



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0035367
(43) 공개일자 2024년03월15일

(51) 국제특허분류(Int. Cl.)
G01N 33/574 (2006.01) C12Q 1/6886 (2018.01)
(52) CPC특허분류
G01N 33/57407 (2013.01)
C12Q 1/6886 (2022.01)
(21) 출원번호 10-2023-0119290
(22) 출원일자 2023년09월07일
심사청구일자 2023년09월07일
(30) 우선권주장
1020220113349 2022년09월07일 대한민국(KR)

(71) 출원인
재단법인 아산사회복지재단
서울특별시 송파구 올림픽로43길 88 (풍납동)
울산대학교 산학협력단
울산광역시 남구 대학로 93(무거동)
(72) 발명자
신동명
서울특별시 강동구 고덕로 131, 105동 2904호(압사동, 강동 롯데캐슬퍼스트아파트)
조영미
서울특별시 강남구 압구정로 151, 101동 407호(압구정동, 현대아파트)
김용환
서울특별시 강동구 구천면로11길 18-16, 201호(천호동)
(74) 대리인
특허법인 이룸리온

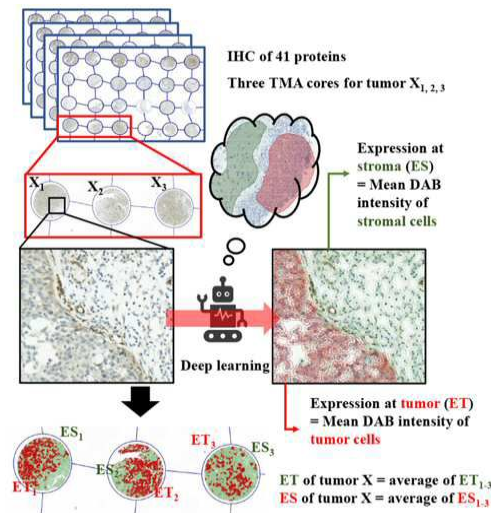
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 선행화학요법 내성의 고형암 환자 진단용 바이오마커 및 이를 이용한 선행화학요법 내성 진단을 위한 정보제공방법

(57) 요약

본 발명은 고형암(예컨대, 근치윤성 방광암(MIBC))의 선행화학요법(NAC)으로 투여되는 백금제제(예컨대, 시스플라틴)의 내성을 진단하기 위한 바이오마커에 관한 것으로, 본 발명에 따르면, 글루타티온(GSH) 경로(pathway)는 고형암의 화학요법 내성과 연관되어 있음을 확인하였다. 이에, 본 발명에서는 글루타티온(GSH) 대사체를 고형암의 화학요법 내성 진단용 바이오마커로서 제공한다.

대표도 - 도5a



(52) CPC특허분류

C12Q 2600/106 (2013.01)

G01N 2500/00 (2013.01)

G01N 2800/52 (2021.08)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711192152
과제번호	2021R1A2C2005790
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)
연구과제명	TFCP2L1 전사 인자 아세틸화 기반 전분화능 줄기세포와 방광암 세포의 줄기세포성
조절 기전 규명	
기 여 율	1/1
과제수행기관명	울산대학교
연구기간	2023.03.01 ~ 2024.02.29

명세서

청구범위

청구항 1

글루타티온 대사체 또는 이의 조합을 포함하는, 고형암 환자의 화학요법 내성 진단용 바이오마커 조성물.

청구항 2

제1 항에 있어서,

상기 고형암은 근침윤성 방광암인, 바이오마커 조성물.

청구항 3

제1 항에 있어서,

상기 글루타티온 대사체는 GLS1, PSAT1, CBS, GCLC, GCLM, Gluta (GSR), GlnRS (QARS1), GGT7, Perox (PRDX1), PLOD2, RPAP1, RPL9, MITF, CD44v6, CDK1, FZD9, GAD2, PPP2R5A, non-p (b-catenin), B-catenin, SALL4, SOX2, TFCEP2L1, TFEB, ICAM1, TRAF2, TRAF6, IL15RA, AFAP1, CARD16, CD11c, CD73, CD3Z, EBI3 (IL27B), EMX1, DNMT3L, DPH2, EGR2, FYB1, GADD45B, HCFC1R1, KCTD14, MTCH1, OCT2, PCDHB9, PPIL2, RFX7, SLC15A3, TNFAIP8, ANPEP, BDN, EOGT, FOXA1, KIFC2, KIR3DL1, NOTUM, USP2, CCDC6, CK5(KRT5), CK20 (KRT20), GATA3, BHLHE40, C11orf95, REX7, COX17, NTSE, ITGAX, KIAA1671, NT5E, POU2F2, PLEK, POSTN, PPP1R35, THBS2 및 INFAIP8으로 이루어진 군에서 선택된 적어도 어느 하나의 유전자인, 바이오마커 조성물.

