

(19)



**Евразийское
патентное
ведомство**

(11) **039167**

(13) **B1**

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ

(45) Дата публикации и выдачи патента
2021.12.13

(51) Int. Cl. **C12Q 1/68** (2018.01)
G06F 19/00 (2018.01)

(21) Номер заявки
201791612

(22) Дата подачи заявки
2008.07.23

**(54) ДИАГНОСТИКА ФЕТАЛЬНОЙ ХРОМОСОМНОЙ АНЭУПЛОИДИИ С
ИСПОЛЬЗОВАНИЕМ ГЕНОМНОГО СЕКВЕНИРОВАНИЯ**

(31) 60/951,438

(32) 2007.07.23

(33) US

(43) 2017.11.30

(62) 201300072; 2008.07.23

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

(72) Изобретатель:
**Ло Йук-Минг Деннис, Чну Росса Вай
Квун, Чан Кван Чее (НК)**

(74) Представитель:
Агуреев А.П., Фелицына С.Б. (RU)

(56) DENNIS LO Y.M. et al. Increased Fetal DNA Concentrations in the Plasma of Pregnant Women Carrying Fetuses with Trisomy 21. *Clinical Chemistry*, 1999, 45:10, pp. 1747-1751

RONALD WAPNER M.D. et al. First-Trimester Screening for Trisomies 21 and 18. *The New England Journal of Medicine*, 2003, Vol. 349, No. 15, с 1405-1413

US-B1-6440706

US-A1-20060252071

TONG Yu K. et al. Noninvasive Prenatal Detection of Fetal Trisomy 18 by Epigenetic Allelic Ratio Analysis in Maternal Plasma: Theoretical and Empirical Considerations. *Clinical Chemistry*, 2006, 52:12, pp.2194-2202

(57) Воплощения изобретения относятся к способам, системам и аппаратуре для определения того, имеется ли фетальная хромосомная анэуплоидия из биологического образца, взятого у беременной женщины. Молекулы нуклеиновых кислот биологического образца секвенируют, так что секвенируется часть генома. Соответствующие количества клинически релевантной хромосомы и фоновых хромосом определяют из результатов секвенирования. Параметр, полученный из таких количеств (например, соотношение), сравнивают с одним или несколькими пороговыми значениями, причем таким образом определяют классификацию того, имеется ли фетальная хромосомная анэуплоидия.

039167 B1

039167 B1

Притязание на приоритет

Настоящая заявка притязает на приоритет и является непредварительной заявкой предварительной заявки на патент США № 60/951438, озаглавленной "Определение дисбаланса нуклеотидной последовательности", зарегистрированной 23 июля 2007 (досье поверенного № 016285-005200US), включенной в данное описание в качестве ссылки.

Перекрестные ссылки на родственные заявки

Настоящая заявка также является родственной одновременно находящейся на регистрации непредварительной заявке, озаглавленной "Определение дисбаланса нуклеотидной последовательности" (досье поверенного № 016285-005210US), включенной в данное описание в качестве ссылки.

Область техники, к которой относится изобретение

Изобретение вообще относится к диагностическому исследованию фетальной хромосомной анеуплоидии посредством определения дисбаланса между различными нуклеотидными последовательностями, и конкретнее, к идентификации трисомии 21 (синдром Дауна) и других хромосомных анеуплоидий через исследование материнского образца (например, крови).

Уровень техники

Фетальная хромосомная анеуплоидия является результатом присутствия аномальной(ых) дозы(доз) хромосомы или хромосомного участка. Аномальная(ые) доза(ы) может(могут) быть аномально высокой(ими), например присутствие лишней хромосомы 21 или хромосомного участка при трисомии 21, или аномально низкой(ими), например отсутствие хромосомы X при синдроме Тернера.

Обычные пренатальные диагностические методы фетальной хромосомной анеуплоидии, например трисомии 21, включают взятие образцов материалов плода инвазивными процедурами, такими как амниоцентез, или взятие образцов хориальной ворсины, которые ограничиваются опасностью потери плода. Неинвазивные процедуры, такие как скрининг ультразвуковой эхографией и биохимические маркеры, используют для устранения опасности для беременной женщины от характерных инвазивных диагностических процедур, однако такие методы скрининга вместо коровой хромосомной аномальности обычно определяют вторичные патологические явления, которые ассоциируются с хромосомной анеуплоидией, например трисомией 21, и, таким образом, имеют ограниченную диагностическую точность и другие недостатки, такие как слишком высокое влияние гестационного возраста.

Открытие циркулирующей внеклеточной фетальной ДНК в материнской плазме в 1977 предоставило новые возможности для неинвазивной пренатальной диагностики (Lo Y.M.D. and Chiu R.W.K., 2007, *Nat. Rev. Genet.*, 8, 71-77). Хотя данный метод легко применим для пренатальной диагностики связанных с полом (Costa J.M. et al., *N. Engl. J. Med.*, 346, 1502) и некоторых отдельных генных расстройств (Lo Y.M.D. et al., *N. Engl. J. Med.*, 339, 1734-1738), его применение для пренатальной диагностики хромосомных анеуплоидий представляется весьма сомнительным (Lo Y.M.D. and Chiu R.W.K., 2007, *цит. выше*). Во-первых, фетальные нуклеиновые кислоты сосуществуют в материнской плазме с высоким фоном нуклеиновых кислот материнского происхождения, которые часто могут препятствовать анализу фетальных нуклеиновых кислот (Lo Y.M.D. et al., 1998, *Am. J. Hum. Genet.*, 62, 768-775). Во-вторых, фетальные нуклеиновые кислоты циркулируют в материнской плазме преимущественно во внеклеточной форме, что затрудняет получение информации о количестве генов или хромосом в геноме плода.

Недавно осуществлены существенные разработки, преодолевающие указанные сомнения (Benachi A. and Costa J.M., 2007, *Lancet*, 369, 440-442). В одном подходе детектируют фетальноспецифические нуклеиновые кислоты в материнской плазме, причем таким образом преодолевается проблема влияния материнского фона (Lo Y.M.D. and Chiu, RWK, 2007, *цит. выше*). Дозу хромосомы 21 выводят из соотношений полиморфных аллелей в молекулах ДНК/РНК, полученных из плаценты. Однако такой метод менее точен, когда образцы содержат небольшое количество целевой нуклеиновой кислоты, и может применяться только к плодам, гетерозиготным для целевых полиморфизмов, которые являются только подмножеством популяции, если используется один полиморфизм.

Dhallan et al. (Dhallan R. et al., 2007, *цит. выше*, Dhallan R. et al., 2007, *Lancet*, 369, 474-481) описали альтернативную стратегию обогащения доли циркулирующей фетальной ДНК посредством добавления формальдегида к материнской плазме. Пропорцию последовательностей хромосомы 21 в материнской плазме, вносимую плодом, определяют посредством оценки отношения наследуемых от отца фетальноспецифических аллелей к фетально неспецифическим аллелям для однонуклеотидных полиморфизмов (SNP) на хромосоме 21. Подобным образом отношения SNP вычисляют для эталонной хромосомы. Затем выводят дисбаланс фетальной хромосомы 21 посредством обнаружения статистически значимого различия между отношениями SNP для хромосомы 21 и эталонной хромосомы, где значимость определяют с использованием заданного значения $p \leq 0,05$. Для того чтобы добиться высокого охвата популяции, намечают более 500 SNP на хромосому. Однако имеются расхождения, касающиеся эффективности формальдегида для обогащения фетальной ДНК до высокой доли (Chung G.T.Y. et al., 2005, *Clin. Chem.*, 51, 655-658), и поэтому воспроизводимость метода нуждается в дополнительной оценке. Также, так как каждый плод и мать будут давать информацию о различном числе SNP для каждой хромосомы, степень статистического критерия для сравнения отношений SNP может изменяться от случая к случаю (Lo Y.M.D. and Chiu R.W.K., 2007, *Lancet*, 369, 1997). Более того, так как такие подходы зависят от детекции генети-

ческих полиморфизмов, они ограничены плодами, гетерозиготными для указанных полиморфизмов.

С использованием полимеразной цепной реакции (ПЦР, PCR) и количественного определения ДНК локуса хромосомы 21 и эталонного локуса в культуре аминокитов, полученной из плодов с трисомией 21 и эуплоидных плодов, Zimmermann et al. (2002, Clin. Chem., 48, 362-363) смогли различить две группы плодов на основании 1,5-кратного увеличения в последовательностях ДНК хромосомы 21 в первых. Так как 2-кратное различие в концентрации матрицы ДНК составляет различие только в один пороговый цикл (Ct), установление 1,5-кратного различия является пределом обычной ПЦР в реальном времени. Для того чтобы добиться более точной степени количественного определения, необходимы другие стратегии.

Разработана численная (digital) ПНП для детекции отклонения аллельного соотношения в образцах нуклеиновых кислот (Chang H.W. et al., 2002, J. Natl. Cancer Inst, 94, 1697-1703). Численная ПЦР представляет собой амплификацию на основе метода анализа нуклеиновых кислот, при котором требуется распределение образца, содержащего нуклеиновые кислоты, на множество отдельных образцов, где каждый образец содержит в среднем не более чем примерно одну последовательность-мишень на образец. Специфические нуклеиновые кислоты-мишени амплифицируются с последовательностью специфическими праймерами для образования специфических ампликонов численной ПЦР. Локусы нуклеиновых кислот, являющихся мишенями, и виды или панель последовательностей специфических праймеров, включенных в реакции, определяют или отбирают перед анализом нуклеиновых кислот.

Клинически показана применимость детекции утраты гетерозиготности (LOH) в образцах опухолевых ДНК (Zhou W. et al., 2002, Lancet, 359, 219-225). Для результатов анализов методом численной ПЦР последовательный критерий отношения вероятностей (SPRT) адаптирован предварительными исследованиями для классификации экспериментальных результатов как приводящих к выводу о наличии или отсутствии LOH в образце (El Karoui et al., 2006, Stat. Med., 25, 3124-3133).

В методах, используемых в предыдущих исследованиях, количество данных, собранных при численной ПЦР, слишком мало. Таким образом, точность может быть поставлена под сомнение из-за небольшого числа полученных точек и типичных статистических флуктуаций.

Поэтому желательно, чтобы неинвазивные тесты имели высокую чувствительность и специфичность для минимизации ложно-отрицательных и ложно-положительных результатов соответственно. Однако фетальная ДНК в материнской плазме и сыворотке присутствует в низкой абсолютной концентрации и представляет незначительную часть всех последовательностей ДНК. Поэтому также желательно иметь методы, которые допускают неинвазивную детекцию фетальной хромосомной анеуплоидии посредством максимизации количества генетической информации, которую можно получить из ограниченного количества фетальных нуклеиновых кислот, которые присутствуют как минимальная популяция в биологическом образце, содержащем материнские фоновые нуклеиновые кислоты.

