

(19)



**Евразийское
патентное
ведомство**

(21) **202090325** (13) **A1**

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

(43) Дата публикации заявки
2020.05.12

(22) Дата подачи заявки
2017.07.26

(51) Int. Cl. *C12Q 1/6869* (2018.01)
G06F 19/22 (2011.01)
G06F 19/24 (2011.01)
C12Q 1/6883 (2018.01)

(54) **СПОСОБ НЕИНВАЗИВНОГО ПРЕНАТАЛЬНОГО ВЫЯВЛЕНИЯ ХРОМОСОМНОЙ АНЕУПЛОИДИИ ПЛОДА ПО КРОВИ МАТЕРИ НА ОСНОВЕ БАЙЕСОВСКОЙ СЕТИ**

(86) PCT/EP2017/068881

(87) WO 2019/020180 2019.01.31

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

(72) Изобретатель:

Будиш Ярослав, Дюриш Франтишек,
Кухарик Марцель, Газдарица Юрай,
Сземес Томаш (SK)

(74) Представитель:

Носырева Е.Л. (RU)

(57) В настоящем изобретении предусмотрен эффективный и точный способ определения анеуплоидии и пола у плода, а также соответствующей фетальной фракции по образцу крови, плазмы крови или сыворотки крови матери, содержащему смесь фрагментов ДНК, происходящих от плода и матери. Способ является особенно применимым для выявления наличия или отсутствия анеуплоидий, таких как трисомия и, более конкретно, трисомия по хромосомам 21, 18 и 13. Основным компонентом способа является байесовская сеть, описывающая смесь фрагментов ДНК, происходящих от плода и матери, циркулирующих в крови матери, с помощью которой распределения по длинам фрагментов, обнаруживаемых в крови, используют для определения фетальной фракции.

202090325

A1

A1

202090325

**СПОСОБ НЕИНВАЗИВНОГО ПРЕНАТАЛЬНОГО ВЫЯВЛЕНИЯ
ХРОМОСОМНОЙ АНЕУПЛОИДИИ ПЛОДА ПО КРОВИ МАТЕРИ НА ОСНОВЕ
БАЙЕСОВСКОЙ СЕТИ**

ОБЛАСТЬ ТЕХНИКИ, К КОТОРОЙ ОТНОСИТСЯ ИЗОБРЕТЕНИЕ

Настоящее изобретение в целом относится к области неинвазивного пренатального скрининга и диагностики. В настоящем изобретении предусмотрен способ выявления наличия или отсутствия анеуплоидий, в частности, трисомии и, более конкретно, трисомии по хромосомам 21, 18 и 13, пола и доли фрагментов ДНК плода по образцу крови, взятого у матери на ранней стадии беременности.

УРОВЕНЬ ТЕХНИКИ

В настоящее время пренатальное тестирование представляет собой неотъемлемый компонент акушерской практики. Основной целью пренатального тестирования является скрининг анеуплоидий плода, таких как трисомия по хромосоме 21 (Т21, синдром Дауна), трисомия по хромосоме 18 (Т18, синдром Эдвардса) и трисомия по хромосоме 13 (Т13, синдром Патау). Хотя в большинстве случаев анеуплоидия у плода приводит к прерыванию беременности во время развития плода, Т21 характеризуется наивысшим уровнем выживаемости, следовательно, пренатальное выявление Т21 считается наиболее важным пренатальным генетическим тестом для целей пренатального скрининга или диагностики. Доступны надежные инвазивные пренатальные тесты, однако из-за их рискованного характера в настоящее время их проводят только в случаях беременности с высокой степенью риска. Разработка надежного способа неинвазивного пренатального тестирования (НИПТ) трисомий плода, по-видимому, является наиболее важной из современных задач наблюдения беременности.

Обнаружение циркулирующей внеклеточной ДНК плода (cffDNA) в крови матери (Lo и соавт., Lancet 350:485-487, 1997) предоставило возможность разработки неинвазивных способов, в которых применяются нуклеиновые кислоты плода из образца периферической крови матери для определения хромосомных аномалий плода. CffDNA составляет примерно менее чем 10% от общей циркулирующей внеклеточной ДНК

(cfDNA) в плазме крови матери, однако недавно было обнаружено, что в крови матери присутствует полный геном плода в форме cffDNA, который, таким образом, представляет собой весьма перспективный материал для NIPT.

В некоторых документах (Chiu и соавт. 2008, Fan и соавт. 2008, Chiu и соавт. 2011, Sehnert и соавт. 2011, Lau и соавт. 2012 и Bianchi и соавт. 2012, EP2183693, EP2366031, US8296076) раскрыты способы, в которых секвенирование ДНК из крови матери применяют для получения информации о количестве aberrантных хромосом у плода. Во всех указанных выше способах применяется анализ общей cfDNA в плазме крови матери без необходимости в выделении специфичной для плода ДНК cffDNA. Эти способы основаны на выявлении дополнительной копии хромосомы для проведения различий между нормальными случаями и случаями трисомии. В случае трисомии у плода количество копий трисомической хромосомы в крови матери является несколько более высоким по сравнению с другими аутосомами. Появление массового параллельного секвенирования ДНК (MPS) позволило проводить секвенирование чрезвычайно больших количеств молекул ДНК и, таким образом, с недавних пор применяется секвенирование ДНК нового поколения (NGS) для неинвазивного выявления трисомии плода по крови матери.

В целом, выявление трисомии плода с применением NGS осуществляют с помощью следующего способа. Вначале проводят секвенирование короткого участка на одном конце каждой молекулы ДНК из плазмы крови матери и картируют его на эталонный человеческий геном, чтобы определить хромосомное происхождение каждой последовательности. Далее количество маркерных последовательностей в хромосоме, представляющей интерес (например, хромосоме 21), сравнивают с некоторой эталонной хромосомой (обычно другой отдельной хромосомой или выбранной группой хромосом), и такое значение сравнивают между исследуемым образцом и эуплоидным контрольным образцом.

Ключевой этап анализа состоит в оценивании доли фрагментов ДНК, происходящих от плода, называемой фетальной фракцией. В случае низкой доли диагностируемый образец может быть ошибочно классифицирован как здоровый на основании показателя обилия фрагментов от матери.

В EP 2183693 (Lo, Y.M.D. и соавт., права переданы Китайскому университету Гонконга, Гонконг) раскрыт способ, в котором соответствующие количества клинически значимой хромосомы и фоновой хромосомы определяют по результатам MPS и рассчитывают процентную представленность последовательностей, картированных на хромосому 21. Было обнаружено, что такая процентная представленность является в значительной степени более высокой в образце, полученном от беременной женщины, вынашивающей плод с трисомией по хромосоме 21, по сравнению с женщиной с нормальным плодом. Было определено, что в этом способе z-оценка не имела решающего значения.

В US8296076 (Fan, Ch. H.M. и Quake, S.R., права переданы Совету попечителей Университета имени Леланда Стэнфорда-младшего, США) раскрыт способ диагностики анеуплоидии плода на основе секвенирования. Путем подсчета количества маркерных последовательностей, картированных в предварительно определенное окно в каждой хромосоме, можно выявить избыточную или недостаточную представленность любой хромосомы в ДНК из плазмы крови матери, вызванную анеуплоидией плода.

Способы согласно Chiu и соавт. 2008, Chiu и соавт. 2011, Sehnert и соавт. 2011, Lau и соавт. 2012 основаны на полногеномном секвенировании с применением MPS и определения z-оценки. Хотя z-оценка является общепринятой в качестве стандартного параметра, используемого для выявления анеуплоидных образцов, существуют различия в его расчете. У Chiu и соавт. 2011 раскрыт подход с применением ридов, картируемых на все хромосомы, применяемые в качестве эталона (за исключением хромосом 13, 18 и 21), для расчета z-оценки, у Lau и соавт. 2012 раскрыт подход с применением ридов, картируемых на некоторую конкретную хромосому, например, 14, в качестве эталона для определения трисомии T21, и Sehnert и соавт. 2011 выбрали хромосому 9 в качестве оптимального внутреннего эталона для хромосомы 21.

В целом существует три основных 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 о последовательностях. В последние несколько лет был достигнут значительный прогресс в секвенировании, однако на первой и третьей стадии имеется простор для улучшения, которое может привести к снижению стоимости и оказать большое влияние на качество тестирования.

Существуют разнообразные подходы для улучшения первой стадии, раскрытые, например, в EP2366031 (Rava, R.P. и соавт., права переданы Verinata Health, Inc., США). В данном документе раскрыт способ пренатального скрининга и диагностики хромосомной анеуплоидии плода на основе NGS, включающий новый протокол для получения библиотек для секвенирования из образца от матери. Новый подход к получению библиотек для секвенирования включает последовательные этапы репарации концов, присоединения dA-хвостов и лигирования адаптеров в указанных нуклеиновых кислотах, и при этом указанные последовательные этапы не включают очистку продуктов репарации концов перед этапом присоединения dA-хвостов и не включают очистку продуктов присоединения dA-хвостов перед этапом лигирования адаптеров. Способ позволяет определить вариации числа копий (CNV) любой последовательности, представляющей интерес. В этом способе z-оценка не имела решающего значения.

Другое улучшение было раскрыто в WO 2011/051283 (Benz, M. и соавт., права переданы Lifecodexx, AG, Германия). Раскрытый в ней способ неинвазивной диагностики хромосомной анеуплоидии улучшен посредством обогащения и количественной оценки выбранных последовательностей cfDNA в образце крови матери.

Поскольку плод мужского пола имеет пары половых хромосом (X и Y), отличающиеся от материнских (две копии X), то фетальная фракция обычно оценивается на основании показателя обилия фрагментов в этих хромосомах. Данный подход, однако, ограничен образцами, где плод имеет мужской пол. Фракция для плода женского пола может быть оценена на основании характеристик, отличающихся между фрагментами от плода и матери.

Различия в локализации фрагментов во всем геноме использовались в способе SeqFF (Kim и соавт. 2015). Они демонстрируют, что фетальные фракции отличаются в различных участках генома и в значительной степени коррелируют с содержанием GC и наличием экзонных участков. Таким образом, в способе используются нормализованные, сгруппированные числовые значения в качестве прогностических переменных и оценивается фетальная фракция с применением стандартных моделей многомерной регрессии.

В способе SANEFALCON (Straver и соавт. 2016) используется более подробная информация о происхождении фрагментов на основании различных механизмов разрушения фрагментов. 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 ДНК, происходящие от плода, как правило, являются более короткими, то профили длин фрагментов от матери и плода в значительной степени различаются, и их можно использовать в качестве прогностического параметра фетальной фракции. Способ, основанный на длине фрагментов, был предложен Yu и соавт. 2014.

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 и соавт. 2016) используется эта информация, однако необходим еще один лабораторный анализ для создания карты генома родителей. Применение данного способа, таким образом, является слишком времязатратным и дорогостоящим для повседневной диагностики.

Несмотря на существование нескольких способов неинвазивного выявления анеуплоидии плода, все еще существует потребность в альтернативном способе, который был бы по меньшей мере таким же чувствительным и специфичным, как и существующие на сегодняшний день способы, и в то же время менее затратным. Кроме того, существующие в настоящее время способы прогнозирования по фетальной фракции все еще ограничены их точностью или необходимостью в дополнительном лабораторном анализе.

В настоящем изобретении предусмотрен альтернативный и надежный способ, который применим для практической реализации неинвазивного пренатального скрининга анеуплоидий, таких как трисомия по хромосоме 13, 18 или 21. В нем предусмотрено

одновременное выявление анеуплоидии, пола и фетальной фракции с применением единой модели, которую можно легко расширить на другие хромосомные анеуплоидии и нарушения, таких как вариации числа копий. В модели используются различные источники информации о происхождении фрагментов ДНК, такие как локализация в хромосоме и длина фрагмента ДНК, и модель можно постепенно расширить для включения других источников для достижения еще большей точности.

