(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0123159
(43) 공개일자 2020년10월28일

(51) 국제특허분류(Int. Cl.)
C12Q 1/6869 (2018.01) *C12Q 1/6858* (2018.01)
C12Q 1/6886 (2018.01)
(52) CPC특허분류
C12Q 1/6869 (2018.05)
C12Q 1/6858 (2018.05)
(21) 출원번호 10-2020-7026362
(22) 출원일자(국제) 2019년02월13일
심사청구일자 없음
(85) 번역문제출일자 2020년09월11일
(86) 국제출원번호 PCT/US2019/017908
(87) 국제공개번호 WO 2019/160998
국제공개일자 2019년08월22일
(30) 우선권주장
62/630,228 2018년02월13일 미국(US)
62/737,097 2018년09월26일 미국(US)

(71) 출원인
트윈스트랜드 바이오사이언시스, 인코포레이티드
미국, 워싱턴 98121, 시애틀, 스위트 750, 3131 엘리엇 애비뉴
(72) 발명자
솔크 제시 제이.
미국, 워싱턴 98125, 시애틀, 10704 덜랜드 애비뉴 엔이
발렌타인 찰스 클린턴 3세
미국, 워싱턴 98122, 시애틀, 963 22 애비뉴
(74) 대리인
특허법인한얼

전체 청구항 수 : 총 65 항

(54) 발명의 명칭 **유전독성을 검출하고 평가하기 위한 방법 및 시약**

(57) 요 약

유전독성을 평가하기 위한 방법, 시스템 및 시약을 갖는 키트가 본원에 개시된다. 유전독성 및 이의 작용 기전은 대상체의 노출의 몇일 내에 결정될 수 있다. 본 기술내용의 일부 실시형태는 노출된 대상체에서 화합물(예를 들어, 화학적 화합물)의 유전독성 가능성을 평가하기 위해 듀플렉스 시퀀싱을 사용하는 것에 관한 것이다. 본 기술내용의 다른 실시형태는 유전독성 물질과 연관된 돌연변이 서명; 및/또는 유전독소 노출의 안전한 역치 수준을 검출하기 위해 듀플렉스 시퀀싱을 사용하는 것에 관한 것이다. 본 기술내용의 추가 실시형태는 대상체의 DNA 돌연변이 스펙트럼을 알려진 돌연변이성 화합물의 돌연변이 스펙트럼과 비교함으로써 대상체가 노출될 수 있는 하나 이상의 유전독성 물질을 확인하는 것에 관한 것이다. 대상체에서 유전독소 노출이 확인되거나 확증되면, 이후 예방학적 및/또는 억제성 치료학적 치료 과정이 제공된다.

(52) CPC특허분류
C12Q 1/6886 (2018.05)
C12Q 2525/191 (2013.01)
C12Q 2600/142 (2013.01)

명 세 서

청구범위

청구항 1

돌연변이원에 대한 대상체의 노출 후에 대상체에서 생체내 발생한 게놈 돌연변이를 검출하고 정량화하는 방법으로서,

상기 대상체로부터의 샘플을 제공하는 단계이되, 상기 샘플은 이중-가닥 DNA 분자를 포함하는 상기 제공하는 단계;

샘플에서 복수의 이중-가닥 DNA 분자의 각각에 대해 오류-보정된 서열 리드를 생성하는 단계이되,

어댑터-DNA 분자의 원래의 제1 가닥의 카피의 세트 및 어댑터-DNA 분자의 원래의 제2 가닥의 카피의 세트를 생성하는 단계;

원래의 제1 가닥의 카피의 세트 및 원래의 제2 가닥의 카피의 세트를 시퀀싱하여 제1 가닥 서열 및 제2 가닥 서열을 제공하는 단계; 및

제1 가닥 서열과 제2 가닥 서열을 비교하여 제1 가닥 서열과 제2 가닥 서열 사이의 하나 이상의 관련성을 확인하는 단계를 포함하는 상기 생성하는 단계; 및

하나 이상의 관련성을 분석하여 샘플에서의 이중-가닥 DNA 분자에 대한 돌연변이 스펙트럼을 결정하는 단계를 포함하는 방법.

청구항 2

제1항에 있어서, 시퀀싱된 듀플렉스 염기-쌍마다 고유한 돌연변이의 수를 계산함으로써 표적 이중-가닥 DNA 분자에 대한 돌연변이체 빈도를 계산하는 단계를 추가로 포함하는, 방법.

청구항 3

제1항에 있어서, 상기 표적 이중-가닥 DNA 분자는 대상체의 간, 비장, 혈액, 폐 또는 골수로부터 추출되는, 방법.

청구항 4

제1항에 있어서, 상기 대상체는 표적 이중-가닥 DNA 분자가 대상체로부터 제거되기 30일 이하 전에 돌연변이원에 노출되는, 방법.

청구항 5

제1항에 있어서, 상기 돌연변이 스펙트럼은 비지도된 계층적 돌연변이 스펙트럼 클러스터링에 의해 생성되는, 방법.

청구항 6

제1항에 있어서, 상기 돌연변이 스펙트럼은 삼중항 돌연변이 스펙트럼인, 방법.

청구항 7

제1항에 있어서, 복수의 이중-가닥 DNA 분자의 각각에 대해 오류-보정된 서열 리드를 생성하는 단계는 하나 이상의 표적화된 게놈 영역의 오류-보정된 서열 리드를 생성하는 것을 포함하는, 방법.

청구항 8

제7항에 있어서, 상기 하나 이상의 표적화된 게놈 영역은 게놈에서의 돌연변이-경향성 부위인, 방법.

청구항 9

제7항에 있어서, 상기 하나 이상의 표적화된 게놈 영역은 알려진 암 유발자 유전자인, 방법.

청구항 10

제1항에 있어서, 상기 대상체는 형질전환 동물이고, 상기 표적 이중-가닥 DNA 분자의 적어도 일부는 전이유전자의 하나 이상의 부분을 포함하는, 방법.

청구항 11

제1항에 있어서, 상기 대상체는 비형질전환 동물이고, 상기 표적 이중-가닥 DNA 분자는 내인성 게놈 영역을 포함하는, 방법.

청구항 12

제1항에 있어서, 상기 대상체는 인간이고, 상기 표적 이중-가닥 DNA 분자는 인간에서 채혈된 혈액으로부터 추출되는, 방법.

청구항 13

시험 물질의 돌연변이성 서명을 생성하는 방법으로서,

상기 시험 물질에 노출된 시험 대상체로부터 추출된 DNA 단편을 듀플렉스 시퀀싱하는 단계; 및

상기 시험 물질의 돌연변이성 서명을 생성하는 단계이되,

시퀀싱된 듀플렉스 염기-쌍마다 고유한 돌연변이의 수를 계산함으로써 복수의 DNA 단편에 대한 돌연변이체 빈도를 계산하는 단계; 및

복수의 DNA 단편에 대한 돌연변이 패턴을 결정하는 단계이되, 돌연변이 패턴은 돌연변이 유형, 돌연변이 트리뉴클레오타이드 상황 및 돌연변이의 게놈 분포를 포함하는 결정하는 단계를 포함하는 상기 생성하는 단계를 포함하는, 방법.

청구항 14

제13항에 있어서, 상기 시험 물질의 돌연변이 서명을 하나 이상의 알려진 유전독소의 돌연변이 서명과 비교하는 단계를 추가로 포함하는, 방법.

청구항 15

제13항에 있어서, 상기 시험 물질의 돌연변이 서명은 조직 유형, 시험 물질에 대한 노출의 수준, 게놈 영역 및 대상체 유형 중 하나 이상에 기초하여 변하는, 방법.

청구항 16

제15항에 있어서, 상기 대상체 유형은 배양물에서 성장한 인간 세포인, 방법.

청구항 17

제13항에 있어서, 상기 시험 동물은 동물이 희생되기 30일 이하 전에 시험 화합물에 노출되는, 방법.

청구항 18

제13항에 있어서, 상기 돌연변이성 서명은 컴퓨터를 사용한 패턴 매칭에 의해 생성되는, 방법.

청구항 19

제13항에 있어서, 상기 돌연변이 서명은 삼중항 돌연변이 서명인, 방법.

청구항 20

제13항에 있어서, DNA 단편의 듀플렉스 시퀀싱은 하나 이상의 표적화된 게놈 영역의 듀플렉스 시퀀싱을 포함하는, 방법.

청구항 21

제20항에 있어서, 상기 하나 이상의 표적화된 게놈 영역은 게놈에서의 돌연변이-경향성 부위인, 방법.

청구항 22

제20항에 있어서, 상기 하나 이상의 표적화된 게놈 영역은 알려진 암 유발자 유전자인, 방법.

청구항 23

제13항에 있어서, 상기 시험 동물은 형질전환 동물이고, 상기 DNA 단편의 적어도 일부는 전이유전자의 하나 이상의 부분을 포함하는, 방법.

청구항 24

제13항에 있어서, 상기 시험 동물은 비형질전환 동물이고, 상기 DNA 단편은 내인성 게놈 영역을 포함하는, 방법.

청구항 25

시험 물질의 유전독성 가능성을 평가하는 방법으로서,

(a) 상기 시험 물질에 노출된 생물학적 소스로부터 복수의 이중-가닥 DNA 단편을 포함하는 샘플로부터 시퀀싱 라이브러리를 제조하는 단계이되, 서열 라이브러리의 제조는 비대칭적 어댑터 분자를 복수의 이중-가닥 DNA 단편에 결찰하여 복수의 어댑터-DNA 분자를 생성하는 것을 포함하는 상기 제조하는 단계;

(b) 상기 어댑터-DNA 분자의 제1 가닥 및 제2 가닥을 시퀀싱하여 각각의 어댑터-DNA 분자에 대한 제1 가닥 서열 리드 및 제2 가닥 서열 리드를 제공하는 단계;

(c) 각각의 어댑터-DNA 분자에 대해, 상기 제1 가닥 서열 리드와 제2 가닥 서열 리드를 비교하여 제1 가닥 서열 리드와 제2 가닥 서열 리드 사이의 하나 이상의 관련성을 확인하는 단계; 및

(d) 각각의 어댑터-DNA 분자에 대해 제1 가닥 서열 리드와 제2 가닥 서열 리드 사이의 하나 이상의 관련성을 확인함으로써 시험 물질의 돌연변이 서명을 결정하여 샘플에서 돌연변이 패턴, 돌연변이 유형, 돌연변이체 빈도, 돌연변이 유형 분포 및 돌연변이의 게놈 분포 중 적어도 하나를 결정하는 단계; 및

(e) 상기 시험 물질의 돌연변이 서명을 알려진 유전독소로부터 유래된 복수의 돌연변이 스펙트럼과 비교하여 돌연변이 서명이 알려진 유전독소로부터의 돌연변이 스펙트럼과 충분히 유사한지를 결정하는 단계; 또는

(f) 상기 돌연변이체 빈도, 돌연변이 유형 또는 돌연변이 유형 분포 중 적어도 하나가 안전한 역치 수준보다 높은지를 평가하는 단계; 또는

(g) 상기 돌연변이체 빈도가 안전한 역치 돌연변이체 빈도를 초과하는지를 결정하는 단계를 포함하는, 방법.

청구항 26

제25항에 있어서, 상기 시험 물질의 돌연변이 서명은 안전 역치 빈도 초과의 돌연변이체 빈도를 포함하는, 방법.

청구항 27

제25항에 있어서, 상기 시험 물질의 돌연변이 서명은 알려진 암-연관된 돌연변이 패턴과 충분히 유사한 돌연변이 패턴을 포함하는, 방법.

청구항 28

제25항에 있어서, 상기 생물학적 소스는 배양물에서 성장한 세포, 동물, 인간, 인간 세포주, 형질전환 동물, 비형질전환 동물, 인간 조직 샘플 또는 인간 혈액 샘플 중 적어도 하나인, 방법.

청구항 29

제25항에 있어서, 상기 생물학적 소스는 복수의 이중-가닥 DNA 단편을 포함하는 샘플을 추출하기 30일 이하 전에 시험 물질에 노출되는, 방법.

청구항 30

제25항에 있어서, 상기 돌연변이 서명은 삼중항 돌연변이 서명인, 방법.

청구항 31

제25항에 있어서, 상기 제1 가닥 서열 리드와 제2 가닥 서열 리드를 비교하기 전에, 상기 방법은 어댑터 서열, 서열 리드 길이 및 원래의 가닥 정보 중 하나 이상을 사용하여 제1 가닥 서열 리드를 제2 가닥 서열 리드와 연관시키는 단계를 포함하는, 방법.

청구항 32

제25항에 있어서, 상기 시퀀싱 라이브러리를 제조하기 전에, 상기 방법은 생물학적 소스를 시험 물질에 노출시키는 단계를 추가로 포함하는, 방법.

청구항 33

제32항에 있어서, 상기 생물학적 소스를 시험 물질에 노출시키기 전에, 상기 생물학적 소스는 암 조직이거나 이를 포함하는, 방법.

청구항 34

제32항에 있어서, 상기 생물학적 소스를 시험 물질에 노출시키기 전에, 생물학적 소스는 건강한 조직이거나 이를 포함하는, 방법.

청구항 35

제25항에 있어서, 상기 샘플은 혈액 샘플이거나 이를 포함하는, 방법.

청구항 36

제25항에 있어서, 상기 샘플은 암 세포주이거나 이를 포함하는, 방법.

청구항 37

제25항에 있어서, 상기 생물학적 소스는 암성 세포를 포함하고, 상기 물질은 암성 세포의 적어도 일부에 대해 선택적 유전독성에 대해 시험되는, 방법.

청구항 38

제37항에 있어서, 상기 물질은 치료 화합물인, 방법.

청구항 39

제38항에 있어서, 상기 치료 화합물의 선택적 유전독성에 민감한 것으로 나타난 암성 세포의 일부에 대해, 상기 방법은 치료 화합물에 대한 노출 전에 암성 세포의 일부에 대해 돌연변이체 빈도 및 돌연변이 스펙트럼 중 하나 이상을 결정하는 단계를 추가로 포함하는, 방법.

청구항 40

제25항에 있어서, 상기 시험 물질은 식품, 약물, 백신, 화장용 물질, 산업용 첨가제, 산업 부산물, 석유 증류물, 중금속, 가정용 세척제, 공기 매개 미립자, 제조 부산물, 오염물질, 가소제, 세제, 방사선-방출 생성물, 담배 제품, 화학 물질 또는 생물학적 물질을 포함하는, 방법.

청구항 41

유전독성 물질에 대한 대상체의 노출을 결정하는 방법으로서,

대상체의 DNA 돌연변이 스펙트럼을 알려진 돌연변이성 화합물의 돌연변이 스펙트럼과 비교하는 단계; 및

대상체의 DNA 돌연변이 스펙트럼과 가장 유사한 알려진 돌연변이성 화합물의 돌연변이 스펙트럼을 확인하는 단

계를 포함하는, 방법.

청구항 42

제41항에 있어서, 상기 대상체의 DNA 돌연변이 스펙트럼은 듀플렉스 시퀀싱에 의해 평가되는, 방법.

청구항 43

제41항에 있어서, 상기 대상체의 DNA 돌연변이 스펙트럼은 환자의 혈액으로부터 추출된 DNA로부터 생성되는, 방법.

청구항 44

제41항에 있어서, 상기 대상체의 DNA 돌연변이 스펙트럼은 삼중항 돌연변이 스펙트럼인, 방법.

청구항 45

제41항에 있어서, 상기 대상체의 DNA를 시퀀싱하여 대상체의 DNA 돌연변이 스펙트럼을 생성하는 단계를 추가로 포함하는, 방법.

청구항 46

제45항에 있어서, 상기 대상체의 DNA의 시퀀싱은 하나 이상의 알려진 암 유발자 유전자의 시퀀싱을 포함하는, 방법.

청구항 47

유전독소를 확인하기 위해 이중 가닥 폴리뉴클레오타이드의 오류 보정된 듀플렉스 시퀀싱에 사용될 수 있는 키트로서,

중합효소 연쇄 반응(PCR: polymerase chain reaction) 프라이머의 적어도 하나의 세트 및 어댑터 분자의 적어도 하나의 세트이되, 상기 프라이머 및 어댑터 분자는 오류 보정된 듀플렉스 시퀀싱 실험에 사용될 수 있는, 상기 프라이머 및 어댑터 분자; 및

대상체가 적어도 하나의 유전독소에 노출되었는지를 확인하기 위해 대상체의 샘플로부터 추출된 DNA의 오류 보정된 듀플렉스 시퀀싱을 수행하는 데 있어서 키트를 사용하는 방법에 대한 명령을 포함하는, 키트.

청구항 48

제47항에 있어서, 상기 시약은 DNA 복구 효소를 포함하는, 키트.

청구항 49

제47항에 있어서, 상기 어댑터 분자의 세트에서의 각각의 어댑터 분자는 적어도 하나의 단일 분자 식별자(SMI: single molecule identifier) 서열 및 적어도 하나의 가닥 한정 요소를 포함하는, 키트.

청구항 50

제47항에 있어서, 컴퓨터에서 실행될 때, 샘플에서 하나 이상의 이중-가닥 DNA 분자에 대한 오류-보정된 듀플렉스 시퀀싱 리드를 결정하는 단계 및 오류-보정된 듀플렉스 시퀀싱 리드를 사용하여 적어도 하나의 유전독소의 돌연변이체 빈도, 돌연변이 스펙트럼, 및/또는 삼중항 스펙트럼을 결정하는 단계를 수행하는 비일시적 컴퓨터 판독 가능한 매체에 구현된 컴퓨터 프로그램 제품을 추가로 포함하는, 키트.

청구항 51

제50항에 있어서, 상기 컴퓨터 프로그램 제품은 대상체의 DNA를 돌연변이시키는 데 있어서의 유전독소의 작용 기전; 및 유전독소 작용 기전에 기초하여 대상체에게 투여하기에 적합한 치료학적 치료 또는 예방학적 치료를 추가로 결정하는, 키트.

청구항 52

유전독소에 노출된 대상체를 진단하고 치료하는 방법으로서,

a) 대상체가

ⅰ) 대상체로부터 생물학적 샘플을 수득하는 것;

ii) 샘플로부터 추출된 복수의 이중 가닥 DNA 서열에 대한 듀플렉스 오류 보정된 시퀀싱 리드를 제공하는 것;

iii) DNA 서열의 돌연변이체 빈도, 돌연변이 스펙트럼, 및/또는 삼중항 돌연변이 스펙트럼을 결정하는 것;

iv) 돌연변이체 빈도, 돌연변이 스펙트럼 및/또는 삼중항 돌연변이 스펙트럼이 유전독소에 노출되었던 대상체를 나타내는지를 결정하는 것

에 의해 유전독소에 노출되었는지를 결정하는 단계;

b) 상기 대상체가 유전독소에 노출되었던 경우, 상기 유전독소와 연관된 질병 또는 장애의 발생을 예방하거나 억제하기 위한 예방학적 치료 및/또는 치료학적 치료를 제공하는 단계를 포함하는, 방법.

청구항 53

유전독소에 대한 안전한 노출의 역치 수준을 확인하고 치료를 제공하는 방법으로서,

a) 안전한 노출의 유전독소의 역치 수준을 결정하는 단계;

b) 대상체가

ⅰ) 대상체로부터 생물학적 샘플을 수득하는 것;

ii) 생물학적 샘플로부터 추출된 복수의 이중 가닥 DNA 서열에 대한 듀플렉스 오류 보정된 시퀀싱 리드를 제공하는 것;

iii) DNA 서열의 돌연변이체 빈도, 돌연변이 스펙트럼, 및/또는 삼중항 돌연변이 스펙트럼을 결정하는 것;

iv) 돌연변이체 빈도, 돌연변이 스펙트럼 및/또는 삼중항 돌연변이 스펙트럼이 특정 유전독소에 노출되었던 대상체를 나타내는지를 결정하는 것;

v) 돌연변이체 빈도, 돌연변이 스펙트럼 및/또는 삼중항 돌연변이 스펙트럼에 기초하여 유전독소에 대한 대상체의 노출의 수준을 컴퓨팅하는 것

에 의해 안전한 노출의 역치 수준보다 높은 수준에서 유전독소에 노출되었는지를 결정하는 단계; 및

c) 대상체가 안전한 노출의 유전독소의 역치 수준 초과에 노출되었던 경우, 유전독소와 연관된 질병 또는 장애의 발생을 예방하거나 억제하기 위한 예방학적 치료 및/또는 치료학적 치료를 제공하는 단계를 포함하는, 방법.

청구항 54

샘플의 유전독성 노출로부터 생긴 돌연변이성 사건 및/또는 핵산 손상 사건을 검출하고 확인하기 위한 시스템으로서,

시퀀싱 데이터 및 유전독성 데이터와 관련된 정보를 전송하기 위한 컴퓨터 네트워크이되, 상기 정보는 원시 시퀀싱 데이터, 듀플렉스 시퀀싱 데이터, 샘플 정보 및 유전독소 정보 중 하나 이상을 포함하는 컴퓨터 네트워크;

하나 이상의 사용자 컴퓨팅 장치와 연관되고 컴퓨터 네트워크와 통신하는 클라이언트 컴퓨터;

복수의 유전독소 프로파일 및 사용자 결과 기록을 저장하기 위한 컴퓨터 네트워크에 연결된 데이터베이스;

상기 컴퓨터 네트워크와 통신하고, 원시 시퀀싱 데이터 및 듀플렉스 시퀀싱 데이터를 생성하기 위한 클라이언트 컴퓨터로부터의 요청, 원래의 이중-가닥 핵산 분자를 나타내는 패밀리로부터의 그룹 서열 리드를 수신하고, 듀플렉스 시퀀싱 데이터를 생성하기 위해 개별 가닥으로부터의 대표적인 서열을 서로 비교하도록 구성된 듀플렉스 시퀀싱 모듈; 및

상기 컴퓨터 네트워크와 통신하고, 듀플렉스 시퀀싱 데이터를 기준 서열 정보와 비교하여 돌연변이를 확인하고, 돌연변이체 빈도, 돌연변이 스펙트럼 및 삼중항 돌연변이 스펙트럼 중 적어도 하나를 포함하는 유전독소 데이터를 생성하도록 구성된 유전독소 모듈을 포함하는, 시스템.

청구항 55

제54항에 있어서, 상기 유전독소 프로파일은 복수의 알려진 유전독소로부터의 유전독소 돌연변이 스펙트럼을 포함하는, 시스템.

청구항 56

비일시적 컴퓨터 판독 가능한 저장 매체로서,

하나 이상의 프로세서에 의해 실행될 때, 대상체가 적어도 하나의 유전독소에 노출되는지를 결정하고/하거나, 적어도 하나의 유전독소의 정체를 결정하기 위한 제1항 내지 제53항 중 어느 한 항의 방법을 수행하는 명령을 포함하는, 비일시적 컴퓨터 판독 가능한 저장 매체.

청구항 57

제56항에 있어서, 적어도 하나의 유전독소의 정체가 결정되는 검출된 물질의 돌연변이 스펙트럼, 돌연변이체 빈도, 및/또는 삼중항 돌연변이 스펙트럼을 컴퓨팅하는 것을 추가로 포함하는, 비일시적 컴퓨터 판독 가능한 저장 매체.

청구항 58

컴퓨터 시스템으로서,

대상체가 적어도 하나의 유전독소에 노출되는지 및/또는 이의 정체를 결정하기 위한 제1항 내지 제53항 중 어느 한 항의 방법을 수행하기 위한 것이되, 상기 시스템은 프로세서, 메모리, 데이터베이스 및 프로세서(들)에 대한 명령을 포함하는 비일시적 컴퓨터 판독 가능한 저장 매체를 갖는 적어도 하나의 컴퓨터를 포함하고, 상기 프로세서(들)는 제1항 내지 제53항 중 어느 한 항의 방법을 포함하는 연산을 수행하기 위해 상기 명령을 실행하도록 구성된, 시스템.

청구항 59

제58항에 있어서,

a. 유선 네트워크 또는 무선 네트워크;

b. 대상체의 샘플의 폴리뉴클레오타이드 서열을 추출하고 증폭시키고 제조하기 위해, 그리고 폴리뉴클레오타이드 서열을 네트워크를 통해 원격 서버에 전송하기 위해 시약을 포함하는 키트의 사용으로부터 도출된 데이터를 수신할 수 있는 복수의 사용자 전자 컴퓨팅 장치; 및

c. 프로세서, 메모리, 데이터베이스 및 프로세서(들)에 대한 명령을 포함하는 비일시적 컴퓨터 판독 가능한 저장 매체를 포함하는 원격 서버이되, 상기 프로세서(들)는 제1항 내지 제53항 중 어느 한 항의 방법을 포함하는 연산을 수행하기 위해 상기 명령을 실행하도록 구성된 원격 서버

를 포함하는 네트워크 컴퓨터 시스템을 추가로 포함하고;

d. 상기 원격 서버는 샘플의 유전독성 노출로부터 생긴 돌연변이성 사건 및/또는 핵산 손상 사건을 검출하고 확인할 수 있는, 시스템.

청구항 60

제59항에 있어서, 상기 네트워크를 통해 접근 가능한 데이터베이스 및/또는 제3자 데이터베이스는 알려진 유전독소의 유전독소 프로파일, 적어도 하나의 대상체의 샘플의 유전독소 프로파일 중 하나 이상을 포함하는 복수의 기록을 추가로 포함하고, 상기 유전독소 프로파일은 돌연변이 또는 DNA 손상 부위를 포함하는, 시스템.

청구항 61

비일시적 컴퓨터 판독 가능한 매체로서,

이 매체의 컨텐츠는 적어도 하나의 컴퓨터가 유전독성 스크리닝 검정으로부터 샘플에서 이중-가닥 핵산 분자에 대한 듀플렉스 시퀀싱 데이터를 제공하는 방법을 수행하게 하고, 상기 방법은

사용자 컴퓨팅 장치로부터 원시 서열 데이터를 수신하는 단계; 및

샘플에서 복수의 핵산 분자로부터 도출된 복수의 원시 서열 리드를 포함하는 샘플-특정 데이터세트를 생성하는 단계;

원래의 이중-가닥 핵산 분자를 나타내는 패밀리로부터 서열 리드를 그룹화하는 단계이되, 그룹화는 공유된 단일 분자 식별자 서열에 기초하는 상기 그룹화하는 단계;

원래의 이중-가닥 핵산 분자로부터 제1 가닥 서열 리드와 제2 가닥 서열 리드를 비교하여 제1 가닥 서열 리드와 제2 가닥 서열 리드 사이의 하나 이상의 관련성을 확인하는 단계; 및

샘플에서 이중-가닥 핵산 분자에 대한 듀플렉스 시퀀싱 데이터를 제공하는 단계를 포함하는, 비일시적 컴퓨터 판독 가능한 매체.

청구항 62

제58항에 있어서, 비교된 제1 서열 리드와 제2 서열 리드 사이에 비상보성의 뉴클레오타이드 위치를 확인하는 것을 추가로 포함하고, 상기 방법은

비상보성의 위치에서, 공정 오류를 확인하고 제거하거나 무시하는 단계; 및

공정 오류로서 확인되지 않은 비상보성의 위치에서, 유전독소에 대한 노출로부터 생긴 가능한 생체내 DNA 손상 부위로서 비상보성의 남은 위치를 확인하는 단계를 추가로 포함하는, 컴퓨터 판독 가능한 매체.

청구항 63

비일시적 컴퓨터 판독 가능한 매체로서,

이 매체의 컨텐츠는 적어도 하나의 컴퓨터가 샘플의 유전독성 노출로부터 생긴 돌연변이성 사건을 검출하가 확인하는 방법을 수행하게 하고, 상기 방법은

듀플렉스 서열 데이터를 기준 서열 정보와 비교하는 단계;

듀플렉스 서열 데이터에서 돌연변이를 확인하는 단계이되, 돌연변이는 기준 정보와 비동의의 영역으로 확인되는 상기 확인하는 단계;

듀플렉스 서열 데이터에서 돌연변이체 빈도를 결정하는 단계;

듀플렉스 서열 데이터로부터 돌연변이 스펙트럼을 생성하는 단계;

듀플렉스 서열 데이터로부터 삼중항 돌연변이 스펙트럼을 생성하는 단계; 및

돌연변이 스펙트럼 및/또는 삼중항 돌연변이 스펙트럼을 복수의 알려진 유전독소 데이터세트와 비교하는 단계를 포함하는, 비일시적 컴퓨터 판독 가능한 매체.

청구항 64

비일시적 컴퓨터 판독 가능한 매체로서,

이 매체의 컨텐츠는 적어도 하나의 컴퓨터가 대상체에서 발암물질 또는 발암물질 노출을 검출하가 확인하는 방법을 수행하게 하고, 상기 방법은

대상체로부터 샘플로부터 생성된 듀플렉스 시퀀싱 데이터를 사용하여 표적 게놈 영역에서 서열 변이체를 확인하는 단계;

시험 샘플 및 대조군 샘플의 변이체 대립유전자 빈도(VAF: variant allele frequency)를 계산하는 단계;

VAF가 대조군에서보다 시험 그룹에서 더 높은지를 결정하는 단계;

VAF가 더 높은 샘플에서, 서열 변이체가 비일중항인지를 결정하는 단계;

VAF가 더 높은 샘플에서, 서열 변이체가 유발자 돌연변이인지를 결정하는 단계; 및

비일중항 및/또는 유발자 돌연변이를 갖는 샘플을 발암물질임에 의심되는 것으로 규명하는 단계를 포함하는, 비일시적 컴퓨터 판독 가능한 매체.

청구항 65

제68항에 있어서, 상기 발암물질에 대한 안전성 역치를 평가하고/하거나, 대상체에서 노출 후에 유전독소 연관된 질병 또는 장애를 발생시키는 것과 연관된 위험을 결정하는 것을 추가로 포함하는, 비일시적 컴퓨터 판독 가능한 매체.

발명의 설명

기 술 분 야

[0001] **관련 출원의 상호 참조**

[0002] 본원은 2018년 2월 13일자에 출원된 미국 가특허 출원 제62/630,228호 및 2018년 9월 26일자에 출원된 미국 가특허 출원 제62/737,097호의 우선권 및 이익을 주장하고, 이의 개시내용은 그 전문이 본원에 참조로 포함된다.

배 경 기 술

[0003] 유전독성은 유전 물질(예를 들어, DNA, RNA)에 손상을 야기하는 물질 또는 과정(즉, 유전독소)의 파괴적 특성을 지칭한다. 생식 세포주에서, 핵산 물질의 손상은 유전 가능한 생식선 돌연변이를 발생시킬 가능성을 갖지만, 체세포에서의 핵산 물질의 손상은 체성 돌연변이를 발생시킬 수 있다. 일부 경우에, 이러한 체성 돌연변이는 악성 상태 또는 다른 질병으로 이어질 수 있다. 유전독소 노출이 이러한 핵산 손상을 직접적으로 또는 간접적으로 야기할 수 있거나, 일부 경우에 핵산 손상을 직접적으로 및 간접적으로 둘 다로 촉발함에 책임이 있을 수 있다고 규명되었다. 예를 들어, 유전독성 물질은 뉴클레오타이드 서열 자체 또는 이의 구조의 변화를 야기하거나, 카피되거나 복구되거나 그렇지 않으면 세포 기계에 의해 처리되도록 시도될 때 뉴클레오타이드 서열의 변경을 유도하는(또는 유도 확률을 증가시키는) 화학 변형(예를 들어, 부가물 또는 파괴)을 생성시키기 위해 유전 물질과 직접적으로 상호작용할 수 있다. 유전독소는 자연 발생 화학물질 또는 과정(예를 들어, 석탄, 라듐 또는 UV 광) 또는 인공적으로 생성된 화학물질 또는 과정 또는 처리(예를 들어, 산업용 메탄, X선 기계, 많은 화학요법 약물 및 일부 형태의 유전자 치료의 치료)일 수 있다.

[0004] 다른 유전독소는 DNA 복제의 충실도를 감소시키는 세포 경로를 활성화함으로써 핵산 손상을 간접적으로 촉발할 수 있다. 예를 들어, 이는 정상 관문을 우회하거나 핵산의 정상 복구를 감소시키는 세포-주기 기계의 직접적인 활성화 또는 간접적인 활성화(예컨대, 다른 것들 중에서 미스매치 복구(MMR: mismatch repair), 뉴클레오타이드 절제 복구(NER: nucleotide excision repair), 염기 절제 복구(BER: base excision repair), 이중-가닥 파괴 복구(DSBR: double-strand break repair), 전사-커플링 복구(TCR: transcription-coupled repair), 비상동성 말단 봉합(NHEJ: non-homologous end-joining)을 포함하는 많은 핵산 복구 경로 중 어느 하나의 직접적인 기능이상 또는 간접적인 기능이상)일 수 있다. 다른 유전독소는 자체가 유전독성인 세포 환경을 촉진함으로써 간접적으로 작용할 수 있다. 이러한 환경의 하나의 예는 서열 화학 조성 자체를 변형시키거나 핵산 가닥을 구조적으로 변경함으로써 유전 물질에 손상을 야기할 수 있는 (예를 들어, 면역 매개된 염증의 자극을 통해) 유기체 또는 세포에서의 반응성 산소 종 생산을 증가시킴으로써 생성될 수 있는 "산화 스트레스"이다. 또 다른 간접적인 형태의 유전독소는 유기체의 면역계의 소정의 양태를 억제하는 물질 또는 과정이다. 이러한 면역 감시의 감소는 (예를 들어, 소정의 조직에서의 염증의 야기 또는 세포-주기 진행의 촉진에 의해) 몇몇 기전 중 어느 하나를 통해 유전독성일 수 있는 미생물의 증식을 허용함으로써 유기체에서의 유전독성을 야기할 수 잇다. 더욱이, 이러한 물질 또는 과정은 유전자 비정상을 보유하는 세포를 없애는 정상 능력의 감소를 통해 유기체의 유전독성 로드에 기여할 수 있고, 이 기전을 통해 발암성일 수 있는데, 이 세포는 그렇지 않으면 제거될 것이다. 많은 유전독소의 기전이 발견되어 있다.

[0005] 유전독소는 다양한 외부 소스 및 내부 소스로부터 기원할 수 있다. 예를 들어, 외부(즉, 외인성) 소스는 화학물질 또는 화학물질의 혼합물(예를 들어, 의약품, 산업용/제조 부산물, 화학 폐기물, 화장품, 가정용 세척제, 가소제, 흡연, 용매 등); 자연 환경 또는 장치로부터의 중금속, 공기 매개 입자, 오염물질, 식품 제품, 방사선(예를 들어, 광자, 예컨대 감마 방사선, X-방사선, 입자 방사선 또는 이의 혼합), 물리적 힘(예를 들어, 자기장, 중력장, 가속력 등); 다른 유기체(예를 들어, 바이러스, 기생충, 박테리아, 원생동물, 진균)를 포함하거나, 다른 자연-발생 유기체(예를 들어, 진균, 식물, 동물, 박테리아, 박테리아, 원생동물 등)에 의해 제조될 수 있다. 소정의 작물 자체(예를 들어, 담배)는 이의 자연 형태에서 알려진 유전독소를 함유한다. 주된 식품 작물은 성장 동안(예를 들어, 산업용 폐기물에 의한 관개 용수의 오염), 수확 동안(예를 들어, 돌연변이원 아리스토로크산을 제조하는 아리스토콜리아와 작물의 우연의 동시수확), 저장 동안(예를 들어, 돌연변이원 아플라톡신을 생성하는

아스페르길루스 종을 성장시키는 축축한 콩과식물 및 곡물 사일로), 또는 제조 동안(예를 들어, 많은 형태의 유전독소를 생성하는 흡연 및 일부 다른 육류 보존 방법 또는 돌연변이원 아크릴아미드를 생성할 수 있는 전분의 고온 조리) 유전독소로 오염될 수 있다. 내부(즉, 내인성) 소스의 일부 예는 생화학적 과정 또는 생화학적 과정의 결과를 포함할 수 있다. 예를 들어, 화학 물질은 대사 활성화로부터 생긴 돌연변이원의 전구체이면 유전독소인 것으로 결정될 수 있다. 다른 예는 염증성 경로(예를 들어, 스트레스, 자가면역 질병)의 자극제, 또는 아폽토시스 또는 면역 감시의 저해제를 포함한다. 소스와 무관하게, 다수의 인자는 물질 또는 과정이 잠재적으로 유전독성, 돌연변이성 또는 발암성(즉, 암을 야기)인지를 결정하는 데 역할을 한다.

[0006] 소정의 분야에서, 돌연변이 과정을 검출하고 정량화하는 능력은 인간에서 암 위험을 평가하고 발암성 노출의 영향을 예측하는 데 중요하다. 마찬가지로, 화학적 화합물 또는 다른 물질이 핵산 돌연변이를 야기할 가능성의 평가는 (예를 들어, 의약품, 화장품, 식품 제품, 제조 부산물 등) 판매 전 제품 안전성 시험의 필수 요소이다. 유전독소를 확인하기 위한 현재의 방법은 힘들고, 비싸고, 시간이 지연되고(예를 들어, 노출과 증상 사이의 몇 년), (오직 소정의 모델 유기체에 대해) 진정한 인간내 효과를 대표하지 않을 수 있고, 일부 경우에 정확한 원인 물질을 집어내는 데 어려움을 제시한다. 예를 들어, 가끔 아픈 대상체의 집단(예를 들어, 암 클러스터)의 발생률 증가의 검출은 유전독소의 조사(예를 들어, 의약품 및 식품 안전성 분석, 환경 오염물질 또는 환경 투기의 조사 등)가 개시되기 전에 필요하다.

[0007] 생체내 체성 돌연변이의 종래의 수단은 박테리아, 세포 배양물 또는 형질전환 동물에서 선택-기반 검정으로부터 간접적으로 추론되는데, 여기서 전장-게놈 효과는 작은 인공 리포터로부터 추정된다. 따라서, 현재 사용되는 검정은 생체내 화합물의 진정한 유전독성 가능성에 대한 불완전한 대리물이고, 이것은 화합물의 돌연변이 가능성에 대한 정보의 제한된 하위집단을 오직 제공하면서 노동 집약적이다. 인공 박테리아 시스템에서 돌연변이 가능성을 나타내는 많은 화합물(즉, Ames 검정)은 인간에서 진짜의 위험을 정확하게 반영하지 않고, 그렇지 않으면 치료학적으로 유망한 화합물이 개발 또는 상업적 사용으로부터 불필요하게 취소되게 할 가능성이 크다. 유사하게, 발암 가능성을 갖는 일부 화합물은 박테리아에서 검출 불가능한 비직접적인 돌연변이 기전을 통해 그렇게 한다. 위험이 조기에 적절히 인식될 수 없으면서, 이러한 화합물은 대상체에 해를 야기할 수 있다.

[0008] 생체내 포유류 리포터 시스템, 예컨대 형질전환 설치류 검정(예를 들어, BigBlue® 마우스 및 래트 및 Muta™ Mouse)은 박테리아보다 인간 약물의 더 양호한 근사치를 제공한다. 동물이 인간을 완벽히 표시하지 않는 한 이 시스템이 제한되지만, 포유류 형질전환 검정은 초기 전임상 안전성 시험에 귀중하지만, 이 검정은 복합하고 여전히 다소 인공적이다. BigBlue® 검정은 예를 들어 리포터-기반 시스템에 의존하고, 이로써 다중카피 람다-파지 전이유전자에서 생기는 돌연변이의 하위집단은 셔틀 벡터에 의한 리포터의 회수 후 표현형적으로 확인될 수 있고, 이 벡터는 이후 박테리아로 형질주입된다. 294 BP 리포터 유전자에서 생기는 모든 돌연변이가 검출될 수 있는 것은 아닌데, 많은 것이 표현형을 부여하지 않기 때문이다. 전이유전자 자체는 고도로 압축되고 메틸화되고 더 넓은 게놈의 고도로 가변적인 전사 및 압축 상태를 나타내지 않는다. 바이러스 기계 및 박테리아 기계를 통한 돌연변이체 분자의 통과는 인공 돌연변이를 도입할 가능성을 갖고, 각각의 단계에서 발생하는 고유한 병목현상은 돌연변이의 대립유전자 분획이 비정량적임을 의미한다. 더욱이, 시험은 제한된 종 하위집단의 특정 균주의 사용을 필요로 한다. 그리고 설치류 자체는 인간을 완벽히 대표하지 않는다. 예를 들어, 아플라톡신은 인간에서 고도로 돌연변이원성이지만, 소정의 대사 효소가 발현될 때 이의 해독작용을 촉진하는 성 성숙 후에 마우스에서 의미있게 발암성이 아니다. 형질전환 설치류가 일부 시험 상황에서 발암성 대리물로서 사용될 수 있는 유효한 유전독성 메트릭으로서 미국 식품의약청(FDA: Food and Drug Administration) 및 다른 관리 기관에 의해 인정된 현재의 황금 표준이지만, 그것은 화합물이 인간에서 암을 야기할 가능성을 평가하기 위한 대략적으로 유용한 도구로서 전혀 최적이 아니다.

[0009] 소정의 건강 위험(즉, 암/악성상태/신생물, 신경독성, 신경퇴행, 불임, 선천적 결함 등)에 기여하는 핵산 돌연변이 및 손상을 야기하는, 대상체가 노출될 수 있는, 인자/물질/환경의 유전독성 가능성의 직접적인 측정을 허용하는 빠르고 유연하고 신뢰성 있는 방법이 필요하다. 이 방법은 (종래 기술의 황금 표준 시험에 필요한 바대로) 어떤 클론성 선택의 필요 없이 어떻게 발암 인자가 생체내 돌연변이 또는 다른 유전독성 손상을 야기하여 대상체/유기체, 또는 대상체/유기체에 의해 모델링된 다른 유기체에서 암 발생 또는 다른 질병 또는 장애로 이어지는지의 작용 기전에 대한 (추론적으로 또는 직접적으로) 정보를 제공하면서 임의의 유형의 유기체에서의 임의의 조직 유형 및/또는 세포 유형의 임의의 게놈 유전좌위에서 이용 가능해야 한다.

[0010] 이들 특징을 갖는 충분히 정확하고 편리한 도구가 이용 가능한 경우, 이것은 예를 들어 전임상 및 임상 약물 안전성 시험 둘 다에서의; 유전독소 연관된 질병 및 장애의 예방, 진단 및 치료에서의; 돌연변이 원인 인자/물질

및 이의 작용 기전의 검출 및 확인에서의 많은 적용; 및 다른 산업-전반 영향(예를 들어, 환경 오염 시험 및 독성 발생의 역치 수준의 결정, 고속 소비재 안전성 시험, 독성 노출이 의심되는 경우 환자 진단 및 치료, 유전독소의 의도적인 방출 또는 의도치 않은 방출의 국가 보안 위험 평가 등)을 가질 것이다.

발명의 내용

[0011] 본 기술내용은 유전독성을 평가하기 위한 방법, 시스템 및 시약의 키트에 관한 것이다. 특히, 본 기술내용의 일부 실시형태는 노출된 대상체에서 화합물(예를 들어, 화학적 화합물) 및/또는 환경 물질(예를 들어, 방사선)의 유전독성 가능성을 평가하기 위한 듀플렉스 시퀀싱(Duplex Sequencing)의 이용에 관한 것이다. 예를 들어, 본 기술내용의 다양한 실시형태는 어떤 클론성 선택의 필요 없이 임의의 유기체의 임의의 게놈 상황에서 화합물-유발된 돌연변이의 직접적인 측정을 허용하는 듀플렉스 시퀀싱 방법을 수행하는 것을 포함한다. 본 기술내용의 추가의 예는 듀플렉스 시퀀싱 및 연관된 시약을 사용하여 게놈 생체내 돌연변이유발을 검출하고 평가하는 방법에 관한 것이다. 본 기술내용의 다양한 양태는 전임상 및 임상 약물 안전성 시험 둘 다에서 많은 적용, 및 다른 산업-전반 영향을 갖는다.

[0012] 일 실시형태에서, 본 기술내용은 (1) 돌연변이원에 노출된 대상체로부터 추출된 하나 이상의 표적 이중-가닥 DNA 분자를 듀플렉스 시퀀싱하는 단계; (2) 표적화된 이중-가닥 DNA 분자에 대한 오류-보정된 공통 서열을 생성하는 단계; 및 (3) 표적화된 이중-가닥 DNA 분자에 대한 돌연변이 스펙트럼을 확인하는 단계; (4) 시퀀싱된 하나 이상의 유형의 듀플렉스 염기-쌍마다 고유한 돌연변이의 수를 계산함으로써 표적 이중-가닥 DNA 분자에 대한 돌연변이체 빈도를 계산하는 단계를 포함하는, 돌연변이원에 대한 대상체의 노출 후에 대상체에서 생체내 발생한 게놈 돌연변이를 검출하고 정량화하는 방법을 포함한다.

[0013] 다른 실시형태에서, 본 기술내용은 (1) 시험 화합물에 노출된 살아 있는 유기체, 예를 들어 시험 동물로부터 추출된 DNA 단편을 듀플렉스 시퀀싱하는 단계; 및 (2) 시험 화합물의 돌연변이성 서명을 생성하는 단계를 포함하는, 시험 화합물의 돌연변이성 서명을 생성하는 방법을 포함한다. 그리고 상기 방법은 시퀀싱된 듀플렉스 염기-쌍마다 고유한 돌연변이의 수를 계산함으로써 복수의 DNA 단편에 대한 돌연변이체 빈도를 계산하는 단계를 추가로 포함할 수 있다.

[0014] 다른 실시형태에서, 본 기술내용은 (1) 화합물에 노출된 시험 동물로부터 추출된 표적화된 DNA 단편을 듀플렉스 시퀀싱하여 표적화된 DNA 단편의 오류-보정된 공통 서열을 생성하는 단계; (2) 오류-보정된 공통 서열로부터 화합물의 돌연변이성 서명을 생성하는 단계; 및 (3) 화합물에 대한 노출이 충분히 유전독성 화합물을 대표하는 돌연변이성 서명을 생성시키는지를 결정하는 단계를 포함하는, 화합물의 유전독성 가능성을 평가하는 방법을 포함한다.

[0015] 다른 실시형태에서, 본 기술내용은 유전독소를 검출하고 정량화하기 위한 본원에 개시된 방법을 수행하기 위한 설명서를 갖는 시약을 포함하는 키트를 포함한다. 이 키트는 전자 컴퓨팅 장치(예를 들어, 랩탑/데스크탑 컴퓨터, 태블릿 등)에 설치되거나 네트워크(예를 들어, 대상체 기록 및 검출된 유전독소의 데이터베이스를 갖는 원격 서버)를 통해 접근 가능한 컴퓨터 프로그램 제품을 추가로 포함할 수 있다. 컴퓨터 프로그램 제품은, 컴퓨터에서 실행될 때, 유전독소를 검출하고 확인하기 위한 본원에 개시된 키트를 사용하는 방법의 단계를 수행하는 비일시적 컴퓨터 판독 가능한 매체에서 구현된다.

[0016] 다른 실시형태에서, 본 기술내용은 (1) 원격 서버; (2) 대상체의 샘플을 추출하고 증폭시키고 시퀀싱하기 위해 본원에 개시된 키트를 사용할 수 있는 복수의 사용자 전자 컴퓨팅 장치; (3) 알려진 유전독소 프로파일을 갖는 제3자 데이터베이스(선택적); 및 (4) 전자 컴퓨팅 장치, 데이터베이스와 원격 서버 사이에 전자 통신을 전송하기 위한 유선 네트워크 또는 무선 네트워크를 포함하는, 적어도 하나의 유전독소에 대한 대상체의 노출을 확인하거나 확정하기 위한 네트워크 컴퓨터 시스템을 포함한다. 원격 서버는 (a) 사용자 유전독소 기록 결과, 및 유전독소 프로파일(예를 들어, 스펙트럼, 빈도, 작용 기전 등)의 기록을 저장하는 데이터베이스; (b) 메모리에 통신 연결된 하나 이상의 프로세서; 및 프로세서(들)에 대한 명령을 포함하는 하나 이상의 비일시적 컴퓨터 판독 가능한 저장 장치 또는 매체를 추가로 포함하고, 여기서 상기 프로세서는 단편의 듀플렉스 시퀀싱에서 오류를 보정하는 단계; 및 적어도 하나의 유전독소의 정체가 결정될 수 있는 검출된 물질의 돌연변이 스펙트럼, 돌연변이체 빈도 및 삼중항 돌연변이 스펙트럼을 컴퓨팅하는 단계를 포함하는 연산을 수행하기 위해 상기 명령을 실행하도록 구성된다.

[0017] 본 기술내용은, 하나 이상의 프로세서에 의해 실행될 때, 대상체가 적어도 하나의 유전독소에 노출되는지 및/또는 이의 정체를 결정하기 위한 방법을 수행하는 명령을 포함하는 비일시적 컴퓨터 판독 가능한 저장 매체를 추

가로 포함하고, 상기 방법은 단편의 듀플렉스 시퀀싱에서 오류를 보정하는 단계; 및 적어도 하나의 유전독소의 정체가 결정되는 검출된 물질의 돌연변이 스펙트럼, 돌연변이체 빈도 및 삼중항 스펙트럼을 컴퓨팅하는 단계를 포함한다.

[0018] 본 기술내용은 대상체가 적어도 하나의 유전독소에 노출되는지 및/또는 이의 정체를 결정하기 위한 전산화 방법을 추가로 포함하고, 상기 방법은 단편의 듀플렉스 시퀀싱에서 오류를 보정하는 단계; 및 적어도 하나의 유전독소의 정체가 결정되는 검출된 물질의 돌연변이 스펙트럼, 돌연변이체 빈도 및 삼중항 스펙트럼을 컴퓨팅하는 단계를 포함한다.

[0019] 다른 실시형태에서, 본 기술내용은 유전독소에 노출된 대상체를 진단하고 치료하기 위한 방법, 시스템 및 키트를 포함한다. 진단은 대상체가 노출되고/되거나 소비한 적어도 하나의 유전독소를 검출하는 것을 포함하고; 치료는 미래의 유전독소(들)의 노출 및/또는 소비를 제거하는 것 및/또는 유전독소(들)의 생물학적 효과를 차단하고/하거나 그렇지 않으면 이에 대응하기 위해 치료 프로토콜(예를 들어, 의약품)을 투여하는 것을 포함한다.

[0020] 다른 실시형태에서, 본 기술내용은 전임상 및 임상 약물 안전성 시험 둘 다를 위한; 발암물질 및 이의 작용 기전을 검출하고 사용하기 위한; 그리고 다른 산업-전반 영향을 위한(예를 들어, 독성 환경 오염물질, 고속 소비재 및 약물 안전성 시험 등) 방법, 전산 시스템 및 키트를 제공한다.

[0021] 다른 실시형태에서, 본 기술내용은 오류 보정된 듀플렉스 시퀀싱을 사용하여 신규의 유전독소를 확인하고/하거나, 이후 대상체가 유전독소 연관된 질병 또는 장애를 발생시킬 위험에 있기 전에 (예를 들어, 미국 환경 보건국(Environmental Protection Agency) 기준을 설정하는 데 사용되는; 유전독소에 노출된 대상체를 진단하고 치료하는 데 사용되는 등의) 대상체가 노출될 수 있는 유전독소의 안전성 역치 양(중량, 부피, 농도 등) 및/또는 안전성 역치 돌연변이체 빈도를 결정하기 위한 방법, 시스템 및 키트를 포함한다.

[0022] 다른 실시형태에서, 본 기술내용은 대상체가 안전성 역치 수준(즉, 유전독소 양 및/또는 유전독소 돌연변이체 빈도 및 삼중항 서명) 초과로 유전독소에 노출되는지를 결정하고; 그렇다면, 이후 질병 발생을 예방하거나 억제하거나 방해하기 위한 예방학적 치료를 제공함으로써 대상체가 돌연변이 연관된 질병 또는 장애를 발생시키는 것을 예방하기 위한 방법, 시스템 및 키트를 포함한다.

[0023] 본 기술내용의 일 양태는 유전독소를 야기하는 돌연변이에 대한 노출 후 몇 일 또는 몇 주 또는 몇 달 또는 몇 년 내이지만 질병을 야기하는 돌연변이를 검출하는 능력을 포함한다. 보통, 완전한 질병 발생은 여러 해(예를 들어, 석면에 대한 노출 후 폐암 발생의 경우 10년 내지 20년) 동안 진단되지 않는다. 본원에 개시된 방법 및 키트는 증상이 나타나는 데 여러 해 대기하는 것에 비해 노출 직후 질병 발생을 야기하는 게놈 돌연변이의 검출이 가능하게 한다.

[0024] 본 기술내용의 다른 양태는 대상체가 유전독소에 대한 가능한 노출 후 최소 약 2일 내지 5일 내지는 몇 년 후 내에 유전독소 야기된 돌연변이로 인한 질병 또는 장애를 발생시킬 위험이 증가하는지를 예측하고; 그렇다면, 초기 단계에서 질병 발생을 검출하기 위해 예방학적 치료 및 주기적인 스크리닝을 제공하는 능력을 포함한다.

[0025] 다른 양태는 복수의 이중-가닥, 단리된 게놈 DNA 단편을 포함하는 DNA 라이브러리 및 이의 제조 방법을 포함하고, 여기서 각각의 단편은 하나 이상의 원하는 어댑터 분자에 결찰된다.

[0026] 다른 양태는 어떤 화합물이 유전독성인지를 확인하기 위해 복수의 화합물을 신속히 스크리닝하기 위한 고속 방법을 포함한다.

[0027] 다른 양태는 대상체가 어떤 유전독소에 노출되는지를 결정하기 위해 동일한 대상체의 복수의 상이한 조직/세포 유형을 신속히 스크리닝하기 위한 고속 방법을 포함한다.

[0028] 다른 양태는 임의의 유전독소에 노출된 집단의 백분율을 결정하기 위해 상이한 대상체로부터 유래된 복수의 조직 및 세포를 신속히 스크리닝하기 위한 고속 방법을 포함한다.

[0029] 다른 양태는 유전독소의 노출에 의해 특정 질병 또는 장애와 연관된 돌연변이를 생성시키는 유전독소의 "작용 기전"을 직접적으로 또는 추론적으로 결정하는 것을 포함한다.

[0030] 본 기술내용의 다른 실시형태, 양태 및 이점은 하기 상세한 설명에 추가로 기재된다.

도면의 간단한 설명

[0031] 본 개시내용의 많은 양태는 하기 도면을 참조하여 더 잘 이해될 수 있다. 도면의 구성성분은 비율조정될 필요는

없다. 대신에, 본 개시내용의 원칙을 명확히 예시하는 데 강조가 이루어진다.

도 1a는 본 기술내용의 일부 실시형태와 사용하기 위한 핵산 어댑터 분자 및 본 기술내용의 실시형태에 따른 이중-가닥 핵산 단편에 대한 어댑터 분자의 결찰로부터 생긴 이중-가닥 어댑터-핵산 복합체를 예시한다.

도 1b 및 도 1c는 본 기술내용의 실시형태에 따른 다양한 듀플렉스 시퀀싱 방법 단계의 개념적 예시이다.

도 2a는 종래의 장기간 설치류 발암성 연구(왼쪽 도식), 생체외 선택에 의한 종래의 형질전환 설치류 돌연변이원성 연구(중간 도식) 및 본 기술내용의 양태에 따른 직접적인 DNA 시퀀싱 계획을 통한 돌연변이유발 평가(오른쪽 도식)를 포함하는 시험 화합물의 인간 암 위험을 예측하기 위한 생체내 동물 연구를 사용하기 위한 다양한 방법 계획의 개념적 예시이다.

도 2b 및 도 2c는 본 기술내용의 양태에 따른 배양물에서 성장한 인간 세포에서 시험 화합물의 시험관내 돌연변이유발을 평가하기 위한(2b) 그리고 야생형 마우스에서 시험 화합물의 생체내 돌연변이유발을 평가하기 위한(2c) 듀플렉스 시퀀싱을 사용하기 위한 방법 계획의 개념적 예시이다.

도 3a 내지 도 3d는 본 기술내용의 실시형태에 따른 돌연변이원 치료 후에 간 및 골수에서 듀플렉스 시퀀싱(도 3a 및 도 3b) 및 BigBlue® *cII* 플라크 검정(도 3c 및 도 3d)에 대해 계산된 돌연변이체 빈도를 보여주는 상자 그림 그래프이다.

도 3e는 본 기술내용의 실시형태에 따른 BigBlue® *cII* 플라크 검정에서 도 3a 내지 도 3d의 듀플렉스 시퀀싱 검정에 비하여 상대 *cII* 돌연변이체 배수 증가를 예시하는 선도이다.

도 3f는 본 기술내용의 실시형태에 따른 BigBlue® 마우스 조직으로부터 생성된 개별적으로 선별된 돌연변이체 플라크 및 BigBlue® 마우스 조직으로부터의 *cII*의 gDNA의 듀플렉스 시퀀싱에 대한 *cII* 유전자 내의 단일 뉴클레오타이드 변이체(SNV)의 비율을 보여준다.

도 3g 및 도 3h는 본 기술내용의 실시형태에 따른 코돈 위치 및 기능적 결과에 의한 모든 BigBlue® 조직 유형 및 치료 그룹에 걸쳐 *cII*의 직접적인 듀플렉스 시퀀싱(도 3g)에 의해 그리고 개별적으로 수집된 돌연변이체 플라크(도 3h) 중에서 확인된 돌연변이의 분포를 보여준다.

도 4는 본 기술내용의 실시형태에 따른 각각의 치료 그룹의 다수의 샘플에서 듀플렉스 시퀀싱에 의해 측정된 돌연변이체 빈도를 보여주는 막대 그래프이다.

도 5a 및 도 5b는 본 기술내용의 실시형태에 따른 듀플렉스 시퀀싱에 의해 측정된 바와 같은 간(도 5a) 및 골수(도 5b)에서의 *cII* 전이유전자와 비교된 내인성 유전자의 돌연변이체 빈도를 보여주는 막대 그래프이다.

도 5c는 본 기술내용의 실시형태에 따른 표시된 치료 카테고리에 대해 간 및 골수에 대한 유전자 영역에 의한 듀플렉스 시퀀싱에 대해 계산된 SNV 돌연변이체 빈도(MF)를 보여주는 상자 그림 그래프이다.

도 5d는 본 기술내용의 실시형태에 따른 도 5c에 도시된 집합체 데이터의 개별 측정치를 보여주는 산점도이다.

도 6은 본 기술내용의 실시형태에 따른 듀플렉스 시퀀싱에 의해 측정된 바와 같은 돌연변이 스펙트럼을 보여주는 막대 그래프이다.

도 7a 내지 도 7c는 본 기술내용의 실시형태에 따른 비히클 대조군(7a), 벤조[a]피렌(7b) 및 N-에틸-N-니트로소우레아(7c)에 대한 트리뉴클레오타이드 돌연변이 스펙트럼을 보여주는 그래프이다.

도 8은 본 기술내용의 실시형태에 따른 대조군 및 우레탄에 처리된 실험 동물에 대한 폐, 비장 및 혈액 샘플의 돌연변이체 빈도를 보여주는 막대 그래프이다.

도 9는 본 기술내용의 실시형태에 따른 조직 샘플의 그룹에 걸친 평균 최소 점 돌연변이체 빈도를 보여주는 막대 그래프이다.

도 10a는 본 기술내용의 실시형태에 따른 표시된 치료 카테고리에 대해 폐, 비장 및 혈액에 대한 유전자 영역에 의한 듀플렉스 시퀀싱에 대해 계산된 SNV MF를 보여주는 상자 그림 그래프이다.

도 10b는 본 기술내용의 실시형태에 따른 도 10a에 도시된 집합체 데이터의 개별 측정치를 보여주는 산점도이다.

도 11은 본 기술내용의 실시형태에 따른 듀플렉스 시퀀싱에 의해 측정된 바와 같은 시험된 조직 내에 우레탄 및 비히클 대조군의 돌연변이 스펙트럼을 보여주는 막대 그래프이다.

도 12a 및 도 12b는 본 기술내용의 실시형태에 따른 비히클 대조군(12a) 및 우레탄(12b)에 대한 인접한 뉴클레오타이드의 상황에서의 돌연변이 스펙트럼(즉, 트리뉴클레오타이드 스펙트럼)을 보여주는 그래프이다.

도 13은 본 기술내용의 실시형태에 따른 우레탄 치료된 샘플에서의 단일 뉴클레오타이드 변이체(SNV) 스펙트럼 가닥 바이어스를 보여준다.

도 14는 본 기술내용의 실시형태에 따른 듀플렉스 시퀀싱에 의해 검출된 바와 같은 변이체 대립유전자 분획의 초기 단계 신생물성 클론성 선택을 예시하는 그래프이다.

도 15a는 본 기술내용의 실시형태에 따른 Tg-rasH2 마우스 모델에서 인간 형질전환 유전좌위를 포함하는 유전자의 *Ras* 패밀리로부터 포획된 엑손에 대한 게놈 간격에 걸쳐 작도된 SNV를 예시하는 그래프이다.

도 15b는 본 기술내용의 실시형태에 따른 인간 *HRAS* 전이유전자의 엑손 3에 정렬된 단일 뉴클레오타이드 변이체를 예시하는 그래프이다.

도 16a 내지 도 16b는 종래의 DNA 시퀀싱(도 16a) 및 본 기술내용의 실시형태에 따른 듀플렉스 시퀀싱(도 16b)을 사용한 우레탄 치료 후에 마우스 폐에서의 인간 *HRAS*의 대표적인 400개의 염기 쌍 절편으로부터의 시퀀싱 데이터의 그래프 표현을 보여준다.

도 17a 내지 도 17c는 COSMIC로부터의 서명 1(도 17a), 서명 4(도 17b) 및 서명 29(도 17c)에 대한 인접한 뉴클레오타이드의 상황에서의 돌연변이 스펙트럼(즉, 트리뉴클레오타이드 스펙트럼)을 보여주는 그래프이다.

도 18은 본 기술내용의 실시형태에 따른 실시예 1 및 실시예 2로부터의 모든 30개의 공개된 COSMIC 서명 및 4개의 코호트 스펙트럼의 비지도된 계층적 클러스터링을 보여준다.

도 19는 본 기술내용의 실시형태에 따른 유전독성 노출로부터 생긴 돌연변이성 사건 및/또는 핵산 손상 사건을 확인하기 위해 본원에 개시된 방법 및/또는 키트와 사용하기 위한 네트워크 컴퓨터 시스템의 도식적 다이어그램이다.

도 20은 본 기술내용의 실시형태에 따른 본 기술내용의 실시형태에 따른 듀플렉스 시퀀싱 공통 서열 데이터를 제공하기 위한 루틴을 예시하는 흐름 다이어그램이다.

도 21은 본 기술내용의 실시형태에 따른 샘플의 유전독성 노출로부터 생긴 돌연변이성 사건을 검출하고 확인하기 위한 루틴을 예시하는 흐름 다이어그램이다.

도 22는 본 기술내용의 실시형태에 따른 샘플의 유전독성 노출로부터 생긴 DNA 손상 사건을 검출하고 확인하기 위한 루틴을 예시하는 흐름 다이어그램이다.

도 23은 본 기술내용의 실시형태에 따른 대상체에서 발암물질 또는 발암물질 노출을 검출하고 확인하기 위한 루틴을 예시하는 흐름 다이어그램이다.

발명을 실시하기 위한 구체적인 내용

[0032] 본 기술내용의 몇몇 실시형태의 상세한 설명은 도 1a 내지 도 20과 관련하여 하기에 기재된다. 실시형태는 예를 들어 유전독성을 평가하기 위한 방법, 시스템, 키트를 포함할 수 있다. 본 기술내용의 일부 실시형태는 노출된 대상체, 모델 유기체 또는 모델 세포 배양 시스템에서 물질(예를 들어, 화학적 화합물) 또는 임의의 다른 유형의 노출(예를 들어, 방사선 소스)의 유전독성 가능성을 평가하기 위해 듀플렉스 시퀀싱을 사용하는 것에 관한 것이다. 본 기술내용의 다른 실시형태는 유전독성 물질과 연관된 돌연변이 서명을 결정하기 위해 듀플렉스 시퀀싱을 사용하는 것에 관한 것이다. 본 기술내용의 추가 실시형태는 대상체의 DNA 돌연변이 스펙트럼을 알려진 돌연변이성 화합물의 돌연변이 스펙트럼과 비교함으로써 대상체가 노출될 수 있는 하나 이상의 유전독성 물질을 확인하는 것에 관한 것이다. 본 기술내용의 추가 실시형태는 하나 이상의 조직에서의 하나 이상의 세포 유형으로부터의 대상체의 DNA 돌연변이 스펙트럼을 알려진 환경 또는 대상체가 노출될 수 있는 하나 이상의 위치 또는 환경에 존재하는 것으로 알려진 화합물의 돌연변이 스펙트럼과 비교함으로써 이러한 위치 또는 환경을 확인하는 것에 관한 것이다. 본 기술내용의 추가 실시형태는 하나 이상의 조직에서의 하나 이상의 세포 유형으로부터의 대상체의 DNA 돌연변이 스펙트럼을 알려진 개체 또는 이 개체가 노출되는 것으로 알려진 위치 또는 환경 또는 이러한 위치 또는 환경에 존재하는 것으로 알려진 화합물의 돌연변이 스펙트럼을 비교함으로써 대상체를 확인하

는 것에 관한 것이다. 소정의 실시형태에서, 유전독소는 발암 가능성에 평가될 수 있다. 추가 실시형태는 암 유발자 돌연변이로 생기는 돌연변이-보유 클론을 확인함으로써 돌연변이성 발암물질 또는 비돌연변이성 발암물질로부터 생긴 발암현상 위험을 확인하고 평가하는 것을 포함한다. 추가 실시형태는 돌연변이가 암 유발자(대개 "패신저" 돌연변이 또는 "히치하이커" 돌연변이로 알려짐)인 것으로 생각되지 않고, 실질적으로 클론을 고유하게 마킹하는 돌연변이-보유 클론의 비상상태를 확인함으로써 돌연변이성 또는 비돌연변이성 발암물질로부터 생긴 발암현상 위험을 확인하고 평가하는 것을 포함한다(Salk and Horwitz Sem Cancer Bio 2010 PMID: 20951806). 본 기술내용의 다른 실시형태는 유전독소 노출 또는 다른 내인성 유전독성 과정(예를 들어, 노화)으로부터 생긴 핵산 손상(특히 DNA 손상, 예컨대 부가물)을 검출하고 평가하기 위한 듀플렉스 시퀀싱에 관한 것이다.