Сущность изобретения

Воплощения данного изобретения относятся к способам, системам и аппаратуре для определения того, существует ли дисбаланс нуклеотидных последовательностей (например, хромосомный дисбаланс) в биологическом образце, взятом у беременной женщины. Такое определение можно осуществить с использованием параметра количества клинически релевантного хромосомного участка по отношению к другим клинически нерелевантным хромосомным участкам (фоновые участки) в биологическом образце. В одном аспекте количество хромосом определяют из секвенирования молекул нуклеиновых кислот в материнском биологическом образце, таком как моча, плазма, сыворотка, и других подходящих биологических образцах. Молекулы нуклеиновых кислот биологического образца секвенируют, так что секвенируют часть генома. Выбирают одно или несколько пороговых значений для определения того, имеется ли изменение при сравнении с эталонным количеством (т.е. дисбаланс), например, в отношении соотношения количеств двух хромосомных участков (или наборов участков).

Согласно примеру одного воплощения биологический образец, взятый у беременной женщины, анализируют для осуществления пренатальной диагностики фетальной хромосомной анеуплоидии. Биологический образец включает молекулы нуклеиновых кислот. Часть молекул нуклеиновых кислот, содержащихся в биологическом образце, секвенируют. В одном аспекте количество полученной генетической информации является достаточным для точного диагноза, не будучи пока чрезмерно избыточным, так как ограничивается ценами и количеством исходного требуемого биологического образца.

На основании секвенирования определяют первое количество первой хромосомы из последовательностей, идентифицированных как происходящие от первой хромосомы. Второе количество одной или нескольких вторых хромосом определяют из последовательностей, идентифицированных как происходящие от одной из вторых хромосом. Затем параметр от первого количества и второго количества сравнивают с одним или несколькими пороговыми значениями. На основании сравнения определяют, существует ли фетальная хромосомная анеуплоидия для первой хромосомы. Секвенирование выгодно максимизирует количество генетической информации, которую можно получить из ограниченного количества фетальных нуклеиновых кислот, которые существуют как незначительная популяция в биологическом образце, содержащем материнские фоновые нуклеиновые кислоты.

Согласно одному из примеров воплощения биологический образец, полученный от беременной

женщины, анализируют для осуществления пренатального диагноза фетальной хромосомной анеуплоидии. Биологический образец включает молекулы нуклеиновых кислот. Идентифицируют процент фетальных ДНК в биологическом образце. На основании процента вычисляют число N анализируемых последовательностей с учетом желательной точности. Секвенируют рандомизированно, по меньшей мере, N молекул нуклеиновых кислот, содержащихся в биологическом образце.

На основании рандомизированного секвенирования определяют первое количество первой хромосомы из последовательностей, идентифицированных как происходящие от первой хромосомы. Второе количество одной или нескольких вторых хромосом определяют из последовательностей, идентифицированных как происходящие от одной из вторых хромосом. Затем параметр от первого количества и второго количества сравнивают с одним или несколькими пороговыми значениями. На основании сравнения определяют, существует ли фетальная хромосомная анеуплоидия для первой хромосомы. Рандомизированное секвенирование выгодно максимизирует количество генетической информации, которую можно получить из ограниченного количества фетальных нуклеиновых кислот, которые существуют как незначительная популяция в биологическом образце, содержащем материнские фоновые нуклеиновые кислоты.

Другие воплощения изобретения относятся к системам и носителям, считываемым на компьютере, связанным со способами, описанными в данном описании.

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

Краткое описание чертежей

Фиг. 1 представляет собой схему способа 100 для осуществления пренатальной диагностики фетальной хромосомной анеуплоидии в биологическом образце, полученном от беременной женщины, согласно воплощению настоящего изобретения;

фиг. 2 - схему способа 200 для осуществления пренатальной диагностики фетальной хромосомной анеуплоидии с использованием рандомизированного секвенирования согласно воплощению настоящего изобретения;

фиг. 3А показывает график представленности в процентах последовательностей хромосомы 21 в образцах материнской плазмы, включающих плоды с трисомией 21 или эуплоидные плоды, согласно воплощению настоящего изобретения;

фиг. 3В - корреляцию между концентрациями фракции фетальной ДНК в материнской плазме, определенными массивным параллельным секвенированием и микрогидродинамической численной ПЦР, согласно воплощению настоящего изобретения;

фиг. 4А - диаграмму представленности в процентах выровненных последовательностей на хромосому согласно воплощению настоящего изобретения;

фиг. 4А - график различия (%) в представленности в процентах на хромосому между случаем трисомии 21 и случаем эуплоидии, показанными на фиг. 4А;

фиг. 5 - степень корреляции между сверхпредставленностью в последовательностях хромосомы 21 и концентрациями фракции фетальной ДНК в материнской плазме, включая плоды с трисомией 21, согласно воплощению настоящего изобретения;

фиг. 6 - таблицу части генома человека, которую анализируют согласно воплощению настоящего изобретения. T21 отмечает образец, полученный при беременности, включающей плод с трисомией 21;

фиг. 7 - таблицу числа последовательностей, требуемых для того, чтобы отличить эуплоидный плод от плода с трисомией 21 согласно воплощению настоящего изобретения;

фиг. 8А - таблицу верхних десяти начальных позиций секвенированных меток, выровненных с хромосомой 21, согласно воплощению настоящего изобретения;

фиг. 8В - таблицу верхних десяти начальных позиций секвенированных меток, выровненных с хромосомой 22, согласно воплощению настоящего изобретения;

фиг. 9 - 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(ых) кислоты(кислот), представляющих интерес.

Термин "нуклеиновая кислота" или "полинуклеотид" относится к дезоксирибонуклеиновой кислоте (ДНК) или рибонуклеиновой кислоте (РНК) и их полимеру или в одно- или в двухцепочечной форме. Если конкретно не указано иное, термин охватывает нуклеиновые кислоты, содержащие известные аналоги природных нуклеотидов, которые имеют свойства связывания, схожие с эталонной нуклеиновой кислотой, и метаболизируются подобно нуклеотидам, встречающимся в природе. Если не указано иное, определенная нуклеотидная последовательность также безоговорочно охватывает ее консервативно модифицированные варианты (например, замещения вырожденного кодона), аллели, ортологи, SNP и комплементарные последовательности, а также точно указанную последовательность. Конкретно замещений вырожденного кодона можно достичь, генерируя последовательности, в которых третья позиция одного или нескольких выбранных (или всех) кодонов замещена смешанными остатками оснований и/или де-

зоксинозинов (Batzer et al., *Nucleic Acids 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" используется как взаимозаменяемый с геном, к ДНК, мРНК, малой некодирующей РНК, микроРНК (миРНК), взаимодействующей с Рiwi РНК и короткой шпилечной РНК (shRNA), кодируемой геном или локусом.

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

Термин "реакция", используемый в данном описании, относится к любому процессу, включающему химическое, ферментативное или физическое действие, которое указывает на присутствие или отсутствие определенной полинуклеотидной последовательности, представляющей интерес. Примером "реакции" является реакция амплификации, такая как полимеразная цепная реакция (ПЦР). Другим примером "реакции" является реакция секвенирования или путем синтеза или путем лигирования. "Информативная реакция" представляет собой реакцию, которая указывает на присутствие одной или нескольких определенных полинуклеотидных последовательностей, представляющих интерес, и в одном случае, когда присутствует только одна последовательность, представляющая интерес. Термин "лунка", используемый в данном описании, относится к реакции в предварительно обусловленном месте в пределах ограниченной структуры, например в сосуде, клетке или камере в форме лунки в ряду операций ПЦР.

Термин "клинически релевантная нуклеотидная последовательность", используемый в данном описании, может относиться к полинуклеотидной последовательности, соответствующей сегменту более крупной геномной последовательности, возможный дисбаланс которой проверяют, или самой более крупной геномной последовательности. Одним из примеров является последовательность хромосомы 21. Другие примеры включают хромосомы 18, 13, X и Y. Еще другие примеры включают мутированные генетические последовательности или генетические полиморфизмы или вариации числа копий, которые плод может унаследовать от одного или обоих родителей. Еще другие примеры включают последовательности, которые мутированы, делетированы или амплифицированы в злокачественной опухоли, например последовательности, в которых происходит потеря гетерозиготности или дупликации генов. В некоторых воплощениях несколько клинически релевантных нуклеотидных последовательностей или, равнозначно, несколько маркеров клинически релевантной нуклеотидной последовательности можно использовать для получения данных для детекции дисбаланса. Например, данные о пяти не следующих подряд друг за другом последовательностях на хромосоме 21 можно использовать вместе для определения возможного дисбаланса хромосомы 21, причем эффективно уменьшается потребность в объеме образца до 1/5.

Термин "фоновая нуклеотидная последовательность", используемый в данном описании, относится к нуклеотидной последовательности, нормальное отношение которой к клинически релевантной нуклеотидной последовательности известно, например, отношение 1 к 1. Как один из примеров, фоновая нуклеотидная последовательность и клинически релевантная нуклеотидная последовательности представляют собой два аллеля из одной и той же хромосомы, различающиеся по гетерозиготности. В другом примере фоновая нуклеотидная последовательность представляет собой один аллель, гетерозиготный к другому аллелю, который представляет собой клинически релевантную нуклеотидную последовательность. Более того, некоторые из любых - фоновой нуклеотидной последовательности и клинически релевантной нуклеотидной последовательности могут происходить от различных индивидуумов.

Термин "эталонная нуклеотидная последовательность", используемый в данном описании, относится к нуклеотидной последовательности, средняя концентрация которой на реакцию известна или, что равнозначно, измерена.

Термин "сверхпредставленная нуклеотидная последовательность", используемый в данном описании, относится к одной из двух нуклеотидных последовательностей, представляющих интерес (например, клинически релевантной нуклеотидной последовательности и фоновой нуклеотидной последовательности), которой в биологическом образце больше, чем другой последовательности.

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

Термин "количественные данные", используемый в данном описании, означает, что данные получены из одной или нескольких реакций и что получены одно или несколько численных значений. Например, число лунок, которое показывает флуоресцентный маркер для определенной последовательности, может представлять собой количественные данные.

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

Термин "пороговое значение", обозначает численное значение, которое используют для того, чтобы вынести решение о двух или большем числе состояний (например, болезненное или неболезненное) для классификации биологического образца. Например, если параметр больше порогового значения, количественным данным дают первую классификацию (например, болезненное состояние); или если параметр меньше порогового значения, количественным данным дают вторую классификацию (например, неболезненное состояние).

Термин "дисбаланс", используемый в данном описании, обозначает значимое отклонение, определенное по меньшей мере по одному пороговому значению в количестве клинически релевантной нуклеотидной последовательности в сравнении с эталонной последовательностью. Например, эталонное количество может представлять собой отношение 3/5, и, таким образом, дисбаланс может иметь место, если измеренное отношение равно 1:1.

Термин "хромосомная анеуплоидия", используемый в данном описании, обозначает отклонение поддающегося измерению количества хромосомы от количества диплоидного генома. Отклонение может представлять собой прирост или потерю. Оно может включать всю одну хромосому или участок хромосомы.

