



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0011140  
(43) 공개일자 2022년01월27일

(51) 국제특허분류(Int. Cl.)  
G16B 20/20 (2019.01) C12Q 1/6886 (2018.01)  
G16B 40/20 (2019.01)  
(52) CPC특허분류  
G16B 20/20 (2019.02)  
C12Q 1/6886 (2022.01)  
(21) 출원번호 10-2021-7040903  
(22) 출원일자(국제) 2020년05월20일  
심사청구일자 없음  
(85) 번역문제출일자 2021년12월14일  
(86) 국제출원번호 PCT/US2020/033821  
(87) 국제공개번호 WO 2020/236941  
국제공개일자 2020년11월26일  
(30) 우선권주장  
62/850,474 2019년05월20일 미국(US)

(71) 출원인  
파운데이션 메디신 인코포레이티드  
미국 메사추세츠 02141 케임브리지 세컨드 스트리트 150  
(72) 발명자  
팬들러, 베르나드  
미국 02141 메사추세츠주 케임브리지 세컨드 스트리트 150 파운데이션 메디신 인코포레이티드 내  
휴즈, 제이슨 디.  
미국 02141 메사추세츠주 케임브리지 세컨드 스트리트 150 파운데이션 메디신 인코포레이티드 내  
로엘스, 스티븐  
미국 02141 메사추세츠주 케임브리지 세컨드 스트리트 150 파운데이션 메디신 인코포레이티드 내  
(74) 대리인  
양영준, 이상남

전체 청구항 수 : 총 59 항

(54) 발명의 명칭 **종양 분획 평가를 위한 시스템 및 방법**

(57) 요약

적어도 부분적으로 대상체로부터의 샘플의 종양 분획을 결정하는 방법이 본원에 개시된다. 상기 방법은, 예를 들어 상기 샘플에서 하위 유전체 간격과 연관된 목표 변수에 대한 값을 획득하는 단계; 상기 목표 변수로부터 확실성 메트릭을 결정하는 단계; 저장된 확실성 메트릭 및 저장된 종양 분획 사이의 결정된 관계에 액세스하는 단계; 및 상기 확실성 메트릭 및 상기 결정된 관계를 참조하여 상기 샘플의 종양 분획을 결정하는 단계를 포함할 수 있다.

(52) CPC특허분류

**G16B 40/20** (2019.02)

C12Q 2600/112 (2013.01)

C12Q 2600/156 (2013.01)

---

## 명세서

### 청구범위

#### 청구항 1

대상체로부터의 샘플의 종양 분획을 결정하는 방법으로서, 상기 방법은:

복수의 값을 수득하는 단계로서, 각 값은 상기 샘플에서 하위 유전체 간격 내 대응하는 유전자좌의 대립유전자 분획을 나타내는 단계;

상기 복수의 값의 분산을 나타내는 확실성 메트릭을 결정하는 단계;

하나 이상의 저장된 확실성 메트릭 및 하나 이상의 저장된 종양 분획 사이의 사전 결정된 관계에 액세스하는 단계; 및

상기 확실성 메트릭 및 상기 사전 결정된 관계로부터 상기 샘플의 종양 분획을 결정하는 단계를 포함하는, 방법.

#### 청구항 2

제1항에 있어서,

상기 복수의 값 내의 각 값은 대립유전자 분획인, 방법.

#### 청구항 3

제1항에 있어서,

상기 복수의 값 내의 각 값은 대응하는 유전자좌의 모계 대립유전자 또는 부계 대립유전자의 존재량에 대비하여 상기 모계 대립유전자 및 상기 부계 대립유전자 간의 존재량 차이의 비를 포함하는, 방법.

#### 청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서,

상기 확실성 메트릭은 기대값으로부터 상기 복수의 값의 각각의 편차를 나타내는, 방법.

#### 청구항 5

제4항에 있어서,

상기 기대값은 유전자좌 특이적 기대값인, 방법.

#### 청구항 6

제4항 또는 제5항에 있어서,

상기 확실성 메트릭은 상기 기대값으로부터 평균 제공된 편차인, 방법.

#### 청구항 7

제4항 내지 제6항 중 어느 한 항에 있어서,

상기 기대값은 비종양 샘플에 대한 예상되는 대립유전자 빈도인, 방법.

#### 청구항 8

제4항 내지 제7항 중 어느 한 항에 있어서,

상기 복수의 값 내의 각 값은 대립유전자 분획이고, 상기 기대값은 0.5인 방법.

#### 청구항 9

제4항 내지 제6항 중 어느 한 항에 있어서,

상기 복수의 값 내의 각 값은 대응하는 유전자좌에서 모계 대립유전자 또는 부계 대립유전자의 존재량에 대비하여 상기 모계 대립유전자 및 상기 부계 대립유전자 간의 존재량 차이의 비를 포함하고, 상기 기대값은 모계 대립유전자 또는 부계 대립유전자의 존재량에 대비하여 상기 모계 대립유전자 및 상기 부계 대립유전자 간의 존재량 차이의 기대비를 포함하며, 상기 기대값은 비종양 샘플에 대한 기대비인, 방법.

#### 청구항 10

제9항에 있어서,

상기 기대값은 0인, 방법.

#### 청구항 11

제9항에 있어서,

상기 기대값은 0인, 방법.

#### 청구항 12

제1항에 있어서,

상기 복수의 값에 대한 확률 분포 함수를 결정하는 단계로서, 상기 확실성 메트릭은 상기 확률 분포 함수를 이용하여 결정되는 단계를 추가로 포함하는, 방법.

#### 청구항 13

제12항에 있어서,

상기 확실성 메트릭은 상기 확률 분포 함수의 엔트로피인, 방법.

#### 청구항 14

제1항 내지 제13항 중 어느 한 항에 있어서,

상기 대응하는 유전자좌는 상이한 모계 대립유전자 및 부계 대립유전자를 갖는 하나 이상의 유전자좌를 포함하는, 방법.

#### 청구항 15

제1항 내지 제14항 중 어느 한 항에 있어서,

상기 대응하는 유전자좌는 상이한 모계 대립유전자 및 부계 대립유전자를 갖는 유전자좌로 구성되는, 방법.

#### 청구항 16

제1항 내지 제14항 중 어느 한 항에 있어서,

상기 대응하는 유전자좌는 동일한 모계 대립유전자 및 부계 대립유전자를 갖는 하나 이상의 유전자좌를 포함하는, 방법.

#### 청구항 17

대상체로부터의 샘플의 종양 분획을 결정하는 방법으로서, 상기 방법은:

복수의 값을 수득하는 단계로서, 각 값은 하위 유전체 간격 내 복수의 유전자좌에서 종양 샘플 내 유전자좌의 대립유전자 범위 및 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위 간의 차이를 나타내는 단계;

상기 복수의 값의 분산을 나타내는 확실성 메트릭을 결정하는 단계;

하나 이상의 저장된 확실성 메트릭 및 하나 이상의 저장된 종양 분획 사이의 사전 결정된 관계에 액세스하는 단계; 및

상기 확실성 메트릭 및 상기 사전 결정된 관계로부터 상기 샘플의 종양 분획을 결정하는 단계를 포함하는,

방법.

#### 청구항 18

제17항에 있어서,

상기 복수의 값 내의 각 값은 상기 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위에 대비한 상기 종양 샘플 내 유전자좌의 대립유전자 범위의 비를 포함하는, 방법.

#### 청구항 19

제17항에 있어서,

상기 복수의 값 내의 각 값은 상기 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위에 대비한 상기 종양 샘플 내 유전자좌의 대립유전자 범위의 로그비를 포함하는, 방법.

#### 청구항 20

제19항에 있어서,

상기 로그비는  $\log_2$  비인, 방법.

#### 청구항 21

제17항에 있어서,

상기 복수의 값 내의 각 값은 상기 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위에 대비하여 상기 종양 샘플 내 유전자좌 및 상기 비종양 샘플 내 동일한 유전자좌 간의 대립유전자 범위의 차이의 비를 포함하는, 방법.

#### 청구항 22

제17항 내지 제21항 중 어느 한 항에 있어서,

상기 확실성 메트릭은 상기 대응하는 유전자좌를 통틀어 기대값으로부터 상기 복수의 값 내 각 값의 편차를 나타내고, 상기 기대값은 상기 종양 샘플이 비종양 샘플인 경우 예상되는 값인, 방법.

#### 청구항 23

제22항에 있어서,

각 값은 상기 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위와 비교하여 상기 종양 샘플 내 유전자좌의 대립유전자 범위의 비를 포함하고, 상기 기대값은 1이고;

각 값은 상기 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위와 비교하여 상기 종양 샘플 내 유전자좌의 대립유전자 범위의 로그비를 포함하고, 상기 기대값은 0이거나; 또는

각 값은 상기 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위와 대비하여 상기 종양 샘플 내 유전자좌 및 상기 비종양 샘플 내 동일한 유전자좌 간의 대립유전자 범위의 차이의 비를 포함하고, 상기 기대값은 0인, 방법.

#### 청구항 24

제17항 내지 제23항 중 어느 한 항에 있어서,

상기 확실성 메트릭은 상기 기대값의 평균 제곱근 편차인, 방법.

#### 청구항 25

제17항에 있어서,

상기 복수의 값에 대한 확률 분포 함수를 결정하는 단계로서, 상기 확실성 메트릭은 상기 확률 분포 함수를 이용하여 결정되는 단계를 추가로 포함하는, 방법.

#### 청구항 26

제25항에 있어서,  
상기 확실성 메트릭은 상기 확률 분포 함수의 엔트로피인, 방법.

#### 청구항 27

제17항 내지 제26항 중 어느 한 항에 있어서,  
상기 유전자좌는 모계 대립유전자 및 부계 대립유전자의 대립유전자 범위를 포함하는, 방법.

#### 청구항 28

제17항 내지 제27항 중 어느 한 항에 있어서,  
상기 대립유전자 범위는 모계 대립유전자 및 부계 대립유전자의 대립유전자 범위로 구성되는, 방법.

#### 청구항 29

제1항 내지 제28항 중 어느 한 항에 있어서,  
상기 복수의 유전자좌는 단일 염기 다형성(SNP)과 관련된 적어도 하나의 뉴클레오티드를 포함하는, 방법.

#### 청구항 30

제29항에 있어서,  
상기 복수의 유전자좌는 각각 단일 염기 다형성(SNP)과 관련된 2개 이상의 뉴클레오티드를 포함하는, 방법.

#### 청구항 31

제29항 또는 제30항에 있어서,  
상기 SNP는 암과 관련된, 방법.

#### 청구항 32

제1항 내지 제31항 중 어느 한 항에 있어서,  
상기 복수의 유전자좌 중 적어도 일부는 복제수 변이(CNV)와 관련되는, 방법.

#### 청구항 33

제32항에 있어서,  
상기 CNV는 암과 관련되는, 방법.

#### 청구항 34

제1항 내지 제33항 중 어느 한 항에 있어서,  
각 유전자좌에서 대립유전자 존재량 또는 범위를 결정하기 위해 상기 샘플을 서열분석하는 단계를 추가로 포함하는, 방법.

#### 청구항 35

제1항 내지 제33항 중 어느 한 항에 있어서,  
각 유전자좌에서 대립유전자 존재량 또는 범위를 결정하기 위해 상기 샘플 상에 배열 혼성화(array hybridization)를 실시하는 단계를 추가로 포함하는, 방법.

#### 청구항 36

제1항 내지 제35항 중 어느 한 항에 있어서, 상기 방법은:

복수의 훈련 확실성 메트릭 및 연관된 훈련 종양 분획들 사이의 복수의 관계를 포함하는 훈련 데이터세트에 액세스하는 단계; 및

상기 훈련 확실성 메트릭 및 상기 훈련 종양 분획 사이의 소정의 관계를 결정하기 위해 상기 훈련 데이터세트에 기계 학습 공정을 적용하는 단계를 추가로 포함하는, 방법.

#### 청구항 37

제1항 내지 제36항 중 어느 한 항에 있어서,

상기 대상체 및 상기 결정된 종양 분획을 식별하는 정보를 포함하는 보고서를 생성하는 단계를 포함하는, 방법.

#### 청구항 38

제37항에 있어서,

상기 대상체 또는 의료인에게 상기 보고서를 제공하는 단계를 포함하는, 방법.

#### 청구항 39

제37항 또는 제38항에 있어서,

전자 건강 기록에 대한 상기 보고서의 형식을 만드는 단계를 포함하는, 방법.

#### 청구항 40

대상체에서 종양을 치료하는 방법으로서, 상기 방법은:

결정된 종양 분획에 반응하여 유효량의 종양 요법을 상기 대상체에게 투여하는 단계로서, 상기 종양 분획은 제1항 내지 제39항 중 어느 한 항의 방법에 따라 결정되는, 방법.

#### 청구항 41

제40항에 있어서,

상기 결정된 종양 분획에 기초하여 상기 환자에서 상기 종양의 존재를 결정하는 단계를 포함하는, 방법.

#### 청구항 42

제40항 또는 제41항에 있어서,

상기 종양 치료는 화학 요법, 방사선 요법 또는 수술을 포함하는, 방법.

#### 청구항 43

대상체에서 종양의 진행 또는 재발을 모니터링하는 방법으로서, 상기 방법은:

(a) 제1항 내지 제39항 중 어느 한 항의 방법에 따라 제1 시점에서 상기 대상체로부터 수득한 제1 샘플의 제1 종양 분획을 결정하는 단계;

(b) 제2 시점에서 상기 대상체로부터 수득한 제2 샘플의 제2 종양 분획을 결정하는 단계; 및

(c) 상기 제1 종양 분획을 상기 제2 종양 분획과 비교하여 상기 종양 진행을 모니터링하는 단계를 포함하는, 방법.

#### 청구항 44

제43항에 있어서,

상기 제2 종양 분획을 결정하는 단계는:

제2 복수의 값을 수득하는 단계로서, 각 값은 상기 제2 종양 샘플의 하위 유전체 간격 내 대응하는 유전자좌에서 대립유전자 분획을 나타내며, 상기 제2 샘플의 하위 유전체 간격은 상기 제1 샘플의 하위 유전체 간격과 동일하거나 상이한 단계;

상기 제2 복수의 값의 분산을 나타내는 제2 확실성 메트릭을 결정하는 단계;

하나 이상의 저장된 확실성 메트릭 및 하나 이상의 저장된 종양 분획 사이의 소정의 관계에 액세스하는 단계;  
및

상기 제2 확실성 메트릭 및 상기 소정의 관계로부터 상기 제2 샘플의 제2 종양 분획을 결정하는 단계를 포함하는, 방법.

#### 청구항 45

제43항에 있어서,

상기 제2 종양 분획을 결정하는 단계는:

제2 복수의 값을 수득하는 단계로서, 각 값은 상기 샘플의 하위 유전체 간격 내 복수의 유전자좌에서 상기 제2 종양 샘플 내 유전자좌의 대립유전자 범위 및 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위 간의 차이를 나타내며, 상기 제2 종양 분획을 결정하는데 이용되는 하위 유전체 간격은 상기 제1 종양 분획을 결정하는데 사용되는 하위 유전체 간격과 동일하거나 상이한 단계;

상기 제2 복수의 값의 분산을 나타내는 제1 확실성 메트릭을 결정하는 단계;

하나 이상의 저장된 확실성 메트릭 및 하나 이상의 저장된 종양 분획 사이의 소정의 관계에 액세스하는 단계;  
및

상기 제2 확실성 메트릭 및 상기 소정의 관계로부터 상기 제2 종양 샘플의 제2 종양 분획을 결정하는 단계를 포함하는, 방법.

#### 청구항 46

제43항 내지 제 45항 중 어느 한 항에 있어서,

상기 종양 진행에 반응하여 종양 요법을 조절하는 단계를 포함하는, 방법.

#### 청구항 47

제46항에 있어서,

상기 종양 진행에 반응하여 상기 종양 요법의 투여량을 조절하거나 상이한 종양 요법을 선택하는 단계를 포함하는, 방법.

#### 청구항 48

제46항 또는 제47항에 있어서,

상기 조절된 종양 요법을 상기 대상체에게 투여하는 단계를 포함하는, 방법.

#### 청구항 49

제43항 내지 제48항 중 어느 한 항에 있어서,

상기 제1 시점은 상기 대상체에게 종양 요법이 투여되기 전이고, 상기 제2 시점은 상기 대상체에게 상기 종양 요법이 투여된 이후인, 방법.

#### 청구항 50

제1항 내지 제49항 중 어느 한 항에 있어서,

상기 대상체는 암을 가지거나, 암을 가질 위험이 있거나, 암을 가질 것으로 의심되는, 방법.

#### 청구항 51

제50항에 있어서,

상기 암은 고형 종양인, 방법.

#### 청구항 52

제50항에 있어서,  
상기 암은 혈액암인, 방법.

#### 청구항 53

제1항 내지 제52항 중 어느 한 항에 있어서,  
상기 샘플은 액체 샘플인, 방법.

#### 청구항 54

제 1항 내지 제52항 중 어느 한 항에 있어서,  
상기 샘플은 고형 샘플인, 방법.

#### 청구항 55

제1항 내지 제53항 중 어느 한 항에 있어서,  
상기 샘플은 무세포 DNA(cfDNA) 또는 순환 종양 DNA(ctDNA)를 포함하는, 방법.

#### 청구항 56

제1항 내지 제55항 중 어느 한 항에 있어서,  
상기 하나 이상의 저장된 확실성 메트릭은 복수의 저장된 확실성 메트릭을 포함하고, 상기 하나 이상의 저장된 종양 분획은 복수의 저장된 종양 분획을 포함하는, 방법.

#### 청구항 57

컴퓨터 시스템으로서:

프로세서; 및

상기 프로세서에 통신 가능하게 결합되고, 하기를 저장하도록 구성된 메모리로서:

하나 이상의 저장된 확실성 메트릭 및 하나 이상의 연관된 저장된 종양 분획 사이의 소정의 관계를 저장하도록 구성된 메모리; 및

상기 프로세서에 의해 실행될 때 상기 프로세서가 하기의 단계를 실시하도록 하는 명령어를 포함하는 컴퓨터 시스템으로서, 상기 명령어들은:

(a)(i) 복수의 값을 수득하는 단계로서, 각 값은 상기 샘플의 하위 유전체 간격 내 대응하는 유전자좌에서 대립 유전자 분획을 나타내는 단계, 또는 (ii) 복수의 값을 수득하는 단계로서, 각 값은 하위 유전체 간격 내 복수의 유전자좌에서 종양 샘플 내 유전자좌의 대립유전자 범위 및 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위의 차이를 나타내는 단계;

(b) 상기 복수의 값의 분산을 나타내는 확실성 메트릭을 결정하는 단계;

(c) 상기 저장된 소정의 관계에 액세스하는 단계; 및

(d) 상기 확실성 메트릭 및 상기 소정의 관계로부터 상기 샘플의 종양 분획을 결정하는 단계를 실시하도록 하는, 컴퓨터 시스템.

#### 청구항 58

제58항에 있어서,

상기 메모리는 상기 프로세서에 의해 실행될 때 상기 프로세서가 하기의 단계를 실시하도록 하는 명령어를 추가로 포함하는 컴퓨터 시스템으로서, 상기 명령어들은:

복수의 훈련 확실성 메트릭 및 연관된 훈련 종양 분획 사이의 복수의 관계를 포함하는 훈련 데이터세트에 액세스

스하는 단계; 및

상기 훈련 확실성 메트릭 및 상기 훈련 종양 분획 사이의 소정의 관계를 결정하기 위해 상기 훈련 데이터세트에 기계 학습 공정을 적용하는 단계를 실시하는, 컴퓨터 시스템.

## 청구항 59

제57항 또는 제58항에 있어서,

상기 명령어들은, 상기 프로세서에 의해 실행될 때, 상기 프로세서로 하여금 제1항 내지 제39항 중 어느 한 항의 방법을 실시하도록 하는, 컴퓨터 시스템.

## 발명의 설명

### 기술 분야

[0001] 관련 출원의 상호 참조

[0002] 본 출원은 2019년 5월 20일에 제출한 미합중국 가출원 특허 제62/850,474호에 기초한 우선권의 이익을 주장하며, 그 전체 내용은 본원에 원용된다.

### 배경 기술

[0003] 암세포는 암의 발병 및 진행 중에 돌연변이를 축적한다. 상기 돌연변이는 DNA 복구, 복제 또는 변형의 본질적인 기능부전, 외부 돌연변이원에 대한 노출의 결과일 수 있다. 특정 돌연변이는 암세포에 성장 우위를 부여하고 암이 발생하는 조직의 미세 환경에서 긍정적으로 선택된다. 하지만, 유전체 연구를 일상적인 임상 실습으로서 실시하는 것은 여전히 비용이 많이 들고 시간 집약적이며 기술적으로 어려운 일이다.

[0004] 따라서, 암과 관련된 샘플을 분석하기 위해 유전체 프로파일링을 포함한 신규의 접근법이 여전히 필요하다.

### 발명의 내용

[0005] 본원에 기재된 방법 및 시스템은 샘플, 생검 또는 대상체에서 종양 분획 수준의 평가를 가능하게 한다. 전형적으로, 종양 분획은 샘플 내의 참조, 예컨대 비종양 DNA 또는 모든 DNA에 대비한 샘플 내의 종양 유래 DNA의 수준 또는 비율로서 표현되거나 측정된다. 본원에 기재된 방법에서, 상기 샘플에 대한 확실성 메트릭(certainty metric)의 값을 획득하고 상기 값은, 예컨대 참조와 비교함으로써 참조의 측면에서 평가할 수 있다. 확실성 메트릭 자체는 하위 유전체 간격에서 대립유전자의 수준을 반영하는 목표 변수의 함수일 수 있다. 표적 변수는 대립유전자 분획의 함수인 변수뿐만 아니라 하위 유전체 간격의 리드의 함수인 변수를 포함할 수 있다.

[0006] 일부 구현예들에서, 상기 목표 변수에 대한 값은 상기 샘플로부터 획득, 예컨대 직접 획득한다. 전형적으로, 상기 샘플에 대한 확실성 메트릭이 비교되는 참조는 종양 분획의 수준과 연관되는, 예컨대 상관관계가 있는 확실성 메트릭 값(또는 복수의 확실성 메트릭 값)이다. 상기 참조에 포함된 확실성 메트릭 값은, 예컨대 상기 샘플 내의(예컨대, 이중 하위 유전체 간격의 대립유전자의 경우 0.5) 또는 상기 샘플 외부의(예컨대, 하나 이상의 다른 대상체로부터 작성된 표준 곡선) 개체 또는 관계에 기초할 수 있다.

[0007] 일부 예시들에서, 상기 목표 변수는 하나 이상의 하위 유전체 간격에서 대립유전자 분획일 수 있다. 목표 변수의 다른 예들은 하나 이상의 하위 유전체 간격에서 리드 수의 함수인 로그2비와 같은 변수를 포함한다. 전형적으로, 복수의 하위 유전체(예컨대, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300개 이상의 하위 유전체 간격)를 분석하여 종양 분획을 결정한다. 상기 복수의 하위 유전체 간격은 동일한 염색체 또는 상이한 염색체 상에 존재할 수 있다(예컨대, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22개 이상의 염색체 사이에 분포할 수 있다). 한 구현예에서, 상기 복수의 하위 유전체 간격의 적어도 일부는 (하위 유전체 간격에서 대립유전자의 관점에서) 이형접합성이다.

[0008] 한 구현예에서, 대상체로부터의 샘플에 대한 확실성 메트릭은 확실성 메트릭을 종양 분획과 관련시키는 곡선과 비교되고, 샘플 종양 분획에 대한 값이 획득된다.

[0009] 한 구현예에서, 상기 확실성 메트릭은 목표 변수, 예컨대 대립유전자 분획의 함수이다. 예로서, 상기 확실성 메트릭은 관찰된 대립유전자 분획이 참조, 예컨대 예상 대립유전자 분획 또는 로그2비로부터 벗어나는 정도와 관련될 수 있고, 종양 분획의 수준과 관련된 참조와 비교될 수 있다. 다른 예들에서, 상기 확실성 메트릭은 목표

변수의 상대적 확실성, 예컨대 본원에 기재된 설명된 엔트로피 척도를 측정할 수 있다.

- [0010] 따라서, 본원에 기재된 방법은 샘플의 종양 분획을 평가, 예컨대 추정하는 방법을 포함한다. 상기 방법은:
- [0011] 샘플의 목표 변수에 대한 값을 수득하는 단계;
- [0012] 참조에 대한 값, 예컨대 상기 목표 변수의 함수로서 확실성 메트릭을 수득하는 단계; 및
- [0013] 상기 샘플 값을 상기 참조 값과 비교하여 상기 샘플의 종양 분획에 대한 값을 수득하는 단계를 포함한다.
- [0014] 일부 구현예들에서, 대상체로부터의 샘플의 종양 분획을 결정하는 방법은 복수의 값을 수득하는 단계로서, 각 값은 상기 샘플의 하위 유전체 간격 내 대응하는 유전자좌에서 대립유전자 분획을 나타내는 단계; 상기 복수의 값의 분산을 나타내는 확실성 메트릭을 결정하는 단계; 하나 이상의 저장된 확실성 메트릭 및 하나 이상의 저장된 종양 분획 사이의 소정의 관계에 액세스하는 단계; 및 상기 확실성 메트릭 및 상기 소정의 관계로부터 상기 샘플의 종양 분획을 결정하는 단계를 포함한다.
- [0015] 일부 구현예들에서, 상기 복수의 값 내의 각 값은 대립유전자 분획이다. 일부 구현예들에서, 상기 복수의 값 내의 각 값은 대응하는 유전자좌에서 모계 대립유전자 또는 부계 대립유전자의 존재량에 대비하여 상기 모계 대립유전자 및 상기 부계 대립유전자 간의 존재량 차이의 비를 포함한다. 일부 구현예들에서, 상기 확실성 메트릭은 기대값으로부터 상기 복수의 값의 각각의 편차를 나타낸다. 일부 구현예들에서, 상기 기대값은 유전자좌 특성의 기대값이다.
- [0016] 일부 구현예들에서, 상기 확실성 메트릭은 상기 기대값으로부터의 평균 제곱근 편차이다. 일부 구현예들에서, 상기 기대값은 비종양에 대한 예상 대립유전자 빈도이다. 일부 구현예들에서, 상기 복수의 값 내의 각 값은 대립유전자 분획이고, 상기 기대값은 0.5이다.
- [0017] 일부 구현예들에서, 상기 복수의 값 내의 각 값은 대응하는 유전자좌에서 모계 대립유전자 또는 부계 대립유전자의 존재량에 대비하여 상기 모계 대립유전자 및 상기 부계 대립유전자 간의 존재량 차이의 비를 포함하고, 상기 기대값은 모계 대립유전자 또는 부계 대립유전자의 존재량에 대비하여 상기 모계 대립유전자 및 상기 부계 대립유전자 간의 존재량 차이의 기대비를 포함하며, 상기 기대값은 비종양 샘플에 대한 기대비이다. 일부 구현예들에서, 상기 기대값은 0이다.
- [0018] 일부 구현예들에서, 상기 복수의 값은 복수의 대립유전자 범위를 포함한다.
- [0019] 일부 구현예들에서, 상기 방법은 상기 복수의 값에 대한 확률 분포 함수를 결정하는 단계로서, 상기 확실성 메트릭은 상기 확률 분포 함수를 이용하여 결정되는 단계를 추가로 포함한다. 일부 구현예들에서, 상기 확실성 메트릭은 상기 확률 분포 함수의 엔트로피이다.
- [0020] 일부 구현예들에서, 상기 대응하는 유전자좌는 상이한 모계 대립유전자 및 부계 대립유전자를 갖는 하나 이상의 유전자좌를 포함한다. 일부 구현예들에서, 상기 대응하는 유전자좌는 상이한 모계 대립유전자 및 부계 대립유전자를 갖는 유전자좌로 구성된다. 일부 구현예들에서, 상기 대응하는 유전자좌는 동일한 모계 대립유전자 및 부계 대립유전자를 갖는 하나 이상의 유전자좌를 포함한다.
- [0021] 일부 구현예들에서, 대상체로부터의 샘플의 종양 분획을 결정하는 방법은: 복수의 값을 수득하는 단계로서, 각 값은 하위 유전체 간격 내 복수의 유전자좌에서 종양 샘플 내 유전자좌의 대립유전자 범위 및 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위 간의 차이를 나타내는 단계; 상기 복수의 값의 분산을 나타내는 확실성 메트릭을 결정하는 단계; 하나 이상의 저장된 확실성 메트릭 및 하나 이상의 저장된 종양 분획 사이의 소정의 관계에 액세스하는 단계; 및 상기 확실성 메트릭 및 상기 소정의 관계로부터 상기 샘플의 종양 분획을 결정하는 단계를 포함한다.
- [0022] 일부 구현예들에서, 상기 복수의 값 내의 각 값은 상기 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위에 대비한 상기 종양 샘플 내 유전자좌의 대립유전자 범위의 비를 포함한다.
- [0023] 일부 구현예들에서, 상기 복수의 값 내의 각 값은 상기 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위에 대비한 상기 종양 샘플 내 유전자좌의 대립유전자 범위의 로그비를 포함한다. 일부 구현예들에서, 상기 로그비는  $\log_2$ 이다.
- [0024] 일부 구현예들에서, 상기 복수의 값 내의 각 값은 상기 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위에 대비하여 상기 종양 샘플 내 유전자좌 및 상기 비종양 샘플 내 동일한 유전자좌 간의 대립유전자 범위의 차이의

비를 포함한다.

- [0025] 일부 구현예들에서, 상기 확실성 메트릭은 상기 대응하는 유전자좌를 통틀어 기대값으로부터 상기 복수의 값 내 각 값의 편차를 나타내고, 상기 기대값은 상기 종양 샘플이 비종양 샘플인 경우 예상되는 값이다.
- [0026] 일부 구현예들에서, 각 값은 상기 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위에 대비한 상기 종양 샘플 내 유전자좌의 대립유전자 범위의 비를 포함하고, 상기 기대값은 1이며; 각 값은 상기 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위에 대비한 상기 종양 샘플 내 유전자좌의 대립유전자 범위의 로그비를 포함하고, 상기 기대값은 0이거나; 또는 각 값은 상기 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위에 대비하여 상기 종양 샘플 내 유전자좌 및 상기 비종양 샘플 내 동일한 유전자좌 간의 대립유전자 범위의 차이의 비를 포함하고 상기 기대값은 0이다.
- [0027] 일부 구현예들에서, 상기 확실성 메트릭은 상기 기대값으로부터의 평균 제곱근 편차이다.
- [0028] 일부 구현예들에서, 상기 방법은 상기 복수의 값에 대한 확률 분포 함수를 결정하는 단계로서, 상기 확실성 메트릭은 상기 확률 분포 함수를 이용하여 결정되는 단계를 추가로 포함한다. 일부 구현예들에서, 상기 확실성 메트릭은 상기 확률 분포 함수의 엔트로피이다.
- [0029] 일부 구현예들에서, 상기 대립유전자 범위는 모계 대립유전자 및 부계 대립유전자의 유전자좌를 포함한다.
- [0030] 일부 구현예들에서, 상기 대립유전자 범위는 모계 대립유전자 및 부계 대립유전자의 대립유전자 범위로 구성된다.
- [0031] 상기 방법의 일부 구현예들에서, 상기 복수의 유전자좌는 단일 염기 다형성(SNP)과 관련된 적어도 하나의 뉴클레오티드를 포함한다. 일부 구현예들에서, 상기 복수의 유전자좌는 각각 단일 염기 다형성(SNP)과 관련된 2개 이상의 뉴클레오티드를 포함한다. 일부 구현예들에서, 상기 SNP는 암과 관련된다.
- [0032] 상기 방법의 일부 구현예들에서, 상기 복수의 유전자좌 중 적어도 일부는 복제수 변이(CNV)와 관련된다. 일부 구현예들에서, 상기 CNP는 암과 관련된다.
- [0033] 상기 방법의 일부 구현예들에서, 상기 방법은 각 유전자좌에서 대립유전자 존재량 또는 범위를 결정하기 위해 상기 샘플을 서열 분석하는 단계를 추가로 포함한다.
- [0034] 상기 방법의 일부 구현예들에서, 상기 방법은 각 유전자좌에서 대립유전자 존재량 또는 범위를 결정하기 위해 상기 샘플 상에 배열 혼성화(array hybridization)를 실시하는 단계를 추가로 포함한다.
- [0035] 상기 방법의 일부 구현예들에서, 상기 방법은 복수의 훈련 확실성 메트릭 및 연관된 훈련 종양 분획들 사이의 복수의 관계를 포함하는 훈련 데이터세트에 액세스하는 단계; 및
- [0036] 상기 훈련 확실성 메트릭 및 상기 훈련 종양 분획 사이의 소정의 관계를 결정하기 위해 상기 훈련 데이터세트에 기계 학습 공정을 적용하는 단계를 추가로 포함한다.
- [0037] 상기 방법의 일부 구현예들에서, 상기 방법은 상기 대상체 및 상기 결정된 종양 분획을 식별하는 정보를 포함하는 보고서를 생성하는 단계를 추가로 포함한다. 일부 구현예들에서, 상기 방법은 상기 대상체 또는 의료인에게 상기 보고서를 제공하는 단계를 추가로 포함한다. 일부 구현예들에서, 상기 방법은 전자 건강 기록에 대한 상기 보고서의 형식을 만드는 단계를 추가로 포함한다.
- [0038] 일부 구현예들에서, 대상체에서 종양을 치료하는 방법은 결정된 종양 분획에 반응하여 유효량의 종양 요법을 상기 대상체에게 투여하는 단계로서, 상기 종양 분획은 상기 기재된 방법들 중 어느 한 방법에 따라 결정되는 단계를 포함한다. 일부 구현예들에서, 상기 방법은 상기 결정된 종양 분획에 기초하여 상기 환자에서 상기 종양의 존재를 결정하는 단계를 포함한다. 일부 구현예들에서, 상기 종양 치료는 화학 요법, 방사선 요법 또는 수술을 포함한다.
- [0039] 일부 구현예들에서, 대상체에서 종양의 진행 또는 재발을 모니터링하는 방법은 (a) 상기 기재된 방법들 중 어느 한 방법에 따라 제1 시점에서 상기 대상체로부터 수득한 제1 샘플의 제1 종양 분획을 결정하는 단계; (b) 제2 시점에서 상기 대상체로부터 수득한 제2 샘플의 제2 종양 분획을 결정하는 단계; 및 (c) 상기 제1 종양 분획을 상기 제2 종양 분획과 비교하여 상기 종양 진행을 모니터링하는 단계를 포함한다.
- [0040] 종양의 진행 또는 재발을 모니터링하는 방법의 일부 구현예들에서, 상기 제2 종양 분획을 결정하는 단계는 제2 복수의 값을 수득하는 단계로서, 각 값은 상기 제2 종양 샘플의 하위 유전체 간격 내 대응하는 유전자좌에서 대

립유전자 분획을 나타내며, 상기 제2 샘플의 하위 유전체 간격은 상기 제1 샘플의 하위 유전체 간격과 동일하거나 상이한 단계; 상기 제2 복수의 값의 분산을 나타내는 제2 확실성 메트릭을 결정하는 단계;

[0041] 하나 이상의 저장된 확실성 메트릭 및 하나 이상의 저장된 종양 분획 사이의 소정의 관계에 액세스하는 단계; 및 상기 제2 확실성 메트릭 및 상기 소정의 관계로부터 상기 제2 샘플의 제2 종양 분획을 결정하는 단계를 포함한다.

[0042] 종양의 진행 또는 재발을 모니터링하는 방법의 일부 구현예들에서, 상기 제2 종양 분획을 결정하는 단계는 제2 복수의 값을 수득하는 단계로서, 각 값은 상기 샘플의 하위 유전체 간격 내 복수의 유전자좌에서 상기 제2 종양 샘플 내 유전자좌의 대립유전자 범위 및 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위 간의 차이를 나타내고, 상기 제2 종양 분획을 결정하기 위해 사용되는 하위 유전체 간격은 상기 제1 종양 분획을 결정하기 위해 사용되는 하위 유전체 간격과 동일하거나 상이한 단계; 상기 제2 복수의 값의 분산을 나타내는 제2 확실성 메트릭을 결정하는 단계; 하나 이상의 저장된 확실성 메트릭 및 하나 이상의 저장된 종양 분획 사이의 소정의 관계에 액세스하는 단계; 및 상기 제2 확실성 메트릭 및 상기 소정의 관계로부터 상기 제2 종양 샘플의 제2 종양 분획을 결정하는 단계를 포함한다.

[0043] 종양의 진행 또는 재발을 모니터링하는 방법의 일부 구현예들에서, 상기 방법은 상기 종양 진행에 반응하여 종양 요법을 조절하는 단계를 추가로 포함한다. 일부 구현예들에서, 상기 방법은 상기 종양 진행에 반응하여 상기 종양 요법의 투여량을 조절하거나 상이한 종양 요법을 선택하는 단계를 포함한다. 일부 구현예들에서, 상기 방법은 상기 조절된 종양 요법을 상기 대상체에게 투여하는 단계를 포함한다.

[0044] 종양의 진행 또는 재발을 모니터링하는 방법의 일부 구현예들에서, 상기 방법은 상기 제1 시점은 상기 대상체에게 종양 요법이 투여되기 전이고, 상기 제2 시점은 상기 대상체에게 상기 종양 요법이 투여된 이후인 것을 포함한다.

[0045] 상기 기재된 방법들 중 어느 한 방법의 일부 구현예들에서, 상기 대상체는 암을 가지거나, 암을 가질 위험이 있거나, 암을 가질 것으로 의심된다. 일부 구현예들에서, 상기 암은 고형 종양이다. 일부 구현예들에서, 상기 암은 혈액암이다.

[0046] 상기 기재된 방법들 중 어느 한 방법의 일부 구현예들에서, 상기 샘플은 액체 샘플이다.

[0047] 상기 기재된 방법들 중 어느 한 방법의 일부 구현예들에서, 상기 샘플은 고체 샘플이다.

[0048] 상기 기재된 방법들 중 어느 한 방법의 일부 구현예들에서, 상기 샘플은 무세포 DNA(cfDNA) 또는 순환 종양 DNA(ctDNA)를 포함한다.

[0049] 상기 기재된 방법들 중 어느 한 방법의 일부 구현예들에서, 상기 하나 이상의 저장된 확실성 메트릭은 복수의 저장된 확실성 메트릭을 포함하고, 상기 하나 이상의 저장된 종양 분획은 복수의 저장된 종양 분획을 포함한다.

[0050] 컴퓨터 시스템으로서: 프로세서; 및 상기 프로세서에 통신 가능하게 결합되고, 하기를 저장하도록 구성된 메모리로서: 하나 이상의 저장된 확실성 메트릭 및 하나 이상의 연관된 저장된 종양 분획 사이의 소정의 관계를 저장하도록 구성된 메모리; 및 상기 프로세서에 의해 실행될 때 상기 프로세서가 하기의 단계를 실시하도록 하는 명령어를 포함하는 컴퓨터 시스템으로서, 상기 명령어들은: (a)(i) 복수의 값을 수득하는 단계로서, 각 값은 상기 샘플의 하위 유전체 간격 내 대응하는 유전자좌에서 대립유전자 분획을 나타내는 단계, 또는 (ii) 복수의 값을 수득하는 단계로서, 각 값은 하위 유전체 간격 내 복수의 유전자좌에서 종양 샘플 내 유전자좌의 대립유전자 범위 및 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위 간의 차이를 나타내는 단계; (b) 상기 복수의 값의 분산을 나타내는 확실성 메트릭을 결정하는 단계; (c) 상기 저장된 소정의 관계에 액세스하는 단계; 및 (d) 상기 확실성 메트릭 및 상기 소정의 관계로부터 상기 샘플의 종양 분획을 결정하는 단계를 실시하도록 하게 한다.

[0051] 상기 컴퓨터 시스템의 일부 구현예들에서, 상기 메모리는 상기 프로세서에 의해 실행될 때 상기 프로세서가 하기의 단계를 실시하도록 하는 명령어를 추가로 포함하는 컴퓨터 시스템으로서, 상기 명령어들은: 복수의 훈련 확실성 메트릭 및 연관된 훈련 종양 분획들 사이의 복수의 관계를 포함하는 훈련 데이터세트에 액세스하는 단계; 및

[0052] 상기 훈련 확실성 메트릭 및 상기 훈련 종양 분획 사이의 소정의 관계를 결정하기 위해 상기 훈련 데이터세트에 기계 학습 공정을 적용하는 단계를 실시하도록 하게 한다.

[0053] 상기 컴퓨터 시스템의 일부 구현예들에서, 상기 명령어들은 상기 프로세서에 의해 실행될 때 상기 프로세서가

상기 기재된 방법들 중 어느 한 방법을 실시하도록 하게 한다.

### 도면의 간단한 설명

[0054] 하나 이상의 예시한 대한 다양한 양태는 축적으로 그려지도록 의도되지 않은 첨부 도면을 참조하여 아래에서 논의된다. 도면은 다양한 양태 및 예시에 대한 설명 및 추가적인 이해를 제공하기 위해 포함되며, 본 명세서에 통합되고 이의 일부를 구성하지만 특정 예시의 제한을 정의하는 것으로 의도되지 않는다. 명세서의 나머지 부분과 함께 도면은 기재되고 청구된 양태들 및 예시들의 원리 및 작동을 설명하는 역할을 한다. 다양한 도면에 예시된 각각의 동일하거나 거의 동일한 구성요소는 도면에서 유사한 번호로 표시된다. 명확성을 위해, 모든 구성 요소가 모든 도면에 표시되지 않을 수도 있다.

도 1은 일 구현예에 따른 공정을 도시한다. 상기 개시된 공정은 샘플로부터 종양 분획을 추정하기 위해 사용될 수 있다.

도 2는 본 개시의 다양한 양태들이 실시될 수 있는 예시적인 컴퓨터 시스템을 도시한다.

도 3은 본 개시의 다양한 양태들을 구현할 수 있는 예시적인 저장 시스템을 도시한다.

도 4는 연속 희석된 여러 암 샘플을 사용하여 결정된 바와 같이 관련 종양 분획(최대 체세포 대립유전자 빈도로 표시됨)을 갖는 샘플에서 SNP 대립유전자 분획에 대한 확률 분포 함수의 엔트로피 간의 예시적인 관계를 도시한다.

### 발명을 실시하기 위한 구체적인 내용

[0055] 대상체로부터의 샘플의 종양 분획을 결정하기 위한 방법 및 시스템이 본원에 기재된다. 결정된 종양 분획에 반응하여 대상체에서 종양을 치료하는 방법, 및 2개 이상의 시점에서 상기 대상체로부터 수득한 샘플의 종양 분획을 결정하는 단계를 포함하여 대상체에서 종양 진행 또는 재발에 모니터링하기 위한 방법 및 시스템이 또한 기재된다. 특히 낮은 종양 분획 수준에서 빠르고 정확한 종양 분획 결정은 상기 종양 또는 종양 재발의 초기 단계에서 상기 대상체가 효과적인 치료를 받도록 함으로써 종양 요법을 실질적으로 향상시킬 수 있다. 종양 분획에 대한 다른 사용이 또한 고려되고 본원에서 추가로 논의된다. 예를 들어, 상기 종양 분획은 일부 구현예들에서 종양 생검을 분석하기 위해 사용될 수 있다. 일부 구현예들에서, 상기 종양 분획은, 예를 들어 체세포-생식세포 계열-접합성(SGZ) 알고리즘을 사용하여 변이체(예를 들어, 체세포 또는 생식세포 계열로서, 또는 동형 접합체, 이형 접합체 또는 하위 클론으로서)를 특성화하는 데 사용된다. 본원에 기재된 방법 및 시스템은 낮은 종양 분획 수준에서도 정확한 종양 분획 결정을 제공한다.

[0056] 본원에 추가로 기재된 바와 같이, 종양 분획은 복수의 분석된 유전자좌에 걸친 대립유전자 분획 분산과 밀접하게 연관된다. 상기 분산은 "확실성 메트릭"이라고 할 수 있다. 하나 이상의 확실성 메트릭 및 하나 이상의 대응하는 종양 분획 사이의 관계는 상기 대상체로부터의 샘플의 결정된 확실성 메트릭으로부터 샘플의 종양 분획을 결정하는 데 사용될 수 있다. 상기 관계는 상기 결정된 확실성 메트릭을 입력으로 받고 상기 샘플에 대한 종양 분획을 출력한다. 상기 관계는 효과적인 종양 요법, 종양 진행 또는 재발에 대한 대상체의 모니터링, 및/또는 종양 샘플의 분석을 가능하게 할 수 있는 상기 대상체로부터의 샘플의 종양 분획을 결정하는 데 적용될 수 있다.

[0057] 일부 구현예들에서, 상기 샘플의 종양 분획은 상기 종양 샘플 및 비종양 샘플(예컨대, 건강한 조직 샘플)을 사용하여 종양 샘플에 대해 결정된다. 상기 종양 샘플 및 상기 비종양 샘플은 동일한 개체(즉, 일치하는 정상 대조군) 또는 상이한 개체들로부터 수득할 수 있다. 상기 확실성 메트릭은 복수의 값에 대한 분산일 수 있고, 각각의 값은 복수의 유전자좌에서 상기 종양 샘플 내 유전자좌의 범위 및 상기 비종양 샘플 내 동일한 유전자좌의 범위 간의 차이를 나타낸다. 상기와 같이, 확실성 메트릭 및 종양 분획 사이의 관계는 상기 대상체로부터의 샘플의 결정된 확실성 메트릭으로부터 상기 샘플의 종양 분획을 결정하는 데 사용될 수 있다. 상기 관계는 상기 결정된 확실성 메트릭을 입력으로 받고 상기 샘플에 대한 종양 분획을 출력한다. 상기 관계는 효과적인 종양 요법, 종양 진행 또는 재발에 대한 대상체의 모니터링, 및/또는 종양 샘플의 분석을 가능하게 할 수 있는 상기 대상체로부터의 샘플의 종양 분획을 결정하는 데 적용될 수 있다.

[0058] 종양 분획 결정

[0059] 암의 모니터링, 진단 및 치료에 있어 중요한 지표는 종양 분획이다. 일부 구현예들에서, 종양 분획은 세포 기원에 관계없이 총 유전체 함량에 비례하여, 예를 들어 샘플(예컨대, 생검)에서 종양 유전체 함량의 척도이다. 일

반적으로, 샘플로부터 종양 함량 또는 종양 함량의 변화를 결정(예컨대, 추정)하는 것이 유리하다. 이는 변경을 보고하고 질병의 존재 또는 진행을 알리는 데 도움이 될 수 있기 때문이다. 예를 들어, 일반적으로 암 환자로부터의 혈액 샘플을 활용하는 액체 생검은 고체 생검이 가능하지 않거나 권장되지 않을 때 유용할 수 있다. 본원에 기재된 방법은 다양한 유형의 샘플, 예를 들어 고체 및 액체 샘플에서 종양 분획을 결정하는 데 사용될 수 있다. 일부 구현예들에서, 본원에 기재된 방법은, 예컨대 대안으로서 또는 시각적 선별 방법과 조합하여 고체 샘플에 사용된다. 다른 구현예들에서, 본원에 기재된 방법은, 예컨대 시각적 선별 방법이 효과적이지 않거나 이용 가능하지 않은 경우 액체 샘플에 사용된다.

[0060] 일부 구현예들에서, 무세포 샘플의 종양 분획은 혈류로 흘러들어 가서 혈액순환을 통해 몸 전체에 운반되고 있는 총 DNA(예컨대, 종양 및 정상)의 양에 대비하여 원발성 종양으로부터 혈관계 또는 림프관으로 배출된 종양 DNA의 측정치를 포함한다. 종양 분획은 암을 가질 위험이 있는(현재 진단이 있거나 없는) 환자를 모니터링하는 데; 암 진단에 사용되는 인자로서; 또는 현재 치료 요법이 효과, 예컨대 유익한 효과를 갖는지 여부를 결정하기 위해 사용될 수 있다.

[0061] 종양 분획을 측정하기 위한 전통적인 접근 방식은 일반적으로 모델링된 매개변수인 순도 및 배수성 모두를 로그비 또는 대립유전자 빈도 측정 또는 이의 둘 다, 또는 병리학 검토를 통해 추론할 것을 요구한다. 일부 구현예들에서, 종양 분획은 이중 종양 샘플에서 암 세포 분획의 모델링된 변수로서 고려할 수 있고 종양 순도 또는 다른 측정을 고려할 수 있다. 일부 구현예들에서, 종양 세포 배수성은 모든 염색체(또는 이의 일부)의 평균 가중복제 수를 지칭할 수 있다. 샘플에서 관찰된 배수성은 다양한 정도의 종양 세포 이수성, 샘플의 이질성(예컨대, 정상 세포에 대한 종양 세포의 상이한 비율), 또는 둘 다에 의해 영향을 받을 수 있다.

[0062] 종양 분획을 예측하기 위한 전통적인 접근 방식은 적합하지 않은 모델에 따른 낮은 종양 함량으로 인해 상당히 신뢰할 수 없다. 일부 구현예들에서, 본원에 기재된 방법은, 예컨대 샘플에서 하나 이상의 하위 유전체 간격으로 대립유전자 범위 또는 대립유전자 분획에 의해 측정되는 바와 같이, 예를 들어 종양 세포 이수성의 효과에 기초하여 종양 분획(및 관련 신뢰 수준)을 결정함으로써 전통적인 접근법의 특정 결점들을 극복할 수 있다. 일부 구현예들에서, 상기 하위 유전체 간격은 이형접합의 단일 염기 다형성(SNP) 부위를 포함한다. 다른 구현예들에서, 상기 하위 유전체 간격은 하나 초과 뉴클레오타이드 위치를 포함한다.

[0063] 본원에서 사용된 용어 "대립유전자 범위", 또는 간단히 "범위" 또는 " $Cvg$ "는 샘플에서 하위 유전체 간격의 DNA 서열분석으로부터 생성된 리드(예컨대, 고유 리드)의 수를 지칭한다. 본원에서 사용된 용어 "대립유전자 강도", 또는 간단히 "강도"는 샘플에서 하위 유전체 간격의 유전체 혼성화로부터 생성된 신호(예컨대, 고유 신호)의 수를 지칭한다. "리드" 또는 "신호"는 동일한 "고유 리드" 또는 "고유 신호"의 중복이 존재할 수 있지만(즉, 본원에 기재된 방법을 실시하기 전에 중복이 제거되지 않은), 상기 중복은 분자 및 분모 모두로 표시되기 때문에 기재된 방법을 사용하여 계산된 임의의 비율이 "고유한" 리드 또는 신호 비율과 매우 유사한 값을 생성하는 상황을 포함하도록 의도되는 것으로 이해될 것이다.

[0064] 본원에서 사용된 용어 "대립유전자 분획"은 샘플에서 하위 유전체 간격의 대립유전자의 상대적 수준(예컨대, 존재량)을 지칭한다. 대립 유전자 분획은 분수 또는 백분율로 나타낼 수 있다. 예를 들어, 대립유전자 분획은 하위 유전체 간격의 모든 상이한 대립유전자 수에 대비한 상기 하위 유전체 간격의 한 특정 대립유전자(예컨대, A, T, C 또는 G) 수의 비율로 표현될 수 있다. 일부 구현예들에서, 대립유전자 분획은 주어진 하위 유전체 간격의 모든 상이한 대립유전자로부터의 총 범위 또는 강도에 대한 하나의 특정 대립유전자(예컨대, A, T, C, 또는 G)로부터의 범위 또는 강도의 비율을 계산함으로써 측정된다. 때때로, 용어 "대립유전자 분획" 및 "대립유전자 빈도"는 본원에서 상호교환가능하게 사용된다. 본원에서 사용된 바와 같이, 로그비는 전형적으로  $\log_2(T/R)$ 로 측정되며, 여기서 T는 샘플의 하위 유전체 간격과 관련된 하나 이상의 대립유전자의 수준(예컨대, 존재량)이고, R은 참조 샘플의 하위 유전체 간격과 관련된 하나 이상의 대립유전자의 수준(예컨대, 존재량)이다. 본원에서 사용된 용어 "대립유전자"는 유전체 서열(예컨대, 유전자 또는 이의 임의의 부분)의 둘 이상의 대안적 형태 중 하나를 지칭한다. 예를 들어, "C" 내지 "T" SNP가 하위 유전체 간격과 관련되어 있으면, 상기 하위 유전체 간격은 상기 SNP에 대한 대립유전자 "C" 및 "T"와 관련되는 것으로 설명될 수 있다.

[0065] 일부 구현예들에서, 하위 유전체 간격과 관련된 2개 이상의 상이한 대립유전자가 존재한다. 만약 상기 2개 이상의 상이한 대립 유전자가 샘플에 존재하는 경우, 상기 하위 유전체 간격은 상기 샘플에 대해 이형 접합으로 간주된다. 상기 하위 유전체 간격이 상기 샘플에 대해 이형접합이 아닌 경우, 일부 구현예들에서는 동형접합, 반접합(semizygous) 또는 반접합(hemizygous)일 수 있다.

[0066] 본원에서 사용된 용어 "존재량"은 대상의 양, 수 또는 분량을 지칭한다. 예를 들어, 하위 유전체 간격과 관련된

대립유전자의 존재량은, 예를 들어 서열분석 또는 배열 기반 포괄적 유전체 혼성화(array-based comprehensive genomic hybridization, aCGH)에 의해 결정된 바와 같이 샘플에서 하위 유전체 간격과 관련된 대립유전자의 양, 수 또는 분량을 의미할 수 있다. 예를 들어, 특정 하위 유전체 간격과 관련된 2개의 대립유전자 "A" 및 "G"가 있고, 샘플에 대립유전자 "A"의 복제 10개 있고 대립유전자 "G"의 복제 20개 있는 경우, 대립유전자 "A"의 존재량은 10으로 간주될 수 있고 대립 유전자 "G"의 존재량은 20으로 간주될 수 있다. 일부 구현예들에서, 대립유전자의 존재량은 대립유전자 범위 또는 대립유전자 강도에 의해 측정된다. 예를 들어, 대립유전자 "A" 또는 "G"에 대한 고유 리드 수는 상기 샘플 내에 얼마나 많은 대립유전자 "A" 또는 "G"의 복제 수가 존재하는지를 반영한다.

[0067] 본원에서 사용되는 용어 "확실성 메트릭(certainty metric)"은 목표 변수의 측정 또는 값으로부터 유래된 척도를 지칭한다. 일부 구현예들에서, 상기 목표 변수는 샘플에서 하위 유전체 간격의 존재량, 또는 상기 하위 유전체 간격과 관련된 대립유전자를 나타낼 수 있다. 일부 예시들에서, 상기 확실성 메트릭은 예상 대립유전자 분획으로부터 대립유전자 분획의 편차일 수 있다. 다른 예시들에서, 상기 확실성 메트릭은 대립유전자 강도의 척도일 수 있다. 상기 예시들은 설명을 위한 것이며 다른 확실성 메트릭들이 사용될 수 있다.

[0068] 일례로써, 이형접합 SNP의 경우, 대립 유전자 분획 값 0.50은 전형적인 이배체 하위 유전체 간격을 나타낼 수 있고; 기대값 0.50에서 벗어난 대립 유전자 분획은 해당 부위에서 이수성을 나타낸다. 상기 예시들에서, 상기 대립유전자의 편차는 대립유전자 범위에 기초한 중앙 분획을 결정(예컨대, 예측 또는 추정)하는 모델을 구축하기 위해 훈련 세트의 중앙 분획과 상관관계가 있을 수 있다. 일부 구현예들에서, 본원에 기재된 방법은 대립유전자 분획 또는 로그비의 편차가 중앙 분획과 상관관계가 있음을 보여줌으로써, 중앙 순도 및 배수성을 모델링할 필요성을 제거한다. 일부 구현예들에서, 본원에 기재된 방법은 낮은 수준, 예컨대 30% 미만의 중앙 분획에 대하여 보다 정확한 결정을 가능하게 한다. 구현예에서, 상기 대립유전자 분획 또는 로그비는 서열분석, 예컨대 차세대 서열분석(NGS)을 포함하는 방법에 의해 결정된다. 대립유전자 분획 또는 로그비를 결정하기 위한 방법은 서열분석에 제한되지 않음에 이해될 것이다. 예를 들어, SNP 범위 또는 SNP의 상대적 수준(예컨대, 존재량)을 측정하는 임의의 방법 뿐만 아니라 더 큰 유전체 영역으로부터의 범위를 측정하는 임의의 방법을 사용할 수 있다. 한 구현예에서, 상기 대립유전자 분획 또는 로그비는 서열분석 이외의 방법에 의해 결정되며, 예컨대 배열 기반 포괄적 유전체 혼성화(array-based comprehensive genomic hybridization, aCGH)에 의해 결정된다. 한 구현예에서, 상기 중앙 분획은 0.25 이하, 0.2 이하, 0.15 이하, 또는 0.1 이하, 예컨대 0.1 내지 0.3, 0.1 내지 0.2, 0.2 내지 0.3, 또는 0.15와 0.25이거나 그러할 것으로 기대된다.

[0069] 일부 구현예들에서 본원에 기재된 방법은 예상되는 범위의 비율을 나타내기 위해 대립유전자 분획 또는 로그비를 사용하지만, 본 개시는 일반적으로 대립유전자 분획, 로그비 또는 기타 특정 메트릭에 제한되지 않고 예상되는 범위 편차에 대한 중앙 분획의 상관관계를 기재하도록 의도되는 것으로 이해될 것이다.

[0070] 본원에서 사용된 "단일 염기 다형성" 또는 SNP는 유전체의 특정 위치에서 발생하는 단일 염기의 변경을 지칭한다. 일부 구현예들에서, 상기 변경은 집단 내에서 어느 정도 인식할 수 있는 정도(예컨대, >1%)로 존재한다. 일반적으로, SNP는 생식선 변경이며 체세포 단일 염기 변이체(SNV)가 아니다.

[0071] 한 구현예에서, 상기 중앙 분획은 상기 샘플 내 DNA(예컨대, 중앙 및 비중앙 DNA)의 총량에 대비한 중앙 세포로부터의 DNA의 양을 나타내는 수치적 표현(예컨대, 분획 또는 백분율)이다. 한 구현예에서, 상기 샘플은 액체 생검이다. 한 구현예에서, 상기 샘플은 고형 조직 샘플이다. 한 구현예에서, 상기 암은 고형 종양이다. 한 구현예에서, 상기 종양은 혈액암이다. 한 구현예에서, 액체 생검에서 중앙 분획은 체내에서 검출가능한 종양의 존재 또는 수준을 나타낸다.

[0072] 대상체로부터의 샘플 내 중앙 분획을 결정하는 예시적인 방법은: 복수의 값을 수득하는 단계로서, 각 값은 상기 샘플에서 하위 유전체 간격 내 대응하는 유전자좌의 대립유전자 분획을 나타내는 단계; 상기 복수의 값의 분산을 나타내는 확실성 메트릭을 결정하는 단계; 저장된 확실성 메트릭 및 저장된 중앙 분획 사이의 사전 결정된 관계에 액세스하는 단계; 및 상기 확실성 메트릭 및 상기 소정의 관계로부터 상기 샘플의 중앙 분획을 결정하는 단계를 포함한다.

[0073] 대립유전자 분획을 나타내는 값은 각각의 대응하는 유전자좌에 대해 결정될 수 있다. 상기 유전자좌는 하나 이상의 뉴클레오티드를 포함할 수 있다. 일부 구현예들에서, 상기 대응하는 유전자좌는 상이한 모계 대립유전자 및 부계 대립유전자를 갖는 하나 이상의 유전자좌를 포함한다. 일부 구현예들에서, 상기 대응하는 유전자좌는 상이한 모계 대립유전자 및 부계 대립유전자를 갖는 유전자좌로 구성된다. 일부 구현예들에서, 상기 대응하는 유전자좌는 동일한 모계 대립유전자 및 부계 대립유전자를 갖는 하나 이상의 유전자좌를 포함한다.

[0074] 일부 구현예들에서, 상기 샘플 내 복수의 대응하는 유전자좌에서 대립유전자 분획을 나타내는 복수의 값은 상기 샘플 내 복수의 대응하는 유전자좌에서의 복수의 대립유전자 분획이다. 각각의 대응하는 유전자좌에서의 대립유전자 분획은, 예를 들어 상기 중앙 샘플에서 핵산 분자를 서열분석하고 각각의 유전자좌에서 각 대립유전자에 대한 대립유전자 범위를 할당함으로써 결정될 수 있다. 예를 들어, 유전자좌  $i$  ( $af_i$ )에서의 대립유전자 분획은 하기에 의해 결정될 수 있고:

$$af_i = \frac{Cvg_{i,a}}{Cvg_{i,a} + Cvg_{i,b}}$$

[0075]

[0076] 여기서  $Cvg_{i,a}$  은 유전자좌  $i$ 에서 대립유전자  $a$ 의 범위이고,  $Cvg_{i,b}$  은 유전자좌  $i$ 에서 대립유전자  $b$ 의 범위이다. 일부 구현예들에서, 대립유전자  $a$  및 대립유전자  $b$ 가 할당되어  $Cvg_{i,a} \leq Cvg_{i,b}$  이 되고,  $af_i \leq 0.5$  이 된다.

[0077] 일부 구현예들에서, 예상되는 대립유전자 분획은 건강한 개체 또는 건강한 샘플(즉, 비종양 샘플)에서 예상되는 대립유전자 분획이다. 예를 들어, 이형접합 유전자좌(즉, 상이한 모계 대립유전자 및 부계 대립유전자를 가짐)에서의 대립유전자 분획은 0.5일 것으로 예상되고, 동형접합 유전자좌(즉, 상기 모계 대립유전자 및 상기 부계 대립유전자가 동일함)는 1.0일 것으로 예상된다.

[0078] 대립유전자 분획은, 일부 구현예들에서 대립유전자 분획을 나타내는 다른 값들이 사용될 수 있지만, 본원에 기재된 방법에 따라 종양 분획을 결정하기 위한 예시적인 값이다. 일부 구현예들에서, 상기 대립유전자 분획을 나타내는 값은 대립유전자 빈도의 상대적 차이이다. 예를 들어, 상기 대립유전자 분획을 나타내는 값은 모계 대립유전자 또는 부계 대립유전자의 존재량에 대비하여 상기 모계 대립유전자 및 상기 부계 대립유전자 간의 존재량(예컨대, 범위 또는 서열분석의 정도)의 차이의 비율 수 있다. 즉, 일부 구현예들에서, 상기 값은 하기에 의해 상대\_차이일 수 있고:

$$\text{상대_차이} = \frac{Cvg_{i,a} - Cvg_{i,b}}{Cvg_{i,b}}$$

[0079]

[0080] 여기서  $Cvg_{i,a}$  은 유전자좌  $i$ 에서 대립유전자  $a$ 의 범위이고,  $Cvg_{i,b}$  은 유전자좌  $i$ 에서 대립유전자  $b$ 의 범위이다. 건강한 개체 또는 건강한 샘플에서, 대립유전자 빈도 간의 차이 뿐만 아니라 상대적 차이는 0일 것으로 예상된다. 일부 구현예들에서, 대립유전자 분획을 나타내는 복수의 값에 대해 확률 분포 함수가 결정된다. 예를 들어, 일부 구현예들에서, 상기 확률 분포 함수는 상기 샘플 내의 복수의 대응하는 유전자좌에서 복수의 대립유전자 분획에 대해 결정된다. 일부 구현예들에서, 상기 복수의 대립유전자 분획에 대한 확률 분포 함수는 하기와 같이 정의된다:

$$P(af) = P\left(\frac{Cvg_{i,a}}{Cvg_{i,a} + Cvg_{i,b}}\right)$$

[0081]

[0082] 여기서  $Cvg_{i,a}$  은 유전자좌  $i$ 에서 대립유전자  $a$ 의 범위이고,  $Cvg_{i,b}$  은 유전자좌  $i$ 에서 대립유전자  $b$ 의 범위이다.

[0083] 분산(또는 확실성 메트릭)은, 예를 들어 상기 복수의 유전자좌에 걸친 예상되는 대립유전자 분획(또는 예상되는 대립유전자 분획을 나타내는 값)으로부터의 편차일 수 있다. 일부 구현예들에서, 상기 확실성 메트릭은 상기 예상되는 대립유전자 분획(또는 이를 나타내는 값)로부터의 평균 제곱근 편차이다. 예를 들어, 일부 구현예들에서, 상기 확실성 메트릭은 하기에 의해 정의되는 평균 제곱근 편차(RMSD)이다:

$$RMSD = \left[ \frac{1}{N} \sum_{i=0}^N (af_i - af_{i,\neq\neq})^2 \right]^{(1/2)}$$

[0084]

[0085] 여기서,  $af_i$  은 유전자좌  $i$ 에서의 대립유전자 빈도(또는 대립유전자 빈도를 나타내는 값, 예컨대 상대 차이 비율)이고,  $af_{\neq\neq}$  은 유전자좌  $i$ 에서의 예상되는 대립유전자 빈도이며,  $N$ 은 복수의 대응하는 유전자좌 내의 유전

자좌의 수이다. 예를 들어, 일부 유전자좌의 경우,  $af_{\text{좌}}$  은 0.5일 수 있고, 다른 유전자좌에서  $af_{\text{좌}}$  은 1일 수 있다. 일부 구현예들에서, 상기 유전자좌는 상이한 모계 대립유전자 및 부계 대립유전자를 갖는 유전자좌만을 포함한다. 따라서, 상기  $af_{\text{좌}}$  은 모든 유전자좌에 걸쳐 0.5로 정의될 수 있고, 상기 RMSD는 하기와 같이 정의될 수 있다:

$$RMSD = \left[ \frac{1}{N} \sum_{i=0}^N (af_i - 0.5)^2 \right]^{(1/2)}$$

일부 구현예들에서, 상기 대립유전자 분획을 나타내는 값은 모계 대립유전자 또는 부계 대립유전자의 존재량에 대비하여 상기 모계 대립유전자 및 상기 부계 대립유전자 간의 존재량(예컨대, 범위 또는 서열분석의 정도)의 차이의 비일 수 있고, 상기  $af_{\text{좌}}$  은 0으로서 정의될 수 있다. 따라서, 상기 RMSD는 하기와 같이 정의될 수 있다:

$$RMSD = \left[ \frac{1}{N} \sum_{i=0}^N \left( \frac{Cvg_{i,a} - Cvg_{i,b}}{Cvg_{i,b}} \right)^2 \right]^{(1/2)}$$

여기서  $Cvg_{i,a}$  은 유전자좌  $i$ 에서 대립유전자  $a$ 의 범위이고,  $Cvg_{i,b}$  은 유전자좌  $i$ 에서 대립유전자  $b$ 의 범위이다.

