



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2018년08월24일
(11) 등록번호 10-1891847
(24) 등록일자 2018년08월20일

(51) 국제특허분류(Int. Cl.)
C12Q 1/68 (2018.01) C12N 15/11 (2006.01)
G06F 19/18 (2011.01)
(21) 출원번호 10-2013-7016780
(22) 출원일자(국제) 2011년11월30일
심사청구일자 2016년11월14일
(85) 번역문제출일자 2013년06월27일
(65) 공개번호 10-2014-0091650
(43) 공개일자 2014년07월22일
(86) 국제출원번호 PCT/AU2011/001562
(87) 국제공개번호 WO 2012/071621
국제공개일자 2012년06월07일
(30) 우선권주장
61/418,391 2010년11월30일 미국(US)
61/529,877 2011년08월31일 미국(US)
(56) 선행기술조사문헌
Invest Ophthalmol Vis Sci. 2010
Oct;51(10):4884-91.
Genome Res. 2001 Jun;11(6):981-93

(73) 특허권자
더 차이니스 유니버시티 오브 홍콩
중국 홍콩 엔. 티. 새턴 더 차이니스 유니버시티
오브 홍콩
(72) 발명자
로 욱 밍 데니스
중국 홍콩 코우룬 호만틴 7 킹 탁 스트리트 4층
찬 판 치
중국 홍콩 코우룬 마이 푸 선 추엔 브로드웨이 스
트리트 블록 34 플랫 에이 13/에프
(뒷면에 계속)
(74) 대리인
김진희

전체 청구항 수 : 총 23 항

심사관 : 이준혁

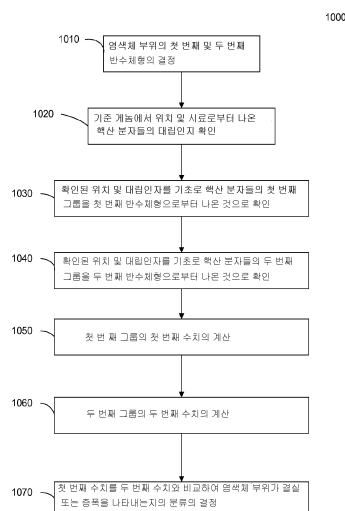
(54) 발명의 명칭 암과 연관된 유전적 또는 분자적 이상들의 검출

(57) 요약

본 발명은 유기체로부터 나온 생물학적 시료에서 유전적 또는 분자적 이상들 (aberrations)을 결정하는 시스템들, 장치, 및 방법들이 제공된다. 무세포 DNA 단편들을 포함하는 생물학적 시료들이, 예로 종양에서 결실들 및/또는 증폭들로 인한 염색체 부위들에서 불균형들을 확인하도록 분석된다. 복수의 좌위들이 각 염색체 부

(뒷면에 계속)

대표도 - 도10



위로 사용된다. 다음으로 이러한 불균형들은 환자를 암에 대해 진단 (검색)할 뿐만 아니라 암을 가진 환자를 예후 판단 (prognosticate)하거나, 환자에서 전암성 병태의 존재를 검출하거나 이의 진행을 감시하는 데 사용될 수 있다. 불균형의 중증도뿐만 아니라 불균형을 나타내는 부위들의 수도 사용될 수 있다. 게놈의 비-중복 분절들의 체계적인 분석은 시료를 위한 일반적인 검색 도구를 제공할 수 있다. 추가적으로, 환자는 진행을 (예로, 치료 이후에) 검색하고 예후를 판단할 뿐만 아니라 감시하는 것을 허용하는 하나 이상의 염색체 부위들 각각의 중증도 및 염색체 부위들의 수를 추적하도록 계속 (over time) 테스트될 수 있다.

(72) 발명자

치우 와이 쿤 로사

중국 홍콩 뉴 테리토리즈 샤틀린 더블 하벤 52 마 록
페트 하우스 31

지앙 페이용

중국 홍콩 뉴 테리토리즈 샤틀린 프린스 오브 웨일즈
하스퍼탈 30-32 연간 썽 스트리트 스태프 쿼터즈
플랫 1106 블록 에이

명세서

청구범위

청구항 1

유기체의 생물학적 시료를 암과 연관된 염색체 결실들 또는 증폭들에 대해 분석하는 방법으로서, 생물학적 시료가 정상세포들로부터 및 잠재적으로 암과 연관된 세포들로부터 기원하는 핵산 분자들을 포함하고, 상기 핵산 분자들 중 적어도 일부는 상기 생물학적 시료에서 무세포 상태로 있으며,

하나 이상의 프로그램화된 프로세서를 포함하는 컴퓨터 시스템에 의하여 첫 번째 다수의 좌위들을 포함하는 첫 번째 염색체 부위에서 유기체의 정상 세포들에 대한 첫 번째 및 두 번째 반수체형들을 결정하는 단계로서, 첫 번째 및 두 번째 반수체형들이 첫 번째 다수의 좌위들 각각에서 이중접합인 단계;

생물학적 시료 중의 다수의 핵산 분자들 각각에 대해, 하나 이상의 프로그램화된 프로세서를 포함하는 컴퓨터 시스템에 의하여 유기체의 기준 게놈에서 핵산 분자의 위치를 확인하고, 하나 이상의 프로그램화된 프로세서를 포함하는 컴퓨터 시스템에 의하여 핵산 분자의 각각의 대립인자를 결정하는 단계;

하나 이상의 프로그램화된 프로세서를 포함하는 컴퓨터 시스템에 의하여 확인된 위치들 및 결정된 대립인자들에 기초하여 핵산 분자들의 첫 번째 그룹을 첫 번째 반수체형으로부터 나온 것으로서 확인하는 단계로서, 상기 첫 번째 그룹은 첫 번째 다수의 좌위들 각각에 위치하는 적어도 하나의 핵산 분자를 포함하는 것인 단계;

하나 이상의 프로그램화된 프로세서를 포함하는 컴퓨터 시스템에 의하여 확인된 위치들 및 결정된 대립인자들에 기초하여 핵산 분자들의 두 번째 그룹을 두 번째 반수체형으로부터 나온 것으로서 확인하는 단계로서, 상기 두 번째 그룹은 첫 번째 다수의 좌위들 각각에 위치하는 적어도 하나의 핵산 분자를 포함하는 것인 단계;

하나 이상의 프로그램화된 프로세서를 포함하는 컴퓨터 시스템에 의하여 핵산 분자들의 첫 번째 그룹의 첫 번째 수치를 계산하는 단계로서, 상기 첫 번째 수치는 첫 번째 그룹의 핵산 분자들의 수 또는 크기를 정의하며, 상기 핵산 분자들의 크기는 핵산 분자의 길이, 질량 또는 무게인 것인 단계;

하나 이상의 프로그램화된 프로세서를 포함하는 컴퓨터 시스템에 의하여 핵산 분자들의 두 번째 그룹의 두 번째 수치를 계산하는 단계로서, 상기 두 번째 수치는 두 번째 그룹의 핵산 분자들의 수 또는 크기를 정의하며, 상기 핵산 분자들의 크기는 핵산 분자의 길이, 질량 또는 무게인 것인 단계; 및

하나 이상의 프로그램화된 프로세서를 포함하는 컴퓨터 시스템에 의하여 첫 번째 수치를 두 번째 수치와 비교하여 첫 번째 염색체 부위가 암과 연관된 임의의 세포들에서 결실 또는 증폭을 나타내는지 여부의 분류를 결정하는 단계

를 포함하는 분석 방법.

청구항 2

제1항에 있어서, 첫 번째 염색체 부위는 암과 연관된 세포들에서 결실 또는 증폭을 나타내는 것으로 분류되고, 암과 연관된 세포들은 악성 종양 및/또는 전암성 병변을 포함하는 것인 분석 방법.

청구항 3

제1항에 있어서, 첫 번째 다수의 좌위들의 각 좌위가 첫 번째 다수의 좌위들의 또 다른 좌위로부터 적어도 500 개 염기들이 떨어져 있는 것인 분석 방법.

청구항 4

제1항에 있어서, 상기 비교는 첫 번째 수치와 두 번째 수치 간의 차이 또는 비율을 결정하는 것을 포함하고, 분류는 상기 차이 또는 비율을 적어도 하나의 역치 수치와 비교하는 것을 포함하며, 역치 수치는 초과 또는 미만 시 특정한 분류가 적용되는 수치인 것인 분석 방법.

청구항 5

제4항에 있어서, 상기 적어도 하나의 역치 수치는 건강한 유기체로부터 또는 결실 또는 증폭을 가지지 않는 부위로부터 유래하고, 순차 확률비 테스트, t-테스트, 또는 카이-제곱 테스트가 적어도 하나의 역치 수치를 결정하는 데 사용되는 것인 분석 방법.

청구항 6

제1항에 있어서, 상기 첫 번째 수치는 첫 번째 그룹의 핵산 분자들의 크기 분포의 통계학적 수치에 해당하고, 상기 두 번째 수치는 두 번째 그룹의 핵산 분자들의 크기 분포의 통계학적 수치에 해당하며, 상기 크기 분포는 핵산 분자의 길이, 질량 또는 무게인 것인 분석 방법.

청구항 7

제6항에 있어서, 상기 첫 번째 수치는 첫 번째 그룹의 핵산 분자들의 평균 크기이고, 상기 두 번째 수치는 두 번째 그룹의 핵산 분자들의 평균 크기인 분석 방법.

청구항 8

제6항에 있어서, 상기 첫 번째 수치 Q_{HapI} 는 첫 번째 그룹에서 컷오프 크기보다 짧은 핵산 분자들의 분율이고, 상기 두 번째 수치 Q_{HapII} 는 두 번째 그룹에서 컷오프 크기보다 짧은 핵산 분자들의 분율인 분석 방법.

청구항 9

제8항에 있어서, 상기 비교는 첫 번째 수치와 두 번째 수치 간의 차이를 결정하는 것을 포함하고, 상기 차이는 $\Delta Q = Q_{HapI} - Q_{HapII}$ 이며, 역치 수치보다 큰 ΔQ 의 양의 수치는 두 번째 반수체형이 유기체의 종양 조직에서 결실을 포함하거나 첫 번째 반수체형이 유기체의 종양 조직에서 증폭을 포함하는 것을 나타내는 것인 분석 방법.

청구항 10

제8항에 있어서, 상기 비교는 첫 번째 수치와 두 번째 수치 간의 차이를 결정하는 것을 포함하고, 상기 차이는 $\Delta Q = Q_{HapI} - Q_{HapII}$ 이며, 역치 미만인 ΔQ 의 수치는 첫 번째 염색체 부위에 결실 또는 증폭이 존재하지 않는 것을 나타내는 것인 분석 방법.

청구항 11

제6항에 있어서, 상기 첫 번째 수치 F_{HapI} 및 두 번째 수치 F_{HapII} 는 각각의 반수체형에 대해 $F = \Sigma^w \text{길이} / \Sigma^N \text{길이}$ 로서 정의되고, 여기에서 Σ^w 길이는 컷오프 크기 w 이하의 길이를 가진 해당 그룹의 핵산 분자들의 길이의 합을 나타내고, 또한 Σ^N 길이는 N 개 염기 이하의 길이를 가진 해당 그룹의 핵산 분자들의 길이의 합을 나타내며, N 은 w 보다 큰 것인 분석 방법.

청구항 12

제11항에 있어서, 상기 비교는 첫 번째 수치와 두 번째 수치 간의 차이를 결정하는 것을 포함하고, 상기 차이는 $\Delta F = F_{HapI} - F_{HapII}$ 이며, 역치 수치보다 큰 ΔF 의 양의 수치는 두 번째 반수체형이 유기체의 종양 조직에서 결실을 포함하는 것을 나타내는 것인 분석 방법.

청구항 13

제11항에 있어서, 상기 비교는 첫 번째 수치와 두 번째 수치 간의 차이를 결정하는 것을 포함하고, 상기 차이는 $\Delta F = F_{HapI} - F_{HapII}$ 이며, 역치 수치보다 큰 ΔF 의 음의 수치는 두 번째 반수체형이 유기체의 종양 조직에서 증폭을 포함하는 것을 나타내는 것인 분석 방법.

청구항 14

제1항에 있어서, 상기 첫 번째 그룹의 첫 번째 수치는 첫 번째 다수의 좌위들에 위치한 핵산 분자들의 수에 해당하고, 상기 두 번째 그룹의 두 번째 수치는 첫 번째 다수의 좌위들에 위치한 핵산 분자들의 수에 해당하는 것

인 분석 방법.

청구항 15

제14항에 있어서, 첫 번째 수치와 두 번째 수치의 비율을 계산하여 생물학적 시료에서 암 DNA의 분획 농도를 결정하는 단계를 더 포함하는 분석 방법.

청구항 16

제15항에 있어서,

하나 이상의 프로그램화된 프로세서를 포함하는 컴퓨터 시스템에 의하여 생물학적 시료에서 암 DNA의 분획 농도를 다수의 시점에서 결정하는 단계; 및

다수의 시점에서의 분획 농도를 사용하여 하나 이상의 프로그램화된 프로세서를 포함하는 컴퓨터 시스템에 의하여 유기체에서 암 수준을 진단하거나, 병기를 결정하거나, 예후를 판단하거나, 진행을 모니터링하는 단계

를 더 포함하는 분석 방법.

청구항 17

제1항에 있어서,

다수의 다른 염색체 부위들 각각에 대해,

제1항의 방법을 반복하는 단계;

결실 또는 증폭을 나타내는 염색체 부위들의 첫 번째 수를 결정하는 단계; 및

상기 첫 번째 수를 하나 이상의 역치 수치와 비교하여 유기체에서 암의 수준을 결정하며, 상기 역치 수치는 초과 또는 미만 시 특정한 분류가 적용되는 수치이고, 상기 암의 수준은 암의 존재, 암의 단계, 종양의 크기, 염색체 부위의 결실들 또는 증폭들의 수 및 암의 중증도의 기타 척도에서 선택되는 것인 단계

를 더 포함하는 분석 방법.

청구항 18

제17항에 있어서,

결실 또는 증폭을 나타내는 것으로서 확인된 각 염색체 부위에 대해 결실 또는 증폭의 양을 결정하는 단계; 및

상기 양을 하나 이상의 역치 수치와 비교하여 유기체에서 암의 수준을 결정하는 단계

를 더 포함하는 분석 방법.

청구항 19

제17항에 있어서,

제17항의 방법을 다수의 시점에서 반복하는 단계; 및

하나 이상의 프로그램화된 프로세서를 포함하는 컴퓨터 시스템에 의하여 상기 다수의 시점에서의 상기 첫 번째 수를 사용하여 유기체에서 암의 수준을 진단하거나, 병기를 결정하거나, 예후를 판단하거나, 진행을 모니터링하는 단계

를 더 포함하는 분석 방법.

청구항 20

제19항에 있어서, 상기 다수의 시점에서의 상기 첫 번째 수를 사용하여 유기체에서 암의 수준을 진단하거나, 병기를 결정하거나, 예후를 판단하거나, 진행을 모니터링하는 단계는 전암성 병태의 존재 또는 진행을 결정하는 것을 포함하는 것인 분석 방법.

청구항 21

제17항에 있어서, 각 염색체 부위가 선결정된 길이인 분석 방법.

청구항 22

실행시에 제1항 내지 제21항 중 어느 한 항의 방법을 수행하도록 컴퓨터 시스템을 제어하는 다수의 명령들을 저장하는 비전송형 컴퓨터 판독가능한 기록 매체.

청구항 23

제22항의 비전송형 컴퓨터 판독가능한 기록 매체를 포함하는 컴퓨터 시스템.

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

발명의 설명

기술 분야

[0001]

본 발명은 유기체(생물)로부터 나온 생물학적 시료에서 유전적 또는 분자적 이상들 (aberrations)을 결정하는 시스템들, 장치, 및 방법들이 제공된다. 무세포 DNA 단편들을 포함하는 생물학적 시료들이, 예로 종양에서 결실들 및/또는 증폭들로 인한 염색체 부위들에서 불균형들을 확인하도록 분석된다. 복수의 좌위들이 각 염색체 부위로 사용된다. 다음으로 이러한 불균형들은 환자를 암에 대해 진단 (검색)할 뿐만 아니라 암을 가진 환자를

예후 판단 (prognosticate)하거나, 환자에서 전암성 병태의 존재를 검출하거나 이의 진행을 감시하는 데 사용될 수 있다. 불균형의 중증도뿐만 아니라 불균형을 나타내는 부위들의 수도 사용될 수 있다. 게놈의 비-중복 분절들의 체계적인 분석은 시료를 위한 일반적인 검색 도구를 제공한다. 추가적으로, 환자는 진행을 (예로, 치료 이후에) 검색하고 예후를 판단할 뿐만 아니라 감시하는 것을 허용하는 하나 이상의 염색체 부위들 각각의 중증도 및 염색체 부위들의 수를 추적하도록 계속 (over time) 테스트될 수 있다.

배경 기술

- [0002] 암은 많은 사람들을 침범하는 공통적인 질환이다. 종종 암은 심각한 증상들의 소견을 보일 때까지 확인되지 않는다. 공통적인 유형들의 암의 경우, 암을 가질 수 있는 환자들을 확인하는 검색 기법들이 있다. 그러나, 이러한 기법들은 종종 신뢰할만하지 못하거나 환자를 방사선 조사에 놓이게 한다. 다른 유형들의 많은 암들의 경우는, 효과적인 검색 기법들이 전혀 없다.
- [0003] 이형접합의 소실 (LOH)이 폐, 및 두경부 암들을 앓고 있는 환자들의 순환 DNA (circulating DNA)에 있는 특정한 좌위의 경우에 검출된 바 있었다 (Chen XQ, et al. Nat Med 1996; 2: 1033-5; Nawroz H, et al. Nat Med 1996; 2: 1035-7). 그러나, 이러한 기법들은 검출가능하였던 LOH의 양이 비교적 적음으로 인해 특정한 좌위를 조사하는 것이 방해로 되었다. 디지털 PCR을 사용할 때에도, 이들 방법들은 적은 양들의 LOH를 검출하는 무능력으로 인해 여전히 실패하고 있다. 게다가, 이러한 기법들은 특이적 유형의 암에서 발생하는 것으로 알려진 특정한 좌위를 연구하는 데는 여전히 한계가 있었다. 따라서, 보편적으로 암을 검색하는 것은 가능하지 않거나 효과적이지 않았다.
- [0004] 암의 존재를 검색하는 것 이외에도, 최신의 기법들은 암을 가진 환자의 예후를 판단하고 치료의 효과들을 감시하는 데 있어서는 부족하다 (예로, 수술 또한 화학요법 또는 면역요법 또는 표적화 요법 이후의 회복). 이러한 기법들은 종종 비싸거나 (예로, 영상화 기법들), 부정확하거나, 비효과적이거나, 민감도가 떨어지거나 영상화 기법들에 사용되는 방사선 조사에 환자를 노출시킬 수 있다.
- [0005] 따라서, 환자에게 암을 검색하고, 예후를 판단하고 또한 감시하는 새로운 기법들을 제공하는 것이 바람직하다.

발명의 내용

해결하려는 과제

- [0006] 본 출원은 2010년 11월 30일자로 제출된 "암과 연관된 유전적 이상들의 검출 (Detection of genetic aberrations associated with cancer)"이라는 명칭을 가지는 미국 특허출원 제 61/418,391호, 및 2011년 8월 31일자로 제출된 "암과 연관된 유전적 또는 분자적 이상들의 검출 (Detection of genetic or molecular aberrations associated with cancer)"이라는 명칭을 가지는 미국 특허출원 제 61/529,877호로부터 우선권을 주장하고 또한 이들의 비-예비적 출원이며, 그들 내용들 전부가 본 명세서에서 모든 목적들을 위한 참고문헌으로 통합되어 있다.
- [0007] 본 출원은 공동 소유되어 있는 2010년 11월 5일자로 제출된 로 등 (Lo et al.)에 의한 "크기-기초 게놈 분석 (Size-Based Genomic Analysis)"(변리사 서류번호 제 80015-794101/006610US호)이라는 명칭을 가지는 미국 특허출원 제 12/940,992호 (미국 공고번호 제 2011/0276277호), 및 2010년 11월 5일자로 제출된 로 등에 의한 "모성의 생물학적 시료로부터 태아 게놈 분석 (Fetal Genomic Analysis From A Maternal Biological Sample)"(변리사 서류번호 제 80015-794103/006710US호)이라는 명칭을 가지는 미국 특허출원 제 12/940,993호 (미국 공고번호 제 2011/0105353호)에 관한 것이고, 그들의 기재내용들은 전부가 참고문헌으로 통합되어 있다.
- [0008] 구현예들은 암과 연관된 이상들 (aberrations)을 결정하는 시스템들, 장치, 및 방법들을 제공한다. 무세포 DNA (cell-free DNA) 단편들을 포함하는 생물학적 시료들이, 예로 종양에서 결실들 및/또는 증폭들로 인한 염색체 부위들에서 불균형들 (imbalances)을 확인하도록 분석된다. 복수의 좌위들을 가지는 염색체 부위를 사용하는 것은 더 큰 효율 및/또는 정확도를 허용할 수 있다. 다음으로 이러한 불균형들은 환자를 암에 대해 진단하거나 검색할 뿐만 아니라 암을 가진 환자를 예후 판단 (prognosticate)하는 데 사용될 수 있다. 불균형의 중증도 (severity)뿐만 아니라 불균형을 나타내는 부위들의 수도 사용될 수 있다. 추가적으로, 환자는 진행을 (예로, 치료 이후) 검색하고 예후를 판단할 뿐만 아니라 감시하는 것을 허용하는 하나 이상의 염색체 부위들 각각의 중증도 및 염색체 부위들의 수를 추적하도록 계속 (over time) 테스트될 수 있다.

[0009] 한 가지 구현예에 따르면, 암과 연관된 염색체 결실들 또는 증폭들에 대해 유기체(organism)의 생물학적 시료를 분석하는 방법이 제공된다. 생물학적 시료는 정상 세포들로부터, 및 잠재적으로는 암과 연관된 세포들로부터 기원한 핵산 분자들을 포함한다. 적어도 일정의(some) 핵산 분자들은 시료에서 무세포 상태로 존재한다. 첫 번째 및 두 번째 반수체형들(haplotypes)이 첫 번째 염색체 부위에서 유기체의 정상 세포들을 위해 결정된다. 첫 번째 염색체 부위는 첫 번째 다수의 이중접합 좌위들을 포함한다. 시료에서 각(each) 다수의 핵산 분자들은 확인된 유기체의 기준 게놈에서 위치를 가지고 또한 결정된 개별(respective) 대립인자를 가진다. 위치들 및 결정된 대립인자들은 첫 번째 반수체형으로부터 나온 핵산 분자들의 첫 번째 그룹 및 두 번째 반수체형으로부터 나온 핵산 분자들의 두 번째 그룹을 결정하는 데 사용된다. 컴퓨터 시스템은 첫 번째 그룹의 첫 번째 수치 및 두 번째 그룹의 두 번째 수치를 계산한다. 각 수치는 핵산 분자들의 개별 그룹의 특성을 정의하였다 (예로, 그룹에서 분자들의 평균 크기 및 수). 첫 번째 수치는 첫 번째 염색체 부위가 암과 연관된 임의의(any) 세포들에서 결실 또는 증폭을 나타내지는 여부의 분류를 결정하도록 두 번째 수치와 비교된다.

[0010] 또 다른 구현예에 따르면, 유기체의 생물학적 시료를 분석하는 방법이 제공된다. 생물학적 시료는 정상 세포들로부터, 및 잠재적으로는 암과 연관된 세포들로부터 기원한 핵산 분자들을 포함한다. 적어도 일정의 핵산 분자들은 시료에서 무세포 상태로 존재한다. 유기체의 다수의 비-중복 염색체 부위들이 확인된다. 각 염색체 부위는 다수의 좌위들을 포함한다. 시료에서 각 다수의 핵산 분자들은 확인된 유기체의 기준 게놈에서 위치를 가진다. 각 염색체 부위를 위해, 핵산 분자들의 개별 그룹은 확인된 위치들을 기초로 하여 염색체 부위로부터 나온 것으로서 확인된다. 각 개별 그룹은 염색체 부위의 각 다수의 좌위들에 위치한 적어도 하나의 핵산 분자를 포함한다. 컴퓨터 시스템은 개별 그룹의 핵산 분자들의 특성을 정의하는 개별 그룹의 개별 수치를 계산한다. 개별 수치는 염색체 부위가 결실 또는 증폭을 나타내지는 여부의 분류를 결정하도록 기준 수치와 비교된다. 다음으로 결실 또는 증폭을 나타내는 것으로서 분류된 염색체 부위들의 양이 결정된다.

[0011] 또 다른 구현예에 따르면, 정상 세포들로부터, 및 잠재적으로는 암과 연관된 세포들로부터 기원한 핵산 분자들을 포함하는 생물학적 시료들을 사용하여 유기체에서 염색체 이상들의 진행을 결정하는 방법이 제공된다. 적어도 일정의 핵산 분자들은 생물학적 시료들에서 무세포 상태로 존재한다. 하나 이상의 비-중복 염색체 부위들이 유기체의 기준 게놈을 위해 확인된다. 각 염색체 부위는 다수의 좌위들을 포함한다. 서로 다른 시간대에서 채취된 시료들은 진행을 결정하도록 분석된다. 시료를 위해, 시료에 있는 각 다수의 핵산 분자들은 확인된 유기체의 기준 게놈에서 위치를 가진다. 각 염색체 부위를 위해, 핵산 분자들의 개별 그룹은 확인된 위치들을 기초로 하여 염색체 부위로부터 나온 것으로서 확인된다. 각 개별 그룹은 염색체 부위의 각 다수의 좌위들에 위치한 적어도 하나의 핵산 분자를 포함한다. 컴퓨터 시스템은 핵산 분자들의 개별 그룹의 개별 수치를 계산한다. 개별 수치는 개별 그룹의 핵산 분자들의 특성을 정의한다. 개별 수치는 첫 번째 염색체 부위가 결실 또는 증폭을 나타내지는 여부의 분류를 결정하도록 기준 수치와 비교된다. 다음으로 다수의 시간대에서의 각 염색체 부위들의 분류들은 유기체에서 염색체 이상들의 진행을 결정하는 데 사용된다.

[0012] 본 발명의 다른 구현예들은 본 명세서에서 기술된 방법들과 연관된 시스템들, 휴대용 사용자 장치들, 및 컴퓨터 관독가능한 매체들을 지향한다.

[0013] 본 발명의 본질 및 장점들의 보다 나은 이해는 다음의 상세한 설명 및 첨부된 도면들을 참고하여 구할 수 있다.

파제의 해결 수단

[0014] 정의들

[0015] 본 명세서에서 사용되는 바, 용어 "생물학적 시료 (biological sample)"는 개체 (예로, 인간, 암을 가진 사람, 암을 가진 것으로 의심되는 사람, 또는 기타 유기체들)로부터 채취되고 하나 이상의 관심 있는 핵산 분자들을 포함하는 시료라면 모두를 말한다.

[0016] 용어 "핵산 (nucleic acid)" 또는 "폴리뉴클레오타이드 (polynucleotide)"는 단일- 또는 이중가닥 둘 중 하나의 형태로 데옥시리보핵산 (DNA) 또는 리보핵산 (RNA) 및 그의 중합체를 말한다. 특정하게 제한되지 않는 경우라면, 본 용어는 기준 핵산과 유사한 결합 특성들을 가지고 자연적으로 발생하는 뉴클레오타이드들과 유사한 방식으로 대사되는 천연의 뉴클레오타이드들의 기지의 유사체들을 포함하는 핵산들을 포괄한다. 달리 표시되지 않는 경우라면, 특정한 핵산 서열은 보존적으로 변형된 그의 변이체들 (예로, 중복성 (degenerate) 코돈 치환들), 대립인자들, 오소로그체 (orthologs), 단일 뉴클레오타이드 다형성들 (SNPs), 사본수 변이체들, 및 상보적 서열들뿐만 아니라 분명하게 표시된 서열도 역시 명확하게 포괄하고 있다. 상세하게, 중복성 코돈 치환들은 하나 이상의 선택된 (또는 모든) 코돈들의 세 번째 자리가 혼합-염기 및/또는 데옥시이노신 잔기들로 치환된

서열들을 생성하여 달성될 수 있다 (Batzer *et al.*, *Nucleic Acid Res.* **19**: 5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* **260**: 2605-2608 (1985); 및 Rossolini *et al.*, *Mol. Cell. Probes* **8**:91-98 (1994)). 용어 핵산은, 이에 제한되는 것은 아니지만 유전자, cDNA, mRNA, 작은 비코딩 RNA, 마이크로 RNA (miRNA), 피위-상호작용 (Piwi-interacting) RNA, 또한 유전자 또는 좌위에 의해 인코딩되는 짧은 헤어핀 RNA (shRNA):를 포괄한다.

[0017] 용어 "유전자 (gene)"는 폴리펩타이드 사슬 또는 전사된 RNA 산물을 생산하는 데 관여하는 DNA의 분절을 의미한다. 이것은 선행하는 부위들 및 이어지는 코딩 부위 (리더 및 트레일러)뿐만 아니라 개별적 코딩 분절들 (엑손들) 사이의 개입 서열들 (인트론들)을 포함할 수 있다.

[0018] 본 명세서에서 사용되는 바, 용어 "임상적으로 적절한 핵산 서열" 또는 "임상적으로 적절한 염색체 부위" (또는 테스트된 부위/분절)는 잠재적인 불균형이 테스트되고 있는 더 큰 게놈 서열의 분절 또는 더 큰 게놈 서열 자체에 해당하는 폴리뉴클레오타이드 서열을 말할 수 있다. 예들로는 결실 또는 증폭되거나 잠재적으로 결실 또는 증폭되는 (단순한 증폭을 포함) 게놈 분절들, 또는 분절의 소부위 (subregion)을 포함하는 더 큰 부위를 포함한다. 일정 구현예들에서, 복수의 임상적으로 적절한 핵산 서열, 또는 동등하게 임상적으로 적절한 핵산 서열의 복수의 마커들은 부위에서 불균형을 검출하는 데이터를 제공하는 데 사용될 수 있다. 예를 들어, 염색체 상의 5개 비연속 서열들로부터 나온 데이터는 가능한 불균형의 결정을 위한 추가적인 방식에 사용될 수 있고, 효과적으로 필요한 시료 부피를 1/5로 감소시킨다.

[0019] 본 명세서에서 사용되는 바, 용어 "기준 핵산 서열 (reference nucleic acid sequence)" 또는 "기준 염색체 부위 (reference chromosomal region)"는 테스트 부위에 대해 정량적 프로파일 또는 크기 분포를 비교하는 데 사용되는 핵산 서열을 말한다. 기준 핵산 서열의 예들로는 결실 또는 증폭을 포함하지 않는 염색체 부위, 게놈 전부 (예로, 전체 서열결정된 태그 계수들에 의한 정상화를 통한), 정상이라고 알려진 하나 이상의 시료들로부터 나온 부위 (테스트되는 시료의 경우 동일한 부위일 수 있음), 또는 염색체 부위의 특정한 반수체형을 포함한다. 이러한 기준 핵산 서열들은 시료에 내인적으로 존재하거나, 시료 가공 또는 분석 동안 외인적으로 첨가될 수 있다. 일정 구현예들에서, 기준 염색체 부위는 질환이 없는 건강한 상태를 대표하는 크기 프로파일을 보여준다. 보다 다른 구현예들에서, 기준 염색체 부위는 질환이 없는 건강한 상태를 대표하는 정량적 프로파일을 보여준다.

[0020] 본 명세서에서 사용되는 바, 용어 "기초로 하는 (based on)"은 "적어도 부분적으로는 기초로 하는"을 의미하고, 방법의 입력정보 및 방법의 출력정보의 관계에서 발생하는 것과 같은 또 다른 수치의 결정에 사용되는 하나의 수치 (또는 결과)를 말한다. 본 명세서에서 사용되는 바, 용어 "유래하다 (derive)"는 유래가 공식의 계산일 때 발생하는 것과 같은 방법의 입력정보 및 방법의 출력정보의 관계도 역시 말한다.

[0021] 본 명세서에서 사용되는 바, 용어 "매개변수 (parameter)"는 정량적 데이터 집합 및/또는 정량적 데이터 집합들 간의 수적인 관계를 특징 분석하는 수적인 수치를 의미한다. 예를 들어, 첫 번째 핵산 서열의 첫 번째 양 및 두 번째 핵산 서열의 두 번째 양의 비율 (또는 비율의 함수)이 매개변수이다.

[0022] 본 명세서에서 사용되는 바, 용어 "좌위 (locus)" 또는 그의 복수 형태 "좌위들 (loci)"은 게놈들을 통틀어 변이를 가질 수 있는 뉴클레오타이드들 (또는 염기쌍들)의 길이라면 모두의 위치 또는 주소이다.

[0023] 본 명세서에서 사용되는 바, 용어 "서열 불균형 (sequence imbalance)" 또는 "이상 (aberration)"은 임상적으로 적절한 염색체 부위의 정량에서 적어도 하나의 컷오프 수치에 의해 정의되는 바와 같은 기준 정량으로부터 유의한 편차라면 모두를 의미한다. 서열 불균형은 염색체 용량 불균형, 대립인자 불균형, 돌연변이 용량 불균형, 사본수 불균형, 반수체 용량 불균형, 및 기타 유사한 불균형들을 포함할 수 있다. 한 가지 예로서, 대립인자 불균형은 종양이 게놈에서 결실된 유전자의 하나의 대립인자 또는 증폭된 유전자의 하나의 대립인자 또는 두 개의 대립인자들의 차별적인 증폭을 가지고, 이로 인해 시료에 있는 특정한 좌위에서 불균형을 생성할 때 발생할 수 있다. 또 다른 예로서, 환자는 종양 억제 유전자 (tumor suppressor gene)에서 유전된 돌연변이를 가질 수 있다. 그러면 환자는 종양 억제 유전자의 비-변이된 대립인자가 결실된 종양을 계속해서 발생시킬 수 있다. 따라서, 종양 내에서는 돌연변이 용량 불균형이 존재한다. 종양이 환자의 혈장 내로 그의 DNA를 방출할 때, 종양 DNA는 혈장에서 환자의 전신발현 DNA (정상 세포들로부터 나온)와 혼합될 것이다. 본 명세서에서 기술된 방법들의 사용을 통하여, 혈장에 있는 본 DNA 혼합물의 변이성 용량 불균형은 검출될 수 있다.

[0024] 본 명세서에서 사용되는 바, 용어 "반수체형 (haplotype)"은 동일한 염색체 또는 염색체 부위 상에 다함께 전달되는 복수의 좌위들에 있는 대립인자들의 조합을 말한다. 반수체형은 한 쌍의 좌위들 만큼이나 작은 부위 또는 염색체 부위 또는 전체 염색체를 말한다. 용어 "대립인자들 (alleles)"는 서로 다른 표현형 형질들을 가져오는

나 가져올 수 없는, 동일한 생리학적 계층 좌위에서 대안의 DNA 서열들을 말한다. 각 염색체의 두 개 사본들 (수컷 인간 개체에서 성 염색체를 제외함)을 가진 특정한 이수성 (diploid) 유기체라면 모두에서, 각 유전자의 유전형은 동형접합체들에서는 동일하고 이형접합체들에서는 서로 다른, 해당 좌위에 존재하는 대립인자들의 쌍을 포함한다. 유기체들의 집단 또는 종은 전형적으로 다양한 개인들 중에 있는 각 좌위에서 복수의 대립인자들을 포함한다. 하나 이상의 대립인자가 집단에서 발견되는 계층 좌위는 다형성 부위라고 말한다. 좌위에서 대립인자 다양성 (variation)은 존재하는 대립인자들의 수 (예로, 다형성의 정도), 또는 집단에서 이형접합들의 비율 (예로, 이형접합율)로서 측정가능하다. 본 명세서에서 사용되는 바, 용어 "다형성 (polymorphism)"은 그의 빈도와는 상관없이 인간 계층에서 개인 간의 다양성이라면 모두를 말한다. 이러한 다양성의 예들로는, 이에 제한되는 것은 아니지만 단일 뉴클레오타이드 다형성, 단순 일렬 반복서열 다형성들, 삽입-결실 다형성들, 돌연변이들 (질환을 유발할 수 있음) 및 사본수 다양성들을 포함한다.

[0025] 용어 "서열결정된 태그 (sequenced tag)"는 핵산 분자, 예로 DNA 단편의 전부 또는 일부로부터 결정된 서열을 말한다. 종종, 단편의 한쪽 말단만, 예로 약 30 bp가 서열결정된다. 다음으로 서열결정된 태그는 기준 계층에 정렬될 수 있다. 임의적으로, 단편의 말단들 둘 다가 서열결정되어 두 개의 서열결정된 태그들이 생성될 수 있고, 이는 정렬에서 더 큰 정확도를 제공하고 단편의 길이도 역시 제공할 수 있다.

[0026] 용어 "보편적 서열결정 (universal sequencing)"은 어댑터들이 단편의 말단에 첨가되고, 서열결정을 위한 프라이머들이 어댑터들에 부착되는 서열결정을 말한다. 따라서, 단편이라면 모두가 동일한 프라이머들로 서열결정될 수 있고, 따라서 서열결정은 무작위적일 수 있다.

