любые перестановки и комбинации из всех описанных элементов в данном описании следует считать раскрытыми в описании настоящей заявки, если в контексте не указано иное.

Определения

Далее приведены определения некоторых терминов, часто используемых в настоящем описании. Данные термины в каждом случае их использования в остальной части настоящего описания будут иметь соответствующие определенные значения и предпочтительные значения.

В настоящем описании и в прилагаемой формуле изобретения формы единственного числа включают ссылку на формы множественного числа, если в контексте явно не указано иное.

В настоящем описании термин "нуклеиновая кислота" включает полимерные или олигомерные макромолекулы или большие биологические молекулы, необходимые для всех известных форм жизни. Нуклеиновые кислоты, которые включают ДНК (дезоксирибонуклеиновую кислоту) и РНК (рибонуклеиновую кислоту), состоят из мономеров, известных как нуклеотиды. Большинство встречающихся в природе молекул ДНК состоит из двух комплементарных биополимерных цепей, закрученных по спирали вокруг друг друга с образованием двойной спирали. Цепь ДНК также известна как полинуклеотиды, состоящие из нуклеотидов. Каждый нуклеотид состоит из содержащего азот нуклеосооснования, а также моносахарида, названного дезоксирибозой или рибозой, и фосфатной группы. Встречающиеся в природе нуклеосооснования включают гуанин (G), аденин (A), тимин (T), урацил (U) или цитозин (C). Нуклеотиды соединены друг с другом в цепочки посредством ковалентных связей между сахаром одного нуклеотида и фосфатом следующего, в результате чего образуется чередующийся сахарофосфатный остов. Если сахар представляет собой дезоксирибозу, то полимер называют ДНК. Если сахар представляет собой рибозу, то полимер называют РНК. Обычно полинуклеотид образуется с помощью фосфодиэфирных связей между отдельными мономерами нуклеотидов. В контексте настоящего изобретения термин "нуклеиновая кислота" включает, но не ограничен, рибонуклеиновую кислоту (РНК), дезоксирибонуклеиновую кислоту (ДНК) и их смеси, такие как, например, гибриды РНК-ДНК (в одной цепи), а также кДНК, геномную ДНК, рекомбинантную ДНК, кРНК и мРНК. Нуклеиновая кислота может состоять из полноразмерного гена или из его части, нуклеиновая кислота также может представлять собой микроРНК, миРНК или рРНК. МикроРНК представляют собой короткие молекулы рибонуклеиновой кислоты (РНК), длина которых в среднем составляет 22 нуклеотида, но может быть и больше, которые обнаружены во всех эукариотических клетках, т.е. в растениях, животных и в некоторых вирусах, и которые участвуют в регуляции экспрессии гена на транскрипционном и посттранскрипционном уровне. МикроРНК представляют собой посттранскрипционные регуляторы, которые связываются с комплементарными последовательностями на целевых транскриптах информационных РНК (мРНК), что обычно приводит к подавлению трансляции и замалчиванию гена. Малые интерферирующие молекулы РНК (миРНК), иногда известные как короткие интерферирующие РНК или замалчивающие РНК, представляют собой короткие рибонуклеиновые кислоты (молекулы РНК) длиной 20-25 нуклеотидов. Они участвуют в пути РНК-интерференции (РНКи), в котором они препятствуют экспрессии определенных генов. РРНК также представляют собой короткие молекулы РНК, которые обычно содержат 26-31 нуклеотид, название которых происходит от так называемых белков *riwi*, с которыми они связываются. Нуклеиновая кислота также может представлять собой искусственную нуклеиновую кислоту. Искусственные нуклеиновые кислоты включают полиамидо- или пептидо-нуклеиновую кислоту (ПНК), морфолино-нуклеиновую кислоту и запертую нуклеиновую кислоту (ЗНК), а также гликоль-нуклеиновую кислоту (ГНК) и треозо-нуклеиновую кислоту (ТНК). Каждая из данных искусственных нуклеиновых кислот отличается от встречающейся в природе ДНК или РНК изменениями в остове данной молекулы.

Термин "одноцепочечная нуклеиновая кислота" (оц нуклеиновая кислота), используемый в настоящем описании, относится к нуклеиновой кислоте, которая состоит только из одной полинуклеотидной цепи. Напротив, "двухцепочечная нуклеиновая кислота" (дц нуклеиновая кислота) состоит из двух полинуклеотидных цепей, в которых большинство нуклеотидов спарены по правилам спаривания оснований (А с Т и С с G в случае ДНК, А с U и С с G в случае РНК и А с U, Т с А или С с G в гибридах РНК/ДНК), водородные связи соединяют азотистые основания двух отдельных полинуклеотидных цепей с образованием двухцепочечной нуклеиновой кислоты. В двойных цепях также допустимы несовпадения. Несовпадение в двойной цепи возникает, если два нуклеотида, которые находятся в одном и том же положении в противоположных цепях, не соответствуют правилам спаривания оснований. Количество несовпадений, допустимых в данной двойной цепи, определяется длиной двойной цепи, составом оснований, условиями температуры и буфера, например концентрацией соли. В данной области хорошо известно, как данные параметры влияют на образование двойной цепи.

Термин "неоднозначное основание" или "вырожденное основание", используемый в контексте настоящего описания, относится к конкретному положению нуклеотида в синтетическом олигонуклеотиде ДНК или РНК, в котором может находиться более чем одно основание. "Неоднозначное основание" или "вырожденное основание" представляет собой комбинацию dA, dT, dG, dC, dU, A, T, G, C или U во всех

возможных молярных соотношениях. Широко используемые термины "неоднозначные основания" или "вырожденные основания" представляют собой последовательность идущих друг за другом вырожденных (неоднозначных) оснований ДНК, предпочтительно выбранных из N, V, H, D, B и J, где N представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT, dC и dG, т.е. он обозначает любой из dA, dT, dC и dG; B представляет собой продукт включения нуклеотида из эквимольной смеси dT, dC и dG, т.е. он обозначает любой из dT, dC и dG; D представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dG, т.е. он обозначает любой из dA, dT, и dG; H представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dC, т.е. он обозначает любой из dA, dT, и dC; V представляет собой продукт включения нуклеотида из эквимольной смеси dA, dC и dG, т.е. он обозначает любой из dA, dC и dG, J представляет собой продукт включения нуклеотида из смеси (0-100% dA) с (0-100% dG), с (0-100% dC), с (0-100% dT), с (0-100% dU), с (0-100% rA), с (0-100% rG), с (0-100% rC), с (0-100% rT), с (0-100% rU). Таким образом, олигонуклеотид, который содержит неоднозначное основание в некотором положении, будет содержать один конкретный нуклеотид из соответствующей указанной смеси. С другой стороны, смесь олигонуклеотидов будет включать различные олигонуклеотиды, которые содержат в соответствующем положении все нуклеотиды, входящие в соответствующую смесь. Соотношение олигонуклеотидов, включающих различные указанные нуклеотиды, определяют по соответствующему соотношению нуклеотидов, включенных в данное положение. Это можно проиллюстрировать с помощью последовательности ANG, которая является аббревиатурой для эквимольной смеси четырех различных олигонуклеотидов, а именно AAG, ACG, AGG и ATG. Таким образом, если указано, что праймер или олигонуклеотид содержит неоднозначное основание, то это подразумевает, что присутствует смесь праймеров или олигонуклеотидов, содержащая различные нуклеотиды в данном положении.

Термин "образец" относится к части или кусочку из ткани, органа или индивида, и он обычно меньше, чем такая ткань, орган или индивид, предположительно представляющие собой целые ткань, орган или индивид. После анализа образца получают информацию о статусе ткани или статусе здоровья или болезни органа или индивида. Примеры образцов включают, но не ограничены перечисленными: образцы жидкости, такой как кровь, сыворотка, плазма, синовиальная жидкость, лимфатическая жидкость, спинномозговая жидкость, менингеальная жидкость, секрет (жидкость) железы, тонкоигльный пунктат, спинномозговая жидкость и другие биологические жидкости (моча, слюна). Дополнительные примеры образцов включают культуры клеток или культуры ткани. Дополнительные примеры также включают образцы жидких и твердых биоптотов или твердые образцы, такие как экстракты из ткани. Образцы могут включать ископаемые остатки, останки вымерших организмов, образцы из растений, плодов, животных, микробов, бактерий, вирусов, грибов или клетки, полученные из них.

Термин "последовательные нуклеотиды" используют в настоящем описании по отношению к последовательности, состоящей из нуклеотидов, непрерывно следующих друг за другом.

Термин "нуклеотид без азотистого основания" используют в настоящем описании по отношению к соединению, которое может соединить два нуклеотида, образуя фосфодиэфирные связи с 3'-концом одного из нуклеотидов и 5'-концом другого нуклеотида, у которого отсутствует структура, позволяющая спаривание оснований с любым из встречающихся в природе нуклеотидов, т.е. с производным пиримидина или пурина, и который создает расстояние между 5'-ОН и 3'-ОН фланкирующих нуклеотидов, которое составляет по меньшей мере 90% от расстояния между 5'-ОН и 3'-ОН встречающегося в природе нуклеотида. Предпочтительно указанное расстояние составляет по меньшей мере 91, 92, 93, 94, 95, 96, 97, 98, 99 или 100% от расстояния между 5'-ОН и 3'-ОН встречающегося в природе нуклеотида. "Нуклеотид без азотистого основания" служит в качестве так называемого "заполнителя места" встречающегося в природе нуклеотида. Для специалиста очевидно, что заполнитель должен удлинять цепь нуклеотидов на длину, которая сходна с удлинением путем добавления встречающегося в природе нуклеотида.

Таким образом, нуклеотид без азотистого основания позволяет идущему перед ним и следующему за ним нуклеотидам образовывать пары оснований по Уотсону-Крику с тремя последовательными нуклеотидами, при этом первая и последняя пара оснований образуется с идущим перед ним и следующим за ним нуклеотидом. Для специалиста также очевидно, что обозначения 3'-ОН и 5'-ОН относятся к химическим группам, которые будут присутствовать в 3'-положении сахарного остова предшествующего нуклеотида и 5'-ОН последующего нуклеотида в отсутствие нуклеотида без азотистого основания. Если присутствует нуклеотид без азотистого основания, то предпочтительно, чтобы он был связан с предшествующим и последующим нуклеотидом посредством фосфодиэфирных связей. В ДНК сайты без азотистого основания получают путем гидролиза гликозидной связи с нуклеотидным основанием, оставляя только сахарофосфатный остов в данном положении. В клетке образование сайта без азотистого основания происходит в результате спонтанного события депуринизации/депиримидинизации под действием ионизирующего УФ-излучения или как промежуточный продукт эксцизионной репарации оснований. Так как такие сайты хрупкие, они легко подвергаются одноцепочечному/двухцепочечному разрыву, и если они не репарируются с помощью механизма эксцизионной репарации оснований, то удаление азотистого основания часто приводит к мутации в результате синтеза через повреждения в процессе репликации. Конкретное основание, включенное напротив данного повреждения, изменяется в зависимости от

организма и условий окружающей среды. Широко используемый синтетический нуклеотид без азотистого основания включает фуран без азотистого основания, называемый dSpacer (1,2-дидезоксирибоза), который представляет собой производное тетрагидрофурана, в котором метиленовая группа занимает 1 положение в 2-дезоксирибозе. dSpacer широко используют, чтобы имитировать сайт без азотистого основания в олигонуклеотиде. Другие доступные нуклеотиды без азотистых оснований включают rSpacer, Spacer 18, Spacer 9, Spacer C3, Spacer C12.

Термин "гибридизация" относится к присоединению одноцепочечной нуклеиновой кислоты, предпочтительно олигонуклеотида с известной последовательностью, к частично или полностью комплементарной последовательности одноцепочечной нуклеиновой кислоты при определенных температурных условиях, которые определяют на основании состава нуклеосоставов и длины нуклеотидов. "Гибридизацию" также можно понимать как относящуюся к процессу обнаружения определенных последовательностей нуклеиновых кислот. Последовательность нуклеиновой кислоты, кодирующую последовательность комплементарную последовательности, которую нужно обнаружить, можно использовать в качестве гибридационного зонда в соответствии со стандартными методиками гибридализации. При "гибридизации in situ" используют меченую комплементарную молекулу нуклеиновой кислоты, например цепь ДНК или РНК (т.е. зонд), чтобы выявить специфическую молекулу нуклеиновой кислоты, например последовательность ДНК или РНК, в образце, например в части или срезе ткани (in situ). Условия гибридализации известны специалистам в данной области, и их можно найти, например, в *Current Protocols in Molecular Biology*, John Wiley & Sons, Нью-Йорк, 6.3.1-6.3.6, 1991. Термин "умеренные условия гибридализации" используется в контексте настоящего изобретения по отношению к гибридализации в 2X хлориде натрия/цитрате натрия (SSC) при 30°C, а затем промывке в 1X SSC, 0,1% ДСН при 50°C. "Условия высокой строгости" относятся к гибридализации в 6X хлориде натрия/цитрате натрия (SSC) при 45°C, а затем промывке в 0,2 X SSC, 0,1 % ДСН при 65°C.

Термин "комплементарный", используемый в настоящем описании, относится к последовательности нуклеотидов, которая спаривается посредством нековалентных связей со всей целевой нуклеиновой кислотой или с ее участком. В каноническом спаривании оснований по Уотсону-Крику аденин (А) образует пару с тиминном, а гуанин - с цитозином в ДНК. В РНК тимин замещен на урацил. Поэтому А комплементарен Т и G комплементарен С. В РНК А комплементарен U и наоборот. Обычно, "комплементарный" относится к последовательности нуклеотидов, которая, по меньшей мере частично, комплементарна. В объем термина "комплементарный" также могут входить дуплексы, которые полностью комплементарны, так что каждый нуклеотид в одной цепи комплементарен каждому нуклеотиду в другой цепи в соответствующих положениях. В некоторых случаях нуклеотид может быть частично комплементарен мишени, при этом не все нуклеотиды комплементарны каждому нуклеотиду в целевой нуклеиновой кислоте во всех соответствующих положениях. Например, праймер может быть идеально (т.е. на 100%) комплементарным целевой нуклеиновой кислоте или у праймера и целевой нуклеиновой кислоты может быть некоторая степень комплементарности, которая меньше, чем идеальная (т.е. составляет 70, 75, 80, 85, 90, 95, 99%).

Термин "комплементарная ДНК" (кДНК), используемый в настоящем описании, относится к ДНК, синтезированной по матрице РНК в реакции, катализируемой ферментами, такими как, например, обратная транскриптаза и ДНК-полимераза. кДНК часто используют для клонирования эукариотических генов в проکاریотах. кДНК также получают природным способом с помощью ретровирусов (таких как ВИЧ-1, ВИЧ-2 или вирус иммунодефицита обезьян), а затем встраивают в геном хозяина, в котором она создает провирус. Термин кДНК также используется, обычно в контексте биоинформатики, по отношению к последовательности транскрипта мРНК, выраженного в виде ДНК-оснований (GCAT), а не РНК-оснований (GCAU). "Комплементарную РНК" (кРНК) понимают как цепь РНК, комплементарную данной матрице РНК.

В настоящем описании термин "зависящая от матрицы ДНК- или РНК-полимераза" относится к ферментам, которые обладают каталитической активностью, способной использовать матричную цепь нуклеиновой кислоты для синтеза второй цепи нуклеиновой кислоты, комплементарной цепи матрицы. Для данных ферментов требуется матрица, которая используется как основа для синтезируемой цепи. Предпочтительный пример представляет собой термин "обратная транскриптаза" (ОТ), который относится к ферменту, также называемому РНК-зависимой ДНК-полимеразой, и широко применяется для получения комплементарной ДНК по матрице РНК в процессе, который называют обратной транскрипцией. Каталитические активности данного фермента превращают одноцепочечную геномную РНК на первом этапе в гибрид РНК/ДНК, а на втором этапе - в двухцепочечную ДНК. Источниками обратной транскриптазы являются ретровирусы, например вирус иммунодефицита человека (ВИЧ), которому обратная транскриптаза необходима для репликации. Активность обратной транскриптазы также связана с репликацией концов хромосомы (ферментами теломеразы) и некоторыми мобильными генетическими элементами (транспозонами). Обычно у обратной транскриптазы есть две последовательные биохимические активности: РНК-зависимая ДНК-полимераза и ДНК-полимераза, которые вместе осуществляют транскрипцию. Дополнительно к функции транскрипции ретровирусные ОТ содержат домен, принадлежащий

к семейству РНКазы Н, который незаменим для репликации. Предпочтительно используют ОТ, которые обладают активностью РНКазы Н. ОТ применяют в лаборатории для молекулярного клонирования, секвенирования РНК, полимеразной цепной реакции и анализа генома. Было показано, что ОТ обладает активностью переключения матрицы, что означает, что она способна переключаться с одной матрицы на другую. ОТ, которые особенно подходят для способа, наборов и применений согласно настоящему изобретению, включают, но не ограничены перечисленными: обратную транскриптазу ВИЧ-1 из вируса иммунодефицита человека типа 1 (PDB 1HNV), обратную транскриптазу M-MLV из вируса мышиного лейкоза Молони, обратную транскриптазу AMV из вируса миелобластома птиц и теломеразы. ОТ могут включать обратную транскриптазу MMLV, которую можно получить от NEB, обратную транскриптазу Superscript II или Superscript III, которую можно получить от Invitrogen, обратную транскриптазу Multiscribe, которую можно получить от Applied Biosystems, обратную транскриптазу SMART MMLV или обратную транскриптазу SMARTScribe, которые можно получить от Clontech. Теломераза представляет собой другой пример обратной транскриптазы, обнаруженной во многих эукариотах, включая людей, которая содержит собственную матрицу РНК; данная РНК используется в качестве матрицы для репликации ДНК, и ее можно применять в контексте настоящего изобретения.

