



(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ

(45) Дата публикации и выдачи патента
2021.07.08

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

(21) Номер заявки
201792570

(22) Дата подачи заявки
2016.05.20

(54) МУЛЬТИПЛЕКСНЫЙ ПАРАЛЛЕЛЬНЫЙ АНАЛИЗ ЦЕЛЕВЫХ ГЕНОМНЫХ ОБЛАСТЕЙ С ЦЕЛЬЮ НЕИНВАЗИВНОГО ПРЕНАТАЛЬНОГО ТЕСТИРОВАНИЯ

(31) 62/165,593; 62/263,320

(32) 2015.05.22; 2015.12.04

(33) US

(43) 2018.05.31

(86) PCT/IB2016/000833

(87) WO 2016/189388 2016.12.01

(71)(73) Заявитель и патентовладелец:

**ЭнАйПиДи ДЖЕНЕТИКС ПАБЛИК
КОМПАНИ ЛИМИТЕД (СУ)**

(72) Изобретатель:

**Кумбарис Джордж, Кипри Елена,
Цангарас Кириакос, Ахиллеос
Ахиллеас, Мина Петрос, Папагеоргиу
Елисавет А., Пацалис Филиппос К.
(СУ)**

(74) Представитель:

Медведев В.Н. (РУ)

(56) BERNHARD ZIMMERMANN ET AL.: "Noninvasive prenatal aneuploidy testing of chromosomes 13, 18, 21, X, and Y, using targeted sequencing of polymorphic loci", PRENATAL DIAGNOSIS, vol. 32, no. 13, 30 October 2012 (2012-10-30), pages 1233-1241, XP055119823, ISSN: 0197-3851, DOI: 10.1002/pd.3993, cited in the application, abstract; figures 1-5, Methods Conclusion

K.H. NICOLAIDES ET AL.: "Validation of targeted sequencing of single-nucleotide polymorphisms for non-invasive prenatal detection of aneuploidy of chromosomes 13, 18, 21, X, and Y", PRENATAL DIAGNOSIS, vol. 33, no. 6, 24 June 2013 (2013-06-24), pages 575-579,

XP055257233, GB, ISSN: 0197-3851, DOI: 10.1002/pd.4103, cited in the application, abstract; figure 1; table 1 "patients and methods" on page 576

ANDREW B. SPARKS ET AL.: "Selective analysis of cell-free DNA in maternal blood for evaluation of fetal trisomy", PRENATAL DIAGNOSIS, vol. 32, no. 1, 1 January 2012 (2012-01-01), pages 3-9, XP055024132, ISSN: 0197-3851, DOI: 10.1002/pd.2922, cited in the application, abstract; figures 1-4, Methods Conclusion

GARY J.W. LIAO ET AL.: "Noninvasive Prenatal Diagnosis of Fetal Trisomy 21 by Allelic Ratio Analysis Using Targeted Massively Parallel Sequencing of Maternal Plasma DNA", PLOS ONE, vol. 7, no. 5, 29 May 2012 (2012-05-29), page e38154, XP055062111, DOI: 10.1371/journal.pone.0038154, abstract; figures 1-4

ELISAVET A. PAPAGEORGIU ET AL.: "Non-invasive prenatal diagnosis of aneuploidies: new technologies and clinical applications", GENOME MED, vol. 4, no. 5, 1 January 2012 (2012-01-01), page 46, XP055305603, London, UK, ISSN: 1756-994X, DOI: 10.1186/gm345, abstract

R.W.K. CHIU ET AL.: "Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, vol. 105, no. 51, 23 December 2008 (2008-12-23), pages 20458-20463, XP055284693, US ISSN: 0027-8424, DOI: 10.1073/pnas.0810641105, cited in the application, abstract; figure 1

KOUMBARIS GEORGE ET AL.: "Cell-Free DNA Analysis of Targeted Genomic Regions in Maternal Plasma for Non-Invasive Prenatal Testing of Trisomy 21, Trisomy 18, Trisomy 13, and Fetal", CLINICAL CHEMISTRY, AMERICAN ASSOCIATION FOR CLINICAL CHEMISTRY, WASHINGTON, DC, vol. 62, no. 6, 1 June 2016 (2016-06-01), pages 848-855, XP009191834, ISSN: 0009-9147, DOI: 10.1373/CLINCHEM.2015.252502, the whole document

(57) В изобретении предложены способы неинвазивного пренатального тестирования, которые позволяют обнаруживать риск хромосомных и субхромосомных аномалий, включая, но не ограничиваясь ими, анеуплоидии, микроделеции и микродупликации, вставки, транслокации, инверсии и мутации небольшого размера, включая точечные мутации и мутационные подписи. В способах по данному изобретению применяют пул последовательностей целевого захвата для обогащения последовательностей, представляющих интерес, в смешанном образце, содержащем как материнскую ДНК, так и ДНК плода, с последующим массово-параллельным секвенированием и статистическим анализом обогащенной популяции, чтобы таким образом обнаружить риск генетической аномалии в ДНК плода. Также предложены наборы для реализации способов по данному изобретению.

Перекрестная ссылка на родственные заявки

Заявка на данное изобретение испрашивает приоритет по предварительной заявке на патент США № 62/263320 (поданной 4 декабря 2015 г.) по предварительной заявке на патент США № 62/165593 (поданной 22 мая 2015 г.), которые включены в данный документ посредством ссылки.

Уровень техники

Открытие свободной ДНК плода (сэДНК) в кровообращении матери (Lo, Y.M. et al. (1997), *Lancet* 350:485-487) было важным этапом в развитии неинвазивного пренатального тестирования на анеуплоидию плода и открыло новые возможности для клиницистов. сэДНК успешно применяется для определения пола плода и резус-D статуса плода в материнской плазме (см., например, Bianchi, D. et al. (2005), *Obstet. Gynecol.* 106:841-844; Lo, Y.M. et al. (1998), *N. Engl. J. Med.* 339:1734-1738; патент США № 6258540; публикация РСТ WO 91/07660). Эти методы стали широко применяемыми исследованиями в ряде диагностических лабораторий по всему миру. Однако прямой анализ ограниченного количества сэДНК в присутствии избытка материнской ДНК является большой проблемой при оценке неинвазивного пренатального тестирования (НИПТ) анеуплоидии плода.

По первоначальным оценкам, процент сэДНК в циркулирующей крови матери составляет около 3-6% от общей ДНК (Lo, Y.M. et al. (1998), *Am. J. Hum. Genet.* 62:768-775). Однако недавние исследования показывают, что эмбриональная ДНК может достигать 10-20% от общей ДНК в циркулирующей крови матери (Lun, F.M. et al. (2008), *Clin. Chem.* 54:1664-1672). При анеуплоидиях одна из хромосом присутствует с дополнительным или меньшим количеством копий. Например, в случае трисомии 21 хромосома 21 присутствует в трех копиях вместо двух. Таким образом, способность отличать нормальные случаи от случаев трисомии 21 зависит от способности обнаруживать дополнительную копию хромосомы 21. Однако высокий уровень материнской ДНК в циркулирующей крови матери по сравнению с ограниченным количеством ДНК плода еще больше усложняет количественную оценку.

За последнее десятилетие было применено большое количество различных методов, направленных на распознавание сэДНК от циркулирующей материнской ДНК или на обогащение сэДНК (Chan, K.C. et al. (2004), *Clin. Chem.* 50:88-92; Papageorgiou, E.A. et al. (2009), *Am. J. Pathol.* 174:1609-1618). К ним относятся подходы, основанные на ДНК, такие как подходы на основе анализа последовательностей Chiu, R.W. et al. (2008), *Proc. Natl. Acad. Sci. USA*, 105:20458-20463; Fan, H.C. et al. (2008), *Proc. Natl. Acad. Sci. USA* 105:16266-16271) или эпигенетические подходы, которые фокусируются на исследовании статуса метилирования ДНК плода либо с применением обработки ДНК бисульфитом натрия (Chim, S.S. et al. (2005), *Proc. Natl. Acad. Sci. USA* 102:14753-14758; публикация РСТ WO 2003/020974; публикация РСТ WO 2005/028674), чувствительных к метилированию рестрикционных ферментов (Old, R.W. et al. (2007), *Reprod. Biomed. Online* 15:227-235; РСТ Publication WO 2005/035725), либо антител, специфичных к остаткам 5-метилцитозина CpG-динуклеотидов по всему геному (Papageorgiou, E.A. et al. (2009), *Am. J. Pathol.* 174:1609-1618, Papageorgiou, E.A. et al. (2011), *Nature Medicine* 17:510-513; Tsaliki, E. et al. (2012), *Prenat. Diagn.* 32: 996-1001; публикация РСТ WO 2011/0 92 592). Альтернативные подходы нацелены на эмбрион-специфическую мРНК (Ng, E.K. et al. (2003), *Proc. Natl. Acad. Sci. USA* 100:4748-4753) или сосредоточены на исследовании эмбрион-специфических белков (Avent, N.D. et al. (2008), *Semin. Fetal Neonatal Med.* 13:91-98).

Внедрение технологий секвенирования следующего поколения (ССП) в развитие НИПТ анеуплоидий осуществило революцию в этой области. В 2008 году две независимые группы исследователей продемонстрировали, что НИПТ трисомии 21 может быть достигнуто с применением метода массово-параллельного дробного секвенирования (МПДС) следующего поколения (Chiu, R. W. et al. (2008), *Proc. Natl. Acad. Sci. USA* 105:20458-20463; Fan, H.C. et al. (2008), *Proc. Natl. Acad. Sci. USA* 105:16266-16271). Новая эра НИПТ для оценки анеуплоидий открыла новые возможности для внедрения этих технологий в клиническую практику. Биотехнологические компании, которые частично или полностью заняты разработкой тестов для НИПТ, начали широкомасштабные клинические исследования по их внедрению (Palomaki, G.E. et al. (2011), *Genet. Med.* 13:913-920; Ehrich, M. et al. (2011), *Am. J. Obstet. Gynecol.* 204:205e1-11; Chen, E.Z. et al. (2011), *PLoS One* 6:e21791; Sehnert, A.J. et al. (2011), *Clin. Chem.* 57:1042-1049; Palomaki, G.E. et al. (2012), *Genet. Med.* 14:296-305; Bianchi, D.W. et al. (2012), *Obstet. Gynecol.* 119:890-901; Zimmerman, B. et al. (2012), *Prenat. Diagn.* 32:1233-1241; Nicolaidis, K.H. et al. (2013), *Prenat. Diagn.* 33:575-579; Sparks, A.B. et al. (2012), *Prenat. Diagn.* 32:3-9). В настоящее время четыре компании в Соединенных Штатах (SEQUENOM Inc., Verinata Health, Inc., Natera и Ariosa) предлагают тестирование НИПТ с применением подходов секвенирования следующего поколения.

В первоначальных подходах к НИПТ применялись такие методики SSP, как массово-параллельное дробное секвенирование (МПДС) (см., например, патенты США № 7888017, 8008018, 8195415, 8296076, 8682594, публикацию заявки на патент США 20110201507, публикацию заявки на патент США 20120270739). Таким образом, эти подходы основаны на целом геноме, в которых весь материнский образец, содержащий как материнскую ДНК, так и свободную эмбриональную ДНК, подвергаются амплификации, секвенированию и анализу.

Совсем недавно были разработаны целенаправленные подходы SSP для НИПТ, при которых секвенируют только конкретные последовательности, представляющие интерес. Например, описан подход

ССП на основе SNP, включающий целевую амплификацию и анализ SNP на хромосомах 13, 18, 21, X и Y в одной реакции (Zimmerman, B. et al. (2012), Prenat. Diagn. 32:1233-1241; Nicolaidis, K.H. et al. (2013), Prenat. Diagn. 33: 575-579; публикация РСТ WO 2011/041485; патент США № 8825412). Кроме того, был разработан подход на основе ССП, при котором секвенируют только конкретные области, представляющие интерес, при этом три зонда на целевой локус гибридизируют с дополнительной матрицей. Сразу после гибридизации указанных трех зондов их лигируют, вследствие чего образуется один непрерывный длинный зонд, который затем амплифицируют и секвенируют (Sparks, A.B. et al. (2012), Prenat. Diagn. 32:3-9; публикация патента США 20120034603). Образцы анализируют с помощью мультиплексного анализа, называемого цифровым анализом выбранных областей (ЦАВО). Такие целевые подходы требуют значительно меньшего количества этапов секвенирования, чем подходы МПДС, поскольку секвенирование выполняется только на определенных локусах в интересующей хромосоме, а не на всем геноме.

На сегодняшний день по-прежнему существует необходимость в дополнительных методиках для подходов к НИПТ, основанных на ССП, в частности подходов, которые могут быть нацелены на конкретные последовательности, представляющие интерес, в результате чего значительно уменьшается количество необходимых этапов секвенирования по сравнению с подходами, основанными на вовлечении всего генома.

Краткое описание сущности изобретения

В данном изобретении предлагаются способы неинвазивного пренатального тестирования, которые позволяют обнаруживать риск хромосомных аномалий и используют целевой подход для обогащения последовательностей, представляющих интерес, перед массово-параллельным секвенированием и статистическим анализом, что позволяет проводить высокоточный подсчет и оценку хромосомных составляющих материнской плазмы в областях, представляющих интерес. Таким образом, способы согласно данному изобретению уменьшают количество этапов секвенирования, необходимое для массово-параллельного секвенирования, обеспечивают высокую пропускную способность анализа со снижением стоимости и очень высокой степенью точности. В способах по данному изобретению применяется пул последовательностей целевого захвата (ПЦЗ) для обогащения последовательностей, представляющих интерес, в смешанном образце, содержащем как материнскую, так и эмбриональную ДНК. В частности, пул ПЦЗ разработан таким образом, что последовательности в пуле имеют функции, которые оптимизируют эффективность, специфичность и точность оценки хромосомных аномалий. Если более конкретно, то оптимизируются размеры ПЦЗ, количество ПЦЗ, их размещение на хромосоме(ах), представляющей интерес, и их содержание GC. Гибридизация ПЦЗ с библиотекой последовательностей, полученной из смешанного образца ДНК матери и эмбриона (например, образца материнской плазмы, содержащего сДНК), с последующим выделением последовательностей в пределах библиотеки, которые связываются с ПЦЗ, позволяет обогащать только те области хромосом, которые представляют интерес, перед массово-параллельным секвенированием и анализом.

Соответственно, в одном аспекте данного изобретения предлагается способ анализа риска хромосомной аномалии в хромосоме, представляющей интерес, в ДНК плода, в смешанном образце ДНК матери и эмбриона, при этом указанный способ включает:

- (a) подготовку библиотеки последовательностей из смешанного образца;
- (b) гибридизацию библиотеки последовательностей с пулом последовательностей целевого захвата (ПЦЗ), при этом пул ПЦЗ содержит последовательности, которые связываются с одной или более хромосомами, представляющими интерес, и при этом:
 - (i) каждая последовательность в пуле составляет 100-260, и/или 100-300, и/или 100-350 п.о. в длину, причем каждая последовательность имеет 5'-конец и 3'-конец;
 - (ii) каждая последовательность в пуле связывается с хромосомой(ами), представляющей интерес, по меньшей мере на расстоянии 150 пар оснований как на 5'-конце, так и на 3'-конце от областей, содержащих вариации числа копий, сегментные дублирования или повторяющиеся элементы ДНК; а также
 - (iii) содержание GC в ПЦЗ составляет 19-50, и/или 19-60, и/или 19-70, и/или 19-80%;
- (c) выделение элементов библиотеки последовательностей, которые связываются с ПЦЗ для получения обогащенной библиотеки;
- (d) амплификацию и секвенирование обогащенной библиотеки; а также
- (e) проведение статистического анализа на выходе секвенирования обогащенных последовательностей библиотеки, чтобы тем самым определить риск хромосомной и/или другой генетической аномалии в ДНК плода.

В одном варианте реализации данного изобретения хромосомная аномалия представляет собой анеуплоидию, такую как трисомия. Хромосомой, представляющей интерес, может быть любая хромосома, хотя предпочтительные хромосомы включают хромосомы 13, 18, 21, X и Y. Предпочтительной анеуплоидией для обнаружения является трисомия 21 (T21). В дополнение к количественным аномалиям, таким как анеуплоидия, данное изобретение позволяет обнаруживать другие типы хромосомных аномалий, таких как структурные нарушения, включая, но не ограничиваясь ими, изменения количества копий, включая, но не ограничиваясь ими, микроделеции и микродупликации, вставки, транслокации, инверсии и мутации небольшого размера, включая точечные мутации и мутационные подписи.

В одном варианте реализации данного изобретения пул ПЦЗ фиксируют на твердой подложке. Например, ПЦЗ могут быть биотинилированы и связаны с магнитными гранулами, покрытыми стрептавидином. В другом варианте реализации данного изобретения пул ПЦЗ может свободно перемещаться в растворе.

В одном варианте реализации данного изобретения ПЦЗ предназначены для связывания с хромосомой, представляющей интерес, и одной или более эталонными последовательностями для определения риска хромосомной аномалии в хромосоме, представляющей интерес. В альтернативном варианте пул ПЦЗ может быть сконструирован таким образом, чтобы он мог связываться с множеством хромосом, представляющих интерес, и тем самым была возможность обнаруживать риск множественных хромосомных аномалий, а также, например, пол плода, причем все это в рамках одного анализа образца. Например, в одном варианте реализации данного изобретения пул ПЦЗ содержит различные последовательности, которые связываются с хромосомами 13, 18, 21 и X или с хромосомами 13, 18, 21, X и Y.

В различных вариантах реализации данного изобретения содержание GC в ПЦЗ составляет от 19 до 80%, от 19 до 70%, от 19 до 60%, от 19 до 50%, от 19 до 49%, от 19 до 48%, от 19 до 47%, от 19 до 46%, от 19 до 45%, от 19 до 44%, от 19 до 43%, от 19 до 42%, от 19 до 41% или от 19 до 40%.

В различных вариантах реализации данного изобретения каждая последовательность в пуле ПЦЗ составляет от 100 до 350 пар оснований, 150-260 пар оснований, 100-200 пар оснований или 200-260 пар оснований в длину. В одном варианте реализации данного изобретения каждая последовательность в пуле ПЦЗ составляет 250 пар оснований в длину.

В различных вариантах реализации данного изобретения пул ПЦЗ может содержать 800 или более, 1500 или более отличающихся последовательностей, 2000 или более отличающихся последовательностей, 2500 или более отличающихся последовательностей или 3000 или более отличающихся последовательностей. В одном варианте реализации данного изобретения пул ПЦЗ содержит 1600 отличающихся последовательностей.

В одном варианте реализации данного изобретения в результате секвенирования обогащенной библиотеки получают данные о глубине считывания локусов, обнаруживаемых на хромосоме, представляющей интерес, и глубине считывания эталонных локусов, при этом статистический анализ выполняют с применением алгоритма, который последовательно анализирует глубину считывания локусов обнаруживаемых на хромосоме, представляющей интерес, по сравнению с глубиной считывания для эталонных локусов, в результате чего обнаруженные различия могут указывать на наличие генетических вариантов. Этапы алгоритма могут включать, но не ограничиваются, (а) удаление локусов, секвенированных ненадлежащим образом; (b) уменьшение погрешности показателей содержания GC и (c) классификацию статуса плоидности. В одном варианте реализации данного изобретения погрешность содержания GC уменьшают путем группирования локусов с сопоставимым содержанием GC.

В другом варианте реализации данного изобретения секвенирование обогащенной библиотеки обеспечивает размер фрагментов не содержащего клеток генетического материала, захваченного ПЦЗ, при этом статистический анализ выполняют с применением алгоритма, который сравнивает и сопоставляет распределение размеров фрагментов из анализируемых локусов и эталонных локусов, в результате чего различия в распределении указывают на наличие генетических вариантов. Этапы алгоритма могут включать, но не ограничиваются, (а) удаление резко выделяющихся размеров фрагментов; (b) создание бинарного распределения размеров фрагментов и (c) тестирование бинарного распределения размеров фрагментов, происходящих из области, представляющей интерес, по сравнению с соответствующим распределением эталонных локусов, с целью классификации состояния плоидности.

Как правило, статус плоидности классифицируют путем применения одного или более статистических методов. Например, статистический метод можно выбрать из группы, состоящей из t-критерия, двумерного непараметрического бутстреп-критерия, стратифицированного критерия перестановки и биномиального критерия пропорций и/или их комбинаций. В одном варианте реализации данного изобретения к образцу применяют все четыре вышеупомянутых статистических метода. Как правило, в результате применения статистического метода получают балльную оценку для смешанного образца, при этом, когда балльная оценка для смешанного образца является выше эталонного порогового значения, подозревают риск хромосомной аномалии в ДНК плода. Способ по данному изобретению может дополнительно включать оценку фракции ДНК плода в смешанном образце.

В другом варианте реализации данного изобретения статистический метод может быть выбран из группы, состоящей из t-критерия, двумерного непараметрического бутстреп-критерия и стратифицированного критерия перестановки. В одном варианте реализации данного изобретения к образцу применяют все вышеупомянутые статистические методы. Как правило, в результате применения статистического метода получают балльную оценку для смешанного образца, при этом, когда балльная оценка для смешанного образца является выше эталонного порогового значения, подозревают риск хромосомной аномалии в ДНК плода. Способ по данному изобретению может дополнительно включать оценку фракции ДНК плода в смешанном образце.

В другом аспекте данного изобретения предлагаются наборы для выполнения способа по данному изобретению. В одном варианте реализации данного изобретения набор содержит контейнер, содержа-

щий пул ПЦЗ, и инструкции для выполнения указанного способа. В различных других вариантах реализации данного изобретения набор содержит дополнительные компоненты для выполнения других этапов указанного способа.