[0027] 용어 "크기 분포 (size distribution)"은 특정한 그룹에 해당하는 분자들 (특정한 반수체형으로부터 또는 특정한 염색체 부위로부터 나온 단편들)의 길이, 질량, 무게, 또는 크기의 척도를 대표하는 하나의 수치 또는 수치들의 집합이라면 모두를 말한다. 다양한 구현예들은 다양한 크기 분포들을 사용할 수 있다. 일정 구현예들에서, 크기 분포는 다른 염색체들의 단편들과 비교하여 하나의 염색체의 단편들의 크기들 (예로, 평균, 중앙값, 또는 평균값 (mean))의 순위매김에 관한 것이다. 다른 구현예들에서, 크기 분포는 염색체 단편들의 실제 크기들의 통계학적 수치에 관한 것일 수 있다. 한 가지 시행예에서, 통계학적 수치는 염색체 단편의 평균, 평균값 (mean), 또는 중앙값이라면 모두를 포함할 수 있다. 또 다른 시행예에서, 통계학적 수치는 컷오프 수치 미만의 단편들의 전체 길이를 포함할 수 있고, 이는 모든 단편들, 또는 적어도 더 큰 컷오프 수치 미만의 단편들의 전체 길이에 의해 구분될 수 있다.

[0028] 본 명세서에서 사용되는 바, 용어 "분류 (classification)"는 시료의 특정한 특성과 연관된 수(들) 또는 기타 특징(들)이라면 모두를 말한다. 예를 들어, "+" 기호 (또는 단어 "양성")은 시료가 결실들 또는 증폭들을 가지는 것으로 분류되는 점을 의미할 수 있다. 분류는 이분적 (예로, 양성 또는 음성)이거나 더 많은 분류 수준들을 가질 수 있다 (예로, 1로부터 10까지 또는 0 내지 1의 규모). 용어 "컷오프 (cutoff)" 및 "역치 (threshold)"는 적용 (operation) 시 사용되는 선결정된 수를 말한다. 예를 들어, 컷오프 크기는 초과 시 단편들이 배제되는 크기를 말할 수 있다. 역치 수치는 초과 또는 미만 시 특정한 분류가 적용되는 수치일 수 있다. 이들 용어들의 둘 중 하나는 이들 문맥들의 둘 중 하나에서 사용될 수 있다.

[0029] 용어 "암의 수준 (level of cancer)"은 암이 존재하는지 여부, 암의 단계, 종양의 크기, 염색체 부위의 결실들 또는 증폭들이 관여하는 횟수 (예로, 중복 또는 삼중복), 및/또는 기타 다른 암의 중증도의 척도를 말할 수 있다. 암의 수준은 수 또는 기타 특징들일 수 있다. 수준은 제로일 수 있다. 암의 수준은 결실들 또는 증폭들과 연관된 전악성 (premalignant) 또는 전암성 (precancerous) 병태도 역시 포함한다.

[0030] 양성 조직 (종양)은 염색체 부위의 결실 또는 증폭과 같은 이상들을 가질 수 있다. 종양은 신체에 있는 체액들 내로 DNA 단편들을 방출할 수 있다. 구현예들은 염색체 부위에서 DNA의 정상 (기대) 수치들과 대비하여 이상들을 확인하도록 DNA 단편들을 분석하여 종양을 확인할 수 있다.

[0031] 결실 또는 증폭의 정확한 크기뿐만 아니라 위치는 달라질 수 있다. 특정한 부위가 암들 또는 특정한 유형의 암에 대해 일반적으로 이상을 나타내는 것으로 (이로 인해 특정한 암의 진단을 유도함) 알려져 있는 시간대가 존재할 수 있다. 특정한 부위가 알려져 있지 않을 때, 계층 전부 또는 계층의 큰 부분들을 분석하는 체계적인 방법이 계층 전체에 분산될 수 있고 크기가 (예로, 결실되거나 증폭된 염기들의 수)가 다양한 이상 (aberrant) 부위들을 검출하는 데 사용될 수 있다. 본 염색체 부위(들)은 이상의 중증도 또는 이상을 보이는 부위들의 수에서의 변화들을 확인하도록 계속 추적될 수 있다. 본 추적은 종양을 검색하고, 예후를 판단하고, 감시하는 (예

로, 치료 이후 또는 재발 또는 종양 진행을 검출하는) 핵심 정보를 제공할 수 있다.

[0032] 본 명세서는 먼저 암에서 염색체 이상들의 예들로 시작한다. 다음으로, 생물학적 시료에서 무세포 DNA를 검출하고 분석하여 염색체 이상을 검출하는 방식들의 예들이 논의된다. 일단 하나의 염색체 부위에서 이상을 검출하는 방법들이 확인되고 나면, 체계적인 방식으로 환자들을 검색 (진단)하고 예후를 판단하는 데 사용되는 많은 염색체 부위들에서 이상들을 검출하는 방법들이 기술된다. 본 명세서는 환자들을 검색하고 예후를 판단하고 감시하는 것을 계속 제공하도록 하나 이상의 부위들에서 염색체 이상에 대한 테스트들로부터 획득된 수적인 지시자들 (indicators)을 추적하는 방법들도 역시 기술하고 있다. 예들은 다음에 논의된다.

I. 암에서 염색체 이상들의 예들

[0034] 염색체 이상들은 암 세포들에서 보편적으로 검출된다. 게다가, 염색체 이상들의 특징적인 양상은 선택된 유형의 암에서 발견될 수 있다. 예를 들어, 염색체 팔들 1p, 1q, 7q, 15q, 16p, 17q 및 20q에서 DNA의 증가들 또한 3p, 4q, 9p 및 11q에서 DNA의 소실들은 간세포 암종 (HCC)에서 공통적으로 검출된다. 이전의 연구들은 이러한 유전적 이상들이 암 환자들의 순환 DNA에서도 역시 검출될 수 있는 점을 기술하여 왔다. 예를 들어, 이종접합 (LOH)의 소실이 폐, 및 두경부 암들을 앓고 있는 환자들의 순환 DNA 분자들에 있는 특정한 좌위에 대해 검출되었다 (Chen XQ, et al. Nat Med 1996; 2: 1033-5; Nawroz H, et al. Nat Med 1996; 2: 1035-7). 혈장 또는 혈청에서 검출되는 유전적 변형들이 종양 조직들에서 발견되는 것들과 부합하였다. 그러나, 종양-유래 DNA는 단지 전체 순환 무세포 DNA의 소수 분획에 기여하기 때문에, 종양 세포들의 LOH에 의해 초래되는 대립인자 불균형은 보통 작다. 많은 연구자들이 순환 DNA 분자들 중에서 좌위의 서로 다른 대립인자들의 정확한 정량을 위해 (Chang HW, et al. J Natl Cancer Inst. 2002; 94: 1697-703), 디지털 중합효소 연쇄반응 (PCR) 기술학을 개발하였다 (Vogelstein B, Kinzler KW. Proc Natl Acad Sci U S A. 1999; 96: 9236-41; Zhou W, et al. Nat Biotechnol 2001; 19: 78-81; Zhou W, et al. Lancet. 2002; 359: 219-25). 디지털 PCR은 종양 DN에 있는 특정한 좌위에서 LOH에 의해 초래되는 작은 대립인자 불균형을 검출하는 실시간 PCR 또는 기타 DNA 정량 방법들보다 더욱 더 민감하다. 그러나 디지털 PCR은 특정한 좌위에서 매우 작은 대립인자 불균형을 확인하는 데 여전히 어려움들을 가질 수 있고, 따라서 본 명세서에서 기술된 구현예들은 조합적 방식으로 염색체 부위들을 분석하고 있다.

[0035] 본 명세서에서 기술된 기술학은 비악성 또는 비암성 병태의 검출을 위한 적용들도 역시 가진다. 이러한 병태의 예들로는 간의 경변 및 자궁경부 상피내종양을 포함한다. 전자의 병태는 간세포 암종의 비악성 병태인 한편 후자의 병태는 자궁경부 암종의 비악성 병태이다. 이러한 비악성 병태들은 악성 종양이 되도록 그들의 진화에서 여러 분자적 변형들을 이미 소유하는 것으로 보고되어 왔다. 예를 들어, 염색체 팔들 1p, 4q, 13q, 18q에서 LOH의 존재 및 3개 이상의 좌위에서 동시적인 소실들은 간경변을 가진 환자들에서 HCC 발달의 증가된 위험과 연관되어 있다 (Roncalli M et al. Hepatology 2000; 31: 846-50). 이러한 비악성 병변들은 더 낮은 농도들일 수는 있더라도 순환 내로 DNA도 역시 방출할 것이다. 기술학은 혈장에서 DNA 단편들을 분석하여 결실들 또는 증폭들의 검출을 허용하고, 혈장에서 순환하는 비악성 DNA의 농도 (분획 농도를 포함함)를 측정하도록 허용할 수 있다. 이러한 변형들이 검출되는 용이성 (예로, 서열결정의 깊이 또는 검출되는 이러한 변화들의 수) 및 농도들은 만기 중만 암성 병태로의 진행 가능성 또는 신속도를 예측할 것이다.

A. 염색체 부위의 결실

[0037] 도 1은 결실의 이상을 보여주는 암세포의 염색체 부위를 도시한다. 정상 세포는 두 개의 반수체형들, Hap I 및 Hap II로 관찰된다. 관찰된 바와 같이, Hap I 및 Hap II 둘 다는 각 다수의 이종접합 좌위들 110에서 서열을 가진다 (단일 뉴클레오타이드 다형성 SNPs라고도 함). 암과 연관된 세포에서, Hap II은 결실된 염색체 부위 120을 가진다. 예로서, 암과 연관된 세포는 종양 (예로, 악성 종양)으로부터, 종양의 전이성 초점 (예로, 국부적 림프절, 또는 먼 기관에서)으로부터, 또는 비암성 또는 비악성 병변으로부터, 예로 상기 언급된 바와 같이 나올 수 있다.

[0038] 두 개의 동종 반수체형들의 하나가 결실된 암세포의 염색체 부위 120에서, 모든 이종접합 SNPs 110은 해당하는 결실된 동종의 염색체 상에서 다른 대립인자의 소실 때문에 동종접합으로서 나타날 것이다. 따라서, 이러한 유형의 염색체 이상은 이종접합의 소실 (LOH)이라고 부른다. 부위 120에서, 이들 SNPs의 비-결실된 대립인자들은 정상 조직들에서 발견될 수 있는 두 개의 반수체형들의 하나를 대표할 수 있다. 도 1에 나타난 예에서, LOH

부위 120 에서 반수체형 I (Hap I)은 종양 조직을 유전형 분석하여 결정될 수 있다. 다른 반수체형 (Hap II)은 정상 조직들 및 암 조직들의 외관상 유전형들을 비교하여 결정될 수 있다. Hap II은 모든 결실된 대립인자들을 결합시켜서 제작될 수 있다. 즉, 암세포의 부위 120에서는 출현하지 않는 부위 120을 위한 정상 세포에서의 모든 대립인자들은 동일한 반수체형, 예로 Hap I 상에 있도록 결정된다. 본 분석을 통하여, 환자들 (예로, 간세포 암종 HCC 환자들)의 반수체형들은 종양 조직에서 LOH를 보이는 모든 염색체 부위들에 대해 결정될 수 있다. 이러한 방법은 개체가 암세포들을 가지는 경우에만 유용하고, 부위 120에서 반수체형을 결정하기 위해서만 작용하지만, 결실된 염색체 부위의 좋은 도시를 제공한다.

[0039] B. 염색체 부위의 증폭

[0040] 도 2는 증폭의 이상을 보여주는 암세포의 염색체 부위를 도시한다. 정상 세포는 두 개의 반수체형들, Hap I 및 Hap II로 관찰된다. 관찰된 바와 같이, Hap I 및 Hap II 둘 다는 각 다수의 이중접합 좌위들 110에서 서열을 가진다. 종양 세포에서, Hap II은 2배로 증폭된 (복제된) 염색체 부위 220을 가진다.

[0041] 유사하게, 종양 조직들에서 단일-대립인자 증폭을 가지는 부위들의 경우 SNPs 210에서 증폭된 대립인자들은 마이크로어레이 (microarray) 분석과 같은 방법들에 의해 검출될 수 있다. 두 개 반수체형들의 하나 (도 2에서 나타난 예에서 Hap II)는 염색체 부위 220에 있는 모든 증폭된 대립인자들을 결합시켜서 결정될 수 있다. 특정한 좌위에서 증폭된 대립인자는 좌위에 있는 각 대립인자들의 수를 비교하여 결정될 수 있다. 다음으로, 다른 반수체형 (Hap I)은 비-증폭된 대립인자들을 결합시켜서 결정될 수 있다. 이러한 방법은 개체가 암세포들을 가지는 경우에만 유용하고, 부위 220에서 반수체형을 결정하기 위해서만 작용하지만, 증폭된 염색체 부위의 좋은 도시를 제공한다.

[0042] 증폭은 2개 이상의 염색체들을 가지는 것 또는 하나의 염색체에서 유전자의 반복으로부터 유발될 수 있다. 하나의 부위는 일렬로 복제될 수 있거나, 부위가 해당 부위의 하나 이상의 사본들을 포함하는 미세한 염색체일 수 있다. 증폭은 복제되고 서로 다른 염색체 또는 동일한 염색체의 서로 다른 부위 내로 삽입된 염색체의 유전자로부터도 역시 유발될 수 있다. 이러한 삽입들은 증폭의 한 유형이다.

[0043] II. 염색체 부위의 선택

[0044] 암 조직은 적어도 일정의 이들 무세포 DNA (및 잠재적으로 세포성 DNA)에 기여하기 때문에, 암 조직의 게놈 이상들은 혈장 및 혈청과 같은 시료에서 검출될 수 있다. 이상을 검출하는 데 문제점은 종양 또는 암이 매우 작을 수 있어, 암 세포들로부터 비교적 적은 DNA를 제공하는 것이다. 따라서 이상을 가진 순환 DNA의 양은 매우 적어, 검출을 매우 어렵게 만든다. 게놈의 단일 좌위에는 이상을 검출할만한 충분한 DNA가 존재하지 않을 수 있다. 본 명세서에서 기술된 방법들은 다수의 좌위 (반수체형)를 포함하는 염색체 부위에서 DNA를 분석하여 본 어려움을 극복할 수 있고, 따라서 한 좌위에서의 작은 변이가 반수체형 위로 모일 때 감지가 가능한 차이로 변화된다. 따라서, 부위의 다수의 좌위를 분석하는 것은 더 큰 정확도를 제공할 수 있고, 위양성 및 위음성을 감소시킬 수 있다.

[0045] 또한, 이상의 부위는 매우 작을 수 있어, 이상을 확인하는 것을 어렵게 만든다. 단 하나의 좌위 또는 특정한 좌위들이 사용되는 경우라면, 이들 좌위들이 아닌 이상들은 놓칠 것이다. 본 명세서에서 기술된 바와 같이, 일정 방법들은 부위의 소집합 내에서 이상을 찾도록 전체 부위들 조사할 수 있다. 분석된 부위들이 게놈으로 확장될 때, 향후 게놈 전부가 하기에 보다 자세하게 기술되는 바와 같이 다양한 길이 및 위치의 이상들을 찾도록 분석될 수 있다.

[0046] 상기에서 나타난 바와 같이 이들 요점들을 설명하기 위하여, 부위는 이상을 가질 수 있다. 그러나, 부위는 분석을 위해 선택되어야 한다. 부위의 길이 및 위치는 결과들을 변경시켜서 분석에 영향을 줄 수 있다. 예를 들어, 도 1에서 첫 번째 부위가 분석되는 경우라면, 이상이 전혀 검출되지 않을 것이다. 두 번째 부위가 분석되는 경우라면, 이상이 예로 본 명세서에 기술된 방법들을 사용하여 검출될 수 있다. 첫 번째 및 두 번째 부위들 다를 포함하는 더 큰 부위가 분석되는 경우라면, 이상이라면 모두를 확인하는 것을 더욱 어렵게 만들 수 있는 더 큰 부위의 단지 일부가 이상을 가지는 어려움, 뿐만 아니라 이상의 정확한 위치 및 길이를 확인하는 문제점과 만난다. 다양한 구현예들이 이들 난점들의 일정 및/또는 모두를 해결할 수 있다. 부위를 선택하는 설명은 동일한 염색체 부위의 반수체형들을 사용하거나 두 개의 서로 다른 염색체 부위들을 사용하는 방법들에 동등

하게 적용가능하다.

[0047] A. 특정한 염색체 부위를 선택하기

[0048] 한 가지 구현예에서, 특정한 부위는 암 또는 환자의 지식을 기초로 하여 선택될 수 있다. 예를 들어, 부위는 많은 암들 또는 특정한 환자에서 공통적으로 이상을 나타내는 것으로 알려져 있을 수 있다. 부위의 정확한 길이 및 자리는 암 유형 또는 특정한 위험 요인들을 가진 환자들에게 잘 알려진 점에 관한 문헌을 참조하여 결정될 수 있다. 추가적으로, 환자의 종양 조직들이 획득되고 상기에서 기술된 바와 같이 이상들의 부위들을 확인하도록 분석될 수 있다. 현재 이러한 기법은 암세포를 획득하는 것을 (바로 진단된 환자들에게는 실용적이지 않을 수 있음) 요구할 것이지만, 이러한 기법은 동일한 환자에서 계속 감시하기 위해 (예로, 암성 조직을 제거하는 수술 이후, 또는 화학요법 또는 면역요법 또는 표적화 요법 이후, 또는 종양 재발 또는 진행의 검출을 위해) 부위들을 확인하는 데 사용될 수 있다.

[0049] 하나 이상의 특정한 부위를 확인할 수 있다. 이러한 부위들의 각기 하나의 분석은 독립적으로 사용될 수 있거나, 서로 다른 부위들은 종합적으로 분석될 수 있다. 추가적으로, 부위는 이상을 위치시키는 데 더 큰 정확도를 제공하도록 소분할될 수 있다.

[0050] 도 3은 서로 다른 유형들의 암들 및 연관된 부위들 또한 그들의 해당 이상들을 도시하는 표 300을 나타낸다. 컬럼 310은 서로 다른 암 유형들을 나열하고 있다. 본 명세서에서 기술된 구현예들은 이상이 관여하는 암 유형이라면 모두에 사용될 수 있고, 따라서 본 리스트는 단지 예시적이다. 컬럼 320은 증가 (증폭)이 동일한 행의 특정한 암과 연관된 부위들 (예로, 7p번와 같은 큰 부위들 또는 17q25번와 같은 더욱 특이적인 부위들)을 보여준다. 컬럼 330은 소실들 (결실들)이 발견될 수 있는 부위들을 보여준다. 컬럼 340은 특정한 암을 가진 이들 부위들의 연관성을 논의하는 참고문헌들을 나열하고 있다.

[0051] 잠재적인 염색체 이상들을 가진 이들 부위들은 본 명세서에서 기술된 방법들에 따른 분석을 위한 염색체 부위 (들)로서 사용될 수 있다. 암에서 변경되는 다른 게놈 부위들의 예들은 암 게놈 해부 프로젝트의 데이터베이스 (cgap.nci.nih.gov/Chromosomes/RecurrentAberrations), 유전학의 아틀라스, 또한 종양학 및 혈액학의 세포 유전학 (atlasgeneticsoncology.org/Tumors/Tumorliste.html)에서 찾을 수 있다.

[0052] 관찰할 수 있는 바와 같이, 확인된 부위들은 매우 클 수 있는 한편 다른 부위들은 더 특이적일 수 있다. 이상들은 표에서 확인되는 부위 전부를 포함할 수 없다. 따라서, 이상들의 유형에 대하여 이러한 단서들은 특정한 환자의 경우 이상이 실제로 존재하는 곳을 정확하게 겨냥하지는 못하지만, 분석을 위한 큰 부위들에 관한 대략적인 가이드로서 더 사용될 수 있다. 이러한 큰 부위들 (본 명세서에 상세한 설명이 기술됨)은 더 큰 부위에서 개별적뿐만 아니라 종합적으로 분석되는 많은 소부위들 (동등한 크기일 수 있음)을 포함할 수 있다. 따라서, 구현예들은 테스트될 암의 특정한 사항들을 기초로 하여 더 큰 부위를 선택하는 것의 관점들을 조합할 수 있지만, 다음에 기술된 바와 같이 더욱 일반적인 기법들 (예로, 소부위들을 테스트하는 것)을 사용할 수도 있다.

[0053] B. 임의적 염색체 부위를 선택하기

[0054] 또 다른 구현예에서, 분석되는 염색체 부위는 임의적으로 선택된다. 예를 들어, 게놈은 길이가 1 메가염기 (Mb)인 부위들 또는 500 Kb 또는 2 Mb와 같은 기타 선결정된 분절 길이들로 분리될 수 있다. 부위들이 1 Mb인 경우라면, 다음으로 반수체 인간 게놈에는 약 30억 개 염기들이 있기 때문에, 인간 게놈에는 대략 3,000개의 부위들이 존재하게 된다. 다음으로 이들 부위들은, 이후에 보다 자세하게 논의되는 바와 같이 각각 분석될 수 있다.

[0055] 이러한 부위들은 암 또는 환자의 지식이라면 모두를 기초로 하지 않고 분석될 부위 내로 게놈의 체계적인 분배를 기초로 하여 결정될 수 있다. 한 가지 시행예에서, 염색체가 복수의 선결정된 분절이 되는 (1 백만 개 염기들로 분할가능하지 않음) 길이를 가지지 않을 때, 염색체의 마지막 부위는 선결정된 길이 이하 (예로, 1 MB 이하)일 수 있다. 또 다른 시행예에서, 각 염색체는 염색체의 전 길이 및 만들어질 분획들의 수 (전통적으로 염색체들 중에서 다양해짐)를 기초로 하여 동등한 길이 (또는 대략적으로 동등 - 주변 오차 이내)의 부위들로 분리될 수 있다. 이러한 시행예에서, 각 염색체의 분절들의 길이는 서로 다를 수 있다.

[0056] 상기 언급된 바와 같이, 특정한 부위는 테스트된 특이적 암을 기초로 하여 확인될 수 있지만, 다음으로 특정한 부위는 더 작은 부위들로 소분할될 수 있다 (예로, 특정한 더 큰 부위로 확장되는 동등한 크기의 소부위들).

본 방식으로, 이상들은 지적 (pinpoint)될 수 있다. 하기 논의에서, 염색체 부위에 관한 일반적인 기준이라면 모두가 특이적으로 확인된 부위, 임의적으로 선택된 부위, 또는 이 둘의 조합일 수 있다.

III. 특정한 반수체형에서 이상들의 검출

본 섹션은 무세포 DNA를 포함하는 생물학적 시료를 분석하여 단일 염색체 부위에서 이상을 검출하는 방법들을 기술하고 있다. 본 섹션의 구현예들에서, 단일 염색체 부위는 부위에 있는 다수의 좌위들에서 이중접합 (서로 다른 대립인자)이고, 이로 인해 주어진 좌위에서 특정한 대립인자를 알아내어 구별될 수 있는 두 개의 반수체형을 제공한다. 따라서, 주어진 핵산 분자 (예로, 무세포 DNA의 단편)는 두 개의 반수체형들의 특정한 하나로부터 나온 것으로서 확인될 수 있다. 예를 들어, 단편은 염색체 부위에 정렬하는 서열 태그를 획득하도록 서열 결정될 수 있고, 다음으로 대립인자가 속하는 이중접합 좌위에서 반수체형이 확인될 수 있다. 특정한 반수체형 (Hap)에서 이상들을 결정하는, 특이적으로 태그 계수 및 크기 분석을 위한 두 가지 일반적 유형들의 기법들이 하기에 기술된다.

A. 반수체형들을 결정하는 것

두 개의 반수체형들을 분간하기 위하여, 염색체 부위의 두 개 반수체형들이 먼저 결정된다. 예를 들어, 도 1의 정상 세포에서 관찰되는 두 개의 반수체형들 Hap I 및 Hap II이 결정될 수 있다. 도 1에서, 반수체형들은 이중접합이고 두 개 반수체형들의 분간을 허용하는 첫 번째 다수의 좌위 110를 포함한다. 서로 다른 이중접합 좌위 상의 대립인자들 (hets)이 먼저 결정되고, 다음으로 환자의 반수체형들을 결정하도록 단계가 정해질 수 있다.

SNP 대립인자들의 반수체형이 단일 분자 분석 방법들에 의해 결정될 수 있다. 이러한 방법들의 예들은 팬 등 (Fan et al., Nat Biotechnol. 2011;29:51-7), 양게트 등 (Yanget et al., Proc Natl Acad Sci U S A. 2011;108:12-7) 및 기즈마네트 등 (Kitzmanet et al., Nat Biotechnol. 2011 Jan;29:59-63)에 의해 기술되어 왔다. 임의적으로, 개인의 반수체형들은 가족 구성원들 (예로, 부모들, 형제자매들, 및 아이들)의 유전형들의 분석에 의해 결정될 수 있다. 예들로는 로취 등 (Roach et al., Am J Hum Genet. 2011;89(3):382-97) 및 로 등 (Lo et al., Sci Transl Med. 2010;2:61ra91)에 의해 기술된 방법들을 포함한다. 보다 또 다른 구현예에서, 개인들의 반수체형은 종양 조직들 및 게놈 DNA의 유전형분석 결과들을 비교하여 결정될 수 있다. 이들 개체들의 유전형은 애피매트릭스사 SNP 6.0, 일루미나사 인피니움 HD 플랫폼 또는 기타 SNP 유전형 분석 플랫폼들을 사용하는 것과 같은 마이크로레이 분석에 의해 수행될 수 있다.

반수체형들은 당업자라면 잘 숙지하고 있는 기타 방법들에 의해서도 역시 작성될 수 있다. 이러한 방법들의 예들로는 디지털 PCR (Ding C and Cantor CR. Proc Natl Acad Sci USA 2003; 100: 7449-7453; Ruano G et al. Proc Natl Acad Sci USA 1990; 87: 6296-6300), 염색체선별 또는 분리 (Yang H et al. Proc Natl Acad Sci U S A 2011; 108: 12-17; Fan HC et al. Nat Biotechnol 2011; 29: 51-57), 정자 반수체형 분석 (Lien S et al. Curr Protoc Hum Genet 2002; Chapter 1:Unit 1.6) 및 영상화 기법들 (Xiao M et al. Hum Mutat 2007; 28: 913-921)과 같은 단일 분자 분석을 기초로 하는 것들을 포함한다. 다른 방법들로는 대립인자-특이적인 PCR (Michalatos-Beloin S et al. Nucleic Acids Res 1996; 24: 4841-4843; Lo YMD et al. Nucleic Acids Res 19: 3561-3567), 클로닝 및 제한효소 소화 (Smirnova AS et al. Immunogenetics 2007; 59:93-8) 등을 기초로 하는 것들을 포함한다. 보다 다른 방법들은 개체의 반수체형이 통계학적 평가들로부터 추론되도록 허용하는 집단에서 반수체형 블록들의 분포 및 연관 불평형 구조를 기초로 한다 (Clark AG. Mol Biol Evol 1990; 7:111-22; 10:13-9; Salem RM et al. Hum Genomics 2005; 2:39-66).

LOH 부위의 반수체형을 결정하는 또 다른 방법은 종양성 조직들이 입수가능한 경우 개체의 정상 조직들 및 종양성 조직들을 유전형 분석하는 것에 의한다. LOH의 존재 시, 종양 세포들의 매우 높은 분획 농도를 가진 종양 조직들은 LOH를 보이는 부위 내의 모든 SNP 좌위들에 대해 외관상 동종접합을 보여줄 것이다. 이들 SNP 좌위들의 유전형들은 하나의 반수체형 (도 1에서 LOH 부위의 Hap I)을 포함할 것이다. 한편, 정상 조직들은 개체가 LOH 부위 내의 SNP 좌위에 대해 이중접합인 점을 가리킬 것이다. 정상 조직들에서는 존재하지만 종양 조직들에서는 아닌 대립인자들은 다른 반수체형 (도 1에서 LOH 부위의 Hap II)를 포함할 것이다.

B. 상대적 반수체형 용량 (RHDO)분석

- [0066] 상기에 언급된 바와 같이, 염색체 부위의 반수체형들 하나의 증폭 또는 결실을 가진 염색체 이상들은 종양 조직들에 있는 염색체 부위의 두 개 반수체형들의 용량 불균형을 유도할 것이다. 종양성 성장을 가진 사람의 혈장에서, 순환 DNA의 분획은 종양 세포들로부터 유래한다. 암 환자들의 혈장에서 종양-유래 DNA의 존재로 인해, 이러한 불균형은 그들의 혈장에서도 역시 존재할 것이다. 두 개 반수체형들의 용량에서 불균형은 각 반수체형으로부터 나온 분자들의 수를 계수하여 검출될 수 있다.
- [0067] LOH가 종양 조직들에서 관찰되는 염색체 부위들의 경우 (예로, 도 1의 부위 120), Hap I은 조직들로부터 나온 Hap II의 기여가 부족하기 때문에 Hap II와 비교될 때 순환 DNA 분자들 (단편들) 중에서 과다-발현될 것이다. 사본수 증폭이 관찰되는 염색체 부위들의 경우, Hap II는 종양 조직들로부터 나온 Hap II의 추가적인 용량 방출 때문에 Hap II의 단일-대립인자 증폭에 의해 영향을 받는 부위들의 경우 Hap I과 비교될 때 과다-발현될 것이다. 과다 또는 과소 발현을 결정하기 위하여, 시료에서 소정의 DNA 단편들은 Hap I 또는 Hap II로부터 나온 것으로 결정되고, 이는 다양한 방법들에 의해, 예로 보편적인 서열결정 및 정렬을 수행하거나 디지털 PCR 및 서열-특이 탐침들을 사용하여 시행될 수 있다.
- [0068] 서열결정된 태그들을 생성하도록 암 환자들의 혈장 (또는 기타 생물학적 시료)로부터 다수의 DNA 단편들의 서열 결정 이후에, 두 개의 반수체형들 상의 대립인자들에 해당하는 서열결정된 태그들이 확인되고 계수될 수 있다. 다음으로 각 두 개의 반수체형들에 해당하는 서열결정된 태그들의 수들이 두 개 반수체형들이 혈장에서 동등하게 발현되는지 여부를 결정하도록 비교될 수 있다. 한 가지 구현예에서, 연속적 확률 테스트 (SPRT)가 혈장에서 두 개 반수체형들의 발현들이 유의하게 서로 다른지 여부를 결정하는 데 사용될 수 있다. 통계학적으로 유의한 차이는 분석된 염색체 부위에서 염색체 이상의 존재를 제시한다. 또한, 혈장에서 두 개 반수체형들의 정량적 차이는 상기 기술된 바와 같이 혈장에서 종양-유래 DNA의 분획 농도의 추정에 사용될 수 있다.
- [0069] 본 출원에서 기술된 DNA 단편의 정체 (identity) (예로, 인간 게놈에서 그의 위치)를 결정하는 진단적 접근법들은 검출 플랫폼으로서 대규모 병렬 서열결정을 사용하는 것에 제한되지 않는다. 이들 진단적 접근법들은, 예를 들어 이에 제한되는 것은 아니지만 마이크로플루이드스 (microfluidics) 디지털 PCR 시스템들 (예로, 플루이다임 (Fluidigm) 디지털 어레이 시스템), 마이크로드롭 (microdroplet) 디지털 PCR 시스템 (예로, 레인댄스 (RainDance) 및 퀀타라이프사 (QuantaLife)로부터 출시된 제품들), BEAM-ing 시스템 (예로, 비드들, 현탁 PCR, 증폭 및 자석들) (Diehl et al. ProcNatI AcadSci USA 2005; 102: 16368-16373), 실시간 PCR, 질량 분광분석법 기초한 시스템들 (예로, the 시퀀노매스어레이 (SequenomMassArray) 시스템) 또한 복합체 라이게이션-의존성 탐침 증폭 (MLPA) 분석에도 역시 적용될 수 있다.
- [0070] **정상 부위들**
- [0071] 도 4는 본 발명의 구현예들에 따라 혈장에서 시행된 측정들과 함께 이상을 보여주지 않는 암세포 내의 염색체 부위들을 도시한다. 염색체 부위 410은, 예로 테스트될 특이적 암을 기초로 하거나 게놈의 큰 절편들로 확장되는 선결정된 분절들을 사용하는 일반적인 검색을 기초로 하는 방법이라면 모두에 의해 선택될 수 있다. 두 개 반수체형들을 분간하기 위하여, 두 개의 반수체형들은 먼저 결정된다. 도 4는 염색체 부위 410에 대한 정상 세포의 두 개 반수체형들 (Hap I 및 Hap II)을 보여준다. 반수체형들은 첫 번째 다수의 좌위들 420을 포함한다. 본 첫 번째 다수의 좌위들 420은 분석된 염색체 부위 410으로 확장된다. 나타난 바와 같이, 이들 좌위들은 정상 세포에서 이중접합이다. 암세포를 위한 두 개 반수체형들도 역시 관찰된다. 암세포에서 부위들은 전혀 결실되거나 증폭되지 않는다.
- [0072] 도 4는 각 좌위들 420를 위한 각 반수체형 상의 대립인자 계수들의 수를 나타낸다. 누적 합계 (total)가 염색체 부위 410의 소정의 소부위들을 위해 제공된다. 대립인자 계수들의 수는 각 특정한 좌위에서 반수체형에 해당하는 DNA 단편들의 수에 해당한다. 예를 들어, 첫 번째 좌위 421를 포함하고 대립인자 A를 가지는 DNA 단편은 Hap I로 계수되게 될 것이다. 또한, 대립인자 T를 가지는 DNA 단편은 Hap II로 계수되게 될 것이다. 단편이 정렬하는 곳 (예로, 이것이 특정한 좌위를 포함하는지 여부) 및 이것이 포함하는 대립인자의 종류의 결정은 본 명세서에서 언급된 바와 같이 다양한 방식으로 결정될 수 있다. 두 개 반수체형들 상의 계수들의 비율은 통계학적으로 유의한 차이가 존재하는지 여부를 결정하는 데 사용될 수 있다. 두 개 수치들 간의 차이도 역시 사용될 수 있고; 차이는 단편들의 전체 수에 의해 정상화될 수 있다. 비율 및 차이 (또한 그의 함수들)는 이상이 존재하는지 여부의 분류를 결정하도록 역치들과 비교되는 매개변수들의 예들이다.
- [0073] RHDO 분석이 혈장에서 두 개 반수체형들의 불균형이라면 모두가 존재하는지 여부를 결정하기 위해 동일한 반수

체형 (예로, 누적 계수들) 상의 모든 대립인자들, 예로 상기 언급된 로 (Lo)의 특허 출원들 제 12/940,992호 및 제 12/940,993호에 기술된 바와 같이 모성의 혈장에서 시행될 수 있는 바와 같이 사용될 수 있다. 본 방법은 불균형이라면 모두가 존재하는 지 여부를 결정하는 데 사용된 DNA 분자들의 수를 유의하게 증가시킬 수 있고, 따라서 암의 존재로 인한 불균형을 암 또는 비악성 병태의 부재 시 대립인자 계수들의 확률적 분포와 분간하는 데 나은 통계학적 능력을 유도한다. 복수의 SNP 좌위들을 분리하여 분석하는 것과는 대조적으로, RHDO 접근법은 두 개 염색체들 상의 대립인자들의 상대적 자리 (반수체형 정보)를 사용할 수 있어 동일한 염색체 상에 위치하는 대립인자들이 다함께 분석될 수 있다. 반수체형 정보의 부재 시, 서로 다른 SNP 좌위들의 대립인자 계수들은 반수체형이 혈장에서 과다 또는 과소 발현되는지 여부를 통계학적으로 결정하도록 함께 첨가될 수 없다. 대립인자 계수들의 정량은, 이에 제한되는 것은 아니지만 대규모 병렬 서열결정 (예로, 합성 시스템에 의한 일루미나 (Illumina) 서열결정을 사용함), 라이프 테크놀로지사 (Life Technologies)에 의해 라이게이션 기술학에 의한 서열결정 (SOLiD), 이온 토렌트사 (Ion Torrent) 및 라이프 테크놀로지사에 의한 이온 토렌트 서열결정 시스템, 나노포아 서열결정 (nanoporetech.com), 및 454 서열결정 기술학 (로슈사 (Roche)), 디지털 PCR (예로, 마이크로플루이딕스사 디지털 PCR에 의함 (예를 들어, 플루이다임 (fluidigm.com)) 또는 BEAMing (비드들, 현탁 PCR, 증폭, 자석들 (inostics.com)) 또는 드롭 PCR (예로, 퀀타라이프사 (quantalife.com) 및 레인댄스사 (raindancetechnologies.com)에 의함) 또한 실시간 PCR에 의해 수행될 수 있다. 본 기술학의 다른 시행예에서, 용액-내 포획을 사용하는 강화된 표적 서열결정 (예로, 애질런트사 (Agilent) SureSelect 시스템, 일루미나사 TruSeq 맞춤 강화 키트를 사용함) (illumina.com/applications/sequencing/targeted_resequencing.ilmn)을 사용하거나, 마이제노스틱사 (MyGenostics) GenCap 맞춤 강화 시스템 (mygenostics.com/) 또는 어레이-기초 포획 (예로, 로슈사 NimbleGen 시스템을 사용함)에 의해 사용될 수 있다.

