(19) Евразийское патентное ведомство

(21) 202090348 (13) A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2020.05.06

(22) Дата подачи заявки
2017.08.04

(51) Int. Cl. *C12Q 1/6869* (2018.01)
G06F 19/18 (2011.01)
G06F 19/22 (2011.01)
C12Q 1/6879 (2018.01)

(54) СПОСОБ НЕИНВАЗИВНОГО ПРЕНАТАЛЬНОГО ВЫЯВЛЕНИЯ АНОМАЛИЙ ПОЛОВЫХ ХРОМОСОМ У ПЛОДА И ОПРЕДЕЛЕНИЯ ПОЛА ПЛОДА В СЛУЧАЕ ОДНОПЛОДНОЙ И ДВУХПЛОДНОЙ БЕРЕМЕННОСТИ

(86) **PCT/EP2017/069795**
(87) **WO 2019/025004 2019.02.07**

(71) Заявитель:
ТРИСОМЫТЕСТ, С.Р.О. (SK)

(72) Изобретатель:
Дюриш Франтишек, Газдарица Юрай, Кухарик Марцель, Гиблова Михаэла, Сземес Томаш, Будиш Ярослав (SK)

(74) Представитель:
Носырева Е.Л. (RU)

(57) В целом настоящее изобретение относится к области неинвазивного пренатального скрининга и диагностики. В настоящем изобретении предусмотрен надежный способ, который применим в практической реализации неинвазивного пренатального скрининга анеуплоидий половых хромосом, таких как моносомия по X (X0, синдром Тернера), XXY (синдром Клайнфельтера), XXX (синдром трисомии X) и XYY (синдром Джейкобс) с использованием образца крови, взятого у матери на ранней стадии беременности. Более того, в настоящем изобретении предусмотрен новый подход к расчету фетальной фракции фрагментов внеклеточной ДНК. В частности, настоящее изобретение относится к одноплодным или двухплодным беременностям, однако предполагается расширение изобретения в отношении трехплодных или четырехплодных беременностей.

202090348 A1

202090348 A1

P51639668EA

СПОСОБ НЕИНВАЗИВНОГО ПРЕНАТАЛЬНОГО ВЫЯВЛЕНИЯ АНОМАЛИЙ ПОЛОВЫХ ХРОМОСОМ У ПЛОДА И ОПРЕДЕЛЕНИЯ ПОЛА ПЛОДА В СЛУЧАЕ ОДНОПЛОДНОЙ И ДВУХПЛОДНОЙ БЕРЕМЕННОСТИ

ОБЛАСТЬ ИЗОБРЕТЕНИЯ

В целом настоящее изобретение относится к области неинвазивного пренатального скрининга и диагностики. В настоящем изобретении предусмотрен способ выявления наличия или отсутствия анеуплоидий половых хромосом, в частности, моносомии и трисомии по хромосоме X, синдрома Клайнфельтера и синдрома XYY по хромосоме Y, определения пола и доли фрагментов ДНК плода с использованием образца крови, взятого у матери на ранней стадии беременности. Более того, в настоящем изобретении предусмотрен новый подход к расчету показателя фетальной фракции внеклеточных фрагментов ДНК, применение которого лежит в его основе и который можно применять отдельно в других областях неинвазивного пренатального скрининга и диагностики. Настоящее изобретение относится к одноплодным или двухплодным беременностям. Предполагается расширение настоящего изобретения в отношении трехплодных или четырехплодных беременностей с помощью технологий будущего.

УРОВЕНЬ ТЕХНИКИ

На сегодняшний день пренатальное исследование является неотъемлемым компонентом акушерской практики. Основной целью пренатального исследования является скрининг в отношении анеуплоидий плода, таких как трисомия по хромосоме 21 (синдром Дауна), трисомия 18 (синдром Эдвардса) и трисомия 13 (синдром Патау). Другая большая группа аномалий включает аберрации половых хромосом (SCA), такие как моносомия по X (X0, синдром Тернера), XXY (синдром Клайнфельтера), XXX (синдром трисомии X) и XYY (синдром Джейкобс). Хотя большинство случаев анеуплоидии у плода приводят к прерыванию беременности во время развития плода, SCA редко смертельны, и их фенотипические свойства являются менее серьезными, чем таковые аберраций аутосомных хромосом. Наиболее распространенная моносомия X проявляется в 1-1,5% случаев беременности и часто обуславливает прерывание беременности в течение первого триместра (прим. 23%). Следовательно, пренатальное

выявление SCA является важным пренатальным генетическим тестом для пренатального скрининга или диагностики. Доступны надежные инвазивные пренатальные тесты, однако из-за связанного с ними риска в настоящее время их проводят только в случаях беременности с высокой степенью риска. Разработка надежного способа неинвазивного пренатального исследования (NIPT) анеуплоидий у плода на данный момент является наиболее важной задачей наблюдения беременности.

Обнаружение циркулирующей внеклеточной эмбриональной ДНК (cffDNA) в материнской крови (*Lo et al., Lancet 350:485-487, 1997*[1]) создало возможность разработки неинвазивных способов, в которых используются нуклеиновые кислоты плода из образца материнской периферической крови для выявления хромосомных аномалий у плода. cffDNA составляет примерно менее 10% общей циркулирующей внеклеточной ДНК (cfDNA) в плазме материнской крови, при этом недавно было обнаружено, что в материнской крови присутствует полный геном плода в форме cffDNA, который, таким образом, является перспективным материалом для NIPD.

В некоторых документах (*Chiu et al 2008*[2], *Fan et al. 2008*[3], *Chiu et al 2011*[4], *Sehnert et al 2011*[5], *Lau et al 2012*[6], *Bianchi et al. 2012*[7], *Straver et al. 2014*[8], *Yu et al. 2014*[9], *Tynan et al. 2016*[10], *EP2183693, EP2366031, US8296076*) раскрыты способы, в которых применяют секвенирование ДНК из материнской крови для получения информации в отношении содержания аберрантных хромосом у плода. Особый интерес в отношении способов выявления SCA представляют собой раскрытые документы *Zimmermann et al. 2012*[11], *Mazloom et al. 2013*[12], *Liang et al. 2013*[13], *Wang et al. 2014*[14]. Во всех вышеуказанных способах применяется анализ общей cfDNA в плазме материнской крови без необходимости выделения специфической для плода ДНК, cffDNA. Эти способы основаны на выявлении дополнительной копии или потери хромосомы для определения отличий нормальных случаев от случаев трисомии или моносомии. В случае трисомии (или моносомии) у плода число копий трисомной (или моносомной) хромосомы в материнской крови является немного большим (или меньшим) по сравнению с другими хромосомами. В основном, это же применимо к аберрациям половых хромосом, где дополнительная или отсутствующая копия принадлежит либо хромосоме X, либо хромосоме Y. Появление массового параллельного секвенирования ДНК (MPS) позволило проводить секвенирование чрезвычайно больших количеств молекул ДНК. Таким образом, секвенирование ДНК нового поколения (NGS) с

недавних пор применяется для неинвазивного выявления трисомий у плода с использованием материнской крови.

В общем, выявление SCA плода, таких как моносомия или трисомия X хромосомы, у плода женского пола с применением NGS выполняют посредством следующего способа. Вначале проводят секвенирование короткого участка на одном конце каждой молекулы ДНК из плазмы материнской крови и картируют их относительно эталонного человеческого генома, чтобы определить хромосомное происхождение каждой последовательности. Далее количество секвенированных меток представляющей интерес хромосомы (например, хромосомы X) сравнивают с каким-либо эталонным значением, полученным из группы эуплоидных образцов. Ключевой этап анализа состоит в оценке доли фрагментов ДНК, происходящих от плода, называемой фетальной фракцией. В случае низкой доли диагностируемый образец может быть ошибочно классифицирован как здоровый на основании показателя содержания фрагментов от матери.

ПРЕДПОСЫЛКИ ИЗОБРЕТЕНИЯ

Способы, описанные *Chiu et al. 2008, Chiu et al. 2011, Sehnert et al. 2011, Bianchi et al. 2012, Lau et al. 2012,* для выявления аутосомных анеуплоидий, таких как трисомия 21 хромосомы, основаны на секвенировании полного генома с помощью MPS и определении z-показателя. Хотя z-показатель является общепринятым в качестве стандартного параметра, используемого для выявления анеуплоидных образцов, существуют различия в его расчете. *Chiu et al. 2011*, раскрыли подход с применением ридов, картированных на всех хромосомах, применяемых в качестве эталона для расчета z-показателя, *Lau et al. 2012,* раскрыли подход с применением ридов, картированных на конкретной хромосоме, например, 14, в качестве эталона для трисомии T21, и *Sehnert et al. 2011*, выбрали хромосому 9 в качестве оптимального внутреннего стандарта для хромосомы 21. Авторы изобретения рассматривают такие способы в качестве подходов, которые можно также применять в отношении SCA, что, в частности, отмечено в некоторых документах.

Способ, описанный в *Zimmermann et al. 2012*, основан на нацеленном секвенировании конкретного полиморфного локуса (SNP). Способ порождает множество гипотез в отношении того, как должны выглядеть данные секвенирования в случае моносомной,

дисомной или трисомной хромосомы с применением конкретной фетальной фракции. Вероятность каждой гипотезы затем рассчитывают с учетом наблюдаемых данных и выбирают наиболее вероятную гипотезу. Часть способа основана на адаптировании проприетарного алгоритма выравнивания Novoalign (Novocraft, Селангор, Малайзия), проприетарного статистического алгоритма Parental Support, применяемого в MATLAB (MathWorks, Natic, Массачусетс, США), и данных *Haptap Database*[15]. Возможным недостатком данного способа в отношении стоимости теста является необходимость секвенирования родительской ДНК для измерения родительских генотипов.

Способ *Mazloom et al. 2013* основан на чрезмерной и недостаточной представленности хромосом X и Y по сравнению с группой эуплоидных образцов от плода женского пола в случае хромосомы X и группой эуплоидных образцов от плода мужского пола в случае хромосомы Y. Способ является схожим со способом, описанным *Chiu et al. 2008*. О z-показателе или вероятности гипотезы в диагностике образца не сообщается. Однако применение z-показателя в отношении хромосомы X лежит в основе определения классификационных областей (см. вспомогательные данные). Древовидную схему решений применяют для классификации образца (45 X, 46 XX, 47 XXX, 47 XXY, 47 XYY или 46 XY).

Согласно способу *Liang et al. 2013*, как сообщается, рассчитывают вероятность трисомии 9-й, 13-й, 18-й, 21-й хромосом, а также вероятность аберрации половых хромосом X0, XXX, XXY и XYY, который основан на нормализованных хромосомных значениях NCR, равняющихся количеству уникально картированных последовательностей хромосомы, представляющей интерес/общему количеству уникально картированных последовательностей относительно всех аутосомных хромосом. Значения NCR используются для получения z-показателя общепринятым в данной области техники способом. Отнесение образцов к эуплоидным или абберантным выполняют с помощью данного z-показателя и различных пороговых значений, различных для каждого случая анеуплоидии.

Способ *Wang et al. 2014* основан на разделении каждой хромосомы на непрерывные бины размером 20 т. п. о. С учетом обучающего набора эуплоидных беременностей рассчитывают специальное нормализованное число ридов для каждого бина и сохраняют медианное значение. Далее для установления приобретения или потери хромосомных областей сходные бины тестируемого образца сравнивают с

сохраненными медианными значениями посредством объединенного алгоритма LASSO (наименее абсолютное уменьшение и оператор выбора). Также описан уровень хромосомного мозаицизма, и его рассчитывают посредством нормализированных хромосомных представленностей NCR как (NCRj - NCRf)/NCRf, где NCRj представляет собой NCR любой хромосомы тестируемого образца, и NCRf представляет собой среднее значение NCR такой же хромосомы в группе стандартных образцов.

В целом существует три основные стадии определения анеуплоидии плода по образцу крови, плазмы крови или сыворотки крови матери: 1) получение образца ДНК и библиотеки ДНК, 2) секвенирование и 3) анализ данных в отношении последовательностей. В последние несколько лет был достигнут значительный прогресс в секвенировании, однако на первой и третьей стадии имеется простор для улучшения, что может привести к снижению стоимости и оказать большое влияние на качество тестирования.

Существуют различные подходы для улучшения первой стадии, как раскрыто, например, в *EP2366031 (Rava, R.P. et al.,* права переданы *Verinata Health, Inc., США).* В этом документе раскрыт способ пренатального скрининга и диагностики хромосомной анеуплоидии плода на основе NGS, включающий новый протокол для получения библиотек секвенирования с использованием материнского образца. Новый подход для получения библиотек секвенирования предусматривает последовательные этапы репарации концов, присоединения dA и лигирования адапторов в указанных нуклеиновых кислотах, и указанные последовательные этапы не включают очистку продуктов репарации концов перед этапом присоединения dA и не включают очистку продуктов присоединения dA перед этапом лигирования адапторов. Способ позволяет определить вариации количества копий (CNV) любой представляющей интерес последовательности. В этом способе z-показатель не имел решающего значения.

Другой вариант улучшения раскрыт в *WO 2011/051283 (Benz, M. et al.,* права переданы *Lifecodexx, AG, Германия).* Раскрытый в нем способ неинвазивной диагностики хромосомной анеуплоидии улучшен путем обогащения и количественной оценки выбранных последовательностей cfDNA в образце материнской крови.

<u>Описание уровня техники для оценки фетальной фракции</u>

Поскольку плод мужского пола имеет пару половых хромосом (X и Y), отличную от материнской (две копии X), то фетальную фракцию обычно оценивают на основании показателя содержания фрагментов в этих хромосомах. Данный подход, однако, ограничен образцами от плода мужского пола. Фракция для плода женского пола может быть оценена на основании характеристик, отличающихся между фрагментами от плода и матери.

Отличия в локализации фрагмента по всему геному применяли в способе SeqFF (*Kim et al. 2015*[16]). Они демонстрируют, что фетальные фракции отличаются в различных областях генома и в значительной степени коррелируют с содержанием GC и наличием экзонных областей. Следовательно, в способе используют нормализированные сгруппированные значения в качестве переменных прогностического параметра и оценивают фетальную фракцию с применением стандартных моделей многомерной регрессии.

В способе Sanefalcon (*Straver et al. 2016*[17]) применяется более подробная информация о происхождении фрагмента на основе различных механизмов разрушения фрагмента. В нем используется доля соответствующих фрагментов с предварительно рассчитанными местами локализации нуклеосом в геноме в качестве прогностического параметра фетальной фракции на основании предположения о том, что фрагменты от матери появляются в положениях нуклеосом более часто, чем фрагменты от плода.

Другой перспективной характеристикой является длина фрагмента. Поскольку фрагменты от плода обычно короче, профили длины фрагментов от матери и плода различаются в значительной степени и могут применяться в качестве прогностического параметра фетальной фракции. Основанный на длине фрагментов способ был предложен в *Yu et al. 2014*[18].

Геном плода наследует геномную информацию в равной степени от отца и матери. В зависимости от происхождения он различается в определенных положениях, особенно по точечным изменениям нуклеотидов, называемым SNP. Фрагменты, которые соответствуют SNP, характерным для отца, наиболее вероятно принадлежат к фетальной фракции и могут применяться в прогнозировании. В способе FetalQuant (*Jiang et al. 2016*[19]) применяется данная информация; однако необходим еще один лабораторный анализ для создания карты генома родителей. Применение способа,

таким образом, является слишком времязатратным и дорогостоящим для массовой диагностики.

Несмотря на существование нескольких способов неинвазивного выявления анеуплоидии половых хромосом плода, до сих пор существует потребность в альтернативном способе, который был бы по меньшей мере таким же чувствительным и специфическим, как существующие на сегодняшний день способы, и в то же время менее дорогостоящим. Кроме того, существующие способы для прогнозирования с помощью фетальной фракции до сих пор ограничены их точностью или необходимостью дополнительных лабораторных анализов.

ОПИСАНИЕ НАСТОЯЩЕГО ИЗОБРЕТЕНИЯ

В настоящем изобретении предусмотрен альтернативный и надежный способ, который применим к практической реализации неинвазивного пренатального скрининга анеуплоидий половых хромосом, таких как моносомия по X (X0, синдром Тернера), XXY (синдром Клайнфельтера), XXX (синдром трисомии X) и XYY (синдром Джейкобс). Он обеспечивает одновременную диагностику анеуплоидии, определение пола и фетальной фракции с применением одной модели, которая может быть с легкостью расширена в отношении других хромосомных анеуплоидий и расстройств, таких как моносомия или трисомия 13-й, 18-й, 21-й или любой другой хромосомы. Кроме того, данным способом предусматривается применение лежащего в его основе способа расчета доли фетальной фракции фрагментов cfDNA в материнской крови, который может применяться отдельно от определения SCA в других областях, таких как выявление (без ограничений) трисомий T21, T18 или T13. Кроме того, данный способ требует относительно небольшого количества данных секвенирования, следовательно, способ является относительно недорогим и будет доступным даже для небольших учреждений здравоохранения.

В следующем описании будут разъяснены основные признаки способа согласно настоящему изобретению; однако это не означает, что изобретение должно включать все описанные в данном документе признаки и аспекты. Настоящее изобретение станет полностью понятно для специалиста после изучения следующего описания совместно с примерами, в которых некоторые конкретные признаки и аспекты будут разъяснены более подробно.

Используемые в данном документе технические и научные термины имеют то же значение, которое обычно подразумевается специалистами в области медицины, молекулярной генетики, пренатальной неинвазивной диагностики, молекулярной биологии и биоинформатики. Некоторые конкретные термины разъяснены ниже.

Определения

Термин *анеуплоидия* применяется в общем смысле, известном специалисту в данной области техники, он означает дисбаланс генетического материала, обусловленный потерей или приобретением целой хромосомы или части хромосомы. Другими словами, он означает наличие целой избыточной хромосомы или отсутствие целой хромосомы, или частичную хромосомную дупликацию, или делецию значительного размера (> 1 т. п. о.). В частности, *T21*, *T18* и *T13* обозначают наиболее распространенные типы аутосомных трисомий, которые сохраняются до рождения у людей, при этом трисомия по хромосоме 21 приводит к синдрому Дауна, трисомия по хромосоме 18 приводит к синдрому Эдвардса и трисомия по хромосоме 13 приводит к синдрому Патау. Кроме того, *X0, XXX, XXY* и *XYY* обозначают наиболее распространенные типы анеуплоидий половых хромосом, которые сохраняются до рождения у людей, при этом моносомия по хромосоме X приводит к синдрому Тернера, трисомия по хромосоме X приводит к синдрому трисомии X, кариотип 47 XXY приводит к синдрому Клайнфельтера и кариотип 47 XYY приводит к синдрому Джейкобс.

Термин *массовое параллельное секвенирование* (MPS), применяемый в комбинации с *секвенированием нового поколения* (NGS), означает широко известные специалистам в данной области техники методики секвенирования огромного количества, т. е. порядка миллионов или десятков миллионов фрагментов ДНК, как это практикуют в данной области техники с применением анализаторов Illumina или IonTorrent. Протоколы полногеномного секвенирования известны специалисту в данной области техники и могут быть найдены в примерах ниже.

Риды последовательностей представляют собой короткие последовательности ДНК, полученные с помощью NGS-секвенирования, однако достаточно длинные (например, по меньшей мере приблизительно 30 п. о.), чтобы служить в качестве *меток* последовательности, т. е. с их помощью можно однозначно обозначить одну из

хромосом (1-22, X, Y). Обычно может допускаться небольшая степень несовпадения (1 п. о.). В действительности метки представляют собой относимые или, вернее, картируемые риды. Метки уникальным образом картированы на эталонный геном, т. е. они отнесены к одному местоположению в эталонном геноме. Метки, которые можно картировать в более чем одном местоположении в эталонном геноме, т. е. метки, которые не картируются уникальным образом, исключают из анализа.

Термин «*картирование*» означает выравнивание последовательности после NGS (т. е. фрагмента ДНК, положение которого в геноме неизвестно) с совпадающей последовательностью в эталонном человеческом геноме. Это может быть осуществлено несколькими путями, авторы изобретения применяли способ Liu et al 2014. Применяемый в данном документе термин *эталонный геном человека* или *эталонный геном* означает *последовательность hg19*[20].

Применяемые в данном документе термины «*выровненный*» или «*выравнивание*» относятся к одной или более последовательностям, которые с точки зрения порядка их молекул нуклеиновых кислот идентифицируются как совпадающие с известной последовательностью из эталонного генома. Такое выравнивание может быть произведено в ручную или с помощью компьютерных алгоритмов, которые широко известны специалистам в области молекулярной биологии и биоинформатики. Совпадение рида последовательности при выравнивании может составлять 100% совпадения последовательностей или менее 100% (неидеальное совпадение).

Фетальная фракция представляет собой долю фрагментов cfDNA, происходящих от плода, по сравнению со всеми секвенированными фрагментами.

Нормальное распределение (или *распределение Гаусса*) является распространенным непрерывным распределением вероятностей. Оно определяется функцией плотности вероятностей $f(x|\mu,\sigma)f(x|\mu,\sigma)$

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Авторы изобретения будут часто обозначать данную функцию как $N(\mu,\sigma^2)N(\mu,\sigma^2)$. Если некоторая стохастическая переменная X имеет нормальное распределение, авторы изобретения будут обозначать ее с помощью $X \sim N(\mu,\sigma^2)$ $X \sim N(\mu,\sigma^2)$.

Искусственная нейронная сеть представляет собой вычислительный подход, применяемый в компьютерных науках и других исследовательских дисциплинах, который основан на большой совокупности нейронных процессоров (искусственных нейронов), приблизительно имитирующих путь, с помощью которого биологический головной мозг решает проблемы с помощью крупных кластеров биологических нейронов, соединенных с помощью аксонов. Каждый нейронный процессор соединен с множеством других, и эффект связей может быть усиливающим или подавляющим по отношению к состоянию активации соединенных нейронных процессоров. Каждый отдельный нейронный процессор может иметь функцию суммирования, которая объединяет значения всех его входных данных вместе. Каждое соединение и сам процессор могут предусматривать пороговую функцию или предельную функцию таким образом, что сигналы должны преодолеть предел перед распространением к другим нейронам.

Формат *FASTQ*[21] представляет собой текстовый формат файлов для хранения как биологической последовательности (обычно нуклеотидной последовательности), так и ее соответствующих показателей качества. Как код последовательности, так и показатель качества для краткости закодированы одним символом ASCII. Он изначально был разработан в Институте Сенгера Wellcome Trust для объединения последовательности FASTA и ее данных о качестве, но с недавних пор фактически стал стандартным для хранения выходных результатов приборов высокопроизводительного секвенирования, таких как геномный анализатор Illumina.

SAM[22] представляет собой текстовый формат файлов для хранения биологических последовательностей, выравненных относительно эталонной последовательности, разработанный Heng Li. Сокращение SAM означает выравнивание последовательностей/карта. Он широко применяется для хранения данных, таких как нуклеотидные последовательности, полученных с помощью технологий секвенирования нового поколения.

Сдвиг содержания GC описывает зависимость между числом фрагментов (охватом ридов) и содержанием GC, обнаруживаемым в данных секвенирования Illumina. Данный сдвиг может подавлять сигнал, представляющий интерес для анализов, которые сосредоточены на измерении показателя содержания фрагментов в геноме, такого как оценка количества копий. Сдвиг не является единообразным среди образцов;

и нет единого мнения о лучших методах его устранения в одном образце. Существует множество предложенных подходов к коррекции этого сдвига, таких как в *Benjamini and Speed 2012*[23] или *Liao et al. 2014*[24].

Термин *гипотеза* в контексте данного патента относится к возможному статусу аберраций половых хромосом плода совместно с фетальной фракцией фрагментов cfDNA в материнской крови, плазме крови или сыворотке крови. Часть способа в соответствии с настоящим изобретением представляет собой формулирование такой гипотезы с помощью теории вероятностей и расчет вероятности каждой такой гипотезы на основе наблюдений, сделанных на основе фрагментов cfDNA, обнаруженных в вышеупомянутых крови, плазме крови или сыворотке крови.

КРАТКОЕ ОПИСАНИЕ СПОСОБА В СООТВЕТСТВИИ С НАСТОЯЩИМ ИЗОБРЕТЕНИЕМ

Настоящее изобретение относится к способу определения анеуплоидии половых хромосом плода с использованием материнского образца крови, плазмы крови или сыворотки крови содержащих смесь фрагментов ДНК, происходящих от плода и матери, где смесь фрагментов ДНК, происходящих от плода и матери, представляет собой циркулирующие молекулы внеклеточной ДНК, при этом указанный способ включает четыре основные стадии: 1) получение и обработку образцов материнской крови, 2) получение образца ДНК и библиотеки ДНК, 3) секвенирование и 4) анализ данных в отношении последовательностей с получением прогноза или, иными словами, диагноза. Важной частью способа является получение обучающих данных, которые необходимо обработать с помощью такого же лабораторного способа, что и тестируемые образцы (образцы, подвергаемые изучению).