Термин "рандомизированное секвенирование", используемый в данном описании, относится к секвенированию, при котором секвенируемые фрагменты нуклеиновой кислоты специфически не идентифицируют или не намечают перед процедурой секвенирования. Последовательность специфических праймеры для локализации специфических генных локусов не требуются. Секвенированные пулы нуклеиновых кислот изменяются от образца к образцу и даже от анализа к анализу одного и того же образца. Идентичность секвенированных нуклеиновых кислот обнаруживается только из секвенирования полученной продукции. В некоторых воплощениях настоящего изобретения рандомизированному секвенированию могут предшествовать процедуры обогащения биологического образца определенными популяциями молекул нуклеиновых кислот, разделяющих некоторые общие особенности. В одном воплощении каждый из фрагментов в биологическом образце имеет равную вероятность быть секвенированным.

Термин "фракция генома человека" или "часть генома человека", используемый в данном описании, относится менее чем к 100% нуклеотидных последовательностей в геноме человека, который включают около 3 миллиардов пар оснований нуклеотидов. В контексте секвенирования он относится менее чем к 1-кратному охвату нуклеотидных последовательностей в геноме человека. Термин может выражаться в процентах от абсолютного числа нуклеотидов/пар оснований. Как пример использования, термин может использоваться как относящийся к фактическому объему секвенирования. Воплощения могут определять требуемое минимальное значение секвенированной фракции генома человека для того, чтобы поставить точный диагноз. Как другой пример использования, термин можно отнести к количеству данных по секвенированию, используемых для получения параметра или всей суммы данных для классификации заболевания.

Термин "секвенированная метка", используемый в данном описании, относится к ряду секвенированных нуклеотидов из любой части или всей молекулы нуклеиновой кислоты. Например, секвенированная метка может представлять собой короткий ряд секвенированных нуклеотидов из фрагмента нуклеиновой кислоты, короткий ряд нуклеотидов по обоим концам фрагмента нуклеиновой кислоты или секвенирование всего фрагмента нуклеиновой кислоты, который существует в биологическом образце. Фрагмент нуклеиновой кислоты представляет собой любую часть более крупной молекулы нуклеиновой кислоты. Фрагмент (например, ген) может существовать отдельно (т.е. не быть связанным) от других частей более крупной молекулы нуклеиновой кислоты.

Осуществление изобретения

Воплощения данного изобретения относятся к способам, системам и аппаратуре для определения того, имеется ли увеличение или уменьшение (болезненное состояние) клинически релевантного хромосомного участка по сравнению с неболезненным состоянием. Такое определение можно осуществить с использованием параметра количества клинически релевантного хромосомного участка относительно других клинически нерелевантных хромосомных участков (фоновые участки) в биологическом образце. Молекулы нуклеиновых кислот биологического образца секвенируют, так что секвенируют часть генома, и количество можно определить из результатов секвенирования. Выбирают одно или несколько пороговых значений для определения того, имеется ли изменение по сравнению с эталонным количеством (т.е. дисбаланс), например, в отношении отношения количеств двух хромосомных участков (или наборов участков).

Изменение, обнаруженное в эталонном количестве, может представлять собой любое отклонение (в большую или меньшую сторону) в отношении клинически релевантной нуклеотидной последовательности к другим клинически нерелевантным нуклеотидным последовательностям. Таким образом, эталонное состояние может представлять собой любое отношение или другую количественную характеристику (например, иное соотношение, чем 1÷1), и измеренное состояние, обозначающее изменение, может представлять собой любое отношение или другую количественную характеристику, отличающиеся от эталонной количественной характеристики, определенной по одному или нескольким пороговым значениям.

Клинически релевантный хромосомный участок (также называемый клинически релевантной нук-

леотидной последовательностью) и фоновая нуклеотидная последовательность могут происходить от первого типа клеток и от одного или нескольких вторых типов клеток. Например, фетальные нуклеотидные последовательности, полученные из клеток плода/плаценты, присутствуют в биологическом образце, таком как материнская плазма, который содержит фоновые материнские нуклеотидные последовательности, происходящие от материнских клеток. В одном воплощении пороговое значение определяют на основании, по меньшей мере частично, процента первого типа клеток в биологическом образце. Данные по проценту последовательностей плода в образце можно определить с помощью любых локусов, полученных от плода, и не ограничиваться измерением клинически релевантных нуклеотидных последовательностей. В другом воплощении пороговое значение определяют, по меньшей мере частично, по проценту опухолевых последовательностей в биологическом образце, таком как плазма, сыворотка, слюна или моча, который содержит фон нуклеотидных последовательностей, полученных из незлокачественных клеток организма.

I. Общий способ.

Фиг. 1 представляет собой схему способа 100 для осуществления пренатальной диагностики фетальной хромосомной анеуплоидии в биологическом образце, полученном от беременной женщины, согласно воплощению настоящего изобретения.

На стадии 110 получают биологический образец от беременной женщины. Биологический образец может представлять собой плазму, мочу, сыворотку или любой другой подходящий образец. Образец содержит молекулы нуклеиновых кислот от плода и беременной женщины. Например, молекулы нуклеиновых кислот могут представлять собой фрагменты хромосом.

На стадии 120 секвенируют, по меньшей мере, часть нескольких молекул нуклеиновых кислот, содержащихся в биологическом образце. Секвенированная часть представляет фракцию генома человека. В одном воплощении молекулы нуклеиновых кислот представляют собой фрагменты соответствующих хромосом. Секвенированы могут быть один конец (например, 35 пар оснований (п.о.)), оба конца или весь фрагмент. Могут быть секвенированы все молекулы нуклеиновых кислот в образце, или может быть секвенировано только подмножество. Такое подмножество может быть выбрано произвольно, а также как описано подробнее ниже.

В одном воплощении секвенирование осуществляют с использованием массированного параллельного секвенирования. Массированное параллельное секвенирование, такое, которого можно достичь на платформе 454 (Roche) (Margulies M. et al., 2005, Nature, 437, 376-380), с анализатором Illumina Genome (или на платформе Solexa) или с системой SOLiD (Applied Biosystems) или технологией секвенирования Helicon True Single Molecule DNA (Harris T.D. et al., 2008, Science, 320, 106-109), одномолекулярной технологией в реальном времени (SMRT™) от Pacific Biosciences и нанопористым секвенированием (Soni G.V. and Meller A., 2007, Clin. Chem., 53: 1996-2001), позволяет параллельно секвенировать множество молекул нуклеиновых кислот, выделенных из образца, при высоких порядках мультиплексирования (Dear Brief Funct. Genomic Proteomic, 2003, 1:397-416). На каждой из таких платформ секвенируют клонально размноженные или даже неамплифицированные отдельные молекулы фрагментов нуклеиновых кислот.

Так как при каждом прогоне от каждого образца получают большое число данных секвенирования - порядка от сотен тысяч до миллионов или даже, возможно, сотни миллионов или миллиарды, полученные результаты секвенирования образуют характерный профиль смеси видов нуклеиновых кислот в исходном образце. Например, гаплотип, транскриптома и профили метилирования по данным секвенирования имеют сходство с показателями исходного образца (Brenner et al., Nat. Biotech., 2000, 18:630-634; Taylor et al., Cancer Res., 2007, 67: 8511-8518). Из-за большого числа отобранных последовательностей из каждого образца число идентичных последовательностей, таких как полученные при секвенировании пула нуклеиновых кислот при многократном охвате или высокой избыточности, также является хорошей количественной представленностью количества определенных видов нуклеиновых кислот или локуса в исходном образце.

На стадии 130 на основании секвенирования (например, результатов секвенирования) определяют первое количество первой хромосомы (например, клинически релевантной хромосомы). Первое количество определяют из последовательностей, идентифицированных как происходящие от первой хромосомы. Например, затем можно использовать процедуры биоинформатики для определения местонахождения каждой из таких последовательностей ДНК в геноме человека. Возможно, что доля таких последовательностей будет отброшена при последующем анализе, поскольку они присутствуют в повторяющихся участках генома человека или в участках, подвергнутых интериндивидуальным вариациям, например вариациям числа копий. Таким образом, можно определить количество хромосомы, представляющей интерес, и одной или нескольких других хромосом.

На стадии 140 на основании секвенирования определяют второе количество одной или нескольких вторых хромосом из последовательностей, идентифицированных как происходящие от одной из вторых хромосом. В одном воплощении вторые хромосомы представляют собой все другие хромосомы, кроме одной первой (т.е. испытывают одну). В другом воплощении вторая хромосома представляет собой только одну другую хромосому.

Существует ряд способов определения количеств хромосом, в том числе, но без ограничения, подсчет числа секвенированных меток, числа секвенированных нуклеотидов (пар оснований) или общей длины секвенированных нуклеотидов (пар оснований), происходящих от определенной(ых) хромосомы(хромосом) или участков хромосомы.

В другом воплощении могут быть установлены критерии для результатов секвенирования для определения того, что подсчитывать. В одном аспекте количество можно получить, основываясь на доле секвенированного продукта. Например, после анализа методами биоинформатики можно выбрать продукт секвенирования, соответствующий фрагментам нуклеиновой кислоты в определенном интервале по размеру. Примерами интервалов размера являются примерно <300 п.о., <200 п.о. или <100 п.о.

На стадии 150 определяют параметр из первого количества и второго количества. Параметр может представлять собой, например, простое отношение первого количества ко второму количеству или первого количества ко второму количеству плюс первое количество. В одном аспекте каждое количество может представлять собой аргумент функций или отдельных функций, где затем из таких отдельных функций может быть взято отношение. Специалисту в данной области техники будет понятно значение ряда различных подходящих параметров.

В одном воплощении параметр (например, представленность фракции) хромосомы, потенциально вовлеченный в хромосомную анеуплоидию, например хромосомы 21, или хромосомы 18, или хромосомы 13, затем можно вычислить из результатов процедуры биоинформатики. Представленность фракции можно получить на основании количества всех последовательностей (например, некоторой меры всех хромосом, включая клинически релевантную хромосому) или определенного подмножества хромосом (например, только одной другой хромосомы кроме испытываемой).

На стадии 150 параметр сравнивают с одним или несколькими пороговыми значениями. Пороговые значения можно определить любым подходящим путем. Такие пути включают метод правдоподобия по байесовскому типу, последовательный критерий отношения вероятностей (SPRT), ложное обнаружение, доверительный интервал, оперативную характеристику получателя (ROC). Примеры применений таких методов и специфических для образцов методов описаны в одновременно рассматриваемой заявке "Определение дисбаланса нуклеотидных последовательностей" (досье поверенного № 016285-005210US), включенной в данное описание в качестве ссылки.

В одном воплощении параметр (например, представленность фракции клинически релевантной хромосомы) затем сравнивают с эталонным интервалом, установленным при беременностях, включающих здоровые (т.е. euploidные) плоды. Возможно, что в некоторых вариантах процедуры эталонный интервал (т.е. пороговые значения) можно регулировать в соответствии с концентрацией фракции фетальной ДНК (f) в определенном образце материнской плазмы. Величину f можно определить из набора данных секвенирования, например, с использованием последовательностей, картируемых к хромосоме Y, если пол плода мужской. Величину f также можно определить в отдельном анализе, например, с использованием эпигенетических маркеров плода (Chan K.C.A. et al., 2006, Clin. Chem., 52, 2211-8) или из анализа однонуклеотидных полиморфизмов.

