

(19)



**Евразийское
патентное
ведомство**

(11) **047457**

(13) **B1**

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ

(45) Дата публикации и выдачи патента
2024.07.23

(21) Номер заявки
202091459

(22) Дата подачи заявки
2019.01.08

(51) Int. Cl. *C12N 9/02* (2006.01)
C12P 19/18 (2006.01)
C12Q 1/26 (2006.01)
C12Q 1/6806 (2018.01)
C12Q 1/6827 (2018.01)
C12Q 1/6869 (2018.01)

(54) МОДИФИКАЦИЯ ЦИТОЗИНА БЕЗ ПРИМЕНЕНИЯ БИСУЛЬФИТА С РАЗРЕШЕНИЕМ ДО ОДНОГО ОСНОВАНИЯ

(31) 62/614,798; 62/660,523; 62/771,409

(32) 2018.01.08; 2018.04.20; 2018.11.26

(33) US

(43) 2020.11.13

(86) PCT/US2019/012627

(87) WO 2019/136413 2019.07.11

(71)(73) Заявитель и патентовладелец:
**ЛЮДВИГ ИНСТИТЬЮТ ФОР
КЭНСЕР РИСЕРЧ ЛТД (СН)**

(72) Изобретатель:
Сун Чуньсяо, Лю Ибинь (GB)

(74) Представитель:
Нилова М.И. (RU)

(56) US-A1-20140322707
SCHULER et al. "Sequencing the Sixth Base (5-Hydroxymethylcytosine): Selective DNA Oxidation Enables Base-Pair Resolution," *Angewandte Chemie International Edition*, 22 October 2012 (22.10.2012), Vol. 51, No. 43, Pgs. 10704-10707. entire document

US-A1-20140127678
SONG et al. "Genome-wide Profiling of 5-Formylcytosine Reveals Its Roles in Epigenetic Priming," *Cell*, 25 April 2013 (25.04.2013), Vol. 153, No. 3, Pgs. 678-691. entire document

LU et al. "Chemical Modification-Assisted Bisulfite Sequencing (CAB-Seq) for 5-Carboxylcytosine Detection in DNA," *J Am Chem Soc*, 26 June 2013 (26.06.2013), Vol. 135, No. 25, Pgs. 9315-9317. entire document

LIU et al. "Bisulfite-Free, Base-Resolution, and Quantitative Sequencing of Cytosine Modifications," *bioRxiv*, 25 April 2018 (25.04.2018), Pgs. 1-18. entire document

US-A1-20170176421
PAIS et al. "Biochemical Characterization of a Naegleria TET-Like Oxygenase and its Application in Single Molecule Sequencing of 5-Methylcytosine," *Proceedings of the National Academy of Sciences*, 07 April 2015 (07.04.2015), Vol. 112, No. 14, Pgs. 4316-4321. entire document

(57) В настоящем изобретении представлены способы модификации нуклеиновой кислоты и набор для модификации нуклеиновой кислоты, включающие применение восстанавливающего агента на основе борана для превращения 5-карбоксилцитозина (5caC) или 5-формилцитозина (5fC) в дигидроурацил (DHU).

B1

047457

047457

B1

Настоящая заявка испрашивает приоритет предварительных заявок на патент США № 62/614798, поданной 8 января 2018 года, 62/660523, поданной 20 апреля 2018 года, и 62/771409, поданной 26 ноября 2018 года, каждая из которых включена в данный документ посредством ссылки во всей своей полноте.

Область техники, к которой относится изобретение

В настоящем изобретении представлены способы идентификации в последовательности нуклеиновой кислоты местоположений 5-метилцитозина, 5-гидроксиметилцитозина, 5-карбоксилцитозина и/или 5-формилцитозина. В настоящем изобретении также представлены способы модификации нуклеиновой кислоты и набор для модификации нуклеиновой кислоты, включающие применение восстанавливающего агента на основе борана для превращения 5-карбоксилцитозина (5caC) или 5-формилцитозина (5fC) в дигидроурацил (DHU).

Уровень техники

5-Метилцитозин (5mC) и 5-гидроксиметилцитозин (5hmC) являются двумя основными эпигенетическими метками, обнаруживаемыми в геноме млекопитающих. 5hmC образуется из 5mC под действием диоксигеназ из семейства ферментов транслокации десять-одиннадцать (TET). Tet может дополнительно окислять 5hmC до 5-формилцитозина (5fC) и 5-карбоксилцитозина (5caC), который присутствует в геноме млекопитающих со значительно меньшей представленностью по сравнению с 5mC и 5hmC (в от 10 раз до 100 раз меньшей, чем 5hmC). 5mC и 5hmC совместно играют решающую роль в широком спектре биологических процессов от регуляции генов до нормального развития. Было установлено, что аберрантное метилирование и гидроксиметилирование ДНК ассоциированы с различными заболеваниями и являются общепринятыми характерными признаками рака. Таким образом, определение 5mC и 5hmC в последовательности ДНК не только важно для фундаментальных исследований, но также представляет ценность для клинических путей применения, в том числе диагностики и терапии.

5fC и 5caC являются двумя конечными окисленными производными 5mC и могут превращаться в немодифицированный цитозин под действием тимин-ДНК-гликозилазы (TDG) в пути эксцизионной репарации оснований. Следовательно, 5fC и 5caC являются двумя важными ключевыми промежуточными соединениями в процессе активного деметилирования, который играет важную роль в эмбриональном развитии. 5fC и 5caC обнаруживаются в этом окружении и могут служить индикатором практически полного деметилирования 5mC. 5fC и 5caC также могут выполнять дополнительные функции, такие как связывание специфических белков и влияние на скорость и специфичность действия РНК-полимеразы II.

5mC также представляет собой посттранскрипционную модификацию РНК, которая была идентифицирована как в стабильных, так и характеризующихся высокой представленностью тРНК и рРНК, а также в мРНК. Кроме того, 5mC был выявлен в snRNA (малой ядерной РНК), miRNA (микроРНК), lncRNA (длинной некодирующей РНК) и eRNA (энхансерной РНК). Тем не менее, по-видимому, существуют различия в частоте встречаемости 5mC в определенных типах РНК у разных организмов. Например, 5mC, по-видимому, отсутствует в тРНК и мРНК бактерий, при этом он был обнаружен в тРНК и мРНК эукариот и архей.

5hmC также был выявлен в РНК. Например, было обнаружено, что мРНК дрожозофилы и мыши содержит 5hmC. Сообщалось, что то же семейство ферментов, которые окисляют 5mC в ДНК, катализируют образование 5hmC в общей РНК млекопитающих. У мух в ходе широкого исследования транскриптома с применением иммунопреципитации и секвенирования метилированной РНК (MeRIP-seq) с антителами к 5hmC выявили присутствие 5hmC во многих кодирующих последовательностях мРНК с особенно высокими уровнями в головном мозге. Также сообщалось, что активная трансляция ассоциирована с высокими уровнями 5hmC в РНК, а мухи, у которых отсутствует фермент TET, отвечающий за накопление 5hmC в РНК, характеризуются нарушением развития головного мозга.

Существующим золотым стандартом и наиболее широко применяемым способом анализа метилирования и гидроксиметилирования ДНК является бисульфитное секвенирование (BS) и производные способы на его основе, такие как бисульфитное секвенирование с использованием Tet (TAB-Seq) и окислительное бисульфитное секвенирование (oxBS). Во всех этих способах используют обработку бисульфитом для превращения неметилированного цитозина в урацил, оставляя при этом интактными 5mC и/или 5hmC. Посредством ПЦР-амплификации ДНК, обработанной бисульфитом, в которой урацил прочитывается как тимин, информация о модификации каждого цитозина может быть получена при разрешении до одного основания (где превращение с заменой С на Т позволяет определить местоположение неметилированного цитозина). Однако существуют по меньшей мере два основных недостатка бисульфитного секвенирования. Во-первых, обработка бисульфитом является агрессивной химической реакцией, в ходе которой разрушается более 90% ДНК вследствие депуринизации в необходимых кислотных и температурных условиях. Это разрушение значительно ограничивает его применение для образцов с низким вводимым количеством, таких как клинические образцы, содержащие циркулирующую внеклеточную ДНК, и секвенирования отдельных клеток. Во-вторых, бисульфитное секвенирование основывается на полном превращении немодифицированного цитозина в тимин. Немодифицированный цитозин составляет примерно 95% общего цитозина в геноме человека. При превращении оснований во всех этих положениях в тимин значительно снижается сложность последовательности, что приводит к низкому качеству секвенирования, низким показателям картирования, неравномерному охвату генома и увеличе-

нию стоимости секвенирования. Способы бисульфитного секвенирования также чувствительны к ложному выявлению 5mC и 5hmC вследствие неполного превращения немодифицированного цитозина в тимин.

Бисульфитное секвенирование также применялось для выявления метилирования цитозина в РНК. В отличие от других способов выявления 5mC в РНК, таких как иммунопреципитация метилированной РНК, бисульфитное секвенирование РНК (RNA-BS-seq) обладает преимуществом, заключающимся в возможности определения степени метилирования С в конкретном положении в РНК. RNA-BS-seq, однако, страдает теми же недостатками, что описаны выше для бисульфитного секвенирования ДНК. В частности, условия реакции могут вызывать существенное разрушение РНК.

Существует потребность в способе анализа метилирования и гидроксиметилирования ДНК, который представлял бы собой мягкую реакцию, с помощью которой можно было бы проводить количественное выявление модифицированного цитозина (5mC и 5hmC) при разрешении до одного основания без воздействия на немодифицированный цитозин. Аналогичным образом, существует потребность в способе анализа метилирования и гидроксиметилирования РНК, в котором использовались бы мягкие условия реакции и можно было бы проводить количественное выявление модифицированного цитозина при разрешении до одного основания без воздействия на немодифицированный цитозин.

Краткое описание изобретения

В настоящем изобретении представлены способы идентификации местоположения одного или более из 5-метилцитозина, 5-гидроксиметилцитозина, 5-карбоксилцитозина и/или 5-формилцитозина в нуклеиновой кислоте. В способах, описанных в данном документе, предусмотрен анализ метилирования и гидроксиметилирования ДНК или РНК, включающий мягкие реакции, с помощью которых проводят количественное выявление модифицированного цитозина с разрешением до одного основания без воздействия на немодифицированный цитозин. В данном документе представлен новый способ идентификации 5mC и 5hmC путем объединения окисления с помощью ТЕТ и восстановления с помощью борановых производных (например, пиридинборана и 2-пиколинборана (пиколин-ВН₃)), называемый в данном документе TAPS (пиридинборановое секвенирование с использованием ТЕТ) (табл. 1). TAPS обеспечивает выявление модификаций напрямую с высокой чувствительностью и специфичностью без воздействия на немодифицированные цитозиновые основания и может применяться для выявления других модификаций цитозина. Этот способ является неразрушающим и обеспечивает сохранение РНК и ДНК длиной до 10 т.о. По сравнению с бисульфитным секвенированием TAPS обеспечивает более высокие показатели картирования, более равномерный охват и более низкую стоимость секвенирования, что позволяет проводить более высококачественные, более полные и более дешевые анализы метилома. Варианты этого способа, в которых не используется стадия окисления, применяют для идентификации 5fC и/или 5caC, как описано в данном документе.

В предпочтительном аспекте в настоящем изобретении представлен способ модификации нуклеиновой кислоты, включающий стадии:

а) обеспечения образца нуклеиновой кислоты, содержащего один или более остатков 5-карбоксилцитозина (5caC) или 5-формилцитозина (5fC); и

б) осуществления контакта указанного образца с восстанавливающим агентом на основе борана в условиях, которые обеспечивают восстановление 5caC и 5fC в дигидроурацил (DHU). Необязательно, стадия обеспечения (а) указанного способа включает подготовку указанного образца нуклеиновой кислоты посредством осуществления контакта целевой нуклеиновой кислоты, которая содержит один или более остатков 5-метилцитозина (5mC) или 5-гидроксиметилцитозина (5hmC), с ферментом транслокации десять-одиннадцать (ТЕТ), так что образуется один или более остатков 5caC или 5fC.

В предпочтительном аспекте в настоящем изобретении представлен способ модификации целевой нуклеиновой кислоты, включающий стадии:

а) осуществления контакта целевой нуклеиновой кислоты, которая содержит один или более остатков 5mC или 5hmC *in vitro* с ферментом ТЕТ с образованием препарата окисленной нуклеиновой кислоты; и затем

б) осуществления контакта указанного препарата окисленной нуклеиновой кислоты *in vitro* с восстанавливающим агентом на основе борана в условиях, которые обеспечивают восстановление 5caC и 5fC в дигидроурацил (DHU).

Необязательно, указанный восстанавливающий агент выбран из группы, состоящей из пиридинборана, 2-пиколинборана (пиколин-ВН₃), борогидрида натрия и триацетоксиборогидрида натрия. Необязательно, способ модификации дополнительно включает стадию секвенирования, причем предпочтительно стадия секвенирования позволяет различить модифицированный и немодифицированный цитозин в указанном образце нуклеиновой кислоты, более предпочтительно стадия секвенирования позволяет определить количественный уровень одной или более модификаций цитозина в указанном образце нуклеиновой кислоты. Стадия секвенирования может включать одно или более из секвенирования методом обрыва цепи, микроматричного анализа, высокопроизводительного секвенирования и анализа с помощью ферментов рестрикции.

Необязательно, способ модификации дополнительно включает стадию блокирования одного или

более модифицированных цитозинон, предпочтительно указанная стадия блокирования включает добавление сахара к 5hmC, где указанный сахар предпочтительно представляет собой глюкозу или модифицированную глюкозу.

В способе модификации, фермент TET предпочтительно выбран из группы, состоящей из TET1, TET2 и TET3 человека; Tet1, Tet2 и Tet3 мыши; TET Naegleria (NgTET); и TET Coprinopsis cinerea (CcTET).

Необязательно, стадия обеспечения (а) включает подготовку указанного образца нуклеиновой кислоты посредством осуществления контакта целевой нуклеиновой кислоты, которая содержит один или более остатков 5hmC, с химическим окисляющим агентом, так что образуется один или более остатков 5fC, где указанный окисляющий агент предпочтительно представляет собой перрутат калия или Cu (II)/TEMPO.

Необязательно, стадия блокирования включает осуществление контакта образца нуклеиновой кислоты, содержащего один или более остатков 5fC, с альдегид-реактивным соединением, выбранным из производных гидросиламина, производных гидразина и производных гидразида, причем производное гидросиламина предпочтительно представляет собой О-этилгидросиламин.

Необязательно, стадия блокирования включает осуществление контакта образца нуклеиновой кислоты, содержащего один или более остатков 5caC, с реагентом для дериватизации карбоновой кислоты; и соединением амина, гидразина или гидросиламина, предпочтительно указанная стадия блокирования включает осуществление контакта образца нуклеиновой кислоты, содержащего один или более остатков 5caC, с 1-этил-3-(3-диметиламинопропил) карбодииимидом (EDC) и этиламином.

Необязательно, указанный способ дополнительно включает стадию амплификации количества копий одной или более последовательностей нуклеиновых кислот.

В предпочтительном аспекте в настоящем изобретении представлен набор для модификации нуклеиновой кислоты, содержащий в отдельных контейнерах:

а) фермент семейства TET; и

б) восстанавливающий агент на основе борана, который способен восстанавливать 5caC и 5fC до дигидроурацила (DHU).

Предпочтительно, набор дополнительно содержит по меньшей мере одно из:

i) фермент глюкозилтрансфераза и подходящий реагент для блокирования 5hmC;

ii) один или более реагентов для блокирования 5fC; и/или

iii) один или более реагентов для блокирования 5caC.

В одном аспекте в настоящем изобретении представлен способ идентификации 5-метилцитозина (5mC) в целевой нуклеиновой кислоте, включающий стадии:

а) получения образца нуклеиновой кислоты, содержащего целевую нуклеиновую кислоту;

б) модификации нуклеиновой кислоты, включающей стадии:

i) добавления блокирующей группы к 5-гидроксиметилцитозину (5hmC) в образце нуклеиновой кислоты;

ii) превращения 5mC в образце нуклеиновой кислоты в 5-карбоксилцитозин (5caC) и/или 5-формилцитозин (5fC) и

iii) превращения 5caC и/или 5fC в дигидроурацил (DHU) с получением модифицированного образца нуклеиновой кислоты, содержащего модифицированную целевую нуклеиновую кислоту; и

с) выявления последовательности модифицированной целевой нуклеиновой кислоты; где превращение с заменой цитозина (C) на тимин (T) в последовательности модифицированной целевой нуклеиновой кислоты по сравнению с целевой нуклеиновой кислотой позволяет определить местоположение 5mC в целевой нуклеиновой кислоте.

В вариантах осуществления способа идентификации 5mC в целевой нуклеиновой кислоте значения процентного содержания T в каждом местоположении превращения позволяют определить количественный уровень 5mC в каждом местоположении в целевой нуклеиновой кислоте. В вариантах осуществления нуклеиновая кислота представляет собой ДНК. В других вариантах осуществления нуклеиновая кислота представляет собой РНК.

В другом аспекте в настоящем изобретении представлен способ идентификации 5mC или 5hmC в целевой нуклеиновой кислоте, включающий стадии:

а) получения образца нуклеиновой кислоты, содержащего целевую нуклеиновую кислоту;

б) модификации нуклеиновой кислоты, включающей стадии:

i) превращения 5mC и 5hmC в образце нуклеиновой кислоты в 5-карбоксилцитозин (5caC) и/или 5fC и

ii) превращения 5caC и/или 5fC в DHU с получением модифицированного образца нуклеиновой кислоты, содержащего модифицированную целевую нуклеиновую кислоту; и

с) выявления последовательности модифицированной целевой нуклеиновой кислоты; где превращение с заменой цитозина (C) на тимин (T) в последовательности модифицированной целевой нуклеиновой кислоты по сравнению с целевой нуклеиновой кислотой позволяет определить местоположение 5mC либо 5hmC в целевой нуклеиновой кислоте.

В вариантах осуществления способа идентификации 5mC или 5hmC значения процентного содержания Т в каждом местоположении превращения позволяют определить количественный уровень 5mC или 5hmC в каждом местоположении в целевой нуклеиновой кислоте. В вариантах осуществления нуклеиновая кислота представляет собой ДНК. В других вариантах осуществления нуклеиновая кислота представляет собой РНК.

В другом аспекте в настоящем изобретении представлен способ идентификации 5mC и идентификации 5hmC в целевой нуклеиновой кислоте, включающий:

- a) идентификацию 5mC в целевой нуклеиновой кислоте, включающую стадии:
 - i) получения первого образца нуклеиновой кислоты, содержащего целевую нуклеиновую кислоту;
 - ii) модификации нуклеиновой кислоты в первом образце, включающей стадии:
 - 1) добавления блокирующей группы к 5-гидроксиметилцитозину (5hmC) в первом образце нуклеиновой кислоты;
 - 2) превращения 5mC в первом образце нуклеиновой кислоты в 5caC и/или 5fC и
 - 3) превращения 5caC и/или 5fC в DHU с получением модифицированного первого образца ДНК, содержащего модифицированную целевую нуклеиновую кислоту;
 - iii) необязательно амплификации числа копий модифицированной целевой нуклеиновой кислоты и
 - iv) выявления последовательности модифицированной целевой нуклеиновой кислоты; где превращение с заменой цитозина (С) на тимин (Т) в последовательности модифицированной целевой нуклеиновой кислоты по сравнению с целевой нуклеиновой кислотой позволяет определить местоположение 5mC в целевой нуклеиновой кислоте;
 - b) идентификацию 5mC или 5hmC в целевой нуклеиновой кислоте, включающую стадии:
 - i) получения второго образца нуклеиновой кислоты, содержащего целевую нуклеиновую кислоту;
 - ii) модификации нуклеиновой кислоты во втором образце, включающей стадии:
 - 1) превращения 5mC и 5hmC во втором образце нуклеиновой кислоты в 5caC и/или 5fC и
 - 2) превращения 5caC и/или 5fC в DHU с получением модифицированного второго образца нуклеиновой кислоты, содержащего модифицированную целевую нуклеиновую кислоту;
 - iii) необязательно амплификации числа копий модифицированной целевой нуклеиновой кислоты;
 - iv) выявления последовательности модифицированной целевой нуклеиновой кислоты из второго образца; где превращение с заменой цитозина (С) на тимин (Т) в последовательности модифицированной целевой нуклеиновой кислоты по сравнению с целевой нуклеиновой кислотой позволяет определить местоположение 5mC либо 5hmC в целевой нуклеиновой кислоте; и
 - c) сравнение результатов стадий (a) и (b), где превращения с заменой С на Т, присутствующие на стадии (b), но не на стадии (a), позволяют определить местоположение 5hmC в целевой нуклеиновой кислоте.

В вариантах осуществления для идентификации 5mC и идентификации 5hmC в целевой нуклеиновой кислоте на стадии (a) значения процентного содержания Т в каждом местоположении превращения позволяют определить количественный уровень 5mC в целевой нуклеиновой кислоте; на стадии (b) значения процентного содержания Т в каждом местоположении превращения позволяют определить количественный уровень 5mC или 5hmC в целевой нуклеиновой кислоте; и на стадии (c) различия в значениях процентного содержания, соответствующие превращениям с заменой С на Т, идентифицированным на стадии (b), но не на стадии (a), позволяют определить количественный уровень 5hmC в каждом местоположении в целевой нуклеиновой кислоте. В вариантах осуществления нуклеиновая кислота представляет собой ДНК. В других вариантах осуществления нуклеиновая кислота представляет собой РНК.

В вариантах осуществления настоящего изобретения блокирующая группа, добавляемая к 5hmC в образце нуклеиновой кислоты, представляет собой сахар. В вариантах осуществления сахар представляет собой встречающийся в природе сахар или модифицированный сахар, например, глюкозу или модифицированную глюкозу. В вариантах осуществления настоящего изобретения блокирующую группу добавляют к 5hmC путем приведения образца нуклеиновой кислоты в контакт с UDP, связанным с сахаром, например, UDP-глюкозой или UDP, связанным с модифицированной глюкозой, в присутствии фермента глюкозилтрансферазы, например, β-глюкозилтрансферазы бактериофага Т4 (βGT) и α-глюкозилтрансферазы бактериофага Т4 (αGT) и их производных и аналогов.

В вариантах осуществления настоящего изобретения каждая из стадий превращения 5mC в образце нуклеиновой кислоты в 5caC и/или 5fC и стадии превращения 5mC и 5hmC в образце нуклеиновой кислоты в 5caC и/или 5fC включает приведение образца нуклеиновой кислоты в контакт с ферментом транслокации десять-одиннадцать (TET). В дополнительных вариантах осуществления фермент TET представляет собой один или более из TET1, TET2 и TET3 человека; Tet1, Tet2 и Tet3 мыши; TET *Naegleria* (NgTET); *Coprinopsis cinerea* (CcTET) и их производных или аналогов. В вариантах осуществления фермент TET представляет собой NgTET.

В другом аспекте в настоящем изобретении представлен способ идентификации 5caC или 5fC в целевой нуклеиновой кислоте, включающий стадии:

- a) получения образца нуклеиновой кислоты, содержащего целевую нуклеиновую кислоту;

b) превращения 5caC и 5fC в DHU с получением модифицированного образца нуклеиновой кислоты, содержащего модифицированную целевую нуклеиновую кислоту;

c) необязательно амплификации числа копий модифицированной целевой нуклеиновой кислоты и

d) выявления последовательности модифицированной целевой нуклеиновой кислоты; где превращение с заменой цитозина (C) на тимин (T) в последовательности модифицированной целевой нуклеиновой кислоты по сравнению с целевой нуклеиновой кислотой позволяет определить местоположение 5caC либо 5fC в целевой нуклеиновой кислоте.

В вариантах осуществления способа идентификации 5caC или 5fC в целевой нуклеиновой кислоте значения процентного содержания T в каждом местоположении превращения позволяют определить количественный уровень 5caC или 5fC в каждом местоположении в целевой нуклеиновой кислоте.

В другом аспекте в настоящем изобретении представлен способ идентификации 5caC в целевой нуклеиновой кислоте, включающий стадии:

a) получения образца нуклеиновой кислоты, содержащего целевую нуклеиновую кислоту;

b) добавления блокирующей группы к 5fC в образце нуклеиновой кислоты;

c) превращения 5caC в DHU с получением модифицированного образца нуклеиновой кислоты, содержащего модифицированную целевую нуклеиновую кислоту;

d) необязательно амплификации числа копий модифицированной целевой нуклеиновой кислоты и

e) определения последовательности модифицированной целевой нуклеиновой кислоты; где превращение с заменой цитозина (C) на тимин (T) в последовательности модифицированной целевой нуклеиновой кислоты по сравнению с целевой нуклеиновой кислотой позволяет определить местоположение 5caC в целевой нуклеиновой кислоте.

В вариантах осуществления способа идентификации 5caC в целевой нуклеиновой кислоте значения процентного содержания T в каждом местоположении превращения позволяют определить количественный уровень 5caC в каждом местоположении в целевой нуклеиновой кислоте. В вариантах осуществления нуклеиновая кислота представляет собой ДНК. В других вариантах осуществления нуклеиновая кислота представляет собой РНК.

В вариантах осуществления настоящего изобретения добавление блокирующей группы к 5fC в образце нуклеиновой кислоты включает приведение нуклеиновой кислоты в контакт с альдегид-реактивным соединением, в том числе, например, производными гидроксилamina (такими как О-этилгидроксилamin), производными гидразина и производными гидразидов.

В другом аспекте в настоящем изобретении представлен способ идентификации 5fC в целевой нуклеиновой кислоте, включающий стадии:

a) получения образца нуклеиновой кислоты, содержащего целевую нуклеиновую кислоту;

b) добавления блокирующей группы к 5caC в образце нуклеиновой кислоты;

c) превращения 5fC в DHU с получением модифицированного образца нуклеиновой кислоты, содержащего модифицированную целевую нуклеиновую кислоту;

d) необязательно амплификации числа копий модифицированной целевой нуклеиновой кислоты;

e) выявления последовательности модифицированной целевой нуклеиновой кислоты; где превращение с заменой цитозина (C) на тимин (T) в последовательности модифицированной целевой нуклеиновой кислоты по сравнению с целевой нуклеиновой кислотой позволяет определить местоположение 5fC в целевой нуклеиновой кислоте.

В вариантах осуществления способа идентификации 5fC в целевой нуклеиновой кислоте значения процентного содержания T в каждом местоположении превращения позволяют определить количественный уровень 5fC в каждом местоположении в целевой нуклеиновой кислоте. В вариантах осуществления нуклеиновая кислота (в образце и целевая) представляет собой ДНК. В других вариантах осуществления нуклеиновая кислота (в образце и целевая) представляет собой РНК.

В вариантах осуществления стадия добавления блокирующей группы к 5caC в образце нуклеиновой кислоты включает приведение образца нуклеиновой кислоты в контакт с реагентом для дериватизации карбоновых кислот, в том числе, например, 1-этил-3-(3-диметиламинопропил)карбодимидом (EDC), и (ii) аминным (таким как этиламин), гидразиновым или гидроксилaminовым соединением.

В вариантах осуществления настоящего изобретения вышеуказанные способы дополнительно включают стадию амплификации числа копий модифицированной целевой нуклеиновой кислоты. В вариантах осуществления такую стадию амплификации проводят перед стадией выявления последовательности модифицированной целевой нуклеиновой кислоты. Стадию амплификации числа копий в случае, если модифицированная целевая нуклеиновая кислота представляет собой ДНК, можно осуществлять путем проведения полимеразной цепной реакции (ПЦР), достройки праймера и/или клонирования. В случае, если модифицированная целевая нуклеиновая кислота представляет собой РНК, стадию амплификации числа копий можно осуществлять с помощью RT-PCR с использованием праймера олиго(dT) (для мРНК), случайных праймеров и/или геноспецифичных праймеров.

