

(19)



**Евразийское  
патентное  
ведомство**

(21) **202293523** (13) **A2**

(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ**

(43) Дата публикации заявки  
2023.02.28

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

(22) Дата подачи заявки  
2010.11.05

---

(54) **ГЕНОМНЫЙ АНАЛИЗ НА ОСНОВЕ РАЗМЕРОВ**

---

(31) 61/259,076; 61/360,399

(32) 2009.11.06; 2010.06.30

(33) US

(62) 201992444; 2010.11.05

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

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

(74) Представитель:  
**Фелицына С.Б. (RU)**

---

(57) Описываются системы, способы и аппаратура для выполнения пренатальной диагностики дисбаланса последовательностей. Признаком дисбаланса в некоторых обстоятельствах может быть сдвиг (например, в сторону более узкого распределения по размерам). Например, распределение по размерам фрагментов нуклеиновых кислот из хромосомы "в группе риска" можно использовать для определения фетальной хромосомной анеуплоидии. Ранжировку по размеру различных хромосом можно использовать для определения отклонений ранжира хромосомы "в группе риска" от ожидаемой ранжировки. Также различие между статистическим показателем размеров для одной хромосомы можно сравнить со статистическим показателем размеров другой хромосомы для идентификации значимого сдвига в размере. Также можно определить генотип и гаплотип плода, используя распределение по размерам для определения того, имеет ли место дисбаланс последовательностей в материнском образце относительно генотипов или гаплотипа матери, причем посредством этого получают генотип или гаплотип плода.

---

202293523

A2

A2

202293523

## ГЕНОМНЫЙ АНАЛИЗ НА ОСНОВЕ РАЗМЕРОВ

### Перекрестные ссылки на родственные заявки

Настоящая заявка притязает на приоритет и является окончательным вариантом предварительной заявки на патент США № 61/259076, озаглавленной «Обнаружение хромосомного дисбаланса», поданной 6 ноября 2009, и предварительной заявки на патент США № 61/360399, озаглавленной «Геномный анализ на основе размеров», поданной 30 июня 2010, включенных в данное описание в качестве ссылок для всех целей.

Настоящая заявка также является родственной заявке на патент США № 12/178181, озаглавленной «Диагностика фетальной хромосомной анеуплоидии с использованием массивного параллельного секвенирования генома», поданной 23 июля 2008 (номер в патентном реестре № 016285-005220US); заявке на патент США № 12/614350, озаглавленной «Диагностика фетальной хромосомной анеуплоидии с использованием массивного параллельного секвенирования генома с обогащением» (номер в патентном реестре № 016285-005221US), и одновременно поданной заявке на патент США, озаглавленной «Геномный анализ плода из материнского биологического образца» (номер в патентном реестре № 016285-006710US), включенных в данное описание в качестве ссылок для всех целей.

### Уровень техники

Открытие фетальной ДНК в материнской плазме в 1997 открывает новые возможности для неинвазивной пренатальной диагностики (Lo et al., *Lancet*, 1997, 350: 485-487). Такая технология быстро переместилась в клинические применения с детекцией происходящих от плода унаследованных от отца генов или последовательностей, например, для определения пола плода и для определения статуса RhD плода. Однако применение в пренатальной диагностике, включающее анализ геномных мишеней, которые присутствуют как в материнском геноме, так и геноме плода, например, 21 хромосомы, является гораздо более многообещающим.

Недавно показано, что методы подсчета отдельных молекул с их превосходной количественной точностью могут быть перспективным решением такой проблемы (Lo et al., *Proc. Natl. Acad. Sci. USA*, 2007, 104: 13116-13121; Fan et al., *Anal. Chem.*, 2007, 79: 7576-7579; заявка на патент США 11/701686; Chiu et al., *Trends Genet.*, 2009, 25: 324-331; Chiu et al., *Proc. Natl. Acad. Sci. USA*, 2008, 105: 20458-20463; Fan et al., *Proc. Natl. Acad. Sci. USA*, 2008, 105: 16266-16271). В таких методах целей диагностики достигают через

наблюдение количественных различий в числе молекул из выбранных геномных локаций между болезненным и здоровым состоянием. Например, для диагноза синдрома Дауна у плода число молекул из хромосомы 21 будет возрастать, когда плод страдает от трисомии 21 (Chiu et al., Proc. Natl. Acad. Sci. USA, 2008, 105: 20458-20463; Fan et al., Proc. Natl. Acad. Sci. USA, 2008, 105: 16266-16271).

Однако такие методы подсчета могут страдать от ограниченного числа полученных данных или других недостатков. Поэтому желательно предоставить новые методы, системы и приборы для выполнения пренатальной диагностики, имеющие определенные преимущества перед существующими методами.

### Сущность изобретения

Некоторые воплощения настоящего изобретения могут предоставить системы, методы и аппаратуру, которые можно использовать в анализе на основе размеров для выполнения пренатальной диагностики дисбаланса последовательностей (например, фетальной хромосомной анеуплоидии) в биологическом образце, полученном от беременной женщины. Например, распределение по размерам фрагментов молекул нуклеиновых кислот для хромосомы «в группе риска» можно использовать для определения фетальной хромосомной анеуплоидии. Некоторые воплощения также могут позволить обнаруживать другие дисбалансы последовательностей, такие как дисбаланс в биологическом образце (содержащем ДНК матери и плода), где дисбаланс соотносится с генотипом, мутационным статусом или гаплотипом матери. Такой дисбаланс можно определить с помощью распределения по размерам фрагментов (молекул нуклеиновых кислот), соответствующих определенной последовательности, относительно с ожидаемым распределением по размерам, когда образец является чисто материнским, а не от плода и матери. Сдвиг (например, к более узкому распределению по размерам) может быть признаком дисбаланса при некоторых обстоятельствах.

В одном воплощении для определения дисбаланса используют ранжировку распределения по размеру (например, статистического показателя, представляющего распределение по размеру) фрагментов из соответствующих хромосом относительно друг друга. Например, ранжировку по размеру фрагментов хромосомы «в группе риска» в испытываемом образце можно сравнить с ранжировкой для хромосомы «в группе риска», которую получают из эталонного биологического образца. На основе сравнения можно выполнить диагностику. Как пример, если ранжировка изменяется (например, показывая уменьшение размера фрагментов нуклеиновых кислот) на определенную величину, тогда можно поставить диагноз, что в хромосоме «из группы риска» существует фетальная

хромосомная анеуплоидия. В различных воплощениях с использованием такого анализа ранжировки можно использовать все 22 аутосомы и половые хромосом или подмножество хромосом.

В другом воплощении используют различие между размером (т.е. статистическим показателем, представляющим распределение по размерам) фрагментов хромосомы «из группы риска» и размером фрагментов эталонной хромосомы. Например, если различие в размере больше или меньше, чем предельная величина (также называемая пороговой величиной), тогда можно диагностировать, что в хромосоме «из группы риска» существует фетальная хромосомная анеуплоидия.

Согласно одному примеру воплощения предоставляется способ выполнения пренатальной диагностики дисбаланса последовательностей в биологическом образце, полученном от беременной женщины. Биологический образец включает молекулы нуклеиновых кислот, которые являются частью нуклеотидных последовательностей. Измеряют размер каждой молекулы нуклеиновой кислоты из множества молекул нуклеиновых кислот в биологическом образце, и из этого идентифицируют нуклеотидную последовательность, из которой происходит молекула нуклеиновой кислоты. Компьютерная система определяет распределение по размерам молекул нуклеиновых кислот, соответствующих первой последовательности. На основании установленного распределения по размерам проводят классификацию, существует ли дисбаланс последовательности для первой последовательности.

Согласно другому примеру воплощения предоставляется способ выполнения пренатальной диагностики фетальной хромосомной анеуплоидии в биологическом образце, полученном от беременной женщины. Измеряют размер каждой молекулы нуклеиновой кислоты из множества молекул нуклеиновых кислот в биологическом образце, и из этого идентифицируют хромосому, из которой происходит молекула нуклеиновой кислоты. Компьютерная система вычисляет статистический показатель исходя из размеров молекул нуклеиновых кислот, соответствующих хромосоме. Статистический показатель вычисляют для каждой из множества хромосом. На основании статистического показателя определяют ранжировку хромосом. Установленную ранжировку первой хромосомы сравнивают с другой ранжировкой первой хромосомы, полученной из эталонного биологического образца. На основании сравнения проводят классификацию, существует ли фетальная хромосомная анеуплоидия для первой последовательности.

Согласно другому примеру воплощения предоставляется способ выполнения пренатальной диагностики фетальной хромосомной анеуплоидии в биологическом

образце, полученном от беременной женщины. Измеряют размер каждой молекулы нуклеиновой кислоты из множества молекул нуклеиновых кислот в биологическом образце, и из этого идентифицируют хромосому, из которой происходит молекула нуклеиновой кислоты. Компьютерная система вычисляет первый статистический показатель исходя из размеров молекул нуклеиновых кислот, соответствующих первой хромосоме. Компьютерная система вычисляет второй статистический показатель исходя из размеров молекул нуклеиновых кислот, соответствующих одной или нескольким другим хромосомам. Определяют показатель разделения между первым статистическим показателем и вторым статистическим показателем. Показатель разделения сравнивают с одной или несколькими пороговыми величинами. На основании сравнения проводят классификацию, существует ли фетальная хромосомная анеуплоидия для первой последовательности.

Другие воплощения изобретения относятся к системам и машиночитаемым носителям информации, связанным со способами, описанными в данном описании. В одном воплощении машиночитаемый носитель информации содержит команды по восприятию данных и анализу данных, но отсутствуют команды по управлению механизмом для получения данных (например, секвенированию молекул нуклеиновых кислот). В другом воплощении машиночитаемый носитель информации содержит команды по управлению механизмом для получения данных. В одном воплощении компьютерный программный продукт включает машиночитаемый носитель информации, содержащий набор команд по регулированию процессора для выполнения операции, для способов, описанных в данном описании. Воплощения также относятся к компьютерным системам, сформированным для выполнения стадий любого из способов, описанных в данном описании, возможно, с различными компонентами, выполняющими соответствующую стадию или соответствующую группу стадий.

Обращение к остальным частям описания, включая чертежи и формулу изобретения, делает наглядными другие особенности и преимущества воплощений настоящего изобретения. Другие особенности и преимущества, а также структура и работа различных воплощений настоящего изобретения подробно описаны ниже с учетом прилагаемых чертежей. На чертежах схожие номера для ссылки могут указывать на идентичные или функционально схожие элементы.

#### Краткое описание чертежей

Фиг. 1 представляет собой блок-схему, иллюстрирующую способ 100 выполнения пренатальной диагностики дисбаланса последовательностей в биологическом образце,

полученном от беременной женщины, согласно воплощениям настоящего изобретения.

Фиг. 2 представляет собой график, показывающий срединную ранжировку хромосом распределенную в соответствии с размерами последовательностей, выровненных с ними (когда используют вариант 1 набора реагентов генерации кластеров Illumina), согласно воплощениям настоящего изобретения.

Фиг. 3 представляет собой график, показывающий, что анализ размеров последовательностей, происходящих от различных хромосом в материнской плазме, согласно воплощениям настоящего изобретения можно использовать для неинвазивного пренатального обнаружения фетальных анеуплоидий хромосомы 21.

Фиг. 4 представляет собой блок-схему, иллюстрирующую способ выполнения пренатальной диагностики фетальной хромосомной анеуплоидии в биологическом образце, полученном от беременной женщины, с использованием ранжировки статистического показателя размера, согласно воплощениям настоящего изобретения.

Фиг. 5 представляет собой таблицу, иллюстрирующую сравнение хромосомы 21 с хромосомой 7 и хромосомой 14 по размеру последовательностей, выровненных с ними, согласно воплощениям настоящего изобретения.

Фиг. 6 представляет собой блок-схему, иллюстрирующую способ выполнения пренатальной диагностики дисбаланса последовательностей в биологическом образце, полученном от беременной женщины, согласно воплощениям настоящего изобретения, с использованием сравнения статистического показателя размера фрагментов для геномной локации.

Фиг. 7 показывает график 700 доли общей длины, вносимой короткими фрагментами (F), против порогового значения (w), согласно воплощениям настоящего изобретения.

Фиг. 8 показывает график 800 доли общей длины, вносимой короткими фрагментами (F), для хромосомы 21 (сплошная линия) и эталонных хромосом (все аутосомы за исключением хромосом 13, 18 и 21) (пунктирная линия), против порогового значения в случае зуплоидной беременности, согласно воплощениям настоящего изобретения.

Фиг. 9 показывает график 900 величин F для хромосомы 21 (сплошная линия) и эталонных хромосом (все аутосомы за исключением хромосом 13, 18 и 21) (пунктирная линия) против порогового значения в случае беременности с трисомией 21, согласно воплощениям настоящего изобретения.

Фиг. 10 показывает график 1000  $\Delta F_{(21-Ref)}$  между хромосомой 21 и эталонными хромосомами (все аутосомы за исключением хромосом 13, 18 и 21) против порогового

значения в случае эуплоидной беременности и беременности с трисомией 21, согласно воплощениям настоящего изобретения.

Фиг. 11 показывает таблицу 1100 для анатомического пола и классификацию 120 беременностей согласно воплощениям настоящего изобретения.

Фиг. 12 иллюстрирует ранжировку различных хромосом для четырех образцов с различными заболеваниями или состояний без заболеваний согласно воплощениям настоящего изобретения.

Фиг. 13 показывает ранжировку хромосомы 13 для 120 случаев эуплоидии, трисомии 13, трисомии 18 и трисомии 21 согласно воплощениям настоящего изобретения.

Фиг. 14 показывает ранжировку хромосомы 18 для 120 случаев эуплоидии, трисомии 13, трисомии 18 и трисомии 21 согласно воплощениям настоящего изобретения.

Фиг. 15 показывает ранжировку хромосомы 21 для 120 случаев эуплоидии, трисомии 13, трисомии 18 и трисомии 21 согласно воплощениям настоящего изобретения.

Фиг. 16 представляет собой таблицу, иллюстрирующую сравнение хромосомы 13 с хромосомой 5 и хромосомой 6 по размеру последовательностей, выровненных с ними, согласно воплощениям настоящего изобретения. Сравнение хромосом 5 и 6 с хромосомой 13 используют для обнаружения изменений в размере последовательностей хромосомы 21 при беременности с трисомией 13. Результаты при беременности с эуплоидией и трисомией 18 включены для сравнения.

Фиг. 17 представляет собой таблицу, иллюстрирующую сравнение хромосомы 18 с хромосомой 12 и хромосомой 14 по размеру последовательностей, выровненных с ними, согласно воплощениям настоящего изобретения. Результаты при беременности с эуплоидией и трисомией 13 включены для сравнения согласно воплощениям настоящего изобретения.

Фиг. 18 показывает различие в долях общей длины, вносимых короткими фрагментами, между хромосомой 18 и эталонными хромосомами ( $\Delta F_{(18-Ref)}$ ) в 150 п.о., согласно воплощениям настоящего изобретения.

Фиг. 19 показывает различие в долях общей длины, вносимых короткими фрагментами, между хромосомой 21 и эталонными хромосомами ( $\Delta F_{(21-Ref)}$ ), в 150 п.о., согласно воплощениям настоящего изобретения.

Фиг. 20 показывает GC-состав различных хромосом (NCBI build 36, версия 48) согласно воплощениям настоящего изобретения.

Фиг. 21 показывает различие в долях общей длины, вносимых короткими фрагментами, между хромосомой 13 и эталонными хромосомами ( $\Delta F_{(13-ref)}$ ), в 150 п.о., согласно воплощениям настоящего изобретения.

Фиг. 22 представляет собой график, показывающий корреляцию между средним размером последовательностей, выровненных с хромосомой 21, и процентом последовательностей, выровненных с хромосомой Y, согласно воплощениям настоящего изобретения.

Фиг. 23А-23С представляют собой графики, показывающие корреляцию между средним размером последовательностей, выровненных соответственно с хромосомами 18, 13 и 21, и процентом последовательностей, выровненных с хромосомой Y, согласно воплощениям настоящего изобретения.

Фиг. 24 показывает сравнение точности воплощения настоящего изобретения и другого способа неинвазивного обнаружения фетальной анеуплоидии (трисомия 13 и трисомия 18) с использованием анализа ДНК материнской плазмы.

Фиг. 25А-25С показывают диаграммы различных вариантов генотипов беременной женщины и плода согласно воплощениям настоящего изобретения.

Фиг. 26 показывает пример, когда мать является гетерозиготной, а отец является гомозиготным, согласно воплощениям настоящего изобретения.

Фиг. 27 показывает пример, когда плод наследует  $\text{Hap I}$  от матери, когда родительский гаплотип такой, как показанный на фиг. 26, согласно воплощениям настоящего изобретения.

Фиг. 28 показывает таблицу, иллюстрирующую анализ размеров для однонуклеотидного полиморфизма (SNP) типа  $\alpha$  на хромосоме 22 согласно воплощениям настоящего изобретения.

Фиг. 29 показывает таблицу, иллюстрирующую анализ размеров для SNP типа  $\beta$  на хромосоме 22 согласно воплощениям настоящего изобретения.

Фиг. 30 показывает график  $\Delta F_{(\text{Hap I}-\text{Hap II})}$  для SNP типа  $\alpha$  и типа  $\beta$  на хромосоме 22 согласно воплощениям настоящего изобретения.

Фиг. 31А представляет собой таблицу, которая представляет анализ размеров ДНК плазмы без обогащения мишени согласно воплощениям настоящего изобретения.

Фиг. 31В представляет собой таблицу, которая представляет анализ размеров ДНК плазмы с обогащением мишени согласно воплощениям настоящего изобретения.

Фиг. 32 представляет собой график  $\Delta F$  для T21 и эуплоидных образцов с и без обогащения мишени согласно воплощениям настоящего изобретения.

Фиг. 33 представляет собой блок-схему примера компьютерной аппаратуры, применимой с системой и способами согласно воплощениям настоящего изобретения.



## Определения

Термин «биологический образец», используемый в данном описании, относится к любому образцу, взятому у субъекта (например, человека, такого как беременная женщина), и содержащему одну или несколько молекул нуклеиновой кислоты, представляющих интерес.

Термин «нуклеиновая кислота» или «полинуклеотид» относится к дезоксирибонуклеиновой кислоте (ДНК) или рибонуклеиновой кислоте (РНК) и их полимерам или в одно- или в двухцепочечной форме. Если нет особых ограничений, термин охватывает нуклеиновые кислоты, содержащие известные аналоги природных нуклеотидов, которые имеют схожие свойства связывания с эталонной нуклеиновой кислотой и метаболизируются схоже с нуклеотидами, встречающимися в природе. Если не указано иное, определенная нуклеотидная последовательность также потенциально охватывает ее консервативно модифицированные варианты (например, замены вырожденных кодонов), аллели, ортологи, SNP, варианты копийности и комплементарные последовательности, а также точно указанные последовательности. Конкретно, замены вырожденных кодонов можно достичь путем генерации последовательностей, в которых третья позиция одного или нескольких выбранных (или всех) кодонов заменена смешанными основными и/или дезоксиинозиновыми остатками (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» используют как взаимозаменяемый с геном, кДНК, мРНК, небольшой некодирующей РНК, микро-РНК (миРНК), взаимодействующей с Рiwi РНК и короткой шпилечной РНК (sh-РНК), кодированной геном или локусом.

Термин «ген» обозначает сегмент ДНК, вовлеченной в продуцирование полипептидной цепи или транскрибированного продукта РНК. Он может включать участки, предшествующие и следующие за кодирующим участком (лидерные и трейлерные), а также вставочные последовательности (интроны) между отдельными кодирующими сегментами (экзонами).

Термин «клинически релевантная нуклеотидная последовательность» (также называемая последовательностью-мишенью или хромосомой), используемый в данном описании, может относиться к нуклеотидной последовательности, соответствующей сегменту более крупной геномной последовательности, потенциальный дисбаланс которой проверяют, или к самой более крупной геномной последовательности. Одним из примеров является последовательность хромосомы 21. Другие примеры включают хромосому 18, 13, X и Y. Еще другие примеры включают мутированные генетические

последовательности или генетические полиморфизмы или варианты копийности, которые плод может унаследовать от одного или обоих родителей, или как мутацию *de novo* у плода. В некоторых воплощениях несколько клинически релевантных нуклеотидных последовательностей или, равнозначно, несколько маркеров клинически релевантной нуклеотидной последовательности могут быть использованы для получения данных для обнаружения дисбаланса. Например, данные по пяти несмежным последовательностям на хромосоме 21 можно использовать в дополнение для определения возможного дисбаланса хромосомы 21, причем объем образца необходимый для анализа эффективно уменьшается до 1/5.

Термин эталонная «нуклеотидная последовательность, используемый в данном описании, относится к нуклеотидной последовательности, распределение по размерам которой используют для сравнения против последовательности-мишени. Примеры эталонной нуклеотидной последовательности включают хромосому, часть хромосомы, определенный аллель (например, матери), определенный гаплотип, геном или искусственно синтезированную нуклеотидную последовательность. Такие эталонные нуклеотидные последовательности могут или существовать эндогенно в образце или добавляться экзогенно во время обработки или анализа образца. В некоторых воплощениях эталонные нуклеотидные последовательности демонстрируют профиль размеров, который характерен для здорового состояния в отсутствие заболевания.

Термин «основан на», используемый в данном описании, означает «основанный, по меньшей мере, частично, на» и относится к одному показателю (или результату), используемому при определении другого показателя (величины), как происходит в соотношении входа способа и выхода указанного способа. Термин «происходить», используемый в данном описании, также относится к соотношению входа способа и выхода указанного способа, такого, какое имеет место, когда происхождение является вычислением по формуле.

Термин «параметр», используемый в данном описании, означает численную величину, которая характеризует набор количественных данных, и/или численное соотношение между наборами количественных данных. Например, параметром является отношение (или функция отношения) между первым количеством первой нуклеотидной последовательности и вторым количеством второй нуклеотидной последовательности.

Используемый в данном описании термин «локус» или его множественное число «локусы» отображает локацию или местоположение любого отрезка нуклеотидов (или пар оснований), который имеет вариацию в геномах.

Термин «дисбаланс последовательностей», используемый в данном описании,

означает любое значимое отклонение, определяемое как, по меньшей мере, одна пороговая величина, в количестве клинически релевантной нуклеотидной последовательности от количества эталона. Дисбаланс последовательностей может включать дисбаланс дозы хромосом, дисбаланс аллелей, дисбаланс дозы мутаций, дисбаланс копийности, дисбаланс дозы гаплотипов и другие подобные дисбалансы. Как пример, дисбаланс аллелей или дозы мутаций может иметь место, когда плод имеет генотип, отличный от материнского, причем посредством этого создается дисбаланс в образце в определенном локусе.

Термин «хромосомная анеуплоидия», используемый в данном описании, означает изменение в количестве хромосом от количества диплоидного генома. Изменение может быть увеличением или уменьшением. Оно может включать одну хромосому целиком или участок хромосомы.

Термин «гаплотип», используемый в данном описании, относится к комбинации аллелей в нескольких локусах, которые переданы по наследству вместе на одной и той же хромосоме или участке хромосомы. Гаплотип может относиться к такому небольшому количеству, как одна пара локусов, или к участку хромосомы или ко всей хромосоме. Термин «аллели» относится к альтернативным последовательностям ДНК в одном и том же физическом геномном локусе, которые могут или не могут привести к различным фенотипическим признакам. В любом определенном диплоидном организме с двумя копиями каждой хромосомы (за исключением половых хромосом у мужчины) генотип для каждого гена включает пару аллелей, присутствующих в указанном локусе, которые являются одинаковыми в гомозиготах и различными в гетерозиготах. Популяция или вид организмов типично включает несколько аллелей для каждого локуса у различных индивидуумов. Геномный локус, в котором в популяции находят более одного аллеля, называют полиморфным сайтом. Изменение аллелей в локусе можно измерить как число присутствующих аллелей (т.е. степень полиморфизма) или долю гетерозигот (т.е. степень гетерозиготности) в популяции. Используемый в данном описании термин «полиморфизм» относится к любому межличностному изменению в геноме человека независимо от его повторяемости. Примеры таких изменений включают, но не ограничиваются указанным, однонуклеотидный полиморфизм, полиморфизмы простых tandemных повторов, инсерционные-делеционные полиморфизмы, мутации (которые могут вызывать заболевания) и изменения копийности.

### Подробное описание

Обнаружено, что молекулы фетальной ДНК, которые присутствуют в плазме беременной женщины, как правило, короче, чем молекулы материнского происхождения (Chan et al., Clin. Chem., 2004, 50: 88-92; Li et al., Clin. Chem., 2004, 50: 1002-1011; заявка на патент США 20050164241). В некоторых воплощениях настоящего изобретения из ДНК материнской плазмы по изменению размера молекул ДНК, полученных из указанной части генома, можно определить, имеет ли плод избыточное или недостаточное представление этой части генома. Так как фетальная ДНК представляет минорную фракцию ДНК в материнской плазме, общая степень изменения размеров в материнской плазме должна быть, вероятно, небольшой, и таким образом, может быть обнаружена с трудом. В некоторых воплощениях размеры ряда молекул измеряют для достижения статистически значимого различия между болезнью и здоровым состоянием.

#### I. Определение размера фрагмента

Одним из способов, которыми можно измерить размеры ряда молекул ДНК, является массивное параллельное секвенирование генома. Его можно осуществить, например, на платформе геномного анализатора Illumina (с использованием секвенирования путем синтеза) (Bentley D.R. et al., Nature, 2008, 456: 53-59), ABI SOLiD (с использованием секвенирования путем лигирования) (McKernan et al., Genome Res., 2009, 19: 1527-1541), платформе Roche 454 (Marguelis et al., Nature, 2005, 437: 376-380) и платформе одномолекулярного секвенирования Helicos (Harris et al., Science, 2008, 320: 106-109). Также предполагается, что также можно использовать другие платформы массивного параллельного секвенирования, например, Pacific Biosciences (одномолекулярная технология в реальном времени (SMRT™)) (Eid et al., Science, 2009, 323: 133-138), секвенирование в нанопорах (Clarke J. et al., Nat. Nanotechnol., 2009, 4: 465-470), полупроводниковое секвенирование (например, с Ion Torrent ([www.iontorrent.com](http://www.iontorrent.com))), и т.д.

Одним из путей получения информации о размерах ДНК из такого секвенирования является выполнение секвенирования спаренных концов (paired-end) (PE), при котором секвенируют оба конца молекулы ДНК. Затем последовательности, соответствующие обоим концам молекулы, можно «обратно» картировать в эталонном геноме (например, эталонном геноме человека или эталонном геноме лошади или геноме любого животного, представляющего интерес). В одном воплощении оба конца секвенируют, каждый, на расстояние, которое является достаточно протяженным для обратного картирования в эталонном геноме человека (например, примерно 10-24 основания или 25-36 оснований). В другом воплощении только часть последовательностей может быть обратно

картирована без ошибок в участке генома человека не содержащем повторы. В одном аспекте картирование может быть однозначным, если обе последовательности используют при картировании вместе. При таком варианте, даже если оба конца без сомнения могут быть слишком короткими для обратного картирования, использование обеих последовательностей может обеспечить однозначное картирование. Размер молекулы может быть вычислен путем вычитания геномных координат концов двух последовательностей.

В другом воплощении размер молекулы можно получить путем полного или почти полного секвенирования всей молекулы ДНК вместо только ее двух концов. Это можно эффективно осуществить с помощью платформ секвенирования с относительно протяженными отрезками считывания, такими как платформа Roche 454, одномолекулярная технология в реальном времени (SMRT™) Pacific Biosciences и технология Ion Torrent ([www.iontorrent.com](http://www.iontorrent.com)).

Производительность методов секвенирования на основе вышеуказанного можно повысить путем использования индексирования или штрихового кода (Cronn et al., *Nucleic Acids Res.*, 2008, 36: e122). Таким образом, к фрагментам нуклеиновой кислоты в определенной библиотеке секвенирования нуклеиновых кислот может быть добавлен индекс или штриховой код, специфический для образца или пациента. Затем ряд таких библиотек, каждая с индексом или штриховым кодом, специфическим для образца или пациента, смешивают вместе и вместе секвенируют. После реакций секвенирования можно собрать данные секвенирования для каждого образца или пациента на основании кода или индекса. Такая стратегия может повысить производительность и, таким образом, экономическую эффективность данного изобретения.

В другом воплощении молекулы нуклеиновых кислот в биологическом образце перед анализом размеров можно отобрать или фракционировать. В одном варианте молекулы нуклеиновых кислот обрабатывают с помощью устройства (например, микроматрицы или растворов, содержащих зонды), которое может преимущественно связывать молекулы нуклеиновых кислот из выбранных локусов в геноме (например, одной из хромосом 21, 18, 13 или X), затем можно выполнить анализ размеров на подмножестве связанных молекул нуклеиновых кислот. В таком воплощении можно использовать систему захвата последовательностей Nimblegen ([www.nimblegen.com/products/seqcap/index.html](http://www.nimblegen.com/products/seqcap/index.html)) или систему Agilent SureSelect Target Enrichment ([www.opengenomics.co.SureSelect\\_Target\\_Enrichment\\_System](http://www.opengenomics.co.SureSelect_Target_Enrichment_System)) или подобные платформы. В другом воплощении подмножество несвязанных молекул нуклеиновых кислот может быть дифференцированно удалено или разрушено или гидролизировано.

