



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2018-0087244  
(43) 공개일자 2018년08월01일

(51) 국제특허분류(Int. Cl.)  
G06F 19/24 (2011.01) G06F 19/26 (2011.01)  
(52) CPC특허분류  
G06F 19/24 (2013.01)  
G06F 19/26 (2013.01)  
(21) 출원번호 10-2018-7013060  
(22) 출원일자(국제) 2018년10월12일  
심사청구일자 없음  
(85) 번역문제출일자 2018년05월08일  
(86) 국제출원번호 PCT/US2016/056692  
(87) 국제공개번호 WO 2017/066357  
국제공개일자 2017년04월20일  
(30) 우선권주장  
62/240,494 2015년10월12일 미국(US)

(71) 출원인  
난토믹스, 엘엘씨  
미국 캘리포니아 컬버 시티 제퍼슨 블리바드 9920  
(우:90232)  
(72) 발명자  
웅유엔, 앤드류  
미국 95112 캘리포니아 산호세 사쿠라 드라이브  
929  
니아치, 카이반  
미국 90049 캘리포니아 로스앤젤레스 키오와 드라이브 #202 11928  
(뒷면에 계속)  
(74) 대리인  
특허법인에이아이피

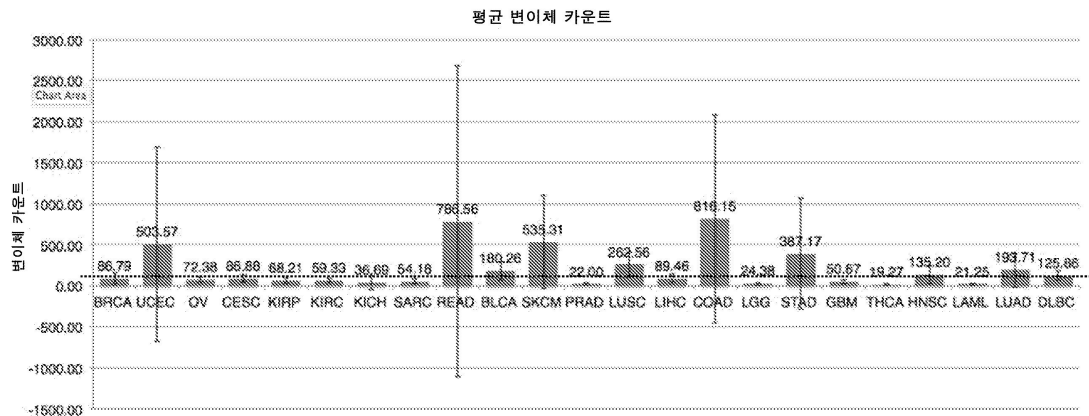
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 체크포인트 억제제에 대한 감도를 예측하는 MSI 및 네오에피토프를 발견하기 위한 시스템, 조성물 및 방법

(57) 요약

체크포인트 억제제에 대한 종양의 치료 반응을 예측할 수 있는 시스템 및 방법이 제공된다. 하나의 예시적인 양태에서, 상기 치료 반응은 비교적 다수의 환자- 및 종양-특이적 면역학적으로 가시적인 네오에피토프와 직접 연관된다. 상기 네오에피토프를 코딩하는 핵산의 특정 돌연변이 패턴은 치료 반응을 추가로 나타낼 수 있다.

대표도



(72) 발명자

**순-시웅, 패트릭**

미국 90232 캘리포니아 컬버 시티 제퍼슨 블리바드  
9920

**벤츠, 스티븐 찰스**

미국 95062 캘리포니아 산타크루즈 파크 웨이 436

**라비자데호, 샤흐루즈**

미국 90077 캘리포니아 로스앤젤레스 앤젤로 드라  
이브 2816

## 명세서

### 청구범위

#### 청구항 1

면역요법을 사용하여 암 치료를 개선하는 방법으로서,

환자로부터 종양 조직 및 매칭된 정상 조직으로부터의 오믹스(omics) 데이터를 획득하고, 상기 오믹스 데이터를 사용하여 복수의 미스센스(missense)-기반 환자- 및 종양-특이적 네오에피토프(neoepitope)를 결정하는 단계;

상기 환자- 및 종양-특이적 네오에피토프 중으로부터 발현된 네오에피토프를 동정하는 단계;

상기 발현된 네오에피토프를 여과하여 HLA-매칭된 네오에피토프를 획득하고, 상기 HLA-매칭된 네오에피토프를 정량화하는 단계; 및

상기 HLA-매칭된 네오에피토프의 양이 소정의 임계량을 초과한 것으로 결정되는 경우 환자에게 체크포인트 억제제(checkpoint inhibitor)를 투여하는 단계

를 포함하는, 방법.

#### 청구항 2

청구항 1에 있어서, 상기 발현된 네오에피토프를 여과하는 단계는, 복수의 구별되는 개별적인 네오에피토프 서열을 사용하여 상기 네오에피토프 각각에 대해 수행되며, 여기서 변화된 아미노산은 상기 네오에피토프 서열 내에 구별되는 위치를 갖는, 방법.

#### 청구항 3

청구항 2에 있어서, 상기 개별적인 네오에피토프 서열은 7 내지 20개의 아미노산 길이를 갖는, 방법.

#### 청구항 4

청구항 1에 있어서, 상기 여과 단계는, 단일 뉴클레오타이드 다형성, 짧은 결실 및 삽입 다형성, 마이크로위성(microsatellite) 마커, 짧은 탠덤 반복, 이중접합 서열, 다중 뉴클레오타이드 다형성 및 명명된 변이체로 이루어진 군으로부터 선택되는 선택적으로 알려진(a priori known) 분자 변이 중 적어도 하나에 의해 여과하는 단계를 추가로 포함하는, 방법.

#### 청구항 5

청구항 1에 있어서, 상기 여과 단계는, 환자의 적어도 하나의 MHC 클래스 I 하위-유형(sub-type) 및 적어도 하나의 MHC 클래스 II 하위-유형에 대한 상기 발현된 네오에피토프의 친화도의 결정 단계를 포함하는, 방법.

#### 청구항 6

청구항 1에 있어서, 상기 여과 단계는 상기 발현된 네오에피토프의 발현 수준의 결정 단계를 추가로 포함하는, 방법.

#### 청구항 7

청구항 1에 있어서, 상기 HLA-매칭된 네오에피토프는, 환자의 적어도 하나의 MHC 클래스 I 하위-유형 또는 적어도 하나의 MHC 클래스 II 하위-유형에 대해 150 nM 이하의 친화도를 갖는, 방법.

#### 청구항 8

청구항 1에 있어서, 상기 HLA-매칭된 네오에피토프를 정량화하는 단계는, 환자의 적어도 하나의 MHC 클래스 I 하위-유형 또는 적어도 하나의 MHC 클래스 II 하위-유형에 대한 상기 네오에피토프의 친화도의 정량화 및 HLA-매칭된 네오에피토프의 총 수의 결정 단계를 포함하는, 방법.

## 청구항 9

청구항 1에 있어서, 돌연변이 시그니처(signature)에 의해 상기 HLA-매칭된 네오에피토프를 여과하는 단계를 추가로 포함하는 방법.

## 청구항 10

청구항 9에 있어서, 상기 돌연변이 시그니처가, UV-유도된 DNA 손상 또는 흡연-유도된 DNA 손상에 대한 시그니처 특징인, 방법.

## 청구항 11

청구항 1에 있어서, 상기 HLA-매칭된 네오에피토프의 소정의 임계량은 적어도 100개의 HLA-매칭된 네오에피토프인, 방법.

## 청구항 12

청구항 11에 있어서, 상기 적어도 100개의 HLA-매칭된 네오에피토프는, 환자의 적어도 하나의 MHC 클래스 I 하위-유형 또는 적어도 하나의 MHC 클래스 II 하위-유형에 대해 150 nM 이하의 친화도를 갖는, 방법.

## 청구항 13

청구항 1에 있어서, 상기 오믹스 데이터를 사용하여 상기 질환 조직 내의 마이크로위성 불안정성(MSI)을 검출하는 단계를 추가로 포함하는 방법.

## 청구항 14

청구항 1에 있어서, 상기 오믹스 데이터를 사용하여 상기 질환 조직 내의 결함성 부정합 교정(MMR)을 검출하는 단계를 추가로 포함하는 방법.

## 청구항 15

청구항 1에 있어서, 상기 체크포인트 억제제가 CTLA-4 억제제 또는 PD-1 억제제인, 방법.

## 청구항 16

체크포인트 억제제에 대한 종양의 양성 치료 반응을 예측하는 방법으로서,

환자로부터 종양 조직 및 매칭된 정상 조직으로부터의 오믹스 데이터를 획득하고, 상기 오믹스 데이터를 사용하여 복수의 미스센스-기반 환자- 및 종양-특이적 네오에피토프를 결정하는 단계;

상기 환자- 및 종양-특이적 네오에피토프 중으로부터 발현된 네오에피토프를 동정하는 단계;

상기 발현된 네오에피토프를 여과하여 HLA-매칭된 네오에피토프를 획득하고, 상기 HLA-매칭된 네오에피토프를 정량화하는 단계; 및

상기 HLA-매칭된 네오에피토프의 양이 소정의 임계량을 초과했는지 확인한 후, 상기 종양이 상기 체크포인트 억제제에 의한 치료에 반응성인 것을 결정하는 단계를 포함하는, 방법.

## 청구항 17

청구항 16에 있어서, 돌연변이 시그니처에 의해 상기 HLA-매칭된 네오에피토프를 여과하는 단계를 추가로 포함하는 방법.

## 청구항 18

청구항 16에 있어서, 상기 오믹스 데이터를 사용하여 상기 질환 조직 내의 마이크로위성 불안정성(MSI) 및 결함성 부정합 교정(MMR) 중 적어도 하나를 검출하는 단계를 추가로 포함하는 방법.

## 청구항 19

체크포인트 억제제에 대한 종양의 양성 치료 반응을 예측하는 방법으로서,

환자로부터 종양 조직 및 매칭된 정상 조직으로부터의 오믹스 데이터를 획득하고, 상기 오믹스 데이터를 사용하여 복수의 미스센스-기반 환자- 및 종양-특이적 네오에피토프를 결정하는 단계;

상기 환자- 및 종양-특이적 네오에피토프 중으로부터 발현된 네오에피토프를 동정하는 단계;

상기 발현된 네오에피토프를 여과하여 HLA-매칭된 네오에피토프를 획득하고, 상기 HLA-매칭된 네오에피토프를 정량화하는 단계;

상기 정량화된 HLA-매칭된 네오에피토프에 대한 돌연변이 시그니처를 동정하는 단계; 및

상기 체크포인트 억제제에 대한 상기 종양의 양성 치료 반응에 대한 결정인자로서 네오에피토프의 양 및 돌연변이 시그니처를 사용하는 단계를 포함하는, 방법.

## 청구항 20

청구항 19에 있어서, 상기 돌연변이 시그니처가, UV-유도된 DNA 손상 또는 흡연-유도된 DNA 손상에 대한 특징인, 방법.

## 발명의 설명

### 기술 분야

[0001] 본원은 2015년 10월 12일자로 출원된 미국 가출원 제62/240494호를 우선권으로 주장하며, 이를 본원에 참고로 인용한다.

[0002] 기술분야

[0003] 본 발명의 분야는 치료 옵션을 예측하기 위한 오믹스(omics) 데이터의 컴퓨터 분석으로서, 특히 하나 이상의 체크포인트 억제제에 대한 종양의 양성 치료 반응을 예측하는 것에 관한 것이다.

### 배경 기술

[0004] 배경 기술은 본 발명을 이해하는 데 유용할 수 있는 정보를 포함한다. 여기에 제공된 임의의 정보가 선행 기술이거나 본원에 청구된 발명과 관련이 있거나 또는 구체적으로 또는 암시적으로 언급된 임의의 공보가 선행 기술임을 인정하는 것은 아니다.