청구항 4

제1 항에 있어서,

상기 화학요법 내성은 백금제제에 대한 내성인, 바이오마커 조성물.

청구항 5

제4 항에 있어서,

상기 백금제제는 시스플라틴인, 바이오마커 조성물.

청구항 6

제4 항에 있어서,

상기 백금제제는 선행화학요법으로 투여되는 것인, 바이오마커 조성물.

청구항 7

글루타티온 대사체의 발현을 측정하는 제제를 포함하는, 고형암 환자의 화학요법 내성 진단용 조성물.

청구항 8

제7 항에 있어서,

상기 글루타티온 대사체는 GLS1, PSAT1, CBS, GCLC, GCLM, Gluta (GSR), GlnRS (QARS1), GGT7, Perox (PRDX1), PLOD2, RPAP1, RPL9, MITF, CD44v6, CDK1, FZD9, GAD2, PPP2R5A, non-p (b-catenin), B-catenin, SALL4, SOX2, TFCEP2L1, TFEB, ICAM1, TRAF2, TRAF6, IL15RA, AFAP1, CARD16, CD11c, CD73, CD3Z, EBI3 (IL27B), EMX1, DNMT3L, DPH2, EGR2, FYB1, GADD45B, HCFC1R1, KCTD14, MTCH1, OCT2, PCDHB9, PPIL2, RFX7, SLC15A3, TNFAIP8, ANPEP, BDN, EOGT, FOXA1, KIFC2, KIR3DL1, NOTUM, USP2, CCDC6, CK5(KRT5), CK20 (KRT20), GATA3, BHLHE40, C11orf95, REX7, COX17, NTSE, ITGAX, KIAA1671, NT5E, POU2F2, PLEK, POSTN, PPP1R35, THBS2 및 INFAIP8 으로 이루어진 군에서 선택된 적어도 어느 하나의 유전자인, 고형암 환자의 화학요법 내성 진단용 조성

물.

청구항 9

제7 항에 있어서,

상기 글루타티온 대사체의 발현을 측정하는 제제는 상기 대사체의 유전자에 특이적으로 결합하는 펩타이드, 항체, 프라이머를 포함하는 것인, 고형암 환자의 화학요법 내성 진단용 조성물.

청구항 10

제7 항의 조성물을 포함하는, 고형암 환자의 화학요법 내성 진단용 키트.

청구항 11

- a) 고형암 환자에서 분리된 시료로부터 글루타티온 대사체의 발현량을 측정하는 단계;
- b) 상기 글루타티온 대사체의 발현량을 정상 대조군 시료의 발현량과 비교하는 단계; 및
- c) 고형암 환자에서 분리된 시료로부터 측정된 글루타티온 대사체의 발현량이 대조군 시료의 수준과 차이가 있는 경우, 화학요법제 내성이 있는 것으로 판단하는 단계를 포함하는, 고형암 환자의 화학요법 내성 진단을 위한 정보제공 방법.

청구항 12

제11 항에 있어서,

상기 글루타티온 대사체의 발현량은 웨스턴 블랏, 효소-면역화학 검출법(ELISA), 면역조직화학 염색법, 면역침강, 면역형광법, 전자현미경, 후성유전체 및 정량적 실시간 PCR 기법으로 구성된 군으로부터 선택되는 어느 하나 이상의 방법으로 측정되는, 정보제공 방법.

청구항 13

제11 항에 있어서,

상기 a) 단계의 글루타티온 대사체는 GLS1, PSAT1, CBS, GCLC, GCLM, Gluta (GSR), GlnRS (QARS1), GGT7, Perox (PRDX1), PLOD2, RPAP1, RPL9, MITF, CD44v6, CDK1, FZD9, GAD2, PPP2R5A, non-p (b-catenin), B-catenin, SALL4, SOX2, TFCP2L1, TFEB, ICAM1, TRAF2, TRAF6, IL15RA, AFAP1, CARD16, CD11c, CD73, CD3Z, EBI3 (IL27B), EMX1, DNMT3L, DPH2, EGR2, FTYL1, GADD45B, HCFC1R1, KCTD14, MTCH1, OCT2, PCDHB9, PPIL2, RFX7, SLC15A3, TNFAIP8, ANPEP, BDN, EOGT, FOXA1, KIFC2, KIR3DL1, NOTUM, USP2, CCDC6, CK5(KRT5), CK20 (KRT20), GATA3, BHLHE40, C11orf95, REX7, COX17, NTSE, ITGAX, KIAA1671, NT5E, POU2F2, PLEK, POSTN, PPP1R35, THBS2 및 TNFAIP8으로 이루어진 군에서 선택된 적어도 어느 하나의 유전자인, 정보제공 방법.