КРАТКОЕ ОПИСАНИЕ ИЗОБРЕТЕНИЯ

В следующем описании будут разъяснены основные признаки способа по настоящему изобретению, однако это не означает, что изобретение должно включать все признаки и аспекты, описанные в данном документе. Настоящее изобретение станет полностью понятно для специалиста в данной области после изучения следующего описания вместе с примерами, где 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 специфические термины разъяснены ниже.

Термин «анеуплоидия» используется в общем смысле, известном специалисту в данной области, и он означает дисбаланс генетического материала, обусловленный утратой или приобретением целой хромосомы или части хромосомы. Другими словами, он означает наличие целой избыточной хромосомы, или отсутствие полной хромосомы, или частичную хромосомную дупликацию или делецию значительного размера (> 1 т. п. о.). Конкретно, *T21*, *T18* и *T13* обозначают наиболее распространенные типы аутомных 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.

Термин «массовое параллельное секвенирование» (MPS), используемый в комбинации с «секвенированием нового поколения» (NGS), означает широко известные специалистам в данной области методики секвенирования огромного количества, т. е. порядка миллионов или десятков миллионов, фрагментов ДНК, осуществляемые на практике в

данной области техники с применением анализаторов Illumina или IonTorrent.

Протоколы полногеномного секвенирования известны специалисту в данной области.

Последовательности *ридов* представляют собой короткие последовательности ДНК, получаемые в результате NGS-секвенирования, которые, однако, являются достаточно длинными (например, длиной по меньшей мере приблизительно 30 п. о.), чтобы служить в качестве маркерных последовательностей, т. е. их можно однозначно отнести к одной 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 картируются на эталонный геном, т. е. их относят к одному месту локализации в эталонном геноме. Метки, которые могут быть картированы в более чем одно место локализации 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 можно осуществлять несколькими путями, и авторы настоящего изобретения применяли способ согласно Liu и соавт. 2014. Используемый в данном документе термин «*эталонный человеческий геном*» или «*эталонный геном*» относится к последовательности hg19.

Используемые 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 можно осуществлять вручную или с помощью компьютерных алгоритмов, которые широко известны специалистам в области молекулярной биологии и биоинформатики. Совпадение последовательности рида при выравнивании может представлять собой 100% совпадение последовательности или менее чем 100% (неабсолютное совпадение).

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

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 сосредотачиваются на измерении показателя обилия фрагментов в геноме, такого как оценка числа 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 приведенные в Benjamini and Speed 2012 или Liao и соавт. 2014.

Байесовская сеть является вероятностной моделью, которая представляет совокупность случайных переменных и их условных зависимостей с помощью направленного ациклического графа.

Апостериорная вероятность случайного события означает условную вероятность, которая присваивается после принятия во внимания соответствующего фактического материала или фонового значения. Аналогичным образом, *распределение апостериорных вероятностей* представляет собой распределение вероятностей неизвестной величины, рассматриваемое в качестве случайной переменной, при условии наличия фактического материала, полученного в результате эксперимента или опроса. «Апостериорный» в этом контексте означает после принятия во внимание соответствующего фактического материала, связанного с конкретным исследуемым случаем.

Алгоритм максимизации ожидания (EM) представляет собой итерационный способ, позволяющий находить оценку максимального правдоподобия или оценку апостериорного максимума (MAP) параметра в статистических моделях, где модель зависит от ненаблюдаемых скрытых переменных.

Настоящее изобретение относится к способу определения анеуплоидии и пола у плода по образцу крови, плазмы крови или сыворотки крови матери, содержащему смесь молекул нуклеиновых кислот плода и матери, где смесь молекул нуклеиновых кислот плода и матери представляет собой молекулы внеклеточной ДНК, при этом указанный способ включает четыре основные стадии: 1) получение и обработку образцов крови матери, 2) получение образца ДНК и библиотеки ДНК, 3) секвенирование и 4) анализ данных о последовательностях с получением прогноза. Важной частью способа является получение обучающих данных, т. е. анализ набора эуплоидных образцов, которые обрабатывают тем же способом, что и исследуемые образцы. Минимальный

обучающий набор образцов должен содержать мать со здоровым, т. е., эуплоидным, плодом мужского пола и мать со здоровым, т. е., эуплоидным, плодом женского пола, предпочтительно в количестве приблизительно ста в каждой из этих категорий.

Отправной точкой способа по настоящему изобретению является анализ образца от матери, т. е. периферической крови, поэтому способ является неинвазивным. Периферическая кровь содержит циркулирующую внеклеточную ДНК (cfDNA), которая представляет собой смесь фрагментов ДНК матери и ДНК плода (cffDNA), далее называемых также фрагментами ДНК или, вкратце, фрагментами. Значение имеет CffDNA, следовательно, ДНК плода можно обогатить (посредством отбора более коротких фрагментов в условиях *in silico* либо путем физического отбора). Затем образец общей ДНК (по-прежнему содержащий ДНК как матери, так и плода, т. е. это смешанный образец) подвергают массовому параллельному секвенированию с помощью подхода NGS с получением огромного количества коротких последовательностей ридов. Эти риды служат в качестве маркерных последовательностей, т. е. их картируют на определенный участок генома или хромосому.

В способе по настоящему изобретению назначают апостериорную вероятность происхождения от плода каждому фрагменту ДНК с применением различных отличительных признаков. В данном случае авторы настоящего изобретения задействовали длину фрагмента и положение в хромосоме в качестве выбранных признаков фрагментов ДНК, однако не вызывает затруднений итерационное расширение модели на другие типы признаков (например, расстояние от фрагмента до ближайшего кодирующего участка, его содержание GC и/или информацию об однонуклеотидных полиморфизмах), что делает ее более точной. Фрагменты, с высокой вероятностью происходящие от плода, оказывают больший эффект на результат тестирования, чем фрагменты с низкой вероятностью этого.

Основное улучшение и преимущество способа по настоящему изобретению заключается в этапе анализа и обработки данных о последовательностях. Прогнозирование пола, эуплоидии/анеуплоидии и фетальной фракции является в высокой степени взаимозависимым, однако в описанных существующих способах эти данные обрабатываются независимо, что может приводить к их несоответствиям. В отличие от этого, способ по настоящему изобретению позволяет обрабатывать свойства

наблюдаемых фрагментов cfDNA одновременно, что приводит к более надежному, унифицированному прогнозированию.

Байесовская сеть (называемая также байесовской моделью), которая является основой способа по настоящему изобретению, позволяет естественным образом разрабатывать способ выявления других геномных аномалий, таких как абберрации половых хромосом, микроабберрации и нестандартные анеуплоидии, на основе положения, диапазона и типа аномалии, определенных специалистом в данной области.

Один аспект настоящего изобретения относится к способу определения анеуплоидии и пола у плода по образцу крови, плазмы крови или сыворотки крови матери наряду с фетальной фракцией, как подробно объясняется ниже и как определено в прилагаемой формуле изобретения.

В настоящее время существует практическая необходимость в автоматизации способов пренатальных тестирований 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 последовательностей ридов (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, пользователь (потребитель) может загрузить всю компьютерную программу с помощью Интернета.

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 (пол, наличие/отсутствие анеуплоидии, фетальную фракцию).

Еще один аспект настоящего изобретения относится к компьютерному программному продукту, содержащему 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. Указанную выше компьютерную программу можно преимущественно установить на компьютер системы секвенирования.

Объект настоящего изобретения определен в прилагаемой формуле изобретения.

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

КРАТКОЕ ОПИСАНИЕ ГРАФИЧЕСКИХ МАТЕРИАЛОВ

Фиг. 1. Общая схема способа, на которой продемонстрированы основные этапы и итерационный характер обработки данных, полученных в результате секвенирования.

Фиг. 2. Схема байесовской сети (байесовской модели).

Фигура 3. Оценочные распределения по длинам фрагментов для фрагментов ДНК матери и плода. Демонстрируется эмпирическое распределение для секвенированного образца с оценочной фетальной фракцией 25,6%.

Фигура 4. Сравнение величин фетальной фракции, оцениваемых на основании показателя обилия в эталонных хромосомах и с помощью байесовского метода в соответствии с настоящим изобретением.

ПОДРОБНОЕ ОПИСАНИЕ ИЗОБРЕТЕНИЯ

1. СПОСОБ

В этом новом способе авторы настоящего изобретения рассматривают фрагменты cfDNA (фрагменты cfDNA или, вкратце, фрагменты), секвенированные из крови матери (или плазмы крови, или сыворотки крови), в качестве смеси фрагментов cfDNA, происходящих от плода и матери. Авторы настоящего изобретения исследуют несколько признаков этих фрагментов, которые по-разному распределены у матери и плода, и авторы настоящего изобретения используют эти различия для определения наиболее вероятного состава этой смеси.

Для определения этого состава авторы настоящего изобретения используют алгоритм максимизации ожидания (EM). Байесовская сеть является вероятностной графической моделью, в которой представлен набор случайных переменных, которые можно наблюдать (длина фрагмента и положение фрагмента в хромосоме), что позволяет прогнозировать апостериорную вероятность для ненаблюдаемой переменной (происхождения фрагмента от плода). Авторы настоящего изобретения разработали байесовскую сеть для описания смеси фрагментов cfDNA плода и матери в плазме крови матери с несколькими признаками, каждый из которых характеризуется различным распределением значений в зависимости от происхождения фрагмента (т. е. от плода или матери). Распределения по хромосоме, являющейся источником происхождения (= положению фрагмента в хромосоме или локализации фрагмента), и длине фрагмента для образцов от плода и матери вначале выводят в результате обучения из набора обучающих образцов и затем корректируют с помощью алгоритма EM. Результатом применения нового способа является классификация исследуемого образца как относящегося к одной из четырех категорий, таких как эуплоиды мужского пола, анеуплоиды мужского пола, эуплоиды женского пола, анеуплоиды женского пола,

т. е. к наиболее вероятному классу для случая с наиболее вероятной фетальной фракцией фрагментов cfDNA с учетом данных NGS. Дополнительно, способ можно расширять для включения в рассмотрение также близнецов из двойни или тройни.

Авторы настоящего изобретения отмечают, что способ может включать дополнительные признаки фрагментов (такие как расстояние от фрагмента до ближайшего кодирующего участка, его содержание GC и т. д.) аналогичным образом, и рассматриваемые хромосомные aberrации охватывают все аутосомы, а также половые хромосомы.

Более конкретно, способ состоит из следующих этапов. Для каждого случая, например, выявления одной конкретной трисомии (например, T21) и каждой возможной категории (анеуплоиды мужского пола, анеуплоиды женского пола, эуплоиды мужского пола, эуплоиды женского пола), осуществляют следующее.

1. Сбор и обработка обучающих образцов.
2. Выведение в результате обучения исходных распределений, касающихся длины фрагмента и хромосомы, являющейся источником происхождения.
3. Построение байесовской сети, которая включает признаки фрагмента: длину фрагмента (LEN) и хромосому, являющуюся источником происхождения (CHROM).
4. Прогон алгоритма EM в этой байесовской сети по обучающим образцам.
5. На основании результата EM:
 - a. классификация фрагментов из образца как происходящих от матери или плода;
 - b. корректировка распределений по хромосомам, являющимся источниками происхождения, и длинам фрагментов.
6. Повторение этапов 4 и 5 до тех пор, пока параметры выведенного в результате обучения распределения не будут изменяться между итерациями незначительно, например, пока сумма различий не будет меньше 0,000001.

