



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0014121  
(43) 공개일자 2019년02월11일

- |   |  |
|---|--|
| <p>(51) 국제특허분류(Int. Cl.)<br/><b>C12Q 1/6886</b> (2018.01)</p> <p>(52) CPC특허분류<br/><b>C12Q 1/6886</b> (2018.05)<br/><b>C12Q 2545/113</b> (2013.01)</p> <p>(21) 출원번호 <b>10-2019-7002653(분할)</b></p> <p>(22) 출원일자(국제) <b>2016년01월13일</b><br/>심사청구일자 <b>없음</b></p> <p>(62) 원출원 <b>특허 10-2017-7022238</b><br/>원출원일자(국제) <b>2016년01월13일</b><br/>심사청구일자 <b>2017년08월09일</b></p> <p>(85) 번역문제출일자 <b>2019년01월25일</b></p> <p>(86) 국제출원번호 <b>PCT/CN2016/070785</b></p> <p>(87) 국제공개번호 <b>WO 2016/112850</b><br/>국제공개일자 <b>2016년07월21일</b></p> <p>(30) 우선권주장<br/>62/102,867 2015년01월13일 미국(US)<br/>62/111,534 2015년02월03일 미국(US)</p> | <p>(71) 출원인<br/><b>더 차이나이즈 유니버시티 오브 홍콩</b><br/>중국 홍콩 엔. 티. 새턴 더 차이나이즈 유니버시티 오브 홍콩</p> <p>(72) 발명자<br/><b>로 유크-밍 테니스</b><br/>중국 999077 홍콩 주룽 호만틴 킹 탁 스트리트 7 포쓰 플로어<br/><b>치우 로사 와이 쿤</b><br/>중국 999077 홍콩 뉴 테리토리즈 샤틀린 마 록 패스 52 더블 해븐 하우스 31<br/>(뒷면에 계속)</p> <p>(74) 대리인<br/><b>김진희, 김태홍</b></p> |
|---|--|

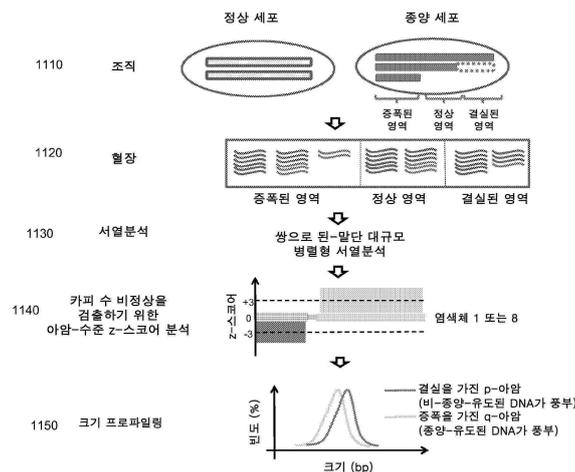
전체 청구항 수 : 총 16 항

(54) 발명의 명칭 **암 검출을 위한 혈장 DNA의 크기 및 수 비정상성의 이용 방법**

(57) 요약

본 발명은 암 검출을 위하여 순환형 무세포 DNA의 크기 및 카피-수 비정상성을 분석하는 방법을 제공한다. 상기 방법은 전체 게놈 방식으로 단일 염기 해상도에서 병렬 서열분석을 포함한다. 혈장 DNA 분자는 염색체 아암-수준 z-스코어 분석을 이용하여 추가로 확인된다.

대표도



(52) CPC특허분류

C12Q 2600/112 (2013.01)

C12Q 2600/156 (2013.01)

(72) 발명자

**찬 관 치**

중국 999077 홍콩 뉴 테리토리즈 샬틴 메이 턴 로  
드 피크 원 블럭 5 피프티스 플루어 플랫 씨

**지양 페이지용**

중국 999077 홍콩 뉴 테리토리즈 샬틴 시우 렉 유  
엔 로드 넘버스 62-66(넘버 62) 쉥 램 코트 블럭  
비 1층 플랫 7

## 명세서

### 청구범위

#### 청구항 1

생물의 생물학적 샘플의 분석 방법에 있어서, 상기 생물학적 샘플이 정상 세포 및 잠재적으로 암과 관련된 세포에서 유래하는 핵산 분자를 포함하되, 상기 생물학적 샘플에서 상기 핵산 분자의 적어도 일부가 무세포이고, 상기 방법이 하기 단계를 포함하는, 방법:

상기 생물의 복수의 염색체 영역을 확인하는 단계로서, 각 염색체 영역이 복수의 유전자자리를 포함하되, 상기 복수의 염색체 영역이 하위염색체인, 상기 복수의 염색체 영역을 확인하는 단계;

상기 생물의 상기 생물학적 샘플에서 상기 핵산 분자의 복수의 핵산 분자의 각각에 대하여:

상기 생물의 표준 게놈에서 상기 핵산 분자의 위치를 확인하는 단계;

상기 복수의 염색체 영역의 각각에 대하여:

확인된 위치에 기반하여 상기 복수의 핵산 분자 내의 핵산 분자의 각자의 그룹을 상기 염색체 영역 기원인 것으로서 확인하는 단계로서, 상기 각자의 그룹이 상기 염색체 영역의 상기 복수의 유전자자리의 각각에 위치한 적어도 하나의 핵산 분자를 포함하는, 상기 핵산 분자의 각자의 그룹을 확인하는 단계;

컴퓨터 시스템으로, 상기 핵산 분자의 각자의 그룹의 각자의 값을 계산하는 단계로서, 상기 각자의 값이 상기 각자의 그룹의 상기 핵산 분자의 특성을 한정하는, 상기 각자의 값을 계산하는 단계; 및

상기 염색체 영역이 결실 또는 증폭을 나타내는지의 분류를 결정하기 위해 상기 각자의 값을 각자의 기준치와 비교하는 단계; 및

결실 또는 증폭을 나타내는 상기 염색체 영역의 시험 패턴을 결정하는 단계로서, 상기 시험 패턴은

결실 또는 증폭을 나타내는 상기 염색체 영역의 세트;

증폭을 나타내는 상기 세트의 제1 서브세트; 및

결실을 나타내는 상기 세트의 제2 서브세트를 포함하는, 상기 시험 패턴을 결정하는 단계;

상기 시험 패턴을 상이한 유형의 암의 복수의 참조 패턴과 비교하는 단계;

상기 비교에 기반하여, 제1 유형의 암에 대응하는 제1 참조 패턴과 동일한 결실 또는 증폭을 나타내는 상기 시험 패턴의 영역의 제1 양을 결정하는 단계; 및

상기 생물학적 샘플이 상기 제1 유형의 암을 나타내는지의 제1 분류를 결정하기 위해 상기 제1 양을 제1 한계값과 비교하는 단계.

#### 청구항 2

제1항에 있어서, 상기 복수의 염색체 영역이 비-중첩인, 방법.

#### 청구항 3

제1항에 있어서, 상기 각자의 그룹 중 적어도 하나의 그룹의 핵산 분자의 특성이 상기 염색체 영역의 하나의 일배체형이고, 상기 각자의 기준치가 상기 염색체 영역의 또 다른 일배체형인, 방법.

#### 청구항 4

제1항에 있어서, 상기 각자의 그룹 중 적어도 하나의 그룹의 핵산 분자의 특성이 상기 염색체 영역에서 위치를 가진 수많은 핵산 분자를 포함하는, 방법.

#### 청구항 5

제1항에 있어서, 상기 각자의 그룹 중 적어도 하나의 그룹의 핵산 분자의 특성이 크기 분포를 포함하는, 방법.

**청구항 6**

제5항에 있어서, 상기 각자의 값이 상기 크기 분포의 평균, 상기 크기 분포의 중앙값, 상기 크기 분포의 형태, 또는 크기 한계값 미만의 크기를 갖는 핵산 분자의 비율을 포함하는, 방법.

**청구항 7**

제1항에 있어서, 상기 복수의 참조 패턴이 무세포 핵산 분자의 조직 및/또는 혼합물의 참조 샘플로부터 결정되는, 방법.

**청구항 8**

제1항에 있어서, 상기 제1 유형의 암이 간세포 암종, 대장암, 유방암, 폐암, 또는 비인두암종인, 방법.

**청구항 9**

제1항에 있어서, 상기 제1 참조 패턴이, 각각 증폭, 결실, 또는 무 비정상의 한정된 상태를 갖는 영역의 제1 수를 포함하고, 상기 영역의 제1 양은 상기 시험 패턴이 매칭되는 영역의 상기 제1 수의 백분율인, 방법.

**청구항 10**

제1항에 있어서,

하기 단계들에 의해 상기 제1 한계값을 결정하는 단계를 더 포함하는, 방법:

참조 패턴의 세트를 클러스터링하는 단계로서, 각 클러스터가 상이한 유형의 암에 대응하는, 상기 클러스터링하는 단계;

상기 제1 유형의 암에 대응하는 제1 클러스터용 센트로이드를 한정하는 단계로서, 상기 센트로이드가 상기 제1 클러스터의 참조 패턴의 적어도 예정된 수만큼 공유되는 상기 제1 클러스터의 상기 참조 패턴의 영역에 대응하는, 상기 한정하는 단계; 및

상기 제1 클러스터 및 상기 센트로이드의 경계로부터의 거리를 상기 제1 한계값으로서 한정하는 단계.

**청구항 11**

제1항에 있어서,

상기 복수의 참조 패턴의 각 참조 패턴에 대하여:

상기 참조 패턴과 동일한 결실 또는 증폭을 나타내는 상기 시험 패턴의 영역의 각자의 양을 결정하는 단계; 및

상기 각자의 양을 서로 비교함으로써 각 유형의 암에 대하여 상대적 우도(relative likelihood)를 결정하는 단계; 및

상기 상대적 우도에 기반하여 암의 매칭 유형을 확인하는 단계를 더 포함하는, 방법.

**청구항 12**

제1항에 있어서, 상기 생물학적 샘플이 혈장 또는 혈청인, 방법.

**청구항 13**

제1항에 있어서, 상기 핵산 분자의 위치를 확인하는 단계가 서열 판독을 얻기 위해 상기 핵산 분자를 서열분석하고, 참조 게놈에 상기 서열 판독을 정렬시키는 것을 포함하는, 방법.

**청구항 14**

제1항 내지 제13항 중 어느 한 항의 방법의 작동을 수행하도록 컴퓨터 시스템을 제어하기 위한 복수의 명령어를 저장하는 컴퓨터 판독 가능한 기록매체 및, 상기 컴퓨터 판독 가능한 기록매체 상에 저장된 명령어를 실행하기 위한 하나 이상의 프로세서를 포함하는, 컴퓨터 시스템.

**청구항 15**

하기를 포함하는, 시스템:

제14항의 컴퓨터 시스템; 및

생물학적 샘플 중 복수의 핵산 분자를 서열분석 하도록 구성된 서열분석 시스템.

**청구항 16**

실행시 제1항 내지 제13항 중 어느 한 항의 방법을 수행하도록 컴퓨팅 시스템을 제어하는 복수의 명령어를 저장하는, 컴퓨터 판독 가능하고 전송 불가능한 기록매체.

**발명의 설명**

**기술 분야**

[0001] 관련 출원에 대한 교차 참조

[0002] 본 출원은, 2015년 1월 13일 출원된, Lo 등 (변리사 정리 번호 80015-015800US)에 의해 발명의 명칭 "Using Size And Number Aberrations In Plasma DNA For Detecting Cancer"인 미국 특허 출원 번호 62/102,867; 및 2015년 2월 3일 출원된, Lo 등 (변리사 정리 번호 80015-015801US)에 의해 발명의 명칭 "Using Size and Number Aberrations in Plasma DNA for Detecting Cancer"인 62/111,534에 우선권을 주장하고, 이들의 개시내용은 모든 목적으로 그 전체가 참고로 편입된다.

**배경 기술**

[0003] 순환형 무세포 DNA의 분석은 암 (1-3)의 검출 및 모니터링에 점점 더 사용되어 왔다. 카피 수 비정상 (4-7), 메틸화 변화 (8-11), 단일 뉴클레오타이드 돌연변이 (4, 12-15), 암-유도된 바이러스 서열 (16, 17) 및 염색체 재배열 (18, 19)을 포함한, 상이한 암-관련 분자 특징은 다양한 유형의 암을 가진 환자의 혈장에서 검출될 수 있다. 임상 적용의 급속 팽창에도 불구하고, 암 환자내 순환형 DNA의 많은 근본적인 분자 특징은 불확실하게 남아 있고, 이로써 상기 분석의 가장 유효한 임상 용도를 제한시킨다.

[0004] 특히, 암 환자내 순환형 DNA의 크기에 관한 이전의 연구는 불일치된 결과를 제공하였다. 연구는 순환형 DNA의 전체 완전성 (크기의 측정)이 악성 병태 (20-23) 없이 대상체와 비교된 경우 암 환자에서 증가한다는 것을 입증하였다. 상이한 앰플리콘 크기를 가진 PCR을 이용하여, 더 긴 DNA의 비율이 암 환자에서 더 높다는 것이 보여졌다. DNA 완전성에서 이러한 비정상은 치료 이후 되돌려지는 것으로 보여졌고 그러한 변화의 지속성은 좋지 못한 예측 (20, 24)과 관련되었다. 다른 한편으로, 종양 조직으로부터 유도된 순환형 DNA가 비-악성 세포로부터 유도된 것보다 더 짧을 수 있다는 의견상으로 모순되는 증거가 또한 있다. 예를 들어, 더 짧은 앰플리콘 (12, 25)을 가진 PCR을 이용하여 그 돌연변이가 검출된 경우 암-관련 돌연변이를 담지하는 DNA 분자의 비율이 더 높다는 것이 보여져 왔다.

[0005] 추가로, 종양-유도된 혈장 DNA가 혈장에서 비-종양-유도된 배경 DNA와 쉽게 식별될 수 없기 때문에 HCC 환자의 혈장에서 종양-유도된 DNA의 크기 프로파일 연구는 도전적 시도이다. 암-특이적 돌연변이의 검출은 비-종양성 혈장 DNA와 종양성을 식별하기 위한 유전자형 수단을 제공한다. 그러나, 게놈 (29-32)에 대해 암-특이적 돌연변이가 상대적으로 소수이다. 따라서, 특히 종양-유도된 DNA의 크기 분포의 넓은, 상세한 및 더욱 비용-효율적인 관점 생성의 목적으로, 혈장에 종양-유도된 DNA를 정확하게 확인하는 것이 어려울 수 있다.

[0006] 상기 어려움은 종양성 및 비-종양성 DNA의 혼합물을 가능하게는 함유한 샘플에서 정확한 측정의 수득에서 장애물을 제공한다.

**발명의 내용**

**해결하려는 과제**

**과제의 해결 수단**

[0007] 간단한 요약

[0008] 구현예는, 영역이 암과 관련될 수 있는, 비정상 (예를 들면, 증폭 또는 결실)을 나타내는지의 결정용 시스템 및 방법을 제공할 수 있다. 예를 들어, 구현예는 카운트-기반 분석을 이용한 비정상을 가능하게는 갖는 것으로서 영역을 확인할 수 있고 영역이 크기-기반 분석을 이용한 비정상을 갖는지를 확증할 수 있다.

[0009] 다른 구현예에서, 비정상을 나타내는 영역은 암의 공지된 유형에 대응하는 참조 패턴에 비교될 수 있다. 암의 유형은 영역의 충분한 수가 매칭 비정상을 갖는 것으로서 확인될 수 있다. 상기 매칭 영역은 종양 DNA의 분석을 위하여, 예를 들면, 크기 분석을 위하여 암에 관련된 것으로서 추가로 확인될 수 있다.

[0010] 더욱 다른 구현예에서, 샘플 (예를 들면, 종양 및 비-종양 DNA 둘 모두를 가능하게는 함유한 혼합물)에서 DNA 단편의 크기 분석은 샘플내 종양 DNA의 측정된 분획에 의존할 수 있다. 예를 들어, 건강한 대조군보다 더 긴 DNA 단편은 낮은 종양 DNA 분획에 대하여 초기 암을 나타낼 수 있고, 건강한 대조군보다 더 짧은 DNA 단편은 더 높은 종양 DNA 분획에 대하여 후기 암을 나타낼 수 있다.

[0011] 다른 구현예는 본원에서 기재된 방법과 관련된 시스템 및 컴퓨터 판독가능한 기록매체에 관한 것이다.

[0012] 본 발명의 성질 및 구현예의 이점에 대한 더 나은 이해는 하기 상세한 설명 및 수반되는 도면들을 참조하여 이득될 수 있다.

### 도면의 간단한 설명

[0013] 도 1은 본 발명의 구현예에 따라 비정상을 나타낸 것으로서 염색체 영역 확인의 방법 100을 예시하는 흐름도이다.

도 2는 본 발명의 구현예에 따라 대표적인 간세포 암종 (HCC) 환자의 혈장 및 조직 샘플에서 증폭 및 결실을 나타내는 영역 확인의 시르코스(Circos) 플롯 200을 보여준다.

도 3은 본 발명의 구현예에 따라 다양한 대상체에 대하여 혈장 카피 수 비정상 (CNA)을 보여준다.

도 4는 본 발명의 구현예에 따라 HCC 환자, B형 간염 바이러스 (HBV) 캐리어, 간경변증을 가진 환자 및 건강한 대상체의 혈장에서 CNA의 검출가능성을 보여주는 표 400이다.

도 5는 12 HCC 환자의 종양 및 대응하는 혈장에서 검출된 CNAs의 표 500을 보여준다.

도 6은 생물학적 샘플이 본 발명의 구현예에 따라 암의 제1 유형을 나타내는지를 결정하기 위해 생물의 생물학적 샘플의 분석 방법을 예시하는 흐름도를 보여준다.

도 7은 구현예에 따라 표 700에서 암의 상이한 유형에 대하여 상이한 패턴을 나타내는 염색체 아암을 보여준다.

도 8A, 8B, 및 8C는 암의 상이한 유형에 대하여 염색체 영역의 패턴의 표 800을 보여준다.

도 9는 본 발명의 구현예에 따라 생물의 생물학적 샘플의 분석 방법을 예시하는 흐름도를 보여준다.

도 10은 혈장내 종양 DNA 분획에 대하여 (A) 150 bp보다 더 짧은, (B) 150 내지 180 bp, 및 (C) 180 bp보다 더 긴 혈장 DNA 단편의 비율의 플롯을 보여준다.

도 11은 암 환자에서 혈장 DNA 크기 분석의 원리의 도식적 예시이다.

도 12는 대표적인 케이스 H291의 증폭된 8q 및 결실된 8p에서 유래하는 혈장 DNA의 크기 분포를 보여준다. (A) 8p (적색) 및 8q (녹색)에 대한 혈장 DNA의 크기 분포. (B) 8p (적색) 및 8q (녹색)에 대한 혈장 DNA 크기용 누적 빈도의 플롯. (C) HCC 케이스 H291에 대한 누적 빈도의 차이.

도 13은 8q와 8p ( $\Delta S$ ) 사이 크기에 대하여 누적 빈도의 차이를 보여준다. (A) 혈장내 8p 및 8q상에서 상이한 CNAs를 가진 전체 HCC 사례에 대하여 크기에 대한  $\Delta S$ 의 플롯. (B) 상이한 그룹 중에서  $\Delta S_{166}$ 의 값.

도 14는 대표적인 HCC 환자에 대하여 크기에 대하여 1q와 1p 사이  $\Delta S$ 의 값의 플롯이다.

도 15는 건강한 대조군, HBV 캐리어, 간경변 환자 및 HCC 환자에 대하여 1q와 1p 사이  $\Delta S_{166}$ 의 값의 플롯이다.

도 16은 본 발명의 구현예에 따라 생물의 생물학적 샘플을 분석하기 위해 염색체 아암-수준 z-스코어 분석 (CAZA) 및 크기 분석 수행의 방법을 예시하는 흐름도이다.

도 17은 본 발명의 구현예에 따라 생물의 생물학적 샘플 분석의 방법을 예시하는 흐름도이다.

도 18은 혈장내 종양-유도된 DNA의 상이한 단편적 농도를 가진 HCC 환자에서 혈장 DNA 단편의 크기 분포를 보여준다.

도 19는 (A) 건강한 대조군, (B) 만성 HBV 캐리어, 및 (C) 간경변 환자에 대하여 혈장 DNA의 크기 프로파일을 보여준다.

도 20은 건강한 대조군 대상체, 혈장내 2% 미만의 종양 DNA 분획을 가진 HCC 환자, 및 6% 초과 종양 DNA 분획을 가진 HCC 환자에 대하여 짧은 단편의 비율의 박스플롯을 보여준다.

도 21은 건강한 대조군 대상체와 2% 미만 종양 DNA 분획을 가진 HCC 환자를 구분하기 위한  $P(<150)$  적용에 대하여 수신자 조작 특성 (ROC) 곡선이다.

도 22는 건강한 대상체와 6% 초과 종양 DNA 분획을 가진 HCC 환자를 구분하기 위한  $P(<150)$  적용에 대하여 수신자 조작 특성 (ROC) 곡선이다.

도 23은 건강한 대조군 대상체 및 혈장내 2% 미만의 종양 DNA 분획을 가진 HCC 환자에 대하여 긴 단편의 비율의 박스플롯을 보여준다.

도 24는 건강한 대조군 대상체와 2% 미만 종양 DNA 분획을 가진 HCC 환자를 구분하기 위한  $P(>180)$  사용에 대하여 ROC 곡선이다.

도 25는 건강한 대조군 대상체, 2% 미만 종양 DNA 분획을 가진 HCC 환자, 및 6% 초과 종양 DNA 분획을 가진 HCC 환자의 종양 단편 크기의 박스플롯을 보여준다.

도 26은 2% 미만 종양 DNA 분획을 가진 HCC 환자와 건강한 대조군 대상체 사이를 구분하기 위한 종양 단편 크기 사용에 대하여 ROC 곡선이다.

도 27은 6% 초과 종양 DNA 분획을 가진 HCC 환자와 건강한 대조군 대상체를 구분하기 위한 종양 단편 크기 사용에 대하여 ROC 곡선이다.

도 28은 6% 초과 종양 DNA 분획을 가진 HCC 환자에 대하여 및 건강한 대조군 대상체에 대하여 염색체 1q에 정렬된 150 bp 미만의 짧은 혈장 DNA 단편의 비율의 박스플롯을 보여준다.

도 29는 6% 초과 종양 DNA 분획을 가진 HCC 환자와 건강한 대조군 대상체 사이를 구분하기 위한 150 bp 미만의 짧은 혈장 DNA 단편의 비율의 이용에 대하여 ROC 곡선이다.

도 30은 HCC 환자의  $\Delta S$  대 종양 크기의 플롯이다.

도 31은 종양 크기에 대한 특정 크기의 DNA 단편의 백분율의 플롯이다.

도 32는 본 발명의 구현예에 따라 시스템 및 방법으로 사용가능한 예 컴퓨터 시스템 10의 블록 선도를 보여준다.

### 발명을 실시하기 위한 구체적인 내용

[0014]

용어

[0015]

용어 "생물학적 샘플"은 본원에서 사용된 바와 같이 대상체 (예를 들면, 인간, 예컨대 임신한 여성)로부터 취해지는 그리고 당해 하나 이상의 핵산 분자(들)을 함유하는 임의의 샘플을 지칭한다. 그 예는 혈장, 타액, 늑막 유체, 땀, 복수액, 담즙, 소변, 혈청, 체액, 대변, 자궁경부 세척 유체, 및 자궁경부 도말표본 샘플을 포함한다.

[0016]

용어 "핵산" 또는 "폴리뉴클레오타이드"는 단일- 또는 이중-가닥 형태의 데옥시리보핵산 (DNA) 또는 리보핵산 (RNA) 및 이들의 폴리머를 지칭한다. 구체적으로 제한되지 않는 한, 상기 용어는 참조 핵산과 유사한 결합 특성을 갖는 천연 뉴클레오타이드의 공지된 유사체를 함유한 핵산을 포함하고 천연 발생 뉴클레오타이드와 유사한 방식으로 대사작용된다. 달리 명시되지 않는 한, 특정한 핵산 서열은 또한 전적으로 이들의 보존적으로 변형된 변이체 (예를 들면, 퇴화 코돈 치환), 대립유전자, 오쏘로그, 단일 뉴클레오타이드 다형성 (SNPs), 및 상보성 서열 뿐만 아니라 명백하게 제시된 서열을 포함한다. 구체적으로, 퇴화 코돈 치환은 하나 이상의 선택된 (또는 모든) 코돈의 제3 위치가 혼합된-염기 및/또는 데옥시이노신 잔기로 치환되는 서열의 생성에 의해 달성될 수 있다 (Batzer MA 등, *Nucleic Acids Res* 1991; **19**:5081; Ohtsuka E 등, *J Biol Chem* 1985; **260**:2605-2608; and Rossolini GM 등, *Mol Cell Probes* 1994; **8**:91-98). 용어 핵산은 유전자, cDNA, mRNA, 소형 넌코딩 RNA, 마이

크로RNA (miRNA), Piwi-상호작용 RNA, 및 유전자 또는 유전자자리에 의해 인코딩된 짧은 헤어핀 RNA (shRNA)와 상호교환적으로 사용된다.

- [0017] 용어 "유전자"는 폴리펩타이드 사슬 생산에 관여된 DNA의 세그먼트를 의미한다. 코딩 영역을 선행 및 후행하는 영역 (리더 및 트레일러) 뿐만 아니라 개별적인 코딩 세그먼트 (엑손) 사이 개입 서열 (인트론)을 포함할 수 있다.
- [0018] 본원에서 사용된 바와 같이, 용어 "유전자자리" 또는 그의 복수 형태 "유전자자리"는 게놈을 거쳐 변동을 갖는 뉴클레오타이드 (또는 염기쌍)의 임의의 길이의 위치 또는 어드레스이다.
- [0019] (또한 서열 관독으로 불리는) 용어 "서열분석된 태그"는 핵산 분자의 전체 또는 일부, 예를 들면, DNA 단편으로부터 수득된 서열을 지칭한다. 하나의 구현예에서, 단편의 단 하나의 말단은 예를 들면, 약 30 bp 서열분석된다. 서열분석된 태그는 그 다음 표준 게놈으로 정렬될 수 있다. 대안적으로, 단편의 양쪽 말단은, 정렬에서 더 큰 정확도를 제공할 수 있고 또한 단편의 길이를 제공할 수 있는, 2개의 서열분석된 태그를 생성하기 위해 서열분석될 수 있다. 더욱 또 다른 구현예에서, 선형 DNA 단편은, 예를 들면, 결찰에 의해 원형화될 수 있고, 결찰 부위를 미치는 부분은 서열분석될 수 있다.
- [0020] 용어 단편적 종양 DNA 농도는 용어 종양 DNA 비율 및 종양 DNA 분획과 상호교환적으로 사용되고, 종양으로부터 유도되는 샘플에서 존재하는 DNA 분자의 비율을 지칭한다.
- [0021] 용어 "크기 프로파일"은 일반적으로 생물학적 샘플에서 DNA 단편의 크기에 관련한다. 크기 프로파일은 다양한 크기에서 DNA 단편의 양의 분포를 제공하는 히스토그램일 수 있다. (또한 크기 파라미터 또는 단지 파라미터로서 지칭된) 다양한 통계적 파라미터는 하나의 크기 프로파일을 또 다른 것과 식별하기 위해 사용될 수 있다. 하나의 파라미터는 전체 DNA 단편에 비하여 또는 또 다른 크기 또는 범위의 DNA 단편에 비하여 특정한 크기 또는 크기의 범위의 DNA 단편의 백분율이다.
- [0022] 용어 "파라미터"는 본원에서 사용된 바와 같이 정량적 데이터 세트 및/또는 정량적 데이터 세트 사이에서 수치 관계를 특징으로 하는 수치를 의미한다. 예를 들어, 제1 핵산 서열의 제1 양과 제2 핵산 서열의 제2 양 사이의 비 (또는 비의 함수)는 파라미터이다.
- [0023] 용어 "분류"는 본원에서 사용된 바와 같이 샘플의 특정한 특성과 관련되는 (단어를 포함하는) 임의의 수(들) 또는 다른 특징(들)을 지칭한다. 예를 들어, "+" 기호는 샘플이 결실 또는 증폭 (예를 들면, 중복)을 갖는 것으로서 분류되는 것을 의미할 수 있다. 용어 "컷오프" 및 "한계값"은 작업에서 사용된 예정된 수를 지칭한다. 예를 들어, 컷오프 크기는 단편이 제외되는 초과 크기를 지칭할 수 있다. 역치는 특정한 분류가 적용하는 값 초과 또는 미만일 수 있다. 이들 용어들의 어느 한쪽은 이들 문맥의 어느 한쪽에서 사용될 수 있다.
- [0024] 용어 "암의 수준"은 암의 존재 여부, 암의 병기, 종양의 크기, 염색체 영역의 많은 결실 또는 증폭이 (예를 들면 이중으로 또는 삼중으로) 관여되는 방식, 및/또는 암의 중증도의 다른 측정을 지칭할 수 있다. 암의 수준은 수 또는 다른 특징일 수 있다. 수준은 0일 수 있다. 암의 수준은 또한 결실 또는 증폭과 관련된 암이전 또는 전 암성 상태를 포함한다.
- [0025] "하위염색체 영역"은 염색체보다 더 작은 영역이다. 하위염색체 영역의 예는 100 kb, 200 kb, 500 kb, 1 Mb, 2 Mb, 5 Mb, 또는 10 Mb이다. 하위염색체 영역의 또 다른 예는 하나 이상의 밴드, 또는 하위밴드, 또는 염색체의 아암 중 하나에 대응하는 것이다. 밴드 또는 하위밴드는 세포유전적 분석에서 관측된 특징이다. 하위염색체 영역은 참조 인간 게놈 서열과 관련하여 그의 게놈 배위에 의해 지칭될 수 있다.
- [0026] 상세한 설명
- [0027] 암은 종종 사람의 정상 게놈에 비하여 카피 수 비정상 (증폭 또는 결실)을 가진 영역을 갖는다. 기술은 종양 DNA 단편 및 비-종양 DNA 단편을 포함하는 샘플 (예를 들면, 혈장 또는 혈청)에서 무세포 DNA 단편을 계수할 수 있다. 계수는 과대-표현되는 (증폭을 나타내는) 또는 과소-표현되는 (결실을 나타내는) 영역을 확인할 수 있다. 그러나, 상기 카운트-기반 기술이 통계적 성격임에 따라, 부정확한 징후가 발생할 수 있다. 구현예는 카운트-기반 분석을 이용하여 카피 수 비정상 (또한 비정상으로서 지칭됨)을 가능하게는 갖는 것으로서 영역을 확인할 수 있고 영역이 크기-기반 분석을 이용하여 비정상을 갖는지를 입증할 수 있다. 그와 같은 확증은 비정상을 가진 영역 확인에서 추가의 정확도를 제공한다.
- [0028] 비정상을 갖는 영역은 샘플이 수득된 생물에서 암의 존재를 확인하기 위해 사용될 수 있다. 그러나, 암의 존재는 암의 유형을 나타내지 않는다. 이러한 문제를 다루기 위해, 구현예는 공지된 암을 가진 샘플로부터 영역내

비정상적 참조 패턴을 이용할 수 있다. 영역이 비정상인 시험 패턴은 시험된 주어진 샘플에 대하여 결정될 수 있고, 시험 패턴은 암의 유형을 결정하기 위해 참조 패턴에 비교될 수 있다. 암의 특정한 유형에 대응하는 참조 패턴으로서 동일한 결실 또는 증폭을 나타내는 시험 패턴의 영역의 양은 결정될 수 있고, 양은 암의 특정한 유형이 존재하는지의 분류를 결정하기 위해 한계값과 비교될 수 있다. 일단 영역이 비정상을 갖는 것 및 암의 특정한 유형에 대응하는 것 모두로서 확인되면, 종양 DNA에 대한 영역 분석에서 더 큰 신뢰를 가질 수 있다. 예를 들어, 영역은 샘플에서 종양 DNA 분획을 측정하기 위해 사용될 수 있다.