일부 구현예들에서, 확률 분포(예컨대, 확률 분포 함수)는 복수의 유전자좌에 걸친 대립유전자 분획에 대해 결정될 수 있다. 상기 확실성 메트릭(예컨대, 분산)은 확률 분포의 메트릭, 예컨대 확률 분포의 엔트로피일 수 있다. 예를 들어, 일부 구현예들에서, 대립유전자 분획 확률 분포 함수( $S[P(af)]$ )의 엔트로피는 하기와 같이 정의될 수 있다:

$$S[P(af)] = \sum_{af=0}^{0.5} P(af) \log_n(P(af))$$

여기서  $P(af)$  은 상기 대립유전자 분획 확률 분포 함수이고,  $n$ 은 로그 염기이다. 일부 구현예들에서, 상기 로그 염기는 2(즉,  $\log_2$ )이다. 따라서, 일부 구현예들에서, 대립유전자 분획 확률 분포 함수( $S[P(af)]$ )의 엔트로피는 하기와 같이 정의될 수 있다:

$$S[P(af)] = \sum_{af=0}^{0.5} P(af) \log_2(P(af))$$

일부 구현예들에서, 대상체로부터의 샘플의 종양 분획을 결정하는 방법은: 복수의 값을 획득하는 단계로서, 각 값은 하위 유전체 간격 내 복수의 유전자좌에서 종양 샘플 내 유전자좌의 대립유전자 범위 및 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위 간의 차이를 나타내는 단계; 상기 복수의 값의 분산을 나타내는 확실성 메트릭을 결정하는 단계; 저장된 확실성 메트릭 및 저장된 종양 분획 사이의 소정의 관계에 액세스하는 단계; 및 상기 확실성 메트릭 및 상기 소정의 관계로부터 상기 샘플의 종양 분획을 결정하는 단계를 포함한다. 일부 구현예들에서, 종양 샘플 및 비종양 샘플은 동일한 개체(즉, 일치된 정상 대조군)로부터 획득한다. 일부 구현예들에서, 상기 종양 샘플 및 상기 비종양 샘플은 상이한 개체들로부터 획득한다. 범위는 원시 범위(예를 들어, 서열 분석 리드의 원시 수), 정규화 범위(예를 들어, 평균 또는 중앙 서열분석 정도로 정규화), 및/또는 편향 보정된 범위(예를 들어, GC 편향 보정된 범위의 정도)이다. 일부 구현예들에서, 상기 대립유전자 범위는 모계 대립유전자 범위 및 부계 대립유전자 범위(예컨대, 상기 모계 대립유전자의 범위 및 상기 부계 대립유전자의 범위의 합)를 포함한다. 일부 구현예들에서, 상기 대립유전자 범위는 모계 대립유전자 범위 및 부계 대립유전자 범위(예컨대, 상기 모계 대립유전자의 범위 및 상기 부계 대립유전자의 범위의 합)로 구성된다.

일부 구현예들에서, 종양 샘플 내 유전자좌의 대립유전자 범위 및 비종양 샘플 내 동일한 유전자좌의 대립유전

자 범위 간의 차이를 나타내는 각 값은 상기 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위에 대비한 상기 종양 샘플 내 유전자좌의 대립유전자 범위의 비를 포함한다. 일부 구현예들에서, 상기 대립유전자 범위는 모계 대립유전자 범위 및 부계 대립유전자 범위(예컨대, 상기 모계 대립유전자의 범위 및 상기 부계 대립유전자의 범위의 합)를 포함한다. 일부 구현예들에서, 상기 대립유전자 범위는 모계 대립유전자 범위 및 부계 대립유전자 범위(예컨대, 상기 모계 대립유전자의 범위 및 상기 부계 대립유전자의 범위의 합)로 구성된다. 예를 들어, 일부 구현예들에서, 상기 비는 하기와 같이 정의될 수 있다:

$$b_j = \frac{(Cvg_{i,a}^{af} + Cvg_{i,b}^{af})}{(Cvg_{i,a}^{ff} + Cvg_{i,b}^{ff})}$$

여기서  $Cvg_{i,a}^{af}$ 은 상기 종양 샘플 내 유전자좌  $i$ 에서 모계 대립유전자의 범위이고,  $Cvg_{i,b}^{af}$ 은 상기 종양 샘플 내 유전자좌  $i$ 에서 부계 대립유전자의 범위이고,  $Cvg_{i,a}^{ff}$ 은 상기 비종양 샘플 내 유전자좌  $i$ 에서 모계 대립유전자의 범위이며,  $Cvg_{i,b}^{ff}$ 은 상기 비종양 샘플 내 유전자좌  $i$ 에서 부계 대립유전자의 범위이다.

일부 구현예들에서, 종양 샘플 내 유전자좌의 대립유전자 범위 및 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위 간의 차이를 나타내는 각 값은 상기 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위에 대비한 상기 종양 샘플 내 유전자좌의 대립유전자 범위의 로그비(예컨대, 로그<sub>2</sub>비)를 포함한다. 일부 구현예들에서, 상기 대립유전자 범위는 모계 대립유전자 범위 및 부계 대립유전자 범위(예컨대, 상기 모계 대립유전자의 범위 및 상기 부계 대립유전자의 범위의 합)를 포함한다. 일부 구현예들에서, 상기 대립유전자 범위는 모계 대립유전자 범위 및 부계 대립유전자 범위(예컨대, 상기 모계 대립유전자의 범위 및 상기 부계 대립유전자의 범위의 합)로 구성된다. 예를 들어, 상기 로그비는 일부 구현예들에서 하기와 같이 정의될 수 있다:

$$\log_n b_j = \log_n \frac{(Cvg_{i,a}^{af} + Cvg_{i,b}^{af})}{(Cvg_{i,a}^{ff} + Cvg_{i,b}^{ff})}$$

여기서 로그 <sub>$n$</sub> 은 염기  $n$ 에서의 로그이고,  $Cvg_{i,a}^{af}$ 은 상기 종양 샘플 내 유전자좌  $i$ 에서 모계 대립유전자의 범위이고,  $Cvg_{i,b}^{af}$ 은 상기 종양 샘플 내 유전자좌  $i$ 에서 부계 대립유전자의 범위이고,  $Cvg_{i,a}^{ff}$ 은 상기 비종양 샘플 내 유전자좌  $i$ 에서 모계 대립유전자의 범위이며,  $Cvg_{i,b}^{ff}$ 은 상기 비종양 샘플 내 유전자좌  $i$ 에서 부계 대립유전자의 범위이다. 예를 들어, 상기 로그비는 로그<sub>2</sub>비일 수 있다. 일부 구현예들에서, 상기 로그비는 하기와 같이 정의된다:

$$\log_2 b_j = \log_2 \frac{(Cvg_{i,a}^{af} + Cvg_{i,b}^{af})}{(Cvg_{i,a}^{ff} + Cvg_{i,b}^{ff})}$$

여기서  $Cvg_{i,a}^{af}$ 은 상기 종양 샘플 내 유전자좌  $i$ 에서 모계 대립유전자의 범위이고,  $Cvg_{i,b}^{af}$ 은 상기 종양 샘플 내 유전자좌  $i$ 에서 부계 대립유전자의 범위이고,  $Cvg_{i,a}^{ff}$ 은 상기 비종양 샘플 내 유전자좌  $i$ 에서 모계 대립유전자의 범위이며,  $Cvg_{i,b}^{ff}$ 은 상기 비종양 샘플 내 유전자좌  $i$ 에서 부계 대립유전자의 범위이다.

일부 구현예들에서, 종양 샘플 내 유전자좌의 대립유전자 범위 및 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위 간의 차이를 나타내는 각 값은 상기 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위에 대비하여 상

기 비중앙 샘플 내 동일한 유전자좌의 대립유전자 범위에 대비한 상기 중앙 샘플 내 유전자좌의 대립유전자 범위 간의 차이의 비를 포함한다. 일부 구현예들에서, 상기 대립유전자 범위는 모계 대립유전자 범위 및 부계 대립유전자 범위(예컨대, 상기 모계 대립유전자의 범위 및 상기 부계 대립유전자의 범위의 합)를 포함한다. 일부 구현예들에서, 상기 대립유전자 범위는 모계 대립유전자 범위 및 부계 대립유전자 범위(예컨대, 상기 모계 대립유전자의 범위 및 상기 부계 대립유전자의 범위의 합)로 구성된다. 예를 들어, 일부 구현예들에서, 상기 비는 하기와 같이 정의된다:

$$\frac{(Cvg_{i,a}^{\text{모}} + Cvg_{i,b}^{\text{모}}) - (Cvg_{i,a}^{\text{부}} + Cvg_{i,b}^{\text{부}})}{(Cvg_{i,a}^{\text{부}} + Cvg_{i,b}^{\text{부}})}$$

[0104]

여기서  $Cvg_{i,a}^{\text{모}}$ 은 상기 중앙 샘플 내 유전자좌  $i$ 에서 모계 대립유전자의 범위이고,  $Cvg_{i,b}^{\text{모}}$ 은 상기 중앙 샘플 내 유전자좌  $i$ 에서 부계 대립유전자의 범위이고,  $Cvg_{i,a}^{\text{부}}$ 은 상기 비중앙 샘플 내 유전자좌  $i$ 에서 모계 대립유전자의 범위이며,  $Cvg_{i,b}^{\text{부}}$ 은 상기 비중앙 샘플 내 유전자좌  $i$ 에서 부계 대립유전자의 범위이다.

[0106]

일부 구현예들에서, 확률 분포 함수는 중앙 샘플 내 유전자좌의 대립유전자 범위 및 비중앙 샘플 내 동일한 유전자좌의 대립유전자 범위 사이의 차이를 나타내는 복수의 값에 대해 결정된다. 일부 구현예들에서, 상기 대립유전자 범위는 모계 대립유전자 범위 및 부계 대립유전자 범위(예컨대, 상기 모계 대립유전자의 범위 및 상기 부계 대립유전자의 범위의 합)를 포함한다. 일부 구현예들에서, 상기 대립유전자 범위는 모계 대립유전자 범위 및 부계 대립유전자 범위(예컨대, 상기 모계 대립유전자의 범위 및 상기 부계 대립유전자의 범위의 합)로 구성된다. 예를 들어, 일부 구현예들에서, 상기 확률 분포 함수는 비중앙 샘플 내 동일한 유전자좌의 대립유전자 범위에 대비한 중앙 샘플 내 유전자좌의 대립유전자 범위의 복수의 비에 대해(예컨대, 로그비, 예를 들어 로그<sub>2</sub>비) 결정된다. 일부 구현예들에서, 상기 복수의 대립유전자 분획에 대한 확률 분포 함수는 하기와 같이 정의된다:

$$P\left(\log_n\left(\frac{Cvg_{i,a}^{\text{모}} + Cvg_{i,b}^{\text{모}}}{Cvg_{i,a}^{\text{부}} + Cvg_{i,b}^{\text{부}}}\right)\right)$$

[0107]

여기서 로그 <sub>$n$</sub> 은 염기  $n$ 에서의 로그이고,  $Cvg_{i,a}^{\text{모}}$ 은 상기 중앙 샘플 내 유전자좌  $i$ 에서 모계 대립유전자의 범위이고,  $Cvg_{i,b}^{\text{모}}$ 은 상기 중앙 샘플 내 유전자좌  $i$ 에서 부계 대립유전자의 범위이고,  $Cvg_{i,a}^{\text{부}}$ 은 상기 비중앙 샘플 내 유전자좌  $i$ 에서 모계 대립유전자의 범위이며,  $Cvg_{i,b}^{\text{부}}$ 은 상기 비중앙 샘플 내 유전자좌  $i$ 에서 부계 대립유전자의 범위이다. 일부 구현예들에서, 상기 로그비는 로그<sub>2</sub>비이다. 예를 들어, 일부 구현예들에서, 상기 복수의 대립유전자 분획에 대한 확률 분포 함수는 하기와 같이 정의된다:

$$P\left(\log_2\left(\frac{Cvg_{i,a}^{\text{모}} + Cvg_{i,b}^{\text{모}}}{Cvg_{i,a}^{\text{부}} + Cvg_{i,b}^{\text{부}}}\right)\right)$$

[0109]

여기서  $Cvg_{i,a}^{\text{모}}$ 은 상기 중앙 샘플 내 유전자좌  $i$ 에서 모계 대립유전자의 범위이고,  $Cvg_{i,b}^{\text{모}}$ 은 상기 중앙 샘플 내 유전자좌  $i$ 에서 부계 대립유전자의 범위이고,  $Cvg_{i,a}^{\text{부}}$ 은 상기 비중앙 샘플 내 유전자좌  $i$ 에서 모계 대립유전자의 범위이며,  $Cvg_{i,b}^{\text{부}}$ 은 상기 비중앙 샘플 내 유전자좌  $i$ 에서 부계 대립유전자의 범위이다.

[0111]

분산(또는 확실성 메트릭)은, 예를 들어 대응하는 유전자좌에 걸친 기대값으로부터 복수의 값 내의 각 값의 편차일 수 있다. 상기 기대값은 상기 중앙 샘플이 비중앙(예컨대, 건강한) 샘플인 경우 예상되는 값이다. 일부 구

현예들에서, 상기 확실성 메트릭은 상기 기대값으로부터의 평균 제곱근 편차이다. 예를 들어, 일부 구현예들에서, 상기 확실성 메트릭은 하기에 의해 정의되는 평균 제곱근 편차(RMSD)이다:

$$RMSD = \sqrt{\frac{1}{N} \sum_i \left[ \left( \log_2 \frac{(Cvg_{i,a}^{af} + Cvg_{i,b}^{af})}{(Cvg_{i,a}^{af,af} + Cvg_{i,b}^{af,af})} \right) - 0 \right]^2}$$

일부 구현예들에서, 상기 대립유전자 분획을 나타내는 값은 상기 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위에 대비하여, 상기 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위에 대비하여 상기 종양 샘플 내 유전자좌의 대립유전자 범위 간의 차이의 비이다. 따라서, 상기 RMSD는 하기와 같이 정의될 수 있다:

$$RMSD = \left[ \frac{1}{N} \sum_{i=0}^N \left( \frac{(Cvg_{i,a}^{af} + Cvg_{i,b}^{af}) - (Cvg_{i,a}^{af,af} + Cvg_{i,b}^{af,af})}{(Cvg_{i,a}^{af,af} + Cvg_{i,b}^{af,af})} \right)^2 \right]^{(1/2)}$$

일부 구현예들에서, 확률 분포(예컨대, 확률 분포 함수)는 종양 샘플 내 유전자좌의 대립유전자 범위 및 비종양 샘플 내 동일한 유전자좌의 대립유전자 범위 사이의 차이를 나타내는 복수의 값에 대해 결정될 수 있다. 상기 확실성 메트릭(예컨대, 분산)은 확률 분포의 메트릭, 예컨대 확률 분포의 엔트로피일 수 있다. 예를 들어, 일부 구현예들에서, 대립유전자 분획 확률 분포 함수( $S[P(af)]$ )의 엔트로피는 하기와 같이 정의될 수 있다:

$$S[P(lnr)] = \sum_{lnr=0} P(lnr) \log_n(P(lnr))$$

여기서:

$$P(lnr) = P \left( \log_n \left( \frac{Cvg_{i,a}^{af} + Cvg_{i,b}^{af}}{Cvg_{i,a}^{af,af} + Cvg_{i,b}^{af,af}} \right) \right)$$

여기서  $\log_n$ 은 염기  $n$ 을 갖는 로그이고,  $Cvg_{i,a}^{af}$ 은 상기 종양 샘플 내 유전자좌  $i$ 에서 모계 대립유전자의 범위이고,  $Cvg_{i,b}^{af}$ 은 상기 종양 샘플 내 유전자좌  $i$ 에서 부계 대립유전자의 범위이고,  $Cvg_{i,a}^{af,af}$ 은 상기 비종양 샘플 내 유전자좌  $i$ 에서 모계 대립유전자의 범위이며,  $Cvg_{i,b}^{af,af}$ 은 상기 비종양 샘플 내 유전자좌  $i$ 에서 부계 대립유전자의 범위이다. 일부 구현예들에서, 상기 로그 염기는 2(즉,  $\log_2$ )이다. 따라서, 일부 구현예들에서, 대립유전자 분획 확률 분포 함수( $S[P(af)]$ )의 엔트로피는 하기와 같이 정의될 수 있다:

$$S[P(l2r)] = \sum_{l2r=0} P(l2r) \log_2(P(l2r))$$

여기서:

$$P(l2r) = P \left( \log_2 \left( \frac{Cvg_{i,a}^{af} + Cvg_{i,b}^{af}}{Cvg_{i,a}^{af,af} + Cvg_{i,b}^{af,af}} \right) \right)$$

여기서  $Cvg_{i,a}^{af}$ 은 상기 종양 샘플 내 유전자좌  $i$ 에서 모계 대립유전자의 범위이고,  $Cvg_{i,b}^{af}$ 은 상기 종양 샘플 내 유전자좌  $i$ 에서 부계 대립유전자의 범위이고,  $Cvg_{i,a}^{af,af}$ 은 상기 비종양 샘플 내 유전자좌  $i$ 에서 모계 대립유전자의 범위이고,  $Cvg_{i,b}^{af,af}$ 은 상기 비종양 샘플 내 유전자좌  $i$ 에서 부계 대립유전자의 범위이다.

의 범위이며,  $Cvg_{ib}^{정상}$  은 상기 비종양 샘플 내 유전자좌  $i$ 에서 부계 대립유전자의 범위이다.

- [0124] 하나 이상의 저장된 확실성 메트릭 및 하나 이상의 저장된 종양 분획 사이의 관계는 상기 결정된 확실성 메트릭에 기초하여 종양 분획을 결정하는 데 사용될 수 있다. 일부 구현예들에서, 모델은 확실성 메트릭 및 종양 분획 간의 관계를 결정하기 위해 훈련 확실성 메트릭 및 관련된 종양 분획을 포함하는 훈련 데이터셋을 사용하도록 훈련된다. 상기 훈련 데이터셋은, 예를 들어 공지된(즉, 훈련) 종양 분획으로 복수의 임상 샘플을 사용하여 결정될 수 있다(예를 들어, 최대 체세포 대립유전자 빈도(MSAF)에 의해 결정되는 바와 같이, 종양 샘플 내 모든 호출로부터의 생식선 변이체 호출을 여과하고 남은 변이체(즉, 최대 체세포 변이체)를 총 변이체(최대 체세포 변이체 및 생식선 변이체)와 비교하여 상기 최대 체세포 대립 유전자 빈도를 결정한다). 임상 샘플 내의 핵산 분자는 복수의 유전자좌에 걸친 대립유전자 빈도(또는 대립유전자 빈도를 나타내는 값) 뿐만 아니라 연관된 훈련 확실성 메트릭을 결정하기 위해 서열분석될 수 있다. 상기 훈련 확실성 메트릭은 훈련 종양 분획과 상관관계를 보여 확실성 메트릭 및 종양 분획 사이의 관계를 결정할 수 있다. 다른 방법에서, 연속 회색은 하나 이상의 임상 샘플로부터 만들어져 복수의 상이한 종양 분획을 수득할 수 있으며, 이는 연속 회색 샘플에 대한 확실성 메트릭과 상관되어 관계를 결정할 수 있다.
- [0125] 일부 구현예들에서, 종양 분획을 결정(예컨대, 추정)하기 위해 훈련 하위 공정이 먼저 실시된다. 데이터셋은 임상 표본으로부터 구성할 수 있다. 훈련 세트 및 상기 훈련 세트의 인실리코 회색을 사용하여, 종양 분획은 종양에서 일반적으로 관찰되는 이수성에 해당하는 대립유전자 분획 또는 로그비의 변화와 상관될 수 있다. 다른 예들에서, 세포주/임상 샘플 회색이 실시될 수 있다.
- [0126] 일부 구현예들에서, 상기 확실성 메트릭은 특정 대립유전자 및/또는 대립유전자 빈도(예컨대, 0 내지 0.5의 범위에서)에 대한 특정 SNP 빈에서의 범위의 함수일 수 있다. 일부 예들에서, 상기 훈련 데이터는 편차 메트릭(예컨대, 대립유전자 분획 편차 또는 로그비 편차)을 입력으로 사용하고 하한 및 상한과 함께 추정된 종양 분획을 반환한다. 0 및 1 사이에서 벗어나고 0.5가 아닌 값(배타적)은 "잡음"으로 간주될 수 있으며, 평균 잡음은 예상되거나 추정된 종양 비율과 상관관계가 있을 수 있다. 다른 예들에서, 상기 훈련 데이터는 로그비 편차 메트릭, 또는 일반적으로 기대치로부터의 범위 편차를 정량화하는 임의의 메트릭을 입력으로 제공한다. 어느 경우든, 대립유전자 범위 편차 메트릭 또는 로그비 편차 메트릭은 종양 분획의 척도일 수 있다.
- [0127] 훈련 중에 도출된 이러한 상관 관계를 활용하여, 환자의 종양 분획을 상한 및 하한으로 추정하거나 평가할 수 있다. SNP 대립유전자 범위 변동 메트릭과 같은 범위 메트릭은 상관관계를 생성하는데 사용될 수 있다.
- [0128] 본원에 기재된 방법은, 예를 들어 종양이 생물학적 샘플에 존재하는지를 확인하는 능력을 개선하고 공지된 추정 한도로 종양 분획 결정(예컨대, 추정)을 제공할 수 있고; 체세포 변이체를 평가하기 위한 체계적이고 직교적인 접근법을 제공하며; 새로운 저렴한 종양 추적/식별 분석을 위한 틀을 제공한다.
- [0129] 일부 구현예들에서, 본원에 기재된 방법은 또한 액체 생검의 특정 경우에 있어서 이점을 제공한다(비록 본 개시는 액체 생검으로 제한되지 않음). 고형 종양에는 병리학 검토, 체세포 대립유전자 빈도(MSAF) 및 분석적 복제 수 변경(CNA) 모델링을 포함하여 종양 함량을 추정하기 위한 여러 상이한 수단이 있다. 하지만, 액체 생검은 일반적으로 이러한 방법에 적합하지 않거나 상당한 재조정을 필요로 한다. 무세포 DNA는 혈액 내에 자유롭게 떠 있기 때문에 이의 존재는 나노 수준이므로 병리학자가 검토할 수 없다. 또한, 종양이 혈류로 흘러들어가는 경향이 있는 DNA의 양은 정상 DNA에 비해 미미할 수 있다. 따라서, 분석 CNA 모델링은 낮은 종양 함량으로 인해 실패할 수 있다.
- [0130] 본원에 기재된 방법은 일반적으로 병리학 검토를 요구하지 않고; 충분히 민감하고 분석 방정식을 쓰지 않기 때문에 종양의 존재 또는 함량을 식별하기 위해 분석적 CNA 모델링을 필요로 하지 않고; 짧은 변종 호출과 무관하게 짧은 변종에 대한 직교 평가를 제공하며; CNA 이벤트가 있을 때 개선된다(예컨대, 혼동하지 않는다).
- [0131] 본원에 기재된 방법은 새로운 저렴한 종양 추적(예컨대, 모니터링) 분석의 개발을 가능하게 한다. 예를 들어, 환자가 충분한 수의 하위 유전체 간격(예컨대, 하나 이상의 SNP를 포함하는 하위 유전체 간격)을 포함하는 분석(예컨대, 포괄적인 분석)에서 종양 함량을 나타내는 경우, 상기 방법은 SNP 변형에만 기초할 수 있기 때문에 종양 진행은 상당히 적은 비용을 위해 제2 분석으로 시간 경과에 따라 추적될 수 있다. 일부 구현예들에서, 제1 분석은 제2 분석보다 더 많은 하위 유전체 간격을 포함한다. 다른 구현예들에서, 제1 분석은 제2 분석보다 더 적은 하위 유전체 간격을 포함한다. 특정 구현예들에서, 제1 분석 및 제2 분석은 본질적으로 동일한 수의 하위 유전체 간격을 포함한다.

- [0132] 제1 및 제2 분석에 포함된 유전자 패널은 크기가 동일하거나 상이할 수 있다. 예를 들어, 적어도 약 100, 150, 200, 250, 300, 350, 400, 450, 500개 이상의 유전자 패널을 포함하는 분석은 대형 패널로 간주될 수 있으며, 약 100, 90, 80, 70, 60, 50, 40, 30, 20 또는 10개 미만의 유전자를 포함하는 분석은 소형 패널로 간주될 수 있다. "대형" 및 "소형" 패널 크기는 일반적으로 분석의 목적에 따라 결정되며 상기 예시적인 크기로 제한되어서는 안 된다. 일부 구현예들에서, 상기 제1 분석은 대형 패널을 포함하고 상기 제2 분석은 동일하거나 상이한 대형 패널을 포함한다. 다른 구현예들에서, 상기 제1 분석은 소형 패널을 포함하고 상기 제2 분석은 동일하거나 상이한 소형 패널을 포함한다. 특정 구현예들에서, 상기 제1 분석은 대형 패널을 포함하고 상기 제2 분석은 소형 패널을 포함하거나, 또는 그 반대이다. 상기 제1 및 제2 분석은 동일한 분석 유형일 필요는 없다. 예를 들어, 상기 제1 분석은 서열분석(예컨대, NGS)을 기반으로 할 수 있고 상기 제2 분석은 유전체 혼성화를 기반으로 할 수 있거나, 또는 그 반대이다.
- [0133] 일부 구현예들에서, 상기 제2 분석에 의해 포함되는 하위 유전체 간격은 상기 제1 분석에 의해 포함되는 하위 유전체 간격의 하위 집합일 수 있다. 일부 구현예들에서, 상기 제2 분석에 의해 포함되는 하위 유전체 간격은 상기 제1 분석에 의해 포함되는 하위 유전체 간격의 하위 집합일 수 있다. 다른 구현예에서, 제2 분석에 의해 포함된 하위 유전체 간격은 제1 분석에 의해 포함된 하위 유전체 간격과 중첩되지만 동일하지는 않다. 특정 구현예에서, 제1 분석은 제2 분석에서 다루지 않는 하나 이상의 하위 유전체 간격을 포함한다.
- [0134] 특정 구현예에서, 제1 분석은 제1 분석에서 다루지 않는 하나 이상의 하위 유전체 간격을 포함한다.
- [0135] 일부 구현예에서, 추정된 종양 분획이 환자에 걸쳐 넓은 오차 한계를 가질 수 있음에도 불구하고, 임의의 환자 내 비교는 작은 오차 한계를 제공하여 포괄적 검정(예컨대, FoundationOne, FoundationOne CDx 또는 FoundationOne Liquid 분석). 두 번째 분석은 종합 분석보다 훨씬 저렴할 수 있기 때문에 환자가 암에 걸렸는지 여부에 대한 질문에 답하기 위해 위험에 처한 환자와 같은 환자의 하위 집합에 대한 표준 스크리닝 기술로 사용할 수 있습니다.
- [0136] 도 1은 샘플로부터 종양 분획을 추정하는 방법(100)을 도시한다. 방법(100)은 단계(102)에서 시작한다. 단계(104)에서, 하위 유전체 간격과 연관된 표적 변수에 대한 값이 예를 들어, 대상체로부터의 샘플로부터 직접 획득된다. 상기 목표 변수는, 예를 들어 대립유전자 분획일 수 있다. 상기 샘플은, 예컨대 액체 샘플 또는 고체 샘플일 수 있다.
- [0137] 일부 예시들에서, 적어도 하나의 이형접합 단일 염기 다형성(SNP) 부위에 대한 환자 대립유전자 분획은 환자에서 채취한 생검으로부터 결정된다. 일례에서, 상기 생검은 액체 생검, 즉 비교형 생물학적 조직, 예를 들어 혈액의 샘플일 수 있다. 하지만, 본 개시는 이에 제한되지 않으며 제한 없이 임의의 고체 또는 액체 분석 또는 생검을 포함하도록 의도된다. 한 구현예에서, 상기 액체 생검은 혈액 샘플을 포함한다. 한 구현예에서, 상기 액체 생검은 무세포 DNA(cfDNA)를 포함한다. 한 구현예에서, 상기 액체 생검은 순환 종양 DNA(ctDNA)를 포함한다. 한 구현예에서, 상기 액체 생검은 종양으로부터 유출된 DNA를 포함한다. 한 구현예에서, 상기 액체 생검은 DNA 이외의 핵산, 예컨대 RNA를 포함한다. 한 구현예에서, 상기 액체 생검은 순환 종양 세포(CTC)를 포함한다. 다른 유형의 액체 생검이, 예컨대 Crowley et al. Nat Rev Clin Oncol. 2013; 10(8): 472-484에 기재되어 있으며, 그 내용은 전체가 참조로서 인용된다.
- [0138] 단계(106)에서, 확실성 메트릭이 목표 변수로부터 결정될 수 있고, 단계(108)에서, 저장된 확실성 메트릭 및 저장된 종양 분획 사이에 결정된 관계가 액세스된다. 상기 결정된 관계는 적어도 하나의 이형접합 SNP 부위에 대한 확실성 메트릭(예컨대, 샘플링된 대립유전자 분획 편차)을 대응하는 샘플링된 종양 분획과 관련시키는 과거 샘플 데이터(환자 또는 다른 시험 대상체로부터 수집됨)를 포함할 수 있다. 일부 예들에서, 샘플링된 대립유전자 범위 편차는 대립유전자 분획이 기대값으로부터 가변되는 정도를 반영하는 "잡음" 메트릭이다. 일부 예들에서, 대립유전자 분획으로부터 계산된 잡음 메트릭과 종양 분획의 상관관계를 보여주는 데이터 포인트의 수는 백(100), 천(1,000), 만(10,000) 또는 그 이상을 초과할 수 있다.
- [0139] 일례에서, 상기 결정된 관계는 인실리코(*in silico*) 공정으로부터 도출될 수 있고, 분석은 기계 학습 공정에 의해 실시될 수 있다. 상기 공정은 하나 이상의 하위 유전체 간격(예컨대, SNP, SNP 빈 및/또는 염색체)에 걸쳐 하나 이상의 범위 편차 메트릭(예컨대, 대립유전자 분획 값)의 상관관계를 보여주기 위해 특정 종양 분획에서 시작하여 샘플 회식(예컨대, 일치된 정상 사용)을 실시할 수 있다. 상기 메트릭은 종양 분획이 0 또는 1의 값 사이에 속하는 빈도 및 정도의 척도일 수 있다. 0 및 1(배타적) 사이의 평균적인 "잡음" 메트릭은 예상되거나 추정된 종양 분획과 상관 관계가 있을 수 있다.

- [0140] 종양 분획과 상관관계가 있는 확실성 메트릭 값의 계산에 기여하는 하위 유전체 간격과 관련된 요소들의 수는 십(10), 일백(100), 일천(1,000), 일만(10,000) 또는 그 이상의 순서일 수 있다.
- [0141] 상관관계에서의 확실성 메트릭 계산에 기여하는 하위 유전체 간격과 관련된 많은 요소들로 인해, 일부 예들에서 상기 요소들은 하위 유전체 간격의 위치 또는 다른 특성들에 의해 "비닝"되거나 집계될 수 있다. 비닝은 추정된 종양 분획에 부정적인 영향을 미치는 확실성 메트릭에서의 상관 관계에 불균형적으로 가중치를 주는 단일(또는 작은 집합의) 요소(들)를 피할 수 있다. 예를 들어, 단일 하위 유전체 간격에 있는 한 요소가 5,000개 복제를 갖는 복제 변이를 나타내는 경우, 추정된 종양 분획이 부정확하게 높을 수 있다. 따라서, 일부 예들에서, 확실성 메트릭에 기여하는 요소들은, 예를 들어 22개의 관련 염색체 각각에 대해 염색체에 의해 평균화되거나 그렇지 않으면 집계된다. 그런 다음, 22개의 집계된 염색체 값을 사용하여 종양 분획과 상관관계를 보이는 확실성 메트릭을 계산하여, 단일 하위 유전체 간격(예컨대, SNP 부위)이 상기 상관 관계에 불균형적으로 영향을 미치지 않도록 한다. 이상값 요소가 확실성 메트릭 계산에 포함되는 것을 방지하는 것을 포함하지만 이에 제한되지는 않는 극단적인 복제 수 이벤트의 영향을 제한하기 위해 다른 방법들이 사용될 수 있다.
- [0142] 일부 예들에서, 상기 상관관계는 상한 및 하한 상관관계도 마찬가지로 계산된 평균(즉, 보통)의 상관관계일 수 있다. 상기 방식으로, 평균의 상관 관계는 95% 신뢰 구간으로 제한된다.
- [0143] 하위 유전체 간격은 하나 또는 여러 개의 하위 유전체 간격을 포함할 수 있고, 일부 예들에서는 적어도 하나의 이형접합 SNP 부위일 수 있다. 하위 유전체 간격은 다양한 기준에 따라 선택될 수 있다. 예를 들어, 하위 유전체 간격은 상기 하위 유전체 간격이 일반적인 건강한 집단뿐만 아니라 건강한 하위집단(상이한 성별, 연령 또는 민족적 배경을 포함함)에서 얼마나 다형성인지를 기반으로 선택될 수 있다. 상기 하위 유전체 간격이 건강한 집단에서 상당히 가변되는 것이 유리할 수 있다. 상기 하위 유전체 간격의 서열분석 특성은 "양호하게", 즉 0, 0.5 및 1.0과 같이 예상되는 대립유전자 빈도에 가깝게 되는 것을 기반으로 선택될 수도 있다. 또한, 영역은 "충분히 포함되는", 즉 해당 부위의 집단 전체에 걸쳐 일반적인 범위를 갖는 것을 기반으로 선택될 수 있다. 하위 유전체 간격은 유전자군의 단순 반복 또는 일반적으로 반복되는 DNA 서열에서 발생하는 경우 상기 특성이 정렬 방법론에 위배될 수 있기 때문에 제외될 수 있다. 한 구현예에서, 하위 유전체 간격은 높은 동일성, 단순 반복 또는 유전자군이 없거나 본질적으로 없는 유전체 영역에 위치할 수 있다.
- [0144] 한 구현예에서, 상기 하위 유전체 간격은 소수 대립유전자를 포함한다. 본원에서 사용된 "소수 대립유전자"는 주어진 집단에서 특정 하위 유전체 간격과 관련된 가장 흔한 대립유전자(예컨대, 두 번째로 가장 흔한 대립유전자 또는 가장 덜 흔한 대립유전자) 이외의 대립유전자이다. 한 구현예에서, 적어도 10, 20, 50, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 또는 10000개의 이형접합 하위 유전체 간격이 선택된다. 일례에서, 10, 20, 50, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 10000개 이하의 이형접합 SNP 부위가 선택된다.
- [0145] 일례에서, 선택된 하위 유전체 간격 및/또는 상관관계는 광범위한 선별검사 기술을 제공하기 위해, 즉 모든 질병 온톨로지에 걸쳐 보편적일 수 있다. 다른 예들에서, 하위 유전체 하위 유전체 간격이 선택될 수 있고, 질병 온톨로지(예를 들어, 종양 유형)에 기초하여 상관관계가 조정될 수 있다.
- [0146] 목표 변수(예컨대, 대립유전자 범위 편차 및/또는 대립유전자 분획 변동)를 종양 분획과 상관시키는 데 하나 이상의 확실성 메트릭이 사용될 수 있다. 예를 들어, 대립유전자 분획과 관련된 메트릭이 적용될 수 있다. 일례에서, 대립유전자 빈도 엔트로피 메트릭 또는 평균 제공근 편차(RMSD) 메트릭이 사용될 수 있다:
- [0147] 대립유전자 빈도 엔트로피:
- $$S[P(af)] = - \sum_{af=0}^{0.5} P(af) \log_2(P(af))$$
- [0148]
- [0149] 평균 제공근 편차:
- $$RMSD = \left[ \frac{1}{N} \sum_{i=0}^N (af^i - 0.5)^2 \right]^{(1/2)}$$
- [0150]
- [0151] 여기서  $i$  = SNP 빈 및 af 0 내지 0.5 범위의 대립유전자 빈도. 본원에서는 점진 SNP 대립유전자 빈도가 관례에 따라 사용되지만(예컨대, Nielsen. Hum Genomics. 2004; 1(3): 218-224 and Marth et al. Genetics. 2004; 6(1): 351-372에 기재된 바와 같이), 상기 방법론은 0 내지 1의 전체 범위가 활용되는 경우 유지된다. 로그2비

를 기반으로 하는 메트릭과 같은 다른 메트릭들도 사용할 수 있다. 상기 메트릭 중 임의의 것은 특정 SNP 빈에서의 범위와 같은 요인들을 통합할 수 있으며, 상기 "빈"은 1개 이상의 염기쌍으로 정의될 수 있다. 일부 구현예들에서, 상기 확실성 메트릭은  $\text{확실성\_메트릭} = f(Cvg)$ 가 되도록 범위의 함수로서 기록될 수 있다. 또한, 상기 확실성\_메트릭에 작용하는 모든 수학적 변환 또는 연산도 확실성\_메트릭으로 간주될 수 있다.

[0152] 일부 예들에서, 상기 확실성 메트릭은 적어도 하나의 하위 유전체 간격에 대한 예상 로그2비로부터의 편차일 수 있다. 다른 예들에서, 상기 확실성 메트릭은 이형접합인 것으로 공지된 적어도 하나의 하위 유전체 간격(예컨대, SNP)에 대한 건강한 집단의 예상되는 대립유전자 분획으로부터의 편차일 수 있다. 다른 예들에서, 상기 확실성 메트릭은 이형접합인 것으로 공지된 적어도 하나의 하위 유전체 간격(예컨대, SNP)에 대한 상기 건강한 집단의 예상되는 대립유전자 범위로부터의 편차일 수 있다.

[0153] 표 1은 임의의 p 모멘트 또는 이들의 조합을 포함하여 사용될 수 있는 예시적인 확실성 메트릭을 도시한다.

표 1

| 메트릭과 중앙 분획을 관련시키는 조건                                                                                                                                                                                                                                                                                                                                     | 변수의 상대적 확실성을 계산하는 메트릭                                                                                                                                                                                                                                              | 비고                                                                               |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|
| <p>샘플 내-비교</p> <p>고려 중인 모든 범위는 <math>af</math> 샘플에서 가져온 것이다. <math>af</math>는 모성 및 부계 염색체의 범위를 비교하는 메트릭이다.</p> <p><math>Cvg_a</math>은 모계 또는 부계이고, <math>Cvg_b</math>은 다른 대립유전자이다.</p> <p><math>N</math>은 하위 유전체 간격의 수이고 <math>i</math>는 <math>N</math>의 지수이다.</p> <p>생식계열 모계 및 부계 염색체가 동일한 유전체 위치에서 상이하도록 임의의 유전자좌가 사용될 수 있으며; SNP 또는 더 큰 것이 될 수 있다.</p> | <p>변수 = <math>af</math></p> $af = \frac{Cvg_a}{Cvg_a + Cvg_b}$ <p><math>af_{avg} = 0.5</math></p> $\text{확실성\_메트릭} = \sqrt{\frac{1}{N} \sum_i^N (af - af_{avg})^2}$ $= \sqrt{\frac{1}{N} \sum_i^N \left( \frac{Cvg_{i,a}}{Cvg_{i,a} + Cvg_{i,b}} - 0.5 \right)^2}$ | <p>샘플내 비교</p> <p><math>Cvg_a &lt; Cvg_b</math>, 따라서 <math>af \leq 0.5</math></p> |

|                                                                                                                                                                                                                                            |                                                                                                                                                                                                                          |                                                                                                                                                                                                                  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>고려 중인 모든 범위는 암 샘플에서 가져온 것이다.</p> <p><math>Cvg_a</math>은 모계 또는 부계이고, <math>Cvg_b</math>은 다른 대립유전자이다.</p> <p>생식계열 모계 및 부계 염색체가 동일한 유전체 위치에서 상이하도 록 임의의 유전자좌가 사용될 수 있으며; SNP 또는 더 큰 것이 될 수 있다.</p>                                         | <p>변수 = <math>Cvg</math></p> <p>상대_차이 = <math>\frac{Cvg_a - Cvg_b}{Cvg_b}</math></p> <p>상대_차이_정상 = 0</p> <p>확실성_메트릭 = <math>\sqrt{\frac{1}{N} \sum_i^N \left( \frac{Cvg_{ia} - Cvg_{ib}}{Cvg_{ib}} \right)^2}</math></p> | <p>샘플내 비교</p> <p><math>Cvg_a &gt; Cvg_b</math></p> <p>대립유전자가 증폭되어 "b" 대립유전자가 정상이라고 가정하기 때문에 이전의 선택이 선택된다.</p> <p>일반적으로, 삭제로 전환되므로 문제가 되지 않는다. 따라서, <math>Cvg_a &lt; Cvg_b</math>이거나, 또는 상기 전환을 완전히 무시할 수 있다.</p> |
| <p>고려 중인 모든 범위는 암 샘플에서 가져온 것이다.</p> <p>af는 모성 및 부계 염색체의 범위를 비교하는 메트릭이다.</p> <p><math>Cvg_a</math>은 모계 또는 부계이고, <math>Cvg_b</math>은 다른 대립유전자이다.</p> <p>생식계열 모계 및 부계 염색체가 동일한 유전체 위치에서 상이하도 록 임의의 유전자좌가 사용될 수 있으며; SNP 또는 더 큰 것이 될 수 있다.</p> | <p>변수 = <math>af</math></p> <p>대립유전자 빈도의 확률 = <math>P(af)</math></p> <p><math>= P\left(\frac{Cvg_{i,a}}{Cvg_{i,a} + Cvg_{i,b}}\right)</math></p> <p>확실성_메트릭 = <math>S[P(af)]</math></p>                                  | <p>샘플내 비교</p> <p><math>S</math> = 엔트로피는 본질적으로 변수의 상대적 확실성을 측정하는 메트릭이다.</p>                                                                                                                                       |

|                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                        |        |
|------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|
| 샘플 내-<br>비교                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                        |        |
| <p>고려 중인 모든 범위는 암 및 참조 샘플 모두에서 가져온 것이다. 비는 참조의 모계 및 부계에 대한 암 샘플의 모계 및 부계의 총 범위에서 결정된다.</p> <p>1에서 무한 염기 집합에 이르기까지 유전체의 임의의 유전자좌를 사용할 수 있다.</p> | <p>변수 = 로그2비 = <math>l2r</math></p> $\mu = \frac{(Cvg_a^{\text{암}} + Cvg_b^{\text{암}})}{(Cvg_a^{\text{정상}} + Cvg_b^{\text{정상}})}$ $l2r = \log_2(\mu)$ $l2r_{\text{정상}} = 0$ $\text{확실성\_메트릭} = \sqrt{\frac{1}{N} \sum_i^N [l2r - l2r_{\text{정상}}]^2}$ $= \sqrt{\frac{1}{N} \sum_i^N \left[ \log_2 \left( \frac{(Cvg_{i,a}^{\text{암}} + Cvg_{i,b}^{\text{암}})}{(Cvg_{i,a}^{\text{정상}} + Cvg_{i,b}^{\text{정상}})} \right) - 0 \right]^2}$ | 샘플내 비교 |
| <p>고려 중인 모든 범위는 암 및 참조 샘플 모두에서 가져온 것이다. 비는 참조의 모계 및 부계에 대한 암 샘플의 모계 및 부계의 총 범위에서 결정된다.</p> <p>1에서 무한 염기 집합에 이르기까지 유전체의 임의의 유전자좌를 사용할 수 있다.</p> | <p>변수 = <math>Cvg</math></p> $Cvg_{\text{암}} = (Cvg_a^{\text{암}} + Cvg_b^{\text{암}})$ $Cvg_{\text{정상}} = (Cvg_a^{\text{정상}} + Cvg_b^{\text{정상}})$ $\text{상대\_차이} = \frac{Cvg_{\text{암}} - Cvg_{\text{정상}}}{Cvg_{\text{정상}}}$ $\text{상대\_차이}_{\text{정상}} = 0$ $\text{확실성\_메트릭} = \sqrt{\frac{1}{N} \sum_i^N \left( \frac{Cvg_{i,\text{암}} - Cvg_{i,\text{정상}}}{Cvg_{i,\text{정상}}} \right)^2}$                                           | 샘플내 비교 |

|                                                                                             |                                                                                                                                                                                                                                                                                                 |                                                                                                          |
|---------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|
| <p>암 샘플의 모계 및 부계의 총 범위는 참조의 모계 및 부계와 비교한다. 1에서 무한 염기 집합에 이르기까지 유전체의 임의의 유전자좌를 사용할 수 있다.</p> | <p>변수 = 로그2비 = <math>l2r</math></p> <p><b>로그의 확률2 비 = <math>P(l2r)</math></b></p> $= P \left( \log_2 \left( \frac{Cvg_{i,a}^{\frac{a}{a+b}} + Cvg_{i,b}^{\frac{b}{a+b}}}{Cvg_{i,a}^{\frac{a}{a+b}} + Cvg_{i,b}^{\frac{b}{a+b}}} \right) \right)$ <p><b>확실성 메트릭 = <math>S[P(l2r)]</math></b></p> | <p>샘플내 비교</p> <p><math>S</math> = 엔트로피는 기대치로부터의 편차를 계산하지 않는 메트릭이다. 이는 본질적으로 변수의 상대적 확실성을 측정하는 메트릭이다.</p> |
|---------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|

- [0155] 단계(110)에서, 샘플의 중앙 분획은 확실성 메트릭 및 결정된 관계를 참조하여 결정(예컨대, 추정)된다. 일부 예들에서, 상기 결정된 관계의 계수는 환자 샘플, 및 평가된(예컨대, 추정된) 중앙 분획에 도달하도록 합산된 결과물로부터 결정된 확실성 메트릭에 적용된다. 최종 추정된 중앙 분획을 산출하기 위해 다른 기능들이 실시될 수 있음이 이해될 것이다. 예를 들어, 상기 추정된 중앙 분획은 초기 또는 원시 추정된 중앙 분획 측정치로부터 축적되거나, 정규화되거나, 그렇지 않으면 조정될 수 있다.
- [0156] 단계(112)에서, 방법(100)은 종료된다.
- [0157] 상기 추정된 중앙 분획은 의료인에 의해 여러 가지 방법으로 사용할 수 있다. 예를 들어, 상기 추정된 중앙 분획은 하나 이상의 유형의 암에 대한 위험이 있는 환자를 모니터링하는 데 사용할 수 있다. 상기 추정된 중앙 분획은 또한 암을 진단하거나 또는 암 치료가 중앙에 성공적으로 영향을 미치는지 여부를 결정하는 데 사용할 수 있다.
- [0158] 상기 추정된 중앙 분획은 시험 결과를 확인하거나 검증하기 위해 다른 선별 검사 기술과 관련하여 사용할 수도 있다. 예를 들어, CNA 선별 검사는 환자, 특히 낮은 중앙 분획(예컨대, 30% 미만)을 갖는 환자에 대한 순도 및 배수성의 다수의 가능한 조합을 산출할 수 있다. 본 기술은 상기 결과를 명확하게 하는 데 사용할 수 있다.
- [0159] 일부 구현예들에서, 상기 추정된 중앙 분획을 포함하는 보고서가 생성될 수 있다. 한 구현예에서, 상기 보고서는 상기 추정된 중앙 분획에 기초한 치료 옵션을 추가로 포함한다. 한 구현예에서, 상기 보고서는 상기 추정된 중앙 분획에 기초한 예후를 추가로 포함한다.
- [0160] 중앙 치료 및 모니터링 방법
- [0161] 대상체에서 질병을 치료하는 방법이 또한 개시된다. 상기 방법은 중앙 분획(예컨대, 본원에 기재된 방법에 따라 결정됨)의 결정(예컨대, 추정)에 반응하여, 대상체에게 유효량의 요법을 투여함으로써 질환을 치료하는 단계를 포함하고, 상기 중앙 분획의 추정은 추정 중앙 분획은 상기 샘플에서 하위 유전체 간격과 관련된 목표 변수에 대한 값을 획득하는 단계; 상기 목표 변수로부터 확실성 메트릭을 결정하는 단계; 저장된 확실성 메트릭 및 저장된 중앙 분획 사이의 결정된 관계에 액세스하는 단계; 및 상기 확실성 메트릭 및 상기 결정된 관계를 참조하여 상기 샘플의 중앙 분획을 결정하는 단계를 포함한다.
- [0162] 한 구현예에서, 상기 방법은 상기 대상체에게 제2 요법을 투여하는 단계를 추가로 포함한다. 한 구현예에서, 상기 방법은 상기 대상체에게 제2 요법을 중지하는 단계를 추가로 포함한다. 한 구현예에서, 상기 방법은 상기 대상체에서 체세포 변경(예컨대, 질병과 관련된 체세포 변경)의 존재를 결정하는 단계를 추가로 포함한다.
- [0163] 한 구현예에서, 상기 대립유전자 분획은 서열분석, 예컨대 차세대 서열분석(NGS)을 포함하는 방법에 의해 결정된다. 한 구현예에서, 상기 대립유전자 분획은 표적 선택을 추가로 포함하는 방법에 의해, 예컨대 용액 혼성화에 의해 결정된다. 다른 구현예들에서, DNA(예컨대, cfDNA, ctDNA 등)를 검출하기 위해 사용되는 다른 방법론들, 예컨대 미세배열이 사용될 수 있다.
- [0164] 대상체의 질병을 평가하는 방법이 또한 기재되며, 중앙 분획(예컨대, 본원에 기재된 방법에 따라 결정됨)의 결정(예컨대, 추정)은 상기 샘플에서 하위 유전체 간격과 관련된 목표 변수에 대한 값을 획득하는 단계; 상기 목

표 변수로부터 확실성 메트릭을 결정하는 단계; 저장된 확실성 메트릭 및 저장된 종양 분획 사이의 결정된 관계에 액세스하는 단계; 및 상기 확실성 메트릭 및 상기 결정된 관계를 참조하여 상기 샘플의 종양 분획을 결정함으로써 상기 질병을 평가하는 단계를 포함한다. 한 구현예에서, 상기 대립유전자 분획은 서열분석, 예컨대 NGS를 포함하는 방법에 의해 결정된다. 한 구현예에서, 상기 대립유전자 분획은 표적 선택을 추가로 포함하는 방법에 의해, 예컨대 용액 혼성화에 의해 결정된다. 다른 구현예들에서, DNA(예컨대, cfDNA, ctDNA 등)를 검출하기 위해 사용되는 다른 방법론들, 예컨대 미세배열이 사용될 수 있다. 한 구현예에서, 상기 방법은 상기 질병에 대한 요법을 선택하는 단계를 추가로 포함한다. 한 구현예에서, 상기 방법은 상기 대상체에 요법을 중지하는 단계를 추가로 포함한다. 한 구현예에서, 상기 방법은 임상 시험을 위한 대상체를 선택하는 단계를 추가로 포함한다. 한 구현예에서, 상기 방법은 질병 상태, 예컨대 관해, 안정, 재발 등을 결정하는 단계를 추가로 포함한다. 한 구현예에서, 상기 질병은 주기적으로, 예컨대 매월, 2개월마다, 3개월마다, 6개월마다, 또는 매년 평가한다. 한 구현예에서, 상기 방법은 상기 대상체에서 체세포 변경(예컨대, 질병과 관련된 체세포 변경)의 존재를 결정하는 단계를 추가로 포함한다.

[0165] 대상체를 평가하는 방법이 또한 기재되며, 종양 분획(예컨대, 본원에 기재된 방법에 따라 결정됨)의 결정(예컨대, 추정)은 상기 샘플에서 하위 유전체 간격과 관련된 목표 변수에 대한 값을 획득하는 단계; 상기 목표 변수로부터 확실성 메트릭을 결정하는 단계; 저장된 확실성 메트릭 및 저장된 종양 분획 사이의 결정된 관계에 액세스하는 단계; 및 상기 확실성 메트릭 및 상기 결정된 관계를 참조하여 상기 샘플의 종양 분획을 결정함으로써 상기 대상체를 평가하는 단계를 포함한다. 한 구현예에서, 상기 대립유전자 분획은 서열분석, 예컨대 NGS를 포함하는 방법에 의해 결정된다. 한 구현예에서, 상기 대립유전자 분획은 표적 선택을 추가로 포함하는 방법에 의해, 예컨대 용액 혼성화에 의해 결정된다. 다른 구현예들에서, DNA(예컨대, cfDNA, ctDNA 등)를 검출하기 위해 사용되는 다른 방법론들, 예컨대 미세배열이 사용될 수 있다.

[0166] 한 구현예에서, 상기 방법은 요법을 위한 대상체를 선택하는 단계를 추가로 포함한다. 한 구현예에서, 상기 방법은 상기 대상체에 요법을 중지하는 단계를 추가로 포함한다. 한 구현예에서, 상기 방법은 임상 시험을 위한 대상체를 선택하는 단계를 추가로 포함한다.

[0167] 한 구현예에서, 상기 대상체는 주기적으로, 예컨대 매월, 2개월마다, 3개월마다, 6개월마다, 또는 매년 평가한다.

[0168] 한 구현예에서, 상기 방법은 상기 대상체에서 체세포 변경(예컨대, 질병과 관련된 체세포 변경)의 존재를 결정하는 단계를 추가로 포함한다.

[0169] 한 구현예에서, 상기 목표 변수(예컨대, 대립유전자 분획)는 서열분석, 예컨대 NGS를 포함하는 방법에 의해 결정된다. 한 구현예에서, 상기 대립유전자 분획은 표적 선택을 추가로 포함하는 방법에 의해, 예컨대 용액 혼성화에 의해 결정된다. 다른 구현예들에서, DNA(예컨대, cfDNA, ctDNA 등)를 검출하기 위해 사용되는 다른 방법론들, 예컨대 미세배열이 사용될 수 있다.

[0170] 요법을 평가하는 방법이 또한 기재되며, 종양 분획(예컨대, 본원에 기재된 방법에 따라 결정됨)의 결정(예컨대, 추정)은 상기 샘플에서 하위 유전체 간격과 관련된 목표 변수에 대한 값을 획득하는 단계; 상기 목표 변수로부터 확실성 메트릭을 결정하는 단계; 저장된 확실성 메트릭 및 저장된 종양 분획 사이의 결정된 관계에 액세스하는 단계; 및 상기 확실성 메트릭 및 상기 결정된 관계를 참조하여 상기 샘플의 종양 분획을 결정함으로써 상기 요법을 평가하는 단계를 포함한다.

[0171] 한 구현예에서, 상기 목표 변수(예컨대, 대립유전자 분획)는 서열분석, 예컨대 NGS를 포함하는 방법에 의해 결정된다. 한 구현예에서, 상기 대립유전자 분획은 표적 선택을 추가로 포함하는 방법에 의해, 예컨대 용액 혼성화에 의해 결정된다. 다른 구현예들에서, DNA(예컨대, cfDNA, ctDNA 등)를 검출하기 위해 사용되는 다른 방법론들, 예컨대 미세배열이 사용될 수 있다.

[0172] 한 구현예에서, 상기 방법은 상기 대상체에 대한 요법을 선택하는 단계를 추가로 포함한다.

[0173] 한 구현예에서, 상기 요법은 주기적으로, 예컨대 매월, 2개월마다, 3개월마다, 6개월마다, 또는 매년 평가한다.

[0174] 보고서를 제공하는 방법(예컨대, 본원에 기재된 방법에 따라 결정된 종양 분획을 보고하기 위해)이 기재된다. 상기 방법은 상기 샘플에서 하위 유전체 간격과 관련된 목표 변수에 대한 값을 획득하는 단계; 상기 목표 변수로부터 확실성 메트릭을 결정하는 단계; 저장된 확실성 메트릭 및 저장된 종양 분획 사이의 결정된 관계에 액세스하는 단계; 상기 확실성 메트릭 및 상기 결정된 관계를 참조하여 상기 샘플의 종양 분획을 결정하는 단계; 및

보고서에 추정된 종양 분획을 기록함으로써 상기 보고서를 제공하는 단계를 포함한다.

- [0175] 한 구현예에서, 상기 대립유전자 분획은 서열분석, 예컨대 NGS를 포함하는 방법에 의해 결정된다. 한 구현예에서, 상기 대립유전자 분획은 표적 선택을 추가로 포함하는 방법에 의해, 예컨대 용액 혼성화에 의해 결정된다. 다른 구현예들에서, DNA(예컨대, cfDNA, ctDNA 등)를 검출하기 위해 사용되는 다른 방법론들, 예컨대 미세배열이 사용될 수 있다.
- [0176] 한 구현예에서, 상기 방법은 상기 대상체 또는 제3자에게 상기 보고서를 전송하는 단계를 추가로 포함한다. 한 구현예에서, 상기 보고서는 상기 추정된 종양 분획에 기초한 치료 옵션을 추가로 포함한다.
- [0177] 한 구현예에서, 상기 보고 단계는 상기 대상체의 유전체 프로파일(예컨대, 상기 질병과 관련된 유전체 프로파일)을 추가로 포함한다.
- [0178] 생검을 평가하는 방법(예컨대, 본원에 기재된 방법에 따라 결정된 종양 분획을 결정하는 단계를 포함함)이 기재된다. 상기 방법은 상기 생검으로부터의 샘플에서 하위 유전체 간격과 관련된 목표 변수에 대한 값을 획득하는 단계; 상기 목표 변수로부터 확실성 메트릭을 결정하는 단계; 저장된 확실성 메트릭 및 저장된 종양 분획 사이의 결정된 관계에 액세스하는 단계; 및 상기 확실성 메트릭 및 상기 결정된 관계를 참조하여 상기 샘플의 종양 분획을 결정함으로써 상기 생검을 평가하는 단계를 포함한다.
- [0179] 한 구현예에서, 임계값을 초과하는 추정된 종양 분획은 상기 생검이 유전체 프로파일링에 적합하다는 것을 나타낸다.
- [0180] 예시적인 컴퓨터 구현
- [0181] 상기 기재된 공정은 단지 종양 분획을 추정하는 데 사용될 수 있는 시스템의 예시적인 구현예들이다. 상기 예시적인 구현예들은 본 개시의 범위를 제한하도록 의도되지 않는다. 본원에 제시된 구현예들 및 청구항들 중 어느 것도, 상기 청구항이 특정 구현을 명시적으로 인용하는 제한을 포함하지 않는 한, 임의의 특정 구현으로 제한되도록 의도되지 않는다.
- [0182] 다양한 구현예와 관련된 공정 및 방법, 이의 작동, 및 상기 방법 및 작동의 다양한 구현예 및 변형은 개별적으로 또는 조합하여 컴퓨터 판독가능 매체, 예를 들어 비휘발성 기록 매체, 집적 회로 메모리 소자, 또는 이들의 조합 매체 상에 유형적으로 구현된 컴퓨터 판독가능 신호에 의해 정의될 수 있다. 일 구현예에 따르면, 상기 컴퓨터 판독가능 매체는 컴퓨터 실행 가능 명령어가 상기 매체 상에 영구적으로 또는 반영구적으로 저장될 수 있다는 점에서 비밀시적일 수 있다. 상기 신호는, 예를 들어 컴퓨터에 의해 실행된 결과로서 상기 컴퓨터에 본원에 기재된 방법 또는 작동, 및/또는 다양한 구현예, 변형 및 이들의 조합 중 하나 이상을 실시하도록 지시하는 하나 이상의 프로그램의 일부로서 명령을 정의할 수 있다. 상기 명령어는 복수의 프로그래밍 언어 중 임의의 것, 예를 들어 자바(Java), 비주얼 베이직(Visual Basic), C, C# 또는 C++, 포트란(Fortran), 파스칼(Pascal), 에펠(Eiffel), 베이직(Basic), COBOL 등 또는 이들의 다양한 조합 중 임의의 것으로 작성될 수 있다. 상기 명령어가 저장되는 컴퓨터 판독가능 매체는 상기 기재된 범용 컴퓨터의 하나 이상의 구성요소에 상주할 수 있고, 상기 구성요소 중 하나 이상에 걸쳐 분산될 수 있다.
- [0183] 컴퓨터 판독가능 매체는 그 위에 저장된 명령어가 본원에서 논의된 본 개시의 양태들을 구현하기 위해 임의의 컴퓨터 시스템 리소스 상에 로딩될 수 있도록 전송가능할 수 있다. 또한, 상기 기재된 컴퓨터 판독가능 매체에 저장된 명령어는 호스트 컴퓨터에서 실행되는 응용 프로그램의 일부로서 구현된 명령어에 제한되지 않는다는 것이 이해되어야 한다. 오히려, 상기 명령어는 본 개시의 상기 논의된 양태들을 구현하기 위해 프로세서를 프로그래밍하는데 사용될 수 있는 임의의 유형의 컴퓨터 코드(예컨대, 소프트웨어 또는 마이크로코드)로서 구현될 수 있다.
- [0184] 본 개시에 따른 다양한 구현예가 하나 이상의 컴퓨터 시스템에서 구현될 수 있다. 상기 컴퓨터 시스템은, 예를 들어 범용 컴퓨터, 예컨대 인텔 펜티엄(Intel PENTIUM) 형 프로세서, 모토롤라 파워PC(Motorola PowerPC), 선울트라스파크(Sun UltraSPARC), 휴렛-패커드(Hewlett-Packard) PA-RISC 프로세서, ARM 코어텍스(Cortex) 프로세서, 퀄컴 스콜피온(Qualcomm Scorpion) 프로세서 또는 기타 유형의 프로세서에 기반한 것일 수 있다. 임의의 유형의 컴퓨터 시스템 중 하나 이상이 본 개시의 다양한 구현예에 따라 사용자에게 제안을 확장하고 제안을 상환하는 것을 부분적으로 또는 완전히 자동화하는 데 사용될 수 있음이 이해되어 한다. 또한, 소프트웨어 설계 시스템은 단일 컴퓨터에 위치하거나 또는 통신 네트워크에 의해 연결된 복수의 컴퓨터에 분산될 수 있다.
- [0185] 컴퓨터 시스템은 특별히 프로그래밍된 특수 목적 하드웨어, 예를 들어 애플리케이션 별 집적 회로(application-

specific integrated circuit, ASIC)를 포함할 수 있다. 본 개시의 양태들은 소프트웨어, 하드웨어 또는 펌웨어, 또는 이들의 임의의 조합으로 구현될 수 있다. 또한, 상기 방법, 작동, 시스템, 시스템 요소 및 이들의 구성요소는 상기 기재된 컴퓨터 시스템의 일부로서 또는 독립 구성요소로서 구현될 수 있다.

[0186] 컴퓨터 시스템은 고급 컴퓨터 프로그래밍 언어를 사용하여 프로그래밍할 수 있는 범용 컴퓨터 시스템일 수 있다. 컴퓨터 시스템은 또한 특별히 프로그래밍된 특수 목적 하드웨어를 사용하여 구현될 수도 있다. 컴퓨터 시스템에는 일반적으로 상업적으로 입수 가능한 프로세서, 예컨대 인텔 사(Intel Corporation)에서 입수 가능한 일반적으로 공지된 펜티엄(Pentium) 급 프로세서가 있을 수 있다. 다른 많은 프로세서를 입수 가능하다. 상기 프로세서는 일반적으로, 예를 들어 마이크로소프트 사(Microsoft Corporation)에서 입수 가능한 윈도우(Windows) NT, 윈도우 2000(Windows ME), 윈도우 XP, 윈도우 비스타(Windows Vista) 또는 윈도우 7 운영 체제, 애플 컴퓨터(Apple Computer)에서 입수 가능한 MAC OS X 스노우 레오파드(Snow Leopard), MAC OS X 라이언(Lion) 운영 체제, 오라클 사(Oracle Corporation)에서 입수 가능한 솔라리스(Solaris) 운영 체제, 다양한 공급원에서 입수 가능한 iOS, 블랙베리(Blackberry) OS, 윈도우 7 모바일(Mobile) 또는 안드로이드(Android) OS 운영 체제 또는 UNIX일 수 있는 운영 체제를 실행한다. 다른 많은 운영 체제가 사용될 수 있다.

[0187] 본 개시의 일부 양태들은 컴퓨터 네트워크를 통해 결합된 다수의 상이한 유형들의 시스템들 상에서 실행될 수 있는 분산된 애플리케이션 구성요소들로서 구현될 수 있다. 일부 구성요소들은 모바일 장치, 서버, 태블릿 또는 기타 시스템 유형 상에 위치하여 실행될 수 있다. 데이터베이스 또는 기타 구성요소 유형과 같은 분산 시스템의 다른 구성요소들도 사용할 수 있다.

[0188] 프로세서 및 운영 체제는 함께 고급 프로그래밍 언어의 응용 프로그램이 작성되는 컴퓨터 플랫폼을 정의한다. 본 개시는 특정 컴퓨터 시스템 플랫폼, 프로세서, 운영 체제, 알고리즘의 계산 세트, 코드, 또는 네트워크로 제한되지 않는다는 것이 이해되어야 한다. 또한, 본 개시의 다양한 양태를 구현하는 분산 컴퓨터 시스템에서 복수의 컴퓨터 플랫폼 유형이 사용될 수 있음이 이해되어야 한다. 또한, 본 개시가 특정 프로그래밍 언어, 알고리즘의 계산 세트, 코드 또는 컴퓨터 시스템에 제한되지 않는다는 것이 당업자에게 명백해야 한다. 또한, 다른 적절한 프로그래밍 언어 및 다른 적절한 컴퓨터 시스템이 또한 사용될 수 있음이 이해되어야 한다.

[0189] 컴퓨터 시스템의 하나 이상의 부분은 통신 네트워크에 연결된 하나 이상의 컴퓨터 시스템에 걸쳐 분산될 수 있다. 상기 컴퓨터 시스템은 또한 범용 컴퓨터 시스템일 수 있다. 예를 들어, 본 개시의 다양한 양태는 서비스(예컨대, 서버)를 하나 이상의 클라이언트 컴퓨터에 제공하거나 분산 시스템의 일부로서 전체 작업을 실시하도록 구성된 하나 이상의 컴퓨터 시스템 사이에 분산될 수 있다. 예를 들어, 본 개시의 다양한 양태는 본 개시의 다양한 구현예에 따라 다양한 기능을 실시하는 하나 이상의 서버 시스템들 사이에 분산된 구성요소들을 포함하는 클라이언트-서버 시스템 상에서 실시될 수 있다. 상기 구성요소는 통신 프로토콜(예컨대, TCP/IP)을 사용한 통신 네트워크(예컨대, 인터넷)를 통해 통신하는 실행, 중간(예컨대, IL) 또는 해석(예컨대, 자바) 코드일 수 있다. 본 개시의 특정 양태는 또한 클라우드 기반 컴퓨터 시스템(예컨대, 아마존닷컴(Amazon.com)에 의해 제공되는 EC2 클라우드 기반 컴퓨팅 플랫폼), 클라이언트 및 서버를 포함하는 분산 컴퓨터 네트워크, 또는 시스템들의 임의의 조합에서 구현될 수 있다.

[0190] 본 개시는 임의의 특정 시스템 또는 시스템 군 상에서 실행하는 것으로 제한되지 않음이 이해되어야 한다. 또한, 본 개시는 임의의 특정 분산 아키텍처, 네트워크, 또는 통신 프로토콜에 제한되지 않음이 이해되어야 한다.

[0191] 본 개시의 다양한 구현예는 객체 지향 프로그래밍 언어, 예컨대 스몰토크(SmallTalk), 자바, C++, 아다(Ada), 또는 C#(C-Sharp)을 사용하여 프로그래밍될 수 있다. 다른 객체 지향 프로그래밍 언어도 사용할 수 있다. 대안적으로, 기능적, 스크립팅 및/또는 논리적 프로그래밍 언어가 사용될 수 있다. 본 개시의 다양한 양태는 프로그래밍되지 않은 환경(예컨대, 브라우저 프로그램의 윈도우에서 볼 때 그래픽 사용자 인터페이스(GUI)의 양태들을 제공하고 다른 기능들을 실시하는 HTML, XML 또는 다른 포맷으로 생성되는 문서)에서 구현될 수 있다. 본 개시의 다양한 양태는 프로그래밍된 또는 프로그래밍 되지 않은 요소들, 또는 이들의 임의의 조합으로서 구현될 수 있다.

[0192] 또한, 장치의 하나 이상의 구성 요소를 포함하는 하나 이상의 컴퓨터 시스템 각각에서 상기 구성 요소의 각각은 시스템 상의 하나 이상의 위치에 상주할 수 있다. 예를 들어, 상기 장치의 구성 요소 중 상이한 부분들이 하나 이상의 컴퓨터 시스템 상의 상이한 메모리 영역(예컨대, RAM, ROM, 디스크 등)에 상주할 수 있다. 상기 하나 이상의 컴퓨터 시스템 각각은 다른 구성요소들 중에서 복수의 공지된 구성요소, 예컨대 하나 이상의 프로세서, 메모리 시스템, 디스크 저장 시스템, 하나 이상의 네트워크 인터페이스, 및 상기 다양한 구성요소를 상호연결하는

하나 이상의 버스 또는 기타 내부 통신 링크를 포함할 수 있다.