Термин "независимые от матрицы ДНК/РНК-полимеразы" относится к ферменту, катализирующему добавление нуклеотидов к 3'-концу молекулы ДНК и/или РНК. В отличие от большинства ДНК- и/или РНК-полимераз, для данных полимераз не требуется матрица, которая используется в качестве основы для синтеза соответствующей цепи. Предпочтительные примеры таких ферментов представляют собой ДНК/РНК-лигазы, концевые трансферазы и поли(А, U или С)-полимеразы. Предпочтительный субстрат для данных ферментов представляет собой 3'-выступ двухцепочечной нуклеиновой кислоты или 3'-конец одноцепочечной нуклеиновой кислоты, но они также могут добавлять нуклеотиды к тупым или обрезанным 3'-концам. Кобальт представляет собой кофактор, необходимый для некоторых из данных ферментов, в частности для концевых трансфераз, тем не менее указанный фермент также катализирует реакцию при добавлении Mg и Mn *in vitro*. Предпочтительные примеры концевых трансфераз для применения в контексте настоящего изобретения представляют собой концевую дезоксинуклеотидилтрансферазу (TdT), также называемую ДНК-нуклеотидилэкзотрансферазой (DNTE), или поли(N)-полимеразы, где N означает А, G или U. Поли(N)-полимеразы представляют собой предпочтительные ферменты в контексте настоящего изобретения и включают поли(А)-полимеразы, которые представляют собой класс ферментов, способных добавлять поли(А)-хвост к одноцепочечной нуклеиновой кислоте. В природе реакция добавления поли(А)-хвоста происходит на 3'-конце первичного транскрипта РНК. Поли(А)-хвост состоит из множества аденозинмонофосфатов и представляет собой участок, который состоит исключительно из оснований аденина. Происходящая в природе реакция добавления поли(А)-хвоста приводит к образованию зрелой мРНК для трансляции. Поли(А)-полимераза может использовать цитозин в качестве субстрата для получения поли(С)-хвостов. Более того, можно применять поли(U)-полимеразу и поли(G)-полимеразу, которые обладают такой же функцией, но используют урацил, аденин и гуанин для реакции добавления хвоста соответственно. Например, поли(U)-полимеразу можно применять для катализа независимого от матрицы добавления UMP из UTP или AMP из ATP к 3'-концу РНК и, таким образом, можно применять для реакции добавления поли-А-или поли-U-хвоста. "ДНК-лигаза" представляет собой другой предпочтительный пример независимой от матрицы полимеразы и относится к определенному типу ферментов, к лигазе, который способствует соединению цепей ДНК путем катализа образования фосфодиэфирных связей между 3'-гидроксильной группой на конце одной ДНК и 5'-фосфорильной группой другой. РНК также можно лигировать аналогичным образом. Как правило, в данной реакции участвует кофактор, и он обычно представляет собой ATP или NAD⁺. ДНК-лигазы могут использовать мононуклеотиды, ди-, три- или п-нуклеотиды для получения хвоста, состоящего из моно-, ди-, три-, п-нуклеотидов, где "n" предпочтительно составляет от 4 до 100 нуклеотидов.

В контексте настоящего изобретения предпочтительно лигировать олигонуклеотид с известной последовательностью с 3'-гидроксильным концом одноцепочечной ДНК. Аналогично РНК-лигазы представляют собой определенный тип ферментов, которые катализируют образование одной или более фосфодиэфирных связей между 3'-гидроксильной группой одного конца РНК или ДНК и 5'-фосфорильной группой РНК или ДНК. Предпочтительная РНК-лигаза для применения в контексте настоящего изобретения представляет собой РНК-лигазу T4 или РНК-лигазу T7, которая катализирует лигирование донорной нуклеиновой кислоты с фосфорильной группой на 3'-конце с акцепторной нуклеиновой кислотой с гидроксильной группой на 3'-конце путем образования 3'→5' фосфодиэфирной связи с одновременным гидролизом ATP в AMP и PP_i. РНК-лигазы могут использовать динуклеозидпирофосфаты в качестве субстратов для создания хвоста из мононуклеотидов, а также могут использовать ди-, три-, п-нуклеотиды для создания хвоста, состоящего из ди-, три-, п-нуклеотидов.

Термин "иммобилизация" используют в настоящем описании по отношению к любому способу, позволяющему зафиксировать нуклеиновую кислоту на поверхности. Иммунобилизированная на поверхности ДНК необходима для разработки ДНК-чипов и -матриц, ДНК-сенсоров или других распознающих устройств, включая микрофлюидные устройства, дополнительно к устройствам доставки генов. Широкий диапазон применения всех данных систем на основе ДНК в основном принадлежал к области медицины,

хотя данные устройства также применялись для секвенирования ДНК и, кроме того, для анализа продуктов питания, анализа окружающей среды или криминалистического анализа. В зависимости от различных используемых поверхностей были разработаны и оптимизированы различные методики иммобилизации (например, посредством физической адсорбции, ковалентного, аффинного связывания и захвата матрицей), которые описаны для углеродистых материалов (например, нанотрубок углерода), кремниевых и силиконовых поверхностей, поверхностей из золота, а также для недавно разработанных сложных биосовместимых поверхностей (например, полимерных гелей).

"Полимеразная цепная реакция" (ПЦР) представляет собой биохимическую методику в молекулярной биологии, используемую для амплификации одной или нескольких копий кусочка ДНК в несколько порядков величины, позволяя получить от тысяч до миллионов копий определенной последовательности ДНК. Почти во всех применениях ПЦР используют термоустойчивую ДНК-полимеразу, такую как Taq-полимераза (фермент, исходно выделенный из бактерии *Thermus aquaticus*). Данная ДНК-полимераза в ходе ферментативной реакции собирает новую цепь ДНК из структурных элементов ДНК - нуклеотидов, используя одноцепочечную ДНК в качестве матрицы и ДНК-олигонуклеотиды (также называемые ДНК-праймерами), которые необходимы для запуска синтеза ДНК. В большинстве способов ПЦР используют циклическое нагревание, т.е. попеременное нагревание и охлаждение ПЦР-образца с помощью определенной серии температурных режимов. Для проведения обычной ПЦР требуется несколько компонентов и реагентов. Данные компоненты включают ДНК-матрицу, которая содержит участок ДНК (мишень), который нужно амплифицировать, два праймера, которые комплементарны 3'-концам каждой из смысловой и антисмысловой цепи ДНК-мишени, Taq-полимеразу или другую ДНК-полимеразу с оптимальной активностью при температуре около 70°C, дезоксинуклеозидтрифосфаты (dNTP), структурные элементы, из которых ДНК-полимераза синтезирует новую цепь ДНК, буферный раствор, обеспечивающий подходящую химическую среду для оптимальной активности и стабильности ДНК-полимеразы, двухвалентные катионы - ионы магния или марганца; как правило используют Mg^{2+} , но Mn^{2+} можно применять для ПЦР-опосредованного мутагенеза ДНК, так как более высокая концентрация Mn^{2+} повышает количество ошибок в процессе синтеза ДНК, или одновалентные катионы - ионы калия. Описанный выше способ может включать мечение нуклеиновой кислоты. Специалисту известен целый ряд методик, позволяющих пометить ДНК, РНК или олигонуклеотиды. Такие методики включают, например, мечение путем никт-трансляции, мечение ДНК со случайными праймерами, мечение ДНК-зондов с помощью ПЦР и мечение олигонуклеотидов по 3'/5'-концу, транскрипционное мечение РНК-зондов, мечение олигонуклеотидов по 3'/5'-концу и добавление хвоста к олигонуклеотиду ПЦР можно использовать в некоторых предпочтительных вариантах реализации способа согласно настоящему изобретению, предпочтительно после синтеза двухцепочечной нуклеиновой кислоты.

Термин "определение последовательности", используемый в настоящем описании, относится к различным способам определения точного порядка нуклеотидов в молекуле ДНК или РНК, другими словами, определения порядка четырех оснований - аденина, гуанина, цитозина и тимина в цепи ДНК или урацила вместо тимина в случае РНК. Секвенирование ДНК можно применять для определения последовательности отдельных генов, генетических участков большего размера (т.е. кластеров генов или оперонов), целых хромосом или полных геномов. Секвенирование позволяет установить порядок отдельных нуклеотидов в ДНК или РНК, выделенной из клеток животных, растений, бактерий, архей или практически любого другого источника генетической информации.

Термин "чип" используют в настоящем описании по отношению к микрочипу с нуклеиновыми кислотами (который также принято называть ДНК-чипом или биочипом, если ДНК иммобилизована), и он представляет собой упорядоченно расположенные на твердой поверхности области, каждая из которых содержит одинаковые или различные нуклеиновые кислоты. Предпочтительно каждая область содержит только идентичные молекулы нуклеиновых кислот. Области могут принимать любую форму, предпочтительно круглую или квадратную. Такие микрочипы используют для одновременного измерения уровней экспрессии большого количества генов или для генотипирования множества участков генома. Каждая область обычно содержит пикомоли (10-12 пмоль) ДНК с определенной последовательностью, известной как зонды (или репортеры, или олигонуклеотиды). Они, возможно, представляют собой короткий участок гена или другого элемента ДНК, который используют для гибридизации с образцом (называемым мишенью) кДНК или кРНК (также называемой антисмысловой РНК) при условиях высокой строгости. Гибридизацию зонд-мишень обычно детектируют и осуществляют ее количественный анализ путем обнаружения меченых флуорофором, серебром или хемилюминесцентной меткой мишеней, чтобы определить относительное содержание последовательностей нуклеиновых кислот в мишени. Указанные зонды синтезируют, а затем присоединяют путем модифицирования поверхности к твердой поверхности с помощью ковалентной связи с химической матрицей (с помощью эпокисилана, аминсилана, лизина, полиакриламида или других соединений). Указанная твердая поверхность может представлять собой стекло или кремниевый чип. ДНК-микрочипы можно применять для измерения изменения уровней экспрессии, для обнаружения однонуклеотидных полиморфизмов (SNP) или для генотипирования или направленного повторного секвенирования.

Варианты реализации изобретения

В следующих параграфах различные аспекты настоящего изобретения описаны более подробно. Каждый аспект, описанный таким образом, можно комбинировать с любым другим аспектом или аспектами, если явно не указано противоположное. В частности, любое свойство, указанное как предпочтительное или дающее преимущество, можно комбинировать с любым другим свойством или свойствами, указанными как предпочтительные или дающие преимущество.

В работе, которая привела к настоящему изобретению, неожиданно было показано, что из одноцепочечных нуклеиновых кислот можно синтезировать двухцепочечную нуклеиновую кислоту с определенными 3'- и 5'-концами быстрым способом, при этом полученная двухцепочечная нуклеиновая кислота будет готова к проведению секвенирования с помощью современных методик секвенирования нового поколения без каких-либо дополнительных этапов согласно способу, описанному в настоящем изобретении.

На основании данных результатов в первом аспекте настоящего изобретения предложен способ синтеза двухцепочечной нуклеиновой кислоты с определенной 3'- и 5'-концевой последовательностью нуклеотидов из образца, содержащего одноцепочечную нуклеиновую кислоту, включающий следующие этапы:

- a) обеспечение образца, содержащего одноцепочечную или двухцепочечную нуклеиновую кислоту, возможно денатурирование указанной двухцепочечной нуклеиновой кислоты;
- b) добавление по меньшей мере 5 последовательных нуклеотидов к 3'-концу указанной одноцепочечной или двухцепочечной нуклеиновой кислоты;
- c) гибридизация иницирующего олигонуклеотида, комплементарного к добавленной последовательности нуклеотидов, и синтез кДНК или кРНК с помощью зависимой от матрицы ДНК- или РНК-полимеразы с получением двухцепочечной нуклеиновой кислоты;
- d) гибридизация олигонуклеотида, переключающего матрицу, с указанной двухцепочечной нуклеиновой кислотой; и
- e) удлинение 3'-конца цепи кДНК или кРНК для синтеза двухцепочечной нуклеиновой кислоты, при этом одна цепь указанной нуклеиновой кислоты содержит иницирующий олигонуклеотид и кДНК или кРНК, которая комплементарна указанной одноцепочечной нуклеиновой кислоте и олигонуклеотиду переключающему матрицу.

Одна из целей способа согласно настоящему изобретению состояла в добавлении известной последовательности нуклеотидов, также называемой в контексте настоящего изобретения определенной последовательностью, к обоим 3'- и 5'-концам в одно- или двухцепочечной нуклеиновой кислоте с неизвестной последовательностью. Данные добавленные последовательности нуклеотидов обеспечивают возможность специфичного отжига олигонуклеотидов с идентичной и/или комплементарной последовательностью с двухцепочечной нуклеиновой кислотой, которая представляет собой продукт способа согласно настоящему изобретению и, следовательно, множество последующих манипуляций над двухцепочечной нуклеиновой кислотой, включая захват, амплификацию, удлинение и т.д. Предпочтительно, каждая из 3'-концевой и 5'-концевой "определенных последовательностей" не гибридизуется с другой, а также маловероятно будет гибридизоваться с каким-либо из нуклеотидов, присутствующих в образце, при условиях, выбранных для последующих манипуляций над двухцепочечной нуклеиновой кислотой, которая представляет собой продукт способа согласно настоящему изобретению.

В предпочтительном варианте реализации первого аспекта настоящего изобретения образец получен из жидкого или твердого биоптата или произведен из него, более предпочтительно представляет собой образец крови, образец плазмы, образец сыворотки, образец биологической жидкости, образец слюны, образец мочи, образец семенной жидкости, образец жидкости из плевральной полости, образец жидкости из перитонеальной полости, образец спинномозговой жидкости, мазок с поверхности эпителия, образец мокроты, образец кала, образец эякулята, образец слезы, образец пота, образец лимфатической жидкости, образец бронхиального лаважа, образец плеврального выпота, образец менингеальной жидкости, образец секрета железы, образец тонкоигольного пунктата, микропрепарированные клетки, жидкий образец аспирата наконечником, образец спинномозговой жидкости, образец конъюнктивальной жидкости, образец вагинальной жидкости, образец дуоденальной жидкости, образец сока поджелудочной железы или образец желчи. В дополнительном предпочтительном варианте реализации образец представляет собой криминалистический образец или археологический образец. Более предпочтительно образец получен из ископаемых остатков, останков вымерших организмов, растений, плодов и животных, микробов, бактерий, вирусов. В другом более предпочтительном варианте реализации образец получен из млекопитающего, более предпочтительно из человека. В дополнительном предпочтительном варианте реализации образец получен из человека с некоторым расстройством. Более предпочтительно образец включает венозную кровь человека, еще более предпочтительно плазму человека. В другом предпочтительном варианте реализации образец, содержащий одноцепочечную или двухцепочечную нуклеиновую кислоту, предпочтительно кровь человека, образец сыворотки или образец плазмы крови, напрямую подвергают способу согласно настоящему изобретению без предшествующего этапа выделения нуклеиновой кислоты из образца, взятого из пациента. Это предпочтительный вариант реализации, если одноцепочечная или двухцепочечная нуклеиновая кислота представляет собой ДНК. Более предпочтительно в случае

когда образец напрямую подвергают способу согласно настоящему изобретению, указанный образец обрабатывают ферментом, способным расщеплять пептидные связи в белках, предпочтительно протеазой, в частности протеиназой K, и инкубируют при подходящей температуре в течение подходящего времени. Предпочтительно, чтобы указанный образец был получен с помощью способа, который не несет существенного риска для здоровья пациента, например путем взятия крови из периферической вены или артерии. Образец, используемый на этапе а), может содержать одноцепочечные и/или двухцепочечные нуклеиновые кислоты. Если указанный образец содержит двухцепочечную ДНК, то предпочтительно, чтобы перед этапом а) осуществляли этап денатурации. Такой этап может включать тепловую или химическую денатурацию.

