

명세서

발명의 명칭: 개인맞춤형 항암백신 개발을 위한 신항원 선정 방법

기술분야

- [1] 관련 출원들과의 상호 인용
- [2] 본 출원은 2023년 8월 23일자 대한민국 특허출원 제10-2023-0110887호에 기초한 우선권의 이익을 주장하며, 해당 한국 특허 출원의 문헌에 개시된 모든 내용은 본 명세서의 일부로서 포함된다.
- [3] 본 출원은 신항원(종양-특이 면역원성 펩타이드)의 선정 방법, 및 상기 선별된 종양-관련 신항원의 개인맞춤형 항암 백신 제조를 위한 용도를 제공한다.

[4]

배경기술

- [5] 본 명세서에서, 용어 “신항원(Neoantigen)”은 암 특이적 변이로 인해 발생하는 항원, 보다 구체적으로, 암 특이적으로 세포 표면에 발현되는 변이 펩타이드로서 항원결정기를 포함하여 항원으로 인식될 수 있는 펩타이드(표면 단백질)를 의미하는 것으로, 암세포에서만 발현되기 때문에 표적 항암제 개발에 적용이 가능하다. 본 명세서에서, 신항원은 면역원성 펩타이드의 한 종류로, “종양-특이 신항원”, “종양-특이 항원”, “종양-특이 면역원성 펩타이드”로도 표현될 수 있다.
- [6] “개인맞춤형 항암백신(Personalized Cancer Vaccine)”은 환자 개개인에서 발현되는 암 특이적 항원들 중 면역 증폭 능력을 가지고 있는 신항원을 선별하여 단백질 또는 mRNA 형태로 제작하고, Lipid Nanoparticles (LNP 및 Solid Lipid Nanoparticle (SLN), Nanostructured Lipid Carrier (NLC) 등), 리포솜-핵산 복합체(e.g., lipoplex, liposomes 등) 등의 classification을 포함하는 지질 기반 나노입자(Lipid-based Nanoparticles)등과 같은 전달물질을 이용하여 환자에게 투여하여, 환자의 체내 면역체계를 활성화시켜 암세포를 제거하는 치료제 또는 상기 치료제를 사용하는 치료법을 의미한다.
- [7] 개인맞춤형 항암백신의 효능을 기대하기 위해서는 환자 개개인의 신항원을 발굴하여 면역원성이 있는 신항원을 Prioritization 및 선별하는 것이 Key Factor이다. 면역원성이 있는 신항원은 환자의 항원자극에 대한 면역반응, 세포성면역과 체액성면역의 조절 등에 기능하는 조직적합성항원 (Human Leukocyte Antigen: HLA)과 결합력을 가지고 있어야 하고, HLA와 결합된 특정 신항원 서열을 체내 T 세포의 T 세포 수용체 (T cell receptor)가 인식할 수 있을 때, 신항원에 대한 면역활성화가 이루어진다.
- [8] 하지만, 환자마다 발생하는 암 특이적인 변이와 개인마다 발현하는 HLA type이 다르기 때문에 개인맞춤형 항암백신 개발에 적용할 수 있는 면역원성이 있는 신항원을 발굴하는데 어려움이 있다. 더욱이, 면역원성 있는 신항원을 예측할 수 있는 신항원 AI 예측모델의 학습을 위한 Data 수집 및 선별, 개인 맞춤형 항암백

신 개발 시 환자 적용 가능한 신항원을 예측하고 선정하는 방법에 대한 확립이 부재하여, 신항원 예측모델을 효율적으로 사용하는데 어려움이 있다. 종래에는 항암백신에 적용 시 항암효능을 기대할 수 있는 신항원 선정을 위해 환자 암세포의 돌연변이를 포함하는 면역원성 펩타이드와 HLA와의 결합력을 측정하는 AI 예측 모델이 개발되었다. 다만, 환자 에피토프와 HLA와의 결합력이 높다고 해서 반드시 T세포 면역반응이 반드시 높게 나타난 것은 아니어서 도출된 면역원성 펩타이드를 백신으로 사용하는데 어려움이 있었다. 또한 T 세포 면역원성을 예측하기 위해 암세포의 돌연변이 포함 펩타이드와 비암세포의 비돌연변이 포함 펩타이드 간의 아미노산 이화학적 특성을 비교하는 기술도 있었으나 아미노산의 서열 배열 정보만으로는 3차원의 단백질 구조를 반영하지 못하여 암세포의 돌연변이 포함 신항원 펩타이드와 결합한 HLA-펩타이드 Complex가 T 세포 수용체가 종합적으로 인식하여 면역반응을 발생시키기 때문에 HLA와 신항원의 결합력 만으로는 다양한 면역반응을 포괄하지 못해 면역반응 예측 성공률이 높지 않다는 한계가 있었다.

[9]

발명의 상세한 설명

기술적 과제

- [10] 본 출원의 일 예는 종양-특이 면역원성 펩타이드 (신항원, 종양-특이 항원)의 선정 방법(예측 방법, 선택 방법, 확인 방법, 또는 선정에 정보를 제공하는 방법)을 제공한다.
- [11] 상기 종양-특이 면역원성 펩타이드 선정 방법은, (I) 면역원성 펩타이드 예측 모델 구축 단계 및 (II) 종양-특이 면역원성 펩타이드의 예측 및/또는 선별 단계를 포함할 수 있다.
- [12] 상기 (I) 면역원성 펩타이드 예측 모델 구축 단계는 다음의 (a) 내지 (d) 단계를 포함할 수 있다:
- [13] (a) 펩타이드 데이터베이스로부터 펩타이드 정보를 획득하는 단계, 이 때, 상기 펩타이드 정보는 종양세포 또는 종양조직 유래 펩타이드의 서열정보 및 정상세포 또는 정상조직 유래 펩타이드의 아미노산 서열정보, 상기 펩타이드가 결합하는 클래스 I HLA 타입 정보, 상기 펩타이드와 상기 클래스 I HLA 간의 결합력 정보 및/또는 상기 펩타이드의 면역원성 정보 등으로부터 선택된 하나 이상의 정보를 포함할 수 있음;
- [14] (b) 생물학적 서열 데이터베이스로부터 상기 클래스 I HLA의 단백질 서열정보를 획득하는 단계;
- [15] (c) 상기 (a) 및 (b) 단계에서 획득된 정보를 HLA에 대한 결합점수(binding affinity score; BA)를 예측할 수 있는 모델에 학습하여 결합력 예측모델을 구축하고 결합점수를 얻는 단계; 및

- [16] (d) 상기 (a) 및 (b) 단계에서 수득된 정보를 면역원성 점수(immunogenicity score; IMM)를 예측할 수 있는 모델에 학습하여 면역원성 예측모델을 구축하고 면역원성 점수를 얻는 단계.
- [17] 상기 단계 (I)의 (a)에서 사용 가능한 펩타이드 데이터베이스는 IEDB(Immune Epitope Database and Analysis Resource), MHCBN(Comprehensive Database of MHC Binding and Non-binding Peptides) 등으로 이루어진 군에서 선택된 하나 이상의 펩타이드 데이터베이스일 수 있으나, 이에 제한되는 것은 아니며, 펩타이드 정보, 예컨대, 인간 유래(정상세포/정상조직 유래 및 종양세포/종양조직 유래)의 펩타이드의 아미노산 서열 정보와 HLA에 대한 결합 정보(예컨대, IC50, KD 값 등) 및/또는 면역원성 관련 정보를 포함하는 데이터베이스라면 특별한 제한 없이 사용될 수 있다.
- [18] 상기 단계 (I)의 (b)에서 사용 가능한 생물학적 서열 데이터베이스는 NCBI(National Center for Biotechnology Information), Protein 데이터베이스, MHC Motif Atlas 등으로 이루어진 군에서 선택된 하나 이상의 단백질 데이터베이스일 수 있으나, 이에 제한되는 것은 아니며, HLA 단백질의 아미노산 서열을 구할 수 있는 데이터베이스라면 특별한 제한 없이 학습에 사용될 수 있다.
- [19] 상기 펩타이드 데이터베이스 및/또는 생물학적 서열 데이터베이스로부터 수득한 정보는 단계 (c) 및 (d)의 예측 모델 학습에 사용될 수 있다.
- [20]
- [21] 상기 (II) 종양-특이 면역원성 펩타이드 예측 및/또는 선별 단계는 다음의 (1) 내지 (7) 단계를 포함할 수 있다:
- [22] (1) 환자 유래 종양세포 또는 종양조직 및 정상세포 또는 정상조직으로부터 NGS 서열정보를 수득하는 단계;
- [23] (2) 정상세포 또는 정상조직 NGS 서열과 비교하여 상기 환자 유래 종양세포 또는 종양조직 NGS 서열 중 하나 이상의 아미노산 변이를 가진 아미노산 7개 이상의 길이의 종양-특이 펩타이드의 서열정보를 수득하는 단계;
- [24] (3) 상기 종양-특이 펩타이드 또는 상기 종양-특이 펩타이드의 코딩 유전자의 발현수준(TPM; Transcripts per Million)을 측정하는 단계;
- [25] (4) 상기 종양-특이 펩타이드 서열정보를 상기 (c) 단계에서 구축된 결합력 예측 모델에 적용하여 HLA 분자에 대한 결합 점수(binding affinity score; BA)를 측정(확인, 예측)하는 단계;
- [26] (5) 상기 종양-특이 펩타이드 서열정보를 상기 (d) 단계에서 구축된 면역원성 예측 모델에 적용하여 면역원성 점수(immunogenicity score; IMM)를 측정(확인, 예측)하는 단계;
- [27] (6) 하기의 수식으로 계산된 수치를 기준으로 종양-특이 펩타이드의 순위를 정하는 우선순위 결정 단계:
- [28] Neoantigen prioritization score = $\min(\text{TPM}, 1) * (W1 * \text{BA} + W2 * \text{IMM}) / (W1 + W2)$

- [29] $[0 < W1 < 1$ (또는 $0.1 \leq W1 \leq 0.9$), $0 < W2 < 1$ (또는 $0.1 \leq W2 \leq 0.9$) 범위에서 각각 독립적으로 설정; “ $\min(TPM, 1)$ ”은 $0 < TPM \leq 1$ 인 경우 해당 값을 그대로 사용하고, $1 \leq TPM$ 인 경우 $TPM=1$ 로 하여 계산함(TPM 이 1 이상(예컨대, 100, 1000)인 경우 Score에 Bias가 커질 수 있어서 이로 인한 결과 해석의 오류를 줄이기 위함)을 의미함]; 및
- [30] (7) 상기 결정된 종양-특이 펩타이드의 우선순위를 바탕으로 2 이상의 종양-특이 면역원성 펩타이드를 선정하는 단계.
- [31] 상기 단계 (1) 및 (2)의 NGS(next-generation sequencing) 서열정보는 핵산(유전자) 서열 및/또는 이를 아미노산(단백질) 서열로 변환한 서열 정보일 수 있다. 일 예에서, 상기 단계 (1)은 NGS 서열 정보를 취득하는 단계 이후에, 상기 취득된 NGS 서열을 아미노산 서열로 변환하는 단계를 추가로 포함할 수 있다.
- [32] 상기 단계 (1) 및/또는 (2)의 종양세포 또는 종양조직과 정상세포 또는 정상조직은 동일한 환자 유래의 것일 수 있으나, 이에 제한되는 것은 아니다.
- [33] 상기 단계 (3)의 TPM 측정 단계는 종양세포 및/또는 종양조직에서 수행하는 것일 수 있고, 상기 종양세포 및/또는 종양조직은 환자유래 시료로부터 얻어진 것 또는 입수가 가능한 종양(암) 세포주 (cell line)일 수 있다. 상기 단계 (3)의 TPM 측정이 환자 유래 종양세포 또는 종양조직에서 수행되는 경우, 상기 단계 (1) 및/또는 (2)의 종양세포 또는 종양조직과 정상세포 또는 정상조직과 동일한 환자 유래의 것일 수 있으나, 이에 제한되는 것은 아니다.
- [34] 상기 단계 (3)의 TPM 측정 단계, 단계 (4)의 결합 점수 측정 단계, 및 단계 (5)의 면역원성 점수 측정 단계는 수행 순서에 특별한 제한이 없으며, 예를 들어, 순서에 상관없이 순차적으로 수행되거나, 2 이상이 동시에 수행될 수 있으나, 이에 제한되는 것은 아니다.
- [35] 상기 선정된 2 이상의 종양-특이 면역원성 펩타이드는 최종적인 종양-특이 면역원성 펩타이드(신항원)의 후보물질로서 사용될 수 있다.
- [36]
- [37] 일 예에서, 본 명세서에서 제공되는 종양-특이 면역원성 펩타이드 선정 방법은 앞서 설명한 (II) 종양-특이 면역원성 펩타이드의 예측 및/또는 선별 단계 (즉, 단계 (1) 내지 (7))를 포함할 수 있다.
- [38] 다른 예에서, 상기 종양-특이 면역원성 펩타이드 선정 방법은 상기 (II) 종양-특이 면역원성 펩타이드의 예측 및/또는 선별 단계에 더하여, (I) 면역원성 펩타이드 예측 모델 구축 단계 (즉, 단계 (a) 내지 (d))를 추가로 포함할 수 있다. 이 경우, 상기 추가로 포함되는 (I) 면역원성 펩타이드 예측 모델 구축 단계는 상기 (II) 종양-특이 면역원성 펩타이드의 예측 및/또는 선별 단계 전에 포함될 수 있다.
- [39]
- [40] 상기 선정 방법의 상기 단계 (I)의 예측모델 구축에 사용된 펩타이드 데이터베이스에 대하여 하기 i) 내지 iv) 중 선택된 하나 이상의 선별과정을 거쳐서 얻어진 데이터를 사용하는 것일 수 있다:

- [41] i) HLA-A, HLA-B, 및 HLA-C에 해당하지 않는 데이터의 삭제,
- [42] ii) HLA 분자에 돌연변이가 있는 데이터의 삭제,
- [43] iii) 펩타이드의 길이가 7 이하 또는 15 이상인 데이터의 삭제, 및
- [44] iv) 아미노산 서열에 인간 유래 아미노산 20종 외의 아미노산 잔기가 포함된 것의 제거.
- [45]
- [46] 상기 단계 (I) (단계 (a) 내지 (d) 중 하나 이상), 단계 (II) (단계 (1) 내지 (7) 중 하나 이상)는 컴퓨터 등의 전산 처리 디바이스, 또는 전산처리 장치에서 수행되는 것일 수 있다.
- [47] 상기 단계 (I)의 결합력 예측모델의 학습에 사용 가능한 펩타이드 데이터베이스는 IEDB(Immune Epitope Database and Analysis Resource), MHCBN(Comprehensive Database of MHC Binding and Non-binding Peptides) 등으로 이루어진 군에서 선택된 하나 이상의 펩타이드 데이터베이스일 수 있으나, 이에 제한되는 것은 아니며, 펩타이드 정보, 예컨대, 인간 유래의 펩타이드의 아미노산 서열 정보와 HLA에 대한 결합 정보(예컨대, IC50, KD 값 등) 및/또는 면역원성 관련 정보를 포함하는 데이터베이스라면 특별한 제한 없이 학습에 사용될 수 있다.
- [48] 일 구체예에서는 면역원성 점수(IMM)를 확인(측정, 결정)하는 단계에서 펩타이드-MHC(HLA) 결합체가 T세포 수용체(TCR)에 결합할 때 면역반응과 관련된 2 이상의 면역원성 모델 학습 데이터를 포함한다.
- [49] 상기 면역원성 모델 학습 데이터는 펩타이드 데이터베이스 또는 관련 문헌으로부터 수득된 펩타이드 후보군 서열에 대한 모든 면역원성 관련 정보, 예컨대, INF-g, TNF-a, IL-2, IL-1b 등을 포함하는 T세포 유래 사이토카인 분비량, 면역원성 활성화로 인한 T세포 proliferation, CCL4, CXCL9을 포함하는 Chemokine 수준, Granzyme B 등 활성화된 T세포의 cytotoxicity를 측정한 값 등으로 이루어진 군에서 선택된 2종 이상일 수 있다. 구체적으로, 상기 펩타이드 데이터베이스는 IEDB(Immune Epitope Database and Analysis Resource), MHCBN(Comprehensive Database of MHC Binding and Non-binding Peptides)로 이루어진 군에서 선택된 하나 이상의 펩타이드 데이터베이스일 수 있으나, 이에 제한되는 것은 아니며 펩타이드 정보, 예컨대, 인간 유래의 펩타이드의 아미노산 서열 정보와 면역원성 측정값을 포함하는 데이터베이스라면 특별한 제한 없이 학습에 사용될 수 있다.
- [50] 상기 면역원성 측정값은 ELISPOT, ELISA, Flow cytometry intracellular staining, binding assay, Proliferation assay (예, T cell Proliferation assay 등), cytotoxicity assay (예, in vitro/ex vivo/in vivo Cytotoxicity assay 등) 등을 포함하는 방법으로 측정된 값이다.
- [51] 상기 면역원성 모델 학습 데이터는 정성적으로 측정된 값 또는 정량적으로 측정된 값을 모두 포함하며 그 중 정성적으로 측정된 값은 음성대조군 대비 값을 베타 분포 변환 등 통계적으로 수치화하여 데이터로 포함한다.