- [0193] 본 개시는 도 2 및 도 3과 관련하여 하기 기재되는 컴퓨터 시스템 상에서 구현될 수 있다. 특히, 도 2는 다양한 양태를 구현하는데 사용되는 예시의 컴퓨터 시스템 200을 도시한다. 도 3은 사용될 수 있는 예시적인 저장 시스템을 도시한다.
- [0194] 시스템(200)은 본 개시의 다양한 양태를 구현하기에 적합한 컴퓨터 시스템의 예시적인 구현예에 불과하다. 그러한 예시적인 구현예는, 예를 들어 시스템의 다수의 다른 구현들 중 임의의 것이 가능하고 본 개시의 범위 내에 속하는 것으로 의도되기 때문에 범위를 제한하도록 의도되지 않는다. 예를 들어, 가상 컴퓨팅 플랫폼이 사용될 수 있다. 하기에 제시된 청구항들 중 어느 것도, 상기 청구항이 특정 구현을 명시적으로 인용하는 제한을 포함하지 않는 한, 상기 시스템의 임의의 특정 구현으로 제한되도록 의도되지 않는다.
- [0195] 본 개시에 따른 다양한 구현예가 하나 이상의 컴퓨터 시스템에서 구현될 수 있다. 상기 컴퓨터 시스템은, 예를 들어 범용 컴퓨터, 예컨대 인텔 펜티엄(Intel PENTIUM)형 프로세서, 모토로라 파워PC(Motorola PowerPC), 선 울트라스파크(Sun UltraSPARC), 휴렛-패커드(Hewlett-Packard) PA-RISC 프로세서, 또는 기타 유형의 프로세서에 기반한 것일 수 있다. 임의의 유형의 컴퓨터 시스템 중 하나 이상이 본 개시의 다양한 구현예에 따라 보안 서비스를 다른 시스템 및 서비스와 통합하는 것을 부분적으로 또는 완전히 자동화하는 데 사용될 수 있음이 이해되어 한다. 또한, 소프트웨어 설계 시스템은 단일 컴퓨터에 위치하거나 또는 통신 네트워크에 의해 연결된 복수의 컴퓨터에 분산될 수 있다.
- [0196] 예를 들어, 본 개시의 다양한 양태는 도 2에 도시된 것과 같은 범용 컴퓨터 시스템(200)에서 실행되는 특수 소프트웨어로서 구현될 수 있다. 상기 컴퓨터 시스템(200)은 디스크 드라이브, 메모리, 또는 데이터를 저장하기 위한 기타 장치와 같은 하나 이상의 메모리 장치(204)에 연결된 프로세서(203)를 포함할 수 있다. 메모리(204)는 일반적으로 컴퓨터 시스템(200)의 작동 동안 프로그램 및 데이터를 저장하는 데 사용된다. 컴퓨터 시스템(200)의 구성요소는 하나 이상의 버스(예컨대, 동일한 기계 내에 통합된 구성요소 간) 및/또는 네트워크(예컨대, 별도의 개별 기계 상에 상주하는 구성요소 간)를 포함할 수 있는 상호접속 메커니즘(205)에 의해 연결될 수 있다. 상기 상호접속 메커니즘(205)은 통신(예컨대, 데이터, 명령)이 시스템(200)의 시스템 구성요소 사이에서 교환될 수 있게 한다. 컴퓨터 시스템(200)은 또한 하나 이상의 입력 장치(202), 예를 들어 키보드, 마우스, 트랙볼, 마이크로폰, 터치 스크린, 및 하나 이상의 출력 장치(201), 예를 들어 인쇄 장치, 디스플레이 스크린, 및/또는 스피커를 포함한다. 또한, 컴퓨터 시스템(200)은 컴퓨터 시스템(200)을 통신 네트워크에 연결하는 하나 이상의 인터페이스(미도시)를 포함할 수 있다(상호접속 메커니즘(205)에 추가로 또는 대안으로서).
- [0197] 도 3에 더 상세히 도시된 저장 시스템(206)은 일반적으로 신호가 저장되어 상기 프로세서에 의해 처리될 프로그램을 저장하프로세서에 의해 실행될 프로그램을 정의하는 신호 또는 상기 프로그램에 의해 처리될 매체(301) 상에 또는 그 내부에 저장된 정보가 저장된 컴퓨터 판독 및 기록 가능한 비휘발성 기록 매체(301)를 포함한다. 상기 매체는, 예를 들어 디스크 또는 플래시 메모리일 수 있다. 전형적으로, 동작시에, 상기 프로세서는 데이터가 상기 비휘발성 기록 매체(301)로부터 상기 매체(301)보다 프로세서에 의해 정보에 대한 더 빠른 액세스를 허용하는 다른 메모리(302)로 판독되게 한다. 상기 메모리(302)는 일반적으로 동적 랜덤 액세스 메모리(DRAM) 또는 정적 메모리(SRAM)와 같은 휘발성 랜덤 액세스 메모리이다.
- [0198] 데이터는 도시된 바와 같이 저장 시스템(206) 또는 메모리 시스템(204) 내에 위치할 수 있다. 상기 프로세서(203)는 일반적으로 집적 회로 메모리(204, 202) 내의 데이터를 조작하고 처리가 완료된 후 데이터를 상기 매체(301)에 복사한다. 상기 매체(301) 및 상기 집적 회로 메모리 요소(302) 사이의 데이터 이동을 관리하기 위해 다양한 메커니즘이 공지되어 있으며, 본 개시는 이에 제한되지 않는다. 본 개시는 특정 메모리 시스템(204) 또는 저장 시스템(206)으로 제한되지 않는다.
- [0199] 컴퓨터 시스템은 특별히 프로그래밍된 특수 목적 하드웨어, 예를 들어 애플리케이션 별 집적 회로(application-specific integrated circuit, ASIC)를 포함할 수 있다. 본 개시의 양태들은 소프트웨어, 하드웨어 또는 펌웨어, 또는 이들의 임의의 조합으로 구현될 수 있다. 또한, 상기 방법, 작동, 시스템, 시스템 요소 및 이들의 구성요소는 상기 기재된 컴퓨터 시스템의 일부로서 또는 독립 구성요소로서 구현될 수 있다.
- [0200] 컴퓨터 시스템(200)이 본 개시의 다양한 양태가 실시될 수 있는 컴퓨터 시스템의 한 유형으로서 예시적으로 도시되어 있지만, 본 개시의 양태는 도 2에 도시된 바와 같이 컴퓨터 시스템 상에서 구현되는 것으로 제한되지 않는다는 것이 이해되어야 한다. 본 개시의 다양한 양태들은 도 2에 도시된 것과 상이한 아키텍처 또는 구성요소

들을 갖는 하나 이상의 컴퓨터 상에서 실시될 수 있다.

- [0201] 컴퓨터 시스템(200)은 고급 컴퓨터 프로그래밍 언어를 사용하여 프로그래밍할 수 있는 범용 컴퓨터 시스템일 수 있다. 컴퓨터 시스템(300)은 또한 특별히 프로그래밍된 특수 목적 하드웨어를 사용하여 구현될 수도 있다. 컴퓨터 시스템(200)에서, 프로세서(203)는 상업적으로 입수 가능한 프로세서, 예컨대 일반적으로 공지된 펜티엄(Pentium), 코어(Core), 코어 Vpro, 제온(Xeon) 또는 인텔 사(Intel Corporation)로부터 입수 가능한 아이테니엄(Itanium) 급 프로세서이다. 다른 많은 프로세서를 입수 가능하다. 상기 프로세서는 일반적으로, 예를 들어 마이크로소프트 사(Microsoft Corporation)에서 입수 가능한 리눅스(Linux), 윈도우(Windows) NT, 윈도우 2000(Windows ME), 윈도우 XP, 윈도우 비스타(Windows Vista), 윈도우 7, 또는 윈도우 10 운영 체제, 애플 컴퓨터(Apple Computer)에서 입수 가능한 MAC OS 스노우 레오파드(Snow Leopard), MAC OS X 라이언(Lion) 운영 체제, 선 마이크로시스템즈(Sun Microsystems)에서 입수 가능한 솔라리스(Solaris) 운영 체제, 다양한 공급원에서 입수 가능한 iOS, 블랙베리(Blackberry) OS, 윈도우 7 모바일(Mobile) 또는 안드로이드(Android) OS 운영 체제 또는 UNIX일 수 있는 운영 체제를 실행한다. 다른 많은 운영 체제가 사용될 수 있다.
- [0202] 프로세서 및 운영 체제는 함께 고급 프로그래밍 언어의 응용 프로그램이 작성되는 컴퓨터 플랫폼을 정의한다. 본 개시는 특정 컴퓨터 시스템 플랫폼, 프로세서, 운영 체제, 또는 네트워크로 제한되지 않는다는 것이 이해되어야 한다. 또한, 본 개시가 특정 프로그래밍 언어 또는 컴퓨터 시스템에 제한되지 않는다는 것이 당업자에게 명백해야 한다. 또한, 다른 적절한 프로그래밍 언어 및 다른 적절한 컴퓨터 시스템이 또한 사용될 수 있음이 이해되어야 한다.
- [0203] 컴퓨터 시스템의 하나 이상의 부분은 통신 네트워크에 연결된 하나 이상의 컴퓨터 시스템(미도시)에 걸쳐 분산될 수 있다. 상기 컴퓨터 시스템은 또한 범용 컴퓨터 시스템일 수 있다. 예를 들어, 본 개시의 다양한 양태는 서비스(예컨대, 서버)를 하나 이상의 클라이언트 컴퓨터에 제공하거나 분산 시스템의 일부로서 전체 작업을 실시하도록 구성된 하나 이상의 컴퓨터 시스템 사이에 분산될 수 있다. 예를 들어, 본 개시의 다양한 양태는 본 개시의 다양한 구현예에 따라 다양한 기능을 실시하는 하나 이상의 서버 시스템들 사이에 분산된 구성요소들을 포함하는 클라이언트-서버 시스템 상에서 실시될 수 있다. 상기 구성요소는 통신 프로토콜(예컨대, TCP/IP)을 사용한 통신 네트워크(예컨대, 인터넷)를 통해 통신하는 실행, 중간(예컨대, IL) 또는 해석(예컨대, 자바) 코드일 수 있다.
- [0204] 본 개시는 임의의 특정 시스템 또는 시스템 군 상에서 실행하는 것으로 제한되지 않음이 이해되어야 한다. 또한, 본 개시는 임의의 특정 분산 아키텍처, 네트워크, 또는 통신 프로토콜에 제한되지 않음이 이해되어야 한다.
- [0205] 본 개시의 다양한 구현예는 객체 지향 프로그래밍 언어, 예컨대 스몰토크(SmallTalk), 자바, C++, 아다(Ada), 또는 C#(C-Sharp)을 사용하여 프로그래밍될 수 있다. 다른 객체 지향 프로그래밍 언어도 사용할 수 있다. 대안적으로, 기능적, 스크립팅 및/또는 논리적 프로그래밍 언어가 사용될 수 있다. 본 개시의 다양한 양태는 프로그래밍되지 않은 환경(예컨대, 브라우저 프로그램의 윈도우에서 볼 때 그래픽 사용자 인터페이스(GUI)의 양태들을 제공하고 다른 기능들을 실시하는 HTML, XML 또는 다른 포맷으로 생성되는 문서)에서 구현될 수 있다. 본 개시의 다양한 양태는 다양한 인터넷 기술, 예를 들어 일반적으로 공지된 공통 게이트웨이 인터페이스(Common Gateway Interface, CGI) 스크립트, 하이퍼-텍스트 전처리기(Hyper-text Preprocessor, PHP PHP), 액티브 서버 페이지(Active Server Pages, ASP), 하이퍼텍스트 마크업 언어(HyperText Markup Language, HTML), 확장성 마크업 언어(Extensible Markup Language, XML), 자바(Java), 자바스크립트(JavaScript), 비동기 자바스크립트와 XML(AJAX), 플래시(Flash) 및 기타 프로그래밍 방법들을 사용하여 구현될 수 있다. 또한, 본 개시의 다양한 양태는 무엇보다도 클라우드 기반 컴퓨팅 플랫폼, 예컨대 아마존닷컴(Amazon.com, 워싱턴주 시애틀 소재)에서 상업적으로 입수가 가능한 일반적으로 공지된 EC2 플랫폼에서 구현될 수 있다. 본 개시의 다양한 양태는 프로그래밍된 또는 프로그래밍 되지 않은 요소들, 또는 이들의 임의의 조합으로서 구현될 수 있다.
- [0206] 정의
- [0207] 특정 용어들이 정의된다. 명세서 전반에 걸쳐 추가 용어가 정의된다.
- [0208] 본원에서 사용된 관사 "a" 및 "an"은 관사의 문법적 대상의 하나 또는 하나 초과(예컨대, 적어도 하나)를 지칭한다.
- [0209] "약" 및 "대략"은 일반적으로 측정의 특성 또는 정밀도를 고려할 때 측정된 양에 대해 허용 가능한 오차의 정도를 의미한다. 예시적인 오차의 정도는 주어진 값 또는 값의 범위의 20%(%) 이내, 일반적으로 10% 이내, 보다 일

반적으로 5% 이내이다.

[0210] 본원에서 사용된 용어 "수득하다" 또는 "수득한"은 물리적 실체 또는 값, 예컨대 수치 값의 보유를 상기 물리적 실체 또는 값을 "직접적으로 수득" 또는 "간접적으로 수득"함으로써 수득하는 것을 지칭한다. "직접적으로 수득한"은 상기 물리적 개체 또는 값을 수득하기 위한 공정을 실시하는 것(예컨대, 합성 또는 분석 방법을 실시하는 것)을 의미한다. "간접적으로 수득한"은 다른 당사자 또는 공급원(예컨대, 상기 물리적 실체 또는 값을 직접 수득한 제3자 실험실)으로부터 상기 물리적 실체 또는 값을 수령하는 것을 지칭한다. 물리적 개체를 직접 수득하는 것은 물리적 물질, 예컨대 출발 물질의 물리적 변화를 포함하는 공정을 실시하는 것을 포함한다. 예시적인 변경은 2개 이상의 출발 물질로부터 물리적 개체 만드는 것, 물질의 전단 또는 단편화, 물질의 분리 또는 정제, 2개 이상의 개별 개체를 혼합물로 결합, 공유 또는 비-공유 결합을 파괴 또는 형성하는 것을 포함하는 화학 반응을 실시하는 것을 포함한다. 값을 직접 수득하는 것은 샘플 또는 다른 물질의 물리적 변화를 포함하는 공정을 실시하는 것, 예컨대 물질, 예컨대 샘플, 분석물 또는 시약의 물리적 변화를 포함하는 분석적 공정(때로는 "물리적 분석"으로 본원에서 언급됨)을 실시하는 것, 분석적 방법, 예컨대, 하기 중 하나 이상을 포함하는 방법을 실시하는 것: 다른 물질로부터 하나의 물질, 예컨대 분석물, 이의 단편 또는 기타 유도체를 분리 또는 정제하는 것; 분석물, 또는 이의 단편 또는 기타 유도체를 다른 물질, 예컨대 완충제, 용매 또는 반응물과 결합하는 것; 또는 분석물의 제1 및 제2 원자 간의 공유 또는 비공유 결합을 끊거나 형성함으로써 상기 분석물, 또는 이의 단편 또는 기타 유도체의 구조를 변경하는 것; 또는, 예컨대 시약의 제1 및 제2 원자 간의 공유 또는 비공유 결합을 끊거나 형성함으로써 상기 시약, 또는 이의 단편 또는 기타 유도체의 구조를 변경하는 것을 포함하는 공정을 실시하는 것을 포함한다.

[0211] 본원에서 사용된 용어 "서열을 수득하는" 또는 "리드(read)를 수득하는"은 서열 또는 리드를 "직접적으로 수득" 또는 "간접적으로 수득"함으로써 뉴클레오타이드 서열 또는 아미노산 서열의 소유를 수득하는 것을 지칭한다. 서열 또는 리드를 "직접적으로 수득하는" 것은 서열분석 방법(예컨대, 차세대 서열분석(NGS) 방법)을 실시하는 것과 같이 서열을 수득하기 위한 공정(예컨대, 합성 또는 분석적 방법을 실시)을 실시하는 것을 의미한다. 서열 또는 리드를 "간접적으로 수득하는" 것은 다른 당사자 또는 공급원(예컨대, 서열을 직접 수득한 제 3의 실험실)으로부터 상기 서열에 관한 정보 또는 지식을 수령하거나 상기 서열을 수령하는 것을 지칭한다. 수득한 서열 또는 리드는 전체 서열일 필요는 없으며, 예컨대 적어도 하나의 뉴클레오타이드를 서열분석하는 것, 또는 본원에 개시된 변경 중 하나 이상을 샘플, 생검 또는 대상체에 존재하는 것으로 식별하는 정보 또는 지식을 수득하는 것은 서열을 수득하는 것으로 여겨진다.

[0212] 서열 또는 리드를 직접 수득하는 것은 물리적 물질, 예컨대 출발 물질, 예컨대 본원에 기재된 샘플의 물리적 변화를 포함하는 공정을 실시하는 것을 포함한다. 예시적인 변경은 2개 이상의 출발 물질로부터 물리적 개체 만드는 것, 물질, 예컨대 유전체 DNA 파편의 전단 또는 단편화, 물질의 분리 또는 정제(예컨대, 조직으로부터 핵산 샘플을 분리하는 것), 2개 이상의 개별 개체를 혼합물로 결합, 공유 또는 비-공유 결합을 파괴 또는 형성하는 것을 포함하는 화학 반응을 실시하는 것을 포함한다. 값을 직접 수득하는 것은 상기에 기재된 샘플 또는 다른 물질의 물리적 변화를 포함하는 공정을 실시하는 것을 포함한다. 상기 단편의 크기(예컨대, 상기 단편들의 평균 크기)는 2500 bp 이하, 2000 bp 이하, 1500 bp 이하, 1000 bp 이하, 800 bp 이하, 600 bp 이하, 400 bp 또는 200 bp 이하일 수 있다. 일부 구현예들에서, 상기 단편(예컨대, cfDNA)의 크기는 약 150 bp 내지 약 200 bp(예컨대, 약 160 bp 내지 약 170 bp)이다. 일부 구현예들에서, 상기 단편(예컨대, FFPE 샘플로부터의 DNA 단편)의 크기는 약 150 bp 내지 약 250 bp이다. 일부 구현예들에서, 상기 단편(예컨대, FFPE 샘플의 RNA로부터 수득된 cDNA 단편)의 크기는 약 100 bp 내지 약 150 bp이다.

[0213] 본원에서 사용된 용어 "샘플을 수득하는"은 샘플, 예컨대 본원에 기재된 샘플을, 상기 샘플을 "직접적으로 수득" 또는 "간접적으로 수득"함으로써 이의 보유를 수득하는 것을 지칭한다. "샘플을 직접적으로 수득하는"은 샘플을 수득하기 위한 공정을 실시하는(예컨대, 수술 또는 추출과 같은 물리적 방법을 실시하는) 것을 의미한다. "샘플을 간접적으로 수득하는" 것은 다른 당사자 또는 공급원(예컨대, 샘플을 직접 수득한 제 3의 실험실)으로부터 상기 샘플을 수령하는 것을 지칭한다. 샘플을 직접 수득하는 것은 물리적 물질, 예컨대 출발 물질, 예컨대 조직, 예컨대 인간 환자의 조직 또는 환자로부터 이전에 분리된 조직의 물리적 변화를 포함하는 공정을 실시하는 것을 포함한다. 예시적인 변경은 출발 물질로부터 물리적 개체 만드는 것, 조직의 절개 또는 폐기; 물질(예컨대, 샘플 조직 또는 핵산 샘플)의 분리 또는 정제; 2개 이상의 개별 개체를 혼합물로 결합; 공유 또는 비-공유 결합을 파괴 또는 형성하는 것을 포함하는 화학 반응을 실시하는 것을 포함한다. 샘플을 직접 수득하는 것은 샘플 또는 다른 물질, 예컨대 상기에 기재된 것의 물리적 변화를 포함하는 공정을 실시하는 것을 포함한다.

- [0214] 유전자 또는 유전자 생성물(예컨대, 마커 유전자 또는 유전자 산물)의 본원에 사용된 "변경" 또는 "변경된 구조"는 상기 유전자 또는 유전자 산물 내의 돌연변이 또는 돌연변이들, 예컨대 정상 또는 야생형 유전자와 비교하여 상기 유전자 또는 유전자 산물의 무결성, 서열, 구조, 양 또는 활성에 영향을 미치는 돌연변이의 존재를 지칭한다. 상기 변경은 정상 또는 건강한 조직 또는 세포(예컨대, 대조군)에서의 양, 구조 및/또는 활성과 비교하여 암 조직 또는 암 세포의 양, 구조 및/또는 활성에서 이루어질 수 있으며, 암과 같은 질병 상태와 관련이 있다. 예를 들어, 암과 연관되거나 항암 치료제에 대한 반응성을 예측하는 변경은 변경된 뉴클레오티드 서열(예컨대, 돌연변이), 아미노산 서열, 염색체 전좌, 염색체 내 역위, 복제 수, 발현 수준, 단백질 수준, 단백질 활성, 유전 외적 변형(예컨대, 정상의 건강한 조직 또는 세포와 비교하여 암 조직 또는 암세포에서의 메틸화 또는 아세틸화 상태, 또는 번역후 변형)을 가질 수 있다. 예시적인 돌연변이는 점 돌연변이(예컨대, 침묵, 과오 또는 무의미), 결실, 삽입, 역위, 복제, 증폭, 전좌, 염색체간 및 염색체내 재배열을 포함하지만 이에 제한되지는 않는다. 돌연변이는 유전자의 부호화 또는 비부호화 영역에 존재할 수 있다. 특정 구현예들에서, 상기 변경(들)은 재배열, 예컨대 하나 이상의 인트론 또는 이의 단편을 포함하는 유전체 재배열(예컨대, 5'- 및/또는 3'-UTR에서 하나 이상의 재배열)로서 검출된다. 특정 구현예들에서, 상기 변경은 표현형, 예컨대 암성 표현형(예컨대, 암 위험, 암 진행, 암 치료 또는 암 치료에 대한 내성 중 하나 이상)과 연관된다(또는 연관되지 않는다). 일 구현예에서, 상기 변경(또는 중앙 돌연변이 부담)은 암에 대한 유전적 위험 인자, 양성 치료 반응 예측자, 음성 치료 반응 예측자, 양성 예후 인자, 음성 예후 인자, 또는 진단 인자 중 하나 이상과 연관된다.
- [0215] 본원에 사용된 용어 "삽입-결실(indel)"은 세포의 핵산에서 하나 이상의 뉴클레오티드의 삽입, 결실 또는 둘 다를 지칭한다. 특정 구현예들에서, 삽입-결실은 하나 이상의 뉴클레오티드의 삽입 및 결실 둘 모두를 포함하며, 상기 삽입 및 결실 둘 모두는 핵산 상에서 가깝게 있다. 특정 구현예들에서, 상기 삽입-결실은 뉴클레오티드의 총 수에서 순 변화를 유발한다. 특정 구현예들에서, 상기 삽입-결실은 약 1 내지 약 50개의 뉴클레오티드에서 순 변화를 유발한다.
- [0216] 본원에서 사용된 용어 "클론 프로파일"은 대상체 간격(또는 이를 포함하는 세포)의 하나 이상의 서열, 예컨대 대립유전자 또는 시그니처(signature)의 발생, 동일성, 가변성, 분포, 발현(하위 유전체 시그니처의 전사된 복제의 발생 또는 수준), 또는 존재량, 예컨대 상대적 존재량을 지칭한다. 한 구현예에서, 상기 클론 프로파일은 대상체 간격에 대한 복수의 서열, 대립유전자 또는 시그니처가 샘플 내에 존재할 때 상기 대상체 간격(또는 이를 포함하는 세포)에 대한 하나의 서열, 대립유전자 또는 시그니처에 대한 상대적 존재량의 값이다. 예컨대, 한 구현예에서, 클론 프로파일은 대상체 간격에 대한 복수의 VDJ 또는 VJ 조합 중 하나 이상의 상대적 존재량에 대한 값을 포함한다. 한 구현예에서, 클론 프로파일은 대상체 간격 동안 선택된 V 분절의 상대적 존재량에 대한 값을 포함한다. 한 구현예에서, 클론 프로파일은 대상체 간격의 서열 내에서, 예컨대 체세포 초돌연변이로부터 발생하는 바와 같은 다양성에 대한 값을 포함한다. 한 구현예에서, 클론 프로파일은, 예컨대 서열, 대립유전자 또는 시그니처를 포함하는 발현된 하위 유전체 간격의 발생 또는 수준에 의해 입증되는 바와 같이, 서열, 대립유전자 또는 시그니처의 발현의 발생 또는 준에 대한 값을 포함한다.
- [0217] 본원에 사용된 용어 "발현된 하위 유전체 간격"은 하위 유전체 간격의 전사된 서열을 지칭한다. 한 구현예에서, 상기 발현된 하위 유전체 간격의 서열은, 예컨대 일부 서열이 전사되지 않을 수 있기 때문에 전사되는 하위 유전체 간격과 상이할 것이다.
- [0218] 본원에 사용된 용어 "돌연변이 대립유전자 빈도"(MAF)는, 예컨대 샘플 내 특정 유전자좌에서 돌연변이 대립유전자의 상대 빈도를 지칭한다. 일부 구현예들에서, 돌연변이 대립유전자 빈도는 분획 또는 백분율로 표현된다.
- [0219] 본원에서 사용되는 "시그니처"는 대상체 간격의 서열을 지칭한다. 시그니처는 대상체 간격에서 복수의 가능성 중 하나의 발생을 진단할 수 있으며, 예컨대 시그니처는 재배열된 중쇄 또는 경쇄 가변 영역 유전자에서 선택된 V 분절의 발생; 선택된 VJ 접합의 발생, 예컨대 재배열된 중쇄 가변 영역 유전자에서 선택된 V 및 선택된 J 분절의 발생을 진단할 수 있다. 한 구현예에서, 시그니처는 복수의 특정 핵산 서열을 포함한다. 따라서, 시그니처는 특정 핵산 서열에 제한되지 않고, 대상체 간격에서 제1 군의 서열 또는 가능성 및 대상체 간격에서 제2 군의 가능성을 서로 구별할 수 있을 만큼, 예컨대 제1 V 분절 및 제2 V 분절을 서로 구별하여, 예컨대 다양한 V 분절의 사용을 평가할 수 있을 만큼 충분히 독특하다. 상기 용어 시그니처라는 특정 핵산 서열인 특이적 시그니처라는 용어를 포함한다. 한 구현예에서, 상기 시그니처는 특정 이벤트, 예컨대 재배열 이벤트를 나타내거나 이의 결과이다.
- [0220] 본원에서 사용된 용어 "하위 유전체 간격"은 유전체 서열의 일부를 지칭한다. 한 구현예에서, 하위 유전체 간격은 단일 뉴클레오티드 위치일 수 있으며, 예컨대 해당 위치의 변이체는 중앙 표현형과 (양성 또는 음성으로)연

관되어 있다. 한 구현예에서, 하위 유전체 간격은 하나 초과와 뉴클레오타이드 위치를 포함한다. 상기 구현예들은 적어도 2, 5, 10, 50, 100, 150, 또는 250개의 뉴클레오타이드 위치의 길이를 갖는 서열을 포함한다. 하위 유전체 간격은 전체 유전자 또는 이의 일부, 엑손 부호화 영역(또는 이의 일부), 인트론(또는 이의 일부) 또는 엑손(또는 이의 일부)을 포함할 수 있다. 하위 유전체 간격은 자연 발생, 예컨대 유전체 DNA, 핵산의 단편의 전부 또는 일부를 포함할 수 있다. 예컨대, 하위 유전체 간격은 서열분석 반응을 거치는 유전체 DNA의 단편에 해당할 수 있다. 한 구현예에서, 하위 유전체 간격은 유전체 공급원로부터의 연속적인 서열이다. 한 구현예에서, 하위 유전체 간격은 유전체에서 인접하지 않은 서열을 포함하고, 예컨대 cDNA의 하위 유전체 간격은 스플라이싱의 결과로 형성된 엑손-엑손 접합을 포함할 수 있다. 한 구현예에서, 상기 하위 유전체 간격은 종양 핵산 분자를 포함한다. 한 구현예에서, 상기 하위 유전체 간격은 비종양 핵산 분자를 포함한다.

[0221] 한 구현예에서, 하위 유전체 간격은 재배열된 서열, 예컨대 V 분절에서 D 분절로, D 분절에서 J 분절로, V 분절에서 J 분절로. 또는 J 분절에서 클래스 분절로 결합한 결과로서 발생한 B 또는 T 세포 내 서열에 해당한다.

[0222] 한 구현예에서, 상기 하위 유전체 간격은 하나의 서열로 표현된다. 한 구현예에서, 상기 하위 유전체 간격은 하나 초과와 서열로 표시되며, 예컨대 VD 서열을 포함하는 하위 유전체 간격은 하나 초과와 시그니처로 표시될 수 있다.

[0223] 한 구현예에서, 하위 유전체 간격은 단일 뉴클레오타이드 위치; 유전자내 영역 또는 유전자간 영역; 엑손 또는 인트론, 또는 이의 단편, 전형적으로 엑손 서열 또는 이의 단편; 부호화 영역 또는 비부호화 영역, 예컨대, 프로모터, 인핸서, 5' 비번역 영역(5' UTR), 또는 3' 비번역 영역(3' UTR), 또는 이의 단편; cDNA 또는 이의 단편; SNP; 체세포 돌연변이, 생식계열 돌연변이 또는 둘 다; 변형, 예컨대 점 또는 단일 돌연변이; 결실 돌연변이(예컨대, 틀널 결실, 유전자내 결실, 전체 유전자 결실); 삽입 돌연변이(예컨대, 유전자내 삽입); 역위 돌연변이(예컨대, 염색체내 역위); 역위 중복 돌연변이; 직렬 중복(예컨대, 염색체내 직렬 중복); 전좌(예컨대, 염색체 전좌, 비가역적 전좌); 재배열(예컨대, 유전체 재배열(예컨대, 하나 이상의 인트론의 재배열, 하나 이상의 엑손의 재배열, 또는 이들의 조합 및/또는 단편; 재배열된 인트론은 5' - 및/또는 3' -UTR을 포함할 수 있음); 유전자 복제 수의 변화; 유전자 발현의 변화; RNA 수준의 변화; 또는 이들의 조합을 포함하거나 이로써 구성된다. 상기 "유전자의 복제 수"는 특정 유전자 산물을 부호화하는 세포 내 DNA 서열의 수를 지칭한다. 일반적으로, 주어진 유전자에 대해 포유동물은 각 유전자에 대해 2개의 복제를 가진다. 상기 복제 수는, 예컨대 유전자 중복 또는 중복에 의해 증가되거나, 또는 결실에 의해 감소될 수 있다.

[0224] 본원에서 사용된 용어 "대상체 간격"은 하위 유전체 간격 또는 발현된 하위 유전체 간격을 지칭한다. 한 구현예에서, 하위 유전체 간격 및 발현된 하위 유전체 간격은 서로 대응하며, 이는 발현된 하위 유전체 간격이 대응하는 하위 유전체 간격으로부터 발현된 서열을 포함함을 의미한다. 한 구현예에서, 하위 유전체 간격 및 발현된 하위 유전체 간격은 서로 대응하지 않으며, 이는 발현된 하위 유전체 간격이 대응하지 않는 하위 유전체 간격으로부터 발현된 서열을 포함하지 않고, 오히려 상이한 하위 유전체 간격에 대응함을 의미한다. 한 구현예에서, 하위 유전체 간격 및 발현된 하위 유전체 간격은 서로 부분적으로 대응하며, 이는 발현된 하위 유전체 간격이 대응하는 하위 유전체 간격으로부터 발현된 서열 및 상이한 대응하는 하위 유전체 간격으로부터 발현된 서열을 포함함을 의미한다.

[0225] 본원에 사용된 용어 "라이브러리"는 핵산 분자의 집합체를 지칭한다. 일 구현예에서, 상기 라이브러리는 핵산 분자들의 집합체, 예컨대 전체 유전체, 하위 유전체 단편들, cDNA, cDNA 단편들, RNA, 예컨대 mRNA, RNA 단편들, 또는 이들의 조합의 집합체를 포함한다. 전형적으로, 핵산 분자는 DNA 분자, 예컨대 유전체 DNA 또는 cDNA이다. 핵산 분자는 단편화될 수 있고, 예컨대 전단되거나 효소적으로 제조된 유전체 DNA일 수 있다. 핵산 분자는 대상체로부터의 서열을 포함하고 또한 상기 대상체로부터 유래되지 않은 서열, 예컨대 어댑터 서열, 프라이머 서열, 또는 식별을 허용하는 기타 서열들, 예컨대 "바코드" 서열을 포함할 수 있다. 일 구현예에서, 라이브러리 핵산 분자들의 일부 또는 전부는 어댑터 서열을 포함한다. 상기 어댑터 서열은 한쪽 또는 양쪽 끝에 위치할 수 있다. 상기 어댑터 서열은, 예컨대 서열분석 방법(예컨대, NGS 방법), 증폭, 역전사 또는 벡터 내로의 클로닝에 유용할 수 있다. 상기 라이브러리는 핵산 분자들의 집합체, 예컨대 표적 핵산 분자(예컨대, 종양 핵산 분자, 참조 핵산 분자, 또는 이들의 조합)를 포함할 수 있다. 상기 라이브러리의 핵산 분자들은 단일 개체로부터 유래할 수 있다. 구현예들에서, 라이브러리는 하나 초과와 대상체(예컨대, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30개 이상의 대상체)로부터의 핵산 분자들을 포함할 수 있고, 예컨대 상이한 대상체들로부터의 2개 이상의 라이브러리는 1개 초과와 대상체로부터의 핵산 분자들을 포함하는 라이브러리를 형성하기 위해 조합될 수 있다. 일 구현예에서, 상기 대상체는 암 또는 종양을 갖거나 가질 위험이 있는 인간이다.

- [0226] "라이브러리 캐치(Library catch)"는 라이브러리의 하위 집합, 예컨대 대상체 간격에 대해 농축된 하위 집합, 예컨대 표적 포획 시약과의 혼성화에 의해 포획된 생성물을 지칭한다.
- [0227] 본원에 사용된 "표적 포획 시약"은 표적을 포획할 수 있는 분자를 지칭한다. 표적 포획 시약(예컨대, 미끼 또는 표적 포획 올리고뉴클레오타이드)은 핵산 분자, 예컨대 DNA 또는 RNA 분자를 포함할 수 있으며, 혼성화되어(예컨대, 상보적이 되어) 표적 핵산의 포획을 가능하게 할 수 있다. 한 구현예에서, 표적 포획 시약은 DNA 분자(예컨대, 자연 발생 또는 변형된 DNA 분자), RNA 분자(예컨대, 자연 발생 또는 변형된 RNA 분자), 또는 이들의 조합을 포함한다. 일 구현예에서, 표적 포획 시약은 용액상 혼성화에 적합하다.
- [0228] "상보적"은 2개의 핵산 가닥의 영역 사이 또는 동일한 핵산 가닥의 2개 영역 사이의 서열 상보성을 지칭한다. 제1 핵산 영역의 아데닌 잔기는 상기 잔기가 티민 또는 우라실인 경우 상기 제1 영역에 역평행인 제2 핵산 영역의 잔기와 특이적 수소 결합("염기쌍 형성")을 형성할 수 있는 것으로 공지되어 있다. 유사하게 제1 핵산 가닥의 시토신 잔기는 상기 잔기가 구아닌인 경우 상기 제1 가닥에 역평행인 제2 핵산 가닥의 잔기와 염기쌍을 형성할 수 있는 것으로 공지되어 있다. 핵산의 제1 영역은 동일하거나 상이한 핵산의 제2 영역에 상보적이다. 만약 상기 두 영역이 역평행 방식으로 배열되고, 상기 제1 영역의 적어도 하나의 뉴클레오타이드 잔기가 상기 제2 영역의 잔기와 염기쌍을 형성할 수 있는 경우에 그러하다. 특정 구현예들에서, 상기 제1 영역은 제1 부분을 포함하고 상기 제2 영역은 제2 부분을 포함함에 따라, 상기 제1 및 제2 부분이 역평행 방식으로 배열될 때, 상기 제1 부분의 뉴클레오타이드 잔기의 약 50% 이상, 약 75% 이상, 약 90% 이상, 또는 약 95% 이상이 상기 제2 부분의 뉴클레오타이드 잔기와 염기쌍을 형성할 수 있다. 다른 구현예들에서, 상기 제1 부분의 모든 뉴클레오타이드 잔기는 상기 제2 부분의 뉴클레오타이드 잔기와 염기쌍을 형성할 수 있다.
- [0229] 용어 "암" 및 "종양"은 본원에서 상호교환적으로 사용된다. 상기 용어는 암 유발 세포의 전형적인 특징들, 예컨대 통제되지 않는 증식, 불멸, 전이 가능성, 빠른 성장 및 증식 속도 그리고 독특한 특정 형태학적 특징들을 갖는 세포들의 존재를 지칭한다. 암세포는 종종 종양의 형태로 존재하지만, 상기 세포는 동물 내에 단독으로 존재할 수 있거나, 또는 백혈병 세포와 같은 비종양유발성 암세포일 수 있다. 상기 용어는 고형 종양, 연조직 종양 또는 전이성 병변을 포함한다. 본원에서 사용된 용어 "암"은 전암성 뿐만 아니라 악성 암을 포함한다.
- [0230] 본원에서 사용된 "가능성이 있는" 또는 "가능성이 증가된"은 항목, 대상, 사물 또는 사람이 발생할 확률의 증가를 지칭한다. 따라서, 일례에서, 치료에 반응할 가능성이 있는 대상체는 참조 대상체 또는 대상체군에 대비하여 치료에 반응할 확률이 증가한다.
- [0231] "가능성이 낮은"은 참조와 관련하여 이벤트, 항목, 대상, 사물 또는 사람이 발생할 확률의 감소를 지칭한다. 따라서, 치료에 반응할 가능성이 낮은 대상체는 참조 대상체 또는 대상체군에 대비하여 치료에 반응할 확률이 감소한다.
- [0232] "대조군 핵산 분자"는 비종양 세포로부터의 서열을 갖는 핵산 분자를 지칭한다.
- [0233] 본원에서 사용된 "차세대 서열분석" 또는 "NGS" 또는 "NG 서열분석"은 개별 핵산 분자(예컨대, 단일 분자 서열 분석에서) 또는 개별 핵산에 대한 클론 확장된 대리물의 뉴클레오타이드 서열을 고처리량 방식으로(예컨대,  $10^3$ ,  $10^4$ ,  $10^5$ 개 이상의 분자가 동시에 서열분석되는) 결정하는 임의의 서열분석 방법을 지칭한다. 일 구현예에서, 라이브러리 내 핵산 종의 상대적 존재량은 서열분석 실험에 의해 생성된 데이터에서 이의 동족 서열의 상대적 발생 횟수를 계수함으로써 추정할 수 있다. 차세대 서열분석 방법은 당업계에 공지되어 있고, 예컨대 Metzker, M. (2010) *Nature Biotechnology Reviews* 11:31-46에 기재되어 있으며, 본원에 참조로서 인용된다. 차세대 서열분석은 샘플에서 핵산의 5% 미만 또는 1% 미만에 존재하는 변이체를 검출할 수 있다.
- [0234] 본원에서 언급된 "뉴클레오타이드 값"은 뉴클레오타이드 위치를 점유하거나 이에 할당된 뉴클레오타이드(들)의 동일성을 나타낸다. 전형적인 뉴클레오타이드 값은: 누락(예컨대, 삭제); 추가(예컨대, 하나 이상의 뉴클레오타이드의 삽입, 이의 동일성은 포함되거나 포함되지 않을 수 있음); 또는 존재(점유); A; T; C; 또는 G를 포함한다. 다른 값들은, 예컨대 Y가 아닐 수 있고(Y는 A, T, G 또는 C임); A 또는 X(X는 T, G 또는 C 중 1개 또는 2개임); T 또는 X(X는 A, G 또는 C 중 1개 또는 2개임); G 또는 X(X는 T, A 또는 C 중 1개 또는 2개임); C 또는 X(X는 T, G 또는 A 중 1개 또는 2개임); 피리미딘 뉴클레오타이드; 또는 퓨린 뉴클레오타이드. 뉴클레오타이드 값은 뉴클레오타이드 위치에서 1개 이상, 예컨대 2, 3, 또는 4개의 염기(또는 본원에 기재된 다른 값들, 예컨대 누락 또는 추가)에 대한 빈도일 수 있다. 예컨대, 뉴클레오타이드 값은 뉴클레오타이드 위치에서 A에 대한 빈도 및 G에 대한 빈도를 포함할 수 있다.

- [0235] "또는"은 문맥상 명백하게 다르게 나타나지 않는 한, 용어 "및/또는"을 의미하기 위해 본원에 사용되며, 서로 상호교환적으로 사용된다. 본원의 일부 위치에서 용어 "및/또는"의 사용은 문맥상 명백하게 다르게 나타나지 않는 한, 용어 "또는"의 사용이 용어 "및/또는"과 상호 교환될 수 없음을 의미하지 않는다.
- [0236] "일차 대조군"은 샘플 내 정상 인접 조직(NAT) 이외의 비종양 조직을 지칭한다. 혈액은 전형적인 1차 대조군이다.
- [0237] 본원에서 사용된 "샘플"은 본원에 기재된 바와 같이 관심 공급원으로부터 수득하거나 유래된 생물학적 샘플을 지칭한다. 일부 구현예들에서, 관심 공급원은 유기체, 예컨대 동물 또는 인간을 포함한다. 상기 샘플의 공급원은 신선, 냉동 및/또는 보존된 장기의 고형 조직, 조직 샘플, 생검, 절제, 도말 또는 흡인물; 혈액 또는 임의의 혈액 성분; 체액, 예컨대 뇌척수액, 양수, 복막액 또는 간질액; 또는 대상체의 임신 또는 발달 중 어느 시점의 세포일 수 있다. 일부 구현예들에서, 상기 샘플의 공급원은 혈액 또는 혈액 성분이다.
- [0238] 일부 구현예들에서, 상기 샘플은 생물학적 조직 또는 유체이거나 이를 포함한다. 상기 샘플은 보존제, 항응고제, 완충제, 고정제, 영양소, 항생제 등과 같이 사실상 자연적으로는 상기 조직과 혼합되지 않는 화합물을 함유할 수 있다. 한 구현예에서, 상기 샘플은 동결 샘플로서 또는 폼알데하이드 또는 파라포름알데하이드 고정 파라핀 포매(FPPE) 조직 표본으로서 보존된다. 예를 들어, 상기 샘플은 매트릭스, 예컨대 FPPE 블록 또는 동결 샘플 내에 포함될 수 있다. 다른 구현예에서, 상기 샘플은 혈액 또는 혈액 성분 샘플이다. 또 다른 구현예에서, 상기 샘플은 골수 흡인물 샘플이다. 다른 구현예에서, 상기 샘플은 무세포 DNA(cfDNA)를 포함한다. 일부 구현예들에서, cfDNA는 세포자멸을 겪고 있는 세포 또는 괴사 세포로부터의 DNA이다. 일반적으로, cfDNA는 단백질(예컨대, 히스톤)에 의해 결합되고 뉴클레아제에 의해 보호된다. CfDNA는 비침습적 산전 검사(NIPT), 장기 이식, 심근병증, 미생물군유전체 및 암에 대한 바이오마커로 사용될 수 있다. 다른 구현예에서, 상기 샘플은 순환 종양 DNA(ctDNA)를 포함한다. 일부 구현예들에서, ctDNA는 종양 세포 대 비종양 세포로부터 기원하는 것을 구별할 수 있는 유전적 또는 후성적 변경(예컨대, 체세포 변경 또는 메틸화 특징)을 갖는 cfDNA이다. 다른 구현예에서, 상기 샘플은 순환 종양 세포(CTCs)를 포함한다. 일부 구현예들에서, CTC는 원발성 또는 전이성 종양으로부터 순환계로 배출된 세포이다. 일부 구현예들에서, CTC 세포자멸사는 혈액/림프 내에서 ctDNA의 공급원이다.
- [0239] 일부 구현예들에서, 생물학적 샘플은 골수; 혈액; 혈액 세포; 복수; 조직 또는 미세 바늘 생검 샘플; 세포 함유 체액; 자유 부유 핵산; 담; 타액; 소변; 뇌척수액, 복막액; 흉막액; 대변; 림프; 부인과 체액; 피부 면봉; 질 면봉; 구강 면봉; 비강 면봉; 관 세척액 또는 기관지폐포 세척액과 같은 세척액; 흡인물; 찰과 표본; 골수 표본; 조직 생검 표본; 수술 표본; 대변, 기타 체액, 분비물 및/또는 배설물; 및/또는 이들로부터의 세포 등일 수 있거나 이를 포함할 수 있다. 일부 구현예들에서, 생물학적 샘플은 개체로부터 수득한 세포이거나 이를 포함한다. 일부 구현예들에서, 수득한 세포는 상기 샘플을 수득한 개체로부터의 세포이거나 이를 포함한다.
- [0240] 일부 구현예들에서, 샘플은 임의의 적절한 수단에 의해 관심 공급원으로부터 직접 수득한 "일차 샘플"이다. 예를 들어, 일부 구현예들에서, 1차 생물학적 샘플은 생검(예컨대, 미세 바늘 흡인 또는 조직 생검), 수술, 체액 수집(예컨대, 혈액, 림프 또는 대변) 등으로부터 선택된 방법에 의해 수득한다. 일부 구현예들에서, 문맥상 명백한 바와 같이, 용어 "샘플"은 1차 샘플, 예컨대 1차 샘플을 처리함으로써(예컨대, 1개 이상의 성분을 제거하고/하거나 1개 이상의 체제를 추가함으로써), 예컨대 반투과막을 이용한 여과에 의해 수득한 제조물을 지칭한다. 상기 "가공된 샘플"은, 예를 들어 샘플로부터 추출되거나 mRNA의 증폭 또는 역전사, 특정 성분의 분리 및/또는 정제 등과 같은 기술에 1차 샘플을 적용함으로써 수득한 핵산 또는 단백질을 포함할 수 있다.
- [0241] 한 구현예에서, 상기 샘플은 종양, 예컨대 종양 세포 또는 종양 침윤 림프구(TIL)와 관련된 세포이다. 일 구현예에서, 상기 샘플은 하나 이상의 전암성 또는 악성 세포를 포함한다. 한 구현예에서, 상기 샘플은 혈액 악성종양(또는 전암), 예컨대 본원에 기재된 혈액 악성종양(또는 전암)으로부터 수득한다. 특정 구현예들서, 상기 샘플은 고형 종양, 연조직 종양, 또는 전이성 병변으로부터 수득한다. 다른 구현예들에서, 상기 샘플은 수술 절제면으로부터의 조직 또는 세포를 포함한다. 다른 구현예에서, 상기 샘플은 하나 이상의 순환 종양 세포(CTC)(예컨대, 혈액 샘플로부터 수득한 CTC)를 포함한다. 한 구현예에서, 상기 샘플은 종양, 예컨대 비종양 세포 또는 말초 혈액과 관련되지 않은 세포이다.
- [0242] 본원에서 사용된 "민감도"는 이중 서열 집단에서 서열 변이체를 검출하는 방법의 능력의 척도이다. 만약 상기 서열 변이체가 샘플 내에서 서열들의 적어도 F%로서 존재하는 샘플을 고려할 때, 한 방법이 해당 시간의 C% ST%의 신뢰도로 상기 서열을 검출할 수 있다면, 상기 방법은 F%의 변이체에 대해 ST%의 민감도를 가진다. 예로서, 만약 상기 변이체 서열이 샘플 내에서 서열들의 적어도 5%로서 존재하는 샘플을 고려할 때, 한 방법이 10번 중 9번에서 99%의 신뢰도로 상기 서열을 검출할 수 있다면, 상기 방법은 5%의 변이체에 대해 90%의 민감도를 가진

다( $F=5\%$ ;  $C=99\%$ ;  $ST=90\%$ ). 예시적인 민감도는  $C=90\%$ ,  $95\%$ ,  $99\%$  및  $99.9\%$ 의 신뢰 수준에서  $F=1\%$ ,  $5\%$ ,  $10\%$ ,  $20\%$ ,  $50\%$ ,  $100\%$ 에서 서열 변이체에 대한  $ST=90\%$ ,  $95\%$ ,  $99\%$ 의 민감도를 포함한다.

[0243] 본원에서 사용된 "특이성"은 서열분석 결과물 또는 기타 밀접하게 관련된 서열로부터 실제로 발생하는 서열 변이체를 구별하는 방법의 능력의 척도이다. 위양성 검출을 피하는 능력이다. 위양성 검출은 샘플 제조 중 관심 서열에 도입된 오류, 서열분석 오류 또는 유전자군의 유사 유전자 또는 핵산 분자와 같이 밀접하게 관련된 서열의 부주의한 서열분석으로 인해 발생할 수 있다.  $X_{\text{진실}}$  서열이 진정한 변이체이고  $X_{\text{진실}}$  아님이 진정한 변이체가 아닌  $N_{\text{진실}}$  서열의 샘플 세트에 적용할 때, 한 방법이 진정한 변이체가 아닌 것의 적어도  $X\%$ 를 선택하는 경우, 상기 방법은  $X\%$ 의 특이성을 가진다. 예컨대, 500개 서열이 진정한 변이체이고 500개는 진정한 변이체가 아닌 1,000개 서열의 샘플 세트에 적용할 때, 한 방법이 500개의 진정한 변이체가 아닌 서열 중  $90\%$ 를 변이체가 아닌 것으로 선택하는 경우, 상기 방법은  $90\%$ 의 특이성을 가진다. 예시적인 특이성은  $90$ ,  $95$ ,  $98$ , 및  $99\%$ 를 포함한다.

[0244] 본원에서 사용된 "대조군 핵산" 또는 "참조 핵산"은 대조군 또는 기준 샘플로부터의 핵산 분자를 지칭한다. 전형적으로, 이는 유전자 또는 유전자 산물의 변경 또는 변이를 포함하지 않는 DNA, 예컨대 유전체 DNA, 또는 RNA로부터 유래된 cDNA이다. 특정 구현예들에서, 상기 참조 또는 대조군 핵산 샘플은 야생형 또는 돌연변이되지 않은 서열이다. 특정 구현예들에서, 상기 참조 핵산 샘플은 정제되거나 단리된다(예컨대, 자연 상태에서 제거된다). 다른 구현예들에서, 상기 참조 핵산 샘플은 혈액 대조군, 정상 인접 조직(NAT), 또는 동일하거나 상이한 대상체로부터의 임의의 다른 비암성 샘플로부터의 것이다. 일부 구현예들에서, 상기 참조 핵산 샘플은 정상 DNA 혼합물을 포함한다. 일부 구현예들에서, 상기 정상 DNA 혼합물은 공정 일치된 대조군이다. 일부 구현예들에서, 상기 참조 핵산 샘플은 생식계열 변이체를 갖는다. 일부 구현예들에서, 상기 참조 핵산 샘플은 체세포 변이를 갖지 않으며, 예컨대 음성 대조군으로서 작용한다.

[0245] 핵산 분자를 "서열분석"하려면 분자(예컨대, DNA 분자, RNA 분자 또는 RNA 분자로부터 유래된 cDNA 분자)에서 1개 이상의 뉴클레오타이드의 동일성을 결정해야 한다. 구현예들에서, 분자 내의 모든 뉴클레오타이드보다 적은 뉴클레오타이드의 동일성이 결정된다. 다른 구현예들에서, 상기 분자 내의 다수 또는 모든 뉴클레오타이드의 동일성이 결정된다.

[0246] 본원에서 사용된 "임계값"은 대상체 간격(예컨대, 하위 유전체 간격 또는 발현된 하위 유전체 간격)에 뉴클레오타이드 값을 할당하기 위해 존재하는 데 필요한 리드 수의 함수인 값이다. 예컨대, 이는 하위 유전체 간격에서 뉴클레오타이드 위치에 뉴클레오타이드 값을 할당하는데 필요한, 상기 뉴클레오타이드 위치에서 특정 뉴클레오타이드 값, 예컨대 "A"를 갖는 리드 수의 함수이다. 상기 임계값은, 예컨대 리드의 수, 예컨대 정수로서(또는 이의 함수로서), 또는 상기 값을 갖는 리드의 비율로서 표현될 수 있다. 예로서, 상기 임계값이  $X$ 이고 "A"의 뉴클레오타이드 값을 갖는  $X+1$  리드가 존재하는 경우 "A"의 값은 대상체 간격(예컨대, 하위 유전체 간격 또는 발현된 하위 유전체 간격) 내 위치에 할당된다. 상기 임계값은 또한 돌연변이 또는 변이 기대치, 돌연변이 빈도, 또는 베이저안 사전(Bayesian prior)의 함수로서 표현될 수 있다. 구현예에서, 돌연변이 빈도는 해당 뉴클레오타이드 값을 호출하기 위해 위치에서 뉴클레오타이드 값, 예컨대 A 또는 G를 갖는 리드의 수 또는 비율을 필요로 한다. 구현예들에서, 상기 임계값은 돌연변이 기대치, 예컨대 돌연변이 빈도 및 종양 유형의 함수일 수 있다. 예컨대, 뉴클레오타이드 위치의 변이는 환자가 제1 종양 유형을 갖는 경우 제1 임계값을 가질 수 있고 환자가 제2 종양 유형을 갖는 경우 제2 임계값을 가질 수 있다.

[0247] 본원에서 사용된 "표적 핵산 분자"는 핵산 라이브러리로부터 단리하고자 하는 핵산 분자를 지칭한다. 일 구현예에서, 상기 표적 핵산 분자는 본원에 기재된 바와 같은 종양 핵산 분자, 참조 핵산 분자, 또는 대조군 핵산 분자일 수 있다.

[0248] 본원에서 사용된 "종양 핵산 분자" 또는 기타 유사한 용어(예컨대, "종양 또는 암 관련 핵산 분자")는 종양 세포로부터의 서열을 갖는 핵산 분자를 지칭한다. 용어 "종양 핵산 분자" 및 "종양 핵산"은 때때로 본원에서 상호 교환적으로 사용될 수 있다. 일 구현예에서, 상기 종양 핵산 분자는 암성 표현형과 관련된 변경(예컨대, 돌연변이)을 갖는 서열(예컨대, 뉴클레오타이드 서열)을 갖는 대상체 간격을 포함한다. 다른 구현예들에서, 상기 종양 핵산 분자는 야생형 서열(예컨대, 야생형 뉴클레오타이드 서열)을 갖는 대상체 간격을 포함한다. 예를 들어, 이형 접합 또는 동형접합 야생형 대립유전자로부터의 대상체 간격이 암 세포 내에 존재한다. 종양 핵산 분자는 참조 핵산 분자를 포함할 수 있다. 전형적으로, 이는 샘플로부터의 DNA, 예컨대 유전체 DNA, 또는 RNA로부터 유래된 cDNA이다. 특정 구현예들에서, 상기 샘플은 정제되거나 단리된다(예컨대, 자연 상태에서 제거된다). 일부 구현예들에서, 상기 종양 핵산 분자는 cfDNA이다. 일부 구현예들에서, 상기 종양 핵산 분자는 ctDNA이다. 일부 구현

예들에서, 상기 종양 핵산 분자는 CTC로부터의 DNA이다.

- [0249] 본원에서 사용된 "참조 핵산 분자" 또는 다른 유사한 용어(예컨대, "대조군 핵산 분자")는 양성 표현형과 관련이 없는 서열(예컨대, 뉴클레오티드 서열)을 갖는 대상체 간격을 포함하는 핵산 분자를 지칭한다. 일 구현예에서, 상기 참조 핵산 분자는 돌연변이된 경우 양성 표현형과 관련된 유전자 또는 유전자 산물의 야생형 또는 돌연변이되지 않은 뉴클레오티드 서열을 포함한다. 상기 참조 핵산 분자는 암세포 또는 비암세포 내에 존재할 수 있다.
- [0250] 본원에서 사용된 "변이"는 1개 초과 구조, 예컨대 다형성 유전자좌에서의 대립유전자를 가질 수 있는 하위 유전체 간격에 존재할 수 있는 구조를 지칭한다.
- [0251] "분리된" 핵산 분자는 상기 핵산 분자의 천연 공급원에 존재하는 다른 핵산 분자로부터 분리된 것이다. 특정 구현예들에서, "분리된" 핵산 분자는 핵산이 유래된 유기체의 유전체 DNA에서 상기 핵산(즉, 상기 핵산의 5' 및 3' 말단에 위치하는 서열)의 측면에 자연적으로 인접하는 서열(예컨대, 단백질 부호화 서열)이 없다. 예를 들어, 다양한 구현예에서, 상기 분리된 핵산 분자는 상기 핵산이 유래된 세포의 유전체 DNA에서 상기 핵산 분자의 측면에 자연적으로 인접하는 약 5 kB 미만, 약 4 kB 미만, 약 3 kB 미만, 약 2 kB 미만, 약 1 kB 미만, 약 0.5 kB 미만 또는 약 0.1 kB 미만의 뉴클레오티드 서열을 함유할 수 있다. 또한, RNA 분자 또는 cDNA 분자와 같은 "분리된" 핵산 분자는, 예컨대 재조합 기술에 의해 생성될 때 다른 세포 물질 또는 배양 배지가 실질적으로 없을 수 있거나 또는, 예컨대 화학적으로 합성될 때 화학적 전구체 또는 기타 화합물이 실질적으로 없을 수 있다.
- [0252] "다른 세포 물질 또는 배양 배지가 실질적으로 없는"이라는 표현은 이로부터 분리되거나 재조합적으로 생성되는 세포의 세포 성분으로부터 상기 분자가 분리된 핵산 분자의 제제를 포함한다. 따라서, 세포 물질이 실질적으로 없는 핵산 분자는 약 30% 미만, 약 20% 미만, 약 10% 미만, 또는 약 5% 미만(건조 중량 기준)의 다른 세포 물질 또는 배양 배지를 갖는 핵산 분자의 제제를 포함한다.
- [0253] 본원에서 사용된, "X는 Y의 함수이다"는, 예컨대 하나의 변수 X가 다른 변수 Y와 연관됨을 의미한다. X와 Y 사이의 연관은 직접적 또는 간접적일 수 있다. 일 구현예에서, X가 Y의 함수인 경우, X와 Y 사이의 인과 관계가 암시될 수 있지만, 반드시 존재하는 것은 아니다.
- [0254] 예컨대, (a), (b), (i) 등의 표제는 명세서 및 청구범위를 쉽게 읽을 수 있도록 제공된다. 명세서 또는 청구범위에서 표제를 사용하는 것은 단계나 요소를 알파벳순이나 숫자순 또는 제시된 순서로 실시할 것을 요구하지 않는다. 명세서 또는 청구범위에서 표제를 사용한다고 해서 모든 단계 또는 요소의 실시가 필요한 것은 아니다.
- [0255] 다중 유전자 분석
- [0256] 본원에 기재된 방법은, 예컨대 본원에 기재된 유전자 또는 유전자 산물의 세트로부터 대상체 간격 세트를 평가하기 위한 방법과 조합하여 또는 그 일부로서 사용될 수 있다.
- [0257] 특정 구현예들에서, 상기 유전자 세트는 돌연변이 형태로 세포 분열, 성장 또는 생존에 대한 효과와 연관되거나, 또는 암, 예컨대 본원에 기재된 암과 연관되는 복수의 유전자를 포함한다.
- [0258] 특정 구현예들에서, 상기 유전자 세트는, 예컨대 본원에 기재된 바와 같이 적어도 약 50개 이상, 약 100개 이상, 약 150개 이상, 약 200개 이상, 약 250개 이상, 약 300개 이상, 약 350개 이상, 약 400개 이상, 약 450개 이상, 약 500개 이상, 약 550개 이상, 약 600개 이상, 약 650개 이상, 약 700개 이상, 약 750개 이상, 또는 약 800개 이상의 유전자를 포함한다. 일부 구현예들에서, 상기 유전자 세트는 표 2A-5B에 기재된 적어도 약 50개 이상, 약 100개 이상, 약 150개 이상, 약 200개 이상, 약 250개 이상, 약 300개 이상, 또는 모든 유전자를 포함한다.
- [0259] 특정 구현예들에서, 상기 방법은 샘플로부터 복수의 종양 핵산 분자를 포함하는 라이브러리를 수득하는 단계를 포함한다. 특정 구현예들에서, 상기 방법은 라이브러리를 표적 포획 시약과 접촉시켜 선택된 종양 핵산 분자를 제공하는 단계를 추가로 포함하며, 상기 표적 포획 시약은 상기 라이브러리로부터의 종양 핵산 분자와 혼성화하여 라이브러리 캐치(catch)를 제공한다. 특정 구현예들에서, 상기 방법은 라이브러리 또는 라이브러리 캐치로부터의 종양 핵산 분자로부터의 변경(예컨대, 체세포 변경)을 포함하는 대상체 간격에 대한 리드를 수득함으로써, 예컨대 차세대 서열분석 방법에 의해 상기 대상체 간격에 대한 리드를 수득하는 단계를 포함한다. 특정 구현예들에서, 상기 방법은 정렬 방법, 예컨대 본원에 기재된 정렬 방법에 의해 대상체 간격에 대한 리드를 정렬하는 단계를 추가로 포함한다. 특정 구현예들에서, 상기 방법은, 예컨대 본원에 기재된 돌연변이 호출 방법에 의해

대상체 간격에 대한 리드로부터 뉴클레오티드 위치에 대한 뉴클레오티드 값을 할당하는 단계를 추가로 포함한다.

- [0260] 특정 구현예들에서, 상기 방법은 하기의 단계들로서:
- [0261] (a) 샘플로부터 복수의 종양 핵산 분자를 포함하는 라이브러리를 수득하는 단계;
- [0262] (b) 상기 라이브러리를 복수의 표적 포획 시약과 접촉시켜 선택된 종양 핵산 분자들을 제공하는 단계로서, 상기 복수의 표적 포획 시약은 상기 종양 핵산 분자들과 혼성화되어 라이브러리 캐치를 제공하는 단계;
- [0263] (c) 상기 라이브러리 캐치로부터의 종양 핵산 분자로부터의 변경(예컨대, 체세포 변경)을 포함하는 대상체 간격에 대한 리드를 수득함으로써, 예컨대 차세대 서열분석 방법에 의해 상기 대상체 간격에 대한 리드를 수득하는 단계;
- [0264] (d) 정렬 방법, 예컨대 본원에 기재된 정렬 방법에 의해 상기 리드를 정렬하는 단계; 또는
- [0265] (e) 예컨대, 본원에 기재된 돌연변이 호출 방법에 의해 상기 리드로부터 뉴클레오티드 위치에 대한 뉴클레오티드 값을 할당하는 단계로부터 1, 2, 3, 4개의 단계 또는 모두를 포함한다.
- [0266] 특정 구현예들에서, 상기 대상체 간격에 대한 리드를 수득하는 단계는 적어도 약 50개 이상, 약 100개 이상, 약 150개 이상, 약 200개 이상, 약 250개 이상, 약 300개 이상, 350개 이상, 약 400개 이상, 약 450개 이상, 약 500개 이상, 약 550개 이상, 약 600개 이상, 약 650개 이상, 약 700개 이상, 약 750개 이상, 또는 약 800개 이상의 유전자로부터 대상체 간격을 서열분석하는 단계를 포함한다. 특정 구현예들에서, 상기 대상체 간격으로부터 리드를 수득하는 단계는 표 2A-5B에 기재된 적어도 약 50개 이상, 약 100개 이상, 약 150개 이상, 약 200개 이상, 약 250개 이상, 약 300개 이상, 또는 모든 유전자로부터 대상체 간격을 서열분석하는 단계를 포함한다.
- [0267] 특정 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 100배 이상의 평균 정도로 서열분석하는 것을 포함한다. 특정 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 약 250배 이상의 평균 정도로 서열분석하는 것을 포함한다. 다른 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 약 500배 이상의 평균 정도로 서열분석하는 것을 포함한다. 특정 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 약 800배 이상의 평균 정도로 서열분석하는 것을 포함한다. 다른 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 약 1,000배 이상의 평균 정도로 서열분석하는 것을 포함한다. 다른 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 약 1,500배 이상의 평균 정도로 서열분석하는 것을 포함한다. 다른 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 약 2,000배 이상의 평균 정도로 서열분석하는 것을 포함한다. 다른 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 약 2,500배 이상의 평균 정도로 서열분석하는 것을 포함한다. 특정 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 약 3,000배 이상의 평균 정도로 서열분석하는 것을 포함한다. 특정 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 약 3,500배 이상의 평균 정도로 서열분석하는 것을 포함한다. 특정 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 약 4,000배 이상의 평균 정도로 서열분석하는 것을 포함한다. 특정 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 약 4,500배 이상의 평균 정도로 서열분석하는 것을 포함한다. 특정 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 약 5,000배 이상의 평균 정도로 서열분석하는 것을 포함한다. 특정 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 약 5,500배 이상의 평균 정도로 서열분석하는 것을 포함한다. 특정 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 약 6,000배 이상의 평균 정도로 서열분석하는 것을 포함한다.
- [0268] 특정 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 서열분석된 유전자(예컨대, 엑손)의 약 99%를 초과하는 약 100배 이상의 평균 정도로 서열분석하는 것을 포함한다. 특정 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 서열분석된 유전자(예컨대, 엑손)의 약 99%를 초과하는 약 250배 이상의 평균 정도로 서열분석하는 것을 포함한다. 다른 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 서열분석된 유전자(예컨대, 엑손)의 약 95%를 초과하는 약 500배 이상의 평균 정도로 서열분석하는 것을 포함한다. 다른 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 서열분석된 유전자(예컨대, 엑손)의 약 95%를 초과하는 약 800배 이상의 평균 정도로 서열분석하는 것을 포함한다. 다른 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 서열분석된 유전자(예컨대, 엑손)의 약 90%를 초과하는 약 1,000배 초과 평균 정도로 서열분석하는 것을 포함한다. 다른 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 서열분석된 유전자(예컨대, 엑손)의 약 90%를 초과하는 약 2,000배 이상의 평균 정도로 서열분석하는 것을 포함한다. 다른 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 서열분석된 유전자(예컨대, 엑손)의 약 90%를 초과하는 약 3,000배 이상의 평균 정도로 서열분석하는 것을 포함한다. 다른 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 서열분

석된 유전자(예컨대, 엑손)의 약 90%를 초과하는 약 3,500배 이상의 평균 정도로 서열분석하는 것을 포함한다. 다른 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 서열분석된 유전자(예컨대, 엑손)의 약 90%를 초과하는 약 4,000배 이상의 평균 정도로 서열분석하는 것을 포함한다. 다른 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 서열분석된 유전자(예컨대, 엑손)의 약 90%를 초과하는 약 4,500배 이상의 평균 정도로 서열분석하는 것을 포함한다. 다른 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 서열분석된 유전자(예컨대, 엑손)의 약 90%를 초과하는 약 5,000배 이상의 평균 정도로 서열분석하는 것을 포함한다. 다른 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 서열분석된 유전자(예컨대, 엑손)의 약 90%를 초과하는 약 5,500배 이상의 평균 정도로 서열분석하는 것을 포함한다. 다른 구현예들에서, 대상체 간격에 대한 리드를 수득하는 것은 서열분석된 유전자(예컨대, 엑손)의 약 90%를 초과하는 약 6,000배 이상의 평균 정도로 서열분석하는 것을 포함한다. 특정 구현예들에서, 상기 대상체 간격에 대한 리드를 수득하는 단계는 서열분석된 약 99% 초과 유전자(예컨대, 엑손)에서 약 100X 이상, 약 250X 이상, 약 500X 이상, 약 1,000X 이상, 약 1,500X 이상, 약 2,000X 이상, 약 2,500X로 이상, 약 3,000X 이상, 약 3,500X 이상, 약 4,000X 이상, 약 4,500X 이상, 약 5,000X 이상, 약 5,500X 이상, 또는 약 6,000X 이상의 평균 정도로 서열분석하는 단계를 포함한다.