[0074] 도 4에 나타난 예에서, 첫 번째 두 개 SNP 좌위들 (첫 번째 SNP의 경우 24 vs 26 및 두 번째 SNP의 경우 18 vs 20)은 약간의 대립인자 불균형이 관찰된다. 그러나, 대립인자 계수들의 수는 실제 대립인자 불균형이 존재하는 지 여부를 결정하는 데 통계학적으로 충분하지 않다. 따라서, 동일한 반수체형 상의 대립인자의 계수들은 두 개 반수체형들에 대한 누적 대립인자 계수들이 두 개 반수체형들 간의 대립인자 불균형이 염색체 부위 410 (본 예의 경우 다섯 번째 SNP)에 전혀 존재하지 않는 것으로 통계학적으로 판단하는 데 충분한 때까지 첨가된다. 통계학적으로 유의한 분류에 도달한 이후, 누적 계수는 재설정된다 (본 예의 경우 여섯 번째 SNP). 다음으로 누적 계수는 두 개 반수체형들에 대한 누적 대립인자 계수들이 두 개 반수체형들 간의 대립인자 불균형이 부위 410의 특정한 소부위에 없는 것으로 통계학적으로 판단하는 데 다시 충분한 때까지 결정된다. 전체 누적 계수가 부위 전부에 마찬가지로 사용될 수 있지만, 이전의 방법은 서로 다른 소부위들이 테스트되도록 허용하고, 이는 부위 410 전부와 반대되는 이상의 위치를 결정하는 데 더 큰 정확성 (예로, 소부위)을 제공한다. 실제 대립인자 불균형이 존재하는지 여부를 결정하는 통계학적 테스트들의 예들로는, 이에 제한되는 것은 아니지만 연속적 확률 테스트 (Zhou W, et al. Nat Biotechnol 2001; 19: 78-81; Zhou W, et al. Lancet. 2002; 359: 219-25), t-테스트, 및 카이-제곱 테스트를 포함한다.

[0075] 결실을 검출하는 것

[0076] 도 5는 본 발명의 구현예들에 따라 결실된 부위를 결정하도록 혈장에서 작성된 측정들과 함께 암세포 내의 염색체 부위 510의 결실을 도시한다. 도 5는 염색체 부위 510에 대한 정상 세포의 두 개 반수체형들 (Hap I 및 Hap II)을 보여준다. 반수체형들은 분석된 염색체 부위 510를 확장시키는 첫 번째 다수의 이중접합 좌위520을 포함한다. 암세포의 두 개 반수체형들도 역시 관찰된다. 암세포에서, 부위 510는 Hap II가 결실된다. 도 4에서와 같이, 도 5도 역시 각 좌위들 520을 위한 대립인자 계수들의 수를 보여준다. 누적 합계도 역시 염색체 부위 510 내의 소정의 소부위들을 위해 유지된다.

[0077] 종양 조직들은 전형적으로 종양 세포들 및 비종양 세포들의 혼합물을 포함하기 때문에, LOH는 부위 510 내의 좌위들에 있는 두 개 대립인자들의 양의 비율을 측정하여 표현될 수 있다. 이러한 상황에서, 부위 510에서 결실된 반수체형 Hap II는 정상 조직들 상의 해당 좌위들과 비교될 때 DNA 양들의 상대적 감소를 보여주는 좌위들 520의 좌표들에 의해 결정될 수 있다. 더욱 자주 출현하는 단편들을 가진 반수체형은 Hap I이고, 이는 종양 세포들에서 보유된다. 소정의 구현예들에서, 결실된 및 보유된 반수체형들이 바로 결정되게 하도록, 종양 시료에서 종양 세포들의 비율을 강화할 절차를 수행하는 것이 바람직할 수 있다. 이러한 절차의 한 가지 예가 미세절개 (수작업 또는 레이저 포획 기법들에 의함)이다.

[0078] 이론적으로, Hap I 상의 각 대립인자들은 종양 조직들에서 LOH를 보이는 염색체 부위의 경우 순환 DNA에서 과다-발현될 것이고 대립인자 불균형의 정도는 혈장에서 종양 DNA의 분획 농도에 의존할 것이다. 그러나 동시에, 순환 DNA 시료라면 모두에서 두 개 대립인자들의 상대적인 과다 (abundance)도 역시 포아송 분포에 의해 관리될 것이다. 통계학적 분석은 관찰된 대립인자 불균형이 암 조직들에서 LOH 존재로 인한 것 또는 기회로 인한 것인지 여부를 결정하도록 수행될 수 있다. 암에서 LOH와 연관된 실제 대립인자 불균형을 검출하는 능력은 분석되는 순환 DNA 분자들의 수 및 종양성 DNA의 분획 농도에 의존적이다. 종양성 DNA의 더 높은 분획 농도 및 분석되는 분자들의 더 많은 수는 실제 대립인자 불균형을 검출하는 더 높은 민감도 및 특이도를 생기게 할 것이다.

[0079] 도 5에 나타난 예에서, 첫 번째 두 개 SNP 좌위들 (첫 번째 SNP의 경우 24 vs 22 및 두 번째 SNP의 경우 18 vs 15)의 경우 약간의 대립인자 불균형이 관찰된다. 그러나, 대립인자 계수들의 수는 실제 대립인자 불균형이 존재하는지 여부를 결정하는 데 통계학적으로 충분하지 않다. 따라서, 동일한 반수체형 상의 대립인자의 계수들은 두 개 반수체형들에 대한 누적 대립인자 계수들이 두 개 반수체형들 간의 대립인자 불균형이 염색체 부위 510 (본 예의 경우 다섯 번째 SNP)에 존재하지 것으로 통계학적으로 판단하는 데 충분할 때까지 함께 첨가된다. 일정 구현예들에서는, 불균형만이 알려져 있고 특이적 유형 (증폭의 결실)은 결정되지 않는다. 다음으로 누적 계수가 두 개 반수체형들에 대한 누적 대립인자 계수들이 부위 510의 특정한 소부위에 대한 두 개 반수체형들 간의 대립인자 불균형이 통계학적으로 판단하는 데 다시 충분할 때까지 결정된다. 전체 누적 계수가 부위 전부에 본 명세서에서 기술된 방법이라면 모두에서 시행될 수 있는 바와 같이 마찬가지로 사용될 수 있다.

[0080] 염색체 부위의 증폭을 검출하는 것

[0081] 도 6은 본 발명의 구현예들에 따라 증폭된 부위를 결정하도록 혈장에서 작성된 측정들과 함께 암세포 내의 염색체 부위 610의 증폭을 도시한다. LOH에 추가하여, 염색체 부위들의 증폭도 역시 암 조직들에서 빈번하게 관찰된다. 도 6에 나타난 예에서, 염색체 부위 610의 Hap II는 암세포에서 세 개의 사본들로 증폭된다. 나타난 바와 같이, 부위 610은 이전의 도면들에서 관찰되는 더 긴 부위들과는 반대되는 바와 같이 6개의 이중접합 좌위들을 포함한다. 증폭은 과다-발현이 통계학적으로 유의한 유의한 것으로 결정되는 여섯 번째 좌위들에서 통계학적으로 유의한 것으로서 검출되었다. 일정 구현예들에서, 하나의 불균형만이 알려져 있고 특이적 유형 (결실 또는 증폭)은 결정되지 않는다. 다른 구현예들에서, 암세포는 획득되고 분석될 수 있다. 이러한 분석은 불균형이 결실 (암세포가 결실된 부위에 대해 동종접합임) 또는 증폭 (암세포가 증폭된 부위에 대해 이종접합임)으로 인한 것인지 여부에 대한 정보를 제공한다. 다른 시행예들에서, 결실 또는 증폭이 존재하는지 여부는 부위 전부를 (예로, 개별적으로 반수체형들이 아님) 분석하도록 섹션 IV의 방법들을 사용하여 결정될 수 있다. 부위가 과다-발현된 경우라면, 이상은 증폭이고; 부위가 과소-발현된 경우라면, 이상은 결실이다. 부위 620도 역시 분석되고 누적 계수들은 불균형이 전혀 존재하지 않는 점을 입증한다.

[0082] 혈청 RHDO 분석을 위한 SPRT 분석

[0083] 이중접합성 좌위들을 가진 염색체 부위들이라면 모두를 위해, RHDO 분석이 혈장에서 두 개 반수체형들의 용량 불균형이라면 모두가 있는지 여부를 결정하는 데 사용될 수 있다. 이들 부위들에서, 혈장에서 반수체형 용량 불균형의 존재는 혈장 시료에서 종양-유래 DNA의 존재를 제시하고 있다. 한 가지 구현예에서, SPRT 분석은 Hap I 및 Hap II에 대한 서열결정된 해독 (reads)의 수에서 차이가 통계학적으로 유의한지 여부를 결정하는 데 사용될 수 있다. SPRT 분석의 본 예에서, 우리는 먼저 각 두 개의 반수체형들로부터 나온 서열결정된 해독의 수를 결정한다. 다음으로, 우리는 잠재적으로 과다-발현된 반수체형이 기여하는 서열결정된 해독의 비례적 양을 나타내는 매개변수 (예로, 분획)를 결정할 수 있다 (예로, 다른 반수체형의 해독 수로 나눈 반수체형의 해독 수의 분획). 잠재적으로 과다-발현된 반수체형은 LOH의 시나리오에서는 비-결실된 반수체형이고 염색체 부위의 단일-대립인자 증폭의 시나리오에서는 증폭된 반수체형일 것이다. 다음으로, 본 분획은 귀무가설, 예로 반수체형 용량 불균형의 부재 및 대안의 가설, 예로 반수체형 용량 불균형의 존재를 기초로 하여 구축된 두 개의 역치 수치들과 비교될 것이다. 분획이 상부 역치 이상인 경우라면, 이것은 혈장에서 두 개의 반수체형들의 통계학적으로 유의한 불균형의 존재를 가리킨다. 분획이 저부 역치 미만인 경우라면, 이것은 두 개의 반수체형들의 통계학적으로 유의한 불균형이 전혀 존재하지 않는 점을 가리킨다. 분획이 상부 및 저부 역치들 사이에 있는 경우라면, 이것은 판단하도록 충분한 통계학적 능력이 없는 점을 가리킨다. 분석되는 부위에 대한 이중접합성 좌위들의 수에서의 연속적 증가는 성공적인 SPRT 분류가 작성될 때까지 수행될 수 있다.

[0084] SPRT의 상부 및 저부 경계들을 계산하는 방정식들은:

[0085] 상부 역치 = $[(\ln 8)/N - \ln \delta]/\ln \gamma$; 저부 역치 = $[(\ln 1/8)/N - \ln \delta]/\ln$,

[0086] 여기서 $\delta = (1 - \theta_1)/(1 - \theta_2)$ 및 $\gamma = \frac{\theta_1(1-\theta_2)}{\theta_2(1-\theta_1)}$, θ_1 은 대립인자 불균형이 혈장에 존재할 때 잠재적으로 과다-발현되는 반수체형으로부터 나온 서열결정된 태그들의 예상된 분획이고, θ_2 는 대립인자 불균형이 혈장에 존재하지 않는, 예로 0.5일 때 두 개의 반수체형들이라면 모두의 예상된 분획이고, N은 Hap I 및 Hap II를 위한 서열결정된 태그들의 전체 수이고, $\ln$ 은 자연 로그, 예로 $\log_e$ 를 나타내는 수학적 기호이다. θ_1 은 혈장 시료에 존재할 것으로 기대하는 (또는 알고 있는) 종양-유래 DNA (F)의 분획 농도에 의존적일 것이다.

[0087] LOH의 시나리오에서, $\theta_1 = 1/(2 - F)$. 단일-대립인자 증폭의 시나리오에서, $\theta_1 = (1 + zF)/(2 + zF)$, 여기서 z는 종양에서 증폭되는 염색체 부위의 여분의 사본 수를 나타낸다. 예를 들어, 하나의 염색체가 복제되는 경우라면, 특정한 염색체의 하나의 여분의 사본이 있을 것이다. 그러면, z는 1과 같다.

[0088] 도 7은 본 발명의 구현예들에 따라 종양 조직에서 단일-대립인자 증폭을 보였던 염색체 1p에 위치하는 분절에 대한 간세포 암종 (HCC) 환자의 혈장 DNA의 RHDO 분석을 나타낸다. 녹색의 삼각형들은 환자의 데이터를 나타낸다. 서열결정된 해독의 전체 수는 분석되고 있는 SNPs의 증가하는 수와 함께 증가하였었다. 종양에서 증폭된 반수체형으로부터 나온 전체 서열결정된 해독의 분획은 분석된 서열결정된 해독의 전체 수와 함께 변화하였고 궁극적으로 상부 역치보다 더 높은 수치에 도달하였다. 이것은 유의한 반수체형 용량 불균형을 가리키고, 따라서 혈장에서 암-연관된 염색체 이상의 존재를 지지하고 있다.

[0089] SPRT를 사용하는 RHDO 분석은 종양 조직에서 증폭들 및 결실들을 보이는 HCC 환자의 모든 염색체 부위들을 위해 수행되었다. 결과들은 LOH를 가지는 것으로 알려진 922개 분절들 및 증폭을 가지는 것으로 알려진 105개 분절들의 경우 다음과 같다. LOH의 경우, 922개 분절들이 SPRT로 분류되었고, 분절들의 921개는 혈장에서 반수체형 용량 불균형을 가지는 것으로 정확하게 확인되어, 99.99%의 정확도를 제공하였다. 단일-대립인자 증폭의 경우, 105개 분절들이 SPRT로 분류되었고, 분절들의 105개는 혈장에서 반수체형 용량 불균형을 가지는 것으로 정확하게 확인되어, 100%의 정확도를 제공하였다.

[0090] C. 상대적 반수체형 크기 분석

[0091] 두 개의 반수체형들에 정렬된 단편들의 용량을 계수하는 것의 대안으로서, 각각의 반수체형들의 단편들의 크기가 사용될 수 있다. 예를 들어, 특정한 염색체 부위를 위해, 하나의 반수체형으로부터 나온 DNA 단편들의 크기가 다른 반수체형의 DNA 단편들의 크기와 비교될 수 있다. 부위의 첫 번째 반수체형의 이중접합 좌위들에서 대립인자들이라면 모두에 해당하는 DNA 단편들의 크기 분포를 분석하고, 이를 두 번째 반수체형의 이중접합 좌위들에서 대립인자들이라면 모두에 해당하는 DNA 단편들의 크기 분포와 비교한다. 크기 분포에서 통계학적으로 유의한 차이는 계수들의 수와 유사한 방식으로 이상을 확인하는 데 사용될 수 있다.

[0092] 전체 (예로, 종양성과 비종양성을 더함) 혈장 DNA의 크기 분포는 암 환자들에서 증가하는 것으로 보고되어 왔다 (Wang BG, et al. Cancer Res. 2003; 63: 3966-8). 그러나, 종양-유래 DNA (DNA 전체량 (예로, 종양과 비종양을 더함) 대신에)를 상세하게 연구하고 있는 경우라면, 다음으로 종양-유래 DNA 분자들의 크기 분포가 비종양 세포들로부터 유래한 분자들의 분포보다 더 짧은 점이 관찰되었다 (Diehl et al. Proc Natl Acad Sci U S A. 2005; 102: 16368-73). 따라서, 순환 DNA의 크기 분포는 암-연관된 염색체 이상들이 존재하는지 여부를 결정하는 데 사용될 수 있다. 크기 분석의 원리는 도 8에 나타나 있다.

[0093] 도 8은 본 발명의 구현예들에 따라 결실을 포함하는 종양이 존재할 때 염색체 부위의 두 개 반수체형들에 대한 단편들의 크기 분포에서 변화를 나타낸다. 도 8에 도시된 바와 같이, T 대립인자가 종양 조직들에서 결실된다. 그 결과로, 종양 조직들은 A 대립인자의 짧은 분자들만을 혈장으로 방출할 것이다. 종양-유래 짧은 DNA 분자들은 혈장에서 A 대립인자에 대한 크기 분포의 전반적인 단축화를 유도하여, 혈장에서 T 대립인자와 비교하여 더 짧은 A 대립인자의 크기 분포를 가져올 것이다. 이전의 섹션들에서 논의된 바와 같이, 동일한 반수체형 상에 위치하는 모든 대립인자들은 다함께 분석될 수 있다. 다른 말로 하면, 하나의 반수체형 상에 위치하는 대립인자들을 보유하는 DNA 분자들에 대한 크기 분포는 다른 반수체형 상에 위치하는 대립인자들을 보유하는 DNA 분자

들에 대한 것과 비교될 수 있다. 종양조직들에서 결실된 분수체형은 혈장에서 더 긴 크기 분포를 보여줄 것이다.

[0094] 크기 분석은 암과 연관된 염색체 부위들의 증폭을 검출하는데도 역시 적용될 수 있다. 도 9는 본 발명의 구현 예들에 따라 증폭을 포함하는 종양이 존재할 때 염색체 부위의 두 개 분수체형들에 대한 단편들의 크기 분포에서 변화를 나타낸다. 도 9에서 나타난 예에서, 대립인자 T를 보유하는 염색체 부위는 종양에서 복제된다. 그 결과로, A 대립인자의 크기 분포와 비교하여 T 대립인자의 크기 분포의 전반적인 단축화를 가져올 것이다. 크기 분석은 동일한 분수체형 상에 위치하는 모든 대립인자들에 종합적으로 적용될 수 있다. 다른 말로 하면, 종양 조직들에서 증폭된 분수체형에 대한 크기 분포는 종양에서 증폭되지 않은 분수체형에 대한 크기 분포보다 더 짧을 것이다.

[0095] 순환 DNA의 크기 분포의 단축화의 검출

[0096] 두 개의 분수체형들, 즉 Hap I 및 Hap II으로부터 나온 DNA 단편들의 크기는, 이에 제한되는 것은 아니지만 쌍을 이룬 말단 대규모 병렬 서열결정에 의해 결정될 수 있다. DNA 단편의 말단들을 서열결정한 이후에, 서열결정된 해독 (태그들)은 기준 인간 게놈에 정렬될 수 있다. 서열결정된 DNA 분자들의 크기는 각 말단에서 최외부 뉴클레오타이드의 좌표들로부터 추론될 수 있다. 분자의 서열결정된 태그들은 서열결정된 DNA 단편이 Hap I 또는 Hap II로부터 나오는지를 결정하는 데 사용될 수 있다. 예를 들어, 서열결정된 태그들의 하나는 분석되고 있는 염색체 부위에서 이중접합 좌위들을 포함할 수 있다.

[0097] 따라서, 각 서열결정된 분자를 위해 우리는 크기 및 Hap I 또는 Hap II로부터 나오는지 여부 둘 다를 결정할 수 있다. 각 분수체형에 정렬된 단편들의 크기를 기초로 하여, 컴퓨터 시스템은 Hap I 및 Hap II 둘 다에 대한 크기 분포 프로파일들 (예로, 평균 단편 크기)를 계산할 수 있다. Hap I 및 Hap II로부터 나온 DNA 단편들의 크기 분포는 크기 분포들이 이상을 확인하도록 충분히 서로 다른 시점을 결정하도록 적절한 통계학적 분석을 사용하여 비교될 수 있다.

[0098] 쌍을 이룬 말단 대규모 병렬 서열결정과는 별도로, 이에 제한되는 것은 아니지만 전체 DNA 단편들을 서열결정하는 것, 질량 분광분석법, 및 관찰된 DNA 분자들의 길이들을 관찰하고 표준과 비교하는 광학적 방법들을 포함하는 다른 방법들이 DNA 단편들의 크기를 결정하는 데 사용될 수 있다.

[0099] 다음, 우리는 종양들의 유전적 이상들과 연관된 순환 DNA의 단축화를 검출하는 방법들의 두 가지 예를 소개한다. 이들 두 방법들은 DNA 단편들의 두 개 집단들에 대한 크기 분포에서 차이의 정량적 측정을 제공하는 것을 목적으로 한다. DNA 단편들의 두 개 집단은 Hap I 및 Hap II에 해당하는 DNA 분자들을 말한다.

[0100] 짧은 DNA 단편들의 분획에서 차이

[0101] 한 가지 시행예에서, 짧은 DNA 단편들의 분획이 사용된다. 짧은 DNA 분자들을 정의하도록 컷오프 크기 (w)를 설정한다. 컷오프 크기는 서로 다른 진단적 목적들을 맞추도록 변화되고 선택될 수 있다. 컴퓨터 시스템은 크기 컷오프와 동등하거나 이보다 짧은 분자들의 수를 결정할 수 있다. 다음으로 DNA 단편들의 분획 (Q)가 짧은 DNA의 수를 DNA 단편들의 전체 수로 나누어 계산될 수 있다. Q의 수치는 DNA 분자들 집단의 크기 분포에 의해 영향을 받을 것이다. 더 짧은 전반적인 크기 분포는 DNA 분자들의 더 높은 비율이 짧은 단편들일 것이고, 따라서 더 높은 Q 수치를 주는 점을 의미한다.

[0102] 짧은 DNA 단편들의 분획에서 Hap I 및 Hap II 간의 차이가 다음으로 사용될 수 있다. Hap I 및 Hap II로부터 나온 DNA 단편들의 크기 분포에서 차이는 Hap I 및 Hap II (Q)에 대한 짧은 단편들의 분획에서 차이에 의해 반영될 수 있다. $\Delta Q = Q_{\text{Hap I}} - Q_{\text{Hap II}}$, 여기에서 $Q_{\text{Hap I}}$ 은 Hap I DNA 단편들에 대한 짧은 단편들의 분획이고; $Q_{\text{Hap II}}$ 는 Hap II에 대한 DNA 단편들에 대한 짧은 단편들의 분획이다. $Q_{\text{Hap I}}$ 및 $Q_{\text{Hap II}}$ 은 각 분수체들로부터 나온 단편들의 크기 분포들의 두 개 그룹들의 통계학적 수치의 예들이다.

[0103] 이전의 섹션에서 도시된 바와 같이, Hap II가 종양 조직들에서 결실될 때 Hap I DNA 단편들의 크기 분포는 Hap II DNA 단편들의 것보다 짧을 것이다. 그 결과로, ΔQ 의 양성 수치가 관찰될 것이다. ΔQ 의 양성 수치는 ΔQ 가 결실이 존재하는 것으로 고려되기 충분히 큰지를 결정하도록 역치 수치와 비교될 수 있다. Hap I의 증폭도 역시 ΔQ 의 양성 수치를 나타낼 것이다. 종양 조직들에서 Hap II의 복제가 존재할 때, Hap II DNA 단편들의

크기 분포는 Hap I DNA 단편들의 크기 분포보다 짧을 것이다. 따라서, ΔQ 의 수치는 음성이 될 것이다. 염색체 이상의 부재 시, 혈장/혈청에서 Hap I 및 Hap II DNA 단편들의 크기 분포가 유사할 것이다. 따라서, ΔQ 의 수치는 대략 제로일 것이다.

[0104] 환자의 ΔQ 는 수치가 정상인지를 결정하도록 정상 개인들과 비교될 수 있다. 추가적으로 또는 임의적으로, 환자의 ΔQ 수치가 비정상인지를 결정하도록 유사한 암들을 가진 환자들로부터 획득된 수치들과 비교될 수 있다. 이러한 비교는 본 명세서에서 기술된 바와 같은 역치 수치들과 비교(들)이 관여할 수 있다. 질환의 감시의 문맥에서, 수치 ΔQ 는 일련으로 계속 감시될 수 있다. ΔQ 의 수치에서 변화는 혈장/혈청에서 종양성 DNA의 증가된 분획 농도를 가리킬 수 있다. 본 기술학의 선택된 시행예에서, 종양성 DNA의 증가된 농도는 종양 단계, 예후 판단 및 질환의 진행과 상관될 수 있다. 서로 다른 시간대에서의 측정들을 사용하는 이러한 시행예는 이후에 보다 자세하게 논의된다.

[0105] 짧은 DNA 단편들이 기여하는 전체 길이 분획에서 차이

[0106] 본 시행예에서, 짧은 DNA 단편들이 기여하는 전체 길이의 분획이 사용된다. 컴퓨터 시스템은 시료에서 DNA 단편들 그룹의 전체 길이를 결정할 수 있다 (예로, 주어진 부위의 특정한 반수체형으로부터 또는 주어진 부위로부터 바로 나온 단편들). DNA 단편들이 "짧은 단편들"로서 정의하는 것 미만의 컷오프 크기가 선택된다. 컷오프 크기는 서로 다른 진단적 목적들을 맞추도록 변화되고 선택될 수 있다. 다음으로, 컴퓨터 시스템은 짧은 DNA 단편들의 전체 길이를 컷오프 크기와 동등하거나 이보다 짧은 DNA 단편들의 무작위 선택의 길이를 합산하여 결정할 수 있다. 다음으로 짧은 DNA 단편들이 기여하는 전체 길이 분획이 다음과 같이 계산될 수 있다: $F = \sum^N \text{길이} / \sum^N \text{길이}$, 여기에서 $\sum^N \text{길이}$ 는 w (bp)와 동등하거나 이하의 길이를 가진 DNA 단편들의 길이들의 합을 나타내고; $\sum^N \text{길이}$ 는 선결정된 길이 N과 동등하거나 이하의 길이를 가진 DNA 분자들의 길이들의 합계를 나타낸다. 한 가지 구현예에서, N은 600개 염기들이다. 그러나 다른 크기 한계들, 예로, 150개 염기들, 180개 염기들, 200개 염기들, 250개 염기들, 300개 염기들, 400개 염기들, 500개 염기들, 및 700개 염기들이 "전체 길이"를 계산하는 데 사용될 수 있다.

[0107] 600개 염기들 미만의 수치가 일루미나사 게놈 분석기 시스템이 600개 염기들 이상의 DNA 단편들을 증폭시키고 서열결정하는 데 효과적이지 않기 때문에 선택될 수 있다. 또한, 분석을 600개 염기들 이하의 DNA 단편들로 제한하는 것은 게놈의 구조적 다양성들로부터 발생하는 편향들도 역시 피할 수 있다. 구조적 다양성들, 예를 들어 재배열들 (Kidd JM et al, Nature 2008; 453:56-64)의 존재 시, DNA 단편의 크기는 DNA 단편의 말단들을 기준 게놈에 맵핑하여 생물정보학적으로 추정될 때 과다추정될 수 있다. 또한, 성공적으로 서열결정되고 기준 게놈으로 맵핑된 모든 DNA 단편들의 > 99.9%가 600개 염기들 이하이고, 따라서 모든 600개 염기들과 동등하거나 더 짧은 단편들을 포함하는 것이 시료에서 DNA 단편들의 크기 분포의 대표적인 추정을 제공할 것이다.

[0108] 따라서, Hap I 및 Hap II 간의 짧은 DNA 단편들이 기여하는 전체 길이의 분획들에서 차이가 사용될 수 있다. Hap I 및 Hap II DNA 단편들 간의 크기 분포에서 교란은 그들의 F 수치들에서 차이에 의해 반영될 수 있다. 여기에서 우리는 $F_{\text{Hap I}}$ 및 $F_{\text{Hap II}}$ 를 Hap I 및 Hap II에 대한 짧은 DNA 단편들이 기여하는 전체 길이의 분획들로서 각각 정의한다. Hap I 및 Hap II 간의 짧은 DNA 단편들이 기여하는 전체 길이의 분획들에서 차이는 다음과 같이 계산될 수 있다: $\Delta F = F_{\text{Hap I}} - F_{\text{Hap II}}$. $F_{\text{Hap I}}$ 및 $F_{\text{Hap II}}$ 는 각 반수체형들로부터 나온 단편들의 크기 분포들의 두 개 그룹들의 통계학적 수치의 예들이다.

[0109] 이전의 섹션에서 도시된 구현예들과 유사하게, 종양 조직들에서 Hap II의 결실은 Hap II DNA 단편들과 비교할 때 Hap I DNA 단편들에 대한 크기 분포의 외관상 단축화를 유도할 것이다. 이것은 ΔF 의 양성 수치를 유도할 것이다. Hap II가 복제될 때, 음성 ΔF 수치가 관찰될 것이다. 염색체 이상의 부재 시, ΔF 의 수치는 대략 제로일 것이다.

[0110] 환자의 ΔF 는 수치가 정상인지를 결정하도록 정상 개인들과 비교될 수 있다. 이러한 비교는 본 명세서에서 기술된 바와 같은 역치 수치들과 비교(들)이 관여할 수 있다. 질환의 감시의 문맥에서, 수치 ΔQ 는 일련으로 감시될 수 있다. ΔF 의 수치에서 변화는 혈장/혈청에서 종양성 DNA의 증가된 분획 농도를 가리킬 수 있다.

[0111] D. 일반적인 방법

- [0112] 도 10은 본 발명의 구현예들에 따라 염색체 부위가 결실 또는 증폭을 보여주는지 여부를 결정하도록 유기체의 생물학적 시료의 반수체형들을 분석하는 방법을 도시하는 흐름도이다. 생물학적 시료는 정상 세포들로부터, 잠재적으로는 암과 연관된 세포들로부터 기원한 핵산 분자들 (단편들이라고도 불림)을 포함한다. 이들 분자들은 생물학적 시료들에서 무세포일 수 있다. 유기체는 염색체의 하나 이상의 사본을 가지는 유형이라면 모두, 예로 적어도 이수체 유기체들일 수 있지만, 고등 다배수체 유기체들을 포함할 수 있다.
- [0113] 본 방법 및 본 명세서에서 기술된 기타 방법이라면 모두의 한 가지 구현예에서, 생물학적 시료는 무세포 DNA 단편들을 포함한다. 현장 DNA의 분석이 본 출원에 기술된 서로 다른 방법들을 설명하는 데 사용되었다라도, 이들 방법들도 역시 정상 및 종양-유래 DNA의 혼합물을 포함하는 시료들에서 종양-연관된 염색체 이상들을 검출하는 데 사용될 수 있다. 기타 동일한 유형들은 침, 눈물들, 흉막액, 복수액, 담즙, 소변, 혈청, 체장액, 대변 및 경부 도말 시료들을 포함한다.
- [0114] 단계 1010에서, 첫 번째 및 두 번째 반수체형들은 첫 번째 염색체 부위에서 유기체의 정상 세포들을 위해 결정된다. 반수체형들은 본 명세서에서 언급된 것들과 같은 적합한 방법이라면 모두에 의해 결정될 수 있다. 염색체 부위는 방법이라면 모두, 예로 본 명세서에서 기술된 방법들에 의해 선택될 수 있다. 첫 번째 염색체 부위는 이중접합인 첫 번째 다수의 좌위들 (예로, 부위 410의 좌위들 420)을 포함한다. 이중접합 좌위들 (hets)는 서로로부터 멀리 떨어져 있고, 예로 좌위들은 첫 번째 다수의 좌위들의 또 다른 좌위로부터 500개 또는 1,000개 (또는 이상) 염기들이 떨어져 있다. 다른 hets는 첫 번째 염색체 부위에 존재할 수 있지만, 사용되지는 않는다.
- [0115] 단계 1020에서, 생물학적 시료에서 다수의 핵산 분자들은 각 분자의 위치 및 대립인자에 관하여 특징 분석된다. 예를 들어, 유기체의 기준 게놈에서 핵산 분자의 위치는 확인될 수 있다. 본 위치 설정은, 분자의 하나 또는 두 개 (쌍을 이룬 말단) 서열결정된 태그들을 획득하도록 분자의 서열결정을 (예로, 보편적 서열결정에 의해) 수행한 다음 서열결정된 태그(들)을 기준 게놈에 정렬시키는 것을 포함하는 다양한 방식으로 수행될 수 있다. 이러한 정렬은 기본 로컬 정렬 탐색 도구 (BLAST)와 같은 도구들을 사용하여 수행될 수 있다. 위치는 염색체의 팔에 있는 수로서 확인될 수 있다. 이중접합 좌위들 (hets)의 하나에 있는 대립인자는 단편이 나온 반수체형을 결정하는 데 사용될 수 있다.
- [0116] 단계 1030에서, 핵산 분자들의 첫 번째 그룹은 확인된 위치들 및 결정된 대립인자들을 기초로 하여 첫 번째 반수체형으로부터 나온 것으로서 확인된다. 예를 들어, 대립인자 A를 가지는 도 4의 좌위들 421을 포함하는 단편은 Hap I로부터 나온 것으로서 확인될 것이다. 첫 번째 그룹은 각 첫 번째 다수의 좌위들에 위치하는 적어도 하나의 핵산 분자를 포함하여 첫 번째 염색체 부위로 확장될 수 있다.
- [0117] 단계 1040에서, 핵산 분자들의 두 번째 그룹은 확인된 위치들 및 결정된 대립인자들을 기초로 하여 두 번째 반수체형으로부터 나온 것으로서 확인된다. 예를 들어, 대립인자 T를 가지는 도 4의 좌위들 421을 포함하는 단편은 Hap II로부터 나온 것으로서 확인될 것이다. 두 번째 그룹은 각 첫 번째 다수의 좌위들에 위치하는 적어도 하나의 핵산 분자를 포함한다.
- [0118] 단계 1050에서, 컴퓨터 시스템은 핵산 분자들의 첫 번째 그룹의 첫 번째 수치를 계산한다. 첫 번째 수치는 첫 번째 그룹의 핵산 분자들의 특성을 정의한다. 첫 번째 수치의 예들로는 첫 번째 그룹에서 분자들의 수의 태그 계수 및 첫 번째 그룹에서 분자들의 크기 분포를 포함한다.
- [0119] 단계 1060에서, 컴퓨터 시스템은 핵산 분자들의 두 번째 그룹의 두 번째 수치를 계산한다. 두 번째 수치는 두 번째 그룹의 핵산 분자들의 특성을 정의한다.
- [0120] 단계 1070에서, 첫 번째 수치는 첫 번째 염색체 부위가 결실 또는 증폭을 나타내는지 여부의 분류를 결정하도록 두 번째 수치와 비교된다. 결실 또는 증폭이 존재하는 것의 분류는 암과 연관된 세포들을 가지는 유기체에 관한 정보를 제공할 수 있다. 비교의 예들은 본 명세서에서 기술된 바와 같이 두 개 수치들의 차이 또는 비율을 취하여 하나 이상의 역치 수치들과 결과들을 비교하는 것을 포함한다. 예를 들어, 비율은 SPRT 분석에서 역치 수치들과 비교될 수 있다. 분류 예들은 양성 (예로, 검출된 증폭 또는 결실), 음성 및 미분류됨, 뿐만 아니라 양성 및 음성의 다양화 정도들 (예로, 1 및 10 사이의 정수들, 또는 0 및 1 사이의 실수를 사용함)을 포함할 수 있다. 증폭은 단순한 복제를 포함할 수 있다. 이러한 방법은 종양 DNA 및 전종양증식 병변들, 예로 암의 전구체들로부터 나온 DNA를 포함하는, 암-연관된 핵산들의 존재를 검출할 수 있다.

[0121] E. 깊이

[0122] 분석의 깊이는 특정된 정확도 내의 분류 또는 기타 결정을 제공하도록 분석되는 것이 필요한 분자들의 양을 말한다. 한 가지 구현예에서, 깊이는 기지의 이상을 기초로 하여 계산될 수 있고, 다음으로 해당 깊이를 가진 측정 및 분석이 수행될 수 있다. 또 다른 구현예에서, 분석은 분류가 작성될 때까지 계속될 수 있고 분류가 작성되는 깊이는 암의 수준 (예로, 암의 단계 또는 종양의 크기)을 결정하는 데 사용될 수 있다. 다음은 깊이가 관여하는 일정 계산들의 예들을 제공한다.

[0123] 편차는 본 명세서에서 기술되는 바와 같이 차이 또는 비율이라면 모두를 말할 수 있다. 예로서, 편차는 본 명세서에서 기술되는 바와 같이 역치 또는 종양 농도로부터 나온 매개변수의 첫 번째 및 두 번째 수치 사이에 있을 수 있다. 편차가 두 배가 되는 경우라면, 측정될 필요가 있는 단편들의 수는 $1/4$ 로 감소된다. 보다 일반적으로, 편차가 팩터 N 으로 증가되는 경우라면, 측정될 필요가 있는 단편들의 수는 $1/N^2$ 이다. 그 결과로, 편차가 팩터 $1/N$ 으로 감소되는 경우라면, 테스트될 필요가 있는 단편들의 수는 N^2 이다. N 은 실수 또는 정수일 수 있다.

[0124] 종양 DNA가 시료 (예로, 혈장)의 10%인 사례를 가정하고, 통계학적으로 유의한 차이가 1천만 개 단편들을 서열 결정하여 관찰되는 것으로 추측하라. 다음으로 예를 들어, 농축 절차가 수행되어 현재 시료에 20%의 종양 DNA가 존재하고, 다음으로 필요한 단편들의 수는 2,500,000개 단편들이 될 것이다. 본 방식으로, 깊이는 시료에 있는 종양 DNA의 백분율과 상관될 수 있다.