- [52] 상기 단계 (II)는 개인맞춤 대상 환자(포유류; 예컨대, 인간)에서 유래한 시료(종양 세포/조직 및 정상 세포/조직)에서 얻어진 NGS 서열 정보 및/또는 이를 펩타이드 정보로 변환한 정보를 사용하여 수행하는 것일 수 있다.
- [53] 상기 종양-특이 면역원성 펩타이드는 암세포 돌연변이에서 유래한 펩타이드일 수 있다. 일 예에서, 상기 종양-특이 면역원성 펩타이드는 환자 유래 시료(예컨대, 암세포)의 유전체 서열 분석(예컨대, 차세대 유전체 서열분석(NGS; Next Generation Sequencing) 등)에 의하여 얻어진 서열로부터 도출된 것(예컨대, NGS로부터 얻어진 핵산 서열 또는 이로부터 변환된 아미노산 서열)일 수 있다. 보다 구체적으로, 상기 종양-특이 면역원성 펩타이드는 환자 유래 시료(예컨대, 암세포)의 유전체 서열 분석(예컨대, 차세대 유전체 서열분석(NGS; Next Generation Sequencing) 등)에 의하여 얻어진 서열(핵산 서열 또는 이로부터 변환된 아미노산 서열) 중에서 정상 세포와 상이한 서열(예, 정상과 비교하여 돌연변이가 발생한 서열)로부터 도출된 것일 수 있다.
- [54] 상기 HLA 분자는 인간 유래의 주조직적합성복합체(major histocompatibility complex: MHC) 분자들로 이루어진 군에서 선택된 하나 이상일 수 있다. 상기 HLA 분자는 상이한 HLA 분자 유형 및/또는 상이한 HLA 대립유전자를 포함하는 것일 수 있다. 일 예에서, 상기 HLA 분자는 HLA 타입 I 분자일 수 있다.
- [55]
- [56] 상기 면역원성 펩타이드 예측 모델 구축 단계(단계 (I))에 있어서, 상기 단계 (c)의 결합 점수는 상기 면역원성 펩타이드 및 HLA 분자의 아미노산 서열 정보를 이용하여 2 가지 이상의 방식으로 학습시켜 얻어진 예측값을 조합(ensemble)하여 얻어지는 것일 수 있다. 상기 HLA 분자의 아미노산 서열 정보는 앞서 설명한 후보 펩타이드의 데이터베이스로부터 얻어진 것일 수 있으나, 이에 제한되는 것은 아니다.
- [57] 상기 단계 (d)의 면역원성 점수는 상기 면역원성 펩타이드를 면역세포를 포함하는 시료에 처리한 경우의 사이토카인 분비량에 의하여 얻어지는 것일 수 있다. 상기 면역세포를 포함하는 시료는 혈액, 백혈구, 및 말초혈액단핵구(PBMC)로 이루어진 군에서 선택되고, 상기 사이토카인은 IFN-g, IL-2, TNF-a, 및 granzyme B로 이루어진 군에서 선택된 1종 이상일 수 있으나, 이에 제한되는 것은 아니다. 일 예에서, 상기 단계 (2)의 면역원성 점수는 상기 후보 펩타이드를 면역세포(예컨대, PBMC 등) 또는 이를 포함하는 시료에 처리한 경우의 사이토카인 분비량 정보에 의하여 결정되는 것일 수 있다. 상기 사이토카인 분비량 정보는, 예컨대, IEDB, MHCBN 등으로 이루어진 군에서 선택된 데이터베이스로부터 얻을 수 있으나, 이에 제한되는 것은 아니다.
- [58] 상기 면역원성 점수는 전체 실험 횟수 대비 양성 실험의 비율로서 간접적으로 측정될 수 있다. 일 예에서, 상기 IEDB 데이터베이스는 사이토카인 분비량의 정량 측정값 대신 정성 측정값, 실험 횟수(tested), 반응 횟수(responded)와 비율(frequency)를 데이터로 제공하며, 이를 이용하여 면역원성 점수의 수치화를 위한

학습데이터로 사용할 수 있다. 실제 실험값은 실험횟수 대비 반응횟수 비율을 반영하여 수치로 변환하여, 학습 데이터로서 면역원성 예측모델의 재학습시 이용될 수 있다. 상기 수치화는 평균, 중간값, 베타 분포 변환으로 이루어진 군에서 선택된 하나 이상의 방법으로 수행될 수 있다.

- [59] 상기 단계 (d)의 면역원성 점수는 상기 면역원성 학습 데이터의 펩타이드 아미노산 서열 정보와 상기 면역세포의 HLA 분자에 해당하는 아미노산 서열 및 상기 펩타이드의 면역세포 처리시의 사이토카인 분비량 정보를 이용하여 2 가지 이상의 방식으로 학습시켜 얻어진 예측값을 조합(ensemble)하여 얻어지는 것일 수 있다.
- [60] 상기 단계 (c)와 (d)의 조합(ensemble)은 2개 이상의 서로 다른 학습 방식으로 훈련된 2개 이상의 예측 모델 중에서, 학습 데이터와 다른 검증(Validation) 데이터 셋에 대한 예측 오차가 가장 작은 모델 2개 이상을 선별하여 예측값을 조합하는 것일 수 있다. 구체적으로, 상기 검증 데이터셋은 상기 모델 학습 데이터의 일부를 미리 분리해 놓은 것일 수 있다.
- [61] 상기 단계 (c)(결합 예측 모델링 및 점수 측정) 및/또는 단계 (d)(면역원성 예측 모델링 및 점수 측정)에 있어서,
- [62] - 상기 학습은 GRU(Gated Recurrent Unit), Transformer, FFN(Feed Forward Neural-network), GNN(Graph Neural Network) 등 통상적으로 사용되는 AI(인공지능) 학습모델들로부터 선택된 1가지 이상의 모델로 수행될 수 있고,
- [63] - 상기 단계 (c)(결합 점수 측정) 및 단계 (d)(면역원성 점수 측정)의 학습은 관련 분야에서 통상적으로 사용되는 학습방식의 적용 및 미적용 중에서 선택한 1개의 학습 방법과 상기 AI 모델을 조합하여 수행될 수 있으며,
- [64] - 상기 결합 점수 및/또는 면역원성 점수는 상기 1가지 이상의 AI 모델 및 1가지 이상의 학습방법으로 학습시켜 얻어진 각각의 예측모델로부터 얻어진 예측값을 통상의 ensemble 모델에 의하여 조합(ensemble)한 최종 예측값일 수 있고, 일 예에서 상기 최종 예측값은 [(예측값의 평균)-(예측값의 분산)]으로 얻어지는 것일 수 있다
- [65]
- [66] 상기 종양-특이 면역원성 펩타이드 예측 및/또는 선별 단계 (단계 (II))에 있어서, 단계 (1)은 환자 유래 종양세포 또는 종양조직의 NGS 서열과 환자 유래 정상 세포 또는 정상조직의 NGS 서열 간의 생물정보 분석 과정을 통해 수행될 수 있다. 상기 종양세포 또는 종양조직과 정상세포 또는 정상조직은 동일한 환자로부터 유래한 것일 수 있다. 상기 생물정보 분석 과정은 다음과 같은 i)에서 vii)의 과정을 포함할 수 있으나, 이에 제한되지 않고, 세부 과정이 추가되거나 제거될 수 있다.
- [67] i) NGS 데이터의 무결성 확인; 일 예로서 FastQC와 같은 software를 사용할 수 있으나 이에 제한되지 않음.

- [68] ii) 저품질 데이터의 제거; 일 예로서 Cutadapt와 같은 software를 사용할 수 있으나 이에 제한되지 않음.
- [69] iii) 참조유전체(reference genome)로의 서열 정렬(sequence alignment); 일 예로서 hg19 참조 유전체와 bwa software를 사용할 수 있으나 이에 제한되지 않음.
- [70] iv) 서열 정렬 결과의 품질 확인 및 보정; 일 예로서 GATK와 같은 software를 이용하여 Indel realignment, BQSR (Base Quality Score Recalibration) 등을 적용할 수 있으나 이에 제한되지 않음.
- [71] v) 종양-특이 변이(mutation) 발굴; 일 예로서 GATK, Strelka, Mutect2 등과 같은 software를 사용할 수 있으나 이에 제한되지 않음;
- [72] vi) 종양-특이 변이의 분류; 일 예로서 SnpEff와 같은 software로 변이를 분류하고, 이 중에서 펩타이드 서열에 영향을 주는 missense, insertion, deletion 변이만을 선별할 수 있으나 이에 제한되지 않음
- [73] vii) 종양-특이 변이를 포함하여 생성 가능한 길이 7 이상의 펩타이드 서열 추출
- [74] 상기 단계 (3)의 TPM(transcriptome per million)은 염기서열로 구성된 유전자가 전사되어 전사체를 형성하는 과정인 발현에 있어서, 그 발현된 정도를 측정한 발현값으로, 발현값의 측정 방법에는 mRNA 농도의 측정, 생성 전사체 농도의 측정 등이 있으나, 이에 제한되지 않고 다양한 방법으로 발현값을 측정할 수 있다. 본 발명의 일 실시예에 따른, 상기 발현값은 발현체(transcriptome)에 맵핑된 NGS 결과 얻어진 read 개수로서, TPM 단위로 산출될 수 있으나, 이에 제한되지 않고 더 다양한 단위로 산출되어 발현값을 나타낼 수 있다. 돌연변이 염기서열 또는 전체 RNA 염기서열 중에서 TPM (transcriptome per million)이 1 이상인 RNA 염기서열을 필터링하는 단계를 더 포함할 수 있다.
- [75] 상기 단계 (3)의 종양-특이 펩타이드 또는 이의 코딩 유전자의 발현수준(예컨대, TPM, Transcripts Per Million; Li et al. (2010); 각 transcript(맵핑된 NGS read 개수)가 라이브러리에서 얼마만큼의 비율을 차지하는지를 기반으로 정규화 함, 한 라이브러리에서 TPM 값을 전부 더하면 1,000,000이 됨, 서로 다른 샘플 간 비교가 가능함)은 상기 종양-특이 펩타이드가 실제 환자 시료(예컨대, 암세포 등)에 존재할 가능성(예컨대, 암세포의 유전자 돌연변이가 전사(Transcription) 및 번역(Translation)을 거쳐 단백질로 생성될 가능성, 빈도, 또는 정도)를 의미하는 것일 수 있다. 상기 단계 (3)의 종양-특이 펩타이드 또는 이의 코딩 유전자의 발현수준은 맞춤 치료 대상인 환자 시료 데이터의 서열 분석을 통하여 얻어지는 것일 수 있고, 보다 구체적으로 microarray 또는 RNA-seq 데이터를 처리하여 계산된 것일 수 있다. 일 예에서 환자 시료 RNA-seq 데이터를 분석하여 상기 종양-특이 펩타이드 또는 이의 코딩 유전자의 TPM을 계산할 수 있으며 상기 TPM값이 1 이상인 상기 종양-특이 펩타이드에 우선순위를 부여할 수 있다. 예컨대, TPM은 유전자 길이와 데이터 생산량으로 정규화된 RNA-seq 데이터 백만 개 당 해당 유전자의 발현량을 의미하므로, $TPM \geq 1$ 를 우선순위 부여에 사용함으로써 전체 RNA-seq

데이터 중 최소 백만 분의 일 이상의 지분을 차지하여 상기 후보 펩타이드 중 실제 환자 시료에 존재할 가능성이 높은 것에 우선순위를 부여할 수 있다.

- [76] 상기 단계 (6)의 우선순위 결정단계는 아래의 수식을 사용하여 수행된다:
- [77] $\text{Neoantigen prioritization score} = \min(\text{TPM}, 1) * (\text{W1} * \text{BA} + \text{W2} * \text{IMM}) / (\text{W1} + \text{W2})$.
- [78] 상기 수식에서 W1 및 W2는 각각 결합점수(BA) 및 면역원성 점수(IMM)의 가중치를 의미하는 것으로, $0 < \text{W1} < 1$ (또는 $0.1 \leq \text{W1} \leq 0.9$), $0 < \text{W2} < 1$ (또는 $0.1 \leq \text{W2} \leq 0.9$) 범위서 각각 독립적으로 설정할 수 있으며, 일 예에서, $\text{W1} + \text{W2} = 1$ 일 수 있으나, 이에 제한되는 것은 아니다.
- [79] 일 예에서, 상기 수식은 $\text{TPM} > 1$, $0.6 > \text{결합력 점수(BA)} \geq 0.1$, 및 면역원성 점수(IMM) ≥ 0.25 의 조건을 만족하는 후보 펩타이드에 대하여 수행되는 것일 수 있으나, 이에 제한되는 것은 아니다.
- [80]
- [81] 상기 단계 (7)의 종양-특이 면역원성 펩타이드(신항원)의 선정 단계는 상기 단계 (6)의 결과가 높은 순서대로 정렬하여 상위 2~100개, 2~75개, 2~50개, 2~40개, 2~30개, 2~20개, 2~10개, 5~100개, 5~75개, 5~50개, 5~40개, 5~30개, 5~20개, 5~10개, 10~100개, 10~75개, 10~50개, 10~40개, 10~30개, 10~20개, 20~100개, 20~75개, 20~50개, 20~40개, 또는 20~30개의 펩타이드 중에서 선택된 하나 이상을 종양-특이 면역원성 펩타이드로 선택하는 단계를 추가로 포함할 수 있다.
- [82]
- [83] 본 명세서에서 제공되는 신항원(종양-특이 면역원성 펩타이드) 예측 및 개인 맞춤형 항암백신이 mRNA로 제작될 경우에는 특별히 Hydrophobicity에 대한 고려가 필요하지 않을 수 있지만, 펩타이드로 항암백신을 만들거나 면역원성 시험의 진행 시에는, 펩타이드의 Hydrophobicity가 높으면 (예컨대, hydrophobic 아미노산 함량이 높음) 생산이 불가능하기 때문에, 펩타이드의 Hydrophobicity를 측정하는 것이 필요할 수 있다. 일 구체예에서, 상기 Hydrophobicity는 Gravy (Grand average of hydropathy) score를 계산하여 측정할 수 있으나, 이에 제한되는 것은 아니다.
- [84]
- [85] 본 명세서에서 제공되는 종양-특이 면역원성 펩타이드의 선정 방법에 의하여 선별된 종양-관련 면역원성 펩타이드는 암(종양) 시료에 특이적으로 존재하는 것으로, 암(종양)을 가지지 않는 정상 시료에는 존재하지 않는 펩타이드 (예컨대, 변이 펩타이드)일 수 있다. 따라서, 상기 종양-특이 면역원성 펩타이드는 암의 예방 및/또는 치료용 약학 조성물, 예컨대, 항암 백신, 특히 상기 선정 방법에 사용된 환자의 개인맞춤형 항암백신의 제조에 유용하게 사용될 수 있다.
- [86] 다른 예는 상기 종양-관련 면역원성 펩타이드의 선정 방법으로 선별된 종양-특이 면역원성 펩타이드(신항원) 및/또는 이를 암호화하는 유전자(mRNA, DNA 등)를 포함하는 항암 백신 조성물을 제공한다. 상기 항암 백신 조성물은 약학적으로

허용가능한 담체 및/또는 부형제 및/또는 1종 이상의 아쥬반트, 안정화제 등을 추가로 포함할 수 있다.