- [0269] 특정 구현예들에서, 본원에 기재된 일련의 대상체 간격(예컨대, 대상체 간격을 부호화함)의 서열, 예컨대 뉴클레오타이드 서열이 본원에 기재된 방법에 의해 제공된다. 특정 구현예들에서, 상기 서열은 일치된 정상 대조군(예컨대, 야생형 대조군), 일치된 종양 대조군(예컨대, 원발성 대 전이성), 또는 둘 모두를 포함하는 방법을 사용하지 않고 제공된다.
- [0270] 유전자 선택
- [0271] 분석을 위한 대상체 간격, 예컨대 하위 유전체 간격, 발현된 하위 유전체 간격, 또는 둘 모두, 예컨대 유전자 및 기타 영역의 세트 또는 군에 대한 하위 유전체 간격의 군 또는 세트가 본원에 기재되어 있다.
- [0272] 일부 구현예들에서, 상기 방법은, 상기 수득된 핵산 샘플로부터 적어도 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500개 이상의 유전자 또는 유전자 산물로부터의 대상체 간격을, 예컨대 차세대 서열분석 방법에 의해 서열분석하는 단계를 포함하고, 상기 유전자들은 표 2A-5B로부터 선택된다.
- [0273] 일부 구현예들에서, 상기 방법은, 상기 샘플로부터 적어도 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500개 이상의 유전자 또는 유전자 산물격을, 예컨대 차세대 서열분석 방법에 의해 서열분석하는 단계를 포함하고, 상기 유전자들은 표 2A-5B로부터의 대상체 간부터 선택된다.
- [0274] 다른 구현예에서, 하기의 세트 또는 군 중 하나의 대상체 간격이 분석된다. 예컨대, 종양 또는 암 유전자 또는 유전자 산물 및 참조(예컨대, 야생형) 유전자 또는 유전자 산물과 관련된 대상체 간격은 상기 샘플로부터의 하위 유전체 간격의 군 또는 세트를 제공할 수 있다.
- [0275] 한 구현예에서, 상기 방법은 상기 샘플로부터 리드, 예컨대 서열, 대상체 간격의 세트를 수득하고, 상기 대상체 간격은 하기 중 적어도 1, 2, 3, 4, 5, 6, 7개 또는 모두로부터 선택된다:
- [0276] A) 표 2A-5B에 따른 돌연변이된 또는 야생형 유전자로부터의 적어도 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500개 이상의 대상체 간격, 예컨대 하위 유전체 간격, 또는 발현된 하위 유전체 간격, 또는 둘 모두;
- [0277] B) 종양 또는 암과 관련된 유전자 또는 유전자 산물로부터의 적어도 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500개 이상의 대상체 간격(예컨대, 종양 또는 암, 예컨대 표 2A-5B에 따른 유전자에 대한 양성 또는 음성 치료 반응 예측인자, 양성 또는 음성 예후 인자이거나 감별 진단을 가능하게 함);
- [0278] C) 표 2A-5B로부터 선택된 유전자 내에 존재하는 하위 유전체 간격의 돌연변이된 또는 야생형 유전자 또는 유전자 산물(예컨대, 단일 염기 다형성(SNP))로부터의 적어도 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500개 이상의 대상체 간격;
- [0279] D) 표 2A-5B로부터 선택된 유전자 내에 존재하는 하위 유전체 간격의 돌연변이된 또는 야생형 유전자(예컨대, 단일 염기 다형성(SNP))로부터의 적어도 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500개 이상의 대상체 간격으로서: (i) 약물로 치료된 암 환자의 더 나은 생존(예컨대, 파클리탁셀로 치료된 유방암 환자의 더 나은 생존); (ii) 파클리탁셀 대사; (iii) 약물에 대한 독성; 또는 (iv) 약물

대한 부작용 중 하나 이상과 관련된 대상체 간격;

- [0280] E) 표 2A-5B에 따른 적어도 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500개 이상의 유전자 또는 유전자 산물을 수분하는 복수의 전좌 변경;
- [0281] F) 표 2A-5B로부터 선택된 적어도 5개의 유전자로서, 예컨대 한 위치에서의 대립유전자 변이는 종양 유형과 연관되고, 상기 대립유전자 변이는 상기 종양 유형의 세포의 5% 미만에 존재하는 유전자;
- [0282] G) GG 풍부 영역에 수반된 표 2A-5B로부터 선택된 적어도 5개의 유전자; 또는
- [0283] H) 암 발병에 대한 유전(예컨대, 생식계열 위험) 인자를 나타내는 적어도 5개의 유전자(예컨대, 유전자 또는 유전자 산물은 표 2A-5B로부터 선택됨).
- [0284] 또 다른 구현예에서, 상기 방법은 상기 샘플로부터 대상체 간격 세트에 대한 리드, 예컨대 서열을 획득하고, 상기 대상체 간격은 표 2A-2C에 기재된 유전자들 중 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400개, 또는 모두로부터 선택된다.
- [0285] 또 다른 구현예에서, 상기 방법은 상기 샘플로부터 대상체 간격 세트에 대한 리드, 예컨대 서열을 획득하고, 상기 대상체 간격은 표 3A-3B에 기재된 유전자들 중 5, 6, 7, 8, 9, 10, 15, 20, 25, 30개 또는 모두로부터 선택된다.
- [0286] 또 다른 구현예에서, 상기 방법은 상기 샘플로부터 대상체 간격 세트에 대한 리드, 예컨대 서열을 획득하고, 상기 대상체 간격은 표 4A-4C에 기재된 유전자들 중 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300개, 또는 모두로부터 선택된다.
- [0287] 또 다른 구현예에서, 상기 방법은 상기 샘플로부터 대상체 간격 세트에 대한 리드, 예컨대 서열을 획득하고, 상기 대상체 간격은 표 5A-5B에 기재된 유전자들 중 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80개 또는 모두로부터 선택된다.
- [0288] 상기 선택된 유전자 또는 유전자 산물(본원에서 "표적 유전자 또는 유전자 산물"로도 지칭됨)은 유전자내 영역 또는 유전자간 영역을 포함하는 대상체 간격을 포함할 수 있다. 예를 들어, 상기 대상체 간격은 엑손 또는 인트론, 또는 이의 단편, 전형적으로 엑손 서열 또는 이의 단편을 포함할 수 있다. 상기 대상체 간격은 부호화 영역 또는 비부호화 영역, 예컨대 프로모터, 인핸서, 5' 비번역 영역(5' UTR), 또는 3' 비번역 영역(3' UTR), 또는 이의 단편을 포함할 수 있다. 다른 구현예들에서, 상기 대상체 간격은 cDNA 또는 이의 단편을 포함한다. 다른 구현예들에서, 상기 대상체 간격은, 예컨대 본원에 기재된 바와 같은 SNP를 포함한다.
- [0289] 다른 구현예들에서, 상기 대상체 간격은 유전체의 실질적으로 모든 엑손, 예컨대 본원에 기재된 바와 같은 대상체 간격 중 하나 이상(예컨대, 선택된 유전자 또는 관심 유전자 산물(예컨대, 본원에 기재된 암성 표현형과 관련된 유전자 또는 유전자 산물)을 포함한다. 일 구현예에서, 상기 대상체 간격은 체세포 돌연변이, 생식계열 돌연변이 또는 둘 모두를 포함한다. 일 구현예에서, 상기 대상체 간격은 변경, 예컨대 점 또는 단일 돌연변이, 결실 돌연변이(예컨대, 틀내 결실, 유전자내 결실, 전체 유전자 결실), 삽입 돌연변이(예컨대, 유전자내 삽입), 역위 돌연변이(예컨대, 염색체내 역위), 연결 돌연변이, 연결된 삽입 돌연변이, 역위 중복 돌연변이, 직렬 중복(예컨대, 염색체내 직렬 중복), 전좌(예컨대, 염색체 전좌, 비가역적 전좌), 재배열, 유전자 복제 수의 변화, 또는 이들의 조합을 포함한다. 특정 구현예들에서, 상기 대상체 간격은 샘플에서 종양 세포 유전체의 부호화 영역의 5%, 1%, 0.5%, 0.1%, 0.05%, 0.01%, 0.005%, 또는 0.001% 미만을 구성한다. 다른 구현예들에서, 상기 대상체 간격은 질병에 연루되지 않고, 예컨대 본원에 기재된 바와 같은 암성 표현형과 연관되지 않는다.
- [0290] 일 구현예에서, 상기 표적 유전자 또는 유전자 산물은 바이오마커이다. 본원에서 사용된 "바이오마커" 또는 "마커"는 변경될 수 있는 유전자, mRNA 또는 단백질이며, 상기 변경은 암과 관련이 있다. 상기 변경은 정상 또는 건강한 조직 또는 세포(예컨대, 대조군)에서의 양, 구조 및/또는 활성과 비교하여 암 조직 또는 암 세포의 양, 구조 및/또는 활성에서 이루어질 수 있으며, 암과 같은 질병 상태와 관련이 있다. 예를 들어, 암과 연관되거나 항암 치료제에 대한 반응성을 예측하는 마커는 변경된 뉴클레오타이드 서열, 아미노산 서열, 염색체 전좌, 염색체내 역위, 복제 수, 발현 수준, 단백질 수준, 단백질 활성, 유전 외적 변형(예컨대, 정상외의 건강한 조직 또는 세포와 비교하여 암 조직 또는 암세포에서의 메틸화 또는 아세틸화 상태, 또는 번역후 변형)을 가질 수 있다. 또한, "마커"는 구조가 변경된, 예컨대 돌연변이된(돌연변이를 함유함), 예컨대 암과 같은 질병 상태와 관련된 조직이나 세포 내에 존재할 때 치환, 결실 또는 삽입에 의해 뉴클레오타이드 또는 아미노산 수준에서 야생형 서열과 상이한 분자를 포함한다.

- [0291] 일 구현예에서, 상기 표적 유전자 또는 유전자 생성물은 단일 염기 다형성(SNP)을 포함한다. 다른 구현예에서, 상기 유전자 또는 유전자 생성물은 작은 결실, 예컨대 작은 유전자내 결실(예컨대, 틀내 또는 틀이동 결실)을 갖는다. 또 다른 구현예에서, 상기 표적 서열은 전체 유전자의 결실로부터 발생한다. 또 다른 구현예에서, 상기 표적 서열은 작은 삽입, 예컨대 작은 유전자내 삽입을 갖는다. 일 구현예에서, 상기 표적 서열은 역위, 예컨대 염색체내 역위로부터 발생한다. 다른 구현예에서, 상기 표적 서열은 염색체간 전좌로부터 발생한다. 또 다른 구현예에서, 상기 표적 서열은 순차 중복을 갖는다. 일 구현예에서, 상기 표적 서열은 바람직하지 않은 특징(예컨대, 높은 GC 함량 또는 반복 요소)을 갖는다. 다른 구현예에서, 상기 표적 서열은, 예컨대 이의 반복적인 성질 때문에 그 자체가 성공적으로 표적화될 수 없는 뉴클레오타이드 서열의 일부를 갖는다. 일 구현예에서, 상기 표적 서열은 대안적 스플라이싱으로부터 발생한다. 다른 구현예에서, 상기 표적 서열은 표 2A-5B에 따른 유전자 또는 유전자 산물, 또는 이의 단편으로부터 선택된다.
- [0292] 한 구현예에서, 상기 표적 유전자 또는 유전자 산물, 또는 이의 단편은 항체 유전자 또는 유전자 산물, 면역글로불린 상과 수용체(예컨대, B 세포 수용체(BCR) 또는 T 세포 수용체(TCR)) 유전자 또는 유전자 산물, 또는 이의 단편이다.
- [0293] 인간 항체 분자(및 B 세포 수용체)는 적어도 하나의 3개 유전자좌에 있는 유전자에 의해 부호화되는 불변(C) 및 가변(V) 영역을 모두 갖는 중쇄 및 경쇄로 구성된다:
- [0294] 1. 면역글로불린 중쇄에 대한 유전자 분절을 함유하는 14번 염색체 상의 면역글로불린 중쇄 유전자좌(IGH@);
- [0295] 2. 면역글로불린 경쇄에 대한 유전자 분절을 함유하는 2번 염색체 상의 면역글로불린 카파( $\kappa$ ) 유전자좌(IGK@);
- [0296] 3. 면역글로불린 경쇄에 대한 유전자 분절을 함유하는 22번 염색체 상의 면역글로불린 람다( $\lambda$ ) 유전자좌(IGL@).
- [0297] 각각의 중쇄 및 경쇄 유전자는 항체 단백질의 가변 영역에 대한 3가지 상이한 유형의 유전자 분절의 다중 복제를 포함한다. 예를 들어, 상기 면역글로불린 중쇄 영역은 5개의 상이한 부류  $\gamma$ ,  $\delta$ ,  $\alpha$ ,  $\mu$  및  $\epsilon$ , 44개의 가변(V) 유전자 분절, 27개의 다양성(D) 유전자 분절 및 6개의 결합(J) 유전자 분절 중 하나를 함유할 수 있다. 상기 경쇄는 또한 수많은 V 및 J 유전자 분절을 가질 수 있지만 D 유전자 분절은 없다. 상기 람다 경쇄에는 7개의 가능한 C 영역이 있고 상기 카파 경쇄에는 1개의 가능한 C 영역이 있다.
- [0298] 면역글로불린 중쇄 유전자좌(IGH@)는 인간 항체(또는 면역글로불린)의 중쇄에 대한 유전자를 함유하는 인간 염색체 14의 영역이다. 예를 들어, 상기 IGH 유전자좌는 IGHV(가변), IGHD(다양성), IGHI(결합) 및 IGHC(불변) 유전자들을 포함한다. 상기 면역글로불린 중쇄를 부호화하는 예시적인 유전자는 IGHV1-2, IGHV1-3, IGHV1-8, IGHV1-12, IGHV1-14, IGHV1-17, IGHV1-18, IGHV1-24, IGHV1-45, IGHV1-46, IGHV1-58, IGHV1-67, IGHV1-68, IGHV1-69, IGHV1-38-4, IGHV1-69-2, IGHV2-5, IGHV2-10, IGHV2-26, IGHV2-70, IGHV3-6, IGHV3-7, IGHV3-9, IGHV3-11, IGHV3-13, IGHV3-15, IGHV3-16, IGHV3-19, IGHV3-20, IGHV3-21, IGHV3-22, IGHV3-23, IGHV3-25, IGHV3-29, IGHV3-30, IGHV3-30-2, IGHV3-30-3, IGHV3-30-5, IGHV3-32, IGHV3-33, IGHV3-33-2, IGHV3-35, IGHV3-36, IGHV3-37, IGHV3-38, IGHV3-41, IGHV3-42, IGHV3-43, IGHV3-47, IGHV3-48, IGHV3-49, IGHV3-50, IGHV3-52, IGHV3-53, IGHV3-54, IGHV3-57, IGHV3-60, IGHV3-62, IGHV3-63, IGHV3-64, IGHV3-65, IGHV3-66, IGHV3-71, IGHV3-72, IGHV3-73, IGHV3-74, IGHV3-75, IGHV3-76, IGHV3-79, IGHV3-38-3, IGHV3-69-1, IGHV4-4, IGHV4-28, IGHV4-30-1, IGHV4-30-2, IGHV4-30-4, IGHV4-31, IGHV4-34, IGHV4-39, IGHV4-55, IGHV4-59, IGHV4-61, IGHV4-80, IGHV4-38-2, IGHV5-51, IGHV5-78, IGHV5-10-1, IGHV6-1, IGHV7-4-1, IGHV7-27, IGHV7-34-1, IGHV7-40, IGHV7-56, IGHV7-81, IGHVII-1-1, IGHVII-15-1, IGHVII-20-1, IGHVII-22-1, IGHVII-26-2, IGHVII-28-1, IGHVII-30-1, IGHVII-31-1, IGHVII-33-1, IGHVII-40-1, IGHVII-43-1, IGHVII-44-2, IGHVII-46-1, IGHVII-49-1, IGHVII-51-2, IGHVII-53-1, IGHVII-60-1, IGHVII-62-1, IGHVII-65-1, IGHVII-67-1, IGHVII-74-1, IGHVII-78-1, IGHVIII-2-1, IGHVIII-5-1, IGHVIII-5-2, IGHVIII-11-1, IGHVIII-13-1, IGHVIII-16-1, IGHVIII-22-2, IGHVIII-25-1, IGHVIII-26-1, IGHVIII-38-1, IGHVIII-44, IGHVIII-47-1, IGHVIII-51-1, IGHVIII-67-2, IGHVIII-67-3, IGHVIII-67-4, IGHVIII-76-1, IGHVIII-82, IGHVIV-44-1, IGHD1-1, IGHD1-7, IGHD1-14, IGHD1-20, IGHD1-26, IGHD2-2, IGHD2-8, IGHD2-15, IGHD2-21, IGHD3-3, IGHD3-9, IGHD3-10, IGHD3-16, IGHD3-22, IGHD4-4, IGHD4-11, IGHD4-17, IGHD4-23, IGHD5-5, IGHD5-12, IGHD5-18, IGHD5-24, IGHD6-6, IGHD6-13, IGHD6-19, IGHD6-25, IGHD7-27, IGHI1, IGHI1P, IGHI2, IGHI2P, IGHI3, IGHI3P, IGHI4, IGHI5, IGHI6, IGHA1, IGHA2, IGHG1, IGHG2, IGHG3, IGHG4, IGHGP, IGHD, IGHE, IGHEP1, IGHM, 및 IGHV1-69D를 포함하지만 이에 제한되지는 않는다.

- [0299] 면역글로불린 카파 유전자좌(IGK $\theta$ )는 항체(또는 면역글로불린)의 카파( $\kappa$ ) 경쇄에 대한 유전자를 함유하는 인간 염색체 2의 영역이다. 예를 들어, 상기 IGK 유전자좌는 IGKV(가변), IGKJ(결합) 및 IGKC(불변) 유전자를 포함한다. 상기 면역글로불린 카파 경쇄를 부호화하는 예시적인 유전자는 IGKV1-5, IGKV1-6, IGKV1-8, IGKV1-9, IGKV1-12, IGKV1-13, IGKV1-16, IGKV1-17, IGKV1-22, IGKV1-27, IGKV1-32, IGKV1-33, IGKV1-35, IGKV1-37, IGKV1-39, IGKV1D-8, IGKV1D-12, IGKV1D-13, IGKV1D-16, IGKV1D-17, IGKV1D-22, IGKV1D-27, IGKV1D-32, IGKV1D-33, IGKV1D-35, IGKV1D-37, IGKV1D-39, IGKV1D-42, IGKV1D-43, IGKV2-4, IGKV2-10, IGKV2-14, IGKV2-18, IGKV2-19, IGKV2-23, IGKV2-24, IGKV2-26, IGKV2-28, IGKV2-29, IGKV2-30, IGKV2-36, IGKV2-38, IGKV2-40, IGKV2D-10, IGKV2D-14, IGKV2D-18, IGKV2D-19, IGKV2D-23, IGKV2D-24, IGKV2D-26, IGKV2D-28, IGKV2D-29, IGKV2D-30, IGKV2D-36, IGKV2D-38, IGKV2D-40, IGKV3-7, IGKV3-11, IGKV3-15, IGKV3-20, IGKV3-25, IGKV3-31, IGKV3-34, IGKV3D-7, IGKV3D-11, IGKV3D-15, IGKV3D-20, IGKV3D-25, IGKV3D-31, IGKV3D-34, IGKV4-1, IGKV5-2, IGKV6-21, IGKV6D-21, IGKV6D-41, IGKV7-3, IGKJ1, IGKJ2, IGKJ3, IGKJ4, IGKJ5, 및 IGKC를 포함하지만 이에 제한되지는 않는다. 면역글로불린 람다 유전자좌(IGL $\theta$ )는 항체(또는 면역글로불린)의 람다 경쇄에 대한 유전자를 함유하는 인간 염색체 22의 영역이다. 예를 들어, 상기 IGL 유전자좌는 IGLV(가변), IGLJ(결합) 및 IGLC(불변) 유전자들을 포함한다. 상기 면역글로불린 람다 경쇄를 부호화하는 예시적인 유전자는 IGLV1-36, IGLV1-40, IGLV1-41, IGLV1-44, IGLV1-47, IGLV1-50, IGLV1-51, IGLV1-62, IGLV2-5, IGLV2-8, IGLV2-11, IGLV2-14, IGLV2-18, IGLV2-23, IGLV2-28, IGLV2-33, IGLV2-34, IGLV3-1, IGLV3-2, IGLV3-4, IGLV3-6, IGLV3-7, IGLV3-9, IGLV3-10, IGLV3-12, IGLV3-13, IGLV3-15, IGLV3-16, IGLV3-17, IGLV3-19, IGLV3-21, IGLV3-22, IGLV3-24, IGLV3-25, IGLV3-26, IGLV3-27, IGLV3-29, IGLV3-30, IGLV3-31, IGLV3-32, IGLV4-3, IGLV4-60, IGLV4-69, IGLV5-37, IGLV5-39, IGLV5-45, IGLV5-48, IGLV5-52, IGLV6-57, IGLV7-35, IGLV7-43, IGLV7-46, IGLV8-61, IGLV9-49, IGLV10-54, IGLV10-67, IGLV11-55, IGLV1-20, IGLV1-38, IGLV1-42, IGLV1-56, IGLV1-63, IGLV1-68, IGLV1-70, IGLVIV-53, IGLVIV-59, IGLVIV-64, IGLVIV-65, IGLVIV-66-1, IGLVV-58, IGLVV-66, IGLVVI-22-1, IGLVVI-25-1, IGLVVII-41-1, IGLJ1, IGLJ2, IGLJ3, IGLJ4, IGLJ5, IGLJ6, IGLJ7, IGLC1, IGLC2, IGLC3, IGLC4, IGLC5, IGLC6, 및 IGLC7을 포함하지만 이에 제한되지는 않는다.
- [0300] 상기 B 세포 수용체(BCR)는 두 부분: i) 하나의 동형(예컨대, IgD 또는 IgM)의 막 결합성 면역글로불린 분자로 구성된다. 통합 막 도메인의 존재를 제외하고, 이는 분비된 형태 및 ii) 신호 전달 모이어티: 이항화 가교에 의해 함께 결합된 Ig- $\alpha$ /Ig- $\beta$ (CD79)라고 하는 이종이량체와 동일할 수 있다. 상기 이량체의 각 핵산 분자는 원형 질막에 걸쳐 있으며 면역수용체 티로신 기반 활성화 모티프(ITAM)를 수반하는 세포질 꼬리를 가지고 있다.
- [0301] T 세포 수용체(TCR)는 2개의 상이한 단백질 사슬(즉, 이종이량체)로 구성된다. T 세포의 95%에서, 이는 알파( $\alpha$ ) 및 베타( $\beta$ ) 사슬로 구성되는 반면, T 세포의 5%에서는 감마( $\gamma$ ) 및 델타( $\delta$ ) 사슬로 구성된다. 상기 비율은 개체 발생 동안 및 질병 상태에서 가변될 수 있다. 상기 T 세포 수용체 유전자는 림프구의 발달 중에 재배열되어 각 세포에 고유한 항원 수용체를 제공하는 베타 및 델타 사슬(및 알파 및 감마 사슬의 V 및 J 유전자 분절)에 다수의 V, D 및 J 유전자 분절을 역시 함유한다는 점에서 면역글로불린 유전자와 유사하다.
- [0302] T 세포 수용체 알파 유전자좌(TRA)는 상기 TCR 알파 사슬에 대한 유전자를 함유하는 인간 염색체 14 상의 영역이다. 예를 들어, 상기 TRA 유전자좌는, 예컨대 TRAV(가변), TRAJ(결합) 및 TRAC(불변) 유전자들을 포함한다. T 세포 수용체 알파 사슬을 부호화하는 예시적인 유전자는 TRAV1-1, TRAV1-2, TRAV2, TRAV3, TRAV4, TRAV5, TRAV6, TRAV7, TRAV8-1, TRAV8-2, TRAV8-3, TRAV8-4, TRAV8-5, TRAV8-6, TRAV8-7, TRAV9-1, TRAV9-2, TRAV10, TRAV11, TRAV12-1, TRAV12-2, TRAV12-3, TRAV13-1, TRAV13-2, TRAV14DV4, TRAV15, TRAV16, TRAV17, TRAV18, TRAV19, TRAV20, TRAV21, TRAV22, TRAV23DV6, TRAV24, TRAV25, TRAV26-1, TRAV26-2, TRAV27, TRAV28, TRAV29DV5, TRAV30, TRAV31, TRAV32, TRAV33, TRAV34, TRAV35, TRAV36DV7, TRAV37, TRAV38-1, TRAV38-2DV8, TRAV39, TRAV40, TRAV41, TRAJ1, TRAJ2, TRAJ3, TRAJ4, TRAJ5, TRAJ6, TRAJ7, TRAJ8, TRAJ9, TRAJ10, TRAJ11, TRAJ12, TRAJ13, TRAJ14, TRAJ15, TRAJ16, TRAJ17, TRAJ18, TRAJ19, TRAJ20, TRAJ21, TRAJ22, TRAJ23, TRAJ24, TRAJ25, TRAJ26, TRAJ27, TRAJ28, TRAJ29, TRAJ30, TRAJ31, TRAJ32, TRAJ33, TRAJ34, TRAJ35, TRAJ36, TRAJ37, TRAJ38, TRAJ39, TRAJ40, TRAJ41, TRAJ42, TRAJ43, TRAJ44, TRAJ45, TRAJ46, TRAJ47, TRAJ48, TRAJ49, TRAJ50, TRAJ51, TRAJ52, TRAJ53, TRAJ54, TRAJ55, TRAJ56, TRAJ57, TRAJ58, TRAJ59, TRAJ60, TRAJ61, 및 TRAC를 포함하지만 이에 제한되지는 않는다.
- [0303] T 세포 수용체 베타 유전자좌(TRB)는 상기 TCR 베타 사슬에 대한 유전자를 함유하는 인간 염색체 7 상의 영역이다. 예를 들어, 상기 TRB 유전자좌는, 예컨대 TRBV(가변), TRBD(다양성), TRBJ(결합) 및 TRBC(불변) 유전자들을 포함한다. T 세포 수용체 베타 사슬을 부호화하는 예시적인 유전자는 TRBV1, TRBV2, TRBV3-1, TRBV3-2, TRBV4-1, TRBV4-2, TRBV4-3, TRBV5-1, TRBV5-2, TRBV5-3, TRBV5-4, TRBV5-5, TRBV5-6, TRBV5-7, TRBV6-2, TRBV6-3,

TRBV6-4, TRBV6-5, TRBV6-6, TRBV6-7, TRBV6-8, TRBV6-9, TRBV7-1, TRBV7-2, TRBV7-3, TRBV7-4, TRBV7-5, TRBV7-6, TRBV7-7, TRBV7-8, TRBV7-9, TRBV8-1, TRBV8-2, TRBV9, TRBV10-1, TRBV10-2, TRBV10-3, TRBV11-1, TRBV11-2, TRBV11-3, TRBV12-1, TRBV12-2, TRBV12-3, TRBV12-4, TRBV12-5, TRBV13, TRBV14, TRBV15, TRBV16, TRBV17, TRBV18, TRBV19, TRBV20-1, TRBV21-1, TRBV22-1, TRBV23-1, TRBV24-1, TRBV25-1, TRBV26, TRBV27, TRBV28, TRBV29-1, TRBV30, TRBVA, TRVB, TRBV5-8, TRBV6-1, TRBD1, TRBD2, TRBJ1-1, TRBJ1-2, TRBJ1-3, TRBJ1-4, TRBJ1-5, TRBJ1-6, TRBJ2-1, TRBJ2-2, TRBJ2-2P, TRBJ2-3, TRBJ2-4, TRBJ2-5, TRBJ2-6, TRBJ2-7, TRBC1, 및 TRBC2를 포함하지만 이에 제한되지는 않는다.

[0304] T 세포 수용체 델타 유전자좌(TRD)는 상기 TCR 델타 사슬에 대한 유전자를 함유하는 인간 염색체 14 상의 영역이다. 예를 들어, 상기 TRD 유전자좌는, 예컨대 TRDV(가변), TRDJ(결합) 및 TRDJ(불변) 유전자들을 포함한다. T 세포 수용체 델타 사슬을 부호화하는 예시적인 유전자는 TRDV1, TRDV2, TRDV3, TRDD1, TRDD2, TRDD3, TRDJ1, TRDJ2, TRDJ3, TRDJ4, 및 TRDC를 포함하지만 이에 제한되지는 않는다.

[0305] T 세포 수용체 감마 유전자좌(TRG)는 상기 TCR 감마 사슬에 대한 유전자를 함유하는 인간 염색체 7 상의 영역이다. 예를 들어, 상기 TRG 유전자좌는, 예컨대 TRGV(가변), TRGJ(결합) 및 TRGC(불변) 유전자들을 포함한다. T 세포 수용체 감마 사슬을 부호화하는 예시적인 유전자는 TRGV1, TRGV2, TRGV3, TRGV4, TRGV5, TRGV5P, TRGV6, TRGV7, TRGV8, TRGV9, TRGV10, TRGV11, TRGVA, TRGVB, TRGJ1, TRGJ2, TRGJP, TRGJP1, TRGJP2, TRGC1, 및 TRGC2를 포함하지만 이에 제한되지는 않는다.

[0306] 일 구현예에서, 상기 표적 유전자 또는 유전자 산물, 또는 이의 단편은 표 2A-5B에 기재된 유전자 또는 유전자 산물 중 임의의 것으로부터 선택된다.

【표 2A】 예시적인 DNA-seq 표적 포획 시약에서 완전한 엑손 적용 범위를 갖는 예

시적인 유전자

|                        |                 |        |               |           |             |   |     |     |   |         |
|------------------------|-----------------|--------|---------------|-----------|-------------|---|-----|-----|---|---------|
| ABL1                   | BTX             | CTNMB1 | FAS (TNFRSF6) | HIST1H1C  | DR          | K | MYC | PDK |   | SUFU    |
| ACTB                   | BTX             |        |               | HIST1H1D  | EAP1        | K | MYD | PHF |   | SUZ12   |
| AKT1                   | c11orf30 (EMSY) |        | FBXO1         | HIST1H1E  | IT          | K | MYO |     |   |         |
| AKT2                   | CAD             | UX1    | FBXO3         | HIST1H2AC | LHL6        | K |     |     |   | TAF1    |
| AKT3                   | CARD11          | CXCR4  | FBXW7         | HIST1H2AG | MT2A (MLL)  | K |     | PIK |   | TBL1XR1 |
| ALK                    | P8              | CAS    | FGF10         | HIST1H2AL |             |   | NCO | PIK |   |         |
|                        | B               | CBF    |               | HIST1H2AM | MT2C (MLL3) | K | NCS | PIK | R | TCF3    |
| AMER1 (FAM123B or WTX) | CBL             | DR2    | FGF14         | HIST1H2BC | RAS         | K | NF1 | PIK | R | TCL1A   |
| APC                    | D1              | CCN    | DDX3X         | FGF19     | EF1         | L | NF2 | PIM |   | TET2    |
|                        | D2              | CCN    | FGF23         | HIST1H2BK | MO1         | L | NFE | PLC | S | TGFBR2  |
| APH1A                  | D3              | CCN    | FGF3          | HIST1H2BO | RP1B        | L | NFK | PMS |   |         |
| AR                     | E1              | CCN    | FGF4          | HIST1H3B  | RRK2        | L | NEX | PWR | S | TLL2    |
| ARAF                   | GB              | CCT    | DNMT3A        | FGF6      | AF          | M | NOD | POT | S | TNEM30  |
|                        |                 |        |               |           |             |   |     | DHB |   | A       |

[0307]

|                 |        |        |                 |           |         |     |          |       |                       |
|-----------------|--------|--------|-----------------|-----------|---------|-----|----------|-------|-----------------------|
| ARFRP1          | CD2    | OT1L   |                 | HNFI      | M       | NOT | PPP      | S     | TMSB4XP8 (TMSL3)      |
| ARRGA P26 (GRA) | CD2    | TX1    | FGFR1           | HRAS      | AFB     | CH1 | PRD      | S     | TNFAIP3               |
| ARTD1A          | 6      | CTG    | FGFR2           | HSP90AA1  | AGED1   | CH2 | M        | DMD   | TNFRSF11A             |
| ARID2           | CD5    | USP9   | FGFR3           | ICK       | ALT1    |     | PRK      | S     | TNFRSF14              |
| ASMTL           | 0      | BF1    | FGFR4           | ID3       | AP2K1   | 1   | DC       | ETBP1 | TNFRSF17              |
| ASXL1           | CD7    | CT2L   | FH1T            | IDH1      | AP2K4   | S   | PTC      | S     | TOP1                  |
| ATM             | CD7    | ED     | FLCN            | IDH2      | AP3K1   |     | PTF      | S     | TP53                  |
| ATR             | 73     | GFR    | FLT1            |           | AP3K13  | C2  | NT5      | PTP   | S                     |
| ATRX            | 1      | CDH    | FLT3            | R         | MAP3K14 | K1  | NTR      | PTP   | S                     |
| AURKA           | 12     | CDK    | FLT4            |           | AP3K5   | K2  | NTR      | PTP   | S                     |
| AURKB           | 4      | CDK    | FLTWCH1         | E         | AP3K7   | K3  | NTR      | PTP   | S                     |
| AXIN1           | 6      | CDK    | FOXO1           | 1         | APK1    | 93  | NUP      | RO    | SMARCA4               |
| AXL             | 8      | CDK    | FOXO1           | 2         | CL1     | 98  | NUP      | RO    | SMARCB1               |
| B2M             | CDKN1B | PHB1   | FOXO3           | 3         | DM2     | Y8  | P2R      | RAD   | TSC2                  |
| BAP1            | CDKN2A | RBB2   | FOXO1           | 1L7R      | DM4     | 1   | PAG      |       | TSHR                  |
| BARD1           | CDKN2B | RBB3   | FRS2            | A         | ED12    | 3   | PAK      |       | TUSC3                 |
| BCL10           | CDKN2C | RBB4   | GADD45B         | INPP4B    | EF2B    |     |          | MO    | TYK2                  |
| BCL11           | CEB    | RG     | GATA1           | 5D (SHIP) | EF2C    | B2  | PAL      |       | U2AF1                 |
| BCL2            | 2      | CHD    | GATA2           | IRF1      | EN1     |     |          |       | U2AF2                 |
| BCL2L2          | K1     | CHE    | GATA3           | IRF4      | ET      | 1   | RAF      |       | VHL                   |
| BCL6            | K2     | CHE    | GID4 (c17orf39) | IRF8      | IB1     |     | RAR      |       | VDR90                 |
| BCL7A           |        | EXOSC6 | GNA11           | IRS2      | ITF     |     | RASGEF1A |       | VHSC1 (NMSET or NSD2) |
| BCOR            | CIC    | ZH2    | GNA12           | JAK1      | K167    | K   | PAS      |       | VISP3                 |
| BCORL1          | TA     | AP1    | GNA13           | JAK2      | LH1     | 5   | PAX      |       | WT1                   |
| BIRC3           | 1B     | CKS    | FANAF6C         | GNAQ      | PL      | M1  | PBR      |       | XPB1                  |

[0308]

|               |     |      |        |        |       |       |             |   |              |
|---------------|-----|------|--------|--------|-------|-------|-------------|---|--------------|
| BLM           | CPS | ANCA | GNAS   | JARID2 | RE11A | PC    | RET         | S | XP01         |
| BRAF          | N   | CRB  | ANCC   | 4      | GPR12 | JUN   | SH2         | M | PCB          |
| BRC1A         | BBP | CRE  | FANCD2 | A      | GRIN2 | KAT6  | SH3         | M | PCL          |
| BRC2A         | L   | CRK  | ANCE   |        | GSK3B | KIDM2 | SH5         | M | D1           |
| BRD4          | F2  | CRL  | ANCF   |        | GTSE1 | KIDM4 | TOR         | M | D11          |
| BRIP1 (BACH1) | 1R  | CSF  | ANCG   |        | HDAC1 | KIDM5 | UTYH        | M | D1LG2 (PDL2) |
| BRSK1         | 3R  | CSF  |        |        | HDAC4 | KIDM5 | YC          | M | PDG          |
|               | F   | CTC  | ANCL   |        | HDAC7 | KIDM6 | YCL (MYCL1) | M | FRB          |
| BTG2          | NA1 | CTN  |        |        | HGF   |       | MT2D (MLL2) | K |              |

[0309]

【표 2B】선택된 DNA 재배열

|             |     |     |     |     |     |     |     |     |    |     |   |     |      |     |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|----|-----|---|-----|------|-----|
| ALK         | 2   | BCL | 6   | BCR | F   | BRA | CCN | CRL | B  | EGF | R | EPO | 1    | ETV |
| ETV4        | 5   | ETV | 6   | EWS | R1  | FGF | D1  | IGH | F2 | IGK | 1 | JAX | 2    | JAX |
| KMT2A (MLL) | MYC | K1  | NTR | FRA | PDG | PRB | 1   | RAF | A  | RAR | 1 | ROS | BSS2 | TMP |
| TRG         |     |     |     |     |     |     |     |     |    |     |   |     |      |     |

[0310]

【표 2C】 RNA 유전자 융합 선택

|              |                 |                   |            |                   |             |                    |                        |              |             |
|--------------|-----------------|-------------------|------------|-------------------|-------------|--------------------|------------------------|--------------|-------------|
| ABI<br>1     | ABL<br>1        | ABL<br>2          | ACS<br>L6  | APF<br>1          | APF<br>4    | ALK                | ARH<br>GAP26<br>(GPAF) | ARH<br>GEF12 | ARI<br>D1A  |
| ABN<br>T     | ASX<br>L1       | ATF<br>1          | ATG<br>5   | ATI<br>C          | BCL<br>10   | BCL<br>11A         | BCL<br>11B             | BCL<br>2     | BCL<br>3    |
| BCL<br>6     | BCL<br>7A       | BCL<br>9          | BCO<br>R   | BCR<br>C3         | BIR<br>C3   | BRA<br>F           | BTG<br>1               | CAM<br>TA1   | CAR<br>S    |
| CBF<br>A2T3  | CBF<br>B        | CBL<br>D1         | CCN<br>D2  | CCN<br>D3         | CCN<br>D3   | CD2<br>74 (PD-L1)  | CDK<br>6               | CDX<br>2     | CHI<br>C2   |
| CHN<br>1     | CIC<br>C1C      | CII<br>TA         | CLP<br>1   | CLT<br>C          | CLT<br>CL1  | CMT<br>RL (CEP110) | COL<br>1A1             | CRE<br>B3L1  | CRE<br>B3L2 |
| CRE<br>BBP   | CRL<br>F2       | CSF<br>1          | CTN<br>NB1 | DDI<br>T3         | DDX<br>10   | DDX<br>6           | DEX<br>P22             | DUS<br>P22   | EGF<br>R    |
| EIF<br>4A2   | ELF<br>4        | ELL<br>ELL        | ELN<br>4   | EP3<br>00         | EP3<br>R    | EPO<br>R           | EPS<br>15              | ERB<br>E2    | ERG<br>R    |
| ETS<br>1     | ETV<br>1        | ETV<br>4          | ETV<br>5   | ETV<br>6          | EWS<br>R1   | FCG<br>R2B         | FCR<br>L4              | FEV<br>R1    | FGF<br>R1   |
| FGF<br>R1OP  | FGF<br>R2       | FGF<br>R3         | FLI<br>1   | FNB<br>P1         | FOX<br>O1   | FOX<br>O3          | FOX<br>O4              | FOX<br>P1    | FST<br>L3   |
| FUS<br>A1    | GAS<br>7        | GLI<br>1          | GMP<br>S   | GPH<br>N          | HER<br>PUD1 | HEY<br>1           | HIP<br>1               | HIS<br>T1H41 | HLF<br>HOX  |
| HMG<br>A1    | HMG<br>A2       | HOX<br>A11        | HOX<br>A13 | HOX<br>A3         | HOX<br>A9   | HOX<br>C11         | HOX<br>C13             | HOX<br>D11   | HOX<br>D13  |
| HSP<br>90AA1 | HSP<br>90AB1    | IGH<br>IGH        | IGK<br>IGK | IGL<br>IGL        | IK2<br>F1   | IL2<br>1R          | IL3<br>4               | IRF<br>4     | ITX<br>LCP  |
| JAK<br>1     | JAK<br>2        | JAK<br>3          | JAZ<br>F1  | KAT<br>6A (MYST3) | KDS<br>R    | XIF<br>5B          | KMT<br>2A(MLL)         | LAS<br>P1    | MKL<br>1    |
| LMO<br>1     | LMO<br>2        | LPP<br>1          | LYL<br>1   | MAF<br>B          | MAF<br>B    | MAL<br>T1          | MDS<br>2               | MEC<br>OM    | MSI<br>1    |
| MLF<br>1     | MLL<br>T1 (ENL) | MLL<br>T10 (AF10) | MLL<br>T3  | MLL<br>T4 (AF6)   | MLL<br>T6   | MN1<br>1           | MNX<br>2               | MSI<br>2     | MSN<br>1    |
| MUC<br>1     | MYB<br>MYB      | MYC<br>MYC        | MYH<br>11  | MYH<br>9          | NAC<br>A    | NBE<br>API (BCL8)  | NCO<br>A2              | NDR<br>G1    | NFI<br>NFI  |
| NF2<br>NF2   | NFK<br>E2       | NIN<br>NIN        | NOT<br>CHI | NPM<br>1          | NR4<br>A3   | NSD<br>1           | NTR<br>K1              | NTR<br>K2    | NTR<br>K3   |
| NUM<br>A1    | NUP<br>214      | NUP<br>98         | NUT<br>M2A | OMD<br>OMD        | P2R<br>Y8   | PAF<br>AH1B2       | PAX<br>3               | PAX<br>5     | PAX<br>7    |

|            |             |            |                         |                       |                   |             |               |                  |            |
|------------|-------------|------------|-------------------------|-----------------------|-------------------|-------------|---------------|------------------|------------|
| PEX<br>1   | PCM<br>1    | PCS<br>K7  | PDC<br>D1LG2<br>(PD-L2) | PDE<br>4DIP           | PDG<br>FB         | PDG<br>FRA  | PDG<br>FRB    | PER<br>1         | PHF<br>1   |
| PIC<br>ALM | PIM<br>1    | PLA<br>G1  | PML<br>PML              | POU<br>2AF1           | PPP<br>1CB        | PRD<br>M1   | PRD<br>M16    | PRR<br>X1        | PSI<br>P1  |
| PTC<br>H1  | PTK<br>7    | RAB<br>EP1 | RAF<br>1                | RAL<br>GDS            | RAP<br>1GDS1      | RAR<br>A    | REM<br>15     | RET<br>H         | RHO<br>H   |
| RNF<br>213 | ROS<br>1    | RPL<br>22  | RPN<br>1                | RUN<br>X1             | RUN<br>X1T1 (ETO) | RUN<br>X2   | SEC<br>31A    | SEP<br>T5        | SEP<br>T6  |
| SEP<br>T9  | SET<br>GL1  | SH3<br>GL1 | SLC<br>1A2              | SNX<br>29<br>(BUND2A) | SRS<br>F3         | SS1<br>8    | SSX<br>1      | SSX<br>2         | SSX<br>4   |
| STA<br>T6  | STL<br>STL  | SYK<br>SYK | TAF<br>15               | TAL<br>1              | TAL<br>2          | TBL<br>1XR1 | TCF<br>3(E2A) | TCL<br>1A (TCL1) | TEC<br>TEC |
| TET<br>1   | TFE<br>3    | TFG<br>TFG | TFP<br>T                | TFR<br>C              | TLX<br>1          | TLX<br>3    | TMP<br>RSS2   | TNF<br>RSF11A    | TOP<br>1   |
| TP6<br>3   | TPM<br>3    | TPM<br>4   | TRI<br>M24              | TRI<br>P11            | TTL<br>TTL        | TYK<br>2    | USP<br>6      | WHS<br>C1 (MMS1) | WHS<br>WHS |
| NSD<br>2   | WHS<br>C1L1 | YPE<br>L5  | ZBT<br>B16              | ZMY<br>M2             | ZNF<br>384        | ZNF<br>52   | ZNF<br>ZNF    | ZNF<br>ZNF       | ZNF<br>ZNF |

【표 3A】 예시적인 DNA-seq 표적 포획 시약에서 선택된 인트론이 적용되는 예시적

인 유전자

|      |    |                    |                |    |                |    |           |     |      |    |        |    |                     |     |   |                    |
|------|----|--------------------|----------------|----|----------------|----|-----------|-----|------|----|--------|----|---------------------|-----|---|--------------------|
| L1   | AB | ABL2               | VR1B           | AC | T1             | AK | 2         | AKT | T3   | AK | K      | AL | AME<br>R1 (FAM123B) | PC  | A | AR                 |
| AF   | AR | ARFR               | ID1A           | AR | ID1B           | AR | D2        | ARI | XL1  | AS | M      | AT | ATR                 | TRX | A | AURKA              |
| RKB  | AU | AXIN               | L              | AX | P1             | BA | D1        | BAR | L2   | BC | L2L1   | BC | 2L2                 | BCL | B | BCOR               |
| ORL1 | BC | BLM                | AF             | BR | CA1            | BR | A2        | BBC | D4   | BR | IP1    | BR | 1                   | BTG | B | C11orf3<br>O(EMSY) |
| RD11 | CA | CBFB               | L              | CB | ND1            | CC | D2        | CCN | ND3  | CC | NE1    | CC | 74 (PD-L1)          | CD2 | C | CD79B              |
| C73  | CD | CDH1               | K12            | CD | K4             | CD | 6         | CDK | K8   | CD | KN1A   | CD | N1B                 | CDK | C | CDKN2B             |
| KN2C | CD | CEBP               | D2             | CH | D4             | CH | K1        | CHE | EK2  | CH | C      | CI | N1B                 | CRE | C | CRLF2              |
| F1R  | CS | CTCF               | NNA1           | CT | NNB1           | CT | 3         | CUL | LD   | CY | XX     | DA | 2                   | DDR | D | DNMT3A             |
| T1L  | DO | EGFR               | 300            | EP | HA3            | EP | A5        | EPH | HA7  | EP | HB1    | EP | B2                  | ERB | E | ERBB4              |
| G    | ER | ERRF               | R1             | ES | H2             | EZ | 46C       | FAM | NCA  | FA | NCC    | FA | CD2                 | FAN | F | FANCF              |
| NGG  | FA | FANC               | S              | FA | T1             | FA | W7        | FBX | F10  | FG | F14    | FG | 19                  | FGF | F | FGF3               |
| F4   | FG | FGF6               | FR1            | FG | FR2            | FG | R3        | FGF | FR4  | FG | FH     | FG | N                   | FLC | F | FLT3               |
| T4   | FL | FOXL               | XP1            | FO | S2             | FR | P1        | FUB | BRA6 | GA | TA1    | GA | A2                  | GAT | G | GATA4              |
| TA6  | GA | GID4<br>(C17orf39) | 11             | GL | A11            | GN | 13        | GNA | AQ   | GN | AS     | GN | 124                 | GPR | G | GRM3               |
| K3B  | GS | H3F3               | F              | HG | F1A            | HN | S         | HRA | D3B1 | HS | P90AA1 | HS | 1                   | IDH | I | IGF1R              |
| F2   | IG | IKBK               | ZF1            | IK | 7R             | IL | BA        | INH | PP4B | IN | F2     | IR | 4                   | IRF | I | JAK1               |
| K2   | JA | JAK3               | N              | JU | T6A<br>(MYST3) | KA | 5A        | KDM | M5C  | KD | M6A    | KD |                     | KDR | K | KEL                |
| T    | KI | KLHL               | 6              | KM | T2A (MLL)      | KM | 2D (MLL2) | KMT | AS   | KR | O1     | LM | 1B                  | LRP | L | LZTR1              |
| GI2  | MA | MAP2<br>K1 (MEK1)  | P2K2<br>(MEK2) | MA | P2K4           | MA | 3K1       | MAP | L1   | MC | M2     | MD | 4                   | MDM | M | MEF2B              |
| N1   | ME | MET                | MI             | ML |                | ML |           | MPL | E11A | MR | H2     | MS | 6                   | MSH | M | MUTYH              |
| C    | MY | MYCL<br>(MYCL1)    | TF             | MY | H1             | MY |           | NF1 | 2    | NF | E2L2   | NF | B1A                 | NFK | N | NOTCH1             |

|        |    |      |      |    |       |    |               |     |             |    |      |    |     |     |   |         |
|--------|----|------|------|----|-------|----|---------------|-----|-------------|----|------|----|-----|-----|---|---------|
| TCH2   | NO | NOTC | M1   | NP | AS    | NR | 1             | NSD | RK1         | NT | RE2  | NT | K3  | NTR | N | PAK3    |
| LB2    | PA | PARK | X5   | PA | RM1   | PB | D1LG2 (PD-L2) | PDC | GFRA        | PD | GFRA | PD | 1   | PDK | P | PIK3CA  |
| K3CB   | PI | PIK3 | K3R1 | PI | K3R2  | PI | G2            | PLC | S2          | PM | LD1  | PO | E   | POL | P | PRIM1   |
| EX2    | PR | PRKA | KC1  | PR | KDC   | PR | S8            | PRS | CH1         | PT | EN   | PT | N11 | PTP | Q | RAC1    |
| D50    | RA | RAD5 | F1   | RA | NBP2  | RA | A             | RAR | 1           | RB | M10  | RB |     | RET | R | RNF43   |
| S1     | RO | RPTO | NX1  | RU | NX1T1 | RU | A             | SDH | HB          | SD | HC   | SD | D   | SDH | S | SF3B1   |
| IT2    | SL | SMAD | AD3  | SM | AD4   | SM | BCA4          | SMA | ARCB1       | SM | O    | SM |     | SNC | S | SOX10   |
| X2     | SO | SOX9 | EN   | SP | OP    | SP | A1            | SPT | C           | SR | AG2  | ST | T3  | STA | S | STK11   |
| FU     | SU | SYK  | F1   | TA | X3    | TB | C             | TER | RT(프로모터 단독) | TE |      |    | 2   | TET | T | TNFAIP3 |
| FRSF14 | TN | TOP1 | P2A  | TO | 53    | TP | 1             | TSC | C2          | TS | HR   | TS | F1  | U2A | V | VHL     |
| SP3    | WI | WT1  | O1   | XP | TB2   | ZB | 217           | ZNF | F703        | ZN |      |    |     |     |   |         |

【표 3B】 선택된 재배열

|     |     |     |     |     |     |    |     |     |     |     |     |     |      |     |     |     |
|-----|-----|-----|-----|-----|-----|----|-----|-----|-----|-----|-----|-----|------|-----|-----|-----|
| ALK | 2   | BCL | BCR | F   | BRA | A1 | A2  | BRC | 4   | BRD | R   | EGF | 1    | ETV | 4   | ETV |
| 5   | ETV | 6   | ETV | R1  | FGF | R2 | FGF | R3  | KIT | 2   | MSH | MYB | MYC  | CH2 | NOT |     |
| K1  | NTR | K2  | NTR | FRA | PDG | 1  | RAF | A   | RAR | RET | 1   | ROS | BSS2 | TMP |     |     |

【표 4A】 예시적인 RNA-seq 표적 포획 시약에서 표적화된 예시적인 유전자

|               |       |        |               |
|---------------|-------|--------|---------------|
| BRCA1         | CRKL  | MDM2   | SMO           |
| BRCA2         | EGFR  | MET    | TP53          |
| CCND1         | ERBB2 | MYC    | VEGFA         |
| CD274 (PD-L1) | ERRF1 | MYCN   |               |
| CDH1          | FGFR1 | NF1    |               |
| CDK4          | FGFR2 | L2     | PDCD1LG2 (PD- |
| CDK6          | FOXL2 | PTEN   |               |
| CDKN2A        | KRAS  | PTPN11 |               |

【표 4B】선택된 엑손

|      |    |       |       |    |    |    |    |    |    |      |    |    |    |      |    |      |    |
|------|----|-------|-------|----|----|----|----|----|----|------|----|----|----|------|----|------|----|
| L1   | AB | AKT1  | ALK   | AF | AR | AF | BR | K  | BT | NNB1 | CT | R2 | DD | R1   | ES | H2   | E2 |
| FR3  | FG | FLT3  | GNA11 | AQ | GN | AS | GN | HR | AS | H1   | ID | H2 | ID | K2   | JA | K3   | JA |
| T    | KI | MAP2K | MAP2K | L  | MP | OR | MT | MY | NP | AS   | NP | NR | NR | GFRA | PD | GFRB | PD |
| K3CA | PI | RAF1  | RET   | RT | TE |    |    |    |    |      |    |    |    |      |    |      |    |

【표 4C】선택된 재배열

|    |    |        |   |
|----|----|--------|---|
| K  | AL | FGFR3  | R |
| FR | EG | PDGFRA | R |
|    |    | OS1    |   |

【표 5A】예시적인 DNA-seq 표적 포획 시약에서 완전한 엑손 적용 범위를 갖는 후

가적인 유전자

|       |   |       |   |         |            |    |     |     |      |    |       |    |                     |            |      |
|-------|---|-------|---|---------|------------|----|-----|-----|------|----|-------|----|---------------------|------------|------|
| BL1   | A | CVR1B | A | AKT1    | T2         | AK | 3   | AKT | K    | AL | OX12B | AL | 1 (FAM123B)         | APC        | AR   |
| RAF   | A | RFRP1 | A | ARID1A  | XL1        | AS |     | ATM | R    | AT | RX    | AT | AURK                | KB         | AUR  |
| XL    | A | AP1   | B | BARD1   | L2         | BC | 2L1 | BCL | L2L2 | BC | L6    | BC | BCOR                | RL1        | BCO  |
| RCA1  | B | RCA2  | B | BRD4    | IP1        | BR | 1   | BTG | G2   | BT | K     | BT | C11o<br>rf30 (EMSY) | R          | CAL  |
| ASP8  | C | BFB   | C | CBL     | ND1        | CC | D2  | CCN | ND3  | CC | NE1   | CC | CD22                | 74(PD-L1)  | CD2  |
| D79A  | C | D79B  | C | CDC73   | H1         | CD | 12  | CDK | K4   | CD | K6    | CD | CDK8                | N1A        | CDK  |
| DKN2A | C | DKN2B | C | CDKN2C  | BPA        | CE | K1  | CHE | EK2  | CH | C     | CI | CREB                | L          | CRK  |
| SF3R  | C | TCF   | C | CTNNA1  | MNB1       | CT | 3   | CUL | L4A  | CU | CR4   | CK | CYP1                | X          | DAK  |
| DR2   | D | IS3   | D | DNMT3A  | T1L        | DO |     | EED | FR   | EG | 300   | EP | EPHA                | B1         | EPH  |
| RBB2  | E | RBB3  | E | ERBB4   | CC4        | ER |     | ERG | RF11 | ER | R1    | ES | EZH2                | 45C        | FAM  |
| ANCC  | F | ANCG  | F | FANCL   | S          | FA | W7  | PBX | F10  | FG | F12   | FG | FGF1                | 19         | FGF  |
| GF3   | F | GF4   | F | FGF6    | FR1        | FG | R2  | FGF | FR3  | FG | FR4   | FG | PH                  | N          | FLC  |
| LT3   | F | OXL2  | F | FUBP1   | BRA6       | GA | A3  | GAT | TA4  | GA | TA6   | GA | GID                 | 11         | GNA  |
| NAQ   | G | NAS   | G | GRM3    | K3B        | GS | 3A  | H3F | AC1  | HD | F     | HG | 4(C17orf39)         | A          | HNF1 |
| D3    | I | DH1   | I | IDH2    | F1R        | IG | KE  | IKB | ZF1  | IK | PP4B  | IN | IRF2                | 4          | IRF  |
| AK1   | J | AK2   | J | JAK3    | N          | JU | 5A  | KDM | M5C  | KD | M6A   | KD | KDR                 | P1         | KEA  |
| IT    | K | LHL6  | K | KMT2A   | T2D (MLL2) | KM | S   | KRA | K    | LT | N     | LY | MAF                 | 2K1 (MEK1) | MAP  |
| AP2K4 | M | AP3K1 | M | MAP3K13 | PK1        | MA | 1   | MCL | M2   | MD | M4    | MD | MED1                | 2B         | MEF  |

|       |   |             |   |                  |       |    |       |     |            |    |       |    |      |     |     |     |       |
|-------|---|-------------|---|------------------|-------|----|-------|-----|------------|----|-------|----|------|-----|-----|-----|-------|
| ERTK  | M | ET          | M | MITF             | NK1   | MK | 1     | MLH | L          | MP | E11A  | MR | MSH2 | 3   | MSH | H6  | MS    |
| ST1R  | M | TAP         | M | MTOR             | TYH   | MU |       | MYC | CL (MYCL1) | MY | 8     | MY | MYD8 |     | NBN | 1   | NF    |
| F2    | N | FE2L2       | N | NFKB1A           | X2-1  | NK | CH1   | NOT | TCH2       | NO | TCH3  | NO | NPM1 | S   | NRA | 5C2 | NT    |
| TRK1  | N | TRK2        | N | NTRK3            | RY8   | P2 | B2    | PAL | RK2        | PA | RP1   | PA | PARP | 2   | PAR | X5  | PA    |
| BRM1  | P | DCD1 (PD-1) | P | PDCD1LG2 (PD-L2) | GFRA  | PD | FRB   | PDG | K1         | PD | K3C2B | PI | PIK3 | 2G  | 3CA | PIK | PI    |
| IK3R1 | P | IM1         | P | PMS2             | LD1   | PO | E     | POL | ARG        | PP | P2R1A | PP | PPP2 | R2A | M1  | PRD | KAR1A |
| RKCI  | P | TCH1        | P | PTEN             | PN11  | PT | RO    | PTP | I          | QK | C1    | RA | RAD2 | 1   | RAD | 51B | RA    |
| AD51C | R | AD51D       | R | RAD52            | D54L  | RA | 1     | RAF | RA         | RA | 1     | RB | REB1 | 0   | REL | T   | RE    |
| ICTOR | R | NF43        | R | ROS1             | TOR   | RP | A     | SDH | HB         | SD | HC    | SD | SDHD |     | SET | 3B1 | SF    |
| GK1   | S | MAD2        | S | SMAD4            | ARCA4 | SM | RCB1  | SMA | O          | SM | CAIP  | SM | SOCS | 1   | SOX | X9  | SO    |
| PEN   | S | POP         | S | SRC              | AG2   | ST | T3    | STA | K11        | ST | FU    | SU | SYK  | 3   | TEX | K   | TE    |
| ET2   | T | GFBR2       | T | TIPARP           | FAIP3 | TN | BSF14 | TNF | 53         | TP | C1    | TS | TSC2 | 03  | TYR | AF1 | U2    |
| EGFA  | V | HL          | V | WHSC1 (HMS ET)   | SC1L1 | WH |       | WT1 | O1         | XP | CC2   | XR | ZNF2 | 17  | ZNF |     |       |

【표 5B】 선택된 재배열

|     |    |      |      |     |       |               |      |      |    |    |
|-----|----|------|------|-----|-------|---------------|------|------|----|----|
| K   | AL | BC   | BCR  | BR  | BRCA1 | BRC           | CD   | EGFR | ET | ET |
| V6  | ET | L2   | EW   | FR1 | FG    | FGFR2         | K1   | KMT2 | MS | MY |
| C   | MY | SR1  | NO   | NTR | NT    | PDG           | RA   | RARA | RE | RO |
| P02 | RS | TCH2 | K1   | RK2 | NUTM1 | FRA           | F1   | T    | S1 |    |
|     | C4 | SD   | 34A2 | SLC | TE    | TERT(프<br>단독) | BSS2 |      |    |    |

추가의 예시적인 유전자는, 예컨대 국제특허출원공보 제W02012/092426호의 표 1-11에 기재되어 있으며, 이의 내용은 그 전체가 참고로 인용된다.

전술한 방법의 적용은 의료 표본의 서열분석을 위한 특정 유전자 또는 유전자들의 모든 공지된 서열 변이체(또는 이의 하위 집합)를 함유하는 올리고뉴클레오티드의 라이브러리를 사용하는 것을 포함하지만, 이에 제한되지는 않는다.

## 변경의 유형

본원에 기재된 방법은 본원에 기재된 바와 같은 유전체 변경을 평가하기 위한 방법과 조합하여 또는 그의 일부로서 사용될 수 있다.

다양한 유형의 변경(예컨대, 체세포 변경)은 유전체 변경 분석을 위해 평가되고 사용될 수 있다. 예를 들어, 암 및/또는 종양 돌연변이 부담과 관련된 유전체 변경을 분석할 수 있다. 일부 구현예들에서, 본원에 기재된 방법은 낮은 종양 함량 및/또는 적은 양의 종양 핵산을 갖는 샘플을 분석하는 데 유용하다.

체세포 변경

특정 구현예들에서, 본원에 기재된 방법에 따라 평가된 변경은 체세포 변경이다.

특정 구현예들에서, 상기 변경(예컨대, 체세포 변경)은 짧은 부호화 변이체, 예컨대 염기 치환 또는 인델(삽입 또는 결실)이다. 특정 구현예들에서, 상기 변경(예컨대, 체세포 변경)은 점 돌연변이이다. 다른 구현예들에서, 상기 변경(예컨대, 체세포 변경)은 재배열 이외의 것, 예컨대 전좌가 아닌 것이다. 특정 구현예들에서, 상기 변경(예컨대, 체세포 변경)은 스플라이스 변이체이다.

특정 구현예들에서, 상기 변경(예컨대, 체세포 변경)은 침묵 돌연변이, 예컨대 동의어 변경이다. 다른 구현예들에서, 상기 변경(예컨대, 체세포 변경)은 동의어가 아닌 단일 염기 변이체(SNV)이다. 다른 구현예들에서, 상기 변경(예컨대, 체세포 변경)은 일과성(passenger) 돌연변이, 예컨대 세포 클론의 적합성에 대해 검출가능한 효과가 없는 변경이다. 특정 구현예들에서, 상기 변경(예컨대, 체세포 변경)은 미확인 변이체(VUS), 예컨대 병원성이 확인될 수 없고 배제될 수도 없는 변경이다. 특정 구현예들에서, 상기 변경(예컨대, 체세포 변경)은 암 표현형과 관련된 것으로 확인되지 않았다.

특정 구현예들에서, 상기 변경(예컨대, 체세포 변경)은 세포 분열, 성장 또는 생존에 대한 효과와 연관되지 않거나 연관된 것으로 공지되지 않았다. 다른 구현예들에서, 상기 변경(예컨대, 체세포 변경)은 세포 분열, 성장 또는 생존에 대한 효과와 연관된다.

특정 구현예들에서, 증가된 수준의 체세포 변경은 증가된 수준의 하나 이상의 부류 또는 유형의 체세포 변경(예컨대, 재배열, 점 돌연변이, 삽입-결실 또는 이들의 임의의 조합)이다. 특정 구현예들에서, 증가된 수준의 체세포 변경은 증가된 수준의 하나의 부류 또는 유형의 체세포 변경(예컨대, 재배열 단독, 점 돌연변이 단독, 또는 삽입-결실 단독)이다. 특정 구현예들에서, 증가된 수준의 체세포 변경은 한 위치(예컨대, 뉴클레오티드 위치, 예컨대 하나 이상의 뉴클레오티드 위치)에서, 또는 한 영역(예컨대, 뉴클레오티드 영역, 예컨대 하나 이상의 뉴클레오티드 영역)에서의 증가된 수준의 체세포 변경이다. 특정 구현예들에서, 증가된 수준의 체세포 변경은 증가된 수준의 체세포 변경(예컨대, 본원에 기재된 체세포 변경)이다.

## 기능적 변경

특정 구현예들에서, 상기 변경(예컨대, 체세포 변경)은 하위 유전체 간격의 기능적 변경이다. 다른 구현예들에서, 상기 변경(예컨대, 체세포 변경)은 하위 유전체 간격의 공지된 기능적 변경이 아니다. 예를 들어, 종양 돌연변이 부담이 평가될 때 상기 변경(예컨대, 체세포 변경)의 수는 하나 이상의 기능적 변경을 제외할 수 있다.

일부 구현예들에서, 상기 기능적 변경은 참조 서열, 예컨대 야생형 또는 돌연변이되지 않은 서열과 비교하여 세포 분열, 성장 또는 생존에 영향을 미치는, 예컨대 세포 분열, 성장 또는 생존을 촉진하는 변경이다. 특정 구현

예들에서, 상기 기능적 변경은 기능적 변경의 데이터베이스, 예컨대 COSMIC 데이터베이스에 포함시킴으로써 그 자체로 확인된다(cancer.sanger.ac.uk/cosmic; Forbes *et al. Nucl. Acids Res.* 2015; 43 (D1): D805-D811). 다른 구현예들에서, 상기 기능적 변경은 공지된 기능적 상태를 갖는, 예컨대 COSMIC 데이터베이스에서 공지된 체세포 변경으로 발생하는 변경이다. 특정 구현예들에서, 상기 기능적 변경은 가능성 있는 기능적 상태, 예컨대 종양 억제 유전자의 절단을 갖는 변경이다. 특정 구현예들에서, 상기 기능적 변경은 드라이버(driver) 돌연변이, 예컨대 세포 생존 또는 재생을 증가시킴으로써 미세환경에서 클론에 선택적 우위를 제공하는 변경이다. 다른 구현예들에서, 상기 기능적 변경은 클론 확장을 일으킬 수 있는 변경이다. 특정 구현예들에서, 상기 기능적 변경은 하기로서: (a) 성장 신호의 자극자족 (b) 예컨대 항성장 신호에 대한 둔감성 감소; (c) 세포자멸사 감소; (d) 복제 가능성 증가; (e) 지속적인 혈관신생; 또는 (f) 조직 침습 또는 전이 중에서 1, 2, 3, 4, 5 개 또는 모두를 유발할 수 있는 변경이다.

[0336] 특정 구현예들에서, 상기 변경은 일과성 돌연변이가 아니고, 예컨대 세포 클론의 적합성에 대해 검출가능한 효과가 없는 변경이 아니다. 특정 구현예들에서, 상기 변경은 미확인 변이체(VUS)가 아니고, 예컨대 병원성이 확인될 수 없고 배제될 수도 없는 변경이 아니다.

[0337] 특정 구현예들에서, 표 2A-5B에 기재된 유전자에서 복수의 기능적 변경(예컨대, 약 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% 이상)은 제외된다. 특정 구현예들에서, 표 2A-5B에 기재된 유전자의 모든 기능적 변경은 제외된다. 특정 구현예들에서, 표 2A-5B에 기재된 복수의 유전자에서 복수의 기능적 변경은 배제된다. 특정 구현예들에서, 표 2A-5B에 기재된 모든 유전자의 모든 기능적 변경은 제외된다.

[0338] 생식계열 변경

[0339] 특정 구현예들에서, 상기 변경은 생식계열 변경이다. 다른 구현예들에서, 상기 변경은 생식계열 변경이 아니다. 특정 구현예들에서, 상기 변경은 생식계열 변경과 동일하거나 유사하지 않고, 예컨대 구별가능하다. 예를 들어, 종양 돌연변이 부담이 평가될 때 변경의 수는 생식계열 변경의 수를 제외할 수 있다.

[0340] 특정 구현예들에서, 상기 생식계열 변경은 단일 염기 다형성(SNP), 염기 치환, 인델(예컨대, 삽입 또는 결실), 또는 침묵 변경(예컨대, 동의어 변경)이다.

