



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2022년11월28일
(11) 등록번호 10-2470937
(24) 등록일자 2022년11월22일

(51) 국제특허분류(Int. Cl.)
G16B 5/00 (2019.01) G16B 40/00 (2019.01)
G16C 20/30 (2019.01)
(52) CPC특허분류
G16B 5/00 (2019.02)
G16B 40/00 (2019.02)
(21) 출원번호 10-2022-0040238
(22) 출원일자 2022년03월31일
심사청구일자 2022년03월31일
(30) 우선권주장
1020210135020 2021년10월12일 대한민국(KR)

(73) 특허권자
포항공과대학교 산학협력단
경상북도 포항시 남구 청암로 77 (지곡동)
이뮤노바이옴 주식회사
경상북도 포항시 남구 지곡로 47, 2101호 (효자동, 바이오오픈이노베이션센터)
(72) 발명자
김상욱
경상북도 포항시 남구 지곡로 155, 9동 403호(지곡동, 교수아파트)
공정호
부산광역시 해운대구 해운대해변로 99, 101동 204호(우동, 경남마리나아파트)
(74) 대리인
특허법인엠에이피에스

전체 청구항 수 : 총 14 항

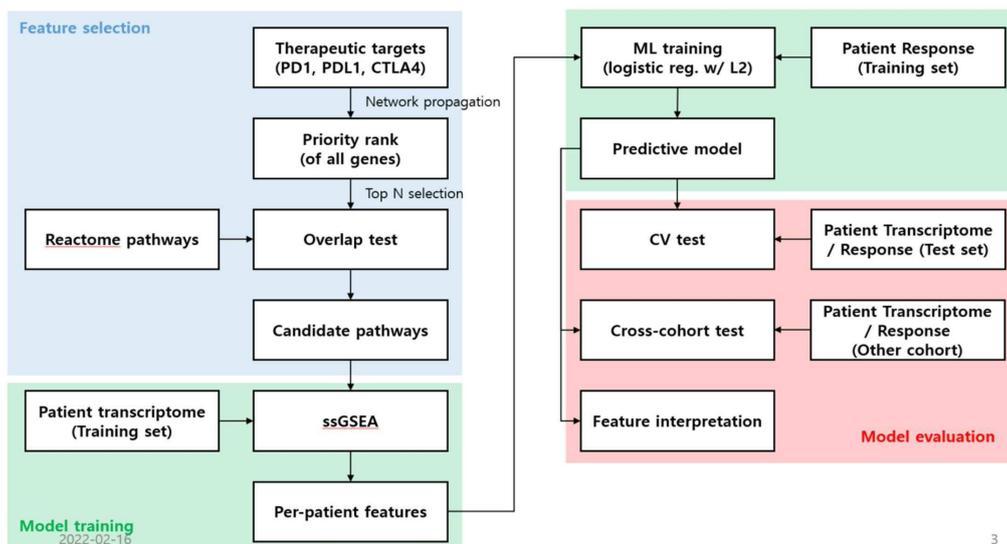
심사관 : 성경아

(54) 발명의 명칭 네트워크에 기반한 머신러닝 기법을 활용하여 암 환자에 대한 ICI 치료 효과와 전체 생존률을 예측할 수 있는 바이오마커의 탐색 장치 및 방법

(57) 요약

본원은 ICI 치료에 있어, 해당 치료에 대한 반응과 해당 치료를 받은 환자의 전체 생존률을 예측할 수 있는 바이오마커의 탐색방법을 제공하고자 한다. 본원의 장치 및 방법을 활용하면 암 환자에 대한 ICI 치료의 효과와 환자의 생존률을 정확하게 예측할 수 있는 바이오마커를 찾아낼 수 있으며, 이에 따라 ICI 치료 효과를 극대화할 수 있다.

대표도



(52) CPC특허분류
G16C 20/30 (2019.02)

(72) 발명자

김인해

경상북도 포항시 남구 지곡로 102, 8동 202호(지곡
동, 낙원아파트)

박창욱

서울특별시 송파구 송파대로 345, 411동 2402호(가
락동, 헬리오시티)

공지예외적용 : 있음

명세서

청구범위

청구항 1

머신 러닝을 이용하여 암환자의 면역 항암제에 대한 반응 유무를 판별하는 장치에 있어서,
 유전자 네트워크 중 면역 항암제의 표적을 포함하는 대상 생물학적 경로를 추출하는 생물학적 경로 추출부;
 상기 면역 항암제를 이용하여 면역 요법을 수행할 대상 암환자의 전사체 데이터(transcriptome data)로부터 유전자 활성 정보를 상기 대상 생물학적 경로의 활성 정보로 변환하는 유전자 활성 정보 변환부; 및
 미리 학습된 면역 항암제 반응 판별 모델에 상기 대상 유전자 정보를 입력하여 상기 대상 암환자의 상기 면역 항암제에 대한 반응 유무를 판별하는 판별부를 포함하는 것인, 장치.

청구항 2

제 1 항에 있어서,
 상기 경로 추출부는 페이지 랭크 알고리즘을 이용한 네트워크 전파를 통한 영향력 점수에 기초하여 상기 유전자 네트워크 중 상기 표적에 대응하는 표적 노드 및 상기 표적 노드에 근접한 복수의 근위 노드를 검출하는 것인, 장치.

청구항 3

제 2 항에 있어서,
 상기 경로 추출부는 유전자 집합 농축 분석(gene set enrichment test) 및 초기하 테스트(hypergeometric test)를 이용한 normalized enrichment score (NES)에 기초하여 복수의 후보 생물학적 경로 중 상기 대상 생물학적 경로를 선정하는 것인, 장치.

청구항 4

제 1 항에 있어서,
 상기 유전자 네트워크는 단백질-단백질 상호작용(Protein-Protein Interaction) 네트워크인 것인, 장치.

청구항 5

제 1 항에 있어서,
 상기 면역 항암제는 항 PD-1 항체, 항 PD-L1 항체 및 항 CTLA4 항체 중 적어도 하나를 포함하는 것인, 장치.

청구항 6

제 1 항에 있어서,
 상기 표적은 PD-1 단백질, PD-L1 단백질 및 CTLA4 단백질 중 적어도 하나를 포함하는 것인, 장치.

청구항 7

제 1 항에 있어서,

상기 면역 항암제 반응 판별 모델은 복수의 암환자의 상기 대상 유전자 정보 및 상기 면역 항암제에 대한 반응 유무에 대한 임상 결과에 기초하여 미리 학습된 것인, 장치.

청구항 8

머신 러닝을 이용하여 암환자의 면역 항암제에 대한 반응 유무를 판별하는 방법에 있어서,

유전자 네트워크 중 면역 항암제의 표적을 포함하는 대상 생물학적 경로를 추출하는 단계;

상기 면역 항암제를 이용하여 면역 요법을 수행할 대상 암환자의 전사체 데이터(transcriptome data)로부터 유전자 활성 정보를 상기 대상 생물학적 경로의 활성 정보로 변환하는 단계; 및

미리 학습된 면역 항암제 반응 판별 모델에 상기 대상 유전자 정보를 입력하여 상기 대상 암환자의 상기 면역 항암제에 대한 반응 유무를 판별하는 단계

를 포함하는 것인, 방법.

청구항 9

제 8 항에 있어서,

상기 대상 생물학적 경로를 추출하는 단계는 페이지 랭크 알고리즘을 이용한 네트워크 전파를 통한 영향력 점수에 기초하여 상기 유전자 네트워크 중 상기 표적에 대응하는 표적 노드 및 상기 표적 노드에 근접한 복수의 근위 노드를 검출하는 단계를 포함하는 것인, 방법.

청구항 10

제 9 항에 있어서,

상기 대상 생물학적 경로를 추출하는 단계는 유전자 집합 농축 분석(gene set enrichment test) 및 초기하 테스트(hypergeometric test)를 이용한 normalized enrichment score (NES)에 기초하여 복수의 후보 생물학적 경로 중 상기 대상 생물학적 경로를 선정하는 단계를 더 포함하는 것인, 방법.

청구항 11

제 8 항에 있어서,

상기 유전자 네트워크는 단백질-단백질 상호작용(Protein-Protein Interaction) 네트워크인 것인, 방법.

청구항 12

제 8 항에 있어서,

상기 면역 항암제는 항 PD-1 항체, 항 PD-L1 항체 및 항 CTLA4 항체 중 적어도 하나를 포함하는 것인, 방법.

청구항 13

제 8 항에 있어서,

상기 표적은 PD-1 단백질, PD-L1 단백질 및 CTLA4 단백질 중 적어도 하나를 포함하는 것인, 방법.

청구항 14

제 8 항에 있어서,

복수의 암환자의 상기 대상 유전자 정보 및 상기 면역 항암제에 대한 반응 유무에 대한 임상 결과에 기초하여 상기 면역 항암제 반응 판별 모델을 학습하는 단계를 더 포함하는 것인, 방법.

발명의 설명

기술 분야

[0001] 본 발명은 네트워크에 기반한 머신러닝 기법을 활용하여 암 환자에 대한 ICI 치료 효과 및 전체 생존률을 예측하는 장치 및 방법에 관한 발명이다.

배경 기술

- [0002] 암은 우리나라 국민의 사망률 1위를 기록하는 질병으로, 항암제 개발의 필요성은 꾸준히 대두되고 있다.
- [0003] 항암제 개발 과정을 살펴보면, 빠르게 증식하는 종양세포의 특성을 이용하여 분열하는 세포를 공격하는 화학항암제와 종양 세포의 특정 분자나 신호전달 체계를 공격하는 표적항암제가 존재하였으나, 여러 부작용이 존재하였고, 체내의 선천 면역을 이용하여 부작용을 최소화 할 수 있는 면역항암제가 등장하였다.
- [0004] 면역항암치료란, 인체의 면역체계를 활성화 시켜서 암세포와 싸우게 하는 암 치료법을 말한다. 면역항암치료는 면역시스템을 이용하여 암세포만 공격해 기존의 항암치료보다 부작용이 적고, 면역시스템의 기억 능력과 적응력을 이용하기 때문에 장기간의 항암효과를 얻을 수 있다는 장점이 있다. 상기와 같이 기존 항암제의 단점을 극복하는 면역항암치료는 암치료의 새로운 패러다임으로 각광받고 있으며, 사이언스지는 2013년 올해의 연구로 면역항암제를 선정한 바 있다.
- [0005] 면역항암제는 종양 항원을 표적하는 항체치료제(Rituximab 등), 면역세포를 다시 활성화 하는 면역관문억제제(Immune checkpoint inhibitor 등), 면역세포를 직접 투여하는 면역세포치료제(Immune cell therapy)등으로 구분할 수 있다(Oiseth *et al.*, 2017).
- [0006] 면역관문억제제(ICI; Immune checkpoint inhibitors)는 수많은 암 환자들의 생존에 기여해왔다. 다른 화학적 요법과 비교했을 때 ICIs 요법은 대체로 부작용이 훨씬 적으며 치료 효과가 오래간다는 장점을 지닌다. 이에 ICIs 요법은 더욱 발전하여 현재는 흑색종(melanoma), 방광암(bladder cancer), 그리고 위장암(gastro-esophageal) 등 적용 가능한 암의 범위가 상당히 넓어졌다.
- [0007] 그럼에도 아직 적은 수의 환자만이 ICI 요법의 효과를 볼 수 있으며 (30% 이하의 치료율) 치료에 따른 독성이 존재할 수 있다. 이에 ICI 요법에 대한 바이오마커(ICI-response-associated biomarker)를 찾아, 치료 전 환자에 대한 치료 효과가 어느 정도인지 예측함으로써 환자의 전체 생존률을 높이는 방법이 시급한 실정이다.
- [0008] 면역학적인 약물 치료에서 중요한 것은 다양한 암 환자 코호트에 대해 치료에 대한 반응을 정확하게 확인할 수 있는 마커를 찾아내는 것이다. 예를 들어, 면역화학에 의한 PD1/PD-L1 발현은 여러 암종에 대해 FDA가 승인한 테스트이다. 이에 더하여 수많은 연구 결과 비소형 폐암에서 PD-L1 발현과 ICI 반응 간의 양적관계가 확인된 바 있다. 그러나 다른 연구들은 PD-L1 발현과 ICI 반응 사이에 특별한 관계가 없다는 결과를 내거나, 심지어 음적 관계를 보인다고 보고하기도 한다. 이와 같이, 종전에 확인되었던 바이오마커들이 일관된 반응을 보이지 않으면서 더 정확한 예측이 가능한 바이오마커를 찾을 필요성이 대두되고 있다. 최근에는 Litchfield는 기존의 바이오마커들이 ICI 반응에 대해 고작 60%정도만 그 효과를 확인할 수 있다는 보고가 있으며 새로운 인자를 찾을 것을 제안하기도 했다.(Litchfield, K. *et al.* Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* 184, (2021).)
- [0009] 네트워크 생물학(Network biology)는 적절한 바이오마커를 찾는 데 큰 도움이 된다. 네트워크 기반의 바이오마커 탐색은 표현형 상 비슷한 기능을 하는 유전자는 PPI(protein-protein interaction)의 특정 구역에서 대체로 같은 곳에 위치한다는 점을 이용한다. 이러한 경향성을 활용하여, 단일 유전자 기반의 탐색에 비해 정확한 표현형 예측이 가능한 유전자 모듈을 찾아왔다. 예를 들어, Hofree *et al.*은 비슷한 네트워크 구역에서 단지 하나의

돌연변이만이 공통인 환자군의 치료 결과가 거의 유사함을 확인하였다.(Hofree, M., Shen, J. P., Carter, H., Gross, A. & Ideker, T. Network-based stratification of tumor mutations. *Nat. Methods* 10, 1108-1115 (2013)) Guney *et al.*은 약효는 약물의 작용 부위와 질병 유전자 간의 거리가 가까울수록 좋은 경향을 보인다는 것을 증명하였다.(Guney, E., Menche, J., Vidal, M. & Barabasi, A.-L. Network-based in silico drug efficacy screening. *Nat. Commun.* 7, 10331 (2016).) 또한 본 발명자들은 환자로부터 유래한 오가노이드 모델 실험의 약리유전체학적 자료를 이용한 실험을 통해, 네트워크 근접성을 이용하여 암환자의 전체 생존률을 예측할 수 있는 약학 반응 바이오마커를 확인할 수 있다고 보고하였다. 결과적으로 네트워크 기반의 탐색으로 정확하고 노이즈가 적은 바이오마커를 찾을 필요성이 있지만, 이러한 접근이 큰 암환자 코호트 사이에서 ICI 치료의 효과를 예측할 수 있다는 점은 아직 증명되지 않았다.

