

(19)



**Евразийское
патентное
ведомство**

(11) **037292**

(13) **B1**

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ

(45) Дата публикации и выдачи патента
2021.03.05

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

(21) Номер заявки
201500027

(22) Дата подачи заявки
2013.06.14

(54) СПОСОБ ДЕТЕКТИРОВАНИЯ РАКА ИЛИ ПРЕДЗЛОКАЧЕСТВЕННОГО ИЗМЕНЕНИЯ ПУТЕМ МУТАЦИОННОГО АНАЛИЗА ДНК В ПЛАЗМЕ

(31) 61/662,878; 61/682,725; 61/695,795;
61/711,172; 13/801,748

(32) 2012.06.21; 2012.08.13; 2012.08.31;
2012.10.08; 2013.03.13

(33) US

(43) 2015.05.29

(86) PCT/IB2013/054898

(87) WO 2013/190441 2013.12.27

(71)(73) Заявитель и патентовладелец:
**ТЕ ЧАЙНИЗ ЮНИВЕРСИТИ ОВ
ГОНКОНГ (CN)**

(72) Изобретатель:
**Чиу Вай Квунь Росса, Ло Юйк-Мин
Деннис, Чань Квань Чэ, Цзян Пэйюн
(CN)**

(74) Представитель:
Бадаева Т.Н., Фелицына С.Б. (RU)

(56) DANIELS Marissa et al.: Whole genome sequencing for lung cancer. Journal of Thoracic Disease, Vol. 4, No 2, April 2012, 155-163 - реферат, с. 160 DOI: 10.3978/j.issn.2072-1439.2012.02.01

WO-A1-2012071621

WO-A1-2009013496

WO-A2-2009013492

GB-A-2485635

WO-A1-2011090557

WO-A1-2011091046

Tao Y. et al.: "Rapid growth of a hepatocellular carcinoma and the driving mutations revealed by cell-population genetic analysis of whole-genome data", PNAS, 2011, Vol. 108, No 29, pages 12042-12047

Chan K.C.A. et al.: "Cancer genome scanning in plasma detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing", Clinical Chemistry, 2013, Vol. 59, No 1, pages 211-224, published online 11 October 2012, See the whole document

Chan K.C.A. et al.: "Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing", PNAS, 2013 November 19, Vol. 110, No 47, pages 18761-18768, See the whole document

Xie C. et al.: "CNV-seq, a new method to detect copy number variation using high-throughput sequencing", BMC Bioinformatics, 2009, Vol. 10, No 80

(57) Изобретение относится к способу детектирования рака или предзлокачественного изменения у субъекта, который основан на том, что частоту соматических мутаций в биологическом образце (например, в плазме или сыворотке) субъекта, проходящего скрининг или мониторинг на наличие рака, можно сравнить с таковой в конститутивной ДНК того же субъекта. Из данных частот можно получить параметр и применять его для определения уровня рака. Ложноположительные результаты можно отсеять, требуя, по меньшей мере, определенного количества считываемых последовательностей (маркеров) в любом измененном локусе, что позволяет получить более точный параметр. Относительные частоты различных измененных локусов можно проанализировать для определения уровня гетерогенности опухолей у пациента.

B1

037292

037292

B1

Ссылки на родственные заявки

Настоящая заявка представляет собой непредварительную заявку и испрашивает приоритет по предварительной заявке на патент США № 61/662878, названной "MUTATIONAL ANALYSIS OF PLASMA DNA FOR CANCER DETECTION", поданной 21 июня 2012 г.; предварительной заявки на патент США № 61/682725, названной "PAKAMUTATIONAL ANALYSIS OF PLASMA DNA FOR CANCER DETECTION", поданной 13 августа 2012 г.; предварительной заявки на патент США № 61/695795, названной "MUTATIONAL ANALYSIS OF PLASMA DNA FOR CANCER DETECTION", поданной 31 августа 2012 г.; и предварительной заявки на патент США № 61/711172, названной "MUTATIONAL ANALYSIS OF PLASMA DNA FOR CANCER DETECTION", поданной 8 октября 2012 г., которые полностью включены в данную заявку посредством ссылки во всех отношениях.

Уровень техники

Было показано, что ДНК опухолевого происхождения присутствует в свободной от клеток плазме/сыворотке пациентов с раковыми заболеваниями (Chen X.Q. и др. *Nat. Med.* 1996; 2: 1033-1035). Наиболее современные способы основаны на непосредственном анализе мутаций, о которых известно, что они связаны с раком (Diehl F. и др. *Proc. Natl. Acad. Sci.* 2005; 102: 16368-16373; Forshew T. и др. *Sci. Transl. Med.* 2012; 4: 136ra68). В другом способе исследуют связанные с раком вариации числа копий, обнаруженные при случайном секвенировании ДНК в плазме (публикация патента США 2013/0040824, Ло и др.).

Известно, что с течением времени более чем одна раковая клетка получит преимущество роста и даст начало нескольким клонам дочерних клеток. В конечном счете, растущая опухоль и/или ее метастазы будут содержать конгломерат из групп клональных раковых клеток. Данное явление обычно называют гетерогенностью опухоли (Gerlinger M. и др. *N. Engl. J. Med.* 2012; 366: 883-892; Yap T.A. и др. *Sci. Transl. Med.* 2012; 4: 127ps10).

Известно, что раковые заболевания высокогетерогенны, т.е. мутационный профиль раковых заболеваний одного и того же типа ткани может значительно различаться. Следовательно, непосредственный анализ конкретных мутаций обычно позволяет обнаружить лишь подгруппу случаев с конкретным типом рака, для которых известна взаимосвязь с данными конкретными мутациями. Кроме того, ДНК опухолевого происхождения обычно составляет лишь небольшую часть ДНК в плазме человека; абсолютная концентрация ДНК в плазме низка. Следовательно, при непосредственном детектировании одной связанной с раком мутации или небольшой группы таких мутаций в плазме или сыворотке аналитическая чувствительность может быть низкой даже среди пациентов с раковыми заболеваниями, при которых, как известно, присутствуют целевые мутации. Более того, было показано, что имеет место существенная внутриопухолевая гетерогенность мутаций, даже в одной опухоли. Мутации могут обнаруживаться только в субпопуляции опухолевых клеток. Различия в мутационных профилях между первичной опухолью и метастатическими поражениями еще больше. Один пример гетерогенности между исходной опухолью и первичными метастазами включает гены KRAS, BRAF и PIK3CA у пациентов, страдающих от колоректального рака (Baldus и др. *Clin. Pak Research* 2010. 16:790-9).

В случае, когда у пациента есть первичная опухоль (несущая мутацию KRAS, но не мутацию PIK3CA) и скрытое метастатическое повреждение (несущее мутацию PIK3CA, но не мутацию KRAS), если сосредоточиться на обнаружении мутации KRAS в первичной опухоли, то скрытое метастатическое поражение невозможно будет обнаружить. Тем не менее, если включить в анализ обе мутации, можно обнаружить как первичную опухоль, так и скрытое метастатическое поражение. Следовательно, анализ, включающий обе мутации, будет обладать большей чувствительностью в отношении обнаружения остаточных опухолевых тканей. Такой простой пример сильно усложняется, если проводят скрининговое обследование на присутствие рака и если информация о типе мутаций, которые могут возникнуть, незначительна или отсутствует.

Следовательно, существует потребность в новых способах осуществления широкого скрининга, детектирования или оценки рака.

Краткое описание изобретения

В вариантах реализации могут исследовать частоту соматических мутаций в биологическом образце (например, в плазме или сыворотке) из субъекта, проходящего скрининг или мониторинг на наличие рака, по сравнению с таковыми в конститутивной ДНК того же субъекта. Для определения данных частот можно применять случайное секвенирование. По данным частотам можно получить параметр и применять его для установления классификации уровня рака. Ложноположительные результаты можно отсеять, выбирая локус только в том случае, если он имеет по меньшей мере определенное количество считываемых последовательностей (маркеров) вариантов последовательности, что позволяет получить более точный параметр. Можно проанализировать относительные частоты различных измененных локусов, чтобы определить уровень гетерогенности опухолей у пациента.

В одном варианте реализации указанный параметр можно сравнить с тем же параметром, полученным для группы субъектов без ракового заболевания или с низким риском развития рака. Существенное различие между данным параметром, полученным для тестируемого субъекта, и тем же параметром, полученным для группы субъектов без ракового заболевания или с низким риском развития рака, может

свидетельствовать о повышенном риске того, что тестируемый субъект страдает раком или предзлокачественным состоянием или что у него в будущем разовьется рак. Таким образом, в одном варианте реализации анализ ДНК в плазме можно осуществить, не обладая предварительной геномной информацией об опухоли. Такой вариант реализации, следовательно, особенно пригоден для скрининга на наличие рака.

В другом варианте реализации описанные варианты реализации также можно применять для мониторинга пациента с раковым заболеванием после лечения и для выявления наличия остаточной опухоли или рецидива опухоли. Например, у пациента с остаточной опухолью или у которого опухоль рецидивировала будет наблюдаться повышенная частота соматических мутаций по сравнению с пациентом, у которого нет остаточной опухоли или у которого не наблюдается рецидив опухоли. Мониторинг может включать получение образцов из пациента с раковым заболеванием в несколько моментов времени после лечения для установления временных изменений связанных с опухолью генетических aberrаций в физиологических жидкостях или других образцах с внеклеточными нуклеиновыми кислотами, например, в плазме или сыворотке.

Согласно одному варианту реализации способ позволяет обнаружить рак или предзлокачественное изменение у субъекта. Получают конститутивный геном субъекта. Получают один или более маркеров последовательности для каждого из множества фрагментов ДНК в биологическом образце субъекта, если биологический образец содержит внеклеточную ДНК. Определяют положения в геноме маркеров последовательности. Маркеры последовательности сравнивают с конститутивным геномом, чтобы определить первое количество первых локусов. В каждом из первых локусов количество маркеров последовательности с вариантной последовательностью относительно конститутивного генома выше порогового значения, где пороговое значение больше единицы. На основании подсчета маркеров последовательности с вариантной последовательностью в первых локусах определяют параметр. Параметр сравнивают с пороговым значением, чтобы определить классификацию уровня рака у субъекта.

Согласно другому варианту реализации в способе анализируют гетерогенность одной или более опухолей у субъекта. Получают конститутивный геном субъекта. Получают один или более маркеров последовательности для каждого из множества фрагментов ДНК в биологическом образце субъекта, если биологический образец содержит внеклеточную ДНК. Определяют положения в геноме маркеров последовательности. Маркеры последовательности сравнивают с конститутивным геномом, чтобы определить первое количество первых локусов. В каждом из первых локусов количество маркеров последовательности с вариантной последовательностью относительно конститутивного генома выше порогового значения, где пороговое значение больше единицы. Меру гетерогенности одной или более опухолей рассчитывают на основании соответствующих первых количеств из набора первых положений в геноме.

Согласно другому варианту реализации в способе определяют фракционную концентрацию опухолевой ДНК в биологическом образце, содержащем внеклеточную ДНК. Получают один или более маркеров последовательности для каждого из множества фрагментов ДНК в биологическом образце. Определяют положения в геноме маркеров последовательности. Для каждой из множества областей генома определяют соответствующее количество фрагментов ДНК, принадлежащих данной области генома, с помощью маркеров последовательности, расположенных в геноме внутри данной области генома. Соответствующее количество нормируют с получением соответствующей плотности. Соответствующую плотность сравнивают с референсной плотностью, чтобы выявить в данной области генома утрату 1 копии или приобретение 1 копии. Первую плотность рассчитывают по соответствующим плотностям с выявленной утратой 1 копии или по соответствующим плотностям с выявленным приобретением 1 копии. Фракционную концентрацию рассчитывают путем сравнения первой плотности с другой плотностью с получением разности, при этом разность нормируют на референсную плотность.

Другие варианты реализации направлены на системы и машиночитаемые носители, связанные со способами, описанными в данной заявке.

Лучшего понимания природы и преимуществ настоящего изобретения можно добиться, основываясь на следующем подробном описании и сопроводительных фигурах.

Краткое описание фигур

На фиг. 1 представлена блок-схема способа 100 детектирования рака или предзлокачественного изменения у субъекта согласно вариантам реализации настоящего изобретения;

на фиг. 2 показана блок-схема способа непосредственного сравнения генома образца (ГО) с конститутивным геномом (КГ) согласно вариантам реализации настоящего изобретения;

на фиг. 3 - блок-схема способа 300 сравнения генома образца (ГО) с конститутивным геномом (КГ) с использованием референсного генома (ЭГ) согласно вариантам реализации настоящего изобретения;

на фиг. 4 представлена таблица 400, в которой показано количество связанных с раком однонуклеотидных мутаций, правильно детектированных с использованием встречаемости различных их количеств в качестве критерия для классификации мутации как присутствующей в образце согласно вариантам реализации настоящего изобретения, когда фракционную концентрацию ДНК опухолевого происхождения в образце принимают равной 10%;

на фиг. 5 - таблица, в которой показано ожидаемое количество ложноположительных локусов и ожидаемое количество мутаций, детектированных, когда фракционную концентрацию ДНК опухолевого

происхождения в образце принимают равной 5%;

на фиг. 6А - диаграмма 600, на которой показана частота детектирования связанных с раком мутаций в плазме с фракционными концентрациями ДНК опухолевого происхождения в плазме, составляющими 10 и 20%, и применение встречаемостей (r), равных четырем и шести, в качестве критериев для определения потенциальных связанных с раком мутаций. На фиг. 6В представлена диаграмма 650, на которой показано ожидаемое количество положений нуклеотидов, которые были ошибочно классифицированы как содержащие нуклеотидную замену с применением критериев встречаемости (r), равных 4, 5, 6 и 7, в зависимости от глубины секвенирования;

на фиг. 7А - диаграмма 700, на которой показано количество истинных связанных с раком мутированных сайтов и ложноположительных сайтов для различных глубин секвенирования, когда фракционную концентрацию ДНК опухолевого происхождения в образце принимают равной 5%. На фиг. 7В представлена диаграмма 750, на которой показано прогнозируемое количество ложноположительных сайтов при проведении анализа полного генома (ПГ) и всех экзонов;

на фиг. 8 - таблица 800, на которой показаны результаты для 4 пациентов с гепатоклеточной карциномой (ГКК) до и после лечения, включающие фракционные концентрации ДНК опухолевого происхождения в плазме согласно вариантам реализации настоящего изобретения;

на фиг. 9 - таблица 900, на которой показано детектирование связанных с ГКК однонуклеотидных вариаций (ОНВ) у 16 здоровых контрольных субъектов согласно вариантам реализации настоящего изобретения;

на фиг. 10А показан график распределения плотностей считываемых последовательностей образца опухоли из пациента с ГКК согласно вариантам реализации настоящего изобретения. На фиг. 10В показан график распределения 1050 z-показателей для всех фрагментов в плазме пациента с ГКК согласно вариантам реализации настоящего изобретения;

на фиг. 11 - график распределения 1100 z-показателей для плазмы из пациента с ГКК согласно вариантам реализации настоящего изобретения;

на фиг. 12 представлена блок-схема способа 1200 определения фракционной концентрации опухолевой ДНК в биологическом образце, содержащем внеклеточную ДНК, согласно вариантам реализации настоящего изобретения;

на фиг. 13А показана таблица 1300 анализа мутаций в плазме пациента с раком яичника и раком молочной железы на момент диагностики согласно вариантам реализации настоящего изобретения;

на фиг. 13В - таблица 1350 анализа мутаций в плазме пациента с двусторонним раком яичников и раком молочной железы после резекции опухоли согласно вариантам реализации изобретения;

на фиг. 14А представлена таблица 1400, на которой показано детектирование однонуклеотидных вариаций в ДНК в плазме для пациента 1 с ГКК (ГКК1). На фиг. 14В представлена таблица 1450, на которой показано детектирование однонуклеотидных вариаций в ДНК в плазме для пациента ГКК2;

на фиг. 15А - таблица 1500, на которой показано детектирование однонуклеотидных вариаций в ДНК в плазме для пациента ГКК3. На фиг. 15В представлена таблица 1550, на которой показано детектирование однонуклеотидных вариаций в ДНК в плазме для пациента ГКК4;

на фиг. 16 - таблица 1600, на которой показано детектирование однонуклеотидных вариаций в ДНК в плазме для пациента с раком яичников (и груди);

на фиг. 17 - таблица 1700, на которой показаны предсказанные чувствительности различных требований встречаемости и глубин секвенирования;

на фиг. 18 - таблица 1800, на которой показаны прогнозируемые количества ложноположительных локусов для различных пороговых значений и различных глубин секвенирования;

на фиг. 19 показана древовидная диаграмма, иллюстрирующая количество мутаций, детектированных в различных локализациях опухоли;

на фиг. 20 представлена таблица 2000, на которой показано количество фрагментов, несущих мутации опухолевого происхождения, в образце плазмы до лечения и после лечения;

на фиг. 21 - диаграмма 2100, на которой показаны распределения встречаемости в плазме мутаций, детектированных в одной локализации опухоли, и мутаций, детектированных во всех четырех локализациях опухоли;

на фиг. 22 - диаграмма 2200, на которой показано прогнозируемое распределение встречаемости в плазме мутаций, происходящих от гетерогенной опухоли;

на фиг. 23 продемонстрирована специфичность вариантов реализации для 16 привлеченных здоровых контрольных субъектов;

на фиг. 24 представлена блок-схема способа 2400 анализа гетерогенности одной или более опухолей субъекта согласно вариантам реализации настоящего изобретения;

на фиг. 25 показана блок-схема типичной компьютерной системы 2500, применимой с системой и способами согласно вариантам реализации настоящего изобретения.

Определения.

В данной заявке термин "локус" или его множественная форма "локусы" представляет собой положение или адрес фрагмента цепочки нуклеотидов (или пар оснований) любой длины, который может

содержать вариации в различных геномах. "Фрагмент" представляет собой участок заранее определенной длины в геноме. У множества фрагментов может быть одинаковая первая длина (разрешение), тогда как у отличного множества может быть одинаковая вторая длина. В одном варианте реализации фрагменты не перекрываются друг с другом.

Термин "случайное секвенирование" в данной заявке относится к секвенированию, при котором секвенированные фрагменты нуклеиновых кислот специально не идентифицировали или заранее не определяли перед процедурой секвенирования. Специфичные к последовательности праймеры для нацеливания на определенные локусы генов не требуются. Термин "универсальное секвенирование" относится к такому секвенированию, при котором процесс секвенирования может начаться с любого фрагмента. В одном варианте реализации к концу фрагмента добавляют адаптеры, и праймеры для секвенирования соединяются с адаптерами. Таким образом, любой фрагмент можно секвенировать с помощью одного и того же праймера и, таким образом, секвенирование может быть случайным.

Термин "маркер последовательности" (также называемый считываемой последовательностью) в данной заявке относится к цепочке нуклеотидов, секвенированной с любой части или по всей молекуле нуклеиновой кислоты. Например, секвенированный маркер может представлять собой короткую цепочку нуклеотидов (например, ~30), секвенированную по фрагменту нуклеиновой кислоты, короткую цепочку нуклеотидов с обоих концов фрагмента нуклеиновой кислоты или секвенирование целого фрагмента нуклеиновой кислоты, который присутствует в биологическом образце. Фрагмент нуклеиновой кислоты представляет собой любую часть более длинной молекулы нуклеиновой кислоты. Фрагмент (например, ген) может существовать отдельно (т.е. в не соединенном виде) от других частей более длинной молекулы нуклеиновой кислоты.

В данной заявке "конститутивный геном" (также называемый КГ) состоит из консенсусных нуклеотидов в локусах внутри генома, и, следовательно, можно считать его консенсусной последовательностью. КГ может охватывать весь геном субъекта (например, геном человека) или только части генома. Конститутивный геном (КГ) можно получить из ДНК клеток, а также из внеклеточной ДНК (например, которую можно обнаружить в плазме). В идеальном случае, консенсусные нуклеотиды должны указывать на то, что локус является гомозиготным по одной аллели или гетерозиготным по двум аллелям. Гетерозиготный локус обычно включает две аллели, которые содержат генетический полиморфизм. В качестве примера, критерием для определения того, является ли локус гетерозиготным, может быть порог появления каждой из двух аллелей по меньшей мере в заранее определенном проценте считываемых последовательностей (например, 30 или 40%), выровненных по данному локусу. Если один гомозиготный нуклеотид появляется с достаточным процентом (например, 70% или больше), тогда локус можно считать гомозиготным в КГ. Хотя геном одной здоровой клетки может отличаться от генома другой здоровой клетки вследствие случайных мутаций, происходящих спонтанно в процессе деления клетки, КГ не должен изменяться, когда используется такой консенсус. В геномах некоторых клеток могут происходить геномные перестройки, например в В- и Т-лимфоцитах, такие как перестройки, затрагивающие гены антител и Т-клеточных рецепторов. Такие крупномасштабные различия все же будут затрагивать относительно небольшую популяцию от всей популяции ядродержащих клеток в крови, и, таким образом, такие перестройки не будут влиять на определение конститутивного генома при достаточной выборке (например, глубине секвенирования) кровяных клеток. Другие типы клеток, включая буккальные клетки, клетки кожи, волосные фолликулы или биоптаты различных нормальных тканей организма, также могут служить источниками КГ.

Термин "конститутивная ДНК" относится к любому источнику ДНК, который отражает генетический состав, с которым субъект родился. Примеры "конститутивных образцов" субъекта, из которых можно получить конститутивную ДНК, включают здоровую ДНК клеток крови, ДНК буккальных клеток и ДНК корня волоса. ДНК из данных здоровых клеток составляет КГ субъекта. Определить, что клетки здоровы, можно множеством способов, например, когда известно, что индивид не страдает раком, или образец можно получить из ткани, которая скорее всего не содержит раковые или предзлокачественные клетки (например, ДНК корня волоса, когда подозревают рак печени). В качестве другого примера образец плазмы можно получить, когда у пациента нет рака, и детектированную конститутивную ДНК сравнить с результатами для полученного впоследствии образца плазмы (например, через год или позже). В другом варианте реализации для выявления конститутивного генома и связанных с опухолью генетических изменений можно использовать единичный биологический образец, имеющий менее 50% всей ДНК, происходящей из опухоли. В таком образце концентрации связанных с опухолью однонуклеотидных мутаций будут ниже, чем таковые для каждой аллели гетерозиготных однонуклеотидных полиморфизмов ОНП в КГ. Такой образец может быть тем же, что и биологический образец, используемый для определения генома образца, описанного ниже.

Термин "биологический образец" в данной заявке относится к любому образцу, который забирают у субъекта (например, у человека, индивида, страдающего раком, индивида, у которого подозревают наличие рака, или у других организмов), и включает одну или более интересующую внеклеточную молекулу (молекулы) нуклеиновой кислоты. Биологический образец может включать внеклеточную ДНК, часть которой могла произойти от здоровых клеток и часть - от опухолевых клеток. Например, опухолевую

ДНК можно найти в крови или других жидкостях, например, в моче, плевральной жидкости, асцитической жидкости, перитонеальной жидкости, слюне, слезах или спинномозговой жидкости. Не относящийся к жидкости пример представляет собой образец кала, который может быть смешан с диарейной жидкостью. Для некоторых из таких образцов биологический образец можно получить неинвазивным способом. В некоторых вариантах реализации биологический образец можно использовать в качестве конститутивного образца.

Термин "геном образца" (также называемый ГО) представляет собой набор считываемых последовательностей, которые выровняли с положениями в геноме (например, геноме человека). Геном образца (ГО) не является консенсусной последовательностью, но включает нуклеотиды, которые могут появиться только при достаточном количестве считываемых последовательностей (например, при пороговых значениях, равных по меньшей мере 2 или 3, или больше). Если аллель появляется достаточное количество раз и не является частью КГ (т.е. не является частью консенсусной последовательности), тогда этот аллель может обозначать "однонуклеотидную мутацию" (также называемую ОНМ). Применяя настоящее изобретение, также можно обнаружить другие типы мутаций, например мутации, включающие два или более нуклеотидов (такие, которые влияют на количество tandemных повторов в микросателлитах, или такие, как полиморфизм простого tandemного повтора), хромосомную транспозицию (которая может быть внутрехромосомной или межхромосомной) и инверсию последовательности.

Термин "референсный геном" (также называемый ЭГ) относится к гаплоидному или диплоидному геному, с которым можно выровнять и сравнить считываемые последовательности из биологического образца и конститутивного образца. В каждом локусе гаплоидного генома может быть только один нуклеотид. В диплоидном геноме можно обнаружить гетерозиготные локусы, при этом такой локус содержит две аллели, где любая аллель может обеспечить совпадение для выравнивания с данным локусом.

Термин "уровень рака" может относиться к любому из перечисленного: наличию рака, стадии рака, размеру опухоли и/или другим мерам тяжести рака. Уровень рака может представлять собой количество или другую характеристику. Уровень может быть равен нулю. Уровень рака также включает предзлокачественные или предраковые состояния (статусы), ассоциированные с мутациями или множеством мутаций. Уровень рака можно использовать различными способами. Например, в результате скрининга можно проверить наличие рака у кого-либо, у кого ранее не был выявлен рак. Оценка поможет исследовать кого-либо, у кого был диагностирован рак. Детектирование может означать "скрининг" или может означать проверку того, страдает ли раком кто-либо, у кого были детектированы признаки, позволяющие предположить наличие рака (например, симптомы или другие положительные анализы).

Подробное описание изобретения

Предложены варианты реализации для детектирования рака путем анализа биологического образца (например, образца плазмы/сыворотки крови), который не берут непосредственно из опухоли и который содержит внеклеточные нуклеиновые кислоты. Внеклеточные нуклеиновые кислоты могут происходить из различных типов ткани по всему организму. Таким образом, можно провести обширный анализ для детектирования различных раковых заболеваний.

Генетические aberrации (включая однонуклеотидные мутации, делеции, амплификации и перестройки) накапливаются в опухолевых клетках в процессе развития раковых заболеваний. В вариантах реализации можно применять массовое параллельное секвенирование для детектирования и определения количества однонуклеотидных мутаций (ОНМ), также называемых однонуклеотидными вариациями (ОНВ), в биологических жидкостях (например, плазме, сыворотке, слюне, асцитической жидкости, плевральной жидкости и спинномозговой жидкости), чтобы обнаружить и контролировать раковые заболевания. Анализ количества ОНМ (или другого типа мутаций) может обеспечить механизм определения ранних стадий рака в рамках скрининговых тестов. В различных вариантах реализации заботятся о том, чтобы отличить ошибки секвенирования и отличить спонтанные мутации, происходящие в здоровых клетках (например, требуя детектирования в конкретном локусе нескольких ОНМ, например, по меньшей мере 3, 4 или 5).

В некоторых вариантах реализации также предложены неинвазивные способы анализа гетерогенности опухоли, которые могут включать клетки из одной и той же опухоли (т.е. внутриопухолевая гетерогенность) или клетки из различных опухолей (либо из того же сайта, либо из различных сайтов) внутри организма. Например, можно неинвазивно проанализировать клональную структуру такой гетерогенности опухоли, включая оценку относительной массы опухолевых клеток, содержащих каждую мутацию. Мутации, которые присутствуют в более высоких относительных концентрациях, присутствуют в большем количестве злокачественных клеток в организме, например в клетках, которые появились раньше в процессе онкогенеза по сравнению с другими злокачественными клетками, также присутствующими в организме (Welch J.S. и др. Cell 2012; 150: 264-278). Ожидается, что такие мутации вследствие их большей относительной распространенности обеспечат большую диагностическую чувствительность для детектирования раковой ДНК, чем таковые с меньшей относительной распространенностью. Периодический мониторинг изменения относительной распространенности мутаций обеспечит возможность неинвазивного контроля изменений в клональном составе опухолей, либо спонтанных по мере прогрессирования заболевания, либо в ответ на лечение. Такая информация будет полезна для оценки прогноза или

для раннего детектирования устойчивости опухоли к лечению.

I. Введение.

Мутации могут возникать в процессе деления клетки из-за ошибок при репликации ДНК и/или репарации ДНК. Один тип таких мутаций включает изменение отдельных нуклеотидов, которое может происходить в нескольких последовательностях из различных областей генома. Как правило, полагают, что раковые заболевания возникают вследствие клональной экспансии отдельной раковой клетки, которая получила преимущество роста. Такая клональная экспансия будет приводить к накоплению мутаций (например, однонуклеотидных мутаций) во всех раковых клетках, которые произошли от исходной раковой клетки. У таких вторичных опухолевых клеток будет одинаковый набор мутаций (например, однонуклеотидных мутаций). В данной заявке описано, что ассоциированные с раком однонуклеотидные мутации обнаруживают в плазме/сыворотке пациентов с раковыми заболеваниями.

Некоторые варианты реализации позволяют эффективно осуществить скрининг всех мутаций в биологическом образце (например, в плазме или сыворотке). Так как количество мутаций не фиксировано (можно обнаружить сотни, тысячи или миллионы связанных с раком мутаций из различных субпопуляций опухолевых клеток), варианты реализации могут обеспечить большую чувствительность, чем методики, которые позволяют обнаружить конкретные мутации. Для детектирования рака можно использовать множество мутаций.

Для того чтобы осуществить такой скрининг множества или всех мутаций, в вариантах реализации могут проводить поиск (например, случайный поиск) генетических изменений в биологическом образце (например, в физиологических жидкостях, включая плазму и сыворотку), который может содержать ДНК опухолевого происхождения. Применение образца, такого как плазма, избавляет от необходимости проведения инвазивной биопсии опухоли или рака. Кроме того, так как скрининг может покрывать весь геном или большие фрагменты генома, такой скрининг не ограничен какими-либо перечисленными и известными мутациями, но в нем может использоваться любая существующая мутация. Более того, поскольку количество мутаций суммируется по всему геному или по большим фрагментам генома, можно получить более высокую чувствительность.