[0005] 본원에 인용된 모든 공보 및 특허 출원은 각각의 개별 공보 또는 특허 출원이 구체적으로 및 개별적으로 참조로 인용되도록 지시된 것과 동일한 정도로 본원에 참고로 인용된다. 인용된 참고문헌에서의 용어의 정의 또는 사용이 일관성이 없거나 본원에 제공된 해당 용어의 정의와 상반되는 경우에는 본원에 제공된 해당 용어의 정의가 적용되며 참고문헌에서의 해당 용어의 정의는 적용되지 않는다.

[0006] 암 면역요법은 일부 환자에서 놀라운 반응을 이끌어 냈지만, 많은 환자가 면역요법 반응자에 비해 명백한 동일한 유형의 암을 가지고 있음에도 불구하고 반응하지 못했다. 이러한 실패에 대한 하나의 가능한 설명은 면역 시스템의 다양한 이펙터(effector) 세포가 하나 이상의 억제 조절 경로와 상호작용하는 화합물(체크포인트 억제제)에 의해 차단될 수 있다는 것이다. 특히, 일부 종양 세포는 억제 조절 경로를 이용하여 면역 시스템에 의한 검출 및 파괴를 피할 수 있다. 다른 성분들 중에서도 PD-1과 CTLA-4는 면역 반응의 억제와 관련하여 가장 많이 연구된 수용체이며, 현재 이 수용체의 활성화를 차단하는 특정 약물이 이용가능하게 되었다. PD-1(예컨대 니볼루맵 및 펌브롤리주맵) 및 CTLA-4(예컨대 이필리무맵)에 대한 항체가 예를 들어, 흑색종, 신장 세포 암종, 비소 세포 폐암 및 다양한 다른 종양 유형의 일부 경우에 중요한 임상 반응을 나타냈다. 불행히도, 모든 유형의 암이 체크포인트 억제제를 사용한 치료에도 똑같이 잘 반응하는 것은 아니다. 더욱이, 동일한 유형의 암 내에서도, 체크포인트 억제제에 대한 양성 반응 예측가능성은 파악하기 어렵다.

[0007] 또한, 부정합 교정(mismatch repair, MMR)의 손실은 직접적으로는 세포가 DNA 손상을 감지하고 아포토시스를 활성화하는 능력을 손상시킴으로써 그리고 간접적으로는 게놈 전체의 돌연변이 속도를 증가시킴으로써 약물 내성을 종종 초래한다. 예를 들어, MMR-결핍 세포는 다양한 메틸화/알킬화제, 특정 백금-함유 약물, 항대사제 및 토포아이스머라제 II 억제제에 대해 내성인 것으로 보고되었다. 또한, MMR-결핍 세포는 증가된 돌연변이 속도를 가지며, 이는 종종 마이크로위성 불안정성(microsatellite instability, MSI)으로 표현된다. 이러한 세포는 종종 통상적인 약물 치료에 덜 민감하므로, 면역요법이 바람직하다. 그러나 MSI 종양에 대한 면역요법의 효능은

체크포인트 억제제에 의해 MSI 종양을 치료하는 것과 마찬가지로 예측할 수 없다.

[0008] 따라서, 환자- 또는 암-특이적 네오에피토프를 표적으로 하는 치료와 단독으로 또는 함께 체크포인트 억제제에 의해 암 치료의 효능을 평가하는 데 도움이 되는 예후 도구를 갖는 것이 바람직할 것이다.

### 발명의 내용

[0009] 본 발명의 대상은 체크포인트 억제제에 대한 종양의 치료 반응을 예측하기 위한 다양한 장치, 시스템 및 방법에 관한 것이다. 특히 바람직한 양태에서, HLA-매칭된 네오에피토프의 존재 및 양(및 패턴)은 체크포인트 억제제에 의한 치료 성공 가능성을 위한 대리 지표로서 사용된다.

[0010] 본 발명의 대상 중 하나의 양태에서, 본 발명자들은 종양 조직 및 매칭된 정상 조직으로부터 환자의 오믹스 데이터(omics data)를 획득하는 단계 및 상기 오믹스 데이터를 이용하여 복수의 미스센스(missense) 기반 환자- 및 종양-특이적 네오에피토프를 결정하는 또 다른 단계를 포함하는 면역요법을 사용하여 암 치료를 개선하는 방법을 고찰한다. 추가 단계에서, 네오에피토프를 여과하고 정량화하여 HLA-매칭된 네오에피토프를 획득한다. 그런 다음, HLA-매칭된 네오에피토프의 양이 소정의 임계량을 초과한 것으로 판정되는 경우 체크포인트 억제제(예컨대 CTLA-4 억제제 또는 PD-1 억제제)를 환자에게 투여한다.

[0011] 가장 전형적으로, 네오에피토프를 여과하는 단계는 복수의 분명한(distinct) 개별적인 네오에피토프 서열(예컨대 각각 7 내지 20개의 아미노산 길이를 갖는)을 사용하여 네오에피토프 각각에 대해 수행되며, 여기서 변화된 아미노산은 네오에피토프 서열 내에서 구별되는 위치를 갖는다. 또한, 여과 단계는 선형적으로 알려진(priori known) 분자 변이 예를 들어 단일 뉴클레오타이드 다형성, 짧은 결실 및 삽입 다형성, 마이크로위성 마커, 짧은 tandem 반복, 이중접합 서열, 다중뉴클레오타이드 다형성 또는 명명된 변이체에 의해 여과하는 단계를 더 포함할 수 있는 것으로 고려된다. 추가로 고려되는 양태에서, 여과 단계는 또한 환자의 적어도 하나의 MHC 클래스 I 하위-유형 및/또는 적어도 하나의 MHC 클래스 II 하위-유형에 대한 네오에피토프의 친화도의 결정 단계를 포함할 수 있으며, 또한 네오에피토프의 발현 수준의 결정 단계를 포함할 수 있다.

[0012] 본 발명의 대상을 한정하는 것은 아니지만, HLA-매칭된 네오에피토프는 환자의 적어도 하나의 MHC 클래스 I 하위-유형 및/또는 적어도 하나의 MHC 클래스 II 하위-유형에 대해 150 nM 이하의 친화도를 가질 것으로 고려된다. 예를 들어, HLA-매칭된 네오에피토프를 정량화하는 단계는 환자의 적어도 하나의 MHC 클래스 I 하위-유형 또는 적어도 하나의 MHC 클래스 II 하위-유형에 대한 네오에피토프의 친화도의 정량화(예컨대 500 nM 이하 또는 250 nM 이하 또는 150 nM 이하 또는 50 nM 이하), 및 HLA-매칭된 네오에피토프의 총 수의 결정(예컨대 적어도 50, 또는 적어도 100, 또는 적어도 200, 또는 적어도 300 등)을 포함할 수 있는 것으로 고려된다.

[0013] 또한, 이러한 방법은 돌연변이 시그니처(예컨대 UV-유도된 DNA 손상 또는 흡연-유도된 DNA 손상에 대한 시그니처 특징)에 의해 HLA-매칭된 네오에피토프를 여과하는 단계를 더 포함할 수 있다. 원하는 경우, 본원에 제시된 방법은 질환 조직에서의 마이크로위성 불안정성(MSI) 및/또는 결함성 부정합 교정(MMR)을 검출하기 위해 오믹스 데이터를 사용하는 단계를 더 포함할 수 있음이 또한 고려된다.

[0014] 따라서, 다른 관점에서 볼 때, 본 발명자들은 또한 체크포인트 억제제에 대한 종양의 양성 치료 반응을 예측하는 방법을 고찰한다. 이러한 방법은 전형적으로 종양 조직 및 매칭된 정상 조직으로부터의 오믹스 데이터를 환자로부터 얻고, 오믹스 데이터를 사용하여 다수의 미스센스-기반 환자- 및 종양-특이적 네오에피토프를 결정하는 단계를 포함할 것이다. 추가 단계에서, 네오에피토프를 여과하고 정량화하여 HLA-매칭된 네오에피토프를 획득한다. 또 다른 단계에서, HLA-매칭된 네오에피토프의 양이 소정의 임계량을 초과했는지 확인한 후 종양이 체크포인트 억제제를 사용한 치료에 반응한다는 것을 결정한다.

[0015] 유사하게, 본 발명자들은 또한 체크포인트 억제제에 대한 종양의 양성 치료 반응을 예측하는 방법을 고찰하고, 여기서 종양 조직 및 매칭된 정상 조직으로부터의 오믹스 데이터를 환자로부터 얻고, 그런 다음 오믹스 데이터를 사용하여 다수의 미스센스-기반 환자- 및 종양-특이적 네오에피토프를 결정한다. 또 다른 단계에서, 네오에피토프를 여과하여 HLA-매칭된 네오에피토프를 얻고 HLA-매칭된 에피토프를 정량화한다. 추가 단계에서, 정량화된 HLA-매칭된 네오에피토프에 대한 돌연변이 시그니처를 결정하고, 그런 다음 네오에피토프 및 돌연변이 시그니처를 체크포인트 억제제에 대한 종양의 양성 치료 반응에 대한 결정인자로서 사용한다.

[0016] 본 발명의 대상의 다양한 목적, 특징, 양태 및 장점은 첨부된 도면과 함께 바람직한 구현예에 대한 하기 상세한 설명으로부터 명백해질 것이며, 도면에서 유사한 번호는 유사한 구성요소를 나타낸다.

## 도면의 간단한 설명

- [0017] 도 1은 각종 암에 따른 네오에피토프 빈도 및 코딩 변이 빈도의 예시적 그래픽 표현뿐만 아니라 각종 암에 따른 고유의 네오에피토프 빈도의 그래픽 표현이다.
- 도 2는 네오에피토프 빈도 및 이들의 암에서의 RNA 내로의 발현의 예시적인 그래픽 표현이다.
- 도 3은 각종 암에 대한 HLA-매칭된 네오에피토프의 빈도의 예시적인 그래픽 표현이다.
- 도 4는 결합된 암에 대한 단일 암 유형(TNBC) 및 HLA-제한된 네오에피토프에 대한 다양한 여과 공정 효과의 예시적인 그래픽 표현이다. 또한 암 구동 유전자 및 비-암 유전자에 관한 네오에피토프의 위치가 예시적으로 도시된다.