발명의 내용

해결하려는 과제

- [0010] 본원은 ICI 치료에 있어, 해당 치료에 대한 반응과 해당 치료를 받은 환자의 전체 생존률을 예측할 수 있는 바이오마커의 탐색방법을 제공하고자 한다.
- [0011] 본원이 해결하고자 하는 과제는 이에 제한되지 않으며, 통상의 기술자가 이해할 수 있는 범위 내의 과제는 모두 포함하는 것으로 해석되어야 한다.

과제의 해결 수단

- [0012] 상기 과제를 해결하기 위해, 본원의 일 측면은 머신 러닝을 이용하여 암환자의 면역 항암제에 대한 반응 유무를 판별하는 장치에 있어서, 유전자 네트워크 중 면역 항암제의 표적을 포함하는 대상 생물학적 경로(표적과 기능적으로 연관됨)를 추출하는 생물학적 경로 추출부, 상기 면역 항암제를 이용하여 면역 요법을 수행할 대상 암환자의 전사체 데이터(transcriptome data)로부터 유전자 활성 정보를 상기 대상 생물학적 경로의 활성 정보로 변환하는 유전자 활성 정보 변환부 및 미리 학습된 면역 항암제 반응 판별 모델에 상기 대상 유전자 정보를 입력하여 상기 대상 암환자의 상기 면역 항암제에 대한 반응 유무를 판별하는 판별부를 포함하는 장치를 제공한다.
- [0013] 본원의 다른 일 측면은 머신 러닝을 이용하여 암환자의 면역 항암제에 대한 반응 유무를 판별하는 방법에 있어서, 유전자 네트워크 중 면역 항암제의 표적을 포함하는 대상 생물학적 경로를 추출하는 단계, 상기 면역 항암제를 이용하여 면역 요법을 수행할 대상 암환자의 전사체 데이터(transcriptome data)로부터 유전자 활성 정보를 상기 대상 생물학적 경로의 활성 정보로 변환하는 단계 및 미리 학습된 면역 항암제 반응 판별 모델에 상기 대상 유전자 정보를 입력하여 상기 대상 암환자의 상기 면역 항암제에 대한 반응 유무를 판별하는 단계를 포함하는 방법을 제공한다.
- [0014] 본원의 과제 해결 수단은 상기한 바에만 제한되지 않으며, 본원 기술분야에 속하는 통상의 기술자가 이해할 수 있는 범위의 모든 수단을 포함하는 것으로 해석되어야 한다.

발명의 효과

- [0015] 본원의 장치 및 방법을 활용하면 암 환자에 대한 ICI 치료의 효과와 환자의 전체 생존률을 정확하게 예측할 수 있는 바이오마커를 찾아낼 수 있으며, 이에 따라 ICI 치료 효과를 극대화할 수 있다.

도면의 간단한 설명

- [0016] 도 1은 본원에 따른 장치를 나타낸 모식도이다.
- 도 2는 본원에 따른 알고리즘의 전반적인 과정을 나타낸 도면이다.
- 도 3는 NetBio 기반 예측과 합성 치사 기반 예측(synthetic lethality-based prediction; SELECT score)을 결합했을 때의 예측 성능을 점수로 환산하여 나타낸 것이다.
- 도 4은 네트워크 기반 머신러닝을 사용하여 면역요법에 관련된 바이오마커를 탐색하는 과정을 나타낸 도면이다.
- 도 5는 네 가지 코호트에서 면역요법을 받은 환자들에 대한 약물 반응 및 전체 생존률 예측 성능을 나타낸 도면이다.
- 도 6는 Monte Carlo 교차 검정을 활용하여 작은 규모의 학습 샘플에 대한 예측 성능을 나타낸 도면이다.

- 도 7은 세 가지 흑색종 데이터세트에 대한 예측 성능을 나타낸 도면이다.
- 도 8은 학습에 사용되지 않은 독립적 흑색종(external melanoma) 데이터세트에 대한 면역 치료 반응 예측 성능을 나타낸 도면이다.
- 도 9은 8 종류의 바이오마커로 22가지 예측 성능 확인 실험을 진행한 결과를 요약한 도면이다.
- 도 10는 유전자 네트워크를 활용(NetBio) 및 비활용(ML-based feature selection)했을 때의 예측 성능을 나타낸 도면이다.
- 도 11은 NetBio 기반 예측으로 종양미세환경의 면역학적 특징을 분석한 도면이다.
- 도 12은 TCGA 코호트에서 NetBio 기반 예측과 면역 특징(immunogenic features) 간의 상관관계를 나타낸 도면이다.
- 도 13는 양적 특징 중요도(positive feature importance) 상위 10개의 면역 특징을 나타낸 도면이다.
- 도 14은 음적 특징 중요도(negative feature importance) 상위 10개의 면역 특징을 나타낸 도면이다.
- 도 15는 NetBio 경로(유사분열 G2기-G2-M기)의 발현 수준이, TCGA 위암의 여포성 보조 T 세포(follicular helper T cell) 비중과 양적 상관 관계에 있음을 나타낸 도면이다.
- 도 16는 NetBio 경로('케모카인 수용체와 케모카인의 결합' 및 'FcγR 활성화')의 발현 수준이, 백혈구 비중과 양적 상관 관계에 있음을 나타낸 도면이다.
- 도 17은 NetBio 경로의 발현 수준이 방광암의 면역조직화학 기반 면역 표현형과 일치함을 나타낸 도면이다.
- 도 18은 네트워크 기반의 전사체 특징과 TMB(tumor mutation burden)을 결합했을 때 PD-L1 억제제(Atezolizumab)를 투여한 환자의 전체 생존률에 대한 예측 성능이 향상됨을 나타낸 도면이다.
- 도 19은 TMB 기반 PD-L1 반응 예측 및, TMB와 NetBio 기반 예측을 비교한 도면이다.
- 도 20는 IMvigor210 데이터세트에서 예상 ICI 반응자 및 비반응자에 대한 TMB 수준을 나타낸 도면이다.
- 도 21은 본원에 따른 알고리즘을 순서도로 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0017] 아래에서는 첨부한 도면을 참조하여 본원이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본원의 실시예를 상세히 설명한다. 그러나 본원은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본원을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.
- [0018] 본원 명세서 전체에서, 어떤 부재가 다른 부재 "상에" 위치하고 있다고 할 때, 이는 어떤 부재가 다른 부재에 접해 있는 경우뿐 아니라 두 부재 사이에 또 다른 부재가 존재하는 경우도 포함한다.
- [0019] 본원 명세서 전체에서, 어떤 부분이 어떤 구성 요소를 "포함" 한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성 요소를 제외하는 것이 아니라 다른 구성 요소를 더 포함할 수 있는 것을 의미한다.
- [0020] 본원 명세서 전체에서 사용하는 정도의 용어 "약", "실질적으로" 등은 언급된 의미에 고유한 제조 및 물질 허용 오차가 제시될 때 그 수치에서 또는 그 수치에 근접한 의미로 사용되고, 본원의 이해를 돕기 위해 정확하거나 절대적인 수치가 언급된 개시 내용을 비양심적인 침해자가 부당하게 이용하는 것을 방지하기 위해 사용된다. 본원 명세서 전체에서 사용하는 정도의 용어 "~(하는) 단계" 또는 "~의 단계"는 "~를 위한 단계"를 의미하지 않는다.
- [0021] 본원 명세서 전체에서, 마쿠시 형식의 표현에 포함된 "이들의 조합(들)"의 용어는 마쿠시 형식의 표현에 기재된 구성 요소들로 이루어진 군에서 선택되는 하나 이상의 혼합 또는 조합을 의미하는 것으로서, 상기 구성 요소들로 이루어진 군에서 선택되는 하나 이상을 포함하는 것을 의미한다.
- [0022] 본원 명세서 전체에서, "A 및/또는 B"의 기재는 "A 또는 B, 또는 A 및 B"를 의미한다.
- [0024] 본원 명세서 전체에서, 유전자 네트워크란 체내 유전자 간의 다양한 유전자 상호작용을 포함하는 용어이다. 예를 들어, 유전자 네트워크는 단백질-단백질 상호작용(Protein-Protein Interaction) 네트워크일 수 있다.

- [0025] 유전자 상호작용은 염색체 상에서 물리적 근접성, 진화 과정에서의 공존성, 발현량의 유사성, 발현 단백질의 물리적 결합성, 질병 등 표현형에 대한 좌위이질성 등을 포함한다. 유전자는 개체의 형태학적, 생리학적 특징을 결정하므로 생물의 건강상태와 큰 관련이 있다. 따라서 유전자 간의 상호작용에 대한 연구는 질병이나 약물에 대한 반응 등 개체의 표현형에 대해 복수의 유전자가 종합적으로 어떤 역할을 하는지 알아낼 수 있다는 점에서 중요하다.
- [0027] 본 출원의 발명자들은 네트워크 기반의 머신러닝 기틀(a network-based machine learning framework)를 제공한다. 본원 발명은 (1) ICI 데이터셋에서 정확한 예측이 가능할 수 있고, 새로운 잠재적 바이오마커를 확인할 수 있다. 구체적으로 본 발명자들은 흑색종(melanoma), 방광암(bladder cancer), 그리고 위장암(gastroesophageal)을 앓고 있으며, PD1/PD-L1에 대한 ICI 치료를 받은 700명 이상의 환자샘플에서, 네트워크 기반 바이오마커의 발현 정도를 이용하여 정확한 반응자와 비반응자를 구별할 수 있었다. 구별을 위하여 본원의 네트워크 기반 탐색을 활용하였고, 그 결과 유전자 네트워크 내 면역요법 표적과 근접한 생물학적 반응 경로를 확인할 수 있었다.
- [0029] 본원의 제1측면은 머신 러닝을 이용하여 암환자의 면역 항암제에 대한 반응 유무를 판별하는 장치에 있어서, 유전자 네트워크 중 면역 항암제의 표적을 포함하는 대상 생물학적 경로를 추출하는 생물학적 경로 추출부, 상기 면역 항암제를 이용하여 면역 요법을 수행할 대상 암환자의 유전자 데이터(transcriptome data)로부터 상기 대상 생물학적 경로에 대응하는 대상 유전자 정보를 추출하는 유전자 활성 정보 변환부 및 미리 학습된 면역 항암제 반응 판별 모델에 상기 대상 유전자 정보를 입력하여 상기 대상 암환자의 상기 면역 항암제에 대한 반응 유무를 판별하는 판별부를 포함하는 장치를 제공한다.
- [0030] 상기 경로 추출부는 유전자 네트워크 준비 및 네트워크 기반 바이오마커 탐색의 과정을 포함할 수 있다.(도 1 참조)
- [0031] **유전자 네트워크 준비(Preparation of genomic network)**
- [0032] STRING database v.11.0.에서 인간 PPI network를 다운로드 받았다(<https://string-db.org/>). 높은 신뢰도의 PPIs를 사용하기 위해 사이트 내 700점 이상의 네트워크만을 사용했다. 네트워크 기반 분석을 위해 16,957개의 노드, 420,381개의 엣지를 포함하는 가장 큰 유전자 네트워크를 사용했다. 해당 네트워크는 NetworkX python module을 통해 계산했다. 네트워크의 시각화는 cytoscape(v.3.7.1)를 사용했다.
- [0034] **네트워크 기반 바이오마커 탐색(Network-based biomarker (NetBio) detection)**
- [0035] 탐색은 (1) 유전자 네트워크 내에서 ICI 표적에 근접한 유전자 탐색 (2) ICI 표적과 근접한 생물학적 경로 (Reactome pathway) 탐색의 두 단계로 수행했다. 먼저 ICI 표적에 가까운 유전자를, NetworkX python module의 personalized page-rank 알고리즘을 사용한 네트워크 전과를 통해 식별했다. ICI 표적에 1을, 다른 유전자에는 0을 할당하여 page-rank 알고리즘의 개별 매개변수에 입력했다. 기타 매개변수는 기본값을 사용했다. 네트워크 전과 후 상위 200개 유전자들을 ICI 표적에 근접한 유전자로 간주했다.
- [0036] 다음, ICI 표적에 근접한 유전자를 사용하여 ICI 표적에 근접한 생물학적 경로를 탐색했다. 이를 위해 ICI 표적 근접 유전자가 각 경로에 어느 정도 포함되어 있는지를 구체적으로 계산하는 gene set enrichment test를 진행했다. 마지막으로 ICI 표적 근접 유전자가 상당히 많이 포함된 경로를, 조정된 P-value가 0.01 이하인지를 기준으로 선택했다. scipy와 statsmodels python modules를 사용하여 각각 Hypergeometric test statistics를 계산하고 P-value를 조정했다.
- [0038] 상기 유전자 활성 정보 변환부는 환자 데이터 처리 과정을 포함할 수 있다.
- [0039] **환자 데이터 처리(curation and preprocessing of patient data)**
- [0040] 본 발명인은 서로 다른 7개의 환자 코호트에 대해 PD/PD-L1을 대상으로 하는 ICIs를 처리했다.
- [0041] (각 환자 코호트:
- [0042] (1) Gide *et al.* (Nivolumab, Pembrolizumab and/or Ipilimumab treated melanoma, n=91; Gide, T. N. *et al.* Distinct Immune Cell Populations Define Response to Anti-PD-1 Monotherapy and Anti-PD-1/Anti-CTLA-4 Combined Therapy. *Cancer Cell* 35, 238-255.e6 (2019))
- [0043] (2) Liu *et al.* (Nivolumab or Pembrolizumab treated melanoma, n=121; Liu, D. *et al.* Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic

melanoma. Nat. Med. 25, 1916-1927 (2019).)

- [0044] (3) Kim *et al.* (Pembrolizumab treated metastatic gastric cancer, n=45; Kim, S. T. *et al.* Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. Nat. Med. 24, 1449-1458 (2018).)
- [0045] (4) IMvigor210 (Atezolizumab treated bladder cancer, n=348; Mariathasan, S. *et al.* TGF β attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature (2018). doi:10.1038/nature25501)
- [0046] (5) Auslander *et al.* (anti-PD-1 and/or anti-CTLA4 treated melanoma, n=37; Auslander, N. *et al.* Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma. Nat. Med. (2018). doi:10.1038/s41591-018-0157-9)
- [0047] (6) Prat *et al.* (Nivolumab or Pembrolizumab treated melanoma, n=25; Prat, A. *et al.* Immune-Related Gene Expression Profiling After PD-1 Blockade in Non-Small Cell Lung Carcinoma, Head and Neck Squamous Cell Carcinoma, and Melanoma. Cancer Res. 77, 3540-3550 (2017).)
- [0048] (7) Riaz *et al.* (Nivolumab treated melanoma, n=49; Riaz, N. *et al.* Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. Cell 171, 934-949.e16 (2017).)
- [0049] 여기서 (6) 코호트는 흑색종 샘플만을, (7) 코호트는 약물 치료 전의 발현 샘플만을 사용했다.)
- [0050] 코호트 정보는 하기 표 1과 같다.