[0029] 추가로, 다양한 연구는 무세포 종양 DNA 단편의 길이: 종양 DNA에 대하여 더 긴 단편을 보여주는 일부 및 종양 DNA에 대하여 더 짧은 단편을 보여주는 다른 것에 대한 불일치된 결과를 보여주었다. 아래 분석은 둘 모두가 상이한 종양 DNA 분획을 제외하고 정확할 수 있다는 것을 보여준다. 구현에는, 비정상을 갖는 것으로서 확인된 영역에서 DNA 단편의 계수를 이용하여 결정될 수 있는, 측정된 종양 DNA 분획에 기반된 크기-분석에서 상이한 크기 한계값을 이용할 수 있다. 따라서, 일부 실험은, 예를 들어 하기를 통해 이들 명백한 불일치를 중재할 수 있다: (a) 대규모 병렬형 서열분석에 의해 가능해진 혈장 DNA의 게놈-전체 고 분해능 크기 프로파일링; 및 (b) (예를 들면, 비정상을 갖는 것으로서 확인된 영역을 이용한) 암 환자의 혈장에서 비-종양성 배경 DNA와 종양-유도된 DNA를 식별하기 위한 효율적인 접근법.

[0030] **I. 도입**

[0031] 대규모 병렬형 서열분석 (26, 27)의 용도를 가진 샘플에서 모든 개별 혈장 DNA 분자의 백만 또는 십억의 길이를 측정하는 것이 실행가능해졌다. 그러므로, 혈장 DNA 크기는 게놈전체 방식으로 및 단일-염기 분해능에서 연구될 수 있다. 상기 접근법을 이용하여, 순환형 DNA의 크기는 혈장 DNA가 세포자멸사 (26, 27)를 통해 생성될 수 있음을 시사하는 모노뉴클레오타이드 DNA의 크기를 닮는다는 것으로 일반적으로 보여지고 있다. 임신한 여성에서, 태아로부터 유도된 혈장 DNA는 모친 (26)으로부터 유도된 DNA의 것보다 더 짧다는 것으로 보여지고 있다. 순환형 태아와 모계 DNA 사이의 크기 차이는 모계 혈장에서 태아 DNA의 정량화 및 혈장 DNA (28)의 크기 분석을 통한 염색체 이수성 검출에 대하여 신규한 개념적 근거를 제공하였다. 또한, 이식된 기관과 환자의 자체 조직으로부터 유도된 순환형 DNA의 크기 분포에서의 차이는 고품 장기 또는 골수 이식 (27)의 수령체에 대하여 관측되었다.

[0032] 암 환자의 혈장은 종양-유도된 DNA 및 비-종양-유도된 DNA의 혼합물을 함유한다. 아래 예는 간세포 암종 (HCC) 을 가진 암 환자에서 혈장 DNA의 크기 분포를 분석한다. HCC 환자, 만성 B형 간염 바이러스 (HBV) 감염을 가진 환자, 간경변증을 가진 환자 및 건강한 대상체에서 혈장 DNA의 크기 분포는 또한 분석되었다. 구현에는 HCC 환자의 혈장에서 종양-유도된 DNA의 크기 프로파일을 분석하기 위해 특정 비정상 영역을 사용할 수 있다. 상기 비정상 영역의 사용은 종양-유도된 혈장 DNA가 혈장에서 비-종양-유도된 배경 DNA와 쉽게 식별되지 않는다는 난제를 극복할 수 있다.

[0033] 일부 구현에는 종양- 및 비-종양-유도된 혈장 DNA의 크기 분포에서 차이를 추론하기 위해 카피 수 비정상 (CNAs)에 의해 영향받는 염색체 아암을 이용한다. 종양 조직에서 증폭되는 염색체 아암에 대하여, 종양-유도된 DNA부터 혈장 DNA까지 비례하는 기여는 증가할 것이고 반면에 종양에서 결실되는 염색체 아암에 대하여, 기여는 감소할 것이다. 따라서, 증폭되고 결실되는 염색체 아암의 크기 프로파일의 비교는 혈장에서 종양-유도된 및 비-종양-유도된 DNA 사이의 크기 차이를 반영할 것이다. 전체의 염색체 아암 또는 염색체 아암의 큰 트렁크를 포함한 CNAs는 상대적으로 흔하다 (33). 염색체 1q 및 8q의 결실 및 염색체 1p 및 8p의 증폭은 HCC 조직 (34-36)에서 통상적으로 관측된다. 따라서, 분석은 혈장 DNA의 CNA 및 크기 프로파일링 분석에 대하여 염색체 1 및 8에 집중한다.

[0034] **II. 비정상 영역을 확인하기 위한 계수 분석**

[0035] 비정상 영역은 증폭 또는 결실을 포함한다. 증폭은 영역내 서열이 참조 서열에서 하는 것보다 더욱 종종 발생하고, 따라서 서열이 증폭되는 것을 의미한다. 증폭은 전형적으로 단 하나의 염색체 카피 (일배체형)에서 발생할 것이다. 결실은 영역내 서열이 참조 서열에 비하여 결실되는 것을 의미하고, 전형적으로 단 하나의 염색체 카피는 2배체 생물에 대하여 결실을 갖는다. 영역은 (서로 분리되는) 적어도 2개 유전자자리에 의해 한정될 수 있고, 이들 유전자자리에서 DNA 단편은 영역에 대하여 집단적인 가치를 획득하기 위해 사용될 수 있다.

[0036] **A. 계수에 의한 비정상 영역의 검출**

[0037] 영역의 비정상은 영역으로부터 유도되는 DNA 단편 (분자)의 양의 계수에 의해 결정될 수 있다. 예로서, 양은 수많은 DNA 단편, DNA 단편이 중첩되는 수많은 염기, 또는 영역내 DNA 단편의 다른 측정일 수 있다. 영역에 대한

DNA 단편의 양은 서열 판독을 수득하기 위한 DNA 단편의 서열분석 및 표준 계놈에 서열 판독의 정렬에 의해 결정될 수 있다. 하나의 구현예에서, 영역에 대하여 서열 판독의 양은 과대표현(증폭) 또는 과소표현(결실)을 결정하기 위해 또 다른 영역에 대하여 서열 판독의 양과 비교될 수 있다. 또 다른 구현예에서, 서열 판독의 양은 하나의 일배체형에 대하여 결정될 수 있고 또 다른 일배체형에 대하여 서열 판독의 양과 비교될 수 있다.

[0038] 따라서, (예를 들면, 그 영역에 정렬된 서열분석된 태그의 계수에 의해 결정된 바와 같이) 하나의 염색체 영역으로부터 DNA 단편의 수는 (참조 염색체 영역으로부터, 또 다른 일배체형에서의 영역으로부터, 또는 건강한 것으로 공지되는 또 다른 샘플내 동일한 영역으로부터 결정될 수 있는) 기준치와 비교될 수 있다. 비교는 양이 통계적으로 상이한 (예를 들면, 초과 또는 미만) 기준치인지를 결정할 수 있다. 집단에서 보이는 값의 분포에서 나타낸 바와 같이, 예를 들면, 3 표준 편차 (SD)에 대응하는, 차이에 대한 한계값은 사용될 수 있다.

[0039] 비교의 일부로서, 태그 카운트는 비교 이전 정규화될 수 있다. 특정한 영역에 대하여 서열 판독 (태그)에 대한 정규화된 값은 전체의 계놈에 대해 정렬가능한 서열분석된 판독의 총 수로 그 영역에 정렬하는 서열분석된 판독의 수를 분할함으로써 계산될 수 있다. 상기 정규화된 태그 카운트는 하나의 샘플로부터의 결과를 또 다른 샘플의 결과와 비교되게 한다. 예를 들어, 정규화된 값은 특정한 영역 기원인 것으로 예상된 서열 판독의 비율 (예를 들면, 백분율 또는 분획)일 수 있다. 그러나, 많은 다른 정규화는, 당해 분야의 숙련가에 명백하듯이, 가능하다. 예를 들어, 참조 영역 (상기 경우에서, 참조 영역은 단지 전체의 계놈이다)에 대하여 카운트의 수로 하나의 영역에 대하여 카운트의 수를 분할함으로써 또는 서열 판독의 동일한 수를 항상 이용함으로써 정규화할 수 있다. 상기 정규화된 태그 카운트는 그 다음, 암을 나타내지 않는 하나 이상의 참조 샘플로부터 결정될 수 있는, 역치에 대해 비교될 수 있다.

[0040] 일부 구현예에서, 역치는 기준치일 수 있다. 다른 구현예에서, 기준치는 정규화에 대하여 사용된 다른 값일 수 있고, 비교는 기준치 및 역치를 포함할 수 있다. 예를 들어, 영역의 양은, 통계적으로 유의미한 상이가 존재하는지를 보기 위해 역치와 비교되는, 파라미터를 수득하기 위해 기준치로 분할될 수 있다. 또 다른 예로서, 영역에 대한 양은 기준치 플러스 역치와 비교될 수 있다.

[0041] 하나의 구현예에서, 비교는 특정한 염색체 영역에 대하여 케이스의  $z$ -스코어 계산에 의해 실시된다.  $z$ -스코어는 하기 방정식을 이용하여 계산될 수 있다:  $z$ -스코어 = (케이스의 정규화된 태그 카운트 - 평균) / SD (식중, "평균"은 참조 샘플에 대하여 특정한 염색체 영역에 정렬한 평균 정규화된 태그 카운트이고; SD는 참조 샘플에 대하여 특정한 영역에 정렬한 정규화된 태그 카운트의 수의 표준 편차이다). 그러므로,  $z$ -스코어는 시험된 케이스에 대하여 염색체 영역의 정규화된 태그 카운트가 하나 이상의 기준 대상체의 동일한 염색체 영역에 대하여 평균 정규화된 태그 카운트에서 떨어지는 표준 편차의 수에 대응할 수 있다. 상기  $z$ -스코어는 한계값, 예를 들면, 증폭에 대하여 3 및 결실에 대하여 -3과 비교될 수 있다. 증폭되는 염색체 영역은 한계값 초과  $z$ -스코어의 양성 값을 가질 수 있다. 결실되는 염색체 영역은 한계값 미만인  $z$ -스코어의 음성 값을 가질 수 있다.

[0042]  $z$ -스코어의 규모는 몇 개의 인자에 의해 결정될 수 있다. 하나의 인자는 생물학적 샘플 (예를 들면 혈장)내 종양-유도된 DNA의 단편적 농도이다. 샘플 (예를 들면 혈장)내 종양-유도된 DNA의 단편적 농도가 높을수록, 시험된 케이스의 정규화된 태그 카운트와 참조 케이스 사이의 차이는 더 클 것이다. 그러므로,  $z$ -스코어의 더 큰 규모가 발생할 것이다.

[0043] 또 다른 인자는 하나 이상의 참조 케이스에서 정규화된 태그 카운트의 변동이다. 시험된 케이스의 생물학적 샘플 (예를 들면 혈장)에서 염색체 영역의 과대-표현의 동일한 정도로, 참조 그룹에서 정규화된 태그 카운트의 더 작은 변동 (즉 더 작은 표준 편차)는 더 높은  $z$ -스코어를 발생시킬 것이다. 유사하게, 시험된 케이스의 생물학적 샘플 (예를 들면 혈장)에서 염색체 영역의 과소-표현의 동일한 정도로, 참조 그룹에서 정규화된 태그 카운트의 더 작은 표준 편차는 더 많은 음성  $z$ -스코어를 발생시킬 것이다.

[0044] 또 다른 인자는 종양 조직에서 염색체 비정상의 규모이다. 염색체 비정상의 규모는 특정한 염색체 영역에 대하여 카피 수 변화 (이득 또는 손실)를 지칭한다. 종양 조직에서 카피 수 변화가 더 클수록, 혈장 DNA에서 특정한 염색체 영역의 과대-표현 또는 과소-표현의 정도는 더 높을 것이다. 예를 들어, 염색체의 모든 카피의 손실은 염색체의 2개 카피 중 하나의 손실보다 혈장 DNA에서 염색체 영역의 더 큰 과소-표현을 초래할 것이고, 그러므로, 더 많은 음성  $z$ -스코어를 초래하였다. 전형적으로, 암에서 다중 염색체 비정상이 있다. 각각의 암에서 염색체 비정상은 추가로 그의 성질 (즉 증폭 또는 결실), 그의 정도 (단일 또는 다중 카피 이득 또는 손실) 및 그의 규모 (염색체 길이에 관하여 비정상의 크기)에 따라 다양할 수 있다.

[0045] 정규화된 태그 카운트 측정의 정확성은 분석된 분자의 수에 의해 영향을 받는다. 예를 들어, 단편적 농도가 대

략 12.5%, 6.3% 및 3.2% 각각인 경우, 15,000, 60,000 및 240,000 분자는 하나의 카피 변화 (이득 또는 손실)로 염색체 비정상 검출을 위해 분석되는 것이 필요할 수 있다. 상이한 염색체 영역에 대하여 암의 검출용 태그 계수의 추가 세부사항은 Lo et al에 의해 발명의 명칭 "Diagnosing Fetal Chromosomal Aneuploidy Using Massively Parallel Genomic Sequencing"인 미국 특허 공개 번호 2009/0029377; 및 Lo 등에 의해 발명의 명칭 "Detection Of Genetic Or Molecular Aberrations Associated With Cancer"인 미국 특허 번호 8,741,811에서 기재되고, 이들의 개시는 전체 목적으로 그 전체가 참고로 편입된다.

[0046] B. 방법

[0047] 도 1은 본 발명의 구현예에 따라 증폭을 잠재적으로 나타내는 것으로서 염색체 영역 확인의 방법 100을 예시하는 흐름도이다. 방법 100, 및 본원에서 기재된 다른 방법은 컴퓨터 시스템을 이용하여 전적으로 또는 부분적으로 수행될 수 있다.

[0048] 단계 110에서, 생물의 복수의 염색체 영역은 확인될 수 있다. 각 염색체 영역은 복수의 유전자자리를 포함할 수 있다. 영역은 1 Mb 크기, 또는 일부 다른 동등 크기일 수 있다. 전체 게놈은 그 다음 약 3,000 영역, 각각의 예정된 크기 및 위치를 포함할 수 있다. 상기 예정된 영역은 사용되는 영역의 특정한 염색체 또는 특정한 수의 길이, 및 본원에서 언급된 임의의 다른 기준을 순응하기 위해 다양할 수 있다. 영역이 상이한 길이를 갖는다면, 상기 길이는, 예를 들면, 본원에서 기재된 바와 같이 결과를 정규화하기 위해 사용될 수 있다.

[0049] 단계 120-140은 각각의 염색체 영역에 대하여 수행될 수 있다. 단계 120에서, 각 염색체 영역에 대하여, 핵산 분자의 각자의 그룹은 염색체 영역 기원인 것으로서 확인될 수 있다. 확인은 표준 게놈에서 핵산 분자의 위치 확인에 기반될 수 있다. 예를 들어, 무세포 DNA 단편은 서열 판독을 획득하기 위해 서열분석될 수 있고, 서열 판독은 표준게놈에 맵핑 (정렬)될 수 있다. 생물이 인간이었다면, 그 다음 표준 게놈은, 잠재적으로 특정한 하위집단으로부터 참조 인간 게놈일 것이다. 또 다른 예로서, 무세포 DNA 단편은 (예를 들면, PCR 또는 다른 증폭 이후) 상이한 프로브로 분석될 수 있고, 여기에서 각 프로브는 상이한 게놈 위치에 대응한다. 일부 구현예에서, 무세포 DNA 단편의 분석은 무세포 DNA 단편에 대응하는 서열 판독 또는 다른 실험적 데이터의 수령, 및 그 다음 컴퓨터 시스템을 이용한 실험적 데이터의 분석에 의해 수행될 수 있다.

[0050] 단계 130에서, 컴퓨터 시스템은 핵산 분자의 각자의 그룹의 각자의 양을 계산할 수 있다. 각자의 값은 각자의 그룹의 핵산 분자의 특성을 한정한다. 각자의 값은 본원에서 언급된 임의의 값일 수 있다. 예를 들어, 값은 그룹내 단편의 수 또는 그룹내 단편의 크기 분포의 통계치일 수 있다. 각자의 값은 또한 정규화된 값, 예를 들면, 샘플에 대하여 태그 카운트의 총 수 또는 참조 영역에 대하여 태그 카운트의 수에 의해 분할된 영역의 태그 카운트일 수 있다. 각자의 값은 또한 또 다른 값으로부터 차이 또는 비일 수 있고, 이로써 영역에 대하여 차이의 특성을 제공할 수 있다.

[0051] 단계 140에서, 각자의 양은 염색체 영역이 비정상 (즉 증폭 또는 결실)을 나타내는지의 분류를 결정하기 위해 기준치와 비교될 수 있다. 일부 구현예에서, 염색체 영역은 비정상을 나타내지 않는 것으로서 분류될 수 있다. 비교는 각자의 양 및 기준치에 기반된 z-스코어 결정을 포함할 수 있다. 예로서, 기준치는 본원에서 기재된 임의의 한계값 또는 기준치일 수 있다. 예를 들어, 기준치는 정상 샘플에 대하여 결정된 역치일 수 있다. 또 다른 예로서, 기준치는 또 다른 영역에 대하여 태그 카운트일 수 있고, 비교는 차이 또는 비 (또는 상기의 함수) 취득 및 그 다음 차이 또는 비가 역치 초과인지의 결정을 포함할 수 있다.

[0052] 기준치는 다른 영역의 결과에 기반하여 다양할 수 있다. 예를 들어, 인접하는 영역이 또한 편차를 보여주면 (비록 한계값과 비교하여 작은, 예를 들면, 3의 z-스코어이어도), 더 낮은 한계값이 사용될 수 있다. 예를 들어, 3개 연속적인 영역이 모두 제1 한계값 초과이면, 암은 더욱 가능성이 있을 수 있다. 따라서, 상기 제1 한계값은 비-연속적인 영역으로부터 암을 확인하기 위해 요구되는 또 다른 한계값보다 작을 수 있다. 심지어 작은 편차를 갖는 3개 (또는 3개 초과) 영역의 가짐은 감수성 및 특이성이 보존될 수 있는 기회 효과의 낮은 충분한 개연성을 가질 수 있다.

[0053] C. 염색체 아암-수준 Z-스코어 분석 (CAZA)

[0054] 일부 구현예에서, 염색체는 많은 하위염색체 영역 (예를 들면, 1 Mb 영역)으로 분열될 수 있다. 상기 고 분해능은 감수성 및 특이성을 최대화할 수 없다. 다른 구현예는 2개 아암, 즉 p 및 q로 염색체를 분열시킬 수 있다. 2개 아암의 분석은 상기 미세 분해능에 의해 야기된 노이즈 감소에 의해 특이성을 개선시킬 수 있다. 염색체 아암-수준 z-스코어 분석의 예는 이제 제공된다.

[0055] 우리는 90 HCC 환자, 만성 HBV 감염을 가진 67 환자, HBV-관련 간경변증을 가진 36 환자 및 32 건강한 대상체로

부터 총 225 혈장 DNA 샘플을 분석하였다. 31 백만 판독 (범위: 17-79 백만)의 중앙값은 각 혈장 샘플로부터 수득되었다. 건강한 대조군의 평균 미만 ( $z$ -스코어  $< -3$ ) 3개 SDs 및 초과 ( $z$ -스코어  $> 3$ ) 3개 SDs가 있는 염색체 아암에서 유래하는 서열 판독의 양은 그 염색체 아암, 각각으로부터 혈장 DNA의 유의미한 과소- 및 과대-표현을 나타낸다고 간주되었다. 이들 정량적 혈장 DNA 비정상은 일반적으로 종양 (4)에서 카피 수 손실 및 카피 수 이득 (CNAs)의 존재의 반영이었다.

[0056] 도 2는 본 발명의 구현예에 따라 대표적인 간세포 암종 (HCC) 환자의 혈장 및 조직 샘플에서 증폭 및 결실을 나타내는 영역을 확인하는 시르코스 플롯 200을 보여준다. 내부에서 외부로: 종양 조직내 CNAs (1-Mb 분해능으로); 종양 조직내 아암-수준 CNAs; 혈장 CNAs (1-Mb 분해능으로); 아암-수준 혈장 CNAs. 이득 및 손실을 가진 영역은 녹색 및 적색, 각각으로 보여진다. 2개 연속적인 수평선 사이의 거리는 5의  $z$ -스코어를 나타낸다. 염색체 이디오그램 (플롯 외부)은 시계 방향으로 pter부터 qter까지 배향된다.

[0057] 도 3은 CAZA의 구현예를 이용하여 모든 연구된 대상체에 대하여 혈장 카피 수 비정상 (CNA) 결과를 보여준다. HCC에서 CNAs에 의해 빈번하게 영향받는 4개 염색체 아암 (1p, 1q, 8p 및 8q)은 분석되었다. 적색 및 녹색 선은 혈장에서 대응하는 염색체 아암의 과소- 및 과대-표현, 각각을 나타낸다. 각 수직선은 하나의 케이스에 대한 데이터를 나타낸다.

[0058] 도 4는 본 발명의 구현예에 따라 HCC 환자, HBV 캐리어, 간경변증을 가진 환자 및 건강한 대상체의 혈장에서 CNA의 검출가능성을 보여주는 표 400이다. 표 400은 제일 왼쪽 칼럼에서 환자의 카테고리를 보여준다. 나머지 칼럼은 환자의 수 및 상이한 염색체 아암에 대하여 혈장에서 검출된 CNA를 가진 백분율을 보여준다. 90 HCC 환자의 76명 (84.4%)은 혈장내 염색체 1 및 8에서 적어도 하나의 염색체 아암-수준 CNA를 가졌다. 12 HCC 환자의 종양 조직은 혈장 DNA 발견을 확증하기 위해 이용가능하였다. 조직 샘플은 서열분석되었고 CNA 패턴은 도 5에서 보여진다.

[0059] 도 5는 12 HCC 환자의 종양 및 대응하는 혈장에서 검출된 CNAs의 표 500을 보여준다. 표 500에서, 환자 케이스 수는 제1 칼럼에서 열거된다. 환자는, 제2 칼럼에 나타난 바와 같이, 혈장내 종양 DNA 분획의 내림차순으로 배열된다. 제3 칼럼은 종양 크기를 보여준다. 나머지 칼럼은 상이한 염색체 아암에 대하여 종양 및 혈장에서 검출된 CNAs를 보여준다. '이득'은 카피 수 이득을 나타낸다. '손실'은 카피 수 손실을 나타낸다. 'nil'은 검출가능한 CNA 없음을 나타낸다. 총 48 염색체 아암은 12 환자에 대하여 분석되었다. 종양과 혈장 사이에서 조화된 및 비조화된 결과를 가진 염색체 아암의 수 (및 백분율)은 보여진다.

[0060] 12 환자에 대하여 분석된 48 염색체 아암 중에서, 혈장 및 종양 조직내 조화된 변화는 30 (63%) 아암에 대하여 관측되었다. CNAs는 10 (21%) 아암에 대하여, 혈장이 아닌, 종양에서만 관측되었다. 이들 케이스는 혈장에서 더 적은 종양 DNA 분획을 갖는 경향이 있었다. CNAs는 7 (15%) 아암에 대하여, 종양이 아닌, 혈장에서 관측되었다. 하나의 케이스 (HOT428)에서, 1q의 이득은 종양에서 관측되었지만, 손실은 혈장에서 관측되었다. 이들 데이터는 혈장 DNA에 기여하는 암 세포의 다른 병소 또는 클론일 수 있는 종양성 이중성의 존재를 시사할 수 있다.

[0061] 간경변증이 있는 및 없는 HBV 캐리어 중에서, 이들 CNA의 검출 속도는 22.2% 및 4.5%, 각각이었다. 간경변증을 가진 한명의 환자 및 간경변이 없는 한명의 만성 HBV 캐리어는, 채혈시에 HCC를 갖는 것으로 공지되지 않은, 혈장에서 CNAs를 나타냈고, 각각, 3 개월 및 4 개월 후에 HCC를 갖는 것으로서 진단되었다. 모든 HBV 캐리어 및 간경변 환자는 적어도 6 개월 동안 추적되었다. 혈장에서 임의의 CNA가 없는 그 대조군 대상체에 대하여, 이들 중 아무도 추적 기간 동안 HCC를 발생시키지 않았다. 32 건강한 대상체의 아무도 CAZA에 의해 혈장내 염색체 1 또는 8에서 검출가능한 CNA를 갖지 않았다. HCC 환자에서, CNA의 존재로 인해 혈장내 서열 판독에서 불균형 증가 또는 감소는 혈장 샘플내 종양 DNA의 단편적 농도를 반영한다. HCC 환자의 혈장에서 종양-유도된 DNA의 종양 단편적 농도는 2.1% (범위: 0% 내지 53.1%; 사분위간 범위: 1.2% 내지 3.8%)이었다.

[0062] CAZA는 비-침습적으로 종양-관련 CNAs를 검출하기 위한 방식을 제공한다. HCC에서, 염색체 1 및 8은 CNAs (34-36)에 의해 통상적으로 영향을 받는다. 사실상, 우리의 데이터는 90 HCC 환자의 76 (84.4%)이 혈장내 염색체 1 및 8에서 어느 하나의 아암을 포함한 적어도 하나의 CNA를 가졌고, 반면에 32 건강한 대상체의 아무도 혈장내 이들 2 염색체에 대하여 임의의 CNA를 나타내지 않았음을 보여주었다. 염색체 1 및 8을 포함한 혈장 CNAs는 간경변 환자 및 HBV 캐리어의 22.2% 및 4.5%로 또한 검출되었다. 한명의 HBV 캐리어 및 한명의 간경변증을 가진 환자에서, HCC는 채혈 직후 진단되었다. 암이 채혈시에 존재할 것이고 혈장내 CNAs와 관련되었고, 이로써 구현예의 초기 선별 능력을 보여줄 것 같다. HCC 환자에서 혈장 CNAs의 상대적으로 높은 검출 속도는 상기 접근법이 HBV 캐리어의 선별에서 미래 가치를 가질 것을 시사한다. 또한, CNAs는 암 (33)의 거의 모든 유형에서 존재한다. 따라서, 상기 접근법은 당해 암의 특이적 CNA 패턴에 적응으로 일반적인 종양 마커로서 적용될 수 있

다.

[0063] **III. 비정상 영역의 패턴에 기반된 암 유형 검출**

[0064] 일부 구현예는 샘플에서 확인된 비정상에 의해 연루된 잠재적인 암을 확인하기 위해 암의 유형의 (증폭 또는 결실과 함께) 공지된 비정상 영역을 이용할 수 있다. 상기 예에서, HCC에 대하여 공지된 비정상 영역은 HCC에 대하여 샘플을 선별하는데 사용되었다. 상기 선별은 비정상 영역의 공지된 세트와 (증폭 또는 결실을 포함한) 확인된 비정상 영역을 비교할 수 있다. 충분히 높은 매치가 결정되면, 암의 그 유형은 가능한 시험 결과로서 표시가 될 수 있다.

[0065] 매칭 기준은 샘플에서 또한 확인되는 세트의 영역의 백분율일 수 있다. 매칭 기준은 특이적 영역을 비정상이도록 요구할 수 있다. 예를 들어, 1p, 1q, 또는 8q가 비정상인 경우, 또는 이들 염색체 아암의 1 초과가 비정상인 경우, 매치는 HCC에 대하여 확인될 수 있다. 따라서, 동일한 매치가 요구되는 특이적 서브세트일 수 있지만, 서브세트는 암의 유형에 대하여 공지된 비정상 영역의 전체 세트보다 더 작을 수 있다.

[0066] 따라서, 시험 샘플에 대하여 비정상 영역의 패턴은, 암의 특정한 유형을 갖도록 공지된 환자로부터 결정될 수 있는, 암의 특정한 유형에 대하여 비정상 영역의 패턴과 비교될 수 있다. 구현예는, 특히 종양이 작을 수 있는 (예를 들면, 2 cm 미만 크기일 수 있는), 관여된 암의 유형을 확인하기 위해 및 암에 대하여 선별하기 위해 사용될 수 있다. 화상형성 기술은 2 cm 미만 크기의 종양 확인에서 어려움을 갖는다. 상기 기술은 또한 치료 이후 환자의 진행을 추적하는데 사용될 수 있다.

[0067] **A. 방법**

[0068] 도 6은 생물학적 샘플이 본 발명의 구현예에 따라 암의 제1 유형을 나타내는지 결정하기 위해 생물의 생물학적 샘플 분석의 방법 600을 예시하는 흐름도이다. 생물학적 샘플은 정상 세포 및 잠재적으로 암과 관련된 세포에서 유래하는 핵산 분자 (또한 단편으로 불림)를 포함한다. 이들 분자의 적어도 일부는 샘플에서 무세포일 수 있다.

[0069] 본원에서 기재된 상기 및 임의의 다른 방법의 하나의 구현예에서, 생물학적 샘플은 무세포 DNA 단편을 포함한다. 혈장 DNA의 분석이 본 출원에서 기재된 상이한 방법을 예시하기 위해 사용되었어도, 이들 방법은 정상 및 종양-유도된 DNA의 혼합물을 함유한 샘플에서 종양-관련 염색체 비정상을 검출하기 위해 또한 적용될 수 있다. 다른 샘플 유형은 타액, 눈물, 늑막 유체, 복수액, 담즙, 소변, 혈청, 채액, 대변 및 자궁경부 도말표본 샘플을 포함한다.

[0070] 단계 610에서, 생물의 복수의 염색체 영역은 확인된다. 복수의 염색체 영역은 하위염색체이고 비-중첩일 수 있다. 계수되는 염색체 영역은 제약일 수 있다. 예를 들어, 적어도 하나의 다른 영역과 인접한 영역만이 계수될 수 있다 (또는 인접 영역이 특정 크기, 예를 들면, 4 이상 영역인 것으로 요구될 수 있다). 영역이 동일하지 않은 구현예에 대하여, 수는 또한 각자의 길이를 설명할 수 있다 (예를 들면, 수는 비정상 영역의 총 길이일 수 있다). 일부 구현예에서, 상기 영역은 염색체의 아암에 대응한다. 다른 구현예에서, 상기 영역은 아암, 예를 들면, 1-Mb 영역보다 더 작을 수 있다.

[0071] 일부 구현예에서, 염색체 영역은 특정한 일배체형일 수 있다 (즉, 특정한 염색체 카피에 대응한다). 상대적 일배체형 복용량 (RHDO) 분석을 이용한 구현예에서, 각 영역은 적어도 2개의 이종접합성 유전자자리를 포함할 수 있다. RHDO에 관한 추가 세부사항은 미국 특허 번호 8,741,811에서 발견될 수 있다.