По меньшей мере, по некоторым воплощениям можно работать с анализом одной молекулы на любой платформе, в котором можно анализировать хромосомное происхождение и длину молекулы, например, с электрофорезом, оптическими методами (например, оптическим картированием и его вариантами, [en.wikipedia.org/wiki/Optical\\_mapping#cite\\_note-Nanocoding-3](http://en.wikipedia.org/wiki/Optical_mapping#cite_note-Nanocoding-3), и Jo et al., Proc. Natl. Acad. Sci. USA, 2007, 104: 2673-2678), методом на основе флуоресценции, методами на основе зондов, цифровой ПЦР (на основе микрожидкостей или на основе эмульсии, например, BEAMing (Dressman et al., Proc. Natl. Acad. Sci. USA, 2003, 100: 8817-8822), RainDance ([www.raindancetech.com/technology/per-genomic-research.asp](http://www.raindancetech.com/technology/per-genomic-research.asp))), амплификацией по типу катящегося кольца, масс-спектрометрией, анализом плавления (или анализом кривой плавления), с молекулярными ситами и т.д. Как пример, в случае масс-спектрометрии более длинная молекула будет иметь бóльшую массу (пример показателя размера).

В одном примере молекулы ДНК плазмы произвольно секвенируют с помощью системы геномного анализатора Illumina с использованием протокола секвенирования спаренных концов. В таком эксперименте используют версию 1 набора реагентов Illumina Paired-end (PE) Cluster Generation. Каждый конец секвенируют для 36 п.о. Оба конца каждой последовательности выравнивают с геномом человека с маскированными повторами (NCBI Build 36, версия 48) с использованием программы `cland_pair` в пакете программ GAPIipeline-1.0, предоставленном Illumina. Для целей выравнивания используют только 32 п.о. из 36 с каждого конца.

В некоторых воплощениях для последующего анализа можно использовать данные PE, удовлетворяющие следующим критериям: (1) отдельные члены каждой предложенной пары оба секвенированы на одной и той же позиции кластера на секвенирующей проточной камере и могут быть выровнены с одной и той же хромосомой с правильной ориентацией, ожидаемой для эталонного генома человека; (2) данные секвенирования обоих членов пары могут быть выровнены с эталонным геномом человека с маскированными повторами без какой-либо ошибки по нуклеотидам; (3) данные секвенирования каждого члена пары имеют показатель однозначности  $>4$ ; и (4) пары, показывающие размер вставки менее 600 п.о. Затем вычисляют размер каждой выровненной последовательности согласно позиции каждого из двух концов.

II. Использование распределения по размерам для определения состояния анеуплоидии

Фиг. 1 представляет собой блок-схему, иллюстрирующую способ 100 осуществления пренатальной диагностики дисбаланса последовательностей в биологическом образце, полученном от беременной женщины, согласно воплощениям

настоящего изобретения. Хотя способ 100 описан, главным образом, в связи с анализом фетальной хромосомной анеуплоидии, другие воплощения способа 100 и другие способы в данном описании могут быть применены к дисбалансам других последовательностей (например, идентификация генотипа или мутаций). Способ 100 и другие способы, указанные в данном описании, могут быть осуществлены полностью или частично с помощью компьютерной системы, включающей один или несколько процессоров.

Способ 100 и любой другой способ, описанный в данном описании, может быть осуществлен полностью или частично с помощью компьютерной системы, включающей процессор, конфигурация которого может быть создана для выполнения стадий. Таким образом, воплощения относятся к компьютерным системам, конфигурация которых создана для выполнения стадий любого из способов, описанных в данном описании, потенциально, с различными компонентами, выполняющими соответствующую стадию или соответствующую группу стадий. Хотя стадии представлены нумерованными, стадии способа по данному изобретению могут быть выполнены все одновременно или в другом порядке. Кроме того, все или части стадий могут быть необязательными. Кроме того, любая из стадий любого из способов может быть выполнена с модулями, схемами или другими средствами для выполнения таких стадий.

На стадии 110 измеряют размер, по меньшей мере, некоторых молекул нуклеиновых кислот (например, ДНК или РНК) в биологическом образце. Молекулы нуклеиновых кислот также называются фрагментами, так как они представляют собой фрагменты полного генома. Размер можно измерить любым подходящим способом, например, способами, описанными выше.

На стадии 120 идентифицируют локацию в геноме, из которого происходит каждая из молекул нуклеиновых кислот. Локацией может являться любая часть генома, который в данных примерах является геномом человека, но может быть частью других геномов. Например, локацией может являться хромосома с определенным номером из которой произошел фрагмент, где часть хромосомы может быть определена геномными координатами (например, конкретной координатой или рядом координат), и даже может представлять собой одну определенную хромосому из хромосомной пары (исходя из того, что организм эуплоид) из которой произошел фрагмент.

В одном воплощении такую идентификацию можно выполнить путем секвенирования и сравнения информации о последовательностях с эталонной последовательностью генома человека. В другом воплощении такую идентификацию можно выполнить путем гибридизации с панелью зондов с известным хромосомным происхождением. Зонды могут быть помечены одной или несколькими флуоресцентными

метками в любом формате микроматриц или в растворе. В еще одном воплощении молекулы нуклеиновых кислот могут быть захвачены панелью зондов или в растворе или на твердой поверхности и, затем захваченные (или оставшиеся незахваченными) молекулы нуклеиновых кислот секвенируют. В некоторых воплощениях, где необходимо идентифицировать дисбаланс последовательностей иной, чем хромосомная анеуплоидия, стадия идентификации хромосомы, из которой происходит фрагмент, может быть необязательной.

На стадии 130 определяют распределение по размерам молекул нуклеиновых кислот, соответствующих первой локации (например, первой хромосоме). Различные воплощения можно использовать для различных распределений по размерам. В некоторых воплощениях распределение по размерам соотносят с ранжированием размеров (например, усредненный (average), срединный (median), средний (mean)) фрагментов одной хромосомы относительно фрагментов других хромосом. В других воплощениях распределение по размерам может соотноситься со статистическим показателем фактических размеров фрагментов хромосомы. В одном воплощении статистический показатель может включать усредненный, срединный или средний размер фрагментов хромосомы. В другом воплощении статистический показатель может включать полную длину фрагментов меньше пороговой величины, которую можно разделить на полную длину всех фрагментов или, по меньшей мере, фрагментов меньше пороговой величины.

На стадии 140 на основании определенного распределения по размерам классифицируют, существует ли дисбаланс последовательностей (например, фетальная хромосомная анеуплоидия) для первой локации. В одном воплощении ранжировку хромосомы сравнивают с ранжировкой эталонного образца (например, ранжировкой эуплоидного образца). Если изменение значимое (например, превышающее пороговое значение), тогда образец можно классифицировать как анеуплоидный. В другом воплощении сравнивают статистические показатели фактических размеров двух хромосом или групп хромосом. Например, можно взять различие между соответствующими статистическими показателями и сравнить с пороговым значением.

## II. Распределение по размеру (ранжировка)

В воплощениях могут быть использованы значения размеров фрагментов нуклеиновых кислот образца для того, чтобы определить существует ли хромосомный дисбаланс. Фрагментами также называют последовательности после выполнения секвенирования. В одном воплощении определяют распределение по размерам фрагментов нескольких хромосом, и хромосомы располагают в определенном порядке на основании статистического показателя распределения (например, усредненного,



срединного или среднего). Для удобства термин «размер» можно использовать в данном описании как синоним статистического показателя размера. Это должно быть ясно, когда термин «размер» относят к размеру конкретного фрагмента и статистической оценке набора фрагментов.

#### А. Ранжировка

Фиг. 2 представляет собой график, иллюстрирующий измеренное распределение по размеру фрагментов ДНК, происходящих из различных хромосом в материнской плазме. Так как измеренный размер может отражать не только размер фрагментов ДНК *in vivo*, но также вклад аналитических стадий, предполагается, что распределение по размерам может изменяться от платформы к платформе (например, для платформы для геномного анализатора Illumina и для платформы ABI SOLID) и даже может изменяться, когда для определенной платформы используют различные реагенты. Однако до тех пор, пока эталонные образцы и испытываемые образцы анализируют с использованием одной и той же платформы или типов реагентов, воплощения можно считать как независимые от платформы/реагентов. Некоторые воплощения также можно использовать с различными платформами и/или типами реагентов, например, если можно определить и скорректировать любые ошибки, или если можно показать, что платформы и/или типы реагентов плотно соответствуют аналитическим действиям.

На фиг. 2 сравнивают 22 аутосомы и хромосому X по размерам последовательностей выровненных с ними фрагментов. По оси Y ранжир предстает в порядке убывания размера последовательностей, т.е. хромосоме с самой длинной последовательностью присваивают ранжир 1, и хромосоме с самой короткой последовательностью присваивают ранжир 23. В одном воплощении сравнение (Kruskal-Wallis One Way Analysis of Variance on Ranks followed by Bonferroni-corrected pairwise comparisons) осуществляют непараметрически с использованием программы SigmaStat (SPSS). Сравнение может быть по любому статистическому показателю, представляющему размер, включая ранжировку каждого из отдельных размеров и статистический анализ отдельных ранжировок для последовательностей каждой хромосомы. В одном воплощении при нумерации по ранжиру допускается нумерация петель и пропусков.

В некоторых воплощениях каждую из последовательностей картируют на определенной хромосоме. Затем для каждой хромосомы определяют один или несколько статистических показателей для последовательностей, картированных на такой хромосоме. Иными словами, для размеров каждой хромосомы могут быть вычислены статистические показатели различных типов (например, усредненный, средний и

срединный). Затем соответствующие статистические показатели можно расположить в определенном порядке. Например, можно сравнить друг с другом усредненные размеры для каждой хромосомы. Если используют более одного статистического показателя, тогда такие статистические показатели можно объединить (например, согласно некоторой формуле, такой как усреднение по массе), и можно ранжировать такой комбинированный статистический показатель. В одном воплощении ранжировки статистических показателей для определенной хромосомы можно объединить (например, так, как указано для статистических показателей выше), и затем комбинированные статистические показатели можно сравнить друг с другом.

В других воплощениях все последовательности по отдельности выстраивают в определенном порядке согласно размеру. Иными словами, если существует 1000000 последовательностей, тогда идет ранжировка с нумерацией от 1 до одного миллиона с возможными петлями и пропусками. Затем могут быть добавлены ранжировки для всех последовательностей, картированных на одной и той же хромосоме. Сумму ранжировок можно разделить на число последовательностей, выровненных с определенной хромосомой, и получить среднюю ранжировку последовательностей для хромосомы. Хромосома с самой высокой средней ранжировкой последовательностей может быть маркирована как самая длинная (ранжиры 1, наивысшие по оси Y), и хромосома с самой низкой средней ранжировкой последовательностей может быть самой короткой (ранжиры 23, самые низкие по оси Y). В другом воплощении может быть определен срединный ранжир. Срединные ранжиры для пациентов с эуплоидным плодом мужского пола, эуплоидным плодом женского пола и плодом мужского пола с трисомией 21 показаны на фиг. 2.

По оси X хромосомы располагают в порядке убывания размеров по распределению по размерам фрагментов ДНК, произошедших от каждой хромосомы (за исключением хромосомы X (см. далее)). В одном воплощении для такого расположения используют только случаи эуплоидии. Хромосому самого большого измеренного размера (например, длины) располагают на левой стороне. Хромосому X размещают на диаграмме справа, поскольку ее ранжировка определяется полом плода.

Как указано выше, измеренный размер может изменяться от платформы к платформе (например, размер изменяется при переходе от системы Illumina к другой системе). Таким образом, в одном аспекте размер может называться «измеренным» размером в отличие от фактического размера. Размер даже может изменяться при переключении от одной версии набора Illumina в другой версии, например, когда размер изменяется при переходе от версии 1 набора для генерации кластера спаренных концов к

версии 2. В одном воплощении пользователи могут выполнять ранжировку для выбранной ими определенной системы.

Из фиг. 2 можно видеть, что средний ранжир хромосомы X ниже (т.е. становится более сжатым) в плазме беременной женщины, вынашивающей плод женского пола, при сравнении с женщиной, вынашивающей плод мужского пола. Объяснение данного наблюдения таково, что фрагменты ДНК, высвобождаемые плодом, более короткие, чем фрагменты от матери. Таким образом, за счет высвобождения двойной дозы хромосомы X плод женского пола в материнской плазме будет уменьшать общий измеренный размер фрагментов, происходящих от хромосомы X. Напротив, плод мужского пола будет способен высвобождать только одну дозу хромосомы X.

Также из фиг. 2 можно видеть, что ранжир хромосомы 21 снижается (т.е. становится более сжатым) в плазме беременной женщины, вынашивающей плод с трисомией 21, при сравнении с женщиной, вынашивающей эуплоидный плод. Объяснение данного наблюдения может быть снова обращено к измеренному размеру фрагментов в материнской плазме, происходящих от хромосомы 21. Напротив, эуплоидный плод будет способен высвобождать только две дозы хромосомы 21 на фетальную клетку.

Фиг. 3 представляет собой график 300, показывающий, что анализ размеров последовательностей в материнской плазме, происходящих от различных хромосом, согласно воплощениям настоящего изобретения можно использовать для неинвазивного пренатального обнаружения фетальных анеуплоидий хромосомы 21. В данном примере размер хромосомы представлен ранжиром по размеру при сравнении с другими хромосомами. Таким образом, большее число ранжира по размеру указывает, что хромосома в материнской плазме имеет более короткие по размерам фрагменты ДНК.

График 300 показывает, что число ранжира по размеру хромосомы 21 в материнской плазме больше (указывая, таким образом, на более короткие фрагменты ДНК) в случае беременности, включающей плод с трисомией 21 (ранжировка от ранжира 18 до 21), чем когда плод является эуплоидным (ранжировка от ранжира 9 до 18). Объяснение данного наблюдения таково, что фетальная ДНК короче материнской ДНК, и что дополнительная доза хромосомы 21 от плода с трисомией 21 может привести к общему сокращению статистического показателя последовательностей хромосомы 21 в материнской плазме.

#### В. Методы с использованием ранжира

Фиг. 4 представляет собой блок-схему, иллюстрирующую способ выполнения пренатальной диагностики фетальной хромосомной анеуплоидии в биологическом образце, полученном от беременной женщины, с использованием ранжировок

статистического показателя размера согласно воплощениям настоящего изобретения.

На стадии 410 измеряют размер каждой из множества молекул нуклеиновых кислот (фрагментов), полученных из биологического образца. Обращает внимание, что такое множество молекул нуклеиновых кислот может представлять собой подмножество всех полученных молекул нуклеиновых кислот. Множество молекул нуклеиновых кислот даже может представлять собой подмножество всех отсеквенированных молекул нуклеиновых кислот, когда осуществляют секвенирование в целях измерения размеров.

На стадии 420 идентифицируют, из каких хромосом происходит множество молекул нуклеиновых кислот. В различных воплощениях порядок стадий 410 и 420 может быть изменен на противоположный, или их осуществляют одновременно. Например, в контексте секвенирования спаренных концов геномное выравнивание последовательности может предоставить ее хромосомную локацию плюс ее длину (путем вычитания начальных и конечных геномных координат). В одном воплощении хромосому можно идентифицировать как на стадии 120.

На стадии 430 для каждой из множества хромосом вычисляют статистический показатель из размеров молекул нуклеиновых кислот, соответствующих хромосоме. Статистический показатель можно вычислить любым из способов, описанных в данном описании. Например, статистический показатель может включать результат начальной стадии ранжировки, описанной выше. В одном воплощении можно использовать только часть молекул нуклеиновых кислот, соответствующих любой определенной хромосоме.

На стадии 440 расположение хромосом в определенном порядке определяют на основании статистических показателей. В одном воплощении ранжировку можно определить с использованием основного алгоритма сортировки. В другом воплощении можно выполнить более сложное сравнение, такое как по Kruskal-Wallis – один из способов анализа дисперсии по ранжирам с последующими попарными сравнениями с учетом поправки по Bonferroni, или другими подходящими способами. В различных воплощениях ранжировки могут быть целыми, частями, реальными числами (например, в интервале) или буквенноцифровыми ранжировками, основанными на инициалах (например, A-X).

На стадии 450 определенную ранжировку первой хромосомы сравнивают с другой ранжировкой первой хромосомы, полученной от эталонного биологического образца. В одном воплощении сравнение представляет собой сравнение определенной ранжировки с пороговым значением (например, одной величиной или интервалом), который определяют из ранжировок одного или нескольких эталонных биологических образцов. Такое пороговое значение может быть, если ранжир 18 (или 19) или выше, такой, какую можно

определить из фиг. 3. В другом воплощении можно определить различие в ранжировках первой хромосомы между двумя образцами, и различие можно сравнить с пороговым значением. В одном воплощении анализируют эталонный биологический образец для того, чтобы определить, что он не содержит заболевания, представляющего интерес, и даже можно определить, что образец не содержит какого-либо релевантного заболевания, которое может вызывать проблемы с ранжировкой.

На стадии 460 на основании сравнения классифицируют, существует ли фетальная хромосомная анеуплоидия для первой хромосомы. В одном воплощении классификация может представлять собой бинарную классификацию заболевания или отсутствия заболевания. В другом воплощении классификация может быть тройной, при которой может быть использована промежуточная классификация. В еще одном воплощении классификация может включать вероятность частной классификации, и таким образом, быть эффективнее, чем просто три классификации.

### III. Распределение по размерам (статистические показатели размера)

В других воплощениях распределение по размерам может включать статистические показатели размера фрагментов (например, статистические показатели фактических или абсолютных величин для определенной геномной локации) как противопоставление ранжировкам. В одном воплощении фактические размеры первой хромосомы можно сравнить с фактическими размерами одной или нескольких эталонных хромосом для одного и того же испытываемого образца. Например, показатель разделения (например, различие или отношение) между такими фактическими размерами первой хромосомы и одной или нескольких эталонных хромосом можно сравнить с пороговым значением. В одном воплощении пороговое значение можно определить из эталонных образцов. В другом воплощении для сравнения наряду с пороговым значением можно использовать показатель разделения между фактическими размерами фрагментов хромосомы между испытываемым образцом и эталонным биологическим образцом. В еще одном воплощении фактические размеры фрагментов хромосомы можно сравнить с пороговым значением и получить надежную классификацию.

#### A. Абсолютный размер

Некоторые примеры показывают, что можно добиться неинвазивного пренатального обнаружения трисомии 21 путем сравнения абсолютных размеров фрагментов, происходящих от хромосомы 21, с абсолютными размерами фрагментов, происходящих от одной или нескольких эталонных хромосом. В одном воплощении в качестве эталонных хромосом можно выбрать хромосомы 7 и 14, так как они могут иметь величины размеров (например, абсолютный размер или ранжировки по размеру) в

материнской плазме, которые относительно близки к хромосоме 21. На практике эталонная хромосома может представлять собой любую хромосому, которая имеет размеры фрагментов, которые соотносятся, например, для конкретной аналитической платформы и/или типа реагента, с хромосомой 21 (или другой хромосомой, представляющей интерес) для эуплоидных образцов.

Фиг. 5 представляет собой таблицу 500, иллюстрирующую сравнение хромосомы 21 с хромосомой 7 и хромосомой 14 по размерам последовательностей, выровненных с соответствующими хромосомами, согласно воплощениям настоящего изобретения. Данные для таблицы 500 получают при испытании 16 образцов. Для каждого образца показан средний размер для фрагментов каждой из хромосом 7, 14 и 21. Также приводится различие между средними значениями. Величина  $p$  показывает вероятность того, что каждое различие может иметь место в образце от здоровой женщины.

Как можно видеть из таблицы 500 на фиг. 5, последовательности, выровненные с хромосомой 21, значительно короче (например, по среднему размеру), чем последовательности, выровненные с хромосомой 7 и хромосомой 14, для всех случаев беременности с трисомией 21 (критерий суммы ранжиров Манна-Уитни, величина  $p < 0,001$ ). Сокращение при такой степени статистической значимости не наблюдают при эуплоидной беременности. Таким образом, таблица 500 показывает, что для всех случаев беременности с трисомией 21 различие в среднем размере фрагментов между хромосомой 21 и хромосомой 7 превышает 1 п.о., в то время как ни один из случаев эуплоидии не показывает различия больше 1 п.о.. Соответственно, 1 п.о. может дать точное пороговое значение для определения классификации. Подобным образом, для всех случаев беременности с трисомией 21 средний размер фрагментов из хромосомы 14 согласованно больше среднего размера фрагментов из хромосомы 21. Действительно, если используют «порог» в 0,5 п.о. для наблюдаемого «удлинения» фрагментов хромосомы 14 при сравнении с фрагментами хромосомы 21, все случаи трисомии 21 можно отличить от случаев без трисомии 21. Так, в одном воплощении пороговую величину можно определить из одного или нескольких эталонных образцов.

#### В. Способы с использованием абсолютного размера

Фиг. 6 представляет собой блок-схему, иллюстрирующую способ выполнения пренатальной диагностики дисбаланса последовательностей в биологическом образце, полученном от беременной женщины, с использованием сравнения статистического показателя размеров фрагментов для геномной локации согласно воплощениям настоящего изобретения. В одном аспекте способ 600 может относиться к определению классификации дисбаланса последовательностей на основании показателя разделения

(например, различия или отношения) для размеров фрагментов первой хромосомы и размеров фрагментов одной или нескольких эталонных хромосом.

На стадии 610 измеряют размер множества молекул нуклеиновых кислот, полученных из биологического образца. Обращает внимание, что множество молекул нуклеиновых кислот может быть получено и может включать схожие фрагменты, как отмечено на стадии 410.

На стадии 620 идентифицируют геномную локацию, из которой происходит каждая из молекул нуклеиновых кислот. Локацией может быть любая часть генома, как описано для стадии 120 и в других местах. Например, идентифицируют, от какой хромосомы происходит каждая из молекул нуклеиновых кислот. Такое определение можно сделать путем картирования в эталонном геноме, как описано в данном описании.

На стадии 630 вычисляют первый статистический показатель из размеров молекул нуклеиновых кислот, соответствующих первой геномной локации (например, первой хромосоме). В одном воплощении первый статистический показатель может представлять собой усредненный, средний или срединный размер фрагментов, соответствующих первой хромосоме. В другом воплощении первый статистический показатель может включать сумму длин фрагментов ниже первого размера, которая может представлять собой пороговую величину. Например, можно суммировать длины каждого из фрагментов, которые меньше 200 п.о. Сумму можно разделить на другое число, такое как сумма длин всех фрагментов, соответствующих первой хромосоме, или сумма длин фрагментов, превышающих пороговую величину второго размера (который может быть таким же, как первый размер). Например, первый статистический показатель может представлять собой отношение общей длины фрагментов ниже пороговой величины первого размера к общей длине фрагментов или отношение общей длины мелких фрагментов к общей длине крупных фрагментов.

На стадии 640 вычисляют второй статистический показатель из размеров молекул нуклеиновых кислот, соответствующих второй геномной локации (например, второй хромосоме). Вторую хромосому можно рассматривать как эталонную хромосому. В одном воплощении можно вычислить статистический показатель для множества эталонных хромосом. В одном воплощении статистические показатели можно объединить, так что может иметь место статистический показатель одной или нескольких вторых хромосом. В другом воплощении статистические показатели для множества эталонных хромосом можно сравнивать по отдельности, как указано ниже.

На стадии 650 первый статистический показатель и второй статистический показатель сравнивают, и получают показатель разделения. В одном воплощении

определяют показатель разделения, который может представлять собой различие между первым статистическим показателем и вторым статистическим показателем. В другом воплощении показатель разделения может представлять собой отношение двух статистических показателей. В еще одном воплощении можно определить множество показателей разделения, например, показатель для каждого второго статистического показателя, который можно вычислить для каждой эталонной хромосомы.

На стадии 660 показатель разделения сравнивают с одной или несколькими пороговыми величинами. В одном воплощении сравнение можно выполнить для каждого из множества показателей разделения. Например, как указано выше, можно определить другой показатель разделения между первым статистическим показателем и каждым вторым статистическим показателем. В различных воплощениях каждый показатель разделения можно сравнить с одной и той же или различными пороговыми величинами. В другом воплощении показатель разделения сравнивают с двумя пороговыми величинами и определяют, находится ли показатель разделения в определенном интервале. Интервал может включать одну пороговую величину для определения, если имеет место один из результатов, отличных от нормального (например, анеуплоидия), и можно использовать для определения вторую пороговую величину, если один из результатов возможно вызван ошибкой измерения или анализа (например, если показатель разделения выше, чем можно было бы ожидать даже для образца при заболевании).

На стадии 670 на основании сравнения определяют классификацию, существует ли дисбаланс последовательностей (например, фетальная хромосомная анеуплоидия) для первой геномной локации. В одном воплощении можно использовать множество пороговых величин (например,  $N$  пороговых величин) для одного показателя разделения. В таком воплощении можно определить  $N+1$  классификаций. Например, можно использовать две пороговые величины для определения классификации как эуплоидной (нормальное или здоровое состояние), недетерминированной и анеуплоидной (болезненное или нездоровое состояние). В другом воплощении, когда осуществляют множество сравнений (например, сравнение для каждого показателя разделения), классификация может быть основана на каждом сравнении. Например, способ, основанный системе правил, можно рассматривать при классификациях, являющихся результатами каждого из сравнений. В одном воплощении окончательная классификация обеспечивается только тогда, когда все классификации совместимы. В другом воплощении используют классификацию большинства. В еще одном воплощении можно использовать более сложные формулы на основании того, насколько близок каждый из



показателей разделения к соответствующей пороговой величине, и такие величины близости можно анализировать для определения классификации. Например, величины близости можно суммировать (наряду с другими факторами, такими как нормализация), и результат можно сравнить с другой пороговой величиной.

В других воплощения также можно применить вариации способа 600 для непосредственного сравнения статистического показателя для первой хромосомы с пороговой величиной, которая может быть получена из эталонного образца. Некоторые воплощения также можно использовать для анализа биологических образцов от небеременных женщин. Такой анализ можно рассматривать именно при статистическом показателе для размера всех фрагментов образца и сравнении статистического показателя или показателя разделения с пороговым значением для того, чтобы определить может ли существовать дисбаланс последовательностей. Если дисбаланс последовательностей классифицируют как существующий, можно выполнить дополнительный анализ локации дисбаланса, например, анализируя статистические показатели размеров и/или показатели разделения для определенной геномной локации (например, хромосомы).

### С. Использование общих длин коротких фрагментов

Как указывалось выше, в некоторых воплощениях распределение по размерам ДНК плазмы также может быть отражено долей общей длины ДНК, в которую вносят вклад короткие фрагменты ДНК. Например, распределение по размерам может включать общую длину фрагментов меньше пороговой величины, которую можно разделить на общую длину всех фрагментов или, по меньшей мере, фрагментов больше пороговой величины. Напротив, распределение по размерам ДНК плазмы также может быть отражено долей полной длины ДНК, в которую вносят вклад длинные фрагменты ДНК. Например, распределение по размерам может включать общую длину фрагментов больше пороговой величины, которую можно разделить на общую длину всех фрагментов или, по меньшей мере, фрагментов меньше пороговой величины. Как другой пример, также можно использовать отношение мелких фрагментов к крупным. В одном воплощении в качестве порога используют 150 п.о. для определения коротких фрагментов ДНК плазмы. Однако также в качестве порогов для определения коротких фрагментов ДНК можно использовать любую пороговую величину, например, в 130 п.о., 140 п.о., 160 п.о. и 170 п.о. Отмечается, что, как используется в данном описании, пары оснований также могут называться, как синоним, числом нуклеотидов (nt) при обращении к длине одноцепочечного фрагмента.

В одном воплощении вычисление доли длины ДНК, вносимой короткими фрагментами ДНК, можно выполнить следующим образом:  $F = \text{доля длины ДНК}$ ,

вносимой короткими фрагментами ДНК;  $S$  = сумма длин всех коротких фрагментов ДНК (длины, равные или меньше порога); и  $T$  = общая длина всех фрагментов ДНК в образце независимо от их длины. Так указанную долю можно представить как  $F = S / T$ , что является одним примером статистического показателя размера. Вычислить  $F$  можно для всех фрагментов из образца или для определенной локации в геноме, например, для определенной хромосомы.

В одном воплощении можно определить общую длину всех фрагментов ДНК. Затем можно выбрать размер порога ( $w$ ), меньше которой фрагменты ДНК определяют как «короткие фрагменты». Размер порога может изменяться и быть выбран как соответствующий определенным диагностическим целям. Общую длину всех фрагментов ДНК можно вычислить, суммируя длину всех фрагментов ДНК, которые равны или короче порога. Долю общей длины, вносимую короткими фрагментами ДНК, можно вычислить следующим образом:

$$F = \Sigma^w \text{ длин} / \Sigma^{600} \text{ длин, где}$$

$\Sigma^w$  длин представляет собой сумму длин фрагментов ДНК с длиной, равной или меньше пороговой длины  $w$  (п.о.); и

$\Sigma^{600}$  длин представляет собой сумму длин фрагментов ДНК с длиной, равной или меньше 600 п.о.

Статистический показатель  $F$  также можно использовать в воплощениях с использованием расположения в определенном порядке. Например, можно вычислить  $F$  для фрагментов из каждого набора соответствующих геномных локаций (например, хромосом).