청구항 14

제11 항에 있어서,

상기 a) 단계의 분리된 시료는 타액, 뇨, 조직, 전혈, 혈청 또는 혈장인, 정보제공 방법.

청구항 15

- a) 고형암 환자에서 분리된 시료로부터 글루타티온 대사체의 발현량을 측정하는 단계;
- b) 상기 글루타티온 대사체의 발현량을 정상 대조군 시료의 발현량과 비교하는 단계; 및
- c) 고형암 환자에서 분리된 시료로부터 측정된 글루타티온 대사체의 발현량이 대조군 시료의 수준과 차이가 있는 경우, 화학요법제 내성이 있는 것으로 판단하는 단계를 포함하는, 고형암 환자의 치료방법 결정을 위한 정보 제공 방법.

발명의 설명

기술 분야

[0001] 본 발명은 고형암 환자에 진행되는 선행화학요법에 내성과 글루타치온 경로 간에 연관성이 있음을 확인하고, 글루타치온 대사체를 고형암 환자의 선행화학요법 내성 진단을 위한 바이오마커로서 제공한다. 본 발명의 바이오마커 조성물, 키트, 정보제공방법을 이용하면, 고형암 환자의 선행화학요법의 내성이 나타나는지 여부를 용이하게 진단할 수 있다.

배경 기술

[0002] 방광암(BC)은 발생률과 사망률이 높으며, 2020년에 약 573,278명의 신규 환자가 발생하고 212,536명이 사망하여 전 세계 남성에서 6번째로 흔한 악성 종양으로 평가된다. 방광암은 상당한 임상적, 병리학적 이질성을 나타낸다. 방광벽 침범 정도에 따라 BC는 비근층고유침윤성 BC(NMIBC; Ta, 상피내암종, T1) 또는 고유근침윤성 BC(MIBC; $\geq T2$)로 분류된다. NMIBC는 초기 진단 당시 대부분의 BC 환자(70~80%)를 차지한다. NMIBC는 일반적으로 비공격성 종양이며 경요도 방광종양 절제술(TURBT)과 방광내 화학요법 또는 BCG 백신으로 관리된다. 그러나 MIBC의 재발률과 진행률이 높기 때문에 평생 감시 비용이 많이 든다. 종양 발병률의 25%를 차지하는 MIBC는 빠르게 진행되어 전이성 BC로 변할 수 있으며 대부분의 환자 사망률을 담당하며 BC 사망률을 낮추려면 MIBC의 적절한 관리가 중요하다. TURBT를 통해 진단한 후, 수술 전 시스플라틴 기반 화학요법, 즉 신보강 화학요법(NAC)을 사용한 근치 방광절제술이 현재 MIBC 치료의 표준이다. NAC는 30~40%의 사례에서만 효과적이며 반응이 없는 나머지 환자는 종양학적 이점 없이 약물 독성을 겪을 수 있으며 최종 치료가 지연될 수 있다. 그러나 현재 MIBC 환자의 임상병리학적 특징은 NAC 반응을 전향적으로 예측하기에는 부적절하다. 반응이 없는 환자에 대한 치료 전략은 확립되어 있지 않다. 따라서 MIBC 임상 관리의 정확성과 효능을 향상시키기 위해 화학요법 저항성과 예측 바이오마커의 기본 메커니즘을 식별하는 것은 상당한 미충족 임상적 요구에 기여할 수 있다.

[0003] 게놈 차원의 유전자 발현 프로파일링은 예후 평가 및 치료 반응을 포함하여 MIBC의 이질적 행동의 복잡성에 대한 정확한 분자 통찰력을 생성했다. 예를 들어, p53 유사 분자 아형을 가진 환자는 NAC에 대한 반응이 좋지 않다. 대조적으로, 기저 유사 종양을 가진 환자는 단독 수술을 받은 타의 추종을 불허하는 환자 코호트에 비해 NAC 후 전체 생존율(OS)이 증가하여 가장 큰 종양학적 이점을 나타냈다. 그러나 이전 연구에서는 기저 편평형(Ba/Sq) 발현 하위 유형의 환자에서 NAC 반응률이 감소한 것으로 보고되었다. 또한 이전에 발표된 주요 데이터 세트를 통합한 연구에서는 합의 분류를 통한 NAC 반응 예측에서 통계적으로 유의미한 연관성이 관찰되지 않았다. 이러한 서로 다른 결과는 MIBC에서 NAC에 대한 대응을 결정하는 복잡성을 강조한다. 따라서 MIBC 임상 관리의 정확성과 효능을 향상시키기 위해 화학요법 저항성과 예측 바이오마커의 기본 메커니즘을 식별하는 것은 상당한 미충족 임상적 요구에 기여한다.