В вариантах осуществления настоящего изобретения образец ДНК содержит пикограммовые количества ДНК. В вариантах осуществления настоящего изобретения образец ДНК содержит от приблизительно 1 пг до приблизительно 900 пг ДНК, от приблизительно 1 пг до приблизительно 500 пг ДНК, от

приблизительно 1 пг до приблизительно 100 пг ДНК, от приблизительно 1 пг до приблизительно 50 пг ДНК, от приблизительно 1 до приблизительно 10 пг ДНК, менее чем приблизительно 200 пг, менее чем приблизительно 100 пг ДНК, менее чем приблизительно 50 пг ДНК, менее чем приблизительно 20 пг ДНК и менее чем приблизительно 5 пг ДНК. В других вариантах осуществления настоящего изобретения образец ДНК содержит нанограммовые количества ДНК. В вариантах осуществления настоящего изобретения образец ДНК содержит от приблизительно 1 до приблизительно 500 нг ДНК, от приблизительно 1 до приблизительно 200 нг ДНК, от приблизительно 1 до приблизительно 100 нг ДНК, от приблизительно 1 до приблизительно 50 нг ДНК, от приблизительно 1 нг до приблизительно 10 нг ДНК, от приблизительно 1 нг до приблизительно 5 нг ДНК, менее чем приблизительно 100 нг ДНК, менее чем приблизительно 50 нг ДНК, менее чем приблизительно 5 нг ДНК или менее чем приблизительно 2 нг ДНК. В вариантах осуществления настоящего изобретения образец ДНК содержит циркулирующую внеклеточную ДНК (cfDNA). В вариантах осуществления настоящего изобретения образец ДНК содержит микрограммовые количества ДНК.

В вариантах осуществления настоящего изобретения стадия превращения 5сaC и/или 5fC в DHU включает приведение образца нуклеиновой кислоты в контакт с восстановителем, в том числе, например, пиридинбораном, 2-пиколинбораном (пиколин-BH₃), бораном, борогидридом натрия, цианоборогидридом натрия и триацетоксиборогидридом натрия. В предпочтительном варианте осуществления восстановитель представляет собой пиколин-BH₃ и/или пиридинборан.

В вариантах осуществления настоящего изобретения стадия определения последовательности модифицированной целевой нуклеиновой кислоты включает секвенирование методом обрыва цепи, микроматричный анализ, высокопроизводительное секвенирование и анализ с помощью ферментов рестрикции.

Краткое описание графических материалов

Фиг. 1. Скрининг бораносодержащих соединений. Бораносодержащие соединения подвергали скринингу в отношении превращения 5сaC в DHU в 11-мерном олигонуклеотиде ("олиго") с частотой превращения, оцениваемой посредством MALDI.

2-Пиколинборан (пиколинборан), пиридинборан и трет-бутиламиноборан могли обеспечивать полное превращение 5сaC в DHU, тогда как этилендиаминборан и диметиламиноборан давали частоту превращения около 30%. При использовании дициклогексиламиноборана, морфолиноборана, 4-метилморфолиноборана и триметиламиноборана отсутствовали измеряемые продукты, поддающиеся выявлению (не выявлено). Другие восстановители, такие как борогидрид натрия и три(ацетокси)борогидрид натрия, быстро разлагались в кислой среде и приводили к неполному превращению. Цианоборогидрид натрия не использовали из-за возможного образования цианистого водорода в кислых условиях. Пиколинборан и пиридинборан были выбраны ввиду обеспечения ими полного превращения, низкой токсичности и высокой стабильности.

Фиг. 2А, В. Реакция пиколинборана с ДНК-олигонуклеотидами. (А) Определение характеристик 11-мерной модельной ДНК, содержащей 5сaC, обработанной пиколинбораном, посредством MALDI. Расчетная масса (масса/заряд) показана над каждым графиком, наблюдаемая масса показана слева от пика. (В) Частота превращения dC и различных производных цитозина определяли количественно с помощью HPLC-MS/MS. Данные показаны как среднее значение ± стандартное отклонение для трех повторностей.

Фиг. 3А, В. Реакция пиколинборана с отдельным нуклеозидом. Результаты ¹H- и ¹³C-ЯМР соответствовали опубликованному ранее сообщению о 2'-дезоксиде-5,6-дигидроуридине (I. Aparici-Espert et al., J. Org. Chem. 81, 4031-4038 (2016)). (А) Диаграмма ¹H-ЯМР (MeOH-d₄, 400 МГц) продукта реакции пиколинборана с отдельным нуклеозидом. δ ppm: 6,28 (t, 1H, J=7 Гц), 4,30 (m, 1H), 3,81 (m, 1H), 3,63 (m, 2H), 3,46 (m, 2H), 2,65 (t, 2H, J=6 Гц), 2,20 (m, 1H), 2,03 (m, 1H). (В) Диаграмма ¹³C-ЯМР (MeOH-d₄, 400 МГц) продукта реакции пиколинборана с отдельным нуклеозидом. δ ppm: 171,56 (CO), 153,54 (CO), 85,97 (CH), 83,86 (CH), 70,99 (CH), 61,92 (CH₂), 36,04 (CH₂), 35,46 (CH₂), 30,49 (CH₂).

Фиг. 4А, В. Диаграмма, на которой показано (А) превращение 5сaC в DHU под действием борана и предполагаемый механизм реакции превращения 5сaC в DHU под действием борана; а также (В) превращение 5fC в DHU под действием борана и предполагаемый механизм реакции превращения 5fC в DHU под действием борана.

Фиг. 5А, В. (А) Диаграмма, на которой показано, что в ходе способа TAPS происходит превращение как 5mC, так и 5hmC в DHU, который при репликации выступает как тимин. (В) Обзор способов TAPS, TAPSβ и CAPS.

Фиг. 6. Определение характеристик модельных ДНК-олигонуклеотидов, содержащих 5fC и 5сaC, обработанных пиколинбораном с блокированием 5fC и 5сaC или без него, посредством MALDI. 5fC и 5сaC превращаются в дигидроурацил (DHU) под действием пиколин-BH₃. 5fC блокировали производными гидроксилamina, такими как О-этилгидроксиламин (EtONH₂), которые преобразовывались в оксим и препятствовали превращению под действием пиколинборана. 5сaC блокировали этиламиноном посредством конъюгирования с EDC и превращали в амид, который блокирует превращение под действием пиколинборана. Расчетный MS (масса/заряд) показан над каждым графиком, наблюдаемый MS показан слева от пика.

Фиг. 7. Определение характеристик модельных ДНК-олигонуклеотидов, содержащих 5mC и 5hmC, обработанных с помощью $KRuO_4$ и пиколинборана с блокированием 5hmC или без него, посредством MALDI. 5hmC можно блокировать глюкозой с помощью βGT и превращать в 5gmC. 5mC, 5hmC и 5gmC не могут превращаться под действием пиколинборана. 5hmC может окисляться под действием $KRuO_4$ до 5fC, а затем превращаться в DHU под действием пиколинборана. Расчетный MS (масса/заряд) показан над каждым графиком, наблюдаемый MS показан слева от пика.

Фиг. 8А, В. Расщепление ферментами рестрикции показало, что в ходе TAPS может происходить эффективное превращение 5mC в Т. (А) Иллюстрация анализа расщепления ферментами рестрикции для подтверждения изменения последовательности, обусловленного TAPS. (В) Тесты расщепления под действием $Taq\alpha I$ для подтверждения превращения С-Т, обусловленного TAPS. TAPS проводили в отношении модельной ДНК размером 222 п.о., имеющей сайт рестрикции для $Taq\alpha I$ и содержащей 5 полностью метилированных сайтов CpG (5mC), и ее неметилированного контроля (С). Амплифицированная посредством ПЦР модельная ДНК размером 222 п.о. может расщепляться под действием $Taq\alpha I$ до фрагментов размером ~ 160 п.о. и ~ 60 п.о., как показано для 5mC, С и С-TAPS. После проведения TAPS в отношении метилированной ДНК последовательность T(mC)GA превращается в TTGA и больше не расщепляется путем расщепления под действием $Taq\alpha I$, как показано на дорожках 5mC-TAPS.

Фиг. 9А, В. TAPS в отношении модельной ДНК размером 222 п.о. и gDNA mESC. (А) Результаты секвенирования по Сэнгеру для модельной ДНК размером 222 п.о., содержащей 5 полностью метилированных сайтов CpG, и ее неметилированного контроля до (5mC, С) и после TAPS (5mC-TAPS, С-TAPS). С помощью способа TAPS только 5mC превращается в Т. (В) Количественная оценка относительных уровней модификаций в контрольной gDNA mESC посредством HPLC-MS/MS после окисления с помощью $NgTET1$ и после восстановления с помощью пиколинборана. Данные показаны как среднее значение \pm стандартное отклонение для трех повторностей.

Фиг. 10А-Д. TAPS не обуславливало существенного разрушения ДНК по сравнению с бисульфитом. Изображения агарозного геля с неметилированной ДНК размером 222 п.о., метилированной ДНК размером 222 п.о. и gDNA mESC (А) до и (В) после охлаждения на ледяной бане. После TAPS не наблюдалось поддающееся выявлению разрушение ДНК, и ДНК оставалась двухцепочечной и могла быть визуализирована без охлаждения. Превращение посредством бисульфита приводило к разрушению, и ДНК становилась одноцепочечной и могла быть визуализирована только после охлаждения на льду. (С) Изображения агарозного геля с gDNA mESC с различной длиной фрагментов, обработанной с помощью TAPS и бисульфита, до (левая панель) и после (правая панель) охлаждения на льду. ДНК после TAPS оставалась двухцепочечной и могла быть напрямую визуализирована в геле. Обработка бисульфитом вызывала большее повреждение и фрагментацию образцов, и ДНК становилась одноцепочечной и могла быть визуализирована только после охлаждения на льду. Превращение посредством TAPS было полным для всех gDNA независимо от длины фрагментов, как показано на фиг. 15. (Д) Визуализация в агарозном геле модельной ДНК размером 222 п.о. до и после TAPS (три независимые повторности) не показала поддающегося выявлению разрушения после реакции.

Фиг. 11. Сравнение кривых амплификации и кривых плавления между модельными ДНК до и после TAPS. Анализ методом qPCR показал незначительное различие в кривых амплификации модельных ДНК до и после TAPS. Кривая плавления метилированной ДНК (5mC) сдвигалась в сторону более низкой температуры после TAPS, что указывало на возможную транзицию С-Т, приводящую к уменьшению T_m , тогда как для неметилированной ДНК (С) сдвиг не наблюдался.

Фиг. 12. Полное превращение С-Т, индуцируемая после TAPS, TAPS β и CAPS, как продемонстрировано с помощью секвенирования по Сэнгеру.

Модельную ДНК, содержащую одиночные метилированные и одиночные гидроксиметилированные сайты CpG, получали, как описано в данном документе. Превращение посредством TAPS выполняли в соответствии с протоколом окисления с помощью $NgTET1$ и восстановления с помощью пиридинборана, как описано в данном документе. Превращение посредством TAPS β выполняли в соответствии с протоколом блокирования 5hmC, окисления с помощью $NgTET1$ и восстановления с помощью пиридинборана. Превращение посредством CAPS выполняли в соответствии с протоколом окисления 5hmC и восстановления с помощью пиридинборана. После превращения 1 нг образца ДНК, подвергнутого превращению, подвергали ПЦР-амплификации с помощью ДНК-полимеразы Taq и обрабатывали для секвенирования по Сэнгеру. В ходе TAPS происходило превращение как 5mC, так и 5hmC в Т. В ходе TAPS β происходило избирательное превращение 5mC, тогда как в ходе CAPS происходило избирательное превращение 5hmC. Ни один из трех способов не обуславливал превращение немодифицированного цитозина и других оснований.

Фиг. 13А, В. (А) TAPS совместимо с различными ДНК- и РНК-полимеразами и индуцирует полную транзицию С-Т, как показано с помощью секвенирования по Сэнгеру. Модельная ДНК, содержащая метилированные сайты CpG для полимеразного теста, и последовательности праймеров описаны в данном документе. После обработки посредством TAPS 5mC превращался в DHU. Полимераза КАРА HiFi Uracil Plus, полимеразы Taq и экзополимераза Vent будут считывать DHU как Т и, следовательно, индуцировать

полную транзицию С-Т после ПЦР. В качестве альтернативы, достройку праймера выполняли с помощью праймера, меченного биотином, и полимераз для изотермической амплификации, в том числе фрагмента Кленова, ДНК-полимеразы Bst и ДНК-полимеразы phi29. Новосинтезированную цепь ДНК отделяли с помощью Dynabeads MyOne C1 со стрептавидином, а затем амплифицировали с помощью ПЦР с использованием полимеразы Taq и обрабатывали для секвенирования по Сэнгеру. РНК-полимераза T7 может эффективно обходить DHU и вставлять аденин напротив сайта DHU, что подтверждается с помощью RT-PCR и секвенирования по Сэнгеру. (В)

Некоторые другие коммерциализированные полимеразы не обеспечивали эффективную амплификацию ДНК, содержащей DHU. После обработки посредством TAPS 5mC превращался в DHU. Полимераза КАРА HiFi Utagil Plus и полимеразы Taq будут считывать DHU как Т и, следовательно, индуцировать полную транзицию С-Т. При использовании некоторых других коммерциализированных полимераз, в том числе полимеразы КАРА HiFi, полимеразы Pfu, полимеразы Phusion и полимеразы NEB Q5, наблюдались низкая частота или отсутствие транзиций С-Т (не показано).

Фиг. 14. Для DHU не демонстрируется ПЦР-сдвиг по сравнению с Т и С.

Модельная ДНК, содержащая одну модификацию DHU/U/T/C, была синтезирована с соответствующими ДНК-олигонуклеотидами, как описано в данном документе. Калибровочные кривые для каждой модельной ДНК с модификацией DHU/U/T/C строили на основании реакций qPCR для введенной модельной ДНК в серийных разведениях 1:10 (от 0,1 пг до 1 нг, каждый эксперимент по qPCR проводили в трех повторностях). Наклон линии регрессии для зависимости между величинами логарифма концентрации (нг) и средними величинами Ct рассчитывали с помощью функции НАКЛОН в Excel. Эффективность ПЦР рассчитывали с помощью следующего уравнения: % эффективности = $(10^{(-1/\text{наклон})} - 1) * 100\%$. Коэффициент амплификации рассчитывали с помощью следующего уравнения: Коэффициент амплификации = $10^{(-1/\text{наклон})}$. Эффективность ПЦР для модельной ДНК с модификацией DHU, или Т, или С была практически одинаковой, что продемонстрировало, что DHU может прочитываться как обычное основание и не будет вызывать ПЦР-сдвиг.

Фиг. 15А, В. В ходе TAPS происходило полное превращение 5mC в Т независимо от длины фрагментов ДНК. (А) Изображения в агарозном геле в анализе расщепления под действием TaqOI подтверждали полное превращение 5mC в Т во всех образцах независимо от длины фрагментов ДНК. Модельную последовательность длиной 194 п.о. из генома лямбда подвергали ПЦР-амплификации после TAPS и расщепляли ферментом TaqOI. Продукт ПЦР, амплифицированный из образца, не подвергнутого превращению, мог расщепляться, тогда как продукты, амплифицированные из образцов, обработанных посредством TAPS, оставались интактными, что позволяет предположить утрату сайта рестрикции и, следовательно, полную транзицию 5mC-Т. (В) Процент превращения С-Т оценивали путем количественной оценки полос геля, и было показано, что он составляет 100% для всех тестируемых длин фрагментов ДНК.

Фиг. 16. Превращение и ложноположительные результаты для различных условий TAPS. При использовании комбинации mTet1 и пиридинборана достигалась наиболее высокая частота превращения метилированного С (96,5%, рассчитанная для ДНК лямбда с полностью метилированными CpG) и наиболее низкая частота превращения немодифицированного С (0,23%, рассчитанная для немодифицированного внутреннего стандарта размером 2 т.о.) по сравнению с другими условиями с использованием NgTET1 или пиколинборана. Над столбцами показана частота превращения +/- SE для всех тестируемых цитозинового сайтов.

Фиг. 17А, В. Частота превращения в коротких внутренних стандартах. (А) 120-мер-1 и (В) 120-мер-2, содержащие 5mC и 5hmC. В сайтах 5mC и 5hmC в обоих цепях достигалось почти полное превращение. Приведены фактические последовательности со статусом модификации, показанным сверху и внизу.

Фиг. 18А-Е. Улучшенное качество секвенирования TAPS по сравнению с полногеномным бисульфитным секвенированием (WGBS). (А) Частота превращения 5mC и 5hmC в ДНК, обработанной посредством TAPS. Слева: синтетические внутренние стандарты (CpN), метилированные или гидроксиметилированные в известных положениях. (В) Частота ложноположительных результатов TAPS для немодифицированного внутреннего стандарта размером 2 т.о. (С) Общее время проведения TAPS и WGBS при обработке 1 миллиона смоделированных ридов на одном ядре одного CPU Intel Xeon. (D) Доля всех секвенированных пар ридов (после обрезания), картированных на геном. (Е) Показатели качества секвенирования в расчете на основание для первого и второго ридов во всех секвенированных парах ридов согласно сообщениям Illumina BaseSpace. Вверху: TAPS. Внизу: WGBS.

Фиг. 19А, В. TAPS обеспечивал более равномерный охват и меньшее количество неохваченных положений, чем WGBS. Сравнение глубины охвата по (А) всем основаниям и (В) сайтам CpG между WGBS и TAPS, вычисленное для обеих цепей. Для "TAPS (с уменьшением размера выборки)" выбирали случайные риды из всех картированных ридов TAPS таким образом, что медианный охват соответствовал медианному охвату WGBS. Положения с охватом выше 50x показаны в последней секции.

Фиг. 20. Распределение уровней модификаций по всем хромосомам.

Средние уровни модификаций в окнах размером 100 т.о. вдоль хромосом мыши, взвешенные по охвату CpG и сглаженные с помощью гауссовского фильтра взвешенного скользящего среднего с размером

окна 10.

Фиг. 21А-Е. Сравнение полногеномных измерений метилома посредством TAPS и WGBS. (А) Средняя глубина охвата секвенирования во всех CpG-островках мыши (сгруппированных в 20 окон) и фланкирующих областях размером 4 т. п.о. (сгруппированных в 50 окон одинакового размера). Чтобы учесть различия в глубине секвенирования, все картированные риды TAPS были подвергнуты уменьшению размера выборки для соответствия медианному количеству картированных ридов WGBS по всему геному. (В) Сайты CpG, охватываемые по меньшей мере тремя ридами, выявляемые посредством TAPS в отдельности, TAPS и WGBS совместно или WGBS в отдельности. (С) Количество сайтов CpG, охватываемых по меньшей мере тремя ридами и характеризующихся уровнем модификаций $>0,1$, выявляемых посредством TAPS в отдельности, TAPS и WGBS или WGBS в отдельности. (D) Пример распределения уровней модификаций (в %) по хромосомам для TAPS и WGBS. Приведена средняя доля модифицированных CpG в расчете на окно размером 100 т.о. вдоль хромосомы 4 мыши, сглаженное с помощью гауссовского фильтра взвешенного скользящего среднего с размером окна 10. (Е) Тепловая карта, на которой представлено количество сайтов CpG, охватываемых по меньшей мере тремя ридами в TAPS и WGBS совместно, с разбивкой по уровням модификаций, измеренным с помощью каждого способа. Для улучшения контраста первая секция, содержащая немодифицированные CpG в обоих способах, была исключена из цветовой шкалы и обозначена звездочкой.

Фиг. 22. Уровни модификаций вокруг CpG-островков. Приведены средние уровни модификаций в CpG-островках мыши (сгруппированных в 20 окон) и фланкирующих областях размером 4 т. п.о. (сгруппированных в 50 окон одинакового размера). Секции с охватом менее 3 ридов не принимались во внимание.

Фиг. 23А, В. TAPS демонстрирует меньший сдвиг зависимости уровня модификаций от охвата, чем WGBS. Все сайты CpG были сгруппированы в секции в соответствии с их охватом, и в каждой секции для WGBS (А) и TAPS (В) показана средняя (кружки) и медианная (треугольники) величина модификации. Сайты CpG, охватываемые более чем 100 ридами, показаны в последней секции. Линиями представлена линейная аппроксимация точек данных.

Фиг. 24А-С. Библиотеки gDNA с низким вводимым количеством и внеклеточной ДНК для TAPS, полученные с помощью наборов для получения библиотек dsDNA и ssDNA. (А) Библиотеки для секвенирования были успешно сконструированы с использованием до 1 нг геномной ДНК (gDNA) эмбриональных стволовых клеток мыши (mESC) с помощью наборов для получения библиотек dsDNA NEBNext Ultra II или наборов KAPA HyperPlus. Набор для получения библиотек ssDNA Accel-NGS Methyl-Seq использовали для дополнительного снижения количества вводимой ДНК до (В) 0,01 нг gDNA mESC или (С) 1 нг внеклеточной ДНК.

Фиг. 25А, В. Библиотеки gDNA с низким вводимым количеством и внеклеточной ДНК для TAPS, полученные с помощью набора для получения библиотек dsDNA KAPA HyperPlus. Библиотеки для секвенирования были успешно сконструированы с использованием всего лишь 1 нг (А) gDNA mESC и (В) внеклеточной ДНК с помощью набора KAPA HyperPlus. Внеклеточная ДНК характеризуется отчетливым распределением длин около 160 п.о. (в соответствии с размером нуклеосомы) благодаря расщеплению нуклеазами плазмы крови. После конструирования библиотеки ее размер становится равным ~ 300 п.о., что соответствует отчетливой полосе в (В).

Фиг. 26А-Д. Высококачественная TAPS для внеклеточной ДНК. (А)

Частота превращения 5mC в cfDNA, обработанной посредством TAPS. (В) Частота ложноположительных результатов в cfDNA, обработанной посредством TAPS. (С) Доля всех секвенированных пар ридов, которые были уникальным образом картированы на геном. (D) Доля всех секвенированных пар ридов, которые были уникальным образом картированы на геном, после удаления ПЦР-дублицированных ридов. CHG и CHH относятся к окружению, отличному от окружения CpG.

Фиг. 27. С помощью TAPS можно выявлять генетические варианты.

Анализ метилирования (MOD, верхний ряд) и SNP С-Т (нижний ряд) показал различные паттерны распределения оснований в исходной верхней цепи (OT)/исходной нижней цепи (OB) в левом столбце и в цепях, комплементарных OT (CTOT) и OB (CTOB), в правом столбце.

Подробное описание изобретения

В настоящем изобретении представлен способ выявления 5mC и 5hmC в последовательности без применения бисульфита с разрешением до одного основания, который в данном документе называется TAPS. TAPS включает мягкие ферментативные и химические реакции для прямого и количественного выявления 5mC и 5hmC при разрешении до одного основания без воздействия на немодифицированный цитозин. В настоящем изобретении также представлены способы выявления 5fC и 5caC при разрешении до одного основания без воздействия на немодифицированный цитозин. Таким образом, способы, представленные в данном документе, обеспечивают картирование 5mC, 5hmC, 5fC и 5caC и преодолевают недостатки ранее известных способов, таких как бисульфитное секвенирование.

Сравнение BS и связанных способов с TAPS, TAPS β и CAPS в отношении секвенирования 5mC и 5hmC

Основание	BS	TAB-Seq	oxBS	TAPS	TAPS β	CAPS
C	T	T	T	C	C	C
5mC	C	T	C	T	T	C
5hmC	C	C	T	T	C	T

Способы идентификации 5mC.

В одном аспекте в настоящем изобретении представлен способ идентификации 5-метилцитозина (5mC) в целевой ДНК, включающий стадии:

- a) получения образца ДНК, содержащего целевую ДНК;
- b) модификации ДНК, включающей стадии:
 - i) добавления блокирующей группы к 5-гидроксиметилцитозину (5hmC) в образце ДНК;
 - ii) превращения 5mC в образце ДНК в 5-карбоксилцитозин (5caC) и/или 5-формилцитозин (5fC) и
 - iii) превращения 5caC и/или 5fC в DHU с получением модифицированного образца ДНК, содержащего модифицированную целевую ДНК;
- c) выявления последовательности модифицированной целевой ДНК; где в сравниваемой последовательности модифицированной целевой ДНК имеет место превращение с заменой цитозина (C) на тимин (T).

В вариантах осуществления способа идентификации 5mC в целевой ДНК способ обеспечивает количественную меру частоты модификации 5mC в каждом местоположении, в котором модификация была идентифицирована в целевой ДНК. В вариантах осуществления значения процентного содержания T в каждом местоположении транзиции позволяют определить количественный уровень 5mC в каждом местоположении в целевой ДНК.

С целью идентификации 5mC в целевой ДНК без включения 5hmC 5hmC в образце блокируется так, чтобы он не подвергался превращению в 5caC и/или 5fC. В способах по настоящему изобретению 5hmC в образце ДНК делают нереакционноспособным для последующих стадий путем добавления блокирующей группы к 5hmC. В одном варианте осуществления блокирующая группа представляет собой сахар, в том числе модифицированный сахар, например, глюкозу или 6-азидоглюкозу (6-азидо-6-дезоксид-глюкозу). Блокирующую сахарную группу добавляют к гидроксиметильной группе 5hmC путем введения образца ДНК в контакт с уридиндифосфат(НЮР)-сахаром в присутствии одного или более ферментов глюкозилтрансфераз.

В вариантах осуществления глюкозилтрансфераза представляет собой β -глюкозилтрансферазу (β GT) бактериофага T4, α -глюкозилтрансферазу (α GT) бактериофага T4 и их производные и аналоги. β GT представляет собой фермент, который катализирует химическую реакцию, в которой бета-D-глюкозильный остаток (глюкоза) переносится от UDP-глюкозы на 5-гидроксиметилцитозиновый остаток в нуклеиновой кислоте.

При определении блокирующей группы как, например, глюкозы она относится к глюкозному фрагменту (например, бета-D-глюкозильному остатку), добавляемому к 5hmC с получением глюкозил-5-гидроксиметилцитозина. Блокирующая сахарная группа может представлять собой любой сахар или модифицированный сахар, который является субстратом фермента глюкозилтрансферазы и блокирует последующее превращение 5hmC в 5caC и/или 5fC. Стадию превращения 5mC в образце ДНК в 5caC и/или 5fC затем осуществляют посредством способов, представленных в данном документе, как, например, путем окисления с помощью фермента TET. Превращение 5caC и/или 5fC в DHU также осуществляют посредством способов, представленных в данном документе, как, например, путем окисления бораном.

Способ идентификации 5-метилцитозина (5mC) можно проводить в отношении образца РНК для идентификации местоположения и обеспечения количественной меры 5mC в целевой РНК.

Способы идентификации 5mC или 5hmC (совместно).

В другом аспекте в настоящем изобретении представлен способ идентификации 5mC или 5hmC в целевой ДНК, включающий стадии:

- a) получения образца ДНК, содержащего целевую ДНК;
- b) модификации ДНК, включающей стадии:
 - i) превращения 5mC и 5hmC в образце ДНК в 5-карбоксилцитозин (5caC) и/или 5fC и
 - ii) превращения 5caC и/или 5fC в DHU с получением модифицированного образца ДНК, содержащего модифицированную целевую ДНК;
- c) выявления последовательности модифицированной целевой ДНК; где превращение с заменой цитозина (C) на тимин (T) в последовательности модифицированной целевой ДНК по сравнению с целевой ДНК позволяет определить местоположение 5mC либо 5hmC в целевой ДНК.

