



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2020년06월04일
(11) 등록번호 10-2112438
(24) 등록일자 2020년05월12일

(51) 국제특허분류(Int. Cl.)
G16B 20/00 (2019.01) C12Q 1/6827 (2018.01)
C12Q 1/6883 (2018.01) G16B 15/00 (2019.01)
G16B 30/00 (2019.01)
(52) CPC특허분류
G16B 20/00 (2019.02)
C12Q 1/6827 (2018.05)
(21) 출원번호 10-2018-7031541(분할)
(22) 출원일자(국제) 2008년07월23일
심사청구일자 2018년11월29일
(85) 번역문제출일자 2018년10월30일
(65) 공개번호 10-2018-0121695
(43) 공개일자 2018년11월07일
(62) 원출원 특허 10-2010-7003969
원출원일자(국제) 2008년07월23일
심사청구일자 2013년07월18일
(86) 국제출원번호 PCT/GB2008/002530
(87) 국제공개번호 WO 2009/013496
국제공개일자 2009년01월29일
(30) 우선권주장
60/951,438 2007년07월23일 미국(US)
(56) 선행기술조사문헌
US20050221341 A1*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
더 차이나이즈 유니버시티 오브 홍콩
중국 홍콩 엔. 티. 새턴 더 차이나이즈 유니버시티
오브 홍콩
(72) 발명자
로 유크-밍 데니스
중국 홍콩 주룽 호만틴 킹 탁 스트리트 7 포쓰 플
로어
츄 로사 와이 켄
중국 홍콩 뉴 테리토리즈 타이 포 형 램 드라이브
1 컨스텔레이션 코브 블록 1 플랫 1에이
찬 판 치
중국 홍콩 주룽 메이 푸 산 춘 브로드웨이 스트리
트 블록 34 13/에프 플랫 에이
(74) 대리인
김진희, 김태홍

전체 청구항 수 : 총 22 항

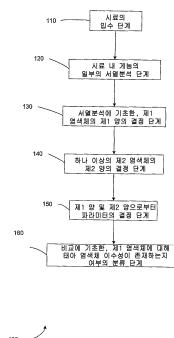
심사관 : 박진아

(54) 발명의 명칭 대규모 병렬 게놈 서열분석을 이용한 태아 염색체 이수성의 진단 방법

(57) 요약

본 발명의 실시양태는 임신한 여성으로부터 입수한 생체 시료로부터 태아 염색체 이수성이 존재하는지 여부를 결정하기 위한 방법, 시스템 및 장치를 제공한다. 게놈의 일부가 서열분석되도록 생체 시료의 핵산 분자가 서열분석된다. 임상적으로 관련성이 있는 염색체 및 배경 염색체의 각각의 양을 서열분석의 결과로부터 결정한다. 이런 양으로부터 도출된 파라미터(예를 들면, 비)를 하나 이상의 컷오프 값과 비교하여, 태아 염색체 이수성이 존재하는지 여부를 결정한다.

대표도 - 도1



(52) CPC특허분류

C12Q 1/6883 (2018.05)

G16B 15/00 (2019.02)

G16B 30/00 (2019.02)

C12Q 2600/112 (2013.01)

C12Q 2600/154 (2013.01)

C12Q 2600/156 (2013.01)

G01N 2800/387 (2013.01)

명세서

청구범위

청구항 1

임신한 여성 피험자로부터 얻은, 핵산 분자를 함유하는 생체 시료에서 태아 염색체 이수성을 산전 진단하는 방법으로서,

상기 생체 시료를 입수하는 단계;

상기 생체 시료에 함유된 복수의 핵산 분자 중 적어도 일부분을 서열분석하는 단계로서, 서열분석된 일부분은 인간 게놈의 일부에 해당하는 것인 단계;

이러한 서열분석에 기초하여,

제1 염색체에서 유래하는 것으로 확인된 서열로부터 제1 염색체의 제1 양을 결정하는 단계;

하나 이상의 제2 염색체 중 하나에서 유래하는 것으로 확인된 서열로부터 하나 이상의 제2 염색체의 제2 양을 결정하는 단계;

상기 제1 양 및 상기 제2 양으로부터 파라미터를 결정하는 단계;

상기 파라미터를 하나 이상의 컷오프 값(cutoff value)과 비교하는 단계; 및

이러한 비교에 기초하여, 제1 염색체에 대해 태아 염색체 이수성이 존재하는지 여부의 분류를 결정하는 단계

를 포함하는 진단 방법.

청구항 2

제1항에 있어서, 서열분석은 생체 시료에 함유된 핵산 분자 중 일부에 대해 무작위로 수행되는 것인 진단 방법.

청구항 3

제1항에 있어서, 생체 시료는 모체 혈액, 혈장, 혈청, 소변 또는 타액인 진단 방법.

청구항 4

제1항에 있어서, 생체 시료는 경자궁경부 세척액(transcervical lavage fluid)인 진단 방법.

청구항 5

제1항에 있어서, 제1 염색체는 염색체 21, 염색체 18, 염색체 13, 염색체 X, 또는 염색체 Y인 진단 방법.

청구항 6

제1항에 있어서, 파라미터는 제1 염색체에서 유래한 서열의 비율인 진단 방법.

청구항 7

제6항에 있어서, 비율은 서열분석된 태그 수의 카운트 비, 서열분석된 뉴클레오타이드 수의 비, 및 누적된 서열의 길이의 비 중 어느 하나 이상으로부터 얻어지는 것인 진단 방법.

청구항 8

제6항에 있어서, 제1 염색체에서 유래한 서열은 염기쌍이 특정 수 미만이 되도록 선택되는 것인 진단 방법.

청구항 9

제8항에 있어서, 염기쌍의 특정 수는 300 bp, 200 bp, 또는 100 bp인 진단 방법.

청구항 10

제1항에 있어서, 생체 시료의 핵산 분자는 하나 이상의 특정 염색체에서 유래한 서열에 대해 농축된 것인 진단 방법.

청구항 11

제1항에 있어서, 생체 시료의 핵산 분자는 300 bp 미만의 서열에 대해 농축된 것인 진단 방법.

청구항 12

제1항에 있어서, 생체 시료의 핵산 분자는 200 bp 미만의 서열에 대해 농축된 것인 진단 방법.

청구항 13

제1항에 있어서, 생체 시료의 핵산 분자는 중합효소 연쇄 반응을 이용하여 증폭된 것인 진단 방법.

청구항 14

제1항에 있어서, 서열분석된 일부분은 적어도 미리 결정된 인간 게놈의 일부에 해당하는 것인 진단 방법.

청구항 15

제1항에 있어서, 일부분은 인간 게놈의 0.1% 이상에 해당하는 것인 진단 방법.

청구항 16

제1항에 있어서, 일부분은 인간 게놈의 0.5% 이상에 해당하는 것인 진단 방법.

청구항 17

제1항에 있어서, 하나 이상의 컷오프 값은 생체 시료 중 태아 DNA 농도의 비와 관련된 것인 진단 방법.

청구항 18

제17항에 있어서, 생체 시료 중 태아 DNA 농도의 비는 Y W 염색체 서열의 비율, 태아의 후생유전학적 마커, 또는 단일 염기 다형성 분석의 이용 중 어느 하나 이상에 의해 결정되는 것인 진단 방법.

청구항 19

제1항에 있어서, 컷오프 값은 정상 생체 시료에서 확립된 기준값인 진단 방법.

청구항 20

제1항에 있어서, 생체 시료에서 태아 DNA의 양을 확인하고, 원하는 정확도에 기초하여 분석하고자 하는 서열의 수(N)를 계산하는 단계를 추가로 포함하는 것인 진단 방법.

청구항 21

삭제

청구항 22

임신한 여성 피험자로부터 얻은, 핵산 분자를 함유하는 생체 시료에서 태아 염색체 이수성을 산전 진단하는 방법으로서,

상기 생체 시료를 입수하는 단계;

원하는 정확도에 기초하여 분석하고자 하는 서열의 수(N)를 계산하는 단계;

상기 생체 시료에 함유된 N개 이상의 핵산 분자를 무작위로 서열분석하는 단계로서, 일부분은 인간 게놈의 일부에 해당하는 것인 단계;

이러한 무작위 서열분석에 기초하여,

제1 염색체에서 유래하는 것으로 확인된 서열로부터 제1 염색체의 제1 양을 결정하는 단계,

하나 이상의 제2 염색체 중 하나에서 유래하는 것으로 확인된 서열로부터 하나 이상의 제2 염색체의 제2 양을 결정하는 단계,

상기 제1 양 및 상기 제2 양으로부터 파라미터를 결정하는 단계,

상기 파라미터를 하나 이상의 컷오프 값(cutoff value)과 비교하는 단계, 및

이러한 비교에 기초하여, 제1 염색체에 대해 태아 염색체 이수성이 존재하는지 여부의 분류를 결정하는 단계

를 포함하는 진단 방법.

청구항 23

제22항에 있어서, 생체 시료에서 태아 DNA의 백분율을 확인하는 단계를 추가로 포함하고, 원하는 정확도에 기초하여 분석하고자 하는 서열의 수(N)를 계산하는 것이 상기 백분율에 기초하는 것인 진단 방법.

발명의 설명

기술 분야

[0001]

우선권 주장

[0002]

본원은 발명의 명칭이 "핵산 서열 불균형의 결정 방법(DETERMINING A NUCLEIC ACID SEQUENCE IMBALANCE)"인 2007년 7월 23일자에 출원된 미국 가출원 제60/951438호(변리사 문서 번호 제016285-005200US호)의 정규 출원으로 이에 대해 우선권을 주장하며, 이의 전체 내용은 사실상 참조로 본원에 인용된다.

[0003]

관련 출원에 대한 상호 참조

[0004]

또한, 본원은 발명의 명칭이 "핵산 서열 불균형의 결정 방법"인 공준 출원된 정규 출원(변리사 문서 번호 제016285-005210US호)에 관한 것이고, 이의 전체 내용은 사실상 참조로 본원에 인용된다.

[0005]

기술분야

[0006]

본 발명은 일반적으로 여러 핵산 서열 간의 불균형을 측정하여 태아 염색체 이수성을 진단 검사하는 방법, 더 구체적으로 모체 시료(예를 들면, 혈액)의 검사를 통해 21번 삼염색체증(다운 증후군) 및 다른 염색체 이수성을 확인하는 방법에 관한 것이다.

배경 기술

[0007]

태아 염색체 이수성은 염색체 또는 염색체 부위의 양(들)이 비정상적으로 존재함으로 인하여 발생한다. 비정상적 양(들)이란 비정상적으로 많은 것(예를 들면, 21번 삼염색체증에서 21번 염색체 또는 염색체 부위가 과잉 존재하는 것), 또는 비정상적으로 적은 것(예를 들면, 터너 증후군에서 X 염색체 카피수가 1개 결손된 것)일 수 있다.

[0008]

21번 삼염색체증과 같은 태아 염색체 이수성의 전통적인 산전 진단 방법에는 양수천자 또는 융모막 융모 생검(chorionic villus sampling)과 같은 침습적 시술에 의한 태아 물질의 시료 채취 단계가 수반되어 한정된 유산 위험이 야기된다. 임신한 여성의 위험을 기준으로 계층화하기 위해 최종적인 침습적 진단 시술 전에 초음파 검사 및 생화학적 마커를 사용한 스크리닝과 같은 비침습적 시술이 이용되어 왔다. 그러나, 이런 스크리닝 방법은 통상적으로 중요한 염색체 이상 대신에 21번 삼염색체증과 같은 염색체 이수성과 관련된 부수 현상만을 측정하여, 부적절한 진단 정확성 및 재태 연령에 의해 매우 영향을 받는다는 것과 같은 다른 단점을 갖는다.

[0009]

1997년에 모체 혈장에서 순환하는 세포 유리 태아 DNA가 발견되어 비침습적 산전 진단에 대한 새로운 가능성을 제공하였다(Lo, YMD and Chiu, RWK 2007 Nat Rev Genet 8, 71-77). 이런 방법이 반성 유전 질환(Costa, JM et al. 2002 N Engl J Med 346, 1502) 및 특정한 단일 유전자 질환(Lo, YMD et al. 1998 N Engl J Med 339, 1734-1738)의 산전 진단에 용이하게 적용되어 왔지만, 태아 염색체 이수성의 산전 진단에 대한 이의 적용에 상당한 문제가 드러났다(Lo, YMD and Chiu, RWK 2007, supra). 첫째, 태아 핵산은 태아 핵산의 분석을 종종 방해할 수

있는 모체 기원의 많은 배경 핵산과 함께 모체 혈장에 공존한다(Lo, YMD et al. 1998 Am J Hum Genet 62, 768-775). 둘째, 태아 핵산은 대개 세포 유리 형태로 모체 혈장에서 순환하기 때문에, 태아 게놈 내의 유전자 또는 염색체의 용량 정보를 도출하기 어려워진다.

[0010] 최근에, 이러한 문제를 극복하는 상당한 발전이 이루어졌다(Benachi, A & Costa, JM 2007 Lancet 369, 440-442). 한 접근법은 모체 혈장 내의 태아 특이적 핵산을 검출하여, 모체 배경 간섭의 문제를 극복하는 것이다(Lo, YMD and Chiu, RWK 2007, supra). 21번 염색체의 용량은 태반 유래 DNA/RNA 분자에서의 다형성 대립유전자의 비로부터 추론된다. 그러나, 이러한 방법은 시료가 표적화된 핵산을 더 적게 함유할 때 덜 정확하고, 하나의 다형성이 이용되는 경우 모집단의 유일한 부분집단인 표적화된 다형성에 대해 이형접합성인 태아에만 적용될 수 있다.

