

(19)



**Евразийское
патентное
ведомство**

(21) **202191888** (13) **A1**

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2022.01.31

(51) Int. Cl. **C12Q 1/6809** (2018.01)

(22) Дата подачи заявки
2020.02.11

(54) СПОСОБ ОПРЕДЕЛЕНИЯ ПРОИСХОЖДЕНИЯ НУКЛЕИНОВЫХ КИСЛОТ В СМЕШАННОМ ОБРАЗЦЕ

(31) **19156966.4**

(32) **2019.02.13**

(33) **EP**

(86) **PCT/EP2020/053497**

(87) **WO 2020/165184 2020.08.20**

(71) Заявитель:
**НИПД ДЖЕНЕТИКС ПАБЛИК
КОМПАНИ ЛИМИТЕД (СУ)**

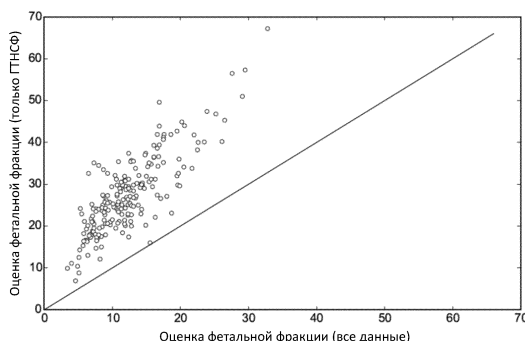
(72) Изобретатель:

**Кумбарис Джордж, Ахиллеос
Ахиллеас, Иоаннидес Мариос,
Патсалис Филиппос (СУ)**

(74) Представитель:

Нилова М.И. (РУ)

(57) Настоящее изобретение относится к способу определения происхождения фрагмента нуклеиновой кислоты или выявления фрагмента нуклеиновой кислоты в смеси фрагментов нуклеиновой кислоты, включающему этапы а) обеспечения смеси фрагментированных нуклеиновых кислот, которые происходят из эукариотического организма, б) приготовления библиотеки секвенирования из смеси фрагментированных нуклеиновых кислот, с) гибридизации одного или более зондов с по меньшей мере одним участком в указанной библиотеке, причем смесь фрагментированных нуклеиновых кислот содержит горячую точку для неслучайной фрагментации (ГТНСФ), и указанный зонд покрывает указанную ГТНСФ, d) выделения из смеси одной или более фрагментированных нуклеиновых кислот, которые являются связанными с одним или большим количеством зондов, e) амплификации и секвенирования обогащенной библиотеки, f) определения размера фрагментированной нуклеиновой кислоты и/или g) определения начального и/или конечного положения фрагментированной нуклеиновой кислоты и h) идентификации происхождения фрагмента нуклеиновой кислоты путем применения информации из этапов (f) и/или (g), что позволяет определить происхождение фрагмента нуклеиновой кислоты и/или выявить его.



A1

202191888

202191888

A1

СПОСОБ ОПРЕДЕЛЕНИЯ ПРОИСХОЖДЕНИЯ НУКЛЕИНОВЫХ КИСЛОТ В СМЕШАННОМ ОБРАЗЦЕ

ОБЛАСТЬ ТЕХНИКИ, К КОТОРОЙ ОТНОСИТСЯ ИЗОБРЕТЕНИЕ

5

Настоящее изобретение относится к области биологии, медицины и химии, в частности, к области молекулярной биологии и, более конкретно, к области молекулярной диагностики.

УРОВЕНЬ ТЕХНИКИ

10

Обнаружение внеклеточной фетальной ДНК (дезоксирибонуклеиновой кислоты) (вкфДНК, cffDNA) в материнской плазме в значительной мере способствовало развитию неинвазивной пренатальной диагностики. Однако, концентрация вкфДНК в материнской плазме варьирует среди индивидуумов, является чрезвычайно низкой, и на ее долю приходится в большинстве случаев 2-19% от общей внеклеточной ДНК (вкДНК) материнской плазмы. Когда доля вкфДНК в материнском кровотоке ниже 4%, достижение достаточной точности для неинвазивного пренатального тестирования (НИПТ, NIPT) даже при помощи методики секвенирования следующего поколения (ССП, NGS), которая обладает высокой чувствительностью, представляет собой сложную задачу.

20

КРАТКОЕ ОПИСАНИЕ ИЗОБРЕТЕНИЯ

Эукариотические геномы организованы в хроматин, который не только обеспечивает возможность уплотнения ДНК, но также регулирует метаболизм ДНК (репликацию, транскрипцию, репарацию, рекомбинацию). Таким образом, задачей в настоящее время является понимание того, (i) каким образом функциональные домены хроматина устроены в ядре, (ii) каким образом структурахроматина/информация в хроматине являются динамичными на протяжении сборки, разборки, модификаций и механизмов ремоделирования, и (iii) каким образом эти события участвуют в развитии, прогрессировании и рецидиве заболевания и/или способствуют им. Понимание этих событий обеспечит возможность идентифицировать новые механизмы прогрессирования заболевания и новые мишени для терапевтического воздействия, а также контролировать эффект терапевтических молекул. Было доказано, что сигнатуры структуры хроматина у эукариотических организмов, в частности, расположение нуклеосом, можно применять для идентификации редких фрагментов нуклеиновой кислоты в сложных смесях, которые

присутствуют в эукариотических организмах (Heitzer E. и соавт. Nat. Rev. Genet. 2019 Feb;20(2):71-88). Такие сложные смеси могут представлять собой, например, смесь вкДНК и материнской вкДНК, или ДНК, происходящей из циркулирующих опухолевых клеток (ЦОК, CTCs) или ткани, и ДНК, происходящей из здоровых циркулирующих клеток.

5

В частности, авторы настоящего изобретения обнаружили новый способ выделения и идентификации редких нуклеиновых кислот в смешанных образцах, который задействует новый направленный подход, использующий длинные синтетические захватывающие мишень последовательности (ЗМП, TACS) (зонды), и новую биоинформатику, а также

10 обнаружили, что в областях, перекрытых этими ЗМП (зондами) существуют неслучайные паттерны фрагментации.

Если бы фрагментация являлась случайной, можно было бы с равной степенью вероятности идентифицировать фрагменты ДНК с начальными и/или конечными положениями в любом положении основания, перекрытом ЗМП (зондами). Это привело бы к равномерному

15 покрытию начальными и/или конечными положениями фрагментов по всему участку зонда. Отклонения от такого покрытия иллюстрируют неслучайные положения фрагментации.

С целью идентифицировать такие отклонения осуществили следующее:

- 20 1. Создали вектор геномных координат для всех начальных и/или конечных участков всех фрагментов, которые выравниваются в пределах специфичной для зонда области (например, одного зонда);
2. На основании вектора, полученного на этапе 1, создали плотность начального и/или конечного участка (например, график, на котором ось Y представляет собой частоту встречаемости, а ось X представляет собой координаты, перекрывающие
- 25 один зонд);
3. Отклонение плотности, созданной на этапе 2, от равномерного покрытия, оценивали в качестве основания предполагать неслучайный механизм фрагментации.

30

Обнаружили некоторое количество таких положений, приходящихся на каждую хромосому. Согласно предположениям, механизм, ответственный за увеличенную частоту неслучайных положений фрагментации в определенных областях, представляет собой защиту ДНК посредством нуклеосомы. Таким образом, отклонение от равномерного покрытия может

свидетельствовать об уменьшенном присутствии какого-либо типа нуклеиновой кислоты в таких положениях (например, часто встречающейся нуклеиновой кислоты, которая включает сложную смесь нуклеиновых кислот), что обусловлено защитой, которую обеспечивает расположение нуклеосом, и, в более широком смысле, об увеличенном присутствии других

5 типов нуклеиновой кислоты (например, редкой нуклеиновой кислоты, которая присутствует в сложной смеси нуклеиновых кислот), что позволяет выявлять области с увеличенной частотой неслучайных положений фрагментации (называемых горячими точками для неслучайной фрагментации [ГНСФ, HSNRF]).

10 ГНСФ в настоящей заявке обозначает геномную область, которая содержит на расстоянии менее 300 п.о. (пар оснований, bp) (предпочтительно менее 200 п.о., в более предпочтительном варианте – менее 100 п.о.) предпочтительные сайты, которые дифференцируют два типа тканей, присутствующих в смеси вкДНК, и в которой указанные предпочтительные сайты присутствуют с более высокой частотой в областях ГНСФ, нежели в

15 других не-ГНСФ областях. Предпочтительные сайты в настоящей заявке обозначают геномные основания, для которых частота нахождения конечной точки прочтения достоверно различается (значение p по меньшей мере менее 0,05) для двух типов тканей, присутствующих в смеси вкДНК.

20 Согласно одному варианту реализации, отклонение от равномерного покрытия оценивали путем количественного определения числа мод в распределении, созданном из начальных и/или конечных положений. На Фигуре 1 наглядно проиллюстрирована такая оценка, причем показано такое положение зонда, в котором не были выявлены ГНСФ (Фигура 1A), и в то же время показано такое положение зонда, в котором ГНСФ были выявлены (Фигура 1B).