[0033] 많은 실시형태가 듀플렉스 시퀀싱과 관련하여 본원에 기재되어 있지만, 본원에 기재된 것 이외에 오류-보정된 시퀀싱 리드를 생성할 수 있는 다른 시퀀싱 양상은 본 기술내용의 범위 내에 있다. 추가적으로, 본 기술내용의 다른 실시형태는 본원에 기재된 것과 상이한 구성, 구성성분 또는 절차를 가질 수 있다. 그러므로, 당업자는 따라서 본 기술내용이 추가 요소를 갖는 다른 실시형태를 가질 수 있고, 본 기술내용이 도 1a 내지 도 20과 관련하여 하기 도시되고 기재된 여러 특징이 없는 다른 실시형태를 가질 수 있다는 것을 이해할 것이다.

[0034] 정의

[0035] 본 개시내용이 보다 용이하게 이해되도록 하기 위해, 소정의 용어가 처음에 하기에 정의된다. 하기 용어 및 다른 용어에 대한 추가 정의가 본 명세서에 걸쳐 제시된다.

[0036] 본원에서, 문맥에서 달리 명확하지 않는 한, 용어 "하나"는 "적어도 하나"를 의미하는 것으로 이해될 수 있다. 본원에 사용된 바와 같이, 용어 "또는"은 "및/또는"을 의미하는 것으로 이해될 수 있다. 본원에서, 용어 "포함하는" 및 "함유하는"은 홀로 제시되든 하나 이상의 추가 성분 또는 단계와 함께 제시되든 항목화된 성분 또는 단계를 포괄하는 것으로 이해될 수 있다. 범위가 본원에 제공되는 경우, 종점이 포함된다. 본원에 사용된 바와 같이, 용어 "포함한다" 및 이 용어의 파생어, 예컨대 "포함하는" 및 "포함"은 다른 첨가제, 성분, 정수 또는 단계를 배제하는 것으로 의도되지 않는다.

[0037] ***약***: 용어 "약"은, 값과 관련하여 본원에 사용될 때, 언급된 값의 맥락에서 유사한 값을 지칭한다. 일반적으로, 그 맥락에 친숙한 당업자는 그 맥락에서 "약"이 포괄하는 변화량의 관련 정도를 이해할 것이다. 예를 들어, 일부 실시형태에서, 용어 "약"은 언급된 값의 25%, 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1% 이하 내의 값의 범위를 포괄할 수 있다. 양 또는 음의 방향의 단일 숫자 값 단계가 그 값의 25%를 초과하는 단일 디지트 정수 값의 변화량에 대해, "약"이 양 또는 음의 방향의 적어도 1, 2, 3, 4 또는 5 정수 값을 포함하는 것으로 당업자에 의해 일반적으로 인정되고, 이는 상황에 따라 0을 가로지르거나 가로지르지 않을 수 있다. 이것의 비제한적인 예는 일부 상황에서 3 센트가 약 5 센트로 생각될 수 있다는 추정인데, 이는 당업자에게는 명확할 것이다.

[0038] ***유사체***: 본원에 사용된 바와 같이, 용어 "유사체"는 기준 물질과 하나 이상의 특정 구조 특징, 요소, 성분 또는 모이어티를 공유하는 물질을 지칭한다. 통상적으로, "유사체"는 예를 들어 코어 또는 공통 구조를 공유하는 기준 물질과 상당한 구조 유사성을 보여주지만, 또한 소정의 별개의 방식에서 다르다 일부 실시형태에서, 유사체는 예를 들어 기준 물질의 화학 조작에 의해 기준 물질로부터 생성될 수 있는 물질이다. 일부 실시형태에서, 유사체는 기준 물질을 생성하는 합성 공정과 실질적으로 유사한(예를 들어, 이 합성 공정과 복수의 단계를 공유하는) 합성 공정의 수행을 통해 생성될 수 있는 물질이다. 일부 실시형태에서, 유사체는 기준 물질을 생성하기 위해 사용되는 합성 공정과 상이한 합성 공정의 수행을 통해 생성되거나 생성될 수 있다.

[0039] ***생물학적 샘플***: 본원에 사용된 바와 같이, 용어 "생물학적 샘플" 또는 "샘플"은 통상적으로 본원에 기재된 바와 같은 생물학적 관심 소스(예를 들어, 조직 또는 유기체 또는 세포 배양물)로부터 수득되거나 유래된 샘플을 지칭한다. 일부 실시형태에서, 관심 소스는 유기체, 예컨대 동물 또는 인간을 포함한다. 다른 실시형태에서, 관심 소스는 미생물, 예컨대 박테리아, 바이러스, 원생동물 또는 진균을 포함한다. 추가의 실시형태에서, 관심 소스는 합성 조직, 유기체, 세포 배양, 핵산 또는 다른 물질일 수 있다. 다른 추가의 실시형태에서, 관심 소스는 식물 기반 유기체일 수 있다. 또 다른 실시형태에서, 샘플은 예를 들어 물 샘플, 토양 샘플, 고고학적 샘플과 같은 환경 샘플, 또는 살아 있지 않은 소스로부터 수집된 다른 샘플일 수 있다. 다른 실시형태에서, 샘플은 다중-유기체 샘플(예를 들어, 혼합된 유기체 샘플)일 수 있다. 일부 실시형태에서, 생물학적 샘플은 생물학적 조직 또는 유체이거나 이를 포함한다. 일부 실시형태에서, 생물학적 샘플은 골수; 혈액; 혈액 세포; 복수; 조직 샘플, 생검 샘플 또는 또는 미세침 흡기 샘플; 세포-함유 체액; 자유 부유하는 핵산; 단백질-결합된 핵산, 리보

단백질-결합된 핵산; 가래; 타액; 뇨; 뇌척수액, 복막액; 흉수; 대변; 림프; 부인과학적 유체; 피부 면봉; 질 면봉; 질세포진(pap smear), 구강 면봉; 코 면봉; 세척액 또는 세척물, 예컨대 젖관 세척물 또는 기관지폐포 세척물; 질액, 흡인물; 부스러기; 골수 시편; 조직 생검 시편; 태아 조직 또는 유체; 수술 시편; 대변, 다른 체액, 분비물 및/또는 배설물; 및/또는 이들로부터의 세포 등이거나 이를 포함할 수 있다. 일부 실시형태에서, 생물학적 샘플은 개체로부터 얻은 세포이거나 이를 포함한다. 일부 실시형태에서, 얻은 세포는 샘플이 얻어진 개체로부터의 세포이거나 이를 포함한다. 일부 실시형태에서, 세포-파생물, 예컨대 세포기관 또는 소낭 또는 엑소좀. 특정 실시형태에서, 생물학적 샘플은 대상체로부터 얻은 액체 생검이다. 일부 실시형태에서, 샘플은 임의의 적절한 수단에 의해 관심 소스로부터 직접 얻은 "1차 샘플"이다. 예를 들어, 일부 실시형태에서, 1차 생물학적 샘플은 생검(예를 들어, 미세침 흡기 또는 조직 생검), 수술, 체액(예를 들어, 혈액, 림프, 대변 등)의 수집 등으로 이루어진 군으로부터 선택된 방법에 의해 얻어진다. 일부 실시형태에서, 상황에서 명확한 것처럼, 용어 "샘플"은 1차 샘플을 처리하여(예를 들어, 이 샘플의 하나 이상의 성분을 제거함으로써 그리고/또는 하나 이상의 물질을 이 샘플에 첨가함으로써) 얻은 제제를 지칭한다. 예를 들어, 반투과성 막을 사용한 여과. 이러한 "처리된 샘플"은 샘플로부터 추출되거나 1차 샘플을 mRNA의 증폭 또는 역전사, 소정의 성분의 단리 및/또는 정제 등과 같은 기법으로 처리함으로써 얻은 예를 들어 핵산 또는 단백질을 포함할 수 있다.

[0040] ***암 질병***: 일 실시형태에서, 유전독성 연관된 질병 또는 장애는 전이할 수 있는 비정상 세포의 이상조절된 성장을 일반적으로 특징으로 하는 것으로 당해 분야에서의 경험자에게 친숙한 "암 질병"이다. 본 기술내용의 하나 이상의 양태를 사용하여 검출 가능한 암 질병은 많은 것들 중에서, 비제한적인 예로서, 전립선암(즉, 선암, 소세포), 난소암(예를 들어, 난소 선암, 장액성 암종 또는 배아 암종, 난황 주머니 종양, 기형종), 간암(예를 들어, HCC 또는 간세포종, 혈관육종), 혈장 세포 종양(예를 들어, 다발성 골수종, 형질구성 백혈병, 형질세포종, 아밀로이드증, 발덴스트롬 마크로글로불린혈증), 결장직장암(예를 들어, 결장 선암, 결장 점액소 선암, 카르시노이드, 림프종 및 직장 선암, 직장 편평 암종), 백혈병(예를 들어, 급성 골수성 백혈병, 급성 림프구성 백혈병, 만성 골수성 백혈병, 만성 림프구성 백혈병, 급성 골수아구성 백혈병, 급성 전골수성 백혈병, 급성 골수단핵구성 백혈병, 급성 단핵구성 백혈병, 급성 적백혈병 및 만성 백혈병, T-세포 백혈병, 세자리 증후군(Sezary syndrome), 전신 비만세포증, 모발 세포 백혈병, 만성 골수성 백혈병성 아세포발증), 골수형성이상 증후군, 림프종(예를 들어, 미만성 거대 B-세포 림프종, 피부 T-세포 림프종, 말초 T-세포 림프종, 호지킨 림프종, 비호지킨 림프종, 소포성 림프종, 외투 세포 림프종, MALT 림프종, 변연 세포 림프종, 리히터 형질전환, 이중유전자이상 림프종(double hit lymphoma), 이식 연관된 림프종, CNS 림프종, 림프절외 림프종, HIV-연관된 림프종, 풍토성 림프종, 버킷 림프종, 이식-연관된 림프증식성 신생물 및 림프구성 림프종 등), 자궁경부암(편평 자궁경부 암종, 투명 세포 암종, HPV 연관된 암종, 자궁경부 육종 등), 식도암(식도 편평 세포 암종, 선암, 소정의 등급의 바렛 식도, 식도 선암), 흑색종(진피 흑색종, 포도막 흑색종, 말단 흑색종, 무색소성 흑색종 등), CNS 종양(예를 들어, 핍지교종, 성상세포종, 교모세포종, 뇌수막종, 신경초종, 두개인두종 등), 췌장암(예를 들어, 선암, 선편평 암종, 반지 고리 세포 암종, 간모양선 암종, 콜로이드 암종, 도세포 암종, 췌장 신경내분비 암종 등), 위장 기질 종양, 육종(예를 들어, 섬유육종, 점액육종, 지방육종, 연골육종, 골원성 육종, 혈관육종, 내피종 육종, 림프관육종, 림프혈관내피종 육종, 평활근육종, 유잉 육종 및 횡문근육종, 방추세포 종양 등), 유방암(예를 들어, 염증성 암종, 대엽성 암종, 유관 암종 등), ER-양성 암, HER-2 양성 암, 방광암(편평 방광암, 소세포 방광암, 요로상피암 등), 두경부암(예를 들어, 두경부의 편평 세포 암종, HPV-연관된 편평 세포 암종, 비인두 암종 등), 폐암(예를 들어, 비소세포 폐암종, 대세포 암종, 기관지성 암종, 편평 세포 암, 소세포 폐암 등), 전이성 암, 구강암, 자궁암(평활근육종, 평활근종 등), 고환암(예를 들어, 정상피종, 비정상피종 및 배아 암종 난황 주머니 종양 등), 피부암(예를 들어, 편평 세포 암종 및 기저 세포 암종, 머켈 세포 암종, 흑색종, 피부 t-세포 림프종 등), 갑상선암(예를 들어, 유두상 암종, 수질성 암종, 미분화 갑상선암 등), 위암, 상피내암, 골암, 담관암, 눈암, 후두암, 신장암(예를 들어, 신장 세포 암종, 윌름스 종양 등), 위암, 아세포종(예를 들어, 신아세포종, 수모세포종, 혈관모세포종, 신경모세포종, 망막아세포종 등), 골수증식성 신생물(진성 다혈구증, 본태성 고혈소판증, 골수섬유증 등), 척삭종, 활막종, 중피종, 선암, 땀샘 암종, 피지선 암종, 낭샘암종, 담도 암종, 융모암종, 상피 암종, 뇌실상의종, 송과체부종양, 속귀 신경집종, 신경초종, 뇌수막종, 뇌하수체 선종, 신경초 종양, 소장의 암, 크롬친화성세포종, 소세포 폐암, 복막 중피종, 부갑상선기능항진 샘종, 부신암, 불명 원발성 암, 내분비계의 암, 음경의 암, 요도의 암, 피부 또는 눈내 흑색종, 부인과학 종양, 소아의 고형 종양, 또는 중추 신경계의 신생물, 원발성 종격동 생식 세포 종양, 부정형 가능성의 클론성 조혈증, 무증상 골수종, 의미 불명 단일클론성 감마글로불린증, 단일클론성 B-세포 림프구증가증, 저등급 암, 클론성 시야 결손, 전신생물성 신생물, 요관암, 자가면역-연관된 암(즉, 궤양성 대장염, 원발성 경화성 담관염, 셀리악병), 유전 소인과 연관된 암(즉, 예컨대 *BRCA1*, *BRCA2*, *TP53*, *PTEN*, *ATM* 등에서 유전 결함을 보유

하는 것) 및 다양한 유전자 증후군, 예컨대 MEN1, MEN2 삼염색체 21 등) 및 자궁에서 화학물질에 노출될 때 발생하는 것(즉, 디에틸스틸베스트롤[DES]에 노출된 여성의 여자 자손에서 투명 세포 암)을 포함한다.

[0041] ***암 유발자 또는 암 유발자 유전자***: 본원에 사용된 바와 같이, "암 유발자" 또는 "암 유발자 유전자"는 세포가 올바른 상황에서 악성 형질전환을 겪게 할 가능성을 갖는 유전자 병변을 지칭한다. 이러한 유전자는 보통 악성 상태 형질전환을 억제하고 돌연변이될 때 소정의 방식으로 더 이상 그렇게 하지 않는 종양 억제자(예를 들어, *TP53*, *BRCA1*)를 포함한다. 다른 유발자 유전자는 돌연변이될 때 소정의 방식에서 구성적으로 활성이 되거나 세포가 악성이 되도록 하는 새로운 특성을 얻는 암유전자(예를 들어, *KRAS*, *EGFR*)일 수 있다. 게놈의 암화화 영역에서 발견된 다른 돌연변이는 암 유발자일 수 있다. 예를 들어, 텔로머라제 유전자(*TERT*)의 프로모터 영역의 돌연변이는 유전자를 과발현시킬 수 있고, 이에 따라 암 유발자가 될 수 있다. 소정의 재배열(예를 들어, *BCR-ABL* 융합)은 하나의 유전자 영역을 다른 것과 병치시켜, 과발현, 억압 손실 또는 키메라 융합 유전자와 관련된 기전을 통해 종양형성을 유발할 수 있다. 대체로, 다른 세포에 비해 증식, 생존 또는 경쟁 이점을 촉진하거나, 능력이 보다 튼튼하게 발달하게 하는 세포에 표현형을 부여하는 유전자 돌연변이(또는 후생변이)는 유발자 돌연변이라 생각될 수 있다. 이것은, 이러한 특징이 결여된 돌연변이가 동일한 유전자에 있도록 발생할 수 있더라도(즉, 동의 돌연변이) 이러한 특징이 결여된 돌연변이와는 대조적일 것이다. 이러한 돌연변이가 종양에서 확인될 때, 이것은 팽창에 의미있게 기여하지 않으면서 클론성 팽창과 함께 "히치하이킹"하므로 패신저 돌연변이라 흔히 칭해진다. 당업자에 의해 인식되는 것처럼, 유발자 및 패신저의 구분이 절대적이지 않고, 그렇게 해석되지 않아야 한다. 일부 유발자는 소정의 상황(예를 들어, 소정의 조직)에서 오직 기능하고, 다른 유발자는 다른 돌연변이 또는 후생변이 또는 다른 인자의 부재 하에 작동하지 않을 수 있다.

[0042] ***대조군 샘플***: 본원에 사용된 바와 같이, "대조군 샘플"은 대조군 샘플이 유전독성 가능성에 대해 평가되는 물질, 환경 또는 공정에 노출되지 않음을 제외하고 이 샘플이 비교되는 샘플과 동일한 방식으로 단리된 샘플을 지칭한다.

[0043] ***결정한다***: 본원에 기재된 많은 방법론은 "결정"의 단계를 포함한다. 본 명세서를 읽는 당업자는 이러한 "결정"이 예를 들어 본원에 명쾌하게 언급된 특정 기법을 포함하여 당업자에게 이용 가능한 임의의 다양한 기법을 사용하거나 이의 사용을 통해 달성될 수 있음을 이해할 것이다. 일부 실시형태에서, 결정은 신체 샘플의 조작을 수반한다. 일부 실시형태에서, 결정은 예를 들어 관련 분석을 수행하도록 적응된 컴퓨터 또는 다른 프로세싱 유닛을 사용하는 데이터 또는 정보의 고려 및/또는 조작을 수반한다. 일부 실시형태에서, 결정은 소스로부터의 관련 정보 및/또는 자료를 수신하는 것을 수반한다. 일부 실시형태에서, 결정은 샘플 또는 집합체의 하나 이상의 특징을 필적하는 기준품과 비교하는 것을 수반한다.

[0044] ***듀플렉스 시퀀싱(DS)***: 본원에 사용된 바와 같이, "듀플렉스 시퀀싱(DS)"은 이의 광의에서 개별 DNA 분자의 가닥 둘 다로부터의 서열을 비교함으로써 예의적 정확도를 달성하는 태그-기반 오류-보정 방법을 지칭한다.

[0045] ***유전독성***: 본원에 사용된 바와 같이, 용어 "유전독성"은 유전 물질(예를 들어, DNA, RNA)에 손상을 야기하는 물질 또는 과정(즉, 유전독소)의 파괴적 특성을 지칭한다. 폴리뉴클레오타이드 손상, 유전자 돌연변이의 형성 및/또는 유전독소에 대한 노출로부터 직접적으로 또는 간접적으로 생긴 일반 핵산 구조의 파괴는 유전독성의 양태이다. 유전독소에 노출된 대상체는 바로 또는 몇년 후 질병 또는 장애(예를 들어, 암)를 잠재적으로 발생시킬 수 있다. 일 실시형태에서, 본 기술내용은 부분적으로 질병 또는 장애 발생의 위험을 예방하거나 감소시키기 위해 그리고/또는 이의 불리한 효과에 대응하기 위해 대상체에서 유전독성을 야기하는 기여하는 사건 및/또는 인자(예를 들어, 물질, 과정)를 확인하는 것에 관한 것이다. 다른 실시형태에서, 유전독성의 개시는 예컨대 유전자 라이브러리의 다양성을 생성하기 위한 설계에 의한다.

[0046] ***유전독소 또는 유전독성 물질 또는 인자***: 본원에 사용된 바와 같이, 용어 "유전독소" 또는 "유전독성 물질 또는 인자"는 예를 들어 핵산 소스(예를 들어, 생물학적 소스, 대상체)가 노출되고/되거나 소비하는 임의의 화학물질, 환경 노출, 및/또는 폴리뉴클레오타이드 손상, 게놈 돌연변이 또는 일반 핵산 구조의 파괴를 야기하는 임의의 촉발 사건(내인성 전구체 돌연변이)을 지칭한다. 일부 실시형태에서, 유전독소는 대상체에서 질병 또는 장애 발생을 직접적으로 또는 간접적으로(예를 들어, 돌연변이성 전구체를 촉발함), 또는 둘 다로 야기하는 능력을 갖는다. 본 기술내용에 의해 검출될 수 있는 유전독성 인자 또는 물질은, 비제한적인 예로서, 화학물질 또는 화학물질의 혼합물(예를 들어, 의약품, 산업용 첨가제 및 부산물-폐기물, 석유 증류물, 중금속, 화장품, 가정용 세척제, 공기 매개 미립자, 식품 제품, 제조 부산물, 오염물질, 가소제, 세제 등); 및 방사선(입자 방사선, 광자 또는 둘 다) 및/또는 자연 환경 또는 (예를 들어, 장치로부터) 인공에 의해 생긴 물리적 힘(예를 들어, 자기장, 중력장, 가속력 등)을 포함한다. 유전독소는 액체, 고체, 및/또는 에어로졸 제형을 추가로 포함

할 수 있고, 이의 노출은 임의의 투여 경로를 통해서일 수 있다. 유전독성 물질 또는 인자는 생물학적 소스의 외부로부터 생긴 노출과 같이 외인성일 수 있거나, 다른 경우에, 유전독성 물질 또는 인자는 생물학적 소스에 내인성일 수 있거나, 또는 이들의 조합일 수 있다. 외인성으로 생긴 물질 또는 인자는 이러한 노출이 내인성으로 처리되면 유전독성이 될 수 있다. 또 다른 예에서, 물질 또는 인자는 하나 이상의 추가 물질 또는 인자와 합해질 때 유전독성이 될 수 있고, 일부 경우에 상승 효과를 가질 수 있다. 유전독성 인자 또는 물질의 추가 예는 (예를 들어, 대상체의 감염을 통해) 노출 시 대상체에서 핵산 손상을 직접적으로 또는 간접적으로 야기할 수 있는 유기체, 예컨대 비제한적인 예로서, 방광암에 기여하는 주혈흡충증, 자궁경부암 또는 두경부암에 기여하는 HPV, 머켈 세포 암종에 기여하는 폴리오마 바이러스, 위암에 기여하는 헬리코박터 파일로리, 편평 세포 암종에 기여하는 피부 상처의 만성 박테리아 감염 등을 추가로 포함할 수 있다. 추가 유전독성 물질 또는 인자는 유전독성 물질, 예컨대 비제한적인 예로서, 아스페르길루스 플라부스(aspergillus flavus)로부터의 아플라톡신 또는 식물의 아리스토콜리아 패밀리(aristocholia family)로부터의 아리스토로크산 등을 생산(예를 들어, 이것 내에 또는 분비)할 수 있는 유기체를 추가로 포함할 수 있다. 본 기술내용의 다양한 양태를 사용하여 검출될 수 있는 유전독성 인자 또는 물질은 정확히 정량화되거나 실험적으로 제어될 수 없는 내인성 유전독소, 예컨대 비제한적인 예로서, 스트레스, 염증, 치료제 치료(예를 들어, 유전자 치료, 유전자 편집 치료, 줄기 세포 치료, 다른 세포 치료, 의약품, 방사선촬영 등)의 효과를 추가로 포함할 수 있다. 내인성 인자는 대상체의 노출의 완전한 효과를 반영하는 대상체의 조직에서의 돌연변이 및 다른 유전독성 사건의 집합체 축적을 또한 나타낼 수 있다.

[0047] ***유전독성 연관된 질병 또는 장애***: 본원에 사용된 바와 같이, 용어 "유전독성-연관된 질병 또는 장애"는 대상체에서 하나 이상의 유전독소에 대한 노출에 의해 직접적으로 또는 간접적으로 생긴 게놈 돌연변이 또는 다른 폴리뉴클레오타이드 손상 또는 재배열로부터 생긴 임의의 의학 질환을 지칭한다. 유전독성-연관된 질병 또는 장애는 암-관련되거나 암-비관련될 수 있다. 추가적으로, 폴리뉴클레오타이드 손상/재배열 또는 돌연변이는 생식 세포 또는 체세포에 있을 수 있다. 생식 세포가 이환된 예에서, 유전독성-연관된 질병 또는 장애가 노출된 대상체의 자손인 대상체에서 표출할 수 있다(또는 그렇지 않으면 이의 위험을 부여한다)고 고려된다.

[0048] ***충분히 유전독성인 물질***: 본원에 사용된 바와 같이, 용어 "충분히 유전독성인 물질"은 노출된 하나 이상의 생물학적 유기체로부터 유래될 수 있는 하나 이상의 분자에서 하나 이상의 뉴클레오타이드 잔기에서 약 50%, 약 40%, 약 30%, 약 20%, 약 10%, 약 5%, 약 4%, 약 3%, 약 2%, 약 1%, 약 0.5%, 약 0.1%, 약 0.01%, 약 0.001%, 약 0.0001%, 약 0.00001%, 약 0.000001% 등의 핵산 손상 또는 돌연변이를 발생시킬 확률을 갖는 것으로 본 기술내용의 시스템, 방법 및 키트에 의해 확인된 물질, 인자, 화합물 또는 과정을 지칭한다. 일부 실시형태에서, 충분히 유전독성인 물질은 대조군 배경 수준을 초과하는 핵산 손상 또는 돌연변이를 야기할 약 50% 초과의 확률을 가질 수 있다. 일부 실시형태에서, 충분히 유전독성인 물질은 유전독소에 노출된 대상체에서 질병 또는 장애를 야기할 약 50%, 약 40%, 약 30%, 약 20%, 약 10%, 약 5%, 약 4%, 약 3%, 약 2%, 약 1%, 약 0.5%, 약 0.1%, 약 0.01%, 약 0.001%, 약 0.0001%, 약 0.00001% 등의 확률을 갖는 것으로 본 기술내용의 시스템, 방법 및 키트에 의해 확인된 물질, 인자, 화합물 또는 과정을 지칭한다.

[0049] ***성장을 억제한다***: 본원에 사용된 바와 같이, 암 질병에서 "성장을 억제한다"의 용어는 치료의 부재 하의 세포의 증식 및/또는 세포 크기 성장에 비해 치료에 노출된 세포의 증식 및/또는 세포의 크기/질량의 감소에 의해 입증된 것처럼 예를 들어 약 5%, 약 10%, 약 20%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 80%, 약 90%, 약 95%, 또는 약 99% 이상만큼 생체내 또는 시험관내 세포 성장(예를 들어, 종양 크기, 암 세포 분열 속도 등)을 감소시키는 것을 지칭한다. 성장 억제는 세포에서 아폽토시스를 유도하거나, 세포에서 괴사를 유도하거나, 세포 주기 진행을 느리게 하거나, 세포 대사를 파괴하거나, 세포 용해를 유도하거나, 세포의 증식 및/또는 세포 크기 성장을 감소시키는 일부 다른 기전을 유도하는 치료의 결과일 수 있다.

[0050] ***발현***: 본원에 사용된 바와 같이, 핵산 서열의 "발현"은 하기 사건들 중 하나 이상을 지칭한다: (1) (예를 들어, 전사에 의한) DNA 서열로부터의 RNA 주형의 생산; (2) (예를 들어, 스플라이싱, 편집, 5' 캡 형성, 및/또는 3' 말단 형성에 의한) RNA 전사체의 처리; (3) 폴리펩타이드 또는 단백질로의 RNA의 번역; 및/또는 (4) 폴리펩타이드 또는 단백질의 번역후 변형.

[0051] ***작용 기전***: 본원에 사용된 바와 같이, 용어 "작용 기전"은 유전독소에 대한 노출 후에 핵산을 변경시키는 생화학적 과정을 지칭한다. 일 실시형태에서, "작용 기전"은 완전한 질병 또는 장애 발생까지 게놈 돌연변이 또는 손상을 따르는 생화학적 경로 및 또는 병리생리학적 과정을 지칭한다. 다른 실시형태에서, "작용 기전"은 유전독소 노출 후에 생물학적 소스에서 생기고, 게놈 손상(예를 들어, 전돌연변이성 병변) 또는 돌연변이를 발생시키는 생화학적 경로 및/또는 생리학적 과정을 포함한다. 또 다른 실시형태에서, 유전독성 물질 또는 과정의 작용 기전은 하기들 중 하나 이상으로부터 추론될 수 있다: 영향을 받는 뉴클레오타이드 염기, 도입된 뉴클레오타

이드 변화, 도입된 DNA 손상의 유형, 도입된 구조 변화, 영향을 받는 뉴클레오타이드(들)의 플랭킹 뉴클레오타이드 서열 상황, 영향을 받는 유전적 상황 또는 서열(들), 영향을 받는 전사 상태 또는 영역, 영향을 받는 영역의 메틸화 상태, 유전독소 노출에 의해 영향을 받는 영역의 단백질 결합된 상태 또는 압축 상태 또는 염색체 위치.

[0052] ***돌연변이***: 본원에 사용된 바와 같이, 용어 "돌연변이"는 핵산 서열 또는 구조의 변경을 지칭한다. 폴리뉴클레오타이드 서열의 돌연변이는 복잡한 멀티뉴클레오타이드 변화 중에서 샘플에서의 DNA 서열의 점 돌연변이(예를 들어, 단일 염기 돌연변이), 멀티뉴클레오타이드 돌연변이, 뉴클레오타이드 결실, 서열 재배열, 뉴클레오타이드 삽입 및 중복을 포함할 수 있다. 상보성 염기 변화(즉, 진성 돌연변이)로서, 또는 복구, 파괴 또는 진정한 이중 가닥 돌연변이로 잘못 복구/전환될 가능성을 갖는 다른 가닥(즉, 헤테로듀플렉스)에서가 아니라 하나의 가닥에서의 돌연변이로서 듀플렉스 DNA 분자의 가닥 둘 다에서 돌연변이가 발생할 수 있다.

[0053] ***돌연변이체 빈도***: 본원에 사용된 바와 같이, 때때로 "돌연변이체 빈도"라고도 칭하는 용어 "돌연변이체 빈도"는 시퀀싱된 듀플렉스 염기-쌍의 전체 수마다 검출된 고유한 돌연변이의 수를 지칭한다. 일부 실시형태에서, 돌연변이체 빈도는 오직 특정 유전자, 또는 유전자의 세트 또는 게놈 표적의 세트 내의 돌연변이의 빈도이다. 일부 실시형태에서, 돌연변이체 빈도는 오직 소정의 유형의 돌연변이(예를 들어, A 염기의 전체 수마다 A>T 돌연변이의 수로서 계산되는 A>T 돌연변이의 빈도)를 지칭할 수 있다. 돌연변이가 세포 또는 분자의 집단으로 도입되는 빈도는 다른 것들 중에서 유전독소, 유전독소에 대한 노출의 시간 또는 수준의 양, 시간에 따른 대상체의 연령, 조직 또는 조직구성 유형, 게놈 영역, 돌연변이 유형, 트리뉴클레오타이드 상황, 유전된 유전 배경에 의해 변할 수 있다.

[0054] ***돌연변이 서명***: 본원에 사용된 바와 같이, 용어 "돌연변이 서명" 및 "돌연변이 스펙트럼 또는 스펙트럼들"은 DNA 복제 불충, 외인성 및 내인성 유전독소 노출, 결함성 DNA 복구 경로 및 DNA 효소 편집과 같은 돌연변이유발 과정으로부터 생긴 돌연변이 유형의 특징적인 조합을 지칭한다. 일 실시형태에서, 돌연변이 스펙트럼은 컴퓨터를 사용한 패턴 매칭(예를 들어, 비지도된 계층적 돌연변이 스펙트럼 클러스터링)에 의해 생성된다.

[0055] ***비암성 질병***: 다른 실시형태에서, 유전독성 연관된 질병 또는 장애는 비암성 질병이고; 대신에 이것은 게놈 돌연변이 또는 손상에 의해 생기거나 이로 기인한 또 다른 유형의 질병 또는 장애이다. 비제한적인 예로서, 본 기술내용의 하나 이상의 양태를 사용하여 검출 가능하거나 예측되는 이러한 비암성 유형의 질병 또는 장애는 당뇨병; 자가면역 질병 또는 장애, 불임, 신경퇴행, 조로증, 심혈관 질병, 다른 유전자-매개된 질병의 치료와 연관된 임의의 질병(즉, 화학요법-매개된 신경병증 및 시스플라틴과 같은 화학요법과 연관된 신부전), 알츠하이머병/치매, 비만, 심장 질병, 고혈압, 관절염, 정신병, 다른 신경학적 장애(신경섬유종증) 및 다인자 유전 장애(예를 들어, 환경 인자에 의해 촉발된 소인)를 포함한다.

[0056] ***핵산***: 본원에 사용된 바와 같이, 이의 광의에서 올리고뉴클레오타이드 사슬로 혼입되거나 혼입될 수 있는 임의의 화합물 및/또는 물질을 지칭한다. 일부 실시형태에서, 핵산은 포스포디에스테르 연결을 통해 올리고뉴클레오타이드 사슬로 혼입되거나 혼입될 수 있는 화합물 및/또는 물질이다. 문맥에서 명확한 것처럼, 일부 실시형태에서, "*핵산*"은 개별 핵산 잔기(예를 들어, 뉴클레오타이드 및/또는 뉴클레오사이드)를 지칭하고; 일부 실시형태에서, "*핵산*"은 개별 핵산 잔기를 포함하는 올리고뉴클레오타이드 사슬을 지칭한다. 일부 실시형태에서, "*핵산*"은 RNA이거나 이를 포함하고; 일부 실시형태에서, "*핵산*"은 DNA이거나 이를 포함한다. 일부 실시형태에서, 핵산은 하나 이상의 자연 핵산 잔기이거나 이를 포함하거나 이것으로 이루어진다. 일부 실시형태에서, 핵산은 하나 이상의 핵산 유사체이거나 이를 포함하거나 이것으로 이루어진다. 일부 실시형태에서, 핵산 유사체는 포스포디에스테르 골격을 사용하지 않는다는 점에서 핵산과 다르다. 예를 들어, 일부 실시형태에서, 핵산은 당해 분야에 알려지고 골격에서 포스포디에스테르 결합 대신에 펩타이드 결합을 갖는 하나 이상의 "*펩타이드 핵산*"이거나 이를 포함하거나 이것으로 이루어지고, 본 기술내용의 범위 내에 고려된다. 대안적으로, 또는 추가적으로, 일부 실시형태에서, 핵산은 포스포디에스테르 결합보다 하나 이상의 포스포로티오에이트 및/또는 5'-N-포스포르아미디트 연결을 갖는다. 일부 실시형태에서, 핵산은 하나 이상의 자연 뉴클레오사이드(예를 들어, 아데노신, 티미딘, 구아노신, 사이티딘, 우리딘, 데옥시아데노신, 데옥시티미딘, 데옥시 구아노신 및 데옥시사이티딘)이거나 이를 포함하거나 이것으로 이루어진다. 일부 실시형태에서, 핵산은 하나 이상의 뉴클레오사이드 유사체(예를 들어, 2-아미노아데노신, 2-티오티미딘, 이노신, 피롤로-피리미딘, 3-메틸 아데노신, 5-메틸사이티딘, C-5 프로피닐-사이티딘, C-5 프로피닐-우리딘, 2-아미노아데노신, C5-브로모우리딘, C5-플루오로우리딘, C5-요오도우리딘, C5-프로피닐-우리딘, C5-프로피닐-사이티딘, C5-메틸사이티딘, 2-아미노아데노신, 7-데아자아데노신, 7-데아자구아노신, 8-옥소아데노신, 8-옥소구아노신, O(6)-메틸구아닌, 2-티오사이티딘, 메틸화 염기, 인터칼레이팅 염기, 및 이들의 조합)이거나 이를 포함하거나 이것으로 이루어진다. 일부 실시형태에서, 핵산은 자연 핵산에서

의 것과 비교하여 하나 이상의 변형된 당(예를 들어, 2'-플루오로리보스, 리보스, 2'-데옥시리보스, 아라비노스 및 헥소스)을 포함한다. 일부 실시형태에서, 핵산은 RNA 또는 단백질과 같은 기능적 유전자 산물을 암호화하는 뉴클레오타이드 서열을 갖는다. 일부 실시형태에서, 핵산은 하나 이상의 인트론을 포함한다. 일부 실시형태에서, 핵산은 자연 소스로부터의 단리, 상보성 주형에 기초한 중합에 의한 효소 합성(생체내 또는 시험관내), 재조합 세포 또는 시스템에서의 생식 및 화학 합성 중 하나 이상에 의해 제조된다. 일부 실시형태에서, 핵산은 적어도 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 15개, 20개, 25개, 30개, 35개, 40개, 45개, 50개, 55개, 60개, 65개, 70개, 75개, 80개, 85개, 90개, 95개, 100개, 110개, 120개, 130개, 140개, 150개, 160개, 170개, 180개, 190개, 200개, 225개, 250개, 275개, 300개, 325개, 350개, 375개, 400개, 425개, 450개, 475개, 500개, 600개, 700개, 800개, 900개, 1000개, 1500개, 2000개, 2500개, 3000개, 3500개, 4000개, 4500개, 5000개 이상의 길이의 잔기이다. 일부 실시형태에서, 핵산은 부분적으로 또는 완전히 단일 가닥이고; 일부 실시형태에서, 핵산은 부분적으로 또는 완전히 이중-가닥이다. 일부 실시형태에서, 핵산은 2차 구조를 가지며 분지될 수 있다. 일부 실시형태에서, 핵산은 폴리펩타이드를 암화하거나 이를 암화하는 서열의 보체인 적어도 하나의 요소를 포함하는 뉴클레오타이드 서열을 갖는다. 일부 실시형태에서, 핵산은 효소 활성을 갖는다. 일부 실시형태에서, 핵산은 예를 들어 리보핵단백질 복합체 또는 운반 RNA에서 기계적 기능을 한다.

[0057] ***약학적 조성물 또는 제형***: 본원에 사용된 바와 같이, 용어 "약학적 조성물"은 약물학적 유효량의 활성 약물 또는 활성제 및 약학적으로 허용 가능한 담체를 포함한다. 일부 예에서, 본 기술내용의 다양한 양태는 약학적 조성물 또는 제형, 또는 이것 내의 활성 약물 또는 물질의 유전독성을 평가하도록 사용될 수 있다.

[0058] ***폴리뉴클레오타이드 손상***: 본원에 사용된 바와 같이, 용어 "폴리뉴클레오타이드 손상" 또는 "핵산 손상"은 유전독소에 의해 직접적으로 또는 간접적으로(예를 들어, 대사물질, 또는 손상을 주거나 돌연변이성인 과정의 유도) 생긴 대상체의 데옥시리보핵산(DNA) 서열("DNA 손상") 또는 리보핵산(RNA) 서열에 대한 손상("RNA 손상")을 지칭한다. 손상된 핵산은 대상체에서 유전독소 노출과 연관된 질병 또는 장애의 발생으로 이어질 수 있다. 일부 실시형태에서, 대상체에서 손상된 핵산의 검출은 유전독소 노출의 표시일 수 있다. 폴리뉴클레오타이드 손상은 세포에서 화학적 및/또는 물리적 DNA 변형을 추가로 포함할 수 있다. 일부 실시형태에서, 그 손상은 비제한적인 예로서 산화, 알킬화, 탈아미노화, 메틸화, 가수분해, 하이드록실화, 닉킹, 가닥내 가교, 가닥간 가교, 무딘 말단 가닥 절단, 엇갈린 말단 이중 가닥 절단, 포스포릴화, 탈포스포릴화, 수모일화, 글라이코실화, 탈글라이코실화, 푸트레시닐화, 카복실화, 할로겐화, 포밀화, 단일-가닥 갭, 열에 의한 손상, 건조에 의한 손상, UV 노출에 의한 손상, X-방사선으로부터의 감마 방사선 손상에 의한 손상, 이온화 방사선에 의한 손상, 비이온화 방사선에 의한 손상, 중입자 방사선에 의한 손상, 핵 붕괴에 의한 손상, 베타-방사선에 의한 손상, 알파 방사선에 의한 손상, 중성자 방사선에 의한 손상, 양성자 방사선에 의한 손상, 은하 방사선에 의한 손상, 높은 pH에 의한 손상, 낮은 pH에 의한 손상, 반응성 산화성 종에 의한 손상, 자유 라디칼에 의한 손상, 퍼옥사이드에 의한 손상, 차아염소산염에 의한 손상, 포르말린 또는 폼알데하이드와 같은 조직 고정에 의한 손상, 반응성 철에 의한 손상, 낮은 이온성 조건에 의한 손상, 높은 이온성 조건에 의한 손상, 비완충 조건에 의한 손상, 뉴클레아제에 의한 손상, 환경 노출에 의한 손상, 화재에 의한 손상, 기계적 스트레스에 의한 손상, 효소 분해에 의한 손상, 미생물에 의한 손상, 예비적 기계적 전단에 의한 손상, 예비적 효소 단편화에 의한 손상, 생체내 자연적으로 생긴 손상, 핵산 추출 동안 생긴 손상, 시퀀싱 라이브러리 제조 동안 생긴 손상, 중합효소에 의해 도입된 손상, 핵산 복구 동안 도입된 손상, 핵산 말단-꼬리화 동안 생긴 손상, 핵산 결찰 동안 생긴 손상, 시퀀싱 동안 생긴 손상, DNA의 기계적 취급에서 생긴 손상, 나노기공을 통한 통과 동안 생긴 손상, 유기체에서의 노화의 일부로 생긴 손상, 개체의 화학적 노출의 결과로서 생긴 손상, 돌연변이원에 의해 생긴 손상, 발암물질에 의해 생긴 손상, 클라스토젠(clastogen)에 의해 생긴 손상, 산소 노출로 인해 생체내 염증 손상에 의해 생긴 손상, 하나 이상의 가닥 파괴로 인한 손상, 및 임의의 이들의 조합 중 적어도 하나이거나 이를 포함한다.

[0059] ***기준품***: 본원에 사용된 바와 같이 비교가 수행되는 표준품 또는 대조군을 기술한다. 예를 들어, 일부 실시형태에서, 관심이 있는 물질, 동물, 개체, 집단, 샘플, 서열 또는 값은 한 위치에 존재하거나 전자 수단을 통해 원격으로 접근될 수 있는 물리적 또는 컴퓨터 데이터베이스에서 기준품 또는 대조군 물질, 동물, 개체, 집단, 샘플, 서열 또는 값 또는 이의 표시와 비교된다. 일부 실시형태에서, 기준품 또는 대조군은 관심이 있는 시험 또는 결정과 실질적으로 동시에 시험되고/되거나 결정된다. 일부 실시형태에서, 기준품 또는 대조군은 선택적으로 실감형 매체에서 구현되는 계층적 기준품 또는 대조군이다. 통상적으로, 당업자가 이해하는 것처럼, 기준품 또는 대조군은 평가되는 것과 필적하는 조건 또는 상황 하에 결정되거나 규명된다. 당업자는 특정한 가능한 기준품 또는 대조군에 대한 의존 및/또는 비교를 정당화하기 위해 충분한 유사성이 존재할 때를 이해할 것이다. "기준품 샘플"은 시험 대상체와는 다르고, 이 샘플이 비교되고 알려진 분량의 동일한 유전독성 물질에 노출된 샘플과 동일한 방식으로 단리되는 대상체로부터의 샘플을 지칭한다. 기준품 샘플의 대상체는 시험 대상체와 유전적

으로 동일할 수 있거나 상이할 수 있다. 또한, 기준품 샘플은 알려진 분량의 동일한 유전독성 물질에 노출된 몇몇 대상체로부터 유래될 수 있다.

[0060] ***안전한 역치 수준***: 본원에 사용된 바와 같이, 용어 "안전한 역치 수준"은 질병 발생으로 이어지는 그럴듯한 게놈 돌연변이가 발생하기 전에 대상체가 노출될 수 있는 특정 유전독소 또는 유전독소의 조합의 양(예를 들어, 중량, 부피, 농도, 질량, 몰 풍부도, 단위*시간 적분 등)을 지칭한다. 예를 들어, 안전한 역치 수준은 0일 수 있다. 다른 예에서, 유전독소 노출의 수준은 관용적일 수 있다. 허용 가능한 노출 위험의 관용은 대상체, 연령, 성별, 조직 유형, 환자의 건강 컨디션 및 당해 분야의 숙련자에게 익숙한 다른 위험-이익 고려사항 등에 따라 다를 수 있다.

[0061] ***안전한 역치 돌연변이체 빈도***: 본원에 사용된 바와 같이, 용어 "안전한 역치 돌연변이체 빈도"는 유전독성 물질 또는 과정에 의해 생긴 허용 가능한 돌연변이율을 지칭하고, 이 돌연변이율 아래에서 대상체는 허용 가능한 유전독성-연관된 질병 또는 장애를 획득할 허용 가능한 위험을 취한다. 허용 가능한 노출 위험의 관용 및 생성된 돌연변이율은 대상체, 연령, 성별, 조직 유형, 환자의 건강 컨디션 등에 따라 다를 수 있다.

[0062] ***단일 분자 식별자(SMI)***: 본원에 사용된 바와 같이, 용어 "단일 분자 식별자" 또는 "SMI"(single molecule identifier)(다른 명칭들 중에서 "태그", "바코드", "분자 바코드", "고유 분자 식별자" 또는 "UMI"라 칭해질 수 있음)는 더 큰 불균질한 분자 집단 중에서 개별 분자를 실질적으로 구별할 수 있는 임의의 물질(예를 들어, 뉴클레오타이드 서열, 핵산 분자 특징)를 지칭한다. 일부 실시형태에서, SMI는 외인성으로 적용된 SMI이거나 이를 포함할 수 있다. 일부 실시형태에서, 외인성으로 적용된 SMI는 축퇴성 서열 또는 반축퇴성 서열이거나 이를 포함할 수 있다. 일부 실시형태에서, 실질적으로 축퇴성인 SMI는 랜덤 고유 분자 식별자(R-UMI: Random Unique Molecular Identifier)로 알려질 수 있다. 일부 실시형태에서, SMI는 알려진 코드의 풀 내로부터 코드(예를 들어, 핵산 서열)를 포함할 수 있다. 일부 실시형태에서, 미리규정된 SMI 코드는 한정 고유 분자 식별자(D-UMI: Defined Unique Molecular Identifier)로 알려져 있다. 일부 실시형태에서, SMI는 내인성 SMI이거나 이를 포함할 수 있다. 일부 실시형태에서, 내인성 SMI는 표적 서열의 특정 전단점, 표적 서열을 포함하는 개별 분자의 말단 끝에 관한 특징, 또는 개별 분자의 말단으로부터 알려진 거리에서의 또는 이것에 인접한 또는 이것 내의 특정 서열에 관한 정보이거나 이를 포함할 수 있다. 일부 실시형태에서, SMI는 랜덤 또는 반랜덤 손상, 화학 변형, 효소 변형 또는 핵산 분자에 대한 다른 변형에 의해 생긴 핵산 분자의 서열 변이와 관련될 수 있다. 일부 실시형태에서, 그 변형은 메틸사이토신의 탈아미노화일 수 있다. 일부 실시형태에서, 그 변형은 핵산 닉의 부위를 포함할 수 있다. 일부 실시형태에서, SMI는 외인성 요소 및 내인성 요소 둘 다를 포함할 수 있다. 일부 실시형태에서, SMI는 물리적으로 인접한 SMI 요소를 포함할 수 있다. 일부 실시형태에서, SMI 요소는 분자에서 공간상 구별될 수 있다. 일부 실시형태에서 SMI는 비핵산일 수 있다. 일부 실시형태에서, SMI는 2개 이상의 상이한 유형의 SMI 정보를 포함할 수 있다. SMI의 다양한 실시형태는 본원에 그 전문이 참조로 포함된 국제 특허 공보 WO 제2017/100441호에 추가로 개시된다.

[0063] ***가닥 한정 요소(SDE)***: 본원에 사용된 바와 같이, 용어 "가닥 한정 요소" 또는 "SDE"는 이중-가닥 핵산 물질의 특정 가닥의 확인 및 이에 따라 다른/상보성 가닥으로부터의 구별이 가능하게 하는 임의의 물질(예를 들어, 표적 이중-가닥 핵산으로부터 생긴 2개의 단일 가닥 핵산의 각각의 증폭 산물이 시퀀싱 또는 다른 핵산 정보획득 후 실질적으로 서로 구별 가능하게 하는 임의의 물질)를 지칭한다. 일부 실시형태에서, SDE는 어댑터 서열 내의 실질적으로 비상보성인 서열의 하나 이상의 분절을 포함할 수 있다. 특정 실시형태에서, 어댑터 서열 내의 실질적으로 비상보성인 서열의 분절은 Y-형상 또는 "루프" 형상을 포함하는 어댑터 분자에 의해 제공될 수 있다. 다른 실시형태에서, 어댑터 서열 내의 실질적으로 비상보성인 서열의 분절은 어댑터 서열 내의 인접한 상보성 서열의 중간에서 쌍을 짓지 않는 "버블"을 형성할 수 있다. 다른 실시형태에서, SDE는 핵산 변형을 포괄할 수 있다. 일부 실시형태에서, SDE는 쌍 지은 가닥이 물리적으로 분리된 반응 구획으로 물리적으로 분리되는 것을 포함할 수 있다. 일부 실시형태에서, SDE는 화학 변형을 포함할 수 있다. 일부 실시형태에서, SDE는 변형된 핵산을 포함할 수 있다. 일부 실시형태에서, SDE는 핵산 분자에 대한 랜덤 또는 반랜덤 손상, 화학 변형, 효소 변형 또는 다른 변형에 의해 생긴 핵산 분자의 서열 변이와 관련될 수 있다. 일부 실시형태에서, 그 변형은 메틸사이토신의 탈아미노화일 수 있다. 일부 실시형태에서, 그 변형은 핵산 닉의 부위를 수반할 수 있다. SDE의 다양한 실시형태는 본원에 그 전문이 참조로 포함된 국제 특허 공보 WO 제2017/100441호에 추가로 개시되어 있다.

[0064] ***대상체***: 본원에 사용된 바와 같이, 용어 "대상체"는 유기체, 통상적으로 포유류, 예컨대 인간(일부 실시형태에서, 태아기 인간 형태를 포함), 비인간 동물(예를 들어, 비제한적인 예로서 비인간 영장류, 말, 양, 개, 고양이, 돼지, 닭, 양서류, 파충류, 해양-생물(일반적으로 바다 원숭이를 배제), 다른 모델 유기체, 예컨대 벌레, 파리 등을 포함하는 포유류 및 비포유류), 및 형질전환 동물(예를 들어, 형질전환 설치류) 등을 지칭한다.

일부 실시형태에서, 대상체는 유전독소 또는 유전독성 인자 또는 물질에 노출되거나, 다른 실시형태에서, 대상체는 잠재적인 유전독소에 노출되었다. 일부 실시형태에서, 대상체는 관련 질병, 장애 또는 질환을 겪는다. 일부 실시형태에서, 대상체는 유전독성 연관된 질병 또는 장애를 겪는다. 일부 실시형태에서, 대상체는 질병, 장애 또는 질환에 걸리기 쉽다. 일부 실시형태에서, 대상체는 질병, 장애 또는 질환의 하나 이상의 증상 또는 특징을 나타낸다. 일부 실시형태에서, 대상체는 질병, 장애 또는 질환의 임의의 증상 또는 특징을 나타내지 않는다. 일부 실시형태에서, 대상체는 질병, 장애 또는 질환의 감수성 또는 위험에 특징적인 하나 이상의 특징을 갖는다. 일부 실시형태에서, 대상체는 질병, 장애 또는 질환의 증상 또는 특징을 나타내고, 일부 실시형태에서, 이러한 증상 또는 특징은 유전독성 연관된 질병 또는 장애와 연관된다. 일부 실시형태에서, 대상체는 환자이다. 일부 실시형태에서, 대상체는 진단 및/또는 치료가 주어지는 개체이다. 또 다른 실시형태에서, 대상체는 유전독소에 노출될 수 있고, 예컨대 생체내 연구를 위해 예를 들어 유기체, 세포, 및/또는 조직, 예를 들어 진균, 원생동물, 박테리아, 고세균, 바이러스, 배양에서 단리된 세포, 의도적으로(예를 들어, 줄기 세포 이식물, 장기 이식물) 또는 의도하지 않고(즉, 태아 또는 모체 마이크로키메리즘) 있는 세포 또는 단리된 핵산 또는 세포기관(즉, 미토콘드리아, 엽록체, 자유 바이러스 게놈, 자유 플라스미드, 압타머, 리보자임 또는 핵산의 유도체 또는 전구체(즉, 올리고뉴클레오타이드, 디뉴클레오타이드 트리포스페이트 등)를 포함할 수 있는 임의의 살아 있는 생물학적 소스 또는 다른 핵산 물질을 지칭한다.

[0065] ***실질적으로***: 본원에 사용된 바와 같이, 용어 "실질적으로"는 관심이 있는 특징 또는 특성의 전체 또는 거의 전체의 규모 또는 정도를 나타내는 정성적 조건을 지칭한다. 생물학적 분야의 당업자는 생물학적 현상 및 화학적 현상이, 설사 그렇더라도, 완전해 지고/지거나 완전성으로 진행하거나 절대 결과를 달성하거나 회피하지 않음을 이해할 것이다. 용어 "실질적으로"는 따라서 본원에서 많은 생물학적 현상 및 화학적 현상에 고유한 완정성의 잠재적인 결여를 포착하도록 사용된다.

[0066] ***치료학적 유효량***: 본원에 사용된 바와 같이, 용어 "치료학적 유효량" 또는 "약물학적 유효량" 또는 단순히 "유효량"은 의도된 약물학적 결과, 치료학적 결과 또는 예방학적 결과를 생성하도록 활성 약물 또는 물질의 양을 지칭한다. 일부 예에서, 본 기술내용의 다양한 양태는 활성 약물 또는 물질(예를 들어, 유전독성-연관된 사건을 목적상 유도하도록 전달된 활성 약물)의 유효량을 평가하거나 결정하도록 사용될 수 있다.

[0067] ***트리뉴클레오타이드 또는 트리뉴클레오타이드 상황***: 본원에 사용된 바와 같이, 용어 "트리뉴클레오타이드" 또는 "트리뉴클레오타이드 상황"은 서열에서 바로 앞의 및 바로 뒤의 뉴클레오타이드 염기(예를 들어, 3개-모노뉴클레오타이드 조합 내의 모노뉴클레오타이드)의 상황 내의 뉴클레오타이드를 지칭한다.

[0068] ***트리뉴클레오타이드 스펙트럼 또는 서명***: 본원에서, 용어 "트리뉴클레오타이드 서명"은 "트리뉴클레오타이드 스펙트럼"과 상호 교환되어 사용되고, "삼중항 서명" 및 "삼중항 스펙트럼"은 트리뉴클레오타이드 상황에서 유전독소 노출과 연관된 것과 같은 돌연변이 서명을 지칭한다. 일 실시형태에서, 유전독소는 고유한, 반고유한 및/또는 그렇지 않으면 확인 가능한 삼중항 스펙트럼/서명을 가질 수 있다.

[0069] ***치료***: 본원에 사용된 바와 같이, 용어 "치료"는 질병, 질병의 증상 또는 질병에 대한 소인을 고치거나 치유하거나 완화하거나 없애거나 변경하거나 구제하거나 경감시키거나 개선하거나 영향을 미칠 목적으로 대상체에 대한 치료제의 도포 또는 투여, 또는 장애, 예를 들어 질병 또는 질환, 질병의 증상, 또는 질병에 대한 소인을 갖는 대상체로부터의 단리된 조직 또는 세포주에 대한 치료제의 도포 또는 투여를 지칭한다. 하나의 예에서, 장애 또는 질병/질환은 유전독성 질병 또는 장애이다. 다른 예에서, 장애 또는 질병/질환은 유전독성 질병 또는 장애가 아니다. 일부 예에서, 본 기술내용의 다양한 양태는 치료 또는 잠재적인 치료의 유전독성을 평가하기 위해 사용된다.

[0070] 듀플렉스 시퀀싱 방법 및 연관된 어댑터 및 시약의 선택된 실시형태

[0071] 듀플렉스 시퀀싱은 이중 가닥 핵산 분자로부터의 오류-보정된 DNA 서열을 제조하는 방법이고, 이것은 원래 국제 특허 공보 WO 제2013/142389호 및 미국 특허 제9,752,188호, 및 WO 제2017/100441호, Schmitt *et. al.*, PNAS, 2012 [1]에; Kennedy *et. al.*, PLOS Genetics, 2013 [2]에; Kennedy et. al., Nature Protocols, 2014 [3]에; 그리고 Schmitt et. al., Nature Methods, 2015 [4]에 기재되어 있다. 각각의 상기 언급된 특허, 특허 출원 및 공보는 그 전문이 본원에 참조로 포함된다. 도 1a 내지 도 1c, 및 본 기술내용의 소정의 양태에 예시된 것처럼, 듀플렉스 시퀀싱은 파생 서열 리드가 차세대 시퀀싱(NGS: next generation sequencing)으로 또한 흔히 알려진 대량 병렬 시퀀싱(MPS: massively parallel sequencing) 동안 동일한 이중-가닥 핵산 모분자로부터 기원한 것으로 인식되지만, 또한 시퀀싱 이후 구별 가능한 집합체로서 서로 구별될 수 있는 방식으로 개별 DNA 분자의 가닥 둘 다를 독립적으로 시퀀싱하도록 사용될 수 있다. 이후, 각각의 가닥으로부터의 생성

된 서열 리드는 듀플렉스 공통 서열(DCS: Duplex Consensus Sequence)로 알려진 원래의 이중-가닥 핵산 분자의 오류-보정된 서열을 얻을 목적을 위해 비교된다. 듀플렉스 시퀀싱의 공정은 원래의 이중 가닥 핵산 분자의 가닥 둘 다가 DCS를 형성하기 위해 사용되는 생성된 시퀀싱 데이터에 나타난다는 것을 명쾌하게 확증할 수 있게 한다.

[0072] 소정의 실시형태에서, DS를 도입하는 방법은 이중-가닥 표적 핵산 복합체를 제조하기 위해 제1 가닥 표적 핵산 서열 및 제2 가닥 표적 핵산 서열을 포함하는 표적 이중-가닥 핵산 분자에 대한 하나 이상의 시퀀싱 어댑터의 결찰을 포함할 수 있다(예를 들어, 도 **1a**).

[0073] 다양한 실시형태에서, 생성된 표적 핵산 복합체는 외인성으로 적용된 축퇴성 서열 또는 반축퇴성 서열(예를 들어, 도 **1a**에 도시된 무작위화된 듀플렉스 태그, 도 **1a**에서 α 및 β로 확인된 서열), 표적 이중-가닥 핵산 분자의 특정 전단점과 관련된 내인성 정보, 또는 이들의 조합을 포함할 수 있는 적어도 하나의 SMI 서열을 포함할 수 있다. SMI는 시퀀싱되는 집단에서 표적-핵산 분자가 복수의 다른 분자로부터 단독으로 또는 이것이 결찰된 핵산 단편의 구별 가능한 요소와 조합되어 실질적으로 구별 가능하게 할 수 있다. SMI 요소의 실질적으로 구별 가능한 특징은 이중-가닥 핵산 분자를 형성하는 각각의 단일 가닥에 의해 독립적으로 보유될 수 있어서, 각각의 가닥의 파생 증폭 산물은 시퀀싱 후 동일한 원래의 실질적으로 고유한 이중-가닥 핵산 분자로부터 나온 것으로 인식될 수 있다. 다른 실시형태에서, SMI는 추가 정보를 포함할 수 있고/있거나, 이러한 분자 구별 기능성이 유용한 다른 방법, 예컨대 상기 언급된 공보에 기재된 것에 사용될 수 있다. 다른 실시형태에서, SMI 요소는 어댑터 결찰 후 도입될 수 있다. 일부 실시형태에서, SMI는 자연에서 이중-가닥이다. 다른 실시형태에서, 이것은 자연에서 단일-가닥이다(예를 들어, SMI는 어댑터의 단일-가닥 부분(들)에 있을 수 있음). 다른 실시형태에서, 이것은 자연에서 단일-가닥 및 이중-가닥의 조합이다.

[0074] 일부 실시형태에서, 각각의 이중-가닥 표적 핵산 서열 복합체는 표적 이중-가닥 핵산 분자를 형성하는 2개의 단일-가닥 핵산의 증폭 산물이 시퀀싱 후 서로 실질적으로 구별 가능하게 하는 요소(예를 들어, SDE)를 추가로 포함할 수 있다. 일 실시형태에서, SDE는 시퀀싱 어댑터 내에 포함된 비대칭적 프라이머 부위를 포함할 수 있거나, 다른 배열에서 서열 비대칭은 프라이머 서열 내가 아닌 어댑터 분자로 도입될 수 있어서, 표적 핵산 서열 복합체의 제1 가닥 및 표적 핵산 서열 복합체의 제2 가닥의 뉴클레오타이드 서열에서의 적어도 하나의 위치는 증폭 및 시퀀싱 후에 서로 상이하다. 다른 실시형태에서, SMI는 정규 뉴클레오타이드 서열 A, T, C, G 또는 U와 상이하지만, 2개의 증폭되고 시퀀싱된 분자에서 적어도 하나의 정규 뉴클레오타이드 서열 차이로 전환되는 2개의 가닥 사이에 다른 생화학적 비대칭을 포함할 수 있다. 또 다른 실시형태에서, SDE는 증폭 전에 2개의 가닥을 물리적으로 분리시키는 수단일 수 있어서, 제1 가닥 표적 핵산 서열 및 제2 가닥 표적 핵산 서열로부터의 파생 증폭 산물은 2개 사이의 구별을 유지시킬 목적을 위해 서로로부터 실질적인 물리적 이격에서 유지된다. 제1 가닥 및 제2 가닥의 구별을 허용하는 SDE 기능을 제공하기 위한 다른 이러한 배열 또는 방법론, 예컨대 상기 언급된 공보에 기재된 것 또는 기재된 기능 목적을 제공하는 다른 방법을 사용할 수 있다.

[0075] 적어도 하나의 SMI 및 적어도 하나의 SDE를 포함하는 이중-가닥 표적 핵산 복합체를 생성한 후에, 또는 이들 요소들 중 하나 또는 둘 다가 후속하여 도입되는 경우, 이 복합체는 예컨대 PCR과 같은 DNA 증폭, 또는 DNA 증폭의 임의의 다른 생화학적 방법(예를 들어, 회전 환 증폭, 다중 변위 증폭, 등온 증폭, 브리지 증폭 또는 표면-결합 증폭)으로 처리될 수 있어서, 제1 가닥 표적 핵산 서열의 하나 이상의 카피 및 제2 가닥 표적 핵산 서열의 하나 이상의 카피가 제조된다(예를 들어, 도 **1b**). 이후, 제1 가닥 표적 핵산 분자의 하나 이상의 증폭 카피 및 제2 표적 핵산 분자의 하나 이상의 증폭 카피는 바람직하게는 "차세대" 대량 병렬 DNA 시퀀싱 플랫폼을 사용하여 DNA 시퀀싱으로 처리될 수 있다(예를 들어, 도 **1b**).