С целью пояснения, в примере, приведенном ниже, можно вычислить общую длину путем суммирования фрагментов короче 600 п.о. Однако для вычисления «общей длины» можно использовать другие пределы размера, например, 400 п.о., 500 п.о. и 700 п.о. В данном примере общую длину вычисляют, основываясь на фрагментах в 600 п.о. или меньше, поскольку система геномного анализатора Illumina (Solexa) не эффективна при амплифицировании и секвенировании фрагментов ДНК длиннее 600 п.о. Кроме того, устанавливая предел анализа фрагментов ДНК короче 600 п.о., также можно избежать смещений, появляющихся из структурных изменений генома. При наличии структурного изменения, например, перестановок (Kidd J.M. et al., Nature, 2008, 453: 56-64), размер фрагмента ДНК может быть переоценен, когда размер оценивают биоинформативно путем картирования концов фрагмента ДНК в эталонном геноме. Кроме того, >99,9% всех фрагментов ДНК, успешно секвенированных и картированных в эталонном геноме, имеют менее 600 п.о., и, таким образом, включение всех фрагментов, равных и короче 600 п.о.,

может обеспечить репрезентативную оценку распределения по размерам фрагментов ДНК в образце.

Как обсуждалось выше, сдвиг распределения по размерам к более коротким фрагментам ДНК можно наблюдать для хромосомы, которая представлена у плода избыточной копией. В одном аспекте измерение различия в долях общей длины, вносимых короткими фрагментами ДНК, между хромосомой с повышенным риском анеуплоидии (хромосома-мишень) и хромосомами, не входящими в группу с повышенным риском анеуплоидии (эталонные хромосомы), может быть осуществлено количественно для определения того, различается ли распределение по размерам фрагментов ДНК, происходящих от указанных хромосом.

В одном воплощении  $F_{(Tar)}$  и  $F_{(Ref)}$  определяют как доли общей длины, вносимые короткими фрагментами ДНК, для хромосомы с повышенным риском анеуплоидии и эталонной хромосомы(хромосом), соответственно. Различие в долях общей длины, вносимых короткими фрагментами ДНК, между хромосомой-мишенью и эталонной хромосомой(хромосомами) ( $\Delta F_{(Tar-Ref)}$ ) можно вычислить как  $\Delta F_{(Tar-Ref)} = F_{(Tar)} - F_{(Ref)}$ . Например,  $\Delta F_{(21-1)}$  представляет собой различие между долями общей длины, вносимыми короткими фрагментами ДНК, для хромосомы 21 и хромосомы 1. Применение  $\Delta F_{(Tar-Ref)}$  в пренатальной диагностике фетальных хромосомных анеуплоидий будет обсуждаться в последующих разделах. В другом воплощении отношение  $F_{(Tar)} / F_{(Ref)}$  также может быть использовано как показатель разделения путем, схожим с путем использования  $\Delta F_{(Tar-Ref)}$ .

Различие между величинами  $F$  хромосомы-мишени и эталонной хромосомы можно использовать в качестве показателя для определения, является ли плод трисомиком по хромосоме-мишени или нет. Когда плод является трисомиком по хромосоме-мишени, трисомическая хромосома будет вносить в материнскую плазму дополнительную дозу короткой фетальной ДНК, приводя, следовательно, к явному сжатию распределения по размерам последовательностей хромосомы-мишени. Такое сжатие распределения по размерам последовательностей хромосомы-мишени может привести к возрастанию доли длины последовательностей, вносимой короткими фрагментами ДНК, для последовательности-мишени ( $F_{target}$ ). В результате может иметь место возрастание в различии  $\Delta F$  между  $F_{target}$  и  $F_{ref}$ .

Фиг. 7 показывает график 700 доли общей длины, вносимой короткими фрагментами ( $F$ ), против размера порога ( $w$ ) для образца материнской плазмы с плодом мужского пола. Величины  $F$  для фрагментов ДНК, выровненных с аутосомами и хромосомой Y, на графике нанесены по вертикальной оси против размера порога,

используемого для определения «коротких фрагментов ДНК». При беременности с плодом мужского пола молекулы ДНК, выровненные с хромосомой Y, представляют ДНК, высвобожденную из плода мужского пола. Так как большинство циркулирующих молекул ДНК в материнской плазме происходят от матери, фрагменты ДНК, выровненные с аутосомами, должны представлять собой преимущественно фрагменты материнской ДНК. Величина F возрастает с размером порога и приближается к величине 1,0, когда все фрагменты ДНК в образце короче или равны пороговому размеру. Различие в распределении по размерам между двумя видами молекул ДНК может быть отражено различием в их величинах F. Более высокая величина F показывает, что большая доля общей длины вносится короткими фрагментами, и, таким образом, указывает на более сжатое распределение по размерам фрагментов ДНК.

Как видно на графике 700, распределение по размерам молекул ДНК из хромосомы Y более узкое, чем распределение по размерам ДНК из аутосом. Конкретно, величина F для хромосомы Y возрастает раньше, чем для аутосом, что приводит к  $F_Y$  более высокой, чем  $F_{\text{аутосомы}}$  между 80 п.о. и 300 п.о. Различия в величинах F между хромосомой Y и аутосомами ( $\Delta F_{(Y\text{-аутосомы})}$ ) откладываются на графике против размера порога и представляют пунктирной линией, которая положительна между 80 п.о. и 300 п.о.. Максимальная величина для  $\Delta F_{(Y\text{-аутосомы})}$  составляет 0,23 и имеет место вблизи 150 п.о. Как поясняется в последующих примерах, различие в доли общей длины фрагментов, вносимой короткими фрагментами ( $\Delta F$ ), между хромосомой «в группе риска» и эталонной хромосомой(ами), является применимым показателем разделения для количественного определения различия в их распределении по размерам. Также, в порядке примера, можно определить величину  $\Delta F$  с любым размером порога, например, между 130 п.о. и 170 п.о.

Фиг. 8 показывает график 800 доли общей длины, вносимой короткими фрагментами (F), для хромосомы 21 (сплошная линия) и эталонными хромосомами (все аутосомы, за исключением хромосом 13, 18 и 21) (пунктирная линия) против размера порога для эуплоидной беременности. Различия в двух величинах F ( $\Delta F_{(21\text{-Ref})} = F_{(xp,21)} - F_{(Ref)}$ ) представлено штриховой линией. Так как распределение по размерам фрагментов ДНК из хромосомы 21 и эталонных хромосом схоже при эуплоидной беременности, величина  $\Delta F_{(21\text{-Ref})}$  близка к нулю для любого размера порога.

Фиг. 9 показывает график 900 величины F для хромосомы 21 (сплошная линия) и эталонных хромосом (все аутосомы, за исключением хромосом 13, 18 и 21) (пунктирная линия) против размера порога для беременности с трисомией 21. Различия  $\Delta F_{(21\text{-Ref})}$  представлено штриховой линией. Из-за дополнительной дозы хромосомы 21 из плода

распределение по размерам фрагментов ДНК в материнской плазме для хромосомы 21 более сжатое, чем распределение по размерам для эталонных хромосом. Такое различие в распределении по размерам фрагментов ДНК отображается положительной величиной  $\Delta F_{(21-Ref)}$ , которая достигает максимума в 0,016 при приблизительно 150 п.о.

Фиг. 10 показывает график  $1000 \Delta F_{(21-Ref)}$  между хромосомой 21 и эталонными хромосомами (все аутосомы, за исключением хромосом 13, 18 и 21) против размера порога для эуплоидной беременности и беременности с трисомией 21. В случае трисомии 21 наблюдают возросшую величину  $\Delta F_{(21-Ref)}$ , но  $\Delta F_{(21-Ref)}$  равна приблизительно нулю при каждом размере порога в случае эуплоидии. Так как максимум  $\Delta F_{(21-Ref)}$  наблюдают вблизи 150 п.о., различие в 150 п.о. можно использовать в качестве показателя разделения для определения, если в распределении по размерам последовательностей хромосомы 21 существует любое значимое сжатие. Однако можно использовать любой размер, при котором возникает различие между случаями эуплоидии и трисомии, например, но без ограничения, 140 п.о., 145 п.о., 155 п.о. и 160 п.о. В данном примере различие между долей общей длины хромосомы 21 и эталонных хромосом, наблюдаемое при 150 п.о., составляет 0,016 и -0,002 для беременности с трисомией 21 и эуплоидной беременности, соответственно.

Такое различие в форме доли общей длины ДНК, вносимой короткими фрагментами ДНК, можно использовать для того, чтобы различить эуплоидный и ануэуплоидный плод. Различие можно проверить различными способами. В одном воплощении величину  $\Delta F$  при определенном размере порога (пример показателя разделения) можно сравнить с пороговой величиной для определения категоризации (классификации) образца. В другом воплощении можно найти пиковую величину в  $\Delta F$ , и эту величину можно сравнить с одной или несколькими пороговыми величинами категоризации. В различных воплощениях пиковая величина может представлять собой максимальную или минимальную величину, усредненную величину вблизи макс/мин величины или другую величину, ассоциированную с или происходящую от макс/мин величины. Также можно использовать другие статистические показатели показателя разделения (например,  $\Delta F$ ), такие как ширина пика или местоположение определенного порога, соответствующей пику.

В одном воплощении величину  $F$  (или другой статистический показатель, описанный выше, такой как длина мелких фрагментов, деленная на длину крупных фрагментов), полученную из множества фрагментов для определенной геномной локации или для полного генома, можно использовать для определения того, существует ли

патология. Например, если статистический показатель превышает величину порога, тогда патологию можно определять как существующую, так как количество мелких фрагментов находится вне нормального интервала. Кроме беременных женщин, это можно осуществлять и для других пациентов для идентификации иных заболеваний, чем заболевания плода.

В некоторых воплощениях перед анализом размеров фрагментов можно выполнить физическое фракционирование по размерам. В одном воплощении молекулы нуклеиновых кислот можно разделить по размерам на две фракции (например, одну крупнее 200 п.о. и одну, равную или меньше 200 п.о.), и затем можно сравнить распределение по размерам выбранных хромосом (например, хромосомы 21) с каждой из таких фракций по размерам. При наличии фетальной трисомии (например, трисомии 21) фракция по размерам молекул меньшего размера может увеличиваться в относительном избытке при сравнении с фракцией по размерам молекул большего размера.

В других воплощениях вместо распределения по размерам можно использовать число фрагментов меньше пороговой длины. Например, можно сравнить число фрагментов меньше пороговой длины (например, по различию или отношению) для хромосомы-мишени (например, хромосомы 21) относительно одной или нескольких эталонных хромосом. В одном воплощении число фрагментов меньше пороговой длины делят на общее число фрагментов, и получают их долю в процентах, и можно сравнить такие доли в процентах для хромосомы-мишени и одной или нескольких эталонных хромосом и получить параметр. Полученный параметр (например, различие или отношение) можно сравнить с пороговой величиной (например, 1%). В одном аспекте пороговую длину можно выбрать при длине, где указанная выше доля в процентах является наибольшей.

#### V. Примеры использования ранжировки

Кроме трисомии 21 анализ размеров фрагментов в материнской плазме также можно использовать для неинвазивного пренатального обнаружения других фетальных хромосомных анеуплоидий, таких как трисомия 13, трисомия 18 и анеуплоидии половых хромосом (такие как синдром Тернера, синдром Кляйнфельтера и ХУУ и т.д.). Воплощения также можно использовать, когда в хромосомную аномальность вовлекается только часть определенной хромосомы (например, при трисомии 21, вызванной хромосомной транслокацией). При таком варианте развития событий аномальности в размерах фрагментов будут отмечаться для фрагментов ДНК из пораженных хромосомных участков.

Фиг. 11 показывает таблицу 1100 для библиотек ДНК материнской плазмы,

которые сконструированы с использованием набора multiplexing sample preparation (Illumina) согласно инструкциям изготовителя. Каждые два образца с различными штриховыми кодами вводят в одну полосу и последовательно подвергают стандартному мультиплексному секвенированию спаренных концов на геномном анализаторе II Illumina. Образцы можно различить на основании штриховых кодов. Анализируют образцы плазмы от 120 беременных женщин. Пол и состояние хромосомной анеуплоидии плодов показаны в таблице 1100.

Фиг. 12 иллюстрирует ранжировку различных хромосом для четырех образцов при различных болезненных состояниях или состояниях в отсутствие заболеваний согласно воплощениям настоящего изобретения. Располагают 22 аутосомы вместе с хромосомой X в порядке размеров их фрагментов, как описано ранее. Как можно видеть, относительная ранжировка хромосом на основе версии 2 набора реагентов для генерации кластеров Illumina отличается от ранжировки при использовании версии 1 набора (фиг. 2). Как показано ниже, воплощения (например, способ 400) позволяют отличить случай анеуплоидии от случая эуплоидии.

#### А. Трисомия 13

В данном примере демонстрируется применение воплощений для пренатальной диагностики трисомии 13. Фиг. 13 показывает график ранжировки хромосомы 13 для 120 случаев беременности, например, который можно получить по способу 400. На графике T13, T18, T21 и Eu относятся к беременности с трисомией 13, трисомией 18, трисомией 21 и эуплоидной беременностью, соответственно. В 18 (78,3%) из 23 случаев беременности с трисомией 13 хромосоме 13 присваивают ранжир 22 или ниже, в то время как только в 2 (2,1%) из 97 случаев беременности без трисомии 13 хромосоме 13 присваивают ранжир 22 или ниже. Следовательно, с использованием порогового ранжира 22 чувствительность и специфичность анализа по ранжиру размеров фрагментов хромосом для пренатальной диагностики фетальной трисомии 13 составляет 87,3% и 97,9%, соответственно.

На фиг. 13 можно видеть, что ранжир хромосомы 13 выше (т.е. с меньшим числом, отмечающим ранжир) для эуплоидной беременности и беременности с трисомией 18 и 21, чем для беременности с трисомией 13. Иными словами, при сравнении с последовательностями других хромосом оказывается, что последовательности хромосомы 13 короче в случаях беременности с трисомией 13, чем в случаях беременности без трисомии 13. Явное укорачивание последовательностей хромосомы 13 при беременности с трисомией 13 имеет место из-за повышенного вклада фетальной ДНК в хромосому 13 из-за дополнительной хромосомы 13 плода.

### В. Трисомия 18

Фиг. 14 показывает ранжировку хромосомы 18 для 120 случаев. В 26 (86,7%) из 30 случаев трисомии 18 ранжиры хромосомы 18 ниже (т.е. с большим числом, отмечающим ранжир), чем 13, в то время как ни в одном из 90 случаев без трисомии 18 ранжир не ниже, чем 13. Следовательно, с использованием ранжира 13 в качестве пороговой величины, чувствительность и специфичность анализа по ранжиру размеров фрагментов хромосом для пренатальной диагностики фетальной трисомии 18 составляет 86,7% и 100%, соответственно.

В таком анализе сравнивают ранжировку по размеру хромосомы 18 при беременности с трисомией 18, трисомией 13, трисомией 21 и эуплоидной беременностью. В контексте хромосомы 18 последние три группы могут рассматриваться как «нормальный» контроль, поскольку они не имеют «дозы» аномальностей, включающих хромосому 18. Как можно видеть из фиг. 14, ранжиры для хромосомы 18 в случаях эуплоидии и трисомии 13 группируются вокруг 1-3. С другой стороны, ранжирами для хромосомы 18 в случаях трисомии 18 являются 13-22, что указывает, что размеры фрагментов хромосомы 18 меньше, чем размеры в случаях эуплоидии, трисомии 21 и трисомии 13. Снова такие наблюдения объясняет дополнительная хромосома 18.

### С. Трисомия 21

Фиг. 15 показывает ранжировку хромосомы 21 для 120 случаев. В 8 (88,9%) из 9 случаев трисомии 21 хромосоме 21 присваивают ранжиры 22 или ниже, в то время как ни в одном из 111 случаев без трисомии 21 хромосома 21 не имеет ранжир 22 или ниже. Следовательно, с использованием ранжира 22 в качестве пороговой величины чувствительность и специфичность анализа по ранжиру размеров фрагментов хромосом для пренатальной диагностики фетальной трисомии 21 составляет 89,9% и 100%, соответственно.

### V. Примеры использования различия в размерах

#### А. Трисомия 13

В последующих примерах демонстрируется, что можно добиться неинвазивного пренатального обнаружения трисомии 13 путем сравнения абсолютных размеров фрагментов, происходящих от хромосомы 13, с абсолютными размерами фрагментов, происходящих от одной или нескольких эталонных хромосом, например, так, как описано для способа 600. В данном примере используют тот же набор данных, что и в предыдущем примере. В порядке иллюстрации, для обнаружения трисомии 13 в качестве эталонных хромосом выбирают хромосому 5 и хромосому 6.

Как можно видеть из таблицы 1600 на фиг. 16, в пределах одного и того же образца



последовательности, выровненные с хромосомой 13, существенно короче, чем последовательности, выровненные с хромосомой 5 и хромосомой 6, для всех случаев беременности с трисомией 13 (критерий суммы ранжиров Манна-Уитни, величина  $p \leq 0,001$ ). В качестве контроля включают эуплоидную беременность и беременность с трисомией 18. Как при эуплоидной беременности, так и при беременности с трисомией 18 доза хромосомы 13 нормальная. Как можно видеть из таблицы 1600, аномальностей размеров такой статистической значимости для последовательностей, происходящих от хромосомы 13, при эуплоидной беременности и беременности с трисомией 18 не видно.

Кроме того, для всех случаев беременности с трисомией 13 в пределах одного и того же образца различие в среднем размере фрагментов между хромосомой 13 и хромосомой 5 больше 0,4 п.о., в то время как ни один из случаев без трисомии 13 не показывает различия больше 0,4 п.о. Подобным образом, для всех случаев беременности с трисомией 13 различие в среднем размере фрагментов между хромосомой 13 и хромосомой 6 больше 0,5 п.о., в то время как ни один из случаев без трисомии 13 не показывает различия больше 0,5 п.о.

#### В. Трисомия 18

В порядке иллюстрации, для обнаружения трисомии 18 в качестве эталонной хромосомы выбирают хромосому 14. Для случаев беременности без трисомии 18 в таблице 1700 на фиг. 17 можно видеть, что последовательности, происходящие от хромосомы 18, статистически значимо длиннее, чем последовательности, происходящие от хромосомы 14 (критерий суммы ранжиров Манна-Уитни, величина  $p \leq 0,005$ ). Однако для случаев трисомии 18 последовательности хромосомы 18 несущественно длиннее, чем последовательности, выровненные с хромосомой 14. Случаи эуплоидии и трисомии 18 можно различить на основании различия между средним размером фрагментов ДНК из хромосом 18 и 14 с использованием пороговой величины 0 п.о. Такие наблюдения можно объяснить тем фактом, что дополнительная доза происходящих от плода последовательностей хромосомы 18, которые короче, чем материнские последовательности, будет снижать общий размер таких последовательностей, когда плод имеет трисомию 18. Это также будет подводить общее распределение по размерам для хромосомы 18 ближе к распределению для хромосомы 14.

#### V. Примеры с использованием вклада в общую длину

В примере, описанном ниже, эталонные хромосомы включают все аутосомы, за исключением хромосом 13, 18 и 21.

Фиг. 18 показывает различие в долях общей длины, вносимых короткими фрагментами, между хромосомой 18 и эталонными хромосомами ( $\Delta F_{(18-Ref)}$ ) в 150 п.о. С

использованием диагностической пороговой величины 0,0003 для  $\Delta F_{(18-Ref)}$  беременность с трисомией 18 можно обнаружить с чувствительностью 93,3% и специфичностью 100%.

Фиг. 19 показывает различие в долях общей длины, вносимых короткими фрагментами, между хромосомой 21 и эталонными хромосомами ( $\Delta F_{(21-Ref)}$ ) в 150 п.о. С использованием диагностической пороговой величины 0,007 для  $\Delta F_{(21-Ref)}$  беременность с трисомией 21 можно обнаружить с чувствительностью 100% и специфичностью 100%.

#### VII. Отбор эталонной хромосомы(хромосом)

Одну или несколько эталонных хромосом для неинвазивного пренатального обнаружения фетальной хромосомной анеуплоидии анализом размеров ДНК в материнской плазме можно отобрать различными способами, например, с использованием способа 600. В различных воплощениях могут быть отобраны различные эталонные хромосомы.

Первым типом эталонных хромосом являются хромосомы, в которых фрагменты ДНК, происходящие от них, показывают распределение по размерам на специфической аналитической платформе (или аналитических платформах, которые близко соответствуют аналитическим действиям), схожее с распределением, полученным для хромосомы в материнской плазме, потенциально вовлеченной в анеуплоидию (например, хромосомы 21, 18 или 13). В одном воплощении в анализе такого типа фетальную хромосомную анеуплоидию обнаруживают, если расположение по размерам или абсолютный размер хромосомы «из группы риска» показывает статистически значимое уменьшение от эталонной хромосомы(хромосом). В других воплощениях можно измерить различие в средних или срединных размерах между фрагментами, происходящими от хромосомы «из группы риска» и фрагментами эталонной хромосомы(хромосом).

Вторым типом эталонных хромосом являются хромосомы, в которых фрагменты ДНК, происходящие от них, статистически короче, чем фрагменты, происходящие от хромосомы в материнской плазме, потенциально вовлеченной в анеуплоидию (например, хромосомы 21, 18 или 13), когда плод является эуплоидным. Такой вариант развития событий может встретиться, когда хромосома «из группы риска» является одной из самых длинных при измерении с использованием определенной платформы. Например, на фиг. 12 фрагменты ДНК в материнской плазме, происходящие от хромосомы 18, при измерении являются самыми длинными среди аутосом. Таким образом, можно отобрать эталонные хромосомы, фрагменты которых статистически значимо короче, чем хромосома 18. В одном воплощении в анализе такого типа фетальную хромосомную анеуплоидию обнаруживают в случае, если нельзя увидеть, что расположение по размерам или абсолютный размер хромосомы «из группы риска» статистически значимо

отличается от эталонной хромосомы(хромосом). Например, такую стратегию используют для анализа, отображенного на фиг. 17, описанного выше.

Третьим типом эталонных хромосом являются хромосомы, в которых фрагменты ДНК в материнской плазме, происходящие от них, статистически длиннее, чем фрагменты, происходящие от хромосомы, потенциально вовлеченной в анеуплоидию (например, хромосомы 21, 18 или 13), когда плод является эуплоидным. Такой вариант может встретиться, когда хромосома «из группы риска» является одной из самых коротких при измерении с использованием определенной платформы. Таким образом, можно отобрать эталонные хромосомы, фрагменты которых статистически значимо длиннее, чем хромосома «из группы риска». В одном воплощении в анализе такого типа фетальную хромосомную анеуплоидию обнаруживают в случае, если ранжировка по размеру, различие в ранжировке или абсолютном размере между эталонной хромосомой и хромосомой «из группы риска» возрастает.

Четвертым типом эталонных хромосом являются хромосомы со схожим содержанием GC. Содержание GC в хромосоме может влиять на считывание количественных данных при реакции секвенирования. По одному способу для минимизации возможной погрешности, являющейся результатом различия в содержании GC между хромосомами, следует отбирать соответствующие эталонные хромосомы со схожим содержанием GC. На фиг. 20 приводится содержание GC в различных хромосомах (NCBI build 36, версия 48). Хромосомы перечислены в порядке возрастания содержания GC. Теперь приводится пример использования эталонных хромосом со схожим содержанием GC.

Фиг. 21 показывает различие в долях общей длины, вносимых короткими фрагментами, между хромосомой 13 и эталонными хромосомами ( $\Delta F_{(13-Ref)}$ ) в 150 п.о. В данном случае в порядке иллюстрации для анализа  $\Delta F_{(13-Ref)}$  в качестве эталонных хромосом используют хромосомы 3, 4, 5 и 6. Как видно на фиг. 20, хромосомы 3-6 имеют содержание GC 36,53%-38,79%, что схоже с содержанием GC 37,10% для хромосомы 13. С использованием диагностической пороговой величины 0,0038, анализ  $\Delta F$  обнаруживает случаи трисомии 13 с чувствительностью 95,7% и специфичностью 99,0%.

Предполагается, что проблема смещения GC может влиять на секвенирование на различных платформах в различной степени. Например, использование платформ, для которых не требуется предварительная амплификация, таких как платформа Helicos (Harris T.D. et al., Single-molecule DNA sequencing of a viral genome, Science, 2008, 320: 106-9), нанопоры (Lund J., Parviz B.A., Scanning probe and nanopore DNA sequencing: core techniques and possibilities, Methods Mol. Biol., 2009, 578: 113-22) или одномолекулярная

система реального времени от Pacific Biosciences (Eid J. et al., Real-time DNA sequencing from single polymerase molecules, Science, 2009, 323: 133-8), может допускать более широкий выбор группы эталонных хромосом.

#### IV. Использование концентрации фетальной ДНК

При беременности с трисомией 13 (или другой трисомией) плод будет высвобождать в материнскую плазму дополнительную дозу фрагментов хромосомы 21, которые могут быть короче, чем фрагменты из материнских клеток. Можно ожидать, что концентрация таких более коротких фрагментов будет коррелировать с концентрацией фетальной ДНК в материнской плазме. Иными словами, при беременности с трисомией 21, чем выше долевая концентрация последовательностей, происходящих от хромосомы Y, тем меньше будет измеренный размер последовательностей, происходящих от хромосомы 21.

Фиг. 22 иллюстрирует результаты, полученные для случаев беременности, включающей плод мужского пола с трисомией 21. Как можно видеть на фиг. 22, действительно существует отрицательная корреляция между средним размером последовательностей хромосомы 21 и процентом последовательностей, выровненных с хромосомой Y ( $r = -0,942$ , корреляция по Пирсону). Подобной тенденции также можно ожидать, если используют ранжировку по размеру хромосомы 21, т.е. номер по ранжиру хромосомы 21 будет повышаться, указывая на более короткие фрагменты, когда долевая концентрация фетальной ДНК возрастает. Так как корреляция существует, в воплощениях измерение концентрации фетальной ДНК можно использовать в качестве параметра в любом из способов, описанных в данном описании.

В анализе такого типа долевую и абсолютную концентрацию фетальной ДНК в материнской плазме можно измерить любым способом, известным специалистам в данной области техники. Если плод мужского пола, концентрацию фетальной ДНК можно измерить по долевой концентрации в материнской плазме последовательностей, происходящих от хромосомы Y. Другим примером является использование наследуемых от отца генетических маркеров, таких как однонуклеотидные полиморфизмы или полиморфизм простых tandemных повторов или инсерционно-делеционные полиморфизмы. Другим примером является использование эпигенетических маркеров, таких как участки, которые дифференцированно метилированы в фетальной и материнской ДНК (Poon et al., Clin. Chem., 2002, 48: 35-41; Chi et al., Am. J. Pathol., 2007, 170: 941-950; Chan et al., Clin. Chem., 2006, 52: 2211-2218; патент США 6927028). Вышеуказанные маркеры можно анализировать с использованием способов, известных специалистам в данной области техники, в том числе, полимеразной цепной реакции

(ПЦР), численной ПЦР, секвенирования, массивным параллельным секвенированием и нацеленным массивным параллельным секвенированием.

В одном воплощении можно изменять диагностический порог (например, пороговую величину для любого способа, описанного в данном описании) для обнаружения хромосомной анеуплоидии в отношении измеренной концентрации фетальной ДНК в материнской плазме. Таким образом, можно ожидать, что для образцов материнской плазмы с относительно высокой концентрацией фетальной ДНК степень укорачивания молекул ДНК в плазме, происходящих от хромосомы, потенциально вовлеченной в анеуплоидию, будет более значительной, чем для образцов материнской плазмы с относительно низкой концентрацией фетальной ДНК.

Так как  $\Delta F$  соотносится с распределением по размерам,  $\Delta F$  также показывает корреляцию с концентрацией фетальной ДНК. Фиг. 23 показывает корреляцию между  $\Delta F_{(18-Ref)}$  (хромосома 18 против эталонных хромосом) и концентрацией фетальной ДНК. В десяти из тридцати случаев Т18 вынашиваются плоды мужского пола, и таким образом, долевую концентрацию фетальной ДНК можно оценить по доле концентрации последовательностей хромосомы Y в указанных образцах. Существует значимая корреляция между  $\Delta F_{(18-Ref)}$  и долевой концентрацией последовательностей хромосомы Y ( $r = 0,879$ , корреляция по Спирману). Такие результаты предполагают, что при беременности с трисомией 18 степень сужения в распределении по размерам последовательностей хромосомы 18 в материнской плазме коррелирует с долевой концентрацией фетальной ДНК в материнской плазме.

Случаи ниже диагностической пороговой величины представлены незаштрихованными кружочками, в то время как случаи с различием больше пороговой величины представлены заштрихованными кружочками. Два случая, в которых величины  $\Delta F$  меньше диагностической пороговой величины (0,0003) (представлены незаштрихованными кружочками), имеют относительно низкие долевые концентрации фетальной ДНК при сравнении со случаями с величинами  $\Delta F$ , превышающими 0,0003 (представлены заштрихованными кружочками). Низкие долевые концентрации фетальной ДНК могут быть причиной, по которой на фиг. 18 указанные два случая пропущены при выполнении анализа. Соответственно, в одном воплощении, если фетальная концентрация в образце низкая, классификацией можно пренебречь или провести ее снова.

Фиг. 23В показывает корреляцию между  $\Delta F_{(21-Ref)}$  (хромосома 21 против референсных хромосом) и концентрацией фетальной ДНК. В пяти из девяти случаев Т21

вынашиваются плоды мужского пола. Существует значимая корреляция между величиной  $\Delta F_{(21-Ref)}$  и долевой концентрацией последовательностей хромосомы Y ( $r = 0,9$ , корреляция по Спирману).