## 발명을 실시하기 위한 구체적인 내용

- [0018] 본 발명자들은 HLA-매칭된 환자- 및 암-특이적 네오에피토프가 하나 이상의 체크포인트 억제제를 갖는 종양의 치료 성공 가능성에 대한 대리 지표로서 사용될 수 있음을 발견하였으며, 여기서 이러한 네오에피토프의 수는 임계치 수준보다 높다. 이러한 비교적 다수의 발현된 네오에피토프는 MMR 및/또는 MSI를 비롯한 다양한 원인에 기인하거나 연관될 수 있다. 또한, 본 발명자들은 환자- 및 암-특이적 네오에피토프가 치료 성능 가능성의 지표일 수 있으며 여기서 네오에피토프는 고유한 종양 특이적 항원을 일으킬 수 있는 특정 돌연변이 패턴(예를 들어, UV-유도된 DNA 손상 또는 흡연-유도된 DNA 손상)과 관련이 있음을 발견하였다.
- [0019] 네오에피토프는 고유한 종양 특이적 항원을 생성하는 종양 세포에서 무작위 돌연변이를 발현한 것을 특징으로 할 수 있다. 따라서, 다른 관점에서 볼 때, 네오에피토프는 침묵 및 다른 비-연관 돌연변이를 제거시키는 제1 내용 필터로서 작용할 수 있는 돌연변이의 유형(예컨대 결실, 삽입, 전환, 전이, 전좌) 및 영향(예컨대 넌센스(non-sense), 미스센스, 프레임 이동 등)을 고려하여 식별할 수 있다. 또한, 네오에피토프 서열은 비교적 짧은 길이(예컨대 7 내지 11mer)의 서열 신장으로 정의될 수 있으며, 여기서 이러한 신장은 아미노산 서열의 변화(들)를 포함할 것이라는 것을 이해해야 한다. 가장 전형적으로, 변화된 아미노산은 중앙 아미노산 위치 또는 그 근처에 있을 것이다. 예를 들어, 전형적인 네오에피토프는  $A_4-N-A_4$ , 또는  $A_3-N-A_5$ , 또는  $A_2-N-A_7$ , 또는  $A_5-N-A_3$ , 또는  $A_7-N-A_2$ 의 구조를 가질 수 있으며, 여기서 A는 단백질생성 아미노산이고 N은 변화된 아미노산(야생형 또는 매칭된 정상군에 대해)이다. 예를 들어, 본원에서 고려되는 네오에피토프 서열은 비교적 짧은 길이(예컨대 5 내지 30mer, 보다 전형적으로는 7 내지 11mer 또는 12 내지 25mer)의 서열 신장을 포함하며, 여기서 이러한 신장은 아미노산 서열의 변화(들)를 포함한다.
- [0020] 따라서, 단일 아미노산 변화가, 변화된 아미노산의 위치에 따라, 변화된 아미노산을 포함하는 다수의 네오에피토프 서열에 제시될 수 있다는 것을 이해해야 한다. 유리하게는, 이러한 서열 가변성은 네오에피토프의 다중 선택을 가능하게 하여 하나 이상의 바람직한 형질(예컨대 환자 HLA-유형에 대한 가장 높은 친화도, 가장 높은 구조적 안정성 등)을 기초로 선택될 수 있는 잠재적으로 유용한 표적의 수를 증가시킨다. 가장 전형적으로, 네오에피토프는 2 내지 50개의 아미노산, 보다 전형적으로는 5 내지 30개의 아미노산, 가장 전형적으로는 9 내지 15개의 아미노산의 길이를 갖도록 계산되며, 변화된 아미노산은 바람직하게는 중앙에 위치하거나 그렇지 않으면 MHC와의 결합을 향상시키는 방식으로 위치한다. 예를 들어, 에피토프가 MHC-I 복합체에 의해 제공되는 경우, 전형적인 네오에피토프 길이는 약 8 내지 11개의 아미노산인 반면에, MHC-II 복합체를 통해 제시된 전형적인 네오에피토프 길이는 약 13 내지 17개의 아미노산을 가질 것이다. 용이하게 이해되는 바와 같이, 네오에피토프에서의 변화된 아미노산의 위치는 중심이 아니기 때문에, 실제 펩타이드 서열 및 네오에피토프의 실제 위상은 상당히 다를 수 있다.
- [0021] 물론, 네오에피토프의 동정 또는 발견은 새로운 생검, 동결 또는 그렇지 않으면 보존된 조직 또는 세포 샘플, 순환형 종양 세포, 엑소솜, 다양한 체액(및 특히 혈액)을 비롯한 다양한 생물학적 물질로 시작될 수 있다는 것을 이해해야 한다. 따라서, 적절한 �믹스 분석 방법은 핵산 시퀀싱, 특히 DNA 상에서 작동하는 NGS 방법(예를 들어, 일루미나 시퀀싱, 이온 토렌트 시퀀싱, 454 파이로시퀀싱, 나노기공 시퀀싱 등), RNA 시퀀싱(예를 들어, RNAseq, 역전사 기반 시퀀싱 등), 및 단백질 시퀀싱 또는 질량 분광학 기반 시퀀싱(예를 들어, SRM, MRM, CRM 등)을 포함한다.
- [0022] 이와 같이, 특히 핵산 기반 시퀀싱의 경우, 종양 조직의 높은 처리량의 게놈 시퀀싱이 특히 네오에피토프의 신속한 동정을 허용할 것이라는 점을 인식해야 한다. 그러나, 이렇게 획득된 서열 정보가 표준 참조와 비교되는



경우, 이중접합뿐만 아니라 정상적으로 발생하는 환자 간 변화(예를 들어, SNP, 짧은 삽입-결실(indel), 다른 반복 횟수 등에 기인함)가 비교적 많은 수의 잠재적인 가양성(false positive) 네오에피토프를 초래한다. 특히, 환자의 종양 샘플이 동일한 환자의 매칭된 정상(즉, 비-종양) 샘플과 비교되는 경우, 이러한 부정확성은 제거될 수 있다.

[0023] 본 발명의 대상 중 하나의 특히 바람직한 양태에서, DNA 분석은 종양 및 매칭된 정상 샘플 둘 다의 전체 게놈 시퀀싱 및/또는 엑손 시퀀싱(전형적으로 적어도 10x, 보다 전형적으로는 적어도 20x의 커버리지(coverage) 깊이에서)에 의해 수행된다. 대안적으로, DNA 데이터는 또한 이전의 서열 결정으로부터 이미 확립된 서열 기록(예를 들어, SAM, BAM, FASTA, FASTQ 또는 VCF 파일)으로부터 제공될 수도 있다. 따라서, 데이터 세트는 처리되지 않거나 처리된 데이터 세트를 포함할 수 있고, 예시적인 데이터 세트는 BAMBAM 포맷, SAMBAM 포맷, FASTQ 포맷 또는 FASTA 포맷을 갖는 것들을 포함할 수 있다. 그러나, 데이터 세트가 BAMBAM 포맷 또는 BAMBAM 디프(diff) 대상으로 제공되는 것이 특히 바람직하다(예를 들어, US2012/0059670A1 및 US2012/0066001A1 참조). 또한, 데이터 세트는 동일한 환자의 종양 및 매칭된 정상 샘플을 반영하여 환자 및 종양 특이적 정보를 얻는다는 점에 유의해야 한다. 따라서, 종양을 유발하지 않는 유전적 배선 변이(예를 들어, 침묵 돌연변이, SNP 등)가 배제될 수 있다. 물론, 종양 샘플은 초기 종양, 치료 시작시 종양, 재발성 종양 또는 전이 부위 등에서 유래될 수 있다는 것을 인식해야 한다. 대부분의 경우, 환자의 매칭된 정상 샘플은 종양과 동일한 조직 유형의 혈액 또는 비-질환 조직일 수 있다.

[0024] 마찬가지로, 서열 데이터의 컴퓨터 분석은 다양한 방식으로 수행될 수 있다. 그러나, 가장 바람직한 방법에서, BAM 파일 및 BAM 서버를 사용하여 예를 들어 US 2012/0059670A1 및 US 2012/0066001A1에 개시된 바와 같이 종양 및 정상 샘플의 위치-유도 동시 정렬에 의해 인-실리코 분석이 수행된다. 이러한 분석은 가양성 네오에피토프를 유리하게 감소시키고 메모리 및 계산 자원에 대한 요구를 상당히 감소시킨다.

[0025] 컴퓨터에 지시된 임의의 언어는 서버, 인터페이스, 시스템, 데이터베이스, 에이전트, 피어(peer), 엔진, 컨트롤러, 또는 개별적으로 또는 집합적으로 작동하는 기타 유형의 컴퓨팅 장치를 포함하는 컴퓨팅 장치들의 임의의 적절한 조합을 포함하도록 관독되어야 한다는 점에 유의해야 한다. 컴퓨팅 장치는 유형의(tangible) 비-일시적인 컴퓨터 관독가능한 저장 매체(예를 들어, 하드 드라이브, 솔리드 스테이트 드라이브, RAM, 플래시, ROM 등) 상에 저장된 소프트웨어 명령을 실행하도록 구성된 프로세서를 포함하는 것으로 이해해야 한다. 소프트웨어 명령은 바람직하게는 개시된 장치와 관련하여 후술되는 바와 같은 역할, 책임 또는 다른 기능을 제공하도록 컴퓨팅 장치를 구성한다. 또한, 개시된 기술은 프로세서로 하여금 컴퓨터-기반 알고리즘, 프로세스, 방법 또는 다른 명령들의 구현과 관련된 개시된 단계들을 실행하게 하는 소프트웨어 명령들을 저장하는 비-일시적인 컴퓨터 관독가능한 매체를 포함하는 컴퓨터 프로그램 제품으로서 구현될 수 있다. 특히 바람직한 구현예에서, 다양한 서버, 시스템, 데이터베이스, 또는 인터페이스는 아마도 HTTP, HTTPS, AES, 공개-개인 키 교환, 웹 서비스 API, 공지의 금융 거래 프로토콜 또는 다른 전자 정보 교환 방법 기반의 표준화된 프로토콜 또는 알고리즘을 사용하여 데이터를 교환한다. 장치간 데이터 교환은 패킷-교환 네트워크, 인터넷, LAN, WAN, VPN, 또는 다른 유형의 패킷 교환 네트워크; 회선 교환 네트워크; 셀 교환 네트워크; 또는 다른 유형의 네트워크를 통해 수행될 수 있다.

[0026] 다른 관점에서 볼 때, 5 내지 25개 아미노산의 소정의 길이를 가지며 적어도 하나의 변화된 아미노산을 포함하는 서열의 환자- 및 암-특이적인 인-실리코 수집을 확립할 수 있다. 이러한 수집은 전형적으로 각각의 변화된 아미노산에 대해, 변화된 아미노산의 위치가 동일하지 않은 2개 이상, 3개 이상, 4개 이상, 5개 이상, 또는 6개 이상의 구성원을 포함할 것이다. 그런 다음, 이러한 수집은 하기에 보다 상세히 설명되는 바와 같이 (예를 들어, 하위-세포 위치, 전사/발현 수준, MHC-1 및/또는 II 친화도 등에 의해) 추가의 여과를 위해 사용될 수 있다.

[0027] 예를 들어, 종양 및 매칭된 정상 서열 데이터에 동기식 위치 유도 분석법을 사용하여, 본 발명자들은 이전에 BLCA, BRCA, CESC, COAD, DLBC, GBM, HNSC, KICH, KIRC, KIRP, LAML, LGG, LIHC, LUAD, LUSC, OV, PRAD, READ, SARC, SKCM, STAD, THCA, 및 UCEC과 같은 암 유형을 포함하는 각종의 암 및 환자로부터 다양한 암 네오에피토프를 동정하였다. 모든 네오에피토프 데이터는 국제 출원 제PCT/US16/29244호에서 찾을 수 있으며, 본원에 참고로 인용한다.

[0028] 암의 유형과 단계에 따라, 체크포인트 억제제를 환자에게 투여할 때, 동정된 모든 네오에피토프가 반드시 환자에게 치료상 동등하게 효과적인 반응을 일으키는 것은 아니라는 점에 유의해야 한다. 실제로, 일부의 네오에피토프만이 면역 반응을 생성한다는 것은 당업계에 잘 알려져 있다. 치료학적으로 바람직한 반응의 가능성을 높이



기 위해, 네오에피토프는 추가로 여과될 수 있다. 물론, 다운스트림 분석은 본원에서 제시된 방법의 목적을 위한 침묵 돌연변이를 고려할 필요가 없다는 것을 이해해야 한다. 그러나, 바람직한 돌연변이 분석은 돌연변이 유형(예를 들어 결손, 삽입, 전환, 전이, 전좌)뿐만 아니라 돌연변이 영향(예를 들어 넌센스, 미스센스 등)에 대한 정보를 제공하며, 침묵 돌연변이를 제거시키는 제1 내용 필터로서 작용할 수 있다. 예를 들어, 네오에피토프는 돌연변이가 프레임-이동, 넌센스 및/또는 미스센스 돌연변이인 추가 고려를 위해 선택될 수 있다.