25 Следует отметить выраженные различия в плотностях распределения начальных и/или конечных положений. ГНСФ демонстрирует бимодальное распределение. Среднему специалисту в данной области техники будет понятно, что существует множество способов выявления таких явлений, которые включают, среди прочего, количественное определение максимумов и минимумов распределения. Согласно другому варианту реализации, ГНСФ

30 можно выявлять при помощи кластерного анализа с применением начальных/конечных положений.

Как таковое, согласно первому аспекту, настоящее изобретение относится к способу определения происхождения фрагмента нуклеиновой кислоты или выявления фрагмента

35 нуклеиновой кислоты в смеси фрагментов нуклеиновой кислоты, включающему этапы (а)

обеспечения смеси фрагментированных нуклеиновых кислот, которые происходят из эукариотического организма, (b) приготовления библиотеки секвенирования из смеси фрагментированных нуклеиновых кислот; (c) гибридизации одного или более длинных зондов с по меньшей мере одним участком в указанной библиотеке, причем смесь

5 фрагментированных нуклеиновых кислот содержит горячие точки для неслучайной фрагментации (ГТНСФ), и указанный зонд покрывает указанную ГТНСФ, (d) выделения из смеси одной или более фрагментированных нуклеиновых кислот, которые являются

гибридизированными с одним или большим количеством зондов; (e) амплификации и секвенирования библиотеки; (f) определения размера фрагментированной нуклеиновой

10 кислоты и/или (g) определения начального и/или конечного положения фрагментированной нуклеиновой кислоты и (h) идентификации происхождения фрагмента нуклеиновой кислоты путем применения информации из этапов (f) и/или (g), что позволяет вычислить вероятность в отношении происхождения фрагмента нуклеиновой кислоты или выявить фрагмент нуклеиновой кислоты.

15

Указанный выше способ предназначен для выявления геномных областей, которые с высокой степенью вероятности являются горячими точками (много сайтов на близком расстоянии) предпочтительных/дифференцирующих сайтов между двумя типами тканей в смеси фрагментов вкДНК.

20 Согласно второму аспекту настоящего изобретения, предложен способ выделения одного или более фрагментов нуклеиновой кислоты из смеси фрагментов нуклеиновых кислот, который включает этапы:

- a. обеспечения смеси фрагментированных нуклеиновых кислот, предпочтительно ДНК, которые происходят из эукариотического организма;
- 25 b. гибридизации одного или более зондов с по меньшей мере одним участком во фрагментах нуклеиновой кислоты, где располагается горячая точка для неслучайной фрагментации (ГТНСФ), или
- c. амплификации одного или более участка из фрагментов нуклеиновой кислоты, причем праймеры для амплификации примыкают к горячей точке для
- 30 неслучайной фрагментации (ГТНСФ).

Согласно третьему аспекту настоящего изобретения, предложен набор для определения происхождения фрагмента нуклеиновой кислоты в смеси фрагментов нуклеиновой кислоты для применения в способе согласно первому аспекту, содержащий:

35

- a. зонды, которые гибридизируются с по меньшей мере одним участком во фрагментах нуклеиновой кислоты, причем указанный по меньшей мере один участок частично или полностью включает указанный фрагмент нуклеиновой кислоты, и, необязательно,
- 5 b. реагенты и/или программное обеспечение для реализации способа определения и/или выявления.

КРАТКОЕ ОПИСАНИЕ ФИГУР

10 На Фигуре 1 показан пример распределения начальных и/или конечных положений фрагмента нуклеиновой кислоты по всем зондам, причем на (А) плотность является унимодальной и равномерной по всем координатам зонда, в то время как на (В) плотность является бимодальной по всем координатам зонда, тем самым указывая на присутствие ГНСФ. Следует отметить различия в распределении плотностей начальных и/или конечных

15 положений.

На Фигуре 2 показано обогащение находящейся в меньшинстве фракции смешанного образца, когда образец подвергают способу ГНСФ на коротких фрагментах, описанному в настоящей заявке, и вслед за этим проводят оценку присутствия находящейся в меньшинстве фракции. Каждая точка на данной фигуре отображает смешанный образец, в котором

20 находящаяся в меньшинстве фракция представляет собой ДНК, происходящую от плода, а мажоритарная фракция представляет собой ДНК, происходящую от матери. На оси X показана оценка находящейся в меньшинстве фракции перед применением способа ГНСФ на коротких фрагментах, а на оси Y показана оценка находящейся в меньшинстве фракции после применения данного способа. Результаты иллюстрируют увеличение находящейся в

25 меньшинстве фракции, что указывает на присутствие большего количества такой ДНК.

ПОДРОБНОЕ ОПИСАНИЕ ИЗОБРЕТЕНИЯ

Как таковое, согласно первому аспекту, настоящее изобретение относится к способу

30 определения происхождения фрагмента нуклеиновой кислоты или выявления фрагмента нуклеиновой кислоты в смеси фрагментов нуклеиновой кислоты, включающему этапы: (а) обеспечения смеси фрагментированных нуклеиновых кислот, которые происходят из эукариотического организма, (b) приготовления библиотеки секвенирования из смеси фрагментов нуклеиновой кислоты; (с) гибридизации одного или более зондов с по меньшей

35 мере одним участком в указанной библиотеке, причем смесь фрагментированных

нуклеиновых кислот содержит горячие точки для неслучайной фрагментации (ГНСФ), и указанный зонд покрывает указанную ГНСФ, (d) выделения из смеси одной или более фрагментированных нуклеиновых кислот, которые связаны одним или большим количеством зондов; (e) амплификации и секвенирования библиотеки; (f) определения размера фрагментированных нуклеиновых кислот и/или (g) определения начального и/или конечного положения фрагментированных нуклеиновых кислот и (h) идентификации возможного происхождения фрагментированной нуклеиновой кислоты путем применения информации из этапов (f) и/или (g), что позволяет определить возможное происхождение фрагмента нуклеиновой кислоты.

В контексте настоящего изобретения этап (c) определяет гибридизацию указанной библиотеки секвенирования с одним или большим количеством зондов, которые покрывают горячие точки для неслучайной фрагментации (ГНСФ), причем указанные области ГНСФ являются областями, имеющими на расстоянии менее 100-300 пар оснований предпочтительные сайты, которые дифференцируют два типа тканей, присутствующих в смеси вкДНК, и где указанные предпочтительные сайты присутствуют с более высокой частотой в областях ГНСФ, нежели в других не-ГНСФ областях. Предпочтительными сайтами в настоящей заявке называют геномные основания, на которых частота нахождения конечной точки прочтения достоверно различается для двух типов тканей, присутствующих в смеси вкДНК.

Указанный выше способ предназначен для выявления геномных областей, которые с высокой степенью вероятности являются горячей точкой (много сайтов на близком расстоянии) предпочтительных и/или дифференцирующих сайтов между двумя типами тканей в смеси фрагментов вкДНК.

ГНСФ в настоящей заявке обозначает геномную область, которая содержит на расстоянии менее 300 п.о. (предпочтительно менее 200 п.о., в более предпочтительном варианте – менее 100 п.о.) предпочтительные сайты, которые дифференцируют два типа тканей, присутствующих в смеси вкДНК, и в при этом указанные сайты присутствуют с более высокой частотой в областях ГНСФ, нежели в других не-ГНСФ областях. Предпочтительными сайтами в настоящей заявке называют геномные основания, на которых частота нахождения конечной точки прочтения достоверно различается (значение p по меньшей мере менее 0,05) для двух типов тканей, присутствующих в смеси вкДНК.

Смесь фрагментов нуклеиновой кислоты в настоящей заявке предпочтительно выделяют из образца, полученного от эукариотического организма, предпочтительно от примата, более предпочтительно – от человека.

- 5 В контексте настоящего изобретения термин “зонд” относится к синтетическим захватывающим мишень последовательностям (ЗМП).

В контексте настоящего изобретения выражения “фрагменты нуклеиновой кислоты” и “фрагментированные нуклеиновые кислоты” можно применять как взаимозаменяемые.

10

Согласно предпочтительному варианту реализации способа в соответствии с настоящим изобретением, фрагменты нуклеиновой кислоты представляют собой циркулирующую внеклеточную ДНК или РНК (рибонуклеиновую кислоту).

- 15 Согласно одному варианту реализации, образец ДНК представляет собой образец материнской плазмы, который содержит материнскую ДНК и внеклеточную фетальную ДНК (вкфДНК).

