

(19)



**Евразийское
патентное
ведомство**

(21) **201991550** (13) **A1**

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2019.11.29

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

(22) Дата подачи заявки
2010.11.05

(54) АНАЛИЗ ГЕНОМА ПЛОДА ИЗ МАТЕРИНСКОГО БИОЛОГИЧЕСКОГО ОБРАЗЦА

(31) 61/258,567; 61/259,075; 61/381,854

(32) 2009.11.05; 2009.11.06; 2010.09.10

(33) US

(62) 201200690; 2010.11.05

(71) Заявитель:

**ТЕ ЧАЙНИЗ ЮНИВЕРСИТИ ОВ
ГОНКОНГ (CN); СЕКЕНОМ, ИНК.
(US)**

(72) Изобретатель:

**Ло Юйк Мин Дэnnис, Чань Квань Чэ,
Чиу Вай Квунь Росса (CN), Кантор
Чарльз (US)**

(74) Представитель:

Фелицына С.Б. (RU)

(57) Представлены системы, способы и приборы для определения по меньшей мере части генома плода. Фрагменты ДНК из материнского образца (ДНК матери и плода) подвергаются анализу для идентификации аллелей по определенным локусам. Можно установить общее количество фрагментов ДНК соответствующих аллелей в этих локусах для определения относительного количества гаплотипов по этим локусам и определить, какие гаплотипы унаследованы из родительских геномов. Можно установить локусы, по которым родители представляют определенные комбинации гомозигот и гетерозигот для определения участков генома плода. Вместе с анализом фрагментов ДНК из материнского образца для определения геномов матери и отца можно использовать контрольные гаплотипы, распространенные в популяции. Также представлено определение мутаций, фракционной концентрации ДНК плода в материнском образце и степени охвата секвенированием материнского образца.

A1

201991550

201991550

A1

АНАЛИЗ ГЕНОМА ПЛОДА ИЗ МАТЕРИНСКОГО БИОЛОГИЧЕСКОГО ОБРАЗЦА

Перекрестные ссылки на родственные заявки

Настоящая заявка претендует на приоритет от и является непредварительным вариантом предварительной заявки на патент США No. 61/258567, озаглавленной “Анализ генома плода” и поданной 5 ноября 2009 г., предварительной заявки на патент США No. 61/259075, озаглавленной “Анализ генома плода из материнского биологического образца” и поданной 6 ноября 2009 г., и предварительной заявки на патент США No. 61/381854, озаглавленной “Анализ генома плода из материнского биологического образца” и поданной 10 сентября 2010 г., все содержание которых включено путем ссылки на все случаи.

Настоящая заявка также связана с заявкой на патент США No. 12/178181, озаглавленной “Диагностика хромосомной анеуплоидии у плода при помощи массивированного параллельного секвенирования генома” и поданной 5 ноября 2009 г. (Attorney Docket No. 016285-005220US); заявкой на патент США No. 12/614350, озаглавленной “Диагностика хромосомной анеуплоидии у плода при помощи секвенирования генома с обогащением” (Attorney Docket No. 016285-005221US), и поданной одновременно с ней заявкой США, озаглавленной “Анализ генома на основе размера” (Attorney Docket No. 016285-006610US), все содержание которых включено путем ссылки на все случаи.

Уровень техники

Настоящее изобретение, в общем, касается анализа генома плода на основе материнского образца, в частности, определения всего или части генома плода на основе анализа генетических фрагментов в материнском образце.

Открытие бесклеточных нуклеиновых кислот зародыша в материнской плазме в 1997 г. открыло новые возможности для неинвазивной пренатальной диагностики (Lo YMD et al., 1997, 350: 485-487; и US Patent 6,285,540). Эта технология быстро воплотилась в клинические применения с детектированием полученных плодом по отцовской линии генов или последовательностей, например, для определения пола у плода, определения статуса RhD у плода и установления того, унаследовал ли он передающуюся по отцовской линии мутацию (Amicucci P et al., Clin Chem 2000, 46: 301-302; Saito H et al., Lancet 2000, 356: 1170; и Chiu RWK et al., Lancet 2002, 360: 998-1000). Недавний прогресс в этой области сделал возможной пренатальную диагностику хромосомной анеуплоидии у плода

типа трисомии 21 из анализа нуклеиновых кислот в материнской плазме (Lo YMD et al., Nat Med 2007, 13: 218-223; Tong YK et al., Clin Chem 2006, 52: 2194-2202; US Patent Publication 2006/0252071; Lo YMD et al., Proc Natl Acad Sci USA 2007, 104: 13116-13121; Chiu RWK et al., Proc Natl Acad Sci USA 2008, 105: 20458-20463; Fan HC et al., Proc Natl Acad Sci 2008, 105: 16266-16271; US Patent Publication 2007/0202525; и US Patent Publication 2009/0029377).

Другая область значительного прогресса последнего времени – это применение способов подсчета отдельных молекул типа цифровой ПЦР для неинвазивной пренатальной диагностики моногенных болезней, при которых и мать, и отец несут одну и ту же мутацию. Это достигается анализом относительной дозы мутации (RMD) в материнской плазме (US Patent Application 2009/0087847; Lun FMF et al., Proc Natl Acad Sci USA 2008, 105: 19920-19925; и Chiu RWK et al., Trends Genet 2009, 25: 324-331).

Однако в таких методах используются уже существующие знания о возможных мутациях для анализа конкретных частей генома, поэтому ими нельзя идентифицировать латентные или редкие мутации или генетические заболевания. Следовательно, нужно предоставить новые методы, системы и приборы, которыми можно идентифицировать весь или часть генома у плода при помощи неинвазивных методов.

Сущность изобретения

В некоторых воплощениях настоящего изобретения предусмотрены способы, системы и приборы для определения по меньшей мере части генома еще не рожденного плода у беременной матери. Можно составить генетическую карту всего генома или отдельных участков генома плода до рождения, используя образцы, содержащие генетический материал плода и матери (например, образец крови беременной матери). Генетическая карта может содержать последовательности, которые плод унаследовал от обоих родителей, или же только последовательности от одного из родителей. Исходя из одной или нескольких таких генетических карт, можно определить риск возникновения у плода генетического заболевания или предрасположенности к генетическому или другому заболеванию или генетической особенности. Также описаны и другие применения воплощений.

В одном воплощении можно подвергнуть анализу фрагменты ДНК из материнского образца (содержащего ДНК матери и плода) для идентификации аллелей по некоторым заданным локусам (ориентирам). Затем можно проанализировать общее количество фрагментов ДНК соответствующих аллелей в этих локусах для определения относительного количества гаплотипов по этим локусам и тем самым определить, какие гаплотипы унаследованы плодом из материнского и/или отцовского генома. Идентифицируя

гаплотипы плода, можно определить генотип плода по индивидуальным локусам в соответствующем участке генома, включая заданные локусы. В различных воплощениях локусы, по которым родители представлены определенными комбинациями гомозиготных и гетерозиготных, можно подвергнуть анализу таким образом, чтобы определить участки генома плода. В одном воплощении для определения материнского и отцовского геномов используются контрольные гаплотипы, представляющие распространенные в популяции гаплотипы, вместе с анализом фрагментов ДНК материнского образца. Также предусмотрены и другие воплощения, как-то определение мутаций, определение доли плода в образце матери и определение охвата секвенированием материнского образца.

Другие воплощения изобретения касаются систем, приборов и считываемых компьютером носителей, связанных с описанными здесь способами. В одном воплощении считываемый компьютером носитель содержит инструкции для получения данных и анализа данных, но не инструкции для направления машины на создание данных (например, секвенирование молекул нуклеиновых кислот). А в другом воплощении считываемый компьютером носитель содержит инструкции для направления машины на создание данных. В одном воплощении компьютерный программный продукт включает считываемый компьютером носитель, хранящий множество инструкций для управления процессором при выполнении операций для описанных здесь способов. Воплощения также касаются компьютерных систем, конфигурированных на выполнение стадий любого из описанных здесь способов, необязательно с различными компонентами, выполняющими соответствующую стадию или соответствующую группу стадий.

При обращении к остальным частям описания, включая рисунки и формулу изобретения, будут реализованы другие особенности и преимущества воплощений настоящего изобретения. Дополнительные особенности и преимущества, а также структура и действие различных воплощений настоящего изобретения подробно описаны ниже с привлечением сопроводительных рисунков. На рисунках одинаковые номера обозначений могут означать идентичные или функционально близкие элементы.

Краткое описание фигур

На фиг. 1 представлена блок-схема способа определения 100 по меньшей мере части генома еще не рожденного плода беременной женщины согласно воплощениям настоящего изобретения.

На фиг. 2 представлены два гаплотипа у отца и два гаплотипа у матери для определенного сегмента соответствующего им геномного кода согласно воплощениям настоящего изобретения.

На фиг. 3 представлены два типа SNP у родительских гаплотипов из фиг. 2

согласно воплощениям настоящего изобретения.

На фиг. 4А и фиг.4В представлен анализ для определения гаплотипов плода для двух типов SNP согласно воплощениям настоящего изобретения.

На фиг. 5А и фиг. 5В представлен сравнительный анализ относительного количества (например, встречаемости) фрагментов по каждому локусу и того, что результат сравнения классифицирует определенный генотип как наследуемый или нет, согласно воплощениям настоящего изобретения.

На фиг. 6 представлен эффект изменения отношения вероятностей при классификации по SPRT согласно воплощениям настоящего изобретения.

На фиг. 7 представлена блок-схема способа определения 700 по меньшей мере части генома, унаследованной от отца, у еще не рожденного плода беременной женщины согласно воплощениям настоящего изобретения.

На фиг.8 представлена блок-схема способа определения 800 по меньшей мере части генома у еще не рожденного плода беременной женщины на участке, в котором отец и мать гетерозиготны, согласно воплощениям настоящего изобретения.

На фиг. 9 представлены гаплотипы отца и матери, которые оба гетерозиготны на определенном участке генома согласно воплощениям настоящего изобретения.

На фиг. 10 представлена блок-схема способа определения 1000 фракционной концентрации материала плода в материнском образце согласно воплощениям настоящего изобретения.

На фиг. 11 представлена блок-схема способа определения того, что локус является информативным согласно воплощениям настоящего изобретения.

На фиг. 12А и 12В представлено прогнозируемое распределение встречаемости для аллеля Т (наименее распространенный аллель в сценариях (а) и (с)) по трем сценариям в предположении, что фракционная концентрация ДНК плода составляет 20% и 5%, соответственно, согласно воплощениям настоящего изобретения.

На фиг. 13А, 13В и 14 представлено прогнозируемое распределение встречаемости для наименее распространенного аллеля при фракционной концентрации ДНК плода в 20% при различном общем числе молекул, соответствующих SNP, согласно воплощениям настоящего изобретения.

На фиг. 15А и 15В представлены примеры контрольных гаплотипов, родительских гаплотипов, полученных от контрольных гаплотипов, и возникающие при этом гаплотипы плода согласно воплощениям настоящего изобретения.

На фиг. 16 представлена блок-схема способа определения 1600 по меньшей мере части генома плода, когда известен набор контрольных гаплотипов, но не известны

родительские гаплотипы, согласно воплощениям настоящего изобретения.

На фиг 17 представлен пример определения информативных локусов при анализе фрагментов ДНК из материнского образца согласно воплощениям настоящего изобретения.

На фиг. 18 представлены три контрольных гаплотипа (Нар А, Нар В и Нар С) и отцовские аллели.

На фиг. 19 представлено определение родительского гаплотипа по отцовским аллелям согласно воплощениям настоящего изобретения.

На фиг. 20 представлено вычисление материнских генотипов при анализе материнского образца согласно воплощениям настоящего изобретения.

На фиг. 21 представлено воплощение для определения материнских гаплотипов из материнских генотипов и контрольных гаплотипов согласно воплощениям настоящего изобретения.

На фиг. 22 представлены материнские гаплотипы и наследуемые по отцовской линии гаплотипы, определенные согласно воплощениям настоящего изобретения.

На фиг. 23 представлены различные типы локусов (альфа (А) и бета (В)) для материнских гаплотипов относительно отцовского гаплотипа согласно воплощениям настоящего изобретения.

На фиг. 24 представлена блок-схема, иллюстрирующая способ идентификации 2400 мутаций de novo в геноме еще не рожденного плода у беременной женщины.

На фиг. 25А представлено абсолютное количество и процентное содержание SNP, представляющих различные комбинации генотипов для отца, матери и плода (CVS) согласно воплощениям настоящего изобретения.

На фиг. 25В представлена таблица, в которой приведена статистика совмещения по первым 20 проточным кюветам.

На фиг. 26 представлена таблица, в которой приведены фракционные концентрации ДНК плода, рассчитанные по SNP двумя способами согласно воплощениям настоящего изобретения.

На фиг. 27А представлен график, иллюстрирующий наблюдаемый процент SNP в той подгруппе, в которой аллель плода наблюдается по данным секвенирования при анализе первых 20 проточных кювет, а на фиг. 27В представлен график охвата от количества прочитанных отрезков согласно воплощениям настоящего изобретения.

На фиг. 28А и 28В представлены графики корреляции между охватом наследуемых по отцовской линии аллелей и числом картируемых прочтений последовательности и количеством секвенируемых проточных кювет, соответственно, согласно воплощениям

настоящего изобретения.

На фиг. 29А представлена корреляция между долей ложно-положительных и количеством секвенируемых проточных кювет, а на фиг. 29В представлена корреляция между долей ложно-положительных и количеством секвенируемых проточных кювет согласно воплощениям настоящего изобретения.

На фиг. 30 представлен охват специфичных для плода SNP при различном числе анализируемых проточных кювет согласно воплощениям настоящего изобретения.

На фиг. 31 представлена точность анализа типа А при использовании данных от 10 проточных кювет согласно воплощениям настоящего изобретения.

На фиг. 32 представлена точность анализа типа В при использовании данных от 10 проточных кювет согласно воплощениям настоящего изобретения.

На фиг. 33 представлена точность анализа типа А при использовании данных от 20 проточных кювет согласно воплощениям настоящего изобретения.

На фиг. 34 представлена точность анализа типа В при использовании данных от 20 проточных кювет согласно воплощениям настоящего изобретения.

На фиг. 35 А и 35В представлены отрезки последовательностей с мутациями и отрезки дикого типа по кодонам 41/42 согласно воплощениям настоящего изобретения.

На фиг. 36 представлена таблица анализа RHDO типа А, а на фиг. 37 представлен анализ RHDO типа В согласно воплощениям настоящего изобретения.

На фиг. 38А и 38В представлены результаты классификации по SPRT для случая PW226 в качестве примера.

На фиг. 39 представлена таблица, суммирующая результаты анализа RHDO для пяти случаев согласно воплощениям настоящего изобретения.

На фиг. 40 представлен график полноты секвенирования от числа секвенируемых проточных кювет согласно воплощениям настоящего изобретения.

На фиг. 41 представлен график размера последовательностей плода и общих последовательностей для всего генома, а на фиг. 42А-42С представлены аналогичные графики индивидуально для каждой хромосомы согласно воплощениям настоящего изобретения.

На фиг. 43 представлена блок-схема примера компьютерной системы 4300, применяемой с системой и способами согласно воплощениям настоящего изобретения.

Определения

Термин “биологический образец” в настоящем изобретении относится к любым образцам, взятым у субъекта (например, человека, как-то беременной женщины) и содержащим одну или несколько представляющих интерес молекул нуклеиновой кислоты.

Термин “нуклеиновая кислота” или “полинуклеотид” относится к дезоксирибонуклеиновой кислоте (ДНК) или рибонуклеиновой кислоте (РНК) и их полимерам как в одноцепочечном, так и двухцепочечном виде. Если не оговорено особо, термин охватывает нуклеиновые кислоты, содержащие известные аналоги природных нуклеотидов, обладающие такими же свойствами связывания, как и стандартная нуклеиновая кислота, и подвергающиеся метаболизму таким же образом, как и природные нуклеотиды. Если не указано иначе, конкретная последовательность нуклеиновой кислоты также косвенным образом охватывает её консервативно модифицированные варианты (например, замены на вырожденные кодоны), аллели, ортологи, SNPs и комплементарные последовательности, а также последовательности, указанные в явном виде. В частности, замены на вырожденные кодоны могут осуществляться путем создания последовательностей, в которых у одного или несколько отдельных (либо всех) кодонов в третьем положении произведены замены на смешанные основания и/или остатки дезоксинозина (Batzner 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 применяется взаимозаменяемо с геном, кДНК, мРНК, небольшими некодирующими РНК, микроРНК (miRNA), взаимодействующими с Рiwi-РНК и короткими шпилечными РНК (shRNA), кодируемыми геном или локусом.

Термин “ген” обозначает сегмент ДНК, участвующий в вырабатывании полипептидной цепи или транскрибируемого продукта РНК. Он может включать участки, предшествующие или последующие за кодирующим участком (лидерные или трейлерные), а также промежуточные последовательности (интроны) между отдельными кодирующими сегментами (экзонами).

Термин “клинически значимая последовательность нуклеиновой кислоты” (также именуется последовательностью мишени или хромосомой) в настоящем изобретении может обозначать полинуклеотидную последовательность, соответствующую сегменту большей геномной последовательности, возможный дисбаланс которой проверяется, или самой этой большей 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 от обоих родителей или в качестве мутации *de novo* у плода. В некоторых воплощениях для получения данных для выявления дисбаланса может использоваться множество клинически значимых последовательностей нуклеиновых кислот или же множество маркеров клинически значимой последовательности

нуклеиновой кислоты. Например, для определения возможного нарушения в хромосоме 21 можно аддитивно использовать данные из пяти несмежных последовательностей на хромосоме 21, при этом требуемый объем образца эффективно снижается до 1/5.

Термин “на основе” в настоящем изобретении означает “по крайней мере частично на основе” и относится к одной величине (или результату), которая используется для определения другой величины, как это происходит на входе способа и на выходе этого способа. Термин “выводить” в настоящем изобретении также относится к взаимоотношению между входными данными способа и результатом этого способа, как это происходит тогда, когда выведение означает вычисление по формуле.

Термин “параметр” в настоящем изобретении обозначает численное значение, характеризующее набор количественных данных и/или количественное взаимоотношение между наборами числовых данных. Например, параметром является соотношение (или функция отношения) между первым количеством первой последовательности нуклеиновой кислоты и вторым количеством второй последовательности нуклеиновой кислоты.

В настоящем изобретении термин “локус” или его множественная форма “локусы” обозначает расположение или адрес любого отрезка нуклеотидов (или пар оснований), имеющего вариации между геномами.

Термин “дисбаланс последовательности” в настоящем изобретении означает любое значительное отклонение, определяемое по меньшей мере одним значением отсечения, по количеству клинически значимой последовательности нуклеиновой кислоты от стандартного количества. Дисбаланс последовательности может включать хромосомный дисбаланс, аллельный дисбаланс, мутационный дисбаланс, дисбаланс по гаплотипу и другие аналогичные дисбалансы. В качестве примера: аллельный или мутационный дисбаланс происходит тогда, когда генотип плода отличается от материнского, при этом возникает дисбаланс по определенному локусу в образце.

Термин “хромосомная анеуплоидия” в настоящем изобретении означает изменение числа хромосом от диплоидного набора в геноме. Изменение может быть приобретением или потерей. Оно может включать целую хромосому или участок хромосомы.

Термин “гаплотип” в настоящем изобретении относится к комбинации аллелей по множественным локусам, которые передаются вместе на одной и той же хромосоме или участке хромосомы. Гаплотип может относиться всего лишь к одной паре локусов или к участку хромосомы либо ко всей хромосоме. Термин “аллели” относится к альтернативным последовательностям ДНК в одном и том же физическом локусе генома, которые могут или не могут приводить к различным фенотипическим признакам. В каком-либо определенном диплоидном организме с двумя копиями каждой хромосомы (за

исключением половых хромосом у мужчин) генотип для каждого гена содержит пару аллелей, присутствующих в таком локусе, которые являются одинаковыми у гомозигот и разными у гетерозигот. Популяция организмов или вид обычно содержит несколько аллелей в каждом локусе у различных индивидов. Геномный локус, по которому в популяции обнаруживается более одного аллеля, называется полиморфным сайтом. Аллельные вариации в локусе измеряются числом аллелей (т.е. степенью полиморфизма) или пропорцией гетерозигот (т.е. гетерозиготностью) в популяции. В настоящем изобретении термин “полиморфизм” относится к любым вариациям между особями в геноме человека, независимо от их частоты. Примеры таких вариаций включают, без ограничения, полиморфизм по одному нуклеотиду, полиморфизмы типа простых tandemных повторов, полиморфизмы типа вставки-делеции, мутации (которые могут вызывать болезни) и вариации по числу копий.

Раскрытие сущности изобретения

Составление частичной генетической карты или полной геномной последовательности еще не рожденного плода может осуществляться на основе гаплотипов полиморфных последовательностей его родителей. Термин “гаплотип” в настоящем изобретении обозначает комбинацию аллелей по множественным локусам, которые переносятся вместе на одну и ту же хромосому или участок хромосомы. Например, в воплощениях можно подвергнуть анализу фрагменты ДНК из материнского образца (содержащие ДНК матери и плода) для идентификации аллелей по некоторым заданным локусам (маркерам). Затем можно проанализировать общее количество фрагментов ДНК соответствующих аллелей в этих локусах для определения относительного количества гаплотипов по этим локусам и тем самым определить, какие гаплотипы унаследованы плодом из материнского и/или отцовского генома. Идентифицируя гаплотипы плода, можно определить генотип плода по индивидуальным локусам в соответствующем участке генома, включая заданные локусы. В различных воплощениях локусы, по которым родители представлены определенными комбинациями гомозигот и гетерозигот, можно подвергнуть анализу таким образом, чтобы определить участки генома плода. В одном воплощении для определения материнского и отцовского геномов используются контрольные гаплотипы, представляющие распространенные в популяции гаплотипы, вместе с анализом фрагментов ДНК материнского образца.

Примером применения одного воплощения для определения по меньшей мере части генома плода может служить определение отцовства путем сравнения выведенного генотипа или гаплотипа плода с генотипом или гаплотипом предполагаемого отца. Другим примером является выявление одной или нескольких приобретенных зародышем

мутаций *de novo* или выявление случаев мейотической рекомбинации, произошедших во время образования гамет у его родителей. Это те гаметы, которые подверглись оплодотворению, а из полученной зиготы развился плод.

Кроме того, некоторые воплощения также дают возможность определить геномную последовательность еще не рожденного плода с требуемым разрешением. Например, в некоторых применениях воплощения можно определить полную или близкую к полной геномную последовательность еще не рожденного плода. В одном воплощении разрешение определяемой геномной последовательности плода зависит от знания геномов отца и матери вместе с информацией по секвенированию материнского биологического образца, содержащего нуклеиновые кислоты плода. В том случае, когда известны полные или почти полные геномные последовательности отца и матери, можно вывести полную или почти полную геномную последовательность еще не рожденного плода.

В других воплощениях выясняются только геномные последовательности отдельных участков генома, например, для пренатальной диагностики выбранных генетических, эпигенетических (таких как нарушения импринтинга) или хромосомных нарушений. Примеры генетических нарушений, к которым может применяться воплощение, включают гемоглобинопатии (как-то бета-талассемия, альфа-талассемия, серповидно-клеточная анемия, болезнь гемоглобина E), кистозный фиброз и связанные с полом заболевания (как-то гемофилия и мышечная дистрофия Дюшенна). Дополнительные примеры мутаций, которые можно определить при помощи воплощения, можно найти в Online Mendelian Inheritance in Man (www.ncbi.nlm.nih.gov/omim/getmorbidity.cgi).

Некоторые воплощения также могут применяться для определения фракционной концентрации ДНК плода, которое может выполняться без предварительного знания конкретных геномов родителей. Аналогичный анализ также может применяться для определения полноты охвата, необходимой для точного определения генома плода. Так, определение охвата может применяться для оценки того, сколько нужно проанализировать данных, чтобы получить точные результаты.

I. Введение

Когда в качестве материала для выяснения гаплотипа плода используется материнский образец (например, плазма или сыворотка), могут возникнуть две основные проблемы. Первая проблема состоит в том, что материнская плазма или сыворотка состоит из смеси ДНК плода и матери, причем ДНК плода составляет незначительную популяцию. Установлено, что средняя/медианная доля ДНК плода составляет от 5% до 10% от общей ДНК в материнской плазме в первые два триместра беременности (Lo YMD

et al., *Am J Hum Genet* 1998, 62: 768-775; Lun FMF et al., *Clin Chem* 2008, 54: 1664-1672). Поскольку ДНК высвобождается материнскими клетками крови в процессе свертывания крови, фракционная концентрация ДНК плода в материнской сыворотке может быть даже меньше, чем в материнской плазме. Поэтому в некоторых воплощениях материнская плазма предпочтительнее, чем материнская сыворотка.

Вторая проблема заключается в том, что ДНК плода и ДНК матери в материнской плазме состоят из коротких фрагментов (Chan KCA et al., *Clin Chem* 2004, 50: 88-92). Так, в материнской плазме ДНК, происходящая от плода, обычно короче, чем ДНК, происходящая от матери. Большая часть ДНК плода в материнской плазме имеет длину менее 200 п.о. При использовании таких коротких фрагментов ДНК в плазме самих по себе будет трудно составить гаплотип генетических полиморфизмов на длинных отрезках генома. Вышеуказанные проблемы в отношении материнской плазмы и сыворотки также применимы и к выявлению ДНК плода в моче матери (Botezatu I et al., *Clin Chem* 2000, 46: 1078-1084). ДНК плода составляет лишь небольшую долю ДНК в моче беременной женщины, и ДНК плода в моче матери тоже состоит из коротких фрагментов ДНК.

А. Секвенирование и анализ материнского образца

Подход, предпринятый в некоторых воплощениях для того, чтобы справиться с первой проблемой, состоит в применении способа, позволяющего с высокой точностью количественно генотипировать нуклеиновые кислоты, полученные из материнского биологического образца. В одном воплощении такого подхода точность достигается путем анализа большого числа (к примеру, миллионов или миллиардов) молекул нуклеиновой кислоты. Более того, точность можно повысить путем анализа отдельных молекул нуклеиновой кислоты или клональной амплификации отдельных молекул нуклеиновой кислоты. В одном воплощении применяется массивированное параллельное секвенирование ДНК типа того, что выполняется на платформе Illumina Genome Analyzer (Bentley DR et al., *Nature* 2008, 456: 53-59), платформе Roche 454 (Margulies M et al., *Nature* 2005, 437: 376-380), платформе ABI SOLiD (McKernan KJ et al., *Genome Res* 2009, 19: 1527-1541), платформе секвенирования отдельных молекул Helicos (Harris TD et al., *Science* 2008, 320: 106-109), секвенирования в реальном времени с использованием отдельных молекул полимеразы (*Science* 2009, 323: 133-138) и нанопорового секвенирования (Clarke J et al., *Nat Nanotechnol.* 2009, 4: 265-70). В одном воплощении массивированное параллельное секвенирование выполняется на случайном подмножестве молекул нуклеиновой кислоты в биологическом образце.

В некоторых воплощениях может оказаться выгодным считывание как можно более длинной последовательности с каждой молекулы. Одним из ограничений длины

считывания последовательности может быть природа молекул нуклеиновой кислоты в материнском биологическом образце. Например, известно, что большинство молекул ДНК в материнской плазме состоят из коротких фрагментов (Chan CA et al., Clin Chem 2004, 50: 88-92). Более того, длина считывания должна уравниваться надежностью системы секвенирования при большой длине считывания. Для некоторых из вышеприведенных систем может оказаться предпочтительным секвенирование с обоих концов молекулы, так называемое секвенирование по спаренным концам. В качестве примера: один из подходов состоит в том, чтобы просеквенировать по 50 п.о. с каждого конца молекулы ДНК, получая в итоге последовательность из 100 п.о. на молекулу. В другом воплощении секвенируют по 75 п.о. с каждого конца молекулы ДНК, получая в итоге последовательность из 150 п.о. на молекулу.

После выполнения секвенирования полученные последовательности выравниваются с контрольным геномом человека. Поскольку воплощения выявляют геномные вариации, унаследованные еще не рожденным плодом от родителей, то алгоритм выравнивания должен учитывать вариации последовательностей. Одним из примеров такого пакета программного обеспечения является программа Efficient Large-Scale Alignment of Nucleotide Databases (ELAND) фирмы Illumina. Другим примером такого пакета программного обеспечения являются программы SOAP (short oligonucleotide alignment program, т.е. программа выравнивания коротких олигонуклеотидов) и SOAP2 (Li R et al., Bioinformatics 2008, 24:713-714; Li R et al., Bioinformatics 2009, 25:1966-1967).

Степень секвенирования ДНК, которая может потребоваться для этого, может зависеть от разрешения, с которым нужно составить генетическую карту плода или геномную последовательность плода. В общем, чем больше молекул подвергается секвенированию, тем выше разрешение. Другим определяющим фактором разрешения генетической карты плода или геномной последовательности плода при данном уровне или полноте секвенирования ДНК является фракционная концентрация ДНК плода в материнском биологическом образце. В общем, чем больше фракционная концентрация ДНК плода, тем выше разрешение генетической карты плода или геномной последовательности плода, которое может быть получено при данном уровне секвенирования ДНК. Поскольку фракционная концентрация ДНК плода в материнской плазме выше, чем в материнской сыворотке, то для некоторых воплощений материнская плазма является более предпочтительным типом материнского биологического образца, чем материнская сыворотка.

Производительность вышеуказанных методов на основе секвенирования может быть повышена при помощи индексирования или штрих-кодирования. Так, к фрагментам

нуклеиновых кислот в определенной библиотеке нуклеиновых кислот для секвенирования можно добавить специфичный для образца или пациента индекс или штрих-код. Затем можно смешать и просеквенировать вместе несколько таких библиотек, каждая со своим специфичным для образца или пациента индексом или штрих-кодом. После реакции секвенирования можно извлечь данные по секвенированию от каждого образца или пациента на основании штрих-кода или индекса. Такая стратегия может повысить производительность и тем самым рентабельность воплощений настоящего изобретения.

В одном воплощении молекулы нуклеиновой кислоты в биологическом образце можно подвергнуть отбору или фракционировать еще до количественного генотипирования (например, секвенирования). В одном варианте молекулы нуклеиновой кислоты обрабатываются с помощью устройства (например, микроматрицы), которое может предпочтительно связывать молекулы нуклеиновой кислоты из выбранных локусов в геноме (например, участка на хромосоме 7, содержащего ген CFTR). После этого секвенирование может выполняться предпочтительно на молекулах нуклеиновой кислоты, захваченных устройством. Такая схема позволяет нацелить секвенирование на представляющий интерес участок генома. В одном воплощении такой схемы можно использовать систему захвата последовательности Nimblegen (www.nimblegen.com/products/seqcap/index.html) либо систему обогащения мишени Agilent SureSelect Target Enrichment System (www.opengenomics.com/SureSelect_Target_Enrichment_System) или похожие платформы. В некоторых воплощениях молекулы нуклеиновой кислоты из выбранных участков генома подвергаются рандомизированному секвенированию.

В другом воплощении представляющий интерес участок генома в биологическом образце сначала может подвергаться амплификации при помощи одного набора или нескольких наборов амплификационных праймеров. После этого может проводиться количественное генотипирование, к примеру, секвенирование продуктов амплификации. В одном исполнении такой схемы может использоваться система RainDance (www.raindancetech.com/technology/pcr-genomics-research.asp). В некоторых воплощениях амплифицированные молекулы нуклеиновой кислоты подвергаются рандомизированному секвенированию.

Молекулы нуклеиновой кислоты в биологическом образце также могут подвергаться фракционированию по размерам. Поскольку известно, что ДНК плода короче ДНК матери в материнском образце (Li et al., Clin Chem 2004, 50: 1002-1011; US Patent Application 20050164241; US Patent Application 20070202525), можно отобрать и затем использовать для количественного генотипирования, например, секвенирования, фракцию более мелких по размерам молекул. Такая фракция будет содержать большую долю ДНК

плода по сравнению с исходным биологическим образцом. Так, секвенирование фракции, обогащенной ДНК плода, дает возможность составить генетическую карту плода или вывести геномную последовательность с более высоким разрешением при заданном уровне анализа (например, полноте секвенирования), чем при использовании не обогащенного образца. При этом технология станет более рентабельной. В качестве примера способов фракционирования по размерам можно использовать: (i) гель-электрофорез с последующим экстрагированием молекул нуклеиновой кислоты из определенных частей геля; (ii) связывающие нуклеиновые кислоты матрицы с дифференциальным сродством к молекулам нуклеиновых кислот разного размера; или (iii) системы фильтрации с дифференциальным удержанием молекул нуклеиновых кислот разных размеров.

А в другом воплощении предпочтительно подвергаются анализу молекулы нуклеиновой кислоты определенного размера или диапазона размеров после секвенирования нуклеиновой кислоты. Например, можно проводить секвенирование по спаренным концам, при котором секвенирование происходит с обоих концов молекулы ДНК. Затем можно картировать геномные координаты обоих этих концов по контрольному геному человека. После этого можно вычислить размер молекулы вычитанием геномных координат двух концов. Один из способов проведения такого секвенирования по спаренным концам состоит в применении протокола секвенирования по спаренным концам Illumina Genome Analyzer. Другой способ выведения размера молекулы ДНК заключается в секвенировании всей молекулы ДНК. Проще всего это делается при помощи секвенирующих платформ с относительно большой протяженностью считывания типа платформы Roche 454 (Marguelis et al., Nature 2005, 437:376-380) и технологии отдельных молекул в реальном времени (SMRT™) фирмы Pacific Biosciences (Eid et al., Science 2009, 323: 133-138). После вычисления размера молекул нуклеиновой кислоты можно сосредоточиться на последующем анализе молекул меньше определенного предела отсечения, при этом увеличивается фракционная концентрация ДНК плода. Анализ этого подмножества молекул позволяет составить генетическую карту плода или геномные последовательности плода с меньшим количеством анализируемых молекул после отбора по размерам, чем без такой процедуры. В одном воплощении используется предел отсечения в 300 п.о. В других воплощениях может использоваться предел отсечения в 250 п.о., 200 п.о., 180 п.о., 150 п.о., 125 п.о., 100 п.о. или 75 п.о.

В. Использование родительских геномов в качестве каркаса

Для решения второй проблемы в некоторых воплощениях гаплотипы хромосом матери могут использоваться в качестве “каркаса”. А гаплотипы хромосом отца могут

использоваться в качестве другого “каркаса”. Этот каркас можно сравнить с генетической информацией плода, полученной из материнского образца, содержащего ДНК плода. Такая генетическая информация плода может использоваться для определения того, как каркас матери и/или отца был развернут в геноме плода, используя при этом составные части каркаса для определения возникающего генома плода.

Родительские гаплотипы могут быть составлены из геномной ДНК отца и матери и других членов семьи, например, брата или сестры плода в текущей беременности. Возможно, что доступность родительских гаплотипов станет все более обычным явлением ввиду снижения стоимости секвенирования генома. В одном сценарии, если геномы одного или обоих родителей уже просеквенированы и их гаплотипы на одном или нескольких участках хромосом определены, то эта информация может использоваться в качестве вышеуказанного каркаса.

Можно использовать любые платформы генотипирования, известные в данной области, которые способны выявлять вариации последовательностей в геноме, включая секвенирование ДНК, микроматрицы, гибридизационные зонды, флуоресцентные методы, оптические методы, молекулярные штрих-коды и формирование изображений отдельных молекул (Geiss GK et al., *Nat Biotechnol* 2008, 26: 317-325), анализ отдельных молекул, ПЦР, цифровую ПЦР, масс-спектрометрию (типа платформы Sequenom MassARRAY) и др. В качестве экстремального примера: можно определить последовательность ДНК отца и матери путем секвенирования всего генома методом массивированного параллельного секвенирования (например, Bentley DR et al., *Nature* 2008, 456: 53-59; McKernan KJ et al., *Genome Res* 2009, 19: 1527-1541). Примером вариаций последовательности, которые могут представлять интерес, являются полиморфизмы по отдельным нуклеотидам (SNP). Особенно предпочтительным способом определения родительских генотипов является анализ SNP на микроматрицах по всему геному или по отдельным участкам генома, например, содержащим гены, мутации которых могут вызывать генетические заболевания (как-то гены в бета-глобиновом кластере или вызывающий кистозный фиброз ген регулятора трансмембранной проводимости (CFTR)). Помимо вариаций последовательности, могут использоваться и вариации числа копий. Вариации последовательности и вариации числа копий именуется полиморфными генетическими признаками (PMF).

В одном аспекте материнские генотипы на представляющих интерес хромосомах или участках хромосом могут быть составлены в гаплотипы. Это может осуществляться путем анализа других членов семейства, родственников матери, например, сына или дочери матери, родителей, братьев и сестер и т.п. Гаплотипы можно составить и другими

методами, которые хорошо известны в данной области и приведены выше.

