

(19)



Евразийское патентное ведомство

(21) 201500327

(13) A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки 2015.07.30

(51) Int. Cl. C12Q 1/68 (2006.01)

(22) Дата подачи заявки 2013.09.20

(54) НЕИНВАЗИВНОЕ ОПРЕДЕЛЕНИЕ МЕТИЛОМА ПЛОДА ИЛИ ОПУХОЛИ ПО ПЛАЗМЕ

(31) 61/703,512; 13/842,209; 61/830,571
(32) 2012.09.20; 2013.03.15; 2013.06.03
(33) US
(86) PCT/AU2013/001088
(87) WO 2014/043763 2014.03.27
(71) Заявитель:

ТЕ ЧАЙНИЗ ЮНИВЕРСИТИ ОВ ГОНКОНГ (CN)

(72) Изобретатель:
Чиу Росса Вай Квун, Чань Кван Чи, Ло Юйк-Мин Деннис, Лунь Мю Фань, Цзян Пэйюн, Чань Вай Мань (CN)

(74) Представитель:
Бадаева Т.Н. (RU)

(57) Системы, способы и устройства позволяют определять и применять профили метилирования различных тканей и образцов. Представлены примеры. Профиль метилирования может быть определен для плодной/опухолевой ткани на основе сравнения метилирования в плазме (или другом образце, содержащем свободную от клеток ДНК) с профилем метилирования матери/пациента. Профиль метилирования может быть определен для плодной/опухолевой ткани с применением тканеспецифических аллелей для идентификации ДНК плода/опухоли, если образец содержит смесь ДНК. Профиль метилирования может применяться для определения вариаций числа копий в геноме плода/опухоли. Маркеры метилирования для плода идентифицировали с помощью различных методов. Профиль метилирования может быть определен путем определения размерного параметра распределения размеров фрагментов ДНК, при этом могут использоваться референсные значения для размерного параметра для определения уровней

метилирования. Кроме того, уровень метилирования может использоваться для определения уровня рака.

Table with columns: Метилирование, OS, Зоны мажор, Метилирование, Метилирование, Метилирование. Rows include various CpG sites and methylation levels.



A1 201500327

201500327 A1

НЕИНВАЗИВНОЕ ОПРЕДЕЛЕНИЕ МЕТИЛОМА ПЛОДА ИЛИ ОПУХОЛИ ПО ПЛАЗМЕ

Перекрестные ссылки на родственную заявку

Настоящая заявка представляет собой заявку РСТ, испрашивающую приоритет по предварительной заявке на патент США № 61/830571, озаглавленной: «Tumor Detection In Plasma Using Methylation Status And Copy Number» («Детекция опухолей в плазме с использованием данных о статусе метилирования и количестве копий»), поданной 3 июня 2013 г.; и заявке на патент США № 13/842209, озаглавленной: «Non-Invasive Determination Of Methylome Of Fetus Or Tumor From Plasma» («Неинвазивное определение метилома плода или опухоли по плазме»), поданной 15 марта 2013 г., которая представляет собой непредварительную заявку и испрашивает приоритет по предварительной заявке на патент США № 61/703512, озаглавленной: «Method Of Determining The Whole Genome DNA Methylation Status Of The Placenta By Massively Parallel Sequencing Of Maternal Plasma» («Способ определения статуса метилирования ДНК полного генома плаценты с применением массивно-параллельного секвенирования материнской плазмы»), поданной 20 сентября 2012 г., которые включены в настоящий документ посредством ссылки полностью во всех отношениях.

Область техники

Раскрытое в настоящем описании изобретение относится в целом к определению паттерна метилирования (метилома) ДНК, и, конкретнее, к анализу биологического образца (например, плазмы), который содержит смесь ДНК разных геномов (например, плода и матери, или опухоли и нормальных клеток), для определения паттерна метилирования (метилома) минорного генома. Также описаны варианты применения установленного метилома.

Уровень техники

Развитие эмбриона и плода представляет собой сложный процесс и включает множество высококоординированных генетических и эпигенетических событий. Развитие ракового заболевания также представляет собой сложный процесс, включающий ряд генетических и эпигенетических этапов. Аномалии эпигенетического контроля процессов развития связаны с бесплодием, самопроизвольными абортами, аномалиями внутриутробного роста и постнатальными последствиями. Метилирование ДНК является

одним из наиболее интенсивно исследуемых эпигенетических механизмов. Метилирование ДНК в основном происходит в контексте добавления метильной группы к 5'-концевому атому углерода остатков цитозина в динуклеотидах CpG. Метилирование цитозина добавляет еще один уровень контроля транскрипции генов и функции ДНК. Например, гиперметилирование генных промоторов, богатых динуклеотидами CpG и называемых CpG-островками, как правило, связано с подавлением функции гена.

Несмотря на важную роль эпигенетических механизмов в опосредовании процессов развития, ткани эмбриона и плода человека не являются легкодоступными для анализа (аналогичным образом могут быть недоступны и опухоли). Исследования динамических изменений таких эпигенетических процессов в норме и при наличии заболеваний во время внутриутробного периода у человека практически невозможны. Одно из главных направлений таких исследований было представлено внеэмбриональными тканями, в частности, плацентой, которые могут быть получены в ходе пренатальных диагностических процедур или после рождения ребенка. Однако для таких тканей требуется осуществление инвазивных процедур.

Профиль метилирования ДНК плаценты человека интересовал ученых в течение многих десятилетий. Плацента человека демонстрирует множество своеобразных физиологических особенностей, задействующих метилирование ДНК. На глобальном уровне плацентарные ткани гипометилированы по сравнению с большинством соматических тканей. На уровне генов статус метилирования отдельных геномных локусов является специфической сигнатурой плацентарных тканей. Как глобальный, так и локус-специфический профили метилирования демонстрируют зависимые от гестационного возраста изменения. Импринтинговые гены, а именно гены, экспрессия которых зависит от родительского происхождения аллелей, осуществляют ключевые функции в плаценте. Плацента была описана как предзлокачественное образование, и наблюдалось гиперметилирование нескольких генов-супрессоров опухолей.

Исследования профиля метилирования ДНК плацентарных тканей обеспечили ценную информацию для понимания патофизиологии ассоциированных с беременностью или связанных с процессами развития заболеваний, таких как преэклампсия и задержка внутриутробного развития. Расстройства геномного импринтинга связаны с такими расстройствами развития, как синдром Прадера-Вилли и синдром Ангельмана. Измененные профили геномного импринтинга и глобального метилирования ДНК в плацентарных и плодных тканях наблюдались при беременностях, наступивших в результате применения вспомогательных репродуктивных технологий (H Niura et al. 2012 Hum Reprod; 27: 2541-2548). Ряд факторов окружающей среды, таких как курение матери

(KE Haworth et al. 2013 Epigenomics; 5: 37-49), питание матери (X Jiang et al. 2012 FASEB J; 26: 3563-3574) и метаболический статус матери, например, диабет (N Hajj et al., Diabetes. ЦИО: 10.2337/db12-0289), были ассоциированы с эпигенетическими абберациями у потомства.

Несмотря на десятилетия работы, отсутствовали какие-либо практические средства для изучения метилома плода или опухоли и для мониторинга динамических изменений на протяжении всей беременности или во время болезненных процессов, например, при злокачественных новообразованиях. Таким образом, существует потребность в неинвазивных способах анализа полного плодного и опухолевого метилома или его фрагментов.

Краткое описание изобретения

В вариантах реализации представлены системы, способы и устройства для определения и применения профилей метилирования различных тканей и образцов. Приведены примеры. Профиль метилирования может быть выведен для плодной/опухолевой ткани на основании сравнения метилирования в плазме (или другом образце, содержащем ДНК, свободную от клеток, например, образце мочи, слюны, смывах с половых органов) с профилем метилирования матери/пациента. Профиль метилирования может быть определен для плодной/опухолевой ткани с применением тканеспецифических аллелей для идентификации ДНК плода/опухоли, если образец содержит смесь ДНК. Профиль метилирования может применяться для определения вариации числа копий в геноме плода/опухоли. Маркеры метилирования у плода идентифицировали с помощью различных методик. Профиль метилирования может быть определен путем определения размерного параметра распределения размеров фрагментов ДНК, при этом референсные значения размерного параметра могут применяться для определения уровней метилирования.

Кроме того, уровень метилирования может применяться для определения уровня рака. В контексте раковых заболеваний измерение изменений метилома в плазме может позволять детектировать рак (например, для целей скрининга), проводить мониторинг (например, детектировать отклик на противораковую терапию; и детектировать рецидив рака) и осуществлять прогнозирование (например, измерение нагрузки организма раковыми клетками или для стадирования, оценки вероятности летального исхода в результате заболевания, прогрессирования заболевания или метастатических процессов).

Лучшее понимание природы и преимуществ вариантов реализации настоящего изобретения может быть достигнуто на основе приведенного ниже подробного описания и

сопроводительных чертежей.

Краткое описание чертежей

На фиг. 1А приведена таблица 100 результатов секвенирования для материнской крови, плаценты и материнской плазмы в соответствии с вариантами реализации настоящего изобретения.

На фиг. 1В показана плотность метилирования участков размером 1 Мб из секвенированных образцов в соответствии с вариантами реализации настоящего изобретения.

На фиг. 2А-2С приведены графики зависимости бета-значений от индексов метилирования: (А) Клетки материнской крови, (В) Образец ворсин хориона, (С) Ткань зрелой плаценты.

На фиг. 3А и 3В приведены гистограммы процентного содержания метилированных CpG-сайтов в плазме и в клетках крови, полученных от взрослого мужчины и небеременной взрослой женщины: (А) Аутосомы, (В) X-хромосома.

На фиг. 4А и 4В приведены графики плотностей метилирования соответствующих локусов в ДНК клеток крови и ДНК плазмы: (А) Небеременная взрослая женщина, (В) Взрослый мужчина.

На фиг. 5А и 5В приведены гистограммы процентного содержания метилированных CpG-сайтов в образцах, полученных при беременности: (А). Аутосомы, (В) X-хромосома.

На фиг. 6 приведена гистограмма уровня метилирования разных классов повторов в геноме человека для материнской крови, плаценты и материнской плазмы.

На фиг. 7А приведен Circos-график 700 для образцов, полученных в первом триместре. На фиг. 7В приведен Circos-график 750 для образцов, полученных в третьем триместре.

На фиг. 8А-8Д приведены графики сравнения плотностей метилирования геномной тканевой ДНК с ДНК материнской плазмы для CpG-сайтов, окружающих информативные однонуклеотидные полиморфизмы.

Фиг. 9 представляет собой блок-схему, иллюстрирующую способ 900 определения первого профиля метилирования по биологическому образцу организма в соответствии с вариантами реализации настоящего изобретения.

Фиг. 10 представляет собой блок-схему, иллюстрирующую способ 1000 определения первого профиля метилирования по биологическому образцу организма в соответствии с вариантами реализации настоящего изобретения.

На фиг. 11А и 11В приведены графические представления эффективности алгоритма прогнозирования с применением данных материнской плазмы и фракционной концентрации плодной ДНК в соответствии с вариантами реализации настоящего изобретения.

Фиг. 12А представляет собой таблицу 1200, где приведены подробные данные для 15 выбранных геномных локусов для прогнозирования метилирования в соответствии с вариантами реализации настоящего изобретения. На фиг. 12В приведено графическое представление 1250 выведенных категорий для 15 выбранных геномных локусов и соответствующих им уровней метилирования в плаценте.

Фиг. 13 представляет собой блок-схему способа 1300 детектирования хромосомной аномалии плода по биологическому образцу, полученному от субъекта – беременной по меньшей мере одним плодом женщины.

Фиг. 14 представляет собой блок-схему способа 1400 идентификации маркеров метилирования путем сравнения профиля метилирования плаценты с материнским профилем метилирования в соответствии с вариантами реализации настоящего изобретения.

Фиг. 15А представляет собой таблицу 1500, представляющую эффективность алгоритма идентификации DMR с применением данных для первого триместра с учетом 33 ранее описанных маркеров первого триместра. Фиг. 15В представляет собой таблицу 1550, представляющую эффективность алгоритма идентификации DMR с применением данных для третьего триместра, сравниваемых с образцом плаценты, полученным при родах.

Фиг. 16 представляет собой таблицу 1600, где приведены количества локусов, гиперметилированных или гипометилированных согласно прогнозу на основе прямого анализа данных бисульфитного секвенирования для материнской плазмы.

Фиг. 17А представляет собой график 1700, демонстрирующий распределение размеров ДНК материнской плазмы, контрольной плазмы небеременной женщины, плацентарной ДНК и ДНК периферической крови. Фиг. 17В представляет собой график 1750 распределения размеров и профиля метилирования материнской плазмы, контрольной плазмы от взрослой женщины, плацентарной ткани и контрольной крови от взрослой женщины.

Фиг. 18А и 18В представляют собой графики плотностей метилирования и размеров молекул ДНК плазмы в соответствии с вариантами реализации настоящего изобретения.

На фиг. 19А приведен график 1900 плотностей метилирования и размеров ридов

секвенирования для взрослой небеременной женщины. Фиг. 19В представляет собой график 1950, демонстрирующий распределение размеров и профиль метилирования специфичных для плода и специфичных для матери молекул ДНК в материнской плазме.

Фиг. 20 представляет собой блок-схему способа 2000 оценки уровня метилирования ДНК в биологическом образце организма в соответствии с вариантами реализации настоящего изобретения.

Фиг. 21А представляет собой таблицу 2100, где приведены плотности метилирования для предоперационных образцов плазмы и ткани пациента с гепатоцеллюлярной карциномой (ГЦК). Фиг. 21В представляет собой таблицу 2150, где приведено количество рядов последовательностей и достигнутая глубина секвенирования на образец.

Фиг. 22 представляет собой таблицу 220, где приведены плотности метилирования в аутосомах, варьирующие от 71,2% до 72,5%, в образцах плазмы здоровых контролей.

На фиг. 23А и 23В приведена плотность метилирования лейкоцитарной пленки, опухолевой ткани, неопухолевой ткани печени, предоперационной плазмы и послеоперационной плазмы пациента с ГЦК.

Фиг. 24А представляет собой график 2400, демонстрирующий плотности метилирования предоперационной плазмы пациента с ГЦК. Фиг. 24В представляет собой график 2450, демонстрирующий плотности метилирования послеоперационной плазмы пациента с ГЦК.

На фиг. 25А и 25В приведены Z-показатели плотностей метилирования ДНК плазмы для предоперационных (график 2500) и послеоперационных (график 2550) образцов плазмы пациента с ГЦК с применением данных о метиломе плазмы четырех здоровых контрольных субъектов в качестве эталона для хромосомы 1.

Фиг. 26А представляет собой таблицу 2600, где приведены данные Z-показателей для предоперационной и послеоперационной плазмы. Фиг. 26В представляет собой Circos-график 2620, отражающий Z-показатель плотностей метилирования ДНК плазмы для предоперационных и послеоперационных образцов плазмы пациента с ГЦК с использованием четырех здоровых контрольных субъектов в качестве эталона для участков размером 1 Мб, проанализированных для всех аутосом. Фиг. 26С представляет собой таблицу 2640, представляющую распределение Z-показателей участков размером 1 Мб для полного генома как в предоперационных, так и в послеоперационных образцах плазмы пациента с ГЦК. Фиг. 26D представляет собой таблицу 2660, где представлены уровни метилирования в образце опухолевой ткани и предоперационной плазмы, перекрывающиеся с некоторыми из контрольных образцов плазмы в контексте СНН и

CHG.

На фиг. 27А-Н представлен Circos-график плотности метилирования у 8 пациентов с раковыми заболеваниями в соответствии с вариантами реализации настоящего изобретения. Фиг. 27I представляет собой таблицу 2780, где приведено количество ридов последовательностей и достигнутая глубина секвенирования на образец. Фиг. 27J представляет собой таблицу 2790, представляющую распределение Z-показателей участков размером 1 Мб для полного генома в плазме пациентов с разными злокачественными новообразованиями. АЛ = аденокарцинома легкого; НГК = носоглоточная карцинома; КРК = колоректальная карцинома; НЭК = нейроэндокринная карцинома; ГМС = гладкомышечная саркома.

Фиг. 28 представляет собой блок-схему способа 2800 анализа биологического образца организма для установления классификации уровня рака в соответствии с вариантами реализации настоящего изобретения.

Фиг. 29А представляет собой график 2900, демонстрирующий распределение плотностей метилирования у референсных субъектов при предположении, что указанное распределение соответствует нормальному распределению. Фиг. 29В представляет собой график 2950, демонстрирующий распределение плотностей метилирования у пациентов с раковым заболеванием при предположении, что указанное распределение соответствует нормальному распределению, и среднее значение уровня метилирования ниже точки отсечения на 2 стандартных отклонения.

Фиг. 30 представляет собой график 3000, демонстрирующий распределение плотностей метилирования ДНК плазмы здоровых субъектов и пациентов с раковыми заболеваниями.

На фиг. 31 приведено графическое представление 3100 распределения различий среднего плотностей метилирования ДНК плазмы здоровых субъектов и опухолевой ткани пациента с ГЦК.

Фиг. 32А представляет собой таблицу 3200, где представлен эффект уменьшения глубины секвенирования в случаях, когда образец плазмы содержал 5% или 2% опухолевой ДНК.

На фиг. 32В приведено графическое представление 3250 плотностей метилирования повторяющихся элементов и областей без повторов в плазме четырех здоровых контрольных субъектов, в лейкоцитарной пленке, нормальной ткани печени, опухолевой ткани, образцах предоперационной плазмы и послеоперационной плазмы пациента с ГЦК.

На фиг. 33 представлена блок-диаграмма примера компьютерной системы 3300,

подходящей для применения с системой и способами в соответствии с вариантами реализации настоящего изобретения.

На фиг. 34А представлено распределение размеров ДНК плазмы у пациента с системной красной волчанкой (СКВ) СКВ04. На фиг. 34В и 34С представлен анализ метилирования для ДНК плазмы пациента с СКВ СКВ04 (фиг. 34В) и пациента с ГЦК TBR36 (фиг. 34С).

Фиг. 35 представляет собой блок-схему способа 3500 установления классификации уровня рака на основе гиперметилирования CpG-островков в соответствии с вариантами реализации настоящего изобретения.

Фиг. 36 представляет собой блок-схему способа 3600 анализа биологического образца организма с применением множества хромосомных областей в соответствии с вариантами реализации настоящего изобретения.

На фиг. 37А представлен анализ CNA (аббераций числа копий) для опухолевых тканей, не обработанной бисульфитом (БС) ДНК плазмы и обработанной бисульфитом ДНК плазмы (от центра к краям) для пациента TBR36. Фиг. 37В представляет собой график рассеяния, отражающий зависимость между Z-показателями для детекции CNA, с применением обработанной бисульфитом и не обработанной бисульфитом плазмы, для участков размером 1 Мб у пациента TBR36.

На фиг. 38А представлен анализ CNA для опухолевых тканей, не обработанной бисульфитом (БС) ДНК плазмы и обработанной бисульфитом ДНК плазмы (от центра к краям) для пациента TBR34. Фиг. 38В представляет собой график рассеяния, отражающий зависимость между Z-показателями для детекции CNA, с применением обработанной бисульфитом и не обработанной бисульфитом плазмы, для участков размером 1 Мб у пациента TBR34.

Фиг. 39А представляет собой Circos-график, демонстрирующий CNA (внутренний круг) и анализ метилирования (внешний круг) для обработанной бисульфитом плазмы пациента с ГЦК TBR240. Фиг. 39В представляет собой Circos-график, демонстрирующий CNA (внутренний круг) и анализ метилирования (внешний круг) для обработанной бисульфитом плазмы пациента с ГЦК TBR164.

На фиг. 40А представлен анализ CNA у пациента TBR36 образца до лечения и образца после лечения. На фиг. 40В представлен анализ метилирования у пациента TBR36 образца до лечения и образца после лечения. На фиг. 41А представлен анализ CNA у пациента TBR34 образца до лечения и образца после лечения. Фиг. 41В представлен анализ метилирования у пациента TBR34 образца до лечения и образца после лечения.

На фиг. 42 представлена диаграмма диагностической эффективности

полногеномного анализа гипометилирования при разном количестве ридов секвенирования.

Фиг. 43 представляет собой диаграмму, демонстрирующую кривые ROC для детекции ракового заболевания на основе полногеномного анализа гипометилирования при разном размере участков (50 кб, 100 кб, 200 кб и 1 Мб).

На фиг. 44А представлена диагностическая эффективность суммарной вероятности (CP) и процентное содержание участков с абберациями. На фиг. 44В представлена диагностическая эффективность анализа плазмы для оценки глобального гипометилирования, CpG-островков гиперметилирования и CNA.

На фиг. 45 приведена таблица результатов для глобального гипометилирования, CpG-островков гиперметилирования и CNA у пациентов с гепатоцеллюлярной карциномой.

На фиг. 46 приведена таблица результатов для глобального гипометилирования, CpG-островков гиперметилирования и CNA у пациентов с раковыми заболеваниями, отличными от гепатоцеллюлярной карциномы.

На фиг. 47 представлен серийный анализ метилирования в плазме для пациента TBR34.

На фиг. 48А представлен Circos-график, демонстрирующий изменения CNA (внутренний круг) и метилирования (внешний круг) в обработанной бисульфитом ДНК плазме пациента с ГЦК TBR36. Фиг. 48В представляет собой график Z-показателей метилирования для областей с добавлением или утратой хромосомного материала, и областей без изменения числа копий для пациента с ГЦК TBR36.

На фиг. 49А представлен Circos-график, демонстрирующий изменения CNA (внутренний круг) и метилирования (внешний круг) в обработанной бисульфитом ДНК плазмы пациента с ГЦК TBR34. Фиг. 49В представляет собой график Z-показателей метилирования для областей с добавлением или утратой хромосомного материала, и областей без изменения числа копий для пациента с ГЦК TBR34.

На фиг. 50А и 50В представлены результаты анализа гипометилирования плазмы и CNA для пациентов с СКВ, СКВ04 и СКВ10.

На фиг. 51А и 51В представлен анализ Z_{meth} для областей с CNA и без CNA в плазме двух пациентов с ГЦК (TBR34 и TBR36). Фиг. 51С и 51D представлен Z_{meth} анализ для областей с CNA и без CNA в плазме двух пациентов с СКВ (СКВ04 и СКВ10).

На фиг. 52А представлен иерархический кластерный анализ для образцов плазмы от пациентов с ГЦК, пациентов с отличным от ГЦК заболеванием и здоровых контрольных субъектов с использованием признаков группы А для CNA, глобального

метилирования и метилирования CpG-островков. На фиг. 52B представлена иерархическая кластеризация с использованием признаков группы В для CNA, глобального метилирования и метилирования CpG-островков.

На фиг. 53A представлен иерархический кластерный анализ для образцов плазмы от пациентов с ГЦК, пациентов с отличным от ГЦК заболеванием и здоровых контрольных субъектов с использованием признаков метилирования CpG-островков группы А. На фиг. 53B представлен иерархический кластерный анализ образцов плазмы от пациентов с ГЦК, пациентов с отличным от ГЦК заболеванием и здоровых контрольных субъектов с использованием плотностей глобального метилирования группы А.

На фиг. 54A представлен иерархический кластерный анализ образцов плазмы от пациентов с ГЦК, пациентов с отличным от ГЦК заболеванием и здоровых контрольных субъектов с использованием глобального содержания CNA группы А. На фиг. 54B представлен иерархический кластерный анализ образцов плазмы от пациентов с ГЦК, пациентов с отличным от ГЦК заболеванием и здоровых контрольных субъектов с использованием плотностей метилирования CpG-островков группы В.

На фиг. 55A представлен иерархический кластерный анализ для образцов плазмы от пациентов с ГЦК, пациентов с отличным от ГЦК заболеванием и здоровых контрольных субъектов с использованием плотностей глобального метилирования группы В. На фиг. 55B представлен иерархический кластерный анализ образцов плазмы от пациентов с ГЦК, пациентов с отличным от ГЦК заболеванием и здоровых контрольных субъектов с использованием плотностей глобального метилирования группы В.

На фиг. 56 представлено среднее плотности метилирования участков размером 1 Мб (красные точки) у 32 здоровых субъектов.

Определения

«Метилом» представляет собой меру величины метилирования ДНК во множестве сайтов или локусов в геноме. Метилом может соответствовать всему геному, существенной части генома или относительно небольшой(им) части(ям) генома. «Плодный метилом» соответствует метилому плода беременной женщины. Плодный метилом может быть определен с использованием разнообразных плодных тканей или источников плодной ДНК, включая плацентарные ткани и свободную от клеток плодную ДНК в материнской плазме. «Опухолевый метилом» соответствует метилому опухоли организма (например, человека). Опухолевый метилом может быть определен с использованием опухолевой ткани или свободной от клеток опухолевой ДНК в материнской плазме. Плодный метилом и опухолевый метилом представляют собой

примеры представляющего интерес метилома. Другими примерами представляющих интерес метиломов являются метиломы органов (например, метиломы клеток головного мозга, костей, легких, сердца, мышц и почек, и т.д.), из которых ДНК может попадать в жидкости организма (например, плазму, сыворотку, пот, слюну, мочу, выделения половых органов, семенную жидкость, жидкость из испражнений, диарейную жидкость, спинномозговую жидкость, выделения из желудочно-кишечного тракта, выделения поджелудочной железы, выделения кишечника, мокроту, слезы, аспирационную жидкость из молочных желез и щитовидной железы, и т.д.). Указанные органы могут представлять собой трансплантированные органы.

«Метилом плазмы» представляет собой метилом, определенный по плазме или сыворотке животного (например, человека). Метилом плазмы представляет собой пример метилом свободной от клеток ДНК, поскольку плазма или сыворотка содержат свободную от клеток ДНК. Метилом плазмы является также примером смешанного метилома, поскольку он представляет собой смесь плодного/материнского метилома или метилома опухоли/пациента. «Плацентарный метилом» может быть определен по образцу ворсин хориона (CVS) или образцу плацентарной ткани (например, полученному после родов). «Клеточный метилом» соответствует метилому, определенному по клеткам (например, клеткам крови) пациента. Метилом в клетках крови называют метиломом клеток крови (или метиломом крови).

«Сайт» соответствует одному сайту, который может представлять собой одно основание или группу связанных оснований, например, CpG-сайт. «Локус» может соответствовать области, которая включает несколько сайтов. Локус может включать всего один сайт, что делает локус эквивалентным сайту в данном контексте.

«Индекс метилирования» каждого геномного сайта (например, CpG-сайта) относится к соотношению ридов секвенирования, демонстрирующих метилирование в указанном сайте, к общему количеству ридов, покрывающих указанный сайт. «Плотность метилирования» области представляет собой число демонстрирующих метилирование ридов на сайтах внутри указанной области, разделенное на общее число ридов, покрывающих сайты в указанной области. Сайты могут иметь специфические характеристики, например, представлять собой CpG-сайты. Соответственно, «плотность метилирования CpG» для области представляет собой число демонстрирующих CpG метилирование ридов, разделенное на общее число ридов, покрывающих CpG-сайты в указанной области (например, конкретный CpG-сайт, CpG-сайты в составе CpG-островка или область большего размера). Например, плотность метилирования для каждого 100-кб участок в геноме человека может быть определена по общему числу цитозинов, не

конвертированных после обработки бисульфитом (что соответствует метилированному цитозину) в CpG-сайтах, как пропорциональной доли от всех CpG-сайтов, покрываемых ридами последовательностей, картируемыми на область размером 100 кб. Указанный анализ может также проводиться для участков других размеров, например, 50 кб или 1 Мб, и т.д. Область может представлять собой весь геном, или хромосому, или часть хромосомы (например, плечо хромосомы). Индекс метилирования CpG-сайта совпадает с плотностью метилирования области, если указанная область включает только указанный CpG-сайт. «Доля метилированных цитозинов» относится к отношению числа цитозиновых сайтов, «С», которые, как показано, метилированы (например, не конвертированы после обработки бисульфитом), к общему количеству анализируемых остатков цитозина, т.е. в том числе, цитозинов вне контекста CpG, в указанной области. Индекс метилирования, плотность метилирования и доля метилированных цитозинов представляют собой примеры «уровней метилирования».

«Профиль метилирования» (также называемый статусом метилирования) включает информацию относительно метилирования ДНК для области. Информация относительно метилирования ДНК может включать, не ограничиваясь перечисленным, индекс метилирования CpG-сайта, плотность метилирования CpG-сайтов в области, распределение CpG-сайтов на протяжении непрерывной области, паттерн или уровень метилирования для каждого индивидуального CpG-сайта в составе области, которая содержит более чем один CpG-сайт, и метилирования по сайтам, не являющимся CpG. Профиль метилирования существенной части генома может считаться эквивалентным метилому. «Метилирование ДНК» в геномах млекопитающих, как правило, относится к добавлению метильной группы к 5'-концевому атому углерода остатков цитозина (т.е. 5-метилцитозинам) в динуклеотидах CpG. Метилирование ДНК по цитозинам может происходить в других контекстах, например, в CHG и CHH, где H представляет собой аденин, цитозин или тимин. Метилирование цитозина может также быть представлено 5-гидроксиметилцитозином. Также было описано метилирование по не являющимся цитозинам основаниям, например, по N6-метиладенину.

«Ткань» соответствует любым клеткам. Разные типы ткани могут соответствовать разным типам клеток (например, печени, легкого или крови), а также могут соответствовать тканям разных организмов (матери/плода) или здоровым клеткам/опухолевым клеткам. «Биологический образец» относится к любому образцу, который получают от субъекта (например, человека, такого как беременная женщина, индивидуум с раковым заболеванием или индивидуум, у которого предположительно имеется раковое заболевание, реципиент трансплантированного органа или субъект, у

которого предположительно имеется болезненный процесс, в который вовлечен орган (например, сердце при инфаркте миокарда, или головной мозг при инсульте), и содержит одну или большее количество представляющих интерес молекул нуклеиновой кислоты. Биологический образец может представлять собой жидкость организма, такую как кровь, плазма, сыворотка, моча, влагалищное отделяемое, смывы из матки или влагалища, плевральная жидкость, асцитическая жидкость, спинномозговая жидкость, слюна, пот, слеза, мокрота, жидкость бронхоальвеолярного лаважа и т.д. Также могут использоваться образцы стула.

Термин «уровень рака» может относиться как к наличию ракового заболевания, так и к стадии ракового заболевания, размеру опухоли, наличию метастазов, общей опухолевой нагрузке на организм и/или к другой мере тяжести ракового заболевания. Уровень рака может быть выражен числом или обозначен иначе. Уровень может быть равен 0. Уровень рака также включает предзлокачественные или предраковые состояния (статусы), связанные с мутациями или рядом мутаций. Уровень рака может быть использован различным образом. Например, скрининг может показывать, имеется ли раковое заболевание у кого-либо, по имеющимся сведениям ранее не страдавшего раковым заболеванием. Оценка может проводиться для кого-либо с диагностированным раковым заболеванием для мониторинга прогрессирования ракового заболевания с течением времени, исследования эффективности терапии или для определения прогноза. Согласно одному варианту реализации прогноз может быть представлен в виде вероятности смерти пациента в результате ракового заболевания, или вероятности прогрессирования ракового заболевания через определенный период или промежуток времени, или вероятности метастазирования рака. Детекция может означать «скрининг» или может означать проверку на наличие рака у кого-либо, демонстрирующего указывающие на раковое заболевание признаки (например, симптомы или другие положительные тесты).

Подробное описание изобретения

Эпигенетические механизмы играют важную роль в развитии эмбриона и плода. Однако ткани эмбриона и плода человека (в том числе плацентарные ткани) не являются легкодоступными (патент США №6927028). Некоторые варианты осуществления направлены на решение указанной проблемы путем анализа образца, содержащего молекулу свободной от клеток плодной ДНК, присутствующей в материнском кровотоке. Плодный метилом может быть выведен различными способами. Например, может быть проведено сравнение метилома материнской плазмы и материнского клеточного метилома

(по клеткам крови матери), и обнаружены различия, которые коррелируют с плодным метилом. Согласно другому примеру специфичные для плода аллели могут применяться для определения метилирования плодного метилома в конкретных локусах. Кроме того, размер фрагмента может применяться в качестве индикатора процента метилирования, так как показана корреляция между размером и процентом метилирования.

Согласно одному варианту реализации полногеномное бисульфитное секвенирование применяется для анализа профиля метилирования (всего метилома или его части) ДНК материнской плазмы с разрешением до одного нуклеотида. При использовании различных полиморфизмов матери и плода можно собрать плодный метилом по образцам материнской крови. Согласно еще одному варианту реализации различия полиморфизмов не использовались, однако может использоваться различие между метиломом плазмы и метиломом клеток крови.

Согласно другому варианту реализации за счет использования однонуклеотидных вариаций и/или аббераций числа копий между геномом опухоли и неопухолевым геномом, а также данных секвенирования для плазмы (или другого образца), может быть получен профиль метилирования опухоли для образца от пациента, у которого предположительно или точно имеется раковое заболевание. Отличие уровня метилирования в образце плазмы тестируемого индивидуума при сравнении с уровнем метилирования плазмы здорового контроля или группы здоровых контролей может обеспечивать идентификацию тестируемого индивидуума как имеющего раковое заболевание. Кроме того, профиль метилирования может выполнять роль сигнатуры типа ракового заболевания, например, показывать, из какого органа происходит рак у данного индивидуума, и имеются ли метастазы.

Благодаря неинвазивной природе указанного способа авторы настоящего изобретения имели возможность провести серийную оценку плодных метиломов и метиломов материнской плазмы по образцам материнской крови, собранным в первом триместре, третьем триместре и после родов. Наблюдались изменения, связанные с гестацией. Указанный способ также может применяться в отношении образцов, полученных на протяжении второго триместра. Плодный метилом, определенный по материнской плазме во время беременности, напоминал плацентарный метилом. Импринтинговые гены и дифференциально метилированные области идентифицировали по данным материнской плазмы.

Таким образом, авторы разработали способ серийного комплексного неинвазивного исследования плодного метилома неинвазивным, предлагающего,

соответственно, возможность идентификации биомаркеров или прямого тестирования связанных с патологиями беременности. Варианты реализации также могут использоваться для серийного комплексного неинвазивного исследования опухолевого метилома, для скрининга или детекции наличия у субъекта ракового заболевания, для мониторинга злокачественных заболеваний у больного раковым заболеванием пациента и для прогнозирования. Варианты реализации могут быть применены к любому типу ракового заболевания, включая, но не ограничиваясь перечисленным, рак легкого, рак молочной железы, рак толстой и прямой кишки, рак предстательной железы, рак носоглотки, рак желудка, рак яичка, рак кожи (например, меланома), поражающий нервную систему рак, рак костей, рак яичника, рак печени (например, гепатоцеллюлярная карцинома), гематологические злокачественные новообразования, рак поджелудочной железы, эндометриоидная карцинома, рак почки, рак шейки матки, рак мочевого пузыря и т.д.

Сначала обсуждаются способы определения метилома или профиля метилирования, затем описываются различные метиломы (такие как плодные метиломы, опухолевый метилом, метиломы матери или пациента, и смешанный метилом, например, плазмы). Затем описано определение профиля метилирования у плода с использованием специфичных для плода маркеров или путем сравнения смешанного профиля метилирования с клеточным профилем метилирования. Плодные маркеры метилирования определяют путем сравнения профилей метилирования. Обсуждается зависимость между размером и метилированием. Также предложены варианты применения профилей метилирования для детекции раковых заболеваний.

I. Определение метилома

Для исследования плацентарного метилома использовалось множество подходов, однако каждый подход имеет свои ограничения. Например, бисульфит натрия, химическое вещество, которое модифицирует неметилированные остатки цитозина, превращая их в урацил, и оставляет метилированный цитозин в неизменном виде, преобразует различия в метилировании цитозина в различия генетических последовательностей для дальнейшего изучения. Золотой стандарт исследований метилирования цитозина основан на обработке тканевой ДНК бисульфитом натрия с последующим прямым секвенированием индивидуальных клонов конвертированных бисульфитом молекул ДНК. После анализа множества клонов молекул ДНК могут быть получены паттерн метилирования цитозина и количественный профиль в расчете на CpG-сайт. Тем не менее, бисульфитное секвенирование клонов представляет собой трудоемкую процедуру с низкой пропускной способностью, которую невозможно с легкостью

применять в полногеномном масштабе.

Чувствительные к метилированию рестрикционные ферменты, которые, как правило, расщепляют неметилированную ДНК, обеспечивают недорогой способ исследования метилирования ДНК. Однако данные, обеспечиваемые такими исследованиями, ограничены локусами, содержащими распознаваемые ферментами мотивы, и результаты не являются количественными. Иммунопреципитация ДНК, связываемой антителами против метилированного цитозина, может применяться для исследования больших сегментов генома, однако для него характерна тенденция к смещению в пользу локусов с высокой плотностью метилирования из-за более интенсивного связывания антител с такими областями. Способы на основе микроматриц зависят от изначальной конструкции аналитических зондов и эффективности гибридизации зондов с целевой ДНК.

Для комплексного изучения метилома в некоторых вариантах реализации используется массивно-параллельное секвенирование (МПС) для получения полногеномной информации и количественной оценки уровня метилирования в расчете на нуклеотид и на аллель. Недавно стало возможным осуществление бисульфитной конверсии с последующим полногеномным МПС (R Lister et al 2008 Cell; 133: 523-536).

Из небольшого числа опубликованных исследований (R Lister et al. 2009 Nature; 462: 315-322; L Laurent et al. 2010 Genome Res; 20: 320-331; Y Li et al. 2010 PLoS Biol; 8: e1000533; и M Kulis et al. 2012 Nat Genet; 44: 1236-1242), где применялось полногеномное бисульфитное секвенирование для исследования метиломов человека, два исследования фокусируются на эмбриональных стволовых клетках и фетальных фибробластах (R Lister et al. 2009 Nature; 462: 315-322; L Laurent et al. 2010 Genome Res; 20: 320-331). В обоих исследованиях исследовали полученную из клеточных линий ДНК.

А. Полногеномное бисульфитное секвенирование

Некоторые варианты реализации позволяют преодолеть вышеописанные проблемы и проводить комплексное неинвазивное серийное исследование плодного метилома. Согласно одному варианту реализации использовали полногеномное бисульфитное секвенирование для анализа свободных от клеток молекул плодной ДНК, присутствующих в кровотоке беременных женщин. Несмотря на незначительное содержание и фрагментированную природу молекул ДНК в плазме, авторам удалось собрать плодный метилом с высоким разрешением по материнской плазме и провести серийные наблюдения изменений с течением беременности. Учитывая, что неинвазивное пренатальное тестирование (НИПТ) представляет значительный интерес, варианты реализации могут обеспечить новый мощный инструмент для поиска фетальных

биомаркеров или служить непосредственно в качестве платформы для проведения NPT заболеваний плода или связанных с беременностью заболеваний. Представлены данные полногеномного бисульфитного секвенирования различных образцов, исходя из которых может быть установлен плодный метилом. Согласно одному варианту реализации указанная технология может применяться для определения профиля метилирования при беременности, осложненной преэклампсией, или внутриутробной задержкой роста, или преждевременными родами. При таких осложненных беременностях указанная технология может применяться серийно благодаря своей неинвазивной природе, позволяя проводить мониторинг, и/или прогнозирование, и/или оценку ответа на лечение.

На фиг. 1А приведена таблица 100 результатов секвенирования для материнской крови, плаценты и материнской плазмы в соответствии с вариантами реализации настоящего изобретения. Согласно одному варианту реализации полногеномное секвенирование осуществляли на конвертированных бисульфитом библиотеках ДНК, полученных с применением адаптеров для библиотек метилированной ДНК (Illumina) (R Lister et al. 2008 Cell; 133: 523-536), клетках крови из образца крови, взятого в первом триместре, CVS, плацентарной ткани, собранной при своевременных родах, образцах материнской плазмы, собранных в первом и третьем триместрах и в послеродовом периоде. Также исследовали образцы ДНК клеток крови и плазмы, полученные от одного взрослого мужчины и одной взрослой небеременной женщины. Всего в указанном исследовании получали 9,5 миллиардов пар необработанных ридов последовательностей. Глубина секвенирования для каждого образца представлена в таблице 100.

Риды последовательностей, уникально картируемые на референсный геном человека, обеспечивали 50-кратное, 34-кратное и 28-кратное среднее покрытие гаплоидного генома, соответственно, для образцов материнской плазмы в первом триместре, третьем триместре и в послеродовой период. Покрытие CpG-сайтов в геноме варьировало от 81% до 92% для образцов, полученных при беременности. Риды последовательностей, захватывающие CpG-сайты, обеспечивали 33-кратное среднее покрытие гаплоидного генома на цепь, 23-кратное покрытие на цепь и 19-кратное покрытие на цепь, соответственно, для образцов материнской плазмы в первом триместре, третьем триместре и в послеродовой период. Эффективность бисульфитной конверсии для всех образцов составляла > 99,9% (таблица 100).

В таблице 100 показатель неоднозначности (помечен индексом «а») относится к доле ридов, картируемых на обе цепи Уотсона-Крика (и цепь 5'→3', и цепь 3'→5') референсного генома человека. Показатель конверсии лямбда относится к доле неметилированных цитозинов во внутреннем контроле, лямбда-ДНК, конвертируемых в

остатки «тимина» в результате бисульфитной модификации. Н – обобщенное обозначение А, С или Т. «а» относится к ридам, которые могут быть картированы на конкретный геномный локус, но не могут быть соотнесены с цепью Уотсона (5'→3') или Крика (3'→5'). «b» относится к парным ридам с идентичными координатами начала и конца. «с»: некоторое количество лямбда-ДНК вносили в каждый образец до бисульфитной конверсии. Показатель конверсии лямбда относится к доле цитозиновых нуклеотидов, остающихся цитозиновыми после бисульфитной конверсии и используемых в качестве индикатора скорости успешной бисульфитной конверсии. «d» относится к числу цитозиновых нуклеотидов, присутствующих в референсном геноме человека и сохраняющих последовательность цитозина после бисульфитной конверсии.

При бисульфитной модификации неметилированные цитозины превращаются в урацилы и впоследствии в тимины после ПЦР-амплификации, тогда как метилированные цитозины остаются интактными (M Frommer et al. 1992 Proc Natl Acad Sci USA;89:1827-31). После секвенирования и выравнивания статус метилирования индивидуального CpG-сайта может, соответственно, быть выведен при помощи подсчета ридов метилированных последовательностей «M» (метилированные) и ридов неметилированных последовательности «U» (неметилированные) по остаткам цитозина в контексте CpG. С применением данных бисульфитного секвенирования конструировали полные метиломы материнской крови, плаценты и материнской плазмы. Среднее плотности метилирования CpG (также называется плотностью метилирования MD) конкретных локусов в материнской плазме может быть рассчитано с применением уравнения:

$$MD = \frac{M}{M + U}$$

где M представляет собой подсчитанное число метилированных ридов и U представляет собой подсчитанное число неметилированных ридов в CpG-сайтах в составе генетического локуса. Если в состав локуса входит более чем один CpG-сайт, M и U соответствуют подсчитанному числу во всех этих сайтах.

В. Различные методики

Согласно приведенному выше описанию, определение профиля метилирования может осуществляться с применением массивно-параллельного секвенирования (МПС) конвертированной бисульфитом ДНК плазмы. МПС конвертированной бисульфитом ДНК плазмы может осуществляться случайным образом или «методом дробовика». Глубина секвенирования может варьировать в соответствии с размером представляющей интерес области.

Согласно другому варианту реализации представляющая(ие) интерес область(и) в конвертированной бисульфитом ДНК плазмы может(гут) быть сначала захвачена(ы) с

применением способа на основе жидкофазной или твердофазной гибридизации с последующим МПС. Массивно-параллельное секвенирование может осуществляться с применением платформ для секвенирования синтезом, таких как Illumina, платформ для секвенирования лигированием, таких как платформа SOLiD от Life Technologies, системы для секвенирования на основе полупроводников, такой как платформы Ion Torrent или Ion Proton от Life Technologies, или система для одномолекулярного секвенирования, такая как Helicos или Pacific Biosciences, или система для секвенирования на основе нанопор. Секвенирование на основе нанопор, в том числе нанопор, конструируемых с применением, например, липидных бислоев, белковых нанопор и твердотельных нанопор (например, на основе графена). Поскольку выбранные платформы для одномолекулярного секвенирования позволяют установить метилирование молекул ДНК (включая N6-метиладенин, 5-метилцитозин и 5-гидроксиметилцитозин) непосредственно, без бисульфитной конверсии (BA Flusberg et al. 2010 Nat Methods; 7: 461-465; J Shim et al. 2013 Sci Rep; 3:1389. ЦИО: 10.1038/srep01389), применение таких платформ позволяет проанализировать статус метилирования в образце неконвертированной бисульфитом ДНК (например, ДНК плазмы).

Наряду с секвенированием могут применяться другие методики. Согласно одному варианту реализации определение профиля метилирования может быть проведено с применением специфической к метилированию ПЦР или расщепления чувствительными к метилированию рестрикционными ферментами с последующей ПЦР или лигазной цепной реакцией с последующим ПЦР. Согласно другим вариантам реализации ПЦР представляет собой форму одномолекулярной или цифровой ПЦР (B Vogelstein et al. 1999 Proc Natl Acad Sci USA; 96: 9236-9241). Согласно дополнительным вариантам реализации ПЦР может представлять собой ПЦР в реальном времени. Согласно другим вариантам реализации ПЦР может представлять собой мультиплексную ПЦР.

II. Анализ метиломов

Некоторые варианты реализации позволяют определять профиль метилирования ДНК плазмы с применением полногеномного бисульфитного секвенирования. Профиль метилирования плода может быть определен посредством секвенирования ДНК из образцов материнской плазмы согласно приведенному ниже описанию. Соответственно, молекулы плодной ДНК (и плодный метилом) оценивали неинвазивным образом во время беременности, и изменения отслеживали серийно по мере развития беременности. Благодаря полноте данных секвенирования авторам удалось исследовать метиломы материнской плазмы в полногеномном масштабе с разрешением до одного нуклеотида.

Поскольку геномные координаты ридов секвенирования были известны, указанные

данные позволяли исследовать общие уровни метилирования метилома или любой представляющей интерес области в геноме, и провести сравнение между разными генетическими элементами. Кроме того, каждый CpG-сайт или локус покрывался несколькими рядами последовательности. В настоящем документе приводится описание некоторых метрик, использованных для измерения метилома.

А. Метилирование молекул ДНК плазмы

Молекулы ДНК присутствуют в плазме человека в низких концентрациях, в виде фрагментов, длина которых, как правило, напоминает мононуклеосомные единицы (YMD Lo et al. 2010 Sci Transl Med; 2: 61ra91; и YW Zheng et al. 2012 Clin Chem; 58: 549-558). Несмотря на указанные ограничения, система полногеномного бисульфитного секвенирования позволила провести анализ метилирования молекул ДНК плазмы. Согласно другим вариантам реализации, так как выбранные платформы для одномолекулярного секвенирования позволяют установить статус метилирования молекул ДНК непосредственно, без бисульфитной конверсии (BA Flusberg et al. 2010 Nat Способы; 7: 461-465; J Shim et al. 2013 Sci Rep; 3:1389. ЦИО: 10.1038/srep01389), применение таких платформ позволяет использовать неконвертированную бисульфитом ДНК плазмы для определения уровней метилирования ДНК плазмы или для определения метилома плазмы. Такие платформы позволяют детектировать N6-метиладенин, 5-метилцитозин и 5-гидроксиметилцитозин, что может обеспечить улучшенные результаты (например, улучшенную чувствительность или специфичность) в отношении разных биологических функций разных форм метилирования. Такие улучшенные результаты могут быть полезны при применении вариантов реализации для детекции или мониторинга конкретных расстройств, например, преэклампсии или ракового заболевания определенного типа.

Бисульфитное секвенирование позволяет также различать разные формы метилирования. Согласно одному варианту реализации можно включить дополнительные этапы, позволяющие различить 5-метилцитозин и 5-гидроксиметилцитозин. Один из таких подходов представлен окислительным бисульфитным секвенированием (oxBS-seq), которое может установить локализацию 5-метилцитозина и 5-гидроксиметилцитозина с разрешением, равным одному основанию (MJ Booth et al. 2012 Science; 336: 934-937; MJ Booth et al. 2013 Nature Protocols; 8: 1841-1851). При бисульфитном секвенировании и 5-метилцитозин, и 5-гидроксиметилцитозин считываются как цитозины и, соответственно, различить их невозможно. С другой стороны, при oxBS-seq специфическое окисление 5-гидроксиметилцитозина до 5-формилцитозина за счет обработки перрутеном калия (KRuO₄) с последующей конверсией новообразованного 5-формилцитозина в урацил с применением бисульфитной конверсии позволяет отличить 5-гидроксиметилцитозин от 5-

метилцитозина. Таким образом, регистрация 5-метилцитозина может быть проведена за один цикл oxBS-seq, а уровни 5-гидроксиметилцитозина выводятся путем сравнения с результатами бисульфитного секвенирования. Согласно другому варианту реализации 5-метилцитозин можно отличить от 5-гидроксиметилцитозина с применением бисульфитного секвенирования с использованием Tet (TAV-seq) (M Yu et al. 2012 Nat Protoc; 7: 2159–2170). TAV-seq позволяет идентифицировать 5-гидроксиметилцитозин с разрешением, равным одному нуклеотиду, а также определить его содержание в каждом участке модификации. Указанный способ задействует опосредованную β -глюкозилтрансферазой защиту 5-гидроксиметилцитозина (глюкозилирование) и опосредованное рекомбинантным Tet1 (mTet1) мыши окисление 5-метилцитозина до 5-карбоксилцитозина. После последующей обработки бисульфитом и ПЦР-амплификации и цитозин, и 5-карбоксилцитозин (полученный из 5-метилцитозина) превращаются в тимин (Т), при этом 5-гидроксиметилцитозин будет считываться как С.

На фиг. 1В представлена плотность метилирования на участках секвенированных образцов размером 1 Мб в соответствии с вариантами реализации настоящего изобретения. График 150 представляет собой Circos-график, отражающий плотность метилирования в ДНК материнской плазмы и геномной ДНК на участках размером 1 Мб по всему геному. От центра к краям: идеограммы хромосом могут быть ориентированы от короткого плеча к длинному плечу (pter-qter) по часовой стрелке (центромеры обозначены красным цветом), материнская кровь (красный), плацента (желтый), материнская плазма (зеленый), общие риды в материнской плазме (голубой), и специфичные для плода риды в материнской плазме (фиолетовый). Общие CpG уровни метилирования (т.е. уровни плотности) клеток материнской крови, плаценты и материнской плазмы приведены в таблице 100. Уровень метилирования клеток материнской крови в целом выше, чем уровень метилирования плаценты, по всему геному.

В. Сравнение бисульфитного секвенирования с другими методиками

Авторы исследовали плацентарный метилом с применением массивно-параллельного бисульфитного секвенирования. Кроме того, авторы исследовали плацентарный метилом с применением платформы на основе олигонуклеотидных матриц, покрывающей приблизительно 480 000 CpG-сайтов в геноме человека (Illumina) (M Kulis et al. 2012 Nat Genet; 44: 1236-1242; и C Clark et al. 2012 PLoS One; 7: e50233). Согласно одному варианту реализации с применением генотипирования на основе BeadChip и анализа метилирования, генотипирование выполняли с применением матрицы для генотипирования Illumina HumanOmni2.5-8 в соответствии с протоколом производителя. Генотипы прогнозировали с применением алгоритма программного обеспечения GenCall

Genome Studio Software (Illumina). Показатели прогнозирования превышали 99%. Для анализа метилирования на основе микроматриц геномную ДНК (500–800 нг) обрабатывали бисульфитом натрия с применением набора Zymo EZ DNA Methylation Kit (Zymo Research, Ориндж, Калифорния, США) в соответствии с рекомендациями производителя для анализа метилирования Illumina Infinium Methylation Assay.

Анализ метилирования осуществляли на 4 мкл конвертированной бисульфитом геномной ДНК в концентрации 50 нг/мкл в соответствии с протоколом анализа метилирования Infinium HD Methylation Assay. Гибридизированный чип (BeadChip) сканировали с помощью инструмента Illumina iScan. Данные метилирования ДНК исследовали с помощью программного обеспечения GenomeStudio (v2011.1) Methylation Module (v1.9.0), с нормализацией по внутренним контролям и вычитанием фона. Индекс метилирования для индивидуального CpG-сайта представлен в виде бета-значения (β), которое рассчитывали с использованием отношения интенсивностей флуоресценции метилированных и неметилированных аллелей:

$$\beta = \frac{\text{Плотность метилированного аллеля}}{\text{Плотность неметилированного аллеля} + \text{Плотность метилированного аллеля} + 100}$$

Для CpG-сайтов, представленных на матрице и секвенированных по меньшей мере с 10-кратной глубиной, авторы сравнивали бета-значение, полученное с помощью матрицы, с индексом метилирования, определенным посредством секвенирования того же сайта. Бета-значения выражают плотность метилированных зондов в виде отношения объединенной плотности метилированных и неметилированных зондов, покрывающих один и тот же CpG-сайт. Индекс метилирования для каждого CpG-сайта относится к отношению метилированных ридов к общему количеству ридов, покрывающих указанный CpG.

На фиг. 2А-2С приведены графики зависимости бета-значений, определенных с помощью матрицы BeadChip Illumina Infinium HumanMethylation 450K, от индексов метилирования, определенных с помощью полногеномного бисульфитного секвенирования, соответствующих CpG-сайтам, исследованных с помощью обеих платформ: (А) Клетки материнской крови, (В) Образец ворсин хориона, (С) Ткань зрелой плаценты. Данные обеих платформ хорошо согласовались, и коэффициенты корреляции Пирсона составляли 0,972, 0,939 и 0,954; значения R^2 составляли 0,945, 0,882 и 0,910 для клеток материнской крови, CVS и ткани зрелой плаценты, соответственно.

Авторы также сравнили собственные данных секвенирования с данными, описанными Chu с коллегами, исследовавших профили метилирования 12 пар образцов ДНК CVS и ДНК клеток материнской крови с применением олигонуклеотидной матрицы, захватывающей приблизительно 27 000 CpG-сайтов (T Chu et al. 2011 PLoS One; 6:

e14723). Данные относительно корреляции результатов секвенирования ДНК CVS и клеток материнской крови и каждого из 12 пар образцов в проведенном ранее исследовании давали средний коэффициент Пирсона (0,967) и R^2 (0,935) для материнской крови, и средний коэффициент Пирсона (0,943) и R^2 (0,888) для CVS. Полученные авторами настоящего изобретения данные относительно CpG-сайтов, представленных на обеих матрицах, хорошо коррелировали с опубликованными данными. Уровни метилирования сайтов, не являющихся CpG-сайтами, составляли < 1% для клеток материнской крови, CVS и плацентарных тканей (таблица 100). Указанные результаты согласуются с современными представлениями о том, что существенные уровни метилирования сайтов, не являющихся CpG-сайтами, в основном ограничены плюрипотентными клетками (R Lister et al. 2009 Nature; 462: 315-322; L Laurent et al. 2010 Genome Res; 20: 320-331).

С. Сравнение метиломов плазмы и крови небеременных субъектов

На фиг. 3А и 3В приведены гистограммы процентного содержания метилированных CpG-сайтов в плазме и клетках крови, полученных от взрослого мужчины и небеременной взрослой женщины: (А) Аутосомы, (В) X-хромосома. Графики демонстрируют сходство метиломов плазмы и крови мужчины и небеременной женщины. Общие отношения метилированных CpG-сайтов в образцах плазмы мужчины и небеременной женщины были почти такими же, как в соответствующей ДНК клеток крови (таблица 100 и фиг. 2А и 2В).

Затем авторы настоящего изобретения исследовали корреляцию профилей метилирования образцов плазмы и клеток крови локус-специфическим образом. Авторы определили плотность метилирования каждого участка размером 100 кб в геноме человека путем определения общего числа неконвертированных цитозинов в CpG-сайтах как доли всех CpG-сайтов, захватываемых ридями последовательностей, картирующимися на область 100 кб. Плотности метилирования хорошо согласовались в образцах плазмы и соответствующих образцах ДНК клеток крови и мужчины, и женщины.

На фиг. 4А и 4В приведены графики плотностей метилирования соответствующих локусов в ДНК клеток крови и ДНК плазмы: (А) Небеременная взрослая женщина, (В) Взрослый мужчина. Коэффициент корреляции Пирсона и значение R^2 для образцов от небеременной женщины составляли, соответственно, 0,963 и 0,927, а для образцов от мужчины, соответственно, 0,953 и 0,908. Эти данные согласуются с полученными ранее результатами, основанными на оценке генотипов молекул ДНК плазмы реципиентов трансплантированных аллогенных гемопоэтических стволовых клеток, которые показали, что гемопоэтические клетки являются преобладающим источником ДНК в плазме

человека (YW Zheng et al. 2012 Clin Chem; 58: 549-558).

D. Уровни метилирования разных метиломов

Затем авторы настоящего изобретения исследовали уровни метилирования ДНК материнской плазмы, клеток материнской крови и плацентарной ткани для определения уровней метилирования. Указанные уровни определяли для областей повторов, областей без повторов, а также определяли общие уровни.

На фиг. 5А и 5В приведены гистограммы процентного содержания метилированных CpG-сайтов в образцах, полученных при беременности: (А). Аутосомы, (В) X-хромосома. Общее содержание метилированных CpG составляло 67,0% и 68,2% для образцов материнской плазмы в первом и третьем триместре, соответственно. В отличие от результатов, полученных от небеременных индивидуумов, указанное содержание было ниже, чем в образце клеток материнской крови в первом триместре, но выше, чем в образцах CVS и ткани зрелой плаценты (таблица 100). Примечательно, что процент метилированных CpG в образце материнской плазмы послеродового периода составлял 73,1%, что аналогично данным для клеток крови (таблица 100). Указанные тренды наблюдались в CpG, встречающихся во всех аутосомах, а также X-хромосоме, и захватывали как области без повторов, так и несколько классов повторяющихся элементов генома человека.

Было обнаружено, что и повторяющиеся, и неповторяющиеся элементы в плаценте гипометилированы относительно клеток материнской крови. Эти результаты согласовались с опубликованными данными, согласно которым плацента гипометилирована относительно других тканей, включая клетки периферической крови.

От 71% до 72% секвенированных CpG-сайтов были метилированы в ДНК клеток крови беременной женщины, небеременной женщины и взрослого мужчины (таблица 100 на фиг. 1). Эти данные сопоставимы с сообщением Y Li et al. (Y Li et al. 2010 PLoS Biol; 8: e1000533) о 68,4% CpG-сайтов мононуклеарных клеток крови. В соответствии с предыдущими сообщениями относительно гипометилированной природы плацентарных тканей, 55% и 59% CpG-сайтов были метилированы в CVS и ткани зрелой плаценты, соответственно (таблица 100).

На фиг. 6 приведена гистограмма уровня метилирования разных классов повторов в геноме человека, для материнской крови, плаценты и материнской плазмы. Указанные классы повторов соответствуют определенным браузерам геномов UCSC. Приведены данные для образцов первого триместра. В отличие от ранее полученных данных, предполагающих, что гипометилированная природа плацентарных тканей наблюдалась главным образом в определенных классах повторов в геноме (B Novakovic et al. 2012

Placenta; 33: 959-970), в настоящем описании авторы изобретения демонстрируют, что плацента на самом деле гипометилирована по большинству классов геномных элементов относительно клеток крови.

Е. Сходство метиломов

Варианты реализации позволяют определять метиломы плацентарных тканей, клеток крови и плазмы с применением одной и той же платформы. Соответственно, можно было проводить прямые сравнения метиломов указанных типов биологических образцов. Высокий уровень сходства метиломов клеток крови и плазмы мужчины и небеременной женщины, а также клеток материнской крови и образца материнской плазмы в послеродовой период дополнительно подтверждают, что гемопоэтические клетки являются основным источником ДНК в плазме человека (YW Zheng et al. 2012 Clin Chem; 58: 549-558).

Сходство очевидно с точки зрения как общей доли метилированных CpG в геноме, так и высокой степени корреляции плотностей метилирования соответствующих локусов в ДНК клеток крови и ДНК плазмы. Однако при этом общая доля метилированных CpG в образцах материнской плазмы в первом триместре и в третьем триместре было понижено по сравнению с данными для клеток материнской крови или образца материнской плазмы в послеродовой период. Пониженные уровни метилирования во время беременности были обусловлены гипометилированной природой молекул плодной ДНК, присутствующих в материнской плазме.

Обращение профиля метилирования в образце материнской плазмы в послеродовой период с приобретением большего сходства с профилем клеток материнской крови предполагает, что молекулы плодной ДНК были устранены из материнского кровотока. Расчет концентраций плодной ДНК на основе маркеров ОНП плода действительно показал, что указанная концентрация изменилась от 33,9% в дородовом образце до всего 4,5% в образце в послеродовой период.

Г. Другие применения

Согласно вариантам реализации были успешно собраны метиломы ДНК с помощью МПС-анализа ДНК плазмы. Возможность определения плацентарного или плодного метилома по материнской плазме обеспечивает неинвазивный способ определения, детекции и мониторинга aberrантных профилей метилирования, ассоциированных со связанными с беременностью состояниями, такими как преэклампсия, задержка внутриутробного развития, преждевременные роды и др. Например, детекция специфической для заболевания сигнатуры aberrантного метилирования позволяет проводить скрининг, диагностику и мониторинг таких связанных с беременностью

состояний. Измерение уровня метилирования материнской плазмы позволяет проводить скрининг, диагностику и мониторинг таких связанных с беременностью состояний. Помимо непосредственного применения для исследования связанных с беременностью состояний, указанный способ может применяться в других областях медицины, для которых анализ ДНК плазмы представляет интерес. Например, метиломы раковых заболеваний могут быть определены по ДНК плазмы страдающих раковым заболеванием пациентов. Анализ метиломов раковых заболеваний по плазме, согласно описанию в настоящем документе, представляет собой потенциальную синергическую технологию анализа ракового генома по плазме (КСА Chan et al. 2013 Clin Chem; 59: 211-224 и RJ Leary et al. 2012 Sci Transl Med; 4:162ra154).

Например, определение уровня метилирования образца плазмы может использоваться для скрининга раковых заболеваний. Если уровень метилирования образца плазмы демонстрирует aberrантные уровни по сравнению со здоровыми контролями, можно предположить наличие рака. В таком случае может быть получено дополнительное подтверждение и проведена оценка типа ракового заболевания или ткани происхождения ракового заболевания путем определения профиля метилирования плазмы в разных геномных локусах или путем геномного анализа плазмы для детекции ассоциированных с опухолью aberrаций числа копий, хромосомных транслокаций и однонуклеотидных вариаций. Так, согласно одному варианту реализации настоящего изобретения, определение профиля ракового метилома и генома в плазме может проводиться одновременно. Как вариант, могут использоваться радиологические исследования и визуализация (например, компьютерная томография, магнитно-резонансная визуализация, позитронно-эмиссионная томография) или эндоскопия (например, эндоскопия верхних отделов желудочно-кишечного тракта или колоноскопия) для дополнительного обследования индивидуумов, у которых на основании анализа уровня метилирования в плазме предполагается наличие ракового заболевания.

При скрининге или детекции ракового заболевания определение уровня метилирования образца плазмы (или другого биологического материала) может применяться в сочетании с другими методами скрининга или детекции ракового заболевания, такими как измерение простатического специфического антигена (например, для рака предстательной железы), карциноэмбрионального антигена (например, для колоректальной карциномы, карциномы желудка, карциномы поджелудочной железы, карциномы легкого, карциномы молочной железы, медуллярной карциномы щитовидной железы), альфа-фетопротеина (например, для рака печени или герминативно-клеточных опухолей), СА125 (например, для рака яичника и рака молочной железы) и СА19-9

(например, для карциномы поджелудочной железы).

Кроме того, могут быть секвенированы другие ткани для получения клеточного метилома. Например, может быть проанализирована ткань печени для определения паттерна метилирования, специфического для печени, который может быть использован для идентификации патологий печени. Другие ткани, которые также могут быть проанализированы, включают клетки головного мозга, кости, легкие, сердце, мышцы и почки и т.д. Профили метилирования различных тканей могут время от времени изменяться, например, в результате развития, старения, болезненных процессов (например, воспаления или цирроза, или аутоиммунных процессов (таких как при системной красной волчанке)) или лечения (например, лечения деметилирующими агентами, такими как 5-азациитидин и 5-азадезоксцитидин). Динамическая природа метилирования ДНК делает такой анализ потенциально очень полезным для мониторинга физиологических и патологических процессов. Например, если обнаруживается изменение метилома плазмы индивидуума относительно исходного значения, полученного, когда указанный индивидуум был здоров, можно детектировать болезненные процессы в органах, которые вносят вклад в ДНК плазмы.

Также могут быть определены метиломы трансплантированных органов по ДНК плазмы реципиентов трансплантированных органов. Метиломный анализ трансплантатов по плазме согласно описанию настоящего изобретения потенциально представляет собой синергическую технологию геномного анализа трансплантатов по плазме (YW Zheng et al, 2012; YMD Lo et al. 1998 Lancet; 351: 1329-1330; и TM Snyder et al. 2011 Proc Natl Acad Sci USA; 108: 6229-6234). Поскольку ДНК плазмы как правило, рассматривается в качестве маркера клеточной смерти, повышение в плазме уровня ДНК, высвобождаемой из трансплантированного органа, может использоваться в качестве маркера повышения клеточной смерти в указанном органе, например, криза отторжения или других патологических процессов с участием указанного органа (например, инфекции или абсцесса). В случае успешного проведения терапии, направленной против отторжения, уровень высвобождаемой трансплантированным органом ДНК в плазме предположительно будет снижаться.

III. Определение метилома плода или опухоли с использованием ОНП

Согласно приведенному выше описанию метилом плазмы соответствует метилому крови у небеременного здорового индивидуума. Однако у беременной женщины указанные метиломы различаются. Молекулы плодной ДНК циркулируют в материнской плазме наряду с составляющим фон большинством молекул материнской ДНК (YMD Lo et al. 1998 Am J Hum Genet; 62: 768-775). Таким образом, у беременной женщины метилом

плазмы в значительной степени представляет собой смесь плацентарного метилома и метилома крови. Соответственно, можно получить плацентарный метилом, используя плазму.

Согласно одному варианту реализации различия однонуклеотидных полиморфизмов (ОНП) между матерью и плодом используют для идентификации молекул плодной ДНК в материнской плазме. Цель заключалась в идентификации ОНП-локусов, по которым мать гомозиготна, а плод гетерозиготен; специфичный для плода аллель может применяться для определения того, какие фрагменты ДНК принадлежат плоду. Геномную ДНК из клеток материнской крови исследовали с применением матрицы для генотипирования ОНП Illumina HumanOmni2.5-8. С другой стороны, для ОНП-локусов, по которым мать гетерозиготна, а плод гомозиготен, аллель ОНП, специфичный для матери, может применяться для определения того, какие фрагменты ДНК плазмы принадлежат матери. Уровень метилирования таких фрагментов ДНК будет отражать уровень метилирования соответствующих геномных областей у матери.

А. Корреляция метилирования специфичных для плода ридов и плацентарного метилома

Локусы, в которых содержится два разных аллеля, причем количество одного аллеля (В) значительно меньше количества другого аллеля (А), идентифицировали по результатам секвенирования биологического образца. Риды, покрывающие аллели В, считали специфичными для плода (специфичные для плода риды). Мать определяли как гомозиготную по А, а плод как гетерозиготный по А/В, и, соответственно, риды, покрывающие аллель А, были общими для матери и плода (общие риды).

В одном проанализированном случае беременности, который использовали для иллюстрации нескольких концепций настоящего изобретения, было обнаружено, что беременная мать гомозиготна по 1945516 локусам на аутосомах. Изучали риды секвенирования ДНК материнской плазмы, покрывающие указанные ОНП. Риды, несущие не являющийся материнским аллель, были обнаружены в 107 750 локусах, и указанные локусы считали информативными. В каждом информативном ОНП аллель, происходящий не от матери, называли специфичным для плода аллелем, а другой называли общим аллелем.

Может быть определена фракционная концентрация плодной/опухолевого ДНК (также называемая процентом плодной ДНК) в материнской плазме. Согласно одному варианту реализации фракционную концентрацию плодной ДНК в материнской плазме, f , определяют из уравнения:

$$f = \frac{2p}{p+q}$$

где p представляет собой число ридов секвенирования со специфичным для плода аллелем, и q представляет собой число ридов секвенирования с общим для матери и плода аллелем (YMD Lo et al. 2010 Sci Transl Med; 2: 61ra91). Доля плодной ДНК в образцах материнской плазмы первом триместре, третьем триместре и в послеродовой период, как было установлено, составляет 14,4%, 33,9% и 4,5%, соответственно. Долю плодной ДНК также рассчитывали с применением числа ридов, выравниваемых по Y-хромосоме. На основе данных Y-хромосомы эти результаты составили 14,2%, 34,9% и 3,7%, соответственно, в образцах материнской плазмы в первом триместре, в третьем триместре и в послеродовой период.

Варианты реализации, где используется отдельный анализ специфичных для плода или общих ридов последовательностей, демонстрируют, что циркулирующие молекулы плодной ДНК в значительно большей степени гипометилированы, чем фоновые молекулы ДНК. Сравнения плотностей метилирования соответствующих локусов в специфичных для плода ридов материнской плазмы и данных для плацентарной ткани, как для первого, так и для третьего триместра, показали высокие уровни корреляции. Эти данные обеспечивают на геномном уровне подтверждение того, что плацента является основным источником происходящих от плода молекул ДНК в материнской плазме, и представляют собой важный шаг вперед по сравнению с имеющимися данными, основанными на информации, полученной для отдельных локусов.

Авторы определяли плотность метилирования каждой области объемом 1 Мб в геноме с использованием либо специфичных для плода, либо общих ридов, захватывающих CpG-сайты, смежные с информативными ОНП. Плодный и неспецифичный для плода метиломы, собранные из ридов последовательностей материнской плазмы, могут быть представлены, например, на Circos-графике (M Krzywinski et al. 2009 Genome Res; 19: 1639-1645). Плотности метилирования в расчете на участок размером 1 Мб также определяли для образцов клеток материнской крови и плацентарной ткани.

На фиг. 7А приведен Circos-график 700 для образцов, полученных в первом триместре. На фиг. 7В приведен Circos-график 750 для образцов, полученных в третьем триместре. Графики 700 и 750 отображают плотность метилирования в расчете на участок 1 Мб. Идеограммы хромосом (внешний круг) ориентированы от короткого плеча к длинному плечу по часовой стрелке (центромеры выделены красным цветом). Второй снаружи круг представляет число CpG-сайтов в соответствующих областях размером 1

Мб. Для красных столбцов приведена шкала до 20 000 сайтов на участок 1 Мб. Плотности метилирования соответствующих областей размером 1 Мб показаны на других кругах с использованием цветовой схемы, приведенной в центре.

Для образцов первого триместра (фиг. 7А), от центра к краям, на кругах представлены: образец ворсин хориона, специфичные для плода риды в материнской плазме, специфичные для матери риды в материнской плазме, комбинация плодных и неплодных ридов в материнской плазме, и клетки материнской крови. Для образцов третьего триместра (фиг. 7В) на кругах представлены: ткань зрелой плаценты, специфичные для плода риды в материнской плазме, специфичные для матери риды в материнской плазме, комбинация плодных и не относящихся к плоду ридов в материнской плазме, в материнской плазме в послеродовой период и клетках материнской крови (из образца крови первого триместра). Видно, что в образцах плазмы и в первом, и в третьем триместре плодные метиломы более гипометилированы, чем в неспецифичных для плода метиломах.

Общий профиль метилирования плодных метиломов больше напоминал профиль образцов CVS или плацентарной ткани. С другой стороны, профиль метилирования ДНК для общих ридов в плазме, преимущественно представляющих материнскую ДНК, больше напоминал профиль клеток материнской крови. Затем авторы настоящего изобретения провели систематическое сравнение плотностей метилирования локусов одного за другим, ридов ДНК материнской плазмы и материнских или плодных тканей. Авторы определили плотности метилирования CpG-сайтов, которые присутствовали на том же риде последовательности, что и информативные ОНП, и которые были покрыты по меньшей мере 5 ридами последовательностей ДНК материнской плазмы.

На фиг. 8А–8D приведены графики сравнения плотностей метилирования геномной тканевой ДНК и ДНК материнской плазмы для CpG-сайтов, окружающих информативные однонуклеотидные полиморфизмы. На фиг. 8А показаны плотности метилирования для специфичных для плода ридов в образце материнской плазмы в первом триместре относительно плотностей метилирования для ридов в образце CVS. Как можно видеть, специфичные для плода значения хорошо согласуются со значениями для CVS.

На фиг. 8В показаны плотности метилирования для специфичных для плода ридов в образце материнской плазмы в третьем триместре относительно плотностей метилирования для ридов в ткани зрелой плаценты. Наборы плотностей также хорошо согласуются, указывая на то, что плодный профиль метилирования может быть получен путем анализа ридов со специфичными для плода аллелями.

На фиг. 8С показаны плотности метилирования для общих ридов в образце материнской плазмы в первом триместре относительно плотностей метилирования для ридов в клетках материнской крови. С учетом того, что большинство общих ридов происходят от матери, указанные два набора значений хорошо согласуются. На фиг. 8D показаны плотности метилирования для общих ридов в образце материнской плазмы в третьем триместре относительно плотностей метилирования для ридов в клетках материнской крови.

Для специфичных для плода ридов в материнской плазме коэффициент корреляции Спирмена между материнской плазмой в первом триместре и CVS составлял 0,705 ($P < 2,2 \cdot 10^{-16}$); между материнской плазмой в третьем триместре и тканью зрелой плаценты указанный коэффициент составлял 0,796 ($P < 2,2 \cdot 10^{-16}$) (фиг. 8А и 8В). Аналогичное сравнение выполняли для общих ридов в материнской плазме с данными клеток материнской крови. Коэффициент корреляции Пирсона составлял 0,653 ($P < 2,2 \cdot 10^{-16}$) для образца плазмы в первом триместре, и 0,638 ($P < 2,2 \cdot 10^{-16}$) для образца плазмы в третьем триместре (фиг. 8С и 8D).

В. Плодный метилом

Согласно одному варианту реализации для сборки плодного метилома по материнской плазме авторы настоящего изобретения отбирали риды последовательностей, захватывающие по меньшей мере один информативный плодный ОНП-сайт и содержавшие по меньшей мере один CpG-сайт в составе того же рида. Риды, где обнаруживались специфичные для плода аллели, включали в сборку плодного метилома. Риды, где обнаруживался общий аллель, т.е. аллель, не являющийся специфичным для плода, включали в сборку неспецифичного для плода метилома, который преимущественно состоит из происходящих от матери молекул ДНК.

Специфичные для плода риды покрывали 218 010 CpG-сайтов на аутосомах в образцах материнской плазмы в первом триместре. Соответствующие значения для образцов материнской плазмы в третьем триместре и в послеродовой период составляли 263 611 и 74 020, соответственно. В среднем, общие риды покрывали указанные CpG-сайты со средней кратностью 33,3, 21,7 и 26,3, соответственно. Специфичные для плода риды покрывали указанные CpG-сайты с кратностью 3,0, 4,4 и 1,8, соответственно, для образцов материнской плазмы первого триместра, третьего триместра и послеродового периода.

Плодная ДНК представляет собой минорную популяцию в материнской плазме и, соответственно, покрытие указанных CpG-сайтов специфичными для плода ридами было пропорционально проценту плодной ДНК в образце. Для образца материнской плазмы в

первом триместре общий процент метилированных CpG в ридов плода составлял 47,0%, тогда как для общих ридов он составлял 68,1%. Для образца материнской плазмы в третьем триместре процент метилированных CpG в ридов плода составлял 53,3%, тогда как для общих ридов он составлял 68,8%. Эти данные показали, что специфичные для плода риды в материнской плазме были более гипометилированы, чем общие риды в материнской плазме.

С. Способ

Описанные выше методики также могут применяться для определения опухолевого профиля метилирования. Ниже описаны способы определения плодных и опухолевых профилей метилирования.

Фиг. 9 представляет собой блок-схему, иллюстрирующую способ 900 определения первого профиля метилирования по биологическому образцу организма в соответствии с вариантами реализации настоящего изобретения. Способ 900 позволяет сконструировать эпигенетическую карту плода по профилю метилирования материнской плазмы. Биологический образец включает свободную от клеток ДНК, содержащую смесь свободной от клеток ДНК, происходящей из первой ткани и из второй ткани. Например, указанная первая ткань может происходить от плода, из опухоли или трансплантированного органа.

На этапе 910 анализируют множество молекул ДНК из указанного биологического образца. Анализ молекулы ДНК может включать определение локализации указанной молекулы ДНК в геноме указанного организма, определение генотипа молекулы ДНК, и определение того, метилирована ли указанная молекула ДНК в одном или большем числе сайтов.

Согласно одному варианту реализации молекулы ДНК анализируют с использованием ридов последовательностей молекул ДНК, при этом секвенирование чувствительно к метилированию. Соответственно, указанные риды последовательностей включают статус метилирования молекул ДНК из указанного биологического образца. Статус метилирования может включать информацию о том, является ли конкретный остаток цитозина 5-метилцитозином или 5-гидроксиметилцитозином. Указанные риды последовательностей могут быть получены с помощью различных методик секвенирования, методик ПЦР, матриц и других подходящих методик идентификации последовательностей фрагментов. Статус метилирования сайтов ридов последовательностей может быть определен согласно описанию в настоящем документе.

На этапе 920 идентифицируют множество первых локусов, отличающихся тем, что первый геном первой ткани гетерозиготен по соответствующему первому аллелю, и

соответствующий второй аллель и второй геном второй ткани гомозиготен по соответствующему первому аллелю. Например, специфичные для плода риды могут быть идентифицированы во множестве первых локусов, или специфичные для опухоли риды могут быть идентифицированы во множестве первых локусов. Тканеспецифичные риды могут быть идентифицированы по ридам при секвенировании, где процент ридов последовательностей второго аллеля попадает в определенный диапазон, например, приблизительно 3–25%, таким образом указывая на минорную популяцию фрагмента ДНК из гетерозиготного генома в указанном локусе и мажорную популяцию из гомозиготного генома в указанном локусе.

На этапе 930 анализируют молекулы ДНК, локализованные в одном или большем числе сайтов каждого из первых локусов. Определяют число молекул ДНК, которые метилированы в сайте и соответствуют соответствующему второму аллелю указанного локуса. Может присутствовать более одного сайта на локус. Например, ОНП может указывать на то, что фрагмент специфичен для плода, и указанный фрагмент может содержать несколько сайтов, статус метилирования которых определяют. Может быть определено число метилированных ридов в каждом сайте и общее число метилированных ридов для локуса.

Локус может быть определен через конкретное число сайтов, конкретный набор сайтов, или конкретный размер области вокруг вариации, которая содержит специфичный для ткани аллель. Локус может содержать всего один сайт. Сайты могут обладать специфическими свойствами, например, быть CpG-сайтами. Определение числа неметилированных ридов эквивалентно определению статуса метилирования и включено в него.

На этапе 940 для каждого из первых локусов рассчитывают плотность метилирования на основе количеств молекул ДНК, метилированных в одном или большем числе сайтов указанного локуса и соответствующих соответствующему второму аллелю указанного локуса. Например, плотность метилирования может быть определена для CpG-сайтов, соответствующих локусу.

На этапе 950 определяют первый профиль метилирования первой ткани по плотностям метилирования для первых локусов. Указанный первый профиль метилирования может соответствовать конкретным сайтам, например, CpG-сайтам. Указанный профиль метилирования может относиться ко всем локусам, содержащим специфичный для плода аллель, или только к некоторым из указанных локусов.

IV. Использование различий метиломов плазмы и крови

Выше было показано, что специфичные для плода риды из плазмы коррелируют с

плацентарным метиломом. Поскольку материнский компонент метилома материнской плазмы в основном определяется клетками крови, различия между метиломом плазмы и метиломом крови могут применяться для определения плацентарного метилома для всех локусов, а не только локализации специфичных для плода аллелей. Различия между метиломом плазмы и метиломом крови также могут применяться для определения метилома опухоли.

А. Способ

Фиг. 10 представляет собой блок-схему, иллюстрирующую способ 1000 определения первого профиля метилирования по биологическому образцу организма в соответствии с вариантами реализации настоящего изобретения. Указанный биологический образец (например, плазма) включает свободную от клеток ДНК, содержащую смесь свободной от клеток ДНК, происходящей из первой ткани и из второй ткани. Указанный первый профиль метилирования соответствует профилю метилирования первой ткани (например, плодной ткани или опухолевой ткани). Способ 1200 может обеспечивать определение дифференциально метилированных областей по материнской плазме.

На этапе 1010 получают биологический образец. Указанный биологический образец может быть получен просто с помощью аппарата (например, секвенатора). Указанный биологический образец может быть представлен формой, полученной из организма, или обработанной формой, например, указанный образец может представлять собой плазму, экстрагированную из образца крови.

На этапе 1020 получают второй профиль метилирования, соответствующий ДНК второй ткани. Указанный второй профиль метилирования может быть считан из памяти, так как возможно он был определен ранее. Указанный второй профиль метилирования может быть определен по второй ткани, например, другого образца, который содержит исключительно или преимущественно клетки второй ткани. Указанный второй профиль метилирования может соответствовать клеточному профилю метилирования и быть получен из клеточной ДНК. Согласно другому примеру указанный второй профиль может быть определен по образцу плазмы, полученному до беременности или до развития ракового заболевания, так как метилом плазмы небеременного индивидуума, не страдающего раковым заболеванием, в значительной мере аналогичен метилому клеток крови.

Указанный второй профиль метилирования может обеспечивать плотность метилирования в каждом из множества локусов в геноме указанного организма. Плотность метилирования в конкретном локусе соответствует доле метилированной ДНК

второй ткани. Согласно одному варианту реализации плотность метилирования представляет собой плотность метилирования CpG, при этом CpG-сайты, связанные с локусом, используют для определения плотности метилирования. Если на один локус приходится один сайт, плотность метилирования может быть равна индексу метилирования. Плотность метилирования также соответствует плотности отсутствия метилирования, так как указанные два значения взаимно дополняют друг друга.

Согласно одному варианту реализации указанный второй профиль метилирования получают путем проведения чувствительного к метилированию секвенирования клеточной ДНК из образца организма. Один из примеров чувствительного к метилированию секвенирования включает обработку ДНК бисульфитом натрия и затем осуществление секвенирования ДНК. Согласно другому примеру чувствительное к метилированию секвенирование может осуществляться без применения бисульфита натрия, с применением платформы для мономолекулярного секвенирования, позволяющей устанавливать статус метилирования молекул ДНК (включая N6-метиладенин, 5-метилцитозин и 5-гидроксиметилцитозин) непосредственно, без бисульфитной конверсии (AB Flusberg et al. 2010 Nat Methods; 7: 461-465; J Shim et al. 2013 Sci Rep; 3:1389. ЦИО: 10.1038/srep01389); или посредством иммунопреципитации метилированного цитозина (например, с применением антитела против метилцитозина или с применением связывающего метилированную ДНК белка или пептида (LG Acevedo et al. 2011 Epigenomics; 3: 93-101) с последующим секвенированием; или путем применения чувствительных к метилированию рестрикционных ферментов с последующим секвенированием. Согласно другому варианту реализации используют методики без секвенирования, такие как матрицы, цифровая ПЦР и масс-спектрометрия.

Согласно другому варианту реализации вторая плотность метилирования второй ткани может быть получена ранее из контрольных образцов от субъекта или от других субъектов. Плотность метилирования от другого субъекта может работать в качестве референсного профиля метилирования, характеризующегося референсными плотностями метилирования. Референсные плотности метилирования могут быть определены по нескольким образцам, при этом среднее уровня (или другая статистическая величина) разных плотностей метилирования в локусе может использоваться в качестве референсной плотности метилирования в указанном локусе.