В предпочтительном варианте реализации первого аспекта настоящего изобретения одно- или двухцепочечная нуклеиновая кислота представляет собой ДНК или РНК. Указанная ДНК или РНК может представлять собой фрагментированную или подвергнутую бисульфитной конверсии РНК или ДНК. В более предпочтительном варианте реализации средняя длина РНК или ДНК, содержащейся в образце, составляет 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, 43, 44, 45, 46, 47, 48, 49, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590 или 600 нуклеотидов. Более предпочтительно указанная одноцепочечная нуклеиновая кислота представляет собой РНК, еще более предпочтительно указанная одноцепочечная нуклеиновая кислота представляет собой миРНК, малую РНК или рiРНК. В дополнительном предпочтительном варианте реализации указанная РНК в естественном виде не содержит непрерывный участок полиаденинов, предпочтительно состоящий по меньшей мере из 30 полиаденинов. В дополнительном предпочтительном варианте реализации одноцепочечная нуклеиновая кислота представляет собой ДНК.

Благодаря чувствительности способа согласно настоящему изобретению количество одно- или двухцепочечной нуклеиновой кислоты, которое необходимо обеспечить на этапе а), может быть очень низким, и все-таки сможет позволить получить двухцепочечные нуклеиновые кислоты. Таким образом, в предпочтительном варианте реализации образец, обеспеченный на этапе а), содержит ДНК и/или РНК в концентрации, меньшей чем 1 мкг/мкл, предпочтительно меньшей чем 0,1 мкг/мкл, более предпочтительно меньшей чем 0,01 мкг/мкл, еще более предпочтительно меньшей чем 1 нг/мкл, еще более предпочтительно меньшей чем 0,1 нг/мкл, еще более предпочтительно меньшей чем 0,01 нг/мкл, еще более предпочтительно меньшей чем 1 пг/мкл, еще более предпочтительно меньшей чем 0,1 пг/мкл, еще более предпочтительно меньшей чем 0,01 пг/мкл, наиболее предпочтительно меньшей чем 1 фг/мкл. Общее содержание ДНК и/или РНК в образце также может быть очень низким и предпочтительно составляет 5 пг. Предпочтительно 5 пг/мкл можно использовать, если нуклеиновая кислота, обеспеченная на этапе а) первого аспекта настоящего изобретения, представляет собой малую РНК, 5 пг можно использовать, если нуклеиновая кислота, обеспеченная на указанном этапе а), представляет собой ДНК, или можно использовать диапазон от 1 пг/мкл до 5 нг/мкл, если нуклеиновая кислота, обеспеченная на указанном этапе а), представляет собой микроРНК или миРНК.

На этапе b) указанного способа требуется добавление по меньшей мере 5 последовательных нуклеотидов к 3'-концу одноцепочечной нуклеиновой кислоты. Данный участок последовательных нуклеотидов служит для цели последующей гибридизации иницирующего олигонуклеотида. Он может служить в качестве 3'-концевой определенной последовательности, которую внедряют в способе согласно настоящему изобретению. Соответственно указанный иницирующий олигонуклеотид и указанные последовательные нуклеотиды должны содержать последовательности, которые комплементарны друг другу. Цель может быть достигнута, если добавлены последовательные нуклеотиды с известной последовательностью, например путем добавления праймера с известной последовательностью или путем добавления последовательного участка из известных моно- или динуклеотидов. Не обязательно, чтобы данный участок нуклеотидов был добавлен непосредственно с 3'-стороны от одноцепочечной нуклеиновой кислоты, если он содержится в добавленном последовательном участке нуклеотидов. Другой предпочтительный вариант реализации первого аспекта настоящего изобретения включает добавление идентичных последовательных нуклеотидов к 3'-концу одноцепочечной нуклеиновой кислоты. Предпочтительно добавляют идентичные последовательные нуклеотиды, выбранные из группы, состоящей из А, Т, G, С или U. Предпочтительно количество идентичных последовательных нуклеотидов находится в диапазоне от 10 до 500 последовательных нуклеотидов, т.е. равно 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490 или 500. Более предпочтительно количество последовательных идентичных нуклеотидов находится в диапазоне от 10 до 100 последовательных идентичных нуклеотидов, более предпочтительно от 15 до 50 последовательных идентичных нуклеотидов, более предпочтительно от 20 до 40 последовательных идентичных нуклеотидов или от 30 до 100 последовательных идентичных нуклеотидов, т.е. равно 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90 или 100. Следовательно, предпочтительно более короткие участки последовательных нуклеотидов. Ограниченный выступ на 3'-конце приводит к (1) пропор-

ционально более высокой способности иницирующего олигонуклеотида, добавленного на этапе с) способа согласно настоящему изобретению, инициировать обратную транскрипцию при тех же концентрациях; (2) возможности точного расчета оптимальных количеств иницирующего олигонуклеотида, добавленного на этапе с) способа согласно настоящему изобретению, что приводит к более низкой частоте встречаемости "пустых" побочных ДНК-продуктов, полученных, когда иницирующий олигонуклеотид взаимодействует непосредственно с ОПМ; (3) более низкой частоте встречаемости ДНК-продуктов, содержащих полинуклеотидный участок длиной более 30 нуклеотидов, вследствие инициации обратной транскрипции с сайтов, отдаленных от 3'-конца матрицы. Преимущества (1), (2) и (3) приводят к статистически значимому повышению чувствительности способа и позволяют синтезировать ДНК по более низким концентрациям матриц. Дополнительно при получении библиотек более короткие участки последовательных нуклеотидов обеспечивают дополнительное преимущество получения лучшей, например более сложной, библиотеки.

В другом предпочтительном варианте реализации указанные идентичные последовательные нуклеотиды включают последовательные динуклеотиды, выбранные из группы, состоящей из AC, AG, AT, AU, CA, CG, CT, CU, GA, GC, GT, GU, TA, TG, TC или TU. В другом предпочтительном варианте реализации добавляют идентичные последовательные три-, quadro- или пентануклеотиды. Для того чтобы добавить только один тип нуклеотидов, предпочтительно, чтобы нуклеотиды, которые добавляют на этапе b) реакции, включали, по существу включали только или состояли только из одного нуклеотидного структурного элемента конкретного типа, например только из A, G, C или T. Тем не менее, можно предвидеть, что указанные нуклеотидные структурные элементы, используемые на этапе b), не полностью гомогенны, но также содержат другие нуклеотидные структурные элементы. В данном случае различные нуклеотидные структурные элементы будут добавляться случайным образом, в зависимости от соответствующей их концентрации в реакционной смеси. Следовательно, в одном варианте реализации способа согласно настоящему изобретению желательно, чтобы концентрация других нуклеотидов была минимальной, чтобы гарантировать образование непрерывного участка с предполагаемой последовательностью нуклеотидов. Тем не менее, способ согласно настоящему изобретению не исключает варианты реализации, в которых используют смеси структурных элементов, при условии, что большинство добавленных нуклеотидов содержит по меньшей мере 10 последовательных нуклеотидов с известной последовательностью.

Существуют различные способы ограничения количества нуклеотидов, добавленных в данной реакции добавления хвоста, известные специалисту в данной области техники. В одном предпочтительном варианте реализации применяют субоптимальные концентрации добавляемого нуклеотида или динуклеотида. Субоптимальная концентрация представляет собой молярность нуклеотида или динуклеотида, которая ниже, чем молярность нуклеотида или динуклеотида, рекомендуемая поставщиком/производителем независимых от матрицы ДНК- и РНК-полимераз; и при которой независимые от матрицы ДНК- и РНК-полимеразы синтезируют полинуклеотидные хвосты короче чем 1000 нт. В случаях когда независимые от матрицы ДНК- и РНК-полимеразы используют для реакции добавления хвоста, специалист в данной области техники может определить для соответствующего фермента концентрацию нуклеотидов или динуклеотидов в соответствующей реакционной смеси, которая приведет к максимальному количеству добавленных нуклеотидов или динуклеотидов за данный промежуток времени, т.е. концентрацию максимальной процессивности фермента. Данную концентрацию потом считают оптимальной концентрацией для данного фермента при данных условиях реакции (например, буфере, pH, температуре и т.д.). "Субоптимальная концентрация" нуклеотидов/динуклеотидов представляет собой концентрацию, которая по меньшей мере в 10 раз ниже, чем оптимальная концентрация нуклеотидов, более предпочтительно в 100 раз ниже. Предпочтительно, чтобы субоптимальная концентрация приводила к уменьшению процессивности фермента, т.е. чтобы количество нуклеотидов/динуклеотидов, добавленных в данный период времени, было по меньшей мере в 10 раз ниже, чем количество нуклеотидов/динуклеотидов, добавленных для оптимальной концентрации нуклеотидов, более предпочтительно в 100 раз ниже. Предпочтительно субоптимальная концентрация находится в диапазоне 0,1-0,01 мМ АТФ в течение 10-20 мин реакции, опосредованной поли(А)-полимеразой *E.coli*, в реакционном буфере для поли(А)-полимеразы (50 мМ Трис-НСl, 250 мМ NaCl, 10 мМ MgCl₂, pH 7,9 при 25°C); 0,001-0,0001 мМ АТФ в течение 10-20 мин реакции, опосредованной поли(А)-полимеразой *E.coli*, в реакционном буфере для обратной транскриптазы MMLV (50 мМ Трис-НСl, 75 мМ KCl, 3 мМ MgCl₂, 10 мМ ДТТ, pH 8,3 при 25°C); 0,1-0,01 мМ АТФ в течение 10-20 мин реакции, опосредованной поли(А)-полимеразой дрожжей, в реакционном буфере для обратной транскриптазы MMLV (50 мМ Трис-НСl, 75 мМ KCl, 3 мМ MgCl₂, 10 мМ ДТТ, pH 8,3 при 25°C); 0,1-0,01 мМ dATP в течение 10-30 мин реакции, опосредованной концевой трансферазой, либо в реакционном буфере для концевой трансферазы (50 мМ ацетат калия, 20 мМ Трис-ацетат, 10 мМ ацетат магния, pH 7,9 при 25°C), либо в реакционном буфере для обратной транскриптазы MMLV (50 мМ Трис-НСl, 75 мМ KCl, 3 мМ MgCl₂, 10 мМ ДТТ, pH 8,3 при 25°C). Оптимальная концентрация представляет собой молярность нуклеотида или динуклеотида, при которой независимые от матрицы ДНК- и РНК-полимеразы добавляют к матрицам ДНК и РНК полинуклеотидные хвосты длиной по

меньшей мере 30 н.т., но не более чем 1000 н.т. В случаях когда независимые от матрицы ДНК- и РНК-полимеразы используют для реакции добавления хвоста, специалист в данной области техники может определить для соответствующего фермента концентрацию нуклеотидов или динуклеотидов в соответствующей реакционной смеси, которая приведет к оптимальному количеству добавленных нуклеотидов или динуклеотидов за данное время. Данную концентрацию тогда считают оптимальной концентрацией для данного фермента при данных условиях реакции (например, буфере, pH, температуре и т.д.). "Субоптимальная концентрация" нуклеотидов/динуклеотидов представляет собой концентрацию, которая по меньшей мере в 10 раз выше, чем оптимальная концентрация нуклеотидов, более предпочтительно в 100 раз выше. Предпочтительно, чтобы оптимальная концентрация приводила к уменьшению процессивности фермента, т.е. чтобы количество нуклеотидов/динуклеотидов, добавленных в данный период времени, было по меньшей мере в 10 раз ниже, чем количество нуклеотидов/динуклеотидов, добавленных для субоптимальной концентрации нуклеотидов, более предпочтительно в 100 раз ниже. Предпочтительно оптимальная концентрация находится в диапазоне 0,1-0,01 mM АТФ в течение 10-20 мин реакции, опосредованной поли(А)-полимеразой *E.coli*, в реакционном буфере для поли(А)-полимеразы (50 mM Трис-HCl, 250 mM NaCl, 10 mM MgCl₂, pH 7,9 при 25°C); 0,001-0,0001 mM АТФ в течение 10-20 мин реакции, опосредованной поли(А)-полимеразой *E.coli*, в реакционном буфере для обратной транскриптазы MMLV (50 mM Трис-HCl, 75 mM KCl, 3 mM MgCl₂, 10 mM ДТТ, pH 8,3 при 25°C); 0,1-0,01 mM АТФ в течение 10-20 мин реакции, опосредованной поли(А)-полимеразой дрожжей, в реакционном буфере для обратной транскриптазы MMLV (50 mM Трис-HCl, 75 mM KCl, 3 mM MgCl₂, 10 mM ДТТ, pH 8,3 при 25°C); 0,1-0,01 mM dATP в течение 10-30 мин реакции, опосредованной концевой трансферазой, либо в реакционном буфере для концевой трансферазы (50 mM ацетат калия, 20 mM Трис-ацетат, 10 mM ацетат магния, pH 7,9 при 25°C), либо в реакционном буфере для обратной транскриптазы MMLV (50 mM Трис-HCl, 75 mM KCl, 3 mM MgCl₂, 10 mM ДТТ, pH 8,3 при 25°C).

В другом предпочтительном варианте реализации этого добиваются путем применения блокирующих нуклеотидов или динуклеотидов. Блокирующие нуклеотиды или динуклеотиды представляют собой нуклеотиды или динуклеотиды, которые предотвращают добавление дополнительных нуклеотидов или динуклеотидов после завершения их добавления. Обычно олигонуклеотиды удлиняют путем добавления следующего нуклеотида к гидроксильной группе, расположенной в 3'-положении рибозы или дезоксирибозы. Если 3'-положение рибозы или дезоксирибозы заблокировано, то нельзя добавить дополнительные нуклеотиды или динуклеотиды. Таким образом, рибоза или дезоксирибоза блокирующего нуклеотида или 3'-концевого нуклеотида в динуклеотиде не позволяет добавление дополнительного нуклеотида или динуклеотида. Предпочтительные блокирующие нуклеотиды представляют собой 3d-АТФ, 3-Ме-АТФ и ddАТФ. Более предпочтительно используют ddАТФ или 3d-АТФ. Если используют смесь блокирующих нуклеотидов и неблокирующих нуклеотидов, то включение первого блокирующего нуклеотида в растущую олигонуклеотидную цепь является случайным событием, и вероятность включения первого блокирующего нуклеотида после включения данного количества неблокирующих нуклеотидов будет зависеть от отношения блокирующих и неблокирующих нуклеотидов, присутствующих в реакционной смеси. Соответственно концентрация данных блокирующих нуклеотидов или динуклеотидов в реакционной смеси ниже, чем концентрация неблокирующих нуклеотидов или динуклеотидов. Чем меньше относительное количество блокирующего нуклеотида или динуклеотида, тем дольше будет происходить удлинение. Поскольку включение первого блокирующего олигонуклеотида является случайным событием, то длина олигонуклеотида, добавленного в ходе реакции добавления хвоста, будет изменяться в данном диапазоне. Предпочтительно отношение концентраций блокирующих нуклеотидов или динуклеотидов к неблокирующим находится между от 1 до 1 и от 1 до 1000. Обычно используемые концентрации находятся в диапазоне от 0,1 до 0,001 mM, т.е. составляют 0,09, 0,08, 0,07, 0,06, 0,05, 0,04, 0,03, 0,02, 0,01, 0,009, 0,008, 0,007, 0,006, 0,005, 0,004, 0,003, 0,002, 0,001 nM. В наиболее предпочтительном случае 3d-АТФ используют при отношении от 1 до 30 к концентрации АТФ для поли(А)-полимеразы дрожжей и при отношении от 1 до 1,7 к концентрации АТФ для поли(А)-полимеразы *E.coli* с получением продуктов удлинения со средней длиной 30 н.т. В наиболее предпочтительном случае используют ddАТФ при отношении от 1 до 30 к концентрации dATP для концевой трансферазы с получением продуктов удлинения со средней длиной 30 н.т. Предпочтительно выбирать условия таким образом, чтобы в среднем добавлялось не более 50 нуклеотидов, предпочтительно не более 40, более предпочтительно не более 35, более предпочтительно не более 30, предпочтительно не более 25, наиболее предпочтительно не более 20 нуклеотидов.

