

(19)



**Евразийское  
патентное  
ведомство**

(11) **038374**

(13) **B1**

**(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ**

**(45)** Дата публикации и выдачи патента  
**2021.08.17**

**(21)** Номер заявки  
**201300649**

**(22)** Дата подачи заявки  
**2011.11.30**

**(51)** Int. Cl. **C12Q 1/68** (2006.01)  
**G06F 19/18** (2011.01)  
**G06F 19/20** (2011.01)

---

**(54) СПОСОБ АНАЛИЗА БИОЛОГИЧЕСКОГО ОБРАЗЦА ОРГАНИЗМА ДЛЯ ВЫЯВЛЕНИЯ ХРОМОСОМНЫХ ДЕЛЕЦИЙ ИЛИ АМПЛИФИКАЦИЙ, АССОЦИИРОВАННЫХ С РАКОМ**

---

**(31)** 61/418,391; 61/529,877

**(32)** 2010.11.30; 2011.08.31

**(33)** US

**(43)** 2013.10.30

**(86)** PCT/AU2011/001562

**(87)** WO 2012/071621 2012.06.07

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

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

**(74)** Представитель:  
**Воробьева Е.В., Фелицына С.Б. (RU)**

**(56)** TABACK, B. et al., "Prognostic Significance of Circulating Microsatellite Markers in the Plasma of Melanoma Patients", Cancer Research, 2001, vol. 61, pages 5723-5726, See Abstract; pg 5724, paragraphs 2-3

MULLER, I. et al., "Identification of Loss of Heterozygosity on Circulating Free DNA in Peripheral Blood of Prostate Cancer Patients: Potential and Technical Improvements", Clinical Chemistry, 2008, vol. 54, no. 4, pages 688-696, See Abstract; page 689, right hand column, paragraph 6 - page 690, left hand column, paragraph 2

US-A1-20090029377

---

**(57)** Предложены способ анализа биологического образца организма для выявления хромосомных делеций или амплификации, ассоциированных с раком, машиночитаемый носитель, включающий компьютерную программу для осуществления этого способа, и компьютерная система, включающая процессор и указанный машиночитаемый носитель. В качестве биологического образца для анализа используют биологический образец, включающий бесклеточную смесь множества молекул нуклеиновых кислот, происходящих из хромосом здоровых клеток и из хромосом клеток, возможно, ассоциированных с раком. При анализе используют множество локусов области хромосомы, в которых аллели первого и второго гаплотипов гетерозиготны, и определяют группы молекул нуклеиновых кислот образца, соответствующие каждому из указанных гаплотипов. Сравнение разницы или отношения значений, рассчитанных для каждой из групп молекул нуклеиновых кислот, с пороговым значением позволяет сделать вывод о том, проявляет ли область хромосомы делецию или амплификацию в клетках, ассоциированных с раком. Выявленный дисбаланс можно применять для диагностики, определения стадии, прогнозирования или отслеживания прогрессирования степени рака.

---

**038374 B1**

**038374 B1**

Настоящее изобретение испрашивает приоритет и является непредварительной заявкой предварительной заявки на патент США № 61/418391, озаглавленной "Detection of genetic aberrations associated with cancer" ("Выявление генетических аббераций, ассоциированных с раком"), поданной 30 ноября 2010 г., и предварительной заявки на патент США № 61/529877, озаглавленной "Detection of genetic or molecular aberrations associated with cancer" ("Выявление генетических или молекулярных аббераций, ассоциированных с раком"), поданной 31 августа 2011 г., полное содержание которых включено в данное описание посредством ссылки для всех целей.

Настоящее изобретение также является родственной заявкам тех же заявителей - заявке на патент США № 12/940992 (публикация 2011/0276277), озаглавленной "Size-Based Genomic Analysis" ("Геномный анализ на основании размера"), на имя Ло с соавт. (досье поверенного № 80015-794101/006610 US), поданной 5 ноября 2010 г., и заявке на патент США № 12/940993 (публикация 2011/0105353), озаглавленной "Fetal Genomic Analysis From A Maternal Biological Sample" ("Анализ генома плода по биологическим образцам, взятым у матери"), на имя Ло с соавт. (досье поверенного № 80015-794103/006710 US), поданной 5 ноября 2010 г., содержание которых включено в данное описание посредством ссылки во всей полноте.

### **Область техники**

Рак является распространенным заболеванием, которое поражает многих людей. Часто рак не удается выявить до тех пор, пока не проявятся тяжелые симптомы. Для распространенных типов рака существуют способы скрининга для идентификации пациентов, у которых может быть рак. Но такие способы часто ненадежны или требуют воздействия облучения на пациента. Для многих других типов рака не существует способов эффективного скрининга.

Была выявлена потеря гетерозиготности (ПГ) в конкретном локусе ДНК, циркулирующей в крови пациентов, страдающих раком легких, а также головы и шеи (Chen XQ, et al. Nat Med 1996; 2: 1033-5; Nawroz H., et al. Nat Med 1996; 2: 1035-7). Однако такие методики сопряжены с затруднениями, связанными с относительно малым количеством ПГ, поддающимся обнаружению при исследовании конкретного локуса. Даже при применении цифровой ПЦР (полимеразной цепной реакции) данные способы по-прежнему не способны обнаруживать малые количества ПГ. Кроме того, такие методики все еще ограничены исследованием конкретного локуса, о котором известно, что он возникает при конкретном типе рака. Таким образом, скрининг для обнаружения раковых заболеваний в целом до сих пор был невозможен или неэффективен.

Помимо скрининга, для определения наличия рака современные способы также не позволяют давать прогноз состояния пациента, страдающего раковым заболеванием, и отслеживать действенность лечения (например, восстановление после операции, или химиотерапии, или иммунотерапии, или направленной терапии). Такие способы часто затратны (например, способы получения изображения), неточны, неэффективны, нечувствительны или могут требовать облучения пациента в случае способов получения изображения.

Соответственно желательно обеспечить новые способы скрининга, прогнозирования и мониторинга пациентов на предмет раковых заболеваний.

### **Сущность изобретения**

Согласно вариантам реализации предложены системы, устройства и способы определения генетических аббераций, ассоциированных с раком. Биологические образцы, включая бесклеточные фрагменты ДНК, анализируют с целью выявления дисбаланса между областями хромосом, например вследствие делеций и/или амплификации в опухоли. Применение области хромосомы с множеством локусов позволяет достичь более высокой эффективности и/или точности. Затем такой дисбаланс можно использовать для диагностики или скрининга рака у пациента, а также прогнозировать состояние пациента, страдающего раковым заболеванием. Могут быть использованы степень дисбаланса, а также число областей, обнаруживающих дисбаланс. Кроме того, состояние пациента можно оценивать во времени, чтобы отслеживать степень тяжести заболевания в каждой одной или более областях хромосом и число областей хромосом, что делает возможным скрининг и прогнозирование, а также отслеживание прогресса (например, после лечения).

Согласно одному варианту реализации предложен способ анализа биологического образца, отобранного из организма, с целью определения хромосомных делеций или амплификаций при раке. Биологический образец включает молекулы нуклеиновых кислот, полученные из здоровых клеток и, возможно, из клеток, ассоциированных с раком. По меньшей мере, некоторые из молекул нуклеиновых кислот в указанном биологическом образце находятся в бесклеточной форме. Первый и второй гаплотипы определяют для здоровых клеток организма в первой области хромосомы. Первая область хромосомы включает первое множество гетерозиготных локусов. Каждое из множеств молекул нуклеиновых кислот в образце имеет определенную локализацию в эталонном геноме данного организма и содержит соответствующий определяемый аллель. Локализацию и определяемые аллели применяют для определения молекул нуклеиновых кислот первой группы из первого гаплотипа и второй группы - из второго гаплотипа. Компьютерная система рассчитывает первое значение для первой группы и второе значение для второй группы. Каждое значение определяет свойство соответствующей группы молекул нуклеиновых кислот

(например, средний размер или количество молекул в группе). Первое значение сравнивают со вторым значением и определяют классификацию или определяют, обнаруживает ли первая область хромосомы делецию или амплификацию в каких-либо клетках, ассоциированных с раком.

Согласно другому варианту реализации предложен способ анализа биологического образца, отобранного из организма. Биологический образец включает молекулы нуклеиновых кислот, полученные из здоровых клеток и, возможно, из клеток, ассоциированных с раком. По меньшей мере, некоторые из молекул нуклеиновых кислот в образце находятся в бесклеточной форме. Выявляют множество неперекрывающихся областей хромосомы. Каждая область хромосомы включает множество локусов. Каждая из множества молекул нуклеиновых кислот в образце имеет локализацию в эталонном геноме данного организма. Для каждой области хромосомы выявляют соответствующую группу молекул нуклеиновых кислот, которые происходят из указанной области хромосомы на основании выявленной локализации. Каждая соответствующая группа включает по меньшей мере одну молекулу нуклеиновой кислоты, локализованную в каждом из множества локусов указанной области хромосомы. Компьютерная система рассчитывает соответствующие значения для соответствующей группы, причем соответствующее значение определяет свойство молекул нуклеиновых кислот соответствующей группы. Соответствующее значение сравнивают с эталонным значением и определяют классификацию или определяют, обнаруживает ли указанная область хромосомы делецию или амплификацию. Затем определяют количество областей хромосомы, классифицированных как обнаруживающие делецию или амплификацию.

Согласно еще одному варианту реализации предложен способ определения прогресса хромосомных aberrаций в организме с применением биологических образцов, включающих молекулы нуклеиновых кислот, полученные из здоровых клеток и, возможно, из клеток, ассоциированных с раком. По меньшей мере, некоторые из молекул нуклеиновых кислот в образце находятся в бесклеточной форме. Для эталонного генома указанного организма выявляют одну или более неперекрывающихся областей хромосомы. Каждая область хромосомы включает множество локусов. Анализируют образцы, отобранные из указанного организма в разные моменты времени, и определяют прогресс. Для определенного образца каждая из множества молекул нуклеиновых кислот в образце обладает локализацией в эталонном геноме данного организма. Для каждой области хромосомы выявляют соответствующую группу молекул нуклеиновых кислот, которые происходят из указанной области хромосомы, на основании выявленной локализации. Каждая соответствующая группа включает по меньшей мере одну молекулу нуклеиновой кислоты, локализованную в каждом из множества локусов указанной области хромосомы. Компьютерная система рассчитывает соответствующие значения для соответствующей группы молекул нуклеиновых кислот. Соответствующее значение определяет свойство молекул нуклеиновых кислот соответствующей группы. Соответствующее значение сравнивают с эталонным значением и определяют классификацию или определяют, обнаруживает ли указанная область хромосомы делецию или амплификацию. Затем классификацию каждой из областей хромосомы в указанном множестве моментов времени применяют для определения прогресса хромосомных aberrаций у указанного организма.

Другие варианты реализации настоящего изобретения относятся к системам, потребительским электронным устройствам и носителям, считываемым на компьютере, ассоциированным со способами, описанными в настоящем изобретении.

Лучшего понимания характера и преимуществ настоящего изобретения можно достичь, обратившись к приведенному далее подробному описанию и прилагаемым рисункам.

#### **Краткое описание чертежей**

Фиг. 1 иллюстрирует область хромосомы раковой клетки, демонстрирующей aberrацию в форме делеции.

Фиг. 2 иллюстрирует область хромосомы раковой клетки, демонстрирующей aberrацию в форме амплификации.

На фиг. 3 показана таблица, иллюстрирующая разные типы рака и ассоциированные с ними области, а также соответствующие им aberrации.

Фиг. 4 иллюстрирует области хромосомы в раковой клетке, которая не обнаруживает aberrации, а также измерения, сделанные в плазме крови согласно вариантам реализации настоящего изобретения.

Фиг. 5 иллюстрирует делецию 510 области хромосомы в раковой клетке, а также измерения, сделанные в плазме для определения делетированной области согласно вариантам реализации настоящего изобретения.

Фиг. 6 иллюстрирует амплификацию 610 области хромосомы в раковой клетке, а также измерения, сделанные в плазме крови для определения амплифицированной области, согласно вариантам реализации настоящего изобретения.

На фиг. 7 показан анализ RHDO (относительной дозировки гаплотипа) для ДНК в плазме пациента, страдающего печеночно-клеточным раком (НСС), для сегмента, локализованного в хромосоме 1p, который продемонстрировал моноаллельную амплификацию в ткани опухоли согласно вариантам реализации настоящего изобретения.

На фиг. 8 показано изменение в распределении по размеру фрагментов для двух гаплотипов области хромосомы в тех случаях, когда имеет место опухоль, содержащая делецию согласно вариантам ре-

лизации настоящего изобретения.

На фиг. 9 показано изменение в распределении по размеру фрагментов для двух гаплотипов области хромосомы в тех случаях, когда имеет место опухоль, содержащая амплификацию согласно вариантам реализации настоящего изобретения.

На фиг. 10 показана блок-схема способа анализа гаплотипов биологической пробы определенного организма с целью определения, обнаруживает ли определенная область хромосомы делецию или амплификацию согласно вариантам реализации настоящего изобретения.

На фиг. 11 показана область 1110 с подобластью 1130, которая подвергается делеции в раковых клетках, а также измерения, сделанные в плазме для определения делетированной области согласно вариантам реализации настоящего изобретения.

На фиг. 12 показано, как локализацию aberrаций можно картировать при помощи анализа RHDO согласно вариантам реализации настоящего изобретения.

На фиг. 13 показана классификация на основе RHDO, начатая с другого направления, согласно вариантам реализации настоящего изобретения.

На фиг. 14 показана блок-схема способа 1400 анализа биологического образца определенного организма с применением множества областей хромосом согласно вариантам реализации настоящего изобретения.

На фиг. 15 показана таблица 1500, иллюстрирующая глубину, нужную для разного количества сегментов и относительной концентрации фрагментов опухолевого происхождения, согласно вариантам реализации настоящего изобретения. На фиг. 15 приведена оценка числа молекул, которое следует анализировать при разном проценте относительной концентрации ДНК опухолевого происхождения в пробе.

На фиг. 16 показан принцип измерения относительной концентрации ДНК опухолевого происхождения в плазме на основании анализа относительной дозы гаплотипов (RHDO) согласно вариантам реализации настоящего изобретения. Нар I и Нар II представляют собой два гаплотипа в неопухолевых тканях согласно вариантам реализации настоящего изобретения.

На фиг. 17 показана блок-схема, иллюстрирующая способ определения прогресса хромосомных aberrаций в организме с применением биологических образцов, включающих молекулы нуклеиновых кислот согласно вариантам реализации настоящего изобретения.

На фиг. 18А показаны кривые последовательного критерия отношения вероятностей (SPRT), полученные при анализе RHDO для сегмента в q-плече 4 хромосомы у пациента с раком. Точками показано отношение накопленной встречаемости после соответствующих гетерозиготных локусов. На фиг. 18В показаны кривые SPRT, полученные при анализе RHDO для сегмента в q-плече 4 хромосомы у пациента после лечения.

На фиг. 19 показаны частые хромосомные aberrации, обнаруживаемые при печеночно-клеточном раке.

На фиг. 20А показаны результаты нормированного отношения количества маркерных последовательностей для пациентов с печеночно-клеточным раком и для здоровых пациентов, полученные путем направленного анализа. На фиг. 20В показаны результаты анализа размеров после направленного обогащения и массивно-параллельного секвенирования для 3 пациентов, страдающих печеночно-клеточным раком, и 4 здоровых субъектов из контрольной группы.

На фиг. 21 показаны круговые диаграммы для пациентов, страдающих печеночно-клеточным раком, на которых изображены данные о подсчете секвенированных маркерных последовательностей в ДНК плазмы крови согласно вариантам реализации настоящего изобретения.

На фиг. 22 показан анализ на основе подсчета секвенированных маркерных последовательностей в образце плазмы хронического носителя вируса гепатита В (HBV), не содержащем печеночно-клеточного рака согласно вариантам реализации настоящего изобретения.

На фиг. 23 показан анализ на основе подсчета секвенированных маркерных последовательностей в образце плазмы пациента, страдающего раком носоглотки (NPC) 3 стадии, согласно вариантам реализации настоящего изобретения.

На фиг. 24 показан анализ на основе подсчета секвенированных маркерных последовательностей в образце плазмы пациента NPC 4 стадии согласно вариантам реализации настоящего изобретения.

На фиг. 25 показан график зависимости накопленной частоты ДНК в плазме от размера для области, проявляющей потерю гетерозиготности (ПГ) в ткани опухоли согласно вариантам реализации настоящего изобретения.

На фиг. 26 показан график зависимости  $\Delta Q$  от размера секвенированной ДНК плазмы для области с ПГ. Согласно вариантам реализации настоящего изобретения  $\Delta Q$  достигает 0,2 при размере 130 по (пар оснований).

На фиг. 27 показан график зависимости накопленной частоты ДНК в плазме от размера для области с дубликацией хромосомы в ткани опухоли согласно вариантам реализации настоящего изобретения.

На фиг. 28 показан график зависимости  $\Delta Q$  от размера секвенированной ДНК плазмы для амплифицированной области согласно вариантам реализации настоящего изобретения.

На фиг. 29 показана блок-схема примерной компьютерной системы 900, применяемой с системой и способами согласно вариантам реализации настоящего изобретения.

#### Определения

Термин "биологический образец", используемый в настоящем описании, относится к любому образцу, взятому у субъекта (например, человека, человека, страдающего раковым заболеванием, человека, у которого предполагается наличие рака, или у других организмов) и содержащему одну или несколько молекул нуклеиновой(ых) кислоты (кислот), представляющих интерес.

Термин "нуклеиновая кислота" или "полинуклеотид" относится к дезоксирибонуклеиновой кислоте (ДНК) или рибонуклеиновой кислоте (РНК) и их полимеру либо в одно-, либо в двухцепочечной форме. Если конкретно не указано иное, данный термин охватывает нуклеиновые кислоты, содержащие известные аналоги природных нуклеотидов, которые обладают свойствами связывания, схожими с эталонной нуклеиновой кислотой, и метаболизируются подобно нуклеотидам, встречающимся в природе. Если не указано иное, определенная последовательность нуклеиновой кислоты также в неявном виде охватывает ее консервативно модифицированные варианты (например, замещения вырожденного кодона), аллели, ортологи, однонуклеотидные полиморфизмы (SNP), варианты по числу копий и комплементарные последовательности, а также точно указанную последовательность. Более конкретно замещений вырожденного кодона можно достичь, генерируя последовательности, в которых третья позиция одного или нескольких выбранных (или всех) кодонов замещена смешанными остатками оснований и/или дезоксинозинов (Batzer et al, *Nucleic Acid Res.*, 19: 5081 (1991); Ohtsuka et al, *J. Biol. Chem.*, 260: 2605-2608 (1985); и Rossolini et al., *Mol. Cell. Probes*, 8: 91-98 (1994)). Термин "нуклеиновая кислота" включает, но не ограничивается указанными, термины "ген, кДНК, мРНК, малая некодирующая РНК, микроРНК (миРНК), взаимодействующая с Рiwi РНК и короткая шпилечная РНК (shRNA), кодируемая геном или локусом".

Термин "ген" обозначает сегмент ДНК, участвующий в продукции полипептидной цепи или транскрибируемого продукта РНК. Он может включать участки, предшествующие и следующие за кодирующей областью (лидерные и концевые), а также промежуточные последовательности (интроны) между отдельными кодирующими сегментами (экзонами).

Термин "клинически релевантная нуклеотидная последовательность" или "клинически релевантная область хромосомы", используемый в настоящем описании, может относиться к полинуклеотидной последовательности, соответствующей сегменту более крупной геномной последовательности, возможный дисбаланс которой исследуют, или еще более крупной геномной последовательности. Примеры включают сегменты генома, которые подверглись или возможно подверглись делеции или увеличению числа копий (включая простую дупликацию), или более крупную область, которая включает подобласть указанного сегмента. Согласно некоторым вариантам реализации несколько клинически релевантных нуклеотидных последовательностей или, эквивалентно, несколько маркеров клинически релевантной нуклеотидной последовательности можно применять для получения данных для выявления дисбаланса в определенной области. Например, данные о пяти не следующих подряд друг за другом последовательностях хромосомы можно применять в совокупности для определения возможного дисбаланса, причем эффективно уменьшается потребность в объеме образца до 1/5.

Термин "эталонная нуклеотидная последовательность" или "эталонная область хромосомы", используемый в настоящем описании, относится к нуклеотидной последовательности, количественный профиль или распределение по размерам которой применяют для сравнения с исследуемой областью. Примеры эталонной нуклеотидной последовательности включают область хромосомы, которая не содержит делеции или амплификации, весь геном (например, посредством нормировки на общее количество секвенированных маркерных последовательностей), область из одного или более образцов, полученных из здоровых клеток (которая может быть той же исследуемой областью образца), или конкретный гаплотип области хромосомы. Такие эталонные нуклеотидные последовательности могут либо существовать эндогенно в образце, либо их можно добавлять экзогенно во время обработки или анализа образцов. Согласно некоторым вариантам реализации указанная эталонная область хромосомы демонстрирует профиль распределения размеров, который характерен для здорового состояния. Согласно некоторым другим вариантам реализации указанная эталонная область хромосомы демонстрирует количественный профиль, который характерен для здорового состояния.

Термин "на основании (основе)", используемый в настоящем описании, обозначает "на основании (основе), по меньшей мере частично" и относится к одному значению (или результату), применяемому при определении другого значения, например, что имеет место в соотношении на входе и выходе данного способа. Термин "получать", используемый в настоящем описании, также относится к соотношению на входе и выходе данного способа, например, когда получение представляет собой расчет по формуле.

Термин "параметр", используемый в настоящем описании, обозначает численное значение, которое характеризует набор количественных данных и/или соотношение между наборами количественных данных. Например, параметром является отношение (или функция отношения) между первым количеством первой нуклеотидной последовательности и вторым количеством второй нуклеотидной последовательности.

Термин "локус" или его множественное число "локусы", используемые в настоящем описании, относятся к локализации или адресу любой длины по количеству нуклеотидов (или пар нуклеотидов), которые могут иметь вариации в разных геномах.

Термин "дисбаланс последовательностей" или "абберрация", используемый в настоящем описании, обозначает значимое отклонение, определенное по меньшей мере по одному пороговому значению в количестве клинически релевантной области хромосомы в сравнении с эталонным количеством. Дисбаланс последовательностей может включать дисбаланс дозировки хромосом, дисбаланс аллелей, дисбаланс дозировки мутаций, дисбаланс дозировки гаплотипов и другие сходные виды дисбаланса. Например, дисбаланс аллелей может возникать, когда в опухоли один аллель гена делетирован, или один аллель гена амплифицирован, или в ее геноме произошла дифференциальная амплификация двух аллелей, что приводит к дисбалансу в конкретном локусе в указанной пробе. В качестве другого примера, у пациента может быть наследственная мутация в гене-супрессоре опухоли. В дальнейшем у пациента может развиться опухоль, в которой немутированный аллель гена-супрессора опухоли делетирован. Таким образом, в указанной опухоли имеет место дисбаланс дозировки мутации. Когда опухоль выделит свою ДНК в плазму указанного пациента, ДНК опухоли будет смешиваться со структурной ДНК (из нормальной клеток) пациента в указанной плазме. Посредством применения способов, описываемых в настоящем описании, можно выявлять дисбаланс дозировки мутации данной смеси ДНК в плазме.

Термин "гаплотип", используемый в настоящем описании, относится к сочетанию аллелей во многих локусах, которые переносятся вместе на одной и той же хромосоме или области хромосомы. Термин "гаплотип" может относиться только лишь к паре локусов или к области хромосомы или же к целой хромосоме. Термин "аллель" относится к альтернативным последовательностям ДНК в одной и той же физической области генома, которые могут приводить, а могут и не приводить к развитию разных фенотипических признаков. В любом конкретном диплоидном организме с двумя копиями каждой хромосомы (за исключением половых хромосом у мужской особи человека) генотип для каждого гена содержит пару аллелей, присутствующих в данном локусе, которые одинаковы у гомозигот и различаются у гетерозигот. Популяция или вид организмов включает обычно множество аллелей в каждом локусе у разных индивидуумов. Геномный локус, в котором обнаруживается более одного аллеля в популяции, называют полиморфным сайтом. Вариации аллелей в определенном локусе можно измерять по количеству присутствующих аллелей (т.е. степень полиморфизма) или по доле гетерозигот (т.е. процент гетерозиготности) в популяции. Термин "полиморфизм" в настоящем описании относится к любой вариации между особями в геноме человека, независимо от их частоты. Примеры таких вариаций включают без ограничений однонуклеотидные полиморфизмы, простые тандемные повторы, полиморфизмы типа вставки-делеции, мутации (которые могут вызывать заболевание) и вариации числа копий.

Термин "секвенированная маркерная последовательность", используемый в настоящем описании, относится к последовательности, определяемой по всей молекуле нуклеиновой кислоты или ее части, например фрагменту ДНК. Часто секвенируют только один конец указанного фрагмента, например приблизительно 30 по. Затем секвенированную маркерную последовательность можно выравнивать относительно эталонного генома. В качестве альтернативы оба конца фрагмента можно секвенировать, чтобы сгенерировать две секвенированные маркерные последовательности, которые могут обеспечить более высокую точность при выравнивании и также обеспечить определенную длину указанного фрагмента.

Термин "универсальное секвенирование" относится к секвенированию, при котором к концу фрагмента добавляют адаптеры, а к указанным адаптерам присоединяют праймеры для секвенирования. Таким образом, любой фрагмент можно секвенировать при помощи одного и того же праймера, и, таким образом, секвенирование может быть случайным.

Термин "распределение по размеру" относится к любому одному значению или набору значений, которые соответствуют длине, массе, весам или другим показателям размера молекул, соответствующих конкретной группе (например, фрагменты из конкретного гаплотипа или из конкретной области хромосомы). В разных вариантах реализации можно применять разные распределения по размеру. Согласно некоторым вариантам реализации распределение по размеру относится к ранжированию по размерам (например, усредненному размеру, медиане или среднему значению) фрагментов одной хромосомы относительно фрагментов других хромосом. Согласно другим вариантам реализации распределение по размеру может относиться к статистическому значению реальных размеров фрагментов хромосомы. Согласно одному варианту реализации статистическое значение может включать усредненное, среднее значение или медиану размера фрагментов хромосомы. Согласно другому варианту реализации статистическое значение может включать общую длину фрагментов ниже порогового значения, которую можно делить на общую длину всех фрагментов, или, по меньшей мере, фрагментов ниже более высокого порогового значения.

Термин "классификация", используемый в настоящем описании, относится к любому числу(ам) или другой характеристике(ам), которые связаны с конкретным свойством образца. Например, символ "+" (или слово "положительный") может обозначать, что образец классифицируется как содержащий делеции или амплификации. Классификация может быть двоичной (например, положительный или отрицательный) или включать больше уровней классификации (например, шкалу от 1 до 10 или от 0 до 1). Тер-

мин "пороговый" и "порог" относится к заранее определенному числу, применяемому при определенной операции. Например, термин "пороговый размер" может относиться, фрагменты более которого исключаются. Пороговым значением может быть значение, выше или ниже которого применима конкретная классификация. Любой из указанных термином может применять в любом из указанных контекстов.

Термин "степень рака" может относиться либо к факту наличия рака, к стадии рака, к размеру опухоли, к количеству вовлеченных делеций или амплификации (например, две или три) и/или к другим показателям тяжести ракового заболевания. Степенью рака может быть число или другие характеристики. Указанная степень может быть нулевой. Степень рака также включает предраковые состояния, ассоциированные с делециями или амплификациями.

#### Подробное описание

Раковая ткань (опухоль) может содержать aberrации, такие как делеция или амплификация области хромосомы. Опухоль может выделять фрагменты ДНК в жидкие среды организма. Варианты реализации изобретения позволяют выявить опухоль посредством анализа фрагментов ДНК, чтобы выявить aberrации по сравнению с нормальными (ожидаемыми) значениями для ДНК в указанной области хромосомы.

Точный размер делеции или амплификации может варьировать так же, как и локализация. В некоторые моменты времени известно, что конкретная область демонстрирует aberrацию при раке вообще или при раке конкретного типа (что приводит к диагностике конкретного рака). Когда конкретная область неизвестна, можно применять систематический способ анализа всего генома или крупных частей генома, чтобы выявить aberrантные области, которые могут быть рассеяны по всему геному и размер которых (например, количество оснований, делетированных, или число копий которых увеличено) варьирует. Указанную область(и) хромосомы можно проследить во времени, чтобы выявить изменения тяжести aberrации или числа областей, демонстрирующих aberrацию. Такое прослеживание может предоставить жизненно важную информацию для скрининга, прогнозирования состояния и мониторинга опухоли (например, после лечения или для выявления рецидива или прогрессирования опухоли).