Тем не менее, в геноме человека существуют полиморфные сайты, включая однонуклеотидные полиморфизмы (ОНП), которые не следует считать мутациями. С помощью вариантов реализации можно установить, похоже ли, что генетические изменения, которые были обнаружены, являются связанными с раком мутациями, или они представляют собой полиморфизмы в геноме. Например, в рамках выявления различия между связанными с раком мутациями и полиморфизмами в геноме, варианты реализации позволяют определить конститутивный геном, который может включать полиморфизмы. Полиморфизмы конститутивного генома (КГ) можно ограничить полиморфизмами, которые представлены с достаточно высоким процентным содержанием (например, 30-40%) в результатах секвенирования.

Последовательности, полученные из биологического образца, можно затем выровнять с конститутивным геномом и обнаружить вариации, которые представляют собой однонуклеотидные мутации (ОНМ), или другие типы мутаций. Данные ОНМ будут представлять собой вариации, не включенные в известные полиморфизмы, и, таким образом, можно отметить, что они связаны с раком и не относятся к конститутивному геному. У здорового индивида может встречаться некоторое количество ОНМ вследствие случайных мутаций в здоровых клетках, например, возникших в процессе деления клетки, но у индивида, страдающего раком, их будет гораздо больше.

Например, у страдающего раком индивида количество ОНМ, детектируемых в физиологической жидкости, будет выше, чем количество полиморфизмов, присутствующих в конститутивном геноме того же индивида. Можно сравнить количества вариаций, детектируемых в образце физиологической жидкости, содержащем ДНК опухолевого происхождения, и в образце ДНК, содержащем преимущественно конститутивную ДНК. В одном варианте реализации термин "преимущественно" будет означать более чем 90%. В другом предпочтительном варианте реализации термин "преимущественно" будет означать более чем 95, 97, 98 или 99%. Если количество вариаций в физиологической жидкости превышает такое в образце с преимущественно конститутивной ДНК, существует повышенная вероятность того, что физиологическая жидкость может включать ДНК опухолевого происхождения.

Одним из способов, которые можно применять для случайного поиска вариаций в образцах ДНК, является случайное секвенирование или секвенирование методом "выстрела из дробового ружья" (например, применяя массовое параллельное секвенирование). Можно применять любую платформу для массового параллельного секвенирования, включая платформу секвенирования путем лигирования (например, платформу Life Technologies SOLiD Platform), Ion Torrent/Ion Proton, полупроводниковое секвенирование, Roche 454, платформы для секвенирования одиночных молекул (например, Helicos, Pacific Biosciences и Nanopore). Кроме того, известно, что могут происходить ошибки секвенирования, которые могут неправильно истолковываться как вариации в конститутивной ДНК или как мутации, происходящие из опухолевой ДНК. Таким образом, для улучшения специфичности предложенного нами подхода, вероятность ошибки секвенирования или других компонентов аналитических ошибок можно учесть, например, применяя подходящую глубину секвенирования наряду с требованием детектирования в локусе, по меньшей мере, определенного количества (например, 2 или 3) аллелей, чтобы считать его содержа-

щим ОНМ.

В данной заявке описано, что в вариантах реализации могут быть представлены доказательства присутствия ДНК опухолевого происхождения в биологическом образце (например, в физиологической жидкости), когда количество случайно детектированных генетических изменений, присутствующих в образце, превышает таковое, ожидаемое для конститутивной ДНК и для вариаций, которые могут быть случайно детектированы вследствие аналитических ошибок (например, ошибок секвенирования). Полученные сведения можно использовать для скрининга, диагностики, прогнозирования и мониторинга раковых заболеваний. В следующих разделах мы описали аналитические этапы, которые можно применять для детектирования однонуклеотидных мутаций в плазме/сыворотке или других образцах (например, в физиологических жидкостях). Физиологические жидкости могут включать плазму, сыворотку, спинномозговую жидкость, плевральную жидкость, асцитную жидкость, выделения из соска, слюну, жидкость бронхоальвеолярного лаважа, мокроту, слезы, пот и мочу. Вдобавок к физиологическим жидкостям указанный способ также можно применять по отношению к образцу кала, так как было показано, что последний может содержать опухолевую ДНК из колоректального рака (Berger B.M., Ahlquist D.A. *Pathology* 2012; 44: 80-88).

II. Основной способ скрининга.

На фиг. 1 представлена блок-схема способа 100 детектирования рака или предзлокачественного изменения у субъекта согласно вариантам реализации настоящего изобретения. Варианты реализации позволяют анализировать внеклеточную ДНК в биологическом образце из субъекта для детектирования вариаций во внеклеточной ДНК, вероятно происходящих из опухоли. В указанном анализе могут использовать конститутивный геном субъекта, чтобы учесть полиморфизмы, которые принадлежат здоровым клеткам, и можно учитывать ошибки секвенирования. Способ 100 и любой из способов, описанных в данной заявке, можно полностью или частично осуществить с помощью компьютерной системы, включающей один или более процессоров.

На этапе 110 получают конститутивный геном субъекта. Конститутивный геном (КГ) можно определить по конститутивной ДНК исследуемого субъекта. В различных вариантах реализации КГ можно считать из памяти или определить активным способом, например путем анализа считываемых последовательностей конститутивной ДНК, которая может находиться в клетках из образца, который включает внеклеточную ДНК. Например, если подозревают наличие негематологического злокачественного новообразования, можно проанализировать клетки крови, чтобы определить конститутивную ДНК субъекта.

В различных вариантах реализации анализ конститутивной ДНК можно осуществить, применяя массовое параллельное секвенирование, гибридизацию на чипах, гибридизацию в растворе, содержащем зонды, анализы на основе лигирования, анализы на основе реакции удлинения праймеров и масс-спектрометрию. В одном варианте реализации КГ можно определить в один момент времени жизни субъекта, например, при рождении или даже в пренатальном периоде (что можно осуществить, применяя клетки зародыша или используя фрагменты внеклеточной ДНК, см. публикацию США 2011/0105353), а затем сравнить с ним результаты для физиологических жидкостей или других образцов, полученных в другие моменты времени жизни субъекта. Таким образом, КГ можно просто считать из памяти компьютера. Конститутивный геном можно считать в виде перечня локусов, по которым конститутивный геном отличается от референсного генома.

На этапе 120 получают один или более маркеров последовательности для каждого из множества фрагментов ДНК в биологическом образце субъекта, если биологический образец содержит внеклеточную ДНК. В одном варианте реализации указанный один или более маркеров последовательности получают в результате случайного секвенирования фрагментов ДНК в биологическом образце. Более чем один маркер последовательности можно получить при осуществлении секвенирования спаренных концов. Один маркер будет соответствовать каждому концу фрагмента ДНК.

Можно провести анализ наличия генетических изменений во внеклеточной ДНК в образце (например, в плазме, сыворотке или другой биологической жидкости). Можно провести анализ внеклеточной ДНК, применяя такую же аналитическую платформу, которую применяли для анализа конститутивной ДНК. В качестве альтернативы, можно применять отличную аналитическую платформу. Например, образец внеклеточной ДНК можно секвенировать, применяя массовое параллельное секвенирование, или можно захватить части генома или обогатить материал частями генома до проведения массового параллельного секвенирования. При использовании обогащения можно, например, применять жидкофазный или твердофазный захват выбранных частей генома. Затем можно осуществить массовое параллельное секвенирование захваченной ДНК.

На этапе 130 определяют положения в геноме маркеров последовательности. В одном варианте реализации маркеры последовательности выравнивают с референсным геномом, который получают из одного или более других субъектов. В другом варианте реализации маркеры геномных последовательностей выравнивают с конститутивным геномом исследуемого субъекта. Выравнивание можно осуществить, применяя методики, известные специалисту в данной области, например применяя основное средство поиска локального выравнивания (BLAST).

На этапе 140 определяют первое количество локусов, в которых по меньшей мере N маркеров по-

следовательности содержат вариант последовательности по сравнению с конститутивным геномом (КГ). N равно или больше двух. Ниже подробнее обсуждается, что ошибки секвенирования, а также соматические мутации, случайным образом возникающие в клетках (например, в результате деления клетки), можно устранить, требуя, чтобы N было равно 2, 3, 4, 5 или больше. Локусы, которые удовлетворяют одному или более определенным критериям, можно считать мутацией (вариантом) или мутантными локусами (измененными локусами), тогда как локус, содержащий вариант, но не удовлетворяющий указанному одному или более критериям (например, только один маркер измененной последовательности), относят к потенциальной или предполагаемой мутации. В последовательности может быть изменен только один нуклеотид или множество нуклеотидов.

N можно определить как процент от всех маркеров для некоторого локуса, а не как абсолютное значение. Например, измененный локус можно обнаружить, когда определенная фракционная концентрация опухолевой ДНК, полученная в результате вариантов считываемых последовательностей, равна или больше 10% (или некоторого другого процента). Другими словами, если покрытие локуса составляет 200 считываемых последовательностей, в качестве критерия для определения того, что данный вариант является мутацией, можно выставить условие, что, по меньшей мере, должно быть 10 считываемых последовательностей, в которых выявлена вариантная аллель. 10 считываемых последовательностей вариантной аллели и 190 считываемых последовательностей аллели дикого типа означает, что фракционная концентрация опухолевой ДНК составляет 10% ($2 \times 10 / (10 + 190)$).

В одном варианте реализации маркеры последовательности (в совокупности называемые геномом образца) можно сравнить непосредственно с КГ, чтобы определить варианты. В другом варианте реализации геном образца (ГО) сравнивают с КГ с помощью референсного генома (ЭГ), чтобы определить варианты. Например, как КГ, так и ГО можно сравнить с ЭГ, чтобы определить соответствующие количества (например, наборы) локусов, в которых детектированы варианты, а затем можно вычислить разность, чтобы получить первое количество локусов. Первое количество можно получить просто в виде количества, или оно может соответствовать определенному набору локусов, которые затем можно дополнительно проанализировать, чтобы определить параметр по маркерам последовательности в первых локусах.

В одном варианте реализации сравнивают результаты секвенирования конститутивной ДНК и ДНК в плазме, чтобы определить, присутствует ли однонуклеотидная мутация в ДНК в плазме. Можно проанализировать области, в которых конститутивная ДНК гомозиготна. С целью иллюстрирования предположим, что генотип в определенном локусе конститутивной ДНК гомозиготен и представляет собой AA. Тогда присутствие в плазме аллеля, отличного от A, будет свидетельствовать о потенциальном присутствии однонуклеотидной мутации (ОНМ) в данном локусе. Локусы, в которых выявлено потенциальное присутствие ОНМ, могут составлять первое количество локусов на этапе 140.

В одном варианте реализации может оказаться полезным нацеливание на части генома, которые, как известно, особенно подвержены мутациям при конкретном типе рака или в конкретной субпопуляции. В отношении последнего из упомянутых аспектов в вариантах реализации можно искать типы мутаций, которые особенно широко распространены в определенной популяционной группе, например мутации, которые особенно распространены у субъектов, которые являются носителями вируса гепатита В (для рака печени) или вируса папилломы человека (для рака шейки матки), или у которых есть генетическая предрасположенность к соматическим мутациям, или у субъектов с генеративными мутациями в гене репарации несовпадений ДНК. Указанный способ также будет полезен для скрининга наличия мутаций при раках яичников и молочной железы у субъектов с мутациями в генах BRCA1 или BRCA2. Способ также будет полезен для скрининга наличия мутаций при колоректальном раке у субъектов с мутациями в гене APC.

На этапе 150 параметр определяют на основании подсчета маркеров последовательности с вариантной последовательностью в первых локусах. В одном примере указанный параметр представляет собой первое количество локусов, в которых, по меньшей мере, N фрагментов ДНК содержат вариант последовательности в локусе по сравнению с конститутивным геномом. Таким образом, можно использовать подсчет, чтобы просто удостовериться в том, что в локусе детектировано более чем N копий конкретного варианта, что определяют перед его включением в первое количество. В другом варианте реализации указанный параметр может представлять собой или включать суммарное количество маркеров последовательности с последовательностью в первых локусах, измененной относительно конститутивного генома.

На этапе 160 параметр для субъекта сравнивают с пороговым значением (например, полученным из одного или более других субъектов), чтобы определить классификацию уровня рака у субъекта. Примеры уровня рака включают наличие рака или предзлокачественного состояния у субъекта, или повышенную вероятность развития рака. В одном варианте реализации пороговое значение можно определить в образце, полученном из данного субъекта ранее.

В другом варианте реализации можно установить, что один или более других субъектов не страдают раком или обладают низким риском развития рака. Таким образом, пороговое значение может представлять собой нормальное значение, нормальный диапазон или обозначать статистически значимое отклонение от нормального значения или диапазона. Например, количество мутаций по сравнению с КГ

конкретного субъекта, обнаруживаемое в плазме субъектов, не страдающих раком, или с низким риском развития рака, можно использовать в качестве нормального диапазона, чтобы определить, является ли количество мутаций, детектированных у исследованного субъекта, нормальным. В другом варианте реализации может быть известно, что другие субъекты страдают раком и, таким образом, аналогичное количество мутаций может указывать на наличие рака.

В одном варианте реализации других субъектов можно выбрать таким образом, чтобы их клинические свойства совпадали с таковыми у тестируемого субъекта, например пол, возраст, режим питания, привычка к курению, лекарственный анамнез, предшествующее заболевание, семейный анамнез, генотипы в выбранных геномных локусах, статус в отношении вирусных инфекций (например, инфекций вирусом гепатита В или С, или вирусом папилломы человека, или вирусом иммунодефицита человека, или вирусом Эпштейна-Барр) или инфекций другими инфекционными агентами (такими как бактерии (например, *Helicobacter pylori*) и паразиты (например, *Clonorchis sinensis*) и т.д.). Например, у субъектов, которые являются носителями вируса гепатита В или С, повышен риск развития печеночно-клеточной карциномы. Таким образом, можно считать, что у тестируемых субъектов, у которых количество мутаций или паттерн мутаций аналогичны таковому у носителя гепатита В или С, повышен риск развития печеночно-клеточной карциномы. С другой стороны, у пациента с гепатитом В или С, у которого обнаружили большее количество мутаций, чем у другого пациента с гепатитом, можно правильно классифицировать уровень рака как более высокий, поскольку используется подходящее исходное состояние (т.е. сравнение с другим пациентом с гепатитом). Аналогично, у субъектов, которые являются носителями инфекции вирусом папилломы человека, повышен риск рака шейки матки и рака головы и шеи. Инфекция вирусом Эпштейна-Барр была связана с карциномой носоглотки, раком желудка, лимфомой Ходжкина и неходжкинской лимфомой. Инфекция *Helicobacter pylori* была связана с раком желудка. Инфекция *Clonorchis sinensis* была связана с холангиокарциномой.

Мониторинг изменений количества мутаций в различные моменты времени можно применять для контроля прогрессирования рака и ответа на лечение. Такой мониторинг также можно применять для подтверждения прогрессирования предзлокачественного состояния или изменения риска того, что у субъекта разовьется рак.

Количество маркеров последовательности, в которых выявлены вариации, также можно использовать для мониторинга. Например, можно применять фракционную концентрацию считываемых последовательностей варианта в некотором локусе. В одном варианте реализации повышение фракционных концентраций связанных с опухолью генетических aberrаций в образцах при осуществлении серийного мониторинга может обозначать прогрессирование заболевания или угрожающий рецидив. Аналогично, уменьшение фракционных концентраций связанных с опухолью генетических aberrаций в образцах при осуществлении серийного мониторинга может обозначать ответ на лечение, и/или ремиссию, и/или благоприятный прогноз.

III. Определение геномов.

Различные геномы, обсуждаемые выше, более подробно объяснены ниже. Например, обсуждаются референсный геном, конститутивный геном и геном образца.

A. Референсный геном.

Референсный геном (ЭГ) относится к гаплоидному или диплоидному геному субъекта или к консенсусу у популяции. Референсный геном известен и, следовательно, его можно применять для сравнения с ним считываемых последовательностей при секвенировании геномов новых пациентов. Считываемые последовательности для образца пациента можно выровнять и сравнить, чтобы определить вариации в данных считываемых последовательностях по сравнению с ЭГ. В каждом локусе гаплоидного генома присутствует только один нуклеотид, и, следовательно, каждый локус можно считать гемизиготным. В диплоидном геноме можно обнаружить гетерозиготные локусы, при этом такие локусы содержат две аллели, и при выравнивании с данным локусом совпадение может быть с любой аллелью.

Референсный геном может быть одинаковым у популяции субъектов. Такой одинаковый референсный геном можно использовать для определения у здоровых субъектов подходящего порога для применения для классификации пациента (например, как имеющего или не имеющего рак). Тем не менее, можно применять различные референсные геномы для различных популяций, например для различных этнических групп или даже для различных семей.

B. Конститутивный геном.

Конститутивный геном (КГ) субъекта (например, человека или другого диплоидного организма) относится к диплоидному геному субъекта. В гетерозиготных локусах КГ можно точно определить, что первая аллель принадлежит первому гаплотипу и отличная вторая аллель принадлежит второму гаплотипу. Следует отметить, что структуры двух гаплотипов, которые охватывают два гетерозиготных локуса, не обязательно должны быть известны, т.е. какая аллель одного гетерозиготного локуса находится в том же гаплотипе, что и аллель другого гетерозиготного локуса. Просто существования двух аллелей в каждом гетерозиготном локусе может быть достаточно.

КГ может отличаться от ЭГ благодаря полиморфизмам. Например, локус в ЭГ может быть гомозиготным по Т, а в КГ он может быть гетерозиготным: Т/А. Таким образом, для КГ будет детектирована

вариация в данном локусе. КГ также может отличаться от ЭГ благодаря наследственным мутациям (например, которые наследуются членами семьи) или мутациям *de novo* (которые возникают у зародыша, но которых нет у его родителей). Наследственную мутацию обычно называют "генеративной мутацией". Некоторые из таких мутаций связаны с предрасположенностью к раку, такие как мутация в BRCA1, которая наследуется членами семьи. Такие мутации отличны от "соматических мутаций", которые могут происходить в результате деления клеток на протяжении всей жизни субъекта и могут способствовать тому, чтобы клетка и ее дочерние клетки стали раковыми.

Целью определения КГ является отделение таких генеративных мутаций и мутаций *de novo* от мутаций генома образца (ГО), чтобы определить соматические мутации. Количество соматических мутаций в ГО затем можно применять для оценки вероятности возникновения рака у субъекта. Данные соматические мутации можно подвергнуть дополнительному отбору, чтобы удалить ошибки секвенирования и потенциально удалить соматические мутации, которые редко встречаются (например, вариант обнаруживают только в одной считываемой последовательности), так как такие соматические мутации вероятно не связаны с раком.

В одном варианте реализации КГ можно определить с использованием клеток (ДНК из лейкоцитарной пленки). Тем не менее, КГ также можно определить по внеклеточной ДНК (например, из плазмы или сыворотки). У типа образца, в котором большинство клеток не являются злокачественными, например в лейкоцитарной пленке из здорового субъекта, большая часть генома или консенсусный геном представляет собой КГ. В КГ каждый геномный локус состоит из последовательности ДНК, которая совпадает у большинства клеток в ткани образца. Глубина секвенирования должна быть достаточной для того, чтобы выявить гетерозиготные сайты в конститутивном геноме.

В качестве другого примера для определения КГ можно использовать плазму в качестве конститутивного образца. Например, в случаях, когда содержание опухолевой ДНК в плазме составляет менее чем 50% и ОНМ находятся в гетерозиготном состоянии, например, мутация представляет собой вставку новой аллели, концентрация новой аллели может составлять менее чем 25%. Тогда как концентрация гетерозиготных аллелей ОНП в КГ должна составлять до приблизительно 50%. Таким образом можно различить соматическую мутацию и полиморфизм КГ. В одном варианте реализации подходящее пороговое значение для того, чтобы отличить соматическую мутацию от полиморфизма, может находиться в диапазоне 30-40%, при использовании плазмы или других смесей с существенной концентрацией ДНК опухоли. Измерение концентрации опухолевой ДНК может оказаться полезным, чтобы удостовериться в том, что концентрация опухолевой ДНК в плазме составляет менее 50%. Примеры определения концентрации опухолевой ДНК описаны в данной заявке.

С. Геном образца.

Геном образца (ГО) представляет собой не просто гаплоидный или диплоидный геном, как в случае ЭГ и КГ. ГО представляет собой совокупность считываемых последовательностей образца и может включать считываемые последовательности конститутивной ДНК, которая соответствует КГ, считываемые последовательности опухолевой ДНК, считываемые последовательности здоровых клеток, у которых обнаружены случайные мутации по сравнению с КГ (например, вследствие мутаций, возникших в результате деления клеток), и ошибки секвенирования. Для того чтобы точно контролировать, какие считываемые последовательности включены в ГО, можно применять различные параметры. Например, требование, чтобы аллель появился по меньшей мере в 5 считываемых последовательностях, позволит уменьшить ошибки секвенирования, присутствующие в ГО, а также устранить часть считываемых последовательностей, связанных со случайными мутациями.

В качестве примера, предположим, что субъект здоров, т.е. не страдает раком. С целью иллюстрирования, ДНК из 1000 клеток находится в 1 мл плазмы (т.е. 1000 геном-эквивалентов ДНК), полученной из данного субъекта. ДНК в плазме обычно состоит из фрагментов ДНК размером приблизительно 150 п.о. Так как геном человека состоит из 3×10^9 п.о., там будет находиться приблизительно 2×10^7 фрагментов ДНК на гаплоидный геном. Так как геном человека диплоидный, в 1 мл плазмы будет находиться приблизительно 4×10^7 фрагментов ДНК.

Так как в единицу времени от миллионов до миллиардов клеток высвобождают ДНК в плазму, и фрагменты ДНК из данных клеток будут смешиваться в кровотоке, 4×10^7 фрагментов ДНК могли произойти от 4×10^7 различных клеток. Если данные клетки не произошли недавно (в противоположность удаленному происхождению, например, от исходной зиготы) от одного клона (т.е. у них нет общей недавней предковой клетки), тогда статистически вероятно, что ни одна мутация не встретится более одного раза среди данных фрагментов.

С другой стороны, если среди 1000 геном-эквивалентов на мл ДНК в плазме присутствует некоторый процент клеток, у которых есть общая недавняя предковая клетка (т.е. они клонально родственны друг другу), то можно видеть, что мутации из данного клона будут преимущественно представлены в ДНК плазмы (например, представляющие клональный мутационный профиль в плазме). Такие клонально родственные клетки могут оказаться раковыми клетками или клетками, которые скоро станут раковыми, но пока таковыми не являются (т.е. преднеопластические). Таким образом, требование появления

мутации более одного раза позволяет устранить такие природные вариации в "мутациях", детектированных в образце, что позволяет оставить мутации, связанные с раковыми клетками или преднеопластическими клетками, что обеспечивает детектирование, особенно раннее детектирование, рака или предраковых состояний.

В одном приближении было установлено, что после каждого деления клетки в геноме будет накапливаться в среднем одна мутация. В предшествующем исследовании показали, что большая часть ДНК в плазме происходит из гематопоэтических клеток (Lui Y.Y. и др. Clin. Chem. 2002: 48: 421-427). Оценили, что гематопоэтические стволовые клетки реплицируются раз в 25-50 недель (Catlin S.N. и др. Blood 2011; 117: 4460-4466). Таким образом, в качестве упрощенного приближения, здоровый субъект к возрасту 40 лет накопит примерно от 40 до 80 мутаций на гематопоэтическую стволовую клетку.

Если в 1 мл плазмы данного индивида содержится 1000 геном-эквивалентов и если каждая из данных клеток произошла от отдельной гематопоэтической стволовой клетки, то можно ожидать выявление от 40000 до 80000 мутаций среди 4×10^{10} фрагментов ДНК (т.е. 4×10^7 фрагментов ДНК на геном и 1000 геном-эквивалентов на мл плазмы). Тем не менее, так как каждая мутация будет встречаться лишь раз, каждая мутация все же может быть ниже предела детектирования (например, если пороговое значение N больше чем 1), и, следовательно, данные мутации можно отсеять, что позволяет сконцентрировать анализ на мутациях, которые с большей вероятностью относятся к раковым состояниям. Пороговое значение может представлять собой любое значение (целое или нецелое) больше единицы, и может изменяться для различных локусов и фрагментов. Глубина секвенирования и фракционная концентрация опухолевой ДНК также могут влиять на чувствительность детектирования мутаций (например, процент детектированных мутаций) из раковых клеток или преднеопластических клеток.

IV. Непосредственное сравнение ГО с КГ.

В некоторых вариантах реализации можно определить положения нуклеотидов, в которых КГ гомозиготный, но в меньшем числе молекул (т.е. в опухолевой ДНК) в ГО данные положения гетерозиготные. При большой глубине секвенирования в некотором положении (например, покрытие более 50 раз), можно обнаружить присутствие одной или двух аллелей в данном положении в смеси ДНК здоровых и раковых клеток. Если обнаруживают две аллели, то либо (1) КГ гетерозиготный, либо (2) КГ гомозиготный, но ГО гетерозиготный. Данные два варианта можно различить, если посмотреть на относительные количества мажорной и минорной аллелей. В первом из упомянутых вариантов количества считанных двух аллелей будут аналогичны; но в последнем из упомянутых вариантов будет большое различие в количествах вариантов считывания каждой. Такое сравнение относительных количеств вариантов считывания аллелей в тестируемом образце представляет собой один вариант реализации для сравнения маркеров последовательности с конститутивным геномом. Можно определить, что первые локусы из способа 100 представляют собой локусы, в которых количество аллелей ниже верхнего порога (порога, соответствующего полиморфизмам в КГ) и выше нижнего порога (порога, соответствующего ошибкам и соматическим мутациям, встречающимся с достаточно низкой частотой, чтобы не быть связанными с раковым состоянием). Таким образом, конститутивный геном и первые локусы можно определить одновременно.

В другом варианте реализации в процессе идентификации мутаций можно сначала определить КГ, а затем определить локусы, содержащие достаточное количество мутаций по сравнению с КГ. КГ можно определить по конститутивному образцу, который отличается от тестируемого образца.

На фиг. 2 показана блок-схема способа 200 непосредственного сравнения генома образца (ГО) с конститутивным геномом (КГ) согласно вариантам реализации настоящего изобретения. В блоке 210 получают конститутивный геном субъекта. Конститутивный геном можно получить, например, из образца, полученного ранее, или из конститутивного образца, который получают и анализируют непосредственно перед осуществлением способа 200.

В блоке 220 получают один или более маркеров последовательности для каждого из множества фрагментов ДНК в биологическом образце субъекта. Секвенирование можно осуществить, применяя различные методики, упомянутые в данной заявке. Маркеры последовательности представляют собой оценку того, какой считают последовательность фрагмента. Но одно или более оснований в маркере последовательности могут быть определены ошибочно.

В блоке 230 по меньшей мере часть маркеров последовательности выравнивают с конститутивным геномом. При выравнивании могут учитывать, что КГ может быть гетерозиготным в различных локусах. Для выравнивания не будет требоваться точное совпадение, чтобы можно было обнаружить варианты.

В блоке 240 обнаруживают маркеры последовательности, которые содержат изменения последовательности в локусе по сравнению с конститутивным геномом. Возможно, что будет детектирован более чем один вариант маркера последовательности. Можно отследить варианты для каждого локуса и для каждого маркера последовательности. Вариант может представлять собой любую аллель, которая не представлена в КГ. Например, КГ может быть гетерозиготным по А/Г и вариант может быть G или C.

В блоке 250 для каждого локуса с вариантом компьютерная система может определить соответствующее первое количество маркеров последовательности, которые выравниваются с данным локусом и содержат изменение последовательности в данном локусе. Таким образом, с каждым локусом может быть связано подсчитанное количество вариантов, наблюдаемых в данном локусе. Как правило, в локусе

будет наблюдаться меньшее количество вариантов по сравнению с маркерами последовательности, которые соответствуют КГ, например вследствие того, что концентрация опухолевой ДНК составляет менее 50%. Тем не менее, в некоторых образцах концентрация опухолевой ДНК может быть больше чем 50%.

В блоке 260 параметр определяют на основании соответствующих первых количеств. Если в одном варианте реализации соответствующее количество больше, чем пороговое значение (например, больше, чем два), то соответствующее количество можно прибавить к сумме, которая представляет собой параметр или используется для определения параметра. В другом варианте реализации в качестве параметра применяют количество локусов, соответствующее количеству для которых больше, чем пороговое значение.

В блоке 270 параметр сравнивают с пороговым значением, чтобы классифицировать уровень рака. Выше описано, что пороговое значение можно определить путем анализа образцов из других субъектов. В зависимости от того, здоровы ли данные другие субъекты или страдают раковыми состояниями, можно определить классификацию. Например, если у других субъектов выявлена 4 стадия рака, то если исследуемый параметр был близок (например, в рамках определенного диапазона) к значению параметра, полученного для других субъектов, значит можно классифицировать стадию рака исследуемого субъекта как 4 стадию рака. Тем не менее, если указанный параметр переходит границы порогового значения (т.е. больше или меньше порогового значения, в зависимости от способа определения параметра), то можно классифицировать стадию рака как меньшую, чем 4 стадия. Аналогичный анализ можно осуществить, когда другие субъекты не страдают раком.