[0072] 단계 620에서, 생물의 생물학적 샘플에서 각각의 복수의 핵산 분자에 대하여, 생물의 표준 게놈에서 핵산 분자의 위치는 확인될 수 있다. 복수의 핵산 분자는 500,000 이상 분자 (단편)를 포함할 수 있다. 상기 위치화는, 분자의 1 또는 2 (쌍으로 된-말단) 서열분석된 태그를 수득하기 위해, (예를 들면 랜덤 서열분석을 통해) 분자의 서열분석 수행 및 그 다음 표준 게놈에 서열분석된 태그(들) 정렬을 포함한다, 다양한 방식으로 수행될 수 있다. 상기 정렬은 예컨대 기본 부분 정렬 탐색 도구 (BLAST)같은 도구를 이용하여 수행될 수 있다. 위치는 염색체의 아암내 수로서 확인될 수 있다.

[0073] 단계 630에서, 핵산 분자의 각자의 그룹은 각각의 복수의 염색체 영역에 대하여 확인된 영역에 기반된 염색체 영역 기원인 것으로서 확인될 수 있다. 각자의 그룹은 염색체 영역의 각각의 복수의 유전자자리에 위치한 적어도 하나의 핵산 분자를 포함할 수 있다.

[0074] 단계 640에서, 컴퓨터 시스템은 각각의 복수의 염색체 영역에 대하여 핵산 분자의 각자의 그룹의 각자의 값을 계산할 수 있다. 각자의 값은 각자의 그룹의 핵산 분자의 특성을 한정할 수 있다. 특성은 핵산 분자의 카운트,

백분율, 또는 크기일 수 있다. 각자의 값은 크기 분포의 평균, 크기 분포의 중앙값, 크기 분포의 형태, 또는 크기 한계값 미만의 크기를 갖는 핵산 분자의 비율을 포함할 수 있다. 특성으로서 크기 이용은 섹션 IV에서 더 상세히 논의된다.

- [0075] 단계 650에서, 각자의 값은 염색체 영역이 결실 또는 증폭을 나타내는지의 분류를 결정하기 위해 각자의 기준치와 비교될 수 있다. 비교는 각자의 값 및 각자의 기준치에 기반된  $z$ -스코어 결정을 포함할 수 있다.  $z$ -스코어는 그 다음 결실 또는 증폭이 존재하는지를 결정하기 위해 하나 이상의 역치와 비교될 수 있다. 상이한 한계값은 결실 및 증폭에 대하여 사용될 수 있다. 다른 구현예에서, 예를 들면,  $z$ -스코어에서 다른 값이 방정식의 다른 측으로 이동하였다면, 기준치는 역치를 포함할 수 있다. 시험된 영역이 제1 일배체형인 경우, 기준치는 건강한 샘플, 또 다른 염색체 영역 (예를 들면, 비정상을 나타내지 않는 것), 또는 다른 일배체형에서 결정된 값에 대응할 수 있다.
- [0076] 단계 660에서, 결실 또는 증폭을 나타내는 염색체 영역의 시험 패턴은 결정될 수 있다. 시험 패턴은 시험된 샘플에서 비정상 영역의 패턴을 지칭한다. 시험 패턴은 결실, 증폭을 나타내는, 또는 정상인 염색체 영역의 세트를 포함할 수 있다. 시험 패턴은 또한 증폭을 나타내는 것으로서 확인되는 세트의 제1 서브세트를 포함할 수 있다. 시험 패턴은 결실을 나타내는 것으로서 확인되는 세트의 제2 서브세트를 추가로 포함할 수 있다. 시험 패턴은 증폭 또는 결실을 나타내지 않은 것으로서 확인되는 세트의 제3 서브세트를 추가로 포함할 수 있다.
- [0077] 단계 670에서, 시험 패턴은 암의 상이한 유형의 복수의 참조 패턴과 비교될 수 있다. 암의 유형에 대한 참조 패턴은 비정상 영역의 공지된 세트를 포함할 수 있다. 참조 패턴은 무세포 핵산 분자의 조직 및/또는 혼합물의 참조 샘플로부터 결정될 수 있다. 참조 패턴은 수많은 영역을 포함할 수 있고, 각각은 증폭, 결실, 또는 비정상 없음의 한정된 상태를 갖는다. 비교는 시험 패턴의 영역이 참조 패턴내영역과 동일한 비정상을 갖는지를 결정할 수 있다. 예를 들어, 동일한 영역이 증폭, 결실을 갖는 것으로서 제시되거나, 또는 시험 패턴 및 참조 패턴 둘 모두에서 정상인지가 결정될 수 있다.
- [0078] 단계 680에서, 비교에 기반하여, 암의 제1 유형에 대응하는 제1 참조 패턴으로서 동일한 결실 또는 증폭을 나타내는 시험 패턴의 영역의 양은 결정될 수 있다. 다양한 구현예에서, 양은 비정상 영역의 공지된 세트와 매칭하는 염색체 영역의 수 또는 백분율일 수 있다.
- [0079] 단계 690에서, 영역의 양은 생물학적 샘플이 암의 제1 유형을 나타내는지의 제1 분류를 결정하기 위해 제1 한계값과 비교된다. 제1 한계값은 암의 제1 유형에 특이적일 수 있거나 또는 암의 다중 유형을 거쳐 사용될 수 있다. 그와 같은 한계값은 확인되는 암의 제1 유형에 대하여 비정상 영역의 공지된 세트와 매칭하는데 필요한 염색체 영역의 최소 양일 수 있다. 다양한 구현예에서, 최소 양은 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 또는 13 염색체 영역일 수 있다. 일부 구현예에서, 특이적 영역은 비정상인 것으로 요구될 수 있고, 따라서 다른 기준은 제1 한계값의 양의 비교 이외에 사용될 수 있다. 상기 특이적 영역은 구속될 수 있거나 또는 다른 영역보다 더 많이 계량될 수 있다. 특이적 비정상 영역은 암의 유형에 대하여 공지된 비정상 영역의 전체 세트의 서브세트일 수 있다. 암의 유형은, 다른 암 중에서, HCC, 대장암, 유방암, 폐암, 또는 비인두 암종을 포함할 수 있다.
- [0080] 분류를 결정하기 위해 사용된 역치는 계수되는 영역의 위치 및 크기에 기반하여 다양할 수 있다. 예를 들어, 염색체의 특정한 염색체 또는 아암에서 영역의 양은 암의 특정한 유형이 연루되는지의 결정에 대한 기준으로서 그 특정한 염색체 (또는 아암)에 대하여 한계값과 비교될 수 있다. 다중 한계값은 사용될 수 있다. 예를 들면, 특정한 염색체 (또는 아암 또는 더 큰 하위염색체 영역)에서 매칭 영역의 양 (즉, 시험 패턴 및 참조 패턴에서 비정상의 동일한 분류)은 제1 역치 초과이도록 요구될 수 있고, 계놈에서 매칭 영역의 총 양은 제2 역치 초과이도록 요구될 수 있다.
- [0081] 매칭 영역의 양에 대한 역치는 불균형이 영역의 분류에 대하여 얼마나 강한지에 또한 의존할 수 있다. 예를 들어, 암의 유형의 분류 결정용 한계값으로서 사용되는 매칭 영역의 양은 각 영역에서 비정상을 검출하기 위해 사용된 특이성 및 감수성 (비정상 한계값)에 의존할 수 있다. 예를 들어, 비정상 한계값이 낮으면 (예를 들면  $z$ -스코어 2), 양 한계값은 높도록 (예를 들면, 15 매칭 영역 또는 80%) 선택될 수 있다. 그러나, 비정상 한계값이 높으면 (예를 들면,  $z$ -스코어 3), 양 한계값은 더 낮을 수 있다 (예를 들면, 5 매칭 영역 또는 60%). 비정상을 보여주는 영역의 양은 또한 계량된 값, 예를 들면, 높은 불균형이 사소한 불균형을 단지 보여주는 영역보다 더 많이 계량될 수 있다는 것을 보여주는 하나의 영역일 수 있다 (즉 비정상에 대하여 단지 양성 및 음성보다 더 많은 분류가 있다). 그와 같은 계량은 확인되는 암의 유형에 대하여 비정상을 갖도록 요구되는 특정 영역과 유사한 방식으로 작용할 수 있다.

- [0082] 일부 구현예에서, 한계값은 암의 다른 유형에 대하여 매칭 영역의 수에 기반하여 동적으로 결정될 수 있다. 예를 들어, 한계값은 확인된 암에 대하여 매칭 영역의 수가 다음 가장 유사한 암 유형에 대하여 매칭 영역 초과적 적어도 특이적 수라는 것일 수 있다. 그와 같은 한계값은 최소 한계값에 더하여 추가의 기준일 수 있다. 따라서, 일부 경우에서, 매칭 영역의 충분한 수가 존재하지 않으면 암 유형이 확인될 수 없다.
- [0083] B. 결과
- [0084] 방법 600은 정확도를 결정하기 위해 복수의 암 유형에 대하여 시험되었다. 방법 600은 공지된 암 유형의 환자로 시험되었다. 추가로, 사용된 한계값은 공지된 암 유형의 샘플을 이용하여 결정될 수 있다. 상이한 한계값은 상이한 암 유형에 대하여 사용될 수 있다.
- [0085] 각각의 17 암 환자 (HCC를 가진 6 환자, 대장암 (CRC)을 가진 4, 유방암 (BrC)을 가진 3, 폐암 (LC)을 가진 2 및 비인두 암종 (NPC)을 가진 2)의 혈장 DNA는 서열분석되었다. 각 염색체 아암에 대하여 카피 수 비정상 (CNAs)은 CAZA 접근법에 기반된 각 환자에 대하여 분석되었다.
- [0086] 도 7은 구현예에 따라 표 700에서 암의 상이한 유형에 대하여 상이한 패턴을 나타내는 염색체 아암을 보여준다. 케이스의  $\geq 50\%$ 에서 발생하는 CNAs는 색상으로 강조된다. 카피 수 손실은 적색으로 강조되고, 카피 수 이득은 녹색으로 강조된다.
- [0087] 표 700은 제일 왼쪽 칼럼 상에서 열거된 염색체 아암을 갖는다. 각각의 다른 칼럼은 암의 유형 및 암 유형에 대한 환자 수를 열거한다. 결실은 '-'으로 확인된다. 증폭은 '+'로 확인된다. 정상 영역은 'NIL'로 확인된다.
- [0088] 혈장 샘플에서 관측된 CNAs의 패턴은 암의 상이한 유형으로 고통받고 있는 환자에 대하여 상이하다. 혈장에서 관측된 CNAs의 가장 흔한 패턴에 기반하여, 구현예는 혈장에서 관측된 CNAs를 가진 환자에서 암의 잠재적인 조직 기원을 추론할 수 있지만 CNAs의 공급원은 미궁지된다. 표 700에서 열거된 CNAs의 패턴은 설명하기 위한 것이고, CNAs의 더욱 포괄적인 목록은 임상 샘플의 훨씬 더 큰 수의 분석에 의해 확립될 수 있다.
- [0089] CNAs의 참조 패턴은 또한 종양 조직의 분석으로부터 결정될 수 있다. 예로서, 1q, 6p, 8q 및 17q에서 이득, 및 4q, 8p, 13q, 16q 및 17p에서 손실은 HCC 종양 조직에서 통상적으로 검출된다 (Moizadeh P 등 Br J Cancer 2005;92:935-941). 5p, 8q, 9p, 13q 및 20q에서 이득, 및 8p, 9p, 17p 및 18q에서 손실은 CRC 종양 조직에서 통상적으로 검출된다 (Farzana 등 PLoS One 2012;2:231968 and Lips EH 등 J Pathol 2007;212:269-77). 5p, 7p, 7q, 8q 14q, 17q 및 20q에서 이득, 및 3p, 8p, 9p, 13q 및 18q에서 손실은 비-소세포 폐암 조직에서 통상적으로 검출되고 반면에 3q, 5p, 14q 및 19q에서 이득, 및 3p, 4p, 4q, 5q, 10p, 10q, 13q, 15q, 17p 및 22q에서 손실은 소세포 폐암 조직에서 통상적으로 검출된다 (Zhao X 등 Cancer Res 2005;65:5561-70). 1q, 8q, 17q 및 20q에서 이득, 및 4p, 5q, 8p, 11q 및 13q에서 손실은 유방암 조직에서 통상적이다 (Andre F 등 Clin Cancer Res 2009;15:441-51). 본원에서 기재된 CNAs의 패턴은 예시적인 예로서 제공되고 본원에서 기재된 방법에서 사용될 수 있는 유일한 패턴이도록 의도되지 않는다.
- [0090] 본 예에서 CNA 패턴에 기반하여, 혈장 DNA 서열분석이 암 검진의 목적을 위하여 환자 BrC2에 대하여 수행되었다는 것을 추정한다. 1q, 3q, 8q, 및 14q에 대하여 카피 수 이득 및 2p, 2q, 3p, 4p, 7q, 8p, 9p, 11p, 12p, 12q, 16q, 및 17p에 대하여 카피 수 손실을 포함하는, CNAs는 관측되었다. 그녀의 혈장에서 CNAs는 유방암에 대하여 13 전형적인 CNAs를 매칭시켰다. 그에 반해서, 그녀의 CNAs는 HCC, CRC, LC, 및 NPC, 각각의 3, 6, 4, 및 1 전형적인 CNAs를 단지 매칭시켰다. 따라서, 그녀의 혈장 DNA의 CNA 패턴에 기반하여, 그녀가 갖는 가장 유사한 암은 유방암인 것으로 추론된다. 선택된 한계값은 관측된 CNAs의 수가 특정 암 유형의 전형적인 CNAs와 양립 가능하는지를 결정하기 위해 사용될 수 있다. 상기 예에서, 7, 8, 9, 10, 11, 12, 또는 13의 한계값은 유방암과 양립가능한 것으로서 CNAs를 분류하기 위해 사용될 수 있다. 매칭 영역의 백분율은 또한 사용될 수 있다. 예를 들어, 통상적으로 비정상 영역을 매칭시키는 영역의 백분율은 사용될 수 있다. 통상적으로 비정상 영역은 참조 샘플의 50% 초과로 특정한 비정상을 갖는 영역으로서 한정될 수 있다.
- [0091] 다른 구현예에서, 다른 통계적 접근법, 예를 들어, 비제한적으로 계층적 클러스터링은, 환자가 갖고 있는 가장 유사한 암 유형을 추론하기 위해 사용될 수 있다. 예를 들어, 각 참조 샘플은 다차원 데이터 포인트로 배정될 수 있고, 여기에서 각각의 치수는 상이한 영역에 대응한다. 하나의 실행에서, 각각의 치수는 -1 (결실에 대하여), 0 (정상), 또는 1 (증폭에 대하여)로 배정될 수 있다. 더 높은 수는 증폭의 상이한 수준에 대하여 가능할 수 있다. 특정한 암 유형에 대한 샘플은 함께 클러스터링할 것이고, 신규 샘플은 클러스터에 배정될 수 있다. 한계값은 신규 샘플이 배정되어야 하는 (있다면) 클러스터를 결정하기 위해 사용된 측정의 기준에 대응할 수 있고, 여기에서 배정된 클러스터는 샘플에 대하여 확인된 암 유형에 대응한다. 예를 들어, 클러스터는 클러

스터의 참조 패턴의 적어도 예정된 수만큼 공유된 클러스터의 참조 패턴의 영역에 대응하는 센트로이드를 가질 수 있다. 클러스터는 시험 패턴이 클러스터의 내부에 놓이는 것을 한정하는 경계를 포함할 수 있다. 경계는 단순히 구형을 넘어 다양한 형상을 가질 수 있다. 경계는 클러스터가 속하는 참조 패턴 결정의 경우 클러스터링 분석의 일부로서 결정될 수 있고, 여기에서 센트로이드에서 가장 멀리 떨어진 그러나 클러스터 내에 있는 참조 패턴이 경계를 한정할 수 있다. 시험 패턴이 클러스터의 일부인지의 결정에 대한 한계값은 센트로이드부터 시험 패턴까지 방향으로 센트로이드부터 경계까지 거리로 고려될 수 있다.

[0092] 더욱 또 다른 구현예에서, 암의 상이한 유형을 가질 상대적 가능성은 결정될 수 있다. 환자의 CNA 패턴은 암의 각 유형에 대하여 CNA의 가능성에 대해 비교될 수 있다. 예를 들어, 환자가 1q 이득을 갖는다는 것은 암의 상이한 유형에 대하여 1q 이득의 개연성에 대해 비교될 것이다. 예시 목적을 위하여, 우리는 1q 이득이 HCC 환자의 70%, LC 환자의 20%, 및 CRC 환자의 1%에서 발생할 수 있다는 것을 가정한다. 이들 가능성으로, 오즈비는 CNA를 갖는 상이한 암 유형을 가진 환자의 상대적 백분율에 기반하여 결정될 수 있다. 예를 들면, 1q 이득에 기반하여, 환자는 3.5배 더 유사하게 LC보다 HCC를 갖는다고 그리고 70배 더 유사하게 CRC보다 HCC를 갖는다고 고려될 수 있다. HCC 대 LC 대 CRC에 대한 오즈비는 70:20:1일 수 있다. 기술 중 하나는 상기 오즈비가 몇 개의 상이한, 더욱 등가인, 형태로 표시될 수 있다는 것을 이해할 것이다. 1q 이외의 염색체 아암에서 상이한 CNAs에 대한 오즈비는 또한 결정될 수 있다. 전체 오즈비는 그 다음 개별적인 CNAs에서 가능성 또는 오즈비로 계산될 수 있다. 환언하면, 환자로부터 CNA 패턴 및 주어진 CNA 패턴을 갖는 암의 상이한 유형의 가능성이 주어지면, 암의 상이한 유형의 가능성은 전체 오즈비에서 서로에 비교될 수 있다. 상기 예가 상이한 염색체 아암에서 CNAs의 가능성을 사용하였어도, 염색체 아암 이외의 상이한 하위염색체 영역에서 CNAs의 가능성은 사용될 수 있다. 일부 구현예에서, CNA이 환자의 염색체 아암 또는 다른 하위염색체 영역에서 발견되지 않으면, 무 CNAs의 패턴은 암의 상이한 유형에 대하여 염색체 아암 또는 하위염색체 영역에서 CNA 발견못함의 가능성에 대해 비교될 수 있다. 환자로부터 CNAs 없는 영역의 패턴은 그 다음 암의 상이한 유형의 가능성을 결정하기 위해 사용될 수 있다. 또한, CNAs을 가진 영역 및 CNAs 없는 영역의 분석 조합은 영역의 단 하나의 유형이 사용된 것보다 잠재적으로 더 높은 정확도에서 암의 유형의 가능성 또는 상대적 가능성을 결정하기 위해 사용될 수 있다.

[0093] 또 다른 예에서, 환자 NPC1이 서열분석된 혈장 DNA를 갖는다는 것을 가정한다. 2q, 12q, 및 22q에 대하여 카피 수 이득 및 6q 및 18q에 대하여 카피 수 손실을 포함한, CNAs는 관측되었다. 상기 환자의 CNA 패턴은 NPC에 대하여 전형적인 CNAs 중 4개를 매칭시켰다. 비교로, 상기 CNA 패턴은 HCC, CRC, BrC, 및 LC의 패턴에 대하여 0, 2, 0, 및 0 전형적인 CNAs를 매칭시켰다. 또 다른 구현예에서, 암 유형에 대하여 전형적인 CNA의 부족은 또한 계수될 수 있다. 예를 들어, NPC에 대하여 전형적인 CNAs의 아무것도 상기 환자에서 부재가 아니었다. 그에 반해서, HCC, CRC, BrC 및 LC에 대하여 7, 16, 13, 및 8 전형적인 CNAs는 상기 환자에서 부재이었다. 따라서, 상기 환자의 CNA 패턴은 HCC, CRC, BrC, 및 LC를 시사하지 않는다.

[0094] 도 8A, 8B, 및 8C는 상기 접근법의 정확도가 표 800에서 더 높은 분해능 CNA 분석을 이용함으로써 추가로 얼마나 향상될 수 있는지를 보여준다. 1-Mb 영역에 영향을 미치는 CNA는 암 환자의 상기 집단에서 확인되었다. 표 800은 제일 왼쪽 칼럼에서 열거된 1-MB 영역의 게놈 배위를 갖는다. 각각의 다른 칼럼은 암의 유형 및 암 유형에 대한 환자 수를 열거한다. 결실은 '-'로 확인된다. 증폭은 '+'로 확인된다. 정상 영역은 'NIL'로 확인된다.

[0095] 상기 예에서, 1 Mb 미친 및 동일한 암 유형을 갖는 모든 환자에서 존재한 CNAs는 확인되었다. 더 높은 분해능으로, 암의 동일한 유형을 가진 환자의 고 비율로 존재하는 하위염색체 CNAs는 확인될 수 있다. 이들 암-유형-특이적 CNAs는 아암-기반 분석에서 확인되지 않는다. 예를 들어, 배위 30-31 Mb 및 44-45 Mb 미치는 염색체 18에서 카피 수 이득은 3명의 폐암을 가진 환자 모두에서 확인되었지만 다른 암 유형을 가진 환자에서 흔치 않았다. 상기에 논의된 바와 같이, 상이한 통계적 시험은 암-특이적 CNA 패턴이 시험된 케이스에 가장 유사한지를 결정하기 위해 사용되었다. 상이한 통계적 시험은, 예를 들어, 상이한 암-관련 CNA 패턴 및 계층적 클러스터링에서 전형적인 CNAs의 수 계수를 포함할 수 있다.

[0096] **IV. 혈장에서 종양-유도된 DNA 단편의 크기 분석**

[0097] DNA 단편의 크기 분포에서 통계적으로 유의미한 차이는, 카운트의 수가 할 수 있는 유사한 방식으로, 비정상을 확인하기 위해 사용될 수 있다. 총 (즉 종양성 플러스 비-종양성) 혈장 DNA의 크기 분포가 암 환자에서 증가되는 것이 보고되었다 (Wang BG, 등 Cancer Res. 2003; 63: 3966-8). 그러나, (DNA의 총 (즉 종양 플러스 비-종양) 양 대신) 종양-유도된 DNA를 구체적으로 연구하고 있으면, 종양-유도된 DNA 분자의 크기 분포가 비-종양 세포로부터 유도된 분자의 것보다 더 짧다는 것이 관측되었다 (Diehl 등 Proc Natl Acad Sci U S A. 2005;102:16368-73). 따라서, 순환형 DNA의 크기 분포는 암-관련 염색체 비정상이 존재하는지의 결정을 위하여

사용될 수 있다.

[0098] 크기 분석은 본원에서 언급된 바와 같이 그리고 미국 특허 번호 8,620,593에서 다양한 파라미터를 사용할 수 있다. 예를 들어, 상기로부터 Q 또는 F 값은 사용될 수 있다. 이들 값이 관독의 수로 규모화하지 않음에 따라 상기 크기 값은 다른 영역으로부터 카운트에 의해 정규화가 필요 없다. 영역의 깊이 및 개량에 관여한 기술은 사용될 수 있다. 일부 구현예에서, 특정한 영역에 대하여 GC 바이어스는 2개 영역 비교 경우 고려될 수 있다. 일부 실행에서, 크기 분석은 단지 DNA 분자를 사용한다.

[0099] A. 방법

[0100] 도 9는 본 발명의 구현예에 따라 생물의 생물학적 샘플 분석의 방법 900을 예시하는 흐름도이다. 생물학적 샘플은 정상 세포 및 잠재적으로 암과 관련된 세포로부터 유래하는 핵산 분자를 포함할 수 있다. 핵산 분자의 적어도 일부는 생물학적 샘플에서 무 세포일 수 있다. 일 측면에서, 방법 900은 제1 염색체의 단편의 크기 및 하나 이상의 참조 염색체의 단편의 크기에 대하여 구분치 (예를 들면 차이 또는 비)에 기반된 서열 불균형의 분류 결정에 관련될 수 있다.

[0101] 단계 910에서, 생물학적 샘플에서 각각의 복수의 핵산 분자에 대하여, 핵산 분자의 크기는 측정될 수 있다. 핵산 분자의 크기 수득은 2013년 3월 7일 출원된 Lo 등에 의해 발명의 명칭 "Size-Based Analysis of Fetal DNA Fraction in Maternal Plasma"인 미국 특허 공개 번호 2013/0237431에서 기재되고, 이의 내용은 모든 목적으로 본원에서 참고로 편입된다.

[0102] 단계 920에서, 생물의 표준 계놈에서 핵산 분자의 위치는 확인될 수 있다. 위치는, 단계 120 및 다른 곳에 기재되는 바와 같이, 계놈의 임의의 일부일 수 있다. 예를 들어, 염색체 각각의 복수의 핵산 분자가 유도되는 것이 확인된다. 상기 결정은 표준 계놈에 맵핑에 의해 실시될 수 있다.

[0103] 단계 930에서, 각각의 복수의 염색체 영역에 대하여, 핵산 분자의 각각의 그룹은 확인된 위치에 기반된 제1 염색체 영역 기원인 것으로서 확인될 수 있다. 제1 염색체 영역은 복수의 제1 유전자자리를 포함할 수 있다.

[0104] 단계 940에서, 컴퓨터 시스템은 핵산 분자의 제1 그룹의 크기 분포의 제1 통계치를 계산할 수 있다. 구현예에서, 제1 통계치는 특정된 크기에서 제1 곡선하 면적 컴퓨팅에 의해 결정될 수 있다. 제1 곡선은 크기의 범위에 걸쳐 제1 염색체 영역에 대하여 핵산 분자의 누적 빈도의 플롯일 수 있다. 하나의 구현예에서, 제1 통계치는 제1 염색체에 대응하는 단편의 크기 분포의 평균값, 평균, 중앙값, 또는 형태일 수 있다. 또 다른 구현예에서, 제1 통계치는, 컷오프의 유형일 수 있는, 제1 크기 미만 단편의 길이의 합계를 포함할 수 있다. 예를 들어, 200 bp 미만인 각각의 단편은 합계된 그의 길이를 가질 수 있다. 합계는 또 다른 수, 예컨대 제1 염색체에 대응하는 모든 단편의 길이의 합계 또는 (제1 크기와 동일할 수 있는) 제2 크기 컷오프 초과 단편의 길이의 합계로 분할될 수 있다. 예를 들어, 제1 통계치는 단편의 총 길이에 비하여 제1 크기 컷오프 미만 단편의 총 길이의 비, 또는 큰 단편의 총 길이에 비하여 작은 단편의 총 길이의 비일 수 있다.

[0105] 단계 950에서, 제1 통계치는 제1 염색체 영역이 비정상성을 나타내는지의 분류를 결정하기 위해 제1 기준치와 비교될 수 있다. 구현예에서, 제1 기준치는 제2 염색체 영역의 핵산 분자의 제2 그룹의 크기 분포의 통계치일 수 있다. 제2 염색체 영역은 참조 염색체 영역으로 고려될 수 있다. 제1 기준치는 특정된 크기에서 제2 곡선하 면적 컴퓨팅에 의해 결정될 수 있다. 제2 곡선은 크기의 범위에 걸쳐 제2 염색체 영역에 대하여 핵산 분자의 누적 빈도의 플롯일 수 있다. 하나의 구현예에서, 제1 기준치는 복수의 참조 염색체에 대하여 통계치일 수 있다. 하나의 실행에서, 통계치는 조합될 수 있어 이로써 통계치는 하나 이상의 제2 염색체일 수 있다. 또 다른 구현예에서, 복수의 참조 염색체에 대한 통계치는 개별적으로 비교될 수 있다. 비교는 제1 염색체 영역이 결실 또는 증폭을 나타내는지의 분류를 결정할 수 있다.

[0106] 제1 통계치 및 제1 기준치는 구분치를 수득하기 위해 비교될 수 있다. 하나의 구현예에서, 구분치는 결정되는 제1 통계치와 제1 기준치 사이 차이일 수 있다. 또 다른 구현예에서, 구분치는 제1 통계치 대 제1 기준치의 비일 수 있다. 더욱 또 다른 구현예에서, 복수의 구분치는, 예를 들면, 각 참조 염색체에 대하여 계산될 수 있는, 각각의 기준치에 대한 것으로 결정될 수 있다.

[0107] 구분치는 하기 방정식을 이용하여 제1 염색체 영역과 참조 염색체 영역 사이 짧은 DNA 단편의 비율에서의 차이일 수 있다:

[0108] 
$$\Delta F = P(\leq 150bp)_{\text{시행}} - P(\leq 150bp)_{\text{참조}}$$

[0109] 식중,  $P(\leq 150\text{bp})_{\text{시험}}$ 은 크기  $\leq 150$  bp로 제1 염색체 영역에서 유래하는 서열분석된 단편의 비율을 나타내고,  $P(\leq 150\text{bp})_{\text{참조}}$ 는 크기  $\leq 150$  bp으로 참조 염색체 영역에서 유래하는 서열분석된 단편의 비율을 나타낸다. 다른 구현예에서, 다른 크기 한계값은, 예를 들어, 비제한적으로 100 bp, 110 bp, 120 bp, 130 bp, 140 bp, 160 bp 및 166 bp로 사용될 수 있다. 다른 구현예에서, 크기 한계값은 염기, 또는 뉴클레오타이드, 또는 다른 단위로 표시될 수 있다. 일부 실행에서, 참조 염색체 영역은 제1 염색체 영역을 제외한 모든 하위염색체 영역으로서 한정될 수 있다. 다른 실행에서, 참조 영역은 제1 염색체 영역을 제외한 하위염색체 영역의 단지 일부분일 수 있다.

[0110] 카운트-기반 분석에서 사용된 대조군의 동일한 그룹은 크기-기반 분석에서 사용될 수 있다. 시험된 영역의 크기-기반 z-스코어는 대조군의  $\Delta F$ 의 평균 및 SD 값을 이용하여 계산될 수 있다:

$$\text{크기-기반 z-스코어} = \frac{\Delta F_{\text{샘플}} - \text{평균 } \Delta F_{\text{대조군}}}{SD \Delta F_{\text{대조군}}}$$

[0111]

[0112] 구분치는 하나 이상의 컷오프 값과 비교될 수 있다. 하나의 구현예에서, 비교는 각각의 복수의 구분치에 대하여 수행될 수 있다. 예를 들어, 상이한 구분치는 제1 통계치와 각 기준치 사이 결정될 수 있다. 다양한 실행에서, 각 구분치는 동일한 또는 상이한 컷오프 값과 비교될 수 있다. 또 다른 구현예에서, 구분치는 구분치가 특정한 범위내인지를 결정하기 위해 2개 컷오프 값과 비교된다. 범위는 비-정상 데이터 포인트가 발생하는지(예를 들면 비정상)를 결정하기 위해 하나의 컷오프를 포함할 수 있고 제2 컷오프는 데이터 포인트가 측정 또는 분석에서 오차에 의해 쉽게 야기되는지(예를 들면, 구분치가 예상되는 것보다, 심지어 이환 샘플에 대하여 더 큰지)를 결정하기 위해 사용될 수 있다.