표 1

Dataset	Cancer type	Drug	Treatment type	Total samples	Responders	Nonresponders
Mariathasan (IMvigor210)	Bladder	Atezolizumab (PD-L1)	Pre	348	60	238
Liu	Melanoma	anti PD1 (Nivolumab, Pembrolizumab)	Pre	121	49	72
Riaz_pre	Melanoma	Nivolumab (PD1)	Pre	49	10	39
Gide	Melanoma	Ipilimumab (CTLA4), Nivolumab (PD1), Pembrolizumab (PD1)	Pre	91	49	42
Kim	Metastatic gastric cancer	anti PD1 (Nivolumab, Pembrolizumab)	Pre	45	12	33
Auslander	me	ICB	Pre / On	37	3	34
Prat_MELANOMA	me	ICB	Pre	25	9	16

- [0051]
- [0052] Pre: pre-treatment; 해당 샘플이 약물 치료 이전에 채취됨
- [0053] On: on-treatment; 해당 샘플이 약물 치료 이후에 채취됨
- [0055] TCGA(The Cancer Genome Atlas) 데이터세트는 (1) TCGA SKCM (melanoma, n=103), (2) TCGA STAD (stomach adenocarcinoma, n=375) and (3) TCGA BLCA (bladder cancer, n=405). Gene expression data (HTSeq - Counts), somatic mutation data and clinical data (i.e. overall survival data)를 TCGAbiolinks R package 를 사용하여 다운로드 받았다. TCGA 암 환자의 TMB(tumor mutation burden)을 계산하기 위해 Wang *et al.*의 아래 식을 차용했다.(Wang, X. & Li, M. Correlate tumor mutation burden with immune signatures in human cancers. BMC Immunol. (2019). doi:10.1186/s12865-018-0285-5)
- [0056] $TMB_{patient} = T_{patient} \times 2.0 + NT_{patient} \times 1.0$
- [0057] 여기서,
- [0058] $T_{patient}$: 유도 돌연변이(truncating mutations)
- [0059] $NT_{patient}$: 비유도 돌연변이(non-truncating mutations)
- [0061] 유도 돌연변이는 nonsense mutation, frame-shift deletion 또는 insertion, 그리고 splice-site mutation을

고려했다. 비유도 돌연변이는 missense mutation, in-frame deletion 또는 insertion, 그리고 nonstop mutation을 고려했다.

- [0062] 유전자 발현정보의 전처리는 유전자 발현수준(gene expression levels)를 계산하는 데 'M-values (TMM normalization from egeR R package'로서 정규화된 IMvigor210, Auslander, Prat, Riaz 및 TCGA datasets를 사용했다. 기타 데이터세트는 Lee *et al.*이 제공한 정규화 발현 값을 사용했다.(<https://zenodo.org/record/4661265>) 유전자 경로(pathway) 발현 수준의 계산에 MsigDB database에서 다운로드 받은 Reactome pathways를 사용했으며(Lee, J. S. *et al.* Synthetic lethality-mediated precision oncology via the tumor transcriptome. Cell (2021). doi:10.1016/j.cell.2021.03.030), GSVA R package를 사용한 single sample GSEA(ssGSEA)를 수행했다.(Hönzelmann, S., Castelo, R. & Guinney, J. GSVA: Gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics (2013). doi:10.1186/1471-2105-14-7) 각 샘플의 경로(pathway) 발현 수준 측정에는 normalized enrichment score(NES)를 사용했다.
- [0063] 샘플을 반응자(responder)와 비반응자(non-responder)로 구분하기 위해 response evaluation criteria in solid tumors (RECIST) criteria를 사용했으며, 여기서 Complete Response(CR)과 Partial Response(PR)은 반응자, Stable Disease(SD)와 Progressive Disease(PD)는 비반응자로 분류했다. RECIST criteria를 사용하지 않거나, 또는 이를 제공하지 않는 데이터세트에 대해서는 각 데이터세트가 제시하는 반응자/비반응자 분류를 사용했다.
- [0065] 상기 관별부는 머신러닝 예측 검정 및, NetBio 기반 예측과 SELECT(synthetic lethal relation) 기반 예측을 결합하여 사용한 모델 활동 검정 과정을 포함할 수 있다.
- [0066] **머신러닝 예측 검정(Measuring performances of machine-learning (ML) predictions)**
- [0067] 머신러닝(ML) 모델을 학습하기 위해 scikit-learn python module을 사용하여 로지스틱 회귀 분석(logistic regression)을 적용했다. 구체적으로 본 출원인은 12개의 정규화 로지스틱 회귀 분석 모델을 사용했다. ML 모델 학습에는 약물 반응(반응자/비반응자로 구분)에 대한 유전자/경로의 발현 수준을 사용했다. 적절한 초 매개변수를 선택하기 위해 0.1부터 1까지 0.1의 간격 동안 정규화 매개 변수(C)를 반복하여 학습 데이터 세트에서 5배 교차 검증을 수행하였다. 클래스 불균형 효과를 줄이기 위해 클래스 가중치 초 매개변수에 '균형' 매개변수를 사용했다. 최적의 초 매개변수를 식별하기 위해 Scikit-learn 모듈의 GridSearchCV 기능을 사용했다. 유전자/경로 발현 수준은 코호트 사이의 배치 효과(batch effect)를 최소화하기 위해 ML 훈련/테스트 전에 z-score 표준화되었다.
- [0068] Leave-one-out cross validation(LOOCV)에서 다음의 기준을 만족하는 코호트를 고려했다: (1) 30개 이상의 샘플 존재 (2) 반응자/비반응자 모두에 각각 최소 10개의 샘플 존재. 그 결과 위 기준에 맞는 데이터세트 4개를 선정했다(Gide, Liu, Kim 및 Imvigor210). Scikit-learn 모듈의 LeaveOneOut 기능을 사용하여 학습/테스트 데이터세트를 분리했다.
- [0069] 유전자 기반 바이오마커(GeneBio)와 종양미세환경 기반 바이오마커(tumor microenvironment-based biomarkers: TME-Bio)에 기반한 예측을 위해 유전자 발현 수준을 이용해 머신러닝 모델을 학습/테스트했다. GeneBio는 PD1, PD-L1 또는 CTLA4의 발현 수준을 사용했다. TME-Bio는 (1) CD8 T cells, (2) T cell exhaustion (3) cancer-associated fibroblasts (4) tumor-associated macrophages(M2 macrophage)의 발현 수준에 대한 마커를 사용했다.
- [0070] 데이터 중심의 머신러닝의 예측성능을 테스트하기 위해 Scikit-learn 키트의 SelectKBest 기능을 이용한 feature selection을 수행했다('f_classif'를 점수 기능 매개변수로서 사용함). Reactome의 K(NetBio 경로의 수) 숫자를 선택했다. 데이터 중심의 머신러닝 모델을 학습과 테스트에 경로 발현 수준을 사용했다.
- [0072] **NetBio 기반 예측과 SELECT(synthetic lethal relation) 기반 예측을 결합하여 사용한 모델 활동 검정 (Calculating prediction performances for the combined model using NetBio-based predictions and predictions from synthetic lethal relationship (SELECT))**
- [0073] 개별 연락을 통해 원작자로부터 SELECT 점수를 제공받았다. SELECT는 ICI 처리되지 않은 암 샘플에서 발견된 두 유전자 사이의 합성치사(synthetic lethal)과 합성구조(synthetic rescue)를 이용한다. SELECT 점수와 NetBio 기반 예측(LOOCV의 예측 확률을 활용)을 결합하기 전에, 먼저 두 예측 점수 간 스피어먼 상관계수(spearman correlation)를 계산했다. Kim *et al.* 코호트(전이성 위암)에서 두 예측 점수는 어떠한 상관관계도 보이지 않았고(spearman correlation rho = 0.28; P-value = 0.16; 도 3의 b 참조) 이는 곧 서로 다른 두 예측 모델이 별

개의 생물학적 신호를 측정했음을 의미한다.

[0074] SELECT 점수와 NetBio 기반 예측의 결합을 위해 본 출원인은 Zhang *et al.*의 선형가중모델을 사용했다(Zhang, N. *et al.* Predicting Anticancer Drug Responses Using a Dual-Layer Integrated Cell Line-Drug Network Model. *PLoS Comput. Biol.* (2015). doi:10.1371/journal.pcbi.1004498):

[0075] $Combined\ score = w \times (NetBio\text{-based\ predictions}) + (1-w) \times (SELECT\ score)$

[0076] 여기서, W는 0부터 1까지 0.1의 간격을 둔 선형 가중치를 의미한다.

[0077] 수신기 작동 특성 곡선의 곡선 아래 영역(area under the curve; AUC)을 성능 지표로 사용했다.

[0079] 본원의 제2측면은 머신 러닝을 이용하여 암환자의 면역 항암제에 대한 반응 유무를 판별하는 방법에 있어서, 유전자 네트워크 중 면역 항암제의 표적을 포함하는 대상 생물학적 경로를 추출하는 단계, 상기 면역 항암제를 이용하여 면역 요법을 수행할 대상 암환자의 유전자 데이터(transcriptome data)로부터 상기 대상 생물학적 경로에 대응하는 대상 유전자 정보를 추출하는 단계 및 미리 학습된 면역 항암제 반응 판별 모델에 상기 대상 유전자 정보를 입력하여 상기 대상 암환자의 상기 면역 항암제에 대한 반응 유무를 판별하는 단계를 포함하는 방법을 제공한다.

[0080] 상기 제1측면에서 제2측면에 공통된 부분은 제 2측면에도 공히 적용된다.

[0081] 전반적인 알고리즘의 단계는 도 2에 도시했다.

[0083] 이하, 첨부된 도면을 참조하여 본원의 구현에 및 실시예를 상세히 설명한다. 그러나, 본원이 이러한 구현에 및 실시예 도면에 제한되지 않을 수 있다.

[0085] **실시예 1. 데이터 전처리 및 머신러닝 모델 학습 진행**

[0086] 16,957개의 노드와 420,381개의 엣지로 이루어진 STRING 유전자 네트워크를 사용했다. 먼저 ICI 표적(PD1-Nivolumab / PD-L1-Atezolizumab)을 시드 유전자(seed gene)로 활용하여 ICI 표적의 영향력을 네트워크 전체에 전파했다.(도 4의 a 참조) 네트워크 전파의 한가지 특징은 ICI 표적에 가까운 노드일수록 영향력 점수(influence score)가 높아는 것이다. 다음으로 영향력 점수로 상위 200개의 유전자를 선정하고 유전자가 풍부한 생물학적 경로(Reactome pathways)를 선별했다.(도 4의 b 참조)선별된 경로로 면역요법 반응을 예측했고, 이러한 경로들을 네트워크 기반 바이오마커(Network-based Biomarkers; NetBio)로 간주했다.

[0087] 머신러닝 기반 면역요법 반응예측에 있어 NetBio를 입력 특징으로, 그리고 면역요법 표적 유전자 등의 유전자 기반 바이오마커(gene-based biomarkers; GeneBio), 종양 미세환경 기반 바이오마커(tumor microenvironment-based biomarkers; TME-Bio) 또는 데이터 중심의 머신러닝 접근으로부터 선택된 경로를 음성 대조군으로 사용했다.(도 4의c 참조) 입력 특징의 발현 수준을 사용해 머신러닝 모델을 로지스틱 회귀분석으로 학습시켰다. 입력 특징의 예측 성능을 테스트하기 위해 (1) 면역요법 치료 후 종양 크기 감소로 측정된 약물 반응 또는 (2) 환자의 전체 생존률에 대한 예측 성능을 확인했다. 머신러닝 모델의 지도 학습(supervised learning)을 위해 서로 다른 학습 및 테스트 데이터셋을 사용하여 예측 성능의 일관성을 측정했다. 구체적으로 (1) 학습 및 테스트 데이터셋이 단일 코호트에서 유래되는 예측 학습 또는 (2) 두 가지 독립된 학습 및 테스트 데이터셋 간 교차 예측 학습을 진행했다. 또한 다양한 모델 학습 상황을 가정하기 위하여 학습 샘플의 크기가 큰 경우와 작은 경우를 번갈아가며 학습을 진행했다.

[0089] **실시예 2. 교차 검정을 통한 NetBio 기반 머신러닝의 예측 성능 증명**

[0090] 본 출원인은 NetBio의 전사체(transcriptome)이 ICI 반응에 대한 일관된 예측 성능을 가지고 있음을 확인했다. 반대로 약물 표적(PD1-Nivolumab / PD-L1-Atezolizumab / CTLA4-Ipilimumab)을 사용했을 경우 예측 성능이 떨어지는 것도 확인했다.

[0091] 성능 검증을 위해 먼저, NetBio와 약물 표적을 포함하여 종전에 알려진 면역요법 관련 바이오마커의 성능을 LOOCV(leave-one-out cross-validation)을 통해 측정했다. 여기서 두 흑색종 코호트(Gide *et al.*, Liu *et al.*), 전이성 위장암 코호트(Kim *et al.*) 그리고 방광암 코호트(Imvigor210)의 총 4가지 코호트를 사용했다. 결론적으로 NetBio를 사용한 학습 결과 모든 코호트에서 일관되고 정확한 예측이 가능했다.(도 5의 a 내지 d 참조; 피셔정확검정 P-value < 0.05 기준) 반대로 약물 표적을 사용한 학습 결과는 일관된 예측이 불가능했고, 오직 흑색종 코호트(Gide *et al.*)에서만 통계적으로 유의한 결과를 보였다. 게다가 약물 표적의 발현 수준을 사용한 예측은 Liu 데이터셋에서 정반대의 결과(inversely predictive)를 보였다.

[0092] 게다가 전체 생존률 증가도 NetBio 기반 머신러닝의 세 데이터세트에서 반응자로 예측된 환자들에서 확인되었다.(Gide, Kim 및 Imvigor210; log-rank 테스트 P-value < 0.05 기준) 약물 표적 발현 기반 머신러닝은 단 하나의 데이터세트에서 전체 생존률 증가가 확인되었다. (도 5의 e 내지 g 참조) 정리하자면 네트워크 기반 바이오마커 탐색은 약물 표적 기반의 탐색에 비해 예측 성능이 증가했음을 확인했다.