Для определения классификации можно применять несколько пороговых значений, где каждое пороговое значение определяют у различной группы субъектов. У каждой группы субъектов может быть одинаковый уровень рака. Таким образом, исследуемый параметр можно сравнить со значениями для каждой группы субъектов, в результате чего можно получить совпадение с одной из групп или получить диапазон. Например, параметр может быть приблизительно одинаковым с параметром, полученным для субъектов, которые находятся на стадии предрака или на 2 стадии рака. В качестве другого примера исследуемый параметр может попасть в диапазон, который возможно может подходить для нескольких различных уровней рака. Таким образом, классификация может включать более чем один уровень рака.

V. Применение референсного генома.

Геномные последовательности как конститутивной ДНК, так и ДНК из биологического образца можно сравнить с референсным геномом человека. Если выявлено больше изменений в образце плазмы, чем в конститутивной ДНК, по сравнению с референсным геномом, то существует повышенная вероятность рака. В одном варианте реализации исследуют гомозиготные локусы в референсном геноме. Сравнивают количества гетерозиготных локусов в конститутивной ДНК и в ДНК из биологического образца. Если количество гетерозиготных сайтов, детектированных в ДНК из биологического образца, превышает таковое в конститутивной ДНК, тогда существует повышенная вероятность рака.

Можно также ограничиться анализом локусов, которые гомозиготны в КГ. ОНМ также можно определить для гетерозиготных локусов, но это, как правило, потребует образования третьего варианта. Другими словами, если гетерозиготный локус представляет собой А/Т, то новый третий вариант будет либо С, либо G. Идентифицировать ОНМ в гомозиготных локусах, как правило, проще.

Степень, до которой увеличивается количество гетерозиготных локусов в биологическом образце ДНК по сравнению с конститутивной ДНК, может предполагать наличие рака или предзлокачественного состояния, если ее сравнить с уровнем изменений, наблюдаемым у здоровых субъектов. Например, если степень увеличения количества таких сайтов превышает таковую, наблюдаемую у здоровых субъектов, на некоторый порог, то можно считать, что результаты свидетельствуют о наличии рака или предзлокачественного состояния. В одном варианте реализации устанавливают распределение мутаций у субъектов без ракового заболевания, и в качестве порога можно взять некоторое количество стандартных отклонений (например, 2 или 3 стандартных отклонения).

В одном варианте реализации может потребоваться, по меньшей мере, определенное количество вариантов в локусе до того, как локус будет учтен. В другом варианте реализации анализируют даже результаты, полученные на основании одной замены.

Например, если суммарное количество вариаций (ошибок + истинных мутаций или полиморфизмов), наблюдаемых в плазме, статистически значимо выше, чем в конститутивной ДНК, то это свидетельствует о наличии рака.

На фиг. 3 показана блок-схема способа 300 сравнения генома образца (ГО) с конститутивным геномом (КГ) с использованием референсного генома (ЭГ) согласно вариантам реализации настоящего изобретения. В способе 300 предполагают, что ЭГ уже получен и что маркеры последовательности в биологическом образце уже получены.

В блоке 310 по меньшей мере часть маркеров последовательности выравнивают с референсным геномом. Выравнивание может позволять наличие несовпадений, так как необходимо обнаружить вариации. Референсный геном можно получить из популяции, сходной с субъектом. Выровненные маркеры последовательности фактически включают геном образца (ГО).

В блоке 320 выявляют первое количество (А) потенциальных вариантов, например однонуклеотид-

ных мутаций (ОНМ). Потенциальные ОНМ представляют собой локусы, в которых в маркере последовательности ГО выявлен нуклеотид, отличный от такового в ЭГ. Можно использовать другие критерии, например, количество маркеров последовательности, в которых выявлена вариация, должно быть больше, чем пороговое значение, и является ли локус в ЭГ гомозиготным. Набор потенциальных ОНМ можно представить как набор А, когда определенные локусы обнаруживают и отслеживают путем хранения локусов в памяти. Можно выявить определенные локусы или можно просто определить количество таких ОНМ.

В блоке 330 конститутивный геном определяют путем выравнивания маркеров последовательности, полученных путем секвенирования фрагментов ДНК из конститутивного образца, с референсным геномом. Данный этап мог быть осуществлен в любой более ранний момент и с применением конститутивного образца, полученного в любой более ранний момент времени. КГ можно просто считать из памяти, если выравнивание было выполнено ранее. В одном варианте реализации конститутивный образец может представлять собой клетки крови.

В блоке 340 обнаруживают второе количество (В) локусов, в которых выровненный маркер последовательности КГ содержит вариант (например, ОНМ) в локусе по сравнению с референсным геномом. Если специально отслеживают набор локусов, тогда В может представлять собой набор, а не просто число.

В блоке 350 набор В вычитают из набора А, чтобы определить варианты (ОНМ), которые присутствуют в геноме образца, но не в КГ. В одном варианте реализации набор ОНМ можно ограничить положениями нуклеотидов, в которых КГ гомозиготный. Для того чтобы осуществить такой отбор, определенные локусы, в которых КГ гомозиготный, можно обозначить как набор С. В другом варианте реализации локус не учитывается в первом количестве А или втором количестве В, если КГ не является гомозиготным в данном локусе. В другом варианте реализации любой известный полиморфизм (например, благодаря его наличию в базе данных ОНП) можно отсеять.

В одном варианте реализации вычитание в блоке 350 может быть просто вычитанием чисел и, следовательно, конкретные потенциальные ОНМ не удаляют, а просто вычитают их значение. В другом варианте реализации вычитание дает разность между набором А и набором В (например, если набор В представляет собой подмножество набора А), чтобы определить конкретные ОНМ, которые не представлены в наборе В. В логических переменных это можно выразить как $[A \text{ И НЕ}(B)]$. Полученный набор детектированных вариантов можно обозначить С. Указанный параметр можно определить как число С или определить из набора С.

В некоторых вариантах реализации можно принимать во внимание природу мутаций и различным классам мутаций приписывать определенный вес. Например, мутациям, которые обычно связаны с раком, можно приписывать больший вес (также называемый уровнем значимости в отношении относительных весов локусов). Такие мутации можно найти в базах данных связанных с опухолью мутаций, например в Каталоге соматических мутаций при раке (COSMIC) (www.sanger.ac.uk/genetics/CGP/cosmic/). В качестве другого примера мутациям, связанным с несинонимичными изменениями, можно приписывать больший вес.

Таким образом, первое количество А можно определить как взвешенную сумму, в которой вес количества маркеров с вариантом в одном локусе может быть отличным от такового для количества маркеров в другом локусе. Первое количество А может отражать такую взвешенную сумму. Аналогичный расчет можно осуществить с В, и, таким образом, количество С и указанный параметр могут отражать такое взвешивание. В другом варианте реализации веса учитывают, когда определяют набор С конкретных локусов. Например, взвешенную сумму можно определить для количества локусов набора С. Такие веса можно применять в других способах, описанных в данной заявке.

Соответственно, параметр, который сравнивают с пороговым значением, чтобы определить классификацию уровня рака, может представлять собой количество локусов, в которых детектирована вариация в ГО и КГ по сравнению с ЭГ. В других вариантах реализации можно сосчитать суммарное количество фрагментов ДНК (которое определяют с помощью маркеров последовательности), содержащих вариацию. В других вариантах реализации такие количества можно использовать в другой формуле для получения указанного параметра.

В одном варианте реализации концентрация варианта в каждом локусе может представлять собой параметр, и ее можно сравнить с пороговым значением. Такое пороговое значение можно применять для определения того, является ли локус потенциально измененным локусом (вдобавок к пороговому значению определенного количества считываемых последовательностей, в которых выявили указанный вариант), а затем обсчитать данный локус. Концентрацию также можно использовать в качестве фактора взвешивания в сумме ОНМ.

VI. Уменьшение количества ложноположительных результатов с помощью пороговых значений.

Выше упоминалось, что однонуклеотидные мутации можно наблюдать в большом количестве внеклеточных фрагментов ДНК (например, циркулирующей в плазме ДНК) из большой области генома (например, из всего генома) или из множества областей генома, чтобы улучшить чувствительность данного подхода. Тем не менее, аналитические ошибки, такие как ошибки секвенирования, могут влиять на осущест्वимость, достоверность и специфичность данного подхода. Здесь мы используем платформу для

массового параллельного секвенирования в качестве примера, чтобы проиллюстрировать значение ошибок секвенирования. Частота ошибок секвенирования платформы для секвенирования путем синтеза от Illumina составляет приблизительно от 0,1 до 0,3% на секвенированный нуклеотид (Minoche и др. *Genome Biol.* 2011, 12:R112). Можно применять любую платформу для массового параллельного секвенирования, включая платформу для секвенирования путем лигирования (например, платформу SOLiD от Life Technologies), платформу Ion Torrent/Ion Proton, платформу для полупроводникового секвенирования, Roche 454, платформу для секвенирования одиночных молекул (например, Helicos, Pacific Biosciences и Nanopore).

В более раннем исследовании печеночно-клеточной карциномы показали, что во всем раковом геноме присутствует приблизительно 3000 однонуклеотидных мутаций (Tao Y. и др. 2011 *Proc. Natl. Acad. Sci. USA*; 108: 12042-12047). Допуская, что лишь 10% всей ДНК в кровотоке произошло из опухолевых клеток и что мы секвенировали ДНК в плазме со средней глубиной секвенирования, составляющей однократное покрытие гаплоидного генома, мы встретим 9 миллионов ($3 \times 10^9 \times 0,3\%$) однонуклеотидных вариаций (ОНВ) вследствие ошибок секвенирования. Тем не менее, ожидается, что большинство из однонуклеотидных мутаций встретится только на одной из двух гомологичных хромосом. При глубине секвенирования, составляющей однократное покрытие гаплоидного генома образца со 100% опухолевой ДНК, мы ожидаем обнаружить лишь половину из 3000 мутаций, т.е. 1500 мутаций. При секвенировании образца плазмы, содержащего 10% ДНК опухолевого происхождения на одно покрытие гаплоидного генома, мы ожидаем обнаружить лишь 150 ($1500 \times 10\%$) связанных с раком однонуклеотидных мутаций. Таким образом, отношение сигнал/шум для детектирования связанных с раком мутаций составляет 1 на 60000. Такое очень низкое отношение сигнал/шум позволяет предположить, что достоверность применения данного подхода для того, чтобы различить нормальных и раковых пациентов, будет очень низкой, если мы используем в качестве параметра просто все однонуклеотидные изменения в биологическом образце (например, в плазме).

Ожидается, что с прогрессированием технологий секвенирования частота ошибок секвенирования будет постоянно уменьшаться. Также можно проанализировать один и тот же образец, применяя более чем одну платформу секвенирования, и посредством сравнения результатов секвенирования на разных платформах точно определить считываемые последовательности, вероятно содержащие ошибки секвенирования. Другой подход состоит в анализе двух образцов, взятых в различные моменты времени из одного и того же субъекта. Тем не менее, такие подходы требуют больших затрат времени.

В одном варианте реализации одним из способов увеличения отношения сигнал/шум при детектировании однонуклеотидных мутаций в плазме пациентов с раковыми заболеваниями является учет только такой мутации, которая встречается в образце несколько раз. В выбранных платформах секвенирования могут быть наиболее распространены ошибки секвенирования, включающие определенные замены нуклеотидов, и они будут влиять на результаты секвенирования тестируемого образца и конститутивного образца ДНК как тестируемого субъекта, так и контрольных субъектов. Тем не менее, как правило, ошибки секвенирования происходят случайным образом.

Вероятность ошибки секвенирования экспоненциально уменьшается, когда наблюдают одну и ту же замену в одном и том же положении нуклеотида в нескольких фрагментах ДНК. С другой стороны, на вероятность детектирования истинного связанного с раком мутационного изменения в образце влияет глубина секвенирования и фракционная концентрация опухолевой ДНК в образце. Вероятность детектирования мутации в нескольких фрагментах ДНК будет повышаться с увеличением глубины секвенирования и фракционной концентрации опухолевой ДНК. В различных вариантах реализации с применением образцов, содержащих внеклеточную опухолевую ДНК (таких как плазма), фракционная концентрация может составлять 5, 10, 20 и 30%. В одном варианте реализации фракционная концентрация составляет менее чем 50%.

На фиг. 4 представлена таблица 400, в которой показано количество связанных с раком однонуклеотидных мутаций, правильно детектированных с применением встречаемости различных их количеств в качестве критерия для классификации мутации как присутствующей в образце согласно вариантам реализации настоящего изобретения. Также показаны количества положений нуклеотидов, в которых ошибочно обнаружили мутацию вследствие ошибки секвенирования, основываясь на тех же критериях классификации. Частоту ошибок секвенирования принимают равной 0,1% (Minoche и др. *Genome Bio* 2011, 12:R112). Фракционную концентрацию ДНК опухолевого происхождения в образце принимают равной 10%.

На фиг. 4 показано, что отношение между количеством связанных с раком мутаций, детектированных в плазме, и количеством ложноположительных сигналов будет экспоненциально возрастать с увеличением количества наблюдений одной и той же замены в образце, в котором выявляют мутацию, когда фракционную концентрацию ДНК опухолевого происхождения в образце принимают равной 10%. Другими словами, как чувствительность, так и специфичность детектирования связанной с раком мутации будет улучшаться. Кроме того, на чувствительность детектирования связанных с раком мутаций будет влиять глубина секвенирования. При секвенировании со 100-кратным покрытием гаплоидного генома можно обнаружить 2205 (73,5%) из 3000 мутаций, даже с применением критерия встречаемости кон-

кретной мутации по меньшей мере в 4 фрагментах ДНК в образце. Можно использовать другие значения минимального количества фрагментов, такие как 3, 5, 8, 10 и больше 10.

На фиг. 5 представлена таблица 500, в которой показано ожидаемое количество ложноположительных локусов и ожидаемое количество мутаций, детектированных когда фракционную концентрацию ДНК опухолевого происхождения в образце принимают равной 5%. При более низкой фракционной концентрации ДНК опухолевого происхождения в образце, будет необходима большая глубина секвенирования для достижения такой же чувствительности детектирования связанных с раком мутаций. Для сохранения специфичности также будет необходим более строгий критерий. Например, потребуются использовать критерий встречаемости в образце отдельной мутации в по меньшей мере 5 фрагментах ДНК, вместо критерия встречаемости в по меньшей мере 4 фрагментах в случае фракции опухолевой ДНК, составляющей 10%. В таблицах 400 и 500 предложено руководство по выбору порогового значения с учетом кратности покрытия и концентрации опухолевой ДНК, которые можно предположить или измерить, как описано в данной заявке.

Другим преимуществом применения критериев более чем однократного детектирования однонуклеотидной замены для определения мутации является то, что в результате этого ожидают минимизацию детектирования ложноположительных результатов, которые дают однонуклеотидные изменения в незлокачественных тканях. Так как нуклеотидные изменения могут происходить в процессе митоза нормальных клеток, каждая здоровая клетка в организме может содержать множество однонуклеотидных изменений. Данные изменения потенциально могут приводить к ложноположительным результатам. Тем не менее, изменения из клетки будут попадать в плазму/сыворотку после гибели клетки. Хотя ожидается, что различные нормальные клетки несут различные наборы мутаций, маловероятно, что мутации, встречающиеся в одной клетке, будут присутствовать в множестве копий в плазме/сыворотке. В противоположность этому в плазме/сыворотке ожидают обнаружить множество копий мутаций из опухолевых клеток, так как рост опухоли по природе клональный. Таким образом, множество клеток из одного клона будет погибать и высвобождать ДНК с характерными мутациями, типичными для данных клонов.

В одном варианте реализации перед проведением секвенирования можно осуществить целевое обогащение определенными областями генома. Данный этап целевого обогащения может увеличить глубину секвенирования интересующих областей при таком же суммарном количестве проведенных секвенирований. В еще одном варианте реализации сначала можно осуществить раунд секвенирования с относительно небольшой глубиной секвенирования. Затем материал можно обогатить областями, в которых выявлена по меньшей мере одна однонуклеотидная замена, а затем провести второй раунд секвенирования с большей кратностью покрытия. Затем к результатам секвенирования с целевым обогащением можно применить критерий множественной встречаемости для определения мутации.

VII. Динамические пороговые значения.

Выше описано, что можно применять пороговое значение N для количества считываемых последовательностей, в которых детектирован вариант (потенциальная мутация), чтобы определить, можно ли признать, что локус содержит мутацию (например, содержит ОНМ), которую необходимо учесть. Применение такого порога может уменьшить количество ложноположительных результатов. В обсуждении ниже предложены способы выбора порогового значения для различных локусов. В следующих вариантах реализации мы предполагаем, что существует один преобладающий раковый клон. Аналогичный анализ можно осуществить для случаев, включающих несколько клонов раковых клеток, высвобождающих различные количества опухолевой ДНК в плазму.

A. Количество связанных с раком мутаций, детектированных в плазме.

На количество связанных с раком мутаций, детектированных в плазме, может влиять множество параметров, например: (1) количество мутаций в опухолевой ткани (N_T) - суммарное количество мутаций, присутствующих в опухолевой ткани, представляет собой максимальное количество связанных с опухолью мутаций, детектируемых в плазме пациента; (2) фракционная концентрация ДНК опухолевого происхождения в плазме (f), чем больше фракционная концентрация ДНК опухолевого происхождения в плазме, тем выше будет вероятность детектирования в плазме связанных с опухолью мутаций; (3) глубина секвенирования (D) - глубина секвенирования относится к количеству раз, которое считывают последовательность при секвенировании области. Например, средняя глубина секвенирования, составляющая 10 раз, означает, что каждый нуклеотид в секвенированной области покрывается в среднем 10 считываемыми последовательностями. Вероятность детектирования связанной с раком мутации будет возрастать при увеличении глубины секвенирования; и (4) минимальное количество раз, которое нуклеотидную замену обнаруживают в плазме, чтобы определить, что она является потенциально связанной с раком мутацией (r), которое представляет собой пороговое значение, используемое для того, чтобы отличить ошибки секвенирования от реальных связанных с раком мутаций.

В одном варианте реализации для того, чтобы прогнозировать количество связанных с раком мутаций, детектированных в плазме, используют пуассоновское распределение. Допуская, что мутация присутствует в некотором положении нуклеотида на одной из двух гомологичных хромосом, при глубине секвенирования D , ожидаемое количество раз, которое мутация присутствует в плазме (M_p), рассчитывают как

$$M_p = D \times f / 2.$$

Вероятность детектирования мутации в плазме (Pb) в отдельном сайте мутации рассчитывают как

$$P_b = 1 - \sum_{i=1}^{r-1} \text{Пуассон}(i, M_p)$$

где r (пороговое значение) представляет собой количество раз, которое нуклеотидная замена детектирована в плазме, для того чтобы определить, что она представляет собой потенциально связанную с опухолью мутацию; Пуассон(i, M_p) представляет собой пуассоновское распределение вероятности встречаемости i раз при среднем количестве M_p .

Суммарное количество связанных с раком мутаций, которые ожидают обнаружить в плазме (N_p), можно рассчитать как

$$N_p = N_T \times P_b,$$

где N_T представляет собой количество мутаций, присутствующих в опухолевой ткани. На следующих диаграммах показаны проценты связанных с опухолью мутаций, которые ожидают обнаружить в плазме, применяя различные критерии встречаемости (r) для определения потенциальной мутации и различные глубины секвенирования.

На фиг. 6А представлена диаграмма 600, на которой показана частота детектирования связанных с раком мутаций в плазме с фракционными концентрациями ДНК опухолевого происхождения, составляющими 10 и 20%, и с применением встречаемостей (r), равных четырем и шести, в качестве критериев для определения потенциальных связанных с раком мутаций. При одной и той же r более высокая фракционная концентрация ДНК опухолевого происхождения в плазме приведет к детектированию в плазме большего количества связанных с раком мутаций. При одной и той же фракционной концентрации ДНК опухолевого происхождения в плазме более высокая r приведет к детектированию меньшего количества мутаций.

В. Количество ложноположительных однонуклеотидных изменений, детектированных вследствие ошибки.

Однонуклеотидные изменения в ДНК плазмы в результатах секвенирования могут встречаться вследствие ошибок секвенирования и выравнивания. Количество положений нуклеотидов с ложноположительными однонуклеотидными изменениями можно предсказать математически на основании биномиального распределения. Параметры, влияющие на количество ложноположительных сайтов (N_{FF}), могут включать: (1) частоту ошибок секвенирования (E) - частоту ошибок секвенирования определяют как долю неправильно определенных при секвенировании нуклеотидов; (2) глубину секвенирования (D) - чем больше глубина секвенирования, тем больше будет количество положений нуклеотидов, в которых выявится ошибка секвенирования; (3) минимальное количество детектирований одной и той же нуклеотидной замены, чтобы определить, что она является потенциально связанной с раком мутацией (r); и (4) суммарное количество положений нуклеотидов внутри интересующего фрагмента (N_I).

Возникновение мутаций, как правило, считают случайным процессом. Следовательно, с увеличением критериев встречаемости для определения потенциальной мутации, количество ложноположительных положений нуклеотидов будет экспоненциально уменьшаться. В некоторых из существующих платформ секвенирования определенное окружение последовательности в большей степени подвержено ошибкам секвенирования. Примеры такого окружения последовательности включают мотив GGC, гомополимеры (например, AAAAAA) и простые повторы (например, ATATATATAT). Такие окружения последовательности по существу будут увеличивать количество ложных однонуклеотидных замен или вставок/делений (Nakamura K. и др. Nucleic. Acids Res. 2011;39,e90 и Minoche A.E. и др. Genome Biol. 2011;12,R112). Кроме того, последовательности повторов, такие как гомополимеры и простые повторы, при вычислениях будут вызывать неоднозначность выравнивания и, следовательно, приводить к ложноположительным результатам для однонуклеотидных вариаций.

Чем больше интересующий участок, тем большее количество ложноположительных положений нуклеотидов будут наблюдать. Если необходимо найти мутации во всем геноме, тогда интересующая область будет представлять собой весь геном и количество исследуемых нуклеотидов будет составлять 3 миллиарда. С другой стороны, если сфокусировать внимание на экзонах, то интересующая область будет состоять из нуклеотидов, кодирующих экзоны, количество которых приблизительно составляет 45 миллионов.

Количество ложноположительных положений нуклеотидов, связанных с ошибками секвенирования, можно определить на основании следующих расчетов. Вероятность (P_{Er}) того, что одна и та же нуклеотидная замена встретится в одном и том же положении вследствие ошибок секвенирования, можно рассчитать как

$$P_{Er} = C(D, r) E \left(\frac{E}{3}\right)^{r-1}$$

где C(D,r) представляет собой количество возможных комбинаций выборки r элементов из общего количества элементов D; r представляет собой значение встречаемости для определения потенциальной мутации; D представляет собой глубину секвенирования; и E представляет собой частоту ошибок секве-

нирования. $C(D, r)$ можно рассчитать как

$$C(D, r) = \frac{D_1}{r_1(D - r_1)}$$

Количество положений нуклеотидов (N_{FP}), ложноположительных по мутациям, можно рассчитать как

$$N_{FP} = N_1 P_{Er}$$

где N_1 представляет собой суммарное количество положений нуклеотидов в интересующей области.

На фиг. 6B представлена диаграмма 650, на которой показано ожидаемое количество положений нуклеотидов, которые ошибочно классифицировали как содержащие нуклеотидную замену с применением критериев встречаемости (r) 4, 5, 6 и 7 в зависимости от глубины секвенирования. В данном расчете за интересующий участок принимают целый геном (3 миллиарда положений нуклеотидов). Частоту ошибок секвенирования принимают равной 0,3% от секвенированных нуклеотидов. Видно, что значение r оказывает существенное влияние на ложноположительные результаты. Но по фиг. 6A можно увидеть, что при большем значении r также уменьшается количество детектированных мутаций, по меньшей мере, до тех пор, пока не будут применяться значительно большие глубины секвенирования.

В. Выбор минимальной встречаемости (r).

Выше обсуждалось, что количество истинных связанных с раком мутированных сайтов и ложноположительных сайтов вследствие ошибок секвенирования будет увеличиваться с увеличением глубины секвенирования. Тем не менее, уровни их увеличения будут различны. Следовательно, можно использовать выбор глубины секвенирования и значения r для того, чтобы максимизировать детектирование истинных связанных с раком мутаций, при этом сохраняя количество ложноположительных сайтов на низком уровне.

На фиг. 7A представлена диаграмма 700, на которой показано количество истинных связанных с раком мутированных сайтов и ложноположительных сайтов для различных глубин секвенирования. Суммарное количество связанных с раком мутаций в опухолевой ткани принимают равным 3000 и фракционную концентрацию ДНК опухолевого происхождения в плазме принимают равной 10%. Частоту ошибок секвенирования принимают равной 0,3%. На подписи к фигуре ИП обозначает истинно положительные сайты, в которых соответствующая мутация присутствует в опухолевой ткани, ЛП обозначает ложноположительные сайты, в которых в опухолевой ткани присутствует не соответствующая мутация и изменения нуклеотидов, присутствующие в результатах секвенирования, выявлены вследствие ошибок секвенирования.

По диаграмме 700 видно, что при 110-кратной глубине секвенирования мы обнаружим приблизительно 1410 истинных связанных с раком мутаций, если в качестве критерия мы используем минимальную встречаемость, равную 6 ($r=6$), чтобы определить потенциальный сайт мутации в плазме. Используя данный критерий, мы обнаружим лишь приблизительно 20 ложноположительных сайтов. Если в качестве критерия мы используем минимальную встречаемость, равную 7 ($r=7$), чтобы определить потенциальную мутацию, количество связанных с раком мутаций, которые можно обнаружить, уменьшится на 470 (до приблизительно 940). Следовательно, критерий $r=6$ позволит с большей чувствительностью обнаружить в плазме связанные с раком мутации.

С другой стороны, при 200-кратной глубине секвенирования количество детектированных истинных связанных с раком мутаций будет составлять приблизительно 2800 и 2600, если мы используем для определения потенциальных мутаций критерии минимальной встречаемости (r), равные 6 и 7, соответственно. Используя два данных значения r , мы получим количества ложноположительных сайтов приблизительно 740 и 20 соответственно. Следовательно, при 200-кратной глубине секвенирования применение более строгого критерия $r=7$ для определения потенциальной мутации позволит существенно уменьшить количество ложноположительных сайтов, при этом не оказывая значительного отрицательного влияния на чувствительность детектирования истинных связанных с раком мутаций.

С. Динамическое пороговое значение для результатов секвенирования для детектирования потенциальных мутаций в плазме.

Глубина секвенирования каждого нуклеотида в интересующей области будет различна. Если мы будем применять фиксированное пороговое значение для встречаемости нуклеотидной замены, чтобы определить потенциальную мутацию в плазме, для нуклеотидов, которые покрыты большим количеством считываемых последовательностей (т.е. секвенированы с большей глубиной), будут выше вероятности ошибочного детектирования в них вариации вследствие ошибки секвенирования, при отсутствии такого изменения в опухолевой ткани, по сравнению с нуклеотидами, которые секвенированы с меньшей глубиной. В одном варианте реализации для преодоления данной проблемы применяют динамическое пороговое значение r к различным положениям нуклеотидов, в соответствии с фактической глубиной секвенирования конкретного положения нуклеотида и в соответствии с желательным верхним пределом вероятности определения ложноположительных вариаций.

В одном варианте реализации максимально допустимый уровень ложноположительных вариаций можно зафиксировать как 1 на $1,5 \times 10^8$ положений нуклеотидов. При данном максимально допустимом

уровне ложноположительных вариаций суммарное количество ложноположительных сайтов, которые будут обнаруживаться во всем геноме, будет менее 20. Значение g для различных глубин секвенирования можно определить по кривым, показанным на фиг. 6В, и данные пороговые значения показаны в табл. 1. В других вариантах реализации можно применять другие различные максимально допустимые уровни ложноположительных вариаций, например 1 на 3×10^8 , 1 на 10^8 или 1 на 6×10^7 . Соответствующее суммарное количество ложноположительных сайтов будет составлять менее чем 10, 30 и 50 соответственно.

Таблица 1. Минимальное количество раз, которое нуклеотидную замену обнаруживают в плазме, чтобы определить, что она является потенциальной мутацией (g), для различных глубин секвенирования конкретного положения нуклеотида

Глубина секвенирования конкретного положения нуклеотида	Минимальное количество раз, которое нуклеотидную замену обнаруживают в результатах секвенирования ДНК в плазме, чтобы определить, что она является потенциальной мутацией (g)
<50	5
50 – 110	6
111 – 200	7
201 – 310	8
311 – 450	9
451 – 620	10
621 – 800	11

Максимальный уровень ложноположительных вариаций зафиксирован как 1 на $1,5 \times 10^8$ нуклеотидов.

D. Секвенирование с целевым обогащением.

На фиг. 7А показано, что большая глубина секвенирования может привести к лучшей чувствительности обнаружения связанных с раком мутаций, при этом количество ложноположительных сайтов сохраняется на низком уровне благодаря возможности применения более высокого значения g . Например, при 110-кратной глубине секвенирования можно детектировать в плазме 1410 истинных связанных с раком мутаций, применяя значение g , равное 6, тогда как количество обнаруженных истинных связанных с раком мутаций будет составлять 2600, если глубину секвенирования увеличить до 200 раз и применить значение g , равное 7. Два указанных набора результатов дадут ожидаемое количество ложноположительных сайтов, приблизительно равное 20.