- [87] 다른 예는 상기 종양-특이 면역원성 펩타이드의 선정 방법으로 선별된 종양-특이 면역원성 펩타이드 및/또는 이를 암호화하는 유전자(mRNA, DNA 등)과, 약학적으로 허용가능한 담체 및/또는 및/또는 1종 이상의 아쥬반트, 안정화제 등을 혼합하는 단계를 포함하는, 항암 백신 제조 방법을 제공한다.
- [88] 상기 담체 및/또는 부형제 및/또는 1종 이상의 아쥬반트, 안정화제 등은 백신 제제화에 통상적으로 사용되는 모든 약학적으로 허용가능한 물질들 중에서 선택된 것일 수 있다.
- [89] 상기 백신은 치료용 백신 또는 예방적 백신의 형태일 수 있다.
- [90] 다른 예는 상기 종양-특이 면역원성 펩타이드의 선정 방법으로 선별된 종양-특이 면역원성 펩타이드 또는 상기 백신을 암의 예방 및/또는 치료를 필요로 하는 환자에게 투여하는 단계를 포함하는, 암의 예방 및/또는 치료 방법을 제공한다.
- [91] 상기 암의 치료는 다른 항암물질과의 병용, 및/또는 외과학적 절제 및/또는 방사능치료 및/또는 전통적인 화학요법과 조합될 수 있다. 상기 다른 항암 물질은 공지된 모든 항암제 (화학 약물, 항체 등) 중에서 선택될 수 있으며, 예를 들어 면역항암제 (immune checkpoint inhibitor (예컨대, PD-1 억제제 및/또는 PD-L1 억제제 등) 등)일 수 있으나, 이에 제한되는 것은 아니다.

[92]

과제 해결 수단

- [93] 암 특이적 서열 돌연변이 중 면역원성이 강한 신항원 발굴을 위해 HLA와 신항원 sequence의 결합력을 예측하는 Binding Affinity 모델과 신항원 Sequence의 면역원성을 예측하는 Immunogenicity 모델을 병합한 AI 학습 기술을 이용한 신항원 예측모델을 개발하였다. 신항원 예측 모델은 Immune Epitope Database and Analysis Resource (IEDB)에 있는 데이터베이스를 기반으로 각 모델의 알고리즘 학습이 진행되고 데이터 업데이트를 통한 *in silico* 개선이 이루어졌다.
- [94]
- [95] AI 신항원 예측모델의 학습을 통해 면역원성 있는 신항원 후보 예측을 가능하게 하고 예측모델을 이용하여 암환자 NGS Data를 적용해 신항원 후보군을 도출하였다. 예측모델 적용시 사용되는 Binding Affinity 모델과 Immunogenicity 모델에서 예측된 신항원 후보군에 대한 Threshold를 정하고 신항원에 대하여 펩타이드를 제작해 Binding Affinity 시험 및 면역원성 시험 (예, IFN γ ELISPOT)을 진행하여 모델에 대한 신항원 예측을 검증 및 개인 맞춤형 항암 백신에 적용 가능한 신항원 선정하였다.
- [96]
- [97] 일 측면에서의 본 출원의 구성을 예시적으로 기재하면 다음과 같다:

- [98] 신항원 예측모델 (Transformer based Binding Affinity Model + Immunogenicity model);
- [99] 신항원 예측모델 Input 데이터 선정;
- [100] 신항원 예측모델 학습방식;
- [101] 환자 Data set의 WES(whole exosome sequencing; mutation selection), RNA-seq (expression level) 데이터 처리 방식;
- [102] Neoantigen Prioritization (TPM, BA Score 와 IMM Score를 이용하여 Neoantigen Ranking Score 확인;
- [103] Binding Affinity 와 면역원성 시험을 진행하여 예측된 신항원의 개인 맞춤형 항암백신 적용 가능성 검증.
- [104]
- [105] 용어의 정의
- [106] 본 발명에서, 용어 "펩타이드"는 펩타이드 결합에 의해 공유적으로 연결된 2개 이상, 3개 이상, 4개 이상, 6개 이상, 8개 이상, 10개 이상, 13개 이상, 16개 이상, 21개 이상 및 최대 8, 10, 20, 30, 40, 50 또는 100개의 아미노산을 포함하는 물질을 의미한다. 용어 "폴리펩타이드" 또는 "단백질"은 펩타이드보다 많은 아미노산 잔기가 펩타이드 결합에 의하여 공유적으로 연결된 물질을 의미할 수 있으나, 본 명세서에서는 "펩타이드", "폴리펩타이드" 및 "단백질"이 호환적으로 사용될 수 있다.
- [107] 본 발명에서, 펩타이드, 폴리펩타이드 또는 단백질과 관련하여 용어 "변형 (modification, mutation)"은 야생형 펩타이드, 폴리펩타이드 또는 단백질의 원래 서열(parental sequence)에 변이가 도입된 것을 의미할 수 있으며, 상기 변이는 아미노산 삽입 변이, 아미노산 부가 변이, 아미노산 결실 변이 및 아미노산 치환 변이 등으로 이루어진 군에서 선택된 것일 수 있으며, 예컨대, 아미노산 치환 변이일 수 있다. 본 명세서에서, 모든 서열 변화는 잠재적으로 새로운 에피토프를 만들 수 있다.
- [108] 본 명세서에서, 용어 "면역원성 펩타이드(immunogenic peptide)"는 항원결정기 (에피토프)를 포함하여 면역반응을 유도할 수 있는 펩타이드를 의미한다.
- [109] 본 명세서에서, 용어 "신항원(neoantigen)"은 모(parental) 펩타이드 또는 단백질에 비해 하나 이상의 아미노산 변형을 포함하는 펩타이드 또는 단백질일 수 있으며, 상기 아미노산 변형은 종양-특이 돌연변이로 인한 아미노산 변형일 수 있다. 본 명세서에서, 신항원은 면역원성 펩타이드의 한 종류로서, 종양(암) 특이적인 돌연변이를 포함하며, 정상세포에서는 존재하지 않거나 검출되지 않는 것일 수 있다. 상기 신항원은 종양-특이 신항원(종양-특이 항원; tumor-specific antigen), 종양-특이 면역원성 펩타이드로도 표현될 수 있다.
- [110] 용어 "면역 반응(immune response)"은 항원과 같은 표적에 대한 통합적인 신체 반응을 의미하며, 예컨대, 세포성 면역 반응 또는 세포성 및 체액성 면역 반응을

의미할 수 있다. 면역 반응은 보호적/예방적/방지적 및/또는 치료적인 반응일 수 있다.

- [111] "면역 반응을 유도하는"이라 함은 유도 전에는 면역 반응이 없었음을 가리킬 수 있으나, 유도 전에 일정 수준의 면역 반응이 있었고 유도 후 상기 면역 반응이 증강된 것을 의미할 수도 있다. 따라서, "면역 반응을 유도하는"은 "면역 반응을 증강하는"의 의미도 포함한다. 예컨대, 소정의 대상(환자)에서 면역 반응 유도 후, 상기 대상은 암 질환과 같은 질환의 발병으로부터 보호되거나 또는 면역 반응의 유도에 의해 그 질환 상태가 개선, 호전 및/또는 경감될 수 있다.
- [112] "세포성 면역 반응(cellular immune response)" 및 "세포성 반응(cellular response)" 또는 이와 유사한 용어 들은 "헬퍼" 또는 "킬러"로서 작용하는 T 세포 또는 T-림프구와 연관된 클래스 I 또는 클래스 II MHC에 의한 항원의 제시에 의해 특징지어지는 세포에 지향된 면역 반응을 의미한다. 헬퍼 T 세포(CD4+ T 세포라고도 칭해짐)는 면역 반응을 조절하는 중추적 역할을 하고 킬러 세포(CD8+ T 세포, 세포독성 T 세포, 세포용해 T 세포, CD8+ T 세포 또는 CTLs이라고도 칭해짐)는 암 세포와 같은 질병 세포를 사멸시켜, 질병 상태의 세포의 생성을 방지한다. 바람직한 구체예에서, 본 발명은 하나 이상의 종양-발현 항원을 발현하는 종양 세포에 대한 항종양 CTL 반응의 자극, 및/또는 HLA 타입 I 또는 클래스 I MHC에 의해 이러한 종양-발현 항원을 제시하는 것을 포함할 수 있다.
- [113] "항원"은 항체 또는 T-림프구 (T 세포)와의 특이 반응과 같은 면역 반응의 표적 및 또는 그러한 면역 반응을 유도하는 모든 물질, 예컨대 펩타이드 또는 단백질을 포괄한다. 예컨대, 항원은 T 세포 에피토프와 같은 적어도 하나의 에피토프를 포함한다. 본 명세서에서 항원은 임의로 프로세싱 후, 그 항원에 특이적인 면역 반응을 유도하는 분자일 수 있다 (항원을 발현하는 세포를 포함함). 항원 또는 그의 T 세포 에피토프는 세포, 예컨대 질병 세포, 특히 암 세포를 포함하는 항원 제시 세포에 의해 MHC 분자 관점에서 제시되어 항원(항원을 발현하는 세포를 포함)에 대해 면역 반응을 일으키는 것일 수 있다.
- [114] 본 명세서에서, 용어 "종양-특이 항원", "종양-특이 발현 항원", "암-특이 항원" 및 "암-특이 발현 항원"은 신항원과 동등한 의미를 가지고 호환적으로 사용될 수 있다.
- [115] 용어 "면역원성(immunogenicity)"은 암에 대한 치료(예컨대, 항체 형성 등)와 같은 치료와 연관된 면역 반응을 유도하는 효능을 의미할 수 있다.
- [116] 용어 "주조직적합 복합체(major histocompatibility complex; MHC)"는 MHC 클래스 I 및 MHC 클래스 II 분자를 포함하며, 모든 척추동물에서 일어나는 유전자 복합체에 관련된다. MHC 단백질 또는 분자는 면역 반응에서 림프구와 항원 제시 세포 또는 병에 걸린 세포 간의 시그널링에 있어서 중요하며, 여기서 MHC 단백질 또는 분자는 펩타이드와 결합하여 이들을 T 세포 수용체에 의한 인식을 위해 제시한다. MHC에 의해 코딩된 단백질 들은 세포 표면에서 발현되어, 자가 항원

(그 세포 자체로부터의 펩타이드 단편) 및 비자기 항원(예컨대 침입 미 생물로부터의 단편) 두 가지 모두를 T 세포에 대해 나타낸다.

- [117] MHC 영역은 3 가지 서브그룹, 즉 클래스 I, 클래스 II, 및 클래스 III로 구분된다. MHC 클래스 I 단백질은 α -사슬 및 β 2-마이크로글로불린(15번 염색체에 의해 코딩된 MHC의 일부가 아님)을 함유한다. 이들은 세포독성 T 세포에 항원 단편을 제시한다. 대부분의 면역계 세포 상에서, 특히 항원-제시 세포 상에서, MHC 클래스 II 단백질은 α 및 β -사슬을 함유하고 이들은 항원 단편을 T-헬퍼 세포에 제시한다. MHC 클래스 III 영역은 보체 성분 및 사이토카인을 코딩하는 것들과 같은, 다른 면역 성분을 코딩한다.
- [118] MHC는 다유전자성(polygenic) (몇몇 MHC 클래스 I 및 MHC 클래스 II 유전자들이 있음)이면서 다형성(polymorphic) (각 유전자마다 복수의 대립유전자들이 있음)이다.
- [119] 본 발명에서, 용어 "일배체형(haplotype)"은 하나의 염색체와 그에 의해 코딩된 단백질 상에서 발견되는 HLA 대립유전자를 의미한다. 일배체형은 또한 MHC 내의 어떤 여하한 위치에 존재하는 대립유전자를 가리킬 수도 있다.
- [120] MHC의 각 클래스는 몇몇 위치에 의해 나타낼 수 있다: 예컨대,
- [121] - 클래스 I의 경우, HLA(Human Leukocyte Antigen)-A, HLA-B, HLA-C, HLA-E, HLA-F, HLA-G, HLA-H, HLA-J, HLA-K, HLA-L, HLA-P, HLA-V 등; 및
- [122] - 클래스 II의 경우, HLA-DR (HLA-DRA, HLA-DRB1-9 등), HLA-DQ (HLA-DQA1, HLA-DQB1 등), HLA-DP (HLA-DPA1, HLA-DPB1 등), HLA-DM (HLA-DMA, HLA-DMB 등), HLA-DO (HLA-DOA, HLA-DOB 등) 등.
- [123] 본 명세서에서, 용어 "HLA 또는 HLA 대립유전자" 및 "MHC 또는 MHC 대립유전자"는 호환적으로 사용될 수 있다.
- [124] MHCs는 매우 강한 다형성을 나타낸다. 인간 집단에서 각 유전자 자리마다, 구별되는 대립유전자를 포함하는 많은 수의 일배체형들이 존재한다. 클래스 I 및 클래스 II 두 가지 모두의 상이한 다형성 MHC 대립유전자들은 상이한 펩타이드 특이성을 갖는다: 각 대립유전자는 특정 서열 패턴을 나타내는 펩타이드와 결합하는 단백질을 코딩 한다.
- [125] 일 구체예에서 MHC 분자는 HLA 분자일 수 있다. HLA 분자를 HLA DNA 유형에 따른 혈청학적 유형으로 표기 시, "HLA dictionary 2008"(John Wiley & Sons A/S; Tissue Antigens 73, 95-170; 본 명세서에 참조로 포함됨)를 참조하여 표기할 수 있다.
- [126] 용어 "MHC 결합 펩타이드"는 MHC 클래스 I 및/또는 클래스 II 결합 펩타이드를 생산하도록 프로세싱될 수 있는 MHC 클래스 I 및/또는 클래스 II 결합 펩타이드 또는 펩타이드를 포함할 수 있다. 클래스 I MHC/펩타이드 복합체의 경우, 상기 펩타이드(결합 펩타이드)는 일반적으로 7-12, 7-10, 8-12, 또는 8-10 아미노산 길이일 수 있으나, 이에 제한되지 않는다. 클래스 II MHC/펩타이드 복합체의 경

우, 상기 펩타이드(결합 펩타이드)는 대체로 9-30 또는 10-25 아미노산 길이, 예컨대, 13-18 아미노산 길이일 수 있으나, 이에 제한되지 않는다.

- [127] 일 예에서, 상기 펩타이드가 직접, 즉 프로세싱 없이, 특히 절단 없이, 제시되는 경우, 그 길이는 MHC 분자, 특히 클래스 I MHC 분자에 결합하기에 적당한 길이, 예컨대, 7-30, 7-20, 7-12, 8-11, 9, 또는 10 아미노산 길이일 수 있다. 다른 예에서, 상기 펩타이드가 부가적인 서열을 포함하는 보다 큰 entity, 예컨대 백신 서열 또는 폴리펩타이드의 일부이고, 프로세싱 특히 절단 후에 제시되는 경우, 프로세싱에 의해 생산되는 그 펩타이드의 길이는 MHC 분자, 특히 클래스 I MHC 분자에 결합하기에 적당한 길이, 예컨대 7-30, 7-20, 7-12, 8-11, 9, 또는 10 아미노산 길이일 수 있다.
- [128] 본 명세서에서 제공되는 방법에 의하여 선정된 종양-특이 면역원성 펩타이드는 클래스 I MHC(HLA) 분자에 결합하는 것일 수 있다.
- [129] 용어 "에피토프"는 항원과 같은 분자 내의 항원 결정기(antigenic determinant), 즉 면역계에 의해 인식되는 분자의 일부 또는 단편을 의미한다. 종양 항원과 같은 단백질의 에피토프는 상기 단백질의 연속 또는 불연속 부분을 포함하며, 그 길이가 5 내지 100, 5 내지 50, 8 내지 30, 또는 10 내지 25 아미노산일 수 있다. 본 명세서에서, 에피토프는 T 세포 에피토프일 수 있다.
- [130] 본 명세서에서 에피토프는 세포 표면에서 MHC 분자에 결합할 수 있는 "MHC 결합 펩타이드"일 수 있다.
- [131] 본 명세서에서 용어 "네오-에피토프(neo-epitope)"는 정상적인 비암 세포 또는 생식세포와 같은 레퍼런스에서는 존재하지 않지만, 암 세포에서는 발견되는 에피토프를 의미한다. 여기에는 특히, 정상적인 비암 세포 또는 생식 세포에서는 대응하는 에피토프가 발견되지만, 암 세포에서의 1 이상의 돌연변이로 인해, 에피토프의 서열이 변화되어 네오-에피토프가 야기되는 상황도 포함된다.
- [132] 본 발명에서, "T 세포 에피토프"라는 용어는 T 세포 수용체에 의해 인식되는 배열(configuration)에서 MHC 분자에 결합하는 펩타이드를 의미한다. 일반적으로, T 세포 에피토프는 항원-제시 세포의 표면에 존재한다.
- [133] T 세포 에피토프는 어떤 항원 단편의 아미노산 서열에 실질적으로 상응하는 아미노산 서열을 포함하는 것일 수 있다. 예컨대, 상기 항원 단편은 MHC 클래스 I 및/또는 클래스 II 제시된 펩타이드일 수 있다.
- [134] 본 발명에 따른 T 세포 에피토프는 면역 반응을 자극할 수 있는 항원의 일부 또는 단편, 또는 그 항원의 발현 및 종기로는 질병 세포 특히 암 세포와 같은 항원의 제시에 의해 특징지어지는 항원이나 세포에 대한 세포 반응을 자극할 수 있는 항원의 일부 또는 단편과 관련된 것일 수 있다. 예컨대, T 세포 에피토프는 클래스 I MHC에 의한 항원 제시에 의해 특징지어지는 세포에 대한 세포성 반응을 자극할 수 있는 것일 수 있고, 항원-반응성 세포독성 T-림프구(CTL)를 자극할 수 있는 것일 수 있다.