В таких вариантах реализации, в которых желателен короткий участок идентичных последовательных нуклеотидов, предпочтительно, чтобы этого добивались, предоставляя смесь либо рибонуклеотидов, либо дезоксирибонуклеотидов и завершающего цепь нуклеотида. Если желательна длина, равная 10 последовательным нуклеотидам, то смесь 1 к 10 завершающего цепь нуклеотида и рибо- или дезоксирибонуклеотидов приведет в среднем к такой длине. Специалист, следовательно, знает, как получить последовательные участки нуклеотидов, длина которых в среднем такая, как описано выше, и предпочтительно находится в диапазоне от 30 до 100 нуклеотидов. Предпочтительные завершающие цепь нуклеотиды

представляют собой дидезоксинуклеотиды.

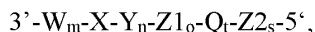
В другом предпочтительном варианте реализации указанные последовательные идентичные нуклеотиды включают смесь из по меньшей мере двух рибонуклеотидов или дезоксирибонуклеотидов.

Включение некоторых количеств одного дополнительного нуклеотида в реакцию добавления полинуклеотидного хвоста может оказать полезный эффект вследствие того, что полинуклеотидный хвост более не будет гомогенным, но при этом сохранит сходную эффективность связывания с иницирующим олигонуклеотидом. В то же время негетерогенные полинуклеотидные хвосты могут быть полезны для секвенирования спаренных концов с применением платформы Illumina, поскольку секвенирование гомонуклеотидов часто приводит к ошибкам. В предпочтительном примере желательно получить смесь, содержащую негетерогенные нуклеотиды, так как полинуклеотидный хвост, возникший в результате добавления данной смеси к одно- или двухцепочечной нуклеиновой кислоте, обеспеченной на этапе а) способа согласно настоящему изобретению, полезен (после завершения этапа е) и использования полученной нуклеиновой кислоты в способах определения ее последовательности) для секвенирования спаренных концов с применением, например, платформы Illumina, поскольку гомополинуклеотиды могут участвовать в нежелательной интерференции.

Предпочтительно А, Т, G или С используют в контексте одноцепочечной ДНК и U или А используют в контексте одноцепочечной РНК.

В другом варианте реализации первого аспекта настоящего изобретения добавление идентичных нуклеотидов на этапе b) осуществляют независимые от матрицы ДНК- и РНК-полимеразы. Предпочтительно данные белки представляют собой концевые трансферазы, ДНК или РНК-лигазы или поли(N)-полимеразы, где N выбран из А, G или U. Фермент, обладающий активностью концевой трансферазы, способен добавлять рибо- или дезоксирибонуклеотиды или их мультимеры к 3'-ОН-концу нуклеиновой кислоты без необходимости наличия комплементарной матричной цепи. Предпочтительные ферменты с данной активностью выбирают из группы, состоящей из концевой дезоксинуклеотидил трансферазы, поли(А)-полимеразы, поли(У)-полимеразы и поли(Г)-полимеразы. РНК-лигазы или ДНК-лигазы могут добавлять мононуклеотиды, динуклеотиды, тринуклеотиды или олигонуклеотиды, предпочтительно они добавляют мононуклеотиды, динуклеотиды, тринуклеотиды. Предпочтительные лигазы представляют собой Т4 РНК-лигазу или Т7 РНК-лигазу. Указанные лигазы могут эффективно добавлять хвост к матрице РНК, которая содержит 2'-О-метил на 3'-концевом нуклеотиде. Предпочтительно, чтобы РНК-лигаза использовала пиррофосфат динуклеотида в качестве субстрата для добавления мононуклеотидов.

Этап с) включает гибридизацию иницирующего олигонуклеотида с ранее добавленной последовательностью нуклеотидов. Данный этап предпочтительно включает повышение температуры, позволяющее образование пары оснований между иницирующим олигонуклеотидом и добавляемыми последовательными нуклеотидами. Дополнительно к последовательности, способной гибридизоваться с добавляемыми последовательными нуклеотидами, иницирующий олигонуклеотид включает дополнительную определенную последовательность, предпочтительно 5'-концевую, которую можно использовать для специфической гибридизации с другим олигонуклеотидом, например с олигонуклеотидом для ПЦР-амплификации. Длина этой части предпочтительно составляет от 5 до 100 нуклеотидов. Предпочтительно эта часть дополнительно содержит так называемую структуру "крючка" на 3'-конце. Крючок предпочтительно представляет собой нуклеотид, который отличен от нуклеотидов, способных гибридизоваться с добавляемыми последовательными нуклеотидами, и служит для позиционирования иницирующего олигонуклеотида непосредственно на 5'-конце последовательных нуклеотидов, добавленных на этапе b), или близко к нему. Предпочтительно иницирующий олигонуклеотид, применяемый в способе согласно настоящему изобретению, содержит следующие элементы последовательности:



где

W в каждом случае независимо выбран из dA, dG, dC, dT и dU;

X выбран из dA, dG, dC, dT, dU, rA, rG, rC, rT и rU;

Y представляет собой полинуклеотид длиной по меньшей мере 10 нуклеотидов, который на 80% последовательности или более состоит из идентичных нуклеотидов или динуклеотидов, выбранных из dA, dG, dC, dT, dU, rA, rG, rC, rT, rU, AC, AG, AT, AU, CA, CG, CT, CU, GA, GC, GT, GU, TA, TC, TG, TU, AA, CC, GG, TT, UU, UA, UC, UG и UT, и который на другие 20% последовательности или менее состоит из нуклеотидов или динуклеотидов, которые отличны от основного нуклеотида или динуклеотида и также выбраны из dA, dG, dC, dT, dU, rA, rG, rC, rT, rU, AC, AG, AT, AU, CA, CG, CT, CU, GA, GC, GT, GU, TA, TC, TG, TU, AA, CC, GG, TT, UU, UA, UC, UG и/или UT, при условии, что X отличен от нуклеотида или динуклеотида, который составляет большую часть Y;

Q представляет собой последовательность идущих друг за другом вырожденных (неоднозначных) оснований ДНК, предпочтительно выбранных из N, V, H, D, B и J, где N представляет собой продукт включения нуклеотида из эквимоларной смеси dA, dT, dC и dG; V представляет собой продукт включения нуклеотида из эквимоларной смеси dT, dC и dG; D представляет собой продукт включения нуклеотида из эквимоларной смеси dA, dT и dG; H представляет собой продукт включения нуклеотида из экви-

молярной смеси dA, dT и dC; V представляет собой продукт включения нуклеотида из эквимольной смеси dA, dC и dG; J представляет собой продукт включения нуклеотида из смеси (0-100% dA) с (0-100% dG), с (0-100% dC), с (0-100% dT), с (0-100% dU), с (0-100% rA), с (0-100% rG), с (0-100% rC), с (0-100% rT), с (0-100% rU);

Z1 представляет собой полинуклеотид длиной по меньшей мере 5 нуклеотидов с определенной последовательностью, при этом указанная последовательность отличается от W_m-X-Y_n , предпочтительно указанная последовательность также отличается от Q_t-Z2_s ;

Z2 представляет собой полинуклеотид длиной по меньшей мере 5 нуклеотидов с определенной последовательностью, при этом указанная последовательность отличается от $W_m-X-Y_n-Z1_o-Q_t$;

m представляет собой целое число от 0 до 6, т.е. 0, 1, 2, 3, 4, 5 или 6;

n представляет собой целое число от 10 до 100, если Y выбран из dA, dG, dC, dT, dU, rA, rG, rC, rT и rU, целое число от 5 до 50, если Y выбран из AC, AG, AT, AU, CA, CG, CT, CU, GA, GC, GT, GU, TA, TC, TG, TU, AA, CC, GG, TT, UU, UA, UC, UG и UT;

o представляет собой 0 или 1;

s представляет собой 0 или 1; и

t представляет собой целое число от 0 до 6, т.е. 0, 1, 2, 3, 4, 5 или 6.

Y представляет собой часть иницирующего олигонуклеотида, которая способна гибридизоваться с присоединенными последовательными нуклеиновыми кислотами. Таким образом, предпочтительно, чтобы комплементарность последовательности Y с добавленными нуклеиновыми кислотами составляла по меньшей мере 90 %. Следовательно, ее длина предпочтительно соответствует длине добавленных последовательных нуклеотидов, более предпочтительно длине от 10 до 100 нуклеотидов, т.е. 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 или 100 нуклеотидов. Авторы настоящего изобретения обнаружили, что короткий Y улучшает достоверность секвенирования. Тем не менее, для того чтобы позволить гибридизацию, предпочтительно при строгих условиях, предпочтительно, чтобы длина Y составляла от 11 до 50, более предпочтительно от 12 до 40, более предпочтительно от 13 до 30 и наиболее предпочтительно от 14 до 20 нуклеотидов.

Авторы настоящего изобретения обнаружили, что присутствие небольшого количества неидентичных нуклеотидов и/или динуклеотидов улучшает достоверность секвенирования. Таким образом, предпочтительно, чтобы последовательность Y состояла из по меньшей мере 80% идентичных нуклеотидов и/или динуклеотидов, выбранных из dA, dG, dC, dT, dU, rA, rG, rC, rT, rU, AC, AG, AT, AU, CA, CG, CT, CU, GA, GC, GT, GU, TA, TC, TG, TU, AA, CC, GG, TT, UU, UA, UC, UG и UT, при этом другие 20% последовательности или менее состоят из нуклеотидов или динуклеотидов, которые отличны от основного нуклеотида и/или динуклеотида и также выбраны из dA, dG, dC, dT, dU, rA, rG, rC, rT, rU, AC, AG, AT, AU, CA, CG, CT, CU, GA, GC, GT, GU, TA, TC, TG, TU, AA, CC, GG, TT, UU, UA, UC, UG и UT. В предпочтительном варианте реализации указанные основные нуклеотиды представляют собой A и/или T. В другом предпочтительном варианте реализации указанные нуклеотиды представляют собой динуклеотиды, предпочтительно AA, TT, AT или TA. В другом предпочтительном варианте реализации второстепенные нуклеотиды представляют собой C и/или G. В другом предпочтительном варианте реализации указанные нуклеотиды представляют собой динуклеотиды, предпочтительно CC, GG, CG и/или GC. В предпочтительном варианте реализации от 80 до 99% последовательности Y состоит из идентичных нуклеотидов и/или динуклеотидов, более предпочтительно от 85 до 95% (для специалиста очевидно, что в данном случае "n" должен быть равен по меньшей мере 20), более предпочтительно от 88 до 92% и наиболее предпочтительно приблизительно 90%. Таким образом, Y в типичном предпочтительном варианте реализации может содержать 9 нуклеотидов T и один нуклеотид G или C или 14 нуклеотидов T и один нуклеотид G или C.

В случаях когда Y содержит один или два различных нуклеотида, предпочтительно, чтобы данный(е) нуклеотид(ы) был(и) расположен(ы) посередине Y или близко к ней (т.е. в пределах от 1 до 4 оснований от нее).

В другом предпочтительном варианте реализации второго аспекта настоящего изобретения предпочтительно, чтобы Y представлял собой последовательный участок нуклеотидов, состоящий только из T, и n находился в диапазоне от 10 до 60, т.е. 50, 45, 40, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11 или 10, более предпочтительно от 11 до 50, более предпочтительно от 12 до 40, более предпочтительно от 13 до 30 и наиболее предпочтительно от 14 до 20. Более предпочтительно n представляет собой 30, 20, 16 или 15. Наиболее предпочтительно, чтобы n представлял собой 20 или 16. В качестве альтернативы данному предпочтительному варианту реализации Y содержит один или два отличных нуклеотида, предпочтительно G или C.

В альтернативном предпочтительном варианте реализации последовательность Y представляет собой последовательный участок нуклеотидов, состоящий только из T, не считая одного или двух остатков G и/или C.

Z1 представляет собой часть иницирующего олигонуклеотида, которую используют после синтеза двухцепочечной молекулы нуклеиновой кислоты, чтобы позволить сайт-специфическую гибридизацию с другим олигонуклеотидом. Таким образом, Z1 предпочтительно представляет собой определенную по-

следовательность, добавляемую к 3'-концу нуклеиновой кислоты, содержащейся в образце. Длина Z1 составляет по меньшей мере 5 нуклеотидов, более предпочтительно находится в диапазоне от 5 до 50 нуклеотидов, более предпочтительно в диапазоне от 10 до 30 нуклеотидов. Длину выбирают таким образом, чтобы праймер мог специфично гибридизоваться с Z1 в последующих реакциях ПЦР-амплификации. В предпочтительном варианте реализации последовательность нуклеиновой кислоты Z1 выбирают из группы, состоящей из SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15 и SEQ ID NO: 16.

Предпочтительно Z1 не идентичен Z2.

Z2 представляет собой часть иницирующего олигонуклеотида, которую используют после синтеза двухцепочечной молекулы нуклеиновой кислоты, чтобы позволить сайт-специфическую гибридизацию с другим олигонуклеотидом. Таким образом, Z2 предпочтительно представляет собой определенную последовательность, добавляемую к 3'-концу нуклеиновой кислоты, содержащейся в образце. Длина Z2 составляет по меньшей мере 5 нуклеотидов, более предпочтительно находится в диапазоне от 5 до 50 нуклеотидов, более предпочтительно в диапазоне от 10 до 30 нуклеотидов. Длину выбирают таким образом, чтобы праймер мог специфично гибридизоваться с Z1 в последующих реакциях ПЦР-амплификации. В предпочтительном варианте реализации последовательность нуклеиновой кислоты Z2 выбирают из группы, состоящей из SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15 и SEQ ID NO: 16 или соответствующей последовательности.

Предпочтительно Z2 не идентичен Z1.

Включение от 1 до 6, более предпочтительно от 2 до 4, т.е. 1, 2, 3, 4, 5 или 6 последовательных неоднозначных оснований в праймер, т.е. между Z1 и Z2, позволит удалять ПЦР-дубликаты из библиотеки. Предпочтительно Q представляет собой последовательность идущих друг за другом вырожденных (неоднозначных) оснований ДНК, предпочтительно в каждом случае независимо, выбранных из N, V, H, D, B и J, где N представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT, dC и dG; B представляет собой продукт включения нуклеотида из эквимольной смеси dT, dC и dG; D представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dG; H представляет собой продукт включения нуклеотида из эквимольной смеси dA, dC и dG; J представляет собой продукт включения нуклеотида из смеси (0-100% dA) с (0-100% dG), с (0-100% dC), с (0-100% dT), с (0-100% dU), с (0-100% rA), с (0-100% rG), с (0-100% rC), с (0-100% rT), с (0-100% rU). Включение последовательных неоднозначных оснований в иницирующий олигонуклеотид предпочтительно, так как это позволяет удалять ПЦР-дубликаты из полученной библиотеки ДНК. Наиболее предпочтительно, чтобы Q был расположен между Z1 и Z2 и представлял собой N. Предпочтительно Q представляет собой N и t равен по меньшей мере 2, более предпочтительно равен по меньшей мере 4.

В другом предпочтительном варианте реализации сумма t и s равна 0, например Z2 и Q отсутствуют.

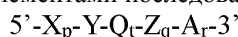
Особенно предпочтительные примеры иницирующего олигонуклеотида представляют собой нуклеотиды с последовательностями, соответствующими SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19 и SEQ ID NO: 20.

Как только иницирующий олигонуклеотид гибридизовался, его 3'-конец удлиняется зависящей от матрицы ДНК- или РНК-полимеразой, предпочтительно ДНК- или РНК-полимеразой, у которой также есть активность концевой трансферазы. Предпочтительные примеры таких ферментов представляют собой обратные транскриптазы (ОТ), в частности ОТ MMLV. После того как достигается конец матрицы, зависящая от матрицы ДНК- или РНК-полимераза использует свою активность концевой трансферазы, чтобы добавить дополнительные нуклеотиды независимым от матрицы способом. Таким образом, продукт этапа с) представляет собой двухцепочечную нуклеиновую кислоту (ДНК/ДНК, РНК/РНК или ДНК/РНК) с выступом на 3'-конце вновь синтезированной цепи. Предпочтительно длина такого выступа составляет по меньшей мере 1 нуклеотид, предпочтительно по меньшей мере 3 нуклеотида. Предпочтительно данные нуклеотиды идентичны. Предпочтительно они выбраны из dA, dC, dG, dT, rA, rC, rG и rU, наиболее предпочтительно представляют собой dC. Таким образом, особенно предпочтительный выступ состоит из непрерывного участка трех нуклеотидов цитозинов.