Краткое описание графических материалов

На фиг. 1 приведена схема мультиплексного параллельного анализа целевых геномных областей с целью неинвазивного пренатального тестирования с применением последовательностей целевого захвата (ПЦЗ).

Фиг. 2 представляет собой перечень типовых хромосомных областей для амплификации ПЦЗ, которые связываются с хромосомами 13, 18, 21 или X.

Фиг. 3 представляет собой график балльной оценки 98 образцов материнской крови, подвергнутых мультиплексному параллельному анализу целевых геномных областей с применением ПЦЗ для обнаружения риска трисомии 21, при этом различия в средней глубине считывания условно парных групп анализировали на статистическую значимость, используя формулу t-критерия (называемого в данном документе статистическим методом 1).

Фиг. 4 представляет собой график балльной оценки 98 образцов материнской крови, подвергнутых мультиплексному параллельному анализу целевых геномных областей с применением ПЦЗ для обнаружения риска трисомии 21, анализируемого с помощью двумерного непараметрического бутстреп-критерия (называемого в данном документе статистическим методом 2).

Фиг. 5 представляет собой график балльной оценки 98 образцов материнской крови, подвергнутых мультиплексному параллельному анализу целевых геномных областей с применением ПЦЗ для обнаружения риска трисомии 21, анализируемого с помощью стратифицированного критерия перестановки (называемого в данном документе статистическим методом 3).

Фиг. 6 представляет собой график балльной оценки 98 образцов материнской крови, подвергнутых мультиплексному параллельному анализу целевых геномных областей для обнаружения риска трисомии 21, анализируемого с помощью биномиального критерия пропорций по размерам фрагментов (называемого в данном документе статистическим методом 4).

Фиг. 7 представляет собой график взвешенных балльных оценок 98 образцов материнской крови, полученных в результате анализа с применением статистических методов 1, 2, 3 и метода взвешенной балльной оценки 1, как показано на фиг. 3-5.

Фиг. 8 представляет собой график взвешенных балльных оценок 98 образцов материнской крови, полученных в результате анализа с применением статистических методов 1, 2, 3, 4 и метода взвешенной балльной оценки 1, как показано на фиг. 3-6.

Фиг. 9 представляет собой график альтернативного взвешенного подхода, метод взвешенной балльной оценки 2, относительно балльных оценок 98 образцов материнской крови, полученных в результате анализа с применением статистических методов 1, 2 и 3, как показано на фиг. 3-5.

Фиг. 10 представляет собой график балльной оценки 9 синтезированных образцов, подвергнутых мультиплексному параллельному анализу целевых геномных областей с применением ПЦЗ для 7q11.23, проанализированных с помощью варианта статистического метода 1 для обнаружения микроделаций.

Фиг. 11 представляет собой график балльной оценки 9 синтезированных образцов, подвергнутых мультиплексному параллельному анализу целевых геномных областей с применением ПЦЗ для 7q11.23, проанализированных с помощью варианта статистического метода 1 для обнаружения микродупликаций.

Подробное описание сущности изобретения

Изобретение относится к способу НИПТ, который включает гибридизационное обогащение выбранных целевых областей генома человека в мультиплексном панельном анализе с последующей количественной оценкой в сочетании с новым биоинформационным подходом и математическим анализом. В прошлом для обогащения конкретных областей, представляющих интерес, осуществляли обогащение гибридизацией в растворах перед секвенированием (см., например, Meyer, M and Kirchner, M. (2010), Cold Spring Harb. Protoc. 2010(6):pdbprot5448; Liao, G.J. et al. (2012), PLoS One 7:e38154; Maricic, T. et al. (2010), PLoS One 5:e14004; Tewhey, R. et al. (2009), Genome Biol. 10:R116; Tsangaras, K. et al. (2014), PLoS One 9:e109101). Однако для способов НИПТ по данному изобретению целевые последовательности, применяемые для обогащения конкретных областей, представляющих интерес, которые необходимы для обнаружения риска хромосомной аномалии, оптимизировали для максимальной эффективности, специфичности и точности. В геноме человека содержится много элементов, которые могут исказить и затруднять любой тип генетического анализа, тем самым подтверждая преимущество целевого подхода в НИПТ. Учитывая это, сложность генома человека и наличие этих искажающих элементов требуют тщательного проектирования последовательностей целевого захвата, применяемых для обогащения. Как описано в данном документе, разработаны оптимальные последовательности целевого захвата (ПЦЗ), которые обеспечивают более простое и более надежное НИПТ, сводя к минимуму риски ложноположительных и ложноотрицательных результатов, которые связаны с применением для НИПТ целого генома и неизбежного секвенирования искажающих элементов.

Способ по данному изобретению для тестирования риска хромосомной аномалии в хромосоме, представляющей интерес, в ДНК плода в смешанном образце материнской и ДНК плода, включает:

- (a) подготовку библиотеки последовательностей из смешанного образца;
- (b) гибридизацию библиотеки последовательностей с пулом последовательностей целевого захвата (ПЦЗ), при этом пул ПЦЗ содержит последовательности, которые связываются с одной или более хромосомами, представляющими интерес, и при этом:
 - (i) каждая последовательность в пуле составляет 100-260 пар оснований в длину, причем каждая последовательность имеет 5'-конец и 3'-конец,
 - (ii) каждая последовательность в пуле связывается с хромосомой(ами), представляющей интерес, по меньшей мере на расстоянии 150 пар оснований как на 5'-конце, так и на 3'-конце от областей, содержащих вариации числа копий (ВЧК), сегментные дублирования или повторяющиеся элементы ДНК, а также
 - (iii) содержание GC в ПЦЗ составляет от 19 до 50%;
- (c) выделение элементов библиотеки последовательностей, которые связываются с ПЦЗ для получения обогащенной библиотеки;
- (d) амплификацию и секвенирование обогащенной библиотеки; а также
- (e) проведение статистического анализа на выходе секвенирования обогащенных последовательностей библиотеки, чтобы тем самым определить риск хромосомной аномалии в ДНК плода.

Также включены наборы для выполнения способа по данному изобретению.

Различные аспекты этого изобретения более подробно описаны в следующих подразделах.

Проектирование последовательности целевого захвата.

В данном контексте термин "последовательности целевого захвата" или "ПЦЗ" относится к коротким последовательностям ДНК, которые являются комплементарными с областью (областями), представляющей интерес, на хромосоме (хромосомах) и которые применяются в качестве "приманки" для захвата и обогащения области, представляющей интерес, из большой библиотеки последовательностей, такой как библиотека последовательностей целого генома, полученная из образца материнской плазмы. Пул ПЦЗ применяется для обогащения, при этом последовательности в пуле оптимизированы в отношении (i) длины последовательностей; (ii) распределения ПЦЗ по области (областям), представляющей интерес; и (iii) содержания GC в ПЦЗ. Также оптимизировано количество последовательностей в пуле ПЦЗ (размер пула).

Обнаружено, что ПЦЗ, имеющие длину 100-260 пар оснований, являются оптимальными для максимизации эффективности обогащения.

В различных других вариантах реализации данного изобретения каждая последовательность в пуле ПЦЗ составляет 150-260 пар оснований, 100-200 пар оснований, 200-260 пар оснований или 100-350 пар оснований в длину. В предпочтительных вариантах реализации данного изобретения длина ПЦЗ в пуле составляет 250 или 260 пар оснований. Специалисту в данной области техники будет понятно, что небольшое изменение размера ПЦЗ обычно может применяться без изменения результатов (например, добавление или удаление нескольких пар оснований на обоих концах ПЦЗ); соответственно, длины пар оснований, приведенные в данном документе, должны рассматриваться с определением "около" или "приблизительно", что допускает небольшое изменение (например, 1-5%) по длине. Таким образом, например, длина "250 пар оснований" предназначена для обозначения "около 250 пар оснований" или "приблизительно 250 пар оснований", таким образом, термином охватываются, например, также 248 или 252 пары оснований.

Распределение ПЦЗ в каждой области или хромосоме, представляющей интерес, оптимизировали во избежание повторов с высокой степенью копий, повторов с низкой степенью копий и вариантов количества копий, и в то же время также для получения возможности нацеливания на информативные однонуклеотидные полиморфизмы (SNP) с тем, чтобы можно было определить как анеуплоидию, так и структурное изменение числа копий, а также осуществить оценку фракции эмбриона (ff). Соответственно, каждая последовательность в пуле ПЦЗ сконструирована таким образом, что 5'-конец и 3'-конец находятся по меньшей мере на расстоянии 150 пар оснований от областей в геноме, которые, как известно, содержат один или более следующих геномных элементов: вариации числа копий (ВЧК), сегментные дублирования и/или повторяющиеся элементы ДНК (такие как транспозированные элементы или области тандемного повтора). В различных других вариантах реализации данного изобретения каждая последовательность в пуле ПЦЗ сконструирована таким образом, что 5'-конец и 3'-конец находятся по меньшей мере на расстоянии 200, 250, 300, 400 или 500 пар оснований от областей в геноме, которые, как известно, содержат один или более из вышеупомянутых элементов.

Термин "вариации числа копий" является термином в области техники, который относится к форме структурных вариаций в геноме человека, при этом у разных индивидуумов могут быть изменения в ДНК генома, что у некоторых индивидуумов может привести к меньшему или большему количеству участка (ов) генома, по сравнению с нормальным количеством. ВЧК соответствуют относительно большому областям генома, которые могут быть удалены (например, участок, который обычно представляет собой А-В-С-Д, может стать А-В-Д) или могут быть продублированы (например, участок, который обычно представляет собой А-В-С-Д, может стать А-В-С-С-Д). На ВЧК приходится около 13% генома человека, причем каждый вариант варьируется в размере от около 1 килобазы до нескольких мегабаз.

Термин "сегментные дублирования" (также известный как "повторы с низкой степенью копий") также является термином в области техники, который относится к блокам ДНК, которые варьируются от около 1 до 400 килобаз в длину, встречаются на более чем одном участке в геноме и обычно характеризуются высоким уровнем (более 90%) идентичности последовательности. Сегментные дублирования описаны, например, в Eichler, E.E. (2001), *Trends Genet.* 17:661-669.

Термин "повторяющиеся элементы ДНК" (также известный как "повторная ДНК" или "повторяющаяся ДНК") также является термином в области техники, который относится к образцам ДНК, встречающимся в множестве копий по всему геному. Термин "повторяющийся элемент ДНК" охватывает терминальные повторы, тандемные повторы и рассеянные повторы, включая транспозируемые элементы. Повторяющиеся элементы ДНК в ССП обсуждаются дополнительно, например, в работе Todd, J. et al. (2012), *Nature Reviews Genet.* 13:36-46.

ПЦЗ спроектированы со специфическими характеристиками содержания GC с целью минимизации погрешности данных GC и получения возможности использовать обычный и инновационный подход к анализу данных. Установлено, что ПЦЗ с содержанием GC, составляющим 19-50%, достигают оптимального обогащения и лучше всего функционируют с эмбриональной внеклеточной ДНК. В пределах пула ПЦЗ различные последовательности могут иметь различное содержание % GC, хотя для их включения в пул, содержание % GC в каждой последовательности выбирают в пределах 19-50%. В некоторых случаях пул ПЦЗ может быть выбран таким образом, чтобы определить другой диапазон содержания % GC, который считается более пригодным для оценки конкретных генетических аномалий. Неограничивающие примеры различных диапазонов содержания % GC могут составлять от 19 до 75%, от 19 до 65%, от 19 до 55%, от 19 до 50%, от 19 до 49%, от 19 до 48%, от 19 до 47%, от 19 до 46, от 19 до 45%, от 19 до 44%, от 19 до 43%, от 19 до 42%, от 19 до 41% или от 19 до 40%.

Как описано более подробно ниже в отношении одного варианта осуществления анализа данных, после амплификации и секвенирования обогащенных последовательностей анализируемые локусы и эталонные локусы затем могут быть "сопоставлены" или сгруппированы вместе в соответствии с содержанием % GC в них (например, анализируемые локусы с содержанием % GC, составляющем 40%, сопоставляются с эталонными локусами с содержанием % GC, составляющем 40%). Следует понимать, что процедура сопоставления содержания % GC может допускать небольшое изменение в разрешенном диапазоне сопоставления % GC. В неограничивающем примере и со ссылкой на ранее описанный пример в тексте анализируемый локус с содержанием % GC, составляющим 40%, может быть сопоставлен с эталонными локусами с содержанием % GC в диапазоне от 39-41%, тем самым охватывая содержание % GC в анализируемом локусе в рамках пригодного диапазона.

Для подготовки пула ПЦЗ, имеющего оптимизированные критерии, изложенные выше, в отношении размера, размещения в геноме человека и содержания % GC, для анализа эталонного генома человека можно применять как ручные, так и компьютеризированные методы анализа, известные в данной области техники. В одном варианте реализации данного изобретения применяют полуавтоматический метод, когда области сначала проектируют вручную на основе конструкции эталонного генома 19 человека (hg19), следя за тем, чтобы избежать вышеупомянутых повторяющихся областей, а затем проверяют содержание GC с применением программного обеспечения, которое вычисляет содержание % GC каждой области на основе ее координат на конструкции эталонного генома 19 человека (hg19). В другом варианте реализации данного изобретения специально разработанное программное обеспечение применяют для анализа эталонного генома человека с целью определения пригодных областей ПЦЗ, которые соответствуют определенным критериям, таким как, но не ограничиваясь ими, содержание % GC, близость к повторяющимся областям и/или близость к другим ПЦЗ.

Количество ПЦЗ в пуле тщательно проверяли и корректировали для достижения наилучшего баланса между устойчивостью результата и стоимостью/пропускной способностью анализа. Пул обычно содержит не менее 800 или более ПЦЗ, но может включать больше, например 1500 или более ПЦЗ, 2000 или более ПЦЗ или 2500 или более ПЦЗ. Обнаружено, что оптимальное количество ПЦЗ в пуле составляет 1600. Специалисту в данной области техники будет понятно, что небольшое изменение размера пула обычно может быть реализовано без изменения результатов (например, добавление или удаление небольшого количества ПЦЗ); соответственно, размеры количества пула, приведенного в данном документе, должны рассматриваться как "приблизительные" или "примерные", что допускает небольшое изменение (например, 1-5%). Так, например, термин размер пула "1600 последовательностей" предназначен для обозначения "около 1600 последовательностей" или "приблизительно 1600 последовательностей", таким образом, термин включает также, например, 1590 или 1610 последовательностей.

Принимая во внимание вышеизложенное, в другом аспекте данного изобретения предлагается способ получения пула ПЦЗ для применения в способе по данному изобретению с целью обнаружения риска хромосомной и/или другой генетической аномалии, при этом способ получения пула ПЦЗ включает выбор областей в одной или более хромосомах, представляющих интерес и соответствующих вышеприведенным критериям (например, имеющих расстояние от вышеупомянутых повторяющихся последовательностей, составляющее по меньшей мере 150 пар оснований, на любом конце, и характеризующихся содержанием GC от 19 до 50%), получение праймеров, которые амплифицируют последовательности,

гибридирующиеся с выбранными областями, и амплификацию последовательности, причем каждая последовательность имеет длину 100-260 пар оснований.

Забор и подготовка образца.

Способы по данному изобретению выполняются на смешанном образце, который содержит как материнскую, так и эмбриональную ДНК. Обычно образец представляет собой образец материнской плазмы, хотя можно применять другие источники ткани, которые содержат как материнскую, так и эмбриональную ДНК. Материнскую плазму можно получить из образца периферической цельной крови беременной женщины, при этом плазму можно получить с помощью стандартных способов. С целью получения пригодного материала ДНК для анализа согласно способу по данному изобретению достаточно 2-4 мл плазмы. Общую, внеклеточную ДНК можно затем экстрагировать из образца с помощью стандартных способов, неограничивающие примеры которых включают протокол Qiasymphony (Qiagen), пригодный для выделения свободной ДНК плода, или любой другой ручной или автоматический способ экстракции, пригодный для выделения внеклеточной ДНК.

После выделения внеклеточную ДНК смешанного образца применяют для конструкции библиотеки последовательностей с целью достижения совместимости образца с технологией секвенирования по ходу транскрипции, такой как, но не ограничиваясь ею, секвенирование следующего поколения Illumina. Как правило, указанная технология включает лигирование адаптеров на концах фрагментов внеклеточной ДНК, с последующей амплификацией. Наборы для подготовки библиотеки последовательностей являются коммерчески доступными. Неограничивающий типовой протокол для подготовки библиотеки последовательностей подробно описан в Примере 1.

Обогащение с помощью гибридизации ПЦЗ.

Представляющую интерес область (области) на представляющей интерес хромосоме (хромосомах) обогащают с помощью гибридизации пула ПЦЗ с библиотекой последовательностей с последующим выделением этих последовательностей в пределах библиотеки последовательностей, которая связывается с ПЦЗ. Чтобы облегчить выделение желаемых обогащенных последовательностей, как правило, последовательности ПЦЗ модифицируют таким образом, что последовательности, которые гибридизуются с ПЦЗ, могут быть отделены от последовательностей, которые не гибридизуются с ПЦЗ. Как правило, это достигается путем фиксации ПЦЗ к твердой подложке. Это позволяет физически разделить те последовательности, которые связывают ПЦЗ, с теми последовательностями, которые не связывают ПЦЗ. Например, каждую последовательность в пуле ПЦЗ можно пометить биотином, и затем пул можно связать с гранулами, покрытыми биотинсвязывающим веществом, таким как стрептавидин или авидин. В предпочтительном варианте реализации данного изобретения ПЦЗ маркируют биотином и связывают с магнитными гранулами, покрытыми стрептавидином. Специалисту в данной области техники, однако, будет понятно, что в данной области техники известны и другие системы с аффинностью связывания, которые могут применяться вместо биотина-стрептавидина/авидина. Например, можно применять систему на основе антител, в которой ПЦЗ маркируют антигеном и затем связывают с гранулами, покрытыми антителами. Кроме того, ПЦЗ могут включать на одном конце тег последовательности и могут быть связаны с твердой подложкой посредством комплементарной последовательности на твердой подложке, которая гибридизуется с тегом последовательности. Кроме того, помимо магнитных гранул, можно применять другие типы твердых подложек, такие как полимерные гранулы и т.п.

В результате обогащения представляющей интерес последовательности (последовательностей) с помощью ПЦЗ формируется обогащенная библиотека, затем элементы обогащенной библиотеки элюируют из твердой подложки, амплифицируют и секвенируют с помощью стандартных способов, известных в данной области техники. Как правило, применяют технологию стандартного секвенирования следующего поколения Illumina, хотя можно также применять другие технологии секвенирования, которые обеспечивают очень точный подсчет в дополнение к информации о последовательности. Обнаружение генетических аномалий, таких как, но не ограничиваясь ими, анеуплоидии или изменения числа структурных копий, требует очень точного подсчета, при этом ССП представляет собой технологию, которая обеспечивает очень точный подсчет. Соответственно, для обнаружения генетических аномалий, таких как, но не ограничиваясь ими, анеуплоидии или изменения числа структурных копий, можно применять другие точные методы подсчета, такие как цифровая ПЦР и микроматричный анализ, вместо ССП. Неограничивающие типовые протоколы для амплификации и секвенирования обогащенной библиотеки подробно описаны в Примере 3.

Анализ данных.

Информацию, полученную из последовательности обогащенной библиотеки, анализируют с применением инновационного биоматематического/биостатистического анализа данных. В подходе к этому анализу используются характеристики ПЦЗ, а высокая эффективность захвата цели позволяет эффективно обнаруживать анеуплоидии или изменения числа структурных копий, а также другие типы генетических аномалий. Детали типового анализа подробно описаны в Примере 4. При проведении анализа на первом этапе секвенированные фрагменты ДНК образца выравнивают с эталонным геномом человека. Показатели ПК применяют для проверки свойств выровненных образцов и определения того, пригоден ли образец для классификации. Эти показатели ПК могут включать, но не ограничиваясь ими, анализ

моделей обогащения локусов, представляющих интерес, такой как, например, общая глубина секвенирования образца, целевой результат секвенирования образца, производительность ПЦЗ, ожидаемая погрешность показателей содержания GC и количественное определение фракции эмбриона. Для определения риска хромосомной аномалии в ДНК плода образца применяют инновационный алгоритм. Этапы алгоритма включают, но не ограничиваются ими, удаление локусов, секвенированных ненадлежащим образом, извлечение информации о глубине считывания и размере фрагмента в ПЦЗ-специфичных координатах, уменьшение погрешности генетических показателей (содержание GC) и классификацию статуса плоидности.

Определение статуса плоидности достигается с применением одного или более статистических методов, неограничивающие примеры которых включают t-критерий, бутстреп-критерий, критерий перестановки и/или биномиальный критерий пропорций и/или их комбинации. Специалисту в данной области техники будет понятно, что выбор и применение методов, необходимых для определения статуса плоидности, основывается на количестве доступных данных наблюдений. Таким образом, пригодность каждого метода определяется различными факторами, такими как, но не ограничиваясь ими, количество применяемых ПЦЗ и соответствующее применение с целью уменьшения погрешности показателей GC, если это применимо. Таким образом, вышеупомянутые методы следует рассматривать в качестве примеров типов статистического анализа, которые могут быть применены, но в тоже время не являются единственными методами, пригодными для определения статуса плоидности. Как правило, в результате применения статистического метода получают балльную оценку для смешанного образца, при этом, когда балльная оценка для смешанного образца является выше эталонного порогового значения, подозревают риск хромосомной аномалии в ДНК плода.