Хотя секвенирование всего генома до 200-кратной глубины относительно дорогостояще в настоящее время, одним возможным способом достижения такой глубины секвенирования является фокусировка на меньшей интересующей области. Анализ целевой области можно осуществить, например, не ограничиваясь перечисленными способами, путем применения затравок ДНК или РНК для захвата интересующих областей генома посредством гибридизации. Захваченные области затем осаждают, например, с помощью магнитных средств, а затем подвергают секвенированию. Такой целевой захват можно осуществить, например, применяя систему целевого обогащения Agilent SureSelect, систему целевого обогащения Roche Nimblegen и систему целевого ресеквенирования Illumina. Другим подходом является проведение амплификации целевых областей с помощью ПНР, а затем проведение секвенирования. В одном варианте реализации интересующая область представляет собой экзом. В таком варианте реализации можно осуществить целевой захват всех экзонов ДНК в плазме, и обогащенную экзонными областями ДНК из плазмы затем можно секвенировать.

Вдобавок к возможности большей глубины секвенирования фокусировка на определенных областях вместо анализа всего генома позволит значительно уменьшить количество положений нуклеотидов в области поиска и приведет к уменьшению количества ложноположительных сайтов при такой же частоте ошибок секвенирования.

На фиг. 7В представлена диаграмма 750, на которой показано прогнозируемое количество ложноположительных сайтов при проведении анализа полного генома (ПГ) и всех экзонов. Для каждого типа анализа использовали два различных значения g , равные 5 и 6. При 200-кратной глубине секвенирования, если для определения мутаций в плазме используют $g = 5$, прогнозируемое количество ложноположительных сайтов составляет приблизительно 23000 и 230 для всего генома и всех экзонов соответственно. Если для определения мутаций в плазме используют $g = 6$, прогнозируемое количество ложноположительных сайтов составляет 750 и 7 соответственно. Следовательно, ограничение количества нуклеотидов в интересующей области может значительно уменьшить количество ложноположительных сайтов при анализе наличия мутаций в ДНК плазмы.

При секвенировании с захватом экзона или даже захватом экзома количество нуклеотидов в области поиска уменьшается. Следовательно, даже если мы позволим более высокий уровень ложноположительных вариаций для детектирования связанных с раком мутаций, абсолютное количество ложноположительных сайтов можно сохранить на относительно низком уровне. Допущение более высокого уровня ложноположительных вариаций позволит применять менее строгий критерий минимальной встречаемости (g) для детектирования однонуклеотидных вариаций в плазме. Это приведет к более высокой чувст-

вительности детектирования истинных связанных с раком мутаций.

В одном варианте реализации мы можем применять максимально допустимый уровень ложноположительных вариаций, равный $1,5 \times 10^6$. При данном уровне ложноположительных вариаций суммарное количество ложноположительных сайтов внутри целевых экзонов будет составлять лишь 20. Значения τ для различных глубин секвенирования с применением максимально допустимого уровня ложноположительных вариаций, равного $1,5 \times 10^6$, показаны в табл. 2. В других вариантах реализации можно применять другие различные максимально допустимые уровни ложноположительных вариаций, например 1 на 3×10^6 , 1 на 10^6 или 1 на 6×10^5 . Соответствующее суммарное количество ложноположительных сайтов будет меньше чем 10, 30 и 50 соответственно. В одном варианте реализации различным классам мутаций можно приписывать различные веса, описанные выше.

Таблица 2. Минимальное количество раз, которое нуклеотидная замена присутствует в плазме, для того чтобы определить, что она является потенциальной мутацией (τ), для различных глубин секвенирования конкретного положения нуклеотида

Глубина секвенирования конкретного положения нуклеотида	Минимальное количество раз, которое нуклеотидную замену обнаруживают в результатах секвенирования ДНК в плазме, чтобы определить, что она является потенциальной мутацией (τ)
<50	4
50 – 125	5
126 – 235	6
236 – 380	7
381 – 560	8
561 – 760	9

Максимальный уровень ложноположительных вариаций зафиксирован на уровне 1 на $1,5 \times 10^6$ нуклеотидов.

VIII. Детектирование рака.

Выше упоминалось, что количество маркеров последовательности в измененных локусах можно применять для определения различными способами параметра, который сравнивают с порогом, чтобы классифицировать уровень рака. Фракционная концентрация вариантных считываемых последовательностей по сравнению со всеми считываемыми последовательностями в локусе или во множестве локусов представляет собой другой параметр, который можно использовать для этой цели. Ниже представлены некоторые примеры вычисления указанного параметра и порога.

A. Определение параметра.

Если КГ гомозиготен в определенном локусе по первой аллели и в биологическом образце (например, в плазме) наблюдается вариантная аллель, то фракционную концентрацию можно рассчитать как $2p/(p+q)$,

где p представляет собой количество маркеров последовательности, содержащих вариантную аллель, и q представляет собой количество маркеров последовательности, содержащих первую аллель КГ. В данной формуле предполагается, что только в одном из гаплотипов опухоли присутствует вариант, что, как правило, будет соответствовать действительности. Таким образом, для каждого гомозиготного локуса можно рассчитать фракционную концентрацию. Фракционные концентрации можно усреднить. В другом варианте реализации число p может включать количество маркеров последовательности для всех локусов и, аналогично числу q , определять фракционную концентрацию. Далее в настоящем разделе представлен пример.

Осуществляли полногеномное исследование наличия однонуклеотидных вариантов (ОНВ) опухолевого происхождения в плазме 4 пациентов с ГМК. Мы секвенировали опухолевую ДНК и ДНК из лейкоцитарной пленки со средними глубинами, равными 29,5-кратному (диапазон от 27- до 33-кратного) и 43-кратному (диапазон от 39- до 46-кратного) покрытию гаплоидного генома соответственно. Сравнивали результаты массового параллельного секвенирования (МПС) опухолевой ДНК и ДНК из лейкоцитарной пленки для каждого из 4 пациентов с ГМК и выявляли ОНВ, присутствующие в опухолевой ДНК, но не в ДНК из лейкоцитарной пленки, с помощью строгого биоинформационного алгоритма. Для данного алгоритма требуется, чтобы предполагаемый ОНВ присутствовал по меньшей мере в пороговом количестве секвенированных фрагментов опухолевой ДНК (т.е. в соответствующем секвенированном маркере), чтобы его можно было классифицировать как истинный ОНВ. Пороговое количество определяют путем учета глубины секвенирования конкретного нуклеотида и частоты ошибок секвенирования, например, описанных в данной заявке.

На фиг. 8 представлена таблица 800, на которой показаны результаты для 4 пациентов с ГМК до и после лечения, включая фракционные концентрации ДНК опухолевого происхождения в плазме согласно вариантам реализации настоящего изобретения. Количество связанных с опухолью ОНВ находилось в

диапазоне от 1334 до 3171 в 4 случаях ГКК. Приведен перечень соотношений таких ОНВ, которые были детектированы в плазме до и после лечения. До лечения в плазме обнаружили 15-94% связанных с опухолью ОНВ. После лечения их процент составлял 1,5-5,5%. Таким образом, количество детектированных ОНВ действительно коррелирует с уровнем рака. Таким образом, показали, что количество ОНВ можно применять в качестве параметра для классификации уровня рака.

Фракционные концентрации ДНК опухолевого происхождения в плазме определяли путем фракционного подсчета количества мутантных последовательностей по сравнению с суммарным количеством последовательностей (т.е. мутантных плюс дикого типа). Формула для такого расчета следующая:

$$2p/(p+q),$$

где 2 учитывает, что в опухоли мутирован только один гаплотип. Полученные фракционные концентрации хорошо коррелировали с таковыми, определенными с помощью полногеномного анализа суммарной потери аллелей (GAAL) (Chan K.C. и др. Clin. Chem. 2013; 59:211-24), и уменьшались после хирургического вмешательства. Таким образом, показали, что фракционная концентрация также является пригодным параметром для определения уровня рака.

Фракционная концентрация, полученная путем анализа ОНВ, может выражать опухолевую нагрузку. У пациента с раковым заболеванием с большей опухолевой нагрузкой (например, с большей выявленной фракционной концентрацией ДНК опухолевого происхождения) будет повышена частота соматических мутаций по сравнению с пациентом с меньшей опухолевой нагрузкой. Таким образом, варианты реализации также можно применять для прогнозирования. Обычно у пациентов с раковыми заболеваниями с большей опухолевой нагрузкой прогноз хуже, чем у пациентов с меньшей опухолевой нагрузкой. Таким образом, у первой из упомянутых групп (у пациентов с раковыми заболеваниями с большей опухолевой нагрузкой) будет повышенный риск умереть от указанного заболевания. В некоторых вариантах реализации, если можно определить абсолютную концентрацию ДНК в биологическом образце, например в плазме (например, применяя ПНР в реальном времени или флуорометрию), то абсолютную концентрацию связанных с опухолью генетических аберраций можно определить и использовать для клинического детектирования, и/или мониторинга, и/или прогнозирования.

В. Определение порога.

Таблицу 800 можно использовать для определения порога. Выше упоминалось, что количество ОНВ и фракционная концентрация, определенная с помощью анализа ОНВ, коррелирует с уровнем рака. Порог можно определить в индивидуальном порядке. Например, для определения порога можно использовать значение до лечения. В различных вариантах реализации порог может представлять собой абсолютную величину изменения относительно значения до лечения. Подходящим порогом может быть уменьшение количества ОНВ или фракционной концентрации на 50%. Такой порог позволит классифицировать более низкий уровень рака для каждого из случаев в таблице 800. Следует отметить, что такой порог может зависеть от глубины секвенирования.

В одном варианте реализации порог можно применять для всех образцов, и он может учитывать или может не учитывать значения данного параметра до лечения. Например, порог из 100 ОНВ можно применять для классификации субъекта как не имеющего рак или имеющего низкий уровень рака. Данный порог из 100 ОНВ удовлетворяется каждым из четырех случаев в таблице 800. Если в качестве параметра применяют фракционную концентрацию, при пороге, составляющем 1,0%, у ГКК1-ГКК3 будут классифицировать практически нулевой уровень рака, и при втором пороге, составляющем 1,5%, у ГКК4 будут классифицировать низкий уровень рака. Таким образом, более чем один порог можно применять для получения более чем двух классификаций.

Чтобы наглядно продемонстрировать другие возможные пороговые значения, мы проанализировали наличие в плазме здоровых контролей связанных с опухолью ОНВ. У здоровых субъектов можно осуществить множество измерений, чтобы определить диапазон количества вариаций, ожидаемых в биологическом образце по сравнению с конститутивным геномом.

На фиг. 9 представлена таблица 900, на которой показано детектирование связанных с ГКК ОНВ у 16 здоровых контрольных субъектов согласно вариантам реализации настоящего изобретения. Таблицу 900 можно применять для оценки специфичности подхода анализа ОНВ. 16 здоровых контролей перечислены в различных рядах. В колонках приведены ОНВ, детектированные для конкретных пациентов с ГКК, и показаны количества считываемых последовательностей в измененных локусах, содержащих вариантную аллель, и количество считываемых последовательностей с аллелью дикого типа (т.е. аллелью из КГ). Например, для ГКК1 в контроле С01 было 40 вариантных считываемых последовательностей в таких измененных локусах, но 31261 считываемых последовательностей аллели дикого типа. В последней колонке показана суммарная фракционная концентрация для всех ОНВ для пациентов с ГКК1. Так как связанные с ГКК ОНВ специфичны для пациентов с ГКК, присутствующие связанные с ГКК ОНВ представляют собой ложноположительные результаты. Если пороговые значения, описанные в данной заявке, применить к данным выявленным вариантам последовательности, все такие ложноположительные результаты будут отфильтрованы.

Присутствие небольшого количества таких предполагаемых связанных с опухолью мутаций в плазме 16 здоровых контролей представляет "стохастический шум" данного способа, и вероятно связано с

ошибками секвенирования. Средняя фракционная концентрация, оцененная для такого шума, составляла 0,38%. Данные значения представляют диапазон для здоровых субъектов. Таким образом, пороговое значение для классификации нулевого уровня рака для ГКК может составлять приблизительно 0,5%, поскольку наибольшая фракционная концентрация составляла 0,43%. Таким образом, если из пациента с ГКК удалить все раковые клетки, будут ожидать присутствие данных низких фракционных концентраций.

Снова возвращаясь к таблице 800, если в качестве порога для нулевого уровня рака использовать значение 0,5%, то результаты для плазмы пациентов с ГКК1 и ГКК3 после лечения определяют как нулевой уровень на основании анализа ОНВ. ГКК2 можно классифицировать на один уровень выше, чем нулевой. ГКК4 также можно классифицировать на один уровень выше, чем нулевой, или еще более высокий уровень, но все же относительно низкий уровень по сравнению с образцами до лечения.

В одном варианте реализации, в котором указанный параметр соответствует количеству измененных локусов, порог может быть нулевым (т.е. один измененный локус может свидетельствовать о ненулевом уровне рака). Тем не менее, при многих параметрах (например, глубины), порог будет выше, например, будет иметь абсолютное значение 5 или 10. В одном варианте реализации, в котором за индивидуум наблюдали после лечения, порог может представлять собой некоторый процент ОНВ (детектированных путем непосредственного анализа опухолей), выявленных в образце. Если пороговое значение для необходимого в некотором локусе вариантных считываемых последовательностей достаточно велико, наличие лишь одного измененного локуса может указывать на ненулевой уровень рака.

Таким образом, количественный анализ вариаций (например, однонуклеотидных вариаций) в ДНК из биологического образца (например, из плазмы) можно применять для диагностики, мониторинга и прогнозирования рака. Для детектирования рака количество однонуклеотидных вариаций, детектированных в плазме исследованного субъекта, можно сравнить с таковым для группы здоровых субъектов. Детектирование в плазме здоровых субъектов однонуклеотидных вариаций может происходить вследствие ошибок секвенирования, неклональных мутаций в кровяных клетках и других органах. Было показано, что клетки у нормальных здоровых субъектов могут нести небольшое количество мутаций (Congrad D.F. и др. *Nat. Genet.* 2011;43:712-4), как показано в таблице 900. Таким образом, суммарное количество детектированных однонуклеотидных вариаций в плазме группы очевидно здоровых субъектов можно применять в качестве референсного диапазона для того, чтобы определить, присутствует ли в плазме исследованного пациента аномально высокое количество однонуклеотидных вариаций, соответствующее ненулевому уровню рака.

Здоровых субъектов, используемых для определения референсного диапазона, можно подобрать по возрасту и полу для соответствия исследуемым субъектам. В более раннем исследовании было показано, что количество мутаций в соматических клетках увеличивается с возрастом (Cheung N.K. и др., *JAMA* 2012;307:1062-71). Таким образом, накопление клонов клеток по мере старения является нормой, даже если они относительно доброкачественные большую часть времени или если им потребуется очень большое время, чтобы стать клинически значимыми. В одном варианте реализации референсные уровни можно получить для различных групп субъектов, например, с различным возрастом, полом, этнической принадлежностью и другими параметрами (например, курение, заболевание гепатитом, алкоголизм, лекарственный анамнез).

Референсный диапазон может изменяться в зависимости от используемого порогового значения (т.е. присутствие в локусе необходимого количества измененных маркеров последовательности), а также предположительного уровня ложноположительных вариаций и других переменных (например, возраста). Таким образом, референсный диапазон можно определить для конкретного набора одного или более критериев, и для определения параметра для образца будут применять такие же критерии. Тогда указанный параметр можно сравнить с референсным диапазоном, поскольку оба определяли, применяя одинаковые критерии.

Выше упоминалось, что в вариантах реализации можно использовать несколько пороговых значений для определения уровня рака. Например, с помощью первого уровня можно определить отсутствие признаков рака для параметров ниже порога и по меньшей мере первый уровень рака, который может представлять собой преднеопластический уровень. Другие уровни могут соответствовать различным стадиям рака.

С. Зависимость экспериментальных переменных.

Глубина секвенирования может быть важна для определения минимального порога детектирования минорного (например, опухолевого) генома. Например, если применять глубину секвенирования, равную 10 гаплоидным геномам, то минимальная концентрация опухолевой ДНК, которую можно обнаружить, даже с помощью технологии секвенирования без каких-либо ошибок, составляет 1/5, т.е. 20%. С другой стороны, если применять глубину секвенирования, равную 100 гаплоидным геномам, то минимальная концентрация опустится до 2%. Данный анализ относится к случаю, когда анализируют только один мутированный локус. Тем не менее, если анализируют большее количество мутированных локусов, то минимальная концентрация опухолевой ДНК может быть ниже, и ее определяют с помощью функции биномиальной вероятности. Например, если глубина секвенирования 10-кратная и фракционная концентрация опухолевой ДНК составляет 20%, то вероятность детектирования мутации составляет 10%. Тем

не менее, если присутствует 10 мутаций, то вероятность детектирования по меньшей мере одной мутации будет составлять

$$1 - (1 - 10\%)^{10} = 65\%.$$

Существует несколько следствий увеличения глубины секвенирования. Чем больше глубина секвенирования, тем больше ошибок секвенирования будет наблюдаться, см. фиг. 4 и 5. Тем не менее, при большей глубине секвенирования будет легче отличить ошибки секвенирования от мутаций вследствие клональной экспансии субпопуляции клеток (например, раковых клеток), так как ошибки секвенирования будут встречаться в геноме случайно, а мутации будут встречаться в одном и том же положении для данной популяции клеток.

Чем больше глубина секвенирования, тем больше мутаций будет детектировано в "здоровых клетках". Тем не менее, если не происходит клональная экспансия данных здоровых клеток и их мутационные профили различны, то мутации в данных здоровых клетках можно отличить от мутаций по частотам встречаемости в плазме (например, применяя пороговое значение N для необходимого количества считываемых последовательностей, в которых детектирована мутация, например N должно быть равно 2, 3, 4, 5 или больше).

Выше упоминалось, что порог может зависеть от количества мутаций в здоровых клетках, которые будут клонально размножаться, и, следовательно, их могут не отличить от других механизмов. Такая изменчивость, которую можно ожидать, может быть получена при анализе здоровых субъектов. Так как клональная экспансия происходит с течением времени, на изменчивость, которую наблюдают у здоровых субъектов, может влиять возраст пациента, и, таким образом, порог может зависеть от возраста.

D. Комбинации целевых подходов.

В некоторых вариантах реализации случайное секвенирование можно применять в комбинации с целевыми подходами. Например, можно осуществить случайное секвенирование образца плазмы при поступлении пациента с раковым заболеванием. Можно проанализировать отклонения числа копий и ОНВ в результатах секвенирования ДНК в плазме. Области, в которых выявлены отклонения (например, амплификация/делеция или высокая плотность ОНВ), можно целенаправленно взять для серийного мониторинга. Мониторинг можно осуществить через какой-либо период времени, или осуществить незамедлительно после случайного секвенирования, фактически как единую процедуру. Для целевого анализа можно успешно применять подходы захвата на основе гибридизации в жидкой фазе для обогащения плазмы ДНК для неинвазивной пренатальной диагностики (Liao G.J. и др. *Clin. Chem.* 2011;57:92-101). Такие методики упоминаются выше. Таким образом, направленные и случайные подходы можно применять в комбинации для детектирования и мониторинга рака.

Таким образом, можно осуществить целевое секвенирование локусов, в которых с применением нецелевого полногеномного подхода, упомянутого выше, обнаружили потенциальные мутации. Такое целевое секвенирование можно осуществить, применяя методики гибридизации в растворе или твердофазной гибридизации (например, применяя Agilent SureSelect, NimbleGen Sequence Capture, или систему целевого ресеквенирования Illumina), а затем массовое параллельное секвенирование. Другой подход представляет собой осуществление амплификации (например, системы на основе ПЦР) для целевого секвенирования (Forshew T. и др. *Sci. Transl. Med.* 2012; 4: 135ra68).

IX. Фракционная концентрация.

Фракционную концентрацию опухолевой ДНК можно применять для определения порогового значения необходимого количества вариаций в локусе до того, как в локусе обнаружили мутацию. Например, если известно, что фракционная концентрация относительно высока, то можно применять высокое пороговое значение для отсеивания большего количества ложноположительных результатов, поскольку известно, что для истинного ОНВ должно присутствовать относительно высокое количество вариантных считываемых последовательностей. С другой стороны, если фракционная концентрация оказалась низкой, то может потребоваться нижнее пороговое значение, чтобы не упустить некоторые из ОНВ. В данном случае фракционную концентрацию будут определять с помощью способа, отличного от анализа ОНВ, где она используется в качестве параметра.

Для определения фракционной концентрации можно применять различные методики, некоторые из которых описаны в данной заявке. Данные методики можно применять для определения фракционной концентрации ДНК опухолевого происхождения в смеси, например в биоптате, содержащем смесь опухолевых клеток и незлокачественных клеток, или в образце плазмы из пациента с раковым заболеванием, содержащем ДНК, высвобожденную из опухолевых клеток, и ДНК, высвобожденную из незлокачественных клеток.

A. GAAL.

Полногеномный анализ суммарной потери аллелей (GAAL) позволяет анализировать локусы, которые утратили гетерозиготность (Chan K.C. и др. *Clin. Chem.* 2013; 59:211-24). Для гетерозиготного сайта конститутивного генома (КГ) в опухоли часто обнаруживают локус, в котором произошла делеция одной из аллелей. Таким образом, считываемые последовательности для такого локуса выявят большее количество одной аллели, чем другой, при этом выявленное различие пропорционально фракционной концентрации опухолевой ДНК в образце. Пример такого расчета приведен далее.

ДНК, выделенную из лейкоцитарной пленки и опухолевых тканей пациентов с ГКК, генотипировали с помощью системы Affymetrix Genome-Wide Human SNP Array 6.0. Результаты, полученные для микрочипа, обрабатывали с помощью Affymetrix Genotyping Console версии 4.1. Анализ генотипирования и определение однонуклеотидных полиморфизмов (ОНП) осуществляли с помощью алгоритма Birdseed версии 2. Результаты генотипирования для лейкоцитарных пленок и опухолевых тканей использовали для определения областей с утратой гетерозиготности (LOH) и для осуществления анализа числа копий. Анализ числа копий осуществляли с использованием параметров по умолчанию Genotyping Console от Affymetrix, с минимальным размером геномных фрагментов 100 п.о. и минимальным количеством генетических маркеров внутри фрагмента, равным 5.

Области с LOH обнаруживали как области с 1 копией в опухолевой ткани и 2 копиями в лейкоцитарной пленке, при этом ОНП в данных областях были гетерозиготными в лейкоцитарной пленке, но гомозиготными в опухолевой ткани. Для области генома, в которой детектирована LOH в опухолевой ткани, аллели ОНП, которые присутствовали в лейкоцитарной пленке, но отсутствовали или проявляли меньшую интенсивность в опухолевых тканях, считали аллелями на удаленном фрагменте фрагмента хромосомы. Аллели, которые присутствовали как в лейкоцитарной пленке, так и в опухолевой ткани, считали полученными из не удаленного фрагмента хромосомы. Для всех фрагментов хромосом с утратой одной копии в опухоли считали суммарное количество считываемых последовательностей с утратой аллелей и без утраты аллелей. Разность данных двух значений использовали для вычисления фракционной концентрации ДНК опухолевого происхождения (F_{GAAL}) в образце, применяя следующее уравнение:

$$F_{GAAL} = \frac{N_{\text{без дел}} - N_{\text{дел}}}{N_{\text{без дел}}}$$

где $N_{\text{без дел}}$ представляет собой суммарное количество считываемых последовательностей без утраты аллелей и $N_{\text{дел}}$ представляет собой суммарное количество считываемых последовательностей с утратой аллелей.

В. Оценка с применением геномного представления.

Проблема методики GAAL состоит в том, что определяют конкретные локусы (т.е. такие локусы, в которых детектированы области с утратой гетерозиготности (LOH)) и используют только считываемые последовательности, выровненных с такими локусами. Такое требование может привести к необходимости дополнительных этапов и, таким образом, к повышению стоимости. Здесь описан вариант реализации, в котором используют только число копий, например плотность считываемых последовательностей.

Хромосомные аберрации, например амплификации и делеции, часто наблюдают в геномах раковых клеток. Хромосомные аберрации, наблюдаемые в раковых тканях, обычно происходят на субхромосомных областях, и такие аберрации могут быть короче, чем 1 мегабаза. Кроме того, связанные с раком хромосомные аберрации гетерогенны у различных пациентов, и, следовательно, у различных пациентов могут быть поражены различные фрагменты. Также в геноме раковой клетки нередко обнаруживают десятки, сотни или даже тысячи вариаций числа копий. Каждый из перечисленных факторов затрудняет определение концентрации опухолевой ДНК.

Варианты реализации включают анализ количественных изменений, происходящих в результате связанных с опухолью хромосомных аберраций. В одном варианте реализации образцы ДНК, включающие ДНК, происходящую из раковых клеток и нормальных клеток, секвенируют, применяя массовое параллельное секвенирование, например, с помощью платформы секвенирования Illumina HiSeq2000. Полученная ДНК может представлять собой внеклеточную ДНК из плазмы или другого подходящего биологического образца.

Области хромосом, которые амплифицировались в опухолевых тканях, с большей вероятностью будут секвенированы, а области, которые удалились в опухолевых тканях, с меньшей вероятностью будут секвенированы. В результате, плотность считываемых последовательностей, выравниваемых с амплифицированными областями, будет повышена и плотность считываемых последовательностей, выравниваемых с удаленными областями, будет понижена. Степень вариации пропорциональна фракционной концентрации ДНК опухолевого происхождения в смеси ДНК. Чем выше доля ДНК из опухолевой ткани, тем большее изменение будет вызвано хромосомными аберрациями.

1) Оценка образца с высокой концентрацией опухоли.

ДНК выделяли из опухолевых тканей четырех пациентов с печеночно-клеточной карциномой. ДНК фрагментировали, применяя систему разрушения ДНК ультразвуком Covaria, и секвенировали, применяя описанную платформу Illumina HiSeq2000 (Chan K.C. и др. Clin. Chem. 2013; 59:211-24). Считываемые последовательности выравнивали с референсным геномом человека (hg18). Геном затем разделяли на фрагменты (области) по 1 мегабазе и для каждого фрагмента рассчитывали плотность считываемых последовательностей после поправки на GC-сдвиг, как было описано (Chen E.Z. и др. PLoS One. 2011;6:e21791).

После выравнивания считываемых последовательностей с референсным геномом можно вычислить плотность считываемых для различных областей. В одном варианте реализации плотность считываемой последовательности представляет собой соотношение, определяемое как количество считываемых по-

следовательностей, картированных на конкретном фрагменте (например, области размером 1 мегабаза), деленное на суммарное количество считываемых последовательностей, которые можно выровнять с референсным геномом (например, с уникальным положением в референсном геноме). Ожидают, что для фрагментов, которые перекрываются с областями хромосом, амплифицированными в опухолевой ткани, плотности считываемых последовательностей будут выше, чем для фрагментов без таких перекрываний. С другой стороны, ожидают, что для фрагментов, которые перекрываются с областями хромосом, содержащими делецию, плотности считываемых последовательностей будут ниже, чем для фрагментов без таких перекрываний. На величину различия плотностей считываемых последовательностей между областями с хромосомными aberrациями и без них преимущественно влияет доля ДНК опухолевого происхождения в образце и степень амплификации/делеции в опухолевых клетках.

Можно применять различные статистические модели, чтобы определить фрагменты с плотностями считываемых последовательностей, соответствующими различным типам хромосомных aberrаций. В одном варианте реализации можно применять модель смеси нормальных распределений (McLachlan G. и Peel D. *Multivariate normal mixtures*. In *Finite mixture models* 2004: стр. 81-116. John Wiley & Sons Press). Также можно применять другие статистические модели, например модель смеси биномиальных распределений и модель пуассоновской регрессии (McLachlan G. и Peel D. *Mixtures with non-normal components*, *Finite mixture models* 2004: стр. 135-174. John Wiley & Sons Press).

Плотность считываемой последовательности фрагмента можно нормировать, применяя плотность считываемой последовательности того же фрагмента, определенную с помощью секвенирования ДНК из лейкоцитарной пленки. На плотности считываемых последовательностей для различных фрагментов может влиять контекст последовательности конкретной хромосомной области, и, таким образом, нормировка позволяет более точно определить области с aberrациями. Например, картируемость (которая относится к вероятности выравнивания последовательности с ее исходным положением) различных областей хромосом может различаться. Кроме того, полиморфизм числа копий (т.е. вариации числа копий) также будут влиять на плотности считываемых последовательностей фрагментов. Следовательно, нормировка на ДНК из лейкоцитарной пленки потенциально может минимизировать количество вариаций, связанных с различием в контексте последовательности между различными областями хромосом.

На фиг. 10А показан график 1000 распределения плотностей считываемых последовательностей образца опухоли из пациента с ГКК согласно вариантам реализации настоящего изобретения. Опухолевую ткань получали после хирургической резекции из пациента с ГКК. Ось x представляет собой \log_2 отношения (R) плотности считываемой последовательности из опухолевой ткани к таковой из лейкоцитарной пленки пациента. Ось y представляет собой количество фрагментов.

Пики можно подогнать под кривую распределения, чтобы показать области с делецией, амплификацией и без хромосомных aberrаций, применяя модель смеси нормальных распределений. В одном варианте реализации количество пиков можно определить с помощью информационного критерия Акаике (AIC) для всех различных правдоподобных значений. Центральный пик с $\log_2 R = 0$ (т.е. $R = 1$) представляет собой области без каких-либо хромосомных aberrаций. Левый пик (относительно центрального) представляет собой области с утратой одной копии. Правый пик (относительно центрального) представляет собой области с приобретением одной копии.