Данное описание начинается с примеров хромосомных aberrаций при раке. Затем обсуждают примеры путей выявления хромосомной aberrации путем выявления и анализа бесклеточной ДНК в биологической пробе. Когда установлены способы выявления aberrации в области хромосомы, обсуждаются способы выявления aberrаций во многих областях хромосом, которые применяются для систематического пути скрининга (диагностики) и прогнозирования состояния пациента. В данном описании также охарактеризованы способы прослеживания количественных показателей, получаемых при тестировании на хромосомные aberrации в одной или более областях во времени для обеспечения скрининга, прогнозирования и мониторинга пациентов. Далее обсуждаются примеры.

#### I. Примеры хромосомных aberrаций при раке.

Хромосомные aberrации часто выявляются в раковых клетках. Кроме того, в избранных типах рака можно обнаружить особый характер хромосомных aberrаций. Например, при печеночно-клеточном раке (НСС) часто выявляется увеличение количества ДНК в плечах хромосом 1p, 1q, 7q, 15q, 16p, 17q и 20q и потеря ДНК в 3p, 4q, 9p и 11q.

Предыдущие исследования продемонстрировали, что такие генетические aberrации можно также выявить в циркулирующей ДНК пациентов, страдающих раковыми заболеваниями. Например, в молекулах, циркулирующих в крови пациентов, страдающих раком легких, а также головы и шеи, была выявлена потеря гетерозиготности (ПГ) в конкретном локусе (Chen XQ, et al. *Nat Med* 1996; 2: 1033-5; Nawroz H, et al. *Nat Med* 1996; 2: 1035-7). Генетические изменения, выявляемые в плазме или сыворотке, были идентичны тем, которые обнаруживались в ткани опухоли. Однако, поскольку ДНК, происходящая из опухоли, составляет лишь незначительную долю полной циркулирующей бесклеточной ДНК, дисбаланс аллелей, вызванный ПГ клеток опухоли, обычно мал. Многие исследователи разработали технологию количественной полимеразной цепной реакции (ПЦР) (Vogelstein B, Kinzler KW. *Proc Natl Acad Sci USA*. 1999; 96: 9236-41; Zhou W, et al. *Nat Biotechnol* 2001; 19: 78-81; Zhou W, et al. *Lancet*. 2002; 359: 219-25) для точного количественного определения разных аллелей локуса среди циркулирующих молекул ДНК (Chang HW, et al. *J Natl Cancer Inst*. 2002; 94: 1697-703). Количественная ПЦР обладает существенно более высокой чувствительностью, чем ПЦР в режиме реального времени или другие способы количественной оценки ДНК при выявлении малого дисбаланса аллелей, вызванного ПГ в конкретном локусе опухолевой ДНК. Однако в методике количественной ПЦР все еще существуют трудности при выявлении очень малого дисбаланса аллелей в конкретном локусе, и, таким образом, варианты реализации, описанные в настоящем описании, позволяют анализировать области хромосом совместно.

Технология, описываемая в настоящем описании, также применима для выявления предраковых состояний. Примеры таких состояний включают цирроз печени и внутриэпителиальную неоплазию шейки матки. Первое состояние является предраковым состоянием для развития печеночно-клеточного рака, а второе состояние является предраковым для развития рака шейки матки. Есть сообщения о том, что такие предраковые состояния уже проявляют некоторые молекулярные изменения в своей эволюции на пути к злокачественной опухоли. Например, наличие ПГ в плечах хромосом 1p, 4q, 13q, 18q и сопутствующие потери более чем в 3 локусах ассоциированы с повышенным риском развития печеночно-клеточного рака у пациентов с циррозом печени (Roncalli M et al. *Hepatology* 2000;31:846-50). Такие

предраковые поражения также могли бы выделять ДНК в кровеносное русло, хотя, вероятно, в более низких концентрациях. Указанная технология может позволить проводить выявление делеций или увеличения числа копий путем анализа фрагментов ДНК в плазме с целью измерения концентрации (включая относительную концентрацию) циркулирующей предраковой ДНК в плазме.

Легкость, с которой выявляются такие аберрации (например, глубина секвенирования или количество таких выявляемых изменений) и концентрации могут предсказать вероятность или скорость прогрессирования развитого ракового состояния.

#### А. Делеция области хромосомы.

Фиг. 1 иллюстрирует область хромосомы раковой клетки, обнаруживающую аберрацию в форме делеции. Показано, что здоровая клетка обладает двумя гаплотипами, Нар I и Нар II. Как показано, и Нар I, и Нар II содержат последовательность в каждом из многих гетерозиготных локусов 110 (также называются однонуклеотидными полиморфизмами - SNP). В клетке, ассоциированной с раком, в Нар II 120 область хромосомы делетирована. В качестве примера раковая клетка может происходить из опухоли (например, злокачественной опухоли), из очага метастазирования (например, в регионарном лимфатическом узле или в отдаленном органе) или из предракового поражения, например, которые упоминались выше.

В 120 области хромосомы раковой клетки, в которой один из двух гомологичных гаплотипов делетирован, все гетерозиготные SNP могут выглядеть гомозиготными вследствие потери другого аллеля в соответствующей делетированной гомологичной хромосоме. Следовательно, данный тип хромосомной аберрации называют потерей гетерозиготности (ПГ). В 120 области неделетированные аллели указанных SNP могут представлять один или два гаплотипа, которые могут обнаруживаться в здоровых тканях. В примере, показанном на фиг. 1, гаплотип I (Нар I) в 120 области с ПГ можно определить путем генотипирования ткани опухоли. Другой гаплотип (Нар II) можно определить путем сравнения наблюдаемых генотипов здоровых тканей и раковых тканей. Нар II можно сконструировать путем присоединения всех делетированных аллелей. То есть все из аллелей в нормальной клетке для 120 области, которые не проявляются в 120 области раковой клетки, определяются как один и тот же гаплотип, т.е. Нар I. Посредством данного анализа можно определить гаплотипы пациентов (например, пациентов с печеночно-клеточным раком) для всех областей хромосом, проявляющих ПГ в ткани опухоли. Такой способ полезен, только если у человека есть раковые клетки, и он работает только для определения гаплотипа в 120 области, но не дает четкой картины для делетированной области хромосомы.

#### В. Амплификация области хромосомы.

Фиг. 2 иллюстрирует область хромосомы раковой клетки, проявляющую аберрацию в форме амплификации. Показано, что здоровая клетка обладает двумя гаплотипами, Нар I и Нар II. Как показано, и Нар I, и Нар II содержат последовательность в каждом из многих гетерозиготных локусов 210. В клетке опухоли Нар II содержит 220 область хромосомы дважды (дублицирована).

Аналогично для областей с моноаллельным увеличением числа копий в тканях опухоли амплифицированные аллели в 210 SNP можно выявить при помощи таких способов как анализ на микроматрицах. Один из двух гаплотипов (Нар II в примере, показанном на фиг. 2) можно определить путем соединения всех амплифицированных аллелей в 220 области хромосомы. Указанный амплифицированный аллель в конкретном локусе можно определить путем сравнения количеств каждого из аллелей в указанном локусе. Затем другой гаплотип (Нар I) можно определить путем соединения неамплифицированных аллелей. Такой способ применим, только если у человека есть раковые клетки, и он действителен только для определения гаплотипа в 220 области, но все же он дает четкую картину амплифицированных областей хромосом.

Амплификация может возникать из-за присутствия более чем 2 хромосом или из-за повторения гена в одной хромосоме. Одна область может удваиваться тандемно, или область может быть микрохромосомой, которая содержит одну или более копий указанной области. Увеличение числа копий также может возникать из-за копирования и вставки гена из одной хромосомы в другую хромосому или другую область в той же хромосоме. Такие инсерции являются определенным типом увеличения числа копий.

#### II. Выбор области хромосомы.

Поскольку раковая ткань будет обуславливать, по меньшей мере, некоторую часть из указанной бесклеточной ДНК (и потенциально клеточной ДНК), геномные аберрации указанной раковой ткани можно выявить в пробе такой плазмы или сыворотки. Проблема выявления указанных аберраций состоит в том, что указанная опухоль или рак могут быть слишком малы, вследствие чего они обеспечивают относительно малое количество ДНК из раковых клеток. Таким образом, количество циркулирующей ДНК с аберрациями слишком мало, что очень затрудняет выявление. ДНК в отдельном локусе генома может быть недостаточно для того, чтобы выявить аберрацию. Способы, описываемые в настоящем описании, позволяют преодолеть указанную трудность путем анализа ДНК в области хромосомы, включающей множество локусов (гаплотип), заменяя маленькие вариации в одном локусе на заметное различие, когда оно собирается со всего гаплотипа. Таким образом, анализ множества локусов некоторой области может обеспечить более высокую точность и может уменьшить ложно-положительные и ложно-отрицательные результаты.

Также область аберрации может быть слишком мала, что затрудняет выявление указанной аберрации. Если применяется только один локус или специфические локусы, аберрации не из указанных локусов не будут учтены. Некоторые способы, такие как описываемые в настоящем описании, позволяют изучать целые области, чтобы обнаружить аберрацию в подмножестве указанных областей. Когда анализируемые области перекрывают геном, таким образом, можно анализировать весь геном с целью обнаружения аберраций разной длины и локализации, как более подробно описано ниже.

Для иллюстрации вышесказанного одна область может содержать одну аберрацию. Но должна быть выбрана область для анализа. Длина и положение области может изменить результаты и, таким образом, повлиять на анализ. Например, если анализируют первую область на фиг. 1, никаких аберраций выявить нельзя. Если анализируют вторую область, можно выявить аберрацию, например, при помощи способов, описываемых в настоящем описании. Если бы анализировалась более крупная область, включающая как первую, так и вторую область, возникла бы та трудность, что только часть более крупной области содержит аберрацию, что может сделать выявление любой аберрации более трудной, а также встает проблема направления точной локализации и длины указанной аберрации. Разные варианты реализации могут быть направлены на решение некоторых и/или всех указанных трудностей. Описание выбора области в равной степени применимо к способам, в которых применяются гаплотипы той же области хромосомы или в которых применяются разные области хромосом.

#### А. Выбор конкретной области хромосомы

Согласно одному варианту реализации конкретную область можно выбрать на основании информации, имеющейся о конкретном типе рака или о пациенте. Например, может быть известно, что указанная область часто проявляет аберрации при многих типах рака или при конкретном типе рака. Точную длину и положение указанной области можно определить, обратившись к литературе, касающейся того, что хорошо известно о данном типе рака или о пациентах с конкретными факторами риска. Кроме того, можно получить и проанализировать ткани опухоли пациента, чтобы выявить области с аберрациями, как описано выше. В настоящее время такая технология может потребовать получения раковой клетки (которая может и не быть полученной конкретно от пациентов, которым только был поставлен диагноз), но такую технологию можно применять для выявления областей для мониторинга одного и того же пациента во времени (например, после операции по удалению раковой ткани или после химиотерапии, иммунотерапии или направленной терапии, или для выявления рецидива или прогрессирования опухоли).

Можно выявить более чем одну конкретную область. Анализ каждой из таких областей можно применять независимо или разные области можно анализировать в совокупности. Кроме того, область можно подразделить, чтобы получить более высокую точность определения локализации аберрации.

На фиг. 3 показана табл. 300, иллюстрирующая разные типы рака, ассоциированные с ними области и соответствующие им аберрации. В столбце 310 перечислены разные типы рака. Варианты реализации, описываемые в настоящем описании, можно применять для любого типа рака, подразумевающего наличие аберрации, и, таким образом, данный перечень является лишь примерным. В столбце 320 показаны области (например, большие области, такие как 7p или более специфические области 17q25), в которых добавление (амплификация) ассоциировано с конкретным раком в той же строке. В столбце 330 показаны области, в которых можно обнаружить потери (делеции). В столбце 340 перечислены ссылки на источники, в которых обсуждается взаимосвязь указанных областей с конкретным типом рака.

Указанные области с потенциальными хромосомными аберрациями можно применять в качестве областей хромосом для анализа согласно способам, описываемым в настоящем описании. Примеры других областей генома, изменяемых при раке, можно найти в базе данных Проекта анатомии генома при раке ([cgap.nci.nih.gov/Chromosomes/RecurrentAberrations](http://cgap.nci.nih.gov/Chromosomes/RecurrentAberrations)) и Атласа генетики и цитогенетики в онкологии и гематологии ([atlasgeneticsoncology.org/Tumors/Tumorliste.html](http://atlasgeneticsoncology.org/Tumors/Tumorliste.html)).

Как можно видеть, выявленные области могут быть достаточно крупными, а другие могут быть более специфичными. Аберрации не могут включать всю область, определенную в таблице. Таким образом, такие подсказки относительно типа аберраций не дают точного указания на то, где указанная аберрация фактически расположена для конкретного пациента, но могут применяться в большей степени как грубое указание на крупные области для анализа. Такие крупные области могут включать много подобластей (которые могут быть равного размера), которые анализируют индивидуально, а также в совокупности в более крупной области (подробности такого анализа описываются в настоящем описании). Таким образом, варианты реализации могут сочетать в себе аспекты выбора крупной области на основании особенностей рака, который требуется оценивать, но могут также включать применение более общих технологий (например, исследование подобластей), как описано далее.

#### В. Выбор произвольной области хромосомы

Согласно еще одному варианту реализации область хромосомы, которую будут оценивать, выбирают произвольно. Например, геном можно поделить на области, которые обладают длиной одна мегабаза (Мб), или другие заранее предопределенные сегменты с длиной 500 Кб или 2 Мб. Если области обладают длиной 1 Мб, то в геноме человека присутствует приблизительно 3000 областей, поскольку в гаплоидном геноме человека всего около 3 млрд оснований. Каждую из указанных областей затем можно анализировать, как более подробно обсуждается далее.

Такие области можно определять не на основании каких-либо знаний о конкретном типе рака, а на основании систематического разбиения генома на области, которые нужно анализировать. Согласно одному варианту реализации, когда хромосома не обладает длиной, которая кратна заранее определенному сегменту (например, не делится на миллион оснований), последняя область хромосомы может быть меньше заранее определенной длины (например, менее чем 1 Мб). Согласно еще одному варианту применения каждую хромосому можно разделить на области равной длины (или приблизительно равной в пределах погрешности округления), исходя из полной длины хромосомы и числа сегментов, которое требуется создать (что обычно может варьировать в разных хромосомах). Согласно таким вариантам применения длина указанных сегментов каждой хромосомы может быть разной.

Как упоминалось выше, конкретную область можно выявить на основании специфического рака, который оценивают, но затем указанную конкретную область можно разделить на более мелкие области (например, подобласти равного размера, которые перекрывают конкретную более крупную область). Таким способом можно точно установить аберрации. В приведенном ниже обсуждении любая общая ссылка на область хромосомы может относиться к области, которая специфически идентифицирована, к области, которая выбрана произвольно, или к их сочетанию.

### III. Выявление аберраций в конкретном гаплотипе.

В данном разделе описаны способы выявления аберрации в одиночной области хромосомы путем анализа биологического образца, которая включает бесклеточную ДНК. Согласно вариантам реализации настоящего раздела одиночная область хромосомы является гетерозиготной (разные аллели) во множестве локусов в указанной области, что обеспечивает наличие двух гаплотипов, которые можно различить, зная конкретный аллель в заданном локусе. Таким образом, заданную молекулу нуклеиновой кислоты (например, фрагмент бесклеточной ДНК) можно идентифицировать как есть из конкретного одного из двух гаплотипов. Например, указанный фрагмент можно секвенировать и получить маркерную последовательность, которая выравнивается с указанной областью хромосомы, а затем можно выявить указанный гаплотип в гетерозиготных локусах, к которым принадлежит указанный аллель. Далее описаны технологии двух общих типов, которые применяют для определения аберраций в конкретном гаплотипе (Нар), а именно подсчет маркерных последовательностей и анализ по размерам.

#### A. Определение гаплотипа.

Чтобы различить два гаплотипа, сначала определяют два указанных гаплотипа области хромосомы. Например, можно определять два гаплотипа Нар I и Нар II, показанные в нормальной клетке на фиг. 1. На фиг. 1 указанные гаплотипы включают первое множество локусов 110, которое является гетерозиготным, что позволяет различать два указанных гаплотипа. Указанное первое множество локусов перекрывает область хромосомы, подлежащую анализу. Можно сначала определить аллели в разных гетерозиготных локусах (полокусное геновое разнообразие), а затем применить их для определения гаплотипов у пациентов.

Гаплотип указанных аллелей SNP можно определить при помощи способов анализа одиночных молекул. Примеры таких способов были описаны у Fan et al. (Nat Biotechnol. 2011; 29:51-7), Yang et al. (Proc Natl Acad Sci USA. 2011; 108:12-7) и Kitzman et al. (Nat Biotechnol. 2011 Jan; 29:59-63). В качестве альтернативы гаплотипы индивидуума можно определить посредством анализа генотипов членов семьи (например, родителей, родных сестер и братьев, детей). Примеры включают способы, описанные у Roach et al. (Am J Hum Genet. 2011; 89(3):382-97) и Lo et al. (Sci Transl Med. 2010; 2:61ra91). Согласно еще одному варианту реализации гаплотип индивидуумов можно определить посредством сравнения результатов генотипирования тканей опухоли и геномной ДНК. Генотипирование указанных субъектов можно проводить путем анализа на микроматрицах.

Также гаплотипы можно сконструировать другими способами, хорошо известными специалистам в данной области техники. Примеры таких способов включают способы, основанные на анализе одиночных молекул, таких как количественная ПЦР (Ding C and Cantor CR. Proc Natl Acad Sci USA 2003; 100: 7449-7453; Ruano G et al. Proc Natl Acad Sci USA 1990; 87: 6296-6300), сортировка или разделение хромосом (Yang H et al. Proc Natl Acad Sci USA 2011; 108: 12-17; Fan HC et al. Nat Biotechnol 2011; 29: 51-57), гаплотипирование спермы (Lien S et al. Curr Protoc Hum Genet 2002; Chapter 1:Unit 1.6) и технологии визуализации (Xiao M et al. Hum Mutat 2007; 28: 913-921). Другие способы включают способы, основанные на аллель-специфичной ПЦР (Michalatos-Beloin S et al. Nucleic Acids Res 1996; 24: 4841-4843; Lo YMD et al. Nucleic Acids Res 19: 3561-3567), клонирование и расщепление рестриктазами (Smirnova AS et al. Immunogenetics 2007; 59:93-8) и др. Еще некоторые способы основаны на распределении и структуре неслучайного распределения блоков гаплотипов в популяции, которые позволяют вывести гаплотип субъекта из статистических оценок (Clark AG. Mol Biol Evol 1990; 7:111-22; 10:13-9; Salem RM et al. Hum Genomics 2005; 2:39-66).

Другой способ определения гаплотипа области ПГ основан на генотипировании нормальных тканей и тканей опухоли указанного субъекта, если ткани опухоли доступны. При наличии ПГ ткани опухоли с очень высокой концентрацией клеток опухоли могут продемонстрировать кажущуюся гомозиготность для всех локусов SNP в пределах области, демонстрирующей ПГ. Генотипы указанных локусов SNP могут содержать один гаплотип (Нар I области ПГ на фиг. 1). С другой стороны, нормальные ткани могут

показать, что указанный субъект гетерозиготен по указанным локусам SNP в пределах области, демонстрирующей ПГ. Аллели, которые присутствуют в нормальных тканях, но не в тканях опухоли, могут содержать другой гаплотип (Нар II области ПГ на фиг. 1).

В. Анализ относительной дозировки гаплотипа (RHDO).

Как упоминалось выше, хромосомные aberrации с амплификацией или делецией одного из гаплотипов области хромосомы могут приводить к дисбалансу дозировки двух гаплотипов в указанной области хромосомы в тканях опухоли. В плазме крови человека с опухолевым ростом определенная доля циркулирующей ДНК происходит из клеток указанной опухоли. Из-за присутствия ДНК из опухоли в плазме пациентов, страдающих раковыми заболеваниями, также может присутствовать такой дисбаланс в плазме. Дисбаланс дозировки двух указанных гаплотипов можно выявить путем подсчета числа молекул, происходящих из каждого гаплотипа.

Для областей хромосом, в которых наблюдается ПГ в тканях опухоли (например, 120 область на фиг. 1) Нар I может быть представлен в избытке среди циркулирующих молекул ДНК (фрагментов) по сравнению с Нар II из-за недостаточной доли Нар II из тканей опухоли. Для областей хромосом, в которых в тканях опухоли наблюдается амплификация, Нар II может присутствовать в избытке по сравнению с Нар I для областей, вовлеченных в моноаллельную амплификацию Нар II, из-за выделения дополнительной дозы Нар II из тканей опухоли. Чтобы определить избыточное и недостаточное присутствие, определяют, происходят ли определенные фрагменты в ДНК в пробе из Нар I или Нар II, что можно сделать при помощи разных способов, например путем проведения универсального секвенирования и выравнивания с применением количественной ПНР сайт-специфичных зондов.

После секвенирования множества фрагментов ДНК из плазмы (или другого биологического образца) пациентов, страдающих раковыми заболеваниями, с целью сгенерировать секвенированные маркерные последовательности указанные секвенированные маркерные последовательности, соответствующие аллелям двух указанным гаплотипам, можно выявить и подсчитать. Количество секвенированных маркерных последовательностей, соответствующих каждому из двух гаплотипов, в равной степени представлено в плазме. Согласно одному варианту реализации для определения того, значительно ли различается присутствие двух указанных гаплотипов, можно применять последовательный критерий отношения вероятностей (SPRT). Статистически значимое различие предполагает наличие хромосомной aberrации в анализируемой области хромосомы. Кроме того, количественное различие двух указанных гаплотипов в плазме можно применять для оценки относительной концентрации ДНК, происходящей из опухоли, как описано ниже.

Диагностические способы для определения идентичности фрагмента ДНК (например, его локализации в геноме человека), описываемые в настоящем описании, не ограничиваются применением массивно-параллельного секвенирования в качестве платформы для выявления. Указанные диагностические способы можно также применять, например, без ограничений, для микрофлюидных систем для количественной ПЦР (например, цифровая матричная система Fluidigm, микрокапельная система для количественной ПЦР (например, система от компании "RainDance" и "QuantaLife"), системы BEAM-ing (например, ПЦР, амплификация и магнитные измерения на гранулах, в эмульсии) (Diehl et al. ProcNatAcadSci USA 2005; 102: 16368-16373), ПЦР в режиме реального времени, системы на основе масс-спектрометрии (например, система SequenomMassArray) и анализ на основе мультиплексной, зависящей от лигирования амплификации зондов (MLPA).

#### Нормальные области

Фиг. 4 иллюстрирует области хромосом в раковой клетке, которые не обнаруживают aberrации, а также измерения, сделанные в плазме согласно вариантам реализации настоящего изобретения. 410 область хромосомы можно выбрать любым способом, например, в зависимости от специфического типа рака, который требуется оценивать, или на основании обычного скрининга, при котором применяют заранее определенные сегменты, перекрывающие большие участки генома. Чтобы различить два гаплотипа, сначала определяют эти два гаплотипа. На фиг. 4 показаны два гаплотипа (Нар I и Нар II) нормальной клетки для 410 области хромосомы. Указанные гаплотипы включают первое множество локусов 420. Указанное первое множество локусов 420 перекрывает 410 область хромосомы, которую анализируют. Как показано, указанные локусы в нормальной клетке гетерозиготны. Показаны два гаплотипа для раковой клетки. В раковой клетке ни одна из областей не подверглась делеции или амплификации.

Также на фиг. 4 показано число встречаемостей аллелей в каждом гаплотипе для каждого локуса 420. Также приведена сумма нарастающим итогом для определенных подобластей 410 области хромосомы. Число встречаемостей аллелей соответствует количеству фрагментов ДНК, соответствующих конкретному гаплотипу в каждом конкретном локусе. Например, фрагмент ДНК, который включает указанные первые локусы 421 и содержит аллель А, можно засчитать как Нар I. А фрагмент ДНК с аллелем Т можно засчитать как Нар II. Выяснить, выравнивается ли фрагмент (т.е. включает ли он конкретный локус) и какой аллель он содержит, можно разными способами, как упоминается в настоящем описании. Можно применять соотношение количеств двух гаплотипов, чтобы определить, есть ли значимое различие. В настоящем описании данное соотношение называется соотношением шансов. Также можно применять различие между двумя значениями; указанное различие можно нормировать на

общее число фрагментов. Отношение и различие (и функции от них) являются примерами параметров, которые сравнивают с пороговыми значениями, чтобы определить классификацию или выяснить, существует ли абберация.

Можно применять анализ RHDO всех аллелей в одном и том же гаплотипе (например, накопленную встречаемость) для определения, есть ли какой-либо дисбаланс двух гаплотипов в плазме, например, как это можно сделать с плазмой матери, как описано в заявках на патенты Lo 12/940 992 и 12/940 993, о которых речь шла выше. Указанный способ может значительно увеличить количество молекул ДНК, применяемых для определения того, есть ли какой-либо дисбаланс и, следовательно, позволяет достичь более высокой статистической мощности для дифференциации дисбаланса в связи с наличием рака по распределению вероятностей встречаемости аллелей в отсутствии рака или предракового состояния. В отличие от раздельного анализа множественных локусов SNP, способ RHDO может основываться на относительном положении аллелей в двух указанных хромосомах (информация о гаплотипе) так, что аллели, расположенные в одной и той же хромосоме, можно анализировать вместе. В отсутствии информации о гаплотипе встречаемости аллелей разных локусов SNP нельзя складывать вместе, чтобы оценить статистически, представлен ли гаплотип в плазме в недостаточной или избыточной доле. Количественное определение встречаемости аллелей можно проводить без ограничений посредством массивно-параллельного секвенирования (например, посредством секвенирования Illumina при помощи системы синтеза, секвенирования при помощи технологии сшивки (SOLiD) от компании "Life Technologies", системы секвенирования Ion Torrent от компании "Life Technologies", секвенирования на нанопорах (nanoporetech.com) и посредством технологии секвенирования 454 ("Roche"), количественной ПЦР (например, посредством микрокапельной количественной ПЦР (например, Fluidigm (fluidigm.com)), или BEAming (например, ПЦР, амплификация и магнитные измерения на гранулах, в эмульсии), или капельной ПЦР (например, от компании "QuantaLife" (quantalife.com) и "RainDance" (raindancetechnologies.com)), а также ПЦР в режиме реального времени. Согласно другому варианту применения указанной технологии можно применять обогащенное направленное секвенирование с применением захвата из раствора (например, с применением системы Agilent SureSelect, пользовательского набора для обогащения Illumina TruSeq (illumina.com/applications/sequencing/targeted\_resequencing.ilmn), или с применением пользовательского набора для обогащения MyGenostics GenCap (mygenostics.com/)), или захвата на основе матриц (например, с применением системы Roche NimbleGen).

В примере, показанном на фиг. 4, наблюдается слабый дисбаланс аллелей для первых двух локусов SNP (24 и 26 для первого SNP, и 18 и 20 для второго SNP). Однако встречаемость аллелей недостаточна со статистической точки зрения для определения, присутствует ли реальный дисбаланс аллелей. Таким образом, встречаемости аллелей в одном и том же гаплотипе складывают до тех пор, пока накопленных встречаемостей аллелей для двух указанных гаплотипов не будет достаточно, чтобы статистически достоверно заключить, что для 410 области хромосомы между двумя указанными гаплотипами дисбаланса аллелей нет (пятый SNP в данном примере). После достижения статистически значимой классификации, накопленная встречаемость обнуляется (в шестом SNP в данном примере). Затем накопленную встречаемость определяют до тех пор, пока накопленных встречаемостей аллелей для двух указанных гаплотипов снова не будет достаточно, чтобы сделать статистическое заключение об отсутствии дисбаланса аллелей между двумя указанными гаплотипами для данной конкретной подобласти 410 области. Также можно применять итоговую накопленную встречаемость для всей области, но предыдущий способ позволяет оценивать разные подобласти, что обеспечивает более высокую точность (т.е. подобласть) при определении локализации абберации, в отличие от целой области 410. Примеры статистических критериев для определения, присутствует ли реальный дисбаланс, включают без ограничений последовательный критерий отношения вероятностей (Zhou W, et al. Nat Biotechnol 2001; 19: 78-81; Zhou W, et al. Lancet. 2002; 359: 219-25), t-критерий Сьюдента и критерий хи-квадрат.