[0125] 증폭의 양도 역시 깊이에 영향을 줄 것이다. 부위에 두 배의 사본량을 가지는 부위를 위해 (예로, 정상 2개와 반대되는 4개), X 개수의 단편들이 분석되도록 요구되는 것으로 가정하라. 부위가 정상 사본량의 4배를 가지는 경우라면, 본 부위는 $X/4$ 양의 단편들을 요구할 것이다.

[0126] F. 역치들

[0127] 정상 수치들로부터 매개변수의 편차의 양은 (예로, 각 반수체형을 위한 수치들의 차이 또는 비율) 상기에 기술된 바와 같이 진단을 제공하는 데 사용될 수 있다. 예를 들어, 편차는 부위의 하나의 반수체형으로부터 나온 단편들의 평균 크기대비 다른 반수체형으로부터 나온 단편들의 평균 크기의 차이가 될 수 있다. 편차가 소정의 양 (예로, 정상 시료들 및/또는 부위들로부터 결정된 역치) 초과인 경우라면, 결실 또는 증폭이 확인된다. 그러나, 역치를 초과하는 정도가 정보가 될 수 있고, 이는 복수의 역치들의 사용을 유도하고 각각은 암의 서로 다른 수준에 해당할 수 있다. 예를 들어, 정상으로부터 더 높은 편차는 암이 존재하는 단계를 제공할 수 있다 (예로, 제 4기는 제 3기보다 더 높은 정도의 불균형을 가질 것이다). 더 높은 편차는 커지고 있어 많은 단편들을 방출하는 종양의 결과, 및/또는 부위가 많은 횡수로 증폭된 점도 역시 될 수 있다.

[0128] 서로 다른 수준들의 암을 제공하는 것과 추가하여, 역치들을 다양화하는 것은 이상을 가진 부위들 또는 특이적 부위들의 효율적인 검출도 역시 허용할 수 있다. 예를 들어, 주로 3배 이상의 증폭을 찾는 높은 역치들을 설정할 수 있고, 이는 하나의 반수체형의 결실보다 큰 불균형을 줄 것이다. 부위의 두 개 사본들의 결실도 역시 검출될 수 있다. 또한, 저부 역치는 이상을 가질 수 있는 부위들을 확인하는 데 사용될 수 있고, 다음으로 이들 부위들은 좀 더 나아가 이상이 존재하는지 여부 및 위치를 입증하도록 분석될 수 있다. 예를 들어, 이종의 탐색 (또는 옥트리 (octree)와 같은 고차의 탐색)이 더 높은 역치들을 사용하는 체계에서 더 낮은 수준들로 수행될 수 있다.

[0129] 도 11은 본 발명의 구현예들에 따라 결실된 부위를 결정하도록 혈장에서 작성된 측정들과 함께 암세포들에서 결실되고 있는 소부위 1130을 가진 부위 1110을 나타낸다. 염색체 부위 1110은 게놈을 동등한 크기의 분절들로 분열시키는 것과 같은 본 명세서에서 언급된 방법이라면 모두에 의해 선택될 수 있다. 도 11은 각 좌위들 1120에 대한 대립인자 계수들의 수도 역시 보여준다. 누적 합계도 역시 부위 1140 (정상 부위) 및 부위 1130 (결실된 부위)의 경우 각각 유지된다.

[0130] 부위 1110은 분석을 위해 선택된 경우라면, 누적 계수의 수는 11개 좌위들 상의 18개 차이를 제공하는 Hap I의 경우 258개 및 Hap II의 경우 240개이다. 이러한 차이는 결실된 소부위 1130이 분석된 경우보다 계수의 전체 수의 백분율과 같이 더 작다. 이것은 부위 1110의 약 절반이 정상인 한편 소부위 1130 모두는 암세포에서 결실되는 점과 일맥 상통한다. 따라서, 부위 1110에서 사용된 역치에 의존하여 이상은 놓칠 수 있다.

- [0131] 소부위들의 결실의 검출을 허용하기 위하여, 구현예들은 상대적으로 큰 부위를 위한 저부 역치를 사용할 수 있다 (본 예의 경우 부위 1110은 확인될 결실된 부위들의 크기와 비교하여 상대적으로 큰 것으로 추측된다). 저부 역치는 더 많은 부위들을 확인할 것이고, 이는 일정 위양성들을 포함할 수 있지만 위음성들을 감소시킬 수 있다. 현재, 위양성들은 이상도 역시 지적할 수 있는 심화 분석을 통하여 제거될 수 있다.
- [0132] 일단 부위가 심화 분석을 위해 표지 (flagged)되었던 경우라면, 부위는 심화 분석을 위해 소부위들로 분할될 수 있다. 도 11에서, 6개 좌위들의 소부위들 1140 및 5개 좌위들의 소부위 1130을 제공하도록 11개 좌위들을 절반으로 (예로, 이중 트리를 사용하여) 분열시킬 수 있다. 이들 부위들은 동일한 역치 수치 또는 더 엄격한 역치 수치로 분석될 수 있다. 본 예에서, 소부위 1140은 다음으로 정상인 것으로서 확인되고 소부위 1130은 결실 또는 증폭을 포함하는 것으로서 확인될 것이다. 본 방식으로, 더 큰 부위들은 이상을 전혀 가지지 않는 것으로서 해결될 수 있고, 의심되는 부위들 (저부 역치의 초과 부위들)은 높은 신뢰도로 이상을 보이는 소부위들을 확인하도록 좀 더 분석하는 데 시간을 쓸 수 있다. 여기에서 RHDO가 사용되었다라도, 크기 기법들은 동등하게 적용 가능하다.
- [0133] 첫 번째 탐색 수준을 위한 부위들의 크기 (및 트리에서 저부 수준들의 소부위들의 크기)는 검출될 이상들의 크기를 기초로 하여 선택될 수 있다. 암들은 10 Mb 길이의 이상들을 가지는 10개 부위들을 보여주는 것으로 확인되어 왔다. 환자들도 역시 이상을 나타내는 100 Mb 부위들을 가진다. 암의 나중 단계들은 이상의 더 긴 절편들을 가질 수 있다.
- [0134] *G. 부위 내에 이상의 위치 결정*
- [0135] 마지막 섹션에서, 트리 탐색을 기초로 하여 소부위들 내로 부위의 분할이 논의되었다. 여기에서, 우리는 부위 내의 이상을 지적하도록 소부위들을 분석하는 다른 방법들을 논의하고 있다.
- [0136] 도 12는 본 발명의 구현예들에 따라 이상들의 위치가 RHDO 분석을 사용하여 맵핑될 수 있는 방식을 나타낸다. 염색체 부위는 Hap I 및 Hap II로 표시된 비-암 세포들의 반수체형들로 수평적으로 나타낸다. 암세포들에서 Hap II의 결실된 부위는 LOH로서 표시된다.
- [0137] 나타낸 바와 같이, RHDO 분석은 이론적인 염색체 부위 1202의 좌측면으로부터 시작되어 우측면까지 이어진다. 각 화살표들은 RHDO 분류 분절을 나타낸다. 각 분절은 그 자신의 부위, 상세하게는 더 큰 부위의 *hets* 소집합을 가진 소부위라고 고려될 수 있다. RHDO 분류 분절의 크기는 분류가 결정될 수 있기 이전에 좌위들의 수 (및 좌위들의 자리들)에 의존적이다. 각 RHDO 분절에 포함된 좌위들의 수는 각 분절에 대해 분석된 분자들의 수, 원하는 정확도 (예로, SPRT 분석에서 우위 비율), 및 시료에서 종양-유래 DNA의 분획 농도에 의존적이다. 분류는 분자들의 수가 도 4 및 도 5에 도시된 예에서와 같이 통계학적으로 유의한 차이가 두 개 반수체형들 간에 존재하는 것을 결정하도록 적절할 때 작성될 수 있다.
- [0138] 각 수평선 화살표들은 반수체형 용량 불균형이 DNA 시료에 부재한 것을 보여주는 RHDO 분류 분절을 나타낸다. 종양에서 LOH 없는 부위 내에, 6개의 RHDO 분류들이 작성되고 각각은 반수체형 용량 불균형의 부재를 가리킨다. 다음 RHDO 분류 분절 1210은 LOH가 있고 없는 부위들 간의 연결부 1205를 교차한다. 도 12의 저부 부분에서, RHDO 분절 1210에 대한 SPRT 곡선이 나타난다. 검은색 수직 화살표는 LOH가 있고 없는 부위들 간의 연결부를 가리킨다. LOH를 가진 부위로부터 얻은 증가하는 데이터의 축적으로, 본 분절의 RHDO 분류는 반수체형 용량 불균형의 존재를 가리킨다.
- [0139] 각 흰색 수평 화살표들은 반수체형 용량 불균형의 존재를 가리키는 RHDO 분류 분절을 나타낸다. 우측면 상의 연속적 4개 RHDO도 역시 DNA 시료에서 반수체형 용량 불균형의 존재를 가리킨다. LOH가 있고 없는 부위들 간의 연결부 위치는 RHDO 분류에서 변화, 즉 반수체형 용량 불균형의 존재로부터 부재까지 또는 그 반대를 보여주는 첫 번째 RHDO 분절에 속하도록 추론될 수 있다.
- [0140] 도 13은 본 발명의 구현예들에 따라 또 다른 방향으로부터 시작된 RHDO 분류를 나타낸다. 도 13에서, 둘 다의 방향으로부터 나온 RHDO 분류가 관찰된다. 좌측면으로부터 시작되는 RHDO 분석으로부터, LOH가 있고 없는 부위들 간의 연결부는 반수체형 용량 불균형의 존재를 보여주는 첫 번째 RHDO 분절 1310에 속하도록 추론될 수 있다. 우측면으로부터 시작되는 RHDO 분석으로부터, 연결부는 반수체형 용량 불균형의 부재를 가리키는 첫 번째 RHDO 분절 1320에 속하도록 추론될 수 있다. 두 가지 방향으로 시행된 RHDO 분석으로부터 얻은 정보를 조합하여, LOH가 있고 없는 부위들 간의 연결부의 위치 1330이 추론될 수 있다.

[0141] **IV. 이상들의 비특이적 반수체형 검출**

[0142] RHD0 방법은 이종접합 좌위들을 사용하는 것에 의존한다. 현재, 이수체 유기체의 염색체들은 두 개 반수체형들을 가져오는 일정 차이들을 가질 것이지만, 이종접합 좌위들의 수는 다양화할 수 있다. 일정 개인들은 상대적으로 적은 이종접합 좌위들을 가질 수 있다. 본 섹션에서 기술된 구현에는 역시 동일한 부위의 두 개 반수체형들이 아닌 두 개 부위들을 비교하여 동종접합이 되는 좌위들에 사용될 수 있다. 따라서, 일정의 장애들이 두 개의 서로 다른 염색체 부위들의 비교로부터 존재할 수 있더라도, 더 많은 데이터 요점들이 획득될 수 있다.

[0143] 상대적 염색체 부위 용량 방법에서, 하나의 염색체 부위로부터 나온 단편들의 수는 (예로, 해당 부위에 정렬된 서열결정된 태그들을 계수하여 결정되는 바) 기대 수치 (예로, 기준 염색체 부위로부터 또는 건강한 것으로 알려진 또 다른 시료에 있는 동일한 부위로부터 나올 수 있음)와 비교된다. 본 방식으로, 단편은 서열결정된 태그가 나온 반수체형과는 상관없이 염색체 부위로 설명될 것이다. 따라서, hets를 전혀 포함하지 않는 서열결정된 태그들이 여전히 사용될 수 있다. 비교를 수행하기 위하여, 구현에는 비교 이전에 태그 수를 정상화할 수 있다. 각 부위는 적어도 두 개의 좌위들 (서로로부터 분리되어 있음)에 의해 정의되고, 이들 좌위들에서 단편들은 부위에 대하여 종합적 수치를 획득하는 데 사용될 수 있다.

[0144] 특정한 부위를 위한 서열결정된 해독 (태그)의 정상화된 수치는 해당 부위에 정렬한 서열결정된 해독의 수를 전체 계놈에 정렬가능한 서열결정된 해독의 전체 수로 나누어 계산될 수 있다. 본 정상화된 태그 계수는 한 시료로부터 나온 결과들을 또 다른 시료의 결과들과 비교하도록 허용한다. 예를 들어, 정상화된 수치는 상기에 진술된 바와 같이 특정한 부위로부터 나올 것으로 기대되는 서열결정된 해독의 비율 (예로, 백분율 또는 분획)일 수 있다. 그러나, 당업자에게는 자명할 바와 같이 많은 다른 정상화가 가능하다. 예를 들어, 한 부위를 위한 계수의 수를 기준 부위를 위한 계수의 수로 정상화할 수 있다 (상기 사례에서, 기준 부위는 바로 전체 계놈이다). 다음으로 본 정상화된 태그 계수는 암을 보이지 않는 하나 이상의 시료들로부터 결정될 수 있는 역치 수치에 대비하여 비교될 수 있다.

[0145] 다음으로 테스트된 사례의 정상화된 태그 계수는 하나 이상의 기준 개체들, 예로 암이 없는 개체들의 정상화된 태그 계수와 비교될 것이다. 한 가지 구현예에서, 특정한 염색체 부위를 위한 사례의 z-점수를 계산하여 비교가 작성된다. z-점수는 다음의 방정식을 사용하여 계산된다: $z\text{-점수} = (\text{사례의 정상화된 태그 계수} - \text{평균}) / S.D.$, 여기에서 "평균값 (mean)"은 기준 시료들의 경우 특정한 염색체 부위에 정렬한 평균 정상화된 태그 계수이고; S.D.는 기준 시료들의 경우 특정한 부위에 정렬한 정상화된 태그 계수의 수의 표준 편차이다. 따라서, z-점수는 테스트된 시료들의 경우 부위의 정상화된 태그 계수가 하나 이상의 기준 개체들의 동일한 염색체 부위의 평균 정상화된 태그 계수로부터 벗어나는 표준 편차의 수치이다.

[0146] 테스트된 유기체가 암을 가지는 상황에서, 종양 조직들에서 증폭된 염색체 부위들은 혈장 DNA에서 과다-발현될 것이다. 이것은 z-점수의 양성 수치를 가져올 것이다. 한편, 종양 조직들에서 결실된 염색체 부위들은 혈장 DNA에서 과소-발현될 것이다. 이것은 z-점수의 음성 수치를 가져올 것이다. z-점수의 규모는 여러 요인들에 의해 결정된다.

[0147] 한 가지 요인은 생물학적 시료 (예로, 혈장)에서 종양-유래 DNA의 분획 농도이다. 시료 (예로, 혈장)에서 종양-유래 DNA의 분획 농도가 높을수록, 테스트된 사례 및 기준 사례들의 정상화된 태그 계수 간 차이가 커진다. 따라서, 더 큰 z-점수가 유도될 것이다.

[0148] 또 다른 요인은 하나 이상의 기준 사례들에서 정상화된 태그 계수의 다양성이다. 테스트된 사례의 생물학적 시료 (예로, 혈장)에서 염색체 부위의 과다-발현의 동일한 정도로, 기준 그룹에서 정상화된 태그 계수의 더 작은 다양성 (예로, 더 작은 표준 편차)은 더 높은 z-점수를 가져올 것이다. 유사하게, 테스트된 사례의 생물학적 시료 (예로, 혈장)에서 염색체 부위의 과소-발현의 동일한 정도로, 기준 그룹에서 정상화된 태그 계수의 더 작은 표준 편차는 더욱 음성의 z-점수를 가져올 것이다.

[0149] 또 다른 요인은 종양 조직들에서 염색체 이상의 규모이다. 염색체 이상의 규모 특정한 염색체 부위를 위한 사본수 변화들 (증가 또는 소실 둘 중 하나)을 말한다. 종양 조직들에서 사본수 변화들이 더 높을수록, 혈장 DNA에서 특정한 염색체 부위의 과다- 또는 과소-발현의 정도가 더 높아질 것이다. 예를 들어, 염색체의 사본들 둘다의 소실은 염색체의 두 개 사본들의 하나의 소실보다 혈장 DNA에서 염색체 부위의 더 큰 과소-발현을 가져오고, 따라서 더욱 음성의 z-점수를 가져올 것이다. 전형적으로, 암들에서는 복수의 염색체 이상들이 존재한다. 각 암에서 염색체 이상들은 그의 본질 (예로, 증폭 또는 결실), 그의 정도 (단일 또는 복수의 사본 증가 또는

소실) 및 그의 범위 (염색체 길이의 견지에서 이상의 크기)에 의해 좀 더 다양화할 수 있다.

[0150] 정상화된 태그 계수를 측정하는 정확도는 분석된 분자들의 수에 의해 영향을 받는다. 우리는 15,000개, 60,000개 및 240,000개 분자들이 분획 농도가 각각 대략 12.5%, 6.3% 및 3.2%일 때 하나의 사본 변화를 가진 염색체 이상들 (증가 또는 소실 둘 중 하나)을 검출하도록 분석하는 것이 필요할 것으로 예상한다. 서로 다른 염색체 부위들에 대한 암 검출을 위한 태그 계수법의 좀 더 상세한 설명들은 로 들에 의해 "대규모 병렬 게놈 서열결정을 사용하여 태아 염색체 비배수체를 진단하는 방법"이라는 명칭으로 된 미국 특허 공고번호 제 2009/0029377호에 기술되고 있고, 그의 내용 전부는 본 명세서에서 모든 목적들 위한 참고문헌으로 통합되어 있다.

[0151] 구현예들은 태그 계수 방법 대신에 크기 분석도 역시 사용할 수 있다. 크기 분석은 정상화된 태그 계수 대신에도 역시 사용될 수 있다. 크기 분석은 본 명세서에서 또한 미국 특허출원 제 12/940,992호에서 언급된 바와 같이 다양한 매개변수들을 사용할 수 있다. 예를 들어, 상기로부터 나온 Q 또는 F 수치들이 사용될 수 있다. 이러한 크기 수치들은 이들 수치들이 해독의 수로 측정되지 않기 때문에 다른 부위들로부터 나온 계수들에 의한 정상화를 필요로 하지 않는다. 반수체형-특이적 방법들의 기법들은 마찬가지로 비특이적 방법들에 사용될 수 있다. 예를 들어, 부위의 깊이 및 위치 결정 (refinement)이 관여하는 기법들이 사용될 수 있다. 일정 구현예들에서, 특정한 부위에 대한 GC 편향은 두 개 부위들을 비교할 때 고려될 수 있다. RHDO 방법은 동일한 부위를 사용하기 때문에, 이러한 교정이 필요없다.

[0152] V. 복수의 부위들

[0153] 소정의 암들이 전형적으로 특정한 염색체 부위들에서 이상들과 함께 존재할 수 있더라도, 이러한 암들은 항상 동일한 부위에서만 존재하지는 않는다. 예를 들어, 추가적인 염색체 부위들이 이상들을 보여줄 수 있고, 이러한 추가적인 부위들의 위치는 미지일 수 있다. 또한, 암의 초기 단계들을 확인하도록 환자들을 검색할 때, 게놈을 통틀어 존재하는 이상들을 보여줄 수 있는 광범위의 암들을 확인하도록 원할 수 있다. 이들 상황들을 겨냥하기 위하여, 구현예들은 이상을 보여주는 부위들을 결정하도록 체계적인 방식으로 다수의 부위들을 분석할 수 있다. 이상들의 수 및 그들의 위치 (예로, 그들이 근접하는지 여부)는, 예를 들어 이상들을 입증하고, 암의 단계를 결정하고, 암의 진단을 제공하고 (예로, 수가 역치 수치 이상인 경우), 또한 이상을 나타내는 다양한 부위들의 수 및 위치를 기초로 하여 예후를 제공하는 데 사용될 수 있다.

[0154] 따라서, 구현예들은 이상을 나타내는 부위들의 수를 기초로 하여 유기체가 암을 가지는지 여부를 확인할 수 있다. 이상을 나타내는 부위들의 수를 확인하도록 다수의 부위들 (예로 3,000개)을 테스트할 수 있다. 부위들은 게놈 전부 또는 게놈의 단지 일부들 예로 비-반복서열 부위를 수용할 수 있다.

[0155] 도 14는 본 발명의 구현예들에 따라 다수의 염색체 부위들을 사용하여 유기체의 생물학적 시료를 분석하는 방법 1400의 흐름도를 나타낸다. 생물학적 시료는 핵산 분자들 (단편들이라고도 불림)을 포함한다.

[0156] 단계 1410에서, 유기체의 다수의 비중복 염색체 부위들이 확인된다. 각 염색체 부위는 다수의 좌위들을 포함한다. 상기에서 언급된 바와 같이, 부위는 1 Mab, 또는 일정의 다른 동등-크기일 수 있다. 다음으로 전체 게놈은 각 선결정된 크기 및 위치인 약 3,000개의 부위들을 포함할 수 있다. 또한 상기에서 언급된 바와 같이, 이러한 선결정된 부위들은 특정한 염색체의 길이 또는 사용될 부위들의 특정된 수, 및 본 명세서에서 언급된 다른 판정기준이라면 모두를 수용하도록 다양화될 수 있다. 부위들이 서로 다른 길이들을 가지는 경우라면, 이러한 길이들은 예로 본 명세서에서 기술된 바와 같이 결과들을 정상화하는 데 사용될 수 있다.

[0157] 단계 1420에서, 유기체의 기준 게놈에서 핵산 분자의 위치는 각 다수의 핵산 분자들에 대해 확인된다. 위치는 본 명세서에서 언급된 방식들이라면 모두로, 예로 단편들을 서열결정하여 서열결정된 태그들을 획득하고 기준 게놈에 서열결정된 태그들을 정렬하여 결정될 수 있다. 분자의 특정한 반수체형도 역시 반수체형-특이적 방법들을 위해 결정될 수 있다.

[0158] 단계들 1430 내지 1450은 각 염색체 부위들을 위해 수행된다. 단계 1430에서, 핵산 분자들의 개별 그룹은 확인된 위치들을 기초로 하여 염색체 부위로부터 나온 것으로서 확인된다. 개별 그룹은 염색체 부위의 각 다수의 좌위들에 위치하는 적어도 하나의 핵산 분자를 포함한다. 한 가지 구현예에서, 본 그룹은 예로 상기 RHDO 방법에서와 같이, 염색체 부위의 특정한 반수체형에 정렬한 단편들일 수 있다. 또 다른 구현예에서, 본 그룹은 섹션 IV에서 기술된 방법들에서와 같이, 염색체 부위에 정렬한 단편이라면 모두일 수 있다.

[0159] 단계 1440에서, 컴퓨터 시스템은 핵산 분자들의 개별 그룹의 개별 수치를 계산한다. 개별 수치는 개별 그룹의

핵산 분자들의 특성을 정의한다. 개별 수치는 본 명세서에서 언급된 수치들이라면 모두일 수 있다. 예를 들어, 본 수치는 본 그룹에 있는 단편들의 수 또는 본 그룹에 있는 단편들의 크기 분포의 통계학적 수치일 수 있다. 개별 수치는 역시 정상화된 수치, 예로 시료를 위한 태그 계수들의 전체 수 또는 기준 부위를 위한 태그 계수들의 수로 나눈 부위의 태그 계수일 수도 있다. 개별 수치는 역시 또 다른 수치로부터 차이 또는 비율일 수도 있고 (예로, RHD0에서), 이로 인해 부위에 대한 차이의 특성을 제공한다.

[0160] 단계 1450에서, 개별 수치는 첫 번째 염색체 부위가 결실 또는 증폭을 나타내는지 여부의 부류를 결정하도록 기준 수치와 비교된다. 본 기준 수치는 본 명세서에서 기술된 역치 또는 기준 수치라면 모두일 수 있다. 예를 들어, 기준 수치는 정상 시료들을 위해 결정된 역치 수치일 수 있다. RHD0를 위해, 개별 수치는 두 개 반수체 형들 위한 태그 계수들의 차이 또는 비율일 수도 있고, 기준 수치는 통계학적으로 유의한 편차가 존재하는 것을 결정하는 역치일 수 있다. 또 다른 예로서, 기준 수치는 또 다른 반수체형 또는 부위를 위한 태그 계수 또는 크기 수치일 수 있고, 비교는 차이 또는 비율 (또는 이러한 함수)을 취한 다음 차이 또는 비율이 역치 수치 이상인지 여부를 결정하는 것을 포함할 수 있다.

[0161] 기준 수치는 다른 부위들의 결과를 기초로 하여 다양화할 수 있다. 예를 들어, 이웃하는 부위들도 역시 편차를 보이는 경우라면 (하나의 역치와 비교하여 작더라도, 예로 z-점수 3), 저부 역치가 사용될 수 있다. 예를 들어, 세 개의 연속적 부위들이 모두 첫 번째 역치 초과인 경우라면, 암은 더욱 가능성이 있을 수 있다. 따라서, 본 첫 번째 역치는 비-연속적 부위들로부터 나온 암을 확인하는 데 요구되는 또 다른 역치보다 낮을 수 있다. 작은 편차라도 가지는 세 개 (또는 세 개 이상) 부위들을 가지는 것은 민감도 및 특이도가 유지될 수 있는 기회 효과의 충분하게 낮은 확률을 가질 수 있다.

[0162] 단계 1460에서, 결실 또는 증폭을 나타내는 것으로서 분류된 염색체 부위들의 양이 결정된다. 계수되는 염색체 부위들은 한계들을 가질 수 있다. 예를 들어, 적어도 하나의 다른 부위와 근접한 부위들만이 계수될 수 있다 (또는 근접한 부위들은 소정의 크기, 예로 4개 이상의 부위들이 되도록 요구될 수 있다). 부위들이 동등하지 않은 구현예들을 위해, 수는 역시 개별 길이들도 설명할 수 있다 (예로, 수는 이상 부위들의 전 길이일 수 있다).

[0163] 단계 1470에서, 양은 시료의 분류를 결정하도록 역치 수치의 양과 비교된다. 예들로서, 분류는 유기체가 암을 가지는 여부, 암의 단계, 및 암의 예후일 수 있다. 한 가지 구현예에서, 모든 이상 부위들이 계수되고 단일 역치 수치가 부위들이 출현하는 곳과는 상관없이 사용된다. 또 다른 구현예에서, 역치 수치는 계수되는 부위들의 위치들 및 크기를 기초로 하여 다양화될 수 있다. 예를 들어, 특정한 염색체 또는 염색체의 팔 상의 부위들의 양은 해당 특정한 염색체 (또는 팔)에 대한 역치와 비교될 수 있다. 복수의 역치들이 사용될 수 있다. 예를 들어, 특정한 염색체 또는 염색체의 팔 상의 이상 부위들의 양은 첫 번째 역치 수치 이상이어야 하고, 계놈에서 이상 부위들의 전체량은 두 번째 역치 수치 이상이어야 한다.

[0164] 부위들의 양을 위한 본 역치 수치는 계수되는 부위들을 위한 불균형이 강한 정도에도 역시 의존할 수 있다. 예를 들어, 암의 분류를 결정하는 역치로서 사용되는 부위들의 양은 각 부위에서 이상을 검출하는 데 사용되는 특이도 및 민감도 (이상 역치)에 의존할 수 있다. 예를 들어, 이상 역치가 낮은 (z-점수 2) 경우라면, 양 역치는 낮도록 (예로, 150) 선택될 수 있다. 그러나, 이상 역치가 높은 (z-점수 3) 경우라면, 양 역치는 더 낮아질 수 있다 (예로, 50). 이상을 보이는 부위들의 양은 역시 평가된 수치 (weighted value)일 수도 있고, 예로 높은 불균형을 보이는 부위는 약간의 불균형만을 보이는 부위보다 높게 평가될 수 있다 (예로, 이상에 대해 단지 양성 및 음성인 것보다 더 많은 분류들이 존재한다).

[0165] 따라서, 정상화된 태그 계수 (또는 그룹의 특성을 위한 다른 개별 수치)의 유의한 과다- 또는 과소-발현을 보여주는 염색체 부위들의 양 (수 및/또는 크기를 포함할 수 있음)이 질환의 중증도를 반영하는 데 사용될 수 있다. 이상 정상화된 태그 계수를 가진 염색체 부위들의 양은 두 가지 요인들, 즉 종양 조직들에서 염색체 이상들의 수 (또는 크기) 및 생물학적 시료 (예로, 혈장)에서 종양-유래 DNA의 분획 농도에 의해 결정될 수 있다. 더 진행된 암들은 더 많은 (및 더 큰) 염색체 이상들을 보이는 경향이 있다. 따라서, 더 많이 암-연관된 염색체 이상들은 시료 (예로, 혈장)에서 잠재적으로 검출가능할 것이다. 더 진행된 암을 가진 환자들에서, 더 높은 종양 부하는 혈장에서 종양-유래 DNA의 더 높은 분획 농도를 유도할 것이다. 그 결과로, 종양-연관된 염색체 이상들은 혈장 시료에서 더 쉽게 검출될 것이다.

[0166] 암 검색 또는 검출의 문맥에서, 정상화된 태그 계수 (또는 다른 수치)의 과다- 또는 과소-발현을 보이는 염색체 부위들의 양은 암을 가지는 테스트된 개체의 가능성을 결정하는 데 사용될 수 있다. 컷오프 ± 2 를 사용하여 (예로, z-점수 > 2 또는 < -2), 테스트된 부위들의 대략 5%는 기회 단독으로 인해 대조군 개체들의 평균으로부터

터 유의하게 벗어난 z -점수를 줄 것으로 예상될 것이다. 전체 게놈이 1 Mb 분절들로 분할될 때, 전체 게놈을 위해 대략 3,000개 분절들이 존재할 것이다. 따라서, 대략 150개 분절들이 z -점수 > 2 또는 < -2 를 가질 것으로 예상될 것이다.

[0167] 따라서, z -점수 > 2 또는 < -2 를 가진 분절들의 수를 위한 150의 컷오프 (역치) 수치는 암이 존재하는지 여부를 결정하는 데 사용될 수 있다. 이상 z -점수 (예로, 100, 125, 175, 200, 250 및 300)을 가진 분절들의 수를 위한 다른 컷오프 수치들은 진단적 목적을 맞추도록 선택될 수 있다. 저부 컷오프 수치, 예로 100은 더 민감한 테스트를 유도할 것이지만 더 낮은 특이도 및 더 높은 컷오프 수치는 더욱 특이적이지만 덜 민감할 것이다. 위양성 분류들의 수는 z -점수의 컷오프 수치들을 증가시켜서 감소될 수 있다. 예를 들어, 컷오프 수치가 3으로 증가되는 경우라면, 분절들의 0.3%만이 위양성일 것이다. 본 상황에서, 이상 z -점수를 가지는 3개 이상의 분절들이 암의 존재를 가리키는 데 사용될 수 있다. 다른 컷오프 수치들도 역시 서로 다른 진단적 목적들을 맞추도록, 예로 1, 2, 4, 5, 10, 20 및 30이 선택될 수 있다. 그러나, 암-연관된 염색체 이상들을 검출하는 민감도는 진단하는 데 요구되는 이상 분절들의 수를 증가시키면서 감소될 것이다.

[0168] 특이도를 희생시키지 않고도 민감도를 개선하는 한 가지 가능한 접근법은 인접한 염색체 분절의 결과를 고려하는 것이다. 한 가지 구현예에서, z -점수를 위한 컷오프는 여전히 > 2 및 < -2 가 된다. 그러나, 염색체 부위는 두 개의 연속적 분절들이 동일한 유형의 이상들을 보이는, 예로 둘 다의 분절들이 z -점수 > 2 를 가질 때에만 잠재적 이상으로서 분류될 것이다. 정상화된 태그 계수의 편차가 무작위적 오류인 경우라면, 동일한 방향으로 위양성이 되는 두 개의 연속적 분절들을 가지는 확률은 0.125% ($5\% \times 5\%/2$)일 것이다. 한편, 염색체 이상이 두 개 연속적 분절들을 둘러싸는 경우라면, 저부 컷오프 수치는 혈장 시료에서 분절들의 과다- 또는 과소-발현의 검출을 더욱 민감하게 만들 것이다. 대조군 개체들의 평균으로부터 정상화된 태그 계수 (또는 다른 계수)의 편차가 무작위 오류로 인하지 않기 때문에, 연속적 분류 요구 조건은 민감도에 유의한 역효과를 가지지 않을 것이다. 다른 구현예들에서, 이웃하는 분절들의 z -점수가 더 높은 컷오프 수치를 사용하여 다함께 첨가될 수 있다. 예를 들어, 세 개 연속적 분절들의 z -점수가 합산될 수 있고 컷오프 수치 5가 사용될 수 있다. 본 개념은 세 개 이상의 연속적 분절들로 확장될 수 있다.

[0169] 양 및 이상 역치들의 조합은 분석의 목적, 및 유기체의 선행 지식이라면 모두 (또는 그의 결여)에도 역시 의존할 수 있다. 예를 들어, 암에 대해 정상의 건강한 집단을 검색하는 경우라면, 전형적으로 잠재적으로 부위들의 둘 다의 양에서 높은 특이도 (예로, 부위들의 수를 위한 높은 역치) 또한 부위가 이상을 가지는 것으로서 확인될 때를 위한 이상 역치를 사용할 것이다. 그러나, 더 높은 위험을 가진 환자 (예로, 종괴 (lump) 또는 가족력, 흡연자, HPV 바이러스, 간염 바이러스, 또는 기타 바이러스들을 호소하는 환자)에서, 다음으로 역치들은 더 높은 민감도 (더 적은 위음성들)를 가지도록 낮아질 수 있다.

[0170] 한 가지 구현예에서, 염색체 이상을 검출하기 위해 종양-유래 DNA의 1 Mab 해상도 및 6.3%의 저부 검출 한계를 사용하는 경우라면, 각 1 Mb 분절에서 분자들의 수는 60,000개가 되는 것이 필요할 것이다. 이것은 전체 게놈을 위한 정렬가능한 대략 180 백만 개 (60,000개 해독/Mb $\times$ 3,000 Mb)로 해석될 것이다.

[0171] 도 15는 본 발명의 구현예들에 따라 종양-유래 단편들의 다양한 수들의 분절들 및 분획 농도를 위해 요구되는 깊이를 도시하는 표 1500을 나타낸 것이다. 컬럼 1510은 시료를 위해 종양 세포들로부터 나온 단편들의 농도를 제공한다. 농도가 높을수록, 이상을 검출하는 것이 더 쉽고, 그래서 더 적은 수의 분자들이 분석되도록 요구된다. 컬럼 1520은 깊이에 관하여 상기 본 섹션에서 기술된 방법을 통하여 계산될 수 있는 분절 당 요구되는 분자들의 추정된 수를 제공한다.

[0172] 분절 크기가 작을수록 더 작은 염색체 이상들을 검출하는 해상도가 더 높아질 것이다. 그러나, 이것은 전체적으로 분석되는 분자들의 수의 요구 조건을 증가시킬 것이다. 분절 크기가 커질수록 해상도를 희생하여 분석에 요구되는 분자들의 수를 감소시킬 것이다. 따라서, 더 큰 이상들만이 검출될 수 있다. 한 가지 시행예에서, 더 큰 부위들이 사용될 수 있고 이상을 보이는 분절들은 소분할되고 이들 소부위들은 더 나은 해상도를 (예로, 상기에 기술된 바와 같이) 획득하도록 분석될 것이다. 컬럼 1530은 각 분절의 크기를 제공한다. 컬럼 1540은 전체 게놈을 위해 분석되는 분자들의 수를 보여준다. 따라서, 추정치 (또는 검출될 최소의 농도)를 가지는 경우라면, 분석될 분자들의 수가 결정될 수 있다.

[0173] VI. 계속적인 진행

[0174] 종양이 진행하면서, 종양 단편들의 양은 종양이 더 많은 DNA 단편들을 방출할 것이기 때문에 (예로, 종양의 성

장, 더 많은 괴사, 또는 더 높은 증가할 것이다. 혈장 내로 종양 조직들로부터 나온 더 많은 DNA 단편들은 혈장에서 불균형 (예로, RHDO에서 두 개 반수체형들 간의 태그 계수들에서 차이)의 정도를 증가시킬 것이다. 추가적으로, 종양 단편들의 수가 증가하기 때문에 이상이 존재하는 부위들의 수는 더 쉽게 검출될 수 있다. 예를 들어, 부위에 대한 종양 DNA의 양이 매우 적을 수 있어, 종양이 작고 적은 양의 암 DNA 단편들의 방출하고 있을 때는 충분치 않은 단편들이 분석되기 때문에 통계학적으로 유의한 차이가 확립될 수 없어서 이상이 검출될 수 없다. 종양이 작을 때에라도 더 많은 단편들이 분석될 수 있지만, 더 큰 시료 (예로, 많은 혈장)를 요구할 수 있다.