Затем информацию по генотипам можно распространить на информацию по гаплотипам родителей путем сравнения информации по генотипам с другими членами семьи, к примеру, братьями и сестрами плода текущей беременности или по генотипам дедушек и бабушек и т.д. Гаплотипы родителей можно составить и другими методами, хорошо известными специалистам. Примеры таких методов включают методы на основе анализа отдельных молекул, как-то метод цифровой ПЦР (Ding C and Cantor CR, *Proc Natl Acad Sci USA* 2003, 100: 7449-7453; Ruano G et al., *Proc Natl Acad Sci USA* 1990, 87: 6296-6300), гаплотипирование спермы (Lien S et al., *Curr Protoc Hum Genet* 2002, Chapter 1 :Unit 1.6) и методы формирования изображений (Xiao M et al., *Hum Mutat* 2007, 28: 913-921). Другие методы включают методы на основе аллеле-специфичной ПЦР (Michalatos-Beloin S et al., *Nucleic Acids Res* 1996, 24: 4841-4843; Lo YMD et al., *Nucleic Acids Res* 1991, 19:3561-3567), клонирования и расщепления рестрикционными ферментами (Smirnova AS et al., *Immunogenetics* 2007, 59: 93-8) и др. Еще другие методы основываются на распределении и структуре неравновесного сцепления блоков гаплотипов в популяции, что позволяет составить материнский гаплотип по статистическим оценкам (Clark AG, *Mol Biol Evol* 1990, 7: 111-22; 10:13-9; Salem RM et al., *Hum Genomics* 2005; 2:39-66).

С. Использование информации по геному из материнского образца для составления каркаса

В одном воплощении для того, чтобы узнать, какая из материнских хромосом была передана плоду, применяется метод относительной дозы гаплотипа (RHDO). Основной принцип этого метода следует из примера, когда мать гетерозиготна по каждому из генетических полиморфизмов. Итак, имеется два гаплотипа и относительная доза этих гаплотипов должна быть 1:1. Однако присутствие в материнском образце небольшой доли ДНК плода может изменить относительную дозу гаплотипа. Это происходит потому, что плод наследует половину своего комплекта гаплотипов от матери, а другую половину от отца. Более того, по каждой хромосоме плод мог унаследовать “мозаику” гаплотипов, происходящих из той или другой гомологичной хромосомы от каждого родителя, в зависимости от наличия мейотической рекомбинации. Все эти факторы могут вызвать отклонение относительной дозы гаплотипов от соотношения 1:1 в составе материнской ДНК. Таким образом, для данной хромосомы или участка хромосомы составляющие аллели этих гаплотипов можно отыскивать из аналитических данных (например, данных по секвенированию), полученных из материнского образца.

Затем можно выполнить статистическую процедуру для определения относительной дозы гаплотипов или того, что один из гаплотипов представлен сильнее, чем другой.

Классификационный порог для этой статистической процедуры может быть установлен в зависимости от фракционной концентрации ДНК плода. В общем случае, более высокая фракционная концентрация ДНК плода способствует достижению порога при меньшем количестве молекул. Классификационный порог также может быть установлен в зависимости от количества успешно классифицированных фрагментов, которое нужно получить по всему геному или по искомым участкам генома. В одном воплощении может применяться последовательный критерий отношения вероятностей (SPRT).

В одном воплощении для определения относительной дозы аллеля по определенным полиморфизмам матери можно использовать относительную дозу мутаций (RMD), как описано в US Patent Application 2009/0087847. Эти относительные количества могут использоваться при определении гаплотипа у плода (например, когда полиморфизмы приходятся на следующие друг за другом или связанные локусы). В одном исполнении этого прицельного подхода применяется полимеразная цепная реакция (ПЦР) для амплификации специфических последовательностей из отдельных частей генома для анализа RMD. Для того, чтобы этот подход к определению наследования плодом по RMD распространялся на большую область генома или целый геном, требуется большой объем материнского образца.

В одном воплощении с использованием рандомизированного секвенирования представляющие интерес участки генома не выбираются прицельно. Таким образом, количество последовательностей, полученных в интересующих участках генома, может быть не столь большим, как при прицельном подходе (если только не проводится очень глубокое секвенирование). Однако при этом можно объединить итоги по нескольким сцепленным полиморфизмам для достижения необходимой статистической значимости для диагностических целей. Практическое значение этого воплощения секвенирования состоит в том, что можно сэкономить расходы, так как отпадает необходимость в слишком глубоком секвенировании. К тому же нужно вносить меньшее количество материнского образца, чем при подходах на основе цифровой ПЦР.

Кроме того, может быть желательно проведение такого анализ RHDO блоками. Иными словами, каждую хромосому можно анализировать в одном или предпочтительно больше чем одном блоке. В одном аспекте последнее дает возможность наблюдать мейотическую рекомбинацию. Например, может оказаться, что гаплотип одного сегмента определенной хромосомы плода происходит из одной из материнских гомологичных хромосом, тогда как другой сегмент той же самой хромосомы плода имеет гаплотип из другой гомологичной хромосомы. Анализ по SPRT может способствовать проведению такого сегментирования.

Например, анализ по SPRT можно провести на соседствующих SNP, проявляющих требуемую конфигурацию родительского генотипа (т.е. гомозиготного отца и гетерозиготной матери), начиная с одного конца хромосомы. Это должно продолжаться до тех пор, пока анализ SPRT не покажет, что в аналитических данных материнской плазмы (например, данных по секвенированию) преобладает один из материнских гаплотипов. Затем анализ SPRT можно “перенастроить” и начать заново со следующего соседнего SNP, проявляющего требуемую конфигурацию родительского генотипа. Это можно опять продолжать до тех пор, пока анализ SPRT не покажет еще раз, что в аналитических данных материнской плазмы (например, данных по секвенированию) преобладает один из материнских гаплотипов. Этот процесс может продолжаться до последнего выбранного SNP на данной хромосоме. Затем эти разнообразные установленные по SPRT сегменты гаплотипов на хромосоме можно сравнить с гаплотипами двух гомологичных хромосом в геноме матери. Мейотическая рекомбинация отмечается тогда, когда окажется, что сегменты гаплотипов у плода переключились с одной материнской гомологичной хромосомы на другую. Эта система срабатывает, даже если имеется больше чем одна гомологичная рекомбинация на хромосому.

Как будет описано позже, анализ RHDO может проводиться и по тем участкам генома, в которых и отец, и мать оба гетерозиготны по входящим в их состав генетическим полиморфизмам. Этот сценарий особенно полезен в тех случаях, когда отец и мать содержат мутантную копию вызывающего заболевание гена от одного и того же предка, как-то когда они близкородственны или когда вызывающая заболевание мутация преобладает вследствие сильного эффекта основателя (т.е. большинство индивидов с этой мутацией унаследовали один и тот же гаплотип от общего предка-основателя популяции). Таким образом, для вычисления гаплотипа плода можно использовать гаплотипы отца и матери на этом участке.

II. Составление генома плода из генома матери

Далее описано составление генетической карты плода или выяснение геномной последовательности плода при явном знании родительских геномов.

A. Способ

На фиг. 1 представлена блок-схема способа определения 100 по меньшей мере части генома еще не рожденного плода беременной женщины. У плода есть отец и беременная женщина-мать. Отец обладает отцовским геномом с двумя гаплотипами, а мать обладает материнским геномом с двумя гаплотипами. Способом 100 проводится анализ молекул (фрагментов) нуклеиновой кислоты из биологического образца, взятого у беременной женщины, для определения генома плода. Способ 100 описан в основном на

примере, когда отец гомозиготен, а мать гетерозиготна по нескольким локусам, тогда как в других примерах описаны другие воплощения.

Способ 100 и любой из описанных здесь способов может целиком или частично выполняться с помощью компьютерной системы, включающей процессор, который может быть настроен на выполнение стадий. Таким образом, воплощения направлены на компьютерные системы, настроенные на выполнение стадий любых из описанных здесь способов, потенциально с различными компонентами, выполняющими соответствующую стадию или соответствующую группу стадий. Хотя они и представлены в виде пронумерованных стадий, стадии способов могут выполняться одновременно или в другом порядке. Кроме того, части этих стадий можно использовать вместе с частями других стадий из других способов. К тому же все стадии или части стадий могут быть необязательными. Кроме того, любые стадии любых способов могут выполняться с помощью модулей, контуров или других средств для выполнения этих стадий.

На стадии 110 определяется первое множество локусов, по которым материнский геном гетерозиготен. В одном воплощении это определение может выполняться как часть генотипирования отца и матери на уровне генома или по выбранным геномным локусам. В других воплощениях определение первого множества локусов может проводиться во время анализа материнского образца, что описано в дальнейших разделах.

На стадии 120 определяется каждый из двух материнских гаплотипов, охватывающих первое множество локусов. Как указано выше, материнский геном может быть получен при прямом секвенировании. В других воплощениях генотипирование может проводиться по множеству локусов с последующим использованием картированного генома лица, у которого должен быть похожий геном, например, члена семьи, или же контрольного генома, общего для той же или сходной популяции. В одном воплощении стадия 120 может выполняться сначала для всего или части материнского генома, а затем материнский геном может быть исследован для выявления локусов, по которым мать гетерозиготна.

В одном аспекте составление гаплотипа хромосом отца не является обязательным. Однако, если можно составить отцовские гаплотипы, то из результатов секвенирования можно получить дополнительную информацию. Такая дополнительная информация включает возможность проведения анализа относительной дозы гаплотипа для участков, по которым оба родителя гетерозиготны. Другая дополнительная информация, которая может быть получена, если доступен отцовский гаплотип, это информация касательно мейотической рекомбинации с участием одной или нескольких отцовских хромосом, чтобы установить, переданы ли плоду вызывающие заболевания аллели, связанные с

такими полиморфизмами.

На стадии 130 определяется аллель, унаследованный плодом от отца по каждому из первого множества локусов. В некоторых воплощениях используются геномные локусы, по которым отец гомозиготен, а мать гетерозиготна (как указано на стадии 110). Таким образом, если отец гомозиготен по этим локусам, то будут известны аллели, унаследованные от отца. Генотипирование отца для определения локусов, по которым отец гомозиготен, может осуществляться любым из описанных здесь способов. В одном воплощении определение первого множества локусов может осуществляться, исходя из генотипирования отца и матери, чтобы обнаружить локусы, по которым отец гомозиготен, а мать гетерозиготна.

В другом воплощении можно использовать второе множество локусов генома отца, по которым он гетерозиготен, для определения отцовского гаплотипа, унаследованного плодом по первому множеству локусов, по которым отец гомозиготен. Например, если материнский геном гомозиготен по второму множеству локусов, то можно идентифицировать аллели, присутствующие в отцовском геноме в соответствующих локусах из второго множества локусов и отсутствующие в материнском геноме. Затем можно идентифицировать унаследованный от отца гаплотип как гаплотип с идентифицированными аллелями и использовать его для определения аллелей, унаследованных от отца по первому множеству локусов. Эти аспекты определения отцовского гаплотипа обсуждаются более подробно ниже.

На стадии 140 проводится анализ множества молекул нуклеиновой кислоты из биологического образца, взятого у беременной женщины. Образец содержит смесь нуклеиновых кислот матери и плода. Материнский биологический образец может быть взят, а затем обработан для анализа. В одном воплощении используется материнская плазма и сыворотка. В других воплощениях можно использовать материнскую кровь, материнскую мочу, материнскую слюну, смывную жидкость матки или клетки плода, полученные из крови матери.

В одном воплощении анализ молекул нуклеиновой кислоты включает определение локализации молекулы нуклеиновой кислоты в геноме человека и определение её аллеля по индивидуальному локусу. Таким образом, в одном воплощении может выполняться количественное генотипирование с использованием установленных аллелей молекул нуклеиновой кислоты из одного и того же локуса. Для определения геномной локализации и аллелей (информации о генотипе) молекул нуклеиновой кислоты в материнском образце можно использовать любой подходящий способ. Некоторые из таких способов описаны в заявках США 12/178,181 и 12/614350 и заявке под названием “Анализ генома на основе

размера”.

На стадии 150, исходя из установленных аллелей молекул нуклеиновой кислоты, определяются количества соответствующих аллелей по каждому из первого множества локусов. В одном воплощении эти количества могут составлять число аллелей каждого типа в первом локусе. Например, шесть А и четыре Т. В другом воплощении количество может означать распределение по размерам молекул нуклеиновой кислоты, содержащих определенные аллели. Например, относительное количество также может включать распределение по размерам фрагментов с определенным генотипом, что может означать относительное количество фрагментов определенной длины. Такие относительные количества могут дать информацию и о том, какой генотип у генома плода, так как у плода фрагменты обычно меньше, чем фрагменты матери. Некоторые примеры количеств и способов описаны в заявках США 12/178,181 и 12/614350 и заявке под названием “Анализ генома на основе размера”.

В одном воплощении относительные количества аллелей в одном локусе могут дать информацию о том, какой генотип унаследован плодом (например, после того, как набор данных достигнет достаточной статистической значимости). Например, относительные количества могут использоваться для определения того, что происходит дисбаланс последовательности относительно генотипа матери по одному локусу. В приведенных выше родственных патентных заявках предоставлены примеры воплощений для выявления дисбаланса последовательности по определенному локусу или участку.

На стадии 160 проводится сравнение относительных количеств соответствующих аллелей молекул нуклеиновой кислоты по более чем одному локусу из первого множества локусов. В некоторых воплощениях перед проведением сравнения суммируют количество каждого аллеля по всем локусам первого множества локусов, составляющих гаплотипы. Затем сравнивают суммарные количества родительских гаплотипов для определения того, что гаплотип избыточно представлен, одинаково представлен или недостаточно представлен. В других воплощениях сравнивают количества аллелей в локусах и проводят сравнение по множественным локусам. Например, можно суммировать значения разделения (например, разности или отношения), которые можно использовать при сравнении с пределом отсечения. Каждое из этих воплощений применимо к любым из описанных здесь операций сравнения.

В различных воплощениях относительное количество может означать встречаемость каждого фрагмента с определенным аллелем в определенном локусе, встречаемость фрагментов из любого локуса (или любых локусов на участке) по определенному гаплотипу и статистическое значение встречаемости (например, среднее) по

определенному локусу или по определенному гаплотипу. Так, в одном воплощении сравнение может означать определение величины разделения (например, разности или отношения) одного аллеля в сравнении с другим аллелем по каждому локусу.

На стадии 170, исходя из сравнения, можно определить гаплотип, унаследованный еще не рожденным плодом от матери в части генома, охваченной первым множеством локусов. В одном воплощении для того, чтобы установить, какая из материнских хромосом перешла к плоду, применяется способ относительных гаплотипов (RHDO), например, как указано выше. Поскольку мать гетерозиготна по каждому из первых локусов, то первые локусы соответствуют двум гаплотипам для области генома первых локусов. Относительная доза этих гаплотипов должна быть 1:1, если образец был только от матери. Отклонение или отсутствие отклонений от этого соотношения можно использовать для определения гаплотипа плода, унаследованного от матери (и от отца, что изложено более подробно далее) Таким образом, для данной хромосомы или участка хромосомы аллели, составляющие эти гаплотипы, можно отыскать из аналитических данных (например, данных по секвенированию), полученных на стадии 130.

Поскольку множество локусов подвергается анализу и сравнению с гаплотипом матери, то последовательности между локусами можно отнести к определенному гаплотипу. В одном воплощении, если несколько локусов соответствуют определенному гаплотипу, то можно заключить, что сегменты последовательности между локусами будут такими же, как и материнский гаплотип. Вследствие явления мейотической рекомбинации окончательный гаплотип, унаследованный плодом, может состоять из мозаики “сегментов гаплотипов”, происходящих из одной из этих двух гомологичных хромосом. Воплощения могут выявлять такую рекомбинацию.

Разрешение, с которым можно выявить такую рекомбинацию, зависит от числа и распределения генетических маркеров, которые были определены в составе ДНК матери и отца, и от порога, который применяется в последующем биоинформационном анализе (к примеру, методом SPRT). Например, если сравнение свидетельствует о том, что унаследованный от матери аллель по каждому из первого набора последовательных локусов соответствует первому гаплотипу, то первый гаплотип считается унаследованным по положению в геноме, соответствующему первому набору локусов. Если второй набор последовательных локусов свидетельствует о том, что унаследован второй гаплотип, то второй гаплотип считается унаследованным по положению в геноме, соответствующему второму набору локусов.

В одном воплощении, поскольку подвергается анализу множество локусов, то гаплотип может быть определен с большей точностью. Например, статистические данные

для одних локусов могут и не быть решающими, но в сочетании со статистическими данными по другим локусам можно установить, какой гаплотип унаследован. В другом воплощении можно подвергнуть анализу каждый локус независимо для составления классификации, а затем проанализировать эти классификации, чтобы установить, какой гаплотип унаследован для данного участка.

В одном воплощении для определения относительной дозы гаплотипа (например, если один из гаплотипов избыточно представлен перед другим гаплотипом) можно провести статистическую процедуру. Классификационный порог для этой статистической процедуры может быть установлен в зависимости от фракционной концентрации ДНК плода. В общем, более высокая фракционная концентрация ДНК плода позволяет достичь порога при меньшем количестве молекул. Классификационный порог также может быть установлен в зависимости от количества успешно классифицированных сегментов, которое нужно получить по всему геному или отдельным участкам генома.

Обращаясь опять к фиг. 1, на стадии 180 можно подвергнуть анализу геном плода на мутации. Например, воплощения можно использовать для поиска на предмет мутаций, вызывающих генетические заболевания в определенной популяции. Примеры мутаций, которые можно установить с помощью воплощений, можно найти на сайте Online Mendelian Inheritance in Man (www.ncbi.nlm.nih.gov/omim/getmorbid.cgi). Эти мутации можно поискать на стадиях 140-160 или в виде отдельной стадии, как описано далее. Например, поиск на мутации в семьях, в которых отец является носителем одной или нескольких мутаций, отсутствующих у матери, можно вести по аналитическим данным (например, данным по секвенированию) из материнского биологического образца.

Помимо выявления существующих мутаций, можно вести поиск на полиморфные генетические маркеры, сцепленные с мутантным аллелем или аллелем дикого типа у отца или матери. Например, при анализе RHDO может оказаться, что плод унаследовал от матери известный гаплотип, который несет мутацию, вызывающую заболевание. Воплощения изобретения также могут применяться для неинвазивной пренатальной диагностики заболеваний, вызванных делециями участков хромосом, например, делеции Southeast Asian, вызывающей альфа-талассемию. В том случае, когда и отец, и мать являются носителями делеции, а плод гомозиготен по делеции, то при проведении массированного параллельного секвенирования ДНК из плазмы матери должно наблюдаться снижение частоты последовательностей ДНК, происходящих из делетированного участка в материнской плазме.

В. Пример

В этом разделе описан пример воплощений (например, способа 100), применимых

к однонуклеотидному полиморфизму (SNP), при котором мать гетерозиготна. Аллели SNP на одной и той же хромосоме образуют гаплотип, причем у матери есть гомологичная пара каждой хромосомы, т.е. два гаплотипа. Для иллюстрации того, как выполняется такое определение, рассмотрим сегмент на хромосоме 3, например, как показано на фиг. 2.

На фиг. 2 представлены два гаплотипа для отца и два гаплотипа для матери по определенному сегменту соответствующего им геномного кода. Внутри этого сегмента обнаружено пять SNP, по которым отец гомозиготен, а мать гетерозиготна по всем 5 из этих SNP. Две гомологичные хромосомы отца имеют один и тот же гаплотип (Нар), т.е. A-G-A-A-G (сверху вниз на фиг. 2). Для простоты отцовские гаплотипы называются Нар I и Нар II, учитывая, что оба они идентичны по этому набору из 5 SNP. У матери отмечены два гаплотипа, а именно Нар III – A-A-A-G-G и Нар IV – G-G-G-A-A.

В этом примере SNPs можно далее классифицировать на два типа. На фиг. 3 представлены два типа SNP согласно воплощениям настоящего изобретения. Тип А состоит из таких SNP, у которых родительские аллели те же, что и в материнском гаплотипе III. Тип В состоит из таких SNP, в которых родительские аллели те же, что и в материнском гаплотипе IV.

Эти два типа SNP могут потребовать слегка различной математической обработки. Так, по сценарию типа А наследование плодом гаплотипа III приведет к избыточной представленности гаплотипа III по отношению к гаплотипу IV в материнской плазме (фиг. 4А). Например, рассматривая для упрощения только один SNP 410, аллель А наследуется от отца, а если от матери наследуется Нар III, то у плода будет два аллеля А в образце, что приведет к избыточной представленности А. Если же плод унаследовал гаплотип IV, то избыточная представленность не будет наблюдаться, так как плод тоже будет гетерозиготным по А и G в этом локусе.

С другой стороны, по сценарию типа В наследование плодом гаплотипа III приведет к равной представленности гаплотипа III и гаплотипа IV в материнской плазме (фиг. 4В). Например, рассматривая SNP 420, наследование G от отца и А в составе Нар III приведет к тому, что у плода будут равные количества А и G по SNP 420, совсем как у матери. Если же плод унаследовал гаплотип IV, тогда будет наблюдаться избыточная представленность, как явствует из вышеприведенного обсуждения.

На фиг. 5А и 5В представлен анализ путем сравнения относительных количеств (например, встречаемости) фрагментов для каждого локуса и будет ли результат сравнения означать, что определенный гаплотип наследуется или нет. Для этого примера можно использовать любое место в геноме, в котором есть SNP, который соответствует одной из этих конфигураций генотипа у отца и матери (например, сценарии типа А или

типа В). Из данных по секвенированию материнской плазмы можно сосредоточиться на том, какое количество просеквенированных молекул соответствует определенному аллелю SNP. Можно использовать анализ по SPRT (или другой метод сравнения) для установления наличия какого-либо аллельного дисбаланса между этими аллелями (Lo YD et al., *Proc Natl Acad Sci USA* 2007, 104: 13116-13121).

На фиг. 5А представлен анализ SNP типа А. Как видим, для каждого SNP сравнение методом SPRT относительных количеств (например, определенных по величине разделения) со значением порога отсечения обеспечивает классификацию. В одном воплощении, если достигнут классификационный порог для SPRT, то можно заключить, что плод унаследовал определенный материнский гаплотип. После этого можно заново настроить параметры для анализа SPRT. Затем можно перейти к анализу соседнего SNP, соответствующего требуемой конфигурации генотипа, как в направлении от теломеры к центромере, так и наоборот; и можно начинать новый анализ SPRT по этому SNP.

С другой стороны, в одном воплощении, если классификация SNP по SPRT не была достигнута, то можно опять же перейти к соседнему SNP аналогичным образом, за исключением того, что баллы для следующего SNP можно добавить к предыдущему, а затем снова провести анализ по SPRT. Этот процесс может продолжаться до тех пор, пока не будет достигнут классификационный порог. На фиг. 5А и фиг. 5В представлено исполнение этого процесса для анализов типа А и типа В. В одном воплощении все классификации анализируются вместе для получения итоговой классификации для участка. Например, если получена классификация для первой группы SNP и для следующей группы SNP, то можно сравнить обе классификации, чтобы узнать, согласуются ли они.

На фиг. 6 представлен эффект изменения отношения вероятностей при классификации по SPRT (Zhou W et al., *Nat Biotechnol* 2001, 19:78-81; Karoui NE et al., *Statist Med* 2006, 25:3124-33). В общем, меньшее отношение вероятностей для классификации, например, 8, может облегчить проведение классификации. При этом может увеличиться количество классифицированных участков внутри генома. Однако следует ожидать, что некоторые из таких участков будут классифицированы неправильно. С другой стороны, при большем отношении вероятностей для классификации, например, 1200, классификация станет возможной только при оценке большего числа SNP. При этом может уменьшиться количество классифицированных участков внутри генома. Следует ожидать, что количество и доля неправильно классифицированных участков будет меньше по сравнению с использованием меньшего классификационного порога.

В одном воплощении классификация засчитывается только в том случае, если две

последовательные классификации по SPRT дают один и тот же гаплотип (что именуется алгоритмом “двух последовательных блоков”). В одном аспекте алгоритм “двух последовательных блоков” может повысить точность классификации. В некоторых воплощениях, для любых отрезков последовательности, сначала может выполняться анализ SPRT для SNP типа А, а затем проводится другой анализ SPRT для SNP типа В. В одном воплощении предусмотрен сценарий для таких отрезков последовательности, в которых SNP типа А и типа В образуют две перемежающиеся группы генетических ориентиров (например, SNP). В воплощениях с применением алгоритма “двух последовательных блоков” эти два блока могут быть различного типа.

Результаты анализа SPRT для типа А и типа В следует проверить на согласованность или несогласованность результатов классификации. Для повышения точности классификации в одном воплощении (“перемежающийся подход”) классификация засчитывается только в том случае, если анализ типа А и типа В для данного участка генома дает непротиворечивые результаты. Если эти два анализа дают противоречивые результаты, то нужно проверить результаты классификации двух смежных участков, следующих за этим, одного со стороны центромеры и другого со стороны теломеры. Если эти два смежных участка дают непротиворечивые результаты, то первый участок можно классифицировать как имеющий одинаковый гаплотип с этими двумя участками. Если же два смежных участка не дают непротиворечивые результаты, то можно перейти к следующим двум смежным участкам, пока не будет наблюдаться согласованность. Один из вариантов этой темы состоит в том, чтобы продвигаться только в одном направлении и принимать результаты классификации следующих одного, двух или же нескольких смежных участков за результаты исходного рассматриваемого участка. Общий принцип заключается в том, чтобы использовать результаты классификации смежных участков генома для проверки результатов классификации данного участка.

III. Определение отцовских аллелей, унаследованных плодом

На фиг. 7 представлена блок-схема способа определения 700 по меньшей мере части генома, унаследованной от отца, у еще не рожденного плода беременной женщины. Способом 700 проводится анализ молекул (фрагментов) нуклеиновой кислоты из биологического образца, взятого у беременной женщины, для определения генома плода. Образец содержит смесь нуклеиновых кислот матери и плода.

На стадии 710 проводится анализ каждой из множества молекул нуклеиновых кислот из биологического образца для определения локализации молекулы нуклеиновой кислоты в геноме человека и определения аллельного типа молекулы нуклеиновой кислоты. Таким образом, в одном воплощении можно определить генотипы молекул

нуклеиновой кислоты в определенном месте (локусе). Для этого анализа можно применять любые способы, описанные выше и по тексту.

На стадии 720 определяется первое множество локусов, по которым отцовский геном гетерозиготен, а материнский геном гомозиготен. В одном воплощении первое множество локусов получают путем определения отцовского и материнского геномов. Эти геномы можно проверить на геномные локусы, по которым отец гетерозиготен, а мать гомозиготна.

На стадии 730 определяется гаплотип, унаследованный еще не рожденным плодом от отца в той части генома, которая охвачена первым множеством локусов, исходя из установленных генотипов по первому множеству локусов. В одном воплощении из аналитических данных (например, данных по секвенированию) по каждому из этих локусов определяется аллель, который представлен у отца, но отсутствует в геноме матери. Комбинация этих аллелей будет указывать на гаплотипы хромосом, которые плод унаследовал от отца.

В другом воплощении, если известны гаплотипы каждой из хромосом или искомых участков хромосом в геноме отца, то можно определить, где во время сперматогенеза у отца произошла мейотическая рекомбинация. Так, мейотическая рекомбинация у отца отмечается тогда, когда гаплотип отрезка ДНК в унаследованной от отца хромосоме отличается между плодом и отцом. Включение такой информации о рекомбинации может быть полезным при использовании аналитических данных (например, данных по секвенированию) для пренатальной диагностики генетических заболеваний по сцеплению с генетическими полиморфизмами.

IV. Отец и мать гетерозиготны на участке генома

Воплощения предусматривают сценарий, при котором отец и мать гетерозиготны на участке генома. Этот сценарий может быть особенно важным для семей, в которых отец и мать являются близкими родственниками, а также когда заболевание связано с преобладающей мутацией вследствие сильного эффекта предка-основателя. При таких обстоятельствах следует ожидать, что если отец и мать еще не рожденного плода являются носителями мутантного гена, то гаплотип хромосомы, несущей мутантную копию гена, будет практически идентичным, за исключением случаев мейотической рекомбинации. Такой тип анализа может оказаться особенно полезным при аутосомных рецессивных заболеваниях, таких как кистозный фиброз, бета-талассемия, серповидно-клеточная анемия и болезнь гемоглобина E.

На фиг. 8 представлена блок-схема способа определения 800 по меньшей мере части генома еще не рожденного плода на участке, в котором отец и мать гетерозиготны

согласно воплощениям настоящего изобретения.

На стадии 810 определяется первое множество локусов, по которым и отец, и мать гетерозиготны. В одном воплощении первые локусы можно определить любым из приведенных здесь способов. Например, можно просеквенировать весь отцовский геном или его участки либо генотипировать различные части для выявления первых локусов. При этом станут известны каждый из двух отцовских и каждый из двух материнских гаплотипов по первому множеству локусов.

В качестве примера на фиг. 9 представлены гаплотипы отца и матери, которые оба гетерозиготны на определенном участке генома. Как видим, оба родителя содержат мутантный ген (аллель) на участке 1. В частности, Нар I отца и Нар III матери содержат мутантный ген. Также видно, что и отец, и мать могут содержать другую копию хромосомы, несущую копию гена дикого типа. В частности, Нар II отца и Нар IV матери содержат ген дикого типа. Таким образом, этот пример относится к определению того, унаследовал ли плод мутантный ген. Хромосомы от отца и матери, несущие ген дикого типа, имеют одинаковый гаплотип в непосредственной близости от гена, но могут иметь разные гаплотипы на большем удалении от гена. Поскольку эта хромосома скорее всего происходит от различных предков, то она вряд ли будет иметь идентичные гаплотипы между отцом и матерью по всей длине хромосомы.

На стадии 820 определяется второе множество локусов, по которым отец гетерозиготен, а мать гомозиготна. Как видим, первое и второе множество локусов находятся на одной и той же хромосоме. На участке 2 представлены такие вторые локусы. Участок 2 выбирается таким образом, чтобы отец был гетерозиготен по одному или нескольким SNP на этом участке, а мать была гомозиготна на этом участке.

На стадии 830 проводится анализ фрагментов из образца беременной женщины для установления локализации в геноме человека и генотипа. Локализация может использоваться для определения того, что фрагмент (молекула нуклеиновой кислоты) включает один или несколько первых локусов либо один или несколько вторых локусов. Затем эта информация может использоваться для определения гаплотипа, унаследованного от отца, и гаплотипа, унаследованного от матери.

На стадии 840 определяется, какой из двух родительских гаплотипов был унаследован плодом, путем анализа установленных генотипов множества молекул нуклеиновой кислоты из биологического образца хотя бы по одному из вторых локусов. Например, можно провести поиск аллелей SNP, представленных исключительно в геноме отца, но отсутствующих в геноме матери, как-то аллеля Т, обозначенного *, и аллеля А, обозначенного + на фиг. 9, из аналитических данных (например, локализации и генотипа,

установленных на стадии 710) материнского биологического образца. Так же, как это делается в способе 700, если в материнской плазме выявляется аллель Т, обозначенный *, то это значит, что плодом от отца унаследован гаплотип II (Нар II). Наоборот, если в материнской плазме определяется аллель А, обозначенный +, то это означает, что плодом от отца унаследован Нар I.

На стадии 850 проводится сравнение относительных количеств установленных генотипов молекул нуклеиновой кислоты по более чем одному из первого множества локусов. В одном воплощении суммируют количества по всем локусам и сравнивают относительные количества материнских гаплотипов. Относительные количества могут означать встречаемость, распределение по размерам и любые другие параметры, дающие информацию о том, какой генотип представлен в геноме плода по определенному локусу.

На стадии 860, исходя из того, какой гаплотип унаследован плодом от отца, и исходя из сравнения относительных количеств, определяется гаплотип, унаследованный еще не рожденным плодом от матери в той части генома, которая охвачена первым множеством локусов. Так, для определения того, какой из двух материнских гаплотипов унаследован плодом, проводится анализ RHDO (например, как описано выше) SNP на участке 1 по аналитических данным материнского биологического образца, с учетом отцовского гаплотипа, унаследованного плодом на участке 2. В одном воплощении предполагается, что между участками 1 и 2 не было рекомбинации при переходе этих участков от родителей к плоду.

Например, рассмотрим сценарий, когда при анализе участка 2 было установлено, что плод унаследовал Нар I от отца. Тогда наследование плодом Нар III (который идентичен Нар I на участке 1) от матери приведет к избыточной представленности Нар III относительно Нар IV в материнской плазме. Наоборот, если плод унаследовал Нар IV от матери, то в материнской плазме будет наблюдаться равная представленность Нар III и Нар IV.

В качестве другого примера рассмотрим сценарий, когда при анализе участка 2 было установлено, что плод унаследовал Нар II от отца. Тогда наследование плодом Нар IV (который идентичен Нар II на участке 1) от матери приведет к избыточной представленности Нар IV относительно Нар III в материнской плазме. Наоборот, если плод унаследовал Нар III от матери, то в материнской плазме будет наблюдаться равная представленность Нар III и Нар IV.

В предыдущих разделах, мы выводили геном плода и фракционную концентрацию ДНК плода, используя данные, полученные при секвенировании ДНК из материнской плазмы, а также информацию о генотипе родителей плода. В следующих разделах

описаны воплощения для выведения фракционной концентрации ДНК плода и генотипа плода без предварительной информации о генотипах/гаплотипах матери и отца.

V. Определение фракционной концентрации ДНК плода

В некоторых воплощениях необязательной стадией является определение фракционной концентрации ДНК плода. В различных аспектах эта фракционная концентрация может определять объем анализа (например, необходимый объем секвенирования) или определять точность анализа для данного количества данных (например, полноту охвата последовательности генома). Определение фракционной концентрации ДНК плода может применяться и для определения предела отсечения при классификации того, какой гаплотип и/или генотип наследуется.

В одном воплощении фракционная концентрация ДНК плода определяется путем извлечения аналитических данных (например, полученных на стадии 140 и 710) для локусов, по которым и отец, и мать гомозиготны, но с различными аллелями. Например, для SNP с двумя аллелями, А и G, у отца может быть AA, а у матери GG, и наоборот. По таким локусам плод будет облигатной гетерозиготой. В вышеприведенном примере у плода будет генотип AG, а доля аллеля А в материнском образце может использоваться для определения фракционной концентрации ДНК плода. В другом воплощении для определения локуса, по которому мать гомозиготна, а плод гетерозиготен, может проводиться статистический анализ. При этом не требуется никакой информации о геноме матери или геноме отца.

В качестве альтернативы извлечению аналитических данных фракционную концентрацию ДНК плода также можно определить при помощи другого подхода, как-то методом ПЦР, цифровым методом ПЦР или методом на основе масс-спектрометрии, используя комплект полиморфных генетических маркеров (Lun FMF et al., Clin Chem 2008, 54: 1664-1672). Другой альтернативой является использование одного или нескольких геномных локусов, отличающихся по метилированию ДНК между плодом и матерью (Poon LLM et al., Clin Chem 2002, 48: 35-41; Chan KCA et al., Clin Chem 2006, 52: 2211-2218; US Patent 6,927,028). Еще одной альтернативой является использование приближенной фракционной концентрации ДНК плода, определенной в контрольной популяции, например, при аналогичном гестационном возрасте. Однако, поскольку фракционная концентрация ДНК плода может отличаться от образца к образцу, можно ожидать, что этот последний подход будет менее точным, чем при измерении концентрации специально для тестируемого образца.

A. Определение фракционной концентрации у облигатных гетерозигот

В тех воплощениях, в которых плод является облигатной гетерозиготой, можно

определить фракционную концентрацию ДНК плода, используя следующие серии расчетов (например, при массивном параллельном секвенировании). Пусть p означает количество аллеля у плода, который отсутствует в материнском геноме. Пусть q означает количество другого аллеля, а именно аллеля, который есть в геноме и матери, и плода. Фракционная концентрация ДНК плода задается следующим уравнением:

$$\frac{2p}{p+q}.$$

В одном воплощении такое вычисление проводится на совокупных данных по всем полиморфным генетическим локусам или полиморфным генетическим признакам, соответствующим конфигурации родительских генотипов (например, оба родителя гомозиготны, но по различным аллелям).

В. Определение по информативным SNP

Фракционную концентрацию ДНК плода также можно определить для любого локуса, по которому мать гомозиготна, а плод гетерозиготен, а не только когда мать гомозиготна по одному аллелю, а отец гомозиготен по другому аллелю. Оба способа предусматривают, что локус является информативным. Термин “информативный SNP” может применяться в разных контекстах в зависимости от того, какая нужна информация. В одном контексте это информация об аллеле в геноме плода в определенном локусе, в котором аллель не представлен в материнском геноме. Так, подмножество SNP, по которым мать гомозиготна, а плод гетерозиготен, можно называть “информативными SNP” в контексте определения концентрации ДНК плода. Случаи, когда и мать, и плод гетерозиготны, но отличаются хотя бы по одному аллелю, также могут использоваться в качестве информативных SNP. Однако триаллельные SNP относительно редко встречаются в геноме.