- Согласно другому варианту реализации, образец ДНК (например, смесь фрагментов нуклеиновой кислоты) содержит внеклеточную опухолевую ДНК (вкОДНК). Согласно одному варианту реализации, образец ДНК выбран из группы, которая состоит из образца плазмы, образца мочи, образца мокроты, образца спинномозговой жидкости, образца асцитической жидкости и образца плевральной жидкости от субъекта, имеющего опухоль или у которого предполагается ее наличие. Согласно одному варианту реализации, образец ДНК происходит из образца ткани субъекта, имеющего опухоль или у которого предполагается ее наличие.
- 20
- 25

- Способы по настоящему изобретению можно применять с разнообразными биологическими образцами. По существу, любой биологический образец, который содержит генетический материал, например, РНК или ДНК, а в частности, внеклеточную ДНК (вкДНК), можно применять в качестве образца в способах, обеспечивающих возможность генетического анализа РНК или ДНК в нем. Например, согласно одному варианту реализации, образец ДНК представляет собой образец плазмы, который содержит внеклеточную ДНК (вкДНК). В частности, образец ДНК для пренатального тестирования содержит фетальную ДНК (например, внеклеточную фетальную ДНК). Согласно одному варианту реализации, образец для НИПТ представляет собой смешанный образец, содержащий как материнскую ДНК, так и
- 30
- 35

фетальную ДНК (например, внеклеточную фетальную ДНК (вкфДНК)), такой как образец материнской плазмы, полученный из материнской периферической крови. В случае смешанных образцов материнской/фетальной ДНК, образец обычно представляет собой образец материнской плазмы, хотя можно применять и другие источники тканей, которые содержат как материнскую, так и фетальную ДНК. Термин “смешанный образец” в настоящем описании относится к смеси по меньшей мере двух биологических образцов, которые происходят из различных источников, например, образцы материнской/фетальной ДНК.

В зависимости от обстоятельств, понятие “биологический образец” охватывает: эмбриональную ДНК и материнскую ДНК, происходящую из опухоли ДНК и не происходящую из опухоли ДНК, ДНК патогена и ДНК хозяина, а также ДНК, происходящую из трансплантированного органа, и ДНК, происходящую от хозяина.

Исходя из вышеуказанного, в контексте неинвазивной диагностики образец представляет собой смешанный образец, причем указанный смешанный образец выбран из группы, которая состоит из (i) эмбриональной ДНК и материнской ДНК, (ii) происходящей из опухоли ДНК и не происходящей из опухоли ДНК, (iii) ДНК патогена и ДНК хозяина и (iv) ДНК, происходящей из трансплантированного органа, и ДНК, происходящей от хозяина.

Согласно одному варианту реализации, вклад фетального генетического материала является меньшим, чем материнский вклад. Согласно одному варианту реализации, средняя длина фрагментов, которые происходят из плаценты, составляет < 140-150 пар оснований, а средняя длина фрагментов, которые происходят от матери, составляет ≥ 160 пар оснований. Вклад фетального материала можно оценить с применением информации о различиях генетических локусов, которые существуют между фетальным и материнским геномами, таких как, среди прочего, однонуклеотидные полиморфизмы (ОНП, SNPs).

Согласно одному варианту реализации, для оценки фетального вклада применяют значения (0-50%) частоты минорного аллеля (ЧМА, MAF). Согласно другому варианту реализации, все выявленные ОНП задействуют с целью оценки вклада фетального компонента.

Материнскую плазму можно получить из образца периферической цельной крови беременного субъекта, а плазму можно получать посредством стандартных способов. Всего лишь 1-4 мл плазмы достаточно для обеспечения подходящего материала ДНК для анализа в соответствии со способом, раскрытым в настоящем изобретении. Затем можно выделить из

образца общую внеклеточную ДНК с применением стандартных методов, неограничивающие примеры которых включают протокол QIAasympyphony (компания QIAGEN), подходящий для выделения внеклеточной фетальной ДНК, или любой другой ручной или автоматизированный способ экстракции, подходящий для выделения внеклеточной ДНК.

5

Согласно еще одному варианту реализации, предназначенному для онкологических целей, образец представляет собой биологический образец, полученный от субъекта, имеющего опухоль или у которого предполагается ее наличие. Согласно одному варианту реализации, образец ДНК содержит внеклеточную опухолевую ДНК (вкоДНК). Согласно другому варианту реализации, образец представляет собой мочу, мокроту, асцитическую, спинномозговую жидкость или плевральный выпот субъекта. Согласно другому варианту реализации, онкологический образец представляет собой образец плазмы субъекта, полученный из периферической крови субъекта. Таким образом, образец может представлять собой жидкий биопсийный образец, который получают неинвазивным способом из образца крови субъекта, тем самым обеспечивая потенциальную возможность раннего выявления рака, которое предшествует развитию поддающейся выявлению или пальпируемой опухоли, или возможность мониторинга прогрессирования заболевания, лечения заболевания или рецидива заболевания.

Термин “субъект” в контексте настоящего изобретения относится к животным, предпочтительно млекопитающим, и, в более предпочтительном варианте, к людям. Упоминаемый в настоящем описании “субъект” является беременным субъектом и, следовательно, предпочтительно субъектом женского пола. Беременный субъект может находиться на любом сроке беременности. “Субъект” может забеременеть естественным способом или посредством искусственных методов. Термин “субъект” в настоящей заявке также относится к субъекту, который страдает от опухоли или у которого подозревают ее наличие. Указанный субъект может подвергаться трансплантации органа или перенести инфекцию, вызываемую патогеном, после трансплантации или независимо от трансплантации.

30

Для приготовления биологического образца ДНК обычно выделяют с применением стандартных методов, известных в данной области техники, неограничивающим примером которых является протокол QIAasympyphony (компания QIAGEN).

После выделения внеклеточную ДНК образца применяют для конструирования библиотеки секвенирования, чтобы сделать образец совместимым с методикой нисходящего секвенирования, такой как секвенирование следующего поколения. Обычно оно включает лигирование адаптеров с концами фрагментов внеклеточной ДНК. Наборы для приготовления библиотеки секвенирования являются коммерчески доступными или их можно разработать.

В предпочтительном варианте в способе согласно настоящему изобретению первый и второй фрагменты нуклеиновой кислоты выбраны из групп, которые включают:

- i. эмбриональную ДНК и материнскую ДНК,
- ii. происходящую из опухоли ДНК и не происходящую из опухоли ДНК,
- iii. ДНК патогена и ДНК хозяина,
- iv. ДНК, происходящую из трансплантированного органа, и ДНК, происходящую от хозяина.

Термины “плод” и “эмбрион” в контексте настоящего изобретения применяют как взаимозаменяемые.

ГНСФ обычно ассоциирована с двумя концами фрагмента, одна ассоциирована с 5'-концом фрагментированной нуклеиновой кислоты и одна – с 3'-концом фрагментированной нуклеиновой кислоты.

Согласно одному варианту реализации, зонд предпочтительно перекрывает сайт ГНСФ таким образом, чтобы только 5'-конец фрагментированной нуклеиновой кислоты был захвачен зондом.

Согласно другому варианту реализации, зонд перекрывает сайт ГНСФ таким образом, чтобы только 3'-конец внеклеточных нуклеиновых кислот, возникающих из ГНСФ, мог связываться с зондом.

Согласно другому предпочтительному варианту реализации, зонд перекрывает оба сайта ГНСФ, ассоциированных с фрагментированной нуклеиновой кислотой, таким образом, чтобы как 5'-конец, так и 3'-конец внеклеточной нуклеиновой кислоты, ассоциированной с данным сайтом ГНСФ, были захвачены зондом.

Согласно другому варианту реализации, применяют смеси из указанных выше.

В идеальном варианте, в способе согласно настоящему изобретению, захватывающие мишень последовательности (ЗМП) (зонды) на этапе (с) описания данного способа, которое находится в начале этого раздела, представляют собой длинные зонды, и (i) длина каждой из ЗМП (зондов) находится в диапазоне от 100 до 500 пар оснований, (ii) каждый зонд обладает 5'-концом и 3'-концом, (iii) предпочтительно каждый зонд связывается с ГТНСФ на расстоянии по меньшей мере 10 пар оснований, как на 5'-конце, так и на 3'-конце, от областей, несущих вариации числа копий (ВЧК, CNVs), сегментные дубликации или элементы ДНК с повторяющимися последовательностями, и (iv) содержание GC в каждом зонде находится в диапазоне между 19% и 80%.

В целом, этап гибридизации зонда можно осуществлять или перед созданием библиотеки секвенирования, или после того, как библиотека была создана.

Представляющую(ие) интерес область(и) на представляющей(их) интерес хромосоме(ах), где располагаются ГТНСФ, обогащают путем гибридизации пула захватывающих ГТНСФ зондов с библиотекой секвенирования с последующим выделением тех последовательностей в пределах библиотеки секвенирования, которые связываются с зондами. Согласно одному варианту реализации, зонд перекрывает сайт ГТНСФ таким образом, чтобы только 5'-конец фрагментированной нуклеиновой кислоты был захвачен зондом. Согласно другому варианту реализации, зонд перекрывает сайт ГТНСФ таким образом, чтобы только 3'-конец фрагментированных внеклеточных нуклеиновых кислот, возникающих из ГТНСФ, мог связываться с зондом. Согласно другому предпочтительному варианту реализации, зонд перекрывает оба сайта ГТНСФ, ассоциированных с фрагментированной нуклеиновой кислотой, таким образом, чтобы как 5'-конец, так и 3'-конец внеклеточной нуклеиновой кислоты, ассоциированной с данным сайтом ГТНСФ, были захвачены зондом.