[0113] 서열 불균형 (예를 들면 비정상)이 제1 게놈 위치에 대하여 존재하는지의 분류는 비교에 기반하여 결정된다. 하나의 구현예에서, 복수의 컷오프 (예를 들면 N 컷오프)는 단일 구분치에 대하여 사용될 수 있다. 그와 같은 구현예에서, N+1 분류는 결정될 수 있다. 예를 들어, 2개 컷오프는 염색체 영역이 정상 또는 건강한, 불확정, 또는 비정상인(예를 들면 증폭 또는 결실)지의 분류를 결정하기 위해 사용될 수 있다. 복수의 비교가 수행되는 또 다른 구현예에서 (예를 들면 각 구분치에 대하여 하나), 분류는 각각의 비교에 기반될 수 있다. 예를 들어, 규칙-기반 방법은 각각의 비교로부터 비롯된 분류를 조사할 수 있다. 하나의 실행에서, 최종적인 분류는 모든 분류가 일치된 경우만 제공된다. 또 다른 실행에서, 다수 분류가 사용된다. 더욱 또 다른 실행에서, 더욱 복잡한 식은 각각의 구분치가 각자의 컷오프 값에 얼마나 근접한지에 기반하여 사용될 수 있고, 이들 근접 값은 분류를 결정하기 위해 분석될 수 있다. 예를 들어, 근접 값은 (다른 인자, 예컨대 정규화와 함께) 합계될 수 있고 결과는 또 다른 컷오프 값과 비교될 수 있다. 다른 구현예에서, 방법 900의 변동은 또한, 참조 샘플로부터 유도될 수 있는, 컷오프 값과 제1 염색체에 대하여 통계치의 직접적인 비교에 적용될 수 있다.

[0114] B. 암과 크기의 상관관계

[0115] 추가 분석을 위하여, 우리는 3개의 상이한 크기 그룹, 즉, 150 bp 미만인 것, 150과 180 bp 사이인 것, 및 180 bp 초과인 것의 혈장 DNA 분자를 개별적으로 탐구하였다. 150 bp 미만 DNA 단편의 비율과 혈장내 종양 DNA 분획 사이 양성 상관관계 (피어슨  $r = 0.6$ ; p-값  $< 0.001$ )가 있다 (도 10A). 도 10A, 10B, 및 10C에서 종양 DNA 분획은 대수 규모로 보여진다. 150과 180 bp 사이 크기를 가진 DNA 단편과 혈장내 종양 DNA 분획 사이 상관관계 ( $r = -0.07$ ; p-값 = 0.95)는 관측되지 않았다 (도 10B). 180 bp 초과 DNA의 비율과 혈장내 종양 DNA 분획 사이 음성 상관관계 ( $r = -0.41$ ; p-값  $< -0.001$ )가 관측되었다 (도 10C).

[0116] 더 낮은 종양 DNA 분획은 암의 초기에서 더 많이 발생할 것 같고, 더 높은 종양 DNA 분획은 암의 후기에서 더 많이 발생할 것 같다. 따라서, DNA 단편에 대하여 정상보다 더 큰 평균 크기 (또는 다른 통계치)의 존재는 초기 암을 나타낼 수 있고, DNA 단편에 대하여 정상보다 더 작은 평균 크기의 존재는 후기 암을 나타낸다.

[0117] 다른 구현예에서, 종양 DNA 분획은 측정될 수 있다. 종양 DNA 분획이 특정 한계값 미만인 경우, 크기 분석은 크기 분포의 통계치가 한계값 초과인지(즉, DNA 단편이 긴지의 시험)를 결정하기 위해 수행될 수 있다. 종양 DNA 분획이 특정 한계값 초과인 경우, 크기 분석은 크기 분포의 통계치가 한계값 미만인지(즉, DNA 단편이 짧은지의 시험)를 결정하기 위해 수행될 수 있다.

[0118] 크기 분석의 방법 및 암과 크기의 관계에 관한 데이터는 2011년 11월 30일 출원된 Lo 등에 의해 발명의 명칭 "Detection of Genetic or Molecular Aberrations Associated with Cancer"인 미국 특허 공개 번호

2013/0040824에서 논의되고, 이의 내용은 모든 목적으로 본원에서 참고로 편입된다.

[0119] **V. 크기 분석을 이용한 CNA 비정상 확증**

[0120] 우리는 단일 염기 분석에서 및 게놈전체 방식으로 혈장 DNA 샘플의 크기 프로파일을 연구하기 위해 대규모 병렬형 서열분석을 사용하였다. 우리는 그의 특이적 크기 프로파일 연구를 위하여 종양-유도된 혈장 DNA를 확인하기 위해 CAZA를 사용하였다.

[0121] 상기 연구에서, 우리는 종양-관련 CNA의 존재를 시사하는 혈장 DNA 정량적 비정상을 보여준 염색체 아암을 확인하기 위해 CAZA 접근법을 사용하였다. 증폭 또는 결실을 가진 염색체 아암 확인 이후, 우리는 (증폭된 영역에서 풍부한) 종양-유도된 및 (결실된 영역에서 풍부한) 비-종양 유도된 혈장 DNA를 비교하기 위한 전략으로서 이들 영역에 집중하였다. 우리는 상기 접근법이 암-관련 돌연변이의 검출에 기반된 것보다 크기 프로파일링 분석에 대하여 종양성 DNA를 확인하기 위해 더욱 강력한 수단을 제공할 수 있다고 믿는다. 후자에 대하여, 평균적으로, 암 게놈 (29-32, 39)에서 수 천의 점 돌연변이 정도가 보고되었다. CAZA에 대하여, 다른 한편으로, 수십의 메가 염기에 관하여 총합하는, CNAs를 나타내는 게놈 영역으로부터 유도된 임의의 무수한 혈장 DNA 분자는 유용할 것이다.

[0122] **A. 조합된 분석**

[0123] 도 11은 암 환자에서 혈장 DNA 크기 분석의 원리의 도식적 예시를 보여준다. 도 11은 병기 1110-2150을 보여준다. 병기 1110은 혈장내 조직의 세포를 보여준다. 종양 세포는, 상기 기재된 바와 같이, 다양한 영역에서 증폭 및/또는 결실을 포함할 수 있다. 예는 특정한 염색체 상에서 증폭된 하나의 영역 및 결실된 또 다른 영역을 보여준다.

[0124] 병기 1120에서, 혈장은 다양한 영역으로부터 기여로 보여진다. DNA 단편은 혈장 샘플에서 보여진다. 암 환자에서, 혈장 DNA는 종양 (적색 분자) 및 비-종양 세포 (청색 분자) 모두로부터 유도된다. 종양 조직에서 증폭되는 게놈 영역은 더 많은 종양성 DNA를 혈장에 기여할 것이다. 종양 조직에서 결실되는 게놈 영역은 더 적은 DNA를 혈장에 기여할 것이다.

[0125] 병기 1130에서, 쌍으로 된-말단 서열분석이 수행된다. 쌍으로 된-말단 서열분석은 혈장 샘플에서 DNA 단편의 크기를 결정하기 위해 사용될 수 있다.

[0126] 병기 1140에서, 카운트-기반 분석은 비정상 영역을 확인하기 위해 사용된다. 보여진 예에서, CAZA 분석은 염색체 아암이, 종양내 염색체 아암의 증폭 또는 결실의 존재를 시사하는, 혈장 DNA에서 과대- 또는 과소-표현되는 지를 결정하기 위해 사용되었다. 큰 양성 z-스코어는 염색체 아암의 증폭의 존재를 나타낼 수 있고, 반면에 큰 음성 z-스코어는 염색체 아암의 결실의 존재를 나타낼 수 있다. 영역의 다른 크기는 아암 이외에 사용될 수 있다.

[0127] 병기 1150에서, 시험 영역의 크기 분포는 분석될 수 있다. 상기에 설명한 바와 같이, 종양 DNA 단편은 건강한 세포의 DNA 단편보다 더 짧다. 비정상 영역의 DNA 단편은 크기 분석이 또한 동일한 비정상을 보여주는지를 확증하기 위해 시험될 수 있다. 보여진 예에서, 증폭을 나타내는 영역의 크기 분포는 결실을 나타내는 영역의 크기 분포와 비교된다. 따라서, 일부 구현예에서, (비-종양 DNA가 풍부한) 과소-표현되는 및 (종양-유도된 DNA가 풍부한) 과대-표현되는 염색체 아암에서 유래하는 혈장 DNA 분자의 크기 프로파일은 아래 더 상세히 기재된 바와 같이 비교될 수 있다.

[0128] **B. 2개 영역 사이 크기 차이**

[0129] 종양 및 비-종양 조직에서 유래하는 혈장 DNA의 크기 프로파일을 비교하기 위해, 우리는 CNAs를 가진 염색체 아암으로부터 혈장 DNA 단편을 분석하였다. 이전의 연구 (34-36) 뿐만 아니라 상기 연구에서 우리의 발견에 기반하여, HCC와 관련된 전형적인 CNAs는 1p 및 8p 결실, 및 1q 및 8q 증폭을 포함한다. 혈장에서 53% 종양-유도된 DNA를 가진 HCC 케이스 (H291)는 원리를 예시하기 위해 사용된다. 상기 케이스는 혈장에서 8p 결실 및 8q 증폭을 보여주었다. 따라서, 종양은 8p의 결실된 영역보다 8q의 증폭된 영역으로부터 더 많은 혈장 DNA를 방출할 것이다. 그 결과, 8q는 종양-유도된 DNA가 상대적으로 풍부할 것이고 8p는 CNA 없는 영역과 비교하여 종양 DNA가 상대적으로 결실될 것이다 (또는 환언하면, 비-종양 DNA가 상대적으로 풍부할 것이다). 8p 및 8q에 대하여 혈장 DNA의 크기 프로파일은 도 12A에서 보여진다. 8q에 대한 크기 프로파일은 8p에 대한 것의 좌측상에 있었고, 8q에 대한 혈장 DNA의 크기 분포가 8p에 대한 것보다 더 짧다는 것을 나타낸다. 8q가 종양 DNA로 풍부하기 때문에, 데이터는 종양에 의해 방출된 DNA가 종양에서 유래하지 않은 DNA보다 더 짧다는 경향을 시사한다.

[0130] 단축의 정도를 정량화하기 위해, 8p 및 8q에 대하여 크기 프로파일용 누적 빈도 플롯 (도 12B)은 각각의 혈장 샘플에 대하여 작제되었다. 이들 플롯은, 샘플내 모든 혈장 DNA 분자의 비율로서, 짧은 크기부터 긴 크기까지, DNA 분자의 진행성 축적을 보여준다. 2개 곡선  $\Delta S$ 에서 차이(도 12C)는 그 다음 하기와 같이 계산되었다:

$$[0131] \Delta S = S_{8q} - S_{8p}$$

[0132] 식 중  $\Delta S$ 는 특정한 크기에서 8p와 8q 사이 누적 빈도에서 차이를 나타내고,  $S_{8p}$  및  $S_{8q}$ 는 8p 및 8q, 각각 상에서 특정한 크기 미만 혈장 DNA 단편의 비율을 나타낸다. 특정한 크기에 대하여  $\Delta S$ 의 양성 값은 8p와 비교된 8q 상에서 그 특정한 크기보다 더 짧은 DNA의 더 높은 존재도를 나타낸다. 상기 방법을 이용하여, 우리는 혈장내 8p 및 8q 상에서 CNAs를 나타낸 모든 HCC 케이스에 대하여 50 bp 내지 250 bp로  $\Delta S$  값을 스캔하였다. HCC 케이스 H291에 대하여 8q와 8p 사이 누적 빈도에서의 차이,  $\Delta S$ 는 도 12C에서 적색선으로서 작도된다. 건강한 대조군 (회색선)과 비교로, 모든 이들 HCC 케이스는 (비-종양 DNA가 풍부한) 8p보다 (종양 DNA가 풍부한) 8q에서 유래하는 200 bp 보다 더 짧은 혈장 DNA의 더 높은 존재도를 보여주었다 (도 13A). 도 13A는 혈장내 8p 및 8q 상에서 상이한 CNAs를 가진 모든 HCC 케이스에 대하여 크기에 대한  $\Delta S$ 의 플롯을 보여주었다. 혈장에서 단편적 종양 DNA 농도의 상이한 범위를 가진 케이스는 상이한 색상으로 보여진다. 단편적 종양 DNA 농도가 증가함에 따라,  $\Delta S$ 는 증가하고, 더 짧은 DNA 단편의 더 높은 존재도를 나타낸다. 이들 데이터는 추가로 종양-유도된 DNA가 비-종양 유도된 DNA의 것보다 더 짧았다는 것을 뒷받침한다.

[0133]  $\Delta S$ 의 값은 종양 및 비-종양 조직으로부터 유도된 혈장 DNA 사이 주요 차이가 DNA < 166 bp 및  $\geq 166$  bp의 상대 존재비인 것을 시사하는 166 bp에서 최대값을 획득하였다. 우리는  $\Delta S_{166}$ 으로서 상기 값을 나타낸다.  $\Delta S_{166}$ 은, HBV 캐리어 및 간경변증을 가진 환자를 포함한, 상기 연구의 모든 대상체에 대하여 작도되었다 (도 13B). HCC 그룹에 대하여, 혈장 CAZA 분석에 의해 결정된 바와 같이 8p 및 8q 상에서 상이한 CNAs 있는 환자 및 없는 환자는 적색 및 흑색 점, 각각으로 나타낸다. 거의 모든 비-HCC 대상체에 대하여,  $\Delta S_{166}$  값은 8p 및 8q로부터 DNA에 대하여 크기 분포가 유사했던 것을 나타내는 0에 근접하였다.  $\Delta S_{166}$  (또는 일부 다른 특정한 크기에서 값)은 한계값과 비교될 수 있고, 차이가 한계값을 초과하면, 그 영역의 적어도 하나는 비정상을 나타내는 것으로서 확인될 수 있다. 하나의 영역이 (예를 들면, CNA 분석으로부터) 비정상 값을 갖지 않는 것으로 공지되면, 다른 영역은 차이가 한계값을 초과하는 경우 비정상 값을 나타내는 것으로서 확인될 것이다. 그와 같은 구현예에서, 차이의 징후는 비정상 유형을 나타낼 수 있다. 예를 들어, 제1 영역이 증폭을 갖고 제2 영역이 갖지 않는 경우, 차이는 양수가 될 것이다. 제1 영역이 결실을 갖고 제2 영역이 갖지 않는 경우, 차이는 음수가 될 것이다. 비정상이 결정되면, 영역 모두는, 각 영역이 가질 수 있는 비정상 유형을 나타내는 징후로, 비정상을 잠재적으로 갖는 것으로서 확인될 수 있다. 차이가 충분히 크면, 하나의 영역이 증폭을 갖고 다른 영역이 결실을 갖는(또는 증폭의 양이 상이한) 것을 나타낼 수 있고, 그때부터 차이는 정상 영역과 비교된 증폭된 영역보다 더 클 것이다. 카피 수 분석은 영역에 초기 분류를 제공할 수 있고, 이로써 적합한 한계값이 선택될 수 있다.

[0134] 1p 및 1q의 혈장 DNA 크기 프로파일에 기반된 크기 분석은 또한 수행되었고 (도 14 및 15) 동일한 추세를 보여주었다. 도 15에서, HCC 그룹에 대하여, 혈장 CAZA 분석에 의해 결정된 바와 같이 1p 및 1q 상에서 상이한 CNAs 있는 및 없는 환자는 적색 및 흑색 점, 각각으로 나타낸다. 상기 크기 분석은 정상 영역에서, 또는 정상 영역 및 결실된 영역에서 증폭된 영역을 이용하여 수행될 수 있다.

[0135] 또 다른 구현예에서, 증폭된 또는 결실된 영역에 대하여 크기 분포는 암을 갖는 것으로 공지된 또는 건강한 것으로 공지된 하나 이상의 기준 대상체의 크기 분포와 비교될 수 있다. 크기 분포는 값, 예를 들면, 통계치, 예컨대 평균 또는 중앙 크기로 나타낼 수 있다.

[0136] 따라서, 염색체 영역의 비정상은 크기 분석을 위하여 특정한 영역을 선택하기 위해 사용될 수 있다. 선택된 영역의 크기 분석은 그 다음 암의 수준의 분류를 결정하기 위해 사용될 수 있다. CNA 및 크기 분석 이용의 조합은 더 큰 정확도를 제공할 수 있다. CNA 분석은 가끔 위 양성, 즉, 암을 갖지 않지만 카피 수 비정상 값을 가진 영역을 갖는 환자를 산출할 수 있다. 따라서, 비정상을 나타내는 영역의 충분한 수 때문에 암을 갖는 것으로 확인된 환자는 그 다음 크기 분석을 이용하여 확인될 수 있다. 하나의 구현예에서, 선택된 영역은 증폭을 갖는 것이다.

[0137] 상기 연구는 종양 조직에 의해 혈장 DNA의 생산 또는 방출에 관련된 기전에 빛을 발할 수 있는 고분해능 및 포괄적인 방식으로 HCC 환자의 혈장 DNA 크기 프로파일을 탐구하기 위한 의도로 설계되었다. 연구의 또 다른 목표는 암-관련 혈장 DNA 크기 프로파일에 관한 문헌에 존재한 명백한 불일치의 일부를 해결하는 것이었다. 연구는

암 환자 (20-23)의 혈장에서 더 긴 DNA의 존재를 보고하였고 반면에 다른 것은 더 짧은 혈장 DNA 분자 (12, 25) 중에서 암-관련 DNA 돌연변이의 더 높은 유병률을 보고하였다. 이들 연구 목표를 달성하기 위해, 2-단계 접근법이 채택되었다. 첫째, 우리는 쌍으로 된-말단 대규모 병렬형 서열분석의 이용과 함께 모집된 대상체의 혈장 샘플에서 모든 DNA 분자의 길이를 측정하였다. 상기 접근법은 개별적인 혈장 DNA 분자 최대 단일 염기 분석의 길이를 측정하게 한다. 더욱이, 계놈을 거쳐 혈장 DNA 분자는 분석될 수 있고 상이한 크기의 DNA 사이 상대 양은 높은 정확도로 결정될 수 있다. 그러므로, 혈장 DNA 크기 프로파일의 넓은 및 깊은 조사는 수득될 수 있다. 둘째, 우리는 상세한 분석을 위하여 종양-유도된 혈장 DNA를 확인하기 위한 수단으로서, CAZA 접근법을, 증폭 또는 결실과 관련된 계놈 위치에서 유래하는 혈장 DNA내 종양성 DNA 함량에서 상대적 차이의 이점으로 가졌다.

[0138] 상기 연구는 혈장 DNA의 방출에 관여될 생물학적 기전에 수많은 통찰력을 제공한다. HBV 캐리어, 간경변증 또는 HCC를 가진 환자를 포함한, 모든 모집된 대상체의 혈장 DNA는 166 bp에서 두드러진 피크를 나타냈다 (도 14 및 16). 상기 패턴은 임신한 여성 및 장기 이식 수령체의 혈장에서 관찰과 유사하다 (26, 27). 연구된 환자의 모든 그룹의 혈장 DNA 크기 프로파일에서 특징적인 166 bp 피크의 존재는, 임신한 여성의 것, 이식 수령체, HCC, 간경변증 또는 만성 HBV를 가진 환자를 포함한, 인간 혈장내 대부분의 순환형 DNA 분자가 모노뉴클레오솜 단위를 닮고 세포자멸사의 과정에서 유래할 것 같다는 것을 시사한다.

[0139] 종양-관련 CNAs를 보유한 혈장 DNA 분자의 크기 프로파일의 연구는 상기 분자가 상기 특징을 담지하지 않는 것보다 더 짧다는 것을 나타낸다 (도 13). 이는 혈장내 종양 DNA의 단편적 농도 증가와 함께 혈장 DNA의 크기 프로파일이 좌측으로 이동할 것이라는 우리의 관찰과 일치한다. 그러나, 혈장내 종양 DNA의 낮은 단편적 농도를 가진 HCC 환자가 건강한 대조군보다 명백하게 더 긴 크기 분포를 가졌다는 사실은 종양-관련 계놈 특징을 담지하지 않았던 혈장 DNA의 추가의 구성요소가 있었다는 것을 시사한다. 상기 구성요소가 종양을 둘러싸는 비-신생물성 간 조직으로부터 유도될 것이라는 것이 가능하다. 이들 긴 DNA 분자는 세포자멸사 대신 괴사로부터 유도될 수 있다. 조직 괴사와 관련된 세포 사멸이 전형적인 올리고뉴클레오솜 DNA 단편 (37, 38)에 더하여 더 긴 DNA 단편을 생성할 수 있다는 것이 보고되었다. 미래의 연구를 위하여, 이들이 간에 대하여 예상된 것에 유사함을 보유하고는 있는지 보기 위해 이들 더 긴 DNA 분자의 DNA 메틸화 프로파일을 연구하는 것이 흥미로운 것이다.

[0140] 우리는 비정상적으로 짧은 및 긴 DNA 분자의 집단이 간세포 암종을 가진 환자의 혈장에서 공존하였다는 것을 보여주었다. 짧은 것이 종양-관련 카피 수 비정상성을 우선적으로 담지하였다.

[0141] 요약하면, 우리는 단일-뉴클레오타이드 분해능에서 HCC를 가진 환자내 혈장 DNA의 크기 분포를 프로파일링하였다. 우리는 종양 및 비-종양 조직으로부터 유도된 혈장 DNA의 크기에서 차이를 입증하였다.

[0142] ΔS와 종양 크기 사이 관계가 또한 분석되었다. 혈장에서 8p 결실 및 8q 증폭을 가진 10 HCC 환자의 혈장 DNA 샘플은 ΔS 분석을 이용하여 분석되었다. ΔS는 8p 및 8q에 대한 혈장 DNA 단편 맵핑 사이 크기 차이에 대하여 결정되었다. ΔS에 대하여 양성 값은 8p와 비교된 8q에 대하여 150 bp 미만 짧은 DNA 단편의 더 많은 존재도를 나타낸다. 도 30에서, ΔS의 값은 HCC 환자의 종양의 가장 긴 치수에 대해 작도되었다.

[0143] ΔS와 종양 크기 사이 양성 상관관계는 관측되었다 ( $r=0.876$ , 피어슨 상관관계). 상기 관찰은 CNAs의 상이한 유형을 나타내는 영역으로부터 혈장 DNA 단편의 크기 분포가 HCC 환자에서 종양의 크기를 반영하기 위해 사용될 수 있다는 것을 시사한다.

[0144] 총 혈장 DNA의 전체 크기 분포는 이들 10 HCC 환자에 대하여 또한 분석되었다. 150 bp 미만 ( $P<150$ )의 혈장 DNA 단편의 백분율은 각 케이스에 대하여 결정되었고 도 31에서 종양 크기에 대해 작도되었다. 짧은 단편의 비율은 최대 치수로 3 cm 초과와 더 큰 암을 가진 환자에서 유의미하게 더 높았다. 하나의 구현예에서, 짧은 단편의 비율은 암의 크기 및 중증도를 반영하기 위해 사용될 수 있다. 다른 실행에서, 크기에 대한 다른 컷오프는, 예를 들어, 비제한적으로 100 bp, 110 bp, 120 bp, 130 bp, 140 bp, 160 bp 및 166 bp가 사용될 수 있다.

[0145] 교정 함수는 종양의 크기와 통계치 사이 관계를 제공하기 위해 사용될 수 있다. 교정 함수는 공지된 크기의 종양을 가진 생물에서 참조 샘플의 교정 데이터 포인트로부터 결정될 수 있다. 교정 데이터 포인트는 종양의 크기의 측정 및 염색체 영역으로부터 핵산 분자의 크기의 대응하는 통계적 측정을 포함할 수 있다. 신규 샘플이 신규 대상체로부터 수득되는 경우, 통계치는 결정될 수 있고, 교정 함수는 통계치를 종양 크기로 전환시키기 위해 사용될 수 있다. 교정 함수의 예는 도 30에서 보여준 선형 적합과 유사한 선형 적합이다. 퇴행 분석의 다른 유형, 예컨대 최소 자승 적합이 교정 함수를 생성하기 위해 사용될 수 있다.

[0146] 교정 함수는 다양한 방식으로, 예를 들면, 특정된 함수, 예컨대 선형 또는 비-선형 함수의 복수의 계수로서 정의되어야 한다. 다른 구현예는 복수의 교정 데이터 포인트 (예를 들면, 교정 함수의 데이터 포인트)를 저장할

수 있어서 이로써 교정 함수가 생성될 수 있다. 추가로, 보간이 상기 교정 데이터 포인트 사이에 수행되어 교정 함수를 획득할 수 있다. 교정 함수는 컴퓨터 메모리에 저장될 수 있고 이로부터 검색될 수 있다.

- [0147] C. 방법
- [0148] 도 16은 본 발명의 구현예에 따라 생물의 생물학적 샘플을 분석하기 위해 CAZA 및 크기 분석 수행의 방법 1600을 예시하는 흐름도이다.
- [0149] 단계 1605에서, 생물의 복수의 염색체 영역은 확인될 수 있다. 각 염색체 영역은 복수의 유전자자리를 포함할 수 있다. 복수의 염색체 영역 중 하나는 제1 염색체 영역으로서 선택될 수 있다. 복수의 염색체 영역 확인은 도 6의 단계 610과 유사할 수 있다.
- [0150] 단계 1610에서, 생물의 표준 계놈에서 핵산 분자의 위치는 각각의 복수의 핵산 분자에 대하여 확인될 수 있다. 핵산 분자의 위치 확인은 도 6의 단계 620과 유사한 방식으로 수행될 수 있다.
- [0151] 단계 1615에서, 핵산 분자의 크기는 생물학적 샘플에서 각각의 복수의 핵산 분자에 대하여 측정될 수 있다. 핵산 분자의 크기는 도 9의 단계 910과 유사하게 측정될 수 있다.
- [0152] 단계 1620에서, 핵산 분자의 각각의 그룹은, 복수의 염색체 영역의 각 염색체 영역에 대하여 염색체 영역 기원인 것으로서, 확인된 위치에 기반하여, 확인될 수 있다. 각각의 그룹은 염색체 영역의 각각의 복수의 유전자자리에 위치한 적어도 하나의 핵산 분자를 포함할 수 있다. 핵산 분자의 각각의 그룹의 확인은 도 1의 단계 120과 유사할 수 있다.
- [0153] 단계 1625에서, 컴퓨터 시스템은 핵산 분자의 각각의 그룹의 각각의 양을 계산할 수 있다. 각각의 양 계산은 도 1의 단계 130에서 계산과 유사할 수 있다.
- [0154] 단계 1630에서, 각각의 양은 염색체 영역이 증폭을 나타내는지의 카운트 분류를 결정하기 위해 카운트 기준치와 비교될 수 있다. 비교에 기반하여, 제1 염색체 영역은 비정상적으로 나타내는 것으로서 확인될 수 있다. 단계 1620-1630은 도 1의 단계 120-140 또는 도 6의 단계 630-650과 유사한 방식으로 수행될 수 있다.
- [0155] 단계 1640에서, 핵산 분자의 제1 그룹은 제1 염색체 영역 기원인 것으로서 확인될 수 있다.
- [0156] 단계 1645에서, 컴퓨터 시스템은 핵산 분자의 제1 그룹의 제1 크기 분포의 제1 통계치를 계산할 수 있다. 제1 통계치는 특정된 크기에서 제1 곡선하 면적의 컴퓨팅에 의해 결정될 수 있다. 제1 곡선은 크기의 범위에 걸쳐 제1 염색체 영역에 대하여 핵산 분자의 누적 빈도의 플롯일 수 있다. 단계 1645에서 제1 통계치의 계산은 도 9 내 단계 940에서 제1 통계치 계산과 유사할 수 있다.
- [0157] 단계 1650에서, 제1 통계치는 제1 염색체 영역이 비정상적으로 나타내는지의 크기 분류를 결정하기 위해 크기 기준치와 비교될 수 있다. 크기 기준치는 특정된 크기에서 제2 곡선하 면적의 컴퓨팅에 의해 결정될 수 있다. 제2 곡선은 크기의 범위에 걸쳐 제2 염색체 영역에 대하여 핵산 분자의 누적 빈도의 플롯일 수 있다. 비교는 2개 곡선 사이 차이에 기반될 수 있다. 일부 구현예에서, 제1 통계치의 크기 기준치와의 비교는 도 9에서 단계 950과 유사할 수 있다.
- [0158] 단계 1655에서, 제1 염색체 영역이 비정상적으로 나타내는지의 최종 분류는 결정될 수 있다. 예를 들어, 크기 분류 및 카운트 분류의 적어도 하나는 비정상이 제1 염색체 영역에 대하여 존재하는지를 결정하기 위해 사용될 수 있다. 일부 구현예에서, 최종 분류는 카운트 분류 및 크기 분류가 동일한 비정상적으로 나타내는 경우에만 제1 비정상이 존재한다는 것일 수 있다. 따라서, 제1 통계치의 크기 기준치와의 비교는 제1 염색체 영역이 비정상적으로 나타내는지를 확증할 수 있다. 일부 구현예에서, 크기 분류의 세트는 대응하는 카운트 분류에 기반된 비정상적으로 확인된 염색체 영역의 세트에 대하여 결정될 수 있다. 크기 분류의 세트에 기반하여, 각각의 염색체 영역은 비정상인 또는 비정상이 아닌 것으로 확증될 수 있다.
- [0159] 일부 구현예에서, 제1 염색체 영역이 비정상적으로 나타내는지의 최종 분류는 다중 카운트 기준치 및 다중 크기 기준치에 기반될 수 있다. 각각의 카운트 기준치는 상이한 카운트 분류 (예를 들면, 카운트 분류의 고유 쌍 사이, 예컨대 수준 1과 수준 2 사이, 또는 수준 2와 수준 3 사이 식별력)에 대응할 수 있다. 유사하게, 각각의 크기 기준치는 상이한 크기 분류에 대응할 수 있다. 최종 분류는 크기 분류 및 카운트 분류의 특별한 조합으로부터 결정될 수 있다.
- [0160] 크기 분류는 크기 분포의 통계치에 의존하는 다중 분류를 포함할 수 있다. 예를 들어, 통계치와 크기 기준치 사이 큰 차이는 비정상의 높은 가능성에 대응하는 크기 분류를 초래할 수 있고, 반면에 통계치와 크기 기준치 사

이 작은 차이는 비정상의 낮은 가능성에 대응하는 크기 분류를 초래할 수 있다. 유사하게, 카운트 분류는 핵산 분자의 그룹의 양에 의존하는 다중 분류를 포함할 수 있다. 예를 들어, 카운트 기준치에 비교된 핵산 분자의 그룹의 양 사이 큰 차이는 비정상의 높은 가능성에 대응하는 카운트 분류를 초래할 수 있고, 반면에 작은 차이는 비정상의 낮은 가능성에 대응하는 카운트 분류를 초래할 수 있다.

[0161] 따라서, 최종 분류는 상이한 크기 분류 및 카운트 분류에 대하여 상이한 한계값에 기반될 수 있다. 예를 들면, 비정상의 높은 가능성을 나타내는 크기 분류는 비정상의 특정한, 가능하게는 낮은, 가능성을 나타내는 카운트 분류가 주어진 비정상을 나타내는 최종 분류를 초래할 수 있다. 크기 분류 또는 카운트 분류의 하나로 제시된 바와 같이 비정상의 가능성이 증가함에 따라, 다른 분류에 의해 제시된 가능성에 대한 한계값은 낮아진다. 일부 케이스에서, 하나의 분류는 비정상의 제1 유형의 높은 가능성을 보여줄 수 있고, 다른 분류는 비정상의 제2 유형의 낮은 가능성을 보여줄 수 있고, 최종 분류는 비정상의 제1 유형이 존재한다는 것을 나타낼 수 있다. 일부 케이스에서, 최종 분류는 비정상의 가능성 또는 개연성에 대응할 수 있다.