Отправной точкой настоящего способа является анализ материнского образца, т. е. периферической крови, следовательно, способ является неинвазивным. Периферическая кровь содержит внеклеточную ДНК (cfDNA), которая представляет собой смесь фрагментов материнской ДНК и ДНК плода (cffDNA). Указанные фрагменты также дополнительно называются фрагментами cfDNA или кратко фрагментами ДНК. Значение имеет CffDNA, следовательно, ДНК плода можно обогатить (посредством отбора более короткого фрагмента либо *in silico*, либо посредством физического отбора, как описано в *Minarik et al. 2015*[25]). Затем образец

общей ДНК (по-прежнему содержащий фрагменты ДНК, происходящие как от матери, так и от плода, т. е. это смешанный образец) подвергают массовому параллельному секвенированию с помощью подхода NGS с получением огромного количества коротких ридов последовательностей. Эти риды служат метками последовательности, т. е. они соответствуют определенной геномной области хромосомы.

В настоящем способе определяют вероятность наблюдения данных NGS-секвенирования для хромосомы X и Y с учетом гипотезы в отношении половых хромосом плода (например, плода женского пола с синдромом Тернера). Строят различные гипотезы по отношению к различным SCA плода и выбирают гипотезу с максимальной вероятностью в качестве наиболее вероятного события. Гипотезы сформулированы для одноплодной или двухплодной беременности, плода мужского или женского пола, с эуплоидными или абберантными половыми хромосомами и для всех фетальных фракций (отдельные значения находятся в диапазоне от 5% до 100% с шагом 0,1%). Для увеличения точности выбора гипотезы априори распределение фетальной фракции (т. е. вероятность того, что конкретное значение является истинной фетальной фракцией cfDNA в данном образце) рассчитывают с использованием других независимых данных секвенирования (но все еще содержащихся в данных NGS-секвенирования того же образца, поэтому стоимость теста не увеличивается). В дополнение к этому, собирают показатели априорной частоты встречаемости SCA, полученные в результате настоящего исследования, для масштабирования предлагаемых гипотез, так что они отражают реальные ожидания (например, ожидание двух плодов женского пола с синдромом Тернера менее вероятно, чем ожидание наличия двух эуплоидных плодов женского пола). Таким образом, согласно способу

1) данные NGS-секвенирования хромосом X и Y,

2) априорное распределение фетальной фракции, полученное с помощью других аспектов тех же данных NGS (например, длины фрагментов),

3) априорная информация относительно распространения SCA в популяции (опубликованные данные или собственные исследования или и то и другое)

включены в одну согласованную модель для получения наиболее вероятного диагноза.

Улучшение по сравнению с уровнем техники многообразно.

1) Новый подход к выявлению SCA в случае отдельных беременностей.

2) Совместное включение выявления SCA в случаях двухплодных беременностей.

3) Способ применим в отношении моносомии или трисомии любой аутосомы плода в случае одноплодной беременности непротиворечивым образом.

4) Легкость теоретического расширения до любого большего числа плодов (например, тройни). Однако существующие технологии секвенирования пока не позволяют осуществить это с разумными затратами в широкой практике.

5) Установление (путем отдельного расчета) *распределения* фетальной фракции согласно способу по сравнению с *одним численным значением* в предыдущем уровне техники. Также данное распределение фетальной фракции может применяться по отдельности в других областях.

6) Введение отдельно рассчитанного распределения фетальной фракции в расчет показателей SCA.

7) Введение данных о распространенности SCA в популяции в вычисление. Это является особенно ключевым в случаях двухплодной беременности, где различные SCA могут компенсировать показатели дефицита/избытка в данных NGS друг друга.

8) Для способа не является необходимым секвенирование родительской ДНК, как в *Zimmermann et al. 2011.*

9) Способ более приспособлен для рассматриваемых проблем при сравнении со способами, адаптированными для выявления трисомий, таких как в *Chiu et al. 2008, Chiu et al. 2011, Sehnert et al. 2011, Bianchi et al. 2012, Lau et al. 2012.*

10) Улучшения в обработке данных перед анализом (относительно применения поправки на сдвиг содержания GC см. ниже).

ПОДРОБНОЕ ОПИСАНИЕ СПОСОБА В СООТВЕТСТВИИ С НАСТОЯЩИМ ИЗОБРЕТЕНИЕМ

Ниже авторы изобретения предполагают, что каждый образец (т. е. тестируемый образец, а также обучающий образец), указанный в данном описании, является

результатом обработки образца крови в лаборатории в соответствии с описанием ниже (см. пример 2) и впоследствии подлежит секвенированию нового поколения, т. е. представляет собой совокупность данных NGS (секвенированные фрагменты cfDNA), сохраненных в формате FASTQ (или эквивалентном ему). Более того, авторы изобретения будут предполагать, что файл FASTQ образца претерпел процесс картирования, где содержание данного файла выровнено относительно эталонного человеческого генома (такого как hg19[26]), и данная информация о картировании хранится в файле формата SAM (или эквивалентном ему). Наконец, авторы изобретения будут предполагать, что содержание файла SAM было обработано с помощью необязательной фильтрации данных, такой как удаление неопределенных результатов картирования, или внесения поправки на сдвиг содержания GC, за счет чего авторы изобретения знают, касательно данного образца, количество фрагментов cfDNA, которые происходят из какой-либо хромосомы в абсолютном, а также относительном значении (в данном документе термин «относительный» понимают в отношении общего количества фрагментов cfDNA в файле SAM, возможно после фильтрации и внесения поправки на сдвиг). Способ, описанный ниже, не является особенно зависимым от осуществления описанной обработки данных (при условии, что полученные данные имеют достаточное качество). Вышеуказанная обработка данных является широко известной в предыдущем уровне техники и, кроме улучшений авторов изобретения (см. раздел *Улучшения в обработке исходных данных*), не будет описана дополнительно. Фрагменты cfDNA также дополнительно взаимозаменяемо называются фрагментами ДНК или просто cfDNA.

Способ в соответствии с настоящим изобретением предусматривает анализ вероятности нескольких нестандартных гипотез о состоянии половых хромосом плода (или плодов в случае двухплодной беременности). Случай беременности, т. е. одноплодная или двухплодная, указывается оператором перед осуществлением способа (тип беременности определяется с помощью ультразвукового исследования у матери). Согласно способу осуществляют стадии, описанные ниже, некоторые из которых необходимо проводить только один раз. Сбор обучающих данных и формулировку гипотезы необходимо проводить только один раз в начале осуществления способа на практике. Обучение модели нейронной сети для расчета распределения фетальной фракции также должно проводиться только один раз. С другой стороны, распределение фетальной фракции должно быть рассчитано с помощью обученной нейронной сети

для каждого тестируемого образца. Подобным образом вероятность каждой гипотезы с учетом конкретных данных NGS должна быть рассчитана для каждого тестируемого образца по отдельности.

Сбор обучающих данных. Во-первых, авторы изобретения собирают группу образцов от женщин (размером по меньшей мере 100) и группу образцов от мужчин (размером по меньшей мере 100). Авторы изобретения будут обозначать первый обучающий набор как *Tf* и второй как *Tm*. С помощью первого обучающего набора *Tf* авторы изобретения получают среднее значение показателя картирования для хромосомы X (т. е. среднее значение фракции секвенированных ридов, картированных на хромосоме X). Однако авторы изобретения в действительности наблюдают среднее значение показателя картирования для двух хромосом X, поскольку у эуплоидных женщин имеются две копии хромосомы X. Авторы изобретения применяют статистический анализ для определения показателя картирования для одной хромосомы (описано ниже). Кроме того, даже в случае женщин обычно существуют некоторые риды, картированные на хромосоме Y, которые являются преимущественно ошибочно картированными из областей, гомологичных хромосоме X. Такие результаты картирования также относятся к совокупности *Tf*. Таким образом, авторы изобретения создают *условные хромосомы*, в данном документе называемые хромосомами Z, которые обеспечивают получение ридов, картированных на хромосоме Y, из хромосомы X. Число хромосом Z будет зависеть от числа хромосом X. Следовательно, образец от женщины будет иметь две хромосомы Z. Плод с синдромом Тернера будет иметь только одну хромосому Z, в то время как плод с синдромом трисомии X будет иметь три. Женщина с эуплоидным плодом женского пола будет иметь в сумме две (одна часть из двух будет происходить от матери, и другая часть будет происходить от плода). Женщина с плодом женского пола, у которого имеется синдром Тернера, будет иметь фракцию, среднюю между одной и двумя, которая будет зависеть от фетальной фракции фрагментов cfDNA, принадлежащих плоду (мать будет являться источником двух хромосом Z с ее стороны, но плод будет являться источником только одной; таким образом, сумма хромосом Z будут зависеть от доли фрагментов cfDNA, источником которых являются мать и плод, - фетальной фракции). Кроме того, мужские образцы или образцы с плодом мужского пола также будут иметь хромосомы Z. Это связано с тем, что такие образцы также содержат хромосомы X, которые также являются источником ошибочных ридов, картированных на хромосоме Y (число хромосом X и, таким образом, хромосом Z

зависит от статуса пола образца и аберраций половых хромосом). Авторы изобретения рассчитывают среднее значение показателя картирования для хромосомы Z с помощью первого обучающего набора *Tf*. Авторы изобретения применяют те же статистические подходы, что и в случае с хромосомой X, так как существуют две хромосомы Z, поскольку существуют две хромосомы X в каждом образце из *Tf*. Наконец, авторы изобретения получают среднее значение показателя картирования для хромосомы Y с помощью обучающего набора *Tm* от мужчин. Таким образом, авторы изобретения имеют три обученных численных значения, с данного момента называемых *muX, muY, muZ* (математические подробности представлены ниже). Кроме того, авторы изобретения рассчитывают значения стандартного отклонения этих численных значений аналогичным образом (математические подробности ниже), которые дополнительно обеспечивают три численных значения, с данного момента называемых *sdX, sdY, sdZ*.

В дополнение к обучающим наборам *Tf* и *Tm* авторы изобретения собирают группу образцов (по меньшей мере 300) с одним эуплоидным плодом мужского пола и различными значениями фетальной фракции *f* (находящимися в диапазоне от 5% до 30%, определенными с помощью хромосомы Y, как указано в предыдущем уровне техники), которая будет применяться для обучающей прогностической модели фетальной фракции (см. ниже). Данный обучающий набор будет называться *Te*.

Следует обратить внимание, что выбор трех обучающих наборов *Tf*, *Tm* и *Te* не зависит от образца(-ов), подлежащего(-их) тестированию в отношении SCA. Авторы изобретения требуют лишь, чтобы отбор образцов для обучающего набора не был предвзятым, т. е. образцы получены случайным образом из большой популяции и ни один образец не повторяется в каком-либо из обучающих наборов.

Формулирование гипотез. Как только авторы изобретения получили шесть численных значений, описанных в предыдущем абзаце (*muX, muY, muZ, sdX, sdY, sdZ*), авторы изобретения определяли гипотезы картирования для каждой аберрации половых хромосом, включая два эуплоидных случая (X0, XX, XXX, XY, XXY, XYY для случаев одноплодной беременности и все их комбинации в случаях двухплодной беременности (например, X0-X0 для близнецов с синдромом Тернера, X0-XX для близнецов, где один является эуплоидным, а у другого имеется синдром Тернера, XX-XY для эуплоидных близнецов мужского и женского пола, и так далее). В каждой такой гипотезе авторы

изобретения предполагают, что мать является эуплоидной, т. е. она не имеет никаких аберраций половых хромосом, а также каких-либо крупных субхромосомальных делеций или вставок. Более того, каждая из таких гипотез параметризована с помощью значения фетальной фракции фрагментов cfDNA в материнской крови. В случае одноплодной беременности данная фетальная фракция будет обозначена символом *f*. В случае двухплодной беременности существуют две фетальные фракции, и авторы изобретения будут обозначать их символами *f1* и *f2*.

Распределение фетальной фракции - одноплодная беременность. Естественно, авторы изобретения ожидают наблюдение фетальной фракции на уровне 10% более вероятным, чем, скажем, на уровне 60%. Для количественного определения такого ожидания с математической строгостью авторы изобретения оценивают распределение фетальной фракции, т. е. функции вероятности, которая приписывает любому дискретизированному значению фетальной фракции вероятность являться истинной фетальной фракцией фрагментов cfDNA (для конкретного образца), с применением других доступных данных секвенирования, а именно значений длины фрагментов cfDNA. Следует обратить внимание, что авторы изобретения применяют дискретизированные значения фетальной фракции, поскольку авторы изобретения не нуждаются в точности, превышающей 0,1% (различие в исходных данных в популяции находится в порядке процентов). Способ определения фетальной фракции с помощью длины фрагмента описан в *Yu et al. 2014*. Однако в данной публикации согласно способу для фетальной фракции данного образца приписывается только одно значение. При таком подходе невозможно принимать во внимание ошибку оценки авторов публикации. При подходе авторов изобретения, с другой стороны, применяется ошибка оценки для получения распределения фетальной фракции, что представляет собой более надежный подход, поскольку он включает больше информации. Таким образом, авторы изобретения не применяют какие-либо алгоритмы, описанные в цитируемой публикации. Скорее авторы изобретения предлагают новый подход, пока еще не описанный в предыдущем уровне техники (см. ниже). Авторы изобретения затем применяют данное расчетное значение распределения фетальной фракции в качестве априорной информации для оценки SCA. Для дальнейших ссылок авторы изобретения обозначают распределение фетальной фракции как $N(muF, sgF^2)$.

Распределение фетальной фракции - двухплодная беременность. В настоящее время не существует известного способа, с помощью которого можно определить фетальные фракции *f1* и *f2* в случае двоен. Пока такой способ находится на стадии разработки, авторы изобретения предполагают, что распределение фетальной фракции является одинаковым для обоих плодов, т. е. $N(muF1, sgF1^2)= N(muF2, sgF2^2)= N(muF/2, sgF^2/2)$. Здесь $N(muF, sgF^2)$ представляет собой распределение фетальной фракции, определенное как для одноплодной беременности.

Выбор гипотезы. Наконец, рассчитывают вероятность в отношении полученных данных NGS-секвенирования согласно каждой гипотезе и для каждой фетальной фракции *f* (соотв. *f1* и *f2*), где фетальная фракция *f* дискретизированна, т. е. она начинается с 5% и с шагом 0,1% доходит до 100% (авторы изобретения требуют наличия по меньшей мере 5% фетальной фракции фрагментов cfDNA в образце). Также принимается во внимание вероятность каждой фетальной фракции с учетом определенного распределения фетальной фракции. Математические подробности приведены ниже. Гипотезу и фетальную фракцию с максимальной вероятностью выбирают в качестве наиболее вероятного случая для данного образца. Более того, вероятность рассматриваемой гипотезы масштабируют на основе ее распространения в популяции. Такая информация может быть извлечена из текущего исследования об анеуплоидии плода.

Интерпретация гипотезы. В случае одноплодной беременности интерпретация выбранной гипотезы понятна. Авторы изобретения обсуждают несколько таких случаев в примерах ниже. В случае двухплодной беременности интерпретация не так понятна, поскольку некоторые результаты не могут интерпретироваться однозначно. Например, два эуплоидных плода женского пола с эквивалентной фетальной фракцией *f1=f2* будут выглядеть очень сходными с плодами женского пола, где у одного имеется синдром Тернера, а у одного имеется трисомия хромосомы X (здесь дополнительная хромосома X одного плода компенсирует отсутствие X хромосомы у другого). В данном случае данные о распространенности SCA в популяции обеспечивают помощь при принятии решения. Кроме того, знание о том, что существует всего две возможности (XX+XX или X0+XXX), может помочь оператору в предложении дополнительных тестов, которые обеспечат весомый ответ. Больше подробностей будет приведено ниже в примерах.

Статистический подход для определения средних значений показателя картирования

Рассматриваем обучающий набор *Tf*. Пусть *S* будет любым ее образцом. Затем фрагменты cfDNA, которые картированы на хромосоме X (и аналогично на хромосоме Z) происходят из любой из двух хромосом X образца. Однако авторы изобретения могут наблюдать только одно совокупное количество всех фрагментов ДНК, картированных на хромосоме X. Формально авторы изобретения могут моделировать данную ситуацию как сумму двух независимых случайных переменных: $X1$ и $X2$. Авторы изобретения предполагают, что обе случайные переменные нормально распределены с неизвестными параметрами muX и sdX (формально $Xi \sim N(muX, sdX^2)$, $i=1,2$). Из теории статистики это указывает, что сумма $a*X1+b*X2$ (где a, b представляют собой некоторые реальные постоянные) нормально распределенных независимых случайных переменных $X1$ и $X2$ представляет собой, опять же, нормально распределенную случайную переменную O. Это также указывает, что $O \sim N(muO, sdO^2)$, где $muO = a*muX + b*muX = (a + b)*muX$ и $sdO^2 = a^2*sdX^2 + b^2*sdX^2 = (a^2 + b^2)*sdX^2$. Поскольку обе хромосомы X из образца S, вероятно, в равной степени вносят вклад в наблюдаемое количество фрагментов, картированных на хромосоме X, авторы изобретения также получают $a = b = 1/2$. Таким образом, $muO = muX$ и $sdO^2 = ½*sdX^2$. Численные значения muO и sdO^2 могут быть определены с помощью обучающего набора *Tf*. Следовательно, авторы изобретения получают искомые численные значения muX и sdX^2.

Численные значения muZ и sdZ^2 определяют аналогично из картированных на хромосоме Y фрагментов в обучающем наборе *Tf*.

С другой стороны, численные muY и sdY^2 могут быть определены непосредственно в мужском обучающем наборе *Tm*, поскольку существует только одна хромосома Y в любом образце из данного набора (авторы изобретения предполагают, что обучающий набор включает только эуплоидных мужчин). Однако для учета ошибок, определенных количественно хромосомой Z, авторы изобретения должны вычесть значение muZ из среднего значения соотношения хромосомы Y, полученного с помощью обучающего набора *Tm*, с целью получить лучшее значение muY. Подобным образом авторы изобретения должны вычесть sdZ^2 из дисперсии соотношения хромосомы Y, полученного с помощью обучающего набора *Tm*, с целью получить лучшее значение sdY^2.

Определение априорного распределения фетальной фракции - одноплодная беременность

Авторы изобретения основывали их модель фетальной фракции на длине фрагментов. Этот аспект данных NGS содержит достаточно информации для удовлетворительной оценки фетальной фракции как показано в *Yu et al. 2014,* и его природа не зависит от данных картирования хромосом. В то время как согласно способу предыдущего уровня техники применяют различные соотношения для обучения и моделирования корреляции фетальной фракции с распределением длины фрагмента с целью получить одно итоговое значение, подход авторов изобретения основан на искусственной нейронной сети. Более того, подход авторов изобретения приводит к вероятности распределения, а не к одному численному значению.

Нейронная сеть авторов изобретения состоит из двух слоев: базового слоя для входных узлов и выходного слоя, состоящего из одного выходного узла (может быть разработана более сложная сеть со скрытыми слоями в случае достаточно большого обучающего набора). Более того, каждый входной узел соединен с выходным узлом. Подробное описание сети можно найти в примере 1 ниже.

Во-первых, нейронную сеть обучают на третьей части образцов из обучающего набора *Te*. Данные секвенирования образцов, а именно картированные фрагменты cfDNA из всех хромосом, классифицируют в соответствии с их длиной, что приводит к получению гистограммы данных (классификационные значения длины ограничены всеми значениями длины, выраженными в парах оснований, от 100 п. о. до 200 п .о., все другие фрагменты отбраковывали). Нейронную сеть обучают обычным способом, известным специалисту в данной области техники. Направляющие значения во время обучения представляют собой значения фетальной фракции, полученные с использованием фрагментов cfDNA, картированных на хромосоме Y, как описано в предыдущем уровне техники.

Во-вторых, как только нейронная сеть обучена, авторы изобретения рассчитывают значение фетальной фракции второй трети образцов из обучающего набора *Te* в соответствии с обученной нейронной сетью, а также хромосомой Y (как описано в предшествующем уровне техники). Посредством сравнения этих двух совокупностей фетальной фракции авторы изобретения могут подобрать линейную модель $f(x) = a*x +$

b в отношении данных для корректировки любого несоответствия между основанными на хромосоме Y и основанными на длине значениями фетальной фракции.

Наконец, последнюю треть образцов из обучающего набора *Te* применяют для оценки ошибок основанных на длине значений фетальной фракции. Авторы изобретения рассчитывают значение фетальной фракции данной последней трети обучающих образцов в соответствии с нейронной сетью с линейным преобразованием, а также с хромосомой Y. Попарно сравнивая такие значения для каждого образца, авторы изобретения рассчитывают стандартное отклонение отличий между основанными на длине и основанными на хромосоме Y значениями фетальной фракции.

Как только нейронная сеть с линейным преобразованием обучена, она может непосредственно применяться для прогнозирования фетальной фракции данного образца. Более того, путем сравнения прогнозированной фетальной фракции с известной, основанной на хромосоме Y, фетальной фракцией в последней трети обучающего набора *Te* авторы изобретения могут рассчитать ошибку прогнозирования. Таким образом, вместо применения конкретного значения для расчетного значения фетальной фракции, как указано в способе предыдущего уровня техники (который может отличаться от истинного значения на некоторое неизвестное количество), авторы изобретения применяют нормальное распределение $N(muF, sgF^2)$, указывая вероятность распределения для значений фетальной фракции *f*, где *muF* представляет собой прогноз, полученный авторами изобретения из нейронной сети для данного тестируемого образца, и sgF^2 представляет собой измерение ошибки, полученное авторами изобретения с помощью обучающего набора. Из данного подхода следует, что ширина sgF^2 распределения $N(muF, sgF^2)$ является одинаковой для всех образцов, в то время как центральная *muF* изменяется с каждым образцом. Данный подход к оценке фетальной фракции с помощью нейронной сети и оценки ошибки еще не был описан в предыдущем уровне техники и обеспечивает дополнительную информацию с повышением таким образом точности способа в соответствии с настоящим изобретением.

Определение априорного распределения фетальной фракции - двухплодная беременность

Поскольку не существует известного способа определения отдельных распределений фетальной фракции в случае двухплодной беременности, авторы изобретения будут предполагать, что они привносят равный вклад в совокупное количество фрагментов cfDNA плода в материнской крови. Авторы изобретения могут рассчитать совокупное распределение фетальной фракции, как в случае одноплодной беременности, с получением таким образом $N(muF, sgF^2)$. Затем, согласно той же теории вероятностей, которую авторы изобретения применяли в разделе *Статистический подход для определения средних значений показателя картирования*, авторы изобретения выяснили, что распределение фетальной фракции для одного плода при двухплодной беременности представляет собой $N(muF/2, sgF^2/2)$. Согласно предположению авторов изобретения то же справедливо и для другого плода. Авторы изобретения будут обозначать функцию плотности распределения фетальной фракции первого плода с применением $N(muF1, sgF1^2)$ и второго плода с применением $N(muF2, sgF2^2)$.

Формулировка гипотезы – одноплодная беременность

Каждая из представленных гипотез будет иметь код, указывающий на то, насколько много каждой из хромосом X, Y или Z присутствует в материнском источнике и источнике плода. Следует обратить внимание, что материнская часть никогда не изменяется (закодирована как Mxxzz), поскольку предполагается, что мать всегда имеет две целые хромосомы X и две целые условные хромосомы Z. Однако, если у авторов изобретения имеется какая-либо конкретная информация относительно распределения хромосомы матери, авторы изобретения могут очень просто корректировать гипотезы для адаптации данного изменения. По данной причине следующие гипотезы также должны быть рассмотрены в качестве примера, согласно которым одна гипотеза может являться источником множества других, возможно конкретных для данного случая, индивидуальных гипотез. Часть кода, относящаяся к плоду (начинается с F), изменяется в зависимости от аберрации половых хромосом плода (например, эуплоидный плод женского пола закодирован Fxxzz, эуплоидный плод мужского пола закодирован Fxzy, поскольку он имеет одну хромосому X, одну условную хромосому Z и одну хромосому Y, и так далее). Опять же, легко скорректировать гипотезы в отношении части, относящейся к плоду, для адаптации другой информации плода, специфической в отношении хромосом (такой как большие делеции или вставки), которую авторы изобретения не хотят включать в анализ.

Авторы изобретения сохраняют материнскую часть кода, несмотря на то, что она не изменяется, для испытания предположения о том, что мать должна быть здоровой для правильной работы данной гипотезы. Однако, как было подчеркнуто выше, гипотезы для конкретного случая могут быть с легкостью разработаны для учета любой абберации материнской части.