[0004] 본 발명자들은, MIBC NAC 반응에서 분자 이질성에 대한 생물학적 기초를 확인하기 위해 4개의 독립적인 코호트에서 MIBC 환자 종양의 유전자 발현 프로파일링을 수행했다. 글루타메이트, 시스테인 및 글리신으로 구성된 매우 풍부한 비단백질 티올 트리펩타이드인 글루타티온(GSH)의 수준과 활성을 조절하기 위한 대사 경로의 조정된 상향 조절의 중요성을 정의한다. 더 나아가 머신러닝 기반의 종양-기질 분류기를 이용한 MIBC 종양 조직의 면역염색을 위한 디지털화된 분석, 생세포 실시간 GSH 모니터링, in vitro 세포배양, in vivo 동물모델 분석을 통합하여, 본 발명자들은 MIBC NAC 반응에 대한 잠재적인 예측 바이오마커로서 그리고 잠재적으로 화학저항성 MIBC를 재감작시키는 새로운 치료 표적으로서 GSH 역학의 생물학적 중요성과 임상적 관련성을 입증했다.

발명의 내용

해결하려는 과제

[0005] 본 발명은 상기의 문제점을 해결하고 상기의 필요성에 의하여 안출된 것으로서, 본 발명의 목적은 고형암 환자에 선행화학요법의 내성과 연관된 매커니즘을 확인하고 이를 내성의 발생 여부 진단을 위한 방법으로 제공하고자 함에 있다.

[0006] 본 발명은, 고형암 환자의 선행화학요법에 대한 내성의 발생 여부를 진단할 수 있는 바이오마커 조성물을 제공함에 있다.

[0007] 본 발명은 또한, 고형암 환자의 선행화학요법에 대한 내성의 발생 여부를 진단할 수 있는 키트를 제공하고자 한다.

과제의 해결 수단

[0008] 상술한 과제를 해결하기 위해, 본 발명은 글루타티온 대사체 또는 이의 조합을 포함하는, 고형암 환자의 화학요

법 내성 진단, 예후 예측용 바이오마커 조성물을 제공한다.

[0009] 본 발명에서, 상기 고형암은 방광암(Bladder cancer), 대장암(Colon cancer), 위암(Stomach cancer), 폐암(Lung cancer), 폐선암(Lung adenocarcinoma), 유방침습성유관암종(Breast invasive ductal carcinoma), 결장선암(Colon adenocarcinoma), 전립선선암(Prostate adenocarcinoma), 방광요로피암종(Bladder urothelial carcinoma), 폐편평세포암종(Lung squamous cell carcinoma), 피부흑색종(Cutaneous melanoma), 원발부위불명암(Cancer of unknown primary), 췌장도관선암종(Pancreatic adenocarcinoma), 교모세포종(Glioblastoma multiforme), 대장선암(Colorectal adenocarcinoma), 고등급장액성난소암(High grade serous ovarian cancer), 위선암(Stomach adenocarcinoma), 신세포암종(Renal clear cell carcinoma), 식도암(Esophageal adenocarcinoma), 고환암(Testicular cancer) 및 간내담관암(Intrahepatic cholangiocarcinoma)을 포함하나, 이에 제한되는 것은 아니며, 가장 바람직하게는 근침윤성 방광암(muscle-invasive bladder cancer, 이하 MIBC라고도 함)이다.

[0010] 본 발명에서, 상기 글루타티온 대사체는 GLS1, PSAT1, CBS, GCLC, GCLM, Gluta (GSR), GlnRS (QARS1), GGT7, Perox (PRDX1), PLOD2, RPAP1, RPL9, MITF, CD44v6, CDK1, FZD9, GAD2, PPP2R5A, non-p (b-catenin), B-catenin, SALL4, SOX2, TFCEP2L1, TFEB, ICAM1, TRAF2, TRAF6, IL15RA, AFAP1, CARD16, CD11c, CD73, CD3Z, EBI3 (IL27B), EMX1, DNMT3L, DPH2, EGR2, FYB1, GADD45B, HCFC1R1, KCTD14, MTCH1, OCT2, PCDHB9, PPIL2, RFX7, SLC15A3, TNFAIP8, ANPEP, BDN, EOGT, FOXA1, KIFC2, KIR3DL1, NOTUM, USP2, CCDC6, CK5(KRT5), CK20 (KRT20), GATA3, BHLHE40, C11orf95, REX7, COX17, NTSE, ITGAX, KIAA1671, NT5E, POU2F2, PLEK, POSTN, PPP1R35, THBS2 및 TNFAIP8를 포함한다.