[0162] *D. 예 케이스*

[0163] 암-관련 CNA의 검출의 특이성은, 하기 2 케이스에 나타난 바와 같이, 혈장 DNA 크기 분석에 의해 개선될 수 있다. 케이스 1은 B형 간염-관련 간경변을 가진 환자였고, 케이스 2는 B형 간염 감염의 만성 보균자이었다. 이들 모두는 모집시에 임의의 암을 갖는 것으로 공지되지 않았다. 이들은 모집 및 암 없음이 검출된 이후로 2년 동안 임상적으로 추적되었다. 정맥 혈액은 모집에서 각각의 2 대상체로부터 수집되었다. 혈장 DNA는 서열분석되었다. 염색체 1q를 연루한 CNA는 각각의 이들 2 환자에서 검출되었다. 케이스 1에 대하여, 1p 및 1q에 대한 z-스코어는 -2.3 및 15.5, 각각이었다. 이들 결과는 1q 증폭의 해석과 일치한다. 혈장 DNA 단편 크기 분석에서, ΔS는 -0.019이었다. ΔS의 음성 값은 짧은 DNA 단편이 1p와 비교된 1q에서 덜 풍부한 것을 나타낸다. 카운트-기반 분석이 1q가 증폭되었다는 것을 시사함에 따라, 크기-기반 분석 결과는 우리가 암-관련 CNAs에 대하여 예상했던 것과 반대이다. 암 환자에서, 카피 수 이득을 가진 영역은 증폭을 가진 영역 또는 임의의 CNA 없는 영역과 비교된 더 많은 암-유도된 짧은 단편의 존재 때문에 전반적인 더 짧은 크기 분포를 보여주는 것이 예상된다. 따라서, 상기 케이스에서 크기 분석은 혈장 DNA에서 암-관련 CNAs의 존재를 시사하지 않는다.

[0164] 케이스 2에 대하여, 1p 및 1q에 대한 z-스코어는 0.4 및 -4.4, 각각이었다. 이들 결과는 1q 결실의 해석과 양립 가능하다. 혈장 DNA 단편 크기 분석에서, ΔS는 0.044이었다. ΔS의 양성 값은 짧은 DNA 단편이 1p와 비교된 1q에서 더욱 풍부하였음을 나타낸다. 카운트-기반 분석이 1q가 결실되었다는 것을 시사함에 따라, 크기-기반 분석 결과는 우리가 암-관련 CNAs에 대하여 예상했던 것과 반대이다. 암 환자에서, 카피 수 손실을 가진 영역은 증폭을 가진 영역 또는 임의의 CNA 없는 영역과 비교된 더 적은 암-유도된 짧은 단편의 존재 때문에 전반적인 더 긴 크기 분포를 보여주는 것이 예상된다. 따라서, 상기 케이스에서 크기 분석은 혈장 DNA에서 암-관련 CNAs의 존재를 시사하지 않는다.

[0165] **VI. 암의 병기의 결정**

[0166] 상기에서 언급된 바와 같이, DNA 단편의 크기는 암의 병기를 나타낼 수 있다. 암의 후기는 증폭을 나타내는 영역에 대하여 더 작은 단편을 나타낸다.

[0167] 고유 생물학적 관심과 별개로, 혈장 DNA 크기 프로파일링은 또한 혈장에서 암-관련 변화 검출을 위하여 진단 접근법의 개발에 유용할 수 있다. 예를 들어, 혈장으로부터 종양성 DNA의 풍부는 짧은 DNA 단편의 분석에 집중함으로써 달성될 수 있다. 또한, 우리는 짧은 DNA 분자의 비율이 혈장에서 종양-유도된 DNA의 단편적 농도와 양성 관계를 보유하고 있다는 것을 관측하였다. 크기 프로파일에서 변화는 치료의 과정 동안 환자의 모니터링에 사용될 수 있다. 더욱이, HCC가 있는 및 없는 환자의 혈장에서 긴 DNA 분자의 집단의 존재는 추가 조사를 정당화한다. 이들 DNA 분자의 방출을 지배하는 조직 공급원 또는 병리적 과정이 더 양호하게 이해되는 경우, 혈장에서 긴 DNA의 비율 측정은 상기 질환의 진단에 유용할 수 있다.

[0168] *A. HCC 환자의 혈장 DNA 크기 분포*

[0169] HCC 환자, HBV 캐리어, 간경변 환자 및 건강한 대조군의 혈장 DNA의 크기 분포는 도 18 및 19에서 보여진다. 도 19에서, 각 개체는 상이한 색상으로 나타낸다. 일반적으로, 가장 두드러진 피크는 각 대상체의 크기 분포 플롯에서 166 bp에 관측되었다. 상기 관찰은 임신한 여성 및 이식 수령자 (26-28)에서 이전의 보고와 일치되고, 대부분의 순환형 DNA 분자가 세포자멸사로부터 유도되는 것을 시사한다. 흥미롭게도, 32 건강한 대조군 (도 18에서 두꺼운 흑색선)에 대하여 중앙 크기 분포 프로파일과 비교된 경우, 낮은 단편적 종양 DNA 농도를 가진 HCC 환자에서 혈장 DNA의 크기는 더 길었다. 그러나, 혈장에서 종양 DNA의 단편적 농도 증가와 함께, 혈장 DNA의 크

기 분포는 좌측으로 점진적으로 이동하였다 (도 18).

[0170] 초기에 기재된 바와 같이, 도 13A는 혈장내 8p 및 8q에서 상이한 CNAs를 가진 모든 HCC 케이스에 대하여 크기에 대한  $\Delta S$ 의 플롯이다. 혈장내 단편적 종양 DNA 농도가 2% 미만에서 8% 넘게 증가함에 따라,  $\Delta S$ 는 증가하고, 더 짧은 DNA 단편의 더 높은 존재도를 나타낸다. 혈장내 단편적 종양 DNA 농도는 암 진행의 병기로서 증가시킬 수 있다. 그 결과, 더 짧은 DNA 단편의 양은 암의 후기를 나타낼 수 있다. 도 13B는  $\Delta S_{166}$ 이 비-HCC 대상체와 비교로 HCC 환자에 대하여 더 높다는 것을 보여주고, DNA < 166 bp 및  $\geq 166$  bp의 상대 존재비가 암의 존재를 나타내기 위해 사용될 수 있다는 것을 나타낸다. 따라서,  $\Delta S_{166}$ 은 또한 암의 병기를 나타낼 수 있다.

[0171] 도 20은 짧은 단편의 비율이 HCC 환자를 건강한 대조군과 구분하기 위해 사용될 수 있는 경우의 예를 보여준다. 150 bp 미만 혈장 DNA 단편의 비율은 32 건강한 대상체, 혈장내 2% 미만의 종양 DNA 분획을 가진 HCC 환자 및 혈장내 6% 초과 종양 DNA 분획을 가진 HCC 환자에 대하여 작도되었다. ('CTR'로서 표지된) 건강한 대조군과 비교로, 2% 미만의 종양 DNA 분획을 가진 HCC 환자는 150 bp ( $p = 0.0002$ , t-시험) 미만의 짧은 DNA 단편의 유의미하게 더 낮은 비율을 가졌고, 6% 초과 종양 DNA 분획을 가진 환자는 짧은 단편 ( $p = 0.003$ , t-시험)의 유의미하게 더 높은 비율을 가졌다. 2% 내지 6%의 종양 DNA 분획을 가진 HCC 환자는 2% 미만의 종양 분획을 가진 HCC 환자와 6% 초과 종양 분획을 가진 HCC 환자 사이 DNA 단편의 비율을 갖는다. 이런 식으로, 2% 내지 6%의 종양 분획을 가진 HCC 환자는 건강한 대조군 대상체와 유사한 분포를 가질 수 있다.

[0172] 도 21은 2% 미만 종양 DNA 분획을 가진 HCC 환자를 건강한 대조군 대상체와 구분하기 위해  $P(<150)$  적용에 대하여 수신자 조작 특성 (ROC) 곡선을 보여준다. 종양 분획은 종양내 카피 수 손실과 양립가능한 혈장에서 과소-표현을 나타내는 염색체 영역의 과소-표현의 규모에 기반하여 결정되었다. 임의의 염색체 아암의 유의미한 과소-표현 없는 케이스에 대하여, 카피 수 이득과 양립가능한 영역에 대하여 과대-표현의 규모는 단일 카피 이득의 가정으로 종양 분획을 결정하기 위해 사용되었다. 종양 분획은 하기 방정식으로 결정될 수 있다:

$$\text{종양 분획} = \frac{|P_{\text{시험}} - P_{\text{정상}}|}{P_{\text{정상}} \times \Delta N / 2}$$

[0173] 식 중  $P_{\text{시험}}$ 은 시험 케이스에 대하여 당해 염색체 아암에 맵핑된 단편의 비율을 나타내고,  $P_{\text{정상}}$ 은 건강한 대조군에 대하여 염색체 아암에 맵핑된 단편의 평균 비율을 나타내고,  $\Delta N$ 은 카피 수 변화의 규모 (예를 들면, 중복 또는 결실에 대하여 1, 및 고차 중복에 대하여 더 높은 수)를 나타낸다. 곡선하 면적 (AUC)은 0.670 및 0.882의 95% 신뢰도 한계로 0.776이었다. 상기 결과는 크기 분석이 혈장내 2% 미만의 종양 분획을 가진 HCC 환자를 확인하기 위해 사용될 수 있다는 것을 나타낸다. ROC 곡선 분석은 상이한 한계값이 상이한 감수성 및 특이성을 달성하기 위해 선택될 수 있다는 것을 나타낸다.

[0175] 도 21과 유사한, 도 22는  $P(<150)$ 를 가진 크기 분석이 또한 혈장에서 6% 초과 종양 분획을 가진 HCC 환자를 검출할 수 있다는 것을 보여준다. 이들 환자의 건강한 대상체와의 식별을 위한 AUC는 0.761 및 1.000의 95% 신뢰도 한계로 0.893이었다.

[0176] 도 23은, 도 20이 짧은 혈장 DNA 단편의 비율을 보여주었듯이, 긴 혈장 DNA 단편의 비율이 HCC 검출을 위하여 사용될 수 있다는 것을 보여준다. 상기 예에서,  $P(>180)$ 로서 표시된, 180 bp 초과 단편의 비율은 혈장내 2% 미만 및 6% 초과 종양 DNA 분획을 가진 HCC 환자 및 건강한 대조군 대상체에 대하여 작도되었다. 상기 비율은 2% 미만 종양 DNA 분획 ( $p < 0.00001$ , t-시험)을 가진 HCC 환자에서 유의미하게 더 높았다.

[0177] 도 24는 2% 미만 종양 DNA 분획을 가진 HCC 환자를 건강한 대조군 대상체와 구분하기 위해  $P(>180)$  이용에 대하여 ROC 곡선을 보여준다. AUC는 0.805 및 0.961의 95% 신뢰도 한계로 0.883이었다.

[0178] 도 25는 상이한 종양 DNA 분획을 가진 DNA 단편의 상이한 크기 분포의 또 다른 예를 제공한다. 도 25는 건강한 대조군 대상체, 2% 미만 종양 DNA 분획을 가진 HCC 환자, 및 6% 초과 종양 DNA 분획을 가진 HCC 환자의 종양 단편 크기의 박스플롯을 보여준다. 2% 미만 종양 DNA 분획을 가진 HCC 환자의 DNA 단편의 종양 크기는 건강한 대조군 대상체보다 유의미하게 더 길었다 ( $P < 0.00001$ , t-시험). 그에 반해서, 6% 초과 종양 DNA 분획을 가진 HCC 환자의 DNA 단편의 종양 크기는 유의미하게 더 짧았다 ( $p = 0.03$ , t-시험). 도 25는 암의 병기를 결정하기 위한 방식으로서 DNA 단편 크기의 용도를 뒷받침한다. 더 긴 종양 크기는 더 작은 종양 DNA 분획과 관련되고, 반면에 더 짧은 종양 크기는 더 큰 종양 DNA 분획과 관련된다. 개체가 제1 컷오프 미만의 더 작은 종양 DNA 분획 및 긴 크기 한계값 초과 종양 크기를 갖는다면, 초기 암은 확증될 수 있다. 다른 한편으로, 개체가 제2 컷

오프 초과와 더 큰 종양 DNA 분획 및 짧은 크기 한계값 미만의 중앙 크기를 갖는다면, 후기 암은 확증될 수 있다.

- [0179] 2% 내지 6%의 종양 DNA 분획을 가진 HCC 환자는 2% 미만의 종양 분획을 가진 HCC 환자와 6% 초과 종양 분획을 가진 HCC 환자 사이 중앙 DNA 단편 크기를 갖는다. 이런 식으로, 2% 내지 6%의 종양 분획을 가진 HCC 환자는 도 25에서 건강한 대조군 대상체와 유사한 분포를 가질 수 있다. 그러므로, 개체가 낮은 컷오프 내지 높은 컷오프의 종양 DNA 분획 및 짧은 크기 한계값 내지 긴 크기 한계값의 중앙 크기를 갖는다면, 중기 암은 확증될 수 있다.
- [0180] 도 26 및 27은 상이한 크기 한계값이 HCC 환자를 건강한 대조군 대상체와 구분하기 위해 사용될 수 있다는 것을 보여주는 ROC 곡선이다. 도 26은 2% 미만 종양 DNA 분획을 가진 HCC 환자와 건강한 대조군 대상체 사이를 식별하기 위해 중앙 단편 크기 이용에 대하여 ROC 곡선이다. AUC는 0.718 및 0.907의 95% 신뢰도 한계로 0.812이었다.
- [0181] 도 27은 2% 초과 종양 DNA 분획을 가진 HCC 환자와 건강한 대조군 대상체 사이를 구분하기 위해 중앙 단편 크기 이용에 대하여 ROC 곡선이다. AUC는 0.627 및 0.963의 95% 신뢰도 한계로 0.795이었다.
- [0182] 크기 분포 (예를 들면, 중앙값, 평균, 백분위)의 다른 통계적 특징은 HCC 환자 및 건강한 대상체의 구분을 위하여 파라미터로서 사용될 수 있다.
- [0183] 모든 게놈 영역에서 발생하는 혈장 DNA 단편의 크기 분포 분석에 더하여, 크기 분석은 또한 특이적 게놈 영역에서 발생하는 DNA 단편에 집중할 수 있다. 특이적 게놈 영역은 염색체 아암일 수 있다.
- [0184] 도 28은 6% 초과 종양 DNA 분획을 가진 HCC 환자에 대하여 그리고 건강한 대조군 대상체에 대하여 염색체 1q에 정렬된 150 bp 미만의 짧은 혈장 DNA 단편의 비율의 박스플롯을 보여준다. 짧은 단편의 비율은 HCC 환자에서 유의미하게 더 높았다 ( $p < 0.00001$ , t-시험).
- [0185] 도 29는 6% 초과 종양 DNA 분획을 가진 HCC 환자와 건강한 대조군 대상체 사이를 구분하기 위해 150 bp 미만의 짧은 혈장 DNA 단편의 비율 이용에 대하여 ROC 곡선이다. AUC는 0.808 내지 1.000의 95% 신뢰 구간으로 0.915이었다.
- [0186] B. 방법
- [0187] 도 17은 본 발명의 구현예에 따라 생물의 생물학적 샘플 분석의 방법 1700을 예시하는 흐름도이다. 생물학적 샘플은 정상 세포 및 암과 관련된 세포에서 유래하는 핵산 분자를 포함할 수 있다. 핵산 분자의 적어도 일부는 생물학적 샘플에서 무세포이다.
- [0188] 단계 1710에서, 생물학적 샘플에서 각각의 복수의 핵산 분자에 대하여, 핵산 분자의 크기는 측정된다. 핵산 분자의 크기는 도 9의 단계 910과 유사하게 측정될 수 있다.
- [0189] 단계 1720에서, 생물의 표준 게놈에서 핵산 분자의 위치는 확인된다. 핵산 분자의 위치 확인은 도 6의 단계 620과 유사한 방식으로 수행될 수 있다.
- [0190] 단계 1730에서, 핵산 분자의 제1 그룹은 확인된 위치에 기반된 제1 염색체 영역 기원인 것으로서 확인된다. 제1 염색체 영역은 복수의 제1 유전자자리를 포함할 수 있다. 핵산 분자의 각자의 그룹의 확인은 도 1의 단계 120과 유사할 수 있다.
- [0191] 단계 1740에서, 컴퓨터 시스템은 핵산 분자의 제1 그룹의 크기 분포의 제1 통계치를 계산할 수 있다. 각자의 양 계산은 도 1의 단계 130에서 계산과 유사할 수 있다.
- [0192] 단계 1750에서, 암과 관련된 세포에서 유래하는 핵산 분자의 분획은 측정될 수 있다. 분획은 2011년 11월 30일 출원된 Lo 등에 의해 발명의 명칭 "Detection of Genetic or Molecular Aberrations Associated with Cancer"인 미국 특허 공개 번호 2013/0040824에서 기재된 방법에 따라 계산될 수 있다. 종양 핵산 분자의 분획은 종양 (들) 기원인 샘플에서 핵산 분자의 비율에 대응한다. 분획/비율은 임의의 백분율 또는 십진 값으로서 표시될 수 있다.
- [0193] 하기 예는 종양 핵산의 분획의 측정용 방법이지만 다른 방법이 사용될 수 있다. 종양 핵산의 분획은 종양 조직에서 카피 수 손실 (또는 카피 수 이득)과 양립가능한 유의미한 과소-표현을 나타내는 영역에 대하여 혈장에서 과소-표현 (또는 과대-표현)의 규모에 기반하여 결정될 수 있다. 또 다른 예는 카피 수 비정상에 의해 영향받은

영역, 예를 들면, 2 상동성 염색체의 1 카피의 손실을 가진 영역에 대하여 2 상동성 염색체에 관한 대립유전자 불균형의 정도를 결정하는 것이다. 또 다른 예는, 단일 뉴클레오타이드 돌연변이를 포함한, 암-관련 돌연변이, 뉴클레오타이드(들)의 결실, 및 전좌의 단편적 농도를 결정하는 것이다. 종양 분획은 상기 도 21로 기재된 방법에 의해 결정될 수 있다.

[0194] 단계 1760에서, 측정된 분획에 기반된 제1 기준치는 선택될 수 있다. 하나의 예에서, 제1 기준치 선택은 측정된 분획이 컷오프 미만인 경우 크기 한계값 선택을 포함할 수 있다. 또 다른 예에서, 제1 기준치 선택은 측정된 분획이 컷오프 초과인 경우 크기 한계값 선택을 포함할 수 있다. 이들 예에서, 컷오프 및 크기 한계값은 상이할 수 있고 측정된 분획의 값에 의존할 수 있다.

[0195] 단계 1770에서, 제1 통계치는 생물학적 샘플의 암의 병기를 결정하기 위해 제1 기준치와 비교될 수 있다. 제1 통계치는 본원에서 기재된 임의의 통계치일 수 있다.

[0196] 암이 존재하는지 여부는 암과 관련된 세포에서 유래하는 핵산 분자의 측정된 분획과 함께 크기 분석에 기반하여 확인될 수 있다. 예를 들어, 측정된 분획이 낮은 컷오프 미만인 경우, 크기 분포가 건강한 대조군에 대해서보다 더 긴지 (예를 들면, 제1 통계치가 크기 한계값 초과인지)가 입증될 수 있다. 크기 분포가 건강한 대조군에 대해서보다 더 길면, 이는 암의 초기를 입증할 수 있다. 낮은 컷오프의 예는 0.01, 0.015, 0.02, 또는 0.025이다. 또 다른 예로서, 측정된 분획이 높은 컷오프 초과인 경우, 크기 분포가 건강한 대조군에 대해서보다 더 짧은지 (예를 들면, 제1 통계치가 크기 한계값 미만인지)가 입증될 수 있다. 크기 분포가 건강한 대조군에 대하여 보다 짧으면, 이는 암의 후기를 입증할 수 있다. 높은 컷오프의 예는 0.03, 0.035, 0.04, 0.045, 0.05, 0.055, 0.06, 0.065, 또는 0.07의 분획일 수 있다.

[0197] 우리는 HCC 환자의 혈장에서 더 짧은 및 더 긴 DNA 분자의 추가의 집단이 있다는 것을 보여주었다. 이들 데이터는 그룹이 암 환자의 혈장내 더 긴 또는 더 짧은 DNA 분자에서 한쪽의 증가의 존재를 보고한 문헌에서 존재한 명백한 불일치를 해결할 수 있었다.

[0198] **VII. 물질 및 방법**

[0199] 도 2-5의 결과 수득에 사용된 기술은 이제 논의된다. 그와 같은 기술은 상기 다른 예에서 사용될 수 있다.

[0200] 연구를 위하여 모집된 대상체는, 종양 절제를 위하여, Department of Surgery of the Prince of Wales Hospital, Hong Kong에 허가된 HCC를 가진 90 환자를 포함하였다. 모든 혈액 샘플은 수술전 수집되었다. 67 HBV 캐리어 및 HBV-관련된 간경변을 가진 36 환자는 Department of Medicine and Therapeutics of the Prince of Wales Hospital, Hong Kong로부터 모집되었다. 모든 환자는 서면 고지에 의한 동의를 주었고 연구는 임상시험 심사 위원회에 의해 승인되었다.

[0201] DNA를 추출하기 위해 그리고 서열 라이브러리를 제조하기 위해, 말초 혈액 샘플은 EDTA-함유 튜브에 수집되었다. 말초 혈액 샘플은 1,600 g으로 10분 동안 4 °C에서 원심분리되었다. 혈장부는 16,000g으로 10분 동안 4 °C에서 재원심분리되어 무세포 혈장을 수득하였다. DNA는 QIAamp DSP DNA Blood Mini Kit (Qiagen)를 이용하여 3 내지 4.8 mL의 혈장 추출되었다. 혈장 DNA는 75-µL 최종 용적 / 샘플에 SpeedVac Concentrator (Savant DNA120; Thermo Scientific)로 농축되었다. 색인된 DNA 라이브러리는 제조자의 지침에 따라 Kapa Library Preparation Kit (Kapa Biosystems) 이용에 의해 제조되었다. 어댑터-결찰된 DNA는 KAPA HiFi HotStart ReadyMix PCR Kit (Kapa Biosystems)를 이용하여 14-사이클 PCR에 의해 풍부해졌다. 라이브러리는 그 다음 2100 Bioanalyzer (Agilent)에 의해 분석되었고 서열분석 이전 Kapa Library Quantification Kit (Kapa Biosystems)에 의해 정량화되었다.

[0202] DNA를 서열화 및 정렬하기 위해, 각 DNA 라이브러리는 희석되었고 쌍으로 된-말단 서열분석 유동 세포 (Illumina)에 혼성화되었다. DNA 클러스터는 TruSeq PE Cluster Generation Kit v3 (Illumina)으로 cBot 클러스터 생성 시스템 (Illumina)상에서, 그 다음 TruSeq SBS Kit v3 (Illumina)으로 HiSeq 2000 시스템 (Illumina)상에서 서열분석의 76 × 2 사이클로 생성되었다. 서열분석은 4-plex 프로토콜을 이용하여 수행되었다. 우리는 각 서열분석된 DNA 분자상에서 지수 서열을 디코딩하기 위해 서열분석의 추가의 7 사이클을 수행하였다. 실시간 이미지 분석 및 염기 소환은 HiSeq Control Software (HCS) v1.4 및 Real Time Analysis (RTA) Software v1.13 (Illumina)를 이용하여 수행되었고, 이에 의해 자동화 매트릭스 및 시기화 계산은 라이브러리로 서열분석된 스파이크트-인 PhiX 컨트롤 v3에 기반되었다. 염기 소환 이후, 어댑터 서열 및 낮은 품질 염기 (즉 품질 스코어 < 5)는 제거되었다.

[0203] 서열분석 데이터 분석에 대하여, 각각의 레인으로부터 서열은 6-염기 지수 서열에 기반된 대응 샘플에 배정되었

다. 서열분석된 판독은 그 다음 짧은 올리고뉴클레오타이드 정렬 프로그램 2 (SOAP2) (40)를 이용하여 비-반복-마스킹된 인간 표준 게놈 (NCBI 빌드 37/hg19)으로 정렬되었다. 최대 2 뉴클레오타이드 미스매치는 쌍으로 된-말단 판독의 각 구성원에 대하여 허용되었지만 삽입 또는 결실은 허용되지 않았다. 고유 게놈 위치에 맵핑된 판독은 다운스트림 분석에 사용되었다. 정확한 배향을 가진 동일한 염색체에 정렬된 및 ≤ 600 bp의 삽입 크기에 미치는 쌍으로 된-말단 판독은 다운스트림 크기 분석에 대하여 유지되었다. 참조 인간 게놈에 정렬 이후, 각 혈장 DNA 단편의 크기는 서열 판독의 각 쌍의 최외부 말단에서 뉴클레오타이드의 배위로부터 추론될 수 있다. 제1 단일-말단 판독은 CNA 분석에 대하여 사용되었다. Bowtie 2 소프트웨어 (41)를 이용한 30 초과 맵핑 품질 (즉 1 잘못된 정렬 / 1,000 정렬)로 판독은 허용되었다.

[0204] CNA에 대하여 CAZA 분석 수행을 위하여, 전체 인간 게놈은 100-kb 빈(bin)으로 분할되었다. GC-정정된 판독 카운트는 이전에 보고된 바와 같이 각 100-kb 빈에 대하여 결정되었다 (42). 당해 각 염색체 아암에 대하여 GC-정정된 판독 카운트의 수는 염색체 아암상에서 각 100-kb 빈의 모든 값의 합계에 의해 결정되었다. z-스코어 통계는 염색체 아암에서 혈장 DNA 표현이 참조 그룹과 비교된 경우 유의미하게 증가 또는 감소될 것인지를 결정하기 위해 사용되었다. 각 염색체 아암에 맵핑된 서열분석 판독의 백분율은 계산되었고 각자의 염색체 아암에 대하여 32 건강한 대조군 대상체의 평균 값과 비교되었다. 아암-수준 z-스코어는 하기와 같이 계산되었다:

[0205] 
$$z\text{-스코어} = \frac{P_{\text{시험}} - P_{\text{정상}}}{SD_{\text{정상}}}$$

[0206] 식 중  $P_{\text{시험}}$ 은 시험 케이스에 대하여 당해 염색체 아암에 맵핑된 단편의 비율을 나타내고;  $P_{\text{정상}}$  및  $SD_{\text{정상}}$ 은 건강한 대조군, 각각에 대하여 염색체 아암에 맵핑된 단편의 비율의 평균 및 SD를 나타낸다. < -3 및 > 3의 z-스코어를 가진 염색체 아암은 결실 및 증폭, 각각에 대응하는 혈장에서 CNAs를 갖는 것으로서 여겨졌다.

[0207] 혈장 (F)에서 종양-유도된 DNA의 단편적 농도는 하기와 같이 계산될 수 있다:

[0208] 
$$F = \frac{|P_{\text{시험}} - P_{\text{정상}}|}{\Delta N / 2 \times P_{\text{정상}}}$$

[0209] 식 중  $P_{\text{시험}}$ 은 시험 케이스에 대하여 당해 염색체 아암에 맵핑된 단편의 비율을 나타내고;  $P_{\text{정상}}$ 은 건강한 대조군에 대하여 염색체 아암에 맵핑된 단편의 평균 비율을 나타내고  $\Delta N$ 은 카피 수 변화를 나타낸다. 적어도 하나의 염색체 아암에서 결실을 보여주는 케이스에 대하여, 우리는 결실된 염색체 아암(들)에 기반된 F를 계산한다. 대부분의 염색체 아암 결실이 2 상동성 염색체 (33)의 단 하나를 포함함에 따라, 우리는 우리의 분석에 대하여 단일 카피 손실을 추정하였다. 결실 없이 염색체 아암 증폭만을 가진 24 케이스에 대하여, F는 단일 카피 이득의 가정으로 증폭된 아암에 기반하여 계산되었다.

[0210] 서열분석 데이터 분석은 Perl 및 R 언어로 쓰여진 생물정보학 프로그램을 이용함으로써 수행되었다. < 0.05의 p-값은 통계적으로 유의미한 것으로서 고려되었고 모든 개연성은 2-테일링되었다.

[0211] **VIII. 컴퓨터 시스템**

[0212] 본원에서 언급된 임의의 컴퓨터 시스템은 하위시스템의 임의의 적합한 수를 사용할 수 있다. 상기 하위시스템의 예는 컴퓨터 장치 10으로 도 32에서 보여진다. 일부 구현예에서, 컴퓨터 시스템은 단일 컴퓨터 장치를 포함하고, 여기에서 하위시스템은 컴퓨터 장치의 구성요소일 수 있다. 다른 구현예에서, 컴퓨터 시스템은, 내부 구성요소를 가진, 각각 하위시스템인, 다중 컴퓨터 장치를 포함할 수 있다. 컴퓨터 시스템은 데스크탑 및 랩탑 컴퓨터, 테블릿, 휴대폰 및 다른 모바일 장치를 포함할 수 있다.

[0213] 도 32에서 보여진 하위시스템은 시스템 버스 75를 통해 상호연결된다. 추가의 하위시스템 예컨대, 어댑터 82를 디스플레이하기 위해 커플링되는, 프린터 74, 키보드 78, 저장 장치(들) 79, 모니터 76, 및 다른 것은 보여진다. I/O 컨트롤러 71에 커플링하는, 주변장치 및 입력/출력 (I/O) 장치는 당해 기술 예컨대 입력/출력 (I/O) 포트 77 (예를 들면, USB, FireWire<sup>®</sup>)에서 공지된 수단의 임의의 수에 의해 컴퓨터 시스템에 연결될 수 있다. 예를 들어, I/O 포트 77 또는 외부 인터페이스 81 (예를 들면 이더넷, 와이파이, 등)은 컴퓨터 장치 10을 광역 네트워크 예컨대 인터넷, 마우스 입력 장치, 또는 스캐너에 연결하기 위해 사용될 수 있다. 시스템 버스 75를 통한 상호연결은 중앙 프로세서 73을 각 하위시스템에 통신하게 하고 시스템 메모리 72 또는 저장 장치(들) 79 (예를 들면, 고정 디스크, 예컨대 하드 드라이브 또는 광학 디스크)로부터 사용설명서의 실행, 뿐

만 아니라 하위시스템 사이 정보를 교환하게 한다. 시스템 메모리 72 및/또는 저장 장치(들) 79는 컴퓨터 관독 가능한 기록매체를 구현할 수 있다. 또 다른 하위시스템은 데이터 수집 장치 85, 예컨대 카메라, 마이크로폰, 가속도계, 등등이다. 본원에서 언급된 임의의 데이터는 하나의 구성요소부터 또 다른 구성요소까지 출력될 수 있고 사용자에게 출력될 수 있다.