В вариантах осуществления способа идентификации 5mC или 5hmC в целевой ДНК способ обеспечивает количественную меру частоты модификаций 5mC или 5hmC в каждом местоположении, в котором модификация была идентифицирована в целевой ДНК. В вариантах осуществления значения процентного содержания T в каждом местоположении транзиции позволяют определить количественный

уровень 5mC или 5hmC в каждом местоположении в целевой ДНК.

Данный способ идентификации 5mC или 5hmC позволяет определить местоположение 5mC и 5hmC, но не позволяет проводить различия между двумя модификациями цитозина. Вместо этого как 5mC, так и 5hmC превращаются в DHU. Присутствие DHU можно выявлять напрямую, или модифицированную ДНК можно подвергать репликации посредством известных способов, в которых DHU превращается в Т.

Способ идентификации 5mC или 5hmC можно проводить в отношении образца РНК для идентификации местоположения и обеспечения количественной меры 5mC или 5hmC в целевой РНК.

Способы идентификации 5tC и идентификации 5hmC.

В настоящем изобретении представлен способ идентификации 5mC и идентификации 5hmC в целевой ДНК путем (i) проведения способа идентификации 5mC в первом образце ДНК, описанном в данном документе, и (ii) проведения способа идентификации 5mC или 5hmC во втором образце ДНК, описанном в данном документе. Определение местоположения 5mC обеспечивается посредством (i). При сравнении результатов (i) и (ii) транзиции с заменой С на Т, выявляемые в (ii), но не в (i), позволяют определить местоположение 5hmC в целевой ДНК. В вариантах осуществления первый и второй образцы ДНК получены из одного и того же образца ДНК. Например, первый и второй образцы могут представлять собой отдельные аликвоты, взятые из образца, содержащего ДНК, подлежащую анализу.

Поскольку 5mC и 5hmC (которые не заблокированы) перед превращением в DHU превращаются в 5fC и 5caC, любые существующие 5fC и 5caC в образце ДНК будут выявляться как 5mC и/или 5hmC. Однако, с учетом крайне низких уровней 5fC и 5caC в геномной ДНК в нормальных условиях, это часто будет приемлемым при анализе метилирования и гидроксиметилирования в образце ДНК. Сигналы 5fC и 5caC можно устранять путем защиты 5fC и 5caC от превращения в DHU, например, посредством конъюгирования с гидроксиламином и связывания с EDC соответственно.

Вышеупомянутый способ позволяет идентифицировать местоположения и значения процентного содержания 5hmC в целевой ДНК путем сравнения местоположений и значений процентного содержания 5mC с местоположениями и значениями процентного содержания 5mC или 5hmC (совместно). В качестве альтернативы, местоположение и частоту модификаций 5hmC в целевой ДНК можно измерить напрямую. Таким образом, в одном аспекте в настоящем изобретении представлен способ идентификации 5hmC в целевой ДНК, включающий стадии:

- a) получения образца ДНК, содержащего целевую ДНК;
- b) модификации ДНК, включающей стадии:
 - i) превращения 5hmC в образце ДНК в 5fC и
 - ii) превращения 5fC в DHU с получением модифицированного образца ДНК, содержащего модифицированную целевую ДНК;
- c) выявления последовательности модифицированной целевой ДНК; где превращение с заменой цитозина (С) на тимин (Т) в последовательности модифицированной целевой ДНК по сравнению с целевой ДНК позволяет определить местоположение 5hmC в целевой ДНК.

В вариантах осуществления стадия превращения 5hmC в 5fC включает окисление 5hmC до 5fC путем приведения ДНК в контакт, например, с перрутеном калия (KRuO₄) (как описано в Science, 2012, 33, 934-937 и WO 2013017853, включенных в данный документ посредством ссылки) или Cu(II)/TEMPO (перхлоратом меди(II) и 2,2,6,6-тетраметилпиперидин-1-оксидом (TEMPO)) (как описано в Chem. Commun., 2017, 53, 5756-5759 и WO 2017039002, включенных в данный документ посредством ссылки). 5fC в образце ДНК затем превращают в DHU посредством способов, раскрытых в данном документе, например, путем реакции с бораном.

Способ идентификации 5mC и идентификации 5hmC можно проводить в отношении образца РНК для идентификации местоположения и обеспечения количественной меры 5mC и 5hmC в целевой РНК.

Способы идентификации 5caC или 5fC.

В одном аспекте в настоящем изобретении представлен способ идентификации 5caC или 5fC в целевой ДНК, включающий стадии:

- a) получения образца ДНК, содержащего целевую ДНК;
- b) превращения 5caC и/или 5fC в DHU с получением модифицированного образца ДНК, содержащего модифицированную целевую ДНК;
- c) необязательно амплификации числа копий модифицированной целевой ДНК;
- d) выявления последовательности модифицированной целевой ДНК; где превращение с заменой цитозина (С) на тимин (Т) в последовательности модифицированной целевой ДНК по сравнению с целевой ДНК позволяет определить местоположение 5caC либо 5fC в целевой ДНК.

Данный способ идентификации 5fC или 5caC позволяет определить местоположение 5fC или 5caC, но не позволяет проводить различия между этими двумя модификациями цитозина. Вместо этого как 5fC, так и 5caC превращаются в DHU, что выявляют посредством способов, описанных в данном документе.

Способы идентификации 5caC.

В другом аспекте в настоящем изобретении представлен способ идентификации 5caC в целевой

ДНК, включающий стадии:

- а) получения образца ДНК, содержащего целевую ДНК;
- б) добавления блокирующей группы к 5fC в образце ДНК;
- в) превращения 5caC в DHU с получением модифицированного образца ДНК, содержащего модифицированную целевую ДНК;
- г) необязательно амплификации числа копий модифицированной целевой ДНК и
- д) определения последовательности модифицированной целевой ДНК; где превращение с заменой цитозина (С) на тимин (Т) в последовательности модифицированной целевой ДНК по сравнению с целевой ДНК позволяет определить местоположение 5caC в целевой ДНК.

В вариантах осуществления способа идентификации 5caC в целевой ДНК способ обеспечивает количественную меру частоты модификации 5caC в каждом местоположении, в котором модификация была идентифицирована в целевой ДНК. В вариантах осуществления значения процентного содержания Т в каждом местоположении транзиции позволяют определить количественный уровень 5caC в каждом местоположении в целевой ДНК.

В этом способе блокируют 5fC (и 5mC и 5hmC не превращаются в DHU), что позволяет идентифицировать 5caC в целевой ДНК. В вариантах осуществления настоящего изобретения добавление блокирующей группы к 5fC в образце ДНК включает приведение ДНК в контакт с альдегид-реактивным соединением, в том числе, например, производными гидроксилamina, производными гидразина и производными гидразидов. Производные гидроксилamina включают гидроксилamin; гидрохлорид гидроксилamina; кислый сульфат гидроксилamina; фосфат гидроксилamina; О-метилгидроксилamin; О-гексилгидроксилamin; О-пентилгидроксилamin; О-бензилгидроксилamin и, в частности, О-этилгидроксилamin (EtONH₂), О-алкилированный или О-ариллированный гидроксилamin, их формы кислоты или соли. Производные гидразина включают N-алкилгидразин, N-арилгидразин, N-бензилгидразин, N,N-диалкилгидразин, N,N-диарилгидразин, N,N-добензилгидразин, N,N-алкилбензилгидразин, N,N-арилбензилгидразин и N-N-алкиларилгидразин. Производные гидразидов включают

-толуолсульфонилгидразид, N-ацилгидразид, N,N-алкилацилгидразид, N,N-бензилацилгидразид, N,N-арилацилгидразид, N-сульфонилгидразид, N,N-алкилсульфонилгидразид, N,N-бензилсульфонилгидразид и N,N-арилсульфонилгидразид.

Способ идентификации 5caC можно проводить в отношении образца РНК для идентификации местоположения и обеспечения количественной меры 5caC в целевой РНК.

Способы идентификации 5fC.

В другом аспекте в настоящем изобретении представлен способ идентификации 5fC в целевой ДНК, включающий стадии:

- а) получения образца ДНК, содержащего целевую ДНК;
- б) добавления блокирующей группы к 5caC в образце ДНК;
- в) превращения 5fC в DHU с получением модифицированного образца ДНК, содержащего модифицированную целевую ДНК;
- г) необязательно амплификации числа копий модифицированной целевой ДНК;
- д) выявления последовательности модифицированной целевой ДНК; где превращение с заменой цитозина (С) на тимин (Т) в последовательности модифицированной целевой ДНК по сравнению с целевой ДНК позволяет определить местоположение 5fC в целевой ДНК.

В вариантах осуществления способа идентификации 5fC в целевой ДНК способ обеспечивает количественную меру частоты модификации 5fC в каждом местоположении, в котором модификация была идентифицирована в целевой ДНК. В вариантах осуществления значения процентного содержания Т в каждом местоположении транзиции позволяют определить количественный уровень 5fC в каждом местоположении в целевой ДНК.

Добавление блокирующей группы к 5caC в образце ДНК можно осуществлять путем (i) приведения образца ДНК в контакт со связывающим средством, например, реагентом для дериватизации карбоновых кислот, таким как производные карбодимидов, например, 1-этил-3-(3-диметиламинопропил)карбодимид (EDC) или N,N'-дициклогексилкарбодимид (DCC), и (ii) приведения образца ДНК в контакт с аминным, гидразиновым или гидроксилaminовым соединением. Так, например, 5caC можно блокировать посредством обработки образца ДНК с помощью EDC, а затем бензиламина, этиламина или другого амина с образованием амида, который блокирует превращение 5caC в DHU под действием, например, пиколин-ВН₃.

Способы связывания с 5caC, катализируемого EDC, описаны в WO2014165770 и включены в данный документ посредством ссылки.

Способ идентификации 5fC можно проводить в отношении образца РНК для идентификации местоположения и обеспечения количественной меры 5fC в целевой РНК.

Образец нуклеиновой кислоты/целевая нуклеиновая кислота.

В настоящем изобретении представлены способы количественной идентификации местоположения одного или более из 5-метилцитозина, 5-гидроксиметилцитозина, 5-карбоксилцитозина и/или 5-

формилцитозина в целевой нуклеиновой кислоте с разрешением до одного основания без воздействия на немодифицированный цитозин. В вариантах осуществления целевая нуклеиновая кислота представляет собой ДНК. В других вариантах осуществления целевая нуклеиновая кислота представляет собой РНК. Аналогично, образец нуклеиновой кислоты, который содержит целевую нуклеиновую кислоту, может представлять собой образец ДНК или образец РНК.

Целевая нуклеиновая кислота может представлять собой любую нуклеиновую кислоту, имеющую модификации цитозина (т.е. 5mC, 5hmC, 5fC и/или 5caC). Целевая нуклеиновая кислота может представлять собой одну молекулу нуклеиновой кислоты в образце или может представлять собой всю популяцию молекул нуклеиновой кислоты в образце (или ее подмножество). Целевая нуклеиновая кислота может представлять собой нативную нуклеиновую кислоту из источника (например, клеток, образцов ткани и т.д.) или может быть предварительно превращена в форму, готовую для высокопроизводительного секвенирования, например, путем фрагментации, репарации и лигирования с адаптерами для секвенирования. Таким образом, целевые нуклеиновые кислоты могут содержать множество последовательностей нуклеиновой кислоты, так что способы, описанные в данном документе, можно применять для создания библиотеки последовательностей целевых нуклеиновых кислот, которые можно анализировать по отдельности (например, путем определения последовательности отдельных целевых молекул) или в группе (например, с помощью способов высокопроизводительного секвенирования или секвенирования нового поколения).

Образец нуклеиновой кислоты может быть получен из организма из царств Дробянки (бактерии), Простейшие, Грибы, Растения и Животные. Образцы нуклеиновой кислоты могут быть получены от пациента или субъекта, из образца окружающей среды или из организма, представляющего интерес. В вариантах осуществления образец нуклеиновой кислоты экстрагируют или получают из клетки или совокупности клеток, биологической жидкости, образца ткани, органа и органеллы.

Образец РНК/целевая РНК.

В настоящем изобретении представлены способы количественной идентификации местоположения одного или более из 5-метилцитозина, 5-гидроксиметилцитозина, 5-карбоксилцитозина и/или 5-формилцитозина в целевой РНК с разрешением до одного основания без воздействия на немодифицированный цитозин. В вариантах осуществления РНК представляет собой одну или более из мРНК (матричной РНК), тРНК (транспортной РНК), рРНК (рибосомальной РНК), snRNA (малой ядерной РНК), miRNA (микроРНК), lncRNA (длинной некодирующей РНК) и eRNA (энхансерной РНК). Целевая РНК может представлять собой одну молекулу РНК в образце или может представлять собой всю популяцию молекул РНК в образце (или ее подмножество). Таким образом, целевая РНК может содержать множество последовательностей РНК, так что способы, описанные в данном документе, можно применять для создания библиотеки последовательностей целевых РНК, которые можно анализировать по отдельности (например, путем определения последовательности отдельных целевых молекул) или в группе (например, с помощью способов высокопроизводительного секвенирования или секвенирования нового поколения).

Образец ДНК/целевая ДНК.

В способах по настоящему изобретению используются мягкие ферментативные и химические реакции, которые позволяют избежать существенного разрушения, ассоциированного с такими способами, как бисульфитное секвенирование. Таким образом, способы по настоящему изобретению применимы при анализе образцов с низким вводимым количеством, как, например, циркулирующей внеклеточной ДНК, и при анализе отдельных клеток.

В вариантах осуществления настоящего изобретения образец ДНК содержит пикограммовые количества ДНК. В вариантах осуществления настоящего изобретения образец ДНК содержит от приблизительно 1 пг до приблизительно 900 пг ДНК, от приблизительно 1 пг до приблизительно 500 пг ДНК, от приблизительно 1 пг до приблизительно 100 пг ДНК, от приблизительно 1 пг до приблизительно 50 пг ДНК, от приблизительно 1 до приблизительно 10 пг ДНК, менее чем приблизительно 200 пг, менее чем приблизительно 100 пг ДНК, менее чем приблизительно 50 пг ДНК, менее чем приблизительно 20 пг ДНК и менее чем приблизительно 5 пг ДНК. В других вариантах осуществления настоящего изобретения образец ДНК содержит нанограммовые количества ДНК. Образец ДНК для применения в способах по настоящему изобретению может представлять собой любое количество, в том числе ДНК из отдельной клетки или совокупные образцы ДНК. В вариантах осуществления способы по настоящему изобретению можно проводить в отношении образца ДНК, содержащего от приблизительно 1 до приблизительно 500 нг ДНК, от приблизительно 1 до приблизительно 200 нг ДНК, от приблизительно 1 до приблизительно 100 нг ДНК, от приблизительно 1 до приблизительно 50 нг ДНК, от приблизительно 1 до приблизительно 10 нг ДНК, от приблизительно 2 до приблизительно 5 нг ДНК, менее чем приблизительно 100 нг ДНК, менее чем приблизительно 50 нг ДНК, менее 5 нг и менее 2 нг ДНК. В вариантах осуществления настоящего изобретения образец ДНК содержит микрограммовые количества ДНК.

Образец ДНК, используемый в способах, описанных в данном документе, может быть получен из любого источника, в том числе, например, из биологической жидкости, образца ткани, органа, органеллы или отдельных клеток. В вариантах осуществления образец ДНК представляет собой циркулирующую внеклеточную ДНК (внеклеточную ДНК или cfDNA), которая представляет собой ДНК, обнаруживае-

мую в крови и отсутствующую в клетке. cfDNA может быть выделена из крови или плазмы крови с помощью способов, известных из уровня техники. Для выделения cfDNA доступны коммерческие наборы, в том числе, например, набор для выделения циркулирующих нуклеиновых кислот (Qiagen). Образец ДНК может быть получен в результате проведения стадии обогащения, в том числе без ограничения иммунопреципитации с помощью антител, иммунопреципитации хроматина, обогащения с использованием расщепления ферментами рестрикции, обогащения с использованием гибридизации или обогащения с использованием химического мечения.

Целевая ДНК может представлять собой любую ДНК, имеющую модификации цитозина (т.е. 5mC, 5hmC, 5fC и/или 5caC), в том числе без ограничения фрагменты ДНК или геномную ДНК, очищенную из тканей, органов, клеток и органелл. Целевая ДНК может представлять собой одну молекулу ДНК в образце или может представлять собой всю популяцию молекул ДНК в образце (или ее подмножество). Целевая ДНК может представлять собой нативную ДНК из источника или быть предварительно превращена в форму, готовую для высокопроизводительного секвенирования, например, путем фрагментации, репарации и лигирования с адаптерами для секвенирования. Таким образом, целевая ДНК может содержать множество последовательностей ДНК, так что способы, описанные в данном документе, можно применять для создания библиотеки последовательностей целевых ДНК, которые можно анализировать по отдельности (например, путем определения последовательности отдельных целевых молекул) или в группе (например, с помощью способов высокопроизводительного секвенирования или секвенирования нового поколения).

Превращение 5mC и 5hmC в 5caC и/или 5fC.

Варианты осуществления настоящего изобретения, такие как способ TAPS, описанный в данном документе, включают стадию превращения 5mC и 5hmC (или только 5mC, если 5hmC заблокирован) в 5caC и/или 5fC. В вариантах осуществления настоящего изобретения эта стадия включает приведение образца ДНК или РНК в контакт с ферментом транслокации десять-одиннадцать (TET). Ферменты TET представляют собой семейство ферментов, которые катализируют перенос молекулы кислорода на метильную группу N5 в 5mC, что приводит к образованию 5-гидроксиметилцитозина (5hmC). TET дополнительно катализирует окисление 5hmC до 5fC и окисление 5fC с образованием 5caC (см. фиг. 5A). Ферменты TET, применимые в способах по настоящему изобретению, включают один или более из TET1, TET2 и TET3 человека; Tet1, Tet2 и Tet3 мыши; TET *Naegleria* (NgTET); *Coprinosopsis cinerea* (CcTET) и их производных или аналогов. В вариантах осуществления фермент TET представляет собой NgTET. В других вариантах осуществления фермент TET представляет собой TET1 человека (hTET1).

Превращение 5caC и/или 5fC в DNU.

Способы по настоящему изобретению включают стадию превращения 5caC и/или 5fC в образце нуклеиновой кислоты в DNU. В вариантах осуществления настоящего изобретения эта стадия включает приведение образца ДНК или РНК в контакт с восстановителем, в том числе, например, борановым восстановителем, таким как пиридинборан, 2-пиколинборан (пиколин-BH₃), боран, борогидрид натрия, цианоборогидрид натрия и триацетоксиборогидрид натрия. В предпочтительном варианте осуществления восстановитель представляет собой пиридинборан и/или пиколин-BH₃.

Аmplification числа копий модифицированной целевой нуклеиновой кислоты.

Способы по настоящему изобретению могут необязательно включать стадию амплификации (увеличения) числа копий модифицированной целевой нуклеиновой кислоты посредством способов, известных из уровня техники. Если модифицированная целевая нуклеиновая кислота представляет собой ДНК, то число копий можно увеличить, например, с помощью ПЦР, клонирования и достройки праймера. Число копий отдельных целевых ДНК можно амплифицировать с помощью ПЦР с использованием праймеров, специфичных по отношению к конкретной последовательности целевой ДНК. В качестве альтернативы, множество различных модифицированных последовательностей целевых ДНК можно амплифицировать путем клонирования в ДНК-вектор с помощью стандартных методик. В вариантах осуществления настоящего изобретения число копий множества различных модифицированных последовательностей целевой ДНК увеличивают с помощью ПЦР для создания библиотеки для секвенирования нового поколения, где, например, двухцепочечная адаптерная ДНК была предварительно лигирована с образцом ДНК (или с модифицированным образцом ДНК), и ПЦР проводят с использованием праймеров, комплементарных адаптерной ДНК.

Выявление последовательности модифицированной целевой нуклеиновой кислоты.

В вариантах осуществления настоящего изобретения способ включает стадию выявления последовательности модифицированной целевой нуклеиновой кислоты. Модифицированная целевая ДНК или РНК содержит DNU в положениях, в которых в немодифицированной целевой ДНК или РНК присутствуют один или более из 5mC, 5hmC, 5fC и 5caC. DNU выступает как T в способах репликации и секвенирования ДНК. Таким образом, модификации цитозина можно выявлять с помощью любого прямого или косвенного способа, известного из уровня техники, который позволяет идентифицировать транзицию с заменой C на T. Такие способы включают способы секвенирования, такие как секвенирование по Сэнгеру, микроматричный анализ и способы секвенирования нового поколения. Транзицию с заменой C на T также можно выявлять посредством анализа с помощью ферментов рестрикции, где при транзиции с за-

меной С на Т удаляется или вводится последовательность распознавания эндонуклеазой рестрикции.

Наборы.

В настоящем изобретении дополнительно представлены наборы для идентификации 5mC и 5hmC в целевой ДНК. Такие наборы содержат реагенты для идентификации 5mC и 5hmC посредством способов, описанных в данном документе. Наборы могут также содержать реагенты для идентификации 5caC и для идентификации 5fC посредством способов, описанных в данном документе. В вариантах осуществления набор включает фермент ТЕТ, борановый восстановитель, а также инструкции по проведению способа. В дополнительных вариантах осуществления фермент ТЕТ представляет собой ТЕТ1, и борановый восстановитель выбран из одного или более из группы, состоящей из пиридинборана, 2-пиколинборана (пиколин-ВНЗ), борана, борогидрида натрия, цианоборогидрида натрия и триацетоксиборогидрида натрия. В дополнительном варианте осуществления фермент ТЕТ1 представляет собой NgTet1 или Tet1 мыши, а борановый восстановитель представляет собой пиридинборан и/или пиколин-ВНЗ.

В вариантах осуществления набор дополнительно содержит блокирующую группу для 5hmC и фермент глюкозилтрансферазу. В других вариантах осуществления блокирующая группа для 5hmC представляет собой уридиндифосфат(UDP)-сахар, где сахар представляет собой глюкозу или производное глюкозы, а фермент глюкозилтрансфераза представляет собой β -глюкозилтрансферазу (β GT) бактериофага T4, α -глюкозилтрансферазу (α GT) бактериофага T4 и их производные и аналоги.

В вариантах осуществления набор дополнительно содержит окислитель, выбранный из перрутената калия (KRuO₄) и/или Cu(II)/ТЕМРО (перхлората меди(II) и 2,2,6,6-тетраметилпиперидин-1-оксила (ТЕМРО)).

В вариантах осуществления набор содержит реагенты для блокирования 5fC в образце нуклеиновой кислоты. В вариантах осуществления набор содержит альдегид-реактивное соединение, в том числе, например, производные гидросиламина, производные гидразина и производные гидразидов, описанные в данном документе. В вариантах осуществления набор содержит реагенты для блокирования 5caC, описанные в данном документе.

В вариантах осуществления набор содержит реагенты для выделения ДНК или РНК. В вариантах осуществления набор содержит реагенты для выделения ДНК с низким вводимым количеством из образца, например, cfDNA из крови, плазмы крови или сыворотки крови.

Примеры

Способы.

Получение модельной ДНК.

ДНК-олигонуклеотиды для тестов методом MALDI и HPLC-MS/MS.

ДНК-олигонуклеотиды ("олиго") с С, 5mC и 5hmC были приобретены у Integrated DNA Technologies (IDT). Все последовательности и модификации можно найти на фиг. 6 и 7. ДНК-нуклеотиды с 5fC синтезировали с помощью способа присоединения С-"хвоста": ДНК-олигонуклеотиды

5'-GTCGACCGGATC-3' и 5'-TTGGATCCGGTTCGACTT-3'

отжигали и затем инкубировали с 5-формил-2'-dCTP (TriLink Biotech) и фрагментом Кленова с 3'→5'-экзонуклеазной активностью (New England Biolabs) в NEBuffer 2 в течение 2 ч. при 37°C. Продукт очищали с помощью колонок Bio-Spin с гелем P-6 (Bio-Rad).

ДНК-олигонуклеотиды с 5caC синтезировали с применением системы для синтеза нуклеиновых кислот Expedite 8900 со стандартными фосфорамидитами (Sigma) и 5-карбоксит-дС-СЕ-фосфорамидитом (Glen Research). Последующее удаление защитных групп и очистку осуществляли с помощью картриджей Glen-Pak (Glen Research) в соответствии с инструкциями производителя. Характеристики очищенных олигонуклеотидов получали с помощью биоспектрометрической рабочей станции Voyager-DE для MALDI-TOF (времяпролетной матрично-активированной лазерной десорбции/ионизации).

Модельная ДНК размером 222 п.о. для теста превращения. Для получения модельной ДНК размером 222 п.о., содержащей пять сайтов CpG, ДНК бактериофага лямбда (Thermo Fisher) подвергали ПЦР-амплификации с помощью ДНК-полимеразы Taq (New England Biolabs) и очищали с помощью гранул AMPure XP (Beckman Coulter). Последовательности праймеров являются следующими:

FW-5'-CCTGATGAAACAAGCATGTC-3', RV-5'-CAUACTCACUTCCCCACUT-3'.

Основание урацил в обратной цепи продукта ПЦР удаляли с помощью фермента USER (New England Biolabs). Затем 100 нг очищенного продукта ПЦР метилировали в 20 мкл раствора, содержащего 1x NEBuffer 2, 0,64 мМ S-аденозилметионина и 20 ед. CpG-метилтрансферазы M.SssI (New England Biolabs), в течение 2 ч. при 37°C с последующей 20-минутной тепловой инактивацией при 65°C. Метилированную модельную ДНК размером 222 п.о. очищали с помощью гранул AMPure XP.

Модельная ДНК для валидации TAPS, TAPSB и CAPS с помощью секвенирования по Сэнгеру. ДНК-олигонуклеотид размером 34 п.о., содержащий один сайт 5mC и один сайт 5hmC, отжигали с другими ДНК-олигонуклеотидами в буфере для отжига, содержащем 5 мМ трис-Cl (pH 7,5), 5 мМ MgCl₂ и 50 мМ NaCl, а затем лигировали в реакционной смеси, содержащей 400 ед. лигазы T4 (NEB), при 25°C в течение 1 ч. и очищали с помощью 1,8X гранул AMPure XP.

ДНК	Последовательность (5'-3')
34 п. о. с mC и hmC	CCCGA ^m CGCATGATCTGTA ^m CTTGTATCGAC ^{hm} CGTGCAAC
Универсальный адаптер TruSeq	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGA CACGACGCTCTTCCGATCT
Адаптер TruSeq (индексная последовательность 6)	/5Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACG CCAATATCTCGTATGCCGTCTTCTGCTTG
Урацилсодержащий линкер	TCTTCCGAUCGTTGCACGGUCGATCAAGUACAGATCAT GCGUCGGGAGAUCGGAAG

Урацилсодержащий линкер удаляли с помощью фермента USER после реакции лигирования, в результате чего получали конечную последовательность продукта (5'-3'):

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGA
TCTCCCGA^mCGCATGATCTGTA^mCTTGTATCGAC^{hm}CGTGCAACGATCGGAAGAGCAC
ACGTCTGAACTCCAGTCACGCCAATATCTCGTATGCCGTCTTCTGCTTG.

ПЦР-праймеры для амплификации модельной ДНК представляли собой: P5: 5'-

AATGATACGGCGACCACCGAG-3' и P7: 5'-CAAGCAGAAGACGGCATAACGAG-3'.