В частности, один из аспектов статистического анализа включает количественное определение и уменьшения погрешности показателей содержания GC. В дополнение к проблеме обнаружения небольших изменений сигнала в ДНК плода в смешанном образце (например, но не ограничиваясь этим, в дополнительном или меньшем генетическом материале из некоторых эмбриональных хромосомных областей), процесс секвенирования сам обуславливает определенные погрешности, которые могут затруднять обнаружение сигнала. Одной из таких погрешностей может характеризоваться показатель содержания GC, на основе которого осуществляется предпочтительное секвенирование/амплификация генетических областей. Таким образом, некоторые способы обнаружения, такие как, но не ограничиваясь ими, способы на основе глубины считывания, должны учитывать такую погрешность при анализе данных секвенирования. Таким образом, необходимо помнить о погрешности данных, и, следовательно, пригодные способы применять с их учетом, в результате чего зависимости генетического содержания не будут иметь возможности влиять на какие-либо статистические методы, которые могут применяться для количественной оценки риска генетической аномалии плода.

Например, одним из способов количественного определения погрешности содержания GC является применение метода локально взвешенного сглаживания диаграммы рассеяния (LOESS) для данных секвенирования. Каждый целевой локус можно определить по его результату глубины считывания при секвенировании и "содержанию GC" в нем. Линия наилучшего соответствия этим двум переменным для большого набора локусов обеспечивает оценку ожидаемой глубины считывания при секвенировании с учетом содержания GC. Как только этот этап количественного определения погрешности показателей содержания GC будет завершен, следующим этапом будет применение этой информации для учета возможных погрешностей в данных. Одним из способов является нормализация глубины считывания всех локусов относительно их ожидаемой глубине считывания (на основе содержания GC в каждом локусе). В принципе, это разъединяет данные глубины считывания и показатели генетического содержания, тем самым обеспечивая сопоставимость всех данных между собой. Таким образом, данные, которые извлекаются из областей с разным содержанием GC, например, но не ограничиваясь этим, из разных хромосом, теперь можно применять в последующих статистических анализах для обнаружения любых аномалий. Таким образом, с помощью процедуры LOESS, погрешности показателей содержания GC разъединяют с данными глубины считывания до проведения статистического анализа. В одном варианте реализации данного изобретения статистический анализ обогащенных библиотек последовательностей включает уменьшение погрешности показателей содержания GC с помощью процедуры LOESS.

В альтернативном предпочтительном варианте реализации данного изобретения погрешность показателя содержания GC определяют количественно и уменьшают путем группирования локусов с аналогичным (сопоставимым) показателем содержания GC. Таким образом, концептуально этот способ уменьшения погрешности показателей содержания GC состоит из трех этапов:

- 1) идентификация и расчет содержания GC в ПЦЗ;
- 2) уменьшение/учет погрешности показателей содержания GC с помощью различных процедур сопоставления/группирования ПЦЗ; а также
- 3) расчет риска любых генетических аномалий, которые могут присутствовать у плода с помощью статистических и математических методов на наборах данных, полученных на этапе 2.

При применении t-критерия набор данных разбивают на две группы: анализируемые локусы и эталонные локусы. Для каждой группы создают подмножества групп, в которых локусы классифицируют

согласно содержанию в них GC, как показано в неограничивающем примере в табл. 1 выборочных данных.

Таблица 1

GC	Глубина считывания эталонных локусов	Глубина считывания анализируемых локусов
40%	$x_1^{40}, x_2^{40}, \dots, x_{n_x}^{40}$	$y_1^{40}, y_2^{40}, \dots, y_{n_y}^{40}$
41%	$x_1^{41}, x_2^{41}, \dots, x_{n_x}^{41}$	$y_1^{41}, y_2^{41}, \dots, y_{n_y}^{41}$
42%	$x_1^{42}, x_2^{42}, \dots, x_{n_x}^{42}$	$y_1^{42}, y_2^{42}, \dots, y_{n_y}^{42}$
...

Специалисту в данной области техники будет понятно, что создание подгруппы может включать охват диапазона соответствующего содержания GC и/или подмножества локусов, которые определяются данным содержанием GC и/или диапазоном содержания GC. Соответственно, содержание % GC, приведенное в неограничивающем примере табл. 1, следует рассматривать как "приблизительное" или "примерное", что допускает небольшое изменение (например, 1-2%). Так, например, содержание % GC, составляющее "40%" может обозначать "около 40%" или "приблизительно 40%"; таким образом, термином охватывается, например, также содержание % GC "39-41%" в локусах, если это будет сочтено целесообразным.

Следовательно, когда речь идет о конкретном содержании GC, понятно, что подгруппы эталонных и анализируемых локусов могут содержать любое количество локусов, характеризующихся определенным % содержания GC и/или диапазоном.

Затем для каждой подгруппы содержания GC вычисляют репрезентативную глубину считывания. Для этого можно применять ряд методов, таких как, но не ограничиваясь ими, метод средних данных, метод медианных данных или варианты каждого метода. Таким образом, создают два вектора репрезентативной глубины считывания, при этом один соответствует эталонным локусам, а другой - анализируемым локусам (например, X_m , Y_m). В одном варианте реализации данного изобретения два вектора можно проанализировать один против другого для выявления значительных различий в глубине считывания. В другом варианте реализации данного изобретения различия двух векторов можно применять для оценки наличия значительных расхождений между анализируемым и эталонным локусом. Образцу присваивают балл согласно проведенному анализу.

Для статистического анализа с применением бутстреп-критерия набор данных разбивают на две группы: анализируемые локусы и эталонные локусы. Затем рассчитывают содержание GC в каждом локусе. После этого выполняют следующую процедуру.

Из эталонных локусов выбирают случайный локус; регистрируют его глубину считывания и содержание GC. Впоследствии из эталонных локусов выбирают случайный локус с единственным условием, заключающимся в том, что содержание GC в нем должно быть аналогично эталонному локусу. Регистрируют его глубину считывания. Специалисту в данной области техники будет понятно, что сходство содержания GC может охватывать диапазон пригодного содержания GC. Таким образом, указание на конкретное содержание % GC может рассматриваться как "приблизительное" или "примерное" или "находящееся в пригодном диапазоне" (например, 1-2%) и охватывать конкретное содержание % GC в исследовании. Таким образом, создают пару "эталонный локус-анализируемый локус" с аналогичным содержанием GC. Различия в паре "эталонный локус-анализируемый локус" записывают, например, E_1 . Затем локусы заменяют на соответствующие группы. Этот процесс повторяют до тех пор, пока не будет создана выборочная совокупность с повторениями (бутстреп) того же размера, что и количество существующих анализируемых ПЦЗ. Оценивают и регистрируют репрезентативную глубину считывания выборочной совокупности с повторениями, например, E_{mi} . Для этого можно применять ряд методов, таких как, но не ограничиваясь ими, метод средних величин, варианты или медианное значение вектора и/или их комбинации.

Процесс, описанный выше, повторяют столько раз, сколько необходимо, и создают распределение E_{mi} . Затем образцу присваивают балл, соответствующий процентилю этого распределения.

Для статистического анализа с применением критерия перестановки сначала набор данных сортируют на две группы: анализируемые локусы и эталонные локусы. В каждой группе создают подмножества групп, в которых локусы классифицируют согласно сходству по содержанию GC (см. столбцы 2 и 3 табл. 2 неограничивающих выборочных данных ниже). Также регистрируют количество локусов, присутствующих в каждой анализируемой подгруппе. Локусы анализируемой группы учитывают при вычислении оценки глубины считывания анализируемой группы, например, Y_{obs} . Для этого можно выбрать репрезентативное количество из каждой подгруппы содержания GC. Для осуществления оценки глубины считывания можно применять любое количество методов, таких как, но не ограничиваясь ими,

метод средних данных, метод медианных данных или варианты выбранных локусов.

Таблица 2

GC	Глубина считывания эталонных локусов	Глубина считывания анализируемых локусов	Ном. анализир уемого локуса	Объединение локусов
40%	$x_1^{40}, x_2^{40}, \dots, x_{nx40}^{40}$	$y_1^{40}, y_2^{40}, \dots, y_{ny40}^{40}$	ny40	$x_1^{40}, \dots, x_{nx40}^{40}, y_1^{40}, \dots, y_{ny40}^{40}$
41%	$x_1^{41}, x_2^{41}, \dots, x_{nx41}^{41}$	$y_1^{41}, y_2^{41}, \dots, y_{ny41}^{41}$	ny41	$x_1^{41}, \dots, x_{nx41}^{41}, y_1^{41}, \dots, y_{ny41}^{41}$
42%	$x_1^{42}, x_2^{42}, \dots, x_{nx42}^{42}$	$y_1^{42}, y_2^{42}, \dots, y_{ny42}^{42}$	ny42	$x_1^{42}, \dots, x_{nx42}^{42}, y_1^{42}, \dots, y_{ny42}^{42}$
...

Затем строили распределение для анализа Yobs с применением локусов независимо от их анализируемого или эталонного статуса следующим образом. Анализируемые или эталонные локусы каждой подгруппы содержания GC (см. последний столбец таблицы выборочных данных 2) объединяют для получения возможности проведения новой оценки глубины считывания. Из каждой объединенной подгруппы случайным образом выбирают количество локусов, при этом это количество ограничено верхним значением количества анализируемых локусов, учитываемых при исходном расчете Yobs (например, для содержания GC 40% и в контексте неограничивающих выборочных данных в табл. 2, это количество локусов может находиться в диапазоне [1, ny40]). Новую оценку глубины считывания проводят с учетом всех выбранных локусов. Процедуру повторяют столько раз, сколько необходимо, чтобы построить распределение эмпирических средних. Затем образцу присваивают балл, соответствующий позиции Yobs в этом распределении, используя подходящую трансформацию, которая учитывает моменты построенного распределения. Как и в случае с уже описанными способами, следует понимать, что незначительное изменение содержания % GC (например, 1-2%) является допустимым, если это будет сочтено целесообразным.

Следовательно, указание на конкретное содержание GC может расцениваться как "примерное" или "приблизительное", в результате чего, например, когда речь идет о содержании GC, составляющем 40%, локусы, которые характеризуются таким "примерным" или "приблизительным" содержанием 40% (например, 39-41%), могут учитываться при осуществлении способа.

В статистическом анализе с помощью биномиального критерия пропорций учитываются размеры фрагментов, выровненные по отношению к ПЦЗ-специфическим геномным координатам. Показано, что фрагменты внеклеточного генетического материала, происходящего из плаценты, имеют меньшую длину по сравнению с другим внеклеточным генетическим материалом (Chan, K.C. (2004), Clin. Chem. 50:88-92). Следовательно, представляет интерес статистически обоснованный ответ на вопрос, существуют ли достоверные различия в пропорции фрагментов небольшого размера, выровненных по отношению к ПЦЗ-специфической анализируемой области, от ожидаемой пропорции при сравнении ее с соответствующей пропорцией фрагментов других ПЦЗ-специфических эталонных областей, поскольку это будет указывать на генетические аномалии плода.

Таким образом, размеры фрагментов делят на две группы. Размеры, относящиеся к анализируемым локусам, определяют в одну группу, а размеры фрагментов, относящихся к эталонным локусам, определяют в другую группу. Впоследствии в каждой группе размеры фрагментов распределяют по двум подгруппам, в результате чего фрагменты небольшого размера относят в одну подгруппу, а все остальные фрагменты относят в оставшуюся подгруппу. На последнем этапе вычисляют пропорцию фрагментов небольшого размера в каждой группе и используют эти величины для расчета биномиального критерия пропорций. Образцу присваивают балл согласно проведенному анализу в исследовании.

Конечный результат образца может быть получен путем объединения одного или более показателей, полученных из разных статистических методов, неограничивающие примеры которых приведены в Примере 4.

Наборы по данному изобретению.

В другом аспекте данного изобретения предлагаются наборы для выполнения способов по данному изобретению. В одном варианте реализации данного изобретения набор содержит контейнер, содержащий пул ЦПЗ, и инструкции для выполнения указанного способа. В одном варианте реализации данного изобретения ПЦЗ предлагаются в форме, которая позволяет связать их с твердой подложкой, такие как биотинилированные ПЦЗ. В другом варианте реализации данного изобретения ПЦЗ предлагаются вместе с твердой подложкой, такие как биотинилированный ПЦЗ, поставляемые вместе с магнитными гранулами, покрытыми стрептавидином. В различных других вариантах реализации данного изобретения набор может содержать дополнительные компоненты для выполнения других аспектов указанного способа. Например, в дополнение к пулу ПЦЗ набор может содержать один или более из следующих элементов: (i) один или более компонентов для выделения внеклеточной ДНК из образца материнской плазмы (на-

пример, как описано в Примере 1); (ii) один или более компонентов для получения библиотеки последовательностей (например, праймеры, адаптеры, линкеры, рестрикционные ферменты, лигирующие ферменты, полимеразные ферменты и тому подобное, как подробно описано в Примере 1); (iii) один или более компонентов для амплификации и/или секвенирования обогащенной библиотеки (например, как описано в Примере 3); и/или (iv) программное обеспечение для выполнения статистического анализа (например, как описано в Примере 4).

Примеры

Данное изобретение дополнительно иллюстрируется следующими примерами, которые не следует рассматривать как дополнительно ограничивающие изобретение. Содержание всех ссылок, приложений, записей Genbank, патентов и опубликованных патентных заявок, приведенных по всему тексту, полностью включено в данное описание посредством ссылки во всей их полноте.

Пример 1. Сбор образцов и подготовка библиотеки.

Сбор образцов.

Образцы плазмы получали анонимно у беременных женщин после 10-й недели беременности. Протоколы, применяемые для сбора образцов для исследования, были одобрены Национальным комитетом по биоэтике Кипра, при этом от всех участников было получено информированное согласие.

Извлечение образцов.

Внеклеточную ДНК экстрагировали из 2-4 мл плазмы каждого индивидуума с помощью ручного или автоматизированного метода экстракции, пригодного для выделения внеклеточной ДНК, например, но не ограничиваясь этим, протокол Qiasymphony, пригодный для выделения свободной ДНК плода (Qiagen).

Подготовка библиотеки.

Извлеченную ДНК из образцов материнской плазмы применяли для конструкции библиотеки последовательностей. Применяли стандартные способы подготовки библиотеки со следующими модификациями (Meyer, M. and Kircher, M. (2010), Cold Spring Harb. Protoc. 2010(6): pdb prot5448). Библиотеку отрицательного контроля экстракции готовили отдельно для мониторинга любой контаминации, возникающей во время эксперимента. На этом этапе заполняли 5'- и 3'-"липкие" концы путем добавления 12 единиц полимеразы T4 (NEB), в то время как 5'-фосфаты присоединяли с помощью 40 единиц полинуклеотид-киназы T4 (NEB) в 100 мкл реакционной смеси с последующей инкубацией при 25°C в течение 15 мин, а затем при 12°C в течение 15 мин. Продукты реакции очищали с применением набора MinElute (Qiagen). Затем адаптеры P5 и P7 (см. подготовку адаптера) лигировали при разведении 1:10 на оба конца ДНК с помощью 5 единиц ДНК-лигазы T4 (NEB) в 40 мкл реакционной смеси в течение 20 мин при комнатной температуре с последующей очисткой с помощью набора MinElute (Qiagen). "Ники" удаляли в полученной реакции с помощью 16 единиц полимеразы Bst (NEB) в 40 мкл реакционной смеси с последующей инкубацией при 65°C в течение 25 мин, а затем при 12°C в течение 20 мин. Продукты реакции очищали с применением набора MinElute (Qiagen). Амплификацию библиотеки проводили с помощью полимеразы Fusion (ДНК-полимераза Herculanase II Fusion (Agilent Technologies) или полимеразы высокой точности Pfusion (NEB)) в 50 мкл реакционных смесей и со следующими условиями циклов: 95°C в течение 3 мин; затем 10 циклов при 95°C в течение 30 с, 60°C в течение 30 с, 72°C в течение 30 с и, наконец, 72°C в течение 3 мин. Конечные продукты библиотеки очищали с помощью набора для очистки MinElute Kit (Qiagen) и измеряли спектрофотометрически.

Подготовка адаптеров.

Отдельно готовили гибридизационные смеси для адаптеров P5 и P7 (26) и инкубировали в течение 10 с при 95°C с последующим снижением температуры от 95 до 12°C со скоростью 0,1°C/с. Реакционные смеси P5 и P7 объединяли для получения готовой к применению смеси адаптеров (100 мкМ каждого адаптера). Гибридизационные смеси готовили следующим образом: реакционная смесь P5, содержащая адаптер P5_F (500 мкМ) при конечной концентрации 200 мкМ, адаптер P5+P7 R (500 мкМ) при конечной концентрации 200 мкМ с 1X олиго-гибридизационным буфером. Кроме того, реакционная смесь P7 содержала адаптер P7_F (500 мкМ) при конечной концентрации 200 мкМ, адаптер P5+P7_R (500 мкМ) при конечной концентрации 200 мкМ с 1X олиго-гибридизационным буфером (30). Последовательности были следующими, где * = фосфоротиоатная связь (PTO) (Integrated DNA Technologies) (Meyer, M. and Kircher, M. (2010), Cold Spring Harb. Protoc. 2010(6): pdb prot5448):

адаптер P5_F:

A*C*A*C*TCTTCCCTACACGACGCTCTTCCG*A*T*C*T (SEQ ID NO: 1)

адаптер P7_F:

G*T*G*A*CTGGAGTTCAGACGTGTGCTCTTCCG*A*T*C*T (SEQ ID NO: 2),

адаптер_P5+P7_R:

A*G*A*T*CGGAA*G*A*G*C (SEQ ID NO: 3)

Пример 2. Разработка и подготовка последовательностей целевого захвата (ПЦЗ).

Специально разработанные ПЦЗ готовили для обнаружения полных или частичных хромосомных

аномалий хромосом 13, 18, 21, X, Y или любой другой хромосомы, а также других генетических аномалий, таких как, но не ограничиваясь ими, синдромы микроделеции/микродуликации, транслокации, инверсии, вставки и другие мутации точечного или малого размера. Геномные целевые локусы, применяемые для разработки ПЦЗ, выбирали на основе содержания в них GC и расстояния от повторяющихся элементов (минимум 150 п.о.). Размер ПЦЗ мог быть переменным. В одном варианте реализации способа ПЦЗ варьируются от 100-260 п.о. по размеру и генерируются с применением подхода на основе ПЦР, как описано ниже. ПЦЗ получали с помощью симплекс-полимеразной цепной реакции с применением стандартной Taq-полимеразы, праймеров, предназначенных для амплификации целевых локусов, и нормальной ДНК, применяемой в качестве матрицы. На фиг. 2 показаны хромосомные области, применяемые с целью разработки праймеров для амплификации пригодных локусов на хромосомах 13, 18, 21 и X, чтобы таким образом подготовить пул TACS для анализа хромосом 13, 18, 21 и X.

Все специально разработанные ПЦЗ создавали в следующих условиях цикла: 95°C в течение 3 мин; 40 циклов при 95°C в течение 15 с, 60°C в течение 15 с, 72°C в течение 12 с; и 72°C в течение 12 с с последующей верификацией с помощью электрофореза в агарозном геле и очисткой с применением стандартных наборов для ПЦР-очистки, таких как набор для очистки Qiaquick PCR Purification Kit (Qiagen) или набор для очистки NucleoSpin 96 PCR (Mackerey Nagel) или набор Agencourt AMPure XP для ПЦР-очистки (Beckman Coulter). Концентрацию измеряли с помощью Nanodrop (Thermo Scientific).

Пример 3. Гибридизация и амплификация ПЦЗ.

Биотинилирование ПЦЗ.

ПЦЗ готовили для гибридизации, как описано ранее (Maricic, T. et al. (2010), PLoS One 5:e14004) с небольшими изменениями, начиная с тупых концов, с применением набора Quick Bunt (NEB), и инкубировали при комнатной температуре в течение 30 мин. Затем продукты реакции очищали с применением набора MinElute (Qiagen) и лигировали с помощью адаптера биотина с применением набора Quick Ligation (NEB) в 40 мкл реакционной смеси при комнатной температуре в течение 15 мин. Продукты реакции очищали с применением набора MinElute (Qiagen) и денатурировали в однопочечную ДНК перед иммобилизацией на магнитных гранулах, покрытых стрептавидином (Invitrogen).

Гибридизация ПЦЗ.

Амплифицированные библиотеки смешивали с блокирующими олигонуклеотидами (Maricic, T., выше) (200 мкМ), 5 мкг ДНК Cot-1 (Invitrogen), 50 мкг ДНК спермы лосося (Invitrogen), буфером для гибридизации 2× Agilent, агентом блокирования 10× Agilent и нагревали при 95°C в течение 3 мин для денатурации цепей ДНК. После денатурации проводили 30-минутную инкубацию при 37°C для блокирования повторяющихся элементов и последовательностей адаптеров. Полученную смесь затем добавляли к биотинилированным ПЦЗ. Все образцы инкубировали во вращающемся инкубаторе в течение 12-48 часов при 66°C. После инкубации гранулы промывали, как описано выше, и ДНК элюировали с помощью нагревания (Maricic, T. выше). Элюированные продукты амплифицировали с применением наружно связанных праймеров адаптеров. Обогащенные амплифицированные продукты эквимольно объединяли и секвенировали на устройстве Illumina или любой другой пригодной платформе.

Пример 4. Биоинформационный анализ образцов.

Выравнивание генома человека.