Фракционную концентрацию ДНК опухолевого происхождения может отражать расстояние между пиками, представляющими амплифицированные и удаленные области. Чем больше указанное расстояние, тем выше будет фракционная концентрация ДНК опухолевого происхождения в образце. Фракционную концентрацию ДНК опухолевого происхождения в образце можно определить с помощью подхода геномного представления, обозначенного F_{GR} , применяя следующее уравнение

$$F_{GR} = R_{\text{правый}} - R_{\text{левый}}$$

где $R_{\text{правый}}$ представляет собой значение R для правого пика и $R_{\text{левый}}$ представляет собой значение R для левого пика. Наибольшее различие будет представлять 1, что соответствует 100%. Оценили, что фракционная концентрация ДНК опухолевого происхождения в образце опухоли, полученном из пациента с ГКК, составляла 66%, где значения $R_{\text{правый}}$ и $R_{\text{левый}}$ составляли 1,376 и 0,712 соответственно.

Чтобы подтвердить полученный результат, также применяли способ полногеномного анализа суммарной потери аллелей (GAAL) для независимого определения фракционной концентрации соотношения опухолевой ДНК (Chan K.C. и др. *Clin. Chem.* 2013; 59:211-24). В табл. 3 показаны фракционные концентрации ДНК опухолевого происхождения в опухолевых тканях четырех пациентов с ГКК, определенные с применением подходов геномного представления (F_{GR}) и GAAL (F_{GAAL}). Значения, определенные с помощью данных двух различных подходов, хорошо согласуются друг с другом.

Таблица 3. Фракционная концентрация, определенная с помощью GAAL и геномного представления (GR)

Опухоль ГКК	F _{GAAL}	F _{GR}
1	60,0%	66,5%
2	60,0%	61,4%
3	58,0%	58,9%
4	45,7%	42,2%

2) Оценка образца с низкой концентрацией опухоли.

Описанный выше анализ показал, что наш способ геномного представления можно применять для измерения фракционной концентрации опухолевой ДНК, когда более чем 50% образца ДНК имеет опухолевое происхождение, т.е. когда опухолевая ДНК представляет собой доминирующую часть. В предыдущем анализе мы показали, что данный способ также можно применять по отношению к образцам, в которых ДНК опухолевого происхождения представляет собой минорную часть (т.е. ниже 50%). Образцы, которые могут содержать минорную часть опухолевой ДНК, включают, но не ограничены перечисленными: кровь, плазму, сыворотку, мочу, плевральную жидкость, спинномозговую жидкость, слезы, слюну, асцитную жидкость и кал пациентов с раковыми заболеваниями. В некоторых образцах фракционная концентрация ДНК опухолевого происхождения может составлять 49, 40, 30, 20, 10, 5, 2, 1, 0,5, 0,1% или менее.

Для таких образцов пики плотности варианта считываемой последовательности, представляющие области с амплификацией и делецией, могут быть не такими явными, как для образцов, содержащих относительно высокую концентрацию ДНК опухолевого происхождения, описанных выше. В одном варианте реализации области с хромосомными aberrациями в раковых клетках можно обнаружить путем сравнения с референсными образцами, в которых, как известно, не содержится раковая ДНК. Например, в качестве эталона можно применять плазму субъектов, не страдающих раком, чтобы определить нормативный диапазон плотностей варианта считываемой последовательностей для областей хромосом. Плотность считываемой последовательности исследуемого субъекта можно сравнить с ее значением для референсной группы. В одном варианте реализации можно определить среднее значение и стандартное отклонение (СО) плотности считываемой последовательности. Для каждого фрагмента плотность считываемой последовательности исследуемого субъекта сравнивают со средним значением в референсной группе, чтобы определить z-показатель, применяя следующую формулу:

$$z\text{-показатель} = \frac{(GR_{\text{тест}} - GR_{\text{этал}})}{CO_{\text{этал}}}$$

где GR_{тест} представляет собой плотность считываемой последовательности пациента с раковым заболеванием;

GR_{этал} представляет собой среднее значение плотности считываемой последовательности референсных субъектов и СО_{этал} представляет СО плотностей считываемых последовательностей референсных субъектов.

Области с z-показателем < -3 указывают на существенную недостаточность представления плотности считываемой последовательности для конкретного фрагмента у пациента с раковым заболеванием, позволяя предположить наличие делеции в опухолевой ткани. Области с z-показателем > 3 указывают на существенную избыточность представления плотности считываемой последовательности для конкретного фрагмента у пациента с раковым заболеванием, позволяя предположить наличие амплификации в опухолевой ткани.

Затем можно построить распределение z-показателей для всех фрагментов, чтобы определить области с различными количествами приобретений и утрат копий, например, с делецией 1 или 2 копий хромосомы; и амплификацией с приобретением 1, 2, 3 и 4 дополнительных копий хромосомы. В некоторых случаях может быть вовлечена более чем одна хромосома или более чем одна область хромосомы.

На фиг. 10В показан график 1050 распределения z-показателей для всех фрагментов в плазме пациента с ГКК согласно вариантам реализации настоящего изобретения. Пики (слева направо), представляющие утрату 1 копии, отсутствие изменения копий, приобретение 1 копии и приобретение 2 копий, подогнали к распределению z-показателей. Области с различными типами хромосомных aberrаций затем можно выявить, например, применяя модель смеси нормальных распределений, описанную выше.

Фракционную концентрацию раковой ДНК в образце (F) затем можно вывести из плотностей считываемых последовательностей фрагментов, в которых детектировано приобретение одной копии или утрата одной копии. Фракционную концентрацию, определенную для конкретного фрагмента, можно рассчитать с помощью уравнения

$$F = \frac{(|GR_{\text{тест}} - GR_{\text{этал}}|) \times 2}{GR_{\text{этал}}} \times 100\%$$

Это уравнение также можно представить в виде

$$F = (|z\text{-показатель} \times CO_{\text{этал}}| / (GR_{\text{этал}})) \times 2$$

а также можно переписать как

$$F = |z - \text{показатель}| \times KB \times 2,$$

где KB представляет собой коэффициент вариации измерения плотности считываемой последовательности референсных субъектов.

В одном варианте реализации результаты, полученные для фрагментов, объединяют. Например, можно усреднить z-показатели фрагментов, в которых выявили приобретение 1 копии, или можно усреднить полученные для них значения F. В другом варианте реализации значение z-показателя, используемое для вывода F, определяют с помощью статистической модели и представляют в виде пиков, показанных на фиг. 10В и 11. Например, z-показатель для правого пика можно применять для определения фракционной концентрации областей, в которых детектировано приобретение 1 копии.

В другом варианте реализации все фрагменты с z-показателем < -3 и z-показателем > 3 можно отнести к областям с утратой одной копии и приобретением одной копии соответственно, так как данные два типа хромосомных аберраций наиболее распространены. Такое приближение наиболее полезно, когда количество фрагментов с хромосомными аберрациями относительно мало и подгонка к нормальному распределению не может быть точной.

На фиг. 11 показан график 1100 распределения z-показателей для плазмы из пациента с ГКК согласно вариантам реализации настоящего изобретения. Хотя количество фрагментов, перекрывающихся с хромосомными аберрациями, относительно мало, все фрагменты с z-показателем < -3 и z-показателем > 3 подогнали к нормальным распределениям утраты одной копии и приобретения одной копии соответственно.

Фракционные концентрации ДНК опухолевого происхождения в плазме четырех пациентов с ГКК определяли, применяя анализ GAAL и подход на основе GR. Результаты представлены в табл. 4. Видно, что полученное фракционное представление хорошо коррелирует между анализом GAAL и анализом GR.

Таблица 4. Фракционная концентрация в плазме ДНК опухолевого происхождения, полученная с помощью анализа хромосомных аберраций

Образцы	Фракционная концентрация в плазме ДНК опухолевого происхождения	
	Анализ GAAL	Анализ GR
Случай 11	4,3%	4,5%
Случай 13	5%	5,5%
Случай 23	52%	62%
Случай 27	7,6%	6,1%

С. Способ определения фракционной концентрации.

На фиг. 12 представлена блок-схема способа 1200 определения фракционной концентрации опухолевой ДНК в биологическом образце, содержащем внеклеточную ДНК согласно вариантам реализации настоящего изобретения. Способ 1200 можно осуществить с помощью различных вариантов реализации, включая варианты реализации, описанные выше.

В блоке 1210 получают один или более маркеров последовательности для каждого из множества фрагментов ДНК в биологическом образце. Блок 1210 можно осуществить, как описано в данной заявке для других способов. Например, можно секвенировать из образца плазмы один конец фрагмента ДНК. В другом варианте реализации можно секвенировать оба конца фрагмента ДНК, что позволяет оценить длину фрагмента.

В блоке 1220 определяют положения в геноме маркеров последовательности. Положения в геноме можно определить, например, как описано в данной заявке, путем выравнивания маркеров последовательности с референсным геномом. Если секвенируют оба конца фрагмента, то спаренные маркеры можно выровнять в виде пары с расстоянием между двумя маркерами, которое должно быть меньше, чем определенное расстояние, например 500 или 1000 оснований.

В блоке 1230 для каждой из множества областей генома определяют соответствующее количество фрагментов ДНК, принадлежащих данной области генома, с помощью маркеров последовательности, расположенных в геноме внутри данной области генома. Области генома могут представлять собой неперекрывающиеся фрагменты равной длины в референсном геноме. В одном варианте реализации можно сосчитать количество маркеров, которые выровняли с фрагментом. Таким образом, с каждым фрагментом может быть выровнено соответствующее количество маркеров. Можно построить гистограмму, иллюстрирующую частоту выравнивания с фрагментами определенного количества маркеров. Способ 1200 можно осуществить для областей генома, каждая из которых имеет одинаковую длину (например, для фрагментов по 1 мегабазе), при этом области не перекрываются. В других вариантах реализации можно применять различные длины, которые могут быть учтены, и области могут перекрываться.

В блоке 1240 соответствующее количество нормируют с получением соответствующей плотности. В одном варианте реализации нормировка соответствующего количества с получением соответствующей

плотности включает применение такого же суммарного количества выровненных референсных маркеров для определения соответствующей плотности и референсной плотности. В другом варианте реализации соответствующее количество можно поделить на суммарное количество выровненных референсных маркеров.

В блоке 1250 соответствующую плотность сравнивают с референсной плотностью, чтобы выявить в данной области генома утрату 1 копии или приобретение 1 копии. В одном варианте реализации вычисляют различие между соответствующей плотностью и референсной плотностью (например, в рамках определения z-показателя) и сравнивают с пороговым значением. В различных вариантах реализации референсную плотность можно получить для образца из здоровых клеток (например, для лейкоцитарной пленки) или для соответствующих количеств таких же клеток (например, получив срединное или среднее значение, при допущении, что в большинстве областей нет утраты или приобретения копий).

В блоке 1260 первую плотность рассчитывают по одной или более соответствующим плотностям, в которых детектирована утрата 1 копии, или по одной или более соответствующим плотностям, в которых детектировано приобретение 1 копии. Первая плотность может соответствовать только одной области генома, или ее можно определить по плотностям нескольких областей генома. Например, первую плотность можно вычислить по соответствующим плотностям с утратой 1 копии. Соответствующие плотности дают возможность измерить величину различия плотностей, возникшего в результате делеции указанной области в опухоли, позволяя получить концентрацию опухоли. Аналогично, если первая плотность относится к соответствующим плотностям с приобретением 1 копии, то можно получить величину различия плотностей, возникшего в результате дупликации данной области в опухоли. В разделах выше описаны различные примеры того, как можно применять плотности нескольких областей для определения средней плотности, чтобы использовать ее для определения первой плотности.

В блоке 1270 фракционную концентрацию рассчитывают путем сравнения первой плотности с другой плотностью с получением разности. Полученную разность нормируют на референсную плотность, что можно осуществить в блоке 1270. Например, разность можно нормировать на референсную плотность путем деления разности на референсную плотность. В другом варианте реализации разность может быть нормирована в описанных ранее блоках.

В одном варианте реализации другая плотность представляет собой референсную плотность, например, описанную выше в разделе 2. Таким образом, вычисление фракционной концентрации может включать умножение разности на два. В другом варианте реализации другая плотность представляет собой вторую плотность, рассчитанную по соответствующим плотностям с выявленной утратой 1 копии (где первую плотность рассчитывают, применяя соответствующие плотности, в которых детектировано приобретение 1 копии), например, описанным выше в разделе 1. В данном случае нормированную разность можно определить путем вычисления первого отношения (например, $R_{\text{правый}}$) первой плотности к референсной плотности и вычисления второго отношения ($R_{\text{левый}}$) второй плотности к референсной плотности, при этом разность находится между первым отношением и вторым отношением. Выше описано, что можно определить область генома, в которой детектирована утрата 1 копии или приобретение 1 копии, путем подгонки пиков к кривой распределения гистограммы соответствующих плотностей.

Вкратце, варианты реализации позволяют анализировать геномное представление ДНК в плазме в различных областях хромосом, чтобы одновременно определить, амплифицирована ли или удалена ли хромосомная область в опухолевой ткани, и, если указанная область амплифицирована или удалена, использовать данное геномное представление для вычисления фракционной концентрации ДНК опухолевого происхождения. В некоторых вариантах реализации применяют модель смеси нормальных распределений, чтобы проанализировать общее распределение геномного представления различных фрагментов для определения геномного представления, связанного с различными типами отклонений, а именно с приобретениями 1, 2, 3 или 4 копий и потерями 1 или 2 копий.

Описанные варианты реализации обладают несколькими преимуществами над другими способами, например над подходом полногеномного анализа суммарной потери аллелей (GAAL) (заявка на патент США 13/308473; Chan K.C. и др. Clin. Chem. 2013; 59:211-24) и анализом связанных с опухолью однонуклеотидных мутаций (Forshew T. и др. Sci. Transl. Med. 2012; 4:136ra68). Все считываемые последовательности, картированные на областях с хромосомными aberrациями, можно применять для определения плотности считываемой последовательности указанной области, и, следовательно, они информативны в отношении фракционной концентрации опухолевой ДНК. С другой стороны, в анализе GAAL информативными будут только считываемые последовательности, покрывающие отдельные нуклеотиды, которые гетерозиготны у индивида и расположены внутри хромосомной области с приобретением или потерей хромосомы. Аналогично, в анализе связанных с раком мутаций только считываемые последовательности, покрывающие мутации, будут полезны для вычисления концентрации опухолевой ДНК. Следовательно, варианты реализации могут обеспечить более экономически эффективное использование результатов секвенирования, так как может потребоваться относительно меньшее количество считываемых последовательностей при секвенировании по сравнению с другими подходами для достижения такой же степени достоверности оценки фракционной концентрации ДНК опухолевого происхождения.

Х. Альтернативные методики.

Помимо использования в качестве критерия для определения того, что локус содержит истинную мутацию, количества раз, которое отдельная мутация наблюдается в маркере последовательности (с помощью чего корректируют положительное прогнозируемое значение), можно применять другие методики вместо или вдобавок к применению порогового значения, чтобы получить большее прогнозируемое значение для идентификации раковой мутации. Например, при обработке результатов секвенирования можно применять биоинформационные фильтры различной строгости, например, учитывая показатель качества секвенированного нуклеотида. В одном варианте реализации можно применять секвенаторы ДНК и химические компоненты для секвенирования с различными профилями ошибок секвенирования. Секвенаторы и химические компоненты с меньшей частотой ошибок секвенирования позволяют получить более высокие положительные предсказательные значения. Можно также осуществлять повторное секвенирование одного и того же фрагмента ДНК, чтобы увеличить достоверность секвенирования. Одна возможная стратегия представляет собой стратегию секвенирования кольцевых консенсусных последовательностей от Pacific Biosciences.

В другом варианте реализации для интерпретации полученных результатов можно учитывать информацию о размере секвенированных фрагментов. Так как в плазме ДНК опухолевого происхождения короче, чем ДНК неопухолевого происхождения (см. заявку на патент США № 13/308473), то положительное прогнозируемое значение для более короткого фрагмента ДНК в плазме, содержащего потенциальную мутацию опухолевого происхождения, будет выше, чем таковое для более длинного фрагмента ДНК в плазме. Данные о размерах можно легко получить, осуществляя секвенирование спаренных концов ДНК в плазме. В качестве альтернативы можно применять секвенаторы ДНК, которые считывают длинные молекулы, таким образом, позволяя получить полную длину фрагмента ДНК в плазме. Также можно осуществить фракционирование ДНК в образце плазмы по размерам перед секвенированием ДНК. Примеры способов, которые можно применять для фракционирования по размерам, включают электрофорез в геле, применение микрофлюидного подхода (например, системы LabChip XT Caliper) и эксклюзионных центрифужных колонок.

В еще одном варианте реализации будут ожидать, что фракционная концентрация связанных с опухолью мутаций в плазме у пациента с негематологическим раком будет выше, если сосредоточиться на более коротких фрагментах ДНК в плазме. В одном варианте реализации можно сравнить фракционную концентрацию связанных с опухолью мутаций в фрагментах ДНК плазмы из двух или более различных распределений по размеру. У пациента с негематологическим раком будут наблюдаться более высокие фракционные концентрации связанных с опухолью мутаций в более коротких фрагментах по сравнению с более длинными фрагментами.

В некоторых вариантах реализации можно объединить результаты секвенирования для двух или более аликвот одного и того же образца крови или для двух или более образцов крови, полученных в то же время или в разные моменты времени. Положительное прогнозируемое значение потенциальных связанных с опухолью мутаций, наблюдаемых в более чем одной аликвоте или в более чем одном образце, будет более высоким. Положительное прогнозируемое значение будет повышаться с количеством образцов, в которых выявлена такая мутация. Потенциальные мутации, которые присутствуют в образцах плазмы, взятых в различные моменты времени, можно считать потенциальными мутациями.

XI. Примеры.

Далее представлены примеры методик и результатов, но не следует считать, что они ограничивают варианты реализации настоящего изобретения.

A. Материалы и методы.

Для сбора образцов привлекали пациентов с печеночно-клеточной карциномой (ГКК), носителей хронического гепатита В и пациента с одновременным раком груди и яичников. У всех пациентов с ГКК из клиники Барселоны был рак печени на стадии заболевания A1. Образцы периферической крови собирали у всех участников в пробирки, содержащие ЭДТА. Опухолевые ткани пациентов с ГКК получали в процессе хирургической резекции рака.

Образцы периферической крови центрифугировали при 1600 g в течение 10 мин при 4°C. Полученную плазму снова центрифугировали при 16000 g в течение 10 мин при 4°C, а затем хранили при 80°C. Выделяли молекулы внеклеточной ДНК из 4,8 мл плазмы согласно протоколу для выделения ДНК из крови и биологической жидкости из набора QIAamp DSP DNA Blood Mini (Qiagen). ДНК из плазмы для каждого из случаев концентрировали с помощью концентратора SpeedVac (Savant DNA120; Thermo Scientific) до конечного объема 40 мкл для последующего получения библиотеки секвенирования ДНК.

Геномную ДНК выделяли из образцов лейкоцитарной пленки пациентов согласно протоколу для выделения ДНК из крови и биологической жидкости с помощью набора QIAamp DSP DNA Blood Mini. ДНК выделяли из опухолевых тканей с помощью набора QIAamp DNA Mini (Qiagen).

Конструировали библиотеки секвенирования образцов геномной ДНК с помощью набора Paired-End Sample Preparation Kit (Illumina), согласно инструкциям производителя. Вкратце, 1-5 мкг геномной ДНК сначала расщепляли на фрагменты размером 200 п.о. с помощью устройства для ультразвуковой фрагментации Covaris S220 Focused-ultrasonicator. Впоследствии застраивали концы молекул ДНК с помощью ДНК-полимеразы T4 и фрагмента Кленова ДНК-полимеразы; затем применяли полинуклеотид-

киназу T4 для фосфорилирования 5'-концов. Получали 3'-выступ с помощью фрагмента Кленова, лишённого 3'-5'-экзонуклеазной активности. Олигонуклеотидные адаптеры Illumina лигировали с липкими концами. Лигированную с адаптером ДНК амплифицировали с помощью 12 циклов ПЦР. Так как молекулы ДНК в плазме представляли собой короткие фрагменты и количества общей ДНК в образцах плазмы были относительно малы, мы пропустили этапы фрагментации и использовали 15 циклов ПЦР для конструирования библиотек ДНК из образцов плазмы.

Биоанализатор Agilent 2100 (Agilent Technologies) применяли для проверки качества и размера библиотек лигированных с адаптерами ДНК. Библиотеки ДНК затем анализировали с помощью набора для количественного анализа библиотек KAPA (Kapa Biosystems), следуя инструкциям производителя. Библиотеку ДНК разбавляли и гибридизовали с проточными кюветами для секвенирования спаренных концов. Кластеры ДНК получали с помощью системы для получения кластеров cBot (Illumina) с помощью 2 версии набора TruSeq PE Cluster Generation (Illumina), после чего проводили 51_2 цикл или 76_2 циклов секвенирования на системе HiSeq 2000 (Illumina) с помощью 2 версии набора TruSeq SBS (Illumina).

Результаты секвенирования спаренных концов анализировали посредством программы для выравнивания коротких олигонуклеотидов Short Oligonucleotide Alignment Program 2 (SOAP2) в режиме спаренных концов. Для каждой считываемой последовательности со спаренными концами по 50 п.о. или 75 п.о. с каждого конца выравнивали с референсным геномом человека без маскировки повторов (hg18). При выравнивании каждого конца допускали несовпадение 2 нуклеотидов. Геномные координаты данных потенциальных выравниваний для 2 концов затем анализировали, чтобы определить, удастся ли выровнять любую комбинацию 2 концов с одной и той же хромосомой в правильной ориентации с размером вставки, меньшим или равным 600 п.о., и с картированием в одном положении в референсном геноме человека. Дублированные считываемые последовательности определяли как считываемые последовательности спаренных концов, в которых начальное и конечное положения в геноме человека вставленной молекулы ДНК были идентичны; дублированные считываемые последовательности удаляли, как описано ранее (Lo и др. *Sci. Transl. Med.* 2010; 2: 61ra91, doi: 10.1126/scitranslmed.3001720).

В некоторых вариантах реализации секвенировали пары образцов опухолевой и конститутивной ДНК, чтобы определить связанные с опухолью однонуклеотидные варианты (ОНВ). В некоторых вариантах реализации мы сосредоточились на ОНВ, встречающихся в гомозиготных сайтах в конститутивной ДНК (в данном примере ДНК получали из лейкоцитарной пленки). В принципе, любой вариант нуклеотида, детектированный в результатах секвенирования опухолевых тканей, но отсутствующий в конститутивной ДНК, может представлять собой потенциальную мутацию (т.е. ОНВ). Вследствие ошибок секвенирования (0,1-0,3% секвенированных нуклеотидов), тем не менее, в геноме будут детектированы миллионы ложноположительных результатов, если однократное появление любой нуклеотидной замены в результатах секвенирования опухолевой ткани рассматривать как связанный с опухолью ОНВ. Одним из способов уменьшения количества ложноположительных результатов будет установление критерия неоднократной встречаемости одной и той же нуклеотидной замены в результатах секвенирования для опухолевой ткани, перед тем, как назвать потенциальную мутацию связанным с опухолью ОНВ.

Так как возникновение ошибок секвенирования представляет собой стохастический процесс, количество ложноположительных результатов вследствие ошибок секвенирования будет экспоненциально уменьшаться с увеличением необходимого количества детектированных ОНВ, чтобы посчитать потенциальную мутацию связанным с опухолью ОНВ. С другой стороны, количество ложноположительных результатов будет увеличиваться с увеличением глубины секвенирования. Такие взаимоотношения можно прогнозировать с помощью функций пуассоновского и биномиального распределения. В вариантах реализации можно определить динамический порог встречаемости, чтобы считать, что наблюдаемый ОНВ связан с опухолью. В вариантах реализации могут учитываться фактическое покрытие конкретного нуклеотида в результатах секвенирования для опухоли, частота ошибок секвенирования, максимально допустимый уровень ложноположительных вариаций и желательная чувствительность для детектирования мутации.

В некоторых примерах мы установили очень строгие критерии для уменьшения количества ложноположительных результатов. Например, можно потребовать полного отсутствия мутации в результатах секвенирования конститутивной ДНК и 20-кратной глубины секвенирования конкретного положения нуклеотида. В некоторых вариантах реализации установление порога встречаемости позволяло добиться частоты ложноположительных результатов, меньшей чем 10^{-7} . В некоторых примерах мы также отсеивали ОНВ, которые были детектированы в центромерных, теломерных областях и областях низкой сложности, чтобы минимизировать ложноположительные результаты вследствие артефактов выравнивания. Кроме того, также удаляли предполагаемые ОНВ, картированные как известные ОНП, представленные в базе данных dbSNP сборки 135.

В. До и после резекции.

На фиг. 13А показана таблица 1300 анализа мутаций в плазме пациента с раком яичника и раком молочной железы на момент диагностики согласно вариантам реализации настоящего изобретения. Здесь мы продемонстрировали пример пациента с двусторонним раком яичников и раком молочной железы. Результаты секвенирования ДНК из плазмы сравнивали с результатами секвенирования конститутивной

ДНК пациента (из лейкоцитарной пленки). Однонуклеотидные изменения, которые присутствовали в плазме, но не в конститутивной ДНК, расценивали как потенциальные мутации. Брали образцы рака яичников с правой и левой стороны пациента из двух локализаций с каждой стороны, т.е. получали всего четыре образца опухоли. Мутациями опухоли считали мутации, детектированные во всех четырех тканях опухоли яичника в четырех различных локализациях.

При секвенировании ДНК плазмы обнаружили более 3,6 миллиона однонуклеотидных изменений за по меньшей мере один раз. Среди данных изменений лишь 2064 также обнаружили в опухолевых тканях, и положительное прогнозируемое значение составило 0,06%. При применении критерия по меньшей мере двукратного детектирования в плазме количество потенциальных мутаций значительно уменьшилось на 99,5% до 18885. Количество опухолевых мутаций уменьшилось лишь на 3% до 2003 и положительное прогнозируемое значение повысилось до 11%.

При применении критерия, по меньшей мере, пятикратного детектирования в плазме обнаружили лишь 2572 потенциальные мутации и среди них 1814 представляли собой мутации, детектированные во всех опухолевых тканях, таким образом, положительное прогнозируемое значение составило 71%. Для детектирования потенциальных мутаций можно применять другие критерии количества детектирований (например, 2, 3, 4, 6, 7, 8, 9, 10 и т.д.) в зависимости от необходимой чувствительности и положительного прогнозируемого значения. Чем большее значение встречаемости используют в качестве критерия, тем больше будет положительное прогнозируемое значение и меньше будет чувствительность.

На фиг. 13В показана таблица анализа мутаций в плазме пациента с двусторонним раком яичников и раком молочной железы после резекции опухоли согласно вариантам реализации настоящего изобретения. Осуществляли хирургическую резекцию опухоли у пациента. Образец крови брали через день после резекции опухолей яичника и рака молочной железы. Затем секвенировали ДНК в плазме. В данном примере анализировали лишь мутации из рака яичников. В образце плазмы обнаружили более 3 миллионов потенциальных мутаций, по меньшей мере, однократно. Тем не менее, при применении критерия по меньшей мере пяти детектирований количество потенциальных мутаций уменьшилось до 238. Существенное уменьшение наблюдали при сравнении с количеством потенциальных мутаций в образце, взятом при диагностике, при применении того же критерия детектирования пяти мутаций.

В одном варианте реализации количество однонуклеотидных изменений, детектированных в плазме, можно применять в качестве параметра для детектирования рака, мониторинга пациента с раковым заболеванием и составления прогноза для него. Встречаемость различных их количеств можно применять в качестве критерия, чтобы добиться желательной чувствительности и специфичности. У пациента с большей опухолевой нагрузкой и, следовательно, худшим прогнозом, будут ожидать больший мутационный груз, наблюдаемый в плазме.

Для такого анализа можно установить профиль мутационного груза для различных типов рака. С целью мониторинга будет видно, что мутационный груз в плазме пациента, который отвечает на лечение, будет уменьшаться. Если опухоль возвратилась, например в процессе рецидива, то будут ожидать увеличение мутационного груза. Такой мониторинг позволит контролировать эффективность выбранного способа лечения пациента и обнаружить появление устойчивости к конкретному лечению.

С помощью анализа конкретных мутаций, которые можно наблюдать в результатах секвенирования ДНК из плазмы, можно также определить мишени, которые будут прогнозировать чувствительность (например, мутации в гене рецептора эпидермального фактора роста и ответ на лечение ингибитором тирозинкиназы) и устойчивость к конкретному целевому лечению (например, мутации KRAS при колоректальном раке и устойчивость к лечению панитумумабом и цетуксимабом), и могут направлять планирование режимов лечения.

Выше был приведен пример для двустороннего рака яичников. Также можно осуществить аналогичный анализ мутаций рака молочной железы, а затем отследить мутации обоих данных типов рака в плазме. Можно также применять аналогичную стратегию для отслеживания мутаций первичного рака и его метастазы или метастаз.