Чтобы облегчить выделение желаемых обогащенных последовательностей (ГТНСФ), последовательности зондов обычно модифицируют таким образом, чтобы последовательности, которые гибридизируются с зондами, можно было отделить от последовательностей, которые не гибридизируются с зондами. Обычно этого достигают путем фиксации зондов на подложке. Это обеспечивает возможность физического отделения тех последовательностей, которые связываются с зондами, от тех последовательностей, которые не связываются с зондами. Например, каждую

последовательность в пределах пула зондов можно пометить при помощи биотина, а затем связать пул с шариками, которые покрыты веществом, связывающим биотин, таким как стрептавидин или авидин. Согласно предпочтительному варианту реализации, зонды помечают биотином и связывают с магнитными шариками, покрытыми стрептавидином, тем самым обеспечивая возможность разделения за счет использования магнитного свойства шариков. При этом среднему специалисту в данной области техники будет понятно, что в данной области техники известны и другие системы аффинного связывания, и их можно применять вместо системы “биотин – стрептавидин/авидин”. Например, можно применять систему на основе антител, в которой зонды помечают антигеном и затем связывают с шариками, которые покрыты антителами. Более того, в зонды можно внедрять на одном конце маркерную последовательность и связывать их с подложкой посредством комплементарной последовательности на подложке, которая гибридизируется с маркерной последовательностью. Помимо этого, наряду с магнитными шариками можно применять другие типы подложек, такие как полимерные шарики и тому подобное.

Согласно определенным вариантам реализации, члены библиотеки секвенирования, которые связываются с пулом зондов, являются полностью комплементарными данному зонду. Согласно другим вариантам реализации, члены библиотеки секвенирования, которые связываются с пулом зондов, являются частично комплементарными данному зонду. Например, при определенных обстоятельствах может быть желательно использовать и анализировать данные, полученные от фрагментов ДНК, которые являются продуктами процесса обогащения, но не обязательно принадлежат к представляющим интерес геномным областям (например, такие фрагменты ДНК могут связываться с зондом по причине частичных гомологий) и при секвенировании производили бы очень малое покрытие всего генома на всем протяжении незондовых координат.

После обогащения представляющей(их) интерес последовательности(ей) с применением зондов, с формированием вследствие этого обогащенной библиотеки ДНК с сайтами ГТНСФ, члены обогащенной библиотеки ГТНСФ подвергают элюированию, амплификации и секвенированию с применением стандартных способов, известных в данной области техники. В качестве нормализаторов применяют эталонные (из лейкоцитарной пленки) обработанные ультразвуком образцы, наряду с содержанием GC, картографируемостью и другими корректировками технических артефактов, с целью очистить данные и отфильтровать ложноположительные результаты.

Обычно применяют секвенирование следующего поколения (ССП), хотя можно задействовать и другие технологии секвенирования, которые обеспечивают очень точный подсчет наряду с информацией о последовательности. Таким образом, вместо ССП можно также применять и другие способы точного подсчета, такие как цифровая ПЦР (полимеразная цепная реакция), одномолекулярное секвенирование, нанопоровое секвенирование и микрочипы.

Как продемонстрировано более подробно, ГНСФ, например, точный размер ее фрагмента и сайт разрезания, служит для валидации происхождения нуклеиновой кислоты.

Данное изобретение относится к способу в соответствии с любым из аспектов или вариантов реализации, в котором фрагменты нуклеиновой кислоты, подлежащие выявлению, или происхождение которых предстоит определить, присутствуют в смеси в более низкой концентрации, чем фрагмент нуклеиновой кислоты из того же генетического локуса, но другого происхождения. Это означает, что если выбран определенный локус, то, например, материнская копия будет присутствовать в растворе, который содержит выделенную вкДНК, 100 раз, а копия от плода – только один раз. В случае вкДНК, происходящей от плода, происходящий от плода компонент смешанного образца может обладать диапазоном возможных значений. Например, спектр фетального материала в смешанном образце может находиться в диапазоне от 2% до 30%. Зачастую происходящие от плода фрагменты представляют собой приблизительно 10% от общей ДНК смешанного образца. Что еще более важно, в некоторых композициях смешанных образцов компонент фетальной ДНК может составлять менее 5% от образца. В частности, в некоторых композициях образцов происходящий от плода материал составляет 3%, или менее, от всего образца.

Настоящий способ в особенности подходит для того, чтобы анализировать такие низкие концентрации целевой вкДНК. В способе согласно настоящему изобретению фрагмент нуклеиновой кислоты, подлежащий выявлению, или происхождение которого предстоит определить, а также фрагмент нуклеиновой кислоты из того же генетического локуса, но другого происхождения присутствуют в смеси в соотношении, которое выбрано из группы 1:2, 1:4, 1:10, 1:20, 1:50, 1:100, 1:200, 1:500, 1:1000, 1:2000 и 1:5000. Данные соотношения следует понимать как приблизительные соотношения, что означает плюс/минус 30%, 20% или 10%. Специалисту в данной области техники известно, что такие соотношения не будут возникать в точных численных значениях, приведенных выше. Соотношения означают отношение числа

локус-специфических молекул редкого типа к числу локус-специфических молекул часто встречающегося типа.

Согласно одному варианту реализации, зонды обеспечивают в форме, которая обеспечивает возможность их связывания с подложкой, как, например, биотинилированные зонды. Согласно другому варианту реализации, зонды обеспечивают совместно с подложкой, как, например, биотинилированные зонды, которые обеспечивают совместно с магнитными шариками, покрытыми стрептавидином.

- 10 Согласно конкретному варианту реализации, содержание GC в зондах находится в диапазоне от 10% до 80%, предпочтительно от 15% до 60%, в более предпочтительном варианте от 20% до 50%.

Согласно второму аспекту настоящего изобретения, предложен способ выделения одного или более фрагментов нуклеиновой кислоты из смеси фрагментов нуклеиновой кислоты, включающий этапы:

- a. обеспечения смеси фрагментированных нуклеиновых кислот, предпочтительно ДНК, которые происходят из эукариотического организма;
- b. гибридизации одного или более зондов с по меньшей мере одним участком во фрагментах нуклеиновой кислоты, где располагается горячая точка для неслучайной фрагментации (ГТНСФ), или
- c. амплификации одного или более участка из фрагментов нуклеиновой кислоты, причем праймеры для амплификации примыкают к горячей точке для неслучайной фрагментации (ГТНСФ).

25

Согласно третьему аспекту настоящего изобретения, предложены наборы для осуществления способов, раскрытых в рамках настоящего изобретения, которые содержат:

- a. зонды, которые гибридизируются с по меньшей мере одним участком во фрагменте нуклеиновой кислоты, причем указанный по меньшей мере один участок частично или полностью включает указанный фрагмент нуклеиновой кислоты и, необязательно,
- b. реагенты и/или программное обеспечение для осуществления способа определения и/или выявления.

35

В контексте настоящего изобретения термин “частично” относится к области (участку) из 10, 20, 30 или 40 оснований фрагмента нуклеиновой кислоты от 5'-конца или от 3'-конца. Следовательно, термин “полностью” в настоящей заявке относится к области (участку), которая охватывает 100% фрагмента нуклеиновой кислоты. В соответствии со способом по

5 настоящему изобретению, зонды гибридизируются с по меньшей мере одним участком в пределах фрагмента нуклеиновой кислоты. Однако более чем один участок в пределах одного и того же фрагмента нуклеиновой кислоты может также являться мишенью для зондов.

- 10 Согласно одному варианту реализации, набор содержит контейнер, который состоит из пула зондов и программного обеспечения и инструкций для выполнения данного способа.

Наряду с пулом зондов набор может содержать одно или более из следующего: (i) один или более компонентов для выделения внеклеточной ДНК из биологического образца, (ii) один

15 или более компонентов для приготовления и обогащения библиотеки секвенирования (например, праймеры, адаптеры, буферы, линкеры, модифицирующие ДНК ферменты, ферменты лигирования, полимеразные ферменты, зонды и тому подобные), (iii) один или более компонентов для амплификации и/или секвенирования обогащенной библиотеки и/или (iv) программное обеспечение для выполнения статистического анализа.

20

Определение происхождения первого фрагмента нуклеиновой кислоты в образцах опухолей может быть весьма важным. Авторы настоящего изобретения обнаружили, что различные

ткани обладают различными “сигнатурами”. Таким образом, представляется возможным выявить происхождение первой нуклеиновой кислоты из конкретной ткани, например, из

25 опухолевого образца.

Зонды для выявления опухолевых биомаркеров сконструированы на основании критериев конструирования, описанных в настоящей заявке, а также известных последовательностей генов опухолевых биомаркеров и генетических мутаций в них, ассоциированных с раком.

30 Согласно одному варианту реализации, множество зондов, применяемых в данном способе, связывается со множеством последовательностей представляющих интерес опухолевых биомаркеров. В данной ситуации зонд может располагаться в горячих точках для неслучайной фрагментации, примыкающих к сайту мутации.