На стадии 160 на основании сравнения выполняют классификацию - существует ли фетальная хромосомная анеуплоидия для первой хромосомы. В одном воплощении классификация представляет собой определенное "да" или "нет". В другом воплощении классификация может быть неподдающейся определению или неопределенной. В еще одном воплощении классификация может представлять собой оценку, которая должна быть интерпретирована позднее, например врачом.

II. Секвенирование, выравнивание и определение количеств.

Как указано выше, секвенируют только часть генома. В одном аспекте, даже когда секвенируют пул нуклеиновых кислот в образце при <100% геномном охвате вместо многократного охвата и из числа захваченных молекул нуклеиновых кислот, большинство каждого вида нуклеиновой кислоты секвенируют только один раз. Также можно определить дисбаланс дозы определенной хромосомы или участков хромосомы. Иными словами, дисбаланс дозы определенной хромосомы или участков хромосомы выводят из представленности в процентах указанного локуса среди других картируемых секвенированных меток образца.

Это контрастирует с ситуациями, когда один и тот же пул нуклеиновых кислот секвенируют несколько раз для того, чтобы достичь высокой избыточности или многократного охвата, посредством чего каждый вид нуклеиновой кислоты секвенируется несколько раз. В таких ситуациях число раз секвенирования определенного вида нуклеиновой кислоты относительно другого вида нуклеиновой кислоты коррелирует с их относительными концентрациями в исходном образце. Стоимость секвенирования возрастает с числом крат охвата, требуемым для достижения точной представленности вида нуклеиновой кислоты.

В одном примере как пояснительном доля таких последовательностей может быть от хромосомы, вовлеченной в анеуплоидию, такой как хромосома 21. Могут быть получены еще другие последовательности из других хромосом при таком осуществлении секвенирования. Учитывая относительный размер хромосомы 21 по сравнению с другими хромосомами, при таком осуществлении секвенирования можно получить нормализованную частоту встречаемости в эталонном интервале последовательностей, специфических для хромосомы 21. Если плод имеет трисомию 21, тогда нормализованная частота встречаемо-

сти последовательностей, образованных хромосомой 21 при таком осуществлении секвенирования, будет возрастать, что дает, таким образом, возможность обнаружения трисомии 21. Степень изменения при нормализации частоты встречаемости будет зависеть от концентрации фракции фетальных нуклеиновых кислот в анализируемом образце.

В одном воплощении используют анализатор Illumina Genome для секвенирования геномной ДНК человека и образцов ДНК плазмы человека по одному концу. Анализатор Illumina Genome секвенирует молекулы клонально размноженной одной ДНК, захваченные на твердой поверхности, названной проточной ячейкой. Каждая проточная ячейка имеет 8 полос для секвенирования 8 отдельных образцов или пулов образцов. Каждая полоса способна генерировать ~200 млн осн. (Mb) последовательности, которая представляет собой только фракцию последовательностей в 3 миллиарда пар оснований в геноме человека. Каждую геномную ДНК или образец ДНК плазмы секвенируют с использованием одной полосы проточной ячейки. Полученные короткие метки последовательностей выравнивают с эталонной последовательностью генома и отмечают хромосомное происхождение. Общее число отдельных секвенированных меток, выровненных с каждой хромосомой, табулируют и сравнивают с относительным размером каждой хромосомы, ожидаемой от эталонного генома человека или характерных образцов без заболевания. Затем идентифицируют прирост или утрату хромосом.

Описанный подход является только одним примером описанной в данном описании стратегии дозы гена/хромосомы. С другой стороны, можно осуществить секвенирование по парным концам. Вместо сравнения длины секвенированных фрагментов с длиной, ожидаемой в эталонном геноме, как описано в Campbell et al. (Nat. genet., 2008, 40:722-729), подсчитывают число выровненных секвенированных меток и сортируют согласно местоположению хромосом. Прирост или утрату хромосомных участков или целых хромосом определяют путем сравнения полученных чисел с ожидаемым размером хромосом в эталонном геноме или в характерных образцах без заболевания. Так как секвенирование по парным концам дает возможность установить размер исходного фрагмента нуклеиновой кислоты, в одном примере фокусируются на подсчете числа парных секвенированных меток, соответствующих фрагментам нуклеиновых кислот обусловленного размера, такого как <300 п.о., <200 п.о. или <100 п.о.

В другом воплощении фракцию пула нуклеиновых кислот, которую секвенируют за прогон, подвергают дополнительному отбору перед секвенированием. Например, можно использовать методы на основе гибридизации, такие как олигонуклеотидная матрица, для первого подбора нуклеотидных последовательностей из определенной хромосомы, например хромосомы, потенциально анеуплоидной, и другая(ие) хромосома(ы) не участвуют в испытании на анеуплоидию. Другим примером является то, что перед секвенированием из пула образцов дополнительно выбирают или обогащают определенное подмножество нуклеотидных последовательностей. Например, как обсуждалось выше, имеется сообщение, что молекулы фетальной ДНК в материнской плазме сравнивают с более короткими фрагментами, чем молекулы материнской фоновой ДНК (Chan et al., Clin. Chem., 2004, 50: 88-92). Таким образом, специалисты в данной области техники могут использовать один или несколько известных способов фракционирования нуклеотидных последовательностей в образце в соответствии с размером молекул, например электрофорез в геле или колонки для исключения по размеру или подход на основе микрогидродинамики. Также, с другой стороны, в примере анализа внеклеточной фетальной ДНК в материнской плазме часть фетальных нуклеиновых кислот можно обогатить методом, при котором подавляется материнский фон, таким как добавление формальдегида (Dhallan et al., JAMA, 2004, 291:1114-9). В одном воплощении часть подмножества предварительно отобранного пула нуклеиновых кислот секвенируют рандомизированно.

Другие стратегии секвенирования отдельных молекул, такие как на платформе 454 Roche, платформе Applied Biosystems SOLiD, технология секвенирования Helicon True Single Molecule DNA, одномолекулярная технология в реальном времени (SMRT) от Pacific Biosciences и нанопористое секвенирование, можно подобным образом использовать в данной заявке.

III. Определение количеств хромосом из продукта секвенирования.

После массивного параллельного секвенирования осуществляют анализ методами биоинформатики для установления места хромосомного происхождения секвенированных меток. После такой процедуры метки, идентифицированные как происходящие от потенциально анеуплоидной хромосомы, т.е. хромосомы 21 в данном исследовании, сравнивают количественно со всеми секвенированными метками или метками, происходящими от одной или нескольких хромосом, не вовлеченных в анеуплоидию. Соотношение между продуктом секвенирования из хромосомы 21 и других не-21 хромосом в испытываемом образце сравнивают с пороговыми значениями, полученными методами, описанными в разделе, приведенном выше, для определения того, получен образец при беременности, включающей эуплоидный плод или плод с трисомией 21.

Число различных количеств включает, но не ограничивается, перечисленное далее, что можно было бы получить из секвенированных меток. Например, число секвенированных меток, т.е. подсчитанных абсолютно, выровненных с определенной хромосомой, можно было бы сравнить с абсолютным числом секвенированных меток, выровненных с другими хромосомами. С другой стороны, подсчет фракции количества секвенированных меток из хромосомы 21 относительно всех или некоторых других секвениро-

ванных меток можно было бы сравнить с соответствующим значением для других неанэуплоидных хромосом. В данном эксперименте, поскольку секвенируют 36 п.о. из каждого фрагмента ДНК, число секвенированных нуклеотидов из определенной хромосомы можно легко получить из 36 п.о. умножением на число секвенированных меток.

Кроме того, так как каждый образец материнской плазмы секвенируют только с использованием одной проточной ячейки, которая может секвенировать только часть генома человека, по статистике, большинство видов фрагментов ДНК материнской плазмы могут быть секвенированы с образованием только одного числа секвенированных меток. Иными словами, фрагменты нуклеиновых кислот, присутствующие в образце материнской плазмы, секвенируют менее чем с 1-кратным охватом. Таким образом, общее число секвенированных нуклеотидов для любой определенной хромосомы будет преимущественно соответствовать количеству, доле или длине части указанной хромосомы, которая секвенирована. Следовательно, количественное определение представленности потенциально анэуплоидной хромосомы можно получить из фракции числа или эквивалентной длины секвенированных нуклеотидов из такой хромосомы относительно полученного подобного образом количества для других хромосом.

IV. Обогащение пулов нуклеиновых кислот для секвенирования.

Как указано выше и установлено в разделе примеров ниже, только часть генома человека требуется для секвенирования для того, чтобы отличить случаи трисомии 21 от эуплоидии. Таким образом, возможно и эффективно по стоимости обогащать пул нуклеиновых кислот для секвенирования перед рандомизированным секвенированием фракции обогащенного пула. Например, молекулы фетальной ДНК в материнской плазме состоят из более коротких фрагментов, чем молекулы материнской фоновой ДНК (Chan et al., Clin. Chem., 2004, 50: 88-92). Таким образом, специалисты могут использовать один или несколько известных способов для фракционирования нуклеотидных последовательностей в образце в соответствии с размером молекул, например с помощью электрофореза в геле или колонок для исключения по размерам или подхода на основе микрогидродинамики.

Также, с другой стороны, в примере анализа внеклеточной фетальной ДНК в материнской плазме долю фетальных нуклеиновых кислот можно обогатить методом, при котором подавляется материнский фон, таким как добавление формальдегида (Dhallan et al., JAMA, 2004, 291:1114-9). Доля полученных фетальных последовательностей может быть обогащена в пуле нуклеиновых кислот, включающих более короткие фрагменты. Согласно фиг. 7 число секвенированных меток, требуемое для того, чтобы отличить случаи трисомии 21 от эуплоидии, будет уменьшаться, так как повышается концентрация фракции фетальной ДНК.

С другой стороны, последовательности, происходящие от потенциально анэуплоидной хромосомы и от одной или нескольких хромосом, не вовлеченных в анэуплоидию, можно обогатить методами гибридизации, например, на олигонуклеотидных микроматрицах. Затем обогащенные пулы нуклеиновых кислот можно подвергнуть рандомизированному секвенированию. Это может создать возможность для уменьшения стоимости секвенирования.

V. Рандомизированное секвенирование.

Фиг. 2 представляет собой схему способа 200 осуществления пренатальной диагностики фетальной хромосомной анэуплоидии с использованием рандомизированного секвенирования согласно воплощению настоящего изобретения. В одном аспекте в случае подхода с массивным параллельным секвенированием одновременно можно получить представительные результаты от всех хромосом. Источник определенного фрагмента заранее не выбирают. Секвенирование осуществляют рандомизированно, и затем можно осуществить поиск в базе данных для того, чтобы увидеть, откуда появился определенный фрагмент. Это противоположно ситуациям, когда амплифицируют специфический фрагмент из хромосомы 21 и другой фрагмент из хромосомы 1.

На стадии 210 получают биологический образец от беременной женщины. На стадии 220 вычисляют число N последовательностей для анализа для желательной точности. В одном воплощении сначала идентифицируют процент фетальной ДНК в биологическом образце. Это можно осуществить любым подходящим способом, известным специалистам в данной области техники. Идентификация может просто зарегистрировать величину, которая измерена с помощью другого объекта. В данном воплощении вычисление числа N последовательностей для анализа основана на проценте. Например, число последовательностей, необходимое для анализа, может возрастать, когда падает процент фетальной ДНК, и может понижаться, когда процент фетальной ДНК растет. Число N может быть фиксированным числом или относительным числом, таким как процент. В другом воплощении может быть известно число последовательностей N , адекватное для точного диагноза заболевания. Можно получить достаточное число N даже при беременностях с концентрациями фетальной ДНК на нижнем пределе нормального интервала.