На фиг. 10 представлена блок-схема способа определения 1000 фракционной концентрации материала плода в материнском образце согласно воплощениям настоящего изобретения. На стадии 1010 фрагменты из образца беременной женщины подвергаются анализу для установления локализации в геноме человека и типа аллелей (что может привести к определению генотипа на этом участке). В одном воплощении фрагменты подвергаются анализу путем секвенирования множества молекул нуклеиновой кислоты из биологического образца, взятого у беременной женщины. В других воплощениях может применяться ПЦР в реальном времени или цифровой метод ПЦР.

На стадии 1020 определяется, что один или несколько первых локусов являются информативными. В некоторых воплощениях материнский геном гомозиготен, но в образце по информативному локусу выявляется не материнский аллель. В одном

воплощении геном плода гетерозиготен по всем первым локусам, а материнский геном гомозиготен по всем первым локусам. Например, в одном первом локусе геном плода может содержать первый и второй аллели (например, ТА), а материнский геном может содержать два соответствующих вторых аллеля (например, АА) в первом локусе. Однако такие локусы могут и не быть заранее известны, например, в ситуациях, когда плод не является облигатной гетерозиготой.

В одном воплощении для определения информативного локуса рассматриваются те SNP, по которым мать гомозиготна. По тем SNP, по которым мать гомозиготна, плод будет либо гомозиготен по тому же аллелю, либо гетерозиготен. Например, если SNP полиморфен по А и Т, а у матери генотип АА, то у плода будет генотип АА или ТА. В этом случае наличие аллеля Т в образце материнской плазмы означает, что у плода будет генотип ТА, а не АА. Некоторые воплощения нацелены на то, в какой степени наличие аллеля Т означает генотип ТА путем вычисления необходимого предела отсечения, как описано ниже.

На стадии 1030 для хотя бы одного из первых локусов определяется первое число p случаев (встречаемость) соответствующего первого аллеля и второе число q случаев (встречаемость) соответствующего второго аллеля. В одном воплощении число специфичных для плода (аллель Т) и совместных (аллель А) аллелей в материнской плазме определяется различными методами, к примеру, без ограничения, ПЦР в реальном времени, цифровым методом ПЦР и массивированным параллельным секвенированием.

На стадии 1040 на основании первых и вторых чисел рассчитывается фракционная концентрация. В одном воплощении для беременной женщины с генотипом АА и её плода с генотипом ТА фракционная концентрация ДНК плода (f) рассчитывается по уравнению: $f = 2 \times p / (p + q)$, где p означает встречаемость специфичного для плода аллеля (аллель Т), а q означает встречаемость аллеля, встречающегося и у матери, и у плода (аллель А).

В другом воплощении фракционная концентрация ДНК плода в материнской плазме может быть установлена с большей точностью при использовании множественных информативных SNP. При использовании встречаемости аллелей по множественным SNP (общее число SNP = n) фракционная концентрация ДНК плода (f) может быть рассчитана по уравнению:

$$f = \frac{\sum_{i=1}^n 2p_i}{\sum_{i=1}^n (p_i + q_i)},$$

где p_i означает встречаемость специфичного для плода аллеля по информативному SNP $_i$; q_i – встречаемость аллеля, совместного для матери и для плода по информативному SNP $_i$; а n – общее число информативных SNP. Использование встречаемости аллелей по

множественным SNP может повысить точность определения фракционной концентрации ДНК плода.

С. Фракционная концентрация без явной генетической информации о родителях

Теперь опишем способ определения фракционной концентрации ДНК плода в образце материнской плазмы, при котором не требуется предварительная информация о генотипах плода и матери. В одном воплощении осуществляется определение информативных SNP по встречаемости различных аллелей в локусах этих SNP в материнской плазме. При этом способ 1000 может применяться вместе с определением информативных SNP на основании воплощений, описанных ниже. Сначала представим описание вероятностей для того, чтобы облегчить понимание расчета порога отсечения, используемого для идентификации информативных SNP.

В одном воплощении вероятность выявления специфичного для плода аллеля подчиняется распределению Пуассона. Вероятность (P) выявления специфичного для плода аллеля можно вычислить по следующему уравнению: $P = 1 - \exp(-f \times N/2)$, где f означает фракционную концентрацию ДНК плода в образце материнской плазмы, N – общее число молекул, соответствующих локусу данного анализируемого SNP, а $\exp()$ означает экспоненциальную функцию. В одном аспекте P рассматривается как ожидаемое распределение, так как это распределение получается не при измерении количества молекул по многим образцам. В других воплощениях можно использовать и другие распределения.

Принимая, что фракционная концентрация ДНК плода составляет 5% (типичное значение для первого триместра беременности) и подвергается анализу 100 молекул (матери + плода), соответствующих этому локусу SNP (что эквивалентно количеству, содержащемуся в 50 диплоидных геномах), вероятность обнаружения специфичного для плода аллеля (аллеля T) равна $1 - \exp(-0,05 \times 100/2) = 0,92$. Вероятность обнаружения специфичного для плода аллеля возрастает с повышением фракционной концентрации ДНК плода и числа молекул, подвергаемых анализу на локус SNP. Например, если концентрация ДНК плода составляет 10% и анализируется 100 молекул, то вероятность обнаружения специфичного для плода аллеля составляет 0,99.

Следовательно, в локусе SNP, по которому мать гомозиготна, присутствие в материнской плазме аллеля, отличающегося от материнского, может означать, что этот SNP “информативен” для расчета фракционной концентрации ДНК плода. Вероятность пропуска любого информативного SNP может зависеть от числа анализируемых молекул. Иными словами, для выявления информативных SNP при любом желательном уровне

достоверности число молекул, которое нужно проанализировать для получения требуемой точности, можно рассчитать по вероятностной функции Пуассона.

Применяя вышеуказанный анализ, в некоторых воплощениях можно определить, является ли локус информативным или нет, когда генотип матери неизвестен. В одном воплощении определяются локусы, по которым в материнской плазме проводится выявление двух различных аллелей. Например, для локуса SNP с двумя возможными аллелями, А и Т, в материнской плазме проводится выявление обоих аллелей, А и Т.

На фиг. 11 представлена блок-схема способа определения 1100 того, что локус является информативным согласно воплощениям настоящего изобретения. В одном воплощении способ 1100 может применяться при выполнении стадии 1020 способа 1000. В другом воплощении одна стадия способа 1100 заключается в определении предела отсечения на основе статистического распределения, а на другой стадии предел отсечения используется для определения того, является ли локус (SNP) информативным.

На стадии 1110 определяется предел отсечения для предсказания встречаемости соответствующего первого аллеля в определенном локусе. В одном воплощении предел отсечения предсказывает, является ли материнский геном гомозиготным, а геном плода гетерозиготным. В другом воплощении предел отсечения определяется на основе статистического распределения встречаемости для различных комбинаций гомозиготности и гетерозиготности в конкретном локусе. Например, распределение частоты аллелей можно предсказать по вероятностной функции Пуассона.

На стадии 1120, исходя из анализа молекул нуклеиновой кислоты из материнского образца (например, на стадии 1010), в локусе проводится выявление первого аллеля и второго аллеля. Например, при картировании на анализируемый локус приходится набор фрагментов, в которых был обнаружен первый аллель или второй аллель. Первый аллель может соответствовать одному из соответствующих первых аллелей из стадии 1020, а второй аллель может соответствовать одному из соответствующих вторых аллелей. В одном воплощении, если не обнаруживаются два различных аллеля, то это значит, что локус не является информативным.

На стадии 1130 определяется фактическая встречаемость соответствующего первого аллеля в локусе на основании анализа молекул нуклеиновой кислоты. Например, можно подсчитать результаты секвенирования множества молекул нуклеиновой кислоты и установить, сколько раз на этот локус приходится фрагмент с генотипом первого аллеля.

На стадии 1140 локус идентифицируется как один из первых локусов, исходя из сравнения фактической встречаемости с пределом отсечения. В одном аспекте предел отсечения используется для различения между тремя возможностями: (а) мать

гомозиготна (AA), а плод гетерозиготен (AT); (b) мать гетерозиготна (AT), и плод гетерозиготен (AT); и (c) мать гетерозиготна (AT), а плод гомозиготен (AA) или (TT). Для иллюстрации в приведенных ниже примерах принимается, что у плода генотип AA по сценарию (c). Однако вычисление будет таким же, если у плода будет генотип TT. Информативному локусу соответствует вариант (a).

В одном воплощении локус определяется как один из первых локусов, если его фактическая встречаемость меньше предела отсечения. В другом воплощении можно использовать и меньшее значение порога, чтобы избежать ложного картирования.

А теперь опишем определение предела отсечения. Исходя из физиологически возможной фракционной концентрации ДНК плода (эта информация доступна из предыдущих исследований) и общего числа молекул, соответствующих локусу SNP, можно предсказать распределение встречаемости аллелей для трех возможных сценариев, приведенных выше. Исходя из предсказанного распределения, можно определить предел отсечения для интерпретации наблюдаемой встречаемости аллелей в материнской плазме для определения того, что SNP является “информативным” (т.е. вариант (a)) или нет.

Фракционная концентрация ДНК плода обычно составляет от 5% до 20% в начале беременности и от 10% до 35% в конце беременности (Lun et al. Micro fluidics digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma. Clin Chem 2008, 54: 1664-72). Так, в одном воплощении определяли прогнозируемое распределение встречаемости аллелей при фракционной концентрации ДНК плода в 5% и 20%.

На фиг. 12A представлено прогнозируемое распределение встречаемости для аллеля Т (менее распространенный аллель по сценариям (a) и (c)) для трех сценариев с предполагаемой фракционной концентрацией ДНК плода в 20%. На фиг. 12B представлено прогнозируемое распределение встречаемости для аллеля Т (менее распространенный аллель по сценариям (a) и (c)) для трех сценариев с предполагаемой фракционной концентрацией ДНК плода в 5%. В обеих моделях предполагается, что для анализа локуса SNP было взято в целом 200 молекул.

Используя в качестве предела отсечения встречаемость менее распространенного аллеля (аллеля Т), составляющую 40 раз, можно статистически отличить три возможности. Иными словами, при выявлении в материнской плазме локуса SNP с двумя аллелями и при анализе в общей сложности 200 молекул, если частота второстепенного аллеля (менее распространенного аллеля) будет меньше 40, то этот локус SNP можно считать “информативным”. При фракционной концентрации ДНК плода в 5% и 20% различие “информативных” SNP (сценарий (a)) от тех SNP, по которым мать гетерозиготна (сценарии (b) и (c)), будет правильным на 100%.

На практике общее число выявляемых молекул может быть различным для различных SNP. Для каждого локуса SNP можно составить свою кривую прогнозируемого распределения с учетом общего числа выявляемых молекул в образце материнской плазмы, охватывающих этот локус SNP. Иными словами, предел отсечения для определения того, что SNP является информативным или нет, может варьировать между различными SNP и зависит от того, сколько раз встречался этот локус SNP.

В следующей таблице представлена встречаемость аллелей из трех локусов SNP в материнской плазме при секвенировании образца материнской плазмы. По каждому из трех SNP в материнской плазме обнаружено два разных аллеля. Общее число выявленных случаев в материнской плазме, соответствующих этим трем SNP, отличается друг от друга.

Локус SNP	Обозначение SNP	Аллель (разы)	Аллель (разы)	Общее число раз
1	rs3107146	A (10)	G (163)	173
2	rs7522344	G (9)	T (112)	121
3	rs2272908	A (72)	G (62)	134

Прогнозируемое распределение встречаемости менее распространенного аллеля при фракционной концентрации ДНК плода в 20% и различном числе молекул, соответствующих SNP, представлено на фиг. 13А, 13В и 14. Прогнозируемое распределение составляли, принимая концентрацию ДНК плода в 20%, так как она представляет верхний предел концентрации ДНК плода в первом триместре. Чем более высокая концентрация ДНК плода, тем большее ожидается перекрытие между кривыми распределения второстепенного аллеля, по которому мать гомозиготна, и главного аллеля, по которому мать гетерозиготна. Таким образом, для предсказания информативных SNP будет более уместным вычисление предела отсечения для второстепенного аллеля с использованием более высокой концентрации ДНК плода.

На фиг. 13А представлено прогнозируемое распределение встречаемости менее распространенного аллеля при общем числе 173 молекул и фракционной концентрации ДНК плода в 20%. В одном воплощении, исходя из этого распределения, для идентификации информативных SNP подходит критерий отсечения менее 40 для встречаемости менее распространенного аллеля. Поскольку аллель А встречается 10 раз, то локус SNP № 1 считается “информативным” для вычисления фракционной концентрации ДНК плода.

На фиг. 13В представлено прогнозируемое распределение встречаемости менее распространенного аллеля при общем числе 121 молекулы и фракционной концентрации ДНК плода в 20%. В одном воплощении, исходя из этого распределения, для иден-

тификации информативных SNP подходит предел отсечения менее 26 для встречаемости менее распространенного аллеля. Поскольку аллель Т встречается 9 раз, то локус SNP № 2 считается “информативным” для вычисления фракционной концентрации ДНК плода.

На фиг. 12 представлено прогнозируемое распределение встречаемости менее распространенного аллеля при общем числе 134 молекул и фракционной концентрации ДНК плода в 20%. В одном воплощении, исходя из этого распределения, для идентификации информативных SNP подходит предел отсечения менее 25 для встречаемости менее распространенного аллеля. Поскольку аллель Т встречается 62 раза, то локус SNP № 3 считается “не информативным” и не должен использоваться для вычисления фракционной концентрации ДНК плода.

В некоторых воплощениях фракционную концентрацию ДНК плода можно рассчитать по формуле $f = 2 \times p / (p + q)$, используя встречаемость аллелей для SNP № 1 и № 2 и объединяя их. Результаты представлены ниже.

Локус SNP, используемый для вычисления	Фракционная концентрация ДНК плода
1	$10 \times 2 / (10 + 163) = 11,6\%$
2	$9 \times 2 / (9 + 112) = 14,9\%$
1 и 2	$(10 + 9) \times 2 / (10 + 9 + 163 + 112) = 12,9\%$

D. Определение полноты охвата генома плода

Помимо получения фракционной концентрации, в воплощениях можно определить степень охвата генома плода при выполнении аналитической процедуры (например, секвенирования) на стадии 1010. В некоторых воплощениях для определения степени охвата можно использовать информативные локусы. Например, можно использовать любой из вышеприведенных примеров. В одном воплощении можно использовать локусы, по которым плод является облигатной гетерозиготой. В другом воплощении можно использовать локусы, по которым плод оказался гетерозиготным, а мать гомозиготна (например, используя способ 1100).

Для определения степени охвата можно использовать фрагменты, которые при картировании попадают на информативные локусы. В одном воплощении определяется доля тех локусов из первого множества локусов, в которых по результатам секвенирования выявляется соответствующий первый аллель. Например, если по одному локусу у плода будет ТА, а у матери АА, то по результатам секвенирования должен выявляться аллель Т, если этот локус подвергнулся секвенированию. Таким образом, можно рассчитать, какая доля генома плода в биологическом образце подвергалась секвенированию, исходя из этого соотношения. В одном воплощении в качестве степени охвата генома

плода принимается доля тех первых локусов, в которых отмечен специфичный для плода аллель. В других воплощениях эта доля может подвергаться модификации в зависимости от того, где находятся эти локусы. Например, можно определить степень охвата для каждой хромосомы. С другой стороны, можно считать, что степень охвата меньше, чем эта доля, если первые локусы не являются хорошей репрезентативной выборкой генома. С другой стороны, можно представить диапазон, в котором эта доля составляет один конец диапазона. В то время, как высокая степень охвата, т.е. близкая к 100%, означает почти полный охват генома плода, большинство генетических заболеваний можно диагностировать при значительном меньшем охвате, чем 100%, например, 80% или 50% или еще меньше.

VI. Отсутствие предварительной информации о геноме матери и отца

В предыдущих разделах, в некоторых воплощениях определяли генетическую карту плода (или части генома плода), когда были известны гаплотипы матери и генотипы отца. В других воплощениях было показано, что фракционную концентрацию ДНК плода можно определить путем анализа ДНК в материнской плазме без предварительного знания о генотипах матери, отца или плода. Еще в других воплощениях мы теперь также опишем способ определения генетической карты плода (или части генома плода) при помощи анализа RHDO без предварительной информации о генотипах/гаплотипах матери и отца.

В одном воплощении используется информация о контрольных (например, распространенных или известных) гаплотипах той популяции, к которой принадлежат родители. Эта информация может использоваться для выведения гаплотипов матери и отца. Покажем принцип этого метода на одном примере. Информацию по таким контрольным гаплотипам можно получить, к примеру, из вебсайта международного проекта International HarMap Project (harmap.ncbi.nlm.nih.gov/).

Для примера предположим, что в популяции есть три контрольных гаплотипа (Нар А, Нар В и Нар С, как показано на фиг. 15А). Каждый из этих трех гаплотипов состоит из 14 локусов SNP, причем в каждом локусе есть два возможных аллеля. В этом примере отец обладает Нар В и Нар С, а мать обладает Нар А и Нар В, как показано на фиг. 15В. В этом примере принимается, что плод наследует Нар А от матери и Нар С от отца. Поэтому у плода будет Нар А и Нар С, как показано на фиг. 15В.

На фиг. 16 представлена блок-схема способа определения 1600 по меньшей мере части генома плода, когда известен набор контрольных гаплотипов, но неизвестны родительские гаплотипы, согласно воплощениям настоящего изобретения.

На стадии 1610 материнский образец подвергается анализу для выявления тех SNP,

по которым мать гомозиготна, а плод гетерозиготен. Такой анализ проводится таким же образом, как и определение информативных локусов, как описано выше. Так, в одном воплощении можно использовать способ 1000 и/или 1100. В других воплощениях, описанных выше, подвергаются анализу геномы матери и отца, чтобы получить данные для картирования генома плода.

На фиг 17 представлен пример определения информативных локусов при анализе фрагментов ДНК из материнского образца. Для каждого из 14 локусов определяется встречаемость двух аллелей по каждому локусу. Встречаемость этих аллелей можно определить, к примеру, без ограничения, методом ПЦР в реальном времени, цифровым методом ПЦР и массивованным параллельным секвенированием. По каждому из этих локусов нужно выявить два разных аллеля в материнской плазме. В отличие от тех SNP, по которым мать гетерозиготна, доля этих двух аллелей будет существенно отличаться. Специфичный для плода аллель (тот, который плод наследует от отца) будет встречаться гораздо реже, чем материнский аллель. Информативные локусы 1710 отмечены на фиг. 17.

На стадии 1620 определяется один или несколько аллелей из отцовского гаплотипа, унаследованных плодом. В одном воплощении каждый из локусов 1710 может использоваться для определения унаследованного отцовского гаплотипа. Например, отцовский аллель, унаследованный плодом, можно идентифицировать как специфичный для плода аллель по локусам 1720, поскольку в материнском образце специфичный для плода аллель встречается гораздо реже, чем материнский аллель.

На стадии 1630 отцовские аллели сравнивают с контрольными гаплотипами, чтобы определить гаплотип, унаследованный от отца. В некоторых воплощениях можно вывести целый ряд возможных гаплотипов плода, каждый со своей вероятностью. Затем один или несколько из наиболее вероятных гаплотипов плода можно использовать для последующего анализа или для клинической диагностики.

В приведенном на фиг. 18 примере в популяции есть три возможных гаплотипа (Нар А, Нар В и Нар С). При анализе материнской плазмы четыре SNPs оказались гомозиготными для матери и гетерозиготными для плода, поэтому они представляют отцовские аллели, унаследованные плодом. Генотипы по этим четырем SNPs соответствуют профилю Нар С. Следовательно, плод унаследовал Нар С от отца, как показано на фиг. 19. Иными словами, для всех SNPs в пределах одного и того же блока гаплотипов можно вычислить отцовские аллели, унаследованные плодом.

На стадии 1640 можно определить локусы (например, SNPs), по которым мать гетерозиготна. В одном воплощении те SNPs, по которым мать гетерозиготна, получают при анализе материнского образца. Например, по каждому из этих SNPs выявляется два

разных аллеля. В отличие от тех SNPs, по которым мать гомозиготна, а плод гетерозиготен, когда специфичные для плода аллели составляют лишь небольшую долю от всех аллелей в материнской плазме, встречаемость двух аллелей будет одинаковой для тех SNPs, по которым мать гетерозиготна. Таким образом, при анализе материнской плазмы можно установить полный генотип матери по всем локусам SNP в пределах одного блока гаплотипов, например, как показано на фиг. 20.

На стадии 1650 составляются материнские гаплотипы из генотипов матери на стадии 1640 путем сравнения генотипов по локусам с информацией о гаплотипах в соответствующей популяции. На фиг. 21 представлено воплощение для определения материнских гаплотипов из генотипов матери и контрольных гаплотипов. В данном примере мать гомозиготна по аллелю G в третьем локусе SNP. Поскольку только Нар А и Нар В соответствуют этому критерию, то это значит, что у матери должна быть одна из трех комбинаций гаплотипов, а именно Нар А/НарА, Нар А/Нар В или Нар В/НарВ. Кроме того, поскольку мать гетерозиготна по А и С для первого SNP, можно заключить, что у матери должна быть комбинация гаплотипов Нар А/Нар В. В одном воплощении может возникнуть более чем одна возможность, причем каждую возможность можно проверить на следующей стадии. Из вышеприведенного анализа были установлены гаплотипы матери и гаплотип, унаследованный плодом от отца. На фиг. 22 представлены установленные гаплотипы матери и унаследованный от отца гаплотип.

На стадии 1660 определяется материнский гаплотип, унаследованный плодом, из гаплотипов матери, установленных на стадии 1650, и унаследованного от отца гаплотипа, установленного на стадии 1630. Используя эту информацию, в одном воплощении для определения того, какой материнский гаплотип перешел к плоду, применяется анализ RHDO. Анализ RHDO может проводиться по любому из описанных здесь воплощений.

В одном воплощении для анализа RHDO те SNPs, по которым мать гетерозиготна, можно разделить на два типа, тип альфа и тип бета (например, как показано на фиг. 23 и описано выше). К SNPs типа альфа относятся те локусы, в которых отцовский аллель, перешедший к плоду, идентичен материнскому аллелю в составе Нар А. Для SNPs типа альфа, если плод от матери унаследует Нар А, то аллели из Нар А будут избыточно представлены в материнской плазме. С другой стороны, если плод от матери унаследует Нар В, то два материнских аллеля будут одинаково представлены в материнской плазме.

К SNPs типа бета относятся те локусы, в которых отцовский аллель, перешедший к плоду, идентичен материнскому аллелю в составе Нар В. Для SNPs типа бета, если плод от матери унаследует Нар В, то аллели из Нар В будут избыточно представлены в материнской плазме. Однако, если плод от матери унаследует Нар А, то два материнских

аллеля будут одинаково представлены в материнской плазме. Возможную избыточную представленность аллелей из Нар А или Нар В можно определить при помощи анализа RHDO.

В некоторых воплощениях при применении анализа RHDO к определенному участку без предварительной информации о гаплотипах матери и генотипах отца может потребоваться сравнительно многократный охват SNPs в пределах блока гаплотипов, к примеру, в одном воплощении нужно проанализировать 200 молекул, соответствующих локусу SNP. Такая информация может быть получена, к примеру, без ограничения, методом ПЦР в реальном времени, цифровым методом ПЦР и массивованным параллельным секвенированием. В одном воплощении для получения репрезентативной и объективной количественной информации по различным аллелям на искомом участке можно использовать прицельное секвенирование (например, сочетанием обогащения мишени и массивованного параллельного секвенирования). Прицельное секвенирование описано в приведенном ниже примере. Таким образом, анализ RHDO может применяться к данным по прицельному секвенированию ДНК в материнской плазме для определения того, какие материнские аллели/гаплотипы передаются плоду, без предварительной информации о генотипах/гаплотипах родителей.

VII. Выявление мутаций de novo

В некоторых воплощениях можно выявить мутации, приобретенные плодом. Мутации de novo – это такие мутации, которые не передаются отцом или матерью, а возникают, к примеру, при гаметогенезе у отца или у матери или у обоих. Выявление их имеет клиническое значение, так как мутации de novo играют важную роль, вызывая целый ряд генетических заболеваний, например, гемофилию А и хондродистрофию.

На фиг. 24 представлена блок-схема, иллюстрирующая способ идентификации 2400 мутаций de novo в геноме еще не рожденного плода у беременной женщины. У плода есть отец и беременная женщина-мать, причем отец обладает отцовским геномом с двумя гаплотипами, а мать обладает материнским геномом с двумя гаплотипами, при этом способ включает:

На стадии 2410 проводится секвенирование множества молекул нуклеиновой кислоты из биологического образца, взятого у беременной женщины. Отметим, что образец содержит смесь нуклеиновых кислот матери и плода.

На стадии 2420 устанавливается локализация каждой из просеквенированных молекул нуклеиновой кислоты в геноме человека. В одном воплощении картирование последовательностей осуществляется путем секвенирования с одного конца или с обоих концов. В одном аспекте картирование для локализации в геноме человека не требует

точного совпадения каждого из нуклеотидов при выявлении локализации.

На стадии 2430 для каждого из по крайней мере части местоположений определяется материнская последовательность и отцовская последовательность в данном положении. Например, если на стадии 2420 установлено 100 местоположений, то можно определить материнские и отцовские геномы в этих 100 положениях. В одном воплощении отцовские последовательности определяются из взятого у отца образца, в отличие от использования контрольных гаплотипов, как это описано выше. При этом можно выявить мутации, не представленные в контрольном геноме. В различных воплощениях материнские последовательности можно получить из образца, содержащего только материнскую ДНК, или же получить из биологического образца, например, с помощью описанных здесь методов.

На стадии 2440 устанавливается первая последовательность во множестве молекул нуклеиновой кислоты, которая отсутствует у выявленных материнских или отцовских последовательностей. В одном воплощении сравнение первой последовательности с выявленными материнскими или отцовскими последовательностями требует точного совпадения. Так, если совпадение не будет точным, то считается, что первая последовательность отсутствует у выявленных материнских или отцовских последовательностей. При этом можно идентифицировать даже незначительные мутации *de novo*, так как мутация *de novo* может представлять собой изменение всего лишь одного нуклеотида. В другом воплощении для того, чтобы последовательность рассматривалась как мутация *de novo*, необходимо определенное количество фрагментов ДНК, проявляющих не материнскую и не отцовскую последовательность. Например, можно использовать предел отсечения в 3 фрагмента ДНК для определения того, что в последовательности есть мутация *de novo* или нет её.

На стадии 2450 определяется первая фракционная концентрация первой последовательности в биологическом образце. Например, количество фрагментов ДНК, проявляющих первую последовательность, можно выразить в виде доли от всех фрагментов ДНК, выявленных по этому локусу.

На стадии 2460 определяется вторая фракционная концентрация нуклеиновой кислоты плода в биологическом образце, используя молекулу нуклеиновой кислоты, унаследованную плодом от отца, которая присутствует в отцовском геноме, но отсутствует в материнском геноме. Такая молекула нуклеиновой кислоты может содержать первый аллель в таком локусе, по которому отец гомозиготен, а мать тоже гомозиготна, но по другому аллелю, поэтому плод будет облигатной гетерозиготой. Для выяснения того, какую молекулу нуклеиновой кислоты следует использовать для определения второй

фракционной концентрации, можно использовать информативные локусы, описанные выше.

В других воплощениях вторая фракционная концентрация определяется при помощи других подходов, как-то методом ПЦР, цифровым методом ПЦР или методом на основе масс-спектрометрии, по Y-хромосоме, по комплекту генетических полиморфизмов, т.е. полиморфизмов по одному нуклеотиду или полиморфизмов типа вставки-делеции (Lun FMF et al., Clin Chem 2008, 54: 1664-1672). Другая альтернатива состоит в использовании одного или нескольких геномных локусов, проявляющих различия по метилированию ДНК между плодом и матерью (Poon LLM et al., Clin Chem 2002, 48: 35-41; Chan KCA et al., Clin Chem 2006, 52: 2211-2218; US Patent 6,927,028).

В одном воплощении различия в эпигенетическом статусе отражаются различиями по метилированию ДНК. Различия по метилированию ДНК могут затрагивать ген 1A из семейства доменов ассоциации с RAS (RASSF1A) или ген лигазы синтетазы голокарбоксилазы (биотин-(пропионил-СоА-карбоксилазы (АТФ-гидролизующей)) (HLCS). Количество фрагментов ДНК со специфичным для плода профилем метилирования ДНК можно выразить в виде доли от всех фрагментов ДНК, возникающих по этому локусу с дифференциальным метилированием.

На стадии 2470 первая последовательность классифицируется как мутация *de novo*, если первая и вторая фракционная концентрация примерно одинаковы. Не материнские и не отцовские последовательности, возникающие из-за ошибок в процессе анализа, например, ошибок при секвенировании, являются случайными явлениями и могут случиться с малой вероятностью. Поэтому множественные фрагменты ДНК, проявляющие одну и ту же не материнскую и не отцовскую последовательность в количестве, близком к измеренной фракционной концентрации ДНК плода в образце, скорее всего представляют мутацию *de novo*, присутствующую в геноме плода, а не возникли из-за ошибки при секвенировании. В одном воплощении для определения того, что фракционные концентрации одинаковы, можно использовать предел отсечения. Например, если концентрации отличаются друг от друга на заданную величину, то первая последовательность классифицируется как мутация *de novo*. В различных воплощениях заданная величина может составлять 5%, 10% или 15%.

ПРИМЕРЫ

Пример 1

Для иллюстрации воплощений настоящего изобретения проанализируем следующий случай. Была привлечена пара, посещающая акушерскую клинику для пренатальной диагностики бета-талассемии. Отец был носителем делеции 4 пар оснований

-СТТТ из кодонов 41/42 гена бета-глобина человека. Беременная мать была носителем мутации A→G по нуклеотиду -28 гена бета-глобина человека. У отца и матери брали образцы крови. У матери образец крови брали перед взятием образца хориальных ворсин (CVS) после 12 недель беременности. Часть образца CVS сохраняли для эксперимента. Целью эксперимента было составление генетической карты всего генома или определение полной или частичной геномной последовательности плода путем массивированного параллельного секвенирования ДНК материнской плазмы.

1. Определение родительских генотипов

ДНК экстрагировали из лейкоцитной пленки отца и матери и из образца CVS. Эти образцы ДНК подвергали анализу на установке Affymetrix Genome-Wide Human SNP Array 6.0 system. Эта система содержит 1,8 млн генетических маркеров, включая ~900 000 полиморфизмов по одному нуклеотиду (SNP) и более 950 000 зондов для выявления вариаций числа копий. Абсолютное количество и проценты SNPs, проявляющих различные комбинации генотипов для отца, матери и плода (CVS), приведены в таблице на фиг. 25А.

Хотя в этом примере использовалась система Affymetrix, на практике можно использовать любую платформу генотипирования, известную в данной области. Так, помимо генотипирования, ДНК из лейкоцитной пленки отца и матери также можно подвергнуть секвенированию, как по всему геному, так и по отдельным участкам генома. Более того, для установления генотипов родителей можно использовать любой источник ДНК (например, ДНК буккальных клеток, ДНК волосных фолликулов и т.д.) из организма отца и матери.

Образец CVS подвергали анализу, чтобы получить стандарт для сравнения с генетической картой плода, составленной при анализе материнской плазмы. Кроме того, для этого эксперимента генотип образца CVS также можно использовать для составления гаплотипа матери для анализа RHDO. При этом использование генотипа CVS с целью составления гаплотипа матери применялось только для иллюстрации. При клиническом применении воплощений гаплотип матери можно составить посредством анализа других членов семьи, к примеру, предыдущих детей, братьев и сестер, родителей или других родственников матери. Гаплотип матери по искомым участкам хромосом также можно составить и другими методами, хорошо известными специалистам, из которых некоторые приведены здесь.

В отдельных воплощениях также определяется гаплотип отца еще не рожденного плода. Такая информация может быть особенно полезной при определении относительной дозы гаплотипа по тем участкам хромосом, по которым и отец, и мать гетерозиготны.

2. Массированное параллельное секвенирование ДНК материнской плазмы

ДНК из плазмы, полученной от матери, подвергали массированному параллельному секвенированию на платформе Illumina Genome Analyzer. Секвенирование молекул ДНК из плазмы проводили с двух концов. Каждую молекулу секвенировали по каждому концу на 50 п.о., в общей сложности 100 п.о. на молекулу. Оба конца последовательности выравнивали с демаскированным по повторам геномом человека (Hg18 NCBI.36 из сайта UCSC <http://genome.ucsc.edu>) с помощью программы SOAP2 из Пекинского института геномики в Шеньчжэне (Beijing Genomics Institute at Shenzhen; soap.genomics.org.cn) (Li R et al., *Bioinformatics* 2009, 25(15): 1966-7). В таблице на фиг. 25В приведена статистика выравнивания для первых 20 проточных кювет. Так, при 20 проточных кюветах с контрольным геномом человека было выровнено свыше 3,932 млрд. прочитанных отрезков.

3. Вычисление фракционной концентрации ДНК плода

Как указано выше, фракционную концентрацию ДНК плода в образце материнской плазмы можно вычислить по данным секвенирования. Так, можно проанализировать те SNPs, по которым и отец, и мать гомозиготны, но по различным аллелям. По таким SNPs плод будет облигатной гетерозиготой по одному отцовскому и одному материнскому аллелю. В одном воплощении можно использовать любой из способов вычисления, описанных в разделе V. А в этом примере вычисления проводились по кумулятивным данным из различных полиморфных генетических локусов, соответствующих конфигурации родительских генотипов (т.е. оба родителя гомозиготны, но по различным аллелям) на разных хромосомах. Фракционные концентрации ДНК плода, рассчитанные по SNPs, локализованным на разных хромосомах, приведены в самом правом столбце на фиг. 26. Как видно из таблицы, фракционные концентрации, рассчитанные по SNPs, локализованным на разных хромосомах, очень хорошо коррелируют друг с другом.

В качестве эксперимента по контролю качества исследовали и те SNPs, по которым мать гомозиготна, а отец гетерозиготен, путем анализа образцов из лейкоцитной пленки методом Affymetrix SNP 6.0 (средний столбец на фиг. 26). Видно, что при достаточной полноте секвенирования ДНК измеренные при этом фракционные концентрации ДНК плода оказались очень близкими тем, которые измеряли по SNPs, по которым и отец, и мать гомозиготны, но по различным аллелям.

В одном воплощении, если по этим двум типам SNPs наблюдаются почти одинаковые фракционные концентрации ДНК плода, то можно заключить, что геном плода был почти полностью охвачен секвенированием. В одном аспекте, при меньшей полноте охвата фракционные концентрации ДНК плода, измеренные по SNPs, по которым

мать гомозиготна, а отец гетерозиготен, будут более высокими, чем при измерении по тем SNPs, по которым и отец, и мать гомозиготны, но по различным аллелям. При меньшей полноте охвата отсутствие уникального по отцовской линии аллеля может быть вызвано двумя возможными факторами: (i) тем, что плод не унаследовал этот аллель от отца; и/или (ii) тем, что плод унаследовал этот аллель от отца, но затем он был утерян при секвенировании, потому что полнота секвенирования была недостаточной.

4а. Вычисление степени охвата генома плода

Также, как указано выше, можно определить, какой процент генома плода был проанализирован при секвенировании ДНК из материнской плазмы, если посмотреть на подгруппу SNPs, по которым и отец, и мать гомозиготны, но по различным аллелям. В этой семье 45 900 SNPs на матрице Affymetrix SNP 6.0 принадлежали к этой подгруппе. Степень охвата генома плода можно вычислить путем анализа данных по секвенированию ДНК из плазмы, чтобы узнать, у какого процента SNPs из этой подгруппы выявляется аллель плода при секвенировании.

На графике из фиг. 27А представлен наблюдаемый процент SNP в этой подгруппе, в которой аллель плода наблюдается по данным секвенирования при анализе первых 20 проточных кювет. Так, аллель плода наблюдается у 94% таких SNPs. Такая степень секвенирования соответствует более 3,932 млрд. прочитанных отрезков при последовательности в 100 п.о. На графике из фиг. 27В представлен охват в зависимости от числа прочитанных отрезков вместо количества проточных кювет. С увеличением производительности различных платформ секвенирования следует ожидать, что количество проточных кювет или запусков, используемых или необходимых для получения такого количества прочитанных последовательностей или отрезков последовательностей, в будущем уменьшится.

В некоторых воплощениях, поскольку в каждом участке хромосомы или хромосом выявляется несколько SNPs, то охват генома плода составит менее 94% и тем не менее обеспечит правильное картирование генома. Например, предположим, что на участке хромосомы имеется 30 информативных SNPs, но аллель плода выявляется только у 20 из этих 30 SNPs. Тем не менее, участок хромосомы можно правильно идентифицировать и при 20 SNPs. Так, в одном воплощении равнозначная точность достигается при охвате, составляющем менее 94%.