7. Расчет вероятности для выбранных признаков фрагментов ДНК в конечной байесовской сети.

После расчета апостериорных вероятностей во всех категориях (анеуплоиды мужского пола, анеуплоиды женского пола, эуплоиды мужского пола, эуплоиды женского пола) выбирается наиболее вероятный случай в качестве результата тестирования или диагноза.

1.1 ОТБОР И ОБРАБОТКА ОБУЧАЮЩИХ ОБРАЗЦОВ

Следующие обучающие наборы необходимы для применения способа по настоящему изобретению для выявления трисомий T13, T18 и T21. Образцы из всех наборов секвенируют, и полученные в результате файлы fastq картируют на человеческий геном. После картирования необязательно применяют поправку на сдвиг содержания GC.

1. Отбирают обучающий набор от беременных женщин с одним эуплоидным плодом женского пола, обозначают его как T_{hx} .
2. Отбирают обучающий набор от беременных женщин с одним эуплоидным плодом мужского пола, обозначают его как T_{hy} .
3. Отбирают обучающий набор от беременных женщин с одним плодом женского пола с трисомией T13, T18 и T21, обозначают его как T_{113x} , T_{118x} и T_{121x} соответственно.
4. Отбирают обучающий набор от беременных женщин с одним плодом мужского пола с трисомией T13, T18 и T21, обозначают его как T_{113y} , T_{118y} и T_{121y} соответственно.
5. Определяют обучающий набор с эуплоидией как $T_h = T_{hx} \cup T_{hy}$.
6. Отбирают обучающий набор из взрослых мужчин, обозначают его как T_m .
7. Отбирают обучающий набор из небеременных женщин, обозначают его как T_f .
8. Для каждого образца из каждого обучающего набора применяют следующие этапы.

- a. Необработанные последовательности картируют на находящийся в открытом доступе эталонный человеческий геном (hg19) с помощью алгоритма Bowtie2 (Langmead and Salzberg, 2012).
- b. Оставляют только те последовательности ридов, которые можно картировать ровно в одно место локализации в геноме с не более чем одним несовпадением (Ehrich et al., 2011; Sehnert et al., 2011).
- c. Эталонный геном разделяют на последовательные и неперекрывающиеся секции размером 20 т. о., и картированные последовательности ридов распределяют по указанным секциям в соответствии с их картированной локализацией.
- d. В количество ридов необязательно вносят поправку на сдвиг содержания GC в соответствии с поправкой с помощью LOESS и PCA (Zhao 2015).

1.2 ПОСТРОЕНИЕ РАСПРЕДЕЛЕНИЙ ПРИЗНАКОВ ФРАГМЕНТОВ

1.2.1 Хромосомы

Можно наблюдать, что существуют различия между распределениями фрагментов cfDNA, картированных на хромосомы, для беременных женщин с плодом мужского и женского пола (наиболее заметно по X- и Y-хромосомам). Аналогичным образом, существуют различия в таких распределениях для эуплоидных и анеуплоидных плодов (например, по хромосоме 21 для плода с трисомией T21). Эти наблюдения можно использовать для определения наиболее вероятных источников смеси cfDNA в плазме крови матери, иными словами, того, вынашивает ли женщина плод мужского или женского пола, и того, является ли он здоровым или aberrантным (например, анеуплоидным, например, с трисомией), посредством рассмотрения распределения фрагментов cfDNA в смеси.

Предположение 1.1. *Предполагается, что распределение фрагментов cfDNA из плазмы крови матери в аутосомах является одинаковым как для матери, так и для эуплоидного плода. Если, дополнительно, плод является девочкой, эта одинаковость распространяется на все хромосомы. Если плод является мальчиком, распределение в X- и Y- хромосомах будет отличаться от распределения у матери.*

Предположение 1.2. Предполагается, что распределение фрагментов cfDNA из плазмы крови матери в хромосомах эуплоидного плода мужского пола и распределение у здорового мужчины являются одинаковыми.

Этапы

1. Расчет фракционной аутомной представленности (FAR) для каждого образца в наборе T_h (см. *Расчет FAR* ниже).
2. Расчет фракционной хромосомной представленности (FCR) для каждого образца в наборах T_m и T_{hx} (см. *Расчет FCR*).
3. Расчет среднего значения FAR для каждой аутосомы (см. *Расчет среднего значения FAR для аутосом*) в T_h .
4. Расчет среднего значения FCR для каждой хромосомы (см. *Расчет среднего значения FCR для хромосом*) в T_m и T_{hx} .
5. Построение распределений по хромосомам CHR-DIST (см. *Построение распределения по хромосомам*).

Расчет FAR

Пусть $M = (m_{i,j})_{k \times 22}$ представляет собой матрицу, связанную с обучающим набором T_h , где k означает размер T_h , и $m_{i,j}$ содержит количество фрагментов, картированных (с поправкой на сдвиг содержания GC) на j^{io} аутосому в i^M образце из T_h . Пусть M_i обозначает сумму элементов $i^{\tilde{i}}$ строки матрицы M . Значения FAR для каждой аутосомы в i^M образце (т. е. в $i^{\tilde{i}}$ строке матрицы M) определяются как $m_{i,x}/M_i$, где x обозначает номер аутосомы (т. е. $x \in \{1, 2, \dots, 22\}$).

Матрица значений FAR для образцов из обучающего набора T_h обозначается авторами настоящего изобретения как F_h .

Расчет FCR

Пусть $N = (n_{i,j})_{k \times 24}$ представляет собой матрицу, связанную с обучающим набором T_m , где k означает размер T_m , и $n_{i,j}$ содержит количество фрагментов, картированных (с поправкой на сдвиг содержания GC) на j^{io} хромосому в i^M образце из T_m (авторы

настоящего изобретения допускают, что X-хромосома имеет номер 23, а Y-хромосома имеет номер 24). Пусть N_i обозначает сумму элементов $i^{\text{-й}}$ строки матрицы N . Значения FCR для каждой хромосомы в $i^{\text{-м}}$ образце (т. е. в $i^{\text{-й}}$ строке матрицы N) определяются как $n_{i,x}/N_i$, где x обозначает номер аутосомы (т. е. $x \in \{1, 2, \dots, 24\}$).

Матрица значений FCR для образцов из обучающего набора T_m обозначается авторами настоящего изобретения как F_m . Значения FCR для образцов из обучающего набора T_{hx} можно рассчитать аналогично, и они обозначаются авторами настоящего изобретения как F_{hx} .

Расчет среднего значения FAR для аутосом

Пусть $F_h = (f_{i,j})_{k \times 22}$ представляет собой матрицу, связанную со значениями FAR для обучающего набора T_h , где k означает размер T_h , и $f_{i,j}$ содержит значение FAR для $j^{\text{-й}}$ аутосомы в $i^{\text{-м}}$ образце из T_h . Среднее значение FAR для $j^{\text{-й}}$ аутосомы определяется как среднее значение для $j^{\text{-го}}$ столбца матрицы F_h .

Средние значения FAR для набора T_h обозначаются авторами настоящего изобретения как FAR_j , где $j \in \{1, 2, \dots, 22\}$ означает номер аутосомы.

Расчет среднего значения FCR для хромосом

Пусть $F_m = (f_{i,j})_{k \times 24}$ представляет собой матрицу, связанную со значениями FCR для обучающего набора T_m , где k означает размер T_m , и $f_{i,j}$ содержит значение FCR для $j^{\text{-й}}$ хромосомы в $i^{\text{-м}}$ образце из T_m . Среднее значение FCR для $j^{\text{-й}}$ хромосомы определяется как среднее значение для $j^{\text{-го}}$ столбца матрицы F_m . Это аналогично справедливо для матрицы F_{hx} .

Средние значения FCR для наборов T_m и T_{hx} обозначаются авторами настоящего изобретения как FCR_j^m и FCR_j^{hx} соответственно, где $j \in \{1, 2, \dots, 24\}$ означает номер хромосомы.

1.2.2 Построение распределений по хромосомам

В модели по настоящему изобретению огромное количество распределений по хромосомам можно определять на основе природы генетического нарушения, которое было бы желательно наблюдать. Внимание авторов настоящего изобретения

предпочтительно сосредоточено на анеуплоидиях по отдельным целым аутосомам (например, T13, T18, T21), но модель может охватывать анеуплоидии по нескольким аутосомам, а также частям этих аутосом. В целях полноты изложения материала построение распределения по хромосомам описывается авторами настоящего изобретения в наиболее общем случае.

Распределение по хромосомам состоит из 24 чисел по одному для каждой хромосомы и описывает вероятность того, что некоторый фрагмент cfDNA будет картирован на конкретную хромосому.

Пусть $FAR = (FAR_1, FAR_2, \dots, FAR_{22})$ означает средние значения FAR для обучающего набора T_h . Пусть $FCR^m = (FCR_1^m, FCR_2^m, \dots, FCR_{24}^m)$ и $FCR^{hx} = (FCR_1^{hx}, FCR_2^{hx}, \dots, FCR_{24}^{hx})$ означают средние значения FCR для обучающих наборов T_m и T_{hx} соответственно.

Распределение по хромосомам как для здоровой женщины (т. е. самой женщины), так и для здорового плода женского пола задается с помощью FCR^{hx} (см. Предположение 1.1 выше). В следующем тексте авторы настоящего изобретения используют символы $CHR - DIST_{x, h}$ и $CHR - DIST_{f, h}$ для распределения по хромосомам у здорового плода женского пола и самой здоровой женщины соответственно. Распределение по хромосомам у здорового плода мужского пола рассчитывают следующим образом. Сначала определяют нормирующую постоянную NK как

$$NK = \sum_{j=1}^{22} FCR_j^m.$$

Пусть

$$t_{23} = \frac{FCR_{23}^m}{NK},$$

$$t_{24} = \frac{FCR_{24}^m}{NK}$$

и

$$d = (FAR_1, FAR_2, \dots, FAR_{22}, t_{23}, t_{24}).$$

При условии, что s обозначает сумму элементов вектора d , распределение у здорового плода мужского пола задается с помощью

$$D = \left(\frac{FAR_1}{s}, \frac{FAR_2}{s}, \dots, \frac{FAR_{22}}{s}, \frac{t_{23}}{s}, \frac{t_{21}}{s} \right).$$

В следующем тексте авторы настоящего изобретения используют символ $CHR - DIST_{y,h}$ для распределения по хромосомам у здорового плода мужского пола.

Построение распределения для хромосомных aberrаций (например, анеуплоидий) осуществляется достаточно просто. Сначала выбирают мишень для aberrации, т. е. мать или плод. Затем выбирают пораженные аутосомы. Наконец, для каждой пораженной аутосомы выбирают степень aberrации, указывающую на то, насколько большая доля пораженной аутосомы утрачивается или мультиплицируется. Таким образом, пусть $t \in \{f, x, y\}$ означает мишень (здесь уже проводятся различия между плодами мужского и женского пола). Кроме того, пусть $a = \mathbb{R}^{24}$, $a = (a_1, a_2, \dots, a_{24})$ представляет собой вектор, указывающий на степень аутосомных aberrаций ($a_i = 0$, если $i^{\text{я}}$ аутосома не поражена, $a_i = -2$, если обе копии $i^{\text{я}}$ аутосомы утрачиваются, $a_i = 1$, если $i^{\text{я}}$ аутосома является трисомической, и $a_i = -1$, если $i^{\text{я}}$ аутосома является моносомической). При условии, что

$$CHR - DIST_{t,h} = (ch_1, ch_2, \dots, ch_{24}),$$

распределение по aberrантным хромосомам задается с помощью

$$CHR - DIST_{t,a} = \left(\frac{ch_1 \left(1 + \frac{a_1}{2}\right)}{s}, \frac{ch_2 \left(1 + \frac{a_2}{2}\right)}{s}, \dots, \frac{ch_{24} \left(1 + \frac{a_{24}}{2}\right)}{s} \right).$$

где

$$s = ch_1 \left(1 + \frac{a_1}{2}\right) + ch_2 \left(1 + \frac{a_2}{2}\right) + \dots + ch_{24} \left(1 + \frac{a_{24}}{2}\right).$$

1.2.3. Длины фрагментов

Предположение 1.3. Наблюдалось (Stephanie и соавт., 2014), что фрагменты cfDNA плода в среднем являются более короткими, чем фрагменты cfDNA матери, иными

словами, длины фрагментов от матери распределяются иначе, чем длины фрагментов от плода.