[0214] 컴퓨터 시스템은, 예를 들면, 외부 인터페이스 81 또는 내부 인터페이스에 의해 함께 연결된, 복수의 동일한 구성요소 또는 하위시스템을 포함할 수 있다. 일부 구현예에서, 컴퓨터 시스템, 하위시스템, 또는 장치는 네트워크를 통해 통신할 수 있다. 상기 경우에서, 하나의 컴퓨터는 클라이언트 및 또 다른 컴퓨터 서버로 고려될 수 있고, 여기에서 각각은 동일한 컴퓨터 시스템의 일부일 수 있다. 클라이언트 및 서버는 각각 다중 시스템, 하위 시스템, 또는 구성요소를 포함할 수 있다.

[0215] 본 발명의 임의의 구현예가 모듈러 또는 통합된 방식으로 하드웨어 (예를 들면 응용 주문형 집적회로 또는 필드 프로그래밍가능 게이트 어레이)를 이용하여 및/또는 일반적으로 프로그래밍가능한 프로세서를 가진 컴퓨터 소프트웨어를 이용하여 제어 논리의 형태로 시행될 수 있다는 것이 이해되어야 한다. 본원에서 사용된 바와 같이, 프로세서는 단일-코어 프로세서, 동일한 집적 칩상의 다중-코어 프로세서, 또는 단일 회로 기판 또는 네트워크 상에서 다중 처리 장치를 포함한다. 본원에서 제공된 개시 및 교시에 기반하여, 당해 분야의 숙련가는 하드웨어 및 하드웨어와 소프트웨어의 조합을 이용하여 본 발명의 구현예를 시행하기 위한 다른 방식 및/또는 방법을 알 것이고 인식할 것이다.

[0216] 본 출원에 기재된 임의의 소프트웨어 구성요소 또는 기능은 임의의 적합한 컴퓨터 언어 예컨대, 예를 들어, 자바, C, C++, C#, 오브젝티브-C, 스위프트, 또는 스크립팅 언어, 예를 들어, 종래의 또는 객체-지향 기술을 이용한 예컨대 Perl 또는 Python을 이용하는 프로세서에 의해 실행되도록 소프트웨어 코드로서 시행될 수 있다. 소프트웨어 코드는 저장 및/또는 전송용 컴퓨터 관독가능한 기록매체 상에서 일련의 명령 또는 커맨드로서 저장될 수 있고, 적합한 기록매체는 랜덤 액세스 메모리 (RAM), 읽기 전용 메모리 (ROM), 자기 기록매체 예컨대 하드-드라이브 또는 플로피 디스크, 또는 광학 기록매체 예컨대 콤팩트 디스크 (CD) 또는 DVD (디지털 다용도 디스크), 플래시 기억, 등등을 포함한다. 컴퓨터 관독가능한 기록매체는 상기 저장 또는 전송 장치의 임의의 조합일 수 있다.

[0217] 상기 프로그램은 또한, 인터넷을 포함한, 다양한 프로토콜에 순응하는 유선, 광학, 및/또는 무선 네트워크를 통해 전송에 적응된 캐리어 신호를 이용하여 인코딩 및 전송될 수 있다. 이와 같이, 본 발명의 구현예에 따라 컴퓨터 관독가능한 기록매체는 상기 프로그램으로 인코딩된 데이터 신호를 이용하여 창출될 수 있다. 프로그램 코드로 인코딩된 컴퓨터 관독가능한 기록매체는 양립가능한 장치로 패키징될 수 있거나 또는 (예를 들면, 인터넷 다운로드를 통해) 다른 장치로부터 개별적으로 제공될 수 있다. 임의의 상기 컴퓨터 관독가능한 기록매체는 단일 컴퓨터 프로덕트 (예를 들면 하드 드라이브, CD, 또는 전체 컴퓨터 시스템)상에서 또는 내에서 존재할 수 있고, 시스템 또는 네트워크내 상이한 컴퓨터 프로덕트상에서 또는 내에서 존재할 수 있다. 컴퓨터 시스템은 사용자에게 본원에서 언급된 임의의 결과를 제공하기 위하여 모니터, 프린터, 또는 다른 적합한 디스플레이를 포함할 수 있다.

[0218] 본원에서 기재된 임의의 방법은, 단계를 수행하기 위해 구성될 수 있는, 하나 이상의 프로세서를 포함한 컴퓨터 시스템으로 전적으로 또는 부분적으로 수행될 수 있다. 따라서, 구현예는, 잠재적으로 각자의 단계 또는 각자의 단계의 그룹을 수행하는 상이한 구성요소와 함께, 본원에서 기재된 임의의 방법의 단계를 수행하기 위해 구성된 컴퓨터 시스템에 관련될 수 있다. 넘버링된 단계로서 제시되어도, 본원에서 방법의 단계는 동일한 시간에서 또는 상이한 순서로 수행될 수 있다. 추가로, 이들 단계의 부분은 다른 방법으로부터 다른 단계의 부분으로 사용될 수 있다. 또한, 단계의 전체 또는 부분은 선택적일 수 있다. 추가로, 임의의 방법의 임의의 단계는 모듈, 회로, 또는 이들 단계 수행용 다른 수단으로 수행될 수 있다.

[0219] 특정한 구현예의 특이적 세부사항은 본 발명의 구현예의 취지 및 범위로부터 이탈 없이 임의의 적합한 방식으로 조합될 수 있다. 그러나, 본 발명의 다른 구현예는 각 개별 측면, 또는 이들 개별 측면의 특이적 조합에 관한 특이적 구현예에 관련될 수 있다.

[0220] 본 발명의 예 구현예의 상기 설명은 예시 및 설명의 목적으로 제시되었다. 본 발명을 기재된 정확한 형태로 제한하거나 또는 포괄하는 의도는 아니고, 많은 변형 및 변동은 상기 교시의 점에서 가능하다.

[0221] 단수의 설명은 반대로 구체적으로 명시되지 않는 한 "하나 이상"을 의미하도록 의도된다. "또는"의 사용은 반대로 구체적으로 명시되지 않는 한 "배타적 또는"이 아닌 "포괄적 또는"을 의미하는 의도이다.

[0222] 전체 특허, 특허 출원, 공개, 및 본원에서 언급된 설명은 모든 목적으로 그 전체가 참고로 편입된다. 어떤 것도 선행기술로 인정되지 않는다.

[0223] IX. 참조문헌

1. Chan KCA (2013) Scanning for cancer genomic changes in plasma: toward an era of personalized blood-based tumor markers. *Clin Chem* 59(11):1553-1555.
2. Dawson SJ, Rosenfeld N, & Caldas C (2013) Circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med* 369(1):93-94.
3. Bidard FC, Weigelt B, & Reis-Filho JS (2013) Going with the flow: from circulating tumor cells to DNA. *Sci Transl Med* 5(207):207ps214.
4. Chan KCA, *et al.* (2013) Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. *Clin Chem* 59(1):211-224.
5. Heitzer E, *et al.* (2013) Establishment of tumor-specific copy number alterations from plasma DNA of patients with cancer. *Int J Cancer* 133(2):346-356.
6. Heitzer E, *et al.* (2013) Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. *Genome Med* 5(4):30.
7. Leary RJ, *et al.* (2012) Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Sci Transl Med* 4(162):162ra154.
8. Chan KCA, *et al.* (2013) Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. *Proc Natl Acad Sci U S A* 110(47):18761-18768.
9. Chan KCA, *et al.* (2008) Quantitative analysis of circulating methylated DNA as a biomarker for hepatocellular carcinoma. *Clin Chem* 54(9):1528-1536.
10. Wong IH, *et al.* (1999) Detection of aberrant p16 methylation in the plasma and serum of liver cancer patients. *Cancer Res* 59(1):71-73.
11. Balgouranidou I, *et al.* (2014) Breast cancer metastasis suppressor-1 promoter methylation in cell-free DNA provides prognostic information in non-small cell lung cancer. *Br J Cancer* 110(8):2054-2062.
12. Diehl F, *et al.* (2005) Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc Natl Acad Sci U S A* 102(45):16368-16373.

[0224]

13. Yung TKF, *et al.* (2009) Single-molecule detection of epidermal growth factor receptor mutations in plasma by microfluidics digital PCR in non-small cell lung cancer patients. *Clin Cancer Res* 15(6):2076-2084.
14. Murtaza M, *et al.* (2013) Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* 497(7447):108-112.
15. Forshew T, *et al.* (2012) Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med* 4(136):136ra168.
16. Lo YMD, *et al.* (1999) Quantitative analysis of cell-free Epstein-Barr virus DNA in plasma of patients with nasopharyngeal carcinoma. *Cancer Res* 59(6):1188-1191.
17. Chan KCA, *et al.* (2013) Early detection of nasopharyngeal carcinoma by plasma Epstein-Barr virus DNA analysis in a surveillance program. *Cancer* 119(10):1838-1844.
18. McBride DJ, *et al.* (2010) Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. *Genes, Chromosomes & Cancer* 49(11):1062-1069.
19. Leary RJ, *et al.* (2010) Development of personalized tumor biomarkers using massively parallel sequencing. *Sci Transl Med* 2(20):20ra14.
20. Chan KCA, Leung SF, Yeung SW, Chan ATC, & Lo YMD (2008) Persistent aberrations in circulating DNA integrity after radiotherapy are associated with poor prognosis in nasopharyngeal carcinoma patients. *Clin Cancer Res* 14(13):4141-4145.
21. Gao YJ, *et al.* (2010) Increased integrity of circulating cell-free DNA in plasma of patients with acute leukemia. *Clin Chem Lab Med* 48(11):1651-1656.
22. Umetani N, *et al.* (2006) Increased integrity of free circulating DNA in sera of patients with colorectal or periampullary cancer: direct quantitative PCR for ALU repeats. *Clin Chem* 52(6):1062-1069.
23. Wang BG, *et al.* (2003) Increased plasma DNA integrity in cancer patients. *Cancer Res* 63(14):3966-3968.
24. Umetani N, *et al.* (2006) Prediction of breast tumor progression by integrity of free circulating DNA in serum. *J Clin Oncol* 24(26):4270-4276.

[0225]

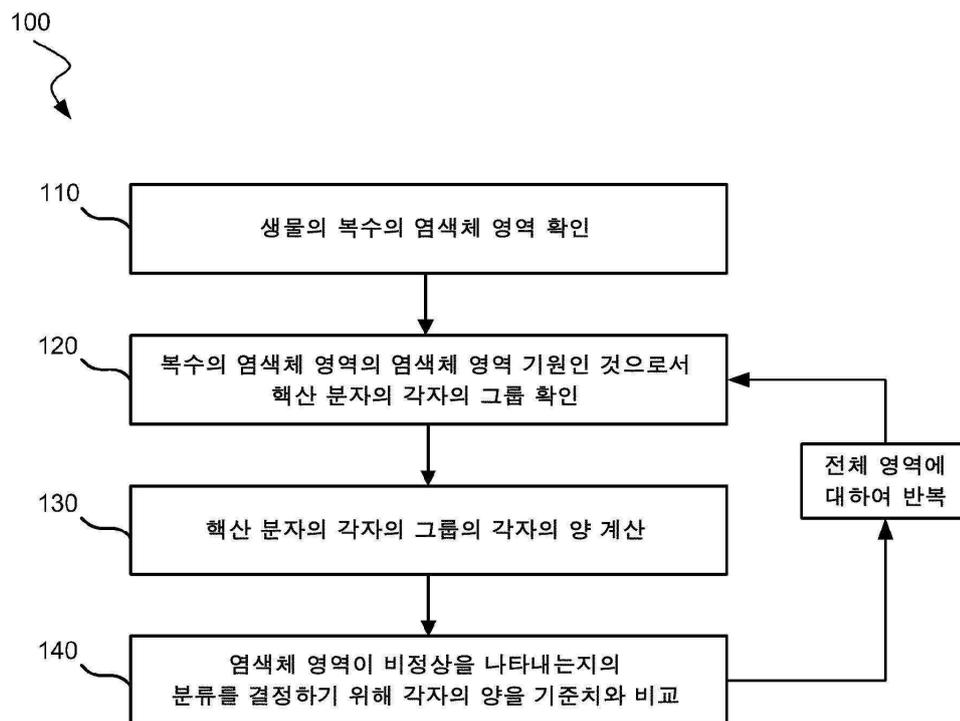
25. Schwarzenbach H, *et al.* (2012) Loss of heterozygosity at tumor suppressor genes detectable on fractionated circulating cell-free tumor DNA as indicator of breast cancer progression. *Clin Cancer Res* 18(20):5719-5730.
26. Lo YMD, *et al.* (2010) Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci Transl Med* 2(61):61ra91.
27. Zheng YWL, *et al.* (2012) Nonhematopoietically derived DNA is shorter than hematopoietically derived DNA in plasma: a transplantation model. *Clin Chem* 58(3):549-558.
28. Yu SCY, *et al.* (2014) Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing. *Proc Natl Acad Sci U S A* 111(23):8583-8588.
29. Pleasance ED, *et al.* (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 463(7278):191-196.
30. Fujimoto A, *et al.* (2012) Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet* 44(7):760-764.
31. Tao Y, *et al.* (2011) Rapid growth of a hepatocellular carcinoma and the driving mutations revealed by cell-population genetic analysis of whole-genome data. *Proc Natl Acad Sci U S A* 108(29):12042-12047.
32. Totoki Y, *et al.* (2011) High-resolution characterization of a hepatocellular carcinoma genome. *Nat Genet* 43(5):464-469.
33. Beroukhim R, *et al.* (2010) The landscape of somatic copy-number alteration across human cancers. *Nature* 463(7283):899-905.
34. Chiang DY, *et al.* (2008) Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. *Cancer Res* 68(16):6779-6788.
35. Kan Z, *et al.* (2013) Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res* 23(9):1422-1433.
36. Kim TM, *et al.* (2008) Clinical implication of recurrent copy number alterations in hepatocellular carcinoma and putative oncogenes in recurrent gains on 1q. *Int J Cancer* 123(12):2808-2815.

[0226]

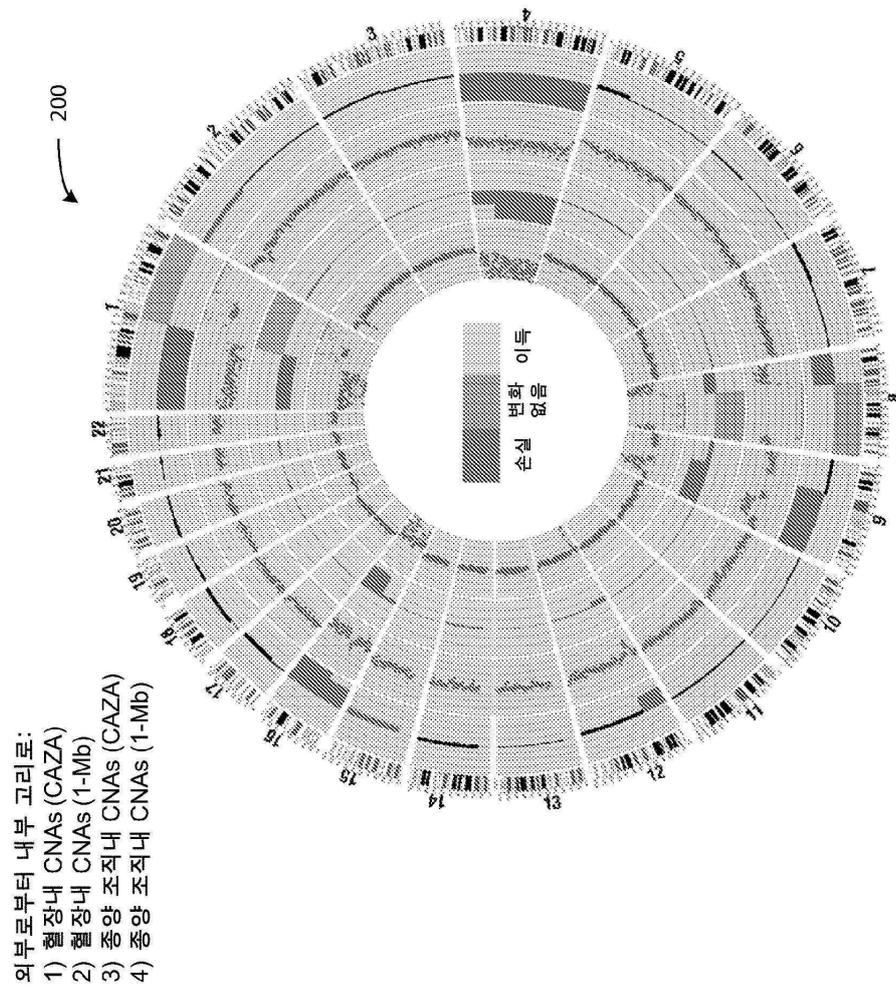
37. Nakano H & Shinohara K (1994) X-ray-induced cell death: apoptosis and necrosis. *Radiation Research* 140(1):1-9.
38. Walker NI, Harmon BV, Gobe GC, & Kerr JF (1988) Patterns of cell death. *Methods and Achievements in Experimental Pathology* 13:18-54.
39. Alexandrov LB, *et al.* (2013) Signatures of mutational processes in human cancer. *Nature* 500(7463):415-421.
40. Li R, *et al.* (2009) SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25(15):1966-1967.
41. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9(4):357-359.
42. Chen EZ, *et al.* (2011) Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing. *PLoS One* 6(7):e21791.

도면

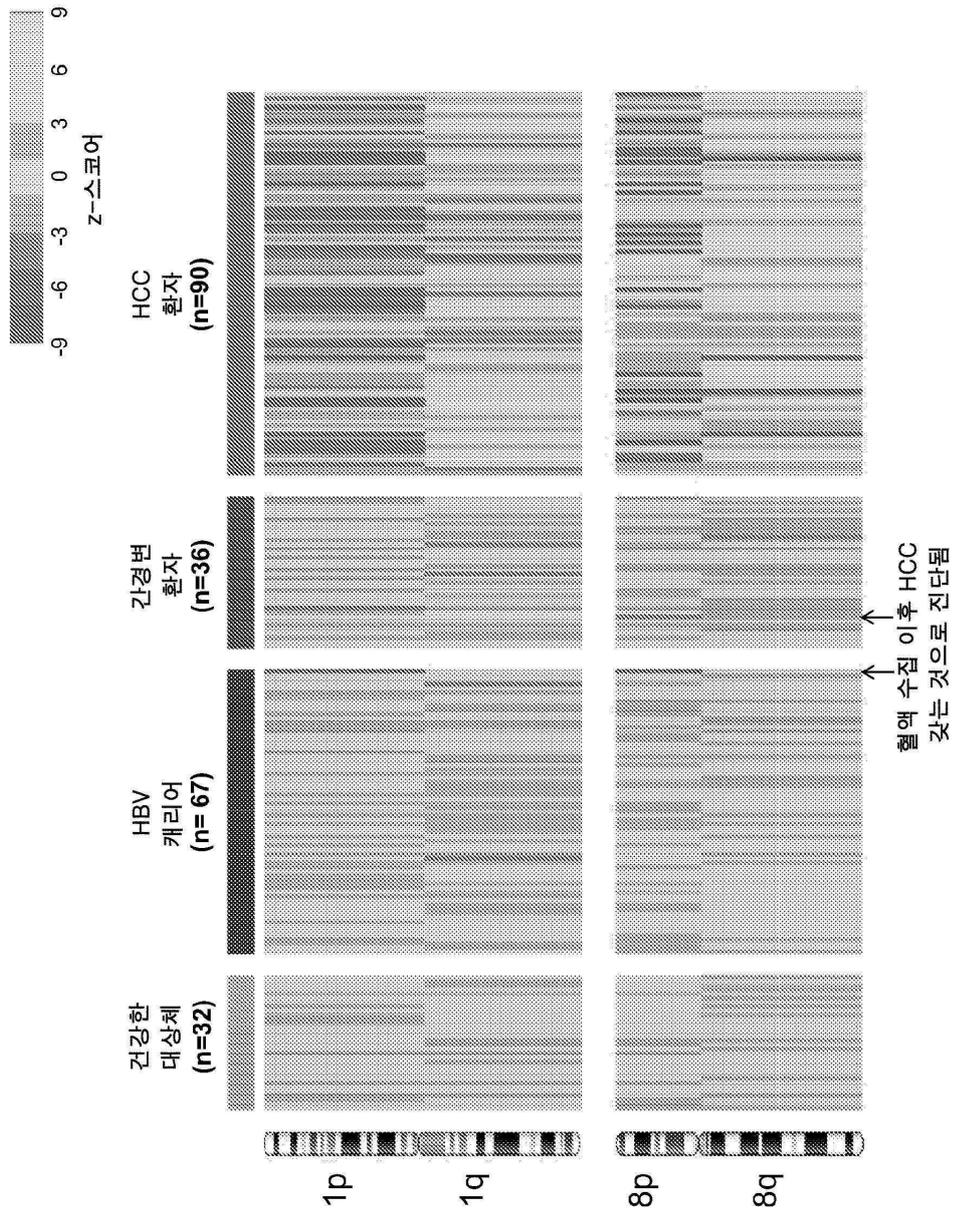
도면1



도면2



도면3



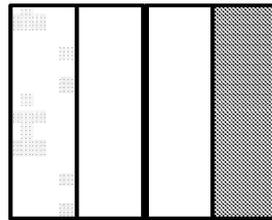
도면4

400

	하기에 대하여 혈장에서 검출된 CNA를 가진 환자의 수 (백분율)					
	Chr 1p	Chr 1q	Chr 8p	Chr 8q	임의의 chr 1p/1q/8p/8q	
HCC 환자 (n=90)	39/90 (43.3%)	46/90 (51.1%)	33/90 (36.7%)	63/90 (70.0%)	76/90 (84.4%)	
간경변 환자 (n=36)	3/36 (8.3%)	5/36 (13.9%)	1/36 (2.8%)	1/36 (2.8%)	8/36 (22.2%)	
HBV 캐리어 (n=67)	1/67 (1.5%)	3/67 (4.48%)	1/67 (1.5%)	1/67 (1.5%)	3/67 (4.5%)	
건강한 대상체 (n=32)	0/32 (0%)	0/32 (0%)	0/32 (0%)	0/32 (0%)	0/32 (0%)	

도면5

케이스 번호	혈장내 중양 DNA 분획 (%)	중양 크기 (cm)	1p		1q		8p		8q	
			중양	혈장	중양	혈장	중양	혈장	중양	혈장
HOT412	27.5	6.5	이득	이득	이득	이득	손실	손실	이득	이득
HOT407	19.4	8	널	널	널	이득	손실	손실	이득	이득
HOT414	18.7	15	손실	손실	이득	이득	손실	손실	이득	이득
HOT426	4.1	6.5	손실	손실	이득	이득	손실	널	널	이득
HOT394	2.5	4	손실	손실	이득	이득	손실	손실	이득	이득
HOT397	2.4	5.5	손실	손실	이득	널	손실	널	이득	이득
HOT393	1.8	3	손실	손실	널	손실	널	이득	널	이득
HOT427	1.7	1.5	손실	널	이득	이득	널	널	널	널
HOT413	1.7	2	널	손실	이득	널	손실	손실	이득	이득
HOT425	1.5	8.5	손실	손실	널	널	손실	널	이득	이득
HOT428	1.5	2	손실	널	이득	손실	손실	손실	이득	이득
HOT432	1.0	2	손실	널	널	널	널	널	널	이득



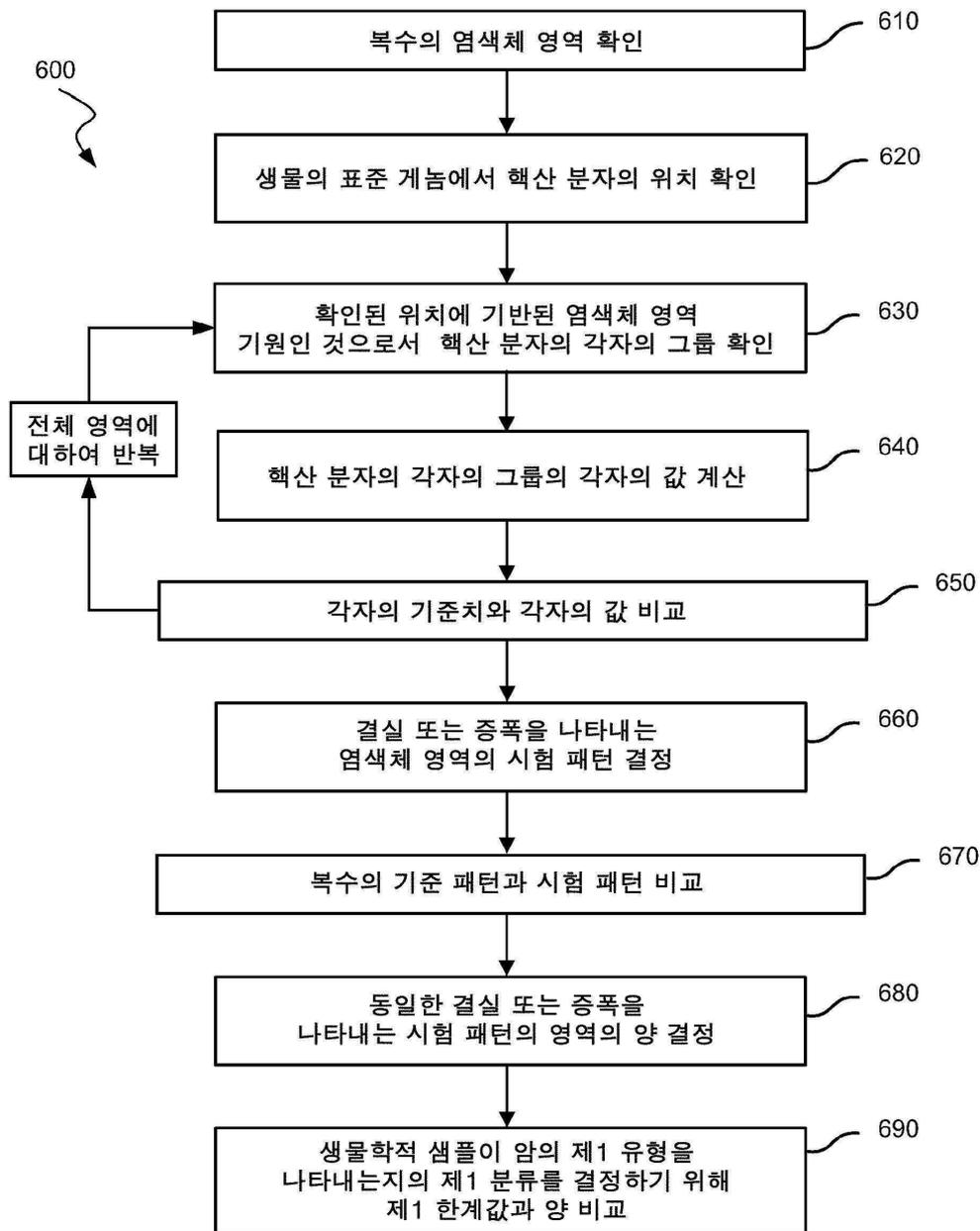
중양 및 혈장내 일치한 결과 30/48 (63%)

혈장내가 아닌 중양내 검출된 CNA 10/48 (21%);  
<2% 중양 DNA를 가진 혈장 샘플에 대하여 주로

중양내가 아닌 혈장내 검출된 CNA 7/48 (15%)

중양 및 혈장내  
상이한 CNA 패턴 1/48 (2%)

도면6



도면7

	HCC1	HCC2	HCC3	HCC4	HCC5	HCC6	CRC1	CRC2	CRC3	CRC4	BrC1	BrC2	BrC3	LC1	LC2	NPC1	NPC2
1p	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	-
1q	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2p	+	+	-	+	+	+	+	+	+	+	+	-	+	+	+	+	+
2q	+	+	-	+	+	+	+	+	+	+	+	-	+	+	+	+	+
3p	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+
3q	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+
4p	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+
4q	+	+	-	+	+	+	+	+	+	+	+	-	+	+	+	+	+
5p	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+
5q	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+
6p	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+
6q	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+
7p	+	+	-	+	+	+	+	+	+	+	+	-	+	+	+	+	+
7q	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+
8p	-	-	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+
8q	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+
9p	+	-	+	+	+	+	+	-	+	-	+	-	+	+	+	+	-
9q	+	-	+	+	+	+	+	-	+	-	+	-	+	+	+	+	-
11p	+	+	+	+	+	+	+	+	-	+	+	-	+	+	+	+	+
12p	-	-	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+
12q	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+
13q	-	+	-	+	+	+	+	+	+	+	+	-	+	+	+	+	+
14q	+	-	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+
15p	+	-	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+
16q	+	-	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+
17p	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+
17q	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+
18p	-	+	-	+	+	+	+	+	+	+	+	-	+	+	+	+	+
18q	+	+	-	+	+	+	+	+	+	+	+	-	+	+	+	+	+
22q	-	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+

700 ↗

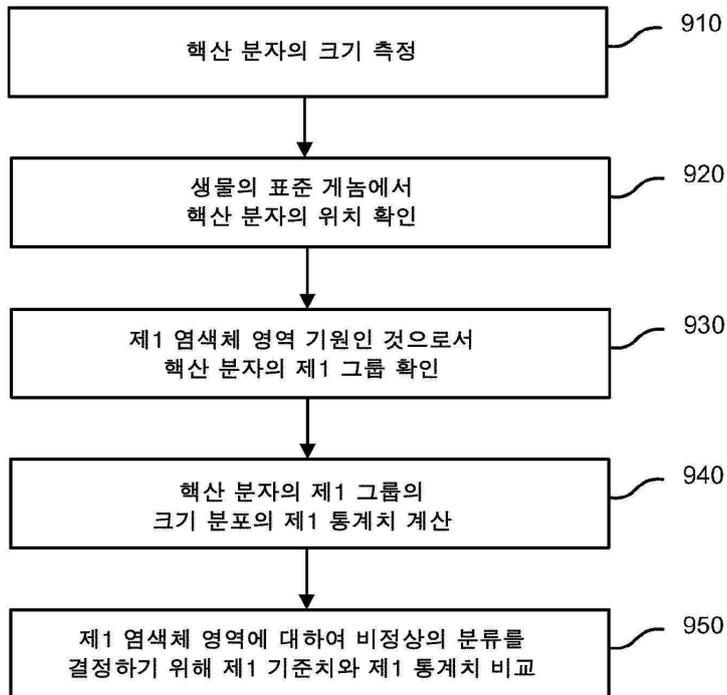






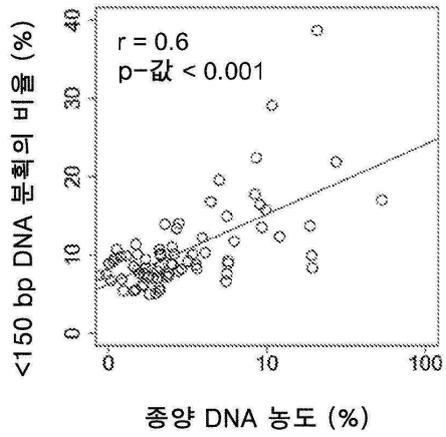
도면9

900

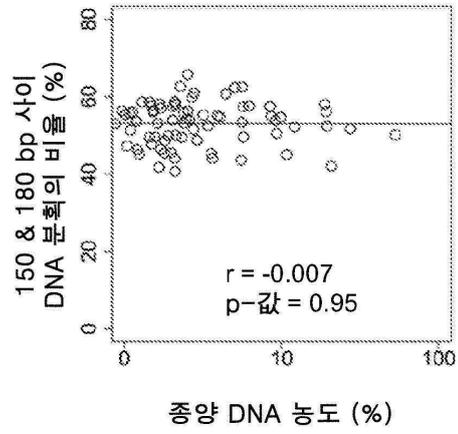


도면10

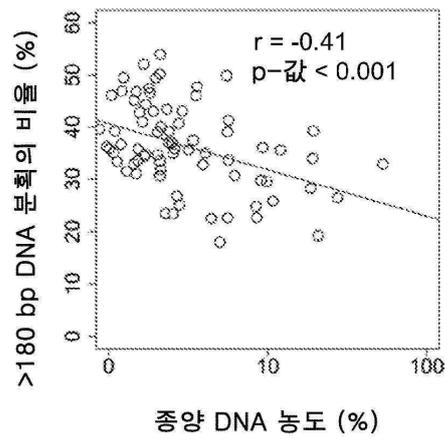
A.