На этапе 1030 определяют профиль метилирования свободной от клеток ДНК по свободной от клеток ДНК из смеси. Указанный профиль метилирования свободной от клеток ДНК определяет плотность метилирования в каждом из множества локусов. Указанный профиль метилирования свободной от клеток ДНК может быть определен

путем получения ридов последовательностей при секвенировании свободной от клеток ДНК, при этом информацию о метилировании получают с ридами последовательностей. Указанный профиль метилирования свободной от клеток ДНК может быть определен тем же способом, что и клеточный метилом.

На этапе 1040 определяют процент свободной от клеток ДНК из первой ткани в биологическом образце. Согласно одному варианту реализации указанная первая ткань представляет собой плодную ткань, и соответствующая ДНК представляет собой плодную ДНК. Согласно другому варианту реализации указанная первая ткань представляет собой опухолевую ткань, и соответствующая ДНК представляет собой опухолевую ДНК. Указанный процент может быть определен различными способами, например, с использованием специфичного для плода аллеля или специфичного для опухоли аллеля. Число копий также может быть использовано для определения указанного процента, например, согласно описанию во включенной посредством ссылки заявке на патент США 13/801748, озаглавленной «Мутационный анализ ДНК плазмы для детекции раковых заболеваний» («Mutational Analysis Of Plasma DNA For Cancer Detection»), которая была подана 13 марта 2013 г.

На этапе 1050 идентифицируют множество локусов для определения первого метилома. Указанные локусы могут соответствовать каждому из локусов, используемых для определения профиля метилирования свободной от клеток ДНК и второго профиля метилирования. Таким образом, соответствовать может множество локусов. Возможно, может быть использовано большее количество локусов для определения профиля метилирования свободной от клеток ДНК и второго профиля метилирования.

Согласно некоторым вариантам реализации могут быть идентифицированы локусы, которые были гиперметилированы или гипометилированы во втором профиле метилирования, например, с применением клеток материнской крови. Для идентификации локусов, которые были гиперметилированы в клетках материнской крови, можно провести сканирование хромосомы с одного конца для обнаружения CpG-сайта с индексом метилирования $\geq X\%$ (например, 80%). Затем можно провести поиск следующего CpG-сайта в расположенной ниже (3') области (например, на расположенном ниже отрезке 200 п.о.). Если расположенный непосредственно ниже CpG-сайт также имеет индекс метилирования $\geq X\%$ (или \geq другой указанной величины), первый и второй CpG-сайты могут быть сгруппированы. Распределение по группам может продолжаться либо до тех пор, пока в составе следующей расположенной ниже области не останется CpG-сайтов; либо до тех пор, пока индекс метилирования расположенного непосредственно ниже CpG-сайта не составит $< X\%$. Область сгруппированных CpG-

сайтов может быть описана как гиперметилированная в клетках материнской крови, если указанная область содержит по меньшей мере пять непосредственно смежных гиперметилированных CpG-сайтов. Аналогичный анализ для поиска локусов, которые были гипометилированы в клетках материнской крови, может осуществляться для CpG-сайтов с индексами метилирования $\leq 20\%$. Плотности метилирования для второго профиля метилирования могут быть рассчитаны для предварительно отобранных локусов и использованы для установления первого профиля метилирования (например, плотности метилирования плацентарной ткани) соответствующих локусов, например, по данным бисульфитного секвенирования материнской плазмы.

На этапе 1060 определяют первый профиль метилирования первой ткани путем расчета дифференциального параметра, который включает разность между плотностью метилирования второго профиля метилирования и плотностью метилирования профиля метилирования свободной от клеток ДНК для каждого из множества локусов. Указанную разность выражают в процентах.

Согласно одному варианту реализации первую плотность метилирования локуса в первой (например, плацентарной) ткани (D) вывели с применением следующего уравнения:

$$D = mbc - \frac{(mbc - mp)}{f * CN} \quad (1)$$

где mbc обозначает плотность метилирования второго профиля метилирования в локусе (например, предварительно отобранном локусе, определенном по данным бисульфитного секвенирования в клетках материнской крови); mp обозначает плотность метилирования соответствующего локуса в материнской плазме по данным бисульфитного секвенирования; f отражает процент свободной от клеток ДНК из первой ткани (например, фракционную концентрацию плодной ДНК), и CN отражает число копий в указанном локусе (например, более высокое значение для амплификаций или более низкое число для делеций относительно нормы). Если амплификации или делеции в первой ткани отсутствуют, CN может быть равен 1. При трисомии (или дупликации области в опухоли или у плода) CN может быть равен 1,5 (поскольку происходит увеличение от 2 копий до 3 копий) и для моносомии может быть равен 0,5. При более высоких уровнях амплификации может происходить увеличение с шагом 0,5. В указанном примере D может соответствовать дифференциальному параметру.

На этапе 1070 преобразуют первую плотность метилирования для получения скорректированной первой плотности метилирования первой ткани. Указанное преобразование позволяет учесть фиксированные различия между дифференциальными

параметрами и фактическим профилем метилирования первой ткани. Например, указанные значения могут отличаться на фиксированную константу или на коэффициент. Указанное преобразование может быть линейным или нелинейным.

Согласно одному варианту реализации было обнаружено, что распределение выведенных значений, D , ниже, чем фактический уровень метилирования плацентарной ткани. Например, выведенные значения могут быть линейно преобразованы с применением данных для CpG-островков, представлявших собой геномные сегменты с преобладанием CpG-сайтов. Положения в геноме CpG-островков, используемых в указанном исследовании, получали из базы данных браузера геномов UCSC (версия NCBI 36/hg18) (PA Fujita et al. 2011 Nucleic Acids Res; 39: D876-882). Например, CpG-островок может быть определен как геномный сегмент с содержанием GC $\geq 50\%$, длиной в геноме >200 п.о. и отношением наблюдаемого/ожидаемого числа CpG $>0,6$ (M Gardiner-Garden et al 1987 J Mol Biol; 196: 261-282).

Согласно одному варианту реализации для получения уравнения линейного преобразования могут быть включены CpG-островки по меньшей мере с 4 CpG-сайтами и средней глубиной секвенирования ≥ 5 на CpG-сайт в секвенированных образцах. После определения линейных зависимостей между плотностями метилирования CpG-островков в CVS или зрелой плаценте, и выведенными значениями, D , использовали следующие уравнения для определения прогнозируемых значений:

$$\text{Прогнозируемые значения для первого триместра} = D \times 1,6 + 0,2$$

$$\text{Прогнозируемые значения для третьего триместра} = D \times 1,2 + 0,05$$

В. Пример с плодом

Как упоминалось выше, способ 1000 может применяться для выведения «ландшафта метилирования» плаценты по материнской плазме. Циркулирующая ДНК в плазме в основном происходит из гемопоэтических клеток. Тем не менее, присутствует неустановленная доля свободной от клеток ДНК, поступающей из других внутренних органов. Кроме того, происходящая из плаценты свободная от клеток ДНК составляет приблизительно 5–40% от всей ДНК в материнской плазме, со средним значением приблизительно 15%. Соответственно, можно сделать предположение, что уровень метилирования в материнской плазме эквивалентен сумме существующего фонового метилирования и вклада плаценты во время беременности, согласно приведенному выше описанию.

Уровень метилирования в материнской плазме, MP , может быть определен с применением следующего уравнения:

$$MP = BKG \times (1 - f) + PLN \times f$$

где BKG представляет фоновый уровень метилирования ДНК в плазме, полученной из клеток крови и внутренних органов, PLN представляет уровень метилирования плаценты и f представляет фракционную концентрацию плодной ДНК в материнской плазме.

Согласно одному варианту реализации уровень метилирования плаценты может теоретически быть выведен исходя из уравнения:

$$PLN = \frac{MP - BKG \times (1 - f)}{f} \quad (2)$$

Уравнения (1) и (2) эквивалентны, если CN равен 1, D равен PLN , и BKG равен mbc . Согласно другому варианту реализации фракционную концентрацию плодной ДНК можно считать принимающей указанное значение, или устанавливать на уровне указанного значения, например, в рамках предположения о присутствии минимальной f .

Уровень метилирования материнской крови принимали за фоновое метилирование материнской плазмы. Помимо локусов, которые были гиперметилированы или гипометилированы в клетках материнской крови, авторы настоящего изобретения дополнительно исследовали способ выведения, фокусируясь на определенных клинически релевантных областях, например, CpG-островках в геноме человека.

Среднее плотности метилирования для общего количества 27 458 CpG-островков (версия NCBI 36/hg18) на аутосомах и X-хромосоме получали из данных секвенирования материнской плазмы и плаценты. Отбирали только островки с покрытием ≥ 10 CpG-сайтов и средней глубиной секвенирования ≥ 5 в расчете на покрытые CpG-сайты во всех анализируемых образцах, включая плаценту, материнскую кровь и материнскую плазму. В результате оставляли 26 698 CpG-островков (97,2%) как удовлетворяющих условию, и для них выводили уровень метилирования с применением данных о метилировании плазмы и фракционную концентрацию плодной ДНК в соответствии с приведенным выше уравнением.

Было отмечено, что распределение выведенных значений PLN было ниже фактического уровня метилирования CpG-островков в плацентарной ткани. Соответственно, согласно одному варианту реализации выведенные значения PLN , или просто выведенные значения (D) использовали в качестве условной единицы оценки уровня метилирования CpG-островков в плаценте. После преобразования выведенные значения линеаризовали, и их распределение начинало в большей степени напоминать фактический набор данных. Преобразованные выведенные значения были названы прогностическими значениями метилирования (MPV) и впоследствии использовались для прогнозирования уровня метилирования генетических локусов в плаценте.

В указанном примере CpG-островки классифицировали на 3 категории на основе их плотностей метилирования в плаценте: Низкая ($\leq 0,4$), Средняя ($>0,4 - <0,8$) и Высокая ($\geq 0,8$). Используя уравнение для выведения, авторы настоящего изобретения рассчитывали MPV для одного и того же набора CpG-островков и затем использовали указанные значения для классификации островков на 3 категории с одними и теми же точками отсечения. Сравнив фактические и выведенные наборы данных, авторы настоящего изобретения обнаружили, что 75,1% предварительно отобранных CpG-островков могут быть корректно сопоставлены по MPV с теми же категориями данных для тканей. Приблизительно 22% CpG-островков были распределены в одну группу при различии на один уровень (высокая/средняя, или средняя/низкая плотность) и менее чем 3% предположительно были полностью неправильно классифицированы (высокая/низкая плотность) (фиг. 12А). Также определяли общую эффективность классификации: 86,1%, 31,4% и 68,8% CpG-островков с плотностями метилирования $\leq 0,4$, $>0,4 - <0,8$ и $\geq 0,8$ в плаценте были корректно соотнесены имеющими «низкую», «среднюю» и «высокую» плотность метилирования (фиг. 12В).

На фиг. 11А и 11В приведены графические представления эффективности алгоритма прогнозирования с применением данных материнской плазмы и фракционной концентрации плодной ДНК в соответствии с вариантами реализации настоящего изобретения. На фиг. 11А приведено графическое представление 1100, демонстрирующее точность классификации CpG-островков с применением классификации с коррекцией по MPV (выведенная категория точно совпадает с фактическим набором данных); различие на 1 уровень (выведенная категория отличается на 1 уровень от фактического набора данных); и ошибочная классификация (выведенная категория противоположна фактическому набору данных). Фиг. 11В представляет собой графическое представление 1150, отражающее долю классифицируемых CpG-островков в каждой из выведенных категорий.

При условии, что материнское фоновое метилирование в соответствующих геномных областях незначительно, присутствие гиперметилированной происходящей из плаценты ДНК в кровотоке повышает общий уровень метилирования плазмы до степени, которая зависит от фракционной концентрации плодной ДНК. Выраженное изменение может наблюдаться в том случае, когда высвободившаяся плодная ДНК полностью метилирована. И напротив, если материнское фоновое метилирование значительно, степень изменения уровня метилирования плазмы станет более значительной при высвобождении гипометилированной плодной ДНК. Таким образом, применение указанного алгоритма выведения может быть более целесообразным тогда, когда уровень

метилирования выводят для генетических локусов, которые, по имеющимся сведениям, различаются для материнского фона и плаценты, в частности, для гиперметилированных и гипометилированных маркеров плаценты.

Фиг. 12А представляет собой таблицу 1200, где приведены подробные данные для 15 выбранных геномных локусов для прогнозирования метилирования в соответствии с вариантами реализации настоящего изобретения. Для подтверждения методик авторы настоящего изобретения выбрали 15 в различной степени метилированных геномных локусов, которые были исследованы ранее. Уровни метилирования выбранных областей выводили и сравнивали с ранее исследованными 15-ю метилированными в различной степени генетическими локусами (RWK Chiu et al. 2007 Am J Pathol; 170: 941-950; S.S.C. Chim et al. 2008 Clin Chem; 54: 500-511; SSC Chim et al. 2005 Proc Natl Acad Sci U S A; 102: 14753-14758; DWY Tsui et al. 2010 PLoS One; 5: e15069).

Фиг. 12В представляет собой графическое представление 1250 выведенных категорий 15 выбранных геномных локусов и соответствующих им уровней метилирования в плаценте. Были выведены следующие категории метилирования: низкое, $\leq 0,4$; среднее, $>0,4 < 0,8$; высокое, $\geq 0,8$. Таблица 1200 и графическое представление 1300 демонстрируют, что уровни метилирования в плаценте могут быть корректно выведены для них с несколькими исключениями: RASSF1A, CGI009, CGI137 и VAPA. Из указанных 4 маркеров только CGI009 показал заметное расхождение с фактическим набором данных. Остальные были в минимальной степени неправильно классифицированы.

В таблице 1200 «1» относится к выведенным значениям (D), рассчитываемым по уравнению:

$$D = \frac{MP - BKG \times (1 - f)}{f}$$

где f представляет собой фракционную концентрацию плодной ДНК. Отметка «2» относится к прогностическим значениям метилирования (MPV), относящимся к линейаризованным выведенным значениям с применением следующего уравнения: $MPV = D \times 1,6 + 0,25$. Отметка «3» относится к точкам отсечения классификации для выведенных значений: Низкая, $\leq 0,4$; Средняя, $>0,4 < 0,8$; Высокая, $\geq 0,8$. Отметка «4» относится к точкам отсечения классификации для фактического плацентарного набора данных: Низкая, $\leq 0,4$; Средняя, $>0,4 < 0,8$; Высокая, $\geq 0,8$. Отметка «5» означает, что плацентарный статус относится к статусу метилирования плаценты относительно статуса метилирования клеток материнской крови.

С. Расчет фракционных концентраций плодной ДНК

Согласно одному варианту реализации для определения процента плодной ДНК из первой ткани можно использовать Y-хромосому для плода мужского пола. Доля последовательностей Y-хромосомы (% chrY) в образце материнской плазмы была

представлена смесью ридов Y-хромосомы, полученных от плода мужского пола, и ряда материнских (женских) ридов, которые ошибочно выравнивали по Y-хромосоме (RWK Chiu et al. 2011 BMJ; 342: c7401). Соответственно, зависимость между %chrY и фракционной концентрацией плодной ДНК (f) в образце может быть задана как:

$$\%chrY = \%chrY_{male} \times f + \%chrY_{female} \times (1 - f)$$

где %chrY_{male} относится к соотношению ридов, выравниваемых по Y-хромосоме, в образце плазмы, содержащем 100% мужской ДНК; и %chrY_{female} относится к соотношению ридов, выравниваемых по Y-хромосоме в образце плазмы, содержащем 100% женской ДНК.

Значение %chrY может быть определено по ридам, которые выравнивались по Y-хромосоме без несовпадений, из образца, полученного от женщины, беременной плодом мужского пола, например, по ридам, полученным из конвертированных бисульфитом образцов. Значение %chrY_{male} может быть получено с помощью бисульфитного секвенирования двух образцов плазмы взрослого мужчины. Значение %chrY_{female} может быть получен с помощью бисульфитного секвенирования двух образцов плазмы небеременной взрослой женщины.

Согласно другим вариантам реализации процент плодной ДНК может быть определен по специфичным для плода аллелям на аутосоме. В другом примере для определения процента плодной ДНК могут использоваться эпигенетические маркеры. Могут также использоваться другие способы определения процента плодной ДНК.

D. Способ применения метилирования для определения числа копий

Плацентарный геном более гипометилирован, чем материнский геном. Как обсуждалось выше, метилирование плазмы беременной женщины зависит от фракционной концентрации происходящей из плаценты плодной ДНК в материнской плазме. Таким образом, посредством анализа плотности метилирования хромосомной области можно детектировать различие вклада плодных тканей в материнскую плазму. Например, у беременной женщины, вынашивающей трисомный плод (например, страдающий трисомией 21, или трисомией 18, или трисомией 13), указанный плод будет вносить дополнительное количество ДНК из трисомической хромосомы в материнскую плазму по сравнению с дисомическими хромосомами. В указанной ситуации плотность метилирования плазмы для трисомической хромосомы (или любой хромосомной области, содержащей амплификацию) будет ниже, чем плотности метилирования дисомических хромосом. Степень различия можно прогнозировать путем математического расчета, с учетом фракционной концентрации плодной ДНК в образце плазмы. Чем выше фракционная концентрация плодной ДНК в образце плазмы, тем значительнее будет

различие плотности метилирования между трисомическими и дисомическими хромосомами. Для областей, содержащих делецию, плотность метилирования будет выше.

Одним из примеров делеции является синдром Тернера, при котором плод женского пола имеет только одну копию X-хромосомы. В указанной ситуации у беременной женщины, вынашивающей плод, страдающий синдромом Тернера, плотность метилирования X-хромосомы в ДНК плазмы будет выше, чем в ситуации, когда та же беременная женщина вынашивает плод женского пола с нормальным числом X-хромосом. Согласно одному варианту реализации указанной стратегии сначала может быть проанализирована материнская плазма на присутствие или отсутствие последовательностей Y-хромосомы (например, с применением МПС или методики на основе ПЦР). Если последовательности Y-хромосомы присутствуют, плод может быть классифицирован как плод мужского пола, и в последующем анализе не будет необходимости. С другой стороны, если последовательности Y-хромосомы в материнской плазме отсутствуют, плод может быть классифицирован как плод женского пола. В указанной ситуации может быть проведен анализ плотности метилирования X-хромосомы в материнской плазме. Более высокая плотность метилирования X-хромосомы по сравнению с нормальной будет указывать на высокий риск наличия синдрома Тернера у плода. Указанный способ также может применяться для других анеуплоидий половых хромосом. Например, у плода XYY плотность метилирования Y-хромосомы в материнской плазме будет ниже, чем у нормального плода XY при аналогичном уровне плодной ДНК в материнской плазме. В другом примере у плода, страдающего синдромом Клайнфельтера (XXY), в материнской плазме присутствуют последовательности Y-хромосомы, однако плотность метилирования X-хромосомы в материнской плазме ниже, чем для нормального плода XY при аналогичном уровне плодной ДНК в материнской плазме.

Исходя из вышесказанного, плотность метилирования в плазме для дисомической хромосомы ($MP_{Non-aneu}$) может быть рассчитана как: $MP_{Non-aneu} = BKG \times (1 - f) + PLN \times f$, где BKG представляет собой фоновый уровень метилирования ДНК в плазме, полученный на клетках крови и внутренних органов, PLN представляет собой уровень метилирования плаценты и f представляет собой фракционную концентрацию плодной ДНК в материнской плазме.

Плотность метилирования в плазме для трисомической хромосомы (MP_{Aneu}) может быть рассчитана как: $MP_{Aneu} = BKG \times (1 - f) + PLN \times f \times 1,5$, где 1,5 соответствует числу копий CN и добавление еще одной хромосомы дает увеличение на 50%. Различие между трисомическими и дисомическими хромосомами (MP_{Diff}) будет составлять

$$MP_{Diff} = PLN \times f \times 0,5$$

Согласно одному варианту реализации сравнение плотности метилирования потенциально анеуплоидной хромосомы (или хромосомной области) с одной или несколькими другими предположительно не являющимися анеуплоидными хромосомами, или общая плотность метилирования генома может применяться для эффективной нормировки концентрации плодной ДНК в образце плазмы. Сравнение может проводиться посредством расчета параметра (например, включающего отношение или разность), связывающего плотности метилирования двух областей, для получения нормированной плотности метилирования. Указанное сравнение может устранять зависимый характер полученного уровня метилирования (например, определенного как параметр по двум плотностям метилирования).

Если плотность метилирования потенциально анеуплоидной хромосомы не нормирована по плотности метилирования одной или нескольких других хромосом, или по другим параметрам, отражающим фракционную концентрацию плодной ДНК, фракционная концентрация будет представлять собой основной фактор, влияющий на плотность метилирования в плазме. Например, плотность метилирования хромосомы 21 в плазме беременной женщины, вынашивающей плод с трисомией 21, при фракционной концентрации плодной ДНК 10% будет такой же, что и у беременной женщины, вынашивающей эуплоидный плод, при фракционной концентрации плодной ДНК 15%, тогда как нормированная плотность метилирования покажет различие.

Согласно другому варианту реализации плотность метилирования потенциально анеуплоидной хромосомы может быть нормирована по фракционной концентрации плодной ДНК. Например, для нормировки плотности метилирования может быть использовано следующее уравнение: $MP_{Normalized} = MP_{non-normalized} + (BKG - PLN) \times f$, где $MP_{Normalized}$ представляет собой плотность метилирования, нормированную по фракционной концентрации плодной ДНК в плазме, $MP_{Non-normalized}$ представляет собой измеренную плотность метилирования, BKG представляет собой фоновую плотность метилирования в клетках материнской крови или тканей, PLN представляет собой плотность метилирования в плацентарных тканях, и f представляет собой фракционную концентрацию плодной ДНК. Плотности метилирования BKG и PLN могут быть основаны на референсных значениях, установленных ранее по клеткам материнской крови и плацентарным тканям, полученным при здоровых беременностях. Для определения фракционной концентрации плодной ДНК в образце плазмы могут применяться различные генетические и эпигенетические способы, например, измерение процента ридов секвенирования от Y-хромосомы с применением массивно-параллельного

секвенирования или ПЦР не конвертированной бисульфитом ДНК.

Согласно одному варианту реализации нормированная плотность метилирования потенциально анеуплоидной хромосомы может сравниваться с референсной группой, которая представлена беременной женщиной, вынашивающей эуплоидные плоды. Может быть определено среднее и SD нормированной плотности метилирования референсной группы. Затем нормированная плотность метилирования в тестируемом случае может быть представлена в виде Z-показателя, отражающего число стандартных отклонений SD

от среднего значения для референсной группы, как:
$$z - score = \frac{MP_{Normalized} - Mean}{SD}$$
, где $MP_{Normalized}$ представляет собой нормированную плотность метилирования для тестируемого случая, $Mean$ представляет собой среднее нормированной плотности метилирования референсных случаев и SD представляет собой стандартное отклонение нормированной плотности метилирования референсных случаев. Точка отсечения, например, Z-показатель < -3 , может применяться для классификации хромосомы как значительно гипометилированной и, таким образом, для определения анеуплоидного статуса образца.

Согласно другому варианту реализации MP_{Diff} может применяться в качестве нормированной плотности метилирования. Согласно подобному варианту реализации PLN может быть выведен, например, с применением способа 1000. Согласно некоторым вариантам реализации референсная плотность метилирования (которая может быть нормирована с использованием f) может быть определена по уровню метилирования не-анеуплоидной области. Например, среднее ($Mean$) может быть определено по одной или нескольким хромосомным областям в одном образце. Точка отсечения может быть масштабирована по f , или просто установлена на уровне, достаточном при условии присутствия минимальной концентрации.

Соответственно, сравнение уровня метилирования для области с точкой отсечения может осуществляться различным образом. Указанное сравнение может включать нормировку (например, согласно приведенному выше описанию), которая может осуществляться эквивалентным образом по уровню метилирования или по предельному значению, в зависимости от способа определения значений. Соответственно, можно различными способами определить, отличается ли статистически уровень метилирования области от референсного уровня (определенный по тому же образцу или другим образцам).

Описанный выше анализ может применяться при анализе хромосомных областей, которые могут включать целую хромосому или части хромосомы, включая

последовательные или разобщенные подобласти хромосомы. Согласно одному варианту реализации потенциально анеуплоидная хромосома может быть разделена на ряд участков. Размеры указанных участков могут быть одинаковыми или разными. Плотность метилирования каждого участка может быть нормирована по фракционной концентрации образца или по плотности метилирования одной или нескольких предположительно неанеуплоидных хромосом, или по общей плотности метилирования генома. Затем нормированную плотность метилирования каждого участка можно сравнить с референсной группой для определения наличия ее значительного гипометилирования. Затем может быть определен процент значительно гипометилированных участков. Точка отсечения, например, более чем 5%, 10%, 15%, 20% или 30% значительно гипометилированных участков, может применяться для классификации анеуплоидного статуса тестируемого случая.

При тестировании на наличие амплификации или делеции плотность метилирования может сравниваться с референсной плотностью метилирования, которая может быть специфической для конкретной тестируемой области. Каждая область может иметь отличную референсную плотность метилирования, так как метилирование может варьировать от области к области, в частности, в зависимости от размера областей (например, области меньшего размера демонстрируют бóльшую вариабельность).

Как упоминалось выше, одна или несколько беременных женщин, каждая из которых вынашивает эуплоидный плод, могут использоваться для определения нормального диапазона плотности метилирования представляющей интерес области или разности плотностей метилирования между двумя хромосомными областями. Нормальный диапазон может также быть определен для PLN (например, путем прямого измерения или выведен согласно способу 1000). Согласно другим вариантам реализации для анализа может применяться отношение двух плотностей метилирования, например, потенциально анеуплоидной хромосомы и неанеуплоидной хромосомы, вместо их разности. Указанный способ анализа метилирования может быть скомбинирован с методом подсчета ридов последовательностей (RWK Chiu et al. 2008 Proc Natl Acad Sci USA;105:20458-20463) и методами, включающими анализ размеров ДНК плазмы (патент США 2011/0276277), для определения или подтверждения анеуплоидии. Метод подсчета ридов последовательностей, применяемый в комбинации с анализом метилирования, может осуществляться либо с применением случайного секвенирования (RWK Chiu et al. 2008 Proc Natl Acad Sci USA;105:20458-20463; DW Bianchi DW et al. 2012 Obstet Gynecol 119:890-901), либо направленного секвенирования (AB Sparks et al. 2012 Am J Obstet Gynecol 206:319.e1-9; B Zimmermann et al. 2012 Prenat Diagn 32:1233-1241; GJ Liao et al.

2012 PLoS One; 7:e38154).

Использование ВКГ позволяет учитывать вариации фона между образцами. Например, у одной женщины уровни метилирования ВКГ могут отличаться от уровней у другой женщины, однако в такой ситуации можно использовать разность ВКГ и PLN для всех образцов. Точка отсечения для разных хромосомных областей может быть разной, например, если плотность метилирования одной области генома отличается от плотности метилирования другой области генома.

Указанный способ может быть обобщен для детекции любых хромосомных aberrаций, включая делеции и амплификации, в плодном геноме. Кроме того, разрешение указанного анализа может быть скорректировано до требуемого уровня, например, геном может быть разделен на участки 10 Мб, 5 Мб, 2 Мб, 1 Мб, 500 кб, 100 кб. Соответственно, указанная технология также может применяться для детектирования субхромосомных дупликаций или субхромосомных делеций. Указанная технология, соответственно, позволяет устанавливать пренатальный молекулярный кариотип плода неинвазивным образом. При применении указанным образом данная технология может применяться в комбинации со способами неинвазивного пренатального тестирования на основе подсчета молекул (A Srinivasan et al. 2013 Am J Hum Genet;92:167-176; SCY Yu et al. 2013 PLoS One 8: e60968). Согласно другим вариантам реализации размер участков не обязательно должен быть идентичен. Например, размер участков может быть скорректирован таким образом, чтобы каждый участок содержал идентичное число динуклеотидов CpG. В указанном случае физический размер участков будет разным.

Уравнение может быть переформулировано для применения для разных типов хромосомных aberrаций как $MP_{Diff} = (BKG - PLN) \times f \times 0.5 \times CN$, где CN – величина изменения числа копий в пораженной области. CN = 1 при добавлении 1 копии хромосомы, 2 при добавлении 2 копий хромосомы и -1 для утраты одной из двух гомологичных хромосом (например, при детектировании синдрома Тернера у плода, при котором у плода женского пола утрачена одна из X-хромосом, что приводит к кариотипу XO). Указанное уравнение не требует внесения поправок при изменении размера участков. Тем не менее, чувствительность и специфичность могут снижаться при использовании участков меньшего размера, так как на участках меньшего размера присутствует меньшее число динуклеотидов CpG (или других комбинаций нуклеотидов, для которых наблюдаются различия в метилировании плодной ДНК и материнской ДНК), что приводит к увеличению случайных колебаний при измерении плотностей метилирования. Согласно одному варианту реализации необходимое число рядов может быть определено путем анализа коэффициента вариации плотности метилирования и требуемого уровня

чувствительности.

Чтобы продемонстрировать осуществимость указанного способа, авторы настоящего изобретения исследовали образцы плазмы 9 беременных женщин. Каждая из пяти беременных женщин вынашивала эуплоидный плод, каждая из остальных четырех женщин вынашивала плод с трисомией 21 (T21). Три из пяти эуплоидных беременностей были выбраны случайным образом для формирования референсной группы. Остальные две эуплоидные беременности (Eu1 и Eu2) и четыре случая T21 (T21-1, T21-2, T21-3 и T21-4) исследовали с применением указанного способа для тестирования потенциального статуса T21. ДНК плазмы конвертировали бисульфитом и секвенировали с применением платформы Illumina HiSeq2000. Согласно одному варианту реализации рассчитывали плотность метилирования индивидуальных хромосом. Затем определяли разность плотности метилирования между хромосомой 21 и средним значением по другим 21 аутосомам для получения нормированной плотности метилирования (таблица 1). Среднее и SD референсной группы использовали для расчета Z-показателя для шести тестируемых случаев.

Таблица 1. При применении точки отсечения <-3 для Z-показателя для классификации образца как T21, классификация всех случаев эуплоидии и T21 была корректной

	Eu1	Eu2	T21-1	T21-2	T21-3	T21-4
Z-показатель для MP_{Diff} между хромосомой 21 и другими аутосомами	-1,48	1,09	-4,46	-5,30	-8,06	-5,69

Согласно другому варианту реализации геном разделяли на участки размером 1 Мб и определяли плотность метилирования для каждого участка размером 1 Мб. Плотность метилирования всех участков на потенциально анеуплоидной хромосоме может быть нормирована по медиане плотности метилирования всех участков, расположенных на предположительно не являющихся анеуплоидными хромосомах. Согласно одному варианту реализации для каждого участка может быть рассчитано различие плотности метилирования от медианы не являющихся анеуплоидными участков. Z-показатель может быть рассчитан для указанных значений с применением значений среднего и SD референсной группы. Можно определить процент участков, где наблюдается гипометилирование (таблица 2), и сравнить с процентом, соответствующим точке отсечения.

Таблица 2. При применении 5% в качестве точки отсечения для значительно более гипометилированных участков на хромосоме 21, все случаи были классифицированы корректно в отношении статуса T21

	Eu1	Eu2	T21-1	T21-2	T21-3	T21-4
Процент участков на хромосоме 21 с Z-показателем для $MP_{Diff} < -3$	0%	0%	33,3%	58,3%	19,4%	52,8%

Указанный способ детектирования плодных хромосомных или субхромосомных aberrаций на основе метилирования ДНК может применяться в сочетании с методами, основанными на подсчета молекул, например, посредством секвенирования (RWK Chiu et al. 2008 Proc Natl Acad Sci USA; 105: 20458-20463), или цифровой ПЦР (YMD Lo et al. 2007 Proc Natl Acad Sci USA; 104: 13116-13121), или определением размеров молекул ДНК (патентная публикация США 2011/0276277). Такое комбинирование (например, метилирование ДНК плюс подсчет молекул, или метилирование ДНК плюс определение размеров, или метилирование ДНК плюс подсчет молекул и плюс определение размеров) будет давать синергетический эффект, который был бы благоприятным в клинических условиях, например, повышал бы чувствительность и/или специфичность. Например, число молекул ДНК, которое необходимо проанализировать, например, посредством секвенирования, может быть уменьшено без отрицательного воздействия на диагностическую точность. Указанная особенность позволяет проводить такие тесты с меньшими затратами. В другом примере для определенного анализируемого числа молекул ДНК комбинированный подход позволяет детектировать плодные хромосомные или субхромосомные aberrации при более низкой фракционной концентрации плодной ДНК.

На фиг. 13 приведена блок-схема способа 1300 детектирования хромосомной аномалии по биологическому образцу организма. Указанный биологический образец включает свободную от клеток ДНК, содержащую смесь свободной от клеток ДНК, происходящей из первой ткани и из второй ткани. Первая ткань может происходить из плода или опухоли, и вторая ткань может происходить от беременной женщины или от пациента.

На этапе 1310 анализируют множество молекул ДНК из указанного биологического образца. Анализ молекулы ДНК может включать определение локализации указанной молекулы ДНК в геноме указанного организма и определение того, метилирована ли указанная молекула ДНК в одном или большем числе сайтов. Анализ может осуществляться путем получения ридов последовательностей с помощью чувствительного к метилированию секвенирования, и, соответственно, анализ может

осуществляться исключительно по ранее полученным для ДНК данным. Согласно другим вариантам реализации анализ может включать фактическое секвенирование или другие активные этапы получения данных.

Определение локализации может включать картирование молекул ДНК (например, с применением ридов последовательностей) на соответствующие части генома человека, например, на специфические области. Согласно одному варианту реализации, если рид не картируется на представляющую интерес область, он может не учитываться.

На этапе 1320 для каждого из множества сайтов определяют относительное количество молекул ДНК, которые метилированы в указанном сайте. Согласно одному варианту реализации указанные сайты представляют собой CpG-сайты, и могут представлять собой только определенные CpG-сайты, выбранные с применением одного или большего количества упоминаемых в настоящем документе критериев. Определение количества метилированной ДНК эквивалентно определению количества неметилированной ДНК, если нормировка проводится с использованием общего числа анализируемых молекул ДНК в конкретном сайте, например, общего числа ридов секвенирования.

На этапе 1330 рассчитывают первый уровень метилирования первой хромосомной области на основании соответствующих количеств молекул ДНК, метилированных в сайтах в составе первой хромосомной области. Указанная первая хромосомная область может быть любого размера, например, иметь размеры, описанные выше. Уровень метилирования может учитывать общее число молекул ДНК, выравниваемых с первой хромосомной областью, например, в рамках процедуры нормировки.

Первая хромосомная область может быть любого размера (например, представлять собой целую хромосому) и может быть состоять из разобщенных подобластей, т.е. подобластей, которые отделены одна от другой. Уровни метилирования каждой подобласти могут быть определены и скомбинированы, например, в виде среднего значения или медианы, для определения уровня метилирования первой хромосомной области.

На этапе 1340 первый уровень метилирования сравнивают с предельным значением (значением отсечения). Предельное значение может представлять собой референсный уровень метилирования или быть связано с референсным уровнем метилирования (например, находится на определенном расстоянии от нормального уровня). Предельное значение может быть определено по другим беременным субъектам женского пола, вынашивающим плоды без хромосомной аномалии по первой хромосомной области, по образцам индивидуумов, не страдающих раковым заболеванием или по локусам

указанного организма, которые, как известно, не связаны с анеупloidией (т.е. по областям, которые являются дисомическими).

Согласно одному варианту реализации предельное значение может быть определено как отличающееся от референсного уровня метилирования на $(BKG - PLN) \times f \times 0,5 \times CN$, где BKG представляет собой фон указанной женщины (или среднее значение или медиану для других субъектов), f представляет собой фракционную концентрацию свободной от клеток ДНК, происходящей из первой ткани, и CN представляет собой тестируемое число копий. CN представляет собой пример масштабирующего коэффициента, соответствующего типу аномалии (делеции или дупликации). Изначально для тестирования всех амплификаций может использоваться точка отсечения для CN, равная 1, и дополнительные точки отсечения могут применяться для определения степени амплификации. Предельное значение может быть основано на фракционной концентрации свободной от клеток ДНК, происходящей из первой ткани, для определения ожидаемого уровня метилирования для локуса, например, если абберация числа копий отсутствует.

На этапе 1350 классифицируют аномалию первой хромосомной области на основе указанного сравнения. Статистически значимое различие уровней может указывать на повышенный риск хромосомной аномалии у плода. Согласно различным вариантам реализации указанная хромосомная аномалия может представлять собой трисомию 21, трисомию 18, трисомию 13, синдром Тернера или синдром Клайнфельтера. Другие примеры представлены субхромосомной делецией, субхромосомной дупликацией или синдромом Ди Джорджи.

V. Определение маркеров

Как было отмечено выше, некоторые части плодного генома метилированы иначе, чем материнский геном. Указанные различия могут быть общими для всех беременностей. Области отличающегося метилирования могут использоваться для идентификации фрагментов ДНК, происходящих от плода.

A. Способ определения DMR по плацентарной ткани и материнской ткани

Плацента содержит тканеспецифичные сигнатуры метилирования. Для детекции в материнской плазме и для применения при неинвазивной пренатальной диагностике были разработаны специфичные для плода маркеры метилирования ДНК на основе локусов, дифференциально метилированных в плацентарных тканях и в клетках материнской крови (SSC Chim et al. 2008 Clin Chem; 54: 500-511; EA Papageorgiou et al 2009 Am J Pathol; 174: 1609-1618; и T Chu et al. 2011 PLoS One; 6: e14723). Предложены варианты реализации для анализа таких дифференциально метилированных областей (DMR) в полногеномном

масштабе.

Фиг. 14 представляет собой блок-схему способа 1400 идентификации маркеров метилирования путем сравнения профиля метилирования плаценты с материнским профилем метилирования (например, определенным по клеткам крови) в соответствии с вариантами реализации настоящего изобретения. Способ 1400 может также применяться для определения маркеров опухоли путем сравнения профиля метилирования опухоли с профилем метилирования, соответствующим здоровой ткани.

На этапе 1410 получают плацентарный метилом и метилом крови. Плацентарный метилом может быть определен по плацентарному образцу, например, CVS или зрелой плаценты. Следует понимать, что метилом может включать плотности метилирования только части генома.

На этапе 1420 идентифицируют область, которая содержит указанное число сайтов (например, 5 CpG -сайтов) и для которой было получено достаточное число ридов. Согласно одному варианту реализации идентификацию начинают с одного конца каждой хромосомы для определения локализации первой области размером 500 п.о., содержащей по меньшей мере пять подходящих CpG-сайтов. CpG-сайт может считаться подходящим, если указанный сайт покрывают по меньшей мере пять ридов последовательностей.

На этапе 1430 для каждого сайта рассчитывают индекс метилирования плаценты и индекс метилирования крови. Например, индекс метилирования рассчитывали индивидуально для всех подходящих CpG-сайтов в каждой области размером 500 п.о.

На этапе 1440 сравнивали индексы метилирования клеток материнской крови и индексы метилирования плацентарного образца, для определения того, различаются ли множества указанных индексов. Например, сравнивали индексы метилирования клеток материнской крови и индексы метилирования CVS или зрелой плаценты с применением, например, критерия Манна-Уитни. *P*-значение, составляющее, например, $\leq 0,01$ считали определяющим статистически значимое различие, хотя могут использоваться и другие значения, при этом меньшее значение приведет к уменьшению количества ложноположительных областей.

Согласно одному варианту реализации в случае, если число подходящих CpG-сайтов составляло менее 5, или критерий Манна-Уитни показывал отсутствие статистической значимости, область размером 500 п.о. сдвигали ниже (в 3'-направлении) на 100 п.о. Область продолжали сдвигать в 3'-направлении, пока критерий Манна-Уитни не показывал статистическую значимость для области размером 500 п.о. Тогда рассматривали следующую область размером 500 п.о. Если выяснялось, что для следующей области наблюдается статистическая значимость согласно критерию Манна-

Уитни, ее добавляли к текущей области, при условии, что размер комбинированной последовательной области не превышал 1000 п.о.

На этапе 1450 смежные области, которые были определены как статистически значимо отличающиеся (например, по критерию Манна-Уитни), могут быть объединены. Отметим, что указанное различие относится к индексам метилирования для двух образцов. Согласно одному варианту реализации в случае, если смежные области находятся в пределах указанного расстояния (например, 1000 п.о.) друг от друга, и если для них наблюдается аналогичный профиль метилирования, их объединяют. Согласно одному варианту реализации сходство профиля метилирования смежных областей может быть определено на основании любого из следующих пунктов: (1) при наблюдаемой одинаковой тенденции в плацентарной ткани относительно клеток материнской крови, например, обе области сильнее метилированы в плацентарных тканях, чем в клетках крови; (2) по различиям плотностей метилирования, составляющим менее чем 10% для смежных областей в плацентарной ткани; и (3) по различиям плотностей метилирования, составляющим менее чем 10% для смежных областей в клетках материнской крови.

На этапе 1460 рассчитывали плотности метилирования для метилома крови по ДНК клеток материнской крови и плацентарного образца (например, CVS или ткани зрелой плаценты) в указанных областях. Плотности метилирования могут быть определены согласно описанию в настоящем документе.

На этапе 1470 определяют предполагаемые DMR, отличающиеся тем, что общая плацентарная плотность метилирования и общая плотность метилирования крови статистически значимо различаются для всех сайтов в указанной области. Согласно одному варианту реализации все подходящие CpG-сайты в составе объединенной области оценивают с применением критерия χ^2 . Критерий χ^2 показывал, присутствовало ли статистически значимое различие между клетками материнской крови и плацентарной ткани по числу метилированных цитозинов, как пропорции метилированных и неметилированных цитозинов во всех подходящих CpG-сайтах в составе объединенной области. Согласно одному варианту реализации для критерия χ^2 P-значение $\leq 0,01$ может считаться определяющим статистически значимое различие. Объединенные сегменты, для которых была показана значимость с применением критерия χ^2 , считали предполагаемыми DMR.

На этапе 1480 идентифицировали локусы, где плотности метилирования ДНК клеток материнской крови были выше верхней точки отсечения или ниже нижней точки отсечения. Согласно одному варианту реализации идентифицировали локусы, где плотности метилирования ДНК клеток материнской крови составляли либо $\leq 20\%$, либо \geq

80%. Согласно другим вариантам реализации могут быть использованы жидкости организма, отличные от материнской крови, включая, не ограничиваясь перечисленным, слюну, смывы с матки или шейки матки из половых путей женщины, слезы, пот, слюну и мочу.

Ключом к успешному получению специфичных для плода маркеров метилирования ДНК в материнской плазме может быть то, что статус метилирования клеток материнской крови соответствует либо высокометилированному, либо настолько неметилированному, насколько это возможно. Это может снизить (например, минимизировать) вероятность влияния молекул материнской ДНК на анализ происходящих из плаценты молекул плодной ДНК, демонстрирующих противоположный профиль метилирования. Соответственно, согласно одному варианту реализации кандидатные DMR выбирали путем дополнительной фильтрации. Кандидатными гипометилированными локусами считали локусы, демонстрирующие плотности метилирования $\leq 20\%$ в клетках материнской крови и по меньшей мере на 20% более высокие плотности метилирования в плацентарных тканях. Кандидатными гиперметилированными локусами считали локусы, демонстрирующие плотности метилирования $\geq 80\%$ в клетках материнской крови и по меньшей мере на 20% меньшие плотности метилирования в плацентарных тканях. Могут быть использованы и другие процентные отношения.

Затем на этапе 1490 идентифицировали DMR в подмножестве локусов, где плацентарные плотности метилирования значительно отличаются от плотностей метилирования крови путем сравнения указанного различия с пороговым. Согласно одному варианту реализации порог составляет 20%, так что плотности метилирования отличались по меньшей мере на 20% от плотностей метилирования клеток материнской крови. Соответственно, может быть рассчитана разность между плацентарными плотностями метилирования и плотностями метилирования крови в каждом из идентифицированных локусов. Указанная разность может представлять собой результат простого вычитания. Согласно другим вариантам реализации могут применяться масштабирующие коэффициенты и другие функции для определения разности (например, разность может быть представлена в виде результата применения функции к результату простого вычитания).

Согласно одному варианту реализации с применением указанного способа идентифицировали 11 729 гиперметилированных и 239 747 гипометилированных локусов по плацентарному образцу первого триместра. 100 первых гиперметилированных локусов перечислены в таблице S2A приложения. 100 первых гипометилированных локусов

перечислены в таблице S2B приложения. В таблицах S2A и S2B указана хромосома, локализация начала и конца, размер области, плотность метилирования в материнской крови, плотность метилирования в плацентарном образце, P-значения (все P-значения очень невысокие) и различия в метилировании. Локализации соответствуют референсному геному hg18, который можно найти по адресу: hgdownload.soe.ucsc.edu/goldenPath/hg18/chromosomes.

11 920 гиперметилированных и 204 768 гипометилированных локусов идентифицировали по плацентарному образцу третьего триместра. 100 первых гиперметилированных локусов для третьего триместра перечислены в таблице S2C, и 100 первых гипометилированных локусов перечислены в таблице S2D. 33 локуса, которые были ранее описаны как дифференциально метилированные в клетках материнской крови и в плацентарных тканях первого триместра, использовали для валидации полученного авторами настоящего изобретения списка кандидатных локусов первого триместра. 79% из указанных 33 локусов были идентифицированы как DMR с применением предложенного авторами настоящего изобретения алгоритма.

Фиг. 15A представляет собой таблицу 1500, отражающую эффективность алгоритма идентификации DMR при применении данных для первого триместра с учетом 33 ранее описанных маркеров первого триместра. В указанной таблице «а» означает, что локусы 1–15 были ранее описаны в (RWK Chiu et al. 2007 Am J Pathol; 170:941-950 и SSC Chim et al. 2008 Clin Chem; 54:500-511); локусы 16–23 были ранее описаны в (KC Yuen, thesis 2007, The Chinese University of Hong Kong, Hong Kong); и локусы 24–33 были ранее описаны в (EA Papageorgiou et al. 2009 Am J Pathol; 174:1609-1618). «b» означает, что указанные данные были получены из перечисленных выше публикаций. «с» означает, что плотности метилирования клеток материнской крови и образца ворсин хориона и их различия наблюдали в данных секвенирования, полученных в настоящем исследовании, но они основаны на геномных координатах, приведенных в оригинальных исследованиях. «d» означает, что данные для локусов идентифицированы с применением вариантов реализации способа 1400 на данных бисульфитного секвенирования без использования референсных данных из упоминаемых выше публикаций Chiu et al (2007), Chim et al (2008), Yuen (2007) и Papageorgiou et al (2009). Указанные локусы содержали ранее описанные геномные области, однако, как правило, захватывали более протяженные области. «е» означает, что кандидатные DMR классифицировали как истинно положительные (TP) или ложноотрицательные (FN) на основании соблюдения требования наблюдаемой разности $> 0,20$ между плотностями метилирования, соответствующими геномным координатам указанных DMR в клетках материнской крови и в образце ворсин

хориона.

Фиг. 15В представляет собой таблицу 1550, отражающую эффективность алгоритма идентификации DMR с применением данных для третьего триместра и сравнения с образцом плаценты, полученным при родах. «а» означает, что использовали тот же список из 33 локусов, что и на фиг. 17А. «b» означает, что поскольку указанные 33 локуса были ранее идентифицированы по образцам, полученным на ранних сроках беременности, они могут не соответствовать данным для третьего триместра. Поэтому данные бисульфитного секвенирования, полученные в настоящем исследовании для ткани зрелой плаценты на основе геномных координат, полученных в оригинальных исследованиях, пересматривались. Разность плотностей метилирования $> 0,20$ между клетками материнской крови и тканью зрелой плаценты использовали для определения того, действительно ли указанные локусы являлись областями DMR в третьем триместре. «с» означает, что данные по локусам были идентифицированы с применением способа 1400 к данным бисульфитного секвенирования без учета упомянутых выше публикаций Chiu et al (2007), Chim et al (2008), Yuen (2007) и Parageorgiou et al (2009). Указанные локусы содержали ранее описанные геномные области, однако, как правило, захватывали более протяженные области. «d» означает, что кандидатные DMR, содержащие локусы, подходящие в качестве дифференциально метилированных в третьем триместре, классифицировали как истинно положительные (TP) или ложноотрицательные (FN) на основании соблюдения требования наблюдающейся разности $> 0,20$ между плотностями метилирования соответствующих геномных координат DMR в клетках материнской крови и в ткани зрелой плаценты. Для локусов, не подходящих в качестве дифференциально метилированных в третьем триместре, отсутствующие в списке DMR или присутствующие DMR, содержащие указанные локусы, но демонстрирующие разность метилирования $< 0,20$, считали областями DMR с истинно отрицательным результатом (TN).

В. DMR по данным секвенирования материнской плазмы

DMR плацентарной ткани можно идентифицировать непосредственно по данным бисульфитного секвенирования ДНК материнской плазмы, при условии, что также известна фракционная концентрация плодной ДНК образца. Это возможно, поскольку плацента является основным источником плодной ДНК в материнской плазме (SSC Chim et al. 2005 Proc Natl Acad Sci USA 102, 14753-14758), и авторы настоящего изобретения показали в настоящем исследовании, что статус метилирования специфичной для плода ДНК в материнской плазме коррелирует с плацентарным метиломом.

Таким образом, аспекты способа 1400 могут быть реализованы с использованием

метилома плазмы для выведения плацентарного метилома вместо использования плацентарного образца. Соответственно, способ 1000 и способ 1400 могут быть скомбинированы для определения DMR. Способ 1000 может применяться для определения прогнозируемых значений профиля метилирования плаценты и их использования в способе 1400. В примере указанного анализа внимание также сфокусировано на локусах, которые были либо $\leq 20\%$, либо $\geq 80\%$ метилированы в клетках материнской крови.

Согласно одному варианту реализации для определения локусов, которые были гиперметилированы в плацентарных тканях относительно клеток материнской крови, авторы настоящего изобретения отбирали локусы, демонстрировавшие метилирование $\leq 20\%$ в клетках материнской крови, и метилирование $\geq 60\%$ в соответствии с предсказанным значением с разностью по меньшей мере 50% между плотностью метилирования клетками крови и предсказанным значением. Для определения локусов, которые были гипометилированы в плацентарных тканях относительно клеток материнской крови, авторы настоящего изобретения отбирали локусы, демонстрировавшие метилирование $\geq 80\%$ в клетках материнской крови, и метилирование $\leq 40\%$ в соответствии с предсказанным значением с разностью по меньшей мере 50% между плотностью метилирования клеток крови и предсказанным значением.

Фиг. 16 представляет собой таблицу 1600, отражающую число локусов, гиперметилированных или гипометилированных согласно прогнозу на основании прямого анализа данных бисульфитного секвенирования материнской плазмы. «Н/О» означает «не определено». «а» означает, что поиск гиперметилированных локусов начинали со списка локусов, демонстрирующих плотности метилирования $< 20\%$ в клетках материнской крови. «b» означает, что поиск гипометилированных локусов начинали со списка локусов, демонстрирующих плотности метилирования $> 80\%$ в клетках материнской крови. «с» означает, что данные бисульфитного секвенирования для образца ворсин хориона использовали для проверки данных материнской плазмы первого триместра, и ткань зрелой плаценты использовали для проверки данных материнской плазмы третьего триместра.

Как видно из таблицы 1600, большинство определенных неинвазивным образом локусов демонстрировали ожидаемый паттерн метилирования в тканях и перекрывались с DMR, установленными по данным тканей и представленными в разделе выше. В приложении приведен список DMR, идентифицированных в плазме. В таблице S3A перечислены 100 первых локусов, для которых установлено гиперметилирование по данным бисульфитного секвенирования для материнской плазмы первого триместра. В

таблице S3B перечислены 100 первых локусов, для которых установлено гипометилирование по данным бисульфитного секвенирования для материнской плазмы первого триместра. В таблице S3C перечислены 100 первых локусов, для которых установлено гиперметилирование по данным бисульфитного секвенирования для материнской плазмы третьего триместра. В таблице S3D перечислены 100 первых локусов, для которых установлено гипометилирование по данным бисульфитного секвенирования для материнской плазмы третьего триместра.

C. Гестационные вариации плацентарного и плодного метиломов

Общая доля метилированных CpG в CVS составляла 55%, и 59% для зрелой плаценты (таблица 100 на фиг. 1). В CVS могут быть идентифицированы более гипометилированные DMR, чем в зрелой плаценте, при этом некоторые гиперметилированные DMR аналогичны в обеих тканях. Соответственно, было очевидно, что CVS более гипометилированы, чем зрелая плацента. Указанная тенденция при гестации была также очевидна для данных по материнской плазме. Доля метилированных CpG в специфичных для плода ридсах составляла 47,0% в материнской плазме в первом триместре, но 53,3% в материнской плазме в третьем триместре. Количества валидированных гиперметилированных локусов были аналогичны в образцах материнской плазмы первого триместра (1457 локусов) и третьего триместра (1279 локусов) однако в образцах первого триместра присутствовали по существу более гипометилированные локусы (21 812 локусов), чем в образцах третьего триместра (12 677 локусов) (таблица 1600 на фиг. 16).

D. Применение маркеров

Дифференциально метилированные маркеры, или DMR, подходят для применения в нескольких аспектах. Присутствие таких маркеров в материнской плазме указывает на наличие и подтверждает присутствие или плодной, или плацентарной ДНК. Указанное подтверждение может применяться для контроля качества при неинвазивном пренатальном тестировании. DMR могут служить универсальными маркерами плодной ДНК в материнской плазме и обладают преимуществами относительно маркеров, основанных на генотипических различиях между матерью и плодом, таких как маркеры на основе полиморфизмов или на основе Y-хромосомы. DMR являются универсальными плодными маркерами, подходящими для всех беременностей. Маркеры на основе полиморфизмов подходят для применения только в подгруппе беременностей, где плод наследует указанный маркер от отца, а геном матери не содержит указанного маркера. Кроме того, можно измерить концентрацию плодной ДНК в образце материнской плазмы путем количественного определения молекул ДНК, происходящих из указанных DMR.

При наличии известного профиля DMR, ожидаемого для нормальных беременностей, могут быть детектированы связанные с беременностью осложнения, в частности, включающие изменения плацентарной ткани, при наблюдении отклонений профиля DMR в материнской плазме или профиля метилирования от ожидаемого для нормальных беременностей. Связанные с беременностью осложнения, включающие изменения плацентарной ткани, включают, не ограничиваясь указанным, плодные хромосомные анеуплоидии. Примеры включают трисомию 21, преэклампсию, внутриутробную задержку роста и преждевременные роды.

Е. Наборы с маркерами

Согласно вариантам реализации могут обеспечиваться композиции и наборы для реализации описанных в настоящем документе и других подходящих способов. Наборы могут применяться для проведения анализов плодной ДНК, например, свободной от клеток плодной ДНК в материнской плазме. Согласно одному варианту реализации набор может включать по меньшей мере один олигонуклеотид, подходящий для специфической гибридизации с одним или большим количеством локусов, идентифицированных согласно настоящему документу. Набор также может включать по меньшей мере один олигонуклеотид, подходящий для специфической гибридизации с одним или большим количеством референсных локусов. Согласно одному варианту реализации измеряют плацентарные гиперметилированные маркеры. Тестируемый локус может представлять собой метилированную ДНК в материнской плазме, и референсный локус может представлять собой метилированную ДНК в материнской плазме. Аналогичный набор может быть собран для анализа опухолевой ДНК в плазме.

В некоторых случаях наборы могут включать по меньшей мере два олигонуклеотидных праймера, которые могут применяться при амплификации по меньшей мере части целевого локуса (например, локуса, представленного в приложении) и референсного локуса. Вместо или помимо праймеров набор может включать меченые зонды для детекции фрагмента ДНК, соответствующего целевому локусу и референсному локусу. Согласно различным вариантам реализации один или большее количество олигонуклеотидов набора соответствуют локусу из таблиц приложения. Как правило, указанные наборы также содержат руководство по эксплуатации для направления пользователей при анализе тестируемых образцов и оценке физиологического состояния или патологии у тестируемого субъекта.

Согласно различным вариантам реализации предложен набор для анализа плодной ДНК в биологическом образце, содержащем смесь плодной ДНК и ДНК субъекта – беременной плодом женщины. Набор может включать один или большее количество

олигонуклеотидов для специфической гибридизации по меньшей мере с частью геномной области из перечисленных в таблицах S2A, S2B, S2C, S2D, S3A, S3B, S3C и S3D. Соответственно, может быть использовано любое число олигонуклеотидов из любых таблиц или только из одной таблицы. Указанные олигонуклеотиды могут функционировать в качестве праймеров, и могут быть представлены в виде пар праймеров, при этом пара соответствует конкретной области из перечисленных в таблицах.

VI. Связь между размером и плотностью метилирования

Известно, что молекулы ДНК плазмы присутствуют в кровотоке в форме коротких молекул, при этом длина большинства молекул составляет приблизительно 160 п.о. (YMD Lo et al. 2010 Sci Transl Med; 2: 61ra91, YW Zheng et al. 2012 Clin Chem; 58: 549-558). Интересно, что данные авторов настоящего изобретения демонстрируют зависимость между статусом метилирования и размером молекул ДНК плазмы. Соответственно, длина фрагмента ДНК плазмы связана с уровнем метилирования ДНК. Характеристические профили размеров молекул ДНК плазмы предполагают, что бóльшая их часть связана с моонуклеосомами, возможно, образующимися при ферментативном расщеплении во время апоптоза.

Циркулирующая ДНК в природе фрагментирована. В частности, циркулирующая плодная ДНК короче, чем происходящая от матери ДНК в образцах материнской плазмы (КСА Chan et al. 2004 Clin Chem; 50: 88-92). Так как выравнивание спаренных концов позволяет проводить анализ размера обработанной бисульфитом ДНК, можно непосредственно оценить наличие какой-либо корреляции между размером молекул ДНК плазмы и соответствующими им уровнями метилирования. Авторы настоящего изобретения исследовали указанную корреляцию в материнской плазме, а также в контрольном образце плазмы небеременной взрослой женщины.

Секвенирование спаренных концов (включающее секвенирование полной молекулы) для обоих концов каждой молекулы ДНК использовали для анализа каждого образца в указанном исследовании. Путем выравнивания пары концевых последовательностей каждой молекулы ДНК с референсным геномом человека и получения геномных координат концов ридов секвенирования, можно определить длины секвенированных молекул ДНК. Молекулы ДНК плазмы естественным образом фрагментированы на небольшие молекулы и библиотеки секвенирования для ДНК плазмы, как правило, получают без каких-либо этапов фрагментации. Соответственно, длины, установленные с помощью секвенирования, отражают размеры исходных молекул ДНК плазмы.

В предыдущем исследовании авторы настоящего изобретения определили профили размеров плодных и материнских молекул ДНК в материнской плазме (YMD Lo et al. 2010 Sci Transl Med; 2: 61ra91). Авторы настоящего изобретения показали, что молекулы ДНК плазмы имеют размеры, напоминающие мононуклеосомы; молекулы плодной ДНК были короче, чем материнские. В указанном исследовании авторы настоящего изобретения определили связь статуса метилирования молекул ДНК плазмы с их размерами.

А. Результаты

На фиг. 17А представлен график 1700, отражающий распределение размеров ДНК материнской плазмы, контрольной плазмы небеременной женщины, плацентарной ДНК и ДНК периферической крови. Два обработанных бисульфитом образца плазмы из образцов материнской плазмы и контрольной плазмы небеременной женщины демонстрировали те же характеристические распределения размеров, что и ранее описанные (YMD Lo et al. 2010 Sci Transl Med; 2: 61ra91), с наиболее распространенной общей длиной последовательностей 166–167 п.о. и периодичностью 10 п.о. для молекул ДНК, длина которых составляет менее 143 п.о.

Фиг. 17В представляет собой график 1750 распределения размеров и профиля метилирования материнской плазмы, контрольной плазмы взрослой женщины, плацентарной ткани и контрольной крови взрослой женщины. Для молекул ДНК одинакового размера и содержащих по меньшей мере один CpG-сайт рассчитывали среднее значение плотности метилирования. Затем авторы настоящего изобретения строили график зависимости между размерами молекул ДНК и их плотностями метилирования. В частности, определяли среднее значение плотности метилирования для каждой длины фрагмента, варьирующей от 50 п.о. до 180 п.о., для ридов секвенирования, покрывающих по меньшей мере 1 CpG-сайт. Интересно, что плотность метилирования увеличивалась при возрастании размера ДНК плазмы, и имела максимальное значение при размерах приблизительно 166–167 п.о. Указанный паттерн, тем не менее, не наблюдался в образцах ДНК плаценты и контрольных образцах крови, фрагментированных с применением ультразвуковой системы.

На фиг. 18 приведены графики плотностей метилирования и размера молекул ДНК плазмы. Фиг. 18А представляет собой график 1800 для материнской плазмы в первом триместре. Фиг. 18В представляет собой график 1850 для материнской плазмы в третьем триместре. Данные для всех ридов секвенирования, которые покрывали по меньшей мере один CpG-сайт, представлены голубой кривой 1805. Данные для ридов, содержащих также специфичный для плода аллель ОНП, представлены красной кривой 1810. Данные для ридов, содержащих также специфичный для матери аллель ОНП, представлены

зеленой кривой 1815.

Риды, содержащий специфичный для плода аллель ОНП, считали относящимися к молекулам плодной ДНК. Риды, содержащий специфичный для матери аллель ОНП, считали относящимися к молекулам материнской ДНК. Как правило, молекулы ДНК с высокими плотностями метилирования имели больший размер. Указанная тенденция наблюдалась для молекул как плодной, так и материнской ДНК, и в первом, и в третьем триместрах. Как ранее уже говорилось, общие размеры молекул плодной ДНК были меньше, чем размеры материнских молекул.

На фиг. 19А приведен график 1900 плотностей метилирования и размеров ридов секвенирования для взрослой небеременной женщины. Для образца ДНК плазмы взрослой небеременной женщины наблюдалась такая же зависимость между размерами и состоянием метилирования молекул ДНК. С другой стороны, на этапе обработки ультразвуком, до анализа с применением МПС, образцы геномной ДНК были фрагментированы. Как видно на графике 1900, в данных для образцов клеток крови и плацентарной ткани указанная тенденция отсутствует. Поскольку производится искусственная фрагментация клеток, можно ожидать, что связь между размером и плотностью будет отсутствовать. Так как для фрагментированных естественным образом молекул ДНК в плазме наблюдается зависимость от размера, можно предположить, что молекулы с более низкими плотностями метилирования с большей вероятностью распадаются на более короткие фрагменты.

Фиг. 19В представляет собой график 1950, отражающий распределение размеров и профиль метилирования специфичных для плода и специфичных для матери молекул ДНК в материнской плазме. Для специфичных для плода и специфичных для матери молекул ДНК плазмы также наблюдалась указанная корреляция между размерами фрагментов и уровнем метилирования. Длины фрагментов как происходящей из плаценты, так и материнской циркулирующей свободной от клеток ДНК увеличивались с увеличением уровня метилирования. Кроме того, распределения их статуса метилирования не перекрывались, предполагая, что указанное явление существует независимо от исходной длины фрагмента из источников циркулирующих молекул ДНК.

В. Способ

Соответственно, распределение размеров может применяться для определения общего процента метилирования образца плазмы. Указанное измерение метилирования затем может отслеживаться во время беременности, при мониторинге ракового заболевания, или при лечении путем серийных измерений распределений размеров ДНК плазмы в соответствии с зависимостью, представленной на фиг. 18А и 18В. Измерение

метиляции также может применяться для отслеживания повышенного или пониженного высвобождения ДНК из представляющего интерес органа или ткани. Например, можно конкретным образом отслеживать сигнатуры метилирования ДНК, специфические для конкретного органа (например, для печени) и измерять концентрации указанных сигнатур в плазме. Поскольку ДНК высвобождается в плазму при гибели клеток, повышение уровней может означать увеличение уровней клеточной смерти или повреждения указанного конкретного органа или ткани. Снижение уровня для конкретного органа может означать, что противодействующее повреждению лечение или патологические процессы в указанном органе находятся под контролем.

Фиг. 20 представляет собой блок-схему способа 2000 оценки уровня метилирования ДНК в биологическом образце организма в соответствии с вариантами реализации настоящего изобретения. Уровень метилирования может быть определен для конкретной области генома или для всего генома. Если требуется конкретная область, могут быть использованы фрагменты ДНК исключительно из этой конкретной области.

На этапе 2010 измеряют количества фрагментов ДНК, соответствующих различным размерам. Для каждого размера из множества размеров может быть измерено количество множества фрагментов ДНК из указанного биологического образца, соответствующих указанному размеру. Например, может быть измерено число фрагментов ДНК с длиной 140 оснований. Указанные количества могут быть сохранены в виде гистограммы. Согласно одному варианту реализации измеряют размер каждой из множества нуклеиновых кислот из указанного биологического образца, что может быть реализовано на индивидуальной основе (например, путем мономолекулярного секвенирования целой молекулы или исключительно концов молекулы) или на групповой основе (например, посредством электрофореза). Размеры могут соответствовать диапазону размеров. Соответственно, количество может относиться к фрагментам ДНК, имеющим размер, попадающий в определенный диапазон. При проведении секвенирования спаренных концов картирование (выравнивание) фрагментов ДНК (определенных по парным ридам последовательностей) по конкретной области может быть использовано для определения уровня метилирования указанной области.

На этапе 2020 рассчитывают первое значение первого параметра на основе количества фрагментов ДНК различных размеров. Согласно одному аспекту первый параметр обеспечивает статистическую меру профиля размеров (например, гистограмму) фрагментов ДНК в биологическом образце. Указанный параметр может называться размерным параметром, поскольку он определяется на основании размеров множества фрагментов ДНК.

Первый параметр может принимать различные формы. Одним из параметров является процент фрагментов ДНК конкретного или попадающего в некоторый диапазон размера относительно всех фрагментов ДНК, или относительно фрагментов ДНК другого размера или из другого диапазона. Такой параметр представляет собой число фрагментов ДНК конкретного размера, разделенное на общее число фрагментов, которое можно получить из гистограммы (любой структуры данных, позволяющей подсчитать абсолютное или относительное количество фрагментов конкретных размеров). В другом примере параметр может представлять собой число фрагментов конкретного размера или относящихся к определенному диапазону, разделенное на число фрагментов другого размер или из другого диапазона. Деление может служить в качестве нормировки для учета разного числа анализируемых фрагментов ДНК для разных образцов. Нормировка может осуществляться путем анализа одинакового числа фрагментов ДНК для каждого образца, что эффективно обеспечивает тот же результат, что и деление на общее число анализируемых фрагментов. Дополнительные примеры параметров и анализа размеров можно найти в заявке на патент США 13/789553, которая включена в настоящий документ посредством ссылки во всех отношениях.

На этапе 2030 сравнивают первое значение размера с референсным значением размера. Референсное значение размера может быть рассчитано по фрагментам ДНК референсного образца. Для определения референсных значений размера может быть рассчитан и количественно определен профиль метилирования для референсного образца, а также значение первого размерного параметра. Соответственно, при сравнении первого значения размера и референсного значения размера может быть определен уровень метилирования.

На этапе 2040 определяют уровень метилирования на основании указанного сравнения. Согласно одному варианту реализации можно определить, выше или ниже первое значение первого параметра, чем референсное значение размера, и таким образом определить, выше или ниже уровень метилирования тестируемого образца, чем уровень метилирования для референсного значения размера. Согласно другому варианту реализации сравнение осуществляют путем введения первого значения в калибровочную функцию. Калибровочная функция позволяет эффективно сравнивать первое значение с калибровочными значениями (набором референсных значений размеров) путем идентификации точки на кривой, соответствующей первому значению. Затем получают расчетный уровень метилирования в качестве выходного значения калибровочной функции.

Соответственно, можно откалибровать размерный параметр по уровню

метилирования. Например, уровень метилирования может быть измерен и связан с конкретным размерным параметром для указанного образца. Затем точки данных для различных образцов могут быть аппроксимированы калибровочной функцией. Согласно одному варианту реализации для разных подмножеств ДНК могут применяться разные калибровочные функции. Соответственно, может применяться некоторая форма калибровки, основанная на предварительном знании зависимости между метилированием и размером для конкретного подмножества ДНК. Например, калибровка для плодной и материнской ДНК может быть разной.

Как было показано выше, плацента более гипометилирована по сравнению с материнской кровью, и, соответственно, плодная ДНК имеет меньший размер ввиду меньшей степени метилирования. Соответственно, для определения плотности метилирования может использоваться средний размер фрагментов образца (или другая статистическая величина). Так как размеры фрагментов могут быть измерены с применением секвенирования спаренных концов, а не потенциально более сложного технически чувствительного к метилированию секвенирования, указанный способ может быть потенциально более экономичным при клиническом применении. Указанный способ может применяться для мониторинга изменений метилирования, связанных с прогрессом беременности, или связанных с беременностью расстройств, таких как преэклампсия, преждевременные роды и расстройства плода (например, вызванные хромосомными или генетическими аномалиями или внутриутробной задержкой роста).

Согласно другому варианту реализации указанный способ может применяться для детектирования и мониторинга ракового заболевания. Например, при успешном лечении ракового заболевания профиль метилирования в плазме или другой жидкости организма по оценке с применением указанного основанного на размерах способа будет изменяться в направлении профиля здоровых индивидуумов, у которых отсутствует раковое заболевание. И напротив, в том случае, если раковое заболевание прогрессирует, профиль метилирования в плазме или другой жидкости организма будет расходиться с профилем здоровых индивидуумов, у которых отсутствует раковое заболевание.

В целом, гипометилированные молекулы в плазме были короче гиперметилированных. Указанная тенденция наблюдалась и для плодных, и для материнских молекул ДНК. Поскольку известно, что метилирование ДНК влияет на упаковку нуклеосом, данные авторов настоящего изобретения предполагают, что, возможно, гипометилированные молекулы ДНК были менее плотно упакованы гистонами и были, таким образом, более восприимчивы к ферментативному расщеплению. С другой стороны, данные, представленные на фиг. 18А и 18В, также показывают, что несмотря на

значительно большее гипометилирование плодной ДНК по сравнению в материнскими ридами, распределения размеров плодной и материнской ДНК не полностью обособлены друг от друга. На фиг. 19В можно видеть, что даже в одной размерной категории уровни метилирования специфичных для плода и для матери ридов отличаются друг от друга. Указанное наблюдение предполагает, что гипометилированное состояние плодной ДНК – не единственный фактор, который обуславливает ее относительно небольшую длину по сравнению с материнской ДНК.

VII. Импринтинговый статус генных локусов

В материнской плазме могут быть детектированы происходящие из плода молекулы ДНК, имеющие тот же генотип, что и мать, но другие эпигенетические сигнатуры (LLM Poon et al. 2002 Clin Chem; 48: 35-41). Чтобы продемонстрировать, что способ с применением секвенирования является чувствительным способом выбора происходящих от плода молекул ДНК в материнской плазме, авторы настоящего изобретения использовали ту же стратегию для детекции импринтинговых плодных аллелей в образце материнской плазмы. Идентифицировали две геномных импринтинговых области: H19 (chr11:1 977 419–1 977 821, версия NCBI 36/hg18) и MEST (chr7:129 917 976–129 920 347, версия NCBI 36/hg18). Обе содержат информативные ОНП для различения материнских и плодных последовательностей. По H19, экспрессируемому у матери гену, мать была гомозиготна (A/A) и плод был гетерозиготен (A/C) по ОНП rs2071094 (chr11:1 977 740) в указанной области. Один из материнских А аллелей был полностью метилирован, а другой не метилирован. В плаценте при этом аллель А был не метилирован, тогда как унаследованный от отца аллель С был полностью метилирован. Авторы настоящего изобретения детектировали два метилированных рида с генотипом С, соответствующих импринтинговым отцовским аллелям, происходящим из плаценты, в материнской плазме.

MEST, также известный как PEG1, представляет собой экспрессируемый у отца ген. И мать, и плод были гетерозиготны (A/G) по ОНП rs2301335 (chr7:129 920 062) в составе указанного импринтингового локуса. В материнской крови аллель G был метилирован, а аллель А был не метилирован. Паттерн метилирования в плаценте был противоположным и включал метилированный материнский А аллель и неметилированный отцовский аллель G. В материнской плазме можно было детектировать три неметилированных аллеля G, которые происходили от отца. Напротив, VAV1, для неимпринтингового генного локуса на хромосоме 19 (chr19:6 723 621–6 724 121) не наблюдалось какого-либо аллельного паттерна метилирования ДНК в образцах ткани и плазмы.

Таким образом, статус метилирования может применяться для определения того, какие фрагменты ДНК принадлежат плоду. Например, само по себе обнаружение аллеля А в материнской плазме не может использоваться в качестве плодного маркера, если мать гетерозиготна (GA). Однако если наблюдается отличный статус метилирования молекул А в плазме, метилированные молекулы А специфичны для плода, а неметилированные молекулы А специфичны для матери, или наоборот.

Далее авторы настоящего изобретения фокусировались на локусах, которые были описаны как демонстрирующие геномный импринтинг в плацентарных тканях. На основе списка локусов, представленного Woodfine et al. (2011 Epigenetics Chromatin; 4: 1), авторы настоящего изобретения дополнительно отбирали такие локусы, которые содержали ОНП в области контроля импринтинга. Четыре локуса удовлетворяли указанным критериям; указанные локусы были представлены H19, KCNQ10T1, MEST и NESP.

Что касается ридов из образца клеток материнской крови для H19 и KCNQ10T1, материнские риды были гомозиготны по ОНП и доли метилированных и неметилированных ридов были приблизительно равны. Образцы CVS и ткани зрелой плаценты показали, что плод был гетерозиготен по обоим локусам, и каждый аллель был строго либо метилирован, либо неметилирован, т.е. наблюдалось моноаллельное метилирование. В образцах материнской плазме были обнаружены унаследованные от отца молекулы плодной ДНК для обоих локусов. Для H19 унаследованные от отца молекулы были представлены ридами секвенирования, которые содержали специфичный для плода аллель и были метилированы. Для KCNQ10T1 унаследованные от отца молекулы были представлены ридами секвенирования, которые содержали специфичный для плода аллель и не были метилированы.

С другой стороны, мать была гетерозиготна как по MEST, так и по NESP. В случае MEST и мать, и плод были гетерозиготны (GA) по ОНП. При этом, что очевидно из данных для цепи 5'-3' для клеток материнской крови и плацентарной ткани, статус метилирования для CpG, смежного с ОНП, был у матери и плода противоположным. Аллель А был неметилирован в ДНК матери, но метилирован в ДНК плода. В случае MEST материнский аллель был метилирован. Соответственно, можно точно определить, что плод унаследовал аллель А от матери (метилирован в CVS), а мать унаследовала аллель А от своего отца (неметилирован в клетках материнской крови). Интересно, что в образцах материнской плазмы все четыре группы молекул можно легко различить, включая каждый из двух аллелей матери и каждый из двух аллелей плода. Таким образом, комбинируя генотипическую информацию со статусом метилирования в импринтинговых локусах, авторы настоящего изобретения могли легко отличить унаследованные от матери

молекулы плодной ДНК от фоновых материнских молекул ДНК (LLM Poon et al. 2002 Clin Chem; 48: 35-41).

Указанный способ может применяться для детекции однородительской дисомии. Например, если известно, что отец данного плода гомозиготен по аллелю G, невозможность детектировать неметилированный аллель G в материнской плазме означает отсутствие вклада отцовского аллеля. Кроме того, в таких обстоятельствах, если в плазме при указанной беременности были детектированы и метилированный аллель G, и метилированный аллель A, у плода предположительно наблюдается материнская гетеродисомия, т.е. наследование двух разных аллелей от матери без наследования от отца. Как вариант, если в материнской плазме были детектированы и метилированный аллель A (плодный аллель, унаследованный от матери) и неметилированный аллель A (материнский аллель, унаследованный матерью от своего отца) без неметилированного аллеля G (отцовский аллель, который должен был быть унаследован плодом), у плода предположительно наблюдается материнская изодисомия, т.е. наследование двух идентичных аллелей от матери без наследования от отца.

В случае NESP мать была гетерозиготна (GA) по ОНП, а плод был гомозиготен по аллелю G. Отцовский аллель был метилирован по NESP. В образцах материнской плазмы унаследованные от отца плодные аллели G, которые были метилированы, могут быть легко отличены от фоновых материнских аллелей G, которые были не метилированы.

VIII. Раковые заболевания/доноры

Некоторые варианты реализации могут применяться для детекции, скрининга, мониторинг (например, рецидива, ремиссии или ответа (например, присутствия или отсутствия ответа) на лечение), стадирования, классификации (например, для содействия при выборе наиболее целесообразного метода лечения) и прогнозирования раковых заболеваний с применением анализа метилирования циркулирующей ДНК плазмы/сыворотки.

Известно, что в раковой ДНК наблюдается aberrантное метилирование (JG Herman et al. 2003 N Engl J Med; 349: 2042-2054). Например, CpG-островковые промоторы генов, например, генов-супрессоров опухолей, гиперметилированы, тогда как CpG-сайты в теле гена гипометилированы по сравнению с не являющимися раковыми клетками. При условии, что профиль метилирования раковых клеток может быть представлен профилем метилирования происходящих из опухоли молекул ДНК плазмы с применением описанных в настоящем документе способов, авторы настоящего изобретения предполагают, что общий профиль метилирования в плазме будет отличаться у индивидуумов, больных раковым заболеванием, по сравнению со здоровыми

индивидуумами, у которых отсутствует раковое заболевание, или по сравнению с индивидуумами, излечившимися от ракового заболевания. Типы различий профиля метилирования могут относиться к количественным различиям плотностей метилирования генома и/или плотностей метилирования сегментов геномов. Например, в плазме страдающих раковым заболеванием пациентов будет наблюдаться уменьшение плотностей метилирования метилома плазмы или сегментов генома ввиду общей гипометилированной природы ДНК раковых тканей (Gama-Sosa MA et al. 1983 *Nucleic Acids Res*; 11: 6883-6894).

Качественные изменения профиля метилирования также должны быть отражены в данных метилома плазмы. Например, молекулы ДНК плазмы, происходящие из генов, которые гиперметилированы исключительно в раковых клетках, будут демонстрировать гиперметилирование в плазме пациента, страдающего раковым заболеванием, относительно молекул ДНК плазмы, происходящих из тех же генов, но присутствующих в образце от здорового контроля. Поскольку aberrантное метилирование происходит при большинстве раковых заболеваний, описанные в настоящем документе способы могут применяться для детекции всех форм злокачественных новообразований с aberrантным метилированием, например, не ограничиваясь перечисленными, злокачественных новообразований в легком, молочной железе, ободочной и прямой кишки, предстательной железы, носоглотки, желудка, яичек, кожи, нервной системы, костей, яичника, печени, гематологических тканей, поджелудочной железы, матки, почек, мочевого пузыря, лимфоидных тканей и т.д. Злокачественные новообразования могут быть представлены разнообразными гистологическими подтипами, такими как карциномы, аденокарциномы, саркомы, фиброаденокарцинома, нейроэндокринные и недифференцированные, и т.д.

С другой стороны, авторы настоящего изобретения предполагают, что происходящие из опухоли молекулы ДНК можно отличить от фоновых происходящих не из опухоли молекул ДНК, так как общий небольшой размерный профиль происходящей из опухоли ДНК более выражен у молекул ДНК, происходящих из локусов с ассоциированным с опухолью aberrантным гипометилированием, которое должно оказывать дополнительный эффект на размер молекулы ДНК. Также происходящие из опухоли молекулы ДНК плазмы можно отличить от фоновых происходящих не из опухоли молекул ДНК плазмы с применением нескольких характерных особенностей, связанных с опухолевой ДНК, включая, но не ограничиваясь перечисленными, однонуклеотидные вариации, увеличение и уменьшение числа копий, транслокации, инверсии, aberrантное гипер- или гипометилирование и размерный профиль. Так как все перечисленные изменения могут происходить независимым образом, комплексное

использование указанных признаков может обеспечить дополнительное преимущество для чувствительной и специфичной детекции раковой ДНК в плазме.

A. Размер и раковые заболевания

Размер происходящих из опухоли ДНК молекул в плазме также напоминает размеры моноклеосомных единиц; указанные молекулы короче фоновых происходящих не из опухоли молекул ДНК, которые также присутствуют в плазме страдающих раковым заболеванием пациентов. Было показано, что размерные параметры коррелируют с раковыми заболеваниями согласно описанию в заявке на патент США 13/789553, включенной в настоящий документ посредством ссылки во всех отношениях.

Поскольку продемонстрирована зависимость между размером и статусом метилирования молекулы и для происходящей от плода, и для происходящей от матери ДНК в плазме, ожидается, что для происходящих из опухоли молекул ДНК будет наблюдаться та же тенденция. Например, гипометилированные молекулы будут короче гиперметилированных молекул в плазме страдающих раковым заболеванием пациентов или субъектов, у которых проводится скрининг на раковые заболевания.

B. Плотности метилирования разных тканей у больных раковыми заболеваниями пациентов

В данном примере авторы настоящего изобретения исследовали образцы плазмы и ткани пациента с гепатоцеллюлярной карциномой (ГЦК). Образцы крови получали от указанного пациента с ГЦК до хирургического удаления опухоли и через 1 неделю после хирургического удаления опухоли. Плазму и лейкоцитарную пленку собирали после центрифугирования образцов крови. Собирали резецированную опухоль и соседнюю неопухолевую ткань печени. Образцы ДНК, экстрагированные из образцов плазмы и ткани, исследовали с применением массивно-параллельного секвенирования с предварительной обработкой бисульфитом и без предварительной обработки бисульфитом. Также исследовали ДНК плазмы четырех здоровых индивидуумов без раковых заболеваний в качестве контролей. Обработка бисульфитом ДНК преобразует неметилированные остатки цитозина в урацил. В ходе последующих полимеразной цепной реакции и секвенирования указанные остатки урацила будут вести себя как тимидин. С другой стороны, обработка бисульфитом не приводит к конверсии метилированных остатков цитозина в урацил. После массивно-параллельного секвенирования ряды последовательностей исследовали с помощью Methy-Pipe (P Jiang, et al. Methy-Pipe: An integrated bioinformatics data analysis pipeline for whole genome methylome analysis («Methy-Pipe: комплексная система анализа данных биоинформатики для полногеномного анализа метилома»), доклад представлен на Международной

практической конференции по биоинформатике и биомедицине Международного института инженеров электротехники и электроники (IEEE International Conference on Bioinformatics and Biomedicine Workshops), Гонконг, 18–21 декабря 2010 г.), для определения статуса метилирования остатков цитозина во всех положениях с динуклеотидами CG, т.е. в CpG-сайтах.

Фиг. 21А представляет собой таблицу 2100, где приведены плотности метилирования предоперационных образцов плазмы и ткани пациента с ГЦК. Плотность метилирования CpG для представляющих интерес областей (например, CpG-сайтов, промотора или повторяющихся областей, и т.д.) относится к доле ридов, демонстрирующих метилирование CpG, в общем количестве ридов, покрывающих геномные динуклеотиды CpG. Плотности метилирования лейкоцитарной пленки и неопухолевой ткани печени аналогичны. Общая плотность метилирования опухолевой ткани, на основе данных для всех аутосом, была на 25% ниже, чем плотность метилирования лейкоцитарной пленки и неопухолевой ткани печени. Гипометилирование было стабильным для каждой индивидуальной хромосомы. Плотность метилирования плазмы находилась между значениями плотности для незлокачественных тканей и раковых тканей. Указанное наблюдение согласуется с тем фактом, что и раковые, и не являющиеся раковыми ткани вносят вклад в циркулирующую ДНК пациента с раковым заболеванием. Было показано, что гемопоэтическая система является главным источником циркулирующей ДНК у индивидуумов без активных злокачественных состояний (YUN Lui, et al. 2002 Clin Chem; 48: 421-7). Поэтому авторы настоящего изобретения также исследовали образцы плазмы, полученные от четырех здоровых контролей. Число ридов секвенирования и достигнутая глубина секвенирования на образец приведены в таблице 2150 на фиг. 21В.