Эуплоидный плод женского пола - MxxzzFxxzz

Для X хромосомы существует четыре источника фрагментов в данном случае: две материнских хромосомы X и две хромосомы X плода. Авторы изобретения связывают такие источники с четырьмя случайными переменными X1, X2, X3 и X4, все из которых имеют одинаковое нормальное распределение N (muX, sdX^2). Следовательно, наблюдаемое совокупное количество фрагментов, картированных на хромосоме X, представляет собой случайную переменную $O = (1-f)*X1 + (1-f)*X2 + f*X3 + f*X4$, которая опять же нормально распределена как $N(muO, sdO^2)$ при $muO = 2*muX$, $sdO^2 = 2*(1-f)^2*sdX^2 + 2*f^2*sdX^2$, и при этом f представляет собой фетальную фракцию всех фрагментов cfDNA (следовательно, $1-f$ представляет собой материнскую фракцию cfDNA).

Для хромосомы Z авторы изобретения получают аналогичную случайную переменную $P = (1-f)*Z1 + (1-f)*Z2 + f*Z3 + f*Z4$ при $muP = 2*muZ$, и $sdP^2 = (2*(1-f)^2*sdZ^2 + 2*f^2*sdZ^2)$. Наконец, не существует истинной хромосомы Y.

Таким образом, авторы изобретения получают ожидаемые показатели распределения картирования для хромосом X и Y в случае эуплоидного плода женского пола, которые позднее будут сравниваться с наблюдаемыми данными. Для короткой ссылки, авторы изобретения будут обозначать такие численные значения с помощью

- MxxzzFxxzz.muX = 2*muX
- MxxzzFxxzz.sdX^2 = $2*(1-f)^2*sdX^2 + 2*f^2*sdX^2$
- MxxzzFxxzz.muY = 2*muZ
- MxxzzFxxzz.sdY^2 = $2*(1-f)^2*sdZ^2 + 2*f^2*sdZ^2$

Следует обратить внимание, что такие формулы зависят от фетальной фракции *f*. Другими словами, в действительности авторы изобретения получают множество гипотез MxxzzFxxzz для различных значений *f*. Данная *f* находится в диапазоне от 5% до 100% с шагом 0,1%. Следует обратить внимание, что данное значение *f* не зависит от фетальной фракции тестируемого образца. Это касается любой гипотезы, определенной ниже.

Эуплоидный плод мужского пола - MxxzzFxzy

Для X хромосомы существует три источника фрагментов: две материнские хромосомы X и одна хромосома X плода. Авторы изобретения связывают такие источники с тремя случайными переменными X1, X2 и X3, все из которых имеют одинаковое нормальное распределение N (muX, sdX^2). Следовательно, наблюдаемое совокупное количество фрагментов, картированных на хромосоме X, представляет собой случайную переменную O = (1-f)*X1 + (1-f)*X2 + f*X3, которая опять же нормально распределена как N(muO, sdO^2) при muO = (2 - f)*muX, $sdO^2 = 2*(1-f)^2*sdX^2 + f^2*sdX^2$, и при этом *f* представляет собой фетальную фракцию всех фрагментов cfDNA (следовательно, *1-f* представляет собой материнскую фракцию cfDNA).

Для хромосомы Z авторы изобретения получают аналогичную случайную переменную $P = (1-f)*Z1 + (1-f)*Z2 + f*Z3$ при $muP = 2*muZ$, и $sdP^2 = (2*(1-f)^2*sdZ^2 + f^2*sdZ^2)$.

Наконец, для истинной хромосомы Y авторы изобретения получают $f*Y \sim N(f*muY, f^2*sdY^2)$.

Таким образом, авторы изобретения получают ожидаемые показатели распределения картирования для хромосом X и Y в случае эуплоидного плода мужского пола, которые позднее будут сравниваться с наблюдаемыми данными. Для короткой ссылки, авторы изобретения будут обозначать такие численные значения с помощью

- MxxzzFxzy.muX = (2-f)*muX
- MxxzzFxzy.$sdX^2 = 2*(1-f)^2*sdX^2 + f^2*sdX^2$
- MxxzzFxzy.muY = (2-f)*muZ + f*muY
- MxxzzFxzy.$sdY^2 = 2*(1-f)^2*sdZ^2 + f^2*sdZ^2 + f^2*sdY^2$

В свою очередь, в действительности существует множество гипотез MxxzzFxzy для различных значений *f*.

Аналогично, другие гипотезы будут представлять собой следующее.

Плод женского пола с синдромом Тернера - MxxzzFxz

- MxxzzFxz.muX = (2-f)*muX
- MxxzzFxz.sdX^2 = 2*(1-f)^2*sdX^2 + f^2*sdX^2
- MxxzzFxz.muY = (2-f)*muZ
- MxxzzFxz.sdY^2 = 2*(1-f)^2*sdZ^2 + f^2*sdZ^2

Плод женского пола с синдромом трисомии X - MxxzzFxxxzzz

- MxxzzFxxxzzz.muX = (2+f)*muX
- MxxzzFxxxzzz.sdX^2 = 2*(1-f)^2*sdX^2 + 3*f^2*sdX^2
- MxxzzFxxxzzz.muY = (2+f)*muZ
- MxxzzFxxxzzz.sdY^2 = 2*(1-f)^2*sdZ^2 + 3*f^2*sdZ^2

Плод мужского пола с синдромом Клайнфельтера - MxxzzFxxzzy

- MxxzzFxxzzy.muX = 2*muX
- MxxzzFxxzzy.sdX^2 = 2*(1-f)^2*sdX^2 + 2*f^2*sdX^2
- MxxzzFxxzzy.muY = 2*muZ + f*muY
- MxxzzFxxzzy.sdY^2 = 2*(1-f)^2*sdZ^2 + 2*f^2*sdZ^2 + f^2*sdY^2

Плод мужского пола с синдромом Джейкобс - MxxzzFxzyy

- MxxzzFxzyy.muX = (2-f)*muX
- MxxzzFxzyy.sdX^2 = 2*(1-f)^2*sdX^2 + f^2*sdX^2 + 2*f^2*sdY^2

- MxxzzFxzyy.muY = 2*muZ + 2*f*muY
- MxxzzFxzyy.sdY^2 = 2*(1-f)^2*sdZ^2 + f^2*sdZ^2 + 2*f^2*sdY^2

Формулировка гипотезы - двухплодная беременность

В случае двухплодной беременности ситуация является аналогичной, за исключением того, что теперь имеют место два плода, каждый со своей собственной фракцией фрагментов cfDNA *f1* и *f2*. Код гипотезы теперь будет изменяться на M_F1_F2_ (например, MxxzzF1xxzzF2xxzz для двух эуплоидных плодов женского пола, MxxzzF1xxzzF2xzy для эуплоидных плодов женского и мужского пола, и так далее). Производные гипотезы являются аналогичными для случаев одноплодной беременности, поэтому авторы изобретения переходят непосредственно к эталонным численным значениям.

Два эуплоидных плода женского пола - MxxzzF1xxzzF2xxzz

- MxxzzF1xxzzF2xxzz.muX = 2*muX
- MxxzzF1xxzzF2xxzz.sdX^2 = 2*(1-f1-f2)^2*sdX^2 + 2*f1^2*sdX^2 + 2*f2^2*sdX^2
- MxxzzF1xxzzF2xxzz.muY = 2*muZ
- MxxzzF1xxzzF2xxzz.sdY^2 = 2*(1-f1-f2)^2*sdZ^2 + 2*f1^2*sdZ^2 + 2*f2^2*sdZ^2

Эуплоидный плод женского пола с эуплоидным плодом мужского пола - MxxzzF1xxzzF2xzy

- MxxzzF1xxzzF2xzy.muX = (2-f2)*muX
- MxxzzF1xxzzF2xzy.sdX^2 = 2*(1-f1-f2)^2*sdX^2 + 2*f1^2*sdX^2 + f2^2*sdX^2
- MxxzzF1xxzzF2xzy.muY = (2-f2)*muZ + f2*muY
- MxxzzF1xxzzF2xzy.sdY^2 = 2*(1-f1-f2)^2*sdZ^2 + 2*f1^2*sdZ^2 + f2^2*sdZ^2 + f2^2*sdY^2

Эуплоидный плод женского пола с плодом женского пола с синдромом Тернера - MxxzzF1xxzzF2xz

- MxxzzF1xxzzF2xz.muX = (2-f2)*muX
- MxxzzF1xxzzF2xz.sdX^2 = 2*(1-f1-f2)^2*sdX^2 + 2*f1^2*sdX^2 + f2^2*sdX^2
- MxxzzF1xxzzF2xz.muY = (2-f2)*muZ
- MxxzzF1xxzzF2xz.sdY^2 = 2*(1-f1-f2)^2*sdZ^2 + 2*f1^2*sdZ^2 + f2^2*sdZ^2

Эуплоидный плод женского пола с плодом женского пола с трисомией по хромосоме X - MxxzzF1xxzzF2xxxzzz

- MxxzzF1xxzzF2xxxzzz.muX = (2+f2)*muX
- MxxzzF1xxzzF2xxxzzz.sdX^2 = 2*(1-f1-f2)^2*sdX^2 + 2*f1^2*sdX^2 + 3*f2^2*sdX^2
- MxxzzF1xxzzF2xxxzzz.muY = (2+f2)*muZ
- MxxzzF1xxzzF2xxxzzz.sdY^2 = 2*(1-f1-f2)^2*sdZ^2 + 2*f1^2*sdZ^2 + 3*f2^2*sdZ^2

Эуплоидный плод женского пола с плод мужского пола с синдромом Клайнфельтера - MxxzzF1xxzzF2xxzzy

- MxxzzF1xxzzF2xxzzy.muX = 2*muX
- MxxzzF1xxzzF2xxzzy.sdX^2 = 2*(1-f1-f2)^2*sdX^2 + 2*f1^2*sdX^2 + 2*f2^2*sdX^2
- MxxzzF1xxzzF2xxzzy.muY = 2*muZ + f2*muY
- MxxzzF1xxzzF2xxzzy.sdY^2 = 2*(1-f1-f2)^2*sdZ^2 + 2*f1^2*sdZ^2 + 2*f2^2*sdZ^2 + f2^2*sdY^2

Эуплоидный плод женского пола с плодом мужского пола с синдромом Джейкобс - MxxzzF1xxzzF2xzyy

- MxxzzF1xxzzF2xzyy.muX = (2-f2)*muX
- MxxzzF1xxzzF2xzyy.sdX^2 = 2*(1-f1-f2)^2*sdX^2 + 2*f1^2*sdX^2 + f2^2*sdX^2
- MxxzzF1xxzzF2xzyy.muY = (2-f2)*muZ + 2*f2*muY

- MxxzzF1xxzzF2xzyy.sdY^2 = 2*(1-f1-f2) ^{2}*sdZ^2 + 2*f1^2*sdZ^2 + f2^2*sdZ^2 + 2*f2^2*sdY^2

Эуплоидный плод мужского пола с эуплоидным плодом женского пола - MxxzzF1xzyF2xxzz

Идентично случаю «Эуплоидный плод женского пола с эуплоидным плодом мужского пола» - MxxzzF1xxzzF2xzy

Два эуплоидных плода мужского пола - MxxzzF1xzyF2xzy

- MxxzzF1xzyF2xzy.muX = (2-f1-f2)*muX
- MxxzzF1xzyF2xzy.sdX^2 = 2*(1-f1-f2) ^{2}*sdX^2 + f1^2*sdX^2 + f2^2*sdX^2
- MxxzzF1xzyF2xzy.muY = (2-f1-f2)*muZ + f1*muY + f2*muY
- MxxzzF1xzyF2xzy.sdY^2 = 2*(1-f1-f2) ^{2}*sdZ^2 + f1^2*sdZ^2 + f1^2*sdY^2 + f2^2*sdZ^2 + f2^2*sdY^2

Эуплоидный плод мужского пола с плодом женского пола с синдромом Тернера - MxxzzF1xzyF2xz

- MxxzzF1xzyF2xz.muX = (2-f1-f2)*muX
- MxxzzF1xzyF2xz.sdX^2 = 2*(1-f1-f2) ^{2}*sdX^2 + f1^2*sdX^2 + f2^2*sdX^2
- MxxzzF1xzyF2xz.muY = (2-f1-f2)*muZ + f1*muY
- MxxzzF1xzyF2xz.sdY^2 = 2*(1-f1-f2) ^{2}*sdZ^2 + f1^2*sdZ^2 + f2^2*sdZ^2

Эуплоидный плод мужского пола с плодом женского пола с трисомией по хромосоме X - MxxzzF1xzyF2xxxzzz

- MxxzzF1xzyF2xzy.muX = (2-f1+f2)*muX
- MxxzzF1xzyF2xzy.sdX^2 = 2*(1-f1-f2) ^{2}*sdX^2 + f1^2*sdX^2 + 3*f2^2*sdX^2
- MxxzzF1xzyF2xzy.muY = (2-f1+f2)*muZ + f1*muY + f2*muY

- MxxzzF1xzyF2xzy.sdY^2 = 2*(1-f1-f2) ^{2}*sdZ^2 + f1^2*sdZ^2 + f1^2*sdY^2 + f2^2*sdZ^2 + f2^2*sdY^2

Эуплоидный плод мужского пола с плодом мужского пола с синдромом Клайнфельтера - MxxzzF1xzyF2xxzzy

- MxxzzF1xzyF2xxzzy.muX = (2-f1)*muX
- MxxzzF1xzyF2xxzzy.sdX^2 = 2*(1-f1-f2) ^{2}*sdX^2 + f1^2*sdX^2 + 2*f2^2*sdX^2
- MxxzzF1xzyF2xxzzy.muY = (2-f1)*muZ + f1*muY + f2*muY
- MxxzzF1xzyF2xxzzy.sdY^2 = 2*(1-f1-f2) ^{2}*sdZ^2 + f1^2*sdZ^2 + f1^2*sdY^2 + 2*f2^2*sdZ^2 + f2^2*sdY^2

Эуплоидный плод мужского пола с плодом мужского пола с синдромом Джейкобс - MxxzzF1xzyF2xzyy

- MxxzzF1xzyF2xzyy.muX = (2-f1-f2)*muX
- MxxzzF1xzyF2xzyy.sdX^2 = 2*(1-f1-f2)^2*sdX^2 + f1^2*sdX^2 + f2^2*sdX^2
- MxxzzF1xzyF2xzyy.muY = (2-f1-f2)*muZ + f1*muY + 2*f2*muY
- MxxzzF1xzyF2xzyy.sdY^2 = 2*(1-f1-f2) ^{2}*sdZ^2 + f1^2*sdZ^2 + f1^2*sdY^2 + f2^2*sdZ^2 + 2*f2^2*sdY^2

Плод женского пола с синдромом Тернера с эуплоидным плодом женского пола - MxxzzF1xzF2xxzz

Идентично случаю «Эуплоидный плод женского пола с плодом женского пола с синдромом Тернера» - MxxzzF1xxzzF2xz

Плод женского пола с синдромом Тернера с эуплоидным плодом мужского пола – MxxzzF1xzF2xzy

Идентично случаю «Эуплоидный плод мужского пола с плодом женского пола с синдромом Тернера» - MxxzzF1xzyF2xz

Два плода женского пола с синдром Тернера - MxxzzF1xzF2xz

- MxxzzF1xzF2xz.muX = (2-f1-f2)*muX
- MxxzzF1xzF2xz.sdX^2 = 2*(1-f1-f2)^2*sdX^2 + f1^2*sdX^2 + f2^2*sdX^2
- MxxzzF1xzF2xz.muY = (2-f1-f2)*muZ
- MxxzzF1xzF2xz.sdY^2 = 2*(1-f1-f2)^2*sdZ^2 + f1^2*sdZ^2 + f2^2*sdZ^2

Плод женского пола с синдромом Тернера с плодом женского пола с трисомией по хромосоме X - MxxzzF1xzF2xxxzzz

- MxxzzF1xzF2xxxzzz.muX = (2-f1+f2)*muX
- MxxzzF1xzF2xxxzzz.sdX^2 = 2*(1-f1-f2)^2*sdX^2 + f1^2*sdX^2 + 3*f2^2*sdX^2
- MxxzzF1xzF2xxxzzz.muY = (2-f1+f2)*muZ
- MxxzzF1xzF2xxxzzz.sdY^2 = 2*(1-f1-f2)^2*sdZ^2 + f1^2*sdZ^2 + 3*f2^2*sdZ^2

Плод женского пола с синдромом Тернера с плодом мужского пола с синдромом Клайнфельтера - MxxzzF1xzF2xxzzy

- MxxzzF1xzF2xxzzy.muX = (2-f1)*muX
- MxxzzF1xzF2xxzzy.sdX^2 = 2*(1-f1-f2)^2*sdX^2 + f1^2*sdX^2 + 2*f2^2*sdX^2
- MxxzzF1xzF2xxzzy.muY = (2-f1)*muZ + f2*muY
- MxxzzF1xzF2xxzzy.sdY^2 = 2*(1-f1-f2)^2*sdZ^2 + f1^2*sdZ^2 + 2*f2^2*sdZ^2 + f2^2*sdY^2

Плод женского пола с синдромом Тернера с плодом мужского пола с синдромом Джейкобс - MxxzzF1xzF2xzyy

- MxxzzF1xzF2xzyy.muX = (2-f1-f2)*muX
- MxxzzF1xzF2xzyy.sdX^2 = 2*(1-f1-f2)^2*sdX^2 + f1^2*sdX^2 + f2^2*sdX^2
- MxxzzF1xzF2xzyy.muY = (2-f1-f2)*muZ + 2*f2*muY
- MxxzzF1xzF2xzyy.sdY^2 = 2*(1-f1-f2)^2*sdZ^2 + f1^2*sdZ^2 + f2^2*sdZ^2 + 2*f2^2*sdY^2

Плод женского пола с трисомией по хромосоме X с эуплоидным плодом женского пола - MxxzzF1xxxzzzF2xxzz

Идентично случаю «Эуплоидный плод женского пола с плодом женского пола с трисомией по хромосоме X» - MxxzzF1xxzzF2xxxzzz

Плод женского пола с трисомией по хромосоме X с эуплоидным плодом мужского пола - MxxzzF1xxxzzzF2xzy

Идентично случаю «Эуплоидный плод мужского пола с плодом женского пола с трисомией по хромосоме X» - MxxzzF1xzyF2xxxzzz

Плод женского пола с трисомией по хромосоме X с плодом женского пола с синдромом Тернера - MxxzzF1xxxzzzF2xz

Идентично случаю «Эуплоидный плод мужского пола с плодом женского пола с трисомией по хромосоме X» - MxxzzF1xzF2xxxzzz

Два плода женского пола с синдромом трисомии X - MxxzzF1xxxzzzF2xxxzzz

- MxxzzF1xxxzzzF2xxxzzz.muX = (2+f1+f2)*muX
- MxxzzF1xxxzzzF2xxxzzz.sdX = $2*(1-f1-f2)^2*sdX^2 + 3*f1^2*sdX^2 + 3*f2^2*sdX^2$
- MxxzzF1xxxzzzF2xxxzzz.muY = (2+f1+f2)*muZ
- MxxzzF1xxxzzzF2xxxzzz.sdY = $2*(1-f)^2*sdZ^2 + 3*f1^2*sdZ^2 +3*f2^2*sdZ^2$

Плод женского пола с трисомией по хромосоме X с плодом мужского пола с синдромом Клайнфельтера - MxxzzF1xxxzzzF2xxzzy

- MxxzzF1xxxzzzF2xxzzy.muX = (2+f1)*muX
- MxxzzF1xxxzzzF2xxzzy.sdX^2 = $2*(1-f1-f2)^2*sdX^2 + 3*f1^2*sdX^2 + 2*f2^2*sdX^2$
- MxxzzF1xxxzzzF2xxzzy.muY = (2+f1)*muZ + f2*muY
- MxxzzF1xxxzzzF2xxzzy.sdY^2 = $2*(1-f1-f2)^2*sdZ^2 + 3*f1^2*sdZ^2 + 2*f2^2*sdZ^2 + f2^2*sdY^2$

Плод женского пола с трисомией по хромосоме X с плодом мужского пола с синдромом Джейкобс - MxxzzF1xxxzzzF2xzyy

- MxxzzF1xxxzzzF2xzyy.muX = (2+f1-f2)*muX
- MxxzzF1xxxzzzF2xzyy.sdX^2 = 2*(1-f1-f2)^2*sdX^2 + 3*f1^2*sdX^2 + f2^2*sdX^2
- MxxzzF1xxxzzzF2xzyy.muY = (2+f1-f2)*muZ + 2*f2*muY
- MxxzzF1xxxzzzF2xzyy.sdY^2 = 2*(1-f1-f2)^2*sdZ^2 + 3*f1^2*sdZ^2 + f2^2*sdZ^2 + 2*f2^2*sdY^2

Плод мужского пола с синдромом Клайнфельтера с эуплоидным плодом женского пола - MxxzzF1xxzzyF2xxzz

Идентично случаю «Эуплоидный плод женского пола с плодом мужского пола с синдромом Клайнфельтера» - MxxzzF1xxzzF2xxzzy

Плод мужского пола с синдромом Клайнфельтера с эуплоидным плодом мужского пола - MxxzzF1xxzzyF2xzy

Идентично случаю «Эуплоидный плод мужского пола с плодом мужского пола с синдромом Клайнфельтера» - MxxzzF1xzyF2xxzzy

Плод мужского пола с синдромом Клайнфельтера с плодом женского пола с синдромом Тернера - MxxzzF1xxzzyF2xz

Идентично случаю «Плод женского пола с синдромом Тернера с плодом женского пола с трисомией по хромосоме X» - MxxzzF1xzF2xxzzy

Плод мужского пола с синдромом Клайнфельтера с плодом женского пола с трисомией по хромосоме X - MxxzzF1xxzzyF2xxxzzz

Идентично случаю «Плод женского пола с трисомией по хромосоме X с плодом мужского пола с синдромом Клайнфельтера» - MxxzzF1xxxzzzF2xxzzy

Два плода мужского пола с синдромом Клайнфельтера – MxxzzF1xxzzyF2xxzzy

- MxxzzF1xxzzyF2xxzzy.muX = 2*muX

- MxxzzF1xxzzyF2xxzzy.sdX^2 = 2*(1-f1-f2)^2*sdX^2 + 2*f1^2*sdX^2 + 2*f2^2*sdX^2
- MxxzzF1xxzzyF2xxzzy.muY = 2*muZ + f1*muY + f2*muY
- MxxzzF1xxzzyF2xxzzy.sdY^2 = 2*(1-f1-f2)^2*sdZ^2 + 2*f1^2*sdZ^2 + f1^2*sdY^2 + 2*f2^2*sdZ^2 + f2^2*sdY^2

Плод мужского пола с синдромом Клайнфельтера с плодом мужского пола с синдромом Джейкобс - MxxzzF1xxzzyF2xzyy

- MxxzzF1xxzzyF2xzyy.muX = (2-f2)*muX
- MxxzzF1xxzzyF2xzyy.sdX^2 = 2*(1-f1-f2)^2*sdX^2 + 2*f1^2*sdX^2 + f2^2*sdX^2
- MxxzzF1xxzzyF2xzyy.muY = (2-f2)*muZ + f1*muY+ 2*f2*muY
- MxxzzF1xxzzyF2xzyy.sdY^2 = 2*(1-f1-f2)^2*sdZ^2 + 2*f1^2*sdZ^2 + f1^2*sdY+ f2^2*sdZ^2 + 2*f2^2*sdY^2

Плод мужского пола с синдромом Джейкобс с эуплоидным плодом женского пола - MxxzzF1xzyyF2xxzz

Идентично случаю «Эуплоидный плод женского пола с плодом мужского пола с синдромом Джейкобс» - MxxzzF1xxzzF2xzyy

Плод мужского пола с синдромом Джейкобс с эуплоидным плодом мужского пола - MxxzzF1xzyyF2xzy

Идентично случаю «Эуплоидный плод мужского пола с плодом мужского пола с синдромом Джейкобс» - MxxzzF1xzyF2xzyy

Плод мужского пола с синдромом Джейкобс с плодом женского пола с синдромом Тернера - MxxzzF1xzyyF2xz

Идентично случаю «Плод женского пола с синдромом Тернера с плодом мужского пола с синдромом Джейкобс» - MxxzzF1xzF2xzyy

Плод мужского пола с синдромом Джейкобс с плодом женского пола с трисомией по хромосоме X - MxxzzF1xzyyF2xxxzzz

Идентично случаю «Плод женского пола с трисомией по хромосоме X с плодом мужского пола с синдромом Джейкобс» - MxxzzF1xxxzzzF2xzyy

Плод мужского пола с синдромом Джейкобс с плодом мужского пола с синдромом Клайнфельтера - MxxzzF1xzyyF2xxzzy

Идентично случаю «Плод мужского пола с синдромом Клайнфельтера с плодом мужского пола с синдромом Джейкобс» - MxxzzF1xxzzyF2xzyy

Два плода мужского пола с синдромом Джейкобс - MxxzzF1xzyyF2xzyy

- MxxzzF1xzyyF2xzyy.muX = (2-f1-f2)*muX
- MxxzzF1xzyyF2xzyy.sdX == $2*(1-f1-f2)^2*sdX^2 + f1^2*sdX^2 + 2*f1^2*sdY^2 + f2^2*sdX^2 + 2*f2^2*sdY^2$
- MxxzzF1xzyyF2xzyy.muY = 2*muZ + 2*f1*muY + 2*f1*muY
- MxxzzF1xzyyF2xzyy.sdY = $2*(1-f1-f2)^2*sdZ^2 + f1^2*sdZ^2 + 2*f1^2*sdY^2 + f2^2*sdZ^2 + 2*f2^2*sdY^2$

Формулировка гипотезы - случаи беременности с наличием более чем двух плодов

Не существует теоретической причины останавливаться на гипотезах для двухплодных беременностей – авторы изобретения могут с легкостью расширять теорию до, скажем, трехплодной беременности с фетальными фракциями f1, f2 и f3. Однако большая глубина секвенирования, необходимая для таких случаев, не доступна по разумной цене (относительно масштабного применения) при текущих технологиях секвенирования, хотя теоретически (или в будущем) это легко возможно.