[0341] 특정 구현예들에서, 상기 생식계열 변경은 일치된 정상 서열과의 비교를 사용하지 않는 방법의 사용에 의해 확인된다. 다른 구현예들에서, 상기 생식계열 변경은 SGZ 알고리즘의 사용을 포함하는 방법에 의해 확인된다. 특정 구현예들에서, 상기 생식계열 변경은 생식계열 변경의 데이터베이스, 예컨대 dbSNP 데이터베이스에 포함시킴으로써 그 자체로 확인된다(www.ncbi.nlm.nih.gov/SNP/index.html; Sherry *et al. Nucleic Acids Res.* 2001; 29(1): 308-311). 다른 구현예들에서, 상기 생식계열 변경은 ExAC 데이터베이스의 2개 이상의 카운트에 포함시킴으로써 그 자체로 확인된다(exac.broadinstitute.org; Exome Aggregation Consortium *et al.* "Analysis of protein-coding genetic variation in 60,706 humans," bioRxiv preprint. October 30, 2015). 일부 구현예들에서, 상기 생식계열 변경은 1000 유전체 프로젝트 데이터베이스에 포함시킴으로써 그 자체로 확인된다(www.1000genomes.org; McVean *et al. Nature.* 2012; 491, 56-65). 일부 구현예들에서, 상기 생식계열 변경은 ESP 데이터베이스에 포함시킴으로써 그 자체로 확인된다(엑솜 변이체 서버(Exome Variant Server), NHLBI GO 엑솜 서열분석 프로젝트(Exome Sequencing Project, ESP), 워싱턴주 시애틀 소재(avs.gs.washington.edu/EVS/)).

[0342] 샘플

[0343] 본원에 기재된 방법은 다수의 상이한 공급원으로부터의 다양한 유형의 샘플에서 종양 분석을 평가하는 데 사용될 수 있다.

[0344] 일부 구현예들에서, 상기 샘플은 핵산, 예컨대 DNA, RNA 또는 둘 모두를 포함한다. 특정 구현예들에서, 상기 샘플은 종양으로부터의 하나 이상의 핵산을 포함한다. 특정 구현예들에서, 상기 샘플은 종양, 예컨대 세포, 단백질, 탄수화물 또는 지질로부터의 하나 이상의 비핵산 성분을 추가로 포함한다. 특정 구현예들에서, 상기 샘플은 비종양 세포 또는 조직으로부터의 하나 이상의 핵산을 추가로 포함한다.

[0345] 특정 구현예들에서, 상기 샘플은 액체 생검으로부터 수득한다. 특정 구현예들에서, 상기 샘플은 조직 생검으로부터 수득하지 않는다. 특정 구현예에서, 상기 샘플은 액체 샘플이다. 특정 구현예들에서, 상기 샘플에는 고체가 없거나 본질적으로 없다.

[0346] 특정 구현예들에서, 상기 샘플은 고형 종양, 혈액암, 또는 이의 전이성 형태를 갖는 대상체로부터 수득한다. 특정 구현예들에서, 상기 샘플은 암을 가지거나 또는 암을 가질 위험이 있는 대상체로부터 수득한다. 특정 구현예

들에서, 상기 샘플은 본원에 기재된 바와 같이 암을 치료하기 위한 요법을 받은 적이 없거나, 암을 치료하기 위한 요법을 받고 있거나, 암을 치료하기 위한 요법을 받은 대상체로부터 수득한다.

- [0347] 일부 구현예들에서, 상기 샘플은 전암성 또는 악성 세포로부터의 하나 이상의 핵산, 예컨대 DNA, RNA 또는 둘 모두, 고형 종양, 연조직 종양 또는 전이성 병변으로부터의 세포, 혈액암으로부터의 세포, 조직학적으로 정상인 세포, 순환 종양 세포(CTC), 또는 이들의 조합을 포함한다. 일부 구현예들에서, 상기 샘플은 전암성 또는 악성 세포로부터 선택된 하나 이상의 세포, 고형 종양, 연조직 종양 또는 전이성 병변으로부터의 세포, 혈액암으로부터의 세포, 조직학적으로 정상인 세포, 순환 종양 세포(CTC), 또는 이들의 조합을 포함한다.
- [0348] 특정 구현예들에서, 상기 샘플은 무세포 DNA(cfDNA)를 포함한다. 특정 구현예들에서, 상기 샘플은 순환 종양 DNA(ctDNA)를 포함한다. 특정 구현예들에서, 상기 샘플은 혈액, 혈청 또는 혈장을 포함한다. 특정 구현예들에서, 상기 샘플은 뇌척수액(CSF)을 포함한다. 특정 구현예들에서, 상기 샘플은 흉수를 포함한다. 특정 구현예들에서, 상기 샘플은 복수를 포함한다. 특정 구현예들에서, 상기 샘플은 소변을 포함한다. 특정 구현예들에서, 상기 샘플은 절제술, 바늘 생검, 미세 바늘 흡인물, 또는 세포검사 도말을 포함한다. 특정 구현예들에서, 상기 샘플은 포르말린 고정 파라핀 포매(FFPE) 샘플이다.
- [0349] 다양한 조직이 본 방법에 사용되는 샘플의 공급원이 될 수 있다. 유전체 또는 하위 유전체 핵산(예컨대, DNA 또는 RNA)은 대상체의 샘플(예컨대, 종양 세포를 포함하는 샘플, 혈액 샘플, 혈액 성분 샘플, 무세포 DNA(cfDNA))를 포함하는 샘플, 순환 종양 DNA(ctDNA)를 포함하는 샘플, 순환 종양 세포(CTC)를 포함하는 샘플, 또는 임의의 정상 대조군(예컨대, 정상 인접 조직(NAT))으로부터 분리될 수 있다.
- [0350] 일부 구현예들에서, 상기 샘플은 종양으로부터의 핵산, 예컨대 DNA, RNA 또는 둘 모두를 포함한다. 상기 핵산은 DNA 또는 RNA일 수 있다. 특정 구현예들에서, 상기 샘플은, 예컨대 상기 종양으로부터의 비핵산 성분, 예컨대 세포, 단백질, 탄수화물, 또는 지질을 추가로 포함한다. 특정 구현예들에서, 상기 샘플은 정상 세포 또는 조직으로부터의 핵산을 추가로 포함한다.
- [0351] 특정 구현예들에서, 상기 샘플은 동결 샘플로서 또는 포르말데히드 또는 파라포르말데히드 고정 파라핀 포매(FFPE) 조직 표본으로서 보존된다. 예를 들어, 상기 샘플은 매트릭스, 예컨대 FFPE 블록 또는 동결 샘플 내에 포함될 수 있다. 특정 구현예들에서, 상기 샘플은 혈액 샘플이다. 특정 구현예들에서, 상기 조직 샘플은 혈액 성분 샘플이다. 특정 구현예들에서, 상기 샘플은 cfDNA 샘플이다. 특정 구현예들에서, 상기 샘플은 ctDNA 샘플이다. 특정 구현예들에서, 상기 샘플은 CTC 샘플이다. 다른 구현예들에서, 상기 조직 샘플은 골수 흡인물(BMA) 샘플이다. 분리 단계는 개별 염색체의 흐름 분류; 및/또는 대상체의 샘플(예컨대, 본원에 기재된 샘플)의 미세 해부를 포함할 수 있다.
- [0352] 다른 구현예들에서, 상기 샘플은 하나 이상의 전암성 또는 악성 세포를 포함한다. 특정 구현예들에서, 상기 샘플은 고형 종양, 연조직 종양, 또는 전이성 병변으로부터 획득된다. 특정 구현예들에서, 상기 샘플은 혈액 악성종양 또는 전암으로부터 수득한다. 다른 구현예들에서, 상기 샘플은 수술 절제면으로부터의 조직 또는 세포를 포함한다. 특정 구현예들에서, 상기 샘플은 종양 침윤성 림프구를 포함한다. 상기 샘플은 조직학적 정상 조직일 수 있다. 한 구현예에서, 상기 샘플은 하나 이상의 비악성 세포를 포함한다.
- [0353] 특정 구현예들에서, 상기 FFPE 샘플은 하기 특성들 중 1개, 2개, 또는 모두를 갖는다: (a) 약  $10\text{ mm}^2$  이상, 약  $25\text{ mm}^2$  이상, 또는 약  $50\text{ mm}^2$  이상의 표면적을 갖고; (b) 약  $0.1\text{ mm}^3$  이상, 약  $0.2\text{ mm}^3$  이상, 약  $0.3\text{ mm}^3$  이상, 약  $0.4\text{ mm}^3$  이상, 약  $0.5\text{ mm}^3$  이상, 약  $0.6\text{ mm}^3$  이상, 약  $0.7\text{ mm}^3$  이상, 약  $0.8\text{ mm}^3$  이상, 약  $0.9\text{ mm}^3$  이상, 약  $1\text{ mm}^3$  이상, 약  $2\text{ mm}^3$  이상, 약  $3\text{ mm}^3$  이상, 약  $4\text{ mm}^3$  이상, 또는 약  $5\text{ mm}^3$  이상의 샘플 부피를 갖고; (c) 약 50% 이상, 약 60% 이상, 약 70% 이상, 약 80% 이상, 또는 약 90% 이상의 세포충실도를 갖고; 및/또는 (d) 약 10,000개 이상의 세포, 약 20,000개 이상의 세포, 약 30,000개 이상의 세포, 약 40,000개 이상의 세포, 또는 약 50,000개 이상의 세포의 유효 세포의 수를 갖는다.
- [0354] 일 구현예에서, 상기 방법은 샘플, 예컨대 본원에 기재된 샘플을 수득하는 단계를 추가로 포함한다. 상기 샘플은 직접 또는 간접적으로 수득할 수 있다. 한 구현예에서, 상기 샘플은 cfDNA를 포함하는 샘플로부터, 예컨대 분리 또는 정제에 의해 수득한다. 한 구현예에서, 상기 샘플은 ctDNA를 포함하는 샘플로부터, 예컨대 분리 또는 정제에 의해 수득한다. 한 구현예에서, 상기 샘플은 악성 세포 및 비악성 세포(예컨대, 종양 침윤 림프구) 둘 모두를 포함하는 샘플로부터, 예컨대 분리 또는 정제에 의해 수득한다. 한 구현예에서, 상기 샘플은 CTC를 포함하는 샘플로부터, 예컨대 분리 또는 정제에 의해 수득한다.

- [0355] 다른 구현예들에서, 상기 방법은 본원에 기재된 방법을 사용하여, 예컨대 수술 절제면으로부터의 샘플, 예컨대 조직학적 정상인 샘플을 평가하는 단계를 포함한다. 일부 구현예들에서, 조직학적 정상 조직(예컨대, 그렇지 않으면 조직학적 정상 조직의 절제면)으로부터 수득한 샘플은 여전히 본원에 기재된 바와 같은 변경을 여전히 가질 수 있다. 따라서, 상기 방법은 검출된 변경의 존재에 기초하여 샘플을 재분류하는 단계를 추가로 포함할 수 있다. 한 구현예에서, 예컨대 상이한 대상체로부터의 다수의 샘플이 동시에 처리된다.
- [0356] 한 구현예에서, 상기 방법은 샘플로부터 핵산을 분리하여 단리된 핵산 샘플을 제공하는 단계를 포함한다. 한 구현예에서, 상기 방법은 대조군으로부터 핵산을 분리하여 분리된 핵산 샘플을 제공하는 단계를 포함한다. 한 구현예에서, 상기 방법은 검출 가능한 핵산이 없는 샘플을 거부하는 단계를 추가로 포함한다.
- [0357] 한 구현예에서, 상기 방법은 1차 대조군이 입수 가능한지 여부를 결정하고, 만약 그렇다면 상기 1차 대조군으로부터 대조군 핵산(예컨대, DNA)을 분리하는 단계를 추가로 포함한다. 한 구현예에서, 상기 방법은 NAT가 상기 샘플에 존재하는지 여부를 결정하는 단계를 추가로 포함한다(예컨대, 어떠한 1차 대조군 샘플도 입수 가능하지 않은 경우). 한 구현예에서, 상기 방법은, 예컨대 1차 대조군을 수반하지 않는 샘플에서 상기 NAT로부터 비종양 조직을 거대해부함으로써 비종양 세포에 대해 농축된 하위 샘플을 수득하는 단계를 추가로 포함한다. 한 구현예에서, 상기 방법은 그 어떤 1차 대조군 및 NAT도 입수 가능하지 않음을 결정하고 일치된 대조군 없이 분석을 위해 상기 샘플을 표시하는 단계를 추가로 포함한다.
- [0358] 한 구현예에서, 방법은 상기 샘플에서 핵산 수율에 대한 값을 수득하고 상기 수득한 값을 참조 기준과 비교하며, 예컨대 상기 수득한 값이 상기 참조 기준보다 작으면 라이브러리 구성 이전에 핵산을 증폭시키는 단계를 추가로 포함한다. 한 구현예에서, 방법은 상기 샘플에서 핵산 단편의 크기에 대한 값을 수득하고, 수득한 값을 참조 기준, 예컨대 적어도 300, 600 또는 900 bps의 크기, 예컨대 평균 크기와 비교하는 단계를 추가로 포함한다. 본원에 기재된 매개변수는 상기 결정에 응답하여 조정되거나 선택될 수 있다.
- [0359] 특정 구현예들에서, 상기 방법은 숙성된 샘플, 예컨대 숙성된 FFPE 샘플로부터 핵산을 분리하는 단계를 포함한다. 상기 숙성된 샘플은 수년, 예컨대 1년, 2년, 3년, 4년, 5년, 10년, 15년, 20년, 25년, 50년, 75년 또는 100년 이상일 수 있다.
- [0360] 다양한 크기의 샘플로부터 핵산을 수득할 수 있다. 예를 들어, 핵산은 5 내지 200  $\mu\text{m}$  또는 그 이상의 샘플로부터 분리할 수 있다. 예를 들어, 상기 샘플은 5  $\mu\text{m}$ , 10  $\mu\text{m}$ , 20  $\mu\text{m}$ , 30  $\mu\text{m}$ , 40  $\mu\text{m}$ , 50  $\mu\text{m}$ , 70  $\mu\text{m}$ , 100  $\mu\text{m}$ , 110  $\mu\text{m}$ , 120  $\mu\text{m}$ , 150  $\mu\text{m}$  또는 200  $\mu\text{m}$  이상으로 측정할 수 있다.
- [0361] 샘플로부터의 DNA 분리를 위한 프로토콜은, 예컨대 국제특허출원공보 제WO 2012/092426호의 실시예 1에 제공된 바와 같이 당업계에 공지되어 있다. 포름알데히드 또는 파라포름알데히드 고정, 파라핀 포매(FFPE) 조직으로부터 핵산(예컨대, DNA)을 분리하는 추가적인 방법이, 예컨대 Cronin M. *et al.*, (2004) *Am J Pathol.* 164(1):35-42; Masuda N. *et al.*, (1999) *Nucleic Acids Res.* 27(22):4436-4443; Specht K. *et al.*, (2001) *Am J Pathol.* 158(2):419-429, 앰비온 리커버올(Ambion RecoverAll)<sup>TM</sup> 전체 핵산 분리 프로토콜(Total Nucleic Acid Isolation Protocol, 앰비온 Cat. No. AM1975, 2008년 9월), 맥스웰(Maxwell)<sup>®</sup> 16 FFPE 플러스 LEV DNA 정제 키트 기술 설명서(프로메가(Promega) 문헌 #TM349, 2011년 2월), E.Z.N.A.<sup>®</sup> FFPE DNA 키트 소책자(오메가 바이오-텍(OMEGA bio-tek), 노르크로스(Norcross), GA, 제품 번호 D3399-00, D3399-01, 및 D3399-02; 2009년 6월), and QIAamp<sup>®</sup> DNA FFPE 조직 소책자(키아젠(Qiagen), Cat. No. 37625, 2007년 10월)에 개시되어 있다. 리커버올(Ambion RecoverAll)<sup>TM</sup> 전체 핵산 분리 키트는 고온에서 자일렌을 사용하여 파라핀 포매된 샘플을 가용화하고 유리 섬유 필터를 사용하여 핵산을 캡처한다. 맥스웰<sup>®</sup> 16 FFPE 플러스 LEV DNA 정제 키트는 FFPE 조직의 1 내지 10  $\mu\text{m}$  섹션에서 유전체 DNA를 정제하기 위해 맥스웰<sup>®</sup> 16 인스트루먼트(Instrument)와 함께 사용한다. DNA는 실리카 피복 상자성 입자(PMP)를 사용하여 정제하고 낮은 용리 부피로 용리한다. E.Z.N.A.<sup>®</sup> FFPE DNA 키트는 유전체 DNA 분리를 위해 스핀 컬럼(spin column) 및 완충계를 사용한다. QIAamp<sup>®</sup> DNA FFPE 조직 키트는 유전체 및 미토콘드리아 DNA 정제를 위해 QIAamp<sup>®</sup> DNA 마이크로 기술을 사용한다. 혈액으로부터 DNA 분리를 위한 프로토콜은, 예컨대 맥스웰<sup>®</sup> 16 LEV 혈액 DNA 키트 및 맥스웰 16 구강 면봉 LEV DNA 정제 키트 기술 매뉴얼(프로메가 문헌 #TM333, 2011년 1월 1일)에 개시되어 있다.
- [0362] RNA 분리를 위한 프로토콜은, 예컨대 맥스웰<sup>®</sup> 16 전체 RNA 정제 키트 기술 게시판(프로메가 문헌 #TB351, 2009년 8월)에 개시되어 있다.
- [0363] 분리된 핵산(예컨대, 유전체 DNA)은 일상적인 기술을 실행하여 단편화 또는 전단될 수 있다. 예를 들어, 유전체 DNA는 물리적 전단 방법, 효소적 절단 방법, 화학적 절단 방법, 및 당업자에게 일반적으로 공지된 기타 방법에

의해 단편화될 수 있다. 핵산 라이브러리는 유전체의 복잡성을 모두 또는 실질적으로 모두 함유할 수 있다. 상기 맥락에서 "실질적으로 모두"라는 용어는 절차의 초기 단계에서 실제로 원치 않는 유전체 복잡성의 손실이 있을 수 있는 가능성을 지칭한다. 본원에 기재된 방법은 또한 핵산 라이브러리가 상기 유전체의 일부인 경우, 예컨대 상기 유전체의 복잡성이 설계에 의해 감소되는 경우에 유용하다. 일부 구현예들에서, 상기 유전체의 임의의 선택된 부분이 본원에 기재된 방법과 함께 사용될 수 있다. 특정 구현예들에서, 전체 엑솜 또는 이의 하위 집합이 분리된다.

[0364] 특정 구현예들에서, 상기 방법은 상기 샘플로부터 핵산을 분리하여 라이브러리(예컨대, 본원에 기재된 바와 같은 핵산 라이브러리)를 제공하는 단계를 추가로 포함한다. 특정 구현예들에서, 상기 샘플은 전체 유전체, 하위 유전체 단편, 또는 둘 모두를 포함한다. 상기 분리된 핵산을 사용하여 핵산 라이브러리를 제조할 수 있다. 전체 유전체 또는 하위 유전체 단편으로부터 라이브러리를 분리하고 제조하기 위한 프로토콜은 당업계에 공지되어 있다(예컨대, 일루미나(Illumina)의 유전체 DNA 샘플 제조 키트). 특정 구현예들에서, 상기 유전체 또는 하위 유전체 DNA 단편은 대상체의 샘플(예컨대, 본원에 기재된 샘플)로부터 분리된다. 일 구현예에서, 상기 샘플은 보조된 표본이며, 예컨대 매트릭스, 예컨대 FFPE 블록 또는 동결된 샘플에 매립된다. 특정 구현예들에서, 상기 분리 단계는 개별 염색체의 흐름 분류; 및/또는 샘플의 미세해부를 포함한다. 특정 구현예들에서, 핵산 라이브러리를 생성하기 위해 사용된 핵산의 양은 5 마이크로그램 미만, 1 마이크로그램 미만, 또는 500 ng 미만, 200 ng 미만, 100 ng 미만, 50 ng 미만, 10 ng 미만, 5 ng 미만, 또는 1 ng 미만이다.

[0365] 또 다른 구현예에서, 상기 라이브러리를 생성하기 위해 사용된 핵산은 RNA 또는 RNA로부터 유래된 cDNA를 포함한다. 일부 구현예들에서, 상기 RNA는 전체 세포 RNA를 포함한다. 다른 구현예들에서, 특정 풍부한 RNA 서열(예컨대, 리보솜 RNA)이 고갈되었다. 일부 구현예들에서, 전체 RNA 제제 중 폴리(A)-꼬리 mRNA 분획이 농축되었다. 일부 구현예들에서, cDNA는 무작위 프라이밍된 cDNA 합성 방법에 의해 생성된다. 다른 구현예들에서, 상기 cDNA 합성은 올리고(dT) 함유 올리고뉴클레오타이드에 의한 프라이밍에 의해 성숙한 mRNA의 폴리(A) 꼬리에서 개시된다. 고갈, 폴리(A) 농축 및 cDNA 합성을 위한 방법은 당업자에게 일반적으로 공지되어 있다.

[0366] 다른 구현예들에서, 상기 핵산은 물리적 또는 효소적 방법에 의해 단편화 또는 전단되고, 선택적으로 합성 어댑터에 결합되고, 크기가 선택되고(예컨대, 예비 겔 전기영동에 의해) 증폭된다(예컨대, PCR에 의해). DNA 전단을 위한 대안적인 방법은, 예컨대 국제특허출원공보 제WO 2012/092426호의 실시예 4에 기재된 바와 같이 당업계에 공지되어 있다. 예를 들어, 대안적인 DNA 전단 방법은 더 자동화되고/되거나 더 효율적일 수 있다(예컨대, 분해된 FFPE 샘플을 사용). DNA 전단 방법에 대한 대안은 라이브러리 제조 중 결합 단계를 피하기 위해 사용할 수도 있다.

[0367] 다른 구현예들에서, 상기 분리된 DNA(예컨대, 유전체 DNA)는 단편화되거나 전단된다. 일부 구현예들에서, 상기 라이브러리는 50% 미만의 유전체 DNA, 예컨대 다른 수단에 의해 하위분획화된 유전체의 감소된 표현 또는 정의된 부분인 유전체 DNA의 하위분획을 포함한다. 다른 구현예들에서, 상기 라이브러리는 모든 또는 실질적으로 모든 유전체 DNA를 포함한다.

[0368] 다른 구현예들에서, 핵산의 단편화되고 어댑터 결합된 군은 혼성 선택 전에 명시적인 크기 선택 또는 증폭 없이 사용된다. 일부 구현예들에서, 상기 핵산은 당업자에게 일반적으로 공지된 특이적 또는 비특이적 핵산 증폭 방법에 의해 증폭된다. 일부 구현예들에서, 상기 핵산은, 예컨대 무작위 프라이밍된 가닥 치환 증폭과 같은 전체 유전체 증폭 방법에 의해 증폭된다.

[0369] 본원에 기재된 방법은, 예컨대 공급원 DNA 또는 RNA의 양이 제한적일 때(예컨대, 전체 유전체 증폭 후에도) 소량의 핵산을 사용하여 실시할 수 있다. 일 구현예에서, 상기 핵산은 약 5  $\mu$ g, 4  $\mu$ g, 3  $\mu$ g, 2  $\mu$ g, 1  $\mu$ g, 0.8  $\mu$ g, 0.7  $\mu$ g, 0.6  $\mu$ g, 0.5  $\mu$ g, 또는 400 ng, 300 ng, 200 ng, 100 ng, 50 ng, 10 ng, 5 ng, 1 ng 미만의 핵산 샘플을 포함한다. 예를 들어, 일반적으로 50-100 ng의 유전체 DNA로 시작할 수 있다. 하지만, 혼성화 단계, 예컨대 용액 혼성화 이전에 유전체 DNA를 증폭하는 경우(예컨대, PCR을 사용) 더 적게 시작할 수 있다. 따라서, 혼성화, 예컨대 용액 혼성화 이전에 유전체 DNA를 증폭하는 것이 가능하지만 필수적인 것은 아니다.

[0370] 한 구현예에서, 상기 샘플은 비암세포 또는 비악성 세포, 예컨대 종양 침윤 림프구로부터의 DNA, RNA(또는 RNA로부터 유래된 cDNA), 또는 둘 모두를 포함한다. 한 구현예에서, 상기 샘플은 비암세포 또는 비악성 세포, 예컨대 종양 침윤 림프구로부터의 DNA, RNA(또는 RNA로부터 유래된 cDNA), 또는 둘 모두를 포함하고, 암세포 또는 악성 세포로부터의 DNA, RNA(또는 RNA에서 파생된 cDNA) 또는 둘 모두를 포함하지 않거나 본질적으로 없다.

[0371] 한 구현예에서, 상기 샘플은 암세포 또는 악성 세포로부터의 DNA, RNA(또는 RNA로부터 유래된 cDNA)를

포함한다. 한 구현예에서, 상기 샘플은 암세포 또는 악성 세포로부터의 DNA, RNA(또는 RNA로부터 유래된 cDNA), 또는 둘 모두를 포함하고, 비암세포 또는 비악성 세포, 예컨대 종양 침윤 림프구로부터의 DNA, RNA(또는 RNA에서 파생된 cDNA) 또는 둘 모두를 포함하지 않거나 본질적으로 없다.

[0372] 한 구현예에서, 상기 샘플은 비암세포 또는 비악성 세포, 예컨대 종양 침윤 림프구로부터의 DNA, RNA(또는 RNA로부터 유래된 cDNA), 또는 둘 모두를 포함하고, 암세포 또는 악성 세포로부터의 DNA, RNA(또는 RNA에서 파생된 cDNA) 또는 둘 모두를 포함한다.

[0373] 특정 구현예들에서, 상기 샘플은 암을 가진 대상체로부터 취득된다. 예시적인 암은 B 세포 암, 예컨대 다발성 골수종, 흑색종, 유방암, 폐암(예컨대, 비소세포 폐암 또는 NSCLC), 기관지암, 결장직장암, 전립선암, 췌장암, 위암, 난소암, 방광암, 뇌 또는 중추신경계암, 말초신경계암, 식도암, 자궁경부암, 자궁암 또는 자궁내막암, 구강암 또는 인두암, 간암, 신장암, 고환암, 담도암, 소장암 또는 맹장암, 타액선암, 갑상선암, 부신암, 골육종, 연골육종, 혈액조직암, 선암종, 염증성 근섬유모세포종, 위장관기질종양(GIST), 결장암, 다발성 골수종(MM), 골수이형성 증후군(MDS), 골수증식성 장애(MPD), 급성 림프성 백혈병(ALL), 급성 골수성 백혈병(AML), 만성 골수성 백혈병(CML), 만성 림프구성 백혈병(CLL), 진성다혈구증, 호지킨 림프종, 비호지킨 림프종(NHL), 연조직 육종, 섬유육종, 점액육종, 지방육종, 골육종, 척색종, 혈관육종, 내피육종, 림프관육종, 림프관내피육종, 활막종, 증피종, 유잉종양, 평활근육종, 횡문근육종, 편평세포암종, 기저세포암종, 선암종, 땀샘암종, 피지선암종, 유도모양암종, 유도모양샘암종, 수질암종, 기관지암종, 신세포암종, 간종양, 신세포암종, 담낭암종, 용모암종, 정상피종, 배아암종, 윌름스 종양, 방광암종, 상피암종, 신경교종, 성상세포종, 수모세포종, 두개인두종, 뇌실막종, 송과체종, 혈관모세포종, 청각신경종, 희소돌기아교종, 수막종, 망막모세포종, 여포성 림프종, 미만성 거대B세포림프종, 외투세포림프종, 간세포암종, 갑상선암, 위암, 두경부암, 소세포암, 본태성 혈소판증가증, 원인불명골수화생, 과호산구성 증후군, 전신 비만세포증, 친숙한 과호산구 증가증, 만성 호산구성 백혈병, 신경 내분비암, 카르시노이드 종양 등을 포함하나 이에 제한되지는 않는다.

[0374] 한 구현예에서, 상기 암은 혈액암(또는 전암)이다. 본원에서 사용된 혈액암은 조혈 또는 림프 조직의 종양, 예컨대 혈액, 골수 또는 림프절에 영향을 미치는 종양을 지칭한다. 예시적인 혈액암은 백혈병(예컨대, 급성 림프모구성 백혈병(ALL), 급성 골수성 백혈병(AML), 만성 림프구성 백혈병(CLL), 만성 골수성 백혈병(CML), 모세포 백혈병, 급성 단핵구 백혈병(AMoL), 만성 골수단핵구 백혈병(CMML), 연소형 골수단구성 백혈병(JMML) 또는 거대과립 림프구성 백혈병, 림프종(예컨대, AIDS 관련 림프종, 피부 T 세포 림프종, 호지킨 림프종(예컨대, 고전적 호지킨 림프종 또는 결절성 림프구 우위성 호지킨 림프종), 균상 식육종, 비호지킨 림프종(예컨대, B 세포 비호지킨 림프종(예컨대, 버킷 림프종, 소형림프구성 림프종(CLL/SLL), 미만성 거대 B 세포 림프종, 여포성 림프종, 면역아구성 대세포 림프종, 전구 B 림프모구 림프종 또는 외투 세포 림프종) 또는 T 세포 비호지킨 림프종(균상 식육종, 역형성 거대 세포 림프종 또는 전구 T 림프모구 림프종)), 원발성 중추신경계 림프종, 세자리 증후군, 발덴스트롬 거대글로블린혈증), 만성 골수증식성 신생물, 랑게르한스 세포 조직구증, 다발성 골수종/형질 세포 신생물, 골수이형성 증후군 또는 골수이형성/골수증식성 신생물을 포함하지만 이에 제한되지는 않는다. 본원에서 사용된 전암은 아직 악성은 아니지만 악성이 될 태세인 조직을 지칭한다.

[0375] 일부 구현예들에서, 본원에 기재된 샘플은 표본이라고도 한다. 일부 구현예들에서, 상기 샘플은 조직 샘플, 혈액 샘플 또는 골수 샘플이다.

[0376] 일부 구현예들에서, 상기 혈액 샘플은 무세포 DNA(cfDNA)를 포함한다. 일부 구현예들에서, cfDNA는 건강한 조직, 예컨대 비질환 세포, 또는 종양 조직, 예컨대 종양 세포로부터의 DNA를 포함한다. 일부 구현예들에서 종양 조직으로부터의 cfDNA는 순환하는 종양 DNA(ctDNA)를 포함한다. 일부 구현예들에서, ctDNA 샘플은, 예컨대 고형 종양, 예컨대 폐암, 유방암 또는 결장암을 가진 환자로부터 취득하고 수집된다.

[0377] 일부 구현예들에서, 상기 샘플, 예컨대 표본은 포르말린 고정 파라핀 포매(FFPE) 표본이다. 일부 구현예들에서, 상기 FFPE 표본은 중심부 바늘 생검, 미세 바늘 흡인물, 또는 삼출 세포검사에서 선택된 표본을 포함하지만 이에 제한되지는 않는다. 일부 구현예들에서, 상기 샘플은 FFPE 블록 및 하나의 원래 헤마톡실린 및 에오신(H&E) 염색된 슬라이드를 포함한다. 일부 구현예들에서, 상기 샘플은 비착색 슬라이드(예컨대, 양으로 하전된, 베이킹되지 않은 및 4-5 마이크론 두께; 예컨대, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 상기 슬라이드) 및 하나 이상의 H&E 착색 슬라이드를 포함한다.

[0378] 일부 구현예들에서, 상기 샘플은 FFPE 블록 또는 비착색 슬라이드, 예컨대 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16개 이상의 비착색 슬라이드 및 하나 이상의 H&E 슬라이드를 포함한다. 일부 구현예들에서, 상기 샘플은, 예컨대 본원에 기재된 바와 같이, 예컨대 표준 고정 방법을 사용하여 포르말린 고정되고 파라

핀 블록에 매립된 조직을 포함한다.

- [0379] 일부 구현예들에서, 상기 샘플은 적어도  $1-30\text{mm}^2$ , 예컨대 약  $5-25\text{mm}^2$ 의 표면적을 포함한다. 일부 구현예들에서, 상기 샘플은 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는  $10\text{mm}^2$ , 예컨대  $5\text{mm}^2$ 의 표면적을 포함한다. 일부 구현예들에서, 상기 샘플은 적어도  $5\text{mm}^2$ 의 표면적을 포함한다. 일부 구현예들에서, 상기 샘플은 약 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 또는  $30\text{mm}^2$ , 예컨대  $25\text{mm}^2$ 의 표면적을 포함한다. 일부 구현예들에서, 상기 샘플은  $25\text{mm}^2$ 의 표면적을 포함한다.
- [0380] 일부 구현예들에서, 상기 샘플은 적어도  $1-5\text{mm}^3$ , 예컨대 약  $2^3$ 의 표면 부피를 포함한다. 일부 구현예들에서, 약  $2\text{mm}^3$ 의 표면 부피는 약 80 미크론, 예컨대 80 미크론 이상 또는 초과와 같은 깊이에서 약  $25\text{mm}^2$ 의 표면적을 갖는 샘플을 포함한다.
- [0381] 일부 구현예들에서, 상기 샘플은, 예컨대 종양 핵을 포함하는 종양 함량을 포함한다. 일부 구현예들에서, 상기 샘플은 적어도 5-50%, 10-40%, 15-25%, 또는 20-30% 종양 핵을 갖는 종양 함량을 포함한다. 일부 구현예들에서, 상기 샘플은 20% 이상의 종양 핵의 종양 함량을 포함한다. 일부 구현예들에서, 상기 샘플은 약 30%의 종양 핵의 종양 함량을 포함한다. 일부 구현예들에서, 종양 핵 퍼센트는, 예컨대 종양 세포의 수를 핵을 갖는 모든 세포의 총수로 나눔으로써 계산된다. 일부 구현예들에서, 상기 샘플이 간 샘플, 예를 들어 간세포를 포함하는 경우, 더 높은 종양 함량이 요구될 수 있다. 일부 구현예들에서, 간세포는 다른, 예컨대 비간세포, 체핵의 DNA 함량이 2배인 핵을 갖는다. 일부 구현예들에서, 예컨대 본원에 기재된 바와 같은 변경의 검출 감도는 상기 샘플의 종양 함량에 의존하며, 예컨대 종양 함량이 낮을수록 더 낮은 검출 감도를 유발할 수 있다.
- [0382] 일부 구현예들에서, DNA는 상기 샘플의 유핵 세포로부터 추출된다. 일부 구현예들에서, 샘플은, 예컨대 상기 샘플이 주로 적혈구, 과도한 세포질을 함유하는 병변 세포, 또는 섬유증이 있는 조직으로 구성되는 경우 낮은 유핵 세포충실도를 갖는다. 일부 구현예들에서, 유핵 세포충실도가 낮은 샘플은 DNA 추출을 위해 더 많은, 예컨대 더 큰 조직 부피, 예컨대  $2\text{mm}^3$  초과를 요구할 수 있다.
- [0383] 일부 구현예들에서, 상기 FPPE 샘플, 예컨대 표본은 핵산 완전성을 보존하기 위해 표준 고정 방법을 사용하여 제조된다. 일부 구현예들에서, 상기 표준 고정 방법은, 예컨대 6-72시간 동안 10% 중성 완충 포르말린을 사용하는 것을 포함한다. 일부 구현예들에서, 상기 방법은 네덜란드의 보우인(Bouins), B5, AZF와 같은 고정제를 포함하지 않는다. 일부 구현예들에서, 상기 방법은 탈석회화를 포함하지 않는다. 일부 구현예들에서, 상기 방법은 탈석회화를 포함한다. 구현예들에서, 탈석회화는 EDTA로 실시된다. 일부 구현예들에서, 강산, 예컨대 염산, 황산 또는 피크르산은 탈석회화에 사용되지 않는다.
- [0384] 일부 구현예들에서, 상기 샘플은 FPPE 블록 또는 비착색 슬라이드, 예컨대 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16개 이상의 비착색 슬라이드 및 하나 이상의 H&E 슬라이드를 포함한다. 일부 구현예들에서, 상기 샘플은, 예컨대 본원에 기재된 바와 같이, 예컨대 표준 고정 방법을 사용하여 포르말린 고정되고 파라핀 블록에 매립된 조직을 포함한다.
- [0385] 일부 구현예들에서, 상기 샘플은 말초 전혈 또는 골수 흡인물을 포함한다. 일부 구현예들에서, 상기 샘플, 예컨대 병변 조직은 적어도 20%의 유핵 요소를 포함한다. 일부 구현예들에서, 상기 말초 전혈 샘플 또는 골수 흡인물 샘플은 약 2.5 ml의 부피로 수집된다. 일부 구현예들에서, 상기 혈액 샘플은 수집과 동일한 날에, 예컨대 주위 온도, 예컨대  $43-99^\circ\text{F}$  또는  $6-37^\circ\text{C}$ 에서 배송된다. 일부 구현예들에서, 상기 혈액 샘플은 냉동 또는 냉장되지 않는다.
- [0386] 일부 구현예들에서, 상기 샘플은 분리된, 예컨대 추출된 핵산, 예컨대 DNA 또는 RNA를 포함한다. 일부 구현예들에서, 상기 분리된 핵산은, 예컨대 뉴클레아제가 없는 물에서 DNA 또는 RNA를 포함한다.
- [0387] 일부 구현예들에서, 상기 샘플은 혈액 샘플, 예컨대 말초 전혈 샘플을 포함한다. 일부 구현예들에서, 상기 말초 전혈 샘플은, 예컨대 튜브당 약 8.5 ml의 혈액이 있는, 예컨대 2개의 튜브 내에 수집된다. 일부 구현예들에서, 상기 말초 전혈 샘플은, 예컨대 CLSI H3-A6에 따라 정맥천자에 의해 수집된다. 일부 구현예들에서, 상기 혈액은, 예컨대 약 8-10회 동안, 예컨대 완전한 반전으로 즉시 혼합된다. 일부 구현예들에서, 반전은, 예컨대 손목의 완전한, 예컨대 완전한  $180^\circ$  회전에 의해 실시된다. 일부 구현예들에서, 상기 혈액 샘플은 수집과 동일한 날에, 예컨대 주위 온도, 예컨대  $43-99^\circ\text{F}$  또는  $6-37^\circ\text{C}$ 에서 배송된다. 일부 구현예들에서, 상기 혈액 샘플

은 냉동 또는 냉장되지 않는다. 일부 구현예들에서, 상기 수집된 혈액 샘플은, 예컨대 43-99° F 또는 6-37° C에서 보관된다.

[0388]

대상체

[0389]

일부 구현예들에서, 상기 샘플은 병태 또는 질병, 예컨대 과증식성 질환(예컨대, 본원에 기재된 바와 같음) 또는 비암 징후가 있는 대상체, 예컨대 환자로부터 획득, 예컨대 수집된다. 일부 구현예들에서, 상기 질병은 과증식성 질환이다. 일부 구현예들에서, 상기 과증식성 질환은 암, 예컨대 고형암 또는 혈액암이다. 일부 구현예들에서, 상기 암은 고형 종양이다. 일부 구현예들에서, 상기 암은 혈액암, 예컨대 백혈병 또는 림프종이다.

[0390]

일부 구현예들에서, 상기 대상체는 암을 갖는다. 일부 구현예들에서, 상기 대상체는 대상체는 암에 대한 치료를 받았거나 받고 있다. 일부 구현예들에서, 상기 대상체는, 예컨대 암 요법으로 치료를 받은 후 암 진행 또는 퇴행에 대해 모니터링될 필요가 있다. 일부 구현예들에서, 상기 대상체는 암의 재발에 대해 모니터링될 필요가 있다. 일부 구현예들에서, 상기 대상체는 암을 가질 위험이 있다. 일부 구현예들에서, 상기 대상체는 암 요법으로 치료받은 적이 없다. 일부 구현예들에서, 상기 대상체는 암에 대한 유전적 소인(암 발병에 대한 대상체의 기저 선 위험을 증가시키는 돌연변이가 있는 경우)이 있다. 일부 구현예들에서, 상기 대상체는 암 발병에 대한 대상체의 위험을 증가시키는 환경(예컨대, 방사선 또는 화학물질)에 노출되어 있다. 일부 구현예들에서, 상기 대상체는 암 발병에 대해 모니터링될 필요가 있다.

[0391]

일부 구현예들에서, 상기 환자는 표적 요법, 예컨대 하나 이상의 표적 요법으로 이전에 치료를 받은 적이 있다. 일부 구현예들에서, 표적 요법으로 이전에 치료를 받은 환자의 경우, 표적 요법 후 샘플, 예컨대 표본이 획득되고, 예컨대 수집된다. 일부 구현예들에서, 상기 표적 요법 후 샘플은 표적 요법의 완료 후에 획득된, 예컨대 수집된 샘플이다.

[0392]

일부 구현예들에서, 상기 환자는 암 요법으로 이전에 치료받은 적이 없다. 일부 구현예들에서, 표적 요법으로 이전에 치료를 받은 적이 없는 환자의 경우, 상기 샘플은 절제술, 예컨대 원래 절제술, 또는 재발, 예컨대 요법, 예컨대 비표적 요법 후 질병 재발을 포함한다. 일부 구현예들에서, 상기 샘플은 원발성 종양 또는 전이, 예컨대 전이 생검이거나 이의 일부이다. 일부 구현예들에서, 상기 샘플은 인접 부위, 예컨대 종양 세포가 있는 인접 부위와 비교하여 가장 높은 퍼센트의 종양, 예컨대 종양 세포를 갖는 부위, 예컨대 종양 부위로부터 획득된다. 일부 구현예들에서, 상기 샘플은 인접 부위, 예컨대 종양 세포가 있는 인접 부위와 비교하여 가장 큰 종양 집중을 보이는 부위, 예컨대 종양 부위로부터 획득된다.

[0393]

일부 구현예들에서, 상기 질병은 비소세포 폐암(NSCLC), 흑색종, 유방암, 결장직장암(CRC), 또는 난소암으로부터 선택된다. 일부 구현예들에서, 본원에 기재된 NSCLC는, 예컨대 EGFR 변경(예컨대, 엑손 19 결실 또는 엑손 21 L858R 변경), ALK 재배열 또는 BRAF V600E를 갖는 NSCLC를 포함한다. 일부 구현예들에서, 본원에 기재된 흑색종은 BRAF 변경, 예컨대 V600E 및/또는 V600K를 갖는 흑색종을 포함한다. 일부 구현예들에서, 본원에 기재된 유방암은 ERBB2(HER2) 증폭을 갖는 유방암을 포함한다. 일부 구현예들에서, 본원에 기재된 결장직장암은 야생형 KRAS를 갖는 결장직장암, 예컨대 코돈 12 및/또는 13에서의 돌연변이 부재, 또는 코돈 2, 3 및/또는 4에서의 돌연변이 부재를 포함한다. 일부 구현예들에서, 본원에 기재된 결장직장암은 야생형 NRAS를 갖는 결장직장암, 예컨대 코돈 2, 3 및/또는 4에서의 돌연변이 부재를 포함한다. 일부 구현예들에서, 본원에 기재된 결장직장암은, 예컨대 본원에 기재된 바와 같은 야생형 KRAS 및, 예컨대 본원에 기재된 바와 같은 야생형 NRAS를 갖는 결장직장암을 포함한다. 일부 구현예들에서, 본원에 기재된 난소암은 BRCA1 및/또는 BRCA2 변경을 갖는 난소암을 포함한다.

[0394]

표적 포획 시약

[0395]

본원에 기재된 방법은 표적 포획 시약, 예컨대 서열분석할 표적 핵산 분자의 선택을 위한 용액 혼성화에 사용하기 위한 표적 포획 시약의 적절한 선택에 의해 하나 이상의 대상체로부터의 샘플, 예컨대 본원에 기재된 암으로부터 다수의 유전자 및 유전자 산물의 최적화된 서열분석을 제공한다.

[0396]

2, 3, 4, 5개 이상의 복수의 표적 포획 시약의 임의의 조합, 예를 들어 제1 및 제2 복수의 표적 포획 시약의 조합; 제1 및 제3 복수의 표적 포획 시약; 제1 및 제4 복수의 표적 포획 시약; 제1 및 제5 복수의 표적 포획 시약; 제2 및 제3 복수의 표적 포획 시약; 제2 및 제4 복수의 표적 포획 시약; 제2 및 제5 복수의 표적 포획 시약; 제3 및 제4 복수의 표적 포획 시약; 제3 및 제5 복수의 표적 포획 시약; 제4 및 제5 복수의 표적 포획 시약; 제1, 제2 및 제3 복수의 표적 포획 시약; 제1, 제2 및 제4 복수의 표적 포획 시약; 제1, 제2 및 제5 복수의 표적 포획 시약; 제1, 제2, 제3 및 제4 복수의 표적 포획 시약; 제1, 제2, 제3, 제4 및 제5 복수의 표적 포획 시약.

획 시약 등이 사용될 수 있다.

[0397] 일부 구현예들에서, 상기 방법은:

[0398] (a) 샘플로부터의 복수의 핵산 분자(예컨대, 표적 핵산 분자), 예컨대 샘플, 예컨대 본원에 기재된 샘플로부터의 복수의 중앙 핵산 분자를 포함하는 라이브러리를 수득하는 단계;

[0399] (b) 상기 라이브러리를 2, 3개 이상의 복수의 표적 포획 시약과 접촉시켜 선택된 핵산 분자(예컨대, 라이브러리 캐치)를 제공하는 단계;

[0400] (c) 예컨대, 서열분석을 포함하는 방법에 의해, 예컨대 차세대 서열분석 방법을 이용하여 상기 라이브러리 또는 라이브러리 캐치로부터의 핵산 분자, 예컨대 중앙 핵산 분자로부터 대상체 간격에 대한 리드를 수득하는 단계;

[0401] (d) 정렬 방법, 예컨대 본원에 기재된 정렬 방법에 의해 상기 리드를 정렬하는 단계; 및

[0402] (e) 뉴클레오티드 위치에 대한 상기 리드로부터 뉴클레오티드 값을 할당(예컨대, 돌연변이를, 예컨대 베이지안 방법 또는 본원에 기재된 방법으로 호출)하는 단계를 포함한다.

[0403] 일부 구현예들에서, 본원에서 사용된 바와 같은 서열분석 정도의 수준(예컨대, 서열분석 정도의 X배 수준)은 중복 리드, 예컨대 PCR 중복 리드의 검출 및 제거 후 리드의 수(예컨대, 고유 리드)를 지칭한다. 다른 구현예들에서, 예컨대 복제 수 변경(CNA)의 검출을 지원하기 위해 중복 리드가 평가된다.

[0404] 일 구현예에서, 상기 표적 포획 시약은 하나 이상의 재배열을 함유하는 대상체 간격, 예컨대 유전체 재배열을 함유하는 인트론을 선택한다. 상기 구현예들에서, 상기 표적 포획 시약은 반복 서열이 선택 효율을 증가시키기 위해 마스킹되도록 설계된다. 상기 재배열이 공지된 집합 서열을 갖는 구현예들에서, 선택 효율을 증가시키기 위해 집합 서열에 대하여 상보적인 표적 포획 시약을 설계할 수 있다.

[0405] 일부 구현예들에서, 상기 방법은 2개 이상의 상이한 표적 범주를 포획하도록 설계된 표적 포획 시약의 사용을 포함하고, 각각의 범주는 상이한 설계 전략을 갖는다. 일부 구현예들에서, 본원에 개시된 방법(예컨대, 혼성 포획 방법) 및 조성물은 표적 서열(예컨대, 표적 핵산 분자)의 하위 집합을 포획하고 상기 표적 서열의 균질한 적용 범위를 제공하면서 상기 하위 집합 외부의 적용 범위를 최소화한다. 일 구현예에서, 상기 표적 서열은 유전체 DNA로부터의 전체 엑솜, 또는 이의 선택된 하위 집합을 포함한다. 다른 구현예에서, 상기 표적 서열은 큰 염색체 영역, 예컨대 전체 염색체 팔을 포함한다. 본원에 개시된 방법 및 조성물은 복잡한 표적 핵산 서열(예컨대, 핵산 라이브러리)에 대해 상이한 서열분석 정도 및 적용 범위의 패턴을 달성하기 위한 상이한 표적 포획 시약을 제공한다.

[0406] 한 구현예에서, 상기 방법은 하나 또는 복수의 핵산 라이브러리(예컨대, 라이브러리 캐치)의 선택된 핵산 분자를 제공하는 단계를 포함한다. 예를 들어, 상기 방법은:

[0407] 복수의 핵산 분자, 예컨대 표적 핵산 분자(예컨대, 복수의 중앙 핵산 분자 및/또는 참조 핵산 분자를 포함)를 포함하는 하나 또는 복수의 라이브러리(예컨대, 하나 또는 복수의 핵산 라이브러리)를 제공하는 단계;

[0408] 예컨대, 용액 기반 반응에서 하나 또는 복수의 라이브러리를 2, 3개 이상의 복수의 표적 포획 시약(예컨대, 올리고뉴클레오티드 표적 포획 시약)과 접촉시켜 복수의 표적 포획 시약/핵산 분자 혼성체를 포함하는 혼성화 혼합물을 형성하는 단계;

[0409] 예컨대, 상기 혼성화 혼합물을 상기 혼성화 혼합물로부터의 상기 복수의 표적 포획 시약/핵산 분자 혼성체의 분리를 허용하는 결합 물질과 접촉시킴으로써, 상기 혼성화 혼합물로부터 복수의 표적 포획 시약/핵산 분자 혼성체를 분리시키는 단계;

[0410] 이에 의해 라이브러리 캐치(예컨대, 하나 또는 복수의 라이브러리로부터 핵산 분자의 선택되거나 농축된 하위군)를 제공하는 단계를 포함한다.

[0411] 일 구현예에서, 상기 제1, 제2 또는 제3 복수의 표적 포획 시약의 각각은 고유한 회수 효율을 갖는다. 일부 구현예들에서, 적어도 2개 또는 3개의 복수의 표적 포획 시약은 상이한 회수 효율 값을 갖는다.

[0412] 특정 구현예들에서, 회수 효율에 대한 값은 상이한 표적 포획 시약의 차등 표현, 표적 포획 시약 하위 집합의 차등 중첩, 차등 표적 포획 시약 파라미터, 상이한 표적 포획 시약의 혼합, 및/또는 상이한 유형의 표적 포획 시약의 사용 중 하나 이상에 의해 수정된다. 예를 들어, 회수 효율의 변화(예컨대, 각 표적 포획 시약/표적 범주의 상대적 서열 적용 범위)는, 예컨대 복수의 표적 포획 시약 내에서 및/또는 상이한 복수의 표적 포획 시약

중에서 하기 중 하나 이상을 변경함으로써 조정할 수 있다:

- [0413] (i) 상이한 표적 포획 시약의 차등 표현 - 주어진 표적(예컨대, 표적 핵산 분자)을 포획하기 위한 표적 포획 시약 설계는 상대적 표적 서열분석 정도를 향상/감소시키기 위해 더 많거나 적은 수의 복제에 포함될 수 있다;
- [0414] (ii) 표적 포획 시약 하위 집합의 차등 중첩 - 주어진 표적(예컨대, 표적 핵산 분자)을 포획하기 위한 표적 포획 시약 설계는 상대적 표적 서열분석 정도를 향상/감소시키기 위해 이웃하는 표적 포획 시약 사이의 더 길거나 더 짧은 중첩을 포함할 수 있다;
- [0415] (iii) 차등 표적 포획 시약 매개변수 - 주어진 표적(예컨대, 표적 핵산 분자)을 포획하기 위한 표적 포획 시약 설계는 포획 효율을 감소시키고 상대적 표적 서열분석 정도를 낮추기 위해 서열 변형/짧은 길이가 포함될 수 있다;
- [0416] (iv) 상이한 표적 포획 시약의 혼합 - 상이한 표적 세트를 포획하도록 설계된 표적 포획 시약은 상대적 표적 서열분석 정도를 향상/감소시키기 위해 상이한 물비로 혼합될 수 있다;
- [0417] (v) 상이한 유형의 올리고뉴클레오타이드 표적 포획 시약의 사용 - 특정 구현예들에서, 상기 표적 포획 시약은 다음을 포함할 수 있다:
- [0418] (a) 하나 이상의 화학적으로(예컨대, 비효소적으로) 합성된(예컨대, 개별적으로 합성된) 표적 포획 시약,
- [0419] (b) 배열에서 합성된 하나 이상의 표적 포획 시약,
- [0420] (c) 하나 이상의 효소적으로 제조된, 예컨대 시험관내 전사된 표적 포획 시약;
- [0421] (d) (a), (b) 및/또는 (c)의 조합,
- [0422] (e) 하나 이상의 DNA 올리고뉴클레오타이드(예컨대, 천연 또는 비천연 발생 DNA 올리고뉴클레오타이드),
- [0423] (f) 하나 이상의 RNA 올리고뉴클레오타이드(예컨대, 천연 또는 비천연 발생 RNA 올리고뉴클레오타이드),
- [0424] (g) (e) 및 (f)의 조합, 또는
- [0425] (h) 상기의 임의의 것의 조합.
- [0426] 상이한 올리고뉴클레오타이드 조합은 상이한 비, 예컨대 1:1, 1:2, 1:3, 1:4, 1:5, 1:10, 1:20, 1:50; 1:100, 1:1000 등으로부터 선택된 비로 혼합될 수 있다. 일 구현예에서, 화학적으로 합성된 표적 포획 시약 대 배열 생성 표적 포획 시약의 비는 1:5, 1:10, 또는 1:20으로부터 선택된다. DNA 또는 RNA 올리고뉴클레오타이드는 자연 발생 또는 비자연 발생일 수 있다. 특정 구현예들에서, 상기 표적 포획 시약은, 예컨대 용융 온도를 증가시키기 위해 하나 이상의 비자연 발생 뉴클레오타이드를 포함한다. 예시적인 비자연 발생 올리고뉴클레오타이드는 변형된 DNA 또는 RNA 뉴클레오타이드를 포함한다. 예시적인 변형된 뉴클레오타이드(예컨대, 변형된 RNA 또는 DNA 뉴클레오타이드)는 잠금 핵산(LNA)을 포함하지만 이에 제한되지 않으며, LNA 뉴클레오타이드의 리보스 모이어티는 2' 산소 및 4' 탄소를 연결하는 추가 가교; 펩티드 핵산(PNA), 예컨대 펩티드 결합에 의해 연결된 반복 N-(2-아미노에틸)-글리신 단위로 구성된 PNA; 낮은 GC 영역을 포획하도록 변형된 DNA 또는 RNA 올리고뉴클레오타이드; 두고리 핵산(BNA); 가교된 올리고뉴클레오타이드; 변형된 5-메틸 데옥시시티딘; 및 2,6-디아미노퓨린으로 변형된다. 다른 변형된 DNA 및 RNA 뉴클레오타이드는 당업계에 공지되어 있다.
- [0427] 특정 구현예들에서, 표적 서열(예컨대, 표적 핵산 분자)의 실질적으로 균일하거나 균질한 적용 범위가 수득된다. 예를 들어, 각 표적 포획 시약/표적 범주 내에서, 적용 범위의 균일성은 표적 포획 시약 매개변수를 변경함으로써, 예를 들어 다음 중 하나 이상에 의해 최적화될 수 있다:
- [0428] (i) 증가/감소하는 표적 포획 시약의 표현 또는 중첩은 표적(예컨대, 표적 핵산 분자)의 적용 범위를 향상/감소시키는 데 사용될 수 있으며, 이는 동일한 범주의 다른 표적에 비해 과소/과도하게 적용된다;
- [0429] (ii) 적용 범위가 낮고 표적 서열을 포획하기 어려운 경우(예컨대, 높은 GC 함량 서열), 표적 포획 시약으로 표적화되는 영역을 확장하여, 예컨대 인접 서열(예컨대, GC가 덜 풍부한 인접 서열)을 포함한다;
- [0430] (iii) 표적 포획 시약 서열을 수정하면 표적 포획 시약의 2차 구조를 줄이고 회수 효율을 높일 수 있다;
- [0431] (iv) 표적 포획 시약의 길이를 수정하면 동일한 범주 내에서 상이한 표적 포획 시약의 용융 혼성화 역학을 균등화할 수 있다; 표적 포획 시약의 길이는 직접(다양한 길이의 표적 포획 시약의 생성) 또는 간접적(일관된 길이의 표적 포획 시약의 생성 및 표적 포획 시약 말단을 임의의 서열로 교체)으로 수정할 수 있다;

- [0432] (v) 동일한 표적 영역(즉, 정방향 및 역방향 가닥)에 대해 상이한 방향의 표적 포획 시약을 수정하면 상이한 결합 효율을 가질 수 있다. 각 표적에 대해 최적의 적용 범위를 제공하는 방향 중 하나를 사용하여 표적 포획 시약을 선택할 수 있다;
- [0433] (vi) 각 표적 포획 시약에 존재하는 결합 물질, 예컨대 포획 태그(예컨대, 비오틴)의 양을 수정하면 결합 효율에 영향을 미칠 수 있다. 특정 표적을 표적으로 하는 표적 포획 시약의 태그 수준을 증가/감소시키는 것은 상대적 표적 범위를 향상/감소시키기 위해 사용될 수 있다;
- [0434] (vii) 상이한 표적 포획 시약에 사용되는 뉴클레오타이드 유형을 수정하여 표적에 대한 결합 친화도에 영향을 미치고 상대적 표적 적용 범위를 향상/감소시킬 수 있습니다; 또는
- [0435] (viii) 변형된 올리고뉴클레오타이드 표적 포획 시약을 사용하여, 예컨대 보다 안정적인 염기 쌍을 갖는 것은 높은 GC 함량에 비해 낮거나 정상적인 GC 함량 영역 간의 용융 혼성화 동역학을 균등화하는 데 사용할 수 있다.
- [0436] 한 구현예에서, 상기 방법은 종양 핵산 분자, 예컨대 종양 세포로부터 대상체 간격을 포함하는 핵산 분자를 선택하는 표적 포획 시약을 포함하는 복수의 표적 포획 시약의 사용을 포함한다. 상기 종양 핵산 분자는 종양 세포에 존재하는 임의의 뉴클레오타이드 서열, 예컨대 종양 또는 암세포에 존재하는 본원에 기재된 바와 같은 돌연변이, 야생형, 참조 또는 인트론 뉴클레오타이드 서열일 수 있다. 일 구현예에서, 상기 종양 핵산 분자는 낮은 빈도로 나타나는 변형(예컨대, 하나 이상의 돌연변이)을 포함하며, 예컨대 상기 샘플로부터의 세포의 약 5% 이하가 이의 유전체의 변형을 수반한다. 다른 구현예들에서, 상기 종양 핵산 분자는 상기 샘플로부터 세포의 약 10%의 빈도로 나타나는 변형(예컨대, 하나 이상의 돌연변이)을 포함한다. 다른 구현예들에서, 상기 종양 핵산 분자는 인트론 서열, 예컨대 본원에 기재된 바와 같은 인트론 서열, 종양 세포에 존재하는 참조 서열로부터의 하위 유전체 간격을 포함한다.
- [0437] 다른 구현예들에서, 상기 방법은 (예컨대, PCR에 의해)라이브러리 캐치를 증폭시키는 단계를 포함한다. 다른 구현예들에서, 상기 라이브러리 캐치는 증폭되지 않는다.
- [0438] 다른 양태에서, 본 발명은 본원에 기재된 표적 포획 시약 및 본원에 기재된 개별 복수의 표적 포획 시약의 조합을 특징으로 한다. 상기 표적 포획 시약은 지시, 표준, 완충액 또는 효소 또는 기타 시약을 선택적으로 포함할 수 있는 키트의 일부일 수 있다.
- [0439] 표적 포획 시약의 설계 및 구성
- [0440] 일부 구현예들에서, 표적 포획 시약은 표적 분자에 결합하여 표적 분자의 포획을 허용할 수 있는 분자이다. 예를 들어, 표적 포획 시약은 미끼, 예컨대 핵산 분자, 예컨대 DNA 또는 RNA 분자일 수 있으며, 혼성화되어(예컨대, 상보적이 되어) 표적 핵산의 포획을 가능하게 할 수 있다. 일부 구현예들에서, 상기 표적 포획 시약, 예컨대 미끼는 포획 올리고뉴클레오타이드이다. 특정 구현예들에서, 상기 표적 핵산은 유전체 DNA 분자이다. 다른 구현예들에서, 상기 표적 핵산은 RNA 분자 또는 RNA 분자로부터 유래된 cDNA 분자이다. 일 구현예에서, 상기 표적 포획 시약은 DNA 분자이다. 일 구현예에서, 상기 표적 포획 시약은 RNA 분자이다. 일 구현예에서, 상기 표적 포획 시약은 용액상 혼성화에 적합하다. 일 구현예에서, 상기 표적 포획 시약은 고체상 혼성화에 적합하다. 일 구현예에서, 상기 표적 포획 시약은 용액상 및 고체상 혼성화 둘 모두에 적합하다.
- [0441] 일반적으로, DNA 분자는 표적 포획 시약 서열로서 사용되지만 RNA 분자도 사용할 수 있다. 일부 구현예들에서, DNA 분자 표적 포획 시약은 단일 가닥 DNA(ssDNA) 또는 이중 가닥 DNA(dsDNA)일 수 있다.
- [0442] 일부 구현예들에서, RNA-DNA 이중체는 DNA-DNA 이중체보다 더 안정적이고, 따라서 핵산의 잠재적으로 더 나은 포획을 제공한다. RNA 표적 포획 시약은 신규 화학 합성 및 DNA 의존성 RNA 중합효소를 사용한 DNA 분자의 전사를 포함하지만 이에 제한되지 않는 당업계에 공지된 방법을 사용하여 본원의 다른 곳에서 기재된 바와 같이 제조될 수 있다. 일 구현예에서, 상기 표적 포획 시약 서열은, 예컨대 인간 DNA 또는 통합된 인간 DNA 샘플을 주형으로 사용하는 PCR과 같은 공지된 핵산 증폭 방법을 사용하여 생성된다. 그런 다음, 올리고뉴클레오타이드는 RNA 표적 포획 시약으로 전환될 수 있다. 일 구현예에서, 시험관내 전사는, 예를 들어 올리고뉴클레오타이드의 한 쪽 말단에 RNA 중합효소 프로모터 서열을 부가하는 것을 기초로 사용된다. 일 구현예에서, 상기 RNA 중합효소 프로모터 서열은, 예컨대 PCR 또는 다른 핵산 증폭 방법을 사용하여, 예컨대 각각의 표적 특이적 프라이머 쌍의 하나의 프라이머를 RNA 프로모터 서열로 테일링함으로써 상기 표적 포획 시약 서열을 증폭 또는 재증폭함으로써 상기 표적 포획 시약의 말단에 첨가된다. 일 구현예에서, 상기 RNA 중합효소는 T7 중합효소, SP6 중합효소, 또는 T3 중합효소이다. 일 구현예에서, RNA 표적 포획 시약은 태그, 예컨대 친화성 태그로 표지된다. 일 구현예에

서, RNA 표적 포획 시약은, 예컨대 비오틴화된 UTP를 사용하여 시험관내 전사에 의해 제조된다. 다른 구현예에서, RNA 표적 포획 시약은 비오틴 없이 생성된 다음, 비오틴이 소랄렌 가교와 같은 당업계에 일반적으로 공지된 방법을 사용하여 RNA 분자에 가교된다. 일 구현예에서, 상기 RNA 표적 포획 시약은, 예컨대 RNase 분해에 저항하는 RNA 분자를 생성하기 위해 전사 동안 변형된 뉴클레오타이드를 사용함으로써 제조될 수 있는 RNase 저항성 RNA 분자이다. 일 구현예에서, 상기 RNA 표적 포획 시약은 이중 가닥 DNA 표적의 한 가닥에만 해당한다. 일반적으로, 상기 RNA 표적 포획 시약은 자가 보완적이지 않으며 혼성화 드라이버로서 더 효과적이다.

[0443] 상기 표적 포획 시약은 표적 포획 시약이 참조 서열의 표적을 선택하는 데 최적이 되도록 참조 서열로부터 설계될 수 있다. 일부 구현예들에서, 표적 포획 시약 서열은 혼합 염기(예컨대, 축퇴)를 사용하여 설계된다. 예를 들어, 혼합 염기(들)는 공통 SNP 또는 돌연변이의 위치(들)에서 표적 포획 시약 서열에 포함되고, 상기 표적 포획 시약 서열을 최적화하여 두 대립 유전자(예컨대, SNP 및 비 SNP; 돌연변이 및 비돌연변이)를 포착할 수 있다. 일부 구현예들에서, 모든 공지된 서열 변이(또는 이의 하위 집합)는 혼합된 축퇴 올리고뉴클레오타이드를 사용하기 보다는 다중 올리고뉴클레오타이드 표적 포획 시약으로 표적화될 수 있다.

[0444] 특정 구현예들에서, 상기 표적 포획 시약은 길이가 약 100개 뉴클레오타이드 내지 300개 뉴클레오타이드인 올리고뉴클레오타이드(또는 복수의 올리고뉴클레오타이드)를 포함한다. 전형적으로, 상기 표적 포획 시약은 길이가 약 130개 뉴클레오타이드 내지 230개 뉴클레오타이드, 또는 약 150개 내지 200개 뉴클레오타이드인 올리고뉴클레오타이드(또는 복수의 올리고뉴클레오타이드)를 포함한다. 다른 구현예들에서, 상기 표적 포획 시약은 길이가 약 300개 뉴클레오타이드 내지 1000개 뉴클레오타이드인 올리고뉴클레오타이드(또는 복수의 올리고뉴클레오타이드)를 포함한다.

[0445] 일부 구현예들에서, 상기 올리고뉴클레오타이드 내의 표적 핵산 분자 특이적 서열은 약 40 내지 1000개의 뉴클레오타이드, 약 70 내지 300개의 뉴클레오타이드, 약 100 내지 200개의 뉴클레오타이드 길이, 전형적으로 약 120 내지 170개의 뉴클레오타이드 길이를 갖는다.

[0446] 일부 구현예들에서, 상기 표적 포획 시약은 결합 물질을 포함한다. 상기 결합 물질은 친화성 태그일 수 있다. 일부 구현예들에서, 상기 친화성 태그는 비오틴 분자 또는 합텐이다. 특정 구현예에서, 상기 결합 물질은 아비딘 분자, 또는 합텐 또는 이의 항원 결합 단편에 결합하는 항체와 같은 짝에 결합함으로써 혼성화 혼합물로부터 표적 포획 시약/핵산 분자 혼성체의 분리를 허용한다.

[0447] 다른 구현예들에서, 상기 표적 포획 시약 내의 올리고뉴클레오타이드는 동일한 표적 핵산 분자 서열에 대한 정방향 및 역방향 보체 서열을 함유하고, 이에 의해 역상보화된 핵산 분자 특이적 서열을 갖는 올리고뉴클레오타이드는 또한 역상보 보편적 꼬리를 보유한다. 이는 동일한 가닥, 즉 서로 상보적이지 않은 RNA 전사체로 이어질 수 있다.

[0448] 다른 구현예들에서, 상기 표적 포획 시약은 하나 이상의 위치에 축퇴 또는 혼합 염기를 함유하는 올리고뉴클레오타이드를 포함한다. 또 다른 구현예들에서, 상기 표적 포획 시약은 단일 종의 집단 또는 유기체의 군집에 존재하는 다중 또는 실질적으로 모든 공지된 서열 변이체를 포함한다. 일 구현예에서, 상기 표적 포획 시약은 인간 집단에 존재하는 다중 또는 실질적으로 모든 공지된 서열 변이체를 포함한다.

[0449] 다른 구현예들에서, 상기 표적 포획 시약은 cDNA 서열을 포함하거나 cDNA 서열로부터 유래된다. 다른 구현예들에서, 상기 표적 포획 시약은 유전체 DNA, cDNA 또는 클로닝된 DNA로부터 증폭되는 증폭 생성물(예컨대, PCR 생성물)을 포함한다.

[0450] 다른 구현예들에서, 상기 표적 포획 시약은 RNA 분자를 포함한다. 일부 구현예들에서, 상기 세트는 RNase에 대해 더 안정하고 내성인 것을 포함하지만 이에 제한되지는 않는 화학적으로, 효소적으로 변형된, 또는 시험관내 전사된 RNA 분자를 포함한다.