#### Выявление делеций

Фиг. 5 иллюстрирует делецию 510 области хромосомы в раковой клетке, а также измерения, сделанные в плазме для определения делетированной области согласно вариантам реализации настоящего изобретения. На фиг. 5 показаны два гаплотипа (Нар I и Нар II) нормальной клетки для 510 области хромосомы. Указанные гаплотипы включают первое множество гетерозиготных локусов 520, которые перекрывают 510 область хромосомы, подвергаемую анализу. Показаны два гаплотипа для раковой клетки. В раковой клетке для гаплотипа II область 510 подверглась делеции. Как и на фиг. 4, на фиг. 5 также показано число встречаемости аллелей, для каждого локуса 520. Также приведена сумма нарастающим итогом для определенных подобластей 510 области хромосомы.

Поскольку ткани опухоли обычно содержат смесь опухолевых и неопухолевых клеток, ПГ может проявляться по отклонению соотношения количеств двух указанных аллелей в локусах в 510 области. В такой ситуации делетированный гаплотип Нар II в 510 области можно определить на основании сочетания локусов 520, которые демонстрируют относительное уменьшение количества фрагментов ДНК по сравнению с соответствующими локусами в здоровых тканях. Гаплотип с фрагментами, которые появляются чаще, является Нар I, который сохранен в клетках опухоли. Согласно определенным вариантам реализации может быть желательно провести процедуру, которая могла бы обогатить долю раковых кле-

ток в пробе опухоли, чтобы было легче определять и делеированный, и сохраненный гаплотипы. Одним примером такой процедуры является микродиссекция (либо ручная, либо при помощи технологий лазерного захвата).

Теоретически каждый из аллелей в Нар I может быть представлен в избытке в циркулирующей ДНК для области хромосомы, проявляющей ПГ в тканях опухоли, и степень дисбаланса аллелей может зависеть от относительной концентрации ДНК опухоли в плазме крови. Однако в то же время относительная распространенность двух указанных аллелей в любой пробе циркулирующей ДНК также должна подчиняться закону распределения Пуассона. Можно проводить статистический анализ, чтобы определить, связан ли наблюдаемый дисбаланс аллелей с наличием ПГ в раковых тканях или с вероятностью. Мощность выявления реального дисбаланса аллелей, ассоциированного с ПГ в раковой опухоли, зависит от количества циркулирующих молекул ДНК, которые анализируют, и от относительной концентрации опухолевой ДНК. Более высокая относительная концентрация опухолевой ДНК и большее количество анализируемых молекул могут обеспечить более высокую чувствительность и специфичность выявления реального дисбаланса аллелей.

В примере, показанном на фиг. 5, наблюдается слабый дисбаланс аллелей для первых двух локусов SNP (24 и 22 для первого SNP, и 18 и 15 для второго SNP). Однако значения встречаемости аллелей не достаточно со статистической точки зрения для определения, присутствует ли реальный дисбаланс аллелей. Таким образом, встречаемости аллелей в одном и том же гаплотипе складываются до тех пор, пока накопленных встречаемостей аллелей для двух указанных гаплотипов не будет достаточно, чтобы статистически достоверно заключить, что в 510 области присутствует дисбаланс аллелей между двумя указанными гаплотипами (пятый SNP в данном примере). Согласно некоторым вариантам реализации известен только дисбаланс, а специфический тип (делеция или увеличение числа копий) не известен. Затем накопленную встречаемость определяют до тех пор, пока накопленных встречаемостей аллелей для двух указанных гаплотипов снова не будет достаточно, чтобы сделать статистическое заключение о дисбалансе аллелей между двумя указанными гаплотипами для данной конкретной подобласти 510 области. Также можно применять итоговую накопленную встречаемость для всей области, как можно делать при любом способе, описываемом в настоящем описании.

#### Выявление амплификации хромосомы

Фиг. 6 иллюстрирует амплификацию 610 области хромосомы в раковой клетке, а также измерения, сделанные в плазме для определения амплифицированной области согласно вариантам реализации настоящего изобретения. Помимо ПГ в раковых тканях также часто наблюдается амплификация областей хромосом. В примере, показанном на фиг. 6, в раковой клетке число копий Нар II в 610 области хромосомы увеличено до трех. Как показано, 610 область включает только шесть гетерозиготных локусов, в противоположность более длинным областям, показанным на предыдущих фигурах. Увеличение числа копий выявляются как статистически значимое в шести локусах, где определяется статистически значимое избыточное присутствие. Согласно некоторым вариантам реализации известен только дисбаланс, а специфический тип (делеция или амплификация) не определяется. Согласно другим вариантам реализации можно получить и анализировать раковую клетку. Такой анализ может предоставить информацию о том, связан ли указанный дисбаланс с делецией (раковая клетка гомозиготна по делетированной области) или с амплификацией (раковая клетка гетерозиготна по амплифицированной области). Согласно другим вариантам применения определить, присутствует ли делеция или амплификация, можно при помощи способов, описанных в разделе IV, для анализа целой области (т.е. не отдельно взятых гаплотипов). Если указанная область присутствует в избытке, то абберрация является увеличением числа копий; а если указанная область присутствует в недостатке, то абберрация является делецией. Также анализируют 620 область, и накопленная встречаемость подтверждает, что дисбаланса не существует.

#### Анализ SPRT для анализа RHDO плазмы

Для любых областей хромосом, которые содержат гетерозиготные локусы, анализ RHDO можно применять для определения того, существует ли дисбаланс дозировок двух указанных гаплотипов в плазме. В указанных областях наличие дисбаланса дозировок гаплотипов в плазме свидетельствует о наличии ДНК опухолевого происхождения в указанной пробе плазмы. Согласно одному варианту реализации анализ SPRT можно применять для определения того, является ли статистически значимым различие в числе секвенированных считываний для Нар I и Нар II. В данном примере анализа SPRT мы сначала определяли число секвенированных считываний, поступающих от каждого из двух указанных гаплотипов. Затем можно определить параметр (например, долю), которая соответствует пропорциональному количеству секвенированных считываний, в которое вносит вклад потенциально представленный в избытке гаплотип (например, доля числа считываний для одного гаплотипа, разделенная на число считываний другого гаплотипа). Потенциально представленный в избытке гаплотип может быть неделетированным гаплотипом в сценарии ПГ и амплифицированным гаплотипом в сценарии моноаллельного увеличения числа копий области хромосомы. Затем указанную долю можно сравнивать с пороговыми значениями (верхним и нижним пороговым значением), которые сконструированы на основании нулевой гипотезы, т.е. отсутствия дисбаланса дозировки гаплотипов, и альтернативной гипотезы, т.е. присутствия дисбаланса дозировки гаплотипов. Если указанная доля больше, чем верхнее пороговое значение, это

указывает на наличие статистически значимого дисбаланса двух указанных гаплотипов в плазме. Если указанная доля ниже, чем нижнее пороговое значение, это указывает на то, что нет статистически значимого дисбаланса двух указанных гаплотипов. Если указанная доля попадает в диапазон между верхним и нижним пороговым значением, это указывает на то, что статистическая мощность недостаточна, чтобы сделать заключение. Можно проводить последовательное увеличение количества гетерозиготных локусов для области, которую анализируют, до тех пор, пока не будет сделана успешная классификация SPRT. Уравнения для расчета верхних и нижних границ SPRT следующие:

$$\text{Верхнее пороговое значение} = [(\ln 8)/N - \ln \delta] / \ln \gamma ;$$

$$\text{Нижнее пороговое значение} = [(\ln 1/8)/N - \ln \delta] / \ln \gamma ,$$

где

$$\delta = (1 - \theta_1)/(1 - \theta_2) \text{ и } \gamma = \frac{\theta_1(1 - \theta_2)}{\theta_2(1 - \theta_1)},$$

$\theta_1$  представляют собой ожидаемую долю секвенированных маркерных последовательностей из потенциально представленного в избытке гаплотипа, когда в плазме имеет место дисбаланс аллелей,

$\theta_2$  представляет собой ожидаемую долю любых двух гаплотипов, когда нет дисбаланса аллелей, т.е. 0,5,  $N$  - общее количество секвенированных маркерных последовательностей для Нар I и Нар II,  $\ln$  - математический символ, обозначающий натуральный логарифм (т.е.  $\log_e$ ).  $\theta_1$  может зависеть от относительной концентрации ДНК опухолевого происхождения ( $F$ ), которая, как предполагается (или известно), присутствует в пробе плазмы.

В варианте ПГ  $\theta_1 = 1/(2-F)$ . В варианте моноаллельной амплификации  $\theta_1 = (1+zF)/(2+zF)$ , где  $z$  представляет собой количество дополнительных копий области хромосомы, которая амплифицирована в опухоли. Например, если одна хромосома дуплицирована, будет одна дополнительная копия конкретной хромосомы. Тогда  $z$  равно 1.

На фиг. 7 показан анализ RHDO ДНК в плазме пациента с НСС для сегмента, локализованного в хромосоме 1p, которая демонстрирует амплификацию одного аллеля в ткани опухоли согласно вариантам реализации настоящего изобретения. Зеленными треугольниками показаны данные пациента. Общее количество секвенированных считываний увеличилось при увеличении количества анализируемых SNP. Доля всех секвенированных считываний и амплифицированного гаплотипа в опухоли варьирует при увеличении количества всех анализируемых секвенированных считываний и в итоге достигает значения, превышающего верхнее пороговое значение. Это указывает на значимый дисбаланс дозировки гаплотипов и, следовательно, подтверждает наличие в плазме данной ассоциированной с раком аберрации.

Для всех областей хромосом пациента с НСС, демонстрирующих амплификации и делеции в тканях опухоли, проводили анализ RHDO с применением SPRT. Результаты приведены далее для 992 сегментов, которые, как известно, демонстрируют ПГ, и для 105 сегментов, которые, как известно, демонстрируют амплификации. Для ПГ 922 сегмента были классифицированы при помощи SPRT, и 921 сегмент был правильно идентифицирован как демонстрирующий дисбаланс дозировки гаплотипов в плазме, и была достигнута точность 99,99%. Для моноаллельной амплификации 105 сегментов были классифицированы при помощи SPRT, и 105 сегментов были правильно идентифицированы как демонстрирующие дисбаланс дозировки гаплотипов в плазме, и была достигнута точность 100%.

С. Анализ относительного размера гаплотипа.

В качестве альтернативы подсчету дозировки фрагментов, выровненных с двумя указанными гаплотипами, можно применять размер фрагментов для соответствующих гаплотипов. Например, для конкретной области хромосомы размер фрагментов ДНК из одного гаплотипа можно сравнивать с размером фрагментов ДНК другого гаплотипа. Можно проанализировать распределение по размерам фрагментов ДНК, которые соответствуют любому из аллелей гетерозиготных локусов первого гаплотипа указанной области, и сравнить его с распределением по размерам фрагментов ДНК, которые соответствуют любому из аллелей гетерозиготных локусов второго указанного гаплотипа. Статистически значимое различие распределений по размеру можно применять для идентификации аберрации, аналогично с применением количества встречаемости.

Сообщалось, что распределение по размеру всей ДНК плазмы (т.е. опухолевой и неопухолевой) у пациентов, страдающих раковыми заболеваниями, увеличено (Wang BG, et al. Cancer Res. 2003; 63: 3966-8). Однако если специально изучать ДНК опухолевого происхождения вместо количества всей ДНК (т.е. опухолевой и неопухолевой), было отмечено, что распределение по размеру молекул ДНК опухолевого происхождения более узкое, чем распределение молекул, полученных из неопухолевых клеток (Diehl et al. Proc Natl Acad Sci USA. 2005;102:16368-73). Следовательно, распределение по размеру циркулирующей ДНК можно применять для определения того, присутствуют ли ассоциированные с раком хромосомные аберрации. Принцип анализа по размерам показан на фиг. 8.

На фиг. 8 показано изменение распределения по размеру фрагментов двух гаплотипов области хромосомы, когда присутствует опухоль, содержащая делецию согласно вариантам реализации настоящего изобретения. Как проиллюстрировано на фиг. 8, аллель T в указанной ткани опухоли делетирован. Как

следствие, ткани опухоли могут выделять в плазму только короткие молекулы указанного аллеля А. Короткие молекулы ДНК опухолевого происхождения могут приводить к общему сужению распределения по размеру в плазме для аллеля А, таким образом, приводя к сужению распределения по размеру аллеля А по сравнению с распределением аллеля Т в плазме. Как обсуждалось выше, все аллели, локализованные в одном и том же гаплотипе, можно анализировать вместе. Иными словами, распределение по размеру молекул ДНК, несущих аллели, локализованные в одном гаплотипе, можно сравнивать с распределением молекул ДНК, несущих аллели, локализованные в другом гаплотипе. Гаплотип, делетированный в тканях опухоли, может демонстрировать более широкое распределение по размеру в плазме.

Анализ размеров также можно применять для выявления амплификации областей хромосомы, ассоциированных с раком. На фиг. 9 показано изменение распределения по размеру фрагментов двух гаплотипов области хромосомы, когда присутствует опухоль, содержащая амплификацию согласно вариантам реализации настоящего изобретения. В примере, показанном на фиг. 9, аллель Т в указанной опухоли дублирован. Как следствие, в плазму может выделяться увеличенное количество коротких молекул ДНК опухолевого происхождения, таким образом, приводя к сужению распределения по размеру аллеля Т по сравнению с распределением аллеля А. Анализ размеров можно применять совместно для всех аллелей, локализованных в одном и том же гаплотипе. Иными словами, распределение по размеру для гаплотипа с увеличенным числом копий в тканях опухоли может быть уже, чем распределение по размеру для гаплотипа без увеличенного числа копий в тканях опухоли.

#### Выявление сужения распределения по размеру циркулирующей ДНК

Размер фрагментов ДНК, возникающих из двух гаплотипов, а именно Нар I и Нар II, можно определять без ограничений при помощи массивно-параллельного секвенирования с парными концами. После секвенирования концов некоторого фрагмента ДНК секвенирование считывания (маркерные последовательности) можно выровнять с эталонным геном человека. О размере секвенированных молекул ДНК можно сделать вывод на основании координат крайнего нуклеотида на каждом конце. Секвенированные маркерные последовательности указанной молекулы можно применять для определения того, происходит ли указанный фрагмент ДНК из Нар I или Нар II. Например, одна из секвенированных маркерных последовательностей может включать гетерозиготный локус в анализируемой области хромосомы.

Следовательно, для каждой секвенируемой молекулы мы можем определить как размер, так и то, происходит ли она из Нар I или Нар II. На основании размера указанных фрагментов, выровненных с каждым гаплотипом, компьютерная система может рассчитать профиль распределения по размерам (например, средний размер фрагмента) как для Нар I, так и для Нар II. Распределение по размерам фрагментов ДНК из Нар I и Нар II можно сравнивать при помощи соответствующего статистического анализа, чтобы определить, когда распределение по размерам достаточно различается, чтобы идентифицировать аберрацию. Помимо массивно-параллельного секвенирования с парными концами для определения размера фрагментов ДНК можно применять другие способы, включая без ограничений секвенирование целых фрагментов ДНК, масс-спектрометрию и оптические способы для выявления и сравнения длины наблюдаемых молекул ДНК со стандартом.

Далее мы внедрили два примерных способа для выявления укорочения циркулирующей ДНК, ассоциированного с генетическими аберрациями опухолей. Указанные два способа направлены на проведение количественных измерений различий в распределении по размеру для двух популяций фрагментов ДНК. Указанные две популяции фрагментов ДНК приписывают молекулам ДНК, соответствующим Нар I и Нар II.

#### Различие в доле коротких фрагментов ДНК

Согласно одному способу применения применяют долю коротких фрагментов ДНК. Устанавливают размеры отсечки ( $w$ ), чтобы определить, какие молекулы ДНК являются короткими. Размер отсечки может варьировать, и можно выбирать его, чтобы удовлетворить разным диагностическим целям. Компьютерная система может определить количество молекул, которые имеют размер меньше либо равный размеру отсечки. Затем можно рассчитать долю фрагментов ДНК ( $Q$ ) путем деления числа коротких ДНК на общее число фрагментов ДНК. На значение  $Q$  может влиять распределение по размеру указанной популяции молекул ДНК. Более узкое общее распределение по размеру показывает, что большая доля молекул ДНК может представлять собой короткие фрагменты, что дает более высокое значение  $Q$ .

Затем можно применять различие в доле коротких фрагментов ДНК между Нар I и Нар II. Различие в распределении по размеру фрагментов ДНК из Нар I и Нар II может отражаться на разнице в доле коротких фрагментов для Нар I и Нар II ( $\Delta Q$ ).

$\Delta Q = Q_{\text{Нар I}} - Q_{\text{Нар II}}$ , где  $Q_{\text{Нар I}}$  представляет собой долю коротких фрагментов для фрагментов ДНК Нар I; а  $Q_{\text{Нар II}}$  представляет собой долю коротких фрагментов для фрагментов ДНК Нар II.  $Q_{\text{Нар I}}$  и  $Q_{\text{Нар II}}$  являются примерами статистического показателя двух групп распределений по размеру фрагментов из каждого гаплотипа.

Как показано в предыдущем разделе, когда Нар I выявляется в тканях опухоли, распределение по размеру фрагментов ДНК Нар I может быть уже, чем для фрагментов ДНК Нар II. Как следствие, может наблюдаться положительное значение  $\Delta Q$ . Положительное значение  $\Delta Q$  можно сравнивать с пороговым

значением, чтобы определить, достаточно ли велико AQ, чтобы делецию считать существующей. Амплификация Нар I также может демонстрировать положительное значение AQ. Когда в тканях опухоли имеет место дупликация Нар II, распределение по размеру фрагментов ДНК Нар II может быть уже, чем для фрагментов ДНК Нар I. Следовательно, значение AQ может стать отрицательным. В отсутствии хромосомной аберрации распределение по размерам фрагментов ДНК Нар I и Нар II в плазме/сыворотке может быть сходным. Следовательно, значение AQ может быть приблизительно нулевым.

AQ определенного пациента можно сравнивать с AQ нормальных индивидуумов, чтобы определить, нормальное ли данное значение. Дополнительно или в качестве альтернативы значение AQ пациента можно сравнивать со значениями, полученными у пациентов, страдающих аналогичным типом рака, чтобы определить, аномальное ли данное значение. Такое сравнение может включать в себя сравнение с пороговыми значениями, описываемыми в настоящем описании. В контексте мониторинга заболевания значение AQ можно отслеживать последовательно во времени. Изменение значения AQ может указывать на увеличение относительной концентрации опухолевой ДНК в плазме/сыворотке крови. Согласно избранному варианту применения указанной технологии относительная концентрация опухолевой ДНК может быть коррелирована со стадией опухоли, прогнозированием и прогрессированием заболевания. Такие варианты применения, основанные на измерениях в разные времена, более подробно обсуждаются далее.

Различие в доле в общей длине, приписываемой коротким фрагментам ДНК

Согласно данному варианту применения применяют долю в общей длине, которую приписывают коротким фрагментам ДНК. Компьютерная система может определить общую длину группы фрагментов ДНК в пробе (например, фрагменты из конкретного гаплотипа данной области или просто из данной области). Можно выбрать размер отсечки ( $w$ ), фрагменты ДНК меньше которого определяют как "короткие фрагменты". Размер отсечки может варьировать, и можно выбирать его, чтобы удовлетворить разным диагностическим целям. Затем компьютерная система может определить общую длину коротких фрагментов ДНК путем суммирования длины случайно выбранных фрагментов ДНК, которые равны или короче размера отсечки. Затем можно рассчитать долю в общей длине, приписываемую коротким фрагментам ДНК, которую можно рассчитать следующим образом:

$F = \sum^w \text{length} / \sum^N \text{length}$ , где  $\sum^w \text{length}$  представляет собой сумму длин фрагментов ДНК с длиной, равной или меньше  $w$ (bp); и  $\sum^N \text{length}$  представляет собой сумму длин фрагментов ДНК с длиной, равной или меньше, чем заранее установленная длина  $N$ . Согласно одному варианту реализации  $N$  составляет 600 оснований. Однако для расчета "общей длины" можно применять другие ограничения размеров, например 150 оснований, 180 оснований, 200 оснований, 250 оснований, 300 оснований, 400 оснований, 500 оснований и 700 оснований.

Может быть выбрано значение 600 оснований или менее, поскольку система Illumina Genome Analyzer не эффективна при амплификации и секвенировании фрагментов ДНК более 600 оснований. Кроме того, ограничение анализа фрагментами ДНК короче 600 оснований также позволяет избежать погрешности, возникающей из-за структурных вариаций генома. В присутствии структурных вариаций, например перестроек (Kidd JM et al, Nature 2008; 453:56-64), размер фрагментов генома можно переоценить, когда размер оценивают способами биоинформатики, путем картирования концов указанного фрагмента ДНК относительно эталонного генома. Кроме того, >99,9% всех фрагментов ДНК, успешно секвенированных и картированных относительно эталонного генома, короче 600 оснований и, таким образом, включение всех фрагментов, равных или короче 600 оснований, могут обеспечить репрезентативную оценку распределения по размеру фрагментов ДНК в пробе.

Соответственно можно применять различие в долях в общей длине, приписываемых коротким фрагментам ДНК, между гаплотипами Нар I и Нар II. Отклонение распределения по размеру между фрагментами ДНК Нар I и Нар II может отражаться в различии их значений  $F$ . Здесь мы определяем  $F_{\text{Нар I}}$  и  $F_{\text{Нар II}}$  как доли в общей длине, приписываемые коротким фрагментам ДНК Нар I и Нар II соответственно. Разницу в долях в общей длине, приписываемых коротким фрагментам ДНК, между Нар I и Нар II

$\Delta F = F_{\text{Нар I}} - F_{\text{Нар II}}$ .  $F_{\text{Нар I}}$  и  $F_{\text{Нар II}}$  представляют собой примеры статистического показателя двух групп распределений по размеру фрагментов из каждого гаплотипа.

Аналогично вариантам реализации, проиллюстрированным в предыдущем разделе, делеция Нар II в тканях опухоли может приводить к кажущемуся сужению распределения по размеру фрагментов ДНК Нар по сравнению с распределением фрагментов ДНК Нар II. Это может приводить к тому, что значение  $\Delta F$  будет положительным. Когда Нар II дуплицирован, может наблюдаться отрицательное значение  $\Delta F$ . В отсутствии хромосомной аберрации значение  $\Delta F$  может быть приблизительно нулевым.

Значение  $\Delta F$  определенного пациента можно сравниваться со значением  $\Delta F$  у здоровых индивидуумов, чтобы определить нормальное ли указанное значение. Значение  $\Delta F$  определенного пациента можно сравниваться со значением  $\Delta F$ , полученным у пациента с аналогичными типами рака, чтобы определить аномальное ли указанное значение. Такое сравнение может включать в себя сравнение с пороговыми значениями, описываемыми в настоящем описании. В контексте мониторинга заболевания значение  $\Delta Q$  можно отслеживать во времени последовательно. Изменение значения  $\Delta F$  может указывать на увеличе-

ние относительной концентрации опухолевой ДНК в плазме/сыворотке крови.

D. Общий способ.

На фиг. 10 показана блок-схема способа анализа гаплотипов биологической пробы организма с целью определения, обнаруживает ли определенная область хромосомы делецию или амплификацию согласно вариантам реализации настоящего изобретения. Указанный биологический образец включает молекулы нуклеиновых кислот (также называемые фрагментами), полученные из здоровых клеток и, возможно, из клеток, ассоциированных с раком. Указанные молекулы могут присутствовать в указанном образце в бесклеточной форме. Указанный организм может быть любого типа, который содержит более одной копии хромосомы, т.е., по меньшей мере, диплоидным организмом, но может включать и организмы с более высокой степенью полиплоидности.

Согласно одному варианту реализации данного и других способов, описываемых в настоящем описании, указанный образец содержит бесклеточные фрагменты ДНК. Хотя анализ ДНК в плазме применялся для иллюстрации разных способов, описанных в настоящем описании, указанные способы можно также применять для определения ассоциированных с опухолью хромосомных aberrаций в пробах, содержащих смесь здоровой ДНК и ДНК опухолевого происхождения. Другие типы проб включают слюну, слезную жидкость, плевральную жидкость, мочу, сыворотку, панкреатический сок, кал и пробы мазков из шейки матки.

На этапе 1010 определяют первый и второй гаплотипы для здоровых клеток указанного организма в первой области хромосомы. Гаплотипы можно определять любыми подходящими способами, такими как способы, упоминаемые выше в настоящем описании. Указанную область хромосомы можно выбрать посредством любого способа, например посредством способов, описываемых в настоящем описании. Указанная первая область хромосомы включает первое множество локусов (например локусы 420 из 410 области), которые гетерозиготны. Гетерозиготные локусы (полокусное генное разнообразие) могут находиться далеко друг от друга, например, локусы могут быть на расстоянии 500 или 1000 оснований (или более) от другого локуса указанного первого множества локусов. Могут существовать другие типы локусного генного разнообразия в первой области хромосомы, но их не применяют.

На этапе 1020 множество молекул нуклеиновых кислот в указанной биологической пробе характеризуют с точки зрения локализации и аллеля каждой молекулы. Например, можно выявить локализацию молекулы нуклеиновой кислоты в эталонном геноме указанного организма. Определение локализации можно проводить разными способами, включая проведение секвенирования молекул (например, посредством универсального секвенирования), чтобы получить одну или две (с парными концами) секвенированных маркерных последовательностей указанной молекулы, а затем провести выравнивание секвенированной маркерной последовательности (последовательностей) с эталонным геномом. Такое выравнивание можно проводить при помощи таких средств как средство поиска основного локального выравнивания (BLAST). Локализацию можно идентифицировать как число в плече определенной хромосомы. Аллель в одном из гетерозиготных локусов локусы (полокусное генное разнообразие) можно применять для определения того, из какого гаплотипа произошел фрагмент.

На этапе 1030 идентифицируют первую группу молекул нуклеиновых кислот как происходящую из первого гаплотипа на основании выявленных локализаций и определенных аллелей. Например, фрагмент, включающий локус 421 на фиг. 4, содержащий аллель А, можно идентифицировать как происходящий из Нар I. Указанная первая группа может перекрывать первую область хромосомы при включении по меньшей мере одной молекулы нуклеиновой кислоты, расположенной в каждом из первого множества локусов.

На этапе 1040 идентифицируют вторую группу молекул нуклеиновых кислот как происходящую из второго гаплотипа на основании выявленных локализаций и определенных аллелей. Например, фрагмент, включающий локус 421 на фиг. 4, содержащий аллель Т, можно идентифицировать как происходящий из Нар II. Указанная вторая группа включает по меньшей мере одну молекулу нуклеиновой кислоты, расположенную в каждом из первого множества локусов.

На этапе 1050 компьютерная система рассчитывает первое значение для первой группы молекул нуклеиновых кислот. Первое значение определяет свойство молекул нуклеиновых кислот первой группы. Примеры первого значения включают подсчет маркерных последовательностей из числа молекул первой группы и распределение по размеру указанных молекул в первой группе.

На этапе 1060 компьютерная система рассчитывает второе значение для второй группы молекул нуклеиновых кислот. Второе значение определяет свойство молекул нуклеиновых кислот второй группы.

На этапе 1070 указанное первое значение сравнивают с указанным вторым значением, чтобы классифицировать, обнаруживает ли указанная первая область хромосомы делецию или амплификацию. Классификация, основанная на наличии делеции или амплификации, может дать информацию о том, содержит ли указанный организм клетки, ассоциированные с раком. Примеры сравнения включают оценку различия или отношения двух значений и сравнение результатов с одним или более пороговыми значениями, как описано в настоящем описании. Например, отношение можно сравнивать с пороговыми значениями при анализе SPRT. Пример классификации может включать положительное (т.е. увеличение числа копий или делеция определяются), отрицательное значение или неопределенное значение, а также

разные степени положительного и отрицательного значения (например, с применением целых чисел от 0 до 10 или действительных чисел от 0 до 1). Увеличение числа копий может включать простую дубликацию. При помощи такого способа можно выявлять наличие ассоциированных с раком нуклеиновых кислот, которые включают ДНК опухолевого происхождения и ДНК из новообразований, например из предшественников рака.

Е. Глубина.

Термин "глубина анализа" относится к количеству молекул, которые требуется анализировать, чтобы получить классификацию или другое определение с заданной точностью. Согласно одному варианту реализации глубину можно рассчитать на основании известной аберрации, а затем можно проводить измерение и анализ с такой глубиной. Согласно еще одному варианту реализации указанный анализ можно продолжать до тех пор, пока не удастся произвести классификацию, которую можно применять для определения степени рака (например, стадии рака или размера опухоли). Далее приведены примеры некоторых расчетов, относящихся к глубине.