Ограничения

Определенные комбинации SCA при случаях двухплодной беременности могут компенсировать показатели дефицита или избытка друг друга. Например, один плод женского пола с синдромом Тернера с плодом женского пола с трисомией по хромосоме X компенсирует аберрации по хромосоме X таким образом, что образец будет выглядеть как два эуплоидных плода женского пола, если фетальные фракции обоих плодов эквивалентны или близки. В следующей таблице перечислены все

неопределенные случаи, где фетальная фракция является эквивалентной для обоих плодов при двухплодной беременности. Симметричная часть таблицы опущена (пустые ячейки). Прочерк (-) поставлен там, где неопределенность отсутствует.

	XX	XY	X0	XXX	XXY	XYY
XX	X0 + XXX	X0 + XXY	-	-	XXX + XY	XXY + XY
XY		X0 + XYY	-	XXY + XX	XYY + XX	-
X0			-	XX + XX	XX + XY	XY + XY
XXX				-	-	XXY + XXY
XXY					XXX + XYY	-
XYY						-

Таким образом, до тех пор, пока не будет существовать способа определения отдельных фетальных фракций *f1* и *f2* в случае двухплодных беременностей, или данный способ не будет указывать, что *f1* эквивалентна или близка к эквивалентной с *f2*, данные NGS в целом не могут быть интерпретированы однозначно. Однако даже в данном неопределенном состоянии тест обеспечивает пользователя ценной информацией, поскольку всегда существуют только две возможные интерпретации. Таким образом, пользователь может предположить дополнительные тесты, конкретно нацеленные на одну из возможных интерпретаций с целью доказать или опровергнуть ее. Кроме того, распространенность SCA в популяции также обеспечивает помощь при интерпретации. Например, намного более вероятно иметь эуплоидную двойню женского пола, чем один плод с синдромом Тернера и другой плод с синдромом трисомии X.

Распространенность SCA в популяции

Распространенность SCA в популяции было описано в *Snijders et al. 1995*[27], *Zimmermann et al. 2012*, *Lau et al. 2012*, *Mazloom et al. 2013*, *Norton et al. 2014*[28], *Hook et al. 2014*[29], *Grati et al. 2014*[30], *Stumm et al. 2014*[31]. В предыдущем уровне техники было показано, что распространенность SCA зависит как от возраста матери, так и от недели гестации во время проведения теста.

Оператор может свободно выбрать любую модель распространенности, какую пожелает. В примерах ниже применяли модель распространенности, описанную в *Snijders et al. 1995*. Более конкретно авторы изобретения применяли таблицу 6 в отношении распространенности XXX, XXY и XYY, и они применяли таблицу 7 в

отношении распространенности X0. Дополнительно авторы изобретения предполагали, что распространенность плодов мужского и женского пола является эквивалентным.

Что касается двухплодных беременностей, авторы изобретения считали, что оба плода являются независимыми. По данной причине вероятности распространения умножаются. Например, вероятность наличия одного эуплоидного плода женского пола (XX) и одного плода женского пола с трисомией по хромосоме X (XXX) равняется распространенности XX, умноженной на распространенность XXX.

Анализ гипотез методом максимального правдоподобия

Входные данные для анализа методом максимального правдоподобия представляют собой два числа, x и y, которые указывают на соотношение фрагментов ДНК, картированных на хромосоме X и Y соответственно, для данного образца. Помимо этого вторая часть входных данных представляет собой априорное распределение фетальной фракции с учетом функции плотности $N(muF, sgF^2)$ или распределения двух фетальных фракций с учетом их функций плотности $N(muF1, sgF1^2)$ и $N(muF2, sgF2^2)$ в случае двухплодной беременности. Авторы изобретения снова обращают внимание, что с учетом текущего статуса технологий они вынуждены предполагать, что $N(muF1, sgF1^2)=N(muF2, sgF2^2)$.

В случае одноплодной беременности на данной стадии для каждой гипотезы *h* (MxxzzFxxzz, MxxzzFxzy, MxxzzFxz, MxxzzFxxxzzz, MxxzzFxxzzy, MxxzzFxzyy) и каждого значения фетальной фракции *f* (дискретизированные значения, находящиеся в диапазоне от 5% до 100% с шагом 0,1%) рассчитывают вероятность наблюдения данных x и y при *h* и *f*, т. е. $Pr[x|h,f]$ и $Pr[y|h,f]$. Вероятность $Pr[x|h, f]$ (соотв. $Pr[y|h, f]$) рассчитывают как определенный интеграл от функции плотности вероятности $N(h.muX, h.sdX^2)$ (соотв. $N(h.muY, h.sdY^2)$) в интервале $[x - 1e\text{-}6, x + 1e\text{-}6]$ (соотв. $[y - 1e\text{-}6, y + 1e\text{-}6]$), где $h.muX$ (соотв. $h.muY$) и $h.sdX^2$ (соотв. $h.sdY^2$) указаны гипотезой *h*. Кроме того, вероятность конкретного значения фетальной фракции *f* рассчитывают как определенный интеграл функции плотности вероятности $N(h.muF, h.sdF^2)$ в интервале $[f - 1e\text{-}3, f + 1e\text{-}3]$. Конечная вероятность гипотезы h_2 с учетом данных NGS, распределений фетальной фракции и распространенности *h* в популяции равняется $Pr[h,f|x,y] = Pr[x|h,f]* Pr[y|h,f]* Pr[f]*Pr[h]$. Данное значение рассчитывают и наносят на график для каждой гипотезы *h* и фетальной фракции *f*.

В случае двухплодной беременности на данной стадии для каждой гипотезы *h* (MxxzzF1xxzzF2xxzz, MxxzzF1xxzzF2xzy, MxxzzF1xxzzF2xz, MxxzzF1xxzzF2xxxzzz, MxxzzF1xxzzF2xxzzy, MxxzzF1xxzzF2xzyy, MxxzzF1xzyF2xzy, MxxzzF1xzyF2xz, MxxzzF1xzyF2xxxzzz, MxxzzF1xzyF2xxzzy, MxxzzF1xzyF2xzyy, MxxzzF1xzF2xz, MxxzzF1xzF2xxxzzz, MxxzzF1xzF2xxzzy, MxxzzF1xzF2xzyy, MxxzzF1xxxzzzF2xxxzzz, MxxzzF1xxxzzzF2xxzzy, MxxzzF1xxxzzzF2xzyy, MxxzzF1xxzzyF2xxzzy, MxxzzF1xxzzyF2xzyy, MxxzzF1xzyyF2xzyy, в целом 21 гипотеза) и для каждой фетальной фракции *f1, f2* (две фетальные фракции *f1* и *f2* находятся в диапазоне от 5% до 100% с шагом 0,1%, но для которых авторы изобретения дополнительно получили *f1*+*f2*<=100%) рассчитывают вероятность наблюдения данных x и y при *h* и *f1, f2,* т. е. *Pr[x|h, f1, f2]* и *Pr[y|h, f1, f2]*. Вероятность *Pr[x|h, f1, f2]* (соотв. *Pr[y|h, f1, f2]*) рассчитывают как определенный интеграл функции плотности вероятности $N(h.muX, h.sdX^2)$ (соотв. $N(h.muY, h.sdY^2)$) в интервале *[x - 1e-6, x + 1e-6]* (соотв. *[y - 1e-6, y + 1e-6]*), где *h.muX* (соотв. *h.muY*) и $h.sdX^2$ (соотв. $h.sdY^2$) указаны гипотезой *h.* Кроме того, вероятность конкретного значения фетальной фракции *f1(*соотв. *f2)* рассчитывают как определенный интеграл функции плотности вероятности $N(h.muF1, h.sdF1^2)$ (соотв. $N(h.muF2, h.sdF2^2)$) в интервале *[f - 1e-3, f + 1e-3].* Конечная вероятность гипотезы *h* и фетальных фракций *f1* и *f2*, с учетом данных NGS, распределений фетальной фракции и распространенности *h* в популяции равняется*Pr[h, f1, f2|x, y] = Pr[x|h, f1, f2]* Pr[y|h, f1, f2]* Pr[f1]*Pr[f2]*Pr[h]*, где *Pr[h]* представляет собой распространенность гипотезы *h* в популяции. Данное значение рассчитывают и наносят на график для каждой гипотезы *h* и фетальной фракции *f1* и *f2*.

Улучшения в обработке исходных данных

Это относится к применению поправки на сдвиг содержания GC в отношении картированных данных NGS. Обычно выбирают модель поправки на сдвиг содержания GC и применяют ко всем хромосомам. Однако авторы изобретения наблюдали, что для анализа SCA лучше применять поправку на сдвиг содержания GC в отношении хромосомы X и оставлять хромосому Y без какой-либо коррекции.

Чтобы объяснить, данная стадия обеспечивала заметное улучшение корреляции между показателями картирования хромосом X и Y для образцов с эуплоидным плодом мужского пола. Следует обратить внимание, что авторы изобретения ожидают обнаружить высокий уровень корреляции вследствие различной организации хромосом

X и Y эуплоидного плода мужского пола по сравнению с организацией у его матери, и масштабирование изменения прямо пропорционально фетальной фракции. Действительно, такая корреляция существует, и она была в значительной степени улучшена благодаря отсутствию поправки на сдвиг содержания GC хромосомы Y. Подобное преимущество в корреляции наблюдалось, когда авторы изобретения применяли поправку на сдвиг содержания GC в отношении хромосомы Y и не применяли в отношении хромосомы X. С другой стороны, применение поправки на сдвиг содержания GC в отношении обоих или ни к одной из хромосом приводило к сравнимым, но меньшим показателям корреляции. Подробности приведены в таблице ниже (численные значения жирным шрифтом соответствуют применению смешанной поправки на сдвиг содержания GC, численные значения с подчеркиванием соответствуют общей поправке на сдвиг содержания GC).

	non.X	non.Y	GC.X	GC.Y
non.X	1	-0,73336	0,788391	**-0,93994**
non.Y	-0,73336	1	**-0,92773**	0,818423
GC.X	0,788391	**-0,92773**	1	-0,75308
GC.Y	**-0,93994**	0,818423	-0,75308	1

Следует обратить внимание, что авторы изобретения применяли поправку на сдвиг содержания GC в соответствии с Liao et al. 2014.

Поскольку SCA также предусматривают различные организации хромосом X и Y по сравнению с таковыми у матери, авторы изобретения обосновали, что применение поправки на сдвиг содержания GC только в отношении хромосомы X будет улучшать результаты. Однако поскольку отсутствует достаточно случаев, скажем, плодов с синдромом Тернера, авторы изобретения не могут установить улучшение в корреляции с помощью эксперимента.

Применение способа по настоящему изобретению в отношении распространенных аутосомных анеуплоидий

Данная часть ограничена только случаями одноплодной беременности. Указанный способ можно с легкостью отрабатывать и применять в отношении распространенных аутосомных анеуплоидий, таких как (без ограничения) трисомия или моносомия по

хромосомам 13, 18 или 21. Такой случай непосредственно аналогичен случаю, когда у плода женского пола имеется синдром Тернера (моносомия) или синдром трисомии (трисомия) X. Это связано с тем, что у здорового плода женского пола имеется две X хромосомы, что в данной аналогии соответствует двум хромосомам 21 (или любой другой аутосоме). Кроме того, моносомия по хромосоме 21 (или любой другой аутосоме) аналогична синдрому Тернера, и трисомия по хромосоме 21 (или любой другой хромосоме) аналогична синдрому трисомии X. Таким образом, гипотезы для плода с синдромом Тернера и синдромом трисомии X могут быть с легкостью скорректированы, чтобы соответствовать аутосомной моносомии или трисомии. Единственное отличие будет заключаться в числе рассматриваемых гипотез: 1) моносомия (также называемая синдромом Тернера), 2) трисомия (также называемая синдромом трисомии X) и 3) эуплоидия (также называемая эуплоидным плодом женского пола). Остальные детали легки для осуществления любым специалистом в данной области техники (распределение фетальной фракции такое же, а также уравнения для расчета конечных вероятностей).

Компьютеризация способа

В настоящее время существует практическая необходимость в автоматизации способов пренатальных исследований или диагностики. Способ согласно настоящему изобретению можно в значительной степени автоматизировать. По меньшей мере биоинформатическую часть способа (т. е. обработку данных секвенирования и все последующие определения и расчеты) можно осуществлять с применением подходящей компьютерной системы, такой как ПК, оснащенный процессором, периферическими устройствами ввода/вывода (например, портами, интерфейсами), запоминающими устройствами (например, системной памятью, жестким диском), клавиатурой, монитором, мышью и т. д. и конкретным программным обеспечением, программой для сообщения компьютерной системе инструкций по осуществлению конкретного этапа. Компьютерная система предпочтительно имеет возможность обмена данными с системой секвенирования, передающей данные о последовательностях, предпочтительно в форме множества ридов последовательностей (с помощью проводной или беспроводной сети, технологии Bluetooth, Интернета, системы облачных вычислений и т. д.). Это означает, что компьютерная система выполнена с возможностью получать данные о последовательностях от системы

секвенирования. Подходящие компьютерные системы, а также средства для соединения с системой секвенирования хорошо известны специалистам в данной области.

По меньшей мере часть способа, в частности, биоинформатическая часть способа, может быть реализована в виде программного кода, т. е. множества инструкций (компьютерной программы), предназначенных для выполнения процессором компьютерной системы. Код может содержаться на машиночитаемом носителе для хранения или передачи, таком как, например, RAM, ROM, жесткий диск, SDS, CD, DVD, флэш-память и т. д. Кроме того, код можно передавать посредством любой подходящей проводной, оптической или беспроводной сети, например, с помощью Интернета. Например, оператор (пользователь) может загрузить целую компьютерную программу через Интернет.

Следовательно, другой аспект настоящего изобретения относится к выполняемому с помощью компьютера способу, включающему все этапы расчетов и определений, необходимые для обработки входных данных (данных секвенирования), с получением выходного(выходных) параметра(параметров), характеризующих диагностируемое состояние (фетальная фракция и наиболее вероятные гипотезы, из которых может быть логически выведены наличие или отсутствие анеуплоидии, а также пол).

Еще один аспект настоящего изобретения относится к компьютерному программному продукту, предусматривающему машиночитаемый носитель, содержащий множество инструкций по управлению компьютерной системой для осуществления по меньшей мере части способа в соответствии с настоящим изобретением, предпочтительно его части, начинающейся с этапа получения информации о последовательностях после этапа случайного секвенирования, осуществляемого с помощью автоматической системы секвенирования.

Интерпретация результатов

Интерпретация результатов подробно рассмотрена в примерах ниже. В общем результаты вычисления способа по настоящему изобретению могут быть преобразованы в графики, на которых изображена вероятность гипотез относительно половых хромосом плода и то, как данная вероятность изменяется с фетальной фракцией (см. примеры 3-11 и фигуры 3-11).

Следует обратить внимание, что может существовать разногласие между фетальной фракцией из распределения входных данных фетальной фракции и фетальной фракции из гипотез. Это связано с тем, что обе из них рассчитываются на основе независимых данных. В то время как фетальную фракцию для распределения входных данных фетальной фракции рассчитывают исходя из распределения длин фрагментов cfDNA, фетальную фракцию для гипотез рассчитывают исходя из данных NGS для хромосом X и Y.

Интерпретация результатов - одноплодная беременность

Существует три графика для каждого образца, организованных в три строки (см. примеры 3-8 и фигуры 3-8). На графике в нижнем ряду показано, как вероятность каждой гипотезы изменяется со значением фетальной фракции. Значение фетальной фракции, при котором каждая гипотеза достигает максимальной вероятности, приведено в обозначениях. На графике в среднем ряду показана максимальная вероятность среди всех рассматриваемых значений фетальной фракции для каждой гипотезы. Значение фетальной фракции, при котором достигается данный максимум, показано на оси x в круглых скобках (вероятности для значений фетальных фракций, составляющие ниже 5%, не включены). Данный график служит для сравнения максимальных вероятностей всех гипотез. Наконец, в графике в верхнем ряду добавлена распространенность каждой гипотезы в популяции, как указано *Snijders et al. 1995.*

Оператор должен принять решение, основываясь на верхнем графике, так как в нем содержится вся доступная информация (данные NGS секвенирования для хромосом X и Y, распределение фетальной фракции и распространенность SCA в популяции). Однако на других графиках также может быть предоставлена некоторая подтверждающая информация, особенно в случаях наличия мозаицизма, когда распределение фетальной фракции может очень отличаться от наиболее вероятной фетальной фракции из гипотез SCA.

Анализ типичных образцов в случае эуплоидного плода мужского или женского пола продемонстрирован в примерах 3 и 4, и образцов в случае анеуплоидного плода продемонстрирован в примерах 5-8.

Интерпретация результатов - двуплодная беременность

В случае двуплодной беременности авторы изобретения демонстрируют два графика для настоящих образцов (см. примеры 9-10 и фигуры 9-10) благодаря содержанию данных, более точно, только верхние два графика получены с применением тестов для одноплодной беременности (вероятность каждого случая приводится или сверху, или снизу каждого прямоугольника). Как было указано выше в разделе *Ограничения*, интерпретация результатов для двуплодной беременности является более сложной, поскольку те же данные NGS в отношении хромосом X и Y могут приводить к неопределенным результатам (для получения больших подробностей см. таблицу в разделе *Ограничения*).

В соответствии с таблицей в разделе *Ограничения,* обе гипотезы из «эуплоидный плод мужского пола - эуплоидный плод мужского пола» и «плод женского пола с синдромом Тернера - плод мужского пола с синдромом Джейкобс» обеспечивают почти идентичные данные NGS для хромосом X и Y. Таким образом, обе гипотезы имеют высокую вероятность. С другой стороны, вероятность последнего случая незначительна, как показано на верхнем графике, где принята во внимание распространенность SCA в популяции. Вывод состоит в том, что нижний график предоставляет пользователю информацию о конкретных измеренных данных (распределение фетальной фракции плюс данные NGS для хромосом X и Y), в то время как на верхнем графике для него предоставлена информация о реальных ожиданиях. На основе этого пользователь может предлагать последующий тест для определения различий между двумя конкретными возможностями.

Кроме того, один искусственный образец продемонстрирован в примере 11 (см. также фигуру 11).

Объект настоящего изобретения представляет собой способ определения анеуплоидии половых хромосом, пола и фетальной фракции одного или множества плодов с использованием тестируемого образца материнской крови, плазмы крови или сыворотки крови, как определено в прилагаемой формуле изобретения.

Еще один объект настоящего изобретения представляет собой выполняемый с помощью компьютера способ, включающий стадии вышеописанного способа, которые следуют после стадии осуществления секвенирования и получения информации о последовательности, как определено в прилагаемой формуле изобретения.

Еще один объект настоящего изобретения представляет собой компьютерный программный продукт, предусматривающий машиночитаемый носитель, содержащий множество инструкций по управлению компьютерной системой для осуществления указанного выполняемого с помощью компьютера способа, как определено в прилагаемой формуле изобретения.

Признаки и преимущества настоящего изобретения станут понятны специалисту в данной области техники из нижеследующего подробного описания, примеров и фигур.

КРАТКОЕ ОПИСАНИЕ ГРАФИЧЕСКИХ МАТЕРИАЛОВ

На фиг. 1 показана блок-схема, на которой продемонстрированы основные стадии способа в соответствии с настоящим изобретением. Добавлены ссылки на соответствующие части описания.

На фиг. 2 продемонстрирована нейронная сеть для прогнозирования фетальной фракции с использованием гистограммы длины фрагментов образца. Фетальную фракцию, измеренную исходя из содержания хромосомы Y, применяли как направляющие значения для обучения весов w100, w101 ,..., w200, а также функции преобразования $f(x) = ax + b$.

На фиг. 3 показаны результаты анализа образца при одноплодной беременности в случае эуплоидного плода женского пола.

На фиг. 4 показаны результаты анализа образца при одноплодной беременности в случае эуплоидного плода мужского пола.

На фиг. 5 показаны результаты анализа образца при одноплодной беременности в случае плода женского пола с синдромом Тернера.

На фиг. 6 показаны результаты анализа образца при одноплодной беременности в случае плода мужского пола с синдромом Клайнфельтера.

На фиг. 7 показаны результаты анализа образца при одноплодной беременности в случае плода женского пола с синдромом трисомии X.

На фиг. 8 показаны результаты анализа образца при одноплодной беременности в случае плода мужского пола с синдромом Джейкобс.

На фиг. 9 показаны результаты анализа образца при двухплодной беременности в случае либо двух эуплоидных плодов мужского пола, либо одного плода женского пола с синдромом Тернера и одного плода мужского пола с синдромом Джейкобс.

На фиг. 10 показаны результаты анализа образца при двухплодной беременности в случае либо двух эуплоидных плодов женского пола, либо одного плода женского пола с синдромом Тернера и одного плода женского пола с синдромом трисомии X.

На фиг. 11 показаны результаты анализа образца при двухплодной беременности, при которой у одного из двух плодов женского пола имеется синдром Тернера.

ПРИМЕРЫ

Пример 1.

Нейронная сеть для расчета фетальной фракции

Нейронная сеть авторов изобретения, показанная на **фиг. 2**, состоит из двух слоев: базового слоя для 101 входного узла и выходного слоя, состоящего из одного узла (может быть разработана более сложная сеть со скрытыми слоями в случае достаточно большого обучающего набора). Более того, каждый входной узел соединен с выходным узлом. Данные секвенирования образцов, а именно картированные фрагменты cfDNA из всех хромосом, классифицируют в соответствии с их значениями длины, что приводит к получению гистограммы данных (значения длины фрагментов ограничены диапазоном от 100 п. о. до 200 п .о., все другие фрагменты отбракованы). Входные данные нейронной сети тогда представляют собой относительные количества рассмотренных значений длины для каждого образца. Набор образцов с различной и известной фетальной фракцией (например, полученной на основе хромосомы Y в случае образцов с плодом мужского пола) применяют для обучения нейронной сети обычным способом[32]. Фетальные фракции, измеренные исходя из содержания хромосомы Y, применяли в качестве направляющих значений для обучения весов w100, w101 ,..., w200, а также функции преобразования f(x) = ax + b.

Как только модель обучена, авторы изобретения могут прогнозировать основанные на длине фетальные фракции для обучающих образцов, сравнивая их с направляющими, основанными на хромосоме Y, фетальными фракциями, и определить ошибку прогнозирования. Если авторы изобретения обозначат модель прогнозирования как f_{len}

= f_Y + *эпсилон*, где *эпсилон* представляет собой случайную ошибку, имеющую нормальное распределение, то

$$\sqrt{Var[\text{эпсилон}]} = \sqrt{\frac{\sum_{i=1}^{N}\left(f_{lsn(i)} - f_{Y(i)}\right)^2}{N}}$$

представляет собой стандартное отклонение ошибки прогнозирования (N равняется количеству образцов в обучающем наборе). Следует обратить внимание, что среднее значение *эпсилон*, т. е. среднее значение ошибки прогнозирования, равняется нулю, поскольку иначе нейронная сеть не является достаточно обученной.

Пример 2.

Методики получения и секвенирования образцов, применяемые при реализации настоящего изобретения

Анализ последовательностей

Массовое параллельное секвенирование необходимо для применения способа в соответствии с настоящим изобретением. Способ был специально разработан и валидирован для небольших настольных систем секвенирования нового поколения для обеспечения возможности снижения первоначальных затрат на создание сервисной лаборатории NIPT. Способ был валидирован на системе NextSeq500 system (Illumina, Inc., Сан-Диего, Калифорния, США).

Коммерчески доступные устройства для секвенирования совместно с соответствующими протоколами и реагентами, рекомендуемыми поставщиком, использовали в качестве иллюстративного примера, однако специалисту в данной области известен целый ряд различных способов секвенирования и их видоизменений, которые также можно применять при практической реализации настоящего изобретения.