[0029] 추가적인 여과 접근법에서, 네오에피토프는 또한 하위-세포 위치 파라미터에 대한 상세한 분석의 대상이 될 수 있다. 예를 들어, 네오에피토프 서열은 네오에피토프가 막 관련 위치를 갖는 것으로 확인된 경우(예를 들어, 세포의 세포막 외부에 위치하는 경우) 및/또는 인-실리코 구조 계산에 의해 네오에피토프가 용매에 노출되거나 구조적으로 안정한 에피토프(예컨대 J Exp Med 2014)를 나타낼 수 있는 경우에 추가 고려를 위해 선택될 수 있다.

[0030] 네오에피토프를 여과하는 것과 관련하여, 네오에피토프는 네오에피토프가 실제로 발현되는 것으로 밝혀진 오믹스(또는 다른) 분석에 사용하기에 특히 적합하다고 생각된다. 네오에피토프의 발현 및 발현 수준의 동정은 당 업계에 공지된 모든 방식으로 수행될 수 있으며, 바람직한 방법은 정량적 RNA(hnRNA 또는 mRNA) 분석 및/또는 정량적 프로테오믹스 분석을 포함한다. 가장 일반적으로, 네오에피토프를 포함하는 것에 대한 임계치 수준은 상응하는 매칭된 정상 서열의 발현 수준의 적어도 20%, 적어도 30%, 적어도 40%, 또는 적어도 50%의 발현 수준이 될 것이므로, (네오)에피토프가 면역 시스템에 잠재적으로 "가시적"인 것으로 보인다. 따라서, 오믹스 분석은 또한 유전자 발현 분석(전사체 분석)을 포함하여 돌연변이가 있는 유전자 발현 수준을 동정하는 것이 일반적인 것으로 바람직하다.

[0031] 당해 분야에 공지된 전사체 분석의 수많은 방법이 존재하며, 공지된 모든 방법이 본원에서의 사용에 적합한 것으로 간주된다. 예를 들어, 바람직한 물질은 mRNA 및 1차 전사체(hnRNA)를 포함하고, RNA 서열 정보는 동일한 환자의 종양 샘플 및 매칭된 정상(건강한) 샘플로부터 차례로 얻어지는 역전사된 폴리 A<sup>+</sup>-RNA로부터 얻을 수 있다. 마찬가지로, 폴리 A<sup>+</sup>-RNA는 전형적으로 전사체의 표현으로서 바람직하지만, 다른 형태의 RNA(hn-RNA, 비-폴리아데닐화된 RNA, siRNA, miRNA 등)도 본원에서의 사용에 적합하다고 간주됨을 유의해야 한다. 바람직한 방법은 정량적 RNA(hnRNA 또는 mRNA) 분석 및/또는 정량적 프로테오믹스 분석, 특히 예를 들어 RNAseq 등을 포함한다. 다른 양태에서, 다양한 대안적인 방법(예를 들어, 고상 하이브리드화-기반 방법)이 또한 적합하다고 여겨지지만, RNA 정량화 및 시퀀싱은 RNA-seq, qPCR 및/또는 rtPCR 기반 방법을 사용하여 수행된다. 또 다른 관점에서 볼 때, 전사체 분석은 암 및 환자-특이적 돌연변이를 갖는 유전자를 동정하고 정량화하기 위해 (단독으로 또는 게놈 분석과 함께) 적합할 수 있다.

[0032] 유사하게, 프로테오믹스 분석은 네오에피토프의 RNA의 실제적인 번역을 확인하기 위해 다양한 방식으로 수행될 수 있으며, 프로테오믹스 분석의 모든 공지된 방식이 본원에서 고려된다. 그러나, 특히 바람직한 프로테오믹스 방법은 항체-기반 방법 및 질량 분광분석 방법을 포함한다. 또한, 프로테오믹스 분석은 단백질 그 자체에 관한 정성적 또는 정량적 정보를 제공할 뿐만 아니라, 단백질이 촉매 작용 또는 다른 기능적 활성을 갖는 단백질 활성 데이터베이스를 또한 포함할 수 있음을 주목해야 한다. 프로테오믹(proteomic) 분석을 수행하기 위한 하나의 예시적인 기술은 본원에 참고로 인용된 US 7473532에 기재되어 있다. 단백질 발현의 더 적절한 동정 및 심지어 정량화의 또 하나의 적합한 방법은 다양한 질량분광 분석(예를 들어, 선택적 반응 모니터링(SRM), 다중 반응 모니터링(MRM) 및 연속 반응 모니터링(CRM))을 포함한다.

[0033] 여과의 또 다른 양태에서, 네오에피토프는 인간-동일 서열의 사용을 피하기 위해 (예를 들어, 환자 또는 환자 집단의) 공지된 인간 서열을 함유하는 데이터베이스와 비교될 수 있다. 또한, 여과는 SNP가 종양 및 매칭된 정상 서열 둘 다에 존재하는 환자의 SNP로 인한 네오에피토프 서열의 제거를 포함할 수도 있다. 예를 들어, dbSNP(단일 뉴클레오타이드 다형성 데이터베이스(Single Nucleotide Polymorphism Database))는 국립 인간 게놈 연구소(NHGRI)와 공동으로 국립 생물공학 정보 센터(NCBI)가 개발하고 호스팅한 여러 종에 걸친 유전적 변이에 대한 무료 공개 아카이브이다. 데이터베이스의 명칭이 한 종류의 다형성(단일 뉴클레오타이드 다형성(SNP))의 집합만을 의미하지만, 실제로는 (1) SNP, (2) 짧은 결실 및 삽입 다형성(인델/DIP), (3) 마이크로위성 마커 또는 짧은 탠덤 반복(STR), (4) 다중 뉴클레오타이드 다형성(MNP), (5) 이중접합 서열, 및 (6) 명명된 변이체와 같은 비교적 광범위한 분자적 변이를 포함한다. dbSNP는 명백하게도 중성 다형성, 공지된 표현형에 상응하는 다형성, 및 변화가 없는 영역을 수용한다.

[0034] 상기한 바와 같은 데이터베이스 및 다른 여과 옵션을 사용하여, 환자 및 종양 특이적 네오에피토프를 여과하여 이들 공지된 서열을 제거함으로써, 실질적으로 감소된 가양성을 갖는 다수의 네오에피토프 서열을 갖는 서열 세

트를 산출할 수 있다.

- [0035] 그럼에도, 여과에도 불구하고, 네오에피토프가 환자의 MHC 복합체에 제시되어야 하므로 모든 네오에피토프가 면역 시스템에서 가시적일 수는 없다는 것을 인식해야 한다. 실제로, 네오에피토프의 일부만이 표현에 충분한 친화도를 나타낼 것이며, MHC 복합체의 다양성으로 인해 모두는 아니더라도 대부분의 공통적인 네오에피토프의 사용은 배제될 것이다. 결과적으로, 면역 요법의 맥락에서, 네오에피토프가 MHC 복합체에 결합되고 제시될 때 네오에피토프가 보다 효과적일 것이라는 것은 너무나도 분명하다. 또 다른 관점에서 볼 때, 체크포인트 억제제를 이용한 치료 성공은 네오에피토프가 환자의 HLA-유형에 최소 친화도를 가져야 하는 MHC 복합체를 통해 여러 네오에피토프가 제시되어야 한다. 따라서, 효과적인 결합 및 표현은 환자의 네오에피토프 서열과 특정 HLA-유형의 조합된 기능이다. 가장 전형적으로, HLA-유형 결정은 적어도 3개의 MHC-I 하위-유형(예컨대, HLA-A, HLA-B, HLA-C) 및 적어도 3개의 MHC-II 하위-유형(예컨대, HLA-DP, HLA-DQ, HLA-DR)을 포함하며, 여기서 바람직하게는 각각의 하위-유형은 적어도 4-자리 깊이(4-digit-depth)로 결정된다. 그러나, 더 큰 깊이(예를 들어, 6자리, 8자리)도 또한 본원에서 고려된다.
- [0036] (공지의 화학 또는 인-실리코 결정을 사용하여) 일단 환자의 HLA-유형이 확인되면, HLA-유형에 대한 구조 용액을 데이터베이스에서 계산하거나 얻은 다음, HLA 구조 용액에 대한 (전형적으로 여과된) 네오에피토프의 결합 친화도를 결정하기 위해 인-실리코(in silico) 도킹 모델에 사용한다. 하기에서 추가로 논의되는 바와 같이, 결합 친화도를 결정하기에 적합한 시스템은 NetMHC 플랫폼을 포함한다(예를 들어 문헌[Nucleic Acids Res. 2008 Jul 1; 36(Web Server issue): W509-W512.] 참조). 그런 다음, MHC-I/II 하위-유형에 대한 지식과 함께, 사전 결정된 HLA-유형에 대해 높은 친화도(예를 들어, 100 nM 미만, 75 nM 미만, 50 nM 미만)를 갖는 네오에피토프가 치료 생성을 위해 선택된다.
- [0037] HLA 결정은 당해 분야에 잘 공지된 습식-화학에서의 다양한 방법을 사용하여 수행될 수 있고, 이들 방법 모두는 본원에서의 사용에 적합한 것으로 간주된다. 그러나, 특히 바람직한 방법에서, HLA-유형은 또한 하기에 보다 상세히 도시된 바와 같이 공지된 및/또는 공통 HLA-유형의 대부분 또는 전부를 함유하는 참조 서열을 사용하여 인-실리코의 옴믹스 데이터로부터 예측될 수 있다.
- [0038] 예를 들어, 본 발명의 대상에 따른 하나의 바람직한 방법에서, 염색체 6p21.3(또는 HLA 대립 형질이 발견되는 곳/근처에 있는 임의의 다른 위치)에 매핑하는 비교적 많은 수의 환자 서열 판독이 데이터베이스 또는 시퀀싱 기계에 의해 제공된다. 가장 전형적으로, 서열 판독은 약 100 내지 300개의 염기의 길이를 가지며 판독 품질, 정렬 정보, 배향, 위치 등을 포함한 메타데이터를 포함한다. 예를 들어 적절한 포맷에는 SAM, BAM, FASTA, GAR 등이 포함된다. 본 발명의 대상을 제한하는 것은 아니지만, 일반적으로 환자 서열 판독은 적어도 5x, 보다 전형적으로는 적어도 10x, 더욱 전형적으로는 적어도 20x, 가장 일반적으로는 적어도 30x의 커버리지의 깊이를 제공하는 것이 바람직하다.
- [0039] 환자 서열 판독에 추가하여, 숙고된 방법은 공지된 별개의 HLA 대립 형질의 다수의 서열을 포함하는 하나 이상의 참조 서열을 추가로 이용한다. 예를 들어, 전형적인 참조 서열은 상기 HLA-유형의 다중 HLA-대립 형질을 갖는 적어도 하나의 HLA-유형의 서열 단편을 포함하는 합성(상응하는 인간 또는 다른 포유동물 대응물을 갖지 않음) 서열일 수 있다. 예를 들어, 적합한 참조 서열은 HLA-A의 적어도 50개의 상이한 대립 형질에 대한 공지된 게놈 서열의 집합을 포함한다. 대안적으로 또는 부가적으로, 상기 참조 서열은 또한 HLA-A의 적어도 50개의 상이한 대립 형질에 대한 공지된 RNA 서열의 집합을 포함할 수 있다. 물론, 이하에서 보다 상세히 논의되는 바와 같이, 참조 서열은 HLA-A의 50개의 대립 형질에 한정되지 않고 HLA-유형 및 대립 형질의 수/조성에 대한 대안적인 조성물을 가질 수 있다. 가장 전형적으로, 참조 서열은 컴퓨터 판독가능한 포맷으로 존재하고 데이터베이스 및 다른 데이터 저장 장치로부터 제공될 것이다. 예를 들어, 적합한 참조 서열 포맷은 FASTA, FASTQ, EMBL, GCG 또는 GenBank 포맷을 포함하고, 공용 데이터 저장소(예컨대 국제 면역유전학 정보 시스템(IMG) 또는 대립 형질 빈도 네트워크 데이터베이스(The Allele Frequency Net Database), EUROSTAM, URL: [www.allelefrequencies.net](http://www.allelefrequencies.net))의 데이터로부터 직접 얻거나 구축할 수 있다. 대안적으로, 참조 서열은 또한 대립 형질 빈도, 종족 대립 형질 분포, 공통 또는 희귀 대립 형질 유형 등과 같은 하나 이상의 사전 결정된 기준에 기초한 개별적인 공지의 HLA-대립 형질로부터 구축될 수 있다.
- [0040] 참조 서열을 사용하여, 환자 서열 판독은 이제 드브루인(de Bruijn) 그래프를 통해 쓰레딩(threading)되어 가장 적합한 대립 형질을 식별할 수 있다. 이러한 맥락에서, 각 개인은 각각의 HLA-유형에 대해 2개의 대립 형질을 갖고 있으며 이러한 대립 형질은 매우 유사하거나 어떤 경우에는 동일할 수도 있다는 점에 주목해야 한다. 이러한 고도의 유사성은 전통적인 정렬 방식에 상당한 문제를 제기한다. 본 발명자들은 이제 HLA 대립 형질 및