4б. Охват генетической картой аллелей, унаследованных плодом от отца

В этом примере анализа изучали те аллели SNPs, по которым отец гетерозиготен, а мать гомозиготна. В этой семье 131 037 SNPs на матрице Affymetrix SNP 6.0 относились к этой категории. Подгруппа таких SNPs состояла из 65 875 SNPs, по которым мать гомо-

зиготна, а отец и плод гетерозиготны. При использовании 20 проточных кювет аллели, наследуемые по отцовской линии, наблюдались у 61 875 таких SNPs, что означает охват в 93,9%. Этот процент хорошо соответствует данным по степени охвата, полученным в предыдущем параграфе. Корреляции между охватом наследуемых по отцовской линии аллелей и количеством поддающихся картированию прочитанных последовательностей и количеством последовательностей из проточных кювет приведены на фиг. 28А и фиг. 28В, соответственно.

Для выяснения специфичности такого подхода к выявлению истинных наследуемых по отцовской линии аллелей у плода подвергали анализу 65 162 (т.е. 131 037 – 65 875) SNPs, по которым плод унаследовал те же самые аллели, что и мать. Для таких SNPs кажущееся выявление аллелей, отличающихся от аллелей матери, должно означать ложно-положительные результаты. Так, из этих 65 162 SNPs при анализе 20 проточных кювет отмечено только 3 225 ложно-положительных (4,95%). Эти ложно-положительные результаты могли возникнуть в результате ошибок при секвенировании или ошибок при генотипировании ДНК отца или матери либо мутаций *de novo* у плода. Корреляция между долей ложно-положительных и количеством просеквенированных проточных кювет представлена на фиг. 29А.

Долю ложно-положительных можно определить и при рассмотрении подгруппы SNPs, по которым и отец, и мать гомозиготны по одному и тому же аллелю. Наличие любого альтернативного аллеля в определенном локусе считалось ложно-положительным результатом. Эти ложно-положительные могли возникнуть в результате ошибок при секвенировании или ошибок при генотипировании ДНК отца или матери либо мутаций *de novo* у плода. В этой подгруппе было 500 673 SNPs. По данным секвенирования из 20 проточных кювет ложно-положительные результаты отмечались у 48 396 SNPs (9,67%). Корреляция между долей ложно-положительных и количеством просеквенированных проточных кювет представлена на фиг. 29В. Доля ложно-положительных при этом была выше, чем при оценке по подгруппе SNPs, по которым и мать, и плод гомозиготны, а отец гетерозиготен. Это вызвано тем, что в последней подгруппе SNPs ложно-положительным считается только наличие в материнской плазме наследуемых по отцовской линии аллелей, тогда как в первой подгруппе ложно-положительным результатом считается любой аллель, отличающийся от общих для отца и матери аллелей.

На фиг. 30 представлен охват специфичных для плода SNPs при различном числе анализируемых проточных кювет. В этот анализ включены те SNPs, по которым и отец, и мать гомозиготны, но по различным аллелям. По оси X представлена степень охвата специфичных для плода SNPs, а по оси Y – процент SNPs с заданной степенью охвата.

При повышении числа анализируемых проточных кювет возрастает средняя степень охвата специфичных для плода SNPs. Например, при одной анализируемой проточной кювете средняя степень охвата SNPs составляет 0,23 раз. А при анализе 20 проточных кювет средняя степень охвата возрастает до 4,52 раз.

5. Правильность генетической карты, унаследованной от матери

На фиг. 31 представлена точность анализа типа А при использовании данных из 10 проточных кювет. В разделе II.В описаны воплощения анализа типа А и типа В (также именуется альфа и бета). Точность относится к правильному определению гаплотипа, унаследованного от матери. Точность представлена отдельно для каждой хромосомы.

Используя соотношение правдоподобия = 1200 при анализе SPRT (Zhou W et al., Nat Biotechnol 2001, 19:78-81; Karoui NE et al., Statist Med 2006, 25:3124-33), точность составляла от 96% до 100%. Как видим, даже при таком высоком соотношении правдоподобия для классификации по SPRT, в общей сложности классифицировали 2760 сегментов по всему геному. Такая степень разрешения достаточна для большинства задач, если учесть, что мейотическая рекомбинация происходит с частотой от одного до нескольких раз на одно плечо хромосомы за поколение. Кроме того, видно, что все ошибочные классификации можно было предотвратить при использовании перемежающегося подхода (справа на фиг. 31). Как описано выше, в перемежающемся подходе применяется анализ и типа А, и типа В.

На фиг. 32 представлена точность анализа типа В при использовании данных из 10 проточных кювет. Используя соотношение правдоподобия = 1200 при анализе SPRT, точность составляла от 94,1% до 100%. Все ошибочные классификации можно было предотвратить при использовании перемежающегося подхода (справа на фиг. 32), как это отмечалось и на фиг. 31.

На фиг. 33 представлена точность анализа типа А при использовании данных из 20 проточных кювет. Используя соотношение правдоподобия = 1200 при анализе SPRT и алгоритм “двух последовательных блоков”, в общей сложности было сделано 3780 классификаций, и только 3 (0,1%) классификации оказались неправильными. На фиг. 34 представлена точность анализа типа В при использовании данных из 20 проточных кювет. Используя соотношение правдоподобия = 1200 при анализе SPRT и алгоритм “двух последовательных блоков”, в общей сложности было сделано 3355 классификаций, и только 6 (0,2%) классификаций оказались неправильными. В этих примерах анализ SPRT проводился по целому ряду генетических маркеров типа SNPs.

Пример 2. Пренатальное определение риска бета-талассемии

В одном воплощении для определения риска бета-талассемии у плода (аутосомного

рецессивного заболевания) можно определить, унаследовал ли плод мутантные аллели, переносимые его отцом и матерью. В данном случае, приведенном выше, отец является носителем делеции 4 пар оснований -СТТТ из кодонов 41/42 гена бета-глобина человека. Беременная мать является носителем мутации А→G по нуклеотиду -28 гена бета-глобина человека.

Для того, чтобы определить, унаследовал ли плод отцовскую мутацию по кодонам 41/42, проводили поиск среди данных секвенирования ДНК из материнской плазмы с использованием первых 10 проточных кювет на предмет этой мутации. Всего обнаружили 10 сигналов с этой мутацией (фиг. 35А). Таким образом, плод унаследовал отцовскую мутацию. Кроме того, 62 сигнала содержали последовательность дикого типа по кодонам 41/42 (фиг. 35В). Следовательно, на этом участке процент сигналов, содержащих эту мутацию, равен 0,1389. Эта цифра очень близка фракционной концентрации ДНК плода, установленной на фиг. 26. В одном воплощении риск наследования плодом отцовской мутации можно определить и по наследованию им генетических полиморфизмов, сцепленных с отцовской мутацией.

В одном воплощении для определения риска того, что плод унаследовал материнскую мутацию -28, проводили анализ RHDO. В этой семье мутация -28 локализована на гаплотипе IV, тогда как аллель дикого типа локализован на гаплотипе III. Результаты анализа RHDO типа А представлены на фиг. 36, а результаты анализа RHDO типа В представлены на фиг. 37. При обоих типах анализа установлено наследование плодом гаплотипа III от матери. Иными словами, плод унаследовал от матери аллель дикого типа. Окончательный диагноз у плода состоит в том, что он унаследовал мутацию по кодонам 41/42 от отца и аллель дикого типа от матери. Таким образом, плод является гетерозиготным носителем бета-талассемии и поэтому должен быть клинически здоровым.

Пример 3. Обогащение мишени и прицельное секвенирование

Как изложено в предыдущих разделах, точность определения фракционной концентрации ДНК плода и разрешение генетической карты, составленной по анализу ДНК в материнской плазме, может зависеть от полноты охвата искомых локусов. Например, мы показали, что для определения с высокой точностью фракционной концентрации ДНК плода без предварительной информации о генотипе матери должно хватить и 200 молекул, соответствующих локусу SNP. Встречаемость аллелей для SNP в материнской плазме можно получить, к примеру, без ограничения, методом ПЦР в реальном времени, цифровым методом ПЦР и массивированным параллельным секвенированием.

Поскольку при массивированном параллельном секвенировании ДНК из материнской плазмы можно одновременно определить встречаемость аллелей для миллионов SNP по

всему геному, то это идеальная платформа для анализа всего генома по различным локусам. Основной формат массивированного параллельного секвенирования позволяет охватить различные участки генома с аналогичной полнотой. Однако для того, чтобы просеквенировать определенный участок с высокой полнотой секвенирования методом случайного массивированного параллельного секвенирования, нужно в такой же степени просеквенировать и остальные части генома (не предназначенные для анализа). Так что этот подход может оказаться дорогостоящим. Для улучшения рентабельности подхода массивированного параллельного секвенирования можно подвергнуть обогащению искомый участок перед проведением секвенирования. Прицельное секвенирование может проводиться при помощи захвата в растворимой фазе (Gnirke A et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 2009, 27:182-9), захвата на микроматрице (например, на платформе NimbleGen) или прицельной амплификации (Tewhey R et al. Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nat Biotechnol* 2009, 27: 1025-31).

Прицельное секвенирование сначала применялось для генетических вариаций в популяции, например, при исследовании генетической ассоциации. Поэтому его текущее применение в исследованиях по геномике направлено на решение качественных задач (например, генотипирование или выявление мутаций). Однако применение прицельного секвенирования ДНК из материнской плазмы для неинвазивной пренатальной диагностики включает и количественные соображения, осуществимость которых остается неясной. Например, применение прицельного секвенирования может вносить количественные ошибки при выявлении ДНК плода и матери в материнской плазме. Кроме того, в предшествующих работах было показано, что ДНК плода короче, чем ДНК матери (Chan KCA et al. Size distributions of maternal and fetal DNA in maternal plasma. *Clin Chem* 2004, 50: 88-92). Такая разница в размерах также может вносить количественные ошибки или различия в эффективности при захвате ДНК плода и матери в материнской плазме. К тому же нет уверенности в том, с какой эффективностью могут захватываться такие фрагментированные молекулы ДНК. В нижеследующем описании мы покажем, что прицельное секвенирование может осуществляться посредством обогащения мишени с последующим массивированным параллельным секвенированием. Мы также покажем, что обогащение мишени является эффективным способом оценки фракционной концентрации ДНК плода по сравнению с секвенированием всего генома.

А. Определение фракционной концентрации с обогащением мишени

1. Материалы и методы

Были привлечены 4 беременные женщины (M6011, M6028, M6029 и M6043) с

единственными плодами женского пола. Образцы периферической крови матерей собирали в пробирки с ЭДТА перед избирательным кесаревым сечением в третьем триместре, а образцы плаценты брали после избирательного кесарева сечения. После центрифугирования из клеток периферической крови экстрагировали ДНК с помощью набора Blood Mini Kit (Qiagen). ДНК из 2,4 мл плазмы экстрагировали с помощью набора DSP DNA Blood Mini Kit (Qiagen). Геномную ДНК матери экстрагировали из лейкоцитной пленки, а геномную ДНК плода экстрагировали из ткани плаценты. В этом примере образцы брали в третьем триместре только для иллюстрации. Равным образом можно использовать образцы из первого или второго триместра.

Генотипы матери и плода определяли на матрице Affymetrix Genome-Wide Human SNP Array 6.0. В каждом случае 5-30 нг ДНК из плазмы использовали для конструирования библиотеки ДНК с помощью набора Paired-end Sample Preparation kit (Illumina) по методике изготовителя с приготовлением образцов типа Chromatin Immunoprecipitation Sequencing. Лигированную с адаптером ДНК очищали прямо на центрифужных колонках из набора QIAquick PCR Purification kit (Qiagen) без дальнейшего отбора по размерам. Затем лигированную с адаптером ДНК подвергали амплификации за 15 циклов ПЦР с помощью стандартных праймеров, которыми служили PCR Primer PE 1.0 и 2.0 фирмы Illumina. Библиотеки ДНК подвергали количественному определению с помощью спектрофотометра NanoDrop ND-1000 (NanoDrop Technologies) и пропускали через анализатор 2100 Bioanalyzer, используя набор DNA 1000 kit (Agilent) для проверки распределения по размерам. По каждому образцу получали 0,6-1 мкг амплифицированной библиотеки ДНК из плазмы со средним размером около 290 п.о. Захватывающую библиотеку получали от Agilent, она охватывала 85% экзонов на хромосоме chrX человека (кат. № 5190-1993). Во всех 4 случаях в данном исследовании по 500 нг амплифицированной библиотеки ДНК из плазмы инкубировали с захватывающими зондами в течение 24 ч при 65°C согласно инструкции производителя. После гибридизации захваченные мишени отбирали путем притягивания гибридов типа биотинилированный зонд/мишень с помощью покрытых стрептавидином магнитных шариков (DynaL DynaMag-2, Invitrogen) и очищали с помощью набора MinElute PCR Purification Kit (Qiagen). Наконец, искомые библиотеки ДНК подвергали обогащению путем ПЦР-амплификации за 12 циклов с помощью праймеров SureSelect GA PE фирмы Agilent. Продукты ПЦР очищали с помощью набора QIAquick PCR Purification Kit (Qiagen). После этого библиотеки ДНК, полученные с обогащением мишени или без него, подвергали случайному массированному параллельному секвенированию на анализаторе Illumina Genome Analyzer IIx. При секвенировании использовали одну дорожку на стандартной проточной кювете для секвенирования одной библиотеки

ДНК.

2. Фракционная концентрация ДНК плода без обогащения мишени

Фракционная концентрация ДНК плода может быть рассчитана на основе встречаемости аллелей информативных SNP (т.е. SNPs, по которым мать гомозиготна, а отец гетерозиготен). Из нижеприведенной таблицы видно, что для 4 случаев по всему геному идентифицировано 120 184, 110 730, 107 362 и 110 321 информативных SNP, тогда как на искомый участок хромосомы X приходится 63, 61, 69 и 65 (в таком же порядке перечисления, соответственно). Без обогащения мишени фракционные концентрации ДНК плода составили 33,4%, 31,3%, 29,2% и 34,4%, исходя из данных по всем информативным SNP в геноме.

Образец	Обога- щение мишени	Информатив- ные SNP по всему геному	Совместные аллели	Специфич- ные для плода аллели	Фракционная конц. ДНК плода
M6011	нет	120 184	15 309	3 064	33,4%
M6028	нет	110 730	16 778	3 114	31,3%
M6029	нет	107 362	19 889	3 404	29,2%
M6043	нет	110 321	21 070	4 369	34,4%

3. Сравнение образцов с обогащением мишени и без него

В некоторых воплощениях полнота охвата последовательности представлена средним числом раз, когда каждое основание было просеквенировано на определенном участке. В этом воплощении полнота охвата последовательности рассчитывается делением общего числа просеквенированных оснований на искомом участке на длину искомого участка (3,05 Mb). Для участков, охватываемых набором для обогащения, средний охват последовательности составил 0,19 раз для необогащенных образцов и 54,9 раза для обогащенных образцов, что означает в среднем 289-кратное обогащение. При такой полноте секвенирования на искомом участке перед обогащением мишени выявлялось только 4,0% специфичных для плода аллелей (см. таблицу внизу). Для сравнения, после обогащения мишени выявлялось 95,8% из них. Следовательно, обогащение мишени сильно повышает показатель выявления специфичных для плода аллелей на заданном участке.

После этого мы сравнили фракционные концентрации ДНК плода на основе встречаемости всех информативных SNP на заданном участке по каждому образцу, с обогащением и без него. Без обогащения мишени число специфичных для плода аллелей составило от 0 до 6 для четырех образцов (см. таблицу внизу). Вследствие низкого охвата последовательности неадекватный отбор молекул ДНК плода будет мешать точному определению фракционной концентрации ДНК плода. При обогащении мишени на заданном участке наблюдалось гораздо большее число специфичных для плода аллелей (511-776) и

совместных аллелей (2570-3922) (см. таблицу внизу). Содержание ДНК плода составило 35,4%, 33,2%, 26,1% и 33,0%, что согласуется с содержанием ДНК плода, установленным по данным из всего генома в необогащенных образцах (см. таблицу внизу). Эти результаты означают, что на заданном участке молекулы ДНК матери и плода подвергались обогащению в равной степени.

Образец	Обогащение мишени	Число информативных SNP на участке	Выявлено специф. для плода аллелей	Процент специф. для плода аллелей	Число совместных аллелей	Число специф. для плода аллелей	Фракц. конц. ДНК плода
M6011	нет	63	6	9,5%	13	6	63,2%
M6028	нет	61	2	3,3%	6	2	50,0%
M6029	нет	69	2	2,9%	11	2	30,8%
M6043	нет	65	0	0,0%	15	0	0,00%
M6011	да	63	60	95,2%	3072	661	35,4%
M6028	да	61	60	98,4%	2570	511	33,2%
M6029	да	69	66	95,7%	3835	575	26,1%
M6043	да	65	61	93,9%	3922	776	33,0%

В. Определение генома плода с обогащением мишени

Одним из применений метода RHDO является неинвазивная пренатальная диагностика наследуемых по материнской линии генетических заболеваний. С помощью массивированного параллельного секвенирования материнской плазмы без обогащения мишени при анализе RHDO можно правильно определить, какой гаплотип матери был передан плоду, в среднем по 17 SNPs, если полнота секвенирования ДНК материнской плазмы составляет примерно 65-кратный охват генома человека. Для улучшения рентабельности этого подхода можно избирательно направить секвенирование на определенные участки генома, а затем применить анализ RHDO к данным по секвенированию. В примера мы проверили эту концепцию при прицельном секвенировании и анализе RHDO хромосомы X. Однако прицельное секвенирование и анализ RHDO может применяться ко всем хромосомам, например, аутосомам. В одном воплощении анализ RHDO, как описано выше, может применяться для прицельных воплощений.

Было привлечено 5 беременных женщин (PW226, PW263, PW316, PW370 и PW421) с единственными плодами мужского пола. Образцы периферической крови матерей собирали в пробирки с ЭДТА перед взятием образца хориальных ворсин (CVS) в первом триместре. После центрифугирования из клеток периферической крови экстрагировали ДНК с помощью набора Blood Mini Kit (Qiagen). ДНК из 3,2 мл плазмы экстрагировали с помощью набора DSP DNA Blood Mini Kit (Qiagen). Геномную ДНК матери экстрагировали из лейкоцитной пленки, а геномную ДНК плода экстрагировали из хориальных

ворсин. Образцы готовили и анализировали, как описано выше. Затем каждый образец секвенировали случайным образом, используя одну дорожку на проточной кювете фирмы Illumina.

В этом примере для вывода материнских гаплотипов для хромосомы X и того, какой гаплотип унаследован от матери, использовали генотип плода вместе с информацией по секвенированию нуклеиновых кислот матери. Для каждого SNP на хромосоме X, по которому мать гетерозиготна (т.е. это информативные SNP), наследуемый плодом аллель определяется как происходящий из материнского гаплотипа 1 (Нар I), тогда как материнский аллель, который не перешел к плоду, определяется как происходящий из материнского гаплотипа 2 (Нар II). В некоторых воплощениях для клинического применения генотип плода может быть неизвестен заранее, а гаплотипы матери можно определить или установить методами, хорошо известными специалистам, а также методами, описанными здесь. При этом хромосома X используется только для иллюстрации. При таком анализе можно использовать и другие хромосомы, например, аутосомы.

Во всех 5 случаях, описанных здесь, у них было по одному плоду мужского пола. Поскольку мужской плод наследует только одну хромосому X от матери, но никакой хромосомы X от отца, то перешедшая к плоду материнская хромосома X будет избыточно представлена в материнской плазме. Анализ RHDO проводили от pter до qter хромосомы X. Начиная с SNP, ближайшего к pter хромосомы X, анализом SPRT можно определить, что аллель из Нар I или Нар II статистически значимым образом избыточно представлен в материнской плазме. Если ни один из двух гаплотипов не представлен избыточно статистически значимым образом, то можно сложить встречаемость аллелей по следующему SNP для дальнейшего анализа методом SPRT. Можно объединять дополнительные SNPs для анализа до тех пор, пока в процессе SPRT не будет установлено, что один из гаплотипов избыточно представлен статистически значимым образом. Затем процесс классификации можно начать заново по следующему SNP.

На фиг. 38A и 38B представлены результаты классификации по SPRT для случая PW226 в качестве примера. В этом случае всего было 9 успешных классификаций по SPRT для хромосомы X. При каждой классификации по SPRT аллели из Нар I оказались избыточно представленными в образце материнской плазмы, указывая на то, что плод унаследовал от матери Нар I. Поскольку мы определили Нар I как гаплотип, содержащий аллели, перешедшие к плоду, то результаты всех этих классификаций по SPRT были правильными.

На фиг. 39 приведены результаты анализа RHDO для этих 5 случаев. Количество успешных классификаций по SPRT составляло от 1 до 9. Все классификации по SPRT

были правильными. Более высокая фракционная концентрация ДНК плода связана с большим числом классификаций, потому что аллельный дисбаланс, обусловленный присутствием ДНК плода, легче выявляется при более высокой фракционной концентрации ДНК плода. Поэтому для получения успешной классификации RHDO требуется меньше SNPs. При этом определенные участки хромосомы можно разделить на большее число блоков RHDO. Наши результаты подтверждают, что анализ RHDO может проводиться по данным массивированного секвенирования, полученным после обогащения мишени.

Наши данные также показывают, что прицельный подход является более рентабельным способом проведения анализа RHDO. Без обогащения мишени при секвенировании образцов с близкими концентрациями ДНК плода нужно примерно 5 проточных кювет (т.е. 40 дорожек для секвенирования) (фиг. 40), чтобы достичь такой же средней полноты, как у образцов, представленных на фиг. 39. А у нас при обогащении мишени секвенирование всего лишь по одной дорожке дает среднюю полноту секвенирования от 15 до 19 крат для успешной классификации RHDO.

С другой стороны, можно достичь еще более многократного охвата секвенирования при сравнительно небольших дополнительных затратах, если использовать обогащение мишени. Более высокий охват секвенирования может существенно уменьшить размер участка генома, который необходим для успешной классификации RHDO, и тем самым улучшить разрешение при анализе.

Пример 4. Обогащение мишени

С 2004 г. было известно, что циркулирующие молекулы ДНК плода в общем короче, чем ДНК матери в материнской плазме (Chan KCA et al., Clin Chem 2004, 50: 88-92; Li et al., Clin Chem 2004). Однако молекулярная основа этого наблюдения оставалась неизвестной. В настоящем исследовании мы прочитали $3,931 \times 10^9$ отрезков в исследуемом образце плазмы и использовали разбиение в 1 п.о. при биоинформатическом анализе. Размер каждой просеквенированной молекулы ДНК из плазмы выводили из геномных координат концов прочитанных с двух концов отрезков.

Для этого анализа мы сосредоточились на полиморфизмах по одному нуклеотиду (SNPs), по которым и отец, и мать гомозиготны, но по различным аллелям. По таким SNPs плод является облигатной гетерозиготой. Аллели по каждому SNP, унаследованному плодом от отца, можно использовать в качестве специфичного для плода маркера. Определяли размеры последовательностей плода (используя наследуемые по отцовской линии специфичные для плода аллели) и общих последовательностей для всего генома (фиг. 41) и отдельно для каждой хромосомы (фиг. 42A-42C).

Как оказалось, наиболее заметным отличием между ДНК плода и матери в

материнской плазме было уменьшение пика в 166 п.о. по сравнению с пиком в 143 п.о. (фиг. 41). Самые распространенные общие последовательности (главным образом материнские) имели длину в 166 п.о. Наиболее заметным отличием в распределении по размерам между ДНК плода и общей ДНК было то, что у плода проявлялось уменьшение пика ДНК в 166 п.о. (фиг. 41) и относительная заметность пика в 143 п.о. Последний, видимо, соответствует отщеплению соединительного фрагмента (линкера) в ~20 п.о. из нуклеосомы с образованием ядра нуклеосомы в ~146 п.о. (Lewin B., in *Gene IX*, Jones and Bartlett, Sudbury, 2008, pp. 757-795).

Начиная примерно со 143 п.о. и ниже, в распределении ДНК плода и общей ДНК проявлялась периодичность в 10 п.о., как при расщеплении нуклеосом нуклеазой. Это свидетельствует о том, что фрагменты ДНК в плазме возникают при воздействии апоптозных ферментов. Напротив, такой нуклеосомный профиль не проявлялся при анализе по размерам прочитанных отрезков, соответствующих не связанному с гистонами геному митохондрий (фиг. 41). Эти результаты дают ранее неизвестное молекулярное объяснение известным отличиям по размерам между ДНК плода и матери по хромосоме Y и отдельным полиморфным генетическим маркерам (Chan KCA et al., *Clin Chem* 2004, 50: 88-92; Li et al., *Clin Chem* 2004, 50: 1002-1011; US Patent Application 2005/0164241; US Patent Application 2007/0202525) и свидетельствуют о том, что такие отличия по размерам существуют по всему геному. Наиболее вероятное объяснение таких различий состоит в том, что среди циркулирующих молекул ДНК плода содержится больше таких молекул, у которых из нуклеосомы был отщеплен соединительный фрагмент в ~20 п.о.

Учитывая эти наблюдения, существует ряд способов, которыми можно провести обогащение образца ДНК плода. В одном воплощении можно использовать реагенты, которые будут предпочтительно связываться с фрагментом линкера. Такие реагенты должны предпочтительно связываться с ДНК матери по сравнению с ДНК плода в материнской плазме. Одним из примеров таких реагентов являются антитела. Одна из мишеней таких антител должна связываться с гистоном H1. Как известно, гистон H1 связывается с линкерным фрагментом. Одно из применений таких антител заключается в обогащении ДНК плода путем отрицательного отбора, т.е. путем иммунопреципитации содержащей линкер ДНК матери в материнской плазме вместе с гистоном H1. Кроме того, известно, что существует целый ряд вариантов H1, причем некоторые из них проявляют тканеспецифичные вариации в экспрессии (Sancho M et al., *PLoS Genet* 2008, 4: e1000227). Эти варианты можно было бы также использовать для отличия ДНК плода (преимущественно из плаценты) от ДНК матери (преимущественно гемопоэтической) (Lui YYN et al., *Clin Chem* 2002, 48: 421-427). Например, можно нацелиться на вариант гистона

H1, который преимущественно экспрессируется в клетках трофобластов, чтобы проводить избирательный и положительный отбор ДНК плода в материнской плазме. Такая стратегия может применяться и к другим гистоновым белкам, проявляющим тканеспецифичный, особенно специфичный к трофобластам, профиль экспрессии.

Учитывая резкий пик в 166 п.о. у ДНК матери, другая возможность обогащения ДНК плода состоит в разработке системы отрицательного отбора фрагментов ДНК длиной в 166 ± 2 п.о. Например, система на основе капиллярного электрофореза или высокоэффективной жидкостной хроматографии позволила бы точное измерение размеров и разделение молекул ДНК. Другим способом отрицательного отбора является проведение его *in silico* во время биоинформатического анализа данных по секвенированию.

Поскольку и другие разновидности ДНК в плазме, например, ДНК опухолей (Vlassov VV et al., Curr Mol Med 2010, 10: 142-165) и ДНК трансплантированных органов (Lo YMD et al., Lancet 1998, 351: 1329-1330), должны иметь такие же характеристики, как и ДНК плода в материнской плазме, то стратегии, приведенные выше в пунктах (1) и (2), можно использовать и для обогащения этих разновидностей ДНК.

В соответствии с одним воплощением предусмотрен способ дифференциального обогащения разновидностей ДНК в плазме или сыворотке человека через воздействие на линкерный фрагмент нуклеосом. В одном воплощении обогащение осуществляется путем удаления одного из следующего: ДНК матери или ДНК из гемопоэтических клеток. В другом воплощении воздействие включает реагент (как-то антитело или белок другого типа), который будет предпочтительно связываться с компонентом белка или нуклеиновой кислоты линкерного фрагмента нуклеосом. В другом воплощении реагент должен избирательно связываться с гистоном H1 или другим белком, связывающимся с линкерным фрагментом нуклеосом. В другом воплощении реагент должен связываться с материнскими или гематологическими вариантами гистона H1 или другого белка, связывающегося с линкерным фрагментом нуклеосом. В одном воплощении удаление ДНК проводится методом иммунопреципитации или связывания на твердой поверхности.

В соответствии с другим воплощением способ дифференциального обогащения ДНК плода в материнской плазме или сыворотке включает: (а) использование антитела, связывающегося с одним или несколькими компонентами линкерного фрагмента нуклеосом; (b) удаление связавшейся фракции методом иммунопреципитации или захвата на твердой поверхности; и (с) извлечение несвязавшейся фракции, содержащей ДНК плода с повышенной фракционной концентрацией.

Любые компоненты или функции программного обеспечения, описанные в данной заявке, могут выполняться в виде системной программы, исполняемой процессором с

помощью любого подходящего компьютерного языка, такого, к примеру, как Java, C++ или Perl, с использованием, к примеру, стандартных или объектно-ориентированных методов. Системная программа может храниться в виде ряда инструкций или команд на считываемом компьютером носителе для хранения и/или передачи, а подходящими носителями являются запоминающие устройства с произвольной выборкой (RAM), постоянные запоминающие устройства (ROM), магнитные носители типа жесткого диска или дискеты, или же такие оптические носители, как компакт-диск (CD) или DVD (универсальный цифровой диск), флэш-память и др. Считываемый компьютером носитель может представлять собой любую комбинацию таких устройств для хранения или передачи данных.

Такие программы могут кодироваться и передаваться с помощью сигналов-переносчиков, адаптированных для передачи через проводные, оптические и/или беспроводные сети, соответствующие различным протоколам, в том числе Интернет. При этом считываемый компьютером носитель в соответствии с воплощением настоящего изобретения может быть создан с помощью сигнала передачи данных, кодируемого такой программой. Считываемый компьютером носитель, кодируемый управляющей программой, может помещаться в одной упаковке с совместимым устройством или предоставляться отдельно от других устройств (например, загружаться через Интернет). Любой такой считываемый компьютером носитель может находиться на или в одном компьютерном программном продукте (например, жестком диске или целой компьютерной системе), а может находиться на или в различных компьютерных программных продуктах внутри системы или сети. Компьютерная система может включать монитор, принтер или другой подходящий дисплей для предоставления пользователю любых из приведенных здесь результатов.

Пример компьютерной системы представлен на фиг. 43. Приведенные на фиг. 43 подсистемы связаны друг с другом через системную шину 4375. Представлены и другие подсистемы, как-то принтер 4374, клавиатура 4378, несъемный жесткий диск 4379, монитор 4376, который подключен к адаптеру дисплея 4382, и др. Периферические и устройства ввода/вывода (I/O), подключенные к контроллеру ввода-вывода 4371, могут соединяться с компьютерной системой любыми известными способами, как-то через последовательный порт 4377. Например, последовательный порт 4377 или внешний интерфейс 4381 может использоваться для соединения самого компьютера с глобальной сетью типа Интернет, устройством ввода с мыши или сканером. Внутреннее соединение через системную шину дает возможность центральному процессору 4373 соединяться с каждой подсистемой и контролировать исполнение инструкций из системной памяти 4372

или несъёмного жесткого диска 4379, а также обмен информации между подсистемами. Системная память 4372 и/или несъёмный жесткий диск 4379 могут воплощать считываемый компьютером носитель. Любая из приведенных здесь величин может означать вывод данных из одного компонента на другой компонент или же на пользователя.

Компьютерная система может включать несколько одинаковых компонентов или подсистем, например, соединенных через внешний интерфейс 4381 или через внутренний интерфейс. В некоторых воплощениях компьютерные системы, подсистемы или машины могут обмениваться информацией через сеть. В таких случаях один компьютер может рассматриваться как клиент, а другой – как сервер, при этом каждый может входить в состав одной и той же компьютерной системы. И клиент, и сервер может включать множественные системы, подсистемы или компоненты.

Конкретные детали определенных воплощений могут сочетаться любым подходящим образом или отличаться от тех, что представлены и описаны здесь, не отходя от сущности и не выходя за рамки воплощений изобретения.

Вышеприведенное описание типичных воплощений изобретения представлено в целях иллюстрации и описания. Оно не может быть исчерпывающим или ограничивать изобретение в точности описанной формой, причем возможны многие модификации и вариации в свете вышеприведенных положений. Воплощения были выбраны и описаны с тем, чтобы наилучшим образом разъяснить принципы изобретения и его практическое применение и тем самым дать возможность другим специалистам наилучшим образом использовать изобретение в различных воплощениях и с различными модификациями, подходящими для его конкретного предназначения.

Все приведенные здесь публикации, патенты и патентные заявки настоящим включены путем ссылки во всей полноте на все случаи.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ идентификации мутаций *de novo* в геноме еще не рожденного плода беременной женщины, причем у плода есть отец и мать, являющаяся указанной беременной женщиной, который включает:

получение результатов секвенирования множества молекул нуклеиновых кислот из биологического образца, взятого у беременной женщины, где биологический образец содержит смесь нуклеиновых кислот матери и плода;

определение локализации каждой из просеквенированных молекул нуклеиновых кислот в геноме человека;

для каждой из по крайней мере части локализаций – определение одной или нескольких материнских последовательностей в материнском геноме и одной или нескольких отцовских последовательностей в данной локализации;

идентификацию компьютерной системой первой последовательности в множестве молекул нуклеиновых кислот, которая отсутствует у выявленных материнских или отцовских последовательностей в первой локализации;

определение первой фракционной концентрации первой последовательности из первой локализации в биологическом образце;

определение второй фракционной концентрации второй последовательности из второй локализации в биологическом образце, вторая последовательность во второй локализации унаследована плодом от отца, где вторая последовательность присутствует в отцовском геноме во второй локализации, но отсутствует в материнском геноме во второй локализации; и

классификацию первой последовательности как мутации *de novo*, если первая и вторая фракционная концентрация примерно одинаковы.

2. Способ по п.1, где секвенирование молекул нуклеиновых кислот представляет собой секвенирование по спаренным концам, которое обеспечивает пару последовательностей, соответствующих обоим концам молекулы нуклеиновой кислоты, и где определение локализации молекул нуклеиновых кислот в геноме человека включает картирование указанных обоих концов в контрольном геноме человека.

3. Способ по п.1, где определение локализации молекулы нуклеиновой кислоты в геноме человека включает картирование молекулы нуклеиновой кислоты в геноме человека, где картирование не требует точного совпадения последовательностей.

4. Способ по п.1, где по крайней мере для одной из локализаций, отцовская последовательность определяется из образца от отца.

5. Способ по п.1, где по крайней мере для одной из локализаций, материнская

последовательность определяется из биологического образца.

6. Способ по п.1, где идентификация первой последовательности в первой локализации включает:

сравнение последовательностей молекул нуклеиновых кислот, которые идентифицировали как происходящих из первой локализации, с одной или несколькими материнскими последовательностями и с одной или несколькими отцовскими последовательностями в первой локализации; и

идентификацию первой последовательности из последовательностей как не совпадающую ни с одной из одной или нескольких материнских последовательностей и одной или нескольких отцовских последовательностей в первой локализации.

7. Способ по п. 6, далее содержащий:

подсчет количества молекул нуклеиновых кислот, демонстрирующих наличие первой последовательности; и

выставление требования, что если количество выше определенного количества для первой последовательности, то это считается мутацией *de novo*.

8. Способ по п.1, где определение первой фракционной концентрации включает:

определение доли молекул нуклеиновых кислот в первой локализации, которые демонстрируют наличие первой последовательности.

9. Способ по п. 1, где вторая последовательность находится на хромосоме Y или же представляет собой генетический полиморфизм или полиморфизм по одному нуклеотиду или полиморфизм типа вставки-делеции.

10. Способ по п. 1, где мать гомозиготна по первому аллелю во второй локализации и отец гомозиготен по второму аллелю во второй локализации, и второй аллель соответствует второй последовательности.

11. Способ по п. 1, далее содержащий:

определение, что первая и вторая фракционные концентрации примерно одинаковы, посредством:

подсчета разницы между первой и второй фракционными концентрациями; и

сравнение разницы с пороговым значением.

12. Способ идентификации мутаций *de novo* в геноме еще не рожденного плода беременной женщины, причем у плода есть отец и мать, являющаяся указанной беременной женщиной, где отец имеет отцовский геном и мать имеет материнский геном, который включает:

получение результатов секвенирования множества молекул нуклеиновых кислот из биологического образца, взятого у беременной женщины, причем биологический образец

содержит смесь нуклеиновых кислот матери и плода;

определение локализации каждой из просеквенированных молекул нуклеиновых кислот в геноме человека;

для каждой из по крайней мере части локализаций – определение одной или нескольких материнских последовательностей в материнском геноме и одной или нескольких отцовских последовательностей в отцовском геноме в данной локализации;

идентификацию компьютерной системой первой последовательности во множестве молекул нуклеиновых кислот, которая отсутствует у определенных материнских или отцовских последовательностей в первой локализации;

определение первой фракционной концентрации первой последовательности из первой локализации в биологическом образце;

определение второй фракционной концентрации нуклеиновых кислот плода из второй локализации в биологическом образце, которые проявляют эпигенетический статус специфичный для плода, где эпигенетический статус произошедших от плода и произошедших от матери молекул нуклеиновых кислот в биологическом образце отличается во второй локализации; и

классификацию первой последовательности в первой локализации как мутации *de novo*, если первая и вторая фракционная концентрация примерно одинаковы.