Предположение 1.4. Предполагается, что распределение по длинам фрагментов cfDNA, картированных на аутосомы, является одинаковым для зуплоидного плода мужского и женского пола. Эти два распределения, однако, не являются одинаковыми, если рассматриваются все хромосомы. Существует небольшое различие в связи с тем, что две X-хромосомы плода женского пола дают более короткие риды, чем X- и Y-хромосомы плода мужского пола. Таким образом, ожидается, что образец от беременной женщины с плодом женского пола имеет более короткие риды, чем от беременной женщины с плодом мужского пола.

Предположение 1.5. Предполагается, что распределение по длинам фрагментов cfDNA, картированных на аутосомы, является одинаковым для беременных и небеременных женщин (в данном случае авторы настоящего изобретения имеют в виду беременных женщин как таковых, а не в качестве смеси cfDNA плода и матери).

Предположение 1.6. Предполагается, что секвенированные фрагменты имеют длину не более 1000 п. о.

Этапы

1. Расчет фракционного распределения длин (FLD) для образцов из обучающего набора T_f (см. Расчет FLD для T_f ниже).
2. Расчет фракционного распределения по длинам фрагментов, картированных на Y-хромосому, для образцов из обучающего набора T_{hy} (см. Расчет FLD для T_{hy}).
3. Построение распределения по длинам фрагментов для 0% фетальной фракции $LEN - DIST_0$ (см. Построение $LEN - DIST_0$).
4. Построение распределения по длинам фрагментов для 100% фетальной фракции $LEN - DIST_1$ (см. Построение $LEN - DIST_1$).

Следует отметить, что в этих распределениях не проводятся различия между плодами мужского и женского пола, а также зуплоидными и аберрантными плодами. Авторы настоящего изобретения отмечают, что плод женского пола с трисомией будет иметь в

среднем более короткие риды, чем эуплоидный плод мужского пола. Однако, на данном этапе авторы настоящего изобретения решают пренебречь этими различиями.

Расчет FLD для T_f

Пусть $M = (m_{i,j})_{k \times 1000}$ представляет собой матрицу, связанную с обучающим набором T_f , где k означает размер T_f , и $m_{i,j}$ содержит количество картированных (с поправкой на сдвиг содержания GC) фрагментов длиной j для $i^{\text{го}}$ образца из T_f . Пусть M_i обозначает сумму элементов $i^{\text{й}}$ строки матрицы M . Затем фракционное распределение длин $i^{\text{го}}$ набора задается с помощью

$$\left(\frac{m_{i,1}}{M_i}, \frac{m_{i,2}}{M_i}, \dots, \frac{m_{i,1000}}{M_i} \right).$$

Матрица значений FLD для образцов из обучающего набора T_f обозначается авторами настоящего изобретения как L_f .

Расчет FLD для T_{hx}

Этот этап является в сущности идентичным расчету FLD для T_f , с той лишь разницей, что обучающий набор представляет собой T_{hy} , и что интерес представляют только фрагменты, картированные на Y-хромосому.

Матрица значений FLD для образцов из обучающего набора T_{hy} обозначается авторами настоящего изобретения как L_{hy} .

Построение $LEN - DIST_0$

Пусть $L_f = (l_{i,j})_{k \times 1000}$ представляет собой матрицу значений FLD для обучающего набора T_f , где k означает размер T_f . Пусть $L_j, j \in \{1, 2, \dots, 1000\}$, обозначает среднее значение для $j^{\text{го}}$ столбца матрицы L_f . Затем распределение по длинам фрагментов для 0% фетальной фракции задается с помощью

$$LEN - DIST_0 = (L_1, L_2, \dots, L_{1000}).$$

Построение $LEN - DIST_1$

Пусть $L_{hy} = (l_{i,j})_{k \times 1000}$ представляет собой матрицу значений FLD для обучающего набора T_f , где k означает размер T_{hy} . Пусть $K_j, j \in \{1, 2, \dots, 1000\}$, обозначает среднее значение для j^{20} столбца матрицы L_{hy} . Затем распределение по длинам фрагментов для 100% фетальной фракции задается с помощью

$$LEN - DIST_1 = (K_1, K_2, \dots, K_{1000}).$$

1.3 БАЙЕСОВСКАЯ СЕТЬ И АЛГОРИТМ EM

В сети, описывающей смесь фрагментов cfDNA от плода и матери, исследуются длина фрагмента и хромосома, являющаяся источником его происхождения (байесовская сеть, используемая для этой цели, показана на фиг. 2).

Следует отметить, что фактически имеются четыре байесовские сети, которые отличаются по значениям распределений CHROM и LEN для классов, определенных как анеуплоиды мужского пола, анеуплоиды женского пола, эуплоиды мужского пола и эуплоиды женского пола.

Под $CHR - DIST_{\{x,y\},a}$ авторы настоящего изобретения подразумевают то, что используется $CHR - DIST_{x,a}$ для плода женского пола (с аутосомными aberrациями, заданными с помощью a) либо используется $CHR - DIST_{y,a}$ для плода мужского пола (с аутосомными aberrациями, заданными с помощью a).

В этой байесовской сети прогоняется алгоритм EM. Наблюдаемые переменные представляют собой длину фрагмента и хромосому, являющуюся источником происхождения, для всех фрагментов в образце. Происхождение фрагмента (мать, плод) представляет собой одну скрытую или ненаблюдаемую переменную, которую прогнозируют с помощью логического вывода с использованием EM. Параметры вероятностной модели представляют собой $f, LEN - DIST_0, LEN - DIST_1, CHR - DIST_{f,h}$ и $CHR - DIST_{\{x,y\},a}$.

Пусть \mathbf{x} означает все наблюдаемые значения для всех фрагментов в образце, пусть Z обозначает скрытую переменную для всех фрагментов, и пусть θ означает все параметры для вероятностной модели. Затем алгоритм EM можно обобщить данным уравнением:

$$\theta^{(i+1)} = \arg \max_{\theta} \sum_z P(Z = z | x, \theta^{(i)}) L(x, Z = z | \theta^{(i)}), \quad (1.1)$$

где P означает вероятность, а L означает логарифмическое правдоподобие. Более конкретно, пусть $\theta^{(0)}$ означает набор исходных параметров, определенных согласно описанному в разделе 1.2. Также изначально устанавливается $f^{(0)} = 0,5$. Для следующей итерации имеется:

$$f^{(1)} = \sum_{j=1}^N Pr[\text{происхождение фрагмента} = \text{плод} | \text{хромосома} = ch_j, \text{длина} = l_j] / N,$$

где N означает общее количество ридов. Эти показатели вероятности можно вычислить из байесовской сети. Если предположить, что длина фрагмента не зависит от хромосомы, из этого следует, что

$$f^{(1)} = \frac{1}{N} \sum_{j=1}^N \frac{Pr[\text{хромосома} = ch_j | \text{плод}] Pr[\text{длина} = l_j | \text{плод}] Pr[\text{плод}]}{\sum_{\sigma_i \in \{\text{плод}, \text{мать}\}} Pr[\text{хромосома} = ch_j | \text{происх}] Pr[\text{длина} = l_j | \text{происх}] Pr[\text{происх}]}$$

Показатели вероятности $Pr[\text{хромосома} = ch_j | \text{плод}]$ и $Pr[\text{длина} = l_j | \text{плод}]$ в первой итерации можно получать из предварительно выведенных в результате обучения распределений LEN-DIST и CHR-DIST (значения $CHR-DIST_{\{x,y\},a}[ch_j]$ и $LEN-DIST_l[l_j]$). Это аналогично справедливо для значений $Pr[\text{хромосома} = ch_j | \text{мать}]$ и $Pr[\text{длина} = l_j | \text{мать}]$. Наконец, $Pr[\text{плод}] = f^{(0)}$ и $Pr[\text{мать}] = 1 - f^{(0)}$ в этой итерации.

Аналогичным образом можно найти новые ожидаемые значения для векторов $LEN-DIST_0$, $LEN-DIST_l$, $CHR-DIST_{f,h}$ и $CHR-DIST_{\{x,y\},a}$. Например,

$$LEN-DIST_0[1]^{(1)} = \sum_{j: len_j=1} Pr[\text{происхождение фрагмента} = \text{плод} | \text{длина} = 1, \text{хромосома} = ch_j]$$

значение которого можно рассчитать аналогично расчету $f^{(1)}$.

Уравнения чередуются итерационным способом. Во-первых, для каждого фрагмента рассчитывают вероятность происхождения от плода на основе текущих параметров распределений. Во-вторых, по новым показателям вероятности происхождения от плода рассчитывают параметры модели. Затем первый этап выполняют снова с новыми

оценочными значениями параметров, второй этап с улучшением параметров и так далее. Каждая итерация непрерывно улучшает вероятность для параметров модели и наблюдаемых данных. Итерации прекращают, когда изменение параметров становится меньшим, чем некоторая заданная постоянная, например, 0,000001.

1.4 ИНТЕРПРЕТАЦИЯ РЕЗУЛЬТАТОВ

После окончания выполнения алгоритма EM в конкретной байесовской сети можно рассчитать апостериорную вероятность для наблюдаемых данных. Это делается для всех четырех байесовских сетей (следует напомнить, что имеется байесовская сеть для четырех категорий: анеуплоиды мужского пола, анеуплоиды женского пола, эуплоиды мужского пола, эуплоиды женского пола; эти сети отличаются не по структуре, а по установленным исходным параметрам, таким как *LEN – DIST* и *CHR-DIST*. Наконец, для исследуемого случая выбирают наиболее вероятный класс среди этих четырех в качестве результата выполнения способа.

1.5 ФЕТАЛЬНАЯ ФРАКЦИЯ

Фетальную фракцию cfDNA можно непосредственно определить из скорректированной(скорректированных) с помощью EM байесовской(байесовских) сети(сетей) (параметр сети f), выбранной в качестве наиболее вероятного случая.

ПРИМЕРЫ

Пример 1

Методики секвенирования, применяемые при реализации настоящего изобретения

В этом конкретном примере представлен предпочтительный вариант осуществления получения образца и секвенирования. Специалисту в данной области известен тот факт, что большинство описанных далее этапов можно модифицировать с учетом различных факторов, например, реагентов, применяемых в способе, установившейся лабораторной практики и конечной цели анализа.

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 вместе 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 устанавливаемые 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 оценивания с помощью способа.

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

Полный способ получения данных о последовательностях описывается ниже со всеми лабораторными протоколами для конкретных этапов способа. Полный способ, описанный в данном документе, в значительной степени известен специалистам в области молекулярной биологии, биоинформатики и пренатального тестирования, и такому специалисту известны возможные модификации процедуры.

ОТБОР ОБРАЗЦОВ ПЕРИФЕРИЧЕСКОЙ КРОВИ

- a) Образец периферической венозной крови матери объемом 10 мл отбирают в специальную пробирку для отбора образцов и стабилизации ccfDNA крови PAXgene® (PreAnalytiX GmbH, Feldbachstrasse, 8634, Хомбрехтikon, Швейцария) или другую подобную.
- b) Инструкция по взятию крови.

