



(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ

(45) Дата публикации и выдачи патента
2021.03.09

(51) Int. Cl. **C12N 15/09** (2006.01)
C12Q 1/68 (2006.01)

(21) Номер заявки
201891697

(22) Дата подачи заявки
2017.01.13

**(54) СПОСОБ ДЕКОДИРОВАНИЯ ПОСЛЕДОВАТЕЛЬНОСТИ ОСНОВАНИЙ
НУКЛЕИНОВОЙ КИСЛОТЫ, СООТВЕТСТВУЮЩЕЙ КОНЦЕВОМУ УЧАСТКУ РНК,
И СПОСОБ АНАЛИЗА ЭЛЕМЕНТА ДНК**

(31) **2016-013575**

(32) **2016.01.27**

(33) **JP**

(43) **2019.01.31**

(86) **PCT/JP2017/001097**

(87) **WO 2017/130750 2017.08.03**

(71)(73) Заявитель и патентовладелец:
**КАБУСИКИ КАЙСЯ ДиЭнЭйФОРМ
(JP)**

(72) Изобретатель:
**Муракава Ясухиро, Такегами Юдзиро
(JP)**

(74) Представитель:
Медведев В.Н. (RU)

(56) CORE L.J. et al., Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Genet.*, 2014, vol.46, no.12, pp.1311-1320 (Abstract, Figures)
MAGNUSON B. et al., Identifying transcription start sites and active enhancer elements

using BruUV-seq. *Sci. Rep.*, 2015, no.5:17978, pp.1-12 (Abstract, Figures)

ANDERSSON R. et al., An atlas of active enhancers across human cell types and tissues. *Nature*, 2014, vol.507, no.7493, pp.455-461 (Abstract, Figures)

LAM M.T.Y. et al., Enhancer RNAs and regulated transcriptional programs. *Trends Biochem. Sci.*, 2014, vol.39, no.4, pp.170-182 (Figure 3, lowest part)

MURATA M. et al., Detecting expressed genes using CAGE. *Methods Mol. Biol.*, 2014, vol.1164, pp.67-85 (Abstract, Fig.)

KODZIUS R. et al., CAGE: cap analysis of gene expression. *Nat. Methods*, 2006, vol.3, no.3, pp.211-222 (Abstract, Figure 1)

PLESSY C. et al., Linking promoters to functional transcripts in small samples with nanoCAGE and CAGEscan. *Nat. Methods*, 2010, vol.7, no.7, pp.528-534 (Abstract, Figures)

MURAKAWA Y. et al., Enhanced identification of transcriptional enhancers provides mechanistic insights into diseases. *Trends Genet.*, 2016 Jan. 15 (online), vol.32, no.2, pp.76-88 (Abstract, Table 1, Figures)

(57) Изобретение предлагает способ анализа, который позволяет идентифицировать элемент ДНК с повышенной чувствительностью и определить уровень транскрипции элемента ДНК. Изобретение относится к способу декодирования последовательности оснований нуклеиновой кислоты, соответствующей концевому участку РНК, включающему стадию получения-декодирования способа получения с использованием матричной РНК по меньшей мере одной нуклеиновой кислоты, выбранной из группы, состоящей из нуклеиновой кислоты РНК 5'-концевого участка матричной РНК или комплементарной нуклеиновой кислоты ДНК, соответствующей 5'-концевому участку матричной РНК; нуклеиновой кислоты РНК 3'-концевого участка матричной РНК или комплементарной нуклеиновой кислоты ДНК, соответствующей 3'-концевому участку матричной РНК; нуклеиновой кислоты РНК частичного участка антисмысловой цепи или смысловой цепи полноразмерной матричной РНК и их сочетаний, или нуклеиновой кислоты, последовательность которой содержит две или более связанных последовательности, где матричная РНК, используемая на стадии получения-декодирования, представляет собой растущую цепь РНК. Элемент ДНК можно идентифицировать и проанализировать с повышенной точностью и повышенной чувствительностью путем определения информации о декодированной последовательности нуклеиновой кислоты и картирования элемента ДНК в геноме.

Область техники

Настоящее изобретение относится к способу декодирования последовательности оснований нуклеиновой кислоты, соответствующей концевому участку РНК, и к способу анализа элемента ДНК.

Уровень техники изобретения

Предполагается, что биологические события, связанные с заболеваниями, включают изменение контроля экспрессии генов. Для объяснения изменения важную роль играет анализ профиля экспрессии генов. Профиль экспрессии генов включает в себе информацию, например, об экспрессирующихся генах, участках (таких как клетки, ткани) биологических организмов, в которых происходит экспрессия генов, о времени, в течение которого происходит экспрессия генов, и о характере экспрессии генов.

В качестве способа анализа профиля экспрессии в настоящее время используют CAGE (кэп-анализ экспрессии генов) (патентный документ 1, непатентный документ 1). CAGE представляет собой способ, в котором каждое основание 5'-концевого участка РНК, транскрибированной с гена (ДНК), идентифицируют по всему геному и количественно определяют с использованием кэп-структуры на кэпированном 5'-конце. А именно 5'-концевой участок транскрибированной РНК извлекают в виде кДНК-маркера, полученного путем обратной транскрипции 5'-концевого участка, и затем определяют последовательность полученного кДНК-маркера, что позволяет картировать в геноме функциональный элемент ДНК, такой как промотор. Кроме того, путем определения и картирования последовательности перекрывающего маркера кДНК можно измерить уровень экспрессии каждого элемента ДНК исходя из числа маркеров, картированных в одном положении. Как описано выше, CAGE позволяет анализировать профили экспрессии по всему геному и, следовательно, может использоваться, например, для выяснения причин заболеваний, в способах диагностики заболеваний, а также в способах лечения заболеваний.

Целью анализа вышеупомянутого профиля экспрессии является энхансер как элемент ДНК. Энхансер присутствует в вышестоящем участке, нижестоящем участке, или внутри гена, и представляет собой цис-регуляторный участок ДНК длиной в несколько сотен оснований, действие которого заключается в повышении экспрессии гена. Кроме того, отмечено, что энхансер участвует в развитии разных заболеваний. Следовательно, действительно важно анализировать профиль экспрессии энхансера, который, как полагают, играет важную роль в биологических событиях.

Что касается энхансеров, описано, что РНК (далее называемую "энхансерная РНК") синтезируют в двух направлениях с обоих концов энхансера с использованием общего фактора транскрипции или РНК-полимеразы. Кроме того, обнаружено, что энхансерная РНК содержит 5'-концевую кэпирующую структуру, как и мРНК. Таким образом, разработан анализ энхансера методом CAGE с использованием 5'-концевой кэпирующей структуры, идентифицирован участок начала транскрипции энхансерной РНК и идентифицированы примерно 44000 участков энхансера (далее называемых "активный энхансер"), транскрибированных с использованием идентифицированного участка начала транскрипции (непатентный документ 2).

Список литературы.

Патентная литература.

Патентный документ 1: JP 2005-535311 A.

Непатентная литература.

Непатентный документ 1: Shiraki et al., Proc. Natl. Acad. Sci. USA 100, 15776-15781 (2003).

Непатентный документ 2: Andersson et al., Nature 507 (7493) 455-461 (2014).

Сущность изобретения

Техническая проблема.

Однако авторы настоящего изобретения обнаружили новую проблему в анализе энхансеров методом CAGE.

Во-первых, энхансерная РНК, представляющая собой РНК-продукт энхансера, в процессе синтеза (транскрипции) подвергается активной деградации. Поэтому в случае применения образца, полученного из штамма, выращенного из одной клетки индивидуума, подлежащая анализу энхансерная РНК представляет собой только небольшую часть полноразмерной энхансерной РНК, транскрибированной в клетке, что затрудняет проведение чувствительной идентификации. И наоборот, в непатентном документе 2 описано, что при идентификации участка начала транскрипции с использованием общей РНК клетки для энхансера накапливают примерно 1000 видов огромного количества данных, полученных методом CAGE из клеток и тканей, и в результате всеобъемлющего анализа указанных накопленных данных идентифицируют десятки тысяч участков, таких как указанные выше. Однако, поскольку используют огромное количество данных, полученных методом CAGE из разных клеток и тканей, чувствительность идентификации "участка", в котором экспрессируется энхансерная РНК, которая занимает важное место в профиле экспрессии, может быть недостаточной. А именно, поскольку в области медицины полагают, что может существовать связь между мутацией энхансера и развитием рака у индивидуума, страдающего от рака, важно идентифицировать энхансер на основе накопленных данных, полученных методом CAGE не из разных клеток и тканей, а только из одного образца.

Синтез и деградация энхансерной РНК происходят параллельно. Поэтому уровень экспрессии деградируемой энхансерной РНК представляет собой лишь кажущийся уровень экспрессии, являющийся

результатом равновесия между количеством синтезированной РНК и количеством деградированной РНК. Следовательно, невозможно определить, например, изменение уровня экспрессии энхансерной РНК с течением времени и абсолютное значение уровня экспрессии. Однако при анализе профиля экспрессии важно знать абсолютное количество синтезированной энхансерной РНК, чтобы определить активность энхансера.

Данная проблема касается не только энхансера, но и, например, многих не кодирующих белки РНК (нкРНК), которые активно подвергаются деградации, как и энхансер, а также РНК, кодирующих белки, известные как регуляторные факторы, такие как фактор транскрипции.

Таким образом, целью настоящего изобретения является разработка способа кодирования последовательности оснований нуклеиновой кислоты, соответствующей концевому участку РНК, и способа анализа элемента ДНК с целью идентификации, например, функционального элемента ДНК, такого как энхансер, с повышенной чувствительностью, и количественного анализа активации функционального элемента ДНК с повышенной точностью.

Решение проблемы.

Для достижения вышеуказанной цели настоящее изобретение предлагает способ декодирования последовательности оснований нуклеиновой кислоты, соответствующей концевому участку РНК (далее называемый "способ декодирования согласно настоящему изобретению"), включающий стадию получения-декодирования способа получения с использованием матричной РНК по меньшей мере одной нуклеиновой кислоты, выбранной из группы, состоящей из нуклеиновой кислоты РНК 5'-концевого участка матричной РНК или комплементарной нуклеиновой кислоты ДНК, соответствующей 5'-концевому участку матричной РНК; нуклеиновой кислоты РНК 3'-концевого участка матричной РНК или комплементарной нуклеиновой кислоты ДНК, соответствующей 3'-концевому участку матричной РНК; нуклеиновой кислоты РНК частичного участка смысловой цепи или антисмысловой цепи полноразмерной матричной РНК и их сочетаний, или нуклеиновой кислоты, последовательность которой содержит две или более из указанных последовательностей, связанные между собой, где матричная РНК, используемая на стадии получения-декодирования, представляет собой растущую цепь РНК.

Настоящее изобретение также предлагает способ анализа элемента ДНК (далее также называемый "способ анализа согласно настоящему изобретению"), включающий стадию декодирования последовательности оснований нуклеиновой кислоты, соответствующей концевому участку РНК с помощью способа декодирования согласно настоящему изобретению, стадию декодирования последовательности оснований нуклеиновой кислоты, соответствующей концевому участку РНК, и стадию картирования элемента ДНК, который кодирует синтез РНК, в геномной последовательности на основе информации о декодированной последовательности.

Преимущества настоящего изобретения

Настоящее изобретение позволяет идентифицировать с повышенной чувствительностью, например, функциональный элемент ДНК, такой как энхансер, и определить транскрипционное количество функционального элемента ДНК.

Краткое описание чертежей

На фиг. 1 приведено схематическое изображение, демонстрирующее типичные стадии согласно настоящему изобретению.

На фиг. 2 приведено схематическое изображение, демонстрирующее типичные стадии клонирования 5'-концевых специфических маркеров для конкатемера.

На фиг. 3 приведено схематическое изображение, демонстрирующее типичные стадии прямого секвенирования 5'-концевых маркеров.

На фиг. 4 показаны результаты вестерн-блоттинга, полученные в примере 1.

На фиг. 5 приведен график, демонстрирующий число считываний по результатам примера 2.