Фиг. 22 представляет собой таблицу 220, где приведены плотности метилирования в аутосомах, варьирующие от 71,2% до 72,5% в образцах плазмы здоровых контролей. Эти данные показывают ожидаемый уровень метилирования ДНК в образцах плазмы, полученных от индивидуумов, у которых отсутствует источник опухолевой ДНК. У больного раковым заболеванием пациента опухолевая ткань также высвобождает ДНК в кровотоки (КСА Chan et al. 2013 Clin Chem; 59: 211-224); RJ Leary et al. 2012 Sci Transl Med; 4: 162ra154). Из-за гипометилированной природы опухоли ГЦК присутствие как происходящей из опухоли, так и происходящей не из опухоли ДНК в предоперационной плазме пациента приводило к уменьшению плотности метилирования относительно уровней в плазме здоровых контролей. Фактически, плотность метилирования в образце предоперационной плазмы находилась между значениями плотностей метилирования

опухолевой ткани и плазмы здоровых контролей. Причина заключается в том, что на уровень метилирования ДНК плазмы страдающих раковым заболеванием пациентов влияет степень aberrантного метилирования, в указанном случае – гипометилирования, опухолевой ткани и фракционная концентрация происходящей из опухоли ДНК в кровотоке. Меньшая плотность метилирования опухолевой ткани и более высокая фракционная концентрация происходящей из опухоли ДНК в кровотоке приводят к меньшей плотности метилирования ДНК плазмы у больного раковым заболеванием пациента. Большинство опухолей, как сообщается, демонстрируют глобальное гипометилирование (JG Herman et al. 2003 N Engl J Med; 349: 2042-2054; MA Gama-Sosa et al. 1983 Nucleic Acids Res; 11: 6883-6894). Соответственно, данные наблюдения для образцов ГЦК должны быть также применимы к другим типам опухолей.

Согласно одному варианту реализации плотность метилирования ДНК плазмы может использоваться для определения фракционной концентрации происходящей из опухоли ДНК в образце плазмы/сыворотки, если известен уровень метилирования опухолевой ткани. Уровень метилирования, например, плотность метилирования опухолевой ткани, может быть определен, если доступен образец опухоли или биоптат опухоли. Согласно другому варианту реализации информация относительно уровня метилирования опухолевой ткани может быть получена при изучении уровня метилирования в группе опухолей аналогичного типа, и указанную информацию (например, среднее значение уровня или медиана уровня) используют для исследуемого пациента с применением описанной в указанном изобретении технологии. Уровень метилирования опухолевой ткани может быть определен посредством анализа опухолевой ткани пациента или выведен по анализу опухолевых тканей других пациентов с раковым заболеванием того же или аналогичного типа. Метилирование опухолевых тканей может быть определено с применением ряда чувствительных к метилированию платформ, включая, но не ограничиваясь перечисленными, массивно-параллельное секвенирование, одномолекулярное секвенирование, микроматрицы (например, олигонуклеотидные матрицы), или масс-спектрометрия (например, анализ EpiTyper от Sequenom Inc.). Согласно некоторым вариантам реализации такой анализ может предваряться процедурами, чувствительными к статусу метилирования молекул ДНК, включая, но не ограничиваясь перечисленным, иммунопреципитация цитозина и расщепление чувствительными к метилированию рестрикционными ферментами. Если известен уровень метилирования опухоли, после анализа метилома плазмы может быть рассчитана фракционная концентрация опухолевой ДНК в плазме страдающих раковым заболеванием пациентов.

Зависимость между уровнем метилирования в плазме, P , с фракционной концентрацией опухолевой ДНК, f , и уровнем метилирования опухолевой ткани, TUM , может быть описана как: $P = BKG \times (1-f) + TUM \times f$, где BKG представляет собой фоновый уровень метилирования ДНК в плазме, полученной из клеток крови и других внутренних органов. Например, было показано, что общая плотность метилирования всех аутосом составляет 42,9% в биоптате опухолевой ткани, полученном от указанного пациента с ГЦК, т.е. это значение TUM для указанного случая. Среднее значение плотности метилирования образцов плазмы от четырех здоровых контролей составляло 71,6%, т.е. значение BKG для указанного случая. Плотность метилирования в плазме для предоперационной плазмы составляла 59,7%. Используя указанные значения, определяют, что f составляет 41,5%.

Согласно другому варианту реализации уровень метилирования опухолевой ткани может быть установлен неинвазивным образом на основе данных о метиломе плазмы, если известна фракционная концентрация происходящей из опухоли ДНК в образце плазмы. Фракционная концентрация происходящей из опухоли ДНК в образце плазмы может быть определена посредством другого генетического анализа, например, полногеномного анализа потери аллелей (GAAL) и анализа однонуклеотидных мутаций согласно более раннему описанию (заявка на патент США 13/308473; KCA Chan et al. 2013 Clin Chem; 59: 211-24). Указанный расчет основан на тех же взаимосвязях, что и описанные выше, за исключением того, что согласно указанному варианту реализации значение f известно, а значение TUM становится неизвестным. Указанное выведение может осуществляться для полного генома или для частей генома, аналогично данным, рассмотренным в контексте определения уровня метилирования плацентарной ткани по данным материнской плазмы.

Согласно другому варианту реализации может использоваться вариативность между участками или профилями плотностей метилирования для различения субъектов, имеющих раковые заболевания и не имеющих раковых заболеваний. Разрешение анализа метилирования может быть дополнительно увеличено путем деления генома на участки конкретного размера, например, 1 Мб. Согласно подобному варианту реализации рассчитывали плотность метилирования каждого участка размером 1 Мб для собранных образцов, например, лейкоцитарной пленки, ткани резецированной ГЦК, неопухолевой ткани печени, смежной с опухолью и плазмы, собранной до и после удаления опухоли. Согласно другому варианту реализации размеры участков не обязательно должны быть постоянными. Согласно одному варианту реализации число CpG-сайтов остается постоянным в составе каждого участка, при этом размер самого участка может

варьировать.

На фиг. 23А и 23В представлена плотность метилирования лейкоцитарной пленки, опухолевой ткани, неопухолевой ткани печени, предоперационной плазмы и послеоперационной плазмы пациента с ГЦК. Фиг. 23А представляет собой график 2300 результатов для хромосомы 1. Фиг. 23В представляет собой график 2350 результатов для хромосомы 2.

Для большинства участков размером 1 Мб плотности метилирования лейкоцитарной пленки и неопухолевой ткани печени, смежной с опухолью, были аналогичны, тогда как плотности метилирования опухолевых тканей были более низкими. Значения плотности метилирования предоперационной плазмы находятся между значениями для опухоли и незлокачественных тканей. Плотности метилирования исследуемых геномных областей в опухолевых тканях могут быть выведены с применением данных о метилировании предоперационной плазмы и фракционной концентрации опухолевой ДНК. Указанный способ идентичен описанному выше описанию способу с использованием значений плотности метилирования для всех аутосом. Описанное выведение значений метилирования опухоли может также проводиться с применением указанных данных более высокого разрешения относительно метилирования ДНК плазмы. Также могут использоваться участки другого размера, например, 300 кб, 500 кб, 2 Мб, 3 Мб, 5 Мб или более чем 5 Мб. Согласно одному варианту реализации размеры участков не обязательно должны быть постоянными. Согласно одному варианту исполнения число CpG-сайтов остается постоянным в составе каждого участка, при этом размер самого участка может варьировать.

С. Сравнение плотности метилирования в плазме больного раковым заболеванием пациента и здоровых индивидуумов

Как видно из 2100, плотности метилирования ДНК предоперационной плазмы у больного раковым заболеванием пациента были ниже, чем плотности метилирования незлокачественных тканей. Это предположительно обусловлено присутствием ДНК из опухолевой ткани, которая гипометилирована. Указанная более низкая плотность метилирование ДНК плазмы потенциально может использоваться в качестве биомаркера для детекции и мониторинга раковых заболеваний. При мониторинге ракового заболевания в том случае, если рак прогрессирует, со временем будет наблюдаться повышенное количество происходящей из ракового новообразования ДНК в плазме. Согласно указанному примеру повышенное количество циркулирующей происходящей из ракового новообразования ДНК в плазме будет приводить к последующему снижению плотности метилирования ДНК в плазме в полногеномном масштабе.

И напротив, если раковое заболевание реагирует на лечение, количество происходящей из ракового новообразования ДНК в плазме будет уменьшаться со временем. Согласно указанному примеру уменьшение количества происходящей из ракового новообразования ДНК в плазме приведет к увеличению плотности метилирования ДНК в плазме. Например, если проводилось лечение пациента с раком легкого с мутацией рецептора эпидермального фактора роста с применением таргетной терапии, например, ингибирования тирозинкиназ, увеличение плотности метилирования ДНК в плазме свидетельствует об ответе. Последующее появление опухолевого клона, устойчивого к ингибиторам тирозинкиназы, будет связано с уменьшением плотности метилирования ДНК в плазме, что может указывать на рецидив.

Измерения плотности метилирования в плазме могут осуществляться серийно; при таких измерениях могут быть рассчитаны скорости изменения и использованы для прогнозирования или корреляции с клиническим прогрессированием или ремиссией, или прогнозом. Для выбранных геномных локусов, гиперметилированных в раковых тканях, но гипометилированных в нормальных тканях, например, промоторных областей некоторых генов-супрессоров опухолей, зависимость между прогрессированием рака и благоприятным ответом на лечение будет обратной относительно паттернов, описанных выше.

Чтобы продемонстрировать осуществимость указанного способа, авторы настоящего изобретения сравнивали плотности метилирования ДНК образцов плазмы, полученных от больного раковым заболеванием пациента до и после хирургического удаления опухоли, с ДНК плазмы, полученной от четырех здоровых контрольных субъектов.

В таблице 2200 приведены плотности метилирования ДНК каждой аутосомы и объединенные значения для всех аутосом предоперационных и послеоперационных образцов плазмы больного раковым заболеванием пациента, и объединенные значения для всех аутосом четырех здоровых контрольных субъектов. Для всех хромосом плотности метилирования ДНК в образце предоперационной плазмы были ниже, чем плотности в послеоперационном образце и образцах плазмы четырех здоровых субъектов. Различие плотностей метилирования ДНК в плазме между предоперационными и послеоперационными образцами представляет собой доказательство того, что более низкие плотности метилирования в образце предоперационной плазмы были обусловлены присутствием ДНК из опухоли ГЦК.

Возвращение плотностей метилирования ДНК в образце послеоперационной плазмы к уровням, аналогичным уровням в образцах плазмы здоровых контролей,

предполагает, что значительная часть происходящей из опухоли ДНК исчезла благодаря хирургическому удалению источника, т.е. опухоли. Эти данные предполагают, что плотность метилирования предоперационной плазмы, определенная с применением доступных для обширных геномных областей данных, таких как все аутосомы или индивидуальные хромосомы, находилась на более низком уровне, чем плотность метилирования у здоровых контролей, что позволяет идентифицировать тестируемый случай как раковое заболевание, т.е. установить диагноз или провести скрининг.

Данные для предоперационной плазмы также показали значительно более низкий уровень метилирования, чем в послеоперационной плазме, что указывает на то, что уровень метилирования в плазме также можно использовать для мониторинга опухолевой нагрузки и, соответственно, для прогнозирования и мониторинга прогрессирования ракового заболевания у пациента. Референсные значения можно определить по плазме здоровых контролей или индивидуумов, у которых имеется риск развития ракового заболевания, но на текущий момент не имеющих раковых заболеваний. Индивидуумы, у которых имеется риск развития ГЦК, включают страдающих хроническим гепатитом В или инфицированных гепатитом С, индивидуумов с гемохроматозом и индивидуумов с циррозом печени.

Значения плотности метилирования в плазме за пределами, например, ниже указанной точки отсечения, основанной на референсных значениях, может применяться для оценки наличия или отсутствия в плазме небеременного индивидуума опухолевой ДНК. Для детекции присутствия гипометилированной циркулирующей опухолевой ДНК точка отсечения может быть определена как уровень ниже 5-го или 1-го перцентиля значений в контрольной популяции, или может быть определена на основе нескольких стандартных отклонений, например, 2 или 3 стандартных отклонений (SD), ниже средних значений плотности метилирования у контролей, или на основе нескольких медианных значений (MoM). Для гиперметилированной опухолевой ДНК точка отсечения может быть определен как уровень выше 95-го или 99-го перцентиля значений в контрольной популяции, или может быть определена на основе нескольких стандартных отклонений, например, 2 или 3 SD выше средних значений плотности метилирования значений контролей, или на основе определения нескольких медианных значений (MoM). Согласно одному варианту реализации возраст в контрольной популяции совпадает с возрастом тестируемого субъекта. Совпадение по возрасту не обязательно должно быть точным и может быть представлено возрастными группами (например, от 30 до 40 лет, при возрасте тестируемого субъекта, составляющем 35 лет).

Затем авторы настоящего изобретения сравнили плотности метилирования

участков размером 1 Мб в образцах плазмы больного раковым заболеванием пациента и в образцах от четырех контрольных субъектов. Для наглядности представлены указанные результаты для хромосомы 1.

Фиг. 24А представляет собой график 2400, отражающий плотности метилирования предоперационной плазмы от пациента с ГЦК. Фиг. 24В представляет собой график 2450, отражающий плотности метилирования в послеоперационной плазме от пациента с ГЦК. Голубые точки представляют результаты для контрольных субъектов, красными точками обозначены результаты для образца плазмы от пациента с ГЦК.

Как видно на фиг. 24А, плотности метилирования в предоперационной плазме от пациента с ГЦК были ниже, чем плотности метилирования контрольных субъектов для большинства участков. Аналогичные паттерны наблюдались для других хромосом. Как видно на фиг. 24В, плотности метилирования в послеоперационной плазме от пациента с ГЦК были аналогичны плотностям метилирования у контрольных субъектов для большинства участков. Аналогичные паттерны наблюдались для других хромосом.

Для оценки того, имеется ли раковое заболевание у тестируемого субъекта, результат для тестируемого субъекта сравнивают со значениями для референсной группы. Согласно одному варианту реализации референсная группа может быть составлена несколькими здоровыми субъектами. Согласно другому варианту реализации референсная группа может быть составлена субъектами с незлокачественными состояниями, например, инфекцией хронического гепатита В или циррозом. Затем может быть количественно определено различие плотностей метилирования между тестируемым субъектом и референсной группой.

Согласно одному варианту реализации референсный диапазон может быть определен по значениям контрольной группы. Затем могут быть использованы отклонения результата для тестируемого субъекта от верхнего или нижнего предела для референсной группы для определения того, имеется ли у субъекта опухоль. На эту величину будет влиять фракционная концентрация происходящей из опухоли ДНК в плазме и различие уровня метилирования между злокачественными и незлокачественными тканями. Более высокая фракционная концентрация происходящей из опухоли ДНК в плазме приводит к большим различиям плотности метилирования между тестируемым образцом плазмы и контролями. Большая степень различия уровней метилирования злокачественных и незлокачественных тканей также связана с большими различиями плотности метилирования между тестируемым образцом плазмы и контролями. Согласно еще одному варианту реализации для тестируемых субъектов из разных возрастных диапазонов выбирают разные референсные группы.

Согласно другому варианту реализации рассчитывали среднее значение и SD плотностей метилирования четырех контрольных субъектов для каждого участка размером 1 Мб. Затем для соответствующих участков рассчитывали разность между плотностями метилирования пациента с ГЦК и средним значением для контрольных субъектов. Согласно одному варианту реализации указанную разность далее делили на SD соответствующего участка для определения Z-показателя. Другими словами, Z-показатель представляет разность плотностей метилирования тестируемого и контрольного образцов плазмы, выраженную в виде нескольких SD средних значений для контрольных субъектов. Z-показатель >3 для участка означает, что ДНК плазмы пациента с ГЦК более гиперметилирована, чем у контрольных субъектов, более чем на 3 SD на указанном участке, тогда как Z-показатель < -3 на участке означает, что ДНК плазмы пациента с ГЦК более гипометилирована, чем у контрольных субъектов, более чем на 3 SD на указанном участке.

На фиг. 25А и 25В представлены Z-показатели плотностей метилирования ДНК плазмы для предоперационных (график 2500) и послеоперационных (график 2550) образцов плазмы пациента с ГЦК с применением данных о метиломе плазмы четырех здоровых контрольных субъектов в качестве эталона для хромосомы 1. Каждая точка представляет результат для одного участка размером 1 Мб. Черными точками обозначены участки с Z-показателем в диапазоне от -3 до 3. Красными точками обозначены участки с Z-показателем < -3 .

Фиг. 26А представляет собой таблицу 2600, где приведены данные Z-показателей для предоперационной и послеоперационной плазмы. Большинство участков на хромосоме 1 (80,9%) в образце предоперационной плазмы имели Z-показатель < -3 , указывающий на то, что ДНК предоперационной плазмы пациента с ГЦК была значительно более гипометилирована по сравнению с контрольными субъектами. С другой стороны, число красных точек существенно снижалось в образце послеоперационной плазмы (8,3% участков на хромосоме 1), предполагая, что большая часть опухолевой ДНК была устранена из кровотока благодаря хирургическому удалению источника циркулирующей опухолевой ДНК.

Фиг. 26В представляет собой Circos-график 2620, демонстрирующий Z-показатель плотностей метилирования ДНК плазмы для предоперационных и послеоперационных образцов плазмы пациента с ГЦК с использованием четырех здоровых контрольных субъектов в качестве эталона для участков размером 1 Мб, проанализированных для всех аутосом. Внешний круг представляет идеограммы аутосом человека. Средний круг представляет данные для образца предоперационной плазмы. Внутренний круг

представляет указанные данные для образца послеоперационной плазмы. Каждая точка представляет результат для одного участка размером 1 Мб. Черными точками обозначены участки с Z-показателями от -3 до 3. Красными точками обозначены участки с Z-показателями <-3. Зелеными точками обозначены участки с Z-показателями >3.

Фиг. 26С представляет собой таблицу 2640, демонстрирующую распределение Z-показателей участков размером 1 Мб для полного генома как в предоперационных, так и в послеоперационных образцах плазмы пациента с ГЦК. Эти результаты показывают, что ДНК предоперационной плазмы пациента с ГЦК была более гипометилирована, чем у контролей, для большинства областей (85,2% участков размером 1 Мб) во все геноме. С другой стороны, большинство областей (93,5% участков размером 1 Мб) в образце послеоперационной плазмы не демонстрировали значительного гиперметилирования или гипометилирования по сравнению с контролями. Эти данные показывают, что большая часть опухолевой ДНК, по своей природе главным образом гипометилированной для данной ГЦК, больше не присутствовала в образце послеоперационной плазмы.

Согласно одному варианту реализации число, процент или доля участков с Z-показателями <-3 может применяться для определения наличия ракового заболевания. Например, как видно из таблицы 2640, 2330 из 2734 проанализированных участков (85,2%) демонстрировали Z-показатели <-3 в предоперационной плазме, но при этом только 171 из 2734 проанализированных участков (6,3%) демонстрировали Z-показатели <-3 в послеоперационной плазме. Эти данные показывают, что нагрузка опухолевой ДНК в предоперационной плазме была значительно выше, чем в послеоперационной плазме.

Предельные значения числа участков могут быть определены с применением статистических методов. Например, приблизительно 0,15% участков, как ожидается, будут иметь Z-показатель <-3 на основе нормального распределения. Таким образом, точка отсечения для числа участков может составлять 0,15% от общего числа анализируемых участков. Другими словами, если в образце плазмы небеременного индивидуума наблюдается более чем 0,15% участков с Z-показателями <-3, имеется источник гипометилированной ДНК в плазме, а именно, раковое заболевание. Например, 0,15% из 2734 участков размером 1 Мб, которые авторы настоящего изобретения исследовали в данном примере, составляет приблизительно 4 участка. При использовании указанной величины в качестве точки отсечения образцы и предоперационной, и послеоперационной плазмы содержали гипометилированную происходящую из опухоли ДНК, хотя ее количество значительно выше в образце предоперационной плазмы по сравнению с образцом послеоперационной плазмы. У четырех здоровых контрольных субъектов ни один из участков не показал значительного гиперметилирования или

гипометилирования. Могут применяться другие предельные значения (например, 1,1 %) и они могут варьировать в зависимости от требований используемого анализа. В других примерах процент отсека может варьировать на основании статистического распределения, а также требуемой чувствительности и приемлемой специфичности.

Согласно другому варианту реализации точка отсека может быть определена посредством анализа кривых функциональных характеристик (кривых ROC), путем анализа числа страдающих раковым заболеванием пациентов и индивидуумов без раковых заболеваний. Для дополнительной валидации специфичности указанного подхода исследовали образец плазмы от пациента, обратившегося за медицинской консультацией в связи с незлокачественным состоянием (C06). 1,1% участков имели Z-показатель < -3 . Согласно одному варианту реализации разные пороги могут применяться для классификации разных уровней статуса заболевания. Более низкий пороговый процент может применяться для различения статуса здоровья от доброкачественных состояний, и более высокий пороговый процент для различения доброкачественных состояний и злокачественных новообразований.

Диагностическая эффективность анализа гипометилирования плазмы с применением массивно-параллельного секвенирования, по-видимому, выше, чем эффективность при применении полимеразной цепной реакции (ПЦР), основанной на амплификации специфических классов повторяющихся элементов, например, длинного диспергированного ядерного элемента-1 (LINE-1) (P Tangkijvanich et al. 2007 Clin Chim Acta; 379:127-133). Одним из возможных объяснений указанного наблюдения является то, что хотя гипометилирование широко распространено в геноме опухоли, ему все же свойственна некоторая степень неоднородности в разных геномных областях.

Фактически, авторы настоящего изобретения наблюдали, что средние значения плотностей метилирования в плазме у референсных субъектов варьировали по всему геному (фиг. 56). Каждая красная точка на фиг. 56 отражает среднее значение плотности метилирования одного участка размером 1 Мб у 32 здоровых субъектов. На графике представлены все проанализированные участки размером 1 Мб по всему геному. Число в каждой рамке представляет собой номер хромосомы. Авторы настоящего изобретения наблюдали варьирование средних значений плотности метилирования от участка к участку.

Диагностический алгоритм простого анализа на основе ПЦР не позволяет учитывать такую гетерогенность между областями. Такая гетерогенность расширяет диапазон плотностей метилирования, наблюдаемых у здоровых индивидуумов. Поэтому может потребоваться более значительное снижение плотности метилирования, чтобы

образец считался демонстрирующим гипометилирование. Это приведет к снижению чувствительности теста.

Напротив, при использовании способа на основе массивно-параллельного секвенирования геном разделяется на участки размером 1 Мб (или участки других размеров) и плотности метилирования для таких участков измеряются индивидуально. Указанный способ уменьшает влияние вариаций исходных плотностей метилирования в разных геномных областях, поскольку сравнивают каждую область в тестируемом образце и в контролях. Так, для одного и того же участка вариативность между 32 здоровыми контрольными индивидуумами была относительно невелика. Для 32 здоровых контролей 95% участков имели коэффициент вариации (CV), составляющий $\leq 1,8\%$. Однако для дополнительного повышения чувствительности при детектировании связанного с раковым заболеванием гипометилирования сравнение может осуществляться по нескольким геномным областям. Чувствительность будет повышаться при тестировании нескольких геномных областей, поскольку это гарантирует отсутствие влияния биологической изменчивости, когда при тестировании всего одной области гипометилирование образца ракового новообразования для конкретной области не наблюдается.

Способ, включающий сравнение плотности метилирования эквивалентных геномных областей в контролях и тестируемых образцах (например, тестирование каждой геномной области по отдельности, и затем, возможно, объединение таких результатов) и осуществление указанного сравнения для нескольких геномных областей дает более высокое отношение сигнал/шум при детекции гипометилирования, связанного с раковым заболеванием. Указанный способ, включающий массивно-параллельное секвенирование, представлен в иллюстративных целях. Можно ожидать, что и другие методики, позволяющие определять плотности метилирования нескольких геномных областей и проводить сравнение плотностей метилирования соответствующих областей в контролях и тестируемых образцах, будут обеспечивать аналогичный эффект. Например, для достижения требуемого эффекта могут быть сконструированы зонды гибридизации или инвертированные молекулярные зонды, которые могут быть нацелены на молекулы ДНК плазмы, происходящие из конкретных геномных областей, а также определять уровень метилирования указанной области.

Согласно еще одному варианту реализации может применяться сумма Z-показателей всех участков для определения присутствия ракового заболевания или для мониторинга последовательных изменений уровня метилирования ДНК в плазме. В связи с общей гипометилированной природой опухолевой ДНК сумма Z-показателей будет ниже в плазме, полученной от индивидуума, больного раковым заболеванием, чем у

здоровых контролей. Сумма Z-показателей для образцов предоперационной и послеоперационной плазмы пациента с ГЦК составляла -49843,8 и -3132,13, соответственно.

Согласно другим вариантам реализации могут применяться другие способы для исследования уровня метилирования ДНК плазмы. Например, может быть определена доля метилированных остатков цитозина в общем содержании остатков цитозина с применением масс-спектрометрии (ML Chen et al. 2013 Clin Chem; 59: 824-832) или массивно-параллельного секвенирования. Однако так как большинство остатков цитозина находится вне контекста динуклеотидов CpG, доля метилированных цитозинов в общем количестве остатков цитозина будет относительно небольшой по сравнению с уровнями метилирования, определяемыми в контексте динуклеотидов CpG. Авторы настоящего изобретения определили уровень метилирования образцов ткани и плазмы, полученных от пациента с ГЦК, а также четырех образцов плазмы, полученной от здоровых контролей. Уровни метилирования измеряли в контексте CpG, любых цитозинов, в контекстах CHG и CHH с применением данных полногеномного массивно-параллельного секвенирования. Н относится к аденину, тимину или остаткам цитозина.

Фиг. 26D представляет собой таблицу 2660, где представлены уровни метилирования в образце опухолевой ткани и предоперационной плазмы, пересекающиеся с некоторыми из контрольных образцов плазмы в контексте CHH и CHG. Уровни метилирования в образце опухолевой ткани и предоперационной плазмы были стабильно более низкими по сравнению с лейкоцитарной пленкой, неопухолевой тканью печени, образцом послеоперационной плазмы и образцов плазмы здоровых контролей, как для CpG, так и для любых цитозинов. Однако данные на основе метилированных CpG, т.е. плотности метилирования, демонстрировали более широкий динамический диапазон, чем данные на основе метилированных цитозинов.

Согласно другим вариантам реализации статус метилирования ДНК плазмы может быть определен способами с применением антител против метилированного цитозина, например, иммунопреципитацией метилированной ДНК (MeDIP). Однако точность указанных способов, как ожидается, будет ниже, чем точность способов на основе секвенирования, из-за вариабельности связывания антител. Согласно еще одному варианту реализации может быть определен уровень 5-гидроксиметилцитозина в ДНК плазмы. При этом было обнаружено, что снижение уровня 5-гидроксиметилцитозина является эпигенетическим признаком некоторых раковых заболеваний, например, меланомы (CG Lian, et al. 2012 Cell; 150: 1135-1146).

Помимо ГЦК авторы настоящего изобретения также исследовали, возможно ли

применение указанного способа для других типов раковых заболеваний. Авторы настоящего изобретения исследовали образцы плазмы от 2 пациентов с аденокарциномой легкого (АЛ1 и АЛ2), 2 пациентов с носоглоточной карциномой (НГК1 и НГК2), 2 пациентов с раком толстой и прямой кишки (КРК1 и КРК2), 1 пациента с метастатической нейроэндокринной опухолью (НЭК1) и 1 пациента с метастатической гладкомышечной саркомой (ГМС1). ДНК плазмы указанных субъектов конвертировали бисульфитом и секвенировали с применением платформы Illumina HiSeq2000 для 50 п.о. на одном конце. Четырех здоровых контрольных субъектов, упомянутых выше, использовали в качестве референсной группы для анализа у указанных 8 пациентов. Использовали 50 п.о. ридов последовательностей на одном конце. Весь геном разделяли на участки размером 1 Мб. Среднее значение и SD плотности метилирования рассчитывали для каждого участка с применением данных референсной группы. Затем эти результаты для 8 больных раком пациентов выражали в виде Z-показателей, представляющих собой число SD среднего значения референсной группы. Положительное значение означает, что плотность метилирования в тестируемом случае ниже, чем среднее значение для референсной группы, и наоборот. Число ридов секвенирования и достигнутая глубина секвенирования на образец приведены в таблице 2780 на фиг. 27I.

На фиг. 27A-H представлен Circos-график плотности метилирования у 8 пациентов с раковыми заболеваниями в соответствии с вариантами реализации настоящего изобретения. Каждая точка представляет результат для участка размером 1 Мб. Черными точками обозначены участки с Z-показателями в диапазоне от -3 до 3. Красными точками обозначены участки с Z-показателями <-3. Зелеными точками обозначены участки с Z-показателями >3. Интервал между двумя последовательными линиями представляет разность Z-показателей, равную 20.

Значительное гипометилирование наблюдалась в нескольких областях геномов пациентов при большинстве типов раковых заболеваний, включая рак легкого, носоглоточную карциному, рак толстой и прямой кишки и метастатическую нейроэндокринную опухоль. Интересно, что, помимо гипометилирования, наблюдалась значительное гиперметилирование нескольких областей по всему геному в случае с метастатической гладкомышечной саркомой. Гладкомышечная саркома происходит из эмбриональной мезодермы, тогда как другие типы раковых заболеваний у остальных 7 пациентов имеют эктодермальное эмбриональное происхождение. Таким образом, возможно, что паттерн метилирования ДНК саркомы может отличаться от паттерна карциномы.

Как видно из указанного случая, паттерн метилирования ДНК плазмы также может

подходить для дифференциации разных типов раковых заболеваний, в указанном примере – для дифференциации карциномы и саркомы. Эти данные также предполагают, что указанный способ может использоваться для детекции aberrантного гиперметилирования, связанного со злокачественным новообразованием. Во всех указанных 8 случаях были доступны только образцы плазмы и анализ опухолевой ткани не проводился. Это показало, что даже при отсутствии предварительного профиля метилирования или уровней метилирования опухолевой ткани происходящая из опухоли ДНК может быть легко детектирована в плазме с применением описанных способов.

Фиг. 27J представляет собой таблицу 2790, отражающую распределение Z-показателей участков размером 1 Мб полного генома в плазме пациентов с разными злокачественными новообразованиями. Для каждого случая показан процент участков с Z-показателем <-3 , $-3-3$ и >3 . Более чем 5% участков имели Z-показатель <-3 для всех случаев. Таким образом, при использовании авторами настоящего изобретения для классификации образца как положительного по раковому заболеванию точки отсечения, равной 5% значительно гипометилированных участков, все указанные случаи будут классифицированы как положительные по раковому заболеванию. Полученные авторами результаты показывают, что гипометилирование предположительно представляет собой общее явление для разных типов раковых заболеваний, и анализ метилома плазмы подходит для детекции разных типов раковых заболеваний.

D. Способ

Фиг. 28 представляет собой блок-схему способа 2800 анализа биологического образца организма для установления классификации уровня рака в соответствии с вариантами реализации настоящего изобретения. Указанный биологический образец включает ДНК, происходящую из нормальных клеток, и может потенциально включать ДНК из клеток, связанных с раковым заболеванием. По меньшей мере часть ДНК в биологическом образце может быть свободной от клеток.

На этапе 2810 анализируют множество молекул ДНК из указанного биологического образца. Анализ молекулы ДНК может включать определение локализации указанной молекулы ДНК в геноме указанного организма и определение того, метилирована ли указанная молекула ДНК в одном или большем числе сайтов. Указанный анализ может осуществляться путем получения ридов последовательностей с помощью чувствительного к метилированию секвенирования, и, соответственно, анализ может осуществляться исключительно на основе ранее полученных для указанной ДНК данных. Согласно другим вариантам реализации анализ может включать фактическое секвенирование или другие активные этапы получения данных.

На этапе 2820 для каждого из множества сайтов определяют относительное количество молекул ДНК, которые метилированы в указанном сайте. Согласно одному варианту реализации указанные сайты представлены CpG-сайтами, и могут быть представлены только некоторыми CpG-сайтами, выбранными с применением одного или нескольких критериев, упоминаемых в настоящем документе. Число молекул ДНК, которые метилированы, эквивалентно определению числа неметилованных молекул, если нормировка проводится с использованием общего числа анализируемых молекул ДНК в конкретном сайте, например, общего числа ридов секвенирования. Например, увеличение плотности метилирования CpG в области эквивалентно уменьшению плотности неметилованных CpG в той же области.

На этапе 2830 рассчитывают первый уровень метилирования на основании соответствующих количеств молекул ДНК, метилированных во множестве сайтов. Первый уровень метилирования может соответствовать плотности метилирования, которую определяют на основании числа молекул ДНК, соответствующих указанному множеству сайтов. Указанные сайты могут соответствовать множеству локусов или только одному локусу.

На этапе 2840 сравнивают первый уровень метилирования с первым предельным значением. Первое предельное значение может представлять собой референсный уровень метилирования или быть связано с референсным уровнем метилирования (например, находится на определенном расстоянии от нормального уровня). Референсный уровень метилирования может быть определен по образцам индивидуумов без раковых заболеваний, или по локусам организма, которые, как известно, не связаны с раковым заболеванием у указанного организма. Первое предельное значение может быть установлено по референсному уровню метилирования, определенному по предыдущему биологическому образцу от указанного организма, полученному ранее тестируемого биологического образца.

Согласно одному варианту реализации первое предельное значение представляет собой указанное расстояние (например, указанное число стандартных отклонений) от референсного уровня метилирования, установленного по биологическому образцу, полученному от здорового организма. Сравнение может осуществляться путем определения разности между первым уровнем метилирования и референсным уровнем метилирования, и затем сравнения указанной разности с порогом, соответствующим первому предельному значению (например, для определения того, отличается ли статистически уровень метилирования от референсного уровня метилирования).

На этапе 2850 проводят классификацию уровня рака на основании указанного

сравнения. Примеры уровня рака включают наличие или отсутствие у субъекта ракового заболевания или предзлокачественного состояния, или повышенная вероятность развития ракового заболевания. Согласно одному варианту реализации первое предельное значение может быть определено по ранее полученному образцу от субъекта (например, референсный уровень метилирования может быть определен по ранее полученному образцу).

Согласно некоторым вариантам реализации первый уровень метилирования может соответствовать ряду областей, где уровни метилирования превышают пороговое значение. Например, может быть идентифицировано множество областей генома указанного организма. Указанные области могут быть идентифицированы с применением упоминаемых в настоящем документе критериев, например, по определенной длине или определенному числу сайтов. В составе каждой из указанных областей могут быть идентифицированы один или несколько сайтов (например, CpG-сайтов). Для каждой области может быть рассчитан уровень метилирования области. Первый уровень метилирования соответствует первой области. Каждый из уровней метилирования области сравнивают с соответствующим предельным значением для области, которое может быть одинаковым или варьировать для разных областей. Предельное значение для первой области называют первым предельным значением. Соответствующие предельные значения для областей могут быть представлены указанным количеством (например, половиной (0,5)) референсного уровня метилирования, таким образом, подсчитываются только области, значительно отличающиеся от эталона, который может быть определен по субъектам, не имеющим раковых заболеваний.

Может быть определено первое число областей, уровень метилирования которых превышает соответствующее предельное значение для области, и оно может сравниваться с пороговым значением для определения классификации. Согласно одному варианту реализации пороговое значение представляет собой процент. Сравнение указанного первого числа с пороговым значением может включать деление первого числа областей на второе число областей (например, все области) до сравнения с пороговым значением, например, в рамках процесса нормировки.

Согласно приведенному выше описанию фракционная концентрация опухолевой ДНК в биологическом образце может применяться для расчета первого предельного значения. Фракционная концентрация может определяться просто как превышающая минимальное значение, при этом образец с фракционной концентрацией ниже минимального значения может быть помечен, например, как неподходящий для анализа. Минимальное значение может быть определено на основе ожидаемого различия уровней

метилирования опухоли относительно референсного уровня метилирования. Например, если разница составляет 0,5 (например, используемая в качестве предельного значения), требуется определенная опухолевая концентрация, достаточно высокая для того, чтобы проявилась указанная разница.

Специфические методики согласно способу 1300 могут использоваться и в способе 2800. Согласно способу 1300 для опухоли могут быть определены вариации числа копий (например, первая хромосомная область опухоли может быть протестирована на изменение числа копий относительно второй хромосомной области опухоли). Соответственно, способ 1300 может предполагать, что опухоль существует. Согласно способу 2800 образец может быть протестирован на наличие какой-либо опухоли в принципе, независимо от каких-либо характеристик числа копий. Некоторые методики указанных двух способов могут быть аналогичными. Однако предельные значения и параметры метилирования (например, нормированные уровни метилирования) согласно способу 2800 позволяют обнаружить статистическое отличие от референсного уровня метилирования для неопухолевой ДНК, а не отличие от референсного уровня метилирования для смеси раковой ДНК и неопухолевой ДНК, при этом некоторые области, возможно, содержат вариации числа копий. Соответственно, референсные значения для способа 2800 могут быть определены по образцам без раковой ДНК, например, полученным от организмов, не имеющих раковых заболеваний, или по образцам неопухолевой ткани того же пациента (например, плазмы, полученной ранее, или по полученным в то же самое время образцам, которые, как известно, не затронуты раковым заболеванием, что может быть определено по клеточной ДНК).

Е. Прогнозирование минимальной фракционной концентрации опухолевой ДНК, детектируемой с применением анализа метилирования ДНК плазмы

Один из способов измерения чувствительности способа детекции ракового заболевания с использованием уровня метилирования ДНК плазмы связан с минимальной фракционной концентрацией происходящей из опухоли ДНК, необходимой для обнаружения изменения уровня метилирования ДНК плазмы при сравнении с уровнями контролей. Чувствительность тестирования также зависит от степени различия метилирования ДНК между опухолевой тканью и исходными уровнями метилирования ДНК плазмы здоровых контролей или ДНК клеток крови. Клетки крови являются основным источником ДНК в плазме здоровых индивидуумов. Чем больше указанное различие, тем проще отличить больных раковым заболеванием пациентов от не имеющих раковых заболеваний индивидуумов, что находит отражение в более низких порогах детекции происходящей из опухоли ДНК в плазме и более высокой клинической

чувствительности при детекции раковых заболеваний у пациентов. Кроме того, вариации метилирования ДНК в плазме здоровых субъектов или субъектов разного возраста (G Hannum et al. 2013 Mol Cell; 49: 359-367) также влияют на чувствительность детекции изменений метилирования, связанных с наличием ракового заболевания. Меньшая вариативность метилирования ДНК в плазме здоровых субъектов упрощает детекцию изменений, обусловленных наличием небольшого количества происходящей из ракового новообразования ДНК.

Фиг. 29А представляет собой график 2900, отражающий распределение плотностей метилирования у референсных субъектов при предположении, что указанное распределение соответствует нормальному распределению. Указанный анализ основан на наличии для каждого образца плазмы только одного показателя плотности метилирования, например, плотности метилирования всех аутосом или конкретной хромосомы. Показано, как это влияет на специфичность анализа. Согласно одному варианту реализации используется точка отсечения на 3 SD ниже среднего значения плотности метилирования ДНК у референсных субъектов для определения наличия значительно большего гипометилирования тестируемого образца по сравнению с образцами от референсных субъектов. При использовании указанной точки отсечения ожидается, что приблизительно для 0,15% не имеющих раковых заболеваний субъектов будут получены ложноположительные результаты, подтверждающие наличие ракового заболевания, что соответствует специфичности 99,85%.

Фиг. 29В представляет собой график 2950, отражающий распределения плотностей метилирования у референсных субъектов и больных раковым заболеванием пациентов. Предельное значение на 3 SD ниже среднего значения плотностей метилирования у референсных субъектов. Если указанное среднее значение плотностей метилирования больных раковыми заболеваниями пациентов на 2 SD ниже предельного значения (т.е. на 5 SD ниже среднего значения для референсных субъектов), ожидается, что у 97,5% больных раковым заболеванием субъектов плотность метилирования будет ниже предельного значения. Другими словами, ожидаемая чувствительность составит 97,5%, если для каждого субъекта известен один показатель плотности метилирования, например, если анализируется общая плотность метилирования всего генома, всех аутосом или конкретной хромосомы. На разность средних значений плотностей метилирования для двух популяций влияют два фактора, а именно, степень различия уровня метилирования между раковыми и нераковыми тканями, и фракционная концентрация происходящей из опухоли ДНК в образце плазмы. Чем выше значения указанных двух параметров, тем больше разность значений плотностей метилирования для указанных двух популяций.

Кроме того, чем ниже SD для распределений плотностей метилирования для указанных двух популяций, тем меньше пересекаются распределения плотностей метилирования для указанных двух популяций.

В настоящем описании авторы используют гипотетический пример, чтобы проиллюстрировать указанный принцип. Предположим, что плотность метилирования опухолевой ткани составляет приблизительно 0,45, а плотность метилирования ДНК плазмы здоровых субъектов составляет приблизительно 0,7. Указанные предполагаемые значения аналогичны полученным для нашего пациента с ГЦК, при этом общая плотность метилирования аутомом составляет 42,9% и среднее значение плотности метилирования аутомом для образцов плазмы от здоровых контролей составляла 71,6%. При предположении, что CV измерения плотности метилирования ДНК плазмы для полного генома составляет 1%, предельное значение будет составлять $0,7 \times (100\% - 3 \times 1\%) = 0,679$. Для обеспечения чувствительности 97,5%, среднее значение плотности метилирования ДНК плазмы для больных раковыми заболеваниями пациентов должно составлять приблизительно $0,679 - 0,7 \times (2 \times 1\%) = 0,665$. Пусть f представляет фракционную концентрацию происходящей из опухоли ДНК в образце плазмы. Тогда f может быть рассчитан как $(0,7 - 0,45) \times f = 0,7 - 0,665$. Таким образом, f составляет приблизительно 14%. С помощью этого расчета определяют, что минимальная фракционная концентрация, которую можно детектировать в плазме, составляет 14% для обеспечения диагностической чувствительности 97,5%, если в качестве диагностического параметра используется общая плотность метилирования всего генома.

Затем авторы настоящего изобретения проводили указанный анализ на данных, полученных от пациента с ГЦК. Для иллюстративных целей в этом случае осуществлялось только одно измерение плотности метилирования на основе показателя, определенного для всех аутомом, для каждого образца. Среднее значение плотности метилирования составляло 71,6% в образцах плазмы, полученных от здоровых субъектов. SD плотностей метилирования для указанных четырех образцов составляло 0,631%. Таким образом, предельное значение для плотности метилирования в плазме должно составлять $71,6\% - 3 \times 0,631\% = 69,7\%$, чтобы обеспечить Z-показатель < -3 и специфичность 99,85%. Для обеспечения чувствительности 97,5%, среднее значение плотности метилирования в плазме больных раковыми заболеваниями пациентов должно быть на 2 SD ниже точки отсечения, т.е. составлять 68,4%. Поскольку плотность метилирования опухолевой ткани составляла 42,9%, на основании формулы: $P = BK \times (1 - f) + TUM \times f$, f должна составлять по меньшей мере 11,1%.

Согласно другому варианту реализации плотности метилирования разных

геномных областей могут быть проанализированы по отдельности, например, как показано на фиг. 25А или 26В. Другими словами, осуществляли несколько измерений уровня метилирования для каждого образца. Как показано ниже, значительное гипометилирование может быть детектировано при значительно более низкой фракционной концентрации опухолевой ДНК в плазме и, соответственно, диагностическая эффективность анализа метилирования ДНК плазмы для детекции раковых заболеваний будет выше. Число геномных областей, демонстрирующих значительное отклонение плотностей метилирования от референсной популяции, может быть подсчитано. Затем указанное число геномных областей можно сравнить с предельным значением для определения наличия значительного общего гипометилирования ДНК плазмы в популяции для исследуемых геномных областей, например, участков размером 1 Мб по всему геному. Предельное значение может быть установлено путем анализа группы референсных субъектов, не имеющих раковых заболеваний, или выведено математически, например, в соответствии с функцией нормального распределения.

Фиг. 30 представляет собой график 3000, отражающий распределение плотностей метилирования ДНК плазмы здоровых субъектов и пациентов с раковыми заболеваниями. Плотность метилирования каждого участка размером 1 Мб сравнивают с соответствующими значениями для референсной группы. Определяли процент участков, демонстрирующих значительное гипометилирование (на 3 SD ниже среднего значения для референсной группы). Для определения наличия в образце плазмы происходящей из опухоли ДНК использовали точку отсечения, равную 10% значительно гипометилированных областей. Также могут использоваться другие предельные значения, такие как 5%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 60%, 70%, 80% или 90%, в соответствии с требуемыми чувствительностью и специфичностью теста.

Например, для классификации образца как содержащего происходящую из опухоли ДНК авторы настоящего изобретения могут использовать 10% участков размером 1 Мб, демонстрирующих значительное гипометилирование (Z -показатель < -3), в качестве точки отсечения. Если более чем 10% указанных участков значительно более гипометилированы по сравнению с референсной группой, образец классифицируют как положительный по результатам теста на раковое заболевание. Для каждого участка размером 1 Мб используется точка отсечения на 3 SD ниже среднего значения плотности метилирования референсной группы для определения образца как значительно более гипометилированного. Для каждого из участков размером 1 Мб, если среднее значение плотности метилирования ДНК плазмы больных раковыми заболеваниями пациентов на

1,72 SD ниже среднего значения плотностей метилирования ДНК плазмы у референсных субъектов, имеется 10% вероятность, что показатель плотности метилирования любого конкретного участка у пациента с раковым заболеванием будет ниже, чем точка отсечения (т.е. Z-показатель < -3) и даст положительный результат. В таком случае, при рассмотрении всех участков размером 1 Мб для полного генома, ожидается, что приблизительно для 10% участков будут наблюдаться положительные результаты при значительно более низких плотностях метилирования (т.е. Z-показатели < -3). При предположении, что общая плотность метилирования ДНК плазмы здоровых субъектов составляет приблизительно 0,7, и коэффициент вариации (CV) измерения плотности метилирования ДНК плазмы для каждого участка размером 1 Мб составляет 1%, среднее значение плотности метилирования ДНК плазмы больных раковыми заболеваниями пациентов должно составлять $0,7 \times (100\% - 1,72 \times 1\%) = 0,68796$. Пусть f представляет фракционную концентрацию происходящей из опухоли ДНК в плазме, обеспечивающую указанное среднее значение плотности метилирования ДНК плазмы. При предположении, что плотность метилирования опухолевой ткани составляет 0,45, значение f может быть рассчитано с применением следующего уравнения

$$(\overline{M}_{P_{ref}} - M_{tumor}) \times f = \overline{M}_{P_{ref}} - \overline{M}_{P_{cancer}}$$

где $\overline{M}_{P_{ref}}$ представляет среднее значение плотности метилирования ДНК плазмы у референсных индивидуумов; M_{tumor} представляет плотность метилирования опухолевой ткани у больного раковым заболеванием пациента; и $\overline{M}_{P_{cancer}}$ представляет среднее значение плотности метилирования ДНК плазмы у больных раковыми заболеваниями пациентов.

На основании указанного уравнения $(0,7 - 0,45) \times f = 0,7 - 0,68796$. Соответственно, минимальная фракционная концентрация, которую можно детектировать с применением указанного способа, определяется как 4,8%. Чувствительность может быть дополнительно повышена за счет снижения процента отсечения значительно более гипометилированных участков, например, от 10% до 5%.

Как показано в приведенном выше примере, чувствительность указанного способа определяется степенью различия уровня метилирования раковых и нераковых тканей, например, клеток крови. Согласно одному варианту реализации выбирают только хромосомные области, демонстрирующие значительное отличие плотностей метилирования ДНК в плазме субъектов, не имеющих раковых заболеваний, и плотностей метилирования ДНК опухолевой ткани. Согласно одному варианту реализации выбирают только области, различие плотности метилирования которых составляет $> 0,5$. Согласно

другим вариантам реализации для выбора подходящих областей может использоваться различие, составляющее 0,4, 0,6, 0,7, 0,8 или 0,9. Согласно другим вариантам реализации физический размер геномных областей не фиксирован. Вместо этого геномные области определяют, например, на основе фиксированной глубины секвенирования или фиксированного числа CpG-сайтов. Уровни метилирования в нескольких указанных геномных областях оценивают для каждого образца.

Фиг. 31 представляет собой графическое представление 3100, отражающее распределение различий среднего значения плотностей метилирования ДНК плазмы здоровых субъектов и опухолевой ткани пациента с ГЦК. Положительное значение означает, что плотность метилирования выше для ДНК в плазме здоровых субъектов, а отрицательное значение означает, что плотность метилирования выше в опухолевой ткани.

Согласно одному варианту реализации могут выбираться участки с максимальным различием плотности метилирования раковых и нераковых тканей, например, такие, для которых различие составляет $>0,5$, независимо от того, гипометилирована или гиперметилирована опухоль по указанным участкам. Предел детекции фракционной концентрации происходящей из опухоли ДНК в плазме может быть снижен за счет фокусирования на указанных участках, благодаря более значительным различиям распределений уровней метилирования ДНК плазмы между больными раком и не больными раком субъектами, при одинаковой фракционной концентрации происходящей из опухоли ДНК в плазме. Например, при использовании только участков с различиями $>0,5$, и принятии точки отсечения, соответствующей 10% значительно более гипометилированных участков для определения наличия ракового заболевания у тестируемого индивидуума, минимальная фракционная концентрация (f) детектируемой происходящей из опухоли ДНК может быть рассчитана с применением следующего уравнения: $(\overline{M}_{P_{ref}} - M_{tumor}) \times f = \overline{M}_{P_{ref}} - \overline{M}_{P_{cancer}}$, где $\overline{M}_{P_{ref}}$ представляет среднее значение плотности метилирования ДНК плазмы у референсных индивидуумов; M_{tumor} представляет плотность метилирования опухолевой ткани у больного раковым заболеванием пациента; и $\overline{M}_{P_{cancer}}$ представляет среднее значение плотности метилирования ДНК плазмы у больных раковыми заболеваниями пациентов.

При этом, различие плотности метилирования в плазме референсных субъектов и в опухолевых тканях составляет по меньшей мере 0,5. Соответственно, мы имеем $0,5 \times f = 0,7 - 0,68796$ и $f = 2,4\%$. Таким образом, фокусируясь на участках с более значительным различием плотности метилирования в раковых и нераковых тканях, нижний предел для фракционной концентрации происходящей из опухоли ДНК можно снизить с 4,8% до

2,4%. Информацию относительно того, для каких участков наблюдаются бóльшие различия степени метилирования между раковыми и нераковыми тканями, например, клетками крови, можно получить, используя опухолевые ткани из того же органа или ткани того же гистологического типа, полученные от других индивидуумов.

Согласно другому варианту реализации на основе плотности метилирования ДНК плазмы для всех участков может быть получен параметр, учитывающий различие плотностей метилирования между раковыми и нераковыми тканями. Участкам с более значительным различием может придаваться больший вес. Согласно одному варианту реализации различие плотности метилирования между раковой и нераковой тканью для каждого участка может непосредственно использоваться в качестве веса конкретного участка при вычислении конечного параметра.

Согласно еще одному варианту реализации разные типы раковых заболеваний могут иметь разные паттерны метилирования в опухолевой ткани. Может быть получен специфичный для ракового заболевания весовой профиль на основании степени метилирования при конкретном типе ракового заболевания.

Согласно еще одному варианту реализации может быть определена зависимость плотности метилирования между участками у субъектов, у которых имеется или отсутствует раковое заболевание. На фиг. 8 можно наблюдать, что в незначительном числе участков опухолевые ткани более метилированы, чем ДНК плазмы у референсных субъектов. Соответственно, могут выбираться участки с крайними значениями различия, например, различиями $>0,5$ и <0 . Отношение плотности метилирования указанных участков затем можно использовать для определения наличия ракового заболевания у тестируемого индивидуума. Согласно другим вариантам реализации разность и частное плотности метилирования разных участков могут применяться в качестве параметров для определения зависимости между участками.

Авторы также оценивали чувствительность способа детекции или оценки опухоли с применением плотностей метилирования нескольких геномных областей, что иллюстрируют данные, полученные от пациента с ГЦК. Сначала авторы настоящего изобретения объединяли ряды для предоперационной плазмы с рядами, полученными для образцов плазмы здоровых контролей, для имитации образцов плазмы, содержащих фракционные концентрации опухолевой ДНК, варьирующие от 20% до 0,5%. Затем авторы подсчитывали процент участков размером 1 Мб (из 2734 участков всего генома) с плотностями метилирования, эквивалентными Z-показателям <-3 . При фракционной концентрации опухолевой ДНК в плазме, составляющей 20%, для 80,0% участков наблюдалось значительное гипометилирование. Соответствующие данные для

фракционной концентрации опухолевой ДНК в плазме, равной 10%, 5%, 2%, 1% и 0,5%, составляли 67,6%, 49,7%, 18,9%, 3,8% и 0,77% участков, демонстрирующих гипометилирование, соответственно. Поскольку теоретический предел числа участков, демонстрирующих Z-показатели < -3 , в контрольных образцах составляет 0,15%, полученные авторами данные демонстрируют, что даже при фракционной концентрации опухолевой ДНК, составляющей всего 0,5%, имелось больше участков (0,77%) за теоретическим пределом отсечения.

Фиг. 32А представляет собой таблицу 3200, демонстрирующую эффект уменьшения глубины секвенирования в случаях, когда образец плазмы содержал 5% или 2% опухолевой ДНК. Высокое содержание участков ($>0,15%$), демонстрирующих значительное гипометилирование, может быть детектировано даже при среднем значении глубины секвенирования всего $0,022 \times$ гаплоидного генома.

Фиг. 32В представляет собой графическое представление 3250, отражающее плотности метилирования повторяющихся элементов и областей без повторов в плазме четырех здоровых контрольных субъектов, лейкоцитарной пленке, нормальной ткани печени, опухолевой ткани, образцах предоперационной плазмы и послеоперационной плазмы пациента с ГЦК. Видно, что повторяющиеся элементы более метилированы (более высокая плотность метилирования), чем области без повторов как в раковых, так и в нераковых тканях. Однако различие метилирования между повторяющимися элементами и областями без повторов было более значительным в нераковых тканях и ДНК плазмы здоровых субъектов, по сравнению с опухолевыми тканями.

В результате, для ДНК плазмы больного раковым заболеванием пациента наблюдалось большее снижение плотности метилирования для повторяющихся элементов, чем для областей без повторов. Различие плотности метилирования ДНК в плазме между средним значением для четырех здоровых контролей и для пациента с ГЦК составляло 0,163 и 0,088 для повторяющихся элементов и областей без повторов, соответственно. Данные для образцов предоперационной и послеоперационной плазмы также показали, что динамический диапазон изменения плотности метилирования был больше для повторяющихся областей по сравнению с областями без повторов. Согласно одному варианту реализации плотность метилирования повторяющихся элементов ДНК в плазме может использоваться для определения наличия ракового заболевания у пациента или для мониторинга прогрессирования заболевания.

Как обсуждалось выше, вариации плотностей метилирования в плазме у референсных субъектов также влияют на точность дифференциации больных раковым заболеванием пациентов от индивидуумов, не имеющих раковых заболеваний. Чем

плотнее распределение плотностей метилирования (т.е. меньшее стандартное отклонение), тем точнее будет дифференциация больных раком и не имеющих раковых заболеваний субъектов. Согласно другому варианту реализации коэффициент вариации (CV) плотностей метилирования участков размером 1 Мб может применяться в качестве критерия для выбора участков с низкой вариабельностью плотностей метилирования ДНК в плазме в референсной группе. Например, выбирают только участки с CV <1%. Другие значения, например, 0,5%, 0,75%, 1,25% и 1,5% также могут применяться в качестве критериев для выбора участков с низкой вариабельностью плотности метилирования. Согласно еще одному варианту реализации критерии выбора могут включать как CV для указанного участка, так и различие плотности метилирования между раковыми и нераковыми тканями.

Плотность метилирования также может использоваться для определения фракционной концентрации происходящей из опухоли ДНК в образце плазмы, если известна плотность метилирования опухолевой ткани. Указанная информация может быть получена путем анализа опухоли пациента или исследования опухолей нескольких пациентов, у которых имеется раковое заболевание того же типа. Как обсуждалось выше, плотность метилирования в плазме (P) может быть выражена с применением следующего уравнения: $P = BKG \times (1 - f) + TUM \times f$, где BKG представляет собой фоновую плотность метилирования в клетках крови и других органах, TUM представляет собой плотность метилирования в опухолевой ткани, и f представляет собой фракционную концентрацию происходящей из опухоли ДНК в образце плазмы. Указанное уравнение может быть

переформулировано как:
$$f = \frac{BKG - P}{BKG - TUM}$$
.

Значения BKG могут быть определены путем анализа образца плазмы пациента в момент времени, когда рак отсутствует, или путем изучения референсной группы индивидуумов без раковых заболеваний. Таким образом, после измерения плотности метилирования в плазме, может быть определено значение f .

F. Комбинирование с другими способами

Способы анализа метилирования, описанные в настоящем документе, могут применяться в комбинации с другими способами, которые основаны на генетических изменениях происходящей из опухоли ДНК в плазме. Примеры таких способов включают анализ на связанные с раковым заболеванием хромосомные aberrации (КСА Chan et al. 2013 Clin Chem; 59:211-224; RJ Leary et al. 2012 Sci Transl Med; 4:162ra154) и связанные с раковым заболеванием однонуклеотидные вариации в плазме (КСА Chan et al. 2013 Clin Chem; 59:211-224). Применение анализа метилирования имеет некоторые преимущества

по сравнению с указанными генетическими методами.

Как показано на фиг. 21А, гипометилирование опухолевой ДНК представляет собой глобальное явление, охватывающее области, распределенные почти по всему геному. Таким образом, фрагменты ДНК всех хромосомных областей будут информативными в смысле внесения потенциального вклада в наличие происходящей из опухоли гипометилированной ДНК в плазме/сыворотке пациента. В то же время хромосомные aberrации (амплификация или делеция хромосомной области) присутствуют только в некоторых хромосомных областях, и фрагменты ДНК из областей без хромосомных aberrаций в опухолевой ткани не будут информативными для анализа (КСА Chan et al. 2013 Clin Chem; 59: 211-224). Аналогичным образом, всего несколько тысяч однонуклеотидных изменений наблюдается в каждом раковом геноме (КСА Chan et al. 2013 Clin Chem; 59: 211-224). Фрагменты ДНК, не перекрывающиеся с указанными однонуклеотидных изменений не будут информативными для определения наличия в плазме, происходящей из опухоли ДНК. Таким образом, предложенный способ анализа метилирования потенциально представляет собой более экономически эффективный подход, чем указанные генетические методы детекции связанных с раковым заболеванием изменений в кровотоке.

Согласно одному варианту реализации экономическая эффективность анализа метилирования в плазме ДНК может быть дополнительно увеличена путем обогащения фрагментами ДНК из наиболее информативных областей, например, областей с максимальными различиями метилирования между раковыми и не являющимися раковыми тканями. Примеры способов обогащения указанными областями включают применение зондов гибридизации (например, системы Nimblegen SeqCap и системы Agilent SureSelect Target Enrichment), ПЦР-амплификация и твердофазную гибридизацию.

G. Тканеспецифичный анализ/доноры

Происходящие из опухоли клетки внедряются и метастазируют в смежные или отдаленные органы. Пораженные ткани или метастазы вносят вклад в ДНК плазмы в результате клеточной смерти. Путем анализа профиля метилирования ДНК в плазме страдающих раковым заболеванием пациентов и детекции присутствия тканеспецифичных сигнатур метилирования, можно детектировать типы тканей, которые вовлечены в патологический процесс. Указанный способ обеспечивает неинвазивное анатомическое сканирование тканей, вовлеченных в злокачественный процесс, что помогает идентифицировать органы, задействованные в первичных очагах и метастатических сайтах. Мониторинг относительных концентраций сигнатур метилирования вовлеченных органов в плазме также позволит оценить опухолевую

нагрузку в указанных органах и определить, прогрессируют ли обусловленные раковым заболеванием повреждения в указанном органе, или наблюдается улучшение или излечение. Например, предположим, что ген X специфически метилирован в печени. В таком случае можно ожидать, что при метастазировании в печень ракового новообразования (например, рака толстой и прямой кишки) концентрация метилированных последовательностей гена X в плазме будет повышаться. Может существовать также другая последовательность или группы последовательностей с аналогичными гену X характеристиками метилирования. В таком случае результаты для таких последовательностей могут быть объединены. Аналогичные соображения применимы к другим тканям, например, головному мозгу, костям, легким и почкам, и т.д.

С другой стороны, известно, что в ДНК из разных органов наблюдаются тканеспецифичные сигнатуры метилирования (BW Futscher et al. 2002 Nat Genet; 31:175-179; SSC Chim et al. 2008 Clin Chem; 54: 500-511). Соответственно, определение профиля метилирования в плазме может применяться для установления вносимого тканями различных органов вклада в плазму. Установление такого вклада может применяться для оценки повреждения органа, так как считается, что ДНК плазмы высвобождается при гибели клеток. Например, патология печени, такая как гепатит (например, обусловленный вирусами, аутоиммунными процессами и т.п.), или гепатотоксичность (например, передозировка лекарствами (например, парацетамолом) или токсинами (например, алкоголем)), вызванная лекарственными средствами, связана с повреждением клеток печени и предположительно будет связана с повышением уровня происходящей из печени ДНК в плазме. Например, предположим, что ген X специфически метилирован в печени. В таком случае можно ожидать, что при патологии печени концентрация метилированных последовательностей гена X в плазме будет повышаться. И напротив, предположим, что ген Y специфично гипометилирован в печени. В таком случае можно ожидать, что при патологии печени концентрация метилированных последовательностей гена Y в плазме будет снижаться. Согласно еще одному варианту реализации ген X или Y может быть заменен на любые геномные последовательности, которые не обязательно представляют собой ген, демонстрирующие различия метилирования в разных тканях организма.

Методики, описанные в настоящем документе, также могут применяться для оценки происходящей от донора ДНК в плазме реципиентов трансплантации органов (YMD Lo et al. 1998 Lancet; 351:1329-1330). Различия полиморфизмов между донором и реципиентом использовались для дифференциации происходящей от донора ДНК и происходящей от реципиента ДНК в плазме (YW Zheng et al. 2012 Clin Chem; 58: 549-558). Авторы настоящего изобретения предположили, что тканеспецифичные сигнатуры

метилирования трансплантированного органа также можно использовать в качестве способа детекции донорской ДНК в плазме реципиента.

Путем мониторинга концентрации ДНК донора можно неинвазивным образом оценить статус трансплантированного органа. Например, отторжение трансплантата связано с более высоким уровнем клеточной смерти и, соответственно, концентрация ДНК донора в плазме реципиента (или сыворотке), которую отражает сигнатура метилирования трансплантированного органа, будет повышаться относительно концентрации в то время, когда состояние пациента стабильно, или по сравнению с другими стабильными реципиентами трансплантатов или здоровыми контролями, которым не проводилась трансплантация. Аналогично тому, что было описано для раковых заболеваний, происходящая от донора ДНК может быть идентифицирована в плазме реципиентов трансплантации путем детекции всех или некоторых характерных особенностей, включая различия полиморфизмов, ДНК меньших размеров для трансплантированных солидных органов (YW Zheng et al. 2012 Clin Chem; 58: 549-558) и тканеспецифичного профиля метилирования.

Н. Нормировка метилирования по размерам

Согласно приведенному выше описанию и описанию у Lun et al (FMM Lun et al. Clin. Chem. 2013; ЦИО:10.1373/clinchem.2013.212274) плотность метилирования (например, ДНК плазмы) коррелирует с размером фрагментов ДНК. Распределение плотностей метилирования для более коротких фрагментов ДНК плазмы было значительно ниже, чем для более длинных фрагментов. Авторы настоящего изобретения предполагают, что при некоторых не являющихся злокачественными состояниях (например, системной красной волчанке (СКВ)) с аномальными паттернами фрагментации ДНК плазмы может наблюдаться очевидное гипометилирование ДНК плазмы благодаря присутствию большего количества коротких фрагментов ДНК плазмы, которые менее метилированы. Другими словами, распределение размеров ДНК плазмы может представлять собой искажающий фактор для плотности метилирования ДНК плазмы.

На фиг. 34А представлено распределение размеров ДНК плазмы у пациента с СКВ СКВ04. Распределения размеров для девяти здоровых контрольных субъектов показаны пунктирными серыми линиями и распределения для СКВ04 показан в виде сплошной черной линии. Короткие фрагменты ДНК плазмы присутствовали в большем количестве у СКВ04 по сравнению с девятью здоровыми контрольными субъектами, поскольку более короткие фрагменты ДНК обычно менее метилированы, указанный паттерн распределения размеров может исказить результаты анализа метилирования для ДНК плазмы и приводить к более выраженному гипометилированию.

Согласно некоторым вариантам реализации измеренный уровень метилирования может быть нормирован для уменьшения искажающего эффекта распределения размеров на анализ метилирования ДНК плазмы. Например, может быть измерен размер молекул ДНК во множестве сайтов. Согласно разнообразным вариантам реализации указанное измерение может обеспечивать конкретный размер (например, длину) молекулы ДНК или просто определять, что размер попадает в конкретный диапазон, который может также соответствовать размеру. Нормированный уровень метилирования затем может сравниваться с предельным значением. Существует несколько способов выполнения нормировки для снижения искажающего эффекта распределения размеров на анализ метилирования ДНК плазмы.

Согласно одному варианту реализации может осуществляться фракционирование по размеру ДНК (например, ДНК плазмы). Фракционирование по размеру может обеспечить использование фрагментов ДНК аналогичного размера для определения уровня метилирования в соответствии с предельным значением. В рамках фракционирования по размеру могут выбираться фрагменты ДНК первого размера (например, первого диапазона длин), при этом первое предельное значение соответствует первому размеру. Нормировку можно осуществлять путем расчета уровня метилирования с применением исключительно выбранных фрагментов ДНК.

Фракционирование по размеру может достигаться различным образом, например, либо физическим разделением разных по размеру молекул ДНК (например, посредством электрофореза, или с помощью технологий на основе микрофлюидики, или технологий на основе центрифугирования), либо посредством анализа *in silico*. Для анализа *in silico* согласно одному варианту реализации может быть проведено массивно-параллельное секвенирование спаренных концов молекул ДНК плазмы. Затем может быть установлен размер секвенированных молекул путем сравнения локализации каждого из двух концов молекулы ДНК плазмы с референсным геномом человека. Затем может быть проведен последующий анализ путем отбора секвенированных молекул ДНК, соответствующих одному или нескольким критериям выбора размера (например, критериям, определяющим размер в указанном диапазоне). Соответственно, согласно одному варианту реализации плотность метилирования может быть проанализирована для фрагментов аналогичного размера (например, в указанном диапазоне). Предельное значение (например, на этапе 2840 способа 2800) может быть определено на основе фрагментов из того же диапазона размеров. Например, уровни метилирования могут быть определены по образцам, которые, как известно, относятся к раковому заболеванию или не относятся к раковому заболеванию, и по указанным уровням метилирования могут быть определены

предельные значения.

Согласно другому варианту реализации может быть определена функциональная зависимость между плотностью метилирования и размером циркулирующей ДНК. Указанная функциональная зависимость может быть определена точкой данных или коэффициентами функции. Указанная функциональная зависимость может обеспечивать масштабирующие значения, соответствующие соответствующим размерам (например, метилирование при более коротких размерах может соответствующим образом повышаться). Согласно разнообразным вариантам реализации масштабирующее значение может составлять от 0 до 1, или превышать 1.

Можно осуществить нормировку по среднему размеру. Например, может быть вычислен средний размер, соответствующий молекулам ДНК, используемым для расчета первого уровня метилирования, и указанный первый уровень метилирования может быть умножен на соответствующее масштабирующее значение (т.е. соответствующее среднему размеру). В другом примере плотность метилирования каждой молекулы ДНК может быть нормирована в соответствии с размером молекулы ДНК и зависимостью между размером и метилированием ДНК.

Согласно другому варианту реализации нормировка может быть выполнена в расчете на молекулу. Например, может быть получен соответствующий размер молекулы ДНК в конкретном сайте (например, согласно приведенному выше описанию), и масштабирующее значение, соответствующее соответствующему размеру, может быть идентифицировано по функциональной зависимости. При ненормированном расчете каждую молекулу подсчитывают одинаковым образом при определении индекса метилирования в указанном сайте. При нормированном расчете вклад молекулы в индекс метилирования может быть взвешен по масштабирующему коэффициенту, который соответствует размеру молекулы.

На фиг. 34В и 34С представлен анализ метилирования для ДНК плазмы от пациента с СКВ СКВ04 (фиг. 34В) и пациента с ГЦК TBR36 (фиг. 34С). Внешние круги отображают результаты Z_{meth} для ДНК плазмы без фракционирования по размеру *in silico*. Внутренние круги отображают результаты Z_{meth} для ДНК плазмы размером 130 п.о. или более. У пациента с СКВ, СКВ04, наблюдалось гипометилирование 84% участков без фракционирования по размеру *in silico*. Процент участков, демонстрирующих гипометилирование, снижался до 15% при анализе только фрагментов размером 130 п.о. или более. У пациента с ГЦК TBR36, наблюдалось гипометилирование 98,5% и 98,6% участков ДНК плазмы с фракционированием и без фракционирования по размеру *in silico*, соответственно. Указанные результаты предполагают, что фракционирование по размеру

in silico позволяет эффективно уменьшать количество ложноположительных результатов анализа гипометилирования, связанных с повышенной фрагментацией ДНК плазмы, например, у пациентов с СКВ или другими воспалительными состояниями.

Согласно одному варианту реализации могут сравниваться результаты анализов с фракционированием и без фракционирования по размеру для обнаружения наличия какого-либо искажающего эффекта размера на результаты анализа метилирования. Соответственно, вдобавок или вместо нормировки может применяться расчет уровня метилирования для конкретного размера, для определения того, имеется ли вероятность ложноположительного результата, если процент участков со значениями выше предельного значения отличается при фракционировании и без фракционирования по размеру, или отличается только конкретный уровень метилирования. Например, присутствие значимого различия между этими результатами для образцов с фракционированием и без фракционирования по размеру может использоваться для определения вероятности ложноположительного результата, обусловленного аномальным паттерном фрагментации. Порог для определения значимости различий может быть установлен с помощью анализа когорты страдающих раковым заболеванием пациентов и когорты не имеющих раковых заболеваний контрольных субъектов.

I. Анализ полногеномного гиперметилирования CpG-островков в плазме

Помимо общего гипометилирования, при раковых заболеваниях часто наблюдается также гиперметилирование CpG-островков (SB Baylin et al. 2011 Nat Rev Cancer; 11: 726-734; PA Jones et al. 2007, Cell; 128: 683-692; M Esteller et al. 2007 Nat Rev Genet 2007; 8: 286-298; M Ehrlich et al. 2002 Oncogene 2002; 21: 5400-5413). В данном разделе авторы настоящего изобретения описывают применение полногеномного анализа гиперметилирования CpG-островков для детекции и мониторинга раковых заболеваний.

Фиг. 35 представляет собой блок-схему способа 3500 определения классификации уровня рака на основе гиперметилирования CpG-островков в соответствии с вариантами реализации настоящего изобретения. Множество сайтов согласно способу 2800 может включать CpG-сайты, при этом указанные CpG-сайты организованы во множество CpG-островков, где каждый CpG-островок включает один или большее количество CpG-сайтов. Уровни метилирования для каждого CpG-островка могут использоваться для определения классификации уровня рака.

На этапе 3510 идентифицируют подлежащие анализу CpG-островки. В указанном анализе в качестве примера авторы настоящего изобретения сначала определяли набор CpG-островков для анализа, характеризующихся относительно низкими плотностями метилирования в плазме здоровых референсных субъектов. Согласно одному аспекту

вариация плотностей метилирования в референсной группе может быть относительно незначительной, что делает проще детекцию связанного с раковым заболеванием гиперметилирования. Согласно одному варианту реализации среднее значение плотности метилирования CpG-островков меньше первого процента в референсной группе, и коэффициент вариации для плотности метилирования в референсной группе меньше, чем второй процент.

Например, в целях иллюстрации используют следующие критерии для идентификации подходящих CpG-островков:

i. Среднее значение плотности метилирования для CpG-островка в референсной группе (например, здоровых субъектов) <5%

ii. Коэффициент вариации для анализа плотности метилирования в плазме для референсной группы (например, здоровых субъектов) <30%.

Указанные параметры могут быть скорректированы для конкретного применения. В полученном авторами наборе данных 454 CpG-островков в геноме удовлетворяли указанным критериям.

На этапе 3520 рассчитывают плотность метилирования для каждого CpG-островка. Плотности метилирования могут быть определены согласно описанию в настоящем документе.

На этапе 3530 определяют, гиперметилирован ли каждый из CpG-островков. Например, для анализа на гиперметилирование CpG-островка в тестируемом случае, плотность метилирования каждого CpG-островка сравнивали с соответствующими данными референсной группы. Плотность метилирования (пример уровня метилирования) может сравниваться с одним или несколькими предельными значениями для определения того, гиперметилирован ли конкретный островок.

Согласно одному варианту реализации первое предельное значение может соответствовать среднему значению плотностей метилирования для референсной группы плюс указанный процент. Другое предельное значение может соответствовать среднему значению плотностей метилирования для референсной группы плюс указанное число стандартных отклонений. Согласно одному варианту реализации рассчитывали Z-показатель (Z_{meth}) и сравнивали с предельными значениями. Например, CpG-островок у тестируемого субъекта (например, субъекта, у которого проводится скрининг на рак) считался значительно гиперметилированным, если он удовлетворял следующим критериям:

i. его плотность метилирования была выше, чем среднее значение референсной группы, на 2%, и

ii. $Z_{meth} > 3$

Указанные параметры могут также быть скорректированы для конкретного применения.

На этапе 3540 используют плотности метилирования (например, в виде Z -показателей) гиперметилированных CpG-островков для определения кумулятивного показателя. Например, после идентификации всех значительно гиперметилированных CpG-островков может быть рассчитан показатель, включающий сумму Z -показателей или функции Z -показателей всех гиперметилированных CpG-островков. Примером показателя является показатель суммарной вероятности (CP), согласно описанию в другом разделе. Показатель суммарной вероятности использует Z_{meth} для определения вероятности такого наблюдения за счет случайного эффекта в соответствии с распределением вероятности (например, t -распределения вероятности Стьюдента с 3 степенями свободы).

На этапе 3550 кумулятивный показатель сравнивают с кумулятивным порогом для установления классификации уровня рака. Например, если общее гиперметилирование в идентифицированных CpG-островках достаточно велико, организм может быть идентифицирован как пораженный раковым заболеванием. Согласно одному варианту реализации кумулятивный порог соответствует максимальному кумулятивному показателю референсной группы.

IX. Метилирование и CNA (абберации числа копий)

Как упоминалось выше, способы анализа метилирования, описанные в настоящем документе, могут применяться в комбинации с другими способами, которые основаны на генетических изменениях происходящей из опухоли ДНК в плазме. Примеры таких способов включают анализ на связанные с раковым заболеванием хромосомные абберации (KCA Chan et al. 2013 Clin Chem; 59: 211-224; RJ Leary et al. 2012 Sci Transl Med; 4: 162ra154). Аспекты аббераций числа копий (CNA) описаны в заявке на патент США № 13/308473.

A. CNA

Абберации числа копий могут быть обнаружены путем подсчета фрагментов ДНК, которые выравниваются с конкретной частью генома, нормировки полученного подсчета и сравнения подсчитанного значения с предельным значением. Согласно различным вариантам реализации нормировка может осуществляться путем подсчета фрагментов ДНК, выравниваемых по другому гаплотипу той же части генома (метод анализа относительных доз гаплотипа (RHDO)) или путем подсчета фрагментов ДНК, выравниваемых по другой части генома.

Способ RHDO основан на применении гетерозиготных локусов. Варианты

реализации, описанные в данном разделе, также могут применяться для гомозиготных локусов, путем сравнения двух областей, а не двух гаплотипов одной области, и, соответственно, не являются специфичными для гаплотипов. В родственном методе относительных доз хромосомных областей число фрагментов одной хромосомной области (например, определенное с помощью подсчета ридов последовательностей, выравниваемых по указанной области) сравнивают с ожидаемым значением (которое может относиться к референсной хромосомной области или той же области в другом, заведомо здоровом образце). Таким образом, фрагмент хромосомной области будет подсчитан независимо от того, из какого гаплотипа происходит секвенируемая метка. Соответственно, риды последовательностей, которые не содержат гетерозиготных локусов, тем не менее могут использоваться. Для проведения сравнения, согласно варианту реализации, подсчитанное число меток может быть нормировано перед указанным сравнением. Каждая область определяется по меньшей мере двумя локусами (которые отделены друг от друга), и фрагменты в указанных локусах могут применяться для получения общего значения для указанной области.

Нормированное значение для ридов секвенирования (меток) для конкретной области может быть рассчитано путем деления числа ридов секвенирования, выравниваемых с указанной областью, на общее число ридов секвенирования, выравниваемых со всем геномом. Указанное подсчитанное число меток нормировки позволяет сравнить результаты от одного образца с результатами для другого образца. Например, нормированное значение может представлять собой долю (например, процент или фракцию) ридов секвенирования, предположительно происходящих из конкретной области, как указано выше. Согласно другим вариантам реализации возможно применение других способов нормировки. Например, возможна нормировка путем деления подсчитанного числа для одной области на подсчитанное число для референсной области (в описанном выше случае референсная область представлена просто полным геномом). Указанное подсчитанное число меток нормировки затем можно сравнить с пороговым значением, которое может быть определено по одному или нескольким референсным образцам без раковых заболеваний.

Подсчитанное число меток нормировки тестируемого случая затем сравнивают с подсчитанным числом меток нормировки одного или нескольких референсных субъектов, например, субъектов без раковых заболеваний. Согласно одному варианту реализации указанное сравнение осуществляют путем расчета Z-показателя указанного случая для конкретной хромосомной области. Z-показатель может быть рассчитан с применением следующего уравнения: Z-показатель = (подсчитанное число меток нормировки

указанного случая – среднее значение) / SD, где «среднее значение» представляет собой среднее значение подсчитанного числа меток нормировки, выравнивающих по конкретной хромосомной области для референсных образцов; и SD представляет собой стандартное отклонение подсчитанного числа меток нормировки, выравнивающих по конкретной области для референсных образцов. Соответственно, Z-показатель представляет собой число стандартных отклонений, на которое подсчитанное число меток нормировки хромосомной области для тестируемого случая отличается от среднего значения подсчитанного числа меток нормировки для той же хромосомной области у одного или нескольких референсных субъектов.

В ситуации, когда у тестируемого организма имеется раковое заболевание, хромосомные области, амплифицированные в опухолевых тканях, будут избыточно представлены в ДНК плазмы. Это будет приводить к положительному значению Z-показателя. С другой стороны, хромосомные области, которые удалены в опухолевых тканях, будут недопредставлены в ДНК плазмы. Это будет приводить к отрицательному значению Z-показателя. Величина Z-показателя определяется несколькими факторами.

Один фактор представлен фракционной концентрацией происходящей из опухоли ДНК в биологическом образце (например, в плазме). Чем выше фракционная концентрация происходящей из опухоли ДНК в образце (например, в плазме), тем больше будет разность между подсчитанным числом меток нормировки для тестируемого случая и референсных случаев. Соответственно, будет наблюдаться большее значение Z-показателя.

Другой фактор представлен вариацией подсчитанного числа меток нормировки в одном или нескольких референсных случаях. При одинаковом содержании хромосомной области в биологическом образце (например, в плазме) для тестируемого случая, меньшая вариация (т.е. меньшее стандартное отклонение) подсчитанного числа меток нормировки в референсной группе будет приводить к более высокому Z-показателю. Аналогичным образом, при одинаковой степени пониженной представленности хромосомной области в биологическом образце (например, в плазме) для тестируемого случая, меньшее стандартное отклонение подсчитанного числа меток нормировки в референсной группе будет приводить к более отрицательному значению Z-показателя.

Другим фактором является величина хромосомных aberrаций в опухолевых тканях. Величина хромосомных aberrаций относится к изменениям числа копий для конкретной хромосомной области (добавление или утрата). Чем значительнее изменения числа копий в опухолевых тканях, тем выше степень избыточной представленности или пониженной представленности конкретной хромосомной области в ДНК плазмы.

Например, утрата обеих копий хромосомы будет приводить к большей пониженной представленности хромосомной области в ДНК плазмы, чем утрата одной из двух копий хромосомы, и, соответственно, приводить к более отрицательному значению Z-показателя. Как правило, при раковых заболеваниях имеется несколько хромосомных aberrаций. Указанные хромосомные aberrации в каждом случае ракового заболевания могут дополнительно варьировать по характеру (т.е. представлять собой амплификацию или делецию), степени (одно или несколько добавлений или утрат копий) и объему (размеру aberrации относительно длины хромосомы).

На точность измерения подсчитанного числа меток нормировки влияет число анализируемых молекул. Авторы настоящего изобретения предполагают, что необходим анализ 15 000, 60 000 и 240 000 молекул для детекции хромосомных aberrаций с одним изменением числа копий (добавлением или утратой), если фракционная концентрация составляет приблизительно 12,5%, 6,3% и 3,2%, соответственно. Более подробная информация относительно подсчета меток для детекции раковых заболеваний для разных хромосомных областей приведена в патентной публикации США №2009/0029377, авторы Lo et al., озаглавленной "Diagnosing Fetal Chromosomal Aneuploidy Using Massively Parallel Genomic Sequencing" («Диагностика хромосомной анеуплоидии плода с применением массивно-параллельного секвенирования генома»), все содержание которой включено в настоящий документ посредством ссылки во всех отношениях.

Согласно вариантам реализации может также использоваться анализ размеров вместо метода подсчета меток. Может также применяться анализ размеров вместо подсчета меток нормировки. При анализе размеров могут использоваться различные параметры, упоминаемые в настоящем документе и в заявке на патент США № 12/940992. Например, могут использоваться значения Q или F, см. выше. Такие размерные значения не нуждаются в нормировки по подсчетам для других областей, поскольку указанные значения не изменяются в зависимости от числа ридов. Методики, использующие гаплотип-специфические методы, такие как метод RHDO, описанный выше, а также описанный более подробно в заявке на патент США №13/308473, могут применяться также для неспецифических способов. Например, могут быть использованы методики, действующие глубину секвенирования и уточнение области. Согласно некоторым вариантам реализации может учитываться стандартное смещение GC для конкретной области, если сравниваются две области. Поскольку в методе RHDO используется одна и та же область, такая коррекция не нужна.

Хотя некоторые раковые заболевания могут в общем случае сопровождаться aberrациями в конкретных хромосомных областях, такие раковые заболевания не всегда

сопровождаются aberrациями исключительно в таких областях. Например, могут наблюдаться aberrации в дополнительных хромосомных областях, и локализация таких дополнительных областей может быть неизвестна. Кроме того, при скрининге пациентов для выявления ранних стадий рака может требоваться идентификация широкого спектра видов рака, при которых могут наблюдаться aberrации по всему геному. В указанных ситуациях, согласно вариантам реализации, можно анализировать множество областей систематическим образом для определения того, в каких областях наблюдаются aberrации. Число aberrаций и их локализация (например, информация о том, последовательны ли они) может использоваться, например, для подтверждения aberrаций, определения стадии ракового заболевания, диагностирования ракового заболевания (например, если указанное число выше порогового значения), и получения прогноза на основе указанного числа и локализации различных областей, в которых наблюдается aberrация.

Соответственно, варианты реализации позволяют идентифицировать наличие в организме ракового заболевания на основании числа областей, в которых наблюдается aberrация. Соответственно, можно протестировать множество областей (например, 3000) для идентификации числа областей, в которых наблюдается aberrация. Указанные области могут захватывать весь геном или только части генома, например, область без повторов.