- [135] "항원 프로세싱" 또는 "프로세싱"은 펩타이드, 폴리펩타이드 또는 단백질이 상기 펩타이드, 폴리펩타이드 또는 단백질이 분해되어(예컨대 폴리펩타이드가 펩타이드로 분해되는 것) 이들 단편들의 하나 이상이 (예컨대 결합을 통해) 세포, 예컨대 항원 제시세포에 의해 특이 T 세포에 제시되도록 MHC분자와 회합하는 것을 의미한다.
- [136] 용어 "T 세포" 및 "T 림프구"는 본 명세서에서 호환적으로 사용되며 여기에는 세포용해성 T 세포를 포함하는 세포독성 T 세포 (CTLs, CD8+ T 세포) 및 T 헬퍼 세포 (CD4+ T 세포)가 포함된다.
- [137] T 세포는 림프구로 알려진 백혈구 세포군에 속하며, 세포-매개된 면역에서 중추적인 역할을 한다. 이들은 T 세포 수용체(TCR)이라 칭해지는 특별한 수용체가 이들의 세포 표면에 존재한다는 측면에서 여타의 림프구 예컨대 B 세포 및 내츨 킬러 세포와 구별될 수 있다. 흉선은 T 세포의 성숙화에 책임이 있는 일차 기관이다. 몇몇 상이한 T 세포 서브세트들이 발견된 바 있으며 이들 각각은 독특한 기능을 갖는다.
- [138] T 헬퍼 세포는 특히 B 세포의 혈장 세포로의 성숙 및 세포독성 T 세포 및 대식 세포의 활성화와 같은, 면역학적 프로세스에서 다른 백혈구 세포를 도와주는 역할을 한다. 이 세포들은 CD4+ T 세포로도 알려져 있는데 이는 이들이 그 표면에서 CD4 단백질을 발현하기 때문이다. 헬퍼 T 세포는 항원 제시 세포(APC)의 표면에서 발현되는 MHC 클래스 II 분자에 의해 펩타이드 항원과 함께 제시되면 활성화된다. 일단 활성화되면, 이들은 급속히 분열하여 활성 면역 반응을 조절하거나 보조하는 사이토카인이라 칭해지는 작은 단백질을 분비한다.
- [139] 세포독성 T 세포는 바이러스 감염된 세포 및 종양 세포를 파괴하고 이식 거부에도 연루된다. 이 세포들은 또한 CD8+ T 세포라고도 알려져 있는데 이는 이들이 그 표면에서 CD8 당단백질을 발현하기 때문이다. 이 세포들은 신체의 거의 모든 세포 표면에 존재하는, MHC 클래스 I과 연관된 항원에 결합함으로써 그들의 표적을 인식한다.
- [140] 대다수의 T 세포들은 몇몇 단백질의 복합체로서 존재하는 T 세포 수용체 (TCR)을 갖는다. 실제 T 세포 수용체는 독립적인 T 세포 수용체 α 및 β (TCR α 및 TCR β) 유전자로부터 생성되어 α 및 β -TCR 사슬로서 칭해지는 2 개의 별개 펩타이드 사슬로 구성된다. $\gamma\delta$ T 세포 ($\gamma\delta$ T 세포)는 그의 표면에서 독특한 T 세포 수용체 (TCR)를 지니는 T 세포의 작은 서브세트를 나타낸다. 그러나, $\gamma\delta$ T 세포에서, TCR은 하나의 γ -사슬 및 하나의 δ -사슬로 구성된다. T 세포의 이 그룹은 $\alpha\beta$ T 세포에 비해 훨씬 드물다(총 T 세포의 2%).
- [141] T 세포 활성화의 첫 번째 시그널은 T 세포 수용체가 다른 세포 상의 MHC에 의해 제시된 짧은 펩타이드에 결합함으로써 제공된다. 이것은 그 펩타이드에 특이적인 TCR을 갖는 T 세포만이 활성화되도록 해준다. 파트너 세포는 대체로 항원 제시 세포 예컨대 프로페셔널 항원 제시 세포 (Antigen Presenting Cells)이며, 나이

브 반응의 경우 대체로 수지상 세포지만, B 세포 및 대식세포도 중요한 APCs일 수 있다.

- [142] 본 명세서에서, 어떤 분자가 소정의 표적에 대해 유의적인 친화도를 가지고 표준 분석에서 상기 소정의 표적에 결합하면, 그 분자는 그 소정의 표적에 결합할 수 있는 것이다. "친화도(Affinity)" 또는 "결합 친화도(binding affinity)"는 평형 해리 상수 (KD)에 의해 측정될 수 있다. 만일 어떤 분자가 상기 표적에 대해 유의적인 친화도를 갖지 않고 표준 분석에서 상기 표적에 유의적으로 결합하지 않으면 그 분자는 표적에 (실질적으로) 결합하지 못하는 것이다.
- [143] CD4+ 또는 CD8+ T 세포의 특이적인 활성화는 다양한 방식으로 확인할 수 있다. 특이적인 T 세포 활성화를 확인하는 방법들로는 T 세포의 증식, 사이토카인 (예컨대 림포카인)의 생성, 또는 세포용해 활성의 생성을 검색하는 것을 들 수 있다. CD4+ T 세포의 경우, 특이적 T 세포 활성화는 T 세포의 증식을 통하여 확인할 수 있다. CD8+ T 세포의 경우, 특이적 T 세포 활성화는 세포용해 활성에 의하여 확인할 수 있다.
- [144] 본 명세서에서, 용어 "점수/스코어(점수)"는 대개 수치로 표시된 어떤 테스트 또는 시험의 결과이다.
- [145] 본 발명에서, 소정의 펩타이드의 MHC 분자에 대한 결합 점수를 확인하는 것은 MHC 분자에 대한 펩타이드의 결합 가능성을 탐지하는 것을 포함한다.
- [146] 상기 MHC 분자에 대한 펩타이드의 결합 점수는 펩타이드:MHC 결합 예측 도구를 이용함으로써 확인 가능하다. 예를 들어, 면역 에피토프 데이터베이스 분석 리소스(IEDB-AR: <http://tools.iedb.org>)를 이용할 수 있다.
- [147] 본 발명에 따라 그의 면역원성을 탐지하고자 하는 아미노산 변형들은 세포 핵산의 돌연변이로부터 야기되는 것 일 수 있다. 이러한 돌연변이는 공지의 시퀀싱 (서열분석: sequencing) 기술에 의해 동정될 수도 있다.
- [148] 일 구체예에서, 암 특이 체세포 돌연변이 또는 서열 차이는 종양 시료의 엑솜, 예컨대, 전체 엑솜에서 탐지될 수 있다. 따라서, 본 발명은 하나 이상의 암 세포의 엑솜, 종기로는 전체 엑솜의 암 돌연변이 징후를 동정하는 것을 포함할 수 있다. 일 구체예에서, 암 환자의 종양 표본 내의 암 특이 체세포 돌연변이를 동정하는 단계는 엑솜 전반에 걸친(exome-wide) 암 돌연변이 프로파일을 동정하는 것을 포함한다.
- [149] 일 구체예에서, 암 특이 체세포 돌연변이 또는 서열 차이는 종양 시료의 트랜스크립톰, 예컨대 전체 트랜스크립톰에서 탐지될 수 있다. 따라서, 본 발명은 하나 이상의 암 세포의 트랜스크립톰, 예컨대 전체 트랜스크립톰의 암 돌연변이 징후를 동정하는 것을 포함할 수 있다. 일 구체예에서, 암 환자의 종양 시료 내의 암 특이 체세포 돌연변이를 동정하는 단계는 트랜스크립톰 전반에 걸친(transcriptome-wide) 암 돌연변이 프로파일을 동정하는 것을 포함한다.
- [150] 일 구체예에서, 암 특이 체세포 돌연변이를 동정하는 단계 또는 서열 차이를 동정하는 단계는 1개 이상, 예컨대, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,

18, 19, 20개 또는 그 이상의 암 세포의 단일 세포 시퀀싱을 포함한다. 따라서, 본 발명은 상기 하나 이상의 암 세포의 암 돌연변이 징후를 동정하는 것을 포함할 수 있다. 일 구체예에서, 암 세포는 순환 종양 세포이다. 순환 종양 세포와 같은 암 세포는 단일 세포 시퀀싱을 하기 전에 분리될 수 있다.

- [151] 일 구체예에서, 암 특이 체세포 돌연변이를 동정하는 단계 또는 서열 차이를 동정하는 단계는 차세대 시퀀싱 (next generation sequencing: NGS)을 이용하여 수행될 수 있다.
- [152] 일 구체예에서, 암 특이 체세포 돌연변이를 동정하는 단계 또는 서열 차이를 동정하는 단계는 종양 시료의 유전체 DNA 및/또는 RNA를 시퀀싱하는 것을 포함할 수 있다.
- [153] 암 특이 체세포 돌연변이 또는 서열 차이를 밝히기 위해 종양 시료로부터 얻은 서열 정보를 그 환자 또는 다른 개체로부터 획득될 수 있는 생식세포 등의 정상적인 비암 세포의 DNA 또는 RNA와 같은 핵산을 시퀀싱 하여 얻은 서열 정보 등의 레퍼런스와 비교하는 것이 바람직하다. 일 구체예에서, 정상적인 유전체 생식세포 DNA는 말초 혈액 단핵 세포 (PBMCs)로부터 획득될 수 있다.
- [154] "유전체(genome)"이라는 용어는 어떤 생명체 또는 세포의 염색체 내의 유전자 정보의 총량에 관계된 것이다.
- [155] 용어 "엑솜(exome)"은 발현된 유전자의 코딩 부분인, 엑손에 의해 형성된 어떤 생명체의 게놈 부분을 가리키는 것이다. 엑솜은 단백질 및 기타 기능성 유전자 산물의 합성에 사용되는 유전자 청사진을 제공한다. 이것은 유전체의 가장 기능적으로 관련있는 부분이므로, 어떤 생명체의 표현형에 가장 크게 기여할 가능성이 높다. 인간 게놈의 엑솜은 총 게놈의 1.5%를 차지하는 것으로 평가된다.
- [156] 용어 "트랜스크립톰"은 하나의 세포 또는 세포 집단에서 생산되는 mRNA, rRNA, tRNA, 및 기타 비-코딩 RNA를 비롯한 모든 RNA 분자의 세트에 관계된 것이다. 본 명세서에서, 트랜스크립톰은 하나의 세포, 세포 집단, 예컨대, 암 세포 집단 또는 특정 시점에서 주어진 개체의 모든 세포에서 생산되는 모든 RNA 분자의 세트를 의미할 수 있다.
- [157] 본 발명에서 "핵산"이라는 용어는 데옥시리보핵산 (DNA) 또는 리보핵산 (RNA)을 의미할 수 있다. 핵산은 본 발명에 따라 유전체 DNA, cDNA, mRNA, 재조합적으로 생산된 분자 및 화학 합성된 분자를 포함한다. 본 발명에서, 핵산은 단일 가닥 또는 이중 가닥 및 선형 또는 공유적으로 변환된 분자로서 존재할 수 있다.
- [158] "돌연변이"라는 용어는 레퍼런스에 비해서 핵산 서열에 변화가 있거나 차이가 있는 것 (뉴클레오타이드 치환, 부가 또는 결실) 및/또는 이러한 핵산 서열 변화에 의하여 코딩되는 펩타이드(단백질)의 아미노산 서열에 변화가 있는 것을 의미한다. "체세포 돌연변이(somatic mutation)"은 생식세포 (정자 및 난자) 이외의 모든 체 세포에서 일어날 수 있으므로 아들에게는 전달되지 않는다. 이러한 변형은 (항상 그런 것은 아니지만) 암 또는 기타 질환을 유발한다. 종기로는 돌연

변이는 비유사 돌연변이인 것이 바람직하다. "비유사 돌연변이(nonsynonymous mutation)"라는 용어는 번역 산물에서 아미노산 치환과 같은 아미노산 변화를 야기하는 돌연변이, 종기로는 뉴클레오타이드 치환을 의미한다.

- [159] 본 발명에서, "돌연변이"라는 용어는 점 돌연변이, 인델(Indels), 융합, 크로모트립시스(chromothripsis), 염기교정, 및/또는 RNA 에디트(edits)를 포함할 수 있다.
- [160] "암 돌연변이 징후(cancer mutation signature)"라는 용어는 비-암 레퍼런스 세포와 비교시 암 세포에 존재하는 돌연변이 세트를 칭한다.
- [161] 상기 "레퍼런스"는 중앙 시료부터 본 발명의 방법에서 획득된 결과를 연관 및 비교하는데 이용될 수 있다. 일반적으로 "레퍼런스"는 하나 이상의 정상적인 시료, 예컨대 동일 종의 개체들로부터 획득된, 암 질환에 걸리지 않은 시료에 기초하여 얻어질 수 있다.
- [162] 돌연변이 탐지를 위해 모든 적합한 시퀀싱 방법의 사용이 가능하며, 예컨대, 차세대 시퀀싱(Next Generation Sequencing: NGS)을 사용할 수 있다. 시퀀싱 단계의 속도를 높이기 위해 장차 NGS 기술을 3세대 시퀀싱 방법이 대체할 수 있을 것이다. 용어 "차세대 시퀀싱" 또는 "NGS"는, 전체 게놈을 작은 단편들이 되도록 분해함으로써 전체 게놈을 따라 평형하게 무작위적으로 핵산 템플 레이트를 판독하는 쉐거 화학(Sanger chemistry)으로 알려진 "종래의" 시퀀싱 방법론과 대조적으로, 모든 신규 한 고처리능 시퀀싱 기술을 의미한다. 이러한 NGS 기술(대규모 병렬형 시퀀싱 기술(massively parallel sequencing technologies)이라고도 알려짐)은 전체 게놈, 엑솜, 트랜스크립톰 (게놈의 모든 전사된 서열) 또는 메틸롬(게놈의 모든 메틸화된 서열)의 핵산 서열 정보를 매우 단기간 내, 예컨대 1-2주일, 1-7일 또는 24시간 이내에 전달할 수 있으며, 기본적으로 단일 세포 시퀀싱 접근법을 가능하게 한다. 상업적으로 수득 가능하거나 문헌에 언급된 다중 NGS 플랫폼을 사용할 수 있다.
- [163] "RNA"라는 용어는 적어도 하나의 리보뉴클레오타이드 잔기를 포함하고 종기로는 리보뉴클레오타이드 잔기들만으로 이루어지거나 실질적으로 이루어지는 분자에 관계된다. "리보뉴클레오타이드"는 β -D리보퓨라노실기의 2'-위치에서 히드록실기를 갖는 뉴클레오타이드에 관계된다. 용어 "RNA"는 이중-가닥 RNA, 단일-가닥 RNA, 분리된 RNA 예컨대 부분적으로 또는 완전히 정제된 RNA, 기본적으로 순수한 RNA, 합성 RNA 및 재 조합적으로 생성된 RNA 예컨대 하나 이상의 뉴클레오타이드의 부가, 결실, 치환 및/또는 변경에 의해 자연발생적인 RNA와 상이한 변형 RNA를 포함한다. 이러한 변경은 RNA의 말단(들) 또는 내부에, 예컨대 RNA의 하나 이상의 뉴클레오타이드에, 비뉴클레오타이드 물질을 부가하는 것을 포함할 수 있다. RNA 분자 내의 뉴클레오타이드 들은 또한 비-표준 뉴클레오타이드, 예컨대 자연발생적이지 않은 뉴클레오타이드 또는 화학합성된 뉴클레오타이드 또는 테옥시뉴클레오타이드를 포함할 수도 있다. 이들 변경된 RNA 들은 유사체 또는 자연발생적인 RNA의 유사 체라고도 칭할 수 있다.