Использовали набор для высокопроизводительного секвенирования NextSeq 500/550 v2 от Illumina (75 циклов). Секвенирование спаренных концов необходимо в этой реализации для обеспечения возможности интерпретации результатов, связанных с определением размеров фрагментов, которое является частью способа. Использовали устанавленные параметры ридов 2 x 35 п. о. В общей сложности использовали 111

образцов со средним количеством ридов 5979689, стандартным отклонением 3242203, минимальным количеством 1184843, максимальным количеством 17191421 для оценки с помощью способа.

Полная процедура получения образцов для секвенатора NextSeq500 (Illumina Inc., Сан-Диего, Калифорния, США)) и анализа

a) Способ получения данных о последовательностях описан ниже с ссылкой на конкретные наборы, для которых применяют следующие стандартные лабораторные протоколы, описаны только конкретные модификации для стандартных протоколов. Все процедуры в пределах стандартных протоколов по сути известны специалистам в области молекулярной биологии, биоинформатики и пренатального тестирования, и такому специалисту известны возможные модификации процедуры. Рекомендуется, чтобы отбор образца, обработку крови, выделение ДНК и получение библиотеки NGS проводил лаборант женского пола ввиду чувствительности способа к контаминации.

Отбор образца крови и обработка, представляющая собой плазмасепарацию

b) Следует отобрать 10 мл образца периферической крови у беременных женщин после 11-й недели беременности в пробирки общего назначения, содержащие EDTA, или пробирки с возможностью стабилизации циркулирующих внеклеточных нуклеиновых кислот (например, Streck Cell-Free DNA BCT).

c) Двухстадийную плазмасепарацию следует проводить не позднее, чем через 2 дня после сбора с мерами предосторожности для предотвращения переноса лейкоцитов.

d) Рекомендуется обрабатывать образцы плазмы крови непосредственно после разделения, но возможно хранение при -20 или -80C.

Выделение ДНК

a) Для осуществления выделения ДНК с достаточным качеством и выходом следует применять набор для выделения внеклеточной ДНК MagMax^{TM} от Life Technologies.

b) Рекомендуемый объем обработанной плазмы крови составляет по меньшей мере 1 мл.

c) Следует применять стандартный протокол производителя.

d) Для плазмы крови, выделенной из пробирок STRECK, необходимо применение обработки протеиназой K в ходе обработки образца.

Получение библиотеки для секвенирования нового поколения

a) При получении библиотеки следует обработать 30 мкл выделенной ДНК из плазмы материнской крови с применением набора для выделения внеклеточной ДНК MagMax.

b) Библиотеки для секвенирования следует получать с применением набора для получения бииблиотек Illlumina TruSeq Nano.

c) Следует применять стандартный протокол производителя с описанными в данном документе исключениями.

d) Для плазмы крови, выделенной из пробирок STRECK, необходимо применение обработки протеиназой K в ходе обработки образца.

a. РЕПАРАЦИЯ КОНЦОВ

i. В соответствии со стандартным протоколом.

b. Дополнительная стадия ОТБОР ПО РАЗМЕРУ (следует проводить дополнительную стадию отбора по размеру для повторного представления фракции выделенных нуклеиновых кислот по размеру, как описано в данном документе)

i. Магнитные гранулы перемешивают вихревым способом в резервуаре в течение 1 минуты. Добавляют 2x объема магнитных гранул (авторы настоящего изобретения рекомендуют применение гранул Agencourt AMPure от Beckman Coulter, тестирование было валидировано с использованием этого набора). Перемешивают надлежащим образом и осаждают путем центрифугирования в течение непродолжительного времени.

ii. Инкубируют в течение 5 минут при комнатной температуре.

iii. Планшет помещают в магнитный штатив на 5 мин. или до тех пор, пока жидкость не станет прозрачной.

iv. Надосадочную жидкость удаляют и сливают.

v. Первое промывание. Добавляют 200 мкл свежеполученного 80% EtOH и инкубируют в течение 30 с.

vi. Надосадочную жидкость удаляют и сливают.

vii. Второе промывание. Добавляют 200 мкл свежеполученного 80% EtOH и инкубируют в течение 30 с.

viii. Надосадочную жидкость удаляют и сливают.

ix. Пробирки оставляют открытыми в течение 10 минут с обеспечением испарения остаточного этанола.

x. Добавляют 10 мкл буфера для ресуспендирования (RSB), тщательно перемешивают вихревым способом.

xi. Инкубируют в течение 2 мин. при комнатной температуре.

xii. Пробирки помещают в магнитный штатив на 5 мин. или до тех пор, пока жидкость не станет прозрачной.

xiii. Переносят 8,75 мкл прозрачной надосадочной жидкости в новый планшет для ПЦР.

xiv. ТОЧКА ОСТАНОВКИ – образец можно хранить в течение не более 1 недели при -20 ^{0}C.

c. ПРИСОЕДИНЕНИЕ А-ХВОСТОВ

i. В соответствии со стандартным протоколом.

d. ЛИГИРОВАНИЕ АДАПТЕРОВ

i. В соответствии со стандартным протоколом.

e. ПЦР-АМПЛИФИКАЦИЯ

i. Данная стадия полностью исключена из процедуры по сравнению со стандартным протоколом.

ii. Поскольку ПЦР служит источником погрешностей при представлении хромосомного фрагмента, ее исключение представляет собой ключевую стадию в уменьшении уровня погрешностей и повышения уверенности при определении фетальной фракции на основе распределения длины фрагментов ридов.

f. КОЛИЧЕСТВЕННАЯ ОЦЕНКА БИБЛИОТЕКИ

i. Измеряют концентрацию конечной библиотеки ДНК (нг/мкл) с помощью набора для флуориметрического HS-анализа dsDNA Qubit.

ii. Определяют средний размер конечной библиотеки (п. о.) на биоанализаторе Agilent Technologies 2100 с помощью высокочувствительного ДНК-чипа.

iii. Конечные библиотеки нормализуют до концентрации конечной библиотеки 0,5-4 нМ, как описано в протоколе производителя, и объединяют библиотеки в конечную библиотеку.

iv. Получают конечную библиотеку для прогона Illumina NextSeq 500, как описано в протоколе производителя системы секвенирования.

g. АНАЛИЗ ПУТЕМ СЕКВЕНИРОВАНИЯ НА ILLUMINA NEXSEQ500

i. Следует использовать систему массового параллельного секвенирования Illumina NexSeq500.

ii. Для прогона анализа путем секвенирования используют набор для высокопроизводительного секвенирования NextSeq 500/550 v2 от Illumina (75 циклов).

iii. Секвенирование спаренных концов осуществляют с устанавленными параметрами ридов 2 x 35 п. о. со считыванием двух видов индексов.

iv. Только файлы FASTQ необходимы для обработки на последующих этапах интерпретации.

Пример 3.

Анализ биологического образца - одноплодная беременность с эуплоидным плодом женского пола

Отбирали и обрабатывали 10 мл образца периферической крови от матери, как описано в примере 2. Данные секвенирования и впоследствии картирования подвергали процедуре анализов методом максимального правдоподобия в отношении конкретной гипотезы (что подробно описано ранее в данном описании и схематически изображено на **фиг. 1**). Обучающие данные, состав гипотезы и обучение нейронной сети для расчета фетальной фракции проводили один раз перед анализом тестируемых образцов, и эти результаты применяли во всех следующих примерах. Рассчитывали распределение фетальной фракции и вероятность каждой гипотезы и результаты изображали в графической форме. В то же время статус плода подтверждали другим независимым способом (амниоцентез с последующим тестом на анеуплоидию с помощью QF-PCR).

В данном случае на **фиг. 3** можно заметить, что «женская» гипотеза (т. е. MxxzzFxxzz) преобладает над всеми другими гипотезами. Следует рассмотреть серый прямоугольник на нижнем графике, в котором указан диапазон значений фетальной фракции, для которых тест не считается надежным.

Пример 4.

Анализ биологического образца - одноплодная беременность с эуплоидным плодом мужского пола

В данном случае на **фиг. 4** можно заметить, что «мужская» гипотеза (т. е. MxxzzFxzy) преобладает над всеми другими гипотезами.

Пример 5.

Анализ биологического образца - ***одноплодная беременность с плодом женского пола с синдромом Тернера (X0)***

В данном случае на **фиг. 5** можно заметить, что гипотеза «моносомия по X» (т. е. MxxzzFxz) преобладает над всеми другими гипотезами.

Пример 6.

Анализ биологического образца - одноплодная беременность с плодом мужского пола с синдромом Клайнфельтера

В данном случае на **фиг. 6** можно заметить, что гипотеза «Клайнфельтер» (т. е. MxxzzFxxzzy) преобладает над всеми другими гипотезами.

Пример 7.

Анализ биологического образца – одноплодная беременность с плодом женского пола с трисомией по хромосоме X (XXX)

В данном случае на **фиг. 7** можно заметить, что гипотеза «трисомия по X» (т. е. MxxzzFxxxzzz) преобладает над всеми другими гипотезами.

Пример 8.

Анализ биологического образца - одноплодная беременность с плодом мужского пола с синдромом Джейкобс (XYY)

В данном случае на **фиг. 8** можно заметить, что гипотеза «Джейкобс» (т. е. MxxzzFxzyy) преобладает над всеми другими гипотезами.

Пример 9.

Анализ биологического образца - двухплодная беременность либо с двумя эуплоидными плодами мужского пола (XY), либо одним плодом женского пола с синдромом Тернера (X0) и одним плодом мужского пола (XYY) с синдромом Джейкобс

Анализировали образец от матери с двумя плодами мужского пола. Данные NGS в двух случаях, указанных в названии выше, почти идентичны, см. **фиг. 9**, следовательно они характеризуются сравнимой вероятностью на нижнем графике. С другой стороны, принимая во внимание распространенность случаев в популяции, случаи эуплоидии намного более вероятны, чем случаи синдромов Тернера/Джейкобс (данная вероятность легко измеряется количественно с помощью просмотра таблицы распространенности заболевания из раздела *Распространенность SCA в популяции*).

Кроме того, можно наблюдать, что случаи синдромов Тернера/Джейкобс (X0 - XYY) характеризуются немного большими краями при рассмотрении нижнего графика. Это является следствием того факта, что случаи эуплоидии симметричны – оба плода вынуждены иметь одинаковую фетальную фракцию (другими словами, нет причины предпочтения отнесения к другому). С другой стороны, случай синдромов Тернера/Джейкобс является асимметричным и, таким образом, он может быть лучше скорректирован по отношению к конкретным показанным данным с помощью варьирования двух фетальных фракций. Однако это не отражает истинной вероятностной разницы между двумя такими случаями (т. е. случай синдромов Тернера/Джейкобс может лучше подходить в отношении данных, поскольку он имеет еще один свободный параметр). Это касается любой другой пары двух возможностей из раздела *Ограничения.*

Пример 10.

Анализ биологического образца - двухплодная беременность либо с двумя эуплоидными плодами женского пола (XX), либо с одним плодом женского пола с синдромом Тернера (X0) и одним плодом женского пола с трисомией по хромосоме X (XXX)

Анализировали образец от матери с двумя плодами женского пола, как показано на **фиг. 10**. Оценка по сути аналогична предыдущему примеру 9.

Пример 11.

Анализ искусственного образца - двухплодная беременность, при которой у одного из двух плодов женского пола имеется синдром Тернера (X0)

Это искусственный образец, который был создан путем взятия одного образца с двумя плодами XY мужского пола и одного образца с двумя плодами XX женского пола и объединения их таким способом, что картировочные данные хромосомы X взяты от первого образца, и картировочные данные хромосомы Y взяты от второго образца.

Данный случай не имеет каких-либо других альтернатив при тех же данных NGS (см. раздел Ограничения). Это также указано на нижнем графике на **фиг. 11**, где случай X0 - X0 преобладает. С другой стороны, априорная вероятность наличия двух плодов женского пола с синдромом Тернера настолько низкая, что это замещает данные NGS

на верхнем графике, таким образом делая второй наиболее вероятный случай (X0 - XX) из нижнего графика наиболее вероятным случаем на верхнем графике.

Литературные источники

[1] Lo, YM Dennis, et al. "Presence of fetal DNA in maternal plasma and serum." *The Lancet* 350.9076 (1997): 485-4

[2] Chiu, Rossa WK, et al. "Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genom sequencing of DNA in maternal plasma". *Proceedings of the National Academy of Sciences* 105.51 (2008): 20458-20463.

[3] Fan, H. Christina, et al. "Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood". *Proceedings of the National Academy of Sciences* 105.42 (2008): 16266-16271.

[4] Chiu, Rossa WK, et al. "Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study." *Bmj* 342 (2011): c7401.

[5] Sehnert, Amy J., et al. "Optimal detection of fetal chromosomal abnormalities by massively parallel DNA sequencing of cell-free fetal DNA from maternal blood." *Clinical chemistry* 57.7 (2011): 1042-1049.

[6] Lau, Tze Kin, et al. "Noninvasive prenatal diagnosis of common fetal chromosomal aneuploidies by maternal plasma DNA sequencing." *The Journal of Maternal-Fetal & Neonatal Medicine* 25.8 (2012): 1370-1374.

[7] Bianchi, Diana W., et al. "Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing". *Obstetrics & Gynecology* 119.5 (2012): 890-901.

[8] Straver, Roy, et al. "WISECONDOR: detection of fetal aberrations from shallow sequencing maternal plasma based on a within-sample comparison scheme". *Nucleic acids research* 42.5 (2014): e31-e31.

[9] Stephanie, C. Yu, et al. "Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing". *Proceedings of the National Academy of Sciences* 111.23 (2014): 8583-8588.

[10] Tynan, J. A., et al. "Application of risk score analysis to low-coverage whole genome sequencing data for the noninvasive detection of trisomy 21, trisomy 18, and trisomy 13." *Prenatal diagnosis* 36.1 (2016): 56-62.

[11] Zimmermann, Bernhard, et al. "Noninvasive prenatal aneuploidy testing of chromosomes 13, 18, 21, X, and Y, using targeted sequencing of polymorphic loci". *Prenatal diagnosis* 32.13 (2012): 1233-1241.

[12] Mazloom, Amin R., et al. "Noninvasive prenatal detection of sex chromosomal aneuploidies by sequencing circulating cell-free DNA from maternal plasma". *Prenatal diagnosis* 33.6 (2013): 591-597.

[13] Liang, Desheng, et al. "Non-invasive prenatal testing of fetal whole chromosome aneuploidy by massively parallel sequencing". *Prenatal diagnosis* 33.5 (2013): 409-415.

[14] Wang, Yanlin, et al. "Maternal mosaicism is a significant contributor to discordant sex chromosomal aneuploidies associated with noninvasive prenatal testing." *Clinical chemistry* 60.1 (2014): 251-259.

[15] https://www.ncbi.nlm.nih.gov/probe/docs/projhapmap/

[16] Kim, Sung K., et al. "Determination of fetal DNA fraction from the plasma of pregnant women using sequence read counts." *Prenatal diagnosis* 35.8 (2015): 810-815.

[17] Straver, Roy, et al. "Calculating the fetal fraction for noninvasive prenatal testing based on genome-wide nucleosome profiles". *Prenatal diagnosis* 35.8 (2016): 614-621.

[18] Stephanie, C. Yu, et al. "Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing". Proceedings of the National Academy of Sciences 111.23 (2014): 8583-8588.

[19] Jiang, Peiyong, et al. "FetalQuantSD: accurate quantification of fetal DNA fraction by shallow-depth sequencing of maternal plasma DNA". *Genomic Medicine* 1 (2016): 16013.

[20] https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/

[21] Cock, Peter JA, et al. "The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants". *Nucleic acids research* 38.6 (2010): 1767-1771.

[22] Li, Heng, et al. "The sequence alignment/map format and SAMtools". *Bioinformatics* 25.16 (2009): 2078-2079.

[23] Benjamini, Yuval, и Terence P. Speed. "Summarizing and correcting the GC content bias in high-throughput sequencing". *Nucleic acids research* (2012): gks001.

[24] Liao, Can, et al. "Noninvasive prenatal diagnosis of common aneuploidies by semiconductor sequencing." *Proceedings of the National Academy of Sciences* 111.20 (2014): 7415-7420.

[25] Minarik, Gabriel, et al. "Utilization of Benchtop Next Generation Sequencing Platforms Ion Torrent PGM and MiSeq in Noninvasive Prenatal Testing for Chromosome 21 Trisomy and Testing of Impact of In Silico and Physical Size Selection on Its Analytical Performance." *PloS one* 10.12 (2015): e0144811.

[26] Genome Reference Consortium Human Build 37 (GRCh37), Feb. 2009, номер доступа в базе GenBank: GCA_000001405.1, номер доступа в базе RefSeq: GCF_000001405.13

[27] Snijders, R. J. M., N. J. Sebire и K. H. Nicolaides. "Maternal age and gestational age-specific risk for chromosomal defects." *Fetal diagnosis and therapy* 10.6 (1995): 356-367.

[28] Norton, Mary E., Laura L. Jelliffe-Pawlowski, and Robert J. Currier. "Chromosome abnormalities detected by current prenatal screening and noninvasive prenatal testing." *Obstetrics & Gynecology* 124.5 (2014): 979-986.

[29] Hook, Ernest B., and Dorothy Warburton. "Turner syndrome revisited: review of new data supports the hypothesis that all viable 45, X cases are cryptic mosaics with a rescue cell line, implying an origin by mitotic loss." *Human genetics* 133.4 (2014): 417-424.

[30] Grati, Francesca R., et al. "Fetoplacental mosaicism: potential implications for false-positive and false-negative noninvasive prenatal screening results." *Genetics in Medicine* 16.8 (2014): 620-624.

[31] Stumm, Markus, et al. "Diagnostic accuracy of random massively parallel sequencing for non-invasive prenatal detection of common autosomal aneuploidies: a collaborative study in Europe." *Prenatal diagnosis* 34.2 (2014): 185-191.

[32] Russell, Stuart, Peter Norvig, and Artificial Intelligence: A modern approach. *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs* 25 (1995): 27.

Формула изобретения

1. Способ определения анеуплоидии половых хромосом, пола и фетальной фракции у одного или более плодов с использованием тестируемого образца крови, плазмы крови или сыворотки крови матери, содержащего смесь фрагментов ДНК, происходящих от плода и матери, где указанная смесь представляет собой циркулирующие молекулы внеклеточной ДНК, при этом способ предусматривает модель на основе нейронной сети для расчета распределения фетальной фракции и совокупность гипотез в отношении аберраций половых хромосом у плода или плодов в тестируемом образце, где способ дополнительно предусматривает следующие стадии:

a) осуществление случайного секвенирования по меньшей мере части множества молекул внеклеточной ДНК, содержащихся в тестируемом образце и в каждом обучающем образце из обучающих наборов *Tf*, который содержит образцы от женщин, *Tm*, который содержит образцы от мужчин, и *Te*, который содержит образцы в случае одного эуплоидного плода мужского пола, и при различных значениях фетальной фракции *f*, с получением тем самым информации о последовательностях для множества фрагментов ДНК, происходящих от плода и матери, с использованием тестируемого образца от матери и каждого обучающего образца, где информация о последовательностях предусматривает последовательности ридов;

b) картирование последовательностей ридов на эталонный человеческий геном и получение процентной доли последовательностей, картированных на хромосомах X и Y, для указанных тестируемого образца и каждого обучающего образца из указанных обучающих наборов *Tf*, *Tm* и *Te*;

c) внесение поправки на сдвиг содержания GC на хромосоме X, но не на хромосоме Y, или на хромосоме Y и не на хромосоме X для указанных тестируемого образца и каждого обучающего образца из указанных обучающих наборов *Tf*, *Tm* и *Te*;

d) получение средних значений показателя картирования с их соответствующими значениями стандартного отклонения для хромосом X и Y и условной хромосомы Z путем вычисления численных значений muX, muY, muZ, sdX, sdY, sdZ на основе указанных обучающих наборов Tf и Tm и применение указанных численных значений muX, muY, muZ, sdX, sdY, sdZ для получения гипотезы в отношении аберраций половых хромосом у плода или плодов в тестируемом образце, где гипотезы

параметризованы по указанным значениям muX, muY, muZ, sdX, sdY, sdZ и отдельному значению фетальной фракции f, которое находится в диапазоне от 5% до 100% с шагом 0,1%;

e) обучение модели на основе нейронной сети для расчета распределения фетальной фракции на указанном обучающем наборе *Te* и применение обученной нейронной сети для получения расчетного значения фетальной фракции *f* тестируемого образца и ошибки данной оценки, которые совместно применяют для определения распределения фетальной фракции как нормального распределения с применением расчетного значения фетальной фракции *f* в виде его среднего значения и ошибки данной оценки в виде его стандартного отклонения, где определение фетальной фракции основано на длине фрагментов ДНК тестируемого образца;

f) расчет вероятности наблюдения процентной доли последовательностей, картированных на хромосомах X и Y, для тестируемого образца согласно каждой из указанных гипотез, где каждая гипотеза параметризована по значению *f*, которое находится в диапазоне от 5% до 100% с шагом 0,1%, и вычисленным значениям muX, muY, muZ, sdX, sdY, sdZ из стадии d);

g) корректировку указанных вероятностей из стадии f) путем умножения каждой из указанных гипотез, параметризованных по конкретному значению *f*, на вероятность значения *f* в распределении фетальной фракции, полученном на стадии e);

h) корректировку указанных вероятностей из стадии f) в соответствии с априорной информацией относительно распространенности указанной гипотезы в общей популяции и необязательно в соответствии с дополнительной релевантной априорной информацией;

i) выбор наиболее вероятной из указанных гипотез с конкретной параметризацией посредством анализа методом максимального правдоподобия с получением таким образом статуса аберраций половых хромосом, пола и фетальной фракции у плода или плодов для тестируемого образца.

2. Способ по п. 1, где анеуплоидия половых хромосом предусматривает синдром Тернера, X0; синдром Клайнфельтера, XXY; синдром трисомии X, XXX; и синдром Джейкобс, XYY.

3. Способ по п. 1 или п. 2, где определение фетальной фракции основано на длине фрагментов ДНК, ограниченной всеми значениями длины, выраженными в парах оснований, от 100 п. о. до 200 п. о..

4. Способ по любому из пп. 1-3, где анализ методом максимального правдоподобия осуществляют в отношении указанных гипотез, параметризованных по значениям muX, muY, muZ, sdX, sdY, sdZ из стадии d) и по фетальной фракции *f*, дискретизированной от начального значения, составляющего 5%, и с шагом 0,1% до 100%.

5. Способ по любому из пп. 1-4, где дополнительная релевантная априорная информация представляет собой одно или более из возраста матери, недели гестации, случаев анеуплоидной беременности в анамнезе матери и различных биохимических маркеров, таких как маркеры на основе альфа-фетопротеина, хорионического гонадотропина человека или эстриола.

6. Способ по любому из пп. 1-5, где гипотезы в отношении аберраций половых хромосом у плода или плодов обеспечивают определение плода или плодов, выбранных из группы, состоящей из эуплоидного плода женского пола, эуплоидного плода мужского пола, плода женского пола с синдромом Тернера, плода женского пола с синдромом трисомии X, плода мужского пола с синдромом Клайнфельтера, плода мужского пола с синдромом Джейкобс в случае одноплодной беременности и двух эуплоидных плодов женского пола, двух эуплоидных плодов мужского пола, эуплоидного плода женского пола с эуплоидным плодом мужского пола, эуплоидного плода женского пола с плодом женского пола с синдромом Тернера, эуплоидного плода женского пола с плодом женского пола с синдромом трисомии X, эуплоидного плода женского пола с плодом мужского пола с синдромом Клайнфельтера, эуплоидного плода женского пола с плодом мужского пола с синдромом Джейкобс, эуплоидного плода мужского пола с плодом женского пола с синдромом Тернера, эуплоидного плода мужского пола с плодом женского пола с синдромом трисомии X, эуплоидного плода мужского пола с плодом мужского пола с синдромом Клайнфельтера, эуплоидного плода мужского пола с плодом мужского пола с синдромом Джейкобс, двух плодов женского пола с синдромом Тернера, плода женского пола с синдромом Тернера с плодом женского пола с синдромом трисомии X, плода женского пола с синдромом Тернера с плодом мужского пола с синдромом Клайнфельтера, плода женского пола с синдромом Тернера с плодом мужского пола с синдромом Джейкобс,

двух плодов женского пола с синдромом трисомии X, плода женского пола с синдромом трисомии X с плодом мужского пола с синдромом Клайнфельтера, плода женского пола с синдромом трисомии X с плодом мужского пола с синдромом Джейкобс, двух плодов мужского пола с синдромом Клайнфельтера, плода мужского пола с синдромом Клайнфельтера с плодом мужского пола с синдромом Джейкобс и двух плодов мужского пола с синдромом Джейкобс.