Фиг. 23С показывает корреляцию между  $\Delta F_{(13-Ref)}$  (хромосома 13 против хромосом 3, 4, 5 и 6) и концентрацией фетальной ДНК. В четырнадцати из двадцати трех случаев трисомии 13 вынашиваются плоды мужского пола. Долевою концентрацию фетальной ДНК можно оценить по долевой концентрации последовательностей хромосомы Y в образцах. Существует положительная корреляция между  $\Delta F_{(13-Ref)}$  и долевой концентрацией последовательностей хромосомы Y ( $r = 0,644$ , корреляция по Спирману). Случай, когда  $\Delta F_{(13-Ref)}$  меньше диагностической пороговой величины (0,0038), представлен незаштрихованными кружочками. Низкие долевые концентрации фетальной ДНК могут быть причиной, по которой на фиг. 21 указанный случай пропущен при выполнении анализа.

#### IX. Сравнение анализа размеров и способа с подсчетом молекул

Фиг. 24 показывает сравнение точности воплощения настоящего изобретения и другого способа неинвазивного обнаружения фетальных анеуплоидий (трисомии 13 и трисомии 18) с использованием анализа ДНК материнской плазмы. Данный пример иллюстрирует сравнение воплощений с использованием размеров против способа, основанного на подсчете молекул (заявка на патент США 11/701686; Chiu et al., Trends Genet., 2009, 25: 324-331; Chiu et al., Proc. Natl. Acad. Sci. USA, 2008, 105: 20458-20463; Fan et al., Proc. Natl. Acad. Sci. USA, 2008, 105: 16266-16271; публикация патента США 2009/0029377). Восемь образцов материнской плазмы (два эуплоидных, два с трисомией 18 и четыре с трисомией 13) анализируют с использованием способа, описанного Chiu et al. (Proc. Natl. Acad. Sci. USA, 2008, 105: 20458-20463). Для каждого случая описанный ранее способ с подсчетом молекул с использованием z-оценки по Chiu et al. сравнивают с результатами воплощений на основе анализа размеров.

Для вычисления z-оценки сначала в каждом случае вычисляют представленность, в процентах, хромосомы, представляющей интерес. Затем вычисляют среднее и стандартное отклонение хромосомной представленности для случаев для сравнения. В таком наборе данных случаи 1, 2, 5, 6, 7 и 8 используют в качестве эталонной группы для вычисления среднего и стандартного отклонения представленности хромосомы 18. Случаи 1, 2, 3 и 4 используют в качестве эталонной группы для вычисления среднего и стандартного отклонения представленности хромосомы 13. Определяют z-оценку как число стандартных отклонений от среднего группы для сравнения. Значимую сверхпредставленность хромосомы определяют как z-оценку  $< 3$ . В случаях 3 и 4

вынашивается плод с трисомией 18, и в их плазме сверхпредставлены фрагменты хромосомы 18. Случаи 5, 6, 7 и 8 являются случаями вынашивания с трисомией 13, но только 5 и 7 показывают сверхпредставленность хромосомы 13 в их плазме. Случаи 6 и 8 не показывают значимой сверхпредставленности хромосомы 13 в материнской плазме, несмотря на вынашивание плода с трисомией 13.

Значимое уменьшение размеров фрагментов ДНК хромосомы обнаруживают путем сравнения размера всех фрагментов, выровненных с определенной хромосомой, с размерами фрагментов, выровненных с эталонной хромосомой. Используют для сравнения критерий Манна-Уитни, и величину  $P < 0,0001$  определяют как наличие значимого различия. Для анализа размеров фрагментов хромосомы 13 эталонной хромосомой является хромосома 5. Для всех случаев с эуплоидным плодом или плодом с трисомией 18 размер фрагментов хромосомы 13 значимо не отличается от размера фрагментов хромосомы 5. Для четырех случаев с плодом с трисомией 13 фрагменты значимо короче, чем фрагменты хромосомы 5, что подразумевает, что размер фрагментов хромосомы 13 меньше по сравнению с размерами в случае отсутствия трисомии 13. Таким образом, все четыре случая трисомии 13 правильно идентифицируются по настоящему изобретению при сравнении с двумя из четырех по способу с z-оценкой.

Для анализа размеров фрагментов хромосомы 18 эталонной хромосомой является хромосома 12. Для всех случаев вынашивания эуплоидного плода или плода с трисомией 13 фрагменты хромосомы 18 значимо длиннее, чем фрагменты хромосомы 12. Для двух случаев вынашивания плода с трисомией 18 размеры фрагментов хромосомы 18 значимо не отличаются от размеров фрагментов хромосомы 12, что подразумевает, что размер фрагментов хромосомы 18 уменьшается при сравнении со случаями вынашивания плода без трисомии 18. Иными словами, оба случая трисомии 18 классифицированы правильно.

#### Х. Обнаружение полиморфизмов и диагностика генетических нарушений

Анализ размеров ДНК материнской плазмы также можно использовать для неинвазивного анализа генотипа плода. Генотип плода можно использовать для определения того, наследует ли плод мутированный ген, имеет ли дисбаланс определенного аллеля или другой дисбаланс последовательностей, или для других целей. В таких воплощениях один аллель может представлять собой эталонную геномную локацию (последовательность), а отличающийся аллель может представлять собой испытываемую геномную локацию. Таким образом, любой способ с использованием эталонной последовательности также можно применить для определения генотипа и дисбаланса последовательностей.

В одном воплощении дисбаланс последовательностей (и таким образом, возможно,

генотип) можно определить по тому, существует ли или нет различие в размерах (дисбаланс) между аллелями в материнском образце (например, когда мать является гетерозиготной по аллелю). Например, если различия в профиле размеров между аллелями в образце не имеется, тогда плод можно определить как имеющий тот же генотип, что у матери. Как другой пример, если различие в профиле размеров между аллелями в образце имеется, тогда плод можно определить как имеющий генотип, отличающийся от генотипа матери.

В последующих примерах мать является гетерозиготной (т.е. с одной копией аллеля N и одной копией аллеля M, что обозначается NM) для определенного локуса. Буквы N и M номинально представляют аллели дикого типа (N для нормальных) и мутантные (M для мутантных), соответственно. Однако аллели N и M могут соответствовать любым двум различным аллелям, а не обязательно дикому типу и/или мутанту. В одном воплощении M можно рассматривать как последовательность хромосомы «из группы риска» и N как эталонная последовательность. В данном контексте для определения генотипа можно предположить применение любого из вышеуказанных способов с использованием эталонной последовательности.

У небеременной женщины усредненный размер молекул, содержащих оба аллеля, может быть одинаковым. Однако в плазме беременной женщины существует смесь молекул ДНК матери и плода. Молекулы ДНК, происходящие от матери, длиннее, чем происходящие от плода. Если мать и плод имеют два аллеля (т.е. оба аллеля N и M), оба указанных аллеля будут иметь равный вклад длинных и коротких молекул ДНК. Таким образом, полученное распределение по размерам аллелей N и M будет одинаковым. Напротив, если генотипы матери и плода различаются, например, если мать является NM, а плод является MM, тогда распределение по размерам аллелей N и M будет различным. Иными словами, на распределение по размерам двух аллелей в плазме может повлиять генотип плода. Фиг. 25А-25С показывают диаграммы для различных вариантов генотипов беременной женщины и плода согласно воплощениям настоящего изобретения.

На фиг. 25А плод имеет генотип NN, и генотип матери NM. Длина столбца 2510 является показателем усредненного (среднего) размера фрагментов, которые являются фрагментами одного из двух аллелей матери и плода, соответственно. Как описано выше, плод имеет меньший средний размер, чем мать. Таким образом, длинные столбцы представляют материнскую ДНК, и короткие столбцы представляют фетальную ДНК.

Поскольку как мать, так и плод вносят вклад в аллель N, в то время как в аллель M вносит вклад только мать, распределение по размерам молекул с аллелем N будет более узким по сравнению с распределением молекул с аллелем M. Иными словами, у



беременной женщины, которая имеет генотип NM, более узкое распределение по размерам в случае аллеля N по сравнению с аллелем M будет предполагать, что генотип плода NN. Таким образом, можно идентифицировать, что плод имеет два аллеля дикого типа (N), когда распределение по размерам (например, средний размер) N более узкое, чем для M, по определенной пороговой величине (например, величине в процентах или абсолютной величине).

На фиг. 25B плод имеет генотип NM. Как мать, так и плод вносят вклад в аллели N и M. В результате распределение по размерам для молекул с аллелями M и N является одинаковым. У беременной женщины, которая имеет генотип NM, одинаковое распределение по размерам для аллелей M и N будет указывать, что генотип плода NM. Таким образом, можно идентифицировать, что плод имеет один аллель дикого типа (N) и один мутантный (M) аллель, когда распределение по размерам (например, средний размер) N примерно эквивалентно распределению M при определенной пороговой величине (например, в процентах или величине).

На фиг. 25C плод имеет генотип MM. Поскольку как мать, так и плод вносят вклад в аллель M, в то время как в аллель N вносит вклад только мать, распределение по размерам молекул с аллелем M будет более узким по сравнению с аллелем N. У беременной женщины, которая имеет генотип NM, более узкое распределение по размерам в случае аллеля M, чем для аллеля N, может указывать, что генотип плода MM. Таким образом, можно идентифицировать, что плод имеет два мутантных аллеля (M), когда распределение по размерам (например, средний размер) M более узкое, чем для N, при определенной пороговой величине (например, величине в процентах или величине).

Способ также можно использовать для анализа ситуаций, когда мать является гомозиготной, например, NN или MM. Если плод имеет другой генотип, распределение по размерам материнского образца также будет изменяться, и таким образом можно определить генотип плода. Также, если распределение по размерам не изменяется, тогда генотип можно определить как такой же, как у матери, как описано выше для случая, когда мать представляет собой гомозиготный случай.

В некоторых воплощениях определение того, имеется ли дисбаланс, или иначе, отношение к генотипу матери (например, один N для одного M может означать отсутствие дисбаланса) можно выполнить, используя пороговые величины. Например, если имеется достаточно большое отклонение (например, в процентах) от генотипа матери, тогда плод можно определить как наследующий аллель с более узким распределением по размерам. В одном воплощении величина порога может зависеть от процентного содержания фетальных нуклеиновых кислот в материнском образце. Если процент фетальных

нуклеиновых кислот выше, тогда можно ожидать большее отклонение, и таким образом, можно использовать большую величину порога (например, в случае различия в распределении по размерам одного аллеля относительно другого. Если процент фетальных нуклеиновых кислот меньше, тогда можно ожидать меньшего отклонения, и, таким образом, можно использовать меньшую величину порога.

В одном воплощении для определения того, какой аллель матери может иметь распределение по размерам, измененное из-за фетальных нуклеиновых кислот, можно использовать генотип отца. В случаях, когда отец является гомозиготным для генотипа, это может создать возможность для сужения возможного генотипа плода просто до аллеля, наследуемого от матери, так как известно, какой аллель пришел от отца. После сужения определение генома плода может быть более точным, так как может потребоваться испытывать только две возможности. В одном воплощении используют одну и ту же величину порога, несмотря на то, что какой генотип наследуется от отца. В другом воплощении можно использовать различные пороговые величины.

В различных воплощениях любое распределение по размерам, упомянутое в данном описании, можно использовать при таких определениях дисбаланса последовательностей. В некоторых воплощениях также можно обеспечить уровень точности. Например, также можно использовать классификацию «неопределенный», помимо классификации дисбаланса и баланса. В таком случае некоторые определения можно определить с высокой степенью достоверности, в то время как величины в среднем участке могут требовать дополнительных экспериментальных точек.

#### XI. Анализ размеров для определения гаплотипа, унаследованного плодом

Применение анализа размеров также можно расширить до определения, какой материнский гаплотип переходит к плоду. Гаплотип может относиться к аллелям во многих локусах. Определение термина «гаплотип» можно найти в разделе «Определения» настоящей заявки. Гаплотип плода можно использовать для определения того, наследует ли плод мутированный ген, имеет ли дисбаланс определенного аллеля или для других целей. Таким образом, гаплотип можно использовать так же, как генотип, но так как существует больше локусов, можно использовать меньший объем образца крови для достижения такой же, или даже лучшей статистической достоверности определения гаплотипов плода. В одном аспекте дисбаланс последовательностей можно определить относительно гаплотипа матери. В данном случае гаплотип может быть представлен как ряд полиморфизмов, например, SNP, каждого в определенной локации в геноме, при которых известно изменение последовательности.

В одном воплощении предлагается способ определения гаплотипа плода путем

анализа дисбалансов последовательностей в материнской плазме. В одном аспекте для определения дисбаланса последовательностей используют показатели разделения (например, различия) в профилях размеров между гаплотипами. В одном воплощении гаплотипы матери (например, Нар I и Нар II) определяют путем анализа (например, секвенирования (He D. et al., *Bioinformatics*, 2010, 26: i183-i190) или одномолекулярного гаплотипирования (Ding C. et al., *Proc. Natl. Acad. Sci. USA*, 2003, 100: 7449-7453; и Xiao M. et al., *Nat. Methods*, 2009, 6: 199-201) материнского образца (например, образца, который не содержит фетальных нуклеиновых кислот). В другом воплощении гаплотип(ы) матери можно определить с использованием анализа образцов от родителей матери, сиблингов, предыдущего ребенка или других родственников. В еще одном воплощении в случае полиморфизмов при сильном неравновесном сцеплении, зная генотип матери в одном локусе, можно предположить генотип в других локусах, например, когда аллели нормально появляются в одной и той же последовательности, т.е. гаплотип. Таким образом, гаплотип матери может быть потенциально определен из генотипа матери. Генотип в более чем одном локусе также может быть определен, где каждый определенный генотип может предполагать генотипы в других локусах, и следовательно, устанавливается гаплотип.

В одном воплощении также определяют генотип отца. Такую информацию можно использовать для определения, если отец является гомозиготным или гетерозиготным при определенных SNP. Можно осуществить прямое определение отцовского аллеля в каждом локусе. Соответственно, можно определить, если отцовский аллель является тем же, что аллель Нар I или Нар II матери, обозначение SNP типа  $\alpha$  или типа  $\beta$ .

В другом воплощении также определяют гаплотип отца. Гаплотип отца можно определить анализом (например, секвенированием (He D. et al., *Bioinformatics*, 2010, 26: i183-i190) или одномолекулярным гаплотипированием (Ding C. et al., *Proc. Natl. Acad. Sci. USA*, 2003, 100: 7449-7453; и Xiao M. et al., *Nat. Methods*, 2009, 6: 199-201) отцовского образца или анализом образцов от родителей отца, сиблингов, предыдущего ребенка или других родственников. С другой стороны, в случае полиморфизмов при сильном неравновесном сцеплении, зная генотип отца в одном локусе, можно предположить генотип в других локусах, например, когда аллели нормально появляются в одной и той же последовательности, т.е. гаплотип. Генотип в более чем одном локусе также может быть определен, где каждый определенный генотип может предполагать генотипы в других локусах, и, следовательно, устанавливается гаплотип.

Другим применением таких воплощений может быть определение, существует ли или не существует дисбаланс последовательностей (например, путем обнаружения

различия в распределении по размерам) в материнском образце, содержащем фетальные нуклеиновые кислоты, путем анализа SNP, где как отец, так и мать являются гетерозиготными. Генотип отца или гаплотип посредством этого выводится на основании того, существует или не существует дисбаланс последовательностей.

Фиг. 26 показывает пример, когда мать является гетерозиготной, а отец является гомозиготным в анализируемых локусах SNP. Для того чтобы осуществить анализ размеров, средоточение может быть на подмножестве SNP, где мать является гетерозиготной, а отец является гомозиготным. Две гомологичные хромосомы матери обозначают Нар I и Нар II, соответственно. Для каждого из указанных SNP можно определить, который аллель из двух материнских аллелей располагается в Нар I и который в Нар II. SNP определяют как тип  $\alpha$ , если отцовские аллели идентичны материнскому аллелю в Нар I, и как тип  $\beta$ , если отцовские аллели идентичны материнскому аллелю в Нар II. Дополнительное описание определений генома плода можно найти в заявке «Геномный анализ плода из материнского биологического образца», упомянутой выше.

Как только гаплотип матери и генотипы или гаплотип отца известны, можно анализировать распределение по размерам фрагментов, ассоциированных с SNP, для определения гаплотипа плода путем идентификации, существует ли дисбаланс последовательностей в подмножестве SNP. В альтернативном воплощении генотип или гаплотип отца неизвестен наверняка, но, тем не менее, может быть установлен с использованием, например, статистической процедуры, как вероятный генотип или гаплотип отца, на основании, например, повторяемости известных генотипов или гаплотипов в испытываемой популяции. Затем определяют дисбаланс последовательностей (определенный через дисбаланс размеров) для каждого SNP, например, так, как описано в разделе X.

Например, если отец является гомозиготным для аллеля в Нар I (тип  $\alpha$ ), тогда дисбаланса размеров не будет (плод наследует Нар II от матери и, таким образом, является гетерозиготным для Нар I и Нар II именно как мать) или дисбаланс размеров будет иметь место, когда Нар I имеет более узкое распределение по размерам (плод наследует Нар I от матери и, таким образом, является гомозиготным для Нар I). Если отец является гомозиготным для Нар II (тип  $\beta$ , как описано ниже), тогда дисбаланса не будет (плод наследует Нар I от матери и, таким образом, является гетерозиготным для Нар I и Нар II), или дисбаланс будет иметь место, когда Нар II имеет более узкое распределение по размерам (плод наследует Нар II от матери и, таким образом, является гомозиготным для Нар I). Также можно использовать промежуточную классификацию или изменяющиеся

уровни достоверности между балансом и дисбалансом. Типично можно использовать схожие пороговые величины для любого типа. В различных воплощениях можно использовать любое распределение по размерам, указанное в данном описании.

В одном воплощении можно анализировать распределение по размерам фрагментов каждой из множества локаций SNP для определения двух гаплотипов плода. Например, различие между распределением по размерам фрагментов для аллелей SNP в гаплотипе можно сравнить с распределением по размерам фрагментов для аллелей SNP в другом гаплотипе. Статистические результаты можно анализировать различными путями, например, осуществляя определение для каждого SNP и затем, принимая большинство (баланс, дисбаланс, и может быть включая промежуточное состояние) за гаплотип. Как другой пример, размеры можно агрегировать по SNP (например, для получения распределения по усредненным или срединным размерам, которое сравнивают с пороговой величиной). Или можно использовать комбинацию двух способов. Другое воплощение может представлять собой использование экстремума частных значений, например, наименьшее различие для определенного SNP.

Фиг. 27 показывает пример, когда плод наследует Нар I от матери, когда гаплотип отца такой, как показанный на фиг. 26. Для SNP типа  $\alpha$  (заклученных в незаштрихованные прямоугольники) плод будет наследовать от отца аллели, идентичные аллелям, расположенным в материнском Нар I. Следовательно, плод будет гомозиготным для аллелей в Нар I. В результате в материнской плазме распределение по размерам аллелей в Нар I будет более узким, чем аллеля в Нар II. Для SNP типа  $\beta$  (заклученных в заштрихованные прямоугольники) плод будет наследовать от отца аллели, идентичные аллелям, расположенным в материнском Нар II. Следовательно, плод будет гетерозиготным. В результате в материнской плазме распределение по размерам аллелей в Нар I и Нар II будет одинаковым.

В одном воплощении SNP одного и того же типа (типа  $\alpha$  или типа  $\beta$ ) можно анализировать вместе. Для SNP типа  $\alpha$  распределение по размерам молекул, содержащих аллели в Нар I, будет более узким, чем распределение по размерам молекул, содержащих аллели в Нар II. Для SNP типа  $\beta$  распределение по размерам молекул, содержащих аллели в Нар I и Нар II, будет одинаковым. Иными словами, если распределение по размерам молекул, содержащих аллели в Нар I, будет более узким, чем молекул, содержащих аллели в Нар II, плод является гомозиготным для Нар I. Если распределение по размерам молекул, содержащих Нар I и Нар II, будет одинаковым, плод является гетерозиготным.

### Пример

Следующий эксперимент используют для проверки точности анализа определения гаплотипа, унаследованного плодом. Привлекли пару, пришедшую в акушерскую клинику, для проведения пренатальной диагностики  $\beta$ -талассемии. Образцы крови берут у отца и матери. У матери образец крови берут перед взятием образца хориальных ворсинок (CVS) на 12 неделе беременности. Часть образца CVS хранят для эксперимента.

Экстрагируют ДНК из лейкоцитной пленки отца и матери и образца CVS. Эти образцы ДНК подвергают анализу на системе Affymetrix Genome-Wide Human SNP Array 6.0 для определения генотипов отца, матери и плода. В этом эксперименте данные по CVS используют для установления гаплотипа матери. Однако в клиническом осуществлении испытания гаплотип матери можно установить другими способами, как описано выше. Данные по CVS также используют для подтверждения точности определения с использованием способов, для которых CVS не требуется.

В описываемой иллюстрации сосредотачиваются на информативном подмножестве SNP, в которых мать является гетерозиготной, а отец является гомозиготным. В таком подмножестве SNP генотипы пары и плода используют для конструирования гаплотипов матери. Определяют, что гаплотип I (Hap I) представляет собой ряд аллелей, которые мать передает плоду, в то время как гаплотип II (Hap II) представляет собой ряд аллелей, которые плод не получает от матери.

Затем информативные SNP делят на два подтипа, а именно, тип  $\alpha$  и тип  $\beta$ . Для типа  $\alpha$  отцовские аллели идентичны материнскому аллелю в Hap I. В случае таких SNP плод будет наследовать от родителей один и тот же аллель (аллель в Hap I), и, следовательно, плод будет гомозиготным для SNP в Hap I. Для типа  $\beta$  отцовские аллели идентичны материнскому аллелю в Hap II. В случае таких SNP плод будет наследовать аллели в Hap I от матери и другие аллели (идентичные аллелям в Hap II) от отца, и, следовательно, плод будет гетерозиготным.

ДНК, экстрагированную из плазмы матери, подвергают массивному параллельному секвенированию с использованием платформы геномного анализатора Illumina. Выполняют секвенирование спаренных концов молекул ДНК из плазмы. Каждую молекулу секвенируют по каждому концу на 50 п.о., что составляет 100 п.о. на молекулу. Два конца каждой последовательности выравнивают с геномом человека, с немаскированными повторами (Hg18 NCBI.36 downloaded from UCSC genome.ucsc.edu), с использованием программы SOAP2 от Beijing Genomics Institute at Shenzhen (soap.genomics.org.cn/) (Li R. et al., *Bioinformatics*, 2009, 25(15): 1966-7).

В одном воплощении статистический показатель доли общей длины, вносимой

короткими фрагментами для гаплотипов, используют для определения того, какой гаплотип матери передается плоду. Как пример, используют хромосому 22 для иллюстрации того, как можно использовать анализ размеров для установления того, какой гаплотип матери передается плоду. Сначала хромосому 22 делят на несколько сегментов, причем каждый сегмент содержит 50 информативных SNP, в которых мать является гетерозиготной, а отец является гомозиготным (хромосома 22 и сегменты являются примерами последовательностей). Для каждого сегмента фрагменты (примеры молекул, которые являются частью последовательностей), перекрывающие такие информативные SNP, делят на две группы, а именно, Нар I и Нар II, соответственно, которому из двух гаплотипов матери такие фрагменты соответствуют. Для каждого сегмента определяют сумму общей длины всех фрагментов, картирующихся к Нар I матери и Нар II матери. Затем подобным образом в каждом сегменте определяют сумму общей длины коротких фрагментов, картирующихся к Нар I и Нар II матери. С целью иллюстрации, в данном примере для вычисления общей длины коротких фрагментов как короткие определяют фрагменты в 150 п.о. или более короткие. Из полученных длин в каждом сегменте для фрагментов ДНК, картирующихся к Нар I и Нар II, можно вычислить доли, вносимые короткими фрагментами.

Фиг. 28 показывает таблицу, иллюстрирующую анализ размеров для SNP типа  $\alpha$  на хромосоме 22 согласно воплощениям настоящего изобретения.  $\Delta F_{(\text{Нар I} - \text{Нар II})}$  представляет собой различие между Нар I и Нар II в доли общей длины, вносимой короткими фрагментами. Когда последний сегмент, который состоит только из 28 SNP, исключают,  $\Delta F_{(\text{Нар I} - \text{Нар II})}$  колеблется от 0,0288 до 0,0701. Положительные значения  $\Delta F_{(\text{Нар I} - \text{Нар II})}$  показывают, что величины F Нар I всегда больше, чем величины Нар II. Так как величина F определяется как доля общей длины, вносимая короткими фрагментами, более высокая величина F показывает, что более высокая доля общей длины вносится короткими фрагментами. Иными словами, такие результаты показывают, что фрагменты ДНК, содержащие аллели в Нар I, короче, чем фрагменты, содержащие аллели в Нар II, для каждого из участков, состоящих из 50 анализируемых SNP. Это показывает, что распределение по размерам Нар I более узкое, чем распределение по размерам Нар II. Следовательно, можно установить, что плод является гомозиготным для аллелей в Нар I. Как указывалось выше, в одном воплощении можно провести классификацию для каждого сегмента на основании показателя разделения (например,  $\Delta F$ ) между двумя геномными локациями в Нар I и Нар II, и затем можно провести общую классификацию на основании соответствующей классификации сегментов. В другом воплощении можно определить общий показатель разделения (например, усредненный показатель разделения)

из показателей разделения для каждого сегмента, и такой общий статистический показатель можно использовать для определения классификации.

В одном воплощении для определения можно использовать пороговую величину примерно 0,02, если существует дисбаланс. Если используют срединное или усредненное  $\Delta F$ , можно использовать бóльшую пороговую величину и сохранить точность. Пороговые величины также можно использовать для промежуточных результатов, например, при необходимости дополнительного анализа промежуточным может быть интервал от 0,015 до 0,025.

Фиг. 29 показывает таблицу, иллюстрирующую анализ размеров для SNP типа  $\beta$  на хромосоме 22 согласно воплощениям настоящего изобретения.  $\Delta F_{(\text{Нар I} - \text{Нар II})}$  представляет собой разницу в доли общей длины, вносимой короткими фрагментами, для Нар I и Нар II. Величины  $\Delta F_{(\text{Нар I} - \text{Нар II})}$  колеблются от -0,0203 до 0,0207 со средним 0,0003. Небольшая величина  $\Delta F_{(\text{Нар I} - \text{Нар II})}$  совместима с одинаковым распределением по размерам фрагментов, картирующихся к Нар I и Нар II. Таким образом, можно установить, что плод является гетерозиготным для Нар I и Нар II. Так как тип SNP  $\beta$  определяется как SNP, в которых отцовские аллели идентичны материнским аллелям в Нар I, такой результат означает, что плод наследует Нар I от матери.

Фиг. 30 показывает график  $\Delta F_{(\text{Нар I} - \text{Нар II})}$  для SNP типа  $\alpha$  и типа  $\beta$  на хромосоме 22 согласно воплощениям настоящего изобретения. Для SNP типа  $\alpha$  и распределение по размерам фрагментов Нар I более узкое, чем распределение по размерам фрагментов Нар II, что выражается в величинах  $\Delta F_{(\text{Нар I} - \text{Нар II})}$  больше нуля. Для SNP типа  $\beta$  не имеется различия в распределении по размерам фрагментов Нар I и Нар II; следовательно, величины  $\Delta F_{(\text{Нар I} - \text{Нар II})}$  собираются вблизи 0. С использованием порога в 0,025 анализом  $\Delta F_{(\text{Нар I} - \text{Нар II})}$  можно правильно установить наследование плодом для всех 13 сегментов типа  $\alpha$  и 21 сегмента типа  $\beta$ .

## XII. Пример использования прицельного секвенирования

Следующий пример демонстрирует, что диагностический подход на основе размеров по воплощениям настоящего изобретения можно применить в формате прицельного секвенирования. В таком формате для секвенирования специфически намечаются участки генома, представляющие интерес для диагностики. Такой формат имеет преимущество в том, что секвенирование фокусируется на участке, представляющем интерес, в отличие от ситуации, включающей случайное секвенирование, когда часть производительности секвенирования используют для участков, которые не имеют непосредственного отношения к применению в диагностике.



Таким образом, можно ожидать, что формат прицельного секвенирования повысит производительность и снизит затраты на систему. Прицельное секвенирование можно выполнять с использованием любого формата, известного специалистам в данной области техники, включая систему захвата в фазе раствора (например, систему Agilent SureSelect), систему захвата в твердой фазе (например, систему Roche NimbleGen) или мишень-специфическую амплификацию (например, систему RainDance).

Образцы крови берут у восьми беременных женщин во время их первого триместра. В каждом случае экстрагируют ДНК из 3,2 мл плазмы с помощью мининабора DSP DNA Blood (Qiagen). Кариотипирование, выполненное на образцах хориальных ворсин (собранных после взятия образцов материнской крови), показывает, что четыре плода имеют Т21 (UK229, UK510, UK807, PW421), в то время как другие четыре являются эуплоидами мужского пола (PW226, PW263, PW316, PW370).

В каждом случае используют 5-30 нг ДНК для конструирования библиотеки ДНК с помощью набора для получения образцов спаренных концов (Illumina) согласно протоколу получения образцов для секвенирования с иммунопреципитацией с хроматином, представленного изготовителем. ДНК с лигированным адаптером очищают непосредственно с использованием спин-колонок, предоставленных в наборе для очистки продуктов ПЦР QIAquick (Qiagen), без дополнительного отбора по размерам. Затем амплифицируют ДНК с лигированным адаптером с использованием 15-циклов ПЦР со стандартными праймерами. Праймеры представляют собой праймеры для ПЦР PE 1.0 и 2.0 от Illumina. Библиотеки ДНК определяют количественно с использованием спектрофотометра NanoDrop ND-1000 (NanoDrop Technologies) и анализируют в биоанализаторе 2100 с использованием набора DNA 1000 (Agilent) для контроля за распределением по размерам. Получают 0,6-1 мкг библиотеки амплифицированной ДНК плазмы для каждого образца с усредненным размером примерно 290 п.о.