Варианты реализации будут полезны для скрининга на наличие рака у кажущихся здоровыми субъектов или у субъектов с определенными факторами риска (например, статусом курения, статусом носительства вируса (у субъектов-носителей вируса гепатита, субъектов, инфицированных вирусом папилломы человека)). Мутационный груз, который можно наблюдать в плазме таких субъектов, позволит выявить риск того, что у субъекта разовьется симптоматический рак в течение определенного периода времени. Таким образом, будут ожидать, что у субъектов с большим мутационным грузом в плазме риск повышен по сравнению с таковыми с меньшим мутационным грузом. Более того, временной профиль такого мутационного груза в плазме также будет мощным индикатором риска. Например, если у субъекта раз в год проверяют мутационный груз в плазме и если мутационные грузы постепенно увеличиваются, то данного субъекта следует направить на дополнительный скрининг рака, например, с применением рентгенографии органов грудной клетки, ультразвука, компьютерной томографии, магнитно-резонансной визуализации или позитронно-эмиссионной томографии.

С. Динамические пороговые значения для выявления мутаций в результатах секвенирования ДНК плазмы.

Для данного исследования привлекли четырех пациентов с печеночно-клеточной карциномой (ГКК) и одного пациента с раком яичников и молочной железы. Для последнего из упомянутых пациентов мы сконцентрировались на анализе рака яичников. Образцы крови собирали из каждого пациента до и после хирургической резекции опухолей. Также собирали иссеченные опухолевые ткани. ДНК выделяли из опухолевой ткани, белые кровяные клетки из образца крови до операции и образцов плазмы до и после операции секвенировали, применяя систему секвенирования HiSeq2000 (Illumina). Результаты секвенирования выравнивали с последовательностью референсного генома человека (hg18), применяя пакет программ для выравнивания коротких олигонуклеотидов Short Oligonucleotide Alignment Package 2 (SOAP2) (Li R. и др. *Bioinformatics* 2009; 25: 1966-1967). Последовательности ДНК белых кровяных клеток считали конститутивными последовательностями ДНК для каждого исследуемого субъекта.

В данном примере связанные с опухолью ОНМ сначала выявляли по результатам секвенирования ДНК плазмы и КГ, не учитывая опухолевые ткани. Затем полученные для плазмы результаты сравнивали с результатами секвенирования, полученными для опухолевых тканей (в качестве золотого стандарта), чтобы установить достоверность полученных результатов. В этом отношении получали золотой стандарт путем сравнения результатов секвенирования опухолевых тканей с конститутивной последовательностью, чтобы определить мутации в опухолевых тканях. В данном анализе мы сфокусировали внимание на положениях нуклеотидов, в которых конститутивная ДНК исследуемого субъекта была гомозиготной.

1) Ненаправленный полногеномный анализ.

Глубины секвенирования для белых кровяных клеток, опухолевых тканей и ДНК из плазмы каждого пациента показаны в табл. 5.

Таблица 5. Срединные глубины секвенирования различных образцов для четырех случаев ГКК

Случай	Срединная глубина секвенирования (кратность)			
	Белые кровяные клетки	Опухолевая ткань	Плазма до операции	Плазма после операции
ГКК1	39	29	23	24
ГКК2	39	29	25	28
ГКК3	46	33	18	21
ГКК4	46	27	20	23
Пациент с раком яичников	44	53	37	28

Динамические пороговые значения для минимальных встречаемостей для детектирования мутаций в плазме (r), показанные в табл. 1, использовали для определения мутаций в плазме каждого пациента. Так как глубина секвенирования каждого локуса может изменяться, пороговое значение может изменяться, что фактически обеспечивает зависимость порогового значения от суммарного количества считываемых последовательностей локуса. Например, хотя срединная глубина составляет менее 50 (табл. 5), глубина секвенирования отдельных локусов может сильно изменяться и покрытие может составлять >100 раз.

Вдобавок к ошибкам секвенирования другим источником ошибок будет выравнивание. Чтобы минимизировать данный тип ошибок, считываемые последовательности, содержащие мутацию, заново выравнивали с референсным геномом, применяя программу для выравнивания Bowtie (Langmead B. и др. *Genome Biol.* 2009, 10:R25). Только считываемые последовательности, которые можно было выровнять с уникальным положением в референсном геноме с помощью SOAP2 и Bowtie, применяли для последующего анализа мутаций в плазме. Также можно применять другие комбинации пакетов программного обеспечения для выравнивания, основанные на различных алгоритмах.

Для того чтобы дополнительно минимизировать ошибки секвенирования и выравнивания в фактических результатах секвенирования, мы применяли два дополнительных алгоритма фильтрации, чтобы определить положения нуклеотидов, в которых выявили однонуклеотидные вариации в считываемых последовательностях: (1) $\geq 70\%$ считываемых последовательностей, содержащих мутации, можно заново выровнять с такой же геномной координатой, применяя Bowtie с качеством картирования $\geq Q20$ (т.е. вероятность неправильного выравнивания $<1\%$); (2) $\geq 70\%$ считываемых последовательностей, содержащих мутации, находились за пределами 5 п.о. от обоих концов (т.е. 5'- и 3'-концов) считываемых последовательностей. Такой принцип фильтрации установили, так как ошибки секвенирования были более широко распространены в обоих концах считываемых последовательностей.

Мы также исследовали факторы, влияющие на детектирование опухоли, без какой-либо информации о геноме опухоли. Один такой параметр представлял собой фракционную концентрацию ДНК опухолевого происхождения в плазме. Данный параметр можно расценивать как другой параметр золотого стандарта, и его определили с целью сравнения с известной информацией о геноме опухоли, полученной

с применением GAAL.

В табл. 6 показаны вариации нуклеотидов, детектированные в плазме до и после лечения. Для ГКК1 без какой-либо информации о геноме опухоли обнаружили всего 961 однонуклеотидную вариацию. Среди данных вариаций нуклеотидов, детектированных в плазме, 828 представляли собой связанные с раком мутации. После хирургической резекции ГКК суммарное количество вариаций нуклеотидов уменьшилось до 43, и ни одна из них не являлась связанной с раком мутацией.

Для сравнения, фракционная концентрация ДНК опухолевого происхождения в образце плазмы до операции составляла 53%, и ее определили с помощью известной информации о геноме опухоли. У ГКК2, ГКК3 и ГКК4, без какой-либо информации о геномах опухоли, выявили количества однонуклеотидных вариаций в плазме в диапазоне от 27 до 32, для образцов плазмы до операции. Полученные результаты соответствуют математическому прогнозированию того, что при приблизительно 20-кратной глубине секвенирования в плазме можно обнаружить очень низкий процент связанных с раком мутаций, и большинство вариаций последовательности, детектированных в плазме, были следствием ошибок секвенирования. После резекции опухоли не наблюдалось существенного изменения количества детектированных вариаций последовательности. С целью сравнения, фракционные концентрации ДНК опухолевого происхождения в плазме находились в диапазоне от 2,1 до 5%, и их выявили с помощью известной информации о геномах опухолей.

Таблица 6. Вариации нуклеотидов, детектированные в плазме

	Плазма до операции			Плазма после операции		
	Фракционная концентрация ДНК опухолевого происхождения	Суммарное количество однонуклеотидных вариаций	Количество детектированных связанных с раком мутаций	Фракционная концентрация ДНК опухолевого происхождения	Суммарное количество однонуклеотидных вариаций	Количество детектированных связанных с раком мутаций
ГКК1	53%	961	828	0,4%	43	0
ГКК2	5%	32	0	0,6%	49	0
ГКК3	2,1%	29	0	0,2%	32	0
ГКК4	2,6%	27	0	1,3%	35	1
Пациент с раком яичников (и груди)	46%	1718	1502	0,2%	2	0

2) Целевое обогащение экзонами.

Выше обсуждалось, что увеличение глубины секвенирования интересующей области может повысить как чувствительность, так и специфичность определения связанных с раком мутаций в плазме и, следовательно, увеличить дискриминационную мощность между пациентами с раковыми заболеваниями и не страдающими раком субъектами. Хотя увеличение глубины секвенирования всего генома все еще очень дорогостоящее, одной альтернативой будет обогащение некоторыми фрагментами для секвенирования. В одном варианте реализации для секвенирования можно провести направленное обогащение выбранными экзонами или фактически всем экзомом. Данный подход может значительно увеличить глубину секвенирования целевой области без увеличения суммарного количества считываемых последовательностей.

Библиотеки секвенирования ДНК в плазме пациентов с ГКК и пациента, страдающего раком яичников (и груди), захватывали, применяя набор для целевого обогащения экзомом Agilent SureSelect All Exon. Обогащенные экзонами библиотеки секвенирования затем секвенировали, применяя систему секвенирования HiSeq 2000. Считываемые последовательности выравнивали с референсным геномом человека (hg18). После выравнивания анализировали наличие однонуклеотидных вариаций в считываемых последовательностях, уникально картированных на экзонах. Для определения однонуклеотидных вариаций в плазме при анализе захваченного экзома применяли динамические пороговые значения, показанные в табл. 2.

На фиг. 14А представлена таблица 1400, на которой показано детектирование однонуклеотидных вариаций в ДНК из плазмы ГКК1. Без какой-либо информации о геноме опухоли мы выявили в результатах целевого секвенирования всего 57 однонуклеотидных вариаций в плазме. При последующей проверке достоверности результатов секвенирования, полученных для опухолевых тканей, обнаружили, что 55 представляют собой истинные связанные с опухолью мутации. Ранее обсуждалось, что фракционная концентрация ДНК опухолевого происхождения в плазме до операции составляла 53%. После резекции опухоли в результатах целевого секвенирования, полученных для плазмы, не обнаружили однонуклео-

тидных вариаций. Полученные результаты указывают на то, что анализ количества однонуклеотидных вариаций в плазме можно применять для мониторинга прогрессирования заболевания у пациентов с раковыми заболеваниями.

На фиг. 14В представлена таблица 1450, на которой показано детектирование однонуклеотидных вариаций в ДНК из плазмы ГКК2. Без какой-либо информации о геноме опухоли мы выявили в результатах целевого секвенирования плазмы всего 18 однонуклеотидных вариаций. Все данные мутации были детектированы в опухолевых тканях. Ранее обсуждалось, что фракционная концентрация ДНК опухолевого происхождения в плазме до операции составляла 5%. После резекции опухоли в плазме не обнаружили однонуклеотидных вариаций. По сравнению с ГКК1, у которого фракционная концентрация ДНК опухолевого происхождения в плазме была выше, в случае ГКК2 в плазме обнаружили меньшее количество однонуклеотидных вариаций. Полученные результаты позволяют предположить, что количество однонуклеотидных вариаций в плазме можно применять в качестве параметра, который отражает фракционную концентрацию ДНК опухолевого происхождения в плазме и, следовательно, опухолевую нагрузку у пациента, так как показали, что концентрация ДНК опухолевого происхождения в плазме положительно коррелировала с опухолевой нагрузкой (Chan K.C. и др. Clin. Chem. 2005; 51:2192-5).

На фиг. 15А представлена таблица 1500, на которой показано детектирование однонуклеотидных вариаций в ДНК из плазмы ГКК3. Без какой-либо информации о геноме опухоли мы не наблюдали в результатах целевого секвенирования каких-либо однонуклеотидных вариаций в образцах плазмы как до, так и после резекции. Вероятно, это может быть следствием относительно низкой фракционной концентрации (2,1%) ДНК опухолевого происхождения в плазме данного пациента. Предположили, что дополнительное увеличение глубины секвенирования улучшит чувствительность детектирования связанных с раком мутаций в случаях с низкой фракционной концентрацией ДНК опухолевого происхождения.

На фиг. 15В представлена таблица 1550, на которой показано детектирование однонуклеотидных вариаций в ДНК из плазмы ГКК4. Без какой-либо информации о геноме опухоли мы выявили в результатах целевого секвенирования плазмы всего 3 однонуклеотидные вариации. Все данные мутации были детектированы в опухолевых тканях. По сравнению с ГКК1 и ГКК2, у которых фракционные концентрации ДНК опухолевого происхождения в плазме были выше, в случае ГКК4, у которого фракционная концентрация опухолевой ДНК в плазме составляла 2,6%, обнаружили меньшее количество однонуклеотидных вариаций в плазме. Полученные результаты позволяют предположить, что количество однонуклеотидных вариаций в плазме можно применять в качестве параметра, который отражает фракционную концентрацию ДНК опухолевого происхождения в плазме и опухолевую нагрузку у пациента.

На фиг. 16 представлена таблица 1600, на которой показано детектирование однонуклеотидных вариаций в ДНК плазмы пациента с раком яичников (и груди). Без какой-либо информации о геноме опухоли мы выявили в результатах целевого секвенирования плазмы всего 64 однонуклеотидные вариации. 59 из них были детектированы в тканях опухолей яичника. Предполагаемая фракционная концентрация в плазме ДНК, происходящей из опухоли яичника, составляла 46%. После резекции рака яичников обнаружили существенное уменьшение суммарного количества однонуклеотидных вариаций в плазме.

Вдобавок к применению системы целевого обогащения SureSelect (Agilent), мы также применяли систему целевого обогащения Nimblegen SeqCap EZ Exome+UTR (Roche) для обогащения последовательностями из экзонов для секвенирования. Система Nimblegen SeqCap покрывает экзонные области генома, а также 5'- и 3'-нетранслируемые области. Анализировали образцы из плазмы четырех пациентов с ГКК до лечения, двух здоровых контрольных субъектов и двух носителей хронического гепатита В, не страдающих раком (табл. 7). В других вариантах реализации можно применять другие системы целевого обогащения, включая, но не ограничиваясь системами с жидкофазной или твердофазной гибридизацией.

Таблица 7. Результаты секвенирования экзома четырех пациентов с ГКК (ГКК1-4) с применением системы целевого обогащения Nimblegen SeqCap EZ Exome+UTR для захвата последовательности

	Плазма до лечения			Плазма после лечения	
	Фракционная концентрация ДНК опухолевого происхождения в	Количество вариаций последовательности, детектированных в плазме, удовлетворяющих динамич-	Количество вариаций последовательности, которые перекрываются с мутациями, детектированными в	Количество вариаций последовательности, детектированных в плазме, удовлетворяющих	Количество вариаций последовательности, которые перекрываются с мутациями, детектированными в

	плазме, полученная путем анализа GAAL	чиским пороговым значениям	соответствующей опухолевой ткани	динамическим пороговым значениям	соответствующей опухолевой ткани
ГКК1	53%	69	64	1	1
ГКК2	5%	51	47	3	0
ГКК3	2,1%	0	0	1	0
ГКК4	2,6%	8	7	0	0

Анализ секвенирования плазмы ГКК3 до лечения был недостаточно оптимальным вследствие более высокого процента дублированных при ПЦР считываемых последовательностей.

У двух носителей хронического гепатита В и двух здоровых контрольных субъектов обнаружили одну или менее однонуклеотидных вариаций, которые удовлетворяли динамическим пороговым критериям (табл. 8). У трех из четырех пациентов с ГКК количество вариаций последовательности, детектированных в плазме, которые удовлетворяли динамическому пороговому требованию, составляло по меньшей мере 8. У ГКК3 не было детектировано ОНВ, которые бы удовлетворяли динамическому пороговому значению. В секвенированных считываемых последовательностях данного образца наблюдалось повышенное соотношение дублированных при ПЦР считываемых последовательностей, что приводило к меньшему количеству недублированных секвенированных считываемых последовательностей. После хирургической резекции опухоли наблюдалось заметное уменьшение количества ОНВ, детектированных в плазме.

Таблица 8. Результаты секвенирования экзома у 2 носителей хронического гепатита В (ВГВ1 и ВГВ2) и 2 здоровых контрольных субъектов (контроль 1 и контроль2) с применением системы целевого обогащения Nimblegen SeqCap EZ Exome+UTR для захвата последовательности

	Количество вариаций последовательности, детектированных в плазме, удовлетворяющих динамическим пороговым значениям
ВГВ1	0
ВГВ2	1
Контроль1	1
Контроль2	1

ХII. Гетерогенность опухоли.

Количественный анализ однонуклеотидных мутаций в биологическом образце (например, плазме/сыворотке) также полезен для анализа гетерогенности опухоли как внутриопухолевой, так и межопухолевой гетерогенности. Внутриопухолевая гетерогенность относится к существованию нескольких клонов опухолевых клеток внутри одной и той же опухоли. Межопухолевая гетерогенность относится к существованию нескольких клонов опухолевых клеток в двух или более опухолях одного и того же гистологического типа, но присутствующих в различных локализациях (либо в тех же органах, либо в различных органах). При некоторых типах опухолей существование опухолевой гетерогенности является плохим прогностическим показателем (Yoon H.H. и др. J. Clin. Oncol. 2012; 30: 3932-3938; Merlo L.M.F. и др. Cancer Prev. Res. 2010; 3: 1388-1397). При некоторых типах опухолей чем больше степень опухолевой гетерогенности, тем выше будет вероятность прогрессирования опухоли или развития устойчивых клонов после направленного лечения.

Хотя полагают, что раковые заболевания возникают в результате клональной экспансии одной опухолевой клетки, рост и развитие рака будет приводить к накоплению новых и различных мутаций в различных частях рака. Например, когда у пациента с раковым заболеванием развиваются метастазы, в опухоли, расположенной в исходном органе, и в метастатической опухоли будет множество одинаковых мутаций. Тем не менее, раковые клетки из двух локализаций также будут нести уникальный набор мутаций, которые отсутствуют в другой локализации опухоли. Ожидают, что мутации, которые одинаковы в двух локализациях, будут присутствовать в больших концентрациях, чем мутации, которые наблюдаются лишь в одной локализации опухоли.

А. Пример.

Мы проанализировали плазму крови пациента с двусторонним раком яичников и раком молочной железы. Обе опухоли яичника представляли собой серозную аденокарциному. Левая была размером 6 см, и правая была размером 12 см в самом длинном направлении. Также присутствовало несколько метастаз в толстом кишечнике и сальнике. ДНК, выделенную из лейкоцитов, секвенировали, применяя платформу для секвенирования путем синтеза от Illumina в среднем с 44-кратным покрытием гаплоидного генома. Дополнительно анализировали однонуклеотидные мутации в плазме для положений нуклеотидов, в которых присутствовала только одна аллель, т.е. гомозиготная аллель.

ДНК выделяли из четырех различных локализаций левой и правой опухолей и секвенировали, применяя платформу секвенирования Illumina. Две локализации (локализации А и В) были из правой опухоли.

ли и другие две локализации (локализации С и D) были из левой опухоли. Локализации А и В находились приблизительно в 4 см друг от друга. Расстояние между локализациями С и D также составляло приблизительно 4 см. Образцы плазмы собирали из пациента до и после хирургической резекции опухолей яичника. Затем выделяли ДНК из плазмы пациента. Глубины секвенирования опухолей из локализаций А, В, С и D, а также из образцов плазмы показаны в табл. 9.

Таблица 9. Глубина секвенирования опухолей из локализаций А, В, С и D

Образец	Количество первичных считываемых последовательностей при секвенировании	Количество считываемых последовательностей	Кратность покрытия гаплоидного генома
Конститутивная ДНК из лейкоцитарной пленки	1091250072	876269922	43,81
Опухоль правого яичника	1374495256	1067277229	53,36
Опухоль правого яичника	934518588	803007464	40,15
Опухоль левого яичника (локализация С)	1313051122	1036643946	51,83
Опухоль левого яичника (локализация)	1159091833	974823207	48,74
Образец плазмы, собранный до хирургического вмешательства	988697457	741982535	37,10
Образец плазмы, собранный после хирургического вмешательства	957295879	564623127	28,23

В данном примере для детектирования одной связанной с опухолью однонуклеотидной мутации положение нуклеотида секвенировали по меньшей мере 20 раз в опухолевой ткани и 30 раз в конститутивной ДНК. В других вариантах реализации можно применять другие глубины секвенирования, например 35, 40, 45, 50, 60, 70, 80, 90, 100 и >100 раз. Снижение стоимости секвенирования позволит с гораздо большей готовностью осуществлять секвенирование с большей глубиной. Положение нуклеотида гомозиготно в конститутивной ДНК, тогда как в опухолевой ткани наблюдается замена нуклеотида. Критерий встречаемости замены нуклеотида в опухолевой ткани зависит от суммарной глубины секвенирования конкретного положения нуклеотида в опухолевой ткани. Для 20-30-кратного покрытия нуклеотида встречаемость замены нуклеотида (пороговое значение) составит по меньшей мере пять раз. Для 31-50-кратного покрытия встречаемость замены нуклеотида составит по меньшей мере шесть раз. Для 51-70-кратного покрытия необходимая встречаемость составит по меньшей мере семь раз. Данные критерии получены в результате прогнозирования с применением пуассоновского распределения чувствительности детектирования истинных мутаций и ожидаемого количества ложноположительных локусов.

На фиг. 17 представлена таблица 1700, на которой показаны прогнозируемые чувствительности различных требований встречаемости и глубин секвенирования. Чувствительность будет соответствовать количеству истинных мутаций, детектированных при конкретной глубине покрытия с применением конкретного порогового значения. Чем выше глубина секвенирования, тем с большей вероятностью обнаружат мутацию для данного порогового значения, так как будет получено больше считываемых последовательностей с мутацией. Для более высоких пороговых значений вероятность обнаружить мутацию будет меньше, поскольку критерий будет более строгим.

На фиг. 18 представлена таблица 1800, на которой показаны прогнозируемые количества ложноположительных локусов для различных пороговых значений и различных глубин секвенирования. Количество ложноположительных результатов возрастает с увеличением глубины секвенирования, так как осуществляют больше считываемых последовательностей. Тем не менее, для порогового значения, равного пяти или более, не прогнозируется получение ложноположительных результатов, даже при глубине секвенирования, составляющей 70 раз. В других вариантах реализации можно применять различные критерии встречаемости, чтобы добиться желательной чувствительности и специфичности.

На фиг. 19 показана древовидная диаграмма, иллюстрирующая количество мутаций, детектирован-

ных в различных локализациях опухоли. Мутации определяли путем непосредственного секвенирования опухолей. В локализации А обнаружили 71 мутацию, которая специфична для данной опухоли, а в локализации В обнаружили 122 специфичные для данной локализации мутации, даже с учетом того, что они находились на расстоянии всего 4 см друг от друга. 10 мутаций наблюдали одновременно в локализациях А и В. В локализации С обнаружили 168 мутаций, которые специфичны для данной опухоли, а в локализации D обнаружили 248 специфичных для данной локализации мутаций, даже с учетом того, что они находились на расстоянии всего 4 см друг от друга. 12 мутаций наблюдали одновременно в локализациях С и D. Обнаружили существенную гетерогенность в мутационных профилях для различных локализаций опухоли. Например, 248 мутаций были детектированы только в локализации опухоли D, но не детектированы в других трех локализациях опухоли. Всего было детектировано 2129 мутаций во всех локализациях. Таким образом, многие мутации были одинаковы в различных опухолях. Таким образом, было всего семь групп ОНВ. Не наблюдалось различий среди данных четырех областей в отношении вариаций числа копий.

На фиг. 20 представлена таблица 2000, на которой показано количество фрагментов, несущих мутации опухолевого происхождения, в образце плазмы до лечения и после лечения. Также показаны выявленные фракционные концентрации ДНК опухолевого происхождения, несущей соответствующие мутации. Категория мутации относится к локализации(ям) опухоли, в которой обнаружили данную мутацию. Например, категория мутации А относится к мутации, присутствующей только в локализации А, тогда как категория мутации ABCD относится к мутации, присутствующей во всех четырех локализациях опухоли.

Из 2129 мутаций, которые присутствовали во всех четырех локализациях опухоли, 2105 (98,9%) были детектированы по меньшей мере в одном фрагменте ДНК в плазме. С другой стороны, из 609 мутаций, которые присутствовали только в одной из четырех локализаций опухоли, лишь 77 (12,6%) были детектированы по меньшей мере в одном фрагменте ДНК в плазме. Следовательно, количественный анализ однонуклеотидных мутаций в плазме можно применять для выявления относительной распространенности данных мутаций в опухолевых тканях. Данная информация будет полезна для исследования гетерогенности рака. В данном примере вариацию считали потенциальной мутацией, если ее наблюдали в результатах секвенирования хотя бы один раз.

Внутри каждой группы ОНВ определяли фракционные концентрации циркулирующей опухолевой ДНК. Фракционные концентрации опухолевой ДНК в плазме до хирургического вмешательства и после хирургического вмешательства, которые определяли по ОНВ, присутствующим во всех 4 областях (т.е. группа ABCD), составляли 46 и 0,18% соответственно. Данные процентные соотношения хорошо коррелировали с таковыми, полученными при анализе GAAL: 46 и 0,66%. Мутации, присутствующие во всех 4 областях (т.е. группа ABCD), вносили наибольший фракционный вклад в ДНК опухолевого происхождения в плазме.

Фракционные концентрации ДНК опухолевого происхождения в плазме до операции, которые определяли по ОНВ из групп АВ и CD, составляли 9,5 и 1,1% соответственно. Данные концентрации соответствовали относительным размерам опухолей правого и левого яичников. Фракционные концентрации ДНК опухолевого происхождения, которые определяли по уникальным для одной области ОНВ (т.е. из групп А, В, С и D), как правило, были низкими. Полученные результаты позволяют предположить, что для точного измерения суммарной опухолевой нагрузки у пациента с раковым заболеванием применение полногеномного подхода "выстрела из дробового ружья" позволяет получить более типичную картину по сравнению с более традиционным подходом нацеливания на определенные связанные с опухолью мутации. В последнем из упомянутых подходов, если лишь у подмножества опухолевых клеток есть целевые мутации, то можно упустить важную информацию в отношении угрожающего рецидива или прогрессирования заболевания, вызванного опухолевыми клетками, не содержащими целевые мутации, или можно упустить возникновение устойчивого к лечению клона.

На фиг. 21 представлена диаграмма 2100, на которой показаны распределения встречаемости в плазме мутаций, детектированных в одной локализации опухоли, и мутаций, детектированных во всех четырех локализациях опухоли. На столбчатой диаграмме 2100 показаны результаты для двух типов мутаций: (1) мутаций, детектированных лишь в одной локализации, и (2) мутаций, детектированных во всех четырех локализациях опухоли. На горизонтальной оси представлено количество раз, которое мутация детектирована в плазме. На вертикальной оси показан процент мутаций, которые соответствуют конкретному значению на горизонтальной оси. Например, приблизительно 88% мутаций типа (1) встретились в плазме лишь один раз. Можно видеть, что мутации, которые встретились в одной локализации, обнаруживали преимущественно один раз и не более четырех раз. Мутации, присутствующие только в одной локализации опухоли, обнаруживались в плазме гораздо реже по сравнению с мутациями, присутствующими во всех четырех локализациях опухоли.

Одно из применений данного способа состоит в том, что практикующие врачи получают возможность оценить массу опухолевых клеток, несущих различные классы мутаций. Часть данных мутаций потенциально можно будет компенсировать целевыми агентами. Ожидают, что агенты, нацеленные на мутации, которые содержатся в большем количестве опухолевых клеток, будут обладать ярко выражен-

ным терапевтическим действием.

На фиг. 22 представлена диаграмма 2200, на которой показано прогнозируемое распределение встречаемости в плазме мутаций, происходящих от гетерогенной опухоли. Указанная опухоль включает две группы мутаций. Одна группа мутаций присутствует во всех опухолевых клетках, а другая группа мутаций присутствует лишь в $\frac{1}{4}$ опухолевых клеток, на основании приближения, что две локализации представляют опухоли каждого яичника. Суммарную фракционную концентрацию ДНК опухолевого происхождения в плазме принимали равной 40%. Предположили, что образец плазмы секвенировали со средней глубиной 50 раз на положение нуклеотида. Согласно данному прогнозируемому распределению встречаемости в плазме мутации, которые присутствовали во всех опухолевых тканях, можно было отличить от мутаций, присутствующих лишь в $\frac{1}{4}$ опухолевых клеток по их встречаемости в плазме. Например, в качестве порогового значения можно применять встречаемость 6 раз. Для мутаций, присутствующих во всех опухолевых клетках, 92,3% мутаций будет присутствовать в плазме по меньшей мере 6 раз. Напротив, для мутаций, которые присутствуют в $\frac{1}{4}$ опухолевых клеток, лишь 12,4% мутаций будет присутствовать в плазме по меньшей мере 6 раз.

На фиг. 23 представлена таблица 2300, демонстрирующая специфичность вариантов реализации для 16 здоровых контрольных субъектов. Образцы их ДНК из плазмы секвенировали со средним покрытием 30 раз. Осуществляли детектирование мутаций, которые присутствовали в плазме описанного выше пациента с раком яичников, в образцах плазмы данных здоровых субъектов. Мутации, присутствующие в опухоли пациента с раком яичников, очень редко обнаруживали в результатах секвенирования плазмы здоровых контрольных субъектов, и ни одна из категорий мутаций не встречалась с фракционной концентрацией $>1\%$. Полученные результаты показали, что данный способ детектирования высоко специфичен.

В. Способ.

На фиг. 24 представлена блок-схема способа 2400 анализа гетерогенности одной или более опухолей субъекта согласно вариантам реализации настоящего изобретения. Некоторые этапы способа 2400 можно осуществить, как описано в данной заявке.

В блоке 2410 получали конститутивный геном субъекта. В блоке 2420 получали один или более маркеров последовательности для каждого из множества фрагментов ДНК в биологическом образце субъекта, при этом биологический образец содержал внеклеточную ДНК. В блоке 2430 определяли положения в геноме маркеров последовательности. В блоке 2440 маркеры последовательности сравнивали с конститутивным геномом, чтобы определить первое количество первых локусов. В каждом из первых локусов количество маркеров последовательности с вариантной последовательностью относительно конститутивного генома было выше порогового значения, где пороговое значение больше единицы.