Биомаркеры в таком исследовании можно выбирать из группы, которая содержит ABL, AKT, AKT1, ALK, APC, AR, ARAF, ATM, BAP1, BARD1, BCL, BMPR1A, BRAF, BRCA, BRCA1, BRCA2, BRIP1, CDH1, CDKN, CHEK2, CTNNB1, DDB2, DDR2, DICER1, EGFR, EPCAM, ErbB, ErcC, ESR1, FANCA, FANCB, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCL, FANCM, FBXW7, FGFR, FLT, FLT3, FOXA1, FOXL2, GATA3, GNA11, GNAQ, GNAS, GREM1, HOX, HOXB13, HRAS, IDH1, JAK, JAK2, KEAP1, KIT, KRAS, MAP2Ks, MAP3Ks, MET, MLH1, MPL, MRE11A, MSH2, MSH6, MTOR, MUTYH, NBN, NPM1, NRAS, NTRK1, PALB2, PDGFRs, PI3KCs, PMS2, POLD1, POLE, POLH, PTEN, RAD50, RAD51C, RAD51D, RAF1, RB1, RET, RUNX1, SLX4, SMAD, SMAD4, SMARCA4, SPOP, STAT, STK11, TP53, VHL, XPA, XPC и их комбинации.

10

Согласно одному варианту реализации данного способа, после приготовления библиотеки и обогащения ее представляющими интерес последовательностями посредством гибридизации с зондами выполняют последующий этап амплификации обогащенной библиотеки в присутствии блокирующих последовательностей, которые ингибируют амплификацию последовательностей дикого типа. Таким образом, амплификацию смещают в сторону амплификации мутантных последовательностей опухолевых биомаркеров.

15

Согласно другому варианту реализации, для того, чтобы пометить каждый фрагмент ДНК или РНК, можно применять уникальные молекулярные штрихкоды.

20

Пул зондов, который применяют в способе выявления опухолевых биомаркеров, может содержать любую из конструктивных особенностей, описанных в настоящей заявке применительно к конструкции зондов.

25

Подходящие подходы к статистическому анализу для применения с пренатальными образцами и онкологическими образцами, позволяющие выявить биомаркеры ГТНСФ, описаны далее в разделе “Примеры”.

30

Конкретные типы рака можно охарактеризовать фрагментами ДНК в плазме, которые имеют меньший размер, чем ожидаемый размер фрагментов ДНК, происходящих из здоровых тканей, или ассоциировать с ними. Та же самая гипотеза является справедливой в отношении фрагментов, которые происходят из плаценты/плода. В частности, происходящие из плаценты фрагменты обычно обладают меньшим размером в сопоставлении с фрагментами, которые происходят из материнских тканей/клеток. Настоящее изобретение обеспечивает возможность для такого исследования новым образом. Зонды позволяют выделять и

35

обогащать редко встречающиеся, более короткие фрагменты. Таким образом, настоящий способ обеспечивает возможность выявления отклонений от нормы, основанного на фрагментах, в смешанных образцах с низким соотношением “сигнал/шум”.

5 В связи с этим, согласно другому варианту реализации настоящего изобретения, предпочтительно, чтобы фрагмент, который подлежит выявлению или определению, обладал большим или меньшим размером, нежели второй фрагмент. В зависимости от ткани происхождения, по всем участкам ГНСФ следует ожидать различные размеры фрагментов. Учитывая то, что клетки каждой ткани являются дифференцированными для выполнения
10 конкретных функций, следует ожидать, что у каждого типа клеток будут наблюдаться различные паттерны генетической активности. Учитывая то, что активация гена является зависимой от третичной структуры ДНК, различные ткани будут обладать несколько различающимися конформациями хроматина, что приводит к различным сайтам разрезания и, следовательно, различным размерам фрагментов. Например, было продемонстрировано,
15 что внеклеточная ДНК плацентарного происхождения, вероятно, будет обладать меньшим размером в сопоставлении с внеклеточной ДНК материнского происхождения.

Описанный в настоящей заявке способ опирается на гипотезу о том, что предпочтительные сайты существуют, но могут различаться в зависимости от биологических образцов.
20 Соотношение “сигнал/шум” на образец дополнительно улучшено посредством разработки способа для реализации оценки ГНСФ в режиме реального времени, для каждого образца, вместо применения в стандартном режиме эталонного набора предпочтительных сайтов (Фигура 2).

25 ПРИМЕРЫ

Далее настоящее изобретение проиллюстрировано при помощи приведенных ниже примеров, которые никоим образом не следует истолковывать в качестве дальнейшего ограничения.

30

Пример 1: Сбор образцов и приготовление библиотеки

Пояснена общая методика этого подхода, основанного на зондах, к обнаружению ГНСФ в целях неинвазивной пренатальной диагностики. В этом примере описаны способы сбора и обработки образца материнской плазмы (который содержит материнскую и фетальную ДНК).

35 Тому же самому подходу можно следовать для обнаружения ГНСФ в других случаях,

полезных с медицинской точки зрения, таких как, среди прочего, онкология, генетические мутации, трансплантация и оценка патогенной нагрузки.

Сбор образцов

- 5 Образцы плазмы получали анонимным образом от беременных женщин после 10-й недели беременности. Протоколы, которые применяли для сбора образцов, были одобрены Национальным комитетом по биоэтике, а от всех участников получили информированное согласие.

10 *Экстракция образцов*

Внеклеточную ДНК экстрагировали из плазмы от каждого индивидуума с применением ручного или автоматизированного способа экстракции, подходящего для выделения внеклеточной ДНК, такого как, например, но без ограничения этим, протокол QIAasympphony, подходящий для выделения вкДНК (компания QIAGEN).

15

Приготовление библиотеки секвенирования

- Экстрагированные внеклеточные ДНК из образцов материнской плазмы применяли для конструирования библиотеки секвенирования. Библиотеку отрицательного контроля экстрагирования приготавливали отдельно для отслеживания каких-либо контаминаций, вносимых на протяжении эксперимента. На первоначальном этапе выступающие 5'- и 3'-
- 20 концы подвергали репарации, в то время как 5'-концы подвергали фосфорилированию. Продукты реакции очищали с применением шариков AMPure XP (компания Beckman Coulter). Вслед за этим с обоими концами ДНК лигировали адапторы секвенирования, после чего производили очистку с применением шариков AMPure XP (компания Beckman Coulter).
- 25 Одноцепочечные разрывы удаляли в процессе реакции заполнения с полимеразой и вслед за этим очищали их с применением шариков AMPure XP (компания Beckman Coulter). Амплификацию библиотеки осуществляли с применением еще одного фермента полимеразы (Koumbaris *и соавт.* (2016) *Clinical chemistry* 62(6):848-855). Конечные продукты библиотеки очищали с применением шариков AMPure XP (компания Beckman Coulter) и
- 30 подвергали измерению посредством спектрофотометрии.

Пример 2: Конструирование и получение захватывающих мишень последовательностей (ЗМП)

- В этом примере описано получение настраиваемых ЗМП (зондов) для выявления ГТНСФ.
- 35 Целевые геномные локусы, которые применяли для конструирования ЗМП, выбирали на

основании содержания в них GC и расстояния, на котором они находятся от повторяющихся элементов (отстоящих как минимум на 50 п.о.). Размеры ЗМП могут варьировать. Согласно одному варианту реализации данного способа, ЗМП обладают размером в диапазоне от 100 до 500 п.о. и их получают согласно приведенному ниже описанию. ЗМП получали

5 посредством полимеразной цепной реакции (ПЦР) с применением полимеразы Taq (термостабильная ДНК-полимераза бактерии *Thermus aquaticus*), праймеров, сконструированных для амплификации целевых локусов, и нормальной ДНК в качестве матрицы. Согласно предпочтительному варианту реализации, ЗМП перекрывает сайт ГТНСФ

10 таким образом, чтобы только 5'-конец фрагментированной нуклеиновой кислоты был захвачен зондом. Согласно другому варианту реализации, ЗМП перекрывает сайт ГТНСФ таким образом, чтобы только 3'-конец внеклеточных нуклеиновых кислот, возникающих из ГТНСФ, мог связываться с зондом. Согласно другому предпочтительному варианту реализации, ЗМП перекрывает оба сайта ГТНСФ, ассоциированных с фрагментированной нуклеиновой кислотой, таким образом, чтобы как 5'-конец, так и 3'-конец внеклеточной

15 нуклеиновой кислоты, ассоциированной с данным сайтом ГТНСФ, были захвачены ЗМП. Продукты ПЦР верифицировали посредством электрофореза в агарозном геле и очищали с применением стандартных наборов для очистки ПЦР-продуктов. Концентрацию измеряли посредством спектрофотометрии.

20 *Пример 3: Гибридизация и амплификация ЗМП*

В этом примере описан способ целенаправленного захвата нуклеиновых кислот посредством гибридизации с применением ЗМП и последующее секвенирование захваченных последовательностей при помощи секвенирования следующего поколения (ССП).

25 *Биотинилирование ЗМП*

ЗМП подготавливали для гибридизации, начиная с “затупления концов”, с последующей очисткой. Затем их лигировали с биотиновым адаптером и очищали. Перед иммобилизацией на магнитных шариках, покрытых стрептавидином, ЗМП подвергали денатурации.