[0451] 또 다른 구현예들에서, 상기 표적 포획 시약은 US 2010/0029498 및 Gnirke, A. et al. (2009) *Nat Biotechnol.* 27(2):182-189에 기재된 방법에 의해 생성되고, 본원에 참조로서 인용된다. 예를 들어, 바이오틴화된 RNA 표적 포획 시약은 원래 미세배열 상에서 합성된 긴 합성 올리고뉴클레오타이드의 풀(pool)을 수득하고 올리고뉴클레오타이드를 증폭하여 상기 표적 포획 시약 서열을 생성함으로써 생성될 수 있다. 일부 구현예들에서, 상기 표적 포획 시약은 상기 표적 포획 시약 서열의 한쪽 말단에 RNA 증합효소 프로모터 서열을 첨가하고, RNA 증합효소를 사용하여 RNA 서열을 합성함으로써 생성된다. 일 구현예에서, 합성 올리고데옥시뉴클레오타이드의 라이브러리는 애질런트 테크놀로지스사(Agilent Technologies, Inc.)와 같은 상업적 공급업체로부터 수득할 수 있고, 공지된 핵산 증폭 방법을 사용하여 증폭될 수 있습니다.

[0452] 따라서, 전술한 표적 포획 시약의 제조 방법이 제공된다. 상기 방법은, 예를 들어 하나 이상의 표적 포획 시약,

예컨대 표적 특이적 미끼 올리고뉴클레오타이드 서열(예컨대, 본원에 기재된 바와 같은 하나 이상의 돌연변이 포함, 참조 또는 대조군 올리고뉴클레오타이드 서열)을 선택하는 단계; 표적 포획 시약의 풀, 예컨대 표적 특이적 미끼 올리고뉴클레오타이드 서열을 수득하는(예컨대, 표적 특이적 미끼 올리고뉴클레오타이드 서열의 풀을, 예컨대 미세배열 합성에 의해 합성하는) 단계; 및 선택적으로 표적 포획 시약, 예컨대 표적 특이적 미끼 올리고뉴클레오타이드 서열을 증폭하는 단계를 포함한다.

[0453] 다른 구현예들에서, 상기 방법은 하나 이상의 비오틴화된 프라이머를 사용하여 올리고뉴클레오타이드를 증폭(예컨대, PCR에 의해)하는 단계를 추가로 포함한다. 일부 구현예들에서, 상기 올리고뉴클레오타이드는 상기 미세배열에 부착된 각 올리고뉴클레오타이드의 말단에 보편적인 서열을 포함한다. 상기 방법은 상기 올리고뉴클레오타이드로부터 보편적인 서열을 제거하는 단계를 추가로 포함할 수 있다. 상기 방법은 또한 상기 올리고뉴클레오타이드의 상보적 가닥을 제거하고, 상기 올리고뉴클레오타이드를 어닐링하고, 상기 올리고뉴클레오타이드를 연장하는 단계를 포함할 수 있다. 상기 구현예들 중 일부에서, 상기 올리고뉴클레오타이드를 증폭하는 방법(예컨대, PCR에 의해)은 하나 이상의 비오틴화된 프라이머를 사용한다. 일부 구현예들에서, 상기 방법은 증폭된 올리고뉴클레오타이드를 크기 선택하는 단계를 추가로 포함한다.

[0454] 일 구현예에서, RNA 표적 포획 시약이 제조된다. 상기 방법은 본원에 기재된 방법에 따라 표적 포획 시약 서열의 세트를 생성하는 단계, 상기 표적 포획 시약 서열의 한쪽 말단에 RNA 중합효소 프로모터 서열을 첨가하는 단계, 및 RNA 중합효소를 사용하여 RNA 서열을 합성하는 단계를 포함한다. 상기 RNA 중합효소는 T7 RNA 중합효소, SP6 RNA 중합효소, 또는 T3 RNA 중합효소로부터 선택될 수 있다. 다른 구현예들에서, 상기 RNA 중합효소 프로모터 서열은 상기 표적 포획 시약 서열을 증폭함으로써(예컨대, PCR에 의해) 상기 표적 포획 시약 서열의 말단에 첨가된다. 상기 표적 포획 시약 서열이 유전체 DNA 또는 cDNA로부터 특이적 프라이머 쌍을 사용하여 PCR에 의해 증폭되는 구현예들에서, RNA 프로모터 서열을 각 쌍에 있는 2개의 특이적 프라이머 중 하나의 5' 말단에 추가하면 표준 방법을 사용하여 RNA 표적 포획 시약으로 전사될 수 있는 PCR 생성물이 생성될 것이다.

[0455] 다른 구현예들에서, 표적 포획 시약은 인간 DNA 또는 풀링된 인간 DNA 샘플을 주형으로 사용하여 생성될 수 있다. 상기 구현예들에서, 상기 올리고뉴클레오타이드는 중합효소 연쇄 반응(PCR)에 의해 증폭된다. 다른 구현예들에서, 상기 증폭된 올리고뉴클레오타이드는 회전 환 증폭 또는 초분기 회전 환 증폭에 의해 재증폭된다. 상기 동일한 방법은 인간 DNA 또는 풀링된 인간 DNA 샘플을 주형으로 사용하여 표적 포획 시약 서열을 생성하기 위해 사용될 수도 있다. 상기 동일한 방법은 또한 제한효소 분해, 간헐영역 겔 전기영동, 흐름 분류, CsCl 밀도 구배 원심분리, 선택적 동역학 재결합, 염색체 제제의 미세해부, 및 당업자에게 공지된 기타 분별 방법을 포함하지만 이에 제한되지는 않는 다른 방법으로 수득한 유전체의 소분획을 사용하여 표적 포획 시약 서열을 생성하기 위해 사용될 수 있다.

[0456] 특정 구현예들에서, 상기 복수의 표적 포획 시약 중 표적 포획 시약(예컨대, 미끼)의 수는 1,000개 미만이다. 다른 구현예들에서, 상기 복수의 표적 포획 시약 중 표적 포획 시약(예컨대, 미끼)의 수는 1,000 초과, 5,000 초과, 10,000 초과, 20,000 초과, 50,000 초과, 100,000 초과, 또는 500,000 초과이다.

[0457] 상기 표적 포획 시약 서열의 길이는 약 70개 뉴클레오타이드 내지 1000개 뉴클레오타이드일 수 있다. 일 구현예에서, 상기 표적 포획 시약의 길이는 약 100 내지 300개 뉴클레오타이드, 110 내지 200개 뉴클레오타이드, 또는 120 내지 170개 뉴클레오타이드이다. 상기 언급된 것들 외에도, 약 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 300, 400, 500, 600, 700, 800, 및 900개의 뉴클레오타이드 길이를 갖는 중간체 올리고뉴클레오타이드의 길이가 본원에 기재된 방법에 사용될 수 있다. 일부 구현예들에서, 약 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 또는 230개의 염기를 갖는 올리고뉴클레오타이드가 사용될 수 있다.

[0458] 각각의 표적 포획 시약 서열은 표적 특이적(예컨대, 핵산 분자 특이적) 표적 포획 시약 서열 및 한쪽 또는 양쪽 말단에 보편적인 꼬리를 포함할 수 있다. 본원에서 사용한 용어 "표적 포획 시약 서열"은 표적 특이적 표적 포획 시약 서열 또는 표적 특이적 "표적 포획 시약 서열" 및 올리고뉴클레오타이드의 다른 뉴클레오타이드를 포함하는 전체 올리고뉴클레오타이드를 지칭할 수 있다. 상기 표적 포획 시약의 표적 특이적 서열은 길이가 약 40개 뉴클레오타이드 내지 1000개 뉴클레오타이드이다. 일 구현예에서, 상기 표적 특이적 서열은 길이가 약 70개 뉴클레오타이드 내지 300개 뉴클레오타이드이다. 다른 구현예에서, 상기 표적 특이적 서열은 길이가 약 100개 뉴클레오타이드 내지 200개 뉴클레오타이드이다. 또 다른 구현예에서, 상기 표적 특이적 서열은 길이가 약 120개 뉴클레오타이드 내지 170개 뉴클레오타이드, 일반적으로 120개 뉴클레오타이드이다. 상기 언급된 것 이외에 중간 길이, 예컨대 약 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250,

300, 400, 500, 600, 700, 800, 및 900개의 뉴클레오타이드 길이를 갖는 표적 특이적 서열 뿐만 아니라 상기 언급된 길이 사이의 길이를 갖는 표적 특이적 서열이 본원에 기재된 방법에서 또한 사용될 수 있다.

[0459] 일 구현예에서, 상기 표적 포획 시약은 약 50 내지 200개의 뉴클레오타이드 길이(예컨대, 약 50, 60, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 190, 또는 200개의 뉴클레오타이드 길이)를 갖는 올리고머(예컨대, RNA 올리고머, DNA 올리고머, 또는 이의 조합으로 구성된)이다. 일 구현예에서, 각각의 표적 포획 시약 올리고머는 표적 특이적 표적 포획 시약 서열인 약 120 내지 170개, 또는 일반적으로 약 120개의 뉴클레오타이드를 포함한다. 상기 표적 포획 시약은 한쪽 말단 또는 양쪽 말단에 추가적인 비표적 특이적 뉴클레오타이드 서열을 포함할 수 있다. 추가 뉴클레오타이드 서열은, 예컨대 PCR 증폭을 위해 또는 표적 포획 시약 식별자로 사용될 수 있다. 특정 구현예들에서, 상기 표적 포획 시약은 본원에 기재된 바와 같은 결합 물질(예컨대, 비오틴 분자와 같은 친화성 태그)을 추가로 포함한다. 상기 결합 물질, 예컨대 비오틴 분자는, 예컨대 상기 표적 포획 시약의 5'-말단, 3'-말단에서, 또는 내부적으로(예컨대, 비오틴화된 뉴클레오타이드를 통합함으로써) 표적 포획 시약에 부착될 수 있다. 일 구현예에서, 상기 비오틴 분자는 상기 표적 포획 시약의 5'-말단에 부착된다.

[0460] 예시적인 일 구현예에서, 상기 표적 포획 시약은 약 150개 뉴클레오타이드의 길이를 갖는 올리고뉴클레오타이드이고, 그 중 120개의 뉴클레오타이드는 표적 특이적 "표적 포획 시약 서열"이다. 다른 30개의 뉴클레오타이드(예컨대, 각 말단에 15개 뉴클레오타이드)는 PCR 증폭에 사용되는 보편적인 임의의 꼬리이다. 상기 꼬리는 사용자가 선택한 임의의 서열일 수 있다. 예를 들어, 합성 올리고뉴클레오타이드의 풀은 표적 특이적 표적 포획 시약 서열을 나타내는 N<sub>120</sub>을 갖는 5' -ATCGCACCAGCGTGTN<sub>120</sub>CACTGCGGCTCCTCA-3' (서열 번호 1)의 서열의 올리고뉴클레오타이드를 포함할 수 있다.

[0461] 본원에 기재된 표적 포획 시약 서열은 엑손 및 짧은 표적 서열의 선택에 사용될 수 있다. 일 구현예에서, 상기 표적 포획 시약은 길이가 약 100개 뉴클레오타이드 내지 300개 뉴클레오타이드이다. 다른 구현예에서, 상기 표적 포획 시약은 길이가 약 130개 뉴클레오타이드 내지 230개 뉴클레오타이드이다. 또 다른 구현예에서, 상기 표적 포획 시약은 길이가 약 150개 뉴클레오타이드 내지 200개 뉴클레오타이드이다. 예컨대, 엑손 및 짧은 표적 서열의 선택을 위한 표적 포획 시약의 표적 특이적 서열은 약 40개 뉴클레오타이드 내지 1000개 뉴클레오타이드의 길이를 갖는다. 일 구현예에서, 상기 표적 특이적 서열은 길이가 약 70개 뉴클레오타이드 내지 300개 뉴클레오타이드이다. 다른 구현예에서, 상기 표적 특이적 서열은 길이가 약 100개 뉴클레오타이드 내지 200개 뉴클레오타이드이다. 또 다른 구현예에서, 상기 표적 특이적 서열은 길이가 약 120개 뉴클레오타이드 내지 170개 뉴클레오타이드이다.

[0462] 일부 구현예들에서, 긴 올리고뉴클레오타이드는 상기 표적 서열을 포획하는 데 필요한 올리고뉴클레오타이드의 수를 최소화할 수 있다. 예를 들어, 엑손당 하나의 올리고뉴클레오타이드가 사용될 수 있다. 인간 유전체에서 단백질 부호화 엑손의 평균 및 중앙값 길이는 각각 약 164개 및 120개의 염기쌍인 것으로 당업계에 공지되어 있다. 표적 포획 시약 서열이 길수록 더 특이적일 수 있고 짧은 것보다 더 잘 포착할 수 있다. 그 결과, 올리고뉴클레오타이드 표적 포획 시약 서열당 성공률은 짧은 올리고뉴클레오타이드보다 더 높다. 일 구현예에서, 최소 표적 포획 시약 포함 서열은, 예컨대 엑손 크기의 표적을 포획하기 위한 하나의 표적 포획 시약의 크기(예컨대, 120-170개 염기)이다. 상기 표적 포획 시약 서열의 길이를 결정할 때, 불필요하게 긴 표적 포획 시약이 표적에 직접 인접한 원치 않는 DNA를 더 많이 포획한다는 점도 고려할 수 있다. 더 긴 올리고뉴클레오타이드 표적 포획 시약은 또한 짧은 것보다 DNA 샘플의 표적 영역에서 다형성에 더 많은 내성을 가질 수 있다. 일반적으로, 표적 포획 시약 서열은 참조 유전체 서열에서 유래한다. 실제 DNA 샘플의 표적 서열이 참조 서열과 다른 경우, 예를 들어 단일 염기 다형성(SNP)을 함유하는 경우, 표적 포획 시약에 덜 효율적으로 혼성화될 수 있으므로 상기 표적 포획 시약 서열에 혼성화된 서열에서 과소 대표되거나 완전히 부재할 수 있다. 예컨대, 120 내지 170개의 염기에서의 단일 불일치가 다중 증폭 및 미세배열 포획에서 각각 전형적인 표적 포획 시약 또는 프라이머 길이인 20 내지 70개 염기에서의 단일 불일치보다 혼성 안정성에 영향을 덜 미칠 수 있기 때문에 SNP로 인한 대립유전자 탈락은 합성 표적 포획 시약 분자가 길수록 가능성이 적을 수 있다.

[0463] 유전체 영역과 같은 포획 표적 포획 시약의 길이에 비해 긴 표적을 선택하기 위해, 표적 포획 시약 서열 길이는, 인접 서열의 표적화를 최소화하기 위한 단일 목적으로 표적 포획 시약 서열의 최대 크기를 제한할 필요가 없는 경우를 제외하고는, 일반적으로 상기에서 언급한 짧은 표적에 대한 표적 포획 시약과 동일한 크기 범위에 있다. 대안적으로, 올리고뉴클레오타이드는 훨씬 더 넓은 창(일반적으로 600개의 염기)에 걸쳐 타일링될 수 있다. 상기 방법은 일반적인 엑손보다 훨씬 큰(예컨대, 약 500개의 염기) DNA 단편을 포획하는데 사용할 수 있다. 그 결과, 훨씬 더 많은 원치 않는 측면 비표적 서열이 선택된다.

[0464] 표적 포획 시약의 합성

- [0465] 상기 표적 포획 시약은, 예를 들어 임의의 유형의 올리고뉴클레오타이드, 예컨대 DNA 또는 RNA일 수 있다. 상기 DNA 또는 RNA 표적 포획 시약("올리고 표적 포획 시약")은 개별적으로 합성될 수 있거나, 또는 DNA 또는 RNA 표적 포획 시약(예컨대, "배열 미끼")으로서 배열에서 합성될 수 있다. 배열 형식으로 제공되거나 분리된 올리고로 제공되든지 상관없이 올리고 표적 포획 시약은 일반적으로 단일 가닥이다. 상기 표적 포획 시약은 본원에 기재된 바와 같은 결합 물질(예컨대, 비오틴 분자와 같은 친화성 태그)을 추가로 포함한다. 상기 결합 물질, 예컨대 비오틴 분자는, 예컨대 상기 표적 포획 시약의 5'-말단에서 전형적인 표적 포획 시약의 5'-말단 또는 3'-말단에서 상기 표적 포획 시약에 부착될 수 있다. 표적 포획 시약은, 예컨대 국제특허출원공보 제WO 2012/092426호 또는 국제특허출원공보 제WO 2015/021080호에 기재된 바와 같이 당업계에 기재된 방법에 의해 합성될 수 있고, 그 전체 내용은 본원에 참조로서 인용된다.
- [0466] **혼성화 조건**
- [0467] 본 발명에서 특징으로 하는 방법은 라이브러리(예컨대, 핵산 라이브러리)를 복수의 표적 포획 시약과 접촉시켜 선택된 라이브러리 캐치를 제공하는 단계를 포함한다. 상기 접촉 단계는 용액 혼성화에서 실시될 수 있다. 특정 구현예들에서, 상기 방법은 1회 이상의 추가적인 용액 혼성화에 의한 혼성화 단계를 반복하는 것을 포함한다. 일부 구현예들에서, 상기 방법은 상기 라이브러리 캐치에 동일하거나 상이한 표적 포획 시약의 수집물과의 1회 이상의 용액 혼성화를 적용하는 것을 추가로 포함한다. 본원의 방법에 사용하기 위해 조정될 수 있는 혼성화 방법은, 예컨대 국제특허출원공보 제WO 2012/092426에 기재된 바와 같이 당업계에 기재되어 있다.
- [0468] 본 발명의 추가적인 구현예 또는 특징은 하기와 같다:
- [0469] 특정 구현예들에서, 상기 방법은 상기 샘플에서 양성 표현형(예컨대, 본원에 기재된 유전자 또는 유전자 산물의 적어도 10, 20, 30, 50개 이상의 변경)과, 예컨대 양성 또는 음성으로 연관된 변경의 존재 또는 부재를 결정하는 단계를 것을 포함한다. 다른 구현예들에서, 상기 방법은 유전체 시그니처, 예컨대 연속/복합 바이오마커(예컨대, 중앙 돌연변이 부담의 수준)를 결정하는 단계를 포함한다. 다른 구현예들에서, 상기 방법은 하나 이상의 유전체 시그니처, 예컨대 연속/복합 바이오마커, 예컨대 미소부수체 불안정성의 수준, 또는 이형접합체(LOH)의 존재 또는 부재를 결정하는 단계를 포함한다. 상기 방법은 라이브러리 캐치를 수득하기 위해 본원에 기재된 방법 및 표적 포획 시약 중 임의의 것에 따른 용액 기반 반응에서 샘플 내의 핵산을 접촉시키는 단계; 및 상기 라이브러리 캐치의 전체 또는 하위집합을 서열분석(예컨대, 차세대 서열분석에 의해)하여 본원에 기재된 유전자 또는 유전자 산물의 변경의 존재 또는 부재를 결정하는 단계를 포함한다.
- [0470] 특정 구현예들에서, 상기 표적 포획 시약은 길이가 약 100개 뉴클레오타이드 내지 300개 뉴클레오타이드인 올리고뉴클레오타이드(또는 복수의 올리고뉴클레오타이드)를 포함한다. 전형적으로, 상기 표적 포획 시약은 길이가 약 130개 뉴클레오타이드 내지 230개 뉴클레오타이드, 또는 약 150개 내지 200개 뉴클레오타이드인 올리고뉴클레오타이드(또는 복수의 올리고뉴클레오타이드)를 포함한다. 다른 구현예들에서, 상기 표적 포획 시약은 길이가 약 300개 뉴클레오타이드 내지 1000개 뉴클레오타이드인 올리고뉴클레오타이드(또는 복수의 올리고뉴클레오타이드)를 포함한다.
- [0471] 다른 구현예들에서, 상기 표적 포획 시약은 cDNA 서열을 포함하거나 cDNA 서열로부터 유래된다. 일 구현예에서, 상기 cDNA는 RNA 서열, 예컨대 종양 또는 암 세포 유래 RNA, 예컨대 종양-FFPE 샘플, 혈액 샘플, 또는 골수 흡인물 샘플로부터 수득한 RNA로부터 제조된다. 다른 구현예들에서, 상기 표적 포획 시약은 유전체 DNA, cDNA 또는 클로닝된 DNA로부터 증폭되는 증폭 생성물(예컨대, PCR 생성물)을 포함한다.
- [0472] 특정 구현예들에서, 라이브러리(예컨대, 핵산 라이브러리)는 핵산 분자의 집합체를 포함한다. 본원에 기재된 바와 같이, 상기 라이브러리의 핵산 분자는 표적 핵산 분자(예컨대, 종양 핵산 분자, 참조 핵산 분자 및/또는 대조 핵산 분자; 또한 본원에서 각각 제1, 제2 및/또는 제3 핵산 분자라고도 칭함)를 포함할 수 있다. 상기 라이브러리의 핵산 분자들은 단일 개체로부터 유래할 수 있다. 일부 구현예들에서, 라이브러리는 하나 초과(2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30개 이상의 대상체)로부터의 핵산 분자들을 포함할 수 있고, 예컨대 상이한 대상체들로부터의 2개 이상의 라이브러리는 1개 초과(2개 이상의 대상체)로부터의 핵산 분자들을 갖는 라이브러리를 형성하기 위해 조합될 수 있다. 일 구현예에서, 상기 대상체는 암 또는 종양을 갖거나 가질 위험이 있는 인간이다.
- [0473] 일부 구현예들에서, 상기 방법은 하나 또는 복수의 라이브러리(예컨대, 하나 또는 복수의 핵산 라이브러리)를 복수의 표적 포획 시약과 접촉시켜 핵산의 선택된 하위군, 예컨대 라이브러리 캐치를 제공하는 단계를 포함한다. 일 구현예에서, 상기 접촉 단계는 고체 지지체, 예컨대 배열에서 실시된다. 혼성화에 적합한 고체 지지체는, 예컨대 Albert, T.J. *et al.* (2007) *Nat. Methods* 4(11):903-5; Hodges, E. *et al.* (2007) *Nat.*

*Genet.* 39(12):1522-7; and Okou, D.T. *et al.* (2007) *Nat. Methods* 4(11):907-9에 기재되며, 그 내용은 본원에 참조로서 원용된다. 다른 구현예들에서, 상기 접촉 단계는 용액 혼성화에서 실시된다. 특정 구현예들에서, 상기 방법은 1회 이상의 추가적인 혼성화에 의한 혼성화 단계를 반복하는 것을 포함한다. 일부 구현예들에서, 상기 방법은 상기 라이브러리 캐치에 동일하거나 상이한 표적 포획 시약의 수집물과의 1회 이상의 혼성화를 적용하는 것을 추가로 포함한다.

- [0474] 또 다른 구현예들에서, 상기 방법은 상기 라이브러리 캐치를 유전형 분석에 적용함으로써 선택된 핵산의 유전자형을 확인하는 단계를 추가로 포함한다.
- [0475] 특정 구현예들에서, 상기 방법은 하기로서:
- [0476] 샘플의 지문을 채취하는 단계;
- [0477] 상기 샘플에서 유전자 또는 유전자 산물(예컨대, 본 본원에 기재된 바와 같은 유전자 또는 유전자 산물)의 존재량을 정량화(예컨대, 상기 샘플에서 전사체의 상대적 존재량을 정량화)하는 단계;
- [0478] 상기 샘플을 특정 대상체(예컨대, 정상 대조군 또는 암 환자)에 속하는 것으로 확인하는 단계;
- [0479] 상기 샘플에서 유전적 특성(예컨대, 한 명 이상의 대상체의 유전적 구성(예컨대, 민족성, 인종, 가족적 특성))을 확인하는 단계;
- [0480] 핵산 샘플에서 배수성을 결정하는 단계; 상기 샘플에서 이형접합성의 손실을 결정하는 단계;
- [0481] 상기 샘플에서 본원에 기재된 변경, 예컨대 뉴클레오타이드 치환, 복제 수 변경, 삽입-결실 또는 재배열의 존재 또는 부재를 결정하는 단계;
- [0482] 상기 샘플에서 종양 돌연변이 부담 및/또는 미소부수체 불안정성(및/또는 기타 복합 바이오마커)의 수준을 결정하는 단계; 또는
- [0483] 상기 샘플에서 종양/정상 세포 혼합물의 수준을 결정하는 단계 중 하나 이상을 추가로 포함한다.
- [0484] 상이한 올리고뉴클레오타이드 조합은 상이한 비, 예컨대 1:1, 1:2, 1:3, 1:4, 1:5, 1:10, 1:20, 1:50; 1:100, 1:1000 등으로부터 선택된 비로 혼합될 수 있다. 일 구현예에서, 화학적으로 합성된 표적 포획 시약(예컨대, 미끼) 대 배열 생성 표적 포획 시약(예컨대, 미끼)의 비는 1:5, 1:10, 또는 1:20으로부터 선택된다. DNA 또는 RNA 올리고뉴클레오타이드는 자연 발생 또는 비자연 발생일 수 있다. 특정 구현예들에서, 상기 표적 포획 시약(예컨대, 미끼)은, 예컨대 용융 온도를 증가시키기 위해 하나 이상의 비자연 발생 뉴클레오타이드를 포함한다. 예시적인 비자연 발생 올리고뉴클레오타이드는 변형된 DNA 또는 RNA 뉴클레오타이드를 포함한다. 예시적인 변형된 RNA 뉴클레오타이드는 잠금 핵산(LNA)이며, LNA 뉴클레오타이드의 리보스 모이어티는 2' 산소 및 4' 탄소를 연결하는 추가 가교로 변형된다(Kaur, H; Arora, A; Wengel, J; Maiti, S; Arora, A.; Wengel, J.; Maiti, S. (2006). "Thermodynamic, Counterion, and Hydration Effects for the Incorporation of Locked Nucleic Acid Nucleotides into DNA Duplexes". *Biochemistry* 45 (23): 7347-55). 다른 변형된 예시적인 DNA 및 RNA 뉴클레오타이드는 펩티드 결합에 의해 연결된 반복 N-(2-아미노에틸)-글리신 단위로 구성된 펩티드 핵산(PNA)(Egholm, M. *et al.* (1993) *Nature* 365 (6446): 566-8); 낮은 GC 영역을 포획하도록 변형된 DNA 또는 RNA 올리고뉴클레오타이드; 두고리 핵산(BNA) 또는 가교된 올리고뉴클레오타이드; 변형된 5-메틸 데옥시시티딘; 및 2,6-디아미노퓨린을 포함하지만 이에 제한되지는 않는다. 다른 변형된 DNA 및 RNA 뉴클레오타이드는 당업계에 공지되어 있다.
- [0485] 한 구현예에서, 방법은 라이브러리를 수득하는 단계를 추가로 포함하며, 상기 라이브러리 내의 상기 핵산 단편의 크기는 참조 값 이하이고, 상기 라이브러리는 DNA 분리와 라이브러리 제작 사이의 단편화 단계 없이 제조된다.
- [0486] 한 구현예에서, 상기 방법은 핵산 단편을 수득하는 단계를 추가로 포함하고, 상기 핵산 단편의 크기가 참조 값 이상이고 단편화되면, 상기 핵산 단편이 라이브러리로 만들어진다.
- [0487] 한 구현예에서, 상기 방법은, 예컨대 식별가능한 별개의 핵산 서열(바코드)을 복수의 핵산 분자 각각에 추가함으로써 복수의 라이브러리 핵산 분자 각각을 표지하는 단계를 추가로 포함한다.
- [0488] 한 구현예에서, 방법은 프라이머를 복수의 라이브러리 핵산 분자 각각에 부착하는 단계를 추가로 포함한다.
- [0489] 한 구현예에서, 방법은 복수의 표적 포획 시약을 제공하고 복수의 표적 포획 시약을 선택하는 것을 추가로 포함

하며, 상기 선택은 하기로서: 1) 환자 특성, 예컨대 연령, 종양의 병기, 이전 치료 또는 내성; 2) 종양 유형; 3) 샘플의 특성; 4) 대조군 샘플의 특성; 5) 통체의 존재 또는 유형; 6) 분리된 종양(또는 대조군) 핵산 샘플의 특성; 7) 라이브러리 특성; 8) 샘플에서 종양의 유형과 관련된 것으로 공지된 돌연변이; 9) 샘플에서 종양의 유형과 관련된 것으로 공지되지 않은 돌연변이; 10) 서열을 서열분석(또는 혼성화 또는 복구)하거나 돌연변이를 식별하는 능력, 예컨대 높은 GC 영역 또는 재배열을 갖는 서열과 관련된 어려움; 또는 11) 서열분석되는 유전자에 반응한다.

[0490] 한 구현예에서, 방법은, 예컨대 상기 샘플에서 적은 수의 종양 세포의 결정에 반응하여 표적 포획 시약 또는 복수의 표적 포획 시약을 선택하는 단계, 제2 유전자의 핵산 분자와 비교한 제1 유전자의 핵산 분자의 상대적으로 높은 효율성의 포획을 제공하는 단계를 포함하며, 예컨대 상기 제1 유전자의 돌연변이는 상기 샘플의 종양 유형에 대한 종양 표현형과 연관되고, 선택적으로 상기 제2 유전자의 돌연변이는 상기 샘플의 종양 표현형과 연관되지 않는다.

[0491] 한 구현예에서, 상기 방법은 라이브러리 캐치 특성에 대한 값, 예컨대 핵산 농도를 측정하는 단계, 및 측정한 값을 특성에 대한 참조 기준과 비교하는 단계를 추가로 포함한다.

[0492] 한 구현예에서, 방법은 라이브러리 정량화를 위한 참조 기준을 충족하는 라이브러리 특성에 대한 값을 갖는 라이브러리를 선택하는 단계를 추가로 포함한다.

[0493] 서열분석

[0494] 본원에 기재된 방법 및 시스템은 핵산 서열분석을 위한 방법 또는 시스템과 조합하여 또는 그 일부로서 사용될 수 있다.

[0495] 일부 구현예들에서, 라이브러리로부터의 핵산 분자는, 예컨대 용액 혼성화를 사용하여 분리됨으로써 라이브러리 캐치를 제공한다. 상기 라이브러리 캐치 또는 이의 하위군은 서열분석될 수 있다. 따라서, 본원에 기재된 방법은 상기 라이브러리 캐치를 분석하는 단계를 더 포함할 수 있다. 일부 구현예들에서, 상기 라이브러리 캐치는 서열분석 방법, 예컨대 본원에 기재된 바와 같은 차세대 서열분석 방법에 의해 분석된다. 일부 구현예들에서, 상기 방법은 용액 혼성화에 의해 라이브러리 캐치를 분리하고, 상기 라이브러리 캐치를 핵산 서열분석하는 단계를 포함한다. 특정 구현예들에서, 상기 라이브러리 캐치는 재서열분석된다.

[0496] 당업계에 공지된 임의의 서열분석 방법이 사용될 수 있다. 예컨대, 용액 혼성화에 의해 분리된 핵산의 서열분석은 일반적으로 차세대 서열분석(NGS)을 사용하여 실시한다. 본원에서 사용하기에 적합한 서열분석 방법은, 예컨대 국제특허출원공보 제WO 2012/092426에 기재된 바와 같이 당업계에 기재되어 있다.

[0497] 한 구현예에서, 측정 또는 분석된 리드의 적어도 10, 20, 30, 40, 50, 60, 70, 80, 또는 90%는 본원에 기재된 유전자, 예컨대 표 2A-5B의 유전자로부터의 대상체 간격에 대한 것이다. 한 구현예에서, 적어도 0.01, 0.02, 0.03, 0.04, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 2.0, 5.0, 10, 15, 또는 30개의 메가염기(megabase), 예컨대 유전체 염기가 서열분석된다. 한 구현예에서, 상기 방법은 본원에 기재된 샘플로부터 측정한 뉴클레오타이드 서열 리드를 측정하는 단계를 포함한다. 한 구현예에서, 상기 리드는 NGS 서열분석 방법에 의해 제공된다.

[0498] 본원에 개시된 방법은 대상체의 유전체, 전체 엑솜 또는 전사체에 존재하는 변형을 검출하는 데 사용될 수 있고, DNA 및 RNA 서열분석, 예컨대 표적화된 DNA 및/또는 RNA 서열분석에 적용될 수 있다. 일부 구현예들에서, 본원에 기재된 유전자의 전사체가 서열분석된다. 다른 구현예들에서, 상기 방법은 유전자 또는 유전자 산물의 수준에서의 변화(예컨대, 증가 또는 감소), 예컨대 본원에 기재된 유전자 또는 유전자 산물의 발현 변화의 검출을 포함한다. 상기 방법은 선택적으로 표적 RNA에 대한 샘플을 농축하는 단계를 포함할 수 있다. 다른 구현예들에서, 상기 방법은 특정 고존재량 RNA, 예컨대 리보솜 또는 글로빈 RNA의 샘플을 고갈시키는 단계를 포함한다. 상기 RNA 서열분석 방법은 단독으로 또는 본원에 기재된 DNA 서열분석 방법과 조합하여 사용될 수 있다. 일 구현예에서, 상기 방법은 DNA 서열분석 단계 및 RNA 서열분석 단계를 실시하는 단계를 포함한다. 상기 방법은 임의의 순서로 실시될 수 있다. 예를 들어, 상기 방법은 RNA 서열분석에 의해 본원에 기재된 변형의 발현을 확인하는 단계, 예컨대 본 발명의 DNA 서열분석 방법에 의해 검출된 돌연변이 또는 융합의 발현을 확인하는 단계를 포함할 수 있다. 다른 구현예들에서, 상기 방법은 RNA 서열분석 단계 후, DNA 서열분석 단계를 실시하는 것을 포함한다.

[0499] 정렬

- [0500] 본원에 개시된 방법은 다중의 개별적으로 조절된 정렬 방법 또는 알고리즘의 사용을 통합하여 서열분석 방법, 특히 다수의 다양한 유전자에서 다수의 다양한 유전적 이벤트의 대규모 병렬 서열분석에 의존하는 방법, 예컨대 본원에 기재된 암으로부터의 샘플을 분석하는 방법에서 성능을 최적화할 수 있다.
- [0501] 일부 구현예들에서, 리드를 분석하기 위해 사용되는 정렬 방법은 상이한 유전자의 다수의 변이체 각각에 맞게 개별적으로 맞춤화되거나 조정되지 않는다. 일부 구현예들에서, 상이한 유전자에서 다수의 변이체의 적어도 하위 집합에 맞게 개별적으로 맞춤화되거나 조정되는 다중 정렬 방법을 사용하여 리드를 분석한다. 일부 구현예들에서, 상이한 유전자에서 다수의 변이체 각각에 맞게 개별적으로 맞춤화되거나 조정되는 다중 정렬 방법을 사용하여 리드를 분석한다. 일부 구현예들에서, 조정은 서열분석되는 유전자(또는 다른 대상체 간격), 샘플 내의 종양 유형, 서열분석되는 변이체, 또는 샘플 또는 대상체의 특성(중 하나 이상)의 함수일 수 있다. 서열분석할 다수의 대상체 간격에 맞게 개별적으로 조절된 정렬 조건을 선택하거나 사용하면 속도, 감도 및 특이성을 최적화할 수 있다. 상기 방법은 비교적 많은 수의 다양한 대상체 간격에 대한 리드의 정렬이 최적화될 때 특히 효과적이다.
- [0502] 일부 구현예들에서, X개의 고유한 대상체 간격의 각각으로부터의 리드는 고유한 정렬 방법으로 정렬되며, 고유한 대상체 간격(예컨대, 대상체 간격 또는 발현된 대상체 간격)은 다른 X-1 대상체 간격과 상이한 것을 의미하고, 상기 고유한 정렬 방식은 다른 X-1 정렬 방식과 상이한 것을 의미하며, X는 2 이상이다.
- [0503] 한 구현예에서, 적어도 X개의 유전자, 예컨대 표 2A-5B로부터의 적어도 X개의 유전자에서의 대상체 간격은 고유한 정렬 방법으로 정렬되고, X는 2, 3, 4, 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500 이상과 같다.
- [0504] 한 구현예에서, 상기 방법은 리드를 분석, 예컨대 정렬하기 위한 정렬 방법을 선택하거나 사용하는 단계를 포함하며, 상기 정렬 방법은 하기로서:
- [0505] (i) 종양 유형, 예컨대 상기 샘플의 종양 유형;
- [0506] (ii) 서열분석되는 상기 대상체 간격(예컨대, 대상체 간격 또는 발현된 대상체 간격)이 위치하는 유전자 또는 유전자 유형, 예컨대 변이체 또는 변이체의 유형, 예컨대 돌연변이, 또는 빈도의 돌연변이를 특징으로 하는 유전자 또는 유전자 유형;
- [0507] (iii) 분석되는 부위(예컨대, 뉴클레오티드 위치);
- [0508] (iv) 평가되는 대상체 간격(예컨대, 대상체 간격 또는 발현된 대상체 간격) 내의 변이체 유형, 예컨대 치환;
- [0509] (v) 샘플의 유형, 예컨대 본원에 기재된 샘플; 및
- [0510] (vi) 평가되는 상기 대상체 간격 내 또는 근처의 서열, 예컨대 상기 대상체 간격(예컨대, 대상체 간격 또는 발현된 대상체 간격)에 대한 오정렬의 예상되는 경향, 예컨대 상기 대상체 간격 내 또는 근처에서 반복되는 서열의 존재(예컨대, 대상체 간격 또는 발현된 대상체 간격) 중 하나 이상 또는 모두의 함수이거나, 이에 응답하여 선택되거나, 최적화된다.
- [0511] 본원의 다른 곳에서 언급된 바와 같이, 일부 구현예들에서, 방법은 비교적 많은 수의 대상체 간격에 대한 리드의 정렬이 최적화될 때 특히 효과적이다. 따라서, 한 구현예에서, 적어도 X개의 고유한 정렬 방법이 적어도 X개의 고유한 대상체 간격에 대한 리드를 분석하는 데 사용되며, 고유한은 다른 X-1과 상이하고 X는 2, 3, 4, 5, 10, 15, 20, 30, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1,000 이상과 같다.
- [0512] 한 구현예에서, 표 2A-5B로부터의 적어도 X개의 유전자에서의 대상체 간격은 분석되고, X는 2, 3, 4, 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500 이상과 같다.
- [0513] 한 구현예에서, 고유한 정렬 방법이 적어도 3, 5, 10, 20, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 또는 500개의 상이한 유전자 각각에서 대상체 간격에 적용된다.
- [0514] 한 구현예에서, 적어도 20, 40, 60, 80, 100, 120, 140, 160 또는 180, 200, 300, 400, 또는 500개의 유전자, 예컨대 표 2A-5B로부터의 유전자 내의 뉴클레오티드 위치에 뉴클레오티드 값이 할당된다. 한 구현예에서, 고유한 정렬 방법은 분석되는 상기 유전자의 적어도 10, 20, 30, 40, 또는 50% 각각에서 대상체 간격에 적용된다.
- [0515] 본원에 개시된 방법은 문제가 있는 리드, 예컨대 재배열을 갖는 리드의 신속하고 효율적인 정렬을 가능하게 한다. 따라서, 대상체 간격(예컨대, 대상체 간격 또는 발현된 대상체 간격)에 대한 리드가 재배열, 예컨대 전좌가

있는 뉴클레오타이드 위치를 포함하는 구현예에서, 상기 방법은 적절하게 조정되고 하기로서:

- [0516] 리드와의 정렬을 위해 재배열 참조 서열을 선택하는 단계로서, 상기 재배열 참조 서열은 재배열과 정렬되는(일부 구현예들에서, 상기 참조 서열은 상기 유전체 재배열과 동일하지 않음) 단계; 및
- [0517] 리드를 상기 재배열 참조 서열과 비교, 예컨대 정렬하는 단계를 포함하는 정렬 방법을 사용하는 것을 포함할 수 있다.
- [0518] 일부 구현예들에서, 상이한 다른 방법, 예컨대 다른 방법이 문제가 많은 리드를 정렬하는 데 사용된다. 상기 방법은 비교적 많은 수의 다양한 대상체 간격에 대한 리드의 정렬이 최적화될 때 특히 효과적이다. 예로서, 샘플을 분석하는 방법은:
- [0519] 제1 매개변수 세트(예컨대, 제1 맵핑 알고리즘 또는 제1 참조 서열과 함께) 하에 리드의 비교, 예컨대 정렬 비교를 실시하는 단계, 및
- [0520] 상기 리드가 제1 정렬 기준을 충족하는지 여부를 결정하는 단계(예컨대, 상기 리드는 상기 제1 참조 서열과, 예컨대 다수보다 적은 불일치가 정렬될 수 있음);
- [0521] 상기 리드가 상기 제1 정렬 기준을 충족하지 못하는 경우, 매개변수의 제2 세트(예컨대, 제2 맵핑 알고리즘 또는 제2 참조 서열과 함께) 하에 제2 정렬 비교를 실시하는 단계; 및,
- [0522] 선택적으로, 상기 리드가 상기 제2 기준을 충족하는지 여부를 결정하는 단계(예컨대, 상기 리드가 소정의 수 미만의 불일치로 상기 제2 참조 서열과 정렬될 수 있음)를 포함하며,
- [0523] 상기 제2 매개변수 세트는 매개변수 세트, 예컨대 상기 제2 참조 서열의 사용을 포함하며, 이는 상기 제1 매개변수 세트와 비교하여 변이체, 예컨대 재배열, 예컨대 삽입, 삭제 또는 전좌에 대한 리드와 정렬될 가능성이 더 높다..
- [0524] 구현예들에서, 본원의 "정렬" 섹션의 정렬 방법은 본원의 "돌연변이 호출" 섹션의 돌연변이 호출 방법 및/또는 본원의 "표적 포획 시약" 섹션의 및/또는 본원의 "표적 포획 시약의 설계 및 구성"의 섹션의 표적 포획 시약과 조합된다. 상기 방법은 본원의 "유전자 선택" 섹션의 대상체 간격 세트 및/또는 본원의 "대상체" 섹션의 대상체로부터의 "샘플" 섹션의 샘플에 적용될 수 있다.
- [0525] 정렬은 일반적으로 리드를 위치, 예컨대 유전체 위치와 일치시키는 공정이다. 오정렬(예컨대, 유전체의 부정확한 위치에 대한 짧은 리드부터의 염기쌍 배치), 예컨대 실제 암 돌연변이 주변의 리드의 서열 문맥(예컨대, 반복 서열의 존재)으로 인한 오정렬은, 대안적인 대립유전자의 리드가 대안적인 대립유전자 리드의 주요 무리에서 벗어날 수 있기 때문에, 돌연변이 검출의 민감도를 감소시킬 수 있다. 실제 돌연변이가 존재하지 않는 곳에서 문제가 많은 서열 문맥이 발생하면, 오정렬로 인해 참조 유전체 염기의 실제 리드를 잘못된 위치에 배치함으로써 "돌연변이된" 대립유전자의 인위적 리드를 도입할 수 있다. 크게 증가된 다중 유전자 분석을 위한 돌연변이 호출 알고리즘은 낮은 존재량의 돌연변이에도 민감해야 하기 때문에, 상기 오정렬은 위양성 발견 비율을 증가시키거나 특이성을 감소시킬 수 있다.
- [0526] 본원에서 논의된 바와 같이, 실제 돌연변이에 대한 감소된 감도는 분석되는 유전자에서 예상되는 돌연변이 부위 주변의 정렬의 품질을 (수동으로 또는 자동화된 방식으로) 평가함으로써 다루어질 수 있다. 평가할 부위는 암 돌연변이 데이터베이스(예컨대, COSMIC)에서 수득할 수 있다. 문제가 있는 것으로 식별된 영역은, 예컨대 스미스-워터맨(Smith-Waterman)과 같이 더 느리지만 더 정확한 정렬 알고리즘을 사용하여 정렬 최적화(또는 재정렬)를 통해 관련 서열 문맥에서 더 나은 성능을 제공하도록 선택된 알고리즘을 사용하여 수정할 수 있다. 일반 정렬 알고리즘이 상기 문제를 해결할 수 없는 경우, 예컨대 치환을 포함할 가능성이 높은 유전자에 대한 최대 차이 불일치 별점 매개변수의 조정; 특정 종양 유형에서 일반적인 특정 돌연변이 유형(예컨대, 흑색종의 C<sub>â</sub>T)에 기반한 특정 불일치 별점 매개변수의 조정; 또는 특정 샘플 유형에서 공통적인 특정 돌연변이 유형(예컨대, FFPE에서 공통적인 치환)에 기반한 불일치 별점 매개변수를 조정에 의해 맞춤형 정렬 접근법이 생성될 수 있다.
- [0527] 오정렬로 인한 평가된 유전자 영역의 감소된 특이성(위양성 비율 증가)은 서열분석된 샘플의 모든 돌연변이 호출에 대한 수동 또는 자동 검사로 평가할 수 있다. 오정렬로 인해 가짜 돌연변이 호출이 발생하기 쉬운 것으로 밝혀진 영역은 상기와 동일한 정렬 조치를 받을 수 있다. 알고리즘적 조치가 가능하지 않은 경우, 문제 영역의 "돌연변이"를 분류하거나 시험 패널에서 배제할 수 있다.
- [0528] 본원에 개시된 방법은 특히, 예컨대 샘플로부터의 다수의 다양한 유전자 내 다수의 다양한 유전적 사건의 대규모

모 병렬 서열분석에 의존하는 방법에서 재배열, 예컨대 삽입-결실과 관련된 대상체 간격의 서열분석에서 성능을 최적화하기 위해 개별적으로 조정된 다중 정렬 방법 또는 알고리즘의 사용을 가능하게 한다. 일부 구현예들에서, 상이한 유전자의 다수의 재배열 각각에 대해 개별적으로 맞춤화되거나 조정되는 다중 정렬 방법을 사용하여 리드를 분석한다. 일부 구현예들에서, 조정은 서열분석되는 하나 이상의 대상체 간격(예컨대, 하나 이상의 유전자), 샘플과 관련된 종양 유형, 서열분석되는 변이체, 또는 샘플 또는 대상체의 특성 중 하나 이상의 함수일 수 있다. 서열분석할 다수의 대상체 간격에 맞게 미세하게 조정된 정렬 조건을 선택하거나 사용하면 속도, 감도 및 특이성을 최적화할 수 있다. 상기 방법은 비교적 많은 수의 다양한 대상체 간격에 대한 리드의 정렬이 최적화될 때 특히 효과적이다. 구현예들에서, 상기 방법은 재배열에 최적화된 정렬 방법 및 재배열과 관련되지 않은 대상체 간격에 최적화된 다른 방법의 사용을 포함한다.

- [0529] 일부 구현예들에서, 정렬 선택자가 사용된다. 본원에서 사용된 "정렬 선택자"는 대상체 간격의 서열분석을 최적화할 수 있는 정렬 방법, 예컨대 정렬 알고리즘 또는 매개변수의 선택을 허용하거나 지시하는 매개변수를 지칭한다. 정렬 선택자는, 예컨대 하기 중 하나 이상에 대해 특이적이거나 기능으로 선택될 수 있다:
- [0530] 1. 상기 대상체 간격에 대한 리드의 오정렬 경향과 관련된 서열 문맥, 예컨대 대상체 간격(예를 들어, 평가될 뉴클레오티드 위치)의 서열 문맥, 예컨대, 유전체의 다른 곳에서 반복되는 평가될 대상체 간격 내 또는 근처에 서열 요소의 존재는 오정렬을 일으켜 성능을 저하시킬 수 있다. 오정렬을 최소화하는 알고리즘 또는 알고리즘 매개변수를 선택하여 성능을 향상시킬 수 있다. 상기의 경우, 정렬 선택자에 대한 값은 서열 문맥, 예컨대 상기 유전체(또는 분석되는 유전체의 부분)에서 최소 횟수로 반복되는 길이의 서열의 존재 또는 부재의 함수일 수 있다.
- [0531] 2. 분석 중인 종양 유형. 예컨대, 특정 종양 유형은 증가된 결실 비율을 특징으로 할 수 있다. 따라서, 삽입-결실에 더 민감한 알고리즘 또는 알고리즘 매개변수를 선택하여 성능을 향상시킬 수 있다. 상기의 경우, 정렬 선택자에 대한 값은 종양 유형, 예컨대 종양 유형에 대한 식별자의 함수일 수 있다. 한 구현예에서, 상기 값은 종양 유형, 예컨대 고형 종양 또는 혈액 악성종양(또는 전암)의 동일성이다.
- [0532] 3. 분석되는 유전자 또는 유전자 유형, 예컨대 유전자 또는 유전자 유형이 분석될 수 있다. 예로서, 종양유전자는 종종 치환 또는 틀내 삽입-결실을 특징으로 한다. 따라서, 상기 변이체에 특히 민감하고 다른 변이체에 대항하여 특이적인 알고리즘 또는 알고리즘 매개변수를 선택하여 성능을 향상시킬 수 있다. 종양 억제체는 종종 틀 이동 삽입-결실을 특징으로 한다. 따라서, 상기 변이체에 특히 민감한 알고리즘 또는 알고리즘 매개변수를 선택하여 성능을 향상시킬 수 있다. 따라서, 대상체 간격과 일치하는 알고리즘 매개변수를 선택하여 성능을 향상시킬 수 있다. 상기의 경우, 정렬 선택자에 대한 값은 유전자, 유전자 유형, 예컨대 유전자 또는 유전자 유형에 대한 식별자의 함수일 수 있다. 한 구현예에서, 상기 값은 유전자의 동일성이다.
- [0533] 4. 분석되는 부위(예컨대, 뉴클레오티드 위치). 상기의 경우, 정렬 선택자에 대한 값은 부위, 부위 유형, 예컨대 부위 또는 부위 유형에 대한 식별자의 함수일 수 있다. 한 구현예에서, 상기 값은 부위의 동일성이다. (예를 들어, 해당 부위를 포함하는 유전자가 다른 유전자와 고도로 상동성인 경우, 정상/빠른 짧은 리드 정렬 알고리즘(예컨대, BWA)은 두 유전자 간을 구별하는 데 어려움을 겪을 수 있으며 잠재적으로 더 집중적인 정렬 방법(스미스-워터맨) 또는 심지어 조립(ARACHNE)이 필요할 수 있다. 유사하게, 유전자 서열이 낮은 복잡성 영역(예컨대, AAAAAA)을 포함하는 경우, 더 집중적인 정렬 방법이 필요할 수 있다.
- [0534] 5. 평가되는 대상체 간격과 관련된 변이체 또는 이의 변이체 유형. 예컨대, 치환, 삽입, 결실, 전좌 또는 기타 재배열. 따라서, 상기 특이적 변이체 유형에 특히 민감한 알고리즘 또는 알고리즘 매개변수를 선택하여 성능을 향상시킬 수 있다. 상기의 경우, 정렬 선택자에 대한 값은 변이체 유형, 예컨대 변이체 유형에 대한 식별자의 함수일 수 있다. 한 구현예에서, 상기 값은 변이체 유형의 동일성, 예컨대 치환이다.
- [0535] 6. 샘플의 유형, 예컨대 본원에 기재된 샘플. 샘플 유형/품질은 오류(비참조 서열의 가짜 관찰) 비율에 영향을 줄 수 있다. 따라서, 상기 샘플의 실제 오류율을 정확하게 모델링하는 알고리즘 또는 알고리즘 매개변수를 선택하여 성능을 향상시킬 수 있다. 이 경우, 정렬 선택자에 대한 값은 샘플 유형, 예컨대 샘플 유형에 대한 식별자의 함수일 수 있다. 한 구현예에서, 상기 값은 샘플 유형의 동일성이다.
- [0536] 일반적으로, 삽입-결실 돌연변이의 정확한 검출은 정렬의 실행이고, 이는 비활성화된 서열분석의 가짜 삽입-결실 비율이 본원에서 상대적으로 낮기 때문이다(따라서, 올바르게 정렬된 삽입-결실에 대한 소수의 관찰조차도 돌연변이의 강력한 증거가 될 수 있음). 하지만, 삽입-결실의 존재하에서는 (특히, 삽입-결실의 길이가 증가함에 따라)정확한 정렬이 어려울 수 있다. 정렬과 관련된 일반적인 문제, 예컨대 치환에 더하여, 삽입-결실 자체

가 정렬 문제를 일으킬 수 있다. (예를 들어, 디뉴클레오타이드 반복의 2bp의 결실은 쉽게 결정적으로 위치할 수 없다.) 감도 및 특이성 모두가 더 짧고(<15bp) 분명한 삽입-결실이 포함된 리드가 잘못 배치되어 감소될 수 있다. 더 큰 삽입-결실(개별 리드의 길이에 더 가까워짐, 예컨대 36bp 리드)은 리드 정렬에 실패하여 표준 정렬된 리드 세트에서 삽입-결실의 검출을 불가능하게 만들 수 있다.

[0537] 암 돌연변이의 데이터베이스는 이러한 문제를 해결하고 성능을 개선하는 데 사용할 수 있다. 위양성 삽입-결실 발견을 줄이기 위해(특이성 개선), 일반적으로 예상되는 삽입-결실 주변 영역에서 서열 문맥으로 인해 문제가 있는 정렬을 검사하고 위의 치환과 유사하게 처리할 수 있다. 삽입-결실 검출의 민감도를 개선하기 위해 암에서 예상되는 삽입-결실에 대한 정보를 사용하는 여러 상이한 접근 방식을 사용할 수 있다. 예컨대, 예상되는 삽입-결실이 포함된 짧은 리드를 시뮬레이션하고 정렬을 시도할 수 있다. 상기 정렬을 연구할 수 있고 문제가 있는 삽입-결실 영역은, 예컨대 겹 열기/확장 별점을 줄이거나 부분 리드(예컨대, 리드의 전반부 또는 후반부)를 정렬하여 정렬 매개변수를 조정할 수 있다.

[0538] 대안적으로, 초기 정렬은 정상 참조 유전체뿐만 아니라 공지된 또는 가능성이 있는 암 삽입-결실 돌연변이 각각을 함유하는 유전체의 대체 버전으로도 시도할 수 있다. 상기 접근 방식에서, 처음에 정렬에 실패했거나 잘못 정렬된 삽입-결실의 리드는 유전체의 대체(돌연변이) 버전에 성공적으로 배치된다.

[0539] 이러한 방식으로, 삽입-결실 정렬(및 이에 따른 호출)은 예상되는 암 유전자/부위에 대해 최적화될 수 있다. 본원에서 사용된 바와 같이, 서열 정렬 알고리즘은 리드 서열 및 참조 서열 간의 유사성을 평가함으로써 유전체 내의 리드 서열(예컨대, 차세대 서열분석으로부터의 짧은 리드 서열)이 유래되었을 가능성이 가장 높은 곳을 확인하기 위해 사용되는 계산 방법 또는 접근법을 구현한다. 다양한 알고리즘을 서열 정렬 문제에 적용할 수 있다. 일부 알고리즘은 상대적으로 느리지만, 상대적으로 높은 특이성을 가능하게 한다. 여기에는, 예컨대 동적 프로그래밍 기반 알고리즘이 포함된다. 동적 프로그래밍은 복잡한 문제를 더 간단한 단계로 분해하여 해결하는 방법이다. 다른 접근 방식이 상대적으로 더 효율적이지만 일반적으로 철저하지는 않다. 여기에는, 예컨대 대규모 데이터베이스 검색을 위해 설계된 발견적 알고리즘 및 확률적 방법이 포함된다.

[0540] 정렬 매개변수는 알고리즘의 성능을 조정하기 위해, 예컨대 리드 서열과 참조 서열 사이에 최적의 전체 또는 부분 정렬을 생성하기 위해 정렬 알고리즘에서 사용된다. 정렬 매개변수는 일치, 불일치 및 삽입-결실에 대한 가중치를 제공할 수 있다. 예를 들어, 가중치가 낮을수록 불일치 및 삽입-결실이 더 많은 정렬이 허용된다.

[0541] 서열 문맥, 예컨대 반복 서열의 존재(예컨대, 직렬 반복, 산재된 반복), 복잡성이 낮은 영역, 삽입-결실, 유사 유전자 또는 파라로그의 존재는 정렬 특이성에 영향을 미칠 수 있다(예컨대, 오정렬 유발). 본원에서 사용된 바와 같이, 오정렬은 유전체의 부정확한 위치에 대한 짧은 리드로부터 염기쌍의 배치를 지칭한다.

[0542] 정렬의 민감도는 정렬 알고리즘이 선택되거나 정렬 매개변수가 종양 유형, 예컨대 특정 돌연변이 또는 돌연변이 유형을 갖는 경향이 있는 종양 유형을 기반으로 조정되는 경우 증가할 수 있다.

[0543] 특정 유전자 유형(예컨대, 종양 유전자, 종양 억제 유전자)에 따라 정렬 알고리즘을 선택하거나 정렬 매개변수를 조정하면 정렬 민감도가 증가할 수 있다. 상이한 유형의 암 관련 유전자의 돌연변이는 암 표현형에 상이한 영향을 미칠 수 있다. 예를 들어, 돌연변이 종양유전자 대립유전자가 일반적으로 우세하다. 돌연변이 종양 억제 대립유전자는 일반적으로 열성이며, 이는 대부분의 경우 종양 억제 유전자의 두 대립유전자가 모두 영향을 받은 후에야 효과가 나타남을 의미한다.

[0544] 정렬 감도는 정렬 알고리즘이 선택되거나 돌연변이 유형(예컨대, 단일 염기 다형성, 삽입-결실, 역위, 전좌, 직렬 반복)에 따라 정렬 매개변수가 조정될 때 조정(예컨대, 증가)될 수 있다.

[0545] 정렬 감도는 정렬 알고리즘이 선택되거나 돌연변이 부위(예컨대, 돌연변이 다발점)를 기반으로 정렬 매개변수가 조정될 때 조정(예컨대, 증가)될 수 있다. 돌연변이 다발점은 돌연변이가 정상 돌연변이 비율보다 최대 100배 더 자주 발생하는 유전체의 부위를 지칭한다.

[0546] 정렬의 민감도/특이성은 정렬 알고리즘이 선택되거나 샘플 유형(예컨대, cfDNA 샘플, ctDNA 샘플, FFPE 샘플 또는 CTC 샘플)에 따라 정렬 매개변수가 조정될 때 조정(예컨대, 증가)될 수 있다.

[0547] 일부 구현예들에서, NGS 리드는 공지된 참조 서열에 정렬되거나 새로 조립될 수 있다. 예를 들어, 상기 NGS 리드는 참조 서열(예컨대, 야생형 서열)에 정렬될 수 있다. NGS에 대한 서열 정렬 방법이 Trapnell C. and Salzberg S.L. *Nature Biotech.*, 2009, 27:455-457에 기재되어 있다. 새로운 조립의 예는, 예컨대 Warren R. *et al.*, *Bioinformatics*, 2007, 23:500-501; Butler J. *et al.*, *Genome Res.*, 2008, 18:810-820; 및 Zerbino

D.R. and Birney E., *Genome Res.*, 2008, 18:821-829에 기재되어 있다. 서열 정렬 또는 조립은 하나 이상의 NGS 플랫폼으로부터 리드 데이터를 사용하여, 예컨대 로슈(Roche)/454 및 일루미나(Illumina)/솔렉사(Solexa) 리드 데이터를 혼합하여 실시할 수 있다.

[0548] 정렬의 최적화는, 예컨대 국제특허출원공보 제WO 2012/092426호에 기재된 바와 같이 당업계에 기재되어 있다.

[0549] 돌연변이 호출

[0550] 본원에 개시된 방법은 맞춤 또는 조정된 호출 매개변수의 사용을 통합하여 서열분석 방법, 특히 다수의 다양한 유전자에서, 예컨대 샘플에서, 예컨대 본원에 기재된 암으로부터 다수의 다양한 유전적 이벤트의 대규모 병렬 서열분석에 의존하는 방법에서 성능을 최적화할 수 있다.

[0551] 일부 구현예들에서, 다수의 대상체 간격 각각에 대한 돌연변이 호출은 개별적으로 맞춤화되거나 미세 조정되지 않는다. 일부 구현예들에서, 다수의 대상체 간격의 적어도 하위집합에 대한 돌연변이 호출은 개별적으로 맞춤화되거나 미세 조정된다. 일부 구현예들에서, 다수의 대상체 간격 각각에 대한 돌연변이 호출은 개별적으로 맞춤화되거나 미세 조정된다. 맞춤화 또는 조정은 본원에 기재된 하나 이상의 인자, 예컨대 샘플 내 암의 유형, 서열분석될 대상체 간격이 위치한 유전자, 또는 서열분석될 변이체에 기초할 수 있다. 서열분석할 다수의 대상체 간격에 맞게 미세하게 조정된 정렬 조건을 선택하거나 사용하면 속도, 감도 및 특이성을 최적화할 수 있다. 상기 방법은 비교적 많은 수의 다양한 대상체 간격에 대한 리드의 정렬이 최적화될 때 특히 효과적이다.

[0552] 일부 구현예들에서, X개의 고유한 대상체 간격 각각의 뉴클레오타이드 위치에 대해 뉴클레오타이드 값이 지정되고 고유한 호출 방법에 의해 할당되며, 고유한 대상체 간격은 다른 X-1 대상체 간격(예컨대, 하위 유전체 간격, 발현된 하위 유전체 간격 또는 둘 모두)과 상이함을 의미하고, 상기 고유한 호출 방법은 다른 X-1 호출 방법과 상이함을 의미하고, X는 적어도 2이다. 상기 호출 방법은 상이할 수 있으므로, 예컨대 상이한 페이지안 이전 값에 의존함으로써 고유할 수 있다.

[0553] 한 구현예에서, 상기 뉴클레오타이드 값을 할당하는 것은 유형의 종양의 상기 뉴클레오타이드 위치에서 변이체, 예컨대 돌연변이를 나타내는 리드를 관찰하는 것에 대한 사전(예컨대, 문헌) 예상이거나 예상을 나타내는 값의 함수이다.

[0554] 한 구현예에서, 상기 방법은 적어도 10, 20, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 또는 1,000개의 뉴클레오타이드 위치에 대한 뉴클레오타이드 값을 할당(예컨대, 돌연변이를 호출)하는 단계를 포함하고, 여기서 각 할당은 변이체, 예컨대 유형의 종양에서 상기 뉴클레오타이드 위치에서의 돌연변이를 나타내는 리드를 관찰하는 것에 대한 사전(예컨대, 문헌) 예상이거나 예상을 나타내는 고유한(다른 할당에 대한 값과 대조적으로) 값의 함수이다.

[0555] 한 구현예에서, 상기 뉴클레오타이드 값을 할당하는 것은 변이체가 샘플에 일정 빈도(예컨대, 1%, 5%, 10% 등)로 존재하는 경우 및/또는 상기 변이체가 부재한 경우(예컨대, 염기 호출 오류만으로 리드에서 관찰됨) 상기 뉴클레오타이드 위치에서의 상기 변이체를 나타내는 리드를 관찰할 확률을 나타내는 일련의 값들의 함수이다.

[0556] 한 구현예에서, 본원에 기재된 돌연변이 호출 방법은:

[0557] 상기 X 대상체 간격의 각각에서 뉴클레오타이드 위치에 대해:

[0558] (i) X 유형의 종양의 상기 뉴클레오타이드 위치에서 변이체, 예컨대 돌연변이를 나타내는 리드를 관찰하는 것에 대한 사전(예컨대, 문헌) 예상이거나 예상을 나타내는 제1 값; 및

[0559] (ii) 상기 변이체가 샘플에 일정 빈도(예컨대, 1%, 5%, 10% 등)로 존재하는 경우 및/또는 상기 변이체가 부재한 경우(예컨대, 염기 호출 오류만으로 리드에서 관찰됨) 상기 뉴클레오타이드 위치에서의 상기 변이체를 나타내는 리드를 관찰할 확률을 나타내는 제2 세트의 값들을 수득하는 단계;

[0560] 상기 값에 응답하여, 예컨대 본원에 기재된 페이지안 방법에 의해 청량함으로써, 상기 제1 값을 사용하여 상기 제2 세트의 값들 사이의 비교(예컨대, 돌연변이 존재의 사후 확률을 계산)에 의해 상기 뉴클레오타이드 위치의 각각에 대한 상기 리드로부터 뉴클레오타이드 값을 할당(예컨대, 돌연변이를 호출함)함으로써, 상기 샘플을 분석하는 단계를 포함할 수 있다.