Модельная ДНК для полимеразного теста и секвенирования по Сэнгеру. Модельную ДНК для полимеразного теста и секвенирования по Сэнгеру получали с помощью того же способа лигирования, который приведен выше, за исключением того, что использовали другие ДНК-олигонуклеотиды:

ДНК	Последовательность (5'-3')
34 п. о. с mC	AGCAGTCT ^m CGATCAGCTG ^m CTACTGTA ^m CGTAGCAT
Универсальный адаптер TruSeq	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGA CACGACGCTCTTCCGATCT
Адаптер TruSeq (индексная последовательность 6)	/5Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACG CCAATATCTCGTATGCCGTCTTCTGCTTG
Вставка_1_40_п. о.	/5Phos/AGGTGCGCTAAGTTCTAGATCGCCAACGGTTGT GGCCTT
Вставка_2_60_п. о.	/5Phos/CTATAGCCGGCTTGCTCTCTCTGCCTCTAGCAGCT GCTCCCTATAGTGAGTCGTATTAAC
40_п. о.-линкер-1	ATCTAGAACTTAGCGCACCTAGATCGGAAGAGCGTCGT GT
80_п. о.-линкер:	AGAGAGCAAGCCGGCTATAGATGCTACGTACAGTAGCA GCTGATCAAGACTGCTAAGGCCACAACCAGTTGGCG
42_п. о.-линкер-2:	AGACGTGTGCTCTTCCGATCGTTAATACGACTCACTATA GGG

Конечная последовательность продукта (5'-3') представляла собой:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGA
TCTAGGTGCGCTAAGTTCTAGATCGCCAACGGTTGTGGCCTTAGCAGTCT^mCGAT
CAGCTG^mCTACTGTA^mCGTAGCATCTATAGCCGGCTTGCTCTCTCTGCCTCTAGCA
GCTGCTCCCTATAGTGAGTCGTATTAACGATCGGAAGAGCACACGTCTGAACTCC
AGTCACGCCAATATCTCGTATGCCGTCTTCTGCTTG.

ПЦР-праймеры для амплификации модельной ДНК представляют собой праймеры P5 и P7, представленные выше. Последовательность праймера для достройки праймера, меченная биотином, представляет собой биотин, связанный с 5'-концом праймера P7. ПЦР-праймеры для RT-PCR после транскрипции под действием РНК-полимеразы T7 представляли собой праймер P5 и RT:

5'-TGCTAGAGGCAGAGAGAGCAAG-3'.

Модельная ДНК для теста ПЦР-сдвига. Модельную ДНК для теста ПЦР-сдвига получали с помощью того же способа лигирования, который приведен выше, за исключением того, что использовали другие ДНК-олигонуклеотиды:

ДНК	Последовательность (5'-3')
17 п. о. с X	AGCAGTCTXGATCAGCT (X = DHU, или U, или T, или C)
17 п. о. без модификаций	GCTACTGTACGTAGCAT
Универсальный адаптер TruSeq	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACGCTCTTCCGATCT
Адаптер TruSeq (индексная последовательность б)	/5Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACG CCAATATCTCGTATGCCGTCTTCTGCTTG
Вставка_1_40_п. о.	/5Phos/AGGTGCGCTAAGTTCTAGATCGCCAACCTGGTTGT GGCCTT
Вставка_2_60_п. о.	/5Phos/CTATAGCCGGCTTGCTCTCTGCTCTAGCAGCT GCTCCCTATAGTGAGTCGTATTAAC
40_п. о.-линкер-1	ATCTAGAACTTAGCGCACCTAGATCGGAAGAGCGTCGT GT
80_п. о.-линкер:	AGAGAGCAAGCCGGCTATAGATGCTACGTACAGTAGCA GCTGATCAAGACTGCTAAGGCCACAACCAGTTGGCG
42_п. о.-линкер-2:	AGACGTGTGCTCTTCCGATCGTTAATACGACTCACTATA GGG

Конечная последовательность продукта (5'-3'):

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGA
TCTAGGTGCGCTAAGTTCTAGATCGCCAACCTGGTTGTGGCCTTAGCAGTCTXGAT
CAGCTGCTACTGTACGTAGCATCTATAGCCGGCTTGCTCTCTGCTCTAGCAGC
TGCTCCCTATAGTGAGTCGTATTAACGATCGGAAGAGCACACGTCTGAACTCCAG
TCACGCCAATATCTCGTATGCCGTCTTCTGCTTG,

где X = DHU, или U, или T, или C.

ПЦР-праймеры для амплификации модельной ДНК представляют собой праймеры P5 и P7, представленные выше.

Получение метилированной геномной ДНК бактериофага лямбда.

1 мкг неметилированной ДНК бактериофага лямбда (Promega) метилировали в 50 мкл реакционной смеси, содержащей 0,64 мМ SAM и 0,8 ед./мкл фермента M.SssI в буфере, не содержащем Mg²⁺ (10 мМ трис-Cl, pH 8,0, 50 мМ NaCl и 10 мМ EDTA) в течение 2 ч при 37°C. Затем добавляли 0,5 мкл фермента M.SssI и 1 мкл SAM, и реакционную смесь инкубировали в течение дополнительных 2 ч при 37°C. Метилированную ДНК затем очищали на 1X гранулах Ampure XP. Для обеспечения полного метилирования всю процедуру повторяли в буфере NEB 2. Затем метилирование ДНК валидировали посредством анализа расщепления под действием HpaII. 50 нг метилированной и неметилированной ДНК расщепляли в 10 мкл реакционной смеси с 2 ед. фермента HpaII (NEB) в буфере CutSmart (NEB) в течение 1 ч. при 37°C. Продукты расщепления прогоняли в 1% агарозном геле вместе с нерасщепленной контрольной ДНК лямбда. Неметилированную ДНК лямбда расщепляли после анализа, тогда как метилированная ДНК лямбда оставалась интактной, что подтверждало полное и успешное метилирование CpG. Последовательность ДНК лямбда можно найти в GenBank и EMBL под номером доступа J02459.

Получение немодифицированных контрольных внутренних стандартов размером 2 т.о.

Контрольные внутренние стандарты размером 2 т.о. (2 т.о.-1, 2, 3) подвергали ПЦР-амплификации из плазмиды pNIC28-Bsa4 (Addgene, № по кат. 26103) в реакционной смеси, содержащей 1 нг ДНК-матрицы, 0,5 мкМ праймеров, 1 ед. высокоточной ДНК-полимеразы Phusion (Thermo Fisher). Последовательности ПЦР-праймеров перечислены в табл. 2.

Таблица 2

Последовательности ПЦР-праймеров для внутренних стандартов

Название праймера	Последовательность (5'-3')
2 т. о.-3_ прямой	CACAGATGTCTGCCTGTTCA
2 т. о.-3_ обратный	AGGGTGGTGAATGTGAAACC

Продукт ПЦР очищали на колонке Zymo-Spin. Немодифицированная контрольная последовательность размером 2 т.о. (5'-3'):

CACAGATGTCTGCCTGTTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTT
AATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTCTGTTTGGT
CACTGATGCCTCCGTGTAAGGGGATTTCTGTTTCATGGGGGTAATGATACCGATG
AAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCCGGTTA
CTGGAACGTTGTGAGGGTAAACAACCTGGCGGTATGGATGCGGCGGGACCAGAGA
AAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCAC
AGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCG
CTGACTTCCGCGTTTTCCAGACTTTACGAAACACGAAACCGAAGACCATTTCATGT
TGTTGCTCAGGTGCGAGACGTTTTGCAGCAGCAGTCGCTTCACGTTGCTCGCGTA
TCGGTGATTCAATTCTGCTAACAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCT
CAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAAT
GGCTGCTTCTCGCCGAAACGTTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCG
AGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTC
CAGCGAAAGCGGTCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCT
ACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCC
CGCGCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTGCA
GATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGC
CCGCTTCCAGTCGGGAAACCTGTGCGTCCAGCTGCATTAATGAATCGGCCAACG
CGCGGGGAGAGGCGGTTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTACCAG
TGAGACGGGCAACAGCTGATTGCCCTTACCCGCTGGCCCTGAGAGAGTTGCAGC
AAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTA
ACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCACTACCGAGAT
ATCCGCACCAACGCGCAGCCCGACTCGGTAATGGCGCGCATTCGCCCCAGCGCC
ATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTACGATTT
GCATGGTTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATC
GGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCG
CCGAGACAGAACTTAATGGGCCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGC
GACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAACTG
TTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGACAG
GCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCC
CACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCC
GCTTCGTTCTACCATCGACACCACGCTGGCACCCAGTTGATCGGCGCGAGAT
TTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCA
ACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAA
TGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTTCGAGAAACG
TGGCTGGCCTGGTTACCCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACT
CTGCGACATCGTATAACGTTACTGGTTTTACATTCACCACCCT

Получение 120-мерных контрольных внутренних стандартов 120-мерные контрольные внутренние стандарты получали путем достройки праймера. Последовательности олигонуклеотидов и праймеры перечислены в табл. 3.

40°C, и система растворителей представляла собой воду, содержащую 10 mM ацетата аммония (pH 6,0, растворитель А), и смесь вода-ацетонитрил (60/40, об./об., растворитель В) при скорости потока 0,4 мл/мин. Градиент был следующим: 0-5 мин - 0 растворителя В; 5-8 мин - 0-5,63% растворителя В; 8-9 мин - 5,63% растворителя В; 9-16 мин - 5,63-13,66% растворителя В; 16-17 мин - 13,66-100% растворителя В; 17-21 мин - 100% растворителя В; 21-24,3 мин - 100-0% растворителя В; 24,3-25 мин - 0% растворителя В. Для количественной оценки использовали динамический режим мониторинга множественных реакций (dMRM) в MS. Параметры, зависящие от источника, были следующими: температура газа 230°C, расход газа 14 л/мин, давление газа-распылителя 40 фунтов на кв. дюйм, температура защитного газа 400°C, расход защитного газа 11 л/мин, напряжение на капилляре 1500 В в режиме определения положительных ионов, напряжение на форсунке 0 В, RF-напряжение при высоком давлении 110 В и RF-напряжение при низком давлении 80 В, оба в режиме определения положительных ионов. Напряжение на фрагменторе составляло 380 В для всех соединений, тогда как другие параметры, зависящие от соединения, были такими, как обобщено в табл. 5.

Таблица 5

Параметры HPLC-MS/MS, зависящие от соединения, используемые для количественной оценки нуклеозидов. RT: время удерживания, SE: энергия соударений; CAE: напряжение ускорения ячейки. Все нуклеозиды анализировали в режиме выявления положительных ионов

Соединение	Ион-предшественник (масса/заряд)	Ион-продукт (масса/заряд)	RT (мин.)	Дельта RT (мин.)	SE (V)	CAE (V)
dA+H	252	136	13,78	2	10	4
dT+H	243	127	11,07	2	10	4
dT+Na	265	149	11,07	2	10	4
dG+H	268	152	9,64	2	10	4
dC+H	228	112	3,71	1,5	10	4
dC+Na	250	134	3,71	1,5	10	4
mdC+H	242	126	9,05	1,5	10	4
mdC+Na	264	148	9,05	1,5	10	4
hmdC+H	258	142	4,34	2	12	4
hmdC+Na	280	164	4,34	2	12	4
fdC+H	256	140	10,69	2	8	4
fdC+Na	278	162	10,69	2	8	4
cadC+H	272	156	1,75	3	12	4
cadC+Na	294	178	1,75	3	12	4
DHU+H	231	115	3,45	3	10	4
DHU+Na	253	137	3,45	3	10	4

Экспрессия и очистка NgTET1.

Плазмиды pRSET-A, кодирующая His-меченный белок NgTET1 (GG739552.1), была разработана и приобретена в Invitrogen. Белок экспрессировали в бактериях *E. coli* BL21 (DE3) и очищали, как описано ранее, с некоторыми модификациями (J. E. Pais et al., Biochemical characterization of a Naegleria TET-like oxygenase and its application in single molecule sequencing of 5-methylcytosine. Proc. Natl. Acad. Sci. U.S.A. 112, 4316-4321 (2015), включенная в данный документ посредством ссылки). Вкратце, для экспрессии белка бактерии из суточной мелкокомасштабной культуры выращивали в среде LB при 37°C и 200 об/мин до тех пор, пока OD600 не достигала 0,7-0,8. Затем культуры охлаждали до комнатной температуры, и экспрессию целевого белка индуцировали с помощью 0,2 mM изопропил-β-D-1-тиогаляктопиранозида (IPTG). Клетки выдерживали в течение дополнительных 18 часов при 18°C и 180 об./мин. Затем клетки собирали и ресуспендировали в буфере, содержащем 20 mM HEPES (pH 7,5), 500 mM NaCl, 1 mM DTT, 20 mM имидазола, 1 мкг/мл лейпептина, 1 мкг/мл пепстатина А и 1 mM PMSF. Клетки разрушали с помощью гомогенизатора высокого давления EmulsiFlex-C5, и лизат осветляли путем центрифугирования в течение 1 часа при 30000×g и 4°C. Собранную надосадочную жидкость загружали на смолы Ni-NTA, и белок NgTET1 элюировали буфером, содержащим 20 mM HEPES (pH 7,5), 500 mM имидазола, 2 M NaCl, 1 mM DTT. Собранные фракции затем очищали на HiLoad 16/60 Sdx 75 (20 mM HEPES, pH 7,5, 2 M NaCl, 1 mM DTT). Фракции, содержащие NgTET1, затем собирали, подвергали замене буфера на буфер, содержащий 20 mM HEPES (pH 7,0), 10 mM NaCl, 1 mM DTT, и загружали в колонку NiTrap HP SP. Чистый белок элюировали солевым градиентом, собирали и подвергали замене буфера на конечный буфер, содержащий 20 mM трис-Cl (pH 8,0), 150 mM NaCl и 1 mM DTT. Затем белок концентрировали до 130 мкМ, смешивали с глицерином (30% об./об.), и аликвоты хранили при -80°C.

Экспрессия и очистка mTET1CD.

Каталитический домен mTET1CD (NM_001253857.2, 4371-6392) с N-концевой Flag-меткой клони-

рвали в pcDNA3-Flag между сайтами рестрикции для KpnI и BamHI. Для экспрессии белка 1 мг плазмиды вводили путем трансфекции в 1 л культуры клеток Expi293F (Gibco) при плотности 1×10^6 клеток/мл, и клетки выращивали в течение 48 ч. при 37°C, 170 об./мин. и 5% CO₂. Затем клетки собирали путем центрифугирования, ресуспендировали в буфере для лизиса, содержащем 50 mM трис-Cl, pH = 7,5, 500 mM NaCl, 1X коктейль ингибиторов протеаз cOmplete (Sigma), 1 mM PMSF, 1% Triton X-100, и инкубировали на льду в течение 20 мин. Клеточный лизат затем осветляли путем центрифугирования в течение 30 мин. при 30000 × g и 4°C. Собранную надосадочную жидкость очищали на аффинном геле ANTI-FLAG M2 (Sigma), и чистый белок элюировали буфером, содержащим 20 mM HEPES, pH = 8,0, 150 mM NaCl, 0,1 мг/мл 3X пептида Flag (Sigma), 1X коктейль ингибиторов протеаз cOmplete (Sigma), 1 mM PMSF. Собранные фракции концентрировали и подвергали замене буфера на конечный буфер, содержащий 20 mM HEPES, pH = 8,0, 150 mM NaCl и 1 mM DTT.

Концентрированный белок смешивали с глицерином (30% об./об.), замораживали в жидком азоте, и аликвоты хранили при -80°C. Активность и качество рекомбинантного mTET1CD проверяли с помощью масс-спектрометрического анализа MALDI. На основании этого анализа рекомбинантный mTET1CD является полностью активным и способен катализировать окисление 5mC до 5caC. Любое значительное расщепление тестируемого модельного олигонуклеотида выявляли посредством MALDI, что подтверждало то, что белок не содержит нуклеаз.

Окисление с помощью TET.

Окисление с помощью NgTET1. Для окисления модельных ДНК-олигонуклеотидов размером 222 п.о. с помощью Tet 100 нг ДНК размером 222 п.о. инкубировали в 20 мкл раствора, содержащего 50 mM буфера MOPS (pH 6,9), 100 mM сульфата железа(II)-аммония, 1 mM α-кетоглутарата, 2 mM аскорбиновой кислоты, 1 mM дитиотреитола (DTT), 50 mM NaCl и 5 мкМ NgTET в течение 1 ч. при 37°C. После этого к реакционной смеси добавляли 0,4 ед. протеиназы K (New England Biolabs) и инкубировали в течение 30 мин при 37°C. Продукт очищали на колонке Zymo-Spin (Zymo Research) в соответствии с инструкциями производителя.

Для окисления геномной ДНК с помощью NgTET1 500 нг геномной ДНК инкубировали в 50 мкл раствора, содержащего 50 mM буфера MOPS (pH 6,9), 100 mM сульфата железа(II)-аммония, 1 mM α-кетоглутарата, 2 mM аскорбиновой кислоты, 1 mM дитиотреитола, 50 mM NaCl и 5 мкМ NgTET1 в течение 1 ч при 37°C. После этого к реакционной смеси добавляли 4 ед. протеиназы K (New England Biolabs) и инкубировали в течение 30 мин при 37°C. Продукт очищали на 1,8X гранулах Ampure в соответствии с инструкциями производителя.

Окисление с помощью mTET1. 100 нг геномной ДНК инкубировали в 50 мкл реакционной смеси, содержащей 50 mM буфера HEPES (pH 8,0), 100 мкМ сульфата железа(II)-аммония, 1 mM α-кетоглутарата, 2 mM аскорбиновой кислоты, 1 mM дитиотреитола, 100 mM NaCl, 1,2 mM АТР и 4 мкМ mTET1CD в течение 80 мин. при 37°C. После этого к реакционной смеси добавляли 0,8 ед. протеиназы K (New England Biolabs) и инкубировали в течение 1 часа при 50°C. Продукт очищали на колонке Bio-Spin с гелем P-30 (Bio-Rad) и на 1,8X гранулах Ampure XP в соответствии с инструкциями производителя.

Восстановление с помощью боранов.

Восстановление с помощью пиколин-ВН₃. 25 мкл 5 М α-пиколинборана (пиколин-ВН₃, Sigma-Aldrich) в MeOH и 5 мкл 3 М раствора ацетата натрия (pH 5,2, Thermo Fisher) добавляли к 20 мкл образца ДНК и инкубировали при 60°C в течение 1 ч. Продукт очищали на колонке Zymo-Spin (Zymo Research) в соответствии с инструкциями производителя для 222 п.о. или на колонках Micro Bio-Spin 6 (Bio-Rad) в соответствии с инструкциями производителя для олигонуклеотидов.

В качестве альтернативы, 100 мг 2-пиколинборана (пиколинборан, Sigma-Aldrich) растворяли в 187 мкл DMSO с получением около 3,26 М раствора. Для каждой реакции 25 мкл раствора пиколинборана и 5 мкл 3 М раствора ацетата натрия (pH 5,2, Thermo Fisher) добавляли к 20 мкл образца ДНК и инкубировали в течение 3 часов при 70°C. Продукт очищали на колонке Zymo-Spin для геномной ДНК или на колонках Micro Bio-Spin 6 (Bio-Rad) для ДНК-олигонуклеотидов в соответствии с инструкциями производителя.

Восстановление с помощью пиридинборана. 50-100 нг окисленной ДНК в 35 мкл воды восстанавливали в 50 мкл реакционной смеси, содержащей 600 mM раствор ацетата натрия (pH=4,3) и 1 М пиридинборана, в течение 16 ч при 37°C и при 850 об/мин в Eppendorf ThermoMixer. Продукт очищали на колонке Zymo-Spin.

Реакция пиколинборана с отдельным нуклеозидом. 500 мкл 3,26 М 2-пиколинборана (пиколинборан, Sigma-Aldrich) в MeOH и 500 мкл 3 М раствора ацетата натрия (pH 5,2, Thermo Fisher) добавляли к 10 мг натриевой соли 2'-дезокситидин-5-карбоновой кислоты (Beyte & Associates). Смесь перемешивали в течение 1 часа при 60°C. Продукт очищали с помощью HPLC с получением чистого соединения в виде белой пены. Соотношение масса/заряд [M+Na]⁺, определенное с помощью MS высокого разрешения (Q-TOF): расчетное значение для C₉H₁₄N₂O₅Na: 253,0800; найденное значение: 253,0789.

Блокирование 5hmC.

Блокирование 5hmC проводили в 20 мкл раствора, содержащего 50 mM буфера HEPES (pH 8),

25 mM MgCl₂, 200 мкМ уридиндифосфоглюкозы (UDP-Glc, New England Biolabs) и 10 ед. PGT (Thermo Fisher), а также 10 мкМ ДНК-олигонуклеотидов с 5hmC, в течение 1 ч. при 37°C. Продукт очищали на колонках Micro Bio-Spin 6 (Bio-Rad) в соответствии с инструкциями производителя.

Блокирование 5fC.

Блокирование 5fC проводили в 100 mM буфера MES (pH 5,0), 10 mM O-этилгидроксиламина (Sigma-Aldrich) и 10 мкМ ДНК-олигонуклеотидов с 5fC в течение 2 ч при 37°C. Продукт очищали на колонках Micro Bio-Spin 6 (Bio-Rad) в соответствии с инструкциями производителя.

Блокирование 5caC.

Блокирование 5caC проводили в 75 mM буфера MES (pH 5,0), 20 mM N-гидроксисукцинимид (NHS, Sigma-Aldrich), 20 mM гидрохлорида 1-(3-диметиламинопропил)-3-этилкарбодиимида (EDC, Fluorochem) и 10 мкМ ДНК-олигонуклеотидов с 5caC при 37°C в течение 0,5 ч. Затем буфер заменяли на 100 mM фосфата натрия (pH 7,5), 150 mM NaCl с использованием колонок Micro Bio-Spin 6 (Bio-Rad) в соответствии с инструкциями производителя. К олигонуклеотидам добавляли 10 mM этиламина (Sigma-Aldrich) и инкубировали в течение 1 часа при 37°C. Продукт очищали на колонках Micro Bio-Spin 6 (Bio-Rad) в соответствии с инструкциями производителя.

Окисление 5hmC.

46 мкл ДНК-олигонуклеотидов с 5hmC подвергали денатурации с помощью 2,5 мкл 1 M NaOH в течение 30 мин. при 37°C в инкубаторе-встряхивателе, затем окисляли с помощью 1,5 мкл раствора, содержащего 50 mM NaOH и 15 mM перрутената калия (KRuO₄, Sigma-Aldrich) в течение 1 ч на льду. Продукт очищали на колонках Micro Bio-Spin 6 в соответствии с инструкциями производителя.

Валидация превращения посредством TAPS с помощью анализа с использованием TaqαI.

Превращение 5mC после TAPS тестировали путем ПЦР-амплификации целевой области, которая содержит сайт рестрикции для TaqαI (TCGA), и последующего расщепления с помощью TaqαI. Например, превращение 5mC в собственных библиотеках для TAPS можно тестировать с использованием ампликона размером 194 п.о., содержащего один сайт рестрикции для TaqαI, который амплифицирован из контрольного внутреннего стандарта ДНК лямбда с метилированными CpG. Продукт ПЦР, амплифицированный из ампликона размером 194 п.о., расщепляли ферментом рестрикции TaqαI, и продукт расщепления проверяли в 2% агарозном геле. Продукт ПЦР, амплифицированный из контрольной ДНК, не подвергшейся превращению, расщепляли с помощью TaqαI, и он демонстрировал две полосы в геле. Сайт рестрикции в образце, подвергнутому превращению посредством TAPS, утрачивается вследствие транзиции С-Т, поэтому ампликон размером 194 п.о. будет оставаться интактным. Общий уровень превращения можно оценивать на основании количественной оценки полос в геле, соответствующих расщепленному и нерасщепленному продуктам, и для успешных образцов, подвергнутых TAPS, он должен превышать 95%.

Вкратце, образец ДНК, подвергшийся превращению, подвергали ПЦР-амплификации с помощью ДНК-полимеразы Taq (New England Biolabs) с соответствующими праймерами. Продукт ПЦР инкубировали с 4 единицами фермента рестрикции TaqαI (New England Biolabs) в 1X буфере CutSmart (New England Biolabs) в течение 30 мин. при 65°C и проверяли с помощью электрофореза в 2% агарозном геле.

Количественная полимеразная цепная реакция (qPCR).

Для сравнения кривых амплификации и кривых плавления между модельными ДНК до и после TAPS (фиг. 11) 1 нг образца ДНК добавляли к 19 мкл мастер-микса для ПЦР, содержащего 1 × мастер-микс LightCycler 480 для высокоразрешающего плавления (Roche Diagnostics Corporation), 250 нМ праймеров

FW-CCTGATGAAACAAGCATGTC и RV-CATTACTCACTTCCCCACTT

и 3 mM MgSO₄. Для ПЦР-амплификации начальную стадию денатурации проводили в течение 10 мин при 95°C с последующими 40 циклами денатурации в течение 5 с при 95°C, отжига в течение 5 с при индивидуально настроенной температуре отжига и удлинения в течение 5 с при 72°C. Конечная стадия включала 1 мин при 95°C, 1 мин. при 70°C и построение кривой плавления (шаг приращения 0,02°C, удержание в течение 5 с перед каждым сбором данных) от 65°C до 95°C.

Для других анализов qPCR проводили путем добавления необходимого количества образца ДНК к 19 мкл мастер-микса для ПЦР, содержащего 1 × мастер-микс Fast SYBR Green (Thermo Fisher), 200 нМ прямого и обратного праймеров. Для ПЦР-амплификации начальную стадию денатурации проводили в течение 20 с при 95°C с последующими 40 циклами денатурации в течение 3 с при 95°C, отжига и удлинения в течение 20 с при 60°C.

Валидация уровня метилирования C^mCGG в мРНК mESC с помощью анализа методом HpaII-qPCR.

1 мкг gDNA mESC инкубировали с 50 единицами HpaII (NEB, 50 единиц/мкл) и 1X буфером CutSmart в 50 мкл реакционной смеси в течение 16 часов при 37°C. В контрольной реакции HpaII не добавляли. К реакционной смеси добавляли 1 мкл протеиназы К и инкубировали при 40°C в течение 30 минут с последующей инактивацией протеиназы К в течение 10 минут при 95°C. Величину Ct для образца, расщепленного с помощью HpaII, или контрольного образца измеряли с помощью анализа методом qPCR, как указано выше, с соответствующими наборами праймеров для конкретных положений CCGG

(перечислены в табл. 9).

Секвенирование по Сэнгеру.

Продукт ПЦР очищали с помощью экзонуклеазы I и щелочной фосфатазы креветки (New England Biolabs) или на колонке Zymo-Spin и обрабатывали для секвенирования по Сэнгеру.

Тест повреждения ДНК на фрагментах различной длины.

К геномной ДНК mESC добавляли внутренний стандарт в виде 0,5% ДНК лямбда с метилированными CpG, и ее оставляли нефрагментированной или обрабатывали ультразвуком с помощью прибора Covaris M220 и отбирали по размеру в диапазоне от 500 до 1 т.о. или от 1 т.о. до 3 т.о. на гранулах Ampure XP. 200 нг ДНК однократно окисляли с помощью mTET1CD и восстанавливали с помощью комплекса пиридин-боран, как описано выше, или превращали с помощью набора для бисульфитного превращения EpiTect (Qiagen) в соответствии с протоколом производителя. 10 нг ДНК до и после TAPS и бисульфитного превращения прогоняли в 1% агарозном геле. Для визуализации гель с образцами, подвергшимся бисульфитному превращению, охлаждали в течение 10 мин. на ледяной бане. Превращение 5mC в образцах, подвергшихся TAPS, тестировали с помощью анализа расщепления под действием TaqAI, как описано выше.

Культивирование mESC и выделение геномной ДНК.