[0076] 원래의 이중-가닥 표적 핵산 분자로부터 유래된 제1 가닥 표적 핵산 분자 및 제2 가닥 표적 핵산 분자 중 어느 하나로부터 제조된 서열 리드는 관련된 실질적으로 고유한 SMI를 공유함에 기초하여 확인되고, SDE에 의해 반대의 가닥 표적 핵산 분자로부터 구별될 수 있다. 일부 실시형태에서, SMI는 수학적으로-기초한 오류 보정 코드(예를 들어, 해밍 코드(Hamming code))에 기초한 서열일 수 있고, 이로써 소정의 증폭 오류, 시퀀싱 오류 또는 SMI 합성 오류는 원래의 듀플렉스(예를 들어, 이중-가닥 핵산 분자)의 상보성 가닥에서 SMI 서열의 서열을 관련시킬 목적을 위해 관용될 수 있다. 예를 들어, SMI가 정규 DNA 염기의 완전히 축퇴성인 서열의 15개의 염기 쌍을 포함하는 이중 가닥 외인성 SMI에 의해, 추정된 4^15 = 1,073,741,824 SMI 변이체는 완전히 축퇴성인 SMI의 집단에 존재할 것이다. 2개의 SMI가 10,000개의 샘플링된 SMI의 집단 중에서 SMI 서열 내에 1개의 뉴클레오타이드가 다른 시퀀싱 데이터의 판독치로부터 회수되면, 이것은 수학적으로 계산될 수 있고, 이의 확률은 랜덤 선택으로 발생하고, 단일 염기 쌍 차이가 상술된 오류 유형 중 하나를 더욱 반영할 것 같은지 및 SMI 서열이 사실 동일한 원래의 듀플렉스 분자로부터 유래되는 것으로 결정될 수 있는지에 대한 결정이 이루어진다. 일부 실시형

태에서, SMI가 적어도 부분적으로 서열 변이체가 서로 완전히 축퇴성이 아닌 외인성으로 적용된 서열이고 적어도 부분적으로 알려진 서열인 경우, 알려진 서열의 정체는 일부 실시형태에서 상술된 유형의 하나 이상의 오류가 하나의 알려진 SMI 서열의 정체를 다른 SMI 서열의 정체로 전환시키지 않는 방식으로 설계될 수 있어서, 하나의 SMI가 다른 SMI의 정체로 잘못 해석될 가능성이 감소한다. 일부 실시형태에서, 이 SMI 설계 전략은 해밍 코드 접근법 또는 이의 도함수를 포함한다. 제1 가닥 표적 핵산 분자로부터 제조된 하나 이상의 서열 리드는 확인되면 제2 가닥 표적 핵산 분자로부터 제조된 하나 이상의 서열 리드와 비교되어 오류-보정된 표적 핵산 분자 서열을 제조한다(예를 들어, 도 1c). 예를 들어, 제1 가닥 및 제2 가닥 표적 핵산 서열 둘 다로부터의 염기가 동의하는 뉴클레오타이드 위치는 진정한 서열인 것으로 간주되는 한편, 2개의 가닥 사이에 비동의하는 뉴클레오타이드 위치는 무시되거나 제거되거나 보정되거나 그렇지 않으면 확인될 수 있는 기술적 오류의 잠재적인 부위로 인식된다. 원래의 이중-가닥 표적 핵산 분자의 오류-보정된 서열이 따라서 제조될 수 있다(도 1c에 도시됨). 일부 실시형태에서, 제1 가닥 표적 핵산 분자 및 제2 가닥 표적 핵산 분자로부터 제조된 각각의 시퀀싱 리드의 별개의 그룹화 후에, 단일-가닥 공통 서열은 각각의 제1 가닥 및 제2 가닥에 생성될 수 있다. 이후, 제1 가닥 표적 핵산 분자 및 제2 가닥 표적 핵산 분자로부터의 단일-가닥 공통 서열은 오류-보정된 표적 핵산 분자 서열을 제조하도록 비교될 수 있다(예를 들어, 도 1c).

[0077] 대안적으로, 일부 실시형태에서, 2개의 가닥 사이의 서열 비동의의 부위는 원래의 이중 가닥 표적 핵산 분자의 생물학적으로-유래된 미스매치의 잠재적인 부위로 인식될 수 있다. 대안적으로, 일부 실시형태에서, 2개의 가닥 사이의 서열 비동의의 부위는 원래의 이중 가닥 표적 핵산 분자에서의 DNA 합성-유래된 미스매치의 잠재적인 부위로 인식될 수 있다. 대안적으로, 일부 실시형태에서, 2개의 가닥 사이의 서열 비동의의 부위는 손상된 뉴클레오타이드 염기 또는 변형된 뉴클레오타이드 염기가 하나의 가닥 또는 둘 다의 가닥에 존재하고, 효소 과정(예를 들어, DNA 중합효소, DNA 글라이코실라제 또는 다른 핵산 변형 효소 또는 화학 공정)에 의해 미스매치로 전환되는 잠재적인 부위로 인식될 수 있다. 일부 실시형태에서, 이 후자의 발견은 효소 과정 또는 화학 처리 전에 핵산 손상 또는 뉴클레오타이드 변형의 존재를 추론하기 위해 사용될 수 있다.

[0078] 일부 실시형태에서, 본 기술내용의 양태에 따르면, 본원에 기술된 듀플렉스 시퀀싱 단계로부터 생성된 시퀀싱 리드는 DNA-손상된 분자(예를 들어, 저장 동안, 선적 동안, 조직 또는 혈액 추출 동안 또는 후에, 라이브러리 제조 동안 또는 후에 기타 등등에서 손상된)로부터 시퀀싱 리드를 제거하기 위해 추가로 여과될 수 있다. 예를 들어, DNA 복구 효소, 예컨대 우라실-DNA 글라이코실라제(UDG), 포름아미도피리미딘 DNA 글라이코실라제(FPG) 및 8-옥소구아닌 DNA 글라이코실라제(OGG1)는 DNA 손상(예를 들어, 시험관내 DNA 손상 또는 생체내 손상)을 제거하거나 보정하기 위해 사용될 수 있다. 이 DNA 복구 효소는 예를 들어 DNA로부터 손상된 염기를 제거하는 글라이코실라제이다. 예를 들어, UDG는 (사이토신의 자발적 가수분해에 의해 초래된) 사이토신 탈아미노화로부터 생긴 우라실을 제거하고, FPG는 8-옥소-구아닌(예를 들어, 반응성 산소 종으로부터 생긴 흔한 DNA 병변)을 제거한다. FPG는 또한 비염기성 부위에서 1개 염기 갭을 생성할 수 있는 리가제 활성을 갖는다. 예를 들어, 중합효소가 주형을 카피하지 못하므로, 이러한 비염기성 부위는 일반적으로 후속하여 PCR에 의해 증폭하지 못할 것이다. 따라서, 이러한 DNA 손상 복구/제거 효소의 사용은 진성 돌연변이를 갖지 않는 손상된 DNA를 효과적으로 제거할 수 있고, 그렇지 않으면 시퀀싱 및 듀플렉스 서열 분석 후에 오류로서 검출되지 않을 것이다. 상보성 오류가 가닥 둘 다에서 동일한 위치에서 이론적으로 발생하는 드문 경우에 손상된 염기로 인한 오류는 대개 듀플렉스 시퀀싱에 의해 보정될 수 있지만, 이에 따라 오류-증가 손상의 감소는 인공산물의 개연성을 감소시킬 수 있다. 더욱이, 라이브러리 제조 동안, 시퀀싱되는 DNA의 소정의 단편은 이의 소스 또는 프로세싱 단계(예를 들어, 기계적 DNA 전단)로부터의 단일-가닥일 수 있다. 이 영역은 통상적으로 당해 분야에 알려진 "말단 복구" 단계 동안 이중 가닥 DNA로 전환되고, 이로써 DNA 중합효소 및 뉴클레오사이드 기질은 DNA 샘플에 첨가되어 5' 오목한 말단을 연장시킨다. 카피되는 DNA의 단일-가닥 부분에서의 DNA 손상의 돌연변이성 부위(즉, DNA 듀플렉스 또는 내부 단일-가닥 닉 또는 갭의 하나의 말단 또는 둘 다의 말단에서의 단일-가닥 5' 오버행)는 필인(fill-in) 반응 동안 오류를 야기할 수 있는데, 이 반응은 단일-가닥 돌연변이, 합성 오류 또는 핵산 손상의 부위가 진성 돌연변이로서 최종 듀플렉스 공통 서열에서 잘못 해석되는 이중-가닥 형태가 되게 하여서, 진성 돌연변이는 사실 원래의 이중 가닥 핵산 분자에 존재하지 않을 때 이 핵산 분자에 존재한다. "슈도-듀플렉스"라 불리는 이 시나리오는 이러한 손상 파괴/복구 효소의 사용에 의해 감소되거나 방지될 수 있다. 다른 실시형태에서, 이 발생은 원래의 듀플렉스 분자의 단일-가닥 부분을 파괴하거나 이의 형성을 방지하는 전략의 사용(예를 들어, 기계적 전단보다는 원래의 이중 가닥 핵산 물질을 단편화하기 위해 사용되는 소정의 효소 또는 닉 또는 갭을 남길 수 있는 소정의 다른 효소의 사용)을 통해 감소되거나 제거될 수 있다. 다른 실시형태에서, 원래의 이중-가닥 핵산(예를 들어, 단일-가닥 특이적 뉴클레아제, 예컨대 S1 뉴클레아제 또는 녹두 뉴클레아제)의 단일-가닥 부분을 제거하는 과정의 사용은 유사한 목적에 사용될 수 있다.

[0079] 추가의 실시형태에서, 본원에 기술된 듀플렉스 시퀀싱 단계로부터 생성된 시퀀싱 리드는 슈도듀플렉스 인공산물에 가장 경향이 있는 리드의 말단을 손질함으로써 거짓 돌연변이를 제거하도록 추가로 여과될 수 있다. 예를 들어, DNA 단편화는 이중-가닥 분자의 말단 단부에서 단일 가닥 부분을 생성할 수 있다. 이 단일-가닥 부분은 말단 복구 동안 (예를 들어, Klenow 또는 T4 중합효소에 의해) 충전될 수 있다. 일부 경우에, 중합효소는 "슈도듀플렉스 분자"를 생성시키는 이 말단 복구된 영역에서 카피 실수를 만든다. 라이브러리 제조의 이 인공산물은 시퀀싱되면 진성 돌연변이인 것으로 부정확하게 나타날 수 있다. 이 오류는 말단 복구 기전의 결과로서 더 높은 위험 영역에서 발생할 수 있는 임의의 돌연변이를 배제하도록 시퀀싱 리드의 말단을 손질하여서 거짓 돌연변이의 수를 감소시킴으로써 시퀀싱 후 분석으로부터 제거되거나 감소될 수 있다. 일 실시형태에서, 시퀀싱 리드의 이러한 손질은 자동적으로 달성될 수 있다(예를 들어, 일반 공정 단계). 다른 실시형태에서, 돌연변이체 빈도는 단편 말단 영역에 대해 평가될 수 있고, 돌연변이의 역치 수준이 단편 말단 영역에서 관찰되면, 시퀀싱 리드 손질은 DNA 단편의 이중-가닥 공통 서열 리드를 생성하기 전에 수행될 수 있다.

[0080] 특정 예로서, 일부 실시형태에서, 이중-가닥 표적 핵산 물질을 적어도 하나의 어댑터 서열에 결찰하여, 어댑터-표적 핵산 물질 복합체를 형성하는 단계를 포함하는 이중-가닥 표적 핵산 물질의 오류-보정된 서열 리드를 생성하는 방법이 본원에 제공되고, 여기서 적어도 하나의 어댑터 서열은 (a) 이중-가닥 표적 핵산 물질의 각각의 분자를 고유하게 표지하는 축퇴성 또는 반축퇴성 단일 분자 식별자(SMI) 서열, 및 (b) 어댑터-표적 핵산 물질 복합체의 각각의 가닥이 이의 상보성 가닥에 대해 명확하게 확인 가능한 뉴클레오타이드 서열을 갖도록 어댑터-표적 핵산 물질 복합체의 제1 가닥을 태그화하는 제1 뉴클레오타이드 어댑터 서열, 및 어댑터-표적 핵산 물질 복합체의 제2 가닥을 태그화하는 제1 뉴클레오타이드 서열에 적어도 부분적으로 비상보성인 제2 뉴클레오타이드 어댑터 서열을 포함한다. 상기 방법은 다음에 어댑터-표적 핵산 물질 복합체의 각각의 가닥을 증폭시켜 복수의 제1 가닥 어댑터-표적 핵산 복합체 앰플리콘 및 복수의 제2 가닥 어댑터-표적 핵산 복합체 앰플리콘을 제조하는 단계를 포함할 수 있다. 상기 방법은 제1 및 가닥 둘 다를 증폭시켜 제1 핵산 산물 및 제2 핵산 산물을 제공하는 단계를 추가로 포함할 수 있다. 상기 방법은 또한 각각의 제1 핵산 산물 및 제2 핵산 산물을 시퀀싱하여 복수의 제1 가닥 서열 리드 및 복수의 제2 가닥 서열 리드를 제조하는 단계, 및 적어도 하나의 제1 가닥 서열 리드 및 적어도 하나의 제2 가닥 서열 리드의 존재를 확증하는 단계를 포함할 수 있다. 상기 방법은 적어도 하나의 제1 가닥 서열 리드를 적어도 하나의 제2 가닥 서열 리드와 비교하는 단계, 및 동의하지 않는 뉴클레오타이드 위치를 무시함으로써 이중-가닥 표적 핵산 물질의 오류-보정된 서열 리드를 생성하거나, 대안적으로 비교된 제1 가닥 서열 리드 및 제2 가닥 서열 리드가 비상보성인 하나 이상의 뉴클레오타이드 위치를 갖는 비교된 제1 가닥 서열 리드 및 제2 가닥 서열 리드를 제거하는 단계를 추가로 포함할 수 있다.

[0081] 추가 특정 예로서, 일부 실시형태에서, 핵산 물질(예를 들어, 이중-가닥 표적 DNA 분자)의 가닥 둘 다를 적어도 하나의 비대칭적 어댑터 분자에 결찰하여, 이중-가닥 표적 DNA 분자의 제1 가닥과 연관된 제1 뉴클레오타이드 서열(예를 들어, 상부 가닥) 및 이중-가닥 표적 DNA 분자의 제2 가닥과 연관된 제1 뉴클레오타이드 서열에 적어도 부분적으로 비상보성인 제2 뉴클레오타이드 서열(예를 들어, 하부 가닥)을 갖는 어댑터-표적 핵산 물질 복합체를 형성하는 단계, 및 어댑터-표적 핵산 물질의 각각의 가닥을 증폭시켜, 증폭된 어댑터-표적 핵산 산물의 구별되지만 관련된 세트를 생성하는 각각의 가닥을 생성하는 단계를 포함하는, 샘플로부터 DNA 변이체를 확인하는 방법이 본원에 제공된다. 상기 방법은 복수의 제1 가닥 어댑터-표적 핵산 산물 및 복수의 제2 가닥 어댑터-표적 핵산 산물의 각각을 시퀀싱하는 단계, 어댑터-표적 핵산 물질 복합체의 각각의 가닥으로부터 적어도 하나의 증폭된 서열 리드의 존재를 확증하는 단계 및 제1 가닥으로부터 얻은 적어도 하나의 증폭된 서열 리드를 제2 가닥으로부터 얻은 적어도 하나의 증폭된 서열 리드와 비교하여 (예를 들어, 기준 서열과 비교된) 공통 서열 리드에서 특정 위치에서 생긴 변이체가 진정한 DNA 변이체로서 확인되도록 핵산 물질(예를 들어, 이중-가닥 표적 DNA 분자)의 가닥 둘 다의 서열이 동의하는 뉴클레오타이드 염기만을 갖는 핵산 물질의 공통 서열 리드(예를 들어, 이중-가닥 표적 DNA 분자)를 형성하는 단계를 추가로 포함할 수 있다.

[0082] 일부 실시형태에서, 이중-가닥 핵산 물질로부터 고정확성 공통 서열을 생성하는 방법이 본원에 제공되고, 상기 방법은 개별 듀플렉스 DNA 분자를 어댑터 분자로 태그화하여 태그화된 DNA 물질을 형성하는 단계이되, 각각의 어댑터 분자는 (a) 듀플렉스 DNA 분자를 고유하게 표지하는 축퇴성 또는 반축퇴성 단일 분자 식별자(SMI), 및 (b) 각각의 태그화된 DNA 분자에 대해 표지화된 DNA 물질 내에 각각의 개별 DNA 분자의 원래의 하부 가닥으로부터 원래의 상부 가닥을 구별하는 제1 비상보성 뉴클레오타이드 어댑터 서열 및 제2 비상보성 뉴클레오타이드 어댑터 서열을 포함하는 단계 및 태그화된 DNA 분자의 원래의 상부 가닥의 듀플레케이트의 세트 및 태그화된 DNA 분자의 원래의 하부 가닥의 듀플레케이트의 세트를 생성하여 증폭된 DNA 물질을 형성하는 단계를 포함한다. 상기 방법은 원래의 상부 가닥의 듀플레케이트로부터의 제1 단일 가닥 공통 서열(SSCS: single strand consensus

sequence) 및 원래의 하부 가닥의 듀플레케이트로부터의 제2 단일 가닥 공통 서열(SSCS)을 생성하는 단계, 원래의 상부 가닥의 제1 SSCS를 원래의 하부 가닥의 제2 SSCS와 비교하는 단계 및 원래의 상부 가닥의 제1 SSCS 및 원래의 하부 가닥의 제2 SSCS 둘 다의 서열이 상보성인 뉴클레오타이드 염기만을 갖는 고정확성 공통 서열을 생성하는 단계를 추가로 포함할 수 있다.

[0083] 추가의 실시형태에서, 이중-가닥 표적 DNA 분자를 포함하는 샘플로부터 DNA 손상을 검출하고/하거나 정량화하는 방법이 본원에 제공되고, 상기 방법은 각각의 이중-가닥 표적 DNA 분자의 가닥 둘 다를 적어도 하나의 비대칭적 어댑터 분자에 결찰시켜 복수의 어댑터-표적 DNA 복합체를 형성하는 단계이되, 각각의 어댑터-표적 DNA 복합체는 이중-가닥 표적 DNA 분자의 제1 가닥과 연관된 제1 뉴클레오타이드 서열 및 이중-가닥 표적 DNA 분자의 제2 가닥과 연관된 제1 뉴클레오타이드 서열에 적어도 부분적으로 비상보성인 제2 뉴클레오타이드 서열을 갖는 단계 및 각각의 어댑터 표적 DNA 복합체에 대해 어댑터-표적 DNA 복합체의 각각의 가닥을 증폭시켜, 증폭된 어댑터-표적 DNA 앰플리콘의 구별되지만 관련된 세트를 생성하는 각각의 가닥을 생성시키는 단계를 포함한다. 상기 방법은 복수의 제1 가닥 어댑터-표적 DNA 앰플리콘 및 복수의 제2 가닥 어댑터-표적 DNA 앰플리콘의 각각을 시퀀싱하는 단계, 어댑터-표적 DNA 복합체의 각각의 가닥으로부터 적어도 하나의 서열 리드의 존재를 확증하는 단계 및 제1 가닥으로부터 얻은 적어도 하나의 서열 리드를 제2 가닥으로부터 얻은 적어도 하나의 서열 리드와 비교하여 DNA 손상 부위(들)가 검출되고/되거나 정량화될 수 있도록 이중-가닥 DNA 분자의 하나의 가닥의 서열 리드가 이중-가닥 DNA 분자의 다른 가닥의 서열 리드와 비동의하는(예를 들어, 비상보성인) 뉴클레오타이드 염기를 검출하고/하거나 정량화하는 단계를 추가로 포함할 수 있다. 일부 실시형태에서, 상기 방법은 제1 가닥 어댑터-표적 DNA 앰플리콘으로부터의 제1 단일 가닥 공통 서열(SSCS) 및 제2 가닥 어댑터-표적 DNA 앰플리콘으로부터의 제2 단일 가닥 공통 서열(SSCS)을 생성하는 단계, 원래의 제1 가닥의 제1 SSCS를 원래의 제2 가닥의 제2 SSCS와 비교하는 단계 및 제1 SSCS 및 제2 SSCS의 서열이 비상보성인 뉴클레오타이드 염기를 확인하여 샘플에서 이중-가닥 표적 DNA 분자와 연관된 DNA 손상을 검출하고/하거나 정량화하는 단계를 추가로 포함할 수 있다.

[0084] *단일 분자 식별자 서열(SMI)*

[0085] 다양한 실시형태에 따르면, 제공된 방법 및 조성물은 핵산 물질의 각각의 가닥에서 하나 이상의 SMI 서열을 포함한다. SMI는 독립적으로 이중-가닥 핵산 분자로부터 생긴 각각의 단일 가닥에 의해 보유될 수 있어서, 각각의 가닥의 파생 증폭 산물은 시퀀싱 후 동일한 원래의 실질적으로 고유한 이중-가닥 핵산 분자로부터 나온 것으로 인식될 수 있다. 일부 실시형태에서, SMI는 추가 정보를 포함할 수 있고/있거나, 당업자에 의해 인식되는 것처럼 이러한 분자 구별 기능성이 유용한 다른 방법에 사용될 수 있다. 일부 실시형태에서, SMI 요소는 핵산 물질에 대한 어댑터 서열 결찰 전에, 실질적으로 이것과 동시에 또는 이것 후에 혼입될 수 있다.

[0086] 일부 실시형태에서, SMI 서열은 적어도 하나의 축퇴성 핵산 또는 반축퇴성 핵산을 포함할 수 있다. 다른 실시형태에서, SMI 서열은 비축퇴성일 수 있다. 일부 실시형태에서, SMI는 핵산 분자의 단편 말단(예를 들어, 결찰된 핵산 물질의 무작위로 또는 반무작위로 전단된 말단)과 연관되거나 그 근처인 서열일 수 있다. 일부 실시형태에서, 외인성 서열은 예를 들어 단일 DNA 분자를 서로 구별할 수 있는 SMI 서열을 얻기 위해 결찰된 핵산 물질(예를 들어, DNA)의 무작위로 또는 반무작위로 절단된 말단에 상응하는 서열과 함께 생각될 수 있다. 일부 실시형태에서, SMI 서열은 이중-가닥 핵산 분자에 결찰된 어댑터 서열의 일부이다. 소정의 실시형태에서, SMI 서열을 포함하는 어댑터 서열은 이중-가닥이어서, 이중-가닥 핵산 분자의 각각의 가닥은 어댑터 서열에 대한 결찰 후에 SMI를 포함한다. 다른 실시형태에서, SMI 서열은 이중-가닥 핵산 분자에 대한 결찰 전에 또는 후에 단일-가닥이고, 상보성 SMI 서열은 DNA 중합효소로 반대의 가닥을 연장하여 상보성 이중-가닥 SMI 서열을 생성시킴으로써 생성될 수 있다. 다른 실시형태에서, SMI 서열은 어댑터의 단일-가닥 부분(예를 들어, Y-형상을 갖는 어댑터의 아암)에 있다. 이러한 실시형태에서, SMI는 이중-가닥 핵산 분자의 원래의 가닥으로부터 유래된 서열 리드의 패밀리의 그룹화를 용이하게 할 수 있고, 일부 경우에 이중-가닥 핵산 분자의 원래의 제1 가닥과 제2 가닥 사이에 관계(예를 들어, SMI의 전부 또는 일부는 순람표와 관련될 수 있음)를 부여할 수 있다. 실시형태에서, 제1 가닥 및 제2 가닥이 상이한 SMI로 표지되는 경우, 2개의 원래의 가닥으로부터의 서열 리드는 하나 이상의 내인성 SMI(예를 들어, 단편-특이적 특성, 예컨대 핵산 분자의 단편 단편과 연관되거나 그 근처의 서열)를 사용하여, 또는 2개의 원래의 가닥이 공유하는 추가 분자 태그(예를 들어, 어댑터의 이중-가닥 부분에서의 바코드, 또는 이의 조합)의 사용에 의해 관련될 수 있다. 일부 실시형태에서, 각각의 SMI 서열은 약 1개 내지 약 30개의 핵산(예를 들어, 1개, 2개, 3개, 4개, 5개, 8개, 10개, 12개, 14개, 16개, 18개, 20개 이상의 축퇴성 핵산 또는 반축퇴성 핵산)을 포함할 수 있다.

[0087] 일부 실시형태에서, SMI는 핵산 물질 및 어댑터 서열의 하나 또는 둘 다에 결찰될 수 있다. 일부 실시형태에서, SMI는 핵산 물질의 T-오버행, A-오버행, CG-오버행, 탈하이드록실화 염기 및 무딘 말단 중 적어도 하나에 결찰

될 수 있다.

[0088] 일부 실시형태에서, SMI의 서열은 단일 핵산 분자를 서로 구별할 수 있는 SMI 서열을 얻기 위해 예를 들어 핵산 물질(예를 들어, 결찰된 핵산 물질)의 무작위로 또는 반무작위로 전단된 말단에 상응하는 서열과 함께 고려(또는 이에 따라 설계)될 수 있다.

[0089] 일부 실시형태에서, 적어도 하나의 SMI는 내인성 SMI(예를 들어, 전단점 자체를 사용하여 또는 전단점에 바로 인접한 핵산 물질에서의 한정된 수의 뉴클레오타이드[예를 들어, 전단점으로부터의 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개의 뉴클레오타이드]를 사용하여 예를 들어 전단점(예를 들어, 단편 말단)과 관련된 SMI)일 수 있다. 일부 실시형태에서, 적어도 하나의 SMI는 외인성 SMI(예를 들어, 표적 핵산 물질에서 발견되지 않는 서열을 포함하는 SMI)일 수 있다.

[0090] 일부 실시형태에서, SMI는 영상화 모이어티(예를 들어, 형광 또는 달리 광학적으로 검출 가능한 모이어티)이거나 이를 포함할 수 있다. 일부 실시형태에서, 이러한 SMI는 증폭 단계의 필요 없이 검출 및/또는 정량화를 허용한다.

[0091] 일부 실시형태에서, SMI 요소는 어댑터-표적 핵산 복합체에서 상이한 위치에 위치한 2개 이상의 별개의 SMI 요소를 포함할 수 있다.

[0092] SMI의 다양한 실시형태는 본원에 그 전문이 참조로 포함된 국제 특허 공보 WO 제2017/100441호에 추가로 개시된다.

[0093] *가닥-한정 요소(SDE)*

[0094] 일부 실시형태에서, 이중-가닥 핵산 물질의 각각의 가닥은 표적 이중-가닥 핵산 물질을 형성하는 2개의 단일-가닥 핵산의 증폭 산물이 시퀀싱 후 서로 실질적으로 구별 가능하게 하는 요소를 추가로 포함할 수 있다. 일부 실시형태에서, SDE는 시퀀싱 어댑터 내에 포함된 비대칭적 프라이머 부위이거나 이를 포함할 수 있거나, 다른 배열에서 서열 비대칭은 어댑터 서열로 도입되고 프라이머 서열 내에 없을 수 있어서, 표적 핵산 서열 복합체의 제1 가닥 및 표적 핵산 서열 복합체의 제2 가닥의 뉴클레오타이드 서열에서의 적어도 하나의 위치는 증폭 및 시퀀싱 후에 서로 다르다. 다른 실시형태에서, SDE는 정규 뉴클레오타이드 서열 A, T, C, G 또는 U와 다르지만, 2개의 증폭되고 시퀀싱된 분자에서 적어도 하나의 정규 뉴클레오타이드 서열 차이로 전환되는 2개의 가닥 사이에 다른 생화학적 비대칭을 포함할 수 있다. 또 다른 실시형태에서, SDE는 증폭 전에 2개의 가닥을 물리적으로 분리시키는 수단이거나 이를 포함할 수 있어서, 제1 가닥 표적 핵산 서열 및 제2 가닥 표적 핵산 서열로부터의 파생 증폭 산물은 2개의 파생 증폭 산물 사이에 구별을 유지시킬 목적을 위해 서로로부터 실질적인 물리적 이격에서 유지된다. 제1 가닥 및 제2 가닥의 구별을 허용하는 SDE 기능을 제공하기 위한 다른 이러한 배열 또는 방법론을 사용할 수 있다.

[0095] 일부 실시형태에서, SDE는 루프(예를 들어, 헤어핀 루프)를 형성할 수 있다. 일부 실시형태에서, 루프는 적어도 하나의 엔도뉴클레아제 인식 부위를 포함할 수 있다. 일부 실시형태에서, 표적 핵산 복합체는 루프 내에 절단 사건을 용이하게 하는 엔도뉴클레아제 인식 부위를 함유할 수 있다. 일부 실시형태에서, 루프는 비정규 뉴클레오타이드 서열을 포함할 수 있다. 일부 실시형태에서, 함유된 비정규 뉴클레오타이드는 가닥 절단을 용이하게 하는 하나 이상의 효소에 의해 인식 가능할 수 있다. 일부 실시형태에서, 함유된 비정규 뉴클레오타이드는 루프에서 가닥 절단을 용이하게 하는 하나 이상의 화학 공정에 의해 표적화될 수 있다. 일부 실시형태에서, 루프는 루프에서 가닥 절단을 용이하게 하는 하나 이상의 효소적, 화학적 또는 물리적 공정에 의해 표적화될 수 있는 변형된 핵산 링커를 함유할 수 있다. 일부 실시형태에서, 이 변형된 링커는 광 분해 가능한 링커이다.

[0096] 다양한 다른 분자 도구는 SMI 및 SDE로 작용할 수 있다. 전단점 및 DNA-기반 태그 이외에, 쌍 지은 가닥을 물리적으로 근접하게 유지시키는 단일-분자 구획화 방법 또는 다른 비핵산 태그화 방법은 가닥-관련 기능을 제공할 수 있었다. 유사하게, 물리적으로 분리될 수 있는 방식의 어댑터 가닥의 비대칭적 화학 표지화는 SDE 역할을 제공할 수 있다. 듀플렉스 시퀀싱의 최근에 기재된 변형은 사이토신 메틸화의 형태의 자연 발생 가닥 비대칭을 2개의 가닥을 구별하는 서열 차이로 전환시키도록 바이설파이트 전환을 사용한다. 이 실행이 검출될 수 있는 돌연변이의 유형을 제한하지만, 자연 비대칭에서 자본화의 개념은 변형된 뉴클레오타이드를 직접적으로 검출할 수 있는 떠오르는 시퀀싱의 상황에서 주목할 만하다. SDE의 다양한 실시형태는 그 전문이 참조로 포함된 국제 특허 공보 WO 제2017/100441호에 추가로 개시된다.

[0097] *어댑터 및 어댑터 서열*

[0098] 다양한 배열에서, SMI(예를 들어, 분자 바코드), SDE, 프라이머 부위, 유세포 서열 및/또는 다른 특징을 포함하는 어댑터 분자는 본원에 개시된 많은 실시형태와 사용하기 위해 고려된다. 일부 실시형태에서, 제공된 어댑터는 하기 특성 중 적어도 하나를 갖는 PCR 프라이머(예를 들어, 프라이머 부위)에 상보성 또는 적어도 부분적으로 상보성인 하나 이상의 서열이거나 이를 포함할 수 있다: 1) 높은 표적 특이성; 2) 다중화 가능함; 및 3) 튼튼하고 최소로 바이어스된 증폭을 나타냄.

[0099] 일부 실시형태에서, 어댑터 분자는 "Y"-형상, "U"-형상, "헤어핀" 형상이거나, 버블(예를 들어, 비상보성인 서열의 부분), 또는 다른 특징을 가질 수 있다. 다른 실시형태에서, 어댑터 분자는 "Y"-형상, "U"-형상, "헤어핀" 형상 또는 버블을 포함할 수 있다. 소정의 어댑터는 변형된 뉴클레오타이드 또는 비표준 뉴클레오타이드, 제한 부위, 또는 시험관내 구조 또는 기능의 조작을 위한 다른 특징을 포함할 수 있다. 어댑터 분자는 말단 단부를 갖는 다양한 핵산 물질에 결찰할 수 있다. 예를 들어, 어댑터 분자는 T-오버행, A-오버행, CG-오버행, 다중 뉴클레오타이드 오버행, 탈하이드록실화 염기, 핵산 물질의 무딘 말단 및 분자의 말단에 결찰하기에 적합할 수 있고, 표적의 5'는 탈인산화되거나 달리 전통적인 결찰로부터 차단된다. 다른 실시형태에서, 어댑터 분자는 결찰 부위에서 5' 가닥에서 탈인산화되거나 그렇지 않으면 결찰-방지 변형을 함유할 수 있다. 후자의 2개의 실시형태에서, 이러한 전략은 라이브러리 단편 또는 어댑터 분자의 이합체화를 방지하기에 유용할 수 있다.

[0100] 어댑터 서열은 단일-가닥 서열, 이중-가닥 서열, 상보성 서열, 비상보성 서열, 부분 상보성 서열, 비대칭 서열, 프라이머 결합 서열, 유세포 서열, 결찰 서열 또는 어댑터 분자에 의해 제공된 다른 서열을 의미할 수 있다. 특정 실시형태에서, 어댑터 서열은 올리고뉴클레오타이드에 보체의 방식에 의해 증폭에 사용된 서열을 의미할 수 있다.

[0101] 일부 실시형태에서, 제공된 방법 및 조성물은 적어도 하나의 어댑터 서열을 포함한다(예를 들어, 2개의 어댑터 서열, 하나는 핵산 물질의 5' 말단 및 3' 말단의 각각에 있음). 일부 실시형태에서, 제공된 방법 및 조성물은 2개 이상(예를 들어, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개 이상)의 어댑터 서열을 포함할 수 있다. 일부 실시형태에서, 적어도 2개의 어댑터 서열은 (예를 들어, 서열이) 서로 다르다. 일부 실시형태에서, 각각의 어댑터 서열은 (예를 들어, 서열이) 각각의 다른 어댑터 서열과 다르다. 일부 실시형태에서, 적어도 하나의 어댑터 서열은 적어도 하나의 다른 어댑터 서열의 적어도 일부에 적어도 부분적으로 비상보성이다(예를 들어, 적어도 하나의 뉴클레오타이드에 의해 비상보성임).

[0102] 일부 실시형태에서, 어댑터 서열은 적어도 하나의 비표준 뉴클레오타이드를 포함한다. 일부 실시형태에서, 비표준 뉴클레오타이드는 비염기성 부위, 우라실, 테트라하이드로퓨란, 8-옥소-7,8-디하이드로-2'데옥시아데노신(8-옥소-A), 8-옥소-7,8-디하이드로-2'-데옥시구아노신(8-옥소-G), 데옥시이노신, 5'니트로인돌, 5-하이드록시메틸-2'-데옥시사이티딘, 이소-사이토신, 5'-메틸-이소사이토신 또는 이소구아노신, 메틸화 뉴클레오타이드, RNA 뉴클레오타이드, 리보스 뉴클레오타이드, 8-옥소-구아닌, 광 분해 가능한 링커, 바이오티닐화 뉴클레오타이드, 데스티오바이오틴 뉴클레오타이드, 티올 변형된 뉴클레오타이드, 아크리다이트 변형된 뉴클레오타이드 이소-dC, 이소 dG, 2'-O-메틸 뉴클레오타이드, 이노신 뉴클레오타이드 잠김 핵산, 펩타이드 핵산, 5 메틸 dC, 5-브로모데옥시우리딘, 2,6-디아미노퓨린, 2-아미노퓨린 뉴클레오타이드, 비염기성 뉴클레오타이드, 5-니트로인돌 뉴클레오타이드, 아데닐화 뉴클레오타이드, 아지드 뉴클레오타이드, 디곡시게닌 뉴클레오타이드, I-링커, 5' 헥시닐 변형된 뉴클레오타이드, 5-옥타디닐 dU, 광 절단 가능한 스페이서, 광 절단 불가능한 스페이서, 클릭 화학 적합 변형된 뉴클레오타이드, 및 임의의 이들의 조합으로부터 선택된다.

[0103] 일부 실시형태에서, 어댑터 서열은 자기 특성을 갖는 모이어티(즉, 자기 모이어티)를 포함한다. 일부 실시형태에서, 이 자기 특성은 상자성이다. 일부 실시형태에서, 어댑터 서열이 자기 모이어티를 포함하는 경우(예를 들어, 자기 모이어티를 포함하는 어댑터 서열에 결찰된 핵산 물질), 자기장이 인가될 때, 자기 모이어티를 포함하는 어댑터 서열은 자기 모이어티를 포함하지 않는 어댑터 서열(예를 들어, 자기 모이어티를 포함하지 않는 어댑터 서열에 결찰된 핵산 물질)로부터 실질적으로 분리된다.

[0104] 일부 실시형태에서, 적어도 하나의 어댑터 서열은 SMI의 5'에 위치한다. 일부 실시형태에서, 적어도 하나의 어댑터 서열은 SMI의 3'에 위치한다.

[0105] 일부 실시형태에서, 어댑터 서열은 하나 이상의 링커 도메인을 통해 SMI 및 핵산 물질 중 적어도 하나에 연결될 수 있다. 일부 실시형태에서, 링커 도메인은 뉴클레오타이드로 이루어질 수 있다. 일부 실시형태에서, 링커 도메인은 (예를 들어, 본 개시내용에서 그외 기재된 바대로) 적어도 하나의 변형된 뉴클레오타이드 또는 비뉴클레오타이드 분자를 포함할 수 있다. 일부 실시형태에서, 링커 도메인은 루프이거나 이를 포함할 수 있다.

[0106] 일부 실시형태에서, 이중-가닥 핵산 물질의 각각의 가닥의 말단 중 어느 하나 또는 말단 둘 다에서의 어댑터 서열은 SDE를 제공하는 하나 이상의 요소를 추가로 포함할 수 있다. 일부 실시형태에서, SDE는 어댑터 서열 내에 포함된 비대칭적 프라이머 부위이거나 이를 포함할 수 있다.

[0107] 일부 실시형태에서, 어댑터 서열은 적어도 하나의 SDE 및 적어도 하나의 결찰 도메인(즉, 적어도 하나의 리가제의 활성에 수정 가능한 도메인, 예를 들어 리가제의 활성을 통해 핵산 물질에 결찰하기에 적합한 도메인)이거나 이를 포함할 수 있다. 일부 실시형태에서, 5'에서 3'로, 어댑터 서열은 프라이머 결합 부위, SDE 및 결찰 도메인이거나 이를 포함할 수 있다.

[0108] 듀플렉스 시퀀싱 어댑터를 합성하기 위한 다양한 방법은 이전에 예를 들어 본원에 그 전문이 참조로 포함된 미국 특허 제9,752,188호, 국제 특허 공보 WO 제2017/100441호 및 국제 특허 공보 제PCT/US18/59908호(2018년 11월 8일 제출)에 기재되어 있다.

[0109] *프라이머*

[0110] 일부 실시형태에서, 1) 높은 표적 특이성; 2) 다중화 가능함; 및 3) 튼튼하고 최소로 바이어스된 증폭을 나타냄의 특성 중 적어도 하나를 갖는 하나 이상의 PCR 프라이머는 본 기술내용의 양태에 따라 다양한 실시형태에 사용하기에 고려된다. 다수의 이전의 연구 및 상업 제품은 종래의 PCR-CE에 대해 소정의 이들 기준을 만족시키는 설계된 프라이머 혼합물을 갖는다. 그러나, 이 프라이머 혼합물이 MPS와 사용하기에 항상 최적이 아님에 주의한다. 실제로, 고도로 다중화된 프라이머 혼합물의 개발은 도전적이고 시간 소모적인 공정일 수 있다. 편리하게는, Illumina 및 Promega 둘 다는 최근에 다양한 표준 및 비표준 STR 및 SNP 유전좌위의 튼튼하고 효율적인 증폭을 나타낸 Illumina 플랫폼에 대한 다중화 적합 프라이머 혼합물을 개발하였다. 이 키트가 시퀀싱 전에 이의 표적 영역을 증폭시키기 위해 PCR을 사용하므로, 쌍 지은-말단 시퀀싱 데이터에서의 각각의 리드의 5'-말단은 DNA를 증폭시키기 위해 사용된 PCR 프라이머의 5'-말단에 상응한다. 일부 실시형태에서, 제공된 방법 및 조성물은 변하는 반응 농도, 융점 및 2차 구조 및 프라이머내/프라이머간 상호작용의 최소화를 수반할 수 있는 균일한 증폭을 보장하기 위해 설계된 프라이머를 포함한다. 많은 기법은 MPS 분야에 대해 고도로 다중화된 프라이머 최적화에 대해 기술되어 있다. 특히, 이들 기법은 당해 분야에 잘 기재된 것처럼 앰플리세크(ampliseq) 방법으로 대개 알려져 있다.

[0111] *증폭*

[0112] 제공된 방법 및 조성물은 다양한 실시형태에서 적어도 하나의 증폭 단계를 사용하거나 이의 사용에 있고, 여기서 핵산 물질(또는 이의 부분, 예를 들어 특이적 표적 영역 또는 유전좌위)은 증폭된 핵산 물질(예를 들어, 약간의 수의 앰플리콘 산물)을 형성하도록 증폭된다.

[0113] 일부 실시형태에서, 핵산 물질의 증폭은 SMI 서열이 적어도 부분적으로 유지되도록 제1 어댑터 서열에 존재하는 서열에 적어도 부분적으로 상보성인 적어도 하나의 단일-가닥 올리고뉴클레오타이드를 사용하여 원래의 이중-가닥 핵산 물질로부터 각각의 제1 핵산 가닥 및 제2 핵산 가닥으로부터 유래된 핵산 물질을 증폭시키는 단계를 포함한다. 증폭 단계는 각각의 관심 가닥을 증폭시키기 위해 제2 단일-가닥 올리고뉴클레오타이드를 추가로 포함하고, 이러한 제2 단일-가닥 올리고뉴클레오타이드는 적어도 하나의 단일-가닥 올리고뉴클레오타이드 및 제2 단일-가닥 올리고뉴클레오타이드가 핵산 물질을 효과적으로 증폭시키는 방식으로 배향되도록 (a) 관심 표적 서열에 적어도 부분적으로 상보성이거나, (b) 제2 어댑터 서열에 존재하는 서열에 적어도 부분적으로 상보성일 수 있다.

[0114] 일부 실시형태에서, 샘플에서의 핵산 물질의 증폭은 "관"(예를 들어, PCR 관), 에멀션 액적, 마이크로챔버 및 상기에 기재된 다른 예 또는 다른 알려진 용기에서 핵산 물질의 증폭을 포함할 수 있다.

[0115] 일부 실시형태에서, 적어도 하나의 증폭 단계는 적어도 하나의 비표준 뉴클레오타이드이거나 이를 포함하는 적어도 하나의 프라이머를 포함한다. 일부 실시형태에서, 비표준 뉴클레오타이드는 우라실, 메틸화 뉴클레오타이드, RNA 뉴클레오타이드, 리보스 뉴클레오타이드, 8-옥소-구아닌, 바이오티닐화 뉴클레오타이드, 잠김 핵산, 펩타이드 핵산, 높은-Tm 핵산 변이체, 대립유전자 구별 핵산 변이체, 본원에 그외 기재된 임의의 다른 뉴클레오타이드 또는 링커 변이체 및 임의의 이들의 조합으로부터 선택된다.

[0116] 임의의 분야-적절한 증폭 반응이 일부 실시형태와 적합한 것으로 고려되지만, 특정 예로서, 일부 실시형태에서, 증폭 단계는 중합효소 연쇄 반응(PCR: polymerase chain reaction), 회전 환 증폭(RCA: rolling circle amplification), 다중 변위 증폭(MDA: multiple displacement amplification), 등온 증폭, 에멀션 내의 폴로니

증폭, 비드의 또는 하이드로겔 내의 표면인 표면에서의 브리지 증폭, 및 임의의 이들의 조합이거나 이를 포함할 수 있다.

[0117] 일부 실시형태에서, 핵산 물질의 증폭은 핵산 물질의 각각의 가닥의 5' 말단 및 3' 말단에서 어댑터 서열의 영역에 적어도 부분적으로 상보성인 단일-가닥 올리고뉴클레오타이드의 사용을 포함한다. 일부 실시형태에서, 핵산 물질의 증폭은 관심 표적 영역 또는 표적 서열(예를 들어, 게놈 서열, 미토콘드리아 서열, 플라스미드 서열, 합성으로 제조된 표적 핵산 등)에 적어도 부분적으로 상보성인 적어도 하나의 단일-가닥 올리고뉴클레오타이드 및 어댑터 서열(예를 들어, 프라이머 부위)의 영역에 적어도 부분적으로 상보성인 단일-가닥 올리고뉴클레오타이드의 사용을 포함한다.

[0118] 일반적으로, 튼튼한 증폭, 예를 들어 PCR 증폭은 반응 조건에 고도로 의존적일 수 있다. 다중 PCR은 예를 들어 완충액 조성물, 1가 또는 2가 양이온 농도, 세제 농도, 크라우딩제(즉, PEG, 글리세롤 등) 농도, 프라이머 농도, 프라이머 Tm, 프라이머 설계, 프라이머 GC 함량, 프라이머 변형된 뉴클레오타이드 특성 및 사이클링 조건(즉, 온도 및 연장 시간 및 온도 변화 속도)에 민감할 수 있다. 완충액 조건의 최적화는 어렵고 시간 소모적인 공정일 수 있다. 일부 실시형태에서, 증폭 반응은 이전에 알려진 증폭 프로토콜에 따라 완충액, 프라이머 풀 농도 및 PCR 조건 중 적어도 하나를 사용할 수 있다. 일부 실시형태에서, 새로운 증폭 프로토콜이 생성될 수 있고/있거나 증폭 반응 최적화가 사용될 수 있다. 특정 예로서, 일부 실시형태에서, PCR 최적화 키트, 예컨대 다중, 실시간, GC-농후, 및 억제제-내성 증폭과 같은 다양한 PCR 분야에 부분적으로 최적화된 다수의 미리 제제화된 완충액을 함유하는 Promega®로부터의 PCR Optimization Kit를 사용할 수 있다. 이 미리 제제화된 완충액은 상이한 Mg^{2+} 및 프라이머 농도, 및 프라이머 풀 비율로 신속히 보충될 수 있다. 또한, 일부 실시형태에서, 다양한 사이클링 조건(예를 들어, 열 사이클링)이 평가되고/되거나 사용될 수 있다. 특정 실시형태가 원하는 특정 분야에 적절한지를 평가하는 데 있어서, 다른 양태들 중에서 특이성, 이형접합성 유전좌위에 대한 대립유전자 커버리지 비율, 유전좌위간 균형 및 깊이 중 하나 이상이 평가될 수 있다. 증폭 성공의 측정은 산물의 DNA 시퀀싱, 겔 또는 모세관 전기영동에 의한 산물의 평가 또는 HPLC 또는 다른 크기 분리 방법, 이어서 단편 시각화, 이중-가닥 핵산 결합 염료 또는 형광 프로브를 사용한 용융 곡선 분석, 질량 분석법 또는 당해 분야에 알려진 다른 방법을 포함할 수 있다.

[0119] 다양한 실시형태에 따르면, 임의의 다양한 인자는 특정 증폭 단계의 길이(예를 들어, PCR 반응에서의 사이클의 수 등)에 영향을 미칠 수 있다. 예를 들어, 일부 실시형태에서, 제공된 핵산 물질은 손상되거나 그렇지 않으면 준최적(예를 들어, 분해된 및/또는 오염된)일 수 있다. 이러한 경우에, 원하는 산물이 허용 가능한 정도로 증폭되게 보장하기 위해 보다 긴 증폭 단계가 도움이 될 수 있다. 일부 실시형태에서, 증폭 단계는 각각의 출발 DNA 분자로부터 평균 3개 내지 10개의 시퀀싱된 PCR 카피를 제공할 수 있지만, 다른 실시형태에서, 각각의 제1 가닥 및 제2 가닥의 오직 단일 카피가 필요하다. 특정 이론에 구속되고자 바라지 않으면서, 너무 많거나 너무 적은 PCR 카피가 검정 효율을 감소시키고 궁극적으로 깊이를 감소시킬 수 있다. 일반적으로, 증폭(예를 들어, PCR) 반응에 사용된 핵산(예를 들어, DNA) 단편의 수는 동일한 SMI/바코드 서열을 공유하는 리드의 수를 기술할 수 있는 1차의 조정 가능한 변수이다.

[0120] 핵산 물질

[0121] *유형*

[0122] 다양한 실시형태에 따르면, 임의의 다양한 핵산 물질을 사용할 수 있다. 일부 실시형태에서, 핵산 물질은 정규 당-포스페이트 골격 내에 폴리뉴클레오타이드에 대한 적어도 하나의 변형을 포함할 수 있다. 일부 실시형태에서, 핵산 물질은 핵산 물질에서 임의의 염기 내에 적어도 하나의 변형을 포함할 수 있다. 예를 들어, 비제한적인 예로서, 일부 실시형태에서, 핵산 물질은 이중-가닥 DNA, 단일-가닥 DNA, 이중-가닥 RNA, 단일-가닥 RNA, 펩타이드 핵산(PNA: peptide nucleic acid), 잠금 핵산(LNA: locked nucleic acid) 중 적어도 하나이거나 이를 포함한다.

[0123] *변형*

[0124] 다양한 실시형태에 따르면, 핵산 물질은 특정한 제공된 방법 또는 조성물이 사용되는 분야에 따라 임의의 특정 단계 전에, 이외 실질적으로 동시에 또는 이에 후속하여 하나 이상의 변형을 수용할 수 있다.

[0125] 일부 실시형태에서, 변형은 핵산 물질의 적어도 일부의 복구이거나 이를 포함할 수 있다. 임의의 분야-적절한 핵산 복구 방식이 일부 실시형태와 적합한 것으로 고려되지만, 소정의 예시적인 방법 및 조성물은 따라서 하기

에 및 실시예에 기재되어 있다.

[0126] 비제한적인 예로서, 일부 실시형태에서, DNA 복구 효소, 예컨대 우라실-DNA 글라이코실라제(UDG), 포름아미도피리미딘 DNA 글라이코실라제(FPG) 및 8-옥소구아닌 DNA 글라이코실라제(OGG1)는 DNA 손상(예를 들어, 시험관내 DNA 손상)을 보정하기 위해 사용될 수 있다. 상기에 기술된 것처럼, 이 DNA 복구 효소는 예를 들어 DNA로부터 손상된 염기를 제거하는 글라이코실라제이다. 예를 들어, UDG는 (사이토신의 자발적 가수분해에 의해 생긴) 사이토신 탈아미노화로부터 생긴 우라실을 제거하고, FPG는 8-옥소-구아닌(예를 들어, 반응성 산소 종으로부터 생긴 가장 흔한 DNA 병변)을 제거한다. FPG는 또한 비염기성 부위에서 1개 염기 갭을 생성할 수 있는 리가제 활성을 갖는다. 예를 들어, 중합효소가 주형을 카피하지 못하므로, 이러한 비염기성 부위는 후속하여 PCR에 의해 증폭하지 못할 것이다. 따라서, 이러한 DNA 손상 복구 효소의 사용은 진성 돌연변이를 갖지 않는 손상된 DNA를 효과적으로 제거할 수 있고, 그렇지 않으면 시퀀싱 및 듀플렉스 서열 분석 후에 오류로서 검출되지 않을 것이다.

[0127] 상기에 기술된 것처럼, 추가의 실시형태에서, 본원에 기술된 프로세싱 단계로부터 생성된 시퀀싱 리드는 인공산물에 가장 경향이 있는 리드의 말단을 손질함으로써 거짓 돌연변이를 제거하도록 추가로 여과될 수 있다. 예를 들어, DNA 단편화는 이중-가닥 분자의 말단 단부에서 단일-가닥 부분을 생성할 수 있다. 이 단일-가닥 부분은 말단 복구 동안 (예를 들어, Klenow에 의해) 충전될 수 있다. 일부 경우에, 중합효소는 "슈도듀플렉스 분자"를 생성시키는 이 말단-복구된 영역에서 카피 실수를 만든다. 이 인공산물은 시퀀싱되면 진성 돌연변이인 것으로 나타날 수 있다. 이 오류는 말단 복구 기전의 결과로서 발생할 수 있는 임의의 돌연변이를 배제하도록 시퀀싱 리드의 말단을 손질하여서 거짓 돌연변이의 수를 감소시킴으로써 시퀀싱 후 분석으로부터 제거될 수 있다. 일부 실시형태에서, 시퀀싱 리드의 이러한 손질은 자동적으로 달성될 수 있다(예를 들어, 일반 공정 단계). 일부 실시형태에서, 돌연변이체 빈도는 단편 말단 영역에 대해 평가될 수 있고, 돌연변이의 역치 수준이 단편 말단 영역에서 관찰되면, 시퀀싱 리드 손질은 DNA 단편의 이중-가닥 공통 서열 리드를 생성하기 전에 수행될 수 있다.

[0128] 듀플렉스 시퀀싱의 가닥-비교 기술에 의해 제공된 높은 정도의 오류 보정은 표준 차세대 시퀀싱 방법과 비교하여 여러 차수의 규모로 이중-가닥 핵산 분자의 시퀀싱 오류를 감소시킨다. 이 오류 감소는 거의 모든 유형의 서열에서 시퀀싱의 정확도를 개선하지만, 특히 오류 유발인 것으로 당해 분야에서 잘 알려진 생화학적으로 도전하는 서열에 특히 잘 맞을 수 있다. 이러한 유형의 서열의 하나의 비제한적인 예는 동종중합체 또는 다른 미세부수체/짧은-탠덤 반복부이다. 듀플렉스 시퀀싱 오류 보정으로부터 이익인 오류 유발 서열의 다른 비제한적인 예는 예를 들어 가열, 방사선, 기계적 스트레스, 또는 하나 이상의 뉴클레오타이드 중합효소에 의한 카피 동안 오류 유발인 화학 부가물을 생성하는 다양한 화학적 노출에 의해 손상된 분자 및 또한 분자의 말단에서 또는 닉 및 갭으로서 단일-가닥 DNA를 생성하는 것이다. 추가의 실시형태에서, 듀플렉스 시퀀싱은 또한 이중-가닥 핵산 분자의 집단 중에서 소수의 서열 변이체의 정확한 검출에 사용될 수 있다. 본원의 하나의 비제한적인 예는 대상체 내의 비암성 조직으로부터의 더 많은 수의 돌연변이되지 않은 분자들 중에서 암으로부터 유래된 적은 수의 DNA 분자의 검출이다. 듀플렉스 시퀀싱에 의한 희귀 변이체 검출에 대한 다른 비제한적인 분야는 유전독소 노출로부터 생긴 DNA 손상의 조기 검출이다. 듀플렉스 시퀀싱의 추가의 비제한적인 분야는 유발자 돌연변이로 생긴 유전자 클론을 살핌으로써 유전독성 발암물질 또는 비유전독성 발암물질로부터 생긴 돌연변이의 검출을 위한 것이다. 소수의 서열 변이체의 정확한 검출을 위한 또한 추가의 비제한적인 분야는 유전독소와 연관된 돌연변이성 서명을 생성하는 것이다.

[0129] 유전독성의 확인 및 평가

[0130] 본 기술내용은 유전독성을 평가하기 위한 방법, 시스템, 키트 등에 관한 것이다. 특히, 본 기술내용의 일부 실시형태는 생물학적 소스에서 화합물(예를 들어, 화학적 화합물) 또는 다른 물질의 유전독성 가능성을 평가하기 위한 듀플렉스 시퀀싱에 관한 것이다. 예를 들어, 본 기술내용의 다양한 실시형태는 클론성 선택의 필요 없이 임의의 유기체의 임의의 게놈 상황에서 물질-유도된 돌연변이의 직접적인 측정을 허용하는 듀플렉스 시퀀싱 방법을 수행하는 것을 포함한다. 본 기술내용의 추가의 예는 듀플렉스 시퀀싱을 사용하여 생체내 게놈 돌연변이유발을 검출하고 평가하는 방법에 관한 것이다. 본 기술내용의 다양한 양태는 전임상 및 임상 약물 안전성 시험 둘 다에서의 많은 분야, 및 다른 산업-전반 영향을 갖는다. 예를 들어, 본 기술내용은 나중 해에 질병/장애의 발생을 야기하는 초저 빈도 돌연변이를 검출하는 방법을 포함하고, 여기서 돌연변이는 적어도 하나의 유전독소(예를 들어, 방사선, 발암물질)에 대한 노출의 직접적인 결과로서 및/또는 내인성 소스, 예컨대 DNA 중합효소 오류, 자유 라디칼 및 탈푸린화의 결과로서 발생한다. 검출은 초저 빈도 돌연변이를 확인하기 위해 (예를 들어, 노출의 일 내에) 유전독소에 대한 최근의 노출 후 대상체의 시험을 통해 그리고 듀플렉스 시퀀싱을 사용하여 발생할 수 있다. 특정 예에서, 검출된 초저 빈도 돌연변이는 노출 후 몇년 후에 통상적으로 표출한 질병/장애(예를 들어, 석면에 대한 노출 20년 후의 폐암)를 포함하는 특정 질병 또는 장애를 야기하는 것으로 알려진 돌연변

이와 비교될 수 있다. 본 기술내용은 이에 따라 미래의 노출을 방지하기 위해 그리고 조기 의학 치료를 제공하기 위해 유전독소의 존재 및 이들에 노출된 희생자를 확인하기 위한 편리한 방법을 제공한다. 본 기술내용은 또한 시장 또는 환경으로부터 유전독소를 제거하기 위해 불안전한 소비재, 의약품 및 유전독소를 포함하는 다른 산업용/상업용/제조 부산물을 확인하기 위한 다양한 고속 스크리닝 방법에 사용될 수 있다.

[0131] 특정 실시형태에서, 결실, 파괴 및/또는 재배열과 같은 유전독성 효과는 그 손상이 세포사를 즉시 야기하지 않으면 암 또는 다른 유전독성 연관된 질병 또는 장애로 이어질 수 있다. 예를 들어, 핵산 손상은 대상체가 유전독성 연관된 질병 또는 장애를 발생시키기에 충분할 수 있고/있거나, 노출된 대상체에 이미 존재하는 다른 유형의 질병 또는 장애의 활성화 또는 진행에 기여할 수 있다. 취약 부위라 불리는 절단에 민감한 영역은 유전독성 물질(예를 들어, 화학물질, 예컨대 살충제 또는 소정의 화학요법 약물)로부터 생길 수 있다. 일부 화학물질은 발암성 효과를 야기하는 암유전자가 존재하는 염색체의 영역에서 취약 부위를 유도하는 능력을 갖는다. 더욱이, 살충제, 제조 화합물 또는 다른 해로운 물질의 일부 혼합물에 대한 직업상 노출은 노출된 개체에서 유전독성 손상의 증가와 양으로 상관된다. 예를 들어, 인간 노출 전에 유전독성 가능성의 조사는 임의의 잠재적인 유전독소, 예컨대 잠재적인 약물, 화장품, 소비재, 산업용/제조 생성물 또는 부산물 또는 개발 중인 다른 화학적 화합물에 매우 바람직하다. 마찬가지로, 유전독소에 대한 노출이 의심되는 실시형태에서, 유전독소(들)가 확인되면, 대상체는 표적화된 치료학적 치료를 받을 수 있고/있거나, 유전독소는 대상체 및 다른 사람에 대한 미래의 노출을 방지하기 위해 제거될 수 있다.

[0132] 잠재적인 유전독성 물질 또는 인자의 유전독성 효과를 검출하고 시간 효과적이면서 비용 효과적인 방식으로 잠재적으로 생성된 돌연변이 과정을 정량화하는 능력은 상업적으로 그리고 의학적으로 중요하다. 특정 예에서, 잠재적인 유전독소의 돌연변이 과정을 검출하고 정량화하는 능력은 인간에서 암 위험을 평가하고, 발암물질을 확인하고, 노출의 영향을 예측하는 데 중요할 수 있다. 그러나, 현재의 도구는 느리고, 다루기 힘들고/힘들거나, 이것이 제공하는 정보가 제한된다. 상기에 기재된 바대로, 생체내 시험 및 포유류 리포터 시스템, 예컨대 BigBlue® 마우스 및 래트는 DNA 손상을 야기하는 화합물의 가능성을 결정하기 위한 유효한 유전독성 미터법으로서 식품의약청(FDA) 규제 하에 현재 사용된다.

[0133] 도 **2a**는 잠재적인 유전독소(예를 들어, 잠재적인 돌연변이원)의 생체내 돌연변이유발을 평가하기 위한 다양한 방법론을 보여주는 개념적 예시이다. 도 **2a**에 예시된 각각의 도식에서, 시험 대상체(예를 들어, BigBlue® 마우스, 마우스 모델 유기체, 래트 모델 유기체 등)는 적절한 투여 경로를 사용하여 잠재적인 유전독소(예를 들어, 조사 중인 화합물/물질/인자)에 노출된다. 도 **2a**의 가장 왼쪽에 도시된 하니의 종래의 도식에서, 장기간 설치류 발암성 생체검정은 다양한 용량의 시험 물질에 대한 노출 동안에 또는 후에 신생물성 병변의 발생에 대해 장기간(예를 들어, 2년) 동안 시험 동물을 관찰한다. 시험 동물은 예를 들어 예상된 유형의 인간 노출에 기초하여 경구, 진피 또는 흡입 노출에 의해 투약될 수 있다. 종래의 도식에서, 투약은 통상적으로 대략 2년 지속하지만, 투약 매개변수(예를 들어, 투약 기간, 투여 경로, 투약 수준 또는 다른 투약 요법 매개변수)는 원하는 시험 프로토콜에 따라 설정될 수 있다. 도 **2a**, 왼쪽 도식을 참조하면, 소정의 동물 건강 특징은 연구에 걸쳐 언급되지만, 중요한 평가는 연구가 종료될 때 시험 동물의 조직 및 장기의 완전한 병리학적 분석에 있다.

[0134] 도 **2a**의 중간 도식에 도시된 다른 생체내 검정은 형질전환 설치류를 사용한다. 적절한 단기간 투약 요법(예를 들어, 일 또는 주의 차수의) 후에, 시험 동물은 희생되고, 원하는 조직은 수확되고, DNA는 추출된다. 추출된 DNA로부터, 형질전환 단편은 단리되고, 생성된 정제된 플라스미드는 파지 패키징되고, E. 콜라이로 감염된다. 종래의 형질전환 플라크 검정이 수행되고, 기본적인 돌연변이체 빈도가 계산된다.

[0135] 상기에 기재된 도식의 둘 다가 느리고, 시험된 잠재적인 유전독소의 유전독성(예를 들어, 돌연변이유발)에 관한 매우 제한된 정보를 제공한다. 게놈 유전좌위, 조직 또는 유기체에 의해 제한되지 않는 방식으로 체성 돌연변이를 직접적으로 측정하는 것의 가능성은 흥미롭기는 하지만, 정상 조직의 돌연변이체 빈도(약 10^{-7} 내지 10^{-8})보다 매우 높은 오류율(약 10^{-3}) 때문에 표준 DNA 시퀀싱에 의해 현재 불가능하다.

[0136] 대량 병렬 시퀀싱은 돌연변이성 노출의 생체내 효과에 대해 임의의 유기체의 게놈을 완전히 조사할 가능성을 제공하지만, 기술된 바대로, 종래의 방법은 이러한 돌연변이를 검출하기에 훨씬 너무 부정확하고, 이는 백만에 하나 미만의 수준으로 발생할 수 있다. 예를 들어, 대략 0.1%에서의 차세대 시퀀싱(NGS)의 오류율은 희귀 변이체 및 고유한 분자 프로파일 또는 서명의 검출을 모호하게 하는 배경 노이즈를 생성한다. NGS 플랫폼에서의 오류의 일부 흔한 소스는 (증폭 동안 생긴) PCR 효소, 시퀀서 리드 및 프로세싱 동안의 DNA 손상(예를 들어, 8-옥소-구아닌, 탈아미노화 사이토신, 비염기성 부위 및 기타)을 포함한다.

[0137] 본 기술내용의 양태에 따르면, 듀플렉스 시퀀싱 방법 단계는 자세한 돌연변이체 빈도를 추가로 제공할 수 있는 고정확성 DNA 시퀀싱 리드를 생성할 수 있다(예를 들어, 백만에 하나 미만의 유전독소-유도된 돌연변이를 해소하고, 상이한 돌연변이 과정을 객관적으로 규명하고, 작용 기전을 추론하기 위해 돌연변이 스펙트럼 데이터를 제공함). 예를 들어, 도 2a에 도시된 오른쪽 도식은 돌연변이체 빈도, 돌연변이 유형(들)의 스펙트럼 및 게놈 상황 데이터에 대한 자세한 정보를 또한 제공하면서 종래 기술의 도식과 동일한 시험 대상체에서 잠재적인 유전독소(예를 들어, 잠재적인 돌연변이원)의 유전독성을 신속히 검출하고 평가하기 위한 방법을 포함한다. 게다가, 듀플렉스 시퀀싱 분석은 임의의 유기체로부터의 임의의 조직에서의 임의의 유전자 유전좌위에서의 돌연변이유발의 민감한 검출을 제공할 수 있다. 예를 들어, 도 2a 및 도 2b에 예시된 것처럼, 듀플렉스 시퀀싱 방법 도식은 배양물에서 성장한 세포(예를 들어, 인간 세포, 설치류 세포, 포유류 세포, 비포유류 세포 등)에서 시험 화합물의 시험관내 돌연변이유발을 평가하기 위해(도 2b) 그리고 야생형 설치류(예를 들어, 마우스)에서 시험 화합물의 생체내 돌연변이유발을 평가하기 위해(도 2c) 사용될 수 있다. 예를 들어, 일 실시형태에서, 본 기술내용은 시험 유기체(예를 들어, 설치류, 배양물에서 성장한 세포)를 적절한 투여 경로(예를 들어, 경구로, 피하, 국소, 에어로졸, 근육내 등)에 의해 시험 화합물(예를 들어, 잠재적인 유전독소/돌연변이원)에 노출시키는 것을 포함하는 방법 단계를 포함한다. 일 실시형태에서, 시험 유기체는 짧은 기간(예를 들어, 단일 용량, 수분, 수시간, 24시간 미만, 수일, 2일 내지 6일 등), 또는 보통의 기간(예를 들어, 수일, 3일 내지 12일, 대략 1주, 대략 2주, 대략 1달, 대략 2달, 대략 3달 내지 6달 등) 또는 일부 다른 적합한 양의 시간 동안 시험 화합물에 노출될 수 있다. 시험 유기체가 도 1a(오른쪽 도식) 및 도 1c에 예시된 것과 같은 동물(예를 들어, 설치류)이면, 이후 동물은 희생되고/되거나, 원하는 조직은 DNA 추출을 위해 수확될 수 있다. 예를 들어, 소정의 실시형태에서, 시험 동물은 희생되지 않고, (예를 들어, 시험 물질에 대한 투여 또는 노출 후에 동일한 시점 또는 상이한 시점에) 하나 이상의 혈액 샘플은 DNA 추출을 위해 시험 동물로부터 수집될 수 있다. 동물이 희생되는 실시형태에서, 하나 이상의 관심 조직(예를 들어, 간, 골수, 폐, 비장, 혈액 등)은 DNA 추출을 위해 수확될 수 있다. 시험 유기체가 배양물에서 세포를 포함하면(도 1b), 세포의 전부 또는 일부는 DNA 추출을 위해 수집될 수 있다.