[0175] 암의 진행을 추적하는 것은 하나 이상의 부위들에서 이상의 양 (예로, 불균형 또는 요구되는 깊이)에 의해) 또는 이상을 나타내는 염색체 부위들의 양 (수 및/또는 크기)를 사용할 수 있다. 한 가지 구현예에서, 하나의 부위 (또는 여러 부위들)의 이상의 양이 다른 부위들의 이상들보다 더 빠르게 증가하는 경우라면, 해당 부위(들)은 암을 감시하는 바람직한 마커로서 사용될 수 있다. 본 증가는 커지고 있어 많은 단편들을 방출하는 종양의 결과, 및/또는 부위가 많은 횡수로 증폭된 점일 수 있다. 수술 이후에 종양이 적절하게 제거되었던 점을 입증하도록 이상 수치 (예로, 이상의 양 또는 이상들을 보이는 부위들의 수, 또는 그들의 조합)도 역시 감시할 수 있다.

[0176] 기술학의 다양한 시행예들에서, 종양성 DNA의 분획 농도는 암의 단계를 결정하거나, 예후를 판단하거나, 이의 진행을 감시하는 데 사용된다. 측정된 진행은 암의 현재 단계 및 암이 자라거나 퍼지는 속도에 관한 정보를 제공할 수 있다. 암의 "단계"는 다음의 모두 또는 일정과 관련된다: 종양의 크기, 조직학적 외관, 림프절 관여의 존재/부재, 및 먼 전이들의 존재/부재. 암의 "예후 판단"은 질환 진행의 기회 및/또는 암으로부터 생존의 기회를 추정하는 것이 관여한다. 이것은 환자가 임상적 진행이 안될 시간의 추정, 또는 생존의 지속기간도 역시 관여할 수 있다. 암의 "감시"는 암이 진행되었는지 (예로, 크기에서 증가하였는지, 림프절들의 관여가 증가하였는지, 또는 먼 기관으로 퍼졌는지, 예로 전이들) 여부를 관찰하도록 체크하는 것이 관여할 것이다. 감시는 종양이 치료에 의해 통제되었는지 여부를 검토하는 것도 역시 포함할 수 있다. 예를 들어, 치료가 효과적인 경우라면 종양의 크기에서 감소, 전이들 또는 림프절 관여에서의 퇴행, 환자의 일반적인 복지에서의 개선 (예로, 체중의 증가)을 보일 수 있다.

[0177] A. 암 DNA의 분획 농도의 결정

[0178] 하나 이상의 부위들에 대한 이상에서 증가의 양을 추적하는 한 가지 방식은 부위(들)에 대한 암 DNA의 분획 농도를 결정하는 것이다. 다음으로 암 DNA의 분획 농도에서 변화는 계속 종양을 추적하는 데 사용될 수 있다. 본 추적은 진단하는 데 사용될 수 있고, 예로 첫 번째 측정은 (사람의 일반적인 이상의 수준에 해당할 수 있는) 기저 수준을 제공할 수 있고 이후 측정들은 종양 성장 (따라서 암)을 제시할 변화를 보여줄 수 있다. 암 DNA의 분획 농도에서 변화는 치료가 잘 시행되는 정도를 예후 판단하는 데도 역시 사용될 수 있다. 기술학의 다른 시행예들에서, 혈장에서 종양성 DNA의 분획 농도에서 증가는 환자에서 나쁜 예후 또는 종양 부하에서 증가를 가리킬 것이다.

[0179] 암 DNA의 분획 농도는 다양한 방식들로 결정될 수 있다. 예를 들어, 하나의 반수체형으로부터 나온 태그 계수들에서 차이는 또 다른 것과 비교된다 (또는 하나의 부위가 또 다른 것과 비교됨). 또 다른 방법은 통계학적으로 유의한 차이가 관찰되기 이전의 깊이이다 (예로, 분석된 단편들의 수). 초기의 예의 경우, 반수체형 용량에서 차이는 이중집합의 소실을 가진 염색체 부위들을 분석하여 생물학적 시료 (예로, 혈장)에서 종양-유래 DNA의 분획 농도를 결정하는 데 사용될 수 있다.

[0180] 종양-유래 DNA가 암 환자들에서 종양 부하와 양성적으로 상관되는 점이 확인되어 왔다 (Lo et al. Cancer Res. 1999; 59: 5452-5. 및 Chan et al. Clin Chem. 2005; 51: 2192-5). 따라서, RHDO 분석에 의해 생물학적 시료에서 종양-유래 DNA의 분획 농도의 일련의 감시는 환자의 질환 진행을 감시하는 데 사용될 수 있다. 예를 들어, 치료 이후에 일련으로 수집된 시료들 (예로, 혈장)에서 종양-유래 DNA의 분획 농도의 감시는 치료의 성공을 결정하는 데 사용될 수 있다.

[0181] 도 16은 본 발명의 구현예들에 따라 RHDO 분석에 의해 혈장에서 종양-유래 DNA의 분획 농도를 측정하는 원리를 나타낸다. 두 개 반수체형들 간의 불균형이 결정되고 불균형의 정도는 시료에서 종양성 DNA의 분획 농도를 결정하는 데 사용될 수 있다. Hap I 및 Hap II는 비종양 조직들에서 두 개 반수체형들을 대표한다. 소부위 1610에 있는 Hap II는 종양 조직들에서 부분적으로 결실된다. 따라서, 혈장에서 결실된 결실된 부위 1610에 해당하

는 Hap II-연관된 단편들은 비종양 조직이 기여한다. 한편, Hap I에서 부위 1610은 종양 및 비종양 조직들 둘다에 존재한다. 따라서 Hap I 및 Hap II의 해독 계수들 간의 차이는 혈장에서 종양-유래 DNA의 양을 나타낼 것이다.

[0182] 종양-유래 DNA의 분획 농도 (F)는 LOH에 의해 침범된 염색체 부위들을 위한 결실된 및 비결실된 염색체들로부터 나온 서열결정된 해독 (태그들)의 수로부터 다음의 공식을 사용하여 계산될 수 있다: $F = (N_{\text{Hap I}} - N_{\text{Hap II}}) / N_{\text{Hap I}} \times 100\%$, 여기에서 $N_{\text{Hap I}}$ 는 LOH에 의해 침범된 염색체 부위들에 위치하는 이중접합 SNPs를 위한 Hap I 상의 대립인자들에 해당하는 서열결정된 해독의 수이고; $N_{\text{Hap II}}$ 는 LOH에 의해 침범된 염색체 부위 1610에 위치하는 이중접합 SNPs를 위한 Hap II 상의 대립인자들에 해당하는 서열결정된 해독의 수이다.

[0183] 상기 공식은 $F = 1 - q/p$ 로서 계산되는 시료 (F)에서 종양성 DNA의 분획 농도와 함께 p를 결실 (Hap I)을 포함하지 않는 염색체 부위 상에 위치하는 이중접합 좌위들에 대한 누적 태그 계수들로서, 또한 q를 결실 (Hap II) 1610을 포함하는 염색체 부위에 대한 누적 태그 계수들로서 정의하는 것과 동등하다. 도 11에 도시된 예의 경우, 종양성 DNA의 분획 농도는 14% ($1 - 104/121$)이다.

[0184] HCC 환자의 혈장 시료들에서 종양성 DNA의 분획 농도는 종양 절제 이전 및 이후에 수집되었다. 종양 절제 이전에 주어진 염색체 부위의 첫 번째 반수체형을 위한 $N_{\text{Hap I}}$ 는 30,443개이고, 염색체 부위의 두 번째 반수체형을 위한 $N_{\text{Hap II}}$ 는 16,221개이고, 이는 46.7%의 F를 준다. 종양 절제 이후에 $N_{\text{Hap I}}$ 는 31,534개이고, $N_{\text{Hap II}}$ 는 31,089개이고, 이는 1.4%의 F를 준다. 본 감시는 종양 절제가 성공적인 점을 보여준다.

[0185] 순환 DNA 크기 프로파일에서 변화의 정도는 분획 농도를 결정하는 데도 역시 사용될 수 있다. 한 가지 시행예에서, 종양 및 비종양 조직들 둘 다로부터 유래한 혈장 DNA의 정확한 크기 분포가 결정될 수 있고 다음으로 두 개의 기지의 분포들 간에 속하는 측정된 크기 분포는 분획 농도를 (예로, 종양 및 비종양 조직들의 크기 분포들의 두 개 통계학적 수치들 간의 선형 모델을 사용하여) 제공할 수 있다. 임의적으로, 크기 변화들의 일련의 감시가 사용될 수 있다. 한 가지 관점에서, 크기 분포에서 변화는 혈장에서 종양성 DNA의 분획 농도에 비례적인 것으로서 결정된다.

[0186] 서로 다른 부위들 간의 차이도 역시 유사한 방식으로, 예로 상기 기술된 비특이적 반수체형 검출 방법들로 사용될 수 있다. 태그 계수 방법들에서, 여러 매개변수들이 질환 진행의 감시에 사용될 수 있다. 예를 들어, 염색체 이상들을 나타내는 부위들을 위한 z-점수의 규모는 생물학적 시료 (예로, 혈장)에서 종양-유래 DNA의 분획 농도를 반영하는 데 사용될 수 있다. 특정한 부위의 과다- 또는 과소-발현의 정도는 시료에서 종양-유래 DNA의 분획 농도 및 종양 조직들에서 사본수 변화의 범위 또는 수와 비례적이다. z-점수의 규모는 대조군 개체들과 대비하여 시료에서 특정한 염색체 부위의 과다- 또는 과소-발현의 정도의 측정이다. 따라서, z-점수의 규모는 시료에서 종양 DNA의 분획 농도, 또한 향후 환자의 종양 부하를 반영할 수 있다.

[0187] B. 부위들의 수를 추적하는 것

[0188] 상기에서 언급된 바와 같이, 염색체 이상을 나타내는 부위들의 수는 암을 검색하는 데 사용될 수 있고, 이것은 마찬가지로 감시하고 예후를 판단하는 데 사용될 수 있다. 예들로서, 감시는 암의 현재 단계, 암이 재출현하는지 여부 및 치료가 작용하는지 여부를 결정하는 데 사용될 수 있다. 종양이 진행하면서, 종양의 게놈 작성이 더욱 해체될 것이다. 본 계속된 해체를 확인하기 위하여, 부위의 수 (예로, 1 Mb의 미리 정의된 부위들)를 추적하는 방법들은 종양의 진행을 확인하는 데 사용될 수 있다. 다음으로 더 진행된 암의 단계들에서 종양들은 이상을 나타내는 더 많은 부위들을 가질 것이다.

[0189] C. 방법

[0190] 도 17은 본 발명의 구현예들에 따라 핵산 분자들을 포함하는 생물학적 시료들을 사용하여 유기체에서 염색체 이상들의 진행을 결정하는 방법을 도시하는 흐름도를 나타낸다. 한 가지 구현예들에서, 적어도 일정의 핵산 분자들은 무세포이다. 예들로서, 염색체 이상들은 비악성 종양들 또는 전악성 병변들로부터 나올 수 있다. 또한, 이상들에서 증가는 염색체 이상들을 계속 포함하는 더욱 더 많은 세포들을 가지는 유기체로 인해, 또는 세포 당 이상들의 증가하는 수를 포함하는 세포들의 비율을 가지는 유기체로 인할 수 있다. 감소의 예로서, 치료 (예로, 수술 또는 화학요법)는 암과 연관된 세포들의 제거 또는 감소를 초래할 수 있다.

- [0191] 단계 1710에서, 유기체의 하나 이상의 비중복 염색체 부위들이 확인된다. 각 염색체 부위는 다수의 좌위들을 포함한다. 부위들은 적합한 방법이라면 모두, 상기에서 기술된 것들에 의해 확인될 수 있다.
- [0192] 단계들 1720 내지 1750는 각 다수의 시간대를 위해 수행된다. 각 시간대는 시료가 유기체로부터 획득되었던 서로 다른 시간에 해당한다. 현재 시료는 주어진 기간 동안 분석되고 있는 시료이다. 예를 들어, 시료는 6개월 동안 매월 취해질 수 있고, 분석은 시료가 획득된 이후 곧 시행될 수 있다. 임의적으로, 분석은 여러 번의 측정들이 여러 기간들의 시료로 취해진 이후에 시행될 수 있다.
- [0193] 단계 1720에서, 유기체의 현재의 생물학적 시료가 유기체의 기준 계놈에서 핵산 분자의 위치를 확인하도록 분석된다. 위치는 본 명세서에서 언급된 방식들이라면 모두로, 예로 서열결정된 태그들을 획득하도록 단편들을 서열결정하고 기준 계놈에 서열결정된 태그들을 정렬하여 결정될 수 있다. 분자의 특정한 반수체형도 역시 반수체형-특이적 방법들을 위해 결정될 수 있다.
- [0194] 단계들 1730 내지 1750는 하나 이상의 염색체 부위들을 위해 수행된다. 다수의 부위들이 사용될 때, 섹션 V로부터의 구현예들이 사용될 수 있다. 단계 1730에서, 핵산 분자들의 개별 그룹은 확인된 위치들을 기초로 하여 염색체 부위로부터 나온 것으로서 확인된다. 개별 그룹은 염색체 부위의 각 다수의 좌위들에 위치하는 적어도 하나의 핵산 분자를 포함한다. 한 가지 구현예에서, 본 그룹은 예로 상기 RHDO 방법에서와 같이, 염색체 부위의 특정한 반수체형에 정렬하는 단편들일 수 있다. 또 다른 구현예에서, 본 그룹은 섹션 IV에서 기술된 방법들에서와 같이, 염색체 부위에 정렬하는 단편이라면 모두일 수 있다.
- [0195] 단계 1740에서, 컴퓨터 시스템은 핵산 분자들의 개별 그룹의 개별 수치를 계산한다. 개별 수치는 개별 그룹의 핵산 분자들의 특성을 정의한다. 개별 수치는 본 명세서에서 언급된 수치들이라면 모두일 수 있다. 예를 들어, 본 수치는 본 그룹에 있는 단편들의 수 또는 본 그룹에 있는 단편들의 크기 분포의 통계학적 수치일 수 있다. 개별 수치는 역시 정상화된 수치, 예로 시료를 위한 태그 계수들의 전체 수 또는 기준 부위를 위한 태그 계수들의 수로 나눈 부위의 태그 계수일 수도 있다. 개별 수치는 역시 또 다른 수치로부터 차이 또는 비율일 수도 있고 (예로, RHDO에서), 이로 인해 부위에 대한 차이의 특성을 제공한다.
- [0196] 단계 1750에서, 개별 수치는 첫 번째 염색체 부위가 결실 또는 증폭을 나타내는지 여부의 부류를 결정하도록 기준 수치와 비교된다. 본 기준 수치는 본 명세서에서 기술된 역치 또는 기준 수치라면 모두일 수 있다. 예를 들어, 기준 수치는 정상 시료들을 위해 결정된 역치 수치일 수 있다. RHDO를 위해, 개별 수치는 두 개 반수체형들 위한 태그 계수들의 차이 또는 비율일 수도 있고, 기준 수치는 통계학적으로 유의한 편차가 존재하는 것을 결정하는 역치일 수 있다. 또 다른 예로서, 기준 수치는 또 다른 반수체형 또는 부위를 위한 태그 계수 또는 크기 수치일 수 있고, 비교는 차이 또는 비율 (또는 이러한 함수)을 취한 다음 차이 또는 비율이 역치 수치 이상인지 여부를 결정하는 것을 포함할 수 있다. 기준 수치는 적합한 방법 및 판정기준이라면 모두에 따라, 예로 본 명세서에서 기술한 바와 같이 결정될 수 있다.
- [0197] 단계 1760에서, 다수의 시간대에서 각 염색체 부위들의 분류들은 유기체에서 염색체 이상들의 진행을 결정하는데 사용될 수 있다. 본 진행은 유기체가 암을 가지는지 여부, 암의 단계, 및 암의 예후를 결정하는 데 사용될 수 있다. 각 이들 결정은 본 명세서에서 기술된 바와 같이 암의 부류에 관여할 수 있다.
- [0198] 본 암의 분류는 다양한 방식으로 수행될 수 있다. 예를 들어, 이상 부위들의 양은 계수되고 역치와 비교될 수 있다. 부위들을 위한 분류는 수적인 수치 (예로, 서로 다른 반수체형들 또는 서로 다른 부위들에 대한 수치가 되는 개별 및 기준 수치들로의 종양 농도) 및 농도에서의 변화가 결정될 수 있다. 농도의 변화는 유의한 증가가 일어났고 이로 인해 종양의 존재를 신호 전달하는 점을 결정하도록 역치와 비교될 수 있다.

도면의 간단한 설명

- [0199] 도 1은 결실의 이상을 보여주는 암세포의 염색체 부위를 나타낸 것이다.
- 도 2는 증폭의 이상을 보여주는 암세포의 염색체 부위를 나타낸 것이다.
- 도 3은 서로 다른 유형들의 암들 및 연관된 부위들 또한 그들의 해당 이상들을 도시하는 표 300을 나타낸 것이다.
- 도 4는 본 발명의 구현예들에 따라 혈장에서 작성된 측정들과 함께 이상을 보여주지 않는 암세포 내의 염색체 부위들을 나타낸 것이다.
- 도 5는 본 발명의 구현예들에 따라 결실된 부위를 결정하는 혈장에서 작성된 측정들과 함께 암세포 내의 염색체

부위 510의 결실을 나타낸 것이다.

도 6은 본 발명의 구현예들에 따라 증폭된 부위를 결정하는 혈장에서 작성된 측정들과 함께 암세포 내의 염색체 부위 610의 증폭을 나타낸 것이다.

도 7은 본 발명의 구현예들에 따라 종양 조직에서 단일-대립인자 증폭을 보였던 염색체 1p번에 위치하는 분절에 대한 간세포 암종 (HCC) 환자의 혈장 DNA의 RHDO 분석을 나타낸 것이다.

도 8은 본 발명의 구현예들에 따라 결실을 포함하는 종양이 존재할 때 염색체 부위의 두 개 반수체형들에 대한 단편들의 크기 분포에서 변화를 나타낸 것이다.

도 9는 본 발명의 구현예들에 따라 증폭을 포함하는 종양이 존재할 때 염색체 부위의 두 개 반수체형들에 대한 단편들의 크기 분포에서 변화를 나타낸 것이다.

도 10은 본 발명의 구현예들에 따라 염색체 부위가 결실 또는 증폭을 보여주는지 여부를 결정하도록 유기체의 생물학적 시료의 반수체형들을 분석하는 방법을 도시하는 흐름도를 나타낸 것이다.

도 11은 본 발명의 구현예들에 따라 결실된 부위를 결정하는 혈장에서 작성된 측정들과 함께 암세포들에서 결실되고 있는 소부위 1130을 가진 부위 1110을 나타낸 것이다.

도 12는 본 발명의 구현예들에 따라 이상들의 위치가 RHDO 분석을 사용하여 맵핑될 수 있는 방식을 나타낸 것이다.

도 13은 본 발명의 구현예들에 따라 또 다른 방향으로부터 시작된 RHDO 분류를 나타낸 것이다.

도 14는 본 발명의 구현예들에 따라 다수의 염색체 부위들을 사용하여 유기체의 생물학적 시료를 분석하는 방법 1400의 흐름도를 나타낸 것이다.

도 15는 본 발명의 구현예들에 따라 종양-유래 단편들의 다양한 수의 분절들 및 분획 농도를 위해 요구되는 깊이를 도시하는 표 1500을 나타낸 것이다. 도 15는 시료에서 서로 다른 백분율의 암-유래 DNA의 분획 농도를 위해 분석되는 분자들의 수의 추정을 제공한다.

도 16은 본 발명의 구현예들에 따라 혈장에서 종양-유래 DNA의 분획 농도를 상대적 반수체형 용량 (RHDO) 분석에 의해 측정하는 원리를 나타낸 것이다. 본 발명의 구현예들에 따라 Hap I 및 Hap II는 비-종양 조직들에서 두 개의 반수체형들을 표시한다.

도 17은 본 발명의 구현예들에 따라 핵산 분자들을 포함하는 생물학적 시료들을 사용하여 유기체에서 염색체 이상들의 진행을 결정하는 방법을 도시하는 흐름도를 나타낸 것이다.

도 18A는 암을 가진 환자의 염색체 4번의 q 팔 상의 분절을 위한 RHDO 분석의 SPRT 곡선을 나타낸 것이다. 점들은 개별 이종접합 좌위들 이후에 누적 계수들의 비율을 표시한다. 도 18B는 치료 이후에 환자의 염색체 4번의 q 팔 상의 분절을 위한 RHDO 분석의 SPRT 곡선을 나타낸 것이다.

도 19는 HCC에서 발견되는 공통적인 염색체 이상들을 나타낸 것이다.

도 20A는 HCC 및 건강한 환자들을 위한 정상화된 태그 계수들의 결과들을 표적화 분석을 사용하여 나타낸 것이다. 도 20B는 3명의 HCC 환자들 및 4명의 건강한 대조군 개체들을 위한 표적 강화 및 대규모 병렬 서열결정 이후에 크기 분석의 결과들을 나타낸 것이다.

도 21은 본 발명의 구현예들에 따라 혈장 DNA의 서열결정된 태그 계수로부터 얻은 데이터를 보여주는 HCC 환자의 키르코스 좌표들을 나타낸 것이다.

도 22는 본 발명의 구현예들에 따라 HCC가 없는 만성 B형 간염 바이러스 (HBV) 보균자의 혈장 시료에 대한 서열결정된 태그 계수 분석을 나타낸 것이다.

도 23은 본 발명의 구현예들에 따라 제 3기 코인두 암종 (NPC)을 가진 환자의 혈장 시료에 대한 서열결정된 태그 계수 분석을 나타낸 것이다.

도 24는 본 발명의 구현예들에 따라 제 4기 NPC를 가진 환자의 혈장 시료에 대한 서열결정된 태그 계수 분석을 나타낸 것이다.

도 25는 본 발명의 구현예들에 따라 종양 조직에서 이종접합의 소실 (LOH)을 보이는 부위를 위해 크기에 대비한

혈장 DNA의 누적 빈도의 좌표를 나타낸 것이다.

도 26은 LOH 부위를 위해 서열결정된 혈장 DNA의 크기에 대비한 ΔQ 를 나타낸 것이다. 본 발명의 구현예들에 따라 ΔQ 는 130 bp 크기에서 0.2에 도달한다.

도 27은 본 발명의 구현예들에 따라 종양 조직에서 염색체 복제를 가진 부위의 크기에 대비한 혈장 DNA의 누적 빈도의 좌표를 나타낸 것이다.

도 28은 본 발명의 구현예들에 따라 증폭된 부위를 위해 서열결정된 혈장 DNA의 크기에 대비한 ΔQ 를 나타낸 것이다.

도 29는 본 발명의 구현예들에 따른 시스템들 및 방법들로서 사용가능한 컴퓨터 시스템 900 예의 블록 도면을 나타낸 것이다.

발명을 실시하기 위한 구체적인 내용

VII. 실시예들

A. SPRT를 사용하는 RHDO

본 섹션에서, 우리는 간세포 암종 (HCC) 환자를 위해 SPRT를 사용하는 상대적 반수체형 용량 (RHDO) 분석을 사용하는 예를 보여주고 있다. 본 환자의 종양 조직에서, 염색체 4번 두 개 중 하나의 결실이 관찰되었다. 이것은 염색체 4번 상의 SNPs에 대한 이중접합의 소실을 가져온다. 본 환자의 반수체형 분석을 위해, 환자, 그의 아내, 및 그의 아들의 DNA가 분석되었고 3명 개인들의 유전형들이 결정되었다. 다음으로 환자의 전신적 반수체형들은 그들의 유전형들로부터 유래하였다. 대규모 병렬 서열결정이 수행되었고 염색체 4번의 두 개 반수체형들에 해당하는 SNP 대립인자들을 가지는 서열결정된 해독이 확인되고 계수되었다.

RHDO 및 SPRT의 공식들 및 원리들은 상기에 기술되었다. 한 가지 구현예에서, RHDO 분석은 예를 들어 두 개 반수체형들의 하나가 증폭되거나 결실될 때 종양-유래 DNA의 10%의 존재에 해당하고 있는 DNA 시료에서의 반수체형 용량들의 10% 차이를 검출하도록 프로그램될 것이다. 다른 구현예들에서, RHDO 분석의 민감도는 DNA 시료에서 종양-유래 DNA의 2%, 5%, 15%, 20%, 25%, 30%, 40% 및 50% 등을 검출하도록 설정될 수 있다. RHDO 분석의 민감도는 SPRT 분류 곡선들의 상부 및 저부 역치의 계산을 위한 매개변수에서 조정될 수 있다. 조정가능한 매개변수들은 검출 한계 (예로, 분석된 분자의 수에 영향을 주는 검출가능해야 하는 종양 농도의 백분율) 또한 예로 우위 비율 (다른 반수체형을 위한 태그 계수들과 대비한 하나의 반수체형을 위한 태그 계수들의 비율)을 사용하는 분류를 위한 역치의 원하는 수준일 수 있다.

본 RHDO 분석에서, 귀무가설은 염색체 4번을 위한 두 개 반수체형들이 동일한 용량으로 존재하는 점이다. 대안의 가설은 두 개 반수체형들의 용량이 생물학적 시료 (예로, 혈장)에서 10% 이상이 서로 다른 점이다. 두 개 반수체형들에 해당하는 SNP 대립인자들을 가지는 서열분석된 해독의 수가 서로 다른 SNPs로부터 얻은 데이터가 축적된 두 개 가설들에 대해 통계학적으로 비교되었다. SPRT 분류는 축적된 데이터가 두 개 반수체형 용량들이 동등한 양으로 존재하거나 적어도 10%가 통계학적으로 서로 다른지 여부를 결정하는 데 충분할 때 작성된다. 염색체 4번의 q 팔 상의 전형적인 SPRT 분류 블록은 도 18A에 나타낸다. 여기에서 10%의 역치는 설명적 목적으로만 사용된다. 기타 차이의 정도들 (예로, 0.1%, 1%, 2%, 5%, 15% 또는 20%)도 역시 검출될 수 있다. 일반적으로, 검출하기 원하는 차이의 정도가 낮아질수록 분석되는 데 필요한 DNA 분자들이 많아진다. 반대로, 검출하기 원하는 차이의 정도가 클수록 분석되고 통계학적으로 유의한 결과를 달성하는 데 필요한 DNA 분자들의 수가 적어진다. 본 분석을 위해, 우위 비율이 SPRT 분석에 사용되지만, z-점수 또는 p-수치와 같은 기타 매개변수들이 사용될 수 있다.

진단 시점에 취해진 HCC 환자의 본 혈장 시료에서, 염색체 4번의 p 및 q 팔들을 위한 각각 76개 및 148개의 성공적인 RHDO 분류들이 존재하였다. 모든 RHDO 분류들은 진단 시점에 취해진 혈장 시료에서 반수체형 용량 불균형이 존재하는 점을 가리킨다. 비교로서, 종양의 수술적 절제 이후 취해진 환자의 혈장 시료도 역시 도 18B에 나타난 바와 같이 분석되었다. 치료후 시료들의 경우, 염색체 4번의 p 및 q 팔들을 위한 각각 4개 및 9개의 성공적인 RHDO 분류들이 존재하였다. 모두 네 개의 RHDO 분류들은 혈장 시료에서 > 10%의 관찰가능한 반수체형 용량 불균형이 전혀 존재하지 않은 점을 가리킨다. 염색체 4q번을 위한 9개 RHDO 분류들 중에서, 7개는 반수체형 용량 불균형의 부재를 가리키고 2개는 불균형의 존재를 가리킨다. > 10%의 용량 불균형을 보이는 RHDO 블록들의 수는 종양 절제 이후에 유의하게 감소되었고, > 10%의 용량 불균형을 보이는 염색체 부위의 크기가 치료전 시료에서보다 치료후 시료에서 유의하게 작은 점을 가르킨다. 이들 결과들은 혈장에서 종양 DNA의 분획 농도가

종양의 수술적 절제 이후에 감소되었던 점을 제시하고 있다.

[0206] 비-반수체형 특이적 방법들과 비교할 때, RHDO 분석은 종양성 DNA의 분획 농도의 더 정확한 추정을 제공할 수 있고 질환 진행의 감시에 특히 유용하다. 따라서, 질환 진행을 가진 사례들이 혈장에서 종양성 DNA의 분획 농도에서 증가를 나타낼 것인 한편; 안정한 질환을 가진 사례들 또는 종양이 퇴행되거나 크기가 감소되었던 사례들은 혈장에서 종양성 DNA의 분획 농도에서 감소를 가질 것이라고 예상할 것이다.

[0207] B. 표적화된 분석

[0208] 선택된 구현예들에서, DNA 단편들의 보편적 서열결정은 표적 강화 접근법을 따라 수행될 수 있다. 이러한 접근법은 본 명세서에서 강화된 서열결정이라고도 말한다. 이러한 접근법의 한 가지 구현에는 용액 내 포획 시스템 (예로, Agilent SureSelect 시스템, 일루미나사 TruSeq Custom 강화 키트 (illumina.com/applications/sequencing/targeted_resequencing.ilmn))을 사용하거나, 마이게노스틱사의 GenCap Custom 강화 키트 (mygenostics.com/)에 의해 또는 마이크로어레이-기초 포획 시스템 (예로, 로슈사 NimbleGene 시스템)에 의한 단편들의 사전 선별이다. 일정 다른 부위들이 포획될 수 있더라도, 소정의 부위들이 선호적으로 포획된다. 이러한 방법들은 이러한 부위들이 더 큰 깊이로 (예로, 더 많은 단편들이 디지털 PCR로 서열결정되거나 분석될 수 있다) 및/또는 더 낮은 비용으로 분석되도록 허용할 수 있다. 더 큰 깊이는 부위에서 민감도를 증가시킬 수 있다. 다른 강화 방법들이 단편들의 크기 및 메틸화 양상들을 기초로 하여 수행될 수 있다.

[0209] 따라서, 게놈 범위로 DNA 시료를 분석하는 것의 대안은 공통적인 염색체 이상들을 검출하기 위해 관심 있는 부위들을 표적화하는 것이다. 표적화된 접근법은 분석적 공정이 염색체 이상들이 잠재적으로 존재하는 부위들 또는 특정한 종양 유형에 특히 특징적인 변화들을 가진 부위들, 또는 특히 임상적으로 중요할 변화들을 가진 부위들에 주로 중점을 둘 것이기 때문에 본 접근법의 비용-효율성을 잠재적으로 개선시킬 수 있다. 후자의 예들로는 특정한 암 유형의 종양원형성에서 초기에 발생할 변화들 (예로, 1q 및 8q의 증폭, 또한 8q의 결실의 존재는 HCC에서 초기 염색체 변화들이다 - van Malenstein et al. *Eur J Cancer* 2011;47:1789-97); 또는 좋거나 나쁜 예후와 연관된 변화들 (예로, 6q 및 17q에서 증가, 또한 6p 및 9p에서 소실이 종양 진행 동안 관찰되고, 18q, 8p 및 17p에서 LOH의 존재는 결장직장암 환자들에서 더 낮은 생존율과 연관된다 - Westra et al. *Clin Colorectal Cancer* 2004;4:252-9); 또는 치료 반응의 예측하는 변화들 (예로, 7p에서 증가의 존재는 표피성장인자 수용체 돌연변이들을 가진 환자들에서 타이로신 키나제 저해제들에 대한 반응을 예측한다 - Yuan et al. *J Clin Oncol* 2011;29:3435-42)을 포함한다. 암에서 변경되는 게놈 부위들의 다른 예들은 많은 온라인 데이터베이스 (예로, 암 게놈 해부 프로젝트 데이터베이스 (cgap.nci.nih.gov/Chromosomes/RecurrentAberrations)) 또한 유전학의 아틀라스 및 종양학 및 혈액학에서의 세포유전학 (atlasgeneticsoncology.org//Tumors/TumorListe.html)에서 찾을 수 있다. 대조적으로, 비표적화된 게놈범위 접근에서는 염색체 이상들이 발생하지 않을 것 같은 부위들이 잠재적 이상들을 가진 부위들과 동일한 정도로 분석될 것이다.

[0210] 우리는 3명의 HCC 환자들 및 4명의 건강한 대조군 개체들로부터 얻은 혈장 시료들을 분석하도록 표적-강화 전략을 적용하였다. 표적 강화는 애질런트사로부터 구입한 SureSelect 포획 시스템을 사용하여 수행되었다 (Gnirke et al. *Nat. Biotechnol* 2009;27:182-9). SureSelect 시스템은 가능한 표적 강화 기술학의 예로서 선택되었다. 다른 용액상 (IlluminaTruSeq Custom 강화 시스템) 또는 고체상 (예로, Roche-Nimblegen 시스템) 표적 포획 시스템들 및 증폭원점-기초 표적 강화 시스템 (예로, QuantaLife 시스템 및 RainDance 시스템)도 역시 사용될 수 있다. 포획 탐침들은 HCC에서 공통적으로 또한 현저하게도 이상들을 보여주는 염색체 부위들 상에 위치되도록 설계된다. 표적 포획 이후에, 다음으로 각 DNA 시료는 IlluminaGAIIx 분석기 상에서 유동 세포의 한 레인에 의해 서열결정되었다. 증폭 및 결실이 현저하게 발생하는 부위들은 증폭들 및 결실들이 더 공통적으로 존재하는 부위들과 비교하는 기준으로서 사용된다.

[0211] 도 19에서, HCC에서 발견되는 공통적인 염색체 이상들이 관찰된다 (도면은 왕 등으로부터 응용된다 (Wong et al., *Am J Pathol* 1999;154:37-43)). 염색체 핵형도 (chromosome ideogram)의 우측면 상의 선들은 염색체 증가들을 나타내고 좌측면 상의 선들은 개별 환자 시료들의 염색체 소실을 나타낸다. 굵은 선들은 높은 수준 증가들을 나타낸다. 직사각형들은 표적 포획 탐침들의 위치들을 나타낸다.

[0212] 표적화된 태그 계수 분석

[0213] 염색체 이상들의 검출을 위해, 우리는 먼저 잠재적인 이상들을 가진 부위들 및 기준 부위들을 위한 정상화된 태그 계수를 계산하였다. 다음으로 정상화된 태그 계수는 첸 등에 의해 이전에 기술된 바와 같이 부위의 GC 함량으로 교정되었다 (Chen et al., PLoS One 2011;6:e21791). 현재의 예에서, 염색체 8번의 p-팔이 잠재적인 이상의 부위로서 선택되었고 염색체 9번의 q-팔이 기준 부위로서 선택되었다. 세 명 HCC 환자들의 종양 조직들이 염색체 이상들에 대해 AffymetrixSNP 6.0를 사용하여 분석되었다. 종양 조직들에서 8p번 및 9q번에 대한 염색체 용량에서의 변화들이 하기 3명 환자들의 경우 하기에 관찰된다. 환자 HCC 013은 8p번의 소실 및 9q번의 무변화를 가졌다. 환자 HCC 027는 8p번의 증가 및 9q번의 무변화를 가졌다. 환자 HCC 028는 8p번의 소실 및 9q번의 무변화를 가졌다.

[0214] 다음으로 염색체 8p번 및 9q번 간의 정상화된 태그 계수의 비율이 세 명의 HCC 환자들 및 네 명의 건강한 대조군 개체들을 위해 표적화된 분석을 사용하여 계산되었다. 사례들 HCC 013 및 HCC 023의 경우, 감소된 8p번 및 9q번 간의 정상화된 태그 계수 비율이 관찰되었다. 이것은 종양 조직들에서 염색체 8p번의 소실의 연구와 부합한다. 사례 HCC 027의 경우, 증가된 비율이 관찰되었고 본 사례의 종양 조직들에서 염색체 8p번에서 증가와 부합한다. 점선들은 네 명의 정상 사례들의 평균 수치의 두 개 표준 편차들을 가진 부위를 나타낸다.

[0215] 표적화된 크기 분석

[0216] 이전의 섹션들에서, 우리는 암-연관된 변경들을 검출하는 원리를 암 환자들에서 혈장 DNA 단편들의 크기 프로파일을 결정하여 기술하고 있다. 크기 변경들도 역시 표적-강화 접근법으로 검출될 수 있다. 세 명의 HCC 사례들 (HCC 013, HCC 027 및 HCC 023)의 경우, 각 서열결정된 DNA 단편의 크기는 둘 다의 말단들의 최외부 뉴클레오타이드들의 좌표들로부터 추론되었다. 다른 구현예들에서, 전체 DNA 단편이 서열결정될 것이고 다음으로 단편 크기가 서열결정된 길이로부터 직접 결정될 수 있다. 염색체 8p번에 정렬된 DNA 단편들의 크기 분포는 염색체 9p번에 정렬된 DNA 단편들의 크기 분포와 비교되었다. 두 개 집단들의 DNA의 크기 분포에서 차이의 검출을 위해, 150 bp보다 짧은 DNA 단편들의 비율이 현재의 사례에서 각 집단을 위해 먼저 결정되었다. 다른 구현예들에서, 다른 크기의 컷오프 수치들, 예로 80 bp, 110 bp, 100 bp, 110 bp, 120 bp, 130 bp, 140 bp, 160 bp 및 170 bp가 사용될 수 있다. 다음으로 ΔQ 수치들이 두 개 비율들의 차이로서 결정되었다. $\Delta Q = Q_{8p} - Q_{9q}$, 여기서 Q_{8p} 는 150 bp보다 짧은 염색체 8p번에 정렬한 DNA 단편들의 비율이고; Q_{9q} 는 150 bp보다 짧은 염색체 9p번에 정렬한 DNA 단편들의 비율이다.