- [164] 용어 "RNA"에는 "mRNA"이 포함될 수 있다. "mRNA"라는 용어는 DNA 주형을 이용하여 생성되고, 펩타이드 또는 폴리펩타이드를 코딩하는 "메신저-RNA"를 의미하고 "전사체"로 표현될 수 있다. 일반적으로, mRNA는 5'-UTR, 단백질 코딩 영역, 및 3'-UTR을 포함한다. mRNA는 세포와 시험관 내에서 오직 제한적인 반감기만을 갖는다. mRNA는 DNA 주형으로부터 시험관내 전사에 의해 생성될 수 있다. 시험관내 전사 방법론은 통상의 기술자에게 알려져 있다.
- [165] "감소하다" 또는 "억제하다" 등의 용어는 전체적인 감소, 예컨대, 5% 이상, 10% 이상, 20% 이상, 50% 이상, 또는 75% 이상 수준의 전체적인 감소를 일으키는 능력에 관계된다. 용어 "억제하다" 또는 이와 유사한 표현은 완전한 또는 기본적으로 완전한 억제, 즉 0으로의 감소 또는 기본적으로 0으로의 감소를 포함한다.
- [166] "증가하다", "증강시키다", "촉진하다", "또는 연장하다" 등의 용어는, 예컨대, 적어도 약 10%, 적어도 20%, 적어도 30%, 적어도 40%, 적어도 50%, 적어도 80%, 적어도 100%, 적어도 200%, 또는 적어도 300% 수준의 증가, 증강, 촉진 또는 연장에 관계된다. 이 용어들은 또한 0 또는 측정불가하거나 또는 검출불가능한 수준으로부터 0보다 큰 수준 또는 측정가능하거나 검출가능한 수준으로의 증가, 증강, 촉진 또는 연장에 관계된다.
- [167] 용어 "백신"은 투여될 경우 암 세포와 같은 질병 세포 또는 병원균을 인식 및 공격하는, 면역 반응 특히 세포성 면역 반응을 유도하는 의약 조제물(의약 조성물) 또는 생성물에 관계된다. 백신은 질환의 예방 또는 치료에 이용될 수 있다. "개인화화된 암 백신(personalized cancer vaccines) 또는 (individualized cancer vaccines)"이라는 용어는 특정한 암 환자에 관계되어 암 백신이 필요에 따라 개별적인 암 환자의 특수 환경에 알맞게 조정되는 것에 관계된다.
- [168] 일 구체예에서, 본 발명에 따라 제공된 백신은 본 발명의 방법에 따라 면역원성 펩타이드로 선정된 하나 이상의 펩타이드 또는 상기 펩타이드를 코딩하는 핵산(예컨대, RNA)를 포함할 수 있다.
- [169] 본 발명에 따라 제공되는 항암 백신은 환자에게 투여시 환자의 종양에 특이적인 T 세포를 자극, 프라이밍 및/또는 확장하는데 적합한 하나 이상의 T 세포 에피토프를 제공한다. T 세포는 그 T 세포 에피토프가 유래된 항원을 발현하는 세포에 지향된 것일 수 있다. 따라서, 본 발명에서 설명된 백신들은 클래스 II MHC에 의해 하나 이상의 종양-관련 신생항원의 제시를 특징으로 하는 암 질환에 대해 세포 반응, 예컨대, 세포독성 T 세포 활성을 유도 또는 촉진할 수 있는 것일 수 있다. 본 발명에서 제공되는 백신은 암 특이 돌연 변이를 표적으로 할 것이므로, 당해 환자의 종양에 특이적일 수 있다.
- [170] 본 발명에서 제공되는 백신은 환자에게 투여될 경우 본 발명의 방법에 의해 면역원성으로 펩타이드로 선정된 펩타이드와 연관된 T 세포 에피토프를 하나 이상, 예컨대 2 이상, 5 이상, 10 이상, 15 이상, 20 이상, 25 이상, 30 이상 및 60 이하, 55 이하, 50 이하, 45 이하, 40 이하, 35 이하 또는 30개 이하 제공하는 백신에 관한 것이다. 이러한 T 세포 에피토프는 본 발명에서 "네오-에피토프"라고도 칭해진

다. 환자의 세포, 특히 항원 제시 세포에 의한 이들 에피토프의 제시는 종기로는 MHC와 결합시 그 에피토프를 표적화 하는 T 세포를 초래하며 따라서, 환자의 종양, 종기로는 원발성 종양뿐 아니라 종양 전이도 표적화하고, 그 T 세포 에피토프가 유래된 항원을 발현하며 당해 종양 세포 표면에서 동일한 에피토프를 제시한다.

- [171] 용어 "종양" 또는 "종양 질환"은 종기로는 팽윤 또는 병변을 형성하는 세포의 비정상적인 성장 (신생 세포, 종양성 세포 또는 종양 세포라 칭해짐)을 의미한다. "종양 세포"라 함은 급속하게 제어되지 않는 세포 증식에 의해 성장하고 그 새로운 성장을 개시시킨 자극이 멈춘 후에도 지속적으로 성장하는 비정상적인 세포를 의미한다. 종양은 정상 조직과의 기능적 연계 및 구조적 조직화가 부분적으로 또는 완전히 결여되어 있으며 대체로 양성, 전악성 또는 악성일 수 있는 뚜렷한 조직 덩어리를 형성한다.
- [172] 암은 일군의 세포들이 조절되지 않은 성장(정상 한계를 넘어서 분열), 침습 (인접 조직으로의 침투 및 파괴) 및 때때로 전이(림프 또는 혈액을 통해 체내 다른 부위로 확산)를 나타내는 질환 부류이다. 암의 이러한 3 가지 악성적인 특성들은, 자가-제한적이고, 침습 또는 전이하지 않는 양성 종양으로부터 이들은 구별 짓는다. 대부분의 암은 종양을 형성하지만, 백혈병과 같은 몇몇 암은 그렇지 않다. 악성, 악성 신생 물 및 악성 종양은 기본적으로 암과 동의어이다.
- [173] 본 발명의 목적 상, "암" 및 "암 질환"이라는 용어는 "종양" 및 "종양 질환"이라는 용어와 호환적으로 사용될 수 있다.
- [174] 상기 "암"은 암종, 선암종, 모세포종, 백혈병, 정상피종, 흑색종, 기형암종, 림프종, 신경모 세포종, 신경아교종, 직장암, 자궁내막암, 신장암, 부신암, 갑상선암, 혈액암, 피부암, 뇌의 암, 자궁경부암, 장암, 간암, 결장암, 위암, 장암, 두경부암, 위장관암, 림프절암, 식도암, 대장암, 췌장암, 귀, 코 및 인후 (ENT)암, 유방암, 전립선암, 자궁암, 난소암, 폐암 등일 수 있다. 상기 암은 원발성 암 또는 전이암일 수 있다.
- [175] "치료"는 대상에서 종양의 크기 또는 종양의 수를 감소시키는 것; 대상의 질환 진행을 멈추거나 둔화시키는 것; 대상에서 새로운 질병의 발달을 억제하거나 둔화시키는 것; 현재 질병에 걸려 있거나 이전에 질병에 걸렸었던 대상에서 증상 및/또는 재발의 빈도나 위중도를 감소시키는 것; 및/또는 대상의 수명을 연장 등을 의미할 수 있다.
- [176] "면역요법(immunotherapy)"은 특이적인 면역 반응의 활성화가 관여하는 치료 방법에 관계된다. 본 발명의 관점에서, "보호", "방지", "예방", "방지하는", "보호적인" 등의 용어는 대상자에 있어서 어떤 질병의 발 병 및/또는 전파의 예방 또는 치료 또는 양자 모두에 관한 표현으로서, 특히, 대상자에 있어서 질병의 진행 기회를 최소화하거나 질병의 진행을 지연시키는 것과 관련된 표현이다. 예를 들어, 암에 걸릴 위험이 있는 개체는 암 예방 요법의 후보가 된다.

- [177] 면역요법의 예방적 투여, 예컨대 본 발명의 백신의 예방적 투여는 종기로는 수혜자(recipient)를 질환의 발달로부터 보호한다. 면역요법의 치료적 투여, 예컨대 본 발명의 백신의 치료적 투여는 그 질환의 진행/성장의 억제로 이어질 수 있다. 이것은 그 질환의 진행/성장의 감속, 특히 그 질환의 진행의 파괴를 포함하며, 예컨대, 그 질환의 제거로 이어진다.
- [178] 면역요법은 본 발명에서 제공되는 제제가 환자로부터의 병에 걸린 세포를 제거하는 기능을 하는, 다양한 기술을 이용하여 수행될 수 있다. 이러한 제거는 항원 또는 항원을 발현하는 세포에 특이적인 면역 반응을 환자에서 유도하거나 증강시킨 결과로서 발생하는 것일 수 있다.
- [179] 본 발명에서 제공되는 제제 및 조성물은 단독으로 사용되거나 또는 외과수술, 방사능 요법, 화학요법 및/또는 골수이식(자가, 동계, 동종이계 또는 무관련)과 같은 치료법과 조합적으로 사용될 수도 있다.
- [180] 용어 "면역화(immunization)" 또는 "백신접종(vaccination)"는 치료 또는 예방학적 이유로 면역반응을 유도할 목적으로 환자를 치료하는 프로세스를 설명한다.
- [181] "생체내(in vivo)"라는 용어는 대상자 체내의 상황과 관련된다.
- [182] "대상", "개체", 또는 "환자" 등의 용어는 호환적으로 사용되며, 척추동물, 예컨대, 포유동물에 관계된다. 예를 들어, 본 발명의 문맥 상, 포유동물은 인간, 비인간 영장류, 가축화된 동물, 예컨대, 개, 고양이, 양, 소, 염소, 돼지, 말 등, 실험실 동물 예컨대 마우스, 래트, 토끼, 기니 피그 등 뿐만 아니라 동물원에 포획된 동물을 포괄한다. 본 발명에서 "동물"이라는 용어에는 인간도 포함될 수 있다. "대상"이라는 용어는 환자, 즉 동물, 종기로는 질환, 종기로는 본 발명에 설명된 질환에 걸린 인간, 또는 이로부터 분리된 생체 시료(암세포, 혈액 등)를 포함한다.
- [183] 본 명세서에서 제공되는 백신 조성물은 아쥘반트를 추가로 포함할 수 있다. 용어 "아쥘반트(adjuvant)"는 면역 반응을 연장, 증강 또는 가속화하는 화합물에 관한 것이다. 본 발명의 조성물은 종기로는 아쥘반트 첨가 없이 그의 효과를 발휘한다. 그러나 본 발명의 조성물은 공지된 아쥘반트를 함유할 수도 있다. 아쥘반트는 오일 에멀전(예컨대, Freund's 아쥘반트), 미네랄 화합물(예컨대 명반), 박테리아 생성물(예컨대 Bordetella pertussis 독소), 리포좀 및 면역-자극 복합체와 같은 화합물들의 이종 그룹을 포함한다. 아쥘반트의 예로는 모노포스포릴-지질-A (MPL SmithKline Beecham)을 들 수 있다. 사포닌 예컨대 QS21 (SmithKline Beecham), DQS21 (SmithKline Beecham; WO 96/33739), QS7, QS17, QS18, 및 QS-L1, 불완전 Freund's 아쥘반트, 완전 Freund's 아쥘반트, 비타민 E, 몬타니드, 명반, CpG 올리고뉴클레오타이드, 및 스쿠알렌 및/또는 토크페롤과 같은 생물학적으로 분해가능한 오일로부터 제조되는 다양한 유중수형 에멀전을 들 수 있다.
- [184] 상기 백신 조성물은 그 외 환자의 면역 반응을 자극하는 약학적으로 허용 가능한 물질을 추가로 포함할 수 있다. 예컨대, 상기 물질은 사이토카인일 수 있다. 이러한 사이토카인의 예로는, 백신의 보호 활성을 증가시키는 인터류킨-12 (IL-12), GM-CSF, IL-18 등을 들 수 있다.

- [185] 상기 백신 조성물은 면역 반응을 증강시키는 화합물을 추가로 포함할 수 있다. 상기 화합물은 B7-1 및 B7-2 (각각 CD80 및 CD86)와 같은 핵산 또는 단백질 형태로 제공되는 공-자극 분자를 포함할 수 있다.
- [186] 본 명세서에서 제공되는 백신 조성물은 주사 또는 주입을 비롯한 통상적인 경로로 투여될 수 있다. 투여는 예컨대 경구, 정맥내, 복강내, 근육내, 피하 또는 경피 경로로 수행될 수 있다. 일 구체예에서, 투여는 림프절 내로 주사하는 것과 같은 결절내(intranodally)로 수행 가능하다.
- [187] 상기 백신 조성물은 약학적 유효량으로 투여될 수 있다. "약학적 유효량"이라는 용어는 그 단독으로 또는 추가 투여량(doses)과 함께 소망되는 반응 또는 효과를 낼 수 있는 양을 의미한다. 상기 약학적 유효량은 치료하고자 하는 상태, 질병의 위중도, 환자의 개별적인 변수, 예컨대 연령, 생리적 상태, 크기 및 체중, 치료 기간, 부수되는 치료법의 종류(있을 경우), 특정 투여 경로 및 유사 인자들에 따라 달라질 수 있다.
- [188] 상기 백신 조성물은 멸균된 것일 수 있고, 소망되는 반응 또는 소망되는 효과를 생성하기 위해 치료적 활성 물질의 유효량을 함유할 수 있다.
- [189] "약학적으로 허용가능한(pharmaceutically acceptable)"이라는 용어는 의약 조성물의 활성 성분의 작용과 상호반응 하지 않는 비독성 물질을 의미하는 것으로, 예컨대, 염, 완충 물질, 보존제, 담체, 아유반트와 같은 보조 면역-증강물질, 예컨대 CpG 올리고뉴클레오타이드, 사이토카인, 케모카인, 사포닌, GM-CSF 및/또는 RNA를 함유할 수 있으며, 적절한 경우 기타 치료 활성 화합물도 함유할 수 있다.
- [190] "담체"는 활성성분의 적용을 용이화, 증강 또는 가능하게 하기 위해, 활성성분과 조합되는, 천연 또는 합성된 유기 또는 무기 성분을 가리킨다. 본 발명에서 "약학적으로 허용 가능한 담체"라는 용어는 환자에게 투여하기에 적합한 1종 이상의 적합한 고상 또는 액상 충전제, 희석제 또는 캡슐화 물질을 포함한다.
- [191] 상기 백신 조성물은 아세트산 염, 시트르산 염, 붕산염 인산 염과 같은 적절한 완충제를 함유할 수 있다.
- [192] 상기 백신 조성물은 벤잘코늄 클로라이드, 클로로부탄올, 파라벤 및 티메로살과 같은 적절한 보존제를 함유할 수 있다.
- [193] 상기 백신 조성물은 대개 단위 투여 형태로 제공되며 그 자체로 공지인 방법으로 제조될 수 있다. 본 발명의 백신 조성물은 주사제, 캡슐, 정제, 로젠지, 용액, 현탁액, 시럽, 엘릭시르 또는 예컨대 에멀전의 형태일 수 있다.
- [194] 비경구 투여에 적합한 조성물은 대체로 수혜자의 혈액과 등장성인, 활성 화합물의 멸균 수성 또는 비수성 제제를 포함할 수 있다. 적합한 담체 및 용매의 예로는 링거 용액 및 등장성 염화나트륨 용액을 들 수 있다. 이에 더해, 대체로 멸균된 고정 오일을 용액 또는 현탁액 매질로서 이용할 수 있다.
- [195]

발명의 효과

- [196] 상기 발명의 특징이 적용된 경우 환자에 대한 NGS 데이터 프로세싱을 통해 효율적으로 HLA-type I과 Mutation Variant에 대한 Processing이 가능하다
- [197] 본 발명에서는 AI 기술을 이용한 database 학습을 진행하여 Binding Affinity 모델과 Immunogenicity 모델을 통한 신항원 후보군을 선정하고 Neoantigen prioritization 과정을 새롭게 적용하고 검증하여 개인 맞춤형 항암백신에 적용할 수 있는 신항원의 선정을 진행할 수 있는 방법을 제시한다.