심지어 매우 밀접하게 관련된 대립 형질이 드브루인 그래프가 상대적으로 작은 k-mer(전형적으로 10 내지 20개의 염기 길이를 가짐)로 판독된 서열을 분해함으로써 그리고 대립 형질의 서열과 매칭되는 서열 판독의 k-mer에 기초하여 각각의 환자 서열 판독이 각각의 대립 형질에 대한 투표("정량적 판독 지원")를 제공하는 가중 투표 과정을 구현함으로써 구성되는 접근법을 사용하여 해결될 수 있음을 발견하였다. 그런 다음, 대립 형질에 대해 누적적으로 가장 높은 투표가 가장 가능성이 높게 예측된 HLA 대립 형질을 나타낸다. 또한 대립 형질에 매칭되는 각각의 단편을 사용하여 대립 형질의 전체적인 커버리지와 커버리지의 깊이를 계산하는 것이 일반적으로 바람직하다.

[0041] 특히 탑 히트(top hit) 중 많은 부분이 유사한 경우(예를 들어, 스코어의 상당 부분이 고도로 공유된 k-mer 세트의 경우) 스코어를 더 높이거나 세분화할 수 있다. 예를 들어, 스코어 정제는 현재의 탑 히트와 실질적으로 유사한(예를 들어, 99% 초과 또는 다른 소정의 값) 대립 형질이 장래의 고려사항으로부터 제거되는 가중치 방식을 포함할 수 있다. 현재의 탑 히트에 의해 사용되는 k-mer에 대한 카운트는 인자(예컨대, 0.5)에 의해 재-가중되고, 각 HLA 대립 형질에 대한 스코어는 이들 가중된 카운트를 합산함으로써 재계산된다. 이러한 선택 과정은 새로운 탑 히트를 구하기 위해 반복된다. 이러한 방법의 정확성은 심지어 DNA에 존재하는 2개의 대립 형질 중 때로는 단지 1개일 수 있는 종양에 의해 발현되는 대립 형질을 확인할 수 있는 RNA 서열 데이터를 사용하여 더욱 개선될 수 있다. 고려된 시스템 및 방법의 또 다른 유리한 양태에서, DNA 또는 RNA, 또는 DNA와 RNA 둘 다의 조합을 처리하여 고도로 정확하고 종양 또는 혈액 DNA 또는 RNA로부터 유도될 수 있는 HLA 예측을 수행할 수 있다. 인-실리코 HLA 타이핑에서의 고정밀도에 대한 추가의 양태, 적합한 방법 및 고려사항은 본원에 참고로 인용된 국제 출원 제PCT/US16/48768호에 기재되어 있다.

[0042] 일단 환자 및 종양 특이적 네오에피토프 및 HLA-유형이 확인되면, HLA에 네오에피토프를 도킹하고 예를 들어 NetMHC를 사용하여 최상의 결합제(예를 들어, 최저 KD 예컨대 500 nM 미만, 또는 250 nM 미만, 또는 150 nM 미만, 또는 50 nM 미만)를 결정하여 추가적인 컴퓨터 분석을 수행할 수 있다. 이러한 접근법은 환자 및 종양에 진정한 특정 네오에피토프를 동정할 뿐만 아니라 세포 상에 나타날 가능성이 가장 높고 치료 효과에 의해 면역 반응을 이끌어낼 가능성이 가장 큰 네오에피토프를 동정할 수 있음을 이해해야 한다. 물론, 이와 같이 동정된 HLA-매칭된 네오에피토프는 추가로 후술되는 바와 같이 페이로드(payload)로서 에피토프를 코딩하는 핵산을 바이러스 내에 포함시키기 전에 시험관내에서 생화학적으로 유효화될 수 있음을 이해해야 한다.

[0043] 물론, 환자의 HLA-유형과 환자- 및 암-특이적 네오에피토프의 매칭은 NetMHC 이외의 시스템을 사용하여 수행될 수 있으며, 적합한 시스템에는 NetMHC II, NetMHCpan, IEDB 분석 자원(URL [immuneepitope.org](http://immuneepitope.org)), RankPep, PREDEP, SVMHC, Epipredict, HLABinding 및 기타(예컨대, 문헌[J Immunol Methods 2011; 374: 1-4] 참조)가 포함된다. 가장 높은 친화도를 계산할 때, 변형된 아미노산의 위치가 이동된 네오에피토프 서열의 수집(상술됨)이 사용될 수 있다는 점에 주목해야 한다. 대안적으로 또는 부가적으로, 네오에피토프에 대한 변형은 N- 및/또는 C-말단 변형을 가함으로써 발현된 네오에피토프와 환자 HLA-유형의 결합을 더욱 증가시킴으로써 구현될 수 있다. 따라서, 네오에피토프는 특정 HLA-유형에 보다 잘 매칭되도록 동정되거나 추가로 변형된 상태의 원형일 수 있다. 또한, 필요한 경우, 상응하는 야생형 서열(즉, 아미노산 변화가 없는 네오에피토프 서열)의 결합을 계산하여 높은 차등 친화도를 보장할 수 있다. 예를 들어, 네오에피토프와 이의 상응하는 야생형 서열 간의 MHC 결합에서 특히 바람직한 높은 차등 친화도는 적어도 2배, 적어도 5배, 적어도 10배, 적어도 100배, 적어도 500배, 적어도 1000배 등이다.

[0044] 추가 관찰사항(데이터 미도시)에 기초하여, 본 발명자들은 환자의 옴릭스 데이터를 (바람직하게는 엑솜(exome) 또는 전체 시퀀싱의 동기식 위치-유도 정렬을 사용하여) 분석하여 특정 돌연변이의 유형 또는 패턴을 확인하고, (특히 상술한 바와 같은 최소 수의 네오에피토프와 조합된) 이러한 패턴이 체크포인트 억제제로 치료 성공 가능성이 더 있음을 나타낼 수 있다는 것을 고려하였다. 예를 들어, 네오에피토프가 UV 손상(예컨대 탠덤 CC>TT/GG>AA 돌연변이)에 전형적인 돌연변이 패턴과 관련이 있는 경우 및 50, 70, 100개 이상의 HLA-매칭된 환자- 및 암- 특이적 네오에피토프가 존재하는 경우, 체크포인트 억제제에 의한 성공적인 치료는 돌연변이 패턴이 없는 경우 및/또는 HLA-매칭된 환자- 및 암-특이적 네오에피토프가 덜 존재하는 경우보다 가능성이 더 높을 수 있다. 유사하게, 네오에피토프가 흡연-유발 DNA 손상(예컨대 G>T 돌연변이의 고빈도)에 전형적인 돌연변이 패턴과 관련된 경우 및 50 또는 70 또는 100개 이상의 네오에피토프가 존재하는 경우, 체크포인트 억제제에 의한 성공적인 치료는 돌연변이 패턴이 없는 경우 및/또는 HLA-매칭된 환자- 및 암-특이적 네오에피토프가 덜 존재하는 경우보다 가능성이 더 높을 수 있다.

[0045] 또한, 증가된 수의 네오에피토프는 다양한 근본적인 조건 또는 현상에 기인할 수 있는 것으로 생각된다. 예를 들어, 세포 내에서 MMR(DNA 부정합 교정) 시스템의 결함 또는 비정상적인 기능에 의해 더 많은 수의 네오에피토프



프가 유발될 수 있으며, 이로 인해 다중 가닥 파열 및 더 높은 돌연변이율 및 잠재적으로 더 높은 카운트의 HLA-매칭된 환자- 및 암-특이적 네오에피토프를 유도할 수 있으며, 이는 또한 MSI(마이크로위성 불안정성)로도 관찰될 수 있다. 따라서, 전체 게놈 및/또는 엑솜 시퀀싱으로부터의 MMR 및/또는 MSI의 관측이 증가된 수의 HLA-매칭된 환자- 및 암-특이적 네오에피토프에 대한 대리 지표로서 사용될 수 있다는 것도 또한 고려된다. MMR 및/또는 MSI는 공지된 옴믹스 분석 알고리즘 및 데이터 시각화(예를 들어 원형 플롯 다이어그램을 사용)를 사용하여 매칭된 정상 조직에 대해 식별되는 것이 바람직하다.

[0046] 네오에피토프가 위에 기재된 방법 중 하나 이상을 사용하여 적절하게 여과되면, 여과된 네오에피토프를 정량화/계산할 수 있다. 쉽게 이해할 수 있는 바와 같이, 면역 체크포인트 억제제로 암을 치료하는 효과적인 치료법은 발현되고 제시된 네오에피토프의 존재에 의존적이다. 임의의 특정 이론 또는 가설에 구속되기를 바라지는 않지만, 본 발명자들은 일반적으로, 발현되고 제시된 네오에피토프의 일부만이 치료 반응을 유도하고, 종양이 암세포의 이중 집단을 가질 수 있으며, 여기서 각각의 집단이 각각의 개별적인 네오에피토프와 관련됐을 가능성이 있음이 고려된다. 따라서, 본 발명자들은 면역 체크포인트 억제제를 사용하여 암을 치료하기 위해서는 HLA-매칭된 환자- 및 암-특이적 네오에피토프의 최소/임계치 양이 필요할 것으로 예상한다. 따라서, 다양한 암 데이터의 후향적 분석에 기초하고 하기에서 더 상세하게 논의되는 바와 같이, 본 발명자들은 HLA-매칭된 환자- 및 암-특이적 네오에피토프의 양이 사전 결정된 임계치 양을 초과할 때 체크포인트 억제제를 환자에게 투여해야 한다고 생각한다. 가장 전형적으로, 소정의 임계치 양은 적어도 50개의 HLA-매칭된 환자- 및 암-특이적 네오에피토프, 또는 적어도 100개의 HLA-매칭된 환자- 및 암-특이적 네오에피토프, 또는 적어도 150개의 HLA-매칭된 환자- 및 암-특이적 네오에피토프이다.