ESC мыши (mESC) E14 культивировали на планшетах, покрытых желатином, в среде Игла в модификации Дульбекко (DMEM) (Invitrogen) с добавлением 15% FBS (Gibco), 2 mM L-глутамин (Gibco), 1% незаменимых аминокислот (Gibco), 1% смеси пенициллин/стрептавидин (Gibco), 0,1 mM Р-меркаптоэтанол (Sigma), 1000 единиц/мл LIF (Millipore), 1 мкМ PD0325901 (Stemgent) и 3 мкМ CHIR99021 (Stemgent). Культуры поддерживали при 37°C и 5% CO₂ и пересевали каждые 2 дня.

Для выделения геномной ДНК клетки собирали путем центрифугирования в течение 5 мин при 1000 × g и комнатной температуре. ДНК экстрагировали с помощью набора Quick-DNA Plus (Zymo Research) в соответствии с протоколом производителя.

Получение gDNA mESC для TAPS и WGBS.

Для полногеномного бисульфитного секвенирования (WGBS) к gDNA mESC добавляли внутренний стандарт в виде 0,5% неметилированной ДНК лямбда. Для полногеномного TAPS к gDNA mESC добавляли внутренний стандарт в виде 0,5% метилированной ДНК лямбда и 0,025% немодифицированный контрольный внутренний стандарт размером 2 т.о. Образцы ДНК фрагментировали с помощью прибора Covaris M220 и отбирали по размеру в диапазоне 200-400 п.о. на гранулах Ampure XP. К ДНК для TAPS дополнительно добавляли внутренний стандарт в виде 0,25% контрольных олигонуклеотидов с N5mCNN и N5hmCNN после отбора по размеру с помощью гранул Ampure XP.

Полногеномное бисульфитное секвенирование.

Для полногеномного бисульфитного секвенирования (WGBS) использовали 200 нг фрагментированной gDNA mESC с добавлением внутреннего стандарта в виде 0,5% неметилированной ДНК бактериофага лямбда. Реакцию репарации концов и присоединения А-"хвоста" и лигирование метилированных адаптеров (NextFlex) подготавливали с помощью набора KAPA HyperPlus (KAPA Biosystems) в соответствии с протоколом производителя. Затем ДНК подвергали бисульфитному превращению с помощью набора для бисульфитного превращения EpiTect (Qiagen) в соответствии с протоколом Illumina. Конечную библиотеку амплифицировали с помощью полимеразы KAPA HiFi Uracil Plus (KAPA Biosystems) в течение 6 циклов и очищали на 1X гранулах Ampure. Библиотеку для WGBS-секвенирования подвергали секвенированию спаренных концов с размером фрагментов 80 п.о. на секвенаторе NextSeq 500 (Illumina) с использованием набора для высокопроизводительного секвенирования NextSeq с 15% библиотекой PhiX в качестве контрольного внутреннего стандарта.

Полногеномное TAPS.

Для полногеномного TAPS использовали 100 нг фрагментированной gDNA mESC с добавленным к ней внутренним стандартом в виде 0,5% метилированной ДНК лямбда и 0,025% немодифицированным контрольным внутренним стандартом размером 2 т.о. Реакцию репарации концов и присоединения А-"хвоста" и лигирование адаптеров для мультиплексного секвенирования Illumina подготавливали с помощью набора KAPA HyperPlus в соответствии с протоколом производителя. Лигированную ДНК дважды окисляли с помощью mTET1CD, а затем восстанавливали с помощью пиридинборана в соответствии с протоколами, описанными выше. Конечную библиотеку для секвенирования амплифицировали с помощью полимеразы KAPA HiFi Uracil Plus в течение 5 циклов и очищали на 1X гранулах Ampure. Библиотеку для полногеномного TAPS подвергали секвенированию спаренных концов с размером фрагментов 80 п.о. на секвенаторе NextSeq 500 (Illumina) с использованием одного набора для высокопроизводительного секвенирования NextSeq с 1% библиотекой PhiX в качестве контрольного внутреннего стандарта.

Полногеномное TAPS для образцов с низким вводимым количеством с использованием наборов для получения библиотек dsDNA.

gDNA mESC, полученную, как описано выше для полногеномного TAPS, использовали для полногеномного TAPS для образцов с низким вводимым количеством. Вкратце, образцы, содержащие 100 нг,

10 нг и 1 нг gDNA mESC, однократно окисляли с помощью NgTET1 в соответствии с протоколом, описанным выше. Реакцию репарации концов и присоединения А-"хвоста" и лигирование проводили с помощью набора NEBNext Ultra II (New England Biolabs) или КАРА HyperPlus в соответствии с протоколом производителя. Затем ДНК подвергали реакции с пиколинбораном, как описано выше. Библиотеки, подвергшиеся превращению, амплифицировали с помощью полимеразы КАРА HiFi Uracil Plus и очищали на 1X гранулах Ampure.

Полногеномное TAPS для образцов с низким вводимым количеством с использованием набора для получения библиотек ssDNA.

gDNA mESC, полученную, как описано выше для полногеномного TAPS, использовали для полногеномного TAPS для образцов с низким вводимым количеством. Вкратце, образцы, содержащие 100 нг, 10 нг, 1 нг, 100 пг и 10 пг gDNA mESC, однократно окисляли с помощью NgTET1 и восстанавливали с помощью пиколинборана, как описано выше. Библиотеки для секвенирования получали с использованием набора для получения библиотек ДНК Accel-NGS Methyl-Seq (Swift Biosciences) в соответствии с протоколом производителя. Конечные библиотеки амплифицировали с помощью полимеразы КАРА HiFi Uracil Plus в течение 6 циклов (100 нг), 9 циклов (10 нг), 13 циклов (1 нг), 16 циклов (100 пг) и 21 цикла (10 пг) и очищали на 0,85X гранулах Ampure.

В других экспериментах gDNA mESC, полученную, как описано выше для полногеномного TAPS, использовали для полногеномного TAPS для образцов с низким вводимым количеством. Вкратце, образцы, содержащие 100 нг, 10 нг и 1 нг gDNA mESC, использовали для реакции репарации концов и присоединения А-"хвоста" и лигировали с адаптерами для мультиплексного секвенирования Illumina с помощью набора КАРА HyperPlus в соответствии с протоколом производителя. Лигированные образцы затем однократно окисляли с помощью mTET1CD, а затем восстанавливали с помощью пиридинборана в соответствии с протоколами, описанными выше. Библиотеки, подвергшиеся превращению, амплифицировали с помощью полимеразы КАРА HiFi Uracil Plus в течение 5 циклов (100 нг), 8 циклов (10 нг) и 13 циклов (1 нг) и очищали на 1X гранулах Ampure XP.

TAPS для внеклеточной ДНК.

Образцы внеклеточной ДНК для TAPS получали из 10 нг и 1 нг образцов внеклеточной ДНК. Вкратце, образцы однократно окисляли с помощью NgTET1 и восстанавливали с помощью пиколинборана, как описано выше. Библиотеки для секвенирования получали с использованием набора для получения библиотек ДНК Accel-NGS Methyl-Seq (Swift Biosciences) в соответствии с протоколом производителя. Конечные библиотеки амплифицировали с помощью полимеразы КАРА HiFi Uracil Plus в течение 9 циклов (10 нг) и 13 циклов (1 нг) и очищали на 0,85X гранулах Ampure.

В других экспериментах образцы внеклеточной ДНК для TAPS получали из 10 нг и 1 нг образцов внеклеточной ДНК, как описано выше для полногеномного TAPS. Вкратце, образцы внеклеточной ДНК использовали для реакции репарации концов и присоединения А-"хвоста" и лигировали с адаптерами для мультиплексного секвенирования Illumina с помощью набора КАРА HyperPlus в соответствии с протоколом производителя. Лигированные образцы затем однократно окисляли с помощью mTET1CD, а затем восстанавливали с помощью пиридинборана в соответствии с протоколами, описанными выше. Библиотеки, подвергшиеся превращению, амплифицировали с помощью полимеразы КАРА HiFi Uracil Plus в течение 7 циклов (10 нг) и 13 циклов (1 нг) и очищали на 1X гранулах Ampure XP.

Обработка данных WGBS.

Риды со спаренными концами загружали в формате FASTQ из Illumina BaseSpace и затем обрезали в ходе контроля качества с помощью Trim Galore! v0.4.4

(https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/).

Пары ридов, в которых по меньшей мере один рид был короче 35 п.о. после обрезания, удаляли. Обрезанные риды картировали на геном, в котором были объединены версия mm9 генома мыши, фаг лямбда и PhiX (последовательность из Illumina iGENOMES), с использованием Bismark v0.19 с использованием параметра -nooverlap (F. Krueger, S. R. Andrews, Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27, 1571-1572 (2011), включенная в данный документ посредством ссылки). Фильтр "три С" использовали для удаления ридов с чрезмерно высокими показателями частоты отсутствия превращения. ПЦР-дубликаты распознавали с помощью Picard MarkDuplicates v1.119 (<http://broadinstitute.github.io/picard/>). Области, о которых известно, что они склонны к наличию артефактов картирования, загружали (<https://sites.google.com/site/anshulkundaje/projects/blacklists>) и исключали из дальнейшего анализа (E. P. Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57-74 (2012), включенная в данный документ посредством ссылки).

Предварительная обработка данных TAPS.

Риды со спаренными концами загружали из Illumina BaseSpace и затем обрезали в ходе контроля качества с помощью Trim Galore! v0.4.4. Пары ридов, в которых по меньшей мере один рид был короче 35 п.о. после обрезания, удаляли. Обрезанные риды картировали на геном, в котором были объединены последовательности внутренних стандартов, фаг лямбда и версия mm9 генома мыши, с использованием BWA-MEM v.0.7.15 (H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760 (2009), включенная в данный документ посредством ссылки) с

параметрами по умолчанию. Области, о которых известно, что они склонны к наличию артефактов картирования, загружали (<https://sites.google.com/site/anshulkundaje/projects/blacklists>) и исключали из дальнейшего анализа (E. P. Consortium, *Nature* 489, 57-74 (2012)).

Выявление оснований, подвергшихся превращению в ходе TAPS.

Выровненные риды разделяли на исходную верхнюю (OT) и исходную нижнюю (OB) цепи с помощью специально разработанного скрипта python3 (MF-filter.py). Затем удаляли ПЦР-дубликаты с помощью Picard MarkDuplicates на OT и OB по отдельности. Перекрывающиеся сегменты в парах ридов удаляли с помощью clipOverlap в BamUtil (<https://github.com/statgen/bamUtil>) из картированных ридов в OT и OB после удаления дубликатов по отдельности. Модифицированные основания затем выявляли с помощью samtools mpileup и специально разработанного скрипта python3 (MF-caller_MOD.py).

Анализ качества секвенирования для TAPS и WGBS.

Статистические показатели качества для каждого типа нуклеотидов извлекали из исходных файлов FASTQ, загруженных из Illumina BaseSpace, с помощью скрипта python3 (MF-phredder.py).

Анализ охвата для TAPS и WGBS.

Файлы с данными об охвате генома для каждого основания создавали с помощью Bedtools v2.25 genomecov (A. R. Quinlan, I. M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features*. *Bioinformatics* 26, 841-842 (2010), включенная в данный документ посредством ссылки). Для сравнения относительных распределений охвата между TAPS и WGBS из ридов TAPS отбирали подвыборку до достижения медианного охвата, соответствующего WGBS, с использованием параметра -s в samtools view. В анализах сравнения охвата в WGBS и TAPS в подвыборке использовали clipOverlap для bam-файлов, полученных как в TAPS, так и в WGBS.

Анализ модификаций цитозина, измеренных с помощью TAPS и WGBS.

Долю модифицированных ридов на основании рассчитывали из выходных данных Bismark и выходных данных MF-caller_MOD.py соответственно. Нахождение пересечений проводили с помощью Bedtools intersect, а статистические анализы и фигуры создавали в R и Matlab. Области генома визуализировали с помощью IGV v2.4.6 (J. T. Robinson et al., *Integrative genomics viewer*. *Nat. Biotechnol.* 29, 24-26 (2011), включенная в данный документ посредством ссылки). Для представления на графике охвата и уровней модификаций вокруг CGI все координаты CGI для mm9 загружали из геномного браузера UCSC, группировали в 20 окон и увеличивали до 50 окон размером 80 п.о. с обеих сторон (до тех пор, пока они не достигали половины расстояния до следующего CGI). Средние уровни модификаций (в CpG) и охват (во всех основаниях для обеих цепей) в каждой секции вычисляли с помощью Bedtools map. Значения для каждой секции снова усредняли и затем наносили на график в Matlab.

Моделирование времени обработки данных.

Получение синтетических ридов при секвенировании спаренных концов моделировали с помощью ART42 на основе генома фага лямбда (с параметрами -p -ss NS50 -errfree -minQ 15 -k 0 -nf 0 -l 75 -c 1000000 -m 240 -s 0 -ir 0 -ir2 0 -dr 0 -dr2 0 -sam -rs 10). 50% всех положений CpG затем отмечали как модифицированные, и получали две библиотеки, для TAPS (с превращением модифицированных оснований) либо для WGBS (с превращением немодифицированных оснований), с использованием специально разработанного скрипта python3. Риды затем обрабатывали в соответствии с конвейером обработки данных, используемым для каждого из способов в документе. Время обработки измеряли с помощью команды time Linux. Все стадии анализа проводили в однопоточном режиме на одном CPU Intel Xeon с объемом памяти 250 ГБ.

Результаты и обсуждение.

Было обнаружено, что пиколин-ВН₃ может обеспечивать легкое превращение 5fC и 5caC в DHU посредством ранее неизвестной реакции восстановительного декарбокислирования/дезаминирования (фиг. 4). С помощью MALDI было показано, что реакция является количественной как в случае с отдельным нуклеозидом, так и в случае с олигонуклеотидами (фиг. 2-3 и 6-7).

11-мерный ДНК-олигонуклеотид, содержащий 5caC, использовали в качестве модели для скрининга химических веществ, которые могут реагировать с 5caC, что отслеживали с помощью масс-спектрометрии с матрично-активированной лазерной десорбцией/ионизацией (MALDI). Было обнаружено, что некоторые бораносодержащие соединения эффективно реагируют с олигонуклеотидами с 5caC, что приводит к снижению молекулярной массы на 41 Да (фиг. 1 и 2). Пиридинборан и его производное 2-пиколинборан (пиколинборан) были отобраны для дополнительного исследования, поскольку они являются коммерчески доступными и экологически безопасными восстановителями.

Реакцию с отдельным нуклеозидом 5caC повторили, и подтвердили, что пиридинборан и пиколинборан обеспечивают превращение 5caC в дигидроурацил (DHU) (фиг. 3, 4B). Интересно, что пиридинборан и пиколинборан, как было обнаружено, также обеспечивают превращение 5fC в DHU посредством кажущегося механизма восстановительного декарбокислирования/дезаминирования (фиг. 4C и 6). Подробный механизм обеих реакций еще предстоит определить. Количественный анализ реакции боранов с ДНК-олигонуклеотидами посредством HPLC-MS/MS подтверждает, что пиколинборан обеспечивает превращение 5caC и 5fC в DHU с эффективностью, составляющей около 98%, и не обладает активностью в отношении неметилированного цитозина, 5mC или 5hmC (фиг. 2B).

Будучи производным урацила, DHU может распознаваться как ДНК-, так и РНК-полимеразами как тимин. Следовательно, восстановление с помощью боранов можно применять для индукции транзиций как 5caC-T, так и 5fC-T и можно применять для секвенирования 5fC и 5caC с разрешением до одного основания, которое авторы настоящего изобретения назвали пиридинборановым секвенированием ("PS") (табл. 6). Восстановление 5fC и 5caC до Т с помощью боранов можно блокировать путем конъюгирования с гидроксиламином (С. X. Song et al., Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* 153, 678-691 (2013), включенная в данный документ посредством ссылки) и связывания с EDC (X. Lu et al., Chemical modification-assisted bisulfite sequencing (CAB-Seq) for 5-carboxylcytosine detection in DNA. *J. Am. Chem. Soc.* 135, 9315-9317 (2013), включенная в данный документ посредством ссылки) соответственно (фиг. 6). Такое блокирование позволяет применять PS для специфичного секвенирования 5fC или 5caC (табл. 6).

Таблица 6

Сравнение BS и связанных способов с PS в отношении секвенирования 5fC и 5caC								
Основание	B	fCAB-Seq/redBS-Seq	caCAB-Seq	fC-CE-T	P	PS блокирование м 5fC	S	PS блокирование м 5caC
C	T	T	T	C	C	C		C
5mC	C	C	C	C	C	C		C
5hmC	C	C	C	C	C	C		C
5fC	T	C	T	T	T	C		T
5caC	T	T	C	C	T	T		C

Кроме того, ферменты TET можно применять для окисления 5mC и 5hmC до 5caC, а затем 5caC можно подвергать восстановлению с помощью боранов в способе, который в данном документе называется пиридинборановым секвенированием с использованием TET ("TAPS") (фиг. 5A-B, табл. 1). TAPS может индуцировать транзицию C-T в отношении 5mC и 5hmC и, следовательно, может применяться для выявления 5mC и 5hmC с разрешением до одного основания.

Кроме того, β -глюкозилтрансфераза (β GT) может обеспечивать мечение 5hmC глюкозой и тем самым защищать его от окисления под действием TET (M. Yu et al., Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* 149, 1368-1380 (2012)) и восстановления под действием боранов (фиг. 7), что обеспечивает избирательное секвенирование только 5mC в способе, называемом в данном документе

TAPS β (фиг. 5B, табл. 1). Затем сайты 5hmC могут быть расшифрованы путем вычитания измерений TAPS β из измерений TAPS. В качестве альтернативы, перрутат калия (KRuO₄), реагент, ранее использовавшийся в окислительном бисульфитном секвенировании (oxBS) (M. J. Booth et al., Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. *Science* 336, 934-937 (2012)), можно использовать для замены TET в качестве химического окислителя для специфичного окисления 5hmC до 5fC (фиг. 7). Этот подход, называемый в данном документе "пиридинборановым секвенированием с использованием химических веществ" ("CAPS"), можно применять для специфичного секвенирования 5hmC (фиг. 5B, табл. 1). Следовательно, TAPS и связанные с ним способы могут в принципе предоставлять полный комплекс средств для секвенирования всех четырех эпигенетических модификаций цитозина (фиг. 5B, табл. 1, табл. 6).

Применение TAPS в отдельности также обеспечивает выявление существующих 5fC и 5caC в геноме. Однако, с учетом крайне низких уровней 5fC и 5caC в геномной ДНК в нормальных условиях, это будет приемлемым. Если при определенных условиях желательно полностью исключить сигналы 5fC и 5caC, это также может быть легко достигнуто путем защиты 5fC и 5caC посредством конъюгирования с гидроксиламином и связывания с EDC соответственно, за счет чего обеспечивается предотвращение превращения в DHU.

Эффективность TAPS оценивали при сравнении с бисульфитным секвенированием, которое является существующим стандартом и наиболее широко применяемым способом для картирования 5mC и 5hmC с разрешением до одного основания. Использовали TET-подобную оксигеназу *Naegleria* (NgTET1) и Tet1 мыши (mTet1), поскольку они оба могут обеспечивать эффективное окисление 5mC до 5caC *in vitro*. Для подтверждения транзиции 5mC-T TAPS применяли в отношении модельной ДНК, содержащей полностью метилированные сайты CpG, и было показано, что он может обеспечивать эффективное превращение 5mC в Т, как продемонстрировано с помощью расщепления ферментами рестрикции (фиг. 8A-B) и секвенирования по Сэнгеру (фиг. 9A). TAPS β и CAPS также валидировали с помощью секвенирования по Сэнгеру (фиг. 12).

TAPS также применяли в отношении геномной ДНК (gDNA) из эмбриональных стволовых клеток мыши (mESC). Количественная оценка методом HPLC-MS/MS показала, что, как и ожидалось, 5mC составляет 98,5% модификаций цитозина в gDNA mESC; остальная их часть состоит из 5hmC (1,5%) и следовых количеств 5fC и 5caC при отсутствии DHU (фиг. 9B). После окисления с помощью TET приблизительно 96% модификаций цитозина окислялись до 5caC, а 3% окислялись до 5fC (фиг. 9B). После вос-

становления с помощью боранов более 99% модификаций цитозина превращались в DHU (фиг. 9B). Эти результаты демонстрируют, что как окисление с помощью ТЕТ, так и восстановление с помощью боранов эффективно работают применительно к геномной ДНК.

Как окисление с помощью ТЕТ, так и восстановление с помощью боранов являются мягкими реакциями без заметного разрушения ДНК по сравнению с бисульфитом (фиг. 10A-D) и, таким образом, обеспечивают высокую степень извлечения ДНК. Другое примечательное преимущество по сравнению с бисульфитным секвенированием заключается в том, что TAPS является неразрушающим и может обеспечивать сохранение ДНК длиной до 10 т.о. (фиг. 10C). Более того, ДНК остается двухцепочечной после TAPS (фиг. 10A-C), и превращение не зависит от длины ДНК (фиг. 15A-B).

Кроме того, поскольку DHU близок к природному основанию, он совместим с различными ДНК-полимеразами и ДНК- или РНК-полимеразами для изотермической амплификации (фиг. 13A-B) и не демонстрирует сдвиг по сравнению с Т/С в ходе ПЦР (фиг. 14).

Полногеномное секвенирование проводили на двух образцах gDNA mESC, один из которых подвергали превращению с использованием TAPS, а другой - с использованием стандартного полногеномного бисульфитного секвенирования (WGBS) для сравнения.

Для оценки точности TAPS добавляли внутренние стандарты различной длины, которые были полностью немодифицированными либо метилированными *in vitro* с помощью CpG-метилтрансферазы (M.SssI) или GpC-метилтрансферазы (M.CviPI) (с использованием вышеуказанных способов). В случае с короткими внутренними стандартами (120-мер-1 и 120-мер-2), содержащими 5mC и 5hmC, наблюдалось почти полное превращение для обеих модификаций в обеих цепях как в окружении CpG, так и в окружении, отличном от окружения CpG (фиг. 17A-B).

Для TAPS использовали 100 нг gDNA по сравнению с 200 нг gDNA для WGBS. Для оценки точности TAPS авторы настоящего изобретения добавляли три различных типа контрольных внутренних стандартов. ДНК лямбда, где все CpG были полностью метилированными, использовали для оценки частоты ложноотрицательных результатов (частоты отсутствия превращения 5mC); немодифицированный ампликон размером 2 т.о. использовали для оценки частоты ложноположительных результатов (частоты превращения немодифицированного C); синтетические олигонуклеотидные внутренние стандарты, содержащие как метилированный, так и гидроксиметилированный C, окруженный любым другим основанием (N5mCNN и N5hmCNN соответственно), использовали для сравнения частоты превращения 5mC и 5hmC в различных окружениях последовательности. При использовании комбинации mTet1 и пиридинборана достигалась наиболее высокая частота превращения 5mC (96,5% и 97,3% для лямбда и синтетических внутренних стандартов соответственно) и наиболее низкая частота превращения немодифицированного C (0,23%) (фиг. 18A-B и фиг. 16). Частота ложноотрицательных результатов, составляющая от 2,7% до 3,5%, при частоте ложноположительных результатов, составляющей лишь 0,23%, является сравнимой с бисульфитным секвенированием: недавнее исследование показало, что 9 коммерческих наборов для бисульфитного секвенирования характеризовались средними показателями частоты ложноотрицательных и ложноположительных результатов, составляющими 1,7% и 0,6% соответственно (Holmes, E.E. et al. Performance evaluation of kits for bisulfite-conversion of DNA from tissues, cell lines, FFPE tissues, aspirates, lavages, effusions, plasma, serum, and urine. *PLoS One* 9, e93933 (2014)). Использование синтетических внутренних стандартов позволяет предположить, что TAPS работает хорошо применительно как к 5mC, так и к 5hmC, а также что TAPS лишь незначительно хуже функционирует в окружении, отличном от окружения CpG. Частота превращения для 5hmC на 8,2% ниже, чем для 5mC, а частота превращения в окружении, отличном от окружения CpG, на 11,4% ниже, чем в окружении CpG (фиг. 18A).

Для данных WGBS требуется специальное программное обеспечение на стадиях как выравнивания, так и распознавания модификаций. В отличие от этого, в конвейере обработки данных авторов настоящего изобретения используется стандартный геномный выравниватель (bwa), а после него - специально разработанный инструмент для распознавания модификаций, который авторы настоящего изобретения назвали "asTair". При обработке смоделированных ридов WGBS и TAPS (полученных из одной и той же полуметилированной последовательности-источника) TAPS/asTair работал более чем в 3 раза быстрее, чем WGBS/Bismark (фиг. 18C).

Вследствие превращения практически всего цитозина в тимин библиотеки для WGBS имеют крайне искаженный нуклеотидный состав, что может отрицательно влиять на секвенирование Illumina. Таким образом, риды WGBS продемонстрировали существенно более низкие показатели качества секвенирования в парах оснований цитозин/гуанин по сравнению с TAPS (фиг. 18E). Чтобы компенсировать сдвиг нуклеотидного состава, в библиотеки для WGBS обычно добавляют по меньшей мере 10-20% ДНК PhiX (контрольную библиотеку со сбалансированным составом оснований) (см., например, полногеномное бисульфитное секвенирование Illumina в системах HiSeq 3000/HiSeq 4000). Соответственно, авторы настоящего изобретения дополняли библиотеку для WGBS с помощью 15% PhiX. Это в комбинации со сниженным содержанием информации в ридов, подвергшихся превращению посредством BS, и разрушением ДНК в результате обработки бисульфитом приводило к значительно более низким показателям картирования для WGBS по сравнению с TAPS (фиг. 18D и табл. 7).

Таблица 7

Статистические показатели качества картирования и секвенирования для WGBS и TAPS

Количественный показатель	WGBS	TAPS
Общее количество необработанных ридов	376062375	455548210
Обрезанные риды	367860813	453028186
Картированные риды (mm9+внутренние стандарты+PhiX)	251940139	451077132
Риды после удаления ПЦР-дубликатов	232303596	398127851
Показатель картирования (картированные риды/обрезанные риды)	68,49%	99,57%
Показатель уникального картирования (уникальные риды [MAPQ > 0 для TAPS]/обрезанные риды)	68,49%	88,08%
Показатель уникального картирования после удаления ПЦР-дубликатов (уникальные риды после удаления ПЦР-дубликатов [MAPQ > 0 для TAPS]/обрезанные риды)	63,15%	81,31%

Следовательно, при одной и той же стоимости секвенирования (за один прогон высокопроизводительного секвенирования NextSeq) средняя глубина TAPS превышала среднюю глубину WGBS (21× и 13,1 × соответственно; табл. 8). Кроме того, TAPS приводило к меньшему количеству неохваченных областей и в целом демонстрировало более равномерное распределение охвата даже после уменьшения размера выборки до той же глубины секвенирования, что и в WGBS (межквартильный размах: 9 и 11 соответственно; фиг. 19А и табл. 8).

Таблица 8

Статистические показатели охвата для TAPS, WGBS и TAPS с уменьшением размера выборки для достижения примерно такого же среднего охвата, как в WGBS. В данном случае охват вычисляли для обеих цепей во всех положениях в геноме.

Количественный показатель	WGBS	TAPS с уменьшением размера выборки	TAPS без уменьшения размера выборки
Среднее значение	13,078	12,411	21,001
Дисперсия	1988,242	482,242	1371,912
Медиана	13	13	22
qtl25	7	8	15
qtl75	18	17	28
iqr	11	9	13
Максимум	116084	37329	63526

Например, CpG-островки (CGI), в частности, в целом были лучше охвачены при использовании TAPS даже с учетом различий в глубине секвенирования между WGBS и TAPS (фиг. 21А), тогда как оба они продемонстрировали эквивалентное деметилирование внутри CGI (фиг. 22). Кроме того, WGBS продемонстрировало незначительный сдвиг в сторону пониженных уровней модификаций в сайтах CpG с высоким охватом (фиг. 23А), тогда как результаты авторов настоящего изобретения позволяют предположить, что TAPS демонстрирует очень небольшой сдвиг зависимости уровня модификаций от охвата (фиг. 23В). Эти результаты демонстрируют, что TAPS обеспечивал значительное улучшение качества секвенирования по сравнению с WGBS при эффективном двукратном снижении стоимости секвенирования.