[198]

도면의 간단한 설명

- [199] 도 1은 결합점수(BA) 및 면역원성 점수(IMM)의 분포를 보여준다.
- [200] 도 2a는 실시예 4에 따른 면역원성 점수 측정 과정을 모식적으로 보여준다.
- [201] 도 2b 내지 2d는 실시예 4에서 HLA-I 타입별 INFg ELISPOT을 통한 면역원성 확인 결과를 보여준다 (2b: A*02:01, 2c: A*24:02, 2d: A*11:01).
- [202] 도 3은 실시예 5.1의 시험과정을 모식적으로 보여준다.
- [203] 도 4는 실시예 5.1에서 얻어진 splenocyte INFg ELISPOT 결과를 보여주는 그래프이다.
- [204] 도 5는 실시예 5.2의 시험과정을 모식적으로 보여준다.
- [205] 도 6는 실시예 5.2에서 얻어진 splenocyte INFg ELISPOT 결과를 보여주는 그래프이다.
- [206] 도 7은 실시예 6에서 확인된 신항원의 항암효과 (*in vivo*)를 보여주는 그래프이다.

[207]

발명의 실시를 위한 형태

- [208] 이하 본 발명을 다음의 실시예에 의하여 보다 구체적으로 설명하고자 한다. 그러나 이들은 본 발명을 예시하기 위한 것일 뿐이며, 본 발명의 범위가 이들 실시예에 의하여 제한되는 것은 아니다.

[209]

- [210] 실시예 1. 신항원 예측모델의 구축(모델 학습) (신항원 예측모델 I)

- [211] IEDB(Immune Epitope Database and Analysis Resource)(<https://www.iedb.org>)에 있는 HLA(Human Leukocyte Antigen) type I 중 HLA I-A, B, C와 펩타이드의 아미노산 서열 정보를 이용한 학습을 통하여 신항원 후보 펩타이드의 결합력(Binding Affinity) 및 면역원성(Immunogenicity)을 예측하는 인공지능 모델을 준비하였다. 구체적으로, 상기 결합력 및 면역원성 예측 모델은 각각 서로 다른 모델 구조, 학습 데이터 분할, 전이 학습 포함 여부, 학습 최적화 방법, 학습 중단 시점을 조합하여 훈련한 서로 다른 모델 150개로 구성된 앙상블 (ensemble) 모델이다. 상기 모델 구조는 GRU(Gated Recurrent Unit)와 Transformer 중 1개를 선택하였다. 상기 학습 데이터 분할은 5-Fold Cross Validation을 적용하였다. 전이 학습은 결합력 예측 모델의 경우 정성 데이터로 학습한 분류기(Classifier)를 사용하였으며,

면역원성 예측 모델의 경우 결합력 예측 모델 중 Validation 데이터의 예측 오차가 가장 작은 것을 활용하였다. 학습 최적화 방법은 AdamW와 SWA(Stochastic Weights Averaging) 중 1개를 선택하였다. 학습 중단 시점은 전체 학습 과정에서 Validation 데이터의 예측 오차가 가장 낮은 순서로 5개의 시점을 선정하였다.

[212] 이하, 신항원 예측모델의 준비 과정을 보다 상세히 설명한다:

[213]

[214] **1.1. 신항원 예측모델 Input 데이터 선정**

[215] IEDB 데이터 중 HLA type I 학습에 필요한 데이터들을 선별하고 신항원 예측 모델 학습에 이용할 수 있도록 변환하여, Binding Affinity 모델(결합력 예측모델)과 Immunogenicity 모델(면역원성 예측모델)에 각각 적용하였다. 예컨대, 상기 데이터 선별은 HLA-A, HLA-B, HLA-C에 해당하지 않는 데이터를 삭제하는 작업, HLA분자에 돌연변이가 있는 데이터를 삭제하는 작업, 펩타이드의 길이가 7 이하 또는 15 이상인 것을 삭제하는 작업, 아미노산 서열에 인간 유래 아미노산 20종 외의 서열이 포함된 것을 제거하는 작업 등에서 1가지 이상의 작업을 적용하는 것일 수 있다. 예컨대, 상기 데이터 변환은 데이터 수치의 범위가 큰 결합력 측정값(예컨대, IC50)을 0~1 구간으로 정규화하는 작업, 정량적 측정값 대신 [실험 횟수, 반응 횟수]로 제공되는 면역원성 점수 데이터를 베타 분포의 평균값으로 변환하는 작업(L. Guangyuan, et al. "DeepImmuno: deep learning-empowered prediction and generation of immunogenic peptides for T-cell immunity." Briefings in bioinformatics 22.6 (2021)) 등에서 1가지 이상의 작업을 적용하는 것일 수 있다.

[216] 면역원성 예측모델 학습시에는, 상기 IEDB 데이터의 다양한 실험 결과들 중에서 IFN γ , IL-2, TNF α , Granzyme A 등과 같은 사이토카인 분비량 등의 면역원성에 관련된 시험 방식(IEDB 중에서 각 펩타이드의 PBMC 처리시의 INF-g, TNF-a, IL-2, IL-1b 등을 포함하는 T세포 유래 사이토카인, 면역원성 활성화로 인한 T세포 proliferation, CCL4, CXCL9을 포함하는 Chemokine, Granzyme B 등 활성화된 T세포의 cytotoxicity에 해당하는 데이터)을 선별하여 학습 진행하였다.

[217]

[218] **1.2. 신항원 예측모델의 학습 방식**

[219] 상기 실시예 1.1에서 선정된 데이터를 이용하여 아래를 수행하였다:

[220] 결합력 (Binding Affinity) 예측모델에 대하여, 2가지 모델 구조 (GRU (J. Chung et al, <https://arxiv.org/abs/1412.3555>, 2014), transformer (Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).)) 중 1개를 선택하고, 3가지 학습 방법 중 1개를 선택하며, 5-fold Cross Validation의 데이터 분할 방법 중 1개를 선택하고, 학습 중 Validation 데이터에 대한 예측오차가 가장 낮은 5개의 학습 시점에서 모델을 저장하는 방식으로 총 150개의 모델 학습을 진행하였다 ($2 \times 3 \times 5 \times 5 = 150$). 학습한 150개 모델의 예측값의 평균과 분산을 계산하여 최종 결합력을 예측하였다 (예측값 1).

- [221] 면역원성(Immunogenicity) 예측모델은 2가지 모델 구조(GRU, transformer) 중 1개를 선택하고, 3가지 학습 방법 (결합력 예측모델 중 Validation 데이터 예측 오차가 가장 낮은 모델의 학습 포함) 중 1개를 선택하며, 5-fold Cross Validation의 데이터 분할 방법 중 1개를 선택하고, 학습 중 Validation 데이터에 대한 예측오차가 가장 낮은 5개의 학습 시점에서 모델을 저장하는 방식으로 총 150개의 모델 학습을 진행하였다 ($2 \times 3 \times 5 \times 5 = 150$). 학습한 150개 모델의 예측값의 평균과 분산을 계산하여 최종 면역원성을 예측하였다 (예측값 2).
- [222] 상기 예측값 1과 예측값 2의 계산에서는 각각 [(150개 모델 예측값의 평균)-(150개 모델 예측값의 분산)]을 최종 예측값으로 사용하였다. 이는 다른 데이터와 학습 방법을 적용한 150개 모델의 예측값 간의 편차가 작을 수록 더 안정적인 예측이라고 판단할 수 있기 때문이다.
- [223]
- [224] 실시예 2. 신항원 예측 및 선정 단계
- [225] 상기 실시예 1에서 구축된 신항원 예측모델로부터 신항원을 선정하였으며, 상기 신항원 선정은 아래와 같은 단계로 진행되었다:
- [226] 1) 공개 데이터베이스로부터 암 환자의 NGS 데이터 수집;
- [227] 2) 상기 암 환자의 NGS 데이터를 이용하여 HLA 타입과 암세포 유래 펩타이드 판별 및 TPM 계산;
- [228] 3) 상기 실시예 1의 신항원 예측 모델을 이용한 예측 점수 계산;
- [229] 4) 하기 실시예 3의 수식을 이용한 신항원 우선순위 계산 및 실험 대상 펩타이드 선정;
- [230] 5) 상기 2)에서 판별한 HLA타입과 동일한 PBMC를 이용한 결합력 및 면역원성 실험.
- [231] 신항원 예측모델 검증에 위해 사용한 공개 데이터는 pembrolizumab 치료를 받아 complete response를 보인 Melanoma 환자 4명 (HLA type: A*02:01; “John Wiley & Sons A/S, Tissue Antigens 73, 95-170” 참조)의 이터(이하, “환자 Sample data” 또는 “환자 Sample dataset”)를 활용하여 검증을 진행하였다. [Hugo, Willy, et al. "Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma." Cell 165.1 (2016): 35-44.]
- [232]
- [233] 2.1. 신항원 선정을 위한 데이터 준비 (환자 유래 샘플의 Whole exome sequencing, RNA-seq 데이터 처리)
- [234] 상기 환자 Sample dataset에 대하여 Data pre-processing, Variant Discovery, Callset refinement를 위하여 각각의 코드를 조합하여 Mutation 서열 및 HLA-type의 정보 등을 도출할 수 있도록 설정하였다. 보다 구체적으로 설명하면,
- [235] 1) DNA 변이 정보를 Peptide 변이정보로 변환하며, RNA abundance data, HLA type data를 concatenation하였다. 또한, Uniprot의 SwissProt에서 Homo sapience의

알려진 단백질 서열에서 변환하여 얻어진 peptide가 발견되는지 확인하는 과정을 수행하였다.

- [236] 2) 암 특이적 유전적 변이를 찾는 프로그램인 GATK 패키지에 포함된 broad institute의 Mutect2와 illumina사가 개발한 Strelka에서 결과를 얻었다. 그리고 DNA 변이 정보를 가공하여 단백질 수준의 missense 변이와 in-frame deletion 데이터를 각각 툴에서 얻은 정보를 정리한 후, Mutect2와 Strelka에서 중복으로 얻어진 peptide를 단일화하였다 (Koboldt, D. C. et al., VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research* 22, 568-576 (2012); Cai, L., Yuan, W., Zhang, Z., He, L. & Chou, K. C., In-depth comparison of somatic point mutation callers based on different tumor next-generation sequencing depth data. *Scientific Reports* 6, 9, <https://doi.org/10.1038/srep36540> (2016); Kim, S. et al., Strelka2: fast and accurate calling of germline and somatic variants. *Nature Methods* 15, 591-594, <https://doi.org/10.1038/s41592-018-0051-x> (2018); McKenna, A. et al., The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297-1303, <https://doi.org/10.1101/gr.107524.110> (2010); Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., & Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology*, 31(3), 213-219. doi:10.1038/nbt.2514).
- [237] 3) Peptide에 대한 hydrophobicity를 계산하여 추후 신항원 생산에 사용할 데이터를 갖추었다. 신항원 예측 및 개인 맞춤형 항암백신이 mRNA로 제작될 경우에는 특별히 Hydrophobicity에 대한 고려가 필요하지 않을 수 있지만, 펩타이드로 항암백신을 만들거나 면역원성 시험의 진행 시에는, 펩타이드의 Hydrophobicity가 높으면 (예컨대, hydrophobic 아미노산 함량이 높음) 생산이 불가능하기 때문에, 펩타이드의 Hydrophobicity 계산이 필요할 수 있다. 본 실시예에서, 상기 Hydrophobicity는 Gravy (Grand average of hydropathy) score를 계산하여 판단하였다.
- [238]
- [239] **2.2. 신항원 예측모델 (결합력 및 면역원성)을 적용한 신항원의 선정 (개인 맞춤형 항암백신 적용가능성 확인)**
- [240] 상기 실시예 2.1에서 준비된 데이터를 사용하여, 결합력 및 면역원성 Threshold를 설정하여 상기 신항원 예측모델을 적용하여 선정된 신항원에 대한 검증을 수행하였다:
- [241] - 결합력(Binding affinity; BA) 시험은 IC50 값이 알려진 NY-ESO-1 (New York esophageal squamous cell carcinoma 1 peptide (157-165, SLLMWITQC; 애니젠社에 의뢰하여 합성) 대비 기준으로 결합력이 좋은 지 비교하여 예측 신항원의 결합력을 검증
- [242] - IFN γ ELISPOT Immunogenicity 시험법을 적용하여 면역원성 시험 진행

- [243] - HLA-A*02:01 타입의 Normal PBMC를 24시간 동안 Resting 시킨 후 후보군 펩타이드를 50 μ M, IL-2 (10IU/ml) 로 처리하여 Pulsing을 진행 (72시간, 37°C, 5% CO₂). 이후 PBMC를 wash하고 ELISPOT Assay plate로 transfer 하여 같은 펩타이드로 re-stimulation (50 μ M, 24시간, 37°C, 5% CO₂). Immunospot 사의 ELISPOT protocol에 따라 ELISPOT assay 진행. 다른 Donor 3종 PBMC로 재시험 진행 (재현성 및 통계적 유의성 확인)
- [244]
- [245] 상기 HLA-A*02:01은 alpha chain의 일부분의 서열을 사용하였으며, 사용한 서열은 다음과 같다:
- [246] SHSMRYFFTSVSRPGRGEPRIA VGYVDDTQFVRFDSDAA
SQRMEPRAPWIEQEGPEYWDGETRKVKAHSQTHRVDLGTL
RGYYNQSEAGSHTVQRMYGCDVGS DWRFLRGYHQYAYDGK
DYIALKEDLRSWTAADMAAQTTKHKWEAAHVAEQLRAYLE
GTCVEWLRRLRYLENGKETLQRT (서열번호 1)
- [247]
- [248] Cell을 사용한 시험이라서 면역원성 시험의 편차가 존재하여 Min-Max Normalization 및 Weight Value (1 Lot 검출: x1, 2 Lot 검출: x2, 3 Lot 검출: x3)을 적용하여 신항원 Rank 하여 개인 맞춤형 항암 백신 적용 신항원 선정 방법을 확립하였다.
- [249]
- [250] 상기 실시예 2.1에서 환자유래 샘플(pembrolizumab 치료를 받아 complete response를 보인 Melanoma 환자 4명 (HLA type: A*02:01)에서 확인된 118057 개 mutant 펩타이드 정보를 상기 실시예 1의 결합력(BA) 예측모델 및 면역원성(IMM) 예측모델에 적용하여 얻어진 결과를 도 1에 나타냈다.
- [251] 또한, 도 1의 결과 중, BA high group (G1; $0.15 \leq \text{IMM Score} < 0.25$ 및 $0.6 \leq \text{BA Score}$) 및 IMM high group (G2: IMM Score ≥ 0.25 및 $0.6 > \text{BA Score} \geq 0.1$)에 대하여 앞서 기재된 방법으로 실제 결합력과 면역원성을 측정하여 표 1에 나타내었다:
- [252] [표1]

HLA-I	Binding Affinity (BA)	Immunogenicity (IMM)	Neoantigen Selection Test
G1	100%	5.6%	Melanoma 환자 4명 (CR) with Pembrolizumab treatment
G2	49%	23.6%	

- [253] 상기 표 1의 결과에서 보여지는 바와 같이,
- [254] $0.15 \leq \text{IMM Score} < 0.25$ 및 $0.6 \leq \text{BA Score}$ 로 Threshold 설정 시 (G1), 결합력 Positive한 신항원 예측율이 100%임과,