Библиотека SureSubset Human All Exon полученная от Agilent и содержит 37,8 Мм человеческих экзонов (каталожный номер 5190-2310). Для всех восьми случаев в данном исследовании 500 нг библиотеки амплифицированной ДНК плазмы в каждом случае инкубируют с зондами захвата в течение 24 часов при 65°C согласно инструкциям изготовителя. После гибридизации захваченные мишени отбирают путем сбора биотинилированных гибридов зонд/мишень с использованием магнитных шариков, покрытых стрептавидином (DynaMag-2 Invitrogen), и очищают в использовании набора для очистки MinElute PCR (Qiagen). Наконец, меченые библиотеки ДНК обогащают 12-циклами ПЦР амплификации с праймерами SureSelect GA PE от Agilent. Продукты ПЦР очищают с использованием набора для очистки продуктов ПЦР QIAquick

(Qiagen).

Восемь пар библиотек, обогащенных и небогащенных мишенью, загружают на 16 полос двухпроточных ячеек и затем секвенируют с помощью геномного анализатора *Ix* (Illumina) с использованием 36 п.о. x 2 в формате спаренных концов. Все данные по 36 п.о. выравнивают с эталонным геномом человека с немаскированными повторами (Hg18) (<http://genome.ucsc.edu>) с использованием программы выравнивания коротких олигонуклеотидов 2 (<http://soap.genome.org.cn/>). Данные по размерам фрагментов по спаренным концам определяют как колеблющиеся от 40 п.о. до 600 п.о. Размер каждого секвенированного фрагмента ДНК выводят из координат самых дальних нуклеотидов на обоих концах.

Анализ размеров для пренатального обнаружения трисомии 21 у плода

В данном примере доли общей длины, вносимые короткими фрагментами, вычисляют для хромосомы 21 и эталонных хромосом, представленные  $F_{21}$  и  $F_{ref}$ , соответственно. Эталонные хромосомы включают все аутосомы за исключением хромосом 13, 18 и 21. Общую длину вычисляют суммированием всех фрагментов ДНК в 600 п.о. или меньше. Различие между хромосомой 21 и эталонными хромосомами в долях длины ( $\Delta F$ ), вносимых короткими фрагментами, вычисляют как  $F_{21} - F_{ref}$ .

Фиг. 31А представляет собой таблицу, которая предоставляет анализ размеров согласно воплощениям настоящего изобретения ДНК плазмы без обогащения мишенью. В различных столбцах приводится общая длина фрагментов  $\leq 150$  п.о. для хромосомы 21, общая длина фрагментов  $\leq 600$  п.о. для хромосомы 21 и их отношение как  $F_{21}$ . В других столбцах приводится общая длина фрагментов  $\leq 150$  п.о. для эталонных хромосом, общая длина фрагментов  $\leq 600$  п.о. для эталонных хромосом и их отношение как  $F_{ref}$ . В последнем столбце приводится различие между двумя долями  $\Delta F$ .

Фиг 31В представляет собой таблицу, которая предоставляет анализ размеров согласно воплощениям настоящего изобретения ДНК плазмы с обогащением мишенью. Столбцы на фиг. 31 имеют тот же формат данных, как в таблице на фиг. 31А.

Фиг. 32 представляет собой график  $\Delta F$  для T21 и эуплоидных образцов с или без обогащения мишенью. Для образцов без обогащения мишенью с использованием величины порога для  $\Delta F$  0,005 образцы плазмы при беременности с T21 и эуплоидной беременности можно дифференцировать с точностью 100%. Для образцов с обогащением мишенью с использованием величины порога для  $\Delta F$  0,004 образцы плазмы при беременности с T21 и эуплоидной беременности можно дифференцировать с точностью 100%. Данный пример демонстрирует, что анализ на основе размеров можно выполнять с

использованием нацеленного секвенирования. Для обнаружения T21 может быть выгодным использование нацеленного секвенирования для хромосомы 21 и эталонной хромосомы, так что 50% секвенирования направлено на первую хромосому и остальное направлено на последнюю. Такое исполнение может уменьшить «расход» производительности секвенирования на участки, которые не относятся непосредственно к обнаружению T21. Такое исполнение может позволить секвенировать образцы от нескольких пациентов с использованием мультиплексного секвенирования (например, путем использования индексированного секвенирования или секвенирования со штриховым кодом).

Любые из компонентов программного обеспечения или функций, описанных в данной заявке, могут быть осуществлены в виде прикладных программ, выполняемых процессором, с использованием любого подходящего машинного языка, такого как, например, Java, C++ или Perl, с использованием, например, обычных или объектно-ориентированных методов. Прикладные программы могут храниться как набор инструкций или команд на машиночитаемом носителе информации для хранения и/или передачи данных, подходящие носители включают запоминающее устройство с произвольной выборкой (ЗУПВ), постоянное запоминающее устройство (ПЗУ), магнитный носитель, такой как жесткий диск или дискета, или оптический носитель, такой как компактный диск (CD) или DVD (цифровой универсальный диск), флэш-запоминающее устройство и т.п. Машиночитаемый носитель информации может представлять собой любую комбинацию таких устройств для хранения и передачи информации.

Такие программы также могут кодироваться и передаваться с использованием несущих сигналов, адаптированных для переноса по проводным, оптическим и/или беспроводным сетям, соответствующим различным протоколам, включая Интернет. Как таковой, машиночитаемый носитель информации согласно воплощению настоящего изобретения может быть создан с использованием сигнала данных, кодированного такими программами. Удобочитаемые компьютерные носители информации, кодированные прикладными программами, могут быть собраны в пакет с совместимым устройством или предоставлены отдельно от других устройств (например, через загрузку в Интернет). Любой такой машиночитаемый носитель информации может быть размещен на или в рамках одного компьютерного программного продукта (например, жесткого диска или всей компьютерной системы), и может быть представлен на или в рамках различных компьютерных программных продуктов в системе или сети. Компьютерная система может включать монитор или другой подходящий дисплей для предоставления пользователю

любых результатов, указанных в данном описании.

Пример компьютерной системы показан на фиг. 33. Подсистемы, показанные на фиг. 33, взаимосвязаны через системную магистраль 3375. Показаны дополнительные подсистемы, такие как принтер 3374, клавиатура 3378, фиксированный диск 3379, монитор 3376, которые подсоединены к адаптеру дисплея 3382, и другие подсистемы. Периферийные устройства и устройства ввода/вывода (I/O), которые соединены с контроллером I/O, могут быть соединены с компьютерной системой любыми способами, известными в технике, такими как последовательный порт 3377. Например, последовательный порт 3377 или внешний интерфейс 3381 можно использовать для соединения компьютера с глобальной сетью, такой как Интернет, вводным устройством «мышью» или сканером. Взаимосвязь через магистраль позволяет центральному процессору 3373 взаимодействовать с каждой подсистемой и управлять выполнением команд из памяти системы 3372 или с фиксированного диска 3379, а также обмен информацией между подсистемами. Память системы 3372 и/или фиксированный диск 3379 могут заключать в себе машиночитаемый носитель информации.

Компьютерная система может включать несколько одинаковых компонентов или подсистем, например, соединенных вместе наружным интерфейсом 3381 или внутренним интерфейсом. В некоторых воплощениях компьютерные системы, подсистема или приборы могут взаимодействовать по всей сети. В таких случаях один компьютер может рассматриваться как пользователь, а другой компьютер как обслуживающее устройство, где каждый может представлять собой часть компьютерной системы. Пользователь и обслуживающее устройство могут включать, каждый, несколько систем, подсистем или компонентов.

Специфические детали определенных воплощений можно комбинировать любым подходящим способом без отхода от сущности и объема воплощений изобретения. Однако другие воплощения изобретения могут относиться к специфическим воплощениям, касающимся каждого отдельного аспекта, или специфическим комбинациям таких отдельных аспектов.

Приведенное выше описание примеров воплощений изобретения представлено в целях иллюстрации и описания. Оно не предназначено для того, чтобы изобретение точно исчерпывалось или ограничивалось описанным, и в свете вышеизложенного возможны многие модификации и изменения. Воплощения выбраны и описаны с целью наилучшего пояснения принципов изобретения и его практического применения, посредством которых другие специалисты в данной области техники смогут наилучшим образом использовать изобретение в различных воплощениях и с различными модификациями, которые

подходят для определенного рассматриваемого применения.

Все публикации, патенты и заявки на патенты, цитированные в данном описании, включены в него в качестве ссылок.

## ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ пренатальной диагностики фетальной хромосомной анеуплоидии в биологическом образце, полученном от субъекта беременной женщины, при этом этот биологический образец включает молекулы нуклеиновых кислот, причем биологический образец включает молекулы нуклеиновых кислоты, которые происходят из геномов плода и субъекта беременной женщины, включающий:

для каждой из множества молекул нуклеиновых кислот в указанном биологическом образце:

измерение размера молекулы нуклеиновой кислоты,

идентификацию того, из какой хромосомы происходит молекула нуклеиновой кислоты;

для каждой из множества хромосом вычисление с использованием компьютерной системы статистического показателя распределения по размерам молекул нуклеиновых кислот, соответствующих этой хромосоме;

ранжирование хромосом относительно друг друга на основании статистического показателя размеров молекул нуклеиновых кислот соответствующих каждой хромосоме;

определение ранга для определенной хромосомы, основываясь на статистическом показателе размеров молекул нуклеиновых кислот, соответствующих указанной определенной хромосоме относительно статистического показателя размеров молекул нуклеиновых кислот, соответствующих другим хромосомам;

сравнение определенного ранга определенной хромосомы с рангом этой же определенной хромосомы, полученным из эталонного биологического образца; и

определение на основании сравнения существует ли фетальная хромосомная анеуплоидия для указанной определенной хромосомы.

2. Способ по п. 1, где сравнение определенного ранга определенной хромосомы с рангом этой же определенной хромосомы, полученным из эталонного биологического образца, включает:

для биологического образца определение первого различия между определенным рангом хромосомы и рангом эталонной хромосомы;

для эталонного биологического образца определение второго различия между рангом этой же определенной хромосомы и рангом эталонной хромосомы; и

сравнение различий.

3. Способ по п. 1, где измерение размера каждой из множества молекул нуклеиновых кислот включает:

секвенирование, по меньшей мере, части каждой из множества молекул нуклеиновых кислот, содержащихся в биологическом образце.

4. Способ по п. 3, где секвенированная часть каждой молекулы нуклеиновой кислоты включает оба конца соответствующей молекулы нуклеиновой кислоты.

5. Способ по п. 1, где статистический показатель включает медианный размер измеренных размеров молекул нуклеиновых кислот, соответствующих данной хромосоме.

6. Способ по п. 1, где статистический показатель включает усредненный размер измеренных размеров молекул нуклеиновых кислот, соответствующих данной хромосоме.

7. Способ по п. 1, где измеренный размер представляет собой длину или молекулярную массу.

8. Способ по п.7, где измеренный размер является длиной, и где длину измеряют по интенсивности флуоресценции.

9. Способ по п.1, где ранжирование идет от наименьшего до наивысшего статистического показателя.

10. Способ по п.1, где ранжирование идет от наивысшего до наименьшего статистического показателя.

11. Компьютерный программный продукт, включающий машиночитаемый носитель информации, хранящий множество команд для управления процессором, для выполнения любого из способов по п.п. 1-10.

12. Способ пренатальной диагностики фетальной хромосомной анеуплоидии в биологическом образце, полученном от субъекта беременной женщины, при этом этот биологический образец включает молекулы нуклеиновых кислот из плода и женского субъекта, включающий:

для каждой из множества молекул нуклеиновых кислот в биологическом образце:

измерение размера молекулы нуклеиновой кислоты,

идентификацию того, из какой хромосомы происходит эта молекула нуклеиновой кислоты;

вычисления с использованием компьютерной системы первого статистического показателя первого распределения по размерам размеров молекул нуклеиновых кислот, соответствующих определенной хромосоме;

вычисления с использованием компьютерной системы второго статистического показателя второго распределения по размерам размеров молекул нуклеиновых кислот, соответствующих одной или нескольким другим хромосомам;

определение показателя разделения между первым статистическим показателем и вторым статистическим показателем, где показатель разделения представляет собой разницу между первым статистическим показателем и вторым статистическим показателем или отношение первого статистического показателя и второго статистического показателя;

сравнение показателя разделения с одним или несколькими пороговыми значениями; и

определение на основании сравнения того, существует ли фетальная хромосомная анеуплоидия для указанной определенной хромосомы, причем это определение основано на том, будет ли показатель разделения превосходить одно или более пороговых значений.

13. Способ по п. 12, где показатель разделения представляет собой разницу между первым статистическим показателем и вторым статистическим показателем.

14. Способ по п. 12, где показатель разделения представляет собой отношение первого статистического показателя и второго статистического показателя.

15. Способ по п. 12, где первый статистический показатель и второй статистический показатель включают усредненный размер измеренных размеров молекул нуклеиновых кислот, соответствующих соответствующей хромосоме.

16. Способ по п. 12, где первый статистический показатель и второй статистический показатель соответственно включают первую сумму размеров молекул нуклеиновых кислот, которые соответствуют указанной определенной хромосоме и одной или нескольким другим хромосомам, соответственно, и имеют размер ниже первого порога по размеру.

17. Способ по п. 16, где первый статистический показатель включает первую сумму, деленную на вторую сумму, и при этом вторая сумма представляет собой сумму размеров всех молекул нуклеиновых кислот, соответствующих указанной определенной хромосоме, или представляет собой сумму размеров молекул нуклеиновых кислот, соответствующих указанной определенной хромосоме, которые имеют размер ниже второго порога по размеру, причем второй порог по размеру больше, чем первый порог по размеру.

18. Способ по п. 16, где первый статистический показатель включает первую сумму, деленную на вторую сумму, и при этом вторая сумма представляет собой сумму размеров соответствующих молекул нуклеиновых кислот, которые больше второго порога по размеру.



19. Способ по п. 16, также включающий:

повторение вычисления первого статистического показателя и второго статистического показателя для множества значений первого порога по размеру и определение показателя разделения между первым и вторым статистическими показателями для каждого из значений первого порога по размеру; и идентификацию экстремума показателей разделения.

20. Способ по п. 19, также включающий:

сравнение экстремума с одним или несколькими пороговыми значениями.

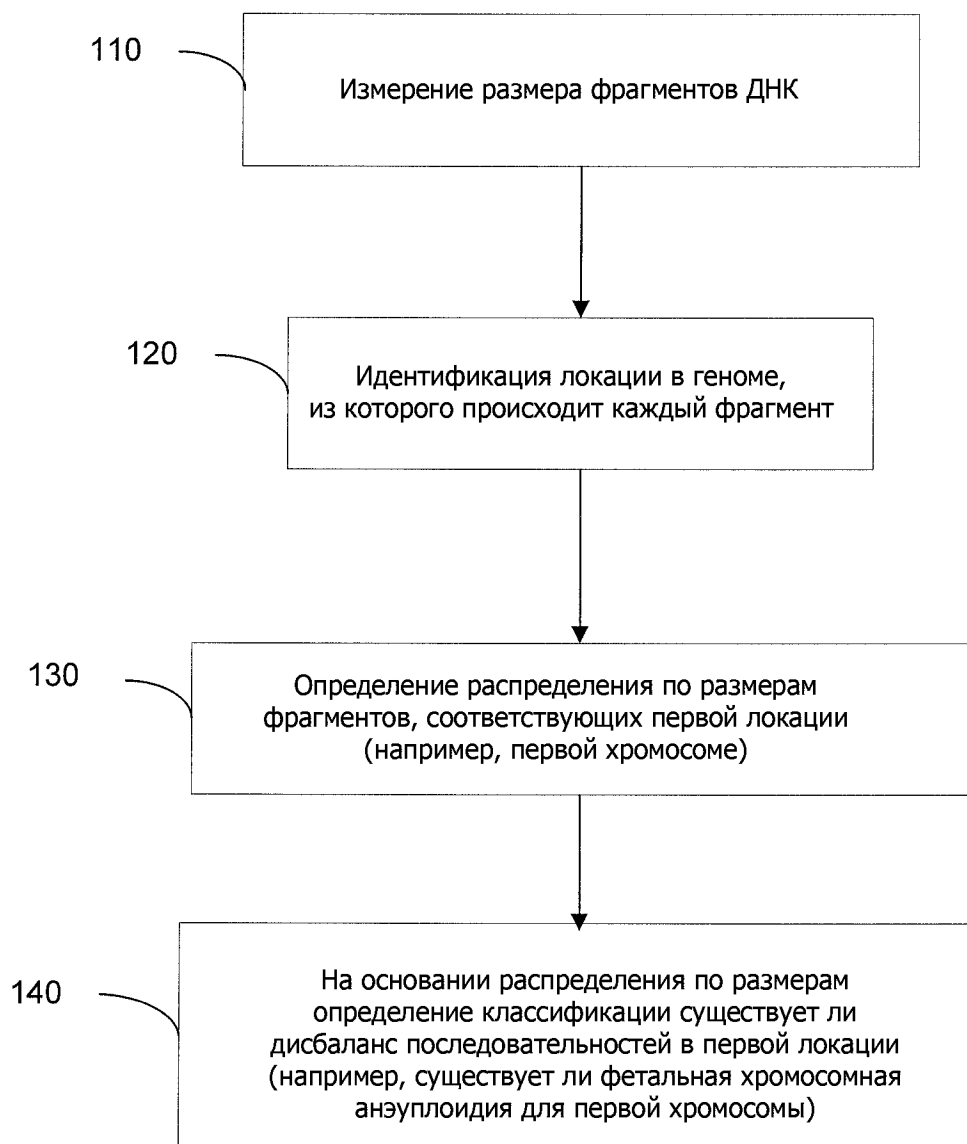
21. Способ по п. 19, также включающий:

идентификацию значения первого порога по размеру, соответствующего экстремуму; и

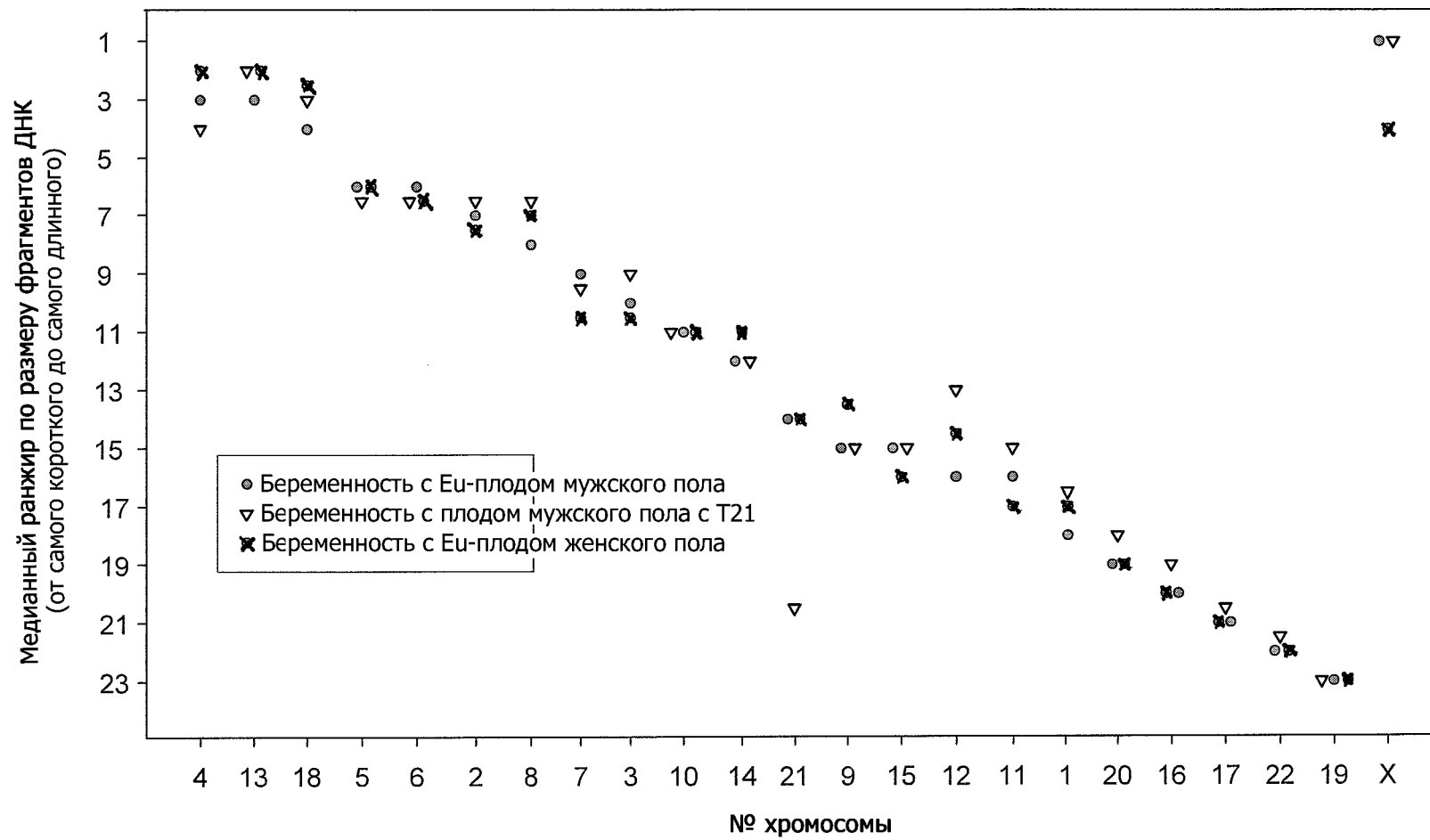
сравнение значения первого порога по размеру, соответствующего экстремуму, с одним или несколькими пороговыми значениями.

22. Способ по п. 12, где одно или несколько пороговых значений определены на основании измерения процентного содержания ДНК плода в биологическом образце, где при большем проценте ДНК плода первому пороговому значению из одного и более пороговых значений определяют более высокое значение, и где при меньшем проценте ДНК плода первому пороговому значению определяют более низкое значение.

23. Компьютерный программный продукт, включающий машиночитаемый носитель информации, хранящий множество команд для управления процессором, для выполнения любого из способов по п.п. 12-22.

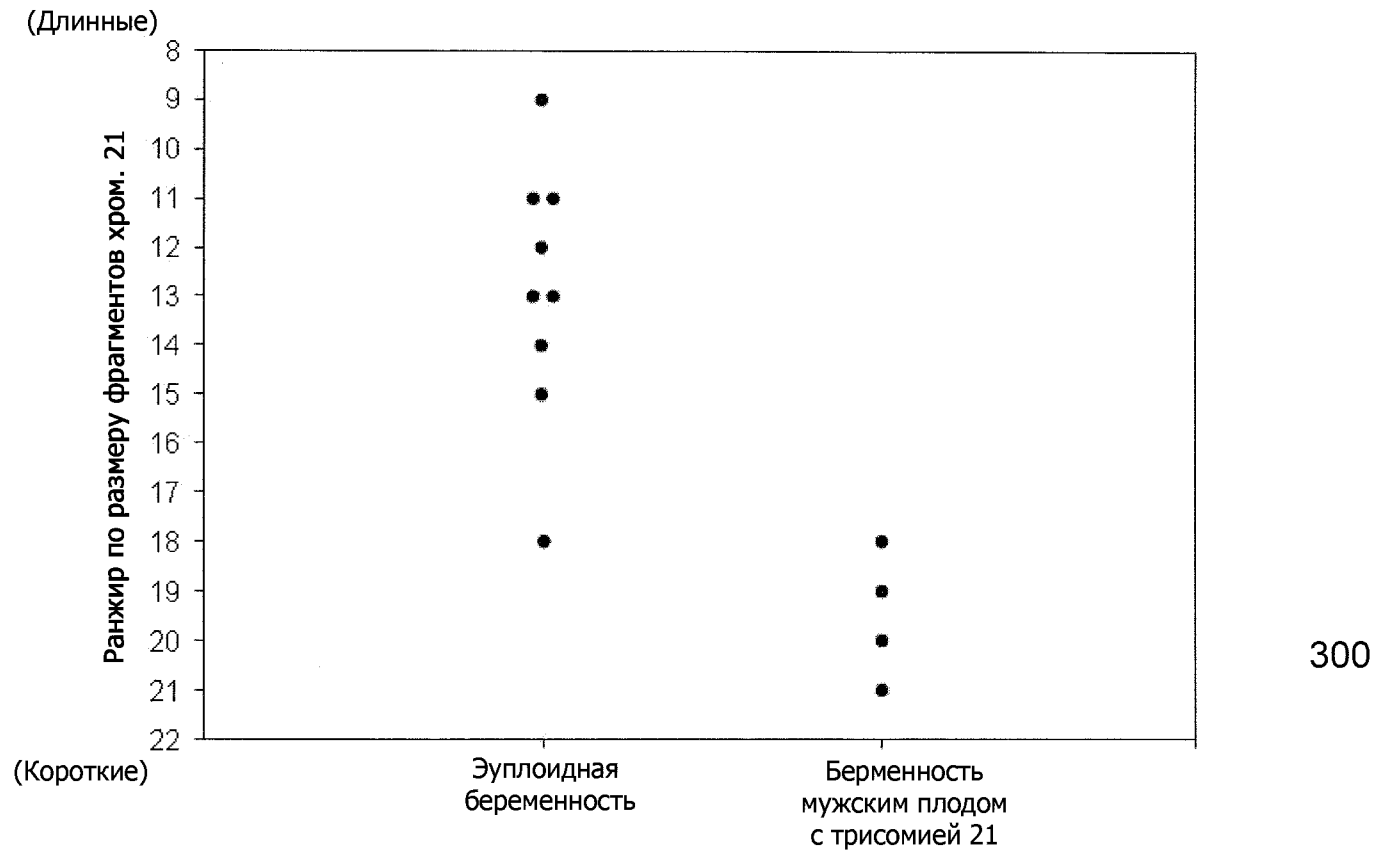


Фиг. 1

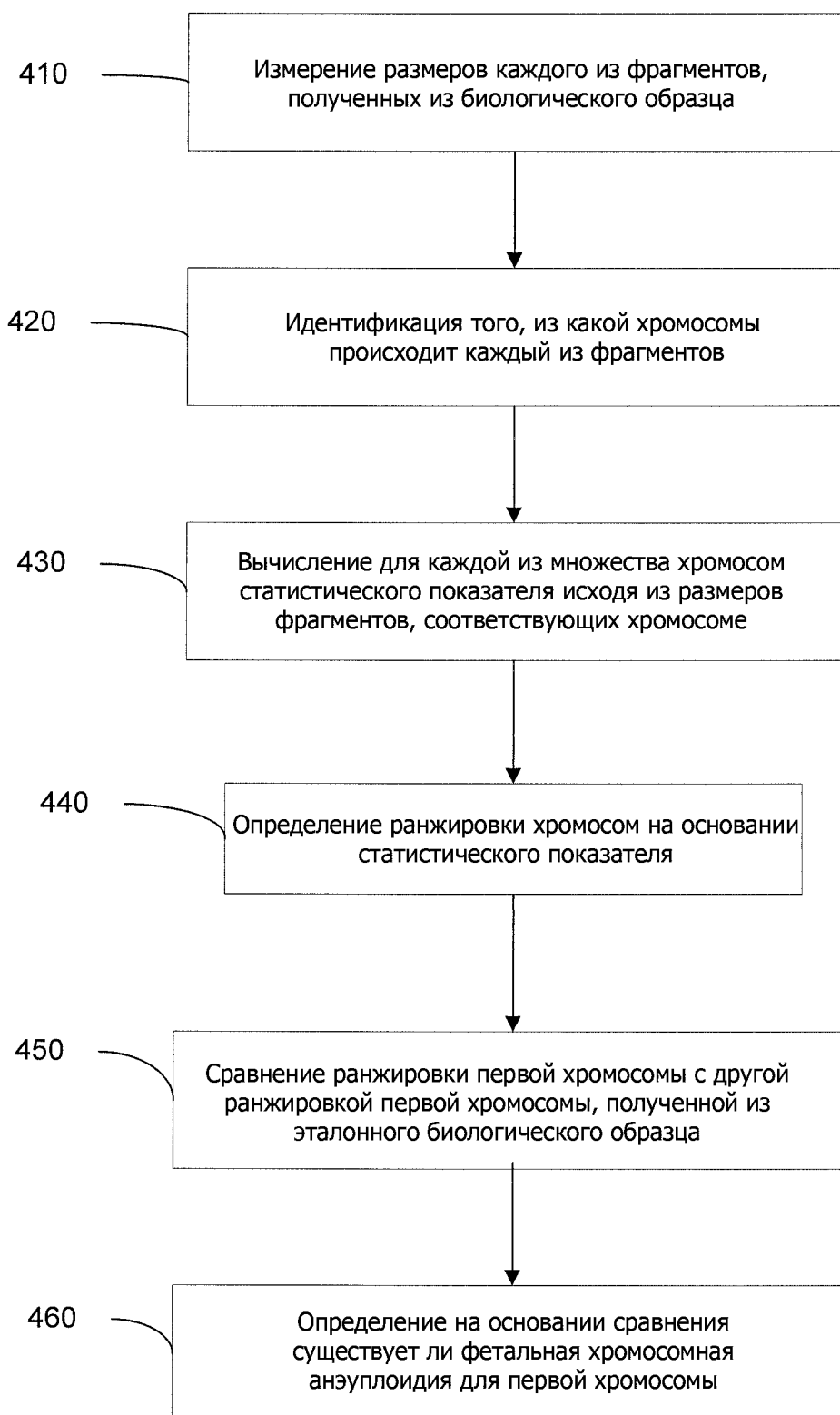


Фиг. 2

**Ранжир хромосомы 21 по размерам последовательностей,  
выровненных с различными хромосомами  
(сравнивают 22 аутосомы вместе с хромосомой X)**