Фиг. 36 представляет собой блок-схему способа 3600 анализа биологического образца организма с применением множества хромосомных областей в соответствии с вариантами реализации настоящего изобретения. Указанный биологический образец включает молекулы нуклеиновых кислот (называемые также фрагментами).

На этапе 3610 идентифицируют множество областей (например, неперекрывающихся областей) генома указанного организма. Каждая хромосомная область включает множество локусов. Размер области может составлять 1 Мб, или быть представлен каким-либо другим равным размером. В ситуации, когда размер области составляет 1 Мб, весь геном может включать приблизительно 3000 областей, каждая из которых имеет заранее установленный размер и локализацию. Такие заранее установленные области могут варьировать, чтобы соответствовать длине конкретной хромосомы или указанному числу областей, которые предполагается использовать, и любым другим критериям, упоминаемым в настоящем документе. Если области имеют разные длины, такие длины могут использоваться для нормировки результатов, например, согласно описанию в настоящем документе. Указанные области могут быть специально отобраны на основании определенных критериев для конкретного организма и/или на

основании информации о раковом заболевании, на которое проводится тестирование. Области также могут быть выбраны произвольно.

На этапе 3620 идентифицируют локализацию молекулы нуклеиновой кислоты в референсном геноме указанного организма для каждой из множества молекул нуклеиновых кислот. Локализация может быть определена любым из способов, упоминаемых в настоящем документе, например, посредством секвенирования фрагментов для получения секвенируемых меток и выравнивания указанных секвенируемых меток по референсному геному. Для гаплотип-специфических способов может также быть определен конкретный гаплотип молекулы.

Этапы 3630–3650 выполняют для каждой из хромосомных областей. На этапе 3630 соответствующую группу молекул нуклеиновых кислот идентифицируют как происходящую из хромосомной области на основе идентифицированных локализаций. Указанная соответствующая группа может включать по меньшей мере одну молекулу нуклеиновой кислоты, локализованную в каждом из множества локусов хромосомной области. Согласно одному варианту реализации указанная группа может быть представлена фрагментами, которые выравниваются по конкретному гаплотипу хромосомной области, например, как в описанном выше методе RHDO. Согласно другому варианту реализации указанная группа может быть представлена любым фрагментом, который выравнивается на хромосомную область.

На этапе 3640 компьютерная система рассчитывает соответствующее значение для соответствующей группы молекул нуклеиновых кислот. Указанное соответствующее значение определяет свойства молекул нуклеиновых кислот соответствующей группы. Указанное соответствующее значение может представлять собой любое из значений, упоминаемых в настоящем документе. Например, указанное значение может представлять собой число фрагментов в группе или статистический показатель распределения размеров фрагментов в группе. Соответствующее значение может также представлять собой нормированное значение, например, подсчитанное число меток указанной области, разделенное на общее число подсчитанных меток для указанного образца или подсчитанное число меток для референсной области. Соответствующее значение может также определяться относительно другого значения как разность или соотношение (например, в методе RHDO), определяя таким образом отличительный признак для области.

На этапе 3650 соответствующее значение сравнивают с референсным значением для определения классификации первой хромосомной области как демонстрирующей делецию или амплификацию. Указанное референсное значение может представлять собой

любое пороговое или референсное значение из описанных в настоящем документе. Например, референсное значение может представлять собой пороговое значение, определенное для нормальных образцов. В методе RHDO соответствующее значение может представлять собой разность или соотношение подсчитанного числа меток для двух гаплотипов, и референсное значение может представлять собой порог для определения наличия статистически значимого отклонения. В другом примере референсное значение может представлять собой подсчитанное число меток или размерную величину для другого гаплотипа или области, и сравнение может включать получение разности или отношения (или их функции) с последующим определением того, превышает ли указанная(ое) разность или отношение пороговое значение.

Референсное значение может варьировать в зависимости от результатов для других областей. Например, если соседние области также демонстрируют отклонение (хотя и незначительное относительно одного порога, например, Z-показателя, равного 3), может применяться более низкий порог. Например, если превышение первого порога наблюдается для трех последовательных областей, раковое заболевание может быть более вероятным. Соответственно, указанный первый порог может быть ниже второго порога, необходимого для идентификации рака по непоследовательным областям. Наличие трех областей (или более чем трех), демонстрирующих даже незначительное отклонение, может иметь достаточно низкую вероятность случайного эффекта для сохранения чувствительности и специфичности.

На этапе 3660 определяют количество геномных областей, классифицированных как демонстрирующие делецию или амплификацию. Подсчитываемые хромосомные области могут иметь ограничения. Например, могут подсчитываться только области, являющиеся последовательными относительно по меньшей мере одной другой области (или может требоваться определенный размер последовательных областей, например, 4 или более области). В вариантах реализации с неравными областями указанное число также может определяться с учетом соответствующих длин (например, указанное число может представлять собой общую длину aberrантных областей).

На этапе 3670 указанное количество сравнивают с пороговым значением количества для определения классификации образца. Например, указанная классификация может определять наличие в организме ракового заболевания, стадию ракового заболевания и прогноз для ракового заболевания. Согласно одному варианту реализации подсчитывают все aberrантные области и используют одно пороговое значение независимо от локализации указанных областей. Согласно другому варианту реализации пороговое значение может варьировать на основании локализации и размера

подсчитываемых областей. Например, количество областей на конкретной хромосоме или плече хромосомы может сравниваться с порогом для указанной конкретной хромосомы (или плеча). Может использоваться несколько порогов. Например, количество aberrантных областей на конкретной хромосоме (или плече) должно быть больше первого порогового значения, а общее количество aberrантных областей в геноме должно быть больше второго порогового значения. Пороговое значение может представлять собой процент областей, для которых было определено присутствие делеции или амплификации.

Указанное пороговое значение для количества областей может также зависеть от того, насколько выражен дисбаланс для подсчитываемых областей. Например, количество областей, используемое в качестве порогового для определения классификации рака, может зависеть от специфичности и чувствительности (aberrантный порог) для детекции aberrации в каждой области. Например, при низком aberrантном пороге (например, Z-показатель равен 2) может быть выбрано высокое пороговое количество (например, 150). Однако при высоком aberrантном пороге (например, Z-показатель равен 3) пороговое количество может быть более низким (например, 50). Количество областей, демонстрирующих aberrацию, может также представлять собой взвешенную величину, например, одной области, которая демонстрирует высокий дисбаланс, может быть придан больший вес, чем области, которая демонстрирует только незначительный дисбаланс (т.е. классификаций больше, чем просто положительные и отрицательные результаты на наличие aberrации). Например, может использоваться сумма Z-показателей, с применением, таким образом, взвешенных значений.

Соответственно, количество (которое может включать число и/или размер) хромосомных областей, демонстрирующих значительную избыточную или недостаточную представленность подсчитанных меток нормировки (или другое соответствующее значение для свойства группы) может использоваться для отражения тяжести заболевания. Количество хромосомных областей с aberrантным подсчитанным числом меток нормировки может определяться по двум факторам, а именно, числу (или размеру) хромосомных aberrаций в опухолевых тканях и фракционной концентрации происходящей из опухоли ДНК в биологическом образце (например, в плазме). При более распространенном раке наблюдается тенденция к большему количеству (и бóльших) хромосомных aberrаций. Соответственно, в образце (например, плазме) потенциально можно детектировать больше связанных с раковым заболеванием хромосомных aberrаций. У пациентов с более распространенным раком более высокая опухолевая нагрузка будет приводить к более высокой фракционной концентрации происходящей из опухоли ДНК в плазме. В результате ассоциированные с опухолью хромосомные

аберрации в образце плазмы проще детектировать.

Один из возможных способов повышения чувствительности без ущерба для специфичности заключается в учете результата для смежного хромосомного сегмента. Согласно одному варианту реализации точки отсечения для Z-показателя остаются на уровне >2 и <-2 . Однако хромосомную область классифицируют как потенциально aberrантную только в том случае, если в двух последовательных сегментах наблюдается один и тот же тип aberrаций, например, оба сегмента имеют Z-показатель >2 . Согласно другим вариантам реализации Z-показатели соседних сегментов могут быть суммированы при использовании более высокого предельного значения. Например, могут быть суммированы Z-показатели трех последовательных сегментов, и может применяться предельное значение, равное 5. Указанный принцип может быть распространен на более чем три последовательных сегмента.

Комбинирование количественных и aberrантных порогов может также зависеть от цели анализа, и любых предварительных данных относительно организма (или их отсутствия). Например, при скрининге нормальной здоровой популяции на наличие раковых заболеваний, как правило, используется высокая специфичность, потенциально как в отношении количества областей (т.е. высокий порог для числа областей), так и в отношении aberrантного порога при идентификации области как содержащей aberrацию. Однако для пациента с более высоким риском (например, для пациента с жалобами на наличие опухоли или пациента с семейным анамнезом, курильщика, носителя хронической инфекции папилломавируса человека (HPV), носителя вируса гепатита или другого вируса) пороги могут быть более низкими, чтобы обеспечить бóльшую чувствительность (меньшее количество ложноотрицательных результатов).

Согласно одному варианту реализации при использовании для детекции хромосомной aberrации разрешения 1 Мб и нижнего предела детекции при 6,3% происходящей из опухоли ДНК, число молекул в каждом сегменте размером 1 Мб должно составлять 60 000. При трансляции это даст приблизительно 180 млн. ($60\,000$ ридов/Мб \times 3000 Мб) выравниваемых ридов для полного генома.

Меньший размер сегмента обеспечивает более высокое разрешение для детекции меньших хромосомных aberrаций. Однако это увеличивает необходимое для анализа общее число молекул. Большой размер сегмента снижает требуемое для анализа число молекул в ущерб разрешению. Соответственно, могут быть обнаружены только aberrации большего размера. Согласно одному варианту реализации могут использоваться области большего размера, демонстрирующие aberrацию сегменты могут быть разделены и полученные подобласти проанализированы для получения лучшего разрешения

(например, согласно описанию выше). При наличии расчетных данных для размера детектируемой делеции или амплификации (или минимальной концентрации для детекции) может быть определено число молекул для анализа.

В. CNA на основе секвенирования обработанной бисульфитом ДНК плазмы

Полногеномное гипометилирование и CNA часто наблюдаются в опухолевых тканях. В данном описании авторы настоящего изобретения демонстрируют, что информация о CNA и связанных с раковым заболеванием изменениях метилирования может быть одновременно получена при бисульфитном секвенировании ДНК плазмы. Поскольку указанные два типа анализа могут быть проведены на одном наборе данных, практически отсутствуют дополнительные расходы на анализ CNA. Согласно другим вариантам реализации могут проводиться отдельные процедуры для получения информации о метилировании и генетической информации. Согласно другим вариантам реализации может проводиться аналогичный анализ связанного с раковым заболеванием гиперметилирования в сочетании с анализом CNA.

На фиг. 37А представлен анализ CNA для опухолевых тканей, на не обработанной бисульфитом (БС) ДНК плазмы и обработанной бисульфитом ДНК плазмы (от центра к краям), для пациента TBR36. На фиг. 37А представлен анализ CNA для опухолевых тканей, на не обработанной бисульфитом (БС) ДНК плазмы и обработанной бисульфитом ДНК плазмы (от центра к краям), для пациента TBR36. Внешний круг представляет собой идеограмму хромосомы. Каждая точка представляет результат для области объемом 1 Мб. Зелеными, красными и серыми точками обозначены области с добавлением копий, утратой копий и без изменения числа копий, соответственно. Приведены Z-показатели для анализа плазмы. Разность между двумя концентрическими линиями составляет 5. Показано число копий для анализа опухолевой ткани. Разность между двумя концентрическими линиями составляет 1 копию. На фиг. 38А представлен анализ CNA для опухолевых тканей, на не обработанной бисульфитом (БС) ДНК плазмы и обработанной бисульфитом ДНК плазмы (от центра к краям), для пациента TBR34. Паттерны CNA, детектированные в обработанных бисульфитом и не обработанных бисульфитом образцах плазмы, согласовались.

Паттерны CNA, детектированные в опухолевых тканях, не обработанной бисульфитом плазме и обработанной бисульфитом плазме, согласовались. Для дальнейшей оценки соответствия между результатами для обработанной бисульфитом и не обработанной бисульфитом плазмы строили график рассеяния. Фиг. 37В представляет собой график рассеяния, отражающий зависимость между Z-показателями для детекции CNA на участках размером 1 Мб с применением обработанной бисульфитом и не

обработанной бисульфитом плазмы, для пациента TBR36. Наблюдалась положительная корреляция между Z-показателями указанных двух анализов ($r=0,89$, $p<0,001$, корреляция Пирсона). Фиг. 38В представляет собой график рассеяния, отражающий зависимость между Z-показателями для детекции CNA на участках размером 1 Мб с применением обработанной бисульфитом и не обработанной бисульфитом плазмы, для пациента TBR34. Наблюдалась положительная корреляция между Z-показателями указанных двух анализов ($r=0,81$, $p<0,001$, корреляция Пирсона).

С. Синергетический анализ связанных с раковым заболеванием CNA и изменений метилирования

Согласно приведенному выше описанию анализ на CNA может включать подсчет числа ридов секвенирования в каждой области размером 1 Мб, а анализ плотности метилирования может включать выявление доли метилированных остатков цитозина в динуклеотидах CpG. Комбинирование указанных двух анализов может обеспечить синергетическую информацию для детекции раковых заболеваний. Например, классификация метилирования и классификация CNA могут использоваться для определения третьей классификации уровня рака.

Согласно одному варианту реализации наличие либо связанных с раковым заболеванием CNA, либо изменений метилирования может использоваться для установления потенциального наличия ракового заболевания. Согласно такому варианту реализации чувствительность детекции ракового заболевания может быть повышена при наличии либо CNA, либо изменений метилирования в плазме тестируемого субъекта. Согласно другому варианту реализации наличие обоих изменений может применяться для установления наличия ракового заболевания. Согласно такому варианту реализации специфичность теста может быть повышена, так как изменения любого из указанных двух типов потенциально могут детектироваться у некоторых не имеющих раковых заболеваний субъектов. Соответственно, третья классификация может быть положительной по раковому заболеванию только в том случае, если и первая классификация, и вторая классификация указывают на раковое заболевание.

В исследование включали 26 пациентов с ГЦК и 22 здоровых субъектов. У каждого субъекта брали образец крови и секвенировали ДНК плазмы после обработки бисульфитом. У пациентов с ГЦК образцы крови брали в момент установления диагноза. Наличие значимого количества CNA определяли, например, как присутствие $>5\%$ участков, демонстрирующих Z-показатель <-3 или >3 . Наличие значимого связанного с раковым заболеванием гипометилирования определяли как присутствие $>3\%$ участков, демонстрирующих Z-показатель <-3 . Количество областей (участков) может быть

представлено, например, в виде подсчитанного без обработки числа участков, процента и длины указанных участков.

В таблице 3 показана детекция значимых количеств CNA и изменений метилирования в плазме 26 пациентов с ГЦК при применении массивно-параллельного секвенирования обработанной бисульфитом ДНК плазмы.

Таблица 3

		CNA	
		Наличие	Отсутствие
Изменение метилирования	Наличие	12	6
	Отсутствие	1	7

Уровни детекции связанного с раковым заболеванием изменения метилирования и CNA составляли 69% и 50%, соответственно. Уровень детекции (т.е. диагностическая чувствительность) повышался до 73% при использовании присутствия любого показателя для установления потенциального наличия ракового заболевания.

Показаны результаты для двух пациентов, отражающие либо наличие CNA (фиг. 39А), либо изменения метилирования (фиг. 39В). Фиг. 39А представляет собой Circos-график, представляющий CNA (внутренний круг) и анализ метилирования (внешний круг) для обработанной бисульфитом плазмы пациента с ГЦК TBR240. В анализе CNA зелеными, красными и серыми точками обозначены области с добавлениями хромосомного материала, утратами хромосомного материала, а также области без изменения числа копий, соответственно. В анализе метилирования зелеными, красными и серыми точками обозначены области с гиперметилированием, гипометилированием и нормальным метилированием, соответственно. У указанного пациента в плазме детектировали связанные с раковым заболеванием CNA, тогда как анализ метилирования не обнаружил значительного связанного с раковым заболеванием гипометилирования. Фиг. 39В представляет собой Circos-график, представляющий CNA (внутренний круг) и анализ метилирования (внешний круг) для обработанной бисульфитом плазмы пациента с ГЦК TBR164. У указанного пациента детектировали связанное с раковым заболеванием гипометилирование в плазме. Однако значительных количеств CNA не наблюдалось. Эти результаты для двух пациентов, отражающие наличие как CNA, так и изменений метилирования, представлены на фиг. 48А (TBR36) и 49А (TBR34).

В таблице 4 показана детекция значительных количеств CNA и изменений метилирования в плазме 22 контрольных субъектов с применением массивно-параллельного секвенирования обработанной бисульфитом ДНК плазмы. Для оценки каждого из контрольных субъектов использовали метод бутстреппинга (т.е. с

исключением по одному образцу). Соответственно, при оценке конкретного субъекта остальные субъекты (21 субъект) использовались для расчета среднего и SD для контрольной группы.

Таблица 4

		CNA	
		Наличие	Отсутствие
Изменение метилирования	Наличие	1	2
	Отсутствие	1	18

Специфичность детекции значительных изменений метилирования и CNA составляла 86% и 91%, соответственно. Специфичность повышалась до 95%, если требовалось присутствие обоих показателей для установления потенциального наличия ракового заболевания.

Согласно одному варианту реализации образцы, положительные по CNA и/или гипометилированию считают положительными по раковому заболеванию, и образцы, где не детектируется ни то, ни другое, считают отрицательными. Применение логической схемы «или» обеспечивает более высокую чувствительность. Согласно другому варианту реализации только положительные и по CNA, и по гипометилированию образцы считают положительными по раковому заболеванию, таким образом обеспечивается более высокая специфичность. Согласно еще одному варианту реализации может использоваться три уровня классификации. Субъектов классифицируют по наличию: i. обоих нормальных показателей; ii. одного аномального показателя; iii. обоих аномальных показателей.

Для указанных трех классификаций могут применяться разные стратегии последующего исследования. Например, для субъектов (iii) может использоваться наиболее интенсивный протокол последующего исследования, например, включающий визуализацию всего организма; для субъектов (ii) может использоваться менее интенсивный протокол последующего исследования, например, повторные секвенирования ДНК плазмы через относительно короткий промежуток времени, составляющий несколько недель; и для субъектов (i) может использоваться наименее интенсивный протокол последующего исследования, например, повторное тестирование через несколько лет. Согласно другим вариантам реализации измерения метилирование и CNA может применяться в сочетании с другими клиническими параметрами (например, результатами визуализации или биохимическими показателями сыворотки) для дополнительного уточнения классификации.

D. Прогностическая ценность анализа ДНК плазмы после направленного на излечение лечения

Наличие связанных с раковым заболеванием CNA и/или изменений метилирования в плазме указывает на наличие происходящей из опухоли ДНК в кровотоке больного раковым заболеванием пациента. После лечения (например, хирургического) ожидается уменьшение или исчезновение указанных связанных с раковым заболеванием изменений. С другой стороны, персистенция указанных изменений в плазме после лечения может указывать на неполное удаление всех опухолевых клеток из организма и может подходить для прогнозирования рецидива заболевания.

Образцы крови брали у двух пациентов с ГЦК, TBR34 и TBR36, через неделю после направленного на излечение хирургического удаления опухолей. Анализ CNA и метилирования проводили на обработанных бисульфитом образцах послеоперационной плазмы.

На фиг. 40А показан анализ CNA на обработанной бисульфитом ДНК плазмы, собранной до (внутренний круг) и после (внешний круг) хирургического удаления опухоли у пациента с ГЦК, TBR36. Каждая точка представляет результат для области размером 1 Мб. Зелеными, красными и серыми точками обозначены области с увеличением числа копий, утратой копий и без изменений числа копий, соответственно. Большинство CNA, наблюдавшихся до лечения, исчезли после удаления опухоли. Доля участков, демонстрирующих Z-показатель <-3 или >3 , снижалась от 25% до 6,6%.

На фиг. 40В показан анализ метилирования обработанной бисульфитом ДНК плазмы собранной до (внутренний круг) и после (внешний круг) хирургического удаления опухоли у пациента с ГЦК TBR36. Зелеными, красными и серыми точками обозначены области с гиперметилированием, гипометилированием и нормальным метилированием, соответственно. Наблюдалось заметное снижение доли участков, демонстрирующих значительное гипометилирование, с 90% до 7,9%, и наблюдалось также заметное снижение степени гипометилирования. У указанного пациента наступила полная клиническая ремиссия через 22 месяца после удаления опухоли.

На фиг. 41А показан анализ CNA обработанной бисульфитом ДНК плазмы собранной до (внутренний круг) и после (внешний круг) хирургического удаления опухоли у пациента с ГЦК, TBR34. Хотя наблюдается снижение как числа участков, демонстрирующих CNA, так и величины CNA в пораженных участках после хирургического удаления опухоли, можно наблюдать остаточные CNA в образце послеоперационной плазмы. Красным кругом обозначена область, где наиболее очевидно выражены остаточные CNA. Доля участков, демонстрирующих Z-показатель <-3 или >3 , снижалась от 57% до 12%.

На фиг. 41В. показан анализ метилирования обработанной бисульфитом ДНК

плазмы собранной до (внутренний круг) и после (внешний круг) хирургического удаления опухоли у пациента с ГЦК TBR34. Величина гипометилирования уменьшалась после удаления опухоли, при уменьшении среднего значения Z-показателя для гипометилированных участков от -7,9 до -4,0. Однако для доли участков с Z-показателем <-3 наблюдалось противоположное изменение, с увеличением от 41% до 85%. Указанное наблюдение потенциально указывает на присутствие остаточных раковых клеток после лечения. Клинически было обнаружено несколько очагов опухолевых узлов в нерезецированной печени через 3 месяца после удаления опухоли. С 4-го месяца после хирургического вмешательства наблюдались метастазы в легкие. Указанный пациент скончался в результате локального рецидива и метастазирования через 8 месяцев после операции.

Наблюдения в случаях указанных двух пациентов (TBR34 и TBR36) предполагают, что наличие остаточных связанных с раковым заболеванием изменений CNA и гипометилирования может использоваться для мониторинга и прогнозирования у больных раковыми заболеваниями пациентов после направленного на излечение лечения. Данные также показали, что степень изменения количества детектируемых CNA в плазме может использоваться синергетически с оценкой степени изменения уровня гипометилирования ДНК плазмы для прогнозирования и мониторинга эффективности лечения.

Соответственно, согласно некоторым вариантам реализации один биологический образец получают до лечения и второй биологический образец получают после лечения (например, хирургического). Для первого образца получают первые значения, например, Z-показатели для областей (например, уровни метилирования области и нормированные значения CNA), и число областей, демонстрирующих гипометилирование и CNA (например, амплификацию или делецию). Для второго образца могут быть получены вторые значения. Согласно другому варианту реализации после лечения может быть получен третий или даже последующие дополнительные образцы. Число областей, демонстрирующих гипометилирование и CNA (например, амплификацию или делецию) может быть получено для третьего или даже последующих дополнительных образцов.

Согласно приведенному выше описанию для фиг. 40А и 41А первое число областей, демонстрирующих гипометилирование, для первого образца может сравниваться со вторым количеством областей, демонстрирующих гипометилирование, для второго образца. Согласно приведенному выше описанию для фиг. 40В и 41В первое количество демонстрирующих гипометилирование областей для первого образца, может сравниваться со вторым количеством демонстрирующих гипометилирование областей для второго образца. Сравнение указанного первого количества с указанным вторым

количеством, и первого числа со вторым числом может использоваться для определения прогноза лечения. Согласно различным вариантам реализации только одно из сравнений может быть определяющим для прогноза, либо могут использоваться оба сравнения. Согласно вариантам реализации, где получают третий или последующие дополнительные образцы, один или большее количество указанных образцов может применяться для определения прогноза лечения, либо сами по себе, либо в сочетании со вторым образцом.

Согласно одному варианту реализации делают более плохой прогноз, если первая разность между первым количеством и вторым количеством ниже порога для первой разности. Согласно другому варианту реализации делают более плохой прогноз, если вторая разности между первым числом и вторым числом ниже порога для второй разности. Порог может быть одинаковым или разным. Согласно одному варианту реализации порог для первой разности и порог для второй разности равны 0. Соответственно, в примере, приведенном выше, разность значений метилирования указывает на худший прогноз для пациента TBR34.

Прогноз может быть лучше, если первая разность и/или вторая разность выше одного и того же порога либо соответствующих порогов. Классификация для прогноза может зависеть от того, насколько выше или ниже порога имеющиеся различия. Для получения различных классификаций может использоваться несколько порогов. При более значительных различиях могут прогнозироваться лучшие исходы, а при меньших различиях (и даже отрицательных значениях) могут прогнозироваться худшие исходы.

Согласно некоторым вариантам реализации могут также учитываться моменты времени, когда отбирают различные образцы. При наличии таких временных параметров возможно определить кинетику или скорость изменения количества. Согласно одному варианту реализации быстрое снижение ассоциированного с опухолью гипометилирования в плазме и/или быстрое уменьшение ассоциированных с опухолью CNA в плазме указывает на хороший прогноз. И напротив, статичность или быстрое повышение ассоциированного с опухолью гипометилирования в плазме и/или статичность или быстрое увеличение ассоциированных с опухолью CNA указывает на плохой прогноз. Измерения метилирования и CNA могут использоваться в комбинации с другими клиническими параметрами (например, результатами визуализации, биохимическими показателями сыворотки или белковыми маркерами) для прогнозирования клинического исхода.

Согласно вариантам реализации могут использоваться другие образцы помимо плазмы. Например, ассоциированные с опухолью аберрации метилирования (например, гипометилирование) и/или ассоциированные с опухолью CNA могут быть измерены в

опухолевых клетках, циркулирующих в крови страдающих раковым заболеванием пациентов, в свободной от клеток ДНК или опухолевых клетках в моче, стуле, слюне, мокроте, билиарной жидкости, жидкости поджелудочной железы, шеечных мазках, секрете половых путей (например, влагалища), асцитической жидкости, плевральной жидкости, семенной жидкости, поте и слезе.

Согласно различным вариантам реализации ассоциированные с опухолью аберрации метилирования (например, гипометилирование) и/или ассоциированные с опухолью CNA могут быть детектированы в крови или плазме пациентов с раком молочной железы, раком легкого, раком толстой и прямой кишки, раком поджелудочной железы, раком яичника, носоглоточной карциномой, раком шейки матки, меланомой, опухолями головного мозга и т.д. Таким образом, поскольку метилирование и генетические изменения, такие как CNA, представляют собой универсальное явление при раковых заболеваниях, описанные способы могут применяться для всех типов рака. Измерения метилирования и CNA могут использоваться в сочетании с другими клиническими параметрами (например, результатами визуализации) для прогнозирования клинического исхода. Варианты реализации также могут применяться для скрининга и мониторинга пациентов с предраковыми состояниями, например, аденомами.

Соответственно, согласно одному варианту реализации указанный биологический образец берут до лечения, и повторяют измерения CNA и метилирования после лечения. Указанные измерения могут давать следующее первое количество областей, для которых было определено присутствие делеции или амплификации, и могут давать следующее второе количество областей, которые, как было определено, имеют уровень метилирования области, превышающий соответствующее предельное значение для области. Первое количество может сравниваться со следующим первым количеством, и второе количество может сравниваться со следующим вторым количеством для определения прогноза для указанного организма.

Сравнение для определения прогноза для указанного организма может включать определение первой разности между первым количеством и следующим первым количеством, и указанная первая разность может сравниваться с одним или несколькими порогами для первой разности для определения прогноза. Сравнение для определения прогноза для указанного организма также может включать определение второй разности между вторым количеством и следующим вторым количеством, и указанная вторая разность может сравниваться с одним или несколькими порогами для второй разности. Пороговое значение может быть равно 0 или другому числу.

Прогноз может быть хуже, если первая разность ниже порога для первой разности,

по сравнению с прогнозом, если первая разность выше порога для первой разности. Прогноз может быть хуже, если вторая разность ниже порога для второй разности, по сравнению с прогнозом, если вторая разность выше порога для второй разности. Примеры лечения включают иммунотерапию, хирургию, радиационную терапию, химиотерапию, терапию на основе антител, генную терапию, эпигенетическую терапию или таргетную терапию.

Е. Эффективность

Ниже описывается диагностическая эффективность анализа CNA и метилирования для разного числа ридов секвенирования и участков разного размера.

1. Число ридов секвенирования

В соответствии с одним вариантом реализации авторы настоящего изобретения исследовали ДНК плазмы 32 здоровых контрольных субъектов, 26 пациентов, страдающих гепатоцеллюлярной карциномой и 20 пациентов, страдающих другими типами раковых заболеваний, включая носоглоточную карциному, рак молочной железы, рак легкого, нейроэндокринный рак и гладкомышечную саркому. 22 из 32 здоровых субъектов были выбраны случайным образом в качестве референсной группы. Среднее значение и стандартное отклонение (SD) для указанных 22 референсных индивидуумов использовали для определения нормального диапазона плотности метилирования и представленности в геноме. ДНК, экстрагированную из образца плазмы каждого индивидуума, использовали для получения библиотеки секвенирования с применением набора для секвенирования парных концов от Illumina. Затем библиотеки для секвенирования обрабатывали бисульфитом, которые конвертирует неметилированные остатки цитозина в урацил. Конвертированную бисульфитом библиотеку для секвенирования для каждого образца плазмы секвенировали на одной дорожке секвенатора Illumina HiSeq2000.

После распознавания оснований адаптерных последовательности и основания низкого качества (т.е. с показателем качества < 5) на концах фрагментов удаляли. Затем обрезанные риды в формате FASTQ обрабатывали с помощью системы анализа данных метилирования Methy-Pipe (P Jiang et al. 2010, IEEE International Conference on Bioinformatics and Biomedicine, ЦИО:10.1109/BIBMW.2010.5703866). Для выравнивания конвертированных бисульфитом ридов секвенирования авторы настоящего изобретения сначала выполняли конверсию *in silico* всех остатков цитозина в тимины на цепях Уотсона-Крика по отдельности с применением референсного генома человека (версия NCBI 36/hg19). Затем авторы настоящего изобретения выполняли конверсию *in silico* каждого цитозина в тимин во всех обработанных ридах и сохраняли информацию о

расположении каждого конвертированного остатка. SOAP2 использовали для выравнивания конвертированных ридов по двум предварительно конвертированным референсным геномам человека (R Li et al. 2009 *Bioinformatics* 25:1966-1967), максимум с двумя несовпадениями для каждого выравниваемого рида. Только риды, картирующиеся с уникальной геномной локализацией, использовали для дальнейшего анализа. Неоднозначные риды, картирующиеся на обе цепи Уотсона-Крика и дублирующиеся (клональные) риды удаляли. Остатки цитозина в контексте динуклеотидов CpG использовали для последующего анализа метилирования. После выравнивания цитозины, исходно присутствовавшие в ридах секвенирования, восстанавливали на основе информации о расположении, сохраненной при конверсии *in silico*. Восстановленные цитозины в динуклеотидах CpG подсчитывали как метилированные. Тимины в динуклеотидах CpG подсчитывали как неметилированные.

Для анализа метилирования геном разделяли на участки равного размера. Протестированные размеры участков включают 50 кб, 100 кб, 200 кб и 1 Мб. Плотность метилирования для каждого участка рассчитывали как число метилированных цитозинов в контексте динуклеотидов CpG, разделенное на общее число цитозинов в положениях CpG. Согласно другим вариантам реализации размер участков может не быть одинаковым по всему геному. Согласно одному варианту реализации каждый участок из таких участков неодинакового размера сравнивают у нескольких субъектов.

Для определения того, является ли нормальной плотность метилирования в плазме в тестируемом случае, плотность метилирования сравнивали с результатами референсной группы. 22 из 32 здоровых субъектов были выбраны случайным образом в качестве референсной группы для расчета Z-показателя метилирования (Z_{meth}).

$$Z_{meth} = \frac{MD_{test} - \overline{MD}_{ref}}{MD_{SD}}$$

где MD_{test} – плотность метилирования в тестируемом случае для конкретного участка размером 1 Мб; \overline{MD}_{ref} – среднее значение плотности метилирования референсной группы для соответствующего участка; и MD_{SD} – SD плотности метилирования референсной группы для соответствующего участка.

Для анализа CNA определяли число ридов секвенирования, картируемых с каждым участком размером 1 Мб (KCA Chan et al. 2013 *Clin Chem* 59:211-24). Плотность ридов секвенирования определяли для каждого участка после коррекции смещения по GC с применением локально взвешенной регрессии для сглаживания графика рассеяния согласно существующему описанию (EZ Chen et al. 2011 *PLoS One* 6: e21791). Для анализа

плазмы плотность ридов секвенирования в тестируемом случае сравнивали с референсной группой для расчета Z-показателя CNA (Z_{CNA}):

$$Z_{CNA} = \frac{RD_{test} - \overline{RD}_{ref}}{RD_{SD}}$$

где RD_{test} – плотность ридов секвенирования для тестируемого случая для конкретного участка размером 1 Мб; \overline{RD}_{ref} – среднее значение плотности ридов секвенирования для референсной группы для соответствующего участка; и RD_{SD} – SD плотность ридов секвенирования для референсной группы для соответствующего участка. Участок определяли как демонстрирующий CNA, если Z_{CNA} указанного участка был <-3 или >3 .

В расчете на один случай получали среднее значение выравниваемых ридов, составляющее 93 млн. (диапазон: 39 млн.–142 млн.). Для оценки эффекта уменьшения числа ридов секвенирования на диагностическую эффективность авторы настоящего изобретения случайным образом выбирали 10 млн. выравниваемых ридов в каждом случае. Один и тот же набор референсных индивидуумов использовали для установления референсного диапазона для каждого участка размером 1 Мб для набора данных с уменьшенным количеством ридов секвенирования. Для каждого случая определяли процент участков, демонстрирующих значительное гипометилирование, т.е. $Z_{meth} < -3$, и процент участков с CNA, т.е. $Z_{CNA} < -3$ или >3 . Кривые функциональных характеристик приемника (ROC) использовали для иллюстрации диагностической эффективности полногеномного анализа гипометилирования и CNA для наборов данных со всеми ридами секвенирования с 1 дорожки и 10 млн. ридов на случай. При анализе ROC использовали всех здоровых субъектов (32 субъекта).

На фиг. 42 приведена диаграмма диагностической эффективности полногеномного анализ гипометилирования при разном количестве ридов секвенирования. Для анализа гипометилирования значения площади под кривой для кривых ROC значительно не отличались между двумя наборами данных, которые исследовали все риды секвенирования с одной дорожки и 10 млн. ридов на случай ($P=0,761$). Диагностическая эффективность анализа CNA ухудшалась, со значительным снижением значений площади под кривой, при уменьшении числа ридов секвенирования от данных одной дорожки до 10 млн. ($P<0,001$).

2. Эффект применения участков различного размера

Помимо деления генома на участки размером 1 Мб авторы настоящего изобретения также исследовали возможность использования участков меньшего размера.

Теоретически, применение участков меньшего размера потенциально может уменьшать вариабельность плотности метилирования на участке. Это обусловлено тем, что плотность метилирования разных геномных областей может широко варьировать. Если размер участка больше, вероятности включения областей с разными плотностями метилирования повышаются и, соответственно, приводят к общему увеличению вариабельности плотности метилирования участков.

Хотя использование участков меньшего размера потенциально может уменьшать вариабельность плотности метилирования, связанную с различиями между областями, с другой стороны, при этом снизится число ридов секвенирования, картируемых на конкретный участок. При уменьшении картирования ридов на индивидуальные участки будет повышаться вариабельность за счет вариаций выборки. Оптимальный размер участка, способный обеспечить минимальную общую вариабельность плотности метилирования, может быть установлен экспериментально с учетом требований при конкретном диагностическом применении, например, с учетом общего числа ридов секвенирования на образец и типа используемого секвенатора ДНК.

На фиг. 43 приведена диаграмма, демонстрирующая кривые ROC для детекции ракового заболевания на основе полногеномного анализа гипометилирования при разных размерах участков (50 кб, 100 кб, 200 кб и 1 Мб). Показаны P-значения для сравнения площади под кривой при размере участка 1 Мб. Наблюдается тенденция к улучшению при уменьшении размера участка от 1 Мб до 200 кб.

F. Показатель суммарной вероятности

Количество областей метилирования и CNA может принимать различные значения. В приведенных выше примерах описано число областей, превышающее предельное значение или процент таких областей, демонстрирующих значительное гипометилирование или CNA, в качестве параметра классификации образца как связанного с раковым заболеванием. Такие способы не учитывают величину aberrации для индивидуальных участков. Например, участок с Z_{meth} , равным -3,5, будет классифицирован так же, как участок с Z_{meth} , равным -30, так как оба они будут классифицированы как значительно гипометилированные. Однако на степень изменений гипометилирования в плазме, т.е. на величину значения Z_{meth} , влияет количество связанной с раковым заболеванием ДНК в образце и, соответственно, она может дополнить информацию о проценте участков, демонстрирующих aberrации, отражая опухолевую нагрузку. Более высокая фракционная концентрация опухолевой ДНК в образце плазмы будет приводить к более низкой плотности метилирования, и это будет обуславливать меньшее значение Z_{meth} .

1. Показатель суммарной вероятности как диагностический параметр

Для использования информации о величине aberrаций авторы настоящего изобретения разработали способ, называемый показателем суммарной вероятности (CP). На основании нормального распределения вероятностной функции каждое значение Z_{meth} переводили в вероятность случайного возникновения такого наблюдения.

Показатель CP рассчитывали как:

$$CP\ score = \sum -\log(\text{Prob}_i)$$

для участка (i) с $Z_{meth} < -3$

где Prob_i представляет собой вероятность Z_{meth} для участка (i) в соответствии с t -распределением Стьюдента с 3 степенями свободы, и \log представляет собой функцию натурального логарифма. Согласно другому варианту реализации может использоваться логарифм по основанию 10 (или другому числу). Согласно другим вариантам реализации могут использоваться другие распределения, например, но не ограниченные перечисленными нормальным распределением и гамма-распределением, для преобразования Z-показателя в CP.

Большой показатель CP указывает на более низкую вероятность случайного возникновения такого отклонения плотности метилирования в нормальной популяции. Таким образом, высокий показатель CP будет указывать на более высокую вероятность аномального гипометилирования ДНК в образце, например, на присутствие связанной с раковым заболеванием ДНК.

По сравнению с процентом участков, демонстрирующих aberrации, измеренный показатель CP имеет более высокий динамический диапазон. Хотя опухолевая нагрузка у разных пациентов может варьировать в широком диапазоне, большой диапазон значений CP может подходить для отражения опухолевой нагрузки у пациентов с относительно высокой и относительно низкой опухолевой нагрузкой. Кроме того, применение показателей CP может потенциально обеспечивать более чувствительную детекцию изменений концентрации связанной с опухолью ДНК в плазме. Это обеспечивает преимущества при мониторинге ответа на лечение и прогнозировании. Соответственно, снижение показателей CP во время лечения указывает на хороший ответ на лечение. Отсутствие снижения или даже увеличение показателей CP во время лечения указывает на неудовлетворительный или отсутствующий ответ. При прогнозировании высокий CP показатель указывает на высокую опухолевую нагрузку и предполагает плохой прогноз (например, более высокую вероятности смерти или прогрессирования опухоли).

На фиг. 44А показана диагностическая эффективность суммарной вероятности (CP) и процентного содержания участков с aberrациями. Не наблюдалось значимой разности

между значениями площади под кривой для двух типов диагностического алгоритма ($P=0,791$).

На фиг. 44В показана диагностическая эффективность анализа плазмы на глобальное гипометилирование, гиперметилирование CpG-островков и CNA. При одной дорожке секвенирования на образец (размер участков 200 кб для анализа гипометилирования и размер участков 1 Мб для CNA; CpG-островки определены в соответствии с базой данных Университета Калифорнии, Санта-Крус (UCSC)) значения площади под кривой для всех трех типов анализа превышали 0,90.

Во всех последующих анализах максимальный показатель CP для контрольных субъектов использовали в качестве точки отсечения для каждого из трех типов анализа. Выбор указанных точек отсечения обеспечивал диагностическую специфичность, составляющую 100%. Диагностическая чувствительность для анализа общего гипометилирования, гиперметилирования CpG-островков и CNA составляла 78%, 89% и 52%, соответственно. У 43 из 46 больных раковыми заболеваниями пациентов детектировали aberrации по меньшей мере одного из трех типов, что, соответственно, давало чувствительность 93,4% и специфичность 100%. Наши результаты показывают, что указанные три типа анализа могут использоваться синергетически для детекции раковых заболеваний.

На фиг. 45 приведена таблица результатов анализа на глобальное гипометилирование, гиперметилирование CpG-островков и CNA у пациентов с гепатоцеллюлярной карциномой. Предельные значения показателя CP для трех типов анализа составляли 960, 2,9 и 211, соответственно. Положительные результаты для показателя CP выделены жирным шрифтом и подчеркнуты.

На фиг. 46 приведена таблица результатов анализа на глобальное гипометилирование, гиперметилирование CpG-островков и CNA у пациентов с раковыми заболеваниями, отличными от гепатоцеллюлярной карциномы. Предельные значения показателя CP для трех типов анализа составляли 960, 2,9 и 211, соответственно. Положительные результаты для показателя CP выделены жирным шрифтом и подчеркнуты.

2. Применение показателя CP для мониторинга раковых заболеваний

От пациента с ГЦК TBR34 до и после лечения получали последовательные образцы. Указанные образцы исследовали на глобальное гипометилирование.

На фиг. 47 представлен серийный анализ метилирования в плазме для пациента TBR34. Внутренний круг отражает плотность метилирования лейкоцитарной пленки (черный) и опухолевых тканей (фиолетовый). Для образцов плазмы показан Z_{meth} для

каждого участка размером 1 Мб. Разность между двумя линиями отражает разность Z_{meth} , равную 5. Красными и серыми точками обозначены участки с гипометилированием и без изменений плотности метилирования по сравнению с референсной группой. Начиная от второго от центра круга в направлении внешнего круга представлены образцы плазмы, полученные до лечения, через 3 дня и через 2 месяца после удаления опухоли, соответственно. До лечения в плазме можно было наблюдать высокую степень гипометилирования, и более 18,5% участков имели $Z_{meth} < -10$. Через 3 дня после удаления опухоли можно наблюдать уменьшение степени гипометилирования в плазме при отсутствии участков с $Z_{meth} < -10$.

Таблица 5

Случай №	Момент времени	Анализ метилирования		
		Процент участков, демонстрирующих значительное гипометилирование	Показатель суммарной вероятности (СР)	Суммативный Z-показатель
ТВР34	До хирургического лечения (ОТ)	62,6%	37,573	14,285
	3 дня после ОТ	80,5%	17,777	9,195
	2 месяца после ОТ	40,1%	15,087	5,201

Таблица 5 демонстрирует, что, несмотря на уменьшение величины изменений гипометилирования через 3 дня после хирургического удаления опухоли, процент участков, где наблюдались aberrации, парадоксальным образом повышался. С другой стороны, показатель СР более точно обнаруживал снижение степени гипометилирования в плазме и может лучше отражать изменения опухолевой нагрузки.

Через 2 месяца после ОТ все еще присутствовал значительный процент участков, демонстрирующих изменения гипометилирования. Показатель СР также оставался статичным на уровне, составляющем приблизительно 15 000. Позже, через 3 месяца, у указанного пациента были диагностированы мультифокальные метастазы опухоли (ранее неизвестные на время хирургического вмешательства) в нерезецированной части печени, и было обнаружено несколько метастазов в легкие через 4 месяца после операции. Пациент скончался в результате метастатического процесса через 8 месяцев после операции. Эти результаты предполагают, что показатель СР может быть более эффективным, чем процент участков с aberrациями, для отражения опухолевой нагрузки.

В целом, СР может подходить для применения в тех случаях, когда требуется измерение количества опухолевой ДНК в плазме. Примеры таких применений включают: прогнозирование и мониторинг у страдающих раковым заболеванием пациентов

(например, наблюдение за ответом на лечение, или наблюдение за прогрессированием опухоли).

Суммативный Z-показатель представляет собой прямую сумму Z-показателей, т.е. без преобразования в вероятность. В указанном примере для суммативного Z-показателя наблюдается такое же поведение, как и для показателя CP. В других случаях CP может быть более чувствителен, чем суммативный Z-показатель, для мониторинга остаточного заболевания, благодаря большему динамическому диапазону показателя CP.

Х. Влияние CNA на метилирование

Использование CNA и метилирования для определения соответствующих классификаций уровня рака, с объединением указанных классификаций для получения третьей классификации, описано выше. Помимо такого объединения CNA может использоваться для изменения предельных значений для анализа метилирования и для идентификации ложноположительных результатов путем сравнения уровней метилирования для групп областей с разными характеристиками CNA. Например, уровень метилирования при избыточном содержании (например, $Z_{CNA} > 3$) может сравниваться с уровнем метилирования для нормального содержания (например, $-3 < Z_{CNA} < 3$). Сначала описано влияние CNA на уровни метилирования.

А. Изменение плотности метилирования в областях с добавлением или утратой хромосомного материала

Поскольку в опухолевых тканях обычно наблюдается общее гипометилирование, наличие происходящей из опухоли ДНК в плазме страдающих раковым заболеванием пациентов будет приводить к снижению плотности метилирования в сравнении с не имеющими раковых заболеваний субъектами. Теоретически, степень гипометилирования в плазме страдающих раковым заболеванием пациентов пропорциональна фракционной концентрации происходящей из опухоли ДНК в образце плазмы.

В случае областей, для которых наблюдается добавление хромосомного материала в опухолевых тканях, из амплифицированных сегментов ДНК в плазму высвобождается дополнительная доза опухолевой ДНК. Указанный повышенный вклад опухолевой ДНК в плазму теоретически приводит более высокой степени гипометилирования пораженной области ДНК плазмы. Дополнительным фактором является то, что геномные области, для которых наблюдается амплификация, предположительно стимулируют рост опухолевых клеток, и, соответственно, предположительно будут экспрессироваться. Такие области обычно гипометилированы.

Напротив, для областей, где наблюдается утрата хромосомного материала в опухолевой ткани, пониженный вклад опухолевой ДНК в плазму будет приводить к более

низкой степени гипометилирования по сравнению с областями без изменения числа копий. Дополнительным фактором является то, что геномные области, которые удалены в опухолевых клетках, могут содержать гены-супрессоры опухолей, и для опухолевых клеток может быть благоприятен сайленсинг таких областей. Соответственно, можно ожидать, что вероятность гиперметилирования таких областей будет выше.

В настоящем изобретении авторы используют результаты двух пациентов с ГЦК (TBR34 и TBR36) для иллюстрации указанного эффекта. На фиг. 48А (TBR36) и 49А (TBR34) кругами выделены области с добавлением или утратой хромосомного материала, и представлен соответствующий анализ метилирования. Фиг. 48В и 49В приведены графики z-показателей метилирования при утрате, норме и добавлении для пациентов TBR36 и TBR34, соответственно.

На фиг. 48А представлен Circos-график, отражающий CNA (внутренний круг) и изменения метилирования (внешний круг) в обработанной бисульфитом ДНК плазмы у пациента с ГЦК TBR36. Красными кругами выделены области с добавлением или утратой хромосомного материала. Области, где наблюдается добавление хромосомного материала, были более гипометилированы, чем области без изменения числа копий. Области, где наблюдается утрата хромосомного материала, были менее гипометилированы, чем области без изменения числа копий. Фиг. 48В представляет собой график Z-показателей метилирования для областей с добавлением или утратой хромосомного материала, и областей без изменения числа копий для пациента с ГЦК TBR36. По сравнению с областями без изменений числа копий, области с добавлением хромосомного материала имели более отрицательные Z-показатели (большее гипометилирование), а области с утратой хромосомного материала имели менее отрицательные Z-показатели (меньшее гипометилирование).

На фиг. 49А представлен Circos-график, отражающий CNA (внутренний круг) и изменения метилирования (внешний круг) в обработанной бисульфитом ДНК плазмы у пациента с ГЦК TBR34. Фиг. 49В представляет собой график Z-показателей метилирования для областей с добавлением или утратой хромосомного материала, и областей без изменения числа копий для пациента с ГЦК TBR34. Различие плотности метилирования между областями с добавлением или утратой хромосомного материала было больше у пациента TBR36 по сравнению с пациентом TBR34, так как фракционная концентрация происходящей из опухоли ДНК у первого пациента была выше.

В указанном примере для определения CNA используются те же области, что и для определения метилирования. Согласно одному варианту реализации соответствующие предельные значения для областей зависят от того, наблюдается ли в соответствующей

области делеция или амплификация. Согласно одному варианту реализации величина соответствующего предельного значения для области (например, точка отсечения Z-показателя, используемая для определения гипометилирования) больше в тех случаях, когда в соответствующей области наблюдается амплификация, чем в тех случаях, когда амплификация не наблюдается (например, указанная величина может превышать 3, и может применяться точка отсечения <-3). Соответственно, при тестировании на гипометилирование соответствующее предельное значение для области может иметь большее отрицательное значение в тех случаях, когда в соответствующей области наблюдается амплификация, чем в тех случаях, когда амплификация не наблюдается. Ожидается, что такой вариант реализации повысит специфичность тестирования для детекции раковых заболеваний.

Согласно другому варианту реализации величина соответствующего предельного значения для области меньше (например, менее чем 3) в тех случаях, когда в соответствующей области наблюдается делеция, чем в тех случаях, когда делеция не наблюдается. Соответственно, при тестировании на гипометилирование соответствующее предельное значение для области может быть менее отрицательным в тех случаях, когда в соответствующей области наблюдается делеция, чем в тех случаях, когда делеция не наблюдается. Ожидается, что такой вариант реализации повысит чувствительность тестирования для детекции раковых заболеваний. Коррекция предельных значений согласно описанным выше вариантам реализации может быть изменена в зависимости от требуемой чувствительности и специфичности для конкретного диагностического сценария. Согласно другим вариантам реализации измерения метилирования и CNA могут применяться в сочетании с другими клиническими параметрами (например, результатами визуализации или биохимическими показателями сыворотки) для прогнозирования раковых заболеваний.

В. Использование CNA для выбора областей

Согласно приведенному выше описанию, авторы настоящего изобретения показали, что плотность метилирования в плазме изменяется в областях, содержащих аберрации числа копий, в опухолевых тканях. В областях с увеличением числа копий в опухолевой ткани повышенный вклад гипометилированной опухолевой ДНК в плазму будет приводить к большей степени гипометилирования ДНК плазмы по сравнению с областями без аберраций числа копий. Напротив, в областях с утратой копий в опухолевой ткани пониженный вклад гипометилированной происходящей из ракового новообразования ДНК в плазму будет приводить к меньшей степени гипометилирования ДНК плазмы. Указанная зависимость между плотностью метилирования ДНК плазмы и относительной

представленностью может потенциально использоваться для дифференциации результатов гипометилирования, связанных с присутствием связанной с раковым заболеванием ДНК и с другими не связанными с раковыми заболеваниями причинами (например, СКВ) гипометилирования ДНК плазмы.

Чтобы проиллюстрировать указанный способ, авторы настоящего изобретения исследовали образцы плазмы двух пациентов с гепатоцеллюлярной карциномой (ГЦК) и двух пациентов с СКВ, не имеющих раковых заболеваний. У указанных двух пациентов с СКВ (СКВ04 и СКВ10) наблюдалось очевидное наличие гипометилирования и CNA в плазме. У пациента СКВ04 наблюдалось гипометилирование 84% участков и CNA в 11,2% участков. У пациента СКВ10 наблюдалось гипометилирование 10,3% участков и CNA в 5,7% участков.

На фиг. 50А и 50В представлены результаты анализа гипометилирования плазмы и CNA для пациентов с СКВ, СКВ04 и СКВ10. Внешний круг отображает Z-показатели метилирования (Z_{meth}) при разрешении 1 Мб. Участки с Z-показателем метилирования $Z_{\text{meth}} < -3$ отмечены красным, с $Z_{\text{meth}} > -3$ – серым цветом. Внутренний круг отображает Z-показатели CNA (Z_{CNA}). Зелеными, красными и серыми точками обозначены участки с $Z_{\text{CNA}} > 3$, < 3 и от -3 до 3 , соответственно. У указанных двух пациентов с СКВ наблюдались изменения гипометилирования и CNA в плазме.

Для определения того, согласовались ли изменения метилирования и CNA с наличием происходящей из ракового новообразования ДНК в плазме, авторы настоящего изобретения сравнивали Z_{meth} для областей с $Z_{\text{CNA}} > 3$, < -3 и от -3 до 3 . При изменениях метилирования и CNA, в которые вносит вклад происходящая из ракового новообразования ДНК в плазме, ожидается, что области с $Z_{\text{CNA}} < -3$ будут менее гипометилированы и будут иметь менее отрицательный Z_{meth} . Напротив, области с $Z_{\text{CNA}} > 3$, как ожидается, будут более гипометилированы и будут иметь более отрицательный Z_{meth} . Для наглядности авторы настоящего изобретения применяли односторонний критерий суммы рангов для сравнения Z_{meth} для областей с CNA (т.е. областей с $Z_{\text{CNA}} < -3$ или > 3) с областями без CNA (т.е. областями с Z_{CNA} от -3 до 3). Согласно другим вариантам реализации могут применяться другие статистические критерии, например, но не ограничиваясь t-критерием Стьюдента, дисперсионным анализом (ANOVA) и критерием Краскела–Уоллиса.

На фиг. 51А и 51В представлен анализ Z_{meth} для областей с CNA и без CNA в плазме двух пациентов с ГЦК (TBR34 и TBR36). Области с $Z_{\text{CNA}} < -3$ и > 3 представляют области, недостаточно и избыточно представленные в плазме, соответственно. И у TBR34, и у TBR36 области, которые были недопредставлены в плазме (т.е. области с $Z_{\text{CNA}} < -3$),

имели значительно более высокий Z_{meth} (P -значение $<10^{-5}$, односторонний критерий суммы рангов) по сравнению с областями, нормально представленными в плазме (т.е. областями с Z_{CNA} , составляющими от -3 до 3). Нормальная представленность соответствуют ожидаемой для эуплоидного генома. Представленные избыточно области в плазме (т.е. области с $Z_{\text{CNA}} >3$) имели значительно более низкий Z_{meth} по сравнению с областями, нормально представленными в плазме (P -значение $<10^{-5}$, односторонний критерий суммы рангов). Все указанные изменения согласовались с присутствием гипометилированной опухолевой ДНК в образцах плазмы.

На фиг. 51C и 51D представлен анализ Z_{meth} для областей с CNA и без CNA в плазме двух пациентов с СКВ (СКВ04 и СКВ10). Области с $Z_{\text{CNA}} <-3$ и >3 представляют области, недостаточно и избыточно представленные в плазме, соответственно. У СКВ04 области, которые были недопредставлены в плазме (т.е. области с $Z_{\text{CNA}} <-3$), не имели значительно более высокого Z_{meth} (P -значение = 0,99, односторонний критерий суммы рангов) по сравнению с областями, нормально представленными в плазме (т.е. областями с Z_{CNA} , составляющими от -3 до 3); и области, избыточно представленные в плазме (т.е. области с $Z_{\text{CNA}} >3$) не имели значительно более низкого Z_{meth} по сравнению с областями, нормально представленными в плазме (P -значение = 0,68, односторонний критерий суммы рангов). Указанные результаты отличаются от ожидаемых изменений из-за присутствия в плазме происходящей из опухоли гипометилированной ДНК. Аналогичным образом, у СКВ10 области с $Z_{\text{CNA}} <-3$ не имели значительно более высокого Z_{meth} по сравнению с областями с Z_{CNA} , составляющими от -3 до 3 (P -значение = 0,99, односторонний критерий суммы рангов).

Причина отсутствия типичного связанного с раковым заболеванием паттерна взаимосвязи между Z_{meth} и Z_{CNA} у пациентов с СКВ заключается в том, что у пациентов с СКВ CNA не присутствуют в конкретном типе клеток, в которых также наблюдается гипометилирование. В данном случае, наблюдаемое очевидное присутствие CNA и гипометилирования обусловлено измененным распределением размеров циркулирующей ДНК у пациентов с СКВ. Измененное распределение размеров потенциально может изменять плотности ридов секвенирования для разных геномных областей, что приводит к очевидным CNA, так как эталоны получены от здоровых субъектов. Согласно описанию в предшествующих разделах, имеется корреляция между размером фрагмента циркулирующей ДНК и плотностью его метилирования. Таким образом, измененное распределение размеров также может приводить к aberrантному метилированию.

Хотя области с $Z_{\text{CNA}} >3$ имели немного более низкие уровни метилирования по сравнению с областями с Z_{CNA} , составляющими от -3 до 3, P -значение для указанного

сравнения было значительно выше, чем наблюдавшиеся для двух больных раковым заболеванием пациентов. Согласно одному варианту реализации P -значение может применяться в качестве параметра для определения вероятности наличия в тестируемом случае ракового заболевания. Согласно другому варианту реализации различие Z_{meth} между нормально и aberrантно представленными областями может использоваться в качестве параметра, отражающего вероятность наличия раковых заболеваний. Согласно одному варианту реализации группа страдающих раковым заболеванием пациентов может быть использована для установления корреляции между Z_{meth} и Z_{CNA} , и для определения порогов для разных параметров, что позволяет установить, согласуются ли изменения с присутствием происходящей из ракового новообразования гипометилированной ДНК в тестируемом образце плазмы.

Соответственно, согласно одному варианту реализации может осуществляться анализ CNA для определения первого набора областей, все из которых демонстрируют что-либо одно из: делеции, амплификации или нормальной представленности. Например, в первом наборе областей все области могут демонстрировать делецию, или все области могут демонстрировать амплификацию, или все области могут демонстрировать нормальную представленность (например, содержать нормальное первое количество областей, например, иметь нормальный Z_{meth}). Для указанного первого набора областей может быть определен уровень метилирования (например, первый уровень метилирования согласно способу 2800 может соответствовать первому набору областей).

Анализ CNA позволяет определить второй набор областей, все из которых демонстрируют что-либо второе из: делеции, амплификации или нормальной представленности. В указанном втором наборе областей будут наблюдаться отличные от первого набора признаки. Например, если области первого набора были нормальными, в областях второго набора может наблюдаться делеция или амплификация. Второй уровень метилирования может быть рассчитан на основании соответствующих количеств молекул ДНК, метилированных в сайтах областей второго набора.

Затем может быть вычислен параметр, связывающий первый уровень метилирования и второе метилирование. Например, может вычисляться разность или отношение и сравниваться с предельным значением. Указанная(ое) разность или отношение могут быть также обработаны с применением распределения вероятностей (например, в рамках статистического тестирования) для определения вероятности получения значения, и указанная вероятность может сравниваться с предельным значением для определения уровня рака на основе уровней метилирования. Такая точка отсечения может быть выбрана для различения образцов, относящихся к раковому

заболеванию и не относящихся к раковому заболеванию (например, СКВ).

Согласно одному варианту реализации может быть определен уровень метилирования для первого набора областей или для смеси областей (а именно, смеси областей, демонстрирующих амплификацию, делецию, а также нормальных). Указанный уровень метилирования затем может сравниваться с первой точкой отсечения в рамках первого этапа анализа. При превышении точки отсечения, что указывает на вероятность наличия ракового заболевания, может осуществляться описанный выше анализ для определения, является ли указание ложноположительным. Результирующая классификация уровня рака может, соответственно, включать сравнение параметра для двух уровней метилирования со второй точкой отсечения.

Первый уровень метилирования может представлять собой статистический показатель (например, среднее значение или медиану) уровней метилирования области, рассчитанных для каждой области из первого набора областей. Второй уровень метилирования может также представлять собой статистический показатель уровней метилирования области, рассчитанный для каждой области второго набора областей. Например, статистический показатель может быть определен с применением одностороннего критерия суммы рангов, t-критерия Стьюдента, дисперсионного анализа (ANOVA) или критерия Краскела–Уоллиса.

XI. Классификация типов раковых заболеваний

Помимо определения наличия или отсутствия ракового заболевания в организме варианты реализации позволяют идентифицировать тип ракового заболевания, связанного с образцом. Указанная идентификация типа ракового заболевания может задействовать паттерны глобального гипометилирования, гиперметилирования CpG-островков и/или CNA. Указанные паттерны могут включать разбиение на кластеры пациентов с известным диагнозом с применением измеренных уровней метилирования областей, соответствующих значений CNA для областей и уровня метилирования CpG-островков. Приведенные ниже результаты демонстрируют, что для организмов с аналогичным типом ракового заболевания наблюдаются аналогичные значения для областей и CpG-островков; также и у пациентов, не имеющих раковых заболеваний, наблюдаются аналогичные значения. При разбиении на кластеры каждое из значений для области или островка может представлять собой отдельное измерение при разбиении на кластеры.

Известно, что при одном и том же типе ракового заболевания наблюдаются аналогичные генетические и эпигенетические изменения (E Gebhart et al. 2004 Cytogenet Genome Res; 104: 352-358; PA Jones et al. 2007 Cell; 128: 683-692). Ниже авторы настоящего изобретения приводят описание того, как паттерны CNA и изменений

метиляции, детектированные в плазме, могут использоваться для установления происхождения или типа ракового заболевания. Образцы ДНК плазмы от пациентов с ГЦК, пациентов с отличным от ГЦК заболеванием и здоровых контрольных субъектов классифицировали с применением, например, иерархического кластерного анализа. Указанный анализ выполняли с применением, например, функции `heatmap.2` пакета скриптов R (cran.r-project.org/web/packages/gplots/gplots.pdf).

Чтобы проиллюстрировать потенциал указанного способа, авторы настоящего изобретения использовали два набора критериев (группа А и группа В) в качестве примеров для идентификации подходящих признаков для классификации образцов плазмы (см. таблицу 6). Согласно другим вариантам реализации могут применяться другие критерии для идентификации указанных признаков. Используемые признаки включали глобальные CNA при разрешении 1 Мб, глобальную плотность метилирования при разрешении 1 Мб и метилирование CpG-островков.

Таблица 6

	Глобальное метилирование при разрешении 1 Мб	
	Группа критериев А	Группа критериев В
Критерии	>20 случаев ракового заболевания с Z-показателем >3 или <-3	>20 случаев ракового заболевания с Z-показателем >2,5 или <-2,5
Кол-во идентифицированных признаков	584	1911

	Признаки CNA	
	Группа критериев А	Группа критериев В
Критерии	>10 случаев ракового заболевания с Z-показателем >3 или <-3	>10 случаев ракового заболевания с Z-показателем >2,5 <-2,5
Кол-во идентифицированных признаков	355	759

	Метилирование CpG-островков	
	Группа критериев А	Группа критериев В
Критерии	>5 случаев ракового заболевания с плотностью метилирования, отличающейся от среднего референсного значения на 2% в конкретных CpG-островках	>1 случая ракового заболевания с плотностью метилирования, отличающейся от среднего референсного значения на 2% в конкретных CpG-островках
Кол-во идентифицированных признаков	110	191

В первых двух примерах авторы настоящего изобретения использовали для классификации все признаки CNA, глобального метилирования при разрешении 1 Мб и метилирования CpG-островков. Согласно другим вариантам реализации могут применяться другие критерии, например, но не ограничиваясь точностью измерения признака в плазме референсной группы.

На фиг. 52А представлен иерархический кластерный анализ для образцов плазмы от пациентов с ГЦК, пациентов с отличным от ГЦК заболеванием и здоровых контрольных субъектов с применением всех из 1130 признаков группы А, включая 355 CNA, 584 признаков глобального метилирования при разрешении 1 Мб и статуса метилирования 110 CpG-островков. Цветной диаграммой в верхней части представлены группы образцов: зеленый, синий и красный цвета представляют здоровых субъектов, ГЦК и пациентов с отличным от ГЦК заболеванием, соответственно. В целом, в указанных трех группах субъектов наблюдалась тенденция к совместной кластеризации. По вертикальной оси представлены классифицирующие признаки. Признаки, характеризующиеся аналогичными паттернами у разных субъектов, кластеризовались вместе. Эти результаты предполагают, что паттерны изменений метилирования CpG-островков, изменений полногеномного метилирования при разрешении 1 Мб и CNA в плазме потенциально могут использоваться для определения происхождения ракового заболевания у пациентов при неизвестных первичных очагах.

На фиг. 52В представлен иерархический кластерный анализ для образцов плазмы от пациентов с ГЦК, пациентов с отличным от ГЦК заболеванием и здоровых контрольных субъектов с применением всех 2780 признаков группы В, включая 759 CNA, 1911 признаков глобального метилирования при разрешении 1 Мб и статус метилирования для 191 CpG-островка. Цветной диаграммой в верхней части представлены группы образцов: зеленый, синий и красный цвета представляют здоровых субъектов, ГЦК и пациентов с отличным от ГЦК заболеванием, соответственно. В целом, в указанных трех группах субъектов наблюдалась тенденция к совместной кластеризации. По вертикальной оси представлены классифицирующие признаки. Признаки, характеризующиеся аналогичными паттернами у разных субъектов, кластеризовались вместе. Эти результаты предполагают, что паттерны изменений метилирования разных наборов CpG-островков, изменений полногеномного метилирования при разрешении 1 Мб и CNA в плазме могут использоваться для определения происхождения ракового заболевания у пациентов при неизвестных первичных очагах. Выбор классифицирующих признаков может быть адаптирован под конкретные варианты применения. Кроме того, прогнозу для типа ракового заболевания может присваиваться вес в соответствии с

предварительно известными вероятностями разных типов раковых заболеваний у субъектов. Например, у пациентов с хроническим вирусным гепатитом имеется склонность к развитию гепатоцеллюлярной карциномы, а у хронических курильщиков имеется склонность к развитию рака легкого. Соответственно, может быть рассчитана взвешенная вероятность типа ракового заболевания с применением, например, но не ограничиваясь перечисленным, логистической, множественной или кластерной регрессии.

Согласно другим вариантам реализации для анализа классификации может применяться один тип признаков. Например, в нижеследующих примерах для иерархического кластерного анализа использовали только глобальное метилирование при разрешении 1 Мб, гиперметилирование CpG-островков или CNA при разрешении 1 Мб. Мощность дифференциации может быть разной при использовании разных признаков. Дополнительное уточнение параметров классифицирующих признаков потенциально может повышать точность классификации.

На фиг. 53А представлен иерархический кластерный анализ для образцов плазмы от пациентов с ГЦК, пациентов с отличным от ГЦК заболеванием и здоровых контрольных субъектов с применением группы признаков А метилирования CpG-островков. В целом, больные раковыми заболеваниями пациенты кластеризовались совместно, а субъекты без раковых заболеваний попадали в другой кластер. Однако разделение пациентов с ГЦК и с отличным от ГЦК заболеванием было менее выраженным при сравнении с применением всех трех типов признаков.

На фиг. 53В представлен иерархический кластерный анализ для образцов плазмы от пациентов с ГЦК, пациентов с отличным от ГЦК заболеванием и здоровых контрольных субъектов с применением группы А плотностей глобального метилирования при разрешении 1 Мб в качестве классифицирующих признаков. Наблюдалось преимущественное разбиение по кластерам пациентов с ГЦК и отличным от ГЦК заболеванием.

На фиг. 54А представлен иерархический кластерный анализ для образцов плазмы от пациентов с ГЦК, пациентов с отличным от ГЦК заболеванием и здоровых контрольных субъектов с применением группы А глобальных CNA при разрешении 1 Мб в качестве классифицирующих признаков. Наблюдалось преимущественное разбиение по кластерам пациентов с ГЦК и с отличным от ГЦК заболеванием.

На фиг. 54В представлен иерархический кластерный анализ для образцов плазмы от пациентов с ГЦК, пациентов с отличным от ГЦК заболеванием и здоровых контрольных субъектов с применением группы В плотностей метилирования CpG-островков в качестве классифицирующих признаков. Можно наблюдать

преимущественное разбиение по кластерам пациентов с ГЦК и с отличным от ГЦК заболеванием.

На фиг. 55А представлен иерархический кластерный анализ для образцов плазмы от пациентов с ГЦК, пациентов с отличным от ГЦК заболеванием и здоровых контрольных субъектов с применением группы В плотностей глобального метилирования при разрешении 1 Мб в качестве классифицирующих признаков. Можно наблюдать преимущественное разбиение по кластерам пациентов с ГЦК и с отличным от ГЦК заболеванием.

На фиг. 55В представлен иерархический кластерный анализ для образцов плазмы от пациентов с ГЦК, пациентов с отличным от ГЦК заболеванием и здоровых контрольных субъектов с применением группы В глобальных CNA при разрешении 1 Мб в качестве классифицирующих признаков. Можно наблюдать преимущественное разбиение по кластерам пациентов с ГЦК и с отличным от ГЦК заболеванием.

Указанные результаты иерархической кластеризации для образцов плазмы предполагают, что комбинации разных признаков потенциально могут использоваться для идентификации типов первичного рака. Дальнейшее уточнение критериев выбора критериев потенциально может дополнительно повышать точность классификации.

Соответственно, согласно одному варианту реализации, если классификация метилирования указывает на раковое заболевание в организме, может быть идентифицирован тип ракового заболевания, связанного с указанным организмом, путем сравнения уровня метилирования (например, первого уровня метилирования согласно способу 2800, или уровня метилирования любой области) с соответствующим значением, определенным для других организмов (т.е. других организмов того же типа, например, человека). Указанные соответствующие значения могут относиться к той же области или набору сайтов, для которой(ого) рассчитывали уровень метилирования. По меньшей мере у двух из указанных других организмов идентифицировано наличие других типов раковых заболеваний. Например, соответствующие значения могут быть организованы в кластеры, при этом два кластера связаны с разными типами рака.

Далее, если CNA и метилирование используют совместно для получения третьей классификации уровня рака, признаки CNA и метилирования могут сравниваться с соответствующими значениями для других организмов. Например, первое количество областей (например, на фиг. 36), где наблюдается делеция или амплификация, может сравниваться с соответствующими значениями, определенными для других организмов, для идентификации типа ракового заболевания, связанного с указанным организмом.

Согласно некоторым вариантам реализации признаки метилирования представлены

уровнями метилирования областей из множества областей генома. Могут использоваться области, которые, как было определено, имеют уровень метилирования областей, превышающий соответствующее предельное значение для области, например, уровни метилирования области указанного организма могут сравниваться с уровнями метилирования тех же областей генома у других организмов. Указанное сравнение может позволить дифференцировать типы рака, или просто обеспечить дополнительный фильтр для подтверждения наличия ракового заболевания (например, для идентификации ложноположительных результатов). Соответственно, на основании указанного сравнения можно определить, имеется ли в организме раковое заболевание первого типа, отсутствует ли раковое заболевание или имеется раковое заболевание второго типа.

Другие организмы (наряду с тестируемым) могут быть кластеризованы с применением уровней метилирования области. Соответственно, сравнение уровней метилирования области может применяться для определения того, к какому кластеру принадлежит организм. При разбиении на кластеры могут также использоваться нормированные значения CNA для областей, для которых было определено присутствие делеции или амплификации, согласно описанию выше. Также при разбиении на кластеры могут использоваться соответствующие плотности метилирования гиперметилированных CpG-островков.

Чтобы проиллюстрировать принцип, на котором основан указанный способ, авторы настоящего изобретения приводят пример применения логистической регрессии для классификации двух неизвестных образцов. Целью указанной классификации было определение того, относятся ли указанные два образца к ГЦК или не являющемуся ГЦК раковому заболеванию. Был составлен обучающий набор образцов, включавший 23 образца плазмы, полученных от пациентов с ГЦК, и 18 образцов от пациентов, страдающих отличным от ГЦК раковым заболеванием. Таким образом, всего в обучающий набор входило 41 случая. В указанном примере выбирали 13 признаков, включая пять признаков метилирования CpG-островков (X1-X5), шесть признаков метилирования областей размером 1 Мб (X6-X11) и 2 признака CNA для областей размером 1 Мб (X12-X13). Признаки метилирования CpG выбирали на основе критерия наличия Z-показателя >3 или <-3 по меньшей мере в 15 случаях из обучающего набора. Признаки метилирования областей размером 1 Мб выбирали на основе критерия наличия Z-показателя >3 или <-3 по меньшей мере в 39 случаях из обучающего набора. Признаки CNA выбирали на основе критерия наличия Z-показателя >3 или <-3 по меньшей мере в 20 случаях. Логистический регрессионный анализ осуществляли на образцах указанного обучающего набора для определения коэффициента регрессии для каждого из признаков

(X1–X13). Признаки с коэффициентами регрессии большей величины (независимо от того, в положительном или в отрицательном смысле) позволяют лучше дифференцировать образцы ГЦК и образцы, не относящиеся к ГЦК. Z-показатели для соответствующих признаков в каждом случае использовали в качестве входных значений независимых переменных. Затем исследовали по 13 признакам два образца плазмы, один от пациента с ГЦК (TBR36) и один от пациента, страдающего раком легкого (TBR177).

В данном анализе для классификации типа рака предполагалось, что указанные два образца получены от пациентов, страдающих раковыми заболеваниями неизвестного происхождения. Для каждого образца Z-показатели соответствующего признака подставляли в уравнение логистической регрессии для определения натурального логарифма отношения шансов ($\ln(\text{отношение шансов})$), где указанное отношение шансов представляет отношение вероятностей наличия ГЦК и отсутствия ГЦК (ГЦК/не-ГЦК).

В таблице 7 приведены коэффициенты регрессии уравнения логистической регрессии для 13 признаков. Также приведены Z-показатели соответствующих признаков для двух тестируемых случаев (TBR36 и TBR177). Логарифм $\ln(\text{отношение шансов})$ ГЦК для TBR36 и TBR177 составлял 37,03 и -4,7, соответственно. Рассчитанная по указанным отношениям шансов вероятность того, что образцы плазмы получены от пациентов с ГЦК, составляла >99,9% и 1%, соответственно. Вкратце, существовала высокая вероятность того, что TBR36 представляет собой образец от пациента с ГЦК, и низкая вероятность того, что TBR177 представляет собой образец от пациента с ГЦК.

Таблица 7

Признак	Коэффициент регрессии	Z-показатель соответствующего признака	
		TBR36	TBR177
X1	-2,9575	14,8	0
X2	2,2534	21,3	0
X3	-1,5099	6,1	0
X4	-0,236	34,0	0
X5	0,7426	17,3	0
X6	-0,6682	-26,3	-1,5
X7	-0,2828	-13,9	-2,6
X8	-0,7281	-9,4	-4,4
X9	1,0581	-7,8	-3,7
X10	0,3877	-20,8	-4,3
X11	0,3534	-15,5	-3,1
X12	-1,1826	4,8	3,3
X13	-0,3805	-11,7	-1,4
	$\ln(\text{отношение шансов})$	37,03	-4,37463

Согласно другим вариантам реализации для определения вероятного первичного

происхождения ракового заболевания может применяться иерархическая кластерная регрессия, анализ с применением деревьев классификации и другие регрессионные модели.

ХИ. Материалы и методы

А. Получение библиотек обработанной бисульфитом ДНК и секвенирование

Геномную ДНК (5 мкг) с добавлением 0,5% (w/w) неметилированной лямбда-ДНК (Promega) фрагментировали с помощью системы Covaris S220 (Covaris) на фрагменты длиной приблизительно 200 п.о. Библиотеки ДНК получали с применением набора для подготовки образцов для секвенирования спаренных концов (Paired-End Sequencing Sample Preparation Kit, Illumina) в соответствии с инструкциями производителя, за исключением того, что во фрагменты ДНК лигировали метилированные адаптеры (Illumina). После двух раундов очищения с применением магнитных гранул AMPure XP (Beckman Coulter) продукты лигирования расщепляли на 2 фрагмента, один из которых подвергали 2 раундам бисульфитной модификации с применением набора EpiTect Bisulfite Kit (Qiagen). Неметилированные цитозины в CpG-сайтах вставок превращались в урацилы, тогда как метилированные цитозины сохранялись в неизменном виде. Молекулы ДНК с лигированными адаптерами, обработанные или необработанные бисульфитом натрия, обогащали путем проведения 10 циклов ПЦР с применением следующего состава: 2,5U ДНК-полимеразы PfuTurboCх Hotstart (Agilent Technologies), 1X реакционного буфера PfuTurboCх, 25 мкМ дНТФ, 1 мкл ПЦР-праймера PE 1.0 и 1 мкл ПЦР-праймера PE 2.0 (Illumina) в реакционном объеме 50 мкл. Использовали следующий профиль термоциклирования: 95°C в течение 2 минут, 98°C в течение 30 с, затем 10 циклов при 98°C в течение 15 с, 60°C в течение 30 с и 72°C в течение 4 минут, с финальным этапом при 72°C в течение 10 минут (R Lister, et al. 2009 Nature; 462: 315-322). Продукты ПЦР очищали с применением магнитных гранул AMPure XP.

В ДНК плазмы, экстрагированной из 3,2–4 мл образцов материнской плазмы, добавляли калибровочную фрагментированную лямбда-ДНК (25 пг на мл плазмы) и конструировали библиотеку согласно приведенному выше описанию (RWK Chiu et al. 2011 BMJ; 342: с7401). После лигирования метилированных адаптеров продукты лигирования расщепляли на 2 половины и часть обрабатывали в 2 раундах бисульфитной модификации. Обработанные бисульфитом или необработанные продукты лигирования затем обогащали путем проведения 10 циклов ПЦР согласно приведенному выше описанию.

Библиотеки обработанной или необработанной бисульфитом ДНК секвенировали по 75 п.о. в формате спаренных концов в применении инструментов HiSeq2000 (Illumina).

Кластеры ДНК получали с применением набора Paired-End Cluster Generation Kit v3 с помощью инструмента cBot (Illumina). Анализ изображений в режиме реального времени и распознавание оснований проводили с применением управляющего программного обеспечения HiSeq Control Software (HCS) v1.4 и программного обеспечения анализа в реальном времени Real Time Analysis (RTA) Software v1.13 (Illumina), с помощью которого проводили автоматизированный расчет матриц и фазирование на основе калибровочного контроля PhiX v3, секвенируемого с библиотеками ДНК.

В. Выравнивание последовательностей и идентификация метилированных цитозинов

После распознавания оснований, адаптерные последовательности и основания низкого качества (т.е. с показателем качества < 20) на концах фрагментов удаляли. Обрезанные риды в формате FASTQ затем обрабатывали с применением системы анализа данных метилирования, называемой Methy-Pipe (P Jiang, et al. Methy-Pipe: An integrated bioinformatics data analysis pipeline for whole genome methylome analysis («Methy-Pipe: комплексная система анализа данных биоинформатики для полногеномного анализа метилома»), доклад представлен на Международной практической конференции по биоинформатике и биомедицине Международного института инженеров электротехники и электроники (IEEE International Conference on Bioinformatics and Biomedicine Workshops), Гонконг, 18–21 декабря 2010 г.). Для выравнивания конвертированных бисульфитом ридов секвенирования авторы настоящего изобретения сначала конвертировали все остатки цитозина в тимины *in silico*, на цепях Уотсона-Крика по отдельности, с использованием референсного генома человека (версия NCBI 36/hg18). Затем авторы настоящего изобретения конвертировали каждый цитозин в тимин *in silico* во всех обработанных ридов и сохраняли информацию о расположении каждого конвертированного остатка. SOAP2 (R Li, et al. 2009 Bioinformatics; 25: 1966-1967) использовали для выравнивания конвертированных ридов по двум предварительно конвертированным референсным геномам человека, максимум с двумя несовпадениями для каждого выравниваемого ридов. Выбирали только ридов, картирующиеся с уникальной геномной локализацией. Неоднозначные ридов, картирующиеся на обе цепи Уотсона-Крика и дублирующиеся (клональные) ридов с одинаковыми стартовыми и конечными положениями в геноме удаляли. Ридов секвенирования с размером вставок ≤ 600 п.о. сохраняли для анализа метилирования и размеров.

Остатки цитозина в контексте динуклеотидов CpG были основной мишенью последующих исследований метилирования ДНК. После выравнивания цитозины, исходно присутствовавшие на ридов секвенирования, восстанавливали на основании

информации о расположении, сохраненной при конверсии *in silico*. Восстановленные цитозины в динуклеотидах CpG подсчитывали как метилированные. Тимины в динуклеотидах CpG подсчитывали как неметилированные. Неметилированная лямбда-ДНК, включенная при подготовке библиотеки, использовалась в качестве внутреннего контроля для оценки эффективности модификации бисульфитом натрия. Все цитозины на лямбда-ДНК должны были быть конвертированы в тимины при эффективности бисульфитной конверсии, составляющей 100%.

XIII. Выводы

Применение описанных в настоящем документе вариантов реализации позволяет проводить скрининг, детекцию, мониторинг или прогнозирование ракового заболевания неинвазивным образом с применением, например, плазмы субъекта. Возможно также осуществление пренатального скрининга, получение диагноза, исследование или мониторинг плода путем установления профиля метилирования плодной ДНК по материнской плазме. Для иллюстрации эффективности указанного способа авторы настоящего изобретения показали, что информация, которую обычно получали путем исследования плацентарных тканей, может быть определена непосредственно по материнской плазме. Например, удалось определить импринтинговый статус генных локусов, идентифицировать локусы с дифференциальным метилированием в плодной и материнской ДНК и оценить гестационные вариации профиля метилирования генных локусов посредством прямого анализа ДНК материнской плазмы. Важным преимуществом предложенного авторами изобретения способа является то, что плодный метилом может комплексно оцениваться во время беременности без ущерба для беременности или без необходимости инвазивного отбора образцов плодных тканей. С учетом известных связей между измененным статусом метилирования ДНК и многими связанными с беременностью состояниями, описанный в указанном исследовании способ может быть ценным инструментом для исследования патофизиологии и идентификации биомаркеров таких состояний. Сфокусировавшись на импринтинговых локусах, авторы настоящего изобретения показали, что как передаваемые от отца, так и передаваемые от матери плодные профили метилирования могут оцениваться по материнской плазме. Указанный способ потенциально может подходить для исследования импринтинговых заболеваний. Варианты реализации также могут применяться непосредственно для пренатальной оценки заболеваний плода или связанных с беременностью заболеваний.

Авторы настоящего изобретения также продемонстрировали, что для исследования профиля метилирования ДНК плацентарных тканей может применяться полногеномное бисульфитное секвенирование. В геноме человека содержится приблизительно 28М CpG-

сайтов (C Clark et al. 2012 PLoS One; 7: e50233). Полученные авторами изобретения данные бисульфитного секвенирования для CVS и образца ткани зрелой плаценты захватывали более чем 80% CpG, что представляет по существу более широкое покрытие, чем обеспечиваемое с применением других высокопроизводительных платформ. Например, набор Illumina Infinium HumanMethylation 27K BeadChip Array, который использовали в предыдущем исследовании на плацентарных тканях (T Chu et al. 2011 PLoS One; 6: e14723), захватывал только 0,1% CpG в геноме. Набор Illumina Infinium HumanMethylation 450K BeadChip, появившийся недавно, захватывал только 1,7% CpG (C Clark et al. 2012 PLoS One; 7: e50233). Поскольку при методе МПС отсутствуют ограничения в отношении конструкции зондов, эффективности гибридизации или эффективности захвата антител, могут оцениваться CpG в составе CpG-островков и за их пределами, в контексте большинства последовательностей.

XIV. Компьютерная система

В любых компьютерных системах, упомянутых в данной заявке, может применяться любое подходящее количество подсистем. Примеры таких подсистем показаны на фиг. 33 в компьютерном устройстве 3300. Согласно некоторым вариантам реализации компьютерная система включает отдельное компьютерное устройство, где подсистемы могут представлять собой компоненты указанного компьютерного устройства. Согласно другим вариантам реализации компьютерная система может включать несколько компьютерных устройств, каждое из которых представляет собой подсистему, с внутренними компонентами.