[0011] 달란(Dhallan) 등은 문헌[Dhallan, R, et al. 2007, supra Dhallan, R, et al. 2007 Lancet 369, 474-481]에서 모체 혈장에 포름알데하이드를 첨가하여 순환하는 태아 DNA 비율을 농축시키는 대안적인 전략을 기재하였다. 21번 염색체에 대한 단일 염기 다형성(SNP; single nucleotide polymorphism)에 대해 태아 비특이적 대립유전자에 대한 부계 유전된 태아 특이적 대립유전자의 비를 측정하여 모체 혈장에서 태아에 의해 제공된 21번 염색체 서열의 비율을 결정하였다. 기준 염색체에 대해서도 SNP 비를 유사하게 산출하였다. 이후, 태아 21번 염색체의 불균형을 21번 염색체에 대한 SNP 비와 기준 염색체의 SNP 비 간의 통계학적 유의차를 검출하여 추론하였으며, 여기서 유의라는 것은 ≤ 0.05 의 고정된 p값을 이용하는 것으로 정의된다. 높은 모집단 커버율(coverage)을 보장하기 위해, 염색체당 500개 이상의 SNP를 표적화하였다. 그러나, 태아 DNA를 높은 비율로 농축시키는 포름알데하이드의 유효성과 관련해서 논쟁이 있어 왔고(Chung, GTY, et al. 2005 Clin Chem 51, 655-658), 따라서 상기 방법의 재현성이 추가로 평가될 필요가 있다. 또한, 각각의 태아 및 모체로부터 각각의 염색체에 대한 상이한 수의 SNP의 정보가 제공되므로, SNP 비 비교를 위한 통계 검정의 검증력은 사례마다 가변적일 것이다(Lo, YMD & Chiu, RWK. 2007 Lancet 369, 1997). 또한, 이런 접근법은 유전자 다형성의 검출에 의존적이므로, 유전자 다형성에 대해 이형접합성인 태아로 제한된다.

[0012] 짐머르마네트(Zimmermann) 등은 21번 삼염색체증 태아 및 정배수체 태아로부터 얻은 양수 세포 배양액에서 21번 염색체 유전자좌 및 기준 유전자좌의 중합효소 연쇄 반응(PCR) 및 DNA 정량검사를 이용하여 21번 삼염색체증 태아에서 21번 염색체 DNA 서열에서의 1.5배 증가에 기초하여 2군의 태아를 구별할 수 있었다(2002 CHn Chem 48, 362-363). DNA 주형 농도에서의 2배 차이는 오직 1회의 역치 사이클(C_t)의 차이를 구성하므로, 1.5배 차이의 변별력은 종래의 실시간 PCR의 한계이었다. 더 우수한 정도의 정량적 변별력을 달성하기 위해, 대안적인 전략이 필요하다.

[0013] 디지털 PCR은 핵산 시료 중의 비대칭 대립유전자 비 검출을 위해 개발되었다(Chang, HW et al. 2002 J Natl Cancer Inst 94, 1697-1703). 디지털 PCR은 핵산 함유 검체를 복수의 별개 시료로 분배할 것을 요하는 증폭 기반 핵산 분석 기술이고, 이 경우 각각의 시료는 시료당 평균 대략 1개 이하의 표적 서열을 함유한다. 특이적 핵산 표적은 서열 특이적 프라이머를 사용하여 디지털 PCR에 의해 증폭시켜 특이적 앰플리콘(amplicon)을 생성한다. 표적화되는 핵산 유전자좌 및 반응에 포함되는 서열 특이적 프라이머의 종 또는 패널은 핵산 분석 전 결정되거나 선택된다.

[0014] 임상적으로, 디지털 PCR은 종양 DNA 시료에서 이형접합성 소실(LOH; loss of heterozygosity)의 검출에 유용한 것으로 밝혀졌다(Zhou, W. et al. 2002 Lancet 359, 219-225). 디지털 PCR 결과를 분석하기 위해, 선행 연구에서는, 실험 결과를 시료 중에 LOH가 존재하는 것 또는 부재하는 것을 제시하는 것으로서 분류하기 위해 순차적 확률비 검정(SPRT)이 채택되었다(El Karoui et al. 2006 Stat Med 25, 3124-3133).

[0015] 선행 연구에 이용된 방법에서, 디지털 PCR로부터 수집된 데이터의 양은 꽤 적다. 따라서, 정확도는 적은 수의 데이터 점 및 통상적인 통계적 변동으로 인해 떨어질 수 있다.

[0016] 따라서, 비침습적 검사는 부정 오류 및 긍정 오류를 각각 최소화할 수 있도록 높은 민감성 및 특이성을 갖는 것이 바람직하다. 그러나, 태아 DNA는 낮은 절대 농도로 존재하고 모체 혈장 및 혈청 중 총 DNA 서열의 작은 비율만을 차지한다. 따라서, 모체 배경 핵산을 함유하는 생체 시료 중에 작은 모집단으로 존재하는 제한된 양의 태아 핵산으로부터 추론될 수 있는 유전자 정보의 양을 최대화함으로써 태아 염색체 이수성을 비침습적으로 검출할 수 있는 방법이 또한 바람직하다.

발명의 내용

도면의 간단한 설명

[0017]

도 1은 본 발명의 실시양태에 따라 임신한 여성 피험자로부터 입수한 생체 시료에서 태아 염색체 이수성을 산전 진단하는 방법(100)의 흐름도이다.

도 2는 본 발명의 실시양태에 따른 무작위 서열분석을 이용하여 태아 염색체 이수성을 산전 진단하는 방법(200)의 흐름도이다.

도 3a는 본 발명의 실시양태에 따라 21번 삼염색체증 태아 또는 정배수체 태아를 갖는 모체 혈장 시료에서 21번 염색체 서열의 표현의 백분율의 도면을 나타낸 것이다.

도 3b는 본 발명의 실시양태에 따른 대규모 병렬 서열분석에 의해 결정된 모체 혈장 내의 태아 DNA의 농도 비와 미세유체 장치 디지털 PCR에 의해 결정된 모체 혈장 내의 태아 DNA의 농도 비 간의 상관관계를 나타낸 것이다.

도 4a는 본 발명의 실시양태에 따라 염색체당 정렬된 서열의 표현의 백분율의 도면을 나타낸 것이다.

도 4b는 도 4a에 도시된 21번 삼염색체증 사례와 정배수체 사례 간의 염색체당 표현의 백분율 차이(%)의 도면을 나타낸 것이다.

도 5는 본 발명의 실시양태에 따라 21번 삼염색체증 태아를 갖는 모체 혈장 시료에서 21번 염색체 서열의 과표현의 정도와 태아 DNA의 농도 비 간의 상관관계를 나타낸 것이다.

도 6은 본 발명의 실시양태에 따라 분석된 인간 게놈의 일부분을 표로 나타낸 것이다. T21은 21번 삼염색체증 태아를 임신한 임산부로부터 입수한 시료를 나타낸다.

도 7은 본 발명의 실시양태에 따라 21번 삼염색체증 태아로부터 정배수체를 구별하는 데에 필요한 서열 수를 표로 나타낸 것이다.

도 8a는 본 발명의 실시양태에 따라 21번 염색체로 정렬되는 서열분석 태그의 처음 10의 출발 위치를 표로 나타낸 것이다.

도 8b는 본 발명의 실시양태에 따라 22번 염색체로 정렬되는 서열분석 태그의 처음 10의 출발 위치를 표로 나타낸 것이다.

도 9는 본 발명의 실시양태에 따른 시스템 및 방법과 함께 사용될 수 있는 예시적인 컴퓨터 장치의 블록 도식을 나타낸 것이다.

발명을 실시하기 위한 구체적인 내용

[0018]

발명의 개요

[0019]

본 발명의 실시양태는 임신한 여성으로부터 입수한 생체 시료 내에 핵산 서열 불균형(예를 들면, 염색체 불균형)이 존재하는지 여부를 결정하기 위한 방법, 시스템 및 장치를 제공한다. 상기 결정은 생체 시료 내에 임상적으로 관련성이 없는 염색체 부위(배경 부위)의 양의 파라미터에 대한 다른 임상적으로 관련성이 있는 염색체 부위의 양의 파라미터를 이용하여 수행할 수 있다. 일 양태에서, 모체 시료, 예컨대 소변, 혈장, 혈청 및 다른 적합한 생체 시료에서 핵산 분자의 서열분석으로부터 염색체의 양을 결정한다. 게놈의 일부가 서열분석되도록 생체 시료의 핵산 분자가 서열분석된다. 예를 들면, 2의 염색체 부위(또는 부위 집단)의 양의 비에 대해, 기준 수량과 비교하여 변화(즉, 불균형)가 존재하는지 여부를 결정하기 위해 하나 이상의 컷오프 값을 선택한다.

[0020]

예시적인 일 실시양태에 따르면, 태아 염색체 이수성을 산전 진단하기 위해 임신한 여성으로부터 입수한 생체 시료를 분석한다. 상기 생체 시료는 핵산 분자를 함유한다. 생체 시료에 함유된 핵산 분자의 일부분을 서열분석한다. 일 양태에서, 필요한 공급 생체 시료의 양 및 비용을 함유하도록, 정확한 진단에 충분하지만 지나치게 과잉은 아닌 유전자 정보의 양이 얻어진다.

[0021]

서열분석에 기초하여, 제1 염색체의 제1 양을 제1 염색체에서 유래하는 것으로 확인된 서열로부터 결정한다. 하나 이상의 제2 염색체의 제2 양을 제2 염색체 중 하나에서 유래하는 것으로 확인된 서열로부터 결정한다. 이후, 제1 양 및 제2 양으로부터 얻은 파라미터를 하나 이상의 컷오프 값과 비교한다. 비교에 기초하여, 제1 염색체에 대해 태아 염색체 이수성이 존재하는지 여부를 분류한다. 이런 서열분석은 모체 배경 핵산을 함유하는 생체 시료에서 작은 모집단으로 존재하는 제한된 양의 태아 핵산으로부터 추론될 수 있는 유전자 정보의 양을 최대화하

는 데에 유리하다.

- [0022] 예시적인 일 실시양태에 따르면, 태아 염색체 이수성을 산전 진단하기 위해 임신한 여성으로부터 입수한 생체 시료를 분석한다. 상기 생체 시료는 핵산 분자를 함유한다. 생체 시료 내의 태아 DNA의 백분율을 확인한다. 원하는 정확도에 기초하여 분석하고자 하는 서열 수(N)를 백분율에 기초하여 계산한다. 생체 시료에 함유된 N개 이상의 핵산 분자를 무작위로 서열분석한다.
- [0023] 무작위 서열분석에 기초하여, 제1 염색체의 제1 양을 제1 염색체에서 유래하는 것으로 확인된 서열로부터 결정한다. 하나 이상의 제2 염색체의 제2 양을 제2 염색체 중 하나에서 유래하는 것으로 확인된 서열로부터 결정한다. 이후, 제1 양 및 제2 양의 파라미터를 하나 이상의 컷오프 값과 비교한다. 비교에 기초하여, 제1 염색체에 대해 태아 염색체 이수성이 존재하는지 여부를 분류한다. 이런 무작위 서열분석은 모체 배경 핵산을 함유하는 생체 시료에서 작은 모집단으로 존재하는 제한된 양의 태아 핵산으로부터 추론될 수 있는 유전자 정보의 양을 최대화하는 데에 유리하다.
- [0024] 본 발명의 다른 실시양태는 본원에 기재된 방법과 관련되는 시스템 및 컴퓨터 판독 가능한 매체에 관한 것이다.
- [0025] 본 발명의 특징 및 이점을 다음의 상세한 설명 및 첨부된 도면을 참조하여 더 잘 이해할 수 있을 것이다.
- [0026] 정의
- [0027] 본원에 사용된 용어 "생체 시료"는 피험자(예를 들면, 임신한 여성과 같은 인간)로부터 입수되고 관심있는 하나 이상의 핵산 분자(들)를 함유하는 임의의 시료를 의미한다.
- [0028] 용어 "핵산" 또는 "폴리뉴클레오타이드"는 단일 가닥 또는 이중 가닥 형태의 데옥시리보핵산(DNA) 또는 리보핵산(RNA) 및 이들의 중합체를 의미한다. 달리 특별히 제한되지 않는 한, 상기 용어는 기준 핵산과 유사한 결합 특성을 갖고 천연 뉴클레오타이드와 유사한 방식으로 대사되는 천연 뉴클레오타이드의 공지된 유사체를 함유하는 핵산을 포함한다. 달리 기재되지 않은 한, 특정 핵산 서열은 또한 명확히 기재된 서열뿐만 아니라 암묵적으로 이의 보존적으로 변형된 변이체(예를 들면, 축퇴성 코돈 치환), 대립유전자, 오소로그, SNP 및 상보적 서열을 포함한다. 구체적으로, 하나 이상의 선택된(또는 모든) 코돈의 3번 위치가 혼합 염기 및/또는 데옥시이노신 잔기로 치환되는 서열을 생성함으로써 축퇴성 코돈 치환이 달성될 수 있다(Batzer et al., Nucleic Acid Res. 19:5081 (1991); Ohtsuka et al., J. Biol. Chem. 260:2605-2608 (1985); and Rossolini et al., Mol. Cell. Probes 8:91-98 (1994)). 상기 용어 핵산은 유전자, cDNA, mRNA, 작은 비코딩 RNA, 마이크로 RNA(miRNA), 피위 상호작용(Piwi-interacting) RNA 및 유전자 또는 유전자좌에 의해 코딩된 짧은 헤어핀 RNA(shRNA)와 상호교환적으로 사용된다.
- [0029] 용어 "유전자"는 폴리펩타이드 사슬을 생성하는 데 관여하는 DNA의 분절을 의미한다. 상기 용어는 코딩 부위에 선행되는 부위(리더) 및 코딩 부위에 후행되는 부위(트레이일러) 및 개별 코딩 분절(엑손) 사이의 개재 서열(인트론)을 포함할 수 있다.
- [0030] 본원에 사용된 용어 "반응"은 관심있는 특정 폴리뉴클레오타이드 서열의 존재 또는 부재를 알려주는 화학 작용, 효소 작용 또는 물리 작용을 수반하는 임의의 과정을 의미한다. 상기 "반응"의 예로 중합효소 연쇄 반응(PCR)과 같은 증폭 반응을 들 수 있다. 상기 "반응"의 다른 예로 합성 또는 결찰 중 어느 하나에 의한 서열분석 반응을 들 수 있다. "정보제공(informative) 반응"은 하나 이상의 관심있는 특정 폴리뉴클레오타이드 서열의 존재를 알려주는 반응이고, 일 예에서 관심있는 서열이 오직 하나 존재한다. 본원에 사용된 용어 "웰"은 한정된 구조물(예를 들면, PCR 어레이에서는 웰 형태의 바이알, 셀 또는 챔버) 내의 소정 위치에서의 반응을 의미한다.
- [0031] 본원에 사용된 용어 "임상적으로 관련성이 있는 핵산 서열"은 가능성이 있는 불균형이 검사되는 게놈 서열의 분절에 해당하는 게놈 서열보다 짧은 폴리뉴클레오타이드 서열 또는 더 큰 게놈 서열 그 자체를 의미할 수 있다. 일 예로 21번 염색체의 서열을 들 수 있다. 다른 예로 18번 염색체, 13번 염색체, X 염색체 및 Y 염색체를 들 수 있다. 또 다른 예로 태아가 이의 편친 또는 양친으로부터 유전받을 수 있는 변이된 유전자 서열 또는 유전자 다형성 또는 카피수 변이를 들 수 있다. 또 다른 예로 악성 종양에서 변이, 결실 또는 증폭되는 서열, 예를 들면 이형접합성 소실 또는 유전자 중복이 발생하는 서열을 들 수 있다. 몇몇 실시양태에서, 복수의 임상적으로 관련성이 있는 핵산 서열 또는 임상적으로 관련성이 있는 핵산 서열의 동등한 복수의 마커를 이용하여 불균형 검출을 위한 데이터를 제공할 수 있다. 예를 들면, 21번 염색체에서 5개의 불연속 서열로부터 얻은 데이터가 가능성이 있는 21번 염색체 불균형을 결정하기 위해 함께 방식으로 사용될 수 있고, 이로써 필요한 시료 부피가 효과적으로 1/5로 줄어든다.