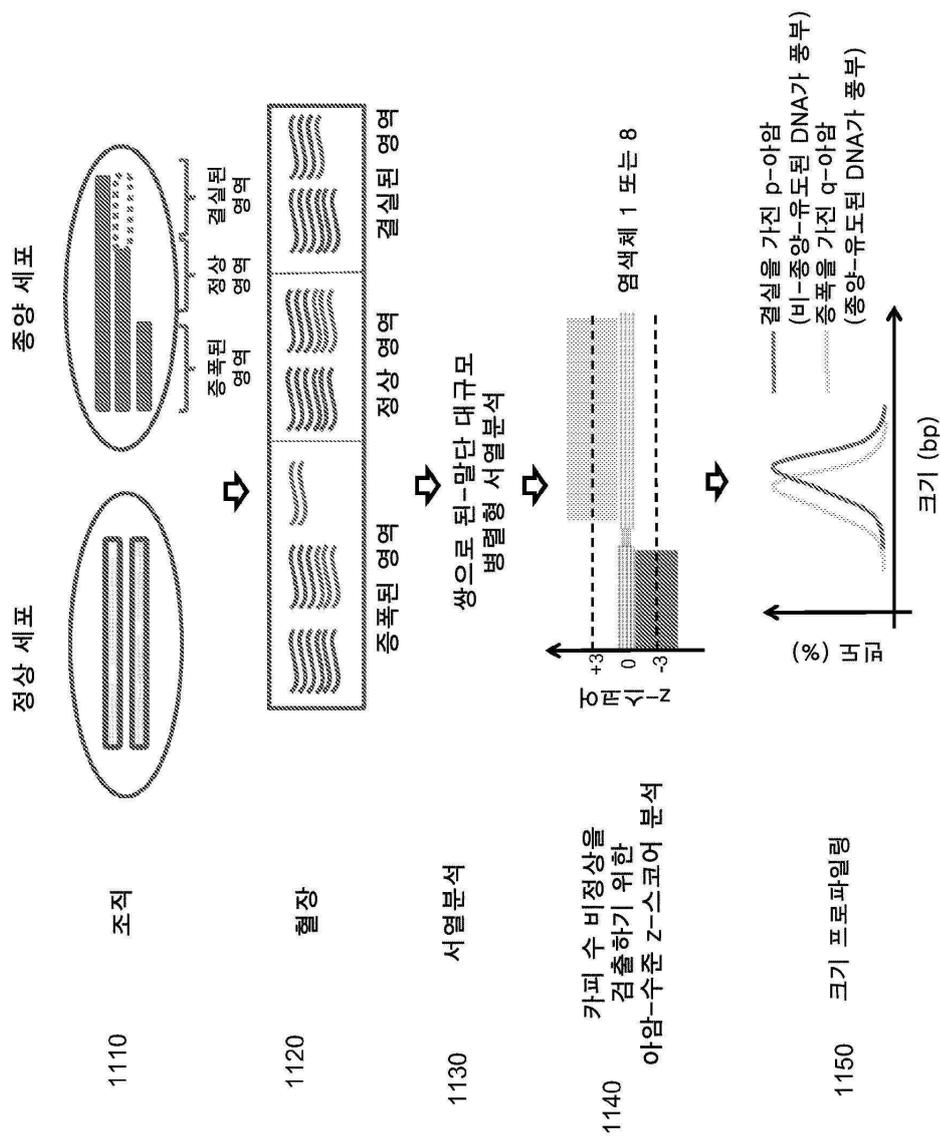
B.



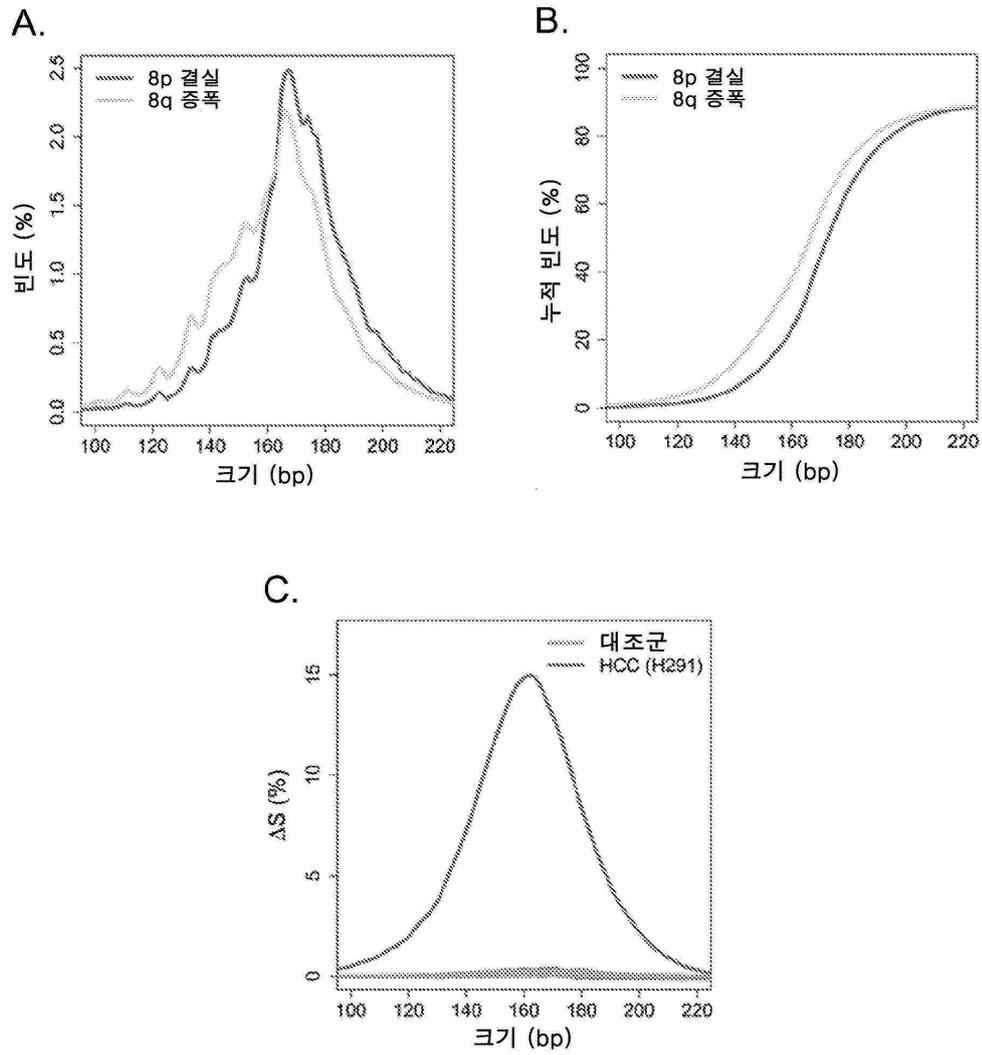
C.