Для анализа каждого образца применяли общепринятую процедуру биоинформационного принципа, описанную ниже, с целью выравнивания фрагментов ДНК последовательности образца относительно эталонного генома человека. Целевые спаренные концевые фрагменты считывания, полученные в результате ССП, обрабатывали для удаления последовательностей адаптеров и считывания низкого качества (Q-score <25) с помощью программного обеспечения cutadapt (Martin, M. et al. (2011), EMB.net Journal 17.1). Качество необработанных и/или обработанных данных считывания, а также любые описательные статистические данные, которые помогают в оценке проверки качества полученной последовательности образца, анализировали с помощью программного обеспечения FastQC (Babraham Institute (2015) FastQC) и/или другого специально разработанного программного обеспечения. Обработанные считывания, которые составляли по меньшей мере 25 оснований в длину, выравнивали относительно конструкции эталонного генома человека hg19 (UCSC Genome Bioinformatics) с помощью алгоритма выравнивания Burrows-Wheel (Li, H. and Durbin, R. (2009), Bioinformatics 25:1754-1760). Если было необходимо, дублирующиеся считывания удаляли после выравнивания. Там, где это применимо, полученные последовательности, относящиеся к одному и тому же образцу, но обработанные на отдельных линиях последовательностей, объединяли в один файл полученных последовательностей. Процедуры удаления дубликатов и объединения выполняли с помощью набора программного обеспечения Picard (Broad Institute (2015) Picard) и/или набора программного обеспечения Sambamba (Sambamba reference, Tarasov, Artem, et al. "Sambamba: fast processing of NGS alignment formats." Bioinformatics 31.12 (2015): 2032-2034.).

С помощью вышеупомянутого программного обеспечения получали окончательную выровненную версию секвенированного образца по отношению к эталонному геному человека, после чего все последующие этапы основывались на этой выровненной версии. Информацию относительно полиморфизмов коротких нуклеотидов (SNP) в локусах, представляющих интерес, получали с применением bcftools из

набора программного обеспечения SAMtools (Li, H. et al. (2009), *Bioinformatics* 25:2078-2079). Данные о глубине считывания на одно основание в локусах, представляющих интерес, получали с помощью функции `mpileup` из набора программного обеспечения SAMtools, упоминаемой в данном документе как файл `mpileup`. Информацию, относящуюся к размеру выровненных фрагментов, получали с помощью функции просмотра из набора программного обеспечения SAMtools, упоминаемой в данном документе как файл размера фрагмента.

Файл `mpileup` и файл размера фрагмента обрабатывали с применением интерфейсов прикладного программирования (API), созданных на языках программирования Python и R (Python Software Foundation, 2015) Python; The R Foundation (2015) The R Project for Statistical Computing). API применяли для определения состояния ploidy хромосом, представляющих интерес, с использованием ряда этапов (далее в совокупности называемых "алгоритмом"), а также для сбора дополнительной описательной статистики, которая предназначена для применения в качестве показателей проверки качества, таких как, но не ограничиваясь ими, количественное определение фракции эмбриона (далее в совокупности называемых как "показатели ПК"). API также можно использовать для оценки генетических аномалий из данных, полученных при применении описанного метода в случаях многоплодной беременности, а также других генетических аномалий, таких как, но не ограничиваясь ими, микроделеции, микродупликации, вариации числа копий, транслокации, инверсии, вставки, точечные мутации и мутационные подписи.

Показатели ПК.

Показатели ПК применяют для проверки свойств выровненных образцов и определения того, пригоден ли образец для классификации. Эти показатели представляли собой, но не ограничивались:

(a) обогащение образца. Профили обогащения свидетельствуют о том, характеризуется ли образец достаточным обогащением в локусах, представляющих интерес, в конкретном эксперименте по секвенированию (в данном документе обозначается как "запуск"). Чтобы оценить это, оценивают различные показатели, неограничивающие примеры которых представляют собой:

- (i) общая глубина целевого считывания в образце,
- (ii) результат целевого секвенирования образца по отношению к общим картированным считываниям,
- (iii) индивидуальные показатели ПЦЗ с точки зрения достигнутой глубины считывания,
- (iv) эксцесс и асимметрия обогащения отдельных ПЦЗ, а также
- (v) эксцесс и асимметрия обогащения, наблюдаемые во всех ПЦЗ.

Приведенные выше проверки также учитывают в отношении обогащения погрешности GC. Образцы, которые не соответствуют одному или более критериям, приведенным выше, помечают для дополнительной проверки перед классифицированием;

(b) фракцию эмбриона образца. Образцы с оцененной фракцией эмбриона, которая ниже определенного порога, не классифицируются.

Алгоритм.

Алгоритм представляет собой набор методов обработки данных, математических и статистических общепринятых моделей, представляющих собой ряд этапов. Этапы алгоритма направлены на определение статуса относительной ploidy хромосомы, представляющей интерес, по отношению ко всем другим хромосомам секвенированного образца и применяются для обнаружения целых или частичных хромосомных аномалий для хромосом 13, 18, 21, X, Y или любых других хромосом, а также других генетических аномалий, таких как, но не ограничиваясь ими, синдромы микроделеции/микродупликации и другие точечные или малые мутации. Таким образом, указанный алгоритм можно применять, но не ограничиваясь этим, для обнаружения целых или частичных хромосомных аномалий для хромосом 13, 18, 21, X, Y или любой другой хромосомы, а также других генетических аномалий, таких как, но не ограничиваясь ими, микроделеции, микродупликации, вариации числа копий, транслокации, инверсии, вставки, точечные мутации и другие мутационные подписи. С помощью указанного алгоритма выполняют, но не ограничиваясь этим, два типа оценок, один из которых относится к информации о глубине считывания для каждого образца, а другой - к распределению размеров фрагментов в ПЦЗ-специфических областях. Один или более статистических критериев могут быть ассоциированы с каждым типом оценки, неограничивающие примеры которых приведены в статистических методах, описанных в данном документе.

При проведении анализов, ассоциированных с глубиной считывания, алгоритм последовательно сравнивает глубину считывания локусов в каждой хромосоме, представляющей интерес (называемой в данном документе анализируемой хромосомой), с глубиной считывания всех других локусов (называемых в данном документе эталонными локусами), чтобы классифицировать ее статус ploidy. Для каждого образца эти этапы представляли собой, но не ограничивались:

(a) Удаление локусов, секвенированных ненадлежащим образом. Получали данные глубины считывания каждого локуса. Локусы, которые не достигали минимального количества считываний, считались недостаточно обогащенными и удалялись перед проведением последующих этапов;

(b) Уменьшение погрешности генетических данных (содержание GC). Процедура секвенирования вводит расхождение в глубине считывания в локусах, представляющих интерес, в зависимости от содержания в них GC. Для учета такой погрешности применяли новый подход, основанный на согласовании последовательностей, который повышает чувствительность и специфичность для выявления хромосом-

ных анеуплоидий. Идентифицировали содержание GC каждого локуса на анализируемой хромосоме, а аналогичные генетические локусы группировали вместе с целью формирования генетически сопоставимых групп. Процедуру повторяли для эталонных локусов.

Затем генетически сопоставимые группы из анализируемой хромосомы условно спаривали с их генетически сопоставимыми аналогами группы на эталонной хромосоме (хромосомах). Группы могли содержать любое количество членов. Затем условно сопоставимые группы использовали для оценки статуса плоидности анализируемых хромосом.

(с) Определение статуса плоидности. Определение статуса плоидности осуществляли с помощью одного статистического метода и/или взвешенной балльной оценки результатов следующих, но не ограничиваясь ими, статистических методов:

Статистический метод 1.

Различия в глубине считывания условно парных групп анализировали на статистическую значимость с помощью формулы t-критерия:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

где t - результат t-критерия,

\bar{x} - среднее значение различий условно парных групп,

μ - ожидаемая глубина считывания и приравненная к значению, которое представляет собой незначительные различия в глубине считывания между двумя группами,

s - стандартное отклонение различий условно парных групп и

n - длина вектора условно парных различий.

Величину t-критерия затем учитывали для идентификации доказательств, если таковые имелись, против нулевой гипотезы об аналогичной плоидности между эталонной и анализируемой хромосомами. В частности, $t \geq c1$ (где c1 - предопределенный порог, принадлежащий множеству всех положительных чисел) демонстрирует доказательства против нулевой гипотезы. Результаты анализа 98 материнских образцов для хромосомы 21, проведенного с помощью этого метода, приведены на фиг. 3.

Статистический метод 2. Двумерный непараметрический бутстреп-критерий.

Метод с применением бутстреп-критерия зависит от соотношения между случайными величинами X (глубина считывания эталонных локусов) и Y (глубина считывания анализируемых локусов). В данном изобретении авторы изобретения рассматривали глубину считывания "приманок" в эталонной группе (случайная величина, обозначаемая X) как независимую коварианту. Первый этап итеративной процедуры включал случайную выборку с заменой (бутстреппинг) глубин считывания локусов на эталонных хромосомах, т.е. (x1, g1), ..., (xn, gn), где параметр g является известным и обозначает содержание GC выбранной "приманки". Затем для каждой выбранной случайным образом эталонной приманки (xi, gi) создавали соответствующую глубину считывания для генетически сопоставимого локуса, т.е. (y1, g1), ..., (yn, gn). Таким образом, получали двумерные данные (x1, y1), (x2, y2), ..., (xn, yn), которые условно сопоставляли с содержанием GC (параметр gi). Различия между глубинами считывания генетически сопоставленных значений xi и yi, полученных в результате бутстреппинга, учитывали при вычислении статистических показателей, представляющей интерес, для каждой итерации. В одном варианте реализации данного изобретения эта статистическая оценка может включать, но не ограничиваясь этим, варианты, среднее или медианное значение зарегистрированных различий и/или их комбинации. Процедуру повторяли по мере необходимости с целью построения распределения статистических показателей, представляющих интерес, из этих различий. Образцу присваивали балл, соответствующий определенному процентилю построенного распределения (например, балл 5-го перцентиля). При нулевой гипотезе плоидности между хромосомами в эталонной и анализируемой группах не отличается. Таким образом, образцы, балл которых для определенной хромосомы превышал предопределенный порог, например c2, статистически классифицировались как таковые, что маловероятно будут иметь одну и ту же плоидность. Можно применять и другие статистические оценки. Результаты анализа 98 материнских образцов для хромосомы 21, проведенного с помощью этого метода, приведены на фиг. 4.

Статистический метод 3. Стратифицированный критерий перестановки.

Статистическими показателями, представляющим интерес, является оценка глубины считывания анализируемой хромосомы, обозначаемая \bar{Y}_{obs} , которая рассчитывается с учетом всех локусов генетически сопоставленных групп анализируемой хромосомы следующим образом:

$$\bar{Y}_{obs} = \frac{\sum_{j=1}^{j=T} \sum_{i=1}^{i=Nj} y_{ij}}{\sum_{j=1}^{j=T} Nj}$$

где y_{ij} представляет собой глубину считывания части i локуса генетически сопоставленной группы j (т.е. локусов, принадлежащих к определенной группе с учетом содержания GC),

Nj представляет собой количество анализируемых локусов в генетически сопоставленной группе j и

T представляет собой количество генетически сопоставленных групп.

Впоследствии строили нулевое распределение для анализируемой \hat{Y}_{obs} .

С этой целью для каждой группы j анализируемые и эталонные локусы объединяли (заменяемость по нулевой гипотезе) и для каждой группы j случайным образом осуществляли выборку до N_j раз без замены (стратифицированная перестановка). Тем самым создавали вектор значений, например, y_i и из этого среднего значения вектора, например, \hat{y}_i , рассчитывали. Процедуру повторяли по мере необходимости для построения нулевого распределения. В заключение, \hat{Y}_{obs} - студентизировали по отношению к нулевому распределению по формуле

$$Z_{Y_{obs}} = \frac{\hat{Y}_{obs} - \hat{Y}}{\sigma_Y}$$

где \hat{Y} , а также σ_Y представляют собой первый и квадратный корень второго момента всех перестановленных \hat{y}_i статистических значений.

Образцы, чьи $Z_{Y_{obs}}$ были больше, чем предопределенный порог, например c_3 , статистически менее вероятно имели аналогичную плоидность в эталонной и анализируемой группах. Результаты анализа 98 материнских образцов для хромосомы 21, проведенного с помощью этого метода, приведены на фиг. 5.

При проведении анализов, ассоциированных с размерами фрагментов, алгоритм вычисляет пропорцию фрагментов небольшого размера, обнаруженных в анализируемых локусах, и сравнивает их с соответствующей пропорцией в эталонных локусах, как описано в разделе "Статистический метод 4".

Статистический метод 4. Пропорции фрагментов в зависимости от размера.

Для каждого образца извлекали количество и размер фрагментов, выровненных относительно эталонного генома человека в соответствующих ПЦЗ-координатах. Затем данные фильтровали, чтобы удалить размеры фрагментов с резко отклоняющимися статистическими значениями с помощью метода обнаружения медианных резко отклоняющихся значений. В частности, фрагментами с резко отклоняющимися размерами определялись те фрагменты, размер которых был выше или ниже пороговых значений, F_{thr} согласно уравнению

$$F_{thr} = F_{median} \pm (X \times IQR)$$

где F_{median} представляет собой медианный размер фрагмента всех фрагментов образца;

X представляет собой переменную, которая может принимать значения из множества R^+ ;

IQR представляет собой межквартильный диапазон размеров фрагментов.

После этого применяют биномиальный критерий пропорций, чтобы проверить наличие подтверждающих доказательств против нулевой гипотезы H_0 , которая определяется как:

H_0 : пропорция небольших фрагментов анализируемой области не отличается от пропорции небольших фрагментов эталонной области.

В различных вариантах реализации данного изобретения небольшие фрагменты определяются как те фрагменты, размер которых меньше или равен подмножеству Z^+ т.е. имеет верхнюю границу - 160 п.о. Если определить множество всех ПЦЗ как T , тогда анализируемая область может быть любым собственным подмножеством S , которое определяет исследуемую область, а эталонная область является относительным дополнением к S в T . Например, в одном варианте реализации данного изобретения множество S определяется всеми ПЦЗ-захваченными последовательностями хромосомы 21, и, таким образом, эталонный набор определяется всеми ПЦЗ-захваченными фрагментами на эталонных хромосомах и/или других эталонных локусах.

Альтернативная гипотеза H_1 определяется как:

H_1 : пропорция небольших фрагментов анализируемой области не эквивалентна пропорции анализируемых фрагментов эталонной области.

Соответственно и с учетом поправки на непрерывность вычисляют следующий балл (Brown et al., Harrel):

$$W_{test} = \frac{\bar{p} - p_{ref}}{\sqrt{\frac{\bar{p}(1-\bar{p})}{N_{test}}}}$$

где

$$\bar{p} = \frac{(\bar{F} + 0.5)}{(N_{test} + 1)}$$

$$p_{ref} = \frac{[(F)_{ref} + 0.5]}{[(N)_{ref} + 1]}$$

\bar{F} - количество фрагментов небольшого размера в анализируемой области,

F_{ref} - количество фрагментов небольшого размера в эталонной области,

N_{test} - количество всех фрагментов в анализируемой области и

N_{ref} - количество всех фрагментов в эталонной области.

Для каждого образца алгоритм последовательно анализирует пропорцию фрагментов в зависимости

от размера в исследуемых областях (например, в хромосоме 21, хромосоме 18, хромосоме 13), но не ограничивается ими, по отношению к эталонным областям; таким, которые не являются исследуемыми во время анализа. Каждому образцу в каждом анализе присваивается определенный балл. Баллы выше установленного порога, например с4, свидетельствуют против нулевой гипотезы. Результаты анализа 98 материнских образцов из хромосомы 21, проведенного с помощью этого метода, приведены на фиг. 6.

Метод взвешенной балльной оценки 1.

В одном варианте реализации способа по данному изобретению взвешенная балльная оценка присваивалась каждому образцу s и вычислялась как взвешенная сумма всех статистических методов с помощью формулы

$$V_s(R, F) = z_1 \max\{R_s, F_s\} + (1 - z_1) \min\{R_s, F_s\}$$

где R_s представляет собой скорректированный, специфический для анализа балл, который получают из взвешенного вклада каждого статистического метода, связанного с глубиной считывания для образца s , и определяют как:

$$R_s = \frac{(\sum_i w_i S_{is} - \overline{R_r})}{\sigma_r}$$

$\overline{R_r}$ представляет собой медианное, специфическое для анализа значение, вычисляемое из вектора всех нескорректированных взвешенных балльных оценок глубины считывания, которые получают в результате одного цикла секвенирования,

σ_r представляет собой величину, кратную среднему квадратичному отклонению балльной оценки R , рассчитанной из эталонного набора 100 эуплоидных образцов,

термины $\max\{R_s, F_s\}$ а также $\min\{R_s, F_s\}$ обозначают максимальное и минимальное значения в скобках соответственно,

F_s представляет собой скорректированный, специфический для анализа балл, который получают в результате применения статистического метода, связанного с размером фрагментов, и определяют, как:

$$F_s = \frac{(W_{test} - \overline{R_f})}{\sigma_f}$$

где W_{test} представляет собой значение, как определено ранее,

$\overline{R_f}$ представляет собой специфическое для анализа медианное значение, вычисляемое из вектора всех нескорректированных статистических балльных оценок, связанных с фрагментами, которые получают из одного цикла секвенирования, и

σ_f представляет собой величину, кратную среднему квадратичному отклонению балльной оценки F , рассчитанной из эталонного набора 100 эуплоидных образцов.

Уникальный классификационный балл, являющийся меньшим от predetermined величины, указывает на отсутствие каких-либо свидетельств из полученных данных о том, что образец характеризуется значительным риском анеуплоидии. Результаты для 98 материнских образцов с применением метода взвешенной балльной оценки для подмножества методов приведены на фиг. 7, а по всем методам - на фиг. 8.

Метод взвешенной балльной оценки 2.

В другом варианте реализации способа взвешенная балльная оценка, полученная в результате применения статистических методов, описанных выше, использовалась для присвоения каждому образцу уникальной балльной оценки риска анеуплоидии по формуле

$$R(t, c) = \sum_{j=0}^{j=N} w_j \frac{t_j}{c_j}$$

где R представляет собой результат взвешенной балльной оценки,

w_j - вес, присвоенный методу j ,

t_j - наблюдаемый балл, полученный в результате применения метода j ,

c_j - порог метода j .

Уникальный классификационный балл, являющийся меньшим от predetermined величины, указывает на отсутствие каких-либо свидетельств из полученных данных о том, что образец характеризуется значительным риском анеуплоидии. Результаты для 98 материнских образцов с применением статистических методов 1-3 и метода взвешенной балльной оценки 2 приведены на фиг. 9.

Поскольку предполагалось, что все глубины считывания от "приманок" в эталонной группе генерируются из одной и той же популяции, и, для того чтобы иметь универсальный порог, с целью уменьшения специфических для анализа погрешностей также применяли специфические для анализа коррективы.

Вышеупомянутый метод(ы) также пригоден для обнаружения других генетических аномалий, таких как, но не ограничиваясь ими, субхромосомные аномалии. Неограничивающим примером является нефрагментированная частичная потеря хромосомного материала, приводящая к состоянию микроделеции, или нефрагментированный частичный прирост хромосомного материала, приводящий к состоянию мик-

родупликации. Известный генетический локус, подверженный обеим аномалиям, представляет собой 7q11.23. В одном варианте реализации статистического метода 1 образцы синтезированной плазмы с 5, 10 и 20% эмбрионального материала анализировали на предмет повышенного риска состояний микроделеции и/или микродупликации для генетического локуса 7q11.23. Результаты показаны на фиг. 10 для случаев микроделеции и на фиг. 11 - для случаев микродупликации.

При точечных мутациях применяют различные биномиальные критерии, которые учитывают оценку эмбриональной фракции образца f , глубину считывания минорного аллеля r и общую глубину считывания последовательности n . Два частых, но не ограничивающих примера включают оценку риска, когда генетическая аномалия является рецессивной точечной мутацией или доминантной точечной мутацией.

В неограничивающем примере рецессивной точечной мутации проверяемая нулевая гипотеза состоит в том, что и мать, и эмбрион являются гетерозиготными (частота минорного аллеля составляет 0,5) против альтернативного варианта, при котором эмбрион является гомозиготным (частота минорного аллеля составляет 0,5- $f/2$). Небольшое значение p , полученное при применении соответствующего критерия отношения правдоподобия, будет свидетельствовать против нулевой гипотезы. В неограничивающем примере доминантной точечной мутации проверяемая нулевая гипотеза заключается в том, что мать и эмбрион являются гомозиготными в данной позиции против альтернативного варианта, при котором только эмбрион является гетерозиготным в данной позиции. Небольшое значение p , полученное при применении соответствующего критерия отношения правдоподобия, будет свидетельствовать против нулевой гипотезы.

В дополнение к вышесказанному, были также разработаны способы определения пола эмбриона, с приведенными ниже неограничивающими примерами. В одном варианте реализации данного изобретения пол эмбриона определяли в образце с применением критерия Пуассона по формуле

$$Pr(r_y \leq k) = e^{-\lambda} \sum_{i=0}^{i=k} \frac{\lambda^i}{i!}$$

где $\lambda = \frac{fB\mu}{2}$ и f - оценка эмбриональной фракции образца,
 B - количество целевых последовательностей на хромосоме Y ,
 μ - глубина считывания образца,
 k - сумма считываний, полученных от всех целевых последовательностей B .

Нулевая гипотеза критерия Пуассона состояла в том, что образец был мужского пола. Значение $Pr(r_y)$ меньше порога c_y считалось достаточным доказательством для отклонения нулевой гипотезы, т.е. образец не был мужского пола. Если какой-либо параметр для вычисления $Pr(r_y)$ был недоступен, тогда пол образца классифицировали как н/д (недоступно для определения).