Подсистемы, показанные на фиг. 33, взаимосвязаны посредством системной шины 3375. Показаны дополнительные подсистемы, такие как принтер 3374, клавиатура 3378, устройство(а) хранения 3379, монитор 3376, который соединен с видеоадаптером 3382, и другие. Периферические устройства и устройства ввода/вывода (I/O), которые соединены с контроллером ввода/вывода 3371, можно подсоединить к компьютерной системе с применением любого числа известных в данной области техники способов, таких как порт последовательного ввода-вывода 3377. Например, порт последовательного ввода-вывода 3377 или внешний интерфейс 3381 (например, Ethernet, Wi-Fi и т.д.) можно применять для подсоединения компьютерной системы 3300 к глобальной компьютерной сети, такой как Интернет, к устройству ввода «мышь» или к сканеру. Взаимосвязь посредством системной шины 3375 позволяет центральному процессору 3373 обмениваться данными с каждой подсистемой и контролировать выполнение инструкций от системной памяти 3372 или устройств(а) хранения 3379 (например, жесткого диска), а также обмениваться информацией между подсистемами. Системная память 3372 и/или устройство(а) хранения

3379 может включать в себя машиночитаемый носитель. Любые значения, упомянутые в настоящем документе, могут передаваться от одного компонента к другому компоненту и могут выводиться для предоставления пользователю.

Компьютерная система может включать множество одинаковых компонентов или подсистем, например, соединенных друг с другом через внешний интерфейс 3381 или через внутренний интерфейс. Согласно некоторым вариантам реализации компьютерные системы, подсистемы или устройства могут обмениваться информацией через сеть. В таких случаях один компьютер можно считать клиентом, а другой компьютер – сервером, при этом каждый из указанных компьютеров может представлять собой часть одной и той же компьютерной системы. Как клиент, так и сервер может включать несколько систем, подсистем или компонентов.

Следует понимать, что любой из вариантов реализации настоящего изобретения можно осуществить в виде логических схем устройства управления с применением аппаратного обеспечения (например, специализированной интегральной микросхемы или программируемой пользователем вентильной матрицы) и/или с применением компьютерного программного обеспечения, с программируемым в широком смысле процессором в модульной или интегральной форме. В настоящем документе процессор включает многоядерный процессор на одном и том же интегральном чипе или несколько обрабатывающих модулей, расположенных на одной печатной плате или объединенных в сеть. На основании настоящего описания и идей, предложенных в настоящем документе, среднему специалисту в данной области техники будут очевидны и понятны другие пути и/или способы осуществления вариантов реализации настоящего изобретения с применением аппаратного обеспечения и комбинаций аппаратного обеспечения с программным обеспечением.

Любые компоненты или функции программного обеспечения, описанные в настоящей заявке, можно осуществить в виде программного кода, который выполнит процессор, с применением любого подходящего языка программирования, такого как, например, Java, C++ или Perl, с использованием, например, обычных или объектно-ориентированных методик. Программный код можно хранить в виде ряда инструкций или команд на машиночитаемом носителе для хранения и/или передачи, подходящие носители включают оперативное запоминающее устройство (RAM), постоянное запоминающее устройство (ROM), магнитный носитель, такой как жесткий диск или дискету, или оптический носитель, такой как компакт-диск (CD) или DVD (универсальный цифровой диск), флеш-память и т.п. Машиночитаемые носители могут представлять собой любую комбинацию таких устройств хранения или передачи.

Такие программы также могут кодироваться и передаваться с помощью несущих сигналов, предназначенных для передачи через проводные, оптические и/или беспроводные сети, соответствующих множеству протоколов, включая Интернет. Таким образом, машиночитаемые носители согласно варианту реализации настоящего изобретения можно получить, применяя информационный сигнал, кодируемый такими программами. Машиночитаемые носители, кодированные с помощью программного кода, могут быть упакованы вместе с совместимым устройством или предоставлены отдельно от других устройств (например, посредством загрузки через Интернет). Любой такой машиночитаемый носитель может храниться на/в едином компьютерном программном продукте (например, на жестком диске, CD-диске или в составе всей компьютерной системы), и может присутствовать в составе различных программных продуктов или на различных программных продуктах внутри системы или сети. Компьютерная система может включать монитор, принтер или другое подходящее устройство отображения для предоставления пользователю любого из результатов, упомянутых в данной заявке.

Любой из способов, описанных в данной заявке, можно полностью или частично осуществить с помощью компьютерной системы, включающей один или большее количество процессоров, которые можно сконфигурировать для осуществления указанных этапов. Таким образом, варианты реализации могут быть направлены на компьютерные системы, сконфигурированные для осуществления этапов любого из способов, описанных в данной заявке, возможно с различными компонентами, осуществляющими соответствующие этапы или соответствующую группу этапов. Хотя этапы согласно описанным в настоящем документе способам и были представлены в пронумерованном виде, указанные этапы могут осуществляться одновременно или в другом порядке. Кроме того, блоки данных этапов можно применять с блоками других этапов из других способов. Также весь этап или блоки этапа могут быть необязательными. Кроме того, любой из этапов любого из способов можно осуществить с помощью модулей, схем или других средств для осуществления указанных этапов.

Специфические детали конкретных вариантов реализации могут быть скомбинированы любым подходящим способом без отступления от сущности и объема вариантов реализации настоящего изобретения. Тем не менее, другие варианты реализации настоящего изобретения могут быть направлены на конкретные варианты реализации, относящиеся к каждому отдельному аспекту, или на конкретные комбинации указанных отдельных аспектов.

Предшествующее описание примеров вариантов реализации настоящего изобретения представлено с целью иллюстрации и описания. Не предполагается, что оно

является исчерпывающим или ограничивает настоящее изобретение точно описанной формой; возможны многие модификации и вариации в свете описанных выше идей. Варианты реализации были выбраны и описаны для того, чтобы наилучшим образом объяснить принципы настоящего изобретения и его практического применения, с тем, чтобы дать возможность другим специалистам в данной области техники наилучшим образом применять настоящее изобретение в различных вариантах реализации и с различными модификациями, подходящими для конкретного предполагаемого применения.

Предполагается, что применение формы единственного числа, в том числе в сочетании с определением «указанный(ая/ое)», означает «один или большее количество», если явным образом не указано иное.

Все патенты, заявки на патент, публикации и описания патентов, упомянутые в настоящем документе, полностью включены посредством ссылки во всех отношениях. Ни один из данных источников не считают известным уровнем техники.

Таблица S2A. Перечень 100 наиболее гиперметилированных областей, идентифицированных в образце ворсин хориона и клетках крови матери в первом триместре

Хромосома	Начало	Конец	Размер (п.о.)	Клетки крови матери	CVS	P-значения	Различие метилирования
chr13	113063600	113064100	500	0,009	0,9	3,67E-15	0,891
chr6	36279700	36280200	500	0,0068	0,8957	2,39E-22	0,8889
chr16	66876000	66876500	500	0,0327	0,9211	3,82E-21	0,8884
chr10	163500	164000	500	0,0195	0,9034	3,60E-35	0,8839
chr9	3518300	3518800	500	0,0263	0,9045	1,32E-26	0,8782
chr12	31877100	31877600	500	0,007	0,8784	3,08E-22	0,8714
chr22	37477400	37478400	1000	0,0152	0,8848	0,00E+00	0,8696
chr4	148940500	148941000	500	0,0055	0,8717	4,40E-29	0,8662
chr5	131836300	131836800	500	0,075	0,9403	1,54E-10	0,8653
chr17	26661700	26663600	1900	0,0187	0,875	2,95E-38	0,8563
chr2	105758600	105759600	1000	0,0305	0,8828	1,19E-53	0,8523
chr22	39188800	39189800	1000	0	0,8514	2,05E-46	0,8514
chr3	153443900	153444900	1000	0,0436	0,8945	5,43E-34	0,8509
chr6	25149600	25150600	1000	0,0135	0,8632	0,00E+00	0,8497
chr5	98296800	98297300	500	0,0432	0,8925	4,97E-23	0,8493
chr7	150679900	150680400	500	0,0496	0,8944	6,50E-17	0,8448
chr7	107563100	107563600	500	0,0495	0,8895	9,58E-26	0,84
chr7	37348300	37349300	1000	0,0012	0,8409	0,00E+00	0,8397
chr14	58837800	58838300	500	0,0097	0,848	3,35E-16	0,8383
chr6	119238100	119238600	500	0,0899	0,928	2,38E-19	0,8381
chr15	93669900	93670400	500	0,0753	0,913	2,19E-10	0,8377
chr17	26669200	26670200	1000	0,0221	0,859	1,44E-29	0,8369
chr2	88108100	88108600	500	0,075	0,9109	3,55E-17	0,8359
chr13	98363800	98364300	500	0,11	0,9457	1,28E-11	0,8357
chr16	66948000	66948500	500	0,0331	0,8685	0,00E+00	0,8354

Хромосома	Начало	Конец	Размер (п.о.)	Клетки крови матери	CVS	P-значения	Различие метилирования
chr6	42098000	42098500	500	0,0484	0,8835	3,73E-16	0,8351
chr3	129876000	129876500	500	0,0565	0,8897	8,81E-17	0,8332
chr3	142700300	142700800	500	0,0063	0,8393	2,59E-22	0,833
chr8	145883800	145884300	500	0,0392	0,872	0,00E+00	0,8328
chr10	8320700	8321200	500	0,0566	0,8871	9,40E-09	0,8305
chr3	120438100	120438600	500	0,102	0,9292	7,09E-16	0,8272
chr3	173792600	173793100	500	0,0182	0,8453	2,84E-39	0,8271
chr17	40320700	40321200	500	0,0539	0,8788	6,50E-30	0,8249
chr15	72076200	72076700	500	0,0299	0,8525	6,48E-10	0,8226
chr16	29663900	29665400	1500	0,0081	0,8305	0,00E+00	0,8224
chr11	66961100	66962100	1000	0,0489	0,8712	0,00E+00	0,8223
chr9	27083100	27084100	1000	0,097	0,9177	2,37E-55	0,8207
chr9	111249600	111250100	500	0,0613	0,8795	1,99E-20	0,8182
chr14	101412400	101412900	500	0	0,8167	8,26E-32	0,8167
chr1	242549200	242549700	500	0	0,8155	3,50E-21	0,8155
chr8	38642800	38643300	500	0,0191	0,8346	3,22E-41	0,8155
chr4	85893600	85894100	500	0,0394	0,8533	1,45E-15	0,8139
chr5	142368600	142369100	500	0,0385	0,8523	1,18E-18	0,8138
chr8	130969500	130970000	500	0,069	0,8824	2,42E-24	0,8134
chr2	196783900	196784400	500	0	0,8123	3,63E-40	0,8123
chr16	49258100	49258600	500	0,0733	0,8851	4,29E-18	0,8118

chr1	232601200	232601700	500	0,0594	0,8707	1,73E-13	0,8113
chr1	109039500	109040000	500	0,0366	0,8471	1,07E-11	0,8105
chr17	59491300	59491800	500	0,0662	0,8758	2,15E-17	0,8096
chr21	42194100	42194600	500	0,11	0,9182	1,61E-12	0,8082
chr9	116174500	116175500	1000	0,0062	0,8132	1,98E-60	0,807
chr15	73429200	73429700	500	0	0,8066	9,81E-33	0,8066
chr6	157462800	157463300	500	0,0758	0,8819	7,94E-16	0,8061
chr3	16858500	16859500	1000	0,0021	0,8068	4,76E-68	0,8047
chr9	96662800	96663300	500	0,0614	0,8651	6,79E-28	0,8037
chr9	88143000	88143500	500	0,1538	0,9559	7,43E-09	0,8021
chr19	16090000	16091000	1000	0,0899	0,8904	1,60E-53	0,8005
chr15	29436300	29437300	1000	0,0553	0,8556	1,18E-80	0,8003
chr11	77816100	77816600	500	0,1069	0,9068	2,31E-17	0,7999
chr10	30346800	30347300	500	0,1212	0,9211	7,48E-07	0,7999
chr1	89510300	89511300	1000	0,0203	0,8191	3,53E-77	0,7988
chr3	125986100	125986600	500	0,1686	0,9674	5,24E-22	0,7988
chr19	60162800	60163300	500	0,0127	0,8113	9,99E-19	0,7986
chr16	73655900	73656900	1000	0,082	0,8806	4,48E-41	0,7986
chr16	30104300	30105800	1500	0,0298	0,8282	0,00E+00	0,7984
chr10	118642400	118642900	500	0,0588	0,8571	8,63E-11	0,7983
chr16	4495000	4496000	1000	0,0632	0,8615	2,27E-44	0,7983
chr1	2048300	2048800	500	0,0309	0,8289	1,19E-80	0,798
chr2	136481800	136482800	1000	0,0554	0,8533	8,50E-48	0,7979
chr10	29959200	29959700	500	0,1429	0,94	2,60E-08	0,7971
chr6	139642400	139642900	500	0,0618	0,8585	2,16E-29	0,7967
chr14	69825300	69825800	500	0,0654	0,8615	6,85E-14	0,7961
chr8	49739700	49740200	500	0,0324	0,828	2,88E-30	0,7956
chr17	42205700	42206200	500	0,057	0,852	2,11E-30	0,795
chr4	77445300	77445800	500	0,0442	0,8377	1,79E-35	0,7935

chr17	53762700	53766300	3600	0,0003	0,7926	0,00E+00	0,7923
chr17	44269900	44270400	500	0,026	0,8182	3,49E-21	0,7922
chr6	42462700	42463200	500	0,0761	0,8678	4,74E-22	0,7917
chr2	23396200	23396700	500	0,0333	0,8235	1,25E-14	0,7902
chr9	100921100	100921600	500	0,0244	0,814	3,32E-21	0,7896
chr7	74016100	74016600	500	0,1442	0,9333	6,74E-10	0,7891
chr6	157879000	157879500	500	0,133	0,9219	6,36E-17	0,7889
chr3	3189400	3190400	1000	0,0693	0,8571	1,38E-24	0,7878
chr16	29581500	29584500	3000	0,0081	0,7956	0,00E+00	0,7875
chr17	42201800	42202800	1000	0,0884	0,8751	0,00E+00	0,7867
chr11	94257000	94257500	500	0,1122	0,8986	4,29E-10	0,7864
chr10	14741600	14742100	500	0,0139	0,8	1,73E-20	0,7861
chr21	33826900	33827400	500	0,0879	0,8739	2,81E-11	0,786
chr4	130057200	130057700	500	0,0893	0,875	1,76E-13	0,7857
chr21	35343400	35343900	500	0	0,7853	7,43E-18	0,7853
chr12	105372800	105373300	500	0,0923	0,8767	8,67E-22	0,7844
chr5	10799800	10800300	500	0,1429	0,9263	8,21E-17	0,7834
chr5	16753100	16753600	500	0,041	0,8241	1,40E-15	0,7831
chr3	135746000	135746500	500	0,1429	0,9259	2,86E-09	0,783
chr6	53708300	53708800	500	0,0412	0,8235	2,74E-31	0,7823
chr2	128122900	128123400	500	0,0634	0,8455	4,82E-21	0,7821
chr5	150574200	150574700	500	0,0876	0,8696	1,56E-20	0,782

chr16	84326000	84327000	1000	0,1071	0,8891	3,58E-61	0,782
chr1	26744500	26745500	1000	0,0336	0,8152	0,00E+00	0,7816
chr2	234882000	234882500	500	0,0392	0,819	7,63E-14	0,7798

Таблица S2B. Перечень 100 наиболее гипометилированных областей, идентифицированных в образце ворсин хориона и клетках крови матери в первом триместре

Хромосома	Начало	Конец	Размер (п.о.)	Клетки крови матери	CVS	P-значения	Различие метилирования
chr18	12217500	12218500	1000	0,9873	0	3,05E-25	0,9873
chr17	22885400	22885900	500	0,9714	0,0161	8,92E-12	0,9553
chr3	184827100	184827600	500	0,9875	0,033	4,79E-16	0,9545
chr5	148968300	148968800	500	0,98	0,0426	6,70E-09	0,9374
chr10	104794500	104795000	500	0,973	0,0385	9,33E-10	0,9345
chr4	84977900	84978400	500	0,9643	0,0417	2,98E-08	0,9226
chr3	180395300	180395800	500	0,9877	0,0667	6,72E-08	0,921
chr2	138908300	138908800	500	0,939	0,0208	1,10E-16	0,9182
chr6	139873100	139873600	500	0,9667	0,0526	1,29E-07	0,914
chr8	59604700	59605200	500	0,9468	0,033	2,88E-14	0,9138
chr6	167622300	167622800	500	0,9452	0,0316	3,86E-14	0,9136
chr3	175701300	175701800	500	0,9846	0,0735	7,43E-10	0,9111
chr13	59246400	59246900	500	0,9402	0,0313	2,31E-11	0,9089
chr12	71263600	71264100	500	0,9296	0,0213	1,08E-08	0,9083
chr5	39459400	39459900	500	0,9219	0,014	5,01E-22	0,9079
chr17	24904700	24905200	500	0,9161	0,0092	5,04E-35	0,9069
chr12	31889900	31890400	500	0,9524	0,0465	6,78E-13	0,9059
chr3	152897800	152898300	500	0,9402	0,0345	1,70E-17	0,9057
chr1	40378700	40379200	500	0,9565	0,0526	3,31E-09	0,9039

Хромосома	Начало	Конец	Размер (п.о.)	Клетки крови матери	CVS	P- значения	Различие метилирования
chr12	43979300	43979800	500	0,952	0,05	6,68E-13	0,902
chr18	1395900	1397400	1500	0,9308	0,0293	0,00E+00	0,9015
chr1	223482900	223483400	500	0,9579	0,0575	3,36E-24	0,9004
chr9	130357000	130357500	500	0,9282	0,0286	9,19E-13	0,8996
chr3	72878300	72878800	500	0,9612	0,0625	8,20E-14	0,8987
chr7	84347200	84348700	1500	0,9401	0,0418	0,00E+00	0,8983
chr15	37317500	37318000	500	0,9358	0,0385	7,58E-14	0,8973
chr8	42528600	42529100	500	0,9302	0,0337	1,73E-14	0,8965
chr6	134914000	134914500	500	0,9037	0,0076	4,84E-21	0,8961
chr13	56207100	56208100	1000	0,9184	0,0245	0,00E+00	0,894
chr2	209074000	209074500	500	0,9309	0,037	6,13E-27	0,8938
chr12	74021100	74022100	1000	0,9513	0,058	0,00E+00	0,8933
chr4	118939300	118939800	500	0,9192	0,0276	6,58E-27	0,8916
chr5	12626600	12628600	2000	0,9266	0,0355	0,00E+00	0,8911
chr5	105517300	105518300	1000	0,927	0,0359	0,00E+00	0,891
chr12	70056300	70057300	1000	0,9488	0,0609	0,00E+00	0,888
chr6	153238200	153239200	1000	0,9123	0,0244	0,00E+00	0,8879
chr17	60374800	60375300	500	0,9655	0,0777	3,64E-14	0,8878
chr14	68272700	68273200	500	0,9389	0,0523	1,23E-22	0,8866
chr19	54533800	54534800	1000	0,9117	0,0262	0,00E+00	0,8855
chr12	15392200	15393200	1000	0,9307	0,0457	0,00E+00	0,885
chr1	212517400	212517900	500	0,9266	0,0417	9,81E-12	0,8849
chr10	49344400	49345400	1000	0,9422	0,0579	0,00E+00	0,8844
chr3	47410400	47410900	500	0,9213	0,0381	5,59E-16	0,8832
chr3	879500	880000	500	0,9455	0,0625	8,06E-06	0,883
chr2	31572400	31573400	1000	0,9176	0,0357	0,00E+00	0,8819

chr1	89131200	89131700	500	0,9314	0,0498	5,15E-70	0,8816
chr8	94832000	94832500	500	0,9156	0,0351	2,16E-65	0,8805
chr7	14008300	14009800	1500	0,9349	0,0545	0,00E+00	0,8804
chr12	12971300	12972300	1000	0,9361	0,0559	0,00E+00	0,8802
chr5	43114700	43115200	500	0,9638	0,0842	1,79E-13	0,8796
chr11	107872400	107872900	500	0,9472	0,0677	2,31E-32	0,8794
chr8	49757600	49758100	500	0,9048	0,0269	3,15E-52	0,8779
chr13	33106400	33106900	500	0,9384	0,0606	9,54E-15	0,8778
chr3	190658800	190659300	500	0,9388	0,0617	2,71E-22	0,877
chr1	181508000	181508500	500	0,9259	0,0495	3,78E-15	0,8764
chr1	180436900	180437400	500	0,9412	0,0652	2,36E-13	0,876
chr6	122642800	122643800	1000	0,9218	0,0458	0,00E+00	0,8759
chr5	166429300	166429800	500	0,9551	0,08	5,26E-05	0,8751
chr12	14972900	14973400	500	0,9483	0,0733	2,10E-18	0,8749
chr5	123933900	123934400	500	0,943	0,0683	1,12E-39	0,8746
chr2	15969400	15970400	1000	0,8939	0,0196	9,43E-46	0,8743
chr3	167635200	167636200	1000	0,9363	0,0625	9,03E-41	0,8738
chr5	159442700	159443200	500	0,9174	0,044	6,27E-14	0,8734
chr4	48027200	48027700	500	0,9839	0,1111	8,89E-06	0,8728
chr6	140071500	140072000	500	0,9234	0,0506	4,39E-33	0,8728
chr10	22356300	22356800	500	0,9548	0,0822	1,04E-18	0,8726
chr8	61007300	61007800	500	0,9197	0,0476	1,24E-15	0,8721
chr11	95463500	95464000	500	0,9348	0,0629	1,16E-20	0,8718
chr2	216399800	216400300	500	0,938	0,0667	5,98E-06	0,8713

chr18	57359700	57360200	500	0,9293	0,0584	1,77E-19	0,871
chr3	102734400	102734900	500	0,8917	0,0207	7,94E-22	0,871
chr1	173605700	173606200	500	0,96	0,0891	5,46E-13	0,8709
chr2	86993700	86995700	2000	0,8965	0,0261	0,00E+00	0,8704
chr3	162621100	162621600	500	0,9226	0,0526	7,89E-38	0,8699
chr12	10144800	10145300	500	0,929	0,0598	3,45E-17	0,8691
chr3	113855100	113855600	500	0,9667	0,0982	3,97E-14	0,8685
chr2	156958200	156959200	1000	0,9252	0,0571	8,89E-50	0,8681
chr2	55775000	55776000	1000	0,9159	0,0483	0,00E+00	0,8676
chr6	124898400	124898900	500	0,8987	0,0313	1,91E-15	0,8675
chr5	42003700	42004700	1000	0,9262	0,0588	0,00E+00	0,8674
chr3	24162200	24162700	500	0,883	0,0161	1,75E-27	0,8668
chr6	35394000	35395000	1000	0,9204	0,0539	0,00E+00	0,8665
chr17	8451800	8453300	1500	0,9376	0,0714	0,00E+00	0,8662
chr14	53487700	53488700	1000	0,9013	0,0353	0,00E+00	0,866
chr7	98572800	98573300	500	0,9651	0,0995	8,37E-26	0,8656
chr6	52298700	52299200	500	0,9427	0,0772	1,31E-28	0,8655
chr6	159047900	159048400	500	0,908	0,0426	3,34E-08	0,8655
chr14	22152600	22153100	500	0,9085	0,0435	4,43E-17	0,865
chr12	103285000	103285500	500	0,9321	0,0674	0,00E+00	0,8647
chr7	43302200	43302700	500	0,968	0,1037	6,40E-16	0,8643
chr14	22247400	22247900	500	0,9804	0,1163	3,83E-06	0,8641
chr2	66780900	66781400	500	0,9355	0,0714	8,37E-09	0,8641
chr12	97393000	97393500	500	0,9045	0,0408	3,46E-21	0,8637
chr5	162797900	162798900	1000	0,9271	0,0635	1,75E-57	0,8636
chr2	83598400	83599400	1000	0,9354	0,0719	0,00E+00	0,8635
chr11	111358800	111359300	500	0,9156	0,0523	4,15E-24	0,8632

chr11	104891100	104892600	1500	0,9164	0,0533	0,00E+00	0,863
chr1	184583600	184584100	500	0,9647	0,1026	3,02E-13	0,8621
chr5	132350500	132351500	1000	0,9042	0,0426	1,86E-36	0,8616
chr5	53268300	53268800	500	0,972	0,1111	4,76E-16	0,8609

Таблица S2C. Перечень 100 наиболее гиперметилированных областей, идентифицированных в третьем триместре в плацентарной ткани и в клетках крови матери

Хромосома	Начало	Конец	Размер (п.о.)	к	Зрелая плацента	P-значения	Различие метилирования
chr4	78129700	78130200	500	0,0488	0,9747	3,97E-33	0,926
chr5	131467400	131467900	500	0,0213	0,9275	7,10E-27	0,9063
chr17	26661700	26663600	1900	0,0187	0,9226	1,79E-41	0,9039
chr4	148940500	148941000	500	0,0055	0,9079	1,82E-29	0,9024
chr9	100921100	100921600	500	0,0244	0,9242	1,38E-25	0,8998
chr6	137114200	137114700	500	0	0,8934	8,87E-14	0,8934
chr3	173792600	173793100	500	0,0182	0,9091	1,70E-42	0,8908
chr5	98296800	98297300	500	0,0432	0,9333	2,58E-23	0,8901
chr12	44898000	44898500	500	0	0,8889	4,47E-11	0,8889
chr3	197328900	197329400	500	0,0169	0,9048	5,55E-10	0,8878
chr8	49739700	49740200	500	0,0324	0,9194	5,71E-34	0,887
chr12	122279300	122279800	500	0,0135	0,8969	3,46E-21	0,8834
chr17	43092200	43092700	500	0	0,8824	4,34E-10	0,8824
chr7	107563100	107563600	500	0,0495	0,931	1,05E-28	0,8815

Хромосома	Начало	Конец	Размер (п.о.)	к	Зрелая плацента	P- значения	Различие метилювания
chr11	72543200	72543700	500	0,0377	0,9167	2,94E-09	0,8789
chr14	58837800	58838300	500	0,0097	0,886	9,16E-18	0,8763
chr3	153443900	153444900	1000	0,0436	0,9197	6,24E-39	0,876
chr3	16953200	16953700	500	0,0896	0,9655	6,78E-09	0,876
chr17	42205700	42206200	500	0,057	0,933	1,13E-31	0,8759
chr6	53217600	53218100	500	0,0818	0,9571	1,54E-19	0,8754
chr3	112749000	112749500	500	0,0403	0,9154	4,11E-22	0,8752
chr8	22453700	22454200	500	0,003	0,8765	1,64E-50	0,8735
chr1	162860900	162861400	500	0,023	0,8932	8,37E-14	0,8702
chr6	36279700	36280200	500	0,0068	0,8762	1,14E-21	0,8694
chr5	80962500	80963000	500	0	0,8679	2,08E-15	0,8679
chr16	11312500	11313000	500	0	0,8679	2,14E-10	0,8679
chr16	29663900	29665400	1500	0,0081	0,8759	0,00E+00	0,8678
chr3	120438100	120438600	500	0,102	0,9639	2,98E-15	0,8618
chr8	134157000	134157500	500	0,0625	0,9219	6,10E-20	0,8594
chr6	42620900	42621400	500	0	0,8571	5,68E-08	0,8571
chr5	131836300	131836800	500	0,075	0,9315	1,26E-10	0,8565
chr14	60290000	60290500	500	0	0,8544	5,63E-14	0,8544
chr6	42850300	42851300	1000	0,0676	0,9211	2,38E-24	0,8534
chr8	28974100	28974600	500	0,0394	0,8927	2,03E-51	0,8533
chr22	22368500	22369000	500	0,0248	0,8778	1,18E-70	0,8529
chr14	69825300	69825800	500	0,0654	0,9174	2,73E-14	0,852
chr3	142700300	142700800	500	0,0063	0,8582	2,56E-23	0,8519
chr17	59491300	59491800	500	0,0662	0,9175	1,81E-16	0,8513
chr15	30881700	30882200	500	0,0493	0,8995	2,38E-26	0,8502
chr15	91496300	91496800	500	0	0,85	3,13E-17	0,85
chr17	18745300	18745800	500	0,0294	0,8775	3,47E-51	0,848
chr15	29436500	29437000	500	0,0336	0,8811	2,62E-66	0,8476
chr2	217795300	217795800	500	0	0,8472	1,78E-22	0,8472
chr11	16328100	16328600	500	0,0278	0,875	3,43E-11	0,8472
chr13	113063500	113064000	500	0,0102	0,8571	1,82E-15	0,8469

chr5	40472400	40472900	500	0,0197	0,8664	7,54E-35	0,8467
chr1	242549200	242549700	500	0	0,8462	8,53E-23	0,8462
chr11	58099100	58099600	500	0,0162	0,8612	4,45E-35	0,845
chr9	16020400	16020900	500	0,0132	0,8555	8,05E-23	0,8423
chr8	37550700	37551200	500	0,0093	0,8512	1,11E-16	0,8419
chr5	75722400	75722900	500	0,1215	0,9627	5,97E-23	0,8411
chr19	60454700	60455200	500	0,0316	0,8722	2,44E-62	0,8405
chr4	99587100	99587600	500	0,0128	0,8526	1,49E-12	0,8398
chr6	25149600	25150600	1000	0,0135	0,8514	0,00E+00	0,8379
chr1	32065200	32065700	500	0	0,8371	1,09E-44	0,8371
chr7	5337200	5337700	500	0,0727	0,9098	2,18E-14	0,8371
chr17	44269900	44270400	500	0,026	0,8621	3,94E-22	0,8361
chr1	36180800	36181300	500	0,0714	0,9067	1,23E-09	0,8352
chr18	10472700	10473700	1000	0,0713	0,9064	9,35E-70	0,8351
chr5	350000	350500	500	0,0297	0,8643	1,49E-16	0,8346
chr2	136481800	136482800	1000	0,0554	0,8887	1,87E-52	0,8332
chr4	89241100	89241600	500	0,1091	0,9423	1,05E-12	0,8332
chr17	40320700	40321200	500	0,0539	0,8859	6,94E-31	0,832
chr7	133897200	133897700	500	0,0769	0,9077	1,64E-24	0,8308
chr8	98060600	98061100	500	0,0741	0,9048	3,11E-07	0,8307

chr8	134141500	134142000	500	0	0,829	2,77E-58	0,829
chr14	80250600	80251100	500	0,0839	0,9122	2,05E-18	0,8283
chr2	100730900	100731400	500	0,0787	0,9067	4,85E-11	0,828
chr2	88108100	88108600	500	0,075	0,901	2,10E-16	0,826
chr19	16338500	16339500	1000	0,0011	0,8259	0,00E+00	0,8247
chr5	141791900	141792900	1000	0,0225	0,8467	0,00E+00	0,8243
chr11	116227400	116227900	500	0	0,8242	1,01E-17	0,8242
chr22	48705500	48706000	500	0,0649	0,8891	4,00E-76	0,8242
chr9	3518300	3518800	500	0,0263	0,8493	6,66E-25	0,823
chr11	16791000	16791500	500	0,1095	0,9322	1,05E-22	0,8228
chr3	135746000	135746500	500	0,1429	0,9651	7,63E-10	0,8223
chr1	19323400	19323900	500	0,0411	0,8624	2,36E-20	0,8213
chr9	96662800	96663300	500	0,0614	0,8826	1,07E-28	0,8212
chr7	37348300	37349300	1000	0,0012	0,821	0,00E+00	0,8198
chr2	234882000	234882500	500	0,0392	0,8591	5,36E-15	0,8198
chr6	44694000	44694500	500	0,1024	0,9222	5,68E-19	0,8198
chr17	18320500	18321000	500	0	0,8197	2,78E-39	0,8197
chr22	28992000	28994000	2000	0,0012	0,8195	0,00E+00	0,8183
chr17	53762700	53766300	3600	0,0003	0,8179	0,00E+00	0,8176
chr1	114215500	114216000	500	0	0,8169	2,24E-20	0,8169
chr6	13381700	13382700	1000	0,0037	0,8206	1,03E-40	0,8169
chr5	17045400	17045900	500	0,0235	0,84	3,24E-13	0,8165
chr12	110924300	110924800	500	0,0855	0,9016	1,01E-18	0,816
chr1	200499800	200500300	500	0,011	0,8269	9,73E-24	0,8159
chr4	8311000	8311500	500	0,053	0,8687	7,82E-18	0,8157
chr8	6535300	6535800	500	0,0667	0,8824	5,25E-09	0,8157
chr6	42462700	42463200	500	0,0761	0,8919	5,78E-23	0,8157
chr1	91969900	91970400	500	0,0172	0,8325	5,23E-18	0,8152
chr2	105758600	105759600	1000	0,0305	0,8455	5,15E-52	0,815
chr21	37538500	37539000	500	0,1595	0,9745	8,81E-16	0,815
chr9	92953000	92953500	500	0,0189	0,8333	1,88E-15	0,8145

chr16	30104400	30105900	1500	0,0505	0,8636	0,00E+00	0,8131
chr1	234184400	234185400	1000	0,0346	0,8477	9,66E-31	0,813
chr8	19116400	19116900	500	0	0,8125	9,33E-11	0,8125
chr4	141194300	141195300	1000	0,0865	0,899	5,46E-30	0,8125

Таблица S2D. Перечень 100 наиболее гипометилированных областей, идентифицированных в третьем триместре в плацентарной ткани и в клетках крови матери

Хромосома	Начало	Конец	Размер (п.о.)	Клетки крови матери	Зрелая плацента	P-значения	Различие метилирования
chr9	40380300	40380800	500	0,9667	0	1,13E-06	0,9667
chr1	31769200	31769700	500	0,9548	0,0256	5,57E-25	0,9291
chr18	12217600	12218100	500	0,9873	0,0602	1,63E-19	0,9271
chr20	19704400	19704900	500	0,9426	0,018	4,34E-18	0,9246
chr15	37317500	37318000	500	0,9358	0,0132	1,90E-25	0,9226
chrX	83368400	83368900	500	0,913	0	3,15E-07	0,913
chr11	27549100	27549600	500	0,9224	0,0123	3,92E-24	0,9101
chr18	58141500	58142000	500	0,9737	0,0645	1,07E-09	0,9092
chr1	159897000	159897500	500	0,9067	0	2,53E-16	0,9067

Хромосома	Начало	Конец	Размер (п.о.)	Клетки крови матери	Зрелая плацента	P- значения	Различие метиляции
chr7	84347200	84348700	1500	0,9401	0,0407	0,00E+00	0,8994
chr2	216916100	216916600	500	0,9695	0,0714	2,13E-16	0,8981
chr7	144200000	144200500	500	0,9294	0,0317	1,24E-10	0,8977
chr1	241331600	241332100	500	0,9198	0,0227	0,00E+00	0,8971
chr7	123190000	123191000	1000	0,9341	0,0384	0,00E+00	0,8957
chr5	12626600	12628600	2000	0,9266	0,0321	0,00E+00	0,8944
chr12	12971300	12972300	1000	0,9361	0,0438	0,00E+00	0,8923
chr22	20936500	20937000	500	0,9528	0,0606	1,87E-06	0,8922
chr13	31321900	31322400	500	0,9231	0,0313	1,43E-06	0,8918
chr22	21701500	21702000	500	0,9579	0,0667	1,30E-09	0,8912
chr10	104794400	104794900	500	1	0,1111	6,10E-09	0,8889
chr7	21835800	21836300	500	0,9156	0,0267	3,85E-13	0,8889
chr10	16134800	16135300	500	0,95	0,0635	4,79E-10	0,8865
chr3	47410400	47410900	500	0,9213	0,0357	6,63E-17	0,8855
chr10	49344400	49345400	1000	0,9422	0,0571	0,00E+00	0,8851
chr2	209073900	209074400	500	0,9196	0,0353	1,63E-22	0,8843
chr1	89131200	89131700	500	0,9314	0,0472	1,05E-75	0,8842
chr3	167118500	167119500	1000	0,9365	0,0527	0,00E+00	0,8838
chr18	1395900	1397400	1500	0,9308	0,0472	0,00E+00	0,8836
chr2	59670300	59670800	500	0,9433	0,0599	5,09E-23	0,8834
chr14	28368900	28369400	500	0,9446	0,0619	5,03E-64	0,8827
chr3	126028800	126029300	500	0,9379	0,0556	1,83E-20	0,8823
chr9	69378900	69379900	1000	0,8816	0	6,02E-51	0,8816
chr5	105517300	105518300	1000	0,927	0,0461	0,00E+00	0,8808
chr2	31572400	31573400	1000	0,9176	0,037	0,00E+00	0,8806
chr5	42003700	42004700	1000	0,9262	0,0462	0,00E+00	0,88
chr14	94718300	94718800	500	0,9548	0,0764	6,67E-19	0,8784
chr19	56417800	56418300	500	0,9615	0,0833	1,71E-06	0,8782
chr2	70183000	70183500	500	0,9694	0,0914	9,49E-39	0,878
chr4	118939300	118939800	500	0,9192	0,0412	2,20E-34	0,878
chr13	59246400	59246900	500	0,9402	0,0633	5,40E-12	0,8769
chr12	74021100	74022100	1000	0,9513	0,0752	0,00E+00	0,8761
chr2	173432500	173433000	500	0,9529	0,0778	5,39E-12	0,8752
chr16	24004400	24004900	500	0,9239	0,0488	3,25E-23	0,8751
chr13	27596300	27597300	1000	0,9538	0,0795	0,00E+00	0,8743
chr15	88904300	88904800	500	0,9212	0,0481	7,69E-27	0,8731

chr18	12720200	12721200	1000	0,9346	0,0618	0,00E+00	0,8728
chr15	60975900	60976900	1000	0,9311	0,0587	0,00E+00	0,8724
chr21	39630100	39631100	1000	0,9423	0,07	4,68E-43	0,8723
chr5	123933900	123934400	500	0,943	0,0707	2,60E-38	0,8722
chr8	77382600	77383600	1000	0,9117	0,0395	0,00E+00	0,8722
chr21	32238800	32239300	500	0,9396	0,0677	1,28E-18	0,8719
chr5	175019600	175020100	500	0,9542	0,0828	4,34E-20	0,8714
chr8	134437400	134438400	1000	0,9083	0,037	4,79E-29	0,8713
chr5	69668800	69669300	500	0,9194	0,0492	2,88E-09	0,8702
chr1	60877900	60878900	1000	0,9378	0,068	0,00E+00	0,8698
chr16	80650400	80650900	500	0,9309	0,0611	2,49E-32	0,8698
chr18	59388800	59389300	500	0,9706	0,1008	4,84E-15	0,8697
chr2	15969400	15970400	1000	0,8939	0,0244	3,07E-52	0,8695
chr13	56207100	56208100	1000	0,9184	0,0505	0,00E+00	0,868

chr3	180395300	180395800	500	0,9877	0,12	1,73E-09	0,8677
chr6	153238200	153239200	1000	0,9123	0,0452	0,00E+00	0,8671
chr18	61635100	61635600	500	0,9268	0,06	2,77E-13	0,8668
chr3	177562200	177563200	1000	0,9121	0,0455	0,00E+00	0,8666
chr4	160368300	160370200	1900	0,9272	0,0606	0,00E+00	0,8665
chr6	144626900	144627400	500	0,9114	0,046	2,10E-12	0,8654
chr16	59885500	59886500	1000	0,9407	0,0757	1,12E-62	0,865
chr1	55667100	55667600	500	0,9095	0,0446	8,62E-39	0,8649
chr2	83598300	83599300	1000	0,9366	0,0718	0,00E+00	0,8648
chr4	105135200	105136200	1000	0,913	0,0486	0,00E+00	0,8644
chr14	32048400	32048900	500	0,9142	0,0499	5,43E-53	0,8643
chr1	223482700	223483700	1000	0,9636	0,0997	2,69E-34	0,864
chr14	47487700	47488200	500	0,915	0,0514	5,45E-33	0,8636
chr3	104515000	104515500	500	1	0,1373	1,08E-06	0,8627
chr7	14008300	14009800	1500	0,9349	0,0725	0,00E+00	0,8624
chr1	243134000	243135500	1500	0,9208	0,0588	0,00E+00	0,8619
chr10	14156400	14156900	500	0,9105	0,0489	0,00E+00	0,8616
chr2	118616200	118617200	1000	0,9178	0,0565	0,00E+00	0,8613
chr17	8455500	8456000	500	0,8941	0,0331	1,94E-18	0,8611
chr12	15392200	15393200	1000	0,9307	0,0697	0,00E+00	0,861
chr8	81275900	81276900	1000	0,9291	0,0684	0,00E+00	0,8606
chr1	234269300	234269800	500	0,9471	0,087	2,25E-25	0,8602
chr1	181970300	181970800	500	0,9167	0,0566	2,15E-08	0,8601
chr2	55775000	55776000	1000	0,9159	0,0559	0,00E+00	0,8599
chr3	88338000	88339000	1000	0,8909	0,0311	0,00E+00	0,8598
chr5	140078700	140079200	500	0,8852	0,0253	0,00E+00	0,8598
chr21	16720900	16721400	500	0,9317	0,0721	1,38E-15	0,8596
chr11	104891100	104892600	1500	0,9164	0,0569	0,00E+00	0,8595
chr1	184204700	184205200	500	0,9194	0,0603	8,16E-16	0,859
chr6	160732500	160733000	500	0,9191	0,0606	1,31E-10	0,8585
chr8	37134300	37134800	500	0,9151	0,0567	1,21E-26	0,8584
chr18	5869800	5870300	500	0,913	0,0548	1,21E-09	0,8582
chr1	98448100	98448600	500	0,9574	0,1	2,08E-05	0,8574
chr3	152897800	152898300	500	0,9402	0,0828	3,28E-18	0,8574
chr1	110304000	110304500	500	0,9783	0,121	2,60E-18	0,8572
chr2	86993600	86995600	2000	0,8965	0,0395	0,00E+00	0,857
chr19	15428100	15430600	2500	0,9424	0,0862	0,00E+00	0,8563

chr13	75176800	75177800	1000	0,9258	0,0697	3,09E-47	0,8561
chr13	24126700	24127200	500	0,9498	0,0938	1,57E-17	0,856
chr16	28238500	28240500	2000	0,9427	0,0868	0,00E+00	0,8559
chr2	158079500	158080500	1000	0,9199	0,0642	0,00E+00	0,8557

Таблица S3A. Перечень первых 100 локусов, определенных как гиперметилированные по данным бисульфитного секвенирования для материнской плазмы в первом триместре

Хромосома	Начало	Конец	Клетки крови матери	CVS	Различие метилирования
chr22	39189067	39189863	0	0,8444	0,8444
chr17	53763065	53764027	0	0,7922	0,7922
chr7	41887694	41888212	0	0,7614	0,7614

Хромосома	Начало	Конец	Клетки крови матери	CVS	Различие метилования
chr2	1,14E+08	1,14E+08	0	0,751	0,751
chr12	25096242	25097206	0	0,7098	0,7098
chr1	66574104	66574793	0	0,7025	0,7025
chr6	11489985	11490755	0	0,7004	0,7004
chr6	1,07E+08	1,07E+08	0	0,6978	0,6978
chr10	30858286	30858871	0	0,6693	0,6693
chr17	21131574	21132167	0	0,6496	0,6496
chr18	13454740	13455292	0	0,5468	0,5468
chr16	11298755	11299326	0	0,5373	0,5373
chr2	1,75E+08	1,75E+08	0	0,5196	0,5196
chr19	44060511	44061036	0	0,5128	0,5128
chr6	1,08E+08	1,08E+08	0	0,5	0,5
chr3	71261611	71262501	0	0,4587	0,4587
chr9	36247847	36248885	0	0,447	0,447
chr19	17819240	17820082	0	0,4279	0,4279
chr17	53769900	53770731	0	0,4102	0,4102
chr1	1,12E+08	1,12E+08	0,0002	0,6167	0,6166
chr7	1,34E+08	1,34E+08	0,0003	0,4351	0,4348
chr3	11658550	11659929	0,0004	0,4299	0,4295
chr17	53764417	53765963	0,0005	0,7967	0,7961
chr10	11246762	11249052	0,0005	0,4002	0,3997
chr22	28992647	28993434	0,0006	0,8092	0,8087
chr15	62460278	62461007	0,0006	0,4334	0,4328
chr1	31002038	31003474	0,0007	0,5926	0,5919
chr19	3129246	3132159	0,0008	0,7725	0,7717
chr12	1,21E+08	1,21E+08	0,0008	0,7303	0,7295
chr19	12304446	12305741	0,0009	0,6986	0,6978
chr3	67788734	67789395	0,001	0,9131	0,9121
chr9	1,32E+08	1,32E+08	0,001	0,7047	0,7037
chr19	6723370	6724479	0,001	0,689	0,688
chr3	1,84E+08	1,84E+08	0,001	0,4384	0,4374
chr2	53848089	53849214	0,001	0,4368	0,4358
chr17	59450886	59452113	0,0012	0,469	0,4678
chr5	1,72E+08	1,72E+08	0,0014	0,578	0,5766
chr21	35342527	35343373	0,0014	0,5392	0,5378
chr21	45164804	45165437	0,0015	0,4251	0,4236
chrX	3742417	3744601	0,0016	0,4486	0,447
chr21	45158293	45159003	0,0017	0,7799	0,7782
chr7	39839340	39839876	0,0017	0,4074	0,4057
chr2	1,75E+08	1,75E+08	0,0018	0,4816	0,4797
chr12	1,24E+08	1,24E+08	0,0019	0,6306	0,6287

chr3	50352688	50353823	0,002	0,624	0,622
chr9	97264382	97265523	0,0021	0,5008	0,4987
chr7	64178628	64179354	0,0021	0,4088	0,4066
chr9	94767202	94767802	0,0023	0,7568	0,7544
chr5	42986308	42988304	0,0023	0,4882	0,4859
chr17	63854127	63854693	0,0024	0,8266	0,8242
chr12	1,22E+08	1,22E+08	0,0024	0,4869	0,4844
chr17	16260170	16260909	0,0026	0,6404	0,6378
chr4	39874787	39875456	0,0027	0,7233	0,7206

chr12	6441080	6441608	0,0027	0,6228	0,6201
chr19	45015653	45016886	0,0028	0,5444	0,5416
chr6	30757752	30758823	0,0028	0,4783	0,4755
chr6	41636176	41637112	0,0028	0,4254	0,4226
chr12	6315199	6315765	0,0029	0,4613	0,4584
chr14	76576283	76577070	0,0029	0,4365	0,4336
chr16	48857790	48858300	0,0031	0,5625	0,5594
chr5	1,7E+08	1,7E+08	0,0031	0,4752	0,4721
chr13	26897813	26898557	0,0032	0,4354	0,4322
chr14	52753948	52754571	0,0032	0,4221	0,4189
chr1	1,66E+08	1,66E+08	0,0033	0,5579	0,5545
chr12	56157424	56158348	0,0033	0,47	0,4667
chr22	16079971	16080532	0,0034	0,6226	0,6193
chr7	1946410	1946975	0,0036	0,6826	0,6789
chr11	258799	259749	0,0036	0,5072	0,5037
chr6	13381944	13382477	0,0037	0,5945	0,5908
chr7	1,27E+08	1,27E+08	0,0037	0,5096	0,5058
chr13	23743886	23744467	0,0037	0,4534	0,4497
chr2	1,21E+08	1,21E+08	0,0038	0,7175	0,7137
chr21	25855853	25857105	0,0039	0,4661	0,4622
chr2	43211724	43212565	0,0039	0,4345	0,4306
chr12	1,08E+08	1,08E+08	0,0041	0,6024	0,5983
chr15	92928924	92929575	0,0041	0,4074	0,4033
chr19	10731043	10731636	0,0042	0,5868	0,5826
chr6	1,45E+08	1,45E+08	0,0043	0,5783	0,574
chr1	52875323	52875907	0,0044	0,4145	0,4101
chr14	75058186	75058956	0,0045	0,602	0,5975
chr12	1,21E+08	1,21E+08	0,0045	0,4821	0,4776
chr17	76873737	76874417	0,0046	0,6012	0,5966
chr2	2,38E+08	2,38E+08	0,0049	0,7654	0,7604
chr2	1,98E+08	1,98E+08	0,0049	0,7228	0,7179
chr6	1,47E+08	1,47E+08	0,0049	0,4967	0,4918
chr9	1,36E+08	1,36E+08	0,0049	0,4584	0,4535
chr1	67545402	67546771	0,005	0,4971	0,4921
chr6	1,58E+08	1,58E+08	0,0052	0,6145	0,6093
chr3	1,7E+08	1,7E+08	0,0052	0,5845	0,5794
chr1	2,34E+08	2,34E+08	0,0053	0,7033	0,6979
chr10	80715722	80716751	0,0053	0,6515	0,6462
chr4	48602901	48603736	0,0053	0,6315	0,6262
chr19	13957965	13958580	0,0053	0,599	0,5937
chr1	90081114	90082367	0,0053	0,4574	0,4521
chr2	1,06E+08	1,06E+08	0,0054	0,8858	0,8804

chr16	29664213	29665369	0,0054	0,8339	0,8285
chr1	1,59E+08	1,59E+08	0,0054	0,7663	0,7608
chr13	97926489	97927025	0,0054	0,6229	0,6175
chr1	41604452	41605277	0,0054	0,6011	0,5956
chr9	1,28E+08	1,28E+08	0,0054	0,5871	0,5818

Таблица S3B. Перечень первых 100 локусов, определенных как гипометилированные по данным бисульфитного секвенирования для материнской плазмы в первом триместре

Хромосома	Начало	Конец	Клетки крови матери	CVS	Различие метилирования
chr1	235771917	235772426	0,9868	0,549	0,4377
chr1	97357972	97358622	0,9835	0,4805	0,503
chr1	4490516	4491074	0,9826	0,4793	0,5032
chr4	181124168	181124671	0,9825	0,4725	0,5099
chr16	71908694	71909213	0,982	0,5581	0,4239
chr3	182727915	182728477	0,981	0,3577	0,6233
chr5	115339535	115340038	0,9802	0,5455	0,4347
chr3	195855575	195856122	0,9801	0,3793	0,6008
chr6	155437621	155438161	0,9799	0,5991	0,3808
chr9	20468093	20468904	0,9798	0,4271	0,5527
chr10	90702298	90702987	0,9787	0,3324	0,6463
chr1	170581654	170582162	0,9785	0,4817	0,4968
chr3	108816849	108817794	0,9783	0,4793	0,4989
chr20	36912749	36913319	0,9783	0,5	0,4783
chr13	72517281	72517839	0,9782	0,4855	0,4927
chr12	103553001	103553677	0,9774	0,492	0,4854
chr22	27638905	27639408	0,9766	0,5385	0,4382
chr7	17290850	17291462	0,9763	0,59	0,3863
chr6	17227866	17228510	0,976	0,4058	0,5703
chr15	56998547	56999107	0,9754	0,3766	0,5988
chr7	70965945	70966842	0,9753	0,5893	0,386
chr3	32159338	32160065	0,9752	0,5379	0,4372
chr16	17043258	17043854	0,9752	0,5521	0,4231
chr16	22776223	22776850	0,9752	0,5735	0,4017
chr5	169344029	169344869	0,9751	0,4211	0,5541
chr11	34324955	34325722	0,975	0,5561	0,4189
chr8	58554745	58555376	0,9747	0,5784	0,3964
chr1	153933389	153934121	0,9746	0,463	0,5116
chr14	88003983	88004485	0,9745	0,5379	0,4366
chr3	151738501	151739120	0,9741	0,4901	0,484
chr14	105618699	105619606	0,974	0,3457	0,6283
chr16	24060085	24060702	0,9738	0,3991	0,5747
chr8	68941792	68942711	0,9738	0,5449	0,429
chr12	53208707	53209304	0,9737	0,4847	0,489
chr7	76892564	76893249	0,9736	0,5664	0,4072
chr3	69464294	69464971	0,9736	0,5893	0,3843
chr19	61401137	61401745	0,9732	0,4933	0,4799
chr11	124569867	124570490	0,9732	0,5136	0,4595
chr18	42618440	42619096	0,9732	0,5942	0,379
chr5	169398896	169399637	0,9731	0,498	0,4751
chr5	169328124	169328983	0,9731	0,572	0,401
chr20	34679880	34680448	0,9731	0,5922	0,3809
chr16	9042198	9042702	0,973	0,4286	0,5444
chr10	90205044	90205701	0,973	0,4407	0,5323
chr13	33236454	33236997	0,973	0,5906	0,3824

chr16	73284579	73285087	0,9729	0,5602	0,4127
-------	----------	----------	--------	--------	--------

chr8	29100691	29101428	0,9728	0,505	0,4678
chr2	202383851	202384447	0,9727	0,5461	0,4267
chr3	179501620	179502300	0,9722	0,5766	0,3956
chr6	107674976	107675906	0,9719	0,4434	0,5285
chr6	107880632	107881161	0,9718	0,5623	0,4095
chr12	56350283	56350933	0,9718	0,5909	0,3809
chr19	40636458	40637339	0,9717	0,4941	0,4776
chr2	223472599	223473287	0,9714	0,1824	0,7891
chr22	20709067	20709787	0,9714	0,5149	0,4565
chr19	46095583	46096190	0,9713	0,5385	0,4328
chr6	90258338	90259318	0,9712	0,3415	0,6297
chr2	54598347	54598933	0,9712	0,5894	0,3819
chr3	114810453	114811493	0,9711	0,5166	0,4545
chr19	15851125	15851654	0,9711	0,5236	0,4476
chr8	42889138	42890084	0,9711	0,5652	0,4059
chr18	52354390	52355064	0,971	0,598	0,373
chr15	38206236	38207010	0,9709	0,4186	0,5523
chr7	99700554	99701110	0,9708	0,305	0,6658
chr12	19487336	19487855	0,9708	0,4105	0,5603
chr7	87996908	87997437	0,9708	0,5462	0,4246
chr6	63628653	63629378	0,9707	0,529	0,4417
chr15	38209108	38209618	0,9706	0,5882	0,3824
chr19	6623769	6624450	0,9704	0,5179	0,4526
chr2	10794513	10795242	0,9704	0,5976	0,3728
chr2	118472785	118474454	0,9704	0,5992	0,3712
chr5	57820209	57820801	0,9701	0,5815	0,3886
chr10	100183380	100184702	0,9701	0,5826	0,3875
chr2	8151989	8152646	0,97	0,4701	0,4999
chr10	3938374	3938914	0,9699	0,1741	0,7958
chr9	123724524	123725439	0,9697	0,57	0,3997
chr14	89085469	89086097	0,9696	0,3278	0,6418
chr16	14129437	14130133	0,9695	0,5304	0,4392
chr5	60746367	60747191	0,9695	0,5571	0,4124
chr1	92002953	92003729	0,9694	0,52	0,4494
chr6	31264677	31265413	0,9693	0,5135	0,4558
chr7	99317013	99318281	0,9692	0,5117	0,4574
chr8	8808867	8809422	0,9692	0,5691	0,4002
chr19	20052165	20052720	0,969	0,2792	0,6898
chr8	129139026	129139573	0,969	0,3458	0,6232
chr11	122314929	122315458	0,969	0,4232	0,5458
chr13	98377663	98378165	0,9688	0,3319	0,6369
chr9	107606194	107606872	0,9688	0,449	0,5198
chr8	56096904	56097736	0,9688	0,5267	0,4422
chr7	128093836	128094339	0,9688	0,5929	0,3758
chr2	103109370	103109916	0,9686	0,3333	0,6352
chr3	101803534	101804063	0,9686	0,5027	0,4659
chr10	69505720	69506278	0,9684	0,2515	0,7169
chr13	26608225	26608754	0,9683	0,3614	0,6069
chr1	90993315	90993828	0,9683	0,5519	0,4164
chr6	11361243	11361801	0,9681	0,2578	0,7103

chr21	36529300	36529981	0,968	0,1944	0,7736
chr21	37813953	37814521	0,9679	0,2175	0,7505
chr2	15226273	15227211	0,9679	0,5134	0,4545

chr19	4102809	4103443	0,9679	0,5646	0,4034
-------	---------	---------	--------	--------	--------

Таблица S3C. Перечень первых 100 локусов, определенных как гиперметилированные по данным бисульфитного секвенирования для материнской плазмы в третьем триместре

Хромосома	Начало	Конец	Клетки крови матери	Зрелая плацента	Различие метилирования
chr17	53763065	53764027	0,0000	0,8680	0,8680
chr22	39189067	39189863	0,0000	0,8233	0,8233
chr10	30858286	30858871	0,0000	0,7713	0,7713
chr7	41887694	41888212	0,0000	0,7578	0,7578
chr2	1,14E+08	1,14E+08	0,0000	0,7500	0,7500
chr12	25096242	25097206	0,0000	0,7332	0,7332
chr6	1,07E+08	1,07E+08	0,0000	0,7229	0,7229
chr1	66574104	66574793	0,0000	0,7136	0,7136
chr16	11298755	11299326	0,0000	0,7005	0,7005
chr6	11489985	11490755	0,0000	0,6935	0,6935
chr18	13454740	13455292	0,0000	0,6594	0,6594
chr6	1,08E+08	1,08E+08	0,0000	0,6231	0,6231
chrX	3627885	3628549	0,0000	0,6133	0,6133
chr12	7979754	7980413	0,0000	0,6118	0,6118
chr3	71261611	71262501	0,0000	0,5938	0,5938
chr17	53769900	53770731	0,0000	0,5586	0,5586
chr11	1,18E+08	1,18E+08	0,0000	0,5558	0,5558
chr19	44060511	44061036	0,0000	0,5464	0,5464
chr2	2,38E+08	2,38E+08	0,0000	0,5330	0,5330
chr1	1,91E+08	1,91E+08	0,0000	0,5294	0,5294
chr1	1,44E+08	1,44E+08	0,0000	0,4857	0,4857
chr2	1,75E+08	1,75E+08	0,0000	0,4785	0,4785
chr4	15366889	15367646	0,0000	0,4729	0,4729
chr2	19537237	19537737	0,0000	0,4599	0,4599
chr1	1,15E+08	1,15E+08	0,0000	0,4351	0,4351
chr1	1,54E+08	1,54E+08	0,0000	0,4299	0,4299
chr14	51383387	51384149	0,0000	0,4186	0,4186
chr1	1,12E+08	1,12E+08	0,0002	0,5350	0,5348
chr3	11658550	11659929	0,0004	0,5579	0,5575
chr17	53764417	53765963	0,0005	0,7894	0,7889
chr22	28992647	28993434	0,0006	0,8053	0,8047
chr6	27214981	27215823	0,0006	0,4593	0,4587
chr1	31002038	31003474	0,0007	0,6309	0,6302
chr12	1,21E+08	1,21E+08	0,0008	0,7360	0,7352
chr19	3129246	3132159	0,0008	0,7257	0,7249
chr19	12304446	12305741	0,0009	0,6397	0,6388
chr6	28723918	28724965	0,0009	0,4344	0,4335
chr19	6723370	6724479	0,0010	0,7280	0,7270
chr2	53848089	53849214	0,0010	0,4060	0,4050
chr9	1,32E+08	1,32E+08	0,0010	0,7558	0,7548
chr3	67788734	67789395	0,0010	0,9219	0,9209
chr19	18276368	18277132	0,0011	0,5136	0,5125
chr17	59450886	59452113	0,0012	0,5196	0,5184
chr17	74243852	74244670	0,0012	0,4117	0,4105

chr3	1,85E+08	1,85E+08	0,0014	0,4961	0,4948
chr21	35342527	35343373	0,0014	0,5126	0,5112
chr5	1,72E+08	1,72E+08	0,0014	0,6531	0,6516
chr21	45164804	45165437	0,0015	0,4364	0,4349
chrX	3742417	3744601	0,0016	0,7517	0,7501
chr21	45158293	45159003	0,0017	0,8180	0,8163
chr2	1,75E+08	1,75E+08	0,0018	0,6214	0,6196
chr12	1,24E+08	1,24E+08	0,0019	0,5906	0,5887
chr3	50352688	50353823	0,0020	0,6082	0,6062
chr9	94767202	94767802	0,0023	0,8327	0,8304
chr17	63854127	63854693	0,0024	0,7886	0,7862
chr12	1,22E+08	1,22E+08	0,0024	0,5021	0,4997
chr17	16260170	16260909	0,0026	0,5780	0,5754
chr12	6441080	6441608	0,0027	0,7471	0,7444
chr4	39874787	39875456	0,0027	0,7962	0,7935
chr18	50536032	50536649	0,0027	0,4920	0,4893
chr6	30757752	30758823	0,0028	0,4029	0,4001
chr19	2571763	2572292	0,0031	0,4200	0,4169
chr5	1,7E+08	1,7E+08	0,0031	0,4218	0,4187
chr13	26897813	26898557	0,0032	0,6485	0,6453
chr12	56157424	56158348	0,0033	0,5541	0,5508
chr1	1,66E+08	1,66E+08	0,0033	0,5147	0,5113
chr22	16079971	16080532	0,0034	0,6265	0,6231
chr6	16820551	16821134	0,0035	0,4800	0,4765
chr11	258799	259749	0,0036	0,5475	0,5439
chr7	1946410	1946975	0,0036	0,8251	0,8215
chr6	13381944	13382477	0,0037	0,8221	0,8183
chr7	1,27E+08	1,27E+08	0,0037	0,4767	0,4730
chr2	1,21E+08	1,21E+08	0,0038	0,6734	0,6697
chr2	43211724	43212565	0,0039	0,4256	0,4217
chr15	92928924	92929575	0,0041	0,5605	0,5564
chr12	1,08E+08	1,08E+08	0,0041	0,7313	0,7271
chr19	10731043	10731636	0,0042	0,5668	0,5626
chr6	1,45E+08	1,45E+08	0,0043	0,5910	0,5867
chr1	52875323	52875907	0,0044	0,6115	0,6071
chr12	1,21E+08	1,21E+08	0,0045	0,5884	0,5839
chr14	75058186	75058956	0,0045	0,6534	0,6489
chr17	76873737	76874417	0,0046	0,5658	0,5612
chr6	1,47E+08	1,47E+08	0,0049	0,5826	0,5777
chr2	1,98E+08	1,98E+08	0,0049	0,7944	0,7895
chr2	2,38E+08	2,38E+08	0,0049	0,7328	0,7278
chr8	1,42E+08	1,42E+08	0,0050	0,7728	0,7679
chr3	1,7E+08	1,7E+08	0,0052	0,7227	0,7176
chr6	1,58E+08	1,58E+08	0,0052	0,6389	0,6337
chr2	2,38E+08	2,38E+08	0,0052	0,4238	0,4185
chr11	56948854	56949496	0,0053	0,4484	0,4431
chr4	48602901	48603736	0,0053	0,5920	0,5867
chr5	1,31E+08	1,31E+08	0,0053	0,4858	0,4805
chr10	80715722	80716751	0,0053	0,5249	0,5196
chr19	13957965	13958580	0,0053	0,4379	0,4326
chr1	2,34E+08	2,34E+08	0,0053	0,8440	0,8387

chr13	97926489	97927025	0,0054	0,7233	0,7179
chr9	1,28E+08	1,28E+08	0,0054	0,7312	0,7258

chr2	1,06E+08	1,06E+08	0,0054	0,8513	0,8459
chr2	96556705	96557637	0,0054	0,4095	0,4041
chr16	29664213	29665369	0,0054	0,8837	0,8783

Таблица S3D. Перечень первых 100 локусов, определенных как гипометилированные по данным бисульфитного секвенирования для материнской плазмы в третьем триместре

Хромосома	Начало	Конец	Клетки крови матери	Зрелая плацента	Различие метилирования
chr10	7548948	7549483	0,9866	0,5685	0,4181
chr1	4490516	4491074	0,9826	0,5015	0,4810
chr4	1,81E+08	1,81E+08	0,9825	0,5981	0,3843
chr3	1,83E+08	1,83E+08	0,9810	0,2925	0,6886
chr3	1,96E+08	1,96E+08	0,9801	0,4643	0,5158
chr6	1,55E+08	1,55E+08	0,9799	0,4610	0,5189
chr1	1,71E+08	1,71E+08	0,9785	0,5122	0,4662
chr20	36912749	36913319	0,9783	0,4513	0,5269
chr22	38583100	38583616	0,9783	0,5428	0,4355
chr1	19391314	19392207	0,9778	0,5273	0,4505
chr5	1,74E+08	1,74E+08	0,9770	0,5852	0,3918
chr19	13678906	13679531	0,9760	0,5812	0,3949
chr14	83650790	83651395	0,9760	0,5378	0,4382
chr15	56998547	56999107	0,9754	0,4691	0,5063
chr16	22776223	22776850	0,9752	0,5114	0,4638
chr5	1,69E+08	1,69E+08	0,9751	0,4809	0,4943
chr8	58554745	58555376	0,9747	0,5977	0,3770
chr14	1,06E+08	1,06E+08	0,9740	0,2069	0,7671
chr8	68941792	68942711	0,9738	0,5872	0,3866
chr16	24060085	24060702	0,9738	0,3470	0,6268
chr12	53208707	53209304	0,9737	0,5278	0,4459
chr5	1,69E+08	1,69E+08	0,9731	0,5057	0,4673
chr16	9042198	9042702	0,9730	0,1860	0,7869
chr10	90205044	90205701	0,9730	0,5922	0,3808
chr3	1,89E+08	1,89E+08	0,9720	0,4949	0,4771
chr6	1,08E+08	1,08E+08	0,9719	0,5825	0,3894
chr2	2,23E+08	2,23E+08	0,9714	0,3333	0,6381
chr19	46095583	46096190	0,9713	0,5065	0,4648
chr8	1,41E+08	1,41E+08	0,9713	0,5753	0,3959
chr6	90258338	90259318	0,9712	0,4357	0,5355
chr13	51403556	51404069	0,9710	0,3980	0,5731
chr18	66875048	66875726	0,9710	0,5259	0,4451
chr7	99700554	99701110	0,9708	0,3757	0,5951
chr7	87996908	87997437	0,9708	0,5720	0,3988
chr19	6623769	6624450	0,9704	0,4774	0,4930
chr1	97639047	97639749	0,9701	0,4148	0,5553
chr16	23892096	23892772	0,9701	0,5000	0,4701
chr10	3938374	3938914	0,9699	0,1148	0,8551
chr14	89085469	89086097	0,9696	0,2964	0,6732
chr8	1,29E+08	1,29E+08	0,9690	0,3565	0,6126
chr13	98377663	98378165	0,9688	0,3123	0,6566
chr8	56096904	56097736	0,9688	0,4562	0,5126
chr2	1,03E+08	1,03E+08	0,9686	0,3459	0,6227

Хромосома	Начало	Конец	Клетки крови матери	Зрелая плацента	Различие метилювания
chr13	26608225	26608754	0,9683	0,4562	0,5121
chr2	22738157	22738760	0,9682	0,5122	0,4560
chr6	11361243	11361801	0,9681	0,2646	0,7035
chr21	36529300	36529981	0,9680	0,1829	0,7852

chr21	37813953	37814521	0,9679	0,3061	0,6619
chr2	2,43E+08	2,43E+08	0,9679	0,5750	0,3929
chr4	12413543	12414103	0,9679	0,5944	0,3735
chr3	1,27E+08	1,27E+08	0,9677	0,4030	0,5648
chr7	33509047	33509556	0,9676	0,4627	0,5048
chr14	59284846	59285553	0,9674	0,5254	0,4420
chr17	42623453	42624024	0,9673	0,4318	0,5355
chr19	6778363	6779377	0,9671	0,4416	0,5255
chr4	41798250	41798788	0,9670	0,5000	0,4670
chr5	88054080	88054588	0,9669	0,2238	0,7431
chr16	24109379	24110289	0,9669	0,5062	0,4607
chr10	13847159	13847895	0,9667	0,3188	0,6479
chr10	1,27E+08	1,27E+08	0,9667	0,5423	0,4244
chr12	1,12E+08	1,12E+08	0,9663	0,3722	0,5941
chr10	17220886	17221845	0,9662	0,4455	0,5207
chr8	5947355	5947862	0,9662	0,5171	0,4491
chr3	73740840	73741439	0,9659	0,3657	0,6002
chr14	57945953	57946875	0,9658	0,5357	0,4301
chr14	50905777	50906333	0,9658	0,3008	0,6650
chr15	90275374	90276000	0,9657	0,5409	0,4248
chr22	24717299	24718197	0,9657	0,5160	0,4497
chr7	36530128	36530987	0,9656	0,5194	0,4462
chr2	1,31E+08	1,31E+08	0,9655	0,4384	0,5271
chr4	42116988	42117788	0,9654	0,5195	0,4459
chr12	1,16E+08	1,16E+08	0,9653	0,5594	0,4059
chr2	7491785	7492736	0,9652	0,4556	0,5097
chr19	6599638	6600187	0,9652	0,5488	0,4163
chr6	25326803	25327398	0,9651	0,3974	0,5677
chr4	1,7E+08	1,7E+08	0,9651	0,4933	0,4718
chr7	99875338	99876155	0,9650	0,2696	0,6953
chr14	97144328	97145208	0,9649	0,5377	0,4272
chr3	11718596	11719163	0,9649	0,5521	0,4128
chr14	1E+08	1E+08	0,9649	0,3794	0,5855
chr7	1,5E+08	1,5E+08	0,9648	0,3327	0,6322
chr12	56357827	56358328	0,9648	0,4217	0,5430
chr10	8275750	8276276	0,9647	0,3100	0,6547
chr11	16999685	17000209	0,9647	0,2765	0,6882
chr22	34419356	34419861	0,9646	0,4245	0,5401
chr18	72453151	72453725	0,9646	0,4700	0,4946
chr5	49919879	49920699	0,9645	0,3169	0,6476
chr1	24580891	24581805	0,9643	0,3565	0,6078
chr22	18233774	18234492	0,9641	0,5205	0,4436
chr14	45356178	45356903	0,9640	0,3934	0,5706
chr3	53007193	53008661	0,9638	0,4902	0,4737
chr4	55027912	55028539	0,9637	0,5254	0,4384
chr5	1,37E+08	1,37E+08	0,9637	0,5290	0,4347

chr1	2,23E+08	2,23E+08	0,9636	0,0997	0,8640
chr7	1,35E+08	1,35E+08	0,9636	0,2959	0,6677
chr5	80350438	80351169	0,9636	0,4969	0,4667
chr12	31889600	31890343	0,9636	0,1745	0,7891
chr12	8365395	8366096	0,9636	0,5721	0,3914

chr19	15424819	15425355	0,9635	0,2836	0,6799
chr10	10985469	10986409	0,9635	0,4877	0,4759

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ анализа биологического образца организма, где указанный биологический образец содержит молекулы нуклеиновых кислот, происходящие из нормальных клеток и потенциально из клеток, связанных с раковым заболеванием, причем по меньшей мере некоторые из молекул нуклеиновых кислот в биологическом образце являются свободными от клеток, при этом указанный способ включает:

анализ множества молекул ДНК из указанного биологического образца, отличающийся тем, что анализ молекулы ДНК включает:

определение локализации указанной молекулы ДНК в геноме указанного организма;

определение того, метилирована ли указанная молекула ДНК в одном или большем числе сайтов;

для каждого из множества сайтов:

определение соответствующего количества молекул ДНК, которые метилированы в указанном сайте;

вычисление первого уровня метилирования на основе соответствующих количеств молекул ДНК, метилированных во множестве сайтов;

сравнение первого уровня метилирования с первым предельным значением;

определение первой классификации уровня рака на основе указанного сравнения.

2. Способ по п. 1, отличающийся тем, что определение того, метилирована ли указанная молекула ДНК в одном или большем числе сайтов, включает:

осуществление чувствительного к метилированию секвенирования.

3. Способ по п. 2, отличающийся тем, что осуществление чувствительного к метилированию секвенирования включает:

обработку молекул ДНК бисульфитом натрия; и

осуществление секвенирования обработанных молекул ДНК.

4. Способ по п. 3, отличающийся тем, что обработка свободных от клеток молекул ДНК бисульфитом натрия представляет собой часть бисульфитной конверсии с использованием Tet или окислительного бисульфитного секвенирования для детекции 5-гидроксиметилцитозина.

5. Способ по п. 2, отличающийся тем, что относительные количества молекул ДНК для каждой области определяют путем выравнивания ридов последовательностей, полученных при чувствительном к метилированию секвенировании.

6. Способ по п. 1, отличающийся тем, что определение того, метилирована ли

указанная молекула ДНК в одном или большем числе сайтов, включает применение расщепления чувствительными к метилированию рестрикционными ферментами, специфической к метилированию ПЦР, зависимой от метилирования преципитации ДНК, связывающего метилированную ДНК белка/пептида или мономолекулярного секвенирования без обработки бисульфитом натрия.

7. Способ по п. 1, также включающий:

для каждой из первого множества областей генома:

определение соответствующего количества молекул ДНК, происходящих из указанной области;

вычисление соответствующего нормированного значения по соответствующему количеству;

сравнение соответствующего нормированного значения с референсным значением для определения того, наблюдается ли в соответствующей области делеция или амплификация;

определение первого количества областей, для которых было определено присутствие делеции или амплификации;

сравнение указанного первого количества с первым пороговым значением для определения второй классификации уровня рака; и

использование первой классификации и второй классификации для определения третьей классификации уровня рака.

8. Способ по п. 7, отличающийся тем, что первое пороговое значение представляет собой процент от первого множества областей, для которых было определено присутствие делеции или амплификации.

9. Способ по п. 7, отличающийся тем, что третья классификация положительна по раковому заболеванию только в том случае, если и первая классификация, и вторая классификация указывают на раковое заболевание.

10. Способ по п. 7, отличающийся тем, что третья классификация положительна по раковому заболеванию в том случае, если либо первая классификация, либо вторая классификация указывают на раковое заболевание.

11. Способ по п. 1 или п. 7, отличающийся тем, что первая классификация указывает на существование ракового заболевания в организме, и включающий дополнительно:

идентификацию типа ракового заболевания, связанного с указанным организмом, путем сравнения первого уровня метилирования с соответствующим значением, определенным для других организмов, при этом по меньшей мере два из указанных

других организмов идентифицированы как имеющие различные типы раковых заболеваний.

12. Способ по п. 11, отличающийся тем, что третья классификация указывает на раковое заболевание в организме, и включающий дополнительно: идентификацию типа ракового заболевания, связанного с указанным организмом, путем сравнения первого количества областей с соответствующими значениями, определенными для других организмов.

13. Способ по п. 1 или 7, отличающийся тем, что вычисление первого уровня метилирования включает:

идентификацию второго множества областей генома;

идентификацию одного или большего количества сайтов в составе каждой из указанных областей;

вычисление уровня метилирования области для каждой области, при этом первый уровень метилирования соответствует первой области, при этом указанный способ дополнительно включает:

сравнение каждого из уровней метилирования области с соответствующим предельным значением для области, включающее сравнение первого уровня метилирования с первым предельным значением;

определение второго количества областей которые, как было определено, имеют уровень метилирования области, превышающий соответствующее предельное значение для области; и

сравнение указанного второго количества областей со вторым пороговым значением для определения первой классификации.

14. Способ по п. 13, отличающийся тем, что указанные области, которые, как было определено, имеют уровень метилирования области, превышающий соответствующее предельное значение для области, соответствуют первому набору областей, и включающий дополнительно: сравнение уровней метилирования области из первого набора областей с соответствующими уровнями метилирования области у других организмов для первого набора областей, при этом у указанных других организмов имеется по меньшей мере два из следующего: раковое заболевание первого типа, отсутствие ракового заболевания и раковое заболевание второго типа; и

определение того, имеется ли у организма раковое заболевание первого типа, отсутствует ли раковое заболевание или имеется ли раковое заболевание второго типа на основании указанного сравнения.

15. Способ по п. 14, также включающий:

разбиение на кластеры указанных других организмов на основании уровней метилирования области из первого набора областей, причем два из указанных кластеров соответствуют любым двум пунктам из: первого типа ракового заболевания, отсутствия ракового заболевания и второго типа ракового заболевания,

при этом сравнение уровней метилирования области определяет, к какому кластеру принадлежит организм.

16. Способ по п. 15, отличающийся тем, что при разбиении на кластеры указанных других организмов используются уровни метилирования области указанного организма.

17. Способ по п. 15, отличающийся тем, что указанные кластеры включают первый кластер, соответствующий первому типу раковых заболеваний, второй кластер, соответствующий второму типу раковых заболеваний, и третий кластер, соответствующий отсутствию раковых заболеваний.

18. Способ по п. 15, отличающийся тем, что разбиение на кластеры указанных других организмов основано дополнительно на соответствующих нормированных значениях второго набора областей для указанных других организмов, причем указанный второй набор областей соответствует областям, для которых было определено присутствие делеции или амплификации, и при этом соответствующее нормированное значение для области определяют по соответствующему количеству молекул ДНК из указанной области; при этом указанный способ дополнительно включает:

для каждой из второго набора областей:

определение соответствующего количества молекул ДНК как происходящих из указанной области;

вычисление соответствующего нормированного значения по соответствующему количеству; и

сравнение соответствующих нормированных значений второго набора областей для указанного организма с соответствующими нормированными значениями для других организмов, в рамках определения того, к какому кластеру принадлежит указанный организм.

19. Способ по п. 18, отличающийся тем, что разбиение на кластеры указанных других организмов основано дополнительно на соответствующих плотностях метилирования гиперметилированных CpG-островков, и указанный способ дополнительно включает:

для каждого из гиперметилированных CpG-островков:

определение соответствующей плотности метилирования,

сравнение соответствующих плотностей метилирования гиперметилированных

СpG-островков у указанного организма с плотностями метилирования у других организмов в рамках определения того, к какому кластеру принадлежит указанный организм.

20. Способ по п. 13, также включающий:

для каждой из второго множества областей:

вычисление соответствующей разности между уровнем метилирования области и соответствующим предельным значением для области; и

вычисление соответствующей вероятности, соответствующей разности;

при этом определение второго количества областей включает:

вычисление кумулятивного показателя, включающего соответствующие вероятности.

21. Способ по п. 20, отличающийся тем, что вычисление указанного кумулятивного показателя включает:

вычисление логарифма соответствующей вероятности для получения соответствующего логарифмического результата; и

вычисление суммы, включающей соответствующие логарифмические результаты.

22. Способ по п. 21, отличающийся тем, что указанный кумулятивный показатель представляет собой отрицательное значение суммы соответствующих логарифмических результатов.

23. Способ по п. 20, отличающийся тем, что каждая разность нормирована по стандартному отклонению, связанному с соответствующим предельным значением для области.

24. Способ по п. 20, отличающийся тем, что соответствующая вероятность связана с вероятностью для соответствующей разности согласно статистическому распределению.

25. Способ по п. 20, отличающийся тем, что второе пороговое значение соответствует максимальному кумулятивному показателю от референсной группы.

26. Способ по п. 13, также включающий:

для каждой из первого множества областей:

вычисление соответствующей разности между соответствующим нормированным значением и референсным значением; и

вычисление соответствующей вероятности, связанной с соответствующей разностью;

при этом определение первого количества областей включает:

вычисление первой суммы, включающей соответствующие вероятности.

27. Способ по п. 13, отличающийся тем, что соответствующие предельные

значения для областей представляют собой указанное количество от референсного уровня метилирования.

28. Способ по п. 13, отличающийся тем, что второе пороговое значение представляет собой процент, и при этом сравнение второго количества областей с вторым пороговым значением включает:

деление указанного второго количества областей на второе число второго множества областей перед сравнением с указанным вторым пороговым значением.

29. Способ по п. 28, отличающийся тем, что указанное второе число соответствует всему второму множеству областей.

30. Способ по п. 13, отличающийся тем, что первое множество областей совпадает с вторым множеством областей, и при этом соответствующие предельные значения для областей зависят от того, присутствует ли в соответствующей области делеция или амплификация.

31. Способ по п. 30, отличающийся тем, что соответствующее предельное значение для области имеет большую величину, если в соответствующей области наблюдается амплификация, чем в тех случаях, когда амплификация не наблюдается, и при этом соответствующее предельное значение для области имеет меньшую величину, если в соответствующей области наблюдается делеция, чем в тех случаях, когда делеция не наблюдается.

32. Способ по п. 31, отличающийся тем, что на основании соответствующих предельных значений для областей определяют гипометилирование второго множества областей, при этом соответствующее предельное значение для области имеет более отрицательное значение, если в соответствующей области наблюдается амплификация, чем в тех случаях, когда амплификация не наблюдается, и при этом соответствующее предельное значение для области имеет менее отрицательное значение, если в соответствующей области наблюдается делеция, чем в тех случаях, когда делеция не наблюдается.

33. Способ по п. 13, отличающийся тем, что указанный биологический образец берут до лечения, и дополнительно включающий:

повторение способа по п. 13 для другого биологического образца, взятого после лечения, для получения:

следующего первого количества областей, для которых было определено присутствие делеции или амплификации; и

следующего второго количества областей, которые, как было определено, имеют уровень метилирования области, превышающий соответствующее предельное значение

для области;

сравнение первого количества со следующим первым количеством и второго количества со следующим вторым количеством для определения прогноза для указанного организма.

34. Способ по п. 33, отличающийся тем, что сравнение первого количества со следующим первым количеством и второго количества со следующим вторым количеством для определения прогноза для указанного организма включает:

определение первой разности между первым количеством и следующим первым количеством;

сравнение указанной первой разности с одним или несколькими порогами для первой разности;

определение второй разности между вторым количеством и следующим вторым количеством; и

сравнение второй разности с одним или несколькими порогами второй разности.

35. Способ по п. 34, отличающийся тем, что делают более плохой прогноз, если первая разность ниже одного порога первой разности, чем в тех случаях, когда указанная первая разность превышает один порог первой разности, и при этом делают более плохой прогноз, если вторая разность ниже одного порога второй разности, чем в тех случаях, когда указанная вторая разность превышает один порог второй разности.

36. Способ по п. 35, отличающийся тем, что один порог первой разности и один порог второй разности равны 0.

37. Способ по п. 33, отличающийся тем, что указанное лечение представляет собой иммунотерапию, хирургию, радиационную терапию, химиотерапию, терапию на основе антител, эпигенетическую терапию или таргетную терапию.

38. Способ по п. 1, отличающийся тем, что первое предельное значение представляет собой указанное расстояние от референсного уровня метилирования, установленного по биологическому образцу, полученному из здорового организма.

39. Способ по п. 38, отличающийся тем, что указанное расстояние представляет собой указанное число стандартных отклонений от референсного уровня метилирования.

40. Способ по п. 1, отличающийся тем, что первое предельное значение устанавливают по референсному уровню метилирования, определенному по предыдущему биологическому образцу из указанного организма, полученному раньше, чем исследуемый биологический образец.

41. Способ по п. 1, отличающийся тем, что сравнение первого уровня метилирования с первым предельным значением включает:

определение разности между первым уровнем метилирования и референсным уровнем метилирования; и

сравнение указанной разности с порогом, соответствующим первому предельному значению.

42. Способ по п. 1, также включающий:

определение фракционной концентрации опухолевой ДНК в биологическом образце;

вычисление первого предельного значения на основе фракционной концентрации.

43. Способ по п. 1, также включающий:

определение того, превышает ли фракционная концентрация опухолевой ДНК в биологическом образце минимальное значение; и

в том случае, если фракционная концентрация не превышает минимальное значение, маркировку указанного биологического образца.

44. Способ по п. 43, отличающийся тем, что минимальное значение определяют на основе ожидаемой разности между уровнями метилирования для опухоли и референсного уровня метилирования.

45. Способ по п. 1, также включающий:

измерение размера молекул ДНК в множестве сайтов; и

нормировку первого уровня метилирования на основе измеренных размеров молекул ДНК перед сравнением первого уровня метилирования с первым предельным значением.

46. Способ по п. 45, отличающийся тем, что нормировка первого уровня метилирования на основе измеренных размеров включает:

отбор молекул ДНК, имеющих первый размер;

использование выбранных молекул ДНК для расчета первого уровня метилирования, первого предельного значения, соответствующего первому размеру.