13. Способ по п.12, где по крайней мере для одной из локализаций отцовская последовательность определяется из образца от отца.

14. Способ по п.12, где по крайней мере для одной из локализаций материнская последовательность определяется из биологического образца.

15. Способ по п.12, где идентификация первой последовательности в первой локализации включает:

сравнение последовательностей молекул нуклеиновых кислот, которые идентифицировали как происходящие из первой локализация с одной или несколькими материнскими последовательностями и с одной или несколькими отцовскими последовательностями в первой локализации; и

идентификацию первой последовательности из последовательностей в первом локализации как не совпадающую ни с одной из одной или нескольких материнских последовательностей и одной или нескольких отцовских последовательностей.

16. Способ по п.15, далее содержащий:

подсчет количества молекул нуклеиновых кислот, демонстрирующих наличие первой последовательности; и

выставление требования, что если количество выше определенного количества для первой последовательности, то это считается мутацией *de novo*.

17. Способ по п.12, где определение второй фракционной концентрации включает:

определение доли молекул нуклеиновых кислот во второй локализации, которые демонстрируют специфичный для плода эпигенетический статус.

18. Способ по п. 12, далее содержащий:

определение, что первая и вторая фракционные концентрации примерно одинаковы посредством:

подсчета разницы между первой и второй фракционными концентрациями; и

сравнение разницы с пороговым значением.

19. Способ по п. 12, в котором различия в эпигенетическом статусе отражаются различиями в профиле метилирования ДНК.

20. Способ по п. 19, в котором различия в профиле метилирования ДНК затрагивают ген 1A из семейства доменов ассоциации с RAS (RASSF1A) или ген лигазы синтетазы голокарбоксилазы (биотин-(пропионил-КоА-карбоксилазы (АТФ-гидролизующей)) (HLCS).

21. Способ определения фракционной концентрации ДНК плода в биологическом образце, взятом у беременной женщины, причем у плода есть отец и мать, являющаяся указанной беременной женщиной, при этом биологический образец содержит смесь нуклеиновых кислот матери и плода, который включает:

анализ множества молекул нуклеиновых кислот из биологического образца, причем анализ молекул нуклеиновых кислот включает:

определение локализации молекулы нуклеиновой кислоты в геноме человека и

определение соответствующего аллеля молекулы нуклеиновой кислоты;

определение компьютерной системой одного или нескольких первых локусов, где геном плода гетерозиготен по каждому из первых локусов, так что геном плода содержит соответствующий первый и соответствующий второй аллель в этом первом локусе, а геном матери гомозиготен по каждому из первых локусов, так что геном матери содержит два соответствующих вторых аллеля в этом первом локусе, причем первый аллель отличается от второго аллеля, при этом определение конкретного локуса как одного из одного или нескольких первых локусов включает:

определение порогового значения для числа предсказанной встречаемости соответствующего первого аллеля в данном локусе, причем пороговое значение предсказывает, является ли геном матери гомозиготным, а геном плода гетерозиготным,

при этом пороговое значение определяется на основании статистического распределения числа встречаемости для различных комбинаций гомозиготности и гетерозиготности в данном локусе;

исходя из анализа множества молекул нуклеиновых кислот, выявление соответствующего первого аллеля и соответствующего второго аллеля в данном локусе;

определение фактического числа встречаемости соответствующего первого аллеля, исходя из анализа множества молекул нуклеиновых кислот из биологического образца;

определение данного локуса как одного из первых локусов, если число фактической встречаемости будет меньше порогового значения; и

по меньшей мере для одного из первых локусов:

определение первого числа встречаемости P соответствующего первого аллеля и второго числа встречаемости Q соответствующего второго аллеля; и

вычисление фракционной концентрации, исходя из первого и второго чисел.

22. Способ по п. 21, в котором фракционная концентрация определяется как $2 \times p / (p + q)$.

23. Способ по п. 21, в котором P и Q определяются для множества первых локусов, причем фракционная концентрация f определяется как:

$$f = \frac{\sum_{i=1}^n 2p_i}{\sum_{i=1}^n (p_i + q_i)},$$

где p_i – первое число для i -ого первого локуса, а q_i – второе число для i -ого первого локуса.

24. Способ по п. 21, в котором определение порогового значения включает определение статистического распределения для максимальной и минимальной фракционной концентрации.

25. Способ по п. 21, где анализ молекул нуклеиновых кислот включает применение в отношении по крайней мере части молекул нуклеиновых кислот по меньшей мере одной методики, выбранной из группы, включающей массивное параллельное секвенирование, микроэррей, гибридизацию, ПЦР в реальном времени, цифровой ПЦР и масс-спектрометрию.

26. Способ по п. 21, где статистическое распределение включает распределение Пуассона.

27. Способ определения первой доли генома плода, которая была просеквенирована в биологическом образце, взятого у беременной женщины, причем у плода есть отец и мать, являющаяся указанной беременной женщиной, при этом биологический образец содержит смесь нуклеиновых кислот матери и плода, который

включает:

получение результатов секвенирования множества молекул нуклеиновых кислот из биологического образца, взятого у беременной женщины;

анализ результатов секвенирования по молекулам нуклеиновых кислот, который включает:

определение локализации молекулы нуклеиновой кислоты в геноме человека и

определение соответствующего аллеля молекулы нуклеиновой кислоты;

определение первого множества локусов, где геном плода гетерозиготен по каждому локусу из первого множества, так что геном плода содержит соответствующий первый и второй аллель в этом локусе, а геном матери гомозиготен по каждому локусу из первого множества, так что геном матери содержит два соответствующих вторых аллеля в этом локусе, причем первый аллель отличается от второго аллеля;

определение компьютерной системой второй доли локусов из первого множества локусов, в которых выявляется соответствующий первый аллель по результатам секвенирования; и

исходя из этой доли, определение доли генома плода, которая была просеквенирована в биологическом образце.

28. Способ по п.27, где идентификация молекулы нуклеиновой кислоты в геноме человека включает картирование молекулы нуклеиновой кислоты в геноме человека.

29. Способ по п. 27, где определение первого множества локусов включает:

определение того, что отцовский геном гомозиготен по соответствующему первому аллелю в каждом локусе из первого множества локусов, а материнский геном гомозиготен по соответствующему второму аллелю в том же самом локусе.

30. Способ по п. 27, где определение конкретного локуса как одного из первого множества локусов включает:

определение порогового значения для числа предсказанной встречаемости соответствующего первого аллеля в данном локусе, причем пороговое значение предсказывает, является ли геном матери гомозиготным, а геном плода гетерозиготным, при этом пороговое значение определяется на основании статистического распределения числа встречаемости для различных комбинаций гомозиготности и гетерозиготности в данном локусе;

исходя из анализа результатов секвенирования, выявление соответствующего первого и второго аллеля в данном локусе;

определение фактического числа встречаемости соответствующего первого аллеля,

исходя из секвенирования множества молекул нуклеиновых кислот из биологического образца; и

определение того, что геном плода гетерозиготен по соответствующему первому и второму аллелю, а геном матери гомозиготен по соответствующему второму аллелю, если фактическое число встречаемости будет меньше порогового значения.

31. Способ по п. 30, в котором статистическое распределение зависит от фракционной концентрации молекул нуклеиновых кислот из биологического образца, происходящих от плода.

32. Способ по п. 31, в котором статистическое распределение дополнительно зависит от количества молекул нуклеиновых кислот, соответствующих данному локусу, в множестве.

33. Способ по п. 27, в котором соответствующие первые аллели по меньшей мере в двух локусах отличаются из первого множества друг от друга.

34. Способ по п.27, где первая доля равна второй доле.

35. Способ по п.27, где первая доля является диапазоном, а вторая доля является одним из концов диапазона.

36. Способ по п. 27, где первое множество локусов находятся на первой хромосоме, и где первая доля является долей первой хромосомы, которая была просеквенирована из биологического образца.

37. Способ по п.36, далее содержащий;

определение доли каждой хромосомы в геноме плода которая была просеквенирована из биологического образца.

38. Способ определения фракционной концентрации ДНК плода в биологическом образце, взятом у беременной женщины, причем у плода есть отец и мать, являющаяся данной беременной женщиной, при этом биологический образец содержит смесь нуклеиновых кислот матери и плода, который включает:

обогащение биологического образца, взятого у беременной женщины, по нуклеиновым кислотам из искомого участка;

секвенирование множества молекул нуклеиновых кислот из обогащенного биологического образца, причем секвенирование специфично для искомого участка, при этом результаты секвенирования подвергаются анализу для:

определения локализации молекулы нуклеиновой кислоты на искомом участке в геноме человека; и

определения соответствующего аллеля молекулы нуклеиновой кислоты;

определение множества первых локусов, причем геном плода гетерозиготен по

каждому из первых локусов, так что геном плода содержит соответствующий первый и второй аллель в этом первом локусе, а геном матери гомозиготен по каждому из первых локусов, так что геном матери содержит два соответствующих вторых аллеля в этом первом локусе, причем первый аллель отличается от второго аллеля;

для множества первых локусов:

определение первого числа встречаемости P соответствующего первого аллеля и второго числа встречаемости Q соответствующего второго аллеля; и

определение фракционной концентрации, исходя из первого и второго чисел.

39. Способ по п. 38, где определение множества первых локусов включает:

определение того, что отцовский геном гомозиготен по соответствующему первому аллелю в каждом локусе из первого множества, а материнский геном гомозиготен по соответствующему второму аллелю в том же самом локусе.

40. Способ по п. 38, в котором определение конкретного локуса как одного из множества первых локусов включает:

определение порогового значения для числа предсказанной встречаемости соответствующего первого аллеля в данном локусе, причем пороговое значение предсказывает, является ли геном матери гомозиготным, а геном плода гетерозиготным, при этом пороговое значение определяется на основании статистического распределения числа встречаемости для различных комбинаций гомозиготности и гетерозиготности в данном локусе;

исходя из анализа результатов секвенирования, выявление соответствующего первого и второго аллеля в данном локусе;

определение числа фактической встречаемости соответствующего первого аллеля, исходя из секвенирования множества молекул нуклеиновых кислот из биологического образца; и

определение того, что данный локус является одним из первых локусов, если число фактической встречаемости будет меньше порогового значения.

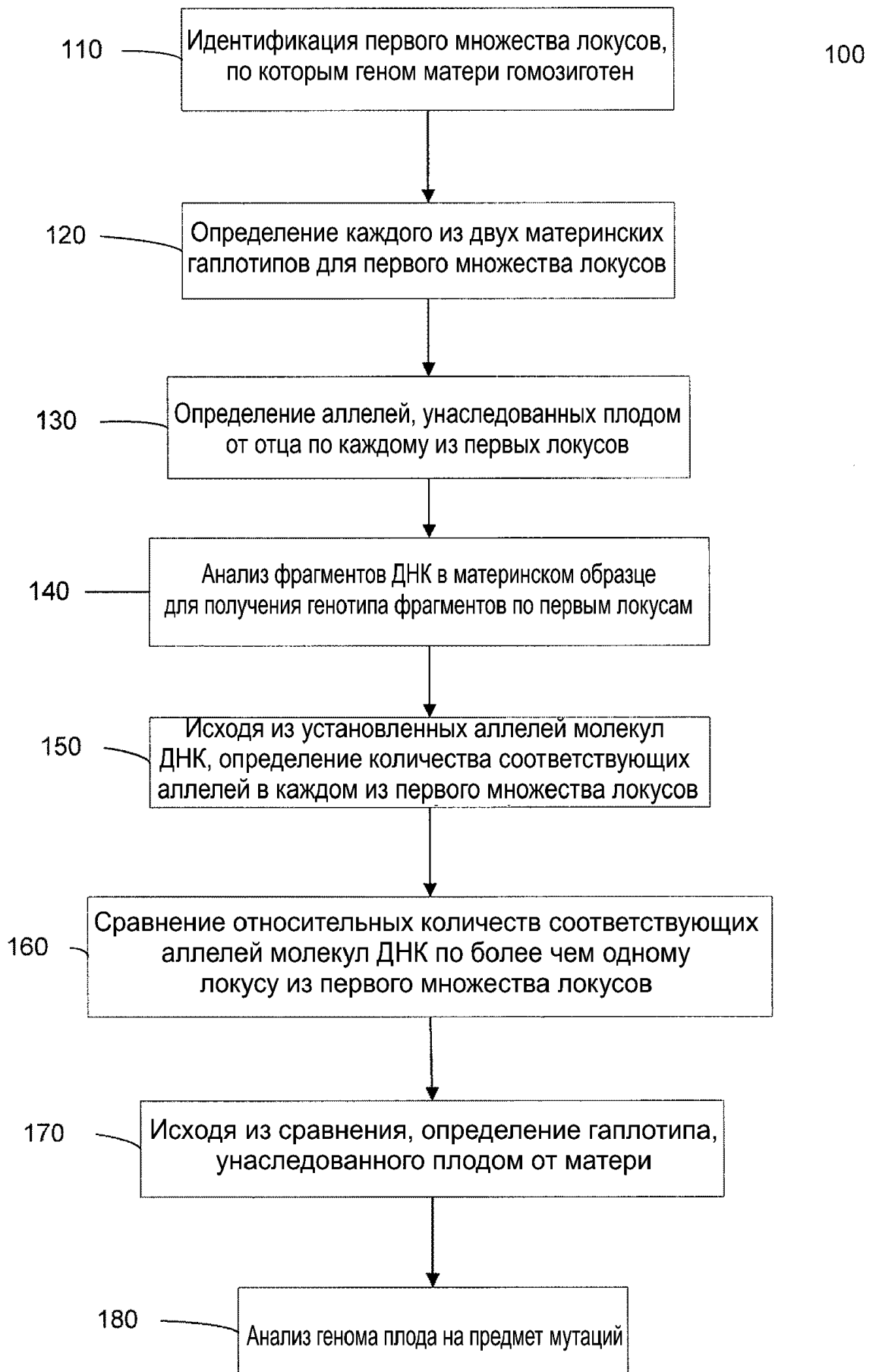
41. Способ по п. 40, где определение порогового значения включает определение статистического распределения для максимальной и минимальной фракционной концентрации.

42. Способ по п.38, где обогащение биологического образца включает прицельное секвенирование при помощи захвата в растворимой фазе, захвата на микроматрице или прицельной амплификации.

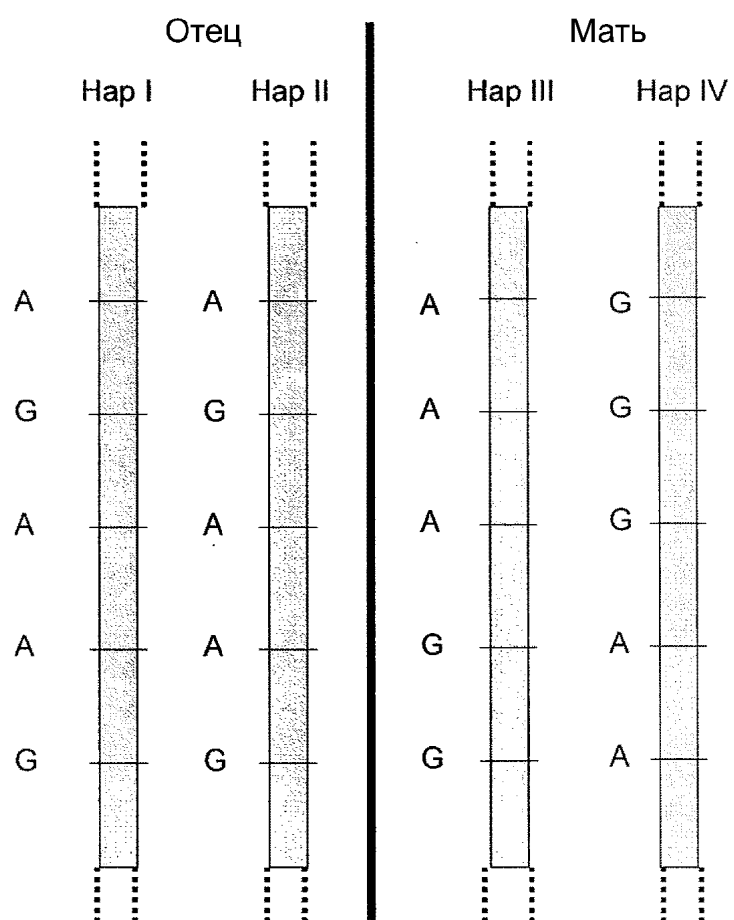
43. Компьютерный программный продукт, включающий машиночитаемый носитель, хранящий инструкции для управления процессором при выполнении способа по

любому из пп. 1-40.

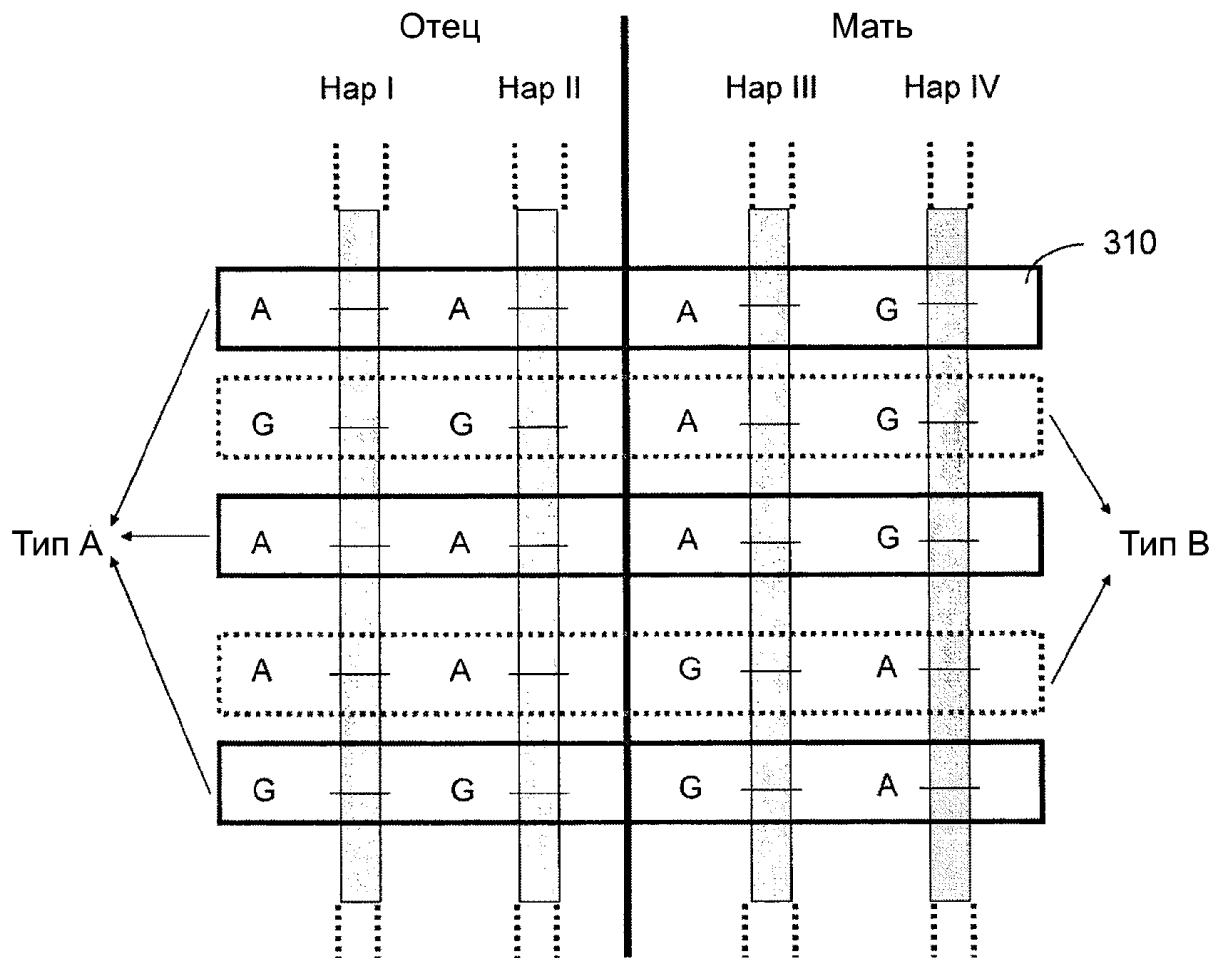
Фиг. 1



Фиг. 2

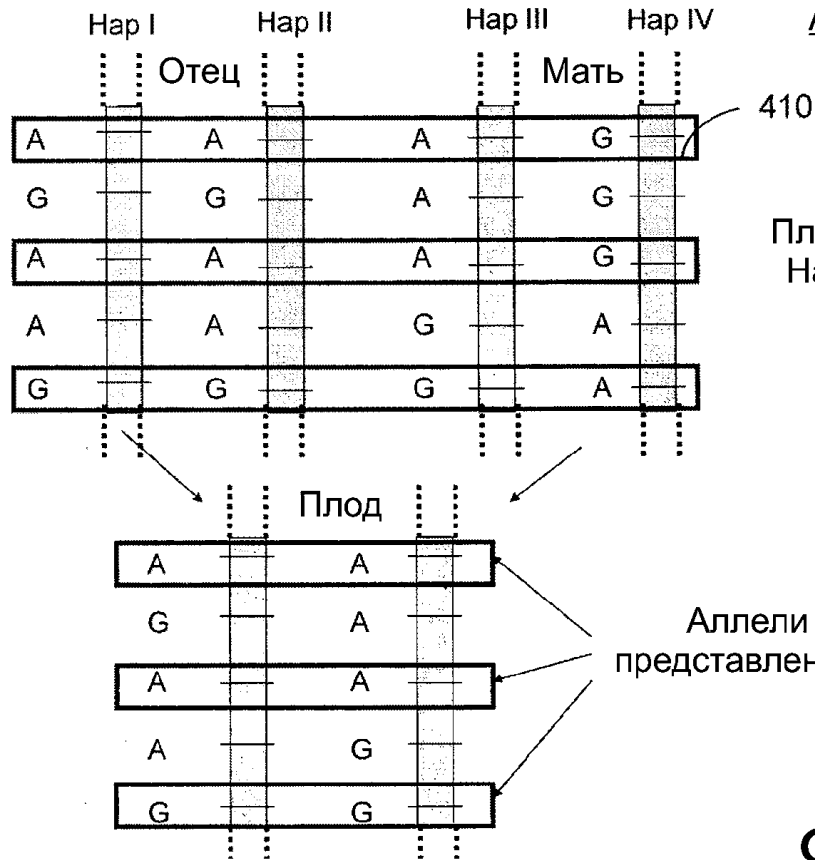


Фиг. 3



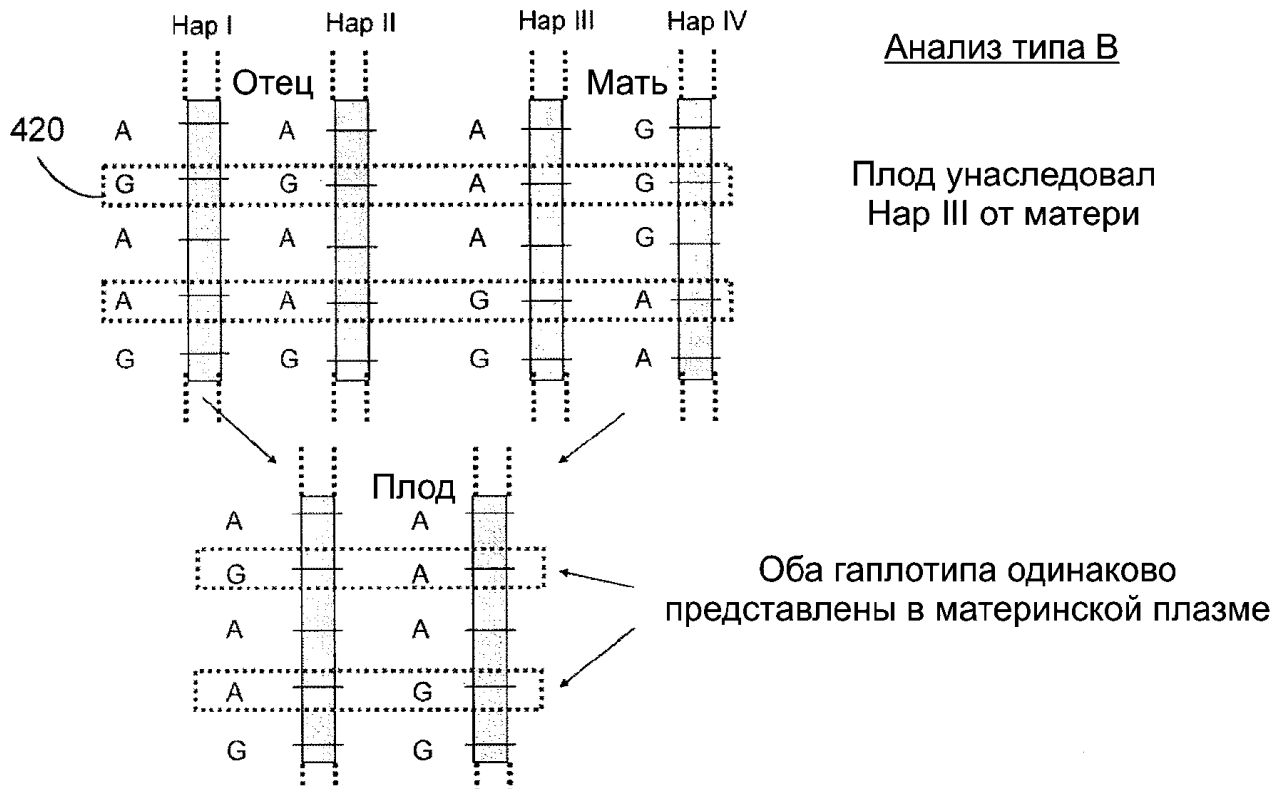
Фиг. 4А

Анализ типа А



Фиг. 4В

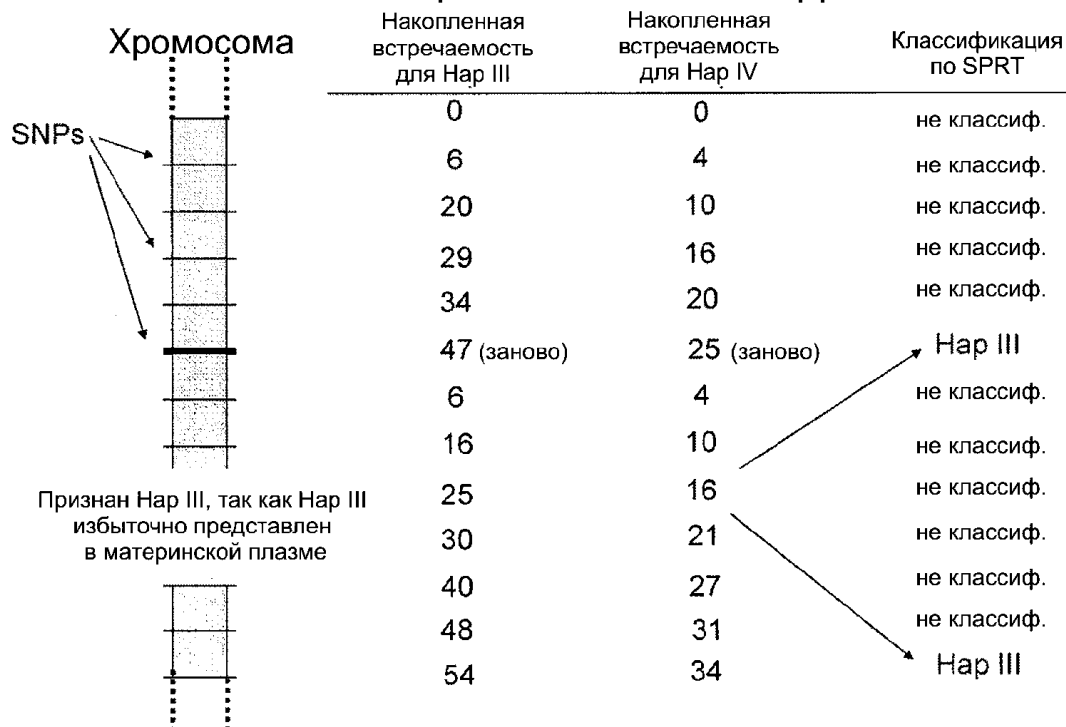
Анализ типа В



Фиг. 5А

Анализ по SPRT на RHDO

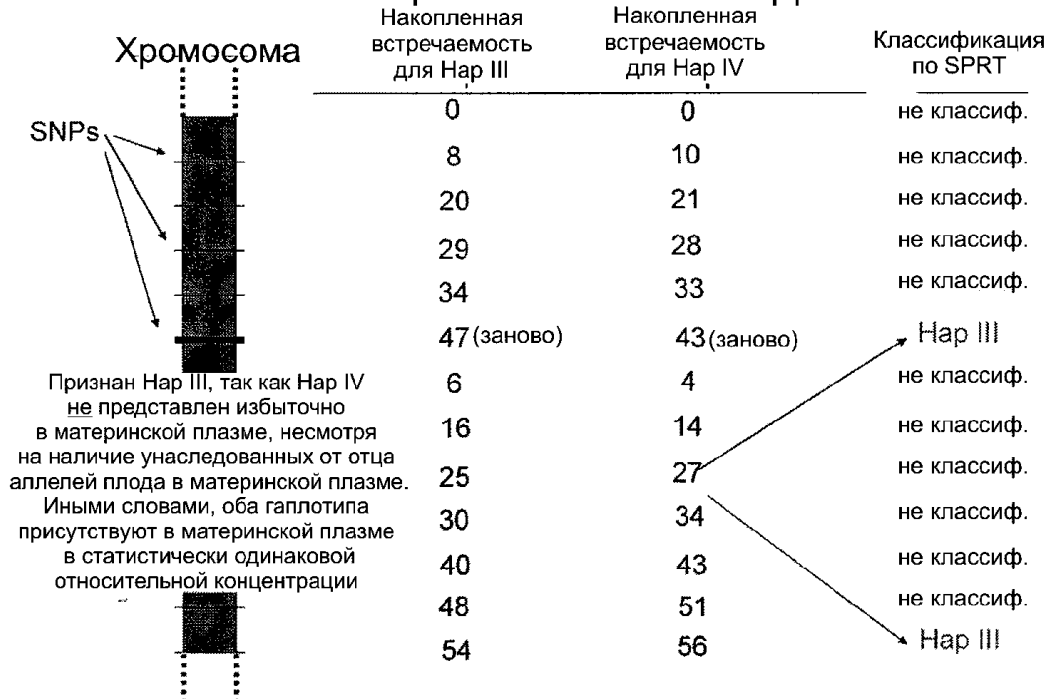
Анализ типа А: отцовские аллели идентичны Har III



Фиг. 5В

Анализ по SPRT на RHDO

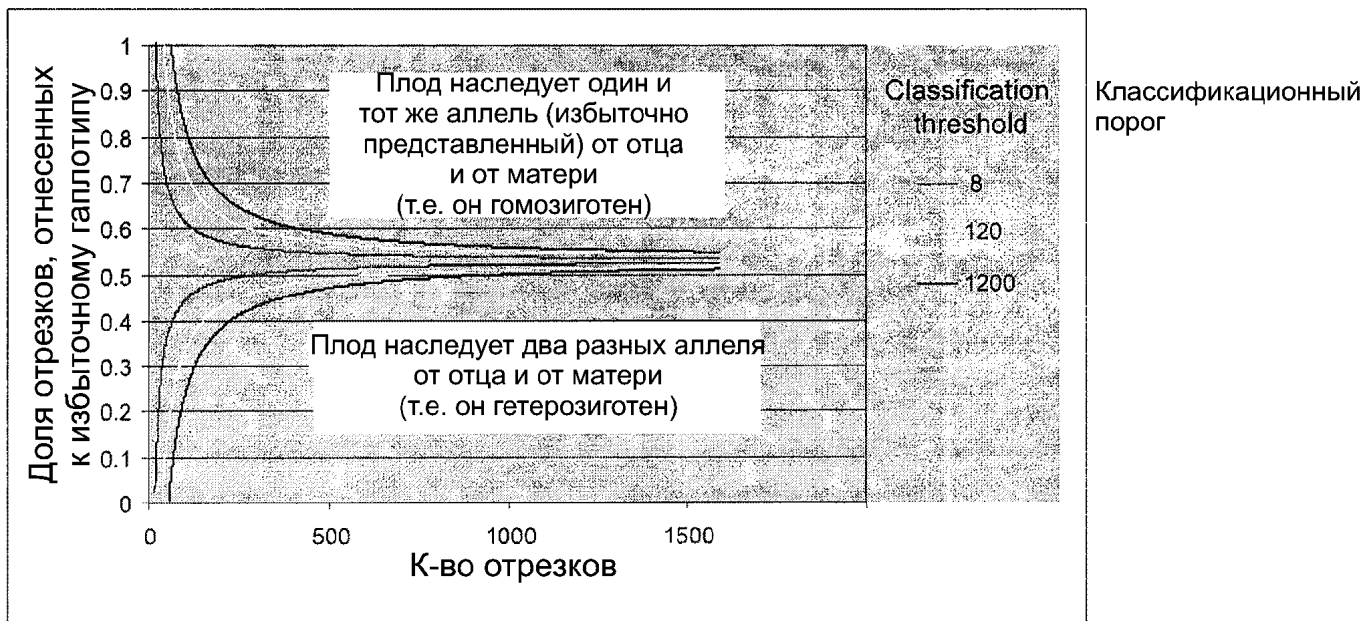
Анализ типа В: отцовские аллели идентичны Har IV



Фиг. 6

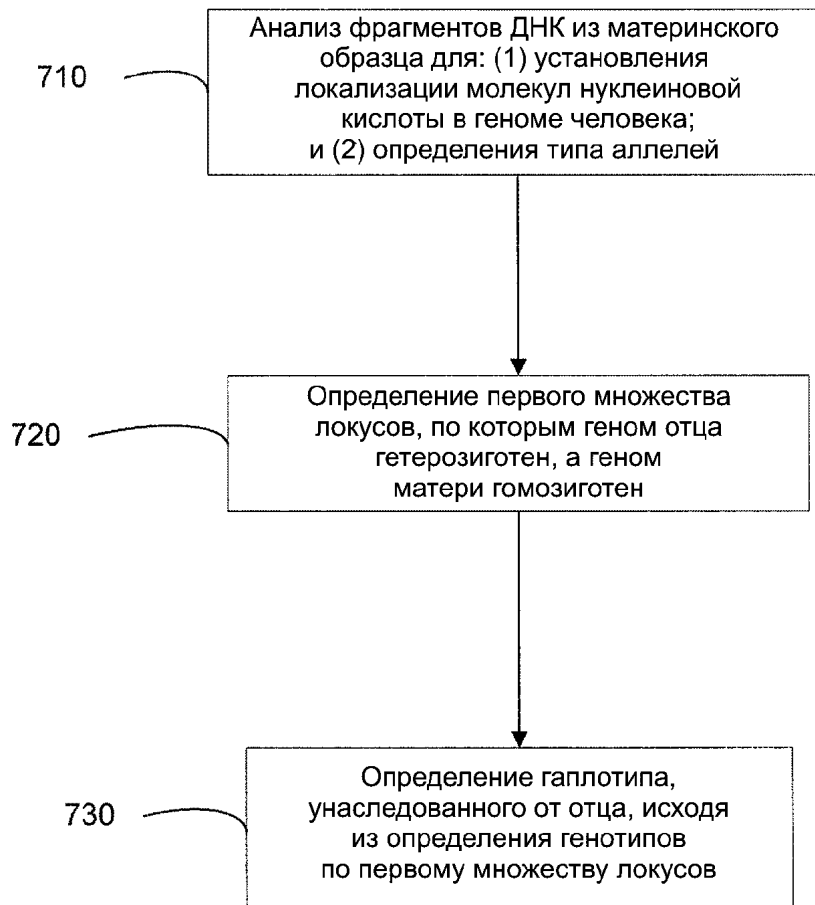
Эффект изменения классификационного порога на кривых SPRT