В другом варианте реализации данного изобретения пол эмбриона определяли с помощью данных средней глубины считывания целевых последовательностей на хромосоме Y . Если средняя глубина считывания целевых последовательностей превышала предопределенный порог, и при этом такой порог можно определить с учетом других образец-специфических характеристик, таких как глубина считывания и оценка фракции эмбриона, пол эмбриона классифицировали как мужской. Если средняя глубина считывания была ниже такого порога, то пол эмбриона образца классифицировали как женский.

Оценка фракции эмбриона.

Разработано несколько способов для оценки фракции эмбриона, которые могут быть применены к одноплодной и/или многоплодной беременности. Соответственно, в зависимости от типа беременности, оценку фракции эмбриона можно осуществить любым способом или в виде взвешенной оценки в результате проведения ряда и/или всех разработанных способов. Ниже приводятся некоторые не ограничивающие примеры.

В одном варианте реализации данного изобретения разработана технология машинного обучения на основе байесовского вывода для вычисления апостериорного распределения фракции ДНК плода с учетом количества аллелей в гетерозиготных локусах в материнской плазме при одноплодных беременностях. В рамках модели применяли три возможные информационные комбинации материнских/эмбриональных генотипов с целью идентификации тех значений фракции ДНК плода, которые в большей степени подтверждаются наблюдаемыми данными.

Пусть f обозначает фракцию ДНК плода. Если мать является гетерозиготной в данном геномном локусе, то генотип эмбриона может быть либо гетерозиготным, либо гомозиготным, что приводит к ожидаемым частотам минорных аллелей, составляющих 0,5 и 0,5- $f/2$ соответственно. Если мать является гомозиготной, а эмбрион является гетерозиготным, то ожидаемая частота минорных аллелей составит $f/2$. Способ Монте-Карло с применением цепи Маркова (алгоритм Метрополиса-Гастингса) (The R Foundation (2015), The R Project for Statistical Computing) применяли либо с неинформативными, либо с априорно информативными данными (т.е. включающими дополнительную информацию, такую как гестационный возраст, вес матери и т.д.) с целью получения последовательности случайных выборок из распределения апостериорной вероятности фракции ДНК плода, которая основана на конечной смешан-

ной модели.

В другом варианте реализации данного изобретения оценку фракции эмбриона рассчитывали только из эмбрион-специфического кластера частоты минорного аллеля (MAF), т.е. кластера, образующегося, когда мать является гомозиготной, а эмбрион - гетерозиготным относительно данного геномного локуса.

Предполагается, что среднее значение оценки фракции эмбриона обычно распределяется как $N(2\bar{x}, \sigma_{\bar{x}}^2)$, где \bar{x} представляет собой среднее значение эмбрион-специфического MAF, а $\sigma_{\bar{x}}$ представляет собой стандартное отклонение эмбрион-специфического MAF. Затем оценку фракции эмбриона получают из процентилей вычисленного распределения, $N(2\bar{x}, \sigma_{\bar{x}}^2)$.

При многоплодных беременностях, неограничивающие примеры которых включают монозиготную и дизиготную беременность двойней, беременность тройней и различные случаи применения донорских яйцеклеток и/или сперматозоидов, фракцию эмбриона можно оценить, используя информацию, полученную из гетерозиготных генетических локусов, значение MAF которых является ниже, чем порог, например, M_{thresh} , а также используя информацию, полученную из потенциальных эмбрион-специфических SNP. Обычному квалифицированному специалисту в данной области техники будет понятно, что эмбрион-специфические SNP могут происходить от любого эмбриона или от любой возможной комбинации эмбрионов или от всех эмбрионов беременности. Таким образом, разработан алгоритм, который оценивает эмбриональную фракцию эмбриона с наименьшим вкладом в общий состав эмбриона, с учетом комбинаторного вклада каждого эмбриона в значения MAF, которые определяют эмбрион-специфические SNP, а также допускает неоднородный вклад эмбрионального материала в общий эмбриональный состав материала, полученного из плазмы. Для этого в указанном алгоритме применяется двухэтапный подход.

В одном варианте реализации указанного алгоритма многоплодную беременность рассматривали как дизиготную беременность двойней. На первом этапе реализации модели алгоритма используются все информативные SNP и допускается неоднородный вклад эмбриона, который можно объяснить разницей в кратности оценок эмбриональной фракции заданного порога, например, cf. В частности, если $f1$ и $f2$ представляют собой эмбриональные фракции "эмбриона один" и "эмбриона два" и $f1 \geq f2$, то предположение состоит в том, что $f2 \geq cf \cdot f1$, где cf - положительное действительное число, большее или равное 1. Согласно этому предположению наблюдаемые данные D , определяемые как подсчеты альтернативных и эталонных аллелей в локусах информативных SNP, считаются генерируемыми из распределения комбинации трех биномов (определяемых параметрами $f1/2$, $f2/2$ и $(f1+f2)/2$), с апостериорным распределением $p(f1, f2|D)$, являющимся пропорциональным наблюдаемой модели, которую можно записать в виде $p(f1|f2, D) p(f2|D)$. Апостериорное распределение $p(f1, f2|D)$ анализировали с помощью алгоритма MCMC Метрополиса-Гастинга с использованием однородного априорных данных. Эмпирический квантильный подход применяли в массиве генерируемых данных для определения фракций эмбриона.

На втором этапе алгоритм обеспечивает выполнение алгоритма кластеризации на основе модели (моделирование конечной гауссовой смеси, скомпонованной с помощью алгоритма EM, R-пакет: `mclust`), чтобы определить, существует ли отдельный резко выделяющийся кластер SNP, который, как полагают, сосредоточен вокруг $f1/2$. Существование такого кластера со средним недействительным предположением $cf \geq f2/f1$ приводит к оценке $f1$ с применением только части SNP идентифицированного кластера.

Ссылки.

Chris Fraley and Adrian E. Raftery (2002). Model-based Clustering, Discriminant Analysis and Density Estimation. Journal of the American Statistical Association, 97:611-631

Chris Fraley, Adrian E. Raftery, T. Brendan Murphy, and Luca Scrucca (2012). `mclust` Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. Technical Report No. 597, Department of Statistics, University of Washington

Пример 5. Результаты анализа материнских образцов.

98 материнских образцов были проанализированы на риск трисомии хромосомы 21 (T21) согласно методик, описанных в Примерах 1-4. Значения балльной оценки обнаружения риска T21 для 98 образцов с применением статистических методов 1-4 приведены на графиках, проиллюстрированных на фиг. 3-6 соответственно. Каждая точка представляет балльную оценку для отдельного образца. Линия иллюстрирует порог "c" ($c=3,00$ для метода 1, $c=5,00$ для метода 2, $c=4,00$ для метода 3 и $c=0,91$ для метода 4). Образцы с баллами, превышающими пороговую линию в положительном направлении (т.е. балльная оценка $>c$, более темные точки), характеризовались высоким риском анеуплоидии, в отличие от образцов с баллами, которые не превышали пороговое значение (т.е. балльная оценка $<c$, более светлые точки). Согласно результатам всех статистических методов четыре образца были определены как характеризующиеся высоким риском анеуплоидии T21. Более того, все статистические методы указывали на те же самые образцы с высоким риском, даже несмотря на то, что балльные оценки соответствовали различным распре-

делениям в разных методах.

Взвешенные балльные оценки, полученные в результате комбинаций статистических методов 1-4 алгоритма (как описано в Примере 4), изображены на графиках на фиг. 7-9, где каждая точка представляет значение балльную оценку для отдельного образца, а линия представляет собой пороговое значение. Взвешенные балльные оценки также указывали на одни и те же образцы, характеризующиеся высоким риском.

Четыре образца, идентифицированные как имеющие высокий риск относительно T21 по результатам вышеописанных анализов, были независимо верифицированы как полученные от беременностей, имеющих эмбрион с трисомией 21, тем самым подтверждая точность метода анализа относительно риска T21.

Кроме того, результаты анализа синтезированных образцов микроделеции и микродупликации приведены на фиг. 10-11.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ анализа риска хромосомной аномалии в ДНК плода в смешанном образце ДНК матери и плода, при этом указанный способ включает:

(a) подготовку библиотеки последовательностей из смешанного образца;
 (b) гибридизацию библиотеки последовательностей с пулом последовательностей целевого захвата (ПЦЗ), при этом пул ПЦЗ содержит последовательности, которые связываются с одной или более хромосомами, представляющими интерес и содержащими хромосомную аномалию, и при этом:

(i) каждая последовательность в пуле составляет 100-260 пар оснований в длину, причем каждая последовательность имеет 5'-конец и 3'-конец,

(ii) каждая последовательность в пуле связывается с хромосомой(ами), представляющей интерес, по меньшей мере на расстоянии 150 пар оснований как на 5'-конце, так и на 3'-конце от областей, содержащих вариации числа копий (ВЧК), удвоения сегментов или повторы элементов ДНК, а также

(iii) содержание GC в ПЦЗ составляет от 19 до 50%;

(c) выделение элементов библиотеки последовательностей, которые связываются с ПЦЗ для получения обогащенной библиотеки;

(d) амплификацию и секвенирование обогащенной библиотеки; а также

(e) проведение статистического анализа обогащенных последовательностей библиотеки, где риск присутствия хромосомной аномалии в ДНК плода определяют по результатам указанного анализа.

2. Способ по п.1, отличающийся тем, что хромосомная аномалия представляет собой анеуплоидию.

3. Способ по п.1, отличающийся тем, что хромосому(ы), представляющую интерес, выбирают из группы, состоящей из хромосом 13, 18, 21, X и Y.

4. Способ по п.1, отличающийся тем, что хромосомная аномалия представляет собой трисомию 21.

5. Способ по п.1, отличающийся тем, что хромосомная аномалия представляет собой структурную аномалию, включая, но не ограничиваясь этим, изменения количества копий, включая микроделеции и микродупликации, вставки, транслокации, инверсии и мутации небольшого размера, включая точечные мутации и мутационные подписи.

6. Способ по п.1, отличающийся тем, что смешанный образец представляет собой образец материнской плазмы.

7. Способ по п.1, отличающийся тем, что пул ПЦЗ фиксируют на твердой подложке или он может быть свободным в растворе.

8. Способ по п.7, отличающийся тем, что ПЦЗ являются биотинилированными и связанными с магнитными гранулами, покрытыми стрептавидином, или могут быть свободными в растворе.

9. Способ по п.1, отличающийся тем, что пул ПЦЗ связывается с множеством хромосом, представляющих интерес, тем самым обеспечивая возможность обнаруживать множество хромосомных аномалий.

10. Способ по п.1, отличающийся тем, что пул ПЦЗ содержит различные последовательности, которые связываются с хромосомами 13, 18, 21, X и Y.

11. Способ по п.1, отличающийся тем, что содержание GC в ПЦЗ составляет от 19 до 46%.

12. Способ по п.1, отличающийся тем, что содержание GC в ПЦЗ составляет от 19 до 43%.

13. Способ по п.1, отличающийся тем, что пул ПЦЗ содержит 800 или более отличающихся последовательностей.

14. Способ по п.1, отличающийся тем, что пул ПЦЗ содержит 1500 или более отличающихся последовательностей.

15. Способ по п.1, отличающийся тем, что пул ПЦЗ содержит 1600 отличающихся последовательностей.

16. Способ по п.1, отличающийся тем, что пул ПЦЗ содержит 2000 или более отличающихся последовательностей.

17. Способ по п.1, отличающийся тем, что пул ПЦЗ содержит 2500 или более отличающихся после-

довательностей.

18. Способ по п.1, отличающийся тем, что пул ПЦЗ содержит 20000 или более отличающихся последовательностей.

19. Способ по п.1, отличающийся тем, что каждая последовательность в пуле ЦПЗ составляет от 150 до 260 пар оснований в длину.

20. Способ по п.1, отличающийся тем, что в результате секвенирования обогащенной библиотеки получают данные о глубине считывания хромосомы, представляющей интерес, и глубине считывания эталонных локусов, при этом статистический анализ включает применение алгоритма, который последовательно анализирует глубину считывания локусов, обнаруживаемых на хромосоме, представляющей интерес, по сравнению с глубиной считывания эталонных локусов, причем алгоритм содержит этапы (а) удаления локусов, секвенированных ненадлежащим образом; (б) уменьшения погрешности показателей содержания GC и (с) классификации статуса ploидности.

21. Способ по п.20, отличающийся тем, что погрешность содержания GC уменьшают путем группирования локусов с сопоставимым содержанием GC.

22. Способ по п.20, отличающийся тем, что определение статуса ploидности осуществляют путем применения одного или более статистических методов.

23. Способ по п.1, отличающийся тем, что в результате секвенирования обогащенной библиотеки получают данные о количестве и размере секвенированных фрагментов для ПЦЗ-специфичных координат, а статистический анализ включает применение алгоритма, который последовательно анализирует пропорцию фрагментов в зависимости от размера в хромосоме, представляющей интерес, по отношению к пропорции фрагментов в зависимости от размера в эталонных локусах, причем алгоритм содержит этапы, направленные на (а) удаление фрагментов с резко выделяющимися размерами; (б) расчет пропорции фрагментов в зависимости от размера и (с) определение статуса ploидности.

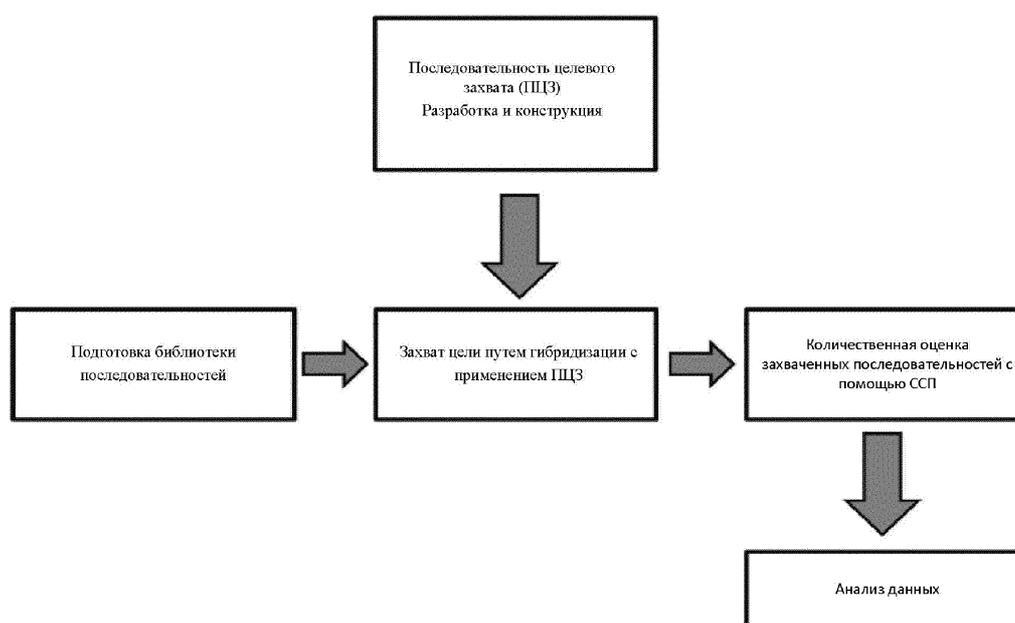
24. Способ по п.23, отличающийся тем, что определение статуса ploидности осуществляют путем применения одного или более статистических методов.

25. Способ по п.1, отличающийся тем, что статистический метод(ы) выбирают из группы, состоящей из t-критерия, двумерного непараметрического бутстреп-критерия, стратифицированного критерия перестановки и/или биномиального критерия пропорций.

26. Способ по п.22 или 24, отличающийся тем, что классификацию уровня ploидности осуществляют путем применения t-критерия, двумерного непараметрического бутстреп-критерия, стратифицированного критерия перестановки и/или биномиального критерия пропорций.

27. Способ по п.20, отличающийся тем, что в результате применения статистического метода получают балльную оценку для смешанного образца, при этом, когда балльная оценка для смешанного образца является выше эталонного порогового значения, подозревают риск хромосомной аномалии в ДНК плода.

28. Способ по п.1, который дополнительно включает оценку фракции ДНК плода в смешанном образце.



Фиг. 1

Хромосома	старт	стоп	Хромосома	старт	стоп	Хромосома	старт	стоп
chr13	91049752	91050001	chr13	21245232	21245481	chr13	28077809	28078058
chr13	92177139	92177388	chr13	21251423	21251672	chr13	28212786	28213035
chr13	93071049	93071298	chr13	21276601	21276850	chr13	28302961	28303210
chr13	94326973	94327222	chr13	21296989	21297237	chr13	28325874	28326123
chr13	94415507	94415756	chr13	21321012	21321261	chr13	28395893	28396142
chr13	95600034	95600283	chr13	21330138	21330386	chr13	28549591	28549840
chr13	112910529	112910778	chr13	21348854	21349102	chr13	28592602	28592851
chr13	53328405	53328654	chr13	21351692	21351941	chr13	28608293	28608542
chr13	20533985	20534234	chr13	21356527	21356775	chr13	28617043	28617292
chr13	20535728	20535977	chr13	21361621	21361869	chr13	28622471	28622720
chr13	20544227	20544476	chr13	21396343	21396591	chr13	28623594	28623843
chr13	20567359	20567607	chr13	21399744	21399992	chr13	28941450	28941699
chr13	20577159	20577407	chr13	21417352	21417601	chr13	28958898	28959147
chr13	20598772	20599021	chr13	21422981	21423229	chr13	28960375	28960624
chr13	20600556	20600804	chr13	21607075	21607324	chr13	29003985	29004234
chr13	20610779	20611027	chr13	21631007	21631256	chr13	29052686	29052935
chr13	20621336	20621585	chr13	22114658	22114907	chr13	52293092	52293341
chr13	20624392	20624641	chr13	22130956	22131205	chr13	52302656	52302905
chr13	20625461	20625710	chr13	22189915	22190164	chr13	53286404	53286653
chr13	20632609	20632857	chr13	22270443	22270692	chr13	53306303	53306552
chr13	20635196	20635444	chr13	22396225	22396474	chr13	53497224	53497473
chr13	21127405	21127653	chr13	22609278	22609527	chr13	53548644	53548893
chr13	21163998	21164247	chr13	22691882	22692131	chr13	53652715	53652964
chr13	21170950	21171199	chr13	22765144	22765393	chr13	53861872	53862121
chr13	21172515	21172764	chr13	27175379	27175628	chr13	53897429	53897678
chr13	21173774	21174022	chr13	27405837	27406086	chr13	53938205	53938454
chr13	21182580	21182828	chr13	27646720	27646969	chr13	54626806	54627055
chr13	21203385	21203634	chr13	27674747	27674996	chr13	54740174	54740423
chr13	21217173	21217422	chr13	27969360	27969609	chr13	54753834	54754083
chr13	21228403	21228652	chr13	27978231	27978480	chr13	54768238	54768487
chr13	21237418	21237666	chr13	28003481	28003730	chr13	54773333	54773582
chr13	21239719	21239968	chr13	28073180	28073429	chr13	54910969	54911218
chr13	54917347	54917596	chr13	111956688	111956937	chr13	60654975	60655224
chr13	54926251	54926500	chr13	112517966	112518215	chr13	61151386	61151635
chr13	55009732	55009981	chr13	112531519	112531768	chr13	62064603	62064852
chr13	59785342	59785591	chr13	112986709	112986958	chr13	71479359	71479608
chr13	60086302	60086551	chr13	113000054	113000303	chr13	39233534	39233783
chr13	60668020	60668269	chr13	21391429	21391678	chr13	39281667	39281916
chr13	60707747	60707996	chr13	21394616	21394865	chr13	39343524	39343773
chr13	60975949	60976198	chr13	21404268	21404516	chr13	39444897	39445146
chr13	61177164	61177413	chr13	21618233	21618482	chr13	39469458	39469707
chr13	61495655	61495904	chr13	21619521	21619769	chr13	39486237	39486486
chr13	67720914	67721163	chr13	21629602	21629851	chr13	39541262	39541511
chr13	67799557	67799806	chr13	21654127	21654376	chr13	39597307	39597556
chr13	70101783	70102032	chr13	22215795	22216044	chr13	39662550	39662799
chr13	71512697	71512946	chr13	22227261	22227510	chr13	39681445	39681694
chr13	71821654	71821903	chr13	22243011	22243260	chr13	39707053	39707302
chr13	72405844	72406093	chr13	22255153	22255402	chr13	39791099	39791348
chr13	74087455	74087704	chr13	22639387	22639636	chr13	39882804	39883053
chr13	74192477	74192726	chr13	27183101	27183350	chr13	39887397	39887646
chr13	74260893	74261142	chr13	27387690	27387939	chr13	40023278	40023527
chr13	78218064	78218313	chr13	30466189	30466438	chr13	40038277	40038526
chr13	88033439	88033688	chr13	30977198	30977447	chr13	40071926	40072175
chr13	90954378	90954627	chr13	31061235	31061484	chr13	40138914	40139163
chr13	91290109	91290358	chr13	31205037	31205286	chr13	40171414	40171663
chr13	91874767	91875016	chr13	32625308	32625557	chr13	40405751	40406000
chr13	92125914	92126163	chr13	34428102	34428351	chr13	40483291	40483540
chr13	92308079	92308328	chr13	36054899	36055148	chr13	40562330	40562579
chr13	92971372	92971621	chr13	36105686	36105935	chr13	40590631	40590880
chr13	94063232	94063481	chr13	40676702	40676951	chr13	40633944	40634193
chr13	94453836	94454085	chr13	40745383	40745632	chr13	40644042	40644291
chr13	94713643	94713892	chr13	49281723	49281972	chr13	40661582	40661831
chr13	95766126	95766375	chr13	53624527	53624776	chr13	40667939	40668188
chr13	95814599	95814848	chr13	60069277	60069526	chr13	40684306	40684555