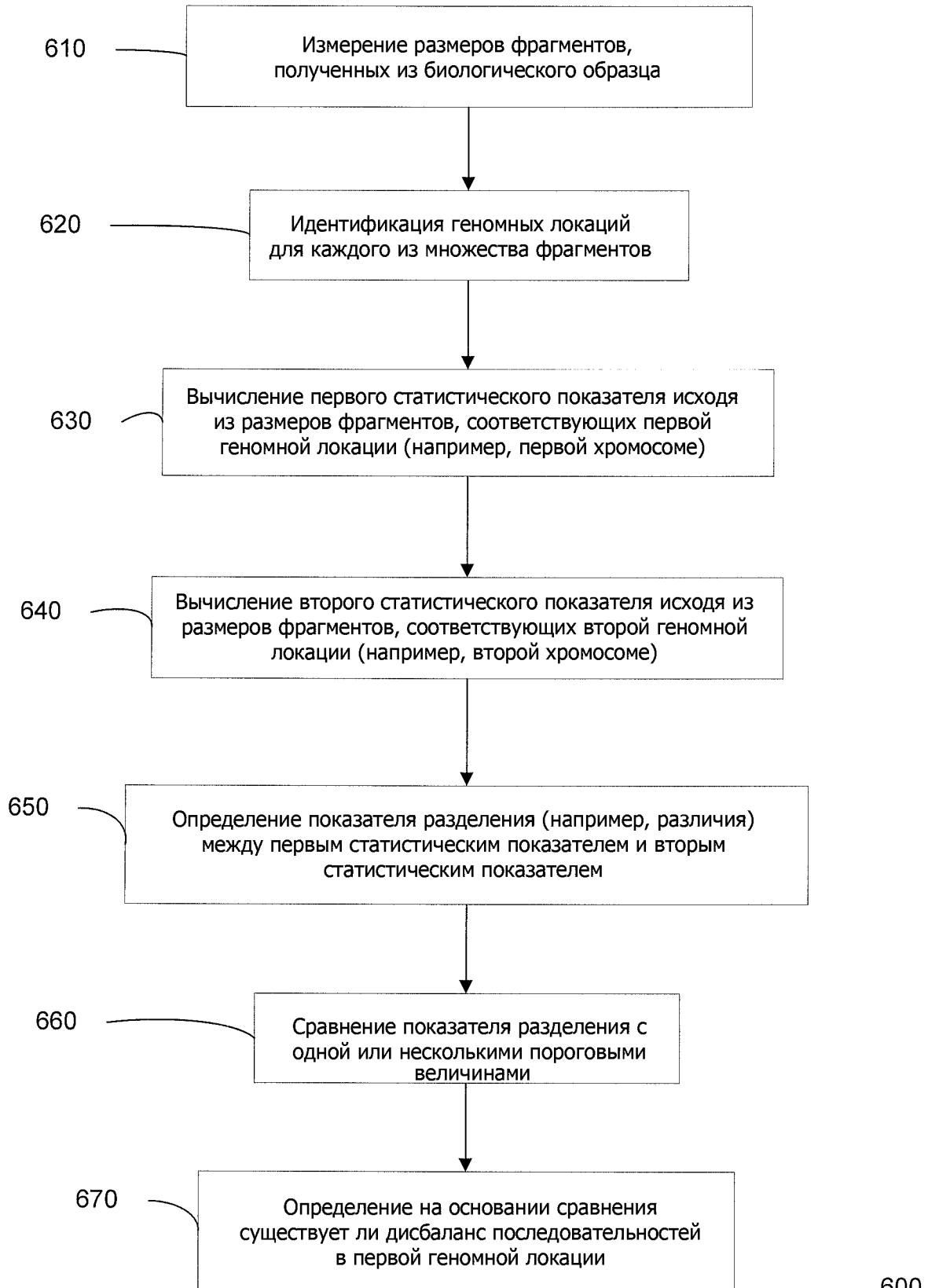


**Фиг. 3**

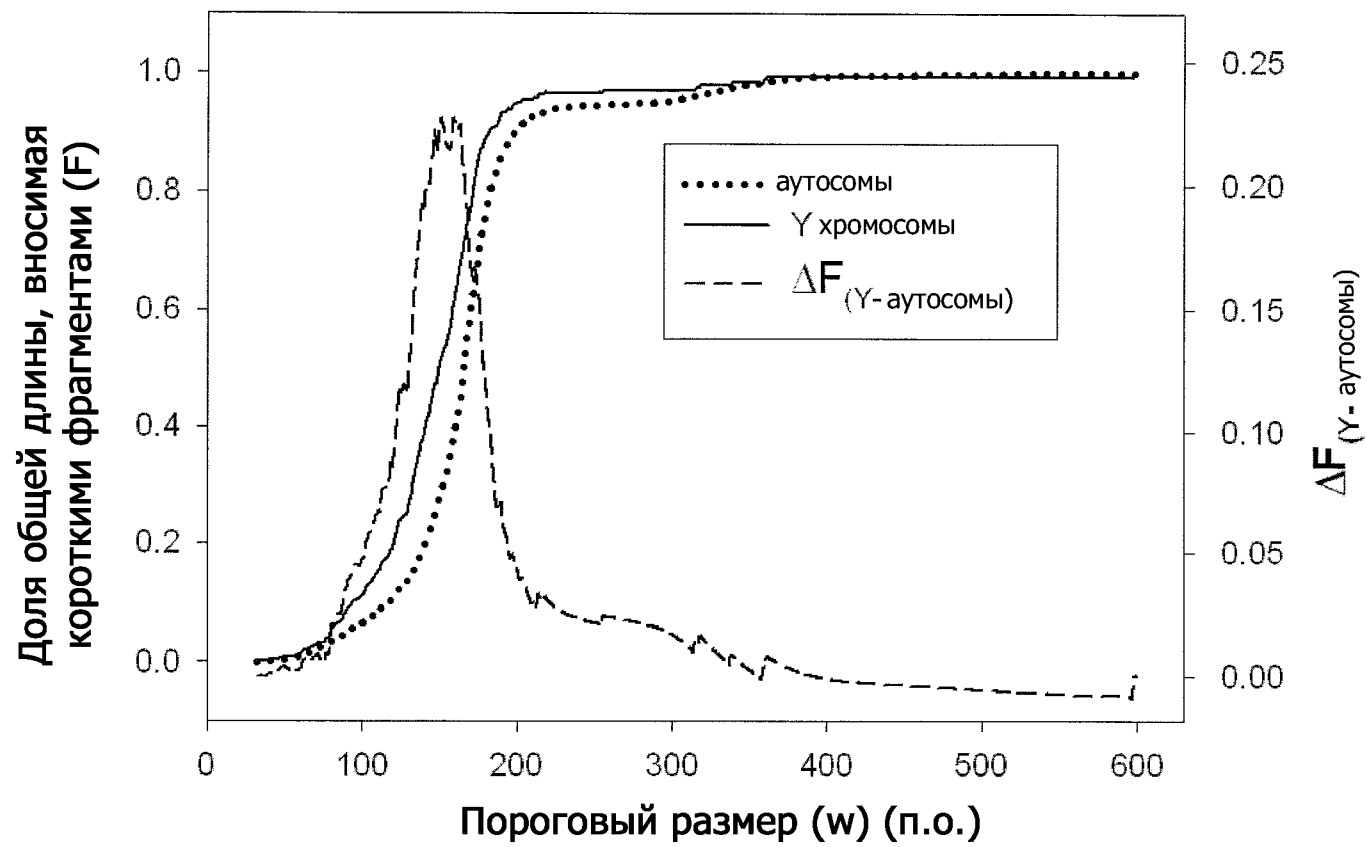


Фиг. 4

Название образца	Триместр	Фетальный кариотип	Средний размер фрагментов хром. 21	Средний размер фрагментов хром. 7	Различие между средними хром. 7 и хром. 21	Величина* Р (хром. 7 против хром. 21)	Средний размер фрагментов хром. 14	Различие между средними хром. 14 и хром. 21	Величина* Р (хром. 14 против хром. 21)
M4800- мужской пол	3-ий	Эуплоидный	156.2	157.0	0.8	0.08	156.1	-0.1	0.822
M4801- мужской пол	3-ий	Эуплоидный	151.2	151.7	0.5	0.085	151.5	0.3	0.564
M4814- женский пол	3-ий	Эуплоидный	157.8	157.3	-0.5	0.106	156.8	-0.9	0.022
PW006-Ец-мужской пол	1-ый	Эуплоидный	146.0	146.1	0.0	0.527	145.9	-0.1	0.988
PW007-Ец-мужской пол	1-ый	Эуплоидный	152.1	152.7	0.6	0.033	152.3	0.2	0.679
PW008-Ец- мужской пол	1-ый	Эуплоидный	147.2	147.3	0.1	0.538	147.0	-0.2	0.688
PW012- Ец-мужской пол	1-ый	Эуплоидный	146.7	146.8	0.1	0.333	146.7	0.1	0.68
PW020- Ец-мужской пол	1-ый	Эуплоидный	139.7	140.0	0.3	0.174	139.7	0.0	0.741
PW009-Ец- женский пол	1-ый	Эуплоидный	151.9	152.2	0.3	0.068	152.1	0.2	0.405
PW010-Ец- женский пол	1-ый	Эуплоидный	148.3	148.5	0.2	0.101	148.4	0.1	0.286
PW016- Ец-женский пол	1-ый	Эуплоидный	141.5	141.1	-0.4	0.782	141.2	-0.3	0.842
PW022- Ец-женский пол	1-ый	Эуплоидный	140.8	140.7	-0.1	0.362	140.5	-0.4	0.94
M2849- T21-мужской пол	1-ый	T21	145.9	147.0	1.1	<0.001	146.8	0.8	<0.001
M4386- T21-мужской пол	1-ый	T21	144.3	146.3	2.0	<0.001	145.9	1.6	<0.001
M4467-T21- мужской пол	1-ый	T21	141.0	142.7	1.7	<0.001	142.3	1.3	<0.001
M4620-T21- мужской пол	1-ый	T21	149.7	151.3	1.6	<0.001	150.9	1.2	<0.001



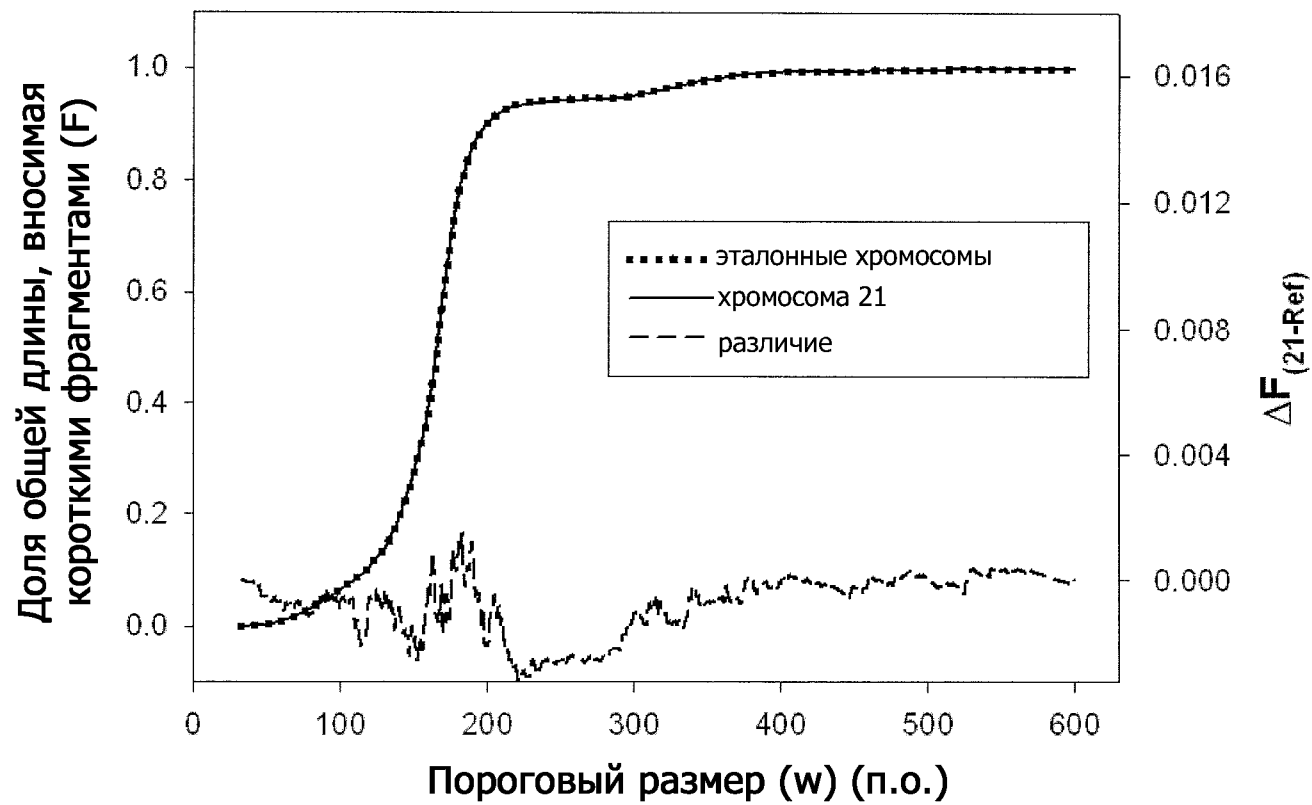
Фиг. 6



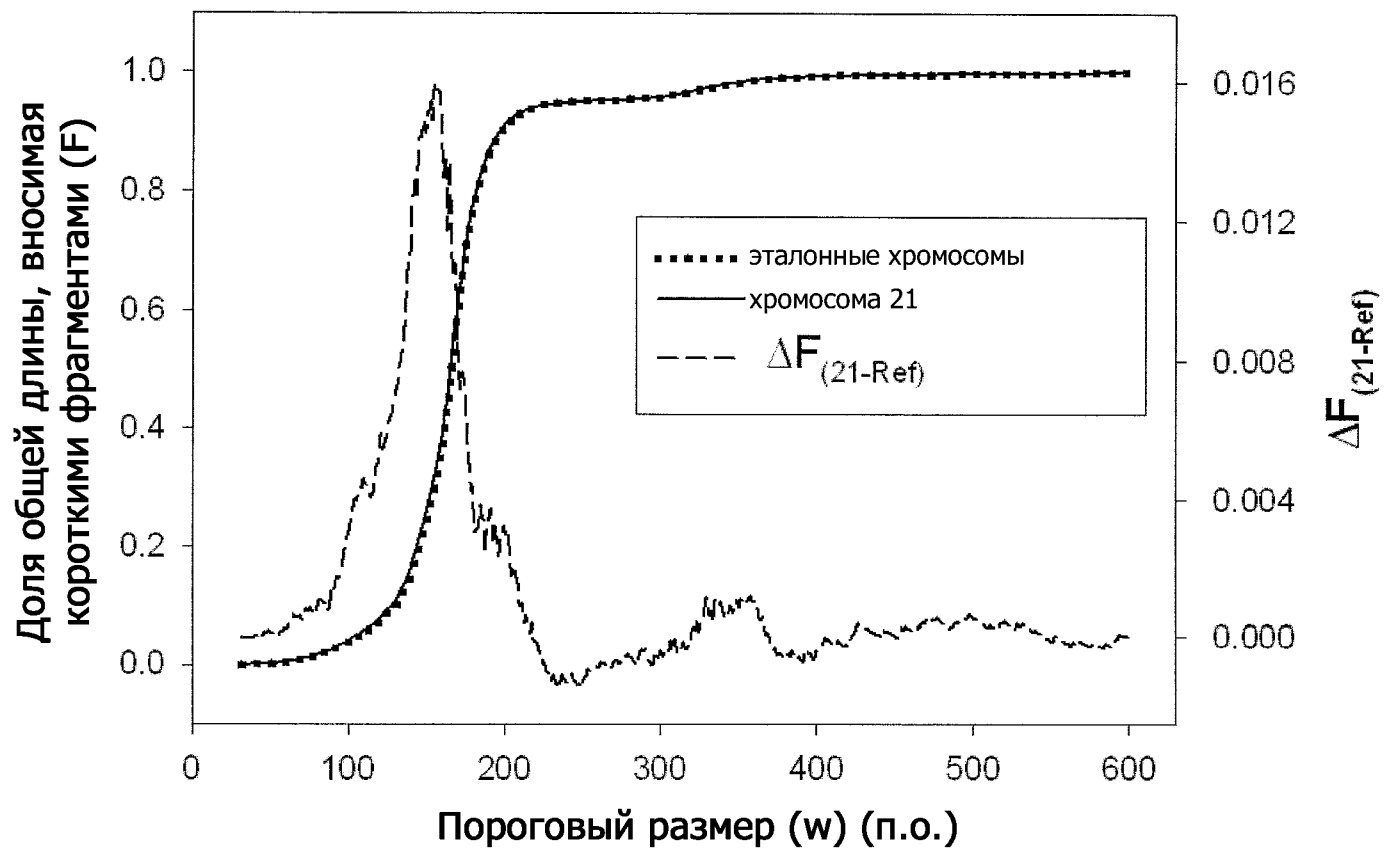
Фиг. 7

700

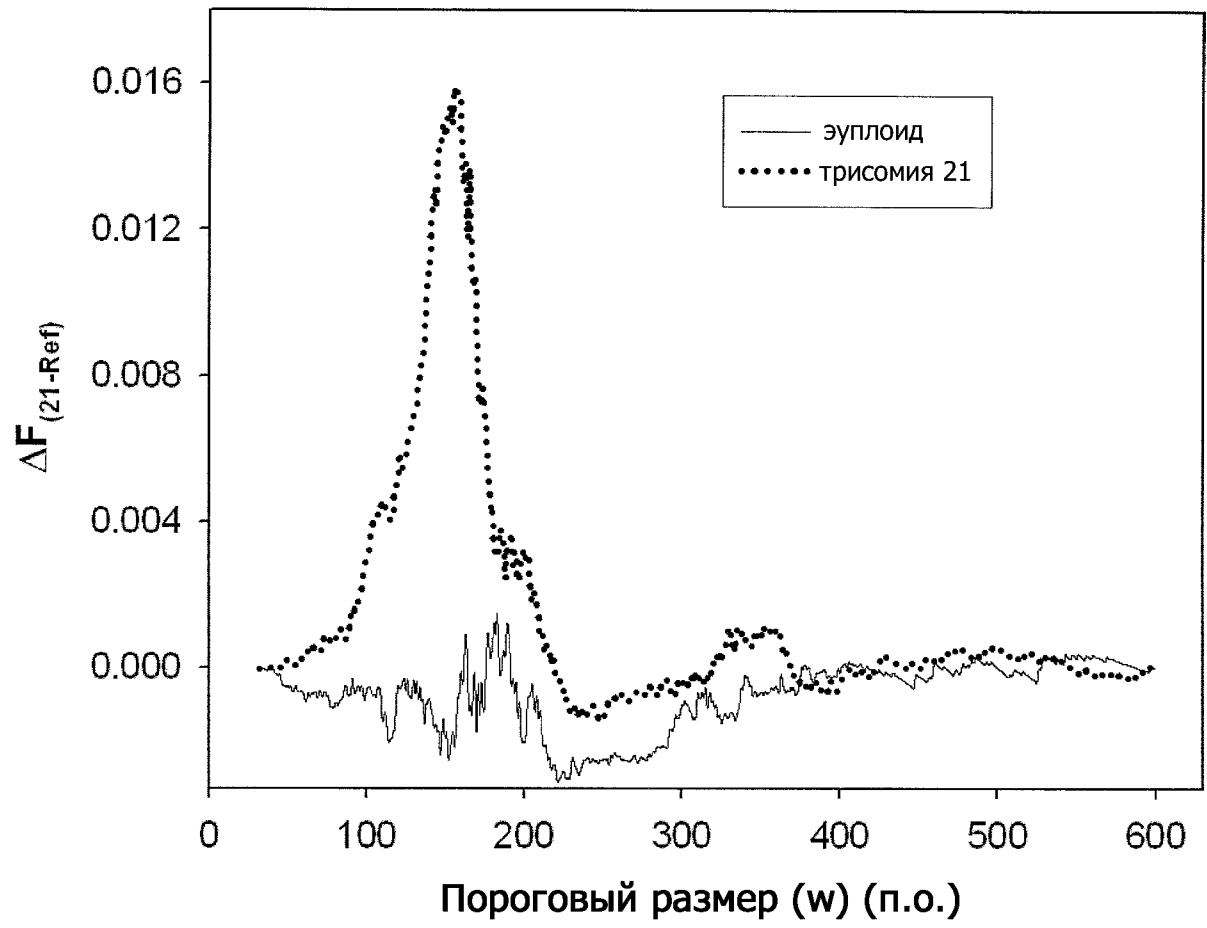




Фиг. 8



Фиг. 9



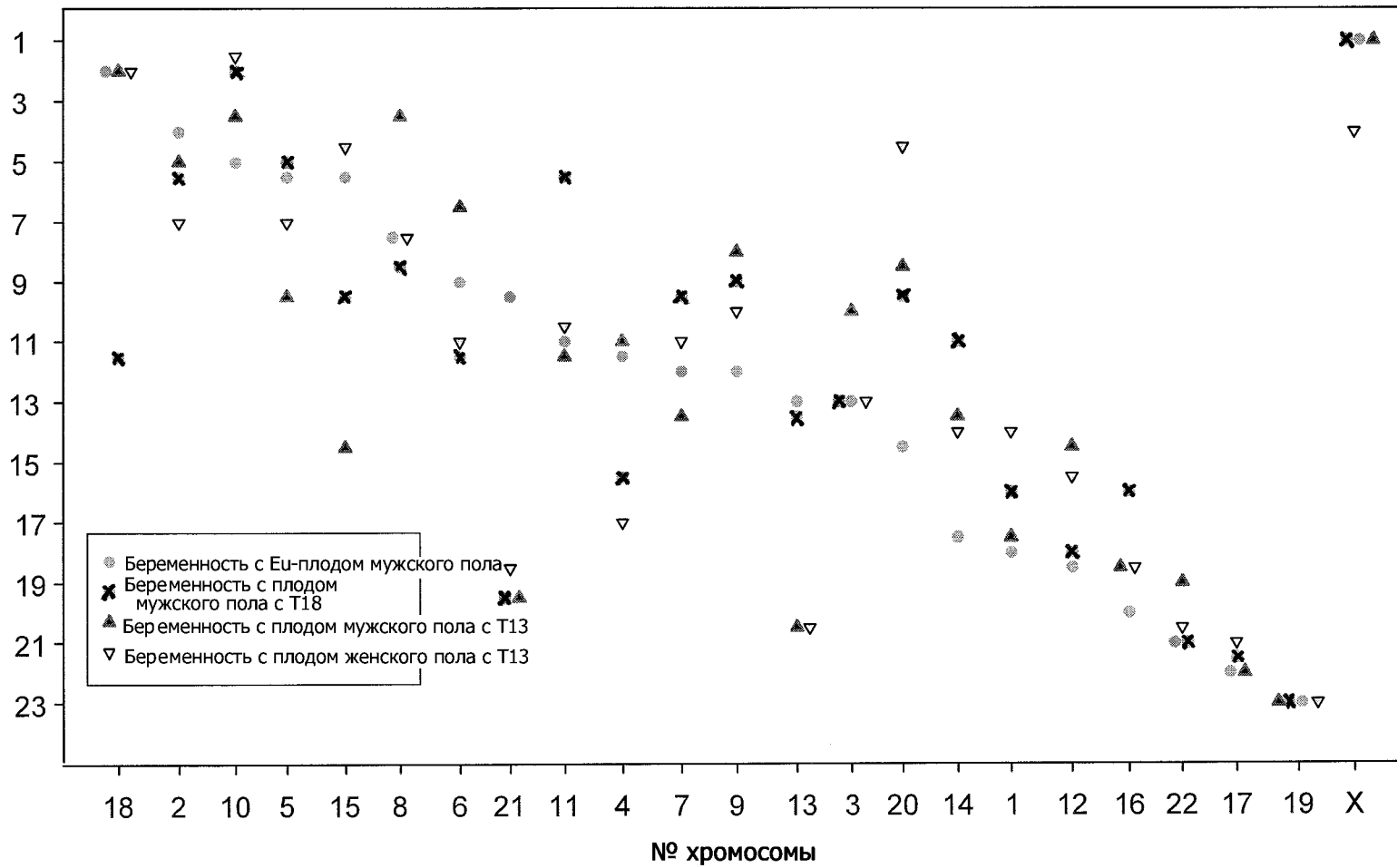
Фиг. 10

1000

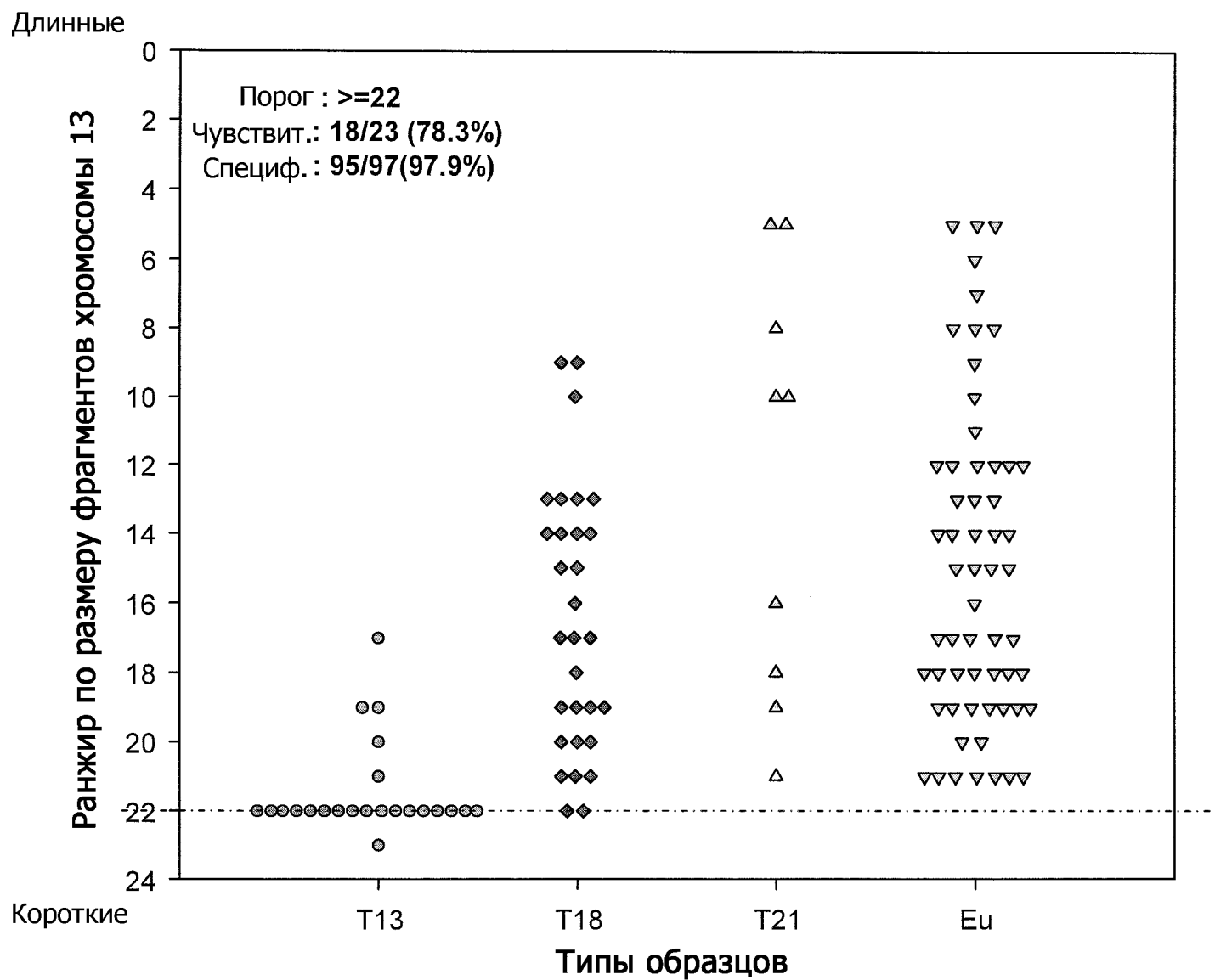
	Число случаев		
	Женский пол	Мужской пол	Общее
Трисомия 13 (Т13)	9	14	23
Трисомия 18 (Т18)	20	10	30
Трисомия 21 (Т21)	4	5	9
эуплоид	23	35	58
Общее	56	64	120