На стадии 230 рандомизированно секвенируют, по меньшей мере, N молекул из множества молекул нуклеиновых кислот, содержащихся в биологическом образце. Особенностью такого описанного подхода является то, что секвенируемые нуклеиновые кислоты специально не идентифицируют или не локализируют перед анализом образца, т.е. секвенированием. Последовательностьспецифические праймеры для локализации специфических генных локусов для секвенирования не требуются. Пулы секвенированных нуклеиновых кислот изменяются от образца к образцу и даже от анализа к анализу одного и того же об-

разца. Более того, из последующего описания (фиг. 6) следует, что количество продукта секвенирования, необходимое в случае диагностики, может изменяться между испытываемыми образцами и эталонной популяцией. Такие аспекты заметно контрастируют с большинством диагностических подходов, таких как основанные на флуоресценции гибридизации *in situ*, количественной флуоресцентной ПЦР, количественной ПЦР в реальном времени, численной ПЦР, сравнительной геномной гибридизации, микроматричной сравнительной геномной гибридизации и т.д., где требуется локализация генных локусов до предварительного определения, причем таким образом необходимо использовать локуспецифические праймеры или набор зондов или их панели.

В одном воплощении рандомизированное секвенирование осуществляют на фрагментах ДНК, которые присутствуют в плазме беременной женщины, и получают геномные последовательности, которые могут происходить или от плода или от матери. Рандомизированное секвенирование включает получение образцов (секвенирование) произвольной части молекул нуклеиновой кислоты, присутствующих в биологическом образце. Так как секвенирование рандомизированное, в каждом анализе могут быть секвенированы различные подмножества (фракции) молекул нуклеиновых кислот (и, таким образом, генома). Воплощения будут работать даже когда такое подмножество изменяется от образца к образцу и от анализа к анализу, что может происходить даже с использованием одного и того же образца. Примеры фракции составляют примерно 0,1, 0,5, 1, 5, 10, 20 или 30% генома. В других воплощениях фракция составляет, по меньшей мере, любую из указанных величин.

Остальные стадии 240-270 могут протекать подобно стадиям в способе 100.

VI. Отбор пулов секвенированных меток после секвенирования.

Как описано ниже в примерах II и III, подмножество данных секвенирования является достаточным для того, чтобы отличить случаи трисомии 21 от случаев эуплоидии. Подмножество данных секвенирования может представлять собой долю секвенированных меток, которые выходят за пределы некоторых качественных параметров. Например, в примере II используют секвенированные метки, которые уникально выравниваются с эталонным геномом человека, замаскированного повторами. С другой стороны, можно секвенировать характерный пул фрагментов нуклеиновых кислот из всех хромосом, но сосредоточиться на сравнении между данными, относящимися к потенциально анеуплоидной хромосоме, и данными, относящимися к ряду неанеуплоидных хромосом.

Также, с другой стороны, подмножество продукта секвенирования, охватывающее секвенированные метки, образованные из фрагментов нуклеиновых кислот, соответствующих окну определенного размера в исходном образце, можно разобрать во время анализа после секвенирования. Например, с использованием анализатора Illumina Genome можно использовать секвенирование по парным концам, которое относится к секвенированию обоих концов фрагментов нуклеиновых кислот. Данные секвенирования каждой пары концов затем выравнивают с эталонной последовательностью генома человека. Затем можно установить расстояние или число нуклеотидов, заключающихся между двумя концами. Также можно установить полную длину исходного фрагмента нуклеиновой кислоты. С другой стороны, платформы секвенирования, такие как платформа 454, и возможно, некоторые другие методы одномолекулярного секвенирования, дают возможность секвенировать полную длину коротких фрагментов нуклеиновых кислот, например, 200 п.о. Таким способом из данных секвенирования можно сразу же узнать фактическую длину фрагмента нуклеиновой кислоты.

Такой анализ парных концов также возможен с использованием других платформ секвенирования, например системы Applied Biosystems SOLiD. В случае платформы Roche 454 из-за повышенной регистрируемой длины по сравнению с другими системами массивного параллельного секвенирования также возможно определить длину фрагмента из его полной последовательности.

Преимущество фокусирования данных анализа на подмножестве секвенированных меток, соответствующих коротким фрагментам нуклеиновых кислот в исходном образце материнской плазмы, имеет место из-за того, что набор данных будет эффективно обогащен последовательностями ДНК, полученными от плода. Это так, поскольку молекулы фетальной ДНК в материнской плазме составляют более короткие фрагменты, чем молекулы фоновой ДНК (Chan et al., Clin. Chem., 2004, 50: 88-92). Согласно фиг. 7 число секвенированных меток, требуемое для того, чтобы отличить случаи эуплоидии от случаев трисомии 21, может быть уменьшено, так как возрастает концентрация фетальной ДНК.

Отбор подмножеств пулов нуклеиновых кислот после секвенирования отличается от других стратегий обогащения нуклеиновых кислот, которые осуществляют перед анализом образца, таких как применение электрофореза в геле или колонок с исключением по размерам, для отбора нуклеиновых кислот определенного размера, которые требуют физического отделения обогащенного пула от фонового пула нуклеиновых кислот. Физические процедуры будут вводить больше экспериментальных стадий и могут привести к проблемам, таким как загрязнение. Отбор *in silico* после секвенирования подмножеств из продукта секвенирования также может допускать отбор в зависимости от чувствительности и специфичности, требуемых для определения заболевания.

Биоинформативные, вычислительные и статистические подходы, используемые для определения того, получен образец материнской плазмы от беременной женщины с плодом с трисомией 21 или эуплоидным плодом, можно исполнить в компьютерном программном продукте, используемом для опреде-

ления параметров из результатов секвенирования. Работа компьютерной программы может включать определение поддающегося измерению количества потенциально анеуплоидной хромосомы, а также количества(количеств) одной или нескольких других хромосом. Параметр можно определить и сравнить с соответствующими пороговыми значениями для определения, в случае потенциально анеуплоидной хромосомы, имеется ли фетальная хромосомная анеуплоидия.

Примеры

Приведенные далее примеры приводятся для пояснения, но не для ограничения изобретения.

I. Пренатальная диагностика фетальной трисомии 21.

Отбирают для исследования восемь беременных женщин. Все беременные женщины находятся на 1- или 2-м триместре беременности и имеют одноплодную беременность. У четырех из них плоды с трисомией 21, а у других четырех эуплоидные плоды. У каждой берут 20 мл периферической венозной крови. Материнскую плазму собирают после центрифугирования при 1600×g в течение 10 мин и затем центрифугируют при 16000×g в течение 10 мин. Затем экстрагируют ДНК из 5-10 мл каждого образца плазмы. Затем ДНК материнской плазмы используют для массивного параллельного секвенирования с помощью анализатора Illumina Genome согласно инструкциям изготовителя. Работники, выполняющие секвенирование, во время секвенирования и анализа данных о последовательностях не знают о диагнозе плода.

Коротко, приблизительно 50 нг ДНК материнской плазмы используют для получения библиотеки ДНК. Можно начинать с меньших количеств ДНК материнской плазмы, таких как 15 или 10 нг. Фрагменты ДНК материнской плазмы затупляют по концам, лигируют с адаптерами Solexa, и отбирают фрагменты в 150-300 п.о. очисткой в геле. С другой стороны, затупленные по концам и адаптерлигированные фрагменты ДНК материнской плазмы можно пропустить через колонки (например, AMPure, Agencourt) для удаления нелигированных адаптеров без отбора по размеру перед получением кластера. Адаптерлигированную ДНК гибридизируют с поверхностью проточных ячеек и получают кластеры ДНК с использованием кластерной системы Illumina, и затем следуют 36 циклов секвенирования на анализаторе Illumina Genome. ДНК из каждого образца материнской плазмы секвенируют только в одной проточной ячейке. Считанные данные секвенирования собирают с использованием конвейера для анализа Solexa Analysis Pipeline. Затем все считанные данные выравнивают с эталонной последовательностью генома человека, маскированной повторами, комплект 36 NCBI (GenBank, инвентарные номера NC_000001 - NC_000024), с использованием программы Eland.

В данном исследовании для уменьшения сложности анализа данных далее рассматривают только последовательности, картированные к единственному участку эталона генома человека, маскированного повторами. С другой стороны, можно использовать другие подмножества или весь набор данных секвенирования. Подсчитывают общее число уникально картируемых последовательностей для каждого образца. Число последовательностей, уникально выровненных с хромосомой 21, выражают для каждого образца в виде доли от общего числа выровненных последовательностей. Так как материнская плазма содержит фетальную ДНК среди фоновой ДНК материнского происхождения, плод с трисомией 21 будет вносить дополнительные секвенированные метки, происходящие от хромосомы 21, из-за присутствия дополнительной копии хромосомы 21 в фетальном геноме. Следовательно, процент последовательностей хромосомы 21 в материнской плазме при беременности с плодом с трисомией 21 будет выше, чем процент при беременности с эуплоидным плодом. Анализ не требует локализации фетальноспецифических последовательностей. Также не требуется предварительное физическое отделение фетальных нуклеиновых кислот от материнских. Также нет необходимости отличать или идентифицировать фетальные и материнские последовательности после секвенирования.

Фиг. 3А показывает процент последовательностей, картированных к хромосоме 21 (представленность в процентах хромосомы 21) для каждого из 8 образцов ДНК материнской плазмы. Представленность в процентах хромосомы 21 существенно выше в материнской плазме при беременности с трисомией 21, чем в плазме при эуплоидной беременности. Такие данные предполагают, что можно осуществить неинвазивную пренатальную диагностику фетальной анеуплоидии, определяя представленность в процентах анеуплоидной хромосомы в сравнении с эталонной популяцией. С другой стороны, сверхпредставленность хромосомы 21 можно детектировать, сравнивая представленность в процентах хромосомы 21, полученную экспериментально, с представленностью в процентах последовательностей хромосомы 21, ожидаемой в случае эуплоидного генома человека. Это можно осуществить, маскируя или не маскируя повторные участки в геноме человека.

У пяти из восьми беременных женщин пол плода мужской. Последовательности, картированные к хромосоме Y, могут быть фетальноспецифическими. Процент последовательностей, картированных к хромосоме Y, используют для вычисления концентрации фракции фетальной ДНК в исходном образце материнской плазмы. Более того, концентрацию фракции фетальной ДНК также определяют, используя микрогидродинамическую численную ПЦП с участием паралолических генов X-связанного цинкопальцевого белка (ZFX) и Y-связанного цинкопальцевого белка (ZFY).

Фиг. 3В показывает корреляцию концентраций фракции фетальной ДНК, выведенных по представ-

ленности в процентах хромоома Y путем секвенирования, и концентраций, определенных микрогидродинамической численной ПЦР ZFY/ZFX . Имеется положительная корреляция между концентрациями фракции фетальной ДНК в материнской плазме, полученными указанными двумя способами. Коэффициент корреляции (r) составляет 0,917 при корреляционном анализе Пирсона.