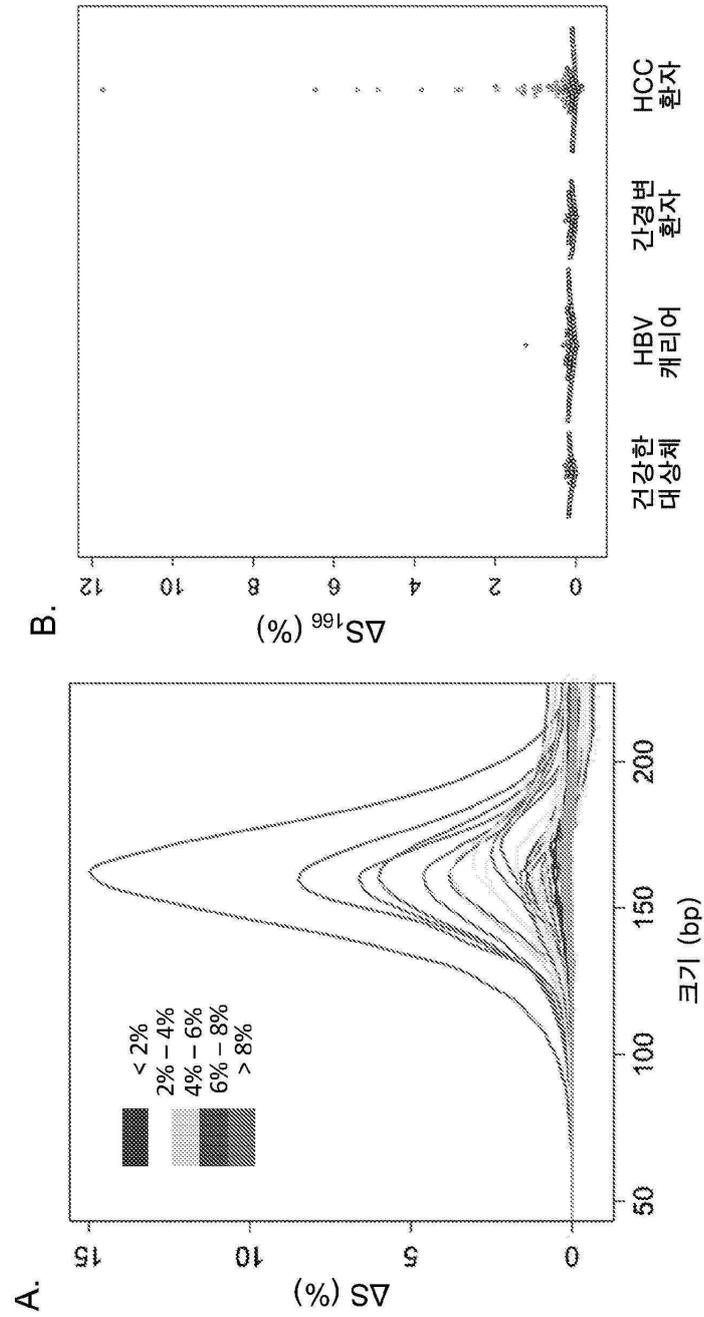
도면11



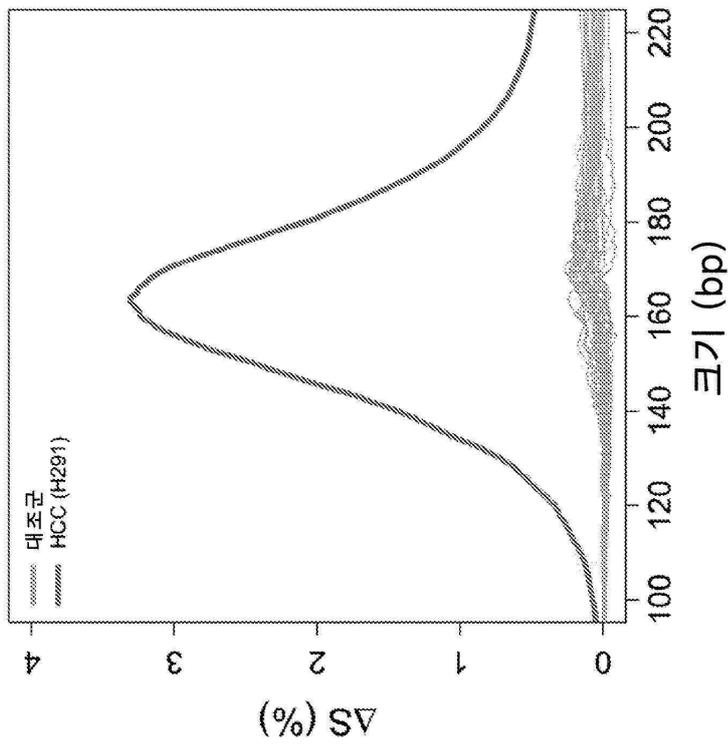
도면12



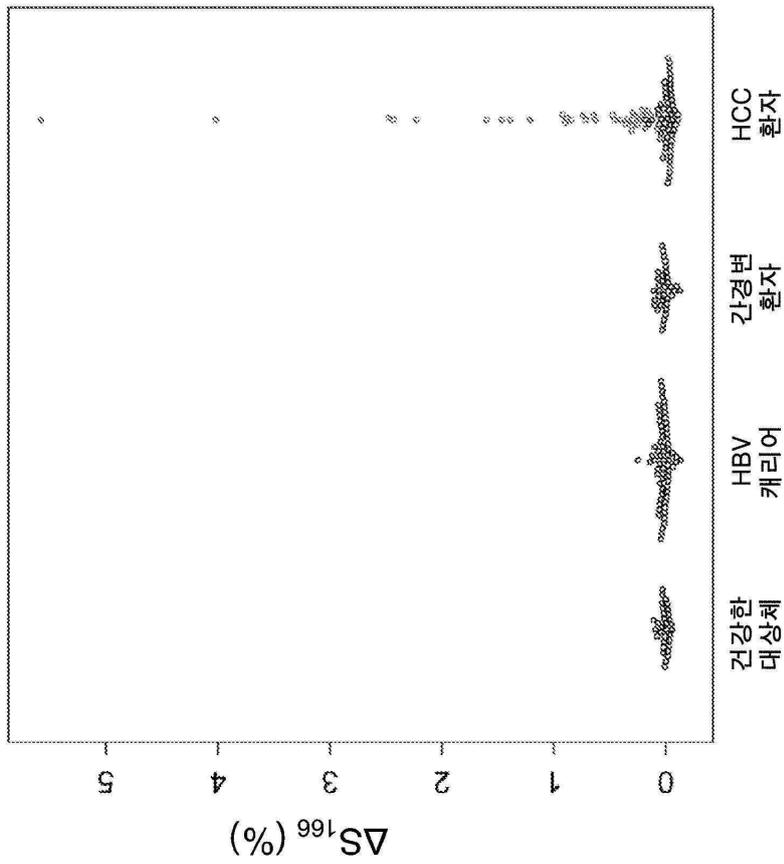
도면13



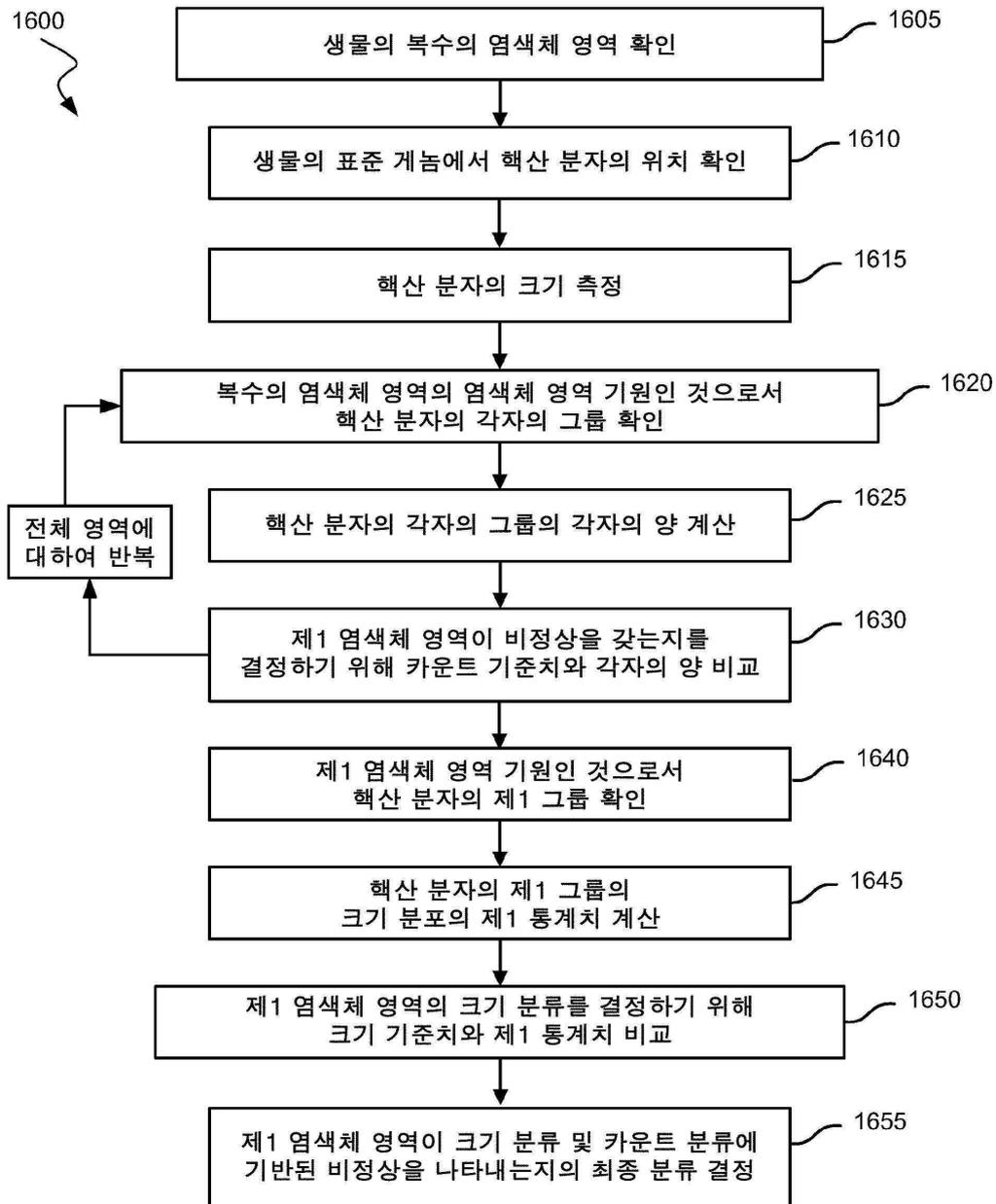
도면14



도면15

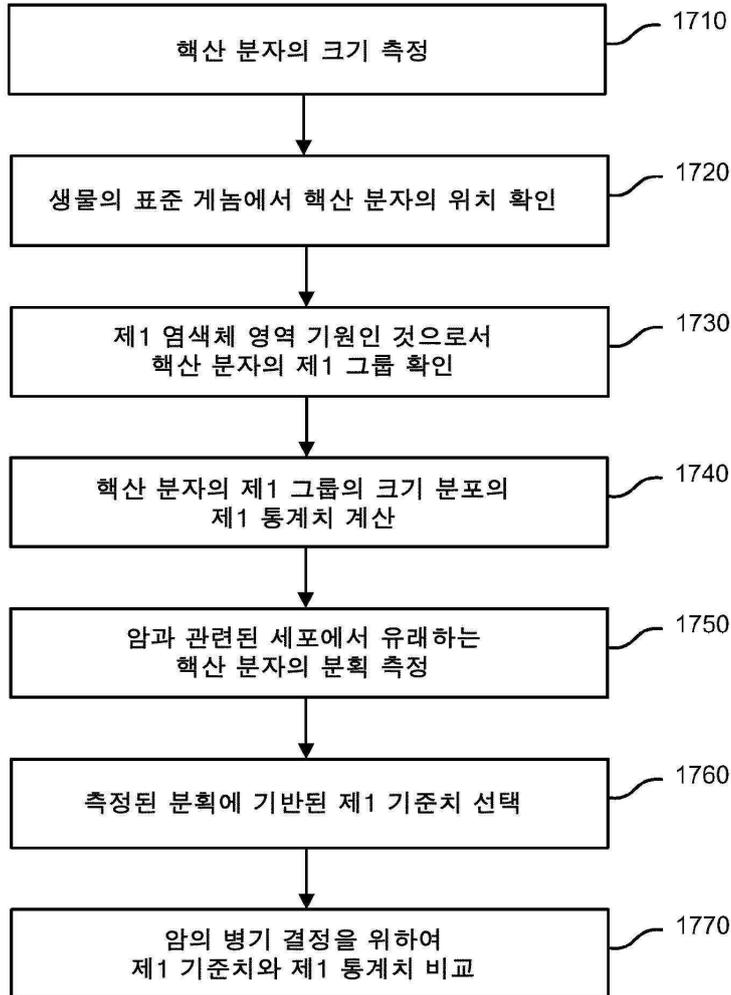


도면16

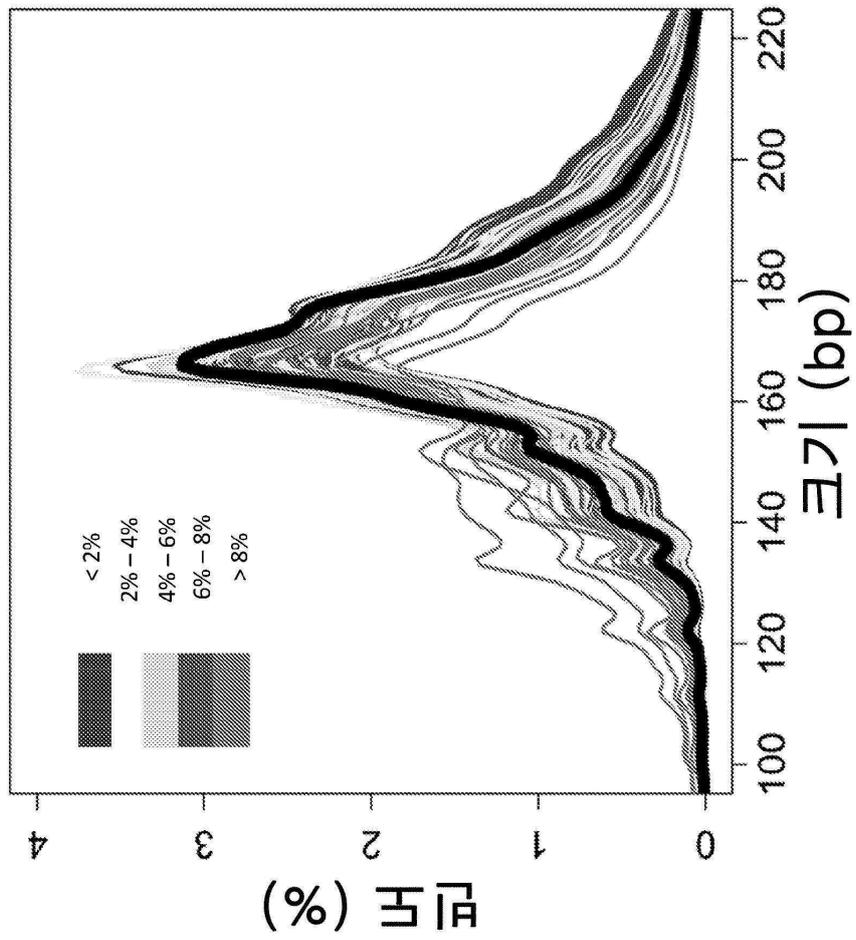


도면17

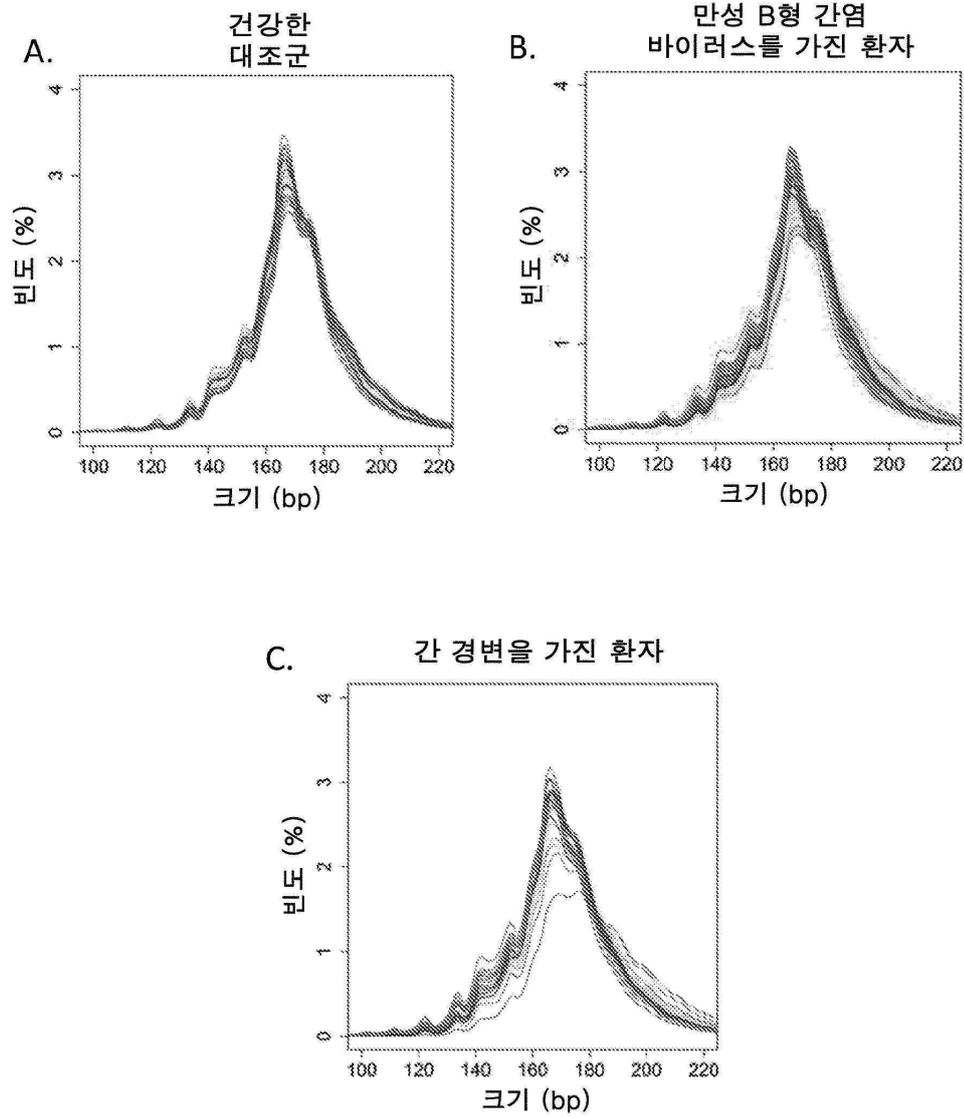
1700



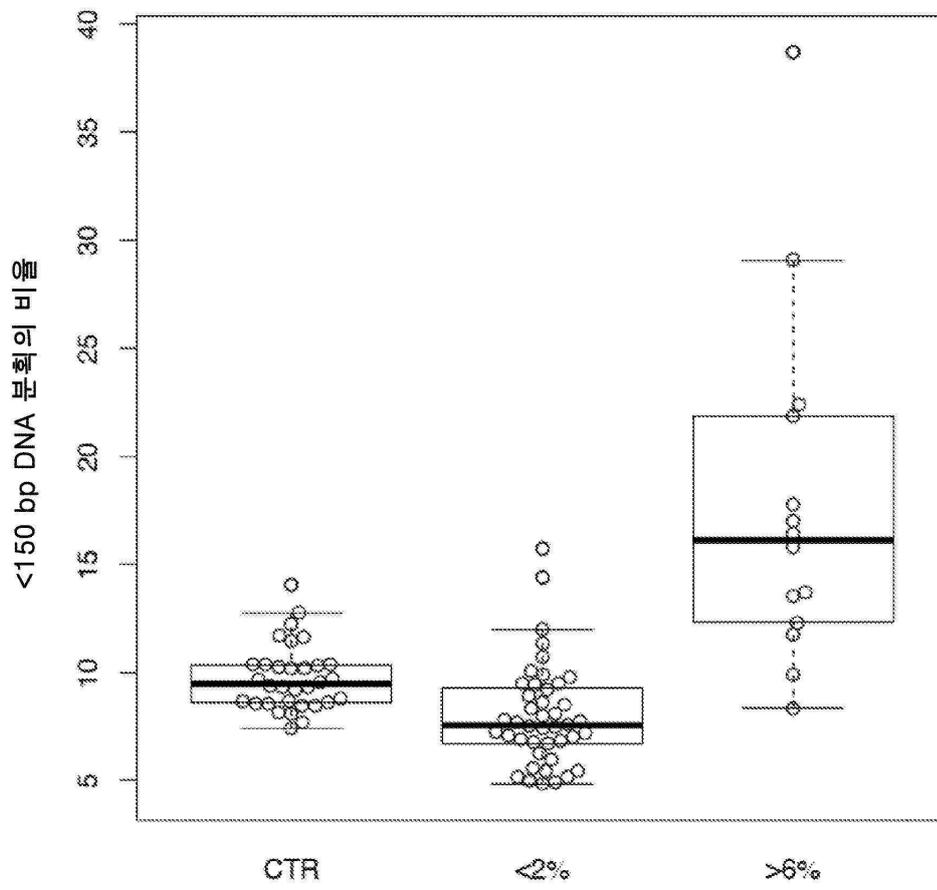
도면18



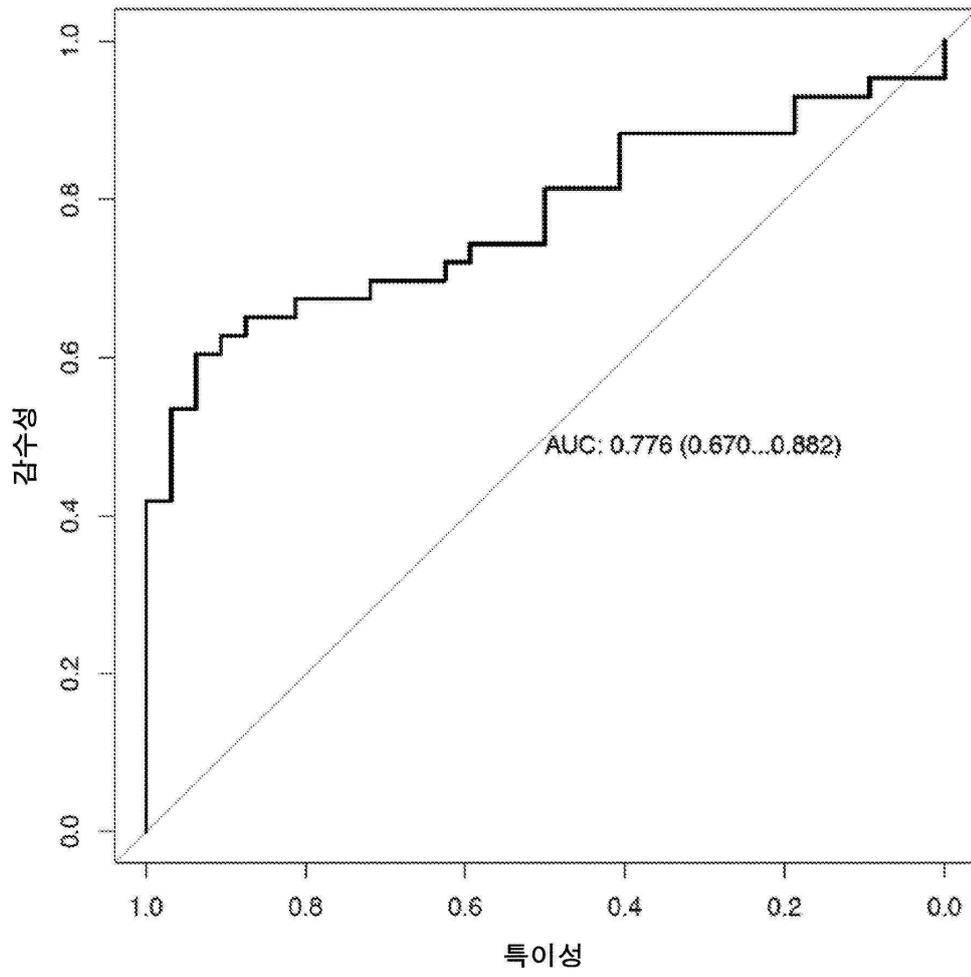
도면19



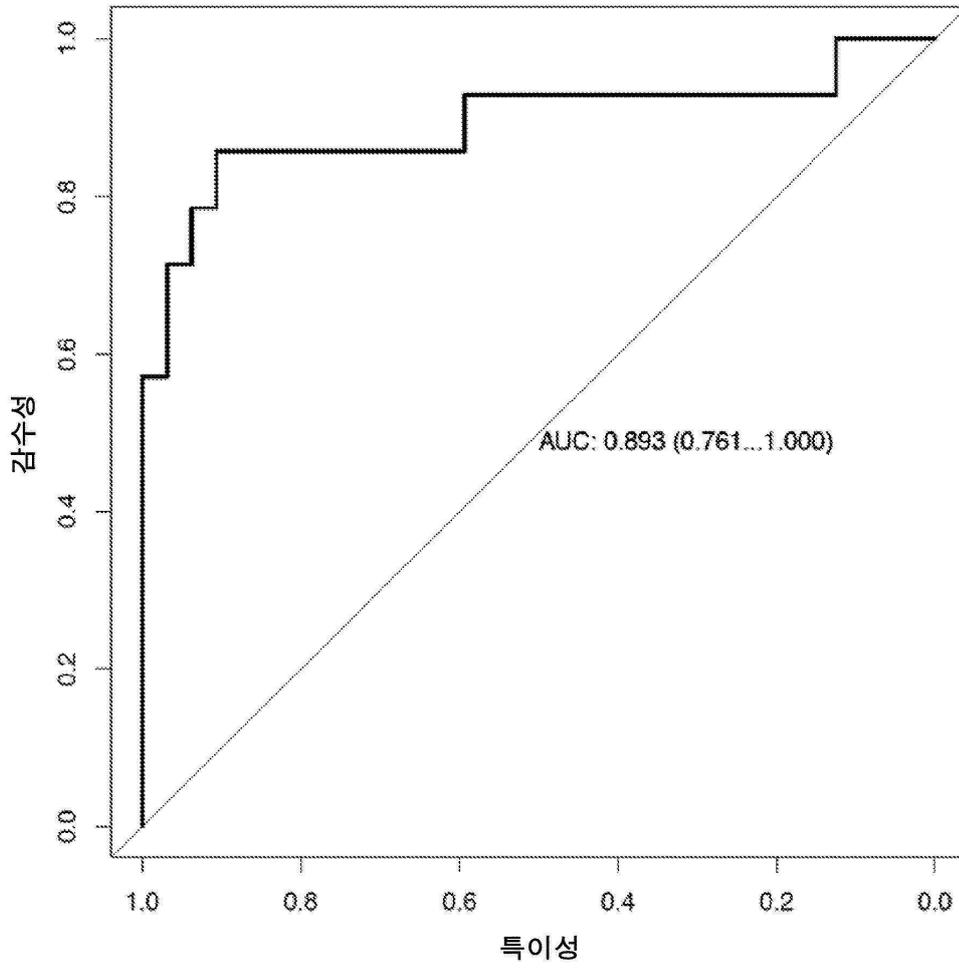
도면20



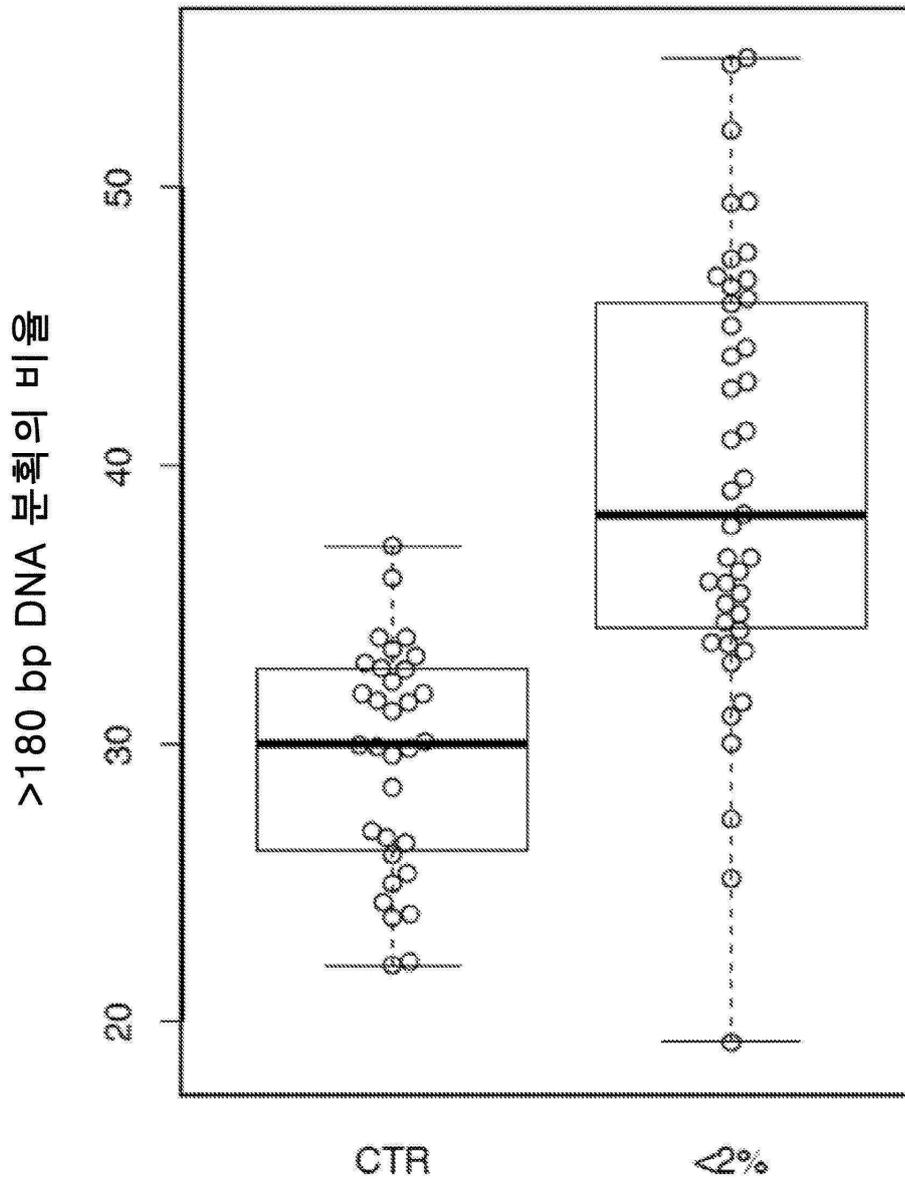
도면21



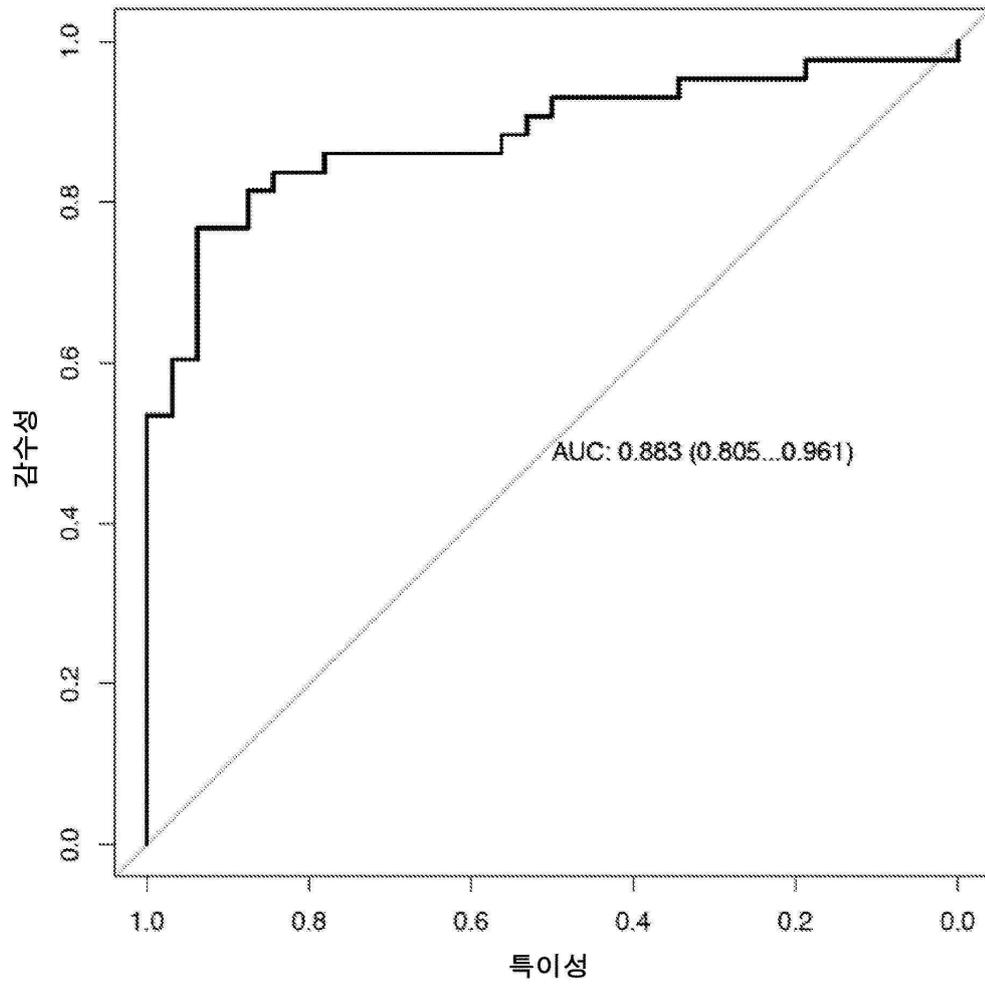
도면22



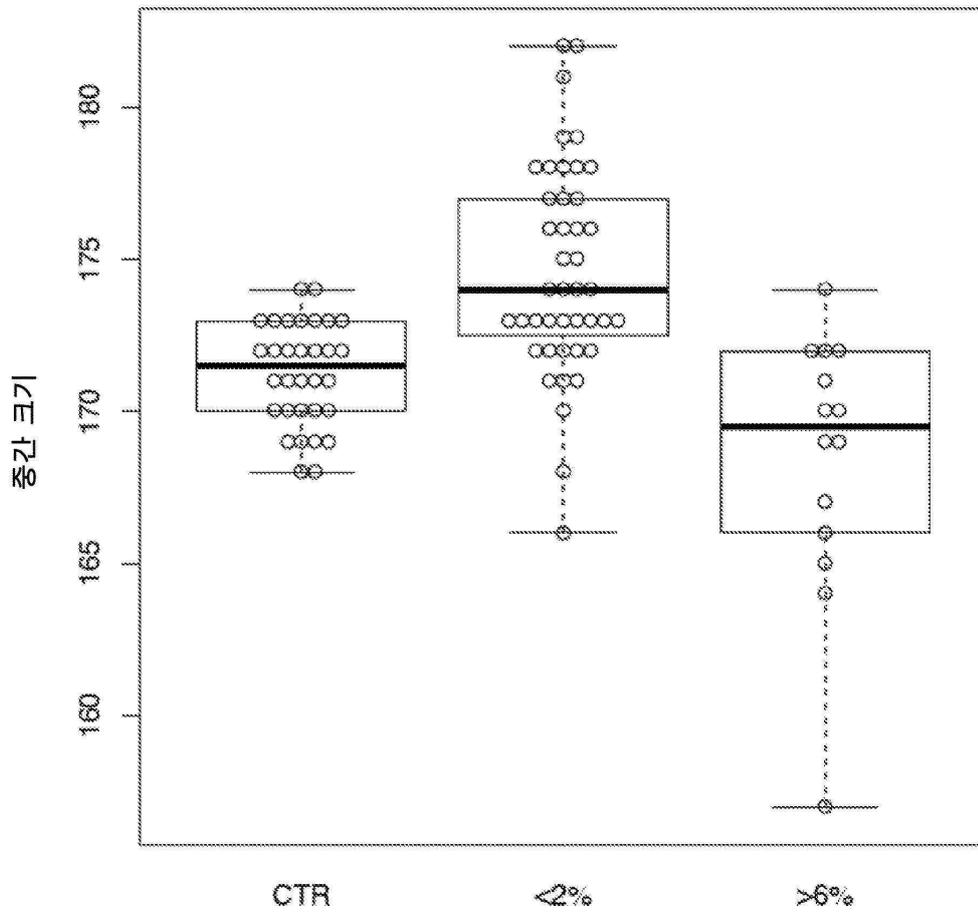
도면23



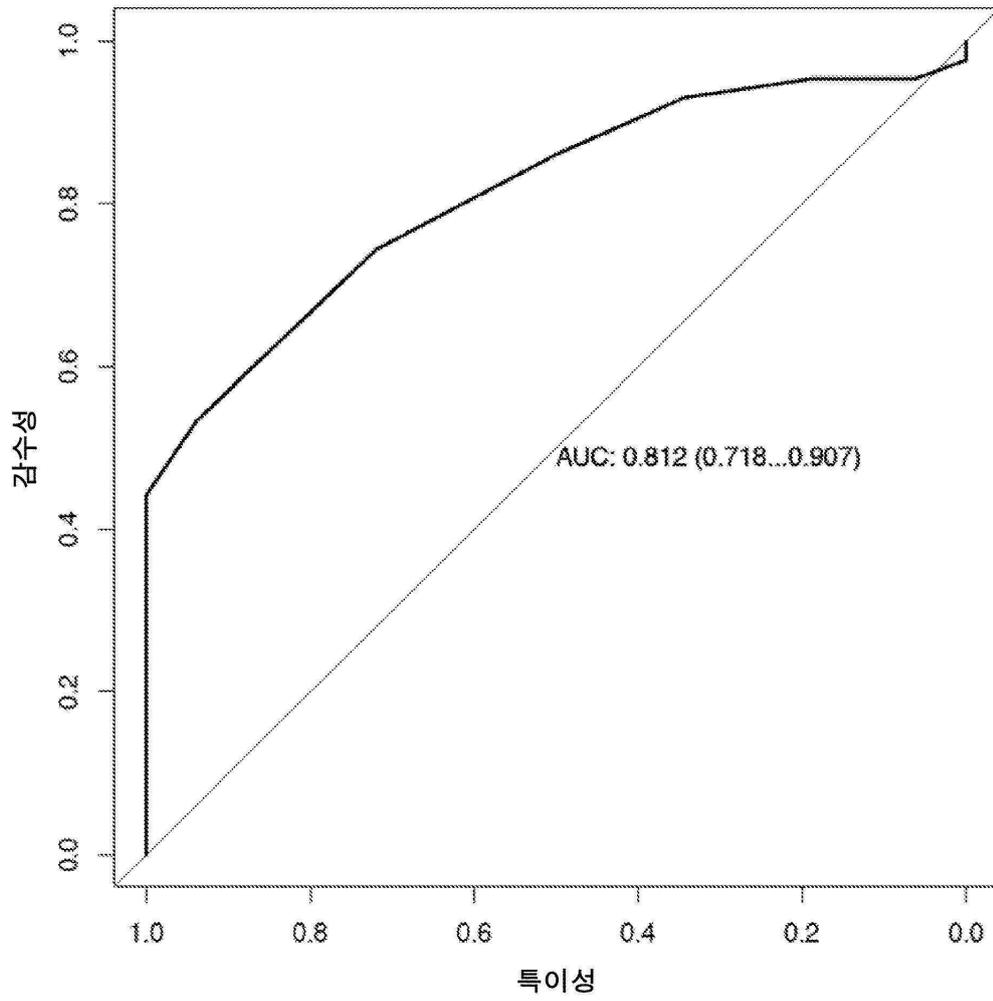
도면24



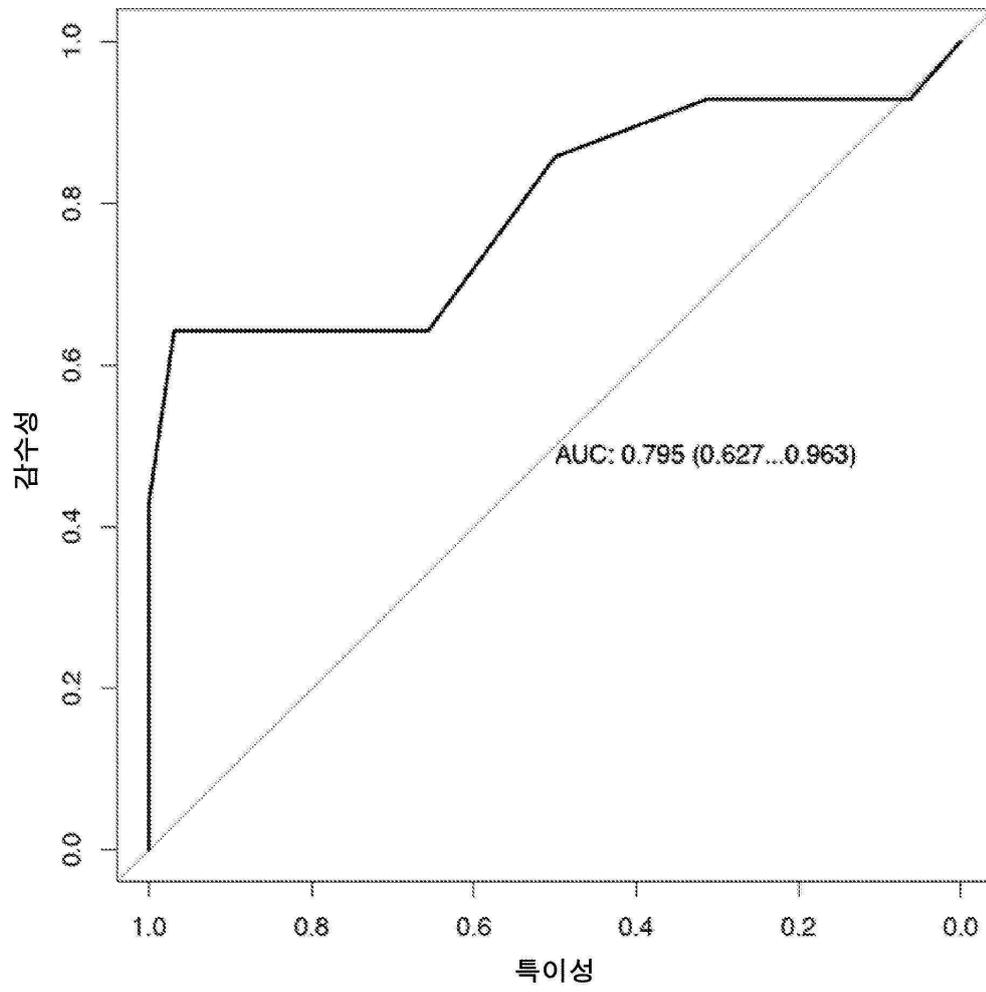
도면25



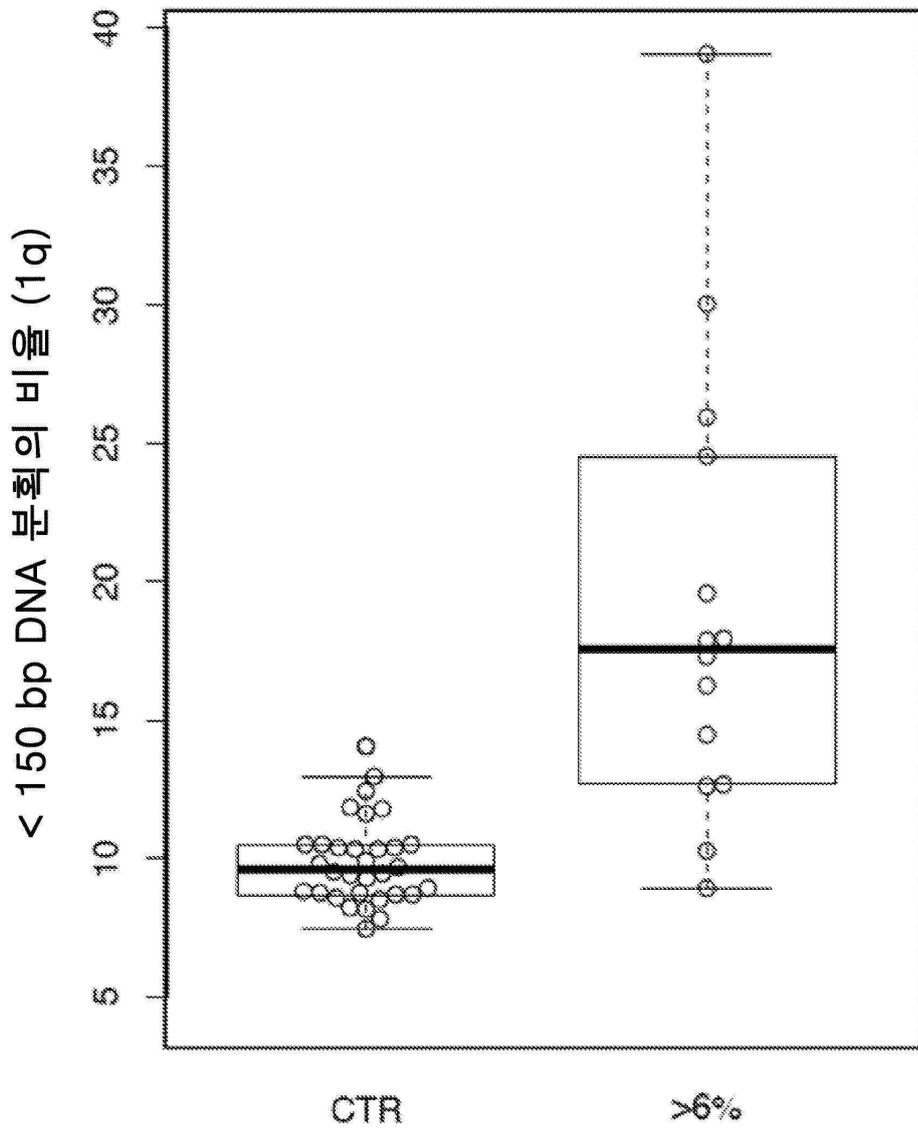
도면26



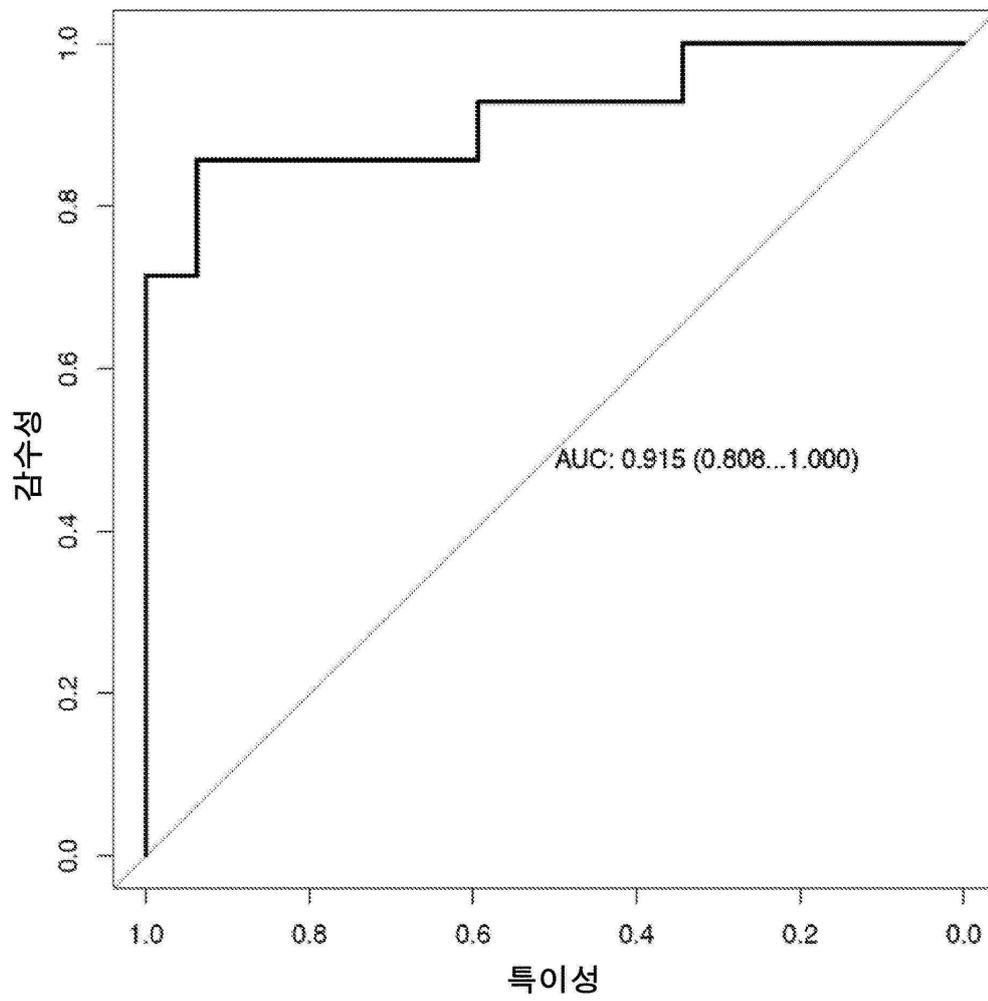
도면27



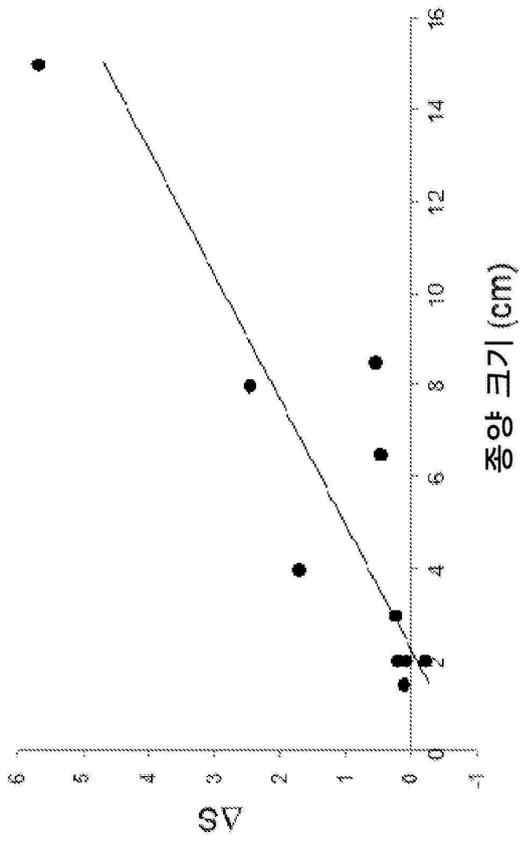
도면28



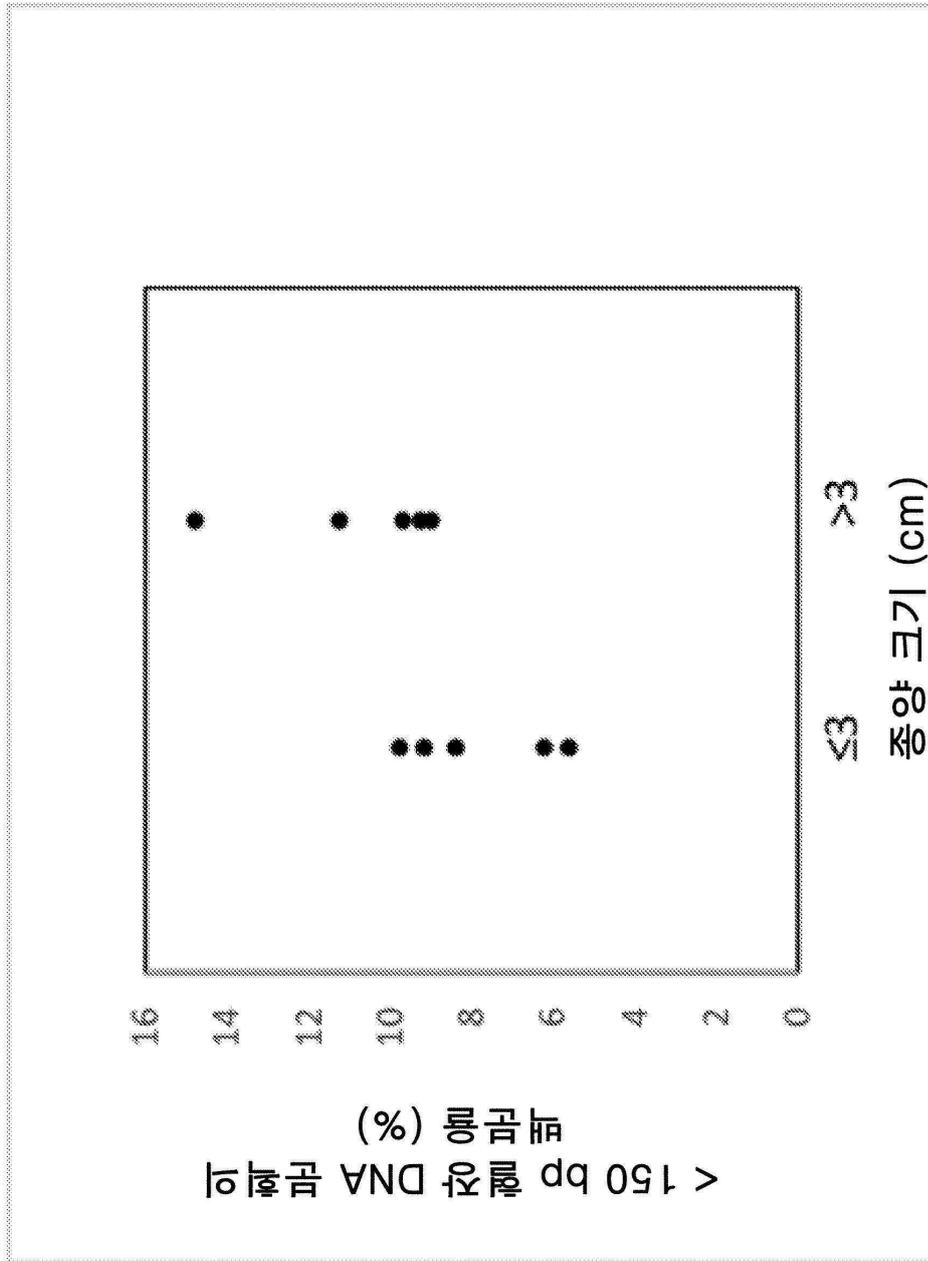
도면29



도면30



도면31



도면32