[0047] 다른 관점에서 보면, 다수의 돌연변이가 (a) 예를 들어 침묵 돌연변이로 인해 네오에피토프를 유발하지 않을 수 있고, (b) 상응하는 매칭된 정상 서열을 가질 수 있어 네오에피토프를 전혀 나타내지 않을 수 있으며, (c) 발현되지 않아 면역 시스템에 보이지 않을 수도 있고, (d) 환자 특이적-MHC-복합체에 결합될 수 없어 면역 시스템에 보이지 않을 수도 있기 때문에, 암에서 비교적 많은 수의 돌연변이를 진단하는 것은 그 자체로 체크포인트 억제제에 의한 암 치료시의 치료 반응을 예측하지 못한다는 것을 인식해야 한다. 대조적으로, 숙고된 시스템 및 방법은 HLA-매칭된 환자- 및 암-특이적 네오에피토프를 높은 신뢰도로 확인한다. 반대로, 비교적 낮은 돌연변이 빈도를 갖는 암은 실제로 환자 돌연변이가 비교적 많은 수의 HLA-매칭된 환자- 및 암-특이적 네오에피토프(예컨대 적어도 50, 또는 100, 150 또는 200개 등)에서 번역되는 환자에서 치료될 수 있음을 인식해야 한다.

[0048] 적합한 체크포인트 억제제와 관련하여, 체크포인트 시그널링(예컨대 CTLA-4(CD152) 또는 PD-1(CD279))을 간섭하는 모든 화합물 및 조성물은 본원에서 사용에 적합한 것으로 간주된다. 예를 들어, 특히 바람직한 체크포인트 억제제는 펩트폴리주맵, 니볼루맵 및 이필리무맵을 포함한다. 가장 전형적으로, 체크포인트 억제제는 통상의 프로토콜에 따라 처방 정보에 기재된 바와 같이 투여될 것이다. 그러나, 체크포인트 억제제가 펩타이드 또는 단백질인 경우, 이러한 펩타이드 및/또는 단백질은 또한 임의의 적합한 발현 시스템(네오에피토프 및/또는 공동-자극 분자를 따라 또는 이들과 함께)으로부터 환자에서 발현될 수 있다는 점에 주목해야 한다. 또한, 본원에서 사용된, 체크포인트 억제제와 관련하여 '투여'라는 용어는 환자에 대한 체크포인트 억제제의 직접 투여(예를 들어, 의사 또는 다른 면허 있는 의료 전문가에 의한 것) 또는 간접 투여(예를 들어, 투여를 유발하거나 투여에 대한 조언을 하는 것)를 지칭한다.

[0049] 또 다른 고려되는 양태에서, 체크포인트 억제제를 하나 이상의 항암 치료제와 함께 사용하는 병용 요법이 적합할 수 있다. 다른 약제들 중에서도, 종양이 동정된 네오에피토프 중의 적어도 하나의 발현을 유도하여 종양에 대한 면역 반응을 강화시키는 핵산 구조에 의해 유전적으로 변형된 바이러스로 종양을 치료할 수 있는 것이 특히 바람직하다. 예를 들어, 적합한 바이러스는 아데노바이러스, 아데노-관련 바이러스, 알파바이러스, 헤르페스 바이러스, 렌티바이러스 등을 포함한다. 그러나, 아데노바이러스가 특히 바람직하다. 또한, 바이러스는, 선택된 바이러스 단백질(예를 들어, E1, E3 단백질)의 표적화된 결실에 의해 전형적으로 달성되는 복제 결핍 및 비-면역원성 바이러스인 것이 더욱 바람직하다. 이러한 바람직한 특성은 E2b 유전자 기능을 결실시킴으로써 더욱 강화될 수 있으며, 최근에 보고된 바와 같이 유전자 변형된 인간 293 세포를 사용하여 높은 역가의 재조합 바이러스를 달성할 수 있다(예를 들어 문헌[J Virol. 1998 Feb; 72(2): 926-933]). 가장 전형적으로, (바이러스 감염된 세포로부터의 발현을 위한) 원하는 핵산 서열은 당업계에 널리 공지된 적절한 조절 요소의 제어하에 있다. 대안적으로, 면역 요법은, 바이러스에 의존할 필요는 없지만, 핵산 백신 접종에 의해 또는 네오에피토프(예를 들어, 단일 펩타이드, 탠덤 미니-유전자 등)의 발현을 유도하는 다른 재조합 벡터에 의해 수행될 수 있다.

[0050] 마찬가지로, 바이러스 발현 벡터 이외의 추가의 면역 치료제가 또한 적합하다고 여겨지며, 키메라 항원 수용체 또는 높은 친화도 CD16 수용체를 발현하는 유전적으로 조작된 세포(및 특히 다양한 면역 적격 세포)를

포함한다. 예를 들어, 고려되는 면역 치료제는 NK 세포(예를 들어 캘리포니아주 90232 컬버 시티 제퍼슨 블러바드 9920 소재의 난트퀘스트(NantKwest)로부터 상업적으로 입수가능한 aNK 세포, haNK 세포 또는 taNK 세포) 또는 유전적으로 변형된 T-세포(예를 들어 T-세포 수용체를 발현하는 세포) 또는 HLA-매칭된 환자- 및 암-특이적 네오에피토프에 의해 생체의 자극된 T-세포를 포함한다. 대안적으로, HLA-매칭된 환자- 및 암-특이적 네오에피토프는 또한 임의적으로 캐리어 단백질에 결합된 펩타이드로서 투여될 수 있다.

[0051] **실시예**

[0052] 데이터 세트: 아래에 표시된 다양한 암에 대한 TCGA WGS 및 RNAseq 데이터는 유니버시티 오브 캘리포니아(University of California) 산타 크루즈(UCSC) 암 유전체학 허브(Cancer Genomics Hub)(<https://cghub.ucsc.edu/>)에서 다운로드되었다. TCGA 샘플은 인-실리코 HLA 타이핑을 돕기 위해 완전한 WGS 데이터 이용성에 기반하여 선택되었다. 가능한 경우 해당 샘플의 RNAseq 데이터를 사용하였다.

[0053] 종양 변이체 및 네오에피토프의 동정: 단일 뉴클레오타이드 변이체(SNV) 및 삽입/결실(인델(indel))은 실질적으로 US 2012/0059670A1 및 US 2012/0066001A1에 개시된 바와 같은 방식으로 BAM 파일을 사용하여 종양 및 정상 샘플의 위치-유도된 동기 정렬에 의해 동정되었다. HLA-A 대립 형질이 9-mer 펩타이드 단편에 우세하게 결합하기 때문에, 본 발명자들은 9-mer 네오에피토프의 동정에 집중했다. 네오에피토프는 동정된 SNV 또는 인델로부터 유래된 9-mer 아미노산 스트링의 모든 가능한 순열을 생성함으로써 동정되었다(즉, 각각의 9-mer는 고유 위치에서 변경된 아미노산을 가졌다). 특정 네오에피토프의 가능한 오프-표적(off-target) 효과를 감소시키는 수단으로서, 본 발명자들은 모든 공지의 인간 유전자로부터 생성된 모든 가능한 9-mer 펩타이드 서열에 대하여 모든 동정된 네오에피토프를 여과하였다. 또한, 본 발명자들은 시퀀싱 데이터 내에서 누락된 희귀한 단백질 서열을 설명하기 위해 dbSNP(URL: [www.ncbi.nlm.nih.gov/SNP/](http://www.ncbi.nlm.nih.gov/SNP/))로부터 단일 뉴클레오타이드 다형성을 여과하였다. 네오에피토프는 종양의 이질성으로 인해 발생하는 문제를 상쇄하기 위해 관찰된 코딩 변이체의 대립 형질 빈도 뿐만 아니라 RNA 발현에 의해 추가로 선정되었다.

[0054] HLA 타이핑: TCGA 샘플에서는 HLA 타이핑 데이터를 이용할 수 없었다; 따라서, 본 발명자들은 실질적으로 PCT/US16/48768에 기재된 바와 같이 WGS, RNAseq 데이터 및 HLA 포레스트 알고리즘을 사용하여 인-실리코 HLA 타이핑을 수행하였다. 간단히 말해, 버로우스-휠러(Burrows-Wheeler) 정렬 알고리즘을 사용하여 시퀀싱 판독을 IMGT/HLA 데이터베이스(URL: [www.ebi.ac.uk/ipd/imgt/hla/](http://www.ebi.ac.uk/ipd/imgt/hla/)) 내의 모든 다른 HLA 대립 형질에 맞춰 정렬하였다. 각각의 정렬에는 판독 품질 스코어를 고려하여 염기 보존에 기초하여 점수가 부여된다. 그 후, 각각의 HLA 대립 형질은 각각의 판독이 특정 HLA 대립 형질에 얼마나 잘 정렬되는지를 나타내는 스코어의 합계를 가지며 가장 높은 스코어를 갖는 대립 형질이 1차 대립 형질 타이핑으로서 선택된다. 그 후, 2차 대립 형질 타이핑은 1차 대립 형질 타이핑에 완벽하게 정렬되는 판독 값을 제거함으로써 수행되며, 이후의 판독은 1차 대립 형질에 대한 정렬없이 다시 스코어링된다. 이러한 방법을 사용하여, 본 발명자들은 모든 샘플에 대해 HLA-A, HLA-B, HLA-C 및 HLA-DRB1에 대한 타이핑 결과를 4자리 이상의 수준으로 얻었다.

[0055] 네오에피토프-HLA 친화도 측정: 네오에피토프가 특정 HLA 대립 형질에 결합하는지를 예측하기 위해 NetMHC 3.4(URL: [www.cbs.dtu.dk/services/NetMHC-3.4/](http://www.cbs.dtu.dk/services/NetMHC-3.4/))를 사용했다. 복잡성 공간을 줄이기 위해, 본 발명자들은 가장 잘 특성화된 HLA 대립 형질이고 가장 우수한 결합 친화도 모델을 가지기 때문에, HLA-A 대립 형질에 대한 결합 분석을 제한하기로 선택했다. NetMHC 3.4 도구에는 동정된 모든 HLA-A 대립 형질에 대한 모델이 없기 때문에, 환자의 HLA-A 타이핑을 NetMHC 3.4에서 사용할 수 없는 경우, 결합 예측을 위해 HLA 상위-유형을 선택하였다. 500 nM 단백질 농도 미만의 예측된 결합 친화도를 갖는 네오에피토프는 추가 분석을 위해 보관되었다. 그러나, 다른 보다 엄격한 결합 기준(<250 nM 또는 <150 nM 또는 <50 nM)도 적절한 것으로 간주된다.

[0056] 암 유형에 따른 코딩 돌연변이 및 네오에피토프 부하: 가능한 경우, WGS 데이터 및 상응하는 RNAseq 데이터를 사용하여 도 1에 도시된 바와 같이 23개의 암 분류에 걸쳐 750개의 환자 샘플에 대한 코딩 DNA의 메가베이스 당 잠재적인 네오에피토프 및 체세포성 코딩 변이체를 확립했다. 여기서는, 네오에피토프 및 변이체 카운트가 TCGA 내의 23개의 암 분류에 걸쳐 750명의 환자 샘플에 대해 도시된다. 패널 (a)는 네오에피토프 카운트를 나타내고; 패널 (b)는 변이체 카운트를 나타낸다. y-축은 코딩 DNA의 메가베이스 당 카운트를 나타낸다(인간 게놈 어셈블리(hg)19에 대해서는 88MB). x-축은 각각의 암 분류를 나타낸 것으로, 괄호 안은 환자 샘플 수를 나타낸다. 종양 샘플 카운트는 정사각형으로 표시된다. 패널 (c)는 모든 암 유형 내에서의 네오에피토프 및 정상 에피토프의 비율을 나타낸다.

[0057] 도 1에서 쉽게 알 수 있는 바와 같이, 돌연변이 및 네오에피토프 부하는 상이한 암 유형에 따라 다양하고, 여기서 흑색종 및 편평 세포 폐암이 가장 높은 네오에피토프 부하를 갖고 갑상선암과 급성 골수성 백혈병이 가장

낮은 네오에피토프 부하를 갖는다. 잠재적 오프-표적 효과를 제거하기 위해 알려진 인간 서열의 데이터베이스에 대한 추정 네오에피토프 여과는 단지 10%의 동정된 네오에피토프만이 공지된 단백질 단편에 매핑되고; 따라서 대부분의 돌연변이는 고유 단백질 서열을 생성한다. 그러나, 고유 네오에피토프의 분율이 비교적 높더라도, 발현 및 표현이 일어날 것으로 추정될 수 없다. 실제로, 하기에서 더 상세하게 나타나는 바와 같이, 발현되고 표시된 네오에피토프의 수는 시퀀싱에 의해서만 확인된 네오에피토프의 수보다 현저히 적다는 것을 인식해야 한다.