30 *Гибридизация ЗМП*

Амплифицированные библиотеки смешивали с блокирующими олигонуклеотидами, ДНК Cot-1, ДНК из молок лососевых, буфером для гибридизации, блокирующим средством, а затем денатурировали. После денатурации следовала инкубация на протяжении 30 минут при 37°C. Затем полученную смесь добавляли к биотинилированным ЗМП и инкубировали на

протяжении 12-48 часов при 60-70°C. По прошествии инкубации обогащенные образцы промывали согласно предшествующему описанию, а ДНК элюировали посредством нагревания. Элюированные продукты амплифицировали с применением праймеров адаптора с внешней связью. Обогащенные амплифицированные продукты объединяли эквимольным образом и подвергали секвенированию на подходящей платформе.

Если это является целесообразным, амплификацию можно намеренно сместить в сторону амплификации конкретных/желаемых последовательностей. Согласно одному варианту реализации, это выполняется в том случае, когда амплификацию осуществляют в присутствии последовательностей, которые гибридизируются с нежелательными последовательностями, представляющими интерес, и как таковые блокируют действие фермента полимеразы на протяжении процесса. Следовательно, действие фермента амплификации на протяжении процесса является направленным на последовательность, представляющую интерес.

Пример 4: Выявление кандидатов ГТНСФ, основанное на начальных/конечных положениях фрагмента

В этом примере проиллюстрировано выявление ГТНСФ по данным секвенирования образцов материнской плазмы, причем образец плазмы представляет собой смешанный образец ДНК, происходящей от матери и происходящей от плода.

20

Данные секвенирования следующего поколения (ССП) обрабатывали с применением способов, общепринятых для специалистов в данной области техники. В кратком изложении, данные ССП подвергали процедуре демультимплексирования с приписыванием секвенированных фрагментов ДНК их образцам, идентифицированным посредством индексной последовательности, и формировали файлы FASTQ.

25

Согласно предпочтительному варианту реализации, файлы FASTQ для каждого образца не выравнивали по варианту эталонного генома человека. Информацию о размере и/или классе фрагментов получали с применением способа выявления количества перекрывания/гомологии спаренных прочтений или длины одиночных прочтений (при размере более 150 п.о.). Идентичные прочтения удаляют и осуществляют или сборку заново, или сопоставление с заранее определенными мишенями коротких последовательностей. Среднему специалисту в данной области техники будет понятно, что методы, основанные на выравнении целого генома, не обладают высокой точностью в областях с низкой идентичностью последовательностей и исходят из предположения о полногеномном

35

линейном порядке гомологии, что не всегда является верным в случае перестроек последовательности.

Разработанный способ осуществляет оценку ГТНСФ в режиме реального времени на образец, вместо применения в стандартном режиме эталонного набора предпочтительных сайтов, что приводит к очень высокому соотношению “сигнал/шум” на образец (Фигура 2).

Согласно другому варианту реализации, файлы FASTQ для каждого образца выравнивали по варианту эталонного генома человека с применением алгоритма выравнивания Burrows-Wheeler, который формирует содержащий информацию файл в текстовом формате; файл SAM (карта выравнивания последовательностей). Перед осуществлением дальнейшей обработки файл SAM конвертировали в двоичный формат с получением файла BAM. Если это являлось целесообразным, данные, происходящие из одного и того же образца, но найденные по различным файлам, объединяли. Там, где это являлось необходимым, данные проверяли на предмет качества секвенирования и присутствия дублированных фрагментов ДНК, чтобы создать конечный файл BAM. Данные, которые относились к глубине прочтения, размеру фрагментов и информации о последовательности, извлекали из конечного файла BAM.

Среднему специалисту в данной области техники будет понятно, что для реализации вышеупомянутых процедур существует множество хорошо известных инструментов, находящихся в свободном доступе.

Выявление горячей точки для неслучайной фрагментации (ГТНСФ)

Допущение для обнаружения ГТНСФ заключается в том, что если фрагментация является случайной, и при отсутствии стандартной ошибки секвенирования можно было бы с равной степенью вероятности найти фрагменты ДНК с начальными и/или конечными положениями на всех координатах, которые содержат область, перекрытую зондом. Более конкретно, строят распределение начальных/конечных координат в пределах зонда, и после введения поправки на секвенирование и на стандартную ошибку содержания GC с применением набора случайным образом фрагментированных образцов, согласно предположениям, любые отклонения от унимодального распределения свидетельствуют о неслучайной фрагментации.

Чтобы оценить распределение начальных/конечных положений в пределах зонда, применяли непараметрическую ядерную оценку плотности с ядром Епанечникова и пропускной способностью, которую оценивали с применением, среди прочего, подхода Sheather и Jones. Вслед за этим полученную информацию оценивали, чтобы определить, существует ли отклонение от унимодального распределения. Учитывая то, что полученная информация описывает распределение начальных и/или конечных положений фрагментов ДНК в пределах зонда, соответствующие количественные показатели распределения можно применять для оценки отклонения от равномерного покрытия. Неограничивающим примером такой оценки является идентификация мод распределения. Мультимодальное распределение представляет собой пример неравномерного распределения, которое является релевантным для настоящего изобретения.

На Фигуре 1 наглядно проиллюстрирована такая оценка. На панели (А) показано положение зонда, в котором не были выявлены ГТНСФ, в то время как на панели (В) показано положение зонда, в котором была выявлена неслучайная фрагментация. Следует отметить различия в распределении плотностей начальных и/или конечных положений. В представленном примере зонд, ассоциированный с неслучайной фрагментацией, иллюстрирует бимодальное распределение, которое дает основания предполагать наличие защищенных областей; областей, в которых имела место уменьшенная фрагментация и, таким образом, неслучайная фрагментация. Это делает данный зонд подходящим кандидатом для существования ГТНСФ.

Специалисту в данной области техники будет понятно, что оценка мод распределения представляет собой способ оценки отклонения от равномерного и/или ожидаемого покрытия, приводимый в качестве примера. Как таковой, пример мультимодального распределения не подлежит истолкованию в качестве ограничивающего объем оценки. Существуют и другие способы выявления таких явлений, которые включают, среди прочего, количественное определение максимумов и минимумов распределения.

Пример 5: Выявление ГТНСФ, основанное на размере фрагмента

Одним из свойств ГТНСФ является размер каждого фрагмента нуклеиновой кислоты, который возникает из таких горячих точек. Размер может представлять собой свойство, которое ассоциировано с тканью происхождения, такой как, например, раковая ткань или относящийся к плоду орган, такой как плацента. Фрагменты нуклеиновой кислоты, которые происходят из раковой ткани, и фрагменты нуклеиновой кислоты, которые происходят из ткани плода, такой как плацента, обычно обладают меньшим размером в сопоставлении с

фрагментами нуклеиновой кислоты, которые обнаруживаются в кровотоке в организме субъекта-человека.

С учетом того, что паттерны фрагментации ДНК ассоциированы с определенной тканью происхождения, сайты ГНСФ должны быть также ассоциированы с определенным размером фрагментов и распределением размеров фрагментов. В случае фетального материала, который происходит из плаценты, горячую точку можно описать как геномную координату, где фрагмент завершает кластер, и размер фрагмента, ассоциированный с положением данного кластера, обычно обладает меньшим значением в сопоставлении с другими фрагментами нуклеиновой кислоты, которые обнаруживают в кровотоке в плазме беременной женщины. Чтобы оценить это, можно уточнить по поводу положения каждого основания в составе зонда, начинается/заканчивается ли фрагмент в этом конкретном положении, и если такой фрагмент существует, то приступить к определению размера данного фрагмента. Размер фрагмента можно получить или перед этапом выравнивания, или после него.

Настоящее изобретение обеспечивает возможность другого определения для малого/большого фрагмента на каждый образец в силу того, что позволяет пороговой величине, задаваемой размеру фрагмента, варьировать от образца к образцу (в зависимости от фетальной фракции и распределения размеров фрагментов). В зависимости от ткани происхождения, по всем участкам ГНСФ следует ожидать различные размеры фрагментов. Учитывая то, что клетки каждой ткани являются дифференцированными для выполнения конкретных функций, следует ожидать, что у каждого типа клеток будут наблюдаться различные паттерны генетической активности. Учитывая то, что активация гена является зависимой от третичной структуры ДНК, различные ткани будут обладать несколько различающимися конформациями, что приводит к различным сайтам разрезания и, следовательно, различным размерам фрагментов.

Согласно одному варианту реализации, размер всех фрагментов, которые начинаются и/или заканчиваются на определенной координате, впоследствии оценивают с получением статистической меры, которая определяет ожидаемый размер фрагмента. Согласно одному варианту реализации, размер фрагмента оценивают с применением статистических свойств распределения размеров фрагментов, таких как среднее значение, срединное значение или другой квантиль, или моды распределения. Специалисту в данной области техники будет понятно, что для оценки распределения можно применять и другие способы. Если

статистическая мера находится ниже или выше данной нижней или верхней границы, то данное расположение основания можно рассматривать как кандидата горячей точки, ассоциированной с нуклеиновой кислотой плацентарного (фетального) происхождения.