Термин "отклонение" может относиться к любому различию или отношению, описываемым в настоящем описании. В качестве примера отклонение может существовать между первым и вторым значениями или может существовать отклонение параметра от порогового значения или концентрации опухоли, описываемых в настоящем описании. Если отклонение удваивается, то количество фрагментов, которые требуется измерить, снижается на  $1/4$ . В более общем случае, если отклонение увеличивается в  $N$  раз, то количество фрагментов, которые требуется измерить, составляет  $1/N^2$ . Дополнительно, если указанное отклонение снижается в  $1/N$ , то количество фрагментов, которые требуется оценивать, увеличивается в  $N^2$  раз.  $N$  может быть действительным или целым числом. Предположим, что концентрация ДНК опухоли в пробе составляет 10% (например, в плазме), и предположим, что статистически значимое различие наблюдается при секвенировании 10 млн фрагментов. Тогда скажем, например, проводится процедура обогащения такая, что теперь концентрация ДНК опухоли в пробе составляет 20%, тогда количество фрагментов, которое требуется, может быть 2500000. Таким образом, глубина может коррелировать с процентом ДНК опухоли в указанном образце.

Также на глубину может влиять объем увеличения числа копий. Для области, в которой произошло удвоение числа копий в данной области (например, 4 вместо 2 в норме), положим  $X$  равным количеству фрагментов, которые требуются для анализа. Если область содержит в четыре раза больше копий, чем в норме, то для данной области потребуется  $X/4$  фрагментов.

Ф. Пороговые значения.

Объем отклонения параметра (например, различия или отношения величин для каждого гаплотипа) от нормальных значений можно применять для проведения диагностики, как описано выше. Например, указанным отклонением может быть различие в среднем размере фрагментов одного гаплотипа области со средним размером фрагментов другого гаплотипа. Если указанное отклонение превышает определенное число (например, пороговое значение, которое определено для нормальных проб и/или областей), то выявляют делецию или амплификацию. Но может быть информативной и степень превышения над пороговым значением, что может приводить к применению многих пороговых значений, каждое из которых соответствует разной степени рака.

Например, более выраженное отклонение от нормы может указывать на то, какая стадия рака присутствует (например, 4 стадия может демонстрировать большую степень дисбаланса, чем 3 стадия). Более выраженное отклонение может также быть следствием того, что опухоль крупная и в результате выделяет много фрагментов и/или что указанная область амплифицирована многократно.

Помимо определения разных степеней рака, разные пороговые значения могут также позволить эффективно выявлять области с аберрациями или специфические области. Например, можно установить высокое пороговое значение, чтобы отследить главным образом амплификацию до трех и более раз, которая приводит к более высокому дисбалансу, чем делеция одного гаплотипа. Также можно выявлять делеции двух копий определенной области. Также можно применять более низкое пороговое значение, чтобы идентифицировать области, которые могут содержать аберрацию, а затем указанные области можно анализировать более подробно, чтобы подтвердить, действительно ли существует аберрация, и ее локализацию. Например, можно проводить бинарный поиск (или поиск более высокого порядка, такой как восьмеричное дерево), причем на более низких уровнях иерархии применять более высокие пороговые значения.

На фиг. 11 показана область 1110 с подобластью 1130, которая подвергается делеции в раковых клетках, а также измерения, сделанные в плазме для определения делетированной области согласно вариантам реализации настоящего изобретения. Область хромосомы 1110 можно выбрать при помощи любого способа, упоминаемого в настоящем описании, такого как разделение генома на сегменты одинакового размера. На фиг. 11 также показано число встречаемости аллелей для каждого из локусов 1120. Сумма нарастающим итогом также показана 1140 области (нормальная область) и области 1130 (делетированная область) соответственно.

Если для анализа выбирают 1110 область, сумма нарастающим итогом составляет 258 для Нар I и 250 для Нар II и различие составляет 18 по одиннадцати локусам. Такое различие меньше по проценту

общего числа отсчетов, чем в случае анализа только делетированной 1130 подобласти. Это следует понимать так, что в раковых клетках приблизительно половина 1110 области нормальна, а вся 1130 подобласть делетирована. Таким образом, аберрацию в 1110 области можно пропустить в зависимости от изменяемого порогового значения.

Чтобы было возможно определить делеции подобластей, можно применять варианты реализации с более низким пороговым значением для относительно крупных областей (в данном примере 1110 область предполагается относительно крупной по сравнению с размером делетированных областей, которые требуется идентифицировать).

Более низкое пороговое значение может позволить идентифицировать больше областей, которые могут включать некоторые ложноположительные результаты, но они могут снизить количество ложноотрицательных результатов. Теперь ложноположительные результаты можно устранить при помощи дальнейшего анализа, который может также указать на аберрации.

После того как определенная область помечена для дальнейшего анализа, указанную область можно разделить на подобласти для дальнейшего анализа. На фиг. 11 можно разделить одиннадцать локусов пополам (например, при помощи бинарного дерева), чтобы получить 1140 подобласть из шести локусов и 1130 подобласть из пяти локусов. Указанные области можно анализировать с тем же пороговым значением или с более строгим пороговым значением. В данном примере 1140 подобласть можно далее определить как нормальную, а 1130 подобласть определить как включающую делецию или амплификацию. Таким образом, более крупные области можно проигнорировать как не содержащие аберраций, и может быть потрачено время на дальнейший анализ подозреваемых областей (областей со значением выше более низкого порогового значения), чтобы выявить подобласти, которые обнаруживают аберрацию с высокой степенью достоверности (например, с применением более высокого порогового значения). Хотя в настоящем описании применяли RHDO, технология с применением размеров применима в той же степени.

Размер областей для поиска первого уровня (и размер подобластей более низких уровней в дереве) можно выбирать на основании размера аберрации, которую требуется выявить. Может быть показано, что типы рака демонстрируют десять областей с аберрациями длиной 10 МБ. У пациентов также имеются области 10 МБ, проявляющие аберрации. На более поздних стадиях рака могут быть более длинные отрезки аберраций.

G. Уточнение локализации аберрации в определенной области.

В последнем разделе обсуждалось подразделение области на подобласти на основании древовидного поиска. Здесь мы обсуждаем другие способы анализа подобластей указания аберраций внутри определенной области.

На фиг. 12 показано, как локализацию аберраций можно картировать при помощи анализа RHDO согласно вариантам реализации настоящего изобретения. Указанная область хромосомы показана горизонтально, а гаплотипы нераковых клеток обозначены как Нар I и Нар II. Делетированная область Нар II в раковой клетке обозначена как ПГ.

Как показано, анализ RHDO начинают с левой стороны в направлении правой стороны гипотетической 1202 области хромосомы. Каждая из стрелок отражает сегмент классификации RHDO. Каждый сегмент можно считать своей собственной областью, особенно подобласть с набором полокусного генного разнообразия более крупной области. Размер сегмента классификации RHDO зависит от количества локусов (и положений указанных локусов) до того, как определена классификация. Количество локусов, включаемых в каждый сегмент RHDO, зависит от количества молекул, анализируемых для каждого сегмента, желаемой точности (например, отношения шансов при анализе SPRT) и относительной концентрации ДНК опухолевого происхождения в пробе. Можно провести классификацию, когда количество молекул адекватно для определения того, есть ли статистически значимое различие между двумя гаплотипами, как в примерах, проиллюстрированных на фиг. 4 и 5.

Каждая из непрерывных горизонтальных стрелок представляет собой сегмент классификации RHDO, и они показывают, что в пробе ДНК дисбаланс дозировки гаплотипов отсутствует. В указанной области без ПГ в опухоли проводится шесть классификаций RHDO, и каждая из них указывает на отсутствие дисбаланса дозировки гаплотипов. Следующий сегмент классификации RHDO 1210 пересекает соединение 1205 между областями с ПГ и без ПГ. В нижней части фиг. 12 показана кривая SPRT для сегмента RHDO 1210. Черные вертикальные стрелки показывают соединение между областями с ПГ и без ПГ. При накоплении возрастающего количества данных из области с ПГ классификация RHDO данного сегмента указывает на наличие дисбаланса дозировки гаплотипов.

Каждая из белых горизонтальных стрелок представляет собой сегмент классификации RHDO, который показывает, что в пробе ДНК присутствует дисбаланс дозировки гаплотипов. Последующие четыре RHDO с правой стороны также указывают на наличие дисбаланса дозировки гаплотипов в указанной пробе ДНК. Можно вывести, что соединение между областями с ПГ и без ПГ находится внутри первого сегмента RHDO, что указывает на изменение в классификации RHDO, а именно с наличия на отсутствие дисбаланса дозировки гаплотипов или наоборот.

На фиг. 13 показана классификация на основе RHDO, начатая с другого направления, согласно вариантам реализации настоящего изобретения. На фиг. 13 показана классификация на основе RHDO с

обоих направлений. На основании анализа RHDO, который начинали с левой стороны, можно заключить, что соединение между областями с ПГ и без ПГ находится внутри первого сегмента RHDO, что указывает на наличие дисбаланса дозировки гаплотипов. На основании анализа RHDO, который начинали с правой стороны, можно заключить, что указанное соединение находится внутри первого сегмента RHDO 1320, что указывает на отсутствие дисбаланса дозировки гаплотипов. Объединяя информацию, полученную при анализе RHDO, проводимого в обоих направлениях, можно сделать заключение о локализации соединения между областями 1330 с ПГ и без ПГ.

#### IV. Неспецифическое выявление аберрации на основании гаплотипа.

Способ RHDO основан на применении гетерозиготных локусов. Теперь, хромосомы диплоидного организма будут обладать некоторыми различиями, что приводит к образованию двух гаплотипов, но количество гетерозиготных локусов может варьировать. У некоторых индивидуумов может быть относительно мало гетерозиготных локусов. Вариант реализации, описываемый в настоящем разделе, можно также применять для локусов, которые гомозиготны, путем сравнения двух областей, а не гаплотипов одной и той же области. Таким образом, можно получить больше экспериментальных точек, хотя могут существовать некоторые недостатки из-за сравнения двух разных областей хромосом.

Согласно способу дозировки областей хромосом количество фрагментов из одной области хромосомы (например, определяемое путем подсчета маркерных последовательностей, выравнивающихся с данной областью) сравнивают с расчетным значением (которое может быть получено для эталонной области хромосомы или для той же области в другом образце, полученной от здорового субъекта). Таким образом, фрагмент можно подсчитывать для определенной области хромосомы, независимо от того, из какого гаплотипа происходит секвенированная маркерная последовательность. Таким образом, по-прежнему можно применять секвенированные маркерные последовательности, которые не содержат полусеквенированного генного разнообразия. Чтобы провести сравнение, в одном варианте реализации можно нормировать количество маркерных последовательностей перед сравнением. Каждая область определяется по меньшей мере по двум локусам (которые отделены друг от друга), и фрагменты в указанных локусах можно применять для получения совокупного значения для указанной области.

Нормированное значение для секвенированных считываний (маркерных последовательностей) для конкретной области можно рассчитать путем деления числа секвенированных считываний, выравнивающихся с данной областью, на общее число секвенированных считываний, которые могут выравниваться со всем геномом. Такое нормированное количество маркерных последовательностей позволяет результаты одной пробы сравнивать с результатами другого образца. Например, нормированное значение может быть долей (например, процентом или частью) секвенированных считываний, предположительно происходящих из конкретной области, как указывалось выше. Но возможны многие другие варианты нормировки, что должно быть понятно специалистам в данной области техники. Например, можно проводить нормировку путем деления количества подсчетов для одной области на количества подсчетов для эталонной области (в случае, описанном выше, эталонной областью является просто весь геном). Такое нормированное количество маркерных последовательностей затем можно сравнить с пороговым значением, которое можно определить из одного или более эталонных образцов, не демонстрирующих рак.

Нормированный подсчет маркерных последовательностей в исследуемом случае можно затем сравнивать с нормированным подсчетом маркерных последовательностей одного или более эталонных субъектов, например субъектов, не страдающих раковыми заболеваниями. Согласно одному варианту реализации, сравнение проводят путем расчета z-показателя случая для конкретной области хромосомы, z-показатель рассчитывают по следующему уравнению:

$$z\text{-показатель} = (\text{нормированный подсчет маркерных последовательностей случая} - \text{среднее}) / \text{С.О.},$$

где "среднее" означает среднюю нормированную встречаемость маркерных последовательностей, выравнивающихся с конкретной областью хромосомы для эталонных проб; а С.О. обозначает стандартное отклонение количества нормированных встречаемостей маркерных последовательностей, выравнивающихся с конкретной областью для эталонных проб. Следовательно, z-показатель представляет собой значение стандартного отклонения, на которое нормированный подсчет маркерных последовательностей области хромосомы для исследуемого случая отклоняется от среднего нормированного подсчета маркерных последовательностей для той же области хромосомы у одного или более эталонных субъектов.

В ситуации, когда исследуемый организм болен раком, области хромосом, которые амплифицированы в тканях опухоли, могут присутствовать в избытке в ДНК плазмы. Это может приводить к тому, что значение z-показателя будет положительным. С другой стороны, области хромосом, которые делетированы в тканях опухоли, могут присутствовать в недостатке в ДНК плазмы. Это может приводить к тому, что значение z-показателя будет отрицательным. Величина z-показателя определяется несколькими факторами.

Одним фактором является относительная концентрация ДНК опухолевого происхождения в указанной биологической пробе (например, плазме). Чем выше относительная концентрация ДНК опухолевого происхождения в указанной биологической пробе (например, плазме), тем больше может быть различие между нормированной встречаемостью маркерных последовательностей исследуемого случая и эталонными случаями. Следовательно, может быть получена более высокая амплитуда z-показателя.

Еще одним фактором является разброс нормированной встречаемости маркерных последовательно-

стей в одном или более эталонных случаях. При одной и той же степени присутствия в избытке области хромосомы в указанном биологическом образце (например, плазме) в исследуемом случае, меньший разброс (т.е. меньшее стандартное отклонение) нормированной встречаемости маркерных последовательностей в эталонной группе может приводить к более высокому z-показателю. Аналогично при одной и той же степени присутствия в недостатке области хромосомы в указанной биологической пробе (например, плазме) в исследуемом случае меньшее стандартное отклонение нормированной встречаемости маркерных последовательностей в эталонной группе может приводить к более отрицательному z-показателю.

Еще одним фактором является объем хромосомной аберрации в тканях опухоли. Термин "объем хромосомной аберрации" относится к изменению числа копий для конкретной области хромосомы (либо добавление, либо утрата). Чем больше изменения числа копий в тканях опухоли, тем выше может быть степень недостаточного или избыточного присутствия конкретной области хромосомы в ДНК плазмы. Например, утрата обеих копий указанной хромосомы может приводить к присутствию в более выраженном недостатке указанной области хромосомы в ДНК плазмы по сравнению с утратой одной из двух копий указанной хромосомы, и, следовательно, приводит к более отрицательному z-показателю. Обычно при раке имеют место множественные хромосомные аберрации. Хромосомные аберрации при каждом типе рака могут также варьировать по своей природе (т.е. амплификации или делеции), степени (добавление или утрата одной или нескольких копий) и по объему (размер аберрации в выражении длины хромосомы).

На точность измерения нормированного подсчета маркерных последовательностей влияет количество анализируемых молекул. Мы предполагаем, что для выявления хромосомных аберраций с изменением одной копии (либо добавлением, либо утратой) может потребоваться 15000, 60000 и 240000 молекул при относительной концентрации приблизительно 12,5%, 6,3% и 3,2% соответственно. Более подробная информация о подсчете маркерных последовательностей для выявления рака для разных областей хромосом представлена в публикации патента США № 2009/0029377 под названием "Diagnosing Fetal Chromosomal Aneuploidy Using Massively Parallel Genomic Sequencing", Lo et al., полное содержание которой включено в настоящее описание посредством ссылки для всех целей.

Также варианты реализации могут быть основаны на применении анализа по размерам, взамен способа подсчета маркерных последовательностей. Анализ по размерам можно также применять вместо подсчета маркерных последовательностей. При анализе по размерам можно применять разные параметры, упоминаемые в настоящем описании, и в заявке на патент США № 12/940992. Например, можно применять значения Q или F, упоминаемые выше. Такие значения, характеризующие размеры, не требуют нормировки на подсчет других областей, поскольку указанные значения не шкалируются в зависимости от числа считываний. Также для неспецифических способов можно применять технологии гаплотип-специфических способов. Например, можно применять технологии, основанные на глубине и уточнении области. Согласно некоторым вариантам реализации при необходимости можно учитывать GC-смещение, когда сравнивают две области. Поскольку при способе RHDO применяют одну и ту же область, такая коррекция не требуется.

#### V. Множество областей.

Хотя могут существовать определенные типы рака с аберрациями в конкретных областях хромосом, такие типы рака не всегда проявляются в одних и тех же областях. Например, аберрации могут демонстрировать дополнительные области хромосом, и локализация таких дополнительных областей может быть неизвестна. Кроме того, когда проводят скрининг пациентов с целью идентификации ранних стадий рака, может быть желательно идентифицировать широкий диапазон типов рака, которые могут демонстрировать аберрации, присутствующие по всему геному. Чтобы справиться с такими ситуациями, варианты реализации могут быть основаны на анализе множества областей систематическим образом, чтобы определить, какие области демонстрируют аберрации. Например, можно применять количество аберраций и их локализацию (т.е. примыкают ли они друг к другу) для подтверждения аберраций, определения стадии рака, проведения диагностики рака (т.е. если количество больше, чем пороговое значение) и прогнозирования на основании количества и локализации разных областей, проявляющих аберрации.

Соответственно варианты реализации позволяют выявить наличие рака в организме на основании некоторого количества областей, которые демонстрируют аберрации. Таким образом, можно тестировать множество областей (например, 3000), чтобы выявить некоторое количество областей, проявляющих аберрацию. Указанные области могут охватывать весь геном или только части генома, например непотворяющуюся область.

На фиг. 14 показана блок-схема способа 1400 анализа биологического образца определенного организма с применением множества областей хромосом согласно вариантам реализации настоящего изобретения. Указанный биологический образец включает молекулы нуклеиновых кислот (также называемые фрагментами).

На этапе 1410 идентифицируют множество неперекрывающихся областей хромосом указанного организма. Каждая область хромосомы включает множество локусов. Как упоминалось выше, область может обладать размером 1 Мб или каким-либо другим одинаковым размером. Тогда весь геном может включать приблизительно 3000 областей, каждая из которых обладает заранее определенным размером и локализацией. Также как упоминалось выше, такие заранее определенные области могут варьировать по

длине, в зависимости от длины конкретной хромосомы или заданного количества областей, которое следует применять, и любых других критериев, упоминаемых в настоящем описании. Если области обладают разной длиной, такие значения длины можно применять для нормировки результатов, например, как описано в настоящем описании.

На этапе 1420 идентифицируют локализацию указанной молекулы нуклеиновой кислоты в эталонном геноме человека для каждой из множества молекул нуклеиновых кислот. Локализацию можно определить при помощи любого из способов, упоминаемых в настоящем описании, например, путем секвенирования фрагментов для получения секвенированных маркерных последовательностей и выравнивания указанных секвенированных маркерных последовательностей с эталонным геномом. Также для гаплотип-специфичных способов можно определить конкретный гаплотип молекулы.

Для каждой из указанных областей хромосом проводят этапы 1430-1450. На этапе 1430 соответствующую группу молекул нуклеиновых кислот определяют как происходящую из указанной области хромосомы на основании выявленных локализаций. Соответствующая группа включает по меньшей мере одну молекулу нуклеиновой кислоты, локализованную в каждом из множества локусов указанной области хромосомы. Согласно одному варианту реализации реализации указанной группой могут быть фрагменты, которые выравниваются с конкретным гаплотипом указанной области хромосомы, например в способе RHDO, описанном выше. Согласно другому варианту реализации группой могут быть фрагмент, который выравнивается с указанной областью хромосомы, как в способе, описанном в разделе IV.

На этапе 1440 компьютерная система рассчитывает соответствующее значение соответствующей группы молекул нуклеиновых кислот. Указанное соответствующее значение определяет свойство молекул нуклеиновых кислот соответствующей группы. Указанное соответствующее значение может представлять собой любое из значений, упоминаемых в настоящем описании. Например, указанное значение может представлять собой количество фрагментов в указанной группе или статистический показатель распределения по размеру фрагментов в указанной группе. Указанное соответствующее значение также может представлять собой нормированное значение, например встречаемость маркерных последовательностей, поделенной на общее количество встречаемости маркерных последовательностей для указанной пробы или количество встречаемости маркерных последовательностей для эталонной области. Указанное соответствующее значение также может представлять собой разницу или отношение к другому значению (например, при RHDO), что позволяет определить свойство различия для указанной области.

На этапе 1450 указанное соответствующее значение сравнивают с эталонным значением и определяют классификацию или определяют, демонстрирует ли указанная первая область хромосомы делецию или амплификацию. Затем определяют количество областей хромосомы, классифицированных как демонстрирующие делецию или амплификацию. Указанное эталонное значение может быть любым пороговым или эталонным значением, обсуждаемым в настоящем описании. Например, указанное эталонное значение может быть пороговым значением, определенным для нормальных проб. Для RHDO указанное соответствующее значение может быть различием или отношением встречаемости маркерных последовательностей для двух гаплотипов, а указанное эталонное значение может быть пороговым значением для определения того, существует ли статистически значимое отклонение. В качестве еще одного примера указанное эталонное значение может быть встречаемостью маркерных последовательностей или показателем размера для другого гаплотипа или области, а сравнение может включать вычисление различия или отношения (или функции от них) и последующее определение того, превышает ли указанное различие или отношение пороговое значение.

Указанное эталонное значение может варьировать в зависимости от результатов других областей. Например, если соседние области также демонстрируют отклонение (хотя и малое по сравнению с одним пороговым значением, например, z-показатель равен 3), можно применять более низкое пороговое значение. Например, если три идущих друг за другом области демонстрируют значение выше первого порогового значения, вероятность наличия рака может быть более высокой. Таким образом, указанное первое пороговое значение может быть ниже, чем другое пороговое значение, которое требуется для идентификации рака на основании не идущих друг за другом областей. Три области (или более трех), обладающих даже малым отклонением, могут иметь низкую достаточную вероятность случайного эффекта, что может сохранить чувствительность и специфичность.

На этапе 1460 определяют количество областей хромосом, классифицированных как демонстрирующие делецию или амплификацию. Области хромосом, которые подсчитываются, могут иметь ограничения. Например, могут быть подсчитаны только области, которые соседствуют по меньшей мере с одной другой областью (может требоваться, чтобы смежные области обладали определенным размером, например 4 или более области). Для тех вариантов реализации, в которых области не равны, число может также обуславливать соответствующие длины (например, число может быть общей длиной аберрантных областей).

На этапе 1470 указанное количество сравнивают с пороговым количеством и определяют классификацию указанной пробы. В качестве примера, классификация может представлять собой определение наличия рака у указанного организма, стадию рака и прогнозирование состояния при раке. Согласно одному варианту реализации подсчитывают все аберрантные области и применяют одно пороговое значе-

ние, независимо от того, где указанные области наблюдаются. Согласно другому варианту реализации пороговое значение может варьировать в зависимости от локализации и размера областей, которые подсчитываются. Например, количество областей конкретной хромосомы или плеча хромосомы можно сравнивать с пороговым значением для данной конкретной хромосомы (или плеча). Можно применять несколько пороговых значений. Например, количество aberrантных областей в конкретной хромосоме (или плече) обязательно должно быть выше, чем первое пороговое значение, а общее количество aberrантных областей в геноме обязательно должно быть выше, чем второе пороговое значение.

Данное указанное пороговое значение для количества областей также может зависеть от того, насколько сильный дисбаланс имеет место для подсчитываемых областей. Например, количество областей, которое применяют в качестве порогового значения для классификации рака, может зависеть от специфичности и чувствительности (aberrантное пороговое значение), применяемых при выявлении aberrации в каждой области. Например, если aberrантное пороговое значение низкое (например, z-показатель равен 2), тогда пороговое значение количества можно выбирать высоким (например, 150). Но если aberrантное пороговое значение высокое (например, z-показатель равен 3), тогда пороговое значение количества может быть ниже (например, 50). Количество областей, демонстрирующих aberrацию, также может быть взвешенным значением, например одна область, которая демонстрирует высокий дисбаланс, может иметь более высокий весовой коэффициент, чем область, которая демонстрирует лишь малый дисбаланс (т.е. существует больше классификаций, чем просто положительная и отрицательная классификация наличия aberrации).

Соответственно количество (которое может включать количество или размер) областей хромосом, демонстрирующих значительный избыток или недостаток нормированной встречаемости маркерных последовательностей (или другого соответствующего значения для оценки свойства указанной группы) можно применять для оценки тяжести заболевания. Количество областей хромосом с aberrантной нормированной встречаемостью маркерных последовательностей может определяться двумя факторами, а именно числом (или размером) хромосомных aberrаций в тканях опухоли и относительной концентрацией ДНК опухолевого происхождения в биологической пробе (например, плазме). Более распространенные типы рака, как правило, обнаруживают больше (и более крупные) хромосомных aberrаций. Следовательно, в указанной пробе (например, плазме) потенциально может определяться больше ассоциированных с раком хромосомных aberrаций. У пациентов с более распространенным раком более высокая опухолевая нагрузка может приводить к более высокой относительной концентрации ДНК опухолевого происхождения в плазме. Как следствие, ассоциированные с опухолью хромосомные aberrации может быть легче определить в пробе плазмы.

В контексте скрининга или выявления рака количество областей хромосом, проявляющих избыток или недостаток нормированной встречаемости маркерных последовательностей (или другого значения), можно применять для определения вероятности того, что исследуемый субъект болен раком. Если применять уровень отсечки  $\pm 2$  (т.е. z-показатель  $> 2$  или  $< -2$ ), приблизительно 5% тестируемых областей предположительно дадут значительное отклонение z-показателя от среднего для контрольных субъектов просто из-за вероятности. Когда весь геном делят на сегменты 1 Мб, во всем геноме будет приблизительно 3000 сегментов. Таким образом, приблизительно 150 сегментов предположительно будут демонстрировать z-показатель  $> 2$  или  $< -2$ .

Таким образом, уровень отсечки (пороговое значение) 150 для числа сегментов с z-показателем  $> 2$  или  $< -2$  можно применять для определения наличия ракового заболевания. Для удовлетворения диагностических целей можно выбрать другие уровни отсечки для числа сегментов с aberrантным z-показателем (например, 100, 125, 175, 200, 250 и 300). Более низкий уровень отсечки, например 100, может приводить к более высокой чувствительности оценки, но к более низкой специфичности, а более высокий уровень отсечки может приводить к большей специфичности, но меньшей чувствительности. Число ложноположительных классификаций можно снизить путем повышения уровней отсечки z-показателя. Например, если уровень отсечки увеличить до 3, то только 0,3% сегментов будут ложноположительными. В данной ситуации для выявления наличия рака можно применять более 3 сегментов с aberrантным z-показателем. Можно выбирать другие уровни отсечки, например 1, 2, 4, 5, 10, 20 и 30, для удовлетворения разных диагностических целей. Однако чувствительность выявления ассоциированных с раком хромосомных aberrаций можно снизить при увеличении количества aberrантных сегментов, требуемых для постановки диагноза.

Один из возможных способов повышения чувствительности без ухудшения специфичности состоит в том, чтобы учитывать результат соседнего сегмента хромосомы. Согласно одному варианту реализации уровень отсечки для z-показателя остается  $> 2$  и  $< -2$ . Однако область хромосомы можно классифицировать как потенциально aberrантную только тогда, когда два идущих друг за другом сегмента будут демонстрировать aberrации одного и того же типа, например оба сегмента демонстрируют z-показатель  $> 2$ . Если отклонение в нормированной встречаемости маркерных последовательностей является случайной погрешностью, вероятность того, что два идущих друг за другом сегмента будут ложноположительными в одном и том же направлении, будет равна  $0,125\%$  ( $5\% \times 5\% / 2$ ). С другой стороны, если хромосомная

аберрация охватывает два идущих друг за другом сегмента, более низкий уровень отсечки сделает выявление избыточного или недостаточного присутствия указанных сегментов в пробе плазмы более чувствительным. Поскольку отклонение нормированной встречаемости маркерных последовательностей (или другого значения) от среднего значения для контрольных субъектов не связано со случайной погрешностью, последовательные требования классификации не окажут значительного неблагоприятного действия на чувствительность. Согласно другому варианту реализации вместе с применением более высокого уровня отсечки можно добавлять z-показатель соседних сегментов. Например, z-показатель трех идущих друг за другом сегментов можно суммировать и можно применить уровень отсечки 5. Данную концепцию можно распространить на более чем три идущих друг за другом сегмента.