[0217] DNA 단편들의 더 짧은 크기 분포가 컷오프 수치 (예로, 현재의 예에서 150 bp)보다 짧은 DNA 비율의 더 높은 수치를 줄 것이기 때문에, ΔF 의 더 높은 (더 양성) 수치는 염색체 9p번에 정렬한 것들과 대비하여 염색체 8p번에 정렬한 DNA 단편들의 더 짧은 분포를 나타낼 것이다. 반대로, 더 작은 (또는 더 음성) 결과는 염색체 9p번에 정렬한 것과 대비하여 염색체 8p번에 정렬한 DNA 단편들의 더 긴 분포를 표시할 것이다.

[0218] 도 20B는 3명의 HCC 환자들 및 4명의 건강한 대조군 개체들을 위한 표적 강화 및 대규모 병렬 서열결정 이후에 크기 분석의 결과들을 나타낸다. 네 명의 건강한 대조군 개체들에서 ΔQ 의 양성 수치들은 염색체 9p번에 정렬한 것들과 대비하여 염색체 8p번에 정렬한 DNA 단편들의 약간 더 짧은 크기 분포를 표시한다. 점선들은 네 명의 대조군 개체들에 대한 평균으로부터 두 개 표준 편차 내에 속하는 ΔQ 의 간격을 나타낸다. 사례들 HCC 013 및 HCC 023은 대조군 개체들의 평균 수치 미만의 두 개 이상의 표준 편차들이었다. 이들 두 사례들은 종양 조직들에서 염색체 8p번의 결실을 가졌다. 종양에서 8p번의 결실은 본 염색체 부위의 경우 종양-유래 DNA의 혈장에 대한 감소된 기여를 가져올 것이다. 순환에서 종양 유래 DNA는 비종양 조직들로부터 유래한 DNA보다 짧기 때문에, 이것은 염색체 8p번에 정렬한 혈장 DNA 단편들의 경우 외관상 더 긴 크기 분포를 유도할 것이다. 이것은 이들 두 사례들에서 ΔQ 의 더 낮은 (더 음성) 수치와 부합된다. 대조적으로, 사례 HCC 027에서 8p번의 증폭은 본 부위에 정렬한 DNA 단편들에 대한 외관상 더 짧은 분포를 유도할 것이다. 따라서, 8p번에 정렬한 혈장 DNA 단편들의 더 높은 비율이 짧은 것으로 고려될 것이다. 이것은 HCC 027의 ΔQ 수치가 건강한 대조군 개체들보다 양성이라는 관찰과 부합된다.

[0219] C. 종양-유래 염색체 이상들의 검출을 위한 복수의 부위들

[0220] 소정의 염색체 부위들의 결실 및 증폭을 포함하는 염색체 이상들은 종양 조직들에서 공통적으로 검출된다. 염

색체 이상들의 특징적인 양상들은 서로 다른 유형들의 암들에서 관찰된다. 여기에서, 우리는 암 환자들의 혈장에서 이들 암-연관된 염색체 이상들을 검출하는 서로 다른 접근법들을 설명하도록 여러 예들을 사용하고 있다. 우리의 접근법은 암을 검색하고 질환 진행 및 치료에 대한 반응을 감시하는 데도 역시 유용하다. 한 명의 HCC 환자 및 두 명의 코인두 (NPC) 환자들로부터 얻은 시료들이 분석되었다. HCC 환자의 경우, 정맥혈 시료들이 종양의 수술절 절제 이전 및 이후에 수집되었다. 두 명의 NPC 환자의 경우 정맥혈 시료들이 진단 시점에 수집되었다. 추가적으로, 한 명의 만성 B형 간염 보균자 및 혈장에 검출가능한 엡스타인-바 바이러스 DNA를 가진 한 명의 개체의 혈장 시료들이 분석되었다. 이들 두 명의 개체들은 암이라면 모두 가지지 않았다.

[0221] 종양-유래 염색체 이상들의 검출은 마이크로어레이 분석으로 수행되었다. 상세하게, HCC 환자의 혈액 세포들 및 종양 시료로부터 추출된 DNA가 애피매트릭스사 SNP6.0 마이크로어레이 시스템을 사용하여 분석되었다. 혈액 세포들 및 종양 조직들의 유전형들이 애피매트릭스사 유전형분석 콘솔 v4.0을 사용하여 결정되었다. 증가들 및 결실들을 포함하는 염색체 이상들은 SNPs의 서로 다른 대립인자들의 강도들을 기초로 하는 Birdseed v2 알고리즘 및 마이크로어레이 상의 사본수 변화 (CNV) 탐침들을 사용하여 결정되었다.

[0222] 계수-기초 분석

[0223] 혈장에서 서열결정된 태그 계수 분석을 수행하기 위하여, 10 밀리리터의 정맥혈이 각 개체들로부터 수집되었다. 각 혈액 시료의 경우, 혈장이 시료의 원심분리 이후에 분리되었다. DNA는 QIAmp 혈액 미니키트 (키아젠사)를 사용하여 4 내지 6 mL의 혈장으로부터 추출되었다. 혈장 DNA 라이브러리는 이전에 기술된 바와 같이 제작되었고 (Lo YMD. Sci Transl Med 2010,2:61ra91), 다음으로 일루미나사 게놈 분석기 플랫폼을 사용하여 대규모 병렬 서열결정에 착수되었다. 각 분자는 50 bp씩 각 두 개 말단들에서, 합하여 분자 당 100 bp이 서열결정되었다. 각 서열의 두 개 말단들은 SOAP2 프로그램 (soap.genomics.org.cn/)을 사용하여 비-반복서열 차단된 인간 게놈 (UCSC genome.ucsc.edu로부터 다운로드되는 Hg18 NCBI.36)에 정렬되었다 (Li R et al. Bioinformatics 2009, 25:1966-7).

[0224] 다음으로 게놈은 복수의 1-메가염기 (1-Mb) 분절들로 분할되었고 각 1 Mb 분절에 정렬된 서열결정된 해독의 수가 결정되었다. 다음으로 각 bin의 태그 계수는 각 bin의 GC 함량에 따른 로칼 측정된 분산좌표 스무딩 (LOESS) 회귀분석을 기초로 하는 알고리즘으로 교정되었다 (Chen E et al. PLoS One 2011, 6:e21791). 본 교정은 서로 다른 게놈 분절들 간의 GC 함량에서 차이 때문에 일어나는 서열결정과 관련된 정량적 편차를 최소화하는 것을 목표로 한다. 상기 언급된 1 Mb 분절들로의 분할은 설명적 목적에 사용된다. 다른 분절 크기들, 예를 2 Mb, 10 Mb, 25 Mb, 또는 50 Mb 등도 역시 사용될 수 있다. 일반적으로 특정한 환자에서 특정한 종양 및 종양의 특정한 유형의 게놈적 특징들을 기초로 하여 분절 크기를 선택하는 것도 역시 가능하다. 또한, 서열결정 공정이 낮은 GC 편향을 가지는 것으로 관찰될 수 있는 경우라면, 예를 들어 헬리코스사 (Helicos) 시스템 (www.helicosbio.com) 또는 퍼시픽 바이오사이언스사 (Pacific Biosciences)의 단일 분자 실시간 시스템 (Single Molecular Real-Time system; www.pacificbiosciences.com)과 같은 단일 분자 서열결정 기법들의 경우 GC 교정 단계가 생략될 수 있다.

[0225] 이전의 연구에서, 우리는 암이 없는 개체들로부터 57개 혈장 시료들을 서열분석하였다. 이들 혈장 서열분석 결과들은 각 1 Mb 분절을 위한 태그 계수들의 기준 범위를 결정하는 데 사용되었다. 각 1 Mb 분절을 위해, 57명 개인들의 태그 계수들의 평균값 및 표준 편차가 결정되었다. 다음으로, 연구 개체들의 결과들이 다음의 방정식을 사용하여 계산된 바와 같이 z-점수로서 표현되었다: $z\text{-점수} = (\text{사례의 서열결정된 태그의 수} - \text{평균}) / \text{S.D.}$, 여기에서 "평균값 (mean)"은 기준 시료들을 위해 특정한 1 Mb에 정렬한 서열결정된 태그들의 평균 수이고; S.D.는 기준 시료들을 위해 특정한 1 Mb에 정렬한 서열결정된 태그들의 표준 편차이다.

[0226] 도 21 내지 도 24는 네 명의 연구 개체들의 서열결정된 태그 계수 분석의 결과들을 나타낸다. 1 Mb 분절들이 좌표들의 모서리에서 관찰된다. 인간 염색체 수들 및 핵형도들 (최외부 고리)은 시계 방향으로 pter-qter로 배치된다. 도 21에서, 내부 고리 2101은 종양을 분석하는 것으로부터 결정되는 바와 같은 이상 (결실 또는 증폭)의 부위들을 보여준다. 내부 고리 2101은 5개의 척도들과 함께 관찰된다. 척도는 -2 (최내부 선)으로부터 +2 (최외부 선)까지이다. 수치 -2는 해당하는 부위에 대한 염색체 사본들 둘 다의 소실을 나타낸다. 수치 -1은 두 개 염색체 사본들의 하나의 소실을 나타낸다. 수치 0은 염색체 증가 또는 소실이 없음을 나타낸다. 수치 +1은 한 개 염색체 사본의 증가를 나타내고, +2는 두 개 염색체 사본들의 증가를 나타낸다.

[0227] 중간 고리 2102는 혈장의 분석으로부터 나온 결과들을 보여준다. 관찰할 수 있는 바와 같이, 결과들은 내부 고

리를 반영한다. 중간 고리 2102는 척도의 더 많은 선들이지만 진행은 동일하다. 외부 고리 2103는 치료 이후에 혈장을 분석하는 것으로부터 얻은 데이터 요점들을 보여주고 이들 데이터 요점들은 회색이다 (과다/과소 발현 없음 - 이상 없음을 입증함).

[0228] 혈장에서 서열결정된 태그들의 과다-발현을 가진 염색체 부위들 (z -점수 > 3)은 녹색 점들 2110에 의해 나타낸다. 혈장에서 서열결정된 태그들의 과소-발현을 가진 부위들 (z -점수 < 3)은 적색 점들 2120에 의해 나타낸다. 혈장에서 검출되는 유의한 염색체 이상이 없는 부위들 (z -점수 -3 및 3 사이)는 회색 점들에 의해 나타낸다. 과다/과소 발현은 계수들의 전체 수에 의해 정상화된다. 서열결정 이전에 증폭으로, 정상화는 GC 편향을 고려할 수 있다.

[0229] 도 21은 본 발명의 구현예들에 따라 혈장 DNA의 서열결정된 태그 계수로부터 얻은 데이터를 보여주는 HCC 환자의 키르코스 좌표들을 나타낸다. 내부로부터 외부까지의 궤적들: 마이크로어레이 분석에 의해 검출된 종양 조직의 염색체 이상들 (적색 및 녹색은 각각 결실 및 증폭을 나타낸다); 종양의 수술적 절제 이전 및 절제 1개월 이후에 획득된 혈장 시료에 대한 z -점수 분석. 종양 절제 이전에, 혈장에서 검출된 염색체 이상들은 마이크로어레이 분석에 의해 종양 조직에서 확인되는 것들과 잘 상관되어 있다. 종양 절제 이후에, 대다수의 암-연관된 염색체 이상들은 혈장에서 사라졌다. 이들 데이터는 질환 진행 및 치료요법 효능을 감시하는 이러한 접근법의 가치를 반영하고 있다.

[0230] 도 22는 본 발명의 구현예들에 따라 HCC가 없는 만성 HBV 보균자의 혈장 시료에 대한 서열결정된 태그 계수 분석을 나타낸다. HCC 환자와는 대조적으로 (도 21), 암-연관된 염색체 이상들이 본 환자의 혈장에서 검출되지 않았다. 이들 데이터는 암 검색, 진단 및 감시를 위한 접근법의 가치를 반영하고 있다.

[0231] 도 23은 본 발명의 구현예들에 따라 제 3기 코인두 암종 (NPC)을 가진 환자의 혈장 시료에 대한 서열결정된 태그 계수 분석을 나타낸다. 염색체 이상들이 치료 이전에 취해진 혈장 시료에서 검출되었다. 상세하게, 유의한 이상들이 염색체들 1번, 3번, 7번, 9번, 및 14번에서 확인되었다.

[0232] 도 24는 본 발명의 구현예들에 따라 제 4기 NPC를 가진 환자의 혈장 시료에 대한 서열결정된 태그 계수 분석을 나타낸다. 염색체 이상들이 치료 이전에 취해진 혈장 시료에서 검출되었다. 제 3기 질환을 가진 환자들과 비교할 때 (도 23), 더 많은 염색체 이상들이 검출되었다. 서열결정된 태그 계수들도 역시 대조군들의 평균값으로부터 더욱 벗어났고, 예로 z -점수는 제로로부터 더욱 벗어나고 있다 (양성 또는 음성적으로 둘 중 하나). 대조군들과 대비하여 염색체 이상들의 증가된 수 및 서열결정된 태그 계수들의 편차의 더 높은 정도는 질환의 더욱 진행된 단계에서 게놈 변경들의 더 심각한 정도를 반영하고 있고, 따라서 암의 단계 결정, 예후판단 및 감시를 위한 이러한 접근법의 가치를 반영한다.

[0233] 크기-기초 분석

[0234] 이전의 연구들에서는, 종양 조직들로부터 유래한 DNA의 크기 분포가 비종양 조직들로부터 유래한 것들의 크기 분포보다 짧은 점을 보여주었다 (Diehl F et al. Proc Natl Acad Sci USA 2005, 102(45):16368-73). 이전의 섹션들에서, 우리는 혈장 DNA의 크기 분석에 의해 혈장 반수체형 불균형을 검출하는 접근법을 개괄하였다. 여기에서, 우리는 본 접근법을 좀 더 설명하도록 HCC 환자의 서열결정 데이터를 사용하였다.

[0235] 설명적 목적을 위해, 우리는 크기 분석을 위한 두 개 부위들을 확인하였다. 한 개 부위에서 (염색체 1번 (chr1); 좌표들: 159,935,347 내지 167,219,158), 두 개 상동 염색체들의 한 개의 복제가 종양 조직에서 검출되었다. 다른 부위에서 (염색체 10번 (chr10); 좌표들: 100,137,050 내지 101,907,356), 두 개 상동 염색체들의 한 개의 결실 (예로, LOH)가 종양 조직에서 검출되었다. 서열결정된 단편이 나온 반수체형을 결정하는 것과 추가하여, 서열결정된 단편의 크기도 역시 기준 게놈에 있는 서열결정된 단편의 최외부 뉴클레오타이드들의 좌표들을 사용하여 생물정보학적으로 결정되었다. 다음으로, 각 두 개의 반수체형들로부터 나온 단편들의 크기 분포들이 결정되었다.

[0236] Chr10의 LOH 부위의 경우, 하나의 반수체형이 종양 조직에서 결실되었다 (결실된 반수체형). 따라서, 본 결실된 반수체형에 정렬한 혈장 DNA 단편들 모두는 비-암 조직들로부터 유래하였다. 한편으로, 종양 조직들에서 결실되지 않았던 반수체형에 정렬한 단편들 (비-결실된 반수체형)은 종양 또는 비종양 조직들로부터 유래할 수 있다. 종양-유래 DNA의 크기 분포가 더 짧기 때문에, 우리는 결실된 반수체형으로부터 나온 것들과 비교할 때 비-결실된 반수체형으로부터 나온 단편들을 위한 더 짧은 크기 분포를 예상할 것이다. 두 개 크기 분포들에서의 차이는 DNA 단편들의 크기에 대비한 단편들의 누적 빈도들을 좌표 표시하여 결정될 수 있다. 더 짧은 크기 분

포를 가진 DNA 집단은 더욱 풍부한 짧은 DNA를 가지고, 따라서 크기 스펙트럼의 짧은 말단에서 누적 빈도의 더 신속한 증가를 보일 것이다.

[0237] 도 25는 본 발명의 구현예들에 따라 종양 조직에서 LOH를 보이는 부위를 위해 크기에 대비한 혈장 DNA의 누적 빈도의 좌표를 나타낸다. X-축은 염기쌍들로의 단편의 크기이다. Y-축은 X-축 상의 수치 미만의 크기를 가지는 단편들의 백분율이다. 비-결실된 반수체형으로부터 나온 서열들은 결실된 반수체형으로부터 나온 서열들과 비교할 때 더 신속한 증가 및 170 bp 크기 미만의 더 높은 누적 빈도를 가진다. 이것은 비-결실된 반수체형으로부터 나온 짧은 DNA 단편들이 더욱 풍부한 점을 가리킨다. 이것은 결실된 반수체형으로부터 나온 짧은 종양-유래 DNA의 기여 때문에 상기 예측과 부합된다.

[0238] 한 가지 구현예에서, 크기 분포에서 차이는 DNA 분자들의 두 개 집단들의 누적 빈도들에서 차이에 의해 정량될 수 있다. 우리는 ΔQ 를 두 개 집단들의 누적 빈도들에서 차이로서 정의한다. $\Delta Q = Q_{\text{비결실}} - Q_{\text{결실}}$, 여기에서 $Q_{\text{비결실}}$ 은 비-결실된 반수체형으로부터 나온 서열결정된 DNA 단편들을 위한 누적 빈도를 나타내고; $Q_{\text{결실}}$ 은 결실된 반수체형으로부터 나온 서열결정된 DNA 단편들을 위한 누적 빈도를 나타낸다.

[0239] 도 26은 LOH 부위를 위해 서열결정된 혈장 DNA의 크기에 대비한 ΔQ 를 나타낸다. 본 발명의 구현예들에 따라 ΔQ 는 130 bp 크기에서 0.2에 도달한다. 이것은 짧은 DNA를 정의하는 컷오프로서 130 bp를 사용하는 것이 상기 방정식들에서 사용에는 최적인 점을 가리킨다. 본 컷오프를 사용하여, 짧은 DNA 분자들이 결실된 반수체형으로부터 나온 집단과 비교할 때 비-결실된 반수체형으로부터 나온 집단에서 20% 더 풍부하다. 다음으로 본 백분율 차이 (또는 유사하게 유래된 수치)는 암이 없는 개인들로부터 유래한 역치 수치와 비교될 수 있다.

[0240] 염색체 증폭을 가진 부위의 경우, 한 개의 반수체가 종양 조직에서 복제된다 (증폭된 반수체형). 본 증폭된 반수체형으로부터 나온 짧은 종양-유래 DNA 분자들의 여분의 양이 혈장 내로 방출될 것이기 때문에, 증폭된 반수체형으로부터 나온 단편들의 크기 분포는 비-증폭된 반수체형으로부터 나온 단편들의 크기 분포보다 더 짧을 것이다. LOH 시나리오와 유사하게, 크기 분포에서 차이는 DNA 단편들의 크기에 대비한 단편들의 누적 빈도들을 좌표 표시하여 결정될 수 있다. 더 짧은 크기 분포를 가진 DNA 집단은 더욱 풍부한 짧은 DNA를 가지고, 따라서 크기 스펙트럼의 짧은 말단에서 누적 빈도의 더 신속한 증가를 보일 것이다.

[0241] 도 27은 본 발명의 구현예들에 따라 종양 조직에서 염색체 복제를 가진 부위의 크기에 대비한 혈장 DNA의 누적 빈도의 좌표를 나타낸다. 증폭된 반수체형으로부터 나온 서열들은 비-증폭된 반수체형으로부터 나온 서열들과 비교할 때 더 신속한 증가 및 170 bp 크기 미만의 더 높은 누적 빈도를 가진다. 이것은 비-증폭된 반수체형으로부터 나온 짧은 DNA 단편들이 더욱 풍부하였던 점을 가리킨다. 이것은 증폭된 반수체형으로부터 나온 짧은, 많은 수의 종양-유래 DNA의 기여 때문에 상기 예측과 부합된다.

[0242] LOH 시나리오와 유사하게, 크기 분포에서 차이는 DNA 분자들의 두 개 집단들의 누적 빈도들에서 차이에 의해 정량될 수 있다. 우리는 ΔQ 를 두 개 집단들의 누적 빈도들에서 차이로서 정의한다. $\Delta Q = Q_{\text{증폭}} - Q_{\text{비증폭}}$, 여기에서 $Q_{\text{증폭}}$ 은 증폭된 반수체형으로부터 나온 서열결정된 DNA 단편들을 위한 누적 빈도를 나타내고; $Q_{\text{비증폭}}$ 은 비-증폭된 반수체형으로부터 나온 서열결정된 DNA 단편들을 위한 누적 빈도를 나타낸다.

[0243] 도 28은 본 발명의 구현예들에 따라 증폭된 부위를 위해 서열결정된 혈장 DNA의 크기에 대비한 ΔQ 를 나타낸다. 본 발명의 구현예들에 따라 ΔQ 는 126 bp 크기에서 0.08에 도달한다. 이것은 짧은 DNA를 정의하는 컷오프로서 130 bp를 사용하는 것이 짧은 DNA 분자들이 비-증폭된 반수체형으로부터 나온 집단과 비교할 때 증폭된 반수체형으로부터 나온 집단에서 8% 더 풍부한 점을 가리킨다.

[0244] D. 추가적인 기법들

[0245] 다른 구현예들에서, 서열-특이적 기법들이 사용될 수 있다. 예를 들어, 올리고뉴클레오타이드들이 특정한 부위의 단편들과 혼성화하도록 설계될 수 있다. 다음으로 올리고뉴클레오타이드들이 서열결정된 태그 계수들과 유사한 방식으로 계수될 수 있다. 본 방법은 특정한 이상들을 나타내는 암들에 사용될 수 있다.

[0246] VIII. 컴퓨터 시스템

[0247] 본 명세서에서 언급된 컴퓨터 시스템들이라면 모두는 적합한 수의 서브시스템들이라면 모두를 사용할 수 있다.

이러한 시스템들의 예들은 도 19에서 컴퓨터 장치 900으로 나타나 있다. 일정 구현예들에서, 컴퓨터 시스템은 서브시스템들이 컴퓨터 장치의 구성성분들일 수 있는 단일한 컴퓨터 장치를 포함한다. 다른 구현예들에서, 컴퓨터 시스템은 내부의 구성성분들을 가진, 각각이 서브시스템이 되는 복수의 컴퓨터 장치들을 포함할 수 있다.

[0248] 도 29에 나타난 부속시스템들은 시스템 버스 2975를 통하여 상호접속된다. 프린터 2974, 키보드 2978, 고정된 디스크 2979, 디스플레이 어댑터 2982와 연결된 모니터 2976와 같은 부수적인 서브시스템들 및 기타 다른 서브시스템들을 도시하고 있다. 주변 장치들 및 입력/출력 (I/O) 컨트롤러 2971와 연결된 I/O 장치들은 직렬 포트 2977와 같은 당해 기술분야에 알려진 수단들이라면 모두에 의해 컴퓨터 시스템에 연결될 수 있다. 예를 들어, 직렬 포트 2977 또는 외부 인터페이스 2981 (예로, 이더넷, Wi-Fi 등)는 컴퓨터 시스템 2900을 인터넷, 마우스 입력 장치, 또는 스캐너와 같은 광면적 네트워크로 연결하는 데 사용될 수 있다. 시스템 버스 2975를 통한 상호접속은 중앙 프로세서 2973을 각 서브시스템과 통신하고 시스템 메모리 2972 또는 고정된 디스크 2979로부터 명령들의 실행, 뿐만 아니라 서브시스템들 간의 정보의 교환을 통제하도록 허용한다. 본 시스템 메모리 2972 또는 고정된 디스크 2979는 컴퓨터 판독가능한 매체를 구현할 수 있다. 본 명세서에서 언급된 수치들이라면 모두는 하나의 구성성분으로부터 다른 구성성분까지 출력될 수 있고 사용자에게 출력될 수 있다.

[0249] 컴퓨터 시스템은 예로 외부 인터페이스 2981에 의해 또는 내부 인터페이스에 의해 다함께 연결된 다수의 동일한 구성성분들 또는 서브시스템들을 포함할 수 있다. 일정 구현예들에서, 컴퓨터 시스템들, 서브시스템, 또는 장치들은 네트워크를 거쳐서 통신할 수 있다. 이러한 예들에서, 하나의 컴퓨터는 고객이라고 고려되고 또 다른 컴퓨터는 각각이 동일한 컴퓨터 시스템의 일부일 수 있는 서버로 고려될 수 있다. 고객 및 서버는 각각 복수의 시스템들, 서브시스템들, 또는 구성성분들을 포함할 수 있다.

[0250] 본 발명의 구현예들이라면 모두가 모듈의 또는 통합된 방식으로 하드웨어 및/또는 컴퓨터 소프트웨어를 사용하는 컨트롤 로직의 형태로 시행될 수 있는 점으로 이해되어야 한다. 본 명세서에서 제공된 기재 및 가르침들을 기초로 하여, 당업자라면 하드웨어 또한 하드웨어 및 소프트웨어의 조합을 사용하여 본 발명의 구현예들을 시행하도록 다른 방식들 및/또는 방법들을 숙지하고 이해할 것이다.

[0251] 본 출원에서 기술된 컴퓨터 소프트웨어 구성성분들 또는 기능들이라면 모두가, 예를 들어 자바, C++ 또는 Perl과 같은 적합한 컴퓨터 언어들이라면 모두를 사용하는 처리장치에 의해 실행되는 소프트웨어 코드로서 시행될 수 있다. 소프트웨어 코드는 저장 및/또는 전송을 위해 컴퓨터 판독가능한 매체에서 일련의 명령들 또는 명령어들이로서 저장될 수 있고, 적합한 매체로는 무작위 접근 메모리 (RAM), 읽기 전용 메모리 (ROM), 하드 드라이브 또는 플로피 디스크와 같은 자기 매체, 또는 콤팩트 디스크 (CD) 또는 DVD (디지털 다기능 디스크)와 같은 광학 매체, 플래시 메모리 등을 포함한다. 컴퓨터 판독가능한 매체는 이러한 저장 또는 전송 장치들의 조합이라면 모두일 수 있다.

[0252] 또한 이러한 프로그램은 인터넷을 포함하는 다양한 프로토콜들과 합치되는 유선 네트워크, 광학 네트워크 및/또는 무선 네트워크를 통한 전송에 적용되는 반송 신호들을 사용하여 코딩되고 전송될 수 있다. 이로써, 본 발명의 구현예에 따른 컴퓨터 판독가능한 매체는 이러한 프로그램들로 암호화된 데이터 신호를 사용하여 생성될 수 있다. 프로그램 코드로 암호화된 컴퓨터 판독가능한 매체는 호환 장치와 함께 패키징화될 수 있거나 다른 장치들로부터 (예로, 인터넷 다운로드를 통해) 별도로 제공될 수 있다. 이러한 컴퓨터 판독가능한 매체라면 모두가 단일 컴퓨터 프로그램 제품 (예로, 하드 드라이브, CD, 또는 전체 컴퓨터 시스템) 위에 또는 이의 내에 공존할 수 있고, 시스템 또는 네트워크 내의 다른 컴퓨터 프로그램 제품 위에 또는 내에 존재할 수 있다. 컴퓨터 시스템은 모니터, 프린터, 또는 본 명세서에서 언급된 결과들이라면 모두를 사용자에게 제공하는 다른 적합한 디스플레이를 포함할 수 있다.

[0253] 본 명세서에서 기술된 방법들이라면 모두는 단계들을 수행하도록 형성될 수 있는 프로세서를 포함하는 컴퓨터 시스템과 함께 전체적으로 또는 부분적으로 수행될 수 있다. 따라서, 구현예들은 잠재적으로 개별 단계들 또는 개별 그룹의 단계들을 수행하는 서로 다른 구성성분들과 함께, 본 명세서에서 기술된 방법들이라면 모두의 단계들을 수행하도록 형성된 컴퓨터 시스템들을 지향할 수 있다. 번호를 매긴 단계들로서 표현되더라도, 본 명세서에서 방법들의 단계들은 동시에 또는 서로 다른 순서로 수행될 수 있다. 추가적으로, 이들 단계들의 일부분들은 다른 방법들로부터 나온 다른 단계들의 일부분들과 함께 사용될 수 있다. 또한, 단계의 모두 또는 일부분들은 조건 선택적이다. 추가적으로, 방법들이라면 모두의 단계들이라면 모두는 모듈들, 회로들, 또는 이들 단계들을 수행하는 기타 수단들로 수행될 수 있다.

[0254] 특정한 구현예들의 자세한 내용은 본 발명의 구현예들의 정신 및 범위를 벗어나지 않는 적합한 방식이라면 모두로 조합될 수 있다. 그러나, 본 발명의 다른 구현예들은 각 개별적 관점, 또는 이들 개별적 관점들의 특정한

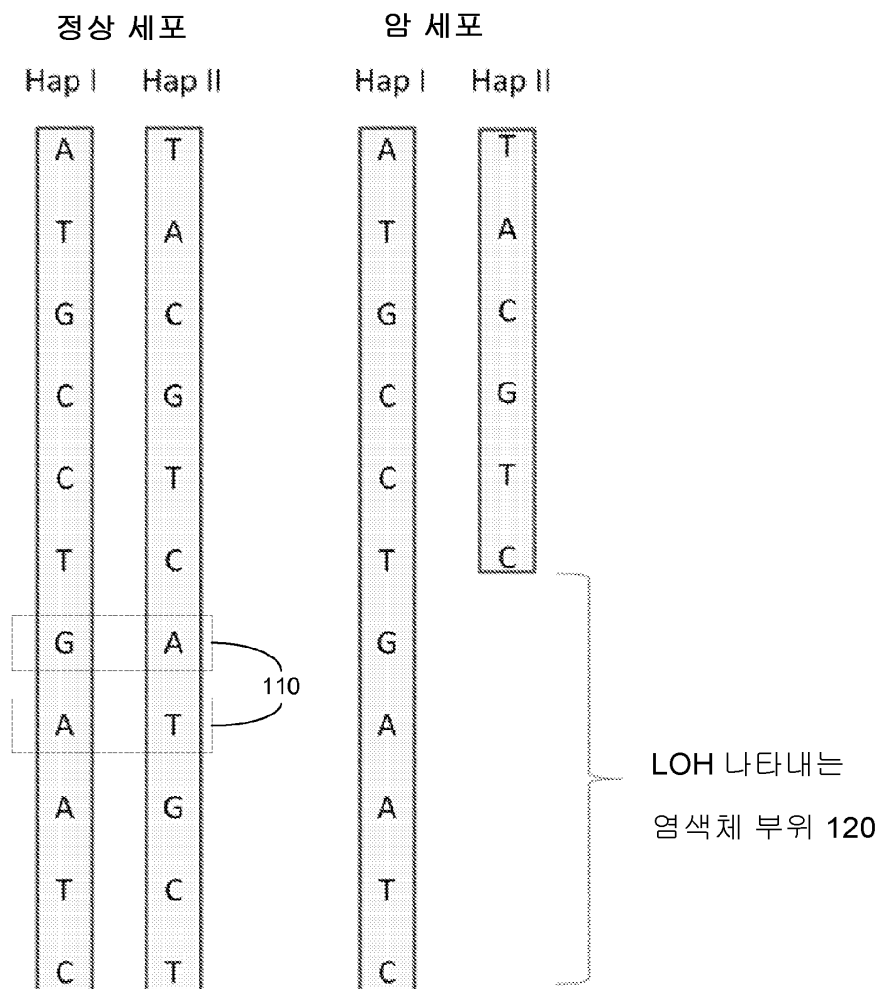
조합에 관한 특정한 구현예들을 지향할 수 있다.

[0255] 본 발명의 대표적인 구현예들의 상기 기재내용은 설명 및 기재의 목적들을 위해 기술되어 왔다. 이것은 본 발명을 기술된 정확한 형태로 나타내거나 제한하도록 의도되지 않고, 많은 변형들 및 변화들이 상기 가르침에 비추어 가능하다. 본 구현예들은 본 발명의 원리들 및 그의 실제적인 응용들을 가장 잘 설명하여 당업자들이라면 다양한 구현예들에서 또한 참작된 특정한 용도에 맞추어진 바와 같은 다양한 변형들과 함께 본 발명을 가장 잘 사용하도록 하기 위하여 선택되고 기술되었다.

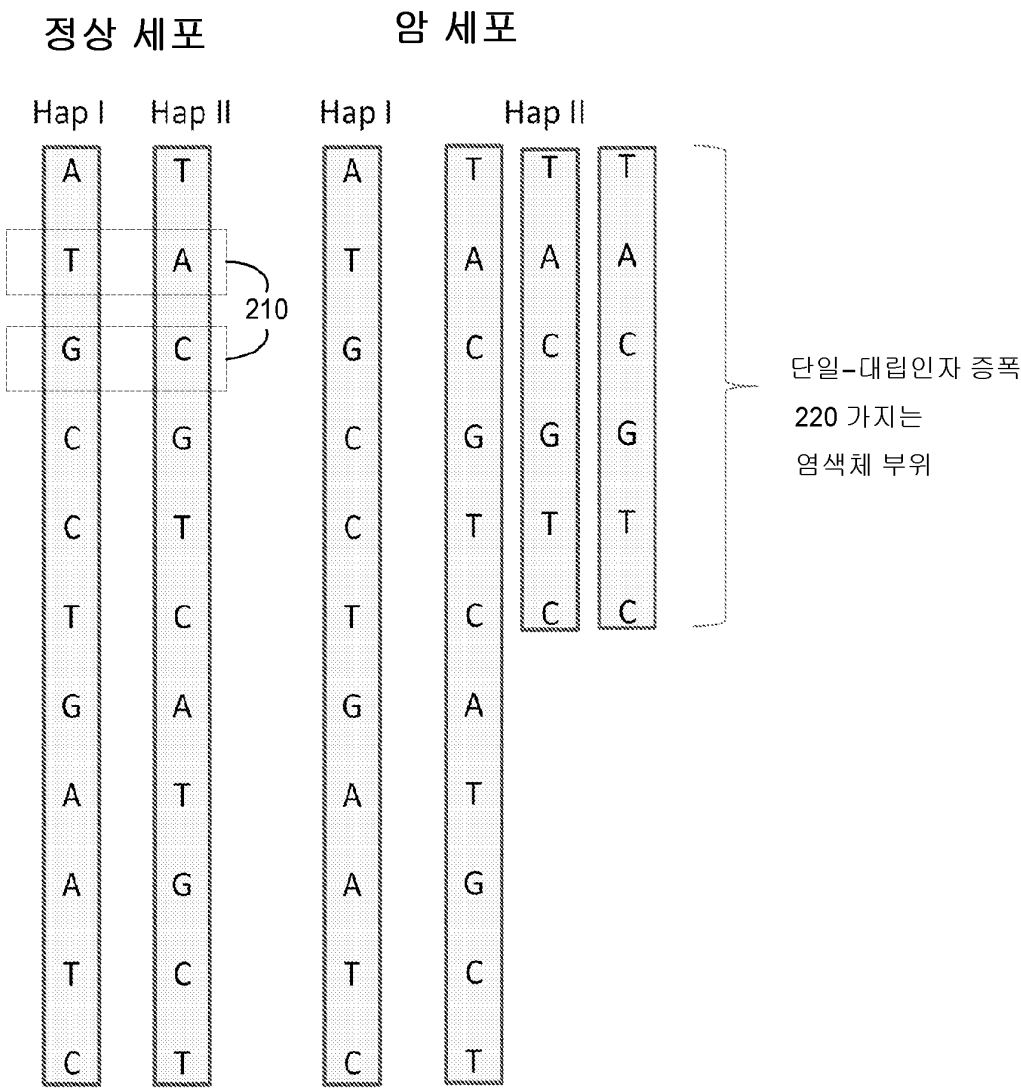
[0256] "하나 (a)", "하나 (an)" 또는 "그 (the)"의 설명은 특정하게 반대로 표시되지 않은 경우라면 "하나 이상 (one or more)"를 의미하도록 의도된다. 상기 언급된 모든 특허들, 특허출원들, 공고들, 및 기재내용들은 본 명세서에서 그들의 전부가 모든 목적들을 위한 참고문헌으로 통합되어 있다.