[0138] 수집되거나 수확된 생물학적 샘플로부터 DNA를 추출한 후에, DNA 라이브러리(예를 들어, 시퀀싱 라이브러리)가 제조될 수 있다. 일 실시형태에서, DNA 라이브러리(또는 다른 핵산 시퀀싱 라이브러리)를 제조하기 위한 접근법은 (예를 들어, 도 1a에 예시된 것과 같은) 듀플렉스 시퀀싱 라이브러리 작제 프로토콜과 관련하여 상기에 기재된 것과 유사한 방식으로 (예를 들어, DNA 샘플로부터의) 단편화된 이중-가닥 핵산 물질을 분자 바코드로 표지(예를 들어, 태그화)하는 것으로 시작할 수 있다. 일부 실시형태에서, 이중-가닥 핵산 물질은 (예를 들어, 무세포 DNA, 손상된 DNA 등에 의해) 단편화될 수 있지만, 다른 실시형태에서 다양한 단계는 기계적 전단, 예컨대 음파처리, 또는 다른 DNA 절단 방법(예를 들어, 효소 분해, 분무 등)을 사용한 핵산 물질의 단편화를 포함할 수 있다. 단편화된 이중-가닥 핵산 물질의 표지화의 양태는 말단-복구 및 3'-dA-꼬리화, 특정 분야에 필요하면, 이어서 (예를 들어, 도 1a에 예시된 바와 같은) SMI를 함유하는 적합한 어댑터의 듀플렉스 시퀀싱과 함께 이중-가닥 핵산 단편의 결찰을 포함할 수 있다. 다른 실시형태에서, SMI는 원래의 핵산 분자의 가닥 둘 다로부터의 정보를 고유하게 관련시키기 위한 내인성 서열 또는 외인성 서열과 내인성 서열의 조합일 수 있다.

[0139] 어댑터 분자를 이중-가닥 핵산 물질에 결찰한 후에, 상기 방법은 증폭(예를 들어, PCR 증폭, 회전 환 증폭, 다중 변위 증폭, 등온 증폭, 브리지 증폭, 표면-결합된 증폭 등)을 계속할 수 있다(도 1b). 소정의 실시형태에서, 예를 들어 하나 이상의 어댑터 서열에 특이적인 프라이머는 원래의 이중 가닥 핵산 분자의 각각의 가닥으로부터 유래된 핵산 앰플리콘의 다수의 카피를 생성시키는 핵산 물질의 각각의 가닥을 증폭시키도록 사용될 수 있고, 각각의 앰플리콘은 원래 연관된 SMI를 보유한다(도 1b). 반응 부산물을 제거하기 위한 증폭 및 연관된 단계 후에, 표적 핵산 영역(들)(예를 들어, 관심 영역, 유전좌위 등)은 혼성화-기반 표적화된 포획을 사용하여, 또는 다른 실시형태에서, 어댑터 서열에 특이적인 프라이머(들) 및 관심 표적 핵산 영역(들)에 특이적인 프라이머(들)(비도시)를 사용하여 다중 PCR로 선택적으로 농후화될 수 있다.

[0140] DNA 라이브러리 제조 및 증폭 단계 후에, 이중-가닥 어댑터-DNA 복합체는 표준 시퀀싱 방법을 사용하여 적절한 대량 병렬 DNA 시퀀싱 플랫폼으로 시퀀싱될 수 있다(도 1b). 제1 가닥의 다수의 카피 및 제2 가닥의 다수의 카피의 시퀀싱 후에, 시퀀싱 데이터는 듀플렉스 시퀀싱 접근법을 사용하여 본원에 기재된 바대로 분석될 수 있고, 이로써 원래의 이중 가닥 표적 핵산 분자의 제1 가닥 또는 제2 가닥으로부터 유래된 동일한 외인성 SMI(예를 들어, 어댑터 서열) 및/또는 내인성 SMI를 공유하는 시퀀싱 리드는 별개로 그룹화된다. 일부 실시형태에서, 제1 가닥(예를 들어, "상부 가닥")으로부터의 그룹화된 시퀀싱 리드는 제1 가닥 공통 서열(예를 들어, 단일-가닥 공통 서열(SSCS))을 형성하도록 사용되고, 제2 가닥(예를 들어, "하부 가닥")으로부터의 그룹화된 시퀀싱 리드는 제2 가닥 공통 서열(예를 들어, SSCS)을 형성하도록 사용된다. 도 1c를 다시 참조하면, 제1 SSCS 및 제2 SSCS는

이후 2개의 가닥 사이에 동의한 뉴클레오타이드를 갖는 듀플렉스 공통 서열(DCS)을 생성하도록 비교될 수 있다(예를 들어, 변이체 또는 돌연변이는 가닥 둘 다로부터 유래된 시퀀싱 리드에 보이면 진성이라고 생각된다)(예를 들어, 도 1c 참조). 마찬가지로, 비교 단계에서, 뉴클레오타이드가 2개의 가닥 사이에 동의하지 않은 DCS의 위치는 유전독소 노출에 의해 야기된 손상과 같은 DNA 손상의 잠재적인 부위로서 추가로 평가될 수 있다.

[0141] 다시 도 2a 내지 도 2c를 참조하면, 본 기술내용의 양태에 따라, 듀플렉스 시퀀싱 분석은 게놈에 걸쳐 유도된 돌연변이의 빈도를 정확히 정량화하기 위해 추가로 사용될 수 있다. 예를 들어, 본 기술내용의 양태는 예를 들어 돌연변이 스펙트럼, 트리뉴클레오타이드 돌연변이 서명을 포함하는 파생 서열 데이터에 포착된 유전독성-연관된 정보, 증식 및 신생물성 선택에 대한 소정의 돌연변이의 기능적 결과에 대한 정보, 알려진 유전독소와 관련된 경험적으로 도출된 유전독성-연관된 정보(예를 들어, 돌연변이 스펙트럼, 트리뉴클레오타이드 돌연변이 서명)와의 비교 등을 생성하는 것에 관한 것이다.

[0142] 본 기술내용은 유전독소에 대한 노출의 결과로서 대상체에서 적어도 하나의 게놈 돌연변이를 검출하는 방법을 추가로 포함하고, 상기 방법은 1) 유전독소 노출 후에 대상체로부터의 샘플을 제공하는 단계이되, 샘플은 복수의 이중-가닥 DNA 분자를 포함하는 단계; 2) 비대칭적 어댑터 분자를 개별 이중-가닥 DNA 분자에 결찰시켜 복수의 어댑터-DNA 분자를 생성하는 단계; 3) 각각의 어댑터-DNA 분자에 대해 (i) 어댑터-DNA 분자의 원래의 제1 가닥의 카피의 세트 및 어댑터-DNA 분자의 원래의 제2 가닥의 카피의 세트를 생성하는 단계; (ii) 원래의 제1 가닥의 카피의 세트 및 원래의 제2 가닥의 카피의 세트를 시퀀싱하여 제1 가닥 서열 및 제2 가닥 서열을 제공하는 단계; 및 (iii) 제1 가닥 서열 및 제2 가닥 서열을 비교하여 제1 가닥과 제2 가닥 서열 사이의 하나 이상의 관련성을 확인하는 단계; 및 4) 각각의 어댑터-DNA 분자에서 하나 이상의 관련성을 분석하여 특정 유전독소, 유전독소의 종류, 및/또는 작용 기전을 나타내는 돌연변이체 빈도 및 돌연변이 스펙트럼 중 적어도 하나를 결정하는 단계를 포함한다. 일부 실시형태에서, 돌연변이 스펙트럼은 삼중항 돌연변이 스펙트럼이다. 다른 실시형태에서, 삼중항 돌연변이 스펙트럼을 결정하기 위해 각각의 어댑터-DNA 분자에서 하나 이상의 관련성을 분석하는 단계는 특정 유전독소에 대해 삼중항 돌연변이 서명을 생성하는 것을 추가로 포함한다. 소정의 실시형태에서, 돌연변이체 빈도를 결정하는 단계는 돌연변이된 염기의 삼중항/트리뉴클레오타이드 상황의 빈도를 결정하는 것을 포함한다.

[0143] 일부 실시형태에서, 삼중항 돌연변이 서명 및/또는 돌연변이 스펙트럼은 (예를 들어, 유사성 및/또는 차이에 기초하여) 대상체가 노출된 유전독소의 유형(알려지지 않은 경우), 유전독소의 작용 기전, 대상체가 유전독소 연관된 질병 또는 장애를 발생시킬 가능성, 및/또는 다른 유전독소 연관된 정보를 결정하기 위해 경험적으로 도출된 유전독소 연관된 정보와 비교된다. 예를 들어, 대상체에서 알려지거나 의심된 유전독소(예를 들어, 시험 유전독소) 노출로부터 생긴 듀플렉스 시퀀싱 트리뉴클레오타이드 스펙트럼 패턴은 다른 알려진 유전독소에 대한 노출과 연관된 경험적으로 도출된 트리뉴클레오타이드 스펙트럼 패턴(예를 들어, 데이터베이스에 저장된 것)과 비교될 수 있다. 소정의 실시형태에서, 듀플렉스 시퀀싱 트리뉴클레오타이드 스펙트럼 패턴은 경험적으로 도출된 트리뉴클레오타이드 스펙트럼 패턴 중 하나 이상과 실질적으로 유사할 수 있어서, 실행자는 하나 이상의 경험적으로 도출된 트리뉴클레오타이드 스펙트럼 패턴과의 유사성에 기초하여 시험 유전독소의 정체, 시험 유전독소에 대한 노출의 수준, 시험 유전독소의 작용 기전 등에 관해 정보가 제공될 수 있다.

[0144] *돌연변이체 빈도*

[0145] 일부 실시형태에서, 듀플렉스 시퀀싱 분석 단계는 다양한 노출 조건 하에 특정 유전독소와 연관된 돌연변이체 빈도를 확인할 수 있다. 예를 들어, 유전독소에 대한 생물학적 샘플의 노출과 연관된 돌연변이체 빈도는 다른 인자들 중에서도 비제한적인 예로서 유기체/대상체, 대상체의 연령, 유전독소 유형, 유전독소에 대한 노출의 시간 또는 수준의 양, 조직 유형, 치료 그룹, 게놈의 영역(예를 들어, 게놈 유전좌위)을 포함하는 다양한 인자에 따라, 돌연변이 유형에 의해, 치환 유형에 의해, 그리고 트리뉴클레오타이드 상황에 의해 달라질 수 있다. 일부 예에서, 돌연변이체 빈도는 시퀀싱된 듀플렉스 염기-쌍마다 검출된 고유한 돌연변이의 수로 측정된다. 다른 실시형태에서, 돌연변이체 빈도는 시간에 따른 단일 유전자 또는 유기체에서의 새로운 돌연변이율이다.

[0146] *돌연변이 스펙트럼*

[0147] 다양한 실시형태에서, 듀플렉스 시퀀싱을 사용하여 생성된 고정확성 (예를 들어, 오류-보정된) 서열 리드는 특정 유전독소 또는 잠재적인 유전독소에 대한 돌연변이 스펙트럼 또는 돌연변이 서명을 생성하기 위해 추가로 분석될 수 있다. 일 실시형태에서, 돌연변이 스펙트럼 또는 돌연변이 서명은 유전독소에 대한 노출로부터 생긴 돌연변이성 과정으로부터 생긴 돌연변이 유형의 특징적인 조합을 포함한다. 이러한 특징적인 조합은 돌연변이의 유형(예를 들어, 핵산 서열 또는 구조의 변경)에 관한 정보를 포함할 수 있다. 예를 들어, 돌연변이 스펙트럼은

샘플에서 점 돌연변이(예를 들어, 단일 염기 돌연변이)의 수, 위치 및 상황, 뉴클레오타이드 결실, 서열 재배열, 뉴클레오타이드 삽입 및 DNA 서열의 중복에 관한 패턴 정보를 포함할 수 있다. 일부 실시형태에서, 돌연변이 스펙트럼은 결정된 돌연변이 패턴을 발생시키는 작용 기전을 결정하기 위해 관련된 정보를 포함할 수 있다. 예를 들어, 돌연변이 스펙트럼은 돌연변이성 과정이 외인성 유전독소 노출 또는 내인성 유전독소 노출에 의해 직접적으로 생기는지 또는 다른 것들 중에서 DNA 복제 불충의 동요, 결함성 DNA 복구 경로 및 DNA 효소 편집을 통해 유전독소 노출에 의해 간접적으로 촉발되는지를 결정할 수 있다. 일부 실시형태에서, 돌연변이 스펙트럼은 컴퓨터를 사용한 패턴 매칭(예를 들어, 비지도된 계층적 돌연변이 스펙트럼 클러스터링, 비부정적 행렬 인수분해 등)에 의해 생성될 수 있다.

[0148] *삼중항 돌연변이 스펙트럼/서명*

[0149] 일 실시형태에서, 듀플렉스 시퀀싱을 사용하여 생성된 고정확성 (예를 들어, 오류-보정된) 서열 리드는 삼중항 돌연변이 스펙트럼(본원에서 트리뉴클레오타이드 스펙트럼 또는 서명이라고도 칭함)을 생성하기 위해 추가로 분석될 수 있다. 예를 들어, 유전독소 및/또는 유전독소 노출의 사건과 연관된 돌연변이 스펙트럼은 트리뉴클레오타이드 또는 트리뉴클레오타이드 상황에서 단일 뉴클레오타이드 변이 또는 돌연변이를 검출하기 위해 추가로 분석될 수 있다. 이론에 의해 구속되지 않으면서, 유전독소 노출 또는 다른 과정(예를 들어, 노화)이 트리뉴클레오타이드 상황(예를 들어, 뉴클레오타이드 염기 및 이의 바로 둘러싼 염기)에 따라 핵산에 가변적이고/이거나 특정한 손상을 야기할 수 있다고 인식된다. 일부 실시형태에서, 유전독소는 고유한, 반고유한 및/또는 그렇지 않으면 확인 가능한 삼중항 스펙트럼/서명을 가질 수 있다. 예를 들어, 제1 유전독소의 트리뉴클레오타이드 스펙트럼은 주로 C·G→A·T 돌연변이를 포함할 수 있고, CpG 부위에 대한 더 높은 편애를 추가로 가질 수 있다. 이러한 트리뉴클레오타이드 스펙트럼은 주로 담배에 대한 노출에 의한 유사한 제안된 병인학 드라이브이고, 여기서 벤조[α]피렌 및 다른 다환식 방향족 탄화수소는 알려진 돌연변이원이다. 다른 예에서, 우레탄은 5'-NTG-3' 트리뉴클레오타이드 상황에서 T·A→A·T의 주기적인 패턴에서 DNA 손상을 생성하는 유전독소이다. 따라서, 일부 실시형태에서, 삼중항 돌연변이 스펙트럼의 결정은 다른 이익들 중에서 대상체에서 유전독소 노출를 확인하는 것, 잠재적인 유전독소의 유전독성을 결정하는 것 및 유전독성 물질 또는 인자의 작용 기전을 확인하는 것에 유리할 수 있다.

[0150] *작용 기전*

[0151] 일부 실시형태에서, 듀플렉스 시퀀싱을 사용하여 생성된 고정확성 (예를 들어, 오류-보정된) 서열 리드는 특정 유전독소에 대한 노출 후에 핵산에 검출된 변경을 발생시키는 생화학적 과정(들)을 추론하도록 사용될 수 있다. 예를 들어, 일 실시형태에서, 듀플렉스 시퀀싱 방법을 사용하여 생성된 돌연변이체 빈도 및 돌연변이 스펙트럼(트리뉴클레오타이드 스펙트럼을 포함)은 관찰된 돌연변이 유형과 연관된 패턴 및 생화학적 특성, 및 유전독소 노출에 의해 야기된 유전자 돌연변이 또는 DNA 손상의 게놈 위치에 관한 경험적으로 도출되거나 선험적으로 도출된 정보와 비교될 수 있다. 검출된 게놈 프리돌연변이, 돌연변이 또는 손상을 따르는 생화학적 경로 및/또는 병리생리학적 과정이 확인되는 실시형태에서, 이러한 정보는 일부 실시형태에서 유전독소에 노출된 대상체에 대한 (예를 들어, 치료학적 또는 예방학적) 치료 옵션을 알려주도록 사용될 수 있거나, 다른 실시형태에서, 이러한 정보는 상업화 효과(예를 들어, 새로운 약물), (예를 들어, 환경 독소 또는 제조 부산물의) 세정 효과의 실행가능성을 알려주도록 사용될 수 있거나, 추가의 실시형태에서, 이러한 정보는 시험된 화합물, 물질 또는 인자와 연관된 유전독성을 제거하고/하거나 감소시키기 위해 이 화합물, 물질 또는 인자가 변경될 수 있다는 것을 알려주도록 사용될 수 있다.

[0152] *유전독성을 평가하기 위한 핵산 물질의 소스*

[0153] 상기에 기술된 것처럼, 핵산 물질이 임의의 다양한 소스로부터 나올 수 있음이 고려된다. 예를 들어, 일부 실시형태에서, 핵산 물질은 적어도 하나의 대상체(예를 들어, 인간 또는 동물 대상체)로부터의 샘플 또는 다른 생물학적 소스로부터 제공된다. 일부 실시형태에서, 핵산 물질은 뱅킹된/저장된 샘플로부터 제공된다. 일부 실시형태에서, 샘플은 혈액, 혈청, 땀, 타액, 뇌척수액, 점액, 자궁 세척액, 질 면봉, 코 면봉, 구강 면봉, 조직 부스러기, 모발, 지문, 뇨, 대변, 유리액, 복막 세척액, 가래, 기관지 세척액, 구강 세척액, 흉막 세척액, 위 세척액, 위액, 담즙, 췌관 세척액, 담관 세척액, 총담관 세척액, 쓸개액, 활액, 감염된 상처, 비감염된 상처, 고고학적 샘플, 법의학적 샘플, 물 샘플, 조직 샘플, 식품 샘플, 바이오반응기 샘플, 식물 샘플, 손톱 부스러기, 정액, 전립샘 분비액, 나팔관 세척액, 무세포 핵산, 세포 내의 핵산, 메타게놈 샘플, 이식된 외래 바디의 세척액, 비강 세척액, 장액, 상피 브러싱, 상피 세척액, 조직 생검, 검시 샘플, 부검 샘플, 장기 샘플, 인간 확인 샘플, 인공 제조된 핵산 샘플, 합성 유전자 샘플, 핵산 데이터 저장 샘플, 종양 조직, 및 임의의 이들의 조합 중 적어

도 하나이거나 이를 포함한다. 다른 실시형태에서, 샘플은 미생물, 식물-기반 유기체, 또는 임의의 수집된 환경 샘플(예를 들어, 물, 흙, 고고학 등) 중 적어도 하나이거나 이를 포함한다. 본원에 추가로 기술된 특정 실시예에서, 핵산 물질은 유전독소 또는 잠재적인 유전독소에 노출된 생물학적 소스로부터 나올 수 있다. 일부 예에서, 유전독소는 돌연변이원 및/또는 발암물질이다. 실시예에서, 핵산 물질은 핵산 물질이 유래된 생물학적 소스가 유전독소에 노출되는지를 결정하기 위해 분석된다.

[0154] 듀플렉스 시퀀싱은 다른 알려지거나 관습적인 독성 검정, 예컨대 Ames 시험(예를 들어, 박테리아에서 돌연변이 유발에 대한 시험), 포유류 세포 배양에서의 시험관내 시험, 형질전환 설치류 검정, Pig-a 검정, 및 생체내 2년 생체검정과 비교될 때 다수의 진전을 제공한다. 예를 들어, 많은 종래 기술의 방법은 시험 물질/인자의 유전독성과 관련된 유익한 정보에 대한 대리물로서 리포터 유전자의 정보획득(예를 들어, Ames 시험, 시험관내 포유류 세포 배양, 생체내 형질전환 설치류 검정) 또는 비인간 소스에서의 시험(예를 들어, Ames 시험, 형질전환 설치류 검정, Pig-a 검정, 2년 생체검정)으로 제한되고, 제공된 매우 적은 정보를 완료하는 데 오랜 기간을 요구할 수 있거나(예를 들어, 야생형 설치류에서의 2년 생체검정), 매우 고비용일 수 있다(예를 들어, 형질전환 설치류 검정, 2년 생체검정). 듀플렉스 시퀀싱 검정은 유전독성에 대한 시험 물질/인자를 스크리닝하기 위한 종래 기술의 검정 및 기법의 많은 단점과 대조적으로 널리 배치 가능하고, 경제적이고, 짧은 기간(예를 들어, 2주 미만)에 고정확성 데이터를 제공하기 위해 사용된 시험 물질/인자의 초기 스크리닝 및 후기 스크리닝 둘 다에 적합할 수 있고, 임의의 유기체/생물학적 소스 또는 임의의 조직/장기로부터의 시험관내 및 생체내 시험된 샘플(즉, 무엇보다도 생체내 인간 샘플을 포함) 둘 다를 스크리닝하기 위해 사용될 수 있거나, 다수의 유전자 유전좌위를 평가하고, 유전독성의 리포터로서 자연 게놈을 사용할 수 있고, 결정된 유전독소 물질/인자의 작용 기전을 알려줄 수 있다.

[0155] 시약을 갖는 키트

[0156] 본 기술내용의 양태는 듀플렉스 시퀀싱 방법의 다양한 양태를 수행하기 위한 키트(본원에서 "DS 키트"라고도 칭함)를 추가로 포괄한다. 일부 실시형태에서, 키트는 핵산 추출, 핵산 라이브러리 제조, (예를 들어, PCR을 통한) 증폭 및 시퀀싱을 위해 본원에 개시된 방법 또는 방법 단계의 하나 이상을 수행하기 위한 명령과 함께 다양한 시약을 포함할 수 있다. 일 실시형태에서, 키트는 본 기술내용의 양태에 따라 예를 들어 돌연변이체 빈도, 돌연변이 스펙트럼, 삼중항 돌연변이 스펙트럼, 샘플과 연관된 알려진 유전독소의 돌연변이 스펙트럼과의 비교 등을 결정하기 위해 시퀀싱 데이터(예를 들어, 원시 시퀀싱 데이터, 시퀀싱 리드 등)를 분석하기 위한 컴퓨터 프로그램 제품(예를 들어, 컴퓨터에서 실행하는 코딩된 알고리즘, 하나 이상의 알고리즘을 실행하기 위한 클라우드-기반 서버에 대한 접근 코드 등)을 추가로 포함할 수 있다.

[0157] 일부 실시형태에서, DS 키트는 샘플 제조(예를 들어, DNA 추출, DNA 단편화), 핵산 라이브러리 제조, 증폭 및 시퀀싱의 다양한 양태를 수행하기에 적합한 시약 또는 시약의 조합을 포함할 수 있다. 예를 들어, DS 키트는 하나 이상의 DNA 추출 시약(예를 들어, 완충액, 칼럼 등) 및/또는 조직 추출 시약을 선택적으로 포함할 수 있다. 선택적으로, DS 키트는 예컨대 물리적 수단(예를 들어, 음향 전단 또는 음파처리가 용이하게 하는 관, 분무기 유닛 등) 또는 효소적 수단(예를 들어, 무작위 또는 반무작위 게놈 전단을 위한 효소 및 적절한 반응 효소)에 의해 이중-가닥 DNA를 단편화하기 위한 하나 이상의 시약 또는 도구를 추가로 포함할 수 있다. 예를 들어, 키트는 표적화된 분해를 위한 효소(예를 들어, 제한 엔도뉴클레아제, CRISPR/Cas 엔도뉴클레아제(들) 및 RNA 가이드, 및/또는 다른 엔도뉴클레아제), 이중-가닥 프레그먼타제 칵테일, DNA의 단편이 주로 이중-가닥이 되고/되거나 단일-가닥 DNA를 파괴하기 위한 단일-가닥 DNase 효소(예를 들어, 녹두 뉴클레아제, S1 뉴클레아제) 중 하나 이상을 포함하는 이중-가닥 DNA를 효소적으로 단편화하기 위한 DNA 단편화 시약, 및 적절한 완충액 및 이러한 효소 반응이 용이하게 하는 용액을 포함할 수 있다.

[0158] 일 실시형태에서, DS 키트는 샘플에서 이중-가닥 핵산 분자의 오류-보정된 (예를 들어, 고정확성) 서열을 생성하기 위해 듀플렉스 시퀀싱 공정 단계를 수행하기에 적합한 샘플로부터 핵산 서열 라이브러리를 제조하기 위한 프라이머 및 어댑터를 포함한다. 예를 들어, 키트는 사용자가 이것을 생성하기 위해 단일 분자 식별자(SMI) 서열을 포함하는 어댑터 분자 또는 도구(예를 들어, 단일-가닥 올리고뉴클레오타이드)의 적어도 하나의 풀을 포함할 수 있다. 일부 실시형태에서, 샘플에서의 복수의 핵산 분자가 단독으로 또는 이것이 결찰된 단편의 고유한 특징과 조합되어 어댑터 분자의 부착 후에 실질적으로 고유하게 표지될 수 있도록 어댑터 분자의 풀은 적합한 수의 실질적으로 고유한 SMI 서열을 포함할 것이다. 분자 태그화의 분야에 경험 있는 사람은 "적합한" 수의 SMI 서열을 포함하는 것이 다양한 특정 인자(유입 DNA, DNA 단편화의 유형, 단편의 평균 크기, 게놈 내에 시퀀싱된 서열의 복잡함 대 반복성 등)에 따라 다수의 차수의 규모로 변한다는 것을 인식할 것이다. 선택적으로, 어댑터 분자는 하나 이상의 PCR 프라이머 결합 부위, 하나 이상의 시퀀싱 프라이머 결합 부위, 또는 둘 다를 추가로 포

함한다. 다른 실시형태에서, DS 키트는 SMI 서열 또는 바코드를 포함하는 어댑터 분자를 포함하지 않고, 대신에 종래의 어댑터 분자(예를 들어, Y-형상 시퀀싱 어댑터 등)를 포함하고, 다양한 방법 단계는 분자 서열 리드와 관련하도록 내인성 SMI를 사용할 수 있다. 일부 실시형태에서, 어댑터 분자는 인덱싱 어댑터이고/이거나, 인덱싱 서열을 포함한다.

[0159] 일 실시형태에서, DS 키트는 비상보성 영역 및/또는 일부 다른 가닥 한정 요소(SDE)를 각각 갖는 어댑터 분자의 세트, 또는 사용자가 이것(예를 들어, 단일-가닥 올리고뉴클레오타이드)을 생성하기 위한 도구를 포함한다. 다른 실시형태에서, 키트는 어댑터 분자의 적어도 하나의 세트 또는 이를 생성하기 위한 도구를 포함하고, 여기서 적어도 어댑터 분자의 하위세트는 각각 적어도 하나의 SMI 및 적어도 하나의 SDE를 포함한다. 듀플렉스 시퀀싱 공정 단계를 수행하기에 적합한 샘플로부터 핵산 시퀀싱 라이브러리를 제조하기 위한 프라이머 및 어댑터에 대한 추가 특징은 상기에 기재되고 또한 미국 특허 제9,752,188호, 국제 특허 공보 WO 제2017/100441호, 및 국제 특허 출원 제PCT/US18/59908호(2018년 11월 8일 출원)에 개시되어 있고, 이들은 모두 본원에 그 전문이 참조로 포함된다.

[0160] 추가적으로, 키트는 예를 들어 SYBR™ 그린 또는 SYBR™ 골드(메사추세츠주 왈탐 소재의 Thermo Fisher Scientific로부터 입수 가능) 또는 Qubit 형광기와 사용하기 위한 유사물(예를 들어, 메사추세츠주 왈탐 소재의 Thermo Fisher Scientific로부터 입수 가능), 또는 적합한 형광 분광기에 사용하기 위한 PicoGreen™ 염료(예를 들어, 메사추세츠주 왈탐 소재의 Thermo Fisher Scientific로부터 입수 가능)와 같은 DNA 결합 염료와 같은 DNA 정량화 물질을 추가로 포함할 수 있다. 다른 플랫폼에서 DNA 정량화에 적합한 다른 시약이 또한 고려된다. 추가의 실시형태는 핵산 크기 선택 시약(예를 들어, 고상 가역적 부동화(SPRI: Solid Phase Reversible Immobilization) 자기 비드, 겔, 칼럼), 미끼/프레이(pray) 혼성화를 사용한 표적 DNA 포획을 위한 칼럼, qPCR 시약 (예를 들어, 카피 수 결정을 위한) 및/또는 디지털 액적 PCR 시약 중 하나 이상을 포함하는 키트를 포함한다. 일부 실시형태에서, 키트는 선택적으로 라이브러리 제조 효소(리가제, 중합효소(들), 엔도뉴클레아제(들), 예를 들어 RNA 정보획득을 위한 역전사효소), dNTP, 완충액, 포획 시약(예를 들어, 비드, 표면, 코팅된 관, 칼럼 등), 인덱싱 프라이머, 증폭 프라이머(PCR 프라이머) 및 시퀀싱 프라이머 중 하나 이상을 포함할 수 있다. 일부 실시형태에서, 키트는 오류-유발 DNA 중합효소 및/또는 고충실도 DNA 중합효소와 같은 DNA 손상의 유형을 평가하기 위한 시약을 포함할 수 있다. 추가 첨가제 및 시약은 특정 조건(예를 들어, 높은 GC 농후 게놈/표적)에서 PCR 또는 결찰 반응에 고려된다.

[0161] 일 실시형태에서, 키트는 (질병으로 이어지는 복구 돌연변이에 대해서) 중합효소 연쇄 반응(PCR) 공정을 방해하는 DNA 서열 오류를 복구하는 DNA 오류 보정 효소와 같은 시약을 추가로 포함한다. 비제한적인 예로서, 효소는 우라실-DNA 글라이코실라제(UDG), 포름아미도피리미딘 DNA 글라이코실라제(FPG), 8-옥소구아닌 DNA 글라이코실라제(OGG1), 인간 비퓨린/비피리미딘 엔도뉴클레아제(APE 1), 엔도뉴클레아제 III(Endo III), 엔도뉴클레아제 IV(Endo IV), 엔도뉴클레아제 V(Endo V), 엔도뉴클레아제 VIII(Endo VIII), N-글라이코실라제/AP-리아제 NEIL 1 단백질(hNEIL1), T7 엔도뉴클레아제 I(T7 Endo I), T4 피리미딘 이합체 글라이코실라제(T4 PDG), 인간 단일-가닥-선택적 단일작용성 우라실-DNA 글라이코실라제((hSMUG1), 인간 알킬아데닌 DNA 글라이코실라제(hAAG) 등 중 하나 이상을 포함하고; DNA 손상(예를 들어, 시험관내 DNA 손상)을 보정하도록 사용될 수 있다. 일부 이러한 DNA 복구 효소는 예를 들어 DNA로부터 손상된 염기를 제거하는 글라이코실라제이다. 예를 들어, UDG는 (사이토신의 자발적 가수분해에 의해 생긴) 사이토신 탈아미노화로부터 생긴 우라실을 제거하고, FPG는 8-옥소-구아닌(예를 들어, 반응성 산소 종으로부터 생긴 가장 흔한 DNA 병변)을 제거한다. FPG는 또한 비염기성 부위에서 1개 염기 갭을 생성할 수 있는 리가제 활성을 갖는다. 예를 들어, 중합효소가 주형을 카피하지 못하므로, 이러한 비염기성 부위는 후속하여 PCR에 의해 증폭하지 못할 것이다. 따라서, 이러한 DNA 손상 복구 효소, 및/또는 여기 기재되고 당해 분야에 알려진 다른 것의 사용은 진성 돌연변이를 갖지 않는 손상된 DNA를 효과적으로 제거할 수 있고, 그렇지 않으면 시퀀싱 및 듀플렉스 서열 분석 후에 오류로 검출되지 않을 것이다.

[0162] 키트는 적절한 대조군, 예컨대 DNA 증폭 대조군, 핵산 (주형) 정량화 대조군, 시퀀싱 대조군, 알려진 유전독소/돌연변이원에 노출된 생물학적 소스로부터 유래된 핵산 분자(예를 들어, 시험 동물로부터 추출된 DNA 또는 유전독소에 노출된 배양물에서 성장한 세포) 및/또는 유전독소/돌연변이원에 노출되지 않은 생물학적 소스로부터 유래된 핵산 분자를 추가로 포함할 수 있다. 다른 실시형태에서, 대조군 시약은 의도적으로 손상된 핵산 및/또는 손상되지 않거나 임의의 손상 물질에 노출되지 않은 핵산을 포함할 수 있다. 추가 실시형태에서, 키트는 또한 제어된 유전독성 실험에서 전달되는 하나 이상의 유전독성 물질 및/또는 비유전독성 물질(예를 들어, 화합물)을 포함할 수 있고, 선택적으로 이러한 물질을 대상체, 조직, 세포 등에 전달하기 위한 프로토콜을 포함한다. 따라서, 키트는 시험 물질(예를 들어, 시험 화합물, 잠재적인 유전독성 물질 또는 인자 등)에 대한 프로토콜 진본성

을 결정하는 듀플렉스 시퀀싱 결과(예를 들어, 예상된 돌연변이 스펙트럼/서명)를 생성시키는 대조군을 제공하기 위한 적합한 시약(시험 화합물, 핵산, 대조군 시퀀싱 라이브러리 등)을 포함할 수 있다. 실시형태에서, 키트는 분석이 대상체 샘플에서의 돌연변이, 패턴 및 유형을 검출하여서 대상체가 어떠한 유전독소에 노출됐는지를 나타내기 위해 대상체 샘플, 예컨대 혈액 샘플을 선적하기 위한 용기를 포함한다. 다른 실시형태에서, 키트는 핵산 오염 대조군 표준품(예를 들어, 시험 유기체 또는 대상체 유기체와 다른 유기체에서 게놈 영역과 친화도를 갖는 혼성화 포획 프로브)을 포함할 수 있다.

[0163] 키트는 PCR 및 시퀀싱 완충액, 희석제, 대상체 샘플 추출 도구(예를 들어, 시린지, 면봉 등)를 포함하는 상업적 및 사용자 견지로부터 바람직한 물질, 및 사용 설명서를 갖는 패키지 인서트를 포함하는 하나 이상의 다른 용기를 추가로 포함할 수 있다. 또한, 라벨은 용기에 사용 지도, 예컨대 상기에 기재된 것이 제공될 수 있고/있거나, 지도 및/또는 다른 정보가 또한 키트와 함께 그리고/또는 이것 내에 제공된 웹사이트 주소를 통해 포함되는 인서트에 포함될 수 있다. 키트는 예를 들어 샘플 관, 플레이트 실러(sealer), 마이크로원심분리 관 오프너, 라벨, 자기 입자 분리기, 폼 인서트, 아이스 팩, 드라이아이스 팩, 절연체 등과 같은 실험실 도구를 또한 포함할 수 있다.

[0164] 키트는 전자 컴퓨팅 장치(예를 들어, 랩탑/데스크탑 컴퓨터, 태블릿 등)에 설치 가능하거나 네트워크(예를 들어, 원격 서버)를 통해 접근 가능한 컴퓨터 프로그램 제품을 추가로 포함할 수 있고, 여기서 컴퓨팅 장치 또는 원격 서버는 듀플렉스 시퀀싱 분석 단계를 포함하는 연산을 수행하기 위한 명령을 실행하도록 구성된 하나 이상의 프로세서를 포함한다. 예를 들어, 프로세서는 듀플렉스 시퀀싱 데이터를 생성하기 위해 원시 시퀀싱 리드 또는 비분석된 시퀀싱 리드를 처기하기 위한 명령을 실행하도록 구성될 수 있다. 추가 실시형태에서, 컴퓨터 프로그램 제품은 대상체 또는 샘플 기록(예를 들어, 특정 대상체 또는 샘플 또는 샘플 그룹에 관한 정보 및 알려진 유전독소에 관한 경험적으로 도출된 정보)을 포함하는 데이터베이스를 포함할 수 있다. 컴퓨터 프로그램 제품은, 컴퓨터에서 실행될 때, 본원에 개시된 방법의 단계를 수행하는 비일시적 컴퓨터 판독 가능한 매체에서 구현된다(예를 들어, 도 19 및 도 20 참조).

[0165] 키트는 데이터(예를 들어, 시퀀싱 데이터, 리포트, 다른 데이터)를 업로딩하고 다운로딩하기 위해 원격 서버(들)(클라우드-기반 서버를 포함)에 접근하기 위한 명령 및/또는 접근 코드/패스워드 등 또는 로컬 장치에 설치되는 소프트웨어를 추가로 포함할 수 있다. 모든 컴퓨팅 작업은 원격 서버에 있고, 인터넷 연결 등을 통해 사용자/키트 사용자에 의해 접근될 수 있다.

[0166] 고속 유전독소 스크리닝

[0167] 본 기술내용은 의심되는 물질 또는 인자(예를 들어, 화합물, 화학물질, 약학 물질, 제조 생성물 또는 부산물, 식품 물질, 환경 인자 등)의 유전독성을 평가하기 위한 고속 스크리닝 계획을 추가로 포함한다. 일 실시형태에서, 알려지지 않은 유전독성 효과를 갖는 물질/인자는 시험 물질/인자가 유전독성 효과를 포함하는지를 결정하기 위해 스크리닝될 수 있다. 일부 실시형태에서, 물질/인자는 유전독성 효과를 갖거나 역치 유전독성 효과를 초과하는 물질/인자의 사용을 제거하고자 하는 바람으로 스크리닝될 수 있다. 예를 들어, 유전독성-연관된 질병 또는 장애를 잠재적으로 야기할 수 있는 방식으로 돌연변이성인 물질/인자가 확인될 수 있어서, 물질/인자는 적절하게 제어되거나 제거되거나 폐기되거나 저장되거나 기타 등등이 될 수 있다. 일부 실시형태에서, 발암성인 물질/인자는 본원에 기재된 바대로 고속 스크리닝 계획을 사용하여 확인될 수 있다. 다른 실시형태에서, 알려지지 않은 유전독성 효과를 갖는 물질/인자는 원하는 유전독성 효과 및 특히 표적 생물학적 소스에 대한 원하는 유전독성 효과를 갖는 물질/인자를 확인하기 위한 의도로 스크리닝될 수 있다. 예를 들어, 질병 또는 장애(예를 들어, 암)를 갖는 환자로부터 유래된 생물학적 샘플은 세포(예를 들어, 암 세포)를 동요시키거나 파괴시킬 수 있는 원하는 유전독성 효과에 대한 다수의 물질/인자를 시험하기 위해 고속 스크리닝 계획에 사용될 수 있다. 이러한 스크리닝은 새로운 약물/치료의 발견을 위해 그리고/또는 개인화된 약제에 사용하기 위한 표적화된 치료를 위해 수행될 수 있다.

[0168] 일부 실시형태에서, 고속 스크리닝은 동시에 및/또는 시간-효과적으로 복수의 샘플을 스크리닝하는 것을 지칭한다. 하나의 예에서, 유전독성을 위한 물질 또는 인자의 시험은 대상체(예를 들어, 생물학적 소스)를 시험 물질 또는 인자에 노출(예를 들어, 치료, 투여, 도포 등)시키는 것을 포함한다. 따라서, 고속 스크리닝 계획을 위해, 다수의 생물학적 소스/샘플은 동시에 동일한 시험 물질/인자와 처리되거나, 다른 실시형태에서 다수의 시험 물질/인자와 처리될 수 있다. 특정 예에서, 복수의 생물학적 샘플(예를 들어, 배양물, 조직 샘플, 혈액 또는 다른 체액 샘플에서 성장한 인간 또는 다른 유기체 세포, 형질전환 동물의 세포, 이종이식에서 성장한 인간 세포, 살아 있는 환자 오가노이드, 피더 세포 등)은 실질적으로 동시에 일관된 조건 하에 시험 물질/인자에 노출될 수

있다. 고속 스크리닝은 장기 칩을 통해, 예컨대 하기 장기 및 조직으로부터 추출된 동일한 대상체로부터의 혈액 또는 조직 샘플을 갖는 10-장기 칩을 사용하여 또한 사용될 수 있다: 내분비; 피부; GI-관; 폐; 뇌; 심장; 골수; 간; 신장; 및 췌장. 고속 스크리닝을 위한 장기 칩의 사용 방법은 당해 분야에서 잘 알려져 있다(예를 들어, Chan et al. [5]). 다른 실시형태에서, 유전자 변형된 세포주(예를 들어, 이러한 세포가 돌연변이성 또는 유전독성 손상 효과에 보다 민감하게 하는 결핍성 또는 손상된 DNA 복구 경로를 가짐)는 고속 스크리닝 계획에 도입될 수 있다.

[0169] 일부 실시형태에서, 복수의 생물학적 샘플은 동일하거나 실질적으로 유사할 수 있다(예를 들어, 배양물에서 성장한 동일한 세포주, 동일한 대상체로부터의 조직 샘플 및/또는 동일한 조직 유형 등). 다른 실시형태에서, 복수의 생물학적 샘플의 하나 이상은 상이할 수 있다. 예를 들어, 시험 물질/인자는 동일한 유기체로부터의 상이한 조직/세포 유형, 상이한 유기체 또는 이들의 조합에 대한 유전독성 효과에 시험될 수 있다. 특정 예에서, 의심된 유전독성 물질 또는 인자(예를 들어, 화합물, 약학적 약물 등)는 동일한 대상체의 다양한 장기로부터의 조직 샘플에서 동시에 시험될 수 있다(예를 들어, 10-장기 칩). 일부 실시형태에서, 고속 스크리닝은 다수의 시험 물질/인자를 동시에 시험하는 것을 포괄할 수 있다. 따라서, 고속 스크리닝 계획이 임의의 원하는 정보를 제공하는 방식으로 다수의 샘플을 효율적으로 스크리닝하기 위해 사용될 수 있도록, 각각의 시험된 샘플이 (예를 들어, 세포 유형, 조직 유형, 세포 또는 조직이 추출되는 대상체, 종 등에 의해) 의도적으로 변하거나 변하지 않을 수 있는 상이한 특성을 가질 수 있고/있거나, (예를 들어, 시험 물질/인자, 용량 수준, 노출 시간 등에 의해) 설계에 따라 변할 수 있는 상이한 시험 체계에 처리될 수 있음이 고려된다.

[0170] 생물학적 샘플이 노출되고/되거나, 원하는 노출 체제가 완료되면, 샘플로부터의 세포/조직은 수확될 수 있고, DNA는 각각의 샘플로부터 유래된 DNA에 대한 시험 물질/인자의 유전독성/돌연변이성 영향을 평가하기 위해 듀플렉스 시퀀싱을 사용하는 것의 목적을 위해 추출될 수 있다. 일부 실시형태에서, (배양 배지에서 방출된 것과 같은) 무세포 DNA는 듀플렉스 시퀀싱 분석을 위해 생물학적 샘플로부터 수집될 수 있다. 본 기술내용에 의해 고려되는 추가의 실시형태는 알려지거나 의심된 유전독소의 DNA 손상, 돌연변이성 또는 발암성을 평가하기 위해 듀플렉스 시퀀싱 데이터를 생성하기 위한 DNA 샘플의 고속 프로세싱을 포함한다.

[0171] 본원에 기재된 고속 스크리닝 공정은 예컨대 생물학적 샘플의 실험 처리, DNA 추출, 라이브러리 제조 단계, 증폭 단계(예를 들어, PCR) 및/또는 DNA 시퀀싱 단계(예를 들어, 다양한 기법 및 대량 병렬 시퀀싱을 위한 장치를 사용) 중 하나 이상을 수행하기 위한 로봇공학의 사용을 통한 자동화를 포함할 수 있다. 유전독성-연관된 돌연변이 및/또는 DNA 손상에 대해 많은 수의 샘플이 빨리 스크리닝되도록 고속 스크리닝의 사용은 복수의 샘플(즉, 동일한 대상체로부터의 상이한 세포 유형 또는 상이한 대상체로부터의 동일한 세포 유형)이 병렬로 시험되게 한다.

[0172] 일 실시형태에서, 웰의 어레이로 각각 이루어지고 각각의 웰이 하나의 샘플을 포함하는 마이크로플레이트는 로봇 취급에 의해 이 시스템을 통해 이동한다. 일례에서, 마이크로플레이트에서의 웰은 자동화 액체 취급 시스템을 통해 충전될 수 있고, 센서는 예를 들어 대개 항온처리의 기간 후 마이크로플레이트에서 샘플을 평가하도록 사용될 수 있다. 실험실 자동화 소프트웨어는 스크리닝 공정의 전체 또는 일부를 제어하여서 공정 내 정확성 및 공정 내 반복성을 보장하도록 사용될 수 있다.

[0173] 환경적/외인성 유전독소

[0174] 본 기술내용의 양태는 예컨대 임의의 상기에 기재된 생체내 또는 시험관내 듀플렉스 시퀀싱 스크리닝 방법을 사용하여 환경적/외인성 물질/인자의 유전독성을 평가하는 것을 포함한다. 본 기술내용의 추가 양태는 대상체/유기체가 환경 지역에서 유전독소에 노출되는지를 평가하는 것을 포함한다. 예를 들어, 생물학적 샘플(예를 들어, 조직, 혈액)은 살아 있는 유기체로부터 수집되거나, 그렇지 않으면 예를 들어 지역이 오염되는지를 결정하기 위해 의심되는 오염 지역에 노출될 수 있다. 다른 실시형태에서, 생물학적 샘플은 더 넓은 지역에 존재하는 유기체로부터 수집되고, 유전독소 오염의 소스(예를 들어, 물 시스템으로 누수된/방출된 산업 부산물)의 특정한 지리학적 위치를 핀포인트하기 위한 스크리닝 공정으로 평가될 수 있다. 본원에 기재된 바와 같은 다양한 방법은 가능한 유전독소의 존재에 대해 조사 중인 환경 지역에 노출된 (예를 들어, 대상체로부터의) 생물학적 샘플을 분석하기 위해 사용될 수 있다. 다른 실시형태에서, 본원에 기재된 바와 같은 다양한 방법은 환경 지역(예를 들어, 지리학적 지역, 주거 지역, 작업상 환경 등)에서 알려진 유전독소에 노출되는 것으로 의심된 대상체로부터 취한 생물학적 샘플(들)을 분석하기 위해 사용될 수 있다. 본 기술내용의 양태에 따르면, 생물학적 샘플은 다수의 유기체(예를 들어, 해양물, 포유류, 필터 피더, 보초 유기체 등) 또는 특정 종(예를 들어, 인간 샘플)으로부터 기원할 수 있다.

[0175] 검출 가능한 환경적 유전독소는 비제한적인 예로서 감마-조사, X선; UV-조사; 마이크로파; 전자 방출; 유독 가스; 유독 공기 미립자(예를 들어, 들이마시는 석면); 및 화학적 화합물 및/또는 병원균 오염된 호수, 강, 시내, 지하수 등과 같은 돌연변이성 물질 중 하나 이상에 대한 노출을 추가로 포함한다. 외인성 유전독소의 추가 소스는 예를 들어 식품 물질, 화장품, 가정용 물품, 건강-관리 관련 제품, 요리 제품 및 도구, 및 다른 제조된 소비재를 포함할 수 있다.

[0176] 듀플렉스 시퀀싱 결과는 질병을 야기하는 오염물질의 존재를 확인하는 다른 방법, 예컨대 암 클러스터의 위치를 처음에 확인하는 유행병학적 연구와 함께 추가로 사용될 수 있다. 일부 실시형태에서, 본원에 개시된 방법은 클러스터의 구성원에 영향을 미치는 특정 유전독소를 확인하기 위해 사용될 수 있다. 이 데이터로부터, 유전독소의 소스가 결정될 수 있다. 대상체의 질병 또는 의학 질환을 원인 사건(예를 들어, 환경 또는 다른 외인성 돌연변이원 또는 발암물질에 대한 노출)과 연결하는 상관관계 정보를 전통적으로 사용한 종래의 조사 수단과 반대로, 듀플렉스 시퀀싱은 고정확성, 재현 가능 데이터, 예컨대 돌연변이 스펙트럼 및 작용 기전을 제공하고, 이의 결과는 원인 사건(들)(예를 들어, 특정 돌연변이원 또는 발암물질에 대한 노출)을 경험적으로 결정하기 위해 사용될 수 있다.

[0177] 내인성 유전독소

[0178] 본 기술내용의 양태는 예컨대 임의의 상기에 기재된 생체내 또는 시험관내 듀플렉스 시퀀싱 스크리닝 방법을 사용함으로써 내인성 물질/인자(예를 들어, 내인성 유전독소 또는 유전독성 과정)의 유전독성을 평가하는 것을 포함한다. 따라서, 본 기술내용의 양태는 대상체/유기체가 DNA 손상을 야기하는 내인성 유전독소 또는 유전독성 과정을 경험하였는지를 평가하는 것을 포함한다. 예를 들어, 생물학적 샘플(예를 들어, 조직, 혈액)은 예를 들어 대상체가 유전독소 연관된 질병 또는 장애를 갖거나 이러한 질병 또는 장애를 발생시킬 위험에 있는지를 결정하기 위해 대상체(예를 들어, 환자)로부터 수집될 수 있다.

[0179] 내인성 인자는 비제한적인 예로서 뉴클레오타이드의 오편입을 야기하는 생물학적 사건, 예컨대 DNA 중합효소 오류, 자유 라디칼 및 탈푸린화를 포함할 수 잇다. 내인성 인자는 예를 들어 스트레스, 염증, 내인성 바이러스의 활성화, 자가면역 질병; 환경 노출; 식품 선택(예를 들어, 발암성 식품 및 음료); 흡연; 자연 유전자 구성; 노화; 신경퇴행; 및 기타 등등과 같은 질병 또는 장애 연관된 폴리뉴클레오타이드 돌연변이에 직접적으로 기여하는 단기간 또는 장기간의 생물학적 질환의 발생을 추가로 포함할 수 있다. 예를 들어, 대상체가 높은 수준의 스트레스에 장기간 노출되는 경우, 대상체는 스트레스 연관된 암(예를 들어, 백혈병, 유방암 등)과 상관된 임의의 돌연변이에 대해 듀플렉스 시퀀싱을 통해 시험될 수 있다.

[0180] 내인성 인자는 또한 개체의 노출의 완전한 효과를 반영하는 개별 인간의 조직에서의 돌연변이 및 다른 유전독성 사건의 집합체 축적을 나타낼 수 있고, 정확히 정량화되거나 실험적으로 제어될 수 없다.

[0181] 안전한 돌연변이체 빈도 수준을 결정하는 방법

[0182] 유전독소에 대한 노출로부터 생긴 DNA 손상의 수준 또는 양은 대상체의 다양한 특징(예를 들어, 건강, 연령, 성별, 유전자 구성, 이전의 유전독소 노출 사건 등의 수준) 이외에 예를 들어 (직접적으로 또는 간접적으로) DNA 손상을 야기하는 데 있어서의 유전독소의 효과, 노출의 용량 또는 양, 노출의 경로 또는 방식(예를 들어, 섭취, 흡입, 경피 흡수, 정맥내 등), 노출 기간(예를 들어, 경시적), 대상체가 노출되는 다른 물질 또는 인자의 상승적 효과 또는 길항적 효과를 포함하는 다양한 인자에 따라 변할 수 있다. 상기에 기술된 바대로, 유전독소에 대한 노출은 알려진 질병-연관된 돌연변이 패턴(예를 들어, 유방암에 대한 구별되는 게놈 돌연변이)과 충분히 유사한 돌연변이 패턴(예를 들어, 트리뉴클레오타이드 상황에서의 돌연변이 유형, 돌연변이체 빈도, 확인 가능한 돌연변이)을 포함할 수 있는 것과 연관된 고유한, 반고유한 및/또는 그렇지 않으면 확인 가능한 돌연변이성 스펙트럼 또는 서명을 결정하기 위해 예를 들어 본원에 기재된 바와 같은 듀플렉스 시퀀싱 방법에 의해 평가될 수 있는 폴리핵산 손상을 발생시킬 수 있다. 본 기술내용의 다양한 양태는 유전독소에 대한 안전한 역치 돌연변이체 빈도를 검출하는 방법을 추가로 포함하는 안전하다고 생각될 수 있는 돌연변이체 빈도 수준를 결정하고/하거나 정량화하는 방법에 관한 것이다. 샘플 내의 돌연변이체 빈도가 안전한 수준보다 높을 때, 이것은 대상체가 시간이 지나면서 질병을 발생시킬 위험이 유의미하게 증가한다는 것을 나타낸다.

[0183] 본 기술내용은 (1) 돌연변이원에 노출된 대상체로부터 추출된 하나 이상의 표적 이중-가닥 DNA 분자를 듀플렉스 시퀀싱하는 단계; (2) 표적화된 이중-가닥 DNA 분자에 대한 오류-보정된 공통 서열을 생성하는 단계; 및 (3) 표적화된 이중-가닥 DNA 분자에 대한 돌연변이 스펙트럼을 확인하는 단계; (4) 시퀀싱된 듀플렉스 염기-쌍마다 고유한 돌연변이의 수를 계산함으로써 표적 이중-가닥 DNA 분자에 대한 돌연변이체 빈도를 계산하는 단계를 포함

하는, 돌연변이원에 대한 대상체의 노출 후에 대상체에서 생체내 발생한 게놈 돌연변이를 검출하고 정량화하는 방법을 추가로 포함한다. 단계 (3)의 실시형태에서, 돌연변이 스펙트럼은 "트루뉴클레오타이드 서명"을 포함하는 샘플의 고유한 프로파일이다.

[0184] 일 실시형태에서, 단계 (1) 및 단계 (2)는 a) 이중-가닥 표적 핵산 분자를 적어도 하나의 어댑터 분자에 결찰하여, 어댑터-표적 핵산 복합체를 형성하는 단계이되, 적어도 하나의 어댑터 분자가 i. 단독으로 또는 표적 핵산 전단점과 조합되어 이중 가닥 표적 핵산 분자를 고유하게 표지하는 축퇴성 또는 반축퇴성 단일 분자 식별자(SMI) 서열; 및 ii. 어댑터-표적 핵산 복합체의 각각의 가닥이 이의 상보성 가닥에 대해 명확히 확인 가능한 뉴클레오타이드 서열을 갖도록 어댑터-표적 핵산 복합체의 각각의 가닥을 태그화하는 뉴클레오타이드 서열을 포함하는 단계, b) 어댑터-표적 핵산 복합체의 각각의 가닥을 증폭시켜 복수의 제1 가닥 어댑터-표적 핵산 복합체 앰플리콘 및 복수의 제2 가닥 어댑터-표적 핵산 복합체 앰플리콘을 제조하는 단계; c) 어댑터-표적 핵산 복합체 앰플리콘을 시퀀싱하여 복수의 제1 가닥 서열 리드 및 복수의 제2 가닥 서열 리드를 제조하는 단계; 및 d) 복수의 제1 가닥 서열 리드로부터의 적어도 하나의 서열 리드를 복수의 제2 가닥 서열 리드로부터의 적어도 하나의 서열 리드와 비교하고, 동의하지 않는 뉴클레오타이드 위치를 무시함으로써 이중 가닥 표적 핵산 분자의 오류 보정된 서열 리드를 생성하는 단계에 의해 달성된다(미국 특허 9,752,188 B2호 및 WO 제2017/100441호 참조).

[0185] 유전독소 양의 안전한 역치 수준을 결정하는 방법

[0186] 본 기술내용은 추가로 특정 유전독소에 대한 대상체의 노출의 안전한 수준(중량 또는 부피 또는 질량 기준의 농도 양 또는 단위*시간 적분 등); 및/또는 화합물 또는 다른 물질(예를 들어, 유선 장치로부터의 라디오파 등)이 임의의 수준의 노출에서 유전독성인지 또는 아닌지를 결정하기 위한 실험적 시험관내 및 생체내 방법을 포함한다. 이 결정은 안전한 역치 돌연변이체 빈도 수준을 처음에 결정함에 의존할 수 있다. 일 실시형태에서, 대조군 대상체의 샘플은 유전독소(또는 이의 결여)에 대해 시험되고, 노출된 대상체의 샘플(예를 들어, 복수의 마우스; 또는 하나의 세트가 대조군 세포인 동일한 대상체로부터의 복수의 세포; 등)의 유전독소 프로파일과 비교된다. 노출된 대상체는 질병 발생에 직접적으로 기여하는 검출된 유전독소 유도된 돌연변이가 발생하기 전에 안전한 노출의 역치 수준을 결정하기 위해 의심된 유전독소의 지정된 미리 결정된 노출 양을 받는다.

[0187] 다른 실시형태에서, 시험 대상체(예를 들어, 실험실 동물, 시험관내 세포 등)는 상이한 시간 기간 동안 상이한 용량에 노출되고, 이로부터 유전독소 노출의 안전한 컷아웃 수준이 결정된다: 1) 어떤 노출 용량에서 폴리뉴클레오타이드 돌연변이가 보이지 않는지: 그리고/또는 2) 어떤 노출 용량에서 검출된 폴리뉴클레오타이드 돌연변이가 있는지, 그러나 어디서 용량 동등 수준이 대상체에서 암을 야기하지 않는지, 그리고 다른 화합물의 동일성을 추론하기 위해 발견된 돌연변이의 수준을 사용하는 것; 그리고/또는 3) 선형 저용량 반응 곡선을 추론하기 위해 유도된 돌연변이의 유전독소 용량 반응 곡선 및 회귀 분석을 결정하는 것; 그리고/또는 4) 대상체 집단에서의 주어진 건강 결과에 대한 어떤 위험 비율이 검출된 유전독소 빈도/검출된 서명과 연관되는지.

[0188] 안전한 노출의 역치 수준은 인간, 개/고양이, 말 등과 같은 종에 의해 추가로 결정될 수 있다. 안전한 역치 수준은 유전독소에 대한 노출 경로에 의해 추가로 결정될 수 있다. 예를 들어, 다양한 양의 유전독소를 사용한 실험은 특정 질병 발생과 연관된 돌연변이 및 삼중항 스펙트럼을 발생시키는 경구, 국소 또는 에어로졸 소모에 의해 양(중량, 부피 등) 및/또는 빈도를 결정하기 위해 본원에 개시된 듀플렉스 시퀀싱 방법에 의해 결정될 수 있다.

[0189] 그리고/또는 본원에 개시된 듀플렉스 시퀀싱 실험적 방법은 시간 및/또는 온도에 기초한 유전독성 노출의 역치 양을 결정하기 위해 사용될 수 있다. 예를 들어, 노출 기간 및 물 온도에 기초한 유전독소를 함유하는 물 중에 샤워 또는 욕조로부터의 피부를 통한 흡수, 및 물 중의 유전독소 농도는 피부를 통해 흡수된 유전독소의 양(용량)을 산출하기 위해 사용될 수 있다.

[0190] 유전독소 안전한 역치 수준을 확인하기 위한 오류-보정된 듀플렉스 시퀀싱 결과는 확립된 표준을 단언하거나 조정하기 위해 다른 안전성 역치 데이터(예를 들어, 기존의 FDA 및 EPA 수준, 독성 물질 질병 등록청(Agency for Toxic Substance Disease Registry levels), 미국 국립 독성물질 관리 프로그램 가이드라인(US National Toxicology Program guideline), OECD 가이드라인, 캐나다 건강 가이드라인, 유럽 규제 가이드라인, ILSI/HESI 가이드라인 등)와 추가로 조합될 수 있다.

[0191] 검출 및 치료의 방법

[0192] 질병 또는 장애 발생은 유전독소 노출 후 수년(예를 들어, 20년)까지 전통적인 시험 및 영상화 기법을 통해 진단될 수 없고; 본 기술내용은 대상체를 예방학적으로 치료하기 위해, 또는 (더 높은 위험 수준에 있음으로

인해) 질병에 대해 대상체를 활발히 스크리닝하기 위해, 그리고 유전독소의 존재를 확인하고, 미래의 노출의 방지를 위해 이를 제거하기 위해 유전독소 노출 후에 수일 또는 수주 또는 수달 내에 질병을 야기하는 돌연변이를 검출하는 방법, 또는 질병을 야기하는 돌연변이 또는 전구체를 돌연변이로 야기할 가능성을 갖는 유전독성 과정의 표시를 제공한다.

[0193] 대상체가 유전독소의 역치 안전한 수준 초과에 노출될 때 그리고/또는 대상체가 잠재적으로 불안전한 수준의 유전독소에 노출되는지가 결정될 때(예를 들어, 위험한 노출 수준을 확인하는 건강 부서), 대상체는 유전독성 연관된 질병 또는 장애의 발생에 대한 위험이 유의미하게 증가된다. 대상체는 이후 유전독소를 차단하고/하거나 이에 대응하는 물질로 예방학적으로 치료되고/되거나, 유전독소 노출은 감소되거나 제거된다(예를 들어, 환경으로부터 유전독소를 제거함 또는 대상체를 이동시킴). 추가적으로, 또는 대안적으로, 대상체는 대상체가 질병 또는 장애의 초기 단계를 발생시키는지를 검출하기 위해 순차적으로 적기의 진단학적 시험(예를 들어, 암 검출을 위한 혈액 시험) 및/또는 영상화(예를 들어, CAT, MRI, PET, 초음파, 혈청 바이오마커 시험 등)를 겪고, 이 단계 동안 이는 거의 효과적으로 치료된다. 비제한적인 예로서, 아플라톡신 또는 아리스토로크산 노출에 대해, 대상체는 다른 간발암물질인 만성 C형 간염을 갖는 환자가 간세포 암종에 대해 스크리닝되는 전형적인 스케줄인 6개월마다의 간 초음파를 겪도록 지도될 것이다. 당해 분야에서 잘 알려진 전통적인 진단학적 시험이 질병(예를 들어, 암)을 검출하는 때에, 이후 치료(예를 들어, 수술, 화학요법, 면역치료 등)가 개시된다.

[0194] 예방학적 치료(즉, 발생 위험을 방지하거나 감소시키는)를 제공하고/하거나, 암의 성장을 억제하고/하거나, 암을 근절하는 방법은 숙련된 임상의에게 잘 알려진 치료 프로토콜을 포함하고, 유전독소 유형에 맞춰질 것이다. 이미 유도된 돌연변이를 역전시키기 위한 치료가 현재 존재하지 않지만, 대상체가 소정의 잔류 유전독소(예를 들어, 킬레이트화를 통한 특정 중금속)를 제거하는 것을 돕는 치료학적 방법은 추가의 유전독성을 감소시킬 수 있다.

[0195] 돌연변이원 유도된 종양(예를 들어, 흡연자에서의 폐암, UV 고 노출자에서의 흑색종, 담배 사용자에서의 구강암 등)에 대해, 이 종양에서의 돌연변이의 부담은 더 높은 경향이 있는데, 이는 신생항원의 더 높은 풍부도로 이어지는 것으로 생각되고, 면역치료제에 양호하게 반응하는 훨씬 더 높은 경향을 설명한다. 면역치료제, 예컨대 관문 억제제(즉, PD1 및 PDL1 억제제, 예컨대 니볼루맙, 펨브롤리주맙 및 아테졸리주맙, CTLA4 억제제, 예컨대 이필리주맙)를 포함하는 것의 예방학적 투여가 대상체의 면역계가 조기 형성하는 종양을 박멸하게 할 것이다. 그러므로, 노출 서명의 확인의 다른 치료 지도된 사용은 공식 임상 실험의 환경에서 조심스런 시험을 필요로 하더라도 면역치료 및 잠재적으로 심지어 예방학적 치료에 의한 질병 예방에 대한 미래의 종양 반응성의 예측이다.

[0196] 검출 및 치료의 방법은 적절한 치료 과정을 결정하는 데 사용될 수 있는 유전독소의 작용 기전을 직접적으로 또는 추론적으로 결정하고/하거나 약물 내성 변이체를 모니터링하는 방법을 추가로 포함한다(Schmitt et al [6] 참조).

[0197] 적어도 하나의 유전독소에 노출된 대상체가 진단되거나 검출되면, 대상체는 유전독소 연관된 질병 또는 장애의 발생을 방지하고/하거나, 발생을 지연하고/하거나, 이의 효과를 감소시키고/시키거나 이를 박멸하기 위한 치료학적 유효량의 약학적 조성물이 투여될 수 있다. 약학적 조성물은 유전독소 연관된 질병 또는 장애의 억제제 또는 박멸제, 및 약학적으로 허용 가능한 담체 또는 염을 포함하는 치료학적 유효량의 조성물을 포함한다. 그리고 치료학적 유효량은 의도된 약물학적, 치료학적 또는 예방학적 결과를 생성하기에 효과적인 유전독소 연관된 질병 또는 장애의 억제제 또는 박멸제를 포함하는 조성물의 치료학적, 비독성, 용량 범위를 포함한다.