- i) Кровь для целей тестирования необходимо предпочтительно отбирать с 12-й недели беременности (не раньше, чем закончится 10-я неделя беременности).
 - ii) Образец должен отбираться медицинским персоналом женского пола.
 - iii) Для отбора образцов необходимо использовать пробирку для отбора образцов, указанную выше; запрещается добавлять любые дополнительные вещества или осуществлять любые другие вмешательства.
 - iv) Во время отбора образцов пробирка должна быть заполнена в течение двух минут. Переполнение или недостаточное наполнение пробирки приводит к неверным результатам анализа.
 - v) Учитывая содержание химических веществ в пробирке для отбора образцов, необходимо принять дополнительные меры во время отбора венозной крови для предотвращения обратного всасывания содержимого пробирки в кровеносную систему пациента.
 - vi) Пробирку для отбора образцов необходимо закрывать непосредственно после взятия крови, и содержимое пробирки необходимо перемешивать посредством плавного вращения от восьми до десяти раз. Ненадлежащее и несвоевременное вращение может приводить к неверным или неинформативным результатам. Содержимое пробирки не следует перемешивать быстро или вихревым способом. Одно вращение означает полное вращение на 180° и обратно.
- c) Транспортировка и хранение образца:
- i) Образец необходимо хранить и транспортировать защищенным от повреждений при температуре от 6°C до 37°C перед дальнейшей обработкой.

ПЛАЗМОСЕПАРАЦИЯ И ВЫДЕЛЕНИЕ ДНК

Лабораторные процедуры

1. Плазмосепарация

- i) Необходимо обеспечить среду, не загрязненную человеческой мужской ДНК, с помощью мер деконтаминации (ПЦР-бокса с УФ-облучением).
Рекомендуется отдельная комната для выделения плазмы крови для эффективной обработки образцов.
- ii) Рекомендуется выполнение всех этапов персоналом женского пола (способ основан на подсчете Y-хромосом с помощью нескольких алгоритмов и, таким образом, он является чувствительным к загрязнению мужской ДНК).
- iii) Необходимо использовать одноразовые наконечники на каждом этапе переливания пипеткой.
- iv) Необходимо отсканировать этикетку со штрих-кодом для цельной крови после доставки во время обработки.
- v) Кровь проверяют визуально на наличие признаков гемолиза.
- vi) Цельную кровь центрифугируют при 2200 x g/8 мин
- vii) Плазму крови осторожно переносят в новую пробирку Eppendorf объемом 2 мл (DNA LoBind).
- viii) Плазму крови центрифугируют при 20000 x g/8 мин.
- ix) Плазму крови осторожно переносят в пробирки Eppendorf объемом 1,5 мл (DNA LoBind) Особое внимание необходимо уделять предотвращению переноса лейкоцитов из лейкоцитарной пленки.
- x) Получают аликвоты объемом 500 мкл для последующих этапов в пробирках Eppendorf объемом 1,5 мл (DNA LoBind).
- xi) Каждую аликвоту необходимо снабдить этикеткой с кодом дня, созданной в приложении NIPT.
- xii) Хранить при -20°C.

2. Выделение ДНК

- a) Авторы настоящего изобретения рекомендуют использование набора для выделения внеклеточной ДНК MagMax™ от Life Technologies ввиду оптимального выхода, сокращения операционного и общего времени, а также простого усовершенствования в направлении автоматизации.
- b) Авторы настоящего изобретения рекомендуют проводить немедленную обработку образцов плазмы крови.
- c) Плазма крови, выделенная из пробирок PAXgene, не требует применения обработки протеиназой *k* в ходе обработки образца.
- d) Получают мастер-микс в пробирке Falcon объемом 15 мл с раствором для лизиса/связывания внеклеточной ДНК MagMax (лизирующим буфером) и магнитными гранулами для выделения внеклеточной ДНК MagMax, состав которого рассчитывают в соответствии со следующей таблицей.

	1 мл плазмы крови	0,5 мл плазмы крови
Раствор для лизиса/связывания	1,25 мл	0,625 мл
Магнитные гранулы	15 мкл	7,5 мкл

- e) Добавляют 633 мкл полученного раствора для лизиса/связывания в пробирку Eppendorf объемом 1,5 мл (DNA LoBind) с аликвотой плазмы крови, перемешивают вихревым способом в течение 10 мин. на встряхивателе Eppendorf MixMate (2000 об/мин.).
- f) Образцы перемещают в магнитный штатив и инкубируют в течение 5 мин.
- g) Надосадочную жидкость сливают и добавляют 800 мкл раствора для промывания внеклеточной ДНК MagMax, перемешивают вихревым способом в течение 30 с на встряхивателе Eppendorf MixMate (2000 об/мин.).
- h) Образцы переносят в магнитный штатив и инкубируют в течение 2 мин.

- i) Надосадочную жидкость сливают и добавляют 800 мкл 80% этанола. Переносят в магнитный штатив и инкубируют в течение непродолжительного времени.
- j) Повторяют предыдущий этап 1х.
- k) Магнитным гранулам дают высохнуть в течение 5 мин., и добавляют 33 мкл сверхчистой воды, перемешивают вихревым способом в течение 5 мин.
- l) Образцы переносят в магнитный штатив и инкубируют в течение 2 мин.
- m) Конечный объем 32 мкл надосадочной жидкости переносят в новую пробирку Eppendorf с кодом дня и переходят к этапу контроля качества (QC: 2 мкл Qubit, 30 мкл набора для получения библиотек ДНК).
- n) Этап контроля качества – концентрацию выделенной ДНК измеряют в системе Qubit 2.0 с помощью набора для HS-анализа dsDNA Qubit®. Концентрация, превышающая 0,3 нг/мкл, может быть обусловлена лизисом клеток матери и снижает эффективность теста. Такие образцы не рекомендуются для использования на последующих этапах.
- o) Выделенную ДНК можно хранить при -20°C.
- p) Выделенную ДНК в пробирках Eppendorf объемом 1,5 мл переносят в лабораторию для получения библиотек для NGS. Если используется транспортировка, всегда следует хранить образцы при +4°C или -20°C. Это необходимо для обеспечения надлежащей физической защиты пробирок с образцами для предотвращения открывания или растрескивания.

ПОЛУЧЕНИЕ БИБЛИОТЕКИ ДЛЯ СЕКВЕНИРОВАНИЯ НОВОГО ПОКОЛЕНИЯ

Процедуры – получение библиотеки (для системы Illumina NextSeq)

- a. Измеряют концентрацию выделенной ДНК с помощью набора для HS-анализа dsDNA Qubit. Если результат измерения концентрации «до низких значений» является приемлемым, продолжают выполнение последующих этапов. Образцы с концентрацией выше 0,3 нг/мкл должны проходить этап контроля качества с использованием Agilent 2100 и набора для высокочувствительного анализа ДНК от

Agilent. Не следует использовать образцы с выявленной ДНК из лизированных клеток. Необходим новый забор крови.

b. На следующем этапе будет использоваться 30 мкл образца.

c. РЕПАРАЦИЯ КОНЦОВ

- i. Смесь для репарации концов 2 (ERP2) оставляют размораживаться при комнатной температуре и центрифугируют в течение непродолжительного времени при 600 g/5 с, инкубируют магнитные гранулы AMPure в течение 30 мин. при комнатной температуре.
- ii. Используют планшет для ПЦР или стрипы для ПЦР для обработки образцов.
- iii. Добавляют 30 мкл очищенного образца ДНК к 20 мкл ERP2, перемешивают вихревым способом и осаждают путем центрифугирования в течение непродолжительного времени.
- iv. Инкубируют в термоциклере при 30°C в течение 30 мин., затем охлаждают до 4°C.
- v. Центрифугируют при 280 x g в течение 1 мин.

d. ОТБОР ПО РАЗМЕРУ

- i. Магнитные гранулы перемешивают вихревым способом в резервуаре в течение 1 минуты. Добавляют 1,1x объема магнитных гранул (авторы настоящего изобретения рекомендуют использование 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 мин. при комнатной температуре.

- iii. Планшет помещают в магнитный штатив на 5 мин. или до тех пор, пока жидкость не станет прозрачной.
- iv. 105 мкл надосадочной жидкости переносят в новую пробирку Eppendorf объемом 1,5 мл.
- v. Магнитные гранулы перемешивают вихревым способом в резервуаре в течение 1 минуты. К надосадочной жидкости добавляют 1,8х (190 мкл) объема магнитных гранул. Перемешивают надлежащим образом и осаждают путем центрифугирования в течение непродолжительного времени.
- vi. Инкубируют в течение 5 мин. при комнатной температуре.
- vii. Пробирки помещают в магнитный штатив на 5 мин. или до тех пор, пока жидкость не станет прозрачной.
- viii. Надосадочную жидкость удаляют и сливают.
- ix. Первое промывание. Добавляют 200 мкл свежеполученного 80% EtOH и инкубируют в течение 30 с.
- x. Надосадочную жидкость удаляют и сливают.
- xi. Второе промывание. Добавляют 200 мкл свежеполученного 80% EtOH и инкубируют в течение 30 с.
- xii. Надосадочную жидкость удаляют и сливают.
- xiii. Пробирки оставляют открытыми в течение 10 минут, позволяя испариться остаточному этанолу.
- xiv. Добавляют 10 мкл буфера для ресуспендирования (RSB), тщательно перемешивают вихревым способом.
- xv. Инкубируют в течение 2 мин. при комнатной температуре.
- xvi. Пробирки помещают в магнитный штатив на 5 мин. или до тех пор, пока жидкость не станет прозрачной.

- xvii. 8,75 мкл прозрачной надосадочной жидкости переносят в новый планшет для ПЦР.
- xviii. ТОЧКА ОСТАНОВКИ – образец можно хранить в течение не более 1 недели при -20°C.

e. ПРИСОЕДИНЕНИЕ А-ХВОСТОВ

- i. Смесь для присоединения А-хвостов извлекают из морозильной камеры и оставляют на 30 мин. при комнатной температуре перед применением.
- ii. Добавляют 6,25 мкл смеси для присоединения А-хвостов, встряхивают в течение непродолжительного времени и осаждают путем центрифугирования в течение непродолжительного времени.
- iii. Планшет для ПЦР помещают в термоциклер и используют следующую программу (предварительное нагревание крышки до 100°C, инкубирование при 37°C в течение 30 минут, инкубирование при 70°C в течение 5 мин., инкубирование при 5°C в течение 5 мин.).
- iv. Центрифугируют при 280 x g в течение 1 мин.

f. ЛИГИРОВАНИЕ АДАПТЕРОВ

- i. Адаптеры и буфер для остановки лигирования (SL) извлекают из морозильной камеры и размораживают их при комнатной температуре, размороженные адаптеры и SL центрифугируют при 600 x g в течение 5 с, смесь для лигирования 2 (LM 2) извлекают из морозильной камеры непосредственно перед применением, перемешивают посредством переворачивания пробирки несколько раз и осаждают путем центрифугирования в течение непродолжительного времени.
- ii. Добавляют 1,25 мкл RSB.