Фиг. 11

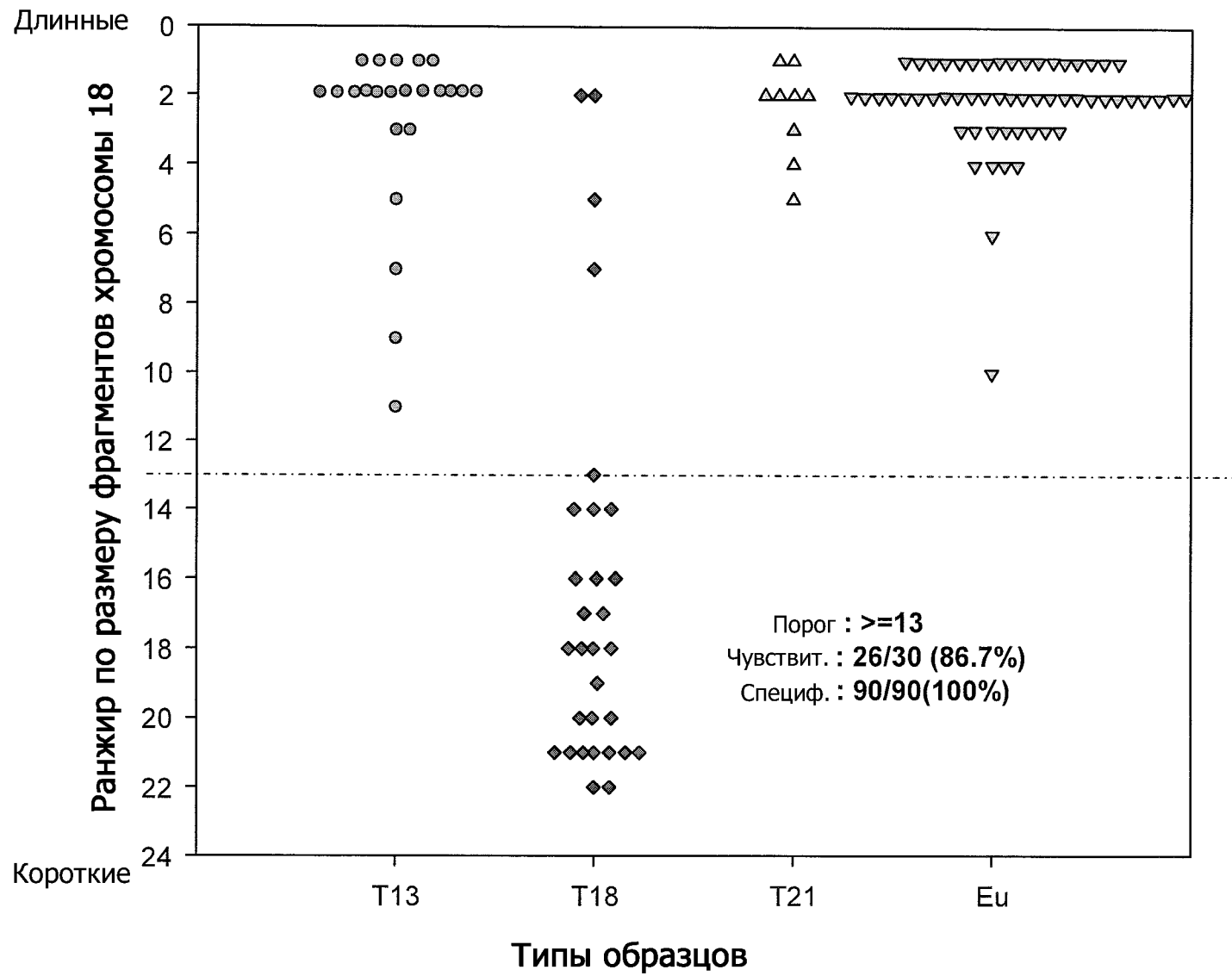
Медианный ранжир по размеру фрагментов ДНК  
(от самого короткого до самого длинного)



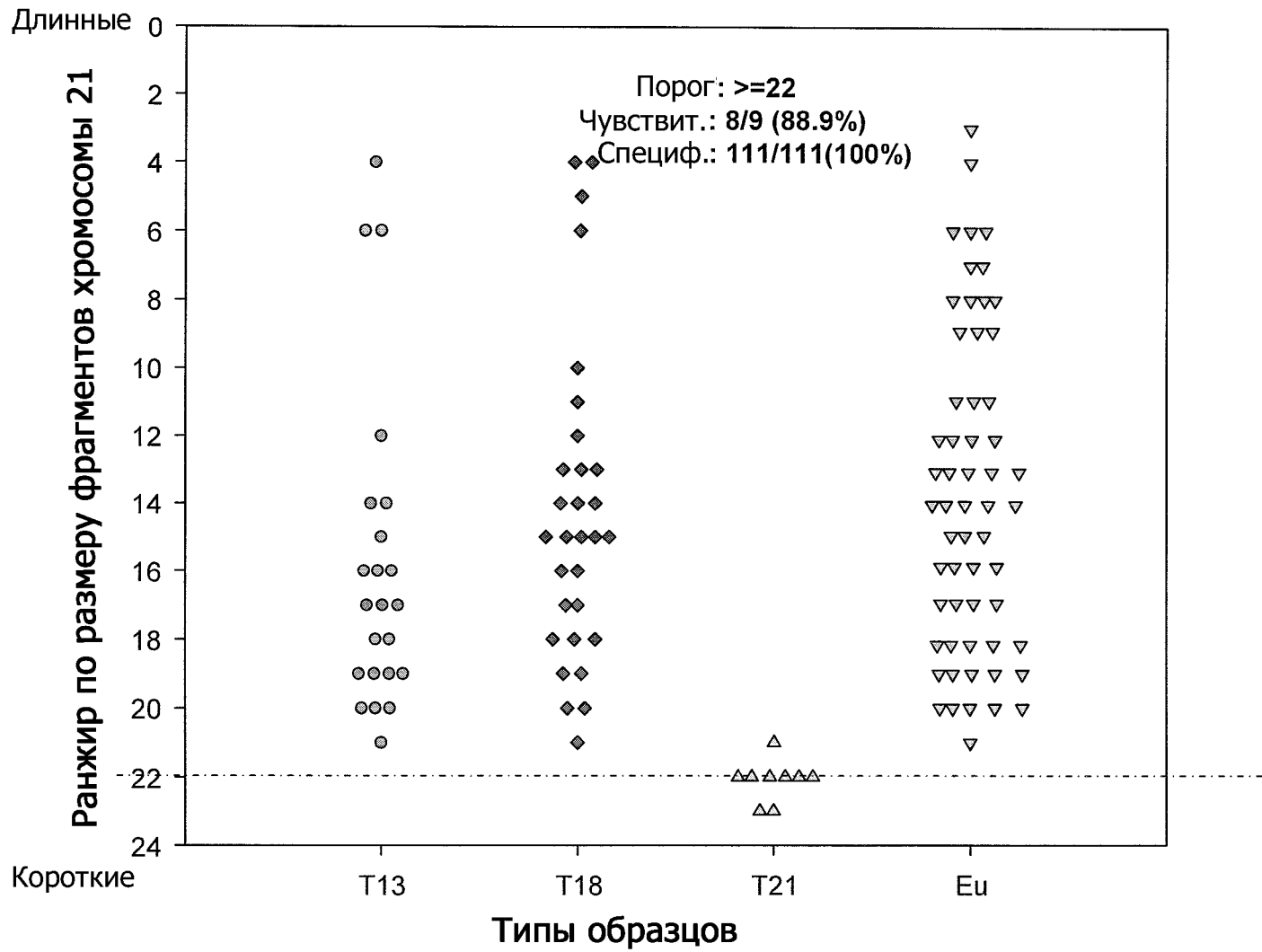
Фиг. 12



Фиг. 13



Фиг. 14



Фиг. 15



Название образца	Кариотип	Средний размер фрагментов хром. 13	Средний размер фрагментов хром. 5	Различие между средними хром. 5 и хром. 13	Величина* Р (хром. 5 против хром. 13)	Средний размер фрагментов хром. 6	Различие между средними хром. 6 и хром. 13	Величина* Р (хром. 6 против хром. 13)
87164 - Eu - мужской пол	Эуплоидный	154.969	155.198	0.229	0.159	155.037	0.068	0.732
87213 - Eu - мужской пол	Эуплоидный	154.965	155.294	0.329	0.09	155.339	0.374	0.256
94355 - T18 - мужской пол	Трисомия 18	163.908	164.11	0.202	0.035	164.143	0.235	0.276
96233 - T18 - мужской пол	Трисомия 18	163.751	164.074	0.323	0.356	163.623	-0.128	0.371
92394 - T13 - мужской пол	Трисомия 13	165.15	165.754	0.604	<0.001	165.901	0.751	<0.001
61175 - T13 - мужской пол	Трисомия 13	156.165	156.581	0.416	<0.001	156.696	0.531	<0.001
96342 - T13 - женский пол	Трисомия 13	157.064	158.127	1.063	<0.001	158.301	1.237	<0.001
PW093 - T13 - женский пол	Трисомия 13	159.038	159.846	0.808	<0.001	159.679	0.641	0.001

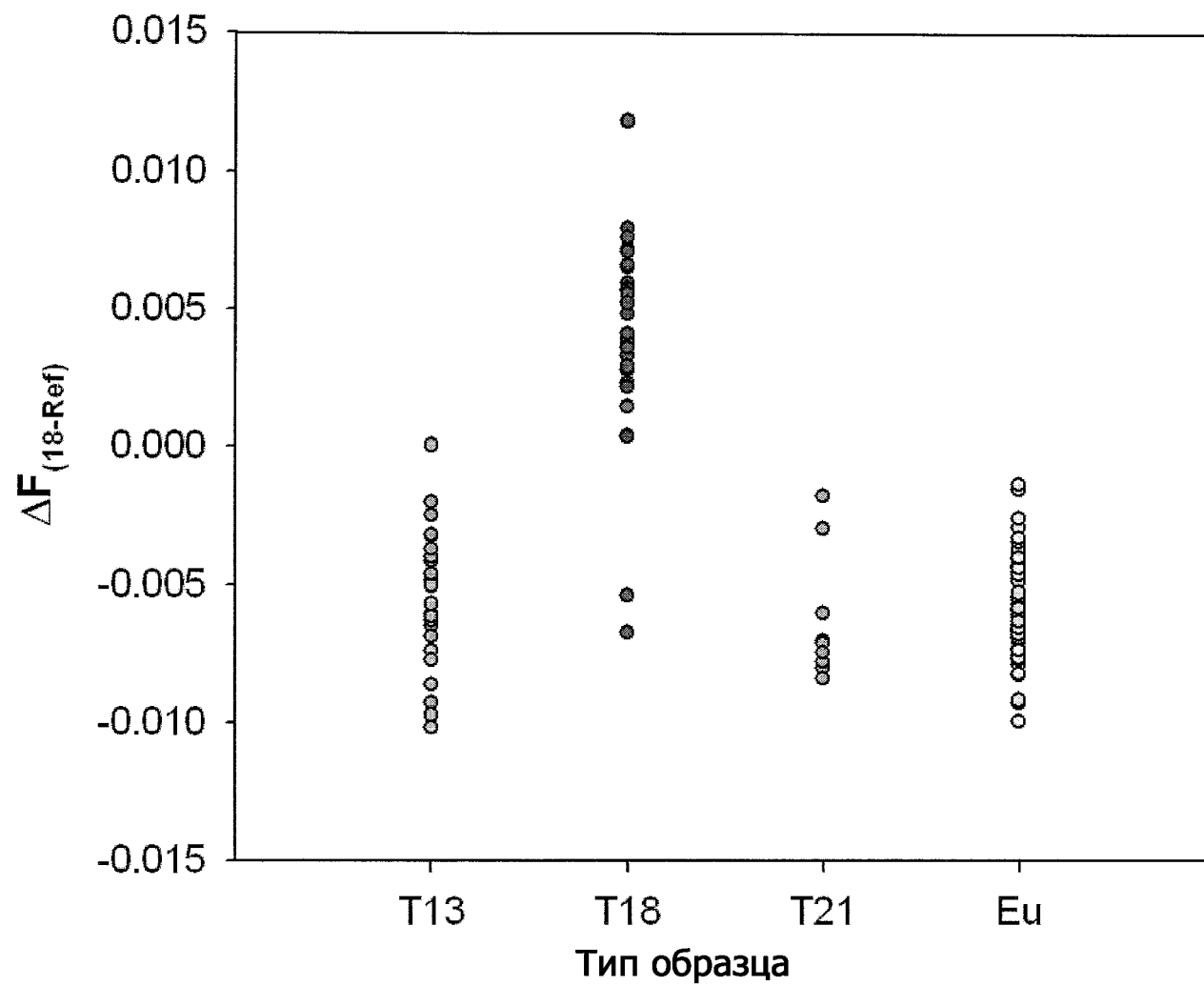
Фиг. 16

1600

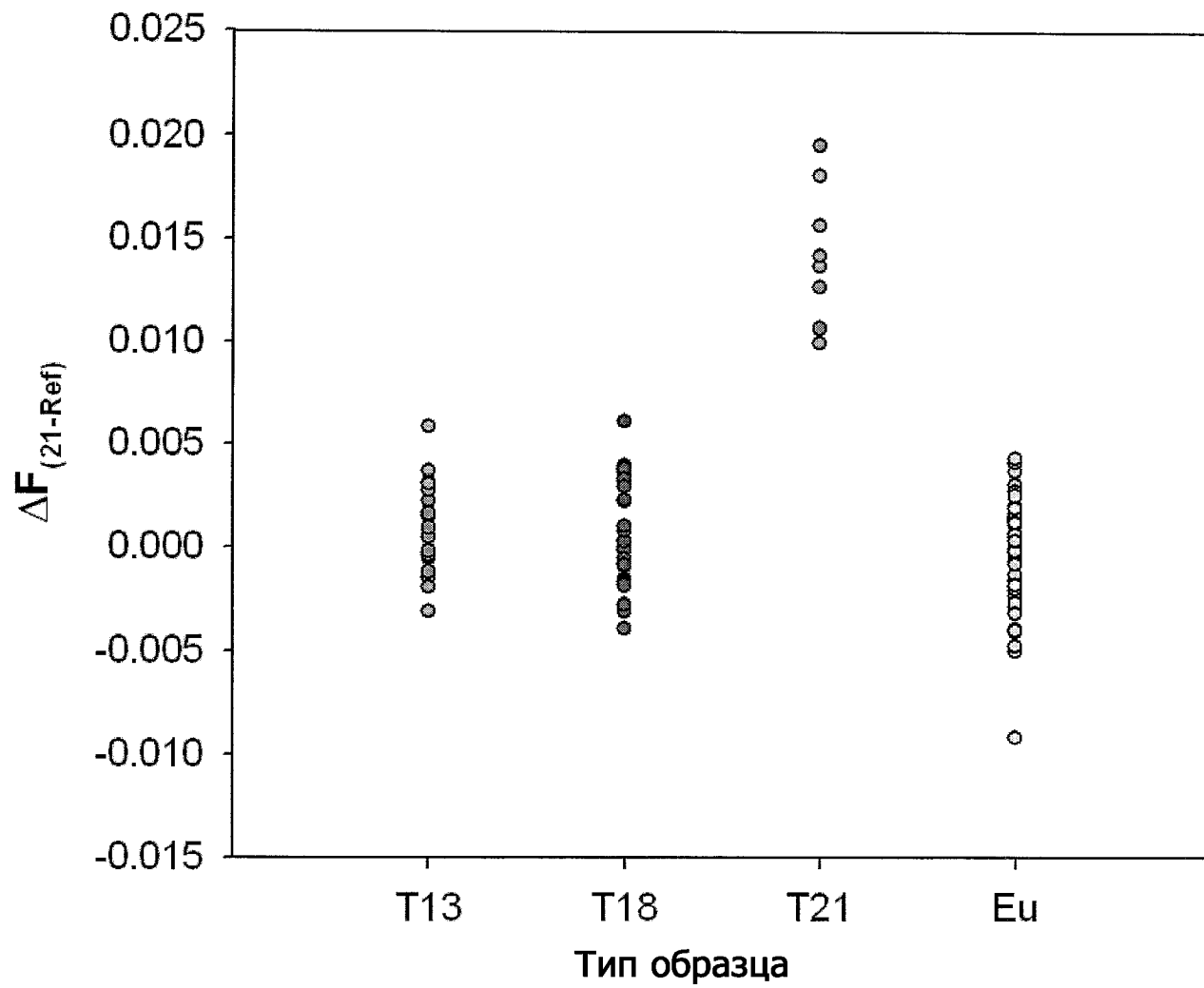
Название образца	Кариотип	Средний размер фрагментов хром. 18	Средний размер фрагментов хром. 14	Различие между средними хром. 18 и хром. 14	Величина* P (хром. 18 против хром. 14)
87164- Eи-мужской пол	Эуплоидный	155.595	154.614	0.981	<0.001
87213- Eи-мужской пол	Эуплоидный	155.977	155.142	0.835	<0.001
94355- T18-мужской пол	Трисомия 18	163.637	163.914	- 0.277	0.541
96233- T18-мужской пол	Трисомия 18	163.897	164.378	- 0.481	0.393
92394- T13-мужской пол	Трисомия 13	166.163	165.715	0.448	<0.001
61175- T13-мужской пол	Трисомия 13	157.186	156.606	0.58	<0.001
96342- T13-женский пол	Трисомия 13	158.326	158.013	0.313	0.003
PW093- T13-женский пол	Трисомия 13	160.469	159.503	0.966	<0.001

Фиг. 17

1700



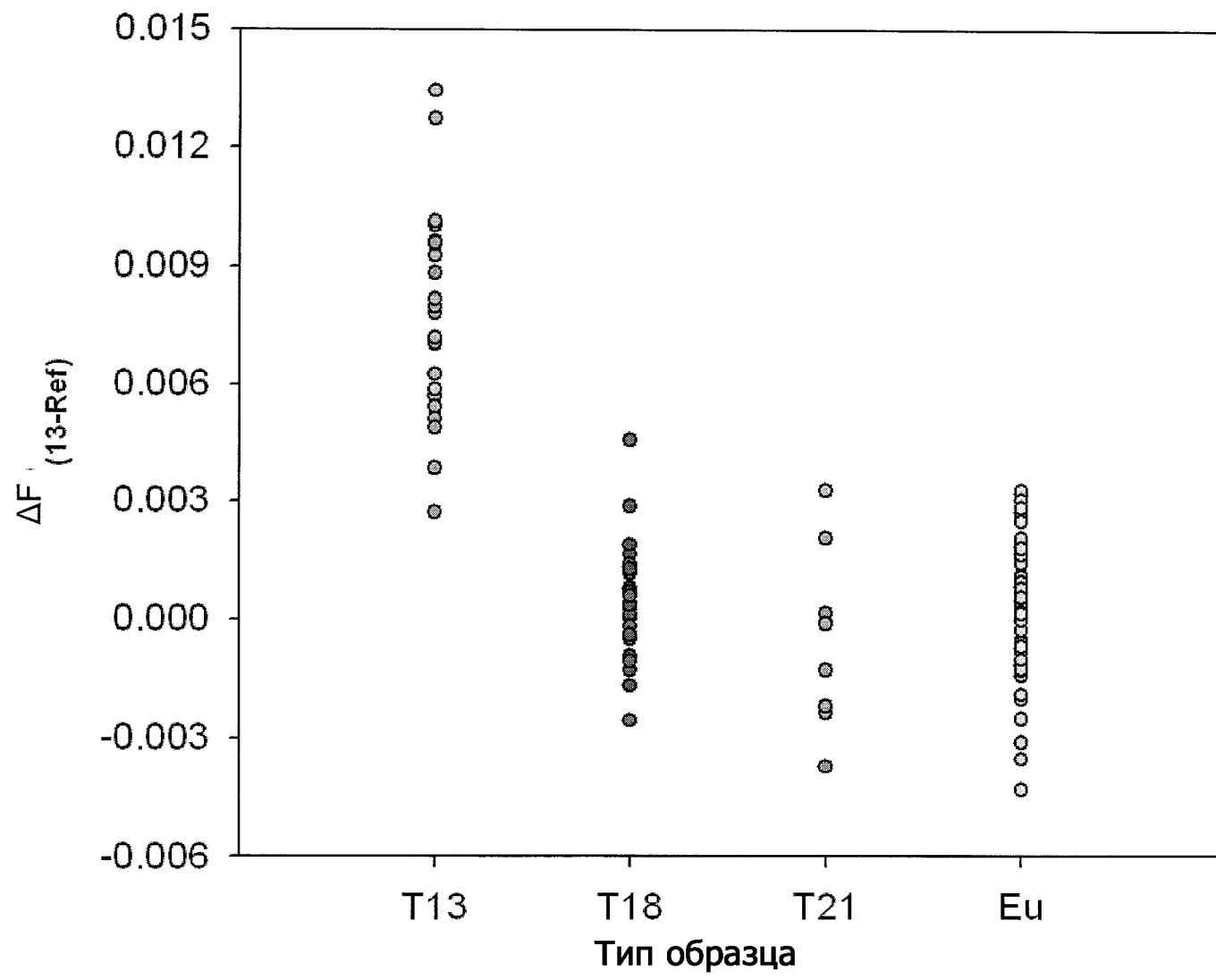
ФИГ. 18



ФИГ. 19

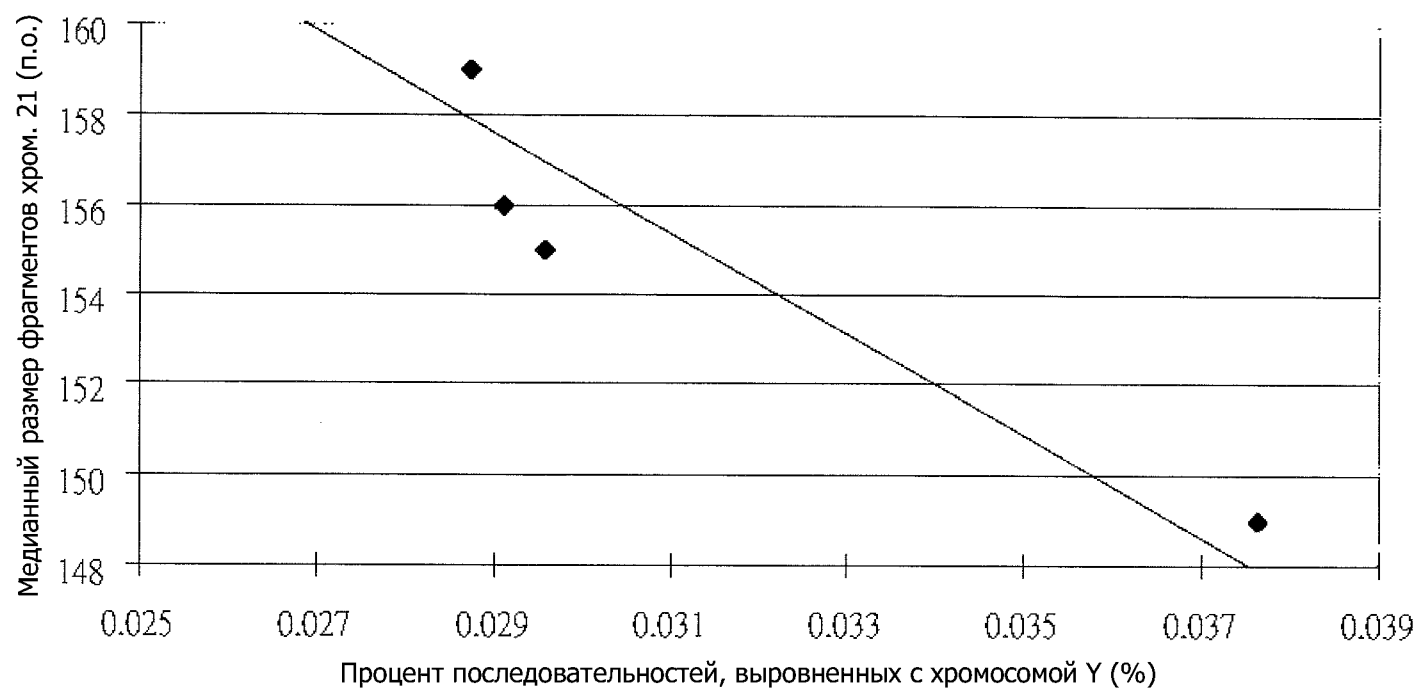
Хром.	Содержание GC (%)
4	36.53
13	37.10
6	38.60
5	38.61
3	38.79
X	38.96
18	39.07
8	39.45
2	39.67
12	39.80
7	39.94
21	40.00
14	40.33
9	41.03
10	41.41
1	41.52
11	41.89
15	42.35
20	44.65
16	45.22
17	46.05
22	49.89
19	50.65

Фиг. 20

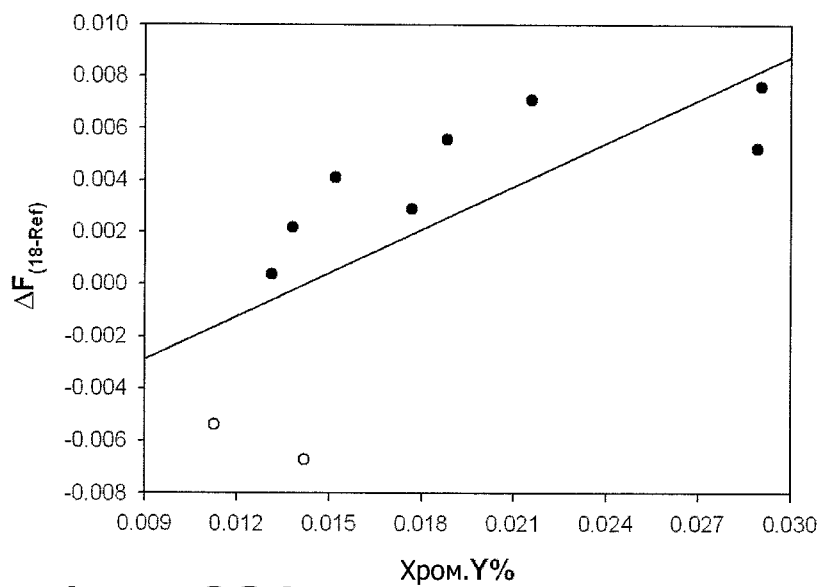


Фиг. 21

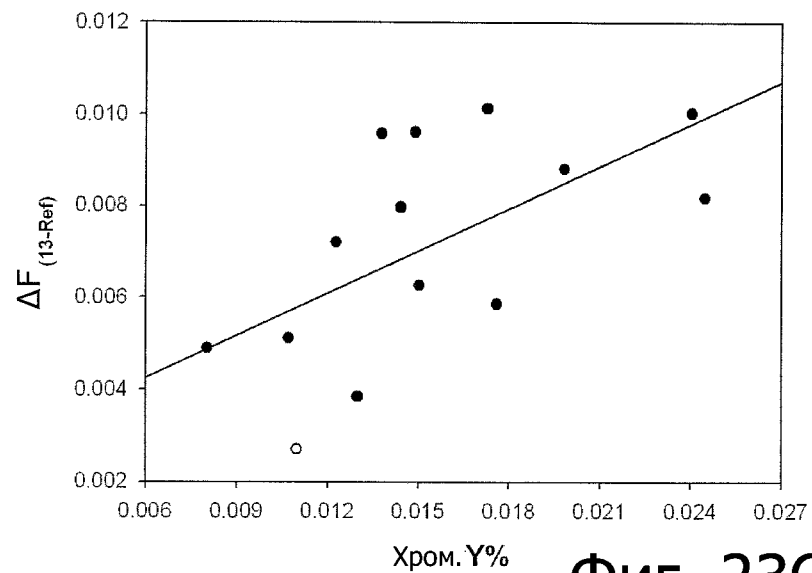
Корреляция между средним размером фрагментов  
хромосомы 21 и процентом последовательностей,  
выровненных с хромосомой Y