[0198] 약학적 조성물은 제형화되고, 경구, 정맥내, 근육내, 피하, 요도내, 직장, 척수내, 국소, 협측 또는 비경구 투여를 포함하는 투여 경로에 의해 투여된다. 약학적 조성물은 종래의 약학적 담체 및 부형제와 혼합되고, 정제, 캡슐, 환제, 액제, 정맥내 용액, 음료 및 식품 제품 등의 형태로 사용될 수 있고; 활성 성분의 중량 또는 부피 기준으로 약 0.1% 내지 약 99.9%, 또는 약 1% 내지 약 98%, 또는 약 5% 내지 약 95%, 또는 약 10% 내지 약 80%, 또는 약 15% 내지 약 60%, 또는 약 20% 내지 약 55%를 함유할 것이다.

[0199] 경구 투여를 위해, 정제, 환제 및 캡슐은 추가적으로 종래의 담체, 예컨대 결합제, 예를 들어 아카시아 검, 젤라틴, 폴리비닐피롤리돈, 소르비톨 또는 트라가칸쓰; 충전제, 예를 들어 인산칼슘, 글리신, 락토스, 옥수수 전분, 소르비톨 또는 수크로스; 활택제, 예를 들어 스테아르산마그네슘, 폴리에틸렌 글리콜, 실리카 또는 탈크: 붕괴제, 예를 들어 감자 전분, 향료 또는 착색제 또는 허용 가능한 습윤제를 포함할 수 있다. 경구 액체 제제는 수성 또는 유성 용액, 현탁액, 에멀션, 시럽 또는 엘릭시르로 제형화될 수 있고, 종래의 첨가제, 예컨대 현탁제, 유화제, 비수성 물질, 보존제, 착색제 및 항료를 함유할 수 있다.

[0200] 정맥내 투여 경로를 위해, 약학적 조성물은 임의의 흔히 사용되는 정맥내 유체에 용해되거나 현탁되고, 점적주사에 의해 투여될 수 있다. 정맥내 유체는 제한 없이 생리학적 식염수 또는 링거액을 포함한다.

[0201] 비경구 투여를 위한 약학적 조성물은 수성 또는 비수성 등장성 무균 주사 용액 또는 현탁액의 형태일 수 있다. 이 용액 또는 현탁액은 경구 투여를 위한 제형에 사용하기 위해 언급된 하나 이상의 담체를 갖는 무균 분말 또는 과립으로부터 제조될 수 있다. 화합물은 폴리에틸렌 글리콜, 프로필렌 글리콜, 에탄올, 옥수수 오일, 벤질 알코올, 염화나트륨 및/또는 다양한 완충액에 용해될 수 있다.

[0202] 치료학적 효과 용량은 유전독성 노출의 양 또는 기간; 대상체의 연령, 체중, 성별 또는 인종; 질병 또는 장애의 발생 단계; 및 숙련된 임상의에게 잘 알려진 다른 방법과 같은 다양한 인자에 기초하여 추가로 산출될 수 있다. 일 실시형태에서, 대상체는 노출이 수년전 발생하더라도 유전독소에 대한 이의 잠재적인 노출 또는 의심된 노출의 발견에 따라 시험된다. 대상체는 안전한 역치 수준보다 높게 노출된다고 진단되면 증상이 나타날 시 또는 증상이 나타남에 바로 약학적 화합물이 투여된다. 모든 실시형태에서, 유전독소는 가능할 때 대상체의 환경으로부터 제거한다.

[0203] 실험적 실시예

[0204] 하기 부문은 듀플렉스 시퀀싱 및 연관된 시약을 사용하여 게놈 생체내 돌연변이유발을 검출하고 평가하기 위한 방법의 예를 제공한다. 하기 실시예는 본 기술내용을 예시하기 위해 그리고 이를 만들고 사용하는 당업자를 보조하도록 제시된다. 실시예는 본 기술내용의 범위를 달리 제한하기 위한 임의의 방식으로 의도되지 않는다.

[0205] 일반적으로, 생체내 돌연변이유발을 측정하기 위한 DS의 효율을 벤치마킹하기 위해, 62개의 샘플에 걸쳐 82억개의 오류-보정된 염기를 생성시킨 일련의 마우스 실험은 2가지의 독립적 동물 균주에서 5개의 건강한 조직으로부터 9개의 유전자에서 3개의 돌연변이원의 효과를 조사하도록 수행되었다. 듀플렉스 시퀀싱은 특정 돌연변이원, 조직 유형 및 게놈 유전좌위에 의해 변하고, 황금-표준 형질전환 설치류 검정의 것을 면밀히 반영하는 정도로 치료된 동물 중에서 증가된 돌연변이체 빈도를 정량적으로 입증하였다. 다양한 실시예에서, 객관적 돌연변이 패턴 단독에 기초한 치료 그룹에 의해 샘플을 확인할 수 있다. 일부 실시예에서, 돌연변이원 민감도는 상이한 유전자 유전좌위 중에서 4배까지 변하고, 이론에 의해 구속되지 않으면서, 스펙트럼 패턴은 이것이 부분적으로 전사 및 메틸화를 포함할 수 있는 구역상 구별되는 공정의 결과라는 것을 제안한다. 다양한 실시예에서, 담배-관련된 발암물질 벤조[a]피렌으로 치료된 동물에서 초저빈도로 DS에 의해 확인된 SNV 중에서 트리뉴클레오타이드 돌연변이 서명은 공공에게 이용 가능한 데이터베이스에서 흡연-연관된 폐암의 게놈에서 클론성 SNV 중에 보인 것과 거의 동일한 것으로 나타났다. 일부 실시예에서, DS는 돌연변이원 치료 후에 단지 4주에 선택적 압력 하에 클론성으로 확장하는 저빈도 종양발생 유발자 돌연변이를 확인하기 위해 사용되었다. 따라서, 본원에 기재된 다양한 실시예에서 입증될 것처럼, DS는 돌연변이 생물학, 독성학 및 암 위험 평가에서 다양한 적용으로 유전독성 과정 및 실시간 신생물성 진화 둘 다를 직접적으로 정량화하기 위해 사용될 수 있다.

[0206] *실시예 1*

[0207] BigBlue® 마우스에서의 *cII* 전이유전자 및 내인성 유전자에서의 생체내 돌연변이 분석을 위한 듀플렉스 시퀀싱의 적용. 이 부문은 오류-보정된 차세대 시퀀싱(NGS)이 BigBlue® 형질전환 설치류(TGR) 돌연변이 검정에 사용된 *cII* 전이유전자 및 자연적 마우스 유전자 둘 다에서 화학적으로-유도된 돌연변이를 직접적으로 측정하기 위해 사용되는 실시예를 기재한다. 현재, TGR 돌연변이 검정은 플라크 형성을 통해 희귀 *cII* 돌연변이체를 검출한다. 표준 NGS는 이의 높은 오류율(시퀀싱된 10^3개의 염기당 약 1개의 오류)로 인해 저빈도 돌연변이 검출에 사용 불가능하다. 오류-보정된 NGS 또는 듀플렉스 시퀀싱은 극적으로 더 낮은 오류율(약 $1/10^8$개의 염기)을 가져서, 초희귀 돌연변이의 검출을 허용한다.

[0208] 이 실시예에서, 듀플렉스 시퀀싱의 적용은 대조군, N-에틸-N-니트로소우레아(ENU) 및 벤조[a]피렌(B[a]P)-노출된 BigBlue® C57BL6 수컷 마우스에서 돌연변이체 빈도(MF) 및 스펙트럼을 평가하도록 사용되었다.

[0209] BigBlue® 형질전환 C57BL/6 수컷 마우스는 1일 내지 28일에 비히클(올리브 오일) 또는 B[a]P(50 mg/kg/일), 또는 1일 내지 3일에 ENU(pH 6 완충액 중의 40 mg/kg/일)(n=6)로 매일 경구 위관영양에 의해 치료되었다. 조직을 수집하고, 31일에 연구에서 동결시켰다. 돌연변이체에 대해 간 및 골수를 분석하였다. DNA를 단리시키고, 돌연변이체를 Agilent Technologies에 의해 기재된 RecoverEase 및 Transpack 방법을 사용하여 *cII* 돌연변이체 플라크에 대해 분석하였다. 듀플렉스 시퀀싱은 간 및 골수에서 돌연변이에 대해 *cII* 및 다른 내인성 유전자를 시

퀀싱하기 위해 사용되었다.

[0210] 평가된 유전자 및 유전자를 선택하도록 사용된 기준은 하기와 같다: (1) 모든 조직 유형에서 편재하여 전사된 *Polr1c*(RNA 중합효소); (2) 망막 이외의 임의의 조직에서 발현되지 않는 *Rho*(로돕신); (3) 거의 다른 어디서도 발현되지 않지만 간에서 고도로 발현되는 *Hp*(합토글로빈); (4) 인간 간세포 암종에서 가장 흔히 돌연변이된 유전자인 *Ctnnb1*(베타-카테닌); 및 (5) BigBlue® 마우스에서 약 80개의 카피에 존재하는 360 bp 형질전환 리포터 유전자인 *cII*.

[0211] 도 **3a** 내지 도 **3d**는 상기에 기재된 바대로 돌연변이원 치료 후에 간 및 골수에서 듀플렉스 시퀀싱(도 3a 및 도 3b) 및 BigBlue® *cII* 플라크 검정(도 3c 및 도 3d)에 계산된 돌연변이체 빈도를 보여주는 상자 그림 그래프이다. 듀플렉스 시퀀싱에 대한 MF는 시퀀싱된 듀플렉스 염기-쌍마다 전체 돌연변이체에 기초한다(n=5 마우스/그룹). BigBlue®에 대한 MF는 돌연변이체 플라크 형성 단위의 수에 대한 돌연변이체 플라크의 수로 계산되었다(n=6 마우스/그룹). 도시된 것처럼, 듀플렉스 시퀀싱 및 전통적인 BigBlue® *cII* 플라크 검정에 의해 측정된 MF는 돌연변이원 둘 다에 유사한 반응을 주었다. 더 빠르게 분열하는 세포인 골수는 방법 둘 다를 사용하여 간보다 더 높은 MF를 나타냈다.

[0212] 도 **3e**는 듀플렉스 시퀀싱에 대한 형질전환 설치류 검정에서 상대 *cII* 돌연변이체 배수 증가를 예시한다. 상기한 바대로, 플라크 검정에서의 MF는 허용적 플레이트에서 형성된 플라크의 총 수로 나눈 선택 플레이트에서 관찰된 표현형으로 활성인 돌연변이체 플라크의 수로 계산된다. 듀플렉스 시퀀싱 검정에서의 MF는 297 BP *cII* 전이유전자 간격 내에 시퀀싱된 염기 쌍의 총 수로 나눈 돌연변이체 염기 쌍 관찰의 수로 계산된다. 도함수 측정의 차이에도 불구하고, 듀플렉스 시퀀싱 검정과 BigBlue® *cII* 플라크 검정 사이의 상관관계는 조직 및 돌연변이원 치료 사이에 강하다.

[0213] 도 **3f**는 BigBlue® 마우스 조직으로부터 생성된 개별적으로 선별된 돌연변이체 플라크 및 BigBlue® 마우스 조직으로부터의 *cII*의 gDNA의 듀플렉스 시퀀싱에 대한 *cII* 유전자 내의 SNV의 비율을 보여준다. SNV는 기준품으로서 피리미딘으로 지정된다. 듀플렉스 시퀀싱은 3,510개의 플라크의 수동 수집에 의해 달성된 바와 같은 각각의 치료 그룹으로부터 동일한 돌연변이 스펙트럼을 생성한다(카이 제곱 시험으로 0.999 초과의 모든 3개의 p-값). 비율은 *cII* 간격 내에 기준 염기의 관찰된 수로 SNV의 총 관찰치를 나누고 1로 정규화함으로서 계산된다.

[0214] 도 **3g**는 코돈 위치 및 기능적 결과에 의한 모든 BigBlue® 조직 유형 및 치료 그룹에 걸쳐 *cII*의 직접적인 듀플렉스 시퀀싱에 의해 확인된 모든 돌연변이의 분포를 보여준다. 도 **3h**는 개별적으로 수집된 돌연변이체 플라크 중에서 확인된 돌연변이에 대한 분포 데이터를 보여준다. 도 **3g** 및 도 **3h**를 함께 참조하면, 직접적인 듀플렉스 시퀀싱(도 **3g**)은 모든 효과 종류를 야기하는 전체 유전자를 따른 돌연변이를 확인하는 반면, 선별된 돌연변이체 플라크로부터의 돌연변이(도 **3h**)는 단백질의 비임계 C 말단 및 N 말단에서의 동의 변이체 및 돌연변이가 없다. 이론에 의해 구속되지 않으면서, 단백질의 비임계 C 말단 및 N 말단에서의 동의 변이체 및 돌연변이가 플라크 검정 내의 선택적 성장 및 점수매김에 필요한 유전자 기능의 파괴를 야기하지 않는다고 생각된다.

[0215] 도 **4**는 듀플렉스 시퀀싱에 의해 측정된 MF가 각각의 치료 그룹 내에 일관된다는 것을 보여주는 막대 그래프이다. 모든 유전자에 걸쳐 집합체화된 MF는 듀플렉스 시퀀싱에 의해 간 및 골수에서 측정되었다. 고유한 돌연변이체의 수는 돌연변이원-노출된 마우스(118개의 돌연변이/26억개의 염기 쌍까지)에 비해 비히클 대조군 동물(1개 내지 13개의 돌연변이/14억개의 염기 쌍)에서 낮았다. 그룹 내의 동물들 사이의 MF는 모든 치료 조건에서 재현 가능하고, 대조군 동물에서의 낮은 수의 돌연변이(1개 내지 13개)는 MF의 튼튼한 추정치를 생성하기 위해 딥 시퀀싱의 필요성을 강조한다.

[0216] 도 **5a** 및 도 **5b**는 듀플렉스 시퀀싱에 의해 측정된 바와 같은 간(도 **5a**) 및 골수(도 **5b**)에서의 *cII* 전이유전자와 비교된 내인성 유전자의 MF를 보여주는 막대 그래프이다. 각각의 유전자(약 3 내지 6 kb)는 대략 5000x의 깊이로 시퀀싱되었고, *cII* 유전자(게놈마다 약 350 bp x 80 카피)는 약 100K 내지 300K의 깊이로 시퀀싱되었다. 돌연변이체 빈도는 도 **3a** 내지 도 **3d**와 관련하여 상기에 기재된 바대로 계산되었다. 도시된 바대로, 내인성 유전자는 *cII* 전이유전자와 유사한 MF 증가를 나타낸다. 듀플렉스 시퀀싱은 MF가 간보다 골수에서 더 높다는 것을 입증한다. 이론에 의해 구속되지 않으면서, 골수에서의 더 높은 속도의 세포 분열은 시험된 돌연변이원 둘 다에 검출된 더 높은 MF 수준을 설명할 수 있다. 더욱이, 도 **5a** 및 도 **5b**에 도시된 내인성 유전자의 반응의 차이가 내인성 유전자의 전사 상태 또는 크로매틱(chromatic) 구조의 차이와 관련될 수 있다.

[0217] 도 **5c**는 간 및 골수에 대한 유전자 영역에 의한 듀플렉스 시퀀싱에 대해 계산된 SNV MF를 보여주는 상자 그림 그래프이고, 도 **5d**는 도 **5c**에 도시된 집합체 데이터의 개별 측정치를 보여주는 산점도이다. 산란 점은 이를 둘러싼 95% CI로 개별 측정치를 보여준다. 도 **5c**에서의 상자 그림은 그 조직 및 치료 카테고리에 대한 모든 데이터 점의 모든 4개 사분위수를 보여준다. Y축 척도는 선형으로 10^{-7} 규모로 제시된다. 도 **5c**를 참조하면, 상자 그림은 도 **5d**에 도시된 BigBlue® 마우스 모델의 4개의 내인성 유전자 및 *cII* 전이유전자에 걸쳐 간 및 골수 조직에서의 SNV 돌연변이 빈도의 집합체를 요약한다. 돌연변이 유도의 정도는 특정 돌연변이원, 조직 유형 및 유전자 유전좌위에 의해 영향을 받는다.

[0218] 도 **6**은 듀플렉스 시퀀싱에 의해 측정된 바와 같은 시험된 조직 내의 각각의 시험 돌연변이원의 돌연변이 스펙트럼(예를 들어, 치료)을 보여주는 막대 그래프이다. 도 **6**을 참조하면, 모든 유전자에 걸쳐 집합체화되고, 각각의 샘플에 대해 계산되고 비지도된 계층적 클러스터 분석에 의해 그룹화된 각각의 돌연변이의 부분은 돌연변이 스펙트럼이 각각의 치료(예를 들어, 시험 돌연변이원)에 고유하다는 것을 나타낸다. 코딩된 데이터의 비지도된 클러스터 분석은 돌연변이 스펙트럼에 기초한 데이터의 그룹화를 허용하고, ENU 샘플이 T→C, T→A 및 C→T 돌연변이의 우세함에 의해 모든 조직에서 쉽게 확인된다는 것을 나타낸다. 마찬가지로, B[a]P 샘플은 C→A 및 G→T 돌연변이와 구별된다.

[0219] 도 **7a** 내지 도 **7c**는 비히클 대조군(7a), B[a]P(7b) 및 ENU(7c)에 대한 인접한 뉴클레오타이드의 상황에서의 돌연변이 스펙트럼(즉, 트리뉴클레오타이드 스펙트럼)을 보여주는 그래프이다. 트리뉴클레오타이드 스펙트럼 포맷에서의 돌연변이 서명은 상이한 돌연변이유발 기전에 관한 정보를 제공하고/하거나, 특정 돌연변이원에 고유한 돌연변이 패턴을 나타낸다. 예를 들어, CCG 및 CGC 상황은 다른 상황보다 B[a]P인 담배-연관된 발암물질에 보다 취약한 것으로 보인다(도 **7b**). 이 서명 패턴은 아플라톡신 노출에 의해 나타난 서명 패턴과 유사할 수 있다(예를 들어, 유사한 돌연변이유발 기전일 수 있음). 도 **7c**는 ENU인 알킬레이터가 S+[G][C]인 IUPAC 코드 GTS와 일치하는 2개의 취약한 상황을 갖고, 전이 돌연변이의 무거운 유도자라는 것을 예시한다.

[0220] 이 실시예에서, ENU 및 B[a]P-치료된 골수 및 간 샘플에서의 돌연변이 하중이 전통적인 BigBlue® *cII* 돌연변이체 플라크 빈도(돌연변이체 빈도 MF)와 필적하게 대조군에 대해 상당히 증가하였고, 조직 유형에 의해 유사하게 변했음이 나타났다. 스펙트럼 평가는 각각의 치료 그룹에서 INDELS 및 단일 염기 치환의 명확한 패턴을 밝혀냈다. 트리뉴클레오타이드 염기 분석은 인접한 뉴클레오타이드 상황이 돌연변이 가능성을 강하게 조절하고; 가장 극심한 핫스팟이 B[a]P에 대해 CCG 및 CGC 및 ENU에 대해 GTG 및 GTC라는 것을 나타냈다. 듀플렉스 시퀀싱은 4개의 내인성 유전자로 연장되었다: *Polr1c*, 로돕신, 합토글로빈 및 베타-카테닌. 다시, MF는 ENU 및 B[a]P에 노출된 동물에서 증가하지만, 게놈 유전좌위에 의해 상당히 변해서, 아마도 전사 상태를 반영한다. 이 실시예에서, 듀플렉스 시퀀싱은 TGR 검정에서 인정된 전임상 안전성 바이오마커인 *cII* 전이유전자에서 돌연변이를 검출하기 위한 성공적인 방법인 것으로 입증되었지만, 추가로, 이 실시예는 듀플렉스 시퀀싱이 내인성 암-관련된 유전자에 기초한 위험 평가 도구의 기초일 수 있다는 것을 입증한다.

[0221] *실시예 2*

[0222] 듀플렉스 시퀀싱을 사용한 포유류 게놈에서의 생체내 화학 돌연변이유발의 직접적인 정량화. 이 부문은 암 유발자 유전자에서의 조기 돌연변이가 시험 돌연변이원의 종양형성 가능성을 반영하는지를 결정하기 위해 듀플렉스 시퀀싱이 사용되는 실시예를 기재한다.

[0223] 이 실시예에서, FDA-허가된 암-경향이 있는 마우스 모델에서 상이한 마우스 조직 유형(폐, 비장, 혈액)에서 우레탄의 영향을 조사하였다: Tg.rasH2(Saitoh et al. Oncogene 1990. PMID 2202951). 이 마우스는 하나의 반접합성 대립유전자에서 발현을 부스팅하기 위해 활성화 인핸서 돌연변이를 갖는 인간 *Hras*의 약 3개의 탠덤 카피를 함유한다. 이 마우스는 비장 혈관육종 및 폐 선암의 경향이 있고, 2년 자연적 동물 연구를 치환하도록 6개월 발암성 연구에 일상적으로 사용된다. 마우스에서 발견된 종양은 보통 인간 Hras 프로토암유전자의 하나의 카피에서 획득된 활성화 돌연변이를 갖는다. 4개의 자연적 마우스 유전자(*Rho, Hp, Ctnnb1, Polr1c*) 이외에, 자연적 마우스 *Hras* 및 인간 *Hras* 전이유전자가 이 실시예에서 또한 분석된다.

[0224] 이 실시예에서, Tg.rasH2 마우스(n=5/그룹)는 우레탄의 비히클 또는 발암성 용량이 투약되고(1일, 3일, 5일), 29일에 표적 조직(폐, 비장) 및 전혈에서 듀플렉스 시퀀싱에 의해 돌연변이 검출을 위해 희생되었다. 내인성 유전자(*Rho, Hp, Ctnnb1, Polr1c*) 및 자연적 마우스 및 인간 *Hras* (전이)유전자를 또한 시퀀싱하였다.

[0225] 종양(비장 혈관육종; 폐 선암)은 이 종양에서 특징적인 암 유발자 돌연변이(CDM: cancer driver mutation)를 확

인하기 위해 우레탄이 투약되고 전장 엑솜 시퀀싱(WES: whole exome sequencing)으로 처리된 동물(n=5/그룹)로부터 11주에 수집되었다.

[0226] 도 8은 대조군 및 우레탄으로 치료된 실험 동물에 대한 폐, 비장 및 혈액 샘플의 돌연변이체 빈도(MF)를 보여주는 막대 그래프이다. 이 분석에서, 모든 고유한 검출된 변이체는 하나의 돌연변이로 계수되고, 이는 샘플마다 합계되었다. 이것은 전체 포착 면적에 걸쳐 시퀀싱된 듀플렉스 염기의 총 수로 나눠졌다. 사건의 수는 각각의 샘플 위에 표기된다. 전체로서, 모든 30개의 샘플에 걸쳐, 3,966,947,832개의 듀플렉스 시퀀싱된 염기 쌍이 생성되었다. 도 8에 도시된 것처럼, 돌연변이 유도는 동일한 치료 그룹에서 동물들 사이에 일관되고, 신뢰도는 시퀀싱 깊이에 따라 증가한다.

[0227] 도 9는 조직 샘플의 각각의 그룹에 걸쳐 평균 최소 점 돌연변이체 빈도를 보여주는 막대 그래프이다(오차 막대는 ± 1 표준 편차임).

표 1

[0228]

조직	치료	돌연변이 빈도	배수 증가	p-값
폐	비히클 대조군	0.67e-07		
폐	우레탄	5.04e-07	7.5x	6.73e-05
비장	비히클 대조군	0.83e-07		
비장	우레탄	2.73e-07	3.3x	1.92e-04
혈액	비히클 대조군	1.11e-07		
혈액	우레탄	2.39e-07	2.2x	0.003025

[0229] 도 9 및 표 1을 함께 참조하면, 비히클 대조군(VC)과 치료 그룹 사이의 차이는 고도로 유의미하였다. (불균등 분산에 대한) Welch t-시험은 돌연변이원 치료된 조직에 대해 대조군의 돌연변이체 빈도에 비해 그 조직의 돌연변이체 빈도의 유의도를 결정하기 위해 사용되었다. 혈액에 의한 약간 더 넓은 신뢰도 간격은 이 특정 예에서 혈액 VC 샘플에서의 시퀀싱의 더 낮은 평균 깊이를 반영한다. 본원에 기재된 방법을 사용하여 이것이 보정될 수 있다고 예측된다.

[0230] 도 10a는 표시된 치료 카테고리에 대해 폐, 비장 및 혈액에 대한 유전자 영역에 의한 듀플렉스 시퀀싱에 대해 계산된 SNV MF를 보여주는 상자 그림 그래프이고, 도 10b는 도 10a에 도시된 집합체 데이터의 개별 측정치를 보여주는 산점도이다. 산란 점은 이들을 둘러싼 95% CI로 개별 측정치를 보여준다. 도 10a에서의 상자 그림은 그 조직 및 치료 카테고리에 대한 모든 데이터 점의 모든 4개의 사분위수를 보여준다. Y축 척도는 선형으로 10^{-7} 규모로 제시된다. 도 10a를 참조하면, 상자 그림은 도 10b에 도시된 Tg-rasH2 마우스 모델의 폐, 비장 및 혈액에서의 SNV 돌연변이 빈도의 집합체를 요약한다. Tg-rasH2 마우스 모델에서 *cII* 전이유전자가 없다. 돌연변이 유도의 정도는 특정 돌연변이원, 조직 유형 및 유전자 유전좌위에 의해 영향을 받는다. 도 11은 듀플렉스 시퀀싱에 의해 측정된 바와 같은 시험된 조직 내에 우레탄 및 VC의 돌연변이 스펙트럼을 보여주는 막대 그래프이다. 도 11을 참조하면, 코딩된 데이터의 비지도된 클러스터 분석은 돌연변이 스펙트럼에 기초한 데이터의 그룹화를 허용하였다. 이 데이터는 뉴클레오타이드 변이의 단순한 스펙트럼 단독이 노출을 확인할 수 있다는 것을 입증한다. 다른 말로, 돌연변이원이 알려지지 않은 경우, 이 돌연변이원은 돌연변이 스펙트럼의 성질에 의해 노출된 유기체의 DNA의 듀플렉스 시퀀싱에 의해 신생 확인될 수 있었다.

[0231] 도 12a 및 도 12b는 비히클 대조군(12a) 및 우레탄(12b)에 대한 인접한 뉴클레오타이드의 상황에서의 돌연변이 스펙트럼(즉, 트리뉴클레오타이드 스펙트럼)을 보여주는 그래프이다. 트리뉴클레오타이드 스펙트럼 포맷에서의 돌연변이 서명은 상이한 돌연변이유발 기전에 관한 정보를 제공하고/하거나, 특정 돌연변이원에 고유한 돌연변이 패턴을 입증한다. 따라서, 트리뉴클레오타이드 상황("삼중항 서명") 내의 각각의 돌연변이 종류의 자세한 내역은 이러한 노출에 의해 생긴 종양으로부터의 클론성 돌연변이의 알려진 서명과 일치하게 각각의 치료 그룹에 대한 고도로 독특한 지문을 밝혀냈다. 비치료된 동물에서, 노화로부터 알려진 패턴인 구아닌의 산화 및 사이토신 및 5-me-사이토신의 탈아미노화에 의해 각각 생긴 C:G→A:T 및 C:G→G:C 돌연변이가 검출되었다. 우레탄 치료 후에, 모티프 "NTG" 내에 T:A→A:T는 가장 흔한 돌연변이로 보인다.

[0232] 도 13은 단일 뉴클레오타이드 변이체(SNV) 가닥 바이어스가 *Hp* 또는 *Rho* 게놈 영역에서가 아니라 *Ctnnb1* 및 *Polr1c*에서 관찰된다는 것을 보여준다. SNV 표기법은 전사된 가닥의 정방향에서 기준 뉴클레오타이드에 정규화된다. 개별 반복검증은 선 분절로 점과 95% 신뢰도 간격으로 보인다. 모든 돌연변이 빈도는 변이체 호출 영역 내에 각각의 기준 염기의 뉴클레오타이드 계수치에 대해 보정되었다. 가닥 무 바이어스에 대한 귀무 가설은 상호간 돌연변이에 동일한 빈도이다. C>N 및 T>N 변이체가 균일한 빈도로 있고, G>N 및 A>N 변이체가 상승된 빈도로 있으므로, 바이어스는 *Ctnnb1* 및 *Polr1c*에서 명확하다. *Hp* 및 *Rho*와 비교하여, 이론에 의해 구속되지 않으면서, 이 차이가 전사-커플링된 뉴클레오타이드 절제 복구 및 이 유전자의 상대 발현 수준으로 인한다고 생각된다.

[0233] 도 14는 듀플렉스 시퀀싱에 의해 검출된 바와 같은 변이체 대립유전자 분획의 초기 단계 신생물성 클론성 선택을 예시하는 그래프이다. 확인된 아주 대부분의 돌연변이는 단일 분자에서 예를 들어 1/10,000의 차수로 매우 낮은 변이체 대립유전자 분획(VAF)으로 발생했다. 적은 변이체가 샘플에서 다수의 분자에서 발견되었고, 상당히 더 높은 VAF를 갖는 것으로 확인되었다.

[0234] 도 15a는 Tg-rasH2 마우스 모델에서 인간 형질전환 유전좌위를 포함하는 유전자의 *Ras* 패밀리로부터 포획된 엑손에 대한 게놈 간격에 걸쳐 작도된 SNV를 예시하는 그래프이다. 일중항은 단일 분자에서 발견된 돌연변이이다. 다중항은 동일한 샘플러 내에 다수의 분자 내에 확인된 동일한 돌연변이이고, 클론성 확장 사건을 나타낼 수 있다. 각각의 점의 높이는 각각의 SNV의 변이체 대립유전자 빈도(VAF: variant allele frequency)에 상응하고, 점의 크기는 오직 다중한 관찰치에 상응한다. COSMIC에서의 *Ras* 패밀리 인간 암 돌연변이 핫스팟의 위치 및 상대 빈도는 각각의 유전자 아래에 표시된다. 도 *15b*는 인간 *HRAS* 전이유전자의 엑손 3에 정렬하는 단일 뉴클레오타이드 변이체(SNV)를 예시하는 그래프이다. 가장 흔한 *HRAS* 암-유발 핫스팟인 인간 *HRAS*의 엑손 3에서 코돈 번호 61에서의 중앙 잔기가 강조된다.

[0235] 도 15a 및 도 15b를 함께 참조하면, T>A 전환의 클러스터는 인간 종양발생 *Hras* 코돈 61 핫스팟에서 5개 중 4개의 우레탄-치료된 폐 샘플 및 5개 중 1개의 우레탄-치료된 비장 샘플에서 관찰되었다. 특히, 5개 중 4개의 치료된 폐 샘플은 0.1% 내지 1.8%의 변이체 대립유전자 빈도로 이 돌연변이를 보유하였다. 특히 이 클론은 상황 NTG에서 전환 T>A를 갖고, 이는 우레탄 돌연변이유발의 특징이다(도 12b에서 NTG 부위의 강한 선호를 지칭). 또한, 2개의 치료된 비장 샘플은 이 코돈에서 돌연변이를 갖는데, 하나는 이 동일한 위치에 있고 하나는 부근의 염기쌍에 있다. 5개 중 4개의 치료된 폐 샘플이 29일에 클론성으로 확장된 병원성 돌연변이를 갖는 한편, 패널에서 어딘가에 보이는 매우 적은 돌연변이는 1개 초과 구성원 클론으로 보이거나 다수의 샘플에서 (잘 확립된 암 유발자에서의 높은 VAF 다중항으로서) 반복하여 보인다는 관찰은 노출 바로 후 양성 선택의 강한 표시이다. 더욱이, 본 기술내용의 실시형태에 따른 듀플렉스 시퀀싱 방법은 이러한 초기 단계 신생물성 클론성 선택을 검출하는 데 필요한 민감도를 제공한다.

표 2

돌연변이 계수치	패밀리의 수	
1	829	
2	8	
4	1	
17	1	우레탄 치료된 폐 조직에서 인간 HRAS 유전자에서의 종양발생 AA61 T>A 돌연변이
58	1	
181	1	
300	1	

[0236]

[0237] 표 2를 참조하면, 97.5%의 돌연변이는 오직 단일 분자에서 확인되고, 1%는 2개의 분자에서 보이고, 약 0.5%는 2개 초과의 분자에서 보였다. 4개의 가장 높은 수준 클론은 모두 인간 *HRAS*에서 재발성 종양 핫스팟인 AA 61에서 종양발생 돌연변이로 발생했다. 가장 높은 수준 클론이 또한 암 핫스팟에 나타난다는 것은 강한 선택적 압력의 규모를 추가로 강조한다.

[0238] 시퀀싱된 듀플렉스 분자로 전환된 것보다 훨씬 더 많은 양의 DNA가 샘플마다 추출되었다. 추출된 조직 샘플의 부분은 거의 5 μg의 게놈 DNA를 생성하였다. 이것을 게놈 당량으로 전환하고, 3을 곱하여 추출에서 tg.HRAS 카

피의 수를 생성한다. 이것의 오직 약 1/3%가 시퀀싱되어서 검출되는 것보다 샘플링된 조직의 원래의 부분에 거의 300배 초과의 돌연변이체가 존재한다.

표 3

[0239]

샘플	ng DNA	게놈	카피 tg.HRAS	AA61에서의 깊이	시퀀싱된 카피 %	돌연변이체	원래의 샘플에서의 돌연변이체 세포
9957-폐 1	5,640	1,692,000	5,076,000	16,425	0.324%	300	92,712
9958-폐 1	4,400	1,320,000	3,960,000	16,319	0.412%	181	43,922
9959-폐 1	4,480	1,344,000	4,032,000	13,692	0.340%	58	17,080
9961-폐 1	4,700	1,410,000	4,230,000	14,706	0.348%	17	4,890

[0240] 이 실시예에서, 선택된 클론은 가장 높은 대립유전자 분획 클론에서 90,000개 초과의 세포를 포함하였다. 그 결과, 계산에 의하면, 예를 들어 돌연변이 노출의 시간으로부터 연구 29일 내에, 세포사가 없음을 추정하여, 이 세포의 배가 시간은 거의 1.8일마다 2^(29/1.8) ~ 90,000였다. 이론에 의해 구속되지 않으면서, 세포 배가의 이 계산된 비율은 짧은 시간(예를 들어, 2주만큼 짧게)에 이 선택된 돌연변이를 검출할 그럴듯한 능력을 제안한다.

[0241] 도 **16a** 내지 도 **16b**는 종래의 DNA 시퀀싱(도 **16a**) 및 듀플렉스 시퀀싱(도 **16b**)을 사용한 우레탄 치료 후에 마우스 폐에서의 인간 *HRAS*의 대표적인 400개의 염기 쌍 절편으로부터의 시퀀싱 데이터의 그래프 표현을 보여준다. 종래의 DNA 시퀀싱은 0.1% 내지 1%의 오류율을 갖고, 이는 진짜의 저빈도 돌연변이의 존재를 모호하게 한다. 도 **16a**는 본 연구에서 하나의 샘플(마우스 폐)의 1개의 유전자(인간 *HRAS*)의 대표적인 400 BP 섹션으로부터의 종래의 시퀀싱 데이터를 보여준다. 각각의 막대는 뉴클레오타이드 위치에 상응한다. 각각의 막대의 높이는 >100,000x 깊이로 시퀀싱될 때 그 위치에서 비기준 염기의 대립유전자 분획에 상응한다. 모든 위치는 약간의 빈도로 돌연변이되는 것으로 보이고; 이들의 거의 모두는 오류이다. 도 **16b**를 참조하면, 이것은 듀플렉스 시퀀싱으로 처리될 때 오직 하나의 돌연변이가 진짜임이 명확해진다.

[0242] 이 실시예의 실험 분석의 결과는 듀플렉스 시퀀싱이 극도로 튼튼하게 그리고 단단한 반복검증 신뢰도 간격으로 우레탄에 의한 돌연변이의 유도를 정량화한다는 것을 입증한다. 추가로, 돌연변이 유도의 정도는 조직-특이적이고, 폐는 비장 및 혈액보다 더 경향성이 있다. 우레탄 노출의 단순한 돌연변이 스펙트럼은 깨끗하고, 비바이어싱된 클러스터링은 그룹들 사이에 식별할 수 있다. 우레탄의 삼중항 돌연변이 스펙트럼은 "NTG"의 상황 내에 T→A 및 T→C 돌연변이에 대한 강한 경향을 보여주고, 돌연변이 스펙트럼은 비히클 대조군(및 다른 돌연변이원; 실시예 1 참조)로부터 구별 가능하다.

[0243] 추가적으로, 말초 혈액에서의 돌연변이 유도는 비장에서 보이는 것을 면밀하 반영하고, 말초 혈액의 생전 샘플링이 일부 돌연변이원에 대해 부검(또는 생검)을 치환한다는 것을 제안한다. 더욱이, 이 실시예는 심지어 29일에 인간 HRAS 전이유전자에서의 종양발생 돌연변이에 대한 선택의 명확한 증거가 듀플렉스 시퀀싱을 사용하여 나타난다는 것을 입증하였다. 이 핫스팟에서의 돌연변이의 스펙트럼은 이 알려진 돌연변이원의 효과를 정확히 반영하였다. 그러므로, 듀플렉스 시퀀싱은 미래의 암 위험의 바이오마커로서 초기 암 유발자 돌연변이를 평가하는 것과 관련하여 초기의 정확한 데이터를 제공할 수 있다. 종간 오염은 극도로 낮은 수준으로 지속하지만, 외래 종 오염의 제거는 자동적으로 자신 있게 수행되었다.

[0244] *실시예 3*

[0245] 듀플렉스 시퀀싱을 사용한 포유류 게놈에서의 돌연변이원 서명의 분석. 이 부문은 듀플렉스 시퀀싱 분석으로부터 생성된 데이터가 돌연변이원의 확인을 위해 돌연변이성 서명을 생성하고 비교하고/하거나 돌연변이원 노출을 확인하기 위해 사용될 수 있는 실시예를 기재한다.

[0246] 암 관련 체성 돌연변이 카탈로그(COSMIC: Catalogue of Somatic Mutations in Cancer) 데이터베이스는 게놈에 존재하여 발견된 돌연변이 유형의 고유한 조합으로 정의된 "돌연변이 서명"의 언급을 제공한다. 체성 돌연변이는 인간 신체의 모든 세포에 존재하고, 생애에 걸쳐 발생한다. 이러한 체성 돌연변이는 예를 들어 DNA 복제 기계의 고유한 약간의 불충, 외인성 또는 내인성 돌연변이원 노출, DNA의 효소 변형 및 결함성 DNA 복구를 포함하는 다수의 돌연변이 과정의 결과이다.

[0247] 도 **17a** 내지 도 **17c**는 COSMIC로부터의 서명 1(도 **17a**), 서명 4(도 **17b**) 및 서명 29(도 **17c**)에 대한 인접한 뉴클레오타이드의 상황에서의 돌연변이 스펙트럼(즉, 트리뉴클레오타이드 스펙트럼)을 보여주는 그래프이다. 도

17a를 참조하면, 서명 1은 CpG 부위에서 C>T 전이를 발생시키는 5-메틸-사이토신의 자발적 탈아미노화에 의해 생긴 제안된 병인론으로 모든 암 유형에서 보인다. 도 17b 내지 도 17c를 참조하면, 서명 4 및 서명 29는 흡연과 상관되고, 담배: 벤조[a]피렌에서의 주요 돌연변이원에 의해 유발된다. 서명 4는 패턴이 유사하지만 흡연자에서 폐암에서 가장 흔히 관찰되는 반면, 서명 29는 흡연자 및 담배를 씹는 사용자에서 가장 흔한 편평 식도암에서 주로 보인다.

표 4

특징	실시예 1	실시예 2	총
마우스 모델	BigBlue	Tg-rasH2	2 개의 균주
조직 (샘플)	간 (15) 골수 (17)	폐 (10) 비장 (10) 혈액 (10)	5 개의 유형
돌연변이원 (그룹마다 동물)	B[a]P (10) ENU (11) VC (11)	우레탄 (15) VC (15)	3 개의 돌연변이원
샘플	32	30	62
내인성 유전자위	*Polr1c* *Rho* *Ctnnb1* *Hp*	*Polr1c* *Rho* *Ctnnb1* *Hp* *Hras* *Kras* *Nras*	7 개의 자연적 유전자
형질전환 유전좌위	*cII*	인간 *Hras*	2 개의 전이유전자
듀플렉스 BP	4,074,604,194	4,348,868,321	8,423,472,515

[0248]

[0249] 표 4는 본원에 기술된 실시예 1 및 실시예 2로부터 도출된 실험 매개변수 및 데이터를 제공한다. 도 18은 실시예 1 및 실시예 2로부터의 모든 30개의 공개된 COSMIC 서명 및 4개의 코호트 스펙트럼의 비지도된 계층적 클러스터링을 보여준다. 클러스터링은 가중 (WGMA) 방법 및 코사인 유사도 메트릭으로 수행되었다. 특히, 벤조[a]피렌(BaP)은 서명 4 및 서명 29 둘 다에 매우 유사하고, 이는 담배 소비 또는 흡입을 통해 BaP 노출과 상관되었다. 비히클 대조군(VC)은 5-메틸-사이토신의 자발적 탈아미노화와 연결된 패턴인 서명 1과 같고, 반응성 산화성 종의 돌연변이성 효과와 5-메틸-사이토신의 자발적 탈아미노화 둘 다의 혼합을 나타내는 것으로 생각된다.

[0250] 이 실시예는 듀플렉스 시퀀싱이 확인 및 다른 분석의 목적을 위해 알려진 돌연변이 서명과 비교되고 이에 참조될 수 있는 돌연변이 스펙트럼 분석을 생성하기 위해 사용될 수 있다는 것을 입증한다.

[0251] 적합한 컴퓨팅 환경

[0252] 하기 논의는 적합한 컴퓨팅 환경의 일반 설명을 제공하고, 여기서 본 개시내용의 양태가 실행될 수 있다. 본 개시내용의 양태 및 실시형태는 컴퓨터-실행 가능한 명령, 예컨대 범용 컴퓨터, 예를 들어 서버 또는 개인용 컴퓨터에 의해 실행되는 루틴의 일반적인 상황에서 기재될 필요는 없지만 기재될 것이다. 당업자는 본 개시내용이 인터넷 어플라이언스, 휴대용 장치, 착용식 컴퓨터, 휴대폰 또는 이동 전화, 멀티-프로세서 시스템, 마이크로프로세서-기반 또는 프로그래밍 가능한 소비자 전자용품, 셋톱 박스, 네트워크 PC, 미니-컴퓨터, 메인프레임 컴퓨터 등을 포함하는 다른 컴퓨터 시스템 구성과 실행될 수 있다는 것을 이해할 것이다. 본 개시내용은 하기 자세히 설명되는 하나 이상의 컴퓨터-실행 가능한 명령을 수행하도록 특별히 프로그래밍되거나 구성되거나 구축된 특수 목적 컴퓨터 또는 데이터 프로세서에서 구현될 수 있다. 실제로, 본원에서 일반적으로 사용된 바와 같은

용어 "컴퓨터"는 임의의 상기 장치, 및 임의의 데이터 프로세서를 지칭한다.

[0253] 본 개시내용은 근거리 통신망("LAN")(Local Area Network), 광역 통신망("WAN")(Wide Area Network) 또는 인터넷과 같은 통신 네트워크를 통해 연결된 원격 프로세싱 장치에 의해 작업 또는 모듈이 수행되는 분산 컴퓨팅 환경에서 또한 실행될 수 있다. 분산 컴퓨팅 환경에서, 프로그램 모듈 또는 하위루틴은 근거리 및 원격 기억 저장 장치 둘 다에 위치할 수 있다. 하기에 기술된 본 개시내용의 양태는 자기 및 광학 판독 가능하고 제거 가능한 컴퓨터 디스크를 포함하는 컴퓨터 판독 가능한 매체에 저장되거나 분산되고, 칩(예를 들어, EEPROM 칩)에서 펌웨어로 저장될 뿐만 아니라, 인터넷 또는 다른 네트워크(무선 네트워크를 포함)에 전자로 분산될 수 있다. 당업자는 본 개시내용의 부분이 서버 컴퓨터에 있을 수 있지만, 상응하는 부분이 클라이언트 컴퓨터에 있다는 것을 인식할 것이다. 본 개시내용의 양태에 특정한 데이터 구조 및 데이터의 전송은 또한 본 개시내용의 범위 내에 포괄된다.

[0254] 개인용 컴퓨터 또는 워크스테이션과 같은 컴퓨터의 실시형태는 하나 이상의 사용자 입력 장치 및 데이터 저장 장치에 연결된 하나 이상의 프로세서를 포함할 수 있다. 컴퓨터는 또한 적어도 하나의 출력 장치, 예컨대 디스플레이 장치 및 하나 이상의 선택적인 추가 출력 장치(예를 들어, 프린터, 플로터, 스피커, 촉각 또는 후각 출력 장치 등)에 연결될 수 있다. 컴퓨터는 예컨대 선택적인 네트워크 연결, 무선 트랜시버 또는 둘 다를 통해 외부 컴퓨터에 연결될 수 있다.

[0255] 다양한 입력 장치는 키보드 및/또는 포인팅 장치, 예컨대 마우스를 포함할 수 있다. 다른 입력 장치, 예컨대 마이크로폰, 조이스틱, 펜, 터치 스크린, 스캐너, 디지털 카메라, 비디오 카메라 등이 가능하다. 추가의 입력 장치는 시퀀싱 기계(들)(예를 들어, 대량 병렬 시퀀서), 형광투시경 및 다른 실험실 설비 등을 포함할 수 있다. 적합한 데이터 저장 장치는 컴퓨터에 의해 접근 가능한 데이터를 저장할 수 있는 임의의 유형의 컴퓨터 판독 가능한 매체, 예컨대 자기 하드 및 플로피 디스크 드라이브, 광학 디스크 드라이브, 자기 카세트, 테이프 드라이브, 플래시 메모리 카드, 디지털 비디오 디스크(DVD: digital video disk), 베르누이 카트리지(Bernoulli cartridge), RAM, ROM, 스마트 카드 등을 포함할 수 있다. 실제로, 근거리 통신망(LAN), 광역 통신망(WAN) 또는 인터넷과 같은 네트워크에 대한 연결 포트 또는 이것 위의 노드를 포함하는 컴퓨터 판독 가능한 명령 및 데이터를 저장하거나 전송하기 위한 임의의 매체를 사용할 수 있다.

[0256] 본 개시내용의 양태는 다양한 다른 컴퓨팅 환경에서 실행될 수 있다. 예를 들어, 네트워크 인터페이스를 갖는 분산된 컴퓨팅 환경은 시스템에서 하나 이상의 사용자 컴퓨터를 포함할 수 있고, 이 시스템에서 이들은 컴퓨터가 인터넷의 월드 와이드 웹(World Wide Web) 부분 내의 웹 사이트를 포함하는 인터넷에 접근하고 인터넷과 데이터를 교환하게 하는 브라우저 프로그램 모듈을 포함할 수 있다. 사용자 컴퓨터는 운영 시스템, 하나 이상의 어플리케이션 프로그램(예를 들어, 워드 프로세싱 또는 스프레드 시트 어플리케이션) 등과 같은 다른 프로그램 모듈을 포함할 수 있다. 컴퓨터는 다양한 유형의 어플리케이션을 실행하도록 프로그래밍될 수 있는 범용 장치일 수 있거나, 특정 기능 또는 기능 종류로 최적화되거나 제한된 단일 목적 장치일 수 있다. 보다 중요하게는, 네트워크 브라우저와 기재되어 있지만, 사용자에게 그래픽 사용자 인터페이스를 제공하기 위한 임의의 어플리케이션 프로그램을 하기 자세히 기재된 것처럼 사용할 수 있고; 웹 브라우저 및 웹 인터페이스의 사용은 여기서 친숙한 예로서 오직 사용된다.

[0257] 인터넷 또는 월드 와이드 웹("웹(Web)")에 연결된 적어도 하나의 서버 컴퓨터는 본원에 기재된 웹 페이지, 데이터 스트림, 오디오 신호 및 전자 영상과 같은 전자 메시지를 수신하고 라우팅하고 저장하기 위한 더 많은 또는 모든 기능을 수행할 수 있다. 인터넷이 기재되어 있지만, 인터넷과 같은 전용 네트워크가 일부 어플리케이션에서 사실 바람직할 수 있다. 네트워크는 클라이언트-서버 구성을 가질 수 있고, 여기서 컴퓨터는 다른 클라이언트 컴퓨터의 서빙에 전용이거나, 피어투피어식(peer-to-peer)과 같은 다른 구성을 가질 수 있고, 여기서 하나 이상의 컴퓨터는 서버 및 클라이언트로서 동시에 작용한다. 서버 컴퓨터(들)에 연결된 데이터베이스 또는 데이터베이스들은 더 많은 웹 페이지를 저장하고, 사용자 컴퓨터 사이에 컨텐츠 교환될 수 있다. 데이터베이스(들)를 포함하는 서버 컴퓨터(들)는 시스템에서 악성 공격을 저해하고 여기에 저장된 메시지 및 데이터의 통합성을 보존하기 위한 보안 조치(예를 들어, 방화벽 시스템, 보안 소켓 계층(SSL: secure socket layer), 패스워드 보호 체계, 부호 매김 등)를 사용할 수 있다.

[0258] 적합한 서버 컴퓨터는 다른 특징들 중에서 서버 엔진, 웹 페이지 관리 구성성분, 컨텐츠 관리 구성성분 및 데이터베이스 관리 구성성분을 포함할 수 있다. 서버 엔진은 기본 프로세싱 및 운영 시스템 수준 작업을 수행한다. 웹 페이지 관리 구성성분은 웹 페이지의 생성 및 디스플레이 또는 라우팅을 다룬다. 사용자는 서버 컴퓨터와 연관된 URL에 의해 서버 컴퓨터에 접근할 수 있다. 컨텐츠 관리 구성성분은 본원에 기재된 실시형태에서 대부분의

기능을 다룬다. 데이터베이스 관리 구성성분은 데이터베이스와 관련한 저장 및 검색 작업, 데이터베이스에 대한 쿼리, 데이터베이스에 대한 리드 및 라이트 기능 및 비디오, 그래픽 및 오디오 신호와 같은 데이터의 저장을 포함한다.

[0259] 본원에 기재된 많은 기능적 유닛은 보다 구체적으로 이의 실행 독립성을 강조하기 위해 모듈로서 표지되었다. 예를 들어, 모듈은 다양한 유형의 프로세서에 의한 실행을 위해 소프트웨어에서 실행될 수 있다. 실행 가능한 코드의 확인된 모듈은 예를 들어 객체, 절차 또는 함수로 체계화될 수 있는 예를 들어 컴퓨터 명령의 하나 이상의 물리적 블록 또는 논리적 블록을 포함할 수 있다. 컴퓨터 명령의 확인된 블록은 물리적으로 함께 배치될 필요는 없고, 논리적으로 함께 연결될 때 모듈을 포함하고 모듈에 대한 기술된 목적을 달성하는, 상이한 위치에 저장된 별개의 명령을 포함할 수 있다.

[0260] 모듈은 또한 커스텀 VLSI 회로 또는 게이트 어레이, 재고품 반도체, 예컨대 로직 칩, 트랜지스터 또는 다른 별개의 성분을 포함하는 하드웨어 회로로서 실행될 수 있다. 모듈은 또한 필드 프로그래밍 가능한 게이트 어레이, 프로그래밍 가능한 어레이 로직, 프로그래밍 가능한 로직 장치 등과 같은 프로그래밍 가능한 하드웨어 장치에서 실행될 수 있다.

[0261] 실행 가능한 코드의 모듈은 단일 명령 또는 많은 명령일 수 있고, 몇몇 메모리 장치에 걸쳐 상이한 프로그램들 중에서 몇몇 상이한 코드 세그먼트 위로 심지어 분포될 수 있다. 유사하게, 운영 데이터는 모듈 내에서 본원에서 확인되고 예시될 수 있고, 임의의 적합한 형태로 구현되고 임의의 적합한 유형의 데이터 구조 내에 체계화될 수 있다. 운영 데이터는 단일 데이터세트로서 수집될 수 있거나, 상이한 저장 장치를 포함하여 상이한 위치 위로 분포될 수 있고, 적어도 부분적으로 시스템 또는 네트워크에서 단순히 전자 신호로 존재할 수 있다.

[0262] 유전독성 시험을 위한 시스템

[0263] 본 발명은 대상체의 샘플을 처리하고, 샘플의 오류-보정된 서열 리드(예를 들어, 듀플렉스 서열 리드, 듀플렉스 공통 서열 등), 돌연변이 스펙트럼, 돌연변이체 빈도, 삼중항 돌연변이 서명, 및 샘플 데이터와 하나 이상의 알려진 유전독소와 연관된 상응하는 데이터 사이의 유사성이 있는지를 결정하기 위해 시퀀싱 데이터를 유선 네트워크 또는 무선 네트워크를 통해 원격 서버에 전송하기 위한 시스템(예를 들어, 네트워크 컴퓨터 시스템, 고속 자동화 시스템 등)을 추가로 포함한다.

[0264] 하기에 그리고 도 19에 예시된 실시형태와 관련하여 보다 자세히 기재된 것처럼, 유전독소 전산화 시스템은 (1) 원격 서버; (2) 시퀀싱 데이터를 생성하고/하거나 전송할 수 있는 복수의 사용자 전자 컴퓨팅 장치; (3) 알려진 유전독소 프로파일 및 연관된 정보를 갖는 데이터베이스(선택적); 및 (4) 전자 컴퓨팅 장치, 데이터베이스와 원격 서버 사이에 전자 통신을 전송하기 위한 유선 네트워크 또는 무선 네트워크를 포함한다. 원격 서버는 (a) 사용자 유전독소 기록 결과, 및 유전독소 프로파일(예를 들어, 스펙트럼, 빈도, 작용 기전 등)의 기록을 저장하는 데이터베이스; (b) 메모리에 통신 연결된 하나 이상의 프로세서; 및 프로세서(들)에 대한 명령을 포함하는 하나 이상의 비일시적 컴퓨터 판독 가능한 저장 장치 또는 매체를 추가로 포함하고, 여기서 상기 프로세서는 도 20 내지 도 23에 기재된 단계 중 하나 이상을 포함하는 연산을 수행하기 위해 상기 명령을 실행하도록 구성된다.

[0265] 일 실시형태에서, 본 기술내용은 하나 이상의 프로세서에 의해 실행될 때, 대상체가 적어도 하나의 유전독소에 노출되는지 및/또는 이의 정체 또는 특성/특징을 결정하는 방법을 수행하는 명령을 포함하는 비일시적 컴퓨터 판독 가능한 저장 매체를 추가로 포함한다. 특정 실시형태에서, 이 방법은 도 20 내지 도 23에 기재된 단계 중 하나 이상을 포함할 수 있다.

[0266] 본 기술내용의 추가 양태는 대상체가 적어도 하나의 유전독소에 노출되는지 및/또는 이의 정체 또는 특성/특징을 결정하기 위한 전산화 방법에 관한 것이다. 특정 실시형태에서, 이 방법은 도 20 내지 23에 기재된 단계 중 하나 이상을 포함할 수 있다.

[0267] 도 19는 유전독성 노출로부터 생긴 돌연변이성 사건 및/또는 핵산 손상 사건을 확인하기 위해 본원에 개시된 방법 및/또는 키트와 사용하기 위한 컴퓨터 프로그램 제품(1950)이 설치된 컴퓨터 시스템(1900)의 블록 다이어그램이다. 도 19가 다양한 컴퓨팅 시스템 구성성분을 예시하지만, 상기에 기술된 것과 같은 당업자에게 알려진 다른 또는 상이한 구성성분이 본 개시내용의 양태가 실행될 수 있는 적합한 컴퓨팅 환경을 제공할 수 있다는 것이 고려된다. 도 20은 본 기술내용의 실시형태에 따른 듀플렉스 시퀀싱 공통 서열 데이터를 제공하기 위한 루틴을 예시하는 흐름 다이어그램이다. 도 21 내지 도 23은 샘플의 유전독성 노출로부터 생긴 돌연변이성 사건 및/또는 핵산 손상 사건을 확인하기 위한 다양한 루틴을 예시하는 흐름 다이어그램이다. 본 기술내용의 양태에 따르면, 도 21 내지 도 23과 관련하여 기재된 방법은 예를 들어 샘플의 돌연변이 스펙트럼, 돌연변이체 빈도, 삼중항 돌

연변이 스펙트럼 및 알려진 유전독소의 데이터세트와의 샘플 데이터의 비교로부터 도출된 정보를 포함하는 샘플 데이터를 제공할 수 있다.

[0268] 도 19에 예시된 것처럼, 컴퓨터 시스템(1900)은 샘플의 유전독성 노출로부터 생긴 돌연변이성 사건 및/또는 핵산 손상 사건을 분석하기 위해 복수의 사용자 컴퓨팅 장치(1902, 1904); 유선 네트워크 또는 무선 네트워크(1910) 및 프로세서를 포함하는 원격 서버("DupSeq™" 서버)(1940)를 포함할 수 있다. 실시형태에서, 사용자 컴퓨팅 장치(1902, 1904)는 시퀀싱 데이터를 생성하고/하거나 전송하도록 사용될 수 있다. 일 실시형태에서, 컴퓨팅 장치(1902, 1904)의 사용자는 유전독성을 평가하기 위한 대상체 샘플의 듀플렉스 시퀀싱 방법 단계과 같은 본 기술내용의 다른 양태를 수행하는 자일 수 있다. 하나의 예에서, 컴퓨팅 장치(1902, 1904)의 사용자는 대상체 샘플에 대한 정보를 얻도록 본 기술내용의 실시형태에 따라 시약 및/또는 어댑터를 포함하는 키트(1, 2)로 소정의 듀플렉스 시퀀싱 방법 단계를 수행한다.

[0269] 각각의 사용자 컴퓨팅 장치(1902, 1904)는 예시된 것처럼적어도 하나의 중앙 처리 장치(1906), 메모리(1907) 및 사용자 및 네트워크 인터페이스(1908)를 포함한다. 일 실시형태에서, 사용자 장치(1902, 1904)는 데스크탑, 랩탑 또는 태블릿 컴퓨터를 포함한다.

[0270] 2개의 사용자 컴퓨팅 장치(1902, 1904)가 도시되어 있지만, 임의의 수의 사용자 컴퓨팅 장치가 포함되거나 이 시스템(1900)의 다른 구성성분에 연결될 수 있음이 고안된다. 추가적으로, 컴퓨팅 장치(1902, 1904)는 또한 샘플을 증폭시키고 시퀀싱하기 위해 사용자 (1) 및 사용자 (2가 사용하는 복수의 장치 및 소프트웨어를 나타낼 수 있다. 예를 들어, 컴퓨팅 장치는 시퀀싱 기계(예를 들어, Illumina HiSeg™, Ion Torrent PGM, ABI SOLiD™ 시퀀서, PacBio RS, Helicos Heliscope™ 등), 실시간 PCR 기계(예를 들어, ABI 7900, Fluidigm BioMark™ 등), 마이크로어레이 기구 등일 수 있다.

[0271] 이 시스템(1900)은 상기에 기재된 구성성분 이외에 유전독소 프로파일 및 연관된 정보를 저장하기 위한 데이터베이스(1930)를 추가로 포함할 수 있다. 예를 들어, 서버(1940)가 접근 가능할 수 있는 데이터베이스(1930)는 복수의 알려진 유전독소에 대한 돌연변이 스펙트럼, 삼중항 돌연변이 스펙트럼/서명, 작용 기전 등의 기록 또는 집합체를 포함할 수 있고, 또한 각각의 저장된 유전독소의 돌연변이 프로파일/패턴에 관한 추가 정보를 포함할 수 있다. 특정 예에서, 데이터베이스(1930)는 유전독소 프로파일을 포함하는 제3자 데이터베이스(1932)일 수 있다. 예를 들어, 암 관련 체성 돌연변이 카탈로그(COSMIC) 웹사이트는 흡연자에서의 폐암과 같은 발암물질에 대한 노출로부터 생기는 종양에서 클론성 돌연변이로 발견되는 "돌연변이 스펙트럼"의 집합체를 포함한다[8,9], 다른 실시형태에서, 데이터베이스는 서버(1940)로부터 별개로 호스팅된 자립형 데이터베이스(1930)(개인용 또는 비개인용)일 수 있거나, 경험적으로 도출된 유전독소 프로파일(1972)을 포함하는 데이터베이스(1970)와 같은 데이터베이스는 서버(1940)에 호스팅될 수 있다. 일부 실시형태에서, 새로운 시험 물질/인자 프로파일을 생성하기 위해 이 시스템(1900)이 사용되면서, 이 시스템(1900) 및 연관된 방법(예를 들어, 본원에서 그리고 예를 들어도 20 내지 23에 기재된 방법)을 사용하여 생긴 데이터는 미래의 비교 활동을 위해 추가 유전독소 프로파일(1932, 1972)이 생성될 수 있도록 데이터베이스(1930 및/또는 1970)에 업로딩될 수 있다.

[0272] 서버(1940)는 네트워크(1910)를 통해 사용자 컴퓨팅 장치(1902, 1904)로부터 시퀀싱 데이터(예를 들어, 원시 시퀀싱 파일) 및 관련된 정보를 수신하고 컴퓨팅하고 분석하도록 구성될 수 있다. 샘플-특이적 원시 시퀀싱 데이터는 장치(1902, 1904)에 설치되거나 네트워크(1910)를 통해 원격 서버(1940)로부터 접근 가능한 컴퓨터 프로그램 제품/모듈(서열 모듈(1905))을 사용하여, 또는 당해 분야에서 잘 알려진 다른 시퀀싱 소프트웨어를 사용하여 근거리에서 컴퓨팅될 수 있다. 이후, 원시 서열 데이터는 네트워크(1910)를 통해 원격 서버(1940)로 전송될 수 있고, 사용자 결과(1974)는 데이터베이스(1970)에 저장될 수 있다. 서버(1940)는 또한 데이터베이스(1970)로부터 원시 시퀀싱 데이터를 수신하도록 구성되고, 예를 들어 본원에 개시된 듀플렉스 시퀀싱 기법을 사용하여 오류 보정된 이중-가닥 서열 리드를 컴퓨터 사용하여 생성하도록 구성된 프로그램 제품/모듈 "DS 모듈"(1912)을 포함한다. 서버(1940)에 DS 모듈(1912)이 도시되어 있지만, 당업자는 DS 모듈(1912)이 대안적으로 장치(1902, 1904)에서 조작되어 호스팅되거나 다른 원격 서버(비도시)에서 호스팅될 수 있다는 것을 인식할 것이다.

[0273] 원격 서버(1940)는 적어도 하나의 중앙 처리 장치(CPU: central processing unit)(1960), 사용자 및 네트워크 인터페이스(1962)(또는 인터페이스가 서버에 연결된 서버-전용 컴퓨팅 장치), 상기에 기재된 것과 같은 데이터베이스(1970)와 알려진 유전독소 및 새로운 유전독소의 돌연변이 프로파일(1972)을 저장하기 위한 복수의 컴퓨터 파일/기록, 및 시험된 샘플에 대한 결과(예를 들어, 원시 시퀀싱 데이터, 듀플렉스 시퀀싱 데이터, 유전독성 분석 등)(1974)를 저장하기 위한 파일/기록을 포함할 수 있다. 서버(1940)는 본 기술내용의 양태에 따라 유전독소 컴퓨터 프로그램 제품(Genotoxin Computer Program Product)(유전독소 모듈)(1950)이 저장되는 저장된 컴퓨

터 메모리(1911)를 추가로 포함한다.

[0274] 컴퓨터 프로그램 제품/모듈(1950)은, 컴퓨터(예를 들어, 서버(1940))에서 실행될 때, 유전독소를 검출하고 확인하기 위한 본원에 개시된 방법의 단계를 수행하는 비일시적 컴퓨터 판독 가능한 매체에서 구현된다. 본 개시내용의 다른 양태는 프로세서가 유전독성 분석(예를 들어, 컴퓨트 돌연변이체 빈도, 돌연변이 스펙트럼, 삼중항 돌연변이 스펙트럼, 유전독소 비교 기록, 역치 수준 기록 등)을 수행하게 하기 위한 컴퓨터 판독 가능한 프로그램 코드 또는 명령이 구현되는 비일시적 컴퓨터 사용 가능 매체를 포함하는 컴퓨터 프로그램 제품/모듈(1950)을 포함한다. 이 컴퓨터 프로그램 명령은 기계를 제조하기 위해 컴퓨터 또는 다른 프로그래밍 가능한 장치에 로딩될 수 있어서, 컴퓨터 또는 다른 프로그래밍 가능한 장치에서 실행하는 명령은 본원에 기재된 기능 또는 단계를 실행하기 위한 수단을 생성한다. 이 컴퓨터 프로그램 명령은 또한 컴퓨터 또는 다른 프로그래밍 가능한 장치가 특정 방식으로 작용하도록 지시할 수 있는 컴퓨터 판독 가능한 메모리 또는 매체에 저장될 수 있어서, 컴퓨터 판독 가능한 메모리 또는 매체에 저장된 명령은 분석을 실행하는 지시 수단을 포함하는 제조 물품을 제조한다. 컴퓨터 프로그램 명령은 또한 일련의 연산 단계가 컴퓨터 또는 다른 프로그래밍 가능한 장치에서 수행되게 하여 컴퓨터 실행된 프로세스를 생성시키는 컴퓨터 또는 다른 프로그래밍 가능한 장치에 로딩될 수 있어서, 컴퓨터 또는 다른 프로그래밍 가능한 장치에서 실행하는 명령은 상기에 기재된 기능 또는 단계를 실행하기 위한 단계를 제공한다.