На этапе d) олигонуклеотид, переключающий матрицу, ОПМ, гибридизуется с продуктом этапа с), который позволяет зависящей от матрицы ДНК- или РНК-полимеразе, предпочтительно обратной транскриптазе, добавлять определенную последовательность к 5'-концу указанной одноцепочечной или двухцепочечной нуклеиновой кислоты, содержащейся в образце. Этого добиваются путем дополнительного удлинения 3'-конца цепи нуклеиновой кислоты, синтезированной на этапе с). Термин "олигонуклеотид, переключающий матрицу", используют по отношению к олигонуклеотидной матрице, на которую переключается активность полимеразы с исходной матрицы (например, с одноцепочечной нуклеиновой кислоты, находящейся в образце согласно настоящему изобретению). В некотором варианте реализации настоящего изобретения олигонуклеотид, переключающий матрицу, представляет собой гибридный оли-

гонуклеотид ДНК/РНК, который используется зависящей от матрицы ДНК- или РНК-полимеразой, предпочтительно ОТ, предпочтительно ОТ MMLV, для продолжения обратной транскрипции после того, как фермент, предпочтительно ОТ MMLV, достигает 5'-конца матрицы нуклеиновой кислоты и добавляет посредством своей активности концевой трансферазы нуклеотиды на 3'-конце синтезированной цепи кДНК или кРНК, т.е. независимо от матрицы. 3'-конец ОПМ гибридизуется с нуклеотидами, добавленными активностью концевой трансферазы зависящей от матрицы ДНК- или РНК-полимеразы, эффективно удлиняя 5'-конец матричной ДНК или РНК и, таким образом, позволяя зависящей от матрицы ДНК- или РНК-полимеразе, предпочтительно ОТ, более предпочтительно ОТ MMLV, осуществлять обратную транскрипцию также остальной 5'-части ОПМ, которая содержит определенную последовательность, которую нужно добавить к 5'-концу матрицы нуклеиновой кислоты. Как описано выше в отношении иницирующего олигонуклеотида, эта определенная последовательность не будет гибридизоваться с иницирующей олигонуклеотидной последовательностью или с комплементарной ей последовательностью и предпочтительно также не будет гибридизоваться с последовательностями, присутствующими в нуклеиновой кислоте, содержащейся в образце. Предпочтительно она не будет гибридизоваться при условиях, которые обычно используются при последующей обработке двухцепочечной нуклеиновой кислоты, которая является продуктом способа согласно настоящему изобретению, в частности при ПЦР или определении последовательности. Специалист в данной области техники хорошо осведомлен о том, как выбрать подходящие последовательности, которые могут служить в качестве определенной последовательности ОПМ. Более того, ОПМ содержит на 3'-конце один или более нуклеотидов, предпочтительно рибонуклеотидов, которые комплементарны нуклеотидам, добавляемым ферментом обратной транскриптазой на этапе с). Предпочтительно ОПМ содержит на 3'-конце от 1 до 10, т.е. 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10, предпочтительно 3 последовательных нуклеотида, предпочтительно рибонуклеотида. Предпочтительно, чтобы при добавлении двух или более нуклеотидов данные нуклеотиды были идентичны.

В предпочтительном варианте реализации ОПМ, применяемый в способе согласно настоящему изобретению, представлен следующими элементами последовательности:



где

X представляет собой химическую группу, выбранную из группы, состоящей из остатка амина, биотина, глицерина, холестерина, дигоксигенина, фтора или производных нуклеотида, включая нуклеотиды без азотистых оснований, дидезоксирибонуклеотиды, 3'-дезоксинуклеотиды, 2'-дезоксиинозин, 2'-дезоксинуридин;

Y представляет собой известную олигонуклеотидную последовательность;

Q представляет собой последовательность идущих друг за другом вырожденных (неоднозначных) оснований ДНК, предпочтительно выбранных из N, V, H, D, B и J, где N представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT, dC и dG; B представляет собой продукт включения нуклеотида из эквимольной смеси dT, dC и dG; D представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dG; H представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dC; V представляет собой продукт включения нуклеотида из эквимольной смеси dA, dC и dG; J представляет собой продукт включения нуклеотида из смеси (0-100% dA) с (0-100% dG), с (0-100% dC), с (0-100% dT), с (0-100% dU), с (0-100% rA), с (0-100% rG), с (0-100% rC), с (0-100% rT), с (0-100% rU);

Z представляет собой рибонуклеотид, выбранный из группы, состоящей из AMP, CMP, GMP, TMP и UMP;

A представляет собой химическую группу, выбранную из группы, состоящей из остатка амина, биотина, глицерина, холестерина, дигоксигенина, фосфата, фтора или производных нуклеотида, включая нуклеотиды без азотистых оснований, дидезоксирибонуклеотиды, 3'-дезоксинуклеотиды, 2'-дезоксиинозин, 2'-дезоксинуридин;

t представляет собой целое число от 0 до 6, т.е. 0, 1, 2, 3, 4, 5 или 6;

p представляет собой целое число от 0 до 10, т.е. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10;

q представляет собой целое число, равное по меньшей мере 1; и

r представляет собой целое число от 0 до 10, т.е. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10.

Для специалиста очевидно, что в случаях когда включено неоднозначное основание, данный пункт формулы изобретения фактически относится к смеси ОПМ, которые отличаются по последовательности неоднозначным основанием, при этом относительный избыток одного нуклеотида над другим определяют по соответствующему молярному отношению нуклеотидов в смеси нуклеотидов, используемой для синтеза данного положения нуклеотида в ОПМ.

Добавление большой химической группы (например, биотина, нескольких нуклеотидов без азотистых оснований, флуоресцентного красителя и т.д.) к 5'-концу ОПМ уменьшает вероятность событий вторичного переключения матрицы и, таким образом, уменьшает встречаемость ДНК-продуктов, содержащих две или более копий 5'-концевой последовательности. Предпочтительно X представляет собой биотин.

Y представляет собой известную последовательность, также называемую определенной последовательностью, и, таким образом, добавляет последовательность нуклеотидов к 5'-концу нуклеиновой кислоты согласно этапу а) и впоследствии в двухцепочечную нуклеиновую кислоту, полученную в способе согласно настоящему изобретению, которую можно использовать в последующих этапах отдельно или в сочетании с определенной последовательностью нуклеиновых кислот, добавленной к 3'-концу одно- или двухцепочечной нуклеиновой кислоты на этапе b), например, чтобы амплифицировать, обнаружить или модифицировать двухцепочечную нуклеиновую кислоту, полученную на этапе e) способа согласно настоящему изобретению. Таким образом, предпочтительно, чтобы у Y была длина, достаточная для того, чтобы позволить специфичную гибридизацию олигонуклеотида, например длина от 15 до 50 нуклеотидов, более предпочтительно от 20 до 40 нуклеотидов. Предпочтительно его последовательность отличается от любой последовательности, находящейся в одно- или двухцепочечной нуклеиновой кислоте согласно этапу а), а также от любой последовательности, добавленной к 3' на этапе b). В предпочтительном варианте реализации Y выбран из группы, состоящей из SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15 и SEQ ID NO: 16 или соответствующей последовательности.

Включение от 1 до 6, более предпочтительно от 2 до 4, т.е. 1, 2, 3, 4, 5 или 6 последовательных неоднозначных оснований в обратный праймер поможет удалить ПЦР-дубликаты из библиотеки. Соответственно в предпочтительном варианте реализации Q представляет собой N, V, H, D, B или J, где J представляет собой смесь, содержащую (0-100% dA), c (0 - 100% dA) dG, c (0-100% dA) dC, c (0-100% dA) dT, c (0-100% dA) dU, c (0-100% dA) rA, c (0-100% dA) rG, c (0-100% dA) rC, c (0-100% dA) rT, c (0-100% dA) rU. Включение последовательных неоднозначных оснований в иницирующий олигонуклеотид предпочтительно, так как это позволяет удалять ПЦР-дубликаты из полученной библиотеки ДНК. Наиболее предпочтительно, чтобы Q был расположен между Z1 и Z2 и представлял собой N. Предпочтительно Q представляет собой N и t равен по меньшей мере 2, более предпочтительно равен по меньшей мере 4.

Добавление последовательных неоднозначных оснований поможет удалить ПЦР-дубликаты из полученной библиотеки. Предпочтительно Q представляет собой N и t равен по меньшей мере 2, более предпочтительно равен по меньшей мере 4.

Добавление химической "блокирующей" группы (например, фосфата, биотина, метила, флуоресцентного красителя и т.д.) к 3'-ОН-группе 3'-конца ОПМ предотвращает добавление полинуклеотидного хвоста к ОПМ, которое произойдет, если добавление полинуклеотидного хвоста и обратную транскрипцию осуществляют одновременно. Также добавление химической "блокирующей" группы к 3'-концу ОПМ устранил необходимость инактивировать нагреванием независимую от матрицы ДНК- или РНК-полимеразу или лигазу, используемую в реакции добавления хвоста перед реакцией обратной транскрипции. Наконец, добавление химической "блокирующей" группы к ОПМ может уменьшить склонность к взаимодействию с матрицами, несущими нуклеотид rG на 5'-конце, данный феномен наблюдают, когда ОПМ с неблокированным 3'-ОН используют на матрицах РНК. Предпочтительно, чтобы A был выбран из группы, состоящей из остатка амина, биотина, глицерина, холестерина, дигоксигенина, фосфата, фтора или производных нуклеотида, включая нуклеотиды без азотистых оснований, дидезоксирибонуклеотиды, 3'-дезоксинуклеотиды, 2'-дезоксиназидин, 2'-дезоксидиридин. Более предпочтительно A представляет собой нуклеотид без азотистого основания, выбранный из группы, состоящей из фурана без азотистого основания, rSpacer, Spacer 18, Spacer 9, Spacer C3 или Spacer C12. Еще более предпочтительно A содержит более чем один сайт без азотистого основания и представляет собой фуран без азотистого основания, т.е. три последовательных фурана без азотистых оснований.

Этап с) согласно настоящему изобретению включает гибридизацию иницирующего олигонуклеотида и синтез кДНК или кРНК с помощью подходящего фермента, такого как обратная транскриптаза, с получением двухцепочечной нуклеиновой кислоты. В предпочтительном варианте реализации реакцию добавления хвоста осуществляют с помощью концевой дезоксирибонуклеотидтрансферазы (этап b)) и реакцию гибридизации (этап с)) с помощью обратной транскриптазы. Более предпочтительно обратная транскриптаза одновременно использует активность полимеразы и активность концевой трансферазы и, таким образом, данный фермент можно применять для осуществления этапа b), а также этапа с) способа согласно настоящему изобретению. Еще более предпочтительно фермент обратную транскриптазу выбирают из группы, состоящей из OT MMLV, которая, например, доступна от NEB, OT Superscript II или OT Superscript III, которая, например, доступна от Invitrogen, OT Multiscribe, которая, например, доступна от Applied Biosystems, OT SMART MMLV или OT SMARTScribe, которая, например, доступна от Clontech. В еще более предпочтительном варианте реализации используют OT M-MLV SuperScribe II или OT SMARTScribe. Предпочтительно, чтобы выбирали полимеразы, которые проявляют как активность полимеразы, т.е. могут синтезировать комплементарную нуклеиновую кислоту на основе матричной нуклеиновой кислоты, так и активность концевой трансферазы, т.е. способны при достижении 5'-конца одноцепочечной нуклеиновой кислоты присоединять дополнительные рибо- и/или дезоксирибонуклеотиды без матрицы. Предпочтительно они способны добавлять 1 или более, предпочтительно от 2 до 20, т.е. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 дополнительных рибо- и/или дезоксирибо-

динуклеотидов к 5'-концу одноцепочечной нуклеиновой кислоты, что позволяет гибридизацию олигонуклеотида, переключающего матрицу, и 3'-конца цепи нуклеиновой кислоты. Предпочтительно, чтобы обратная транскриптаза присоединяла преимущественно гомонуклеотидный участок, предпочтительно гомотринуклеотидный участок, который впоследствии способствует переключению обратной транскриптазы с матричной нуклеиновой кислоты и ее гибридизации с олигонуклеотидом, переключающим матрицу. Более предпочтительно присоединяется три dCTP или gCTP.

В другом варианте реализации настоящего изобретения для синтеза двухцепочечной нуклеиновой кислоты путем удлинения 3'-конца кДНК или кРНК согласно этапу е) способа согласно настоящему изобретению требуется ферментативная активность. Гибридизация ОПМ с присоединенными гомотринуклеотидами двухцепочечной нуклеиновой кислоты, полученной на этапе с) способа согласно настоящему изобретению, позволяет элонгацию синтезированной цепи нуклеиновой кислоты с применением ОПМ в качестве новой матрицы. Предпочтительно данную реакцию осуществляют с помощью обратной транскриптазы, более предпочтительно с помощью обратной транскриптазы MLLV. В дополнительном предпочтительном варианте реализации используемая обратная транскриптаза способна переключиться на матрицу, включающую двухцепочечную нуклеиновую кислоту ДНК/РНК и/или ДНК/ДНК.

Нуклеиновые кислоты, синтезированные с помощью способа согласно настоящему изобретению, можно затем проанализировать в рамках целевых применений, т.е. глубокого секвенирования, генотипирования или клонирования.

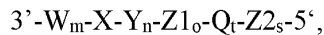
В одном варианте реализации настоящего изобретения двухцепочечные нуклеиновые кислоты иммобилизованы на поверхности, предпочтительно посредством физической адсорбции, ковалентного связывания, аффинного связывания или захвата матрицей. Более предпочтительно нуклеиновые кислоты согласно настоящему изобретению иммобилизованы на микрочипе, поверхности микрочипа, подложках на основе кремния, определенных твердых подложках платформы секвенирования нового поколения.

Способ согласно настоящему изобретению может дополнительно включать этап применения синтезированной двухцепочечной нуклеиновой кислоты в качестве матрицы для ПЦР-амплификации. Согласно одному варианту реализации способ согласно настоящему изобретению дополнительно включает помещение синтезированной двухцепочечной нуклеиновой кислоты или отдельной цепи, полученной из нее, в условия амплификации. Такие условия могут включать добавление прямого или обратного праймера, сконструированного таким образом, чтобы амплифицировать всю или желательную часть синтезированной двухцепочечной нуклеиновой кислоты, dNTP и полимеразу, подходящую для эффективной амплификации, предпочтительно термоустойчивую полимеразу. Первоначальный этап осуществления амплификации может включать денатурацию синтезированной двухцепочечной нуклеиновой кислоты и обеспечение доступности синтезированной нуклеиновой кислоты для связывания праймера. Синтезированная двухцепочечная нуклеиновая кислота предпочтительно содержит по меньшей мере часть иницирующей олигонуклеотидной последовательности, цепь, комплементарную обеспеченной одноцепочечной нуклеиновой кислоте, и по меньшей мере часть ОПМ. Эта информация относительно двух указанных синтезированных цепей нуклеиновой кислоты позволяет предоставить олигонуклеотиды, комплементарные соответствующим последовательностям, для получения больших количеств синтезированных двухцепочечных нуклеиновых кислот. В предпочтительном варианте реализации способ согласно настоящему изобретению включает гибридизацию по меньшей мере одного олигонуклеотида, способного, по меньшей мере, гибридизоваться с частью иницирующего олигонуклеотида согласно этапу с), или гибридизацию олигонуклеотида, переключающего матрицу, согласно этапу d) с двухцепочечной нуклеиновой кислотой, синтезированной на этапе е). Предпочтительно один праймер комплементарен иницирующему олигонуклеотиду и другой праймер комплементарен олигонуклеотиду, переключающему матрицу. Можно применять концентрации праймеров в диапазоне от 200-300 нМ, т.е. 210, 220, 230, 240, 250, 260, 270, 280, 290, 300 нМ.

Продукт амплификации согласно этапу f) способа согласно настоящему изобретению можно затем проанализировать в рамках целевых применений, т.е. глубокого секвенирования, генотипирования или клонирования. В одном варианте реализации настоящего изобретения продукт амплификации иммобилизован на поверхности, предпочтительно посредством физической адсорбции, ковалентного связывания, аффинного связывания или захвата матрицей.

Путем амплификации синтезированной двухцепочечной нуклеиновой кислоты можно получить большие количества нуклеиновой кислоты, позволяющие использовать различные целевые методики обработки. Так как синтезированная нуклеиновая кислота имеет определенные 3'- и 5'-концы, то есть возможность определения интересующей последовательности внутри обеспеченной одноцепочечной нуклеиновой кислоты по способу согласно настоящему изобретению. Таким образом, в предпочтительном варианте реализации способ согласно настоящему изобретению дополнительно включает этап определения по меньшей мере части последовательности одноцепочечной нуклеиновой кислоты. Предпочтительно определяют полную последовательность одноцепочечной нуклеиновой кислоты.