038117

chr13	40747946	40748195	chr13	47314093	47314342	chr13	30145894	30146143
chr13	40767281	40767530	chr13	47623498	47623747	chr13	30187253	30187502
chr13	41165017	41165266	chr13	48562058	48562307	chr13	30421518	30421767
chr13	41176869	41177118	chr13	48698166	48698415	chr13	30531947	30532196
chr13	41573543	41573792	chr13	48762184	48762433	chr13	30778718	30778967
chr13	41581284	41581533	chr13	48806515	48806764	chr13	30905760	30906009
chr13	41588844	41589093	chr13	48811339	48811588	chr13	31011560	31011809
chr13	41628316	41628565	chr13	49119202	49119451	chr13	31089936	31090185
chr13	41633944	41634193	chr13	49479756	49480005	chr13	31227324	31227573
chr13	41827812	41828061	chr13	49505070	49505319	chr13	31430437	31430686
chr13	42280338	42280587	chr13	49530727	49530976	chr13	31575060	31575309
chr13	42349979	42350228	chr13	50387096	50387317	chr13	31641685	31641934
chr13	42589616	42589865	chr13	50960526	50960775	chr13	31718167	31718416
chr13	42773088	42773337	chr13	51783208	51783457	chr13	31723797	31724046
chr13	42811933	42812182	chr13	51816719	51816968	chr13	32249897	32250146
chr13	44211957	44212206	chr13	51825241	51825490	chr13	32296888	32297137
chr13	44239246	44239495	chr13	52025153	52025402	chr13	32311823	32312072
chr13	44247433	44247682	chr13	52249504	52249753	chr13	32316395	32316644
chr13	44359516	44359765	chr13	29111011	29111260	chr13	32321163	32321412
chr13	44388992	44389241	chr13	29192994	29193243	chr13	32348262	32348511
chr13	44493443	44493692	chr13	29233772	29234021	chr13	32565108	32565357
chr13	44557886	44558135	chr13	29310139	29310388	chr13	32578964	32579213
chr13	44959162	44959411	chr13	29563140	29563389	chr13	32584041	32584290
chr13	45043950	45044199	chr13	29580073	29580322	chr13	32596887	32597136
chr13	45115235	45115484	chr13	29929175	29929424	chr13	32651705	32651954
chr13	45147386	45147635	chr13	29934475	29934724	chr13	32760468	32760717
chr13	47067063	47067312	chr13	29939640	29939889	chr13	32829312	32829561
chr13	47120595	47120844	chr13	29957853	29958102	chr13	32852622	32852871
chr13	47154937	47155186	chr13	29995732	29995981	chr13	32872003	32872252
chr13	47258834	47259083	chr13	30003375	30003624	chr13	33535420	33535669
chr13	47261849	47262098	chr13	30014954	30015203	chr13	33634729	33634978
chr13	47286641	47286890	chr13	30114022	30114271	chr13	33660366	33660615
chr13	33742309	33742558	chr13	39264689	39264941			
chr13	33760763	33761012	chr13	39265199	39265451			
chr13	33787480	33787729	chr13	39265501	39265748			
chr13	33803219	33803468	chr13	39266515	39266762			
chr13	33925858	33926107	chr13	39266915	39267157			
chr13	34286418	34286667	chr13	39322100	39322347			
chr13	34301780	34302029	chr13	39338010	39338257			
chr13	34380783	34381032						
chr13	35504924	35505173						
chr13	35532835	35533084						
chr13	35538772	35539021						
chr13	36233165	36233414						
chr13	36343466	36343715						
chr13	36348003	36348252						
chr13	36380590	36380839						
chr13	36386774	36387023						
chr13	36410707	36410956						
chr13	36427833	36428082						
chr13	36445018	36445267						
chr13	36451093	36451342						
chr13	36568336	36568585						
chr13	36576882	36577131						
chr13	36663599	36663848						
chr13	37595672	37595921						
chr13	38283224	38283473						
chr13	38320894	38321143						
chr13	38361974	38362223						
chr13	38380696	38380945						
chr13	38789453	38789702						
chr13	39009512	39009761						
chr13	39096011	39096260						
chr13	39202882	39203131						

Хромосома	старт	стоп	Хромосома	старт	стоп	Хромосома	старт	стоп
chrX	44004660	44004909	chrX	96475190	96475439	chrX	132160888	132161137
chrX	44005572	44005821	chrX	96522258	96522507	chrX	133560531	133560780
chrX	80622102	80622351	chrX	96685785	96686034	chrX	133847304	133847553
chrX	81576088	81576337	chrX	98226881	98227130	chrX	133886077	133886326
chrX	83458870	83459119	chrX	98227467	98227716	chrX	133886778	133887027
chrX	83461989	83462238	chrX	98314697	98314946	chrX	135279738	135279987
chrX	83972195	83972444	chrX	98331718	98331967	chrX	135581031	135581280
chrX	83973091	83973340	chrX	102986351	102986600	chrX	136941543	136941792
chrX	84308155	84308404	chrX	112117885	112118134	chrX	138913313	138913562
chrX	84412933	84413182	chrX	116461543	116461792	chrX	138913920	138914169
chrX	84530826	84531075	chrX	116746667	116746916	chrX	139857034	139857283
chrX	84531356	84531605	chrX	117121287	117121536	chrX	139857318	139857567
chrX	84561384	84561633	chrX	117494627	117494876	chrX	141400062	141400311
chrX	85583338	85583587	chrX	118400460	118400709	chrX	141400814	141401063
chrX	85584098	85584347	chrX	119760145	119760394	chrX	141408548	141408797
chrX	85643367	85643616	chrX	120436725	120436974	chrX	142517923	142518172
chrX	85643847	85644096	chrX	120437307	120437556	chrX	143466384	143466633
chrX	85644765	85645014	chrX	121854317	121854566	chrX	147766073	147766322
chrX	85682212	85682461	chrX	121855104	121855353	chrX	147869477	147869726
chrX	85715710	85715959	chrX	122863038	122863287	chrX	147880766	147881015
chrX	85764768	85765017	chrX	123203756	123204005	chrX	147881811	147882060
chrX	85988749	85988998	chrX	123989030	123989279	chrX	151796091	151796340
chrX	85989214	85989463	chrX	124394967	124395216	chrX	104102171	104102420
chrX	85997574	85997823	chrX	125455385	125455634	chrX	104124449	104124698
chrX	87225841	87226090	chrX	125500101	125500350	chrX	104234605	104234854
chrX	91517109	91517358	chrX	125500582	125500831	chrX	104235195	104235444
chrX	91517578	91517827	chrX	127973758	127974007	chrX	104291380	104291629
chrX	94601042	94601291	chrX	128274367	128274616	chrX	104291786	104292035
chrX	95282210	95282459	chrX	128274689	128274938	chrX	104647680	104647929
chrX	95700693	95700942	chrX	128275238	128275487	chrX	104648691	104648940
chrX	96016317	96016566	chrX	128694864	128695113	chrX	106886942	106887191
chrX	96474722	96474971	chrX	131201343	131201592	chrX	3343487	3343736
chrX	4013011	4013260	chrX	10710087	10710336	chrX	18600682	18600931
chrX	5008870	5009119	chrX	11139863	11140112	chrX	19376384	19376633
chrX	5687782	5688031	chrX	11658912	11659161	chrX	20026004	20026253
chrX	5717003	5717252	chrX	11681358	11681607	chrX	20627108	20627357
chrX	5808998	5809247	chrX	11686196	11686445	chrX	21537202	21537451
chrX	5814187	5814436	chrX	11686913	11687162	chrX	21629967	21630216
chrX	5836574	5836823	chrX	11687182	11687431	chrX	21887062	21887311
chrX	5849776	5850025	chrX	11917847	11918096	chrX	22202443	22202692
chrX	6895474	6895723	chrX	12016474	12016723	chrX	23018625	23018874
chrX	6965226	6965475	chrX	12158815	12159064	chrX	23019006	23019255
chrX	6965573	6965822	chrX	12485152	12485401	chrX	24521525	24521774
chrX	7152743	7152992	chrX	12608017	12608266	chrX	24522047	24522296
chrX	9529739	9529988	chrX	12724960	12725209	chrX	25401534	25401783
chrX	9605323	9605572	chrX	12841791	12842040	chrX	25403300	25403549
chrX	9615158	9615407	chrX	12904558	12904807	chrX	28689766	28690015
chrX	9766329	9766578	chrX	12937788	12938037	chrX	28693600	28693849
chrX	9793372	9793621	chrX	12938100	12938349	chrX	28717026	28717275
chrX	9849298	9849547	chrX	12992266	12992515	chrX	28720131	28720380
chrX	9850182	9850431	chrX	12995926	12996175	chrX	28808128	28808377
chrX	10002581	10002830	chrX	12997273	12997522	chrX	28824813	28825062
chrX	10301512	10301761	chrX	13293562	13293811	chrX	28825844	28826093
chrX	10354762	10355011	chrX	13338207	13338456	chrX	29487025	29487274
chrX	10375361	10375610	chrX	13340422	13340671	chrX	30847194	30847443
chrX	10384260	10384509	chrX	13351532	13351781	chrX	32660941	32661190
chrX	10398645	10398894	chrX	13624857	13625106	chrX	32949975	32950224
chrX	10572719	10572968	chrX	15525635	15525884	chrX	35971182	35971431
chrX	10579800	10580049	chrX	15543469	15543718	chrX	35971544	35971793
chrX	10580482	10580731	chrX	15543853	15544102	chrX	43599991	43600240
chrX	10627502	10627751	chrX	15844810	15845059	chrX	43600262	43600511
chrX	10645064	10645313	chrX	16860721	16860970	chrX	43807492	43807741
chrX	10669664	10669913	chrX	16861729	16861978	chrX	63137429	63137678
chrX	10709674	10709923	chrX	18307677	18307926	chrX	63137928	63138177

chr21	29841662	29841911	chr21	31733443	31733692	chr21	27623692	27623941
chr21	29884231	29884480	chr21	31768996	31769245	chr21	27763679	27763928
chr21	29898255	29898504	chr21	31886189	31886438	chr21	27894000	27894249
chr21	30368230	30368479	chr21	31886808	31887057	chr21	27926498	27926747
chr21	30413906	30414155	chr21	32119833	32120082	chr21	28077177	28077426
chr21	30415555	30415804	chr21	32125263	32125512	chr21	28122462	28122711
chr21	30428953	30429202	chr21	25795207	25795456	chr21	28161337	28161586
chr21	30477650	30477899	chr21	25802273	25802522	chr21	28180925	28181174
chr21	30528161	30528410	chr21	26355452	26355702	chr21	28275754	28276003
chr21	30550132	30550381	chr21	26365770	26366019	chr21	28291449	28291698
chr21	30580738	30580987	chr21	26373440	26373689	chr21	28294820	28295069
chr21	30600446	30600695	chr21	26374774	26375023	chr21	28295351	28295600
chr21	30628668	30628917	chr21	26390279	26390528	chr21	28295857	28296106
chr21	30660609	30660858	chr21	26390951	26391200	chr21	28311866	28312115
chr21	30680391	30680640	chr21	26402802	26403051	chr21	28334784	28335033
chr21	30720214	30720463	chr21	26496769	26497018	chr21	28354814	28355063
chr21	30782462	30782711	chr21	26515400	26515649	chr21	28415341	28415590
chr21	30810199	30810448	chr21	26571662	26571911	chr21	28433186	28433435
chr21	30849402	30849651	chr21	26589786	26590035	chr21	28450226	28450475
chr21	30868525	30868774	chr21	26604824	26605073	chr21	28474083	28474332
chr21	30988776	30989025	chr21	26933670	26933919	chr21	28491062	28491311
chr21	31012406	31012655	chr21	26950703	26950952	chr21	28509439	28509688
chr21	31032720	31032969	chr21	27045134	27045383	chr21	28534176	28534425
chr21	31085356	31085605	chr21	27398390	27398639	chr21	40042441	40042690
chr21	31494755	31495004	chr21	27441346	27441595	chr21	40427338	40427587
chr21	31560667	31560916	chr21	27465490	27465739	chr21	40509976	40510225
chr21	31605973	31606222	chr21	27485182	27485431	chr21	40531291	40531540
chr21	31621826	31622075	chr21	27486006	27486255	chr21	40578323	40578572
chr21	31653822	31654071	chr21	27502429	27502678	chr21	40601158	40601407
chr21	31670830	31671079	chr21	27520192	27520441	chr21	40630577	40630826
chr21	31691347	31691596	chr21	27539803	27540052	chr21	41280723	41280972
chr21	31708936	31709185	chr21	27623130	27623379	chr21	41607273	41607522
chr21	41629820	41630069	chr21	15843963	15844212	chr21	16265484	16265733
chr21	41703462	41703711	chr21	15858450	15858699	chr21	16273734	16273983
chr21	41729089	41729338	chr21	15859015	15859264	chr21	16275865	16276114
chr21	41761636	41761885	chr21	16002625	16002874	chr21	16290234	16290483
chr21	41839021	41839270	chr21	16030414	16030663	chr21	16295877	16296126
chr21	41908902	41909151	chr21	16063119	16063368	chr21	16296687	16296936
chr21	41933321	41933570	chr21	16068358	16068607	chr21	16298842	16299091
chr21	41981868	41982117	chr21	16071495	16071744	chr21	16316213	16316462
chr21	42029142	42029391	chr21	16081282	16081531	chr21	16319836	16320085
chr21	42030642	42030891	chr21	16115078	16115327			
chr21	42080190	42080439	chr21	16122085	16122334			
chr21	42251219	42251468	chr21	16130688	16130937			
chr21	42559559	42559808	chr21	16133865	16134114			
chr21	42601828	42602077	chr21	16139969	16140218			
chr21	42851340	42851589	chr21	16141592	16141841			
chr21	42872713	42872962	chr21	16160517	16160766			
chr21	43032728	43032977	chr21	16165861	16166110			
chr21	43406661	43406910	chr21	16189746	16189995			
chr21	46521300	46521549	chr21	16190806	16191055			
chr21	16323548	16323797	chr21	16192618	16192867			
chr21	16330249	16330498	chr21	16195809	16196058			
chr21	16332622	16332871	chr21	16198337	16198586			
chr21	16333315	16333564	chr21	16203731	16203980			
chr21	16335355	16335604	chr21	16207035	16207284			
chr21	16344361	16344610	chr21	16209341	16209590			
chr21	16349606	16349855	chr21	16211270	16211519			
chr21	16358650	16358899	chr21	16219907	16220156			
chr21	15713251	15713500	chr21	16230785	16231034			
chr21	15714357	15714606	chr21	16236026	16236275			
chr21	15725977	15726226	chr21	16241139	16241388			
chr21	15731366	15731615	chr21	16242812	16243061			
chr21	15816232	15816481	chr21	16246944	16247193			

Хромосома	старт	стоп	Хромосома	старт	стоп	Хромосома	старт	стоп
chr18	64634331	64634580	chr18	73871422	73871671	chr18	50308727	50308976
chr18	64645925	64646174	chr18	73876103	73876352	chr18	50309031	50309280
chr18	64975938	64976187	chr18	73879805	73880054	chr18	50338747	50338996
chr18	64996573	64996822	chr18	73884248	73884497	chr18	50423284	50423533
chr18	65066425	65066674	chr18	73931865	73932114	chr18	50424422	50424671
chr18	65073729	65073978	chr18	73938014	73938263	chr18	50424789	50425038
chr18	65074170	65074419	chr18	73943212	73943461	chr18	50448837	50449086
chr18	66343957	66344206	chr18	73947628	73947877	chr18	50449492	50449741
chr18	66344406	66344655	chr18	73964168	73964417	chr18	50514079	50514328
chr18	66585247	66585496	chr18	73966321	73966570	chr18	50519889	50520138
chr18	66596489	66596738	chr18	73974071	73974320	chr18	56064244	56064493
chr18	66601231	66601480	chr18	73985549	73985798	chr18	56148666	56148915
chr18	69966208	69966457	chr18	73987174	73987423	chr18	56149037	56149286
chr18	70089399	70089648	chr18	48466086	48466335	chr18	56244171	56244420
chr18	70093364	70093613	chr18	48928528	48928777	chr18	56245242	56245491
chr18	70107935	70108184	chr18	48929078	48929327	chr18	56247063	56247312
chr18	70198868	70199117	chr18	48951524	48951773	chr18	56527049	56527298
chr18	70200128	70200377	chr18	48951856	48952105	chr18	56528011	56528260
chr18	70200907	70201156	chr18	48959882	48960131	chr18	56532190	56532439
chr18	70674976	70675225	chr18	48960304	48960553	chr18	56532882	56533131
chr18	70682967	70683216	chr18	49004516	49004765	chr18	56534208	56534457
chr18	73541191	73541440	chr18	49006425	49006674	chr18	56601487	56601736
chr18	73548315	73548564	chr18	49017695	49017944	chr18	56633131	56633380
chr18	73605298	73605547	chr18	49148010	49148259	chr18	56635259	56635508
chr18	73704485	73704734	chr18	49149627	49149876	chr18	56719296	56719545
chr18	73752546	73752795	chr18	49149933	49150182	chr18	58037828	58038077
chr18	73764755	73765004	chr18	49150459	49150708	chr18	58038328	58038577
chr18	73768420	73768669	chr18	50239403	50239652	chr18	58039111	58039360
chr18	73775027	73775276	chr18	50242789	50243038	chr18	58040290	58040539
chr18	73788751	73789000	chr18	50243069	50243318	chr18	59424085	59424334
chr18	73851108	73851357	chr18	50243367	50243616	chr18	59424765	59425014
chr18	73854713	73854962	chr18	50247229	50247478	chr18	59441324	59441573
chr18	59499718	59499967	chr18	60401084	60401333	chr18	60974839	60975088
chr18	59500226	59500475	chr18	60412199	60412448	chr18	60984378	60984627
chr18	59507920	59508169	chr18	60420053	60420302	chr18	61055997	61056246
chr18	59543044	59543293	chr18	60434561	60434810	chr18	61088669	61088918
chr18	59556453	59556702	chr18	60449047	60449296	chr18	61090462	61090711
chr18	59614976	59615225	chr18	60452471	60452720	chr18	61100325	61100574
chr18	59625553	59625802	chr18	60456757	60457006	chr18	61154459	61154708
chr18	59631090	59631339	chr18	60467147	60467396	chr18	61162901	61163150
chr18	59631779	59632028	chr18	60469757	60470006	chr18	61170629	61170878
chr18	59632388	59632637	chr18	60514168	60514417	chr18	61197492	61197741
chr18	59695718	59695967	chr18	60537888	60538137	chr18	61213452	61213701
chr18	59696087	59696336	chr18	60544380	60544629	chr18	61213900	61214149
chr18	59696563	59696812	chr18	60544742	60544991	chr18	61216749	61216998
chr18	59712419	59712668	chr18	60545407	60545656	chr18	61232928	61233177
chr18	59756979	59757228	chr18	60611934	60612183	chr18	61236079	61236328
chr18	59770789	59771038	chr18	60617081	60617330	chr18	61253277	61253526
chr18	59805993	59806242	chr18	60617641	60617890	chr18	61259478	61259727
chr18	59813876	59814125	chr18	60630438	60630687	chr18	61264533	61264782
chr18	59855411	59855660	chr18	60630849	60631098	chr18	61289440	61289689
chr18	59909054	59909303	chr18	60639840	60640089	chr18	61368889	61369138
chr18	59909336	59909585	chr18	60640300	60640549	chr18	61370663	61370912
chr18	59910140	59910389	chr18	60776944	60777193	chr18	63685629	63685878
chr18	59910602	59910851	chr18	60813738	60813987	chr18	63711583	63711832
chr18	59973803	59974052	chr18	60814585	60814834	chr18	63721183	63721432
chr18	59996728	59996977	chr18	60938666	60938915	chr18	64379625	64379874
chr18	60010541	60010790	chr18	60939202	60939451	chr18	64398368	64398617
chr18	60011381	60011630	chr18	60950541	60950790	chr18	64410092	64410341
chr18	60240694	60240943	chr18	60953225	60953474	chr18	64412606	64412855
chr18	60378353	60378602	chr18	60958474	60958723	chr18	64417267	64417516
chr18	60384593	60384842	chr18	60960879	60961128	chr18	64421562	64421811
chr18	60384979	60385228	chr18	60965583	60965832	chr18	64426871	64427120
chr18	60389544	60389793	chr18	60970051	60970300	chr18	64436287	64436536