- iii. Добавляют 1,25 мкл LM 2.
- iv. Добавляют 1,25 мкл адаптеров (содержимое каждой лунки, содержащей два вида индексных последовательностей в планшете для адаптеров, разбавляют в соотношении 1:2,5 сверхчистой водой и замораживают в аликвотах объемом 10 мкл, циклы размораживания-замораживания не рекомендуется повторять более 4 раз). Перемешивают вихревым способом в течение непродолжительного времени и центрифугируют при 280 x g в течение 1 мин.
- v. Планшет для ПЦР помещают в термоциклер и используют следующую программу (предварительное нагревание крышки до 100°C, инкубирование при 30°C в течение 10 минут, инкубирование при 5°C в течение 5 мин.).
- vi. Добавляют 2,5 мкл буфера для остановки лигирования, перемешивают в течение непродолжительного времени, встряхивают и осаждают путем центрифугирования в течение непродолжительного времени.
- vii. Общий объем переносят в новую пробирку Eppendorf объемом 1,5 мл.
- viii. Магнитные гранулы перемешивают вихревым способом в резервуаре в течение 1 минуты. Добавляют 21,3 мкл магнитных гранул, перемешивают надлежащим образом и осаждают путем центрифугирования в течение непродолжительного времени.
- ix. Инкубируют в течение 5 мин. при комнатной температуре.
- x. Пробирки помещают в магнитный штатив на 5 мин. или до тех пор, пока жидкость не станет прозрачной.
- xi. Надосадочную жидкость удаляют и сливают.

- xii. Первое промывание. Добавляют 200 мкл свежеполученного 80% EtOH и инкубируют в течение 30 с.
- xiii. Надосадочную жидкость удаляют и сливают.
- xiv. Второе промывание. Добавляют 200 мкл свежеполученного 80% EtOH и инкубируют в течение 30 с.
- xv. Надосадочную жидкость удаляют и сливают.
- xvi. Пробирки оставляют открытыми в течение 5 минут, позволяя испариться остаточному этанолу.
- xvii. Добавляют 26,3 мкл буфера для ресуспендирования (RSB), встряхивают или перемешивают посредством переливания пипеткой.
- xviii. Инкубируют в течение 2 мин. при комнатной температуре.
- xix. Пробирки помещают в магнитный штатив. Инкубируют в течение 5 мин. или до тех пор, пока жидкость не станет прозрачной.
- xx. 25 мкл надосадочной жидкости переносят в новую пробирку Eppendorf объемом 1,5 мл.
- xxi. Магнитные гранулы перемешивают вихревым способом в резервуаре в течение 1 минуты. Добавляют 25 мкл магнитных гранул, перемешивают надлежащим образом и встряхивают в течение непродолжительного времени.
- xxii. Инкубируют в течение 5 мин. при комнатной температуре.
- xxiii. Пробирки помещают в магнитный штатив на 5 мин. или до тех пор, пока жидкость не станет прозрачной.
- xxiv. Надосадочную жидкость удаляют и сливают.
- xxv. Первое промывание. Добавляют 200 мкл свежеполученного 80% EtOH и инкубируют в течение 30 с.

- xxvi. Надосадочную жидкость удаляют и сливают.
- xxvii. Второе промывание. Добавляют 200 мкл свежеполученного 80% EtOH и инкубируют в течение 30 с.
- xxviii. Надосадочную жидкость удаляют и сливают.
- xxix. Пробирки оставляют открытыми в течение 5 минут, позволяя испариться остаточному этанолу.
- xxx. Добавляют 13,75 мкл буфера для ресуспендирования (RSB), встряхивают или перемешивают надлежащим образом.
- xxxi. Инкубируют в течение 2 мин. при комнатной температуре.
- xxxii. Пробирки помещают в магнитный штатив. Инкубируют в течение 5 мин. или до тех пор, пока жидкость не станет прозрачной.
- xxxiii. 12,5 мкл надосадочной жидкости переносят в новый планшет для ПЦР.
- xxxiv. ВОЗМОЖНАЯ ТОЧКА ОСТАНОВКИ – образец можно хранить в течение не более 1 недели при -20°C.

g. КОЛИЧЕСТВЕННАЯ ОЦЕНКА БИБЛИОТЕКИ

- i. Измеряют концентрацию конечной библиотеки ДНК (нг/мкл), используя набор 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 2-4 нМ с помощью следующей формулы:
концентрация конечной библиотеки (нг)/660 x размер конечной библиотеки (п. о.) x 10⁶

- iv. Конечные библиотеки в концентрации 2-4 нМ объединяют в одну конечную.
- v. Конечную библиотеку денатурируют с помощью 0,2 н. NaOH, перемешивают надлежащим образом и инкубируют в течение 5 мин. при комнатной температуре, осаждают путем центрифугирования в течение непродолжительного времени.
- vi. Добавляют нижеуказанный объем 200 мМ Трис-HCl, pH 7, перемешивают надлежащим образом и осаждают путем центрифугирования в течение непродолжительного времени.
- vii. Добавляют нижеуказанный объем предварительно охлажденного буфера для гибридизации HT1 в пробирку с денатурированными библиотеками, перемешивают вихревым способом в течение непродолжительного времени и центрифугируют при 280 g в течение 1 мин. В результате получают денатурированную библиотеку в концентрации 20 пМ.
- viii. Разбавляют раствор денатурированной библиотеки в концентрации 20 пМ до 1,8 пМ следующим образом:

-117 мкл раствора денатурированной библиотеки;

-1183 мкл предварительно охлажденного HT1.

Общий объем составляет 1,3 мл при концентрации 1,8 пМ. Перемешивают в течение непродолжительного времени и осаждают путем центрифугирования.

- ix. Библиотеку помещают на лед до готовности к переходу на NextSeq500.

АНАЛИЗ ПУТЕМ СЕКВЕНИРОВАНИЯ НА ILLUMINA NEXSEQ500

- a. Следует использовать систему массового параллельного секвенирования Illumina NexSeq500.

- b. Для прогона анализа путем секвенирования используют набор для высокопроизводительного секвенирования NextSeq 500/550 v2 от Illumina (75 циклов).
- c. Секвенирование спаренных концов осуществляют с устанавливаемыми параметрами ридов 2 x 35 п. о. со считыванием двух видов адаптеров.
- d. Только файлы FASTQ необходимы для обработки на следующих этапах интерпретации.

Пример 2

Валидация способа с использованием T21

Для валидации прогностических возможностей модели авторы настоящего изобретения секвенировали 4 образца из T_{121x} (с плодом женского пола с трисомией T21), 6 образцов из T_{121y} (с плодом мужского пола с трисомией T21) и 100 образцов из T_{hy} (с эуплоидными плодами мужского пола), что в совокупности составляет 110 образцов. Авторы настоящего изобретения исключили T_{hx} (эуплоидные плоды женского пола) из-за отсутствия информации об эталонной фетальной фракции. Другую когорту из 8 взрослых небеременных женщин и 4 взрослых мужчин секвенировали для оценивания распределений по хромосомам.

Авторы настоящего изобретения применяли ранее описанный способ (глава 1.2 и глава 1.3, общая схема на фиг. 1) для установления начальных параметров распределений по хромосомам и длинам фрагментов, лежащих в основе байесовской модели, продемонстрированной на фиг. 2. Затем проводили итерационное уточнение параметров с помощью алгоритма EM на 60 случайным образом выбранных образцах со здоровыми плодами мужского пола.

Выведенные в результате обучения распределения использовали для параметризации байесовских моделей для всех четырех категорий, т. е. эуплоидных плодов мужского пола, эуплоидных плодов женского пола, плодов мужского пола с трисомией и плодов женского пола с трисомией. В каждой модели регистрировали фетальную фракцию с максимальной апостериорной вероятностью для наблюдаемых фрагментов. Модель с наивысшей вероятностью, а также с наивысшим рангом фетальной фракции была

установлена в качестве наиболее вероятного результата диагностирования. Наиболее вероятные распределения по длинам фрагментов от матери и плода показаны на фигуре 3.

Пол и плоидность были правильно классифицированы для всех 110 образцов (100% чувствительность и специфичность), см. фиг. 4. Наиболее вероятные фетальные фракции характеризовались высокой корреляцией ($r = 0,76$, $p = 3,12 \times 10^{-22}$) с эталонными фракциями, рассчитанными по показателю обилия в Y-хромосоме (плоды мужского пола) или хромосоме 21 (плоды женского пола).

Для оценивания точности прогнозирования фетальной фракции для эуплоидных плодов женского пола авторы настоящего изобретения получали упрощенную байесовскую модель без информации о хромосомном происхождении фрагментов с использованием только длин фрагментов. Следует отметить, что этим способом невозможно вывести фетальную фракцию по показателю обилия фрагментов в Y-хромосоме или хромосоме 21 для образцов с плодом мужского пола или трисомией соответственно. Для каждого образца авторы настоящего изобретения выбирали фракцию с наивысшей апостериорной вероятностью и сравнивали ее с эталонной фракцией. Корреляция оставалась высокой даже без дополнительной информации ($r = 0,77$, $p = 5,16 \times 10^{-21}$).

Пример 3

Валидация способа с использованием T13, T18 и X0

Прогностическая мощность байесовской сети была продемонстрирована также на дополнительных секвенированных образцах с анеуплоидией; трисомией по хромосоме 13 или синдромом Патау (1 новый образец), трисомией по хромосоме 18 или синдромом Эдвардса (2 новых образца) и моносомией по X-хромосоме или синдромом Тернера (1 образец).

Автоматический логический вывод распределений по хромосомам с помощью алгоритма EM был невозможен из-за небольшого количества целевых образцов. Следовательно, ожидаемые распределения для каждого состояния строили с применением способа, представленного в разделе «Построение распределений по хромосомам». Поскольку распределения по длинам фрагментов не зависят от

клинического состояния, авторы настоящего изобретения использовали параметры, логически выведенные при использовании образцов из примера 2.

Авторы настоящего изобретения тестировали каждый образец в отношении следующих состояний:

- эуплоидный мужчина
- эуплоидная женщина
- мальчик с анеуплоидией по хромосоме 13
- девочка с анеуплоидией по хромосоме 13
- мальчик с анеуплоидией по хромосоме 18
- девочка с анеуплоидией по хромосоме 18
- мальчик с анеуплоидией по хромосоме 21
- девочка с анеуплоидией по хромосоме 21
- девочка с моносомией по X-хромосоме

Клиническое состояние для каждого образца было правильно спрогнозировано в соответствии с наиболее вероятным набором параметров (100% чувствительность и специфичность).

Литературные источники

Bianchi, Diana W, Platt, Lawrence D, Goldberg, James D, Abuhamad, Alfred Z, Sehert, Amy J, & Rava, Richard P. 2012. Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing. *Obstetrics & Gynecology*, 119(5), 890–901.

Ehrich M., C. Deciu, T. Zwiefelhofer, J. A. Tynan, L. Cagasan, R. Tim, V. Lu, R. McCullough E. McCarthy, A. O. Nygren, et al. Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting. *American journal of obstetrics and gynecology*, 204(3):205–e1, 2011.

Fan, H. C., Blumenfeld, Y. J., Chitkara, U., Hudgins, L., & Quake, S. R. (2008). Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proceedings of the National Academy of Sciences*, 105(42), 16266-16271.

Chiu, Rossa WK, Chan, KC Allen, Gao, Yuan, Lau, Virginia YM, Zheng, Wenli, Leung, Tak Y, Foo, Chris HF, Xie, Bin, Tsui, Nancy BY, Lun, Fiona MF, et al. 2008. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. *Proceedings of the National Academy of Sciences*, 105(51), 20458-20463.

Chiu, Rossa WK, Akolekar, Ranjit, Zheng, Yama WL, Leung, Tak Y, Sun, Hao, Chan, KC, Lun, Fiona MF, Go, Attie TJI, Lau, Elizabeth T, To, William WK, et al. 2011. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. *Bmj*, 342.