На фиг. 6 приведен график, демонстрирующий воспроизводимость числа считываний по результатам примера 3.

На фиг. 7 приведено изображение, демонстрирующее картирование энхансеров-кандидатов по результатам примера 4.

Описание вариантов осуществления

Например, на стадии получения-декодирования способа декодирования согласно настоящему изобретению получают нуклеиновую кислоту, которая включает по меньшей мере одну из нуклеиновой кислоты РНК 5'-концевого участка матричной РНК и комплементарной нуклеиновой кислоты ДНК, соответствующей 5'-концевому участку матричной РНК, используя метод, основанный на способе pAnT-iCAGE.

Например, на стадии получения-декодирования способа декодирования preparing-decoding step of the decoding method согласно настоящему изобретению получают нуклеиновую кислоту, которая включает по меньшей мере одну из нуклеиновой кислоты РНК 5'-концевого участка матричной РНК и комплементарной нуклеиновой кислоты ДНК, соответствующей 5'-концевому участку матричной РНК, причем стадия получения-декодирования включает стадию получения нуклеиновой кислоты с добавлением линкера, которая содержит линкер, связанный с нуклеиновой кислотой; стадию расщепления нуклеино-

вой кислоты, содержащей линкер, с помощью рестрикционного фермента, имеющего участок распознавания в линкере и участок расщепления в нуклеиновой кислоте, и стадию собирания фрагмента, соответствующего 5'-концевому участку матричной РНК, из полученного продукта расщепления.

Например, на стадии получения-декодирования способа декодирования согласно настоящему изобретению получают нуклеиновую кислоту, которая включает по меньшей мере одну из нуклеиновой кислоты РНК 5'-концевого участка матричной РНК и комплементарной нуклеиновой кислоты ДНК, соответствующей 5'-концевому участку матричной РНК, используя метод, основанный на способе с переключением матрицы.

Например, на стадии получения-декодирования способа декодирования согласно настоящему изобретению получают нуклеиновую кислоту, которая включает по меньшей мере одну из нуклеиновой кислоты РНК 5'-концевого участка матричной РНК и комплементарной нуклеиновой кислоты ДНК, соответствующей 5'-концевому участку матричной РНК, используя метод, основанный на способе HeliScoreCAGE.

Например, в способе декодирования согласно настоящему изобретению растущую цепь РНК получают из штамма, выращенного из одной клетки, или из одной ткани.

Например, в способе декодирования согласно настоящему изобретению растущую цепь РНК выделяют из комплекса, содержащего ДНК-матрицу и РНК-полимеразу II в процессе транскрипции и РНК в процессе синтеза.

Например, способ декодирования согласно настоящему изобретению дополнительно включает в себя стадию выделения растущей цепи РНК из биологического образца, причем выделенную растущую цепь РНК используют в качестве матричной РНК.

Например, на стадии выделения способа декодирования согласно настоящему изобретению хроматин, содержащий комплекс ДНК-матрицы и РНК-полимеразы II в процессе транскрипции и РНК в процессе синтеза, выделяют из биологического образца, а содержащуюся в комплексе и находящуюся в процессе синтеза РНК выделяют из хроматина в виде растущей цепи РНК.

Например, на стадии картирования способа анализа согласно настоящему изобретению элемент ДНК идентифицируют путем картирования.

Например, способ анализа согласно настоящему изобретению дополнительно включает в себя стадию определения уровня активности элемента ДНК из ряда перекрывающихся участков картируемого фрагмента ДНК.

Например, в способе анализа согласно настоящему изобретению определяемый уровень активности элемента ДНК связан по меньшей мере с одной мутацией или с одним полиморфизмом в геномной ДНК.

Далее настоящее изобретение описывается в деталях. Однако настоящее изобретение никоим образом не ограничивается приведенным описанием.

Способ декодирования.

Способ декодирования согласно настоящему изобретению представляет собой способ декодирования последовательности оснований нуклеиновой кислоты, соответствующей концевому участку РНК, включающий стадию получения-декодирования способа получения с использованием матричной РНК по меньшей мере одной нуклеиновой кислоты, выбранной из группы, состоящей из нуклеиновой кислоты РНК 5'-концевого участка матричной РНК или комплементарной нуклеиновой кислоты ДНК, соответствующей 5'-концевому участку матричной РНК; нуклеиновой кислоты РНК 3'-концевого участка матричной РНК или комплементарной нуклеиновой кислоты ДНК, соответствующей 3'-концевому участку матричной РНК; нуклеиновой кислоты РНК частичного участка антисмысловой цепи или смысловой цепи полноразмерной матричной РНК и их сочетаний, или нуклеиновой кислоты, последовательность которой содержит две или более из указанных последовательностей, связанные между собой, где матричная РНК, используемая на стадии получения-декодирования, представляет собой растущую цепь РНК. В настоящем изобретении получаемую нуклеиновую кислоту иногда называют, например, фрагмент нуклеиновой кислоты или фрагмент, и если получаемая нуклеиновая кислота представляет собой нуклеиновую кислоту ДНК, нуклеиновую кислоту иногда называют фрагмент ДНК.

А именно в способе декодирования согласно настоящему изобретению получают нуклеиновую кислоту с использованием растущей цепи РНК, т.е. РНК в процессе нового синтеза посредством транскрипции, в качестве матричной РНК, и декодируют последовательность нуклеиновой кислоты. Таким образом, способ характеризуется нижеследующими эффектами в анализе элемента ДНК.

Во-первых, как указано выше, например, в случае энхансера, энхансерная РНК после завершения синтеза активно деградирует. Однако в способе декодирования согласно настоящему изобретению в качестве матрицы используют не РНК после завершения синтеза, а растущую цепь РНК в процессе нового синтеза. Таким образом, можно получить, например, не нуклеиновую кислоту, соответствующую РНК, характеризующейся равновесным состоянием между синтезом и деградацией, и, следовательно, присутствующую в кажущемся количестве, а нуклеиновую кислоту, соответствующую РНК до деградации, т.е. нуклеиновую кислоту, соответствующую РНК, которая не является кажущейся, но реально синтезируется. Следовательно, например, анализ элемента ДНК, такого как энхансер, в штамме, выращенном из одной клетки или из одной ткани, идентификацию можно проводить с повышенной чувствительностью.

Можно получить нуклеиновую кислоту, соответствующую реально синтезируемой РНК. Можно измерить не кажущееся количество РНК, отражающее состояние равновесия между синтезом и деградацией, а количество синтезируемой РНК. Таким образом, практически можно осуществить определение абсолютного количества. Следовательно, например, в случае элемента ДНК, такого как энхансер, можно косвенно определить абсолютное количество активности элемента ДНК, такой как энхансерная активность, путем определения абсолютного количества транскрибированной РНК.

Как описано выше, настоящее изобретение позволяет провести с повышенной чувствительностью и повышенной точностью идентификацию неизвестного элемента ДНК и количественное определение активности неизвестного элемента ДНК или идентификацию элемента ДНК в штамме, полученном из одной клетки, или в одной ткани, и количественное определение активности элемента.

Термин "растущая цепь РНК" обозначает РНК в процессе нового синтеза путем транскрипции с использованием ДНК в качестве матрицы.

Растущая цепь РНК представляет собой, например, РНК, входящую в состав комплекса, содержащего ДНК-матрицу и РНК-полимеразу II в процессе транскрипции и РНК в процессе синтеза, причем можно использовать РНК, выделенную из комплекса (РНК в процессе синтеза).

Растущая цепь РНК представляет собой, например, РНК, содержащую 5'-концевую экзистующую структуру (далее также называемую "5'-экзистующая структура"), образующуюся в результате экзистирования 5'-конца в процессе транскрипции.

Способ получения растущей цепи РНК не ограничивается конкретными способами. В качестве примера можно использовать способ, описанный, например, в Jerome Wuarin et al. (Molecular and Cellular Biology, Nov. 1994, p. 7217-7225).

Образец для получения растущей цепи РНК представляет собой, например, биологический образец. Вид биологического образца конкретно не ограничивается, а его примеры включают клетки и ткани. Клетки могут представлять собой клетки одного вида или смесь клеток двух или более видов. Ткани могут представлять собой, например, ткани одного вида или смесь тканей двух или более видов. Настоящее изобретение также можно использовать для анализа элемента ДНК в штамме, выращенном из одной клетки, или в одной ткани, поскольку указанная выше деградация РНК не влияет на фрагмент ДНК. Таким образом, в настоящем изобретении в качестве биологического образца предпочтительно используют, например, клетки одного вида (штамм, выращенный из одной клетки) или ткань одного вида (одна ткань).

Вид клеток конкретно не ограничивается и можно использовать клетки любого вида. С помощью способа анализа согласно настоящему изобретению с использованием фрагмента ДНК, полученного по способу декодирования согласно настоящему изобретению можно анализировать, например, элемент ДНК, участвующий в развитии заболеваний. Конкретные примеры видов клеток включают в себя клетки крови, лимфобластоидные клетки и опухолевые клетки.

Растущую цепь РНК можно получить, например, с помощью описанного ниже способа. Вначале традиционным способом от клеток отделяют ядерную фракцию, содержащую ядра. Затем ядерную фракцию денатурируют, чтобы отделить фракцию хроматина. Фракция хроматина содержит комплекс, содержащий ДНК-матрицу и РНК-полимеразу II в процессе транскрипции и РНК в процессе синтеза. Комплекс является реально стабильным и, следовательно, его структура сохраняется после денатурации. Для денатурации можно использовать буфер, содержащий, например, соль и мочевины в высоких концентрациях. Буфер для денатурации может иметь следующий состав: соль 300 ммоль/л, мочевина 4 моль/л, поверхностно-активное вещество 1%. Соль может представлять собой, например, хлорид натрия, а поверхностно-активное вещество может представлять собой, например, неионное поверхностно-активное вещество, такое как NP-40 (нониidet P-40). Затем из фракции хроматина экстрагируют РНК, чтобы собрать содержащуюся в комплексе РНК, находящуюся в процессе синтеза. Собранная РНК представляет собой растущую цепь РНК. Для экстракции РНК можно использовать, например, коммерчески доступный реагент для экстракции РНК, а в качестве коммерчески доступного реагента для экстракции РНК можно использовать, например, реагент для выделения РНК Trizol (зарегистрированная торговая марка).

Способ декодирования согласно настоящему изобретению может также включать, например, стадию выделения способа выделения растущей цепи РНК из биологического образца. Растущую цепь РНК, выделенную на стадии выделения, можно использовать в качестве матричной РНК. Стадию выделения можно проводить, например, путем отделения хроматина, содержащего комплекс ДНК-матрицы и РНК-полимеразы II в процессе транскрипции и РНК в процессе синтеза, от биологического образца и выделения в качестве растущей цепи РНК, содержащейся в присутствующем в хроматине комплексе РНК, находящейся в процессе синтеза. В качестве способа выделения растущей цепи РНК можно использовать, например, вышеупомянутый способ получения растущей цепи РНК.

В способе декодирования согласно настоящему изобретению получаемая нуклеиновая кислота включает нижеследующие нуклеиновые кислоты (1)-(4) и может представлять собой нижеследующую нуклеиновую кислоту (5):

(1) нуклеиновая кислота РНК 5'-концевого участка РНК или комплементарная нуклеиновая кислота ДНК, соответствующая 5'-концевому участку РНК;

(2) нуклеиновая кислота РНК 3'-концевого участка РНК или комплементарная нуклеиновая кислота ДНК, соответствующая 3'-концевому участку РНК;

(3) нуклеиновая кислота РНК частичного участка антисмысловой цепи или смысловой цепи полно-размерной РНК;

(4) любое сочетание нуклеиновых кислот (1)-(4);

(5) нуклеиновая кислота, последовательность которой получена путем соединения по меньшей мере двух последовательностей друг с другом.