В блоке 2450 рассчитывали меры гетерогенности одной или более опухолей на основании соответствующих первых количеств из набора первых положений в геноме. В одном аспекте указанные меры могут дать значение, которое представляет собой количество мутаций, которые одинаковы среди опухолей, по сравнению с количеством мутаций, которые отличаются среди опухолей. Здесь различные опухоли могут существовать как один объект, с различными опухолями внутри объекта, которые могут представлять собой то, что обычно называют внутриопухолевой гетерогенностью. Указанная мера также может относиться к тому, присутствуют ли мутации в одной или нескольких опухолях, или мутации присутствуют во многих или в большинстве опухолей. Можно рассчитать более чем одну меру гетерогенности.

В блоке 2460 меру гетерогенности можно сравнить с пороговым значением, чтобы определить классификацию уровня гетерогенности. Одну или более мер можно применять различными способами. Например, одну или более мер гетерогенности можно применять для прогнозирования вероятности прогрессирования опухоли. Чем выше гетерогенность некоторых опухолей, тем выше вероятность, что они будут прогрессировать и тем выше вероятность возникновения устойчивого клона после лечения (например, направленного лечения).

С. Меры гетерогенности опухоли.

Одним примером меры гетерогенности является количество "зон концентраций" различных групп мутаций в плазме. Например, если у пациента преобладает два опухолевых клона и если данные клоны присутствуют в различных концентрациях, то мы будем ожидать увидеть две различные мутации с различными концентрациями в плазме. Данные различные значения можно вычислить путем определения фракционной концентрации для различных наборов мутаций, где каждый набор соответствует одной из опухолей.

Каждую из данных концентраций можно назвать "зоной концентрации" или "классом концентрации". Если у пациента присутствует больше клонов, то будет наблюдаться больше зон/классов концентрации. Таким образом, чем больше зон, тем выше гетерогенность. Количество зон концентрации можно наблюдать при нанесении на график фракционных концентраций для различных мутаций. Можно построить гистограмму различных концентраций, где различные пики соответствуют различным опухолям (или различным клонам одной опухоли). У мутаций, которые одинаковы у всех или некоторых опухолей (или клонов опухоли), вероятно будет больший пик. Данные пики можно проанализировать, чтобы определить, какие меньшие пики объединены с получением большего пика. Можно применять процедуру

подгонки, например, аналогичную процедуру подгонки на фиг. 10В и 11.

В одном варианте реализации гистограмма представляет собой график, на котором ось у представляет собой количество (например, число или пропорцию) локусов и ось x представляет собой фракционную концентрацию. Мутации, которые одинаковы у всех или некоторых опухолей, дадут более высокую фракционную концентрацию. Размер пика будет представлять количество локусов, которые дают определенную фракционную концентрацию. Относительный размер пиков при низкой и высокой концентрации будет отражать степень гетерогенности опухолей (или клонов опухоли). Большой пик при высокой концентрации отражает, что большинство мутаций одинаковы в большинстве или во всех опухолях (или клонах опухоли) и свидетельствует о меньшей степени гетерогенности опухоли. Если пик при низкой концентрации больше, то большинство мутаций одинаковы в нескольких опухолях (или в нескольких клонах опухоли). Это будет свидетельствовать о более высокой степени гетерогенности опухоли.

Чем больше присутствует пиков, тем больше присутствует сайт-специфичных мутаций. Каждый пик может соответствовать различному набору мутаций, где набор мутаций происходит из подмножества опухолей (например, только из одной или двух опухолей, как описано выше). В примере на фиг. 19 может быть всего 7 пиков: 4 пика, каждый из которых для единственной локализации, которые обладают, вероятно, наименьшей концентрацией (в зависимости от относительного размера опухолей), два пика для локализации АВ и локализации CD и один пик для мутаций, одинаковых во всех локализациях.

Положение пиков также может указывать на относительный размер опухолей. Большая концентрация будет коррелировать с большей опухолью, так как большая опухоль будет высвобождать больше опухолевой ДНК в образец, например в плазму. Таким образом, можно оценить массу опухолевых клеток, несущих различные классы мутаций.

Другой пример меры гетерогенности представляет собой соотношение мутированных сайтов с относительно малым количеством вариантных считываемых последовательностей (например, 4, 5 или 6) по сравнению с соотношением мутированных сайтов с относительно большим количеством вариантных считываемых последовательностей (например, 9-13). Снова возвращаясь к фиг. 22, можно увидеть, что для специфичных для конкретной локализации мутаций вариантных считываемых последовательностей меньше (что также приводит к меньшей фракционной концентрации). Для общих для локализаций мутаций вариантных считываемых последовательностей больше (что также приводит к большей фракционной концентрации). Первое соотношение при 6 считываемых последовательностях (меньшее число), деленное на второе соотношение при 10 считываемых последовательностях (большее число) выражает меру гетерогенности. Если отношение мало, то присутствует несколько мутаций, специфичных для конкретной локализации, и, таким образом, уровень гетерогенности низкий. Если отношение велико (или, по меньшей мере, больше, чем значения, определенные для известных образцов), то уровень гетерогенности выше.

D. Определение пороговых значений.

Пороговые значения можно определить по субъектам, опухоли которых взяли на биопсию (например, как описано выше), чтобы непосредственно определить уровень гетерогенности. Указанный уровень можно определить различными способами, например, с помощью отношения специфичных для конкретной локализации мутаций к общим для локализаций мутациям. Биологические образцы (например, образцы плазмы) затем можно проанализировать, чтобы определить меры гетерогенности, при этом мера гетерогенности для биологических образцов может быть связана с уровнем гетерогенности, определенным при непосредственном анализе клеток опухолей.

Такая процедура позволит откалибровать пороговые значения относительно уровней гетерогенности. Если мера гетерогенности в данном анализе попадает между двумя пороговыми значениями, то можно оценить, что уровень гетерогенности находится между уровнями, соответствующими пороговым значениям.

В одном варианте реализации калибровочную кривую можно рассчитать по уровням гетерогенности, определенным для биоптатов, и соответствующей мере гетерогенности, определенной для образца плазмы (или другого образца). В таком примере уровни гетерогенности числовые, где данные числовые уровни могут соответствовать различным классификациям. Различные диапазоны числовых уровней могут соответствовать различным диагнозам, например различным стадиям рака.

E. Способ с применением фракционной концентрации из геномного представления.

Гетерогенность опухоли также можно проанализировать, применяя фракционную концентрацию, например, определенную с применением вариантов реализации согласно способу 1200. Области генома, в которых детектирована утрата одной копии, могут происходить из различных опухолей. Таким образом, фракционная концентрация, определенная для различных областей генома, может отличаться в зависимости от того, присутствует ли амплификация (или делеция для утраты 1 копии) только в одной опухоли или в нескольких опухолях. Таким образом, для фракционных концентраций, определенных с помощью вариантов реализации способа 1200, можно применять одинаковые меры гетерогенности.

Например, можно обнаружить, что одна область генома соответствует утрате 1 копии, и можно определить фракционную концентрацию только по соответствующей плотности в данной области генома (соответствующую плотность можно использовать в качестве фракционной концентрации). Можно по-

строить гистограмму по различным соответствующим плотностям путем вычисления количества областей с различными плотностями. Если только у одной опухоли, или одного клона опухоли, или одной метастазы опухоли обнаружили приобретение копии в конкретной области, то плотность данной области будет меньше, чем плотность области, в которой обнаружили приобретение копии в нескольких опухолях, или нескольких клонах опухоли, или нескольких метастазах опухоли (т.е. фракционная концентрация опухолевой ДНК в общей для мутаций области будет больше, чем в области с отдельной мутацией). Таким образом, меры гетерогенности, описанные выше, можно применять к пикам, детектированным по приобретению или утрате числа копий в различных областях, а также к фракционной концентрации в различных локализациях, в которых выявили распределение фракционных концентраций.

В одном варианте реализации, если для построения гистограммы применяют соответствующие плотности, можно получить разделение приобретений и утрат. Области, в которых обнаружили приобретение, можно проанализировать отдельно путем построения гистограммы только для приобретений, и отдельную гистограмму можно построить только для утрат. При применении фракционной концентрации пики утрат и приобретений можно проанализировать вместе. Например, для фракционных концентраций можно использовать разность (например, в виде абсолютного значения) с референсной плотностью, и, следовательно, фракционные концентрации для приобретений и утрат могут вносить вклад в один и тот же пик.

XIII. Компьютерная система.

В любых компьютерных системах, упомянутых в данной заявке, может применяться любое подходящее количество подсистем. Примеры таких подсистем показаны на фиг. 25 в компьютерном устройстве 2500. В некоторых вариантах реализации компьютерная система включает отдельное компьютерное устройство, где подсистемы могут представлять собой компоненты компьютерного устройства. В других вариантах реализации компьютерная система может включать несколько компьютерных устройств, каждое из которых представляет собой подсистему, с внутренними компонентами.

Подсистемы, показанные на фиг. 25, взаимосвязаны посредством системной шины 2575. Показаны дополнительные подсистемы, такие как принтер 2574, клавиатура 2578, жесткий диск 2579, монитор 2576, который соединен с видеоадаптером 2582, и другие подсистемы. Периферические устройства и устройства ввода/вывода (I/O), которые соединены с контроллером ввода/вывода 2571, можно подсоединить к компьютерной системе с помощью любого количества средств, известных в данной области, таких как порт последовательного ввода-вывода 2577. Например, порт последовательного ввода-вывода 2577 или внешний интерфейс 2581 (например, локальную сеть Ethernet, беспроводную сеть Wi-Fi и т.д.) можно применять для подсоединения компьютерной системы 2500 к глобальной компьютерной сети, такой как Интернет, к устройству ввода "мышь" или к сканеру. Взаимосвязь посредством системной шины 2575 позволяет центральному процессору 2573 обмениваться данными с каждой подсистемой и контролировать выполнение инструкций от системной памяти 2572 или жесткого диска 2579, а также обмениваться информацией между подсистемами. Системная память 2572 и/или жесткий диск 2579 может включать машиночитаемые носители. Любые значения, упомянутые в данной заявке, могут передаваться от одного компонента к другому компоненту и могут выводиться пользователю.

Компьютерная система может включать множество одинаковых компонентов или подсистем, например, соединенных друг с другом через внешний интерфейс 2581 или через внутренний интерфейс. В некоторых вариантах реализации компьютерные системы, подсистемы или устройства могут обмениваться информацией через сеть. В таких случаях один компьютер можно считать клиентом, а другой компьютер - сервером, где каждый может представлять собой часть одной и той же компьютерной системы. Как клиент, так и сервер может включать несколько систем, подсистем или компонентов.

Должно быть очевидно, 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, как правило, с программируемым процессором в модульной или интегральной форме. В данной заявке 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 описания и идеи, предложенных в данной заявке, среднему специалисту в данной области будут известны и понятны другие пути и/или способы осуществления вариантов реализации настоящего изобретения с применением аппаратного обеспечения и комбинаций аппаратного обеспечения и программного обеспечения.

Любые компоненты или функции программного обеспечения, описанные в данной заявке, можно осуществить в виде программного кода, который выполнит процессор, применяя любой подходящий язык программирования, такой как, например, Java, C++ или Perl, с применением, например, обычных или объектно-ориентированных методик. Программный код можно хранить в виде ряда инструкций или команд на машиночитаемых носителях 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 тому подобные носители. Машиночи-

таемые носители могут представлять собой любую комбинацию таких устройств хранения или передачи.

Такие программы также могут кодироваться и передаваться с помощью несущих сигналов, пригодных для передачи через проводные, оптические и/или беспроводные сети, соответствующих множеству протоколов, включая Интернет. В связи с этим машиночитаемые носители согласно варианту реализации настоящего изобретения можно получить, применяя информационный сигнал, кодируемый такими программами. Машиночитаемые носители, кодированные с помощью программного кода, можно упаковать вместе с совместимым устройством или предоставить отдельно от других устройств (например, посредством загрузки через Интернет). Любые такие машиночитаемые носители могут находиться внутри или на едином программном продукте (например, жестком диске, CD или всей компьютерной системе), и могут присутствовать внутри 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 более 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 пронумерованы, этапы согласно способам, описанным в данной заявке, можно осуществить 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 данных этапов.

Специфические детали конкретных вариантов реализации можно объединить любым подходящим способом, не отклоняясь 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 данных отдельных аспектов.

Предшествующее описание типичных вариантов реализации настоящего изобретения представлено с целью иллюстрирования и описания. Не предполагается, что оно является исчерпывающим или ограничивает настоящее изобретение точно описанной формой, и возможны многие модификации и вариации в свете описанных выше идей. Варианты реализации были выбраны и описаны для того, чтобы лучше объяснить принципы настоящего изобретения и его практического применения, чтобы с помощью этого дать возможность другим специалистам в данной области лучше применять настоящее изобретение в различных вариантах реализации и с различными модификациями, которые подходят для конкретного предполагаемого применения.

Предполагается, что применение формы единственного числа означает "один или более", если явно не указано иное.

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

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ детектирования рака или предзлокачественного изменения у субъекта, причем указанный способ включает

получение конститутивного генома для генома субъекта, где этот конститутивный геном получен с использованием считываемых последовательностей из образца субъекта, который содержит более 50% ДНК из здоровых клеток;

получение одной или более считываемых последовательностей для каждого из множества фрагментов ДНК в биологическом образце субъекта, причем указанный биологический образец включает внеклеточную ДНК;

определение геномных положений для считываемых последовательностей;

сравнение считываемых последовательностей с указанным конститутивным геномом для определения первого количества первых локусов, при этом

в каждом из данных первых локусов количество считываемых последовательностей с вариантной последовательностью относительно этого конститутивного генома выше значения отсечки, где указанное значение отсечки больше единицы для уменьшения тем самым ложноположительных результатов в отношении истинных ассоциированных с раком мутаций;

определение параметра, как раскрыто в описании, на основании числа считываемых последовательностей с вариантной последовательностью в этих первых локусах; и

сравнение этого параметра с пороговым значением для определения уровня рака у субъекта, где пороговое значение соответствует диапазону параметра для субъектов, имеющих этот уровень рака.

2. Способ по п.1, отличающийся тем, что значение отсечки для локуса зависит от общего количества считываемых последовательностей, которые располагаются в геноме в данном локусе.

3. Способ по п.1, отличающийся тем, что различные значения отсечки используют по меньшей мере для двух из указанных первых локусов, причем способ дополнительно включает динамическое определение первого значения отсечки для одного из этих первых локусов, где указанный один из этих первых локусов находится внутри первой области.

4. Способ по п.3, отличающийся тем, что первое значение отсечки определяют на основании глубины секвенирования одного из первых локусов или на основе частоты ложноположительных результатов, которая зависит от частоты ошибок секвенирования, глубины секвенирования указанной первой области и общего количества положений нуклеотидов в первой области.

5. Способ по п.3, отличающийся тем, что первое значение отсечки определяют на основании частоты ложноположительных результатов, которая зависит от частоты ошибок секвенирования, глубины секвенирования первой области и общего количества положений нуклеотидов в первой области, и где первое значение отсечки далее выбирают на основании ожидаемого числа положений с истинными ассоциированными с раком мутациями в этой первой области, причем способ дополнительно включает расчет указанного ожидаемого числа положений с истинными ассоциированными с раком мутациями для первого значения отсечки на основании глубины секвенирования D первой области и фракционной концентрации f ДНК опухолевого происхождения в биологическом образце.

6. Способ по п.5, отличающийся тем, что расчет числа положений с истинными ассоциированными с раком мутациями делают, используя пуассоновское распределение вероятностей согласно формуле

$$P_b = 1 - \sum_{i=0}^{r-1} \text{Poisson}(i, M_p),$$

где P_b представляет собой вероятность детектирования положений с истинными ассоциированными с раком мутациями, r соответствует значению первой отсечки

$$M_p = D \times f / 2.$$

7. Способ по п.3, отличающийся тем, что указанное первое значение отсечки определяют, применяя любой из следующих критериев:

если глубина секвенирования меньше 50, то первое значение отсечки равно 5,

если глубина секвенирования составляет 50-110, то первое значение отсечки равно 6,

если глубина секвенирования составляет 111-200, то первое значение отсечки равно 7,

если глубина секвенирования составляет 201-310, то первое значение отсечки равно 8,

если глубина секвенирования составляет 311-450, то первое значение отсечки равно 9,

если глубина секвенирования составляет 451-620, то первое значение отсечки равно 10, и

если глубина секвенирования составляет 621-800, то первое значение отсечки равно 11.

8. Способ по п.1, отличающийся тем, что указанный параметр представляет собой взвешенную сумму первого количества первых локусов, при этом вклад каждого из этих первых локусов взвешен по показателю значимости, присвоенному соответствующим первым локусам.

9. Способ по п.1, отличающийся тем, что параметр включает сумму считываемых последовательностей, имеющих вариантную последовательность в первом количестве первых локусов.

10. Способ по п.9, отличающийся тем, что сумма представляет собой взвешенную сумму, и при этом первый вес одного из первых локусов отличается от второго веса второго из первых локусов, где первый вес больше, чем второй вес, и при этом один из первых локусов связан с раком, а второй из первых локусов не связан с раком.

11. Способ по п.1, в котором определение положения в геноме считываемой последовательности включает

выравнивание по меньшей мере части считываемой последовательности с референсным геномом, при этом при выравнивании считываемой последовательности допускается одно или более несовпадений между считываемой последовательностью и указанным конститутивным геномом,

где сравнение считываемых последовательностей с этим конститутивным геномом включает

сравнение конститутивного генома с референсным геномом для определения второго количества вторых локусов, содержащих вариант по сравнению с референсным геномом;

определение на основании выравнивания третьего количества третьих локусов, при этом

в каждом из данных третьих локусов количество считываемых последовательностей с вариантной последовательностью по сравнению с референсным геномом выше значения отсечки; и

вычисление разности между третьим количеством и вторым количеством с получением первого количества первых локусов, где вычисление указанной разности между третьим количеством и вторым количеством идентифицирует первые локусы и где определение указанного параметра включает для каждого локуса из первого количества первых локусов

определение количества считываемых последовательностей, которые выравниваются с локусом и содержат вариантную последовательность в данном локусе; и

определение указанного параметра на основании соответствующих количеств.

12. Способ по п.1, отличающийся тем, что определение положения в геноме для считываемой последовательности включает выравнивание по меньшей мере части считываемой последовательности с

указанным конститутивным геномом, при этом допускается одно или более несовпадений между считываемой последовательностью и этим конститутивным геномом, где сравнение считываемых последовательностей с конститутивным геномом включает идентификацию считываемых последовательностей, которые содержат вариантную последовательность в локусе по сравнению с указанным конститутивным геномом; и для каждого локуса, имеющего вариантную последовательность, подсчет соответствующего первого количества считываемых последовательностей, которые выравниваются с указанным локусом и содержат вариантную последовательность в этом локусе.

13. Способ по п.12, отличающийся тем, что определение указанного параметра дополнительно включает

сложение соответствующих первых количеств с получением первой суммы; и

вычитание количества локусов, демонстрирующих вариантную последовательность, из первой суммы или нормировку первой суммы по количеству выровненных считываемых последовательностей.

14. Способ по п.1, дополнительно включающий

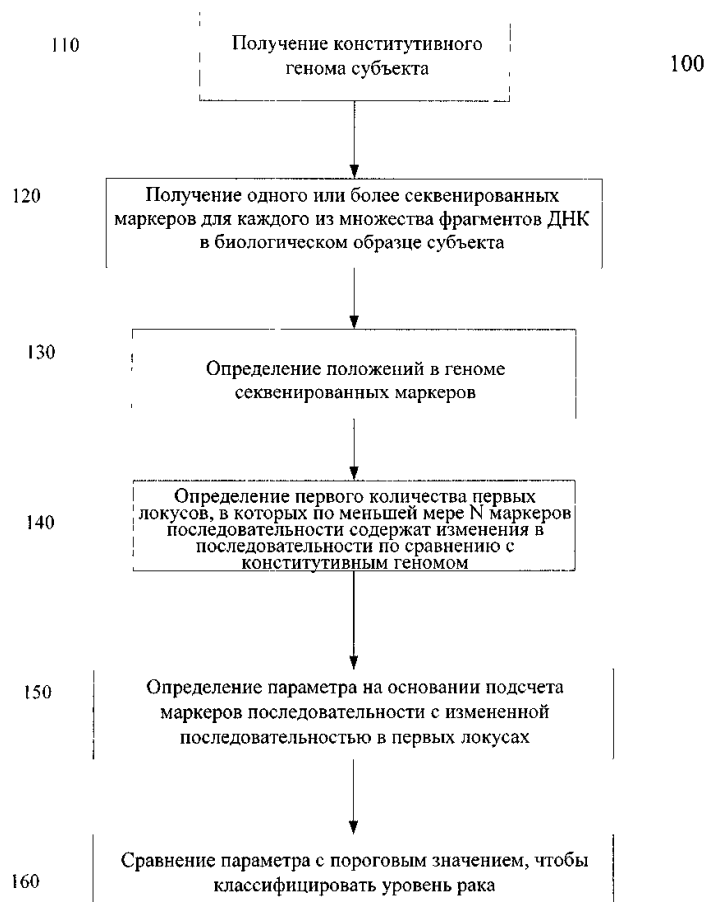
получение конститутивного образца у субъекта, который содержит более 90% конститутивной ДНК;

осуществление случайного секвенирования фрагментов ДНК в конститутивном образце с получением одной или более вторых считываемых последовательностей для каждого из множества фрагментов ДНК в конститутивном образце;

выравнивание по меньшей мере части вторых считываемых последовательностей с референсным геномом, при этом допускается несовпадение между второй считываемой последовательностью и конститутивным геномом в M или меньшем количестве геномных локусов, где M представляет собой целое значение, равное или большее, чем единица; и конструирование конститутивного генома на основании вторых считываемых последовательностей и выравнивания, где конститутивный образец представляет собой биологический образец и где конструирование конститутивного генома включает определение, является ли локус гомозиготным локусом или гетерозиготным локусом, содержащим две аллели.

15. Способ по п.1, отличающийся тем, что одну или более считываемых последовательностей получают в результате случайного секвенирования фрагментов ДНК в биологическом образце.

16. Носитель данных, содержащий инструкции для компьютерного устройства, запрограммированного для осуществления способа по любому из предшествующих пунктов.



Фиг. 1



Фиг. 2



Фиг. 3

Количество обнаружений	1х		2х		3х		4х		5х		6х	
	Ожидаемое количество ложноположительных локусов для ошибок секвенирования	Ожидаемое количество обнаруженных мутаций	Ожидаемое количество ложноположительных локусов для ошибок секвенирования	Ожидаемое количество обнаруженных мутаций	Ожидаемое количество ложноположительных локусов для ошибок секвенирования	Ожидаемое количество обнаруженных мутаций	Ожидаемое количество ложноположительных локусов для ошибок секвенирования	Ожидаемое количество обнаруженных мутаций	Ожидаемое количество ложноположительных локусов для ошибок секвенирования	Ожидаемое количество обнаруженных мутаций	Ожидаемое количество ложноположительных локусов для ошибок секвенирования	Ожидаемое количество обнаруженных мутаций
1	2998500	146	1499	4	0	0	0	0	0	0	0	0
10	29850499	1180	45000	271	40	43	0	5	0	1	0	0
20	59403980	1896	190000	793	380	241	1	57	0	11	0	2
40	117631683	2594	780000	1782	3293	970	10	429	0	158	0	50
60	174706399	2851	1770000	2403	11407	1730	54	1058	0	554	0	252
80	230650961	2945	3160000	2725	27387	2286	176	1700	1	1113	0	645
100	285487746	2980	4950000	2879	53900	2626	436	2205	3	1679	0	1152
120	339238690	2993	7140000	2948	93613	2814	913	2546	7	2145	0	1663
140	391925294	2997	9730000	2978	149193	2911	1703	2755	15	2481	0	2098
160	443568633	2999	12720000	2991	223307	2959	2922	2873	30	2701	0	2426
180	494189366	3000	16110000	2996	318620	2981	4700	2936	55	2835	1	2653
200	543807741	3000	19900000	2999	437800	2992	7187	2969	94	2912	1	2799

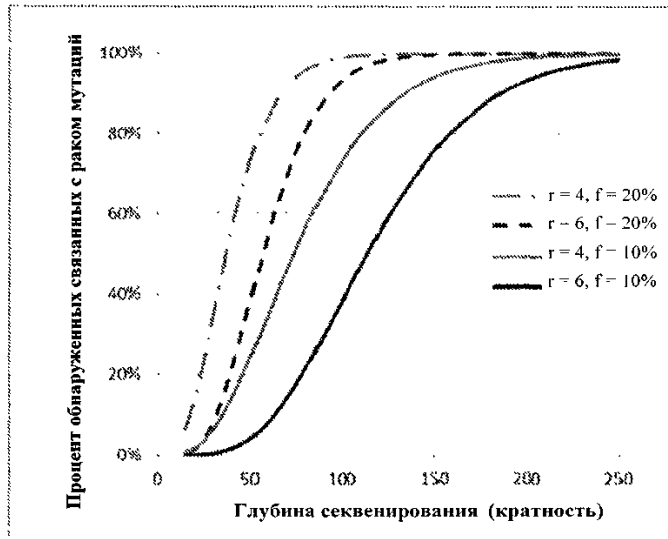
400

Фиг. 4

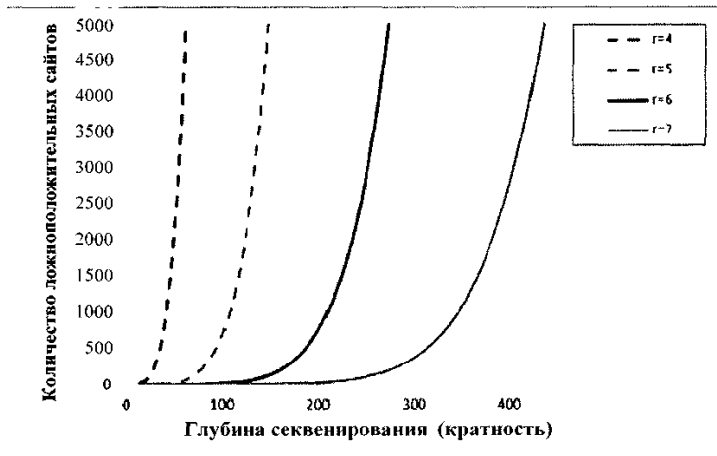
Количество обнаружений	1х		2х		3х		4х		5х		6х	
	Ожидаемое количество ложноположительных локусов для ошибок секвенирования	Ожидаемое количество обнаруженных мутаций	Ожидаемое количество ложноположительных локусов для ошибок секвенирования	Ожидаемое количество обнаруженных мутаций	Ожидаемое количество ложноположительных локусов для ошибок секвенирования	Ожидаемое количество обнаруженных мутаций	Ожидаемое количество ложноположительных локусов для ошибок секвенирования	Ожидаемое количество обнаруженных мутаций	Ожидаемое количество ложноположительных локусов для ошибок секвенирования	Ожидаемое количество обнаруженных мутаций	Ожидаемое количество ложноположительных локусов для ошибок секвенирования	Ожидаемое количество обнаруженных мутаций
1	2998500	74	1499	1	0	0	0	0	0	0	0	0
10	29850499	664	45000	79	40	6	0	0	0	0	0	0
20	59403980	1180	190000	271	380	43	1	5	0	1	0	0
40	117631683	1896	780000	793	3293	241	10	57	0	11	0	2
60	174706399	2331	1770000	1327	11407	573	54	197	0	56	0	13
80	230650961	2594	3160000	1782	27387	970	176	429	1	158	0	50
100	285487746	2754	4950000	2138	53900	1369	436	727	3	326	0	126
120	339238690	2851	7140000	2403	93613	1730	913	1058	7	554	0	252
140	391925294	2909	9730000	2592	149193	2037	1703	1390	15	824	0	427
160	443568633	2945	12720000	2725	223307	2286	2922	1700	30	1113	0	645
180	494189366	2967	16110000	2817	318620	2479	4700	1973	55	1404	1	891
200	543807741	2980	19900000	2879	437800	2626	7187	2205	94	1679	1	1152