chr18	64593727	64593976	chr18	25279010	25279259	chr18	25682079	25682328
chr18	24962594	24962843	chr18	25283817	25284066	chr18	25689906	25690155
chr18	24964748	24964997	chr18	25292042	25292291	chr18	25700540	25700789
chr18	25003553	25003802	chr18	25295429	25295678	chr18	25703648	25703897
chr18	25033345	25033594	chr18	25330345	25330594	chr18	25707188	25707437
chr18	25037434	25037683	chr18	25332963	25333212	chr18	25716100	25716349
chr18	25052992	25053241	chr18	25333284	25333533	chr18	25722951	25723200
chr18	25080042	25080291	chr18	25345252	25345501	chr18	25725055	25725304
chr18	25124915	25125164	chr18	25384702	25384951	chr18	25748105	25748354
chr18	25128057	25128306	chr18	25510767	25511016	chr18	26214441	26214690
chr18	25140257	25140506	chr18	25519108	25519357	chr18	26236788	26237037
chr18	25140722	25140971	chr18	25531724	25531973	chr18	26261336	26261585
chr18	25147759	25148008	chr18	25543168	25543417	chr18	26265937	26266186
chr18	25157828	25158077	chr18	25547425	25547674	chr18	26285286	26285535
chr18	25160244	25160493	chr18	25557343	25557592	chr18	26289006	26289255
chr18	25166354	25166603	chr18	25564483	25564732	chr18	26308405	26308654
chr18	25172986	25173235	chr18	25583087	25583336	chr18	26308962	26309211
chr18	25174972	25175221	chr18	25585083	25585332	chr18	26312054	26312303
chr18	25175640	25175889	chr18	25589902	25590151	chr18	26316149	26316398
chr18	25182855	25183104	chr18	25601966	25602215	chr18	47100905	47101154
chr18	25198531	25198780	chr18	25611986	25612235	chr18	47130278	47130527
chr18	25201886	25202135	chr18	25617498	25617747	chr18	47366936	47367185
chr18	25209395	25209644	chr18	25629105	25629354	chr18	47371131	47371380
chr18	25213880	25214129	chr18	25631180	25631429	chr18	47689868	47690117
chr18	25216888	25217137	chr18	25643639	25643888	chr18	47690267	47690516
chr18	25228617	25228866	chr18	25645371	25645620	chr18	47703689	47703938
chr18	25231535	25231784	chr18	25665025	25665274	chr18	47708139	47708388
chr18	25245662	25245911	chr18	25666614	25666863	chr18	47717445	47717694
chr18	25246300	25246549	chr18	25667057	25667306	chr18	47738531	47738780
chr18	25254843	25255092	chr18	25670741	25670990	chr18	47777438	47777687
chr18	25265476	25265725	chr18	25673208	25673457	chr18	47794212	47794461
chr18	25278532	25278781	chr18	25681018	25681267	chr18	47799400	47799649
chr18	47802662	47802911	chr18	43918519	43918768	chr18	46599847	46600096
chr18	39874986	39875235	chr18	43918998	43919247	chr18	46601866	46602115
chr18	39891490	39891739	chr18	43919544	43919793	chr18	46983278	46983527
chr18	42204240	42204489	chr18	44086123	44086372	chr18	47013103	47013352
chr18	42313146	42313395	chr18	44094299	44094548	chr18	47016802	47017051
chr18	42313939	42314188	chr18	44167214	44167463	chr18	286722	286971
chr18	42314302	42314551	chr18	44167791	44168040	chr18	298393	298642
chr18	42349923	42350172	chr18	44168296	44168545	chr18	338127	338376
chr18	42407165	42407414	chr18	44173512	44173761	chr18	340427	340676
chr18	42418033	42418282	chr18	44174626	44174875	chr18	387135	387384
chr18	42449702	42449951	chr18	44857765	44858014	chr18	391175	391424
chr18	42450400	42450649	chr18	45110525	45110774	chr18	434440	434689
chr18	42461427	42461676	chr18	45197621	45197870	chr18	457527	457776
chr18	42462834	42463083	chr18	45197871	45198120	chr18	460384	460633
chr18	42468595	42468844	chr18	45198760	45199009	chr18	469281	469530
chr18	42468938	42469187	chr18	45234886	45235135	chr18	572716	572965
chr18	42644003	42644252	chr18	45235523	45235772	chr18	598352	598601
chr18	42644494	42644743	chr18	45238370	45238619	chr18	615413	615662
chr18	42645387	42645636	chr18	45335896	45336145	chr18	618280	618529
chr18	42741059	42741308	chr18	45336353	45336602	chr18	901111	901360
chr18	42745733	42745982	chr18	45357461	45357710	chr18	939383	939632
chr18	42746460	42746709	chr18	45357745	45357994	chr18	26321283	26321532
chr18	42746911	42747160	chr18	45358502	45358751	chr18	26337865	26338114
chr18	42747473	42747722	chr18	45362379	45362628	chr18	26345717	26345966
chr18	43796659	43796908	chr18	45397588	45397837	chr18	26365387	26365636
chr18	43820471	43820720	chr18	45864183	45864432	chr18	26367518	26367767
chr18	43841913	43842162	chr18	46464382	46464631	chr18	26392500	26392749
chr18	43842202	43842451	chr18	46473428	46473677	chr18	26406499	26406748
chr18	43842647	43842896	chr18	46474545	46474794	chr18	26711590	26711839
chr18	43845701	43845950	chr18	46481797	46482046	chr18	26729342	26729591
chr18	43846353	43846602	chr18	46483016	46483265	chr18	26733981	26734230
chr18	43917652	43917901	chr18	46589014	46589263	chr18	26739336	26739585

Хромосома	старт	стоп	Хромосома	старт	стоп	Хромосома	старт	стоп
chr18	11701840	11702094	chr18	13040411	13040657	chr18	29705097	29705347
chr18	11742882	11743124	chr18	13049051	13049304	chr18	29707000	29707248
chr18	11743713	11743964	chr18	13053013	13053260	chr18	29711514	29711758
chr18	11743974	11744216	chr18	13057128	13057370	chr18	29772771	29773020
chr18	11745797	11746044	chr18	13725324	13725566	chr18	29773038	29773291
chr18	11766505	11766752	chr18	13731669	13731919	chr18	29790110	29790357
chr18	11784594	11784841	chr18	13930816	13931068	chr18	29794196	29794446
chr18	11784857	11785107	chr18	13936510	13936761	chr18	29814455	29814703
chr18	11789414	11789659	chr18	13939030	13939277	chr18	29816677	29816924
chr18	11791405	11791649	chr18	13941500	13941742	chr18	29824628	29824872
chr18	11831898	11832147	chr18	18709626	18709868	chr18	29941738	29941986
chr18	11840071	11840318	chr18	19345013	19345263	chr18	29942404	29942657
chr18	11853018	11853271	chr18	22314388	22314639	chr18	29947043	29947286
chr18	12786330	12786572	chr18	22317204	22317445	chr18	29963606	29963859
chr18	12793790	12794034	chr18	22318584	22318826	chr18	29967321	29967565
chr18	12813342	12813590	chr18	22474179	22474425	chr18	29988241	29988491
chr18	12813796	12814046	chr18	22503094	22503336	chr18	29989060	29989304
chr18	12836508	12836756	chr18	22514422	22514675	chr18	33702205	33702458
chr18	12837826	12838071	chr18	22527670	22527920	chr18	33703293	33703541
chr18	12839370	12839619	chr18	22713619	22713863	chr18	33703805	33704052
chr18	12839810	12840054	chr18	22715692	22715939	chr18	33710269	33710511
chr18	12840789	12841033	chr18	22746015	22746258	chr18	33881399	33881647
chr18	12955546	12955788	chr18	22751334	22751581	chr18	33884883	33885133
chr18	12968393	12968643	chr18	26362055	26362304	chr18	33965911	33966161
chr18	12984202	12984449	chr18	27204825	27205070	chr18	33967910	33968160
chr18	12984953	12985198	chr18	27683185	27683429	chr18	35703018	35703269
chr18	12986018	12986263	chr18	28529366	28529610	chr18	35705739	35705986
chr18	13007427	13007681	chr18	28538550	28538803	chr18	36487627	36487876
chr18	13007815	13008064	chr18	28550611	28550855	chr18	52428207	52428460
chr18	13019314	13019562	chr18	28551683	28551928	chr18	52472800	52473050
chr18	13029804	13030046	chr18	28584849	28585102	chr18	52702246	52702490
chr18	13036834	13037081	chr18	29704799	29705048	chr18	52703207	52703449
chr18	26750685	26750934						
chr18	26752487	26752736						
chr18	26757573	26757822						
chr18	27244230	27244479						
chr18	35076347	35076596						
chr18	35147912	35148161						
chr18	35282655	35282904						
chr18	35292141	35292390						
chr18	35304247	35304496						
chr18	35311302	35311551						
chr18	36463709	36463958						
chr18	36473033	36473282						
chr18	36475403	36475652						
chr18	36479145	36479394						
chr18	36482038	36482287						
chr18	36483959	36484208						
chr18	36485761	36486010						
chr18	36487627	36487876						
chr18	36489135	36489384						
chr18	36531244	36531493						
chr18	36554006	36554255						
chr18	36558050	36558299						
chr18	36569619	36569868						
chr18	1017928	1018177						
chr18	241538	241787						
chr18	242831	243080						

Хромосома	старт	стоп	Хромосома	старт	стоп	Хромосома	старт	стоп
chr13	19520254	19520499	chr13	26939424	26939672	chr13	32759010	32759261
chr13	19523790	19524042	chr13	26939942	26940196	chr13	32784101	32784351
chr13	20531155	20531403	chr13	26952790	26953038	chr13	32784570	32784820
chr13	20542998	20543246	chr13	26956549	26956797	chr13	32786822	32787069
chr13	21182580	21182828	chr13	26957204	26957453	chr13	32787129	32787381
chr13	22255353	22255601	chr13	26957736	26957984	chr13	32811213	32811463
chr13	26795002	26795250	chr13	26971501	26971749	chr13	32931779	32932029
chr13	26795380	26795628	chr13	26974853	26975100	chr13	33016510	33016761
chr13	26809864	26810112	chr13	26976810	26977059	chr13	33016929	33017182
chr13	26810422	26810669	chr13	27117460	27117707	chr13	33017432	33017681
chr13	26810951	26811198	chr13	27216274	27216522	chr13	36342494	36342743
chr13	26811424	26811674	chr13	29111011	29111260	chr13	36343935	36344179
chr13	26820545	26820794	chr13	29233772	29234021	chr13	36344330	36344577
chr13	26824800	26825048	chr13	29310139	29310388	chr13	36345572	36345822
chr13	26825155	26825405	chr13	29563140	29563389	chr13	36346001	36346253
chr13	26827150	26827400	chr13	31499212	31499458	chr13	36349204	36349453
chr13	26827690	26827939	chr13	31499519	31499772	chr13	36354263	36354513
chr13	26829149	26829396	chr13	31609109	31609361	chr13	36358322	36358572
chr13	26829663	26829911	chr13	31653571	31653822	chr13	36361603	36361850
chr13	26830241	26830491	chr13	32149129	32149380	chr13	36362050	36362297
chr13	26830739	26830987	chr13	32169142	32169392	chr13	36362531	36362778
chr13	26834633	26834880	chr13	32173373	32173622	chr13	36363439	36363689
chr13	26836514	26836761	chr13	32271292	32271539	chr13	36366103	36366353
chr13	26843066	26843315	chr13	32316109	32316362	chr13	36387184	36387431
chr13	26851259	26851507	chr13	32316651	32316898	chr13	39584383	39584630
chr13	26864367	26864615	chr13	32339360	32339608	chr13	39585052	39585299
chr13	26888243	26888485	chr13	32359633	32359883	chr13	39585675	39585922
chr13	26888660	26888907	chr13	32433382	32433629	chr13	40090054	40090298
chr13	26898691	26898936	chr13	32434990	32435237	chr13	40093695	40093942
chr13	26925436	26925684	chr13	32600409	32600659	chr13	40098460	40098707
chr13	26926976	26927224	chr13	32653613	32653857	chr13	40098913	40099157
chr13	26938928	26939175	chr13	32758635	32758877	chr13	40101053	40101303
chr13	40101582	40101832	chr13	68311152	68311402	chr13	73320095	73320339
chr13	40384942	40385189	chr13	68316131	68316378	chr13	77198180	77198433
chr13	40391440	40391687	chr13	68428699	68428952	chr13	77422635	77422880
chr13	40392221	40392468	chr13	68429013	68429258	chr13	77449372	77449622
chr13	40398309	40398557	chr13	68570544	68570791	chr13	77454212	77454458
chr13	40398589	40398840	chr13	71269531	71269778	chr13	77454985	77455228
chr13	40406460	40406711	chr13	71454173	71454420	chr13	77455721	77455971
chr13	40420794	40421037	chr13	71456180	71456429	chr13	77456394	77456643
chr13	40613790	40614037	chr13	71463069	71463321	chr13	77476090	77476339
chr13	42846940	42847187	chr13	71464644	71464887	chr13	77477580	77477833
chr13	42876485	42876732	chr13	71465060	71465307	chr13	77513439	77513692
chr13	42896653	42896897	chr13	71466315	71466560	chr13	77591495	77591742
chr13	43149529	43149781	chr13	71466724	71466967	chr13	77595913	77596160
chr13	43164772	43165022	chr13	71479089	71479339	chr13	77598010	77598257
chr13	43177841	43178091	chr13	71479359	71479608	chr13	77630724	77630968
chr13	52163649	52163899	chr13	71619104	71619349	chr13	77633431	77633680
chr13	52541402	52541652	chr13	71750165	71750409	chr13	77636364	77636611
chr13	52549279	52549529	chr13	71750461	71750708	chr13	77639621	77639873
chr13	53323273	53323520	chr13	71762243	71762493	chr13	77642789	77643041
chr13	53350485	53350727	chr13	71762640	71762893	chr13	77652265	77652509
chr13	53505942	53506191	chr13	71781780	71782031	chr13	77652705	77652952
chr13	53555869	53556121	chr13	71783224	71783469	chr13	77675169	77675423
chr13	53606740	53606987	chr13	71784572	71784822	chr13	77699605	77699849
chr13	53610142	53610389	chr13	73226435	73226679	chr13	77713099	77713352
chr13	53619813	53620063	chr13	73233924	73234168	chr13	77714235	77714481
chr13	53622959	53623209	chr13	73251601	73251851	chr13	77742675	77742920
chr13	59239399	59239649	chr13	73261064	73261309	chr13	77759281	77759528
chr13	62064603	62064852	chr13	73284370	73284617	chr13	78540689	78540943
chr13	68236363	68236610	chr13	73296842	73297089	chr13	89222075	89222318
chr13	68238595	68238842	chr13	73305479	73305731	chr13	94063232	94063481
chr13	68241140	68241393	chr13	73305843	73306093			
chr13	68310701	68310948	chr13	73309111	73309357			

chr18	52719979	52720226
chr18	53843255	53843507
chr18	53843510	53843764
chr18	53843879	53844121
chr18	53870090	53870332
chr18	53873286	53873528
chr18	53913040	53913287
chr18	54141132	54141379
chr18	54154870	54155117
chr18	54183137	54183381
chr18	54271758	54272011
chr18	54352988	54353238
chr18	60378353	60378602
chr18	61164144	61164393

Хромосома	старт	стоп
chr21	15774354	15774598
chr21	15774701	15774951
chr21	15941894	15942144
chr21	15942369	15942621
chr21	16330536	16330783
chr21	16331481	16331728
chr21	16332238	16332485
chr21	16333694	16333941
chr21	16335023	16335271
chr21	16336651	16336901
chr21	16337562	16337816
chr21	16338381	16338633
chr21	16338902	16339154
chr21	16339772	16340019
chr21	16340409	16340656
chr21	16344080	16344330
chr21	16349101	16349354
chr21	16362700	16362954
chr21	16375294	16375539
chr21	16376059	16376308
chr21	16376695	16376942
chr21	16519760	16520004
chr21	16529308	16529552
chr21	16529940	16530189
chr21	16531755	16532002
chr21	16533111	16533355
chr21	16537385	16537631
chr21	16545527	16545774
chr21	16788480	16788722
chr21	16792457	16792708
chr21	17149630	17149877
chr21	45114956	45115203

Хромосома	старт	стоп
chr21	17160410	17160656
chr21	17160972	17161216
chr21	19300496	19300748
chr21	19396568	19396818
chr21	25701930	25702183
chr21	25702313	25702560
chr21	27985131	27985378
chr21	27989556	27989803
chr21	28000089	28000336
chr21	28007209	28007456
chr21	28007550	28007793
chr21	28008308	28008555
chr21	28054591	28054835
chr21	28292245	28292493
chr21	28294400	28294644
chr21	28308980	28309224
chr21	28311566	28311813
chr21	28316801	28317045
chr21	29796060	29796309
chr21	30112410	30112660
chr21	30112772	30113024
chr21	30113106	30113350
chr21	30114870	30115119
chr21	30301073	30301322
chr21	30301466	30301713
chr21	30317970	30318220
chr21	30358601	30358851
chr21	30397555	30397801
chr21	30419470	30419720
chr21	30527461	30527714
chr21	30528513	30528763
chr21	30554566	30554808

Хромосома	старт	стоп
chr21	30568450	30568700
chr21	30901196	30901442
chr21	30950156	30950403
chr21	30950626	30950878
chr21	30952977	30953227
chr21	30953281	30953523
chr21	30953836	30954083
chr21	30966825	30967077
chr21	30968330	30968574
chr21	30969659	30969903
chr21	31015490	31015733
chr21	31022236	31022488
chr21	31091818	31092065
chr21	31109787	31110039
chr21	31110149	31110396
chr21	31113996	31114249
chr21	31120815	31121064
chr21	31144588	31144835
chr21	35154438	35154687
chr21	40549354	40549598
chr21	40759240	40759493
chr21	40839325	40839569
chr21	40846844	40847093
chr21	41160090	41160343
chr21	41171755	41171999
chr21	41298214	41298467
chr21	41592339	41592583
chr21	41596188	41596432
chr21	41741900	41742144
chr21	43219112	43219356
chr21	46468654	46468898
chr21	46495701	46495954

chr21	46573780	46574033							
id	Хромосома	старт	стоп	Хромосома	старт	стоп	Хромосома	старт	стоп
	chr2	50176039	50176288						
	chr2	63596895	63597144						
	chr2	103538910	103539159						

Хромосома	старт	стоп	Хромосома	старт	стоп	Хромосома	старт	стоп
chr3	23866353	23866602						
chr4	21947614	21947863						
chr4	62556697	62556946						
chr4	72418734	72418983						
chr4	122570216	122570465						
chr4	130264593	130264842						

Хромосома	старт	стоп	Хромосома	старт	стоп	Хромосома	старт	стоп
chr7	22725339	22725588						
chr7	42250786	42251035						

Хромосома	старт	стоп	Хромосома	старт	стоп	Хромосома	старт	стоп
chr9	2600491	2600740						
chr9	14299731	14299980						
chr9	72782230	72782479						
chr9	82880926	82881175						

Хромосома	старт	стоп
chr10	52560407	52560656

Хромосома	старт	стоп
-----------	-------	------

Хромосома	старт	стоп
-----------	-------	------

Хромосома	старт	стоп
chr11	10786015	10786264

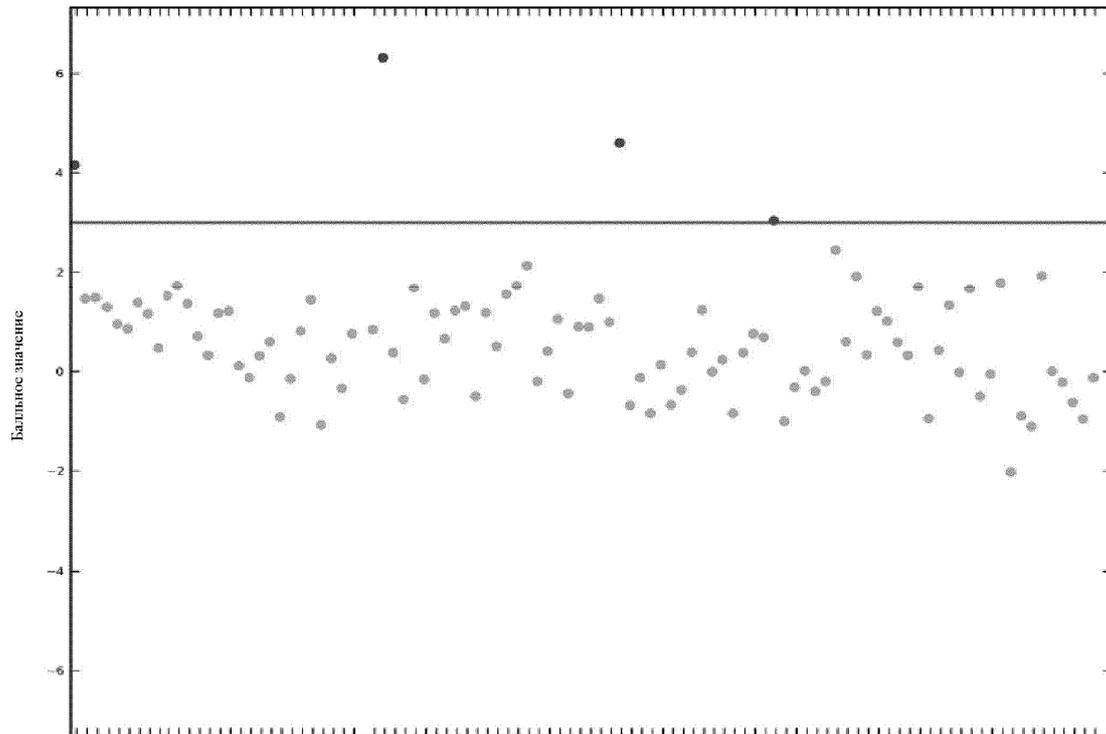
Хромосома	старт	стоп
-----------	-------	------