В нуклеиновой кислоте (5) "последовательности" из "по меньшей мере двух последовательностей" представляют собой по меньшей мере две последовательности из последовательностей нуклеиновых кислот (1)-(4), а "последовательность, полученная путем соединения по меньшей мере двух последовательностей друг с другом" представляет собой последовательность, полученную путем соединения по меньшей мере двух последовательностей из последовательностей нуклеиновых кислот (1)-(4) друг с другом. Соединение может отличаться от непосредственного соединения и может представлять собой косвенное соединение, например, через линкер.

Если нуклеиновая кислота, получаемая на стадии получения-декодирования способа декодирования согласно настоящему изобретению, включает нуклеиновую кислоту РНК 5'-концевого участка матричной РНК или комплементарную нуклеиновую кислоту ДНК, соответствующую 5'-концевому участку матричной РНК, на стадии получения-декодирования можно использовать разные способы, например, на основе CAGE. Нуклеиновая кислота может представлять собой, например, нуклеиновую кислоту РНК 5'-концевого участка матричной РНК или комплементарную нуклеиновую кислоту ДНК, соответствующую 5'-концевому участку РНК, или она может включать и нуклеиновую кислоту РНК 5'-концевого участка матричной РНК, и комплементарную нуклеиновую кислоту ДНК, соответствующую 5'-концевому участку РНК.

В данном случае нуклеиновая кислота, полученная с использованием растущей цепи РНК в качестве матрицы не является нуклеиновой кислотой, соответствующей 3'-концевому участку растущей цепи РНК, а представляет собой фрагмент ДНК, соответствующий 5'-концевому участку растущей цепи РНК. Таким образом, например, можно получить информацию о последовательности фрагмента ДНК, информацию о последовательности 5'-концевого участка растущей цепи РНК, т.е. информацию о последовательности от участка начала транскрипции. Следовательно, можно осуществить с повышенной точностью, например, идентификацию элемента ДНК по всему геному, т.е. идентификацию положения элемента ДНК в геноме. Соответственно можно идентифицировать, например, функциональный элемент ДНК, такой как неизвестный энхансер.

Если нуклеиновая кислота, получаемая на стадии получения-декодирования способа декодирования согласно настоящему изобретению, включает нуклеиновую кислоту РНК 5'-концевого участка матричной РНК или комплементарную нуклеиновую кислоту ДНК, соответствующую 5'-концевому участку матричной РНК, на стадии получения-декодирования можно использовать, например, метод на основе способа nAnT-iCAGE. Способ nAnT-iCAGE не ограничивается конкретными способами. В качестве способа nAnT-iCAGE можно использовать, например, способ, описанный Murata et al., 2014 *Methods Mol Biol.* 1164 67-85.

Если нуклеиновая кислота, получаемая на стадии получения-декодирования способа декодирования согласно настоящему изобретению, включает нуклеиновую кислоту РНК 5'-концевого участка матричной РНК или комплементарную нуклеиновую кислоту ДНК, соответствующую 5'-концевому участку матричной РНК, на стадии получения-декодирования можно использовать, например, метод на основе способа с переключением матрицы. Способ с переключением матрицы (способ nano-CAGE) не ограничивается конкретными способами. В качестве способа с переключением матрицы можно использовать, например, способ, описанный Plessy et al., 2010 *Nature Methods* 7, 528-534.

Если нуклеиновая кислота, получаемая на стадии получения-декодирования способа декодирования согласно настоящему изобретению, включает нуклеиновую кислоту РНК 5'-концевого участка матричной РНК или комплементарную нуклеиновую кислоту ДНК, соответствующую 5'-концевому участку матричной РНК, на стадии получения-декодирования можно использовать, например, метод на основе способа HeliScopeCAGE. Способ HeliScopeCAGE не ограничивается конкретными способами. В качестве способа HeliScopeCAGE можно использовать, например, способ, описанный Kanamori-Katayama et al., 2011 *Genome Res.* 21 (7): 1150-1159.

Если нуклеиновая кислота, получаемая на стадии получения-декодирования способа декодирования, согласно настоящему изобретению включает по меньшей мере одну нуклеиновую кислоту из нуклеиновой кислоты РНК 5'-концевого участка РНК или комплементарной нуклеиновой кислоты ДНК, соответствующей 5'-концевому участку РНК, на стадии получения-декодирования можно использовать, например, метод на основе способа CAGE. Способ CAGE не ограничивается конкретными способами. В качестве способа CAGE можно использовать, например, способ, описанный в WO 2003/106672, JP 2005-535311 A или JP 2009-72062 A.

Стадия получения-декодирования согласно настоящему варианту осуществления на основе способа CAGE включает, например, стадию получения способа получения нуклеиновой кислоты с добавлением

линкера, которая содержит линкер, связанный с нуклеиновой кислотой; стадию расщепления нуклеиновой кислоты, содержащей линкер, с помощью рестрикционного фермента, имеющего участок распознавания в линкере и участок расщепления в нуклеиновой кислоте, и стадию собирания фрагмента, соответствующего 5'-концевому участку РНК, из полученного продукта расщепления. Если нуклеиновая кислота представляет собой нуклеиновую кислоту ДНК, соответствующую 5'-концевому участку РНК, нуклеиновая кислота с добавлением линкера представляет собой, например, нуклеиновую кислоту ДНК с добавлением линкера, а фрагмент, соответствующий 5'-концевому участку РНК представляет собой (комплементарен) фрагмент ДНК, соответствующий 5'-концевому участку РНК. Если нуклеиновая кислота представляет собой комплементарную нуклеиновую кислоту ДНК, соответствующую 5'-концевому участку РНК, нуклеиновая кислота с добавлением линкера представляет собой, например нуклеиновую кислоту РНК с добавлением линкера, а фрагмент, соответствующий 5'-концевому участку РНК представляет собой (комплементарен) фрагмент РНК, соответствующий 5'-концевому участку РНК.

Получение нуклеиновой кислоты ДНК с добавлением линкера на стадии получения не ограничивается конкретными способами получения и может включать, например, следующие способы получения:

(1) синтезируют нуклеиновую кислоту ДНК, соответствующую матричной РНК, и к нуклеиновой кислоте ДНК присоединяют линкер с получением нуклеиновой кислоты ДНК с добавлением линкера;

(2) 5'-концевую кэпирующую структуру матричной РНК заменяют на олигонуклеотид, соответствующий линкеру, синтезируют нуклеиновую кислоту ДНК, которую можно использовать для получения замещенной по 5'-концу матричной РНК, с получением нуклеиновой кислоты ДНК с добавлением линкера.

Стадия получения, описанная в приведенном выше пункте (1), включает, например, нижеследующие стадии (1a) и (1b):

(1a) стадию получения нуклеиновой кислоты ДНК, соответствующей нуклеотидной последовательности 5'-конца матричной РНК; и

(1b) стадию присоединения по меньшей мере одного линкера к нуклеиновой кислоте ДНК.

Линкер представляет собой последовательность, содержащую участок присоединения рестрикционного фермента, используемого на стадии расщепления. Например, участок расщепления рестрикционным ферментом присутствует, например, в нуклеиновой кислоте ДНК, соответствующей 5'-концевому участку матричной РНК, в нуклеиновой кислоте ДНК с добавлением линкера, содержащей присоединенный линкер. Таким образом, нуклеиновая кислота ДНК с добавлением линкера расщепляется рестрикционным ферментом на стадии расщепления с образованием фрагмента ДНК, соответствующего 5'-концевому участку матричной РНК.

Получение нуклеиновой кислоты ДНК на стадии (1a) можно проводить путем обратной транскрипции с использованием матричной РНК, в качестве примера и более конкретно, путем амплификации нуклеиновой кислоты ДНК, соответствующей матричной РНК, с использованием ДНК-полимеразы.

Можно использовать один вид ДНК-полимеразы или смесь двух или более видов ДНК-полимеразы. Предпочтительно ДНК-полимераза является, например, термически стабильной, а ее конкретные примеры включают полимеразу Taq, ДНК-полимеразу Pwo, ДНК-полимеразу Kod, ДНК-полимеразу Pfu, ДНК-полимеразу Vent, ДНК-полимеразу Deep Vent, ДНК-полимеразу rBST и ДНК-полимеразу Mater Amp Ampli Therm.

Для амплификации можно использовать разные реагенты, необходимые для обратной транскрипции, такие как случайные праймеры.

Например, стадия (1a) может включать стадию замены 5'-кэпирующей структуры матричной РНК на олигонуклеотид и стадию синтеза первой цепи кДНК с использованием матричной РНК в качестве матрицы с получением нуклеиновой кислоты ДНК, соответствующей нуклеотидной последовательности 5'-конца матричной РНК.

Например, стадия (1a) может включать стадию синтеза первой цепи кДНК с использованием матричной РНК в качестве матрицы с получением гибрида кДНК-РНК; стадию выбора из гибридов кДНК-РНК конкретного гибрида кДНК-РНК, содержащего 5'-кэпирующую структуру с использованием селективного связывающего вещества, способного специфически распознавать 5'-кэпирующую структуру; и стадию сбора нуклеиновой кислоты ДНК, соответствующей нуклеотидной последовательности 5'-конца матричной РНК.

Селективное связывающее вещество не ограничивается конкретными веществами, а его примеры включают кэп-связывающий белок или кэп-связывающее антитело.

Например, селективное связывающее вещество может быть связано с носителем. Носитель не ограничивается конкретными носителями, его примеры включают магнитные гранулы, агарозные гранулы, латексные гранулы, сефарозный матрикс, силикагелевый матрикс и стеклянные гранулы.

Например, стадия (1a) может включать стадию синтеза первой цепи кДНК с использованием матричной РНК в качестве матрицы с получением гибрида кДНК-РНК и стадию сбора из гибрида кДНК-РНК нуклеиновой кислоты ДНК нуклеиновой кислоты ДНК, соответствующей нуклеотидной последовательности 5'-конца матричной РНК.

Например, стадия (1a) может включать стадию синтеза первой цепи кДНК с использованием мат-

ричной РНК в качестве матрицы с получением гибрида кДНК-РНК; стадию связывания селективного связывающего вещества (первое вещество) с 5'-кэпирующей структурой с получением конъюгата с гибридом кДНК-РНК; стадию приведения гибрида кДНК-РНК в контакт с носителем, на котором иммобилизовано второе соответствующее селективное связывающее вещество (второе вещество); и стадию сбора из гибрида кДНК-РНК, иммобилизованного на носителе, нуклеиновой кислоты ДНК, соответствующей нуклеотидной последовательности 5'-концевой матричной РНК. Например, второе соответствующее селективное связывающее вещество (второе вещество) специфически связывается с селективным связывающим веществом (первое вещество). Таким образом, в результате приведения гибрида кДНК-РНК в контакт с носителем селективное связывающее вещество (первое вещество), связанное с гибридом, и соответствующее селективное связывающее вещество (второе вещество), иммобилизованное на носителе, связываются друг с другом. Соответственно гибрид иммобилизуется на носителе.

Сочетание селективного связывающего вещества (первое вещество) и соответствующего селективного связывающего вещества (второе вещество) не ограничивается конкретными сочетаниями. Сочетание может представлять собой, например, сочетание, в котором селективное связывающее вещество (первое вещество) представляет собой биотин, а соответствующее селективное связывающее вещество (второе вещество) выбрано из группы, состоящей из производных авидина, стрептавидина и биотина, которые специфически связываются с авидином, стрептавидином и биотином. Сочетание также может представлять собой, например, сочетание, в котором селективное связывающее вещество (первое вещество) представляет собой диоксигенин, а соответствующее селективное связывающее вещество (второе вещество) представляет собой антитело против диоксигенина. Носитель представляет собой, например, указанный выше носитель.