Сочетание количества и абберрантных пороговых значений может также зависеть от цели анализа и любого предшествующего знания об организме (или его отсутствия). Например, если проводят скрининг нормальной здоровой популяции для определения наличия рака, то, как правило, применяют высокую специфичность, потенциально как в отношении количества областей (т.е. высокое пороговое значение для количества областей), так и в отношении порогового значения, когда область определяют как обладающую абберрацией. Однако у пациентов с более высоким риском (например, пациент, жалующийся на новообразование или с отягощенным семейным анамнезом, курильщик, вирус ВИЧ, вирус гепатита или другие вирусы) пороговые значения можно снизить, чтобы чувствительность была более высокой (меньше ложноотрицательных результатов).

Согласно одному варианту реализации, если применяют разрешение 1 Мб и более низкий предел выявления 6,3% ДНК опухолевого происхождения для выявления хромосомной абберрации, может потребоваться число молекул 60000 в каждом сегменте 1 Мб. Это можно перевести приблизительно в 180 млн (60000 считываний/Мб×3000 Мб) выравниваемых считываний на весь геном.

На фиг. 15 показана таблица, иллюстрирующая глубину, требуемую для разного числа сегментов и относительной концентрации фрагментов опухолевого происхождения согласно вариантам реализации настоящего изобретения. В столбце 1510 приведена концентрация фрагментов из клеток опухоли в указанной пробе. Чем выше указанная концентрация, тем легче выявить указанную абберрацию, тем меньше молекул требуется для анализа. В столбце 1520 приведено расчетное требуемое число молекул на один сегмент, которое можно рассчитать при помощи способа, описанного выше относительно глубины.

Более маленький размер сегмента может приводить к более высокому разрешению выявления более маленьких хромосомных абберраций. Однако он может повышать требования к количеству молекул, которые требуется анализировать суммарно. Более крупный размер сегмента может снижать требования к количеству молекул для анализа ценой уменьшения разрешения. Следовательно, можно выявлять только более крупные абберрации. Согласно одному варианту реализации более крупные области, сегменты, демонстрирующие абберрацию, можно разделить и указанные подобласти анализировать, получая более высокое разрешение (например, как описано выше). В столбце 1530 приведен размер каждого сегмента. Чем меньше значение, тем больше областей применяется. В столбце 1530 приведено число молекул, которое требуется анализировать для всего генома. Соответственно, если есть оценка (или минимальная концентрация для выявления), можно определить число молекул для анализа.

#### VI. Прогресс во времени.

По мере того как опухоль прогрессирует, количество опухолевых фрагментов может увеличиваться, поскольку опухоль будет выделять больше фрагментов ДНК (например, из-за роста опухоли, расширяющегося некроза или большей васкуляризации). Большее количество фрагментов ДНК из тканей опухоли в плазме будет увеличивать степень дисбаланса в плазме (например, будет увеличиваться различие во встречаемости маркерных последовательностей между двумя гаплотипами при RHD0). Кроме того, поскольку количество фрагментов опухолевого происхождения увеличивается, количество областей, где существует абберрация, можно определять легче. Например, когда опухоль мала и выделяет малое количество фрагментов раковой ДНК, количество ДНК опухолевого происхождения для определенной области может быть столь мало, что абберрацию выявить не удастся, поскольку нельзя достичь статистически значимого различия, поскольку недостаточно фрагментов для анализа. Даже когда опухоль мала, можно анализировать больше фрагментов, но для этого может потребоваться большой образец (например, большое количество плазмы).

При отслеживании прогресса рака можно применять объем абберрации в одной или более областях (например, по дисбалансу или требуемой глубине) или объем (количество или размер) областей хромосом, проявляющих абберрацию. Согласно одному примеру, если объем абберрации одной области (или нескольких областей) увеличивается быстрее, чем объем абберраций других областей, то такую область(и) можно применять в качестве предпочтительного маркера для отслеживания рака. Такое увеличение может быть следствием того, что опухоль стала больше и, таким образом, выделяет много фрагментов и/или что указанная область многократно амплифицирована. Можно также отслеживать показатель абберрации (например, объем абберрации, или количество областей, демонстрирующих абберрации, или их сочетание) после операции для подтверждения того, что опухоль была удалена должным образом.

Согласно разным вариантам применения указанной технологии определение относительной концентрации ДНК опухолевого происхождения применяют для оценки стадии, прогноза или отслеживания

прогресса рака. Оценка прогресса может предоставить информацию о текущей стадии рака и о том, как быстро опухоль растет или распространяется. Термин "стадия" рака относится ко всем или части из следующих пунктов: размер опухоли, гистологическое проявление, наличие/отсутствие вовлечения лимфатических узлов, а также наличие/отсутствие отдаленных метастазов. "Прогнозирование" рака подразумевает оценку вероятности прогрессирования заболевания и/или вероятность выживания при раке. Также оно может подразумевать оценку времени, в течение которого у пациента не будет клинического прогрессирования или длительности выживания. Под "мониторингом" рака можно понимать отслеживание с целью определить, прогрессирует ли заболевание (например, увеличение в размере опухоли, увеличение вовлечения лимфатических узлов или распространение на отделанные органы, т.е. метастазирование). Также мониторинг может подразумевать отслеживание того, удалось ли контролировать опухоль посредством лечения. Например, если лечение эффективно, то можно наблюдать уменьшение размера опухоли, регрессию метастазов или вовлечения лимфатических узлов, улучшение общего состояния пациента (например, повышение массы тела).

А. Определение относительной концентрации ДНК опухолевого происхождения.

Один из способов отслеживания объема увеличения аберрации для одной или более областей состоит в определении относительной концентрации раковой ДНК для определенной области(ей). Изменение относительной концентрации раковой ДНК затем можно применять для отслеживания указанной опухоли во времени. Такое отслеживание можно применять для диагностики, например первое измерение может дать информацию о фоновом уровне (который может соответствовать общему уровню аберрации у субъекта), а более поздние измерения могут показать изменения, которые могут указывать на рост опухоли (то есть, рака). Изменения относительной концентрации раковой ДНК также можно применять для прогнозирования, насколько хорошо пройдет лечение. Согласно другим вариантам применения данной технологии увеличение относительной концентрации ДНК опухолевого происхождения в плазме может указывать на плохой прогноз или увеличение опухолевой нагрузки у указанного пациента.

Относительную концентрацию раковой ДНК можно определять разными способами. Например, различие во встречаемости маркерных последовательностей из одного гаплотипа по сравнению с другим (или одной области по сравнению с другой). Другой способ основан на определении глубины (т.е. числа анализируемых фрагментов) до выявления статистически значимого различия. Для более раннего примера для определения относительной концентрации ДНК опухолевого происхождения в биологическом образце (например, плазме) путем анализа областей хромосомы с потерей гетерозиготности можно применять различие в дозировке гаплотипов.

Было показано, что количество ДНК опухолевого происхождения положительно коррелирует с опухолевой нагрузкой у пациентов с раком (Lo et al. *Cancer Res.* 1999;59:5452-5. and Chan et al. *Clin Chem.* 2005;51:2192-5). Следовательно, для отслеживания прогрессирования заболевания пациента можно применять последовательный мониторинг относительной концентрации ДНК опухолевого происхождения в биологических образцах (например, образцах плазмы) посредством анализа RHDO. Например, мониторинг относительной концентрации ДНК опухолевого происхождения в отбираемых последовательно биологических образцах (например, плазмы) после лечения можно применять для определения успеха лечения.

На фиг. 16 показан принцип измерения относительной концентрации ДНК опухолевого происхождения в плазме на основании анализа RHDO согласно вариантам реализации настоящего изобретения. Определяют дисбаланс между двумя гаплотипами, и степень дисбаланса можно применять для определения относительной концентрации ДНК опухолевого происхождения в указанном образце.

Нар I и Нар II представляют собой два гаплотипа в неопухолевых тканях. Нар II частично делетирован в тканях опухоли в подобласти 1610. Следовательно, Нар II ассоциированные фрагменты, соответствующие делетированной 1610 области, которые определяют в плазме, обусловлены неопухолевыми тканями. С одной стороны, 1610 область в Нар I присутствует как в опухолевых, так в неопухолевых тканях. Следовательно, различие между встречаемостью считываний Нар I и Нар II может представлять собой количество ДНК опухолевого происхождения в плазме.

Относительную концентрацию ДНК опухолевого происхождения (F) можно рассчитать на основании числа секвенированных считываний (маркерных последовательностей) из подвергшихся и не подвергшихся делеции хромосом для областей хромосом, охваченных ПГ, по следующей формуле:

$$F = (N_{\text{Нар I}} - N_{\text{Нар II}}) / N_{\text{Нар I}} \times 100\%$$

где  $N_{\text{Нар I}}$  представляет собой число секвенированных считываний, соответствующих аллелям в Нар I для гетерозиготных SNP, расположенных в областях хромосом, охваченных ПГ; а  $N_{\text{Нар II}}$  представляет собой число секвенированных считываний, соответствующих аллелям в Нар II для гетерозиготных SNP, расположенных в 1610 области хромосомы, охваченной ПГ.

Приведенная выше формула аналогична для определения p как накопленной встречаемости маркерных последовательностей для гетерозиготных локусов, расположенных в областях хромосом, не включающих делецию (Нар I) и для определения q как накопленной встречаемости маркерных последовательностей для области хромосомы, включающей делецию (Нар II) 1610, причем относительная кон-

центрация опухолевой ДНК в указанном образце (F) рассчитывается как  $F=1-q/p$ . Для примера, проиллюстрированного на фиг. 11, относительная концентрация ДНК опухолевого происхождения составляет 14% ( $1-104/121$ ).

Относительную концентрацию ДНК опухолевого происхождения в пробах плазмы пациентов с НСС определяли до резекции опухоли и после нее. До резекции опухоли  $N_{\text{НарI}}$  для первого гаплотипа заданной области хромосомы составляло 30443, а  $N_{\text{НарII}}$  для первого гаплотипа указанной области хромосомы составляло 16221, тогда F составляет 46,7%. После резекции опухоли  $N_{\text{НарI}}$  составило 31534, а  $N_{\text{НарII}}$  составило 31089, тогда F составляет 1,4%. Данный мониторинг показывает, что резекция опухоли была успешной.

Также для определения относительной концентрации можно применять степень изменения профиля размеров циркулирующей ДНК. Согласно одному варианту применения можно применять точное распределение по размеру ДНК плазмы, происходящей как из опухолевых, так из неопухолевых тканей, и затем измеренное распределение по размерам, попадающее между двумя известными распределениями, может дать относительную концентрацию (например, с применением линейной модели между двумя статистическими показателями распределений по размеру опухолевых и неопухолевых тканей). В качестве альтернативы можно применять последовательный мониторинг изменений размеров. Согласно одному аспекту определяют изменение распределения по размеру как пропорциональное относительной концентрации ДНК опухолевого происхождения в плазме.

Аналогичным образом можно также применять различие между разными областями, т.е. способы неспецифического выявления гаплотипа, описываемые выше. Согласно способам подсчета маркерных последовательностей для мониторинга прогрессирования заболевания можно применять несколько параметров. Например, можно применять величину z-показателя для областей, проявляющих хромосомные aberrации, которая отражает относительную концентрацию ДНК опухолевого происхождения в биологическом образце (например, плазме). Степень присутствия в избытке или недостатке конкретной области пропорциональна относительной концентрации ДНК опухолевого происхождения в образце и степени изменения или количеству числа копий в тканях опухоли. Величина z-показателя представляет собой меру степени присутствия в избытке или недостатке конкретной области хромосомы в пробе по сравнению с контрольными субъектами. Следовательно, величина z-показателя может соответствовать относительной концентрации ДНК опухолевого происхождения в пробе и, следовательно, опухолевой нагрузке у пациента.

В. Отслеживание количества областей.

Как упоминалось выше, количество областей, проявляющих хромосомную aberrацию, можно применять для скрининга на рак, а также его можно применять для мониторинга и составления прогноза. В качестве примера мониторинг можно применять для определения текущей стадии рака, если рак появляется вновь, и если действовало лечение. По мере того как опухоль прогрессирует, геномный состав опухоли будет деградировать еще больше. Для определения такой продолжающейся деградации можно применять способы отслеживания некоторых областей (например, заранее определенных областей длиной 1 Мб), чтобы определить прогрессирование опухоли. Тогда опухоли на более поздних стадиях будут содержать больше областей, обнаруживающих aberrации.

С. Способ.

На фиг. 17 показана блок-схема, иллюстрирующая способ определения прогресса хромосомных aberrаций в организме с применением биологических проб, включающих молекулы нуклеиновых кислот согласно вариантам реализации настоящего изобретения.

Согласно одному варианту реализации, по меньшей мере, некоторые молекулы нуклеиновых кислот находятся в бесклеточной форме. В качестве примера указанные хромосомные aberrации могут происходить из злокачественных опухолей или предраковых поражений. Также увеличение aberrаций может быть связано с тем, что в указанном организме со временем становится все больше клеток, содержащих хромосомные aberrации, или с тем, что в указанном организме определенная доля клеток содержит повышенное количество aberrаций на клетку. В качестве примера снижения, лечение (например, операция или химиотерапия) может вызывать удаление или уменьшение числа клеток, ассоциированных с раком.

На этапе 1710 идентифицируют одну или более неперекрывающихся областей хромосом в указанном организме. Каждая область хромосомы включает множество локусов. Указанные области можно выявить при помощи подходящего способа, например способа, описываемого в настоящем описании.

Этапы 1720-1750 проводят для каждого из множества моментов времени. Каждый момент времени соответствует разному времени отбора образца из организма. Текущий образец - это образец, анализируемый для данного периода времени. Например, образец можно отбирать раз в месяц в течение 6 месяцев и анализ можно проводить вскоре после отбора указанного образца. В качестве альтернативы анализ можно проводить после того, как будут сделаны несколько измерений за несколько периодов времени.

На этапе 1720 анализируют текущий биологический образец из указанного организма и определяют локализацию молекулы нуклеиновой кислоты в эталонном геноме указанного организма. Локализацию можно определять при помощи любого из способов, упоминаемых в настоящем описании, например путем секвенирования фрагментов с целью получения секвенированных маркерных последовательностей и

выравнивания указанных секвенированных маркерных последовательностей с эталонным геномом. Также для гаплотип-специфичных способов можно определять конкретный гаплотип молекулы.

Этапы 1730-1750 проводят для каждой из одной или более областей хромосом. Когда применяют множество областей, можно применять варианты реализации из раздела V. На этапе 1730 идентифицируют соответствующую группу молекул нуклеиновых кислот как происходящую из указанной области хромосомы, на основании идентифицированных локализаций. Соответствующая группа включает по меньшей мере одну молекулу нуклеиновой кислоты, локализованную в каждом из множества локусов указанной области хромосомы. Согласно одному варианту реализации указанная группа может быть представлена фрагментами, которые выравниваются с конкретным гаплотипом указанной области хромосомы, например, как в способе RHDO, описанном выше. Согласно другому варианту реализации указанная группа может быть представлена любым фрагментом, который выравнивается с указанной областью хромосомы, как в способах, описанных в разделе V.

На этапе 1740 компьютерная система рассчитывает соответствующее значение для соответствующей группы молекул нуклеиновых кислот. Указанное соответствующее значение определяет свойство молекул нуклеиновых кислот соответствующей группы. Указанное соответствующее значение может быть любым из значений, упоминаемых в настоящем описании. Например, указанное значение может быть количеством фрагментов в указанной группе или статистическим показателем распределения по размеру фрагментов в указанной группе. Указанное соответствующее значение также может быть нормированным значением, например встречаемостью маркерных последовательностей, поделенной на общее количество встречаемости маркерных последовательностей для указанной пробы или количество встречаемости маркерных последовательностей для эталонной области. Указанное соответствующее значение также может быть различием или отношением к другому значению (например, при RHDO), что позволяет определить свойство различия для указанной области.

На этапе 1750 указанное соответствующее значение сравнивают с эталонным значением и определяют классификацию или определяют, демонстрирует ли указанная область хромосомы делецию или амплификацию. Указанное эталонное значение может быть любым пороговым или эталонным значением, описываемым в настоящем описании. Например, указанное эталонное значение может быть пороговым значением, определенным для нормальных проб. Для RHDO указанное соответствующее значение может быть различием или отношением встречаемости маркерных последовательностей для двух гаплотипов, а указанное эталонное значение может быть пороговым значением для определения того, существует ли статистически значимое отклонение. В качестве еще одного примера указанное эталонное значение может быть встречаемостью маркерных последовательностей или показателем размера для другого гаплотипа или области, а сравнение может включать вычисление различия или отношения (или функции от них) и последующее определение того, превышает ли указанное различие или отношение пороговое значение. Указанное эталонное значение можно определять согласно любому подходящему способу и критерию, например, описанному в настоящем описании.

На этапе 1760 указанную классификацию каждой из областей хромосом в указанном множестве моментов времени применяют для определения прогрессирования хромосомных аберраций в организме. Информацию о прогрессировании можно применять для определения наличия ракового заболевания в указанном организме, определения стадии рака и прогнозирования состояния при раке. Каждое из таких определений может подразумевать классификацию рака, как описано в настоящем описании.

Указанную классификацию рака можно проводить разными способами. Например, можно подсчитывать объем аберрантных областей и сравнивать его с пороговым значением. Указанная классификация для областей может принимать числовое значение (т.е. концентрация опухоли, причем соответствующее и эталонное значения являются значениями для разных гаплотипов или разных областей), и можно определять изменение указанной концентрации. Изменение концентрации можно сравнивать с пороговым значением и определять, есть ли значимое увеличение, что должно сигнализировать о существовании опухоли.

## VII. Примеры.

### A. RHDO с применением SPRT.

В данном разделе представлен пример применения анализа относительной дозировки гаплотипа (RHDO) с применением SPRT у пациента с печеночно-клеточным раком (HCC). В ткани опухоли указанного пациента выявили делецию одной из двух 4 хромосом. Это приводит к потере гетерозиготности для SNP в 4 хромосоме. Для гаплотипирования указанного пациента анализировали геномную ДНК самого пациента, его жены и сына и определяли генотипы трех указанных индивидуумов. Затем на основании их генотипов получили структурные гаплотипы указанного пациента. Проводили массивно-параллельное секвенирование, идентифицировали и подсчитывали секвенированные считывания с аллелями SNP, соответствующими двум указанным гаплотипам 4 хромосомы.

Уравнения и принципы RHDO и SPRT были описаны выше. Согласно одному варианту реализации анализ RHDO можно запрограммировать на определение, например, 10% различия дозировки гаплотипов в пробе ДНК, которое соответствует присутствию 10% ДНК опухолевого происхождения, когда один из двух гаплотипов амплифицирован или делетирован. Согласно другим вариантам реализации чувстви-

тельность анализа RHDO можно установить на уровне 2, 5, 15, 20, 25, 30, 40, 50% и т.д., ДНК опухолевого происхождения в указанном образце ДНК. Чувствительность анализа RHDO можно корректировать по параметру для расчета верхнего и нижнего порогового значения кривых SPRT классификации. Таким корректируемым параметром может быть желаемый уровень предела определения (например, какой процент концентрации опухоли должен быть обнаруживаемым, что влияет на количество анализируемых молекул) и пороговое значения для классификации, например, с применением отношения шансов (отношение встречаемости маркерных последовательностей для одного гаплотипа к встречаемости маркерных последовательностей для другого гаплотипа).

При таком анализе RHDO нулевая гипотеза состоит в том, что два гаплотипа для 4 хромосомы присутствуют в одинаковой дозировке. Альтернативная гипотеза состоит в том, что дозировки указанных двух гаплотипов в указанном биологическом образце (например, плазме) различаются более чем на 10%. Число секвенированных считываний с аллелями SNP, соответствующими двум указанным гаплотипам, сравнивали статистическими методами в отношении двух указанных гипотез по мере того, как накапливались разные SNP. Классификацию SNP проводят, когда накопленных данных становится достаточно, чтобы определить, присутствуют ли два указанных гаплотипа в равном количестве или их количества статистически различаются по меньшей мере на 10%. Типичный блок классификации SPRT на q-плече 4 хромосомы показан на фиг. 18. Здесь применяют пороговое значение 10% только в целях иллюстрации. Также можно выявлять другие степени различия (например, 0,1, 1, 2, 5, 15 или 20%). Как правило, чем ниже степень различия, которую желательно выявить, тем больше молекул ДНК требуется для анализа. И наоборот, чем выше степень различия, которую желательно выявить, тем меньше число молекул ДНК, которое требуется для анализа, при котором все же можно достичь статистически значимого результата. Для данного анализа SPRT применяют отношение шансов, но можно применять и другие параметры, такие как z-показатель или p-показатель.

В указанном образце плазмы пациента с НСС, отобранном во время диагностики, было сделано 76 и 148 успешных классификаций RHDO для p и q-плечей 4 хромосомы соответственно. Все классификации RHDO указывают на то, что в образце плазмы, отобранном во время диагностики, был дисбаланс дозировки гаплотипов. Для сравнения в также анализировали образцы плазмы пациента, отобранные после хирургической резекции опухоли, как показано на фиг. 18В. Для образцов, отобранных после лечения, было сделано 4 и 9 успешных классификаций RHDO для p и q-плечей 4 хромосомы соответственно. Все четыре классификации RHDO указывают на то, что в указанном образце плазмы нет обнаруживаемого дисбаланса дозировки гаплотипа >10%. Из 9 классификаций RHDO для хромосомы 4q, 7 указывают на отсутствие дисбаланса дозировки гаплотипа, а 2 указывают, что такой дисбаланс присутствует. Количество блоков RHDO, демонстрирующих дисбаланс дозировки гаплотипа >10, значительно снизилось после резекции опухоли, что указывает на то, что размер области хромосомы, демонстрирующей дисбаланс дозировки гаплотипа >10%, стал значительно меньше в пробе после лечения, чем в пробе до лечения. Данные результаты указывают на то, что относительная концентрация ДНК опухолевого происхождения в плазме снизилась после хирургической резекции опухоли.

По сравнению с не гаплотип-специфичными способами анализ RHDO может обеспечить более точную оценку относительной концентрации ДНК опухолевого происхождения и особенно полезен при мониторинге прогрессирования заболевания. Таким образом, можно ожидать, что случаи прогрессирования заболевания могут проявлять увеличение относительной концентрации ДНК опухолевого происхождения в плазме; тогда как случаи стабильного заболевания или случаи, когда опухоль регрессирует, могут демонстрировать снижение относительной концентрации ДНК опухолевого происхождения в плазме.

#### В. Направленный анализ.

Согласно избранным вариантам реализации можно проводить универсальное секвенирование фрагментов ДНК и последующее применение способа направленного обогащения. Такой способ в настоящем описании также называют обогащенным направленным секвенированием. Один вариант реализации такого способа состоит в предварительном отборе фрагментов с применением захвата из раствора (например, система Agilent SureSelect, пользовательский набор для обогащения Illumina TruSeq ([illumina.com/applications/sequencing/targeted\\_resequencing.ilmn](http://illumina.com/applications/sequencing/targeted_resequencing.ilmn)) или с применением пользовательского набора для обогащения MyGenostics GenCap ([mygenostics.com/](http://mygenostics.com/))) или захвата на основе матриц (например, с применением системы Roche NimbleGen). Хотя можно захватывать некоторые другие области, более предпочтительно захватывать определенные области. Такие способы могут позволить проводить анализ таких областей с большей глубиной (например, можно секвенировать или анализировать больше фрагментов при помощи количественной ПЦР) и/или с меньшими затратами. При большей глубине может увеличиваться чувствительность в указанной области. Можно осуществлять другие способы обогащения на основании размеров фрагментов и характера метилирования.

Соответственно альтернативой анализа пробы ДНК общегеномным способом является нацеливание на рассматриваемые области для выявления распространенных хромосомных aberrаций. Направленный способ может потенциально повысить экономическую эффективность данного способа, поскольку процесс анализа можно главным образом сосредоточить на областях, где потенциально присутствуют хромосомные aberrации, или на областях с изменениями, которые могут быть характерны для определенно-

го типа опухоли, или на областях с изменениями, которые могут обладать особой клинической значимостью. Примеры последнего варианта изменений могут включать в себя изменения, которые могут возникать на ранней стадии онтогенеза конкретного типа рака (например, наличие увеличения числа копий 1q и 8q, и делеции 8q является ранним хромосомным изменением при НСС - van Malenstein et al. *Eur J Cancer* 2011;47:1789-97); или изменения, которые ассоциированы с хорошим и плохим прогнозом (например, во время прогрессирования опухоли наблюдается добавление в 6q и 17q и потеря в 6q и 9q, а с более плохим выживанием у пациентов с раком ободочной и прямой кишки ассоциировано наличие ПГ в 18q, 8p и 17p - Westra et al. *Clin Colorectal Cancer* 2004;4:252-9); или изменения, которые являются прогностическим признаком ответа на лечение (например, наличие добавления в 7p является прогностическим признаком ответа на ингибиторы тирозинкиназы у пациентов с мутациями в рецепторе эпидермального фактора роста - Yuan et al. *J Clin Oncol* 2011;29:3435-42). Другие примеры областей генома, которые изменены при раковом заболевании, можно найти в многочисленных базах данных (например, база данных Проекта анатомии генома при раке ([cgap.nci.nih.gov/Chromosomes/RecurrentAberrations](http://cgap.nci.nih.gov/Chromosomes/RecurrentAberrations)) и Атлас генетики и цитогенетики в онкологии и гематологии ([atlasgeneticsoncology.org/Tumors/Tumorliste.html](http://atlasgeneticsoncology.org/Tumors/Tumorliste.html)). В отличие от этого, при ненаправленном общегеномном способе области, в которых с малой вероятностью возникают хромосомные aberrации, можно анализировать в той же степени, что и области с потенциальными aberrациями.

Мы применили стратегию направленного обогащения для анализа проб плазмы, отобранных у 3 пациентов с НСС и 4 здоровых контрольных субъектов. Направленное обогащение проводили с применением системы захвата SureSelect от компании "Agilent" (Gnrke et al. *Nat. Biotechnol* 2009;27:182-9). Система SureSelect была выбрана в качестве примера возможной технологии направленного обогащения. Также можно применять другие системы направленного захвата на основе жидкой фазы (пользовательская система обогащения Illumina TruSeq) или твердой фазы (например, система Roche-Nimblegen), а также системы направленного обогащения на основе ампликона (например, система QuantaLife и система RainDance). Зонды для захвата разработаны так, чтобы они были локализованы в областях хромосом, которые часто и редко проявляют aberrации при НСС. После направленного захвата каждую пробу ДНК секвенируют при помощи одной полосы измерительной кюветы анализатора IlluminaGAIIx. Области, в которых редко возникают увеличение числа копий и делеции, применяют в качестве эталона для сравнения с областями, в которых амплификации и делеции присутствуют чаще.

На фиг. 19 показаны распространенные хромосомные aberrации, обнаруживаемые при НСС (фигура адаптирована из Wong et al (*Am J Pathol* 1999;154:37-43)). Линии с правой стороны идеограммы хромосомы показывают добавление хромосом, а линии с левой стороны соответствуют потере хромосом в пробах индивидуального пациента. Толстые линии показывают добавление высокого уровня. Квадратами показана локализация зондов для направленного захвата.

#### Направленный анализ с подсчетом маркерных последовательностей

Для выявления хромосомных aberrаций мы сначала рассчитали нормированную встречаемость маркерных последовательностей для областей с потенциальными aberrациями и эталонных областей. Нормированные встречаемости маркерных последовательностей затем корректировали на содержание GC указанной области, как ранее было описано у Chen et al (*PLoS One* 2011;6:e21791). В настоящем примере р-плечо 8 хромосомы было выбрано в качестве области потенциальной aberrации, а q-плечо 9 хромосомы было выбрано в качестве эталонной области. Ткани опухоли трех пациентов с НСС анализировали на наличие хромосомных aberrаций при помощи матрицы AffymetrixSNP 6.0. Изменения дозировки хромосом для 8p и 9q в тканях опухоли показаны ниже для 3 указанных пациентов. У пациента НСС 013 была потеря 8p и не было изменений в 9q. У пациента НСС 027 было добавление в 8p и не было изменений в 9q. У пациента НСС 023 была потеря 8p и не было изменений в 9q.