Jiang, K. C. Allen Chan, Gary J. W. Liao, Yama W. L. Zheng, Tak Y. Leung, Rossa W. K. Chiu, Yuk Ming Dennis Lo, Hao Sun; FetalQuant: deducing fractional fetal DNA concentration from massively parallel sequencing of DNA in maternal plasma. *Bioinformatics* 2012; 28 (22): 2883-2890. doi: 10.1093/bioinformatics/bts549

Kim, S. K., Hannum, G., Geis, J., Tynan, J., Hogg, G., Zhao, C., ... & Boom, D. (2015). Determination of fetal DNA fraction from the plasma of pregnant women using sequence read counts. *Prenatal diagnosis*, 35(8), 810-815.

B. Langmead and S. L. Salzberg. Fast gapped-read alignment with bowtie 2. *Nature methods*, 9(4):357–359, 2012.

Lau, Tze Kin, Chen, Fang, Pan, Xiaoyu, Pooh, Ritsuko K, Jiang, Fuman, Li, Yihan, Jiang, Hui, Li, Xuchao, Chen, Shengpei, & Zhang, Xiuqing. 2012. Noninvasive prenatal diagnosis of common fetal chromosomal aneuploidies by maternal plasma DNA sequencing. *Journal of Maternal-Fetal and Neonatal Medicine*, 25(8), 1370–1374.

Lo, YM, Corbetta, Noemi, Chamberlain, Paul F, Rai, Vik, Sargent, Ian L, Redman, Christopher WG, & Wainscoat, James S. 1997. Presence of fetal DNA in maternal plasma and serum. *The Lancet*, 350(9076), 485–487.

A. J. Sehnert, B. Rhees, D. Comstock, E. de Feo, G. Heilek, J. Burke, and R. P. Rava. Optimal detection of fetal chromosomal abnormalities by massively parallel DNA sequencing of cell-free fetal DNA from maternal blood. *Clinical chemistry*, 57(7):1042–1049, 2011.

Straver, R., Oudejans, C., Sistermans, E. A., & Reinders, M. J. (2016). Calculating the fetal fraction for noninvasive prenatal testing based on genome-wide nucleosome profiles. *Prenatal diagnosis*, 36(7), 614-621.

C. Y. Stephanie, K. A. Chan, Y. W. Zheng, P. Jiang, G. J. Liao, H. Sun, R. Akolekar, T. Y. Leung, A. T. Go, J. M. van Vugt, et al. Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing. *Proceedings of the National Academy of Sciences*, 111(23): 8583–8588, 2014.

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

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.

Формула изобретения

1. Способ определения анеуплоидии и пола у плода, а также соответствующей фетальной фракции по образцу крови, плазмы крови или сыворотки крови матери, содержащему смесь фрагментов ДНК, происходящих от плода и матери, где смесь фрагментов ДНК, происходящих от плода и матери, представляет собой молекулы циркулирующей внеклеточной ДНК, при этом указанный способ включает:

a) осуществление случайного секвенирования по меньшей мере части множества молекул внеклеточной ДНК, содержащихся в образце, за счет чего обеспечивается получение информации о последовательностях для множества фрагментов ДНК, происходящих от плода и матери, из образца от матери, где информация о последовательностях содержит последовательности ридов;

b) картирование последовательностей ридов на эталонный человеческий геном;

c) разделение эталонного генома на последовательные и неперекрывающиеся секции размером 20 т.о. и распределение картированных последовательностей ридов по указанным секциям в соответствии с их картированной локализацией;

d) необязательно внесение в количество ридов поправки на сдвиг содержания GC;

e) выведение в результате обучения исходных распределений выбранных признаков фрагментов ДНК в наборе обучающих образцов;

f) построение байесовской сети, которая включает выбранные признаки фрагментов ДНК по четырем категориям, в которых объединены пол и наличие или отсутствие анеуплоидии у плода, которые представляют собой эуплоиды мужского пола, анеуплоиды мужского пола, эуплоиды женского пола, анеуплоиды женского пола, в образце;

g) прогон итерационного алгоритма EM в байесовской сети;

h) на основании результата EM:

- a. классификацию фрагментов ДНК из образца как происходящих от матери или плода;
- b. корректировку распределений выбранных признаков фрагментов ДНК;
- i) повторение этапов g) и h) до тех пор, пока параметры выведенного в результате обучения распределения не будут изменяться между итерациями незначительно, например, пока сумма различий не будет меньше 0,000001;
- j) расчет вероятности для выбранных признаков фрагментов ДНК в конечной байесовской сети;
- k) нахождение для каждой категории фетальной фракции с наивысшей апостериорной вероятностью;
- l) выбор наиболее вероятной категории для образца с определением классификации образца как относящегося к одной из указанных категорий: эуплоиды мужского пола, анеуплоиды мужского пола, эуплоиды женского пола, анеуплоиды женского пола;

где байесовскую сеть для каждой указанной категории, которая включает выбранные признаки фрагмента ДНК, строят с использованием набора обучающих образцов, содержащего набор от беременных женщин с эуплоидным плодом женского пола и набор от беременных женщин с эуплоидным плодом мужского пола, с применением тех же этапов, что и на указанных этапах а) - i), с получением исходных распределений выбранных признаков фрагментов ДНК.

2. Способ по п. 2, где выбранные признаки фрагмента ДНК представляют собой длину фрагмента LEN и хромосому, являющуюся источником происхождения, CHROM.

3. Способ по п. 1 или п. 2, где распределение по хромосоме, являющейся источником происхождения, определяют на основании фракционной аутосомной представленности FAR и фракционного хромосомного распределения FCR, и распределение по длине фрагмента определяют на основании фракционного распределения длин FLD.

4. Способ по любому из пп. 1-3, где набор обучающих образцов содержит:

набор от беременных женщин с эуплоидным плодом женского пола;

набор от беременных женщин с эуплоидным плодом мужского пола;

набор от беременных женщин с анеуплоидным плодом женского пола;

набор от беременных женщин с анеуплоидным плодом мужского пола;

набор от взрослых мужчин и

набор от небеременных женщин.

5. Способ по любому из пп. 1-4, где анеуплоидия представляет собой трисомию по хромосоме 13, трисомию по хромосоме 18 или трисомию по хромосоме 21.

6. Выполняемый с помощью компьютера способ на основе байесовской сети для определения анеуплоидии и пола у плода, а также соответствующей фетальной фракции, посредством случайного секвенирования по меньшей мере части множества фрагментов ДНК, содержащихся в образце крови, плазмы крови или сыворотки крови матери, содержащем смесь фрагментов ДНК, происходящих от плода и матери, где смесь фрагментов ДНК, происходящих от плода и матери, представляет собой молекулы циркулирующей внеклеточной ДНК, и посредством случайного секвенирования по меньшей мере части множества молекул внеклеточной ДНК, содержащихся в обучающем образце, при этом указанный способ включает этапы b) - l) из способа по п. 1,

где для каждой указанной категории байесовскую сеть, которая включает выбранные признаки фрагмента ДНК, строят по набору обучающих образцов, содержащему по меньшей мере набор от беременных женщин с эуплоидным плодом женского пола и набор от беременных женщин с эуплоидным плодом мужского пола, с применением тех же этапов, что и на указанных этапах а) - i), с получением исходных распределений выбранных признаков фрагмента ДНК.

7. Выполняемый с помощью компьютера способ по п. 6, где выбранные признаки фрагмента ДНК представляют собой длину фрагмента LEN и хромосому, являющуюся источником происхождения, CHROM.

8. Выполняемый с помощью компьютера способ по п. 6 или п. 7, где распределение по хромосоме, являющейся источником происхождения, определяют на основании фракционной аутосомной представленности FAR и фракционного хромосомного распределения FCR, и распределение по длине фрагмента определяют на основании фракционного распределения длин FLD.

9. Выполняемый с помощью компьютера способ по любому из пп. 6-8, где набор обучающих образцов содержит:

набор от беременных женщин с эуплоидным плодом женского пола;

набор от беременных женщин с эуплоидным плодом мужского пола;

набор от беременных женщин с анеуплоидным плодом женского пола;

набор от беременных женщин с анеуплоидным плодом мужского пола;

набор от взрослых мужчин и

набор от небеременных женщин.

10. Выполняемый с помощью компьютера способ по любому из пп. 6-9, где анеуплоидия представляет собой трисомию по хромосоме 13, трисомию по хромосоме 18 или трисомию по хромосоме 21.

11. Компьютерный программный продукт, содержащий машиночитаемый носитель, содержащий множество инструкций по управлению компьютерной системой для осуществления способа по любому из пп. 6-10.

Формула изобретения

1. Способ определения анеуплоидии и пола у плода, а также соответствующей фетальной фракции по образцу крови, плазмы крови или сыворотки крови матери, содержащему смесь фрагментов ДНК, происходящих от плода и матери, где смесь фрагментов ДНК, происходящих от плода и матери, представляет собой молекулы циркулирующей внеклеточной ДНК, при этом указанный способ включает:

- a) осуществление случайного секвенирования по меньшей мере части множества молекул внеклеточной ДНК, содержащихся в образце, за счет чего обеспечивается получение информации о последовательностях для множества фрагментов ДНК, происходящих от плода и матери, из образца от матери, где информация о последовательностях содержит последовательности ридов;
- b) картирование последовательностей ридов на эталонный человеческий геном;
- c) разделение эталонного генома на последовательные и неперекрывающиеся секции размером 20 т. о. и распределение картированных последовательностей ридов по указанным секциям в соответствии с их картированной локализацией;
- d) необязательно внесение в количество ридов поправки на сдвиг содержания GC;
- e) выведение в результате обучения исходных распределений выбранных признаков фрагментов ДНК в наборе обучающих образцов;
- f) построение байесовской сети, которая включает выбранные признаки фрагментов ДНК по четырем категориям, в которых объединены пол и наличие или отсутствие анеуплоидии у плода, которые представляют собой эуплоиды мужского пола, анеуплоиды мужского пола, эуплоиды женского пола, анеуплоиды женского пола, в образце;
- g) прогон итерационного алгоритма EM в байесовской сети;
- h) на основании результата EM:
 - a. классификацию фрагментов ДНК из образца как происходящих от матери или плода;

- b. корректировку распределений выбранных признаков фрагментов ДНК;
- i) повторение этапов g) и h) до тех пор, пока параметры выведенного в результате обучения распределения не будут изменяться между итерациями незначительно, например, пока сумма различий не будет меньше 0,000001;
- j) расчет вероятности для выбранных признаков фрагментов ДНК в конечной байесовской сети;
- k) нахождение для каждой категории фетальной фракции с наивысшей апостериорной вероятностью;
- l) выбор наиболее вероятной категории для образца с определением классификации образца как относящегося к одной из указанных категорий: эуплоиды мужского пола, анеуплоиды мужского пола, эуплоиды женского пола, анеуплоиды женского пола;

где байесовскую сеть для каждой указанной категории, которая включает выбранные признаки фрагмента ДНК, строят с использованием набора обучающих образцов, содержащего набор от беременных женщин с эуплоидным плодом женского пола и набор от беременных женщин с эуплоидным плодом мужского пола, с применением тех же этапов, что и на указанных этапах а) - i), с получением исходных распределений выбранных признаков фрагментов ДНК,

где на этапе e) выбранные признаки фрагмента ДНК представляют собой длину фрагмента LEN и хромосому, являющуюся источником происхождения, CHROM, и

набор обучающих образцов содержит:

набор от беременных женщин с эуплоидным плодом женского пола;

набор от беременных женщин с эуплоидным плодом мужского пола;

набор от взрослых мужчин и

набор от небеременных женщин.

2. Способ по п. 1, где распределение по хромосоме, являющейся источником происхождения, определяют на основании фракционной аутосомной представленности

FAR и фракционного хромосомного распределения FCR, и распределение по длине фрагмента определяют на основании фракционного распределения длин FLD.