Например, стадия (1b) включает стадию связывания линкера с концевым участком нуклеиновой кислоты ДНК, соответствующим нуклеотидной последовательности 5'-конца матричной РНК; стадию синтеза второй цепи кДНК с использованием в качестве матрицы нуклеиновой кислоты ДНК, содержащей присоединенный к ней линкер (далее также называемой "первая цепь кДНК"); стадию обработки рестрикционным ферментом двухцепочечной кДНК (далее также называемой "связанная с линкером двухцепочечная кДНК") между получением первой цепи кДНК, содержащей присоединенный к ней линкер, и получением второй цепи кДНК; и стадию сбора из деградата, полученного под действием рестрикционного фермента, линкера и части кДНК, соответствующим нуклеотидной последовательности 5'-конца матричной РНК. Линкер предпочтительно содержит, например, по меньшей мере один участок распознавания ферментом, а рестрикционный фермент предпочтительно представляет собой, например, фермент, который расщепляет участок, отличающийся от участка распознавания. Во фрагменте линкер может включать, например, часть линкера или полноразмерную последовательность линкера.

На стадии (1b) линкер может включать, например, двухцепочечный олигонуклеотидный участок. В данном случае вторую цепь кДНК можно синтезировать, например, с использованием линкера.

На стадии (1b) вторую цепь кДНК можно синтезировать, например, с использованием другого олигонуклеотида, который частично или полностью соответствует линкеру.

На стадии (1b), например, линкер может быть связан с селективным связывающим веществом (первое вещество) или он может содержать селективное связывающее вещество (первое вещество) в своем составе. В данном случае стадия сбора на стадии (1b) включает, например, стадию деградации рестрикционным ферментом в контакте с носителем, на котором иммобилизовано другое соответствующее селективное связывающее вещество (второе вещество), чтобы обеспечить связывание селективного связывающего вещества (первое вещество) с соответствующим селективным связывающим веществом (второе вещество), и посредством этого сбор носителя. Фрагмент, собранный на стадии (1b) включает линкер, как указано выше. Таким образом, когда селективное связывающее вещество, входящее в состав линкера, и соответствующее селективное связывающее вещество (второе вещество), иммобилизованное на носителе, связываются друг с другом, фрагмент связывается с носителем, и фрагмент можно собрать путем сбора носителя. Примеры селективного связывающего вещества (первое вещество), соответствующего селективного связывающего вещества (второе вещество) и носителя указаны выше.

На стадии (1b) рестрикционный фермент не ограничивается конкретными ферментами, а его примеры включают рестрикционный фермент класса II и рестрикционный фермент класса III. Примеры рестрикционного фермента класса II включают рестрикционный фермент класса IIG и рестрикционный фермент класса IIS. Конкретно рестрикционный фермент может быть выбран, например, из группы, состоящей из GsuI, MmeI, BpmI, BsgI, EcoP15I и XmaJI.

Стадия получения по приведенному выше пункту (2) включает, например, нижеследующие стадии (2a), (2b) и (2c):

- (2a) стадию замены 5'-кэпирующей структуры матричной РНК на олигонуклеотид;
- (2b) стадию синтеза первой цепи кДНК с использованием матричной РНК в качестве матрицы и
- (2c) стадию синтеза второй цепи кДНК с использованием первой цепи кДНК в качестве матрицы.

На стадии получения вышеуказанного пункта (2) получают двухцепочечную кДНК первой цепи кДНК и второй цепи кДНК в виде нуклеиновой кислоты ДНК с добавлением линкера посредством стадий (2b) и (2c). На стадии расщепления указанную двухцепочечную кДНК подвергают обработке рест-

риктионным ферментом.

На стадии получения вышеуказанного пункта (2) олигонуклеотид можно называть линкером. Последовательность олигонуклеотида содержит внутри себя участок распознавания рестриктионным ферментом для использования на стадии расщепления. Первая цепь кДНК, синтезированная на стадии (2b), содержит на 3'-конце последовательность (линкер), соответствующую олигонуклеотиду, а вторая цепь кДНК, синтезированная на стадии (2с), содержит такую же последовательность (линкер), как и олигонуклеотид, причем каждая из первой цепи кДНК и второй цепи кДНК содержит участок расщепления рестриктионным ферментом в нуклеиновой кислоте ДНК, соответствующей 5'-концевому участку матричной РНК. Таким образом, на стадии расщепления двухцепочечная кДНК (нуклеиновая кислота ДНК с добавлением линкера) расщепляется рестриктионным ферментом с отщеплением фрагмента ДНК, соответствующего 5'-концевому участку матричной РНК.

В способе декодирования согласно настоящему изобретению стадия расщепления после получения представляет собой, как указано выше, стадию расщепления нуклеиновой кислоты ДНК с добавлением линкера под действием рестриктионного фермента, участок распознавания которого находится в линкере, а участок расщепления находится в нуклеиновой кислоте.

Рестриктионный фермент на стадии расщепления не ограничивается конкретными ферментами и может быть выбран по обстоятельствам в соответствии с линкером. В качестве рестриктионного фермента можно использовать, например, любой из указанных выше примеров.

В способе декодирования согласно настоящему изобретению стадия, сопровождающая стадию расщепления, представляет собой стадию сбора из полученного продукта расщепления фрагмента ДНК, соответствующего 5'-концевому участку РНК. Способ сбора фрагмента ДНК не ограничивается конкретными способами.

Длина собираемого фрагмента ДНК не ограничивается конкретными значениями длины и составляет, например, от 5 до 100 п.о., от 15 до 30 п.о., от 10 до 30 п.о.

В способе декодирования согласно настоящему изобретению собранный фрагмент ДНК можно присоединить, например, к вектору. Затем, например, можно определить последовательность фрагмента ДНК с использованием вектора, к которому присоединен фрагмент ДНК.

В способе декодирования согласно настоящему изобретению по меньшей мере два собранных фрагмента ДНК можно соединить друг с другом с получением конкатемера, после чего конкатемер можно, например, присоединить к вектору.

Способ анализа элемента ДНК.

Способ анализа согласно настоящему изобретению представляет собой, как указано выше, способ анализа элемента ДНК, включающий стадию декодирования последовательности оснований нуклеиновой кислоты, соответствующей 5'-концевому участку РНК, с помощью способа декодирования согласно настоящему изобретению и стадию картирования элемента ДНК, кодирующего синтез РНК, в геномной последовательности на основе информации по декодированной последовательности оснований.

В способе анализа согласно настоящему изобретению "анализ" включает в себя, например, результаты идентификации положения элемента ДНК в геноме и качественное или количественное определение уровня активности элемента ДНК. В способе анализа согласно настоящему изобретению "число перекрывающихся фрагментов (или уровень экспрессии) нуклеиновой кислоты" можно называть, например, "степень активации" элемента ДНК. Если РНК экспрессируется в результате активации элемента ДНК, уровень экспрессии РНК косвенно соответствует уровню активности элемента ДНК, и уровень экспрессии нуклеиновой кислоты (например, фрагмента, такого как фрагмент ДНК), получаемой с использованием в качестве матрицы РНК, также косвенно соответствует уровню активности элемента ДНК.

В настоящем изобретении "элемент ДНК", который является предметом анализа, может быть, например, известным или неизвестным. Настоящее изобретение позволяет, например, идентифицировать положение неизвестного элемента ДНК в геноме и определить уровень активности неизвестного элемента ДНК.

"Элемент ДНК" в настоящем изобретении, как указано выше, представляет собой "элемент ДНК, который кодирует синтез РНК", а именно элемент ДНК, который кодирует "синтез РНК в результате активации". Например, энхансер не генерирует РНК из самого энхансера в результате активации, но двуправлено генерирует РНК с обоих концов и представляет собой функциональный элемент ДНК, который кодирует синтез РНК, иницируемый активацией, как указано выше. Таким образом, в настоящем изобретении элемент ДНК, который кодирует синтез РНК, также можно называть элементом ДНК, который кодирует "синтез РНК, соответствующей участку, который отличается от элемента ДНК в геноме" в результате активации. Однако настоящее изобретение никоим образом не ограничивается этим, и элемент ДНК может кодировать, например, "синтез РНК, соответствующей самому элементу ДНК".

Способ анализа согласно настоящему изобретению позволяет идентифицировать элемент ДНК посредством, например, стадии картирования. Кроме того, способ анализа согласно настоящему изобретению позволяет определить, например, уровень активности элемента ДНК на основе числа перекрывающихся фрагментов при картировании фрагмента ДНК.

Ниже описаны примеры способа декодирования и способа анализа согласно настоящему изобретению. Однако настоящее изобретение не ограничивается ими.

Стадия 1.

На стадии 1 селективно собирают присутствующую в образце кДНК, содержащую участок, соответствующий 5'-концу РНК. кДНК можно синтезировать, например, с использованием РНК в качестве матрицы.

РНК, собранную из желательных клеток, тканей или организмов, можно использовать в качестве исходного вещества. Для получения РНК можно использовать, например, указанный выше способ.

Кроме того, для выделения 5'-концевых нуклеиновых кислот, соответствующих 5'-концу транскрибируемой части гена, можно использовать библиотеку полноразмерных кДНК.

Саму стадию 1 можно проводить с помощью широко известного способа. Другими словами, способы конструирования полноразмерных кДНК и способы синтеза фрагментов ДНК, по меньшей мере содержащих участок, соответствующий 5'-концу РНК, уже известны, и можно использовать любые из указанных методов. Одним из предпочтительных способов является способ с улавливанием кэпа (например, Piero Carninci et al., *Methods in Enzymology*, Vol. 303, pp. 19-44, 1999). Данный способ с улавливанием кэпа описан ниже. Однако настоящее изобретение не ограничивается применением способа с улавливанием кэпа, для обогащения или отбора кДНК можно использовать и другие подходы.

В способе с улавливанием кэпа вначале синтезируют первую цепь кДНК, используя обратную транскриптазу и РНК в качестве матрицы. Такой синтез можно проводить с помощью известного способа. В качестве затравки для синтеза кДНК можно использовать праймер олиго-dT, или, если в качестве матрицы используют мРНК, - случайный праймер. К реакционному раствору целесообразно добавить трегалозу, поскольку она повышает эффективность реакции обратной транскрипции путем стабилизации обратной транскриптазы (патент США № 6013488). Предпочтительно использовать 5-метил-dСТР вместо стандартного dСТР, поскольку он позволяет избежать внутреннего расщепления кДНК несколькими рестрикционными ферментами и в значительной мере предотвращает нежелательное расщепление рестрикционными ферментами. Кроме того, после синтеза первой цепи кДНК белки и расщепленные пептиды можно удалить путем обработки СТАВ (цетилтриметиламмония бромид) или с помощью других, более общих методов очистки кДНК.

Затем селективное связывающее вещество присоединяют к кэпирующей структуре РНК. "Селективное связывающее вещество" здесь обозначает вещество, селективно связывающееся с конкретным веществом. Такое селективное связывающее вещество предпочтительно включает биотин, но не ограничивается биотином. Кэпирующая структура, как правило, представляет собой структуру 5'-конца мРНК и может также присутствовать в энхансерной РНК, которая представляет собой транскрипт энхансера, позволяя проводить специфическую селекцию энхансерной РНК. Следовательно, если в качестве исходного субстрата используют РНК, селективное связывающее вещество связывается только с кэпирующей структурой РНК. Кроме того, селективное связывающее вещество не связывается с РНК, если кэпирующая структура на 5'-конце отсутствует. Биотин можно присоединить к кэпирующей структуре с помощью известного способа. Например, чтобы биотинилировать кэпирующую структуру, можно окислить диольную группу, присутствующую в кэпирующей структуре, путем обработки РНК окислителем, таким как NaIO_4 , и затем подвергнуть ее взаимодействию с гидразидом биотина.

Одноцепочечную РНК расщепляют, например, путем обработки РНКазой I. Альтернативно можно использовать любую другую РНКазу, способную расщеплять одноцепочечные РНК, но не гибриды кДНК-РНК, или смеси РНКаз, способные расщеплять разные последовательности одноцепочечных РНК с разной специфичностью. В гибриде кДНК-РНК, где первая цепь кДНК удлиняется до участка, соответствующего 5'-концу РНК, окружение 5'-конца РНК является одноцепочечным вследствие неспособности гибридизоваться с кДНК. Таким образом, на этой стадии гибрид расщепляется по одноцепочечному фрагменту и утрачивает кэпирующую структуру. Следовательно, после проведения данной стадии остаются только гибриды РНК-кДНК, содержащие кДНК, которая полностью удлиняется до 5'-конца РНК с сохранением кэпирующей структуры.