[0011] 본 발명에서, 상기 화학요법 내성은 백금제제(예컨대, 시스플라틴 류)에 대한 내성을 의미할 수 있으며, 상기 백금제제는 선행화학요법으로 투여되는 것일 수 있다. 즉, 본 발명은 선행 (항암)화학요법으로 처리되는 백금제제에 대한 내성을 진단하기 위한 바이오마커로서 글루타티온 대사체를 제공한다.

[0012] 본 발명은 또한, 글루타티온 대사체의 발현을 측정하는 체제를 포함하는, 근침윤성 방광암 환자의 화학요법 내성 진단용 조성물을 제공한다. 본 발명에서, 상기 글루타티온 대사체의 발현을 측정하는 체제는 상기 대사체의 유전자에 특이적으로 결합하는 펩타이드, 항체, 프라이머를 포함할 수 있다.

[0013] 본 발명은 또한, 상기 근침윤성 방광암 환자의 화학요법 내성 진단용 조성물을 포함하는, 근침윤성 방광암 환자의 화학요법 내성 진단용 키트를 제공한다.

[0014] 본 발명은 일 측면에서, a) 근침윤성 방광암 환자에서 분리된 시료로부터 글루타티온 대사체의 발현량을 측정하는 단계; b) 상기 글루타티온 대사체의 발현량을 정상 대조군 시료의 발현량과 비교하는 단계; 및 c) 근침윤성 방광암 환자에서 분리된 시료로부터 측정된 글루타티온 대사체의 발현량이 대조군 시료의 수준과 차이가 있는 경우, 화학요법제 내성이 있는 것으로 판단하는 단계를 포함하는, 근침윤성 방광암 환자의 화학요법 내성 진단을 위한 정보제공 방법에 관한 것이다.

[0015] 본 발명에서, 상기 글루타티온 대사체의 발현량은 웨스턴 블랏, 효소-면역화학 검출법(ELISA), 면역조직화학 염색법, 면역침강, 면역형광법, 전사체, 후성유전체 또는 정량적 실시간 PCR 기법으로 측정될 수 있다.

[0016] 본 발명에서, 상기 글루타티온 대사체의 발현을 분석하기 위한 시료로는 타액, 뇨, 조직, 전혈, 혈청 또는 혈장을 이용할 수 있으나, 이에 제한되는 것은 아니다.

[0017] 본 발명은 또한, a) 근침윤성 방광암 환자에서 분리된 시료로부터 글루타티온 대사체의 발현량을 측정하는 단계; b) 상기 글루타티온 대사체의 발현량을 정상 대조군 시료의 발현량과 비교하는 단계; 및 c) 근침윤성 방광암 환자에서 분리된 시료로부터 측정된 글루타티온 대사체의 발현량이 대조군 시료의 수준과 차이가 있는 경우, 화학요법제 내성이 있는 것으로 판단하는 단계를 포함하는, 근침윤성 방광암 환자의 치료방법 결정을 위한 정보 제공 방법에 관한 것이다.

발명의 효과

[0018] 본 발명은 글루타티온 경로가 고형암 환자의 선행화학요법에 대한 내성과 연관이 있음을 확인하고, 고형암 환자의 선행화학요법에 대한 내성의 발생 여부를 진단할 수 있는 바이오마커 조성물을 제공한다.

[0019] 본 발명은 또한, 고형암 환자의 선행화학요법에 대한 내성의 발생 여부를 진단할 수 있는 바이오마커 조합을 제공한다.

[0020]

본 발명에 바이오마커 조성물을 이용하는 경우, 비침습적으로 간단하고 신속하게 선행 화학 요법에 대한 내성이 발생하였는지 여부를 진단할 수 있다.

도면의 간단한 설명

[0021]

도 1은 근침윤성 방광암(MIBC) 환자의 다중 코호트 전사체 및 임상 데이터 분석을 통합하여 머신러닝 기반 종양-기질 분류기를 이용한 면역염색 디지털 분석, 생세포 실시간 세포 모니터링, 체외 세포배양 및 생체 내 동물 모델 실험 절차를 나타내는 모식도이다. 도 1을 참조하면, 본 발명은 수술 전 시스플라틴 기반 신보조 화학요법에 대한 MIBC 반응에 대한 잠재적인 예측 바이오마커로서 그리고 잠재적으로 화학저항성 MIBC를 재감작화하는 새로운 치료 표적으로서 글루타티온 역학의 생물학적 중요성과 임상적 관련성을 확인한다.