Согласно одному варианту реализации данного способа, обогащенный образец подвергают секвенированию с применением способа секвенирования 2x75 п.о., обеспечивая сбор 5 начальных/конечных положений секвенируемого фрагмента как с 5'-конца, так и с 3'-конца, а также получение достаточного количества информации секвенирования данного фрагмента, чтобы обеспечить возможность выравнивания заново/самовыравнивания. Согласно указанному варианту реализации, секвенированные фрагменты группируют в соответствии с размером, на основании протяженности перекрывания прочтений их 10 спаренных концов или отсутствия перекрывания. Таким образом, каждый фрагмент классифицируют в двух группах, короткие (Группа 1) и длинные (Группа 2). ГНСФ идентифицируют при помощи сходства последовательностей по меньшей мере 20 наиболее удаленных от середины нуклеотидов у фрагментов в Группе 1. Фрагменты из Группы 2, концы 15 которых перекрываются с идентифицированными ГНСФ, впоследствии применяют для формирования Группы 3. Затем все фрагменты из Группы 1 и все фрагменты из Группы 3 сохраняют для последующего исследования. Указанные фрагменты вкДНК Группы 1 являются более короткими по размеру и обогащены вкДНК. Указанные фрагменты вкДНК Группы 3 являются более длинными по размеру, но также обогащены вкДНК, поскольку их концы 20 перекрываются с ГНСФ. Фрагменты внеклеточной ДНК Группы 2, концы которых не перекрываются с ГНСФ, с большей степенью вероятности являются происходящими от матери, и их отбрасывают из всех последующих исследований.

Согласно другому варианту реализации, предварительно установленное значение пороговой 25 величины составляет 150 пар оснований. Затем положение каждого основания в составе зонда оценивают, чтобы установить, начинается/заканчивается ли фрагмент в таком положении, и если это так, то размер данного фрагмента сопоставляют с пороговой величиной, чтобы определить, является ли фрагмент большим или малым. Регистрацию повторяют для всех положений зонда с целью получения ассоциации между положением 30 зонда и числом малых фрагментов, которые происходят из такого положения. Положения, ассоциированные с более значительным числом таких малых фрагментов в сопоставлении с другими положениями в пределах зонда, являются кандидатами положений горячих точек, ассоциированных с нуклеиновыми кислотами плацентарного (фетального) происхождения. Среднему специалисту в данной области техники будет понятно, что неограничивающие 35 примеры статистических мер и способов кластерного анализа, которые можно применять для

идентификации кандидатов горячих точек, включают первый и второй моменты распределения.

Зонды могут нести множество таких горячих точек коротких или длинных фрагментов. В смешанном материнском образце можно применять способ, задействующий ядерную оценку плотности, для идентификации ГНСФ, которые состоят из фрагментов, в первую очередь происходящих от матери, и ГНСФ из фрагментов, в первую очередь происходящих от плода. Гипотеза, которую предстоит проверить, заключается в том, что если зонд несет происходящие от матери ГНСФ, например, области, в которых естественный паттерн фрагментации изменен в связи с занятием нуклеосомами, то должна быть возможность с большей легкостью выявить ГНСФ, ассоциированные с фрагментами ДНК фетального происхождения (короткие), в областях, где сайты разрезания происходящих от матери фрагментов являются менее распространенными. Если это является верным, то зонды с происходящими от матери ГНСФ должны статистически обладать более значительным числом ГНСФ, ассоциированных с короткими фрагментами, в сопоставлении с зондом, который не несет ГНСФ фетального происхождения. Данную гипотезу можно проверить с применением анализа соотношения шансов, L_H :

$$L_H = \frac{\left(\frac{p_1}{1-p_1}\right)}{\left(\frac{p_2}{1-p_2}\right)}$$

20

где p_1 представляет собой вероятность встретить горячую точку в составе зонда, который является кандидатом на ГНСФ представляющей интерес ткани (как представлено на Фигуре 1(B) для случая с фетальными ГНСФ), а p_2 представляет собой вероятность встретить горячую точку в составе зонда, который не является кандидатом на ГНСФ представляющей интерес ткани (как представлено на Фигуре 1(A). Значения, при которых $L_H > 1$, свидетельствуют об увеличенной представленности таких горячих точек в составе зондов-кандидатов ГНСФ. При задействовании когорты из 96 образцов получили значение $L_H = 1,27$ (95% CI: 1,08 – 1,50), которое иллюстрирует более высокую вероятность получить представляющие интерес горячие точки у зондов-кандидатов ГНСФ.

30

Пример 6: Подтверждение сайтов ГНСФ

Подтверждение сайтов ГНСФ может происходить посредством количественного определения фракций смешанного образца. Поскольку сайты ГНСФ ассоциированы с тканью

происхождения, количественное определение фракций смешанного образца с применением сайтов ГНСФ и сайтов не-ГНСФ должно обеспечивать различающиеся показатели. В частности, в случае применения ГНСФ на сайтах коротких фрагментов, вклад ткани, ассоциированной с определенной тканью происхождения, должен иметь большее значение

5 в сопоставлении с уровнем, полученным при осуществлении такого же измерения с применением других сайтов.

На Фигуре 2 показаны оценки фетальной фракции с применением информации ОНП из фрагментов, которые происходят из сайтов ГНСФ (ось Y), и проведено сопоставление их с

10 оценками фетальной фракции в том случае, когда информация ОНП получена из всех фрагментов, независимо от их происхождения (ось X). Каждая точка представляет образец. Как видно на фигуре, все точки данных располагаются над линией $x=y$. Это указывает на то, что оценки фетальной фракции из сайтов ГНСФ приводят к более высоким значениям, что указывает на увеличенное присутствие фетального материала и таким образом

15 подтверждает существование сайтов ГНСФ.