Затем для трех указанных пациентов с НСС и четырех здоровых контрольных субъектов рассчитывали отношение нормированных встречаемостей маркерных последовательностей между хромосомами 8p и 9q. На фиг. 20А показаны результаты отношения нормированных встречаемостей маркерных последовательностей для пациентов с НСС и здоровых субъектов. Для случаев НСС 013 и НСС023 наблюдалось сниженное отношение нормированных встречаемостей маркерных последовательностей между 8p и 9q. Это согласуется с наблюдением, касающимся потери 8p в тканях опухоли. Для случая НСС 027 наблюдалось повышенное отношение, что согласуется с добавлением 8p в тканях опухоли в данном случае. Пунктирные линии показывают область с двумя стандартными отклонениями от среднего значения для четырех случаев нормы.

#### Направленный анализ по размеру

В предыдущих разделах описан принцип выявления ассоциированных с раком изменений профиля размеров для фрагментов ДНК в плазме у пациентов, страдающих раковыми заболеваниями. Изменения размеров также можно выявлять при помощи способа направленного обогащения. Для указанных трех случаев (НСС 013, НСС 027 и НСС023) размер каждого секвенированного фрагмента ДНК определяли после выравнивания секвенированных считываний с эталонным геномом человека. Заклчение о размере секвенированных фрагментов ДНК делали на основании координат наиболее удаленных от центра нуклеотидов на обоих концах. Согласно другим вариантам реализации можно секвенировать целый

фрагмент ДНК, и затем размер указанного фрагмента можно определить прямо на основании секвенированной длины. Распределение по размерам фрагментов ДНК, выравнивающихся с хромосомой 8p, сравнивали с распределением по размерам двух фрагментов ДНК, выравнивающихся с хромосомой 9q. Для выявления различия распределений по размеру для двух популяций ДНК сначала для каждой популяции в настоящем примере определяли долю фрагментов ДНК короче, чем 150 по. Согласно другим вариантам реализации можно применять другие уровни отсечки по размеру, например, 80 по, 110 по, 100 по, 110 по, 120 по, 130 по, 140 по, 160 по и 170 по. Затем определяли значения  $\Delta Q$  как разность двух указанных долей.  $\Delta Q = Q_{8p} - Q_{9q}$ , где  $Q_{8p}$  представляет собой долю фрагментов ДНК, выравнивающихся с хромосомой 8p, которые короче чем 150 по; а  $Q_{9q}$ , представляет собой долю фрагментов ДНК, выравнивающихся с хромосомой 9q, которые короче чем 150 по.

Поскольку более узкое распределение по размерам фрагментов ДНК может давать более высокое значение доли ДНК короче уровня отсечки (например, 150 по в настоящем примере), более высокое (более положительное) значение  $\Delta F$  может соответствовать более узкому распределению фрагментов ДНК, выравнивающихся с хромосомой 8p, по сравнению с распределением фрагментов ДНК, выравнивающихся с хромосомой 9q. Напротив, более низкий (или более отрицательный) результат может указывать на более широкое распределение по размерам фрагментов ДНК, выравнивающихся с хромосомой 8p, по сравнению с распределением фрагментов ДНК, выравнивающихся с хромосомой 9q.

На фиг. 20B показаны результаты анализа по размерам после направленного обогащения и массивно-параллельного секвенирования для 3 пациентов с печеночно-клеточным раком и 4 здоровых субъектов из контрольной группы. Положительные значения  $\Delta Q$  у четырех здоровых контрольных субъектов указывает на несколько более узкое распределение по размерам фрагментов ДНК, выравнивающихся с хромосомой 8p, по сравнению с распределением фрагментов ДНК, выравнивающихся с хромосомой 9q. Пунктирные линии показывают интервал  $\Delta Q$  в пределах двух стандартных отклонений от среднего для четырех контрольных субъектов. Значения  $\Delta Q$  для случаев НСС 013 и НСС 023 были ниже среднего значения для контрольных субъектов более чем на два стандартных отклонения. В указанных двух случаях имела место делеция хромосомы 8p в тканях опухоли. Делеция 8p в опухоли может приводить к снижению распределению ДНК опухолевого происхождения в плазме для данной области хромосомы. Поскольку ДНК опухолевого происхождения в кровяном русле короче, чем ДНК, происходящая из неопухолевых тканей, это может приводить к кажущемуся более широкому распределению фрагментов ДНК в плазме, выравнивающихся с хромосомой 8p. Это согласуется с более низким (более отрицательным) значением  $\Delta Q$  в двух указанных случаях. Напротив, увеличение числа копий 8p в случае НСС 027 могло привести к кажущемуся более узкому распределению фрагментов ДНК, выравнивающихся с данной областью. Таким образом, более высокая доля ДНК фрагментов в плазме, выравнивающихся с хромосомой 8p, может считаться короткими. Это согласуется с тем наблюдением, что значение  $\Delta Q$  в случае НСС 027 более положительное, чем у здоровых контрольных субъектов.

С. Множество областей для выявления хромосомных aberrаций опухолевого происхождения.

В тканях опухоли часто выявляются хромосомные aberrации, включая делецию и увеличения числа копий определенных областей хромосом. При разных типах рака наблюдаются характерные картины хромосомных aberrаций. Здесь мы применяем несколько примеров, чтобы проиллюстрировать разные способы выявления указанных ассоциированных с раком хромосомных aberrаций в плазме пациентов, страдающих раковыми заболеваниями. Наш способ также применим для скрининга на рак и мониторинга прогрессирования заболевания, а также ответа на лечение. Анализировали пробы, отобранные у одного пациента с НСС и двух пациентов с раком носоглотки (NPC). У пациента с НСС пробы венозной крови отбирали перед хирургической резекцией опухоли и после нее. У двух пациентов с NPC пробы венозной крови отбирали во время диагностики. Кроме того, анализировали пробы плазмы крови одного хронического носителя вируса гепатита В и одного субъекта с определяемой ДНК вируса Эпштейна-Барр в плазме. У указанных двух субъектов рака не обнаружено.

Определение хромосомных aberrаций опухолевого происхождения проводили при помощи анализа на микроматрицах. Точнее говоря, ДНК, экстрагированную из клеток крови и пробы опухоли пациентов с НСС, анализировали при помощи системы микроматриц Affymetrix SNP6.0. Генотипы клеток крови и тканей опухоли определяли при помощи Affymetrix Genotyping Console v4.0. Хромосомные aberrации, включая добавления и делеции, определяли при помощи алгоритма Birdseed v2, основанного на интенсивностях разных аллелей SNP и зондов вариации числа копий (CNV) на указанной микроматрице.

Анализ, основанный на подсчете

Для проведения анализа на основе подсчета секвенированных маркерных последовательностей у каждого из указанных субъектов отбирали по десять миллилитров венозной крови. Для каждой пробы крови после ее центрифугирования отделяли плазму.

ДНК выделяли из 4-6 мл плазмы при помощи набора QIAmp blood mini ("Qiagen"). Библиотеку ДНК плазмы создавали, как было описано ранее (Lo YMD. Sci Transl Med 2010;2:61ra91), и затем ее подвергали массивно-параллельному секвенированию с применением платформы Illumina Genome Analyzer.

Проводили секвенирование молекул ДНК плазмы с парными концами. Каждую молекулу секвенировали с каждого из двух концов на протяжении 50 по, т.е. суммарно на протяжении 100 по на молекулу. Два конца каждой молекулы выравнивали с маскированным геномом человека без повторов (Hg18 NCBI.36, загружен с сайта UCSC genome.ucsc.edu) при помощи программы SOAP2 (soap.genomics.org.cn/) (Li R et al. *Bioinformatics* 2009, 25:1966-7).

Затем указанный геном разделяли на несколько сегментов длиной 1 мегабаза (1 Мб) и определяли число секвенированных считываний, выравнивающихся с каждым из сегментов 1 Мб. Встречаемость маркерных последовательностей в каждом интервале затем корректировали при помощи алгоритма, основанного на локально взвешенном сглаживании диаграммы рассеяния (LOESS) в соответствии с содержанием GC в каждом интервале (Chen E et al. *PLoS One* 2011, 6:e21791). Такая коррекция направлена на минимизацию количественных погрешностей, связанных с секвенированием, которые возникают из-за различий содержания GC в разных сегментах генома. Упомянутое выше разделение на сегменты 1 Мб применяются в целях иллюстрации. Также можно применять другие размеры сегментов, например 2 Мб, 10 Мб, 25 Мб или 50 Мб. Также возможно выбирать размер сегмента на основании геномных характеристик конкретной опухоли у конкретного пациента и конкретного типа опухоли в целом. Кроме того, если можно показать, что процесс секвенирования обладает низкой погрешностью в связи с GC, например для технологий секвенирования отдельных молекул, таких как система Helicos (www.helicosbio.com) или система Pacific Biosciences Single Molecular Real-Time (www.pacificbiosciences.com), этап коррекции GC можно пропустить.

В предыдущем исследовании мы секвенировали 57 образцов плазмы, отобранных у пациентов, не страдающих раковыми заболеваниями. Результаты указанного секвенирования были применены для определения эталонного диапазона встречаемости маркерных последовательностей для каждого сегмента 1 Мб. Для каждого сегмента 1 Мб определяли среднее и стандартное отклонение встречаемости маркерных последовательностей у 57 индивидуумов. Затем результаты исследования субъектов выражали в виде z-показателя, который рассчитывали по следующей формуле:

$$z\text{-показатель} = (\text{количество секвенированных маркерных последовательностей, в данном случае среднее}) / \text{CO},$$

где термин "среднее" относится к количеству секвенированных маркерных последовательностей, выравнивающихся с конкретным сегментом 1 Мб для эталонных проб; а CO - стандартное отклонение количества секвенированных маркерных последовательностей, выравнивающихся с конкретным сегментом 1 Мб для эталонных проб.

На фиг. 21-24 показаны результаты анализа на основе подсчета секвенированных маркерных последовательностей у четырех исследуемых субъектов. Сегменты 1 Мб показаны на краю графиков. Номера и идеограммы хромосом человека (наружный круг) ориентированы от теломеры короткого плеча к теломере длинного плеча по часовой стрелке (центромеры показаны желтым). На фиг. 21 внутренняя окружность 2101 показывает области аберрации (делеции или увеличения числа копий), определенные при анализе указанной опухоли. Внутренняя окружность 2101 показана с пятью шкалами. Шкала простирается от -2 (самая внутренняя линия) до +2 (самая наружная линия). Значение -2 соответствует потере обеих копий хромосом для соответствующей области. Значение -1 соответствует потере одной из двух копий хромосом. Значение 0 соответствует отсутствию добавления или потери хромосом. Значение +1 соответствует добавлению одной копии хромосомы, а +2 соответствует добавлению двух копий хромосомы.

Средняя окружность 2102 показывает результаты анализа плазмы. Можно видеть, что указанные результаты являются зеркальным отражением внутреннего круга. Средняя окружность 2102 содержит больше делений шкалы, но прогрессирование такое же. Наружная окружность 2103 показывает точки данных, полученные при анализе данных после лечения, и указанные точки данных показаны серым (подтверждение отсутствия избыточного/недостаточного присутствия - отсутствия аберраций).

Области хромосом с присутствием в избытке секвенированных маркерных последовательностей в плазме (z-показатель >3) показаны зелеными точками 2110. Области без значимых хромосомных аберраций, выявленные в плазме (z-показатель от -3 до 3) показаны серыми точками. Присутствие в избытке/недостатке нормируют на общее количество встречаемости. В случае амплификации перед секвенированием при нормировке можно учесть погрешность GC.

На фиг. 21 показаны круговые диаграммы для пациентов с НСС, на которых изображены данные о подсчете секвенированных маркерных последовательностей в ДНК плазмы согласно вариантам реализации настоящего изобретения. Линии, проходящие из внутренней части наружу: хромосомные аберрации в ткани опухоли, выявляемые при анализе на микроматрицах (красный и зеленый цвет соответствует делеции и увеличению числа копий, соответственно); анализ z-показателя для пробы плазмы, полученной до хирургической резекции опухоли и через 1 месяц после резекции. До резекции опухоли хромосомные аберрации, выявляемые в плазме, коррелируют с аберрациями, идентифицируемыми в ткани опухоли посредством анализа на микроматрицах. После резекции опухоли большинство ассоциированных с раком хромосомных аберраций в плазме исчезло. Указанные данные показывают ценность такого способа при мониторинге прогрессирования заболевания и эффективности лечения.

На фиг. 22 показан анализ на основе подсчета секвенированных маркерных последовательностей в

пробе плазмы хронического носителя HBV без НСС согласно вариантам реализации настоящего изобретения. В отличие от пациента с НСС (фиг. 21), ассоциированные с раком хромосомные aberrации в плазме данного пациента не выявлялись. Указанные данные показывают ценность такого способа при скрининге на рак и мониторинге.

На фиг. 23 показан анализ на основе подсчета секвенированных маркерных последовательностей в образце плазмы пациента NPC 3 стадии согласно вариантам реализации настоящего изобретения. В пробе плазмы, отобранной перед проведением лечения, определялись хромосомные aberrации. Точнее говоря, значимые aberrации были идентифицированы в хромосомах 1, 3, 7, 9 и 14.

На фиг. 24 показан анализ на основе подсчета секвенированных маркерных последовательностей в пробе плазмы пациента NPC 4 стадии согласно вариантам реализации настоящего изобретения. В пробе плазмы, отобранной перед проведением лечения, определялись хромосомные aberrации. По сравнению с пациентом с 3 стадией заболевания (фиг. 23) определялось больше хромосомных aberrаций. Встречаемости секвенированных маркерных последовательностей также в большей степени отклонялись от среднего для контрольных субъектов, т.е. z-показатель больше отклонялся от нуля (либо в положительную, либо в отрицательную сторону). Повышенное количество хромосомных aberrаций и более высокая степень отклонения встречаемостей секвенированных маркерных последовательностей по сравнению с контролем отражает более тяжелую степень геномных изменений при более поздней стадии заболевания и, следовательно, показывает ценность такого способа при установлении стадии, прогнозировании и мониторинге рака.

#### Анализ, основанный на размерах

В предыдущих исследованиях было показано, что распределение по размеру фрагментов ДНК, происходящих из тканей опухоли, уже, чем распределение по размеру фрагментов из неопухолевых тканей (Diehl F et al. Proc Natl Acad Sci USA 2005, 102(45): 16368-73). В предыдущих разделах был охарактеризован способ выявления дисбаланса гаплотипов в плазме посредством анализа, основанного на размерах ДНК в плазме. Здесь применялись данные секвенирования для пациентов НСС, чтобы более подробно проиллюстрировать данный способ.

В целях иллюстрации мы идентифицировали две области для анализа по размерам. В одной области (1 хромосома (chr1); координаты 159 935 347 - 167 219 158) в тканях опухоли была выявлена дупликация одной из двух гомологичных хромосом. В одной области (10 хромосома (chr10); координаты 100 137 050 - 101 907 356) в тканях опухоли была выявлена делеция одной из двух гомологичных хромосом (т.е. ПГ). Помимо определения того, из какого гаплотипа произошел секвенированный фрагмент, также способами биоинформатики определяли размер указанного секвенированного фрагмента на основании координат самых удаленных от центра нуклеотидов секвенированного фрагмента в эталонном геноме. Затем определяли распределение по размерам фрагментов из каждого из двух указанных гаплотипов.

Для области с ПГ Chr10 в ткани опухоли гаплотип был делетирован (делетированный гаплотип). Следовательно, все фрагменты ДНК в плазме, выравнивающиеся с данным делетированным гаплотипом, происходили из нераковых тканей. С другой стороны, фрагменты, выравнивающиеся с гаплотипом, который не был делетирован в тканях опухоли (неделетированный гаплотип) мог происходить как из опухолевых, так и из неопухолевых тканей. Поскольку распределение по размерам ДНК опухолевого происхождения уже, мы можем предположить более узкое распределение по размерам для фрагментов из неделетированного гаплотипа по сравнению с распределением фрагментов из делетированного гаплотипа. Различие двух распределений по размеру можно определить путем построения графика зависимости накопленной встречаемости фрагментов от размера фрагментов ДНК. Популяция ДНК с более узким распределением по размерам может быть более богата короткой ДНК и, следовательно, в ней происходит более быстрое увеличение накопленной частоты в короткой части спектра размеров.

На фиг. 25 показан график зависимости накопленной частоты ДНК в плазме от размера для области, проявляющей ПГ в ткани опухоли согласно вариантам реализации настоящего изобретения. По оси X отложен размер фрагмента в парах оснований. По оси Y отложен процент фрагментов, обладающих размером ниже значения на оси X. Последовательности из неделетированного гаплотипа демонстрируют более быстрое увеличение и более высокую накопленную частоту меньше размера 170 по сравнению с последовательностями из делетированного гаплотипа. Это указывает на то, что фрагменты ДНК из неделетированного гаплотипа присутствовали в большем избытке. Это согласуется с предсказанием, сделанным выше исходя из распределения короткой ДНК опухолевого происхождения из неделетированного гаплотипа.

Согласно некоторым вариантам реализации различие в распределении по размеру можно оценить количественно по разности накопленных частот двух популяций молекул ДНК. Мы определяем  $\Delta Q$  как разность накопленных частот двух указанных популяций.

$\Delta Q = Q_{\text{non-deleted}} - Q_{\text{deleted}}$ , где  $Q_{\text{non-deleted}}$  представляет собой накопленную частоту для секвенированных фрагментов ДНК, происходящих из неделетированного гаплотипа; а  $Q_{\text{deleted}}$  представляет собой накопленную частоту для секвенированных фрагментов ДНК, происходящих из делетированного гаплотипа.

На фиг. 26 показан график зависимости  $\Delta Q$  от размера секвенированной ДНК плазмы для области с ПГ.  $\Delta Q$  достигает 0,2 при размере 130 по согласно вариантам реализации настоящего изобретения. Это указывает на то, что применение уровня отсечки 130 по для определения короткой ДНК оптимально, для применения в формулах, приведенных выше. При применении указанного уровня отсечки короткие молекулы ДНК представлены в избытке 20% в популяции из неделетированного гаплотипа по сравнению с популяцией из делетированного гаплотипа. Данное процентное различие (или полученную аналогичным способом величину) можно затем сравнивать с пороговым значением, полученным для людей, не страдающих раковыми заболеваниями.

Для области с увеличением числа копий хромосомы в ткани опухоли один гаплотип был дублирован (амплифицированный гаплотип). Поскольку в плазму могло выделяться избыточное количество коротких молекул ДНК опухолевого происхождения из указанного амплифицированного гаплотипа, распределение по размеру для фрагментов из амплифицированного гаплотипа может быть уже, чем распределение по размеру для фрагментов из неамплифицированного гаплотипа. Аналогично сценарию ПГ, различие в распределении по размеру можно определить путем построения графика зависимости накопленной встречаемости фрагментов от размера фрагментов ДНК. Популяция ДНК с более узким распределением по размерам может быть более богата короткой ДНК и, следовательно, в ней происходит более быстрое увеличение накопленной частоты в короткой части спектра размеров.

На фиг. 27 показан график зависимости накопленной частоты ДНК в плазме от размера для области с дубликацией хромосомы в ткани опухоли согласно вариантам реализации настоящего изобретения. Последовательности из амплифицированного гаплотипа демонстрировали более быстрое увеличение и более высокую накопленную частоту меньше размера 170 по по сравнению с последовательностями из неамплифицированного гаплотипа. Это указывает на то, что фрагменты ДНК из амплифицированного гаплотипа присутствовали в большем избытке. Это согласуется с предсказанием, приведенным ниже, поскольку большее количество короткой ДНК опухолевого происхождения происходило из амплифицированного гаплотипа.

Аналогично сценарию ПГ, различие в распределении по размеру можно оценивать количественно по разности накопленных частот двух популяций молекул ДНК. Мы определяем  $\Delta Q$  как разность накопленных частот двух указанных популяций.  $\Delta Q = Q_{\text{amplified}} - Q_{\text{non-amplified}}$ , где  $Q_{\text{amplified}}$  представляет собой накопленную частоту для секвенированных фрагментов ДНК, происходящих из амплифицированного гаплотипа; а  $Q_{\text{non-amplified}}$  представляет собой накопленную частоту для секвенированных фрагментов ДНК, происходящих из неамплифицированного гаплотипа.

На фиг. 28 показан график зависимости  $\Delta Q$  от размера секвенированной ДНК плазмы для амплифицированной области согласно вариантам реализации настоящего изобретения.  $\Delta Q$  достигает 0,08 при размере 126 по согласно вариантам реализации настоящего изобретения. Это указывает на то, что при применении уровня отсечки 126 по для определения короткой ДНК короткие молекулы ДНК представлены в избытке 8% в популяции из амплифицированного гаплотипа по сравнению с популяцией из неамплифицированного гаплотипа.

#### D. Дополнительные технологии.

Согласно другим вариантам реализации можно применять технологии, специфичные в отношении последовательности. Например, можно разрабатывать олигонуклеотиды для гибридизации фрагментов конкретной области. Затем указанные олигонуклеотиды можно подсчитывать способом, аналогичным подсчету секвенированных маркерных последовательностей. Данный способ можно применять для типов рака, которые проявляют конкретные аберрации.

#### VIII. Компьютерная система.

В любой из компьютерных систем, упоминаемых в настоящем описании, может применяться любое подходящее количество подсистем. Примеры таких подсистем показаны на фиг. 9 в компьютерном устройстве 900. Согласно некоторым вариантам реализации компьютерная система включает одно компьютерное устройство, в котором указанными подсистемами могут быть компоненты указанного компьютерного устройства. Согласно другим вариантам реализации компьютерная система может включать несколько компьютерных устройств, каждый из которых является подсистемой, с внутренними компонентами.

Подсистемы, показанные на фиг. 29, соединяются между собой системной шиной 2975. Показаны дополнительные подсистемы, такие как принтер 2974, клавиатура 2978, фиксированный диск 2979, монитор 2976, который соединен с адаптером дисплея 2982, и другие. Периферические устройства и устройства ввода/вывода (I/O), которые соединены с контроллером I/O 2971, могут быть соединены с компьютерной системой любыми средствами, известными в технике, такими как последовательный порт 2977. Например, для присоединения компьютерной системы 2900 к региональной сети, такой как интернет, устройству ввода данных или сканеру можно применять последовательный порт 2977 или внешний интерфейс 2981 (такой как локальная сеть, Wi-Fi и др.). Соединение через системную шину 2975 позволяет коммутировать центральный процессор 2973 с каждой подсистемой и контролировать выполнение инструкций из запоминающей системы 2972 или фиксированного диска 2979, а также обмен информации.

ей между подсистемами. Запоминающая система 2972 и/или фиксированный диск 2979 могут составлять часть считываемой компьютером среды. Любое из значений, упоминаемых в настоящем описании, можно выводить из компонента в другой компонент и можно выводить для пользователя.

Компьютерная система может включать множество одинаковых компонентов или подсистем, например, соединенных между собой внешним интерфейсом 2981 или внутренним интерфейсом. Согласно некоторым вариантам реализации компьютерные системы, подсистемы или устройства можно соединять друг с другом сетью. Согласно таким примерам один компьютер может считаться клиентом, а другой компьютер - сервером, где каждый из них может быть частью одной и той же компьютерной системы. И клиент, и сервер могут включать многие системы, подсистемы или компоненты.

Следует понимать, что любой из вариантов реализации настоящего изобретения можно реализовать в форме алгоритма управления с применением средств технического обеспечения и/или с применением компьютерных программ в модульном или интегрированном виде. На основании раскрытия изобретения и идей, приведенных в настоящем описании, специалист в данной области техники должен знать и понимать другие способы и/или пути осуществления вариантов реализации настоящего изобретения с применением средств технического обеспечения и сочетания средств технического обеспечения и программ.

Любые из компонентов или функций программ, описываемых в настоящем описании, можно выполнить в виде программного кода, выполняемого процессором с использованием любого подходящего машинного языка, такого как, например, Java, C++ или Perl, с использованием, например, традиционных или ориентированных на объект методов. Программный код может сохраняться как ряд инструкций или команд на носителе, считываемом компьютером, для хранения и/или передачи, подходящие носители включают запоминающее устройство с произвольной выборкой (RAM), постоянное запоминающее устройство (ПЗУ, ROM), магнитный носитель, жесткий дисковод или дискету, или оптический носитель, такой как компакт-диск (CD) или DVD (цифровой универсальный диск), флэш-память и т.п. Носитель, считываемый компьютером, может представлять собой любое сочетание таких устройств хранения или передачи информации.

Такие программы также можно кодировать и передавать с применением сигналов носителя, адаптированных для передачи по проводным, оптическим и/или беспроводным сетям, соответствующим протоколам, включая интернет. Как таковой носитель, считываемый компьютером, согласно одному варианту реализации настоящего изобретения может быть создан с применением сигнала данных, кодированного такими программами. Носитель, считываемый компьютером, кодированный программным кодом, может быть в пакете с совместимым устройством или предоставлен отдельно от других устройств (например, пересылаться через интернет). Любой такой считываемый компьютером носитель может быть размещен на или в пределах отдельного компьютерного программного продукта (например, жесткий дисковод или целая компьютерная система) и может быть представлен на или в рамках различных компьютерных программных продуктов в системе или сети. Компьютерная система может включать монитор, принтер или другой подходящий дисплей для предоставления пользователю любых результатов, указанных в настоящем описании.

Любые из способов, описываемых в настоящем описании, можно реализовать полностью или частично при помощи компьютерной системы, включающей процессор, который можно настроить для реализации указанных этапов. Таким образом, варианты реализации могут направляться в компьютерные системы, настроенные для проведения этапов любого из способов, описанных в настоящем описании, потенциально при помощи разных компонентов, осуществляющих соответствующие этапы или соответствующие группы этапов. Хотя и представленные в форме пронумерованных этапов, этапы способов, описываемых в настоящем описании, можно осуществлять одновременно или в разном порядке. Кроме того, части указанных этапов можно применять с частями других этапов из других способов. Также весь или часть этапа могут быть необязательными. Кроме того, любой из этапов любого из способов можно проводить при помощи модулей, замкнутых циклов или других средств осуществления указанных этапов.

Специфические детали конкретных вариантов реализации настоящего изобретения можно сочетать любым подходящим образом, не отклоняясь от сущности и области вариантов реализации настоящего изобретения. Однако другие варианты реализации настоящего изобретения могут быть направлены на специфические варианты реализации, связанные с каждым отдельным аспектом, или специфическими сочетаниями указанных отдельных аспектов.

Приведенное выше описание примерных вариантов реализации изобретения представлено в целях пояснения и описания. Не предполагается, что оно является исчерпывающим или ограничивает изобретение точно описанной формой, и в свете описанного выше возможны многие модификации и вариации. Варианты реализации выбраны и описаны для наилучшего объяснения принципов изобретения и его практических применений, которые позволят другим специалистам в данной области техники применять изобретение наилучшим образом в различных вариантах реализации и различных модификациях, которые подходят для определенного предполагаемого применения.

Перечисление определенных и неопределенных артиклей должно означать "один или более", если особо не указано другое. Все патенты, заявки на патент, публикации и описания, упоминаемые в данном описании, входят в него в виде ссылок для всех целей. Ни один из них не признается предыдущим уровнем техники.

## ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ анализа биологического образца организма для выявления хромосомных делеций или амплификаций, ассоциированных с раком, причем организм является диплоидным, и указанный биологический образец включает бесклеточную смесь множества молекул нуклеиновых кислот, происходящих из хромосом здоровых клеток и из хромосом клеток, возможно, ассоциированных с раком, причем указанный способ включает:

(1) определение аллелей первого и второго гаплотипов для здоровых клеток указанного организма в области хромосомы, которая включает множество локусов, причем аллели указанного первого и второго гаплотипов гетерозиготны в каждом локусе из указанного множества локусов;

(2) идентификацию локализации в эталонном геноме указанного организма каждой из множества молекул нуклеиновых кислот из указанного биологического образца;

(3) определение аллеля каждой из множества молекул нуклеиновых кислот из указанного биологического образца;

(4) идентификацию первой группы молекул нуклеиновых кислот как происходящей из указанного первого гаплотипа на основании идентифицированных локализаций, определенных аллелей множества молекул нуклеиновых кислот и аллелей первого гаплотипа во множестве локусов, причем указанная первая группа включает по меньшей мере одну молекулу нуклеиновой кислоты, локализирующуюся в каждом из локусов указанного множества в указанной области хромосомы;

(5) идентификацию второй группы молекул нуклеиновых кислот как происходящей из указанного второго гаплотипа на основании идентифицированных локализаций, определенных аллелей множества молекул нуклеиновых кислот и аллелей второго гаплотипа во множестве локусов, причем указанная вторая группа включает по меньшей мере одну молекулу нуклеиновой кислоты, локализирующуюся в каждом из локусов указанного множества в указанной области хромосомы;

(6) расчет при помощи компьютерной системы первого значения для указанной первой группы молекул нуклеиновых кислот, причем указанное первое значение определяет статистический показатель распределения по размеру молекул нуклеиновых кислот первой группы или определяет количество молекул нуклеиновых кислот из первой группы;

(7) расчет при помощи компьютерной системы второго значения для указанной второй группы молекул нуклеиновых кислот, причем указанное второе значение определяет статистический показатель распределения по размеру молекул нуклеиновых кислот второй группы или определяет количество молекул нуклеиновых кислот из второй группы;

(8) сравнение первого значения со вторым значением, причем разница или отношение между указанными первым значением и вторым значением, превышающее пороговое значение, указывает на то, что область хромосомы, которая включает множество локусов, проявляет делецию или амплификацию в клетках, ассоциированных с раком, причем это пороговое значение определяют так, как раскрыто в описании.