Более высокий и более равномерный охват генома при использовании TAPS приводил к большему количеству сайтов CpG, охватываемых по меньшей мере тремя ридами. В случае с TAPS на этом уровне были охвачены 88,3% из всех 43205316 сайтов CpG в геноме мыши по сравнению с лишь 77,5% в случае с WGBS (фиг. 21В и фиг. 19В). TAPS и WGBS приводили к высококоррелированным измеренным показателям метилирования в различных областях хромосом (фиг. 21D и фиг. 20). В расчете на один нуклеотид в обоих способах 32755271 положений CpG были охвачены по меньшей мере тремя ридами (фиг. 21В). В этих сайтах авторы настоящего изобретения определили "модифицированные CpG" как все положения CpG с уровнем модификаций, составляющим по меньшей мере 10% (L. Wen et al, Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. *Genome Biology* 15, R49 (2014)). При использовании этого порогового значения 95,8% CpG продемонстрировали состояния модификации, совпадающие между TAPS и WGBS. 98,5% всех CpG, которые были охвачены по меньшей мере тремя ридами и были обнаружены как модифицированные в WGBS, были повторно распознаны как модифицированные посредством TAPS, что указывает на хорошее соответствие между WGBS и TAPS (фиг. 21С). При сравнении уровней модификации для каждого CpG, охваченного по меньшей мере тремя ридами как в WGBS, так и в TAPS, наблюдалась хорошая корреляция между TAPS и WGBS (коэффициент корреляции Пирсона $r=0,63$, $p<2e-16$, фиг. 21Е). Следует отметить, что с помощью TAPS было идентифицировано подмножество высокомодифицированных положений CpG, которые были пропущены при использовании WGBS (фиг. 21Е, нижний правый угол). Авторы настоящего изобретения дополнительно валидировали 7 из этих CpG с помощью ортогонального рестрикционного рас-

щепления и анализа методом ПЦР в реальном времени и подтвердили, что все они являются полностью метилированными и/или гидроксиметилированными (табл. 9).

Табл. 9. Сравнение уровней метилирования 5mCGG в gDNA mESC, количественно оцененных с помощью TAPS, WGBS и анализа методом HpaII-qPCR. Охват и уровень метилирования ($mC\%$) в TAPS и WGBS вычисляли для каждой цепи. Величина C_t для образца, расщепленного с помощью HpaII (C_{tHpaII}), или контрольного образца (C_{tCtrl}) в анализе методом HpaII-qPCR представляла собой среднее значение для трех повторностей. $mC\%$ рассчитывается с помощью следующего уравнения:

$$mC\% = 2^{-(C_{tCtrl} - C_{tHpaII})} * 100\%.$$

Таблица 9

Положение 5mCGG	TAPS		WGBS		Анализ методом HpaII-qPCR			
	Охват	$mC\%$	Охват	$mC\%$	C_{tHpaII}	C_{tCtrl}	$mC\%$	Прямой и обратный праймеры (5'-3')
chr6: 135868201	17	100%	11	0%	29,628	29,642	101,0%	GCTGCAGATTGG AGCCAAAG <u>TTGATGGTGATG</u> <u>GTGGAGCC</u>
chr3: 31339449	15	100%	10	0%	22,162	22,111	96,5%	TCAGTGTCATG GACTCATACT <u>ATACCCCTGGGAG</u> <u>CAAAGTTGTTG</u>
chr4: 128271030	12	100%	10	0%	31,304	31,279	98,3%	CCCACTAGACAT GCTCTGCC <u>CAAAATGTTGCT</u> <u>TGCCTTCCG</u>
chr1: 58635199	11	100%	8	0%	22,008	22,026	101,3%	TCCCTGAGCCCT GATCTAGT <u>AATACTGGCTGA</u> <u>CCGGTTCT</u>
chr14: 36331351	11	100%	14	0%	21,228	21,053	88,6%	ACACCACAGCA GAAGAGAGC <u>TAGGATTGTTGC</u> <u>ACAGGCCA</u>
chr19: 42893499	11	100%	18	0%	22,515	22,558	103,0%	GCTGAGCTGTAT CCTTGAGGT <u>ACACGTGGGTAT</u> <u>TCCACAGC</u>
chr3: 113611193	10	100%	5	0%	22,439	22,545	107,6%	GTGGATCTTCAG TGGTGGCA <u>ATGCTCCCTCAT</u> <u>CCTTTGCA</u>
Сайт CCGG с отрицательным результатом								
chr19: 9043049	25	0%	17	0%	27,11	21,409	1,9%	AGCCTCTGAACT TGACTGCC <u>GCCTGGA ACTCC</u> <u>TGACAGTC</u>
Сайт CCGG с положительным результатом								
chr15: 39335961	16	100%	4	100%	22,163	22,248	106,1%	GGTCCTTGATC CACCCAGAC <u>ACATGGTGCTG</u> <u>GTCTAACCG</u>

В совокупности эти результаты указывают на то, что TAPS может напрямую заменить WGBS и фактически обеспечивает получение более полного представления о метиломе, чем WGBS.

Наконец, TAPS тестировали с ДНК с низким вводимым количеством, и было показано, что TAPS работает всего лишь с 1 нг gDNA и в некоторых случаях до 10 пг gDNA, приближаясь к уровню отдельных клеток. TAPS также эффективно работает с циркулирующей внеклеточной ДНК в количестве до 1 нг. Эти результаты демонстрируют потенциал TAPS в отношении ДНК с низким вводимым количеством и клинических путей применения (фиг. 24А-С, фиг. 25А-В).

TAPS тестировали на трех образцах циркулирующей внеклеточной ДНК (cfDNA) от одного здоро-

вого субъекта, одного субъекта с пищеводом Барретта (синдромом Барретта) и одного субъекта с раком поджелудочной железы, от которых получали 1-2 мл плазмы крови. Следовали стандартному протоколу TAPS, и каждый образец секвенировали до ~ 10-кратного охвата. Анализ результатов TAPS cfDNA показал, что TAPS обеспечивал такое же высококачественное секвенирование метилома cfDNA с низким вводимым количеством, как и при использовании совокупной геномной ДНК, в том числе высокую частоту превращения 5mC (фиг. 26A), низкую частоту ложноположительных результатов (превращение немодифицированного цитозина, фиг. 26B), высокий показатель картирования (фиг. 26C) и низкую частоту появления ПЦР-дубликатов (фиг. 26D). Эти результаты демонстрируют мощность TAPS при диагностике заболеваний на основании cfDNA.

С помощью TAPS также можно дифференцировать метилирование от генетических вариантов С-Т или однонуклеотидных полиморфизмов (SNP) и, таким образом, можно выявлять генетические варианты. Метилирования и SNP С-Т приводят к различным картинам в TAPS: метилирование приводит к получению ридов T/G в исходной верхней нити (OT)/исходной нижней цепи (OB) и ридам A/C в цепях, комплементарных OT (CTOT) и OB (COB), тогда как SNP С-Т приводят к получению ридов T/A в OT/OB и (COB/CTOT) (фиг. 27). Это дополнительно повышает полезность TAPS для получения информации как о метилировании, так и о генетических вариантах, а следовательно, и мутациях, в одном эксперименте и прогоне секвенирования. Эта возможность способа TAPS, раскрытого в данном документе, обеспечивает интеграцию геномного анализа с эпигенетическим анализом и существенное снижение стоимости секвенирования за счет устранения необходимости в проведении стандартного полногеномного секвенирования (WGS).

Таким образом, авторы настоящего изобретения разработали серию способов секвенирования, являющихся производными PS, без применения бисульфита с разрешением до одного основания для эпигенетических модификаций цитозина и продемонстрировали полезность TAPS для секвенирования всего метилома. Благодаря использованию мягких ферментативных и химических реакций для прямого выявления 5mC и 5hmC при разрешении до одного основания с высокой чувствительностью и специфичностью без воздействия на немодифицированные цитозиновые основания TAPS превосходит бисульфитное секвенирование, обеспечивая высококачественный и более полный анализ метилома при половинной стоимости секвенирования. Таким образом, TAPS может заменить бисульфитное секвенирование в качестве нового стандарта в анализе метилцитозина и гидроксиметилцитозина в ДНК. Вместо введения объемной модификации цитозиновых оснований в способе секвенирования 5fC без применения бисульфита, о котором недавно сообщалось (B. Xia et al., Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale. *Nat. Methods* 12, 1047-1050 (2015); C. Zhu et al., Single-Cell 5-Formylcytosine Landscapes of Mammalian Early Embryos and ESCs at Single-Base Resolution. *Cell Stem Cell* 20, 720-731 (2017)), в ходе TAPS происходит превращение модифицированного цитозина в DHU - почти природное основание, которое может "считываться" как T традиционными полимеразными и является потенциально совместимым с секвенированием ДНК без ПЦР. TAPS совместим с рядом последующих анализов, в том числе без ограничения пиросеквенированием, ПЦР, чувствительной к метилированию, рестрикционным расщеплением, масс-спектрометрией с MALDI, микроматричным анализом и полногеномным секвенированием. Поскольку TAPS может обеспечивать сохранение длинной ДНК, этот способ может быть чрезвычайно полезным в комбинации с технологиями секвенирования длинных ридов, такими как SMRT-секвенирование и нанопоровое секвенирование, для исследования определенных труднокартируемых областей. Также возможно объединить способы осаждения с TAPS, чтобы еще больше снизить стоимость секвенирования и добавить информацию с разрешением до одного основания на аффинные карты с низким разрешением. В данном документе было продемонстрировано, что TAPS может напрямую заменить WGBS в повседневном использовании с уменьшением при этом стоимости, сложности и времени, необходимых для анализа. Это может привести к более широкому внедрению эпигенетических анализов в научные исследования и клиническую диагностику.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ модификации нуклеиновой кислоты, включающий стадии:
 - а) обеспечения образца нуклеиновой кислоты, содержащего один или более остатков 5-карбоксилцитозина (5caC) или 5-формилцитозина (5fC); и
 - б) осуществления контакта указанного образца с восстанавливающим агентом на основе борана в условиях, которые обеспечивают восстановление 5caC и 5fC в дигидроурацил (DHU).
2. Способ по п.1, где стадия обеспечения (а) включает подготовку указанного образца нуклеиновой кислоты посредством осуществления контакта целевой нуклеиновой кислоты, которая содержит один или более остатков 5-метилцитозина (5mC) или 5-гидроксиметилцитозина (5hmC), с ферментом транслокации десять-одиннадцать (TET), так что образуется один или более остатков 5caC или 5fC.
3. Способ модификации целевой нуклеиновой кислоты, включающий стадии:
 - а) осуществления контакта целевой нуклеиновой кислоты, которая содержит один или более остатков 5mC или 5hmC in vitro с ферментом TET с образованием препарата окисленной нуклеиновой кисло-

ты; и затем

b) осуществления контакта указанного препарата окисленной нуклеиновой кислоты *in vitro* с восстанавливающим агентом на основе борана в условиях, которые обеспечивают восстановление 5саС и 5fС в дигидроурацил (DHU).

4. Способ по любому из пп.1-3, где указанный восстанавливающий агент выбран из группы, состоящей из пиридинборана, 2-пиколинборана (пиколин-ВН₃), борогидрида натрия и триацетоксиборогидрида натрия.

5. Способ по п.4, где указанный восстанавливающий агент представляет собой пиридинборан или 2-пиколинборан.

6. Способ по любому из вышеуказанных пунктов, дополнительно включающий стадию секвенирования.

7. Способ по п.6, где указанная стадия секвенирования позволяет различить модифицированный и немодифицированный цитозин в указанном образце нуклеиновой кислоты.

8. Способ по п.6 или 7, где указанная стадия секвенирования позволяет определить количественный уровень одной или более модификаций цитозина в указанном образце нуклеиновой кислоты.

9. Способ по любому из пп.6-8, где указанная стадия секвенирования включает одно или более из секвенирования методом обрыва цепи, микроматричного анализа, высокопроизводительного секвенирования и анализа с помощью ферментов рестрикции.

10. Способ по любому из пп.2-8, дополнительно включающий стадию блокирования одного или более модифицированных цитозинов.

11. Способ по п.10, где указанная стадия блокирования включает добавление сахара к 5hmС.

12. Способ по п.11, где указанный сахар представляет собой глюкозу или модифицированную глюкозу.

13. Способ по любому из пп.2-12, где фермент ТЕТ выбран из группы, состоящей из ТЕТ1, ТЕТ2 и ТЕТ3 человека; Tet1, Tet2 и Tet3 мыши; ТЕТ *Naegleria* (NgTET); и ТЕТ *Coprinopsis cinerea* (CcTET).

14. Способ по п.1, где стадия обеспечения (а) включает подготовку указанного образца нуклеиновой кислоты посредством осуществления контакта целевой нуклеиновой кислоты, которая содержит один или более остатков 5hmС, с химическим окисляющим агентом, так что образуется один или более остатков 5fС.

15. Способ по п.10, где указанная стадия блокирования включает осуществление контакта образца нуклеиновой кислоты, содержащего один или более остатков 5fС, с альдегид-реактивным соединением, выбранным из производных гидросиламина, производных гидразина и производных гидразида.

16. Способ по п.15, где производное гидросиламина представляет собой О-этилгидросиламин.

17. Способ по п.10, где указанная стадия блокирования включает осуществление контакта образца нуклеиновой кислоты, содержащего один или более остатков 5саС, с:

реагентом для дериватизации карбоновой кислоты; и

соединением амина, гидразина или гидросиламина.

18. Способ по п.17, где указанная стадия блокирования включает осуществление контакта образца нуклеиновой кислоты, содержащего один или более остатков 5саС, с 1-этил-3-(3-диметиламинопропил)карбодиимидом (EDC) и этиламином.

19. Способ по п.14, где указанный окисляющий агент представляет собой перруттенат калия или Cu (II)/TEMPO.

20. Способ по любому из вышеуказанных пунктов, отличающийся тем, что указанный способ дополнительно включает стадию амплификации количества копий одной или более последовательностей нуклеиновых кислот.

21. Набор для модификации нуклеиновой кислоты, содержащий в отдельных контейнерах:

а) фермент семейства ТЕТ; и

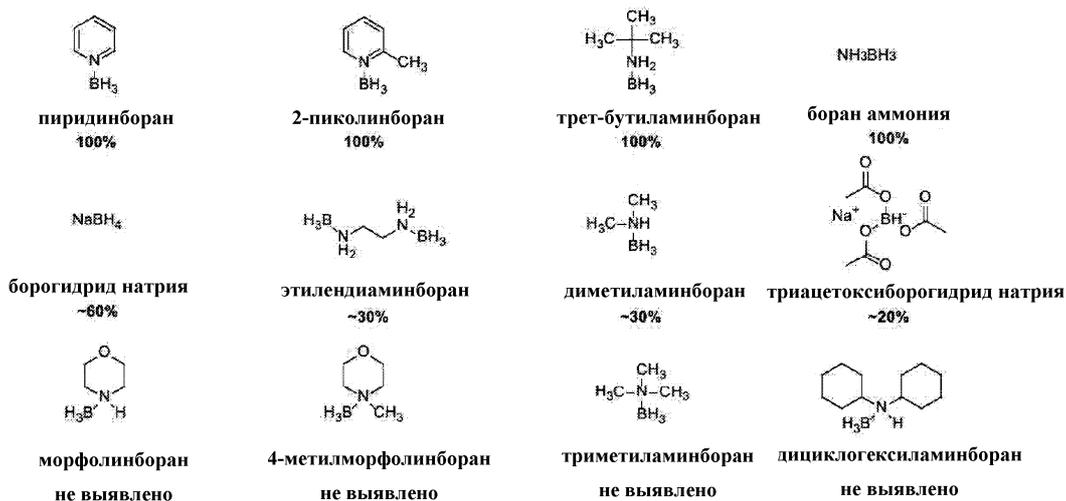
б) восстанавливающий агент на основе борана, который способен восстанавливать 5саС и 5fС до дигидроурацила (DHU).

22. Набор по п.21, дополнительно содержащий по меньшей мере одно из:

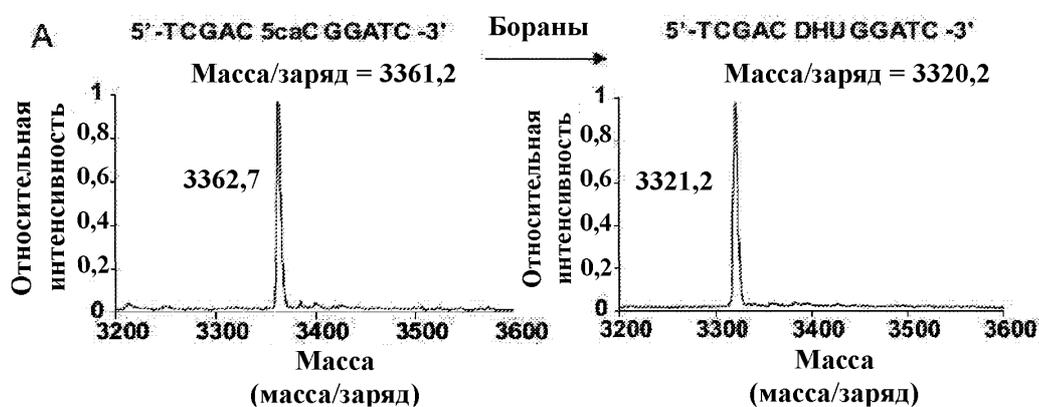
i) фермент глюкозилтрансфераза и подходящий реагент для блокирования 5hmС;

ii) один или более реагентов для блокирования 5fС; и/или

iii) один или более реагентов для блокирования 5саС.

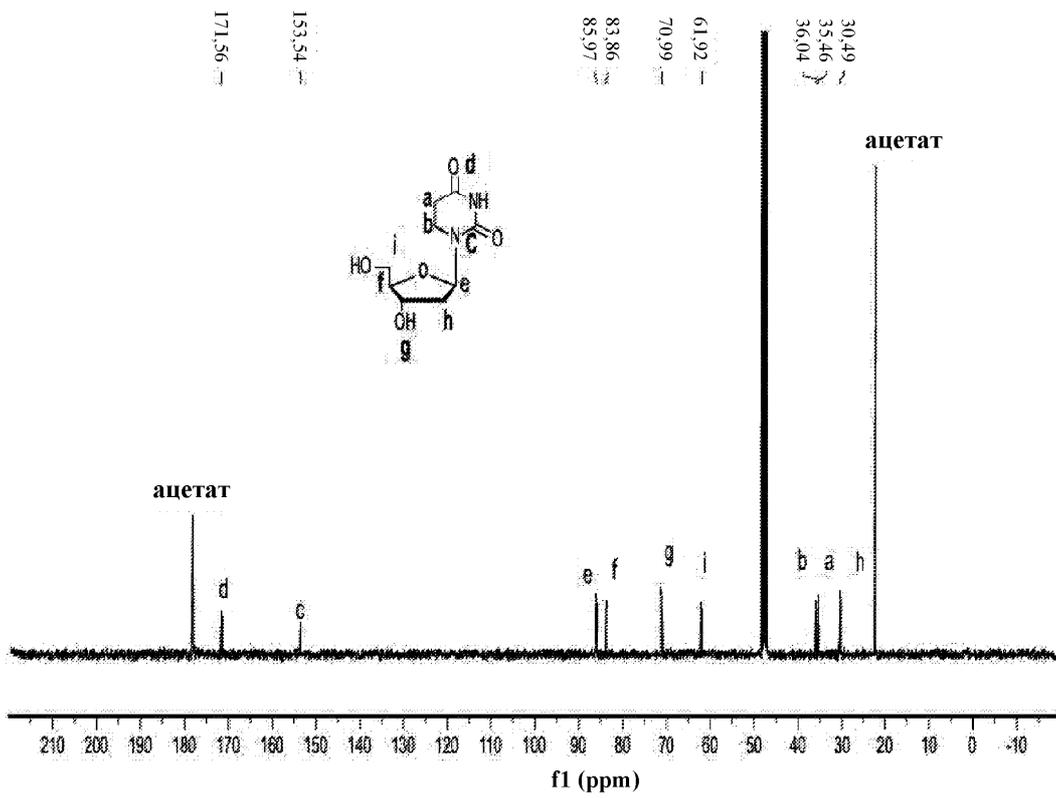
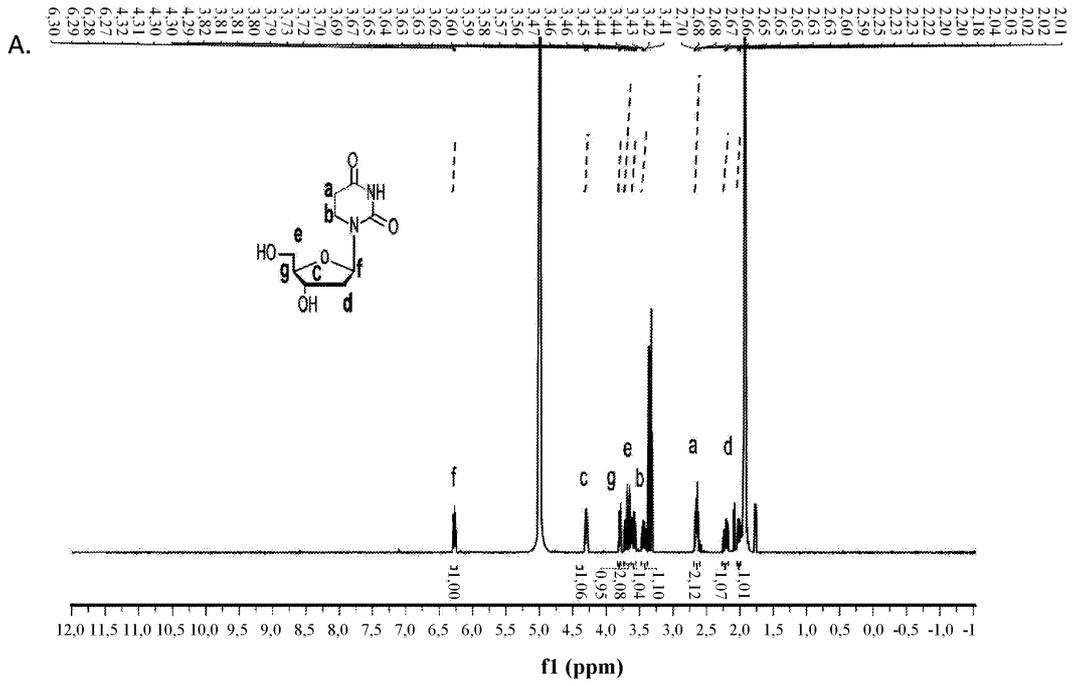


Фиг. 1

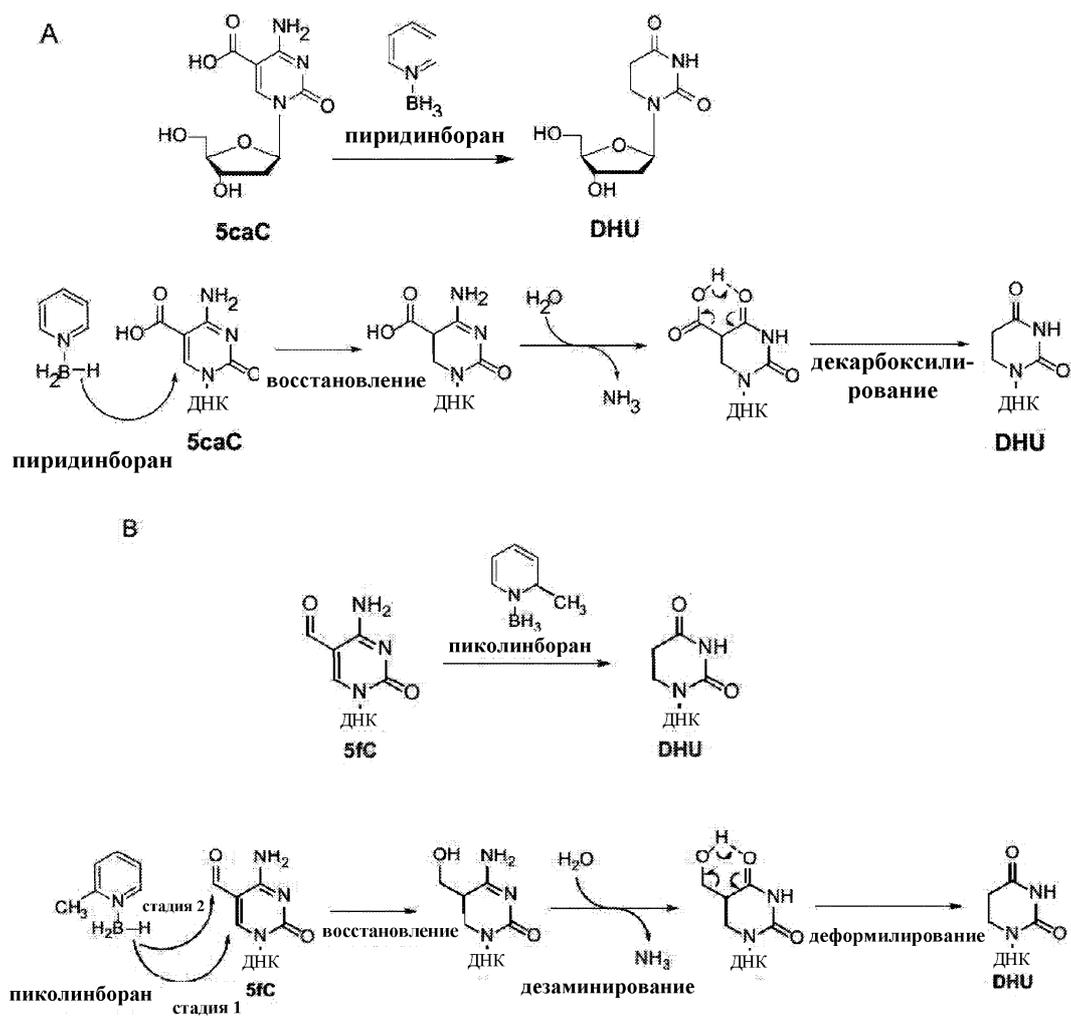
**B**

Модификация цитозина в олигонуклеотидах	Превращение в DHU
C	0%
5mC	0%
5hmC	0%
5fC	97,6 ± 0,1%
5caC	98,1 ± 0,6%