- [255] IMM Score ≥ 0.25 및 $0.6 > \text{BA Score} \geq 0.1$ 로 Threshold 설정 시 (G2), 면역원성 Positive한 신항원 예측율이 23.6%임을 확인하였다.
- [256]
- [257] 실시예 2.3. 추가 변수로서 **RNA expression level(TPM)**의 측정
- [258] 상기에서 예측된 신항원의 평가를 위한 변수로서, TPM (RNA expression level; transcripts per million) 변수를 추가로 측정하였다. TPM은 신항원 후보군의 RNA 서열의 발현율을 의미하는 것으로, 환자 암세포의 돌연변이가 실제 펩타이드로 생성될 가능성을 의미할 수 있다. 그러나 IEDB에서는 상기 TPM 데이터를 제공하지 않기 때문에 모델 학습이 아닌 모델 예측 결과의 평가(prioritization)에 사용할 수 있다. 상기 TPM을 활용한 평가 기준을 개발하기 위하여, 암 환자의 공개 NGS 데이터(TESLA 데이터셋, Wells, Daniel K., et al. "Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction." Cell 183.3 (2020): 818-834.)를 수집하여 암 유래 신항원 후보군(펩타이드 후보)을 선별하였으며, 상기 예측 모델을 사용하여 상기 신항원 후보군의 상기 예측값 1, 2를 예측하였다(실시예 1.2 참조). 상기 암 환자의 공개 NGS 데이터에 포함된 RNA-seq 데이터를 처리하여 상기 신항원 후보군의 RNA 서열의 발현율(TPM)을 계산하였다.
- [259]
- [260] 실시예 3. 선정된 신항원의 평가 (Neoantigen Prioritization)
- [261] 앞선 실시예 1.2에서 얻어진 상기 예측값 1, 2에 대하여 다변량 선형회귀분석 (Multivariate Linear Regression)을 통하여 결합력 예측모델(Binding Affinity Model)과 면역원성 예측모델(Immunogenicity Model)에서 예측모델 변수의 최적화를 진행하고, 최종적으로 TPM (RNA expression level), Binding affinity prediction score(BA; 상기 실시예 3의 예측값 1), Immunogenicity prediction score(IMM; 상기 실시예 3의 예측값 2)를 변수로 확정하고 신항원의 Ranking을 선정하였다.
- [262] 상기 TESLA 데이터셋은 환자 8명의 HLA 타입과 NGS 데이터, 상기 환자들에게서 유래한 신항원 후보 845개의 펩타이드 서열과 면역원성 실험 결과를 포함하고 있으며, 상기 845개 후보 중 38개의 펩타이드에서 면역원성이 있는 것으로 검증되었다. 신항원 예측 기준의 우수성을 비교하기 위하여, 상기 신항원 후보군에 대하여 TPM을 적용하고 Binding affinity prediction score(BA)와 Immunogenicity prediction score(IMM)의 가중치를 다양하게 변화시키면서 상기 환자 1명당 20개의 신항원 후보를 선정하고 (총 160개), 여기에 면역원성이 있는 상기 38개의 신항원 후보가 얼마나 많이 포함되었는가를 아래의 표 2에 기재하였다 (Threshold: TPM (transcripts per million) > 1):

[표2]

Weight	W1: 1.0	W1: 0.8	W1: 0.5	W1: 0.2	W1: 0.0
--------	---------	---------	---------	---------	---------

Distribution	W2: 0.0	W2: 0.2	W2: 0.5	W2: 0.8	W2: 1.0
TESLA	14	18	26	23	23

[264] (상기 표 2에서, W1은 BA의 가중치, W2는 IMM의 가중치를 각각 의미함)
 상기 표 2에서 확인되는 바와 같이, Binding affinity prediction score(BA)와
 Immunogenicity prediction score(IMM)의 가중치를 동일(1:1)하게 한 경우에 가장
 많은 면역원성이 있는 신항원을 찾아낼 수 있었다.

[265] 상기 결과를 기초로, 신항원 순위선정을 위한 수식을 아래와 같이 정하였다:

[266] 신항원의 prioritization scoring scheme: $[\min(\text{TPM}, 1) * (\text{BA} + \text{IMM})/2]$

[267]

[268] 상기한 Prioritization 적용 전후의 검출되는 신항원 개수를 아래의 표 3에 나타
 내었다:

[269] [표3]

	Patient	Positive	Prioritization 적용 전	Prioritization 적용 후
Pilot Round	#4	1	1	1
	#8	1	0	1
	#9	2	1	2
	[310]	4	2	4
Main Round	#1	9	3	7
	#2	4	3	2
	#3	13	2	9
	#12	4	2	2
	#16	4	1	2
	[535]	34	11	22

[270] 표 3에 나타난 바와 같이, Prioritization 적용에 의하여 선별가능한 신항원 개수
 가 증가하여 신항원 예측률이 증진됨을 확인할 수 있다.

[271]

[272]

[273] 실시예 4. 신항원 예측모델 II 및 신항원 선정

[274] 실시예 1 내지 3의 방법을 참조하여, Melanoma 환자 중 3종 HLA type (A*02:01,
 A*24:02, A*11:01)을 보유하고 있는 환자를 선별하여 신항원 예측모델을 통하
 여 신항원 20종을 예측하였고, 예측한 신항원에 대한 면역원성 여부를 IFNg
 ELISPOT을 통하여 확인하였다.

[275]

[276] 보다 구체적으로, 상기 환자군 데이터 선별 기준은 다음과 같다:

- [277] - High Mutation Burden Tumor → Melanoma;
- [278] - Complete WES (whole exome sequencing; Mutation selection), RNA-seq (Expression level) Dataset (펩타이드 서열로 변환하여 사용함);
- [279] - ICI(Immune Checkpoint Inhibitor) treatment (anti-PD-1 therapy: Pembrolizumab, Nivolumab) → CR(Complete Response), PR(Partial Response), PD(Progressive Disease);
- [280] - HLA type: A*02:01, A*24:02, A*11:01 ([HLA-A*02:01]:
 SHSMRYFFTSVSRPGRGEPRFIAVGYVDDTQFVRFDSDAASQRMEP
 RAPWIEQEGPEYWDGETRKYKAHSQTHRVDLGLTRGYYNQ
 SEAGSHTVQRMYGCDVGSDWRFLRGYHQYAYDGKDYLK
 EDLSWTAADMAAQTTKHKWEAAHVAEQLRAYLEGTCVEW
 LRRYLENGKETLQRT (서열번호 1), [HLA-A*11:01]:
 SHSMRYFYTSVSRPGRGEPRFIAVGYVDDTQFVRFDSDA
 ASQRMEPRAPWIEQEGPEYWDQETRNKKAQSQTDRVDLGT
 LRGYYNQSEDSHTIQIMYGCDVGPDGRFLRGYRQDAYDG
 KDYLALNEDLSWTAADMAAQITKRKWEAAHAAEQRAYL
 EGRCVEWLRRYLENGKETLQRT (서열번호 2), [HLA-A*24:02]:
 SHSMRYFSTSVSRPGRGEPRFIAVGYVDDTQFVRFDSDAASQRMEPRAPWIE
 QEGPEYWDEETGKVKKAHSQTDRENLRALRYYNQSEAGSH
 TLQMMFGCDVGSDGRFLRGYHQYAYDGKDYLKEDLSW
 TAADMAAQITKRKWEAAHVAEQQRAYLEGTCVDGLRRYLENGKETLQRT
 (서열번호 3);
- [281] - Source: Hugo et al., 2016, Cell 165, 35-44, Riaz et al., 2017, Cell 171, 934-949 (본 명세서에 참조로서 포함됨).
- [282]
- [283] 상기 선별된 환자군 데이터에 적용한 신항원 예측 모델은 다음과 같다:
- [284] - Binding Affinity Score (BA) = (선별된 환자군의 Binding Affinity score 평균값) - (선별된 환자군의 Binding Affinity Score 분산값);
- [285] - Immunogenicity Score (IMM) = (선별된 환자군의 Immunogenicity score 평균값) - (선별된 환자군의 Immunogenicity Score 분산값)
- [286]
- [287] 상기 BA 점수는 각 환자의 신항원과 상기 환자 고유의 A*02:01, A*24:02, 및 A*11:01의 결합 결과로부터 얻어진 것으로, 실시예 1.2의 예측값 1을 학습한 결합력 예측모델을 사용하여 얻어진 값이다 (실시예 2.1 참조). 상기 IMM 점수는 각 환자의 신항원과 상기 환자 고유의 HLA type의 면역원성 예측 결과로부터 얻어진 것으로, 예측값 2 (실시예 1.2)를 학습한 면역원성 예측모델을 사용하여 얻어진 값이다 (실시예 2.1 참조).
- [288]

[289] TPM (RNA expression level; transcripts per million) score를 적용하여, Neoantigen Prioritization Score를 계산하였다 (TPM threshold: 1; 실시예 3 참조):

[290] $\text{Prioritization Score} = \min(\text{TPM}, 1) * [(\text{BA} + \text{IMM})/2]$

[291] $\text{Min}(\text{TPM}, 1)$: TPM > 1인 경우 TPM = 1로 치환하여 사용하고 TPM < 1인 경우 TPM Value를 그대로 사용하는 것을 의미)

[292]

[293] 상기 얻어진 Prioritization Score로 Rank 진행하여 상위 20 개의 펩타이드를 실험원으로 예측하였다 (표 4 내지 표 6).

[294] [표4]

Rank	Sequence	SEQ ID NO	HLA type	Score
1	YLFNFVVNV	4	A*02:01	0.557830519
2	RLAEFIIGM	5	A*02:01	0.551505094
3	YLFNFVVNVG	6	A*02:01	0.501269973
4	RLAEFIIGMM	7	A*02:01	0.499001865
5	LLTDIIPDGV	8	A*02:01	0.494219272
6	FLALCADA V	9	A*02:01	0.493808672
7	YIAGILARI	10	A*02:01	0.490358034
8	RLAEFIIGMMG	11	A*02:01	0.477305298
9	YLMVALASV	12	A*02:01	0.471998357
10	RLAEFIIGMMGT	13	A*02:01	0.459548693
11	GLVECLPEI	14	A*02:01	0.456333267
12	AYLFNFVVNV	15	A*02:01	0.448139594
13	VL FYLSSAV	16	A*02:01	0.448117179
14	RLFFCFMNV	17	A*02:01	0.443291682
15	CILAVLPSL	18	A*02:01	0.440896908
16	IMFNKGFLA	19	A*02:01	0.438771916
17	KTDPWFLLAV	20	A*02:01	0.437365302
18	SLIGVSFFALPV	21	A*02:01	0.437040423
19	SLWAESLVHV	22	A*02:01	0.436619991
20	FLLAVLGAPV	23	A*02:01	0.435148758

[295] [표5]

Rank	Sequence	SEQ I D NO	HLA type	Score
1	TFYWTFVNAF	24	A*24:02	0.455910057
2	VFYQFSKLIF	25	A*24:02	0.403493367
3	RMYSWRLNL	26	A*24:02	0.371202482
4	TWNVHQQFLW	27	A*24:02	0.366847255
5	RMYSWRL	28	A*24:02	0.36676025
6	FYWTFVNAFSIF	29	A*24:02	0.363375765
7	QWILKTRAYFL	30	A*24:02	0.338967703
8	STWNVHQQF	31	A*24:02	0.338013999
9	QWILKTRAYF	32	A*24:02	0.330919878
10	STWNVHQQFLW	33	A*24:02	0.330000281
11	FYWTFVNAF	34	A*24:02	0.328668696
12	STWNVHQQFLW	35	A*24:02	0.327606417
13	FYWTFVNAF	36	A*24:02	0.324465905
14	STWNVHQQFLW	37	A*24:02	0.320806235
15	FYWTFVNAF	38	A*24:02	0.319582139
16	STWNVHQQFLW	39	A*24:02	0.316823043
17	FYWTFVNAF	40	A*24:02	0.316026847
18	STWNVHQQFLW	41	A*24:02	0.314389744
19	FYWTFVNAF	42	A*24:02	0.313212986
20	LYHFSDRMYSW	43	A*24:02	0.312662407

[296]

[표6]

Rank	Sequence	SEQ I D NO	HLA type	Score
1	STACFHPVTK	44	A*11:01	0.463350479
2	TTFGIQVRYLK	45	A*11:01	0.436937466
3	KIGDFGLATMK	46	A*11:01	0.42843179
4	KIGDFGLATEK	47	A*11:01	0.427919868
5	SSSGSINLLK	48	A*11:01	0.420742704
6	AVAVDFQFSK	49	A*11:01	0.408470424

7	KTAPVEWRK	50	A*11:01	0.392922075
8	STSYLQHREK	51	A*11:01	0.390245612
9	VAVDFQFSK	52	A*11:01	0.382257777
10	SAFTIQEYFVK	53	A*11:01	0.38028147
11	SVSNTTVGSK	54	A*11:01	0.378114721
12	RVQTGTIPK	55	A*11:01	0.377405025
13	SSGSINLLK	56	A*11:01	0.373327626
14	GTA AISVTGALK	57	A*11:01	0.373120201
15	STPNFSRLRRY	58	A*11:01	0.372708181
16	RSVLSSGYQK	59	A*11:01	0.372577947
17	VVSPVSLPRR	60	A*11:01	0.370444394
18	AAISVTGALK	61	A*11:01	0.36963003
19	GSQPPSLGRK	62	A*11:01	0.369425624
20	TVYCRYHNY	63	A*11:01	0.368830969

[297] 상기 예측된 신항원의 검증을 위하여, HLA type 정보가 확보된 환자 성인 donor 3명의 PBMC(Peripheral Blood Mononuclear Cells)를 수득하고 이용하여 면역원성 평가 (IFN γ ELISPOT)를 수행하였다 (상기 3명의 donor는 각기 다른 HLA 타입을 가지며, 상기 수득된 PBMC는 각각 A*02:01, A*24:02, 또는 A*11:01를 가짐). 상기 시험과정을 도 2에 모식적으로 나타내었다.

[298] 상기 얻어진 ELISPOT 결과에 대하여, 하기의 수식으로 Normalization을 수행하여 분석하였다:

[299]

$$\text{Normalization} = \left(\frac{X - X(\text{media})}{X(\text{CPI}) - X(\text{media})} \right)$$

[300] (Threshold >10 으로 설정하여 면역원성 Positive를 선정함;

[301] X: 시험군의 스팟(ELISPOT) 개수;

[302] X(media): 배지만 처리한 군(비처리군; negative control)의 ELISPOT 개수;

[303] X(CPI): (Protein antigens of CMV, Parainfluenza and influenza virus)를 처리한 군 (positive control)의 ELISPOT 개수;

[304] 각 시험마다의 결과의 일정성을 유지하기 위하여 normalization을 수행함)

[305]

[306] 각 HLA type Match된 환자 8명 (A*02:01), 7명 (A*24:02), 5명 (A*11:01)에 대하여 실시예 1 및 2에 기재되어 있는 신항원 예측모델을 이용하여 Top20 및 Top10 신항원 예측을 진행하였고, 실시예 3 및 4에 해당하는 시험 방법을 통하여 20종

예측된 신항원에 대한 면역원성 평가를 상기한 바와 같이 INFg ELISPOT을 통하여 진행하여, 각 환자의 20종 신항원 중 면역원성이 있다고 판단되는 신항원의 수를 확인하였다.

[307]

[308] 상기 얻어진 결과를 표 7 및 도 2b-2d에 나타내었다:

[309] [표7]

HLA-I type	환자 수	Top20 (%)	Top10 (%)
A*02:01	8	26.8	22.5
A*24:02	7	20.7	14.3
A*11:01	5	21.0	18.0

[310] A*02:01 8명, A*24:02 7명, A*11:01 5명에 대하여 면역원성 여부를 실험적으로 평가하였을 때, 각각 26.8%, 20.7%, 21.0%의 면역원성이 있는 신항원 예측율을 확인하였다.