2. Способ по п.1, отличающийся тем, что указанное первое значение для первой группы молекул нуклеиновых кислот соответствует статистическому показателю распределения по размеру молекул нуклеиновых кислот указанной первой группы, а указанное второе значение для второй группы молекул нуклеиновых кислот соответствует статистическому показателю распределения по размеру молекул нуклеиновых кислот указанной второй группы.

3. Способ по п.2, отличающийся тем, что указанное первое значение представляет собой средний размер молекул нуклеиновых кислот указанной первой группы, а указанное второе значение представляет собой средний размер молекул нуклеиновых кислот указанной второй группы.

4. Способ по п.2, отличающийся тем, что указанное первое значение, обозначенное как  $Q_{\text{НарI}}$ , представляет собой долю молекул нуклеиновых кислот в указанной первой группе, которые короче размера отсечки, и тем, что второе значение, обозначенное как  $Q_{\text{НарII}}$ , представляет собой долю молекул нуклеиновых кислот в указанной второй группе, которые короче размера отсечки

5. Способ по п.4, отличающийся тем, что сравнение первого значения со вторым значением включает определение разницы между ними:  $\Delta Q = Q_{\text{НарI}} - Q_{\text{НарII}}$ , причем положительное значение  $\Delta Q$  выше порогового значения указывает на то, что второй гаплотип включает делецию в указанной области хромосомы в ткани опухоли в указанном организме или что первый гаплотип включает амплификацию в указанной области хромосомы в ткани опухоли в этом организме.

6. Способ по п.4, отличающийся тем, что сравнение первого значения со вторым значением включает определение разницы между ними:  $\Delta Q = Q_{\text{НарI}} - Q_{\text{НарII}}$ , причем значение  $\Delta Q$ , приблизительно равное нулю, указывает на отсутствие делеции или амплификации в указанной области хромосомы, которая включает множество локусов.

7. Способ по п.2, отличающийся тем, что указанное первое значение, обозначенное как  $F_{\text{НарI}}$ , и указанное второе значение, обозначенное как  $F_{\text{НарII}}$ , определяют для соответствующей группы по формуле  $F = \sum^w \text{length} / \sum^N \text{length}$ , где  $\sum^w \text{length}$  представляет собой сумму длин молекул нуклеиновых кислот соответствующей группы с длиной, равной или меньше размера отсечки  $w$ , а  $\sum^N \text{length}$  представляет собой

сумму длин молекул нуклеиновых кислот соответствующей группы с длиной, равной или меньшей, чем  $N$  оснований, где  $N$  больше, чем  $w$ .

8. Способ по п.7, отличающийся тем, что сравнение первого значения со вторым значением включает определение разницы между ними:  $\Delta F = F_{\text{Нар I}} - F_{\text{Нар II}}$ , причем если второй гаплотип включает делецию в указанной области хромосомы в раковой ткани в организме, то  $\Delta F$  имеет положительное значение выше порогового значения.

9. Способ по п.7, отличающийся тем, что сравнение первого значения со вторым значением включает определение разницы между ними:  $\Delta F = F_{\text{Нар I}} - F_{\text{Нар II}}$ , причем если второй гаплотип включает амплификацию в указанной области хромосомы в раковой ткани в организме, то  $\Delta F$  имеет отрицательное значение выше порогового значения.

10. Способ по п.1, отличающийся тем, что указанное первое значение для первой группы молекул нуклеиновых кислот соответствует количеству молекул нуклеиновых кислот из первой группы, а указанное второе значение для второй группы молекул нуклеиновых кислот соответствует количеству молекул нуклеиновых кислот из второй группы.

11. Способ по п.10, отличающийся тем, что сравнение первого значения со вторым значением включает определение отношения указанного первого значения к указанному второму значению для определения относительной концентрации раковых молекул нуклеиновых кислот в указанном биологическом образце, причем способ дополнительно включает определение относительной концентрации раковых молекул нуклеиновых кислот в указанном биологическом образце во множество моментов времени для диагностики, определения стадии, прогнозирования или отслеживания прогрессирования степени рака в указанном организме.

12. Способ по любому из пп.1-11, отличающийся тем, что включает анализ множества хромосомных областей; определение количества областей хромосом, которые проявляют делецию или амплификацию; и сравнение указанного количества с одним или более пороговыми значениями для определения степени рака в указанном организме.

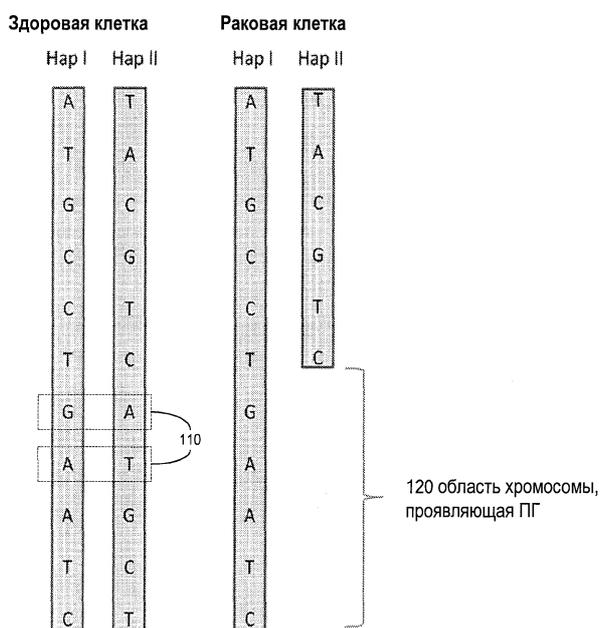
13. Способ по п.12, дополнительно включающий определение для каждой области хромосомы, которая была идентифицирована как проявляющая делецию или амплификацию, степени делеции или амплификации;

сравнение результата с одним или более пороговыми значениями для определения степени рака указанного организма.

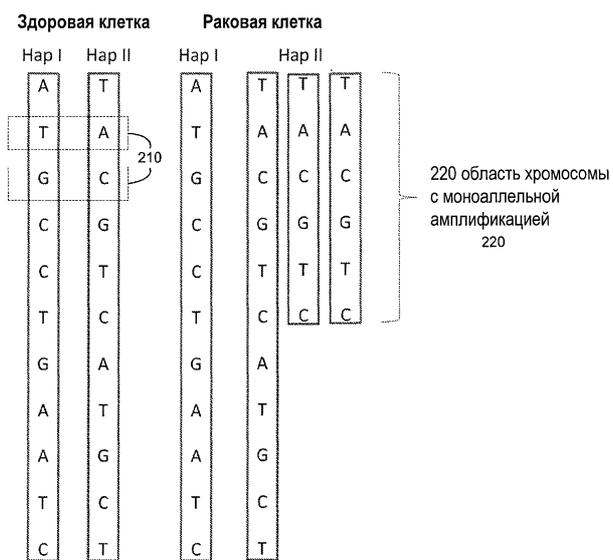
14. Способ по любому из пп.1-11, отличающийся тем, что включает повторение во множество моментов времени стадий анализа множества областей хромосомы и определение количества хромосомных областей, которые проявляют делецию или амплификацию, для диагностики, определения стадии, прогнозирования или отслеживания прогрессирования степени рака в указанном организме.

15. Машиночитаемый носитель, включающий компьютерную программу для осуществления способа по любому из пп.1-14.

16. Компьютерная система, включающая процессор и машиночитаемый носитель по п.15, для осуществления способа по любому из пп.1-14.



Фиг. 1



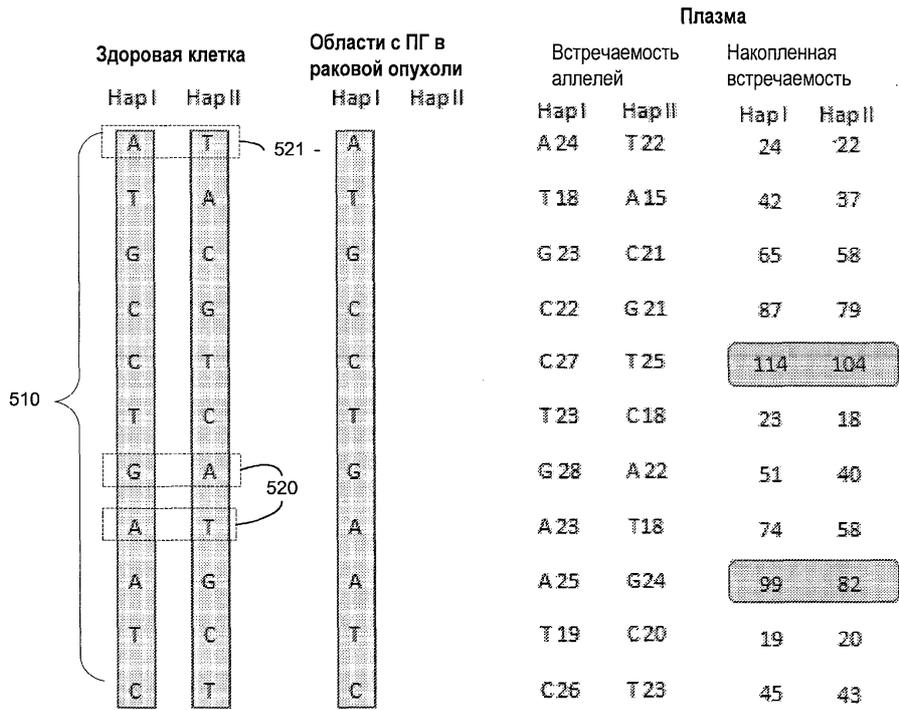
Фиг. 2

310	320	330	340
Тип рака	Добавление	Утрата	Ссылка
Рак щитовидной железы	7p, 8q и 9q	22	Hemmer et al. Am J Pathol, 1999;154:1539-47.
Рак желудка	2q37, 3p21, 5q34-35, 7q34-36, 11q13, 11q23-25, 12q24, 15q23-25, 17q21-25 и 20q12-13	4q, 13q, 5q, 6q и 18p	Noguchi et al. Am J ClinPathol2001; 115:828-34
Рак предстательной железы	1q32, 3q26, 4q26, 7q21, 8q22, 9q33, 17q25 и Xq21	2q22, 4q27-4q28, 5q15, 6q15, 8p21, 10q23, 12p13, 13q21, 15q23, 16q22 и 18q21-22	Sun et al. Prostate 2007;67:692-700.
Мелкоклеточный рак легких	3q26-29, 5p12-13 и 8q23-24	3p13-14, 4q32-35, 5q32-35, 8p21-22, 10q25, 13q13-14 и 17p12-13	Balsara et al. Oncogene 2002;21:6877-83.
Немелкоклеточный рак легких	1q31, 3q25-27, 5p13-14 и 8q23-24	31p21, 8p22, 9p21-22, 13q22 и 17p12-13	Balsara et al. Oncogene 2002;21:6877-83.
Рак носоглотки	1p34, 3q26, 6q25 и 3q26	3p, 9p, 9q, 11q, 13q и 14q	Lo et al. Semin Cancer Biol 2002;12:451-62
Рак мочевого пузыря	1q, 5p, 6p, 8q, 11q, 17q и 20q	3p, 4q, 4q, 6q, 8p, 9p и 18q	El-Rifai et al. Am J Pathol 2000;156:871-8
Рак ободочной и прямой кишки	13q и 20q	4q и 18q	De Angelis et al. Int J Colorectal Dis 2001;16:38-45.
Рак головы и шеи	3q26 и 11q13	3p, 9p и 17p	Smeets et al. Oncogene 2006;25:2558-64
Меланома	1q, 2, 6p, 7, 8, 17 и 20	6q, 8p, 9 и 10	Bastian et al. Cancer Res 1998;58:2170-5.
Лимфома	1q, 3, 6p, 7, 11, 12, 18 и X	1p, 8p и X	Monni et al. Blood 1996;87:5269-78

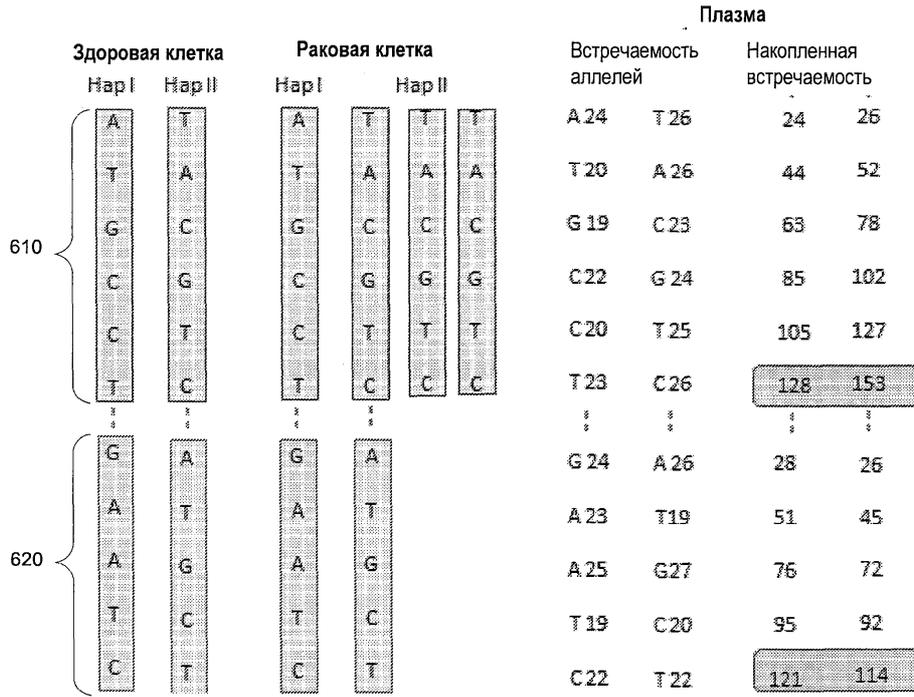
Фиг. 3



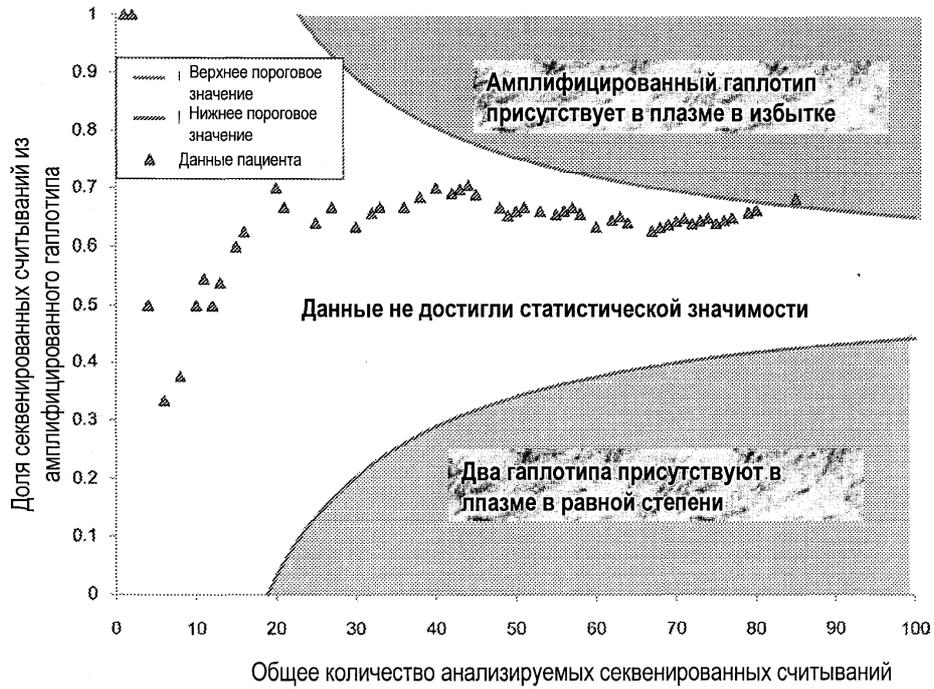
Фиг. 4



Фиг. 5



Фиг. 6



Фиг. 7

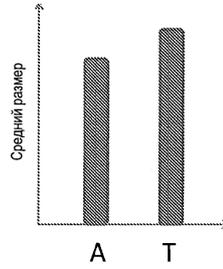
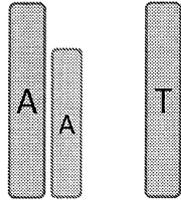
Здоровые ткани



Ткани опухоли

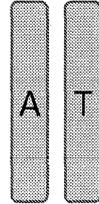


Плазма

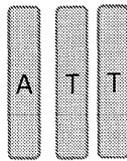


Фиг. 8

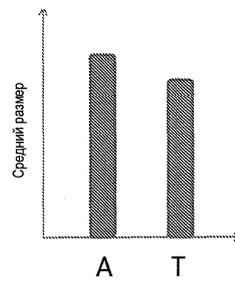
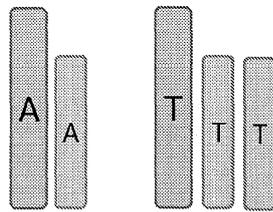
Здоровые ткани



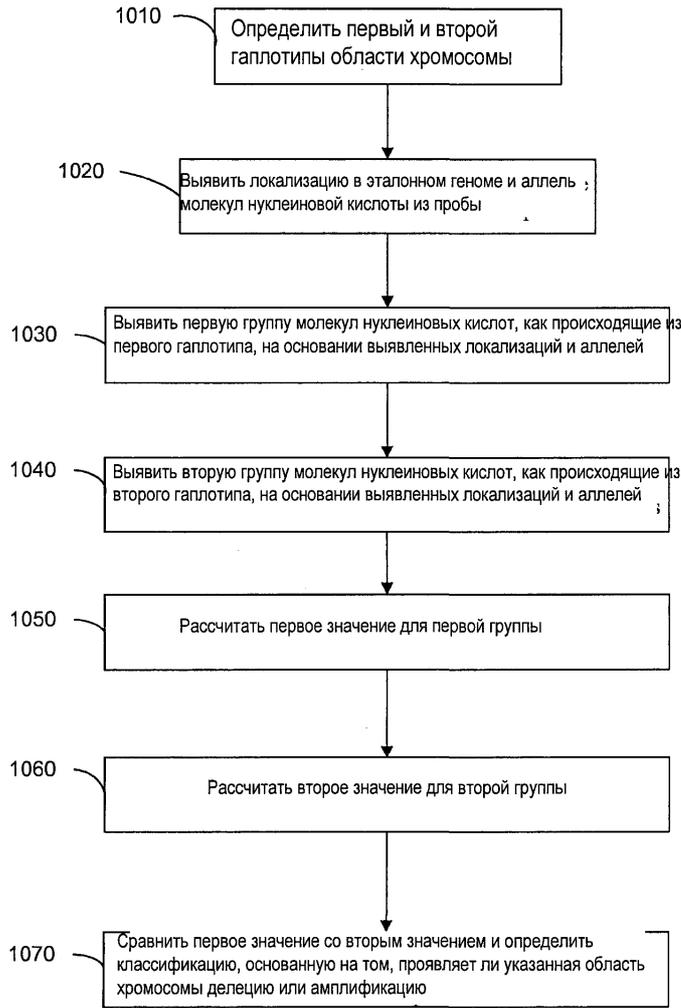
Ткани опухоли



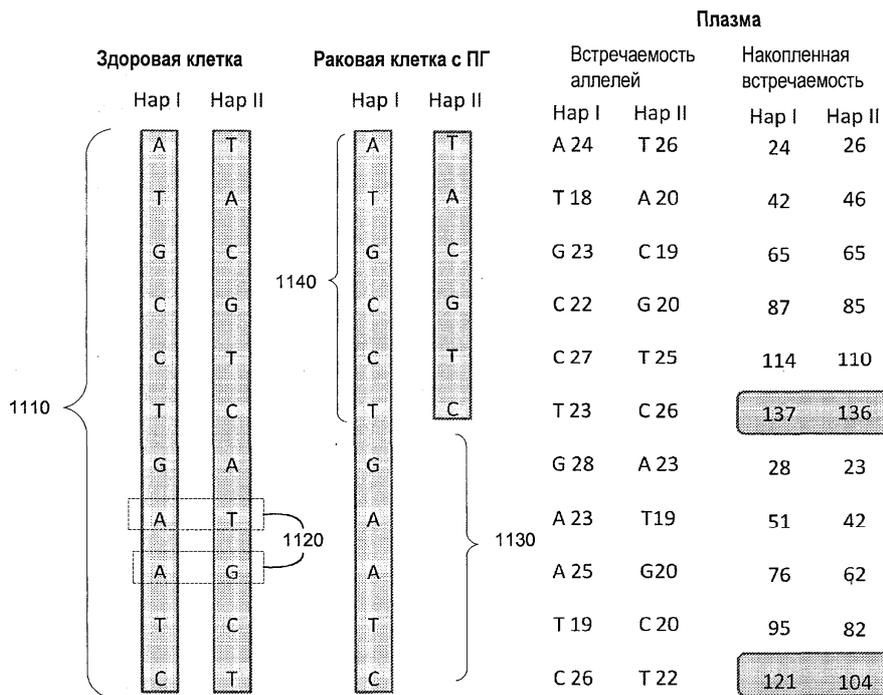
Плазма



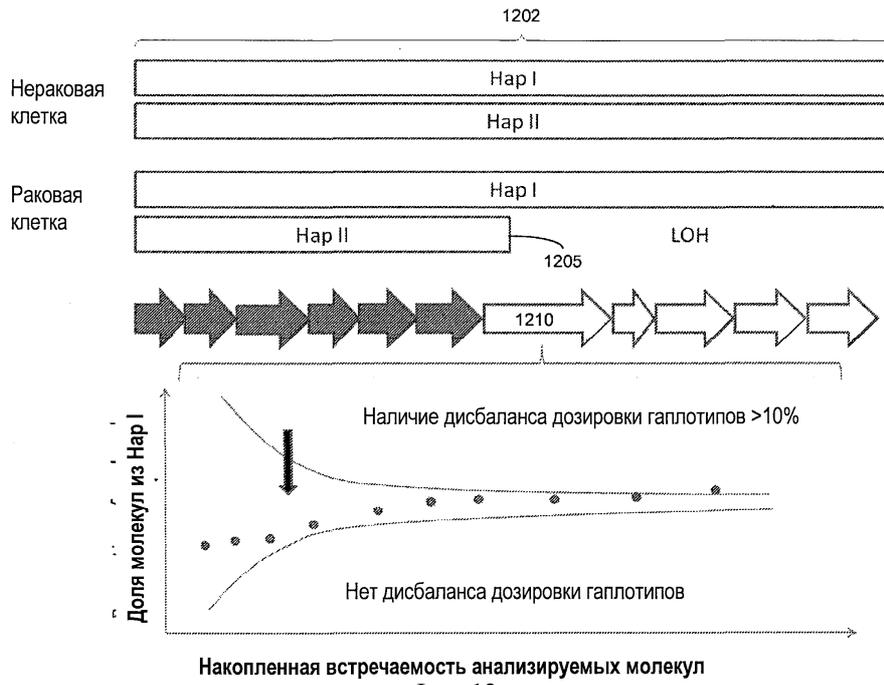
Фиг. 9



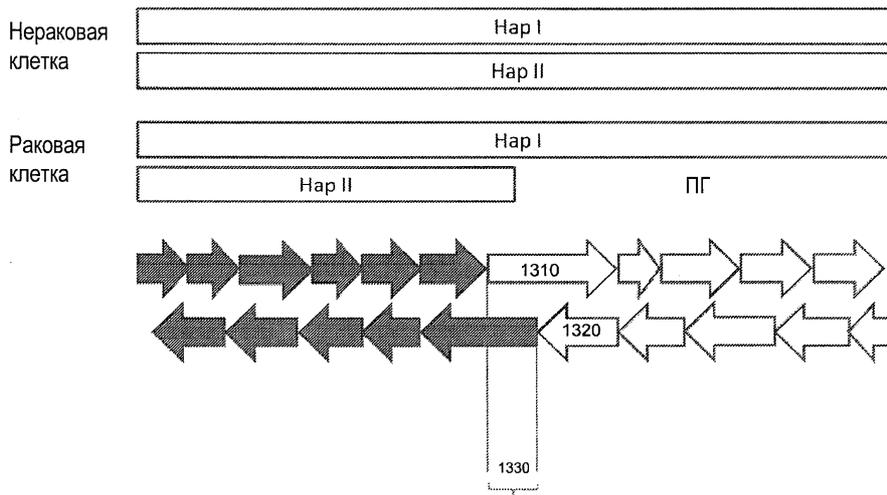
Фиг. 10



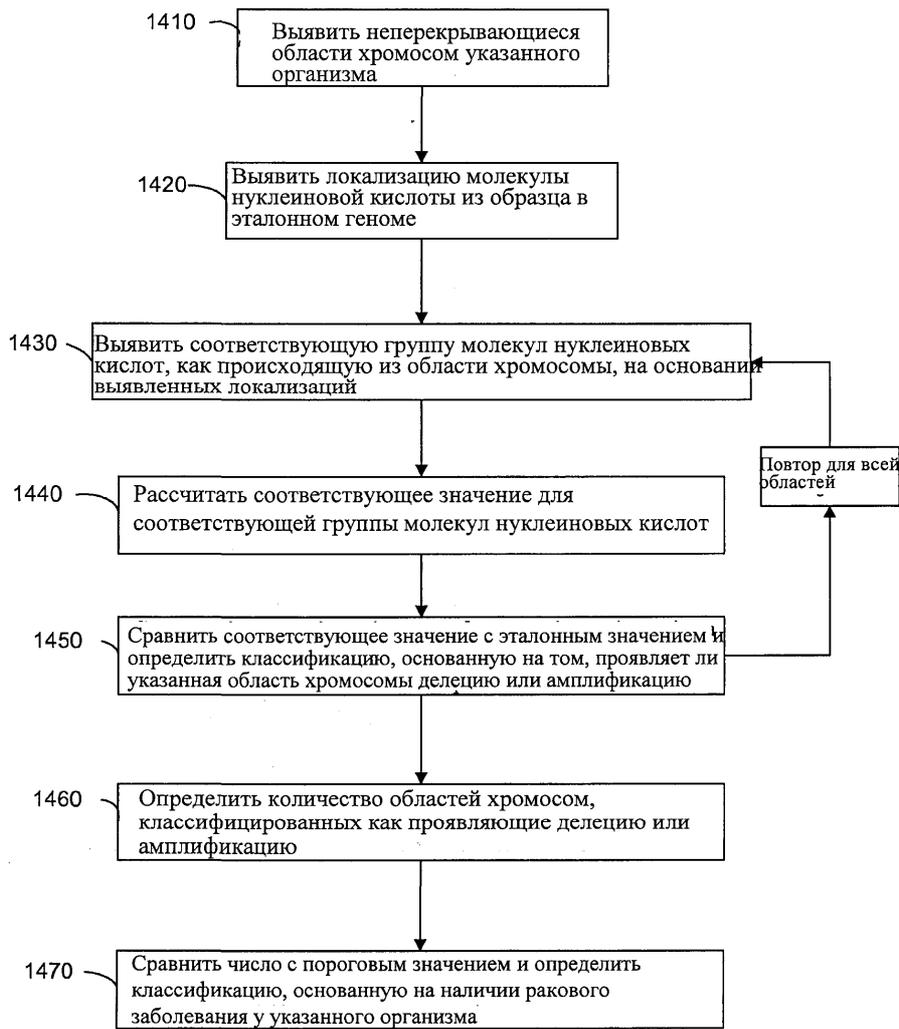
Фиг. 11



Накопленная встречаемость анализируемых молекул  
Фиг. 12



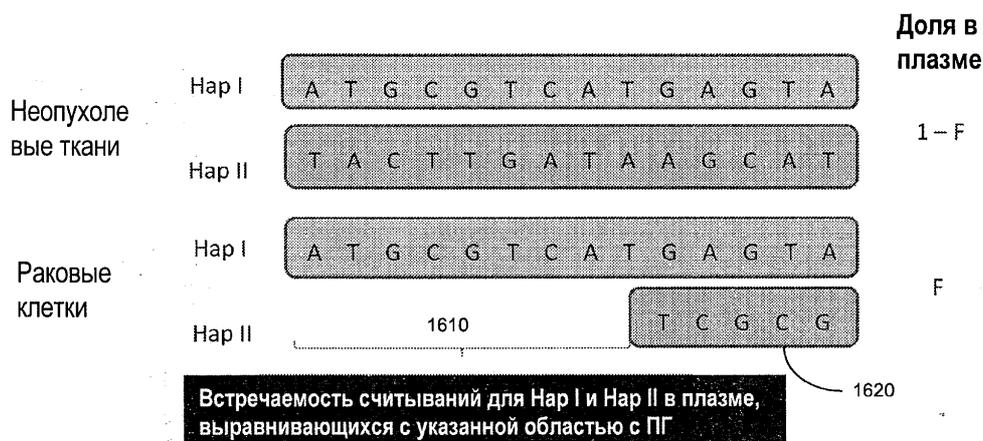
Локализация, где начинается ПГ  
Фиг. 13