- [0032] 본원에 사용된 용어 "배경 핵산 서열"은 임상적으로 관련성이 있는 핵산 서열에 대한 정상 핵산 서열의 비가 공지된 핵산 서열, 예를 들면 1 대 1 비의 핵산 서열을 의미한다. 일 예로, 배경 핵산 서열 및 임상적으로 관련성이 있는 핵산 서열은 이형접합성으로 인해 구별되는 동일 염색체로부터 얻은 2개의 대립유전자이다. 다른 예에서, 배경 핵산 서열은 임상적으로 관련성이 있는 핵산 서열인 하나의 대립유전자에 이형접합성인 다른 대립유전자이다. 또한, 각각의 배경 핵산 서열 및 임상적으로 관련성이 있는 핵산 서열 중 일부는 여러 개인으로부터 입수한 것일 수 있다.
- [0033] 본원에 사용된 용어 "기준 핵산 서열"은 반응당 평균 농도가 공지되어 있거나 동등하게 측정되는 핵산 서열을 의미한다.
- [0034] 본원에 사용된 용어 "과표현된 핵산 서열"은 생체 시료 중의 관심있는 2개의 서열(예를 들면, 임상적으로 관련성이 있는 서열 및 배경 서열) 중에서 다른 서열보다 더 많이 존재하는 핵산 서열을 의미한다.
- [0035] 본원에 사용된 용어 "~에 기초하는"은 "적어도 부분적으로 ~에 기초하는"을 의미하고, 임의의 방법의 입력 값과 그 방법의 출력 값의 관계에서 발생하는 것과 같이, 한 값(또는 결과)이 다른 값의 결정에 사용되는 것을 의미한다. 또한, 본원에 사용된 용어 "도출한다"는 도출이 식의 계산인 경우에 발생하는 것과 같이, 임의의 방법의 입력 값과 그 방법의 출력 값의 관계를 의미한다.
- [0036] 본원에 사용된 용어 "정량적 데이터"는 하나 이상의 반응으로부터 구해지고 하나 이상의 수치를 제공하는 데이터를 의미한다. 예를 들면, 특정 서열에 대해 형광 마커를 나타내는 웰의 수가 정량적 데이터일 것이다.
- [0037] 본원에 사용된 용어 "파라미터"는 정량적 데이터 집합을 규정하는 수치 및/또는 정량적 데이터 집합 사이의 수치 관계식을 의미한다. 예를 들면, 제1 핵산 서열의 제1 양과 제2 핵산 서열의 제2 양 사이의 비(또는 비의 함수)가 파라미터이다.
- [0038] 본원에 사용된 용어 "컷오프 값"은 생체 시료에 대해 2 이상의 상태(예를 들면, 질병 상태 및 비질병 상태) 간의 분류의 중재에 사용되는 수치를 의미한다. 예를 들면, 파라미터가 컷오프 값보다 큰 경우, 정량적 데이터의 제1 분류(예를 들면, 질병 상태)가 이루어지고, 또는 파라미터가 컷오프 값보다 작은 경우, 정량적 데이터의 다른 분류(예를 들면, 비질병 상태)가 이루어진다.
- [0039] 본원에 사용된 용어 "불균형"은 기준 수량으로부터 임상적으로 관련성이 있는 핵산 서열의 수량에서의 하나 이상의 컷오프 값에 의해 정의되는 임의의 유의 편차를 의미한다. 예를 들면, 기준 수량은 3/5의 비일 수 있고, 따라서 측정된 비가 1:1인 경우 불균형이 발생할 것이다.
- [0040] 본원에 사용된 용어 "염색체 이수성"은 이배체 게놈의 정량적 양으로부터 염색체의 정량적 양에서의 변이를 의미한다. 상기 변이는 획득 또는 소실일 수 있다. 이는 하나의 염색체 전체 또는 염색체 부위와 관련될 수 있다.
- [0041] 본원에 사용된 용어 "무작위 서열분석"은 서열분석 절차 전에 서열분석되는 핵산 단편이 특이적으로 확인되거나 또는 표적화되지 않는 서열분석을 의미한다. 특이적 유전자좌를 표적화하는 서열 특이적 프라이머가 필요하지 않다. 서열분석되는 핵산 풀(pool)은 동일 시료에서도 시료마다, 심지어 분석마다 다를 수 있다. 서열분석되는 핵산의 확인은 오로지 생성되는 서열분석 출력 값으로부터 밝혀진다. 본 발명의 몇몇 실시양태에서, 특정한 일반적인 특징을 공유하는 특정한 핵산 분자 모집단을 생체 시료에 농축시키는 절차가 무작위 서열분석보다 선행될 수 있다. 일 실시양태에서, 생체 시료 내의 각각의 단편은 서열분석될 가능성이 동일하다.
- [0042] 본원에 사용된 용어 "인간 게놈의 일부" 또는 "인간 게놈의 일부분"은 30억 개 정도의 뉴클레오타이드 염기쌍으로 이루어지는 인간 게놈에서의 뉴클레오타이드 서열의 100% 미만을 의미한다. 서열분석에 있어서, 상기 용어는 인간 게놈에서 뉴클레오타이드 서열의 1배 미만의 커버율을 의미한다. 상기 용어는 뉴클레오타이드/염기쌍의 백분율 또는 절대 수로서 표현될 수 있다. 일 이용 예로서, 상기 용어는 서열분석이 수행되는 실제량을 의미하는 데에 사용될 수 있다. 실시양태에서 정확한 진단을 얻기 위해 서열분석되는 인간 게놈의 비에 필요한 최소 값을 결정할 수 있다. 다른 이용 예로서, 상기 용어는 질병 분류에 대한 파라미터 또는 양을 도출하는 데에 사용되는 서열분석되는 데이터의 양을 의미할 수 있다.
- [0043] 본원에 사용된 용어 "서열분석 태그"는 핵산 분자의 임의의 일부 또는 전부로부터 서열분석된 뉴클레오타이드의 문자열을 의미한다. 예를 들면, 서열분석 태그는 핵산 단편으로부터 서열분석된 뉴클레오타이드의 단문자열 또는 핵산 단편의 양쪽 말단으로부터 서열분석된 뉴클레오타이드의 단문자열이거나, 또는 생체 시료에 존재하는 핵산 단편 전체를 서열분석한 것일 수 있다. 핵산 단편은 더 큰 핵산 분자의 임의의 일부에 해당한다. 단편(예를 들면, 유전자)은 더 큰 핵산 분자의 다른 일부에 독립적으로 존재할 수 있다(즉, 연결되지 않을 수 있다).

- [0044] 발명의 상세한 설명
- [0045] 본 발명의 실시양태는 임상적으로 관련성이 있는 염색체 부위의 증가 또는 감소가 비질병 상태와 비교하여 존재하는지(질병 상태) 여부를 결정하기 위한 방법, 시스템 및 장치를 제공한다. 상기 결정은 생체 시료 내에 임상적으로 관련성이 없는 염색체 부위(배경 부위)에 대한 다른 임상적으로 관련성이 있는 염색체 부위의 양의 파라미터를 이용하여 수행할 있다. 게놈의 일부가 서열분석되도록 생체 시료의 핵산 분자가 서열분석되며, 상기 양은 서열분석의 결과로부터 결정될 수 있다. 예를 들면, 2의 염색체 부위(또는 부위 집단)의 양의 비에 대해, 기준 수량과 비교하여 변화(즉, 불균형)가 존재하는지 여부를 결정하기 위해 하나 이상의 컷오프 값을 선택한다.
- [0046] 기준 수량에서 검출된 변화는 임상적으로 관련성이 없는 서열에 대한 다른 임상적으로 관련성이 있는 핵산 서열의 관계에서 임의의 (상향 또는 하향) 편차일 수 있다. 따라서, 기준 상태는 임의의 비 또는 (예를 들면, 1-1 대응이 아닌) 다른 수량일 수 있고, 변화를 나타내는 측정된 상태는 하나 이상의 컷오프 값에 의해 결정된 기준 수량과 다른 임의의 비 또는 다른 수량일 수 있다.
- [0047] 임상적으로 관련성이 있는 염색체 부위(소위 임상적으로 관련성이 있는 핵산 서열이라고도 함) 및 배경 핵산 서열은 제1 유형의 세포로부터 그리고 하나 이상의 제2 유형의 세포로부터 얻을 수 있다. 예를 들면, 태아/태반 세포에서 유래하는 태아 핵산 서열은 모체 혈장과 같은 생체 시료에 존재하고, 이 생체 시료는 모체 세포에서 유래하는 모체 핵산 서열의 배경을 함유한다. 일 실시양태에서, 컷오프 값은 적어도 부분적으로 생체 시료 내의 제1 유형의 세포의 백분율에 기초하여 결정된다. 유의할 것은 시료 내의 태아 서열의 백분율은 임의의 태아 유래 유전자좌에 의해 결정될 수 있고 임상적으로 관련성이 있는 핵산 서열을 측정하는 것으로 제한되지 않는다는 것이다. 다른 실시양태에서, 상기 컷오프 값은 적어도 부분적으로 혈장, 혈청, 타액 또는 소변과 같은 생체 시료 내의 중앙 서열의 백분율에 기초하여 결정되고, 이 생체 시료는 생체 내의 비악성 세포에서 유래하는 핵산 서열의 배경을 함유한다.
- [0048] I. 일반적인 방법
- [0049] 도 1은 본 발명의 실시양태에 따라 임신한 여성 피험자로부터 입수한 생체 시료에서 태아 염색체 이수성을 산정 진단하는 방법(100)의 흐름도이다.
- [0050] 단계(110)에서, 임신한 여성으로부터 생체 시료를 입수한다. 생체 시료는 혈장, 소변, 혈청 또는 임의의 다른 적합한 시료일 수 있다. 생체 시료는 태아 및 임신한 여성으로부터 입수한 핵산 분자를 함유한다. 예를 들면, 핵산 분자는 염색체의 단편일 수 있다.
- [0051] 단계(120)에서, 생체 시료에 함유된 복수의 핵산 분자 중 적어도 일부분을 서열분석한다. 서열분석된 일부분은 인간 게놈의 일부에 해당한다. 일 실시양태에서, 핵산 분자는 각각의 염색체의 단편이다. 한쪽 말단(예를 들면, 35개의 염기쌍(bp)), 양쪽 말단 또는 전체 단편을 서열분석할 수 있다. 생체 시료 내의 핵산 분자 모두를 서열 분석할 수 있거나, 또는 부분 집합만을 서열분석할 수 있다. 상기 부분 집합을 무작위로 선택할 수 있고, 이는 하기 더 자세히 기재되어 있다.
- [0052] 일 실시양태에서, 대규모 병렬 서열분석을 이용하여 서열분석을 수행한다. 대규모 병렬 서열분석, 예컨대 454 플랫폼(Roche)(Margulies, M. et al. 2005 Nature 437, 376-380), 일루미나 게놈 애널리저(Illumina Genome Analyzer)(또는 Solexa platform) 또는 SOLiD System(Applied Biosystems)에서 달성 가능한 대규모 병렬 서열 분석 또는 Helicos True Single Molecule DNA 서열분석 기술(Harris TD et al. 2008 Science, 320, 106-109), 퍼시픽 바이오사이언스(Pacific Biosciences)의 단일 분자 실시간(SMRT™) 기술 및 나노기공 서열분석(Soni GV and Meller A. 2007 Clin Chem 53: 1996-2001)에 의해 병렬 방식으로 높은 다중화도로 검체로부터 단리된 많은 핵산 분자를 서열분석할 수 있다(Dear Brief Funct Genomic Proteomic 2003; 1: 397-416). 상기 플랫폼의 각각은 핵산 단편의 클론 증폭된 또는 심지어 비증폭된 단일 분자를 서열분석한다.
- [0053] 10만 내지 100만 또는 심지어 가능하게는 1억 또는 10억의 순으로 각각의 실행에서 각각의 시료로부터 서열분석 관독(lead) 수가 많아지면서, 얻어진 서열분석 관독은 원래 검체에서 핵산 종의 혼합의 전형적인 프로파일을 형성한다. 예를 들면, 서열분석 관독의 일배체형, 전사체 및 메틸화 프로파일은 원래 검체의 프로파일과 유사하다(Brenner et al Nat Biotech 2000; 18: 630-634; Taylor et al Cancer Res 2007; 67: 8511-8518). 각각의 검체로부터 더 많은 서열의 시료 채취로 인해, 이상적인 서열 수, 예컨대 수백의 커버율 또는 높은 중복성(redundancy)에서의 핵산 풀의 서열분석으로부터 생성되는 서열의 수가 또한 원래 시료 내의 특정 핵산 종 또는 유전자좌의 수의 우수한 정량적 표현이다.