Получают соответствующее селективное связывающее вещество, фиксированное на носителе, способное селективно связываться с вышеуказанным селективным связывающим веществом. В настоящем описании термин "соответствующее селективное связывающее вещество" относится к веществу, способному селективно связываться с вышеуказанным селективным связывающим веществом, которое, если селективным связывающим веществом является, например, биотин, может представлять собой авидин, стрептавидин или их производные, которые могут специфически связываться с биотином или его производными. Предпочтительно носитель может представлять собой, без ограничения, магнитные гранулы, в частности магнитные гранулы из пористого стекла. Поскольку магнитные гранулы из пористого стекла с фиксированным на них стрептавидином являются коммерчески доступными, можно использовать такие коммерческие покрытые стрептавидином магнитные гранулы из пористого стекла. Вместо гранул из пористого стекла можно использовать другие материалы, такие как латексные гранулы, латексные магнитные гранулы, агарозные гранулы, полистирольные гранулы, сефарозные гранулы и т.п. Кроме того, настоящее изобретение не ограничивается применением системы биотин-авидин, можно также использо-

вать другие связывающие вещества, такие как диоксигениновый маркер, который можно присоединить к кэпирующей структуре, а диоксигенин может распознавать антитела, присоединенные к твердой основе.

Затем получают вышеуказанный гибрид РНК-кДНК с кэпирующей структурой и подвергают его взаимодействию с вышеуказанным соответствующим селективным связывающим веществом, фиксированным на носителе, чтобы обеспечить связывание селективного связывающего вещества на кэпирующей структуре с соответствующим селективным связывающим веществом на носителе, и посредством этого иммобилизацию гибрида РНК-кДНК с кэпирующей структурой на носителе. Если в качестве носителя используют магнитные гранулы, их можно быстро собрать с помощью силы магнитного поля. Между тем, чтобы предотвратить неспецифическое связывание с носителем, перед проведением этой реакции носитель предпочтительно обрабатывают большим избытком тРНК, не содержащей ДНК, что позволяет блокировать такое связывание. Другими веществами, подходящими для блокирования поверхности, являются, например, нуклеиновые кислоты или их производные, например общая РНК или олигонуклеотиды; белки, например бычий сывороточный альбумин; полисахариды, например гликоген, декстрансульфат, гепарин или другие полисахариды. Для маскировки участков неспецифического связывания можно использовать гибридные молекулы, содержащие фрагменты любых из вышеуказанных веществ.

Выше описано проведение стадии 1 по способу с улавливанием кэпа, однако можно использовать и другие способы, при условии, что с их помощью, например, можно селективно собирать кДНК, содержащую участок, комплементарный 5'-концу РНК.

Альтернативно кэп-отбору можно дефосфорилировать 5'-концы РНК фосфатазой, такой как ВАР (бактериальная щелочная фосфатаза), с последующей обработкой декэпирующим ферментом ТАР (кислая пирофосфатаза табака). Затем вместо исходной кэпирующей структуры к 5'-концу РНК можно присоединить рибонуклеотид или дезоксирибонуклеотид с помощью РНК-лигазы (Maquyama K., Sugano S. *Gene* 138, 171-4 (1994)). В данном способе, например, участок распознавания класса II или класса III можно поместить в олигонуклеотидную или рибонуклеотидную последовательность, используемую на стадии лигирования, которую помещают на 5'-конце кДНК или РНК. Затем можно использовать рестрикционный фермент класса II или класса III для расщепления кДНК с получением 5'-концевого маркера.

Помимо биотина в качестве вышеупомянутого селективного связывающего вещества можно использовать кэп-связывающий белок (Pelletier et al., *Mol Cell Biol* 1995 15:3363-71; Edery I. et al., *Mol Cell Biol* 1995 Jun; 15 (6): 3363-71) или антитело (Theissen H. et al., *EMBO J.* 1986 Dec 1; 5(12):3209-17), способное специфически связываться с кэпирующей структурой.

Альтернативно можно использовать способ, в котором олигонуклеотиды присоединяют к кэпирующей структуре химическими методами, как описано Genset. Такой способ основан на окислении кэпирующей структуры (патент США № 6022715). Он позволяет (1) добавить к кэпу олигонуклеотид, который может содержать участок, распознаваемый рестрикционным ферментом класса IIS или класса III, и (2) получить первую цепь кДНК, которую затем можно использовать для синтеза второй цепи кДНК.

Альтернативно можно использовать способ с переключением кэпа, описанный Clontech (патент США № 5962272). Можно получить первую цепь кДНК в присутствии олигонуклеотида, используемого для переключения кэпа, который несет участок, распознаваемый веществом, способным распознавать нуклеиновые кислоты и отщеплять их от последовательности распознавания, благодаря чему можно использовать рестрикционный фермент класса IIS или класса III. Механизм переключения кэпа позволяет продолжать синтез первой цепи на олигонуклеотидах, используемых для переключения кэпа. Его можно продолжать путем синтеза второй цепи кДНК или путем последующей стадии ПЦР, как описано, например, в системе клонирования SMART™ Clontech.

В другом варианте осуществления в зависимости от качества РНК случайное праймирование и удлинение кДНК до кэпирующей структуры может обеспечить утилизацию 5'-концов. Применение конкретного фермента и конкретных условий реакции позволяет иногда достичь участка кэпа с высокой эффективностью (Carninci et al., *Bio techniques*, 2002). Даже при отсутствии выбора кэпа можно присоединить вместо кэпирующей структуры олигонуклеотиды, которые несут участки, распознаваемые рестрикционным ферментом класса IIS или класса III, которые позже можно использовать для получения конкатемера.

И наконец, кДНК можно расщепить рестрикционным ферментом класса II (класс IIS или класс IIG) или класса III с получением 5'-концевых маркеров. 5'-концевые маркеры используют для последующего получения конкатемера. Также можно использовать любые другие способы, включая механическое расщепление.

На фиг. 1 показаны иллюстративные производственные процессы согласно настоящему изобретению. В соответствии с фиг. 1 для проведения способа согласно настоящему изобретению 5'-концы транскрибируемых участков можно выделить из совокупности молекул РНК, или из совокупности молекул РНК, обогащенных фракциями РНК, соответствующими энхансерам, или из библиотеки полноразмерных кДНК.

Если способ настоящего изобретения применяют к совокупности молекул РНК, молекулы РНК можно использовать в качестве матриц для синтеза комплементарных цепей кДНК. Цепи кДНК подвер-

гают стадии селекции, чтобы обогатить их гибридами РНК-кДНК, содержащими 5'-концы транскрибируемых участков. После удаления или деструкции фрагмента мРНК путем гидролиза щелочью получают совокупность первых цепей кДНК, содержащих 5'-концы транскрибируемых участков.

В другом варианте осуществления согласно настоящему изобретению библиотеку полноразмерных кДНК можно использовать для получения совокупности РНК, содержащих 5'-концы клонов кДНК. Затем синтезируют совокупность одноцепочечных кДНК, используя вышеуказанную совокупность РНК в качестве матрицы. Фракцию одноцепочечных кДНК получают после удаления или деструкции молекул РНК путем гидролиза щелочью, причем полученная совокупность одноцепочечных кДНК содержит 5'-концы транскрибируемых участков. Транскрибируемые участки можно подвергнуть дальнейшему процессингу согласно настоящему изобретению. Следует отметить, что при использовании библиотеки полноразмерных кДНК в качестве исходного вещества селекция по 5'-концам не требуется.

Стадия 2.

Нижеследующую стадию 2 проводят в продолжение стадии 1, чтобы селективно собрать фрагменты, содержащие участок кДНК, который, по меньшей мере, содержит участок, комплементарный 5'-концу РНК.

При использовании способа с улавливанием кэпа высвобождают первую цепь кДНК, иммобилизованную на носителе. Это можно осуществить путем обработки носителя щелочью, такой как гидроксид натрия. Вместо обработки щелочью можно провести ферментативную реакцию с РНКазой H (которая расщепляет только РНК, гибридованную с ДНК). Обработка щелочью приводит к высвобождению кДНК из гибрида РНК-кДНК, связанного с носителем, через кэп на РНК и позволяет отделить кДНК от РНК с получением одноцепочечной кДНК в чистом виде.

Затем к кДНК добавляют линкер, который содержит последовательность, распознаваемую в последовательность-специфической манере веществом, обладающим ферментативной активностью, которое расщепляет распознаваемую ДНК вне распознаваемой последовательности. Такие вещества включают, без ограничения, некоторые рестрикционные ферменты класса II и класса III.

В данном варианте осуществления линкер, который содержит, по меньшей мере, участок рестрикционного фермента класса IIS или класса III и произвольный олигомерный фрагмент на 3'-конце, лигируют к концу указанной одноцепочечной кДНК, который соответствует 5'-концу вышеупомянутой РНК (т.е. к 3'-концу кДНК). Для последующего клонирования маркеров 5'-концевой последовательности в конкатемерах предпочтительно, но не обязательно, ввести в линкер второй участок распознавания. Второй участок распознавания предпочтительно отличается от вышеуказанного участка распознавания, используемого, например, рестрикционным ферментом класса IIS или класса III.

Эту стадию предпочтительно проводят с использованием линкера, несущего участок, распознаваемый рестрикционным ферментом класса IIS или класса III, и произвольный олигомерный фрагмент (SSLLM (способ лигирования одноцепочечного линкера), Y. Shibata et al., *BioTechniques*, Vol. 30, No. 6, pp. 1250-1254, (2001)). Рестрикционные ферменты класса IIS или класса III представляют собой группы рестрикционных ферментов, которые вызывают расщепление по участкам, отличным от участка распознавания. Примером рестрикционного фермента класса IIS является, без ограничения, GsuI. GsuI расщепляет одну из цепей на 16 п.о. ниже участка распознавания, а другую цепь - на 14 п.о. ниже участка распознавания. Другим подходящим примером является MmeI, который расщепляет последовательности соответственно на расстоянии 20 и 18 оснований от распознаваемой им последовательности. Примером рестрикционного фермента класса III является, без ограничения, EcoP15I, который расщепляет последовательности соответственно на расстоянии 25 и 27 п.о. от распознаваемого им участка. Произвольный олигомерный фрагмент находится на 3'-конце линкера, и хотя число оснований конкретно не ограничивается, рекомендуемое число оснований находится в диапазоне от 5 до 9, более предпочтительно от 5 до 6. Участок, распознаваемый рестрикционным ферментом класса IIS или класса III, предпочтительно находится вблизи вышеупомянутого произвольного олигомерного фрагмента, так, чтобы точка расщепления находилась внутри кДНК. Линкер предпочтительно представляет собой линкер двухцепочечной ДНК, вышеупомянутый произвольный олигомерный фрагмент которой выдается до 3'-конца и обеспечивает связывающий конец. Кроме, целесообразно предварительно присоединить к линкеру селективное связывающее вещество, такое как биотин, чтобы облегчить впоследствии его сбор.

Если полученную вышеупомянутую первую цепь кДНК подвергают взаимодействию с таким линкером, произвольный олигомерный фрагмент линкера гибридизуется с 3'-концом первой цепи кДНК (т.е. с 5'-концом матричной РНК). Затем синтезируют вторую цепь кДНК с использованием данного линкера в качестве праймера и первой цепи кДНК в качестве матрицы. Эту стадию можно проводить стандартным способом. В другом варианте осуществления согласно настоящему изобретению первую цепь кДНК можно удалить путем гибридизации с совокупностью нуклеиновых кислот с последующим физическим разделением одноцепочечных молекул и двухцепочечных молекул ДНК-ДНК или двухцепочечных гибридов ДНК-РНК. Такую стадию удаления можно проводить, без ограничения, с помощью способа, раскрытого в публикации патента США № 20020106666. Одноцепочечную кДНК, собранную на стадии удаления, используют в качестве матрицы для синтеза второй цепи с помощью стандартных процедур, подобных вышеуказанному способу без стадии удаления.