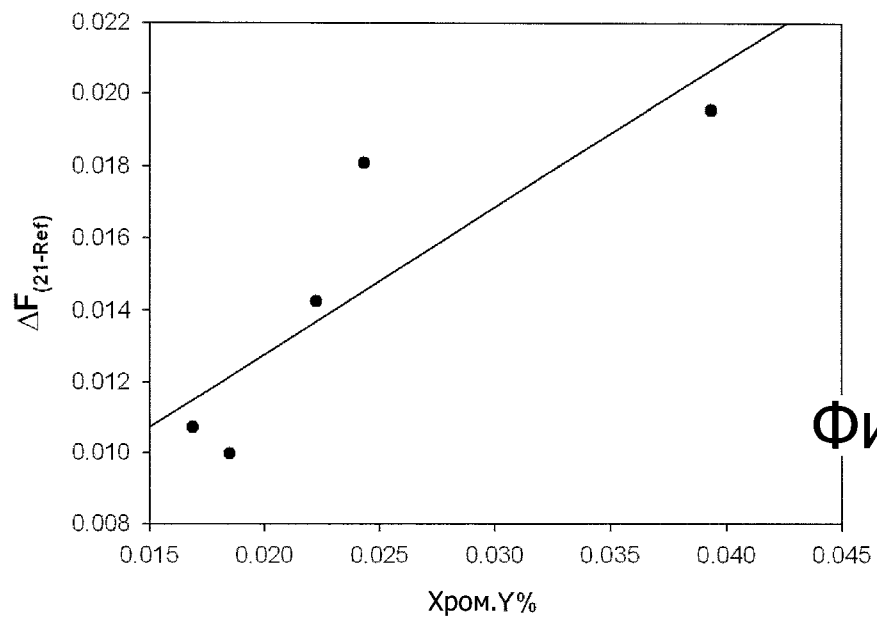
Фиг. 22



Фиг. 23А



Фиг. 23С

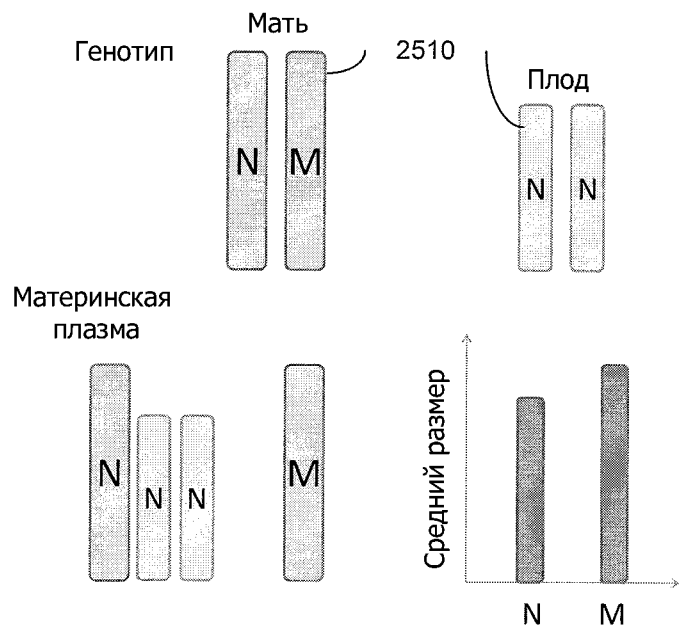


Фиг. 23В

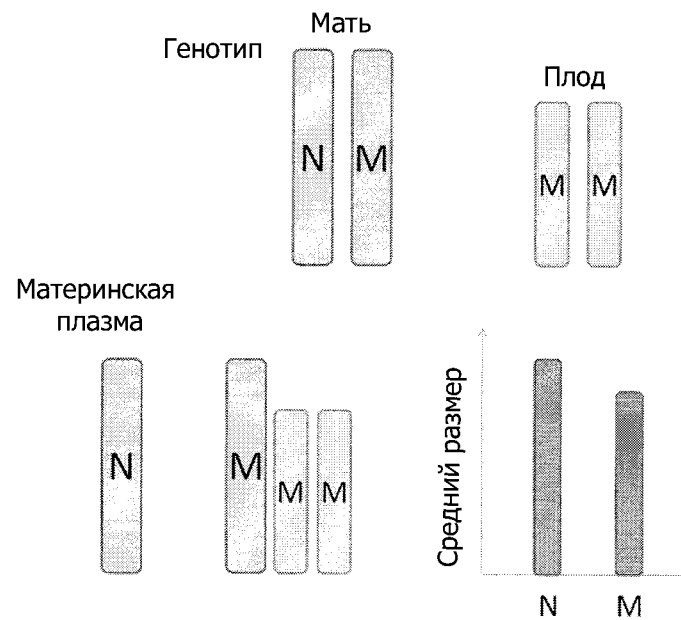


ID образца	Кариотип	Z-оценка для хром. 18	Z-оценка для хром. 13	Обнаруженное значимое укорачивание фрагментов ДНК хром. 18	Обнаруженное значимое укорачивание фрагментов ДНК хром. 13
1	Эуплоидный	1.07	0.84	не имеется	не имеется
2	Эуплоидный	-0.64	0.08	не имеется	не имеется
3	Трисомия 18	<u>3.62</u>	-1.43	имеется	не имеется
4	Трисомия 18	<u>6.07</u>	0.51	имеется	не имеется
5	Трисомия 13	0.61	<u>3.55</u>	не имеется	имеется
6	Трисомия 13	-1.04	2.06	не имеется	имеется
7	Трисомия 13	2.67	<u>5.72</u>	не имеется	имеется
8	Трисомия 13	-1.39	1.55	не имеется	имеется

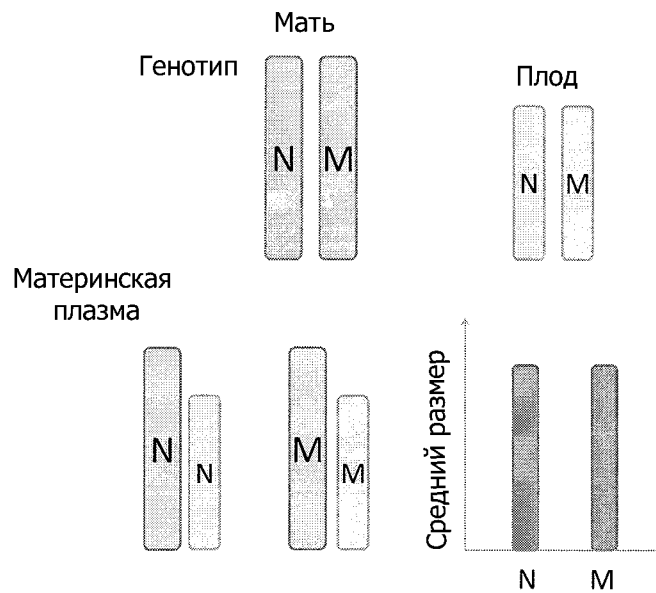
Фиг. 24



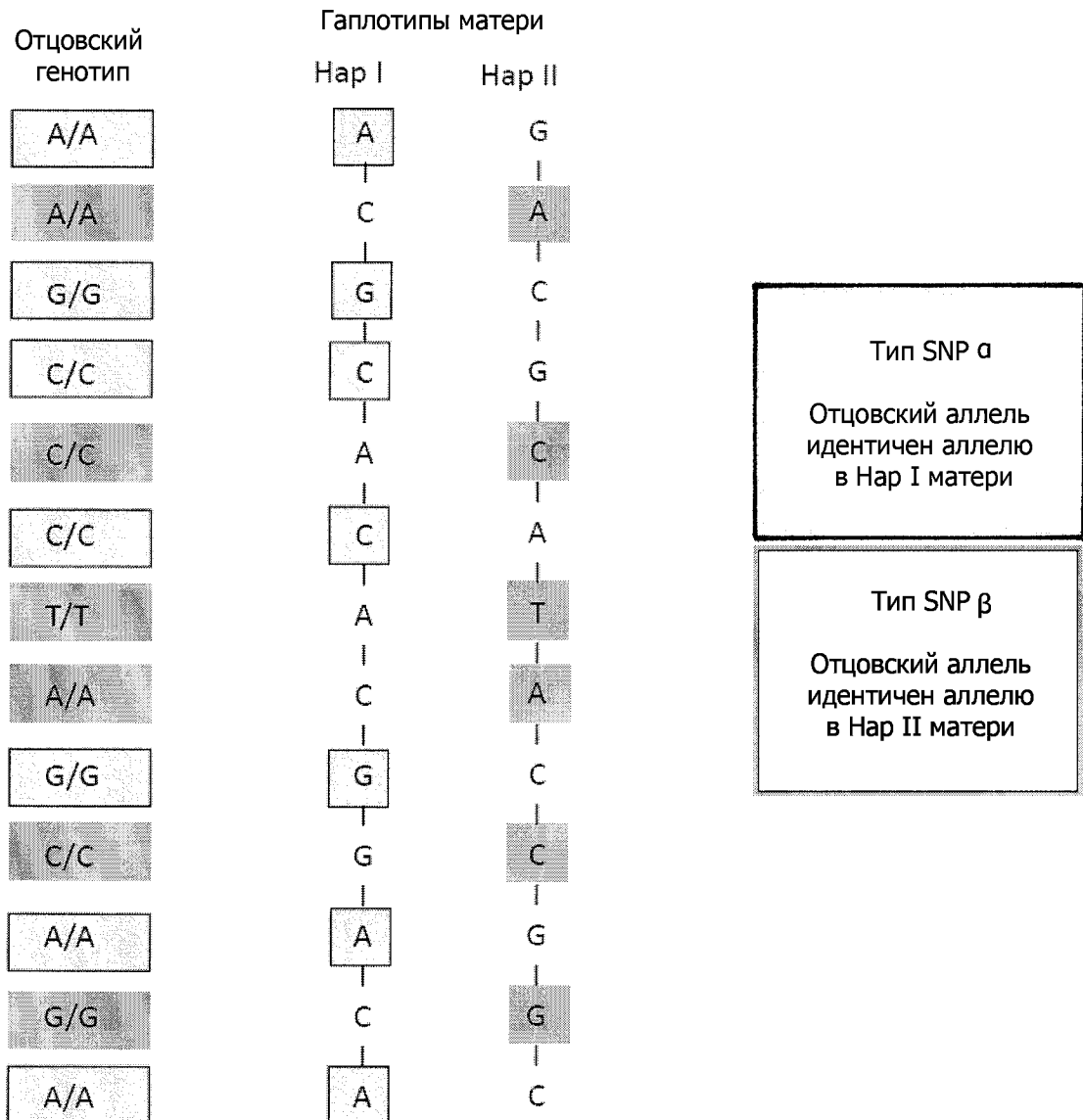
Фиг. 25А



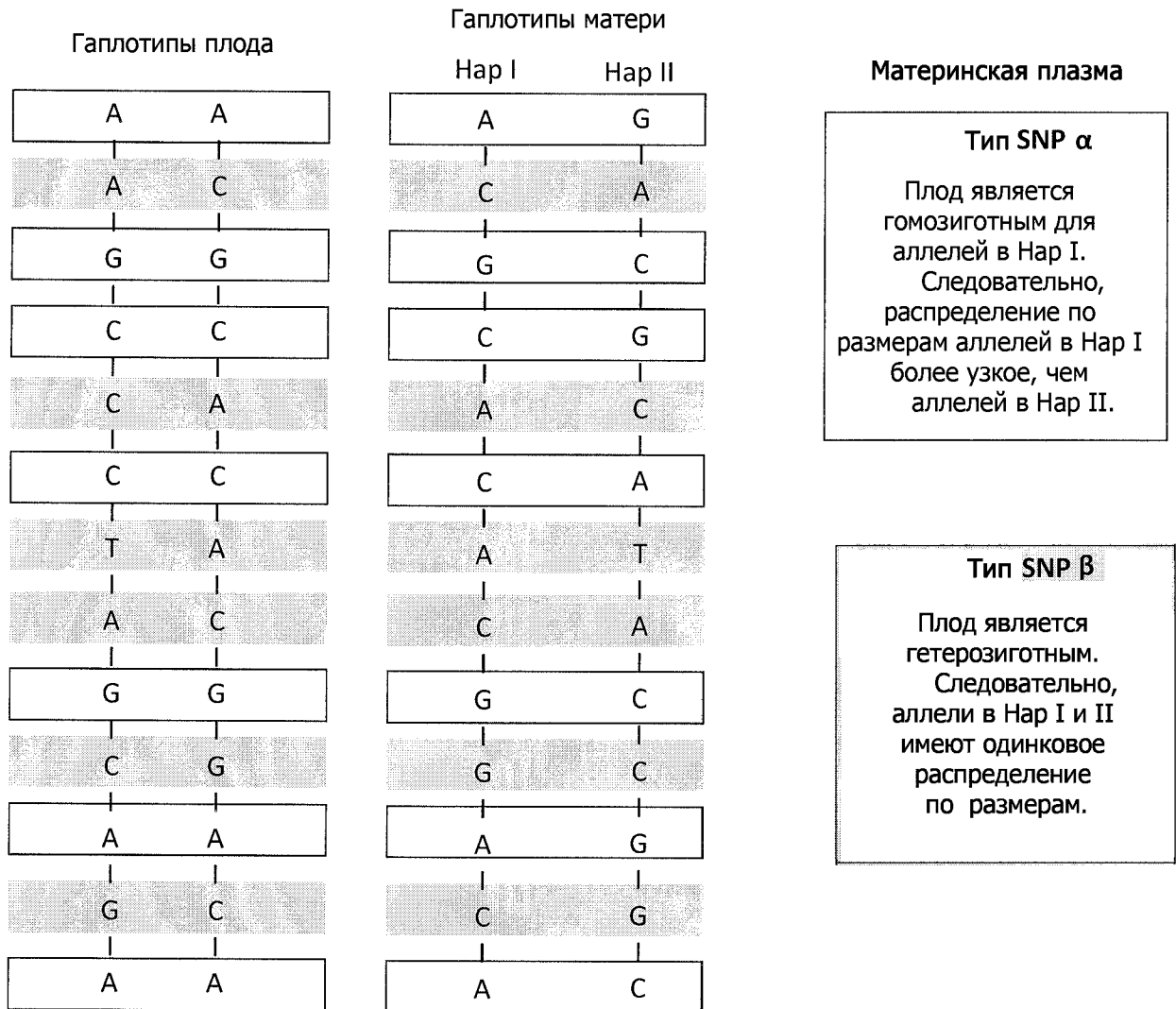
Фиг. 25С



Фиг. 25В



Фиг. 26



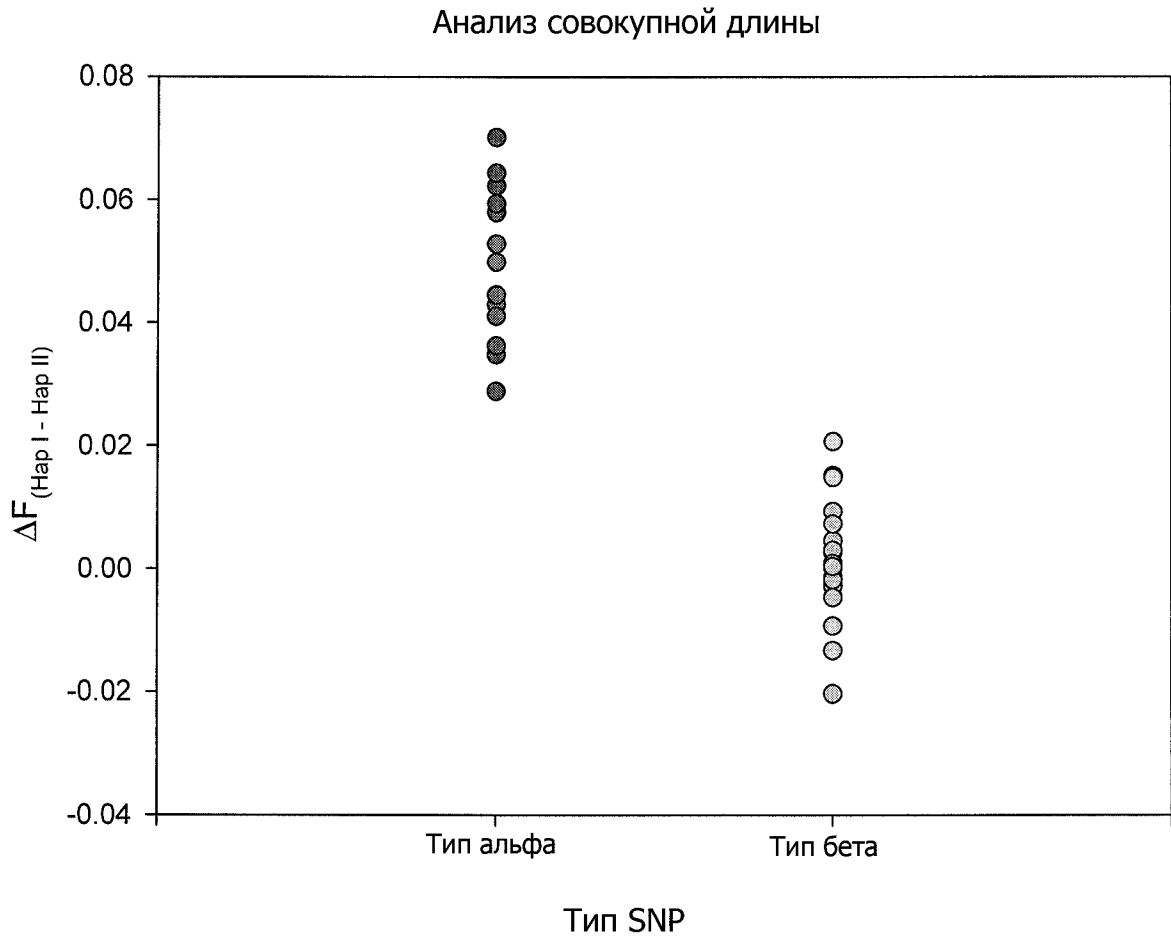
Фиг. 27

Тип SNP $\alpha$					
SNP в сегменте отсутствуют	Исходный SNP	Конечный SNP	Доля общей длины, вносимая короткими фрагментами		$\Delta F_{(\text{Hap I} - \text{Hap II})}$
			Hap I	Hap II	
50	rs2027649	rs5993883	0.2595	0.2096	0.0499
50	rs2239395	rs1002286	0.2648	0.2219	0.0429
50	rs4822458	rs3747134	0.2905	0.2204	0.0701
50	rs5761557	rs4410	0.2659	0.2037	0.0623
50	rs134784	rs4822998	0.2569	0.2221	0.0348
50	rs5762936	rs5998473	0.2882	0.2302	0.0580
50	rs5754086	rs8140669	0.2599	0.2311	0.0288
50	rs5999854	rs5756540	0.2659	0.2249	0.0411
50	rs229535	rs2413637	0.2953	0.2425	0.0528
50	rs4820431	rs11090087	0.2793	0.2199	0.0594
50	rs1023469	rs1972489	0.2884	0.2438	0.0445
50	rs5764858	rs5769218	0.2644	0.2001	0.0643
50	rs6009043	rs5769452	0.2481	0.2118	0.0363
28	rs17825762	rs131815	0.2648	0.2407	0.0241

Фиг. 28

Тип SNPβ					
SNP в сегменте отсутствуют	Исходный SNP	Конечный SNP	Доля общей длины, вносимая короткими фрагментами		$\Delta F_{(\text{Нап I} - \text{Нап II})}$
			Нап I	Нап II	
50	rs2159071	rs873387	0.2565	0.2414	0.0151
50	rs11917	rs78908	0.2548	0.2577	-0.0029
50	rs589089	rs7288450	0.2495	0.2589	-0.0093
50	rs5759868	rs2301492	0.2431	0.2477	-0.0046
50	rs2301497	rs713911	0.2411	0.2204	0.0207
50	rs4820740	rs5752850	0.2181	0.2195	-0.0014
50	rs5752851	rs5752964	0.2344	0.2372	-0.0028
50	rs1894473	rs135472	0.2603	0.2510	0.0093
50	rs135475	rs5754558	0.2324	0.2457	-0.0133
50	rs4821148	rs11705488	0.2381	0.2380	0.0001
50	rs390647	rs362246	0.2270	0.2224	0.0046
50	rs362214	rs6000130	0.2404	0.2401	0.0003
50	rs739206	rs12628179	0.2341	0.2193	0.0148
50	rs17298479	rs2143921	0.2597	0.2643	-0.0047
50	rs5758913	rs926542	0.2655	0.2582	0.0073
50	rs7291629	rs133761	0.2156	0.2360	-0.0203
50	rs133755	rs6007919	0.2359	0.2332	0.0027
50	rs135570	rs136646	0.2496	0.2466	0.0030
50	rs80454	rs8136986	0.2354	0.2372	-0.0018
50	rs5768117	rs133662	0.2339	0.2330	0.0009
49	rs6007851	rs739365	0.2376	0.2372	0.0004

Фиг. 29



Фиг. 30

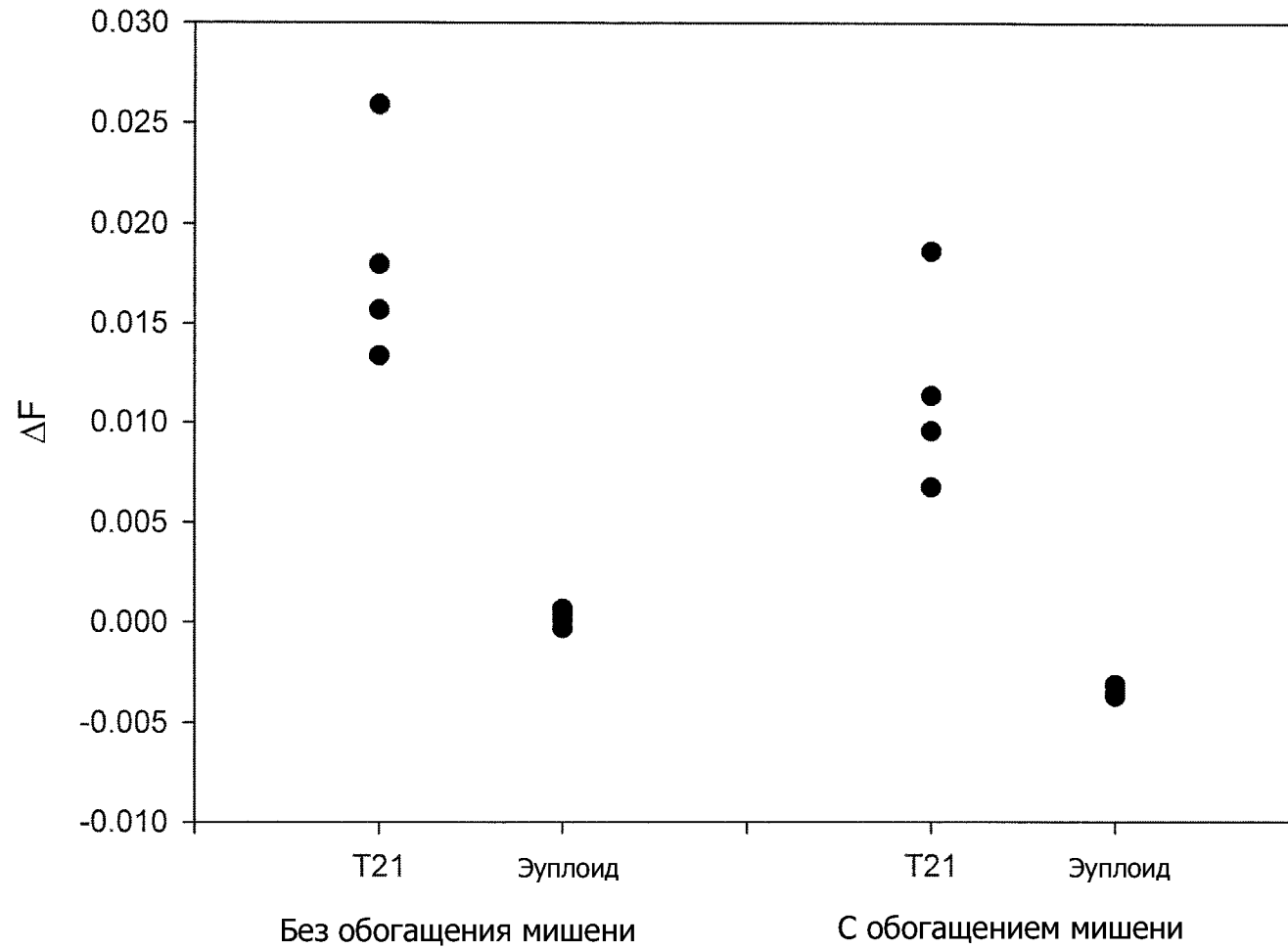
Код образца	Обогащение мишени	Статус плода	Общая длина, вносимая фрагментами ДНК $\leq 150$ п.о., для хромосомы 21 (миллионы п.о.)	Общая длина, вносимая фрагментами ДНК $\leq 600$ п.о., для хромосомы 21 (миллионы п.о.)	$F_{21}$	Общая длина, вносимая фрагментами ДНК $\leq 150$ п.о., для эталонных хромосом (миллионы п.о.)	Общая длина, вносимая фрагментами ДНК $\leq 600$ п.о., для эталонных хромосом (миллионы п.о.)	$F_{ref}$	$\Delta F$
uk 99229	нет	T21	10.2	36.0	0.283	653.1	2422.2	0.270	0.013
pw 421	нет	T21	10.6	35.6	0.299	651.2	2315.5	0.281	0.018
uk 99510	нет	T21	9.8	32.0	0.307	607.1	2082.2	0.292	0.016
uk 99807	нет	T21	12.6	30.1	0.417	745.2	1905.8	0.391	0.026
pw 226	нет	Эуплоид	10.4	33.7	0.310	734.7	2368.4	0.310	0.000
pw 316	нет	Эуплоид	10.9	38.4	0.284	765.2	2691.8	0.284	0.000
pw 263	нет	Эуплоид	7.7	25.9	0.297	537.5	1810.8	0.297	0.001
pw 370	нет	Эуплоид	5.9	23.9	0.247	411.8	1671.0	0.246	0.000

Фиг. 31А

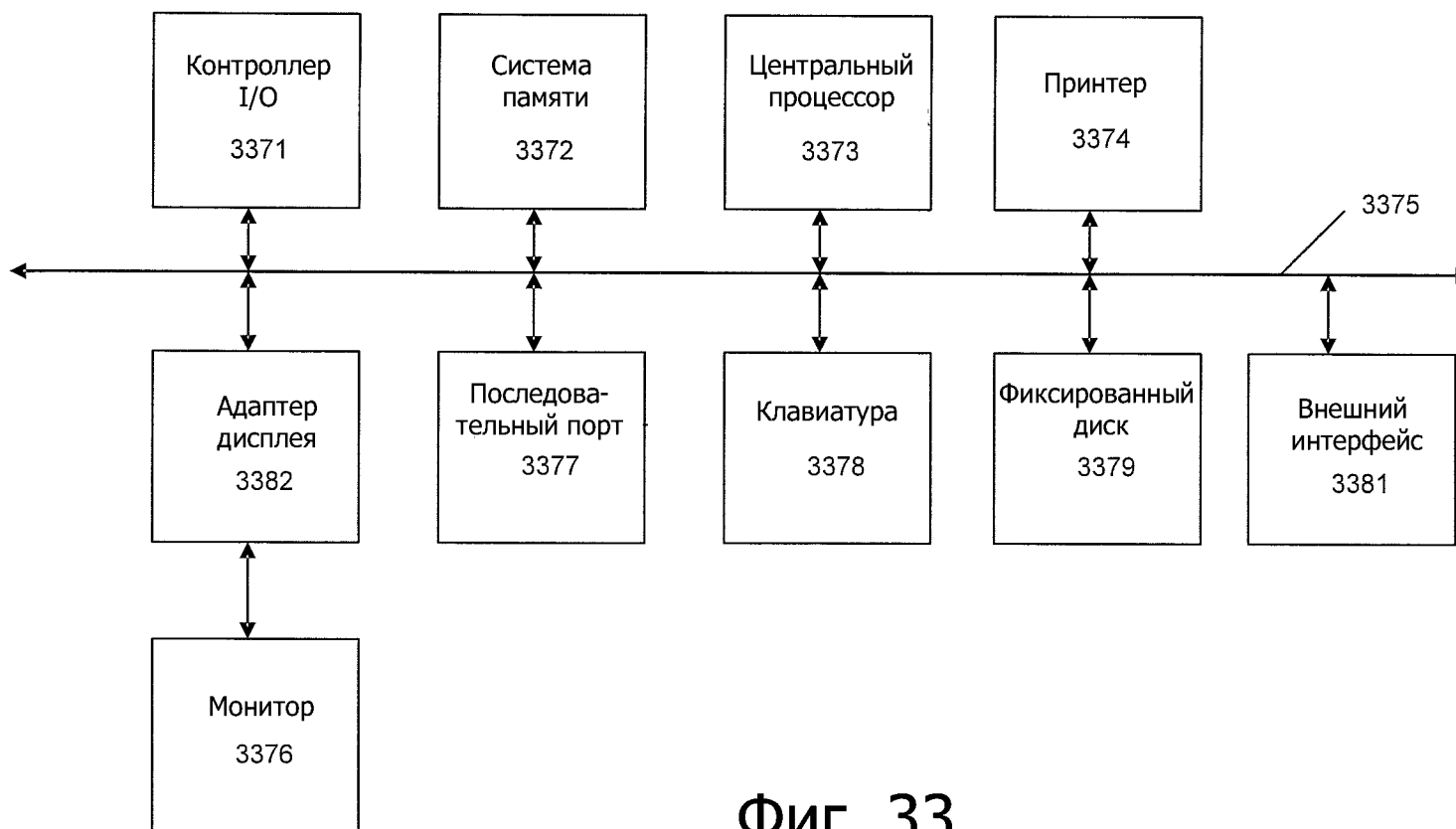
Код образца	Обогащение мишени	Статус плода	Общая длина, вносимая фрагментами ДНК $\leq 150$ п.о., для хромосомы 21 (миллионы п.о.)	Общая длина, вносимая фрагментами ДНК $\leq 600$ п.о., для хромосомы 21 (миллионы п.о.)	$F_{21}$	Общая длина, вносимая фрагментами ДНК $\leq 150$ п.о., для эталонных хромосом (миллионы п.о.)	Общая длина, вносимая фрагментами ДНК $\leq 600$ п.о., для эталонных хромосом (миллионы п.о.)	$F_{ref}$	$\Delta F$
uk 99229	да	T21	5.7	24.9	0.231	434.1	1938.3	0.224	0.007
pw 421	да	T21	5.1	22.4	0.226	362.4	1685.0	0.215	0.011
uk 99510	да	T21	4.4	19.0	0.229	314.4	1435.1	0.219	0.010
uk 99807	да	T21	6.6	21.3	0.312	455.1	1553.2	0.293	0.019
pw 226	да	Эуплоид	5.1	21.4	0.237	413.2	1720.2	0.240	-0.004
pw 316	да	Эуплоид	4.9	22.6	0.215	394.9	1811.5	0.218	-0.003
pw 263	да	Эуплоид	4.1	19.3	0.211	327.0	1526.6	0.214	-0.004
pw 370	да	Эуплоид	3.7	20.4	0.180	296.9	1621.0	0.183	-0.003

Фиг. 31В





Фиг. 32



Фиг. 33