[0561] 한 구현예에서, 상기 방법은 하기의 단계들 중 하나 이상 또는 모두를 포함한다:

[0562] (i) 적어도 10, 20, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 또는 1,000개의

뉴클레오타이드 위치에 대한 뉴클레오타이드 값을 할당(예컨대, 돌연변이 호출)하는 단계로서, 각 할당은 고유한(다른 할당과 대조적으로) 제1 및/또는 제2 값을 기반으로 하는 단계;

- [0563] (ii) (i)의 방법을 할당하는 단계로서, 적어도 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 또는 500개의 할당은, 예컨대 종양 유형의 세포의 5, 10, 또는 20% 미만에 존재하는 변이체의 확률의 함수인 값으로 이루어지는 단계;
- [0564] (iii) 적어도 X개의 뉴클레오타이드 위치에 대한 뉴클레오타이드 값을 할당(예컨대 돌연변이를 호출)하는 단계로서, 이들 각각은 유형의 종양, 예컨대 상기 샘플의 종양 유형 내에 존재할 고유(다른 X-1 할당과 대조적으로) 확률을 갖는 변이체와 관련되고, 선택적으로 각각의 상기 X 할당은 고유한(다른 X-1 할당과 대조적으로) 제1 및/또는 제2 값(X= 2, 3, 5, 10, 20, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 또는 500)에 기초하는 단계;
- [0565] (iv) 제1 및 제2 뉴클레오타이드 위치에서 뉴클레오타이드 값을 할당(예컨대, 돌연변이 호출)하는 단계로서, 상기 제1 뉴클레오타이드 위치의 제1 변이체가 유형의 종양(예컨대, 상기 샘플의 종양 유형)에 존재할 가능성은 상기 제2 뉴클레오타이드 위치의 제2 변이체가 존재할 가능성보다 적어도 2, 5, 10, 20, 30, 또는 40배 더 크고, 선택적으로 각 할당은 고유한(다른 할당과 대조적으로) 제1 및/또는 제2 값에 기초하는 단계;
- [0566] (v) 뉴클레오타이드 값을 복수의 뉴클레오타이드 위치(예컨대, 돌연변이 호출)에 할당하는 단계로서, 상기 복수의 위치는 하기의 확률 백분율 범위: 0.01 이하; 0.01 초과 0.02 이하; 0.02 초과 0.03 이하; 0.03 초과 0.04 이하; 0.04 초과 0.05 이하; 0.05 초과 0.1 이하; 0.1 초과 0.2 이하; 0.2 초과 0.5 이하; 0.5 초과 1.0 이하; 1.0 초과 2.0 이하; 2.0 초과 5.0 이하; 5.0 초과 10.0 이하; 10.0 초과 20.0 이하; 20.0 초과 50.0 이하; 50 초과 100.0% 이하 중에서 하나 이상, 예컨대 적어도 3, 4, 5, 6, 7개, 또는 모두에 속하는 변이체에 대한 할당을 포함하는 단계로서;
- [0567] 확률 범위는 뉴클레오타이드 위치의 변이체가 종양 유형(예컨대, 상기 샘플의 종양 유형)에 존재할 확률의 범위 또는 뉴클레오타이드 위치의 변이체가 미리 선택된 유형(예컨대, 상기 샘플의 종양 유형)에 대한 샘플, 상기 샘플의 라이브러리 또는 해당 라이브러리의 라이브러리 캐치 내 세포의 인용된 백분율(%)로 존재하고; 그리고
- [0568] 선택적으로, 각각의 할당은 고유한 제1 및/또는 제2 값(예컨대, 인용된 확률 범위에서 다른 할당과 대조적으로 고유하거나 다른 나열된 확률 범위의 하나 이상 또는 모두에 대한 제1 및/또는 제2 값과 대조적으로 고유한)에 기초하는 단계;
- [0569] (vi) 적어도 1, 2, 3, 5, 10, 20, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 또는 1,000개의 뉴클레오타이드 위치에 대한 뉴클레오타이드 값을 할당하는(예컨대, 돌연변이 호출) 단계로서, 각각 독립적으로 상기 샘플에서 DNA의 50, 40, 25, 20, 15, 10, 5, 4, 3, 2, 1, 0.5, 0.4, 0.3, 0.2, 또는 0.1% 미만으로 존재하는 변이체를 갖고, 선택적으로 각 할당은 고유한(다른 할당과 대조적으로) 제1 및/또는 제2 값을 기반으로 하는 단계;
- [0570] (vii) 제1 및 제2 뉴클레오타이드 위치에서 뉴클레오타이드 값을 할당(예컨대, 돌연변이 호출)하는 단계로서, 상기 샘플의 DNA 내 상기 제1 위치에서 변이체의 가능성은 상기 샘플의 상기 제2 뉴클레오타이드 위치의 변이체의 가능성보다 적어도 2, 5, 10, 20, 30, 또는 40배 더 크고, 선택적으로 각 할당은 고유한(다른 할당과 대조적으로) 제1 및/또는 제2 값에 기초하는 단계;
- [0571] (viii) 다음 중 하나 이상 또는 모두에 뉴클레오타이드 값을 할당(예컨대, 돌연변이 호출)하는 단계로서:
- [0572] (1) 상기 샘플의 라이브러리에 있는 핵산 또는 상기 라이브러리에 있는 라이브러리 캐치의 핵산 중에서 상기 샘플 내 세포의 1% 미만에 존재하는 변이체를 갖는 적어도 1, 2, 3, 4 또는 5개의 뉴클레오타이드 위치;
- [0573] (2) 상기 샘플의 라이브러리에 있는 핵산 또는 상기 라이브러리에 있는 라이브러리 캐치의 핵산 중에서 상기 샘플 내 세포의 1-2%에 존재하는 변이체를 갖는 적어도 1, 2, 3, 4 또는 5개의 뉴클레오타이드 위치;
- [0574] (3) 상기 샘플의 라이브러리에 있는 핵산 또는 상기 라이브러리에 있는 라이브러리 캐치의 핵산 중에서 상기 샘플 내 세포의 2% 초과 및 3% 이하에 존재하는 변이체를 갖는 적어도 1, 2, 3, 4 또는 5개의 뉴클레오타이드 위치;
- [0575] (4) 상기 샘플의 라이브러리에 있는 핵산 또는 상기 라이브러리에 있는 라이브러리 캐치의 핵산 중에서 상기 샘플 내 세포의 3% 초과 및 4% 이하에 존재하는 변이체를 갖는 적어도 1, 2, 3, 4 또는 5개의 뉴클레오타이드 위치;
- [0576] (5) 상기 샘플의 라이브러리에 있는 핵산 또는 상기 라이브러리에 있는 라이브러리 캐치의 핵산 중에서 상기 샘플 내 세포의 4% 초과 및 5% 이하에 존재하는 변이체를 갖는 적어도 1, 2, 3, 4 또는 5개의 뉴클레오타이드 위치;

- [0577] (6) 상기 샘플의 라이브러리에 있는 핵산 또는 상기 라이브러리에 있는 라이브러리 캐치의 핵산 중에서 상기 샘플 내 세포의 5% 초과 및 10% 이하에 존재하는 변이체를 갖는 적어도 1, 2, 3, 4 또는 5개의 뉴클레오티드 위치;
- [0578] (7) 상기 샘플의 라이브러리에 있는 핵산 또는 상기 라이브러리에 있는 라이브러리 캐치의 핵산 중에서 상기 샘플 내 세포의 10% 초과 및 20% 이하에 존재하는 변이체를 갖는 적어도 1, 2, 3, 4 또는 5개의 뉴클레오티드 위치;
- [0579] (8) 상기 샘플의 라이브러리에 있는 핵산 또는 상기 라이브러리에 있는 라이브러리 캐치의 핵산 중에서 상기 샘플 내 세포의 20% 초과 및 40% 이하에 존재하는 변이체를 갖는 적어도 1, 2, 3, 4 또는 5개의 뉴클레오티드 위치;
- [0580] (9) 상기 샘플의 라이브러리에 있는 핵산 또는 상기 라이브러리에 있는 라이브러리 캐치의 핵산 중에서 상기 샘플 내 세포의 40% 초과 및 50% 이하에 존재하는 변이체를 갖는 적어도 1, 2, 3, 4 또는 5개의 뉴클레오티드 위치; 또는
- [0581] (10) 상기 샘플의 라이브러리에 있는 핵산 또는 상기 라이브러리에 있는 라이브러리 캐치의 핵산 중에서 상기 샘플 내 세포의 50% 초과 및 100% 이하에 존재하는 변이체를 갖는 적어도 1, 2, 3, 4 또는 5개의 뉴클레오티드 위치 중 하나 이상 또는 모두에 뉴클레오티드 값을 할당하는 단계로서,
- [0582] 선택적으로, 각각의 할당은 고유한 제1 및/또는 제2 값(예컨대, 인용된 범위(예컨대, 1% 미만의 (1)의 범위)에 서 다른 할당과 대조적으로 고유하거나 다른 나열된 범위의 하나 이상 또는 모두에서 결정을 위한 제1 및/또는 제2 값과 대조적으로 고유한)에 기초하는 단계; 또는
- [0583] (ix) X개의 뉴클레오티드 위치 각각에 뉴클레오티드 값을 할당(예컨대, 돌연변이 호출)하고, 각 뉴클레오티드 위치는 독립적으로 다른 X-1 뉴클레오티드 위치에서 변이체에 대한 가능성과 비교하여 고유한 (상기 샘플의 DNA 내에 존재하는 변이체의)가능성을 갖고, X는 1, 2, 3, 5, 10, 20, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 또는 1,000 이상이고, 각각의 할당은 고유한(다른 할당과 대조적으로) 제1 및/또는 제2 값을 기반으로 하는 단계를 포함할 수 있다.
- [0584] 일부 구현예들에서, "임계값"은 리드를 평가하고 상기 리드로부터 뉴클레오티드 위치에 대한 값을 선택하는 데 사용되며, 예컨대 유전자의 특정 위치에서 돌연변이를 호출한다. 일부 구현예들에서, 다수의 대상체 간격 각각에 대한 임계값은 맞춤화되거나 미세 조정된다. 맞춤화 또는 조정은 본원에 기재된 하나 이상의 인자, 예컨대 샘플 내 암의 유형, 서열분석될 대상체 간격(하위 유전체 간격 또는 발현된 하위 유전체 간격)이 위치한 유전자, 또는 서열분석될 변이체에 기초할 수 있다. 이는 서열분석될 다수의 대상체 간격의 각각에 대해 미세하게 조정된 호출을 제공한다. 일부 구현예들에서, 상기 방법은 비교적 많은 수의 다양한 하위 유전체 간격이 분석될 때 특히 효과적이다.
- [0585] 따라서, 다른 구현예에서, 상기 방법은 하기의 돌연변이 호출 방법을 포함한다:
- [0586] 상기 X개의 대상체 간격 각각에 대해 임계값을 수득하는 단계로서, 상기 수득한 X 임계값의 각각은 다른 X-1 임계값과 비교하여 고유하며, 이에 따라 X개의 고유 임계값을 제공하는 단계;
- [0587] 상기 X 대상체 간격의 각각에 대해, 뉴클레오티드 위치에서 뉴클레오티드 값을 갖는 리드의 수의 함수인 관찰된 값을 고유 임계값과 비교함으로써, 상기 X 대상체 간격의 각각에 고유 임계값을 적용하는 단계; 및
- [0588] 선택적으로, 상기 비교 결과에 응답하여, 뉴클레오티드 위치에 뉴클레오티드 값을 할당하는 단계로서,
- [0589] X는 2 이상인 단계를 포함하는 방법이다.
- [0590] 한 구현예에서, 상기 방법은 뉴클레오티드 값을 적어도 2, 3, 5, 10, 20, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 또는 1,000개의 뉴클레오티드 위치에 할당하는 단계로서, 각각 독립적으로 0.5, 0.4, 0.25, 0.15, 0.10, 0.05, 0.04, 0.03, 0.02, 또는 0.01 미만인 확률의 함수인 제1 값을 갖는 단계를 포함한다.
- [0591] 한 구현예에서, 상기 방법은 적어도 X개의 뉴클레오티드 위치의 각각에 뉴클레오티드 값을 할당하는 단계로서, 각각 독립적으로 다른 X-1 제1 값과 비교하여 고유한 제1 값을 갖고, 상기 X개의 제1 값의 각각은 0.5, 0.4, 0.25, 0.15, 0.10, 0.05, 0.04, 0.03, 0.02 또는 0.01보다 작은 확률의 함수이고, X는 1, 2, 3, 5, 10, 20,

40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 또는 1,000 이상인 단계를 포함한다.

- [0592] 한 구현예에서, 적어도 20, 40, 60, 80, 100, 120, 140, 160 또는 180, 200, 300, 400, 또는 500개의 유전자, 예컨대 표 2A-5B로부터의 유전자 내의 뉴클레오티드 위치에 뉴클레오티드 값이 할당된다. 한 구현예에서, 고유한 제1 및/또는 제2 값은 분석되는 상기 유전자의 적어도 10, 20, 30, 40, 또는 50% 각각에서 대상체 간격에 적용된다.
- [0593] 상기 방법의 구현예들은, 예컨대 하기의 구현예들로부터 알 수 있는 바와 같이 비교적 많은 수의 대상체 간격에 대한 임계값이 최적화되는 경우에 적용할 수 있다.
- [0594] 한 구현예에서, 고유한 임계값은 적어도 3, 5, 10, 20, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 또는 1,000개의 상이한 유전자 각각에서 대상체 간격, 예컨대 하위 유전체 간격 또는 발현된 하위 유전체 간격에 적용된다.
- [0595] 한 구현예에서, 적어도 20, 40, 60, 80, 100, 120, 140, 160 또는 180, 200, 300, 400, 또는 500개의 유전자, 예컨대 표 2A-5B로부터의 유전자 내의 뉴클레오티드 위치에 뉴클레오티드 값이 할당된다. 한 구현예에서, 고유한 임계값은 분석되는 상기 유전자의 적어도 10, 20, 30, 40, 또는 50% 각각에서 하위 유전체 간격에 적용된다.
- [0596] 한 구현예에서, 표 2A-5B로부터의 적어도 5, 10, 20, 30, 또는 40개의 유전자 내의 뉴클레오티드 위치에 뉴클레오티드 값이 할당된다. 한 구현예에서, 고유한 임계값은 분석되는 상기 유전자의 적어도 10, 20, 30, 40, 또는 50% 각각에서 하위 유전체 간격(예컨대, 하위 유전체 간격 또는 발현된 하위 유전체 간격)에 적용된다.
- [0597] 해당 모듈의 요소는 중앙 분석 방법에 포함될 수 있다. 구현예들에서, "돌연변이 호출" 섹션의 정렬 방법은 본원의 "정렬" 섹션의 정렬 방법 및/또는 본원의 "표적 포획 시약" 섹션의 및/또는 "표적 포획 시약의 설계 및 구성"의 섹션, 및 본원의 "표적 포획 시약의 경쟁"의 표적 포획 시약과 조합된다. 상기 방법은 본원의 "유전자 선택" 섹션의 대상체 간격 세트 및/또는 본원의 "대상체" 섹션의 대상체로부터의 "샘플" 섹션의 샘플에 적용될 수 있다.
- [0598] 염기 호출은 서열분석 장치의 원시 출력을 지칭한다. 돌연변이 호출은 서열분석되는 뉴클레오티드 위치에 대한 뉴클레오티드 값, 예컨대 A, G, T 또는 C를 선택하는 공정을 지칭한다. 일반적으로, 위치에 대한 서열분석 리드(또는 염기 호출)는 하나 이상의 값을 제공하고, 예컨대 일부 리드는 T를 제공하고 일부는 G를 제공한다. 돌연변이 호출은 뉴클레오티드 값을 할당하는 공정이고, 예컨대 상기 서열에 해당 값들 중 하나를 할당한다. "돌연변이" 호출이라고 부르는 것은, 이는 임의의 뉴클레오티드 위치, 예컨대 돌연변이 대립유전자, 야생형 대립유전자, 돌연변이 또는 야생형으로 특성화되지 않은 대립유전자에 상응하는 위치 또는 변동성을 특징으로 하지 않는 위치에 뉴클레오티드 값을 할당하는 데 적용될 수 있다. 돌연변이 호출을 위한 방법은 다음 중 하나 이상을 포함할 수 있다: 참조 서열의 각 위치에 있는 정보를 기반으로 독립적 호출 만들기(예컨대, 서열 리드의 검사; 염기 호출 및 품질 점수의 검사; 관찰된 염기의 확률 및 잠재적인 유전형에 대한 품질 점수의 계산; 및 유전형 할당(예컨대, 베이스(Bayes)의 규칙 사용), 가양성의 제거(예컨대, 예상보다 훨씬 낮거나 높은 리드 정도를 가진 SNP를 거부하기 위해 정도 임계값을 사용, 작은 삽입-결실로 인한 가양성을 제거하기 위한 부분 재정렬); 및 호출을 정제하기 위해 연관 불균형(LD)/대체 기반 분석을 실시하는 단계를 포함한다.
- [0599] 특정 유전자형 및 위치와 관련된 유전자형 가능성을 계산하는 방정식이, 예컨대 Li H. and Durbin R. *Bioinformatics*, 2010; 26(5): 589-95에 기재되어 있다. 특정 암 유형의 특정 돌연변이에 대한 사전 예측은 해당 암 유형의 샘플을 평가할 때 사용할 수 있다. 상기 가능성은 암 돌연변이의 공개 데이터베이스, 예컨대 암의 체세포 돌연변이 목록(Catalogue of Somatic Mutation in Cancer, COSMIC), 인간 유전자 돌연변이 데이터베이스(Human Gene Mutation Database, HGMD), SNP 컨소시엄, 유방암 돌연변이 데이터베이스(Breast Cancer Mutation Data Base, BIC) 및 유방암 유전자 데이터베이스(Breast Cancer Gene Database, BCGD)에서 유래될 수 있다.
- [0600] LD/대체 기반 분석의 예는 Browning B.L. and Yu Z. *Am. J. Hum. Genet.* 2009, 85(6):847-61에 기재되어 있다. 낮은 적용 범위의 SNP 호출 방법의 예는, 예컨대 Li Y. *et al.*, *Annu. Rev. Genomics Hum. Genet.* 2009, 10:387-406에 기재되어 있다.
- [0601] 정렬 후, 호출 방법, 예를 들어 베이지안 돌연변이 호출 방법을 사용하여 치환 검출을 실시할 수 있다. 이는 각 대상체 간격의 각 염기에 적용되고, 예컨대 대체 대립유전자의 존재가 관찰되는 평가할 유전자의 엑손이다. 상기 방법은 돌연변이 존재량의 리드 데이터를 관찰할 확률과 염기 호출 오류만 존재하는 상태에서 리드 데이터를 관찰할 확률을 비교한다. 상기 비교가 돌연변이의 존재를 충분히 강력하게 뒷받침하는 경우 돌연변이를 호출할

수 있다.

- [0602] 암 DNA 분석을 위해 50% 또는 100%의 빈도에서 제한된 편차를 해결하는 방법이 개발되었다. (예컨대, SNVMix-Bioinformatics. 2010 March 15; 26(6): 730-736.) 하지만, 본원에 개시된 방법은 샘플 DNA의 1% 내지 100%, 특히 50% 미만의 수준에서 돌연변이 대립유전자의 존재 가능성을 고려할 수 있다. 상기 접근법은 천연(다중 클론) 종양 DNA의 저순도 FFPE 샘플에서 돌연변이를 검출하는 데 특히 중요하다.
- [0603] 베이지안 돌연변이 검출 접근법의 장점은 돌연변이의 존재 확률과 염기 호출 오류 단독의 확률의 비교가 해당 부위에 돌연변이의 존재에 대한 사전 예측에 의해 가중될 수 있다는 것이다. 특정 암 유형에 대해 자주 돌연변이되는 부위에서 대체 대립유전자의 일부 리드값이 관찰되면 돌연변이의 증거량이 일반적인 임계값을 충족하지 않더라도 돌연변이의 존재를 자신 있게 호출할 수 있다. 그런 다음, 상기 유연성을 사용하여 더 희귀한 돌연변이/순도가 낮은 샘플에 대한 검출 감도를 높이거나 리드 범위의 감소에 대한 시험을 더욱 강력하게 만들 수 있다. 유전체의 무작위 염기쌍의 돌연변이될 가능성은  $\sim 1e-6$ 이다. 전형적인 다유전자 암 유전체 패널의 많은 부위에서 특정 돌연변이의 가능성은 훨씬 더 높을 수 있다. 상기 가능성은 암 돌연변이의 공개 데이터베이스(예컨대, COSMIC)에서 유래될 수 있다. 삽입-결실 호출은 일반적으로 관련 신뢰 점수 또는 통계적 증거 메트릭을 포함하여 삽입 또는 결실에 의해 참조 서열과 상이한 염기서열 분석 데이터에서 염기를 찾는 공정이다.
- [0604] 삽입-결실 호출 방법은 후보 삽입-결실을 식별하고, 국소 재배열을 통한 유전자형 가능성을 계산하고, LD 기반 유전자형 추론 및 호출을 실시하는 단계를 포함할 수 있다. 일반적으로, 베이지안 접근 방식을 사용하여 잠재적인 삽입-결실 후보를 수득한 다음, 상기 후보를 베이지안 프레임워크에서 참조 서열과 함께 시험한다.
- [0605] 후보 삽입-결실을 생성하기 위한 알고리즘이, 예컨대 McKenna A. *et al.*, *Genome Res.* 2010; 20(9):1297-303; Ye K. *et al.*, *Bioinformatics*, 2009; 25(21):2865-71; Lunter G. and Goodson M. *Genome Res.* 2011; 21(6):936-9; and Li H. *et al.*, *Bioinformatics* 2009, *Bioinformatics* 25(16):2078-9에 기재되어 있다.
- [0606] 삽입-결실 호출 및 개별 수준 유전자형 가능성을 생성하는 방법에는, 예컨대 딥델(Dindel) 알고리즘(Albers C.A. *et al.*, *Genome Res.* 2011;21(6):961-73)이 기재되어 있다. 예를 들어, 베이지안 EM 알고리즘은 리드를 분석하고, 초기 삽입-결실을 호출하고, 각 후보 삽입-결실에 대한 유전자형 가능성을 생성한 후, 예컨대 QCALL (Le S.Q. and Durbin R. *Genome Res.* 2011;21(6):952-60)을 사용하여 유전자형을 대체하는데 사용될 수 있다. 삽입-결실 관찰에 대한 사전 기대치와 같은 매개변수는 삽입-결실의 크기 또는 위치에 따라 조정(예컨대, 증가 또는 감소)될 수 있다.
- [0607] 한 구현예에서, 상기 방법에서 만들어진 돌연변이 호출의 적어도 10, 20, 30, 40, 50, 60, 70, 80, 또는 90%는 본원에 기재된 유전자 또는 유전자 산물, 예컨대 표 2A-5B의 유전자 또는 유전자 산물로부터의 대상체 간격에 대한 것이다. 한 구현예에서, 본원에 기재된 고유한 임계값의 적어도 10, 20, 30, 40, 50, 60, 70, 80, 또는 90%는 본원에 기재된 유전자 또는 유전자 산물, 예컨대 표 2A-5B의 유전자 또는 유전자 산물로부터의 대상체 간격에 대한 것이다. 한 구현예에서, 구석이 달리거나 제3자에게 보고된 돌연변이 호출의 적어도 10, 20, 30, 40, 50, 60, 70, 80, 또는 90%는 본원에 기재된 유전자 또는 유전자 산물, 예컨대 표 2A-5B의 유전자 또는 유전자 산물로부터의 대상체 간격에 대한 것이다.
- [0608] 구현예에서, 뉴클레오타이드 위치에 대한 할당된 값은 선택적으로 설명 주석과 함께 제3자에게 전송된다. 한 구현예에서, 뉴클레오타이드 위치에 대해 할당된 값은 제3자에게 전송되지 않는다. 한 구현예에서, 복수의 뉴클레오타이드 위치에 대한 할당된 값은 선택적으로 설명 주석과 함께 제3자에게 전송되고, 제2의 복수의 뉴클레오타이드 위치에 대한 할당된 값은 제3자에게 전송되지 않는다.
- [0609] 한 구현예에서, 상기 방법은, 예컨대 바코드 디콘볼루션에 의해 대상체에 하나 이상의 리드를 할당하는 단계를 포함한다.
- [0610] 한 구현예에서, 상기 방법은, 예컨대 바코드 디콘볼루션에 의해 하나 이상의 리드를 종양 리드 또는 대조군 리드로서 할당하는 단계를 포함한다. 한 구현예에서, 상기 방법은, 예컨대 참조 서열과의 정렬에 의해 상기 하나 이상의 리드 각각을 맵핑하는 단계를 포함한다. 한 구현예에서, 상기 방법은 호출된 돌연변이를 기억하는 단계를 포함한다.
- [0611] 한 구현예에서, 상기 방법은 호출된 돌연변이에 주석을 다는 단계, 예컨대 돌연변이 구조, 예컨대 과오 돌연변이 또는 기능, 예컨대 질병 표현형의 표시로 호출된 돌연변이에 주석을 다는 단계를 포함한다. 한 구현예에서, 상기 방법은 종양 및 대조군 핵산에 대한 뉴클레오타이드 서열 리드를 수득하는 단계를 포함한다. 한 구현예에서,

상기 방법은 각각의 대상체 간격(예컨대, 하위 유전체 간격, 발현된 하위 유전체 간격, 또는 둘 모두)에 대한 뉴클레오티드 값, 예컨대 변이체, 예컨대 돌연변이를, 예컨대 베이지안 호출 방법 또는 비베이지안 호출 방법을 이용하여 호출하는 단계를 포함한다. 한 구현예에서, 상기 방법은 적어도 하나의 SNP를 포함하는 복수의 리드를 평가하는 단계를 포함한다. 한 구현예에서, 상기 방법은 샘플 및/또는 대조군 리드에서 SNP 대립유전자 비율을 결정하는 단계를 포함한다.

[0612] 일부 구현예들에서, 상기 방법은 표적화된 하위 유전체 영역에 대한 서열분석/정렬 인공물의 데이터베이스를 구축하는 단계를 추가로 포함한다. 한 구현예에서, 상기 데이터베이스는 가짜 돌연변이 호출을 걸러내고 특이성을 개선하기 위해 사용될 수 있다. 한 구현예에서, 상기 데이터베이스는 관련되지 않은 샘플 또는 세포주를 서열분석하고 이의 정상 샘플들 중 1개 이상에서만 무작위 서열분석 오류로 인해 예상보다 더 자주 나타나는 비참조 대립유전자 이벤트를 기록함으로써 구축된다. 상기 접근 방식은 생식계열 변이를 인공물로 분류할 수 있지만 체세포 돌연변이와 관련된 방법에서는 허용가능하다. 생식계열 변이를 인공물로 잘못 분류하는 것은 상기 데이터베이스를 공지된 생식계열 변이(일반적인 변이 제거) 및 한 개체에게만 나타나는 인공물(희귀 변이 제거)에 대해 필터링함으로써 원하는 경우 개선될 수 있다.

[0613] 돌연변이 호출의 최적화는, 예컨대 국제특허출원공보 제WO 2012/092426호에 기재된 바와 같이 당업계에 기재되어 있다.

[0614] SGZ 알고리즘

[0615] 다양한 유형의 변경, 예컨대 체세포 변경 및 생식계열 돌연변이는 본원에 기재된 방법(예컨대, 서열분석, 정렬 또는 돌연변이 호출 방법)에 의해 검출할 수 있다. 특정 구현예들에서, 생식계열 돌연변이는 SGZ(체세포-생식계열-접합) 알고리즘을 사용하는 방법에 의해 추가로 확인된다. 예를 들어, 미국특허 제9,792,403호 및 Sun et al., *A computational approach to distinguish somatic vs. germline origin of genomic alteration from deep sequencing of cancer specimens without a matched normal*, PLOS Computational Biology (Feb. 2018)를 참조한다.

[0616] 임상 실습에서, 일치하는 정상 대조군은 일반적으로 수득되지 않는다. 일부 구현예들에서, 잘 특성화된 유전체 변경은 해석을 위해 정상 조직을 필요로 하지 않지만, 일치하는 정상 대조군이 없는 경우 적어도 일부 변경은 생식계열인지 체세포인지의 여부에 대해 공지되지 않을 것이다. SGZ는 암 표본의 차세대 서열분석에서 확인된 변이체의 체세포 대 생식계열 기원 및 동형 접합 대 이형 접합 또는 하위 클론 상태를 예측하기 위한 계산 방법이다.

[0617] SGZ 방법은 일치하는 정상 대조군이 필요하지 않으므로, 임상 환경에서 광범위하게 적용할 수 있다. SGZ는 종양 함량, 종양 배수성 및 국소 복제 수를 고려하여 변경의 대립 유전자 빈도(AF)를 모델링하여 식별된 각 변경의 체세포 대 생식계열 상태를 예측한다. 예측의 정확도는 암 관련 유전자 및 유전체 전체의 단일 염기 다형성(SNP)을 포함하는 높은 정도의 서열분석을 통해 달성할 수 있는 서열분석 및 복제 수 모델 적합도의 정도에 따라 달라진다. 호출은 SNP AF의 리드 정도 및 국소 변동성을 기반으로 한 통계를 사용하여 이루어진다.

[0618] 일부 구현예들에서, 상기 방법은 대상체, 예컨대 인간, 예컨대 암 환자로부터의 조직(예컨대, 종양) 또는 샘플에서 변이체, 예컨대 돌연변이를 특성화하는 단계를 추가로 포함하며, 상기 방법은:

[0619] a) 하기를 수득하는 단계로서:

[0620] i) 복수의 선택된 대상체 간격의 각각에 대하여, 예컨대 엑손, 즉 상기 선택된 대상체 간격에서 정규화된 서열 범위에 대한 값을 포함하는 서열 범위 입력(SCI);

[0621] ii) 복수의 선택된 생식계열 SNP 각각에 대하여, 종양 또는 샘플에서 대립유전자 빈도에 대한 값을 포함하는 SNP 대립유전자 빈도 입력(SAFI);

[0622] iii) 종양 또는 샘플에서 상기 변이체, 예컨대 돌연변이에 대한 대립유전자 빈도를 포함하는 변이 대립유전자 빈도 입력(VAFI)을 수득하는 단계;

[0623] b) SCI 및 SAFI의 함수로서 하기에 대한 값을 수득하는 단계로서:

[0624] 복수의 유전체 분절 각각에 대한 C로서, C는 유전체 분절

[0625] 총 복제 수이고;

- [0626] 복수의 유전체 분절 각각에 대한  $M$ ,  $M$ 은 유전체 분절 소수 대립유전자 복제 수이며;
- [0627]  $p$ , 여기서  $p$ 는 샘플 순도인, 단계; 및
- [0628] c) 하기 중 하나 또는 둘 모두를 수득하는 단계로서:
- [0629] i) 변이체 유형에 대한 값, 예컨대 돌연변이 유형, 예컨대 상기 변이체, 예컨대 체세포인 돌연변이, 하위클론 체세포 변이체, 생식계열 또는 구별불가능한 돌연변이를 나타내고 VAFI,  $p$ ,  $C$  및  $M$ 의 함수인  $g$ ;
- [0630] ii)  $C$  및  $M$ 의 함수로서 종양 또는 샘플에서 변이체, 예컨대 돌연변이의 집합성의 표시를 수득하는 단계를 포함한다.
- [0631] 한 구현예에서, 상기 분석은 상기 대상체로부터 비종양 조직을 분석할 필요 없이 실시될 수 있다. 한 구현예에서, 상기 분석은 상기 대상체로부터의 비종양 조직을 분석하지 않고 실시되며, 예컨대 동일한 대상체로부터의 비종양 조직은 서열분석되지 않는다.
- [0632] 한 구현예에서, 상기 SCI는, 예컨대 샘플로부터의 대상체 간격에 대한 리드 수의 함수, 예컨대 비율의 로그, 및 대조군, 및 대조군, 예컨대 공정 일치된 대조군에 대한 수 또는 리드를 포함한다. 한 구현예에서, 상기 SCI는 적어도 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 1,000, 2,000, 3,000, 5,000, 6,000, 7,000, 8,000, 9,000 또는 10,000개의 대상체 간격, 예컨대 엑손에 대한 값, 예컨대 로그  $r$  값을 포함한다. 한 구현예에서, 상기 SCI는 100개 이상의 대상체 간격, 예컨대 엑손에 대한 값, 예컨대 로그  $r$  값을 포함한다. 한 구현예에서, 상기 SCI는 1,000 내지 10,000, 2,000 내지 9,000, 3,000 내지 8,000, 3,000 내지 7,000, 3,000 내지 6,000, 또는 4,000 내지 5,000개의 대상체 간격, 예컨대 엑손에 대한 값, 예컨대 로그  $r$  값을 포함한다. 한 구현예에서, 상기 SCI는 적어도 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 1,000, 2,000, 3,000, 또는 4,000개의 유전자로부터의 대상체 간격, 예컨대 엑손에 대한 값, 예컨대 로그  $r$  값을 포함한다.
- [0633] 한 구현예에서, 상기 SCI에 포함된 값들 중 적어도 하나, 복수, 또는 실질적으로 모두는 GC 함량과의 상관관계에 대해 보정된다.
- [0634] 한 구현예에서, 상기 샘플로부터의 대상체 간격, 예컨대 엑손은 적어도 10, 20, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700을 갖는다. , 800, 900 또는 1,000개의 리드를 갖는다. 한 구현예에서, 상기 샘플로부터의 복수, 예컨대 적어도 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 또는 10,000개의 대상체 간격, 예컨대 엑손은 다수의 리드를 갖는다. 한 구현예에서, 리드의 수는 적어도 10, 20, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 또는 1,000개이다. 한 구현예에서, 복수의 생식계열 SNP는 적어도 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 1,000, 2,000, 3,000, 4,000, 5000, 6000, 7000, 8000, 9000, 10,000, 또는 15,000개의 생식계열 SNP를 포함한다.
- [0635] 한 구현예에서, 복수의 생식계열 SNP는 적어도 100개의 생식계열 SNP를 포함한다. 한 구현예에서, 복수의 생식계열 SNP는 500 내지 5,000, 1,000 내지 4,000, 또는 2,000 내지 3,000개의 생식계열 SNP를 포함한다. 한 구현예에서, 상기 대립유전자 빈도는 소수 대립유전자 빈도이다. 한 구현예에서, 상기 대립유전자 빈도는 대안적 대립유전자, 예컨대 인간 유전체 참조 데이터베이스에서 표준 대립유전자 이외의 대립유전자이다.
- [0636] 한 구현예에서, 상기 방법은 상기 샘플에서 복수의 변이체, 예컨대 돌연변이체를 특성화하는 단계를 포함한다. 한 구현예에서, 상기 방법은 적어도 2, 3, 4, 5, 6, 7, 8 9, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 또는 500개의 변이체, 예컨대 돌연변이체를 특성화하는 단계를 포함한다. 한 구현예에서, 상기 방법은 적어도 2, 3, 4, 5, 6, 7, 8 9, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450 또는 500개의 상이한 유전자에서 변이체, 예컨대 돌연변이체를 특성화하는 단계를 포함한다.
- [0637] 한 구현예에서, 상기 방법은 적어도 2, 3, 4, 5, 6, 7, 8 9, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 또는 500개의 변이체, 예컨대 돌연변이체에 대한 VAFI를 수득하는 단계를 포함한다. 한 구현예에서, 상기 방법은 적어도 2, 3, 4, 5, 6, 7, 8 9, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 또는 500개의 변이체, 예컨대 돌연변이체에 대한 단계 a), 단계 b) 및 단계 c) 중 하나, 둘 또는 모두를 실시하는 단계를 포함한다. 한 구현예에서,  $C$ ,  $M$  및  $p$ 의 값은 SCI 및 SAFI 중 하나 또는 둘 모두에 유전체 전체 복제 수 모델을 피팅하거나, 피팅을 갖거나 또는 피팅함으로써 수득할 수 있다. 한 구현예에서,  $C$ ,  $M$ , 및  $p$ 의 값은 SCI 및 SAFI의 복수의 유전체 전체 복제 수 모델 입력에 맞는다. 한 구현예에서, 유전체 분절은 복수의 대상체 간격, 예컨대 엑손, 예컨대 SCI 값이 할당된 대상체 간격을 포함한다.

- [0638] 한 구현예에서, 유전체 분절은 적어도 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 225, 250, 275, 300, 400, 또는 500개의 대상체 간격, 예컨대 엑손을 포함한다. 한 구현예에서, 유전체 분절은 10 내지 1,000, 20 내지 900, 30 내지 700, 40 내지 600, 50 내지 500, 60 내지 400, 70 내지 300, 80 내지 200, 80 내지 150, 또는 80 내지 120, 90 내지 110, 또는 약 100개의 대상체 간격, 예컨대 엑손을 포함한다. 한 구현예에서, 유전체 분절은 100 내지 10,000, 100 내지 5,000, 100 내지 4,000, 100 내지 3,000, 100 내지 2,000, 또는 100 내지 1,000개의 대상체 간격, 예컨대 엑손을 포함한다. 한 구현예에서, 유전체 분절은 SAFI 값이 할당된 10 내지 1,000, 20 내지 900, 30 내지 700, 40 내지 600, 50 내지 500, 60 내지 400, 70 내지 300, 80 내지 200, 80 내지 150, 또는 80 내지 120, 90 내지 110, 또는 약 100개의 유전체 SNP를 포함한다. 한 구현예에서, 유전체 분절은 SAFI 값이 할당된 100 내지 10,000, 100 내지 5,000, 100 내지 4,000, 100 내지 3,000, 100 내지 2,000, 또는 100 내지 1,000개의 유전체 SNP를 포함한다.
- [0639] 한 구현예에서, 복수의 유전체 분절 각각은 하기 중 하나 또는 둘 모두를 갖는 것을 특징으로 한다:
- [0640] 미리 선택된 양만큼 차이가 나지 않는 정규화된 서열 적용 범위, 예컨대 로그 r의 측정값, 예컨대 유전체 분절의 경계 내에서 대상체 간격, 예컨대 엑손에 대한  $\log_2 r$  값은 참조값 만큼 차이가 나지 않거나 실질적으로 일정하다; 그리고
- [0641] 미리 선택된 양만큼 차이가 나지 않는 생식계열 SNP에 대한 SNP 대립유전자 빈도, 예컨대 대상체 간격, 예컨대 엑손에 대한 생식계열 SNP 대립유전자 빈도값은 유전체 분절의 경계 내에서 참조값 만큼 차이가 나지 않거나 실질적으로 일정하다; 그리고
- [0642] 한 구현예에서, 유전체 분절에 포함되거나 이를 형성하기 위해 조합되는 대상체 간격, 예컨대 엑손의 수는 유전체 분절의 수의 적어도 2, 5, 10, 15, 20, 50, 또는 100배이다. 한 구현예에서, 대상체 간격, 예를 들어 엑손의 수는 유전체 분절의 수의 적어도 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 또는 15배이다.
- [0643] 한 구현예에서, 유전체 분절에 대한 경계가 제공된다. 한 구현예에서, 상기 방법은 대상체 간격, 예컨대 엑손에 대한 서열을 유전자 분절로 조립하는 단계를 포함한다.
- [0644] 한 구현예에서, 상기 방법은 본원에 기재된 방법, 예컨대 원형 이진 분할(CBS), HMM 기반 방법, 웨이블릿 기반 방법, 또는 염색체에 따른 클러스터 방법을 포함하는 방법으로 대상체 간격에 대한 서열을 조립하는 단계를 포함한다.
- [0645] 한 구현예에서, 상기 SCI에 대한 유전체 전체 복제 수 모델을 피팅하는 것은 하기 방정식을 사용하는 것을 포함한다:
- [0646] 
$$\log \mu_i^j = \log \frac{PC_i + 2(1-p)}{2 \psi + 2(1-p)}, \psi \text{는 중앙 배수성이다.}$$
- [0647] 한 구현예에서,  $\psi = (\sum_i I_i C_i) / \sum_i I_i$ 에 따라  $I_i$ 는 유전체 분절의 길이이다.
- [0648] 한 구현예에서, 상기 SAFI에 대한 유전체 전체 복제 수 모델을 피팅하는 것은 하기 방정식을 사용하는 것을 포함한다:
- [0649] 
$$AF = \frac{pM + 1(1-p)}{pC + 2(1-p)}, AF \text{는 대립유전자 빈도이다.}$$
- [0650] 한 구현예에서, 상기 피팅은 깃스 샘플링을 사용하는 것을 포함한다. 한 구현예에서, 피팅은, 예컨대 마르코브 사슬 몬테 카를로(Markov chain Monte Carlo, MCMC) 알고리즘, 예컨대 대립유전자 특이적 복제수 분석(Allele-Specific Copy Number Analysis of Tumors, ASCAT), 온코(Onco)SNP, 또는 암의 통합 복제수 예측(Predicting Integral Copy Numbers In Cancer, PICNIC)을 사용하는 것을 포함한다. 한 구현예에서, 피팅은 메트로폴리스-헤스팅스(Metropolis-Hastings) MCMC를 사용하는 것을 포함한다. 한 구현예에서, 피팅은 비베이지안 접근법, 예컨대 최소 제곱 피팅을 사용하는 빈도주의적 접근법을 사용하는 것을 포함한다.
- [0651] 한 구현예에서, g는 체세포/생식계열 상태의 모델에 대한 VAFI, p, C 및 M에 대한 값의 적합성을 결정함으로써 결정된다. 한 구현예에서, 상기 방법은 상기 변이체, 예컨대 돌연변이에 대한 이형접합성의 표시를 수득하는 단계를 포함한다. 한 구현예에서, 샘플 순도(p)는 전체 순도이며, 예컨대 모든 유전체 분절에 대해 동일하다.
- [0652] 한 구현예에서, g의 값은 하기에 의해 수득된다:

$AF = \frac{pM+g(1-p)}{pC+2(1-p)}$ , AF는 대립유전자 빈도이다.

[0653]

한 구현예에서, 0에 가까운, 예컨대 0과 크게 다르지 않은  $g$  값은 변이체가 체세포 변이체임을 나타낸다. 한 구현예에서, 0 또는 0에 가까운  $g$  값, 예컨대 0으로부터의 거리 내, 예컨대 0.4 미만의  $g$  값은 변이체가 체세포 변이체임을 나타낸다. 한 구현예에서, 1에 가까운, 예컨대 1과 크게 다르지 않은  $g$  값은 변이체가 생식계열 변이체임을 나타낸다. 한 구현예에서, 1 또는 1에 가까운  $g$  값, 예컨대 1으로부터의 거리 내, 예컨대 0.6 초과의  $g$  값은 변이체가 생식세포 변이체임을 나타낸다. 한 구현예에서,  $g$ 의 값은 1보다 작지만 0보다 크며, 예컨대 양으로 1보다 작고 양으로 0보다 큰 경우, 예컨대  $g$ 가 0.4 내지 0.6 사이인 경우, 이는 구별할 수 없는 결과를 나타낸다.

[0655]

한 구현예에서, 0보다 상당히 작은  $g$  값은 하위클론 체세포 변이체를 나타낸다.

[0656]

한 구현예에서,  $g$ 의 값은 하기에 의해 수득된다:

$$AF = \frac{pM'+g(1-p)}{pC+2(1-p)}$$

[0657]

, 여기서 AF는 대립유전자 빈도이고  $M' = C - M$ (예컨대,  $M$ 이 비소수 대립유전자 빈도인 경우), 예컨대  $g=1$ 인 경우 변이체는 생식계열 다형성이고  $g=0$ 인 경우 변이체는 체세포 돌연변이이다.

[0658]

한 구현예에서, 체세포/생식계열 상태는, 예컨대 샘플 순도가 약 40% 미만, 예컨대 약 10% 내지 30%, 예컨대 약 10% 내지 20%, 또는 약 20% 내지 약 30%일 때 결정된다.

[0659]

한 구현예에서, 0과 동일하고  $C$ 와 동일하지 않은  $M$ 의 값은 변이체, 예컨대 돌연변이의 부재, 예컨대 종양 내 존재하지 않음을 나타내고;  $C$ 와 동일한  $M$ 의 0이 아닌 값은 변이체, 예컨대 돌연변이의 동형접합, 예컨대 이형접합 손실(LOH)을 나타내고; 0과 같은  $M$ 의 값은  $C$ 와 동일한  $M$ 의 값은 변이체, 예컨대 돌연변이의 동형접합 결실, 예컨대 종양 내 존재하지 않음을 나타내며;  $C$ 와 같지 않은  $M$ 의 0이 아닌 값은 변이체, 예컨대 돌연변이의 이형접합성을 나타낸다.

[0660]

한 구현예에서, 상기 방법은 상기 변이체, 예컨대 돌연변이에 대한 접합성의 표시를 수득하는 단계를 포함한다. 한 구현예에서, 상기 돌연변이 상태는  $M = C \neq 0$ 인 경우 동형접합성(예컨대, LOH)인 것으로 결정된다. 한 구현예에서, 상기 돌연변이 상태는  $M = C = 0$ 인 경우 동형접합성 결실인 것으로 결정된다. 한 구현예에서, 상기 돌연변이 상태는  $0 < M < C$ 인 경우 이형접합성인 것으로 결정된다. 한 구현예에서, 상기 돌연변이는  $M = 0$  및  $C \neq 0$ 인 경우 종양에 존재하지 않는다. 한 구현예에서, 상기 접합성은, 예컨대 샘플 순도가 약 80% 초과, 예컨대 약 90% 내지 100%, 예컨대 약 90% 내지 95%, 또는 약 95% 내지 100%일 때 결정된다.

[0661]

한 구현예에서, 대조군은 샘플이 유래된 대상체 이외의 대상체로부터의 정배수체(예컨대, 이배체) 조직의 샘플, 또는 샘플이 유래된 대상체 이외의 하나 이상(예컨대, 적어도 2, 3, 4, 또는 5) 대상체로부터의 혼합된 정배수체(예컨대, 이배체) 조직의 샘플이다. 한 구현예에서, 상기 방법은, 예컨대 차세대 서열분석(NGS)에 의해 선택된 대상체 간격의 각각 및 선택된 생식계열 SNP 각각을 서열분석하는 단계를 포함한다. 한 구현예에서, 정규화 전의 서열 적용 범위는 적어도 약 10X, 20X, 30X, 50X, 100X, 250X, 500X, 750X, 800X, 900X, 1,000X, 1,500X, 2,000X, 2,500X, 3,000X, 3,500X, 4,000X, 4,500X, 5,000X, 5,500X, 6,000X, 6,500X, 7,000X, 7,500X, 8,000X, 8,500X, 9,000X, 9,500X, 또는 10,000X 서열분석의 정도이다.

[0662]

한 구현예에서, 상기 대상체는 항암 요법을 받았다. 한 구현예에서 상기 대상체는 항암 요법을 받았고 상기 요법에 내성이 있거나 질병 진행을 나타낸다. 한 구현예에서 상기 대상체는 FDA, EMA, 또는 다른 규제 기관에 의해 승인된 치료제; 또는 FDA, EMA 또는 기타 규제 기관의 승인을 받지 않은 치료제로부터 선택되는 항암 요법을 받았다. 한 구현예에서, 상기 대상체는 임상 시험, 예컨대 I상, II상 또는 III상 임상 시험(또는 상기 시험의 미국 외 등가물) 과정에서 항암 요법을 받은 적이 있다. 한 구현예에서, 상기 변이체는 상기 대상체에 존재하는 종양의 유형, 예컨대 치료의 발생 또는 치료에 대한 내성과 긍정적으로 연관된다. 한 구현예에서, 상기 변이체는 상기 대상체에 존재하는 종양의 유형과 긍정적으로 연관되지 않는다. 한 구현예에서, 상기 변이체는 상기 대상체에 존재하는 종양 유형 이외의 종양과 긍정적으로 연관된다. 한 구현예에서, 상기 변이체는 상기 대상체에 존재하는 종양의 유형과 긍정적으로 연관되지 않은 변이체이다.

[0663]

한 구현예에서, 상기 방법은, 예컨대 데이터베이스에서, 예컨대 기계 판독 가능한 데이터베이스에서 종양, 예컨대 샘플 내의 종양 유형과 관련된 다른 돌연변이, 샘플 내의 종양 유형과 관련되지 않은 다른 돌연변이, 또는 샘플 내의 종양 유형 이외의 종양과 관련된 다른 돌연변이; 변이체의 특성화; 대립유전자 또는 유전자; 또는 종양 유형, 예컨대 종양이 1차 또는 2차인지 여부에 관계없이 종양 유형의 이름; 대상체의 특성; 또는 치료 대안,

권장 사항 또는 선택 사항들 중 하나 이상에 대한 설명자를 포함하거나 전송하는 보고서를 제공할 수 있다.

- [0664] 한 구현예에서, 상기 변이체의 특성화와 관련된 설명자는 집합체 또는 생식계열 대 체세포 상태에 대한 설명자를 포함한다. 한 구현예에서, 대상체 특성과 관련된 설명자는 상기 대상체의 정체성; 상기 대상체의 나이, 성별, 체중 또는 기타 유사한 특성, 직업 중 하나 이상; 상기 대상체의 병력, 예컨대 종양 또는 기타 장애의 발생; 상기 대상체의 가족 병력, 예컨대 변이체를 공유하거나 공유하지 않는 친척; 또는 대상의 이전 치료 이력, 예컨대 받은 치료, 이전에 투여된 항암 요법에 대한 반응, 예컨대 질병 내성, 반응성 또는 진행 중 하나 이상에 대한 설명자를 포함한다.
- [0665] 상기 SGZ 알고리즘은 또한 Sun *et al.* *PLoS Comput Biol.* 2018; 14(2):e1005965; Sun *et al.* *Cancer Research* 2014; 74(19S):1893-1893; 국제출원공보 제W02014/183078호, 미국특허 제9,792,403호 및 미국특허출원공보 제2014/0336996호에 기재되어 있으며, 그 내용은 전체가 본원에 참조로서 원용된다.
- [0666] 종양 돌연변이 부담
- [0667] 본원에 기재된 방법은 종양 돌연변이 부담(TMB)을 평가하기 위한 방법과 조합하여 또는 그의 일부로서 사용될 수 있다.
- [0668] 특정 구현예들에서, 상기 방법은 샘플(예컨대, 본원에 기재된 샘플)로부터 하위 유전체 간격 세트의 서열을 제공하는 단계; 및 돌연변이 부담에 대한 값을 결정하는 단계를 포함하며, 여기서 상기 값은 하위 유전체 간격 세트에서의 변경의 수의 함수이다. 특정 구현예들에서, 하위 유전체 간격 세트는 유전자 세트, 예를 들어 전체 유전체 또는 엑솜을 포함하지 않는 유전자 세트로부터 유래한다. 특정 구현예들에서, 하위 유전체 간격 세트는 부호화 하위 유전체 간격 세트이다. 다른 구현예들에서, 하위 유전체 간격 세트는 하나 이상의 부호화 하위 유전체 간격 및 하나 이상의 비부호화 하위 유전체 간격을 포함한다. 특정 구현예들에서, 돌연변이 부담에 대한 값은 하위 유전체 간격 세트에서 변경(예컨대, 체세포 변경)의 수의 함수이다. 특정 구현예들에서, 변경의 수는 기능적 변경, 생식계열 변경, 또는 둘 모두의 수를 제외한다.
- [0669] 본원에 기재된 방법은 또한, 예컨대: 샘플로부터 복수의 종양 핵산 분자를 포함하는 라이브러리를 획득하는 단계; 상기 라이브러리를 표적 포획 시약과 접촉시켜 혼성화에 의해 선택된 종양 핵산 분자를 제공함으로써 라이브러리 캐치를 제공하는 단계; 상기 라이브러리 캐치로부터 종양 핵산 분자로부터의 변경을 포함하는 하위 유전체 간격에 대한 리드를 획득하는 단계; 정렬 방법에 의해 상기 리드를 정렬하는 단계; 뉴클레오타이드 위치에 대해 상기 리드로부터 뉴클레오타이드 값을 할당하는 단계; 및 상기 할당된 뉴클레오타이드 위치의 세트로부터 하위 유전체 간격 세트를 선택하는 단계를 포함하는 단계로서, 상기 하위 유전체 간격 세트는 유전자 세트로부터 유래하는 단계 중 하나 이상을 포함한다.
- [0670] 특정 구현예들에서, 상기 돌연변이 부담은 대상체, 예컨대 본원에 기재된 대상체로부터의 샘플에서 측정된다. 특정 구현예들에서, 상기 돌연변이 부담은, 예컨대 참조 집단으로부터의 샘플의 돌연변이 부담 중에서 백분위수로 표현된다. 특정 구현예들에서, 상기 참조 집단은 상기 대상체와 동일한 유형의 암을 갖는 환자를 포함한다. 다른 구현예들에서, 상기 참조 집단은 상기 대상체와 동일한 유형의 요법을 받고 있거나 받은 적이 있는 환자를 포함한다. 특정 구현예들에서, 예컨대 표 1A-4B에 제시된 유전자 세트에서 변경(예컨대, 체세포 변경)의 수준을 평가함으로써 본원에 기재된 방법에 의해 획득한 돌연변이 부담은 전체 유전체 또는 엑솜 돌연변이 부담과 상관관계가 있다.
- [0671] 용어 "돌연변이의 부담", "돌연변이 부담", "돌연변이 부하" 및 "돌연변이의 부하"는 본원에서 상호교환가능하게 사용된다. 종양과 관련하여, 돌연변이의 부하는 또한 본원에서 "종양 돌연변이의 부담", "종양 돌연변이 부담" 또는 "TMB"로 지칭된다. 이론에 얽매이지 않고, 일부 구현예들에서 TMB는 유전체 시그니처의 유형, 예컨대 연속/복합 바이오마커로 간주될 수 있는 것으로 여겨진다.
- [0672] 본원에서 사용된 용어 "돌연변이 부담" 또는 "돌연변이의 부담"은 유전자 세트에서(예컨대, 상기 유전자 세트의 부호화 영역에서) 미리 정의된 단위 당(예컨대, 메가염기 당) 변경(예컨대, 하나 이상의 변경, 예컨대 하나 이상의 체세포 변경)의 수준, 예컨대 수를 지칭한다. 돌연변이의 부담은, 예컨대 전체 유전체 또는 엑솜 기반으로, 또는 유전체 또는 엑솜의 하위 집합을 기반으로 측정할 수 있다. 특정 구현예들에서, 유전체 또는 엑솜의 하위 집합에 기초하여 측정된 돌연변이의 부담은 전체 유전체 또는 엑솜 돌연변이의 부담을 결정하기 위해 외삽될 수 있다.
- [0673] 한 구현예에서, 상기 방법은:

- [0674] a) 샘플로부터 대상체 간격 세트(예컨대, 부호화 대상체 간격)의 서열, 예컨대 뉴클레오타이드 서열을 제공하는 단계로서, 상기 대상체 간격 세트는 유전자 세트로부터 유래하는 것인 단계; 및
- [0675] b) 돌연변이의 부담에 대한 값을 결정하는 단계로서, 상기 값은 대상체 간격 세트에서 변경(예컨대, 하나 이상의 변경), 예컨대 체세포 변경(예컨대, 하나 이상의 체세포 변경)의 수의 함수인 단계를 포함한다.
- [0676] 특정 구현예들에서, 변경의 수는 대상체 간격의 기능적 변경을 제외한다. 다른 구현예들에서, 변경의 수는 대상체 간격의 생식계열 변경을 제외한다. 특정 구현예들에서, 변경의 수는 대상체 간격의 기능적 변경 및 대상체 간격의 생식계열 변경을 제외한다.
- [0677] 특정 구현예들에서, 대상체 간격이 세트는 부호화 대상체 간격을 포함한다. 다른 구현예들에서, 대상체 간격의 세트는 비부호화 대상체 간격을 포함한다. 특정 구현예들에서, 대상체 간격의 세트는 부호화 대상체 간격을 포함한다. 다른 구현예들에서, 대상체 간격의 세트는 하나 이상의 부호화 하위 유전체 간격 및 하나 이상의 비부호화 하위 유전체 간격을 포함한다. 특정 구현예들에서, 상기 대상체 간격의 세트에서 대상체 간격의 약 5% 이상, 약 10% 이상, 약 20% 이상, 약 30% 이상, 약 40% 이상, 약 50% 이상, 약 60% 이상, 약 70% 또는 대상체 간격 세트에서 대상체 간격의 약 80% 이상, 약 90% 이상, 또는 약 95% 이상은 부호화 대상체 간격이다. 다른 구현예들에서, 상기 대상체 간격의 세트에서 대상체 간격의 약 90% 이하, 약 80% 이하, 약 70% 이하, 약 60% 이하, 약 50% 이하, 약 40% 이하, 약 30% 이하, 약 20% 또는 약 10% 이하, 또는 약 5% 이하는 비부호화 대상체 간격이다.
- [0678] 다른 구현예들에서, 상기 대상체 간격의 세트는 전체 유전체 또는 전체 엑솜을 포함하지 않는다. 다른 구현예들에서, 상기 부호화 대상체 간격의 세트는 전체 엑솜을 포함하지 않는다.
- [0679] 특정 구현예들에서, 상기 유전자 세트는 전체 유전자 또는 전체 엑솜을 포함하지 않는다. 다른 구현예들에서, 상기 유전자 세트는 표 2A-5B에 제시된 하나 이상의 유전자를 포함하거나 이로써 구성된다.
- [0680] 특정 구현예들에서, 상기 값은 유전자 세트의 함수로서 표현된다. 특정 구현예들에서, 상기 값은 유전자 세트의 부호화 영역의 함수로서 표현된다. 다른 구현예들에서, 상기 값은 유전자 세트의 비부호화 영역의 함수로서 표현된다. 특정 구현예들에서, 상기 값은 유전자 세트의 엑손의 함수로서 표현된다. 다른 구현예들에서, 상기 값은 유전자 세트의 인트론의 함수로서 표현된다.
- [0681] 특정 구현예들에서, 상기 값은 서열분석된 유전자 세트의 함수로서 표현된다. 특정 구현예들에서, 상기 값은 서열분석된 유전자 세트의 부호화 영역의 함수로서 표현된다. 다른 구현예들에서, 상기 값은 서열분석된 유전자 세트의 비부호화 영역의 함수로서 표현된다. 특정 구현예들에서, 상기 값은 서열분석된 유전자 세트의 엑손의 함수로서 표현된다. 다른 구현예들에서, 상기 값은 서열분석된 유전자 세트의 인트론의 함수로서 표현된다.
- [0682] 특정 구현예들에서, 상기 값은 유전자 세트의 다수의 위치에서 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다. 특정 구현예들에서, 상기 값은 유전자 세트의 부호화 영역의 다수의 위치에서 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다. 다른 구현예들에서, 상기 값은 유전자 세트의 비부호화 영역의 다수의 위치에서 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다. 특정 구현예들에서, 상기 값은 유전자 세트의 엑손의 다수의 위치에서 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다. 다른 구현예들에서, 상기 값은 유전자 세트의 인트론의 다수의 위치에서 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다.
- [0683] 특정 구현예들에서, 상기 값은 서열분석된 유전자 세트의 다수의 위치에서 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다. 특정 구현예들에서, 상기 값은 서열분석된 유전자 세트의 부호화 영역의 다수의 위치에서 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다. 다른 구현예들에서, 상기 값은 서열분석된 유전자 세트의 비부호화 영역의 다수의 위치에서 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다. 특정 구현예들에서, 상기 값은 서열분석된 유전자 세트의 엑손의 다수의 위치에서 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다. 다른 구현예들에서, 상기 값은 서열분석된 유전자 세트의 인트론의 다수의 위치에서 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다.
- [0684] 특정 구현예들에서, 상기 값은 단위 당 변경(예컨대, 체세포 변경)의 수의 함수로서, 예컨대 메가염기 당 체세포 변경의 수의 함수로서 표현된다.
- [0685] 특정 구현예들에서, 상기 값은 유전자 세트에서 메가염기 당 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다. 특정 구현예들에서, 상기 값은 유전자 세트의 부호화 영역에서 메가염기 당 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다. 다른 구현예들에서, 상기 값은 유전자 세트의 비부호화 영역에서 메가염기 당 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다. 특정 구현예들에서, 상기 값은 유전자 세트의 엑손에서 메가염

기 당 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다. 다른 구현예들에서, 상기 값은 유전자 세트의 인트론에서 메가염기 당 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다.

- [0686] 특정 구현예들에서, 상기 값은 서열분석된 유전자 세트에서 메가염기 당 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다. 특정 구현예들에서, 상기 값은 서열분석된 유전자 세트의 부호화 영역에서 메가염기 당 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다. 다른 구현예들에서, 상기 값은 서열분석된 유전자 세트의 비부호화 영역에서 메가염기 당 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다. 특정 구현예들에서, 상기 값은 서열분석된 유전자 세트의 엑손에서 메가염기 당 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다. 다른 구현예들에서, 상기 값은 서열분석된 유전자 세트의 인트론에서 메가염기 당 변경(예컨대, 체세포 변경)의 수의 함수로서 표현된다.
- [0687] 특정 구현예들에서, 상기 돌연변이의 부담은, 예컨대 전체 돌연변이의 부담을 수득하기 위해 유전체의 더 큰 부분, 예컨대 엑솜 또는 전체 유전체로 외삽된다. 특정 구현예들에서, 상기 돌연변이의 부담은 엑솜의 더 큰 부분, 예컨대 전체 엑솜으로 외삽된다.
- [0688] 특정 구현예들에서, 상기 샘플은 대상체로부터 비롯된 것이다. 특정 구현예들에서, 상기 대상체는 장애, 예컨대 암을 가진다. 다른 구현예들에서, 상기 대상체는 요법, 예컨대 면역요법을 받는 중이거나, 받은 적이 있다.
- [0689] 특정 구현예들에서, 상기 돌연변이 부담은, 예컨대 참조 집단으로부터의 샘플의 돌연변이 부담 중에서 백분위로 표현된다. 특정 구현예들에서, 상기 참조 집단은 상기 대상체와 동일한 유형의 암을 갖는 환자를 포함한다. 다른 구현예들에서, 상기 참조 집단은 상기 대상체와 동일한 유형의 요법을 받고 있거나 받은 적이 있는 환자를 포함한다.
- [0690] 특정 구현예들에서, 상기 방법은:
- [0691] (i) 샘플로부터 복수의 종양 핵산 분자를 포함하는 라이브러리를 수득하는 단계;
- [0692] (ii) 상기 라이브러리를 표적 포획 시약과 접촉시켜 선택된 종양 핵산 분자들을 제공하는 단계로서, 상기 표적 포획 시약은 상기 종양 핵산 분자와 혼성화되어 라이브러리 캐치를 제공하는 단계;
- [0693] (iii) 상기 라이브러리 캐치로부터의 종양 핵산 분자로부터의 변경(예컨대, 체세포 변경)을 포함하는 대상체 간격에 대한 리드를, 예컨대 차세대 서열분석 방법에 의해 수득하는 단계;
- [0694] (iv) 정렬 방법에 의해 상기 리드를 정렬하는 단계;
- [0695] (v) 뉴클레오타이드 위치에 대해 상기 리드로부터 뉴클레오타이드 값을 할당하는 단계;
- [0696] (vi) 할당된 뉴클레오타이드 위치의 세트로부터 대상체 간격의 세트(예컨대, 부호화 대상체 간격)를 선택하는 단계로서, 상기 대상체 간격의 세트는 유전자 세트로부터 유래되는 단계; 및
- [0697] (vii) 돌연변이의 부담에 대한 값을 결정하는 단계로서, 상기 값은 대상체 간격 세트에서 변경(예컨대, 하나 이상의 변경), 예컨대 체세포 변경(예컨대, 하나 이상의 체세포 변경)의 수의 함수인 단계를 포함한다.
- [0698] 특정 구현예들에서, 변경(예컨대, 체세포 변경)의 수는 대상체 간격의 기능적 변경을 제외한다. 다른 구현예들에서, 변경의 수는 대상체 간격의 생식계열 변경을 제외한다. 특정 구현예들에서, 변경(예컨대, 체세포 변경)대상 간격의 기능적 변경 및 대상체 간격의 생식계열 변경을 제외한다.
- [0699] 종양 돌연변이의 부담을 평가하기 위한 다른 방법은 국제특허출원공보 제W02017/151524호에 기재되어 있으며, 그 내용은 전체가 참조로서 인용된다.
- [0700] 적용
- [0701] 본원에 개시된 방법은, 예컨대 유전체의 암 관련 분절에 적용되는 바와 같이 최적화된 표적 포획 시약(예컨대, 미끼) 기반 선택, 최적화된 정렬 및 최적화된 돌연변이 호출을 포함하는 다수의 최적화된 요소의 통합을 허용한다. 본원에 기재된 방법은 암별, 유전자별 및 부위별 기반으로 최적화될 수 있는 종양의 NGS 기반 분석을 제공한다. 이는, 예컨대 본원에 기재된 유전자/부위 및 종양 유형에 적용될 수 있다. 상기 방법은 주어진 서열분석 기술을 사용하여 돌연변이 검출을 위한 민감도 및 특이성 수준을 최적화한다. 암별, 유전자별, 부위별 최적화는 임상 제품에 필수적인 매우 높은 수준의 민감도/특이성(예컨대, 둘 다 >99%)을 제공한다.
- [0702] 이론에 얽매이지 않고, 일부 구현예들에서, 본원에 기재된 방법은 선택된 유전체 영역의 검출에서 증가된 민감도로부터 이익을 얻을 일반적인 서열분석 응용에 적용될 수 있다고 믿어진다. 예를 들어, 상기 응용 프로그램에

는 유병률을 기반으로 증가된 적용 범위가 있는 유전성 암 패널, 특정 질병 경로를 대상으로 하는 기타 전체 엑솜 서열분석(WES) 시험 및 후보 실행 가능한 초점 이벤트에 대한 농축을 포함한 산전 테스트가 포함되지만 이에 제한되지는 않는다.

- [0703] 일부 구현예들에서, 상기 방법은 유전체 변경, 예컨대 체세포 변경의 평가에 반응하는 치료제를 선택하는 단계를 추가로 포함한다. 일부 구현예들에서, 상기 방법은 돌연변이의 부담, 예컨대 돌연변이의 부담의 증가 또는 감소된 수준의 평가에 반응하는 치료를 선택하는 단계를 추가로 포함할 수 있다. 일부 구현예들에서, 상기 방법은 유전체 변경의 평가에 반응하는 치료제를 투여하는 단계를 추가로 포함한다. 일부 구현예들에서, 상기 방법은 유전체 변경의 평가에 반응하여 샘플 또는 상기 샘플이 유래된 대상체를 분류하는 단계를 추가로 포함한다. 일부 구현예들에서, 상기 방법은 샘플을 수득한 대상체에 대한 임상 시험 적격성을 결정하는 단계를 추가로 포함한다. 일부 구현예들에서, 상기 방법은 보고서, 예컨대 전자, 웹 기반 또는 서면 보고서를 생성하여 환자 또는 다른 사람 또는 실체, 간병인, 의사, 종양 전문의, 병원, 클리닉, 제3자 지급인, 보험 회사 또는 관공서에 전달하는 단계를 추가로 포함한다. 일부 구현예들에서, 상기 보고서는 본원에 기재된 방법으로부터의 출력을 포함한다.
- [0704] 본원에 기재된 방법은 일상적인 실제 샘플에서 차세대 서열분석 기술을 사용하여 그럴듯하게 실행 가능한 유전자(일반적으로 50 내지 500개의 유전자 범위일 수 있음)의 포괄적인 세트에 대한 유전체 이상에 대한 임상 및 규제 등급 종합 분석 및 해석을 제공함으로써 최적의 치료 및 질병 관리 결정을 알려준다.
- [0705] 본원에 기재된 방법은 최적의 치료 및 질병 관리 결정을 알리기 위해 종양 전문의/병리학자가 샘플을 보내고 포괄적인 분석 및 종양의 유전체 및 기타 분자 변화에 대한 설명을 수신할 수 있는 윈스톱 서비스를 제공한다.
- [0706] 본원에 기재된 방법은 표준의 입수 가능한 샘플을 채취하는 강력한 실제 임상 종양학 진단 도구를 제공하고 단일 시험에서 포괄적인 유전체 및 기타 분자 이상 분석을 제공함으로써 종양 전문의에게 종양을 유발할 수 있는 이상 반응에 대한 포괄적인 설명을 제공하고 종양 전문의의 치료 결정을 알리는 데 유용하다.
- [0707] 본원에 기재된 방법은, 예컨대 차세대 서열분석(NGS)에 의한 임상 등급 품질로 환자의 암 유전체에 대한 포괄적인 분석을 제공한다. 방법은 가장 관련성이 높은 유전자 및 잠재적 변경을 포함하며 돌연변이(예컨대, 삽입-결실 또는 염기 치환), 복제 수, 재배열(예컨대, 전좌, 발현 및 후성유전적 표지자) 분석 중 하나 이상을 포함한다. 유전자 분석의 결과는 실행 가능한 결과의 서술적인 보고와 함께 맥락화될 수 있다. 방법은 일련의 최신 과학 및 의학 지식의 사용을 제공한다.
- [0708] 일부 구현예들에서, 상기 방법은 임의의 질병(예컨대, 암)의 진단, 예방 또는 치료, 또는 장애의 진단, 또는 인간의 건강 평가를 위한 정보를 제공하기 위한 목적으로 인간 신체로부터 유래된 샘플을 분석한다. 일부 구현예들에서, 상기 방법은 임상실험실개선개정안(Clinical Laboratory Improvement Amendment, CLIA) 및/또는 미국 병리학회(College of American Pathologists, CAP)에서 제공하는 지침에 따라 실시한다. 일부 구현예들에서, 상기 방법은 CLIA 및/또는 CAP 인증 시설에서 실시한다. 일부 구현예들에서, 상기 방법은 식품의약품(Food and Drug Administration, FDA), 유럽의약품청(European Medicines Agency, EMA), 품질 시스템 규정(Quality System Regulation, QSR), 유럽 위원회(European Commission, CE), 예컨대 CE 시험관 내 진단(CE-IVD), 중국 식품의약품(Chinese Food and Drug Administration, CFDA) 또는 기타 규제 기관에 의해 제공된 지침에 따라 실시한다. 일부 구현예들에서, 상기 방법은 FDA, QSR, CE 또는 CFDA 인증 시설에서 실시한다. 일부 구현예들에서, 상기 방법은 QSR 인증 시설에서 실시한다. 일부 구현예들에서, 상기 방법은 임상 등급 샘플, 예컨대 임상 실험, 시험 또는 환자 치료의 관리에 적합한 샘플을 분석한다. 일부 구현예들에서, 상기 샘플은 소급 샘플 및/또는 예상 샘플을 포함한다. 일부 구현예들에서, 소급 샘플은 치료가 투여되기 전 또는 후에 분석된 샘플을 포함하거나 연구 샘플이다. 일부 구현예들에서, 예상 샘플은 치료를 받지 않은 대상체의 샘플을 포함한다. 일부 구현예들에서, 예상 샘플을 분석하기 위해 본원에 기재된 방법을 사용하면 상기 샘플을 수득한, 예컨대 유래된 대상체에 대한 요법의 결과를 예측할 수 있다.
- [0709] 일부 구현예들에서, 상기 방법은, 예컨대 본원에 기재된 바와 같은 진단으로서 사용된다. 일부 구현예들에서, 상기 방법은 동반 진단에서 또는 동반 진단과 함께 사용된다. 일부 구현예들에서, 상기 방법은 상보적 진단으로서 사용된다.
- [0710] 일부 구현예들에서, 상기 방법의 유효성은 정확성, 정밀도, 민감도, 특이성, 보고 가능한 범위, 또는 참조 간격의 하나 이상(예컨대, 2, 3, 4, 5 또는 모두)을 결정함으로써 확립된다(예컨대, CLIA 규정에 따라). 특정 구현예들에서, 정확성은 표적 영역에서, 예컨대 공지된 변이체(예컨대, SNP, 삽입-결실)에 대한 적용 범위 및 품질

(예컨대, 프레드(Phred) 스코어)에 의해 결정된다. 특정 구현예들에서, 정밀도는, 예컨대 공지된 변이체에 대해 상이한 조작자 및 기구 사이의 서열 반복 및 적용 범위 분포에 의해 결정된다. 특정 구현예들에서, 특이성은 위양성 비율, 예컨대 잘 특성화된 표적을 갖는 여러 샘플에서 특정 적용 범위의 한계값에서 위변이가 식별되는 정도에 의해 결정된다. 특정 구현예들에서, 민감도는, 예컨대 잘 특성화된 표적을 갖는 여러 샘플에서 공지된 변이체를 검출하는 가능성 시험에 의해 결정된다. 특정 구현예들에서, 보고 가능한 범위는, 예컨대 반복 영역, 삽입-결실 또는 대립유전자 탈락과 함께 하나 이상의 유전자의 인트론 완충액 및 엑손 영역에 의해 결정된다. 특정 구현예들에서, 참조 간격은, 예컨대 영향을 받지 않은 집단에서 서열 변이 배경 측정에 의해 결정된다.