도 2a 내지 도 2g는 근침윤성 방광암(MIBC)의 선행화학요법(NeoAdjuvant Chemotherapy) 반응을 기반으로 한 전사체 프로파일링을 나타낸다. 도 2a는 분자 특징, NAC 반응 및 임상 데이터를 갖춘 MIBC 환자의 다중 코호트 분석에 대한 개략도이다. 도 2b는 MIBC 환자를 NAC 반응(R) 및 무응답(NR) 그룹($n = 8/\text{그룹}$, NR 및 R)으로 분류한 후, AMC 발견 코호트의 전사체의 주요 구성 요소를 분석한 결과이다. 도 2c는 NR과 R 그룹 전사체 사이의 차별적으로 발현된 유전자(DEG)의 히트맵 분석 결과이다. 도 2d는 유전자 집합 농축 분석(GSEA) 및 분자 MIBC 하위 유형을 계층화하기 위해 제시된 분자 분류기와 관련된 유전자 집합을 사용하는 대표적인 농축 플롯 결과이다 (NES: 정규화된 농축 점수, FDR: 잘못된 발견률). 도 e 및 도 f는 본 발명 NAC NR 및 R 그룹 간의 전사체를 비교하여 MetaCore 분석을 통해 확인된 가장 풍부하게 농축된 10개의 경로 맵(도 2e) 및 대표적인 농축 유전자 네트워크(도 2f)를 나타낸 이미지이다. Fold 변화 값은 유전자 네트워크에 중첩되어 빨간색으로 상향 조절된 유전자와 파란색으로 하향 조절된 유전자를 나타낸다. 도 2g는 AMC 발견 코호트에서 NAC 반응을 정의하는 유전자 세트의 대표적인 농축 플롯이 있는 GSEA이다. 도 2g에서 버블 플롯은 $-\text{Log}_2(\text{FDR})$ 에 따른 크기 값을 통해 NES 순서로 표시되었다.

도 3a 내지 도 3i는 다른 MIBC 코호트의 NAC 반응 분자의 특징을 분석한 결과이다. 도 3a 및 도 3b는 AMC 검증 코호트에서 39명의 환자(NR $n = 18$, R $n = 21$)의 전사체 데이터 세트에 대한 GSEA 분석에서 최고 점수를 받은 10개의 유전자 세트(도 3a 왼쪽 패널) 및 대표적인 농축 플롯(도 3a 오른쪽 패널)을 나타낸다. 여기서, 유전자 세트는 표준화된 농축 점수(NES)에 따라 나열되었다. 도 3b는 종양 침윤 림프구(TIL)를 평가하기 위한 AMC 검증 코호트의 TUBRT 종양 표본(X100 배율, 스케일 바 = 100 μm)의 대표 이미지를 나타낸다. 도 3b에서 정량적 데이터는 평균 $\pm$ SEM으로 표시된다 (* $p < 0.05$, unpaired Student's t-tests). 도 3c 내지 도 3e는 NAC NR과 NAC R 사이의 전사체를 비교한 MetaCore 분석 결과로, 당시 가장 풍부해진 경로 맵(도 3c), GO 프로세스(도 3d) 및 무기 화합물의 해독과 관련된 대표적인 강화된 유전자 네트워크(도 3e) AMC 검증 코호트의 R 그룹 결과를 나타낸다. 도 3f는 AMC 검증을 위한 전사체 데이터 세트 및 MIBC 샘플에 대한 설명과 NAC 이전 TURBT 표본의 공개적으로 이용 가능한 유전자 발현 프로파일링 데이터 세트와 NAC에 대한 병리학적 반응 기록을 포함하는 3개의 독립적인 외부 코호트에 대한 설명을 나타낸다. 도 3f에서 GSEA 분석에서 확인된 각 데이터세트와 중요한 유전자 세트에서 사용된 MIBC 샘플의 수는 각각 "MIBC 샘플" 및 "유의한 유전자세트 수" 열에 표시된다. 도 3g 내지 도 3i는 발견 코호트에서 NAC 반응과 차별적으로 연관된 30개의 유전자 세트를 사용하여 3개의 외부 코호트에 대한 GSEA 분석을 나타낸다. 도 3g 및 도 3h는 "MDA MVAC" 코호트에서 최고 점수를 받은 중요한 유전자 세트(도 3g) 및 "MDA DDMVAC" 코호트에서 GSH 대사 유전자 세트(도 3h)의 대표적인 농축 플롯을 나타낸다(ES: 농축 점수; NOM p-값, 명목 p-값; FDR: 잘못된 발견률). 도 3i는 AMC 및 3개의 외부 검증 코호트에 대한 GSEA 분석에서 FDR 값을 요약하는 히트맵 분석 결과이다.