Отцовские аллели идентичны Нар III



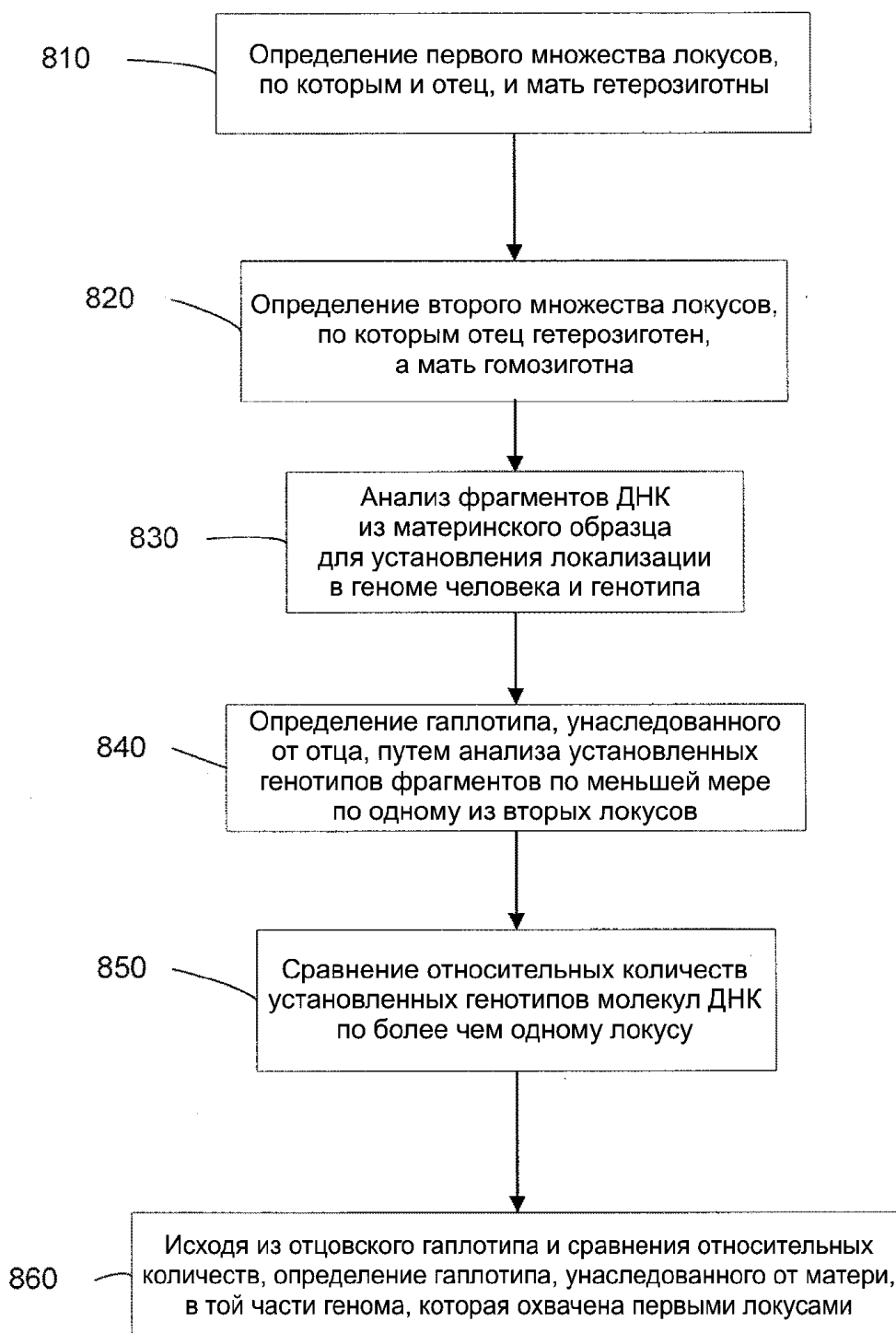
Фиг. 7

700

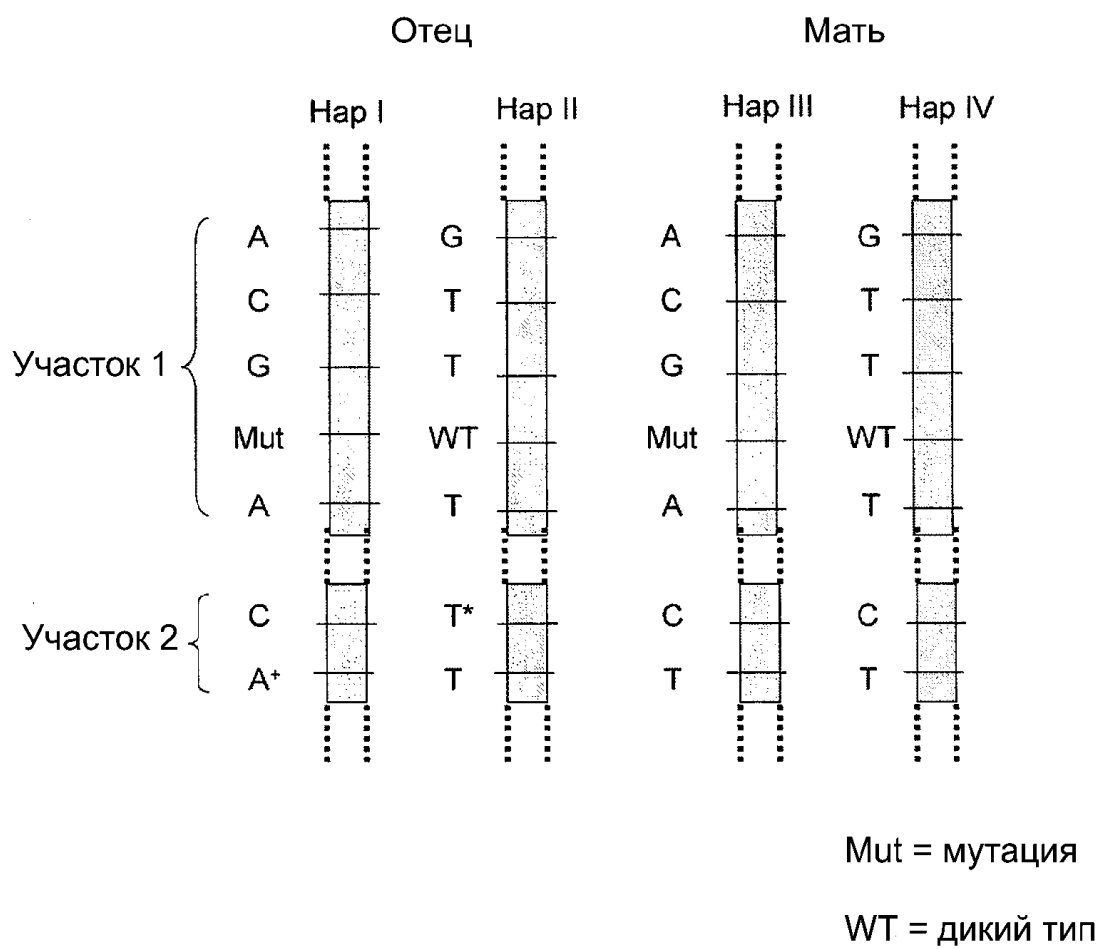


Фиг. 8

800

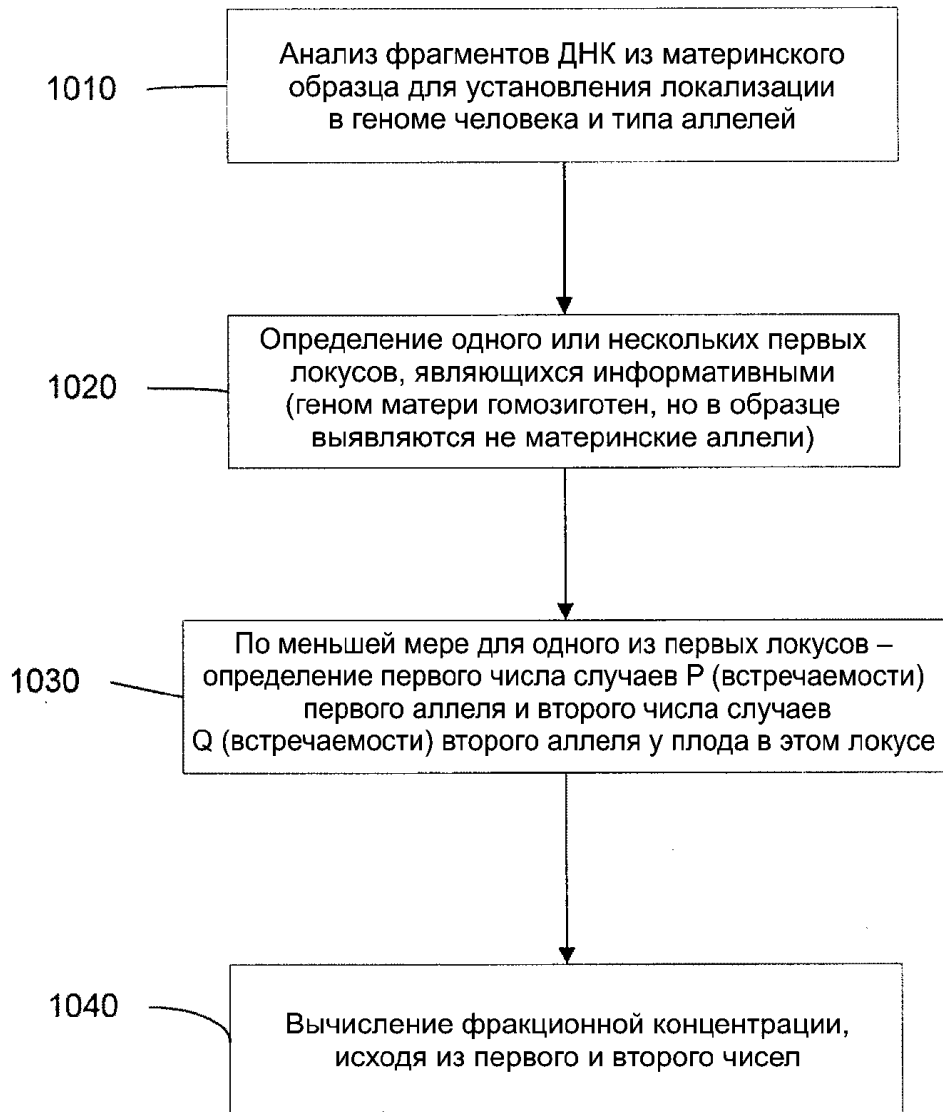


Фиг. 9



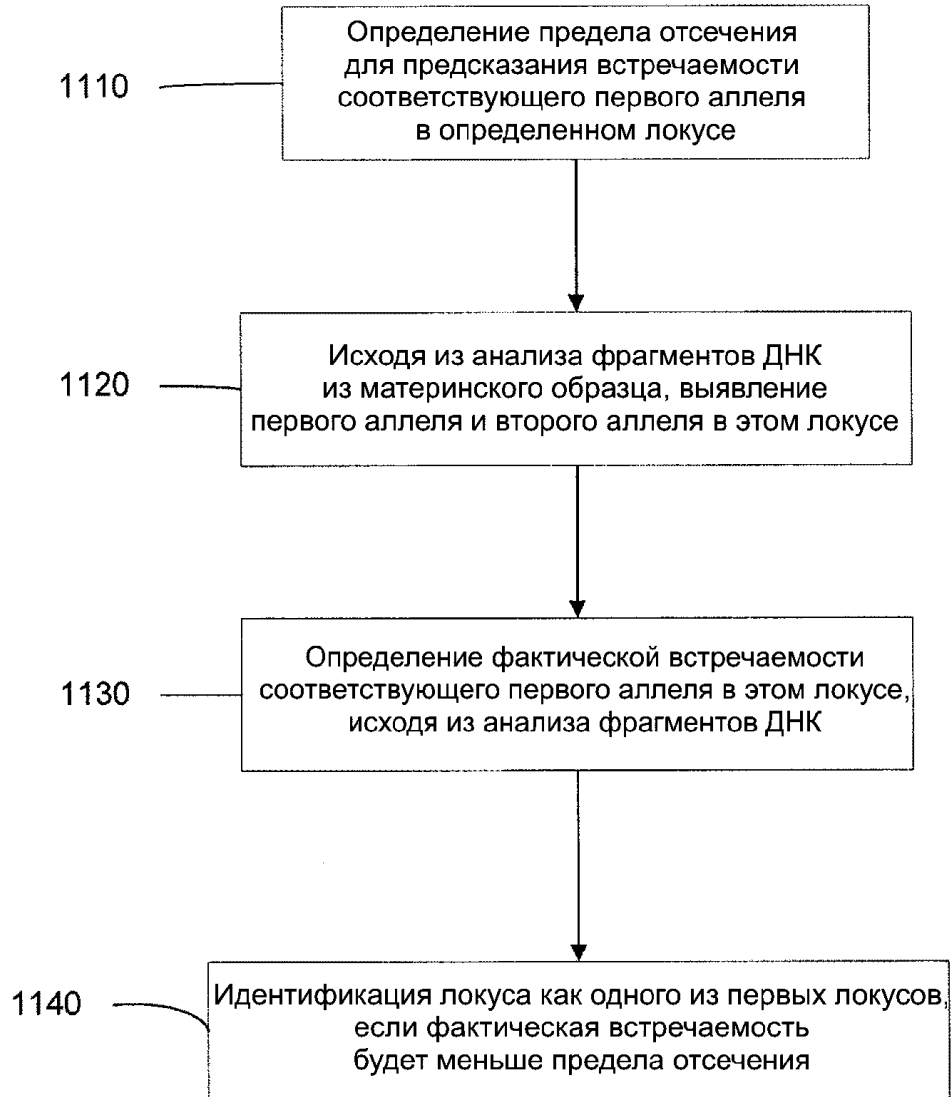
Фиг. 10

1000

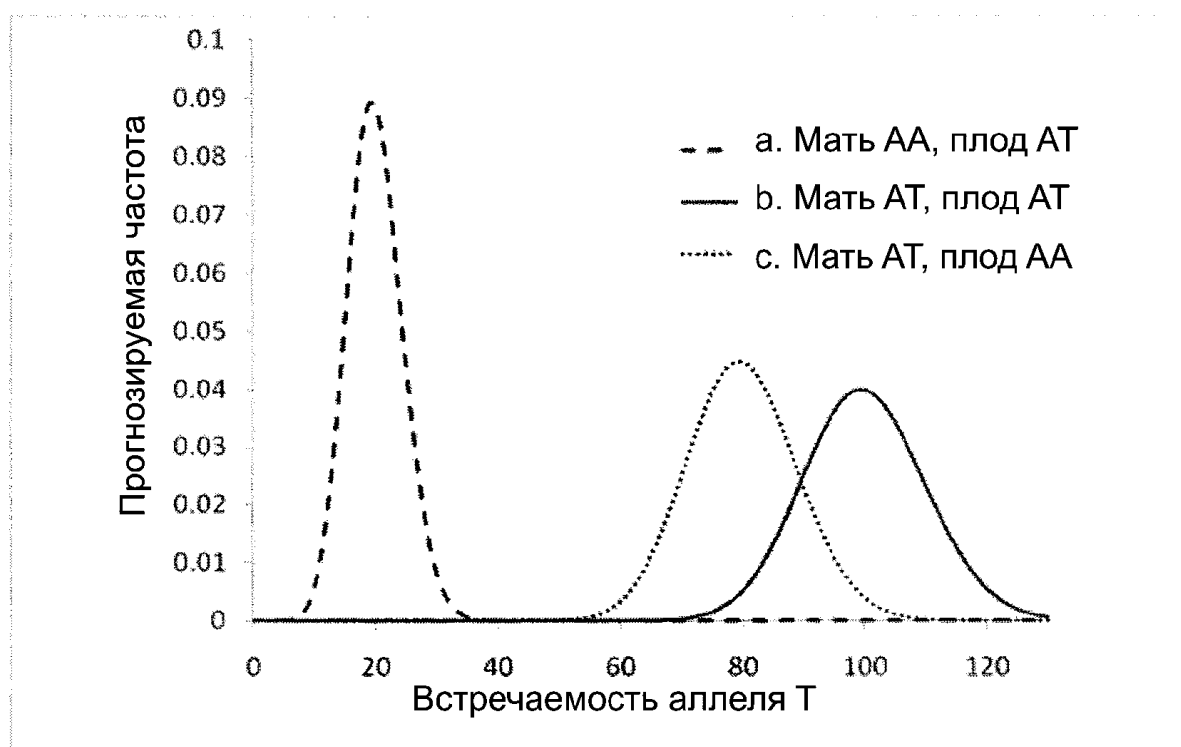


Фиг. 11

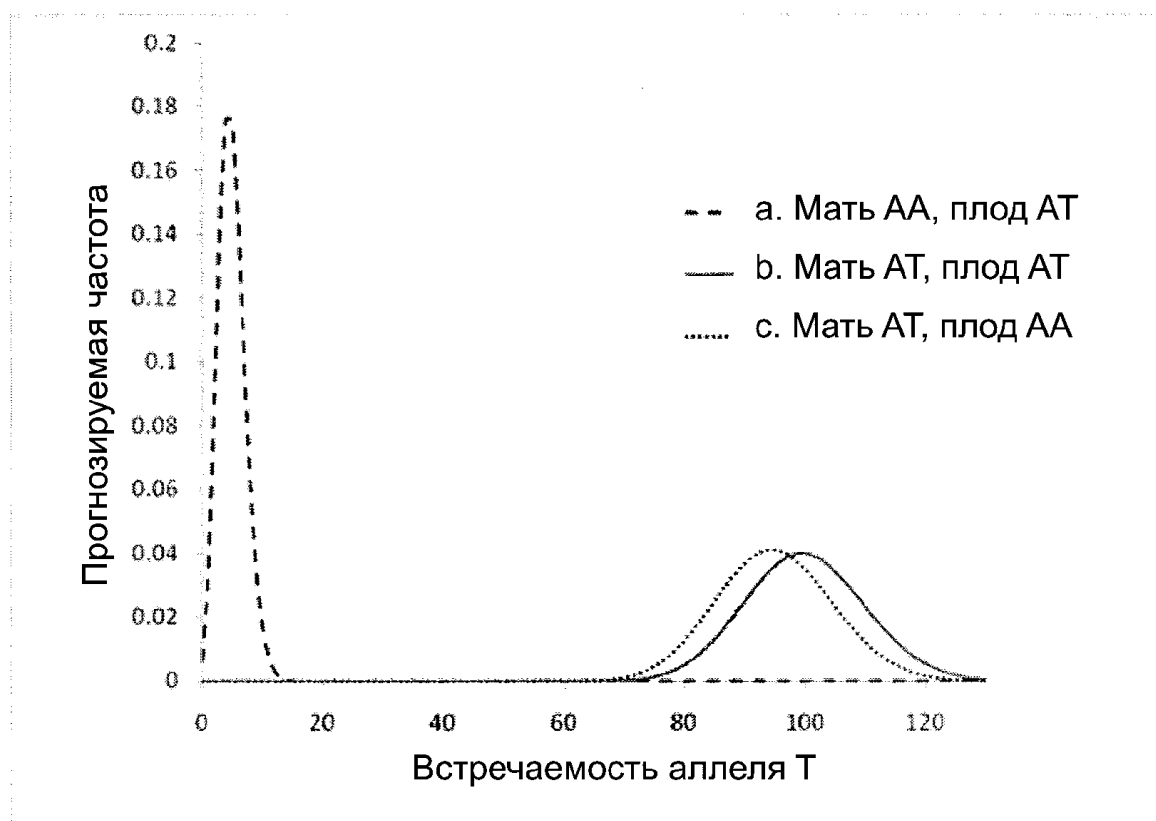
1100



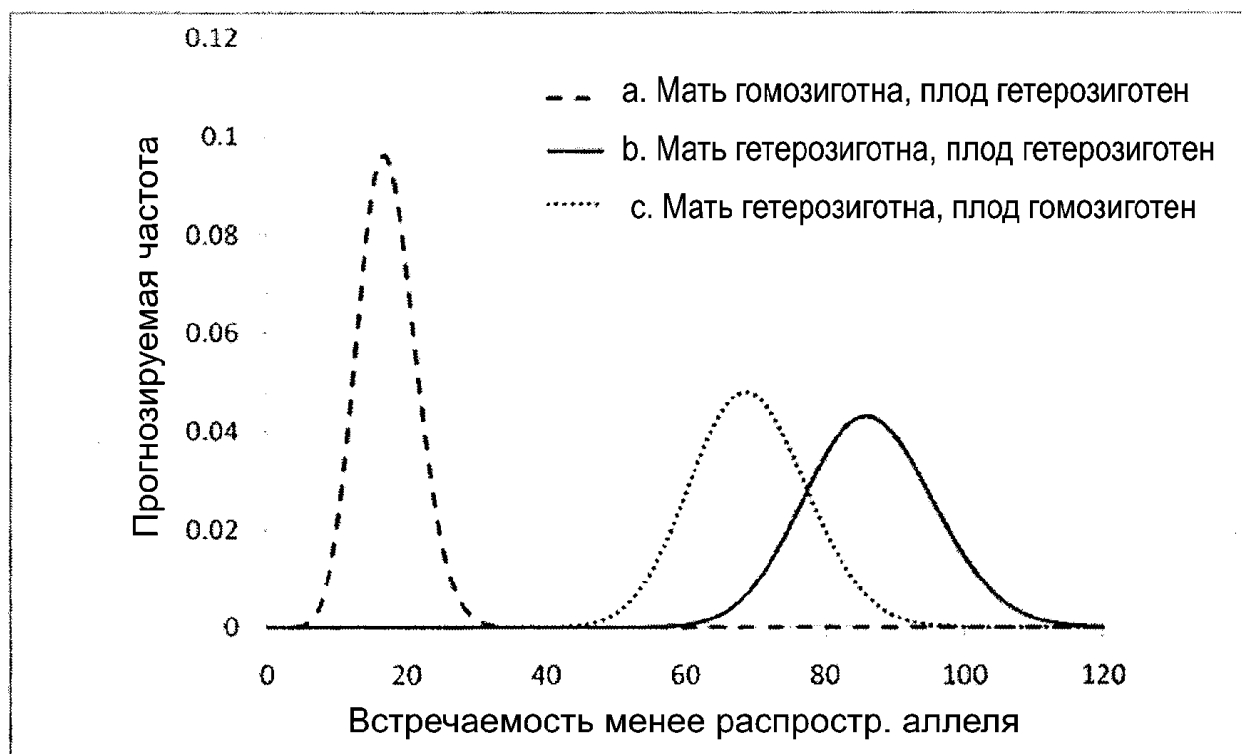
Фиг. 12А



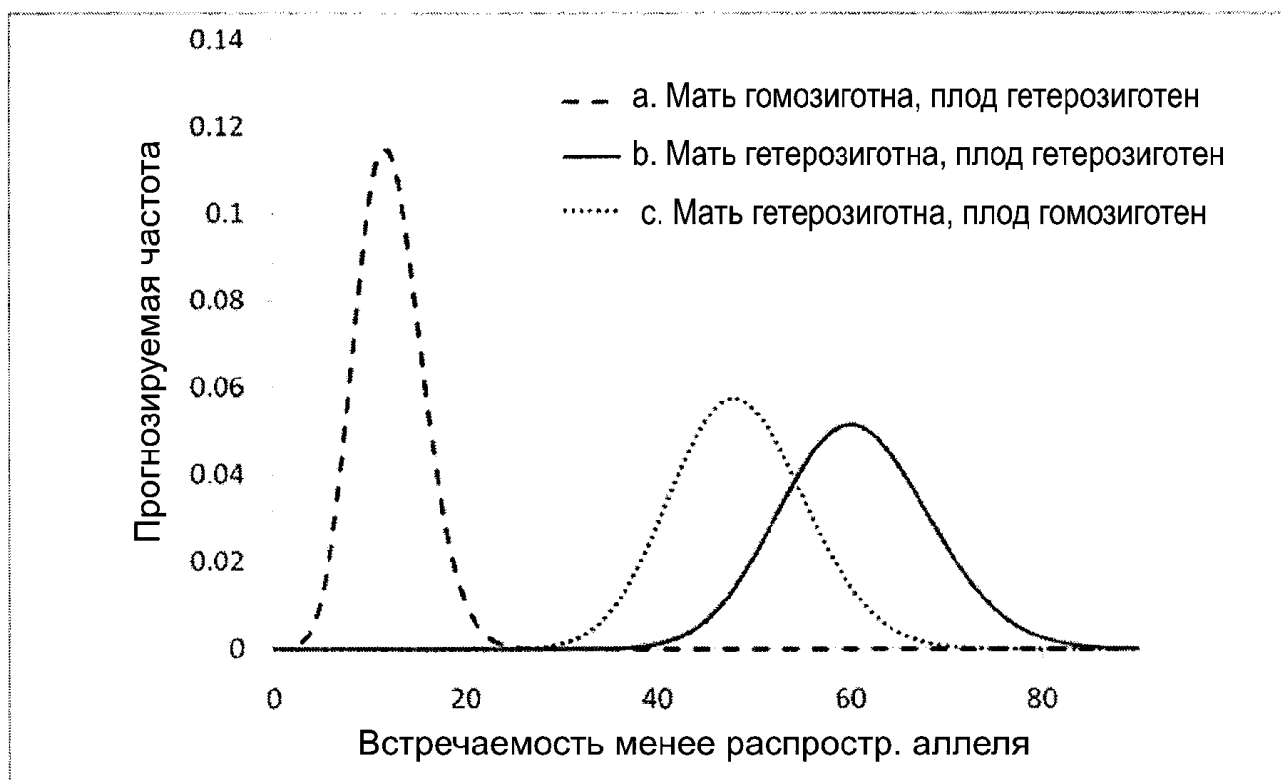
Фиг. 12В



Фиг. 13А



Фиг. 13В



Фиг. 14



Фиг. 15A

Возможные гаплотипы в популяции

Нар А

A
C
G
A
T
A
C
G
A
T
T
A
A
G

Нар В

C
G
G
T
A
G
C
C
T
T
T
A
T
C
G

Нар С

C
C
A
T
T
G
T
C
A
G
T
T
C
A

Гаплотипы отца

Нар В

C
G
G
T
A
G
C
C
T
T
A
T
C
G

Нар С

C
C
A
T
T
G
T
C
A
G
T
T
C
A

Гаплотипы матери

Нар А

A
C
G
A
T
A
C
G
A
T
T
A
A
G

Нар В

C
G
G
T
A
G
C
C
T
T
A
T
C
G

Плод наследует Нар С
от отца и Нар А от матери

Гаплотипы плода

Нар С

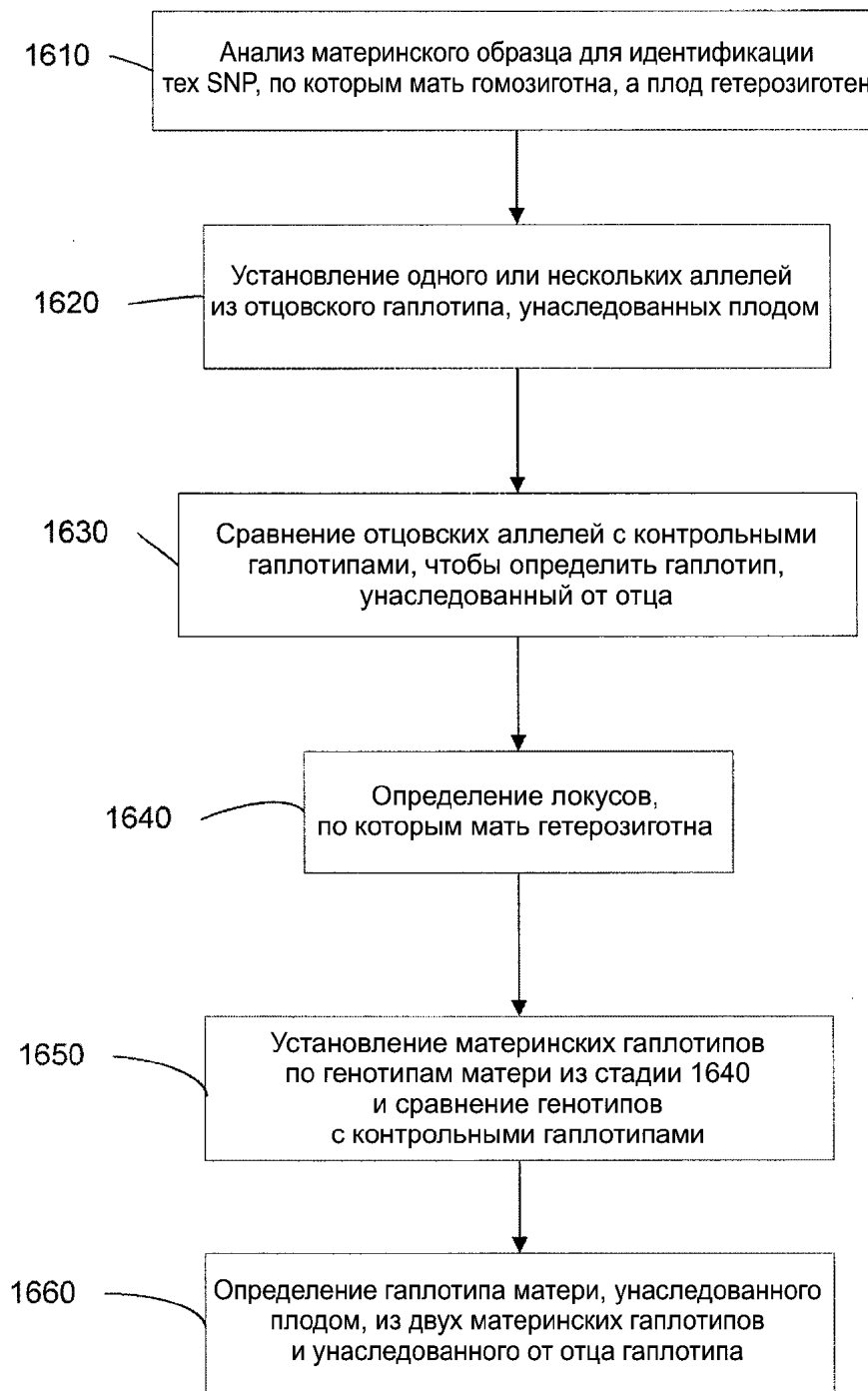
C
C
A
T
T
G
T
C
A
G
T
T
C
A

Нар А

A
C
G
A
T
A
C
G
A
T
T
A
A
G

Фиг. 16

1600



Гаплотипы плода		Гаплотипы матери	
Нар С	Нар А	Нар А	Нар В
C	A	A	C
C	C	C	G
A	G	G	G
T	A	A	T
T	T	T	A
G	A	A	G
T	C	C	C
C	G	G	C
A	A	A	T
G	T	T	T
T	T	T	A
T	A	A	T
C	A	A	C
A	G	G	G

1710

Фиг. 17

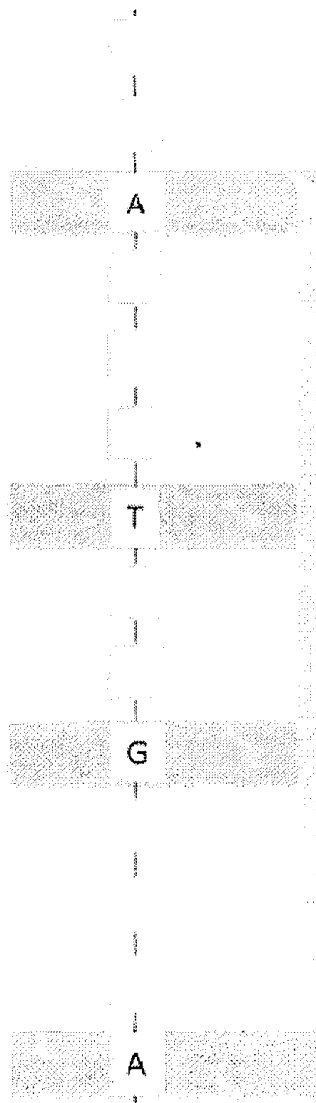
Анализ материнской плазмы

Те SNP, по которым мать гомозиготна, а плод унаследовал от отца другой аллель, чем аллель матери. Такие SNP можно идентифицировать при анализе материнской плазмы путем:

1. выявления специфичного для плода аллеля в материнской плазме; причем
 2. специфичный для плода аллель составляет лишь небольшую долю от всех аллелей, относящихся к этому локусу.
- Напротив, по тем локусам SNP, по которым мать гетерозиготна, встречаемость двух материнских аллелей в материнской плазме будет близкой.