- [0054] 단계(130)에서, 서열분석(예를 들면, 서열분석으로부터 얻은 데이터)에 기초하여, 제1 염색체의 제1 양(예를 들면, 임상적으로 관련성이 있는 염색체)을 결정한다. 제1 양을 제1 염색체에서 유래하는 것으로 확인된 서열로부터 결정한다. 이후, 예를 들면 상기 각각의 DNA 서열을 인간 게놈에 위치시키는 데에 생물정보 절차를 이용할 수 있다. 상기 서열의 부분을 후속 분석에서 제외할 수 있는 데, 왜냐하면 상기 서열이 인간 게놈의 반복 부위 또는 복제수 변이와 같은 개인간 변이가 일어나는 부위에 존재하기 때문이다. 따라서, 관심있는 염색체의 양 및 하나 이상의 다른 염색체의 양을 결정할 수 있다.
- [0055] 단계(140)에서, 서열분석에 기초하여, 하나 이상의 제2 염색체의 제2 양을 제2 염색체 중 하나로부터 유래하는 것으로 확인된 서열로부터 결정한다. 일 실시양태에서, 제2 염색체는 제1 염색체(즉, 검사하고자 하는 염색체)가 아닌 다른 염색체 모두이다. 다른 실시양태에서, 제2 염색체는 바로 단일의 다른 염색체이다.
- [0056] 서열분석 태그 수, 서열분석된 뉴클레오타이드(염기쌍)의 수 또는 특정 염색체(들) 또는 염색체 부위에서 유래하는 서열분석된 뉴클레오타이드(염기쌍)의 누적 길이를 계수하는 것(이에 제한되지 않음)을 비롯하여 염색체 양을 결정하는 여러 방식이 존재한다.
- [0057] 다른 실시양태에서, 계수하여 얻고자 하는 것을 결정하기 위해 서열분석의 결과에 규칙을 도입할 수 있다. 일 양태에서, 서열분석된 출력 값의 비율에 기초하여 일정 양을 얻을 수 있다. 예를 들면, 생물정보 분석 후에 지정된 크기 범위의 핵산 단편에 해당하는 서열분석 출력 값을 선택할 수 있다. 크기 범위의 예로 약 <300 bp, <200 bp 또는 <100 bp를 들 수 있다.
- [0058] 단계(150)에서, 제1 양 및 제2 양으로부터 파라미터를 결정한다. 이 파라미터는 예를 들면 제2 양에 대한 제1 양의 단순한 비이거나, 또는 제1 양 + 제2 양에 대한 제1 양의 단순한 비일 수 있다. 일 양태에서, 각각의 양은 함수 또는 별도의 함수에 대한 독립 변수일 수 있고, 이 별도의 함수로부터 비를 취할 수 있다. 당업자라면 여러 적합한 파라미터의 수를 인지하고 있을 것이다.
- [0059] 일 실시양태에서, 이후 21번 염색체 또는 18번 염색체 또는 13번 염색체와 같은 염색체 이수성에 잠재적으로 관여하는 염색체의 파라미터(예를 들면, 표현 비)를 생물정보 절차의 결과로부터 계산할 수 있다. 표현 비를 서열 모두의 양(예를 들면, 임상적으로 관련성이 있는 염색체를 비롯한 염색체 전부 중 일부의 양) 또는 염색체(예를 들면, 검사되는 염색체가 아닌 단지 하나의 다른 염색체)의 특정 부분 집합에 기초하여 얻을 수 있다.
- [0060] 단계(150)에서, 파라미터를 하나 이상의 컷오프 값과 비교한다. 임의의 수의 적합한 방식으로부터 컷오프 값을 결정할 수 있다. 이러한 방식으로는 베이즈 우도 방법(Bayesian-type likelihood method), 순차 확률비 검정(SPRT; sequential probability ratio testing), 오류 발견, 신뢰 구간, 수용자 작용 특성(ROC; receiver operating characteristic)을 들 수 있다. 상기 방법 및 시료 특이적 방법의 적용의 예는 공존 출원된 "핵산 서열 불균형의 결정 방법"의 특허 출원(변리사 문서 번호 제016285-005210US호)에 기재되어 있고, 이는 참조로 본원에 인용된다.
- [0061] 일 실시양태에서, 이후 파라미터(예를 들면, 임상적으로 관련성이 있는 염색체의 표현 비)를 정상(즉, 정배수체) 태아를 임신한 임산부에서 확립된 기준 범위와 비교한다. 이런 절차의 약간의 변형으로, 기준 범위(즉, 컷오프 값)를 특정 모체 혈장 시료 내의 태아 DNA의 농도 비(f)에 따라 조정할 수 있다. 예를 들면 태아가 스타아인 경우 Y 염색체로 맵핑 가능한 서열을 이용하여 서열분석 데이터 집합으로부터 f의 값을 결정할 수 있다. 또한, 태아의 후생유전학적 마커 등을 이용하는 별도의 분석에서(Chan KCA et al 2006 Clin Chem 52, 2211-8) 또는 단일 염기 다형성의 분석으로부터 f의 값을 결정할 수 있다.
- [0062] 단계(160)에서, 비교에 기초하여, 제1 염색체에 대해 태아 염색체 이수성이 존재하는지 여부에 대해 분류한다. 일 실시양태에서, 이런 분류는 유 또는 무로 제한적이다. 다른 실시양태에서, 이런 분류는 분류할 수 없는 것 또는 불분명한 것일 수 있다. 또 다른 실시양태에서, 이런 분류는 의사 등이 나중에 해석하는 점수일 수 있다.
- [0063] II. 서열분석, 정렬 및 양의 결정
- [0064] 상기 기재된 바대로, 게놈의 일부만을 서열분석한다. 일 양태에서, 검체 내의 핵산 풀은 수배의 커버율 대신에 <100%의 게놈 커버율로 서열분석되고, 포획된 핵산 분자의 부분 중에서 각각의 핵산 종의 대부분은 오직 1회 서열분석된다. 또한, 특정 염색체 또는 염색체 부위의 양 불균형을 정량적으로 결정할 수 있다. 즉, 염색체 또는 염색체 부위의 양 불균형을 검체의 다른 맵핑 가능한 서열분석 태그 중에서 상기 유전자좌의 표현의 백분율로부터 추론한다.
- [0065] 이는 동일 핵산 풀이 높은 중복성 또는 수배의 커버율을 달성하기 위해(이에 의해 각각의 핵산 종이 수회 서열

분석됨) 수회 서열분석되는 경우와 대비된다. 이러한 경우에서, 특정 핵산 종이 서열분석되는 횟수에 대한 다른 핵산 종이 서열분석되는 횟수는 원래 시료 내의 이의 상대 농도와 연관성이 있다. 서열분석 비용은 핵산 종의 정확한 표현을 달성하기 위해 필요한 커버율 배수에 따라 증가한다.

[0066] 일 예에서, 상기 서열의 비율은 본원 실시예에서의 21번 염색체와 같은 이수성에 관여하는 염색체로부터 얻는다. 이런 서열분석 실행으로부터 얻은 또 다른 서열은 다른 염색체에서 유래할 것이다. 다른 염색체와 비교되는 21번 염색체의 상대 크기를 고려하여, 이런 서열분석 실행으로부터 얻은 21번 염색체 특이적 서열의 정규화 빈도를 기준 범위 내에서 얻을 수 있을 것이다. 태아가 21번 삼염색체증을 갖는 경우, 이런 서열분석 실행으로부터 얻은 21번 염색체 유래 서열의 정규화 빈도가 증가할 것이고, 이로써 21번 삼염색체증을 검출할 수 있다. 정규화 빈도의 변화 정도는 분석된 시료 내의 태아 핵산의 농도 비에 따라 달라질 것이다.

[0067] 일 실시양태에서, 본 발명자들은 인간 게놈 DNA 및 인간 혈장 DNA 시료의 단일 말단 서열분석을 위해 일루미나 게놈 애플라이저를 이용하였다. 일루미나 게놈 애플라이저는 유동 셀이라 칭하는 고체 표면에 포획된 클론 증폭된 단일 DNA 분자를 서열분석한다. 각각의 유동 셀은 8개의 개별 검체 또는 검체 풀의 서열분석을 위해 8개의 레인을 갖는다. 각각의 레인은 인간 게놈에서의 30억 염기쌍 서열의 오직 일부인 약 200 Mb의 서열을 생성시킬 수 있다. 각각의 게놈 DNA 또는 혈장 DNA 시료를 유동 셀 중 하나의 레인을 이용하여 서열분석한다. 생성된 짧은 서열분석 태그를 인간 기준 게놈 서열로 정렬하고 염색체 기원을 기록한다. 각각의 염색체로 정렬되는 각각의 서열분석 태그의 총 수를 표로 만들고 기준 인간 게놈 또는 비질병 대표 검체로부터 예상되는 각각의 염색체의 상대 크기와 비교한다. 이후, 염색체 획득 또는 손실을 확인한다.

[0068] 기재된 접근법은 본원에 기재된 유전자/염색체 용량 전략의 오직 하나의 예시이다. 또는, 말단 쌍 서열분석을 수행할 수 있다. 문헌[Campbell et al (Nat Genet 2008; 40: 722-729)]에 기재된 바대로 기준 게놈에서 예상되는 길이로부터 서열분석된 단편의 길이를 비교하는 대신에, 정렬되는 서열분석 태그 수를 계수하고 염색체 위치에 따라 분류한다. 태그 수를 기준 게놈에서 예상된 염색체 크기 또는 비질병 대표 검체의 염색체 크기와 비교하여 염색체 부위 또는 전체 염색체의 획득 또는 손실을 결정한다. 말단 쌍 서열분석에 의해 원래 핵산 단편의 크기를 추론할 수 있으므로, 일 예는 <300 bp, <200 bp 또는 <100 bp와 같은 지정된 크기의 핵산 단편에 해당하는 쌍 서열분석 태그 수를 계수하는 것에 집중한다.

[0069] 다른 실시양태에서, 1회 실행으로 서열분석되는 핵산 풀의 일부를 서열분석 전에 추가로 부분 선택한다. 예를 들면, 올리고뉴클레오타이드 어레이와 같은 하이브리드화 기반 기술을 이용하여 우선 특정 염색체, 예를 들면 이수체 가능성이 있는 염색체 및 검사되는 이수성에 관여하지 않는 다른 염색체(들)로부터 핵산 서열을 부분 선택한다. 다른 예는 서열분석 전에 시료 풀로부터 핵산 서열의 특정 아집단을 부분 선택하거나 또는 농축시키는 것이다. 예를 들면, 상기 기재된 바대로, 모체 혈장 내의 태아 DNA 분자는 모체 배경 DNA 분자보다 더 짧은 단편으로 이루어지는 것으로 보고되어 있다(Chan et al Clin Chem 2004; 50: 88-92). 따라서, 예를 들면 겔 전기영동 또는 크기 배제 칼럼에 의해 또는 미세유체 장치 기반 접근법에 의해 분자 크기에 따라 시료 내의 핵산 서열을 분할하는 당업자에게 공지된 하나 이상의 방법을 이용할 수 있다. 또한, 대안적으로, 모체 혈장 내의 세포 유리 태아 DNA를 분석하는 예에서, 태아 핵산의 일부분을 예컨대 포름알데하이드를 첨가하여 모체 배경을 억제하는 방법에 의해 농축시킬 수 있다(Dhallan et al JAMA 2004; 291: 1114-9). 일 실시양태에서, 미리 선택된 핵산 풀의 일부분 또는 부분 집합을 무작위로 서열분석한다.

[0070] 다른 단일 분자 서열분석 전략, 예컨대 454 플랫폼(Roche) 또는 SOLiD 플랫폼(Applied Biosystems)에 의한 서열분석 전략, Helicos True Single Molecule DNA 서열분석 기술, 퍼시픽 바이오사이언스의 단일 분자 실시간(SMRT™) 기술 및 나노기공 서열분석을 본원에서 유사하게 이용할 수 있다.