[0058]

네오에피토프 돌연변이 부하 및 발현: MHC 클래스 I(MHC-I)에 의해 제시되는 에피토프의 길이로 인해, 단일 돌연변이는 그림에도 불구하고 수많은 별개의 네오에피토프의 발현을 초래할 수 있다. 따라서, 수백 개의 돌연변이가 포함된 개별 환자의 종양에는 수천 개의 네오에피토프가 포함될 가능성이 높다. 많은 종양 돌연변이가 패신저(passenger) 돌연변이가 될 수 있고 암 진행에 책임이 없지만, 잠재적으로는 치료적 개입용 표적으로서 이용될 수 있다. RNAseq 데이터는 결합된 모든 암 분류 내에서 발현되고 또한 다른 돌연변이가 부하를 가지는 암 분류: 자궁 코퍼스 자궁내막 암, 갑상선암 및 유방 침윤성 암에서도 발현되는 네오에피토프를 선택하는 데 사용되었다. 가장 주목할만한 것은, 도 2에 도시된 바와 같이, WGS에 의해 동정된 네오에피토프 카운트가 다양한 암 (결합된 모든 암에 대해 피어슨(Pearson)의  $r = 0.99$ )에 걸친 RNAseq에 의해 동정된 네오에피토프 발현과 상관관계가 있었다. 여기서, 패널 (a)는 모든 암을 나타내고, 패널 (b)는 갑상선암(THCA)을 나타내고, 패널 (c)는 유방 침윤성 암(BRCA)을 나타내고, 패널 (d)는 자궁 코퍼스 자궁내막 암(UCEC)을 나타낸다. y-축은 WGS에 의해 동정된 샘플 당 네오에피토프의 원시 카운트를 보여주고, x-축은 RNAseq에 의해 결정된 발현된 유전자에 대해 여과한 후 네오에피토프의 원시 카운트를 보여준다. 피어슨 상관관계, P-값 및 샘플 수는 각 그래프에 표시된다. 높은 네오에피토프 부하를 가진 암은 평균 돌연변이 부하에 상관없이 일반적으로 높은 네오에피토프 발현을 보였다.

[0059]

삼중 네거티브 유방암에서 네오에피토프 동정: 삼중-네거티브 유방암(TNBC)은 제한적인 치료 옵션을 가진 공격적인 암이며, 표준 화학요법 후에 진행하는 경우 종종 나쁜 예후를 보인다. TCGA 데이터 세트에는 26개의 TNBC 샘플에 대한 WGS 데이터 및 RNAseq 데이터가 포함되어 있다. TNBC에서 네오에피토프 카운트는 다음과 같은 반복 접근법을 사용하여 확인되었다: 모든 가능한 네오에피토프는 WGS에 의해 확인된 코딩 변이체에 근거하여 예측되었고; 네오에피토프의 수는 RNAseq로 확인된 발현된 네오에피토프를 선택하여 좁혀졌으며; 그 목록은 환자의 HLA 유형 내에서 특정 대립 형질에 결합할 것으로 예측되는 네오에피토프를 선택함으로써 더욱 정제되었다. 네오에피토프의 선택적인 가지치기(pruning)는 하기 표 1에 나타난 바와 같이 각 환자에게 고유한 고품질 네오에피토프의 목록을 산출했다.

표 1

TCGA 마크드	HLA-A 타입	HUGO 유전자 명칭	TPM	네오에피토프	단백질 변화	정상	결합된 HLA-A 대립 형질	HLA-A에 대한 친화 도
TCGA-E2-A14X- 01A-11D-A10Y- 09	A*23:01, A*11:01	NAA50	229.85	PTDAHVLQK	p.A145T	PADAHVLQK	A*11:01	146 nM
TCGA-E2-A1LL- 01A-11D-A142- 09	A*02:01, A*02:01	FBX02	187.36	LLLHVLAAL	p.R57H	LLLRVLAAL	A*02:01	18 nM
TCGA-AN-A0G0- 01A-11D-A045- 09	A*11:01, A*11:01	C1orf43	574.04	TQSCYNYLY	p.N94T	NQSCYNYLY	A*11:01	225 nM
TCGA-A2-A0D2- 01A-21D-A128- 09	A*03:01, A*32:01	TBCD	57.13	TVVRWSVAK	p.A380V	TVVRWSAAK	A*03:01	119 nM
TCGA-A7-A26G- 01A-21D-A167- 09	A*01:01, A*01:01	PTEN	41.3	RTGVMKCAY	p.I135K	RTGVMICAY	A*01:01	480 nM

[0060]



[0061] 26명의 모든 환자에 대해 도 3(상단 패널, a)에 도시된 바와 같이, 예측된 네오에피토프, 발현된 네오에피토프 및 각각의 환자-특이적 HLA-A 유형에 대해 친화도를 가진 네오에피토프의 수는 각각 17,925, 8184 및 228이었다. 분명히, 시퀀싱(심지어 매칭된 정상에 대한 종양의 시퀀싱 및 분석)에 의해 측정된 높은 수의 네오에피토프에 대한 의존은 체크포인트 억제제에 의한 치료에 반응하는 의미있는 예측을 제공하지 않을 것이다. 마찬가지로, 실제 HLA-결합체에 대한 결과의 정규화가 발현되고 표현된 환자- 및 종양-특이적 네오에피토프의 비율을 실질적으로 증가시키는 반면, 발현에 대한 추가 고려는 가양성 결과를 단지 적당히 제거할 뿐이다.

[0062] 암 분류를 통한 네오에피토프의 동정: TNBC를 가진 환자들 사이에 공유된 네오에피토프가 없었기 때문에, 본 발명자들은 어떤 새로운 네오에피토프가 TCGA 데이터 세트 내의 다른 암 분류들 사이에서 공유되는지를 결정하려고 했다. 일반적인 네오에피토프가 또한 HLA 복합체에 결합할 수 있도록, 본 발명자들은 북아메리카 전역의 고주파에서 발생하는 HLA-A\*02:01 대립 형질을 함유하는 샘플로 분석을 제한하였다. TNBC에 대해 수행된 동일한 반복 접근법을 사용하여, 본 발명자들은 완전한 WGS 및 RNAseq 데이터를 갖는 12개의 암에 대한 네오에피토프를 동정하였고, 결과를 도 3(하단 패널, b)에 나타내었다. 각 막대 내의 음영은 다른 환자 샘플을 나타낸다.

[0063] 여기서 예측되는 네오에피토프, 발현된 네오에피토프 및 HLA-A\*02:01에 대한 친화도를 갖는 네오에피토프의 수는 각각 211,285, 89,351 및 1,732이었다. 서로 다른 샘플 크기 교정시, 예측되는 네오에피토프, 발현된 네오에피토프 및 HLA-A\*02:01에 대한 친화도를 갖는 네오에피토프의 평균 수는 각각 23,272, 9,619 및 138이었다. 이러한 데이터를 통해, 방광암, 요로 암종, 폐 편평 세포 암종, 폐 선암 및 유방 침윤성 암 등 네 가지 상이한 암 유형을 나타내는 4개의 상이한 환자 샘플에서 발생하는 하나의 네오에피토프가 동정되었다. 다수의 네오에피토프가 두 가지 다른 암 유형을 나타내는 환자 쌍 사이에 공유되었다(표 2).

표 2

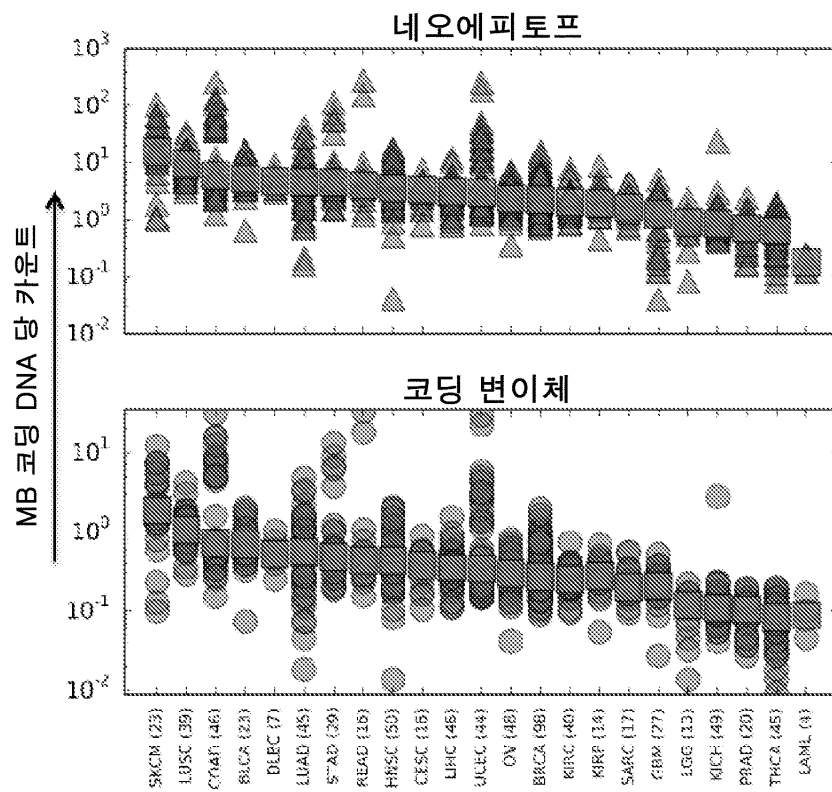
TCGA 바코드	HUGO 유전자명칭	네오에피토프	단백질 변화	정상	암 유형
TCGA-E2-A109-01A-11D-A10M-09, TCGA-CR-5249-01A-01D-2276-10, TCGA-BA-6872-01A-11D-A32X-10, TCGA-CN-6989-01A-11D-A32X-10	SOS2	YIHTHTFYV	p.T390I	YHTHTFYV	HNSC, n=3 BRCA, n=1
TCGA-EW-A1J5-01A-11D-A13L-09, 21-1082-01A-01D-08, TCGA-GD-A2C5-01A-12D-A17V-08, TCGA-75-5147-01A-01D-1625-08	USP8	SQIWNLPV	p.R763W	SQIRNLPV	BLCA, n=1 LUSC, n=1 BRCA, n=1
TCGA-B6-A0RT-01A-21D-A128-09, TCGA-AQ-A04J-01A-02D-A128-09	AKAP11	SCMNPQTFK	p.K400T	SCMNPQKFK	BRCA, n=2
TCGA-55-7281-01A-11D-2036-08, TCGA-AO-A03L-01A-41D-A19H-09	PCMTD1	KLSPESLK	p.P342S	KLPLPESLK	LUAD, n=1 BRCA, n=1
TCGA-EL-A3T0-01A-22D-A22D-08, TCGA-A8-A08L-01A-11D-A19H-09	PKD1	AMPSPEARV	p.T938M	ATPSPEARV	THCA, n=1 BRCA, n=1

[0064]