47. Способ по п. 46, отличающийся тем, что первый размер представляет собой диапазон длин.

48. Способ по п. 46, отличающийся тем, что молекулы ДНК выбирают на основании физического разделения, которое зависит от размера.

49. Способ по п. 46, отличающийся тем, что отбор молекул ДНК первого размера включает:

осуществление массивно-параллельного секвенирования спаренных концов множества молекул ДНК для получения пар последовательностей для каждой из молекул ДНК;

определение размера молекулы ДНК путем сравнения пары последовательностей с референсным геномом; и

отбор молекул ДНК, имеющих первый размер.

50. Способ по п. 45, отличающийся тем, что нормировка первого уровня метилирования на основе измеренных размеров включает:

определение функциональной зависимости между размером и уровнями метилирования; и

использование указанной функциональной зависимости для нормировки первого уровня метилирования.

51. Способ по п. 50, отличающийся тем, что указанная функциональная зависимость обеспечивает масштабирующие значения, связанные с соответствующими размерами.

52. Способ по п. 51, также включающий:

вычисление среднего размера, соответствующего молекулам ДНК, используемым для расчета первого уровня метилирования; и

умножение указанного первого уровня метилирования на соответствующее масштабирующее значение.

53. Способ по п. 51, также включающий:

для каждого из множества сайтов:

для каждой из молекул ДНК, локализованных в указанном сайте:

получение соответствующего размера молекулы ДНК в указанном сайте; и

использование масштабирующего значения, соответствующего указанному соответствующему размеру, для нормировки вклада молекулы ДНК в соответствующее количество молекул ДНК, которые метилированы в указанном сайте.

54. Способ по п. 1, отличающийся тем, что множество сайтов включает CpG-сайты, причем указанные CpG-сайты организованы во множество CpG-островков, каждый CpG-островок включает один или большее количество CpG-сайтов, при этом первый уровень метилирования соответствует первому CpG-островку.

55. Способ по п. 54, отличающийся тем, что среднее значение плотности метилирования CpG-островков ниже первого процента в референсной группе, и при этом коэффициент вариации для плотности метилирования в референсной группе ниже второго процента.

56. Способ по п. 54, также включающий:

для каждого из CpG-островков:

определение того, гиперметилирован ли CpG-островок относительно референсной

группы образцов от других организмов путем сравнения уровня метилирования CpG-островка с соответствующим предельным значением;

для каждого из гиперметилованных CpG-островков:

определение соответствующей плотности метилирования;

вычисление кумулятивного показателя по соответствующим плотностям метилирования; и

сравнение указанного кумулятивного показателя с кумулятивным предельным значением для определения первой классификации.

57. Способ по п. 56, отличающийся тем, что вычисление кумулятивного показателя по соответствующим плотностям метилирования включает:

для каждого из гиперметилованных CpG-островков:

вычисление соответствующей разности между соответствующей плотностью метилирования и референсной плотностью; и

вычисление соответствующей вероятности, связанной с указанной соответствующей разностью; и

применение соответствующих вероятностей для определения кумулятивного показателя.

58. Способ по п. 56, отличающийся тем, что кумулятивный показатель определяют путем:

вычисления логарифма соответствующей вероятности для получения соответствующего логарифмического результата; и

вычисления суммы, включающей соответствующие логарифмические результаты, кумулятивный показатель представляет собой отрицательное значение суммы.

59. Способ по п. 56, отличающийся тем, что каждая соответствующая разность нормирована по стандартному отклонению, связанному с референсной плотностью.

60. Способ по п. 56, отличающийся тем, что кумулятивное предельное значение соответствует максимальному кумулятивному показателю референсной группы.

61. Способ по п. 56, отличающийся тем, что определение того, гиперметилован ли первый CpG-островок, включает:

сравнение первого уровня метилирования с первым предельным значением и с третьим предельным значением,

при этом первое предельное значение соответствует среднему значению плотностей метилирования для референсной группы плюс указанный процент, а третье предельное значение соответствует указанному числу стандартных отклонений плюс среднее значение плотностей метилирования для референсной группы.

62. Способ по п. 61, отличающийся тем, что указанный процент равен 2%.

63. Способ по п. 61, отличающийся тем, что указанное число стандартных отклонений равно трем.

64. Способ по п. 1, также включающий:

для каждой из первого множества областей генома:

определение соответствующего количества молекул ДНК как происходящих из указанной области;

вычисление соответствующего нормированного значения по соответствующему количеству;

сравнение указанного соответствующего нормированного значения с референсным значением для определения того, наблюдается ли в соответствующей области делеция или амплификация;

определение первого набора областей, все из которых, как было определено, демонстрируют что-либо одно из: делеции, амплификации или нормальной представленности, при этом первый уровень метилирования соответствует первому набору областей;

определение второго набора областей, все из которых, как определено демонстрируют что-либо второе из: делеции, амплификации или нормальной представленности;

вычисление второго уровня метилирования на основе соответствующих количеств молекул ДНК, метилированных в сайтах второго набора областей,

при этом сравнение первого уровня метилирования с первым предельным значением включает:

вычисление параметра, связывающего первый уровень метилирования и второе метилирование; и

сравнение указанного параметра с первым предельным значением.

65. Способ по п. 64, отличающийся тем, что первый уровень метилирования представляет собой статистический показатель уровней метилирования области, рассчитанный для каждой области из первого набора областей, и при этом второй уровень метилирования представляет собой статистический показатель уровней метилирования области, рассчитанный для каждой области второго набора областей.

66. Способ по п. 65, отличающийся тем, что указанный статистический показатель определяют с применением t-критерия Стьюдента, дисперсионного анализа (ANOVA) или критерия Краскела—Уоллиса.

67. Способ по п. 64, отличающийся тем, что указанный параметр включает

отношение или разность.

68. Способ по п. 67, отличающийся тем, что вычисление указанного параметра включает применение распределения вероятностей к отношению или разности.

69. Способ определения первого профиля метилирования по биологическому образцу организма, отличающийся тем, что указанный биологический образец включает свободную от клеток ДНК, содержащую смесь молекул свободной от клеток ДНК, происходящих из первой ткани и из второй ткани, при этом указанный способ включает:

получение второго профиля метилирования, соответствующего молекулам ДНК второй ткани, при этом указанный второй профиль метилирования отражает плотность метилирования в каждом из множества локусов в геноме указанного организма, и плотность метилирования в конкретном локусе соответствует доле метилированных молекул ДНК второй ткани;

определение профиля метилирования свободной от клеток ДНК по молекулам свободной от клеток ДНК в указанной смеси, причем указанный профиль метилирования свободной от клеток ДНК отражает плотность метилирования в каждом из множества локусов;

определение процента свободных от клеток молекул ДНК в указанной смеси, происходящих из первой ткани; и

определение первого профиля метилирования первой ткани путем:

для каждого из множества локусов:

вычисления дифференциального параметра, который включает разность между плотностью метилирования второго профиля метилирования и плотностью метилирования профиля метилирования свободной от клеток ДНК, при этом указанную разность выражают в процентах.

70. Способ по п. 69, также включающий:

определение первого уровня метилирования по первому профилю метилирования;

и

сравнение первого уровня метилирования с первым предельным значением для установления классификации уровня рака на основе указанного сравнения.

71. Способ по п. 69, также включающий:

преобразование указанного первого профиля метилирования для получения скорректированного первого профиля метилирования.

72. Способ по п. 71, отличающийся тем, что указанное преобразование представляет собой линейное преобразование.

73. Способ по п. 69, отличающийся тем, что дифференциальный параметр D для

локуса определяют как D , где mbs обозначает плотность метилирования второго профиля метилирования в указанном локусе, tr обозначает плотность метилирования профиля метилирования свободной от клеток ДНК в указанном локусе, f представляет собой процент молекул свободной от клеток ДНК из первой ткани в биологическом образце, и CN представляет число копий в указанном локусе.

74. Способ по п. 73, отличающийся тем, что $CN = 1$ для числа копий в первой ткани, дисомного в указанном локусе.

75. Способ по п. 73, также включающий:

идентификацию областей, где D превышает порог.

76. Способ по п. 69, также включающий:

идентификацию множества локусов путем выбора локусов, отвечающих любому одному или нескольким из следующих критериев:

содержание GC выше 50%;

плотность метилирования второго профиля метилирования ниже первого порога или выше второго порога; и

минимум пять CpG-сайтов в области, определяемой локусом.

77. Способ по п. 69, отличающийся тем, что указанный биологический образец получен от субъекта женского пола, беременного плодом, и при этом первая ткань получена от плода или из плаценты, и вторая ткань получена от указанного субъекта женского пола.

78. Способ по п. 77, также включающий:

определение первого уровня метилирования по первому профилю метилирования;

и

сравнение первого уровня метилирования с первым предельным значением для определения классификации медицинского состояния.

79. Способ по п. 78, отличающийся тем, что указанное медицинское состояние представляет собой преэклампсию или хромосомную аномалию в первой ткани.

80. Способ по п. 69, отличающийся тем, что первая ткань происходит из опухоли внутри организма, а вторая ткань происходит из незлокачественной ткани указанного организма.

81. Способ по п. 69, отличающийся тем, что указанный биологический образец также включает молекулы ДНК из второй ткани, и при этом получение второго профиля метилирования включает:

анализ молекул ДНК из второй ткани для определения указанного второго профиля метилирования.

82. Способ по п. 69, отличающийся тем, что полученный второй профиль метилирования соответствует средним профилям метилирования, полученным в группе контрольных образцов.

83. Способ по п. 69, отличающийся тем, что определение плотности метилирования профиля метилирования свободной от клеток ДНК для первого локуса из множества локусов включает:

определение первого числа свободных от клеток молекул ДНК, происходящих из первого локуса;

определение того, метилирована ли каждая молекула ДНК из первого числа в одном или большем числе сайтов, для получения второго числа метилированных молекул ДНК; и

вычисление плотности метилирования по первому числу и второму числу.

84. Способ по п. 83, отличающийся тем, что один или большее количество сайтов представляют собой CpG-сайты.

85. Способ по п. 69, отличающийся тем, что определение профиля метилирования свободной от клеток ДНК включает:

обработку свободных от клеток молекул ДНК бисульфитом натрия; и

осуществление секвенирования обработанных молекул ДНК.

86. Способ по п. 85, отличающийся тем, что обработка свободных от клеток молекул ДНК представляет собой часть бисульфитной конверсии с использованием Tet или включает обработку свободных от клеток молекул ДНК перрутеном калия ($KRuO_4$).

87. Способ по п. 69, отличающийся тем, что определение указанного профиля метилирования свободной от клеток ДНК включает:

осуществление мономолекулярного секвенирования ДНК, при котором секвенирование одной молекулы включает определение того, метилирован ли по меньшей мере один сайт в молекуле ДНК.

88. Компьютерный продукт, включающий машиночитаемый носитель для длительного хранения данных, где хранится множество инструкций, при выполнении управляющих компьютерной системой для реализации способа по п. 1.

89. Способ определения первого профиля метилирования по биологическому образцу организма, отличающийся тем, что указанный биологический образец включает свободную от клеток ДНК, содержащую смесь свободной от клеток ДНК, происходящей из первой ткани и из второй ткани, включающий:

анализ множества молекул ДНК из указанного биологического образца, при этом анализ молекулы ДНК включает:

определение локализации указанной молекулы ДНК в геноме указанного организма;

определение генотипа молекулы ДНК;

определение того, метилирована ли указанная молекула ДНК в одном или большем числе сайтов;

идентификацию множества первых локусов, в которых первый геном первой ткани гетерозиготен по соответствующему первому аллелю, и соответствующему второму аллелю, и второй геном второй ткани гомозиготен по соответствующему первому аллелю;

для каждого из первых локусов:

для каждого одного или нескольких сайтов, связанных с указанным локусом:

определение числа молекул ДНК, которые метилированы в указанном сайте и соответствуют соответствующему второму аллелю указанного локуса;

вычисление плотности метилирования на основе числа молекул ДНК, метилированных в одном или большем числе сайтов указанного локуса и соответствующих соответствующему второму аллелю указанного локуса;

создание первого профиля метилирования первой ткани по плотностям метилирования для первых локусов.

90. Способ по п. 89, отличающийся тем, что указанный биологический образец происходит от субъекта женского пола, беременного плодом, и при этом первая ткань получена от плода или из плаценты, а вторая ткань соответствует указанному субъекту женского пола.

91. Способ по п. 89, отличающийся тем, что сайты соответствуют CpG-сайтам.

92. Способ по п. 89, отличающийся тем, что каждый локус включает по меньшей мере один CpG-сайт.

93. Способ по п. 89, отличающийся тем, что каждый локус является смежным по отношению к по меньшей мере одному CpG-сайту.

94. Набор для анализа плодной ДНК в биологическом образце, содержащем смесь плодной ДНК и ДНК от субъекта-женщины, беременной плодом, при этом указанный набор содержит:

один или большее количество олигонуклеотидов для специфической гибридизации по меньшей мере с частью геномной области, приведенной в таблицах S2A, S2B, S2C, S2D, S3A, S3B, S3C и S3D.

95. Набор по п. 94, отличающийся тем, что по меньшей мере один из указанных олигонуклеотидов представляет собой праймер, соответствующий по меньшей мере части одной из указанных геномных областей.

96. Набор по п. 94, отличающийся тем, что по меньшей мере один из указанных олигонуклеотидов представляет собой зонд, соответствующий по меньшей мере части одной из указанных геномных областей.

97. Набор по п. 94, также включающий:

один или большее количество чувствительных к метилированию рестрикционных ферментов.

98. Набор по п. 94, отличающийся тем, что по меньшей мере один из одного или нескольких чувствительных к метилированию рестрикционных ферментов расщепляет молекулы ДНК, которые гипометилированы.

99. Набор по п. 94, отличающийся тем, что по меньшей мере один из одного или нескольких чувствительных к метилированию рестрикционных ферментов расщепляет молекулы ДНК, которые гиперметилированы.

100. Набор по п. 94, отличающийся тем, что по меньшей мере один из одного или нескольких олигонуклеотидов выбран из таблицы S2A.

101. Набор по п. 94, отличающийся тем, что по меньшей мере один из одного или нескольких олигонуклеотидов выбран из таблицы S2B.

102. Набор по п. 94, отличающийся тем, что по меньшей мере один из одного или нескольких олигонуклеотидов выбран из таблицы S2C.

103. Набор по п. 94, отличающийся тем, что по меньшей мере один из одного или нескольких олигонуклеотидов выбран из таблицы S2D.

104. Набор по п. 94, отличающийся тем, что по меньшей мере один из одного или нескольких олигонуклеотидов выбран из таблицы S3A.

105. Набор по п. 94, отличающийся тем, что по меньшей мере один из одного или нескольких олигонуклеотидов выбран из таблицы S3B.

106. Набор по п. 94, отличающийся тем, что по меньшей мере один из одного или нескольких олигонуклеотидов выбран из таблицы S3C.

107. Набор по п. 94, отличающийся тем, что по меньшей мере один из одного или нескольких олигонуклеотидов выбран из таблицы S3D.

108. Способ определения уровня метилирования ДНК в биологическом образце организма, включающий:

для каждого размера из множества размеров:

измерение количества множества фрагментов ДНК из указанного биологического образца, соответствующих указанному размеру;

вычисление с применением компьютерной системы первого значения первого параметра, на основании количества фрагментов ДНК нескольких размеров, при это

указанный первый параметр обеспечивает статистическую меру размерного профиля фрагментов ДНК в биологическом образце;

сравнение первой размерной величины с референсным значением размера, при это указанное референсное значение размера рассчитывали по фрагментам ДНК референсного образца; и

оценку уровня метилирования на основе указанного сравнения.

109. Способ по п. 108, отличающийся тем, что измерение количества фрагментов ДНК, соответствующих указанным размерам, включает:

для каждого из множества фрагментов ДНК из указанного биологического образца: измерение размера фрагмента ДНК,

при этом указанное измерение размера фрагмента ДНК включает:

осуществление секвенирования фрагмента ДНК методом спаренных концов для получения парных ридов последовательностей;

выравнивание парных ридов последовательности с референсным геномом; и

использование выравниваемых локализаций для определения размера фрагмента ДНК.

110. Способ по п. 109, отличающийся тем, что уровень метилирования соответствует конкретной области, за счет определения первой размерной величины только по фрагментам ДНК, которые выравниваются с указанной конкретной областью.

111. Способ по п. 108, отличающийся тем, что измерение количества фрагментов ДНК, соответствующих указанным размерам, включает применение электрофореза.

112. Способ по п. 108, отличающийся тем, что референсный образец получают из организма в более раннее время, чем указанный биологический образец, при этом определяемый уровень метилирования представляет собой относительный уровень, соответствующий величине увеличения или уменьшения первого значения размера относительно референсного значения размера.

113. Способ по п. 108, отличающийся тем, что референсное значение размера представляет собой часть калибровочной кривой.

114. Способ по п. 108, отличающийся тем, что уровень метилирования представляет собой плотность метилирования.

115. Способ по п. 114, отличающийся тем, что плотность метилирования представляет собой плотность метилирования CpG-сайтов.

116. Способ детекции хромосомной аномалии по биологическому образцу организма, отличающийся тем, что указанный биологический образец включает свободную от клеток ДНК, содержащую смесь свободной от клеток ДНК, происходящей

из первой ткани и из второй ткани, и включающий:

анализ множества молекул ДНК из указанного биологического образца, причем анализ молекулы ДНК включает:

определение локализации указанной молекулы ДНК в референсном геноме;

определение того, метилирована ли указанная молекула ДНК в одном или большем числе сайтов;

для каждого из множества сайтов:

определение соответствующего количества молекул ДНК, которые метилированы в указанном сайте;

вычисление первого уровня метилирования первой хромосомной области на основе соответствующих количеств молекул ДНК, метилированных в сайтах в составе первой хромосомной области;

сравнение первого уровня метилирования с предельным значением;

определение классификации аномалии в первой ткани для первой хромосомной области на основе указанного сравнения.

117. Способ по п. 116, отличающийся тем, что сравнение первого уровня метилирования с предельным значением включает:

нормировку первого уровня метилирования и сравнение нормированного первого уровня метилирования с предельным значением.

118. Способ по п. 117, отличающийся тем, что при нормировке используется второй уровень метилирования второй хромосомной области.

119. Способ по п. 117, отличающийся тем, что для нормировки используется фракционная концентрация свободной от клеток ДНК из первой ткани.

120. Способ по п. 116, отличающийся тем, что вычисление первого уровня метилирования включает:

идентификацию множества областей генома указанного организма;

идентификацию одного или большего количества сайтов в составе каждой из указанных областей;

вычисление уровня метилирования области для каждой области, при этом первый уровень метилирования соответствует первой области, и

при этом сравнение первого уровня метилирования с первым предельным значением включает

сравнение каждого из уровней метилирования области с соответствующим предельным значением для области;

определение первого числа областей, уровень метилирования которых превышает

соответствующее предельное значение для области; и

сравнение указанного первого числа с пороговым значением для определения классификации.

121. Способ по п. 120, отличающийся тем, что пороговое значение представляет собой процент, и при этом сравнение указанного первого числа с пороговым значением включает:

деление первого числа областей на второе число областей перед сравнением с пороговым значением.

122. Способ по п. 121, отличающийся тем, что второе число областей представляет собой все области из идентифицированного множества областей.

123. Способ по п. 120, отличающийся тем, что соответствующие предельные значения для областей представлены указанным количеством от референсного уровня метилирования.

124. Способ по п. 116, отличающийся тем, что указанный биологический образец происходит от субъекта женского пола, беременного плодом, при этом первая ткань получена из плода или из плаценты, и вторая ткань получена от указанного субъекта женского пола, и при этом хромосомная аномалия представляет собой хромосомную аномалию плода.

125. Способ по п. 116, отличающийся тем, что предельное значение основано на фоновом уровне метилирования, связанном с ДНК второй ткани субъекта женского пола.

126. Способ по п. 124, отличающийся тем, что предельное значение определяют по другим беременным субъектам женского пола, вынашивающим плоды без хромосомной аномалии в первой хромосомной области.

127. Способ по п. 124, отличающийся тем, что хромосомная аномалия представляет собой трисомию 21, или трисомию 18, или трисомию 13, или синдром Тернера, или синдром Клайнфельтера.

128. Способ по п. 116, отличающийся тем, что первая ткань происходит из опухоли внутри организма, и вторая ткань происходит из незлокачественной ткани указанного организма.

129. Способ по п. 116, отличающийся тем, что предельное значение определяют по другим организмам, не имеющим раковых заболеваний.

130. Способ по п. 116, отличающийся тем, что предельное значение основано на концентрации свободной от клеток ДНК, происходящей из первой ткани в биологическом образце.

131. Способ по п. 130, отличающийся тем, что предельное значение основано на

масштабирующем коэффициенте, соответствующем типу аномалии, при этом указанный тип аномалии представляет собой делецию или дупликацию.

132. Способ по п. 116, отличающийся тем, что определение локализации молекулы ДНК в референсном геноме задействует определение того, локализована ли она в первой хромосомной области.

133. Способ по п. 132, отличающийся тем, что определение локализации молекулы ДНК в референсном геноме осуществляют путем определения того, картируется ли указанная молекула ДНК на первую хромосомную область.

134. Способ по п. 116, отличающийся тем, что указанная хромосомная аномалия представляет собой субхромосомную делецию, или субхромосомную дупликацию, или синдром Ди Джорджи

100

Фиг. 1А

	Материнская кровь (1-ый триместр)		CVS		Зрелая плацента		Материнская плазма (1-ый триместр)		Материнская плазма (3-й триместр)		Материнская плазма (после родов)	
	цель 5'-3'	цель 3'-5'	цель 5'-3'	цель 3'-5'	цель 5'-3'	цель 3'-5'	цель 5'-3'	цель 3'-5'	цель 5'-3'	цель 3'-5'	цель 5'-3'	цель 3'-5'
Необработанные риды	547522712	547522712	422873853	422873853	637537170	637537170	4567226854	4567226854	1289337674	1289337674	1277615263	1277615263
Общая картируемость (%)	45,08	44,39	44,22	43,83	46,70	44,40	29,82	29,48	35,31	35,39	37,86	37,40
Общ. число картируемых ридов	246806492	243029086	186985998	185346026	297735873	283050215	1361958504	1346252254	461659066	456249836	483697803	477778078
Неоднозначность (%) ^a	6,72	6,83	5,84	5,89	5,26	5,54	6,35	643	5,86	5,93	5,98	6,05
Дубликации (%) ^b	6,60	5,86	9,52	9,27	32,94	30,39	64,66	64,41	25,70	25,18	41,01	40,56
Глубина для полного генома после удаления неоднозначных оснований, и	11,29	11,19	8,36	8,31	9,43	9,25	25,40	25,29	16,95	16,85	14,08	14,00
показатель конверсии лямбда (%) ^c	99,98	99,98	99,93	99,94	99,95	99,96	99,96	99,97	99,96	99,96	99,96	99,96
Число M в геномн. посл. C ^d	268498966	266662986	190603545	188460370	226378366	223722866	640697054	635295175	466206420	453669856	397672185	395159669
Число U в геномн. посл. C	4484594136	4449649557	4232382985	4201356735	4968445023	4908568234	1,1699E+10	1,1641E+10	8576357118	8538105685	6896711473	6865335640
Покрывтие геномн. C (%)	66,45	66,39	62,66	62,55	57,49	67,40	80,48	80,46	69,34	69,76	67,68	67,61
Глубина секвенир-я геномн. C	814	8,07	7,57	7,51	8,89	8,78	21,12	21,01	15,46	15,39	12,48	12,43
Пл-сть метилир-я геномн. C (%)	5,65	5,63	4,31	4,29	4,36	4,36	5,19	5,17	5,05	5,04	5,45	5,44
Число M в посл. CpG	265203963	262504347	183242316	181245307	213757505	211325313	627866879	622753476	443226176	440725023	389810636	387417582
Число U в посл. CpG	104163667	103976011	146623817	148498340	147248261	146617702	310000676	308133694	206357402	206477046	142858626	142271164
Покрывтие CpG (%)	85,17	85,12	81,02	80,92	84,23	84,15	9205	92,02	86,13	86,06	84,93	84,86
Глубина секвенир-я CpG	13,12	13,01	11,68	11,71	12,82	12,71	33,30	33,05	23,06	22,94	18,91	18,81
Пл-сть метилир-я CpG (%)	71,80	71,63	66,72	64,97	69,21	59,04	66,95	66,90	68,23	68,20	73,18	73,14
Число не-M в посл. CpG	3279043	3147047	7286715	7162226	12502145	12301418	12818574	12532920	12979169	12843942	7859475	7740641
Число не-U в посл. CpG	4302393851	4285094041	3979826287	3971182968	4560807406	4537834784	11356E+10	11306E+10	8367572669	8330532598	6749132596	6719112682
Покрывтие не-CpG (%)	0,13	0,13	0,12	0,12	0,13	0,13	0,16	0,16	0,14	0,14	0,13	0,13
Глубина секвенир-я не-CpG	1,52	1,52	1,41	1,41	1,62	1,61	4,02	4,00	2,96	2,95	2,39	2,38
Пл-сть метилир-я не-CpG (%)	0,08	0,07	0,18	0,18	0,27	0,27	0,11	0,11	0,15	0,15	0,12	0,12
Число M в посл. CHG	1029734	1000287	2100977	2071015	3223342	3176751	3766662	3691190	3625662	3592617	2347392	2314323
Число U в посл. CHG	1345439602	1341393096	1217982855	1216609371	1362682765	1357568840	3416559203	3404225988	2519517702	2509439947	2036338055	2028059603
Покрывтие CHG (%)	77,90	77,82	73,26	73,13	77,50	77,38	88,16	88,12	79,68	79,59	77,84	77,75
Глубина секвенир-я CHG	10,89	10,85	9,86	9,85	11,04	11,00	27,65	27,56	20,40	20,32	16,48	16,42
Пл-сть метилир-я CHG (%)	0,08	0,07	0,17	0,17	0,24	0,23	0,11	0,11	0,14	0,14	0,12	0,11
Число M в посл. CHN	2249309	2146760	5185738	5091211	9278803	9124GG7	9052012	8841730	9353007	9251325	5512083	5426318
Число U в посл. CHN	2956954249	2943700946	2761843432	2764673697	3198124641	3180265944	7939543368	7903584696	5848054967	6821092651	4712844541	4691053079
Покрывтие CHN (%)	61,96	61,90	58,43	58,32	63,54	63,45	77,53	77,52	65,96	65,89	63,65	63,58
Глубина секвенир-я CHN	6,84	6,81	6,40	0,38	7,42	7,36	18,38	18,30	13,55	13,48	10,91	10,86
Пл-сть метилир-я CHN (%)	0,08	0,07	0,19	0,18	0,29	0,29	0,11	0,11	0,16	0,16	0,12	0,12

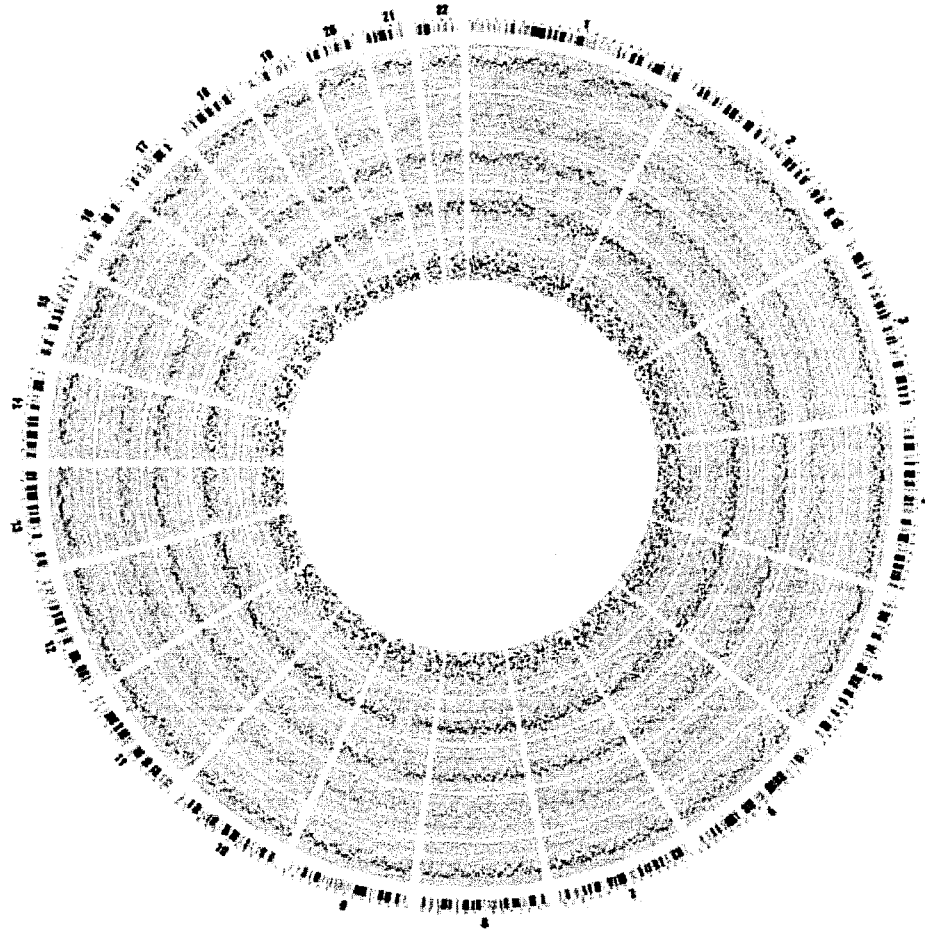
1510341

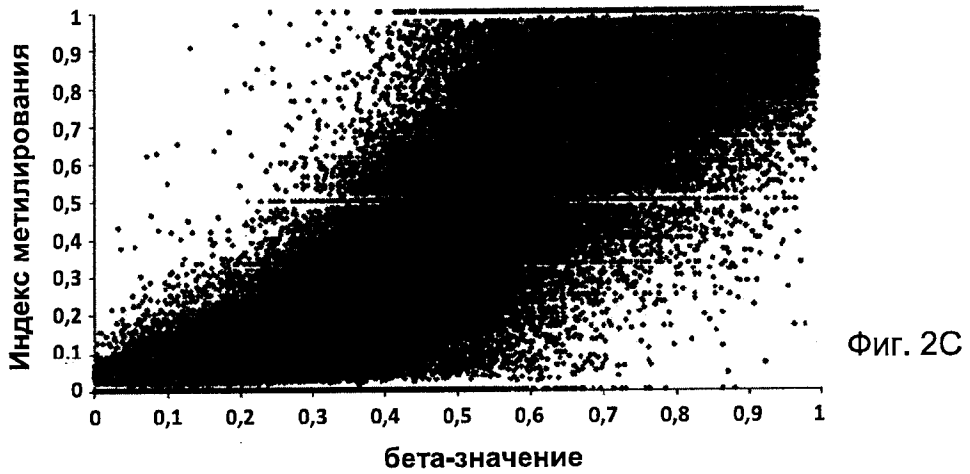
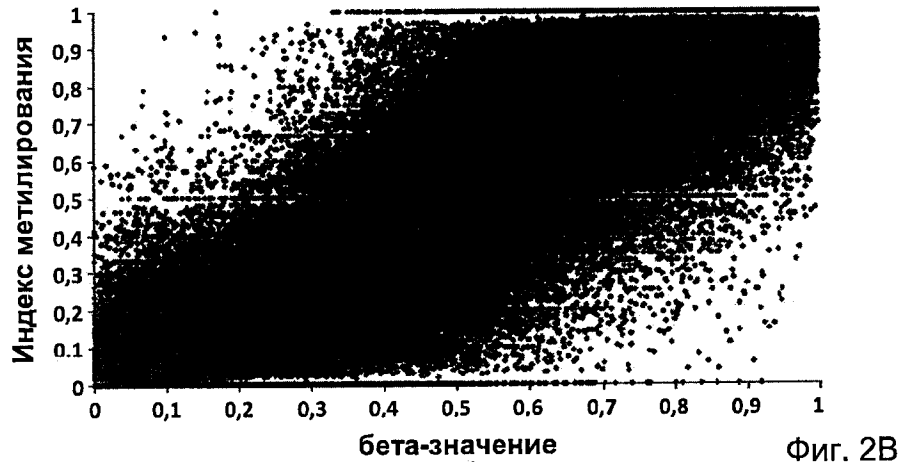
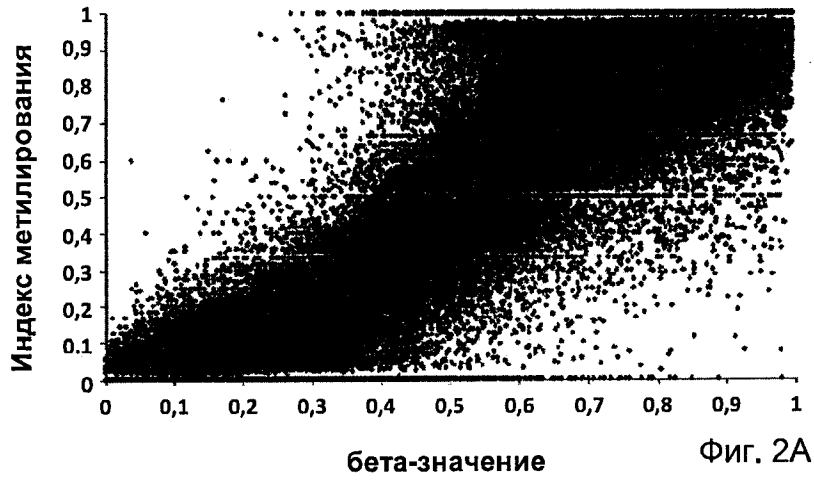
	Кровь взрослого мужчины		Кровь взрослой женщины		Плазма взрослого мужчины		Плазма взрослой женщины	
	цепь 5'–3'	цепь 3'–5'	цепь 5'–3'	цепь 3'–5'	цепь 5'–3'	цепь 3'–5'	цепь 5'–3'	цепь 3'–5'
Необработанные риды	196544792	196544792	197387755	197387755	179688600	179688600	141199439	141199439
Общая картируемость (%)	43,52	42,97	44,94	44,35	22,47	22,29	18,98	18,79
Общее число картируемых ридов	85526774	84462496	88699242	87542953	40380938	40048964	26795699	26537762
Неоднозначность (%) ^a	6,05	6,13	6,29	6,37	6,09	6,14	6,29	6,35
Дубликации (%) ^b	1,99	1,73	1,88	1,53	7,89	7,64	11,71	11,47
Глубина для полного генома после удаления неоднозн. оснований и показатель конверсии лямбда (%) ^c	4,13	4,09	4,28	4,24	1,83	1,82	1,16	1,15
Число M в геномн. посл. C ^d	99,96	99,97	99,97	99,97	99,96	99,97	99,96	99,97
Число U в геномн. посл. C	127154246	125487793	124983462	123238608	55049489	54630306	37403328	37048480
Покрывание геномн. C (%)	2047971831	2026731827	2058153525	2039985644	938186429	933377753	598496591	594974253
Глубина секвенир-я геномн. C	51,66	51,55	55,17	55,06	43,76	43,66	38,39	38,28
Пл-сть метилир-я геномн. C (%)	3,72	3,68	3,74	3,70	1,70	1,69	1,09	1,08
Число M в посл. CpG	5,85	5,83	5,72	5,70	5,54	5,53	5,88	5,86
Число U в посл. CpG	125161042	123578850	123020928	121374309	54155037	53756413	36788819	36453865
Покрывание CpG (%)	49564325	49936003	47562383	47812716	21930864	21330555	13971713	13867179
Глубина секвенир-я CpG	73,37	73,25	76,24	76,12	63,93	63,80	57,87	57,77
Пл-сть метилир-я CpG (%)	6,20	6,16	6,06	6,01	2,70	2,68	1,80	1,79
Число M в посл. не-CpG	71,63	71,22	72,12	71,74	71,18	71,12	72,48	72,44
Число U в посл. не-CpG	1979410	1899750	1951160	1856642	894139	873671	614025	594289
Покрывание не-CpG (%)	1960312392	1950307695	1976031265	1966232257	915375235	910910490	583173346	580050464
Глубина секвенир-я не-CpG	0,10	0,10	0,11	0,11	0,08	0,08	0,07	0,07
Пл-сть метилир-я не-CpG (%)	0,69	0,69	0,70	0,70	0,32	0,32	0,21	0,21
Число M в посл. CHG	0,10	0,10	0,10	0,09	0,10	0,10	0,11	0,10
Число U в посл. CHG	618974	598717	594336	573203	279945	272311	191120	186135
Покрывание CHG (%)	612387336	610126996	607312325	605093927	283916067	282826050	180374803	179612404
Глубина секвенир-я CHG	63,24	63,11	66,75	66,62	54,69	54,56	48,57	48,45
Пл-сть метилир-я CHG (%)	4,96	4,94	4,91	4,90	2,30	2,29	1,46	1,45
Число M в посл. CHH	0,10	0,10	0,10	0,09	0,10	0,10	0,11	0,10
Число U в посл. CHH	1360436	1301033	1356824	1283439	614194	601360	422905	408154
Покрывание CHH (%)	1347925056	1340180699	1368718940	1361138330	631459168	628084440	402798543	400438060
Глубина секвенир-я CHH	46,93	46,83	50,48	50,38	39,32	39,22	34,20	34,11
Пл-сть метилир-я CHH (%)	3,12	3,10	3,17	3,15	1,46	1,45	0,93	0,93
	0,10	0,10	0,10	0,09	0,10	0,10	0,10	0,10

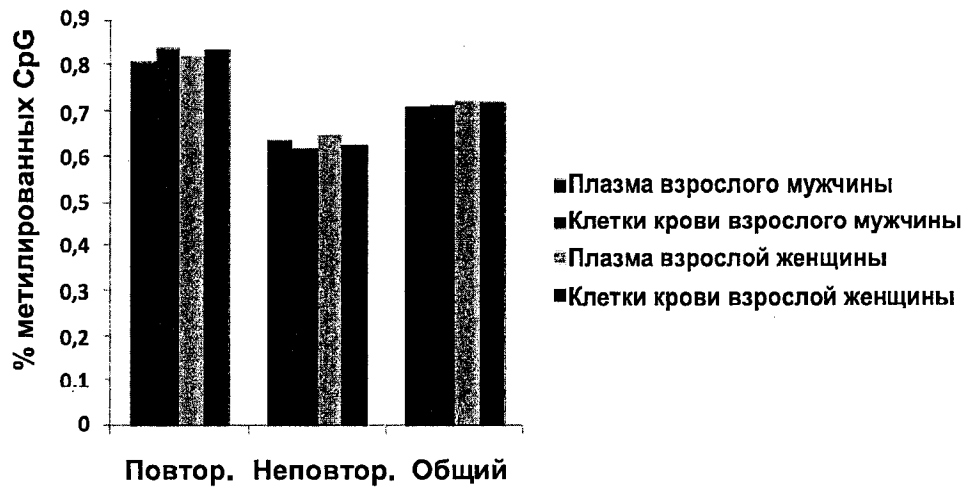
Фиг. 1А (продолжение)

150

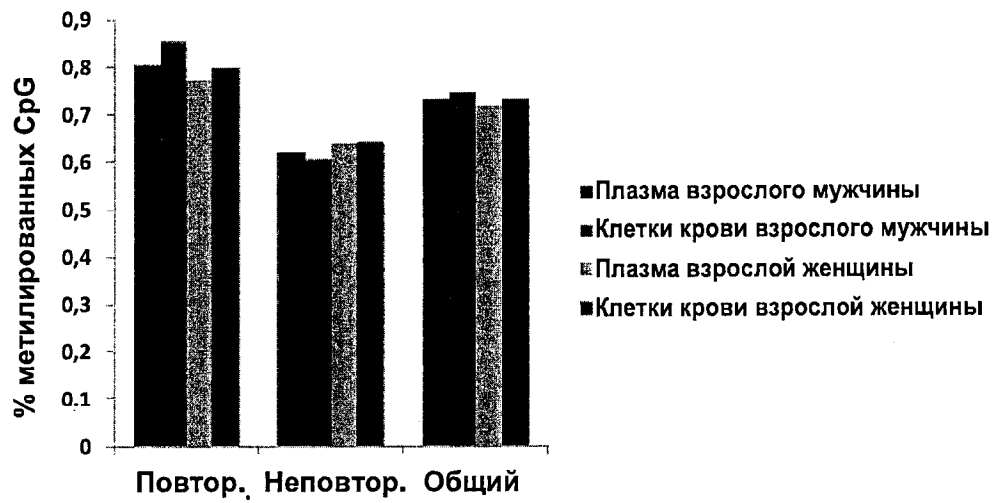
Фиг. 1В



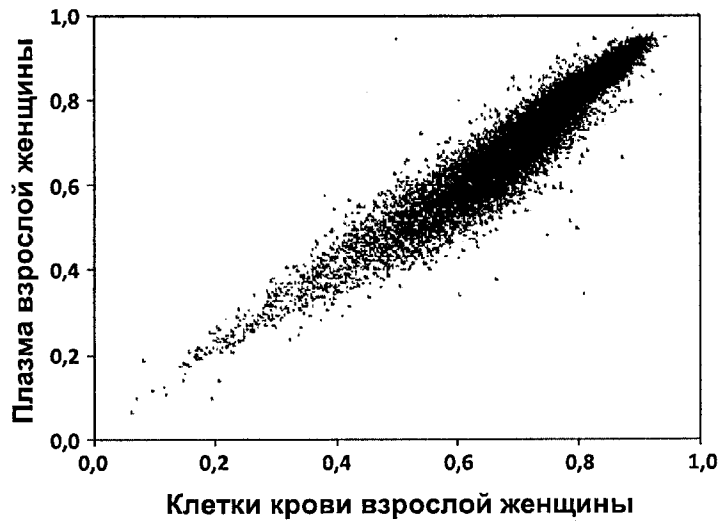




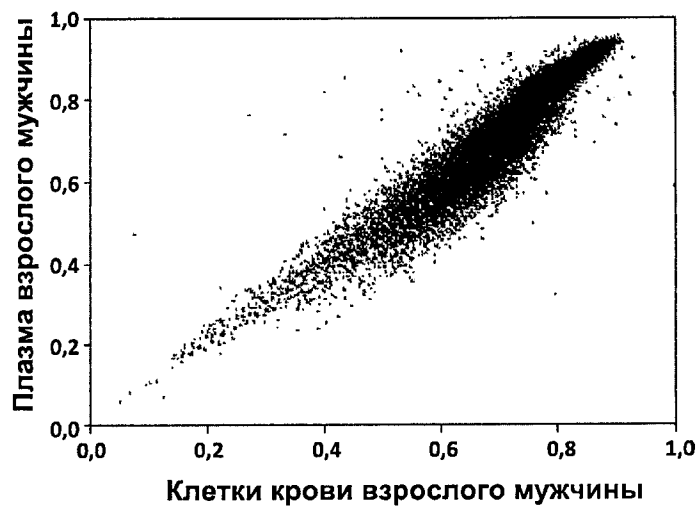
Фиг. 3А



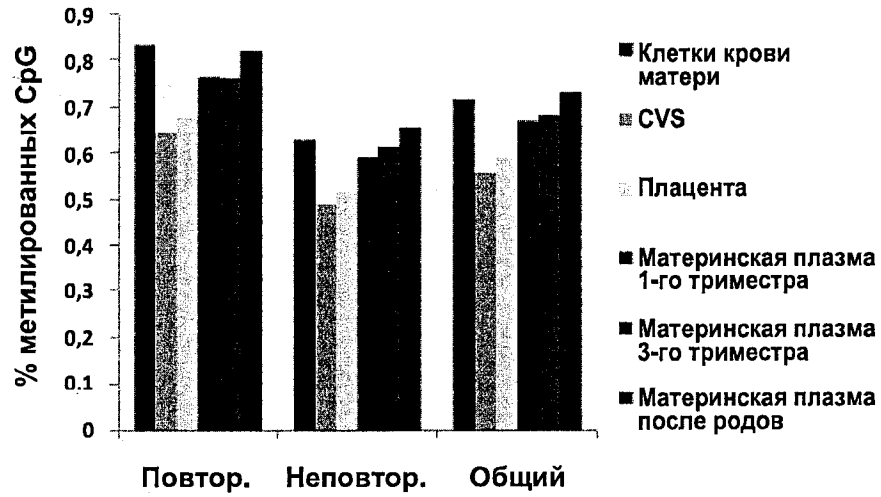
Фиг. 3В



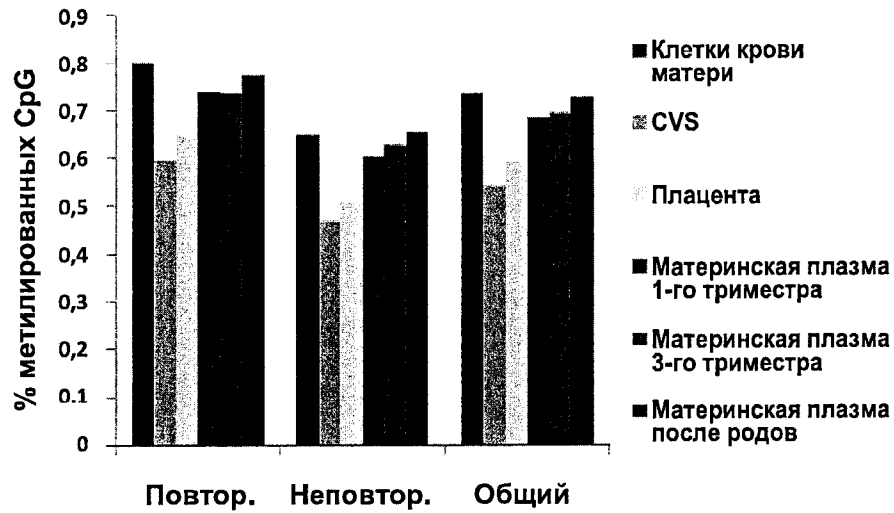
Фиг. 4А



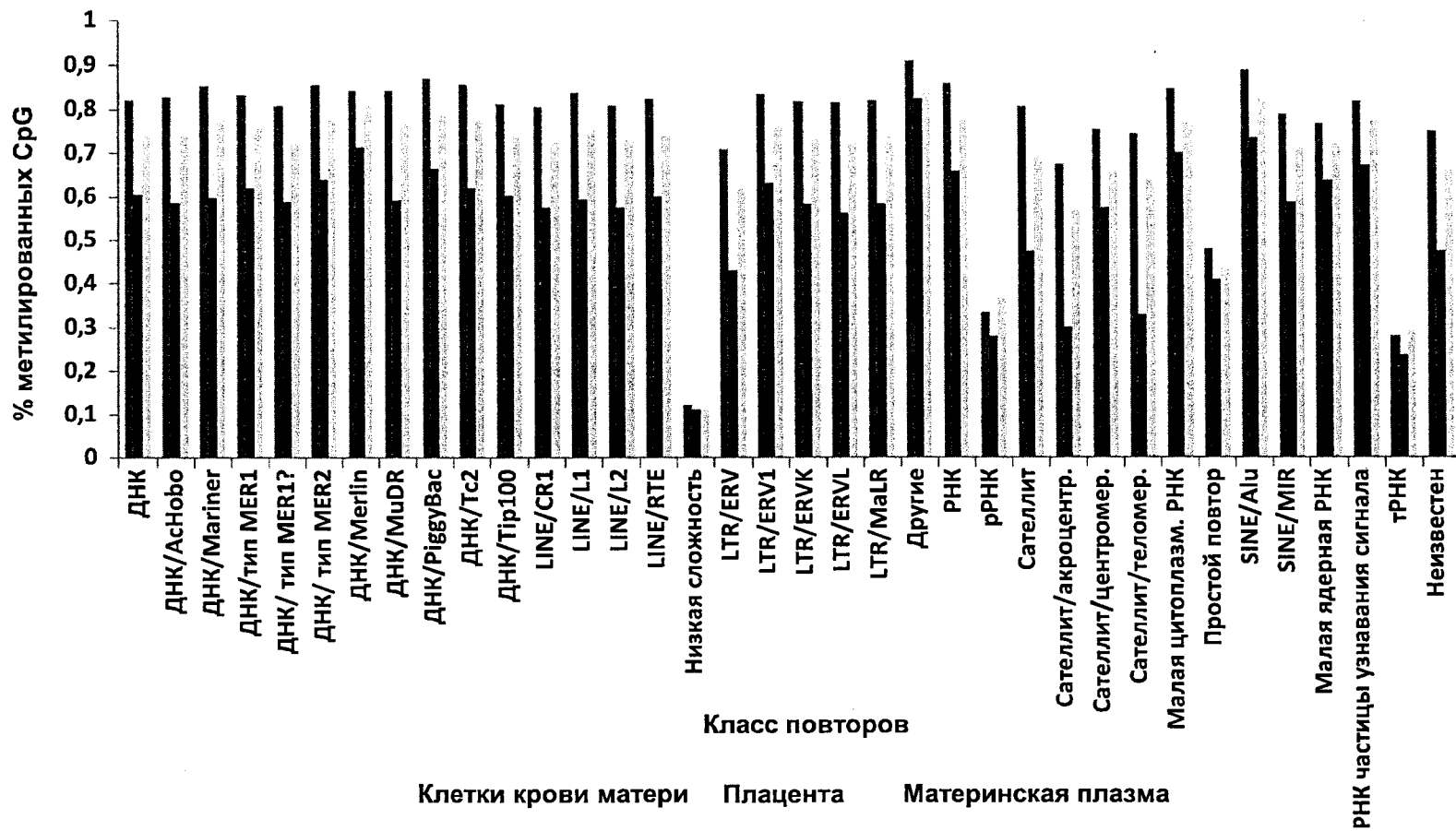
Фиг. 4В



Фиг. 5А



Фиг. 5В

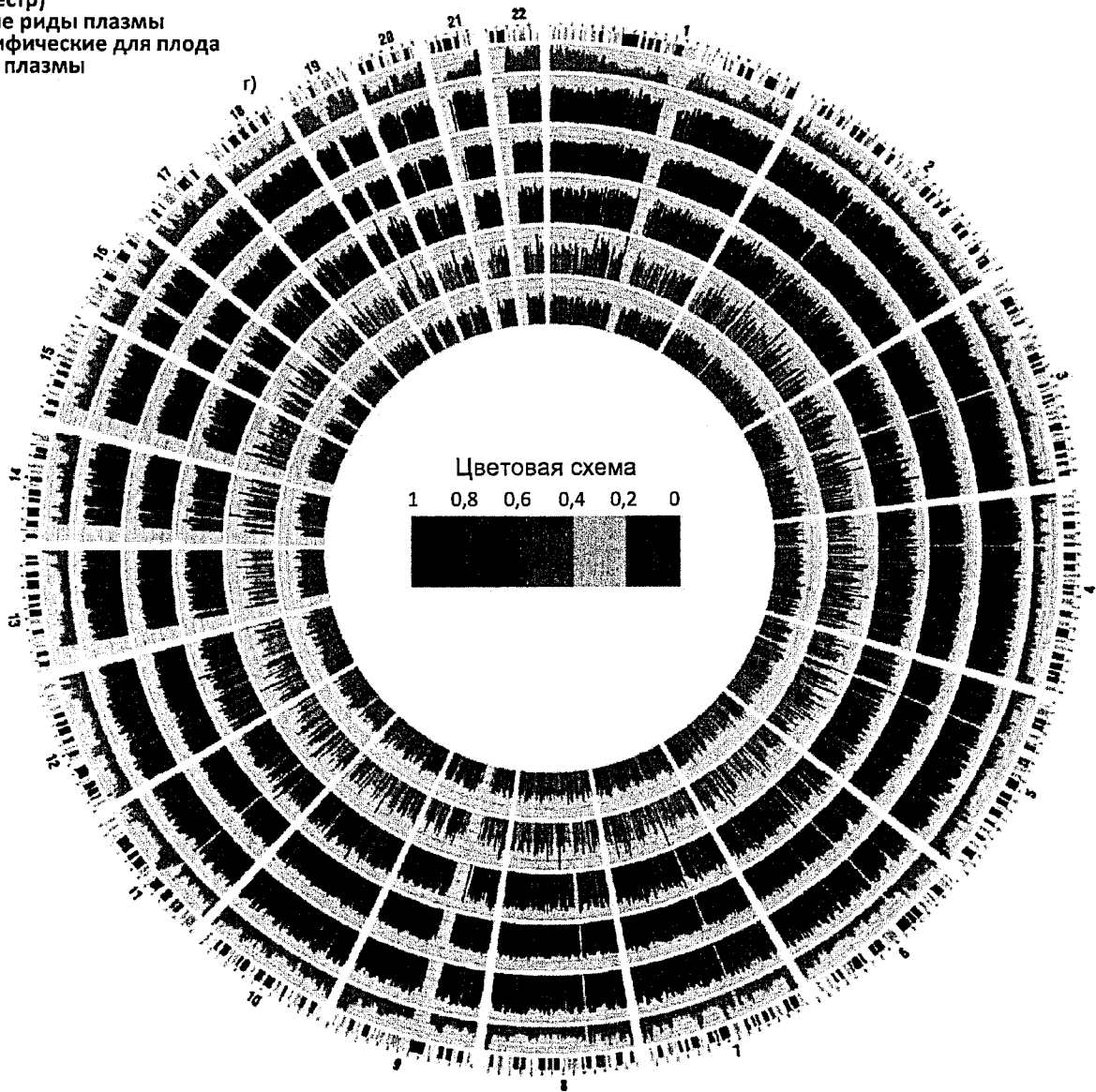


Фиг. 6

От внешнего круга
к внутреннему:

Плотность SrG
Клетки крови матери
Плазма матери (1-й
триместр)
Общие риды плазмы
Специфические для плода
риды плазмы
CVS

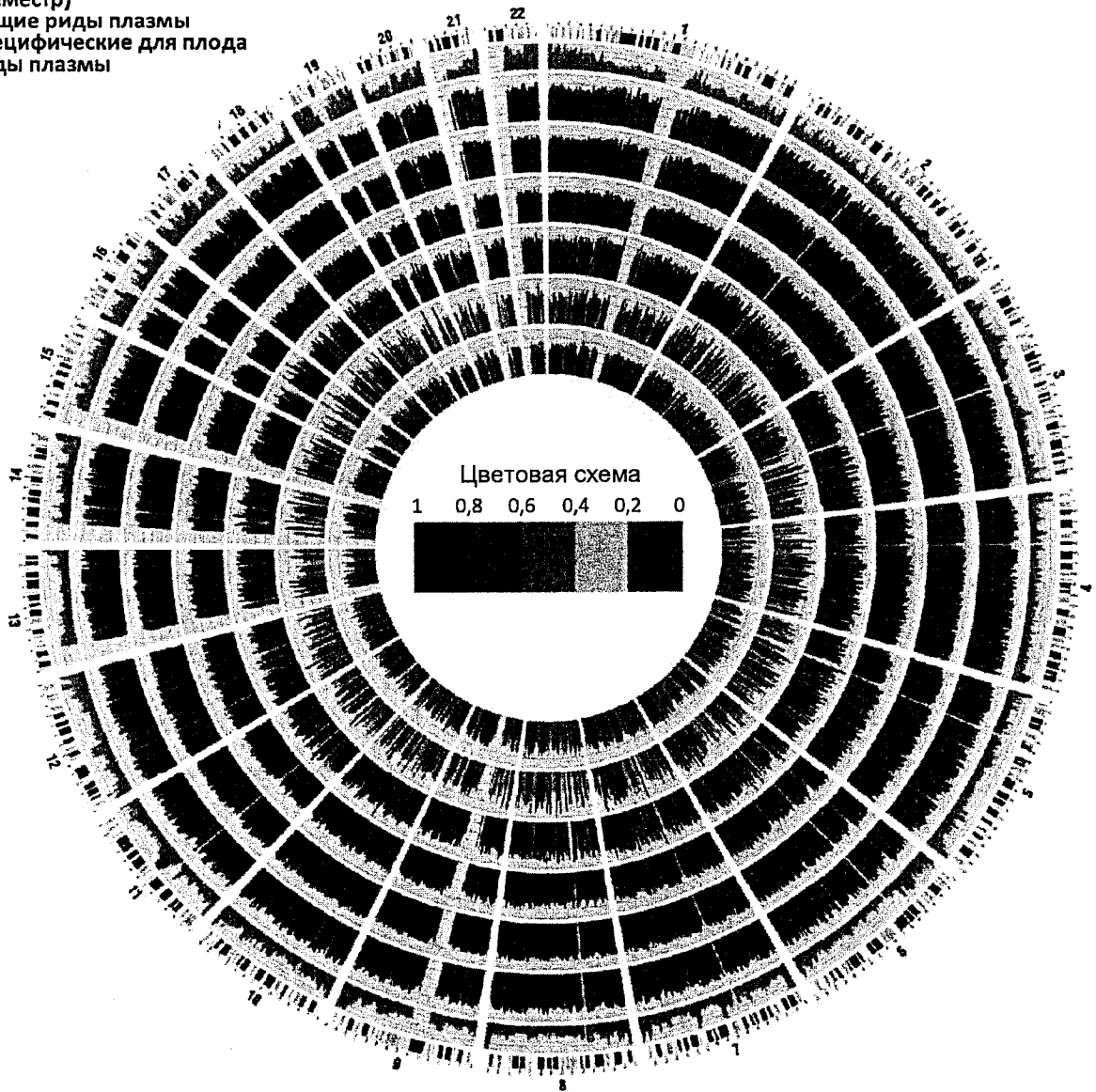
700



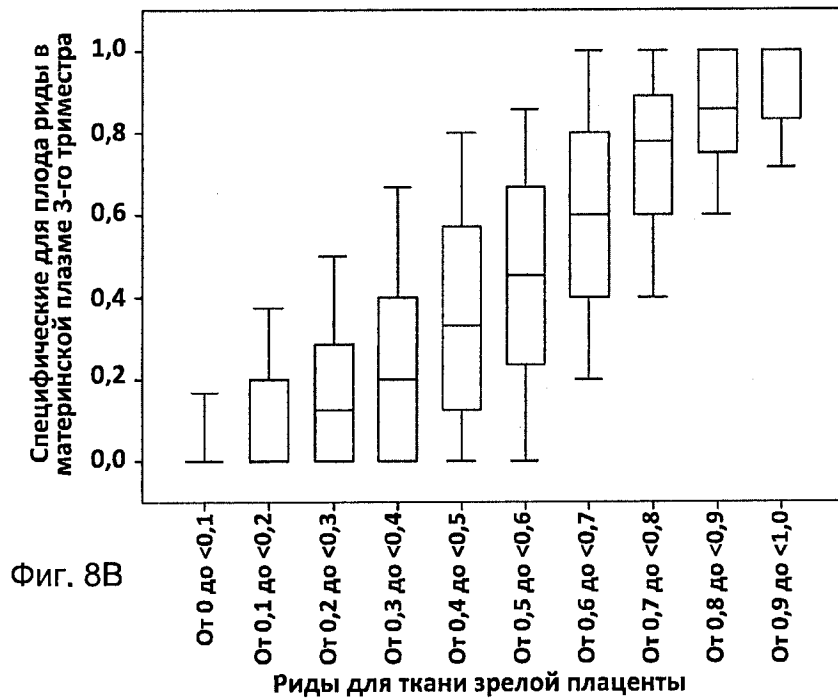
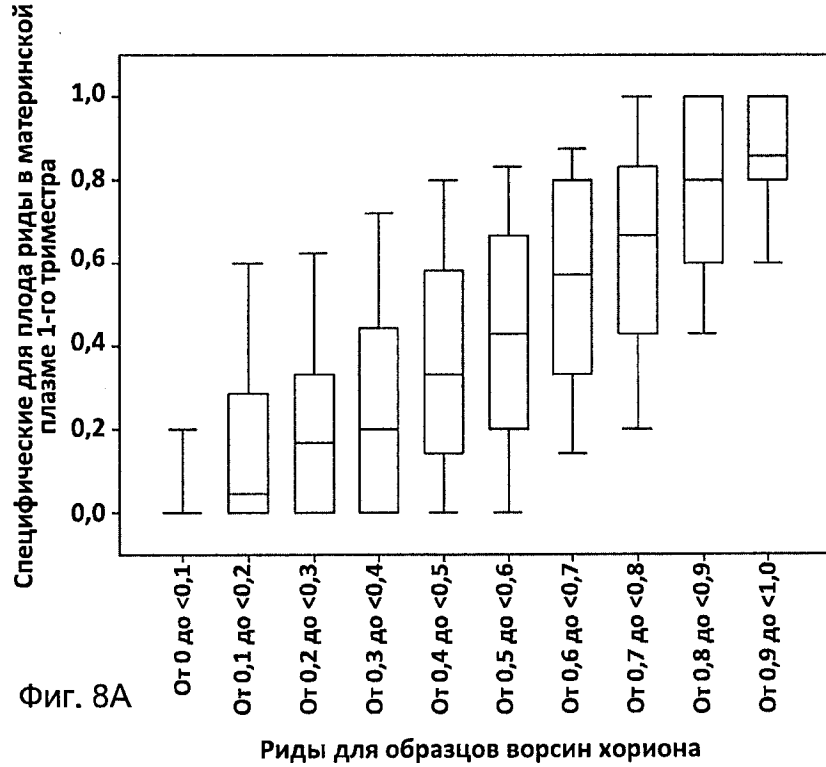
Фиг. 7А

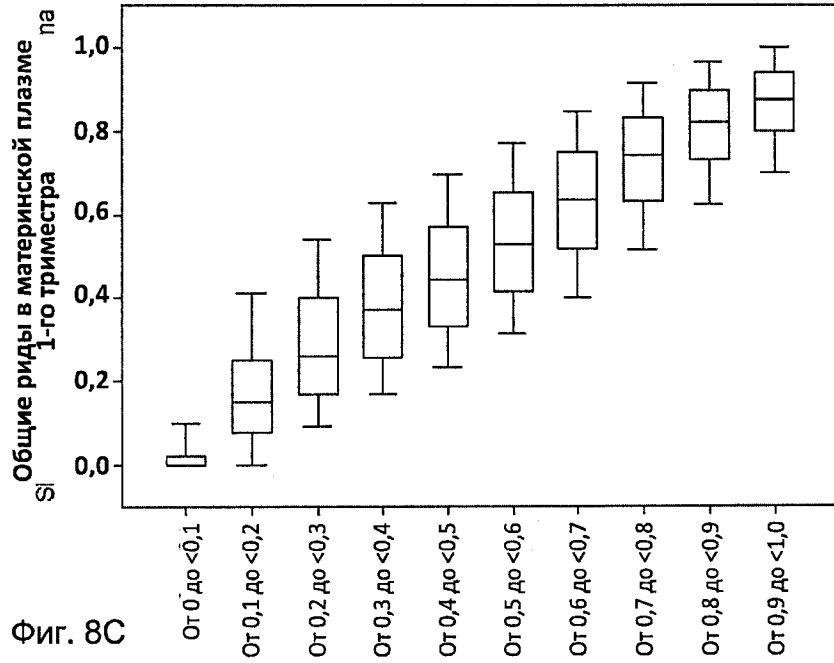
От внешнего круга
к внутреннему:
Плотность СрG
Клетки крови матери
Плазма после родов
Плазма матери (3-й
триместр)
Общие риды плазмы
Специфические для плода
риды плазмы

750



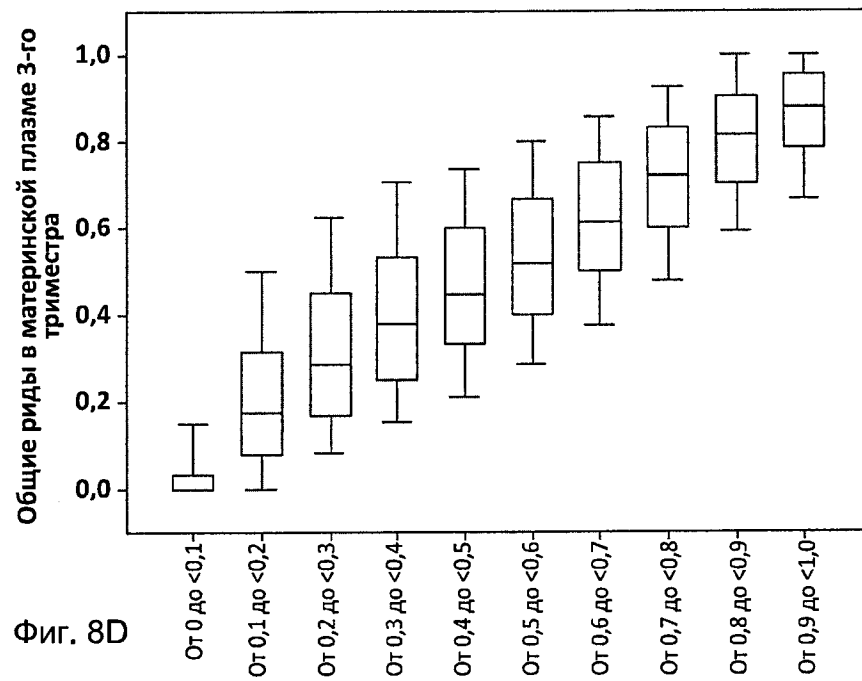
Фиг. 7В





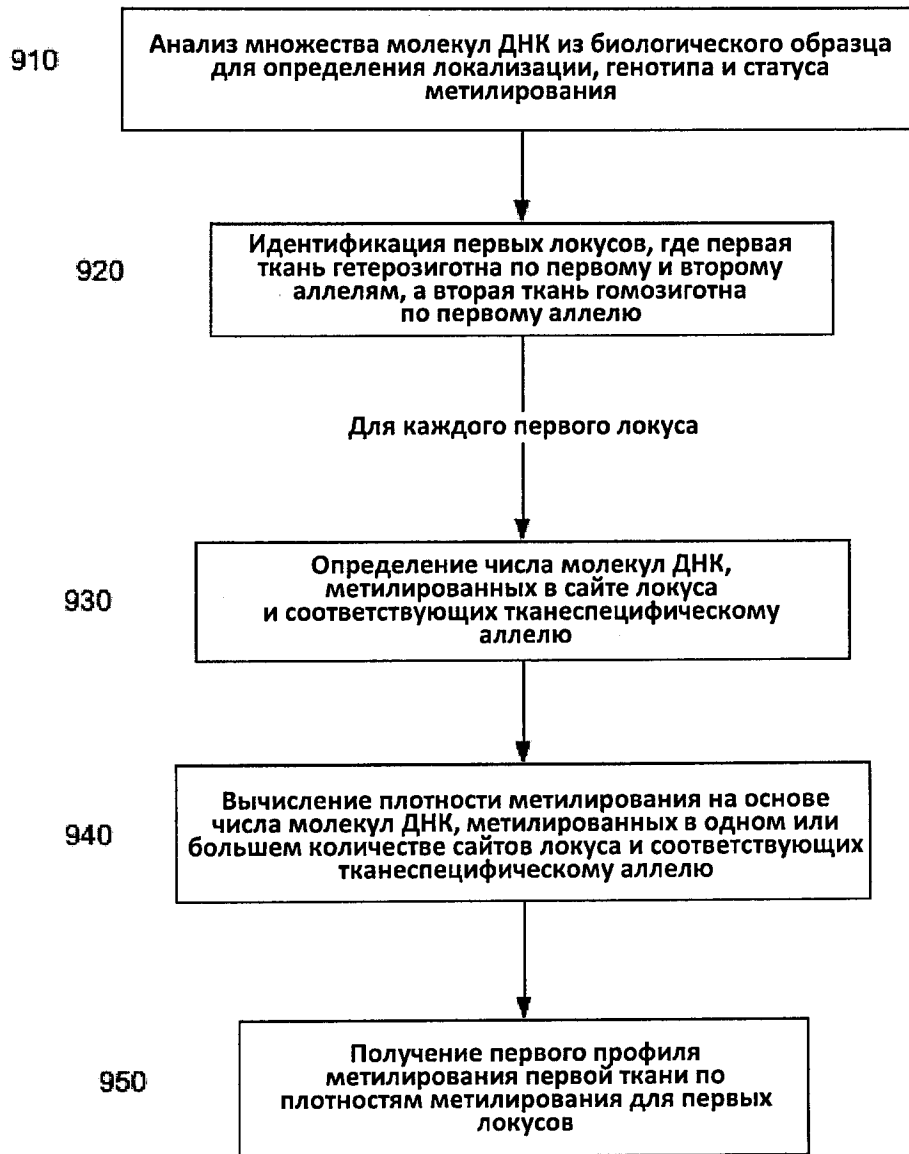
Фиг. 8C

Риды для клеток материнской крови

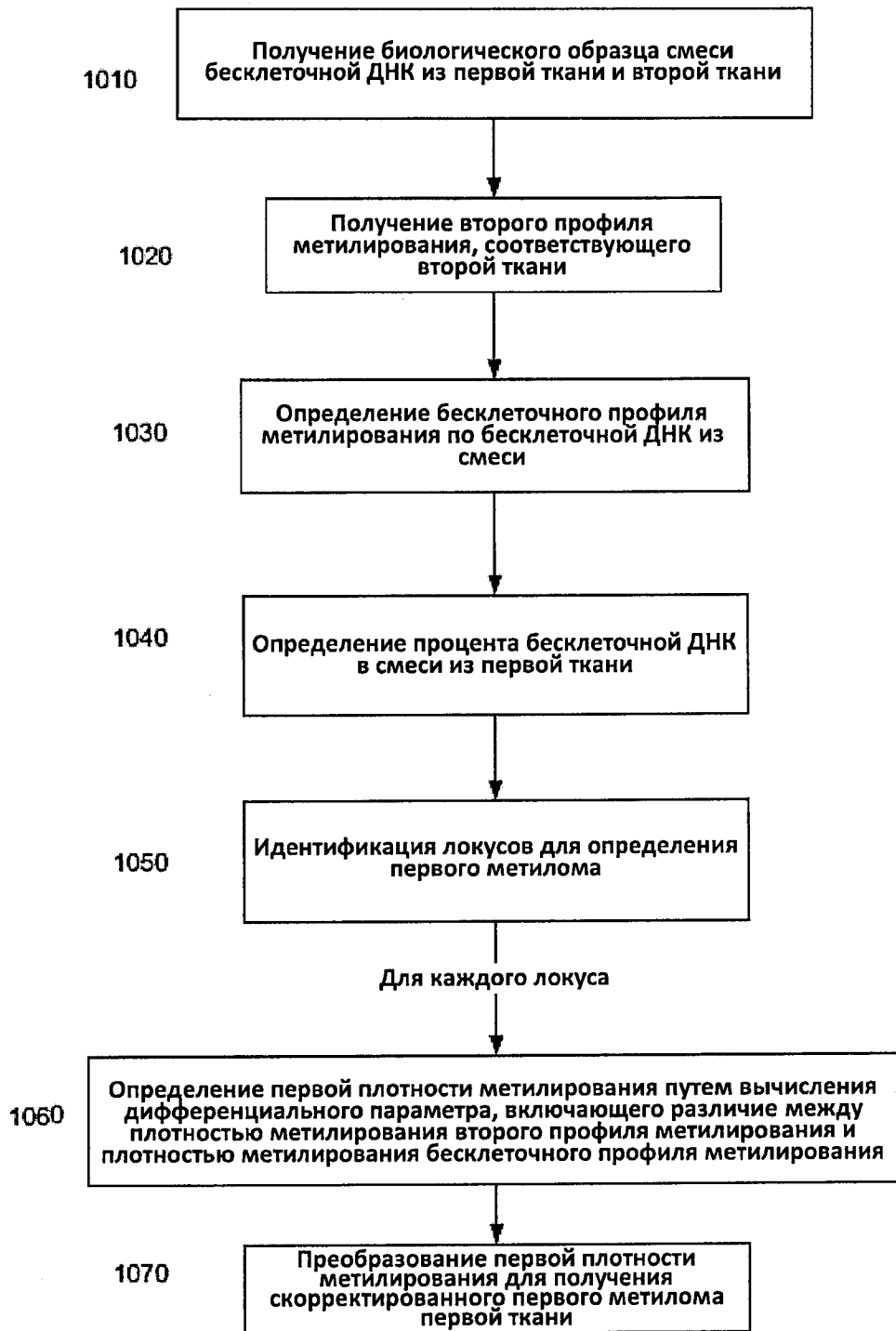


Фиг. 8D

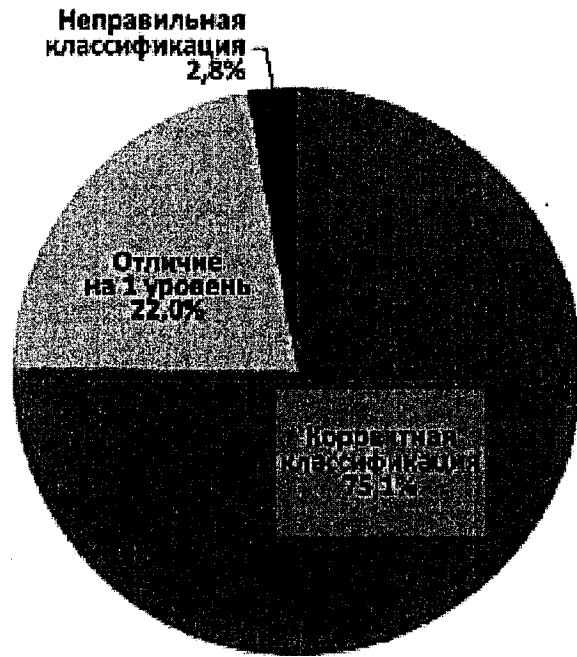
Риды для клеток материнской крови



Фиг. 9

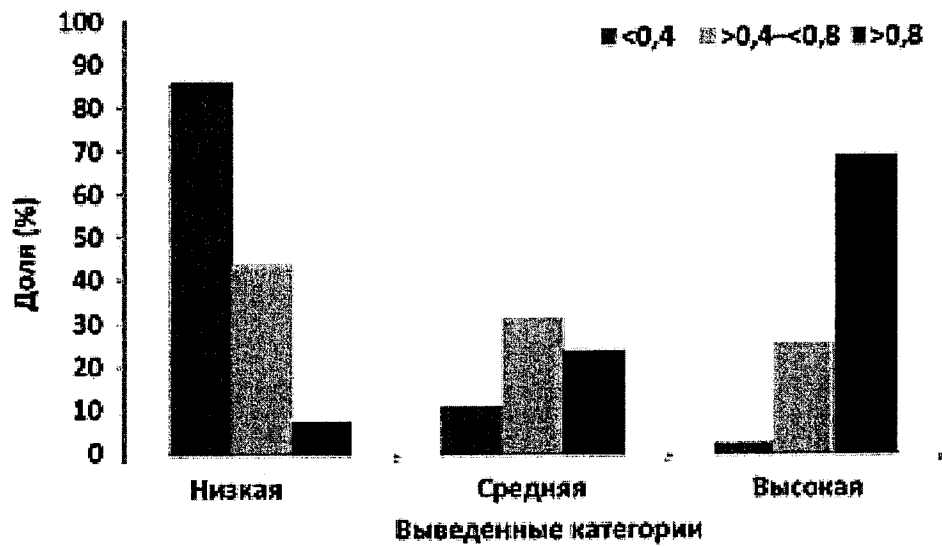


Фиг. 10



1100

Фиг. 11А



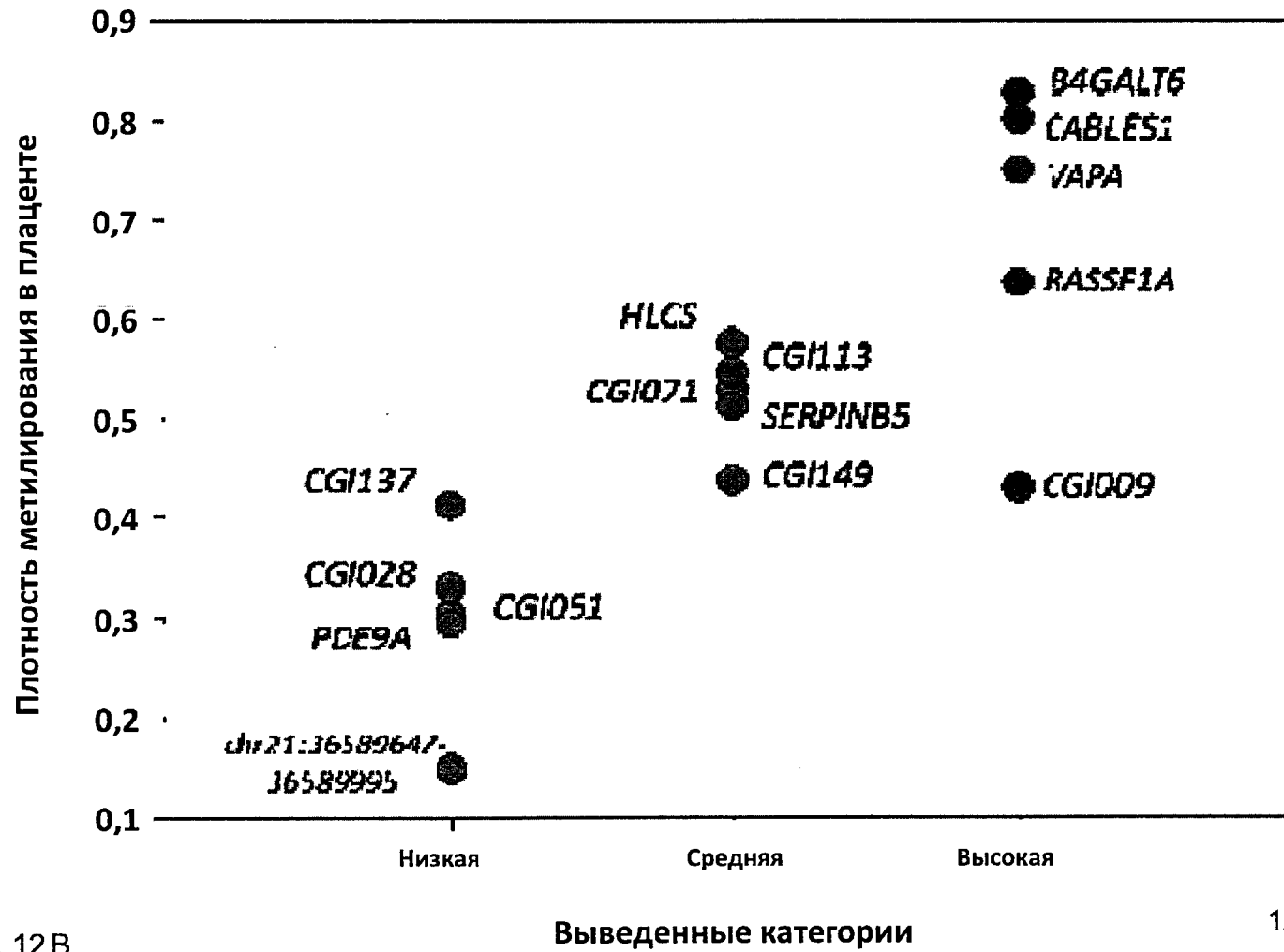
1150

Фиг. 11В

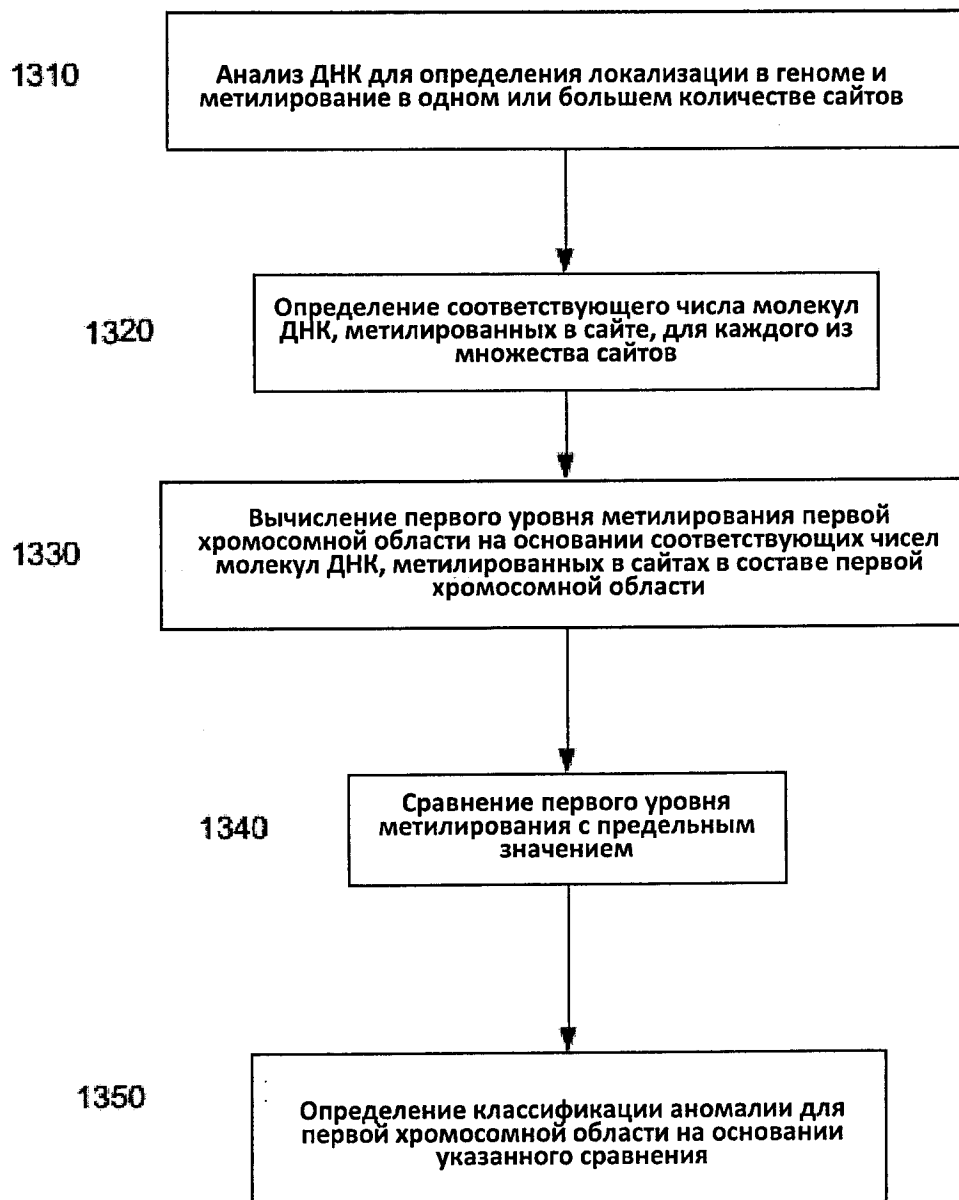
№	Генный локус	Геномная локализация	Плотность метилирования (данные секвенирования)			Выведенные значения (D) ¹	Выведение			Статус метилирования в публикациях	
			Матер. кровь (ВКГ)	Плацента	Матер. плазма (МР)		Прогнозируемый показатель метилирования (MPS) ²	Выведенные категории ³	Уровень метилирования плаценты ⁴	Статус плаценты ⁵	Ссылки
1	RASSF1A	chr3:50352938-50353234	0,000	0,637	0,067	0,506	1,060	Высокий	Средний	Гипермет.	(1)
2	CGI009	chr21:25856489-25856756	0,001	0,432	0,050	0,374	0,849	Высокий	Средний	Гипермет.	(2)
3	CGI028	chr21:33321942-33322182	0,035	0,331	0,040	0,073	0,367	Низкий	Низкий	Гипермет.	(2)
4	CGI051	chr21:36995232-36995509	0,043	0,303	0,046	0,072	0,365	Низкий	Низкий	Гипермет.	(2)
5	CGI071	chr21:41140466-41140898	0,031	0,531	0,046	0,146	0,483	Средний	Средний	Гипермет.	(2)
6	CGI113	chr21:44953820-44953972	0,013	0,547	0,051	0,306	0,739	Средний	Средний	Гипермет.	(2)
7	CGI137	chr21:46249872-46250196	0,940	0,413	0,820	0,030	0,299	Низкий	Средний	Гипомет.	(2)
8	CGI 149	chr21:46911575-46912533	0,004	0,440	0,030	0,202	0,573	Средний	Средний	Гипермет.	(2)
9	HLCS	chr21:37274945-37275031	0,038	0,576	0,059	0,203	0,575	Средний	Средний	Гипермет.	(2)
10	PDE9A	chr21:42979270-42979525	0,950	0,303	0,823	-0,018	0,222	Низкий	Низкий	Гипомет.	(2)
11	-	chr21:36589647-36589995	0,952	0,149	0,827	0,007	0,262	Низкий	Низкий	Гипомет.	Неопубл.
12	SERPINB5	chr18:59294739-59295038	0,936	0,530	0,856	0,329	0,777	Средний	Средний	Гипомет.	(3)
13	VAPA	chr18:10022563-10023186	0,036	0,752	0,082	0,382	0,861	Высокий	Средний	Гипермет.	(4)
14	CABLES1	chr18:19035954-19036069	0,335	0,801	0,491	1,519	2,680	Высокий	Высокий	Гипермет.	(4)
15	B4GALT6	chr18:27486040-27486932	0,113	0,827	0,331	1,764	3,072	Высокий	Высокий	Гипермет.	(4)