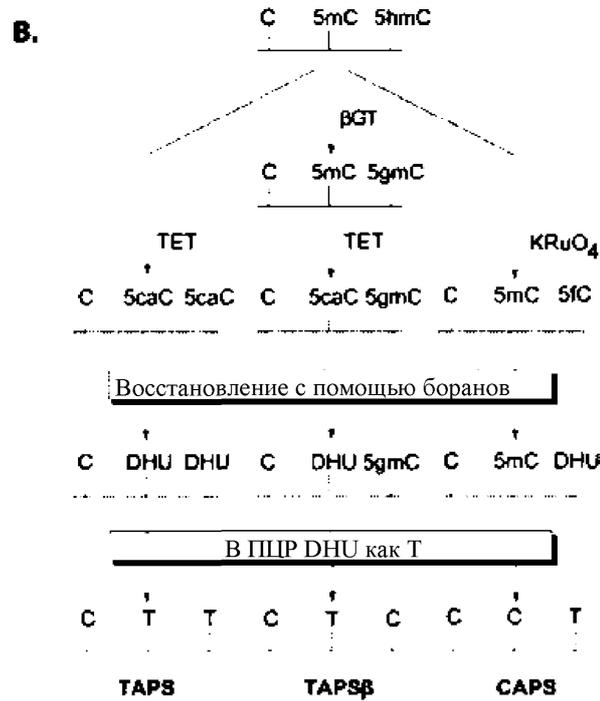
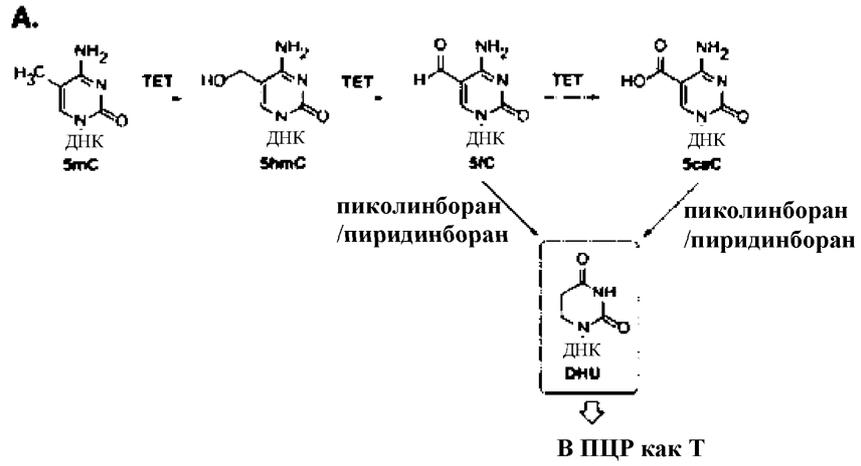
Фиг. 2



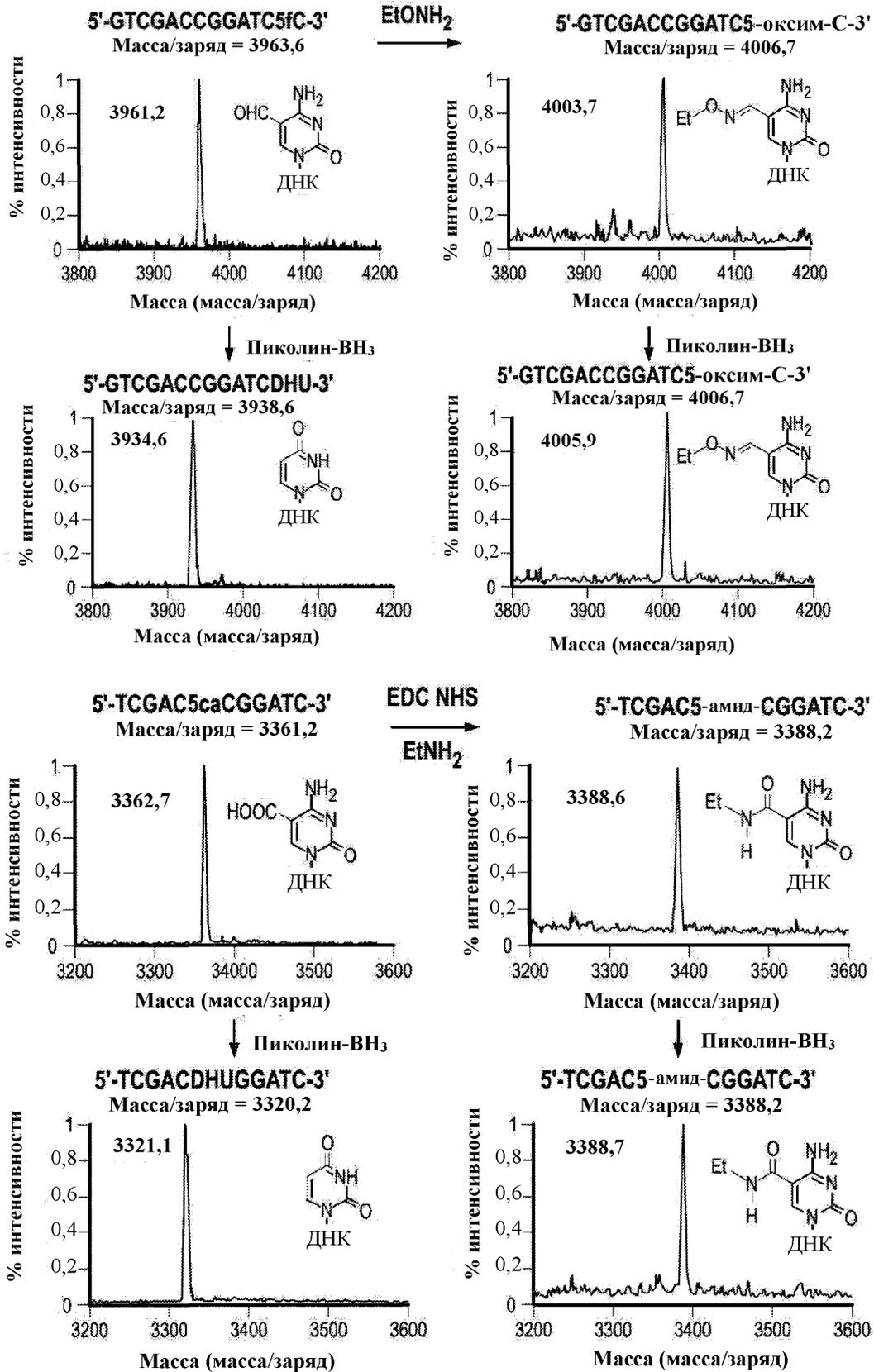
Фиг. 3



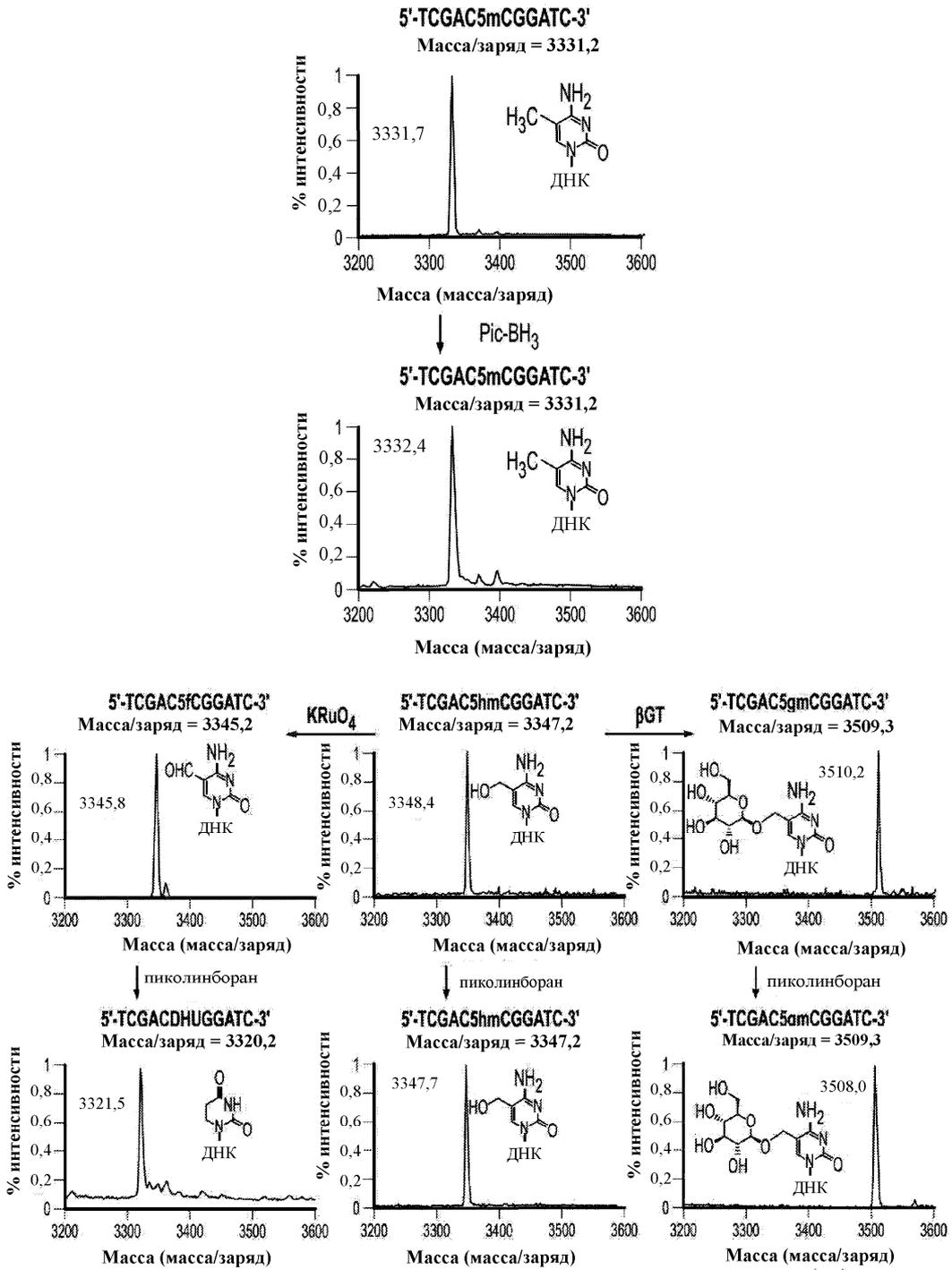
Фиг. 4

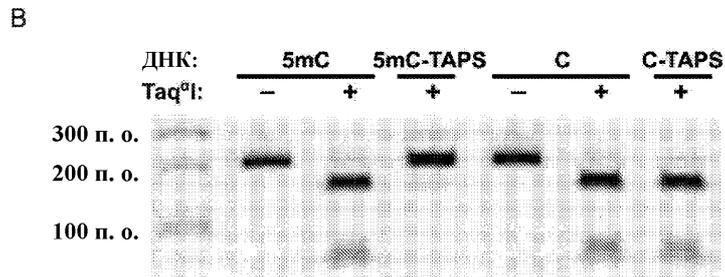
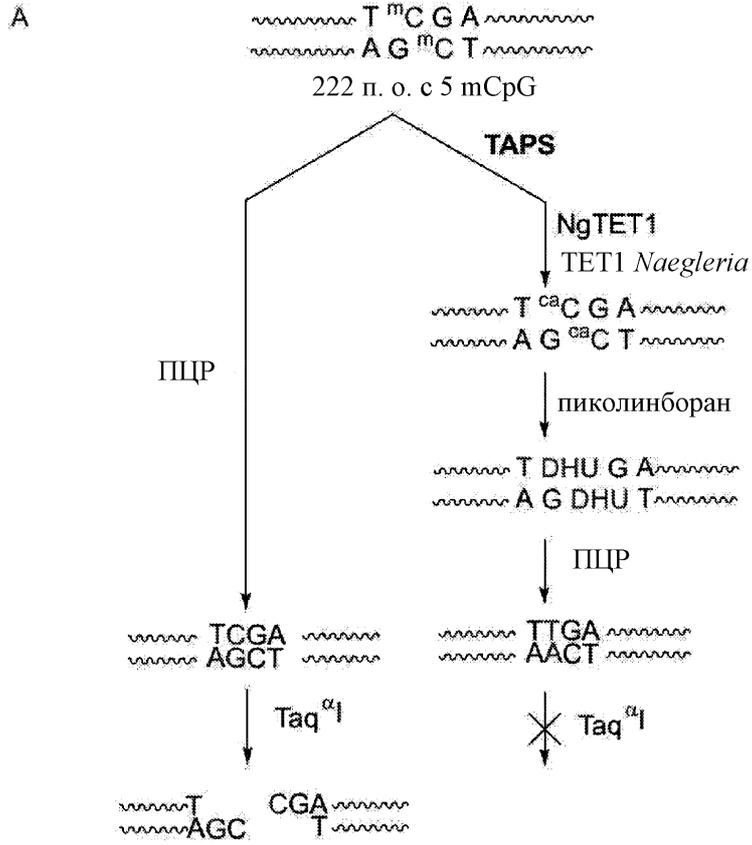


Фиг. 5



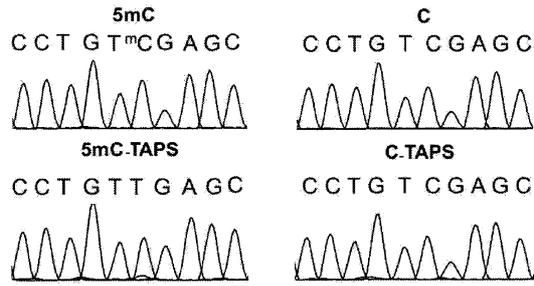
Фиг. 6



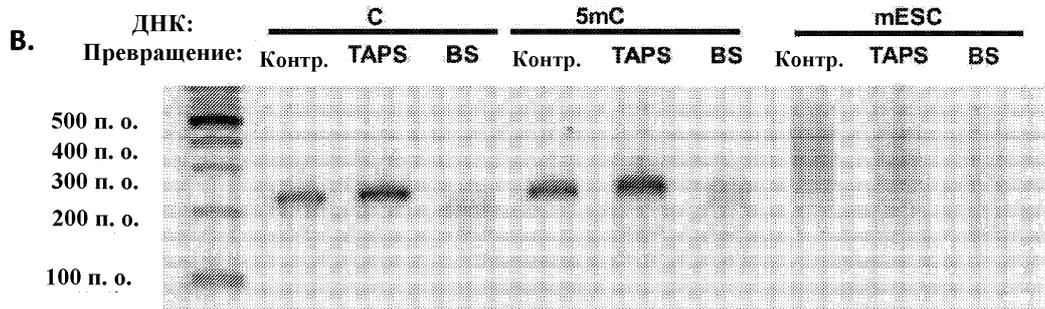
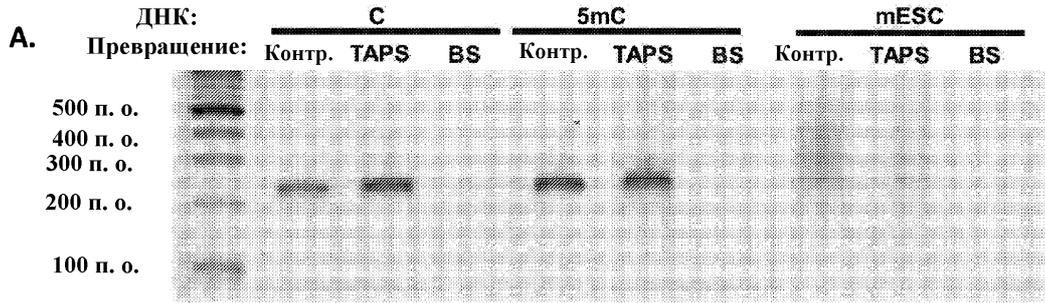
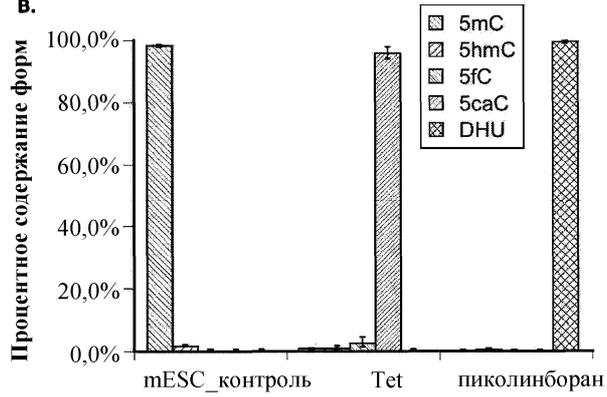


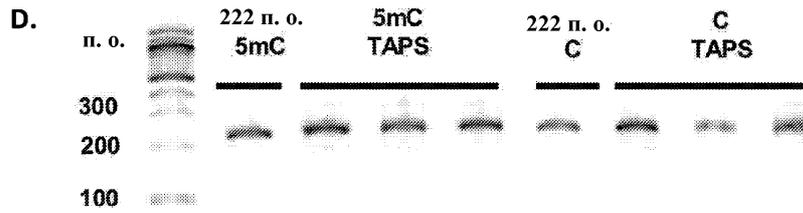
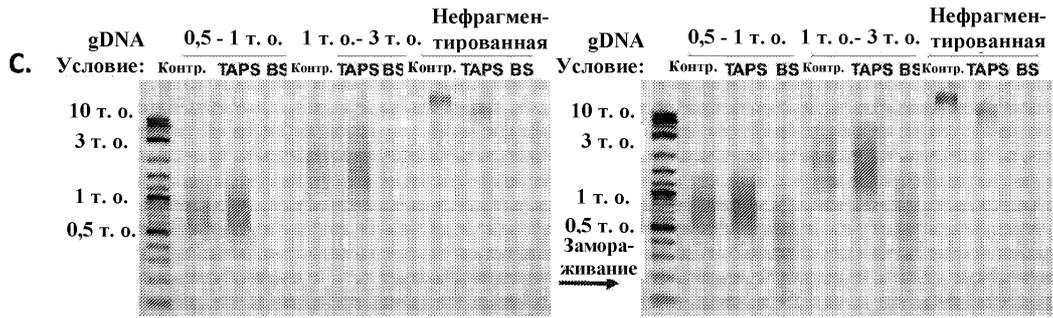
Фиг. 8

A.

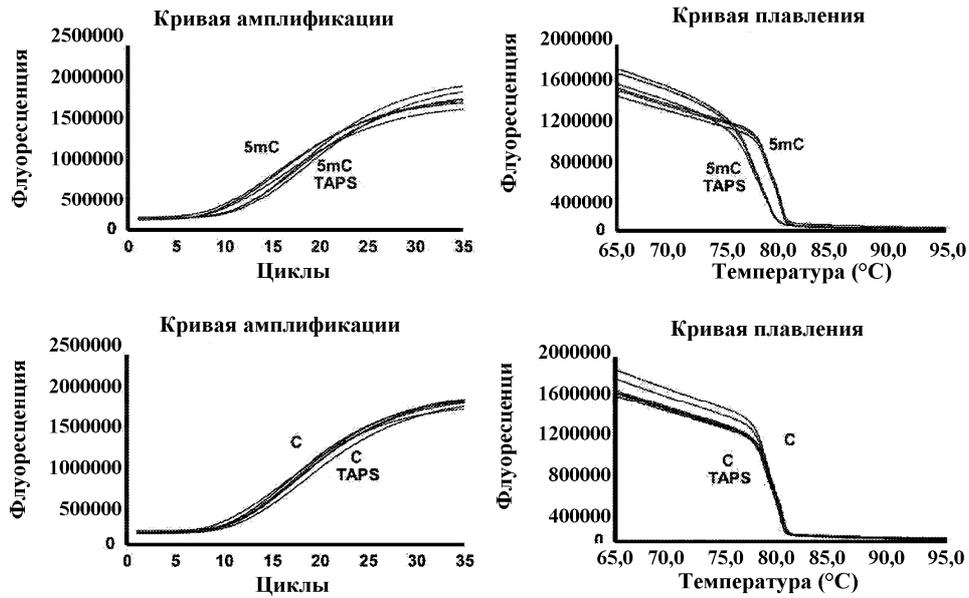


B.