Затем полученную двухцепочечную кДНК обрабатывают вышеупомянутым рестрикционным ферментом класса IIS или класса III. На данной стадии получают двухцепочечный фрагмент кДНК, включающий полученный с использованием линкера фрагмент и фрагмент, полученный из 5'-конца кДНК (5'-конец второй цепи кДНК). Например, если *GsuI* используют в качестве рестрикционного фермента класса IIS, и если линкер конструируют так, чтобы локализовать участок рестрикции сразу за (выше по ходу считывания) вышеупомянутым произвольным олигомерным участком, полученный фрагмент ДНК будет содержать участок, полученный из участка на 5'-конце второй цепи ДНК (т.е. участка на 5'-конце РНК), длиной 16 п.о. (длина комплементарной цепи составляет 14 п.о.). При применении *MmeI* длина второй цепи фрагмента ДНК увеличивается до 20 и 18 п.о. соответственно, а в случае *EcoP15I* - до 25 и 27 п.о. соответственно.

Затем селективно собирают такие фрагменты ДНК. Если селективное связывающее вещество (например, биотин) присоединено к линкеру, как указано выше, сбор можно проводить, как на стадии 1, с использованием носителя, на котором фиксируется соответствующее селективное связывающее вещество (например, стрептавидин). Эта процедура завершает стадию 2, на которой селективно собирают фрагменты, содержащие участок кДНК, принадлежащий к первой цепи кДНК, который, по меньшей мере, содержит участок, комплементарный 5'-концу вышеупомянутой РНК.

Выше описан случай применения SSLLM на стадии 2, однако стадию 2 также можно проводить с помощью любого другого способа, при условии, что способ позволяет селективно собирать фрагменты, содержащие 3'-конец первой цепи кДНК (5'-конец матричной мРНК). Например, можно использовать экзонуклеазу, которая расщепляет нуклеотид в направлении от 5' к 3' с контролируемой скоростью. Обработка экзонуклеазой первой цепи кДНК в течение заданного периода времени позволяет получить одноцепочечный фрагмент, включающий 3'-конец первой цепи кДНК (5'-конец матричной РНК). Путем обработки нуклеазой, которая расщепляет только двухцепочечные фрагменты, можно получить только целевые одноцепочечные фрагменты. Указанные фрагменты можно собрать, соединить адаптерами и клонировать.

Вышеуказанные собранные фрагменты, которые соответствуют 5'-концу, можно дополнительно лигировать с линкерами и затем использовать для амплификации методом ПЦР, если их количество является недостаточным для последующего применения, такого как клонирование.

В одном варианте осуществления фрагменты, соответствующие 5'-части РНК, лигируют с 3'-концом линкера, несущего только участок другого рестрикционного фермента, который может отличаться от участка рестрикции, присутствующего в первом линкере. Соответственно фрагменты, соответствующие 5'-концу мРНК, содержат линкеры, несущие участки, распознаваемые рестрикционными ферментами, на обеих сторонах. Такие фрагменты можно амплифицировать методом ПЦР с последующим расщеплением одним или двумя рестрикционными ферментами и получить фрагменты ДНК, подходящие для клонирования конкатемеров, что описано ниже более подробно.

В другом варианте осуществления, подобно описанному в литературе способу (Velculescu et al., 1995), вышеупомянутый фрагмент ДНК или продукт ПЦР вначале используют для получения димерных молекул, содержащих два фрагмента, специфичных к 5'-концу, лигированных друг с другом в противоположной ориентации. Указанные димеры можно затем использовать непосредственно или после другой ПЦР-амплификации с получением конкатемеров, что описано ниже более подробно.

В очередном варианте осуществления согласно настоящему изобретению альтернативно амплификации ДНК методом ПЦР с помощью РНК-полимеразы можно в первую очередь амплифицировать фрагменты, соответствующие 5'-концам и содержащие подходящие линкеры на обоих концах. Затем фрагменты ДНК воспроизводят на стадии обратной транскрипции и получают вторую цепь, чтобы обеспечить образование конкатемера.

Стадия 3.

На следующей стадии 3 получают конкатемеры путем попарного лигирования собранных фрагментов. Поскольку существует несколько РНК, а линкер гибридизуется с первой цепью кДНК по произвольному олигомерному участку, как указано выше, вышеописанный способ позволяет получить фрагменты, содержащие несколько кДНК, полученных из нескольких РНК, присутствующих в образце. На стадии 3 лигируют несколько указанных фрагментов с получением конкатемеров. Лигирование фрагментов кДНК можно проводить стандартным способом с использованием коммерческих наборов для лигирования, содержащих, без ограничения, ДНК-лигазу Т4. Для надежного проведения лигирования используют, без ограничения, способ, включающий вначале введение второго линкера, несущего участок распознавания рестрикционным ферментом, который отличается от других участков распознавания, используемых на ранних стадиях, затем лигирование двух фрагментов с получением димеров, содержащих два 5'-маркера в противоположном направлении (di-tag), и лигирование таких лигированных фрагментов di-tag с получением конкатемеров. Однако эффективность согласно настоящему изобретению не зависит от клонирования промежуточных di-tags. Мономерные маркеры могут подвергаться самолигированию непосредственно с образованием конкатемеров удовлетворительной длины согласно настоящему изобретению. Таким образом, настоящее изобретение не ограничивается или не зависит от применения di-tags. Число лигированных фрагментов не ограничивается, практически любое число, составляющее два или более,

предпочтительно по меньшей мере от 20 до 30, подходит для проведения настоящего изобретения. Полученные конкатемеры предпочтительно, но без ограничения, амплифицируют или клонируют стандартным способом.

Каждый из конкатемеров, полученных по описанному способу, содержит участок, имеющий такую же последовательность оснований (однако урацил в РНК может соответствовать тимину в ДНК), что и 5'-конец совокупности РНК в образце. Хотя он также содержит фрагмент, полученный из линкера или линкеров, последовательность оснований линкера или линкеров известна из схемы эксперимента, поэтому фрагмент, полученный из линкера или линкеров, и фрагмент, полученный из РНК, можно отчетливо различить путем исследования последовательности оснований конкатемера. Следовательно, путем анализа последовательности оснований полученного конкатемера можно определить последовательность оснований на 5'-конце совокупности РНК в образце. Последовательности оснований, состоящие максимум из 16, 20 или 25 оснований, на 5'-конце каждой РНК предпочтительно определяют с использованием *GsuI*, *MmeI* или *EcoP15I*. Информация по 16, 20 или 25 основаниям может быть достаточной для статистически почти однозначной идентификации РНК и для заключения о том, является ли РНК новой. Кроме того, путем определения последовательности оснований конкатемера можно проанализировать последовательности оснований на 5'-конце РНК для ряда вышеуказанных фрагментов, входящих в состав конкатемера (предпочтительно от 20 до 30), следовательно, можно успешно получить информацию по 5'-концу совокупности РНК. Анализ конкатемеров можно автоматизировать путем применения компьютерного программного обеспечения для различения последовательностей, полученных из 5'-концов, и последовательностей, полученных из линкера или линкеров.

Последовательности специфических 5'-концевых маркеров, полученные из конкатемеров в вышеупомянутой форме, можно проанализировать на идентичность с помощью решений стандартного программного обеспечения для выравнивания последовательностей, такого как NCBI BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>). FASTA присутствует в пакете Genetics Computer Group (GCG), Accelrys Inc. (<http://www.accelrys.com/>) и др. Такие решения программного обеспечения позволяют выравнивать маркеры 5'-концевых специфических последовательностей друг с другом, чтобы идентифицировать уникальные или не избыточные маркеры для кластеризации и дополнительного применения в поиске по базам данных. Все такие не избыточные маркеры последовательностей можно по отдельности сосчитать и дополнительно проанализировать на вклад каждого не избыточного маркера в общее число всех маркеров, полученных из одного образца. Вклад отдельного маркера в общее число всех маркеров позволяет количественно определить транскрипты в совокупности РНК или библиотеке кДНК. Результаты, полученные таким образом для индивидуальных образцов, можно дополнительно сравнить с подобными результатами, полученными из других образцов, чтобы сравнить характеры их экспрессии друг с другом. Таким образом, настоящее изобретение позволяет определить профили экспрессии отдельных транскриптов в одном или более образцах и получить справочную базу данных.

Специфические маркеры 5'-концевых последовательностей, полученные по описанному выше способу, можно также использовать для идентификации транскрибируемых участков в геномах, для которых получены частичные или полноразмерные последовательности. Такой поиск можно проводить с использованием решений стандартного программного обеспечения, такого как NCBI BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>), путем выравнивания специфических маркеров 5'-концевых последовательностей с геномными последовательностями. Хотя обнаружено, что маркеры размером 20 п.о. можно специфически картировать в геномных последовательностях, в некоторых случаях может потребоваться информация об удлиненной исходной последовательности, полученной при анализе конкатемеров, например, с помощью одного из описанных ниже способов. Применение более длинных последовательностей позволяет более точно идентифицировать активно транскрибируемые участки в геноме. Подобным образом, такой же подход и такие же решения программного обеспечения можно использовать для поиска родственных последовательностей в других базах данных, например, таких как NCBI (<http://www.ncbi.nlm.nih.gov/Database/index.html>), EMBL-EBI (<http://www.ebi.ac.uk/Databases/index.html>) или японская база данных ДНК (<http://www.ddbj.nig.ac.jp/>).

Специфические маркеры 5'-концевых последовательностей, которые можно картировать в геномных последовательностях, позволяют идентифицировать регуляторные последовательности (Suzuki Y. et al., EMBO Rep. 2001 May; 2(5): 388-93 и Suzuki Y. et al., Genome Res. 2001 May; 11(5): 677-84). ДНК, находящаяся в гене выше 5'-конца транскрибируемого участка, охватывает большую часть регуляторных элементов, используемых для контроля за экспрессией генов. Указанные регуляторные последовательности можно также анализировать на функциональность путем поиска в базах данных, содержащих информацию об участках связывания факторов транскрипции. Общедоступные базы данных по участкам связывания факторов транскрипции и результатам анализа промоторов, такие как База данных по регуляции транскрипции (TRRD) (<http://www.mgs.bionet.nsc.ru/mgs/dbases/trrd4/>), TRANSFAC (<http://transfac.gbf.de/TRANSFAC/>), TFSEARCH (<http://www.cbrc.jp/research/db/TFSEARCH.html>) и PromoterInspector, предоставляемые Genomatix Software (<http://www.genomatix.de/>), можно использовать для получения информации для компьютерного анализа промоторных участков и энхансерных участков.

Информацию о последовательности, полученную с использованием специфических маркеров 5'-

концевых последовательностей или путем картирования 5'-концевых последовательностей в геноме, также можно использовать для управления регуляцией конкретного гена-мишени. В таком эксперименте информацию, касающуюся промотора, можно использовать для изменения его активности, или для замены его искусственным промотором. Альтернативно 5'-концевые специфические маркеры могут предоставлять информацию о последовательности для конструирования антисмысловых зондов или РНКи-зондов, используемых для инактивации генов.

В другом варианте осуществления согласно настоящему изобретению информацию о последовательности, полученную с использованием конкатемеров, можно использовать для синтеза специфических праймеров, необходимых для клонирования полноразмерных кДНК. В данном подходе последовательность, полученную из конкретного 5'-концевого специфического маркера, используют для конструирования прямого праймера, тогда как выбор обратного праймера зависит от матрицы ДНК, используемой в реакции амплификации. Амплификацию методом полимеразной цепной реакции (ПЦР) можно проводить с использованием матрицы, полученной из совокупности РНК, полученных из биологического образца, и праймера олиго-дТ. На первой стадии для синтеза пула кДНК используют праймер олиго-дТ и обратную транскриптазу. На второй стадии для амплификации полноразмерной кДНК из пула кДНК используют прямой праймер, полученный из 5'-концевого специфического маркера, и праймер олиго-дТ. Подобным образом, специфическую полноразмерную кДНК можно амплифицировать из существующей библиотеки кДНК с использованием прямого праймера, полученного из 5'-концевого маркера, и вставленного в вектор обратного праймера.