7. Выполняемый с помощью компьютера способ определения анеуплоидии половых хромосом, пола и фетальной фракции у одного или более плодов с применением случайного секвенирования по меньшей мере части множества молекул внеклеточной ДНК, содержащихся в образце крови, плазмы крови или сыворотки крови матери, содержащем смесь фрагментов ДНК, происходящих от плода и матери, и случайного секвенирования по меньшей мере части множества молекул внеклеточной ДНК, содержащихся в обучающем образце, при этом указанный способ предусматривает модель на основе нейронной сети для расчета распределения фетальной фракции и совокупность гипотез в отношении аберраций половых хромосом у плода или плодов в тестируемом образце, где способ дополнительно предусматривает стадии b)-h) из способа по п. 1.

8. Способ по п. 7, где анеуплоидия половых хромосом выбрана из синдрома Тернера, X0; синдрома Клайнфельтера, XXY; синдрома трисомии X, XXX; и синдрома Джейкобс, XYY.

9. Способ по п. 7 или п. 8, где определение фетальной фракции основано на длине фрагментов ДНК, ограниченной всеми значениями длины, выраженными в парах оснований, от 100 п. о. до 200 п. о.

10. Способ по любому из пп. 7-9, где анализ методом максимального правдоподобия осуществляют в отношении указанных гипотез, параметризованных по значениям muX, muY, muZ, sdX, sdY, sdZ из стадии d) и по фетальной фракции *f*, дискретизированной от начального значения, составляющего 5%, и с шагом 0,1% до 100%.

11. Способ по любому из пп. 7-10, где дополнительная релевантная априорная информация представляет собой одно или более из возраста матери, недели гестации и случаев анеуплоидной беременности в анамнезе матери, и различных биохимических

маркеров, таких как маркеры на основе альфа-фетопротеина, хорионического гонадотропина человека или эстриола.

12. Способ по любому из пп. 7-11, где гипотезы в отношении аберраций половых хромосом у плода или плодов обеспечивают определение плода или плодов, выбранных из группы, состоящей из эуплоидного плода женского пола, эуплоидного плода мужского пола, плода женского пола с синдромом Тернера, плода женского пола с синдромом трисомии X, плода мужского пола с синдромом Клайнфельтера, плода мужского пола с синдромом Джейкобс в случае одноплодной беременности и двух эуплоидных плодов женского пола, двух эуплоидных плодов мужского пола, эуплоидного плода женского пола с эуплоидным плодом мужского пола, эуплоидного плода женского пола с плодом женского пола с синдромом Тернера, эуплоидного плода женского пола с плодом женского пола с синдромом трисомии X, эуплоидного плода женского пола с плодом мужского пола с синдромом Клайнфельтера, эуплоидного плода женского пола с плодом мужского пола с синдромом Джейкобс, эуплоидного плода мужского пола с плодом женского пола с синдромом Тернера, эуплоидного плода мужского пола с плодом женского пола с синдромом трисомии X, эуплоидного плода мужского пола с плодом мужского пола с синдромом Клайнфельтера, эуплоидного плода мужского пола с плодом мужского пола с синдромом Джейкобс, двух плодов женского пола с синдромом Тернера, плода женского пола с синдромом Тернера с плодом женского пола с синдромом трисомии X, плода женского пола с синдромом Тернера с плодом мужского пола с синдромом Клайнфельтера, плода женского пола с синдромом Тернера с плодом мужского пола с синдромом Джейкобс, двух плодов женского пола с синдромом трисомии X, плода женского пола с синдромом трисомии X с плодом мужского пола с синдромом Клайнфельтера, плода женского пола с синдромом трисомии X с плодом мужского пола с синдромом Джейкобс, двух плодов мужского пола с синдромом Клайнфельтера, плода мужского пола с синдромом Клайнфельтера с плодом мужского пола с синдромом Джейкобс и двух плодов мужского пола с синдромом Джейкобс.

13. Компьютерный программный продукт, содержащий машиночитаемый носитель, содержащий множество инструкций по управлению компьютерной системой для осуществления способа по любому из пп. 7-12.

Формула изобретения

1. Способ определения анеуплоидии половых хромосом, пола и фетальной фракции у одного или более плодов с использованием тестируемого образца крови, плазмы крови или сыворотки крови матери, содержащего смесь фрагментов ДНК, происходящих от плода и матери, где указанная смесь представляет собой циркулирующие молекулы внеклеточной ДНК, при этом способ предусматривает модель на основе нейронной сети для расчета распределения фетальной фракции и совокупность гипотез в отношении аберраций половых хромосом у плода или плодов в тестируемом образце, где способ дополнительно предусматривает следующие стадии:

a) осуществление случайного секвенирования по меньшей мере части множества молекул внеклеточной ДНК, содержащихся в тестируемом образце и в каждом обучающем образце из обучающих наборов *Tf*, который содержит образцы от взрослых женщин, *Tm*, который содержит образцы от взрослых мужчин, и *Te*, который содержит образцы от беременных женщин в случае одного эуплоидного плода мужского пола и при различных значениях фетальной фракции *f*, с получением тем самым информации о последовательностях для множества фрагментов ДНК, происходящих от плода и матери, с использованием тестируемого образца от матери и каждого обучающего образца, где информация о последовательностях предусматривает последовательности ридов;

b) картирование последовательностей ридов на эталонный человеческий геном и получение процентной доли последовательностей, картированных на хромосомах X и Y, для указанных тестируемого образца и каждого обучающего образца из указанных обучающих наборов *Tf*, *Tm* и *Te*;

c) внесение поправки на сдвиг содержания GC на хромосоме X, но не на хромосоме Y, или на хромосоме Y и не на хромосоме X для указанных тестируемого образца и каждого обучающего образца из указанных обучающих наборов *Tf*, *Tm* и *Te*;

d) получение средних значений показателя картирования с их соответствующими значениями стандартного отклонения для хромосомы X и хромосомы Y на основе указанного обучающего набора *Tf*, где в случае обучающего набора *Tf* хромосома Y обозначена как условная хромосома Z, и где полученные средние значения и значения

стандартного отклонения, относящиеся к хромосоме X и условной хромосоме Z, представляют собой muX и sdX и muZ и sdZ соответственно;

e) получение среднего значения показателя картирования с его соответствующим значением стандартного отклонения для хромосомы Y на основе указанного обучающего набора *Tm*, где вычисленные среднее значение и значение стандартного отклонения представляют собой muY и sdY соответственно;

f) применение указанных численных значений muX, muY, muZ, sdX, sdY, sdZ для получения гипотез в отношении аберраций половых хромосом у плода или плодов в тестируемом образце, где гипотезы параметризованы по указанным значениям muX, muY, muZ, sdX, sdY, sdZ и отдельному значению фетальной фракции *f*, которое находится в диапазоне от 5% до 100% с шагом 0,1%;

g) обучение модели на основе нейронной сети для расчета распределения фетальной фракции в указанном обучающем наборе *Te* и применение обученной нейронной сети для получения расчетного значения фетальной фракции *f* тестируемого образца и ошибки данной оценки, которые совместно применяют для определения распределения фетальной фракции как нормального распределения с применением расчетного значения фетальной фракции *f* в виде его среднего значения и ошибки данной оценки в виде его стандартного отклонения, где определение фетальной фракции основано на длине фрагментов ДНК тестируемого образца;

h) расчет вероятности наблюдения процентной доли последовательностей, картированных на хромосомах X и Y, для тестируемого образца согласно каждой из указанных гипотез, где каждая гипотеза параметризована по значению *f*, которое находится в диапазоне от 5% до 100% с шагом 0,1%, и вычисленным значениям muX, muY, muZ, sdX, sdY, sdZ из стадий d) и e);

i) корректировку указанных вероятностей из стадии h) путем умножения каждой из указанных гипотез, параметризованных по конкретному значению *f*, на вероятность значения *f* в распределении фетальной фракции, полученном на стадии g);

j) корректировку указанных вероятностей из стадии i) в соответствии с априорной информацией относительно распространенности указанной гипотезы в общей

популяции и необязательно в соответствии с дополнительной релевантной априорной информацией;

k) выбор наиболее вероятной из указанных гипотез в отношении вероятностей из любой из стадий h), i), j) с конкретной параметризацией посредством анализа методом максимального правдоподобия с получением таким образом статуса аберраций половых хромосом, пола и фетальной фракции у плода или плодов для тестируемого образца.

2. Способ по п. 1, где анеуплоидия половых хромосом предусматривает синдром Тернера, X0; синдром Клайнфельтера, XXY; синдром трисомии X, XXX; и синдром Джейкобс, XYY.

3. Способ по п. 1 или п. 2, где определение фетальной фракции основано на длине фрагментов ДНК, ограниченной всеми значениями длины, выраженными в парах оснований, от 100 п. о. до 200 п. о.

4. Способ по любому из пп. 1-3, где анализ методом максимального правдоподобия осуществляют в отношении указанных гипотез, параметризованных по значениям muX, muY, muZ, sdX, sdY, sdZ из стадий d) и e) и по фетальной фракции *f*, дискретизированной от начального значения, составляющего 5%, и с шагом 0,1% до 100%.

5. Способ по любому из пп. 1-4, где дополнительная релевантная априорная информация представляет собой одно или более из возраста матери, недели гестации, случаев анеуплоидной беременности в анамнезе матери и различных биохимических маркеров, таких как маркеры на основе альфа-фетопротеина, хорионического гонадотропина человека или эстриола.

6. Способ по любому из пп. 1-5, где гипотезы в отношении аберраций половых хромосом у плода или плодов обеспечивают определение плода или плодов, выбранных из группы, состоящей из эуплоидного плода женского пола, эуплоидного плода мужского пола, плода женского пола с синдромом Тернера, плода женского пола с синдромом трисомии X, плода мужского пола с синдромом Клайнфельтера, плода мужского пола с синдромом Джейкобс в случае одноплодной беременности и двух эуплоидных плодов женского пола, двух эуплоидных плодов мужского пола,

эуплоидного плода женского пола с эуплоидным плодом мужского пола, эуплоидного плода женского пола с плодом женского пола с синдромом Тернера, эуплоидного плода женского пола с плодом женского пола с синдромом трисомии X, эуплоидного плода женского пола с плодом мужского пола с синдромом Клайнфельтера, эуплоидного плода женского пола с плодом мужского пола с синдромом Джейкобс, эуплоидного плода мужского пола с плодом женского пола с синдромом Тернера, эуплоидного плода мужского пола с плодом женского пола с синдромом трисомии X, эуплоидного плода мужского пола с плодом мужского пола с синдромом Клайнфельтера, эуплоидного плода мужского пола с плодом мужского пола с синдромом Джейкобс, двух плодов женского пола с синдромом Тернера, плода женского пола с синдромом Тернера с плодом женского пола с синдромом трисомии X, плода женского пола с синдромом Тернера с плодом мужского пола с синдромом Клайнфельтера, плода женского пола с синдромом Тернера с плодом мужского пола с синдромом Джейкобс, двух плодов женского пола с синдромом трисомии X, плода женского пола с синдромом трисомии X с плодом мужского пола с синдромом Клайнфельтера, плода женского пола с синдромом трисомии X с плодом мужского пола с синдромом Джейкобс, двух плодов мужского пола с синдромом Клайнфельтера, плода мужского пола с синдромом Клайнфельтера с плодом мужского пола с синдромом Джейкобс и двух плодов мужского пола с синдромом Джейкобс.

7. Выполняемый с помощью компьютера способ определения анеуплоидии половых хромосом, пола и фетальной фракции у одного или более плодов с применением случайного секвенирования по меньшей мере части множества молекул внеклеточной ДНК, содержащихся в образце крови, плазмы крови или сыворотки крови матери, содержащем смесь фрагментов ДНК, происходящих от плода и матери, и случайного секвенирования по меньшей мере части множества молекул внеклеточной ДНК, содержащихся в обучающем образце, при этом указанный способ предусматривает модель на основе нейронной сети для расчета распределения фетальной фракции и совокупность гипотез в отношении аберраций половых хромосом у плода или плодов в тестируемом образце, где способ дополнительно предусматривает стадии b)-h) из способа по п. 1.

8. Способ по п. 7, где анеуплоидия половых хромосом выбрана из синдрома Тернера, X0; синдрома Клайнфельтера, XXY; синдрома трисомии X, XXX; и синдрома Джейкобс, XYY.

9. Способ по п. 7 или п. 8, где определение фетальной фракции основано на длине фрагментов ДНК, ограниченной всеми значениями длины, выраженными в парах оснований, от 100 п. о. до 200 п. о.

10. Способ по любому из пп. 7-9, где анализ методом максимального правдоподобия осуществляют в отношении указанных гипотез, параметризованных по значениям muX, muY, muZ, sdX, sdY, sdZ из стадии d) и по фетальной фракции *f*, дискретизированной от начального значения, составляющего 5%, и с шагом 0,1% до 100%.

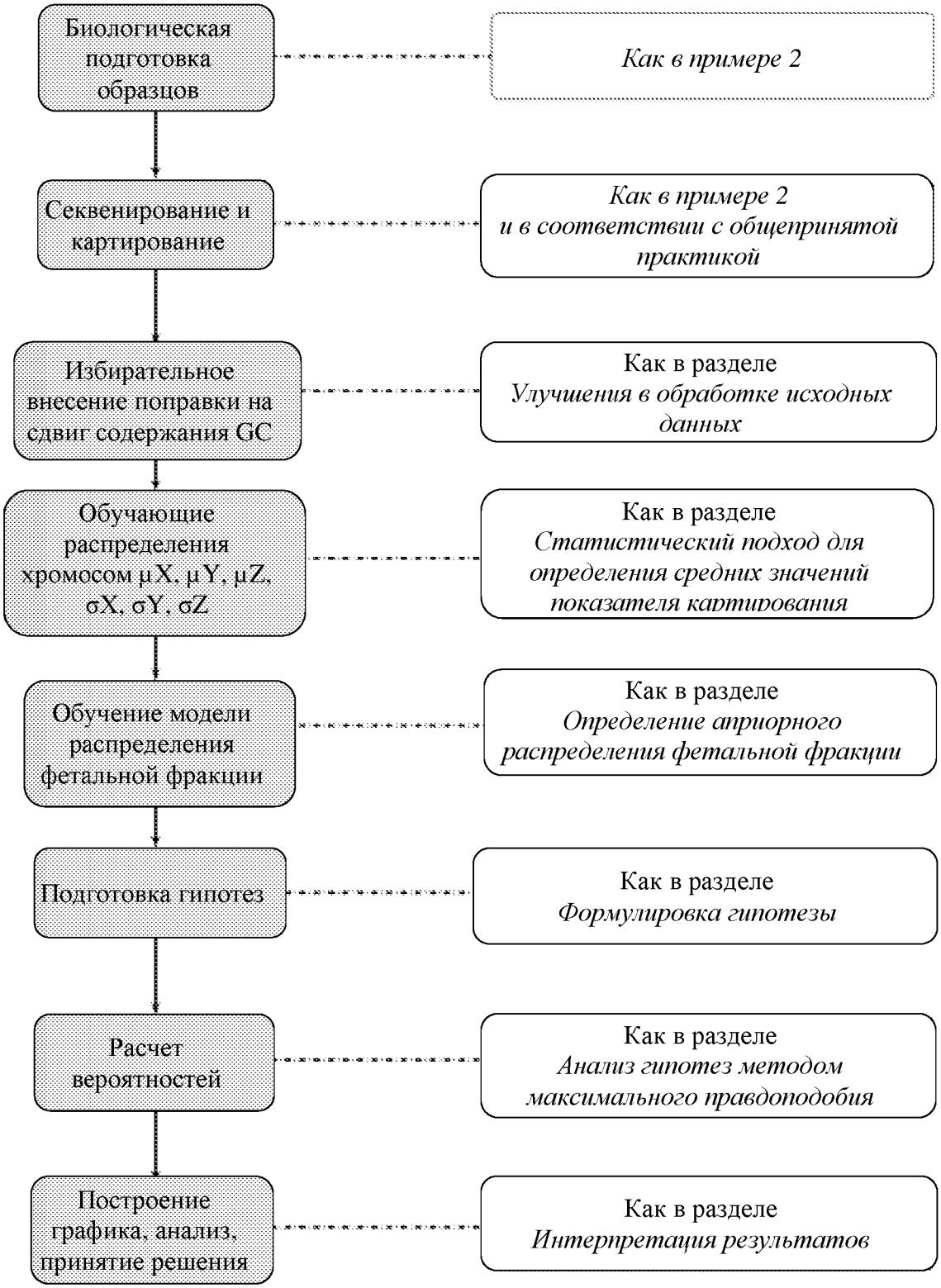
11. Способ по любому из пп. 7-10, где дополнительная релевантная априорная информация представляет собой одно или более из возраста матери, недели гестации, и случаев анеуплоидной беременности в анамнезе матери, и различных биохимических маркеров, таких как маркеры на основе альфа-фетопротеина, хорионического гонадотропина человека или эстриола.

12. Способ по любому из пп. 7-11, где гипотезы в отношении аберраций половых хромосом у плода или плодов обеспечивают определение плода или плодов, выбранных из группы, состоящей из эуплоидного плода женского пола, эуплоидного плода мужского пола, плода женского пола с синдромом Тернера, плода женского пола с синдромом трисомии X, плода мужского пола с синдромом Клайнфельтера, плода мужского пола с синдромом Джейкобс в случае одноплодной беременности и двух эуплоидных плодов женского пола, двух эуплоидных плодов мужского пола, эуплоидного плода женского пола с эуплоидным плодом мужского пола, эуплоидного плода женского пола с плодом женского пола с синдромом Тернера, эуплоидного плода женского пола с плодом женского пола с синдромом трисомии X, эуплоидного плода женского пола с плодом мужского пола с синдромом Клайнфельтера, эуплоидного плода женского пола с плодом мужского пола с синдромом Джейкобс, эуплоидного плода мужского пола с плодом женского пола с синдромом Тернера, эуплоидного плода мужского пола с плодом женского пола с синдромом трисомии X, эуплоидного плода мужского пола с плодом мужского пола с синдромом Клайнфельтера, эуплоидного плода мужского пола с плодом мужского пола с синдромом Джейкобс,

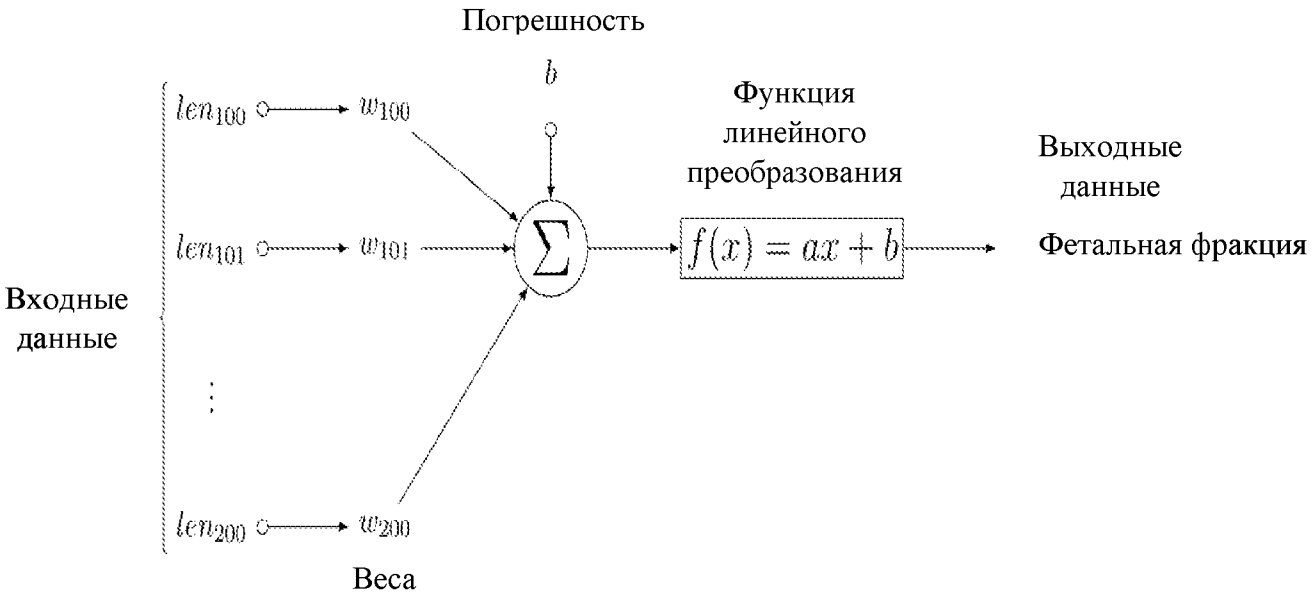
двух плодов женского пола с синдромом Тернера, плода женского пола с синдромом Тернера с плодом женского пола с синдромом трисомии X, плода женского пола с синдромом Тернера с плодом мужского пола с синдромом Клайнфельтера, плода женского пола с синдромом Тернера с плодом мужского пола с синдромом Джейкобс, двух плодов женского пола с синдромом трисомии X, плода женского пола с синдромом трисомии X с плодом мужского пола с синдромом Клайнфельтера, плода женского пола с синдромом трисомии X с плодом мужского пола с синдромом Джейкобс, двух плодов мужского пола с синдромом Клайнфельтера, плода мужского пола с синдромом Клайнфельтера с плодом мужского пола с синдромом Джейкобс и двух плодов мужского пола с синдромом Джейкобс.

13. Компьютерный программный продукт, содержащий машиночитаемый носитель, содержащий множество инструкций по управлению компьютерной системой для осуществления способа по любому из пп. 7-12.