Фиг. 18

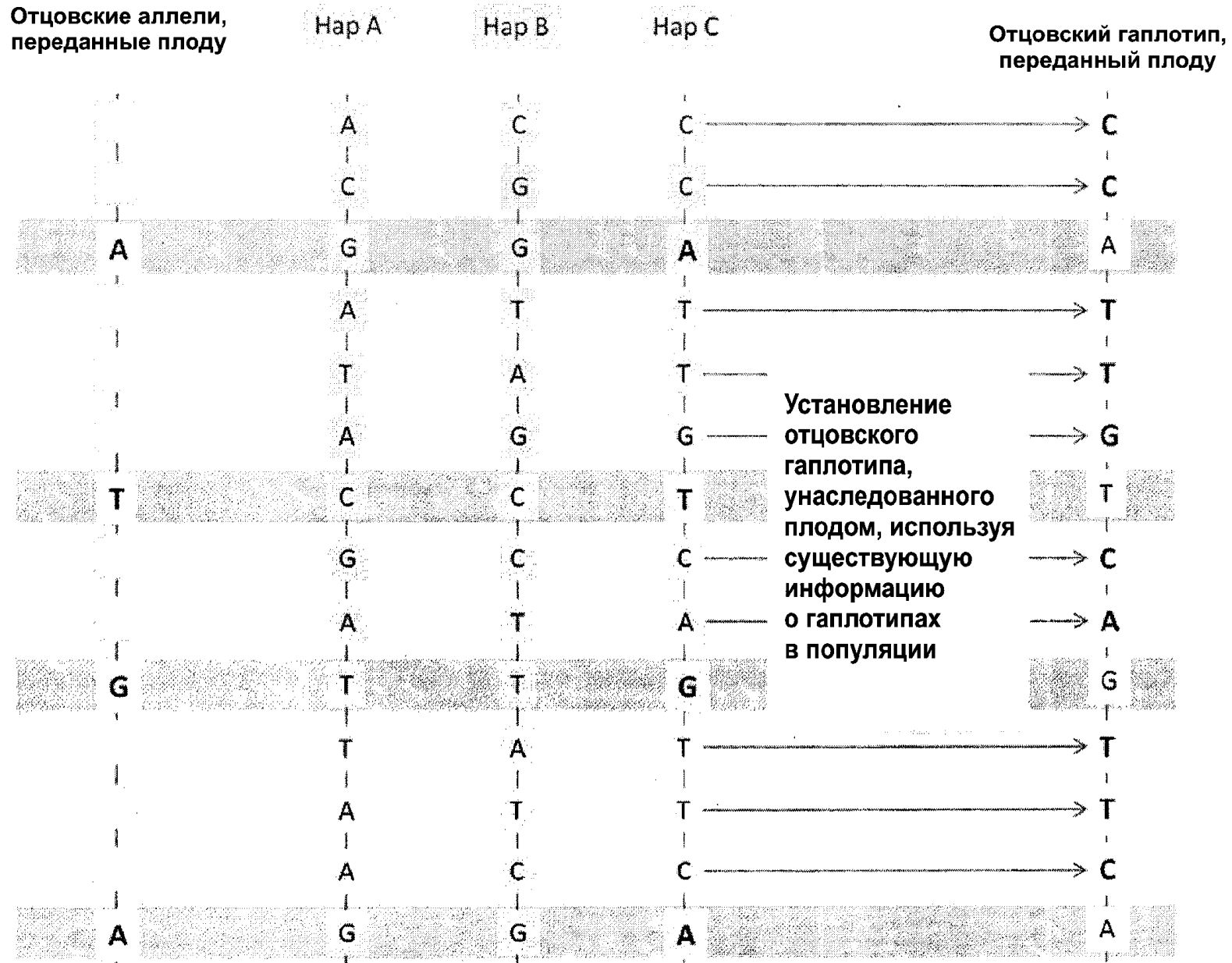
Аллели плода,
унаследованные от отца



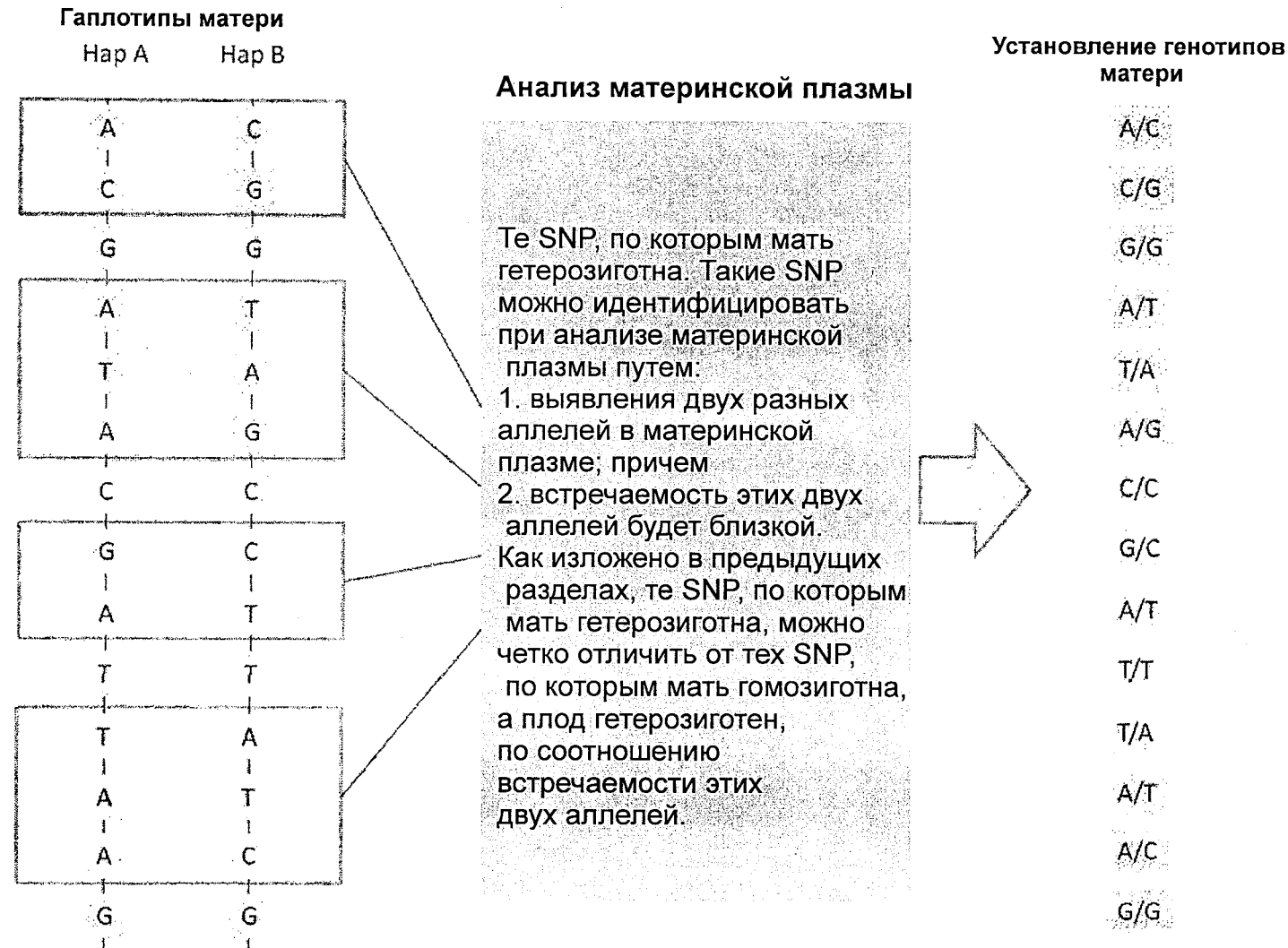
Сравнение
установленных
экспериментально
отцовских аллелей,
унаследованных
плодом (только для
одного подмножества
локусов SNP),
с известными
гаплотипами
в соответствующей
популяции

Нар А	Нар В	Нар С
A	C	C
C	G	C
G	G	A
A	T	T
T	A	T
A	G	G
C	C	T
G	C	C
A	T	A
T	T	G
T	A	T
A	T	T
A	C	C
G	G	A

Фиг. 19

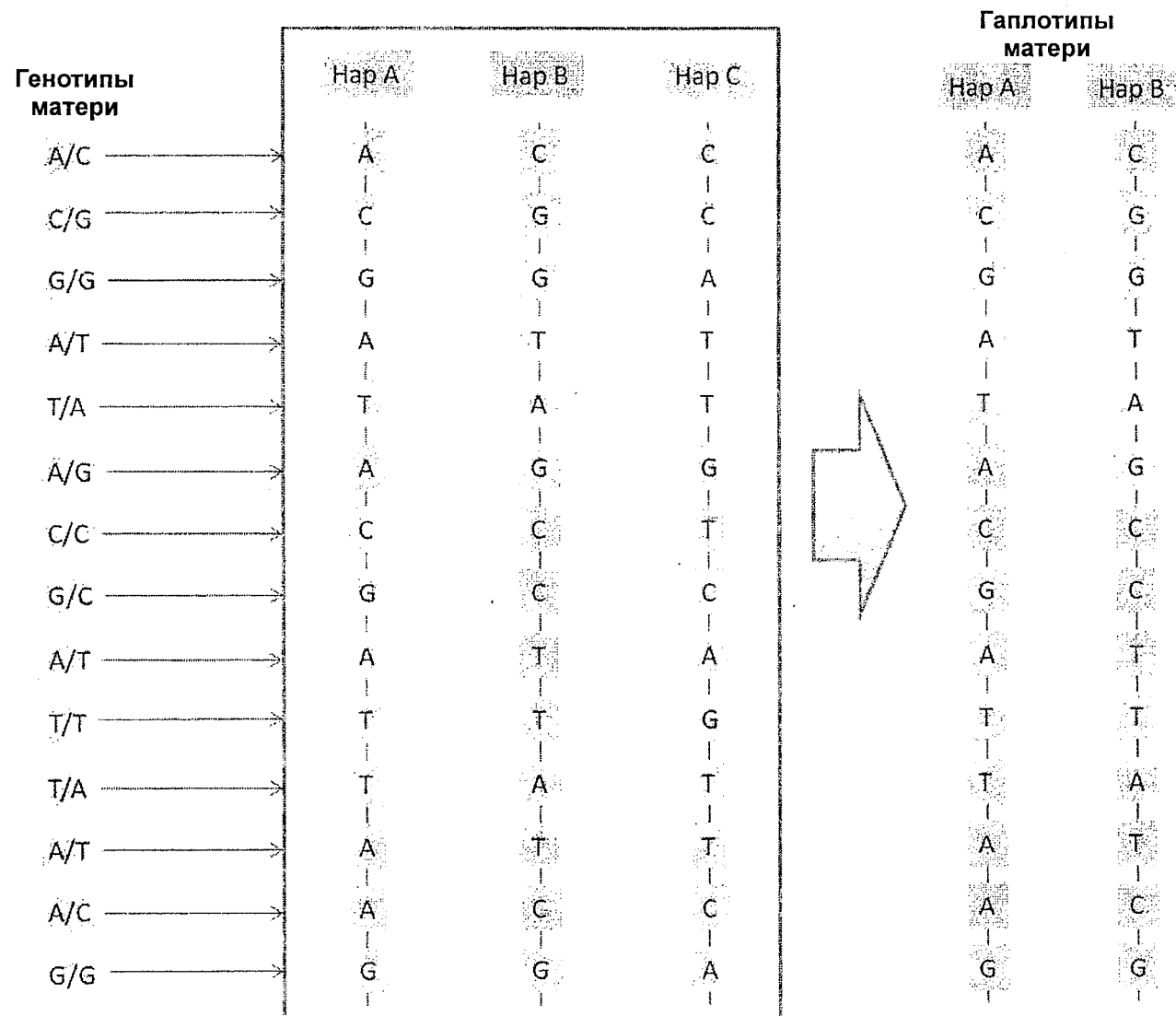


Фиг. 20



Фиг. 21

Определение материнских гаплотипов по генотипам матери



Гаплотипы матери		Гаплотип плода, унаследованный от отца
Нар А	Нар В	
A	C	C
C	G	C
G	G	A
A	T	T
T	A	T
A	G	G
C	C	T
G	C	C
A	T	A
T	T	G
T	A	T
A	T	T
A	C	C
G	G	A

Фиг. 22

Определение того, какой материнский гаплотип перешел к плоду, путем анализа относительной дозы гаплотипа (RHDO).

Фиг. 23

Гаплотипы матери		Гаплотип плода, унаследованный от отца
Нар А	Нар В	
A	C	C
C	G	C
G	G	A
A	T	T
T	A	T
A	G	G
C	C	T
G	C	C
A	T	A
T	T	G
T	A	T
A	T	T
A	C	C
T	C	T
G	G	A

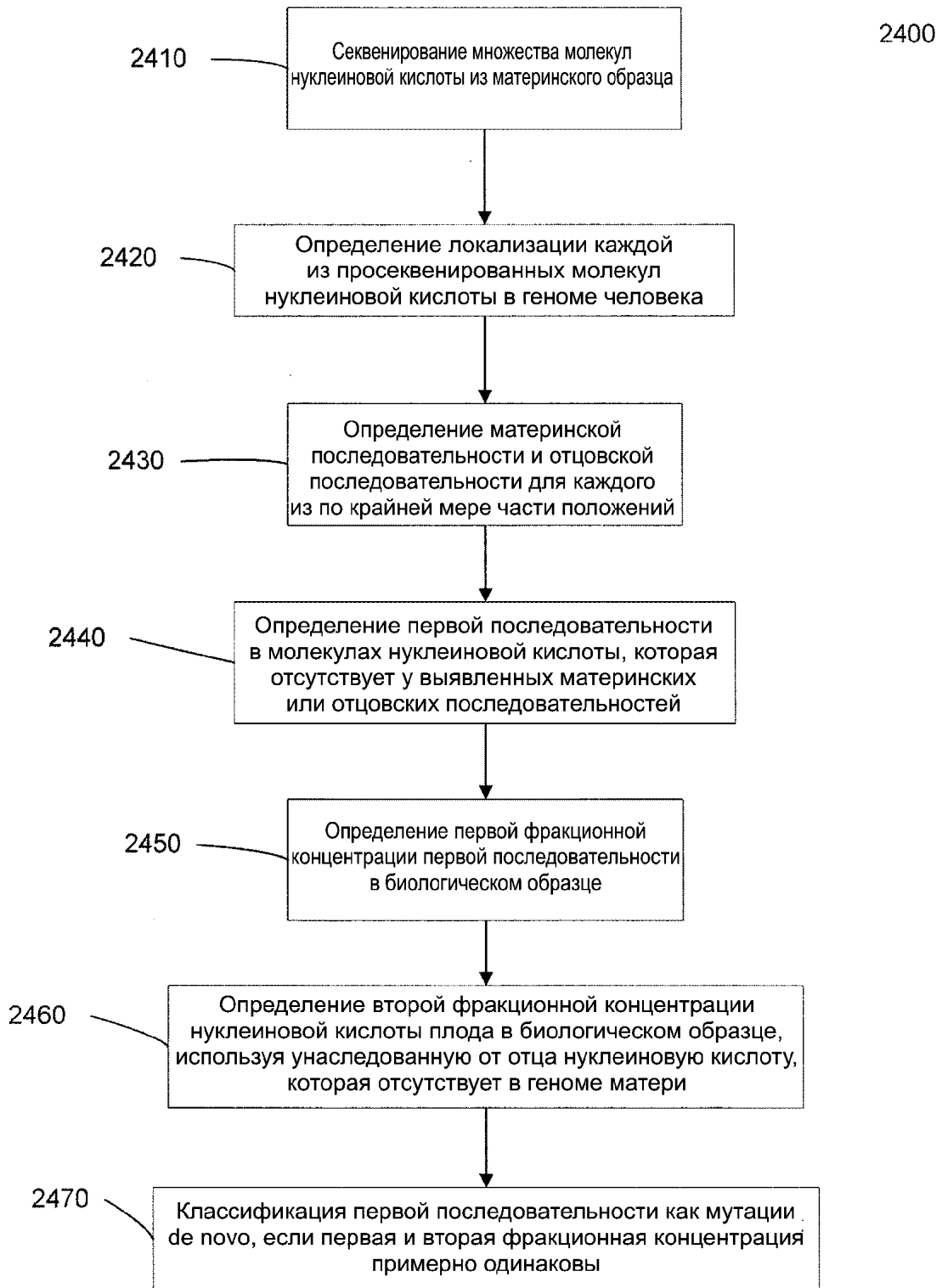
SNPs типа альфа

Те SNP, по которым отцовские аллели, унаследованные плодом, идентичны материнским аллелям из Нар А. По этим SNP, если плод унаследует Нар А от матери, то аллели из Нар А будут избыточно представлены в материнской плазме. Если же плод унаследует Нар В от матери, то два материнских аллеля будут одинаково представлены в материнской плазме.

SNPs типа бета

Те SNP, по которым отцовские аллели, унаследованные плодом, идентичны материнским аллелям из Нар В. По этим SNP, если плод унаследует Нар В от матери, то аллели из Нар В будут избыточно представлены в материнской плазме. Если же плод унаследует Нар А от матери, то два материнских аллеля будут одинаково представлены в материнской плазме.

Фиг. 24



Фиг. 25А

Комбинация генотипа	К-во SNPs	Содержание (%)
Плод гетерозиготен	220,296	24.57
Мать гомозиготна, отец гетерозиготен	131,037	14.61
Мать гомозиготна, плод гетерозиготен	111,775	12.46
Мать гомозиготна, отец гетерозиготен, плод гетерозиготен	65,875	7.35
Мать гомозиготна, отец гетерозиготен, плод гомозиготен	65,162	7.27
Мать гомозиготна, отец гомозиготен, но по другим аллелям	45,900	5.12
Итого	896,756	100

Фиг. 25В

Проточная кювета	Общее к-во отрезков (млн)	К-во выровненных отрезков (млн)	Процент выровненных отрезков (%)
1	265	200	75.29
2	274	206	75.22
3	268	204	75.96
4	254	193	76.10
5	258	196	76.02
6	280	215	76.60
7	272	207	75.97
8	253	192	75.80
9	260	195	75.17
10	266	208	78.11
11	245	188	76.68
12	249	190	76.22
13	250	189	75.32
14	257	192	74.79
15	250	187	75.14
16	243	187	76.93
17	270	200	74.25
18	258	197	76.48
19	258	193	74.98
20	253	193	76.31

Фиг. 26

Хр.	Отец гетеро, мать гомо, плод гетеро (%)	Отец гомо, мать гомо, но по другим аллелям (%)
chr1	11.83	11.92
chr2	12.15	11.99
chr3	11.80	11.92
chr4	11.89	11.84
chr5	12.01	12.10
chr6	12.00	11.80
chr7	12.14	11.75
chr8	12.04	11.95
chr9	11.70	11.82
chr10	11.57	11.75
chr11	11.88	11.90
chr12	12.09	11.80
chr13	12.39	11.94
chr14	11.90	11.84
chr15	11.75	11.61
chr16	11.93	11.36
chr17	11.77	11.49
chr18	11.99	11.74
chr19	11.77	12.02
chr20	11.74	11.56
chr21	11.74	11.37
chr22	11.81	11.34
chrX		12.03

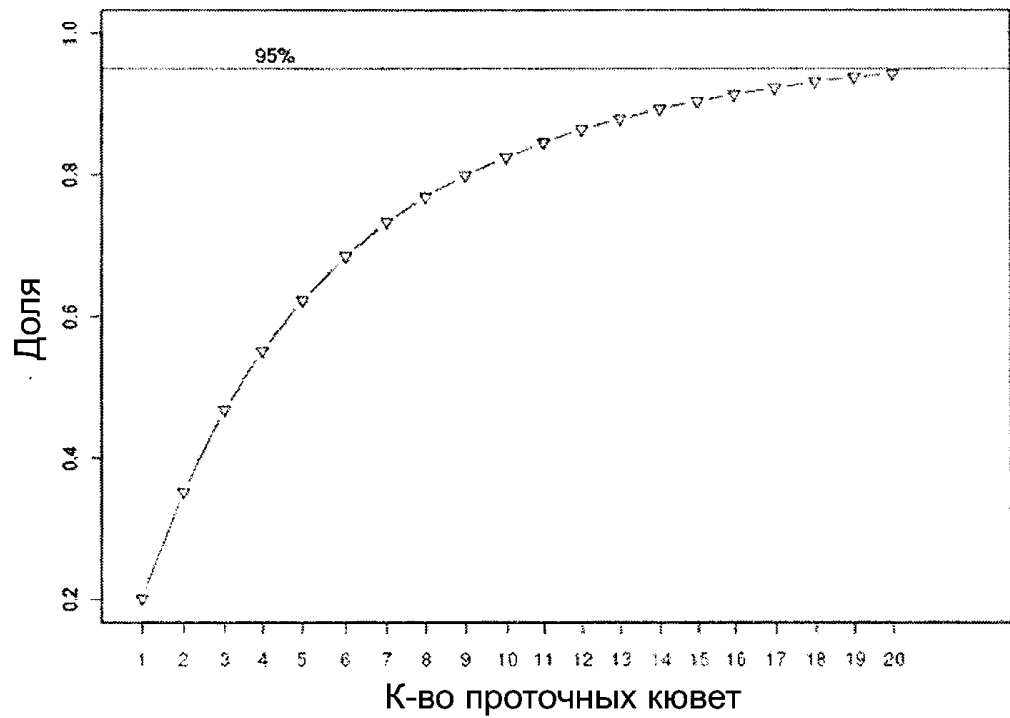
Хр.: хромосома

гомо: гомозигота

гетеро: гетерозигота

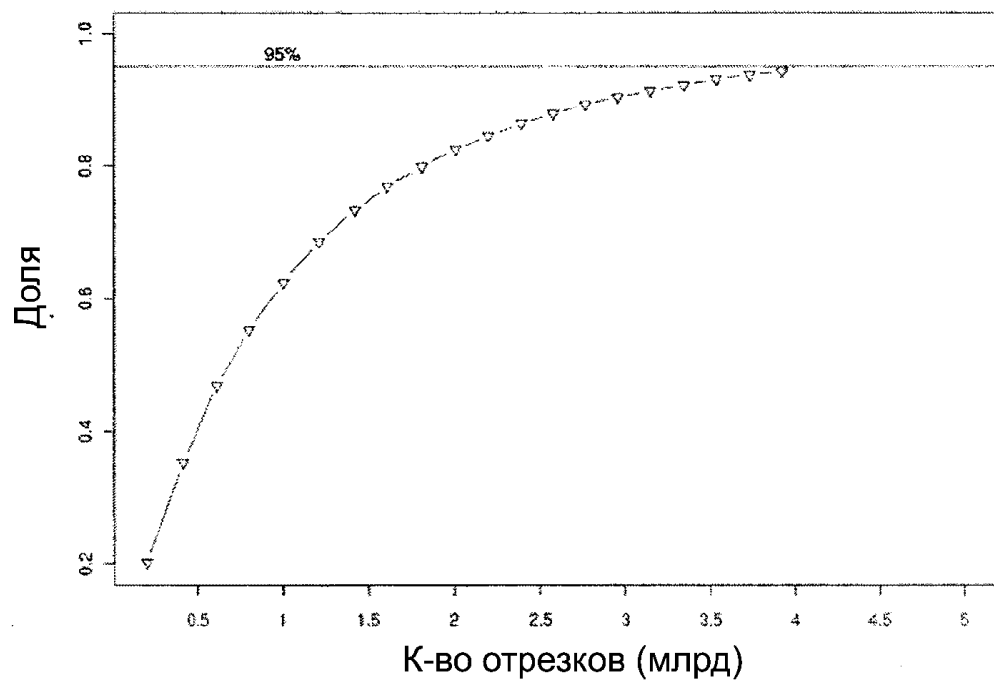
Фиг. 27А

Доля SNPs, охваченных специфичными для плода отрезками
(мать и отец гомозиготны, но по разным аллелям)

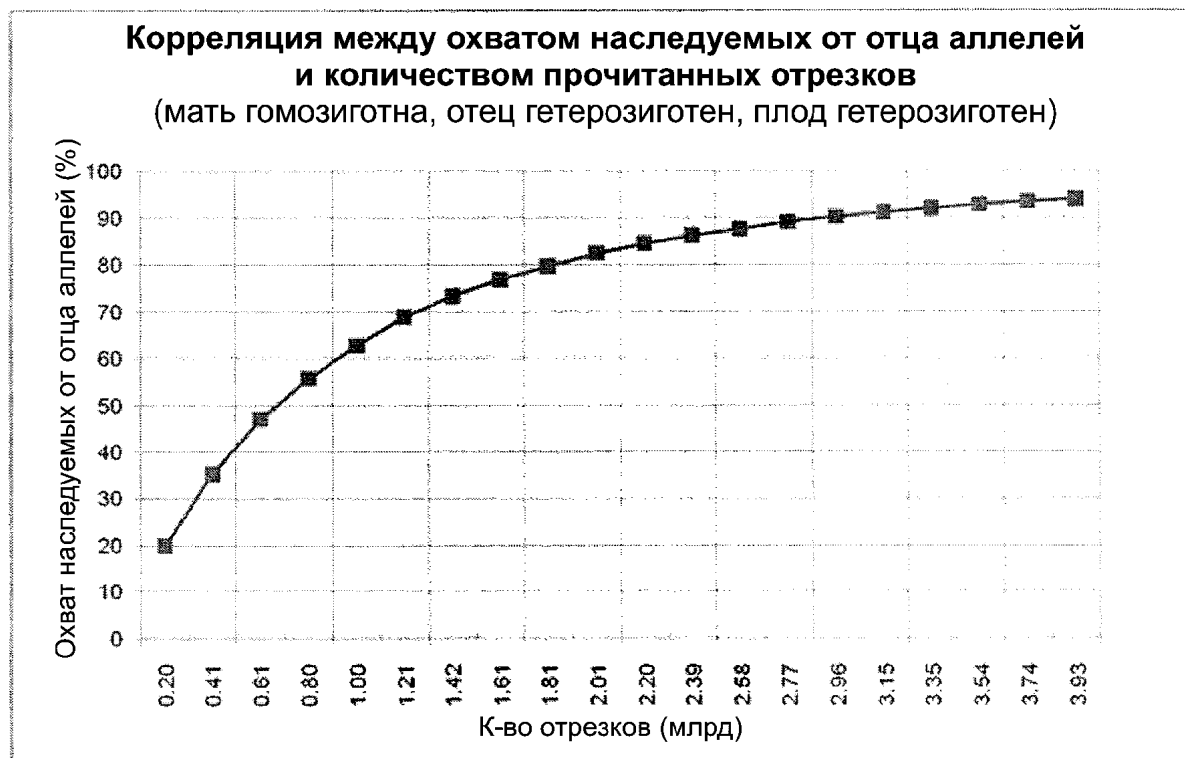


Фиг. 27В

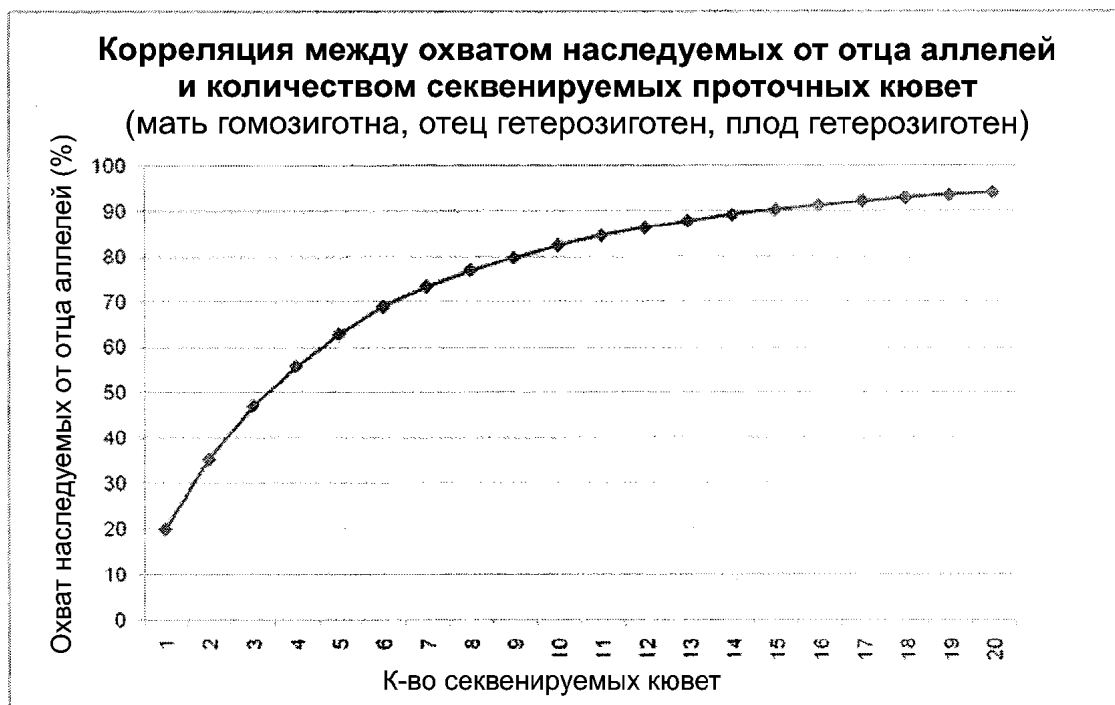
Доля SNPs, охваченных специфичными для плода отрезками
(мать и отец гомозиготны, но по разным аллелям)



Фиг. 28А

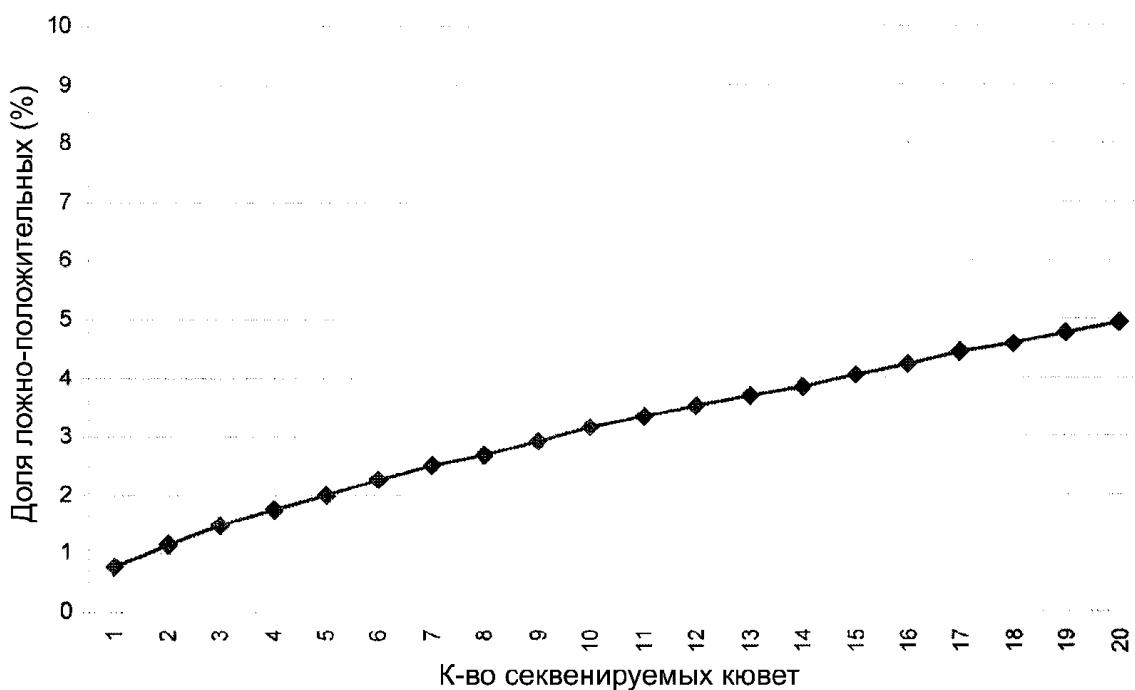


Фиг. 28В



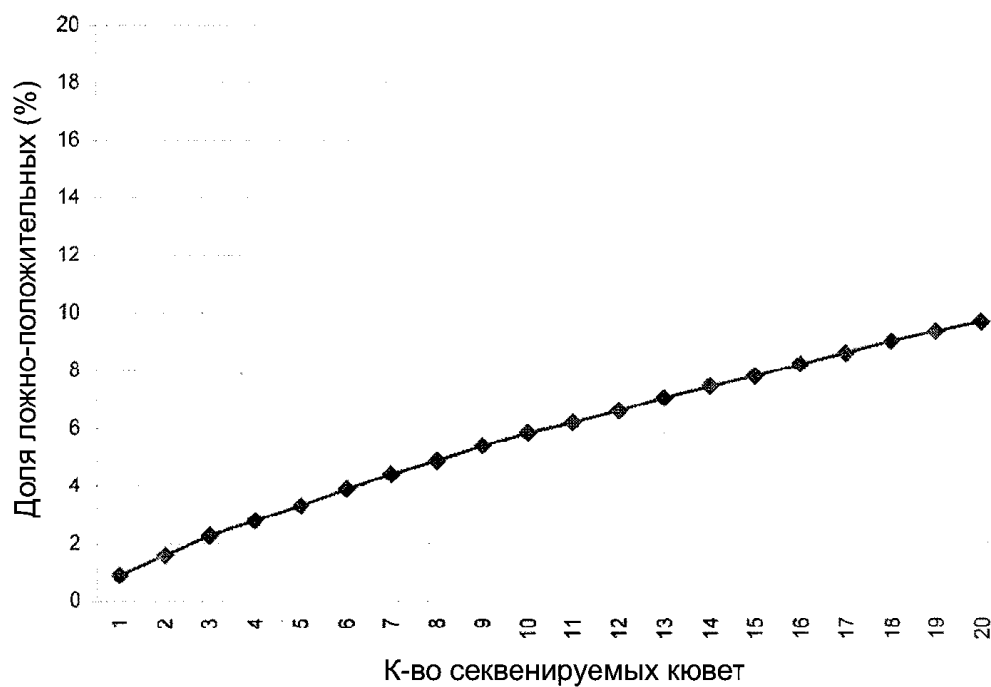
Фиг. 29А

Корреляция между долей ложно-положительных и количеством секвенируемых проточных кювет



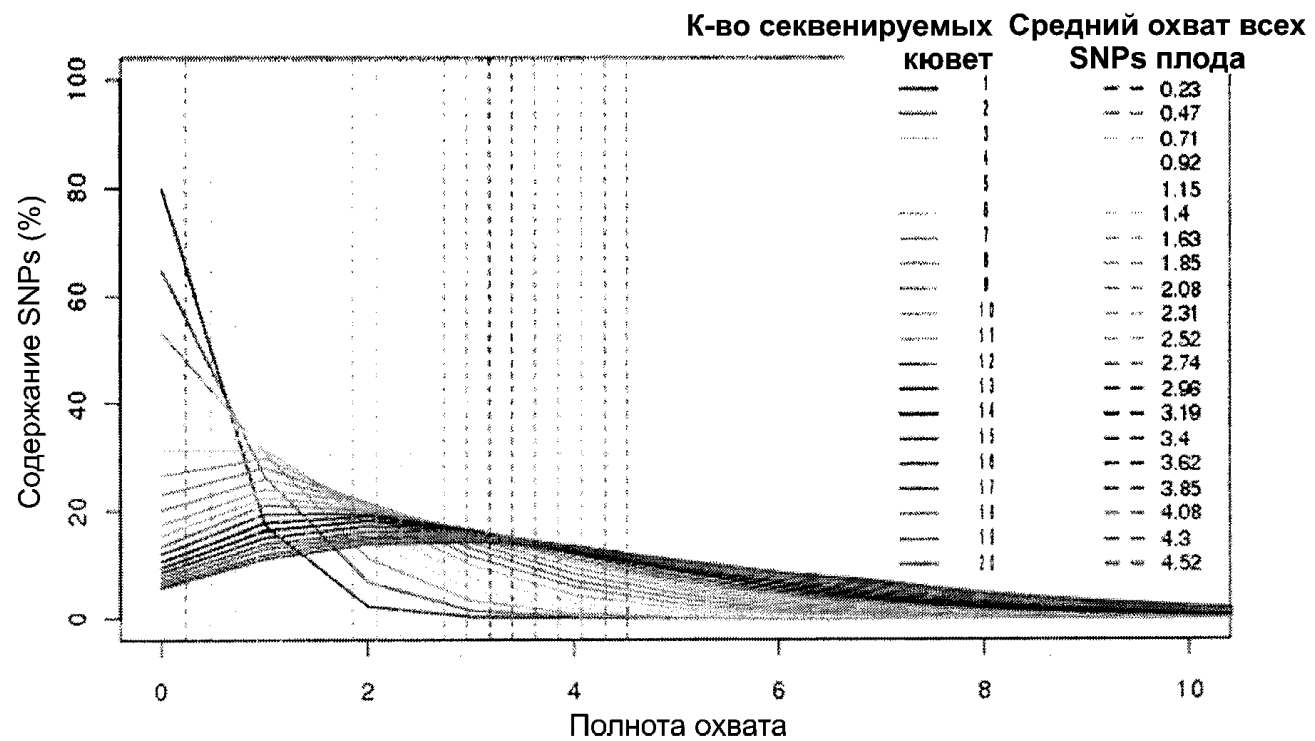
Фиг. 29В

Корреляция между долей ложно-положительных и количеством секвенируемых проточных кювет
(мать гомозиготна, отец гомозиготен по тем же аллелям)



Фиг. 30

Полнота охвата SNPs при различном количестве секвенируемых кювет



Фиг. 31

Точность классификации типа А (10 проточных кювет)

Хр.	Правильная классификация		Неправильная классификация		Неправильная при “перемежающемся” анализе		Всего
	К-во	Процент	К-во	Процент	К-во	Процент	
Chr 1	224	99.1%	2	0.9%	0	0.0%	226
Chr 2	215	96.8%	7	3.2%	0	0.0%	222
Chr 3	176	97.2%	5	2.8%	0	0.0%	181
Chr 4	162	98.8%	2	1.2%	0	0.0%	164
Chr 5	139	99.3%	1	0.7%	0	0.0%	140
Chr 6	170	99.4%	1	0.6%	0	0.0%	171
Chr 7	134	99.3%	1	0.7%	0	0.0%	135
Chr 8	154	98.7%	2	1.3%	0	0.0%	156
Chr 9	134	99.3%	1	0.7%	0	0.0%	135
Chr 10	154	99.4%	1	0.6%	0	0.0%	155
Chr 11	151	98.7%	2	1.3%	0	0.0%	153
Chr 12	136	99.3%	1	0.7%	0	0.0%	137
Chr 13	95	100.0%	0	0.0%	0	0.0%	95
Chr 14	91	100.0%	0	0.0%	0	0.0%	91
Chr 15	76	100.0%	0	0.0%	0	0.0%	76
Chr 16	95	99.0%	1	1.0%	0	0.0%	96
Chr 17	82	100.0%	0	0.0%	0	0.0%	82
Chr 18	67	100.0%	0	0.0%	0	0.0%	67
Chr 19	46	100.0%	0	0.0%	0	0.0%	46
Chr 20	79	97.5%	2	2.5%	0	0.0%	81
Chr 21	45	100.0%	0	0.0%	0	0.0%	45
Chr 22	32	100.0%	0	0.0%	0	0.0%	32
Chr X	103	96.3%	4	3.7%	0	0.0%	107
Итого	2760	98.8%	33	1.2%	0	0.0%	2793

Фиг. 32

Точность классификации типа В (10 проточных кювет)

Хр.	Правильная классификация		Неправильная классификация		Неправильная при “перемежающемся” анализе		Всего
	К-во	Процент	К-во	Процент	К-во	Процент	
Chr 1	210	97.2%	6	2.8%	0	0.0%	216
Chr 2	230	97.9%	5	2.1%	0	0.0%	235
Chr 3	163	99.4%	1	0.6%	0	0.0%	164
Chr 4	127	100.0%	0	0.0%	0	0.0%	127
Chr 5	182	98.9%	2	1.1%	0	0.0%	184
Chr 6	155	96.9%	5	3.1%	0	0.0%	160
Chr 7	115	97.5%	3	2.5%	0	0.0%	118
Chr 8	160	98.8%	2	1.2%	0	0.0%	162
Chr 9	123	98.4%	2	1.6%	0	0.0%	125
Chr 10	130	97.0%	4	3.0%	0	0.0%	134
Chr 11	129	97.7%	3	2.3%	0	0.0%	132
Chr 12	118	100.0%	0	0.0%	0	0.0%	118
Chr 13	90	97.8%	2	2.2%	0	0.0%	92
Chr 14	83	97.6%	2	2.4%	0	0.0%	85
Chr 15	75	98.7%	1	1.3%	0	0.0%	76
Chr 16	80	94.1%	5	5.9%	0	0.0%	85
Chr 17	50	100.0%	0	0.0%	0	0.0%	50
Chr 18	71	97.3%	2	2.7%	0	0.0%	73
Chr 19	48	100.0%	0	0.0%	0	0.0%	48
Chr 20	80	100.0%	0	0.0%	0	0.0%	80
Chr 21	27	100.0%	0	0.0%	0	0.0%	27
Chr 22	44	97.8%	1	2.2%	0	0.0%	45
Chr X	118	97.5%	3	2.5%	0	0.0%	121
Итого	2608	98.2%	49	1.8%	0	0.0%	2657

Фиг. 33

Точность классификации типа В (10 проточных кювет)