Хотя в описанном выше способе в качестве исходного субстрата используют РНК, присутствующую в образце, стадию 1 можно опустить путем использования существующей библиотеки полноразмерных кДНК. В данном способе информацию по последовательности оснований 5'-концов совокупности кДНК (т.е. 5'-концов РНК, используемых в качестве матриц для получения указанных кДНК), содержащуюся в библиотеке полноразмерных кДНК, можно успешно получить по способу, аналогичному вышеуказанной процедуре.

Независимо от исходного вещества, используемого при проведении настоящего изобретения, одноцепочечную первую цепь кДНК можно фракционировать путем вычитающей гибридизации и физического разделения, чтобы обеспечить обогащение 5'-концов генов с разным характером экспрессии, или для концентрирования транскриптов с низкой распространенностью.

В некоторых вариантах осуществления может быть желательно получить информацию о удлинённых последовательностях 5'-концов транскрибируемых участков. С использованием таких удлинённых последовательностей в некоторых случаях можно идентифицировать участок инициации синтеза белка или улучшить картирование геномных последовательностей. Как описано выше, на стадии 2 настоящее изобретение включает лигирование линкера с 5'-концом кДНК. Введение одноцепочечного липкого конца, охватывающего последовательность, полученную из конкатемера, для связывания и лигирования с конкретным фрагментом нуклеиновой кислоты позволяет использовать такой линкер в мишень-специфической манере. После лигирования линкер можно использовать для обогащения фрагмента ДНК путем присоединения линкера к носителю, от которого его можно освободить после обогащения. Линкер также можно использовать в качестве праймера, чтобы получить информацию об удлинённых последовательностях на 5'-концах в жидкой фазе или на твердой фазе, используемой до обогащения.

Путем исследования последовательностей оснований конкатемеров или удлинённых 5'-последовательностей, полученных по способу настоящего изобретения, можно не только клонировать новые гены, как описано выше, но и исследовать профили экспрессии генов в образце. Кроме того, такую технологию можно использовать в разных целях, например для картирования участков инициации транскрипции в геноме, для картирования характера использования промотора или характера использования энхансера, для анализа SNP в участках промотора или участках энхансера, для создания генных сетей путем объединения результатов анализа экспрессии с информацией по промоторам и энхансерам, альтернативному использованию промотора или энхансера и по доступности факторов транскрипции, а также для селективного сбора участка промотора или участка энхансера в фрагментированной геномной ДНК. Для сбора геномных фрагментов, содержащих промоторные участки, фрагмент, содержащий такую же последовательность оснований, как и 5'-конец мРНК, можно присоединить к носителю, например, с использованием вышеупомянутой системы биотина, и гибридизовать с фрагментированной геномной ДНК. Гибридизованные геномные фрагменты ДНК можно затем отделить от смеси геномных фрагментов с использованием, например, магнитных гранул, покрытых стрептавидином, и клонировать в стандартных условиях.

Альтернативно клонирования конкатемера можно избежать путем получения и применения выделенных 5'-концевых маркеров, лигированных со смесью полноразмерных кДНК и присоединённых к магнитным гранулам, несущих гомогенные последовательности олигонуклеотидов, с последующим лигированием, например, к SLLM, получением второй цепи кДНК и расщеплением рестрикционным ферментом класса IIS или класса III. 5'-концевой специфический маркер может специфично фиксироваться на гранулах и может использоваться для специфического секвенирования по способу, подобному описанному Lynx Therapeutics (патенты США №№ 6352828, 6306597, 6280935, 6265163 и 5695934).

Например, олигонуклеотиды могут содержать "произвольный фрагмент I", который связывается с 5'-концами кДНК; и кодовый фрагмент олигонуклеотида, который может "метить" продукт лигирования. Если олигонуклеотид не гибридизован с кДНК, он может разлагаться под действием экзонуклеазы VII. "Декодированные" олигонуклеотиды можно использовать для отбора последовательности. Специфические массивы кДНК на гранулах затем располагают на твердой поверхности по отдельности и затем проводят параллельное секвенирование. Вышеупомянутый способ позволяет создать жидкий формат массива, в котором каждую гранулу можно адресовать с использованием независимой метки и обработать отдельно путем анализа последовательности и т.п.

В другом варианте осуществления согласно настоящему изобретению известные 5'-концевые специфические маркеры можно использовать для альтернативного анализа 5'-концевых специфических последовательностей, опуская клонирование и секвенирование конкатемеров. В таком случае 5'-концевые специфические олигонуклеотиды размером примерно 25 п.о. синтезируют и фиксируют на твердом носителе с получением 5'-концевого специфического микрочипа. Затем путем гибридизации 5'-маркеров, полученных из образца, идентифицируют и количественно определяют транскрипты, присутствующие в образце. Стандартные способы получения и применения микрочипов известны специалистам в области молекулярной биологии (Jordan B., DNA Microarrays: Gene Expression Applications, Springer-Verlag, Berlin Heidelberg New York, 2001; Schena A., DNA Microarrays, A Practical Approach, Oxford University Press, Oxford 1999).

Путем модификаций вышеупомянутых способов непосредственного секвенирования 5'-концов или считывания результатов гибридизации с 5'-концевым специфическим микрочипом настоящее изобретение предлагает разные способы общего анализа 5'-концов в виде конкатемеров или анализа отдельных 5'-концов, которые обогащаются посредством 5'-концевой специфической селекции.

На фиг. 2 приведен пример последовательности операций, соответствующих описанным выше стадиям 2 и 3. На фиг. 2, например, показано, что рестрикционные ферменты XmaII, MmeI и XbaI используют для клонирования фрагментов ДНК размером 33 п.о. Как правило, клонирование 5'-концевых специфических маркеров включает нижеследующие стадии.

На начальной стадии согласно настоящему изобретению, показанной на фиг. 1, получают пул одноцепочечных кДНК. Пул включает в себя 5'-концевые участки, транскрибированные из РНК. Вблизи фрагмента одноцепочечной кДНК, который содержит 5'-концевые участки, транскрибированные с РНК, лигируют специфический линкер, называемый здесь "первый линкер", чтобы ввести участок распознавания рестрикционным ферментом, который расщепляет последовательность вне первого линкера в соответствии с его связывающим участком или в пределах 5'-концевого транскрибируемого участка. В примере, описанном на указанной фигуре, используют рестрикционный фермент MmeI, который расщепляет последовательность на 21 п.о. ниже участка распознавания, обеспечивая терминацию маркеров, которые включают 5'-концы транскрибируемых участков РНК. Кроме того, второй рестрикционный фермент предоставляется для "первого линкера". В целях данного примера XmaII используют для более позднего клонирования 5'-концевых специфических маркеров.

Затем "первый линкер" используют для затравки синтеза второй комплементарной цепи кДНК с получением двухцепочечных молекул кДНК, которые содержат 5'-концы транскрибируемых участков РНК и участок распознавания рестрикционными ферментами, которые расщепляют последовательность по участку, расположенному вне первого линкера, в соответствии с участком связывания, примыкающим к участку, содержащему 5'-концевые участки транскрибируемых РНК.

Вышеупомянутый рестрикционный фермент, который расщепляет последовательность вне участка связывания, представляет собой, в целях этого примера, MmeI. Расщепление MmeI дает двухцепочечные фрагменты ДНК маркеров, которые содержат 5'-концы транскрибируемых участков РНК и "первый линкер", а также они содержат одноцепочечный липкий конец ДНК в участке расщепления MmeI.

К вышеуказанному одноцепочечному липкому концу ДНК в участке расщепления MmeI лигируют "второй линкер", чтобы ввести участок распознавания рестрикционным ферментом, подходящим для клонирования фрагментов ДНК или маркеров, которые можно использовать в качестве матриц при амплификации методом ПЦР.

Фрагмент кДНК, содержащий "первый линкер", фрагменты кДНК, содержащие 5'-концы участков, транскрибированных из РНК, и "второй линкер" очищают путем селективного связывания с носителем посредством селективного связывающего вещества, присоединенного к первому линкеру.

С целью клонирования фрагментов ДНК, содержащих 5'-концы транскрибированных участков или маркеры, вышеупомянутую фракцию кДНК, содержащих "первый линкер", фрагменты кДНК или маркеры, которые содержат 5'-концевые участки, транскрибированные из РНК, и "второй линкер" амплифицируют методом ПЦР, причем линкерные фрагменты отщепляют рестрикционными ферментами, чтобы осуществить лигирование маркеров в конкатемеры. В этом примере используют рестрикционные ферменты XmaII и XbaI, которые отщепляют фрагмент размером 33 п.о. от вышеупомянутых фрагментов кДНК. После соответствующей стадии очистки фрагменты размером 33 п.о. лигируют друг с другом с получением конкатемеров, содержащих, например, до 30 маркеров, включающих 5'-концы транскрибируемых участков указанной РНК, или клонируют их по отдельности.

Конкатемеры можно клонировать в векторе для секвенирования с получением библиотеки, содержащей 5'-концевые участки, транскрибированные из РНК.

На фиг. 3 показан основной поток операций в соответствии с настоящим изобретением, иллюстрирующий альтернативный подход к прямому секвенированию 5'-концевых маркеров. В целях этого варианта осуществления согласно настоящему изобретению одноцепочечные кДНК, которые содержат 5'-концевые участки, транскрибированные из РНК и полученные по способу, показанному на фиг. 1, лигируют с линкером, называемым здесь "первый линкер", который в целях этого примера содержит специфическую метку, обеспечивающую иммобилизацию продукта лигирования на твердом носителе. Этот линкер можно использовать в качестве праймера для синтеза второй цепи кДНК, комплементарной первой цепи. Одноцепочечные ДНК, содержащие двухцепочечный линкер, примыкающий к участку, содержащему 5'-концевые участки, транскрибированные из РНК, или двухцепочечные ДНК, содержащие 5'-концевые транскрибируемые участки, можно подвергнуть индивидуально или параллельному секвенированию, в целях этого примера, посредством способа серийного секвенирования 5'-концов РНК с высокой пропускной способностью.

Примеры

Пример 1.

Вначале из клеток MCF7 собирают ядра обычным способом. Чтобы собрать ядра, буферный раствор для денатурации, содержащий мочевины в заранее определенной концентрации (0,5, 1, 2 или 4 моль/л), добавляют к ядрам, чтобы денатурировать ядра под действием мочевины. Затем ядерные компоненты разделяют на растворимую фракцию (ядерную) и хроматиновую фракцию (хроматин). Затем в растворимой фракции и в хроматиновой фракции методом вестерн-блоттинга детектируют GAPDH, UlsnRNP70 и полимеразу II. GAPDH представляет собой маркер цитоплазмы, а UlsnRNP70 представляет собой маркер внутриядерной растворимой фракции. Результаты вестерн-блоттинга показаны на фиг. 4. Как можно видеть на фиг. 4, в относительном выражении, UlsnRNP70 обнаруживается в растворимой фракции и не обнаруживается в хроматиновой фракции. Это свидетельствует об успешном проведении эффективного фракционирования. В хроматиновой фракции фосфорилированная активная РНК-полимераза (активная Pol II) присутствует в концентрированном виде, указывая на то, что в хроматиновой фракции присутствует РНК-полимераза, находящаяся в процессе синтеза.

Пример 2.

В традиционном способе CAGE в качестве матричной РНК используют общую РНК. Далее данный способ называют обычным способом CAGE. С другой стороны, если способ CAGE используют в контексте настоящего изобретения, в качестве матричной РНК используют растущую цепь РНК. Далее способ CAGE, в котором согласно настоящему изобретению в качестве матричной РНК используют растущую цепь РНК, называют способ NET-CAGE (Native Elongating Transcript-Cap Analysis of Gene Expression) настоящего изобретения. В настоящем примере растущую цепь РНК получают путем сбора РНК из хроматиновой фракции, полученной в примере 1, а способ NET-CAGE проводят с использованием полученной таким образом растущей цепи РНК в качестве матрицы. Затем способ NET-CAGE сравнивают с обычным способом CAGE, в котором используют общую РНК, собранную из клеток MCF7. Если не указано иначе, вышесказанное применяется к описанным ниже примерам.