Фиг. 1

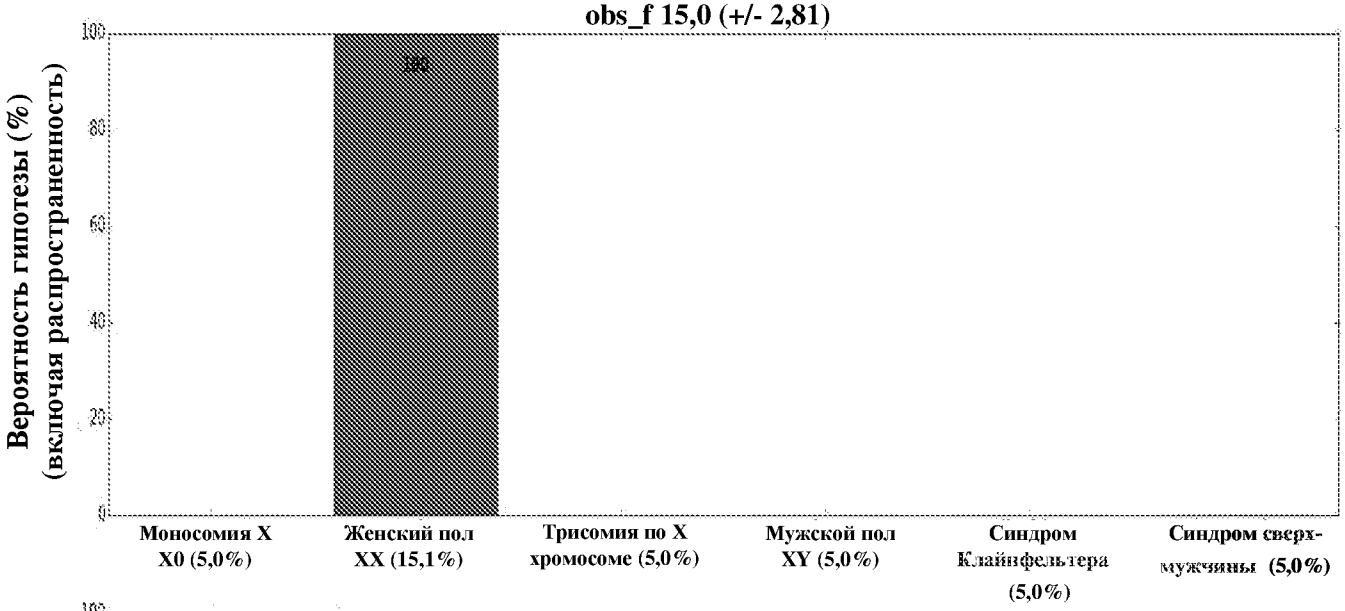

Биологическая подготовка образцов
Как в примере 2
Секвенирование и картирование
Как в примере 2 и в соответствии с общепринятой практикой
Избирательное внесение поправки на сдвиг содержания GC
Как в разделе Улучшения в обработке исходных данных
Обучающие распределения хромосом μX, μY, μZ, σX, σY, σZ
Как в разделе Статистический подход для определения средних значений показателя картирования
Обучение модели распределения фетальной фракции
Как в разделе Определение априорного распределения фетальной фракции
Подготовка гипотез
Как в разделе Формулировка гипотезы
Расчет вероятностей
Как в разделе Анализ гипотез методом максимального правдоподобия
Построение графика, анализ, принятие решения
Как в разделе Интерпретация результатов

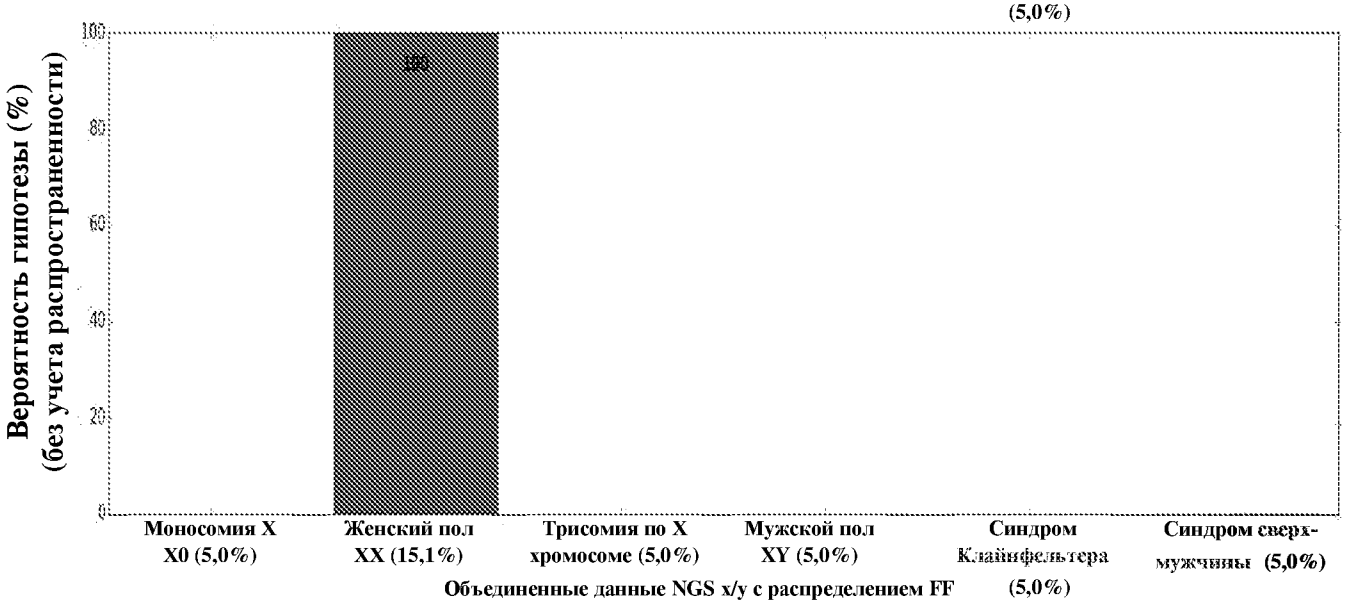
Фиг. 2


Погрешность
b
Функция линейного преобразования
Выходные данные
Фетальная фракция
len_{100}
len_{101}
len_{200}
w_{100}
w_{101}
w_{200}
$\sum$
$f(x) = ax + b$
Входные данные
Веса

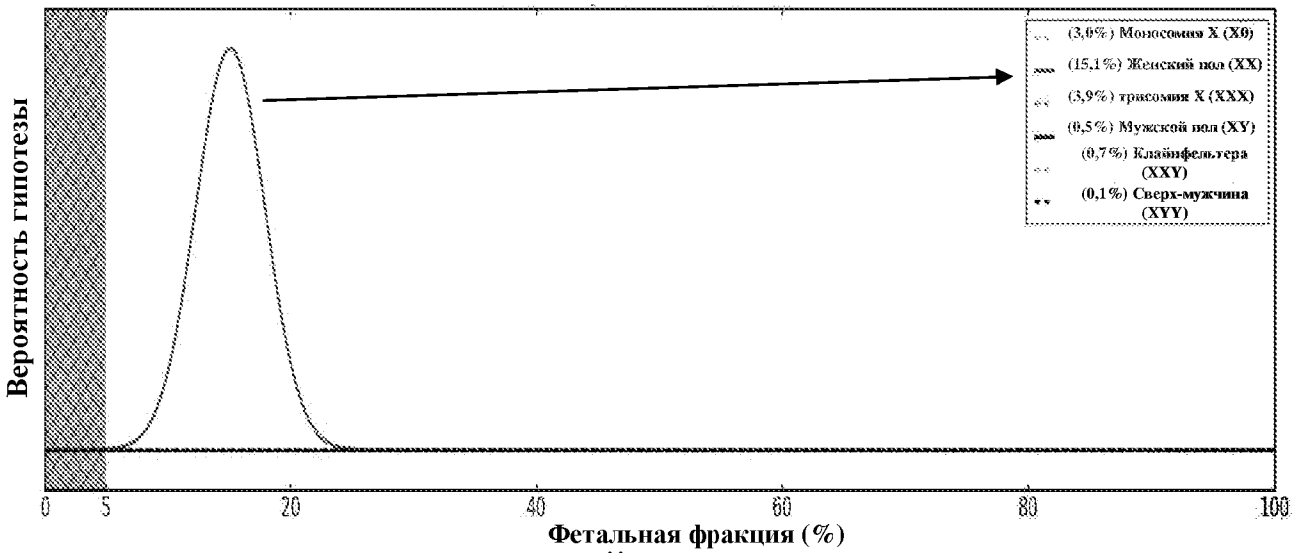
Фиг. 3

Вероятность случая для chrX/chrY
obs x: 0,0491979
obs y: 3,4699e-05
obs_f 15,0 (+/- 2,81)



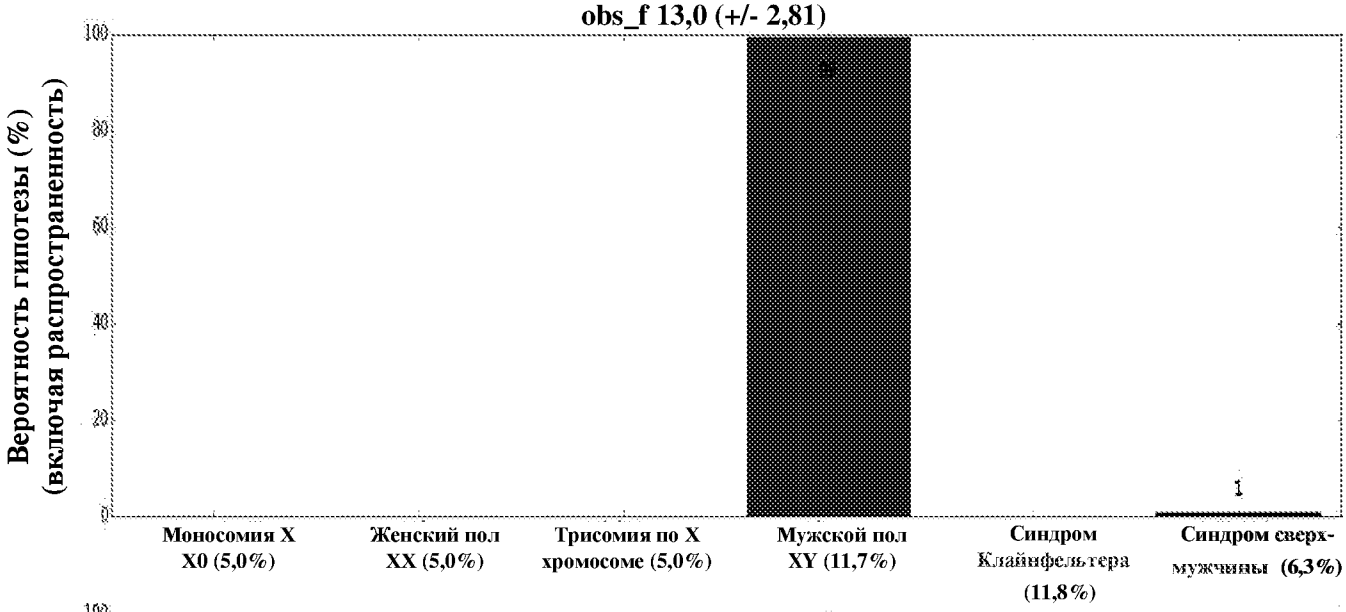


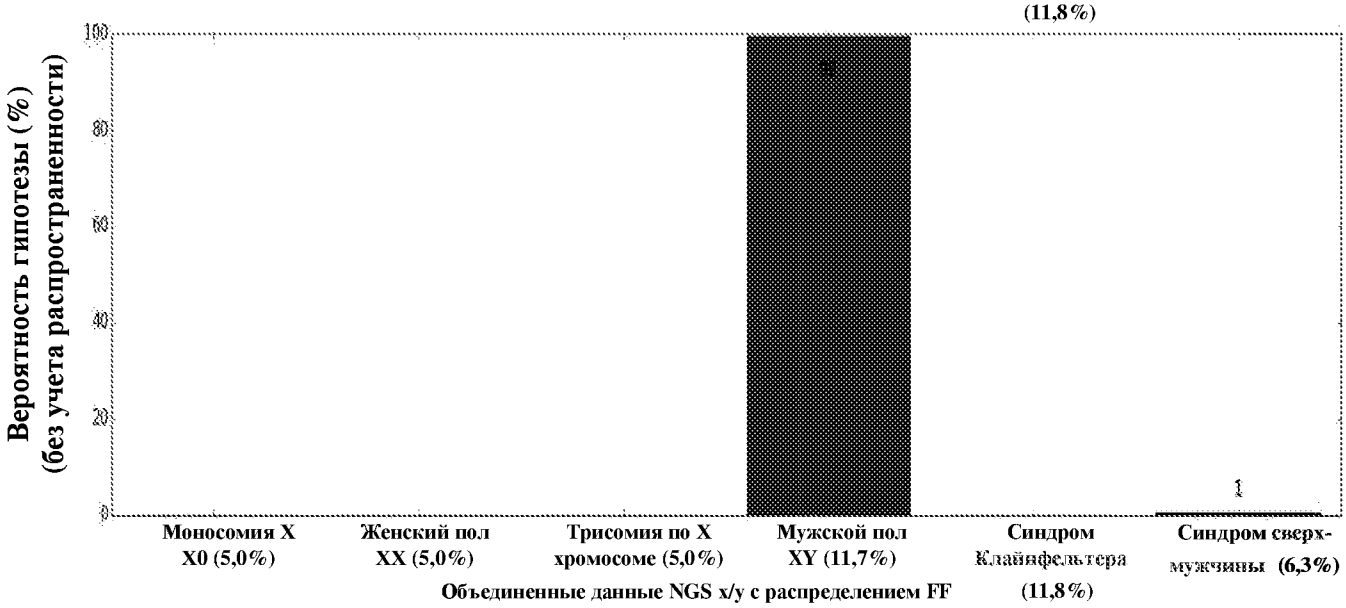
Объединенные данные NGS x/y с распределением FF



Фиг. 4

Вероятность случая для chrX/chrY
obs x: 0,046612
obs y: 0,000240707
obs_f 13,0 (+/- 2,81)

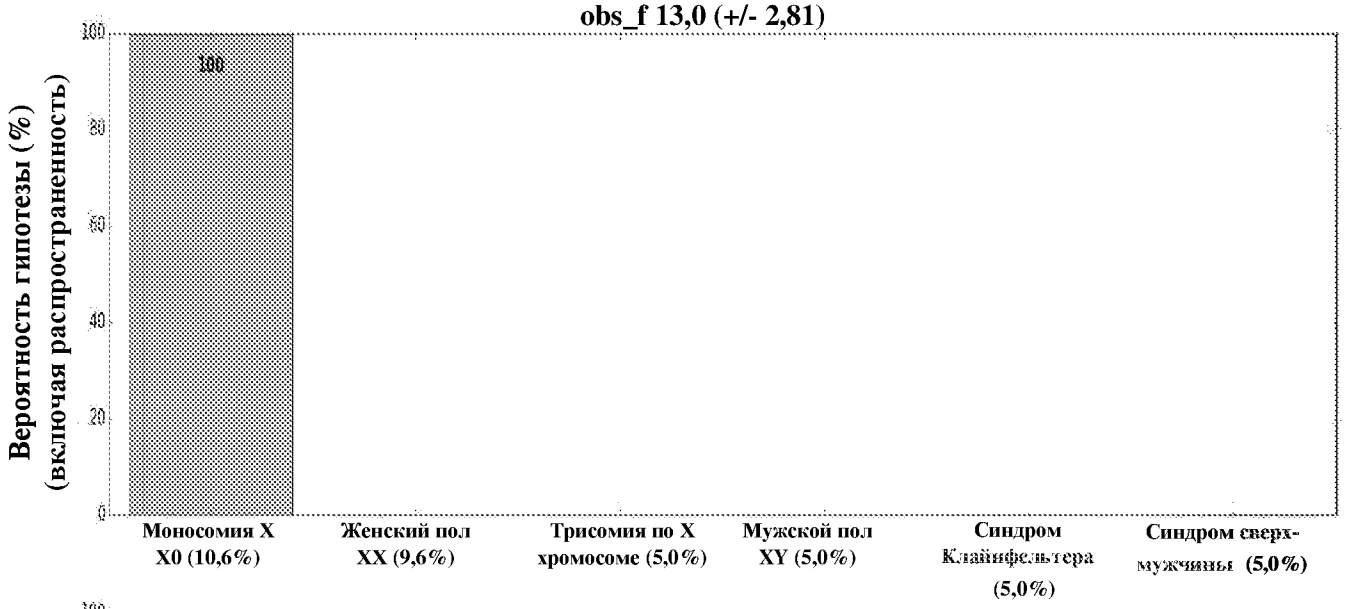


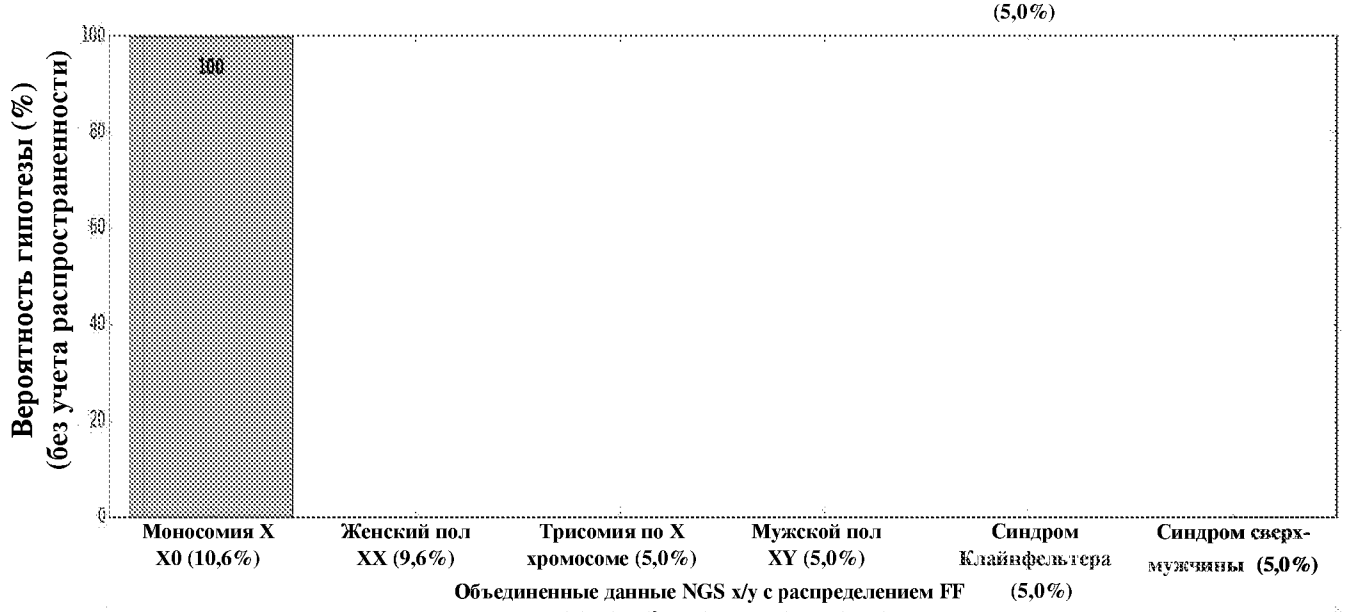


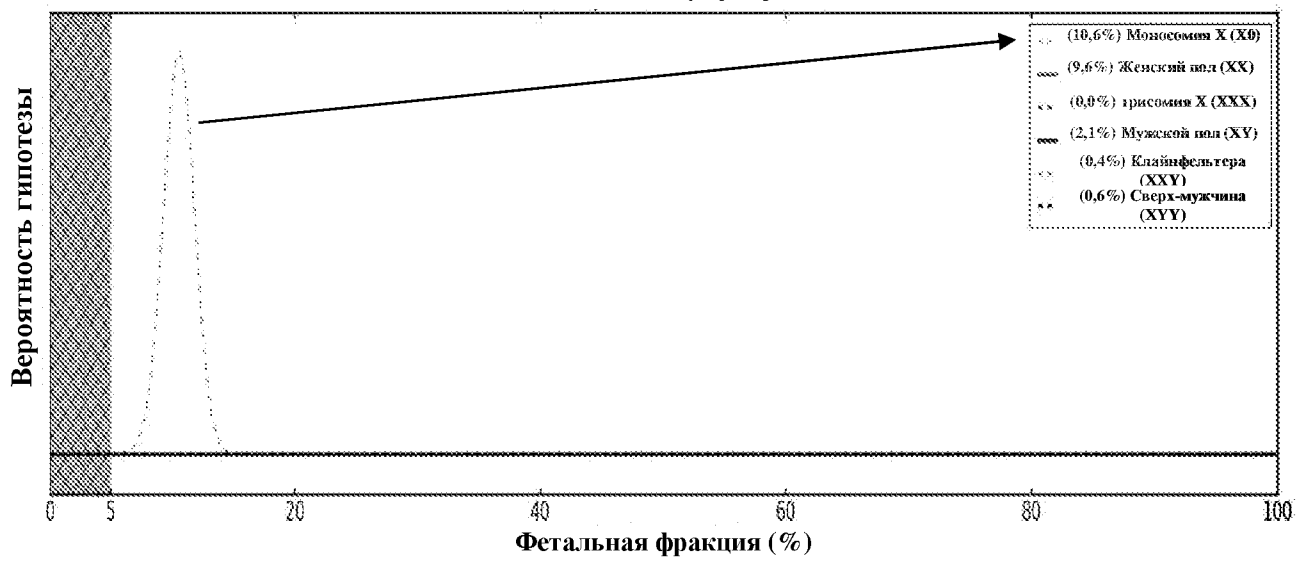
Объединенные данные NGS x/y с распределением FF

Фиг. 5

Вероятность случая для chrX/chrY
obs x: 0,046612
obs y: 3,4699e-05
obs_f 13,0 (+/- 2,81)

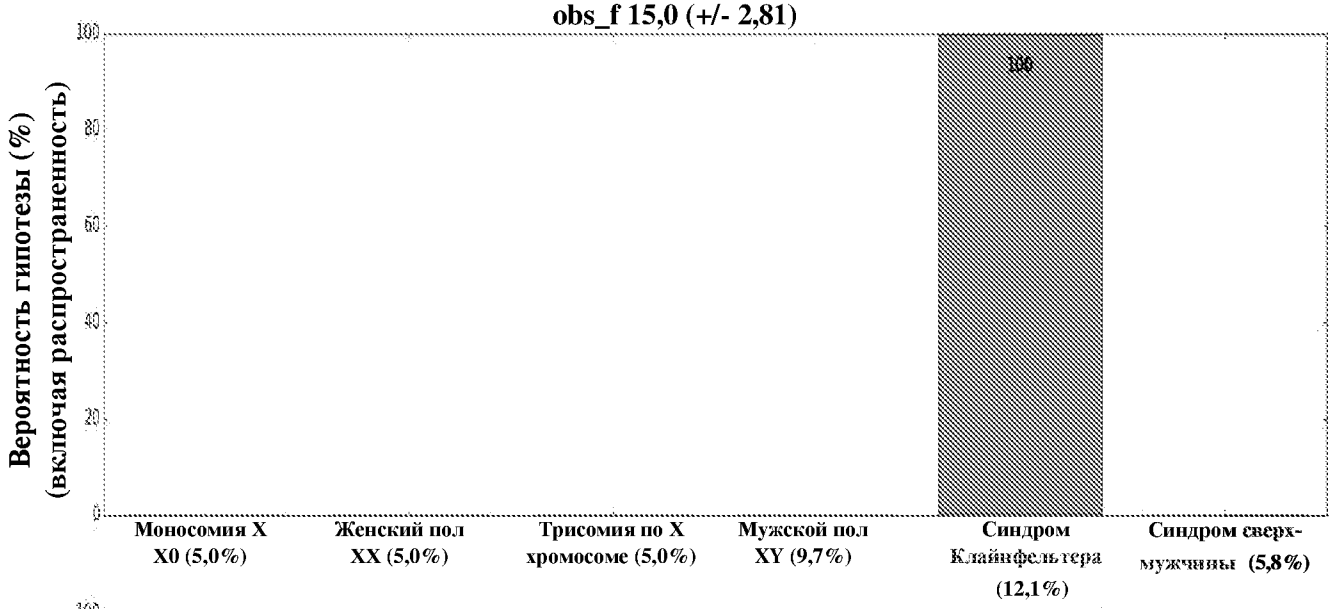


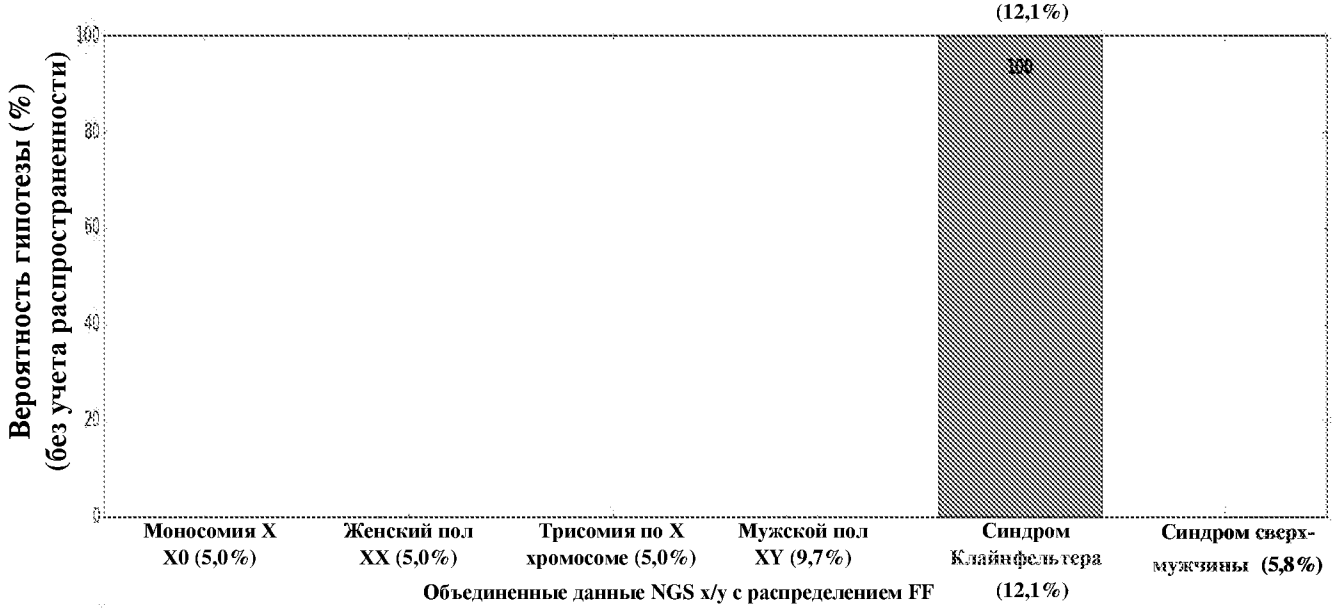




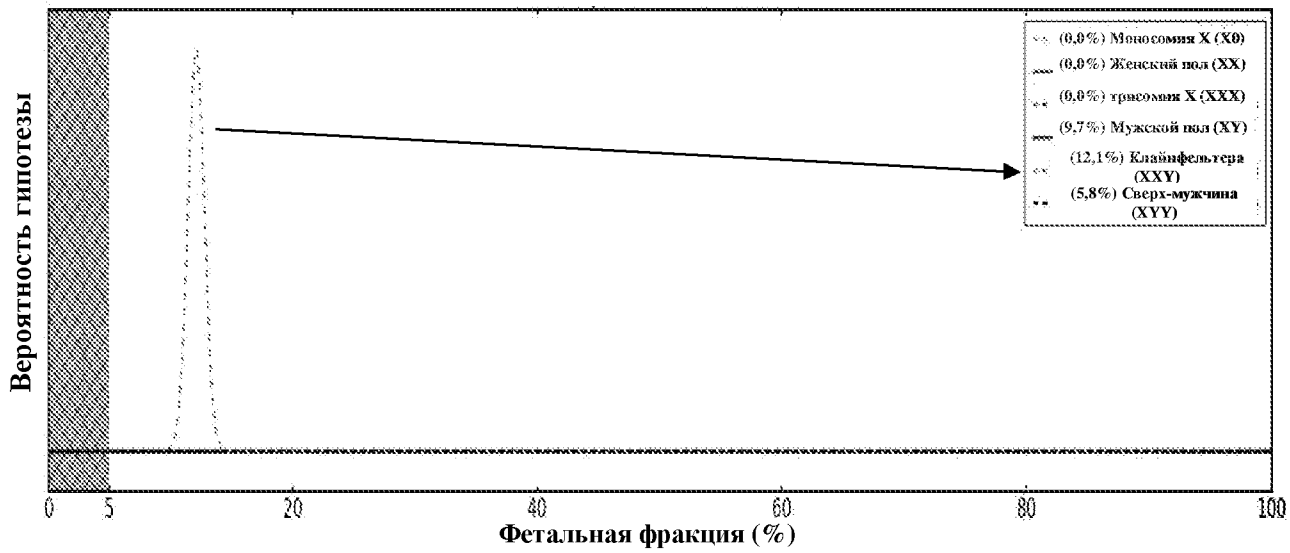
Фиг. 6

Вероятность случая для chrX/chrY
obs x: 0,0491979
obs y: 0,000240707
obs_f 15,0 (+/- 2,81)