도 4a 내지 도 4e는 AMC 발견 코호트에서 NAC 반응을 특성화하는 바이오마커를 확인한 결과이다. 도 4a는 본 발명에 따른 NAC NR 그룹에서 풍부한 상위 10개 유전자 세트(왼쪽)를 사용한 AMC 발견 코호트의 전사체에 대한 최첨단 분석 결과이다. 도 4에서는 유전자 세트의 풍부함(도 4a 우측 이미지)과 해당 표적 유전자의 히트맵(도 4a 하단 이미지)을 나타낸다. 도 4b는 최첨단(GSEA) 및 유전자 네트워크(MetaCore) 분석에 따라 중요한 유전자로 선택된 추정 바이오마커 유전자에 대한 qPCR 결과의 버블 플롯을 나타낸다. 도 4c는 AMC 코호트 환자(NR $n = 17$, R $n = 19$)에서 바이오마커 유전자 하위세트의 발현을 검증하기 위한 qPCR 분석 결과이다. 도 4c에서 발현은 인간 b2-마이크로글로불린(B2M) 발현에 대한 백분율로 표시된다. 도 4d는 AMC 발견 코호트($n = 8/\text{그룹}$, NR 및 R)에서 GSH 및 대사 경로(상단 이미지) 또는 면역 및 세포 접착 과정(하단 이미지)과 관련된 단백질의 발현 수준을 확인한 결과이다. 4e는 NAC NR 및 R 환자의 GSH 역학(GLS1, GSR 및 GCLM) 또는 CD11c 면역 반응 관련 단백질과 관련된 유전자의 발현에 대한 대표 이미지(X40(상단 이미지) X200(하단 이미지) 배율, 스케일 바 = 100 μm). 본 도면에서 정량적 결과는 평균 $\pm$ SEM의 도트 플롯으로 표시된다 (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

0.001, unpaired Student's t-test.). p-값과 반복 횟수는 소스 데이터 세트에 지정된다.

도 5a 내지 5e는 NAC 내성 MIBC에서 GLS1 단백질의 상향 조절을 확인한 이미지이다. 도 5a는 본 발명에 따라 IHC 이미지에서 종양 상피 세포와 간질을 구별할 수 있는 기계 학습 접근 방식을 사용한 디지털 병리학적 분석의 개략적인 워크플로우이다. 도 5b는 AMC 검증 코호트(NR n = 19, R n = 34)의 종양성 상피 세포에서 41개 단백질의 디지털 병리학 분석을 위한 버블 플롯이다. 5c는 AMC 검증 코호트의 종양(상단 이미지) 또는 간질(하단 이미지) 구획에서 GLS1, CD11c 및 EOGT 단백질의 발현 수준을 나타낸다. 정량적 결과는 평균 $\pm$ SEM의 도트 플롯으로 표시된다(*p < 0.05, unpaired Student's t-tests). 5d 및 도 5e는 NAC에 대한 반응을 기반으로 한 연구에서 테스트된 임상병리학적 요인 및 추정 41개 단백질의 발현에 대한 단변량 및 다변량 분석결과이다. 5d는 종양성 상피 세포에서 GLS1의 높은(빨간색) 또는 낮은(파란색) 발현을 기반으로 통계적으로 유의한 것으로 확인된 GLS1 및 6개 단백질 발현의 상관관계(GLS1 높은 n = 14, GLS1 낮은 n = 40)를 나타낸 이미지이다. 각 그룹에서 표시된 단백질의 평균 발현 수준은 p-값으로 표시되며, 정확한 p-값과 반복실험 횟수는 소스 데이터세트에 나타났다.

도 6aa 내지 도 6c는 다중 코호트 MIBC 환자의 전사체 데이터세트에 대한 딥러닝 기반 분석 결과이다. 6aa 및 도 6ab는 MIBC 환자 시료에서 NAC 항암 치료 반응성 예측 딥러닝 기반 분석 모델링을 최적화하는데 사용된 분자 분류자 리스트이다. 6b는 여러 환자 코호트와 분자 분류자를 사용하여 수행한 딥러닝 기반 분석의 주요 결과이다. 6c는 딥러닝 기반 분석 주요 결과 중에서 BC p53-like pathway의 성능과 유효성을 나타내는 그림이다.

도 7a 내지 도 7d는 MIBC 환자의 NAC 반응을 예측하기 위한 추정 바이오마커 식별 결과이다. 7a는 딥러닝 기반의 멀티 코호트 전사체 분석 모델링 결과에서 로지스틱 회귀, 랜덤 포레스트, 그리고 의사 결정 트리 모델링 결과를 나타낸 결과이다. 7b는 두 가지 서로 다른 분자 분류자들이 어떻게 NAC NR과 R그룹을 구분하는지 나타낸 결과이다. 7c는 7b에 사용된 분자 분류자들 내에서 중요한 핵심 유전자들의 결정 트리를 보여준 결과이다.

도 8은 MIBC 환자 시료에서 NAC 반응성을 예측하는데 중요한 핵심 유전자들을 도출하고, 이를 유전자들의 단백질 발현 차이를 디지털 병리 분석을 통해 검증한 결과이다.