Во втором аспекте настоящего изобретения предложен иницирующий олигонуклеотид, содержащий следующие элементы последовательности:



где

W в каждом случае независимо выбран из dA, dG, dC, dT и dU;

X выбран из dA, dG, dC, dT, dU, rA, rG, rC, rT и rU;

Y представляет собой полинуклеотид длиной по меньшей мере 10 нуклеотидов, который на 80% последовательности или более состоит из идентичных нуклеотидов или динуклеотидов, выбранных из dA, dG, dC, dT, dU, rA, rG, rC, rT, rU, AC, AG, AT, AU, CA, CG, CT, CU, GA, GC, GT, GU, TA, TC, TG, TU, AA, CC, GG, TT, UU, UA, UC, UG и UT, и который на другие 20% последовательности или менее состоит из нуклеотидов или динуклеотидов, которые отличны от основного нуклеотида или динуклеотида и также выбраны из dA, dG, dC, dT, dU, rA, rG, rC, rT, rU, AC, AG, AT, AU, CA, CG, CT, CU, GA, GC, GT, GU, TA, TC, TG, TU, AA, CC, GG, TT, UU, UA, UC, UG и/или UT, при условии, что X отличен от нуклеотида или динуклеотида, который составляет большую часть Y;

Q представляет собой последовательность идущих друг за другом вырожденных (неоднозначных) оснований ДНК, предпочтительно выбранных из N, V, H, D, B и J, где N представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT, dC и dG, т.е. представляет собой dA, dT, dC или dG; B представляет собой продукт включения нуклеотида из эквимольной смеси dT, dC и dG, т.е. представляет собой dT, dC и dG; D представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dG, т.е. представляет собой dA, dT или dG; H представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dC, т.е. представляет собой dA, dT или dC; V представляет собой продукт включения нуклеотида из эквимольной смеси dA, dC и dG, т.е. представляет собой dA, dC или dG; J представляет собой продукт включения нуклеотида из смеси (0-100% dA) с (0-100% dG), с (0-100% dC), с (0-100% dT), с (0-100% dU), с (0-100% rA), с (0-100% rG), с (0-100% rC), с (0-100% rT), с (0-100% rU);

Z1 представляет собой полинуклеотид длиной по меньшей мере 5 нуклеотидов с определенной последовательностью, при этом указанная последовательность отличается от $W_m - X - Y_n$, предпочтительно указанная последовательность также отличается от $Q_t - Z2_s$;

Z2 представляет собой полинуклеотид длиной по меньшей мере 5 нуклеотидов с определенной последовательностью, при этом указанная последовательность отличается от $W_m - X - Y_n - Z1_o - Q_t$;

m представляет собой целое число от 0 до 6, т.е. 0, 1, 2, 3, 4, 5 или 6;

n представляет собой целое число от 10 до 100, если Y выбран из dA, dG, dC, dT, dU, rA, rG, rC, rT, и rU, целое число от 5 до 50, если Y выбран из AC, AG, AT, AU, CA, CG, CT, CU, GA, GC, GT, GU, TA, TC, TG, TU, AA, CC, GG, TT, UU, UA, UC, UG и UT;

o представляет собой 0 или 1;

s представляет собой 0 или 1; и

t представляет собой целое число от 0 до 6, т.е. 0, 1, 2, 3, 4, 5 или 6.

Y представляет собой часть иницирующего олигонуклеотида, которая способна гибридизоваться с присоединенными последовательными нуклеиновыми кислотами. Таким образом, предпочтительно, чтобы комплементарность последовательности Y с добавленными нуклеиновыми кислотами составляла по меньшей мере 90%. Следовательно, ее длина предпочтительно соответствует длине добавленных последовательных нуклеотидов, более предпочтительно длине от 10 до 100 нуклеотидов, т.е. 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 или 100 нуклеотидов. Авторы настоящего изобретения обнаружили, что короткий Y улучшает достоверность секвенирования. Тем не менее, для того чтобы позволить гибридизацию, предпочтительно при строгих условиях, предпочтительно, чтобы длина Y составляла от 11 до 50, более предпочтительно от 12 до 40, более предпочтительно от 13 до 30 и наиболее предпочтительно от 14 до 20 нуклеотидов.

Авторы настоящего изобретения обнаружили, что присутствие небольшого количества неидентичных нуклеотидов и/или динуклеотидов улучшает достоверность секвенирования. Таким образом предпочтительно, чтобы последовательность Y состояла из по меньшей мере 80% идентичных нуклеотидов и/или динуклеотидов, выбранных из dA, dG, dC, dT, dU, rA, rG, rC, rT, rU, AC, AG, AT, AU, CA, CG, CT, CU, GA, GC, GT, GU, TA, TC, TG, TU, AA, CC, GG, TT, UU, UA, UC, UG и UT, при этом другие 20% последовательности или менее состоят из нуклеотидов или динуклеотидов, которые отличны от основного нуклеотида и/или динуклеотида и также выбраны из dA, dG, dC, dT, dU, rA, rG, rC, rT, rU, AC, AG, AT, AU, CA, CG, CT, CU, GA, GC, GT, GU, TA, TC, TG, TU, AA, CC, GG, TT, UU, UA, UC, UG и UT. В предпочтительном варианте реализации указанные основные нуклеотиды представляют собой A и/или T. В другом предпочтительном варианте реализации указанные нуклеотиды представляют собой динуклеотиды, предпочтительно AA, TT, AT или TA. В другом предпочтительном варианте реализации второстепенные нуклеотиды представляют собой C и/или G. В другом предпочтительном варианте реализации указанные нуклеотиды представляют собой динуклеотиды, предпочтительно CC, GG, CG и/или GC. В предпочтительном варианте реализации от 80 до 99% последовательности Y состоит из идентичных нуклеотидов и/или динуклеотидов, более предпочтительно от 85 до 95% (для специалиста очевидно, что в данном случае "n" должен быть равен по меньшей мере 20), более предпочтительно от 88 до 92% и наиболее предпочтительно приблизительно 90%. Таким образом, Y в типичном предпочтительном варианте

реализации может содержать 9 нуклеотидов Т и один нуклеотид G или С или 14 нуклеотидов Т и один нуклеотид G или С.

В случаях когда Y содержит один или два различных нуклеотида, предпочтительно, чтобы данный(е) нуклеотид(ы) был(и) расположен(ы) посередине Y или близко к ней (т.е. в пределах от 1 до 4 оснований от нее).

В другом предпочтительном варианте реализации второго аспекта настоящего изобретения предпочтительно, чтобы Y представлял собой последовательный участок нуклеотидов, состоящий только из Т, и n находился в диапазоне от 10 до 60, т.е. 50, 45, 40, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11 или 10, более предпочтительно от 11 до 50, более предпочтительно от 12 до 40, более предпочтительно от 13 до 30 и наиболее предпочтительно от 14 до 20. Более предпочтительно n представляет собой 30, 20, 16 или 15. Наиболее предпочтительно, чтобы n представлял собой 20 или 16. В качестве альтернативы данному предпочтительному варианту реализации Y содержит один или два отличных нуклеотида, предпочтительно G или С.

В альтернативном предпочтительном варианте реализации последовательность Y представляет собой последовательный участок нуклеотидов, состоящий только из Т, не считая одного или двух остатков G и/или С.

Z1 представляет собой часть иницирующего олигонуклеотида, которую используют после синтеза двухцепочечной молекулы нуклеиновой кислоты, чтобы позволить сайт-специфическую гибридизацию с другим олигонуклеотидом. Таким образом, Z1 предпочтительно представляет собой определенную последовательность, добавляемую к 3'-концу нуклеиновой кислоты, содержащейся в образце. Длина Z1 составляет по меньшей мере 5 нуклеотидов, более предпочтительно находится в диапазоне от 5 до 50 нуклеотидов, более предпочтительно в диапазоне от 10 до 30 нуклеотидов. Длину выбирают таким образом, чтобы праймер мог специфично гибридизоваться с Z1 в последующих реакциях ПЦР-амплификации. В предпочтительном варианте реализации последовательность нуклеиновой кислоты Z1 выбирают из группы, состоящей из SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15 и SEQ ID NO: 16.

Предпочтительно Z1 не идентичен Z2.

Z2 представляет собой часть иницирующего олигонуклеотида, которую используют после синтеза двухцепочечной молекулы нуклеиновой кислоты, чтобы позволить сайт-специфическую гибридизацию с другим олигонуклеотидом. Таким образом, Z2 предпочтительно представляет собой определенную последовательность, добавляемую к 3'-концу нуклеиновой кислоты, содержащейся в образце. Длина Z2 составляет по меньшей мере 5 нуклеотидов, более предпочтительно находится в диапазоне от 5 до 50 нуклеотидов, более предпочтительно в диапазоне от 10 до 30 нуклеотидов. Длину выбирают таким образом, чтобы праймер мог специфично гибридизоваться с Z1 в последующих реакциях ПЦР-амплификации. В предпочтительном варианте реализации последовательность нуклеиновой кислоты Z2 выбирают из группы, состоящей из SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15 и SEQ ID NO: 16 или соответствующей последовательности.

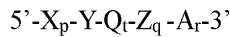
Предпочтительно Z2 не идентичен Z1.

Включение от 1 до 6, более предпочтительно от 2 до 4, т.е. 1, 2, 3, 4, 5 или 6 последовательных неоднозначных оснований в праймер, т.е. между Z1 и Z2, позволит удалять ПЦР-дубликаты из библиотеки. Предпочтительно Q представляет собой последовательность идущих друг за другом вырожденных (неоднозначных) оснований ДНК, предпочтительно в каждом случае независимо, выбранных из N, V, H, D, B и J, где N представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT, dC и dG; V представляет собой продукт включения нуклеотида из эквимольной смеси dT, dC и dG; D представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dG; H представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dC; V представляет собой продукт включения нуклеотида из эквимольной смеси dA, dC и dG; J представляет собой продукт включения нуклеотида из смеси (0-100% dA) с (0-100% dG), с (0-100% dC), с (0-100% dT), с (0-100% dU), с (0-100% rA), с (0-100% rG), с (0-100% rC), с (0-100% rT), с (0-100% rU). Включение последовательных неоднозначных оснований в иницирующий олигонуклеотид предпочтительно, так как это позволяет удалять ПЦР-дубликаты из полученной библиотеки ДНК. Наиболее предпочтительно, чтобы Q располагался между Z1 и Z2 и представлял собой N. Предпочтительно Q представляет собой N и t равен по меньшей мере 2, более предпочтительно равен по меньшей мере 4.

В другом предпочтительном варианте реализации сумма t и s равна 0, например Z2 и Q отсутствуют.

Особенно предпочтительные примеры иницирующего олигонуклеотида представляют собой нуклеотиды с последовательностями, соответствующими SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19 и SEQ ID NO: 20.

В третьем аспекте настоящего изобретения предложен олигонуклеотид, переключающий матрицу, содержащий следующие элементы последовательности:



где

X представляет собой химическую группу, выбранную из группы, состоящей из остатка амина, биотина, глицерина, холестерина, дигоксигенина, фтора или производных нуклеотида, включая нуклеотиды без азотистых оснований, дидезоксирибонуклеотиды, 3'-дезоксинуклеотиды, 2'-дезоксининозин и 2'-дезоксинуридин, предпочтительно биотин;

Y представляет собой известную (определенную) олигонуклеотидную последовательность;

Q представляет собой последовательность идущих друг за другом вырожденных (неоднозначных) оснований ДНК, предпочтительно выбранных из N, V, H, D, B и J, где N представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT, dC и dG; B представляет собой продукт включения нуклеотида из эквимольной смеси dT, dC и dG; D представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dG; H представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dC; V представляет собой продукт включения нуклеотида из эквимольной смеси dA, dC и dG; J представляет собой продукт включения нуклеотида из смеси (0-100% dA) с (0-100% dG), с (0-100% dC), с (0-100% dT), с (0-100% dU), с (0-100% rA), с (0-100% rG), с (0-100% rC), с (0-100% rT), с (0-100% rU);

Z представляет собой рибонуклеотид, выбранный из группы, состоящей из AMP, CMP, GMP, TMP и UMP, предпочтительно GMP;

A представляет собой химическую группу, выбранную из группы, состоящей из остатка амина, биотина, глицерина, холестерина, дигоксигенина, фосфата, фтора или производных нуклеотида, включая нуклеотиды без азотистых оснований, дидезоксирибонуклеотиды, 3'-дезоксинуклеотиды, 2'-дезоксининозин и 2'-дезоксинуридин;

p представляет собой целое число от 0 до 10, т.е. 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10, более предпочтительно 1 или 2, наиболее предпочтительно 1;

q представляет собой целое число от по меньшей мере 1, предпочтительно 1 до 10, т.е. 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10, наиболее предпочтительно 3;

r представляет собой целое число от 0 до 10, т.е. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10, предпочтительно 0 или 1, наиболее предпочтительно 0;

t представляет собой целое число от 0 до 6, т.е. 0, 1, 2, 3, 4, 5 или 6.

Добавление большой химической группы (например, биотина, нескольких сайтов без азотистого основания, флуоресцентного красителя и т.д.) к 5'-концу ОПМ уменьшает вероятность событий вторичного переключения матрицы и, таким образом, уменьшает встречаемость ДНК-продуктов, содержащих две или более копий 5'-концевой последовательности. Предпочтительно X представляет собой биотин.

Y представляет собой известную последовательность, также называемую определенной последовательностью, которая добавляет последовательность нуклеотидов к 5'-концу нуклеиновой кислоты согласно этапу а) и впоследствии в двухцепочечную нуклеиновую кислоту, полученную в способе согласно настоящему изобретению, которую можно использовать в последующих этапах отдельно или в сочетании с определенной последовательностью нуклеиновых кислот, добавляемой к 3'-концу одно- или двухцепочечной нуклеиновой кислоты на этапе б), чтобы, например, амплифицировать, обнаружить или модифицировать двухцепочечную нуклеиновую кислоту, полученную на этапе е) способа согласно настоящему изобретению. Таким образом предпочтительно, чтобы у Y была длина, достаточная для того, чтобы позволить специфичную гибридизацию олигонуклеотида, например длина от 15 до 50 нуклеотидов, более предпочтительно от 20 до 40 нуклеотидов. Предпочтительно его последовательность отличается от любой последовательности, находящейся в одно- или двухцепочечной нуклеиновой кислоте согласно этапу а), а также от любой последовательности, присоединенной к 3'-концу на этапе б). В предпочтительном варианте реализации Y выбран из группы, состоящей из SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15 и SEQ ID NO: 16 или соответствующей последовательности.

В другом предпочтительном варианте реализации Q представляет собой последовательность идущих друг за другом вырожденных (неоднозначных) оснований ДНК, предпочтительно выбранных из N, V, H, D, B и J, где N представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT, dC и dG; B представляет собой продукт включения нуклеотида из эквимольной смеси dT, dC и dG; D представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dG; H представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dC; V представляет собой продукт включения нуклеотида из эквимольной смеси dA, dC и dG; J представляет собой продукт включения нуклеотида из смеси (0-100% dA) с (0 - 100% dA) dG, с (0-100% dA) dC, с (0-100% dA) dT, с (0-100% dA) dU, с (0-100% dA) rA, с (0-100% dA) rG, с (0-100% dA) rC, с (0-100% dA) rT, с (0-100% dA) rU.

A осуществляет функцию, описанную выше в контексте первого аспекта настоящего изобретения. Предпочтительно A выбран из группы, состоящей из остатка амина, биотина, глицерина, холестерина, дигоксигенина, фосфата, фтора или производных нуклеотида, включая нуклеотиды без азотистых осно-

ваний, дидезоксирибонуклеотиды, 3'-дезоксинуклеотиды, 2'-дезоксинозин, 2'-дезоксинуридин. Более предпочтительно А представляет собой нуклеотид без азотистого основания, выбранный из группы, состоящей из фурана без азотистого основания, rSpacer, Spacer 18, Spacer 9, Spacer C3 или Spacer C12. Еще более предпочтительно А содержит более чем один сайт без азотистого основания, т.е. три последовательных фурана без азотистого основания.

В четвертом аспекте настоящего изобретения предложена нуклеиновая кислота, содержащая иницирующий олигонуклеотид согласно второму аспекту настоящего изобретения. Предпочтительно указанная нуклеиновая кислота содержит последовательность использованного иницирующего олигонуклеотида.