Проценты последовательностей ДНК материнской плазмы, выровненных с каждой из 24 хромосом (22 аутосомы и хромосомы X и Y), для двух характерных случаев показаны на фиг. 4А. У одной из беременных женщин плод с трисомией 21, а у другой эуплоидный плод. Представленность в процентах последовательностей, картированных к хромосоме 21, выше у беременной женщины с плодом с трисомией 21 по сравнению с беременной женщиной с эуплоидным плодом.

Различия (%) в представленности в процентах на хромосому между образцами материнской плазмы в вышеуказанных двух случаях показаны на фиг. 4В. Различия в процентах для определенной хромосомы вычисляют с использованием следующей формулы:

$$\text{различие в процентах (\%)} = (P_{21} - P_E) / P_E \times 100\%,$$

где P_{21} - процент последовательностей ДНК плазмы, выровненных с определенной хромосомой, у беременной женщины с плодом с трисомией 21; и

P_E - процент последовательностей ДНК плазмы, выровненных с определенной хромосомой, у беременной женщины с эуплоидным плодом.

Как видно на фиг. 4В, существует сверхпредставленность - на 11% последовательностей хромосомы 21 в плазме беременной женщины с плодом с трисомией 21 в сравнении с беременной женщиной с эуплоидным плодом. В случае последовательностей, выровненных с другими хромосомами, различия между двумя случаями находятся в пределах 5%. Так как представленность в процентах для хромосомы 21 при трисомии 21 возрастает при сравнении с образцами эуплоидной материнской плазмы, различие (%) можно, с другой стороны, обозначить как степень сверхпредставленности в последовательностях хромосомы 21. Кроме различий (%) и абсолютных различий между представленностью в процентах хромосомы 21 также можно вычислить отношения подсчетов опытных и эталонных образцов, которые могут указывать степень сверхпредставленности хромосомы 21 при трисомии 21 при сравнении с эуплоидными образцами.

В случае четырех беременных женщин с эуплоидными плодами в среднем 1,345% последовательностей ДНК в их плазме выравниваются с хромосомой 21. У четырех беременных женщин с плодами с трисомией 21 три плода мужского пола. Представленность в процентах хромосомы 21 вычисляют для каждого из трех указанных случаев. Различия (%) в представленности в процентах хромосомы 21 для каждого из указанных трех случаев трисомии 21 со средней представленностью в процентах хромосомы 21, полученной из значений для четырех случаев эуплоидии, определяют так, как описано выше. Иными словами, среднее из четырех случаев с эуплоидными плодами используют в качестве эталона для такого вычисления. Концентрации фракции фетальной ДНК для указанных трех случаев мужской трисомии 21 выводят из их соответствующей представленности последовательностей хромосомы Y .

Корреляция между степенью сверхпредставленности последовательностей хромосомы 21 и концентрациями фракции фетальной ДНК показана на фиг. 5. Между двумя параметрами существует заметная положительная корреляция. Коэффициент корреляции (r) составляет 0,898 при корреляционном анализе Пирсона. Такие результаты показывают, что степень сверхпредставленности последовательностей хромосомы 21 в материнской плазме связана с концентрацией фетальной ДНК в образце материнской плазмы. Таким образом, для идентификации беременности с плодом с трисомией 21 можно определить пороговые значения степени сверхпредставленности последовательностей хромосомы 21, релевантные концентрациям фракции фетальной ДНК.

Определение концентрации фракции фетальной ДНК в материнской плазме также можно осуществить отдельно от прогона секвенирования. Например, концентрацию ДНК хромосомы Y можно определить предварительно с использованием ПЦР в реальном времени, микрогидродинамической ПЦР или масс-спектрометрии. Например, на фиг. 3В видно, что существует хорошая корреляция между концентрациями фракции фетальной ДНК, установленными на основе данных для хромосомы Y , полученных в прогоне секвенирования, и отношением ZFY/ZFX , полученным отдельно от прогона секвенирования. Действительно, концентрацию ДНК можно определить с использованием иных локусов, чем хромосома Y , и применимых к плоду женского пола. Например, Chan et al. показали, что происходящие от плода метилированные последовательности RASSF1A можно определить в плазме беременной женщины на фоне неметилированных последовательностей RASSF1A материнского происхождения (Chan et al., Clin. Chem., 2006, 52:2211-8). При этом концентрацию фракции фетальной ДНК можно определить путем деления количества метилированных последовательностей RASSF1A на количество всех последовательностей RASSF1A (метилированных и неметилированных).

Ожидается, что материнская плазма может быть предпочтительнее материнской сыворотки для практического осуществления изобретения, поскольку ДНК высвобождается из материнских клеток крови во время свертывания крови. Так, если использовать сыворотку, ожидается, что концентрация фракции фетальной ДНК будет ниже в материнской плазме, чем в материнской сыворотке. Иными словами, если использовать материнскую сыворотку, ожидается, что потребуется получить больше последова-

тельностей для диагностики фетальной хромосомной анеуплоидии по сравнению с образцом плазмы, полученным от той же беременной женщины в то же время.

Еще одним альтернативным путем определения концентрации фракции фетальной ДНК может быть количественное определение полиморфных различий между беременной женщиной и плодом (Dhallan R. et al., 2007, Lancet, 369, 474-481). Примером такого способа может быть локализация полиморфных сайтов, при которой беременная женщина является гомозиготной, а плод является гетерозиготным. Количество фетальноспецифического аллеля можно сравнить с количеством обычного аллеля для определения концентрации фракции фетальной ДНК.

В отличие от существующих методов обнаружения хромосомных aberrаций, включающих сравнительную геномную гибридизацию, микроматричную сравнительную геномную гибридизацию, количественную полимеразную цепную реакцию в реальном времени, которые детектируют и количественно определяют одну или несколько специфических последовательностей, массивное параллельное секвенирование не зависит от детекции или анализа предварительно определенного или предварительно идентифицированного набора последовательностей ДНК. Секвенируют произвольно представленную фракцию молекул ДНК из пула образцов. Число различных секвенированных меток, выровненных с различными участками хромосом, сравнивают между образцами, содержащими или не содержащими виды ДНК, представляющие интерес. Хромосомные aberrации будут выявляться по различиям в числе (или проценте) последовательностей, выровненных с любым заданным участком хромосомы в образцах.

В другом примере метод секвенирования внеклеточной ДНК в плазме можно использовать для детекции хромосомных aberrаций в ДНК плазмы для обнаружения специфического рака. Различные онкозаболевания имеют набор типичных хромосомных aberrаций. Можно использовать изменения (амплификации или делеции) в нескольких участках хромосомы. Так, может иметься повышенная доля последовательностей, выровненных с амплифицированными участками, и пониженная доля последовательностей, выровненных с сокращенными участками. Представленность в процентах на хромосому можно сравнить с размером для каждой соответствующей хромосомы в эталонном геноме, выраженную как процент геномной представленности любой данной хромосомы в отношении всего генома. Также можно использовать прямые сравнения или сравнения с эталонной хромосомой.

II. Секвенирование только фракции генома человека.

В эксперименте, описанном выше в примере I, ДНК материнской плазмы из каждого отдельного образца секвенируют с использованием только одной проточной кюветы. Число секвенированных меток, полученных секвенированием из каждого из испытываемых образцов, показано на фиг. 6. T21 отмечает образец, полученный от беременной женщины с плодом с трисомией 21.

Секвенируют 36 п.о. из каждого из секвенированных фрагментов ДНК материнской плазмы, число секвенированных нуклеотидов/пар оснований из каждого образца можно определить, умножая 36 п.о. на число секвенированных меток, что также показано на фиг. 6. Так как в геноме человека имеется приблизительно 3 миллиарда пар оснований, количество данных секвенирования, полученное от каждого образца материнской плазмы, представляет только часть, колеблющуюся от примерно 10 до 13%.

Более того, при таком исследовании только уникально картированные секвенированные метки, обозначенные U0 в номенклатуре из программы Eland, используют для того, чтобы показать сверхпредставленности в количестве последовательностей хромосомы 21 в образцах материнской плазмы при беременностях с плодами с трисомией 21, как описано выше в примере I. Как видно на фиг. 6, последовательности U0 представляют только подмножество всех секвенированных меток, полученных из каждого образца, и также представляют даже меньшую часть - примерно 2% генома человека. Такие данные показывают, что для успешного выполнения диагностики фетальной анеуплоидии достаточно секвенирования только части последовательностей генома человека, присутствующих в испытываемом образце.

III. Определение числа требуемых последовательностей.

Результат секвенирования ДНК плазмы от беременной женщины с эуплоидным плодом мужского пола используют для такого анализа. Число секвенированных меток, которые можно картировать к эталонной последовательности генома человека без несоответствий, составляет 1990000. Подмножества последовательностей произвольно выбирают из таких 1990000 меток и в каждом подмножестве вычисляют процент последовательностей, выровненных с хромосомой 21. Число последовательностей в подмножествах изменяется от 60000 до 540000 последовательностей. Для каждого размера подмножества составляют несколько подмножеств с тем же числом секвенированных меток случайным отбором секвенированных меток из общего пула до тех пор, пока другая комбинация станет невозможной. Затем из нескольких подмножеств в пределах подмножеств каждого размера вычисляют средний процент последовательностей, выровненных с хромосомой 21, и его стандартное отклонение (SD). Полученные данные сравнивают по всем подмножествам различного размера для определения влияния размера подмножества на распределение процента последовательностей, выровненных с хромосомой 21. Затем вычисляют 5- и 95-е процентиля согласно среднему и SD.

Когда у беременной женщины плод с трисомией 21, секвенированные метки, выровненные с хромосомой 21, должны быть сверхпредставленными в материнской плазме из-за избыточной дозы хромосомы 21 из плода. Степень сверхпредставленности зависит от процента фетальной ДНК в образце ДНК

материнской плазмы в соответствии с приведенным ниже уравнением

$$\text{Per}_{T21} = \text{Per}_{Eu} \times (1 + f/2),$$

где Per_{T21} представляет процент последовательностей, выровненных с хромосомой 21, у беременной женщины с плодом с трисомией 21; и

Per_{Eu} - процент последовательностей, выровненных с хромосомой 21, у беременной женщины с эуплоидным плодом; и

f - процент фетальной ДНК в ДНК материнской плазмы.

Как видно на фиг. 7, SD для процентов последовательностей, выровненных с хромосомой 21, уменьшается с возрастанием числа последовательностей в каждом подмножестве. Следовательно, когда число последовательностей в каждом подмножестве возрастает, интервал между 5- и 95-ми перцентилями уменьшается. Когда интервал 5-95% для случаев эуплоидии и трисомии 21 не перекрывается, тогда дифференциация между двумя группами случаев может быть возможна с точностью >95%.