[0093] 다음으로 NetBio의 예측성능을 종전에 확인된 ICI 관련 바이오마커인 GeneBio 또는 TME-Bio 등의 예측성능과 비교했고, 상기 네 가지 암 데이터세트 모두에서 동등하거나 그 이상의 결과를 확인했다. GeneBio에 대해 면역요법 표적(PD1, PD-L1 또는 CTLA4)의 발현 수준을 고려했고, TME-Bio에 대해 CD8 T 세포 비중, T 세포 소진(T cell exhaustion), 암 관련 섬유아세포(cancer associated fibroblasts; CAF) 그리고 종양 관련 대식세포(tumor associated macrophages; TAM)을 고려했다. LOOCV의 예측 성능 측정에 정확도와 F1 점수를 사용했으며, 그 결과 NetBio 기반 예측이 다른 모든 바이오마커에 비해 56개의 비교예 중 55개(98.2%)에서 더 뛰어났음을 확인했다.(도 5의 h 내지 o 참조)

[0094] 그리고 규모가 작은 학습 데이터세트를 ML 모델 학습에 사용했을 때, NetBio 기반 예측이 다른 바이오마커들 기반 예측과 동등하거나 그 이상의 성능을 보였다. 구체적으로, 임의적으로 학습과 테스트 세트를 8:2 비율로 나누어 100번 반복하여 monte carlo cross 검정을 실시하였다.(도 6의 a 참조) 그 결과 56개의 비교예 중 52개(92.9%)에서 네트워크 기반 예측이 GeneBio 또는 TME-Bio 기반 예측보다 동등하거나 그 이상의 성능을 보임을 확인했다.(도 6의 b 내지 j 참조; two-sided Student t-test P-value < 0.05) 따라서 본원의 알고리즘을 활용했을 때 다른 바이오마커들을 사용했을 때보다 더 정확한 ICI 반응 예측이 가능하다.

[0096] **실시예 3. 또 다른 흑색종 데이터세트에서의 Netbio 기반 예측 성능 실험**

[0097] 정확한 ML 모델의 주요 측면은 (i) 새로운 데이터 세트로 일반화하는 능력 및 (ii) 제한된 수의 훈련 샘플을 사용할 수 있는 경우에도 일관된 성능이다. 먼저 NetBio를 사용하여 훈련된 ML 모델은 독립적인 데이터 세트를 사용할 때 강력한 예측을 할 수 있는 반면 GeneBio 또는 TME-Bio를 사용하여 훈련된 ML 모델은 약물 반응을 덜 예측할 수 있음을 확인했다.(도 7 참조) ML 모델의 일반화 가능성을 테스트하기 위해 Gide *et al.* 흑색종 데이터 세트를 사용하여 ML 모델을 훈련하고 3개의 독립적인 흑색종 데이터 세트에서 예측 성능을 테스트했다(도 7의 a 참조 ; Auslander *et al.*, Prat *et al.*, 및 Riaz *et al.*). 본원 발명에 따른 모델 성능을 계산하기 위해 로지스틱 회귀 모델의 예측 확률을 사용하여 약물 반응을 관찰했다. 수신기 작동 특성 곡선의 곡선 아래 영역(area under the curve; AUC)을 성능 지표로 사용했다. NetBio 기반 ML이 두 개의 외부 데이터 세트(도 7의 b 및 c 참조, Auslander AUC = 0.79, Prat AUC = 0.72)에서 0.7보다 크고 나머지 데이터 세트(도 7의 d 참조; Riaz)에서 0.69보다 큰 AUC를 보여주었다는 것을 확인했다. NetBio 기반 ML과 달리 GeneBio 또는 TME-Bio 기반 예측은 매우 다양한 예측 성능을 보여주었다.(도 7의 b 내지 d 참조) 예를 들어 PD1 발현은 최대 AUC가 0.66에 불과하여 최적의 성능을 나타내지 못했다. 또한, T 세포 소진의 마커를 사용한 예측은 Auslander 및 Riaz 데이터 세트에서 매우 정확했지만(AUC > 0.7), 예측 성능은 Prat 데이터 세트(도 7의 c, AUC = 0.58)의 무작위 예상보다 약간 더 나은 정도였다.

[0098] 다음으로, 더 적은 수의 훈련 샘플을 사용할 수 있는 경우에도 ML 모델이 강력한 예측을 할 수 있는지 테스트했다. 다시 한 번, 샘플 크기가 더 작은 NetBio 기반 ML이 GeneBio 또는 TME-Bio 기반 ML 모델과 비교하여 일관된 예측을 할 수 있음을 확인했다. 이를 테스트하기 위해 100회 반복 동안 훈련 데이터 세트(Gide 데이터 세트)에서 환자의 80%를 무작위로 샘플링하여 ML 모델을 훈련하고 3개의 외부 흑색종 데이터 세트에서 예측 성능을 테스트했다.(도 8의 a 참조) 본원 발명에 따른 바이오마커가 21개 비교 중 18개에서 통계적으로 유의하게 더 좋거나 동일한 성능을 보이는 것을 확인했다.(도 8의 b 참조; 85.7%) Auslander 데이터 세트의 PD-L1 발현, Riaz 데이터 세트의 CTLA4 및 Riaz 데이터 세트의 CD8 T 세포 소진 마커만이 NetBio 기반 예측보다 더 나은 예측 성능을 나타냈지만 이러한 바이오마커(PD-L1, CTLA4 및 CD8 T 소진 마커)는 다른 흑색종 데이터 세트의 예측과 일치하지 않았다.(도 8의 b 내지 e 참조).

[0100] **실시예 4. NetBio 기반 예측과 BeneBio 또는 TME-Bio 기반 예측의 전반적 성능 비교**

[0101] 전반적으로 NetBio 기반 ML 모델이 암 환자의 ICI 반응을 정확하게 예측하는 데 강력하다는 것을 확인했다.(도 9 참조) 본원에서 수행한 22개의 서로 다른 테스트에서 NetBio가 154개 비교 중 143개(92.9%)에서 같거나 더 나은 성능을 보였으며, 전체 평균 예측 순위는 8개의 다른 바이오마커 중 1.5였다.(도 9의 d 참조) 이는 NetBio가 GeneBio 또는 TME-Bio 기반 예측에 비해 향상된 예측을 가능하게 함을 시사한다. CD8 T 세포 소진 및 CD8 T 세포의 마커가 다음으로 우수한 성능을 보였으며(각각 평균 순위 3.09 및 3.55), 이는 ICI가 암세포를 죽이기 위해 CD8 T 세포를 소생시키는 것을 목표로 한다는 점을 고려했을 때 예상된 결과이다. 실제로 종양 주변에 CD8 T

세포가 존재하는 것은 ICI 반응과 상관관계가 있으며, 자연적으로 T 세포를 포함한 종양(hot tumor)과 불포함한 종양(cold tumor)을 식별하는 연구가 임상적 유용성을 위해 활발히 진행되고 있다. 그럼에도 불구하고 CD8 T 세포 마커 또는 CD8 T 세포 소진 마커를 사용하여 만든 예측과 비교하여 NetBio는 각각 22개 테스트 중 20개 (90.9%) 또는 19개(86.3%) 테스트에서 동등하거나 더 나은 성능을 보였다.(도 9 참조) 더욱이 PD-L1 치료를 받은 방광암 환자에서 NetBio 기반 예측은 4가지 다른 예측 작업에서 일관되게 1위를 차지했지만, CD8 T 세포 고갈의 마커는 반응을 제대로 예측하지 못했다. 이 결과는 (1) 다른 암 유형에 대해 뚜렷한 면역 회피 메커니즘이 존재하고 (2) NetBio 기반 예측이 면역 요법 반응에 대해 정확한 예측을 할 수 있음을 시사한다.

[0103] **실시예 5. NetBio 기반 예측과 순수 데이터 기반의 특징 선택의 비교 실험**

[0104] 데이터 기반 ML 모델을 임상 적용에 사용할 때의 주요 제한 사항 중 하나는 훈련 데이터 세트에서 우수한 성능을 보임에도 불구하고 새로운 데이터 세트에서 일관되게 수행할 수 없다는 것이다. 따라서 본 발명에서 유전자 네트워크인 생물학적 사전 지식을 추가하는 것이 순수한 데이터 기반 특징 선택 접근 방식과 비교하여 특징 선택을 향상시킬 수 있는지 여부를 테스트했다. 실제로 NetBio 기반 ML 모델이 순수한 데이터 기반 ML 예측과 비교하여 지속적으로 개선된 예측 성능을 가능하게 한다는 것을 발견했다. 구체적으로, 데이터 기반 ML 모델의 경우 훈련 데이터 세트에서 응답자와 비응답자를 가장 잘 구별하는 K개의 기능(K: NetBio의 수)을 선택하고 선택된 기능을 사용하여 ML 모델을 훈련했다.(도 10의 a 참조) 11개의 다른 작업에서 NetBio 기반 예측은 ML 기반 특징 선택(feature selection)의 기능을 사용할 때와 비교하여 통계적으로 훨씬 더 잘 수행되었다.(도 10의 b 참조; two-sided paired Student t-test P -value = 3.3×10^{-3}) 또한 흑색종 코호트 전반에 걸쳐 예측할 때 일관된 성능 향상을 보였고,(도 10의 c 참조) 이는 네트워크 기반 선택이 ML 모델의 과적합을 줄이는 데 도움이 될 수 있음을 나타낸다. 이러한 결과는 네트워크 기반 기능 선택이 순수한 데이터 기반 기능 선택의 기능에 비해 강력한 기능을 제공할 수 있음을 시사한다. 즉, 본원의 네트워크 기반 바이오마커 선택을 활용하여 강력한 전사체 바이오마커를 발견할 수 있다.

[0106] **실시예 6. TCGA 데이터 세트에서 Netbio 기반 예측의 성능 확인**

[0107] NetBio는 세 가지 다른 암 유형을 포괄하는 별개의 코호트에서 가장 우수한 성능을 보였으므로 NetBio 기반 예측이 면역 요법 반응과 관련된 것으로 알려진 면역 미세 환경에도 적용될 수 있는지 여부를 실험했다. 이를 위해 NetBio 기반 예측이 The Cancer Genome Atlas(TCGA) 데이터 세트(도 11의 a 참조)에서 면역 상황과 어떻게 연관되는지 확인했다. 특히, Gide 또는 Liu 데이터 세트(흑색종 코호트)를 사용하여 TCGA 데이터 세트(TCGA SKCM)의 흑색종 환자들의 ICI 반응을, Kim 데이터 세트(위암 코호트)에서 TCGA 위암(TCGA STAD)의 ICI 반응을, 그리고 IMvigor210 데이터 세트(방광암)에서 방광암 환자의 ICI 반응을 예측하고, TCGA 방광암(TCGA BLCA) 환자를 예측했고, 예측된 약물 반응을 (i) TMB 또는 (ii) TCGA 환자의 면역 환경과 연관시켰다(도 11의 a 참조) 면역 환경의 경우 Thorsson *et al.*이 계산한 면역원성 점수를 사용했다.(Thorsson, V. *et al.* The Immune Landscape of Cancer. *Immunity* 48, 812-830.e14 (2018).) NetBio 기반 예측 대 TMB 또는 면역 맥락에 대한 전체 상관 관계 결과는 도 12에 도시했다.

[0108] NetBio 기반 예측은 면역 미세 환경을 성공적으로 설명할 수 있었다. Gide와 Liu 코호트의 상관 관계 결과는 둘 다 흑색종 환자이므로 공통된 특성을 가질 것으로 예상할 수 있다. 예상대로 두 코호트에서 백혈구 비중(leukocyte fraction) 및 CD8 T 세포 비율과의 높은 양의 상관관계 및 M2 대식세포 비율과의 높은 음의 상관관계를 포함하여 유사한 면역 미세환경 특성을 나타냈다.(도 11의 b 참조)

[0109] 면역 세포 비율과 높은 상관 관계를 나타내는 NetBio 경로를 추가로 조사했다. Gide 데이터 세트(도 13 참조)를 사용한 머신러닝 학습에서 가장 중요한 경로 특징(양의 상관관계가 있는 상위 10개의 특징)으로부터 '클래스 I MHC의 항원 제시 접힘 기작(antigen presentation folding assembly) 및 펩티드 로딩(peptide loading)'이 CD8 T 세포 비율과 가장 높은 양의 상관 관계를 나타냄을 확인했다.(도 11의 c 및 도 13 참조; $PCC = 0.41$). 이는 항원 제시 세포 또는 종양 세포에 의한 항원 제시가 CD8 T 세포의 침윤을 유도할 수 있는 것이 원인으로 보인다. Liu 데이터 세트를 사용할 때 가장 중요한 경로(음의 상관관계가 있는 상위 10개의 특징) 중 'FGFR 신호 전달'이 CD8 T 세포 비율과 가장 높은 상관 관계를 나타냄을 확인했으며,(도 14 참조) 여기서 경로 발현 수준은 세포 비율과 음의 상관관계가 있었다.(도 11의 c 참조; $PCC = -0.29$) 최근 연구에 따르면 섬유아세포 성장 인자 2(FGF2)의 고갈이 T 세포의 수를 증가시켜 종양 퇴행을 가능하게 한다는 결과가 보고 되었다. 따라서 (i) 동일하지 않은 CD8 T 세포 모집 메커니즘이 흑색종에 존재할 수 있고 (ii) NetBio는 다른 흑색종 암 집단이 ML 모델을 훈련하는 데 사용된 경우에도 종양 샘플에서 CD8 T 세포 모집을 강력하게 포착할 수 있을 것이다.

[0110] 본원에서는 위암 및 방광암에서 면역 미세 환경에 관련된 NetBio 경로를 확인했다. 위암에서 NetBio 기반 예측

은 여포 보조 T 세포 비율과 높은 상관 관계가 있었다.(도 11의 b 참조) Kim *et al.*의 코호트에서 가장 중요한 경로 중 우리는 '유사분열 G2-G2-M 단계'의 높은 발현 수준이 높은 여포 보조 T 세포 비율과 관련이 있었다.(도 13 및 도 15 참조) 이러한 실험 결과는 도우미 T 세포의 분화가 세포 주기 경로에 의해 조절된다는 이전 연구결과와 일치한다.