В пятом аспекте настоящего изобретения предложен набор, позволяющий осуществление способа согласно первому аспекту настоящего изобретения. Указанный набор может содержать реагенты, необходимые для осуществления каждого этапа с а) по f) способа согласно настоящему изобретению. Набор может содержать, например, один или более из любых компонентов реакционной смеси, описанных в отношении этапов с а) по f) способа. Например, набор может содержать полимеразу (например, полимеразу, способную к переключению матрицы, термоустойчивую полимеразу или комбинации перечисленных полимераз), иницирующий олигонуклеотид, олигонуклеотид, переключающий матрицу, dNTP, соли, подходящие кофакторы для ферментов, ингибиторы нуклеаз, например ингибитор РНКазы или ингибитор ДНКазы, одно или более вспомогательных веществ для содействия амплификации или репликации GC-богатых последовательностей (например, бетаин, диметилсульфоксид, этиленгликоль или 1,2-пропандиол, или комбинации перечисленных веществ), один или более дестабилизирующих агентов, например дитиотреитол, фермент, способный образовывать двухцепочечную нуклеиновую кислоту с 3'-выступом (например, эндонуклеазы рестрикции, концевая трансфераза или комбинация перечисленных ферментов), и блокирующий нуклеотид, предпочтительно 3d-NTP, 3-Me-NTP и ddNTP, или любой другой желательный компонент набора, такой как пробирки, гранулы, микрожидкостные чипы и тому подобное. В предпочтительном варианте реализации набора согласно настоящему изобретению обсуждаемый набор содержит реагент, способный добавлять нуклеотиды к 3'-концу одноцепочечной нуклеиновой кислоты, предпочтительно фермент, более предпочтительно поли-А-полимеразу или концевую трансферазу. Предпочтительно, чтобы набор также содержал иницирующий олигонуклеотид и олигонуклеотид, переключающий матрицу, и, возможно, фермент, способный расщеплять пептидные связи в белках, более предпочтительно данный фермент представляет собой эндо-или экзопептидазу, наиболее предпочтительно протеиназу К. В другом предпочтительном варианте реализации данного аспекта в указанном наборе могут быть обеспечены реагенты, необходимые для осуществления способов определения последовательности. Более предпочтительно реагенты, поставляемые в данном наборе, обеспечивают средства для секвенирования нового поколения, например для обогащения мишенью с помощью зондов захвата.

В шестом аспекте настоящего изобретения предложен чип, содержащий по меньшей мере одну нуклеиновую кислоту согласно четвертому аспекту настоящего изобретения. Предпочтительно указанный чип позволяет определение последовательности указанной нуклеиновой кислоты. Более предпочтительно чип можно применять для измерения изменений в уровнях экспрессии, чтобы обнаружить однонуклеотидные полиморфизмы (SNP), или для генотипирования или целевого повторного секвенирования, или в качестве средства для глубокого секвенирования.

Методы массового параллельного секвенирования (МПС) проложили путь в новые сферы в нескольких областях исследований, таких как индивидуализированная медицина. Желательно предоставить как последовательность, так и встречаемость молекул нуклеиновых кислот, которые присутствуют в любое конкретное время в определенном типе клеток, ткани или органа. Например, подсчет количества мРНК, которые кодируются отдельными генами (так называемым транскриптом), позволяет узнать возможности кодирования белков, которые вносят основной вклад в фенотип. В седьмом аспекте настоящего изобретения предложено применение двухцепочечной нуклеиновой кислоты, синтезированной с помощью способа согласно настоящему изобретению, или одноцепочечной нуклеиновой кислоты, полученной из нее. В предпочтительном варианте реализации нуклеиновые кислоты, синтезированные с помощью способа согласно настоящему изобретению, можно применять для секвенирования или анализа экспрессии, клонирования, мечения, для идентификации генов или некоторых последовательностей нуклеотидов. Предпочтительно применение включает применение в персонализированной медицине, для контролирования терапии; предсказания, прогноза, раннего обнаружения заболеваний человека или животного или для судебной медицины, для анализа последовательностей нуклеиновых кислот вирусов, бактерий, грибов, животных, или растений, или клеток, полученных из них, предпочтительно для характеристики растений, плодов, проверки скрещивания, обнаружения заболеваний растений, семян или плодов.

Примеры

Пример 1. Образцы РНК и ДНК.

Синтетическую cel-miR-39 (Sigma-Aldrich), микроРНК длиной 22 н.т. из *C.elegans*, использовали в качестве исходного материала для контроля секвенирования малой РНК. Синтетическую ДНК длиной 22 н.т. (Sigma-Aldrich) с последовательностью, эквивалентной cel-miR-39, использовали в качестве исходного материала для контроля секвенирования ДНК. Циркулирующую ДНК выделяли из фракции плазмы из

образцов крови от двух добровольных здоровых доноров (DI - женщина и DI - мужчина). Циркулирующую РНК выделяли из плазмы крови от двух добровольных здоровых женщин (RI и RII). Сбор данных образцов был одобрен Комитетом по этике Медицинского факультета в Гейдельберге. Циркулирующие ДНК и РНК, выделенные из плазмы крови человека, обработанную бисульфитом ДНК из клеток U2OS и тотальную РНК из клеток U2OS, фракционированную в присутствии Mg^{2+} и обогащенную поли(А), использовали в качестве исходных материалов для получения библиотеки кДНК и последующего секвенирования на MiSeq Illumina.

Пример 2. Олигонуклеотиды для синтеза кДНК.

Последовательности всех праймеров, используемых в данной работе, представлены на фиг. 2-5. В процессе разработки данного способа исследовали несколько олигонуклеотидов, переключающих матрицу (ОПМ), с различными структурами.

Пример 3. Синтез первой цепи кДНК и переключение матрицы.

Синтетическую малую РНК или ДНК разбавляли в воде, чтобы получить концентрации 1 нг/мкл и 5 пг/мкл, и использовали в качестве исходного материала для синтеза первой цепи кДНК. Оптимизированный протокол для получения подготовленных к секвенированию библиотек ДНК описан далее. РНК полиаденилировали, применяя поли(А)-полимеразу E.coli (New England Biolabs), в 1х буфере PAP, содержащем 10 единиц рекомбинантного ингибитора РНКаз (Clontech) и 0,1 мМ АТР, в течение 10 мин при 37°C и останавливали реакцию нагревом при 65°C в течение 20 мин. К ДНК присоединяли поли(dA)-хвост, применяя концевую дезоксирибонуклеотидтрансферазу (New England Biolabs), в 1х буфере TdT и 0,1 мМ dATP в течение 30 мин при 37°C, и термоинактивировали фермент в течение 10 мин при 70°C. Перед реакцией добавления поли(dA)-хвоста образцы циркулирующей ДНК и обработанной бисульфитом ДНК денатурировали нагреванием при 95°C в течение 5 мин и быстро охлаждали на льду. В некоторых экспериментах РНК- и ДНК-матрицы предварительно обрабатывали полинуклеотидкиназой T4 (New England Biolabs) в течение 10 мин в 1х буфере PAP/TdT перед реакцией добавления поли(А/dА)-хвоста. Для проведения обратной транскрипции 1 мкл содержащей поли(А)-хвост РНК или содержащей поли(dА)-хвост ДНК смешивали с 2,5 мкл 1х буфера для ОТ первой цепи, содержащего 20% ДМСО и 1 мкл заякоренного на одном основании поли(dT)-праймера Illumina (конечная концентрация 0,1 мкМ для 1 нг РНК и 0,001 мкМ для 5 пг ДНК). Полученный раствор инкубировали при 72°C в течение 2 мин, а затем охлаждали при 42°C в течение 1 мин. На следующем этапе общую реакционную смесь, содержащую 2 мкл 5х буфера для ОТ первой цепи (Clontech), 1 мкл dNTP (по 10 мМ каждого), 1 мкл полимеразы SMARTScribe для ОТ (Clontech), 0,25 мкл ДТТ (100 мМ) и 0,25 мкл рекомбинантного ингибитора РНКаз (Clontech) добавляли в раствор ДНК(РНК)/праймеров и инкубировали в течение 15 мин при 42°C. Затем в реакционную смесь для ОТ добавляли 1 мкл 10 мкМ заблокированного биотином по 5'-концу олигонуклеотида, переключающего матрицу (ОПМ), и инкубировали в течение дополнительных 15 мин при 42°C. Реакцию ОТ останавливали нагреванием при 70°C в течение 10 мин. Либо 1 мкл, либо 10 мкл реакционной смеси для ОТ использовали для амплификации кДНК в общем объеме 100 мкл. Амплификацию кДНК проводили в 2х общей реакционной смеси с Taq-полимеразой (Qiagen) с применением праймеров для амплификации кДНК (фиг. 2А) в конечной концентрации 250 нМ. Амплифицированные кДНК очищали на колонке, применяя набор для очистки продуктов ПЦР QIAquick (Qiagen), и секвенировали, применяя автоматизированное секвенирование по методу Сенгера в GATC GmbH (Констанц, Германия). Для секвенирования нового поколения фрагменты ДНК дополнительно очищали из 4% агарозного геля, применяя набор для экстрагирования из геля PureLink (Life Technologies), и анализировали с помощью высокочувствительных ДНК-чипов Agilent Bioanalyser.

Пример 4. Глубокое секвенирование.

Платформу Illumina MiSeq использовали для секвенирования библиотек ДНК, полученных с помощью способа, описанного выше. Специально разработанный праймер для секвенирования, состоящий из стандартного праймера для секвенирования Illumina и 3'-концевого GGG тринуклеотида, использовали для секвенирования с помощью MiSeq Illumina, чтобы решить проблему с требуемой сложностью первых нескольких оснований, необходимой для успешной идентификации кластеров. Специально разработанный поли(Т)-праймер для секвенирования можно применять для секвенирования в обратном направлении, что, таким образом, позволяет получить результаты секвенирования спаренных концов (в оба направления). Библиотеки ДНК разбавляли до концентрации 5 нМ, денатурировали с помощью 0,2 N NaOH в течение 5 мин и дополнительно разбавляли до 11 пМ незадолго до загрузки в кассету MiSeq. Запуск MiSeq проводили, применяя набор реагентов для MiSeq (на 50 циклов), в течение 77 циклов.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ синтеза двухцепочечной нуклеиновой кислоты с определенной 3'- и 5'-концевой последовательностью нуклеотидов с использованием образца, содержащего одноцепочечную нуклеиновую кислоту, включающий следующие этапы:

а) обеспечение образца, содержащего одноцепочечную или двухцепочечную нуклеиновую кислоту, возможно, денатурирование указанной двухцепочечной нуклеиновой кислоты;

b) добавление по меньшей мере 5 последовательных нуклеотидов к 3'-концу указанной одноцепочечной или двухцепочечной нуклеиновой кислоты;

c) осуществление гибридизации иницирующего олигонуклеотида, комплементарного указанной добавленной последовательности нуклеотидов, и синтез кДНК или кРНК посредством зависимой от матрицы ДНК- или РНК-полимеразы с получением двухцепочечной нуклеиновой кислоты;

d) осуществление гибридизации олигонуклеотида, переключающего матрицу, с указанной двухцепочечной нуклеиновой кислотой; и

e) удлинение 3'-конца цепи кДНК или кРНК для синтеза двухцепочечной нуклеиновой кислоты, при этом одна цепь указанной нуклеиновой кислоты содержит указанный иницирующий олигонуклеотид и кДНК или кРНК, которая комплементарна указанной одноцепочечной нуклеиновой кислоте и указанному олигонуклеотиду, переключающему матрицу.

2. Способ согласно п.1, отличающийся тем, что указанный образец представляет собой:

(i) жидкий или твердый биопатат или получен из жидкого или твердого биоптата, более предпочтительно указанный образец представляет собой образец крови, образец плазмы, образец сыворотки, образец биологической жидкости, образец слюны, образец мочи, образец семенной жидкости, образец жидкости из плевральной полости, образец жидкости из перитонеальной полости, образец спинномозговой жидкости, мазок с поверхности эпителия, образец мокроты, образец кала, образец эякулята, образец слезы, образец пота, образец лимфатической жидкости, образец бронхиального лаважа, образец плеврального выпота, образец менингеальной жидкости, образец секрета железы, образец тонкоигольного пунктата, микропрепарированные клетки, жидкий образец аспирата наконечником, образец спинномозговой жидкости, образец конъюнктивной жидкости, образец вагинальной жидкости, образец дуоденальной жидкости, образец сока поджелудочной железы или образец желчи или

(ii) получен из ископаемых остатков, останков вымерших организмов, растений, плодов и животных, микробов, бактерий или вирусов.

3. Способ согласно любому из пп.1-2, отличающийся тем, что:

(i) указанная одноцепочечная нуклеиновая кислота представляет собой ДНК, предпочтительно фрагментированную и/или обработанную бисульфитом ДНК, или РНК, предпочтительно миРНК, малую РНК, рiРНК или обработанную бисульфитом РНК, и/или

(ii) указанные последовательные нуклеотиды идентичны или состоят из идентичных динуклеотидов, тринуклеотидов, квадрануклеотидов, пентануклеотидов или полинуклеотидов, при этом указанные последовательные идентичные нуклеотиды предпочтительно выбраны из группы, состоящей из рибонуклеотидов или дидезоксирибонуклеотидов А, Т, G, С или U, или выбраны из группы, состоящей из двух или более рибонуклеотидов или дидезоксирибонуклеотидов, и более предпочтительно указанные идентичные рибонуклеотиды или дидезоксирибонуклеотиды присоединяют с использованием фермента, выбранного из группы, состоящей из поли(А)-полимеразы, поли(У)-полимеразы, поли(Г)-полимеразы, концевой трансферазы, ДНК-лигазы, РНК-лигазы, и указанные динуклеотиды и тринуклеотиды присоединяют РНК-лигазы, и/или

(iii) реакцию добавления хвоста полинуклеотидов осуществляют со смесью двух или более рибонуклеотидов или дидезоксирибонуклеотидов, и/или

(iv) иницирующий олигонуклеотид содержит следующие элементы последовательности нуклеотидов:

3'-W_m-X-Y_n-Z1_o-Q_t-Z2_s-5', где

W в каждом случае независимо выбран из dA, dG, dC, dT и dU;

X выбран из dA, dG, dC, dT, dU, rA, rG, rC, rT и rU;

Y представляет собой полинуклеотид длиной по меньшей мере 10 нуклеотидов, последовательность которого на 80% или более состоит из идентичных нуклеотидов или динуклеотидов, выбранных из dA, dG, dC, dT, dU, rA, rG, rC, rT, rU, AC, AG, AT, AU, CA, CG, CT, CU, GA, GC, GT, GU, TA, TC, TG, TU, AA, CC, GG, TT, UU, UA, UC, UG и UT, и на остальные 20% или менее состоит из нуклеотидов или динуклеотидов, которые отличны от основного нуклеотида или динуклеотида и также выбраны из dA, dG, dC, dT, dU, rA, rG, rC, rT, rU, AC, AG, AT, AU, CA, CG, CT, CU, GA, GC, GT, GU, TA, TC, TG, TU, AA, CC, GG, TT, UU, UA, UC, UG и/или UT, при условии, что X отличен от нуклеотида или динуклеотида, который составляет большую часть Y;

Q представляет собой последовательность идущих друг за другом вырожденных (неоднозначных) оснований ДНК, предпочтительно выбранных из N, V, H, D, B и J, где N представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT, dC и dG; B представляет собой продукт включения нуклеотида из эквимольной смеси dT, dC и dG; D представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dG; H представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dC; V представляет собой продукт включения нуклеотида из эквимольной смеси dA, dC и dG; J представляет собой продукт включения нуклеотида из смеси (0-100% dA) с (0-100% dG), с (0-100% dC), с (0-100% dT), с (0-100% dU), с (0-100% rA), с (0-100% rG), с (0-100% rC), с (0-100% rT), с (0-100% rU);

Z1 представляет собой полинуклеотид длиной по меньшей мере 5 нуклеотидов с определенной по-

следовательностью, при этом указанная последовательность отличается от W_m-X-Y_n , предпочтительно указанная последовательность также отличается от Q_t-Z2_s ;

Z2 представляет собой полинуклеотид длиной по меньшей мере 5 нуклеотидов с определенной последовательностью, при этом указанная последовательность отличается от $W_m-X-Y_n-Z1_o-Q_t$;

m представляет собой целое число от 0 до 6, т.е. 0, 1, 2, 3, 4, 5 или 6;

n представляет собой целое число от 10 до 100, если Y выбран из dA, dG, dC, dT, dU, rA, rG, rC, rT и rU, целое число от 5 до 50, если Y выбран из AC, AG, AT, AU, CA, CG, CT, CU, GA, GC, GT, GU, TA, TC, TG, TU, AA, CC, GG, TT, UU, UA, UC, UG и UT;

o представляет собой 0 или 1;

s представляет собой 0 или 1; и

t представляет собой целое число от 0 до 6, т.е. 0, 1, 2, 3, 4, 5 или 6.

при этом предпочтительно Z1 и/или Z2 выбран из группы, состоящей из SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19 и SEQ ID NO: 20.