Как видно на фиг. 7, минимальный размер подмножества для дифференциации случаев трисомии 21 и случаев эуплоидии зависит от процента фетальной ДНК. Минимальные размеры подмножеств для дифференциации случаев трисомии 21 и случаев эуплоидии составляют 120000, 180000 и 540000 последовательностей для процентов фетальной ДНК 20, 10 и 5% соответственно. Иными словами, число последовательностей, требуемых для анализа для определения того, имеет ли плод трисомию 21, будет составлять 120000, когда образец ДНК материнской плазмы содержит 20% фетальной ДНК. Число последовательностей, требуемых для анализа, будет возрастать до 540000, когда процент фетальной ДНК падает до 5%.

Так как данные получают с использованием секвенирования 36 п.о., 120000, 180000 и 540000 последовательностей соответствуют 0,14, 0,22 и 0,65% генома человека соответственно. Так как нижний интервал концентраций фетальной ДНК в материнской плазме, полученной на ранних сроках беременности, как сообщается, примерно 5% (Lo Y.M.D. et al., 1998, Am. J. Hum. Genet., 62, 768-775), секвенирование примерно 0,6% генома человека может представлять минимальный объем секвенирования, требуемый для диагностики с точностью по меньшей мере 95% при детекции фетальной хромосомной анеуплоидии при любых сроках беременности.

IV. Рандомизированное секвенирование.

Для того чтобы пояснить, что секвенированные фрагменты ДНК в ходе секвенирования отбирают рандомизированно, получают секвенированные метки, образованные из восьми образцов материнской плазмы, анализированных в примере I. Для каждого образца материнской плазмы определяют исходные позиции в отношении к эталонной последовательности генома человека комплекту 36 NCBI каждой из секвенированных меток в 36 п.о., которые выравнивают уникально с хромосомой 21 без несоответствий. Затем определяют число позиций для пулов выровненных секвенированных меток из каждого образца в порядке возрастания. Осуществляют подобный анализ для хромосомы 22. В целях пояснения верхние десять исходных позиций для хромосомы 21 и хромосомы 22 для каждого образца материнской плазмы показаны на фиг. 8А и 8В соответственно. Как можно представить из приведенных таблиц, секвенированные пулы фрагментов ДНК между образцами не являются идентичными.

Любые из компонентов или функций программ, описанных в данной заявке, можно выполнить в виде программного кода, выполняемого процессором с использованием любого подходящего машинного языка, такого как, например, Java, C++ или Perl, с использованием, например, традиционных или ориентированных на объект методов. Программный код может сохраняться как ряд инструкций или команд на среде, считываемой компьютером, для хранения и/или передачи, подходящие среды включают запоминающее устройство с произвольной выборкой (RAM), постоянное запоминающее устройство (ПЗУ, ROM), магнитную среду, жесткий дисковод или дискету, или оптическую среду, такую как компакт-диск (CD) или DVD (цифровой универсальный диск), флэш-память и т.п. Среда, считываемая компьютером, может представлять собой любую комбинацию таких устройств хранения или передачи информации.

Такие программы также могут кодироваться и передаваться с использованием сигналов носителя, адаптированных для передачи по проводным, оптическим и/или беспроводным сетям, соответствующим протоколам, включая Интернет. Как таковая, среда, считываемая компьютером, согласно воплощению изобретения может быть создана с использованием сигнала данных, кодированного такими программами. Среда, считываемая компьютером, кодированная программным кодом, может быть в пакете с совместимым устройством или предоставлена отдельно от других устройств (например, пересылаться через Интернет). Любая такая считываемая компьютером среда может быть размещена на или в пределах отдельного компьютерного программного продукта (например, жесткий дисковод или целая компьютерная система), и может быть представлена на или в рамках различных компьютерных программных продуктов в системе или сети. Компьютерная система может включать монитор, принтер или другой подходящий дисплей для предоставления пользователю любых результатов, указанных в данном описании.

Пример компьютерной системы показан на фиг. 9. Подсистемы, показанные на фиг. 9, соединяются между собой системной шиной 975. Показаны дополнительные подсистемы, такие как принтер 974, клавиатура 978, фиксированный диск 979, монитор 976, который соединен с адаптером дисплея 982, и другие. Периферические устройства и устройства ввода/вывода (I/O), которые соединены с контроллером

I/O, могут быть соединены с компьютерной системой любыми средствами, известными в технике, такими как последовательный порт 977. Например, для присоединения компьютера к глобальной сети, такой как Интернет, устройству ввода данных или сканеру можно использовать последовательный порт 977 или внешний интерфейс 981. Соединение через системную шину позволяет коммутировать центральный процессор 973 с каждой подсистемой и контролировать выполнение инструкций из запоминающей системы 972 или фиксированного диска 979, а также обмен информацией между подсистемами.

Запоминающая система 972 и/или фиксированный диск 979 могут составлять часть считываемой компьютером среды.

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

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

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ выполнения пренатальной диагностики фетальной хромосомной анеуплоидии в биологическом образце, полученном от беременной плодом женщины, где биологический образец включает бесклеточные молекулы нуклеиновых кислот указанного беременной женщины и плода, включающий

рандомизированное секвенирование по меньшей мере части молекул нуклеиновых кислот, содержащихся в указанном биологическом образце, для получения предопределенного количества последовательностей, которые представляют собой фракцию генома человека, и где указанная фракция представляет по меньшей мере 0,6% указанного человеческого генома;

выравнивание с помощью компьютерной системы каждой последовательности с человеческим геномом;

определение первого количества последовательностей, идентифицированных как происходящие из первой хромосомы;

определение второго количества последовательностей, идентифицированных как происходящие из одной или более вторых хромосом;

определение параметра из первого количества и второго количества, где указанный параметр определяют из отношения первого количества и второго количества, и сравнение указанного параметра с одним или более пороговыми значениями для определения классификации того, имеется ли фетальная хромосомная анеуплоидия для первой хромосомы, причем пороговое значение является эталонным значением, установленным по одному или большему количеству нормальных биологических образцов.

2. Способ по п.1, где биологический образец представляет собой материнскую кровь, плазму, сыворотку, мочу или слюну.

3. Способ по п.1, где биологический образец представляет собой трансцервикальную промывочную жидкость.

4. Способ по п.1, где первая хромосома представляет собой хромосому 21, хромосому 18, хромосому 13, хромосому X или хромосому Y.

5. Способ по п.1, где указанное отношение представляет собой подсчет фракции количества последовательностей, фракционное число секвенированных нуклеотидов или фракционную длину накопленных последовательностей.

6. Способ по п.1, в котором последовательности, выровненные с первой хромосомой, выбирают так, чтобы число пар оснований в них было менее определенного числа пар оснований.

7. Способ по п.6, в котором указанное определенное число пар оснований равно 300, 200 или 100 п.о.

8. Способ по п.1, в котором молекулы нуклеиновых кислот биологического образца обогащают последовательностями, происходящими по меньшей мере из одной определенной хромосомы.

9. Способ по п.1, в котором молекулы нуклеиновых кислот биологического образца обогащают последовательностями, которые меньше 300 п.о.

10. Способ по п.1, в котором молекулы нуклеиновых кислот биологического образца обогащают последовательностями, которые меньше 200 п.о.

11. Способ по п.1, где молекулы нуклеиновых кислот биологического образца амплифицируют с использованием полимеразной цепной реакции.

12. Способ по п.1, в котором по меньшей мере одно из пороговых значений зависит от фракционной концентрации фетальной ДНК в биологическом образце.

13. Способ по п.12, где указанную фракционную концентрацию фетальной ДНК в биологическом образце определяют с помощью одной или нескольких долей последовательностей хромосомы Y, фе-

тального эпигенетического маркера или с использованием анализа однонуклеотидного полиморфизма.

14. Способ по п.1, дополнительно включающий идентификацию количества фетальной ДНК в биологическом образце; и вычисление числа последовательностей, которое должно быть получено с учетом желательной точности и количества фетальной ДНК в биологическом образце.

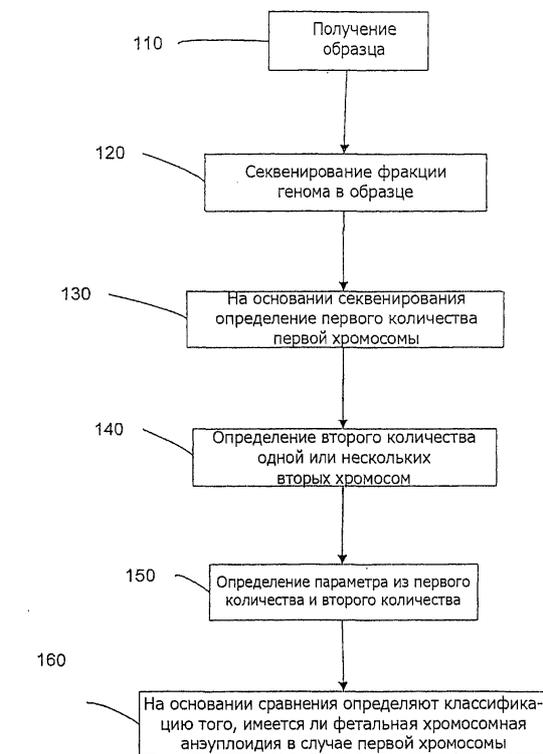
15. Способ по п.1, дополнительно включающий вычисление числа молекул нуклеиновой кислоты, которое должно быть получено с учетом желательной точности.

16. Способ по п.15, где указанная желательная точность составляет по меньшей мере 95%.

17. Способ по п.1, где указанное первое количество и указанное второе количество определено из последовательностей, идентифицированных как выровненных только с первой хромосомой и одной или более вторыми хромосомами соответственно.

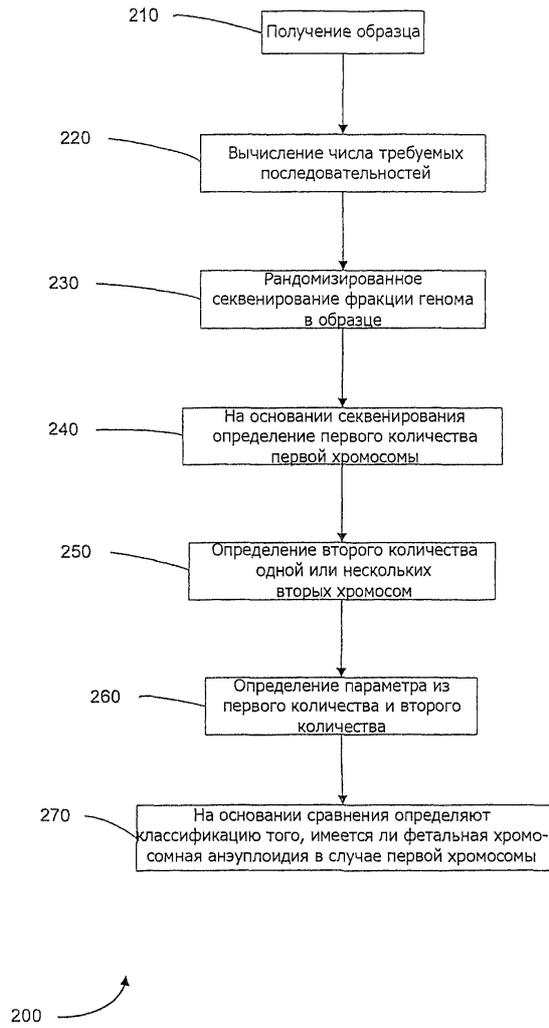
18. Способ по п.1, где указанное определенное число последовательностей составляет по меньшей мере 120000, если биологический образец имеет 20% или более фетальной ДНК, по меньшей мере 180000, если биологический образец имеет по меньшей мере 10% или более фетальной ДНК, или по меньшей мере 540000, если биологический образец имеет 5% и более фетальной ДНК.