[0111] 방광암의 경우 NetBio 기반 예측이 백혈구 비중(leukocyte fractions)과 양의 상관관계가 있음을 확인했다.(도 11의 b 참조). NetBio 경로 또한 면역 침윤 기능과 깊은 관련이 있는 화학주성(chemotaxis; 케모카인 수용체와 케모카인의 결합 등) 및 식균작용(phagocytosis; FcγR의 활성화 등)을 나타냈다.(도 16의 a 및 b 참조; PCC > 0.6) 이러한 결과는 위암과 방광암에 대해 NetBio 경로를 사용하면, 면역 미세 환경까지도 다룰 수 있음을 보여 준다.

[0112] 추가적인 면역조직화학 기반 결과를 사용하여 화학주성 및 식균 작용 경로(예: 케모카인 수용체가 각각 케모카인 및 FcγR 활성화에 결합)가 PD-L1로 치료된 방광암 코호트에서 면역 침윤과 관련이 있음을 확인했다. 확인을 위해IMvigor210 데이터 세트의 면역 표현형을 사용했다. 구체적으로 (1) 10개 미만의 CD8 T 세포(immune desert), (2) 중앙 세포에 인접한 CD8 T 세포 및 (3) 중앙 세포와 접촉하는 CD8 T 세포의 면역 표현형을 사용했고,(도 17의 a 참조) 면역 표현형과, 화학주성 및 식균 작용 경로의 발현 수준을 비교하였다.(도 17의 b 및 c 참조) (3)의 하위 유형이 (1) 또는 (2)의 표현형과 비교했을 때 가장 높은 발현 수준을 보였으며(도 17의 b 및 c 참조; ANOVA P-value < 10⁻¹⁶), 이는 NetBio 경로가 방광암에 대해 백혈구 침투 분절(leukocyte infiltration fractions)을 포착할 수 있음을 의미한다.

[0113] 정리하자면 NetBio 경로는 면역요법 반응과 관련된 면역 미세 환경에 대한 경로를 일관되게 나타낼 수 있다.

[0115] **실시예 7. 기존 바이오마커와 NetBio의 결합**

[0116] 기존에 사용하던 바이오마커인 TMB(Tumor mutation burden)는 ICI 치료의 이점과 연관지어져 왔으나, TMB만으로는 ICI 반응을 충분히 예상할 수 없었다. 따라서 NetBio와 TMB 기반 예측을 결합하면 예측 성능이 향상되는지 확인하는 실험을 진행했다.(도 18의 a 참조) 그 결과 NetBio과 TMB의 발현 수준을 결합하면 PD-L1 억제제인 아테졸리주맙으로 치료받은 방광암 환자의 전체 생존률 예측 성능이 향상되었다.(도 18의 b 및 c 참조) LOOCV를 사용하여 ICI 치료 반응을 예측한 결과, TMB만 사용하여 ML 모델을 훈련한 경우, 예상 반응자 그룹과 예상 비반응자 그룹 간의 1년 생존률 차이는 18%였다.(도 18의 b 참조; log-rank test P-value = 2.0 x 10⁻³, 예상 반응자 및 예상 비반응자 그룹에 대한 1년 생존률은 각각 60.8% 및 42.8%임). 1년 생존률 차이는 TMB와 NetBio를 모두 사용한 경우 25.7%로 증가했으며(도 18의 c 참조; 예상 반응자 및 예상 비반응자 그룹의 1년 생존률은 각각 66.7% 및 40.9%임) 로그 순위 검정 통계도 향상된 결과를 보였다.(P-value = 2.84 x 10⁻⁵).

[0117] 다음으로, NetBio와 TMB가 결합된 예측기가, TMB를 단독으로 사용하여 분류한 예상 반응자(NR2R; 도 19 참조) 및 비반응자 그룹(R2NR; 도 19 참조)으로부터 각각 예상 비반응자와 예상 반응자를 올바르게 재분류하는 것을 확인했다. R2NR 환자의 전반적 1년 예상 생존률은 50%로 낮아졌다.(log-rank test P-value = 0.052) NR2R 환자의 전반적 1년 예상 생존률은 63%로 증가했고 이는 TMB 기반 예측에 의한 예상 비반응자 그룹의 전체 생존률과 비교했을 때 통계적으로 유의한 상승에 해당한다.(도 19의 c 참조; log-rank test P-value = 7.43 x 10⁻³) 즉 NetBio와 TMB를 결합했을 때 반응자와 비반응자의 정확한 분류가 가능했다.

[0118] 다음으로 위 결과에 기반하여, NetBio와 TMB를 결합했을 때 예측 성능이 향상되는 요인을 확인했다. 먼저 TMB 수준이 재분류된 하위 그룹에서 유사하게 유지되었으며(도 20 참조), 이는 TMB 수준이 예측 성능에 큰 영향을 주는 요인이 아님을 의미했다. 높은 TMB 수준을 보인 그룹에서 면역 요법에 대한 내성과 관련된 전사체 특징을 확인하기 위해 높은 TMB 수준을 보인 그룹과 R2NR 그룹에서 예측된 반응자 사이에서 차별적으로 발현된 경로는 Raf 활성화였다.(도 18d 참조, two-sided Student t-test P-value = 2.34 x 10⁻³). 구체적으로, 결합 예측 모델로부터 비반응자로 예측된 환자(R2NR 환자)는 raf 활성화 경로의 더 높은 발현을 나타내었다. 유전자 네트워크에서 HRAS, KRAS 및 JAK2를 포함한 raf 활성화 경로의 구성요소가 PD-L1의 직접적 연관이 있음 확인했으며(도 18e 참조), 이는 해당 경로들이 약물 치료에서 기계적 영향을 미칠 수 있음을 의미한다.

[0119] ICI 치료 바이오마커로서 raf 활성화 경로의 잠재적 유용성을 추가로 조사하기 위해 외부 TCGA 방광암 데이터 세트(n = 405)에서 PD-L1 발현, TMB 및 raf 활성화 발현 수준과 전체 생존률 간의 연관성을 분석했다. 구체적으로, (1) PD-L1이 낮아 PD-L1을 억제하는 경우 (2) TMB 수준이 높은 경우에 raf 활성화가 전체 생존률에 영향을 미치는지 확인했다. 그 결과 raf 활성화 경로는 낮은 PD-L1 발현 및 높은 TMB 수준을 나타내는 방광암 환자의

전체 생존률에 통계적으로 유의한 영향을 미치는 것을 확인했다.(도 18의 f 참조; P -value = 0.025). 특히 raf 활성화 경로의 더 높은 발현은 낮은 전체 생존률과 연관되었으며, 이는 치료에 대한 내성을 나타내는 PD-L1 억제제 치료 환자와 일치했다.(도 18의 d 및 f 참조). 정리하면, 상기 결과는 (1) 네트워크 기반 전사체 바이오마커가 TMB 기반 면역요법 반응 예측을 개선하는 데 도움이 될 수 있고 (2) 새로운 ICI 반응 바이오마커가 네트워크 기반 탐색에서 발견될 수 있음을 의미한다.

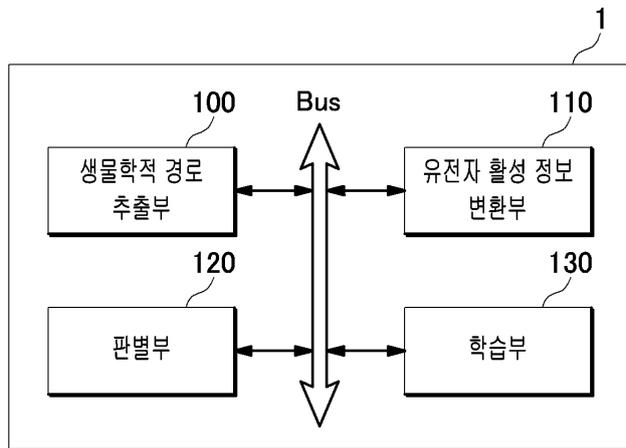
부호의 설명

[0120]

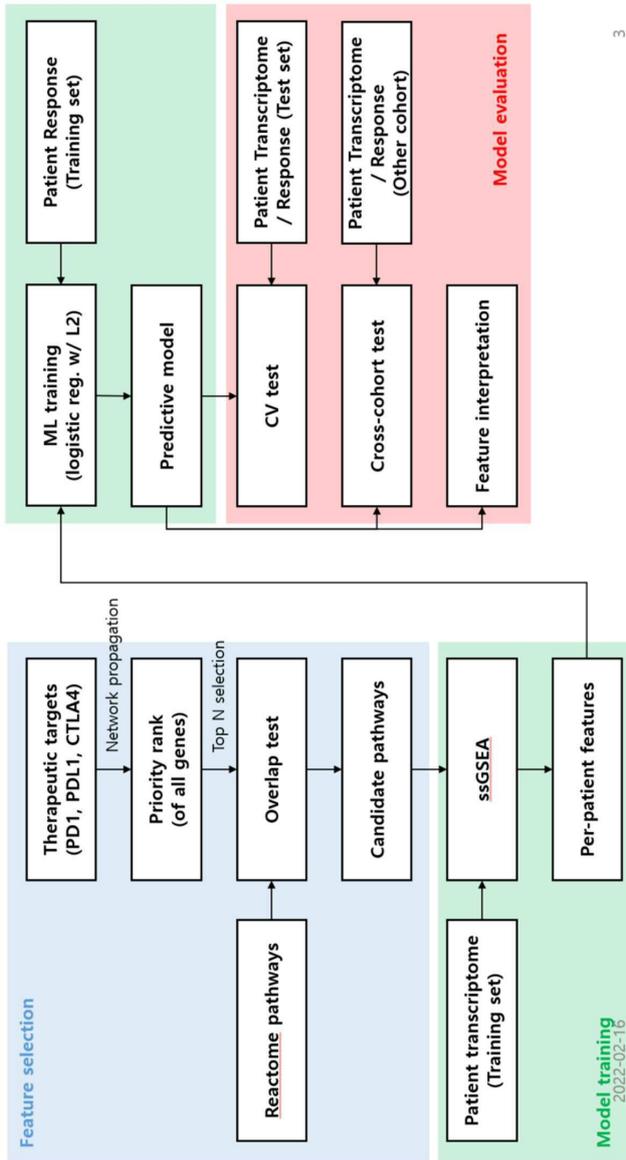
- 1: 면역 항암제 반응 유무 판별 장치
- 100: 생물학적 경로 추출부
- 110: 유전자 활성 정보 변환부
- 120: 판별부
- 130: 학습부