[0071] III. 서열분석 출력 값으로부터 염색체 양의 결정

[0072] 대규모 병렬 서열분석 후, 생물정보 분석을 수행하여 서열분석 태그의 염색체 기원을 좌위시킨다. 이런 절차 후, 이수체 가능성이 있는 염색체(즉, 본 연구에서는 21번 염색체)에서 유래하는 것으로 확인된 태그를 서열분석 태그 또는 이수성에 관여하지 않는 하나 이상의 염색체에서 유래하는 태그의 모두와 정량적으로 비교한다. 검체가 정배수체 태아 또는 21번 삼염색체증 태아를 임신한 임신부로부터 입수된 것인지 여부를 결정하기 위해 검체에 대한 21번 염색체의 서열분석 출력 값과 21번이 아닌 다른 염색체의 서열분석 출력 값 간의 관계식을 상기 II 섹션에 기재된 방법으로부터 도출된 컷오프 값과 비교한다.

[0073] 여러 상이한 양은 서열분석 태그로부터 도출될 수 있는 다음의 양을 포함하지만, 이에 한정되지는 않는다. 예를 들면, 서열분석 태그 수, 즉 특정 염색체로 정렬된 절대 수를 다른 염색체로 정렬되는 서열분석 태그의 절대 수

와 비교할 수 있다. 또는, 다른 서열분석 태그의 전부 또는 일부에 대한 21번 염색체의 서열분석 태그 양의 비를 이수체가 아닌 다른 염색체의 서열분석 태그 양의 비와 비교할 수 있다. 본 실험에서, 각각의 DNA 단편으로부터 36 bp가 서열분석되므로, 36 bp에 서열분석 태그 수를 곱하여 특정 염색체로부터 서열분석된 뉴클레오타이드의 수를 용이하게 도출할 수 있다.

[0074] 또한, 각각의 모체 혈장 검체가 오로지 하나의 유동 셀(오로지 인간 게놈의 일부만 서열분석할 수 있음)을 이용하여 서열분석되므로, 통계에 의해, 모체 혈장 DNA 단편 종의 대부분이 오로지 각각 서열분석되어 하나의 서열분석 태그 수가 생성된다. 즉, 모체 혈장 검체에 존재하는 핵산 단편은 1배 미만의 커버율로 서열분석된다. 따라서, 임의의 특정 염색체에 대해 서열분석된 뉴클레오타이드의 총 수는 서열분석된 상기 염색체의 일부의 양, 비율 또는 길이에 거의 상응할 것이다. 그러므로, 이수체 가능성이 있는 염색체의 표현의 정량적 결정은 다른 염색체에 대해 유사하게 도출된 수량에 대한 이 염색체로부터 서열분석된 뉴클레오타이드의 수 또는 등가 길이의 비로부터 도출될 수 있다.

[0075] IV. 서열분석을 위한 핵산 풀에 대한 농축

[0076] 상기 기재되어 있고 하기 실시예 섹션에 설정된 바대로, 정배수체 사례로부터 21번 삼염색체증 사례를 구별하기 위해 인간 게놈의 오직 일부분만을 서열분석할 필요가 있다. 따라서, 농축된 풀의 일부의 무작위 서열분석 전에 서열분석하고자 하는 핵산 풀을 농축시킬 수 있고 이렇게 하는 것이 비용 효과적이다. 예를 들면, 모체 혈장 내의 태아 DNA 분자는 모체 배경 DNA 분자보다 더 짧은 단편으로 이루어진다(Chan et al Clin Chem 2004; 50: 88-92). 따라서, 예를 들면 겔 전기영동 또는 크기 배제 칼럼에 의해 또는 미세유체 장치 기반 접근법에 의해 분자 크기에 따라 시료 내의 핵산 서열을 분할하는 당업자에게 공지된 하나 이상의 방법을 이용할 수 있다.

[0077] 또한, 대안적으로, 모체 혈장 내의 세포 유리 태아 DNA를 분석하는 예에서, 태아 핵산의 일부분을 예컨대 포름알데하이드를 첨가하여 모체 배경을 억제하는 방법에 의해 농축시킬 수 있다(Dhallan et al JAMA 2004; 291: 1114-9). 태아 유래 서열의 부분은 더 짧은 단편으로 이루어진 핵산 풀에 농축된다. 도 7에 따르면, 태아 DNA의 농도 비가 증가함에 따라 21번 삼염색체증 사례로부터 정배수체 사례를 구별하는 데에 필요한 서열분석 태그 수가 감소한다.

[0078] 또는, 이수체 가능성이 있는 염색체 및 이수성에 관여하지 않는 하나 이상의 염색체에서 유래하는 서열을 하이브리드화 기술에 의해 예를 들면 올리고뉴클레오타이드 마이크로어레이에 농축시킬 수 있다. 이후, 농축된 핵산 풀을 무작위로 서열분석할 수 있다. 이로써 서열분석 비용을 줄일 수 있다.

[0079] V. 무작위 서열분석

[0080] 도 2는 본 발명의 실시양태에 따른 무작위 서열분석을 이용하여 태아 염색체 이수성을 산전 진단하는 방법(200)의 흐름도이다. 대규모 병렬 서열분석 접근법에 대한 일 양태에서, 염색체 모두로부터 얻은 대표적인 데이터를 동시에 생성할 수 있다. 특정 단편의 기원을 미리 선택하지 않는다. 서열분석을 무작위로 수행하고, 이후 특정 단편이 어디에서 유래하는지 알아보기 위해 데이터베이스 조사를 수행한다. 이는 21번 염색체로부터 얻은 특정 단편과 1번 염색체로부터 얻은 다른 단편이 증폭되는 경우와 대비된다.

[0081] 단계(210)에서, 임신한 여성으로부터 생체 시료를 입수한다. 단계(220)에서, 원하는 정확도를 위해 분석하고자 하는 서열 수(N)를 계산한다. 일 실시양태에서, 생체 시료 내의 태아 DNA의 백분율을 우선 확인한다. 이는 당업자에게 공지된 임의의 적합한 수단에 의해 수행할 수 있다. 확인은 단순히 다른 집합체(entity)에 의해 측정된 값을 판독하는 것일 수 있다. 이런 실시양태에서, 분석하고자 하는 서열 수(N)의 계산은 백분율에 기초한다. 예를 들면, 분석될 필요가 있는 서열 수는 태아 DNA 백분율이 감소할 때에 증가할 것이고, 태아 DNA 백분율이 증가할 때에 감소할 것이다. 서열 수(N)는 고정 수 또는 백분율과 같은 상대 수일 수 있다. 다른 실시양태에서, 정확한 질병 진단에 적합한 것으로 공지된 서열 수(N)를 서열분석할 수 있다. 심지어 정상 범위의 하한치인 태아 DNA 농도를 갖는 임신부에서도 서열 수(N)가 충분히 만들어질 수 있다.

[0082] 단계(230)에서, 생체 시료에 함유된 N개 이상의 복수의 핵산 분자를 무작위로 서열분석한다. 상기 기재된 접근법의 특징은 서열분석되는 핵산이 시료 분석 전에, 즉 서열분석 전에 특이적으로 확인되거나 또는 표적화되지 않는다는 것이다. 특이적 유전자좌를 표적화하는 서열 특이적 프라이머가 서열분석에 필요하지 않다. 서열분석되는 핵산 풀은 동일 시료에서도 시료마다, 심지어 분석마다 다르다. 또한, 하기 설명(도 6)으로부터, 사례 진단에 필요한 서열분석 출력 값의 양은 검사되는 검체와 기준 모집단 간에 다르다. 이런 양태는 대부분의 분자 진단 접근법, 예컨대 형광 인 시추 하이브리드화, 정량적 형광 PCR, 정량적 실시간 PCR, 디지털 PCR, 비교 게놈 하이브리드화, 마이크로어레이 비교 게놈 하이브리드화 및 기타 등등에 기초한 접근법(표적화되는 유전자좌가

사전 선결될 필요가 있어, 유전자와 특이적 프라이머 또는 프로브 집합 또는 이들의 패널의 사용을 필요로 함)과 현저히 대비된다.

[0083] 일 실시양태에서, 임신한 여성의 혈장에 존재하는 DNA 단편에서 무작위 서열분석을 수행하여, 원래 태아 또는 모체 중 1명으로부터 유래하는 게놈 서열을 얻을 수 있다. 무작위 서열분석은 생체 시료에 존재하는 핵산 분자의 임의의 일부분의 시료 채취(서열분석)를 수반한다. 서열분석이 무작위로 수행되므로, 핵산 분자(및 이로 따라 게놈)의 상이한 부분 집합(일부)이 각각의 분석에서 서열분석될 수 있다. 심지어 상기 부분 집합이 시료마다 그리고 분석마다 다른 경우에도 이런 실시양태가 수행될 수 있는데, 이는 동일 시료를 이용하는 경우에도 발생할 수 있다. 상기 일부의 예는 게놈의 약 0.1%, 0.5%, 1%, 5%, 10%, 20% 또는 30%에 해당한다. 다른 실시양태에서, 상기 일부는 상기 값들 중 적어도 어느 하나에 해당한다.

[0084] 나머지 단계(240~270)는 방법(100)과 유사한 방식으로 진행될 수 있다.

[0085] VI. 서열분석 태그 풀의 서열분석 후 선택

[0086] 하기 실시예 II 및 실시예 III에 기재된 바대로, 서열분석된 데이터의 부분 집합은 정배수체 사례로부터 21번 삼염색체중 사례를 구별하는 데에 충분하다. 서열분석된 데이터의 부분 집합은 특정 품질 파라미터를 통과한 서열분석 태그의 부분이다. 예를 들면, 실시예 II에서, 반복 마스킹된 기준 인간 게놈으로 유일하게 정렬되는 서열분석 태그를 이용하였다. 또는, 모든 염색체로부터 얻은 핵산 단편의 대표적인 풀을 서열분석할 수 있지만, 이수체 가능성이 있는 염색체와 관련된 데이터와 이수체가 아닌 염색체의 수와 관련된 데이터 간의 비교에 집중한다.

[0087] 또한 대안적으로, 원래 검체 내에 지정된 크기 창에 해당하는 핵산 단편으로부터 생성되는 서열분석 태그를 포함하는 서열분석 출력 값의 부분 집합을 서열분석 후의 분석 동안 부분 선택할 수 있다. 예를 들면, 일루미나 게놈 애널라이저를 이용하여, 핵산 단편의 양쪽 말단의 서열분석인 말단 쌍 서열분석을 이용할 수 있다. 이후, 각각의 말단 쌍으로부터 얻은 서열분석된 데이터를 기준 인간 게놈 서열로 정렬한다. 이후, 양쪽 말단 사이에 이르는 뉴클레오타이드의 거리 또는 수를 추론한다. 또한, 원래 핵산 단편의 전체 길이를 추론할 수 있다. 대안적으로, 서열분석 플랫폼, 예컨대 454 플랫폼 및 가능하게는 몇몇의 단일 분자 서열분석 기술에 의해 200 bp와 같은 짧은 핵산 단편의 전체 길이를 서열분석할 수 있다. 이러한 방식으로, 핵산 단편의 실제 길이를 서열분석된 데이터로부터 즉시 알 수 있다.

[0088] 또한, 상기 말단 쌍 분석은 다른 서열분석 플랫폼, 예를 들면 SOLiD system(Applied Biosystems)을 이용할 수 있다. 454 플랫폼의 경우, 다른 대규모 병렬 서열분석 시스템과 비교하여 이의 판독 길이가 증가하므로, 이의 완전한 서열로부터 단편의 길이를 결정할 수도 있다.

[0089] 원래 모체 혈장 검체 내의 짧은 핵산 단편에 해당하는 서열분석 태그의 부분 집합에 데이터 분석을 집중하는 것이 유리한데, 왜냐하면 데이터 집합은 태아에서 유래한 DNA 서열이 효과적으로 농축되어 있기 때문이다. 이는 모체 혈장 내의 태아 DNA 분자가 모체 배경 DNA 분자보다 짧은 단편으로 이루어지기 때문이다(Chan et al Clin Chem 2004; 50: 88-92). 도 7에 따르면, 태아 DNA의 농도 비가 증가함에 따라 21번 삼염색체중 사례로부터 정배수체 사례를 구별하는 데에 필요한 서열분석 태그 수가 감소한다.

[0090] 핵산 풀의 부분 집합의 서열분석 후 선택은 배경 핵산 풀로부터 농축된 풀의 물리적 분리를 필요로 하는 특정 크기의 핵산의 선택을 위한 겔 전기영동 또는 크기 배제 칼럼과 같은 검체 분석 전 수행되는 다른 핵산 농축 전략과 상이하다. 물리적 절차에는 보다 실험적인 단계가 도입되며 오염과 같은 문제가 일어나기 쉽다. 또한, 서열분석 출력 값의 부분 집합의 서열분석 후 인 실리코 선택에 의해 질병 결정에 필요한 민감성 및 특이성에 따라 선택이 변할 수 있다.

[0091] 모체 혈장 검체가 21번 삼염색체중 태아 또는 정배수체 태아를 임신한 여성으로부터 입수된 것인지 여부를 결정하기 위해 사용된 생물정보 접근법, 컴퓨터 접근법 및 통계 접근법은 서열분석 출력 값으로부터 파라미터를 결정하기 위해 사용되는 컴퓨터 프로그램 제품에 컴파일된다. 컴퓨터 프로그램의 연산은 이수체 가능성이 있는 염색체로부터 얻은 정량적 양 및 하나 이상의 다른 염색체로부터 얻은 양(들)을 결정하는 단계를 포함한다. 파라미터를 결정하고 적절한 컷오프 값과 비교하여 이수체 가능성이 있는 염색체에 대해 태아 염색체 이수성이 존재하는지 여부를 결정할 수 있다.

[0092] [실시예]

[0093] 청구된 발명을 제한하기 위해서가 아니라 예시하기 위해 다음의 실시예를 제공한다.