Хромосома	старт	стоп
-----------	-------	------

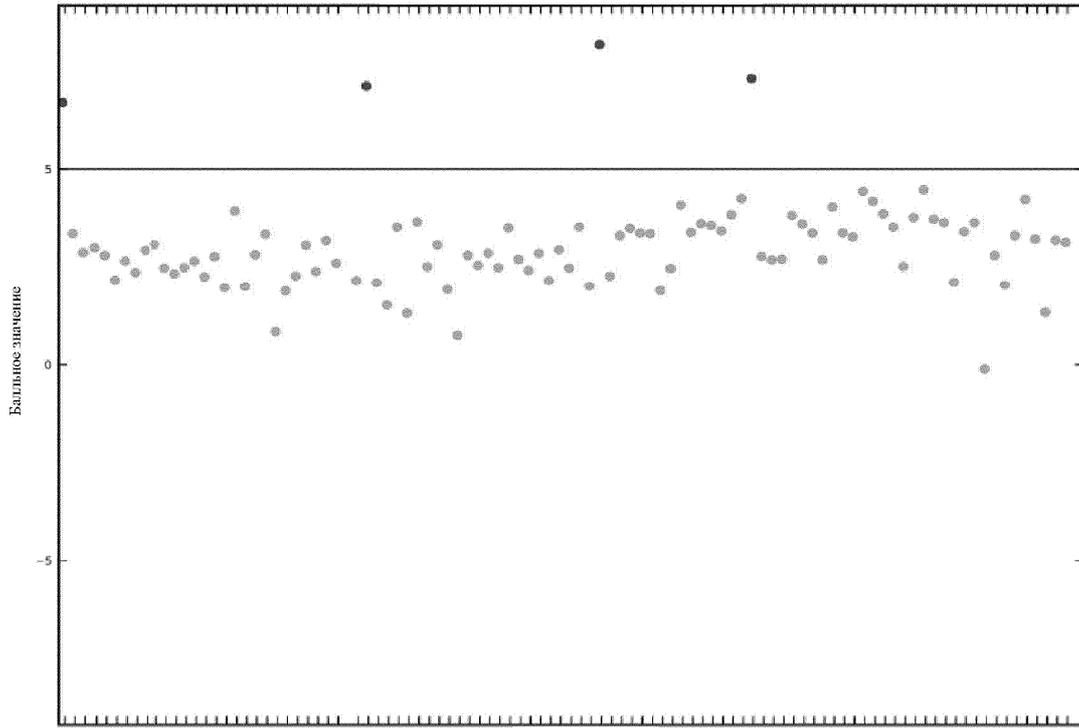
Хромосома	старт	стоп	Хромосома	старт	стоп	Хромосома	старт	стоп
chr12	29714971	29715220						
chr12	40732964	40733213						

Фиг. 2

Балльная оценка с применением метода (i)

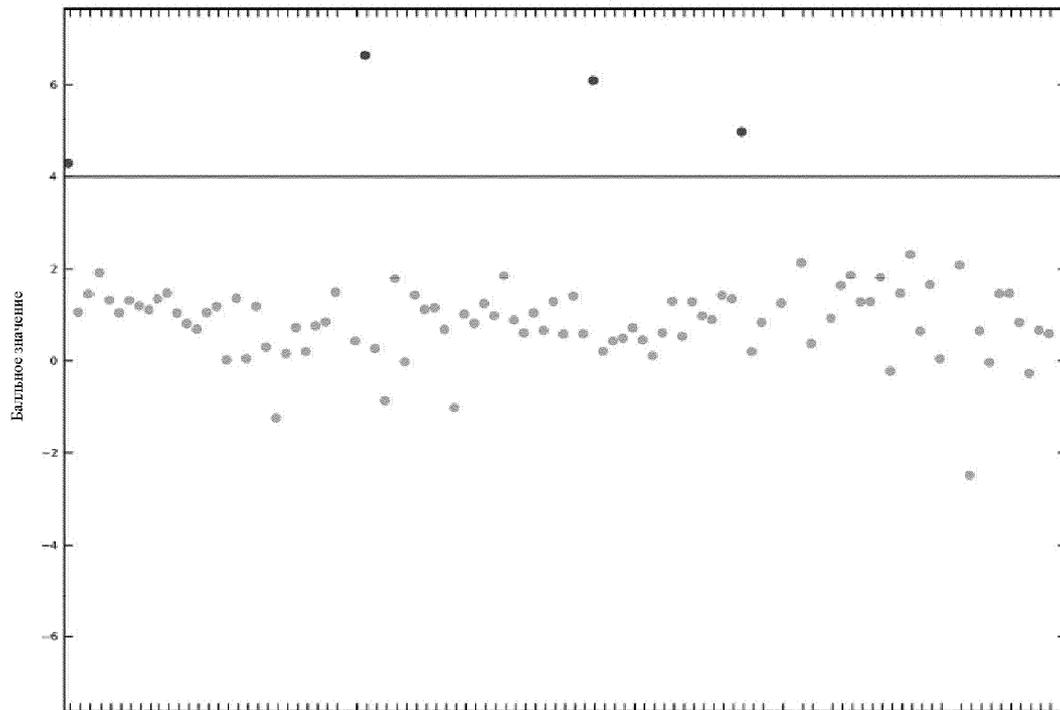
образец
Фиг. 3

Балльная оценка с применением метода (ii)



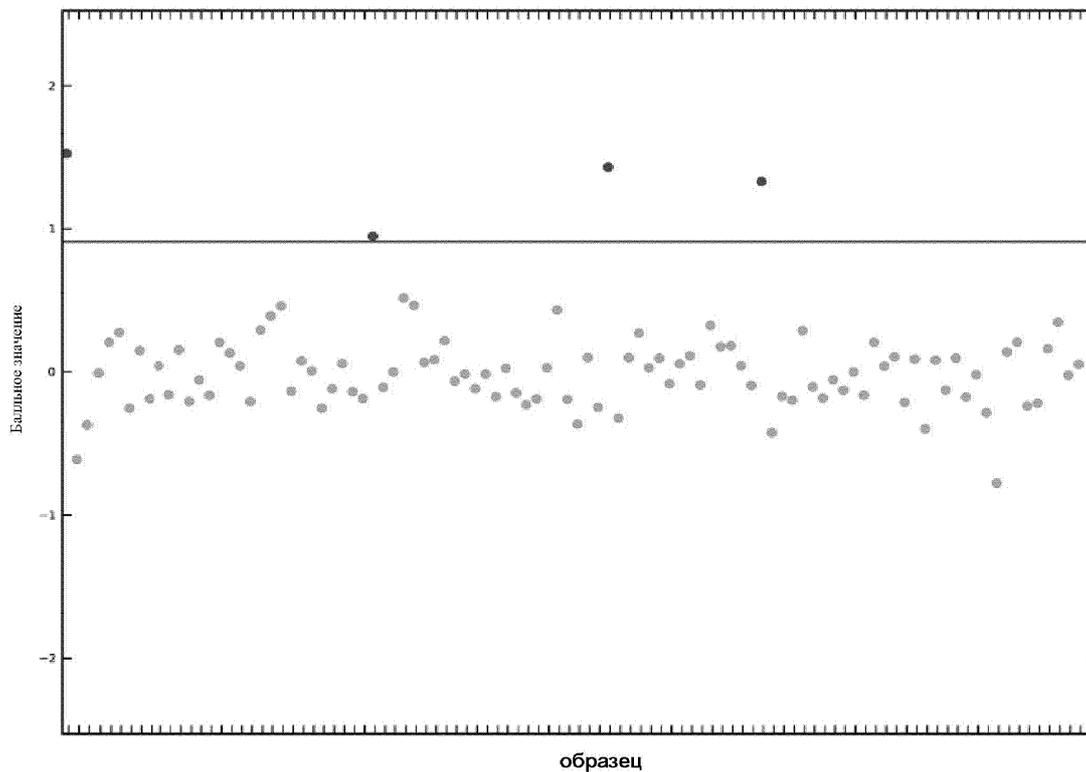
образец
Фиг. 4

Балльная оценка с применением метода (iii)



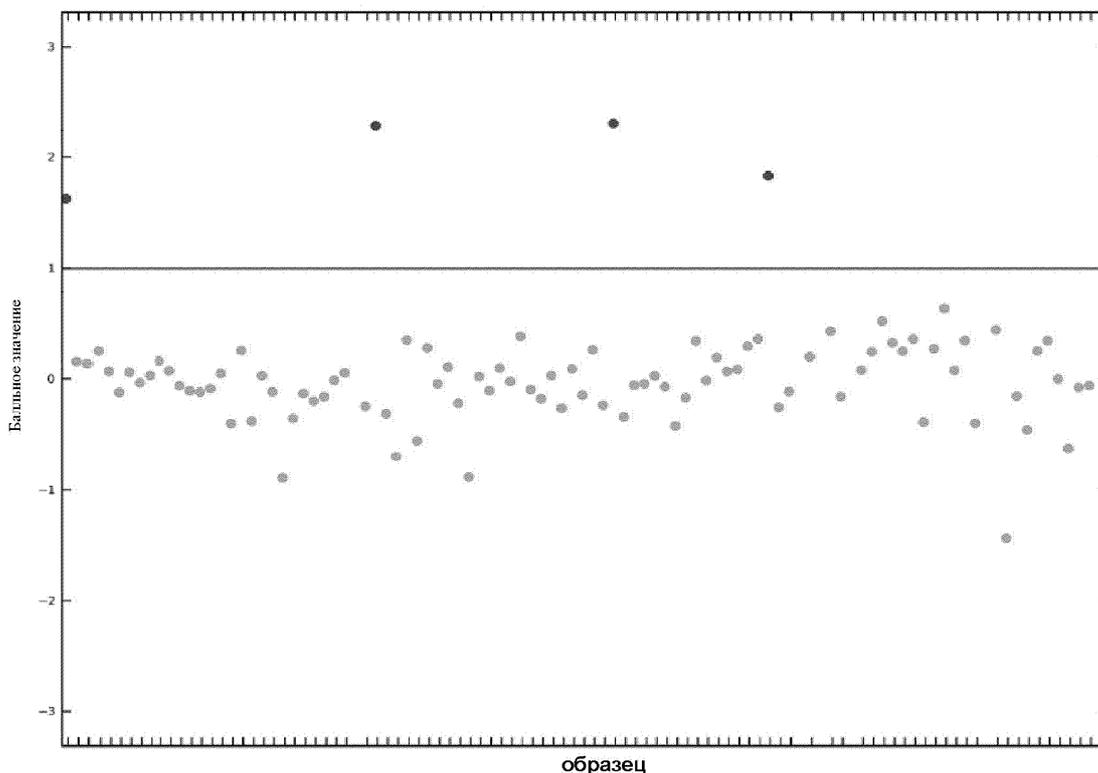
образец
Фиг. 5

Бальная оценка с применением метода (iv)



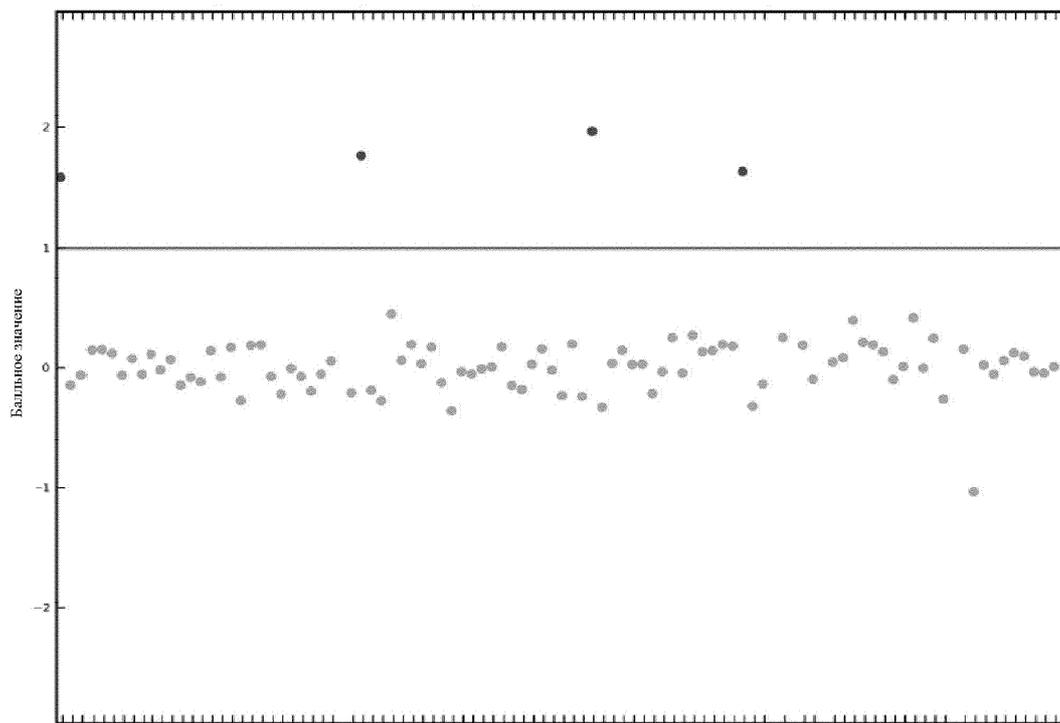
Фиг. 6

Взвешенная бальная оценка с применением статистических методов 1-3 и метода взвешенной бальной оценки 1



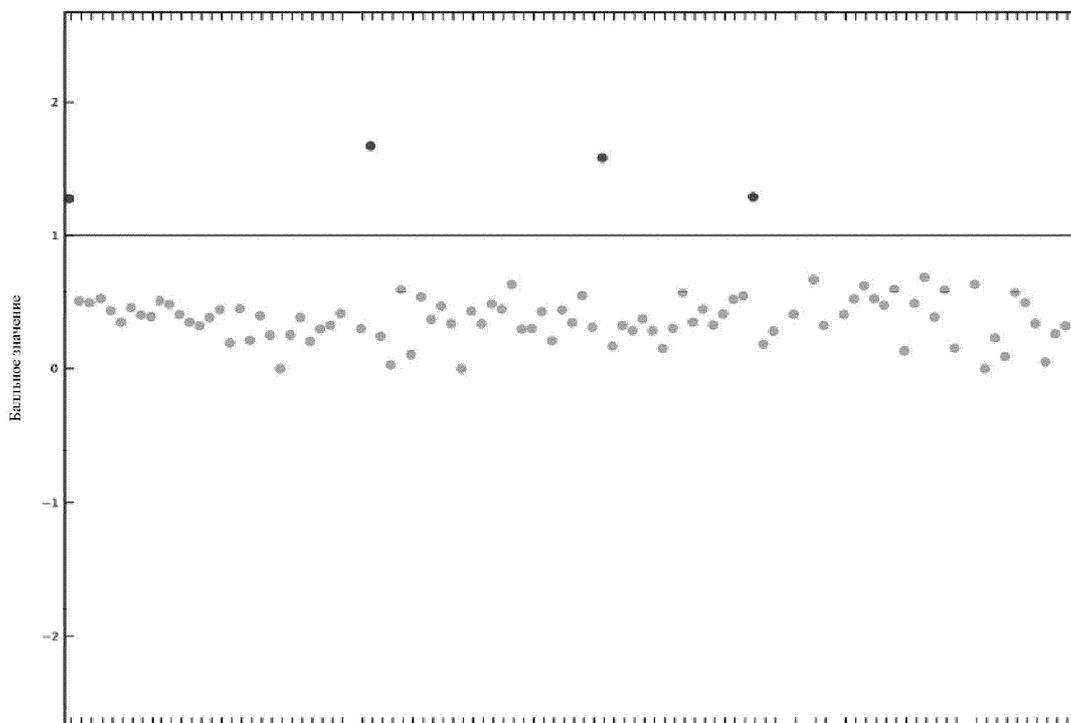
Фиг. 7

Взвешенная балльная оценка с применением статистических методов 1-4 и метода взвешенной балльной оценки 1



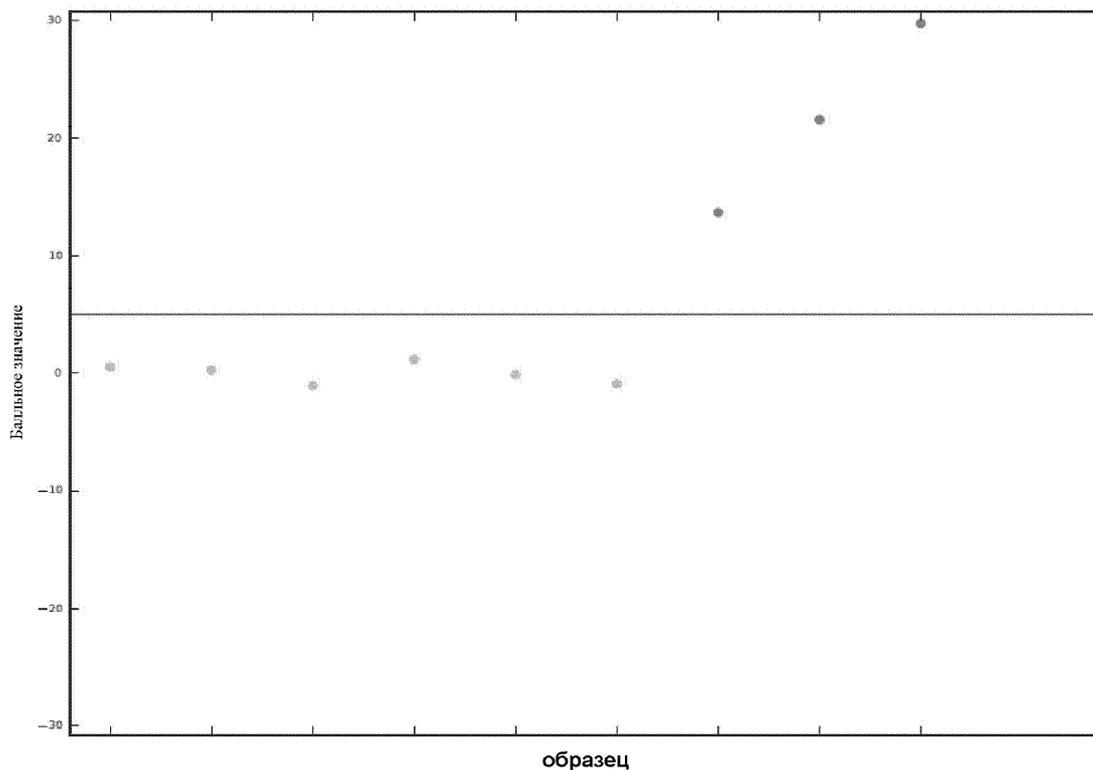
образец
Фиг. 8

Взвешенная балльная оценка с применением статистических методов 1-3 и метода взвешенной балльной оценки 2



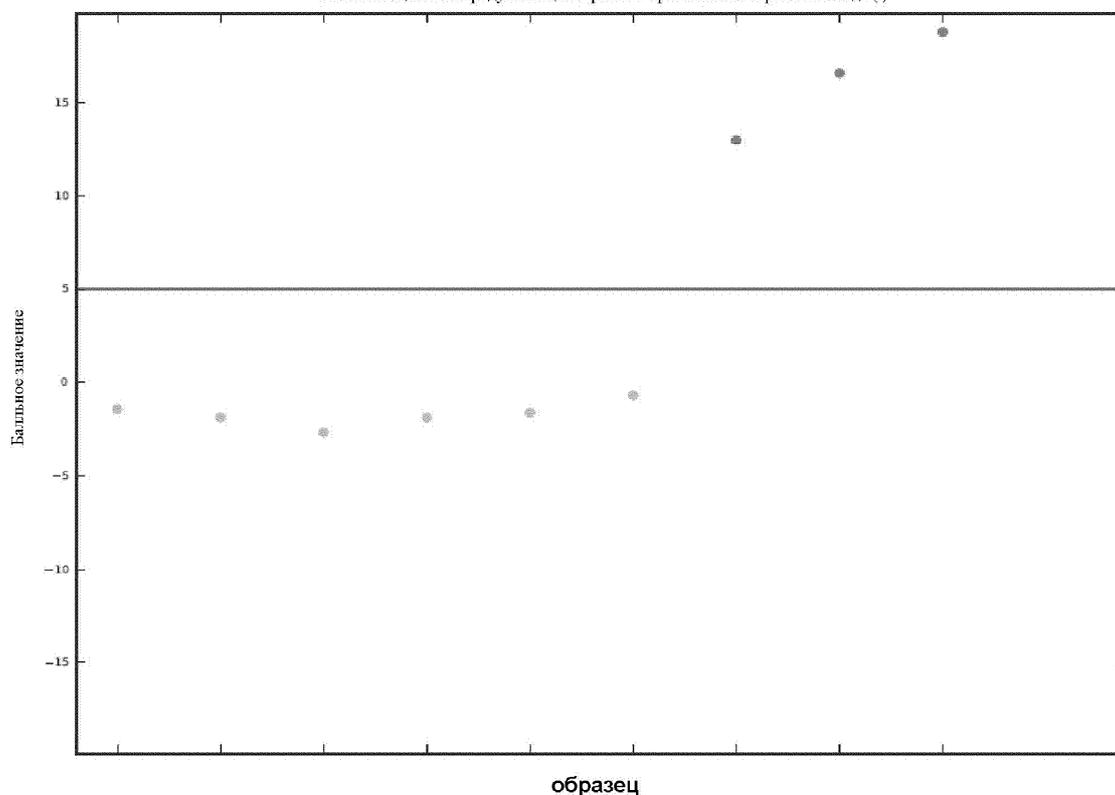
образец
Фиг. 9

Бальная оценка микроделений 7q11.23 с применением варианта метода (i)



Фиг. 10

Бальная оценка микродупликаций 7q11.23 с применением варианта метода (i)



Фиг. 11



Евразийская патентная организация, ЕАПВ

Россия, 109012, Москва, Малый Черкасский пер., 2