도면

도면1



도면2

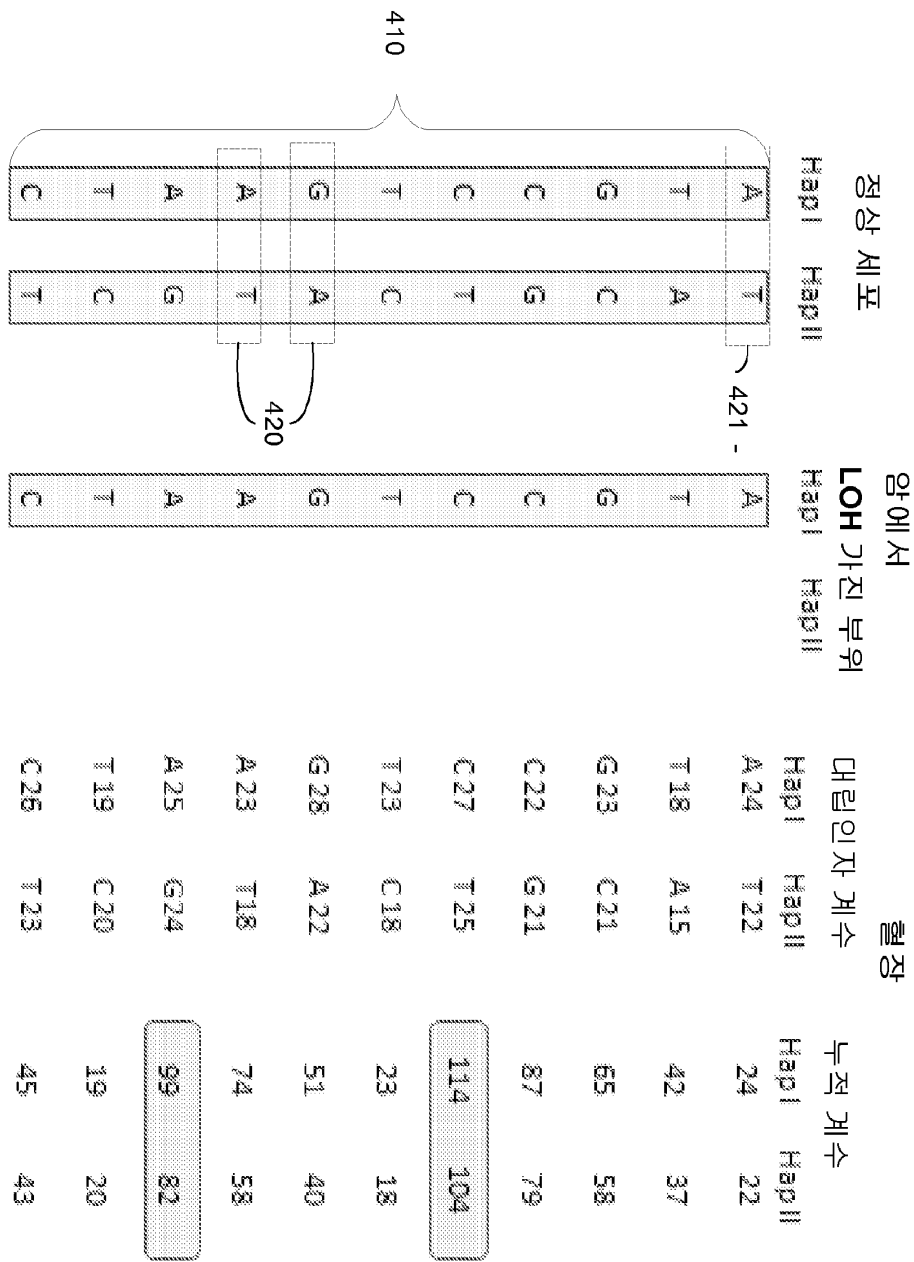


도면3

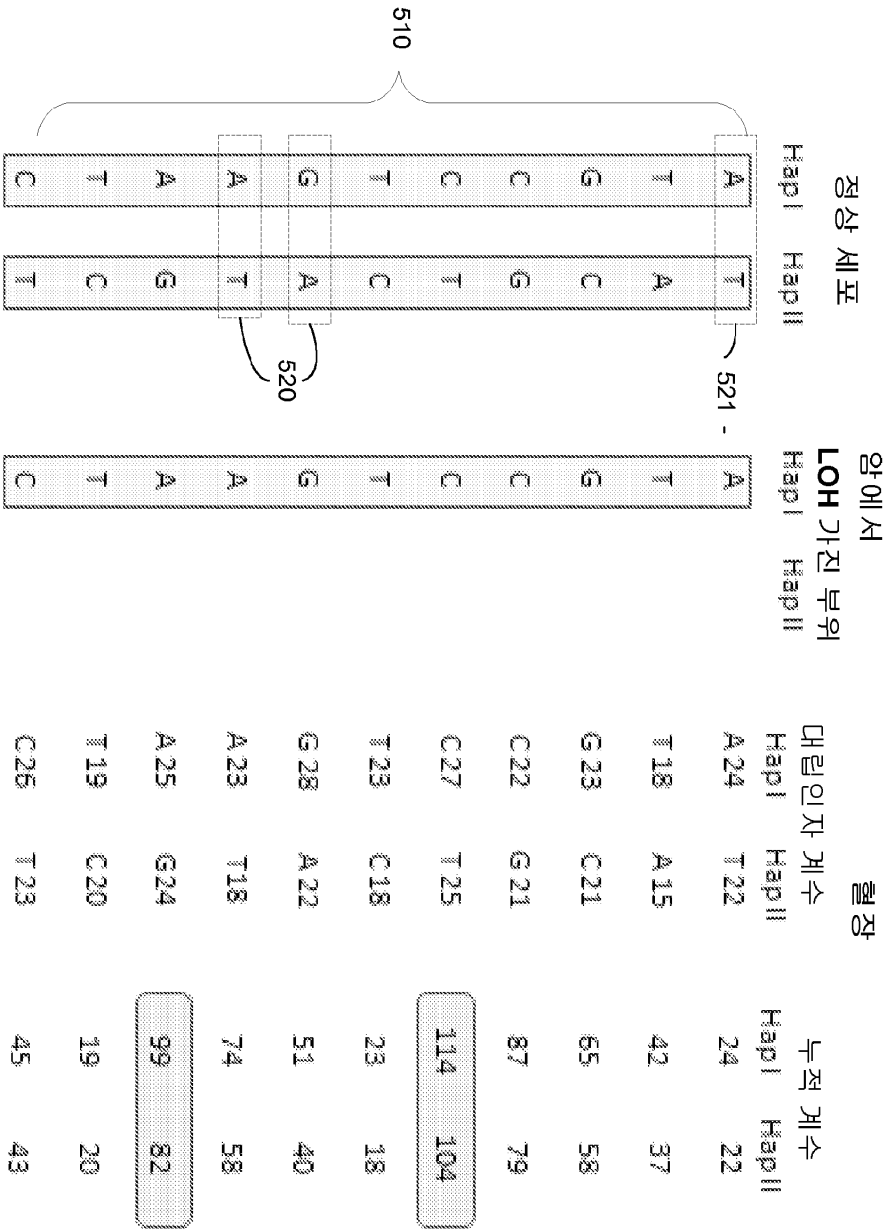
310	320	330	340
암 유형	증가	소실	참고문헌
Thyroid cancer	7p, 8q, and 9q	22	Hemmer et al. Am J Pathol, 1999;154:1539-47.
Gastric cancer	2q37, 3p21, 5q34-35, 7q34-36, 11q13, 11q23-25, 12q24, 15q23-25, 17q21-25, and 20q12-13	4q, 13q, 5q, 6q, and 18p	Noguchi et al. Am J Clin Pathol 2001; 115:828-34
Prostate cancer	1q32, 3q26, 4q26, 7q21, 8q22, 9q33, 17q25, and Xq21	2q22, 4q27-4q28, 5q15, 6q15, 8p21, 10q23, 12p13, 13q21, 15q23, 16q22, and 18q21-22	Sun et al. Prostate 2007;67:692-700.
Small cell lung cancer	3q26-29, 5p12-13, and 8q23-24	3p13-14, 4q32-35, 5q32-35, 8p21-22, 10q25, 13q13-14, and 17p12-13	Balsara et al. Oncogene 2002;21:6877-83.
Non small cell lung cancer	1q31, 3q25-27, 5p13-14, and 8q23-24	31p21, 8p22, 9p21-22, 13q22, and 17p12-13	Balsara et al. Oncogene 2002;21:6877-83.
Nasopharyngeal carcinoma	1p34, 3q26, 6q25, and 3q26	3p, 9p, 9q, 11q, 13q, and 14q	Lo et al. Semin Cancer Biol 2002;12:451-62
Bladder cancer	1q, 5p, 6p, 8q, 11q, 17q, and 20q	3p, 4q, 4q, 6q, 8p, 9p, and 18q	El-Rifai et al. Am J Pathol 2000;156:871-8
Colorectal cancer	13q and 20q	4q and 18q	De Angelis et al. Int J Colorectal Dis 2001;16:38-45.
Head and neck cancer	3q26 and 11q13	3p, 9p, and 17p	Smeets et al. Oncogene 2006;25:2558-64
Melanoma	1q, 2, 6p, 7, 8, 17, and 20	6q, 8p, 9, and 10	Bastian et al. Cancer Res 1998;58:2170-5.
Lymphoma	1q, 3, 6p, 7, 11, 12, 18, and X	1p, 8p, and X	Monni et al. Blood 1996;87:5269-78