[311]

[312] 실시예 5. *Ex vivo* Immunogenicity

[313] 5.1. Human Cell Line (MCF7 cell line) Derived Xenograft Model

[314] MCF7 유방암 세포주는 인간 암세포주로서, HLA-A*02:01을 발현하고 있다. 실시예 4를 참조하여, MCF7 세포주(ATCC, cat# HTB-22)에 대한 WES(whole exosome sequencing; Mutation Selection), RNA-Seq (Express level)의 NGS Sequencing을 진행하였고, 신항원 예측모델 (BA, IMM, TPM)을 사용하여 상위 20개(Top20)를 선택하여 20종의 신항원(mRNA)을 선정하고, 이를 mRNA-lipoplex 항암백신(KR 10-2022-0149461 A, KR 10-2022-0149460 A의 실시예 A(1) 참조)으로 만들어 A02:01 Transgenic 마우스(HLA-A2.1을 발현하도록 만든 C57BL/6 마우스; Taconic biosciences, cat# 8906-F)에 투여(60ug x 3회 QW로 투여, tail-vein IV injection)하여 면역원성 여부를 splenocyte IFNg ELISPOT을 통하여 확인하였다.

[315] 상기 면역원성 검정은 다음과 같은 과정으로 확인하였다: 항-IFNg 항체가 코팅된 96-웰 ELISPOT 플레이트를 준비하였다. 상기 플레이트를 PBS로 한번 세척 후, 상기 면역화된 마우스로부터 채취한 비장세포를 각 well 당 1.0×10^6 개씩 로딩하였다. 각각의 면역 그룹에 대해 세포를 3웰씩 진행하였고, 총 200ul의 부피 중 10uM의 펩타이드 또는 배지 단독으로 재자극하였다. 그 후 플레이트를 37°C 에서 18시간동안 인큐베이션 하였다. PBST로 3번 세척한 후, 각 well을 비오틴(biotin)이 표지된 마우스 IFNg 검출항체로 처리하고 플레이트를 2시간 상온에서 인큐베이션 하였다. 세포를 다시 세척하고 AP-콘쥬게이트된 스트렙타비딘(streptavidin)을 각 well에 첨가하여 30분간 상온에서 인큐베이션 하였다. AP반응은 기질 시약 세트(Immunospot)를 사용하여 전개하였다. Well 당 스팟 형성 단위(SFUs)의 수를 ELISPOT 판독기(CTL ELISPOT 판독기)를 사용하여 자동으로 계수하였다.

[316] 상기 백신 투여 과정을 포함한 시험 과정을 도 3에 모식적으로 나타내었다.

[317] 선정된 20종의 신항원을 표 8에 나타내었다:

[318] [표8]

Rank	Name	Neoantigen	SEQ ID NO
1	MCF7-1	LLFDMTTHSV	64
2	MCF7-2	FQSFFSLIV	65
3	MCF7-3	FLGSGVHGGV	66
4	MCF7-4	YLFFCGHMFHA	67
5	MCF7-5	ITFEHILAL	68
6	MCF7-6	NLANEITYI	69
7	MCF7-7	VLCEPVHGL	70
8	MCF7-8	MLLESLRHAV	71
9	MCF7-9	SLFGLGFEEHV	72
10	MCF7-10	HLLTDELGYV	73
11	MCF7-11	KLSEASMIV	74
12	MCF7-12	ALLDMRSL	75
13	MCF7-13	GLWEALTPCNV	76
14	MCF7-14	SLARLHLTV	77
15	MCF7-15	TLLPLPLL	78
16	MCF7-16	HLFWFADLNYRL	79
17	MCF7-17	STMSFLRSLYL	80
18	MCF7-18	TLNHEIYEA	81
19	MCF7-19	ILWGSISFNLA	82
20	MCF7-20	MVCGAAFFI	83

[319] 상기 얻어진 splenocyte IFNg ELISPOT 결과를 도 4에 나타내었다. 도 4에서 control은 신항원 펩타이드를 처리하지 않은 무처리군의 결과이고, 'MCF7 NEO'는 상기 20종 선정된 신항원을 탑재한 mRNA-lipoplex 항암백신을 투여한 군의 결과이다. 무처리군(control)의 ELISPOT의 평균의 3배 이상을 threshold로 설정하여, ELISPOT 결과가 threshold 이상인 것을 항암 효능이 있는 것으로 판단하였다. 그 결과, 상기 시험된 20종의 신항원 중에서 항암 효능이 있을 것으로 예상되는 신항원은 25%로 평가하였다 (도 4의 * 표시됨; 5/20).

[320]

[321] **5.2. B16F10 (Melanoma) AI Model Predicted Neoantigens**

[322] 실시예 4를 참조하여, B16F10 (Mouse Melanoma; ATCC, cat# CRL-6475) 세포주에 NGS를 수행하고(아미노산 서열로 변환), 신항원 예측모델(BA, IMM, TPM)을 통한 Top 50 및 Top 20 신항원을 예측하였다. 상기 예측된 신항원을 mRNA-lipoplex 항암백신으로 만들어 C57BL/6 마우스에 투여(60ug x 2회 QW IV injection)하여 면역원성 여부를 splenocyte IFN γ ELISPOT을 통하여 확인하였다(도 5 참조). Positive control로서 mTRP2 (Tyrosinase-related protein 2)(Melanoma Tumor-associated Antigen; SVYDFFVWL (서열번호 134))를 앞서 설명한 방법과 동일하게 투여한 마우스 모델을 사용하고, Threshold 설정으로, 상기 예측된 신항원 면역원성 여부를 판단하였다.

[323]

[324] 본 실시예에 사용된 Top50 신항원 서열을 표 9에 기재하였다:

[325] [표9]

B16F10 Top50 Sequence

Rank	B16F10 Neo Sequence	SEQ ID NO
1	MSLPFSFL	84
2	ANFIFRQL	85
3	STYKLTNL	86
4	NMYGFRNV	87
5	VIYSHLQV	88
6	SGFRYNVL	89
7	VVMEYENL	90
8	IVHYFIAL	91
9	VSFAPLVQL	92
10	ATFYFHHPV	93
11	VHYLRSL	94
12	IGIFFQMRV	95
13	VAHVAAFL	96
14	VFLRVRAL	97
15	TIFMMLRL	98
16	KGFLYSAL	99
17	SSVLFKGL	100
18	MQFYLRHL	101

19	ARMMFSGL	102
20	KSLYYASCL	103
21	LVFVVGRL	104
22	SIGVYIPL	105
23	TVYRCGHL	106
24	SLFTLHQSV	107
25	FTFVYPTI	108
26	CMLRYQTCL	109
27	VCFDYTFAL	110
28	LVVYYLPL	111
29	VPYAVFPPL	112
30	VSFANAVV	113
31	VVPSFSAEI	114
32	CNFCYQRNQL	115
33	QMYPAFKRV	116
34	ILVSTYRPL	117
35	SFLAGFRPL	118
36	AIYHHASRAI	119
37	IAMGFPVERL	120
38	VLILHAAAL	121
39	SLVAPFGSV	122
40	MYRYFVTL	123
41	VMMDNFDYL	124
42	VTFQAFIDV	125
43	RAFAAVRL	126
44	KNFEFNKQRSIV	127
45	LMYVYEMLGAGL	128
46	CMLRYQTCLV	129
47	AIALHLHL	130
48	TSYSTAAPL	131

49	SMPHYSSPAI	132
50	TIPFYDNPI	133

[326] 상기 얻어진 결과를 도 6에 나타내었다. 도 6에 나타난 바와 같이, 상위 20개 (Top 20) 예측 신항원 중 45% (9/20)의 면역원성을 확인하였다 (도 6의 상단 참조). Top 50 신항원 중의 면역원성 높은 신항원 12종 중에서 9종 신항원이 Top 20 안에 분포되어 있어, Top 20을 기준으로 신항원 탐색시에 타겟 신항원 포함 가능성이 높음을 확인하였다 (표 10 참조).

[327] [표10]

Rank	Positive
Top 20	9 (45%)
Top21~Top 50	3 (10%)
Total	12 (24%)

[328] (Immunogenicity Threshold: > 3x Media ELISPOT Avg.)

[329]

[330] 실시예 6: **In vivo Efficacy:** 예측 신항원 항암 효능

[331] B16F10 종양세포 5×10^4 개를 C57BL/6 마우스에 subcutaneous 이식방법으로 이식하여, 50mm^3 종양크기에서 군분리를 실시하였다. 시험군은 실시예 5.2에서 확보된 12종의 B16F10 예측 신항원 중 상위 10종을 탑재한 mRNA-lipoplex를 투여하는 군, mTRP2 (Tyrosinase-related protein 2)(Melanoma Tumor-associated Antigen) (서열번호 134)를 탑재한 mRNA-lipoplex를 투여하는 positive control, 신항원 탑재 mRNA 없이 lipoplex만 투여하는 negative control로 구성하였다. mRNA-lipoplex는 1회 60ug씩 5일마다 3회 투여를 진행하고 3주간 종양 크기를 기록하였다.

[332]

[333] 상기 얻어진 결과(종양 크기)를 도 7에 나타내었다 (Neoantigen Cancer Vaccine: 신항원 10종 투여군, Positive control: mTRP2 투여군, control: lipoplex만 투여 (negative control)). 도 7에 나타난 바와 같이, B16F10 예측 신항원 10종을 탑재한 mRNA-lipoplex를 투여군에서는 63%의 종양감소를 확인하였다.

[334]

[335] 이상의 설명으로부터, 본 발명이 속하는 기술분야의 당업자는 본 발명이 그 기술적 사상이나 필수적 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 이와 관련하여, 이상에서 기술한 실시 예들은 모든 면에서 예시적인 것이며 한정적인 것이 아닌 것으로서 이해해야만 한다. 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허 청구범위의 의미 및 범위 그리고 그 등가 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

청구범위

- [청구항 1] 다음의 단계 (1) 내지 (7)을 포함하는, 환자 유래 종양-특이 면역원성 펩타이드 선정 방법:
- (1) 환자 유래 종양세포 또는 종양조직 및 정상세포 또는 정상조직으로부터 NGS 서열정보를 획득하는 단계;
 - (2) 정상세포 또는 정상조직 NGS 서열과 비교하여 상기 종양세포 또는 종양조직 NGS 서열 중 하나 이상의 아미노산 변이를 가진 아미노산 7개 이상의 길이의 종양-특이 펩타이드 서열정보를 획득하는 단계;
 - (3) 종양 세포 또는 종양 조직 내의 종양-특이 펩타이드 또는 이의 코딩 유전자의 발현수준(TPM; Transcripts per Million)을 측정하는 단계;
 - (4) 상기 종양-특이 펩타이드 서열정보를 결합력 예측모델에 적용하여 HLA 분자에 대한 결합 점수(binding affinity score; BA)를 확인하는 단계;
 - (5) 상기 종양-특이 펩타이드 서열정보를 면역원성 예측모델에 적용하여 면역원성 점수(immunogenicity score; IMM)를 확인하는 단계;
 - (6) 하기의 수식으로 계산된 수치를 기준으로 종양-관련 펩타이드의 순위를 정하는 우선순위 결정 단계:

$$\text{Neoantigen prioritization score} = \min(\text{TPM}, 1) * (W1 * \text{BA} + W2 * \text{IMM}) / (W1 + W2)$$

$$[0 < W1 < 1, 0 < W2 < 1 \text{ 범위서 설정}]; \text{ 및}$$
 - (7) 상기 결정된 종양-특이 펩타이드의 우선순위를 바탕으로 2 이상의 종양-특이 면역원성 펩타이드를 선정하는 단계.
- [청구항 2] 제1항에 있어서, 다음의 (a) 내지 (d) 단계를 추가로 포함하는, 환자 유래 종양-특이 면역원성 펩타이드 선정 방법:
- (a) 펩타이드 데이터베이스로부터 종양세포 또는 종양조직 유래 펩타이드 서열정보 및 정상세포 또는 정상조직 유래 펩타이드 서열정보를 획득하는 단계,
 - (b) 상기 생물학적 서열 데이터베이스로부터 클래스 I HLA 펩타이드 서열정보를 획득하는 단계,
 - (c) 상기 (a) 및 (b) 단계에서 획득된 데이터를 HLA에 대한 결합점수(binding affinity score; BA)를 예측할 수 있는 모델에 학습하여 결합력 예측모델을 구축하고 결합점수를 얻는 단계; 및
 - (d) 상기 (a) 및 (b) 단계에서 획득된 데이터를 면역원성 점수(immunogenicity score; IMM)를 예측할 수 있는 모델에 학습하여 면역원성 예측모델을 구축하고 면역원성 점수를 얻는 단계.
- [청구항 3] 제2항에 있어서, 상기 단계 (a)의 예측모델 구축에 사용된 펩타이드 데이터베이스는 하기 i) 내지 iv)의 과정 중 하나 이상의 선별과정을 거치는 것을 특징으로 하는, 환자 유래 종양-특이 면역원성 펩타이드 선정 방법:

- i) HLA-A, HLA-B, HLA-C에 해당하지 않는 데이터의 삭제,
 - ii) HLA 분자에 돌연변이가 있는 데이터의 삭제,
 - iii) 펩타이드의 길이가 7 이하 또는 15 이상인 데이터의 삭제, 및
 - iv) 아미노산 서열에 인간 유래 아미노산 20종 외의 아미노산 잔기가 포함된 것의 제거.
- [청구항 4] 제1항 내지 제3항에 있어서, 상기 HLA는 상이한 HLA 유형, 또는 상이한 HLA 대립유전자를 포함하는 것인, 환자 유래 종양-특이 면역원성 펩타이드 선정 방법.
- [청구항 5] 제1항 내지 제3항 중 어느 한 항에 있어서, 상기 (c) 및 (d) 단계의 학습은 하나 이상의 인공지능 모델 및 하나 이상의 학습방법을 조합하여 수행하는 것인,
환자 유래 종양-특이 면역원성 펩타이드 선정 방법.
- [청구항 6] 제1항 내지 제3항 중 어느 한 항에 있어서, 상기 결합점수는 HLA와의 결합과 관련된 IC50값 또는 Kd값에 의하여 결정되는 것인, 환자 유래 종양-특이 면역원성 펩타이드 선정 방법.
- [청구항 7] 제1항 내지 제3항 중 어느 한 항에 있어서, 상기 면역원성 점수는 상기 후보 펩타이드를 면역세포 또는 이를 포함하는 시료에 처리한 경우의 T세포 유래 사이토카인 분비량, 면역원성 활성화로 인한 T세포 proliferation, 케모카인(Chemokine) 분비량, 활성화된 T세포의 cytotoxicity와 관련된 정보로 이루어진 군에서 선택된 하나 이상의 정보에 의하여 결정되는 것인, 환자 유래 종양-특이 면역원성 펩타이드 선정 방법.
- [청구항 8] 제7항에 있어서, 상기 면역세포를 포함하는 시료는 혈액, 백혈구, 및 말초혈액단핵구(PBMC)로 이루어진 군에서 선택되고, 상기 정보는 IFN- γ , IL-2, TNF- α , CCL4, CXCL9, 및 granzyme B로 이루어진 군에서 선택된 2종 이상인, 환자 유래 종양-특이 면역원성 펩타이드 선정 방법.
- [청구항 9] 제1항 내지 제3항 중 어느 한 항에 있어서, 상기 단계 (3)의 환자 유래 시료 내의 후보 펩타이드 또는 이의 코딩 유전자의 발현수준은 환자 시료 데이터에서 RNA 시퀀싱(RNA-sequencing)을 통하여 얻어지는 것인, 환자 유래 종양-특이 면역원성 펩타이드 선정 방법.
- [청구항 10] 제1항 내지 제3항 중 어느 한 항에 있어서, 상기 환자 유래 종양-특이 면역원성 펩타이드는 상기 환자의 개인맞춤형 항암 백신의 제조에 사용하기 위한 것인, 환자 유래 종양-특이 면역원성 펩타이드 선정 방법.
- [청구항 11] 제1항 내지 제3항 중 어느 한 항의 방법으로 선정된 종양-특이 면역원성 펩타이드 또는 상기 펩타이드를 암호화하는 핵산 분자 및
약학적으로 허용 가능한 담체를 포함하는, 항암 백신 조성물.
- [청구항 12] 제1항 내지 제3항 중 어느 한 항의 방법으로 선정된 종양-특이 면역원성 펩타이드 또는 상기 펩타이드를 암호화하는 핵산 분자 및

약학적으로 허용 가능한 담체
를 혼합하는 단계를 포함하는, 항암 백신 제조 방법.