Фиг. 14

Относительная концентрация ДНК опухолевого происхождения	Расчетное количество молекул, требующихся на один сегмент	Размер каждого сегмента (кб)	Общее количество молекул, которое требуется проанализировать на весь геном (миллион)
1510	1520	1530	1540
50%	950	100	28.5
		1 000	2.85
		5 000	0.570
		10 000	0.285
25%	3 800	100	114
		1 000	11.4
		5 000	2.28
		10 000	1.14
12.5%	15 000	100	450
		1 000	45
		5 000	9
		10 000	4.5
6.3%	60 000	100	1 800
		1 000	180
		5 000	36
		10 000	18
3.2%	240 000	100	7 200
		1 000	720
		5 000	144
		10 000	72

Фиг. 15

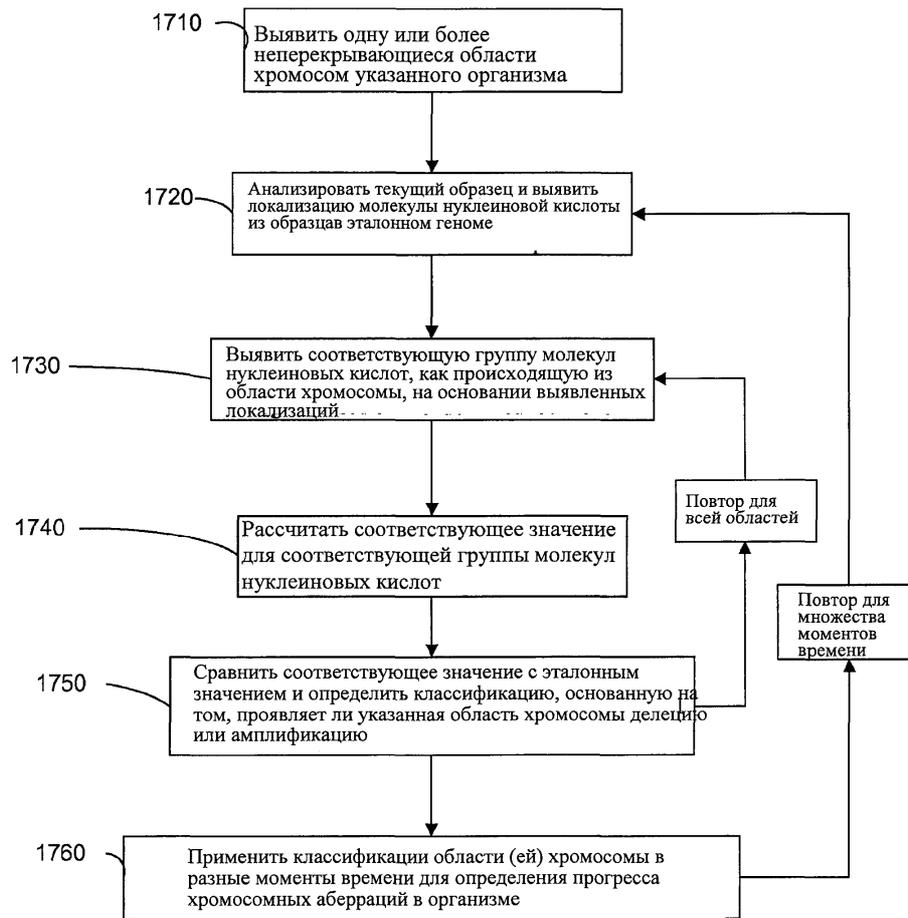


Относительная концентрация ДНК опухолевого происхождения (F):

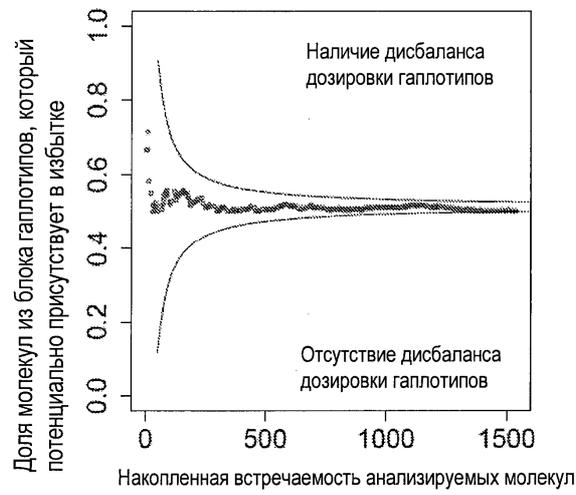
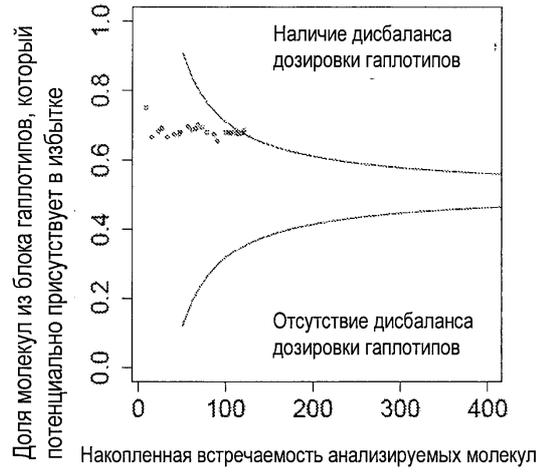
Разница между встречаемостью считываний для Нар I и Нар II

Встречаемость считываний для Нар I

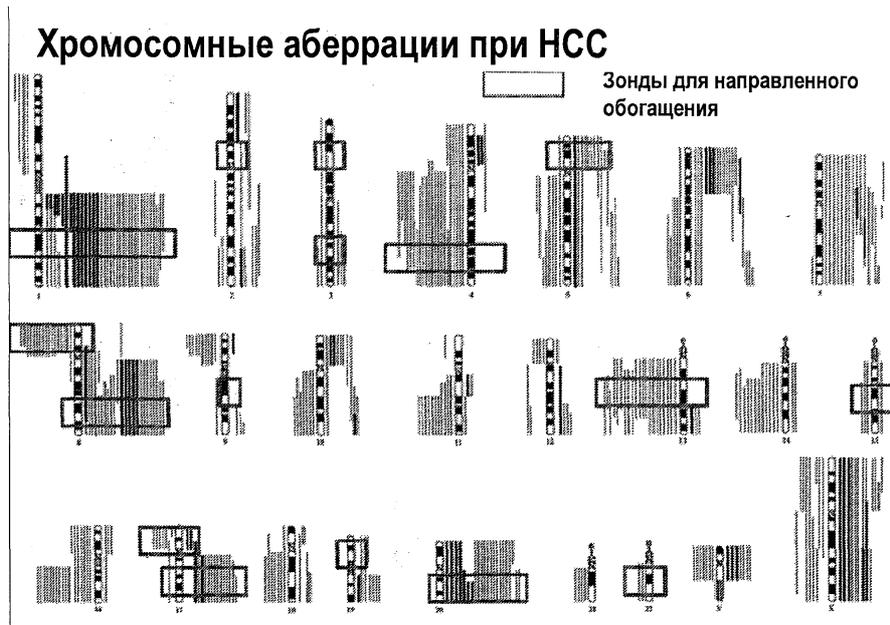
Фиг. 16



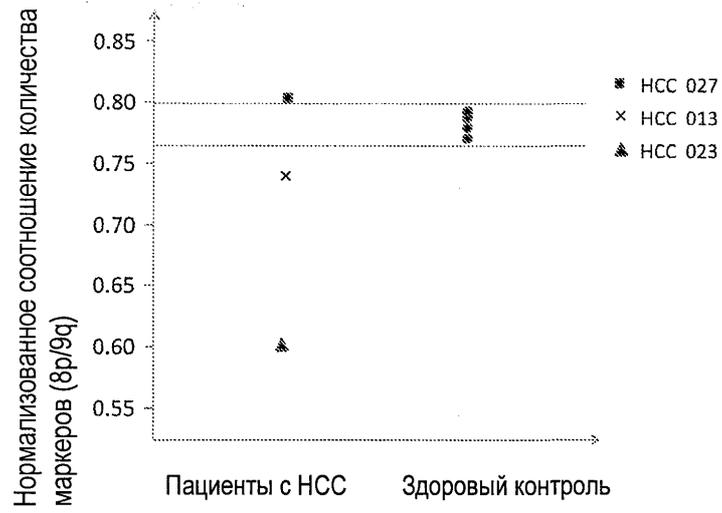
Фиг. 17



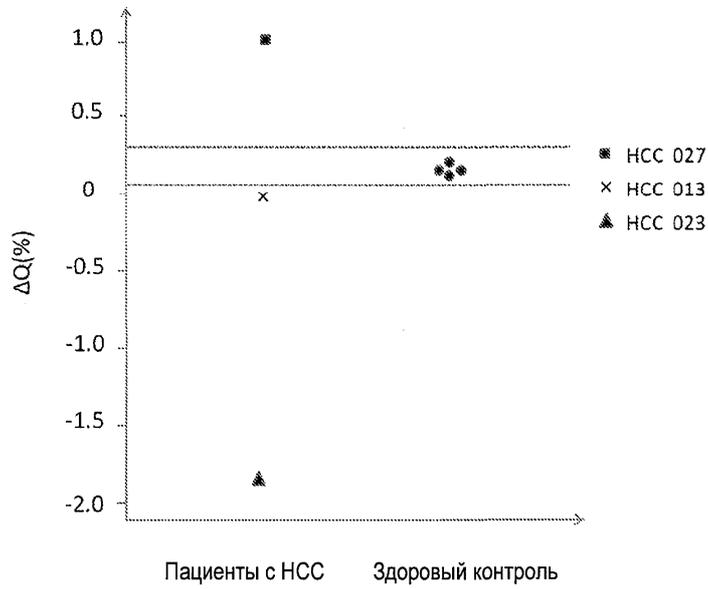
Фиг. 18



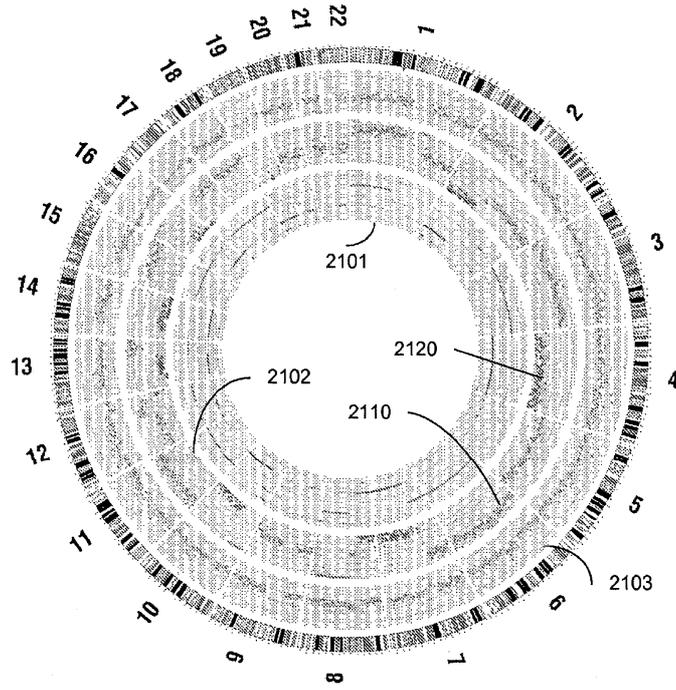
Фиг. 19



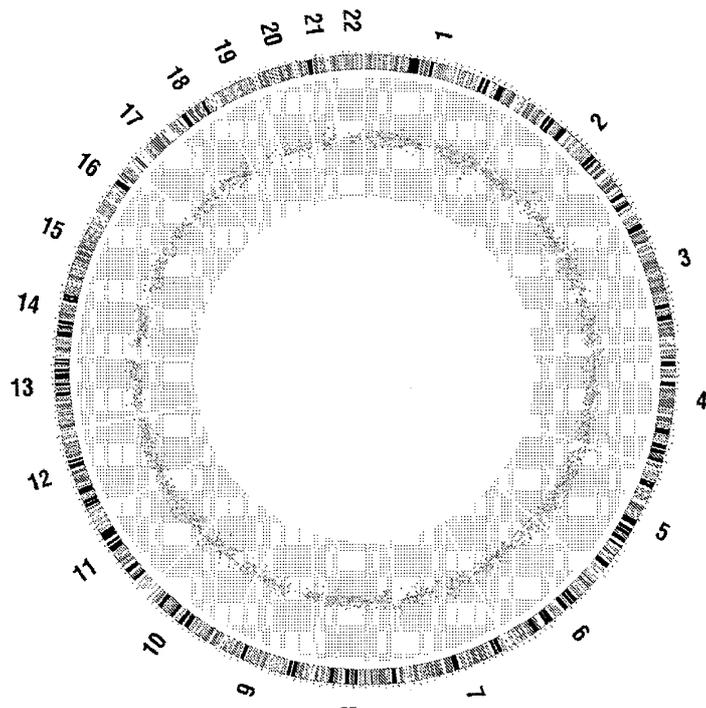
Фиг. 20А



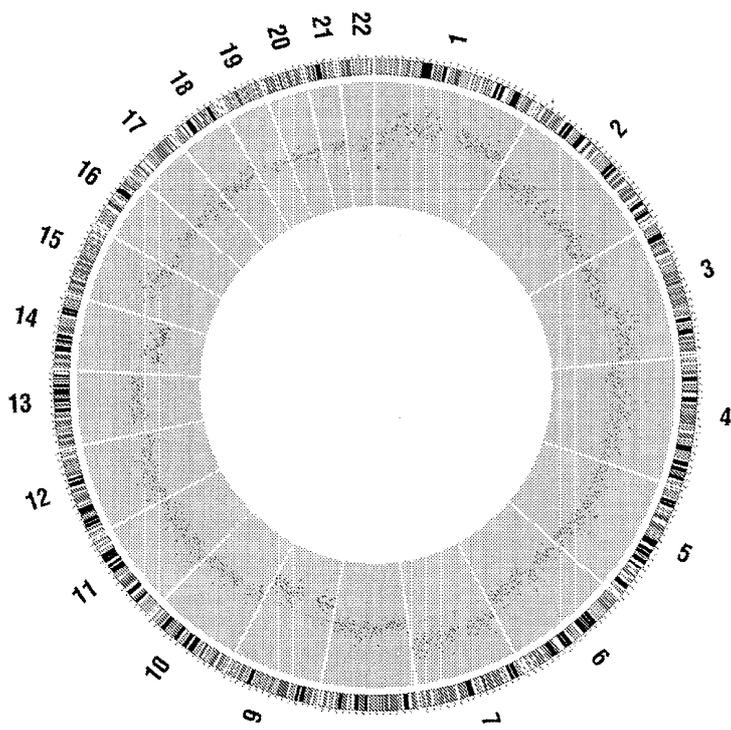
Фиг. 20В



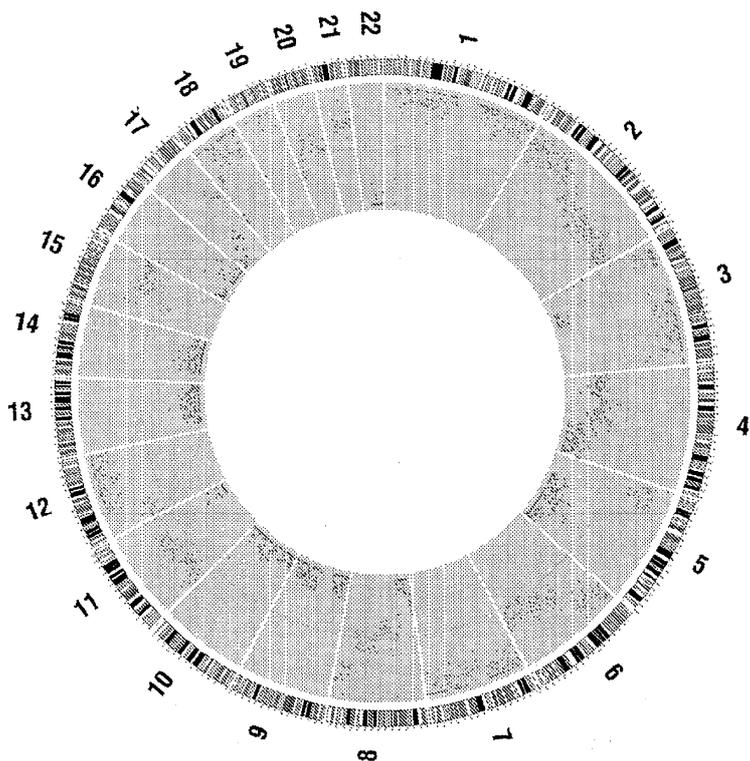
Фиг. 21



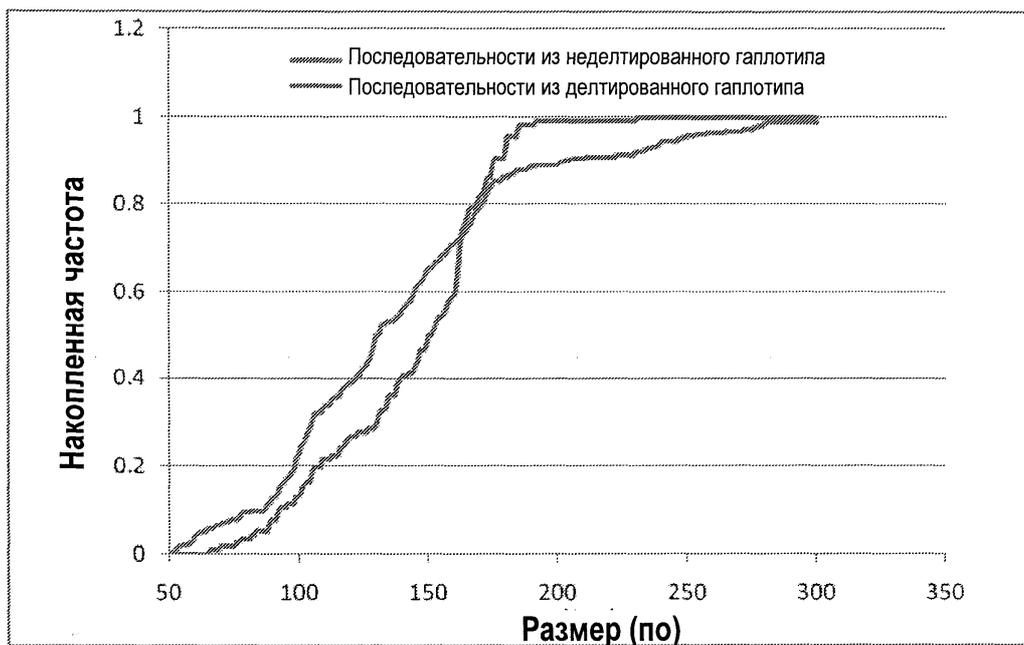
Фиг. 22



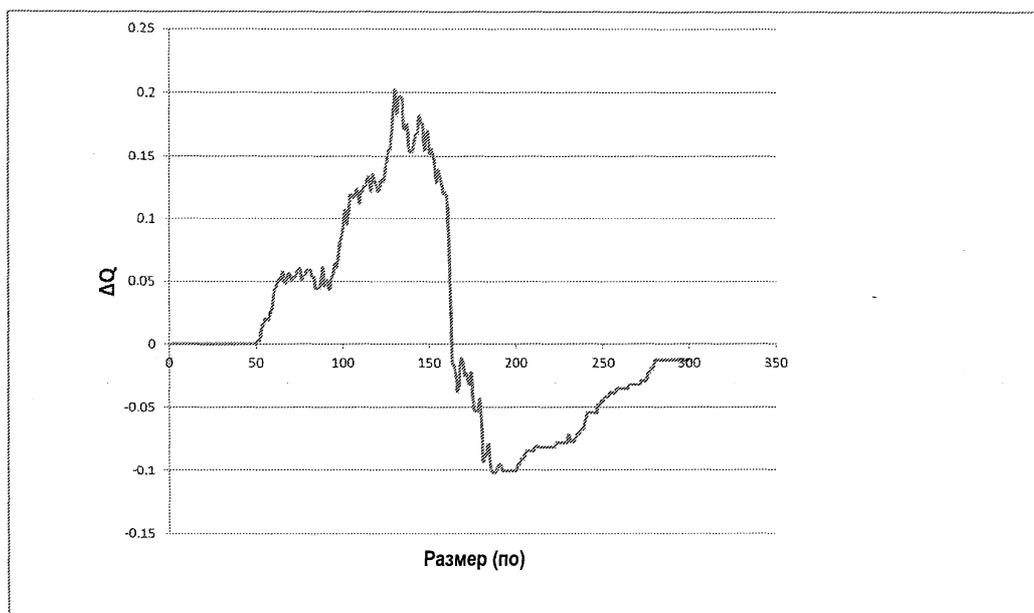
Фиг. 23



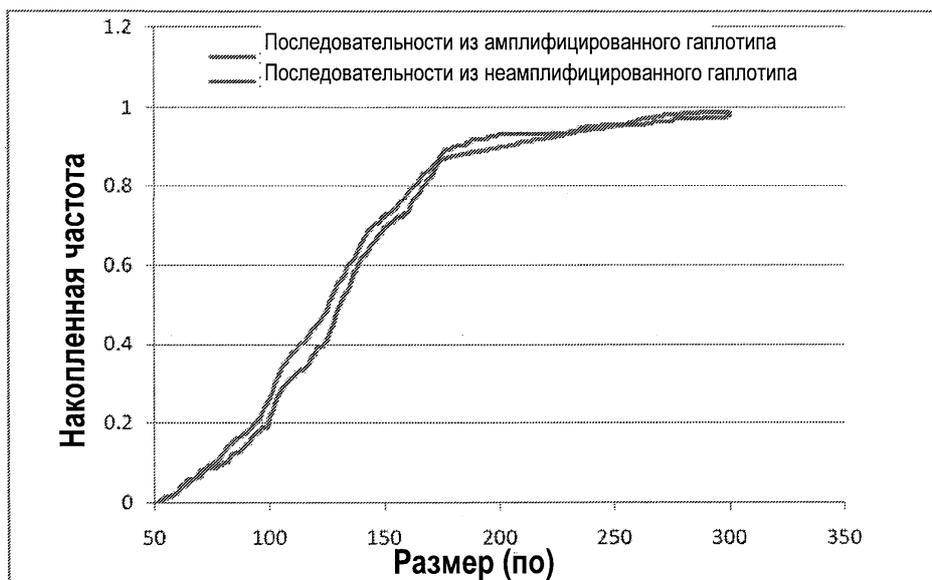
Фиг. 24



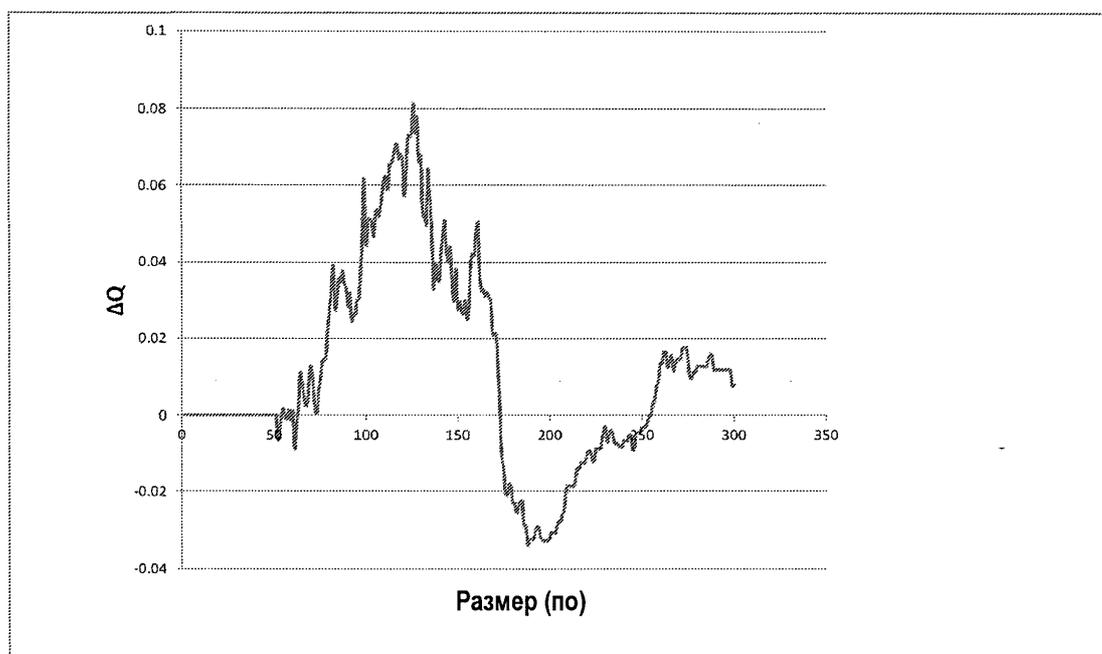
Фиг. 25



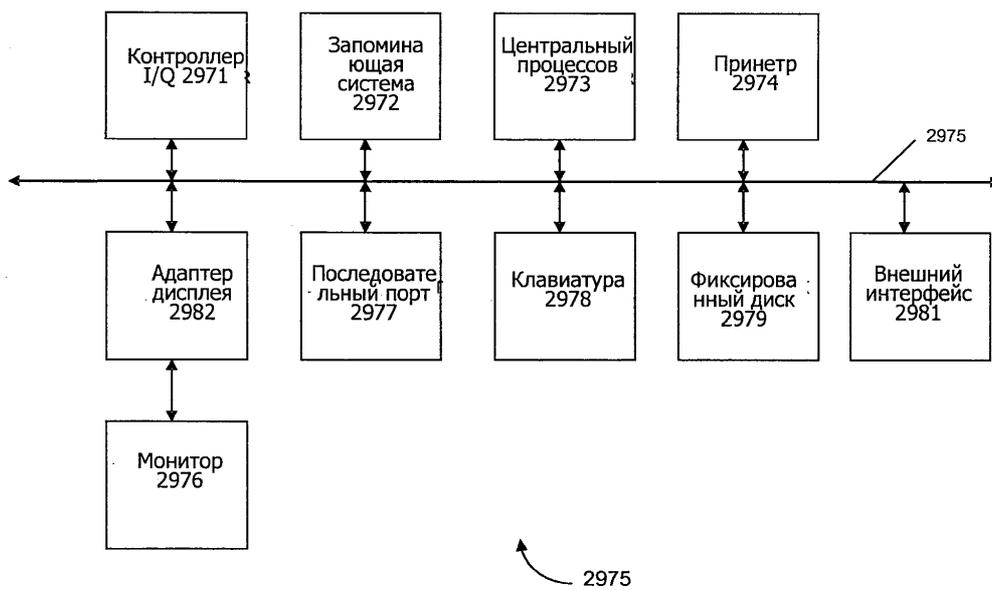
Фиг. 26



Фиг. 27



Фиг. 28



Фиг. 29