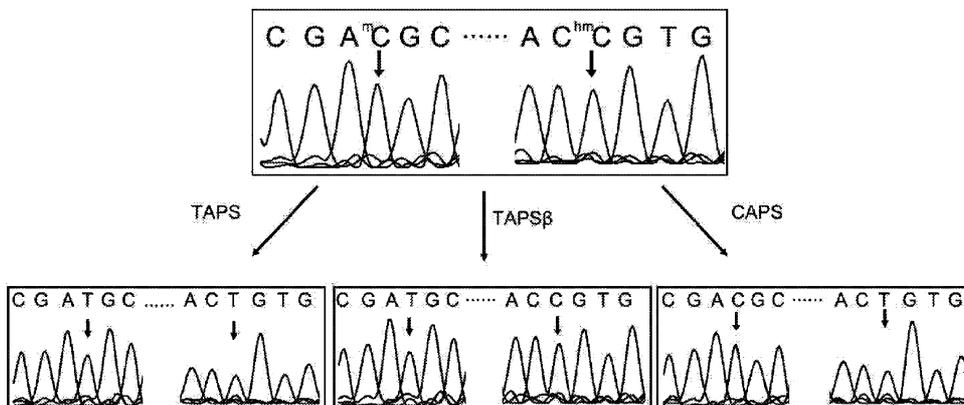




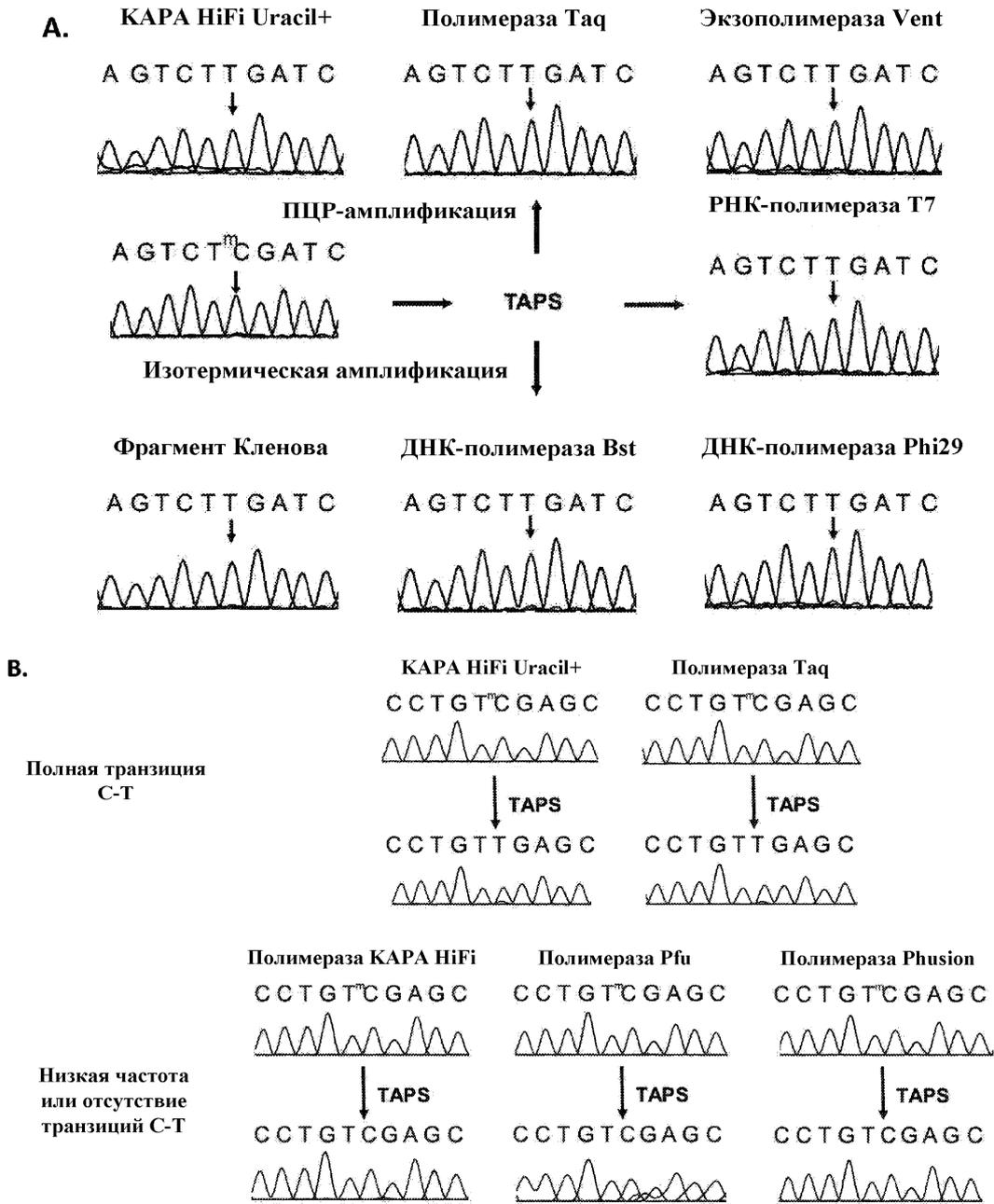
Фиг. 10



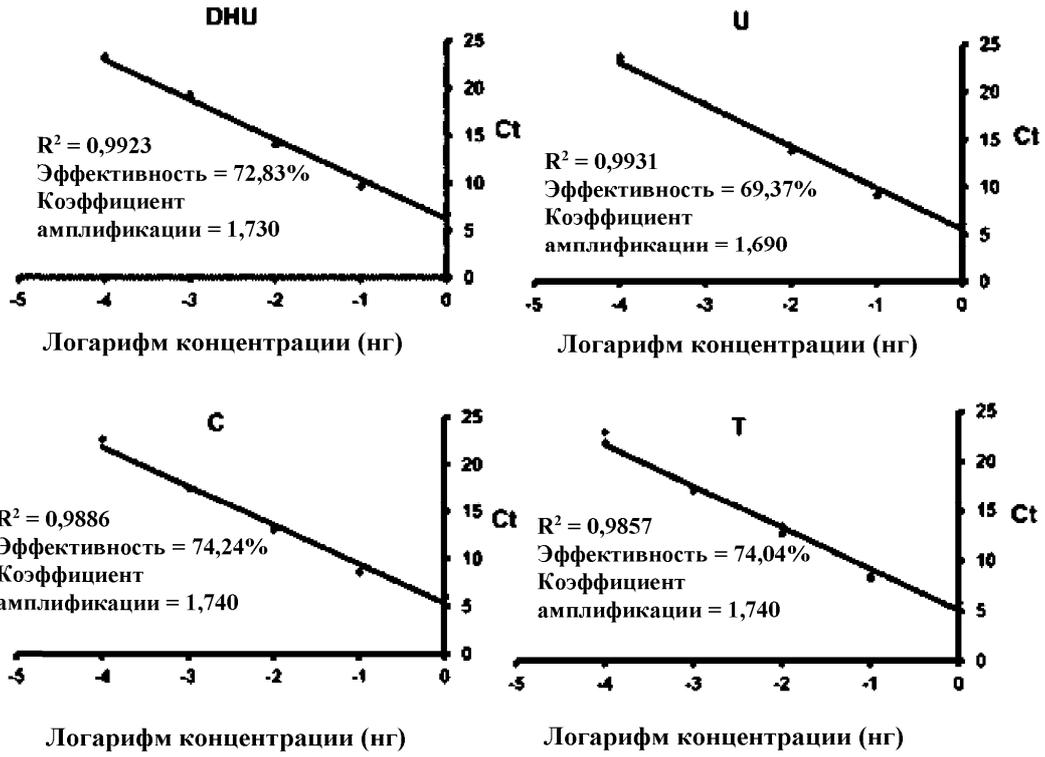
Фиг. 11



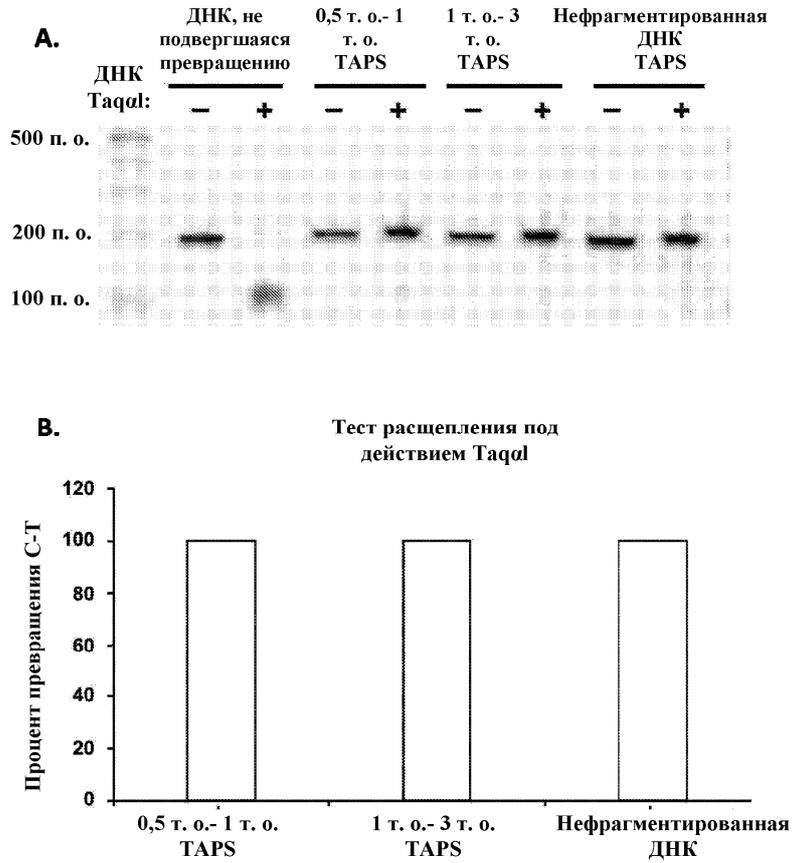
Фиг. 12



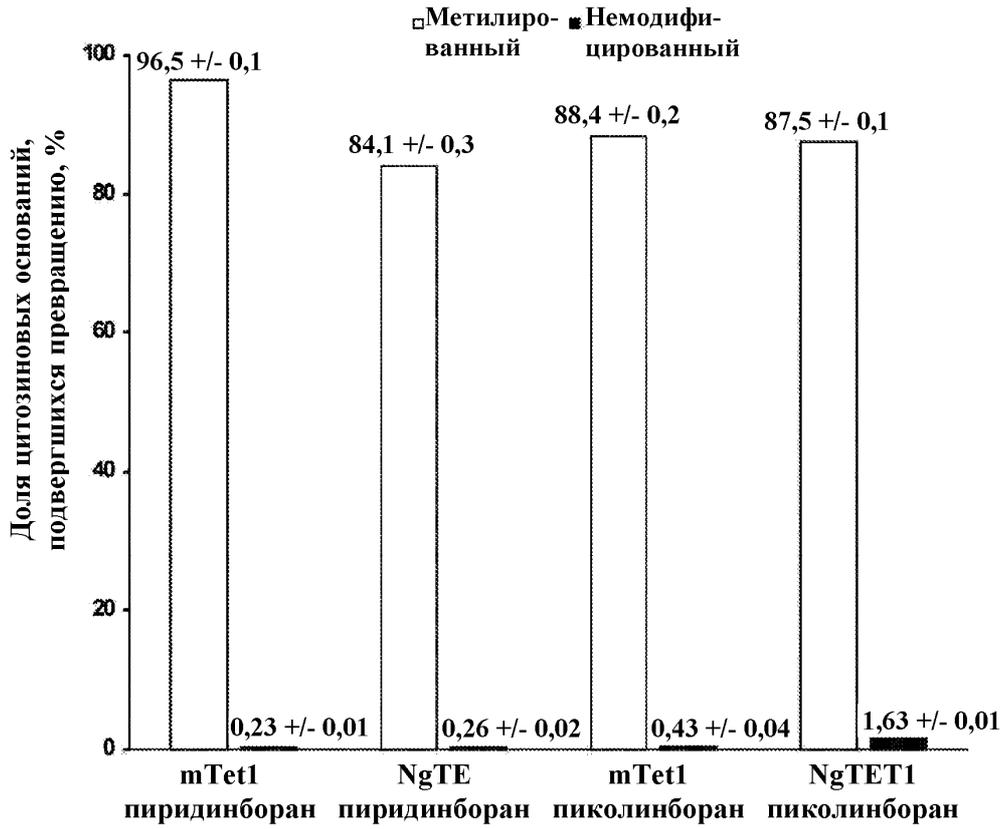
Фиг. 13



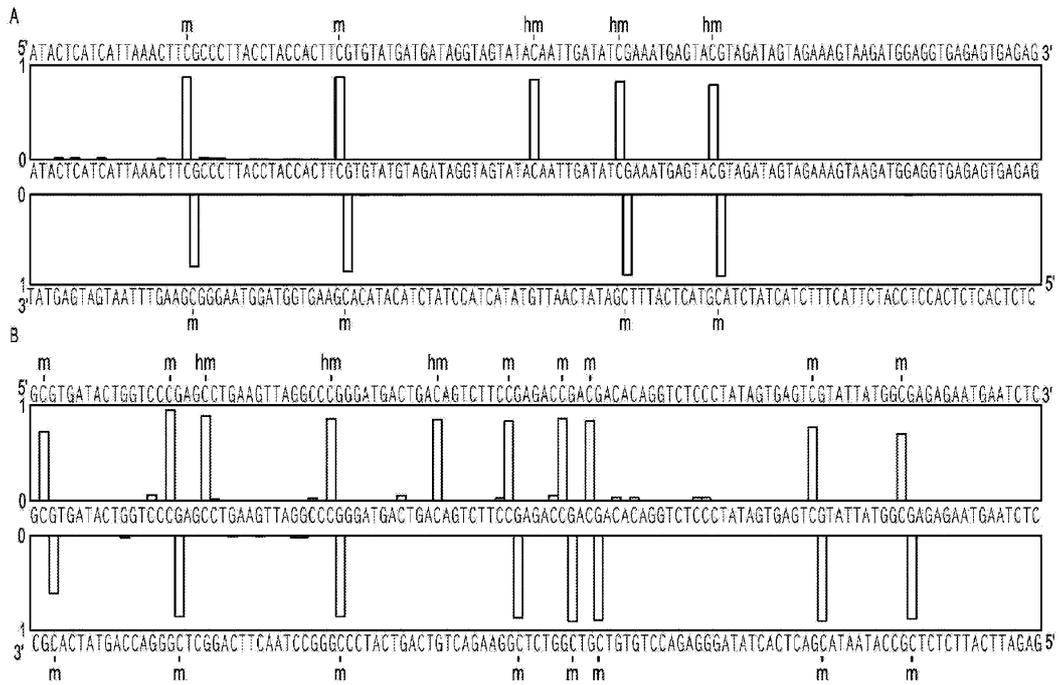
Фиг. 14



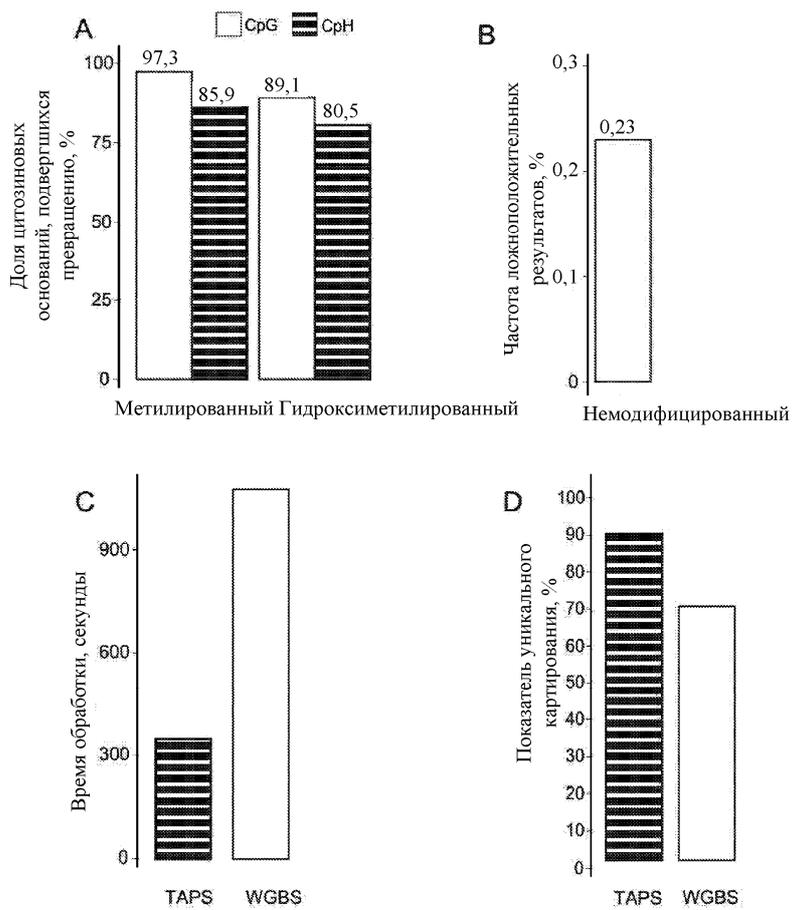
Фиг. 15

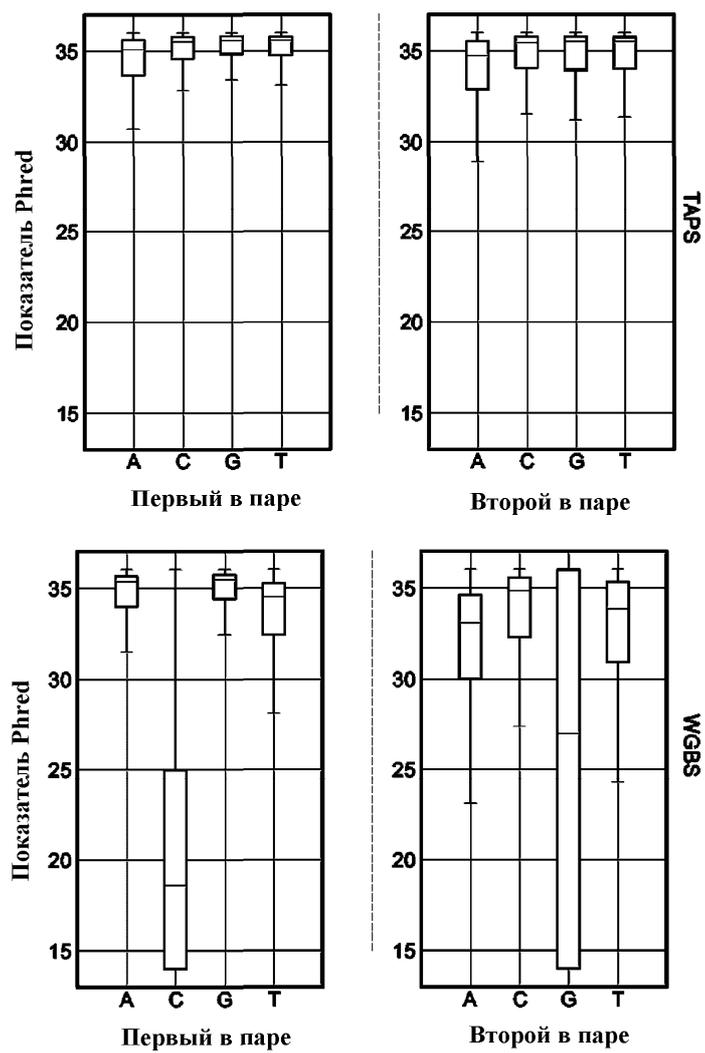


Фиг. 16

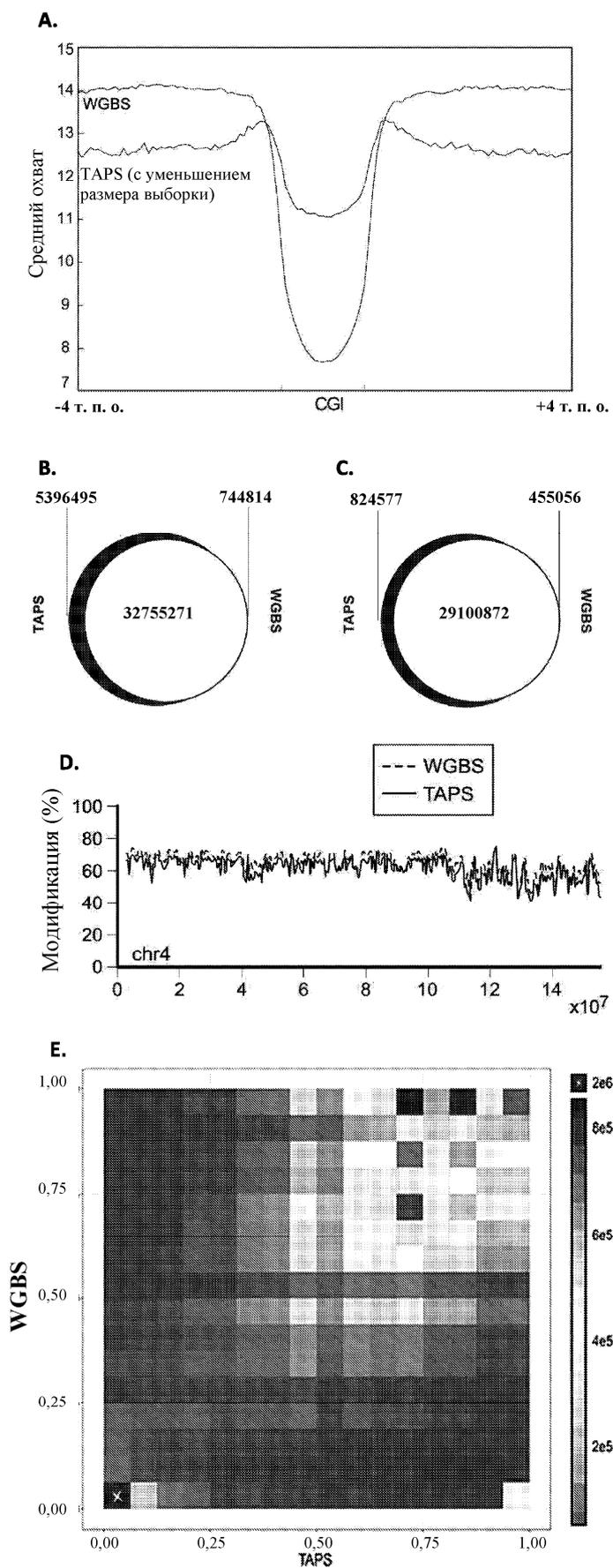


Фиг. 17

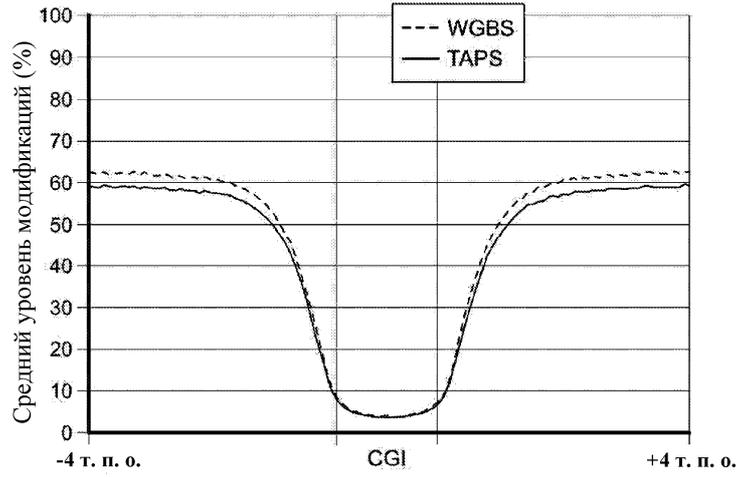


Е Качество секвенирования оснований

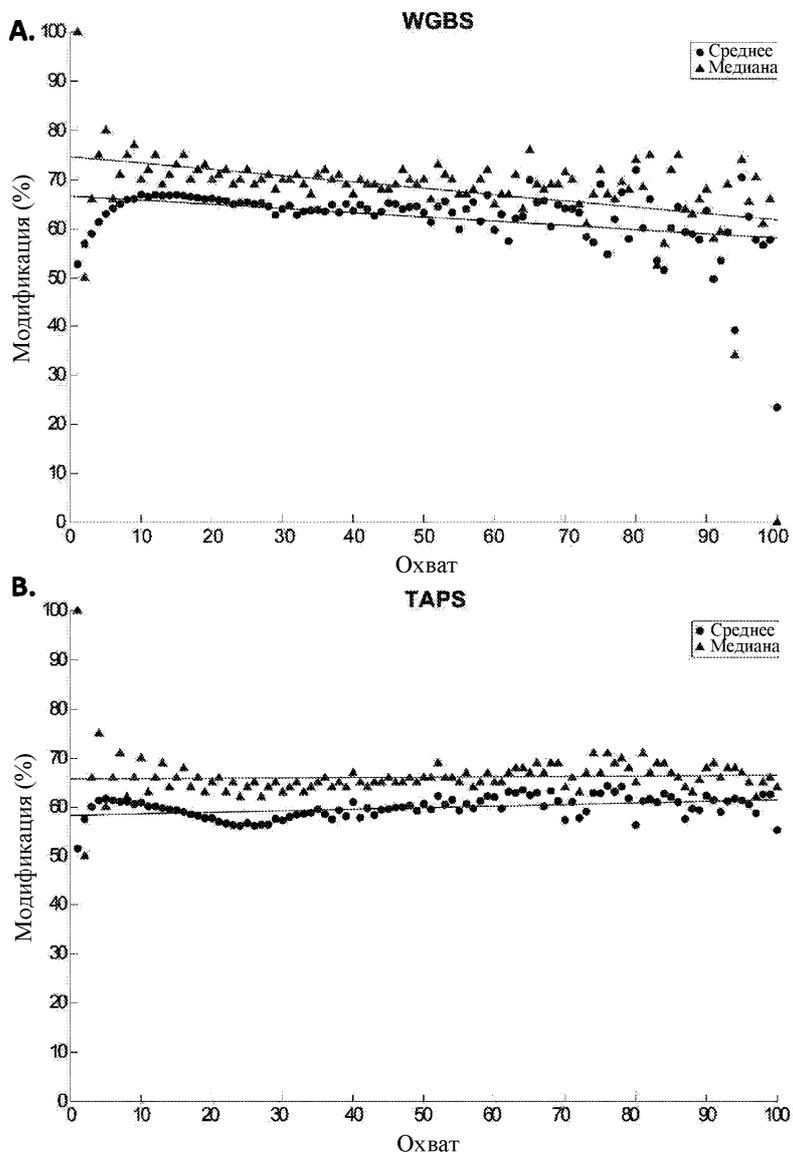
Фиг. 18



Фиг. 21



Фиг. 22



Фиг. 23

А Наборы для получения библиотек dsDNA

Вводимое количество ДНК (нг):	NEBNext Ultra II			KAPA HyperPlus		
	100	10	1	100	10	1
500 п. о.						
400 п. о.						
300 п. о.						
200 п. о.						
Циклы ПЦР:	8	12	16	9	13	17
Концентрация библиотеки в 10 мкл (нМ):	68	43	15	120	75	15

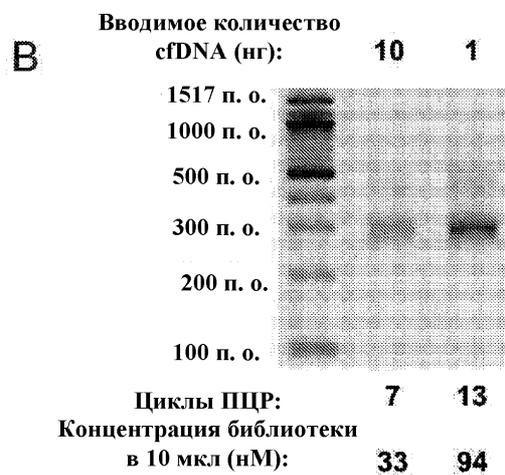
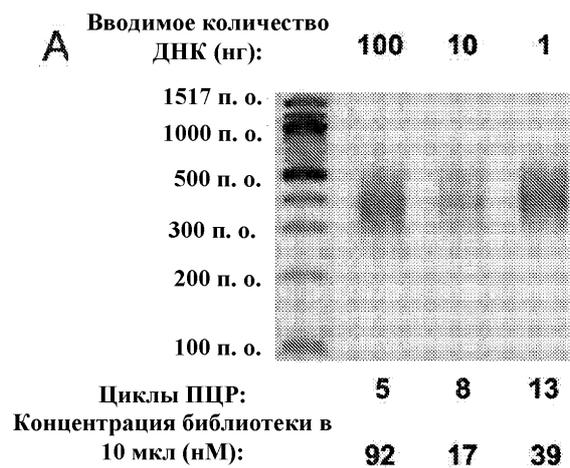
В Набор для получения библиотек ssDNA

Вводимое количество ДНК (нг):	Accel-NGS Methyl-Seq				
	100	10	1	0,1	0,01
500 п. о.					
400 п. о.					
300 п. о.					
200 п. о.					
Циклы ПЦР:	6	9	13	16	21
Концентрация библиотеки в 10 мкл (нМ):	42	22	6,9	6,1	4,9

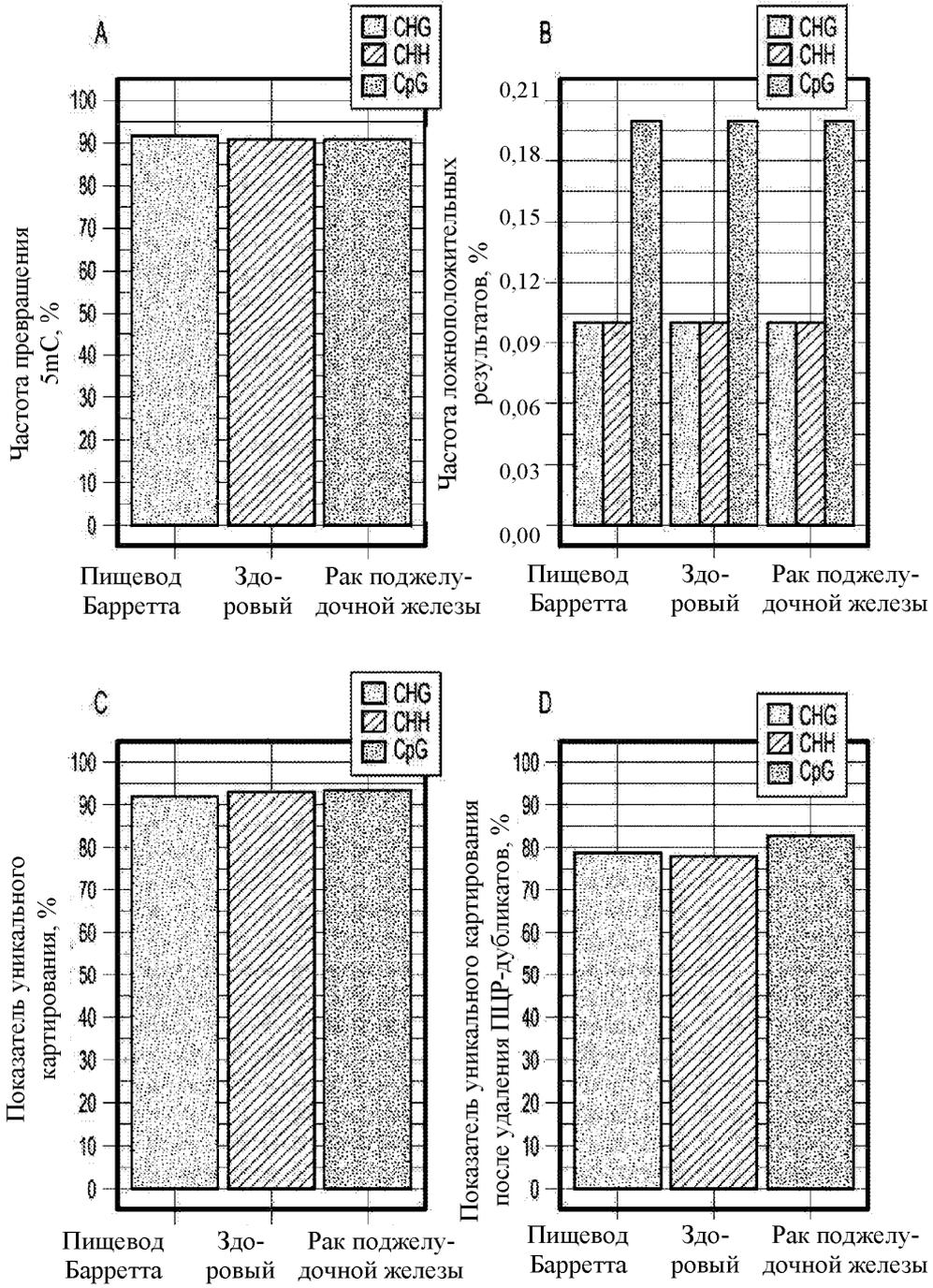
Вводимое количество cfDNA (нг):	:	
	10	1
500 п. о.		
400 п. о.		
300 п. о.		
200 п. о.		
100 п. о.		

Циклы ПЦР:	9	13
Концентрация библиотеки в 10 мкл (нМ):	23	13

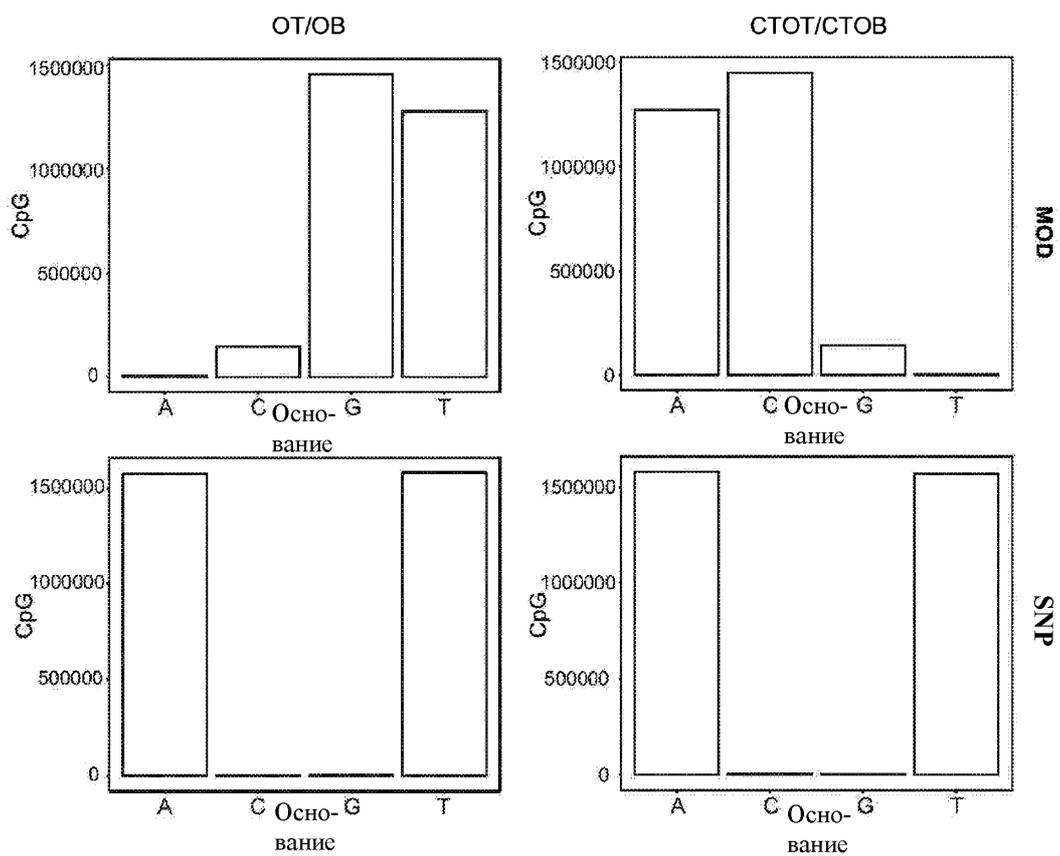
Фиг. 24



Фиг. 25



Фиг. 26



Фиг. 27



Евразийская патентная организация, ЕАПВ

Россия, 109012, Москва, Малый Черкасский пер., 2