4. Способ согласно любому из пп.1-3, отличающийся тем, что указанный олигонуклеотид, переключающий матрицу,

(i) представлен формулой

$5'-X_p-Y-Q_t-Z_q-A_t-3'$ где

X представляет собой химическую группу, выбранную из группы, состоящей из остатка амина, биотина, глицерина, холестерина, дигоксигенина, фтора или производных нуклеотида, включая нуклеотиды без азотистых оснований, дидезоксирибонуклеотиды, 3'-дезоксинуклеотиды, 2'-дезоксининозин, 2'-дезоксинуридин;

Y представляет собой известную олигонуклеотидную последовательность;

Q представляет собой последовательность идущих друг за другом вырожденных (неоднозначных) оснований ДНК, предпочтительно выбранных из N, V, H, D, B и J, где N представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT, dC и dG; B представляет собой продукт включения нуклеотида из эквимольной смеси dT, dC и dG; D представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dG; H представляет собой продукт включения нуклеотида из эквимольной смеси dA, dT и dC; V представляет собой продукт включения нуклеотида из эквимольной смеси dA, dC и dG; J представляет собой продукт включения нуклеотида из смеси (0-100% dA) с (0-100% dG), с (0-100% dC), с (0-100% dT), с (0-100% dU), с (0-100% rA), с (0-100% rG), с (0-100% rC), с (0-100% rT), с (0-100% rU);

Z представляет собой рибонуклеотид, выбранный из группы, состоящей из AMP, CMP, GMP, TMP и UMP;

A представляет собой химическую группу, выбранную из группы, состоящей из остатка амина, биотина, глицерина, холестерина, дигоксигенина, фосфата, фтора или производных нуклеотида, включая нуклеотиды без азотистых оснований, дидезоксирибонуклеотиды, 3'-дезоксинуклеотиды, 2'-дезоксининозин, 2'-дезоксинуридин;

t представляет собой целое число от 0 до 6, т.е. 0, 1, 2, 3, 4, 5 или 6;

p представляет собой целое число от 0 до 10, т.е. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10;

q представляет собой целое число, равное по меньшей мере 1; и

r представляет собой целое число от 0 до 10, т.е. 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 или 10, и/или

(ii) выбран из группы, состоящей из SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19 и SEQ ID NO: 20.

5. Способ согласно любому из пп.1-4, отличающийся тем, что указанная обратная транскриптаза

(i) обладает активностью ДНК-полимеразы и активностью концевой трансферазы и/или

(ii) способна переключаться на матрицу, содержащую двухцепочечную нуклеиновую кислоту ДНК/РНК и/или ДНК/ДНК.

6. Способ согласно любому из пп.1-5, дополнительно включающий этап иммобилизации синтезированной двухцепочечной нуклеиновой кислоты на поверхности.

7. Способ согласно любому из пп.1-6, который дополнительно включает этап f), включающий отжиг с двухцепочечной нуклеиновой кислотой, синтезированной на этапе e), по меньшей мере одного праймера, идентичного, комплементарного или способного гибридизоваться по меньшей мере с частью указанного иницирующего олигонуклеотида или олигонуклеотида, переключающего матрицу, и осуществление ПЦР-амплификации, и предпочтительно включающий этап иммобилизации продукта ПЦР-амплификации и/или предпочтительно дополнительно включающий этап определения по меньшей мере части последовательности указанной одноцепочечной нуклеиновой кислоты.

8. Иницирующий олигонуклеотид, содержащий следующие элементы последовательности:

$3'-W_m-X-Y_n-Z1-Q_t-Z2-5'$, где

лярной смеси dA, dT и dC; V представляет собой продукт включения нуклеотида из эквимольной смеси dA, dC и dG; J представляет собой продукт включения нуклеотида из смеси (0-100% dA) с (0-100% dG), с (0-100% dC), с (0-100% dT), с (0-100% dU), с (0-100% rA), с (0-100% rG), с (0-100% rC), с (0-100% rT), с (0-100% rU); и

t представляет собой целое число от 1 до 6, т.е. 1, 2, 3, 4, 5 или 6.

12. Нуклеиновая кислота, содержащая иницирующий олигонуклеотид по п.8 или его компоненты.

13. Набор, содержащий:

а) реагент, способный добавлять нуклеотиды к 3'-концу одноцепочечной нуклеиновой кислоты, предпочтительно независимую от матрицы ДНК- или РНК-полимеразу, более предпочтительно поли(А)-полимеразу или концевую трансферазу, и блокирующий нуклеотид, предпочтительно 3d-NTP, 3-Me-NTP и ddNTP;

б) фермент обратную транскриптазу;

с) иницирующий олигонуклеотид согласно п.8 и

д) олигонуклеотид, переключаящий матрицу,

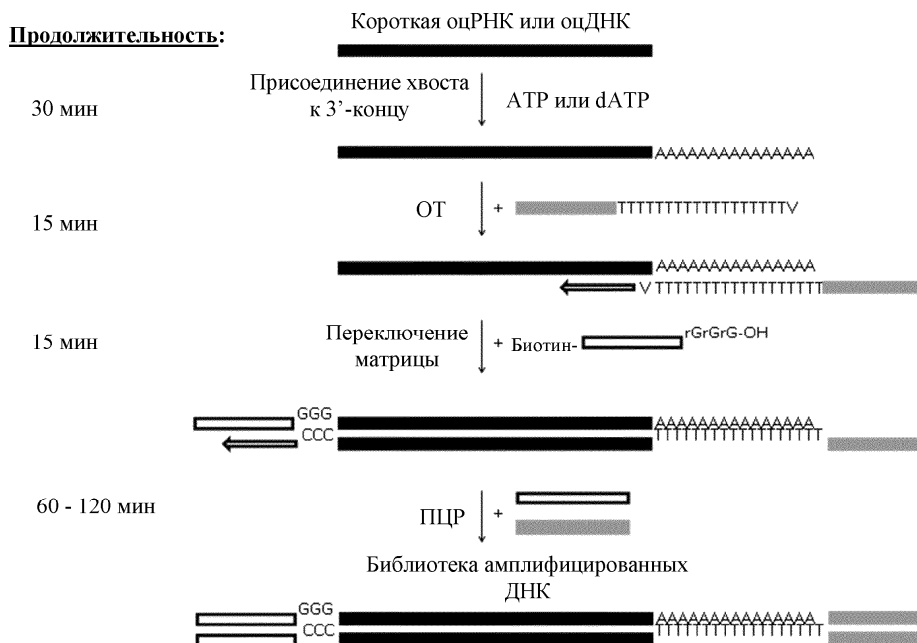
предпочтительно дополнительно содержащий реагенты для ПЦР-амплификации и/или целевого обогащения мишенью и/или предпочтительно дополнительно содержащий фермент, способный расщеплять пептидные связи в белках.

14. Чип, содержащий по меньшей мере одну нуклеиновую кислоту согласно п.12.

15. Применение набора согласно п.13 для получения двухцепочечной ДНК, которую можно напрямую секвенировать.

16. Применение двухцепочечной нуклеиновой кислоты, синтезированной согласно способу по любому из пп.1-7, или одноцепочечной нуклеиновой кислоты, полученной из нее, для секвенирования или анализа экспрессии, клонирования, меченая, для идентификации генов или определенных последовательностей нуклеотидов.

17. Применение двухцепочечной нуклеиновой кислоты, синтезированной согласно способу по любому из пп.1-7, или одноцепочечной нуклеиновой кислоты, полученной из нее, в персонализированной медицине; для мониторинга терапии; предсказания, прогнозирования, раннего обнаружения заболевания человека или животного, или для судебной медицины, для анализа последовательностей нуклеиновых кислот вирусов, бактерий, грибов, животных или растений или клеток, полученных из них, предпочтительно для характеристики растений, плодов, проверки скрещивания, обнаружения заболеваний растений или плодов.



Фиг. 1

A

Праймер для секвенирования по Сэнгеру
aatataccgacccaccacccg

Прямой праймер для ПЦР

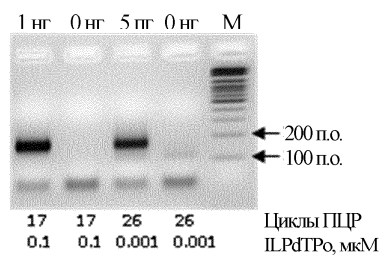
aatgatacggcgaaccaccgaatctacacgttcagaggttctacagtcgaa

[illegible]

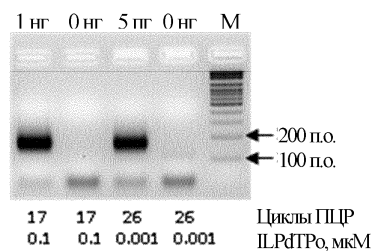
Фиг. 2А

B

cel-miR-39 (синтетическая 22 н.т. РНК): ucaccggguguaaaucagcuug



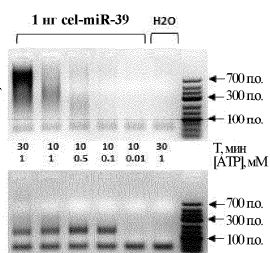
cel-miR-39 (синтетическая 22 н.т. ДНК): tcaccgggtgtaaatacagcttg



Фиг. 2В

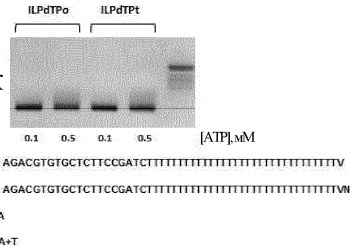
A

Поли(А)-хвост



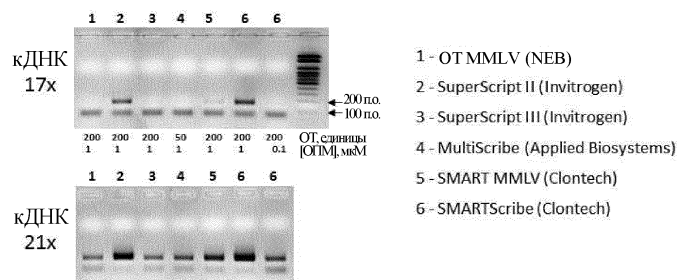
B

кДНК

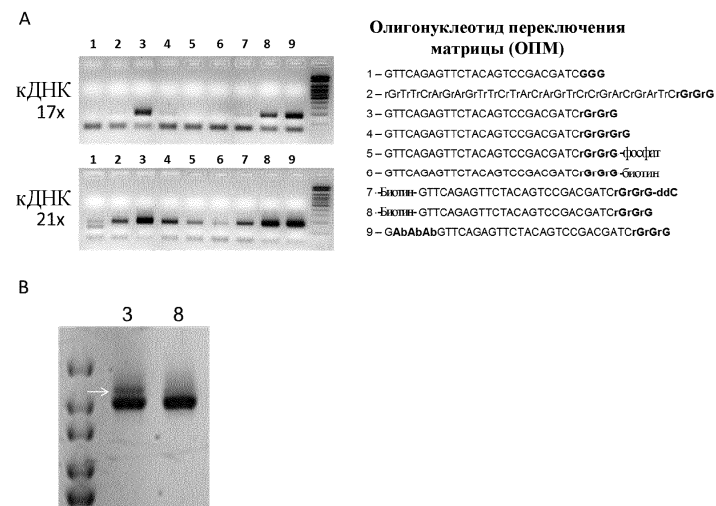


C

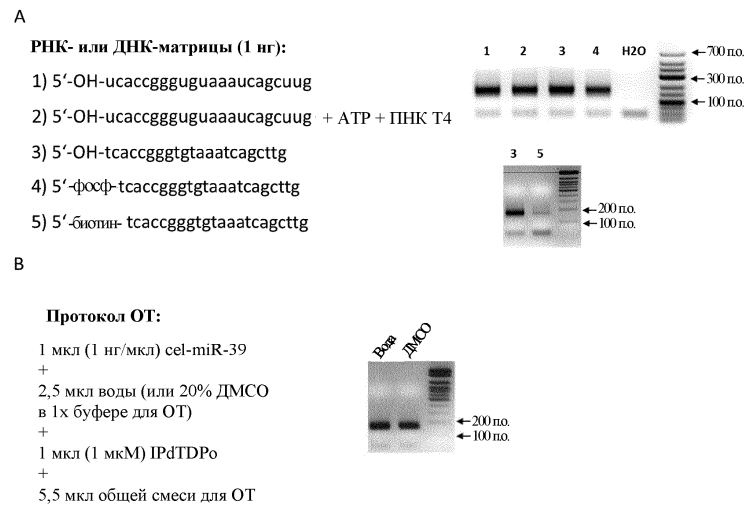
Обратные транскриптазы:



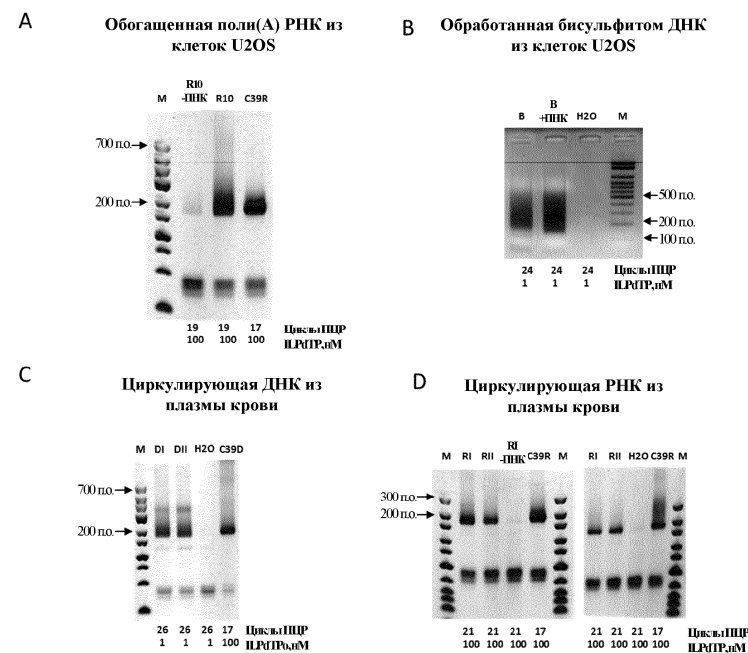
ФИГ. 3



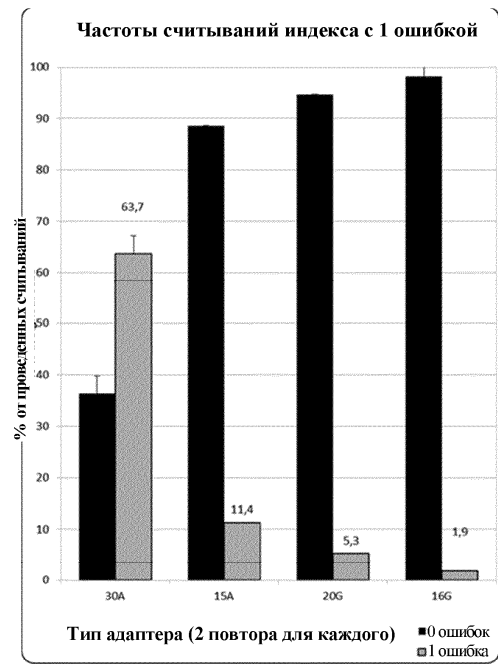
Фиг. 4



Фиг. 5

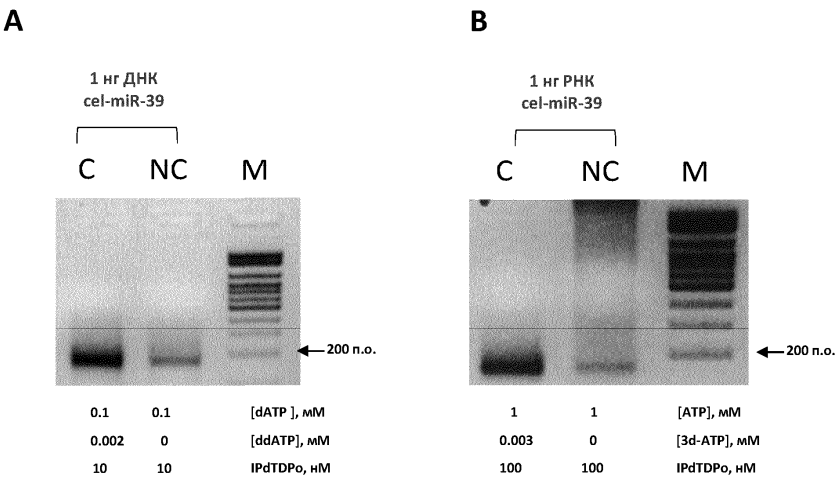


Фиг. 6



30A: AGACGTGTGCTCTCCGATCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
15A: AGACGTGTGCTCTCCGATCTTTTTTTTTTTTTTTTTTTTTTTTTT
20G: AGACGTGTGCTCTCCGATCTTTTTTTTTTTGTTTTTTTTTT
16G: AGACGTGTGCTCTCCGATCTTTTTTTTTGTTTTTTTTTT

Фиг. 7



Фиг. 8