1200

Фиг. 12А

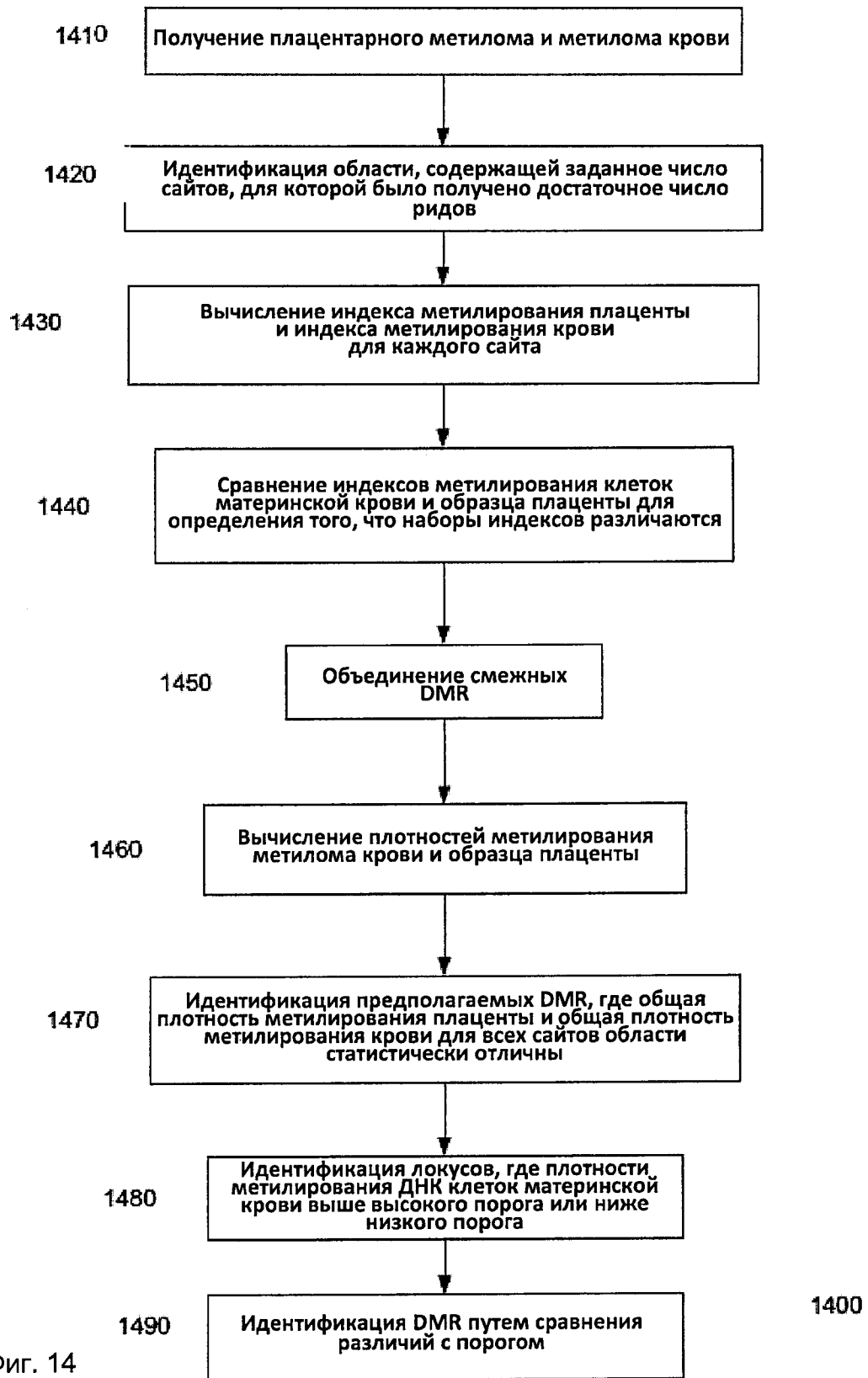


Фиг. 12 В



1300

Фиг. 13



Фиг. 14

№ ^a	Генный локус	Геномная локализация ^b	Размер ^b	Плацентарн. статус ^b	BS-Seq ^c			DMR ^d								
					Клетки крови матери	CVS	Разл. метил.	Хромо-сома	Начало	Конец	Размер	Клетки крови матери	CVS	P-значе-ние	Разл. метил.	Класс ^e
1	RASSF1A	chr3:50352938-50353234	297	Гиперметил.	0,00	63,73	63,73	chr3	50351000	50354000	3000	6,26	66,68	0,00E+00	60,42	TP
2	CGI009	chr21:25856489-25856756	268	Гиперметил.	0,08	43,20	43,13	chr21	25855500	25857500	2000	4,97	48,70	0,00E+00	43,73	TP
3	CGI028	chr21:33321942-33322182	241	Гиперметил.	3,51	33,10	29,59	chr21	33316400	33323900	7500	4,52	23,85	0,00E+00	19,33	FN
4	CGI051	chr21:36995232-36995509	278	Гиперметил.	4,26	30,33	26,07	chr21	36992800	37003900	11100	3,12	21,91	0,00E+00	18,79	FN
5	CGI071	chr21:41140466-41140898	433	Гиперметил.	3,09	53,06	49,97	chr21	41140000	41141000	1000	3,40	52,00	0,00E+00	48,60	TP
6	CGI113	chr21:44953820-44953972	153	Гиперметил.	1,27	54,69	53,42	chr21	44948500	44954700	6200	7,18	57,23	0,00E+00	50,05	TP
7	CGI137	chr21:46249872-46250196	325	Гипометил.	94,00	41,34	52,65	chr21	46227800	46250400	22600	89,74	52,67	0,00E+00	37,07	TP
8	CGI149	chr21:46911575-46912533	959	Гиперметил.	0,40	43,95	43,56	chr21	46911200	46912700	1500	5,31	46,17	0,00E+00	40,86	TP
9	HLCS	chr21:37274945-37275031	87	Гиперметил.	3,77	57,61	53,85	chr21	37274200	37275200	1000	7,33	54,24	0,00E+00	46,91	TP
10	PDE9A	chr21:42979270-42979525	256	Гипометил.	95,01	30,33	64,68	chr21	42978100	43004800	26700	89,97	51,44	0,00E+00	38,53	TP
11	-	chr21:36589647-36589995	349	Гипометил.	95,17	14,89	80,28	chr21	36589200	36590200	1000	94,49	19,27	0,00E+00	75,22	TP
12	SERPINB5	chr18:59294739-59295038	300	Гипометил.	93,64	53,00	40,64	chr18	59294800	59296300	1500	88,47	40,76	3,33E-27	47,71	TP
13	VAPA	chr18:10022563-10023187	625	Гиперметил.	3,63	75,17	71,54	chr18	10021800	10023300	1500	24,08	80,33	0,00E+00	56,25	TP
14	CABLES1	chr18:19035954-19036070	117	Гиперметил.	33,47	80,09	46,62	-	-	-	-	-	-	-	FN	
15	B4GALT6	chr18:27486040-27486933	894	Гиперметил.	11,34	82,75	71,41	chr18	27486100	27487100	1000	11,34	82,75	9,6408E-86	71,41	TP
16	CASP8	chr2:201830342-201831885	1544	Гиперметил.	0,52	47,54	47,03	chr2	201830200	201832200	2000	3,77	49,13	4,22E-42	45,36	TP
17	RARB	chr3:25444840-25445090	251	Гиперметил.	2,28	26,93	24,65	chr3	25443900	25445400	1500	3,44	22,97	9,65E-50	19,53	FN
18	APC	chr5:112100999-112101943	945	Гиперметил.	0,71	45,78	45,06	chr5	112100300	112101800	1500	3,59	48,09	0,00E+00	44,50	TP
19	DAB2IP_ori	chr9:121541221-121541460	240	Гипометил.	84,87	52,63	32,24	chr9	121540400	121541500	1100	85,21	36,86	5,2605E-10	48,35	TP
20	DAB2IP_ex	chr9:123128261-123129077	817	Гипометил.	95,09	32,99	62,10	chr9	123127700	123129200	1500	90,72	50,61	1,4706E-19	40,11	TP
21	THY1	chr11:118798952-118799382	431	Гиперметил.	0,88	39,32	38,44	chr11	118797600	118799600	2000	2,25	27,28	0,00E+00	25,03	TP
22	PTPN6	chr12:6930211-6933772	3562	Гиперметил.	1,52	79,80	78,29	chr12	6930000	6934000	4000	4,26	80,13	0,00E+00	75,87	TP
23	PYCARD_ex	chr16:31121808-31122146	339	Гиперметил.	4,90	46,51	41,61	-	-	-	-	-	-	-	FN	
24	chr18A	chr18:55090284-55090606	323	Гиперметил.	6,14	31,79	25,65	chr18	55089700	55094200	4500	4,78	22,91	0,00E+00	18,13	FN
25	chr21A	chr21:39279723-39280004	282	Гиперметил.	12,80	64,02	51,21	chr21	39279300	39279800	500	4,74	62,07	7,17E-62	57,33	TP
26	chr21B	chr21:44161027-44161371	345	Гиперметил.	0,86	83,90	83,04	chr21	44160700	44161700	1000	11,06	77,32	0,00E+00	66,26	TP
27	chr21C	chr21:33320544-33320829	286	Гиперметил.	5,49	37,92	32,43	chr21	33316400	33323900	7500	4,52	23,85	0,00E+00	19,33	FN
28	chr21D	chr21:42189223-42189683	461	Гиперметил.	13,02	85,84	72,82	chr21	42188700	42189700	1000	7,78	81,67	3,62E-77	73,89	TP
29	chr21E1	chr21:42355712-42355908	197	Гиперметил.	0,00	57,85	57,85	chr21	42355000	42357500	2500	5,21	62,23	0,00E+00	57,02	TP
30	chr21E11	chr21:42357215-42357341	127	Гиперметил.	3,57	48,25	44,68	chr21	42355000	42357500	2500	5,21	62,23	0,00E+00	57,02	TP
31	chr21H	chr21:32268803-32268943	141	Гиперметил.	0,49	78,84	78,35	chr21	32268300	32269300	1000	20,49	85,25	4,9563E-34	64,76	TP
32	chr21I	chr21:44079235-44079535	301	Гиперметил.	1,23	67,44	66,21	chr21	44079400	44080400	1000	15,19	68,06	4,9E-59	52,87	TP
33	SERPINB5_P	chr18:59294811-	465	Гипометил.	91,96	39,71	52,25	chr18	59294800	59296300	1500	88,47	40,76	3,3294E-	47,71	TP

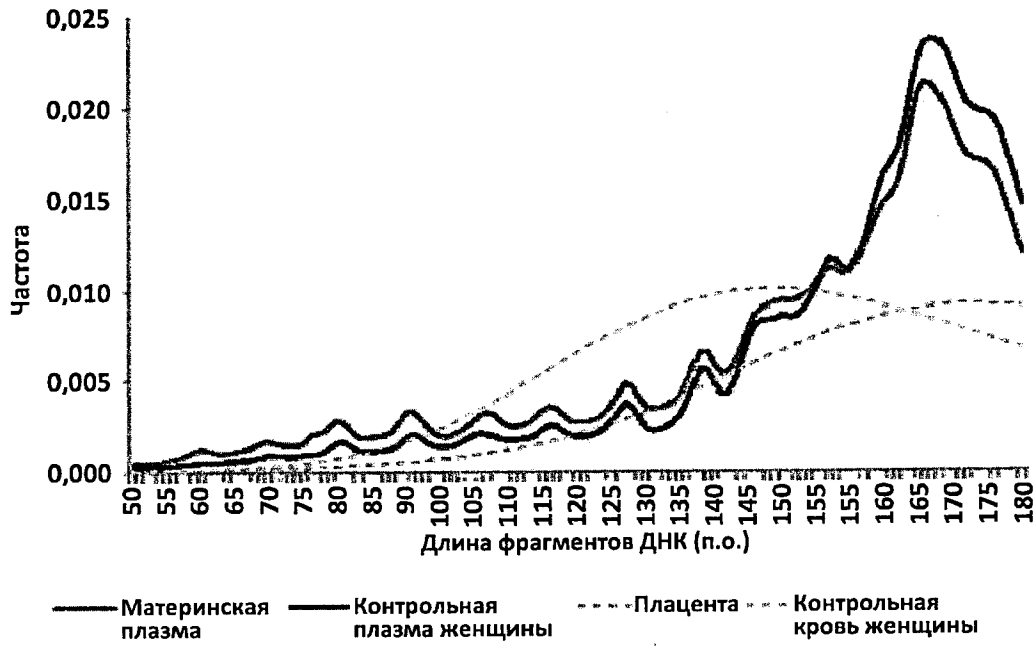
№ ^a	Генный локус	BS-Seq			DMR ^c								
		Клетки крови матери	Плацента	Различие метилир-я ^b	Хромосома	Начало	Конец	Размер	Клетки крови матери	Плацента	p-значение	Различие метилир-я	Класс ^d
1	RASSF1A	0,00	60,84	60,84	chr3	50352100	50354100	2000	2,16	62,21	0,00E+00	60,05	TP
2	CGI009	0,08	20,50	20,43	chr21	25855600	25857100	1500	2,04	24,37	0,00E+00	22,33	TP
3	CGI028	3,51	17,38	13,86	-	-	-	-	-	-	-	-	TN
4	CGI051	4,26	13,93	9,67	-	-	-	-	-	-	-	-	TN
5	CGI071	3,09	24,85	21,76	chr21	41139900	41141400	1500	6,83	29,28	0,00E+00	22,45	TP
6	CGI113	1,27	40,24	38,97	chr21	44948500	44954700	6200	7,18	45,00	0,00E+00	37,82	TP
7	CGI137	94,00	37,10	56,90	chr21	46248400	46250400	2000	92,28	43,85	1,66E-76	48,43	TP
8	CGI149	0,40	22,19	21,79	chr21	46911200	46912700	1500	5,31	25,73	0,00E+00	20,42	TP
9	HLCS	3,77	58,49	54,72	chr21	37274200	37275200	1000	7,33	56,83	0,00E+00	49,50	TP
10	PDE9A	95,01	32,41	62,60	chr21	42978900	42990100	11200	90,89	44,38	0,00E+00	46,51	TP
11	-	95,17	19,92	75,24	chr21	36589200	36590200	1000	94,49	22,66	0,00E+00	71,83	TP
12	SERPINB5	93,64	28,06	65,57	chr18	59294500	59295500	1000	92,11	36,53	3,85E-31	55,58	TP
13	VAPA	3,63	73,88	70,25	chr18	10022100	10023100	1000	11,67	75,76	0,00E+00	64,09	TP
14	CABLES1	33,47	81,40	47,93	-	-	-	-	-	-	-	-	FN
15	B4GALT6	11,34	88,57	77,23	chr18	27485800	27486800	1000	28,97	90,00	4,3779E-59	61,03	TP
16	CASP8	0,52	56,27	55,75	chr2	201830200	201832200	2000	3,77	57,60	1,33E-52	53,83	TP
17	RARB	2,28	38,18	35,90	chr3	25444000	25445000	1000	2,36	24,42	8,00E-54	22,06	TP
18	APC	0,71	34,89	34,18	chr5	112100700	112101700	1000	0,38	32,97	0,00E+00	32,59	TP
19	DAB2IP_ori	84,87	40,28	44,60	chr9	121540400	121541400	1000	84,79	30,63	1,7338E-14	54,16	TP
20	DAB2IP_ex	95,09	16,14	78,95	chr9	123128600	123129100	500	95,14	17,42	9,6247E-47	77,72	TP
21	THY1	0,88	22,28	21,40	chr11	118797800	118799300	1500	1,22	20,50	0,00E+00	19,28	FN
22	PTPN6	1,52	81,33	79,82	chr12	6930000	6934000	4000	4,26	81,58	0,00E+00	77,32	TP
23	PYCARD_ex	4,90	48,73	43,83	chr16	31121700	31122200	500	4,85	32,87	1,1158E-34	28,02	TP
24	chr18A	6,14	16,25	10,10	-	-	-	-	-	-	-	-	TN
25	chr21A	12,80	74,65	61,85	chr21	39278300	39280300	2000	9,96	75,07	0,00E+00	65,11	TP
26	chr21B	0,86	81,56	80,70	chr21	44160700	44161700	1000	11,06	78,6	0,00E+00	67,54	TP
27	chr21C	5,49	14,91	9,42	-	-	-	-	-	-	-	-	TN
28	chr21D	13,02	88,71	75,70	chr21	42188700	42189700	1000	7,78	84,42	3,5806E-77	76,64	TP
29	chr21E1	0,00	73,21	73,21	chr21	42355300	42357800	2500	4,66	79,19	0,00E+00	74,53	TP
30	chr21E11	3,57	84,00	80,43	chr21	42355300	42357800	2500	4,66	79,19	0,00E+00	74,53	TP
31	chr21H	0,49	77,49	77,00	chr21	32268400	32268900	500	18,4	83,85	2,1972E-30	65,45	TP
32	chr21I	1,23	58,62	57,39	chr21	44079300	44080300	1000	12,16	72,55	2,3559E-60	60,39	TP
33	SERPINB5_P	91,96	28,61	63,35	chr18	59294500	59295500	1000	92,11	36,53	3,85E-31	55,58	TP

Фиг. 15В

	Материнская плазма первого триместра		Материнская плазма третьего триместра	
	Гипермет. ^a	Гипомет. ^b	Гипермет. ^a	Гипомет. ^b
Число прогнозируемых локусов	3081	44455	1746	14930
Число локусов с плотностью метилирования > 40% по данным для плацентарной ткани ^c	1,678	Н/О	1525	Н/О
Число локусов с плотностью метилирования < 60% по данным для плацентарной ткани ^c	Н/О	23468	Н/О	13475
Число локусов, перекрывающихся с DMR, установленное по данным для плацентарной ткани ^c и материнской плазмы	1457	21812	1279	12677

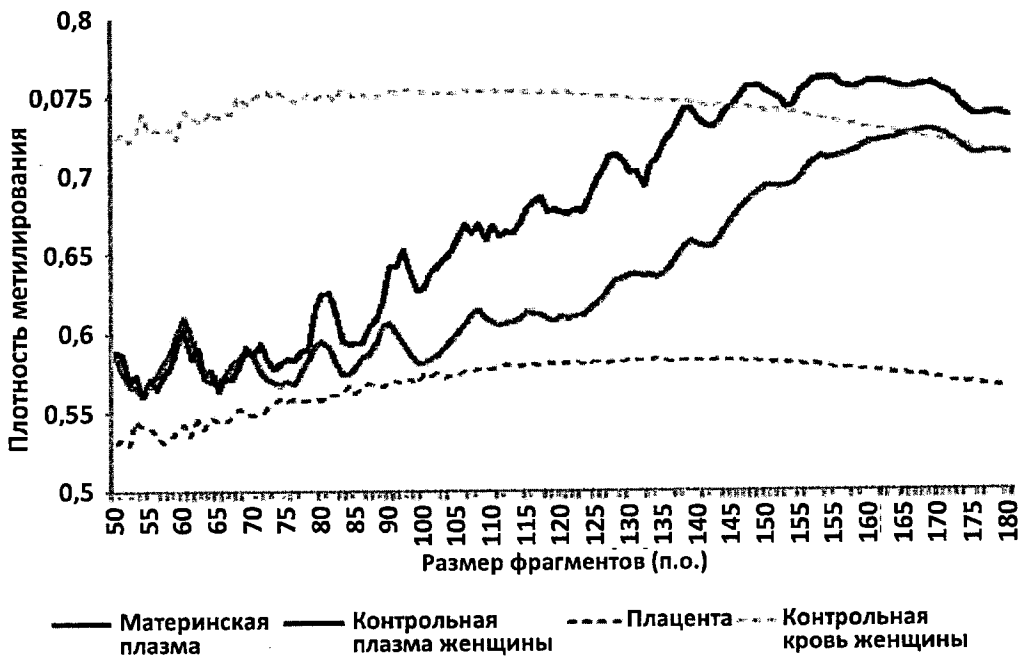
Фиг. 16

1600



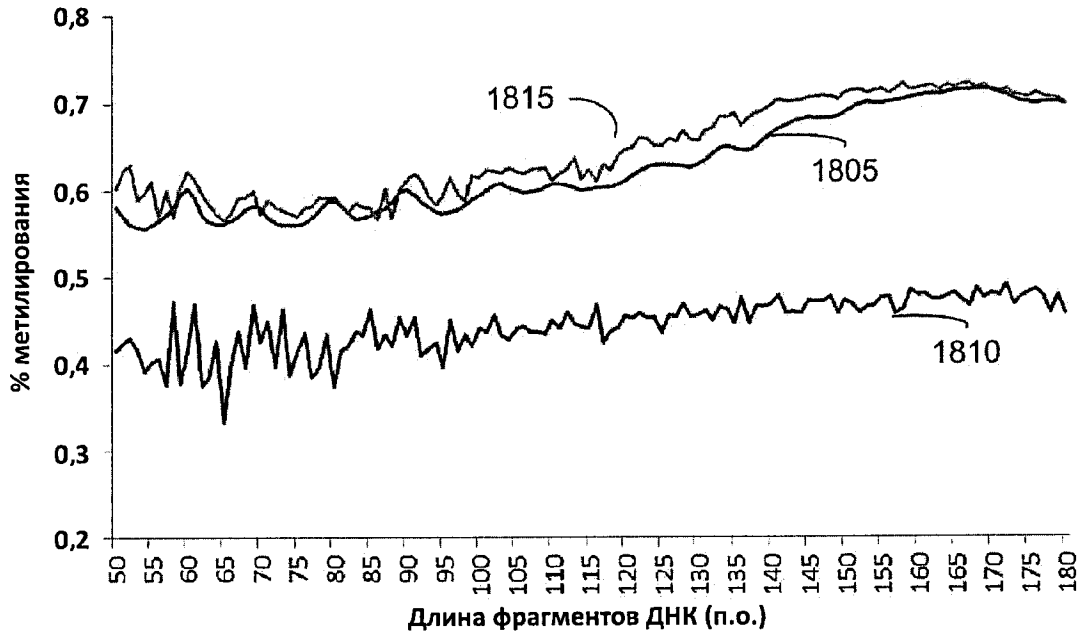
1700

Фиг. 17А



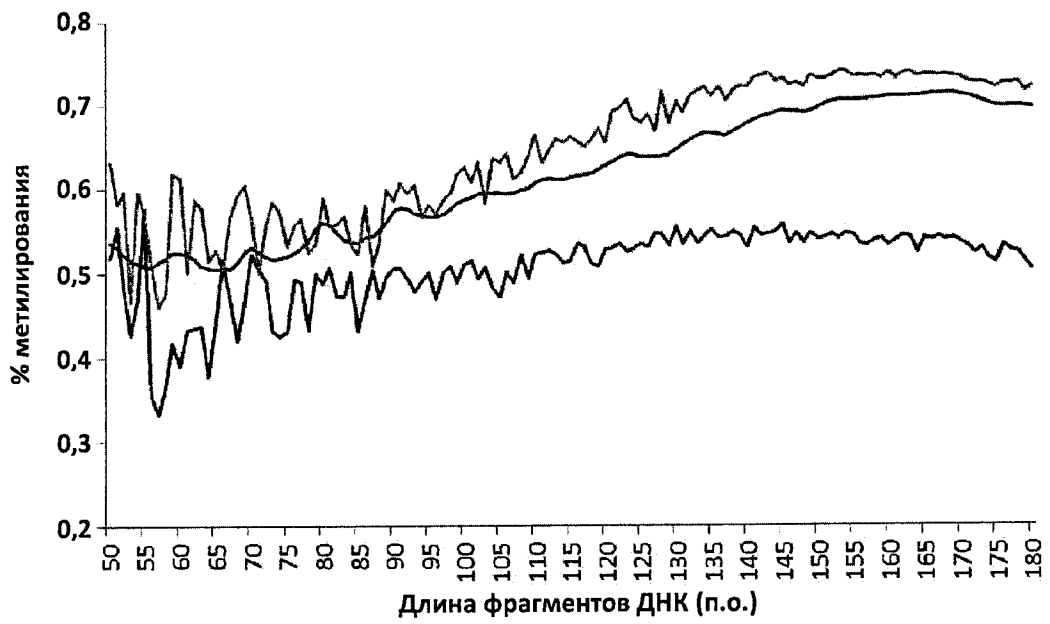
1750

Фиг. 17В



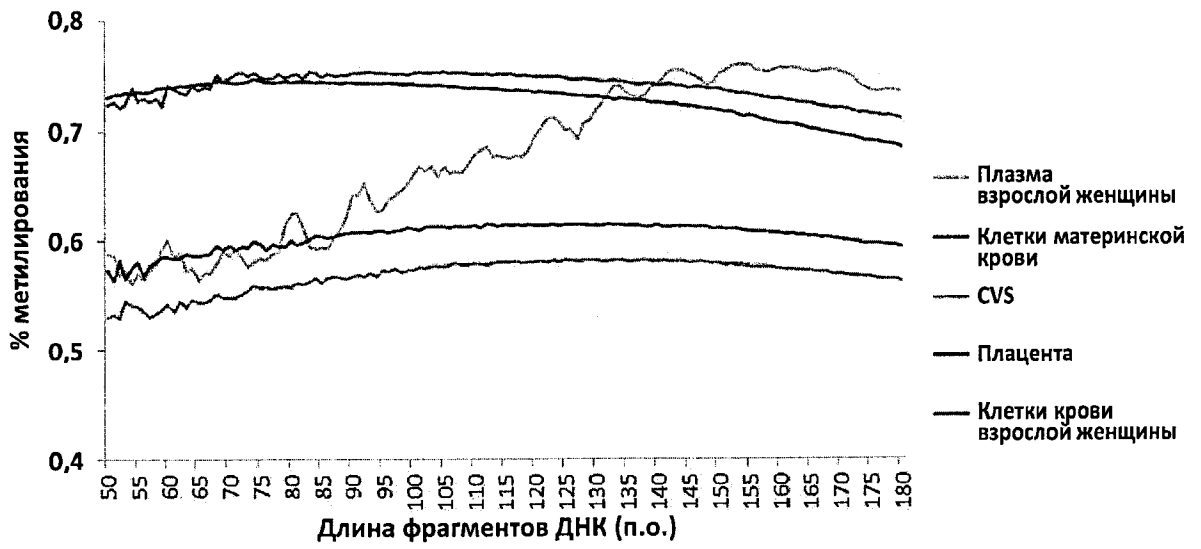
Фиг. 18А

1800



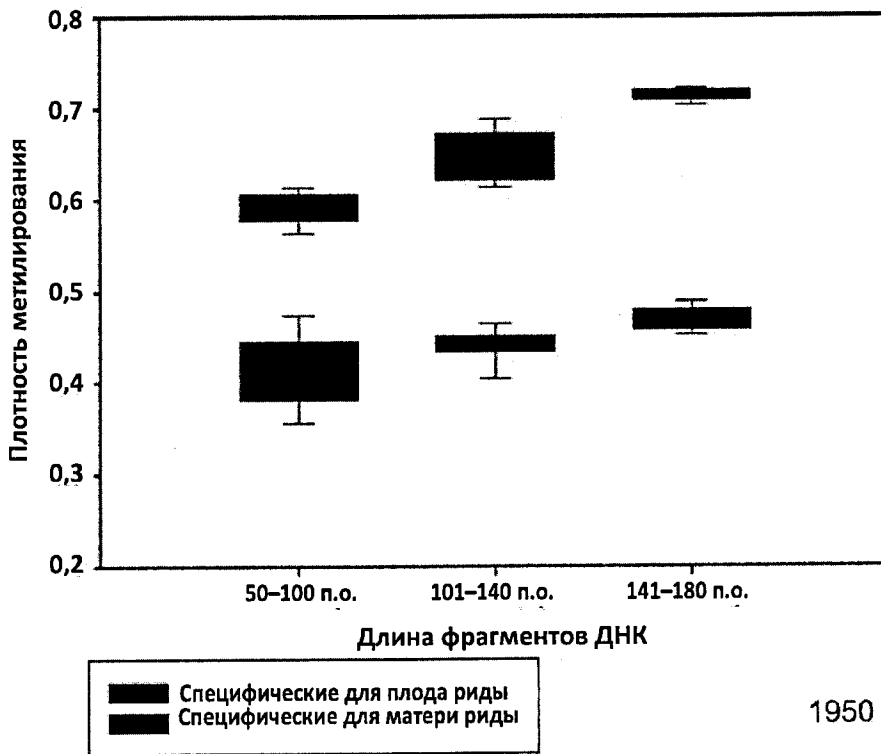
Фиг. 18В

1850



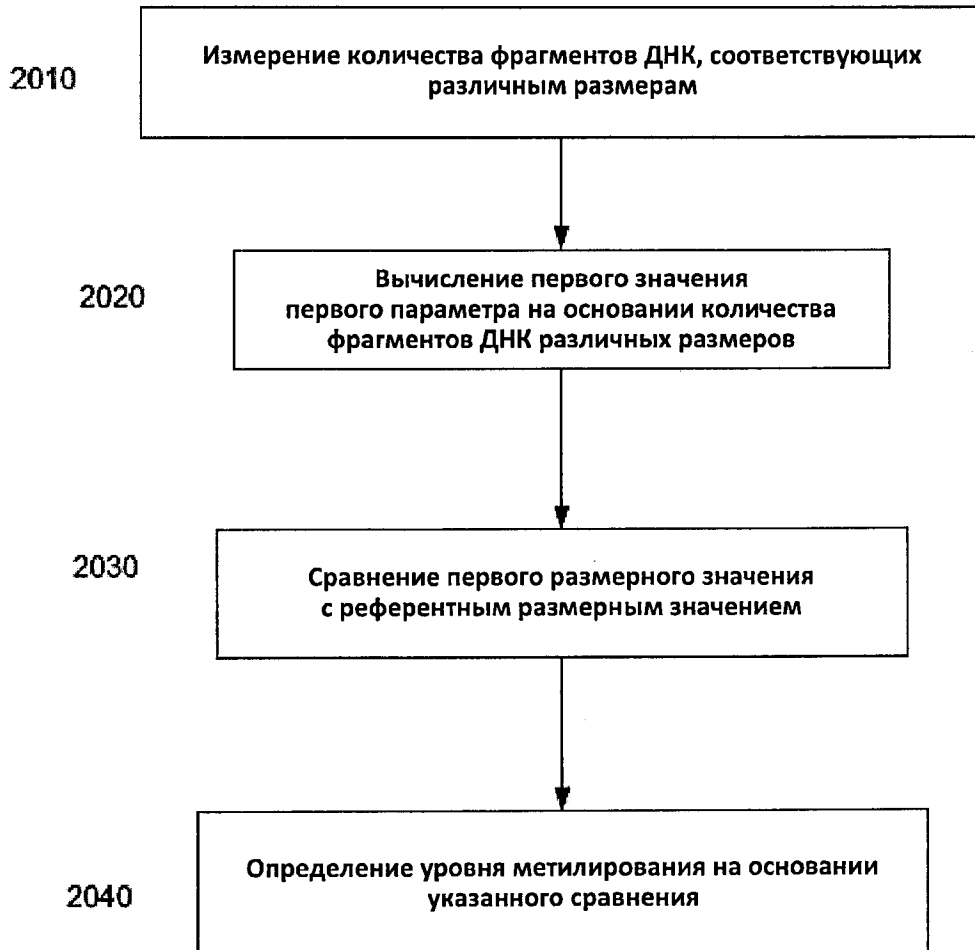
Фиг. 19А

1900



1950

Фиг. 19В



Фиг. 20

Хромосома	Пациент с ГЦК			
	Лейкоцитарная пленка	Неопухолевая ткань печени	Опухолевая ткань	Предоперационная плазма
1	65,4%	65,7%	44,1%	60,2%
2	68,3%	68,2%	42,0%	58,6%
3	68,2%	68,5%	44,6%	60,9%
4	68,9%	67,7%	37,8%	56,9%
5	67,4%	66,7%	37,8%	56,9%
6	67,4%	67,7%	43,2%	59,9%
7	69,1%	68,8%	43,0%	60,0%
8	69,1%	68,0%	40,4%	60,0%
9	66,4%	66,2%	44,9%	60,6%
10	68,9%	68,2%	40,8%	58,5%
11	65,5%	65,6%	41,6%	58,5%
12	67,6%	67,4%	45,9%	62,0%
13	69,8%	68,8%	40,6%	60,2%
14	66,4%	66,2%	41,4%	59,2%
15	66,1%	66,8%	44,6%	61,0%
16	68,9%	67,6%	45,6%	60,7%
17	64,0%	64,9%	45,8%	60,2%
18	69,9%	69,1%	42,9%	60,0%
19	62,3%	62,7%	47,5%	61,0%
20	66,0%	65,3%	39,1%	57,8%
21	61,8%	61,1%	39,9%	57,7%
22	67,9%	67,6%	49,1%	63,8%
Все аутосомы	67,1%	66,9%	42,9%	59,7%

Фиг. 21А

2100

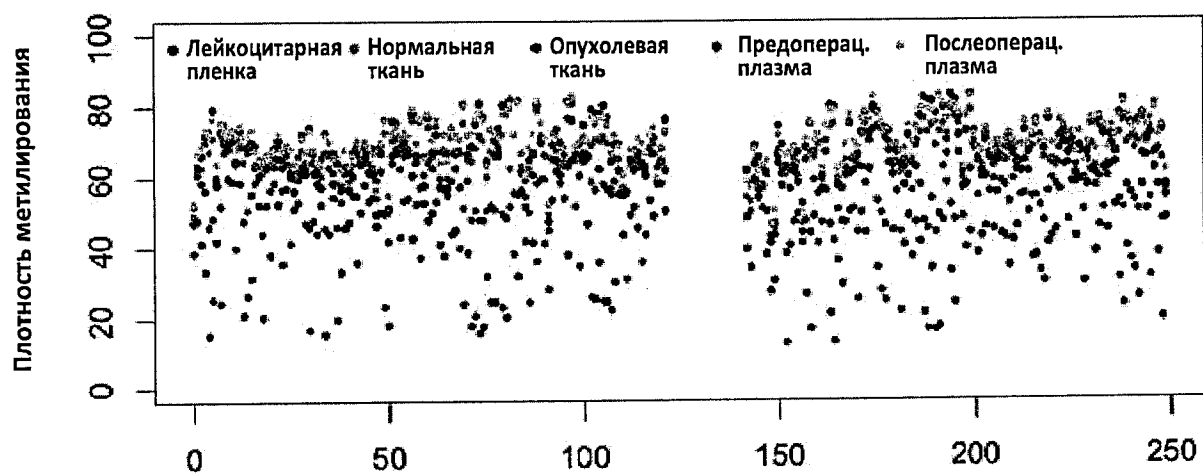
Перечень образцов	Число необработанных ридов	Число картированных ридов	Недублируемые фрагменты (Только аутосомы)	Доля подходящих для использования ридов %	Глубина на цепь	Покрываемость CpG %	Глубина покрытия ридами на сайтах CpG на цепь
Контроль 1	159554637	56323206	41653660	26%	1,12	54,86%	3,56
Контроль 2	118056735	49232079	45741385	39%	1,23	60,26%	3,24
Контроль 3	181913746	75753907	68081010	37%	1,82	65,74%	4,37
Контроль 4	142782870	50195168	42370350	30%	1,13	59,15%	3,14
С06	262763161	197344994	88266088	33%	2,11	64,31%	3,65
Лейкоцитарная пленка	579329822	439663895	375814248	65%	10,07	78,49%	21,67
Неопухолевая ткань печени	549773334	417736311	355154964	65%	9,51	84,89%	18,21
Опухолевая ткань	568004640	487494291	176985129	31%	4,74	83,48%	8,32
Предоперац. плазма	796183998	637735722	474751039	60%	12,72	87,18%	18,69
Послеоперац. плазма	921833848	624851729	550104801	60%	14,73	89,94%	24,81

Фиг. 21В

Хромосома	Плотность метилирования ДНК плазмы					
	Пациент с ГЦК		Контроль 1	Контроль 2	Контроль 3	Контроль 4
	Предоперац.	Послеоперац.				
1	60,2%	69,6%	70,0%	70,3%	70,1%	71,3%
2	58,6%	71,9%	72,2%	72,4%	72,3%	73,5%
5	60,9%	72,3%	72,7%	73,1%	72,9%	73,9%
4	56,9%	71,8%	71,9%	72,2%	72,2%	73,4%
5	56,9%	70,9%	71,0%	71,3%	71,3%	72,6%
6	59,9%	71,2%	71,6%	71,9%	71,7%	72,8%
7	60,0%	72,2%	72,8%	72,8%	72,5%	73,9%
8	60,0%	72,0%	72,6%	72,6%	72,4%	73,7%
9	60,6%	70,4%	71,0%	71,1%	71,0%	72,2%
10	58,5%	71,7%	72,0%	72,1%	72,0%	73,2%
11	58,5%	69,6%	70,2%	70,4%	70,2%	71,6%
12	62,0%	71,0%	71,6%	71,7%	71,5%	72,7%
13	60,2%	72,8%	73,2%	73,4%	73,2%	74,3%
14	59,2%	70,2%	70,7%	70,9%	70,8%	72,0%
15	61,0%	70,5%	70,6%	71,0%	70,9%	72,0%
16	60,7%	71,3%	72,4%	71,8%	71,6%	73,3%
17	60,2%	68,6%	69,3%	69,4%	69,3%	70,8%
18	60,0%	72,7%	72,9%	73,1%	73,0%	74,2%
19	61,0%	66,9%	68,1%	68,1%	67,6%	69,3%
20	57,8%	69,6%	70,3%	70,3%	70,3%	71,6%
21	57,7%	67,0%	64,0%	64,5° о	65,7%	67,4%
22	63,8%	71,4%	72,5%	72,0%	71,9%	73,3%
Все аутосомы	59,7%	70,8%	71,2%	71,3%	71,2%	72,5%

Фиг. 22

2200



Фиг. 23А

Хромосома 1 (Мб)

2300

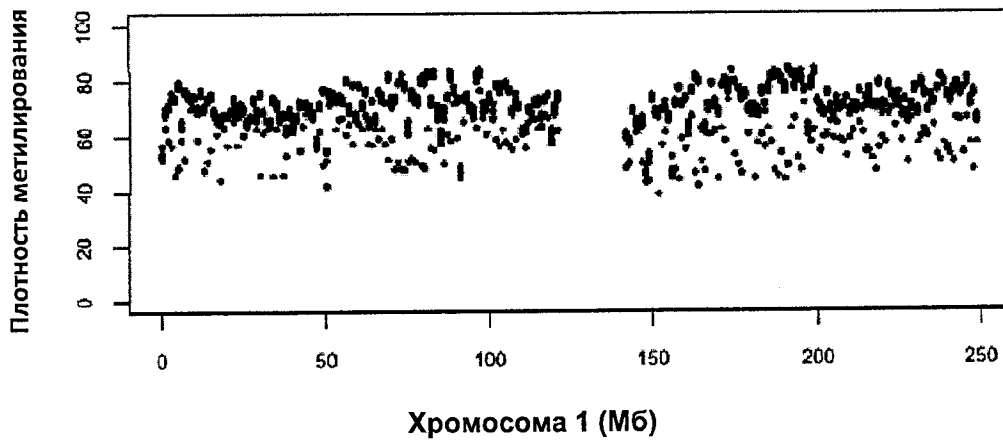


Фиг. 23В

Хромосома 2 (Мб)

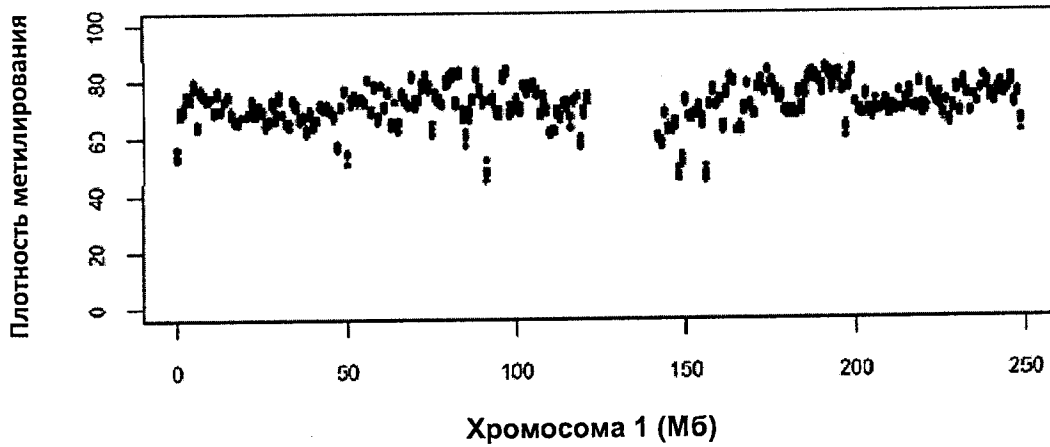
2350

31/68



Фиг. 24А

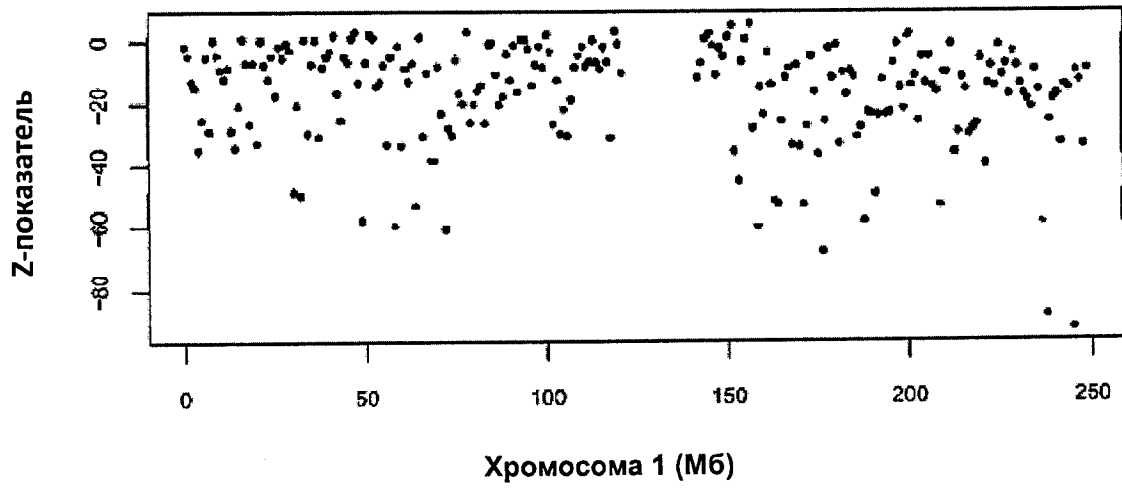
2400



Фиг. 24В

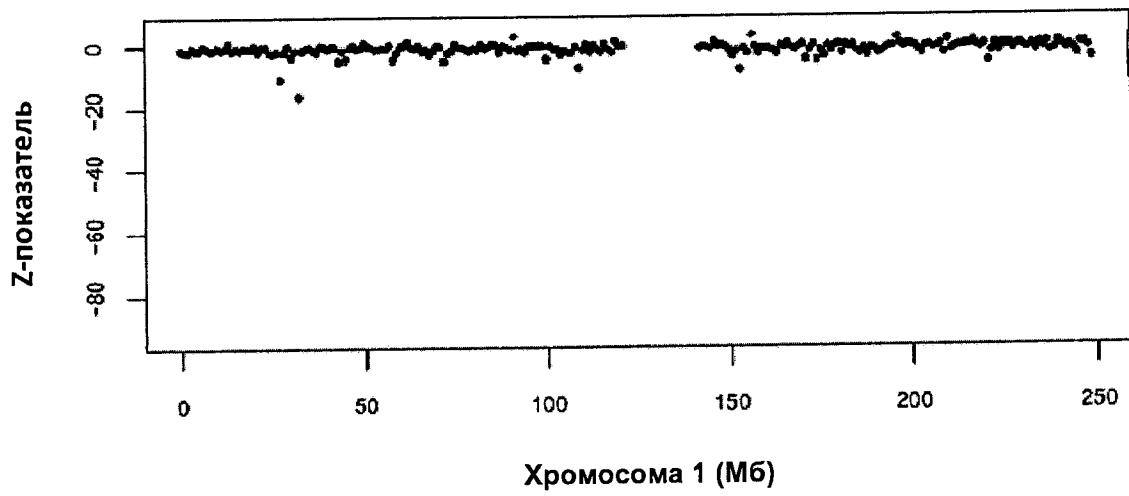
2450

32/68



Фиг. 25А

2500



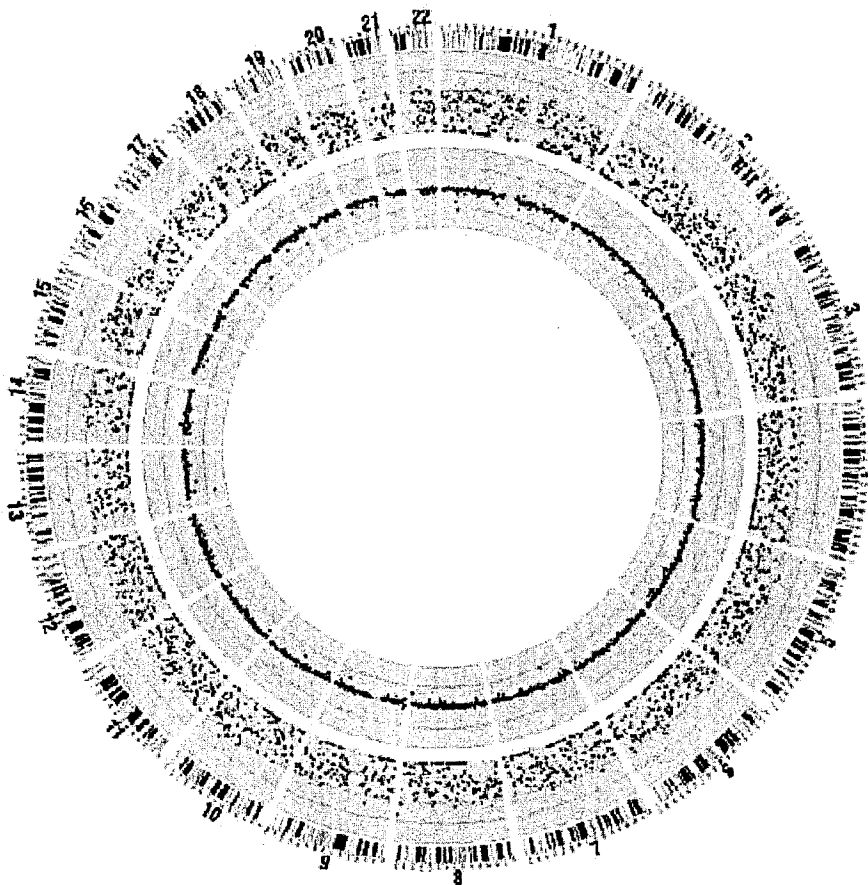
Фиг. 25В

2550

Z-показатели	Предоперационная плазма (число участков)	Послеоперационная плазма (число участков)
<-3	186 (80,9%)	19(8,3%)
От -3 до 3	42 (18,3%)	209 (90,3%)
>3	2 (0,9%)	2 (0,9%)
Всего	230(100%)	230(100%)

Фиг. 26А

2600



Фиг. 26В

2620

z-показатель	Предоперац. плазма (Число участков)	Послеоперац. плазма (Число участков)	С06	Контроль 1 (Число участков)	Контроль 2 (Число участков)	Контроль 3 (Число участков)	Контроль 4 (Число участков)
<-3	2330 (85,2%)	171 (6,3%)	29 (1,1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
От -3 до 3	362 (13,3%)	2556 (93,5%)	2688 (98,3%)	2734 (100%)	2734 (100%)	2734 (100%)	2734 (100%)
>3	42 (1,5%)	7 (0,2%)	17 (0,62%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Всего	2734 (100%)	2734 (100%)	2734 (100%)	2734 (100%)	2734 (100%)	2734 (100%)	2734 (100%)

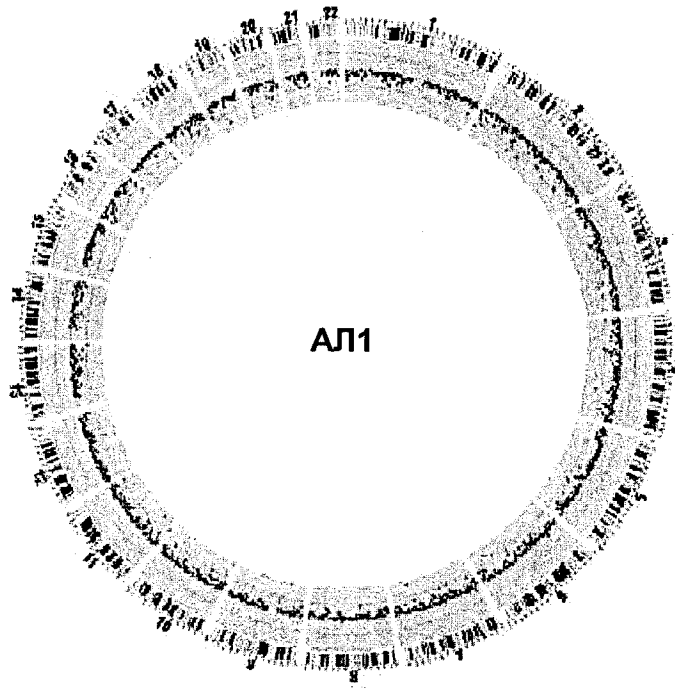
Фиг.26С

2640

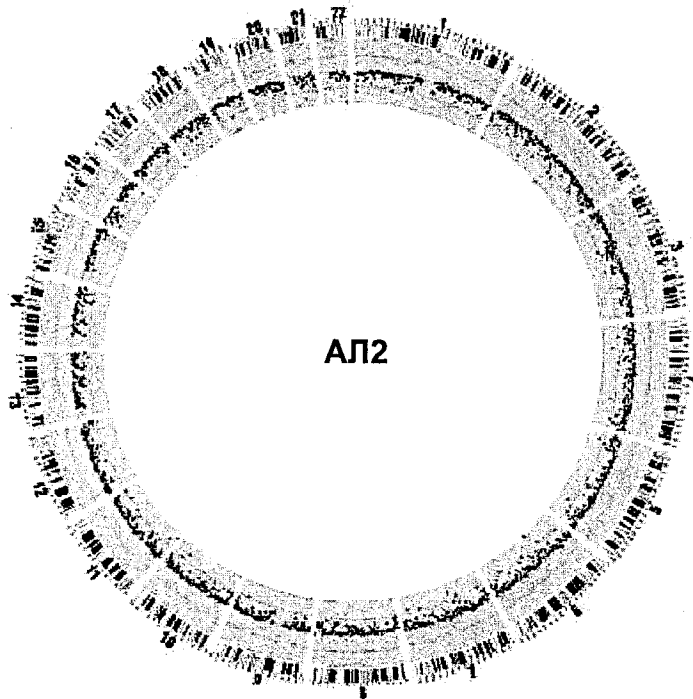
	Доля метилированных СНН	Доля метилированных СНG	Доля метилированных цитозинов	Доля метилированных CpG (плотность метилирования)
Лейкоцитарная пленка от пациента с ГЦК	0,13	0,14	5,74	67,08
Неопухолевая ткань печени	0,16	0,15	5,45	66,85
опухолевая ткань	0,09	0,09	3,29	42,88
Предоперац. плазма	0,09	0,10	3,99	59,67
Послеоперац. плазма	1,25	0,96	6,52	70,75
Контрольная плазма 1	0,12	0,12	6,38	71,17
Контрольная плазма 2	0,10	0,10	5,62	71,33
Контрольная плазма 3	0,09	0,09	5,57	71,20
Контрольная плазма 4	0,10	0,10	5,93	72,48

Фиг. 26D

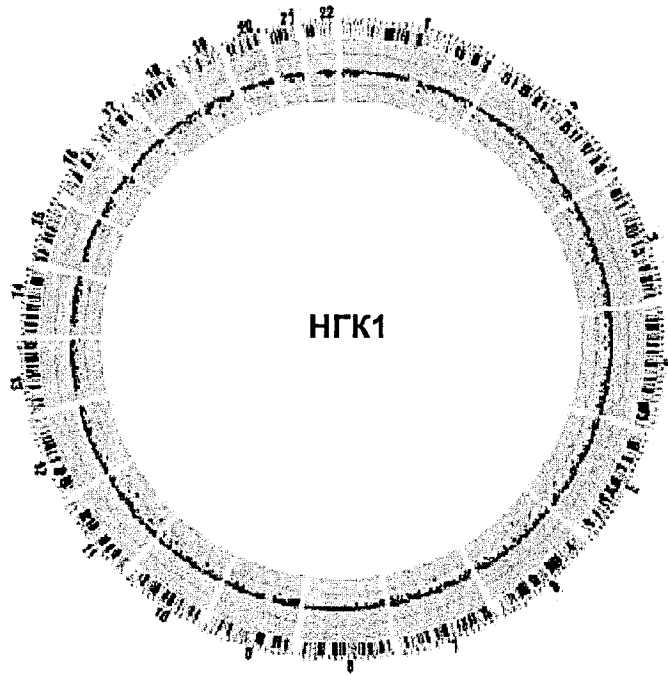
2660



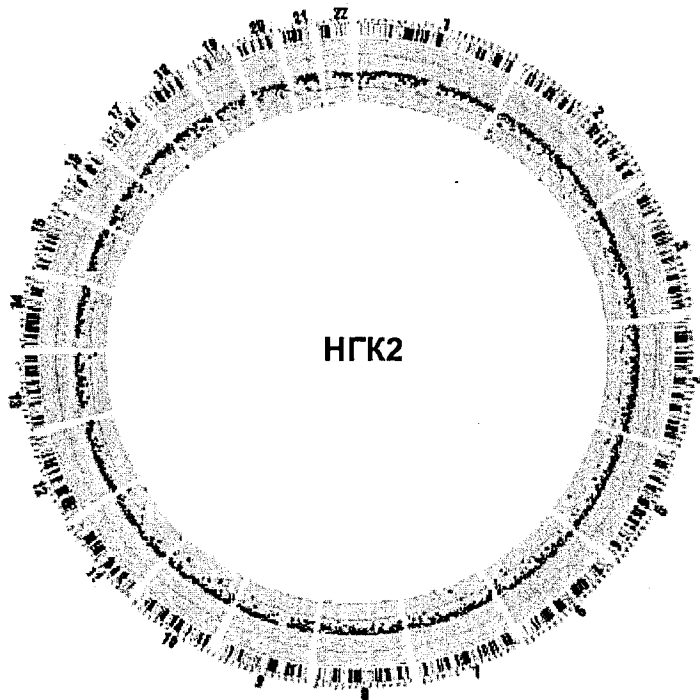
Фиг. 27А



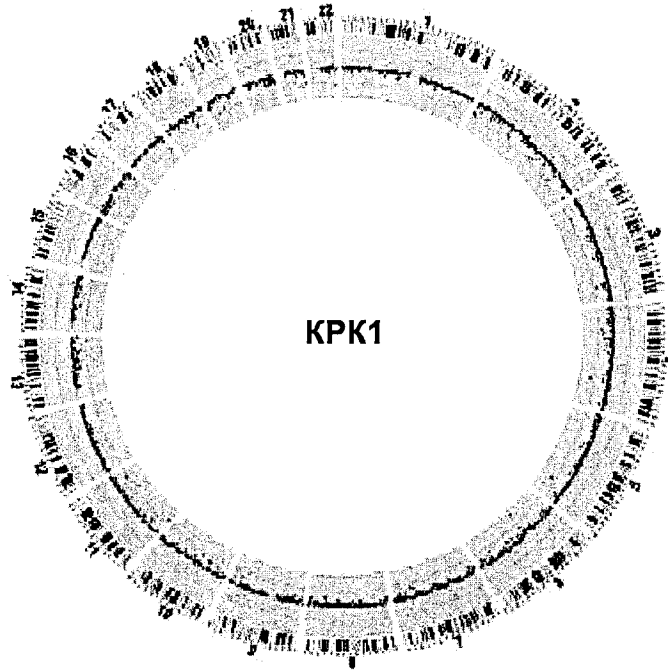
Фиг. 27В



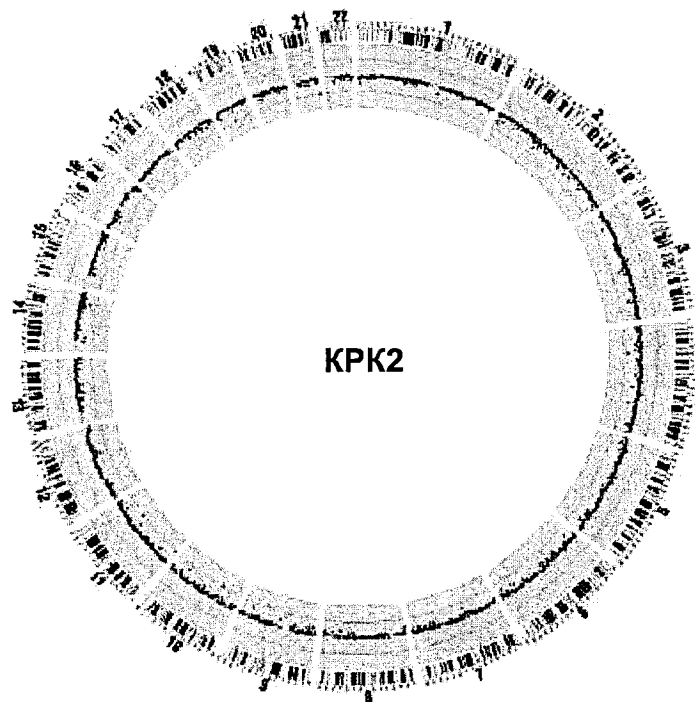
Фиг. 27С



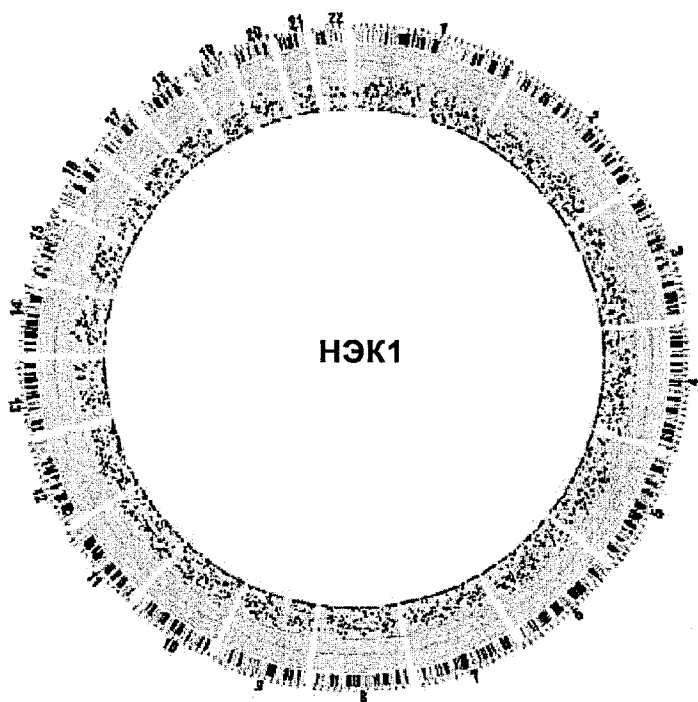
Фиг. 27D



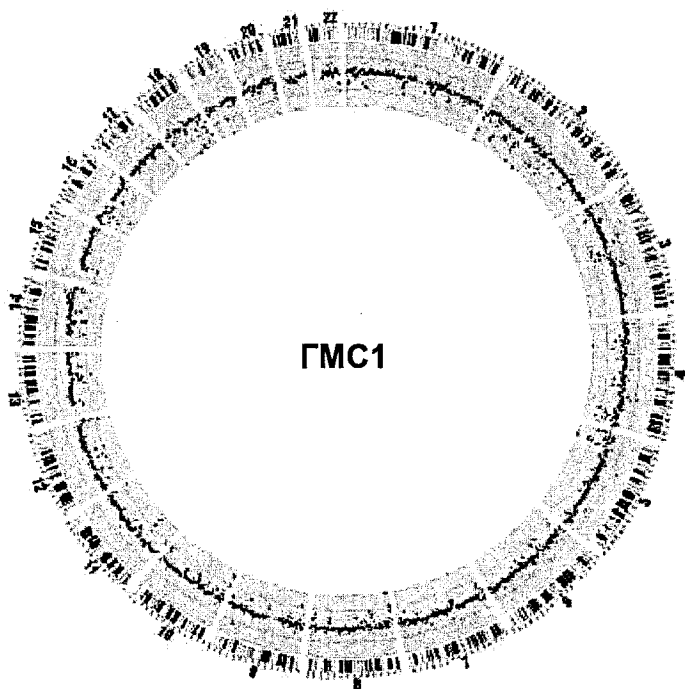
Фиг. 27Е



Фиг. 27F



Фиг. 27G



Фиг. 27H

	Число необработанных ридов	Число картируемых ридов	Недубл. Фрагменты	Доля подходящих для использования ридов %	Глубина на цепь	Покрытие CpG %	Глубина покрытия ридами на сайтах CpG на цепь
CL1	120807588	91750018	76890830	64%	0,69	33,32%	2,87
CL2	101344850	76473792	66246435	65%	0,59	30,26%	2,79
NPC1	101883643	75281330	57236188	56%	0,51	27,87%	2,72
NPC2	102745073	74211781	56841611	55%	0,51	28,74%	2,70
CRC1	87732645	66456695	56112543	64%	0,50	27,38%	2,63
CRC2	109955730	82594007	67253101	61%	0,60	29,91%	2,90
NE1	105514000	79923333	64534726	61%	0,58	27,86%	2,58
SMS1	94727279	71698718	64426529	68%	0,58	29,23%	2,79

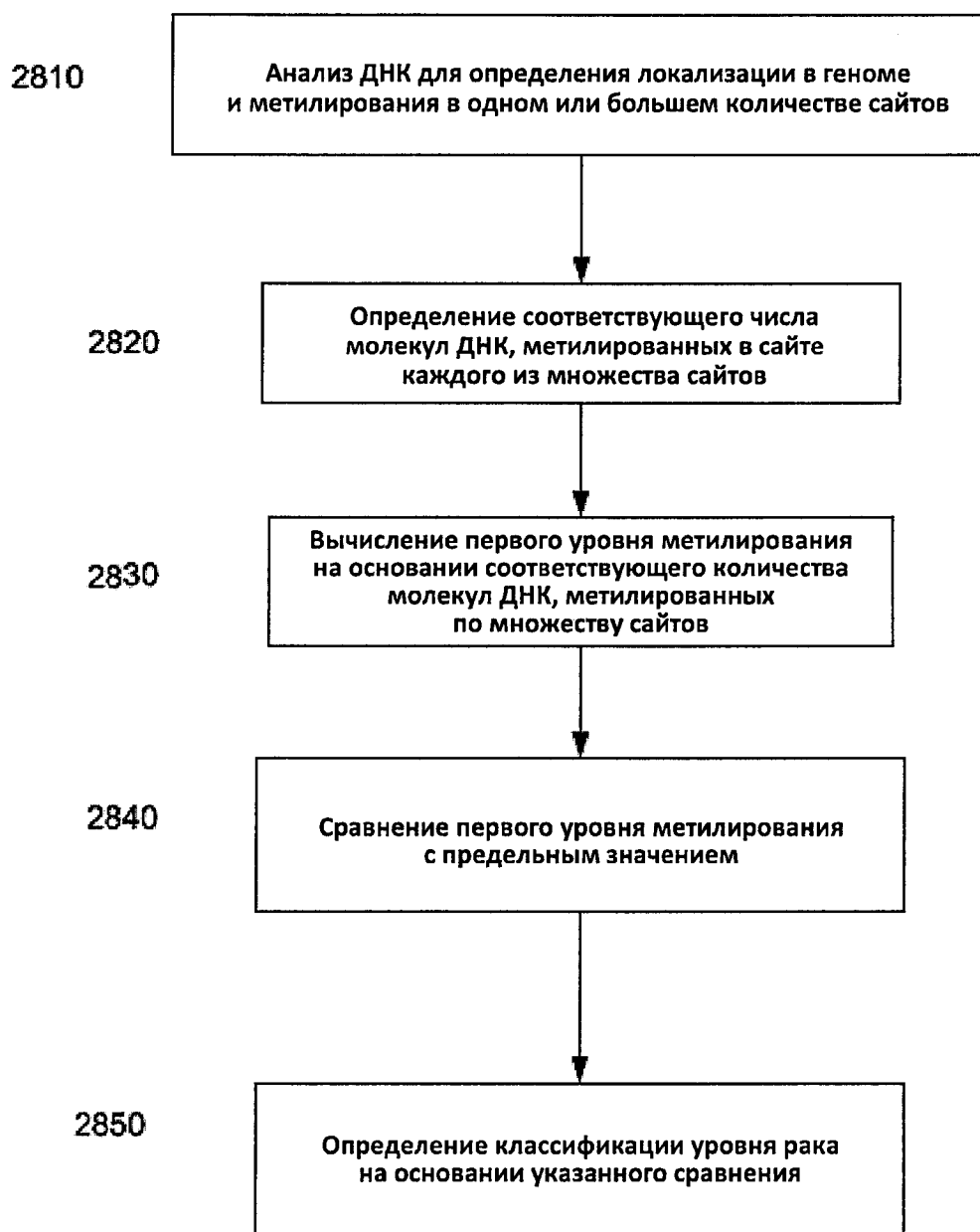
Фиг. 27I

2780

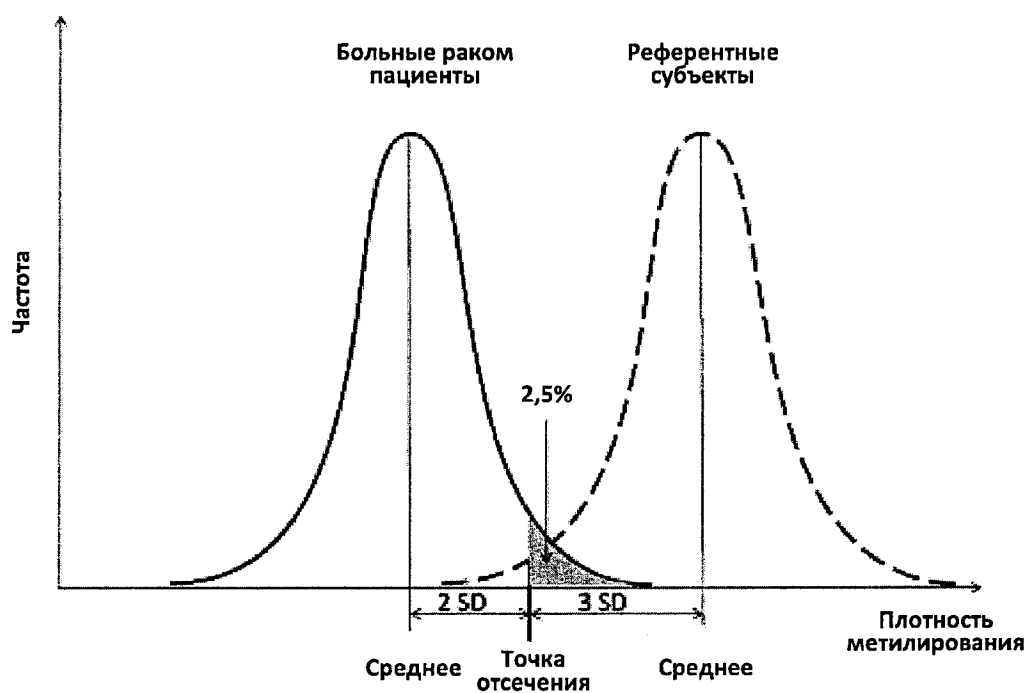
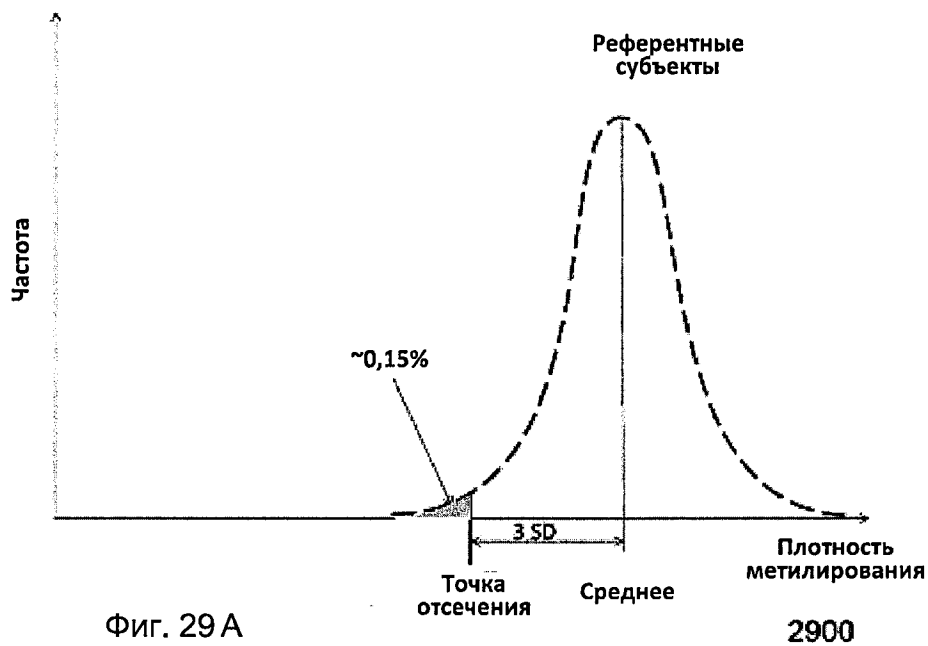
z-показатель	CL1	CL2	NPC1	NPC2	CRC1	CRC2	NE1	SMS1
<-3	34,9%	33,9%	17,6%	54,3%	13,5%	9,8%	98,1%	11,2%
От -3 до 3	65,0%	65,9%	82,4%	45,6%	86,5%	90,2%	1,7%	70,8%
>3	0,07%	0,15%	0,04%	0%	0,07%	0%	0,15%	18,0%

Фиг. 27J

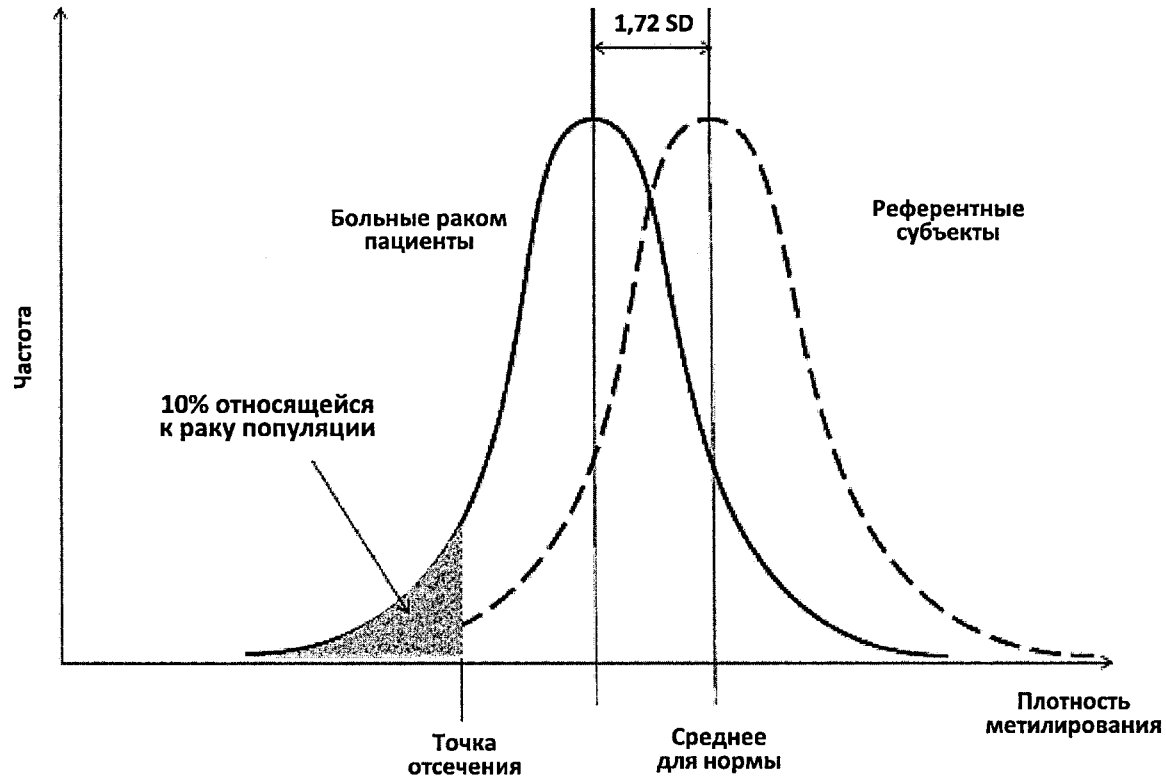
2790



Фиг. 28



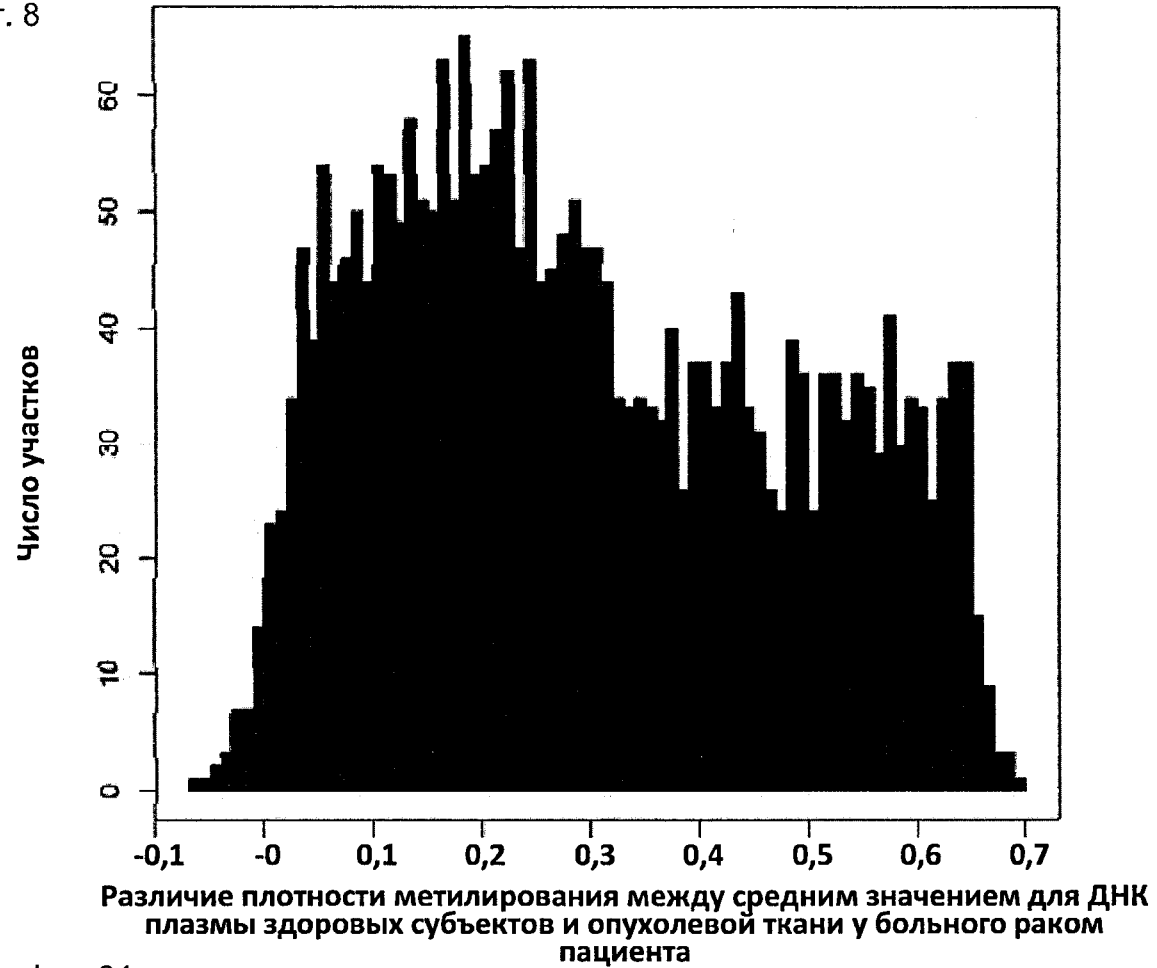
2950



Фиг. 30

3000

Фиг. 8



Фиг. 31

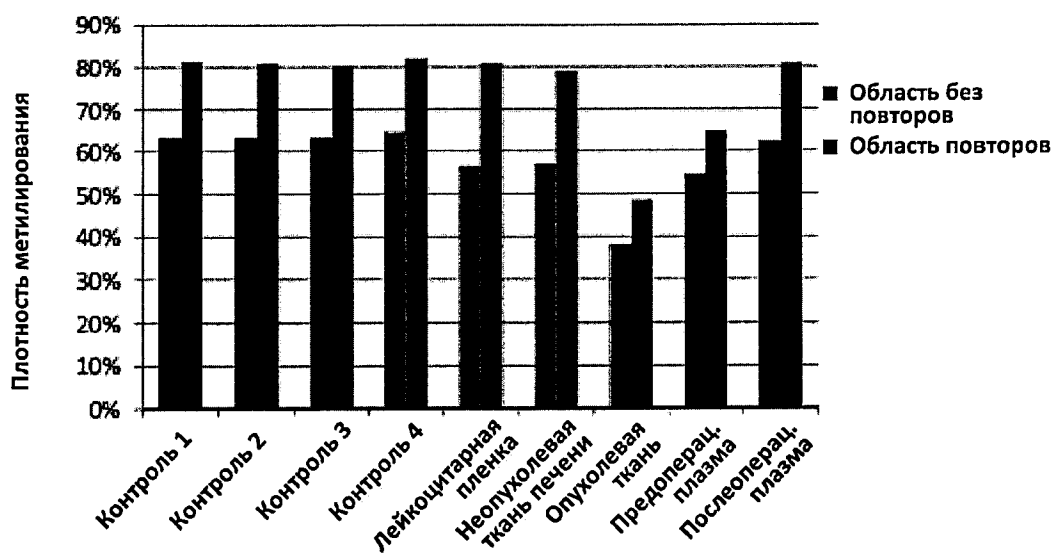
3100

Коэфф. уменьшения глубины при 5%	Глубина секвенирования (кратность секвенирования гаплоидного генома)	Число гипометилированных участков (z-показатели <-3) в предоперационной плазме	% гипометилированных участков (z-показатели <-3) в предоперационной плазме
0,8	1,76	1314	48,06
0,6	1,32	1323	48,39
0,4	0,88	1220	44,62
0,2	0,44	1094	40,01
0,1	0,22	900	32,92
0,05	0,11	645	23,59
0,01	0,022	207	7,57

Коэфф. уменьшения глубины при 2%	Глубина секвенирования (кратность секвенирования гаплоидного генома)	Число гипометилированных участков (z-показатели <-3) в предоперационной плазме	% гипометилированных участков (z-показатели <-3) в предоперационной плазме
0,8	1,76	491	17,96
0,6	1,32	493	18,03
0,4	0,88	460	16,83
0,2	0,44	439	16,06
0,1	0,22	356	13,02
0,05	0,11	297	10,86
0,01	0,022	231	8,45

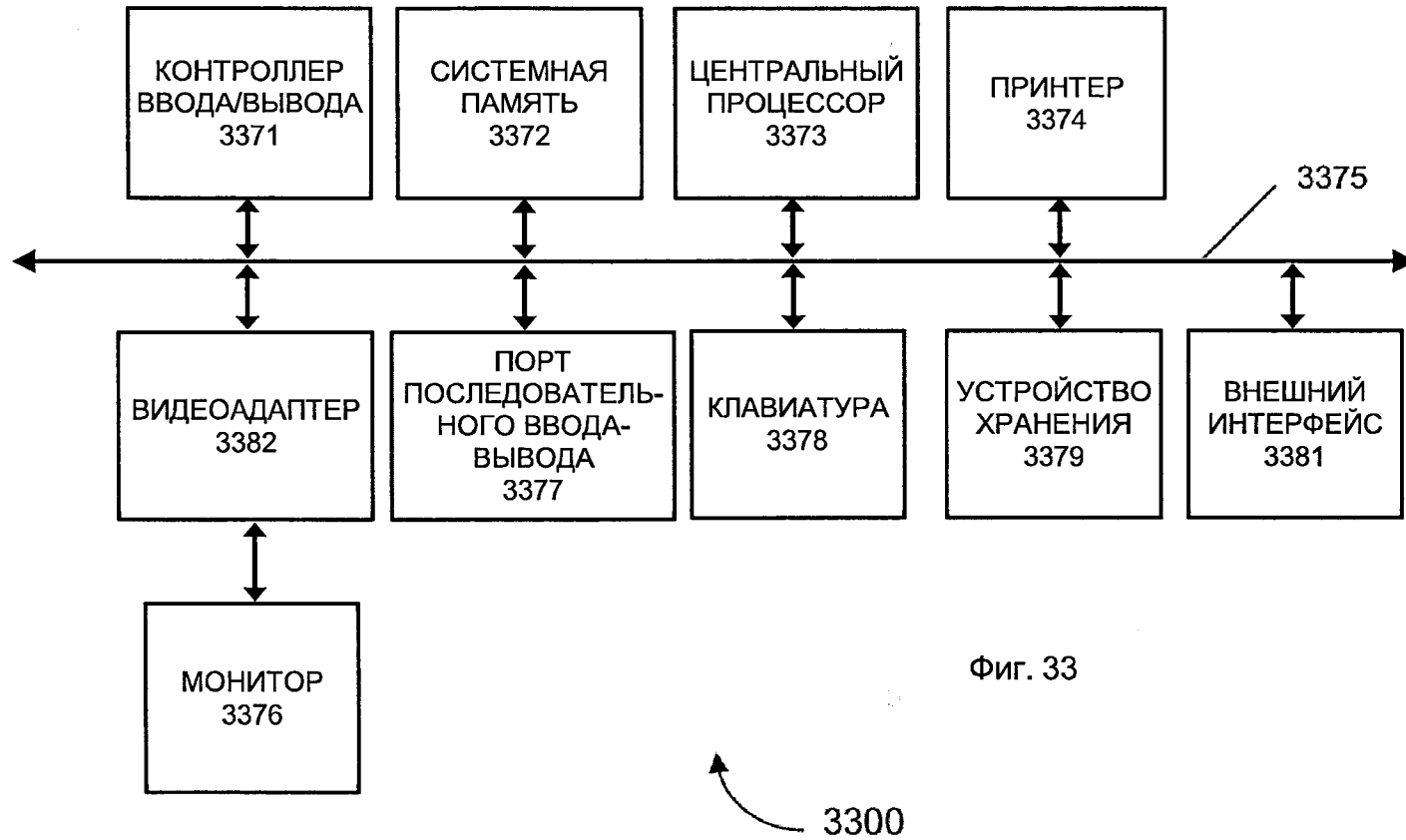
Фиг. 32А

3200

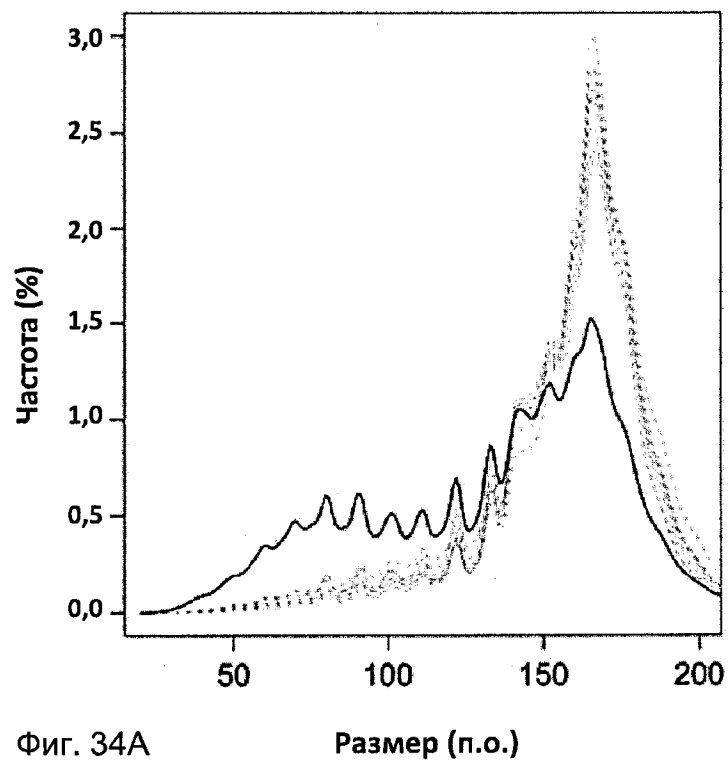


Фиг. 32В

3250

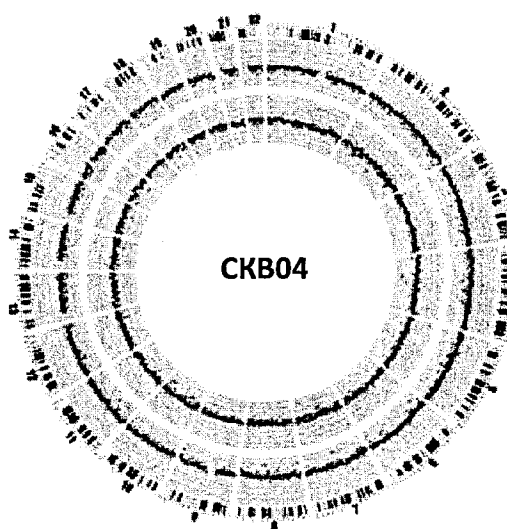


Фиг. 33

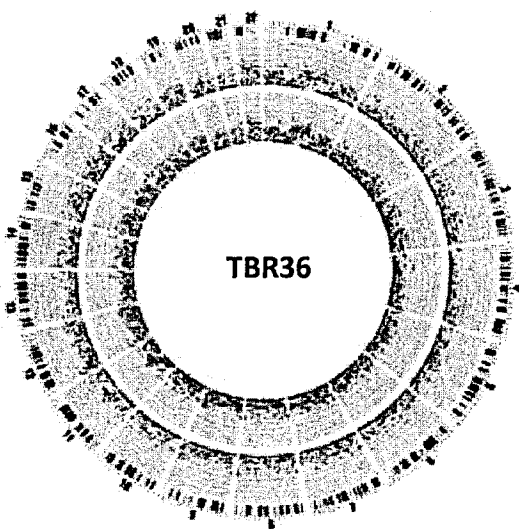


Фиг. 34А

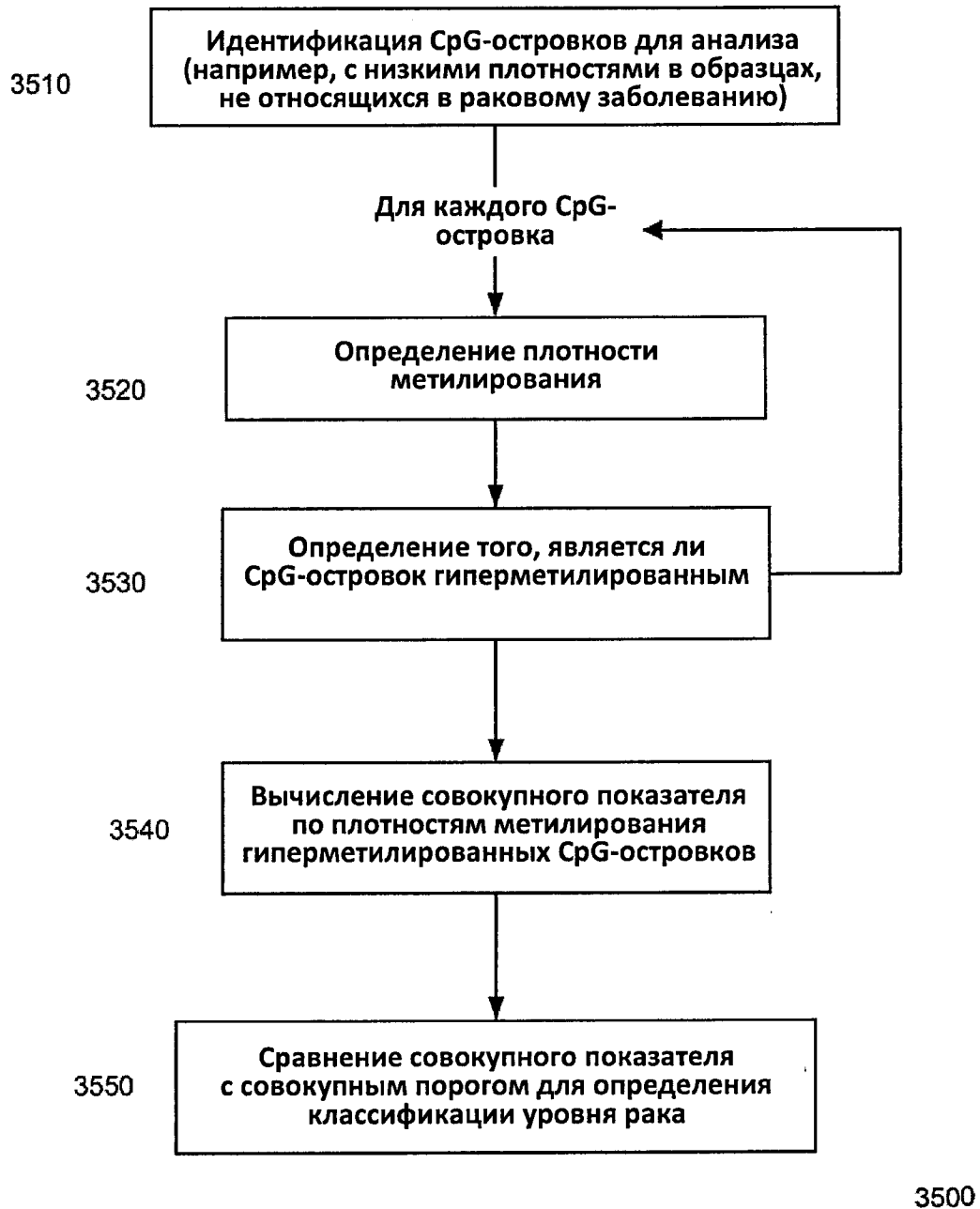
Размер (п.о.)