[0275] 더욱이, 컴퓨터 프로그램 제품/모듈(1950)은 임의의 적합한 언어 및/또는 브라우저에서 실행될 수 있다. 예를 들어, 이것은 바람직하게는 Visual Basic, SmallTalk, C++ 등과 같은 객체-지향 고수준 프로그래밍 언어를 사용하여 Python, C 언어로 실행될 수 있다. 어플리케이션은 Windows™ 98, Windows™ 2000, Windows™ NT 등을 포함하는 Microsoft Windows™ 환경과 같은 환경에 맞도록 쓰여질 수 있다. 또한, 어플리케이션은 MacIntosh™, SUN™, UNIX 또는 LINUX 환경에 대해 또한 쓰여질 수 있다. 또한, 기능적 단계는 범용 또는 플랫폼-독립적 프로그래밍 언어를 사용하여 또한 실행될 수 있다. 이러한 멀티-플랫폼 프로그래밍 언어의 예는 하이퍼텍스트 마크업 언어(HTML: hypertext markup language), JAVA™, JavaScript™, 플래시 프로그래밍 언어, 공통 게이트웨이 인터페이스/구조화 질의 언어(CGI/SQL: common gateway interface/structured query language), 실용적인 추출 및 보고 언어(PERL: practical extraction report language), AppleScript™ 및 다른 시스템 스크립트 언어, 프로그래밍 언어/구조화 질의 언어(PL/SQL: programming language/structured query language) 등을 포함하지만, 이들로 제한되지는 않는다. HotJava™, Microsoft™ Explorer™ 또는 Netscape™과 같은 Java™- 또는 JavaScript™-지원 브라우저를 사용할 수 있다. 액티브 컨텐츠 웹 페이지가 사용될 때, 이것은 Java™ 애플릿 또는 ActiveX™ 컨트롤 또는 다른 액티브 컨텐츠 기술을 포함할 수 있다.

[0276] 이 시스템은 다수의 루틴을 호출한다. 일부 루틴이 본원에 기재되어 있지만, 당업자는 이 시스템이 수행하는 다른 루틴을 확인할 수 있다. 게다가, 본원에 기재된 루틴은 다양한 방식으로 변경될 수 있다. 일례로서, 예시된 로직의 순서가 재배열될 수 있고, 하위단계는 병렬로 수행될 수 있고, 예시된 로직은 생략될 수 있고, 다른 로직이 포함될 수 있고, 기타 등등이다.

[0277] 도 20 내지 도 23은 샘플의 유전독성 노출로부터 생긴 돌연변이성 사건 및/또는 핵산 손상 사건을 검출하고 확인하기 위한 루틴 2000, 2100, 2200, 2300을 예시하는 흐름 다이어그램이다. 도 20은 샘플(예를 들어, 유전독성 검정으로부터의 샘플)에서 이중-가닥 핵산 분자에 대한 듀플렉스 시퀀싱 데이터를 제공하기 위한 루틴 2000을 예시하는 흐름 다이어그램이다. 루틴 2000은 컴퓨터 네트워크에 연결된 클라이언트 컴퓨터 또는 서버 컴퓨터와 같은 컴퓨팅 장치에 의해 호출될 수 있다. 일 실시형태에서, 컴퓨팅 장치는 서열 데이터 생성장치 및/또는 서열 모듈을 포함한다. 일례로서, 컴퓨팅 장치는 운영자가 컴퓨팅 장치와 통신하는 사용자 인터페이스를 연동시킨 후 루틴 2000을 호출할 수 있다.

[0278] 루틴 2000은 블록 2002에서 시작하고, 서열 모듈은 사용자 컴퓨팅 장치로부터 원시 서열 데이터를 수신(블록 2004)하고, 샘플에서 복수의 핵산 분자로부터 도출된 복수의 원시 서열 리드를 포함하는 샘플-특정 데이터세트를 생성(블록 2006)한다. 일부 실시형태에서, 서버는 차후의 프로세싱을 위해 데이터베이스에서 샘플-특정 데이터세트를 저장할 수 있다. 다음에, DS 모듈은 샘플-특정 데이터세트에서 원시 서열 데이터로부터 듀플렉스 공통 시퀀싱(Duplex Consensus Sequencing) 데이터를 생성하기 위한 요청을 수신(블록 2008)한다. DS 모듈은 (예를 들어, SMI 서열에 기초하여) 원래의 이중-가닥 핵산 분자를 나타내는 패밀로부터 서열 리드를 그룹화하고, 개별 가닥으로부터의 대표적인 서열을 서로 비교(블록 2010)한다. 일 실시형태에서, 대표적인 서열은 각각의 원래의 핵산 분자로부터의 하나 또는 하나 초과의 서열 리드일 수 있다. 다른 실시형태에서, 대표적인 서열은 대표적인 가닥 내의 정렬 및 오류-보정으로부터 생성된 단일-가닥 공통 서열(SSCS)일 수 있다. 이러한 실시형태에서, 제1

가닥으로부터의 SSCS는 제2 가닥으로부터의 SSCS와 비교될 수 있다.

[0279] 블록 2012에서, DS 모듈은 비교된 대표적인 가닥들 사이에 상보성의 뉴클레오타이드 위치를 확인한다. 예를 들어, DS 모듈은 뉴클레오타이드 염기 콜이 동의하는 비교된(예를 들어, 정렬된) 서열 리드를 따라 뉴클레오타이드 위치를 확인한다. 추가적으로, DS 모듈은 비교된 대표적인 가닥들 사이에 비상보성의 위치를 확인(블록 2014)한다. 마찬가지로, DS 모듈은 뉴클레오타이드 염기 콜이 동의하지 않는 비교된(예를 들어, 정렬된) 서열 리드를 따라 뉴클레오타이드 위치를 확인할 수 있다.

[0280] 다음에, DS 모듈은 샘플에서 이중-가닥 핵산 분자에 대한 듀플렉스 시퀀싱 데이터를 제공(블록 2016)할 수 있다. 이러한 데이터는 각각의 처리된 서열 리드에 대한 듀플렉스 공통 서열의 형태일 수 있다. 듀플렉스 공통 서열은 일 실시형태에서 원래의 핵산 분자의 각각의 가닥을 형성하는 대표적인 서열이 동의하는 뉴클레오타이드 위치만을 포함할 수 있다. 따라서, 일 실시형태에서, 비동의의 위치는 제거되거나 그렇지 않으면 무시될 수 있어서, 듀플렉스 공통 서열은 오류-보정된 고정확성 서열 리드이다. 다른 실시형태에서, 듀플렉스 시퀀싱 데이터는 (예를 들어, DNA 손상이 평가될 수 있는 경우에) 비동의의 뉴클레오타이드 위치가 추가로 분석될 수 있도록 이러한 위치에서 리포팅 정보를 포함할 수 있다. 이후, 루틴 2000은 블록 2018에 계속 이어질 수 있고, 여기서 이것은 종료한다. 의심

[0281] 도 21은 샘플의 유전독성 노출로부터 생긴 돌연변이성 사건을 검출하고 확인하기 위한 루틴 2100을 예시하는 흐름 다이어그램이다. 이 루틴은 도 20의 컴퓨팅 장치에 의해 호출될 수 있다. 루틴 2100은 블록 2102에서 시작하고, 유전독소 모듈은 (예를 들어, 블록 2016 후에) 도 20으로부터의 듀플렉스 시퀀싱 데이터를 기준 서열 정보와 비교(블록 2104)하고, 돌연변이(예를 들어, 여기서 해당 순서는 기준 순서에서 변함)를 확인(블록 2106)한다. 다음에, 유전독소 모듈은 샘플에 대한 돌연변이체 빈도를 결정(블록 2108)하고, 돌연변이 스펙트럼을 생성(블록 2110)한다. 그러므로, 돌연변이 패턴 분석은 샘플로부터 분석된 핵산 분자에서의 돌연변이 사건의 유형, 위치 및 빈도에 관한 정보가 제공될 수 있다. 선택적으로, 유전독소 모듈은 삼중항 돌연변이 스펙트럼을 생성(블록 2112)하여 노출의 유전독성 결과를 분석하기 위한 트리뉴클레오타이드 상황 및 패턴 정보를 제공할 수 있다.

[0282] 유전독소 모듈은, 예를 들어 샘플이 알려진 유전독소에 노출되는지를 결정하기 위해, 또는 다른 예에서 시험 물질/인자가 이전에 알려진 유전독소와 유사한 유전독성 프로파일을 갖는지를 결정하기 위해, 또한 선택적으로 (결정되는 경우) 돌연변이 스펙트럼 및/또는 삼중항 돌연변이 스펙트럼을 복수의 알려진 유전독소 데이터세트, 예컨대 데이터베이스에서 유전독소 프로파일 기록에 저장된 것과 비교(블록 2114)할 수 있다. 선택적으로, 유전독소 모듈은, 부분적으로 비교 정보에 기초하여, 유전독소의 그럴듯한 작용 기전을 결정(블록 2116)할 수 있다. 다음에, 유전독소 모듈은 데이터베이스에서 샘플-특정 데이터세트에 저장될 수 있는 유전독성 데이터를 제공(블록 2118)할 수 있다. 도시되지 않은 일부 실시형태에서, 유전독성 데이터는 미래의 비교 활동을 위해 데이터베이스에 저장되는 유전독소 프로파일을 생성하기 위해 사용될 수 있다. 이후, 루틴 2100은 블록 2120에 계속 이어질 수 있고, 여기서 이것은 종료한다.

[0283] 도 22는 샘플의 유전독성 노출로부터 생긴 DNA 손상 사건을 검출하고 확인하기 위한 루틴 2200을 예시하는 흐름 다이어그램이다. 이 루틴은 도 20의 컴퓨팅 장치에 의해 호출될 수 있다. 루틴 2200은 도 20의 블록 2014에서 시작하고, 결정 블록 2202에서, 루틴 2200은 비상보성의 뉴클레오타이드 위치가 공정 오류인지를 결정한다. 다양한 실시형태에서, 원래의 DNA 분자의 가닥 둘 다의 서열 리드들 사이의 비동의의 위치인지를 결정하기 위한 매개변수는 운영자, 알려진 DNA 손상 특징, 알려진 공정 오류 특징, 미스매치가 표시되는 최소 서열 리드 수 등에 의해 규정될 수 있다.

[0284] 뉴클레오타이드 위치가 (DNA 추출 전에 생체내 DNA 손상 부위와 반대로) 프로세스 오류인 것으로 결정되는 경우, DS 모듈은 이러한 비상보성 뉴클레오타이드 위치를 제거하거나 무시(블록 2204)할 수 있다. 루틴 2200은 도 20의 블록 2016에 계속 이어질 수 있다.

[0285] 다시 결정 블록 2202를 참조하면, 뉴클레오타이드 위치가 프로세스 오류가 아닌 것으로 결정되는 경우, 유전독소 모듈은 예컨대 유전독소에 대한 노출로부터 생긴 가능한 생체내 DNA 손상 부위로서 이러한 비상보성 위치를 확인(블록 2206)할 수 있다. 확인 후에, 유전독소 모듈은 데이터베이스에서 샘플-특정 데이터세트와 연관된 DNA 손상 기록을 생성(블록 2208)할 수 있다. 일부 실시형태에서, DNA 손상 기록은 잠재적인 유전독소의 작용 기전을 추론하도록 사용될 수 있다(비도시). 루틴 2200은 도 20의 블록 2016에 계속 이어질 수 있다.

[0286] 도 23은 대상체에서 발암물질 또는 발암물질 노출을 검출하고 확인하기 위한 루틴 2300을 예시하는 흐름 다이어

그램이다. 루틴 2300은 도 20의 컴퓨팅 장치에 의해 호출될 수 있다. 루틴 2300은 블록 2302에서 시작하고, 유전독소 모듈은 (예를 들어, 블록 2016 후에) 도 20으로부터의 듀플렉스 시퀀싱 데이터 및 선택적으로 (예를 들어, 블록 2116 후에) 도 21로부터의 유전독성 데이터를 수신하고, 샘플이 유전독소에 노출된다는 것을 확증(블록 2304)한다. 다음에, 유전독소 모듈은 표적 게놈 영역(예를 들어, 유전자)의 서열에서 변이체를 확인(블록 2306)한다. 예를 들어, 유전독소 모듈은 특정 유전자 유전좌위(예를 들어, 암 유발자 유전자, 암유전자 등)에서 듀플렉스 시퀀싱 데이터 및 유전독성 데이터를 분석할 수 있다. 이후, 유전독소 모듈은 변이체 대립유전자 빈도(VAF)를 계산(블록 2308)한다.

[0287] 결정 블록 2310에서, 루틴 2300은 VAF가 대조군보다 시험 그룹에서 더 높은지를 결정한다. 시험 그룹의 VAF가 대조군보다 더 높지 않은 경우, 유전독소 모듈은 그 물질을 발암물질이라는 의심이 감소함을 표지(블록 2312)한다. 이후, 루틴 2300은 블록 2314에 계속 이어질 수 있고, 여기서 이것은 종료한다. VAF가 대조군보다 시험 그룹에서 더 높은 경우, 루틴 2300은 결정 블록 2316에 계속 이어질 수 있고, 여기서 루틴 2300은 돌연변이가 비일중항인지를 결정한다.

[0288] 돌연변이가 일중항인 경우, 유전독소 모듈은 그 물질을 발암물질이라는 중간 의심 수준으로 규명(블록 2318)한다. 돌연변이가 비일중항(즉, 다중항)인 것으로 결정되는 경우, 이 루틴은 결정 블록 2320에 계속 이어질 수 있고, 여기서 루틴 2300은 변이체가 표적 유전자에서 검출되는지 및 변이체가 유발자 돌연변이와 일치하는지(예를 들어, 암 성장/형질전환을 유발하는 것으로 알려진 돌연변이)를 결정한다.

[0289] 돌연변이가 유발자 돌연변이가 아닌 경우, 유전독소 모듈은 그 물질을 발암물질이라는 중간 의심 수준으로 규명(블록 2318)한다. 변이체(들)가 유발자 돌연변이와 일치하는 경우, 유전독소 모듈은 그 물질을 발암물질이라는 높은 의심 수준으로 규명(블록 2322)한다.

[0290] (블록 2318에서) 중간 의심 수준 또는 (블록 2318에서) 높은 의심 수준을 특징으로 하는 물질에 대해, 유전독소 모듈은 발암물질에 대한 안전성 역치를 평가하고/하거나 대상체에서의 노출 후에 유전독소 연관된 질병 또는 장애의 발생과 연관된 위험을 결정(블록 2324)할 수 있다. 이후, 루틴 2300은 블록 2314에 계속 이어질 수 있고, 여기서 이것은 종료한다.

[0291] 다른 단계 및 루틴은 또한 본 기술내용에 의해 고려된다. 예를 들어, 이 시스템 (예를 들어, 유전독소 모듈 또는 다른 모듈)은 대상체가 유전독소에 노출되는지, 시험 물질/인자가 유전독성인지를 결정하기 위해, 어떤 특징하에 유전독소가 돌연변이성 또는 발암성인지 등을 결정하기 위해 유전독소 데이터를 분석하도록 구성될 수 있다. 다른 단계는 특정 대상체의 생물학적 샘플로부터 도출된 유전독소 데이터에 기초하여 대상체가 예방학적으로 또는 치료학적으로 치료되어야 하는지를 결정하는 것을 포함할 수 있다. 예를 들어, 유전독소(들)가 이 시스템을 사용하여 확인되면, 서버는 대상체가 유전독소의 안전한 역치 수준 초과에 노출되는지를 결정할 수 있다. 그렇다면 이후 예방학적 또는 억제제 질병 치료가 개시될 수 있다.

[0292] 추가 실시예

[0293] 1. 돌연변이원에 대한 대상체의 노출 후에 대상체에서 생체내 발생한 게놈 돌연변이를 검출하고 정량화하는 방법으로서,

[0294] 대상체로부터의 샘플을 제공하는 단계이되, 샘플은 이중-가닥 DNA 분자를 포함하는 상기 제공하는 단계;

[0295] 샘플에서 복수의 이중-가닥 DNA 분자의 각각에 대해 오류-보정된 서열 리드를 생성하는 단계이되,

[0296] 어댑터-DNA 분자의 원래의 제1 가닥의 카피의 세트 및 어댑터-DNA 분자의 원래의 제2 가닥의 카피의 세트를 생성하는 단계;

[0297] 원래의 제1 가닥의 카피의 세트 및 원래의 제2 가닥의 카피의 세트를 시퀀싱하여 제1 가닥 서열 및 제2 가닥 서열을 제공하는 단계; 및

[0298] 제1 가닥 서열과 제2 가닥 서열을 비교하여 제1 가닥 서열과 제2 가닥 서열 사이의 하나 이상의 관련성을 확인하는 단계를 포함하는 상기 생성하는 단계; 및

[0299] 하나 이상의 관련성을 분석하여 샘플에서의 이중-가닥 DNA 분자에 대한 돌연변이 스펙트럼을 결정하는 단계를 포함하는 방법.

[0300] 2. 예 1에 있어서, 시퀀싱된 듀플렉스 염기-쌍마다 고유한 돌연변이의 수를 계산함으로써 표적 이중-가닥 DNA 분자에 대한 돌연변이체 빈도를 계산하는 단계를 추가로 포함하는, 방법.

[0301] 3. 예 1에 있어서, 표적 이중-가닥 DNA 분자는 대상체의 간, 비장, 혈액, 폐 또는 골수로부터 추출되는, 방법.

[0302] 4. 예 1에 있어서, 대상체는 표적 이중-가닥 DNA 분자가 대상체로부터 제거되기 30일 이하 전에 돌연변이원에 노출되는, 방법.

[0303] 5. 예 1에 있어서, 돌연변이 스펙트럼은 비지도된 계층적 돌연변이 스펙트럼 클러스터링에 의해 생성되는, 방법.

[0304] 6. 예 1에 있어서, 돌연변이 스펙트럼은 삼중항 돌연변이 스펙트럼인, 방법.

[0305] 7. 예 1에 있어서, 복수의 이중-가닥 DNA 분자의 각각에 대해 오류-보정된 서열 리드를 생성하는 단계는 하나 이상의 표적화된 게놈 영역의 오류-보정된 서열 리드를 생성하는 것을 포함하는, 방법.

[0306] 8. 예 7에 있어서, 하나 이상의 표적화된 게놈 영역은 게놈에서의 돌연변이-경향성 부위인, 방법.

[0307] 9. 예 7에 있어서, 하나 이상의 표적화된 게놈 영역은 알려진 암 유발자 유전자인, 방법.

[0308] 10. 예 1에 있어서, 대상체는 형질전환 동물이고, 표적 이중-가닥 DNA 분자의 적어도 일부는 전이유전자의 하나 이상의 부분을 포함하는, 방법.

[0309] 11. 예 1에 있어서, 대상체는 비형질전환 동물이고, 표적 이중-가닥 DNA 분자는 내인성 게놈 영역을 포함하는, 방법.

[0310] 12. 예 1에 있어서, 대상체는 인간이고, 표적 이중-가닥 DNA 분자는 인간에서 채혈된 혈액으로부터 추출되는, 방법.

[0311] 13. 시험 물질의 돌연변이성 서명을 생성하는 방법으로서,

[0312] 시험 물질에 노출된 시험 대상체로부터 추출된 DNA 단편을 듀플렉스 시퀀싱하는 단계; 및

[0313] 시험 물질의 돌연변이성 서명을 생성하는 단계이되,

[0314] 시퀀싱된 듀플렉스 염기-쌍마다 고유한 돌연변이의 수를 계산함으로써 복수의 DNA 단편에 대한 돌연변이체 빈도를 계산하는 단계; 및

[0315] 복수의 DNA 단편에 대한 돌연변이 패턴을 결정하는 단계이되, 돌연변이 패턴은 돌연변이 유형, 돌연변이 트리뉴클레오타이드 상황 및 돌연변이의 게놈 분포를 포함하는 결정하는 단계를 포함하는 상기 생성하는 단계를 포함하는, 방법.

[0316] 14. 예 13에 있어서, 시험 물질의 돌연변이 서명을 하나 이상의 알려진 유전독소의 돌연변이 서명과 비교하는 단계를 추가로 포함하는, 방법.

[0317] 15. 예 13에 있어서, 시험 물질의 돌연변이 서명은 조직 유형, 시험 물질에 대한 노출의 수준, 게놈 영역 및 대상체 유형 중 하나 이상에 기초하여 변하는, 방법.

[0318] 16. 예 15에 있어서, 대상체 유형은 배양물에서 성장한 인간 세포인, 방법.

[0319] 17. 예 13에 있어서, 시험 동물은 동물이 희생되기 30일 이하 전에 시험 화합물에 노출되는, 방법.

[0320] 18. 예 13에 있어서, 돌연변이성 서명은 컴퓨터를 사용한 패턴 매칭에 의해 생성되는, 방법.

[0321] 19. 예 13에 있어서, 돌연변이 서명은 삼중항 돌연변이 서명인, 방법.

[0322] 20. 예 13에 있어서, DNA 단편의 듀플렉스 시퀀싱은 하나 이상의 표적화된 게놈 영역의 듀플렉스 시퀀싱을 포함하는, 방법.

[0323] 21. 예 20에 있어서, 하나 이상의 표적화된 게놈 영역은 게놈에서의 돌연변이-경향성 부위인, 방법.

[0324] 22. 예 20에 있어서, 하나 이상의 표적화된 게놈 영역은 알려진 암 유발자 유전자인, 방법.

[0325] 23. 예 13에 있어서, 시험 동물은 형질전환 동물이고, DNA 단편의 적어도 일부는 전이유전자의 하나 이상의 부분을 포함하는, 방법.

[0326] 24. 예 13에 있어서, 시험 동물은 비형질전환 동물이고, DNA 단편은 내인성 게놈 영역을 포함하는, 방법.

[0327] 25. 시험 물질의 유전독성 가능성을 평가하는 방법으로서,

[0328] (a) 시험 물질에 노출된 생물학적 소스로부터 복수의 이중-가닥 DNA 단편을 포함하는 샘플로부터 시퀀싱 라이브러리를 제조하는 단계이되, 서열 라이브러리의 제조는 비대칭적 어댑터 분자를 복수의 이중-가닥 DNA 단편에 결찰하여 복수의 어댑터-DNA 분자를 생성하는 것을 포함하는 상기 제조하는 단계;

[0329] (b) 어댑터-DNA 분자의 제1 가닥 및 제2 가닥을 시퀀싱하여 각각의 어댑터-DNA 분자에 대한 제1 가닥 서열 리드 및 제2 가닥 서열 리드를 제공하는 단계;

[0330] (c) 각각의 어댑터-DNA 분자에 대해, 제1 가닥 서열 리드와 제2 가닥 서열 리드를 비교하여 제1 가닥 서열 리드와 제2 가닥 서열 리드 사이의 하나 이상의 관련성을 확인하는 단계; 및

[0331] (d) 각각의 어댑터-DNA 분자에 대해 제1 가닥 서열 리드와 제2 가닥 서열 리드 사이의 하나 이상의 관련성을 확인함으로써 시험 물질의 돌연변이 서명을 결정하여 샘플에서 돌연변이 패턴, 돌연변이 유형, 돌연변이체 빈도, 돌연변이 유형 분포 및 돌연변이의 게놈 분포 중 적어도 하나를 결정하는 단계; 및

[0332] (e) 시험 물질의 돌연변이 서명을 알려진 유전독소로부터 유래된 복수의 돌연변이 스펙트럼과 비교하여 돌연변이 서명이 알려진 유전독소로부터의 돌연변이 스펙트럼과 충분히 유사한지를 결정하는 단계; 또는

[0333] (f) 돌연변이체 빈도, 돌연변이 유형 또는 돌연변이 유형 분포 중 적어도 하나가 안전한 역치 수준보다 높은지를 평가하는 단계; 또는

[0334] (g) 돌연변이체 빈도가 안전한 역치 돌연변이체 빈도를 초과하는지를 결정하는 단계를 포함하는, 방법.

[0335] 26. 예 25에 있어서, 시험 물질의 돌연변이 서명은 안전 역치 빈도 초과의 돌연변이체 빈도를 포함하는, 방법.

[0336] 27. 예 25에 있어서, 시험 물질의 돌연변이 서명은 알려진 암-연관된 돌연변이 패턴과 충분히 유사한 돌연변이 패턴을 포함하는, 방법.

[0337] 28. 예 25에 있어서, 생물학적 소스는 배양물에서 성장한 세포, 동물, 인간, 인간 세포주, 형질전환 동물, 비형질전환 동물, 인간 조직 샘플 또는 인간 혈액 샘플 중 적어도 하나인, 방법.

[0338] 29. 예 25에 있어서, 생물학적 소스는 복수의 이중-가닥 DNA 단편을 포함하는 샘플을 추출하기 30일 이하 전에 시험 물질에 노출되는, 방법.

[0339] 30. 예 25에 있어서, 돌연변이 서명은 삼중항 돌연변이 서명인, 방법.

[0340] 31. 예 25에 있어서, 제1 가닥 서열 리드와 제2 가닥 서열 리드를 비교하기 전에, 상기 방법은 어댑터 서열, 서열 리드 길이 및 원래의 가닥 정보 중 하나 이상을 사용하여 제1 가닥 서열 리드를 제2 가닥 서열 리드와 연관시키는 단계를 포함하는, 방법.

[0341] 32. 예 25에 있어서, 시퀀싱 라이브러리를 제조하기 전에, 상기 방법은 생물학적 소스를 시험 물질에 노출시키는 단계를 추가로 포함하는, 방법.

[0342] 33. 예 32에 있어서, 생물학적 소스를 시험 물질에 노출시키기 전에, 생물학적 소스는 암 조직이거나 이를 포함하는, 방법.

[0343] 34. 예 32에 있어서, 생물학적 소스를 시험 물질에 노출시키기 전에, 생물학적 소스는 건강한 조직이거나 이를 포함하는, 방법.

[0344] 35. 예 25에 있어서, 샘플은 혈액 샘플이거나 이를 포함하는, 방법.

[0345] 36. 예 25에 있어서, 샘플은 암 세포주이거나 이를 포함하는, 방법.

[0346] 37. 예 25에 있어서, 생물학적 소스는 암성 세포를 포함하고, 물질은 암성 세포의 적어도 일부에 대해 선택적 유전독성에 대해 시험되는, 방법.

[0347] 38. 예 37에 있어서, 물질은 치료 화합물인, 방법.

[0348] 39. 예 38에 있어서, 치료 화합물의 선택적 유전독성에 민감한 것으로 나타난 암성 세포의 일부에 대해, 상기 방법은 치료 화합물에 대한 노출 전에 암성 세포의 일부에 대해 돌연변이체 빈도 및 돌연변이 스펙트럼 중 하나 이상을 결정하는 단계를 추가로 포함하는, 방법.

[0349] 40. 예 25에 있어서, 시험 물질은 식품, 약물, 백신, 화장용 물질, 산업용 첨가제, 산업 부산물, 석유 증류물, 중금속, 가정용 세척제, 공기 매개 미립자, 제조 부산물, 오염물질, 가소제, 세제, 방사선-방출 생성물, 담배 제품, 화학 물질 또는 생물학적 물질을 포함하는, 방법.

[0350] 41. 유전독성 물질에 대한 대상체의 노출을 결정하는 방법으로서,

[0351] 대상체의 DNA 돌연변이 스펙트럼을 알려진 돌연변이성 화합물의 돌연변이 스펙트럼과 비교하는 단계; 및

[0352] 대상체의 DNA 돌연변이 스펙트럼과 가장 유사한 알려진 돌연변이성 화합물의 돌연변이 스펙트럼을 확인하는 단계를 포함하는, 방법.

[0353] 42. 예 41에 있어서, 대상체의 DNA 돌연변이 스펙트럼은 듀플렉스 시퀀싱에 의해 평가되는, 방법.

[0354] 43. 예 41에 있어서, 대상체의 DNA 돌연변이 스펙트럼은 환자의 혈액으로부터 추출된 DNA로부터 생성되는, 방법.

[0355] 44. 예 41에 있어서, 대상체의 DNA 돌연변이 스펙트럼은 삼중항 돌연변이 스펙트럼인, 방법.

[0356] 45. 예 41에 있어서, 대상체의 DNA를 시퀀싱하여 대상체의 DNA 돌연변이 스펙트럼을 생성하는 단계를 추가로 포함하는, 방법.

[0357] 46. 예 45에 있어서, 대상체의 DNA의 시퀀싱은 하나 이상의 알려진 암 유발자 유전자의 시퀀싱을 포함하는, 방법.

[0358] 47. 유전독소를 확인하기 위해 이중 가닥 폴리뉴클레오타이드의 오류 보정된 듀플렉스 시퀀싱에 사용될 수 있는 키트로서,

[0359] 중합효소 연쇄 반응(PCR) 프라이머의 적어도 하나의 세트 및 어댑터 분자의 적어도 하나의 세트이되, 프라이머 및 어댑터 분자는 오류 보정된 듀플렉스 시퀀싱 실험에 사용될 수 있는 상기 프라이머 및 어댑터 분자; 및

[0360] 대상체가 적어도 하나의 유전독소에 노출되었는지를 확인하기 위해 대상체의 샘플로부터 추출된 DNA의 오류 보정된 듀플렉스 시퀀싱을 수행하는 데 있어서 키트를 사용하는 방법에 대한 명령을 포함하는, 키트.

[0361] 48. 예 47에 있어서, 시약은 DNA 복구 효소를 포함하는, 키트.

[0362] 49. 예 47에 있어서, 어댑터 분자의 세트에서의 각각의 어댑터 분자는 적어도 하나의 단일 분자 식별자(SMI) 서열 및 적어도 하나의 가닥 한정 요소를 포함하는, 키트.

[0363] 50. 예 47에 있어서, 컴퓨터에서 실행될 때, 샘플에서 하나 이상의 이중-가닥 DNA 분자에 대한 오류-보정된 듀플렉스 시퀀싱 리드를 결정하는 단계 및 오류-보정된 듀플렉스 시퀀싱 리드를 사용하여 적어도 하나의 유전독소의 돌연변이체 빈도, 돌연변이 스펙트럼, 및/또는 삼중항 스펙트럼을 결정하는 단계를 수행하는 비일시적 컴퓨터 판독 가능한 매체에 구현된 컴퓨터 프로그램 제품을 추가로 포함하는, 키트.

[0364] 51. 예 50에 있어서, 컴퓨터 프로그램 제품은 대상체의 DNA를 돌연변이시키는 데 있어서의 유전독소의 작용 기전; 및 유전독소 작용 기전에 기초하여 대상체에게 투여하기에 적합한 치료학적 치료 또는 예방학적 치료를 추가로 결정하는, 키트.

[0365] 52. 유전독소에 노출된 대상체를 진단하고 치료하는 방법으로서,

[0366] a) 대상체가

[0367] ⅰ) 대상체로부터 생물학적 샘플을 수득하는 것;

[0368] ⅱ) 샘플로부터 추출된 복수의 이중 가닥 DNA 서열에 대한 듀플렉스 오류 보정된 시퀀싱 리드를 제공하는 것;

[0369] ⅲ) DNA 서열의 돌연변이체 빈도, 돌연변이 스펙트럼, 및/또는 삼중항 돌연변이 스펙트럼을 결정하는 것;

[0370] ⅳ) 돌연변이체 빈도, 돌연변이 스펙트럼 및/또는 삼중항 돌연변이 스펙트럼이 유전독소에 노출되었던 대상체를 나타내는지를 결정하는 것

[0371] 에 의해 유전독소에 노출되었는지를 결정하는 단계;

[0372] b) 대상체가 유전독소에 노출되었던 경우, 유전독소와 연관된 질병 또는 장애의 발생을 예방하거나 억제하기 위한 예방학적 치료 및/또는 치료학적 치료를 제공하는 단계를 포함하는, 방법.

[0373] 53. 유전독소에 대한 안전한 노출의 역치 수준을 확인하고 치료를 제공하는 방법으로서,

[0374] a) 안전한 노출의 유전독소의 역치 수준을 결정하는 단계;

[0375] b) 대상체가

[0376] ⅰ) 대상체로부터 생물학적 샘플을 수득하는 것;

[0377] ii) 생물학적 샘플로부터 추출된 복수의 이중 가닥 DNA 서열에 대한 듀플렉스 오류 보정된 시퀀싱 리드를 제공하는 것;

[0378] iii) DNA 서열의 돌연변이체 빈도, 돌연변이 스펙트럼, 및/또는 삼중항 돌연변이 스펙트럼을 결정하는 것;

[0379] iv) 돌연변이체 빈도, 돌연변이 스펙트럼 및/또는 삼중항 돌연변이 스펙트럼이 특정 유전독소에 노출되었던 대상체를 나타내는지를 결정하는 것;

[0380] ⅴ) 돌연변이체 빈도, 돌연변이 스펙트럼 및/또는 삼중항 돌연변이 스펙트럼에 기초하여 유전독소에 대한 대상체의 노출의 수준을 컴퓨팅하는 것

[0381] 에 의해 안전한 노출의 역치 수준보다 높은 수준에서 유전독소에 노출되었는지를 결정하는 단계; 및

[0382] c) 대상체가 안전한 노출의 유전독소의 역치 수준 초과에 노출되었던 경우, 유전독소와 연관된 질병 또는 장애의 발생을 예방하거나 억제하기 위한 예방학적 치료 및/또는 치료학적 치료를 제공하는 단계를 포함하는, 방법.

[0383] 54. 샘플의 유전독성 노출로부터 생긴 돌연변이성 사건 및/또는 핵산 손상 사건을 검출하고 확인하기 위한 시스템으로서,

[0384] 시퀀싱 데이터 및 유전독성 데이터와 관련된 정보를 전송하기 위한 컴퓨터 네트워크이되, 정보는 원시 시퀀싱 데이터, 듀플렉스 시퀀싱 데이터, 샘플 정보 및 유전독소 정보 중 하나 이상을 포함하는 컴퓨터 네트워크;

[0385] 하나 이상의 사용자 컴퓨팅 장치와 연관되고 컴퓨터 네트워크와 통신하는 클라이언트 컴퓨터;

[0386] 복수의 유전독소 프로파일 및 사용자 결과 기록을 저장하기 위한 컴퓨터 네트워크에 연결된 데이터베이스;

[0387] 컴퓨터 네트워크와 통신하고, 원시 시퀀싱 데이터 및 듀플렉스 시퀀싱 데이터를 생성하기 위한 클라이언트 컴퓨터로부터의 요청, 원래의 이중-가닥 핵산 분자를 나타내는 패밀리로부터의 그룹 서열 리드를 수신하고, 듀플렉스 시퀀싱 데이터를 생성하기 위해 개별 가닥으로부터의 대표적인 서열을 서로 비교하도록 구성된 듀플렉스 시퀀싱 모듈; 및

[0388] 컴퓨터 네트워크와 통신하고, 듀플렉스 시퀀싱 데이터를 기준 서열 정보와 비교하여 돌연변이를 확인하고, 돌연변이체 빈도, 돌연변이 스펙트럼 및 삼중항 돌연변이 스펙트럼 중 적어도 하나를 포함하는 유전독소 데이터를 생성하도록 구성된 유전독소 모듈을 포함하는, 시스템.

[0389] 55. 예 54에 있어서, 유전독소 프로파일은 복수의 알려진 유전독소로부터의 유전독소 돌연변이 스펙트럼을 포함하는, 시스템.

[0390] 56. 비일시적 컴퓨터 판독 가능한 저장 매체로서,

[0391] 하나 이상의 프로세서에 의해 실행될 때, 대상체가 적어도 하나의 유전독소에 노출되는지를 결정하고/하거나, 적어도 하나의 유전독소의 정체를 결정하기 위한 예 1 내지 53 중 어느 하나의 방법을 수행하는 명령을 포함하는, 비일시적 컴퓨터 판독 가능한 저장 매체.

[0392] 57. 예 56에 있어서, 적어도 하나의 유전독소의 정체가 결정되는 검출된 물질의 돌연변이 스펙트럼, 돌연변이체 빈도, 및/또는 삼중항 돌연변이 스펙트럼을 컴퓨팅하는 것을 추가로 포함하는, 비일시적 컴퓨터 판독 가능한 저장 매체.

[0393] 58. 컴퓨터 시스템으로서,

[0394] 대상체가 적어도 하나의 유전독소에 노출되는지 및/또는 이의 정체를 결정하기 위한 예 1 내지 53 중 어느 하나의 방법을 수행하기 위한 것이되, 상기 시스템은 프로세서, 메모리, 데이터베이스 및 프로세서(들)에 대한 명령을 포함하는 비일시적 컴퓨터 판독 가능한 저장 매체를 갖는 적어도 하나의 컴퓨터를 포함하고, 상기 프로세서(들)는 예 1 내지 53 중 어느 하나의 방법을 포함하는 연산을 수행하기 위해 상기 명령을 실행하도록 구성된, 시스템.

[0395] 59. 예 58에 있어서,

[0396] a. 유선 네트워크 또는 무선 네트워크;

[0397] b. 대상체의 샘플의 폴리뉴클레오타이드 서열을 추출하고 증폭시키고 제조하기 위해, 그리고 폴리뉴클레오타이드 서열을 네트워크를 통해 원격 서버에 전송하기 위해 시약을 포함하는 키트의 사용으로부터 도출된 데이터를 수신할 수 있는 복수의 사용자 전자 컴퓨팅 장치; 및

[0398] c. 프로세서, 메모리, 데이터베이스 및 프로세서(들)에 대한 명령을 포함하는 비일시적 컴퓨터 판독 가능한 저장 매체를 포함하는 원격 서버이되, 상기 프로세서(들)는 예 1 내지 53 중 어느 하나의 방법을 포함하는 연산을 수행하기 위해 상기 명령을 실행하도록 구성된 원격 서버

[0399] 를 포함하는 네트워크 컴퓨터 시스템을 추가로 포함하고;

[0400] d. 상기 원격 서버는 샘플의 유전독성 노출로부터 생긴 돌연변이성 사건 및/또는 핵산 손상 사건을 검출하고 확인할 수 있는, 시스템.

[0401] 60. 예 59에 있어서, 네트워크를 통해 접근 가능한 데이터베이스 및/또는 제3자 데이터베이스는 알려진 유전독소의 유전독소 프로파일, 적어도 하나의 대상체의 샘플의 유전독소 프로파일 중 하나 이상을 포함하는 복수의 기록을 추가로 포함하고, 유전독소 프로파일은 돌연변이 또는 DNA 손상 부위를 포함하는, 시스템.

[0402] 61. 비일시적 컴퓨터 판독 가능한 매체로서,

[0403] 이 매체의 컨텐츠는 적어도 하나의 컴퓨터가 유전독성 스크리닝 검정으로부터 샘플에서 이중-가닥 핵산 분자에 대한 듀플렉스 시퀀싱 데이터를 제공하는 방법을 수행하게 하고, 상기 방법은

[0404] 사용자 컴퓨팅 장치로부터 원시 서열 데이터를 수신하는 단계; 및

[0405] 샘플에서 복수의 핵산 분자로부터 도출된 복수의 원시 서열 리드를 포함하는 샘플-특정 데이터세트를 생성하는 단계;

[0406] 원래의 이중-가닥 핵산 분자를 나타내는 패밀리로부터 서열 리드를 그룹화하는 단계이되, 그룹화는 공유된 단일 분자 식별자 서열에 기초하는 상기 그룹화하는 단계;

[0407] 원래의 이중-가닥 핵산 분자로부터 제1 가닥 서열 리드와 제2 가닥 서열 리드를 비교하여 제1 가닥 서열 리드와 제2 가닥 서열 리드 사이의 하나 이상의 관련성을 확인하는 단계; 및

[0408] 샘플에서 이중-가닥 핵산 분자에 대한 듀플렉스 시퀀싱 데이터를 제공하는 단계를 포함하는, 비일시적 컴퓨터 판독 가능한 매체.

[0409] 62. 예 58에 있어서, 비교된 제1 서열 리드와 제2 서열 리드 사이에 비상보성의 뉴클레오타이드 위치를 확인하는 것을 추가로 포함하고, 상기 방법은

[0410] 비상보성의 위치에서, 공정 오류를 확인하고 제거하거나 무시하는 단계; 및

[0411] 공정 오류로서 확인되지 않은 비상보성의 위치에서, 유전독소에 대한 노출로부터 생긴 가능한 생체내 DNA 손상 부위로서 비상보성의 남은 위치를 확인하는 단계를 추가로 포함하는, 컴퓨터 판독 가능한 매체.

[0412] 63. 비일시적 컴퓨터 판독 가능한 매체로서,

[0413] 이 매체의 컨텐츠는 적어도 하나의 컴퓨터가 샘플의 유전독성 노출로부터 생긴 돌연변이성 사건을 검출하가 확인하는 방법을 수행하게 하고, 상기 방법은

[0414] 듀플렉스 서열 데이터를 기준 서열 정보와 비교하는 단계;

[0415] 듀플렉스 서열 데이터에서 돌연변이를 확인하는 단계이되, 돌연변이는 기준 정보와 비동의의 영역으로 확인되는 상기 확인하는 단계;

[0416] 듀플렉스 서열 데이터에서 돌연변이체 빈도를 결정하는 단계;

[0417] 듀플렉스 서열 데이터로부터 돌연변이 스펙트럼을 생성하는 단계;

[0418] 듀플렉스 서열 데이터로부터 삼중항 돌연변이 스펙트럼을 생성하는 단계; 및

[0419] 돌연변이 스펙트럼 및/또는 삼중항 돌연변이 스펙트럼을 복수의 알려진 유전독소 데이터세트와 비교하는 단계를

포함하는, 비일시적 컴퓨터 판독 가능한 매체.

[0420] 64. 비일시적 컴퓨터 판독 가능한 매체로서,

[0421] 이 매체의 컨텐츠는 적어도 하나의 컴퓨터가 대상체에서 발암물질 또는 발암물질 노출을 검출하가 확인하는 방법을 수행하게 하고, 상기 방법은

[0422] 대상체로부터 샘플로부터 생성된 듀플렉스 시퀀싱 데이터를 사용하여 표적 게놈 영역에서 서열 변이체를 확인하는 단계;

[0423] 시험 샘플 및 대조군 샘플의 변이체 대립유전자 빈도(VAF)를 계산하는 단계;

[0424] VAF가 대조군에서보다 시험 그룹에서 더 높은지를 결정하는 단계;

[0425] VAF가 더 높은 샘플에서, 서열 변이체가 비일중항인지를 결정하는 단계;

[0426] VAF가 더 높은 샘플에서, 서열 변이체가 유발자 돌연변이인지를 결정하는 단계; 및

[0427] 비일중항 및/또는 유발자 돌연변이를 갖는 샘플을 발암물질임에 의심되는 것으로 규명하는 단계를 포함하는, 비일시적 컴퓨터 판독 가능한 매체.

[0428] 65. 예 68에 있어서, 발암물질에 대한 안전성 역치를 평가하고/하거나, 대상체에서 노출 후에 유전독소 연관된 질병 또는 장애를 발생시키는 것과 연관된 위험을 결정하는 것을 추가로 포함하는, 비일시적 컴퓨터 판독 가능한 매체.

[0429] 참고문헌

[0430] 하기에 기술된 참고문헌, 및 상기 명세서에 인용된 특허 및 공개 특허 출원은 본원에서 완전히 제시된 것처럼 본원에 그 전문이 참조로 포함된다.

[0431]
[1] Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, and Loeb LA. Detection of ultra-rare mutations by next-generation sequencing. Proc Natl Acad Sci U S A. 2012; 109(36): 14508–14513.

[2] Kennedy SR, Salk JJ, Schmitt MW, Loeb LA. Ultra-Sensitive Sequencing Reveals an Age-Related Increase in Somatic Mitochondrial Mutations that are inconsistent with oxidative damage. PLOS Genetics. 2013; 9(9): 1-10.

[3] Kennedy SR, Schmitt MW, Fox EJ, Kohrn BF, Salk JJ, Ahn EH, et al. Detecting ultralow-frequency mutations by Duplex Sequencing. Nat Protoc. 2014; 9(11): 2586–2606.

[4] Schmitt MW, Fox EJ, Prindle MJ, Reid-Bayliss KS, True LD, et al. Sequencing small genomic targets with high efficiency and extreme accuracy. Nature Methods. 2015; 12(5): 423-5.

[5] Chan CY, Huang PH, Guo F, Ding X, Kapur V, Mai J D, et al. Accelerating drug discovery via organs-on-chips. Lab Chip. 2013; 12(24): 4697-4710.

[6] Schmitt MW, Loeb LA, and Salk JJ. The influence of subclonal resistance mutations on targeted cancer therapy. Nat Rev Clin Oncol. 2016; 13(6): 335–347.

[7] Salk JJ, Schmitt MW, Loeb L A. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. Nature Reviews Genetics. 2018. 19:269-283.

[0432] 결론

[0433] 본 기술내용의 실시형태의 상기 상세한 설명은 배타적이거나 본 기술내용을 상기 개시된 정확한 형태로 제한하도록 의도되지 않는다. 본 기술내용의 특정 실시형태 및 본 기술내용에 대한 예가 예시적인 목적을 위해 상기에 기재되어 있지만, 당업자가 인식하는 것처럼 본 기술내용의 범위 내에 다양한 동등한 변형이 가능하다. 예를 들어, 단계가 소정의 순서로 제시되지만, 대안적인 실시형태는 상이한 순서로 단계를 수행할 수 있다. 본원에 기재된 다양한 실시형태는 또한 추가의 실시형태를 제공하도록 조합될 수 있다. 본원에서 인용된 모든 참고문헌은 본원에 완전히 기재된 것처럼 참조로 포함된다.

[0434] 상기로부터, 본 기술내용의 특정 실시형태가 예시의 목적을 위해 본원에 기재되어 있지만, 잘 알려진 구조 및 기능이 본 기술내용의 실시형태의 설명을 불필요하게 모호하게 하는 것을 피하도록 자세히 도시되거나 기재되지

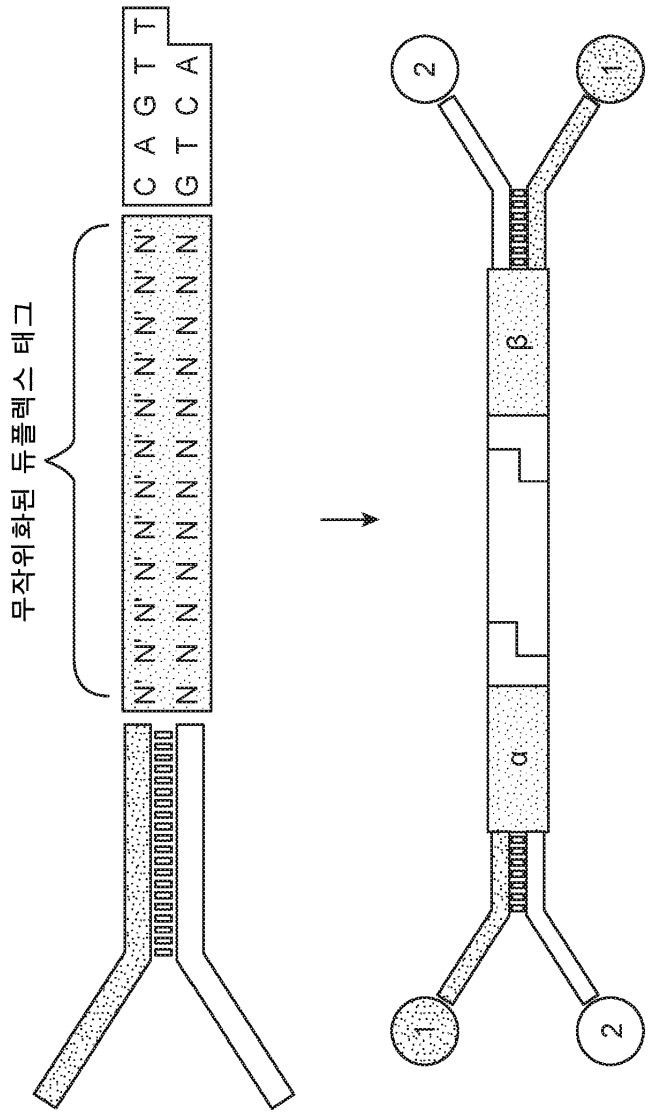
않는다고 이해될 것이다. 상황이 허용하는 경우, 단수 용어 또는 복수 용어는 또한 각각 복수 용어 또는 단수 용어를 포함할 수 있다.

[0435] 더구나, 단어 "또는"이 2개 이상의 항목의 목록과 관련하여 다른 항목을 배제한 단일 항목을 오직 의미하는 것으로 명확히 제한되지 않는 한, 이러한 목록에서의 "또는"의 사용은 (a) 목록에서의 임의의 단일 항목, (b) 목록에서의 모든 항목 또는 (c) 목록에서의 항목의 임의의 조합을 포함하는 것으로 해석되어야 한다. 추가적으로, 용어 "포함하는"은 임의의 더 많은 수의 동일한 특징 및/또는 추가 유형의 다른 특징이 불가능하지 않도록 적어도 인용된 특징(들)을 포함함을 의미하는 것으로 도처에 사용된다. 특정 실시형태가 예시의 목적을 위해 본원에 기재되어 있지만, 본 기술내용으로부터 벗어나지 않으면서 다양한 변형이 이루어질 수 있는 것으로 또한 이해될 것이다. 추가로, 본 기술내용의 소정의 실시형태와 연관된 이점이 이 실시형태의 상황에서 기재되어 있지만, 다른 실시형태가 또한 이러한 이점을 나타낼 수 있고, 모든 실시형태는 본 기술내용의 범위 내에 해당하는 이러한 이점을 반드시 나타낼 필요는 없다. 따라서, 본 개시내용 및 연관된 기술내용은 본원에 명확히 도시되거나 기재되지 않은 다른 실시형태를 포괄할 수 있다.

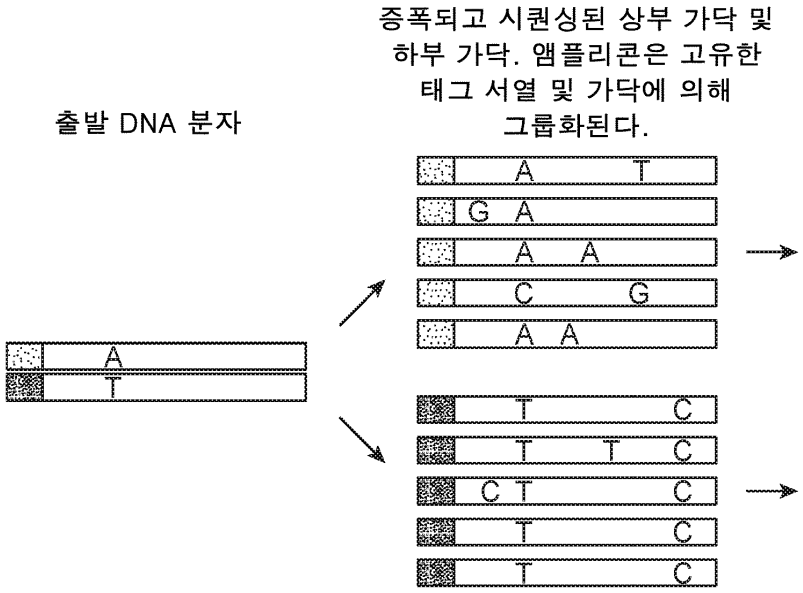
[0436] 본 개시내용에 사용된 제품 명칭은 오직 확인 목적을 위한 것이다. 모든 상표명은 이의 각각의 소유자의 재산이다.

도면

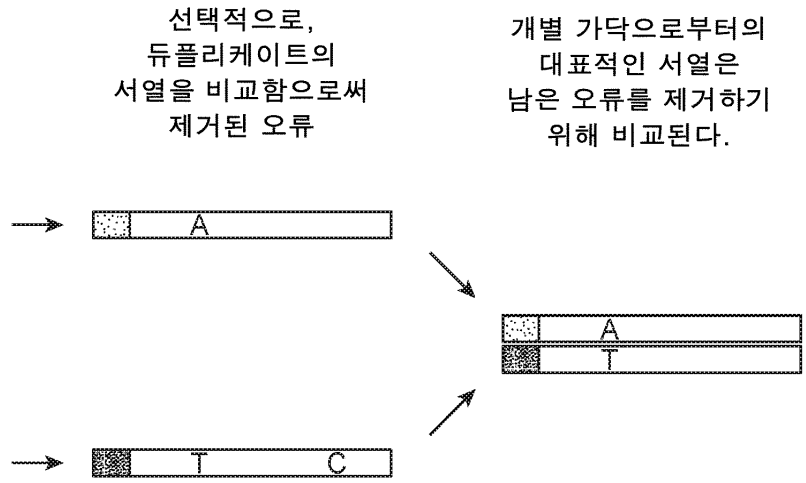
도면1a



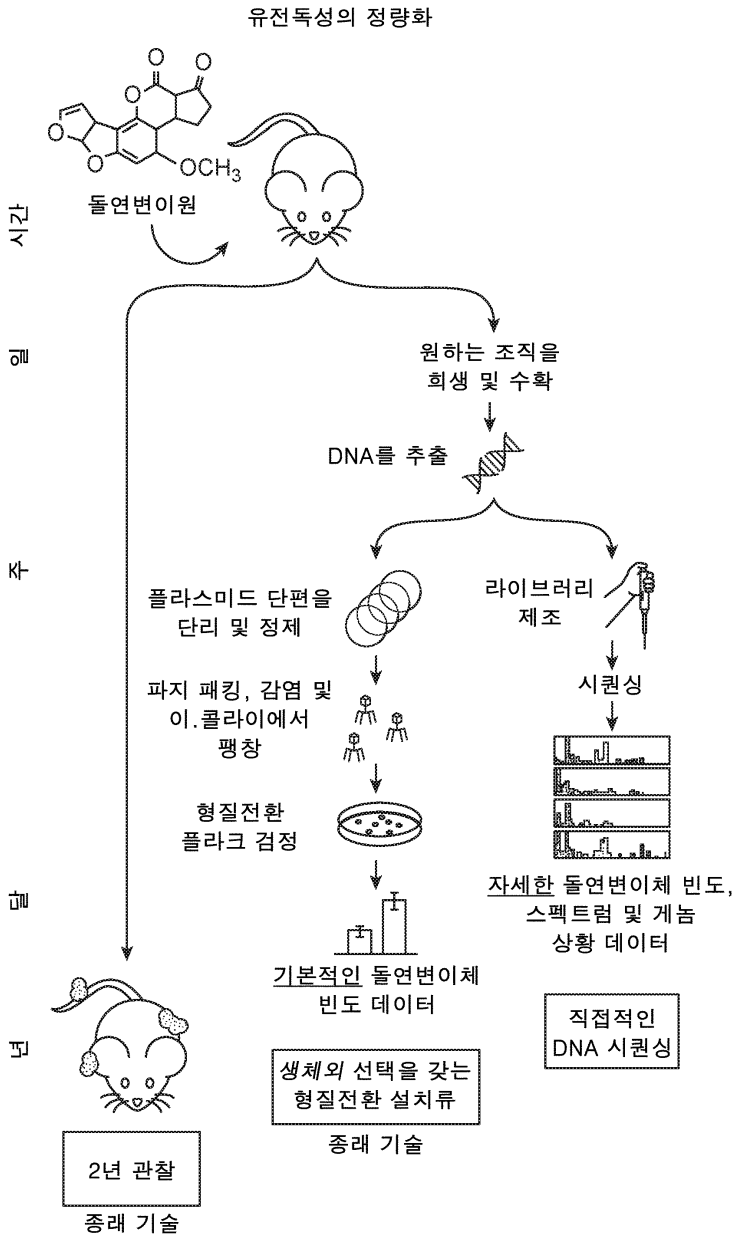
도면1b


출발 DNA 분자
증폭되고 시퀀싱된 상부 가닥 및 하부 가닥. 앰플리콘은 고유한 태그 서열 및 가닥에 의해 그룹화된다.
A
T
A T
G A
A A
C G
A A
T C
T T C
C T C
T C
T C

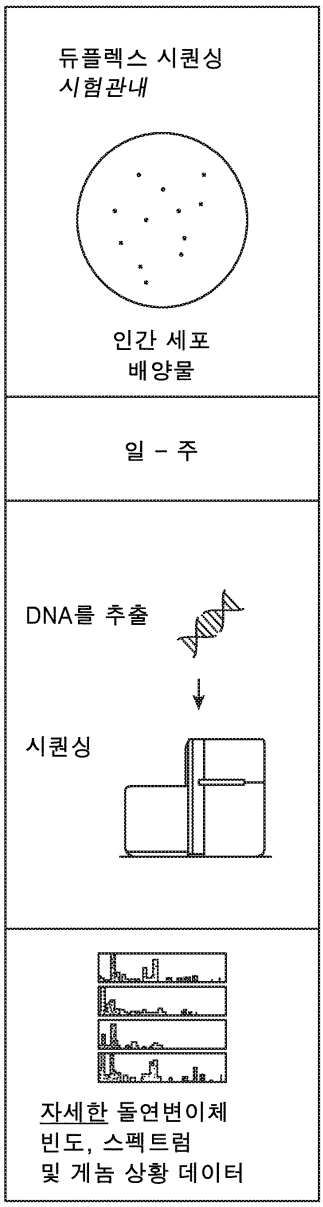
도면1c


선택적으로, 듀플리케이트의 서열을 비교함으로써 제거된 오류
개별 가닥으로부터의 대표적인 서열은 남은 오류를 제거하기 위해 비교된다.
A
T C
A
T

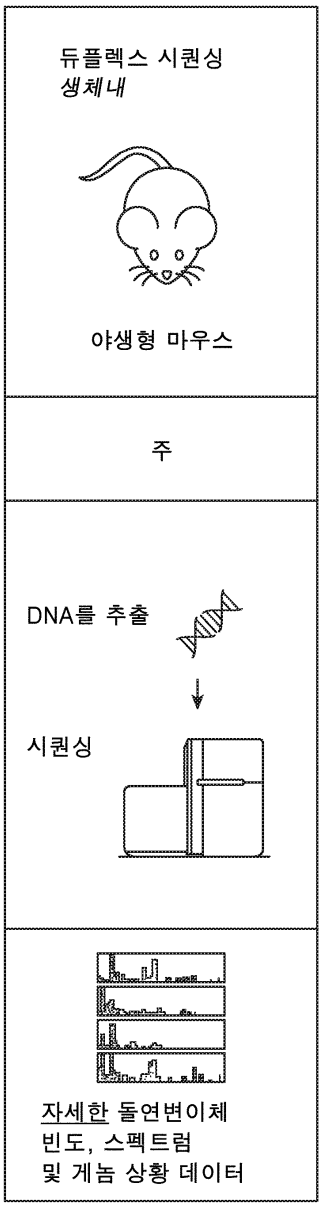
도면2a


유전독성의 정량화
시간
일
주
달
년
돌연변이원
원하는 조직을
희생 및 수확
DNA를 추출
플라스미드 단편을
단리 및 정제
파지 패킹, 감염 및
이.콜라이에서
팽창
형질전환
플라크 검정
기본적인 돌연변이체
빈도 데이터
생체외 선택을 갖는
형질전환 설치류
종래 기술
2년 관찰
종래 기술
라이브러리
제조
시퀀싱
자세한 돌연변이체 빈도,
스펙트럼 및 게놈
상황 데이터
직접적인
DNA 시퀀싱