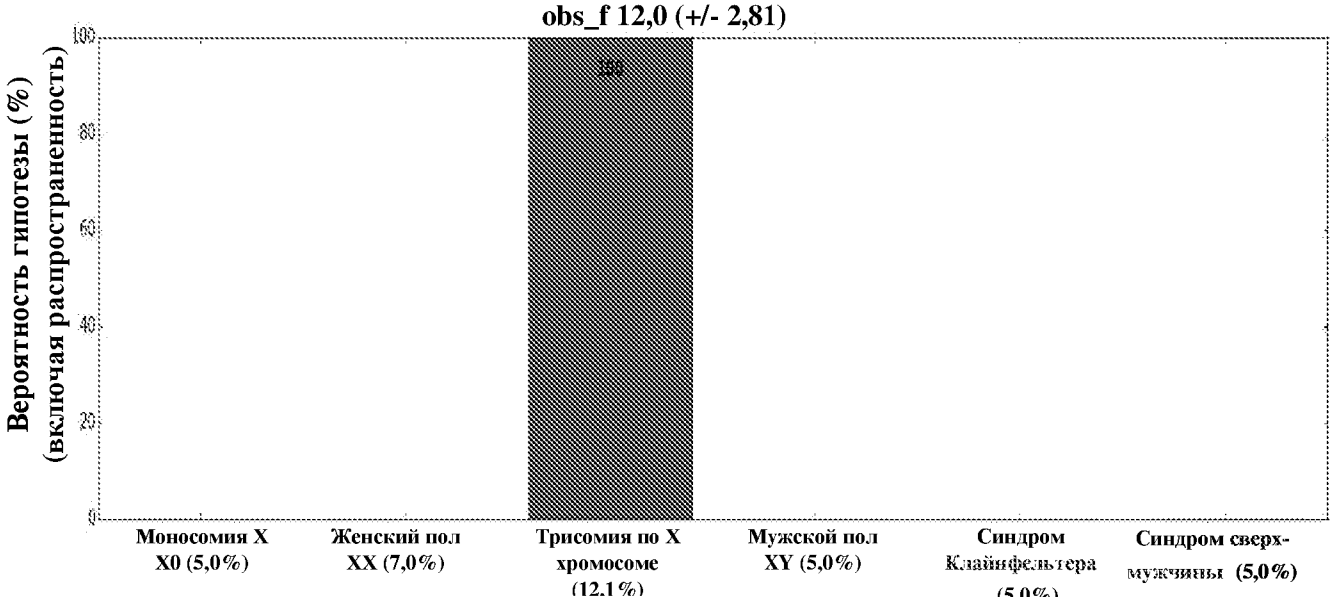


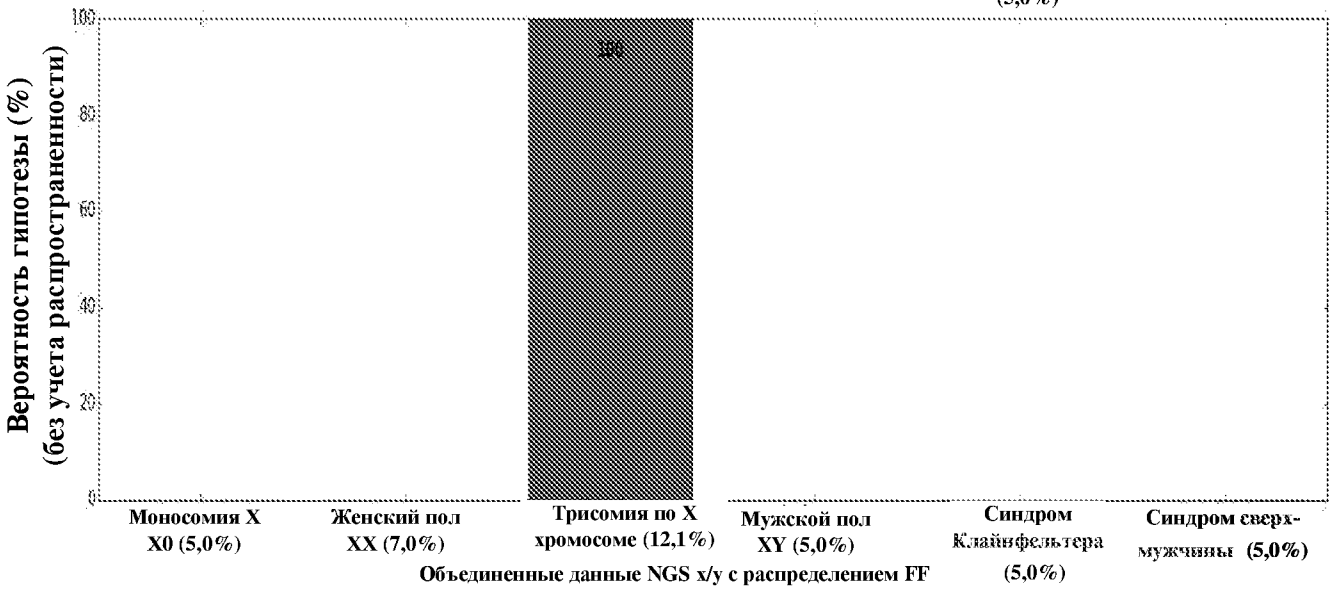
Объединенные данные NGS x/y с распределением FF



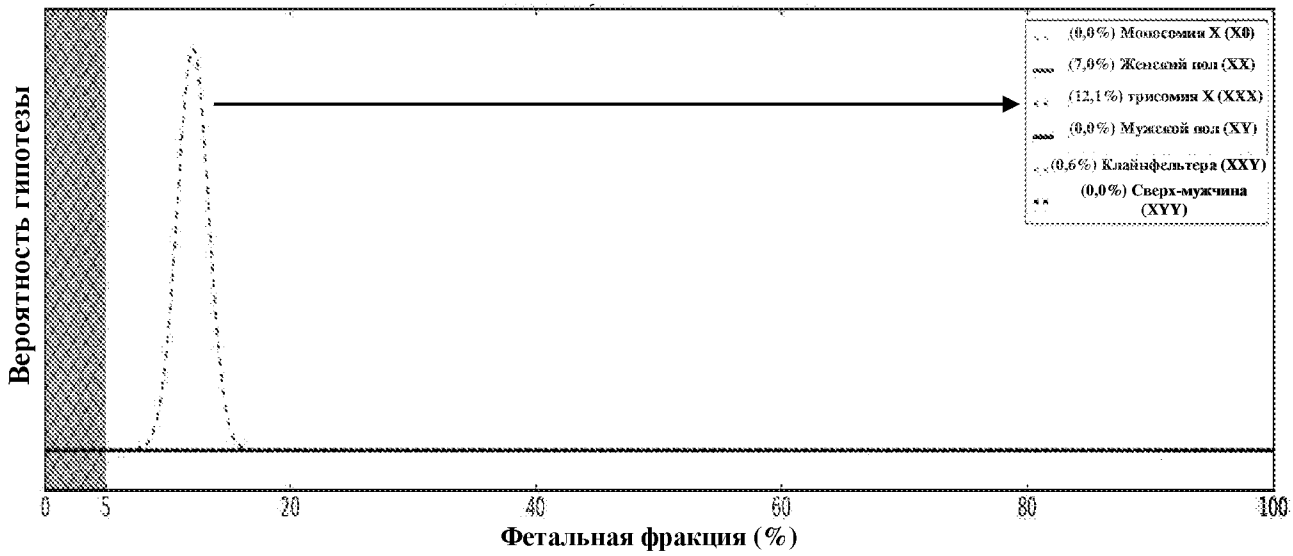
Фиг. 7

Вероятность случая для chrX/chrY
obs x: 0,052006406532
obs y: 4,1981876376e-05
obs_f 12,0 (+/- 2,81)



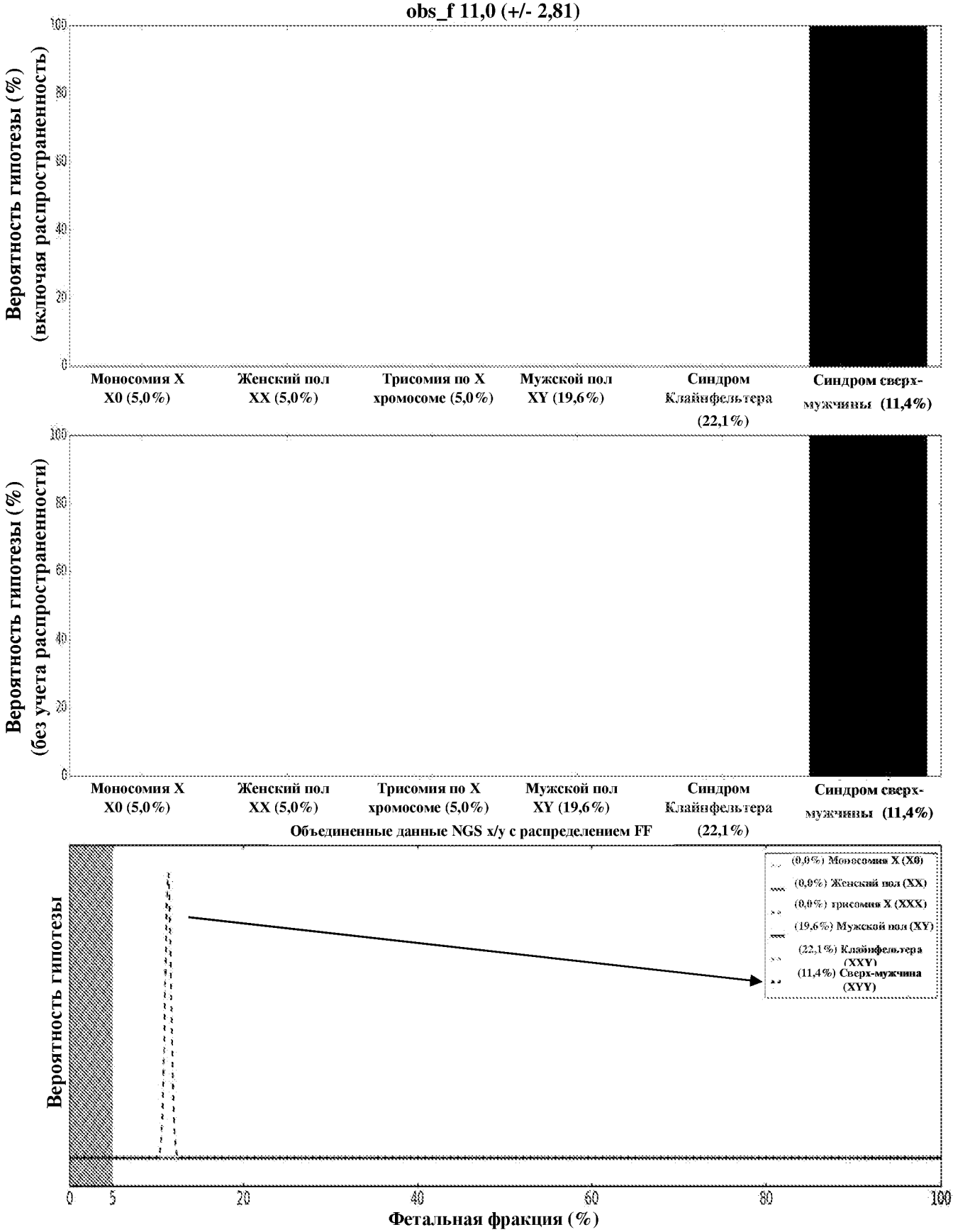


Объединенные данные NGS x/y с распределением FF

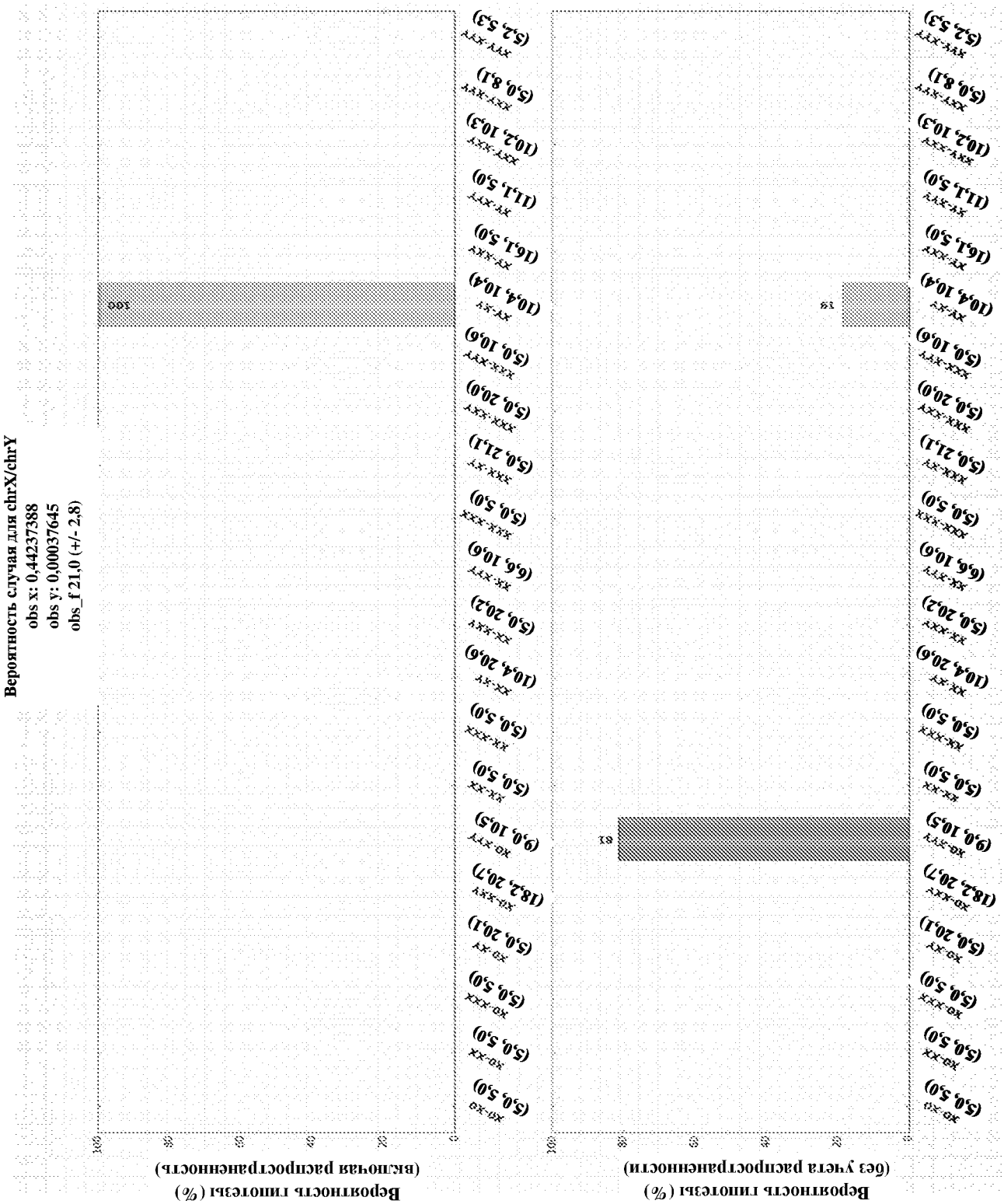


Фиг. 8

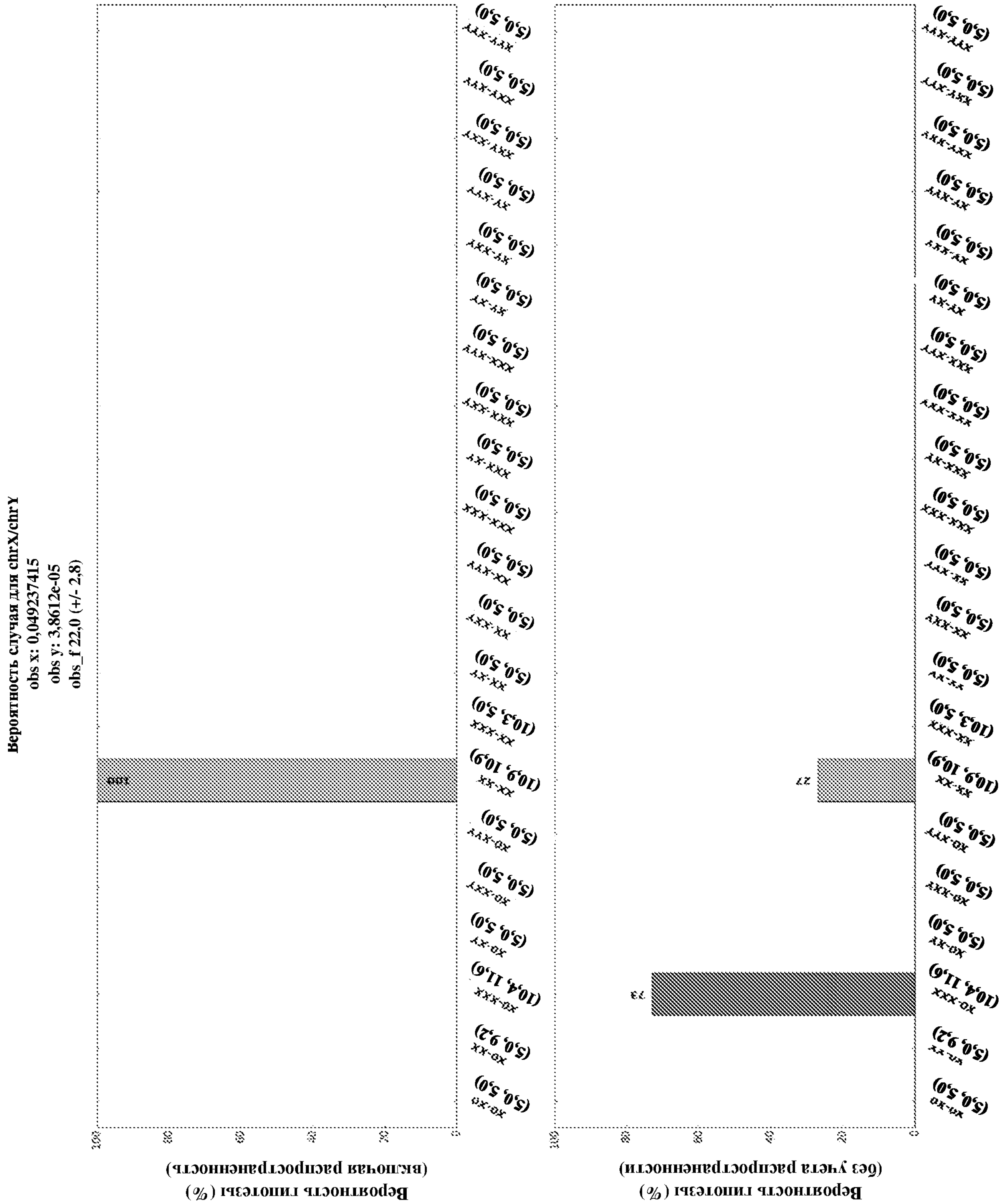
Вероятность случая для chrX/chrY
obs x: 0,0468443571216
obs y: 0,000424924854676
obs_f 11,0 (+/- 2,81)



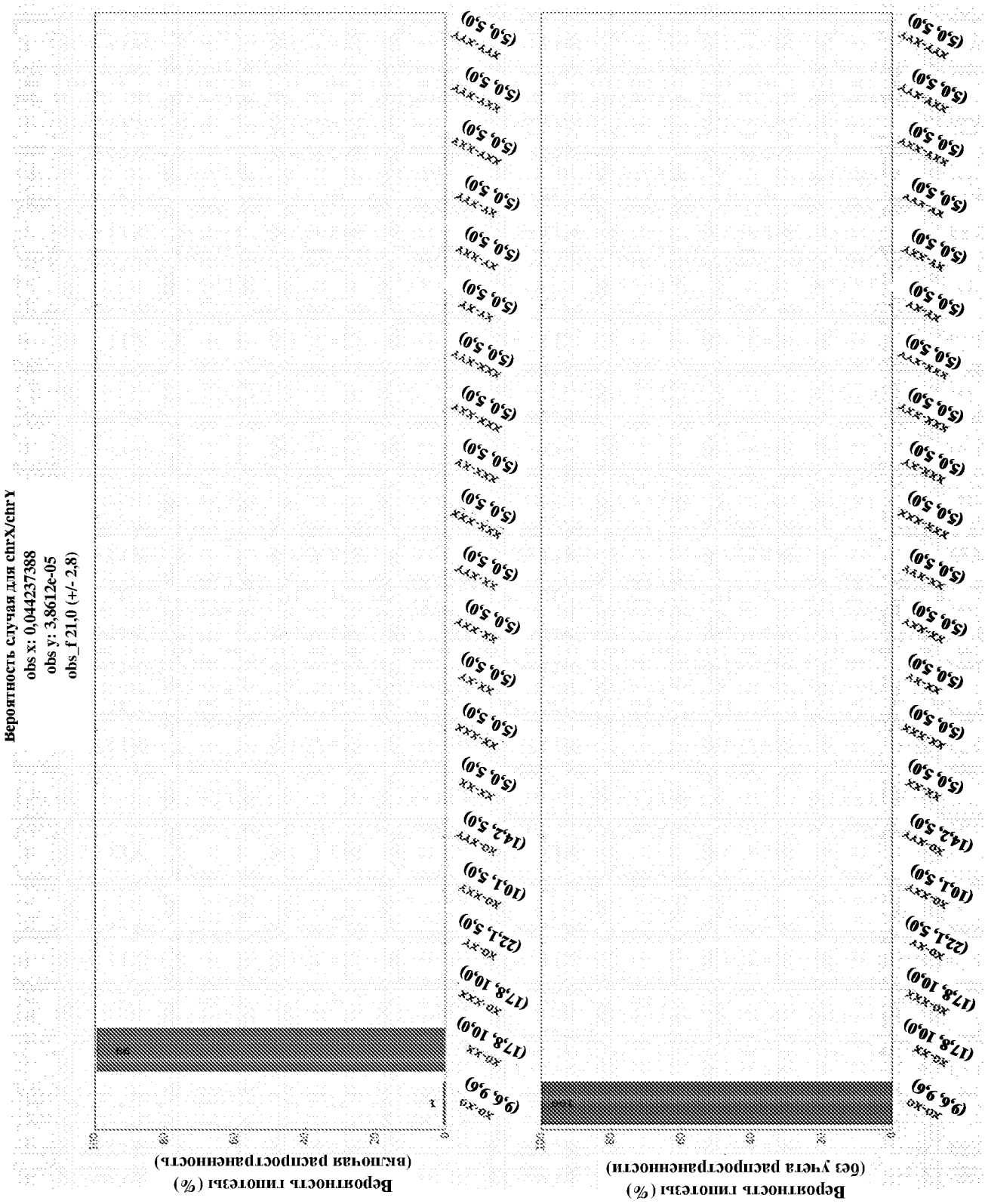
Фиг. 9


Вероятность случая для chrX/chrY
obs x: 0,44237388
obs y: 0,00037645
obs_f 21,0 (+/- 2,8)
Вероятность гипотезы (%)
(включая распространенность)
Вероятность гипотезы (%)
(без учета распространенности)
100
19
61

Фиг. 10


Вероятность случая для chrX/chrY
obs x: 0,049237415
obs y: 3,8612e-05
obs_f 22,0 (+/- 2,8)
Вероятность гипотезы (%)
(включая распространенность)
Вероятность гипотезы (%)
(без учета распространенности)
100
27
73
xo-xx (5,0, 9,2)
xo-xxx (10,4, 11,6)
xx-xx (10,9, 10,9)
xx-xxx (10,3, 5,0)

Фиг. 11


Вероятность случая для chrX/chrY
obs x: 0,044237388
obs y: 3,8612e-05
obs_f 21,0 (+/- 2,8)
Вероятность гипотезы (%)
(включая распространенность)
Вероятность гипотезы (%)
(без учета распространенности)