300

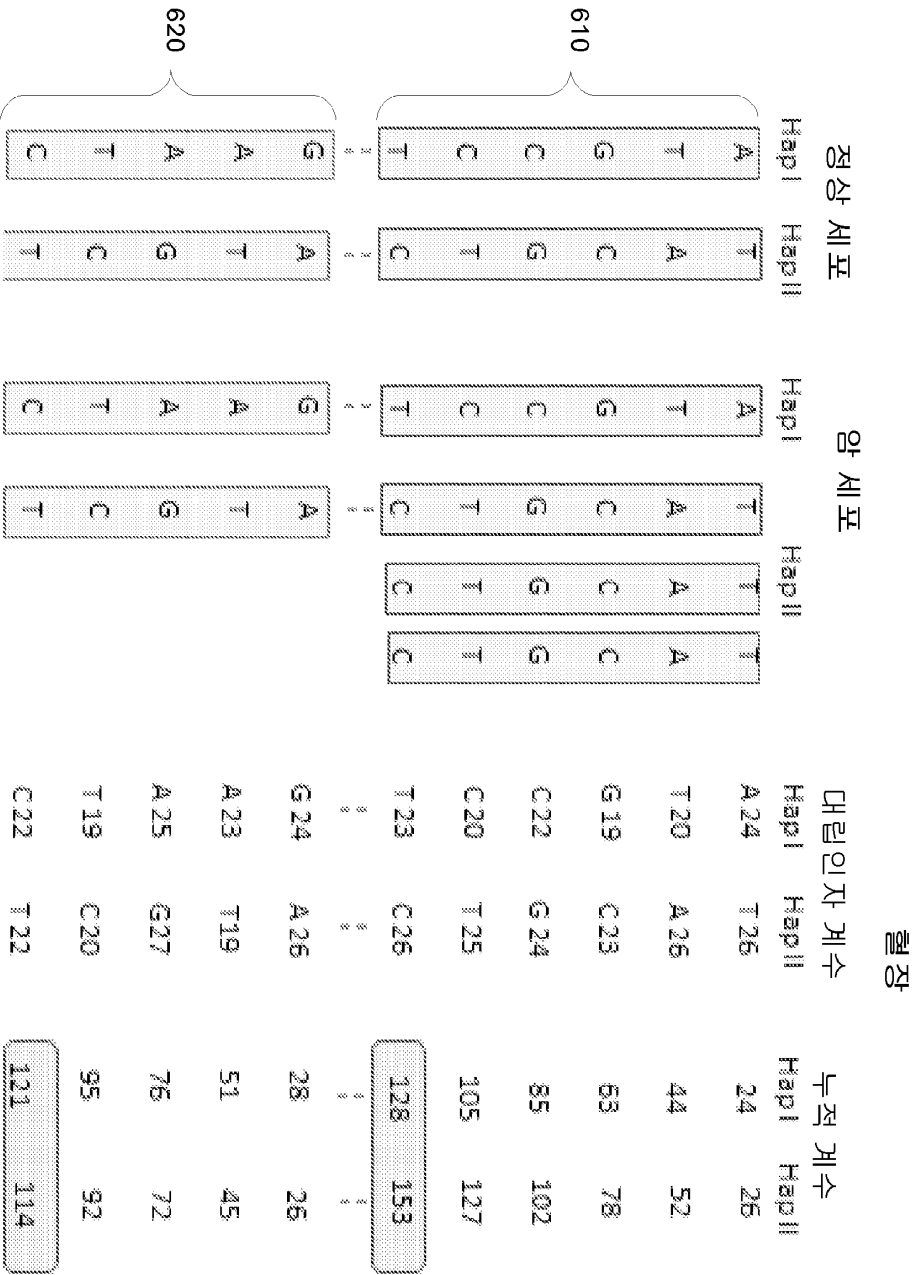
도면4



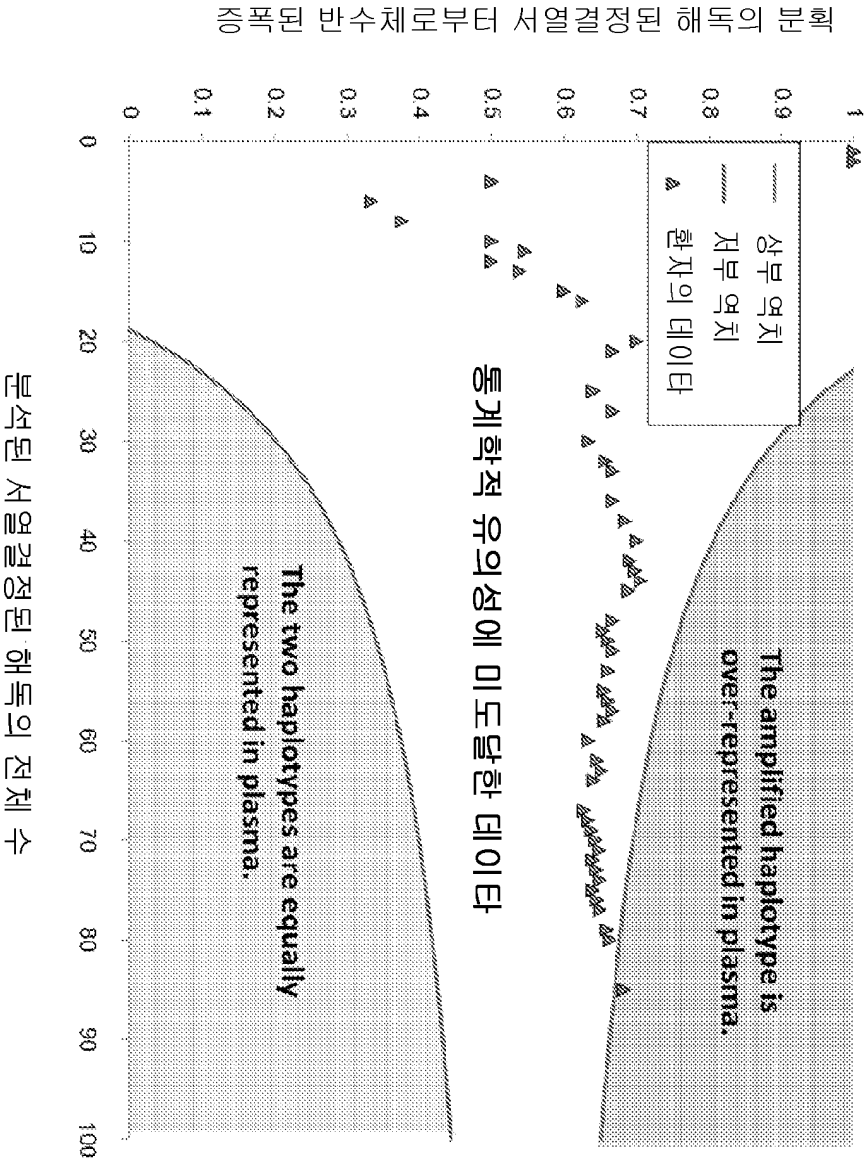
도면5



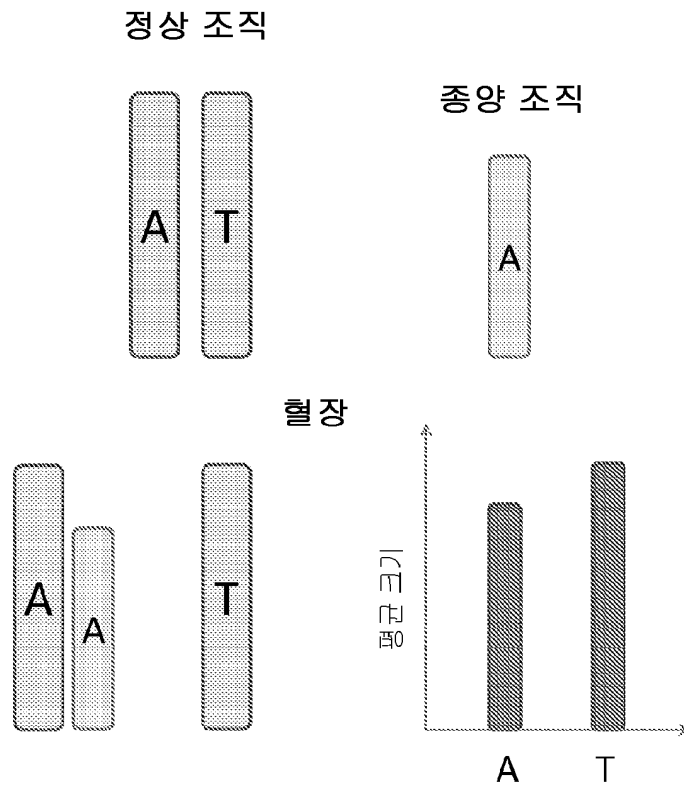
도면6



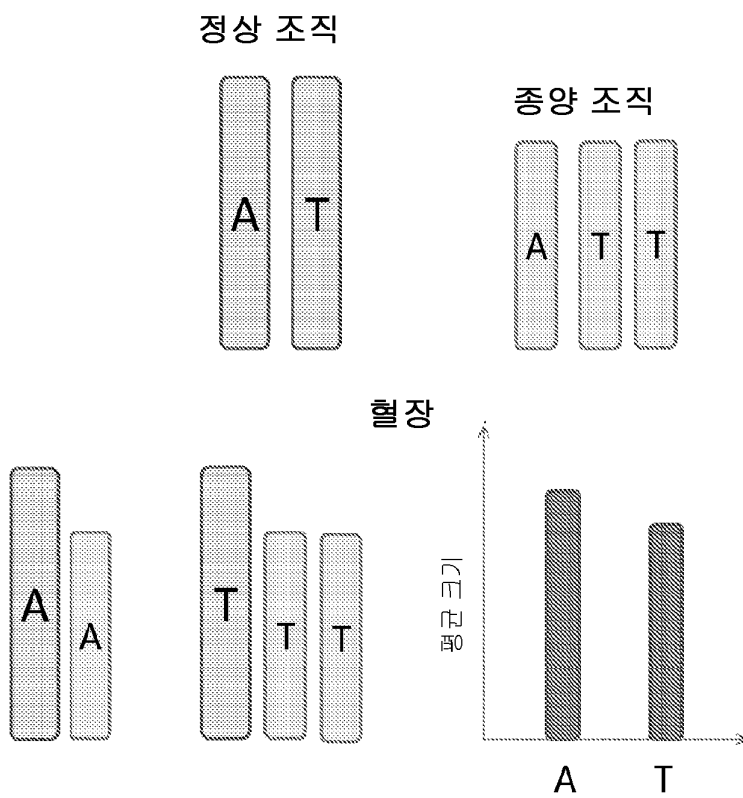
도면7



도면8

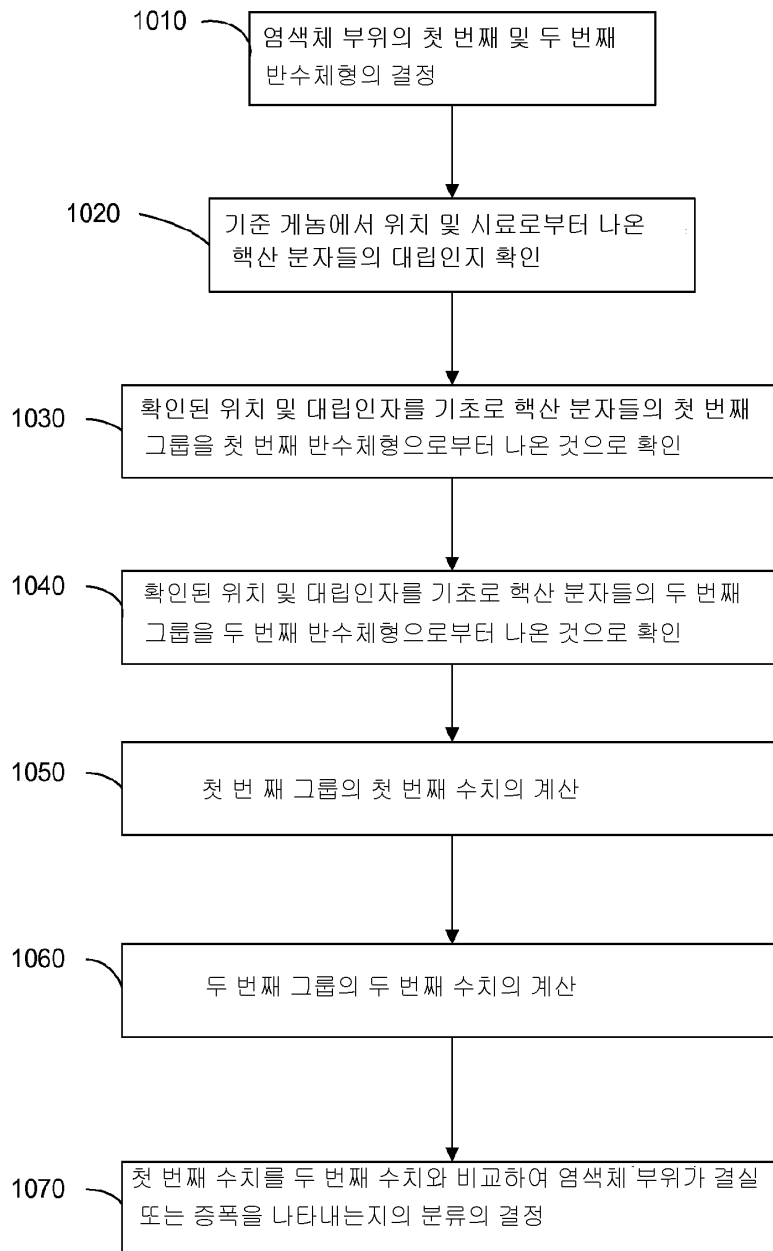


도면9

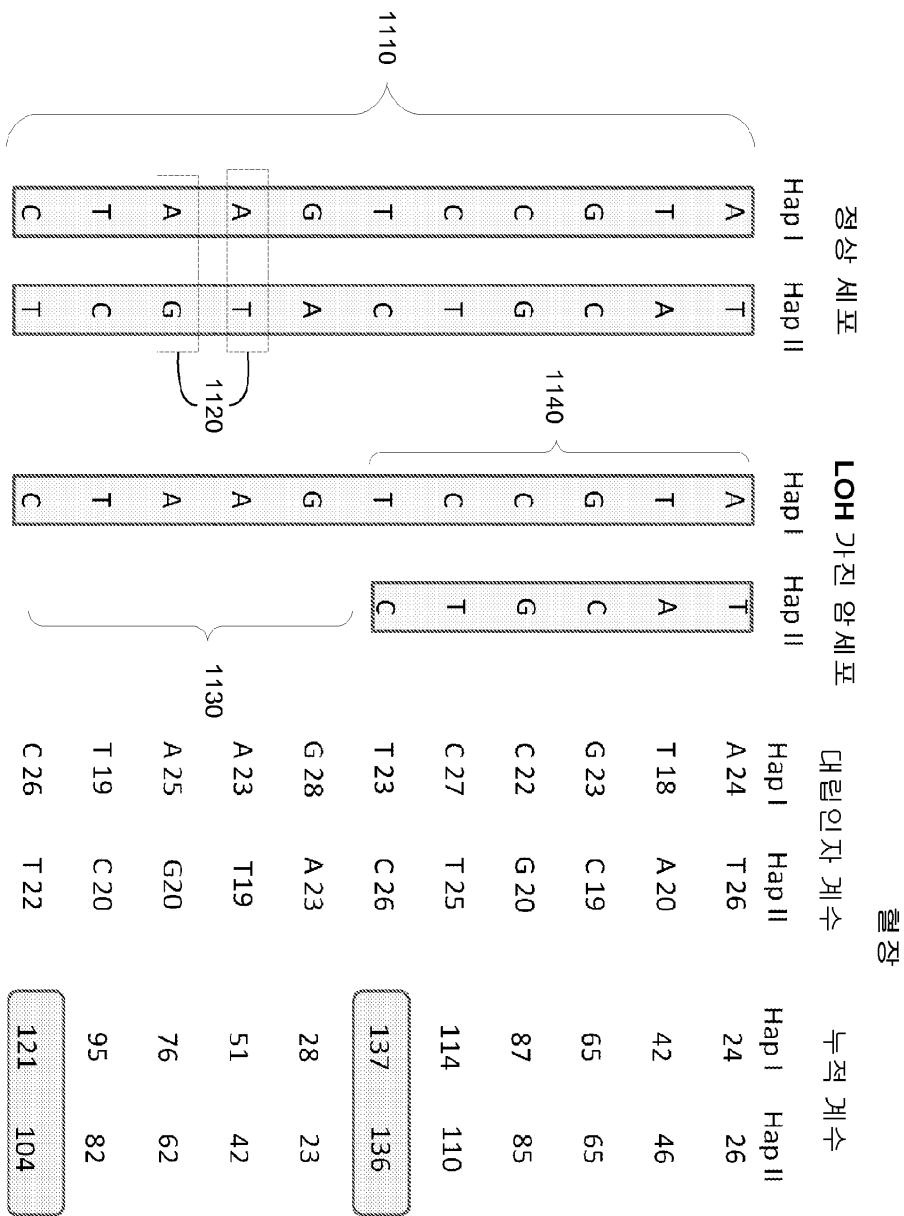


도면10

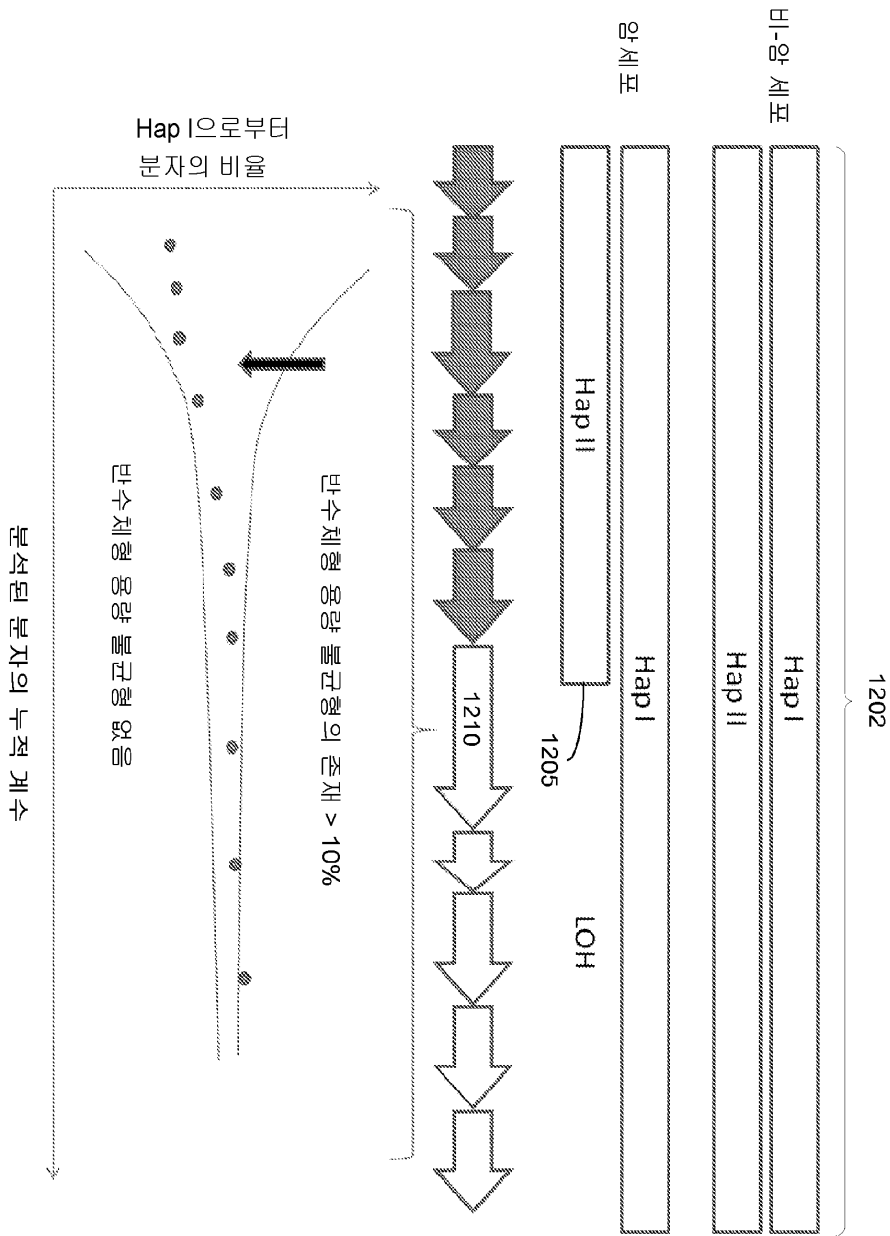
1000



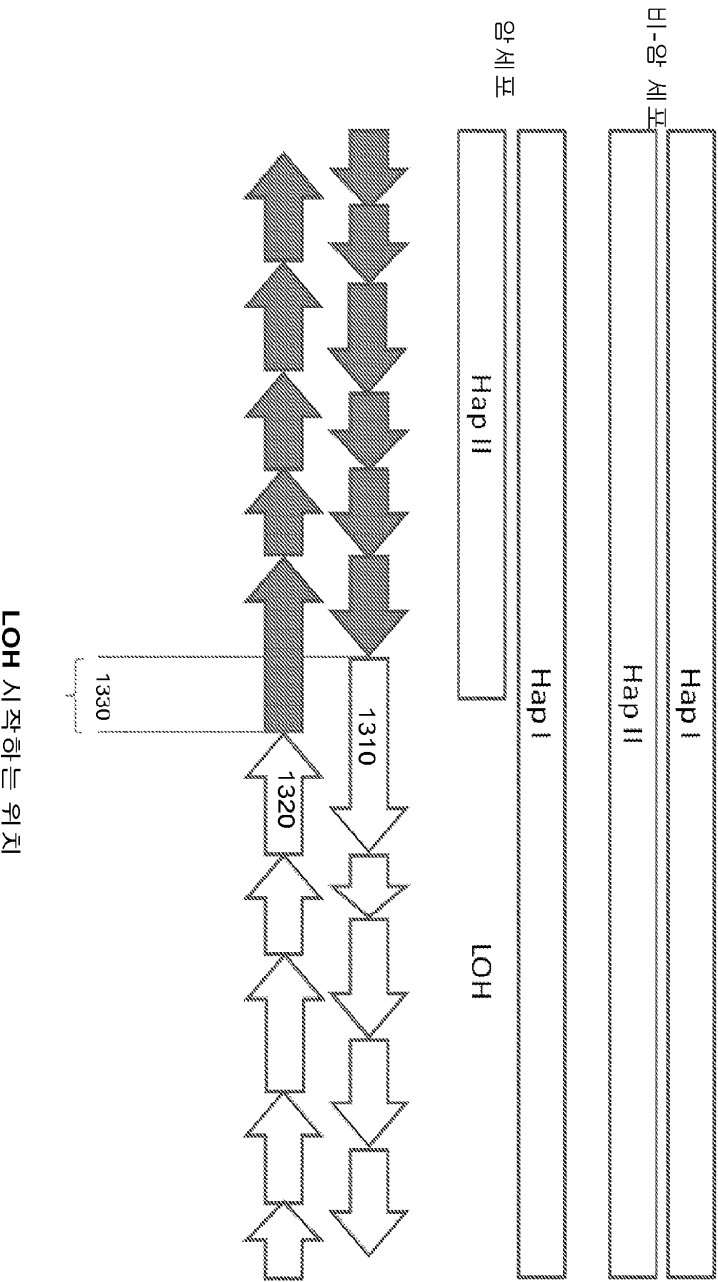
도면11



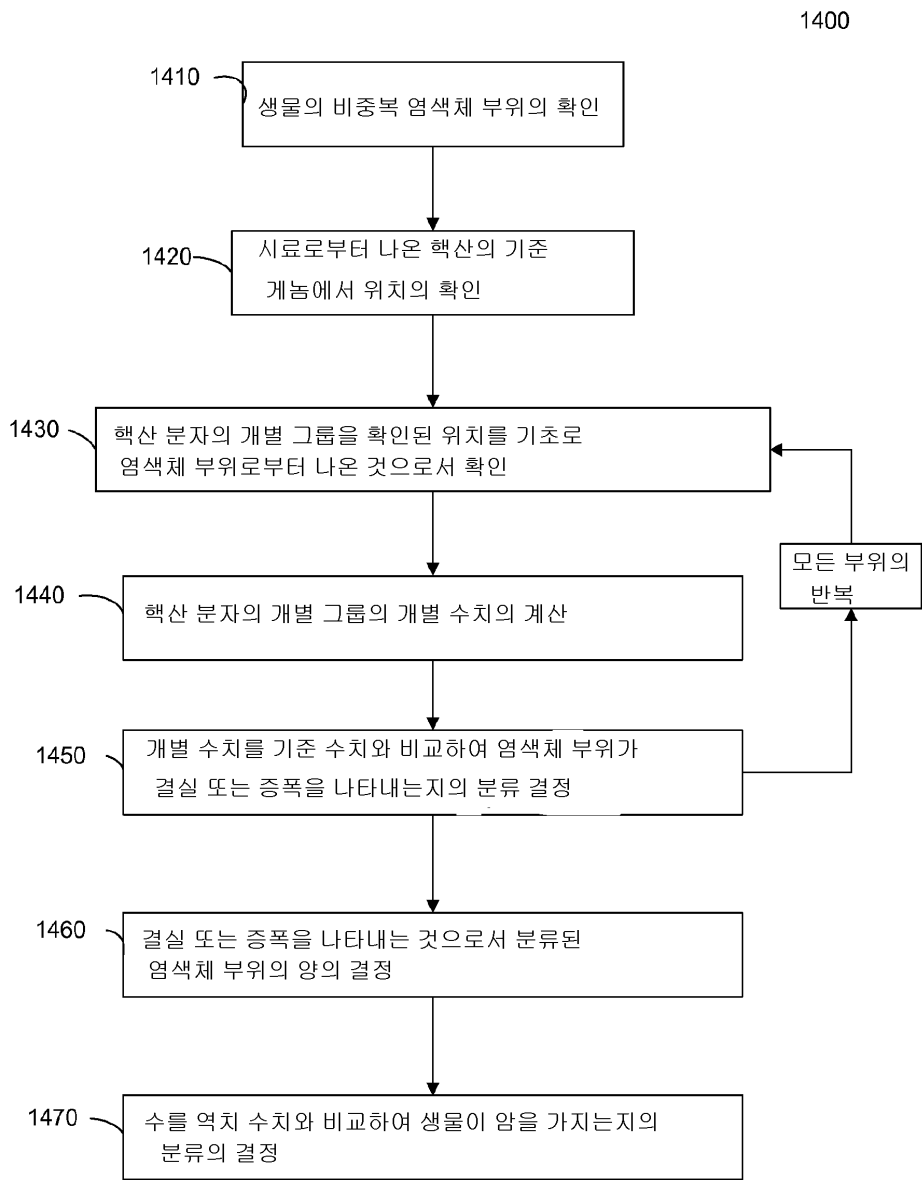
도면12



도면13



도면14

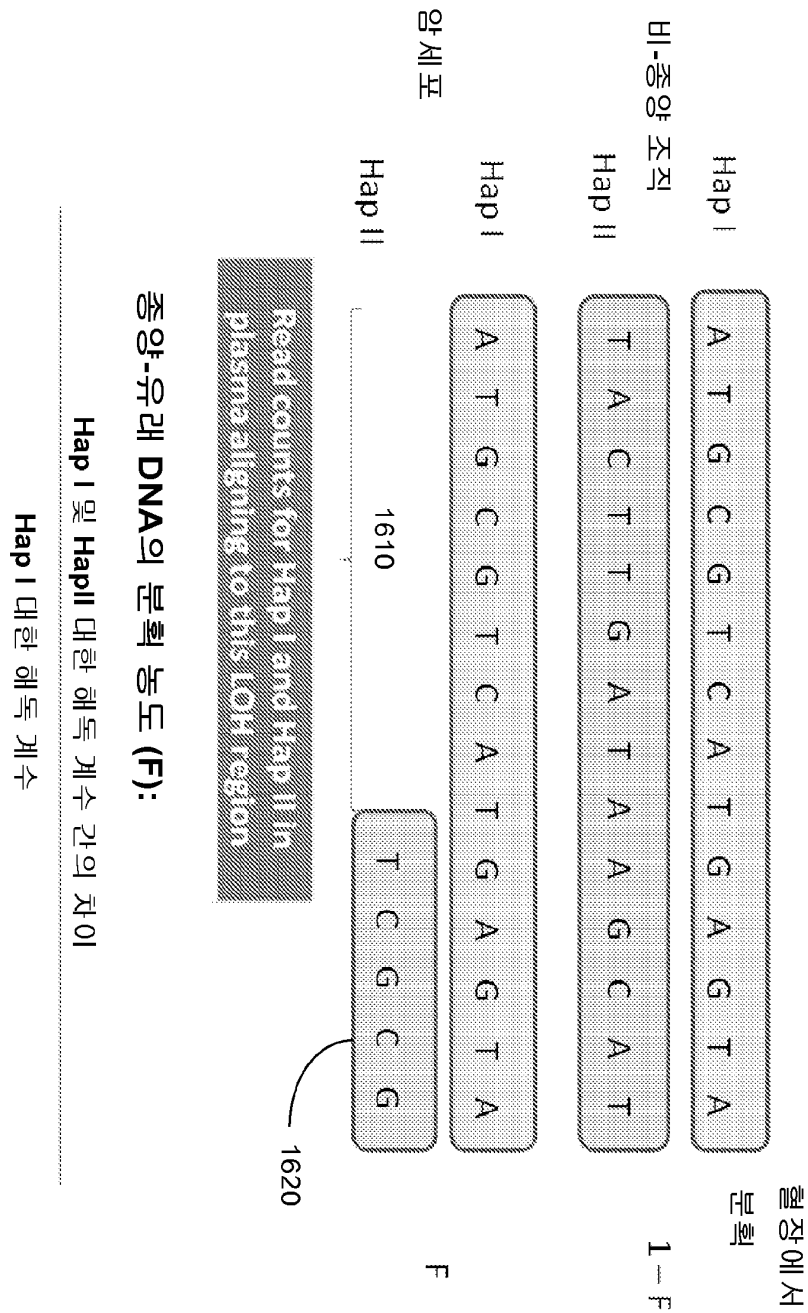


도면15

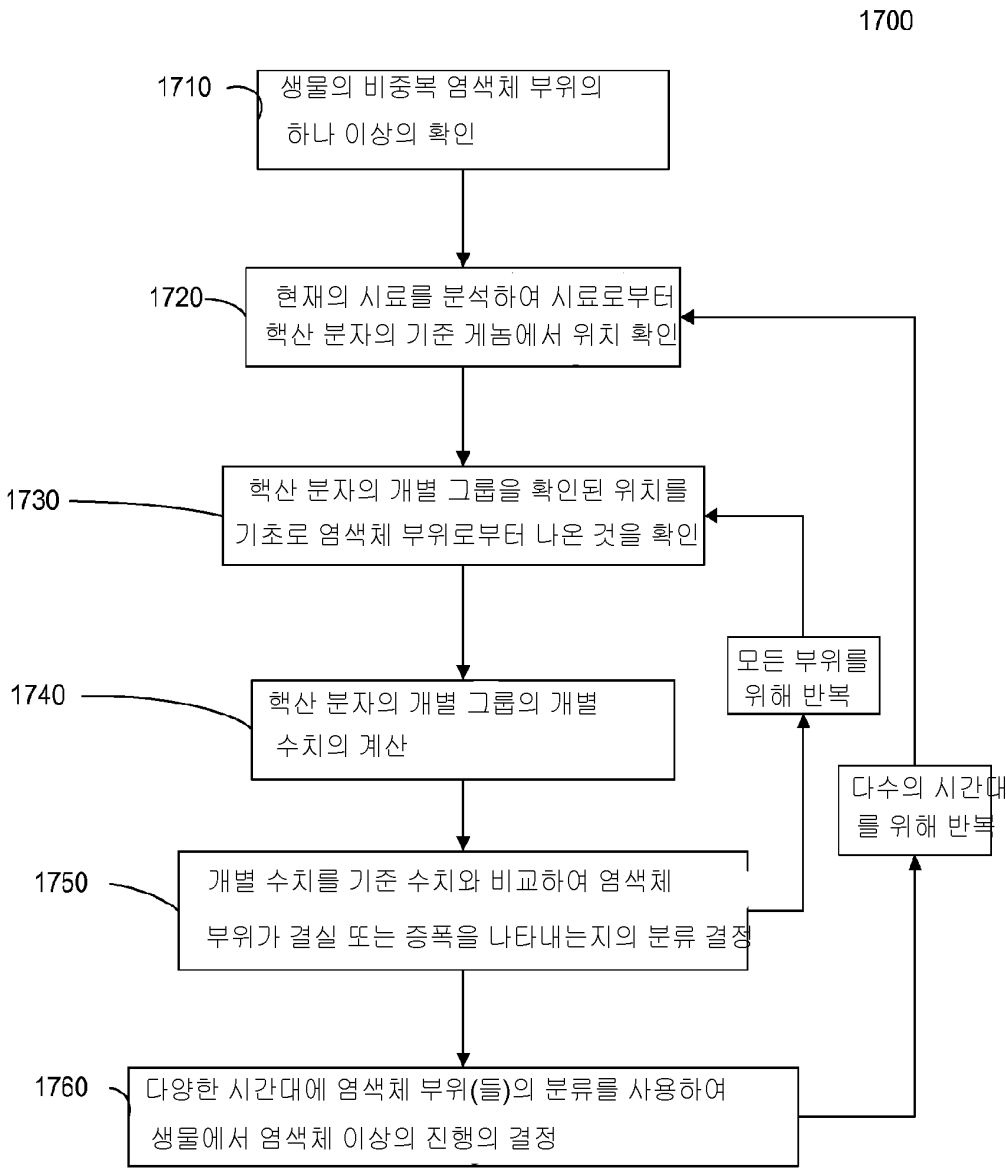
1500

암-유래 DNA의 분획 농도 1510	분절 당 요구되는 분자의 추정된 수 1520	각 분절의 크기 (kb) 1530	전체 게놈에 대해 분석된 분자의 전체 수 (백만) 1540
50%	950	100	28.5
		1,000	2.85
		5,000	0.570
		10,000	0.285
25%	3,800	100	114
		1,000	11.4
		5,000	2.28
		10,000	1.14
12.5%	15,000	100	450
		1,000	45
		5,000	9
		10,000	4.5
6.3%	60,000	100	1,800
		1,000	180
		5,000	36
		10,000	18
3.2%	240,000	100	7,200
		1,000	720
		5,000	144
		10,000	72

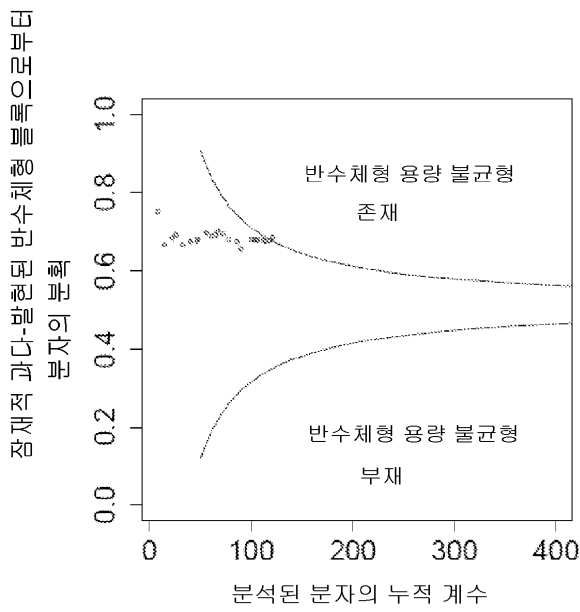
도면16



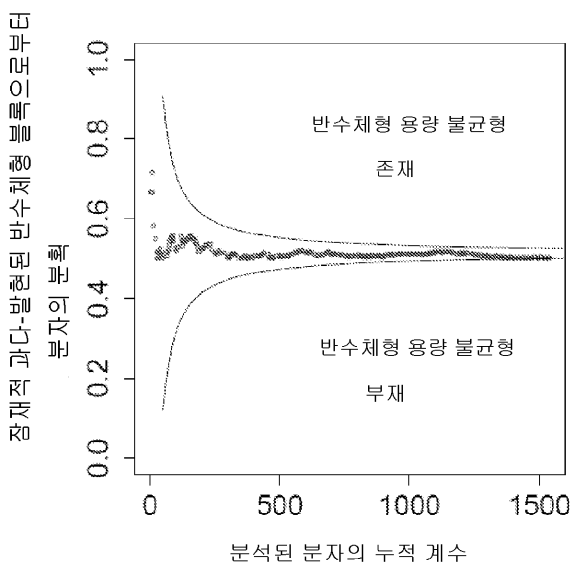
도면17



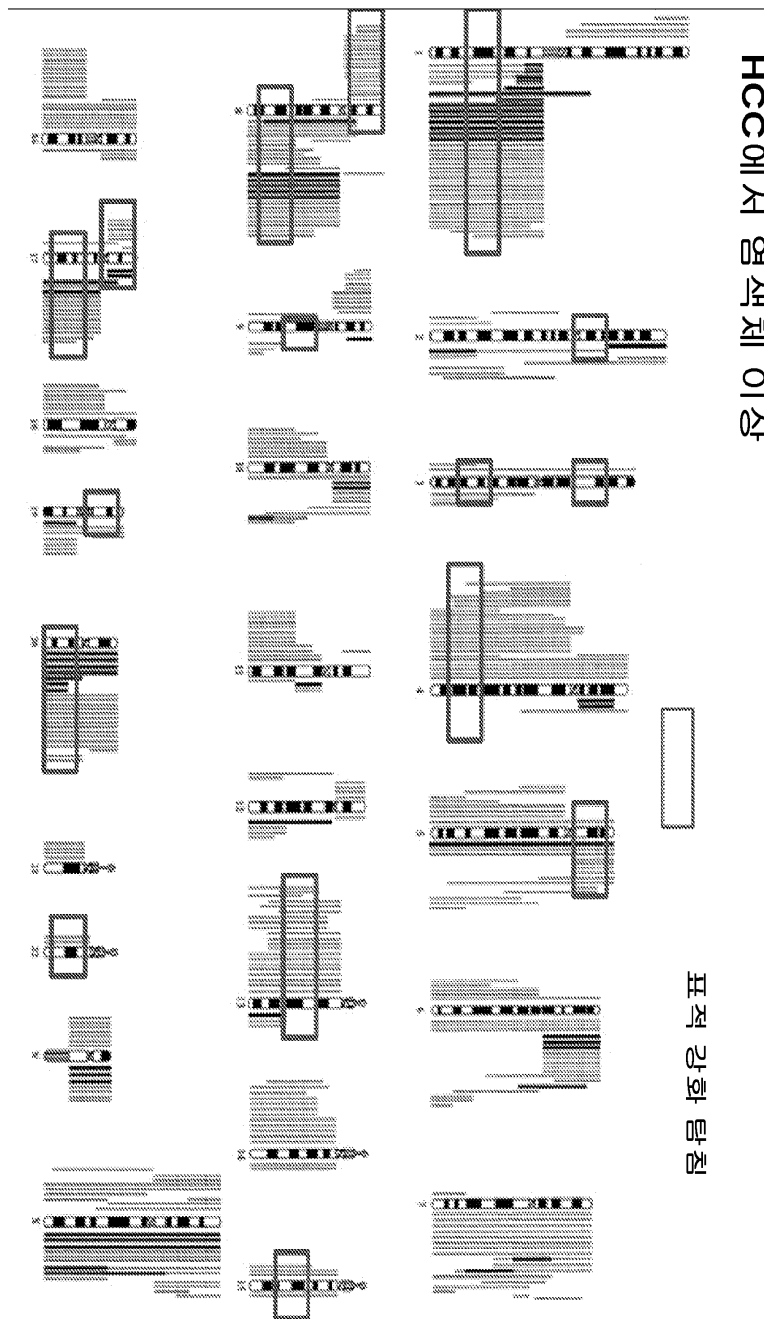
도면18a



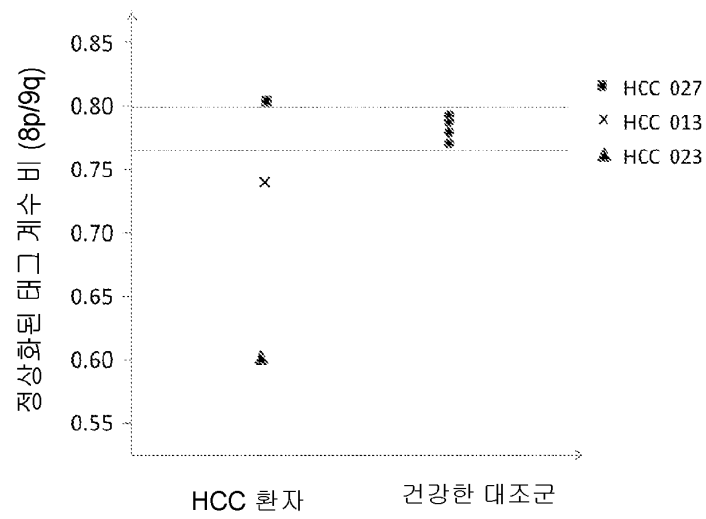
도면18b



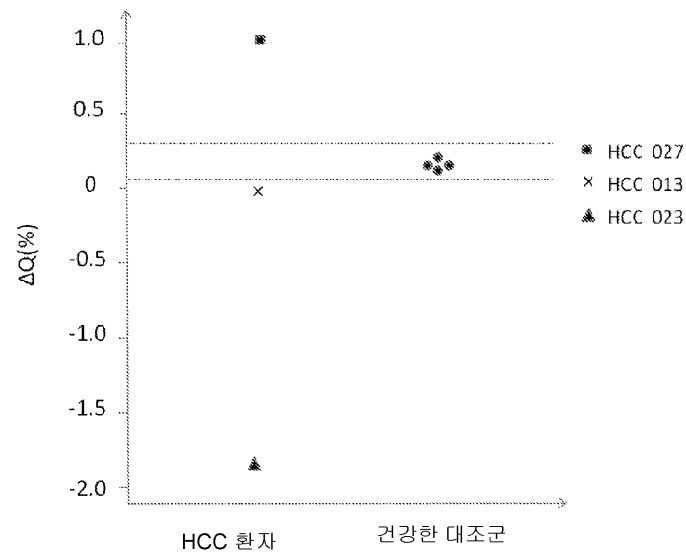
도면19



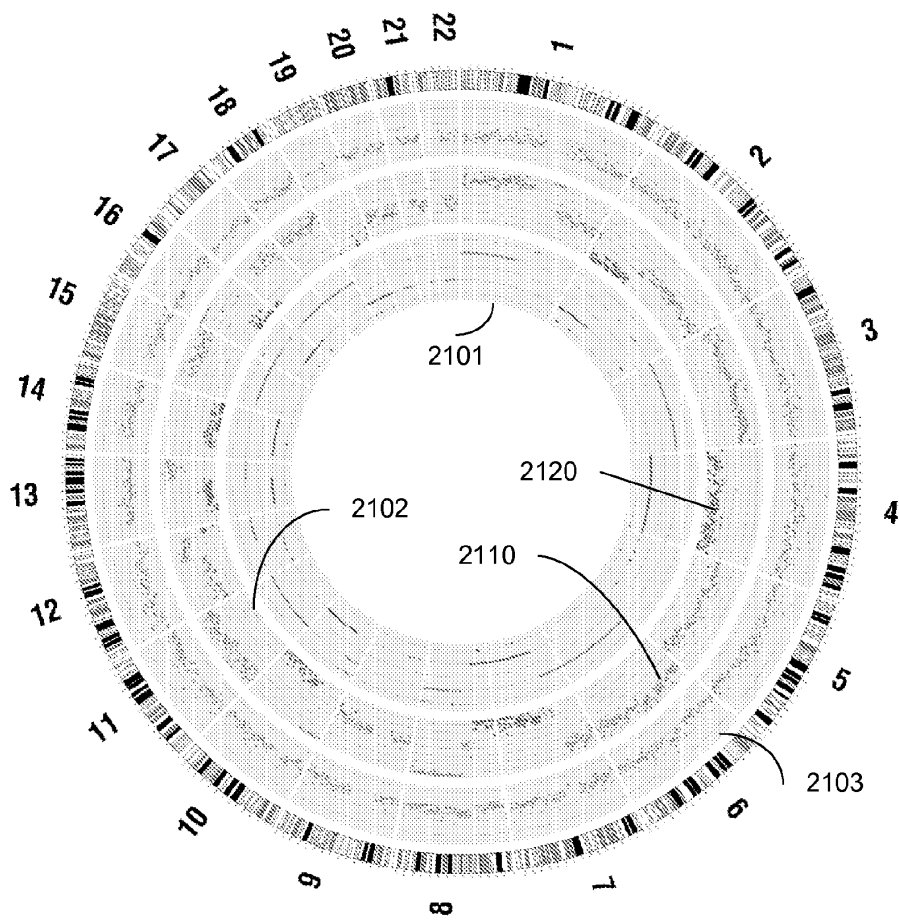
도면20a



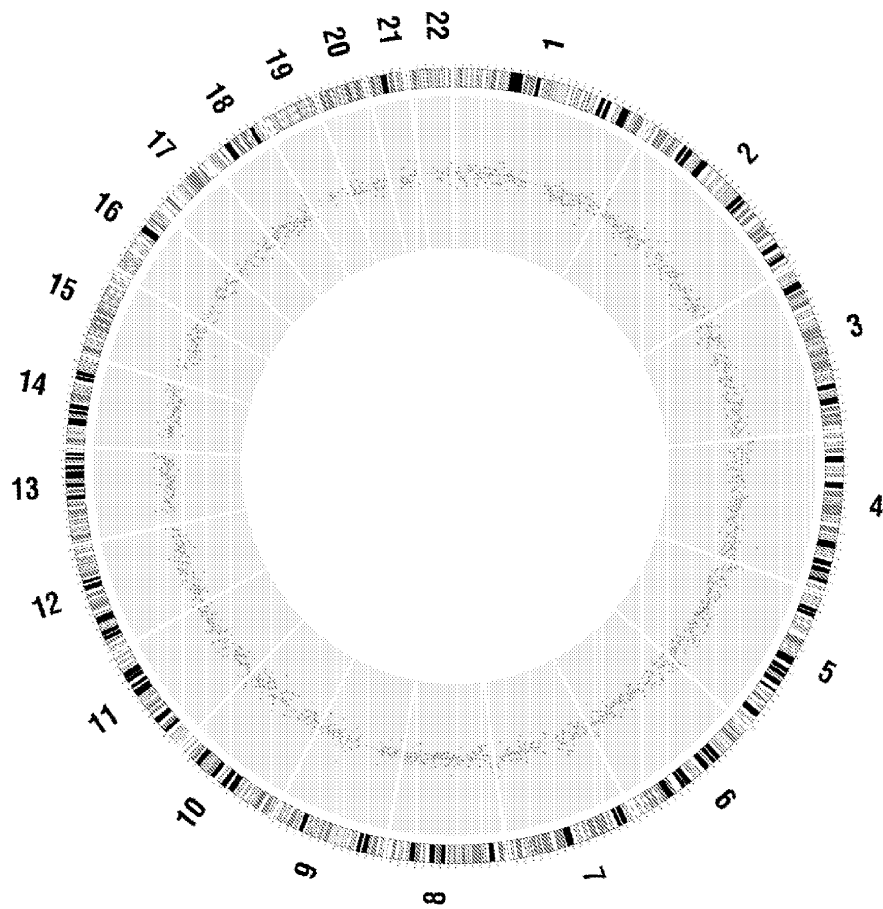
도면20b



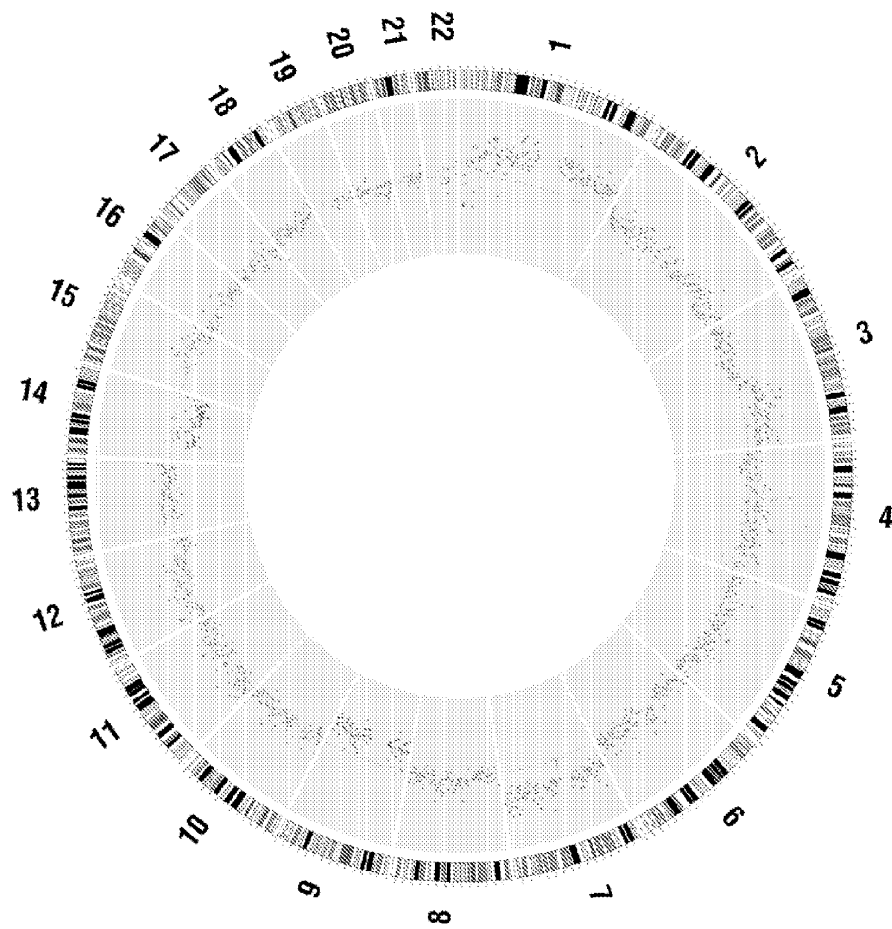
도면21



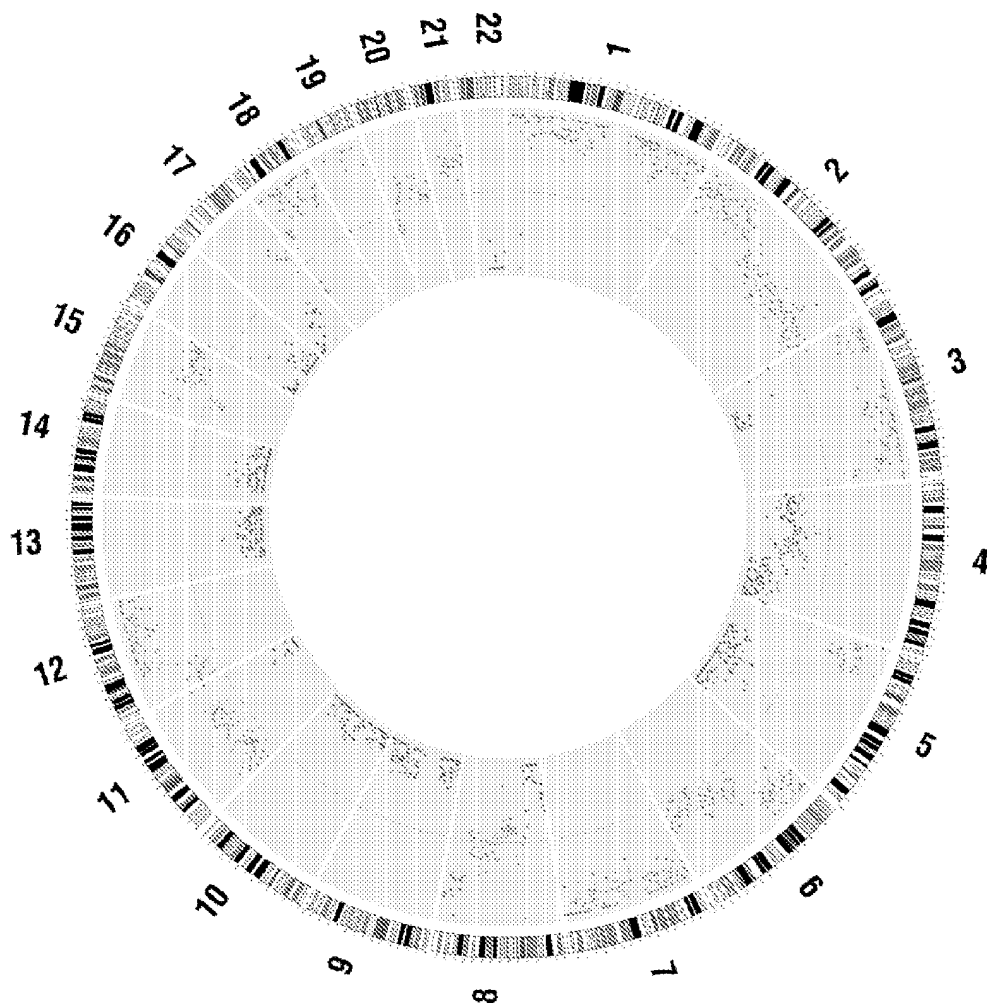
도면22



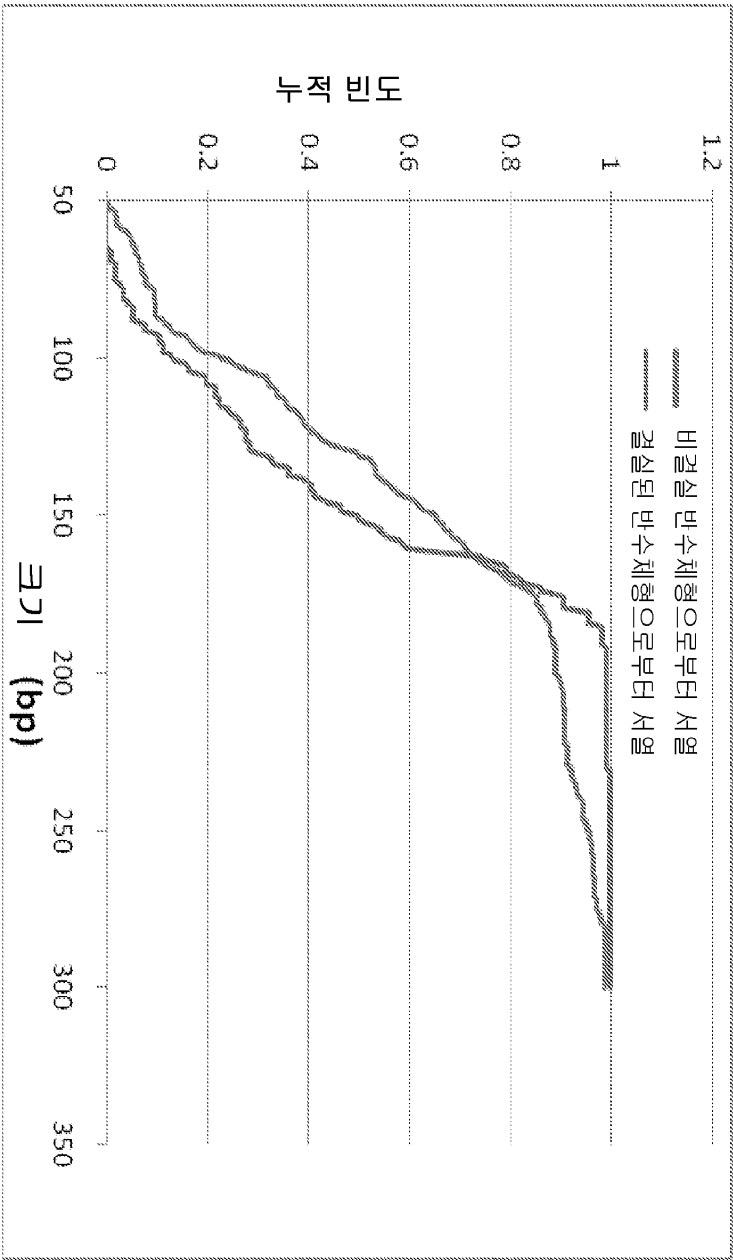
도면23



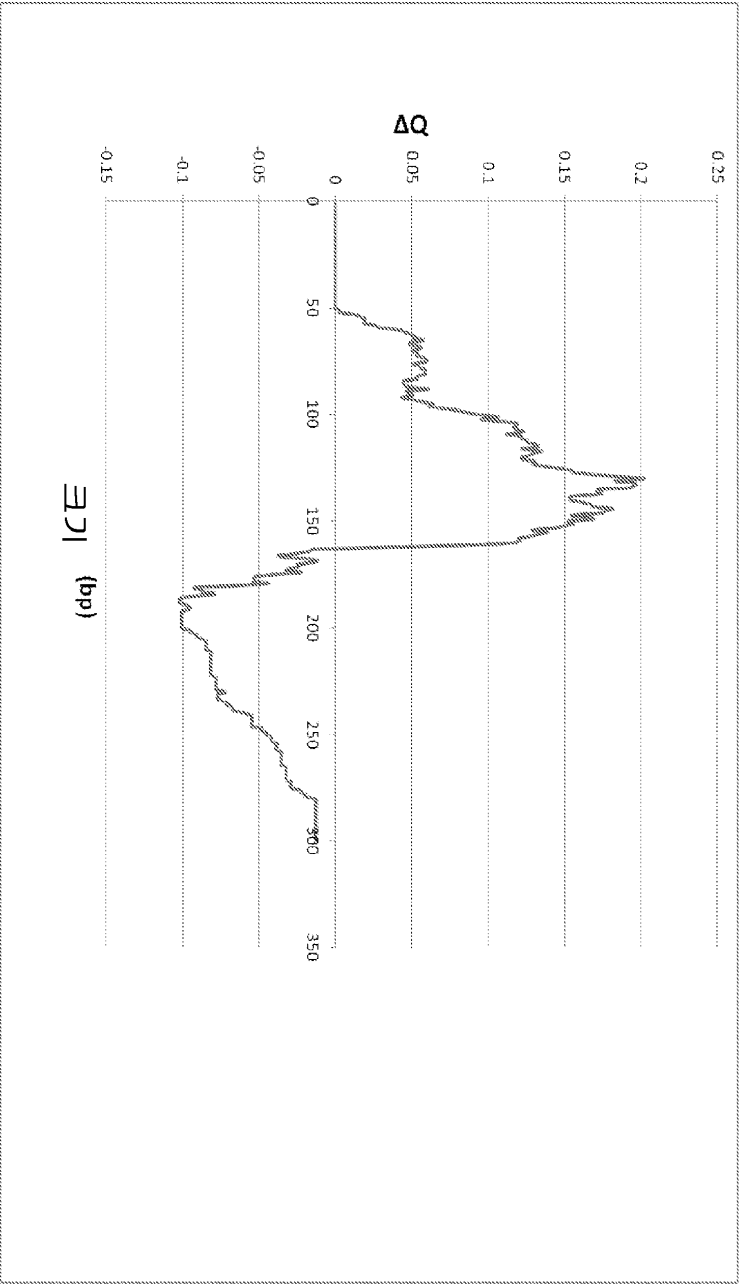
도면24



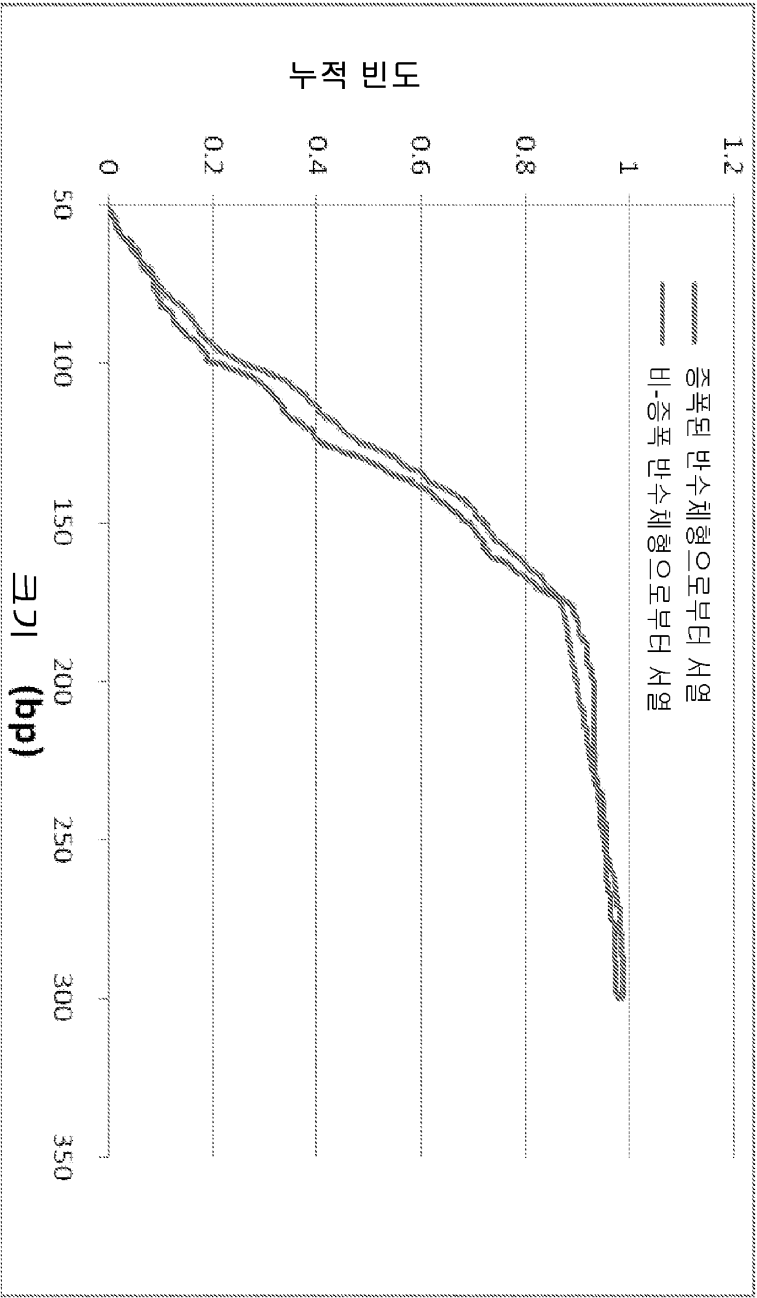
도면25



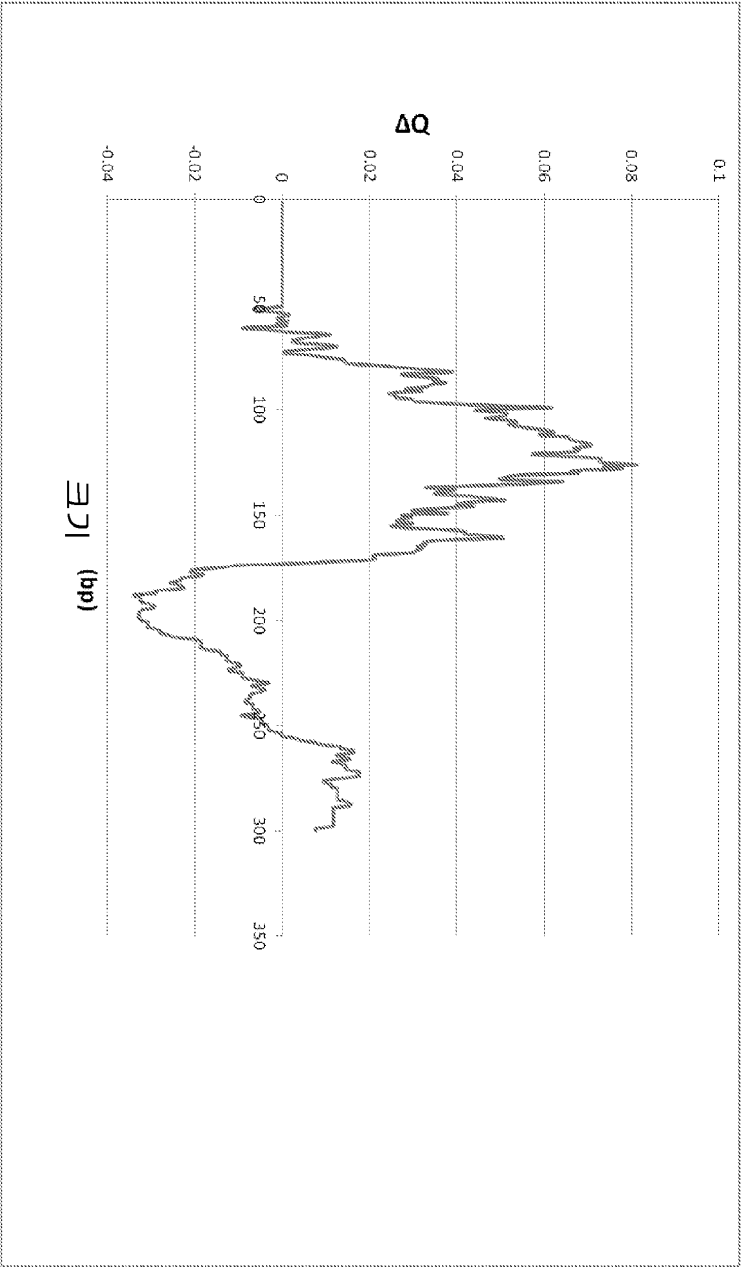
도면26



도면27



도면28



도면29