도면

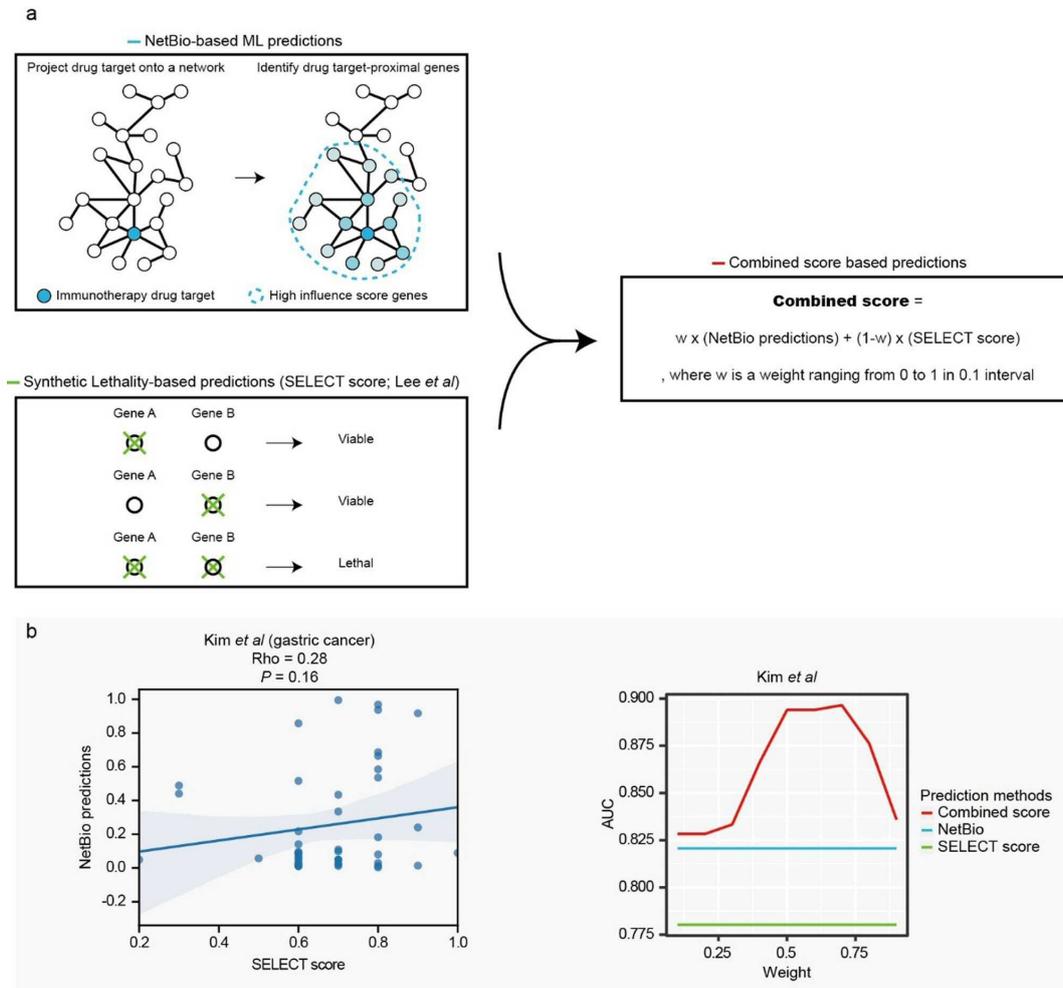
도면1



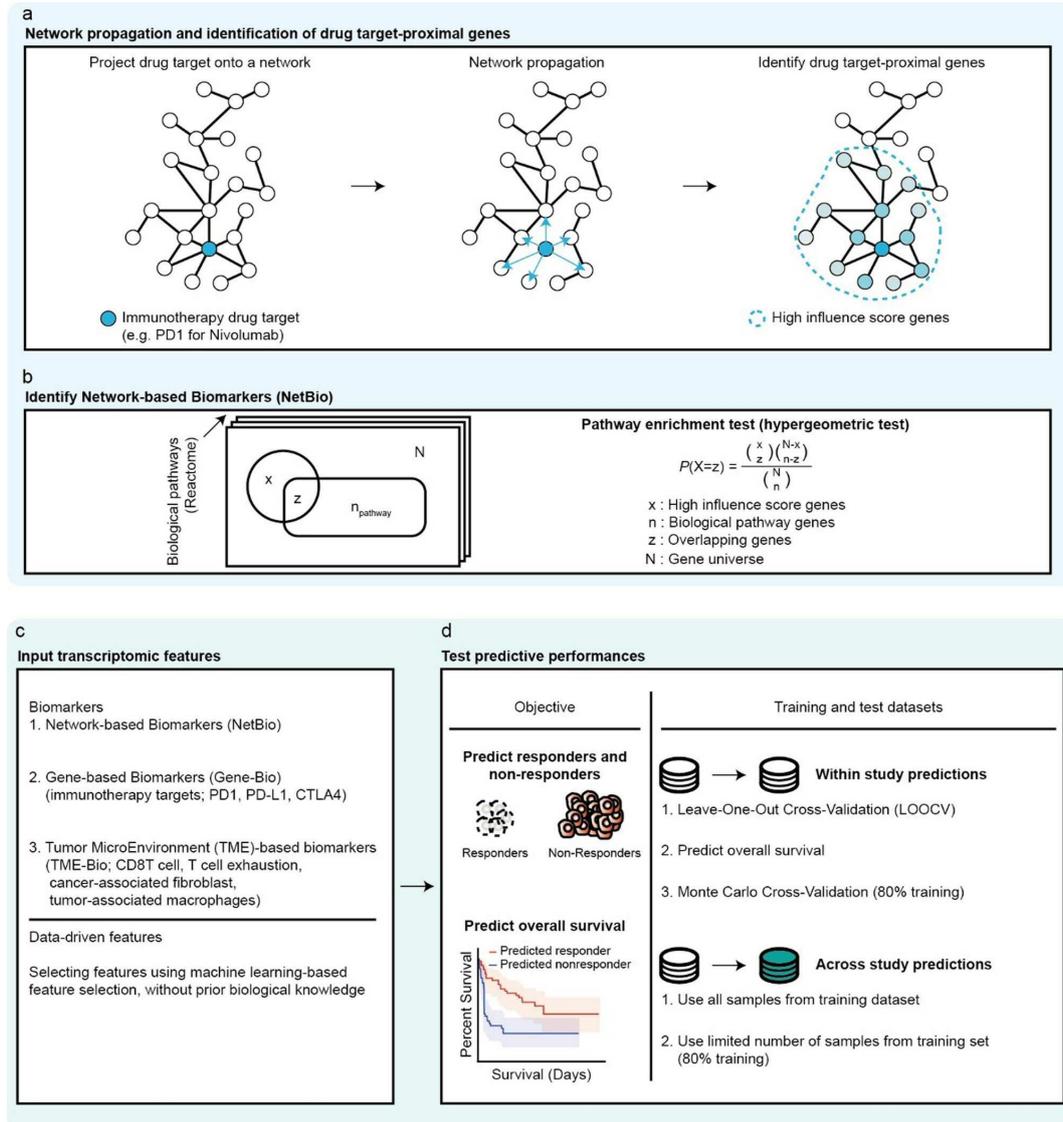
도면2



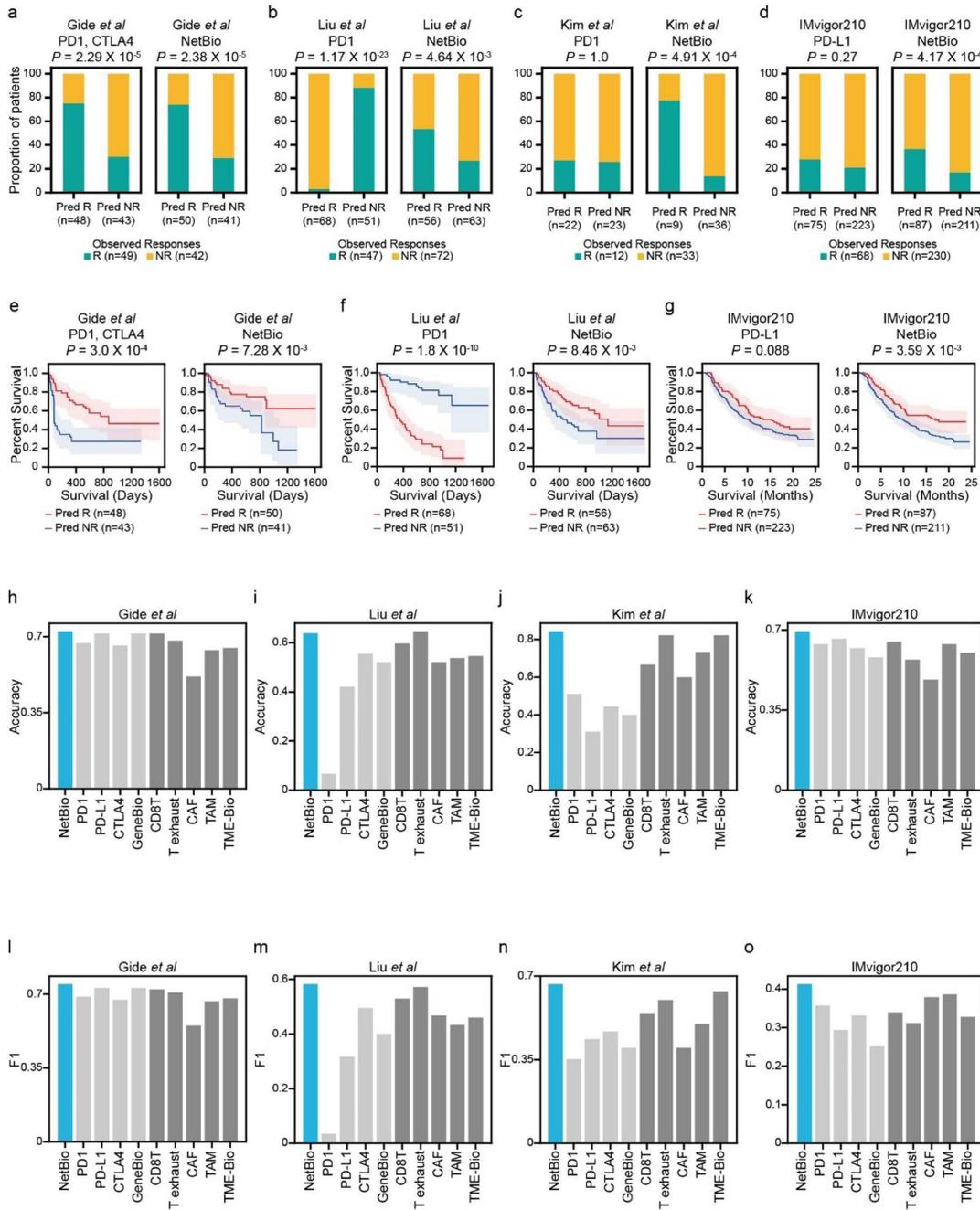
도면3



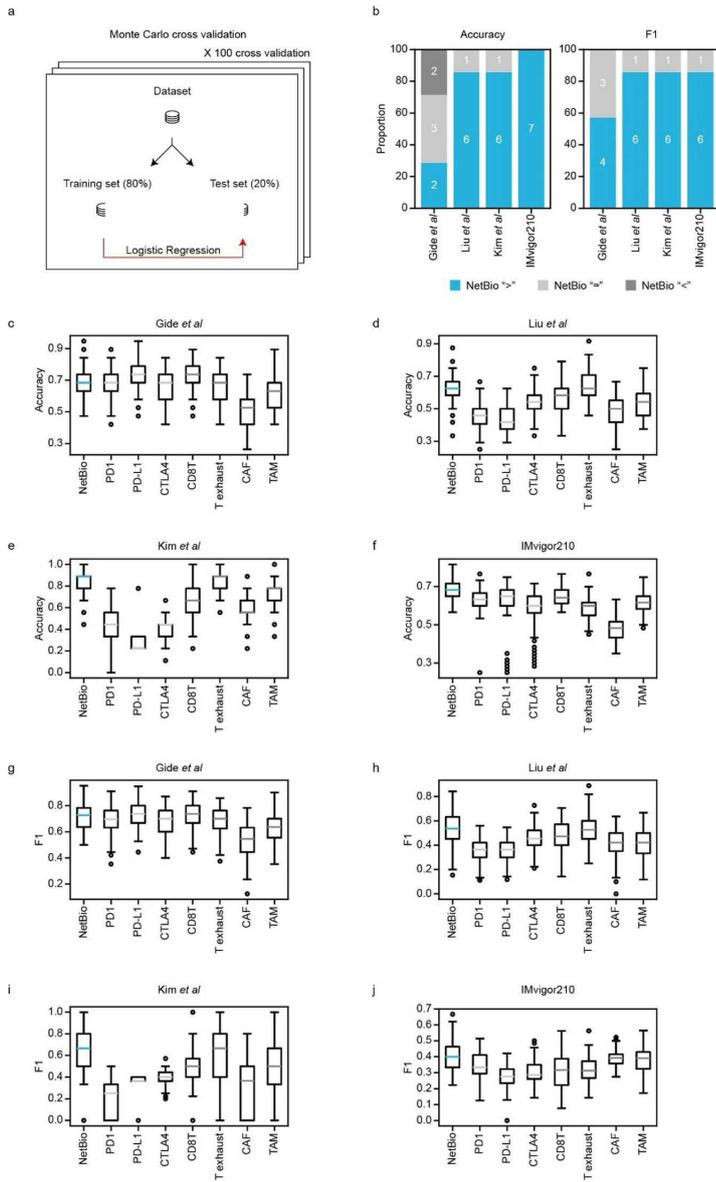
도면4



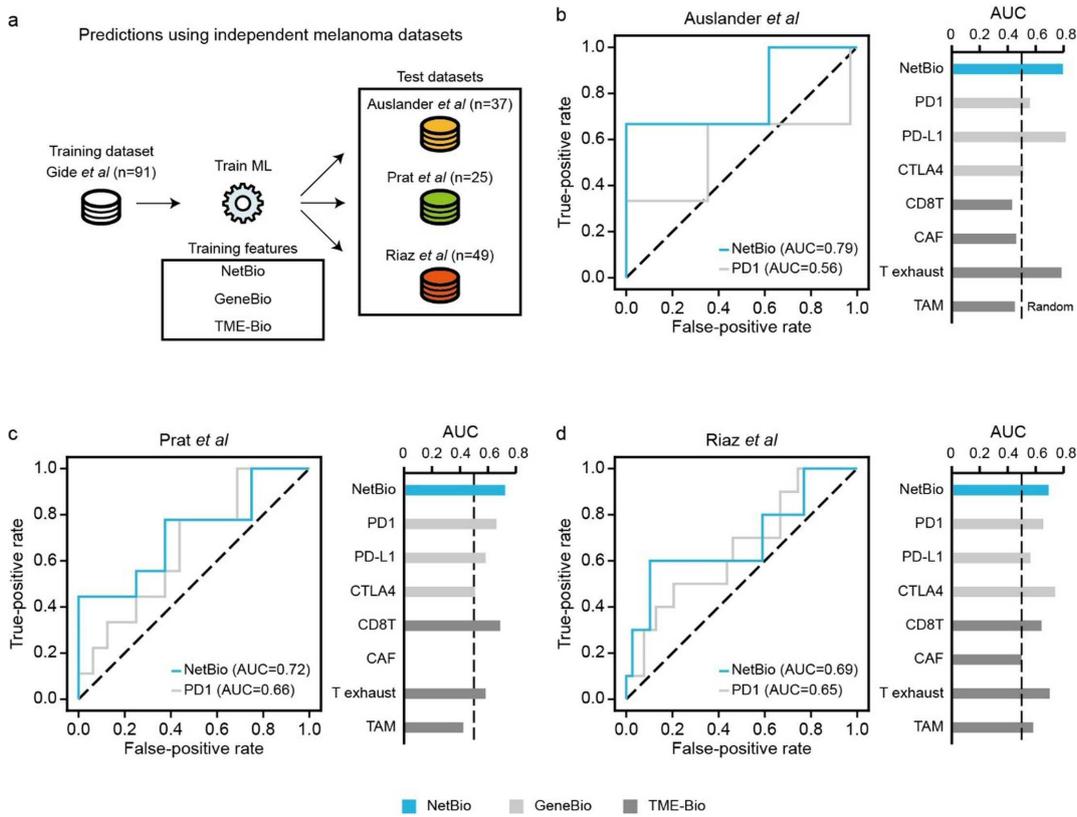
도면5



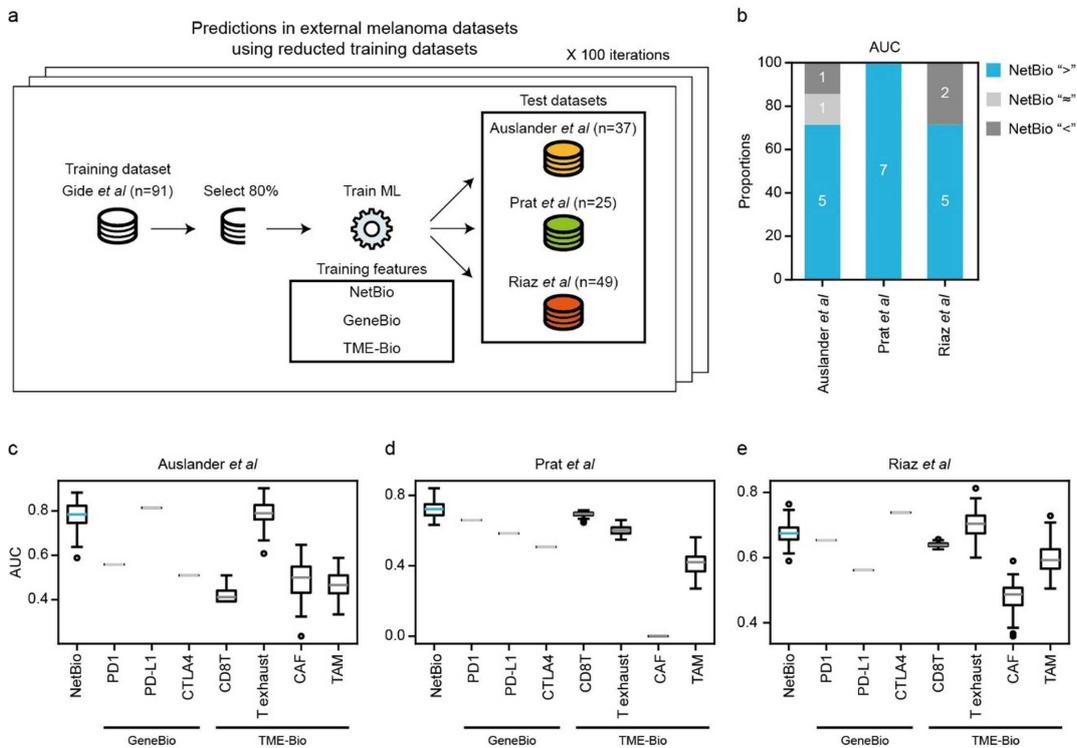
도면6



도면7

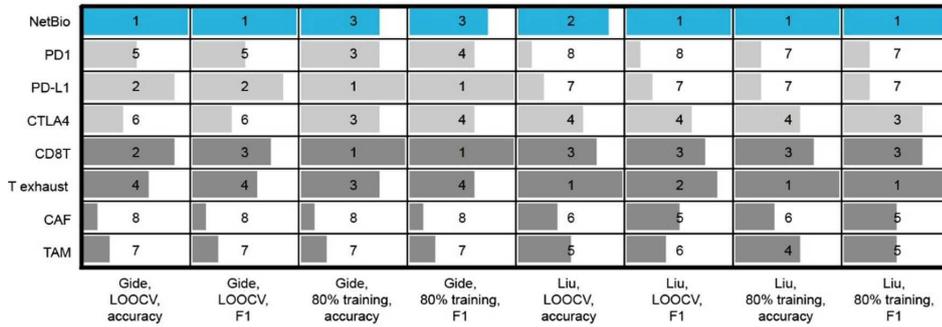


도면8

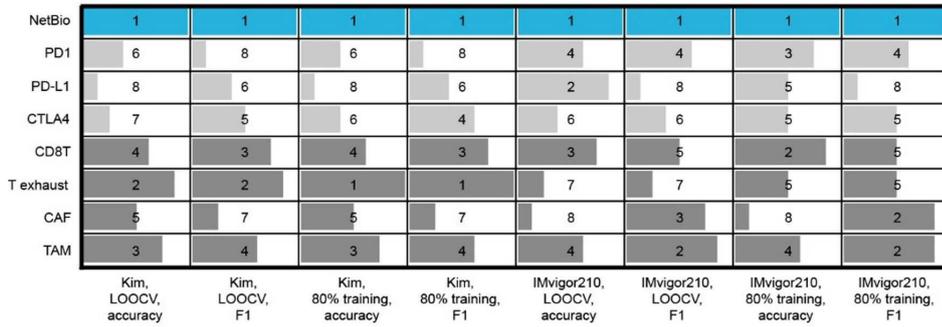


도면9

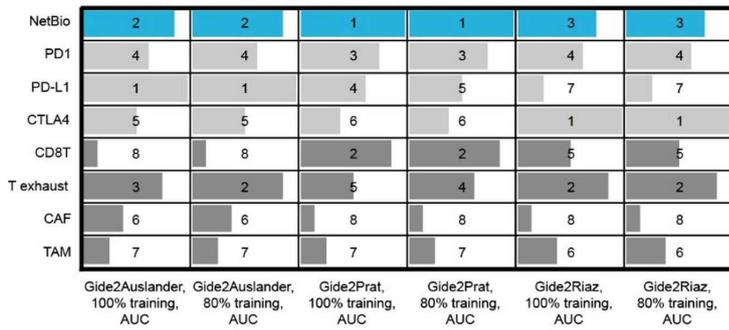
a Within study predictions in melanoma



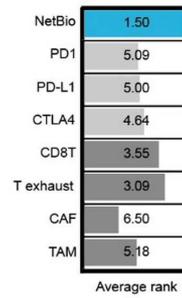
b Within study predictions in gastric cancer and bladder cancer