500

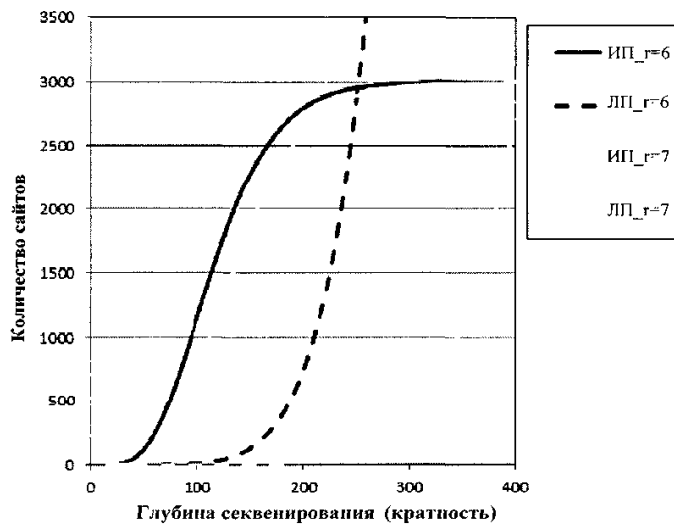
Фиг. 5



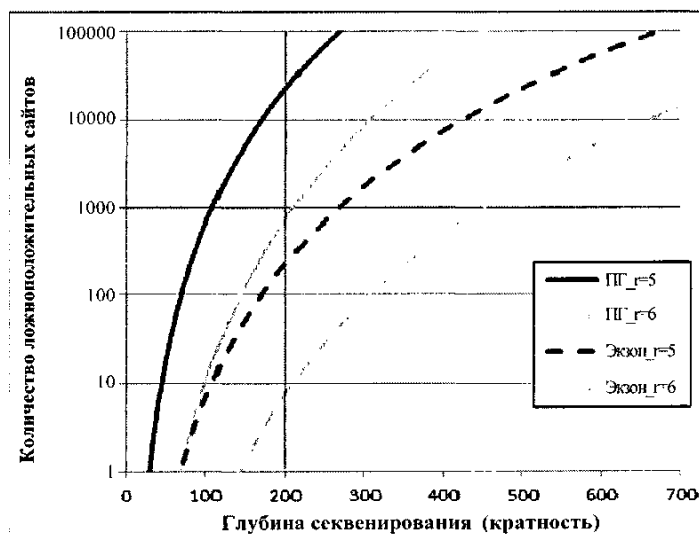
Фиг. 6А



Фиг. 6В



Фиг. 7А



Фиг. 7В

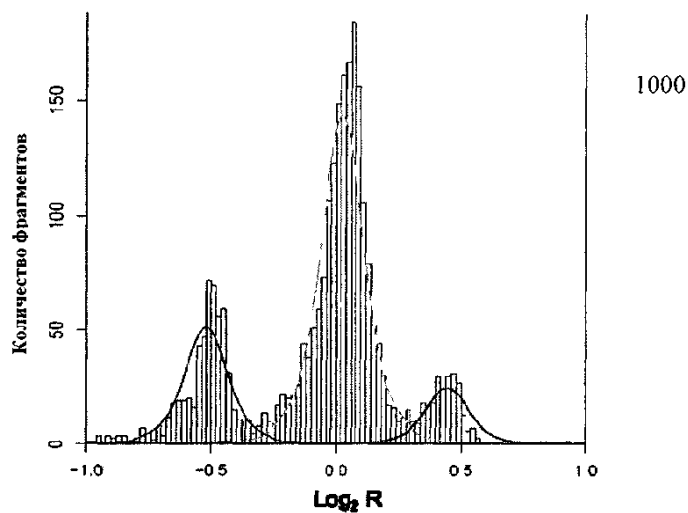
Случай	Количество ОНВ, обнаруженных в опухолевой ткани	Момент времени	Количество связанных с опухолью ОНВ, секвенированных в плазме (% ОНВ, обнаруженных в опухолевой ткани)	Количество прочтений последовательности ДНК в плазме, в которых выявлены ОНВ (p)	Количество прочтений последовательности ДНК в плазме, в которых выявлена последовательность дикого типа (q)	Фракционная концентрация ДНК опухолевого происхождения, выявленная с помощью анализа ОНВ ($\frac{2p}{p+q}$)	Фракционная концентрация ДНК опухолевого происхождения, выявленная с помощью анализа GAAL
ГКК1	2840	До лечения	2569 (94%)	11389	31602	53%	52%
		После лечения	44 (1,5%)	91	46898	0,4%	1,4%
ГКК2	3105	До лечения	1097 (35%)	1490	57865	5,0%	4,3%
		После лечения	72 (2,3%)	206	66692	0,6%	0,9%
ГКК3	3171	До лечения	461 (15%)	525	48886	2,1%	5,6%
		После лечения	31 (1%)	67	58862	0,2%	0,9%
ГКК4	1334	До лечения	201 (15%)	248	18527	2,6%	7,6%
		После лечения	74 (5,5%)	149	22144	1,3%	2,7%

Фиг. 8

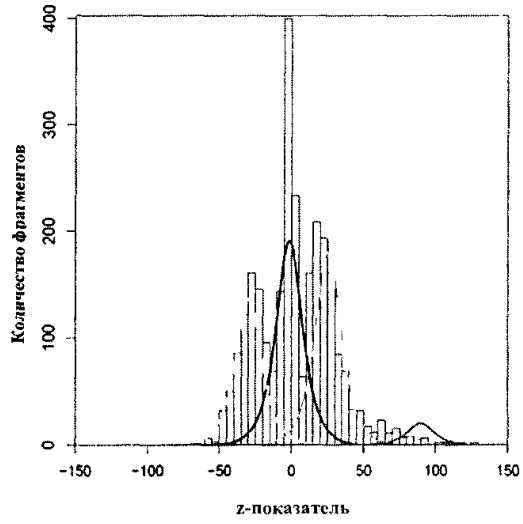
Контроли	ОНВ ГКК1		ОНВ ГКК2		ОНВ ГКК3		ОНВ ГКК4		Все ОНВ		
	Прочтения последовательности ДНК в плазме, в которых выявлены связанные с опухолью ОНВ	Прочтения последовательности ДНК в плазме, в которых выявлена последовательность дикого типа	Прочтения последовательности ДНК в плазме, в которых выявлены связанные с опухолью ОНВ	Прочтения последовательности ДНК в плазме, в которых выявлена последовательность дикого типа	Прочтения последовательности ДНК в плазме, в которых выявлены связанные с опухолью ОНВ	Прочтения последовательности ДНК в плазме, в которых выявлена последовательность дикого типа	Прочтения последовательности ДНК в плазме, в которых выявлены связанные с опухолью ОНВ	Прочтения последовательности ДНК в плазме, в которых выявлена последовательность дикого типа	Прочтения последовательности ДНК в плазме, в которых выявлены связанные с опухолью ОНВ (p)	Прочтения последовательности ДНК в плазме, в которых выявлена последовательность дикого типа (q)	Фракционная концентрация всех связанных с опухолью ОНВ в плазме ($\frac{2p}{p+q}$)
C01	40	31261	66	38172	30	33823	69	14389	205	117645	0,35%
C02	60	34368	112	42660	37	37462	58	16097	267	130587	0,41%
C03	55	37931	100	46297	50	41181	74	17039	279	142448	0,39%
C04	41	40277	87	48441	39	43541	87	17358	254	149617	0,34%
C05	42	31655	71	38787	46	34566	70	14821	229	119829	0,38%
C06	41	34604	97	43999	26	37477	79	16762	243	132842	0,37%
C07	52	38664	118	47200	61	41840	72	17601	303	145305	0,42%
C08	57	38514	131	46464	42	41662	81	17275	311	143915	0,43%
C09	41	34945	115	43760	33	38039	81	16503	270	133247	0,40%
C10	44	35197	89	43622	51	38370	65	16009	249	133198	0,37%
C11	41	36693	95	45108	39	39732	72	16801	247	138334	0,36%
C12	45	39506	104	50085	46	42536	91	18598	286	150725	0,38%
C13	36	33799	71	40309	39	36241	70	14450	216	124799	0,35%
C14	42	38199	104	47832	42	41245	91	17807	279	145083	0,38%
C15	49	36106	90	43129	41	38432	73	16154	253	133821	0,38%
C16	38	33633	66	40023	41	36431	52	14502	197	124589	0,32%
										Среднее	0,38%

900

Фиг. 9

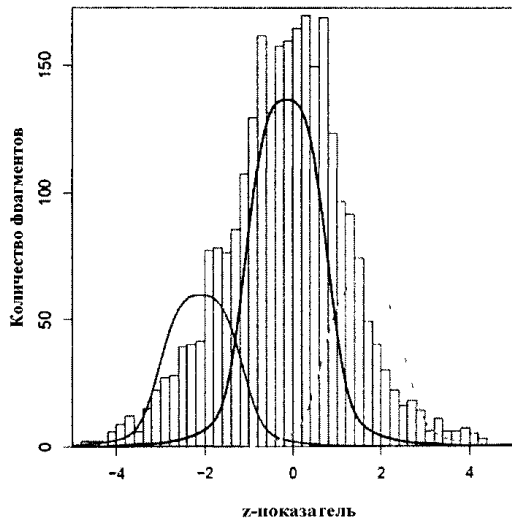


Фиг. 10А



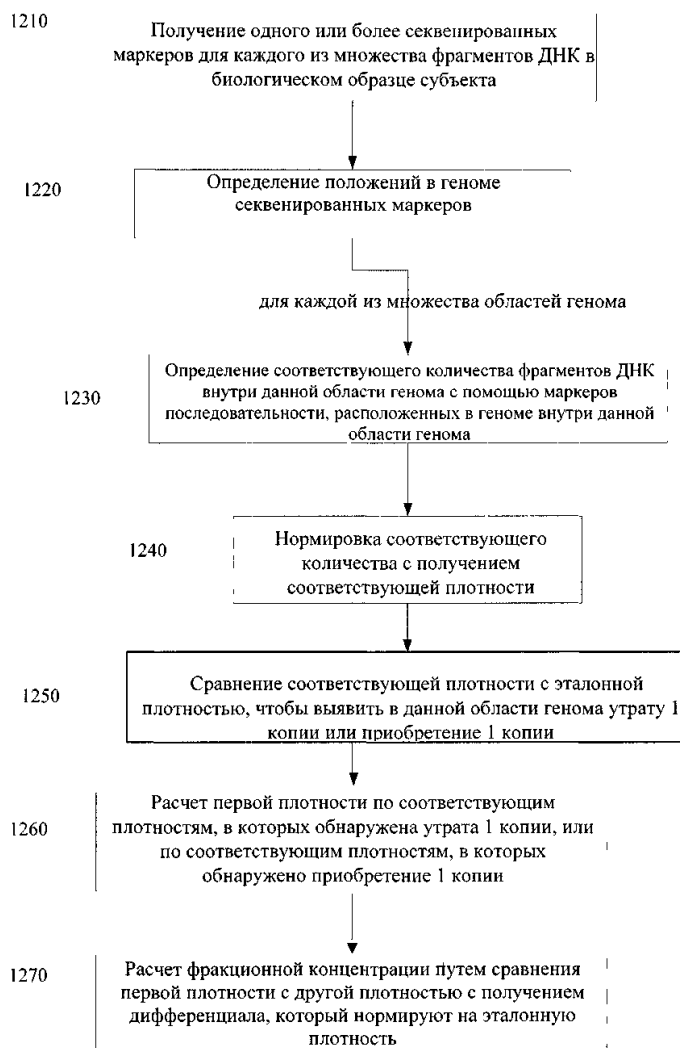
1050

Фиг. 10В



1100

Фиг. 11



Фиг. 12

Количество обнаружений в плазме (N)	Количество потенциальных мутаций, обнаруженных $\times N$ раз в плазме	Количество потенциальных мутаций, обнаруженных $\times \geq N$ раз в плазме	Опухолевые мутации, обнаруженные N раз в плазме	Опухолевые мутации, обнаруженные $\geq N$ раз в плазме	Процент обнаруженных мутаций по сравнению с N=1	Положительное прогнозируемое значение
1	3605085	3623970	61	2064	100%	0,06%
2	13518	18885	35	2003	97%	11%
3	2029	5367	71	1968	95%	37%
4	766	3338	83	1897	92%	57%
5	437	2572	140	1814	88%	71%
6	313	2135	161	1674	81%	78%
7	265	1822	165	1513	73%	83%
8	261	1557	192	1348	65%	87%
9	231	1296	185	1156	56%	89%
10	191	1065	160	971	47%	91%
11	179	874	165	811	39%	93%
12	148	695	137	646	31%	93%
13	122	547	109	509	25%	93%
14	99	425	92	400	19%	94%
15	74	326	71	308	15%	94%
16	54	252	49	237	11%	94%
17	39	198	39	188	9%	95%
18	35	159	33	149	7%	94%
19	24	124	23	116	6%	94%
20	20	100	20	93	5%	93%
21	18	80	17	73	4%	91%
22	11	62	10	56	3%	90%
23	9	51	9	46	2%	90%
24	8	42	5	37	2%	88%
25	6	34	6	32	2%	94%
26	7	28	5	26	1%	93%
27	8	21	8	21	1%	100%
28	1	13	1	13	1%	100%
29	4	12	4	12	1%	100%
30	8	8	8	8	0%	100%

1300

Фиг. 13А

Количество обнаруженных в плазме (N)	Количество потенциальных мутаций, обнаруженных N раз в плазме	Количество потенциальных мутаций, обнаруженных $\geq N$ раз в плазме	Опухолевые мутации, обнаруженные N раз в плазме	Опухолевые мутации, обнаруженные $\geq N$ раз в плазме	Процент обнаруженных мутаций по сравнению с N=1	Положительное прогнозируемое значение
1	3155634	3167144	51	54	100%	<0,01%
2	9835	11510	2	3	6%	0,03%
3	1123	1675	1	1	2%	0,06%
4	314	552	-	-	-	-
5	132	238	-	-	-	-
6	49	106	-	-	-	-
7	22	57	-	-	-	-
8	11	35	-	-	-	-
9	6	24	-	-	-	-
10	5	18	-	-	-	-
11	4	13	-	-	-	-
12	2	9	-	-	-	-
13	2	7	-	-	-	-
14	3	5	-	-	-	-
15	-	2	-	-	-	-
16	-	2	-	-	-	-
17	1	2	-	-	-	-
18	-	1	-	-	-	-
19	1	1	-	-	-	-
20	-	-	-	-	-	-
21	-	-	-	-	-	-
22	-	-	-	-	-	-
23	-	-	-	-	-	-
24	-	-	-	-	-	-
25	-	-	-	-	-	-
26	-	-	-	-	-	-
27	-	-	-	-	-	-
28	-	-	-	-	-	-
29	-	-	-	-	-	-
30	-	-	-	-	-	-

1300

Фиг. 13В

Глубина секвенирования	Плазма до лечения			Плазма после лечения		
	Процент нуклеотидов при определенной глубине секвенирования	Суммарное количество положений нуклеотидов с однонуклеотидными вариациями, обнаруженных в плазме	Количество сайтов с соответствующими изменениями в опухоли	Процент нуклеотидов при определенной глубине секвенирования	Суммарное количество положений нуклеотидов с однонуклеотидными вариациями, обнаруженных в плазме	Количество сайтов с соответствующими изменениями в опухоли
<50	36,3%	9	8	30,5%	0	0
50 – 125	29,9%	15	15	31,2%	0	0
126 – 235	18,6%	21	21	20,8%	0	0
236 – 380	9,4%	7	6	10,9%	0	0
381 – 560	4,2%	2	2	4,8%	0	0
561 – 760	1,6%	3	3	1,7%	0	0
Всего		57	55		0	0

1400

Фиг. 14А

Глубина секвенирования	Плазма до лечения			Плазма после лечения		
	Процент нуклеотидов при определенной глубине секвенирования	Суммарное количество положений нуклеотидов с однонуклеотидными вариациями, обнаруженных в плазме	Количество сайтов с соответствующими изменениями в опухоли	Процент нуклеотидов при определенной глубине секвенирования	Суммарное количество положений нуклеотидов с однонуклеотидными вариациями, обнаруженных в плазме	Количество сайтов с соответствующими изменениями в опухоли
<50	30,0%	0	0	27,1%	0	0
50 – 125	33,6%	2	2	32,5%	0	0
126 – 235	21,6%	6	6	22,2%	0	0
236 – 380	10,2%	5	5	11,5%	0	0
381 – 560	3,7%	4	4	4,9%	0	0
561 – 760	0,9%	1	1	1,7%	0	0
Всего		18	18		0	0

1450

Фиг. 14В

Глубина секвенирования	Плазма до лечения			Плазма после лечения		
	Процент нуклеотидов при определенной глубине секвенирования	Суммарное количество положений нуклеотидов с однонуклеотидными вариациями, обнаруженных в плазме	Количество сайтов с соответствующими изменениями в опухоли	Процент нуклеотидов при определенной глубине секвенирования	Суммарное количество положений нуклеотидов с однонуклеотидными вариациями, обнаруженных в плазме	Количество сайтов с соответствующими изменениями в опухоли
<50	38,8%	0	0	18,5%	0	0
50 – 125	45,7%	0	0	28,6%	0	0
126 – 235	15,2%	0	0	25,1%	0	0
236 – 380	0,2%	0	0	16,2%	0	0
381 – 560	<0,1%	0	0	8,3%	0	0
561 – 760	<0,1%	0	0	3,4%	0	0
Всего		0	0		0	0

1500

Фиг. 15А

Глубина секвенирования	Плазма до лечения			Плазма после лечения		
	Процент нуклеотидов при определенной глубине секвенирования	Суммарное количество положений нуклеотидов с однонуклеотидными вариациями, обнаруженных в плазме	Количество сайтов с соответствующими изменениями в опухоли	Процент нуклеотидов при определенной глубине секвенирования	Суммарное количество положений нуклеотидов с однонуклеотидными вариациями, обнаруженных в плазме	Количество сайтов с соответствующими изменениями в опухоли
<50	26,7%	0	0	24,4%	0	0
50 – 125	35,9%	1	1	31,2%	1	1
126 – 235	24,7%	2	2	23,1%	0	0
236 – 380	10,4%	0	0	13,1%	0	0
381 – 560	2,2%	0	0	6,0%	0	0
561 – 760	0,1%	0	0	2,2%	0	0
Всего		3	3		1	1

1550

Фиг. 15В

Глубина секвенирования	Плазма до лечения			Плазма после лечения		
	Процент нуклеотидов при определенной глубине секвенирования	Суммарное количество положений нуклеотидов с однонуклеотидным и вариациями, обнаруженных в плазме	Количество сайтов с соответствующими изменениями в опухоли	Процент нуклеотидов при определенной глубине секвенирования	Суммарное количество положений нуклеотидов с однонуклеотидным и вариациями, обнаруженных в плазме	Количество сайтов с соответствующими изменениями в опухоли
<50	31,1%	7	7	30,8%	0	0
50 – 125	30,3%	14	14	33,2%	1	0
126 – 235	20,5%	19	19	21,6%	0	0
236 – 380	11,3%	13	13	10,3%	0	0
381 – 560	5,2%	7	4	3,5%	0	0
561 – 760	1,8%	4	2	0,7%	0	0
Всего		64	59		1	0

1600

Фиг. 16

Глубина секвенирования (кратность)	Чувствительность, когда мутанта обнаруживают в результатах секвенирования опухолевой ткани по меньшей мере n раз, где n=					
	4	5	6	7	8	9
	20	73%	56%	38%	24%	13%
21	77%	60%	43%	28%	16%	9%
22	80%	64%	47%	31%	19%	11%
23	83%	68%	51%	35%	22%	13%
24	85%	71%	55%	39%	26%	15%
25	87%	75%	59%	43%	29%	18%
26	89%	78%	63%	47%	33%	21%
27	90%	80%	67%	51%	36%	24%
28	92%	83%	70%	55%	40%	27%
29	93%	85%	73%	59%	44%	30%
30	94%	87%	76%	62%	48%	34%
31	95%	89%	78%	66%	51%	37%
32	96%	90%	81%	69%	55%	41%
33	96%	91%	83%	72%	58%	44%
34	97%	93%	85%	74%	61%	48%
35	97%	94%	87%	77%	65%	51%
36	98%	95%	88%	79%	68%	54%
37	98%	95%	90%	82%	70%	58%
38	99%	96%	91%	84%	73%	61%
39	99%	97%	92%	85%	76%	64%
40	99%	97%	93%	87%	78%	67%
41	99%	98%	94%	88%	80%	69%
42	99%	98%	95%	90%	82%	72%
43	99%	98%	96%	91%	84%	75%
44	100%	98%	96%	92%	86%	77%
45	100%	99%	97%	93%	87%	79%
46	100%	99%	97%	94%	89%	81%
47	100%	99%	98%	95%	90%	83%
48	100%	99%	98%	95%	91%	84%
49	100%	99%	98%	96%	92%	86%
50	100%	99%	99%	97%	93%	88%
51	100%	100%	99%	97%	94%	89%
52	100%	100%	99%	97%	95%	90%
53	100%	100%	99%	98%	95%	91%
54	100%	100%	99%	98%	96%	92%
55	100%	100%	99%	98%	96%	93%
56	100%	100%	99%	99%	97%	94%
57	100%	100%	100%	99%	97%	95%
58	100%	100%	100%	99%	98%	95%
59	100%	100%	100%	99%	98%	96%
60	100%	100%	100%	99%	98%	96%
61	100%	100%	100%	99%	98%	97%
62	100%	100%	100%	99%	99%	97%
63	100%	100%	100%	100%	99%	97%
64	100%	100%	100%	100%	99%	98%
65	100%	100%	100%	100%	99%	98%
66	100%	100%	100%	100%	99%	98%
67	100%	100%	100%	100%	99%	99%
68	100%	100%	100%	100%	99%	99%
69	100%	100%	100%	100%	100%	99%
70	100%	100%	100%	100%	100%	99%

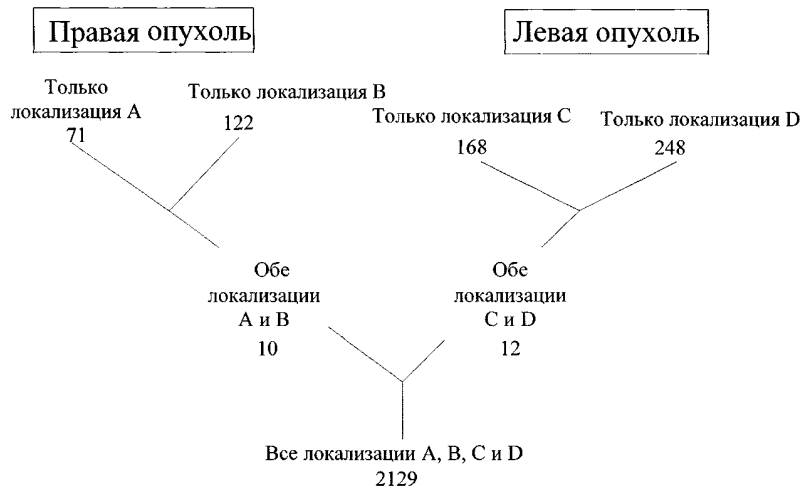
1700

Фиг. 17

Глубина секвенирования (кратность)	Количество ложноположительных локусов во всем геноме, когда мутанта обнаруживают в результатах секвенирования опухоли по меньшей мере n раз, где n=					
	4	5	6	7	8	9
20	1	0	0	0	0	0
21	1	0	0	0	0	0
22	1	0	0	0	0	0
23	1	0	0	0	0	0
24	1	0	0	0	0	0
25	1	0	0	0	0	0
26	2	0	0	0	0	0
27	2	0	0	0	0	0
28	2	0	0	0	0	0
29	3	0	0	0	0	0
30	3	0	0	0	0	0
31	3	0	0	0	0	0
32	4	0	0	0	0	0
33	5	0	0	0	0	0
34	5	0	0	0	0	0
35	6	0	0	0	0	0
36	7	0	0	0	0	0
37	7	0	0	0	0	0
38	8	0	0	0	0	0
39	9	0	0	0	0	0
40	10	0	0	0	0	0
41	11	0	0	0	0	0
42	12	0	0	0	0	0
43	14	0	0	0	0	0
44	15	0	0	0	0	0
45	17	0	0	0	0	0
46	18	0	0	0	0	0
47	20	0	0	0	0	0
48	22	0	0	0	0	0
49	24	0	0	0	0	0
50	26	0	0	0	0	0
51	28	0	0	0	0	0
52	30	0	0	0	0	0
53	33	0	0	0	0	0
54	35	0	0	0	0	0
55	38	0	0	0	0	0
56	41	0	0	0	0	0
57	44	0	0	0	0	0
58	47	0	0	0	0	0
59	51	0	0	0	0	0
60	54	0	0	0	0	0
61	58	0	0	0	0	0
62	62	0	0	0	0	0
63	66	0	0	0	0	0
64	71	0	0	0	0	0
65	75	0	0	0	0	0
66	80	0	0	0	0	0
67	85	0	0	0	0	0
68	90	0	0	0	0	0
69	96	0	0	0	0	0
70	102	0	0	0	0	0

1800

Фиг. 18



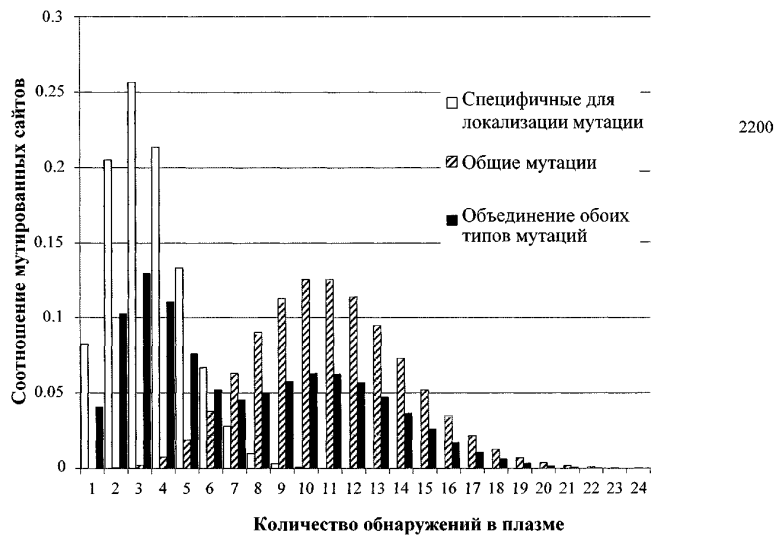
Фиг. 19

Категория	Количество локусов с мутацией	Плазма до лечения			Плазма после лечения		
		Количество фрагментов ДНК, несущих аллели дикого типа	Количество фрагментов в ДНК, несущих мутантные аллели	Фракционная концентрация ДНК опухолевого происхождения	Количество фрагментов ДНК, несущих аллели дикого типа	Количество фрагментов ДНК, несущих мутантные аллели	Фракционная концентрация ДНК опухолевого происхождения
A	71	3321	37	2,20%	2149	5	0,46%
B	122	5633	8	0,28%	3516	1	0,06%
C	168	8507	51	1,19%	5438	2	0,07%
D	248	10880	26	0,48%	7060	2	0,06%
AB	10	423	21	9,46%	297	0	0,00%
CD	12	569	3	1,05%	333	0	0,00%
ABCD	2129	70940	21344	46,26%	61417	54	0,18%

2000
Фиг. 20



Фиг. 21

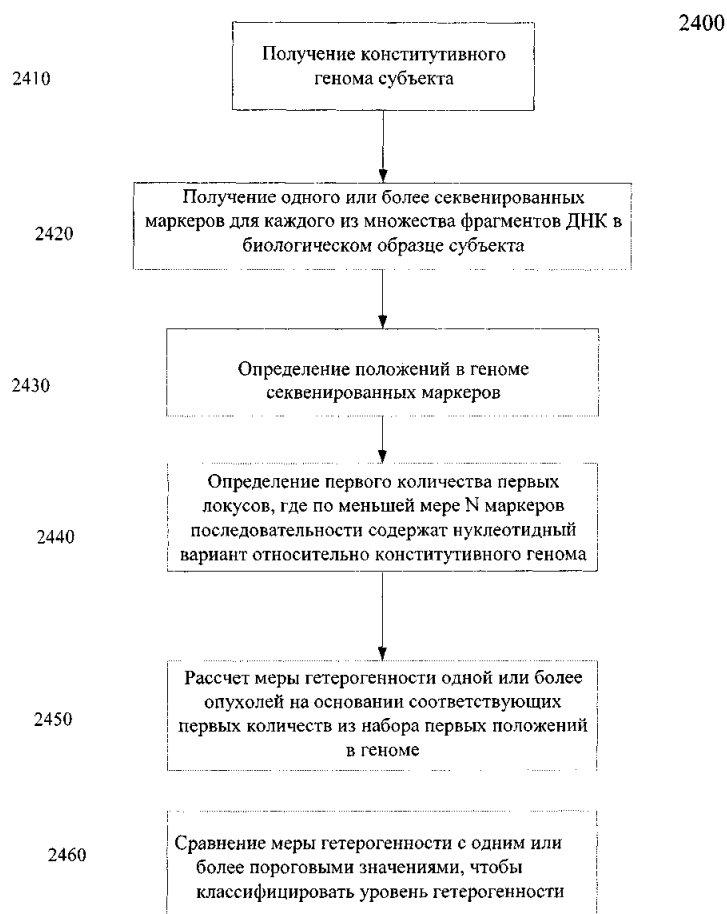


Фиг. 22

Кат.	Количество локусов	Здоровый контрольный субъект															
		TC1	TC2	TC3	TC4	TC5	TC6	TC7	TC8	TC9	TC10	TC11	TC12	TC13	TC14	TC15	TC16
		Количество (фракционная концентрация) мутантных аллелей в плазме															
A	71	0	1 (0,17%)	1 (0,18%)	0	1 (0,17%)	1 (0,18%)	1 (0,07%)	1 (0,20%)	0	6 (0,17%)	0	0	1 (0,17%)	2 (0,38%)	0	2 (0,38%)
B	122	0	0	0	0	0	0	1 (0,04%)	0	1 (0,11%)	3 (0,05%)	1 (0,10%)	1 (0,12%)	0	0	0	1 (0,11%)
C	168	0	0	0	0	0	0	2 (0,06%)	0	2 (0,16%)	0	0	0	1 (0,07%)	0	0	0
D	248	0	0	1 (0,06%)	0	0	0	0	0	0	1 (0,01%)	0	0	1 (0,05%)	0	0	0
AB	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CD	12	0	0	0	0	0	0	0	0	0	1 (0,16%)	0	0	1 (0,92%)	0	0	0
ABCD	2129	11 (0,07%)	3 (0,02%)	2 (0,01%)	8 (0,06%)	5 (0,03%)	5 (0,03%)	10 (0,02%)	4 (0,03%)	4 (0,03%)	19 (0,02%)	3 (0,02%)	0	1 (0,01%)	5 (0,03%)	2 (0,01%)	4 (0,03%)

2300

Фиг. 23



Фиг. 24



Фиг. 25