Хр.	Классификация по единому блоку			Алгоритм “двух последовательных блоков”		
	Правильная (%)	Неправильная (%)	Всего	Правильная (%)	Неправильная (%)	Всего
1	314 (99.1%)	3 (0.9%)	317	309 (100%)	0 (0.0%)	309
2	305 (99.3%)	2 (0.7%)	307	301 (100%)	0 (0.0%)	301
3	240 (99.6%)	1 (0.4%)	241	237 (100%)	0 (0.0%)	237
4	218 (99.5%)	1 (0.5%)	219	215 (100%)	0 (0.0%)	215
5	203 (100%)	0 (0.0%)	203	201 (100%)	0 (0.0%)	201
6	240 (98.0%)	5 (2.0%)	245	237 (98.8%)	3 (1.3%)	240
7	182 (98.9%)	2 (1.1%)	184	179 (100%)	0 (0.0%)	179
8	217 (99.5%)	1 (0.5%)	218	214 (100%)	0 (0.0%)	214
9	177 (99.4%)	1 (0.6%)	178	174 (100%)	0 (0.0%)	174
10	217 (99.1%)	2 (0.9%)	219	213 (100%)	0 (0.0%)	213
11	205 (100%)	0 (0.0%)	205	203 (100%)	0 (0.0%)	203
12	193 (99.5%)	1 (0.5%)	194	190 (100%)	0 (0.0%)	190
13	133 (100%)	0 (0.0%)	133	132 (100%)	0 (0.0%)	132
14	125 (100%)	0 (0.0%)	125	124 (100%)	0 (0.0%)	124
15	103 (100%)	0 (0.0%)	103	102 (100%)	0 (0.0%)	102
16	133 (98.5%)	2 (1.5%)	135	129 (100%)	0 (0.0%)	129
17	104 (100%)	0 (0.0%)	104	102 (100%)	0 (0.0%)	102
18	95 (97.9%)	2 (2.1%)	97	91 (100%)	0 (0.0%)	91
19	67 (100%)	0 (0.0%)	67	65 (100%)	0 (0.0%)	65
20	112 (100%)	0 (0.0%)	112	110 (100%)	0 (0.0%)	110
21	60 (100%)	0 (0.0%)	60	59 (100%)	0 (0.0%)	59
22	39 (100%)	0 (0.0%)	39	38 (100%)	0 (0.0%)	38
X	156 (98.7%)	2 (1.3%)	158	152 (100%)	0 (0.0%)	152
Итого	3838 (99.4%)	25 (0.6%)	3863	3777 (99.9%)	3 (0.1%)	3780

Фиг. 34

Точность анализа по SPRT типа В

Хр.	Классификация по единому блоку			Алгоритм "двух последовательных блоков"		
	Правильная (%)	Неправильная (%)	Всего	Правильная (%)	Неправильная (%)	Всего
1	267 (98.2%)	5 (1.8%)	272	262 (100%)	0 (0.0%)	262
2	283 (99.0%)	3 (1.0%)	286	278 (100%)	0 (0.0%)	278
3	222 (98.2%)	4 (1.8%)	226	216 (100%)	0 (0.0%)	216
4	175 (98.9%)	2 (1.1%)	177	171 (100%)	0 (0.0%)	171
5	255 (99.6%)	1 (0.4%)	256	252 (100%)	0 (0.0%)	252
6	195 (99.0%)	2 (1.0%)	197	192 (100%)	0 (0.0%)	192
7	155 (97.5%)	4 (2.5%)	159	150 (100%)	0 (0.0%)	150
8	211 (100%)	0 (0.0%)	211	209 (100%)	0 (0.0%)	209
9	162 (99.4%)	1 (0.6%)	163	160 (100%)	0 (0.0%)	160
10	172 (96.6%)	6 (3.4%)	178	167 (98.2%)	3 (1.8%)	170
11	171 (99.4%)	1 (0.6%)	172	168 (100%)	0 (0.0%)	168
12	156 (99.4%)	1 (0.6%)	157	153 (100%)	0 (0.0%)	153
13	119 (98.3%)	2 (1.7%)	121	117 (99.2%)	1 (0.8%)	118
14	116 (98.3%)	2 (1.7%)	118	115 (99.1%)	1 (0.9%)	116
15	101 (100%)	0 (0.0%)	101	100 (100%)	0 (0.0%)	100
16	115 (94.3%)	7 (5.7%)	122	107 (99.1%)	1 (0.9%)	108
17	67 (98.5%)	1 (1.5%)	68	65 (100%)	0 (0.0%)	65
18	92 (100%)	0 (0.0%)	92	90 (100%)	0 (0.0%)	90
19	57 (100%)	0 (0.0%)	57	55 (100%)	0 (0.0%)	55
20	95 (100%)	0 (0.0%)	95	93 (100%)	0 (0.0%)	93
21	39 (100%)	0 (0.0%)	39	38 (100%)	0 (0.0%)	38
22	59 (100%)	0 (0.0%)	59	58 (100%)	0 (0.0%)	58
X	142 (99.3%)	1 (0.7%)	143	139 (100%)	0 (0.0%)	139
Итого	3426 (98.8%)	43 (1.2%)	3469	3355 (99.8%)	6 (0.2%)	3361

Таким образом, % мутантных = $10/72 = 0,1389$

Фиг. 36

Анализ типа А по кластеру β-глобина на хромосоме 11

Хр.	db SNPId	Локус	Генотип					Всего НаpIII	Всего НаpIV	Накопленная встречаемость НаpIII	Накопленная встречаемость НаpIII	Классификация
			Отец	Мать	Плод	Наp III	Наp IV					
chr 11		5204905	Мутация: талассемия, бета-глобин, нт. -28									
chr11	rs10742583	5205217	TT	CT	TT	T	C	22	26	22	26	не классиф.
chr11	rs10742584	5205346	TT	CT	TT	T	C	21	12	43	38	не классиф.
chr11	rs1003586	5205946	GG	AG	GG	G	A	24	11	67	49	не классиф.
chr11	rs6578588	5208827	GG	AG	GG	G	A	43	38	110	87	не классиф.
chr11	rs3759076	5213765	CC	CT	CC	C	T	29	0	139	87	не классиф.
chr11	rs10837643	5214614	AA	AT	AA	A	T	23	11	162	98	не классиф.
chr11	rs4283007	5215066	TT	AT	TT	T	A	53	36	215	134	Наp III

Фиг. 37

Анализ типа В по кластеру β-глобина на хромосоме 11

Хр.	dbSNPId	Локус	Отец	Мать	Плод	Нар III	Нар IV	Всего НарIII	Всего НарIV	Накопленная	Накопленная	Классификация
										встречаемость НарIII	встречаемость НарIII	
chr 11		5204905	Мутация: талассемия, бета-глобин, нт. -28									
chr11	rs12275403	5286913	CC	CG	CG	G	C	19	23	19	23	не классиф.
chr11	rs7926089	5289019	CC	CT	CT	T	C	13	21	32	44	не классиф.
chr11	rs12365420	5289330	GG	GT	GT	T	G	5	12	37	56	не классиф.
chr11	rs4910747	5294746	CC	AC	AC	A	C	27	36	64	92	не классиф.
chr11	rs6578604	5299185	AA	AG	AG	G	A	15	18	79	110	не классиф.
chr11	rs10837814	5301704	GG	AG	AG	A	G	39	45	118	155	не классиф.
chr11	rs4910550	5302436	GG	AG	AG	A	G	15	22	133	177	не классиф.
chr11	rs2723381	5304451	TT	CT	CT	C	T	38	34	171	211	не классиф.
chr11	rs6578606	5306671	GG	CG	CG	C	G	40	31	211	242	не классиф.
chr11	rs3824950	5588548	CC	CT	CT	T	C	59	66	270	308	не классиф.
chr11	rs10838441	5601579	GG	CG	CG	C	G	20	10	290	318	не классиф.
chr11	rs11038628	5645516	CC	CT	CT	T	C	65	5	355	323	Нар III

Фиг. 38А

Геномные координаты SNP	Материнские аллели из:		Всего аллелей из:		Накопленная встречаемость аллелей из:		Классификация по SPRT
the SNP	Нар I	Нар II	Нар I	Нар II	Нар I	Нар II	
3251050	A	G	22	15	22	15	не классиф.
3251284	G	A	14	10	36	25	не классиф.
3251791	A	G	18	11	54	36	Нар I
9865342	A	G	5	3	5	3	не классиф.
9872792	A	G	7	5	12	8	не классиф.
9874947	G	C	1	3	13	11	не классиф.
15242479	G	A	17	5	30	16	Нар I
15710521	G	A	1	2	1	2	не классиф.
15710672	A	G	13	13	14	15	не классиф.
21895385	C	T	15	3	29	18	не классиф.
22149641	T	C	5	3	34	21	Нар I
22201527	A	C	11	10	11	10	не классиф.
23806924	C	T	20	9	31	19	не классиф.
31134605	A	G	2	1	33	20	Нар I
32413115	C	T	6	8	6	8	не классиф.
32473184	G	A	8	9	14	17	не классиф.
35730716	A	G	17	14	31	31	не классиф.
35903375	C	T	7	9	38	40	не классиф.
35903407	G	A	6	2	44	42	не классиф.
35917535	T	G	6	7	50	49	не классиф.
43475980	T	G	7	16	57	65	не классиф.
47351305	A	G	7	6	64	71	не классиф.
48561783	A	G	5	1	69	72	не классиф.
50676020	C	T	19	15	88	87	не классиф.
53601143	A	G	8	10	96	97	не классиф.
54854802	T	C	9	3	105	100	не классиф.
74921254	C	A	3	3	108	103	не классиф.
78102680	A	C	5	3	113	106	не классиф.
78313644	C	T	40	24	153	130	Нар I

39/45
Фиг. 38В

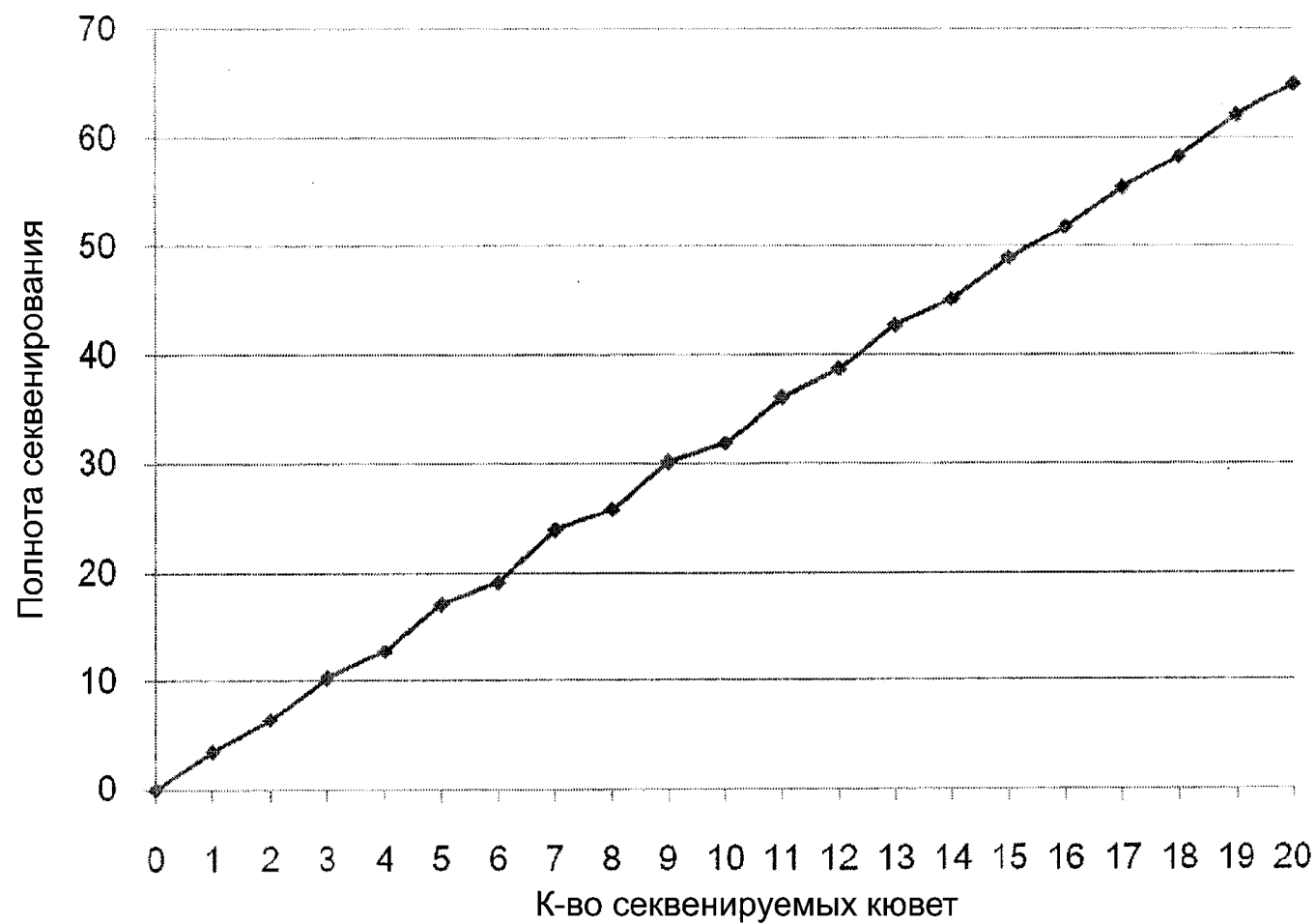
Геномные координаты SNP	Материнские аллели из:		Всего аллелей из:		Накопленная встречаемость аллелей из:		Классификация по SPRT
	Нар I	Нар II	Нар I	Нар II	Нар I	Нар II	
84236666	A	G	3	0	3	0	не классиф.
84249796	A	G	29	12	32	12	Нар I
100273903	T	C	16	10	16	10	не классиф.
100494847	A	G	13	5	29	15	Нар I
106283058	G	A	5	8	5	8	не классиф.
106283111	G	A	13	12	18	20	не классиф.
107283614	G	A	10	4	28	24	не классиф.
107335295	T	G	5	1	33	25	не классиф.
114785655	C	T	11	7	44	32	не классиф.
117584169	G	A	8	2	52	34	Нар I
118471031	T	C	3	5	3	5	не классиф.
118488464	C	T	5	15	8	20	не классиф.
119474561	A	T	5	4	13	24	не классиф.
123022716	C	A	17	12	30	36	не классиф.
128506711	C	G	4	3	34	39	не классиф.
130047404	C	T	4	1	38	40	не классиф.
130243499	C	T	13	19	51	59	не классиф.
134814366	C	T	12	8	63	67	не классиф.
134818744	G	A	27	18	90	85	не классиф.
134821671	C	T	5	9	95	94	не классиф.
138507378	A	T	7	3	102	97	не классиф.
138695073	A	G	10	3	112	100	не классиф.
144713436	C	T	10	16	122	116	не классиф.
149577161	A	G	11	5	133	121	не классиф.
150888835	C	A	11	11	144	132	не классиф.
152381349	T	C	8	3	152	135	Нар I
152694738	C	T	1	0	1	0	не классиф.
152848587	G	A	4	6	5	6	не классиф.

Фиг. 39

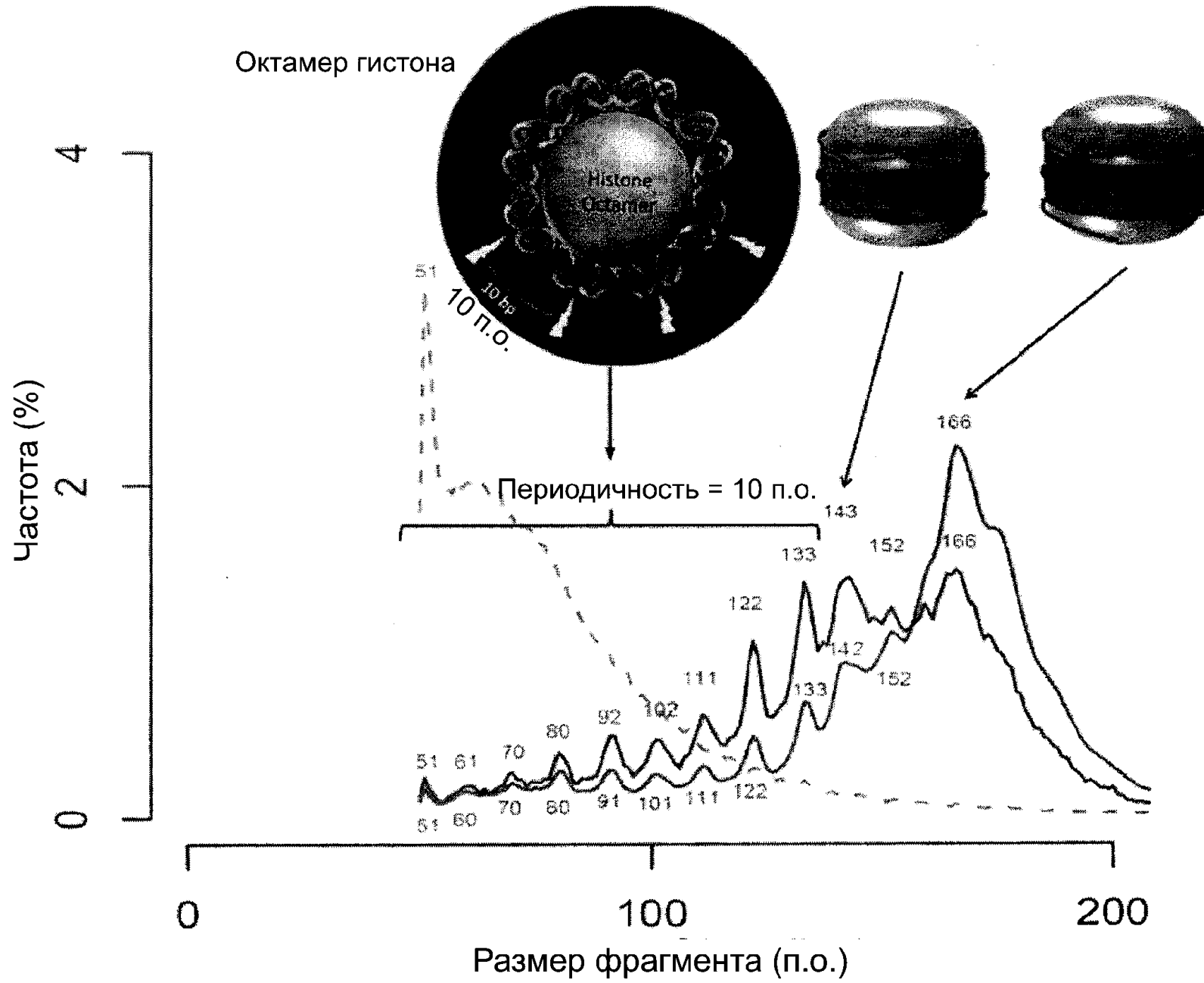
Случай	Всего информативных SNP	Средняя полнота секвенирования для каждого SNP (разы)	Фракц. концентрация ДНК плода (%)	Всего классификаций по SPRT	К-во правильных классификаций	К-во неправильных классификаций
PW226	57	17.6	19.4	9	9	0
PW263	62	15.8	18.4	4	4	0
PW316	50	18.5	17.4	2	2	0
PW370	54	18.9	11.5	1	1	0
PW421	46	19.6	15.3	2	2	0

Фиг. 40

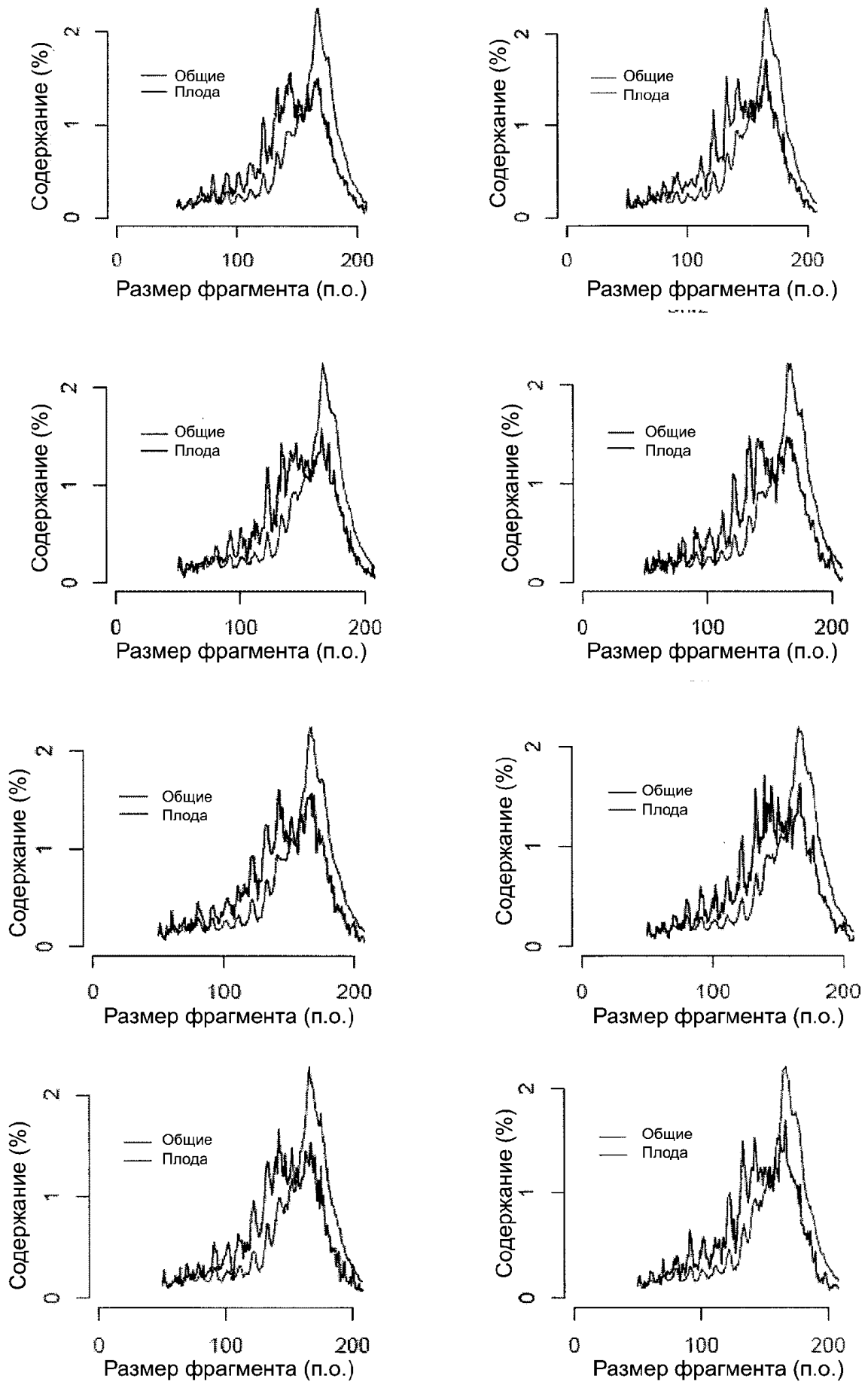
Полнота секвенирования от к-ва секвенируемых проточных кювет



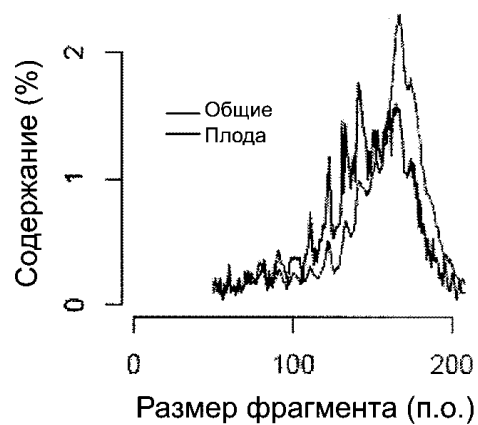
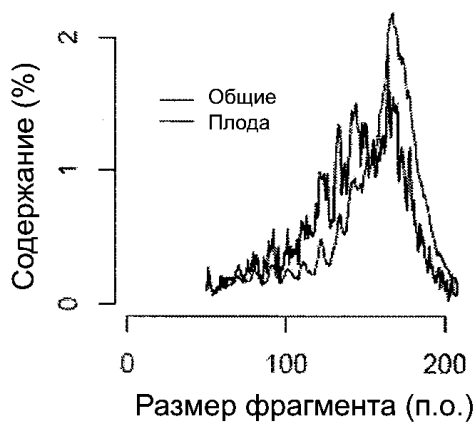
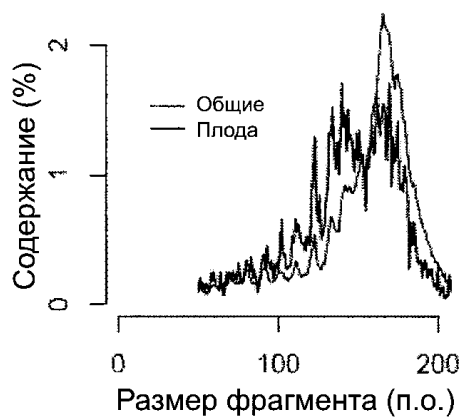
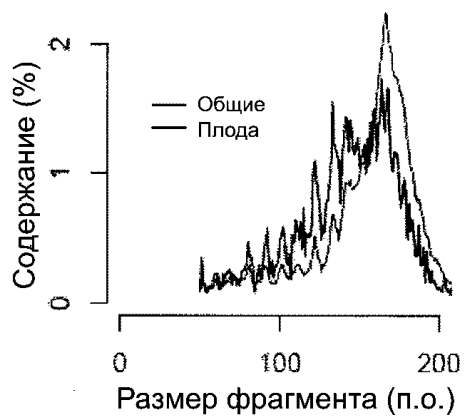
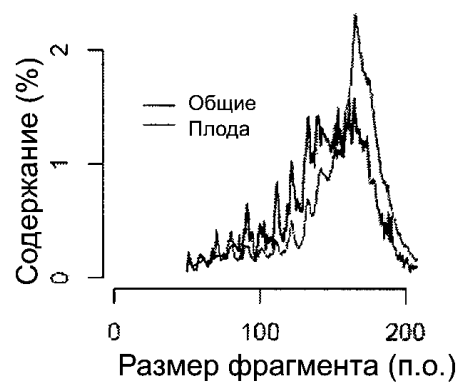
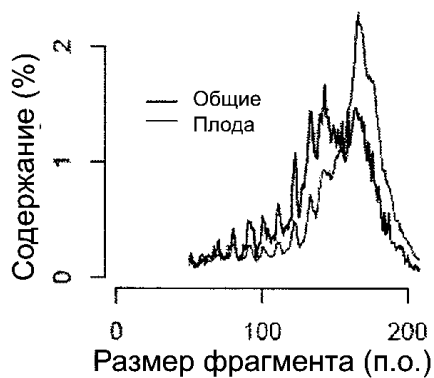
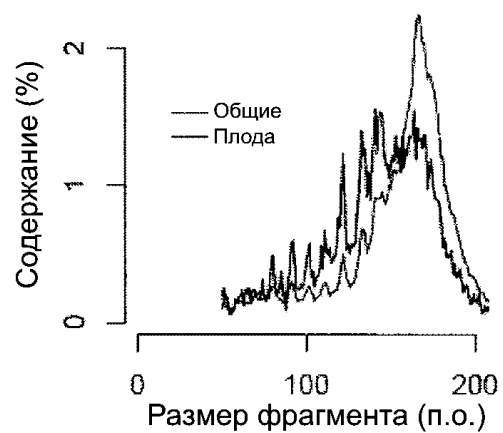
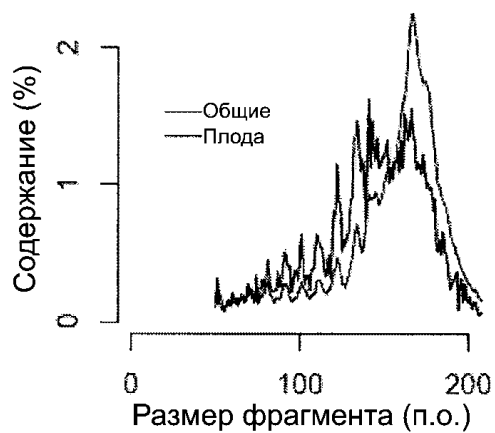
Фиг. 41



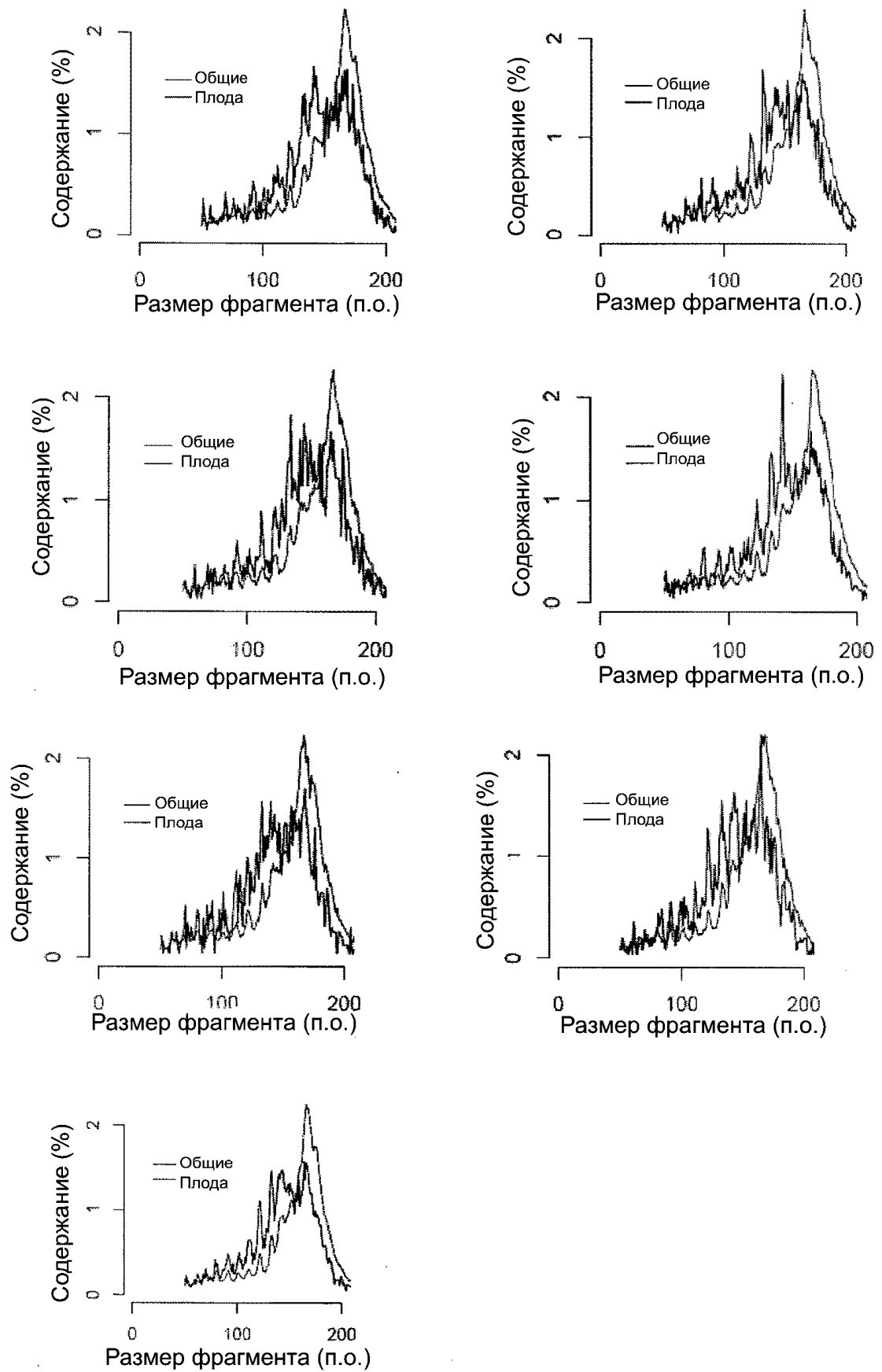
Фиг. 42А



Фиг. 42В



Фиг. 42С



Фиг. 43



INTERNATIONAL SEARCH REPORT

20199/550

International application No

PCT/US2010/055655

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12Q1/68

ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12Q G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2007/100911 A2 (UNIV LOUISVILLE RES FOUND [US]; MITCHELL AOY TOMITA [US]; MITCHELL MIC) 7 September 2007 (2007-09-07) page 1, line 30 - page 2, line 2; figures 1-3; example 4 page 10, line 21 - page 11, line 5 ----- -/--	1-28

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

4 February 2011

Date of mailing of the international search report

20/04/2011

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Nurmi, Jussi

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2010/055655

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DING CHUNMING ET AL: "MS analysis of single-nucleotide differences in circulating nucleic acids: Application to noninvasive prenatal diagnosis", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES (PNAS), NATIONAL ACADEMY OF SCIENCE, US, vol. 101, no. 29, 20 July 2004 (2004-07-20), pages 10762-10767, XP002614099, ISSN: 0027-8424, DOI: DOI:10.1073/PNAS.0403962101 [retrieved on 2004-07-09] abstract; table 3 page 10764, column 2, paragraph 3 page 10765, column 1, paragraph 3 -----	1-28
X	WO 2009/013492 A1 (UNIV HONG KONG CHINESE [CN]; LO YUK-MING DENNIS [CN]; CHIU ROSSA WAI K) 29 January 2009 (2009-01-29) paragraphs [0003], [0012], [0099], [0157], [0228] - [0246] -----	1-28
X	WO 2004/078999 A1 (GENETYPE PTY LTD [AU]; BOHMER RALPH MICHAEL [AU]) 16 September 2004 (2004-09-16) page 2 -----	1-28
X	REED W ET AL: "Non-invasive determination of the paternal HLA haplotype of a fetus using kinetic PCR to detect fetal microchimerism in maternal plasma", BONE MARROW TRANSPLANTATION, NATURE PUBLISHING GROUP, GB, vol. 29, no. 6, 2 March 2002 (2002-03-02), pages 527-529, XP002424270, ISSN: 0268-3369, DOI: DOI:10.1038/SJ.BMT.1703411 abstract; figure 1 -----	1-28
X	CHIU R W K ET AL: "Non-invasive prenatal diagnosis by single molecule counting technologies", TRENDS IN GENETICS, ELSEVIER SCIENCE PUBLISHERS B.V. AMSTERDAM, NL, vol. 25, no. 7, 1 July 2009 (2009-07-01), pages 324-331, XP026301457, ISSN: 0168-9525, DOI: DOI:10.1016/J.TIG.2009.05.004 [retrieved on 2009-06-18] the whole document ----- -/--	1-28

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2010/055655

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LUN FIONA M F ET AL: "Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, vol. 105, no. 50, December 2008 (2008-12), pages 19920-19925, XP002620453, ISSN: 0027-8424 abstract; figure 1 page 19921, column 1, paragraph 3-5	1-28
X	FAN H C ET AL: "Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES (PNAS), NATIONAL ACADEMY OF SCIENCE, US, vol. 105, no. 42, 21 October 2008 (2008-10-21), pages 16266-16271, XP002613056, ISSN: 0027-8424, DOI: DOI:10.1073/PNAS.0808319105 [retrieved on 2008-10-06] the whole document	1-28
X	CHIU ROSSA W K ET AL: "Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, vol. 105, no. 51, December 2008 (2008-12), pages 20458-20463, XP002620454, ISSN: 0027-8424 the whole document	1-28

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2010/055655

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2007100911 A2	07-09-2007	AU 2007220991 A1	07-09-2007
		CA 2647793 A1	07-09-2007
		EP 1996728 A2	03-12-2008
		US 2008020390 A1	24-01-2008
		US 2011059451 A1	10-03-2011

WO 2009013492 A1	29-01-2009	AU 2008278839 A1	29-01-2009
		AU 2008278843 A1	29-01-2009
		CA 2693081 A1	29-01-2009
		CA 2694007 A1	29-01-2009
		CN 101971178 A	09-02-2011
		CN 101849236 A	29-09-2010
		EA 201000231 A1	30-06-2010
		EP 2183692 A1	12-05-2010
		EP 2183693 A1	12-05-2010
		WO 2009013496 A1	29-01-2009
		JP 2010534068 T	04-11-2010
		JP 2010534069 T	04-11-2010
		KR 20100075826 A	05-07-2010
		KR 20100058503 A	03-06-2010
		US 2009087847 A1	02-04-2009
		US 2009029377 A1	29-01-2009

WO 2004078999 A1	16-09-2004	AU 2004217872 A1	16-09-2004
		EP 1599608 A1	30-11-2005

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-28

Methods of determining at least a portion of the genome of an unborn fetus comprising determining a haplotype inherited by the fetus.

2. claims: 29-33

Methods of identifying a de novo mutation in the genome of an unborn fetus.

3. claims: 34-44

Methods of determining a fractional concentration of fetal DNA in a biological sample taken from a pregnant female and methods of determining a proportion of a fetal genome that has been sequenced from a biological sample taken from a pregnant female.