c Across study predictions in melanoma

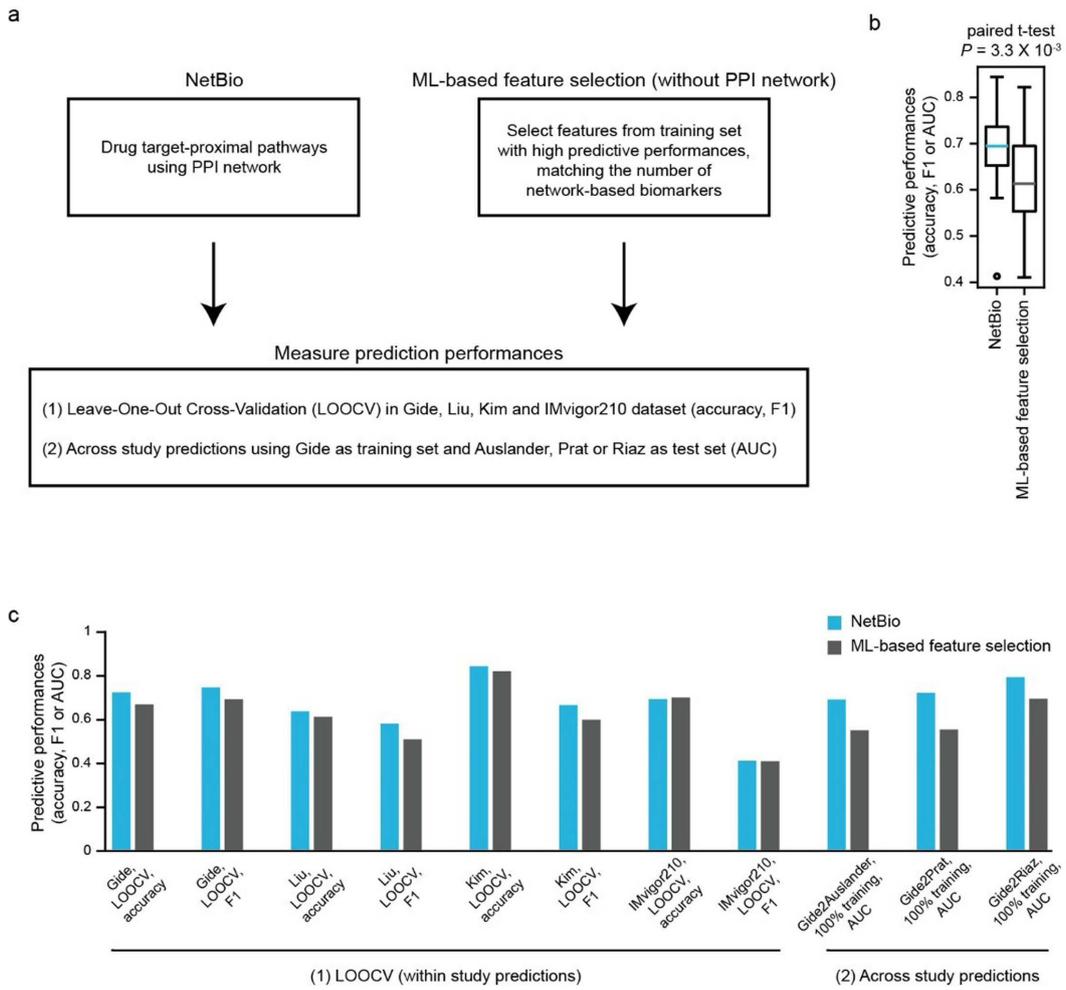


d Overall rank

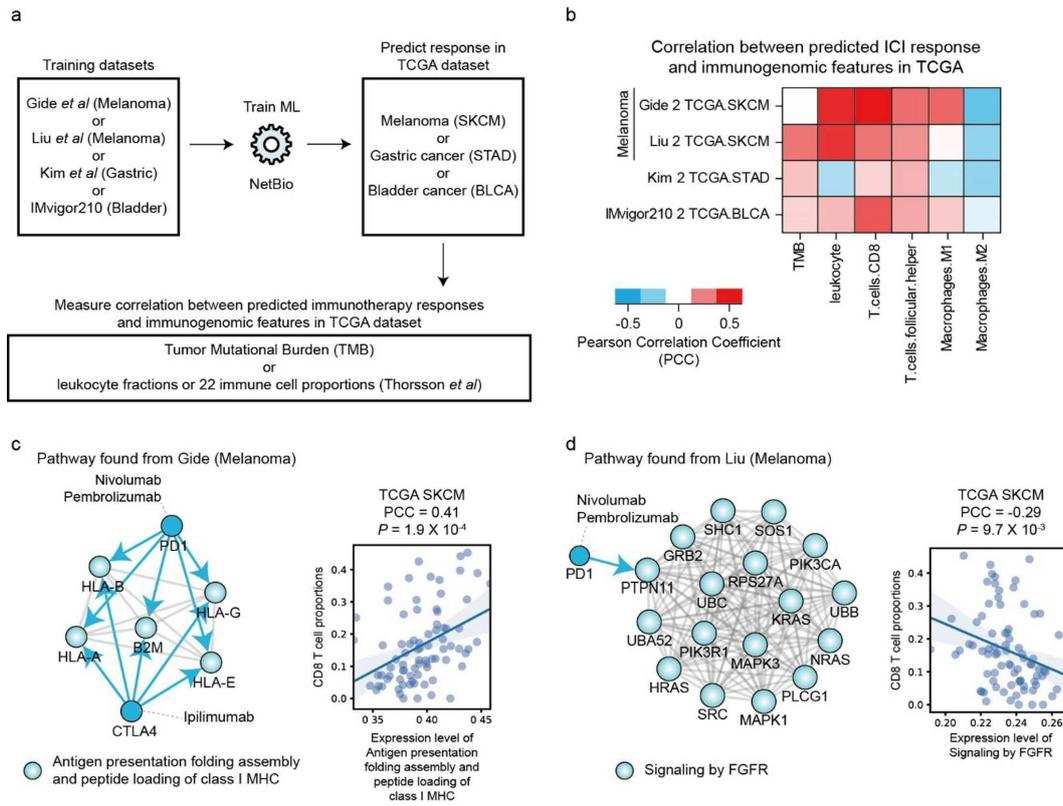


NetBio GeneBio TME-Bio

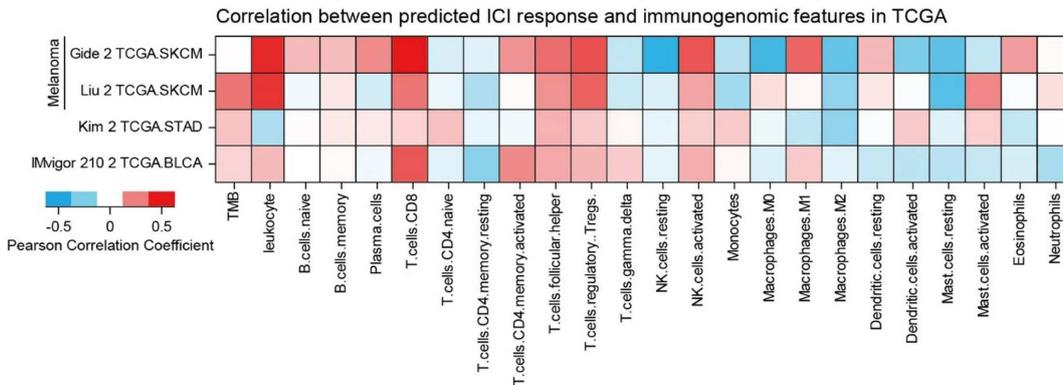
도면10



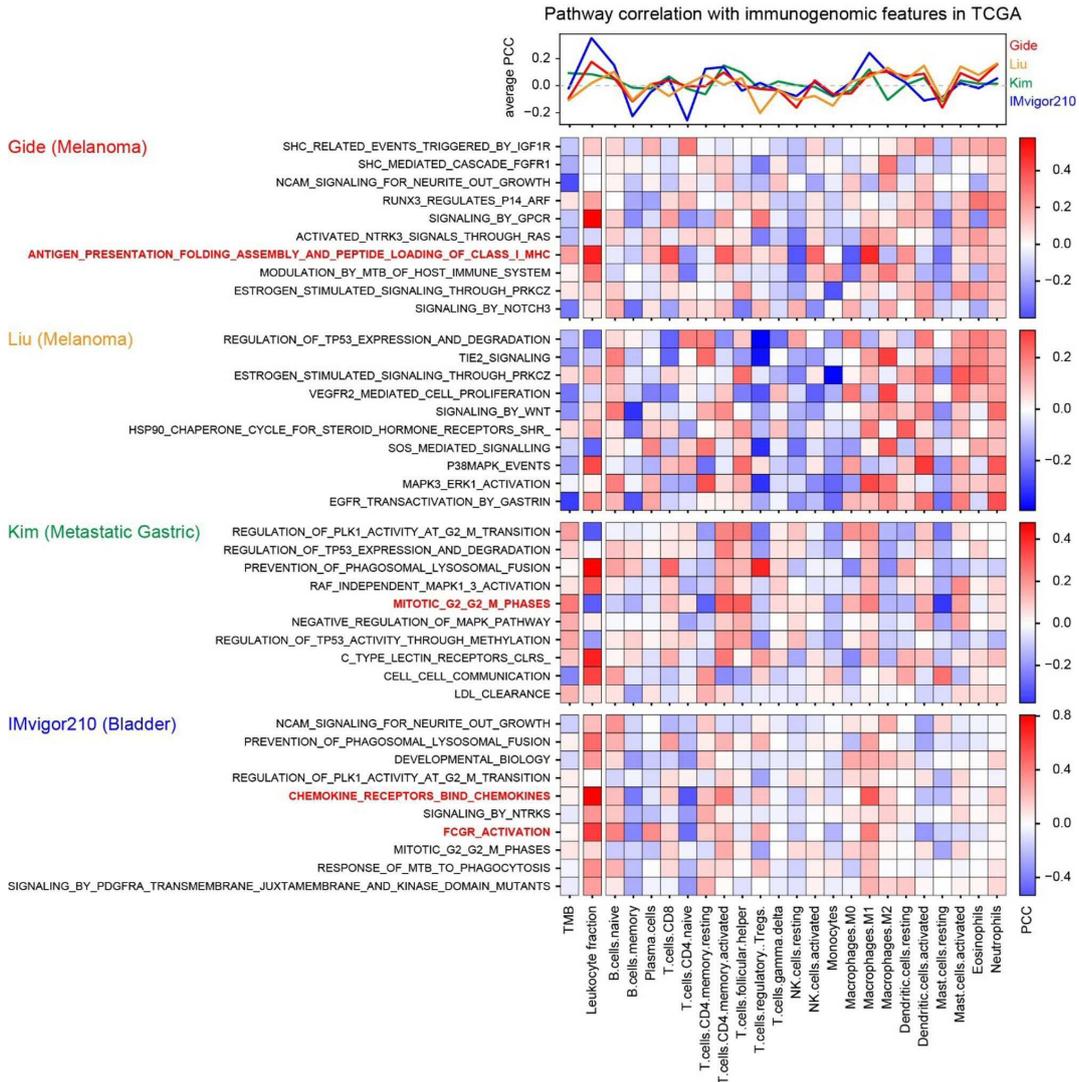
도면11



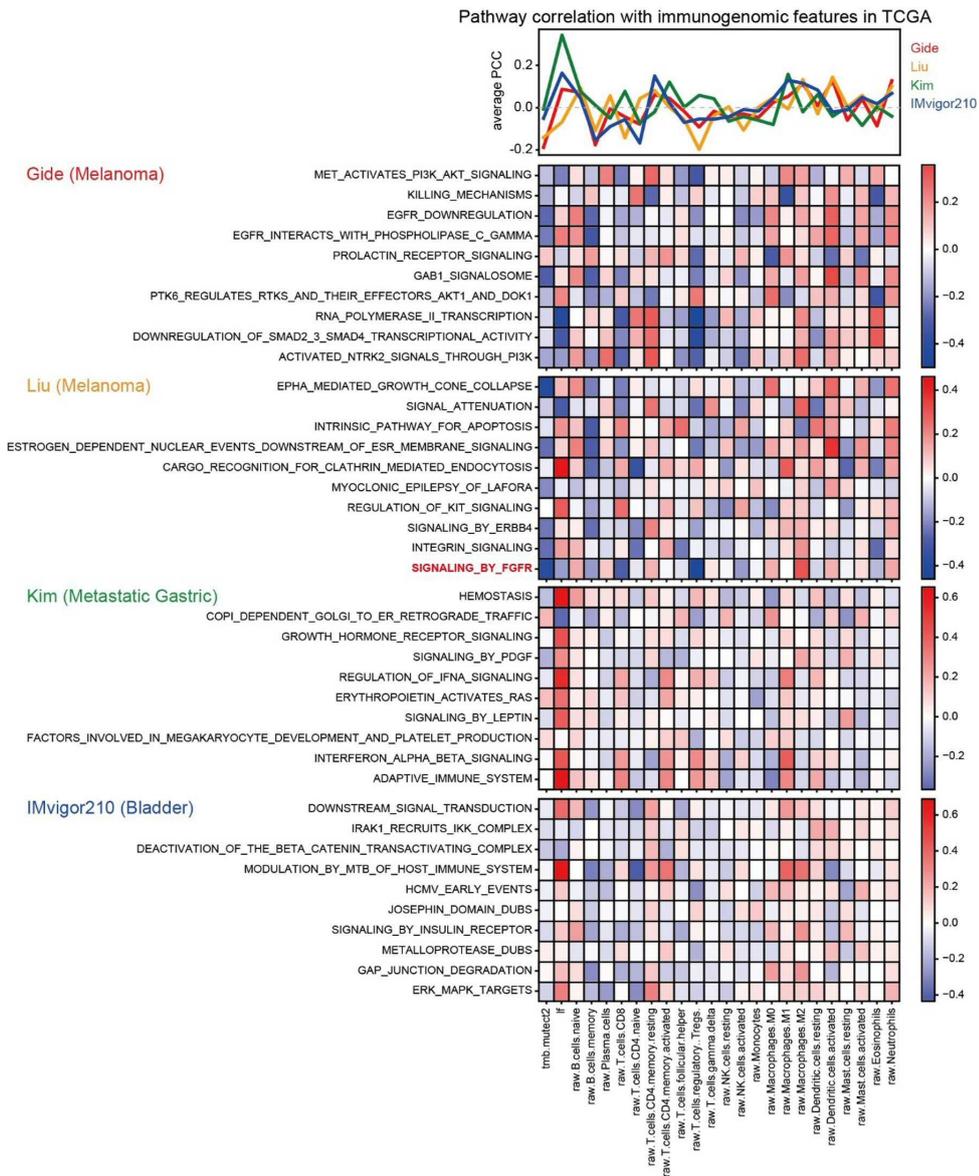
도면12



도면13

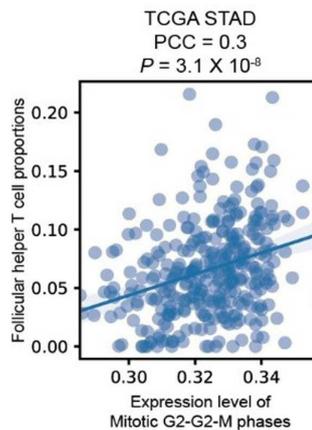