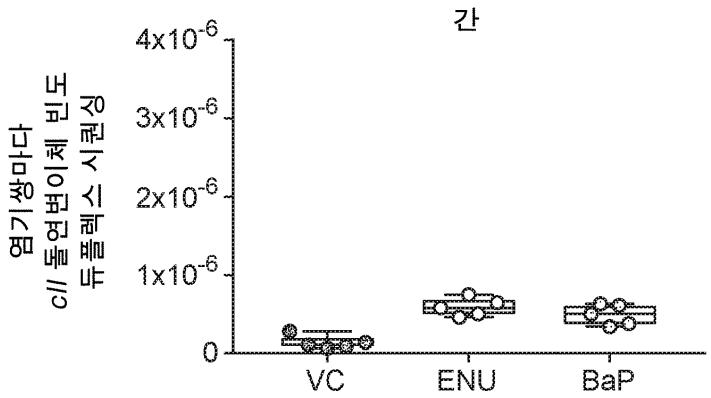
도면2b


듀플렉스 시퀀싱
시험관내
인간 세포
배양물
일 – 주
DNA를 추출
시퀀싱
자세한 돌연변이체
빈도, 스펙트럼
및 게놈 상황 데이터

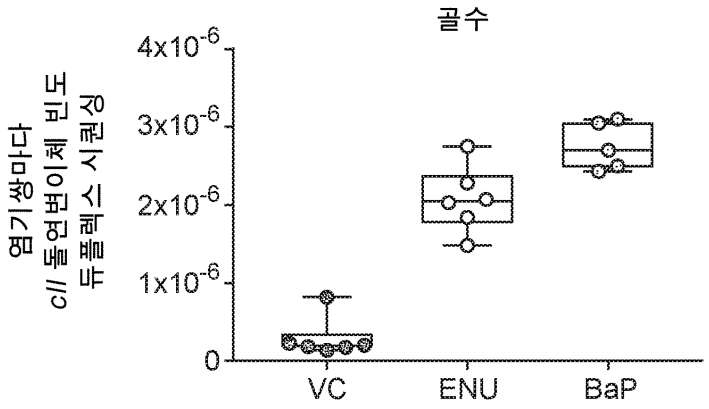
도면2c


듀플렉스 시퀀싱
생체내
야생형 마우스
주
DNA를 추출
시퀀싱
자세한 돌연변이체
빈도, 스펙트럼
및 게놈 상황 데이터

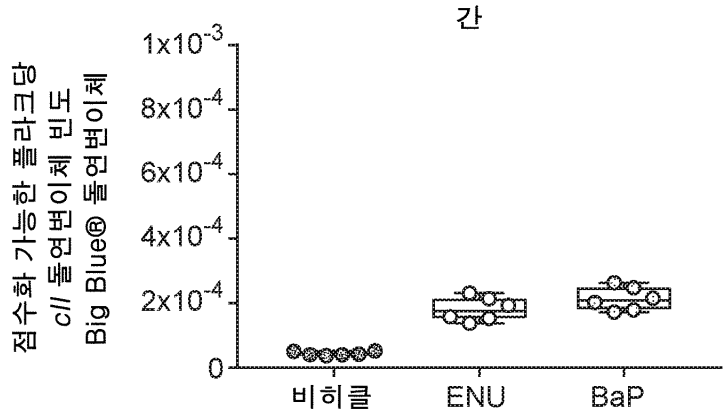
도면3a


간
염기쌍마다
cII 돌연변이체 빈도
듀플렉스 시퀀싱
4x10-6
3x10-6
2x10-6
1x10-6
0
VC
ENU
BaP

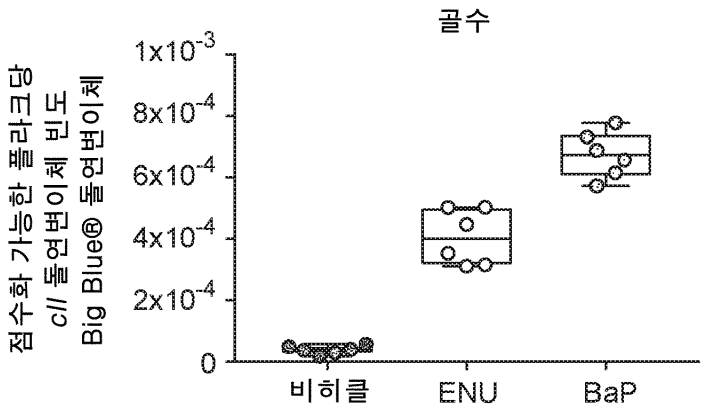
도면3b



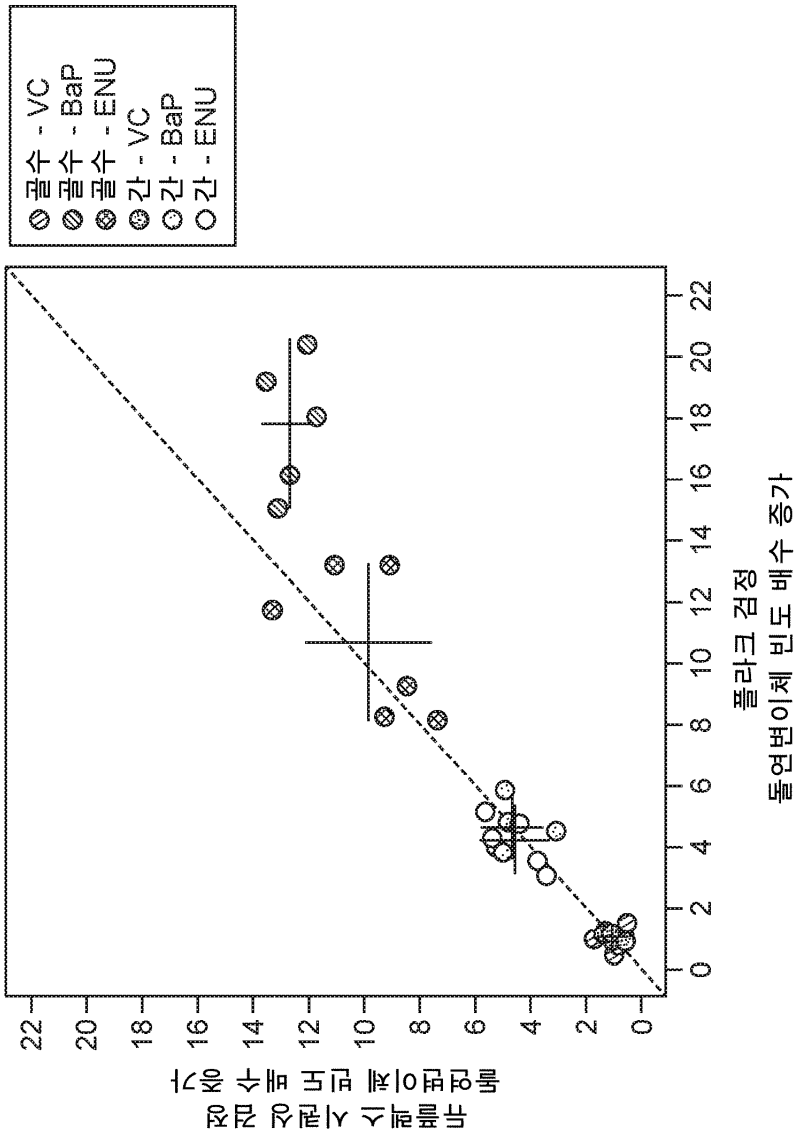
도면3c



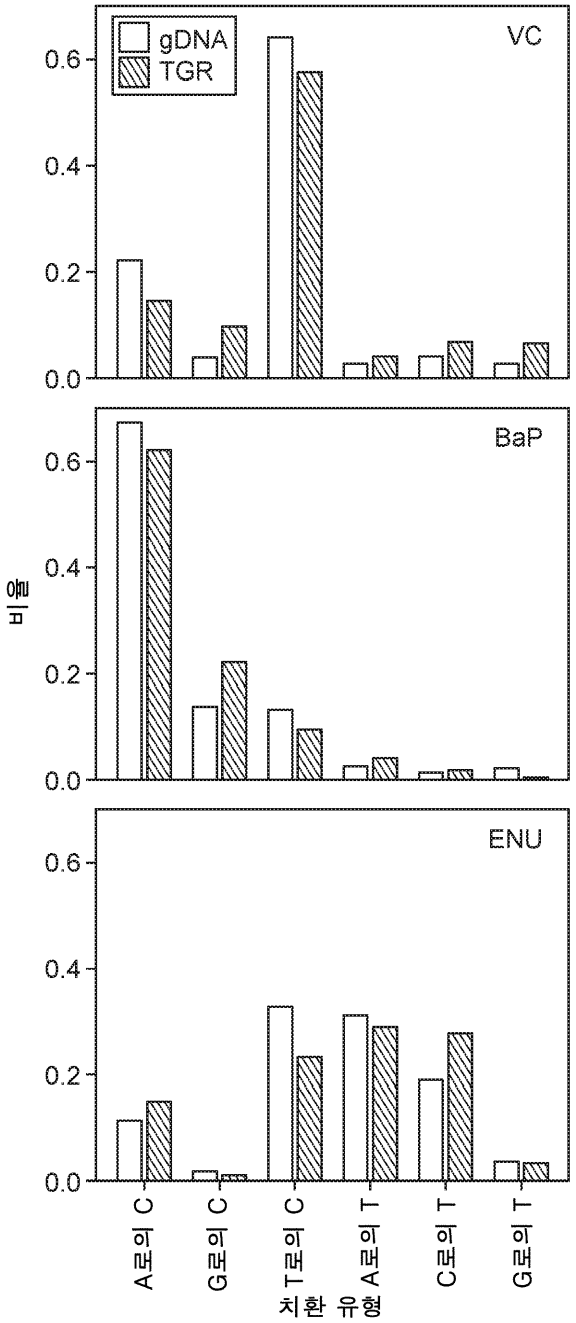
도면3d



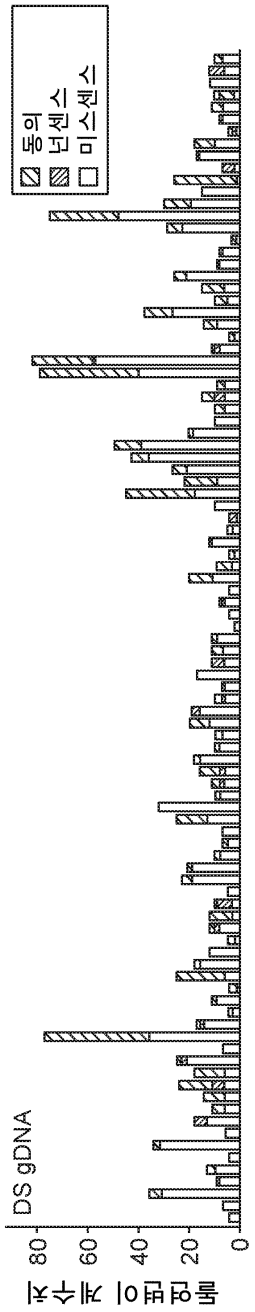
도면3e


골수 - VC
골수 - BaP
골수 - ENU
간 - VC
간 - BaP
간 - ENU
0
2
4
6
8
10
12
14
16
18
20
22
폴라크 검정
돌연변이체 빈도 배수 증가
뉴플렉스 시퀀싱 검정
돌연변이체 빈도 배수 증가

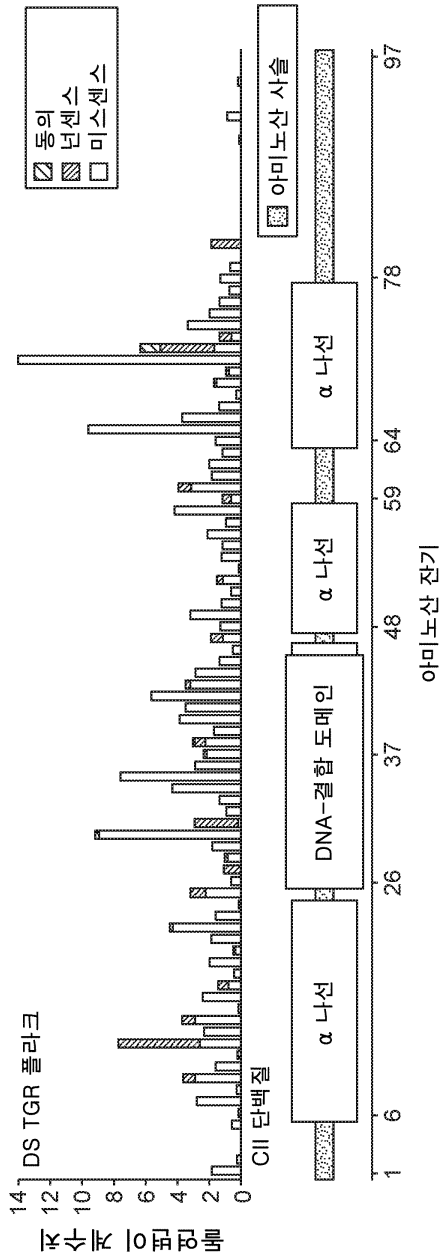
도면3f


gDNA
TGR
VC
BaP
ENU
비율
0.0
0.2
0.4
0.6
A로의 C
G로의 C
T로의 C
A로의 T
C로의 T
G로의 T
치환 유형

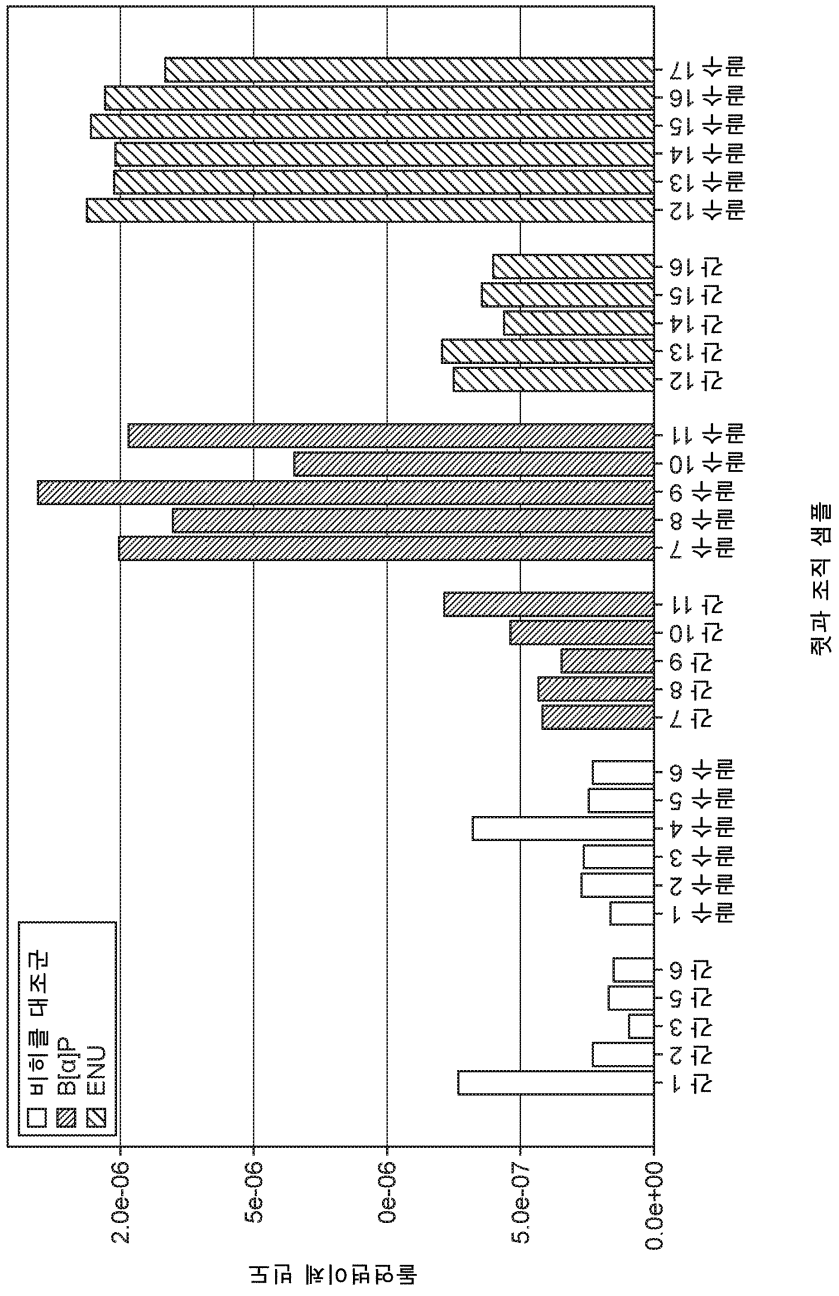
도면3g



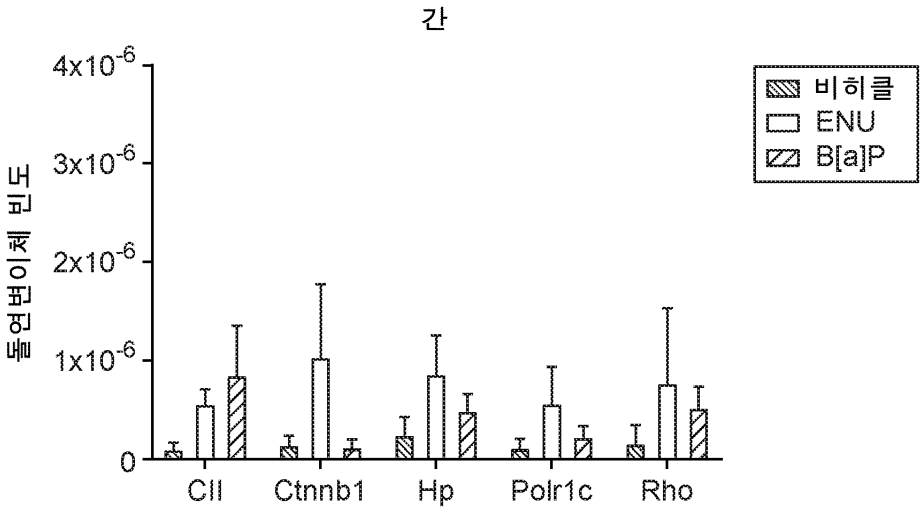
도면3h


동의
넌센스
미스센스
DS TGR 플라크
동일변이 계수치
0
2
4
6
8
10
12
14
CII 단백질
α 나선
DNA-결합 도메인
α 나선
α 나선
아미노산 사슬
1
6
26
37
48
59
64
78
97
아미노산 잔기

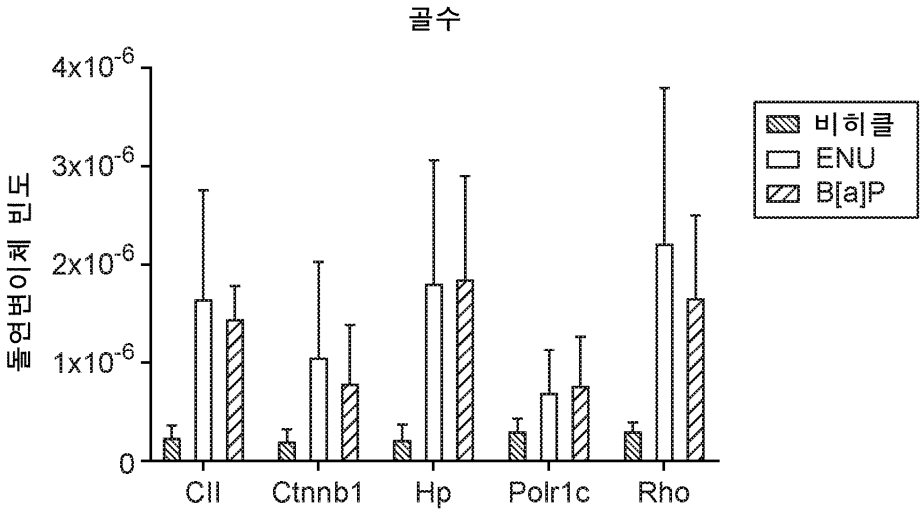
도면4


돌연변이체 빈도
2.0e-06
5e-06
0e-06
5.0e-07
0.0e+00
비히클 대조군
B[a]P
ENU
간 1
간 2
간 3
간 5
간 6
골수 1
골수 2
골수 3
골수 4
골수 5
골수 6
간 7
간 8
간 9
간 10
간 11
골수 7
골수 8
골수 9
골수 10
골수 11
간 12
간 13
간 14
간 15
간 16
골수 12
골수 13
골수 14
골수 15
골수 16
골수 17
적과 조직 샘플

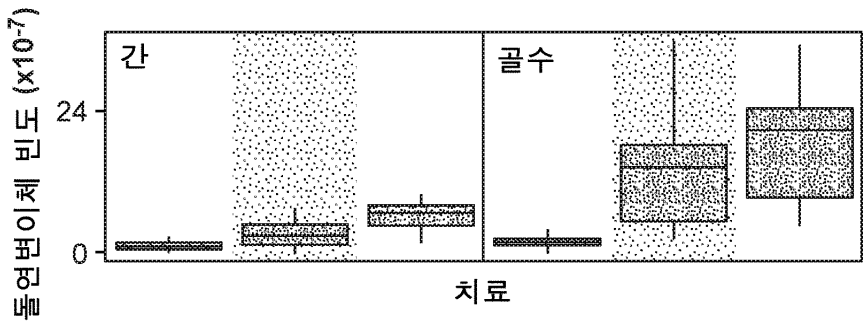
도면5a



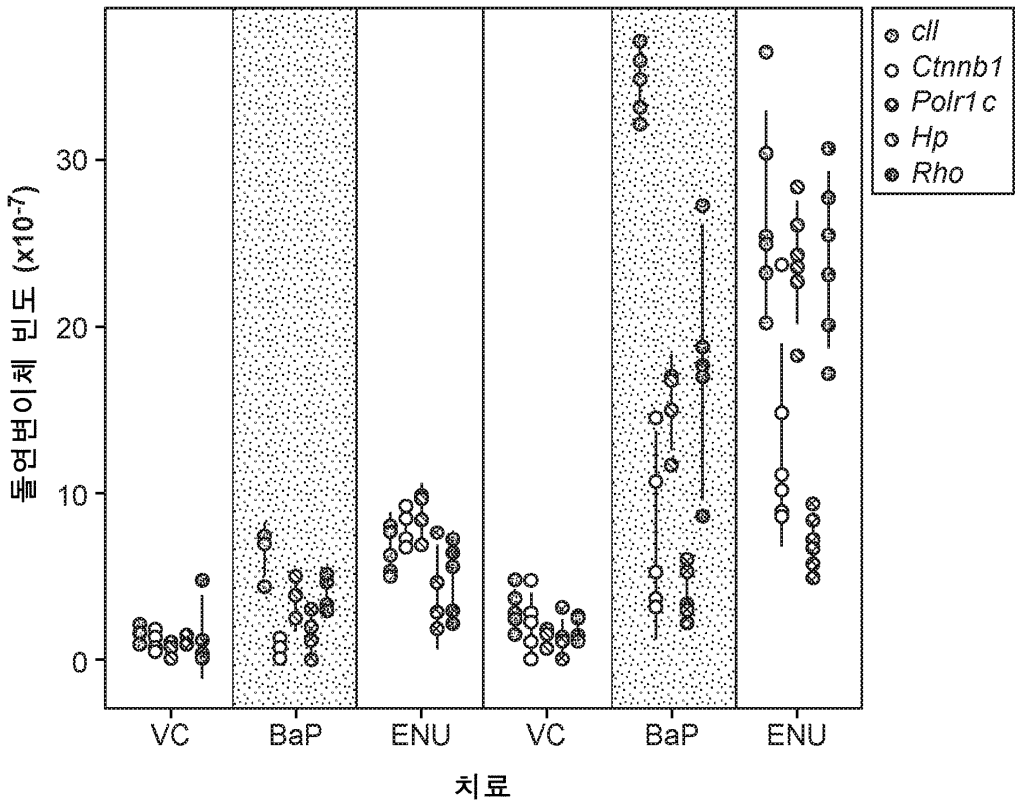
도면5b



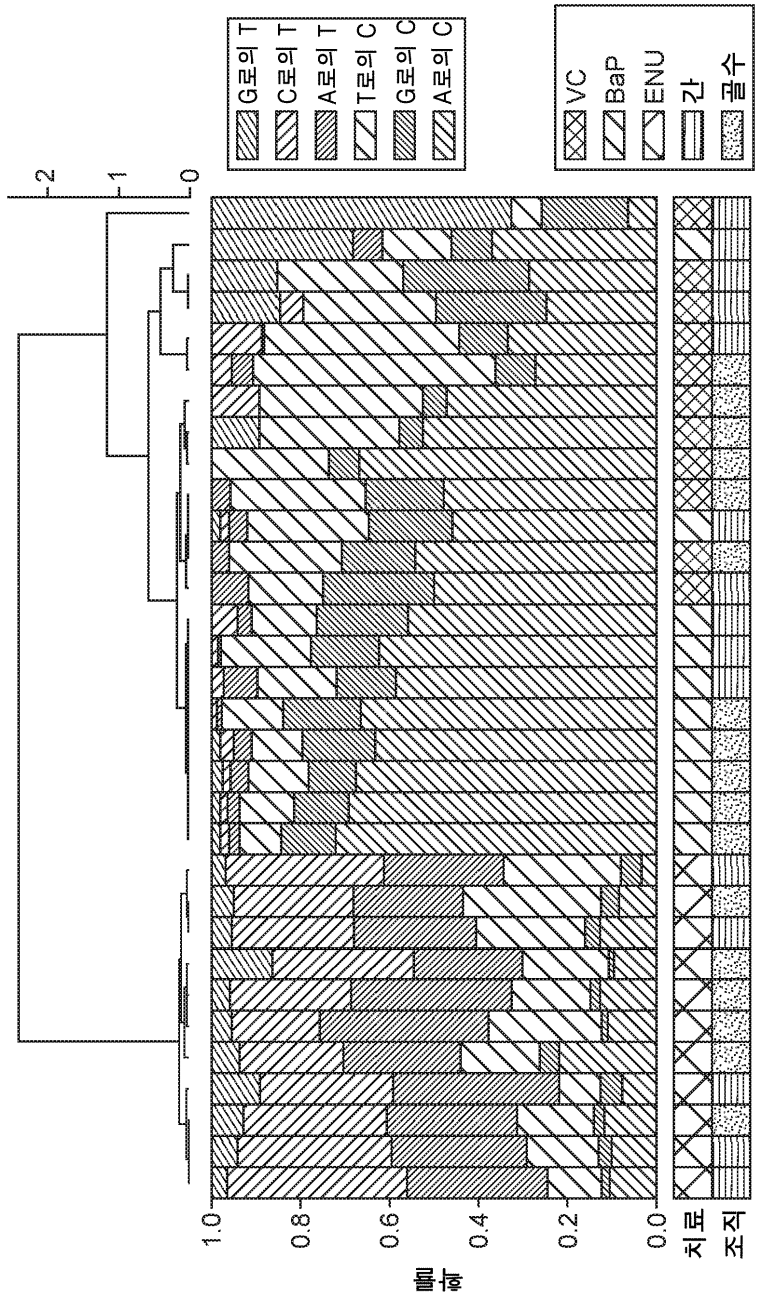
도면5c



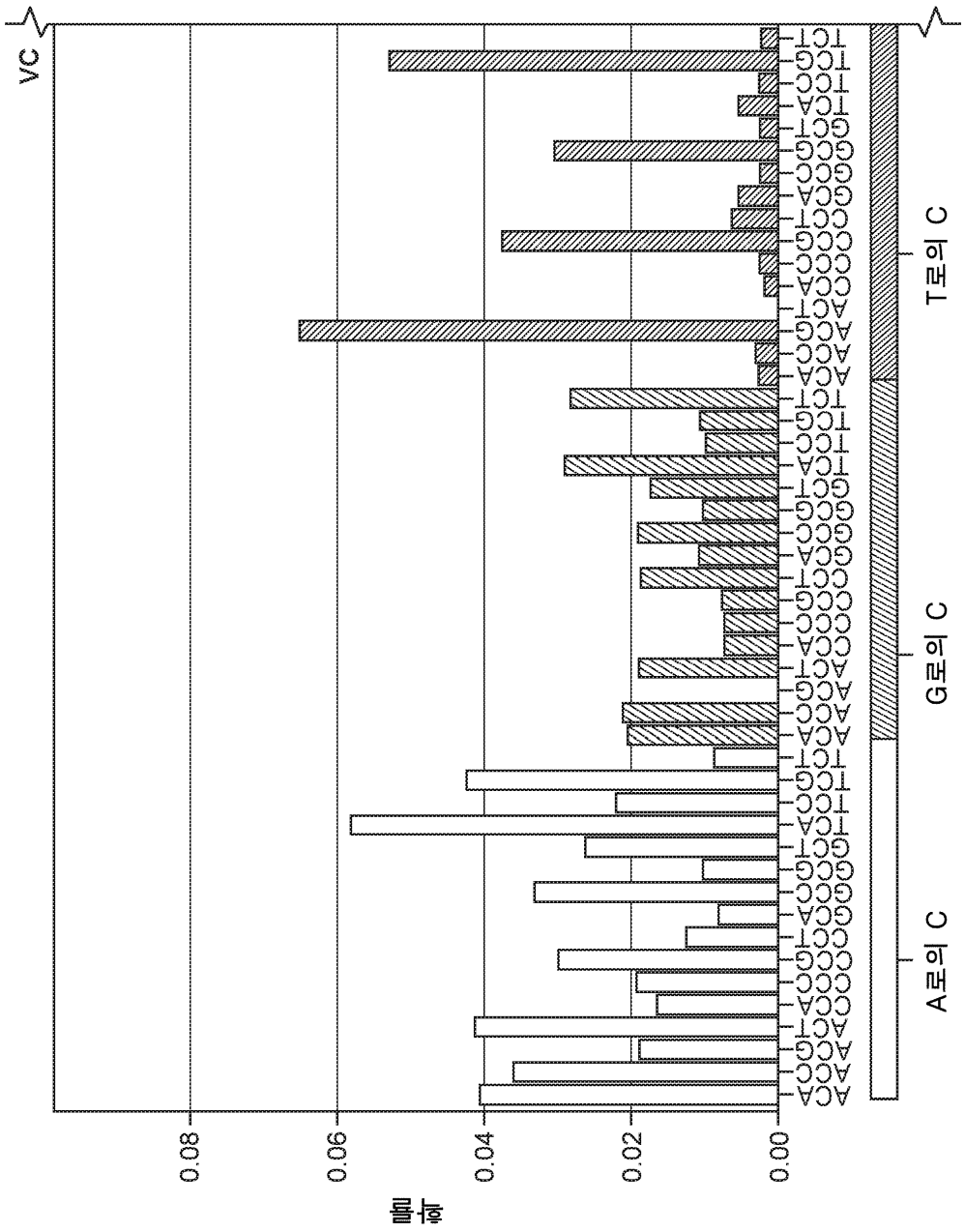
도면5d


돌연변이체 빈도 (x10-7)
30
20
10
0
VC
BaP
ENU
VC
BaP
ENU
치료
cII
Ctnnb1
Polr1c
Hp
Rho

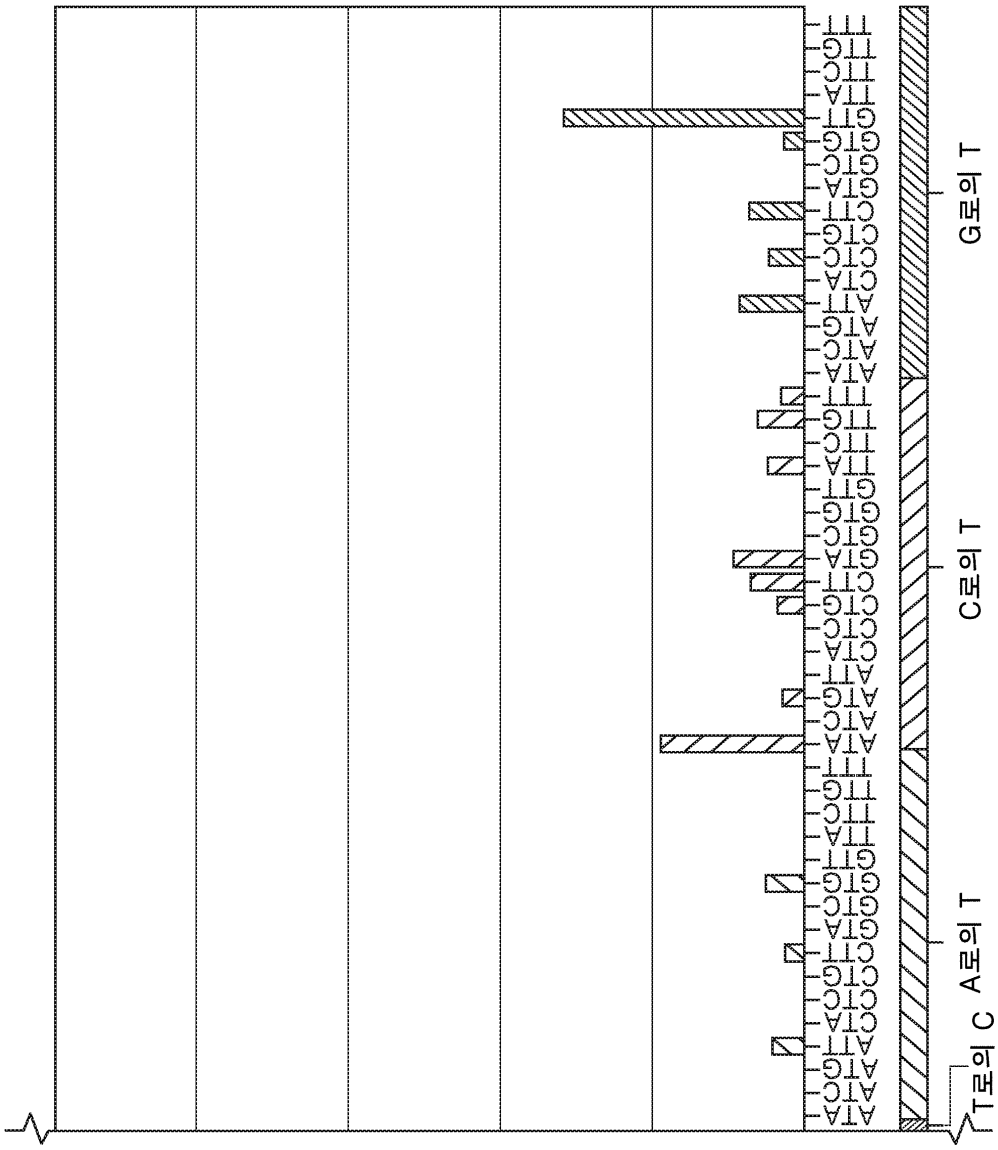
도면6


G로의 T
C로의 T
A로의 T
T로의 C
G로의 C
A로의 C
VC
BaP
ENU
간
골수
2
1
0
1.0
0.8
0.6
0.4
0.2
0.0
확률
치료
조직

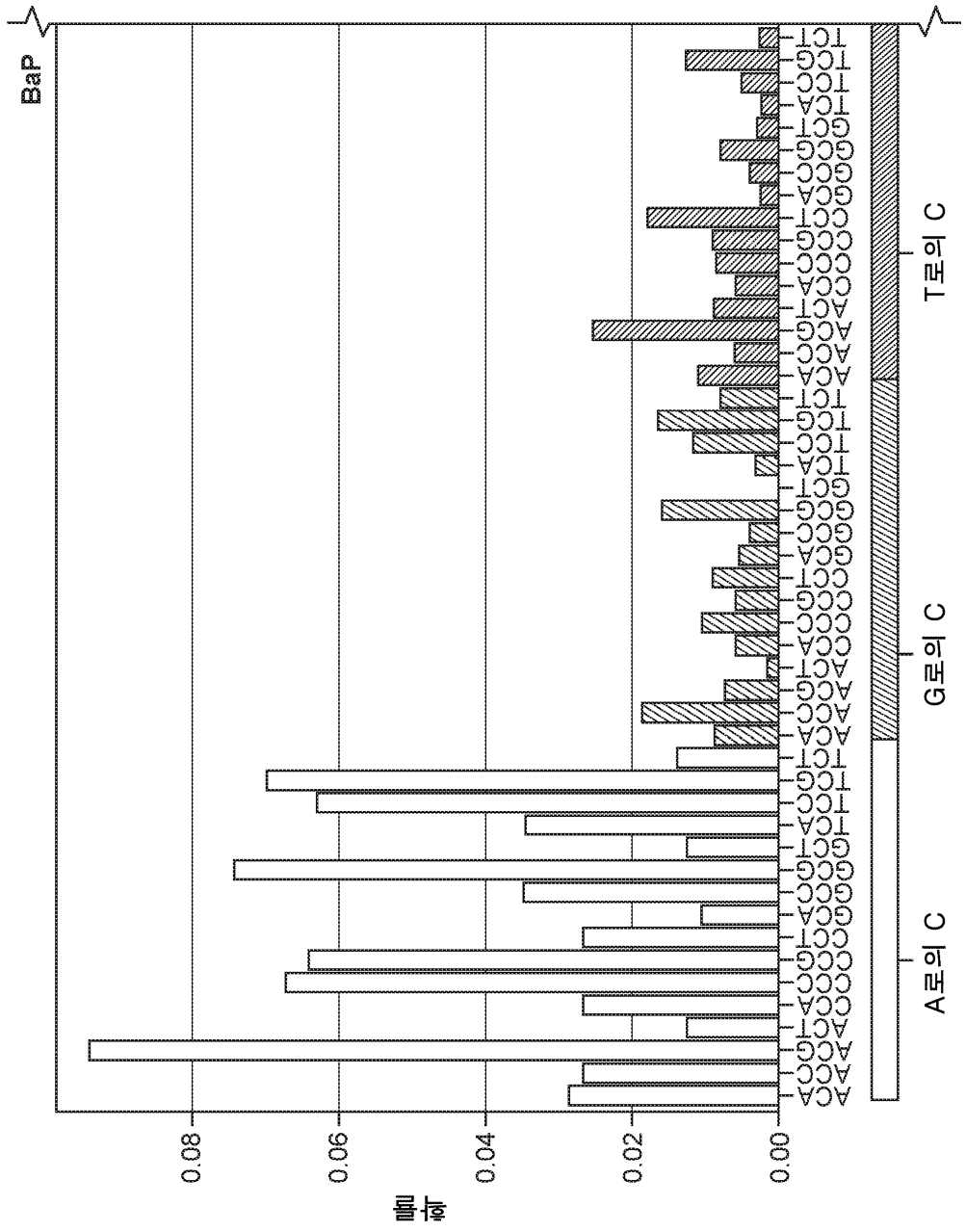
도면7aa


VC
0.08
0.06
0.04
0.02
0.00
확률
ACA
ACC
ACG
ACT
CCA
CCC
CCG
CCT
GCA
GCC
GCG
GCT
TCA
TCC
TCG
TCT
A로의 C
G로의 C
T로의 C

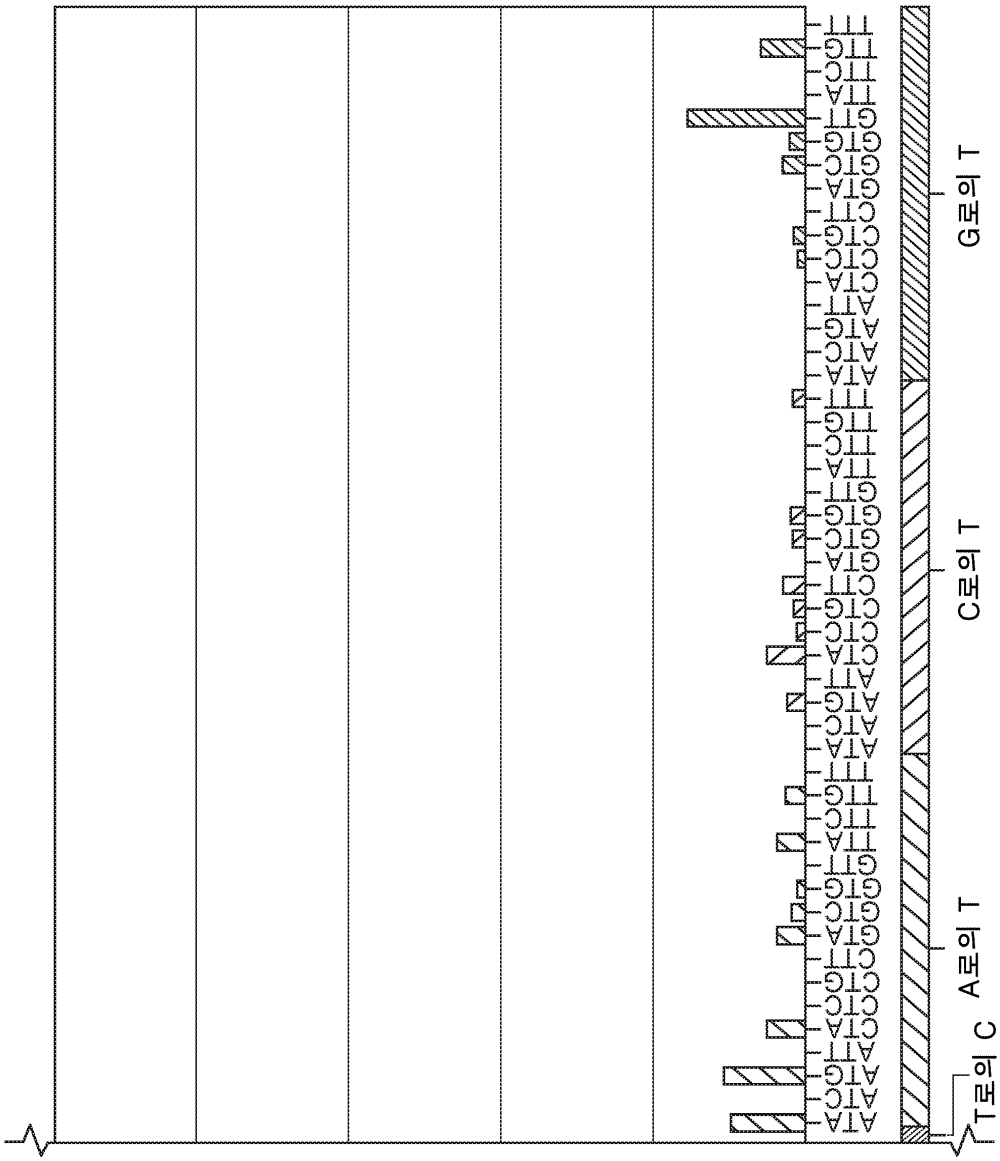
도면7ab


TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
G로의 T
C로의 T
A로의 T
T로의 C

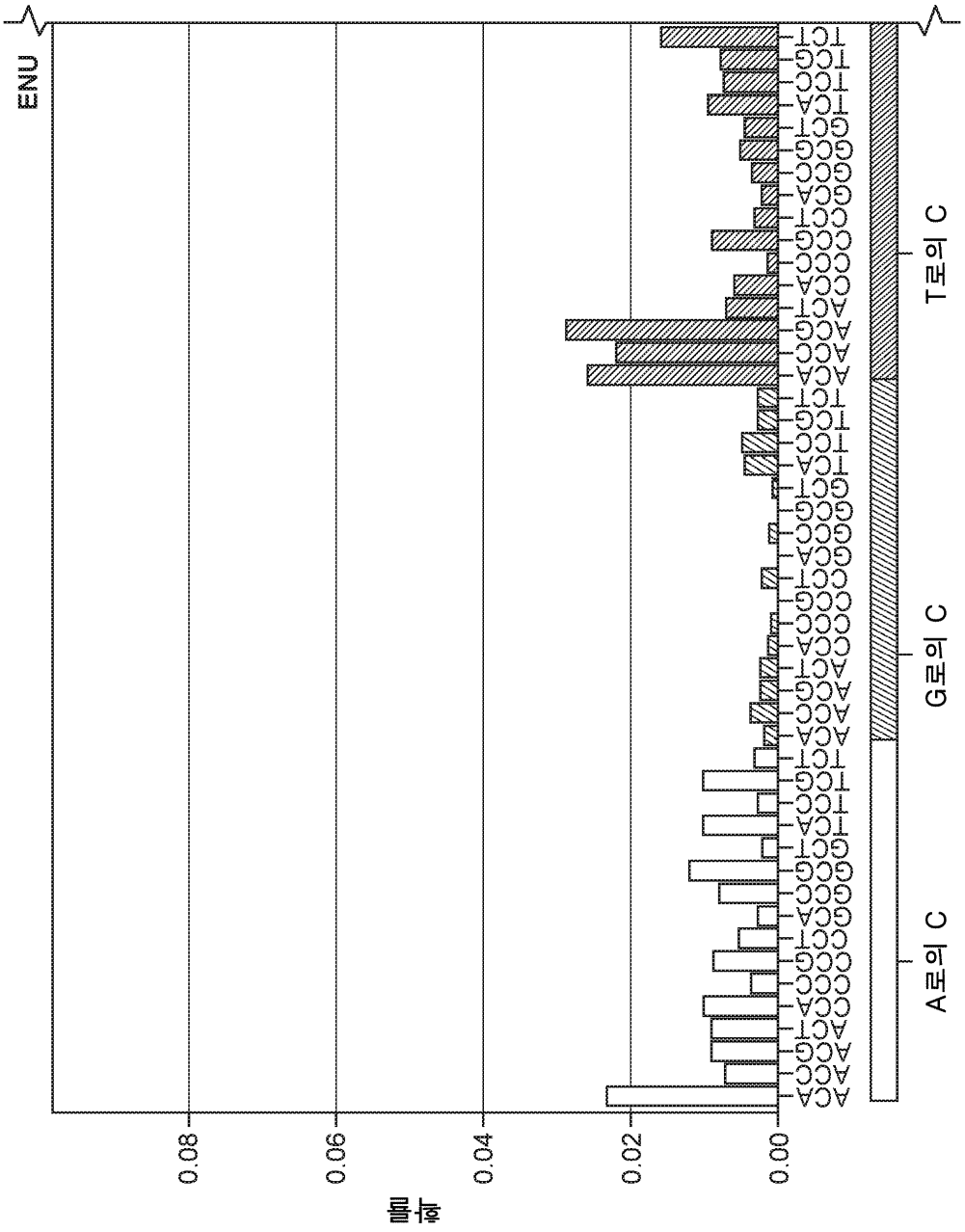
도면7ba


BaP
확률
0.00
0.02
0.04
0.06
0.08
ACA
ACC
ACG
ACT
CCA
CCC
CCG
CCT
GCA
GCC
GCG
GCT
TCA
TCC
TCG
TCT
A로의 C
G로의 C
T로의 C

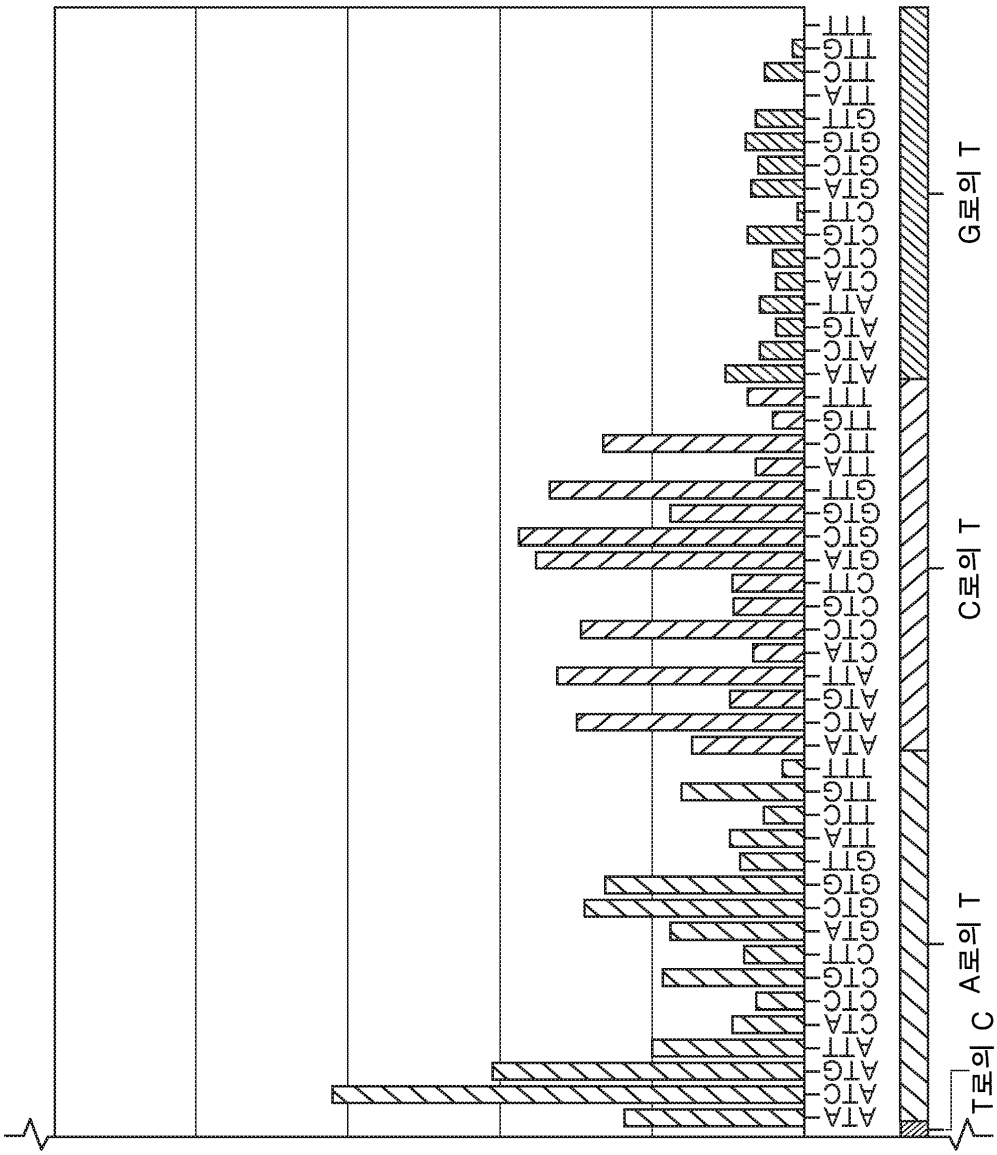
도면7bb


TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
G로의 T
C로의 T
A로의 T
T로의 C

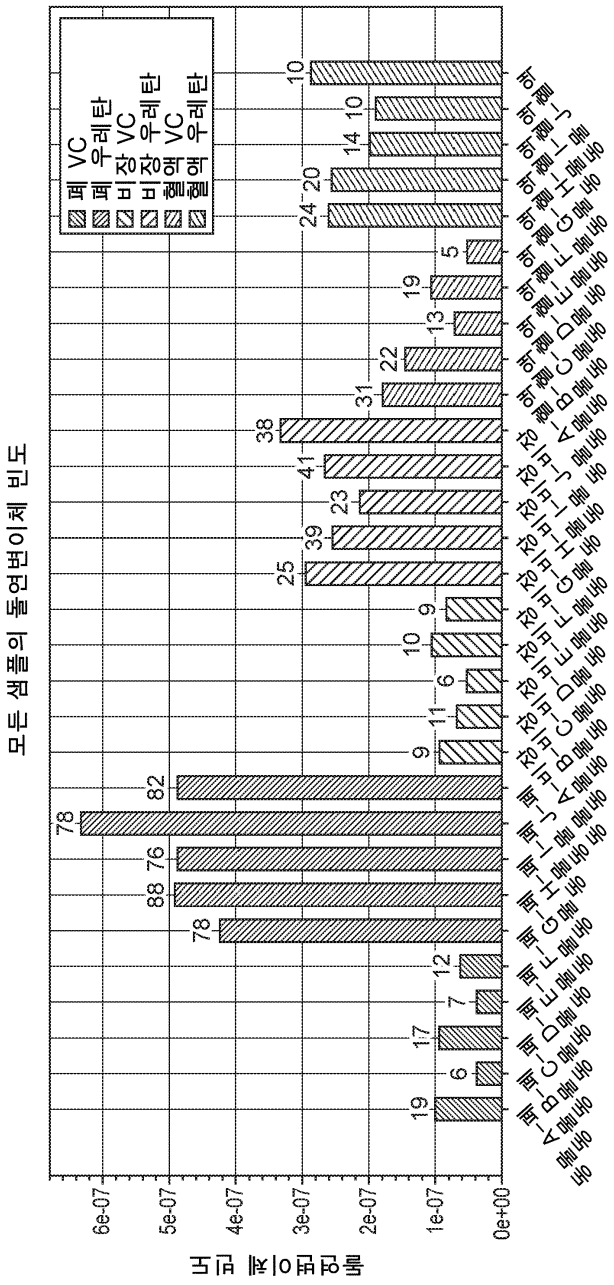
도면7ca


ENU
0.08
0.06
0.04
0.02
0.00
확률
ACA
ACC
ACG
ACT
CCA
CCC
CCG
CCT
GCA
GCC
GCG
GCT
TCA
TCC
TCG
TCT
A로의 C
G로의 C
T로의 C

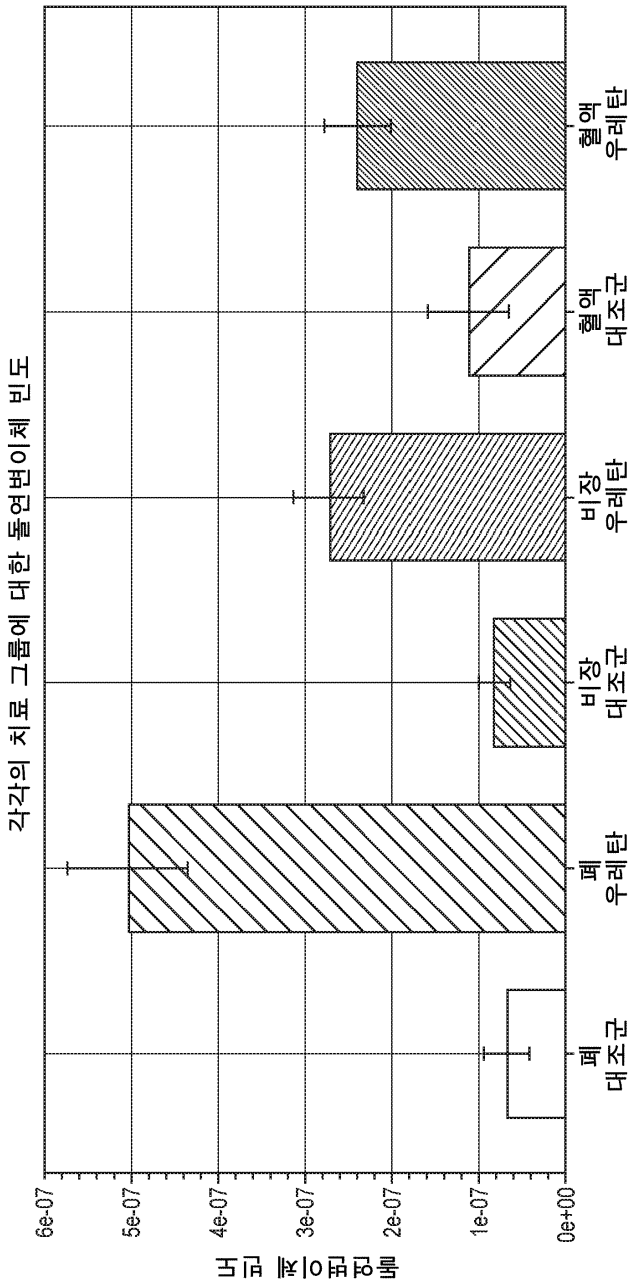
도면7cb


TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
G로의 T
C로의 T
A로의 T
T로의 C

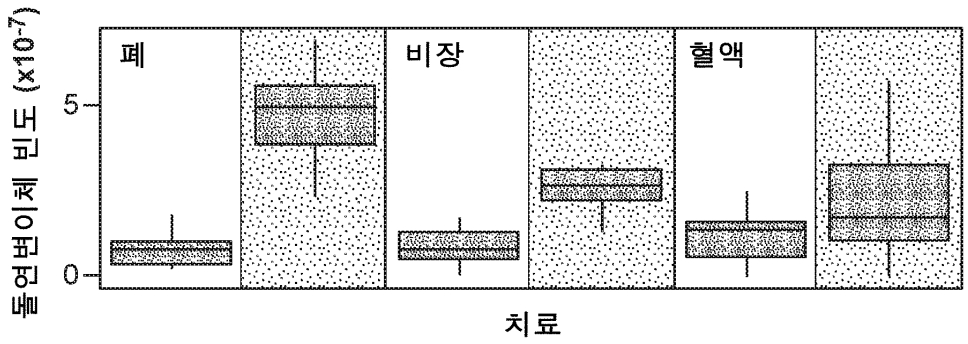
도면8


모든 샘플의 돌연변이체 빈도
돌연변이체 빈도
0e+00
1e-07
2e-07
3e-07
4e-07
5e-07
6e-07
폐 VC
폐 우레탄
비장 VC
비장 우레탄
췌액 VC
췌액 우레탄

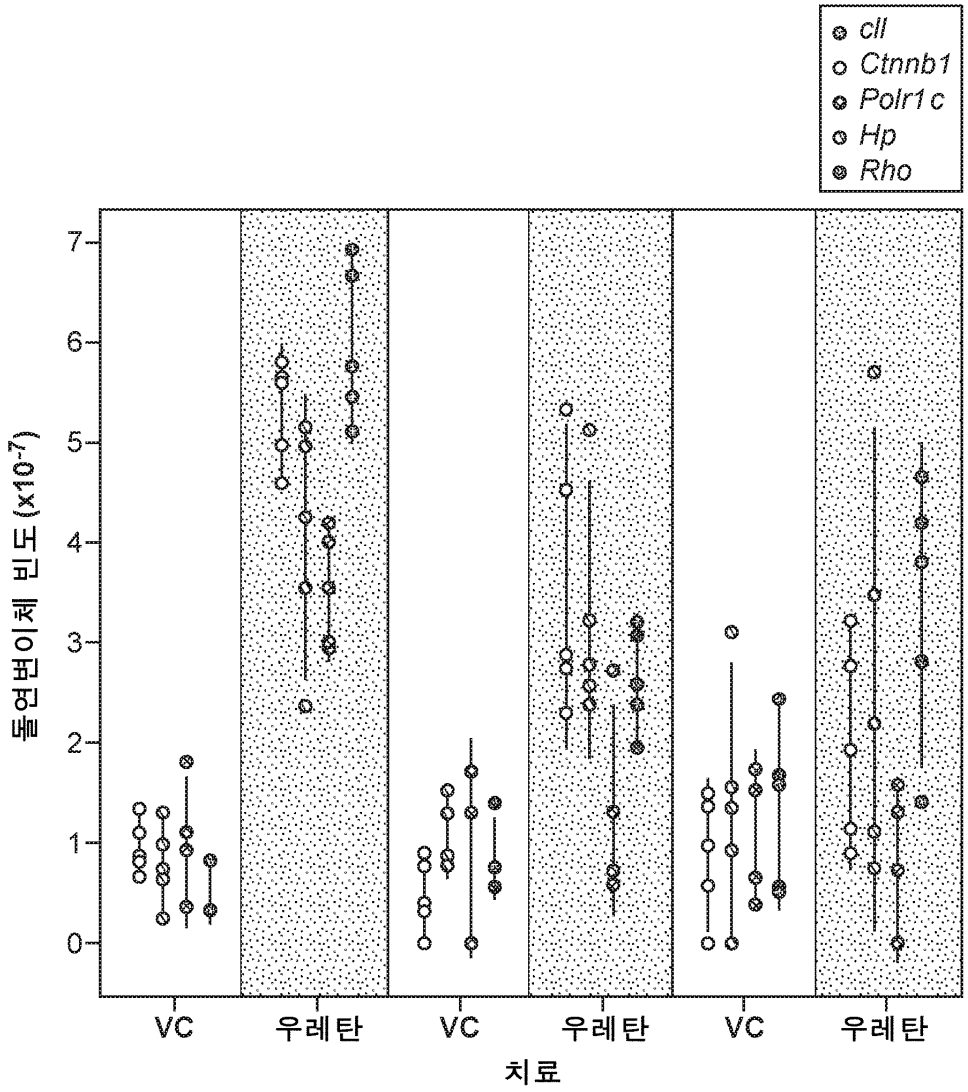
도면9


각각의 치료 그룹에 대한 돌연변이체 빈도
돌연변이체 빈도
6e-07
5e-07
4e-07
3e-07
2e-07
1e-07
0e+00
폐 대조군
폐 우레탄
비장 대조군
비장 우레탄
혈액 대조군
혈액 우레탄

도면10a


돌연변이체 빈도 (x10-7)
5
0
폐
비장
혈액
치료

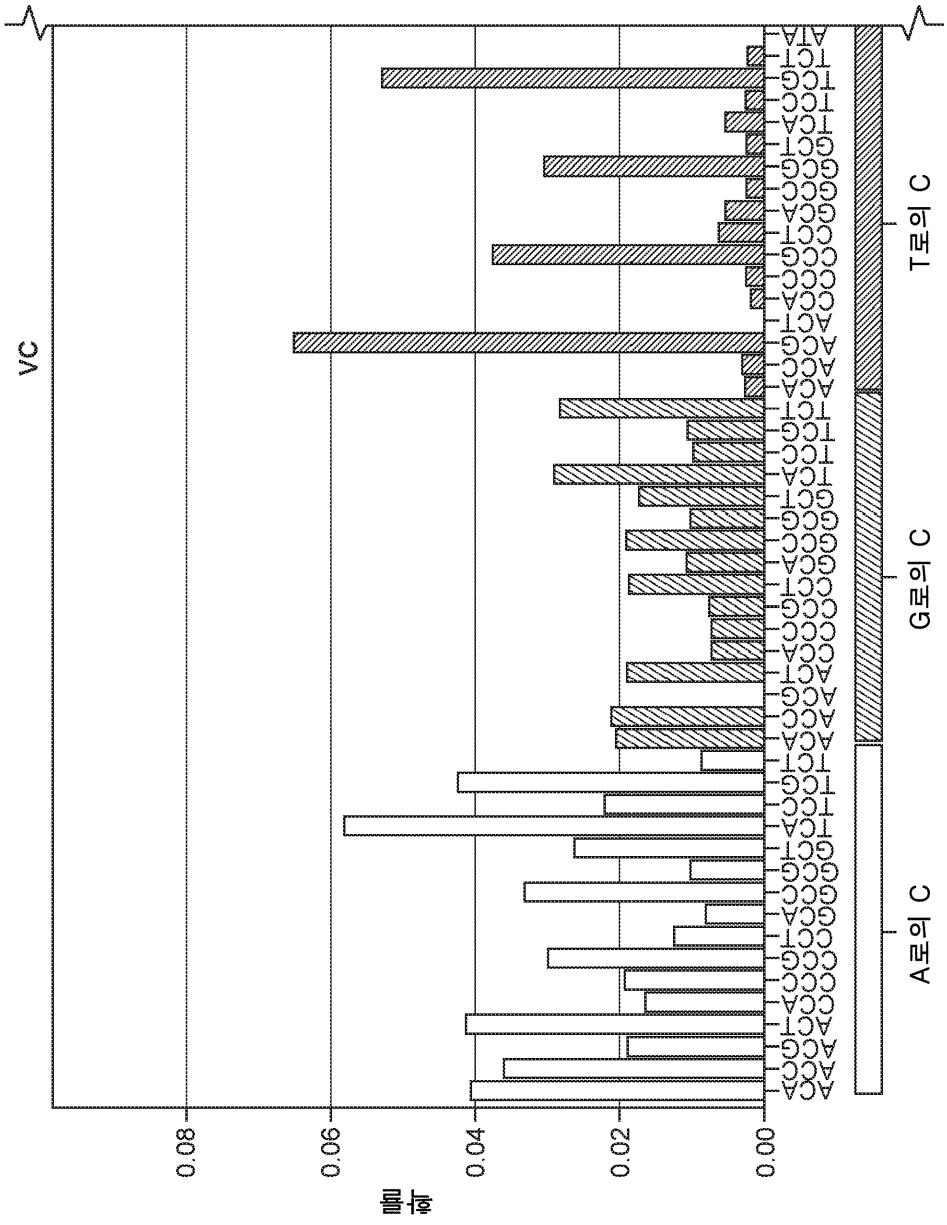
도면10b


cll
Ctnnb1
Polr1c
Hp
Rho
돌연변이체 빈도 (x10-7)
0
1
2
3
4
5
6
7
VC
우레탄
VC
우레탄
VC
우레탄
치료

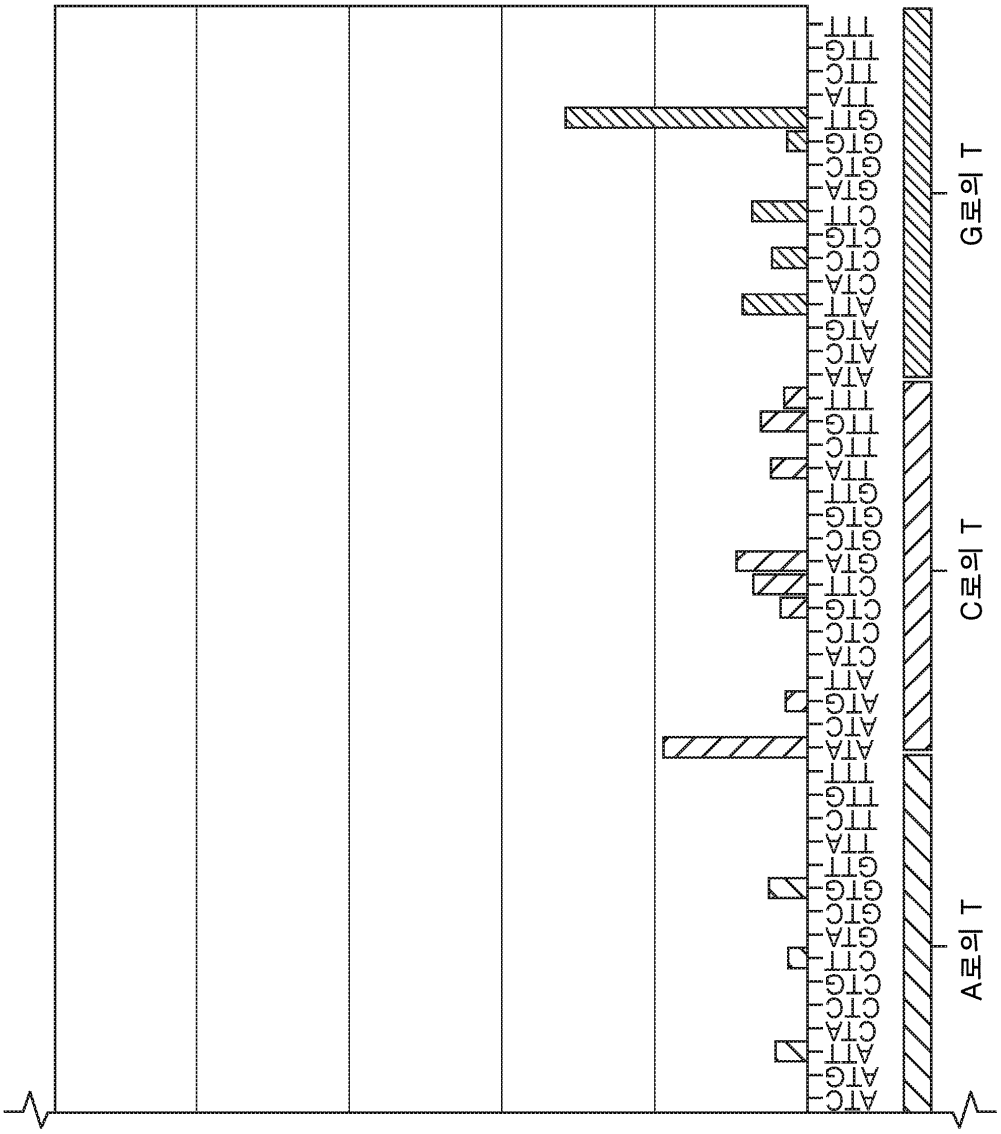
도면11

G로의 T
C로의 T
A로의 T
T로의 C
G로의 C
A로의 C
VC
우레탄
폐
비장
혈액
2
1
0
1.0
0.8
0.6
0.4
0.2
0.0
확률
치료
조직