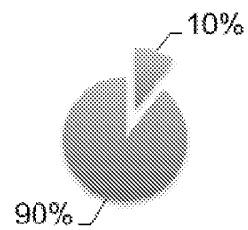
- [0065] 주목할 만하게도, WGS에만 근거한 초기 네오에피토프 예측은 TNBC 환자에서 재발한 몇 가지 네오에피토프를 동정했다(데이터 미도시). 네오에피토프-HLA 결합 분석 후, 재발성 네오에피토프는 모두 환자들 간의 HLA 대립형질과 결합력에서의 차이로 인해 제거되었다. 12개의 암 분류를 조합한 경우에도, 재생되는 네오에피토프가 드물었으며, 특이한 암 유형의 4명의 환자 중 네오에피토프가 하나만 발견되어 포괄적인 분자 프로파일링의 필요성이 한번 더 강조되었다. 도 3(하단 원형 차트, c)은 모든 암에서, 약 6%의 네오에피토프가 암 드라이버(driver) 유전자에서 발생하였는데, 이는 이전의 관찰과 일치한다.
- [0066] 흑색종 및 폐암과 같은 특정 종양은, 항암 반응을 이끌어내고 암이 체크포인트 억제제에 더 민감하게 반응하게 하는 체세포성 네오에피토프의 증가된 발현과 함께 높은 돌연변이 부하를 갖는다. 한편, 비교적 낮은 돌연변이/네오에피토프 부하를 갖는 암은 발현/결합된 네오에피토프가 적어야 하고, 이에 따라 체크포인트 억제제를 사용한 치료에 덜 반응성이어야 한다. 불행하게도, 이러한 가정은 지나치게 단순화되어 있으며, 체크포인트 억제제 치료에 대한 실제 반응은 종양 특이적으로 발현된 네오에피토프와 환자의 HLA-유형에 대한 네오에피토프의 환자 특이적 친화도 사이의 환자 및 종양 특이적 매칭에 크게 좌우될 것이다. 예를 들어, TCGA 내 흑색종 및 폐암 샘플의 평균 돌연변이성 부하가 높았지만, 일부 개별 샘플은 낮은 돌연변이성 부하를 보였다. 그러므로, 질병 유형에 의한 통상적인 분류만이 과도하게 포함될 것이며, 따라서 그 효과가 없을 것 같은 치료 대상 환자들을 임을 이해해야 한다. 본 발명자들은 또한 돌연변이 부하가 높은 다양한 종류의 암 유형에 걸쳐 다수의 개별적인 종양 샘플을 확인함으로써 환자를 체크포인트 억제제로 치료하는 데 민감하게 만들었다. 종합하면, 본 발명자들의 발견은 환자의 종양에 대한 상세한 분자 분석이 이러한 약제에 대해 승인된 적응증 이외의 체크포인트 억제제의 잠재적 이익을 결정하는 데 필요함을 시사한다. 지침 원리로서, 위에서 제공된 데이터 및 고려사항들을 감안하여, 분석은 임계치 수(예를 들어, 적어도 50, 보다 전형적으로는 적어도 100)를 초과하여 존재해야 하는 HLA-매칭된 환자- 및 암-특이적 네오에피토프(즉, 환자 HLA-유형에 높은 친화도, 전형적으로 250 nM 미만 또는 150 nM 미만의 친화도를 갖는 네오에피토프)에 초점을 맞추어야 한다.
- [0067] 도 4는 위에 기술한 바와 같이 동정된 다양한 암 네오에피토프에 대한 변이체 카운트를 예시적으로 도시한다. 쉽게 알 수 있는 바와 같이, 특정 암은 비교적 많은 수의 네오에피토프를 가지고 있는 반면, 다른 암은 단지 적당한 수의 네오에피토프를 가지고 있다. 또한, 같은 유형의 암 내에서 네오에피토프 발생의 다양성은 동질성이 아니라는 점에 주목해야 한다. 실제로, 일부 암은 네오에피토프의 평균 수가 비교적 낮지만 소정의 임계치 값 100(예컨대, HNSC, LUAD; 점선으로 표시된 임계 값)보다 훨씬 높게 확장되는 높은 변동성을 갖는다. 주목할 만한 것은, 임계치를 초과하는 네오에피토프 카운트를 가진 암은 체크포인트 억제제를 사용한 치료에 반응할 확률이 훨씬 높은 것으로 나타났다(예컨대, UCEC, READ, BLCA, SKCM, LUSC, COAD, STAD). 또한, 이들 암은 또한 전형적으로 MMR 및/또는 MSI와 관련되어 있음이 관찰되었다.
- [0068] 본원에서의 값의 범위를 기술하는 것은 단지 그 범위 내에 속하는 각각의 개별적인 값을 독립적으로 참조하는 간단한 방법으로서 제공하기 위한 것이다. 본원에서 달리 지시하지 않는 한, 각각의 개별 값은 본원에서 개별적으로 인용된 것처럼 명세서에 포함된다. 본원에 기술된 모든 방법은 본원에서 달리 지시되지 않는 한 또는 다르게는 문맥상 명백하게 상충되지 않는 한 임의의 적합한 순서로 수행될 수 있다. 본원의 특정 구현예와 관련하여 제공되는 임의의 및 모든 실시예 또는 예시적인 용어(예를 들어, "~와 같은")의 사용은 단지 본 발명을 보다 잘 나타내도록 의도된 것이며, 달리 청구된 본 발명의 범주를 제한하지 않는다. 명세서에서 어떠한 용어도 본 발명의 실시예 필수적인 청구되지 않은 요소를 나타내는 것으로 해석되어서는 안 된다.
- [0069] 본 발명의 개념을 벗어나지 않고 이미 기술된 것들 이외의 많은 수정이 가능하다는 것이 당업자에게 명백할 것이다. 따라서, 본 발명의 대상은 첨부된 청구범위의 범주를 제외하고는 제한되지 않는다. 또한, 명세서 및 청구범위 모두를 해석함에 있어서, 모든 용어는 문맥에 따라 가능한 가장 넓은 방식으로 해석되어야 한다. 특히, 용어 "포함하는"은 요소, 성분 또는 단계를 비-배타적인 방식으로 참조하여, 그 참조된 요소, 성분 또는 단계가 명시적으로 언급되지 않은 다른 요소, 성분 또는 단계들과 함께 존재하거나 이용되거나 조합될 수 있음을 나타내는 것으로 해석되어야 한다. 명세서가 A, B, C ... 및 N으로 이루어진 군으로부터 선택된 것 중 적어도 하나를 가리키는 경우, 본문은 A와 N, 또는 B와 N 등이 아닌, 군으로부터 단지 하나의 요소만을 필요로 하는 것으로 해석되어야 한다.

도면

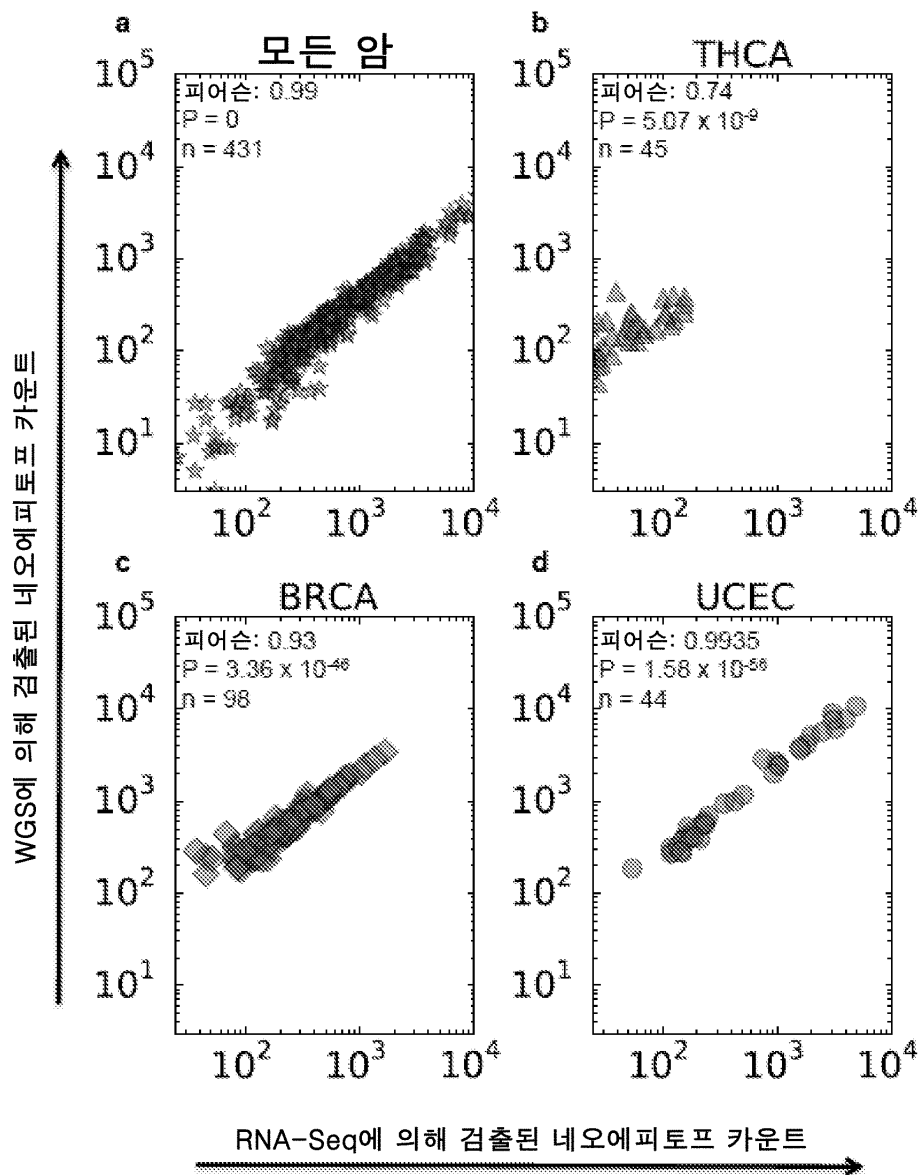
도면1



■ 정상 에피토프   ■ 네오에피토프

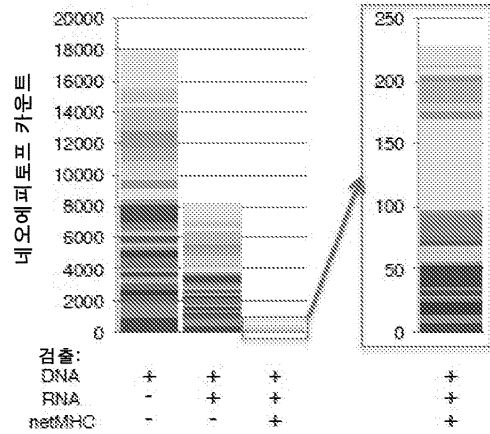


도면2



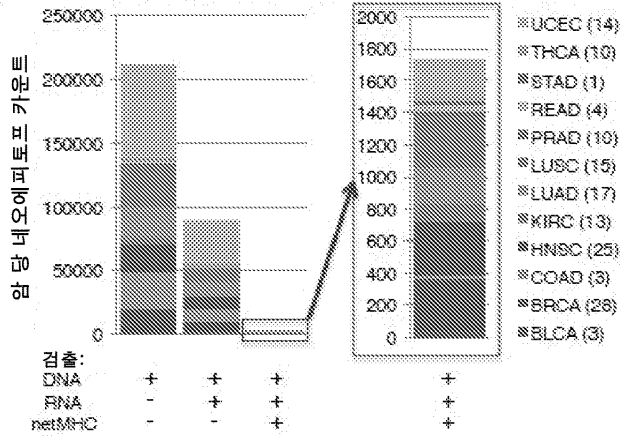
도면3

a 삼중 네거티브 유방암 네오에피토프



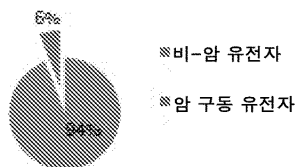
환자 샘플 중에서 공유된 네오에피토프 없음

b A\*02:01 제한된 샘플 네오에피토프



상이한 암의 4개 샘플 중에서 공유된 1개의 네오에피토프

c 네오에피토프의 암 유전자 분류



도면4