[0711] 일부 구현예들에서, 상기 방법은 검증된 샘플 추출, 라이브러리 제조, 바코드, 풀링, 표적 강화 또는 생물정보학(예컨대, 얼마나 정확하고 민감한 변이체가 호출되는지) 중 하나 이상(예컨대, 2, 3, 4, 5개 또는 모두)에 대한 고려를 포함하는 설정(예컨대, CAP 규정 하에)에서 실시한다.

[0712] 본원에 기재된 방법은 환자 치료의 품질 및 효율성 모두를 증가시키는 것을 제공한다. 여기에는 종양이 회귀하거나 제대로 연구되지 않은 유형이어서 치료 표준이 없거나 환자가 확립된 요법에 불응하고 추가 요법 선택 또는 임상 시험 참여를 위한 합리적인 근거가 유용할 수 있는 적용이 포함된다. 예컨대, 상기 방법은 치료의 어느 시점에서도 종양 전문의가 의사 결정에 정보를 제공하는 데 사용할 수 있는 전체 "분자 이미지" 및/또는 "분자 하위 진단"을 제공함으로써 이익을 얻을 수 있는 선택을 가능하게 한다. 결과는 환자가 임상 시험에 등록할 자격이 있는지 여부를 결정하는 데 사용할 수 있다.

[0713] 본원에 기재된 방법은 보고서를, 예컨대 전자, 웹 기반 또는 서면 형식으로 환자 또는 다른 사람 또는 실체, 예컨대 간병인, 예컨대 의사, 예컨대 종양 전문의, 병원, 진료소, 제3자 지불인, 보험 회사 또는 관공서에 제공하는 단계를 포함할 수 있다. 상기 보고서는 상기 방법의 출력, 예컨대 뉴클레오티드 값의 확인, 변경, 돌연변이 또는 야생형 서열의 존재 또는 부재 표시, 예컨대 샘플 유형의 종양과 관련된 대상체 간격에 대한 출력을 포함할 수 있다. 상기 보고서는 또한 종양 돌연변이 부담 수준에 대한 정보를 포함할 수 있다. 상기 보고서는 하나 이상의 유전체 시그니처, 예컨대 연속/복합 바이오마커, 예컨대 미소부수체 불안정성의 수준, 또는 이형접합체(LOH)의 존재 또는 부재에 관한 정보를 포함한다. 상기 보고서는 또한 질병에서 서열, 예컨대 변경, 돌연변이 또는 야생형 서열의 역할에 대한 정보를 포함할 수 있다. 상기 정보에는 예후, 내성 또는 잠재적 또는 제안된 치료 옵션에 대한 정보가 포함될 수 있다. 상기 보고서는 치료 옵션의 가능성 있는 유효성, 치료 옵션의 수용 가능성, 또는 치료 옵션을 환자, 예컨대 보고서에서 확인된 시험 및 구현예에서 확인된 변경을 서열을 가진 환자에게 적용할 만한 타당성에 관한 정보를 포함할 수 있다. 예컨대, 상기 보고서는 환자에 대한 약물의 투여, 예컨대 투여량 또는 치료 섭생의 투여, 예컨대 다른 약물과의 병용에 관한 정보 또는 권고사항을 포함할 수 있다. 한 구현예에서, 상기 방법에서 확인된 모든 돌연변이가 보고서에서 확인된 것은 아니다. 예컨대, 상기 보고서는, 치료에서, 예컨대 치료 옵션과 함께 암의 발생, 예후, 병기 또는 취약성과 상관관계 수준을 갖는 유전자의 돌연변이로 제한될 수 있다. 본원에 소개된 방법을 사용하면 상기 방법을 실시하는 개체가 샘플을 수령한 후 7, 14 또는 21일 이내에 보고서를, 예컨대 본원에 기재된 개체에게 전달하는 것을 가능하게 한다. 따라서, 본 발명에서 소개된 방법은, 예컨대 샘플 수령 후 7, 14 또는 21일 이내에 빠른 처리 시간을 가능하게 한다.

[0714] 본원에 기재된 방법은 조직학적으로 정상인 샘플, 예컨대 수술 절제면으로부터의 샘플을 평가하는 데에도 사용될 수 있다. 본원에 기재된 바와 같은 하나 이상의 변경이 검출되는 경우, 해당 조직은, 예컨대 악성 또는 전암성으로 재분류될 수 있고/있거나 치료 과정이 변경될 수 있다.

[0715] 일부 구현예들에서, 본원에 기재된 방법은 비암 적용, 예컨대 법의학 적용(예컨대, 치과 기록의 사용에 대한 대안으로서 또는 이에 추가하여 식별), 친자 확인 검사, 다른 무엇보다도 감염성 질환, 자가면역 질환, 낭포성 섬유증, 헌팅턴병, 알츠하이머병의 질병 진단 및 예후 유용하다. 예를 들어, 본원에 기재된 방법에 의한 유전적 변경의 확인은 특정 장애를 발병할 개체의 존재 또는 위험을 나타낼 수 있다.

[0716] 시스템

[0717] 다른 양태에서, 본 발명은, 예컨대 본원에 기재된 방법에 따라 샘플에서 유전체 변경을 평가하기 위한 시스템을 특징으로 한다. 상기 시스템은 메모리에 작동 가능하게 연결된 적어도 하나의 프로세서를 포함하고, 실행 시 적어도 하나의 프로세서는 본원에 설명된 샘플을 분석하는 방법을 수행하도록 구성된다.

[0718] 다르게 정의되지 않는 한, 본원에서 사용된 모든 기술 및 과학 용어는 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 갖는다. 본 명세서에 기재된 것과 유사하거나 등가인 방법 및 재료가 본 발명의 실시 또는 시험에 사용될 수 있지만, 적합한 방법 및 재료가 하기에 기재되어

있다. 본원에 언급된 모든 간행물, 특허 출원, 특허 및 기타 참고 문헌은 그 전체가 참고로서 원용된다. 또한, 재료, 방법 및 예는 예시일 뿐이며 제한하려는 의도가 아니다.

- [0719] 본 발명의 다른 특징 및 이점은 상세한 설명, 도면 및 청구범위로부터 명백할 것이다.
- [0720] 기타 구현예들
- [0721] 대안적으로 또는 본원에 기재된 방법과 조합하여, 일부 구현예들에서, 상기 방법은 (a)-(h) 중 하나 이상(예컨대, 2, 3, 4, 5, 6, 7개 또는 모두)을 추가로 포함한다:
- [0722] (a) 예를 들어, 본원에 기재된 복수의 표적 포획 시약을 사용하여 샘플(예컨대, 혈액 샘플)로부터 핵산 분자(예컨대, cfDNA)를 제공하는 단계;
- [0723] (b) 복수의 상이한 바코드 서열을 포함하는 바코드를 포함하는 어댑터를 핵산
- [0724] 분자에 부착함으로써 태깅된 모 핵산 분자를 생성하는 단계;
- [0725] (c) 상기 태깅된 모 핵산 분자를 증폭하여 증폭된 태깅된 자손 핵산 분자를 생성하는 단계;
- [0726] (d) 상기 증폭된 태깅된 자손 핵산 분자를 서열분석하여 각각의 태깅된 모 핵산 분자로부터 복수의 서열 리드를 생성하는 단계로서, 상기 복수의 서열 리드의 각각의 서열 리드는 바코드 서열 및 핵산으로부터 유래된 서열을 포함하는 단계;
- [0727] (e) 상기 복수의 서열 리드의 서열 리드를 하나 이상의 참조 서열에 매핑하는 단계;
- [0728] (f) e)에서 맵핑된 서열 리드를 적어도 서열 리드의 바코드 서열에 기초하여 군으로 분류하는 단계로서, 군의 각각은 동일한 바코드 서열을 포함하는 서열 리드를 포함하고, 이에 의해 군의 각각은 동일한 태깅된 모 핵산 분자로부터 증폭된 서열 리드를 포함하는 단계;
- [0729] (g) 하나 이상의 참조 서열에서 복수의 대상체 간격 각각에서, 상기 대상체 간격에서 각 군에 대한 돌연변이 호출을 생성하기 위해 각 군의 서열 리드를 붕괴시키는 단계; 또는
- [0730] (h) 하나 이상의 대상체 간격에서 하나 이상의 유전체 이상, 예컨대 삽입-결실, 복제 수 변이, 전이, 전좌, 역위, 결실, 이수성, 부분 이수성, 배수성, 염색체 불안정성, 염색체 구조 변경, 유전자 검출 융합, 염색체 융합, 유전자 절단, 유전자 증폭, 유전자 복제, 염색체 병변, DNA 병변, 핵산 화학적 변형의 비정상적 변화, 후성 유전 패턴의 비정상적 변화, 핵산 메틸화의 비정상적 변화, 또는 이들의 조합을 검출하는 단계.
- [0731] 대안적으로 또는 본원에 기재된 방법과 조합하여, 일부 구현예들에서, 상기 방법은, 예컨대 유전체 변경(예컨대, 단일 염기 변이체)를 정량화하기 위해 (a)-(i) 중 하나 이상(예컨대, 2, 3, 4, 5, 6, 7, 8개 또는 모두)을 추가로 포함한다:
- [0732] (a) 예를 들어, 본원에 기재된 복수의 표적 포획 시약을 사용하여 샘플(예컨대, 혈액 샘플)로부터 핵산 분자(예컨대, cfDNA)를 제공하는 단계;
- [0733] (b) 별개의 바코드 서열을 포함하는 바코드를 포함하는 어댑터를 상기 핵산 분자에 부착함으로써 태깅된 모 핵산 분자를 생성하는 단계;
- [0734] (c) 상기 태깅된 모 핵산 분자를 증폭하여 증폭된 태깅된 자손 핵산 분자를 생성하는 단계;
- [0735] (d) 상기 증폭된 태깅된 자손 핵산 분자를 서열분석하여 각각의 모 핵산 분자로부터 복수의 서열 리드를 생성하는 단계로서, 각각의 서열 리드는 바코드 서열 및 핵산으로부터 유래된 서열을 포함하는 단계;
- [0736] (e) (i) 바코드 서열 및 (ii) 핵산으로부터 유래된 서열의 시작 부분에 있는 서열 정보, 핵산으로부터 유래된 서열의 말단에 있는 서열 정보, 또는 서열 리드의 길이 중 하나 이상에 기초하여 각각의 태깅된 모 핵산 분자로부터 생성된 복수의 서열 리드를 군으로 분류하는 단계로서, 각각의 군은 태깅된 모 핵산 분자 중 고유한 핵산 분자로부터 증폭된 태깅된 자손 핵산 분자의 서열 리드를 포함하는 단계;
- [0737] (f) 각 군 내에서 분류된 서열 리드를 서로 비교하여 각 군에 대한 공통 서열을 결정하는 단계로서, 공통 서열 각각은 태깅된 모 핵산 분자 중 고유한 핵산 분자에 대응하는 단계;
- [0738] (g) 하나 이상의 대상체 간격을 포함하는 하나 이상의 참조 서열을 제공하는 단계;
- [0739] (h) 상기 하나 이상의 대상체 간격의 주어진 대상체 간격에 매핑되는 공통 서열을 식별하는 단계; 또는

- [0740] (i) 유전체 변경을 포함하는 주어진 대상체 간격에 매핑되는 다수의 공통 서열을 계산함으로써 상기 샘플의 유전체 변경을 정량화하는 단계.
- [0741] 대안적으로 또는 본원에 기재된 방법과 조합하여, 일부 구현예들에서, 상기 방법은 (a)-(h) 중 하나 이상(예컨대, 2, 3, 4, 5, 6, 7개 또는 모두)을 추가로 포함한다:
- [0742] (a) 예를 들어, 본원에 기재된 복수의 표적 포획 시약을 사용하여 샘플(예컨대, 혈액 샘플)로부터 핵산 분자(예컨대, cfDNA)를 제공하는 단계;
- [0743] (b) 복수의 핵산 분자를 복수의 태깅된 모 핵산 분자로 전환하는 단계로서, 각각의 태깅된 모 핵산 분자는 (i) 복수의 핵산 분자의 핵산 분자로부터의 서열, 및 (ii) 하나 이상의 바코드를 포함하는 식별자 서열을 포함하는 단계;
- [0744] (c) 복수의 태깅된 모 핵산 분자를 증폭하여 대응하는 복수의 증폭된 자손 핵산 분자를 생성하는 단계;
- [0745] (d) 복수의 증폭된 자손 핵산 분자를 서열분석하여 서열 리드 세트를 생성하는 단계;
- [0746] (e) 상기 서열 리드 세트의 서열 리드를 하나 이상의 참조 서열에 매핑하는 단계;
- [0747] (f) 상기 서열 리드를 군으로 분류하는 단계로서, 서열 리드를 포함하는 각각의 군은 동일한 식별자 서열을 포함하고 동일한 시작 및 정지 위치를 갖는 서열 리드를 가지며, 각각의 군은 동일한 태깅된 모 핵산 분자로부터 증폭된 서열 리드를 포함하는 단계;
- [0748] (g) 하나 이상의 참조 서열에서 복수의 대상체 간격의 각각의 대상체 간격에서, 상기 대상체 간격에서 각 군에 대한 돌연변이 호출을 생성하기 위해 각 군의 서열 리드를 붕괴시키는 단계; 또는
- [0749] (h) 군들 중에서 대상체 간격으로 호출된 하나 이상의 돌연변이의 빈도를 결정하는 단계.
- [0750] 대안적으로 또는 본원에 기재된 방법과 조합하여, 일부 구현예들에서, 상기 방법은, 복제 수 변이를 검출하기 위해 (a)-(f) 중 하나 이상(예컨대, 2, 3, 4, 5개 또는 모두)을 추가로 포함한다:
- [0751] (a) 예를 들어, 본원에 기재된 복수의 표적 포획 시약을 사용하여 샘플(예컨대, 혈액 샘플)로부터 핵산 분자(예컨대, cfDNA)를 제공하는 단계;
- [0752] (b) 상기 핵산 분자를 서열분석하는 단계로서, 각각의 핵산 분자는 복수의 서열 리드를 생성하는 단계;
- [0753] (c) 설정된 정확도, 품질 점수 또는 매핑 점수 임계값을 충족하지 못하는 리드를 필터링하는 단계;
- [0754] (d) 복수의 서열 리드를 참조 서열에 매핑하는 단계;
- [0755] (e) 상기 참조 서열의 복수의 영역에서 맵핑된 리드 또는 고유한 서열 리드를 정량화하는 단계; 및
- [0756] (f) i) 복수의 영역에서 리드의 수를 서로 정규화하거나, 또는 상기 복수의 영역에서 고유한 서열 리드의 수를 서로 정규화함으로써, 및/또는 ii) 복수의 영역에서 다수의 리드 또는 복수의 영역에서 다수의 고유한 서열 리드를 대조군 샘플로부터 수득한 숫자로 처리함으로써 복수의 미리 정의된 영역 중 하나 이상에서 복제 수 변이를 결정하는 단계.
- [0757] 대안적으로 또는 본원에 기재된 방법과 조합하여, 일부 구현예들에서, 상기 방법은, 복제 수 변이를 검출하기 위해 (a)-(h) 중 하나 이상(예컨대, 2, 3, 4, 5, 6, 7개 또는 모두)을 추가로 포함한다:
- [0758] (a) 예를 들어, 본원에 기재된 복수의 표적 포획 시약을 사용하여 샘플(예컨대, 혈액 샘플)로부터 핵산 분자(예컨대, cfDNA)를 제공하는 단계;
- [0759] (b) 상기 핵산 분자를 서열분석하는 단계로서, 각각의 핵산 분자는 복수의 서열 리드를 생성하는 단계;
- [0760] (c) 설정된 정확도, 품질 점수 또는 매핑 점수 임계값을 충족하지 못하는 리드를 필터링하는 단계;
- [0761] (d) 서열분석에서 유래된 서열 리드를 참조 서열에 매핑하는 단계;
- [0762] (e) 서열 리드 중에서 핵산 분자에 대응하는 고유한 서열 리드를 결정하는 단계;
- [0763] (f) 각각의 맵핑 가능한 염기 위치에서 참조 서열과 비교하여 변이체를 포함하는 맵핑된 고유 서열 리드의 하위 집합을 확인하는 단계;
- [0764] (g) 각 매핑 가능한 염기 위치에 대해, (a) 참조 서열과 비교하여 변이를 포함하는 맵핑된 고유한 서열 리드의

수 대 (b) 각 매핑 가능한 기본 위치에 대한 총 고유한 서열 리드의 수의 비율을 계산하는 단계; 및

(h) 참조 샘플에서 유사하게 유도된 숫자로 비율을 처리하는 단계.

대안적으로 또는 본원에 기재된 방법과 조합하여, 일부 구현예들에서, 상기 방법은 (a)-(h) 중 하나 이상(예컨대, 2, 3, 4, 5, 6, 7개 또는 모두)을 추가로 포함한다:

(a) 대상체로부터의 샘플(예컨대, 혈액 샘플) 내의 이중 가닥 DNA 분자(예컨대, cfDNA)를 이중 태그 세트로 태깅하는 단계로서, 상기 이중 태그 세트는 복수의 상이한 분자 바코드를 포함하고, 이중 태그 세트의 각각의 이중 태그는 샘플 내 이중 가닥 DNA 분자의 이중 가닥 DNA 분자의 상보적인 가닥을 상이하게 태깅하여 태깅된 가닥을 제공하고, 상기 태깅은 이중 가닥 DNA 분자와 비교하여 적어도 10X 과량의 이중 태그로 실시되며, 상기 이중 가닥 태그의 과량은 대상체의 샘플에서 이중 가닥 DNA 분자의 적어도 20%를 태그하기에 충분한, 단계;

(b) 참조 유전체 내의 하나 이상의 유전자좌 세트 내의 각각의 유전자좌에 대해, 예컨대 본원에 기재된 복수의 표적 포획 시약을 사용하여 농축된 태깅된 가닥을 제공하기 위해 유전자좌에 매핑된 태깅된 가닥의 하위집합에 대한 태깅된 가닥을 선택적으로 농축시키는 단계;

(c) 농축된 태깅된 가닥의 적어도 일부를 서열분석하여 대상체의 샘플로부터 복수의 미가공 서열 리드를 생성하는 단계;

(d) 복수의 원시 서열 리드를 복수의 군으로 분류하는 단계로서, 각 군은 동일한 모 폴리뉴클레오티드로부터 생성된 원시 서열 리드를 포함하며, 상기 분류는 (i) 모 폴리뉴클레오티드와 연관된 분자 바코드 및 (ii) 모 폴리뉴클레오티드의 미가공 서열의 시작 및/또는 말단 부분으로부터의 정보를 기반으로 하는 단계;

(e) 복수의 군으로 분류된 복수의 원시 서열 리드를 복수의 공통 서열 리드로 붕괴시키는 단계로서, 복수의 공통 서열 리드의 각각의 공통 서열 리드는 (i) 하나 이상의 유전자좌의 세트에서 각 유전자좌의 복수의 공통 염기를 포함하고 및 (ii) 이중 가닥 DNA 분자의 단일 가닥을 대표하는 단계;

(f) 하나 이상의 유전자 좌의 세트 내의 각각의 유전자좌에 대해, 상보적 가닥이 복수의 공통 서열 리드에서 검출되는 유전자좌에 매핑되는 농축된 태깅된 가닥의 제1 정량적 측정을 계산하는 단계;

(g) 하나 이상의 유전자 좌의 세트 내의 각각의 유전자좌에 대해, 상보적 가닥 중 한 가닥만이 복수의 공통 서열 리드에서 검출되는 유전자좌에 매핑되는 농축된 태깅된 가닥의 제2 정량적 측정을 계산하는 단계;

(h) 하나 이상의 유전자좌의 세트 내의 각각의 유전자좌에 대해, 상보적 가닥이 복수의 공통 서열 리드에서 검출되지 않는 유전자좌에 맵핑되는 농축된 태깅된 가닥의 제3 정량적 측정을 계산하는 단계로서, 제3 정량적 측정은 제1 및 제2 정량적 측정에 적어도 부분적으로 기초하여 계산되며, 이로써 대상체의 샘플에서 이중 가닥 DNA 분자를 검출하는 단계.

대안적으로 또는 본원에 기재된 방법과 조합하여, 일부 구현예들에서, 방법은, 예컨대 다중 유전체 영역에 대한 농축을 위해 (a)-(b) 중 하나 또는 둘 다를 추가로 포함한다:

(a) 샘플에서 소정의 양의 핵산을 하기를 포함하는 본원에 기재된 복수의 표적 포획 시약과 접촉하도록 하는 단계로서:

(i) 샘플로부터 핵산의 유전체 영역의 제1 세트에 선택적으로 혼성화되는 제1 복수의 표적 포획 시약으로서, 제1의 복수의 표적 포획 시약은 제1의 복수의 표적 포획 시약의 포화점보다 낮은 제1 농도로 제공되는 시약, 및

(ii) 샘플로부터 핵산의 유전체 영역의 제2 세트에 선택적으로 혼성화되는 제2 복수의 표적 포획 시약으로서, 제2의 복수의 표적 포획 시약은 제2의 복수의 표적 포획 시약의 포화점보다 낮은 제2 농도로 제공되는 시약, 및

(b) 제1 세트의 유전체 영역 및 제2 세트의 유전체 영역에 대해 샘플로부터 핵산을 농축하여 농축된 핵산을 생성하는 단계.

대안적으로 또는 본원에 기재된 방법과 조합하여, 일부 구현예들에서, 상기 방법은 (a)-(e) 중 하나 이상(예컨대, 2, 3, 4개 또는 모두)을 추가로 포함한다:

(a) 복수의 표적 포획 시약 혼합물을 제공하는 단계로서, 상기 복수의 표적 포획 시약 혼합물 각각은 제1 세트의 유전체 영역에 선택적으로 혼성화되는 제1 복수의 표적 포획 시약 및 제2 유전체 영역 세트에 선택적으로 혼성화되는 제2 복수 표적 포획 시약을 포함하는 단계로서,

- [0782] 상기 제1의 복수의 표적 포획 시약은 복수의 표적 포획 시약 혼합물에 걸쳐 상이한 농도이고, 제2의 복수의 표적 포획 시약은 복수의 표적 포획 시약 혼합물에 걸쳐 동일한 농도인 단계;
- [0783] (b) 복수의 표적 포획 시약 혼합물 각각을 샘플(예컨대, 혈액 샘플)과 접촉시켜 제1 복수의 표적 포획 시약 및 제2 복수의 표적 포획 시약을 사용하여 샘플로부터 핵산을 포획하는 단계로서, 각각의 표적 포획 시약 혼합물의 제2 복수의 표적 포획 시약은 제2 복수의 표적 포획 시약의 포화점 이상인 제1 농도로 제공되며, 샘플의 핵산은 제1 복수의 표적 포획 시약 및 제2 복수의 표적 포획 시약으로 포획되는 단계;
- [0784] (c) 할당된 수의 서열 리드 내에서 서열 리드 세트를 생성하기 위해 각각의 표적 포획 시약 혼합물로 포획된 핵산의 일부를 서열분석하는 단계;
- [0785] (d) 각각의 표적 포획 시약 혼합물에 대한 제1 복수의 표적 포획 시약 및 제2 복수의 표적 포획 시약에 대한 서열 리드의 리드 정도를 결정하는 단계; 또는
- [0786] (e) 제2 세트의 유전체 영역에 대한 리드 정도를 제공하는 적어도 하나의 표적 포획 시약 혼합물을 확인하는 단계로서;
- [0787] 제2 세트의 유전체 영역에 대한 리드 정도는 적어도 0.0001% 소수 대립유전자 빈도(MAF)의 유전적 변이체의 검출 민감도를 제공한다.
- [0788] 다른 구현예들은 미국특허 제US9,598,731호, US9,834,822호, US9,840,743호, US9,902,992호, US9,920,366호 및 US9,850,523호에 기재되어 있으며, 그 내용은 전체가 참조로서 인용된다.
- [0789] 본원에 기재된 방법의 구현예들에서, 방법의 단계 또는 매개변수는 방법의 하류 단계 또는 매개변수를 수정하는 데 사용된다.
- [0790] 한 구현예에서, 샘플의 특성은 다음 중 하나 이상 또는 모두에서 하류 단계 또는 파라미터를 수정하는 데 사용된다: 상기 샘플로부터 핵산의 분리; 라이브러리 구축; 표적 포획 시약(예컨대, 미끼)의 설계 또는 선택; 혼성화 조건; 서열분석; 리드 매핑; 돌연변이 호출 방법의 선택; 돌연변이 호출; 또는 돌연변이 주석.
- [0791] 한 구현예에서, 분리된 종양 또는 대조군, 핵산의 특성은 다음 중 하나 이상 또는 모두에서 하류 단계 또는 파라미터를 수정하는 데 사용된다: 상기 샘플로부터 핵산의 분리; 라이브러리 구축; 표적 포획 시약(예컨대, 미끼)의 설계 또는 선택; 혼성화 조건; 서열분석; 리드 매핑; 돌연변이 호출 방법의 선택; 돌연변이 호출; 또는 돌연변이 주석.
- [0792] 한 구현예에서, 라이브러리의 특성은 다음 중 하나 이상 또는 모두에서 하류 단계 또는 파라미터를 수정하는 데 사용된다: 상기 샘플로부터 핵산의 분리; 후속 라이브러리 구축; 표적 포획 시약(예컨대, 미끼)의 설계 또는 선택; 혼성화 조건; 서열분석; 리드 매핑; 돌연변이 호출 방법의 선택; 돌연변이 호출; 또는 돌연변이 주석.
- [0793] 한 구현예에서, 라이브러리 캐치의 특성은 다음 중 하나 이상 또는 모두에서 하류 단계 또는 파라미터를 수정하는 데 사용된다: 상기 샘플로부터 핵산의 분리; 후속 라이브러리 구축; 표적 포획 시약(예컨대, 미끼)의 설계 또는 선택; 혼성화 조건; 서열분석; 리드 매핑; 돌연변이 호출 방법의 선택; 돌연변이 호출; 또는 돌연변이 주석.
- [0794] 한 구현예에서, 서열분석 방법의 특성은 다음 중 하나 이상 또는 모두에서 하류 단계 또는 파라미터를 수정하는 데 사용된다: 상기 샘플로부터 핵산의 분리; 후속 라이브러리 구축; 표적 포획 시약(예컨대, 미끼)의 설계 또는 선택; 혼성화 조건의 후속 결정; 후속 서열분석; 리드 매핑; 돌연변이 호출 방법의 선택; 돌연변이 호출; 또는 돌연변이 주석.
- [0795] 한 구현예에서, 매핑된 리드 집합체의 특성은 다음 중 하나 이상 또는 모두에서 하류 단계 또는 파라미터를 수정하는 데 사용된다: 상기 샘플로부터 핵산의 분리; 후속 라이브러리 구축; 표적 포획 시약(예컨대, 미끼)의 설계 또는 선택; 혼성화 조건의 후속 결정; 후속 서열분석; 후속 리드 매핑; 돌연변이 호출 방법의 선택; 돌연변이 호출; 또는 돌연변이 주석.
- [0796] 한 구현예에서, 상기 방법은 샘플 특성에 대한 값을 획득하는 것, 예컨대 상기 샘플에서 종양 세포의 비율에 대해; 상기 샘플의 세포성에 대해; 또는 샘플의 이미지로부터 값을 획득하는 것을 포함한다. 구현예들에서, 상기 방법은 샘플 특성에 대한 상기 획득된 값에 응답하여 다음을 위한 매개변수를 선택하는 단계를 포함한다: 샘플로부터 핵산의 분리, 라이브러리 구축; 표적 포획 시약(예컨대, 미끼)의 설계 또는 선택; 표적 포획 시약(예컨대,

대, 미끼)/라이브러리 핵산 분자 혼성화; 서열분석' 또는 돌연변이 호출.

- [0797] 한 구현예에서, 상기 방법은 상기 샘플에 존재하는 종양 조직의 양에 대한 값을 획득하는 단계, 상기 획득된 값을 참조 기준과 비교하는 단계, 및 상기 참조 기준이 충족되는 경우, 상기 샘플을 수락하는 단계, 예를 들어, 샘플은 30, 40 또는 50% 이상의 종양 세포를 포함한다. 한 구현예에서, 방법은, 예컨대 참고 기준을 충족하지 않는 샘플로부터 상기 샘플의 종양 조직을 거대해부함으로써 종양 세포에 대해 농축된 하위 샘플을 수득하는 단계를 추가로 포함한다.
- [0798] 한 구현예에서, 상기 방법은 상기 샘플에 존재하는 종양 핵산(예컨대, DNA)의 양에 대한 값을 수득하는 단계, 상기 수득된 값을 참조 기준과 비교하는 단계, 및 상기 참조 기준이 충족되는 경우 상기 샘플을 수락하는 단계를 추가로 포함한다. 한 구현예에서, 상기 방법은, 예컨대 참고 기준을 충족하지 않는 샘플로부터 상기 샘플의 종양 조직을 거대해부함으로써 종양 핵산에 대해 농축된 하위 샘플을 수득하는 단계를 추가로 포함한다.
- [0799] 한 구현예에서, 방법은 대상체에 대한 종양 유형, 유전자, 및 유전적 변경(TGA)의 연관성을 제공하는 단계를 추가로 포함한다. 한 구현예에서, 방법은 복수의 요소를 갖는 데이터베이스를 제공하는 단계를 더 포함하고, 여기서 각각의 요소는 TGA를 포함한다.
- [0800] 한 구현예에서, 방법은 대상체의 TGA를 특성화하는 단계를 더 포함하며, 상기 TGA가 데이터베이스, 예를 들어 검증된 TGA의 데이터베이스에 존재하는지 여부를 결정하는 단계; 데이터베이스로부터의 TGA에 대한 정보를 상기 대상체로부터의 상기 TGA(주석)와 연관시키는 단계; 및 선택적으로, 상기 대상체에 대한 제2 또는 후속 TGA가 상기 데이터베이스에 존재하는지 결정하고, 존재한다면 데이터베이스로부터의 두 번째 또는 후속 TGA에 대한 정보를 상기 환자에 존재하는 상기 제2 TGA와 연관시키는 단계를 포함한다. 한 구현예에서, 상기 방법은 보고서를 형성하기 위해 대상체의 TGA의 존재 또는 부재, 및 선택적으로 연관된 주석을 기억하는 단계를 더 포함한다. 한 구현예에서, 방법은 수령자에게 상기 보고서를 전송하는 단계를 추가로 포함한다.
- [0801] 한 구현예에서, 방법은 대상체의 TGA를 특성화하는 단계를 더 포함하며, 상기 TGA가 데이터베이스, 예를 들어 검증된 TGA의 데이터베이스에 존재하는지 여부를 결정하는 단계; 또는 상기 데이터베이스에 없는 TGA가 알려진 임상적으로 관련된 유전자 또는 변경을 가지는지 여부를 결정하고, 그렇다면 상기 데이터베이스에 상기 TGA에 대한 항목을 제공하는 단계를 포함한다. 한 구현예에서, 상기 방법은 보고서를 형성하기 위해 대상체의 DNA에서 발견되는 돌연변이의 존재 또는 부재를 기억하는 단계를 더 포함한다.
- [0802] 예시적인 구현예
- [0803] 하기의 구현예는 예시적인 것이며 본 발명의 범위를 제한하려는 것이 아니다.
- [0804] 구현예 1. 대상체로부터의 샘플의 종양 분획을 결정하는 방법으로서, 상기 방법은:
- [0805] 상기 샘플에서 하위 유전체 간격과 관련된 목표 변수에 대한 값을 수득하는 단계;
- [0806] 상기 목표 변수로부터 확실성 메트릭을 결정하는 단계;
- [0807] 저장된 확실성 메트릭 및 저장된 종양 분획 사이의 결정된 관계에 액세스하는 단계; 및
- [0808] 상기 확실성 메트릭 및 상기 결정된 관계를 참조하여 상기 샘플의 종양 분획을 결정하는 단계를 포함하는, 방법.
- [0809] 구현예 2. 구현예 1에 있어서,
- [0810] 상기 하위 유전체 간격은 적어도 1개의 뉴클레오타이드를 포함하는, 방법.
- [0811] 구현예 3. 구현예 2에 있어서,
- [0812] 상기 적어도 하나의 뉴클레오타이드는 단일 염기 다형성(SNP)과 관련된, 방법.
- [0813] 구현예 4. 구현예 1 내지 구현예 3 중 어느 한 구현예에 있어서,
- [0814] 상기 하위 유전체 간격은 2개 이상의 뉴클레오타이드를 포함하는, 방법.
- [0815] 구현예 5. 구현예 1 내지 구현예 4 중 어느 한 구현예에 있어서,
- [0816] 상기 하위 유전체 간격은 본원에 기재된 유전자의 1개 이상의 뉴클레오타이드를 포함하는, 방법.
- [0817] 구현예 6. 구현예 1 내지 구현예 5 중 어느 한 구현예에 있어서,

- [0818] 상기 확실성 메트릭은 상기 하위 유전체 간격에 대한 예상 로그2비로부터의 편차 또는 상기 하위 유전체 간격에 대한 예상 대립유전자 분획으로부터의 편차 중 하나인, 방법.
- [0819] 구현예 7. 구현예 1 내지 구현예 6 중 어느 한 구현예에 있어서, 예컨대 복수의 하위 유전체 간격에서 목표 변수에 대한 복수의 값이 수득되는, 방법.
- [0820] 구현예 8. 구현예 7에 있어서,
- [0821] 상기 복수의 하위 유전체 간격은 2, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300개 이상의 하위 유전체 간격을 포함하는, 방법.
- [0822] 구현예 9. 구현예 1 내지 구현예 8 중 어느 한 구현예에 있어서,
- [0823] 상기 목표 변수는 상기 샘플의 하위 유전체 간격과 관련된 대립유전자의 존재량의 비교를 포함하는, 방법.
- [0824] 구현예 10. 구현예 1 내지 구현예 9 중 어느 한 구현예에 있어서,
- [0825] 상기 비교는 하나의 대립 유전자의 존재량 및 모든 대립 유전자의 존재량 사이의 비교인, 방법.
- [0826] 구현예 11. 구현예 1 내지 구현예 9 중 어느 한 구현예에 있어서,
- [0827] 상기 비교는 하나의 대립 유전자의 존재량 및 대안적인 대립 유전자의 존재량 사이의 비교인, 방법.
- [0828] 구현예 12. 구현예 1 내지 구현예 11 중 어느 한 구현예에 있어서,
- [0829] 상기 목표 변수는 대립유전자 분획, 또는 모계 및 부계 대립유전자의 존재량에 대비하여 모계 또는 부계 대립유전자의 존재량의 비교(예컨대, 비)를 포함하는, 방법.
- [0830] 구현예 13. 구현예 12에 있어서,
- [0831] 상기 모계 대립유전자는 상기 샘플에서 상기 부계 대립유전자보다 더 풍부한, 방법.
- [0832] 구현예 14. 구현예 12에 있어서,
- [0833] 상기 부계 대립유전자는 상기 샘플에서 상기 모계 대립유전자보다 더 풍부한, 방법.
- [0834] 구현예 15. 구현예 1 내지 구현예 14 중 어느 한 구현예에 있어서,
- [0835] 상기 목표 변수에 대한 값은 0 내지 0.5 사이, 0 내지 1 사이, 또는 0.5 내지 1 사이인, 방법.
- [0836] 구현예 16. 구현예 1 내지 구현예 15 중 어느 한 구현예에 있어서,
- [0837] 상기 목표 변수는 모계 및 부계 대립유전자의 존재량에 대비하여 상기 모계 대립 유전자 또는 상기 부계 대립유전자의 존재량 차이의 비교(예컨대, 비)를 포함하는, 방법.
- [0838] 구현예 17. 구현예 16에 있어서,
- [0839] 상기 모계 대립유전자는 상기 샘플에서 상기 부계 대립유전자보다 더 풍부한, 방법.
- [0840] 구현예 18. 구현예 16에 있어서,
- [0841] 상기 부계 대립유전자는 상기 샘플에서 상기 모계 대립유전자보다 더 풍부한, 방법.
- [0842] 구현예 19. 구현예 1 내지 구현예 18 중 어느 한 구현예에 있어서,
- [0843] 상기 목표 변수는 상기 샘플의 하위 유전체 간격에서의 대립유전자의 존재량과 참조 샘플의 하위유전체 간격에서의 대립유전자의 존재량의 비교를 포함하는, 방법.
- [0844] 구현예 20. 구현예 19에 있어서,
- [0845] 상기 참조 샘플은 건강한 대상체 또는 암을 가지지 않거나 암을 가질 위험이 없는 대상체로부터 수득한, 방법.
- [0846] 구현예 21. 구현예 19 또는 구현예 20에 있어서,
- [0847] 상기 목표 변수는 상기 참조 샘플 내 모계 대립유전자 및 부계 대립유전자의 존재량에 대비하여 상기 샘플 내 상기 모계 대립유전자 및 상기 부계 대립유전자의 존재량의 비교(예컨대, 비)를 포함하는, 방법.
- [0848] 구현예 22. 구현예 19 또는 구현예 20에 있어서,

- [0849] 상기 목표 변수는 상기 참조 샘플 내 모계 대립유전자 및 부계 대립유전자의 존재량에 대비하여 상기 샘플 내 상기 모계 대립유전자 및 상기 부계 대립유전자의 존재량 및 상기 참조 샘플 내 상기 모계 대립유전자 및 상기 부계 대립유전자의 존재량의 비교(예컨대, 비)를 포함하는, 방법.
- [0850] 구현예 23. 구현예 1 내지 구현예 22 중 어느 한 구현예에 있어서,
- [0851] 상기 하위 유전체 간격은 이형접합인(상기 하위 유전체 간격과 관련된 대립유전자의 관점에서), 방법.
- [0852] 구현예 24. 구현예 1 내지 구현예 22 중 어느 한 구현예에 있어서,
- [0853] 상기 하위 유전체 간격은 동형접합, 반접합(semizygous) 또는 반접합(hemizygous)인(상기 하위 유전체 간격과 관련된 대립유전자의 관점에서), 방법.
- [0854] 구현예 25. 구현예 1 내지 구현예 24 중 어느 한 구현예에 있어서,
- [0855] 상기 하위 유전체 간격과 관련된 적어도 하나의 대립유전자는 상기 샘플에서 복제 수 변경에 관여하는, 예컨대 증폭되는, 방법.
- [0856] 구현예 26. 구현예 1 내지 구현예 25 중 어느 한 구현예에 있어서,
- [0857] 상기 확실성 메트릭은 편차 메트릭, 예컨대 본원에 기재된 편차 메트릭, 또는 임의의 p 모멘트 또는 이들의 조합인, 방법.
- [0858] 구현예 27. 구현예 26에 있어서,
- [0859] 상기 편차 메트릭은 참조값, 예컨대 본원에 기재된 기대값으로부터 상기 목표 변수에 대한 값의 편차를 측정하는, 방법.
- [0860] 구현예 28. 구현예 26 또는 구현예 27에 있어서,
- [0861] 상기 확실성 메트릭은 기대비(예컨대, 0.5)로부터 모계 및 부계 대립유전자의 존재량에 대비하여 모계 또는 부계 대립유전자의 존재량의 비의 편차를 측정하는, 방법.
- [0862] 구현예 29. 구현예 26 내지 구현예 28 중 어느 한 구현예에 있어서,
- [0863] 상기 편차 메트릭은 기대비(예컨대, 0)로부터 모계 및 부계 대립유전자의 존재량에 대비하여 상기 모계 또는 부계 대립유전자의 존재량 차이의 비의 편차를 측정하는, 방법.
- [0864] 구현예 30. 구현예 26 내지 구현예 29 중 어느 한 구현예에 있어서,
- [0865] 상기 확실성 메트릭은 기대비(예컨대, 0)로부터 상기 참조 샘플 내 모계 대립유전자 및 부계 대립유전자의 존재량에 대비하여 상기 샘플 내 상기 모계 대립유전자 또는 상기 부계 대립유전자의 존재량의 비의 편차를 측정하는, 방법.
- [0866] 구현예 31. 구현예 30에 있어서,
- [0867] 상기 비는 로그비, 예컨대 로그2비를 포함하는, 방법.
- [0868] 구현예 32. 구현예 26 내지 구현예 31 중 어느 한 구현예에 있어서,
- [0869] 상기 편차 메트릭은 기대비(예컨대, 0)로부터 상기 참조 샘플 내 모계 대립유전자 및 부계 대립유전자의 존재량에 대비하여 상기 샘플 내 상기 모계 대립유전자 및 상기 부계 대립유전자의 존재량 및 상기 참조 샘플 내 상기 모계 대립유전자 및 상기 부계 대립유전자의 존재량 차이의 비의 편차를 측정하는, 방법.
- [0870] 구현예 33. 구현예 26 내지 구현예 32 중 어느 한 구현예에 있어서,
- [0871] 상기 편차 메트릭은 평균 제곱근(p=2 모멘트) 편차 메트릭 또는 p 모멘트 변동 메트릭의 임의의 조합을 포함하는, 방법.
- [0872] 구현예 34. 구현예 26 내지 구현예 32 중 어느 한 구현예에 있어서,
- [0873] 상기 편차 메트릭은 로그2비 메트릭을 포함하는, 방법.
- [0874] 구현예 35. 구현예 26 내지 구현예 32 중 어느 한 구현예에 있어서,
- [0875] 상기 편차 메트릭은 평균 제곱근(p=2 모멘트) 편차 메트릭 또는 p 모멘트 변동 메트릭의 임의의 조합을 포함하

는, 방법.

- [0876] 구현예 36. 구현예 1 내지 구현예 25 중 어느 한 구현예에 있어서,
- [0877] 상기 확실성 메트릭은 참조값, 예컨대 기대값으로부터 상기 목표 변수에 대한 값의 편차를 측정하지 않는, 방법.
- [0878] 구현예 37. 구현예 1 내지 구현예 25 또는 구현예 36 중 어느 한 구현예에 있어서,
- [0879] 상기 확실성 메트릭은 엔트로피 메트릭, 예컨대 목표 변수의 상대적 확실성을 본질적으로 측정하는 메트릭, 예컨대 본원에 기재된 엔트로피 메트릭, 또는 임의의 p 모멘트 또는 이들의 조합인, 방법.
- [0880] 구현예 38. 구현예 37에 있어서,
- [0881] 상기 엔트로피 메트릭은 모계 및 부계 대립유전자의 존재량에 대비하여 모계 또는 부계 대립유전자의 존재량의 비의 확실성을 측정하는, 방법.
- [0882] 구현예 39. 구현예 37 또는 구현예 38에 있어서,
- [0883] 상기 엔트로피 메트릭은 상기 참조 샘플 내 모계 대립유전자 및 부계 대립유전자의 존재량에 대비하여 상기 샘플 내 상기 모계 대립유전자 및 상기 부계 대립유전자의 존재량의 비의 확실성을 측정하는, 방법.
- [0884] 구현예 40. 구현예 39에 있어서,
- [0885] 상기 비는 로그비, 예컨대 로그<sub>2</sub>비를 포함하는, 방법.
- [0886] 구현예 41. 구현예 1 내지 구현예 40 중 어느 한 구현예에 있어서,
- [0887] 예컨대 상기 하위 유전체 간격에서 대립유전자의 존재량을 결정하기 위해, 예컨대 차세대 서열분석(NGS)에 의해 상기 샘플을 서열분석하는 단계를 추가로 포함하는, 방법.
- [0888] 구현예 42. 구현예 1 내지 구현예 41 중 어느 한 구현예에 있어서,
- [0889] 상기 확실성 메트릭은, 예컨대 서열분석이 상기 대립유전자의 존재량을 결정하는 데 사용되는 경우 상기 하위 유전체 간격에서 대립유전자 범위의 함수인, 방법.
- [0890] 구현예 43. 구현예 1 내지 구현예 41 중 어느 한 구현예에 있어서,
- [0891] 예컨대 유전체 유전자좌에서 대립유전자의 존재량을 결정하기 위해, 상기 샘플 상에서 배열 혼성화를 실시하는 단계를 추가로 포함하는, 방법.
- [0892] 구현예 44. 구현예 43에 있어서,
- [0893] 상기 확실성 메트릭은, 예컨대 배열 혼성화가 상기 대립유전자의 존재량을 결정하는 데 사용되는 경우 상기 하위 유전체 간격에서 대립유전자 강도의 함수인, 방법.
- [0894] 구현예 45. 구현예 1 내지 구현예 44 중 어느 한 구현예에 있어서,
- [0895] 상기 하위 유전체 간격은 이의 예상되는 대립유전자 분획에 기초하여 선택되는, 방법.
- [0896] 구현예 46. 구현예 45에 있어서,
- [0897] 상기 예상되는 대립유전자 분획은 건강한 모집단 내 하위 집합의 개체들에서 .50 대립유전자 분획인, 방법.
- [0898] 구현예 47. 구현예 45에 있어서,
- [0899] 상기 예상되는 대립유전자 분획은 비정상적인 종양 성장을 갖는 대상체에서 0, .50, 또는 1 이외인, 방법.
- [0900] 구현예 48. 구현예 1 내지 구현예 47 중 어느 한 구현예에 있어서,
- [0901] 상기 하위 유전체 간격은 이의 각각의 대립유전자 위치에 기초하여 선택되고, 상기 각각의 대립유전자 위치는 특정 질병 온톨로지를 갖는 대상체에서 .50 이외의 대립유전자 분획을 가질 것으로 예상되는, 방법.
- [0902] 구현예 49. 구현예 48에 있어서,
- [0903] 상기 특정 질병 온톨로지는 암 병태 또는 전암 병태 중 하나인, 방법.

- [0904] 구현예 50. 구현예 1 내지 구현예 49 중 어느 한 구현예에 있어서,
- [0905] 임상 표본(또는 세포주, 또는 인실리코(in silico) 시뮬레이션된 샘플 세트)로부터 수득한 정보의 훈련 데이터 세트에 액세스하는 단계로서, 상기 정보는 대상체 집단으로부터 저장된 확실성 메트릭 및 저장된 종양 분획 사이의 복수의 관계를 포함하는 단계; 및
- [0906] 상기 저장된 확실성 메트릭 및 상기 저장된 종양 분획 사이의 결정된 관계를 결정하기 위해 상기 훈련 데이터 세트에 기계 학습 공정을 적용하는 단계를 추가로 포함하는, 방법.
- [0907] 구현예 51. 컴퓨터 시스템으로서:
- [0908] 저장된 확실성 메트릭 및 저장된 종양 분획 사이의 결정된 관계를 저장하도록 구성된 데이터베이스;
- [0909] 프로세서; 및
- [0910] 상기 프로세서에 통신 가능하게 결합되고, 상기 프로세서에 의해 실행될 때 상기 프로세서가 하기의 단계를 실시하도록 하는 명령어를 포함하는 메모리를 포함하는 컴퓨터 시스템으로서, 상기 명령어들은:
- [0911] 상기 샘플 내 하위 유전체 간격에서 목표 변수에 대한 값을 수득하는 단계;
- [0912] 상기 목표 변수로부터 확실성 메트릭을 결정하는 단계;
- [0913] 상기 데이터베이스에서 상기 저장된 확실성 메트릭 및 상기 저장된 종양 분획 간의 결정된 관계에 액세스하는 단계; 및
- [0914] 상기 확실성 메트릭 및 상기 결정된 관계를 참조하여 상기 샘플의 종양 분획을 결정하는 단계를 실시하도록 하는, 컴퓨터 시스템.
- [0915] 구현예 52. 구현예 51에 있어서,
- [0916] 상기 메모리는 상기 프로세서에 의해 실행될 때 상기 프로세서가 하기의 단계를 실시하도록 하는 명령어를 추가로 포함하는 컴퓨터 시스템으로서,
- [0917] 임상 표본(또는 세포주, 또는 인실리코(in silico) 시뮬레이션된 샘플 세트)로부터 수득한 정보의 훈련 데이터 세트에 액세스하는 단계로서, 상기 정보는 저장된 확실성 메트릭 및 대응하는 저장된 종양 분획 사이의 복수의 관계를 포함하고, 상기 복수의 관계는 대상체 집단으로부터 결정되는 단계; 및
- [0918] 상기 저장된 확실성 메트릭 및 상기 대응하는 저장된 종양 분획 사이의 결정된 관계를 결정하기 위해 상기 훈련 데이터 세트에 기계 학습 공정을 적용하는 단계를 실시하도록 하는, 컴퓨터 시스템.
- [0919] 구현예 53. 대상체에서 질병을 치료하는 방법으로서, 상기 방법은:
- [0920] 종양 분획의 추정에 반응하여 유효량의 요법을 상기 대상체에게 투여함으로써 상기 질병을 치료하는 단계를 포함하는 방법으로서,
- [0921] 상기 종양 분획의 추정:
- [0922] 상기 대상체로부터 샘플 내 하위 유전체 간격에서 목표 변수에 대한 값을 수득하는 단계;
- [0923] 상기 목표 변수로부터 확실성 메트릭을 결정하는 단계;
- [0924] 저장된 확실성 메트릭 및 저장된 종양 분획 사이의 결정된 관계에 액세스하는 단계; 및
- [0925] 상기 확실성 메트릭 및 상기 결정된 관계를 참조하여 상기 샘플의 종양 분획을 결정하는 단계를 포함하는, 방법.
- [0926] 구현예 54. 대상체에서 질병을 평가하는 방법으로서, 상기 방법은:
- [0927] 상기 대상체로부터 제1 샘플 내 하위 유전체 간격에서 목표 변수에 대한 제1 값을 수득하는 단계;
- [0928] 상기 목표 변수로부터 제1 확실성 메트릭을 결정하는 단계;
- [0929] 저장된 확실성 메트릭 및 저장된 종양 분획 사이의 결정된 관계에 액세스하는 단계; 및
- [0930] 상기 확실성 메트릭 및 상기 결정된 관계를 참조하여 상기 제1 샘플의 종양 분획을 결정하는 단계;

- [0931] 상기 대상체로부터 제2 샘플 내 하위 유전체 간격에서 목표 변수에 대한 제2 값을 획득하는 단계;
- [0932] 상기 목표 변수로부터 제2 확실성 메트릭을 결정하는 단계;
- [0933] 상기 제2 확실성 메트릭 및 상기 결정된 관계를 참조하여 상기 제2 샘플의 종양 분획을 결정하는 단계; 및
- [0934] 상기 제1 샘플의 종양 분획을 상기 제2 샘플의 종양 분획과 비교함으로써, 상기 대상체에서 상기 질병을 평가하는 단계를 포함하는, 방법.
- [0935] 구현예 55. 구현예 54에 있어서,
- [0936] 상기 제1 샘플은 제1 시점에서 취하고, 상기 제2 샘플은 제2 시점에서 취하는, 방법.
- [0937] 구현예 56. 구현예 55에 있어서,
- [0938] 상기 제1 시점은 상기 대상체에게 요법이 투여되기 전이고, 상기 제2 시점은 상기 대상체에게 상기 요법이 투여된 이후인, 방법.
- [0939] 구현예 57. 대상체를 평가하는 방법으로서, 상기 방법은:
- [0940] 상기 대상체로부터 샘플 내 하위 유전체 간격에서 목표 변수에 대한 값을 획득하는 단계;
- [0941] 상기 목표 변수로부터 확실성 메트릭을 결정하는 단계;
- [0942] 저장된 확실성 메트릭 및 저장된 종양 분획 사이의 결정된 관계에 액세스하는 단계; 및
- [0943] 상기 확실성 메트릭 및 상기 결정된 관계를 참조하여 상기 샘플의 종양 분획을 결정함으로써 상기 대상체를 평가하는 단계를 포함하는, 방법.
- [0944] 구현예 58. 요법을 평가하는 방법으로서, 상기 방법은:
- [0945] 요법을 투여받은 대상체로부터 샘플 내 하위 유전체 간격에서 목표 변수에 대한 값을 획득하는 단계;
- [0946] 상기 목표 변수로부터 확실성 메트릭을 결정하는 단계;
- [0947] 저장된 확실성 메트릭 및 저장된 종양 분획 사이의 결정된 관계에 액세스하는 단계; 및
- [0948] 상기 확실성 메트릭 및 상기 결정된 관계를 참조하여 상기 샘플의 종양 분획을 결정함으로써 상기 투여된 요법의 효능을 평가하는 단계를 포함하는, 방법.
- [0949] 구현예 59. 보고서를 제공하는 방법으로서, 상기 방법은:
- [0950] 대상체로부터 샘플 내 하위 유전체 간격에서 목표 변수에 대한 값을 획득하는 단계;
- [0951] 상기 목표 변수로부터 확실성 메트릭을 결정하는 단계;
- [0952] 저장된 확실성 메트릭 및 저장된 종양 분획 사이의 결정된 관계에 액세스하는 단계; 및
- [0953] 상기 확실성 메트릭 및 상기 결정된 관계를 참조하여 상기 샘플의 종양 분획을 결정하는 단계; 및
- [0954] 보고서에서 상기 종양 분획을 기록하는 단계를 포함하는, 방법.
- [0955] 구현예 60. 대상체에서 생검을 평가하는 방법으로서, 상기 방법은:
- [0956] 상기 대상체로부터 생검 내 하위 유전체 간격에서 목표 변수에 대한 값을 획득하는 단계;
- [0957] 상기 목표 변수로부터 확실성 메트릭을 결정하는 단계;
- [0958] 저장된 확실성 메트릭 및 저장된 종양 분획 사이의 결정된 관계에 액세스하는 단계; 및
- [0959] 상기 확실성 메트릭 및 상기 결정된 관계를 참조하여 상기 생검의 종양 분획을 결정함으로써 상기 생검을 평가하는 단계를 포함하는, 방법.
- [0960] 구현예 61. 구현예 1 내지 구현예 60 중 어느 한 구현예에 있어서,
- [0961] 상기 대상체는 암을 가지거나, 암을 가질 위험이 있거나, 암을 가질 수 있는, 시스템 또는 방법.
- [0962] 구현예 62. 구현예 61에 있어서,

- [0963] 상기 암은 고형암인, 시스템 또는 방법.
- [0964] 구현예 63. 구현예 61에 있어서,
- [0965] 상기 암은 혈액암, 예컨대 백혈병 또는 림프종인, 시스템 또는 방법.
- [0966] 구현예 64. 구현예 1 내지 구현예 63 중 어느 한 구현예에 있어서,
- [0967] 상기 샘플은 액체 샘플, 예컨대 혈액 또는 혈청 샘플인, 방법.
- [0968] 구현예 65. 구현예 1 내지 구현예 63 중 어느 한 구현예에 있어서,
- [0969] 상기 샘플은 고체 샘플, 예컨대 FFPE 샘플인, 시스템 또는 방법.
- [0970] 구현예 66. 구현예 1 내지 구현예 63 중 어느 한 구현예에 있어서,
- [0971] 상기 샘플은 무세포 DNA(cfDNA) 또는 순환 종양 DNA(ctDNA)를 포함하는, 시스템 또는 방법.
- [0972] 구현예 67. 구현예 1 내지 구현예 66 중 어느 한 구현예에 있어서,
- [0973] 상기 대상체는 적어도 하나의 질병에 대해 모니터링을 받고 있는, 시스템 또는 방법.
- [0974] 구현예 68. 구현예 1 내지 구현예 67 중 어느 한 구현예에 있어서,
- [0975] 상기 대상체는 적어도 하나의 질병에 대해 진단을 받고 있는, 시스템 또는 방법.
- [0976] 구현예 69. 구현예 1 내지 구현예 67 중 어느 한 구현예에 있어서,
- [0977] 상기 대상체는 .30 이하의 예상 종양 분획을 갖는, 시스템 또는 방법.
- [0978] 구현예 70. 구현예 1 내지 구현예 69 중 어느 한 구현예에 있어서,
- [0979] 상기 대상체로부터의 샘플의 종양 분획에 기초하여 대상체에 대한 치료를 결정하는 단계를 추가로 포함하는, 시스템 또는 방법.
- [0980] 구현예 71. 구현예 70에 있어서,
- [0981] 상기 대상체에게 치료를 투여하는 단계를 추가로 포함하는, 시스템 또는 방법.
- [0982] 구현예 72. 대상체에서 종양 함량을 발견하는 방법으로서, 상기 방법은:
- [0983] 상기 대상체로부터 생검 내 하위 유전체 간격에서 목표 변수에 대한 값을 수득하는 단계;
- [0984] 상기 목표 변수로부터 확실성 메트릭을 결정하는 단계;
- [0985] 저장된 확실성 메트릭 및 저장된 종양 분획 사이의 결정된 관계에 액세스하는 단계; 및
- [0986] 상기 확실성 메트릭 및 상기 결정된 관계를 참조하여 상기 샘플의 종양 분획을 결정함으로써 상기 대상체 내 종양 함량을 발견하는 단계를 포함하는, 방법.
- [0987] 참조에 의한 인용
- [0988] 본원에 언급된 모든 간행물, 특허 및 특허 출원은 마치 각각의 개별 간행물, 특허 또는 특허 출원이 참조로 포함되는 것으로 구체적이고 개별적으로 표시된 것처럼 그 전체가 참조로서 인용된다. 상충하는 경우, 본원의 정의를 포함하여 본 출원이 우선한다.
- [0989] 공개 데이터베이스의 항목과 관련된 등록 번호를 참조하는 모든 폴리뉴클레오타이드 및 폴리펩티드 서열, 예컨대 웹사이트 [tigr.org](http://tigr.org)의 유전체연구소(The Institute for Genomic Research, TIGR) 및/또는 웹사이트 [ncbi.nlm.nih.gov](http://ncbi.nlm.nih.gov)의 국립 생명공학 정보 센터(NCBI)에서 유지관리 되는 것들 또한 그 전체가 참조에 의해 인용된다.
- [0990] 상호 작용
- [0991] 본원에 기재된 본 발명(들)의 방법 단계는 다른 의미가 명시적으로 제공되거나 문맥에서 명백하지 않는 한, 하나 이상의 다른 당사자 또는 실체가 단계를 수행하도록 하는 임의의 적절한 방법을 포함하도록 의도된다. 상기 당사자 또는 단계는 다른 당사자 또는 실체의 지시나 통제를 받을 필요가 없으며 특정 관할 구역에 위치할 필요도 없다. 따라서, 예를 들어, "제1 숫자를 제2 숫자에 추가하는"에 대한 설명 또는 인용에는 하나 이상의 당사

자 또는 실체가 두 숫자를 함께 추가하도록 하는 것이 포함된다. 예를 들어, 만약 X라는 사람이 Y라는 사람과 정상 거래를 하여 두 숫자를 더하고 Y가 실제로 두 숫자를 더하면 X와 Y 모두 다음과 같이 단계를 수행한다. 즉, 사람 Y는 그가 실제로 숫자를 더했다는 사실 때문에, 사람 X는 사람 Y가 숫자를 더하도록 했다는 사실 때문이다. 또한, 사람 X가 미국 내에 있고 사람 Y가 미국 밖에 있는 경우, 사람 X가 실시되는 단계에 참여함으로써 상기 방법은 미국에서 수행된다.

[0992] 등가물

[0993] 당업자는 단지 일상적인 실험을 사용하여 여기에서 설명된 본 발명의 특정 구체예에 대한 많은 등가물을 인식하거나 확인할 수 있을 것이다. 이러한 등가물은 다음 청구범위에 포함되도록 의도된다.

[0994] 실시예

[0995] 최대 체세포 대립 유전자 빈도(MSAF) 및 대립 유전자 분획(AF)은 Clark et al., *Analytical Validation of a Hybrid Capture-Based Next-Generation Sequencing Clinical Assay for Genomic Profiling of Cell-Free Circulating Tumor DNA*, J. Molecular Diagnostics, vol. 20, pp. 686-702 (2018)에 일반적으로 설명된 방법을 사용하여 TP53 하위 유전체 간격 내 SNP 유전자좌에 걸친 HCC1954 및 HCC1143 세포 배양의 배양물에 대해 결정되었다. MSAF는 각 샘플의 종양 분획에 대한 프록시로 사용되었다. 상이한 종양 분획(즉, MSAF)을 수득하기 위해, 세포주를 한 쌍의 정상 DNA로 연속적으로 희석하였다. 모든 대립유전자 빈도에 대한 확률 분포 함수(PDF)가 각 샘플 세포 배양에 대해 결정되었고, 각 PDF에 대한 해당 엔트로피가 결정되었다.

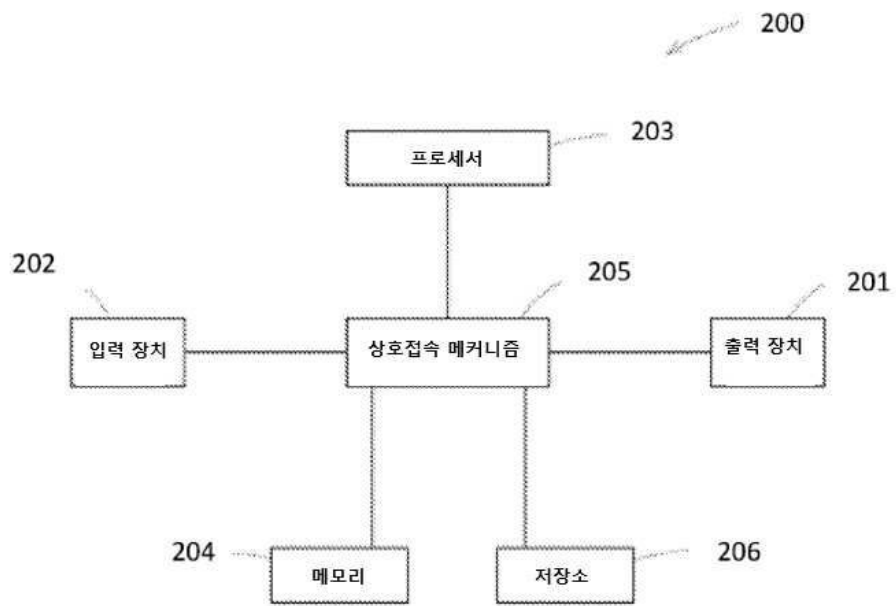
[0996] 종양 분획(MSAF 프록시로 표시됨)은 도 4에 도시된 바와 같이 각 세포에 대해 결정된 엔트로피에 대해 플롯팅되었다. 확률 분포 함수의 엔트로피와 0.05% 초과인 종양 분획에 대한 종양 분획의 로그 사이의 선형 관계가 결정되었다.

## 도면

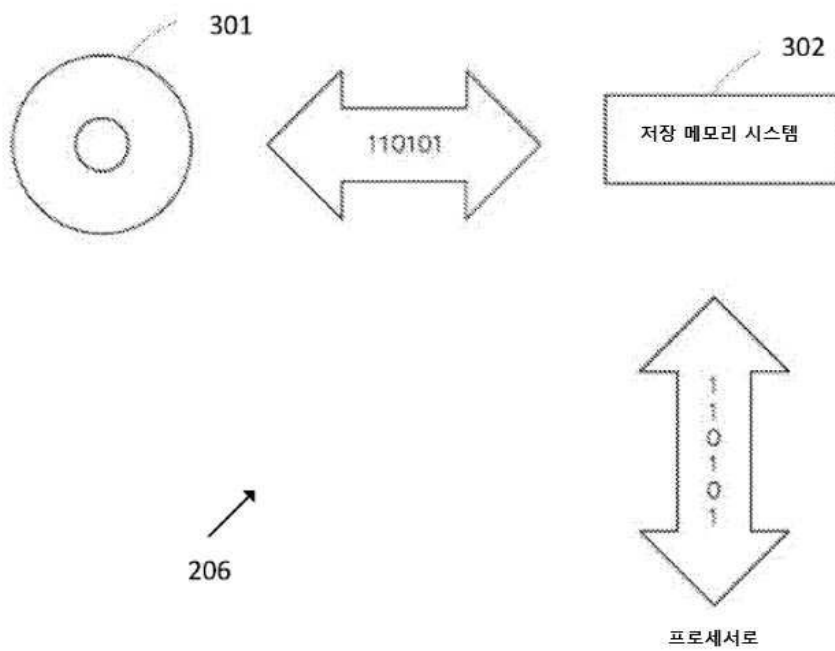
### 도면1



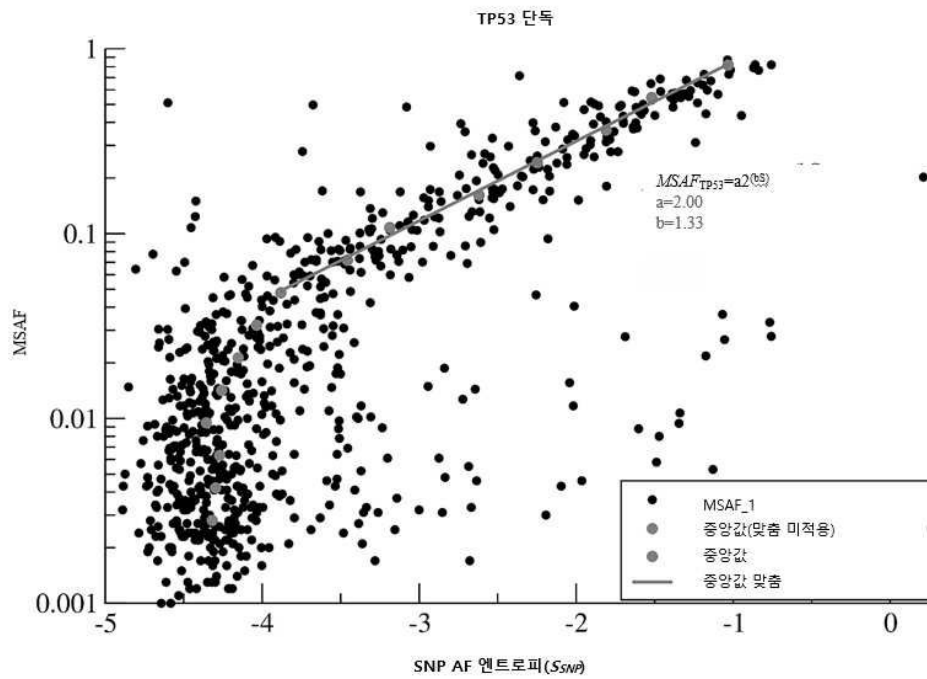
도면2



도면3



#### 도면4



#### 서열 목록

##### SEQUENCE LISTING

<110> Foundation Medicine, Inc.

<120> SYSTEMS AND METHODS FOR EVALUATING TUMOR FRACTION

<130> 19710-20028.40

<140> PCT/US2020/033821

<141> 2020-05-20

<150> US 62/850,474

<151> 2019-05-29

<160> 1

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 150

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic Construct

<220>

<221> misc\_feature

<222> 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30,

31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44,

45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58,

59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72,

73

<223> n = A,T,C or G

<220>

<221> misc\_feature

<222> 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88,

89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102,

103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114,

115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125

<223> n = A,T,C or G

<220>

<221> misc\_feature

<222> 126, 127, 128, 129, 130, 131, 132, 133, 134, 135

<223> n = A,T,C or G

<400> 1

atgcaccag cgtgtnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120

nnnnnnnnnn nnnncactg cggctcctca 150