[0094] I. 태아 21번 삼염색체증의 산전 진단

[0095] 이 연구에 8명의 임신한 여성이 지원하였다. 임신한 여성 모두 임신 제1 삼분기 또는 제2 삼분기였고, 단태아를 임신하였다. 8명 중 4명은 각각 21번 삼염색체증 태아를 임신하였고, 나머지 4명은 각각 정배수체 태아를 임신하였다. 각각의 피험자로부터 말초 정맥혈 20 mL를 수집하였다. 모체 혈장을 1600 × g에서 10 분 동안 원심분리한 후 수확하고, 16000 × g에서 10 분 동안 더 원심분리하였다. 이후, DNA를 각각의 혈장 시료 5~10 mL로부터 추출하였다. 이후, 모체 혈장 DNA를 일루미나 게놈 애널라이저에 의해 제조업자의 지침에 따라 대규모 병렬 서열분석에 사용하였다. 서열분석을 수행하는 테크니션은 서열분석 및 서열 데이터 분석 동안 맹검으로 태아 진단하였다.

[0096] 요약하면, 대략 50 ng의 모체 혈장 DNA를 DNA 라이브러리 제조에 사용하였다. 15 ng 또는 10 ng과 같이 더 적은 양의 모체 혈장 DNA로 시작할 수 있다. 모체 혈장 DNA 단편을 평활 말단하고 Solexa 어댑터에 결합하고, 150~300 bp의 단편을 겔 정제에 의해 선택하였다. 대안적으로, 평활 말단되고 어댑터에 결합된 모체 혈장 DNA 단편을 클러스터 생성 전 결합되지 않은 어댑터를 크기 선택 없이 제거하기 위해 칼럼(예를 들면, AMPure, Agencourt)을 통해 통과시켰다. 어댑터에 결합된 DNA를 유동 셀의 표면에 하이브리드화하였고, DNA 클러스터를 일루미나 클러스터 스테이션을 이용하여 생성시킨 후, 일루미나 게놈 애널라이저에서 36회 서열분석하였다. 각각의 모체 혈장 검체로부터 얻은 DNA를 하나의 유동 셀에 의해 서열분석하였다. 서열분석 판독을 Solexa Analysis Pipeline을 이용하여 컴파일하였다. 이후, 모든 판독을 Eland 응용 소프트웨어를 이용하여 반복 마스킹된 기준 인간 게놈 서열, NCBI 36 어셈블리(GenBank 수탁 번호: NC_00001 내지 NC_000024)로 정렬하였다.

[0097] 본 연구에서, 데이터 분석의 복잡성을 줄이기 위해, 반복 마스킹된 기준 인간 게놈에서 독특한 위치로 맵핑된 서열만을 추가로 고려하였다. 서열분석된 데이터의 다른 부분 집합 또는 전체 집합을 대안적으로 사용하였다. 각각의 검체에 대해 유일하게 맵핑 가능한 서열의 총 수를 계수하였다. 21번 염색체로 유일하게 정렬된 서열 수를 각각의 검체에 대해 정렬된 서열의 총 수에 대한 비율로서 표현하였다. 모체 혈장이 모체 기원의 DNA 배경 중에서 태아 DNA를 함유하므로, 21번 삼염색체증 태아는 태아 게놈 내의 21번 염색체의 과잉 카피수의 존재로 인해 21번 염색체에서 유래하는 과잉 서열분석 태그의 원인이 되었다. 따라서, 21번 삼염색체증 태아를 임신한 임산부로부터 입수한 모체 혈장 내의 21번 염색체 서열의 백분율은 정배수체 태아를 임신한 임산부로부터 입수한 모체 혈장 내의 21번 염색체 서열의 백분율보다 높다. 이런 분석에는 태아 특이적 서열을 표적화할 필요가 없다. 또한, 이런 분석에는 모체 핵산으로부터 태아를 사전에 물리적으로 분리할 필요가 없다. 또한, 이런 분석에는 서열분석 후 모체 서열로부터 태아를 구별하거나 또는 확인할 필요가 없다.

[0098] 도 3a는 8개의 모체 혈장 DNA 시료 각각에 대해 21번 염색체로 맵핑되는 서열의 백분율(21번 염색체의 표현의 백분율)을 나타낸 것이다. 21번 염색체의 표현의 백분율은 정배수체 임산부의 모체 혈장에서보다 21번 삼염색체증 임산부의 모체 혈장에서 유의하게 높았다. 이런 데이터는 태아 이수성의 비침습적 산전 진단을 기준 모집단의 표현의 백분율에 비교되는 이수체 염색체의 표현의 백분율을 결정하여 달성할 수 있다는 것을 제시한다. 또는, 실험으로 얻은 21번 염색체의 표현의 백분율을 정배수체 인간 게놈에 예상되는 21번 염색체 서열의 표현의 백분율과 비교하여 21번 염색체 과표현을 검출할 수 있다. 이는 인간 게놈에서 반복 부위를 마스킹하거나 또는 마스킹하지 않고 수행할 수 있다.

[0099] 8명의 임신한 여성 중 5명은 각각 숫태아를 임신하였다. Y 염색체로 맵핑된 서열은 태아 특이적일 것이다. 원래 모체 혈장 검체 내의 태아 DNA의 농도 비를 계산하기 위해 Y 염색체로 맵핑된 서열의 백분율을 사용하였다. 또한, X-연결 아연 손가락 단백질(ZFX) 및 Y-연결 아연 손가락 단백질(ZFY) 직렬상동 유전자를 포함하는 미세유체 장치 디지털 PCR을 이용하여 태아 DNA의 농도 비도 결정하였다.

[0100] 도 3b는 서열분석에 의한 Y 염색체의 표현의 백분율에 의해 추론된 태아 DNA의 농도 비와 ZFY/ZFX 미세유체 장치 디지털 PCR에 의해 결정된 태아 DNA의 농도 비의 상관관계를 나타낸 것이다. 상기 2의 방법에 의해 결정된 모체 혈장 내의 태아 DNA의 농도 비 간에 양의 상관관계가 존재하였다. 상관관계수(r)는 피어슨 상관관계(Pearson correlation) 분석에서 0.917이었다.

[0101] 대표적인 2례에 대해 24개의 염색체(22개의 상염색체, X 염색체 및 Y 염색체) 각각으로 정렬된 모체 혈장 DNA 서열의 백분율을 도 4a에 도시하였다. 1명의 임신한 여성은 21번 삼염색체증 태아를 임신하였고, 다른 1명은 정배수체 태아를 임신하였다. 21번 염색체로 맵핑된 서열의 표현의 백분율은 정상 태아를 임신한 여성과 비교할 때 21번 삼염색체증 태아를 임신한 여성에서 높았다.

[0102] 상기 2례의 모체 혈장 DNA 검체 간의 염색체당 표현의 백분율 차이(%)를 도 4b에 도시하였다. 특정 염색체에 대

한 백분율 차이를 하기 수학식을 이용하여 계산하였다:

[0103] 백분율 차이(%) = $(P_{21} - P_E)/P_E \times 100\%$

[0104] 상기 식 중,

[0105] P_{21} 은 21번 삼염색체증 태아를 임신한 여성에서 특정 염색체로 정렬된 혈장 DNA 서열의 백분율이고,

[0106] P_E 는 정배수체 태아를 임신한 여성에서 특정 염색체로 정렬된 혈장 DNA 서열의 백분율이다.

[0107] 도 4b에 도시된 바대로, 정배수체 태아를 임신한 여성과 비교하여 21번 삼염색체증 태아를 임신한 여성의 혈장에서 21번 염색체 서열의 과표현이 11% 존재하였다. 다른 염색체로 정렬된 서열의 경우, 2레 간의 차이는 5% 미만이었다. 21번 염색체에 대한 표현의 백분율이 정배수체 모체 혈장 시료와 비교하여 21번 삼염색체증에서 증가하므로, 백분율 차이(%)는 대안적으로 21번 염색체 서열에서 과표현의 정도로서 추론될 수 있었다. 21번 염색체 표현의 백분율 간의 차이(%) 및 절대차 이외에, 검사 시료 및 기준 시료로부터 수의 비도 계산할 수 있고 이 비는 정배수체 시료와 비교되는 21번 삼염색체증에서 21번 염색체 과표현의 정도를 나타낸다.

[0108] 각각 정배수체 태아를 임신한 4명의 여성에서, 이의 혈장 DNA 서열의 평균 1.345%를 21번 염색체로 정렬하였다. 21번 삼염색체증 태아를 임신한 4명의 여성에서, 4명의 태아 중 3명은 숫태아였다. 이런 3레 각각에 대해 21번 염색체의 표현의 백분율을 계산하였다. 정배수체 4레의 값으로부터 도출된 21번 염색체 표현의 평균 백분율과 이들 21번 삼염색체증 3레 각각에 대한 21번 염색체 표현의 백분율의 차이(%)를 상기 기재한 바대로 결정하였다. 즉, 정배수체 태아를 임신한 4레의 평균을 이 계산에서 기준으로서 사용하였다. 이들 숫태아 21번 삼염색체증 3레에 대한 태아 DNA의 농도 비를 이의 각각의 Y 염색체 서열의 표현의 백분율로부터 추론하였다.

[0109] 21번 염색체 서열에 대한 과표현의 정도와 태아 DNA의 농도 비 간의 상관관계를 도 5에 도시하였다. 2의 파라미터 사이에 유의적인 양의 상관관계가 존재하였다. 상관계수(r)는 피어슨 상관관계 분석에서 0.898이었다. 이런 결과는 모체 혈장 내의 21번 염색체 서열의 과표현의 정도가 모체 혈장 시료 내의 태아 DNA의 농도 비와 관련된다는 것을 나타낸다. 따라서, 태아 DNA의 농도 비와 관련된 21번 염색체 서열의 과표현의 정도의 컷오프 값을 결정하여 21번 삼염색체증 태아를 임신한 임산부를 확인할 수 있었다.

[0110] 또한, 모체 혈장 내의 태아 DNA의 농도 비의 결정은 서열분석 실행과 독립적일 수 있다. 예를 들면, Y 염색체 DNA 농도는 실시간 PCR, 미세유체 장치 PCR 또는 질량 분광법을 이용하여 미리 결정하였다. 예를 들면, 본 발명자들은 도 3b에서 서열분석 실행 동안 생성된 Y 염색체 수에 기초하여 확립된 태아 DNA 농도와 서열분석 실행에 외부에서 얻어진 ZFY/ZFX 비 간에 우수한 상관관계가 존재한다는 것을 입증하였다. 사실, 태아 DNA 농도는 Y 염색체가 아닌 유전자좌를 이용하여 결정할 수 있어서 암태아에도 적용할 수 있다. 예를 들면, 찬(Chan) 등은 모체 유래 비메틸화 RASSF1A 서열의 배경에서의 임신한 여성의 혈장에서 태아 유래 메틸화 RASSF1A 서열을 검출할 수 있다는 것을 밝혔다(Chan et al, Clin Chem 2006;52:2211-8). 따라서, 메틸화 RASSF1A 서열의 양을 총 (메틸화 및 비메틸화) RASSF1A 서열의 양으로 나뉘 태아 DNA의 농도 비를 결정할 수 있다.

[0111] 혈액 응고 동안 모체 혈액 세포로부터 DNA가 배출되므로 본 발명의 실행을 위해서는 모체 혈장이 모체 혈정보다 바람직할 것으로 예상된다. 따라서, 혈청을 사용하는 경우, 태아 DNA의 농도 비는 모체 혈정보다 모체 혈장에서 낮을 것으로 예상된다. 즉, 모체 혈청을 사용하는 경우, 동일한 시간에 동일한 임신한 여성으로부터 입수한 혈장 시료와 비교할 때 진단하고자 하는 태아 염색체 이수성에 대해 더 많은 서열이 생성될 필요가 있을 것으로 예상된다.

[0112] 태아 DNA의 농도 비를 결정하는 또 다른 대안적인 방법은 임신한 여성과 태아 간의 다형성 차이를 정량검사하는 것이다(Dhallan R, et al. 2007 Lancet, 369, 474-481). 이러한 방법의 예로 임신한 여성이 동형접합성이고, 태아가 이형접합성인 다형성 부위를 표적화하는 것을 들 수 있다. 태아 특이적 대립유전자의 양을 공통 대립유전자의 양과 비교하여 태아 DNA의 농도 비를 결정할 수 있다.

[0113] 하나 이상의 특이적 서열(들)을 검출하고 정량검사하는 비교 게놈 하이브리드화, 마이크로어레이 비교 게놈 하이브리드화, 정량적 실시간 중합효소 연쇄 반응을 비롯한 염색체 이상을 검출하기 위한 선행 기술과 비교하여, 대규모 병렬 서열분석은 선결된 또는 선정된 DNA 서열 집합의 검출 또는 분석에 의존적이지 않다. 검체 풀로부터 무작위로 선택된 대표적인 DNA 분자의 일부를 서열분석하였다. 다양한 염색체 부위로 정렬되는 상이한 서열 분석 태그 수를 관심있는 DNA 종을 함유하는 검체와 관심있는 DNA 종을 함유하는 하지 않는 검체 간에 비교하였다. 검체에서 임의의 소정의 염색체 부위로 정렬된 서열 수(또는 백분율)의 차이로 염색체 이상이 밝혀질 것이

다.

[0114] 다른 예에서, 혈장 세포 유리 DNA에 대한 서열분석 기술을 이용하여 특정 암의 검출을 위한 혈장 DNA에서의 염색체 이상을 검출할 수 있다. 여러 암은 통상적인 염색체 이상의 집합을 갖는다. 복수의 염색체 부위의 변화(증폭 및 결실)를 이용할 수 있다. 따라서, 증폭된 부위로 정렬된 서열의 부분이 증가할 수 있고, 감소된 부위로 정렬된 서열의 부분이 감소할 수 있다. 염색체당 표현의 백분율을 전체 게놈에 대한 임의의 소정의 염색체의 게놈 표현의 백분율로서 표현되는 기준 게놈에서의 각각의 상응하는 염색체의 크기와 비교할 수 있다. 또한, 직접 비교 또는 기준 염색체에 대한 비교를 이용할 수 있다.