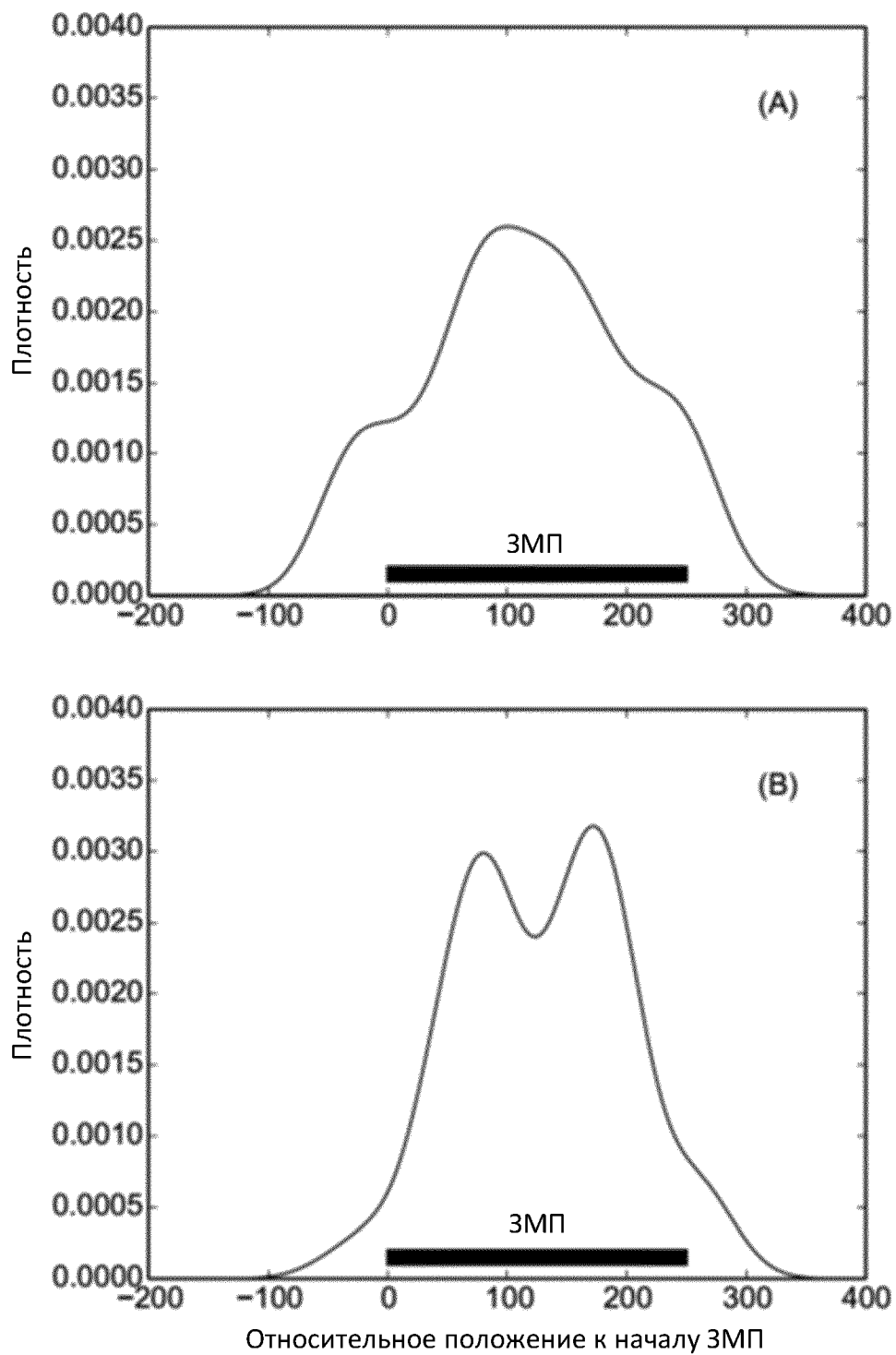
Формула изобретения

1. Способ определения происхождения фрагмента нуклеиновой кислоты или выявления фрагмента нуклеиновой кислоты в смеси фрагментов нуклеиновой кислоты, включающий этапы
 - a. обеспечения смеси фрагментированных нуклеиновых кислот, которые происходят из эукариотического организма,
 - b. приготовления библиотеки секвенирования из указанной смеси фрагментированных нуклеиновых кислот,
 - c. гибридизации одного или более зондов с по меньшей мере одним участком в указанной библиотеке, причем указанная смесь фрагментированных нуклеиновых кислот содержит горячую точку для неслучайной фрагментации (ГНСФ), и указанный зонд покрывает указанную ГНСФ,
 - d. выделения из указанной смеси одной или более фрагментированных нуклеиновых кислот, которые являются связанными с указанным одним или более зондами,
 - e. секвенирования обогащенной библиотеки, причем коэффициент дубликации у указанной библиотеки секвенирования из матричных фрагментов ДНК составляет более 5%,
 - f. определения принадлежности фрагментированной нуклеиновой кислоты к классу малых или больших без выравнивания на каком-либо эталонном геноме, то есть, но не ограничиваясь указанным, с применением сходства последовательностей секвенированных прочтений, и/или,
 - g. определения последовательности нуклеотидов, по меньшей мере 20 п.о., наиболее удаленных от середины указанной фрагментированной нуклеиновой кислоты и
 - h. идентификации происхождения фрагмента нуклеиновой кислоты путем применения информации из этапов (f) и/или (g), с определением тем самым происхождения указанного фрагмента нуклеиновой кислоты или с выявлением тем самым указанного фрагмента нуклеиновой кислоты без необходимости в эталонном и калибровочном значениях.

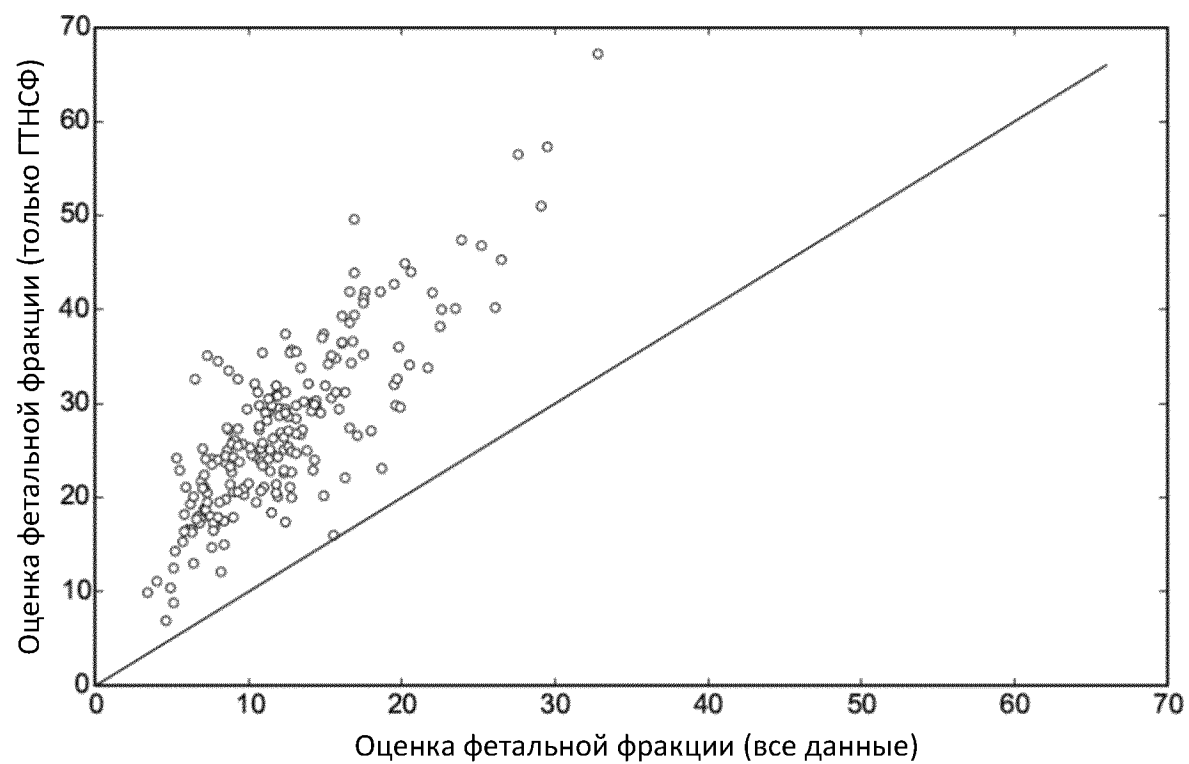
2. Способ по п. 1, отличающийся тем, что указанный фрагмент нуклеиновой кислоты представляет собой циркулирующую внеклеточную ДНК или РНК.
3. Способ по любому из пп. 1-2, отличающийся тем, что указанные фрагменты нуклеиновой кислоты выбраны из групп, которые содержат:
 - i. эмбриональную ДНК и материнскую ДНК,
 - ii. происходящую из опухоли ДНК и не происходящую из опухоли ДНК,
 - iii. ДНК патогена и ДНК хозяина,
 - iv. ДНК, происходящую из трансплантированного органа, и ДНК, происходящую от хозяина.
4. Способ по любому из пп. 1-3, отличающийся тем, что указанные зонды на этапе (с) представляют собой двухцепочечные зонды, и
 - i. длина каждого зонда находится в диапазоне от 100 до 500 пар оснований,
 - ii. каждый денатурированный зонд имеет 5'-конец и 3'-конец,
 - iii. предпочтительно каждый зонд связывается с ГТНСФ на расстоянии по меньшей мере 10 пар оснований, как на 5'-конце, так и на 3'-конце, от областей, несущих вариации числа копий (ВЧК), сегментные дупликации или элементы ДНК с повторяющимися последовательностями, и
 - iv. содержание GC в каждом зонде находится в диапазоне от 10% до 70%, предпочтительно от 15% до 60%, более предпочтительно от 20% до 50%.
5. Способ по любому из предшествующих пунктов, отличающийся тем, что указанный фрагмент нуклеиновой кислоты, подлежащий выявлению или происхождение которого предстоит определить, присутствует в указанной смеси в более низкой концентрации, чем фрагмент нуклеиновой кислоты из того же генетического локуса, но другого происхождения.
6. Способ по п. 5, отличающийся тем, что указанный фрагмент нуклеиновой кислоты, подлежащий выявлению или происхождение которого предстоит определить, и указанный фрагмент нуклеиновой кислоты из того же генетического локуса, но

другого происхождения, присутствуют в указанной смеси в соотношении, которое выбрано из группы 1:2, 1:4, 1:10, 1:20, 1:50, 1:100, 1:200, 1:500, 1:1000, 1:2000 и 1:5000.

7. Способ по любому из предшествующих пунктов, отличающийся тем, что указанные зонды зафиксированы на подложке.
8. Способ по любому из предшествующих пунктов, отличающийся тем, что указанные зонды являются биотинилированными и связаны с магнитными шариками, покрытыми стрептавидином.
9. Способ по любому из предшествующих пунктов, отличающийся тем, что содержание GC в указанных зондах находится в диапазоне от 10% до 70%, предпочтительно от 15% до 60%, в более предпочтительном варианте от 20% до 50%.
10. Способ выделения одного или более фрагментов нуклеиновой кислоты из смеси фрагментов нуклеиновой кислоты, включающий этапы:
 - a. обеспечения смеси фрагментированных нуклеиновых кислот, предпочтительно ДНК, которые происходят из эукариотического организма;
 - b. гибридизации одного или более зондов с по меньшей мере одним участком в указанных фрагментах нуклеиновой кислоты, где располагается горячая точка для неслучайной фрагментации (ГНСФ), или
 - c. амплификации одного или более участка из указанных фрагментов нуклеиновой кислоты, причем праймеры для амплификации примыкают к горячей точке для неслучайной фрагментации (ГНСФ).
11. Набор для определения происхождения фрагмента нуклеиновой кислоты в смеси фрагментов нуклеиновой кислоты для применения в способе по пп. 1-10, содержащий:
 - a. зонды, которые гибридизируются с по меньшей мере одним участком в указанном фрагменте нуклеиновой кислоты, причем указанный по меньшей мере один участок частично или полностью включает указанный фрагмент нуклеиновой кислоты и, необязательно,
 - b. реагенты и/или программное обеспечение для реализации способа, описанного согласно пп. 1-10, и способа определения и/или выявления.



ФИГ. 1



ФИГ. 2