19. Считываемая компьютером среда с программой для осуществления способа по любому из пп.1-18.

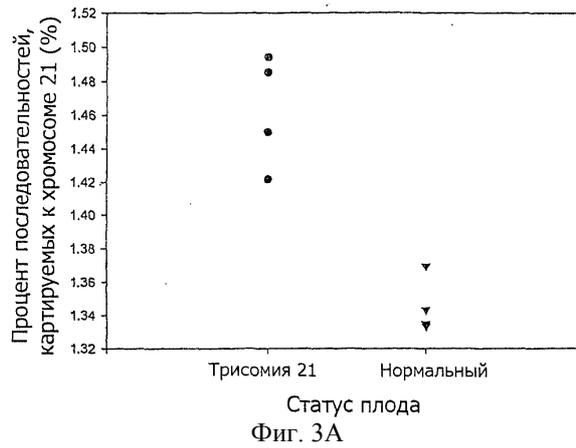


100

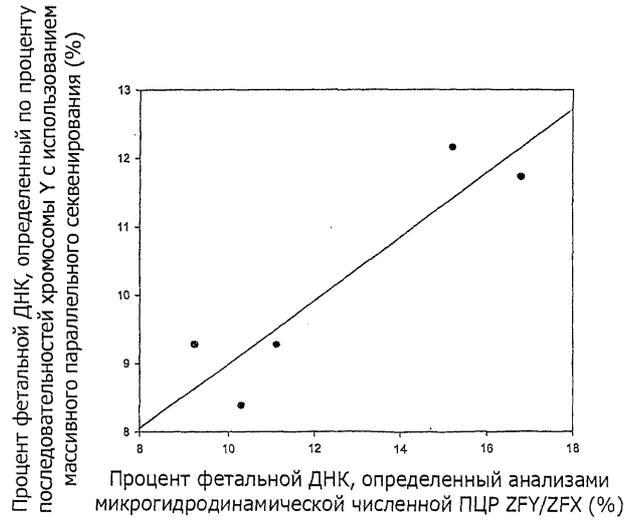
Фиг. 1



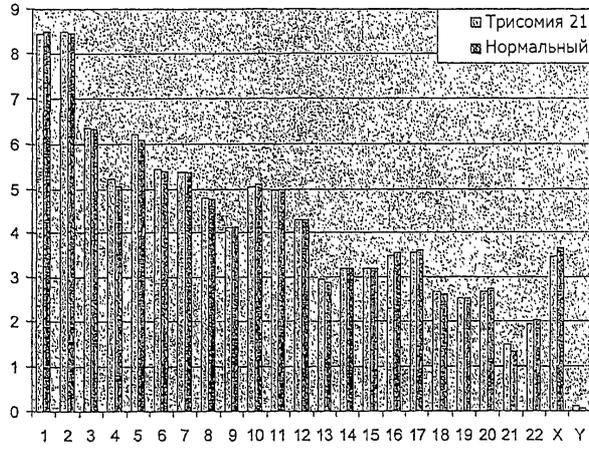
Фиг. 2



Фиг. 3А



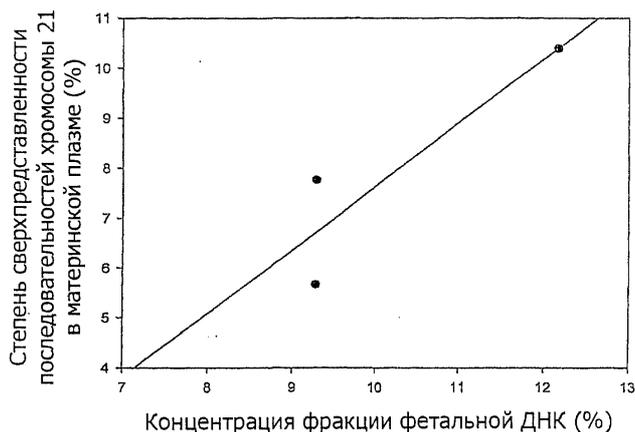
Фиг. 3В



Фиг. 4А



Фиг. 4В



Фиг. 5

Образец	Общее число секвенированных меток	Общее число секвенированных пар оснований	Доля генома человека (%)	Число УО	Число УО/общее число (%)	Число секвенированных пар оснований УО	Доля УО генома человека (%)
Образец 1 T21	1.12E+07	4.02E+08	13.4	1.93E+06	17.3	6.94E+07	2.31
Образец 2 T21	1.04E+07	3.73E+08	12.4	1.80E+06	17.4	6.50E+07	2.17
Образец 3 T21	8.90E+06	3.20E+08	10.7	2.09E+06	23.4	7.51E+07	2.50
Образец 4 T21	1.02E+07	3.69E+08	12.3	2.23E+06	21.7	8.02E+07	2.67
Образец 5 зуплоидный	1.06E+07	3.83E+08	12.8	2.12E+06	19.9	7.62E+07	2.54
Образец 6 зуплоидный	9.58E+06	3.45E+08	11.5	1.91E+06	20.0	6.88E+07	2.29
Образец 7 зуплоидный	9.55E+06	3.44E+08	11.5	2.01E+06	21.0	7.22E+07	2.41
Образец 8 зуплоидный	9.09E+06	3.27E+08	10.9	2.09E+06	23.0	7.53E+07	2.51

Фиг. 6

Число последовательностей в каждом подмножестве	Зуплоидный								T21 (5%)			T21 (10%)			T21 (20%)		
	Среднее	SD	5процентиль	95процентиль	Среднее	5процентиль	95процентиль										
60,000	1.322%	0.036%	1.250%	1.395%	1.355%	1.283%	1.428%	№	1.388%	1.316%	1.451%	№	1.454%	1.392%	1.527%	№	
110,000	1.322%	0.022%	1.277%	1.367%	1.355%	1.311%	1.400%	№	1.388%	1.344%	1.433%	№	1.454%	1.410%	1.499%	Yes	
160,000	1.322%	0.016%	1.291%	1.353%	1.355%	1.324%	1.385%	№	1.388%	1.357%	1.423%	Yes	1.454%	1.423%	1.486%	Yes	
210,000	1.322%	0.013%	1.285%	1.349%	1.355%	1.328%	1.382%	№	1.388%	1.381%	1.415%	Yes	1.454%	1.427%	1.481%	Yes	
360,000	1.322%	0.011%	1.301%	1.344%	1.355%	1.334%	1.377%	№	1.388%	1.367%	1.410%	Yes	1.454%	1.433%	1.476%	Yes	
360,000	1.322%	0.010%	1.303%	1.349%	1.355%	1.335%	1.376%	№	1.388%	1.368%	1.409%	Yes	1.454%	1.434%	1.475%	Yes	
420,000	1.322%	0.009%	1.303%	1.341%	1.355%	1.335%	1.374%	№	1.388%	1.369%	1.407%	Yes	1.454%	1.435%	1.473%	Yes	
480,000	1.322%	0.009%	1.305%	1.338%	1.355%	1.338%	1.372%	№	1.388%	1.371%	1.405%	Yes	1.454%	1.437%	1.471%	Yes	
540,000	1.322%	0.008%	1.306%	1.339%	1.355%	1.339%	1.372%	Yes	1.388%	1.372%	1.405%	Yes	1.454%	1.439%	1.471%	Yes	

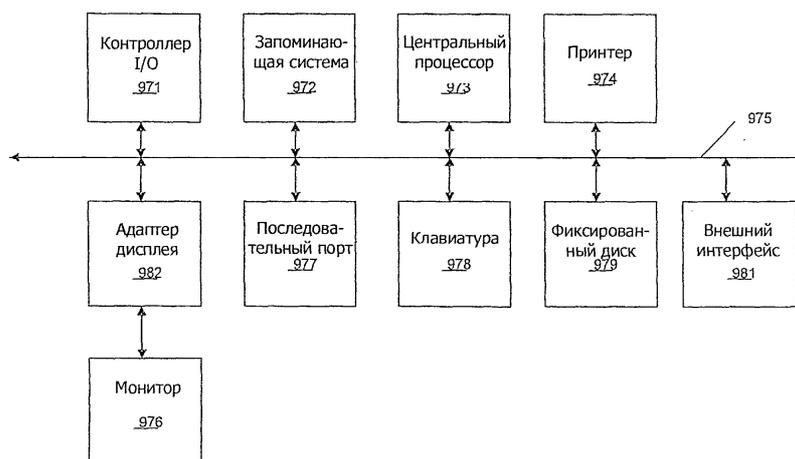
Фиг. 7

Секвенированная метка, по порядку	Образец 1 T21	Образец 2 T21	Образец 3 T21	Образец 4 T21	Образец 5 зуплоидный	Образец 6 зуплоидный	Образец 7 зуплоидный	Образец 8 зуплоидный
1	9796394	9798087	9798123	9795700	9797841	9795972	9796536	9795601
2	9797424	9798402	9798250	9797860	9798176	9796549	9796863	9797404
3	9797438	9798708	9798715	9797864	9798835	9797359	9798161	9798117
4	9798112	9799733	9799467	9798106	9800315	9797418	9798401	9798175
5	9798394	9799852	9799730	9799209	9800385	9797446	9798722	9798387
6	9798729	9800362	9799788	9799440	9800554	9797860	9799752	9798816
7	9798768	9800914	9799834	9799440	9800820	9798062	9800317	9799732
8	9798816	9801528	9800332	9799741	9800829	9798135	9800800	9800144
9	9799421	9801779	9800869	9799812	9800832	9798707	9801380	9800421
10	9799464	9803801	9800881	9799833	9800866	9798715	9801699	9800827

Фиг. 8А

Секвенированная метка, по порядку	Образец 1 T21	Образец 2 T21	Образец 3 T21	Образец 4 T21	Образец 5 эуплоидный	Образец 6 эуплоидный	Образец 7 эуплоидный	Образец 8 эуплоидный
1	14469232	14510976	14510792	14510867	14434681	14510711	14512069	14510996
2	14503781	14511307	14510807	14511312	14506660	14510835	14512407	14522947
3	14510805	14511512	14510886	14511983	14511219	14511308	14522879	14530478
4	14510824	14511755	14511816	14522914	14511328	14511354	14609732	14565255
5	14510965	14511774	14512410	14522978	14511377	14511703	14628248	14609733
6	14511369	14512067	14522944	14523028	14511432	14511703	14675246	14628261
7	14511702	14512190	14523047	14565245	14511597	14512184	14675250	14668784
8	14511738	14522891	14564769	14628257	14511755	14512204	14680184	14680163
9	14523000	14539630	14565205	14635374	14512204	14524806	14692539	14689547
10	14524823	14541260	14565251	14665101	14512204	14565230	14694114	14689592

Фиг. 8В



Фиг. 9