[0115] II. 인간 게놈의 일부에 대한 서열분석

[0116] 상기 실시예 I에 기재된 실험에서, 각각의 개별 검체로부터 얻은 모체 혈장 DNA를 하나의 유동 셀만을 이용하여 서열분석하였다. 서열분석 실행에 의해 각각의 검사된 검체로부터 생성된 서열분석 태그 수를 도 6에 도시하였다. T21은 21번 삼염색체증 태아를 임신한 임산부로부터 입수한 시료를 나타낸다.

[0117] 서열분석된 모체 혈장 DNA 단편 각각으로부터 36 bp가 서열분석되므로, 36 bp에 서열분석 태그 수를 곱하여 각각의 검체로부터 서열분석된 뉴클레오타이드/염기쌍의 수를 결정할 수 있고 또한 도 6에 도시하였다. 인간 게놈에서 대략 30억의 염기쌍이 존재하므로, 각각의 모체 혈장 검체로부터 생성된 서열분석의 데이터 양은 10% 내지 13% 정도에 이르는 일부만을 나타낸다.

[0118] 또한, 본 연구에서, 오직 유일하게 맵핑 가능한 서열분석 태그(Eland 소프트웨어로부터 명명법에서 U0라 칭함)를 상기 실시예 I에 기재된 바대로 21번 삼염색체증 태아를 임신한 임산부로부터 입수한 모체 혈장 검체에서의 21번 염색체 서열의 양의 과표현의 존재를 입증하기 위해 사용하였다. 도 6에 도시된 바대로, U0 서열은 오직 각각의 검체로부터 생성되는 서열분석 태그의 모두의 부분 집합을 나타내고, 더 나아가서는 인간 게놈의 2% 정도의 훨씬 더 낮은 비율을 나타낸다. 상기 데이터는 태아 이수성의 진단을 달성하기 위해 검사된 검체에 존재하는 인간 게놈 서열의 일부분만을 서열분석하는 것으로 충분하다는 것을 나타낸다.

[0119] III. 필요한 서열 수의 결정

[0120] 이 분석에서 정배수체 숫태아를 임신한 여성으로부터 입수한 혈장 DNA의 서열분석 결과를 사용하였다. 기준 인간 게놈 서열로 미스 매칭되지 않고 맵핑될 수 있는 서열분석 태그 수는 1,990,000개이었다. 이들 1,990,000개의 태그로부터 서열의 부분 집합을 무작위로 선택하고, 21번 염색체로 정렬된 서열의 백분율을 각각의 부분 집합 내에서 계산하였다. 부분 집합에서의 서열 수는 60,000개 내지 540,000개의 서열 수로 변할 수 있다. 각각의 부분 집합 크기에 대해, 다른 조합이 가능하지 않을 때까지 총 풀로부터 서열분석 태그를 무작위로 선택하여 동일 수의 서열분석 태그의 복수의 부분 집합을 컴파일하였다. 이후, 각각의 부분 집합 크기 내에서 복수의 부분 집합으로부터 21번 염색체로 정렬된 서열의 평균 백분율 및 이의 표준 편차(SD)를 계산하였다. 21번 염색체로 정렬된 서열의 백분율의 분포에 미치는 부분 집합 크기의 영향을 결정하기 위해 상이한 부분 집합 크기에 걸쳐 이들 데이터를 비교하였다. 이후, 백분율의 5 백분위수 및 95 백분위수를 평균 및 SD에 따라 계산하였다.

[0121] 임신한 여성이 21번 삼염색체증 태아를 임신한 경우, 21번 염색체로 정렬되는 서열분석 태그는 태아로부터의 21번 염색체의 양 과잉으로 인해 모체 혈장에서 과표현될 것이다. 과발현의 정도는 하기 방정식에 따라 모체 혈장 DNA 시료에서 태아 DNA 백분율에 의존적이다:

$$[0122] \text{Per}_{T21} = \text{Per}_{Eu} \times (1 + f/2)$$

[0123] 상기 식 중,

[0124] Per_{T21} 은 21번 삼염색체증 태아를 임신한 여성에서 21번 염색체로 정렬된 서열의 백분율을 나타내고,

[0125] Per_{Eu} 는 정배수체 태아를 임신한 여성에서 21번 염색체로 정렬된 서열의 백분율을 나타내며,

[0126] f 는 모체 혈장 DNA 내의 태아 DNA 백분율을 나타낸다.

[0127] 도 7에 도시된 바대로, 각각의 부분 집합에서 서열 수가 증가함에 따라 21번 염색체로 정렬된 서열의 백분율에 대한 SD가 감소하였다. 따라서, 각각의 부분 집합에서 서열 수가 증가할 때, 5 백분위수와 95 백분위수 간의 간격이 감소하였다. 정배수체 사례 및 21번 삼염색체증 사례에 대한 5%~95% 간격이 중복되지 않을 때, 2군의 사례의 간의 구별이 >95%의 정확도로 가능하다.

[0128] 도 7에 도시된 바대로, 정배수체 사례로부터 21번 삼염색체증 사례를 구별하기 위한 최소 부분 집합 크기는 태

아 DNA 백분율에 의존적이다. 정배수체 사례로부터 21번 삼염색체증 사례를 구별하기 위한 최소 부분 집합 크기는 20%, 10% 및 5%의 태아 DNA 백분율 각각에 대해 120,000개, 180,000개 및 540,000개의 서열이다. 즉, 모체 혈장 DNA 시료가 태아 DNA를 20% 함유할 때 태아가 21번 삼염색체증을 갖는지 여부를 결정하기 위해 분석될 필요가 있는 서열 수는 120,000개가 될 것이다. 분석될 필요가 있는 서열 수는 태아 DNA 백분율이 5% 감소할 때 540,000개 증가할 것이다.

[0129] 36개의 염기쌍 서열분석을 이용하여 데이터가 생성되면서, 120,000개, 180,000개 및 540,000개의 서열이 각각 인간 게놈의 0.14%, 0.22% 및 0.65%에 해당하였다. 임신 초기 임신부로부터 입수한 모체 혈장 내의 태아 DNA 농도의 낮은 범위가 5% 정도인 것으로 보고되면서(Lo, YMD et al. 1998 Am J Hum Genet 62, 768-775), 인간 게놈의 약 0.6%의 서열분석이 어떠한 임신부에 대해서라도 태아 염색체 이수성을 검출할 때에 95% 이상의 정확도의 진단에 필요한 서열분석의 최소 양일 수 있다.

[0130] IV. 무작위 서열분석

[0131] 서열분석된 DNA 단편이 서열분석 실행 동안 무작위로 선택된 것을 예시하기 위해, 본 발명자들은 실시예 I에서 분석된 8개의 모체 혈장 시료로부터 생성되는 서열분석 태그를 얻었다. 각각의 모체 혈장 검체에 대해, 본 발명자들은 미스 매핑되지 않고 21번 염색체로 유일하게 정렬되는 36 bp의 서열분석 태그의 각각의 기준 인간 게놈 서열(NCBI 어셈블리 36)과 관련하여 출발 위치를 결정하였다. 이후, 본 발명자들은 각각의 검체로부터 정렬되는 서열분석 태그 풀에 대한 출발 위치 번호를 오름차순으로 정렬하였다. 본 발명자들은 염색체 22에 대해서도 유사한 분석을 수행하였다. 예시를 위해, 각각의 모체 혈장 검체에 대해 21번 염색체 및 22번 염색체에 대한 처음 10의 출발 위치를 도 8a 및 도 8b에 각각 도시하였다. 도 8a 및 도 8b의 표로부터 DNA 단편의 서열분석된 풀은 시료 간에 동일하지 않다는 것을 이해할 수 있을 것이다.

[0132] 본원에 기재된 임의의 소프트웨어 컴포넌트 또는 기능은 전통적인 또는 객체 지향형 기술을 이용하여 자바(Java), C++ 또는 Perl 등과 같은 임의의 적합한 컴퓨터 언어를 이용하는 처리장치에 의해 실행되는 소프트웨어 코드로서 실행될 수 있다. 소프트웨어 코드는 저장 및/또는 전송을 위해 컴퓨터 판독 가능한 매체에서 일련의 (하드웨어) 명령 또는 (유저) 명령어로서 저장될 수 있고, 적합한 매체로는 임의 접근 메모리(RAM), 읽기 전용 메모리(ROM), 자기 매체, 예컨대 하드 드라이브 또는 플로피 디스크, 또는 광학 매체, 예컨대 콤팩트 디스크(CD) 또는 DVD(디지털 다기능 디스크), 플래쉬 메모리 등을 들 수 있다. 컴퓨터 판독 가능한 매체는 상기 저장 장치 또는 전송 장치의 임의의 조합일 수 있다.

[0133] 또한, 상기 프로그램은 다양한 프로토콜에 합치되는 인터넷을 비롯한 유선 네트워크, 광학 네트워크 및/또는 무선 네트워크를 통한 전송에 조정되는 반송 신호를 이용하여 코딩되고 전송될 수 있다. 그 자체로, 본 발명의 실시양태에 따른 컴퓨터 판독 가능한 매체는 상기 프로그램으로 암호화된 데이터 신호를 이용하여 생성될 수 있다. 프로그램 코드로 암호화된 컴퓨터 판독 가능한 매체는 호환 장치와 함께 패키징화될 수 있거나 또는 (예를 들면, 인터넷 다운로드를 통해) 다른 장치와 별도로 제공될 수 있다. 이러한 임의의 컴퓨터 판독 가능한 매체는 단일 컴퓨터 프로그램 제품(예를 들면, 하드 드라이브 또는 전체 컴퓨팅 시스템) 내에 공존할 수 있고 시스템 또는 네트워크 내의 다른 컴퓨터 프로그램 제품 내에 존재할 수 있다. 컴퓨팅 시스템으로는 모니터, 프린터 또는 사용자에게 본원에 언급된 결과 중 어느 하나를 제공하기 위한 다른 적합한 디스플레이를 들 수 있다.

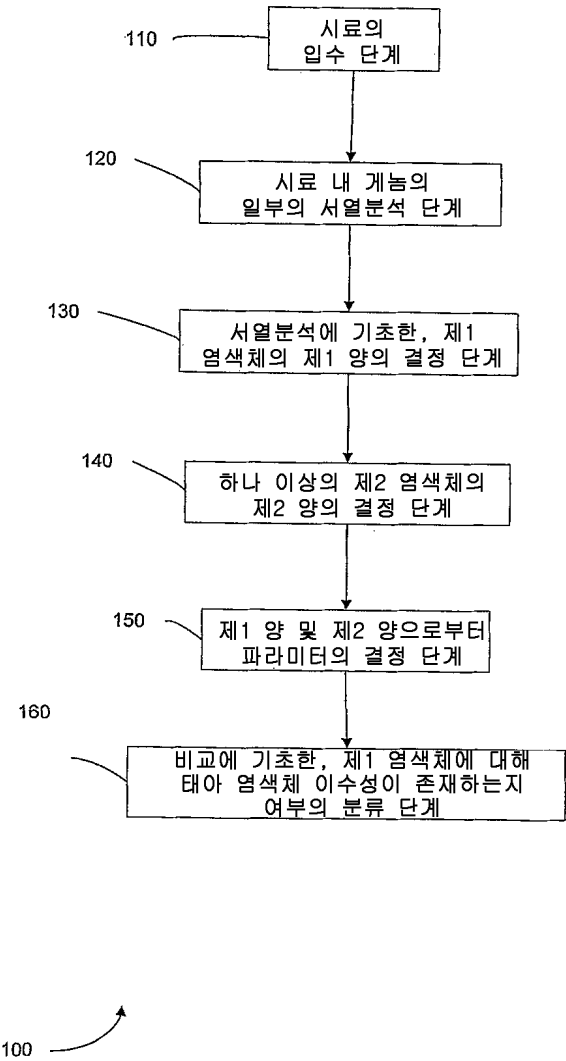
[0134] 컴퓨팅 시스템의 예를 도 9에 도시하였다. 도 9에 도시된 서브시스템은 시스템 버스(975)를 통해 상호접속된다. 부수적인 서브시스템, 예컨대 프린터(974), 키보드(978), 고정 디스크(979), 디스플레이 어댑터(982)에 연결된 모니터(976) 및 다른 서브시스템을 도시하였다. 입력/출력(I/O) 제어기(971)에 연결된 입력/출력(I/O) 장치 및 주변 장치가 직렬 포트(977)와 같은 당해 분야에 공지된 여러 수단에 의해 컴퓨팅 시스템에 연결될 수 있다. 예를 들면, 직렬 포트(977) 또는 외부 인터페이스(981)가 컴퓨터 장치를 광면적 네트워크, 예컨대 인터넷, 마우스 입력 장치 또는 스캐너에 연결하는 데 사용될 수 있다. 시스템 버스를 통한 상호접속에 의해 중앙 처리장치(973)가 각각의 서브시스템과 통신하고 시스템 메모리(972) 또는 고정 디스크(979)로부터 받은 명령의 수행을 제어할 수 있으며, 또한 서브시스템 간의 정보가 교환될 수 있다. 시스템 메모리(972) 및/또는 고정 디스크(979)에는 컴퓨터 판독 가능한 매체가 탑재될 수 있다.