Что касается известных энхансерных участков, идентифицированных в проекте FANTOM 5 и уже описанных в литературе (Andersson et al. Nature 2014), число считываний в клетках MCF7, полученное с помощью обычного способа CAGE, и число считываний в клетках MCF7, полученное с помощью способа NET-CAGE согласно настоящему изобретению, показаны на графике, приведенном на фиг. 5. На фиг. 5, на оси X откладывают число считываний, полученное с помощью обычного способа CAGE, на оси Y откладывают число считываний, полученное с помощью способа NET-CAGE, а каждая точка обозначает один энхансерный участок. Как можно видеть на фиг. 5, в большей части энхансерных участков превышение сигналов (числа считываний), полученных с помощью способа NET-CAGE согласно настоящему изобретению, по сравнению с сигналами, полученными с помощью обычного способа CAGE, составляет, по меньшей мере, от нескольких раз до десяти раз. Кроме того, хотя некоторые энхансеры в целом не детектируются с помощью обычного CAGE, многие из них можно детектировать с помощью способа NET-CAGE с достаточным числом считываний. Из приведенных результатов можно сделать вывод, что способ NET-CAGE согласно настоящему изобретению позволяет детектировать энхансеры с гораздо более высокой чувствительностью, чем обычный способ CAGE.

Пример 3.

Что касается известных энхансерных участков, идентифицированных в проекте FANTOM 5 и уже описанных в литературе (Andersson et al. Nature 2014), воспроизводимость оценивают по числу считываний в клетках MCF7, определенному с помощью обычного способа CAGE, и числу считываний в клетках MCF7, определенному с помощью способа NET-CAGE согласно настоящему изобретению. Полученные результаты показаны на графике, приведенном на фиг. 6. На фиг. 6 на оси X и на оси Y откладывают число считываний, отнесенное к известным энхансерным участкам, идентифицированным в этом документе. Как можно видеть на фиг. 6, если используют обычный способ CAGE, коэффициент корреляции, относящийся к числу считываний, показанному в документе, составляет 0,69, а если используют способ

NET-CAGE согласно настоящему изобретению, коэффициент корреляции, относящийся к числу считываний, показанному в документе, составляет 0,65. Полученные результаты демонстрируют, что, как и обычный CAGE, способ NET-CAGE согласно настоящему изобретению также позволяет идентифицировать и количественно определить энхансеры с высокой воспроизводимостью и с высокой надежностью.

Пример 4.

Обычный CAGE и NET-CAGE проводят с использованием клеток MCF7. Распределения маркеров CAGE в участке, расположенном выше гена GAPDH, полученные с помощью обычного способа CAGE и способа NET-CAGE, показаны отдельно на фиг. 7. На фиг. 7 показаны картины просмотра генома, полученные с помощью интеграционного устройства для просмотра генома (IGV). На фиг. 7 картированные маркеры CAGE показаны в координатах генома, а на горизонтальной оси указаны координаты положения в геноме. Целью настоящего примера является определение, позволяет ли способ NET-CAGE согласно настоящему изобретению идентифицировать новые энхансеры-кандидаты, которые невозможно идентифицировать с помощью обычного CAGE. С этой целью на фиг. 7 показаны фрагменты, экстрагированные из картин просмотров. А именно на фиг. 7 показаны маркеры, присутствующие в основании пика промотора, показанного в верхней части фиг. 7. На фиг. 7 стрелки, не заключенные в рамку, указывают на маркеры CAGE, картированные в смысловом направлении, а стрелки, заключенные в рамку, указывают на маркеры CAGE, картированные в антисмысловом направлении. Участок, указанный тонкой стрелкой, представляет собой участок, в котором CAGE распределяются симметрично в противоположных направлениях. Хотя маркеры CAGE в данном участке практически не детектируются обычным CAGE, способ NET-CAGE позволяет детектировать достаточное число маркеров CAGE. Как описано выше, участок, в котором маркеры CAGE присутствуют симметрично в противоположных направлениях, является отличительным признаком энхансера, и этот характеристичный признак дает веские основания предполагать, что указанный участок является энхансером. Такой участок может не быть идентифицированным как энхансер в проекте FANTOM 5. Соответственно этот участок можно отнести к новому энхансеру-кандидату, идентифицированному с помощью способа NET-CAGE согласно настоящему изобретению. То есть можно сказать, что способ NET-CAGE согласно настоящему изобретению позволяет идентифицировать новые энхансеры, которые невозможно детектировать с помощью обычного способа CAGE.

В данный момент настоящее изобретение описывается посредством примеров. Следует отметить, что настоящее изобретение не ограничивается примерами. Эксперименты, описанные в примерах, может проводить любой специалист, владеющий стандартными методами, используемыми в области молекулярной биологии. Если в тексте не указано иное, технические термины, сокращенные обозначения и решения, используемые в примерах, должны иметь значения, традиционно используемые специалистами в области, к которой относится настоящее изобретение. Общее описание таких терминов, сокращенных обозначений и решений можно найти в разделе Обычно используемые реагенты, в Molecular Cloning (Sambrook and Russel, 2001). Все упомянутые здесь публикации включены в настоящий документ в качестве ссылки с целью раскрытия и описания присутствующих в них методов и/или материалов.

Настоящая заявка исходит из и испрашивает приоритет на основании Японской патентной заявки № 2016-013575, поданной 27 января 2016 г., полное раскрытие которой включено в настоящее описание в качестве ссылки во всей полноте.

Промышленная применимость.

Как описано выше, настоящее изобретение позволяет идентифицировать функциональный элемент ДНК, такой как энхансер, с более высокой чувствительностью и определить уровень транскрипции функционального элемента ДНК.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ декодирования последовательности оснований нуклеиновой кислоты, соответствующей концевому участку РНК, включающий

стадию получения-декодирования, содержащую получение с использованием матричной РНК нуклеиновой кислоты РНК 5'-концевого участка матричной РНК или комплементарной нуклеиновой кислоты ДНК, соответствующей 5'-концевому участку матричной РНК или нуклеиновой кислоты, последовательность которой содержит две или более из указанных последовательности, связанные между собой, и стадию определения последовательности нуклеиновой кислоты РНК, комплементарной нуклеиновой кислоты ДНК или нуклеиновой кислоты, имеющей последовательность, которая имеет две или более ее связанные последовательности, способом CAGE для декодирования последовательности оснований нуклеиновой кислоты, соответствующей концевому участку РНК, где указанное получение осуществляют в ходе следующих подстадий:

получения нуклеиновой кислоты с добавлением линкера;
расщепления нуклеиновой кислоты с добавлением линкера с использованием рестрикционного фермента, имеющего участок распознавания в линкере и участок расщепления в нуклеиновой кислоте, и сбора фрагмента, соответствующего 5'-концевому участку матричной РНК, из полученного продук-

та расщепления, и

где матричная РНК, используемая на стадии получения, представляет собой немеченую растущую цепь РНК.

2. Способ по п.1, где

на стадии получения-декодирования

получаемая нуклеиновая кислота содержит по меньшей мере одну нуклеиновую кислоту РНК 5'-концевого участка матричной РНК или комплементарную нуклеиновую кислоту ДНК, соответствующую 5'-концевому участку матричной РНК, и

используют метод на основе способа nAnT-iCAGE.

3. Способ по п.1, где

на стадии получения-декодирования

получаемая нуклеиновая кислота содержит по меньшей мере одну нуклеиновую кислоту РНК 5'-концевого участка матричной РНК или комплементарную нуклеиновую кислоту ДНК, соответствующую 5'-концевому участку матричной РНК.

4. Способ по п.1, где

на стадии получения-декодирования

получаемая нуклеиновая кислота содержит по меньшей мере одну нуклеиновую кислоту РНК 5'-концевого участка матричной РНК или комплементарную нуклеиновую кислоту ДНК, соответствующую 5'-концевому участку матричной РНК, и

используют метод, основанный на способе с переключением матрицы.

5. Способ по п.1, где

на стадии получения-декодирования

получаемая нуклеиновая кислота содержит по меньшей мере одну нуклеиновую кислоту РНК 5'-концевого участка матричной РНК или комплементарную нуклеиновую кислоту ДНК, соответствующую 5'-концевому участку матричной РНК, и

используют метод, основанный на способе HeliScopeCAGE.

6. Способ по любому из пп.1-5, где

немеченую растущую цепь РНК получают из штамма, выращенного из одной клетки, или из одной ткани.

7. Способ по любому из пп.1-6, где

немеченую растущую цепь РНК выделяют из комплекса, содержащего ДНК-матрицу и РНК-полимеразу II в процессе транскрипции и РНК в процессе синтеза.

8. Способ по любому из пп.1-7, дополнительно включающий

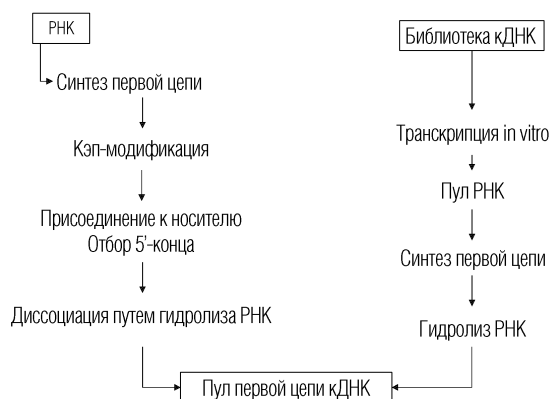
стадию выделения способа выделения немеченой растущей цепи РНК из биологического образца, где

выделенную немеченую растущую цепь РНК используют в качестве матричной РНК.

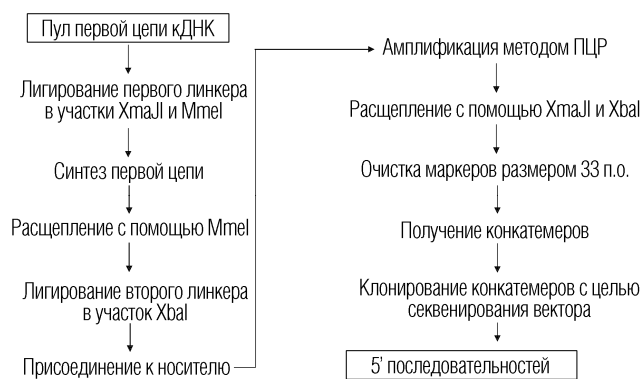
9. Способ по п.8, где

на стадии выделения

хроматин, содержащий комплекс ДНК-матрицы и РНК-полимеразы II в процессе транскрипции и РНК в процессе синтеза, выделяют из биологического образца, а содержащуюся в комплексе РНК, находящуюся в процессе синтеза, выделяют из хроматина в виде растущей цепи РНК.



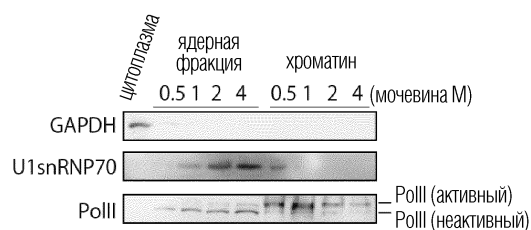
Фиг. 1



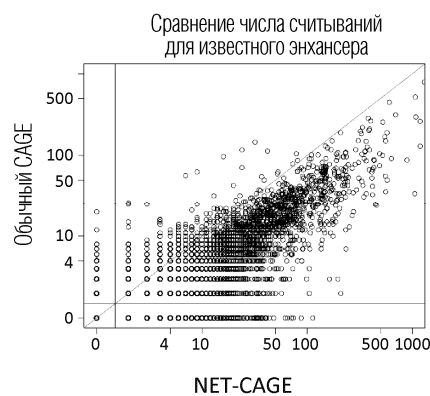
Фиг. 2



Фиг. 3



Фиг. 4



Фиг. 5