3. Способ по п. 1 или п. 2, где анеуплоидия представляет собой трисомию по хромосоме 13, трисомию по хромосоме 18 или трисомию по хромосоме 21.

4. Выполняемый с помощью компьютера способ на основе байесовской сети для определения анеуплоидии и пола у плода, а также соответствующей фетальной фракции, посредством случайного секвенирования по меньшей мере части множества фрагментов ДНК, содержащихся в образце крови, плазмы крови или сыворотки крови матери, содержащем смесь фрагментов ДНК, происходящих от плода и матери, где смесь фрагментов ДНК, происходящих от плода и матери, представляет собой молекулы циркулирующей внеклеточной ДНК, и посредством случайного секвенирования по меньшей мере части множества молекул внеклеточной ДНК, содержащихся в обучающем образце, при этом указанный способ включает этапы b) - l) из способа по п. 1,

где для каждой указанной категории байесовскую сеть, которая включает выбранные признаки фрагмента ДНК, строят по набору обучающих образцов, содержащему по меньшей мере набор от беременных женщин с эуплоидным плодом женского пола и набор от беременных женщин с эуплоидным плодом мужского пола, с применением тех же этапов, что и на указанных этапах а) - i), с получением исходных распределений выбранных признаков фрагмента ДНК,

где на этапе e) выбранные признаки фрагмента ДНК представляют собой длину фрагмента LEN и хромосому, являющуюся источником происхождения, CHROM, и

набор обучающих образцов содержит:

набор от беременных женщин с эуплоидным плодом женского пола;

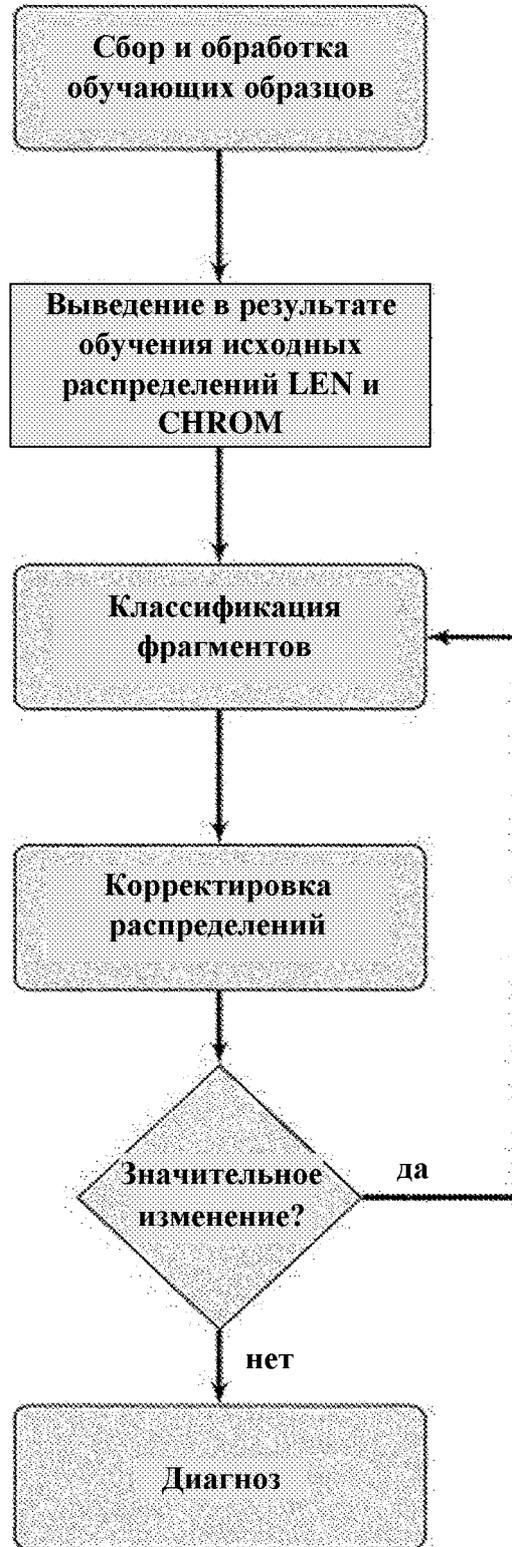
набор от беременных женщин с эуплоидным плодом мужского пола;

набор от взрослых мужчин и

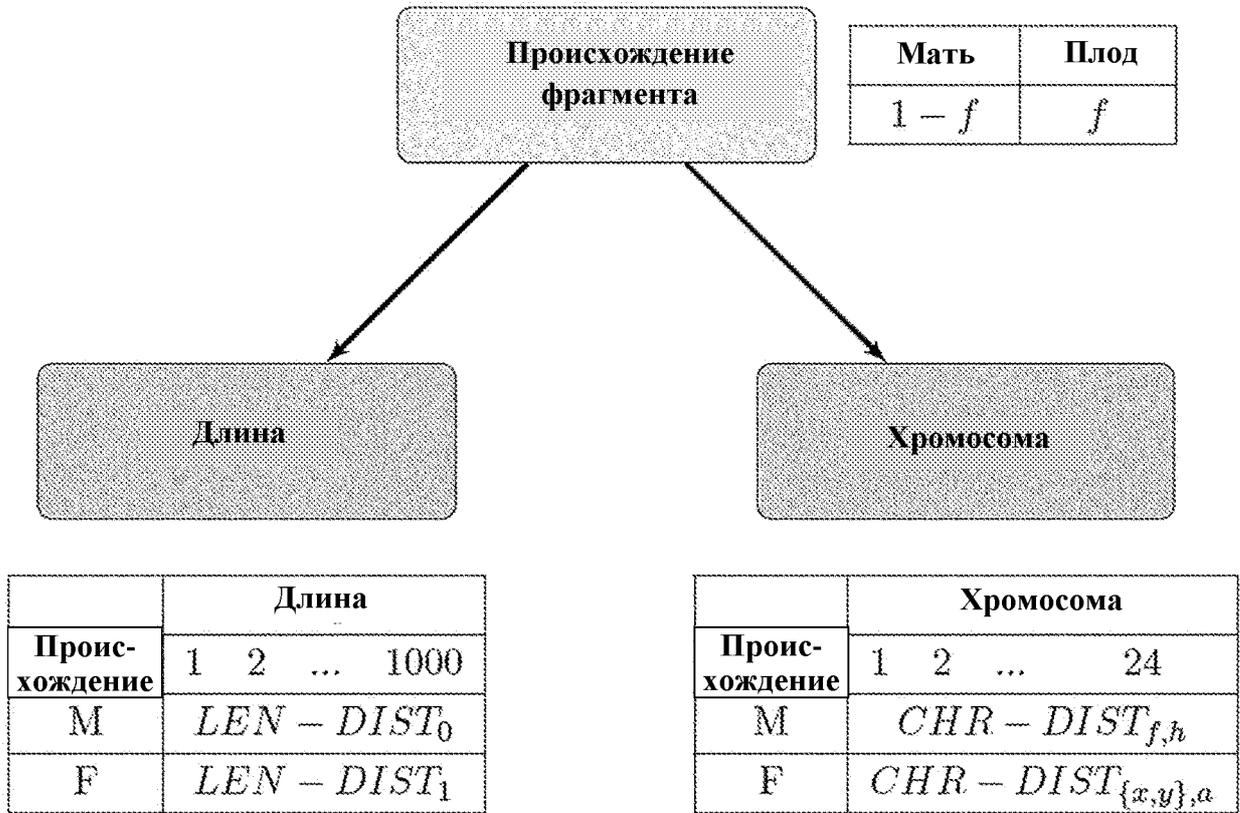
набор от небеременных женщин.

5. Выполняемый с помощью компьютера способ по п. 4, где распределение по хромосоме, являющейся источником происхождения, определяют на основании фракционной аутомной представленности FAR и фракционного хромосомного распределения FCR, и распределение по длине фрагмента определяют на основании фракционного распределения длин FLD.
6. Выполняемый с помощью компьютера способ по п. 4 или п. 5, где анеуплоидия представляет собой трисомию по хромосоме 13, трисомию по хромосоме 18 или трисомию по хромосоме 21.
7. Компьютерный программный продукт, содержащий машиночитаемый носитель, содержащий множество инструкций, которые при выполнении программы на компьютере обеспечивают осуществление компьютером способа по любому из пп. 4-6.

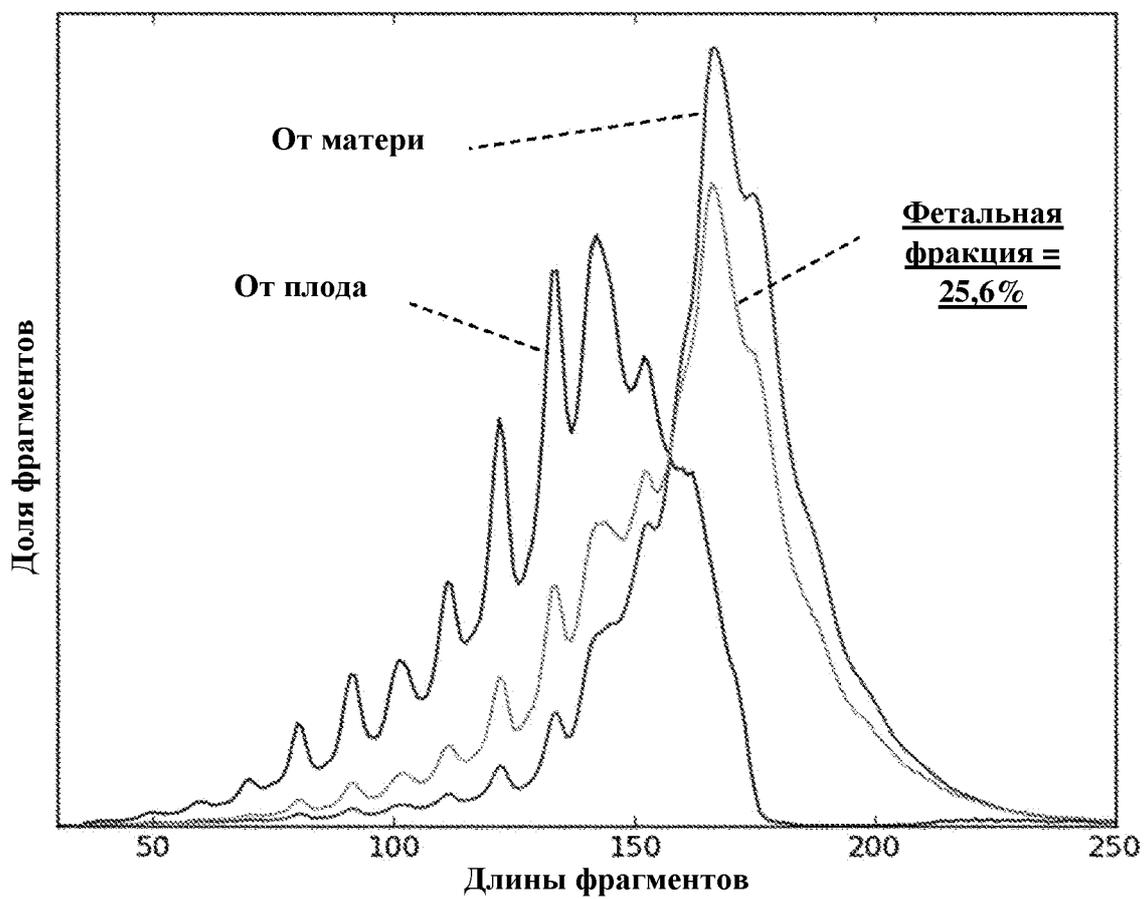
Фиг. 1



Фиг. 2



Фиг. 3



Фиг. 4