[0135] 본 발명을 예시하고 설명하기 위해 본 발명의 예시적인 실시양태에 대해 상기 설명을 제시하였다. 완전한 것으로 또는 본 발명을 기재된 정확한 형태로 제한하도록 의도되지 않고, 많은 변경예 및 변형예가 상기 교시내용의 견지에서 가능하다. 본 발명의 원리 및 이의 실제 적용을 아주 잘 설명하기 위해 실시양태를 선택하고 기재하였으므로, 당업자는 다양한 실시양태에서 그리고 고안된 특정 용도에 적합한 다양한 변경예로 본 발명을 아주 잘 이용할 수 있을 것이다.

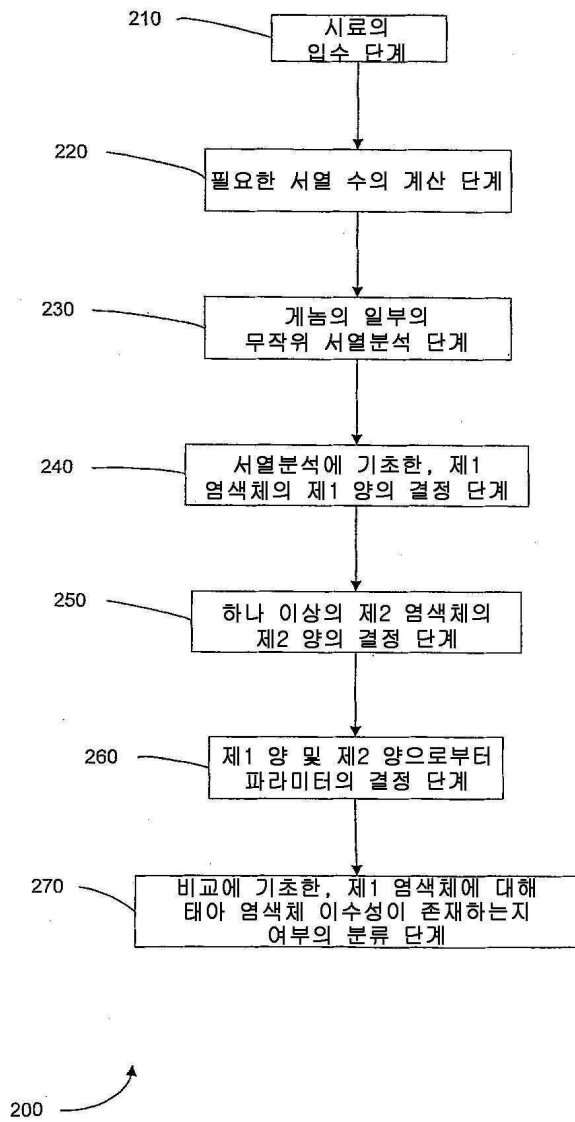
[0136] 본원에 인용된 모든 공보, 특허 및 특허 출원은 사실상 이의 전체로 참조로 본원에 인용된다.

도면

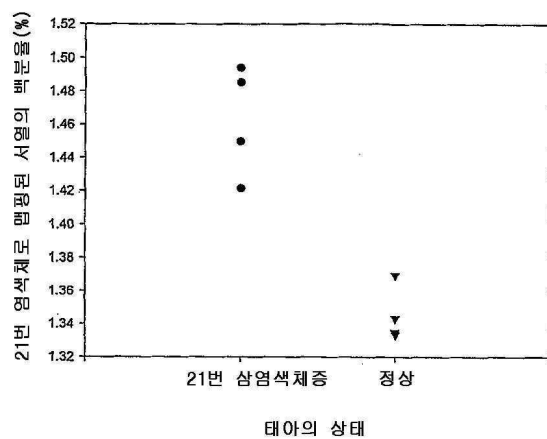
도면1



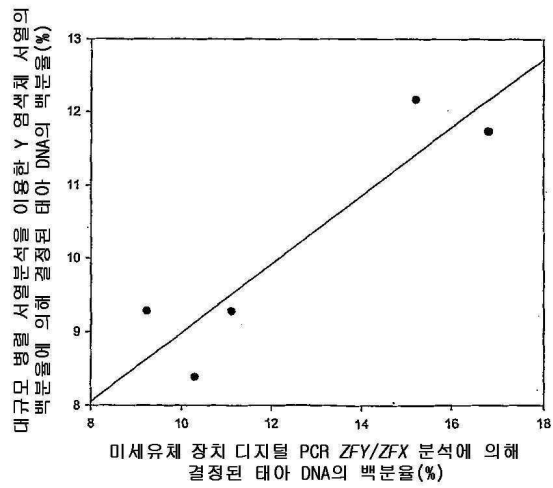
도면2



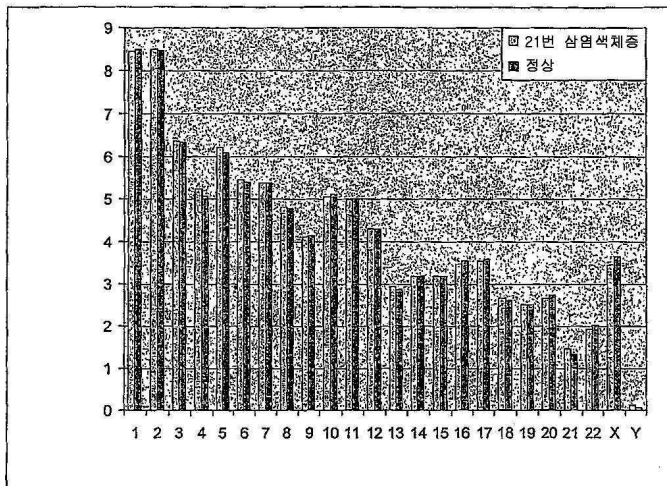
도면3a



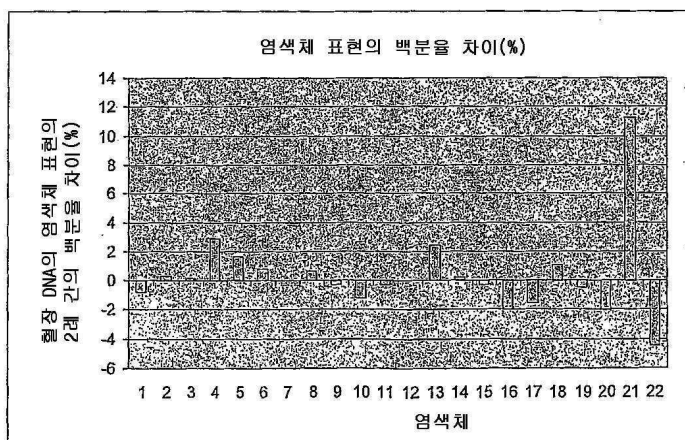
도면3b



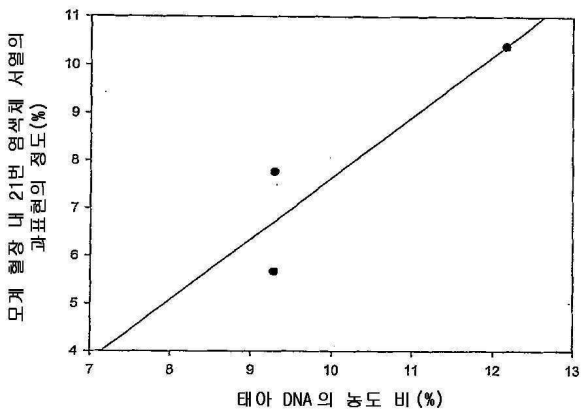
도면4a



도면4b



도면5



도면6

시료	서열분석된 테그의 종 수	서열분석된 염기쌍의 종 수	인간 게놈의 비율 (%)	U0 수	U0 수 / 종 수 (%)	U0 서열분석된 염기쌍의 수	인간 게놈의 U0 비율 (%)
시료 1 T21	1.12E+07	4.02E+08	13.4	1.93E+06	17.3	6.94E+07	2.31
시료 2 T21	1.04E+07	3.73E+08	12.4	1.80E+06	17.4	6.50E+07	2.17
시료 3 T21	8.90E+06	3.20E+08	10.7	2.09E+06	23.4	7.51E+07	2.50
시료 4 T21	1.02E+07	3.69E+08	12.3	2.23E+06	21.7	8.02E+07	2.67
시료 5 정배수체	1.06E+07	3.83E+08	12.8	2.12E+06	19.9	7.62E+07	2.54
시료 6 정배수체	9.58E+06	3.45E+08	11.5	1.91E+06	20.0	6.88E+07	2.29
시료 7 정배수체	9.55E+06	3.44E+08	11.5	2.01E+06	21.0	7.22E+07	2.41
시료 8 정배수체	9.09E+06	3.27E+08	10.9	2.09E+06	23.0	7.53E+07	2.51

도면7

각각의 서브세트 내 서브의 수		정비수치		Tz1 (PS)		Tz1 (100%)		Tz1 (20%)			
평균	SD	5백분위수	95백분위수	평균	5백분위수	95백분위수	대입) 성공적인	평균	5백분위수	95백분위수	대입) 성공적인
69,696	1.322%	0.026%	1.250%	1.355%	1.201%	1.403%	No	1.398%	1.316%	1.461%	No
726,696	1.322%	0.024%	1.277%	1.367%	1.355%	1.403%	No	1.398%	1.344%	1.461%	No
166,696	1.322%	0.019%	1.291%	1.359%	1.324%	1.382%	No	1.398%	1.357%	1.423%	Yes
246,696	1.322%	0.019%	1.289%	1.359%	1.334%	1.372%	No	1.398%	1.351%	1.415%	Yes
326,696	1.322%	0.014%	1.304%	1.355%	1.334%	1.372%	No	1.398%	1.357%	1.403%	Yes
366,696	1.322%	0.010%	1.302%	1.341%	1.333%	1.374%	No	1.398%	1.359%	1.404%	Yes
426,696	1.322%	0.095%	1.303%	1.341%	1.355%	1.374%	No	1.398%	1.359%	1.405%	Yes
446,696	1.322%	0.009%	1.305%	1.355%	1.339%	1.372%	Yes	1.398%	1.372%	1.405%	Yes
546,696	1.322%	0.008%	1.308%	1.359%	1.339%	1.372%	Yes	1.398%	1.372%	1.405%	Yes

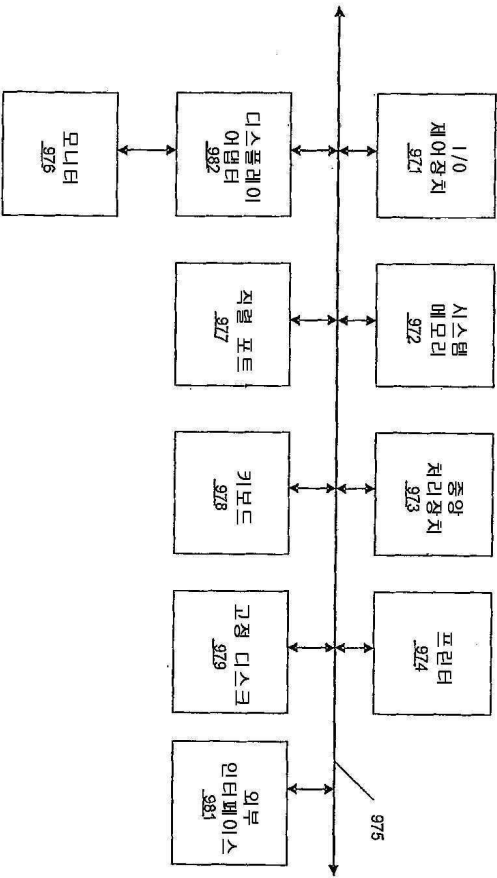
도면8a

서열분석된 타그 순서	시료 1 T21	시료 2 T21	시료 3 T21	시료 4 T21	시료 5 정배수체	시료 6 정배수체	시료 7 정배수체	시료 8 정배수체
1	9796394	9798087	9798123	9795700	9797841	9795972	9796536	9795601
2	9797424	9798402	9798250	9797860	9798176	9796549	9796863	9797404
3	9797438	9798708	9798715	9797864	9798835	9797359	9798161	9798117
4	9798112	9799733	9799467	9798106	9800315	9797418	9798401	9798175
5	9798394	9798852	9799730	9799209	9800385	9797446	9798722	9798387
6	9798729	9800362	9799788	9799440	9800554	9797860	9799752	9798816
7	9798768	9800914	9799834	9799440	9800820	9798062	9800317	9799732
8	9798816	9801528	9800332	9799741	9800829	9798135	9800800	9800144
9	9799421	9801779	9800869	9799812	9800832	9798707	9801380	9800421
10	9799464	9803801	9800881	9799833	9800866	9798715	9801699	9800827

도면8b

서열분석된 단그 순서	시료 1 T21	시료 2 T21	시료 3 T21	시료 4 T21	시료 5 정배수체	시료 6 정배수체	시료 7 정배수체	시료 8 정배수체
1	14469232	14510976	14510792	14510867	14434681	14510711	14512069	14510996
2	14503781	14511307	14510807	14511312	14506660	14510836	14512407	14522947
3	14510806	14511512	14510886	14511983	14511219	14511308	14522879	14530476
4	14510824	14511755	14511816	14522914	14511328	14511354	14609732	14565255
5	14510966	14511774	14512410	14522978	14511377	14511703	14628248	14609733
6	14511369	14512067	14522944	14523028	14511432	14511703	14675246	14628261
7	14511702	14512190	14523047	14566245	14511597	14512184	14675250	14668784
8	14511738	14522891	14564769	14628257	14511755	14512204	14680184	14680163
9	14523000	14539630	14566205	14636374	14512204	14524806	14692535	14689547
10	14524823	14541260	14566251	14665101	14512204	14565230	14694114	14689592

도면9



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 제22항

【변경전】

상기 일부분은

【변경후】

일부분은