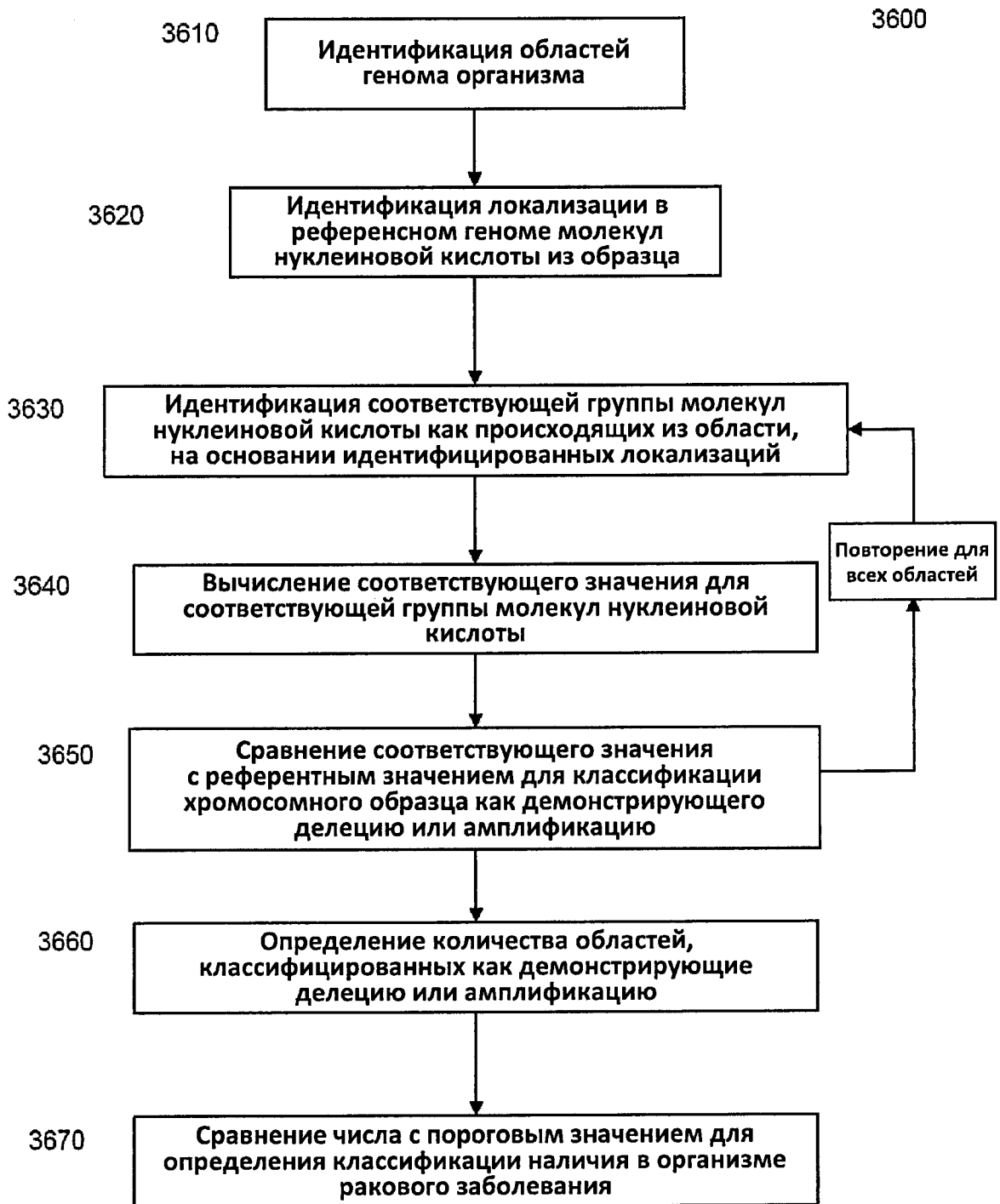
Фиг. 34В



Фиг. 34С



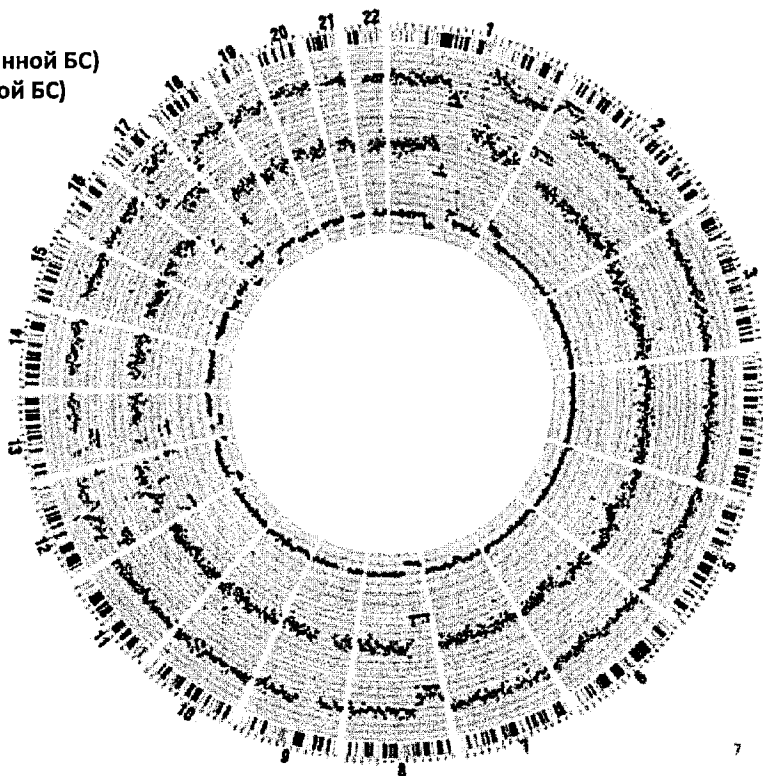
Фиг. 35



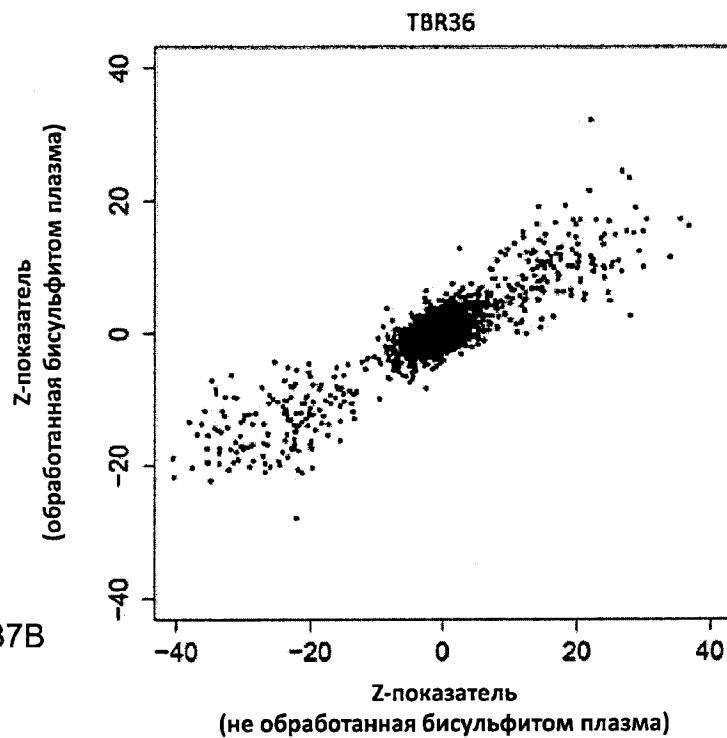
Фиг. 36

Образец: TBR36
Доля опухолевой ДНК: 40%

От центра к краям:
1) Опухолевые CNA
2) CNA плазмы (не обработанной БС)
3) CNA плазмы (обработанной БС)



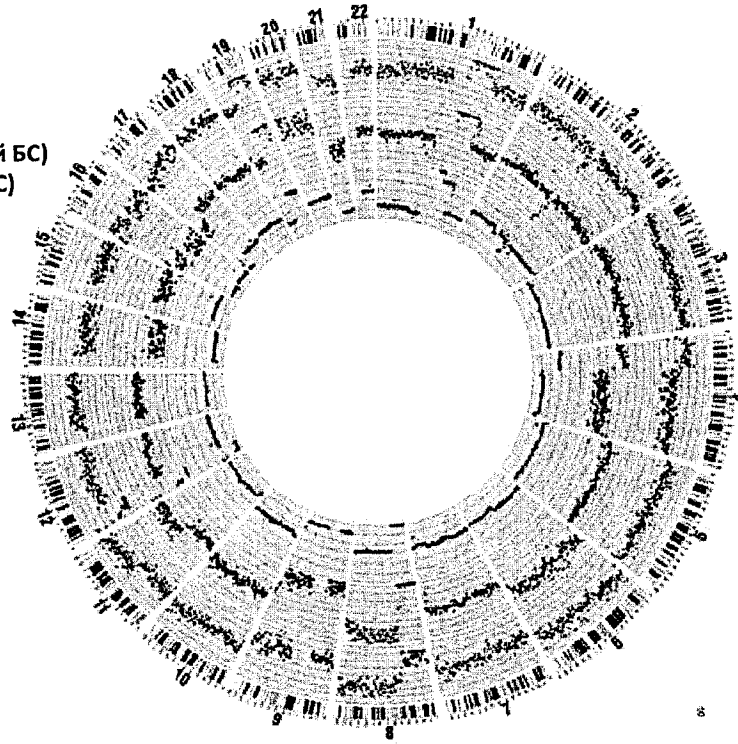
Фиг. 37А



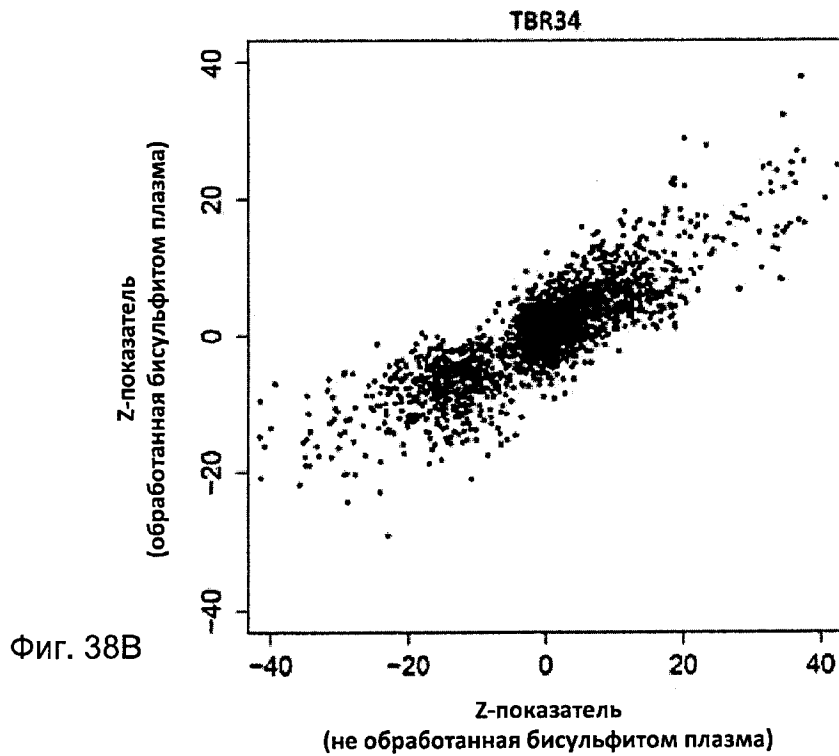
Фиг. 37В

Образец: TBR34
Доля опухолевой ДНК: 30%

- От центра к краям:
1) Опухолевые CNA
2) CNA плазмы (не обработанной БС)
3) CNA плазмы (обработанной БС)



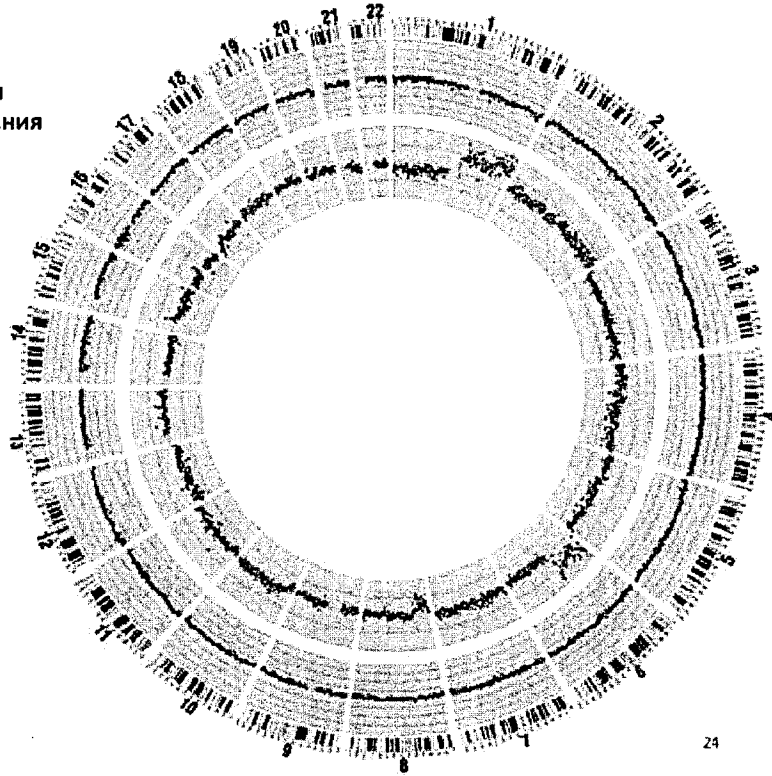
Фиг. 38А



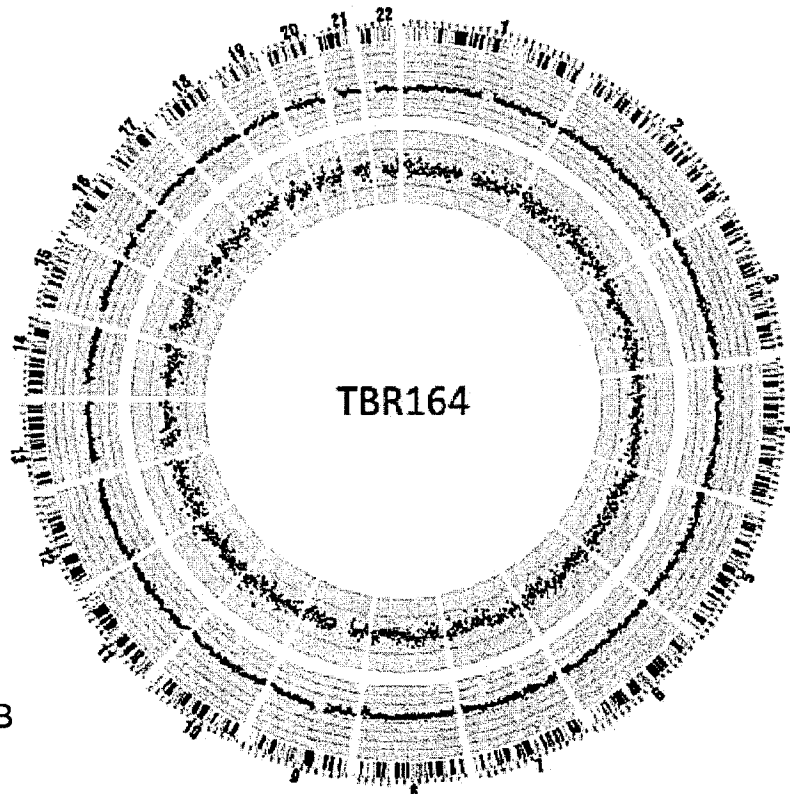
Фиг. 38В

Образец: TBR240
Размер опухоли: 2,8 × 2,3 см

От центра к краям:
1) Z-показатель для SNA плазмы
2) Z-показатель для метилирования плазмы



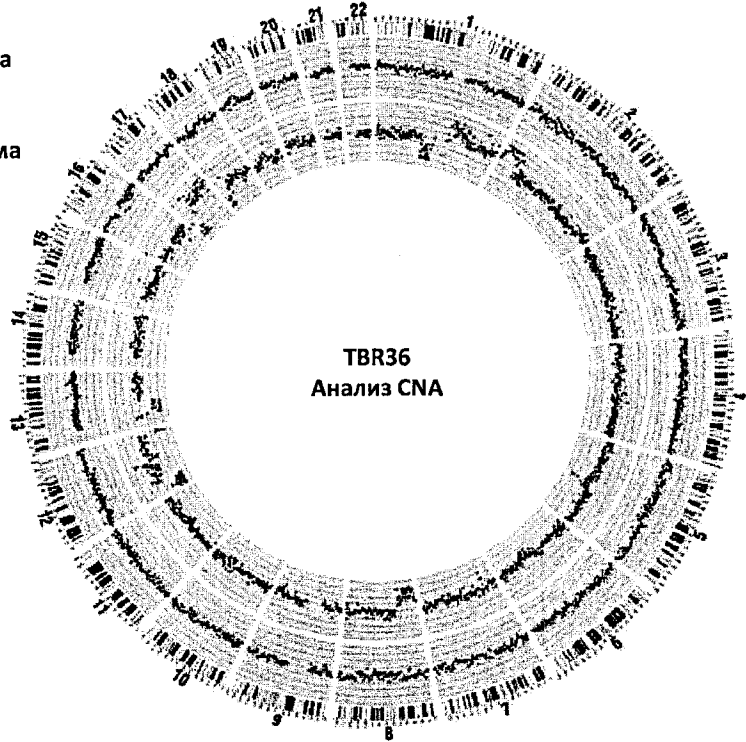
Фиг. 39А



Фиг. 39В

Внутренний круг:
Предоперационная плазма

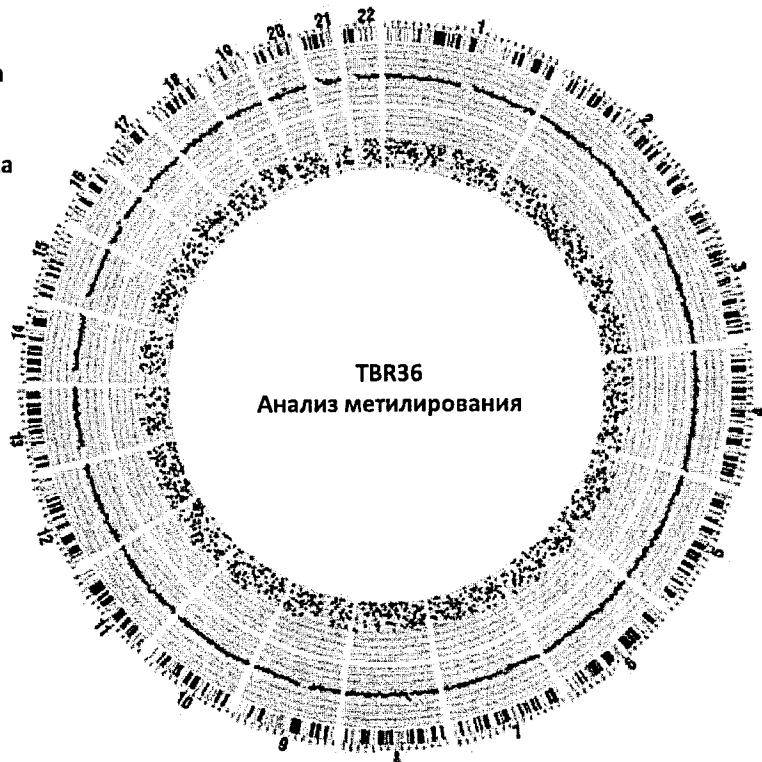
Внешний круг:
Послеоперационная плазма



Фиг. 40А

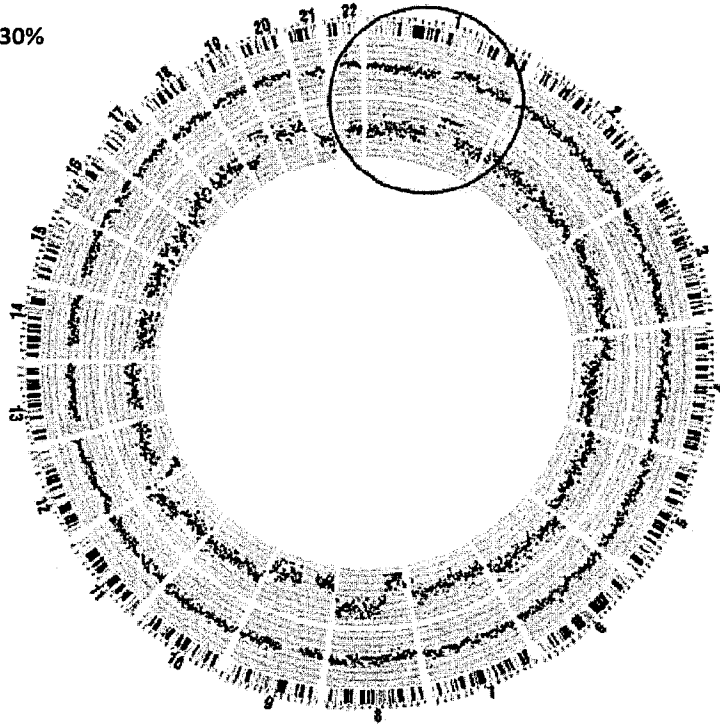
Внутренний круг:
Предоперационная плазма

Внешний круг:
Послеоперационная плазма



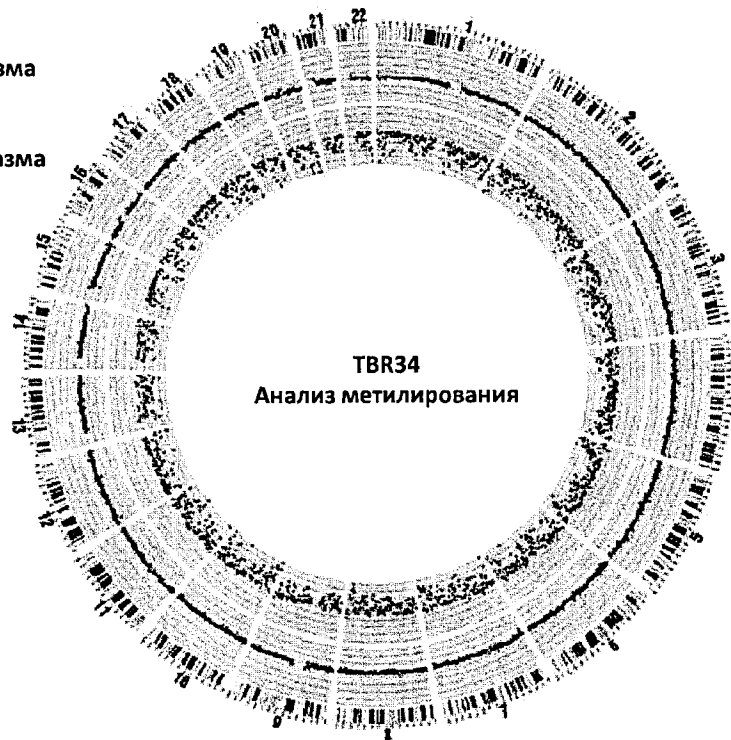
Фиг. 40В

Образец: TBR34
Доля опухолевой ДНК: 30%
От краев к центру:
1) Предоперац. CNV
2) Послеоперац. CNV

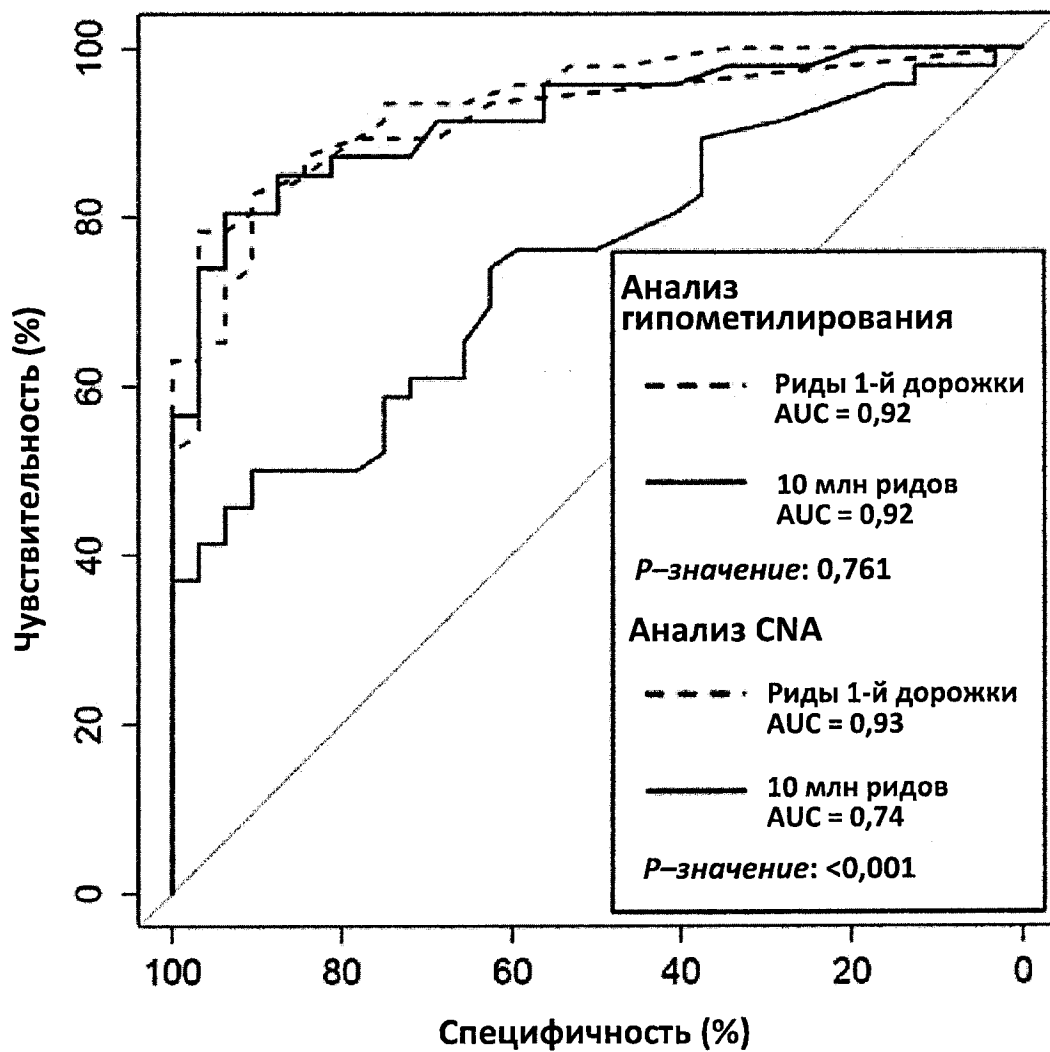


Фиг. 41А

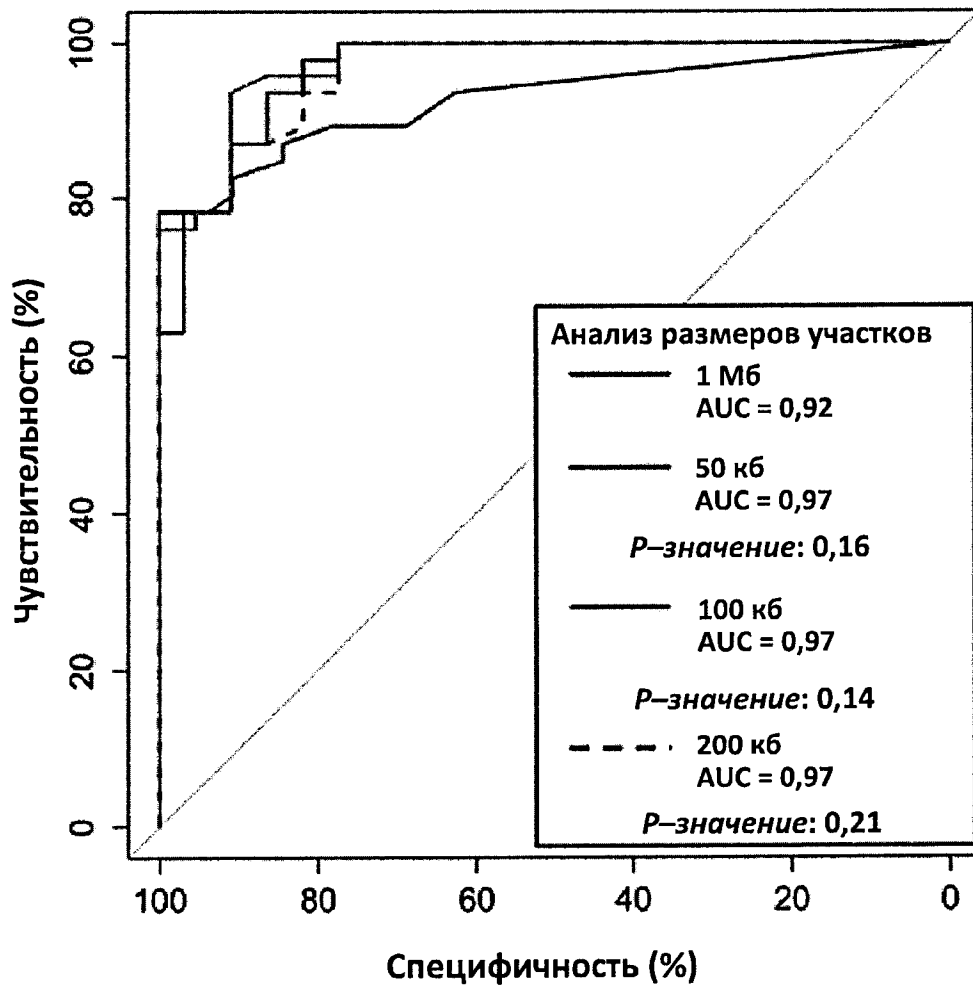
Внутренний круг:
Предоперационная плазма
Внешний круг:
Послеоперационная плазма



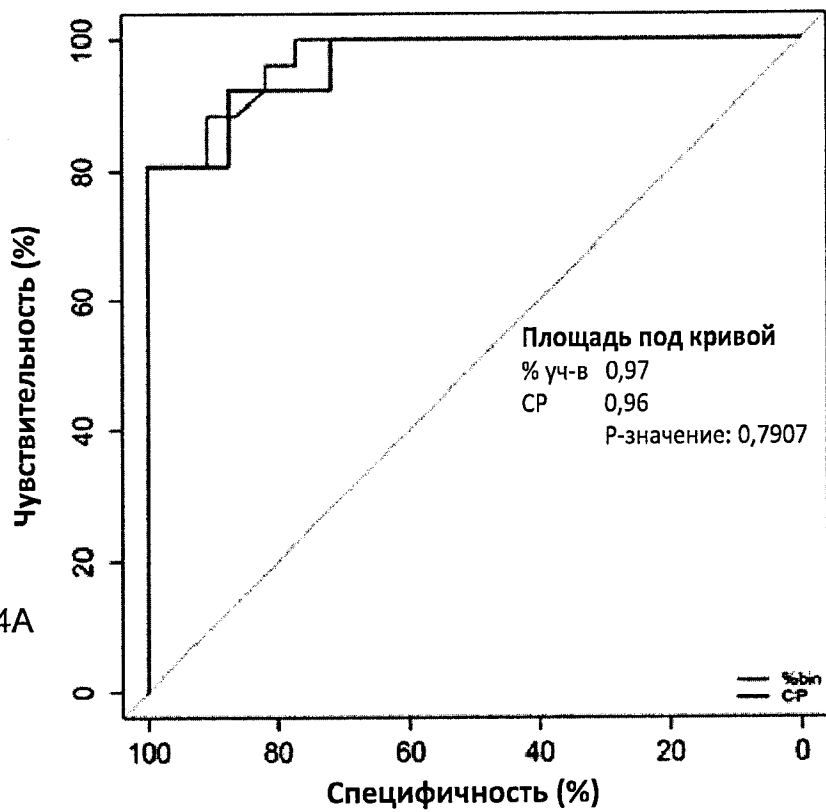
Фиг. 41В



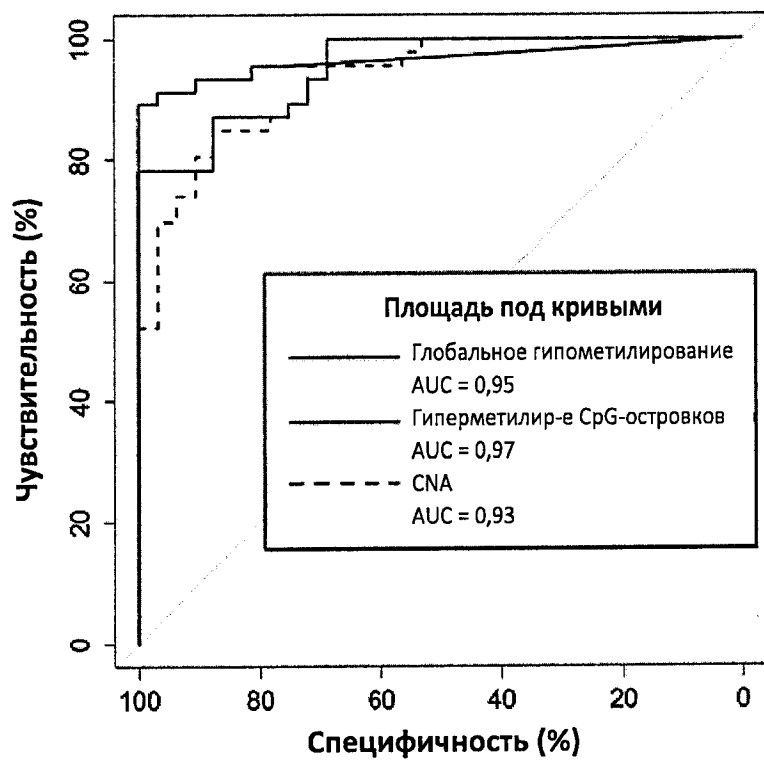
Фиг. 42



Фиг. 43



Фиг. 44В

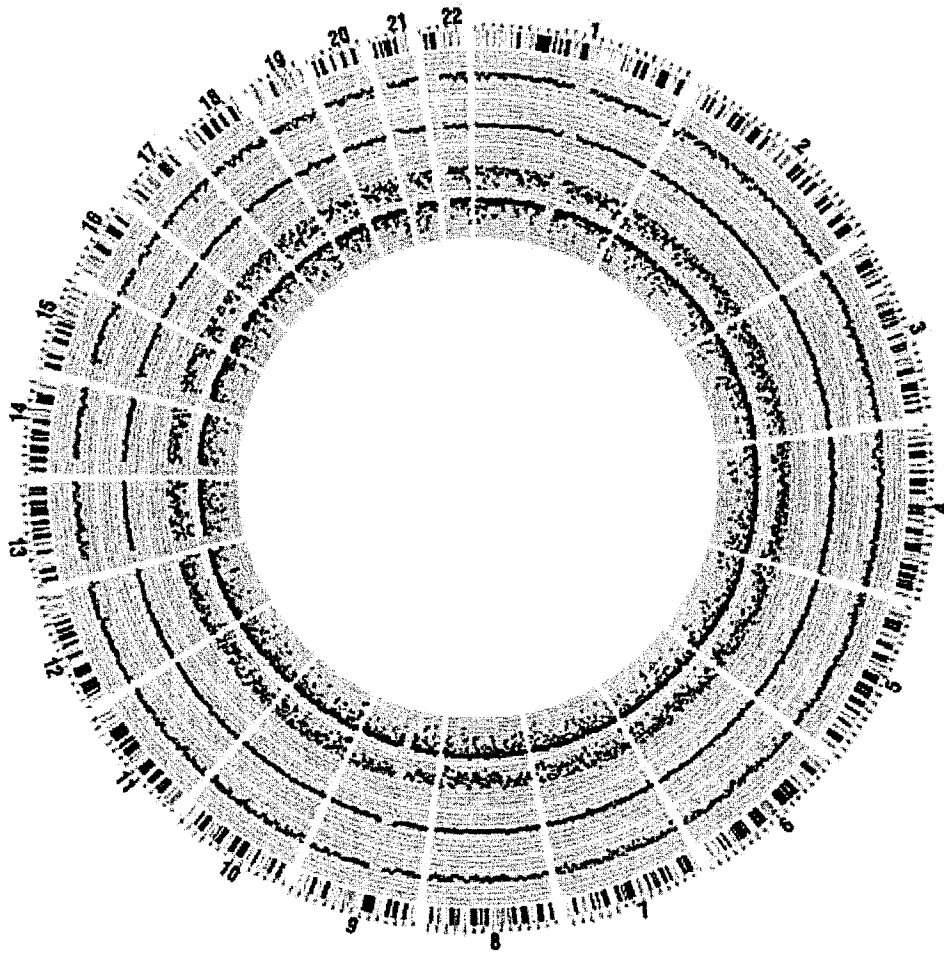


Случай №	Стадия по классификации ВСЛС	Максимальный размер опухоли(ей) (см)	Показатель СР			Присутствие какого-либо типа изменений
			Общее гипометилирование	Гиперметилирование CpG-островков	CNA	
TBR36	A1	16	88585	80,9	3777	Положит.
TBR34	A1	13	37573	221,7	8650	Положит.
HOT198	A1	12	93132	52,5	6803	Положит.
HOT205	A1	6	1146	1,7	175	Положит.
HOT 192	A1	5,5	25914	119,5	1668	Положит.
HOT227	A1	5,5	36333	155,4	5008	Положит.
HOT238	A1	5,5	12187	105,3	1978	Положит.
HOT197	B	5,3	1714	9,5	28	Положит.
HOT156	A1	5	16529	115,9	1947	Положит.
HOT236	A1	3,7	7993	5,9	85	Положит.
HOT 170	A1	3,5	22271	81,9	1542	Положит.
HOT229	A1	3,1	383	5,4	71	Положит.
HOT233	A1	3	189	7,5	322	Положит.
HOT240	A1	2,8	1277	14,0	972	Положит.
HOT222	A1	2,6	18832	10,5	172	Положит.
HOT162	A1	2,5	26112	41,0	3048	Положит.
HOT172	A1	2,5	12909	98,1	1305	Положит.
HOT164	A1	2,3	12604	4,8	210	Положит.
HOT215	A1	2,3	25824	54,8	714	Положит.
HOT207	A1	2,1	3377	12,0	101	Положит.
HOT224	A1	2	367	0,0	13	Отриц.
HOT159	A1	1,5	1367	1,4	26	Положит.
HOT204	A1	1,5	102	4,5	198	Положит.
HOT151	A1	1,5	153	2,8	9	Отриц.
HOT208	A1	1,2	5867	4,7	214	Положит.
HOT167	A1	1	21544	4,5	54	Положит.

Фиг. 45

Тип ракового заболевания	Случай №:	Стадирование/ максимальный размер опухоли	Показатель СР			Присутствие какого-либо типа изменений
			Общее гипометилирование	Гиперметилирование СрG-островков	CNA	
Рак молочной железы	TBR150	Метастатический	<u>14581</u>	<u>146,7</u>	<u>5123</u>	Положит.
	TBR126	T2N1M0 4,5 см	<u>10415</u>	<u>34,9</u>	<u>1951</u>	Положит.
	TBR127	T2N1M0 2,5 см	<u>2598</u>	<u>4,3</u>	54	Положит.
	TBR128	T2N0M0 2 см	157	<u>7,7</u>	<u>2250</u>	Положит.
	TBR111	T1N0M0 0,9 см	298	0,0	128	Отриц.
Гладкомышечная саркома	TBR051	Метастатический	<u>4787</u>	<u>79,8</u>	<u>7986</u>	Положит.
Нейроэндокринная опухоль	TBR052	Метастатический	<u>112518</u>	<u>119,8</u>	<u>14681</u>	Положит.
Рак легкого	TBR164	Метастатический	<u>51818</u>	<u>47,3</u>	<u>4714</u>	Положит.
	TBR012	Метастатический	<u>4351</u>	<u>7,7</u>	<u>2213</u>	Положит.
	TBR014	Метастатический	<u>2356</u>	<u>12,2</u>	<u>825</u>	Положит.
	TBR177	T2N1M0 3,9 см	<u>1072</u>	<u>10,6</u>	25	Положит.
Рак носоглотки	TBR031	Метастатический	<u>1028</u>	<u>4,4</u>	<u>1860</u>	Положит.
	TBR125	Метастатический	<u>20949</u>	<u>162,5</u>	<u>6708</u>	Положит.
	TBR124	T4N1	270	<u>4,9</u>	37	Положит.
	TBR123	T1N2M0	<u>6871</u>	<u>11,8</u>	80	Положит.
	TBR062	T3N1M0	111	<u>30,9</u>	41	Положит.
	TBR108	T3N1M0	93	<u>7,7</u>	26	Положит.
	TBR107	T3N0M0	<u>2439</u>	<u>41,2</u>	205	Положит.
	TBR106	T1N1M0	<u>2281</u>	<u>5,9</u>	97	Положит.
TBRO 99	T1N0M0	<u>6187</u>	<u>10,2</u>	26	Положит.	

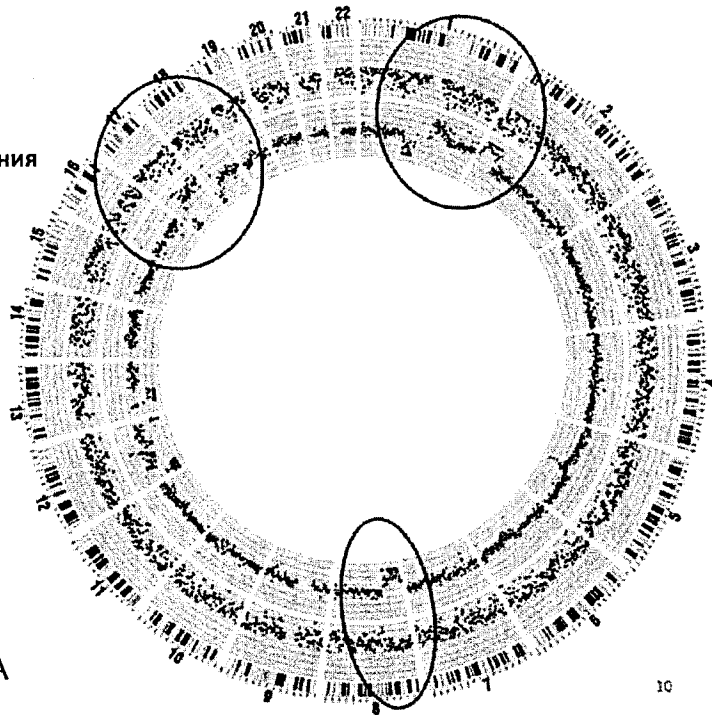
Фиг. 46



Фиг. 47

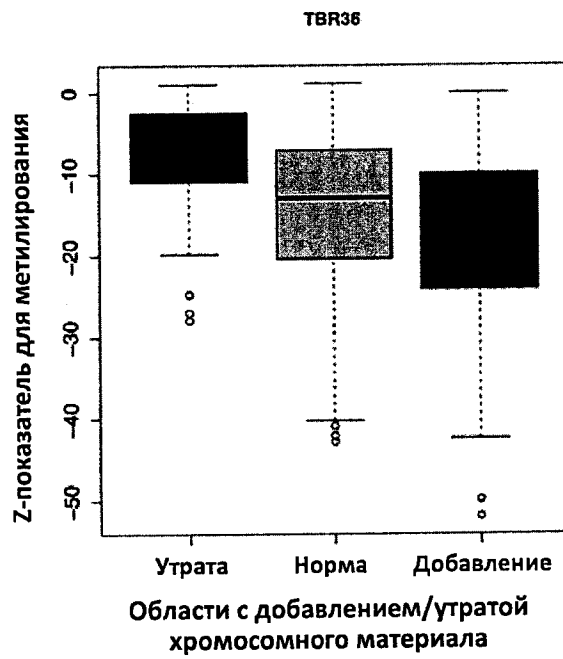
Образец: TBR36
Доля опухолевой ДНК: 40%

От центра к краям:
1) Z-показатель для CNA
2) Z-показатель для метилирования



Фиг. 48А

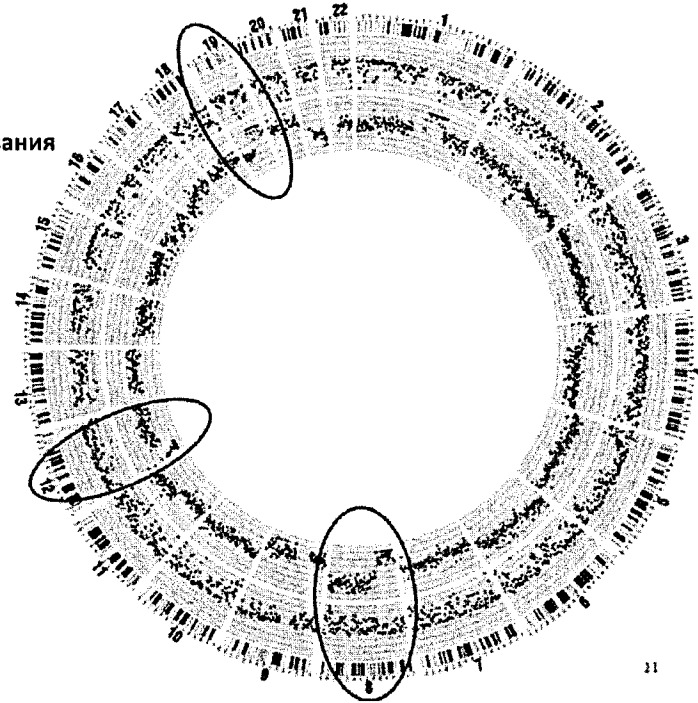
Образец: TBR36
Доля опухолевой ДНК: 40%



Фиг. 48В

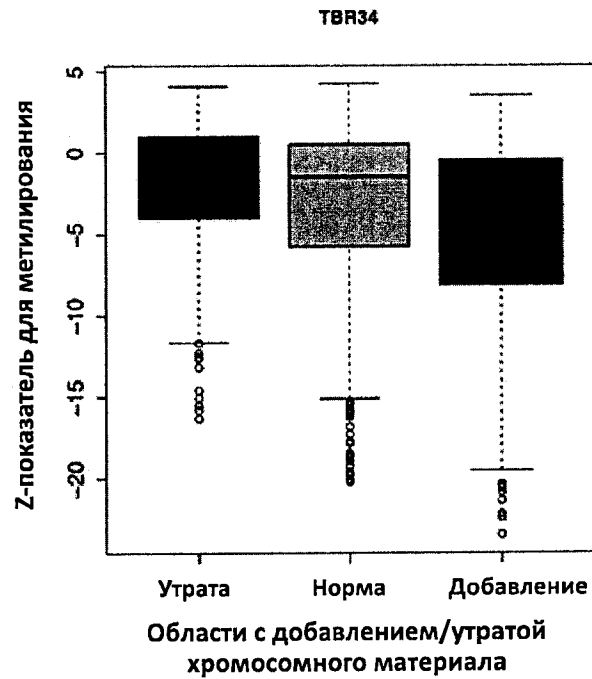
Образец: TBR34
Доля опухолевой ДНК: 30%

От центра к краям:
1) Z-показатель для CNA
2) Z-показатель для метилирования

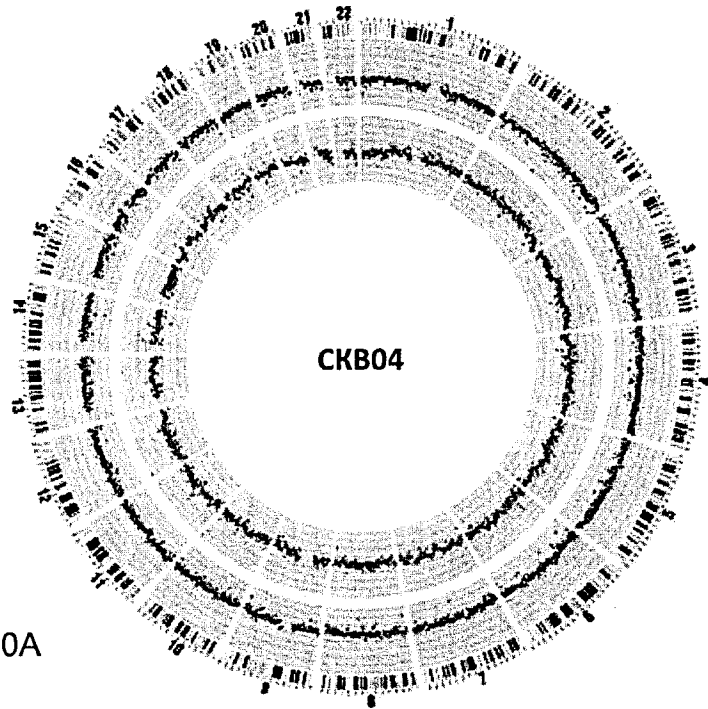


Фиг. 49А

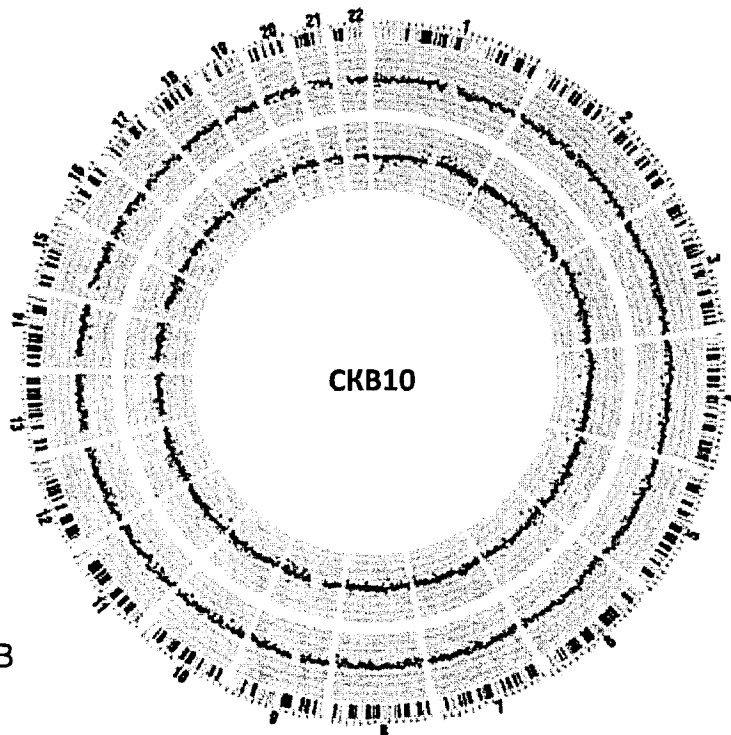
Образец: TBR34
Доля опухолевой ДНК: 30%



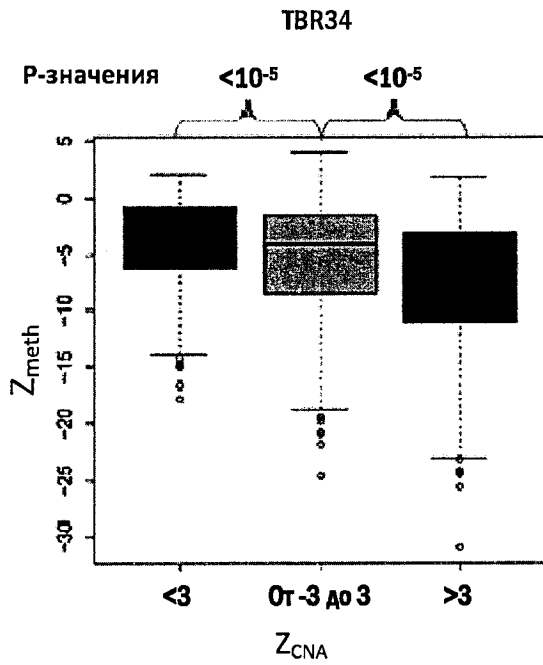
Фиг. 49В



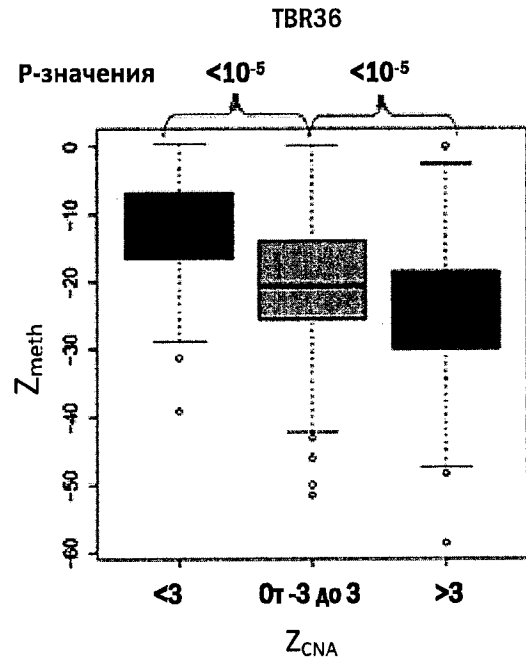
Фиг. 50А



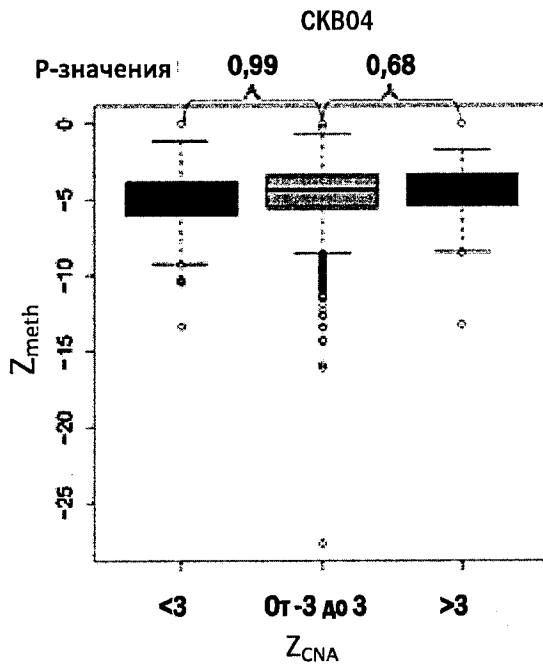
Фиг. 50В



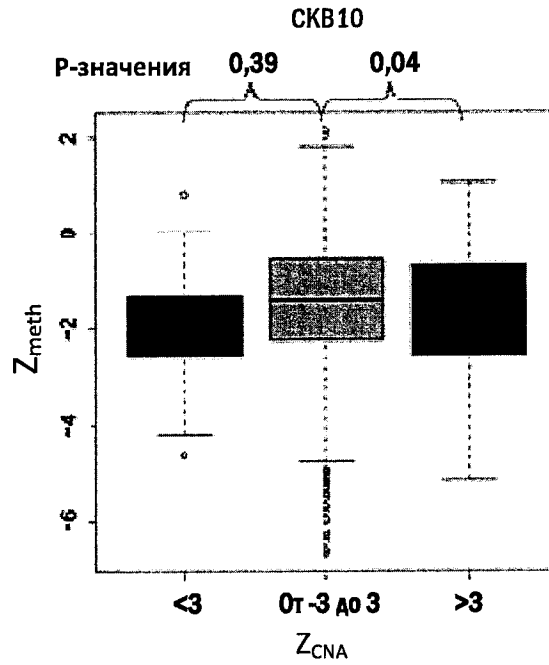
Фиг. 51А



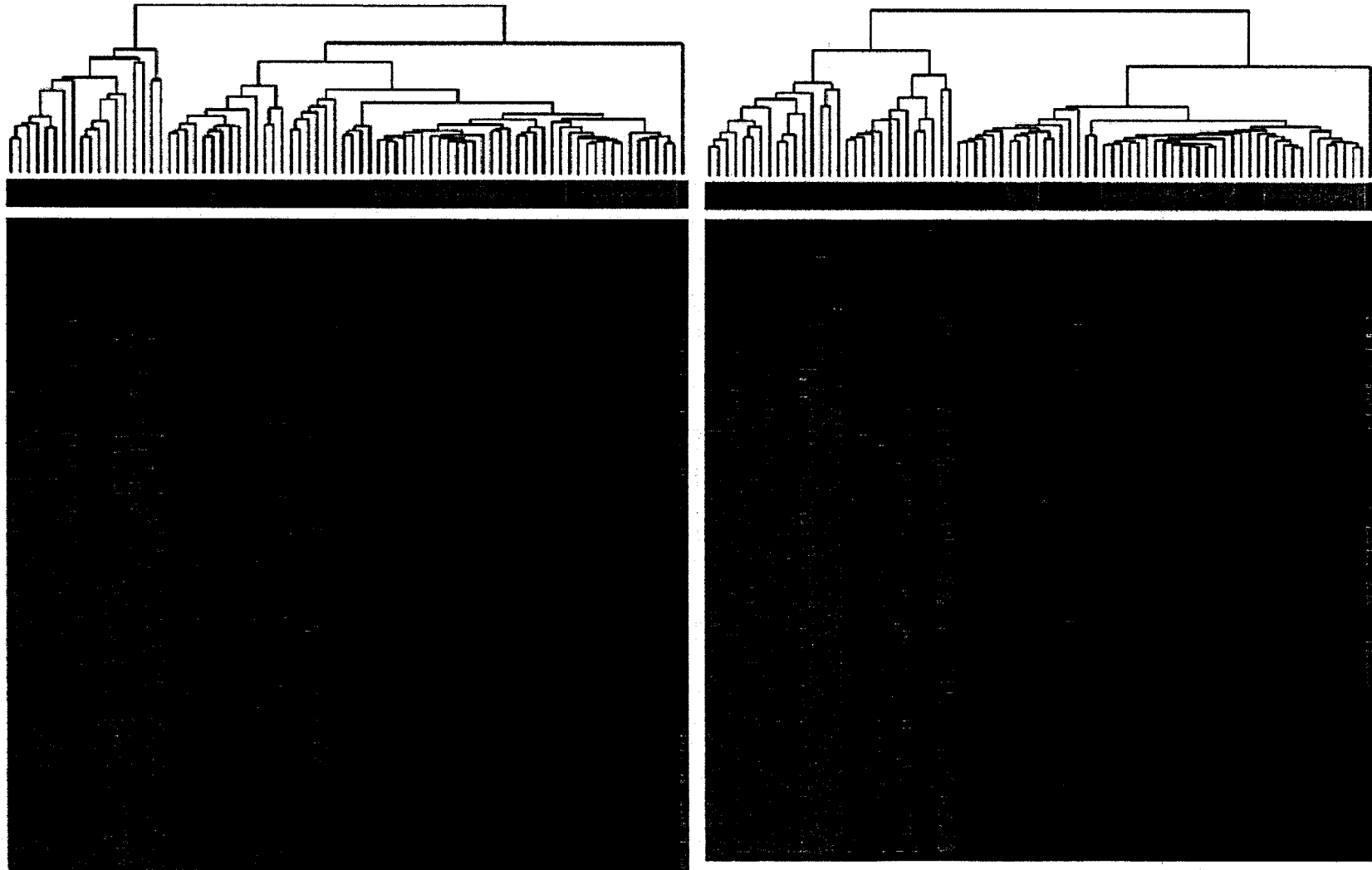
Фиг. 51В



Фиг. 51С

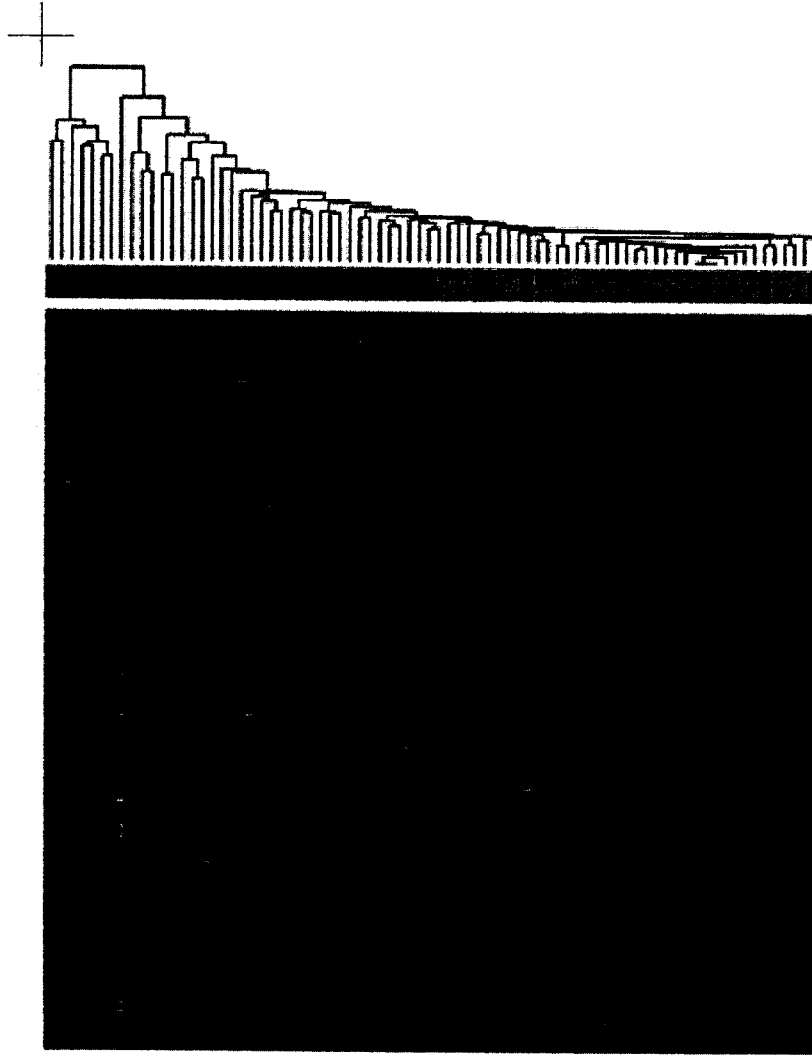


Фиг. 51D

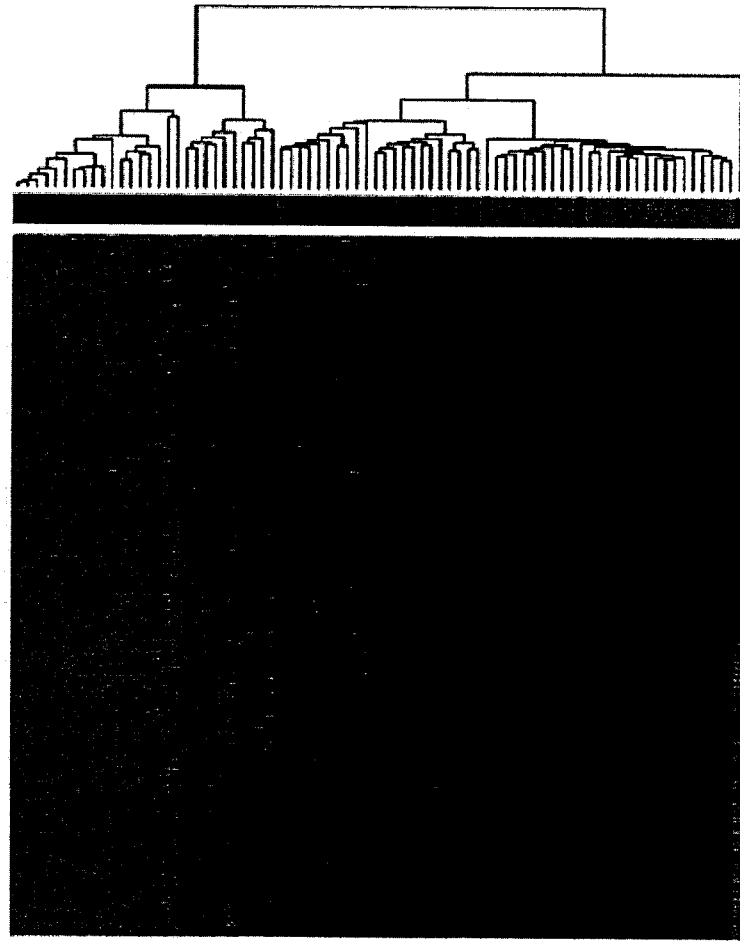


Фиг. 52А

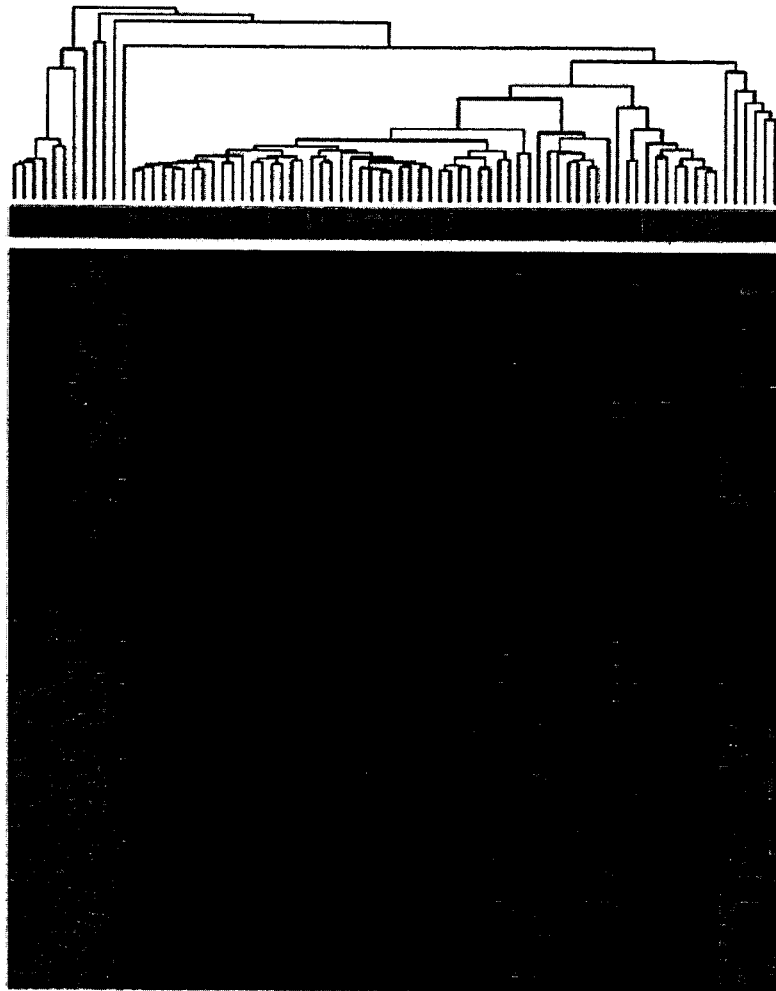
Фиг. 52В



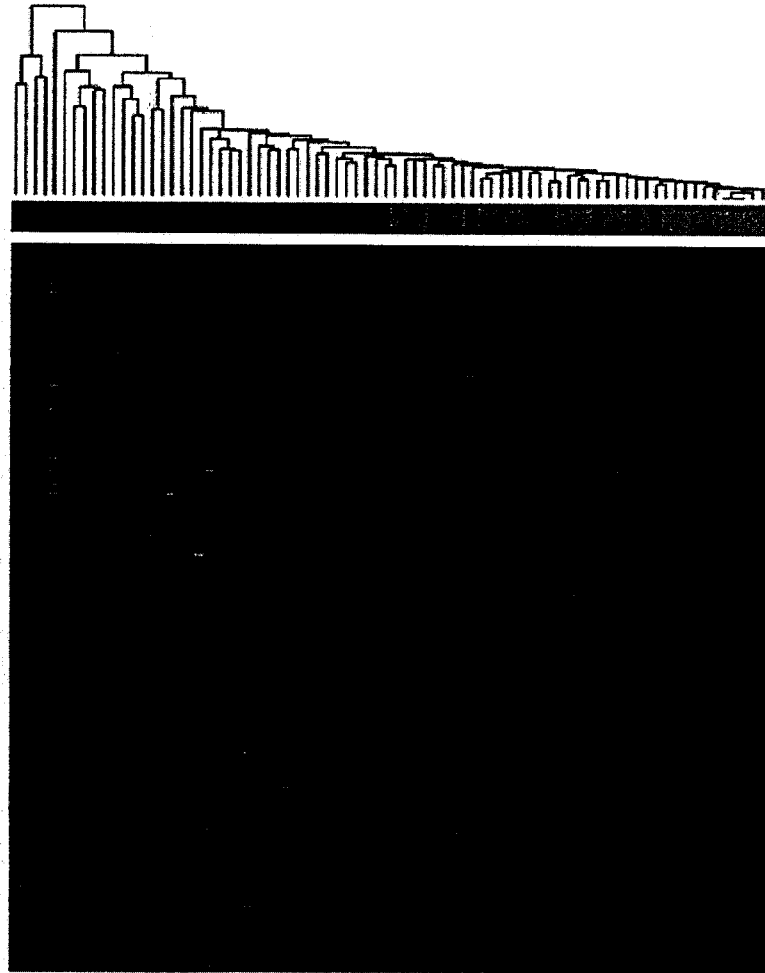
Фиг. 53А



Фиг. 53В

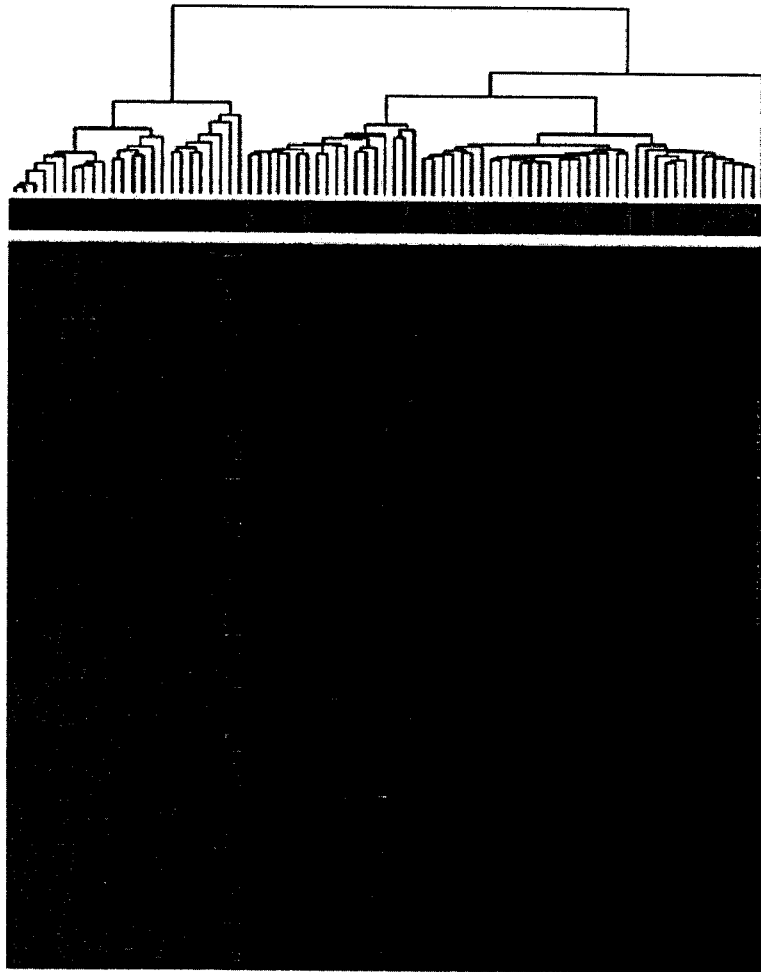


Фиг. 54А

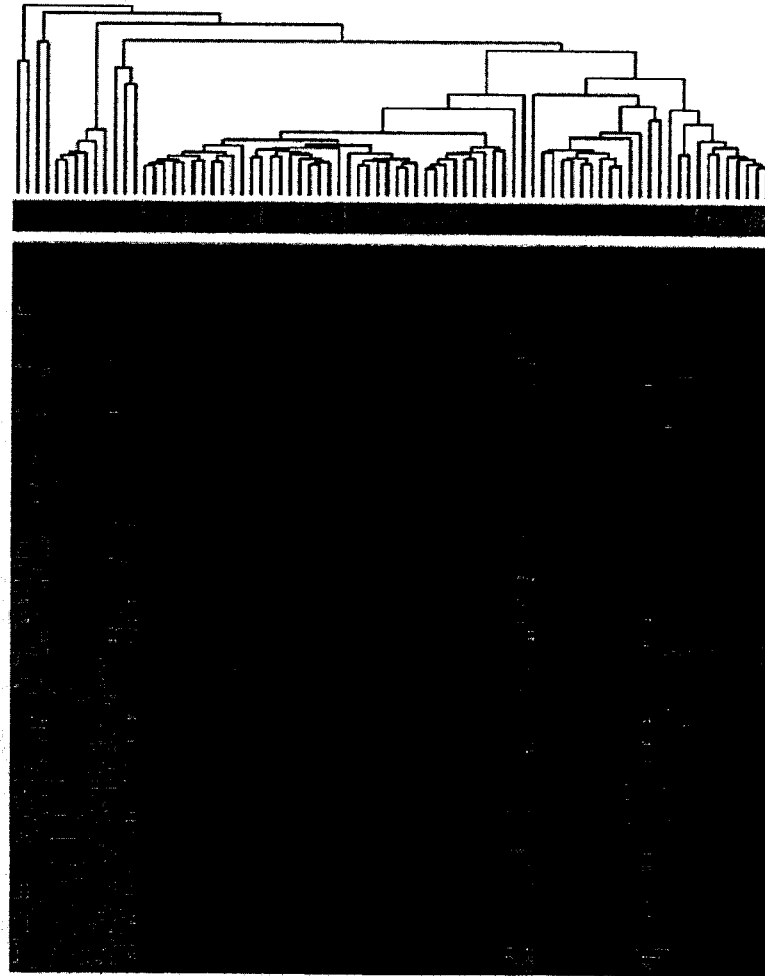


Фиг. 54В



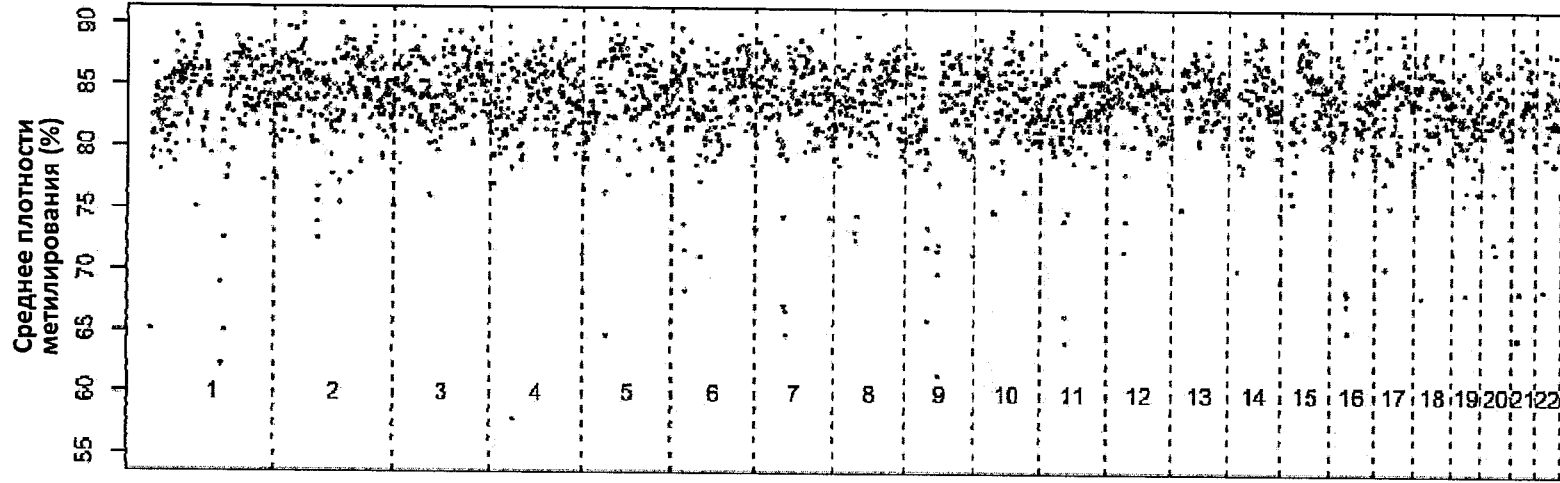


Фиг. 55А



Фиг. 55В





Фиг. 56