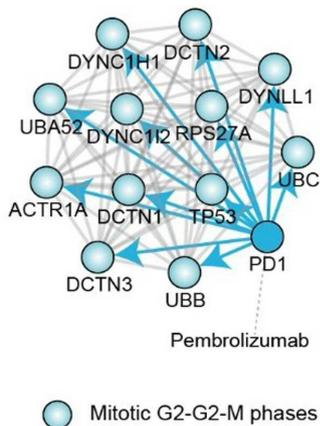


도면14

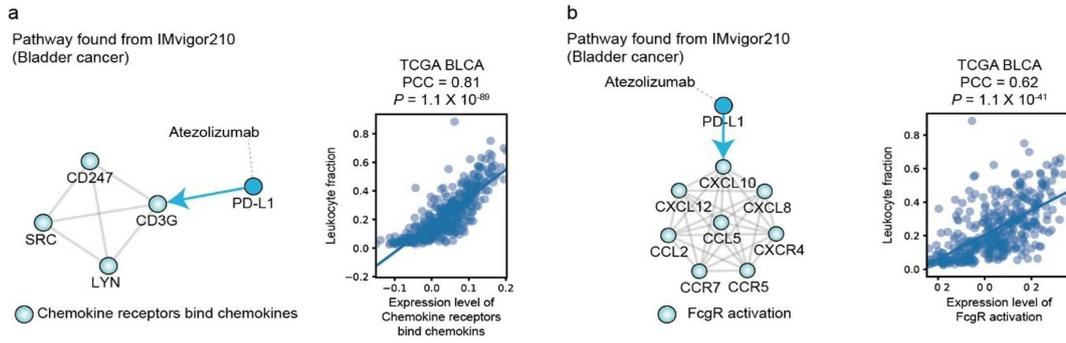


도면15

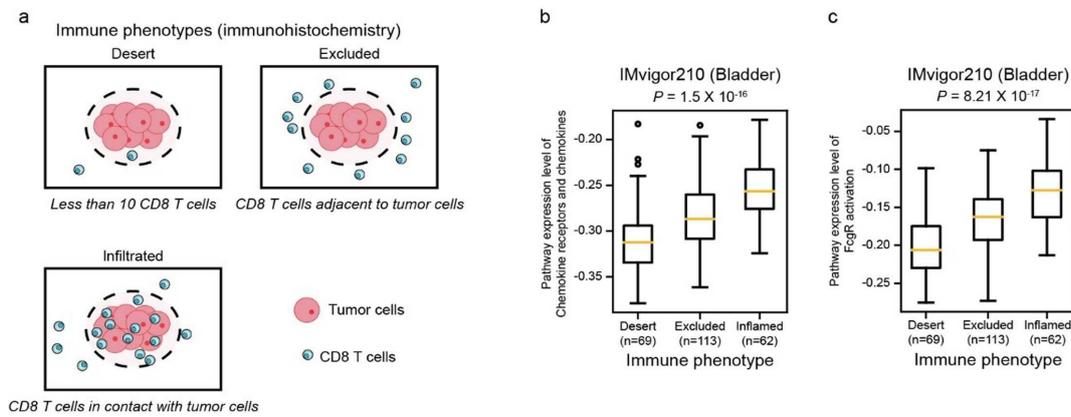
Pathway found from Kim (Gastric cancer)



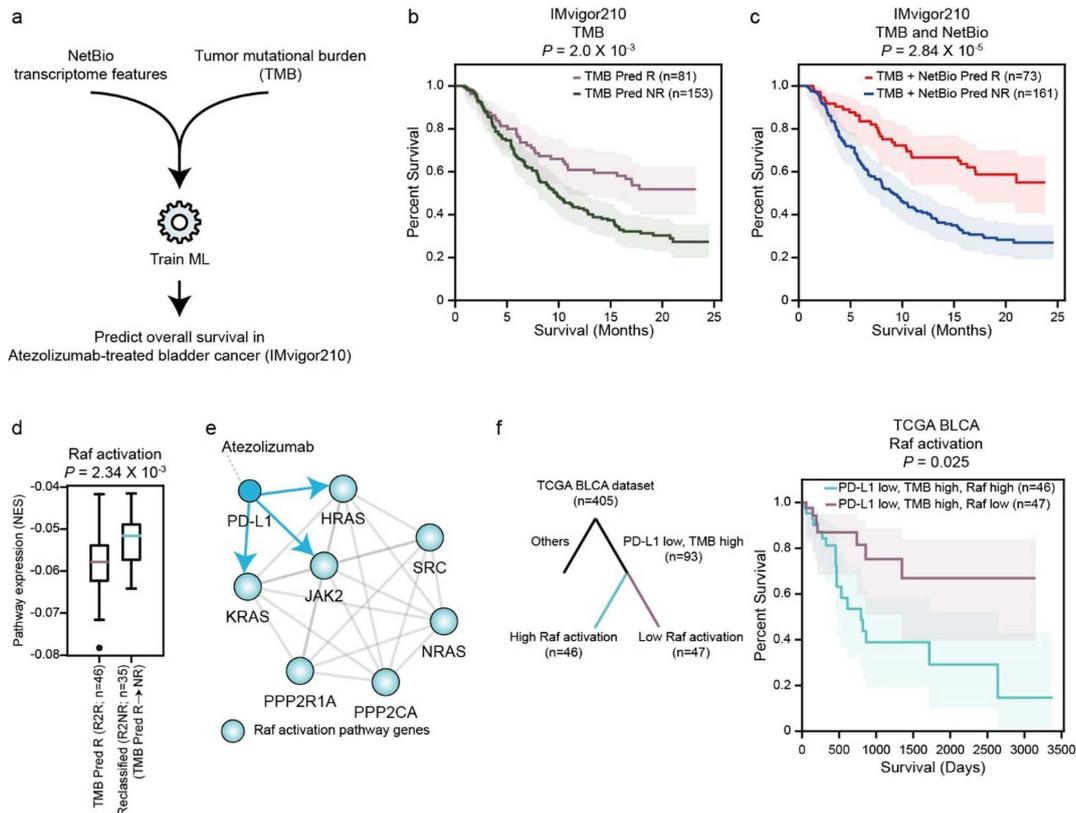
도면16



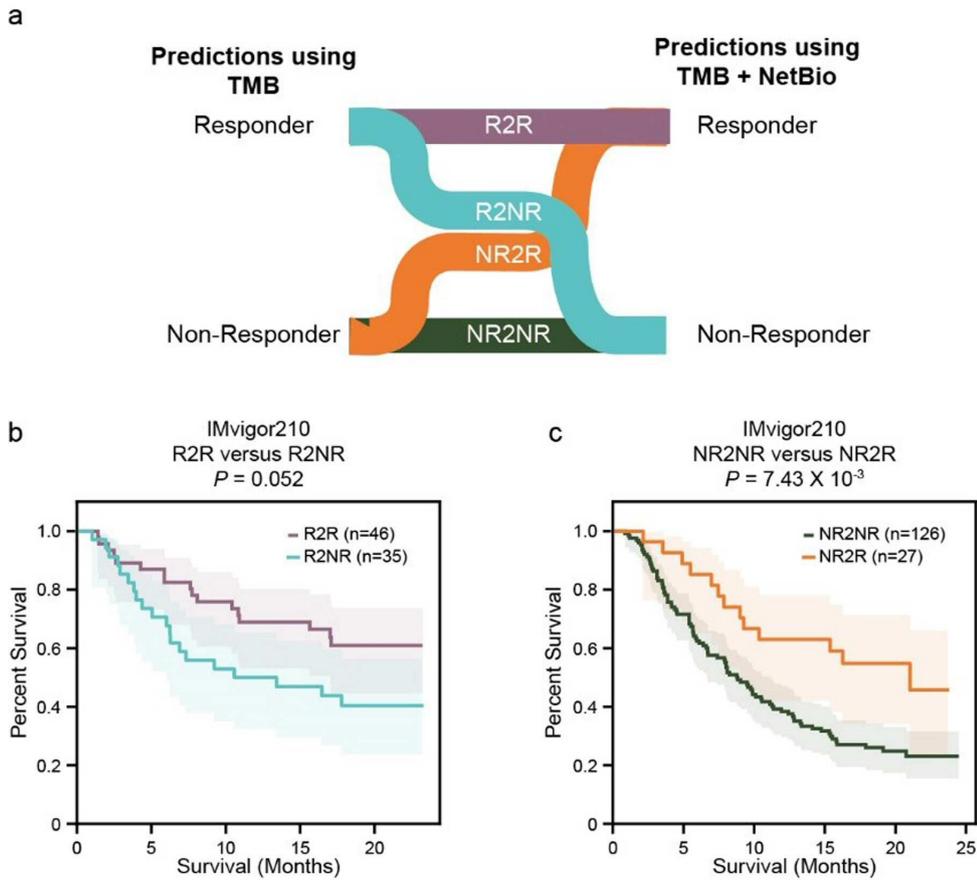
도면17



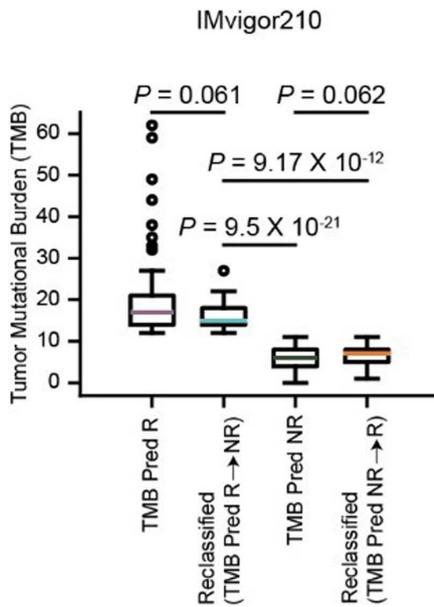
도면18



도면19



도면20



도면21