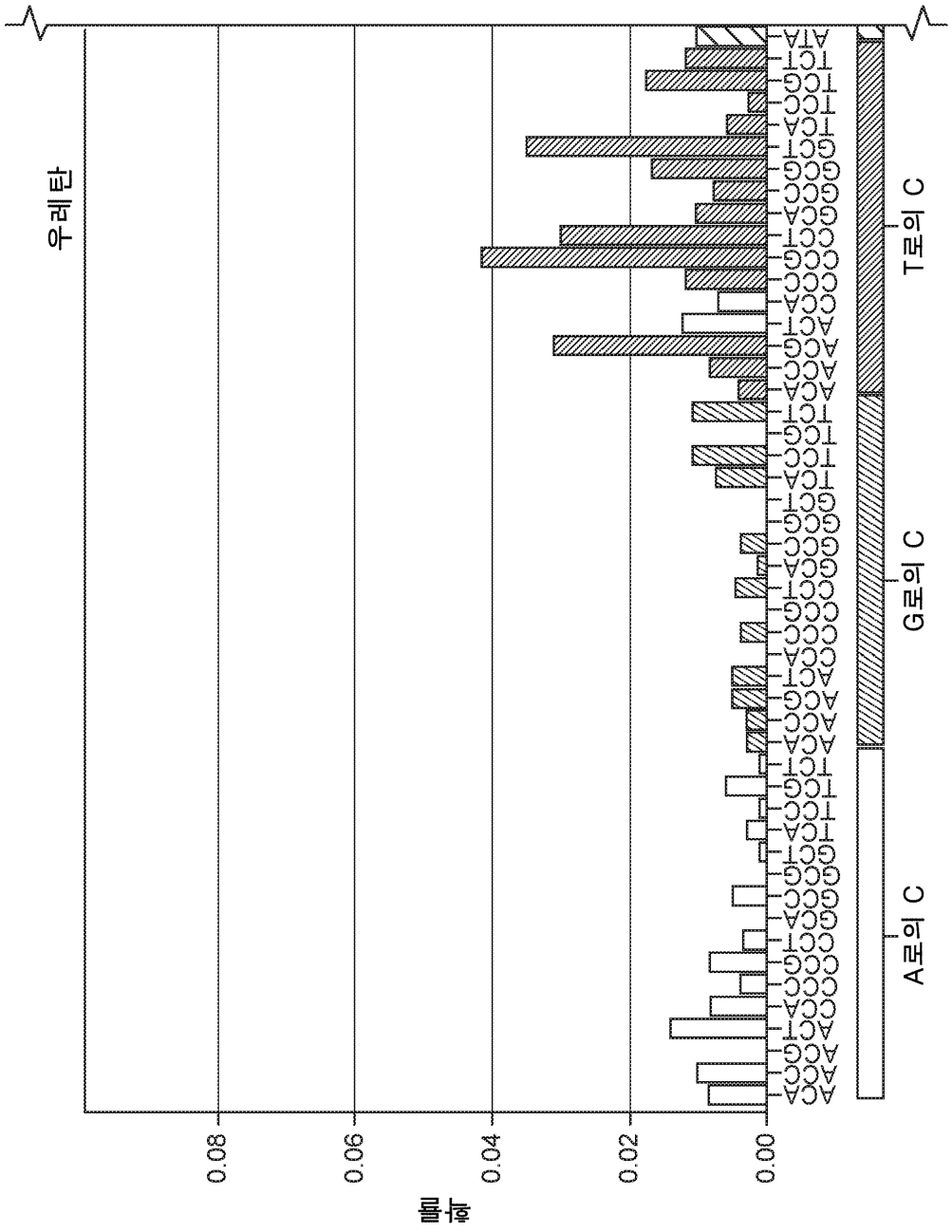
도면12aa


VC
0.08
0.06
0.04
0.02
0.00
확률
ATA
TCT
TCG
TCC
TCA
GCT
GCG
GCC
GCA
CCT
CCG
CCC
CCA
ACT
ACG
ACC
ACA
TCT
TCG
TCC
TCA
GCT
GCG
GCC
GCA
CCT
CCG
CCC
CCA
ACT
ACG
ACC
ACA
TCT
TCG
TCC
TCA
GCT
GCG
GCC
GCA
CCT
CCG
CCC
CCA
ACT
ACG
ACC
ACA
T로의 C
G로의 C
A로의 C

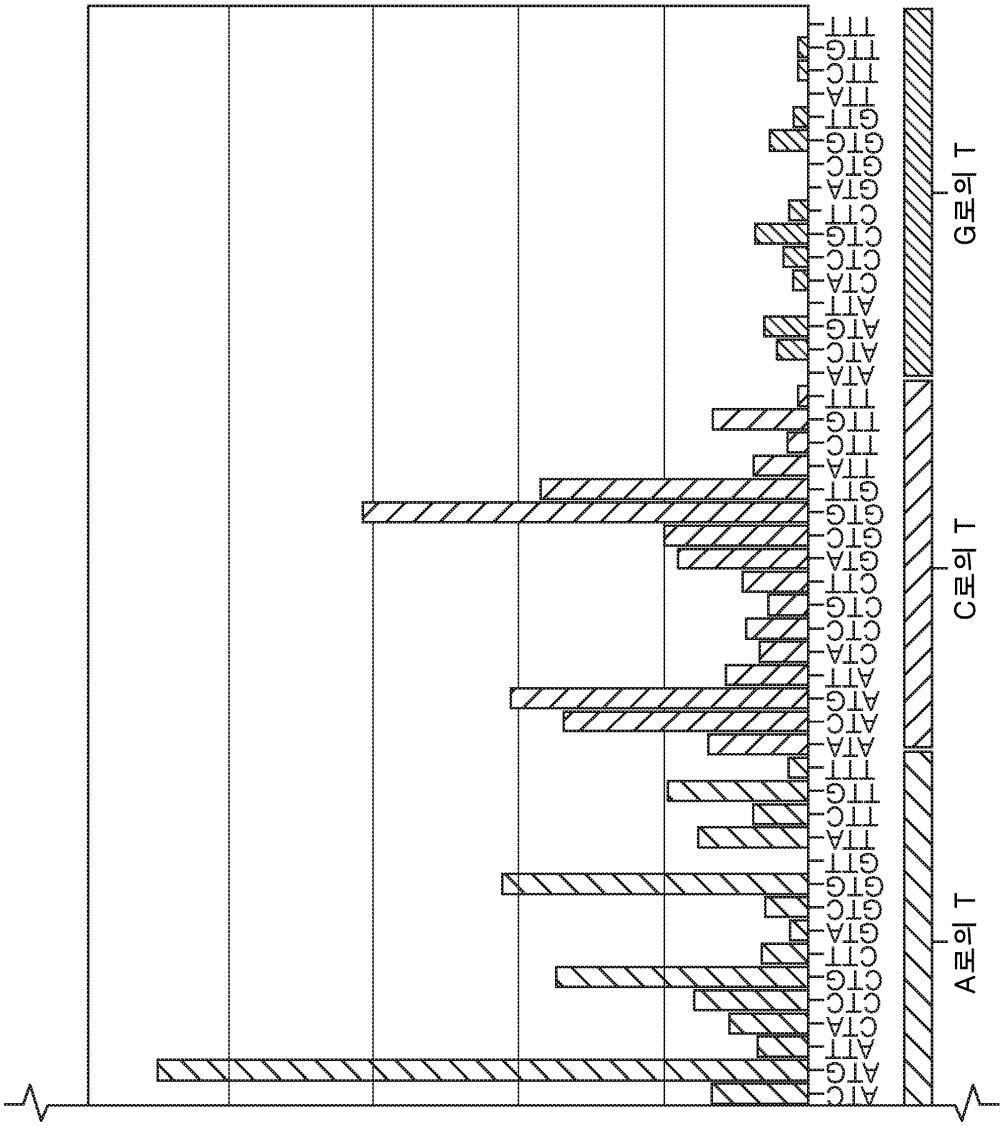
도면12ab


TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
G로의 T
C로의 T
A로의 T

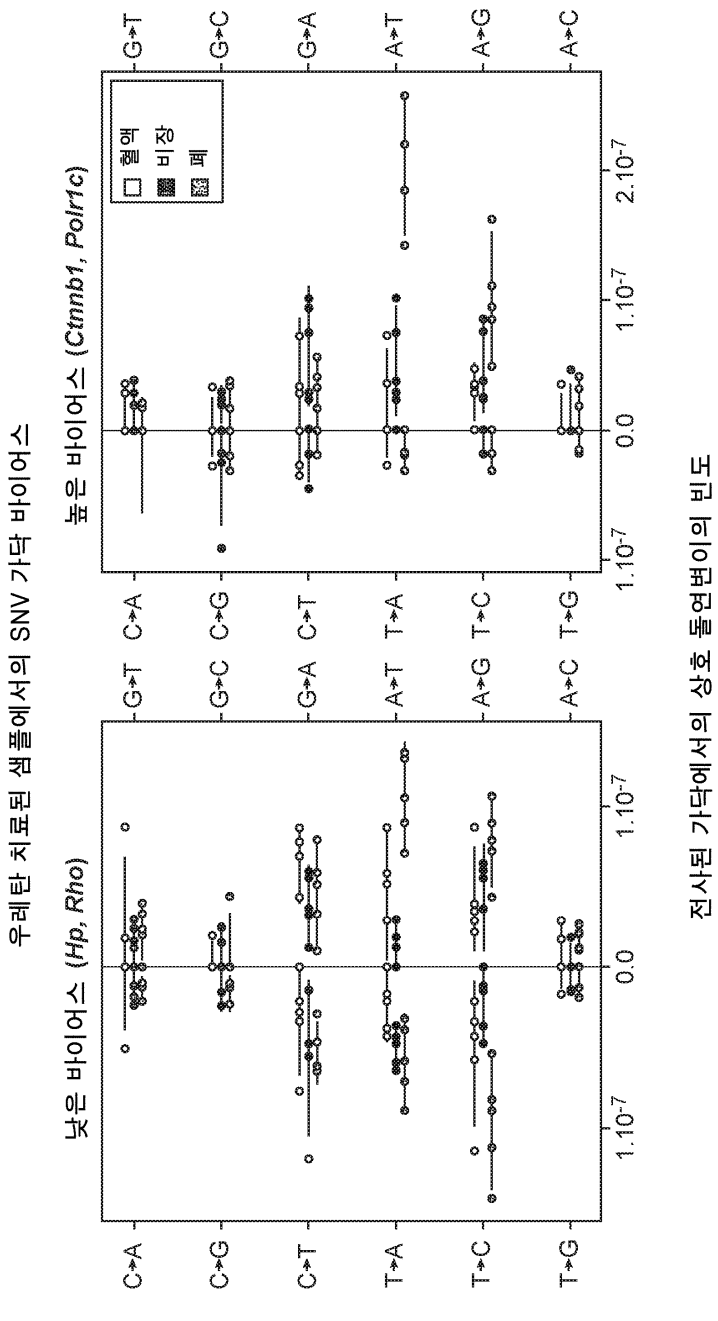
도면12ba



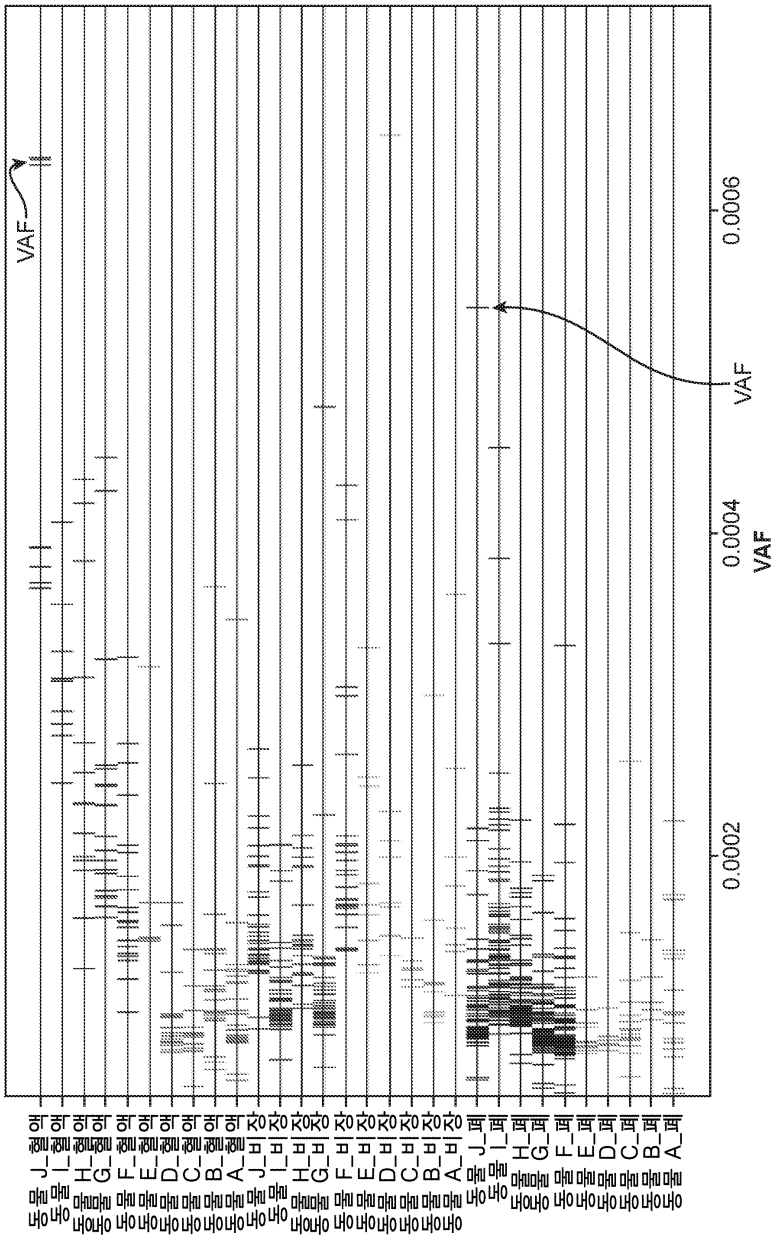
도면12bb


TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
G로의 T
C로의 T
A로의 T

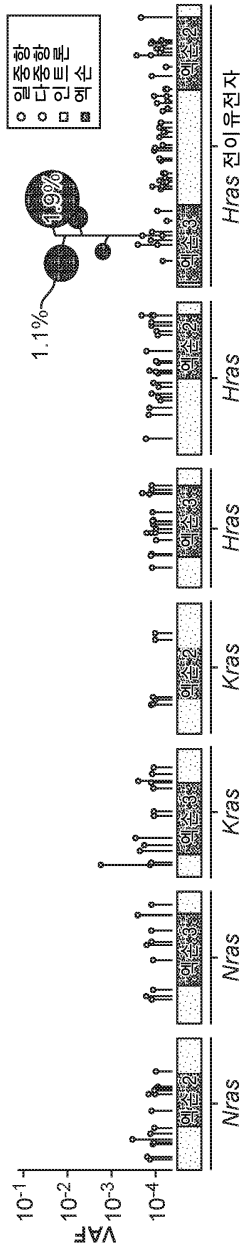
도면13


우레탄 치료된 샘플에서의 SNV 가닥 바이어스
낮은 바이어스 (Hp, Rho)
높은 바이어스 (Ctnnb1, Polr1c)
혈액
비장
폐
C>A
C>G
C>T
T>A
T>C
T>G
G>T
G>C
G>A
A>T
A>G
A>C
1.10-7
0.0
2.10-7
전사된 가닥에서의 상호 돌연변이의 빈도

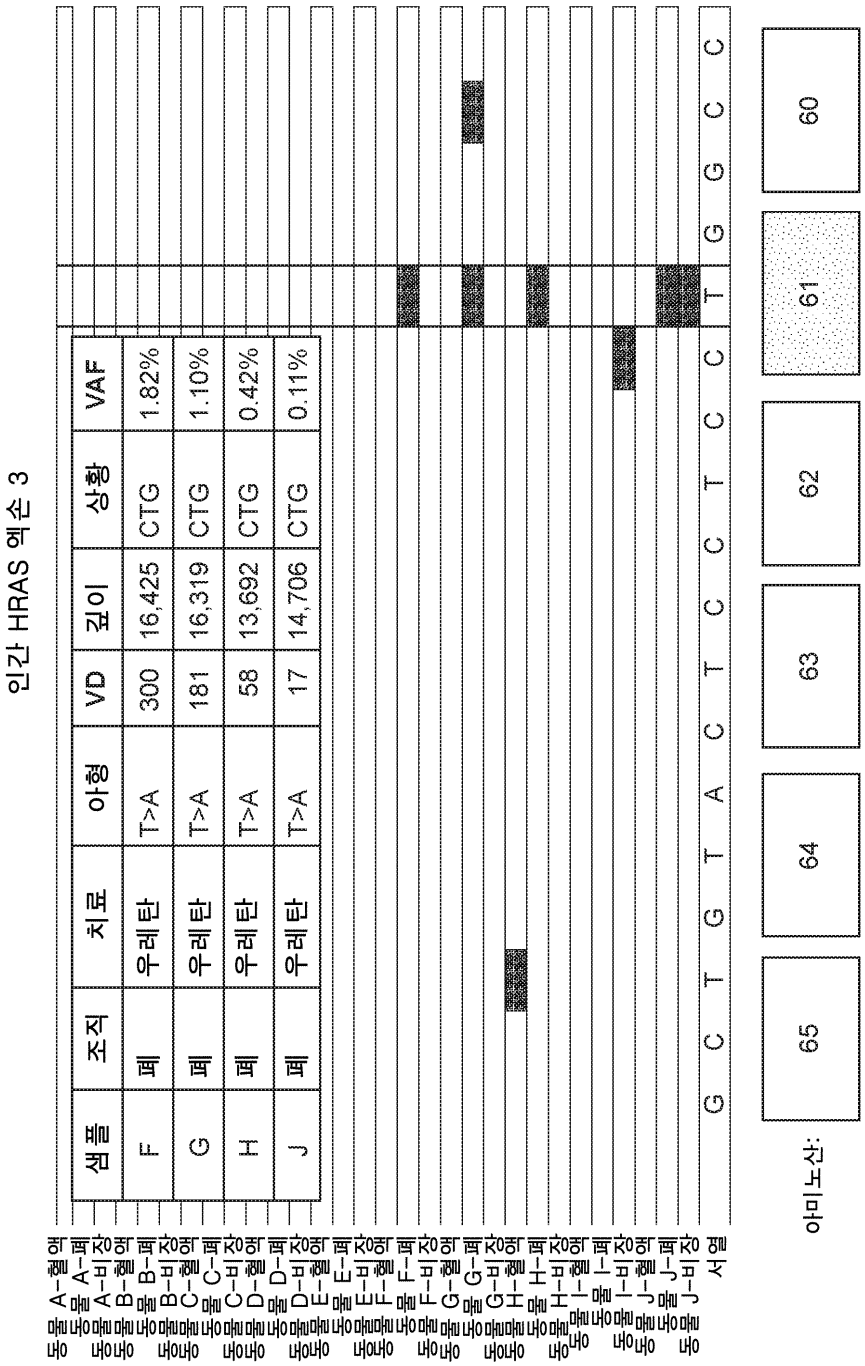
도면14



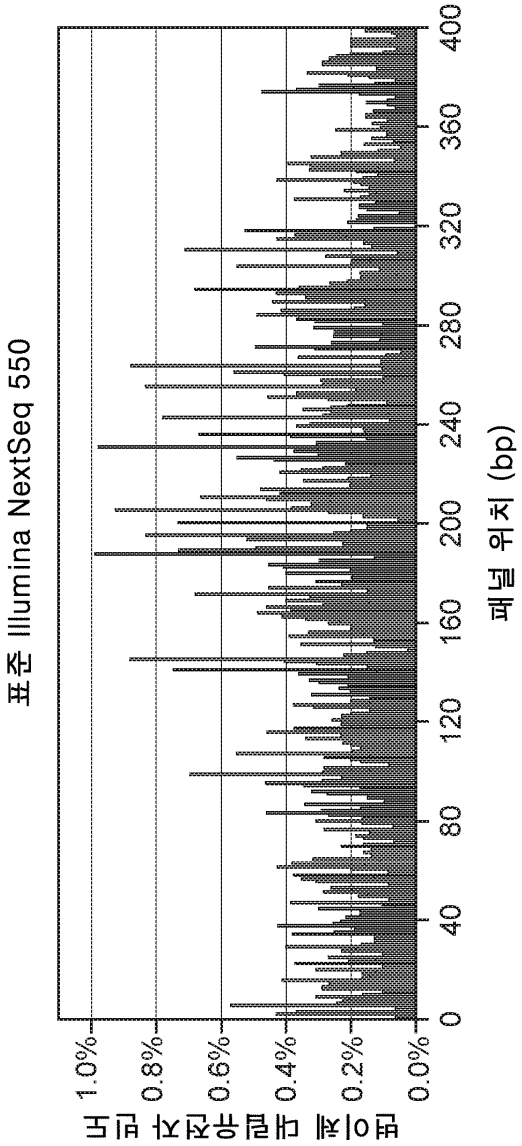
도면15a


일중합
다중합
인트론
엑손
1.1%
1.9%
VAF
10^{-1}
10^{-2}
10^{-3}
10^{-4}
엑손2
엑손3
엑손2
Nras
Nras
Kras
Kras
Hras
Hras
Hras 전이유전자

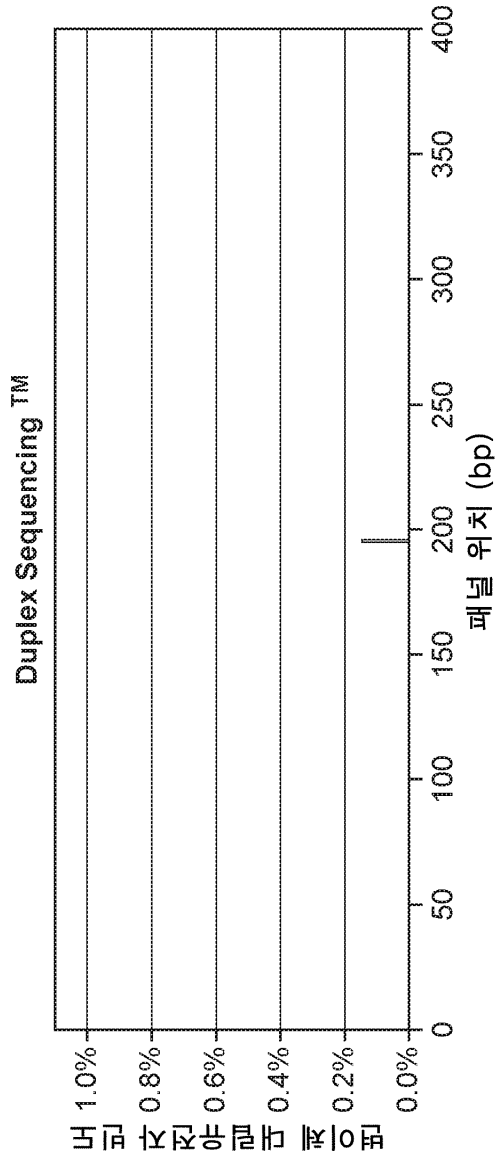
도면15b


인간 HRAS 엑손 3
샘플
조직
치료
아형
VD
깊이
상황
VAF
F
폐
우레탄
T>A
300
16,425
CTG
1.82%
G
폐
우레탄
T>A
181
16,319
CTG
1.10%
H
폐
우레탄
T>A
58
13,692
CTG
0.42%
J
폐
우레탄
T>A
17
14,706
CTG
0.11%
A-혈액
A-폐장
B-혈액
B-폐장
C-혈액
C-비장
D-혈액
D-비장
E-혈액
E-비장
F-혈액
F-폐장
G-혈액
G-비장
H-혈액
H-폐장
I-혈액
I-폐장
J-혈액
J-비장
정상
아미노산:
65
64
63
62
61
60

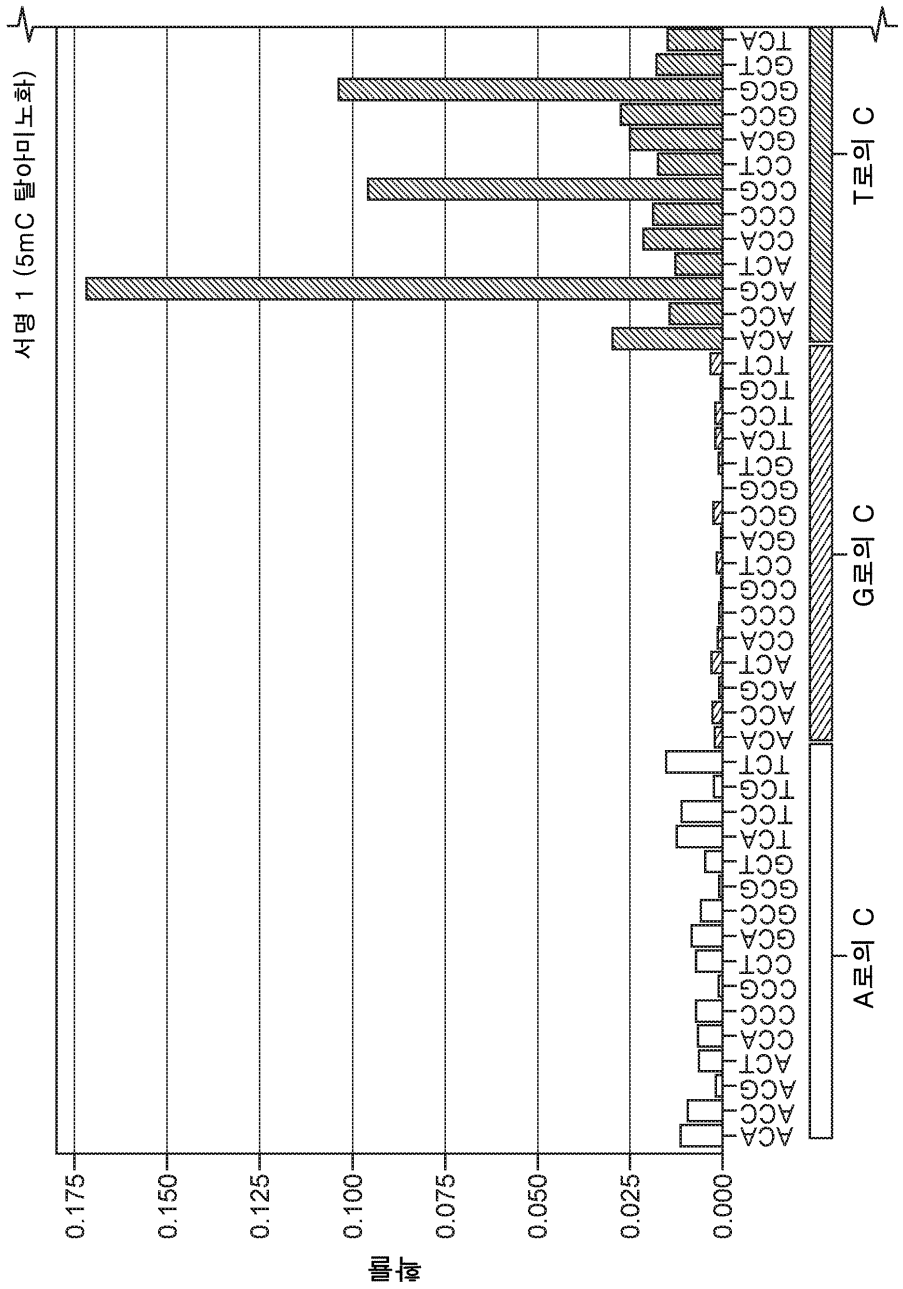
도면16a


표준 Illumina NextSeq 550
변이체 대립유전자 빈도
1.0%
0.8%
0.6%
0.4%
0.2%
0.0%
0
40
80
120
160
200
240
280
320
360
400
패널 위치 (bp)

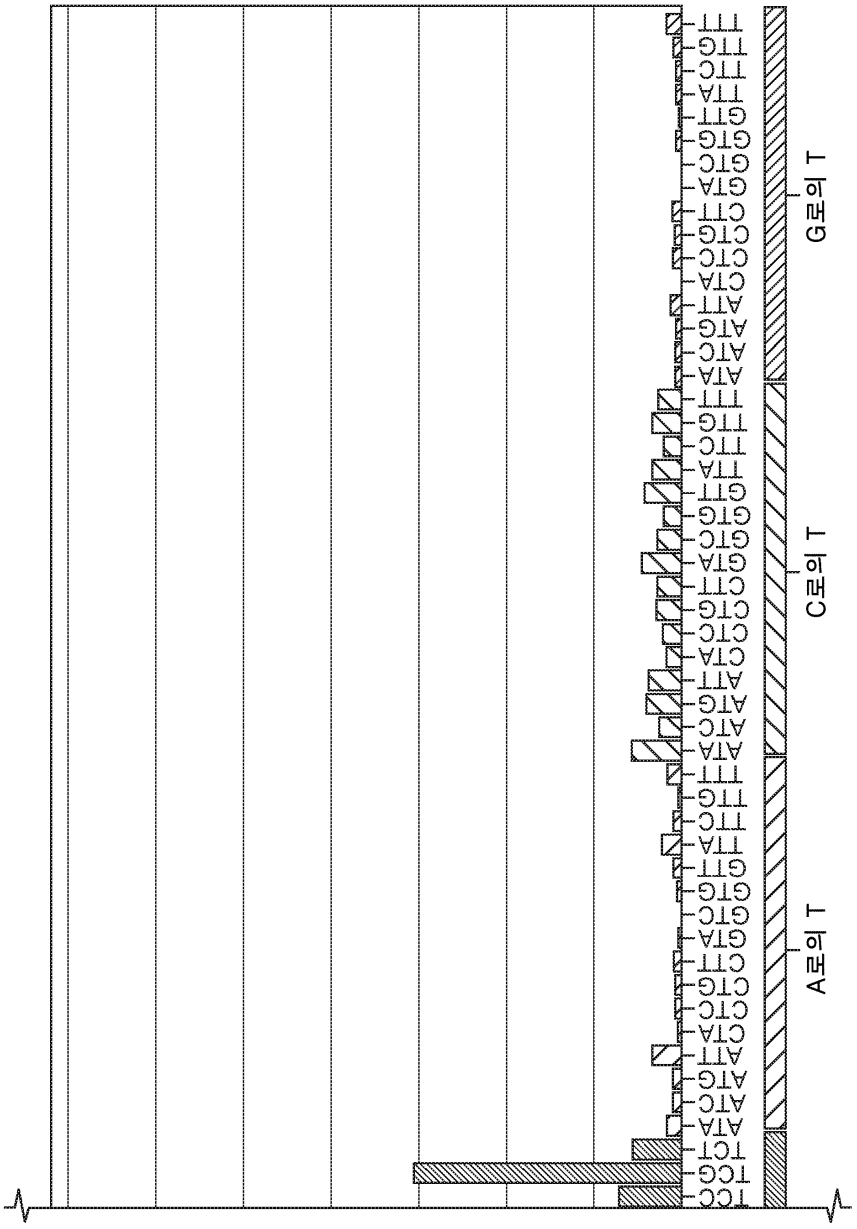
도면16b


Duplex Sequencing ™
변이체 대립유전자 빈도
1.0%
0.8%
0.6%
0.4%
0.2%
0.0%
0
50
100
150
200
250
300
350
400
패널 위치 (bp)

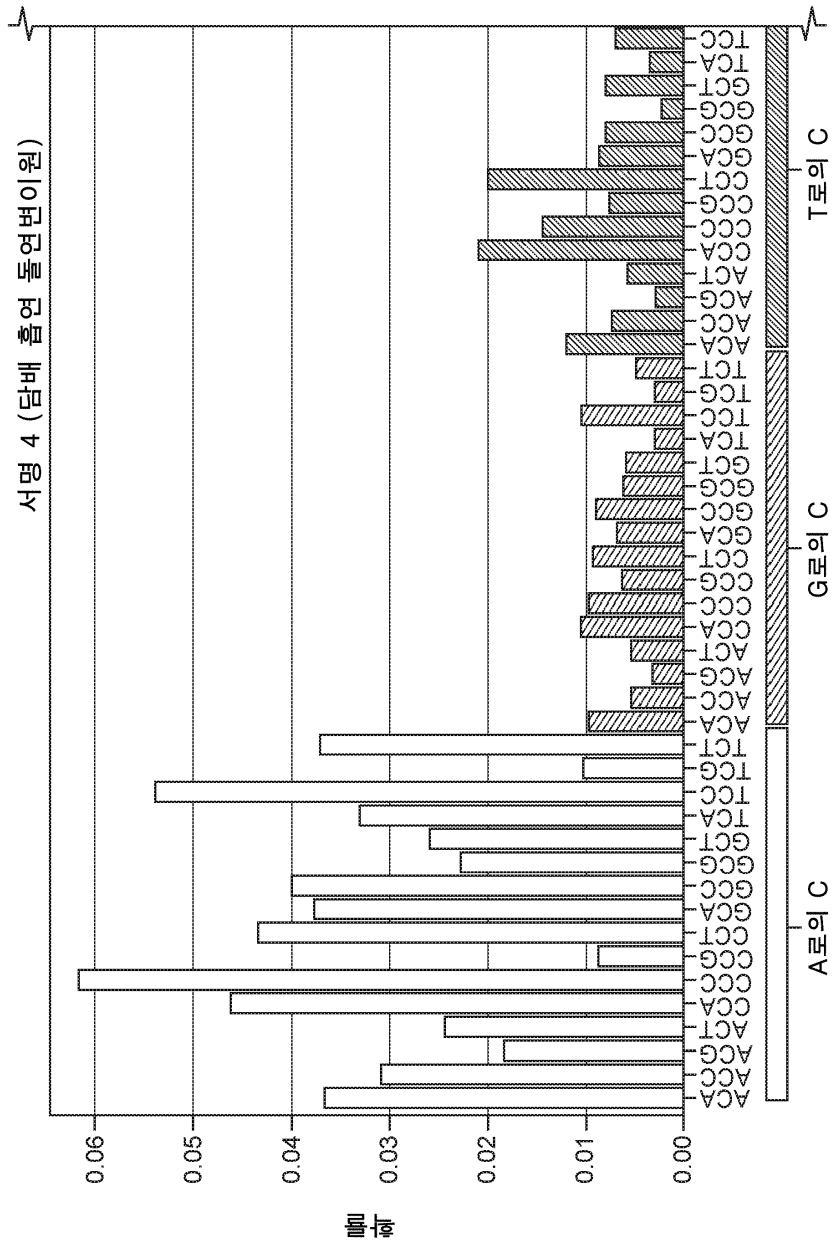
도면17aa


서명 1 (5mC 탈아미노화)
확률
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
ACA
ACC
ACG
ACT
CCA
CCC
CCG
CCT
GCA
GCC
GCG
GCT
TCA
TCC
TCG
TCT
ACA
ACC
ACG
ACT
CCA
CCC
CCG
CCT
GCA
GCC
GCG
GCT
TCA
TCC
TCG
TCT
A로의 C
G로의 C
T로의 C

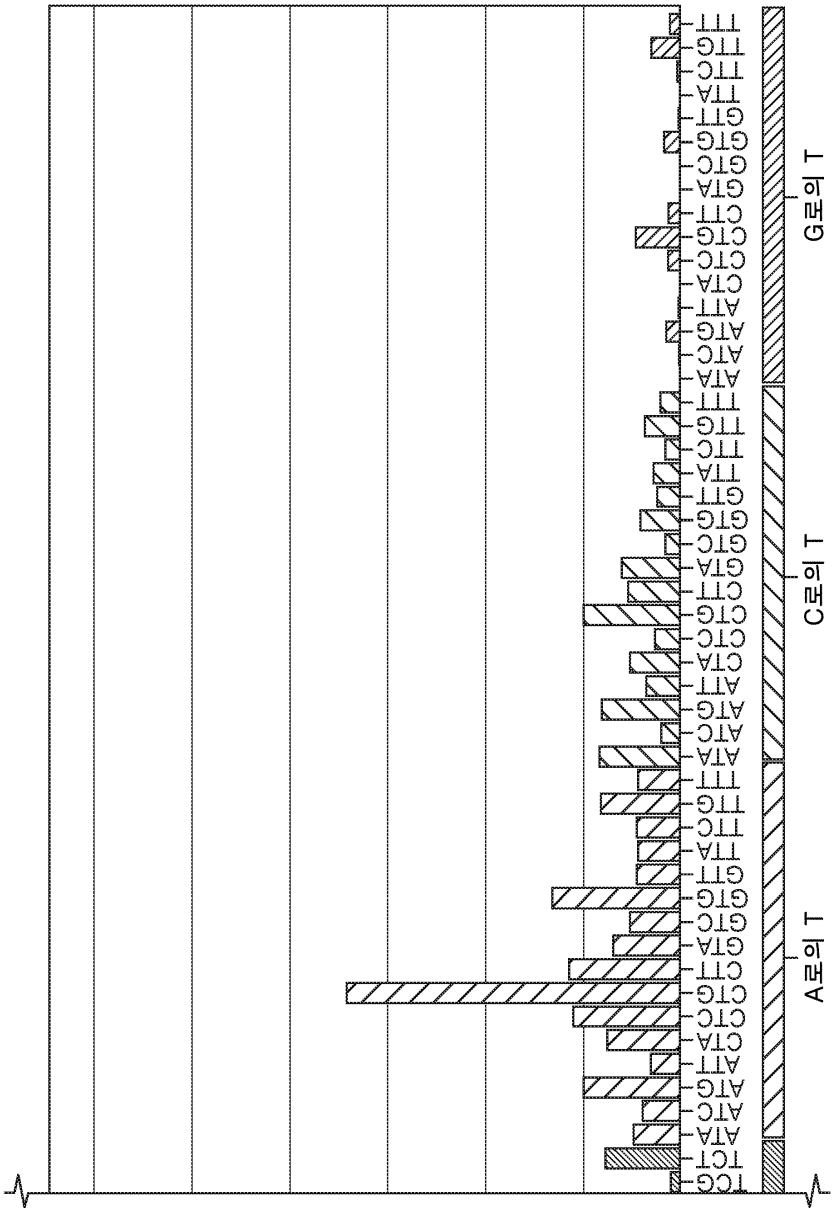
도면17ab


TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
TCT
TCG
TCC
G로의 T
C로의 T
A로의 T

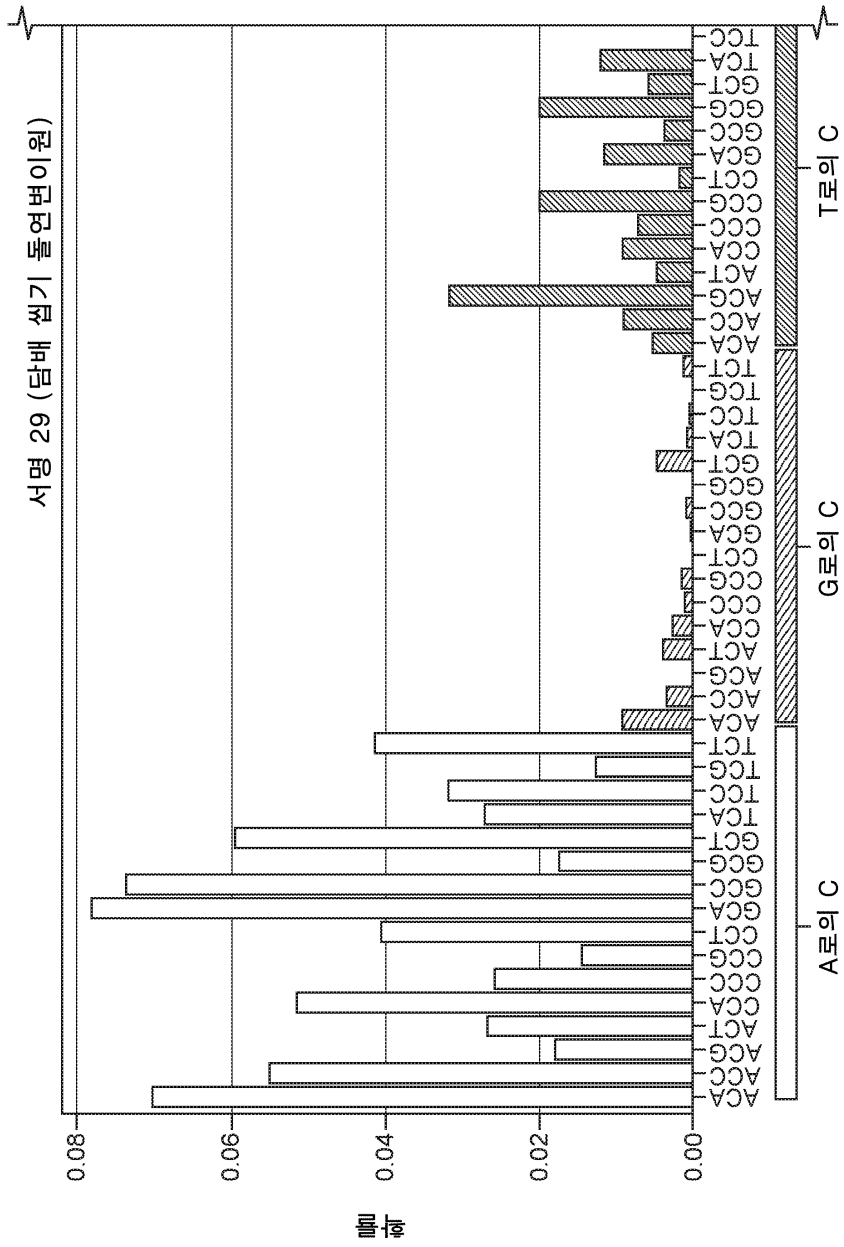
도면17ba


서명 4 (담배 흡연 돌연변이원)
확률
0.00
0.01
0.02
0.03
0.04
0.05
0.06
ACA
ACC
ACG
ACT
CCA
CCC
CCG
CCT
GCA
GCC
GCG
GCT
TCA
TCC
TCG
TCT
ACA
ACC
ACG
ACT
CCA
CCC
CCG
CCT
GCA
GCC
GCG
GCT
TCA
TCC
TCG
TCT
ACA
ACC
ACG
ACT
CCA
CCC
CCG
CCT
GCA
GCC
GCG
GCT
TCA
TCC
A로의 C
G로의 C
T로의 C

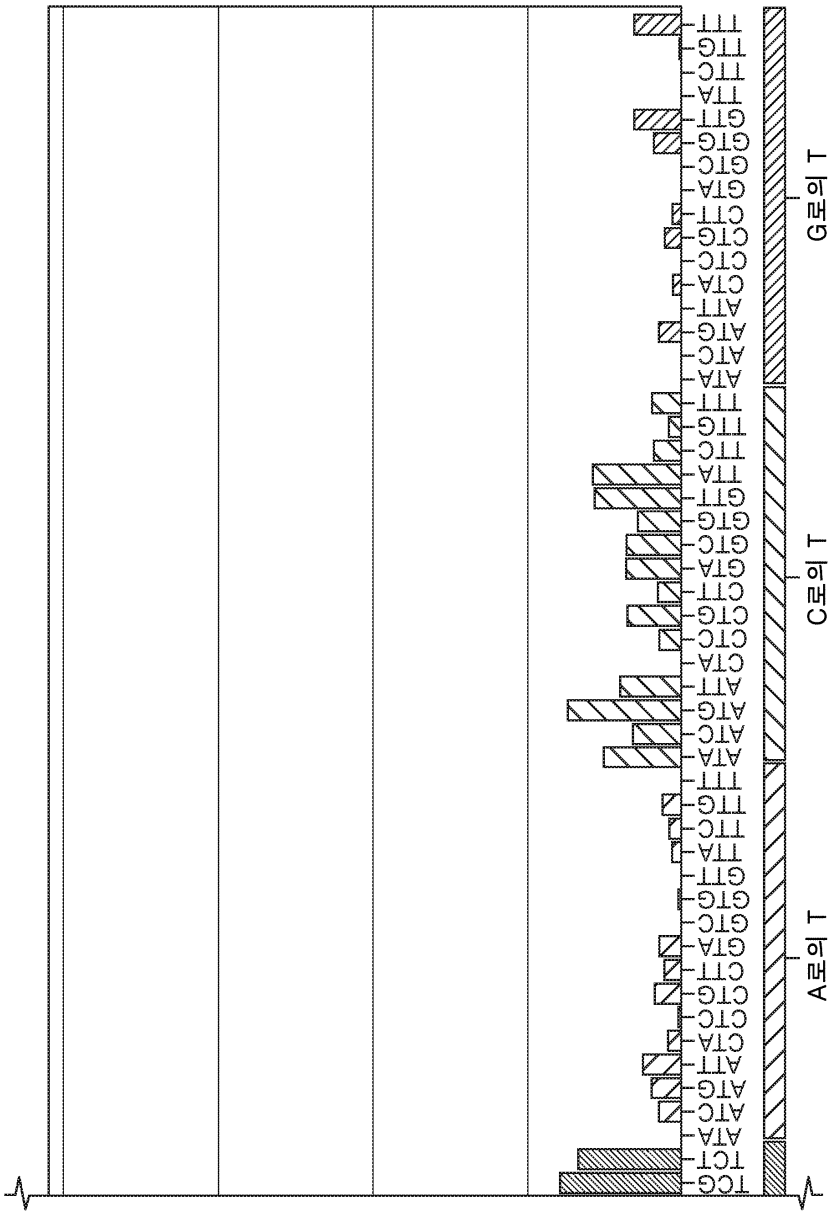
도면17bb


TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
TCT
TCG
G로의 T
C로의 T
A로의 T

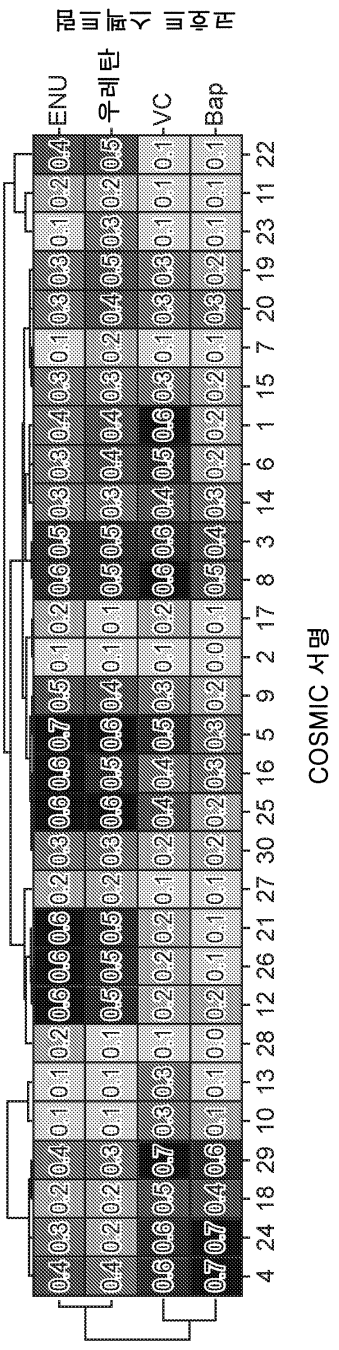
도면17ca


서명 29 (담배 씹기 돌연변이원)
확률
0.00
0.02
0.04
0.06
0.08
ACA
ACC
ACG
ACT
CCA
CCC
CCG
CCT
GCA
GCC
GCG
GCT
TCA
TCC
TCG
TCT
ACA
ACC
ACG
ACT
CCA
CCC
CCG
CCT
GCA
GCC
GCG
GCT
TCA
TCC
TCG
TCT
ACA
ACC
ACG
ACT
CCA
CCC
CCG
CCT
GCA
GCC
GCG
GCT
TCA
TCC
TCG
TCT
A로의 C
G로의 C
T로의 C

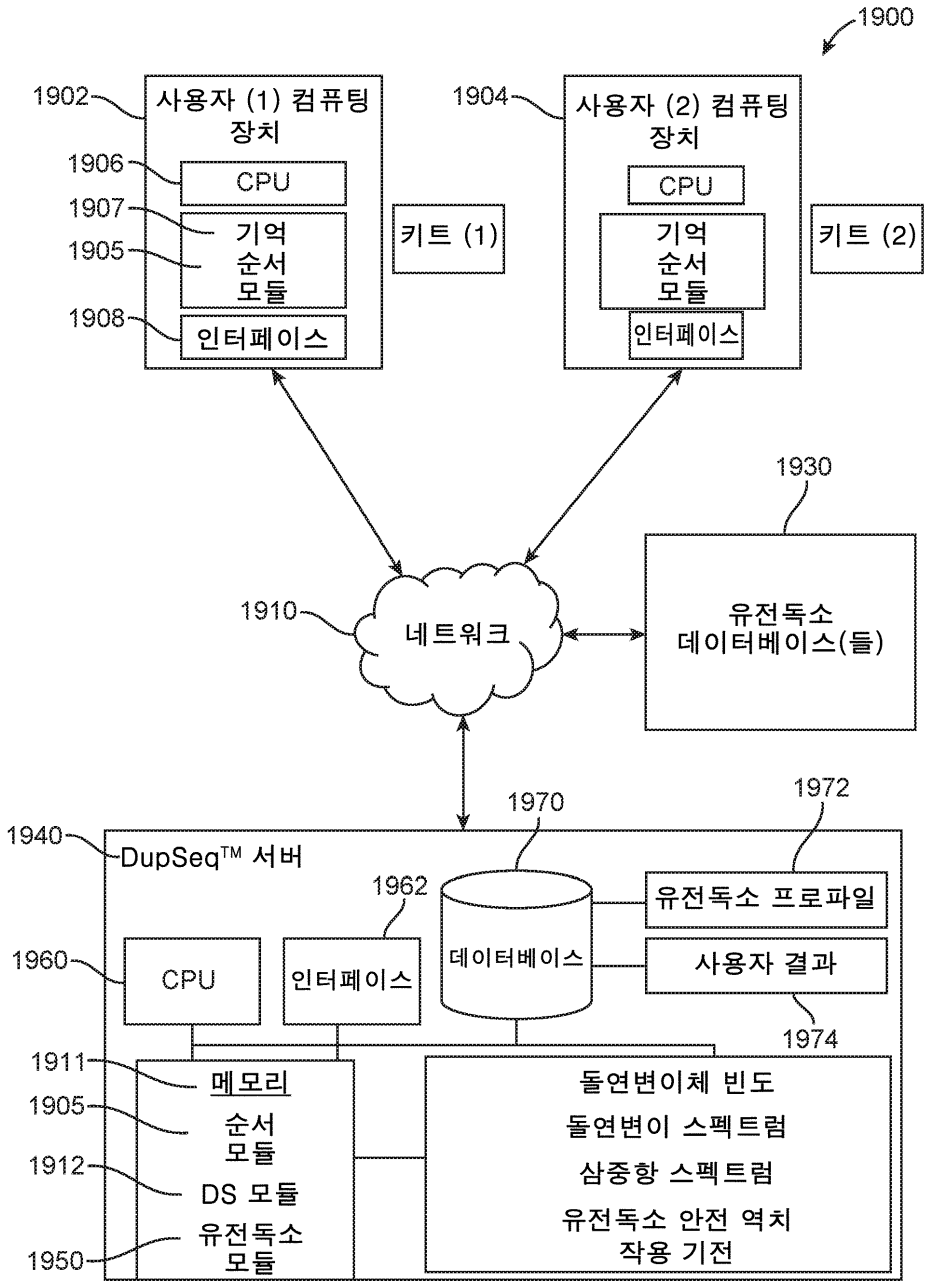
도면17cb


TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
TTT
TTG
TTC
TTA
GTT
GTG
GTC
GTA
CTT
CTG
CTC
CTA
ATT
ATG
ATC
ATA
TCT
TCG
G로의 T
C로의 T
A로의 T

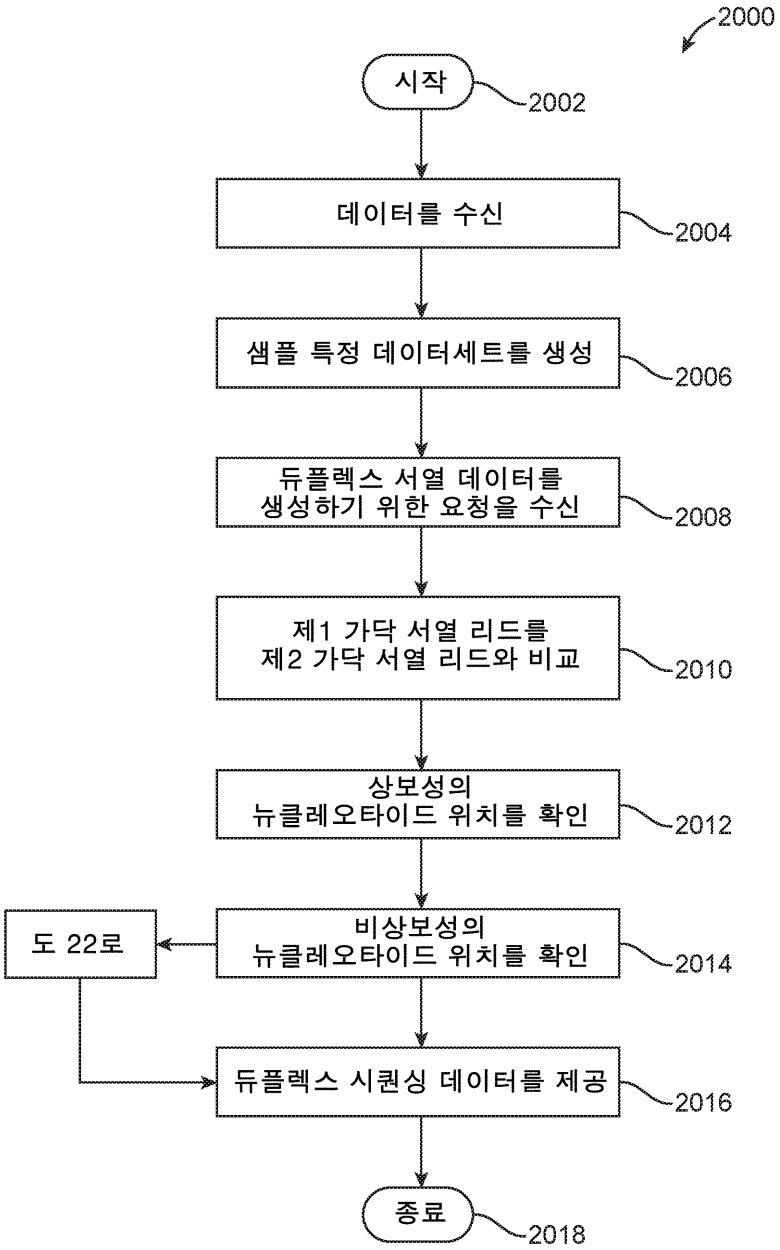
도면18



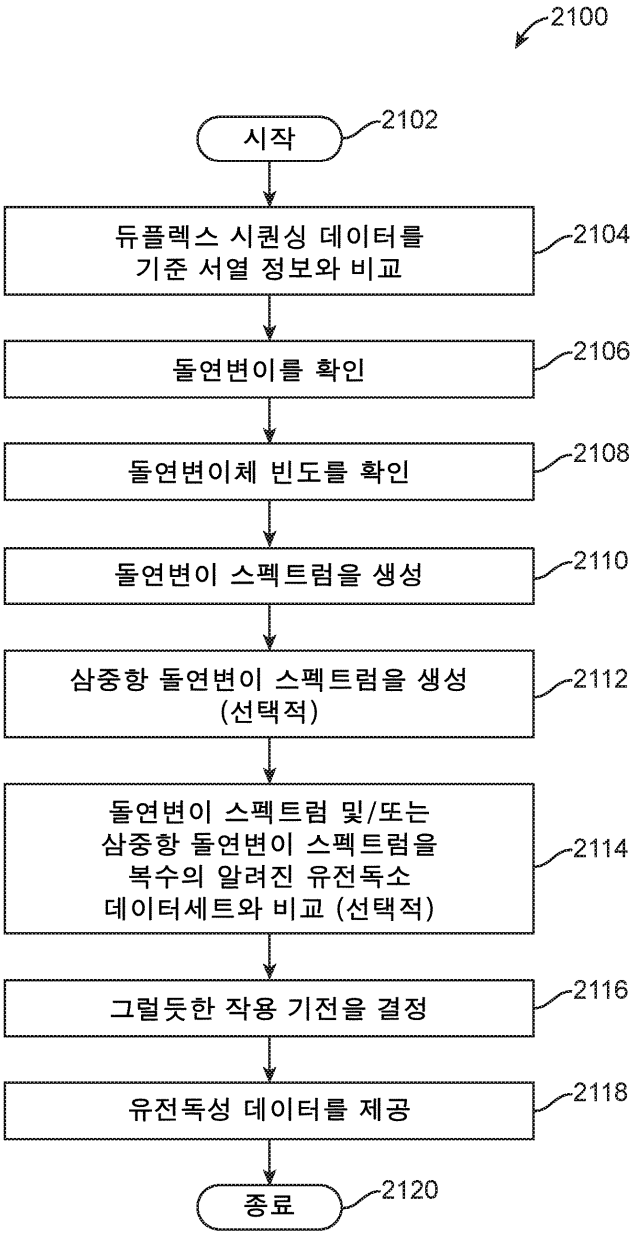
도면19


1900
1902
사용자 (1) 컴퓨팅 장치
1906
CPU
1907
기억
1905
순서
모듈
1908
인터페이스
키트 (1)
1904
사용자 (2) 컴퓨팅 장치
CPU
기억
순서
모듈
인터페이스
키트 (2)
1930
1910
네트워크
유전독소 데이터베이스(들)
1970
1972
1940
DupSeq™ 서버
1962
데이터베이스
유전독소 프로파일
1960
CPU
인터페이스
사용자 결과
1974
1911
메모리
1905
순서
모듈
1912
DS 모듈
1950
유전독소
모듈
돌연변이체 빈도
돌연변이 스펙트럼
삼중항 스펙트럼
유전독소 안전 역치
작용 기전

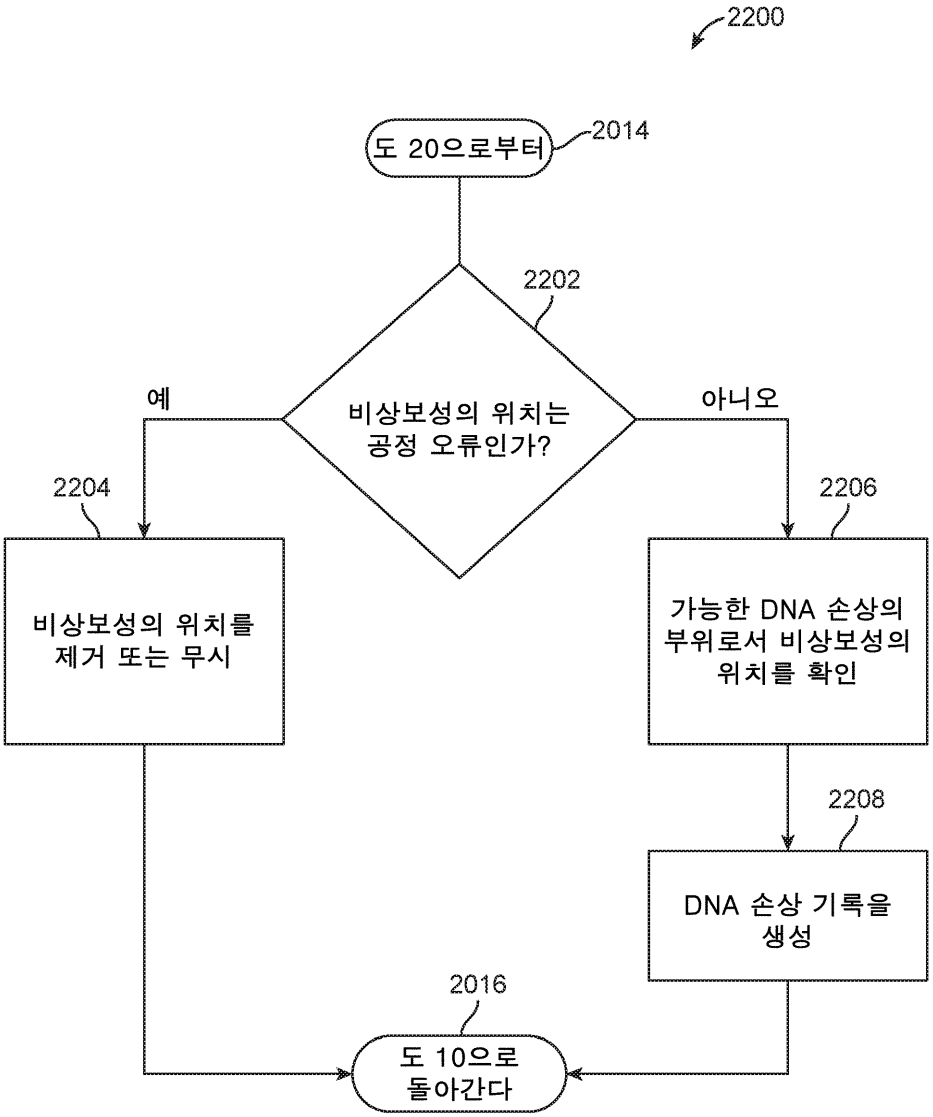
도면20


2000
시작
2002
데이터를 수신
2004
샘플 특정 데이터세트를 생성
2006
듀플렉스 서열 데이터를
생성하기 위한 요청을 수신
2008
제1 가닥 서열 리드를
제2 가닥 서열 리드와 비교
2010
상보성의
뉴클레오타이드 위치를 확인
2012
도 22로
비상보성의
뉴클레오타이드 위치를 확인
2014
듀플렉스 시퀀싱 데이터를 제공
2016
종료
2018

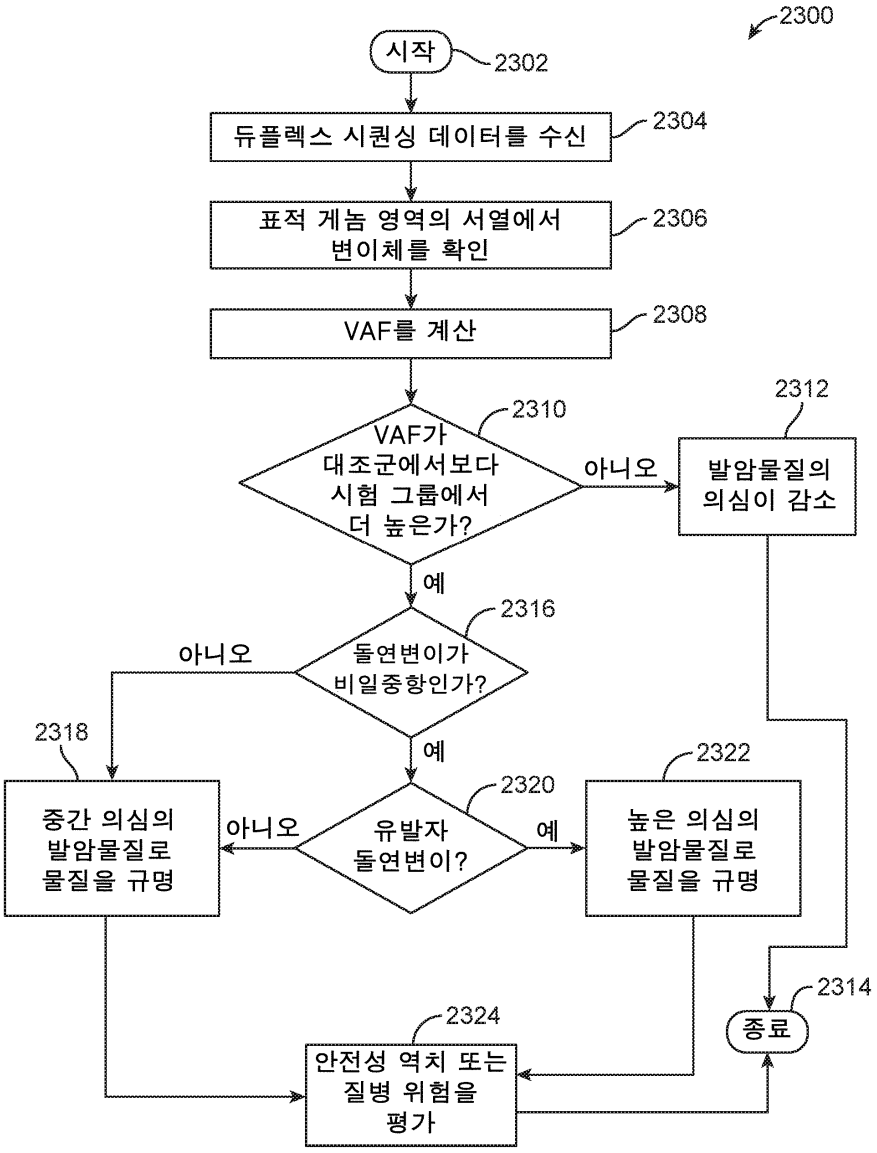
도면21


2100
시작
2102
듀플렉스 시퀀싱 데이터를
기준 서열 정보와 비교
2104
돌연변이를 확인
2106
돌연변이체 빈도를 확인
2108
돌연변이 스펙트럼을 생성
2110
삼중항 돌연변이 스펙트럼을 생성
(선택적)
2112
돌연변이 스펙트럼 및/또는
삼중항 돌연변이 스펙트럼을
복수의 알려진 유전독소
데이터세트와 비교 (선택적)
2114
그럴듯한 작용 기전을 결정
2116
유전독성 데이터를 제공
2118
종료
2120

도면22


2200
도 20으로부터
2014
2202
비상보성의 위치는 공정 오류인가?
예
아니오
2204
비상보성의 위치를 제거 또는 무시
2206
가능한 DNA 손상의 부위로서 비상보성의 위치를 확인
2208
DNA 손상 기록을 생성
2016
도 10으로 돌아간다

도면23


2300
시작
2302
듀플렉스 시퀀싱 데이터를 수신
2304
표적 게놈 영역의 서열에서 변이체를 확인
2306
VAF를 계산
2308
VAF가 대조군에서보다 시험 그룹에서 더 높은가?
2310
아니오
2312
발암물질의 의심이 감소
예
돌연변이가 비일중항인가?
2316
아니오
2318
예
중간 의심의 발암물질로 물질을 규명
유발자 돌연변이?
2320
아니오
예
2322
높은 의심의 발암물질로 물질을 규명
종료
2314
안전성 역치 또는 질병 위험을 평가
2324