도 9a 내지 도 9g는 디지털 병리 데이터셋을 활용하여 NAC 반응성을 예측하기 위한 최적의 단백질 조합을 발굴하는 과정과 결과이다. 9a는 전체 분석 프로세스의 개요이다. 9b는 특정한 데이터셋에서 가장 중요한 변수를 선택하고, 불필요한 변수를 제거하여 모델의 성능을 최적화하기 위해 사용된 교차 검증을 통한 재귀적 특징 제거 분석 방법(RFEV)에 대한 설명이다. 9c는 디지털 병리 데이터셋에서 종양 구획의 결과로부터 선정된 NAC 반응성을 구분할 수 있는 단백질 그룹 리스트이다. 9d는 디지털 병리 데이터셋에서 스트로마 구획의 결과로부터 선정된 NAC 반응성을 구분할 수 있는 단백질 그룹 리스트이다. 9e는 9c, 9d의 결과를 통해 도출된 최적 모델의 결과이다. 9f는 도출된 종양 구획의 최적 모델의 결과를 의사 결정 트리 모델을 사용하여 각 변수의 중요도를 나타낸 결과이다. 9g는 도출된 스트로마 구획의 최적 모델의 결과를 의사 결정 트리 모델을 사용하여 각 변수의 중요도를 나타낸 결과이다.

도 10은 본 연구에서 개발한 딥러닝 기반 디지털 병리 모델의 검증 결과이다.

발명을 실시하기 위한 구체적인 내용

[0022] 이하, 본 발명을 상세히 설명한다. 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 후술되어 있는 실시 예들을 참조하면 명확해질 것이다. 그러나 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다. 명세서 전체에 걸쳐 동일 참조 부호는 동일 구성 요소를 지칭한다.

[0023] 다른 정의가 없다면, 본 명세서에서 사용되는 모든 용어(기술 및 과학적 용어를 포함)는 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 공통적으로 이해될 수 있는 의미로 사용될 수 있을 것이다. 또 일반적으로 사용되는 사전에 정의되어 있는 용어들은 명백하게 특별히 정의되어 있지 않는 한 이상적으로 또는 과도하게 해석되지 않는다. 본 명세서에서 사용된 용어는 실시예들을 설명하기 위한 것이며 본 발명을 제한하고자 하는 것은 아니다. 본 명세서에서, 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함한다.

[0024] 본 발명은 글루타티온 대사체 또는 이의 조합을 포함하는, 고형암 환자의 화학요법 내성 진단, 예후 예측용 바이오마커 조성물을 제공한다. 본 발명은 글루타치온 경로가 고형암 환자의 선행화학요법 내성과 연관이 있음을 확인하고, 글루타티온 대사체의 발현량을 기초로 고형암 환자에서 선행화학요법에 대한 내성 여부를 진단하고자

한다.

- [0025] 본 발명에서, "대사체 (metabolite)"는 대사물질 또는 대사산물이라고도 하며, 물질 대사의 중간 생성물 또는 생성물이다. 이러한 대사체는 연료, 구조, 신호전달, 효소에 대한 촉진 및 저해 효과, 그 자신의 촉매 활성(일반적으로 효소에 대한 보조 인자로서), 방어, 다른 생물체와의 상호작용(예: 색소, 방향 화합물, 페로몬)을 포함하는 다양한 기능을 가지고 있다. 1차 대사체는 정상적인 성장, 발생 및 생식에 직접적으로 관여하며, 2차 대사체는 이러한 과정들에 직접적으로 관여하지 않지만, 대개 중요한 생태학적 기능을 갖는다.
- [0026] 상기 대사체를 수득할 수 있는 생물학적 시료는 전혈, 혈장, 혈청, 적혈구, 백혈구(예를 들어 말초 혈액 단핵구), 유관액, 복수, 늑막 유출물(pleural efflux), 수유관액 (nipple aspirate), 림프액(예를 들어 림프절의 파종성 종양 세포), 골수 흡인물, 타액, 뇨(소변), 대변(즉, 배설물), 가래, 기관지 세척액, 눈물, 미세 바늘 흡인물, 임의의 기타 체액, 조직 샘플(예를 들어 종양 조직), 종양 생검(예를 들어 천자 생검), 림프절 (예를 들어 감시(sentinel) 림프절 생검), 종양의 수술적 절제 및 이의 세포 추출물을 포함하며, 바람직하게는 혈장일 수 있다. 또한, 상기 대사체는 대사 및 대사 과정에 의해 생산된 물질 또는 생물학적 효소 및 분자에 의한 화학적 대사작용으로 발생한 물질 등을 포함할 수 있다.
- [0027] 본 발명에서, 글루타티온(Glutathione) 대사체는 GLS1, PSAT1, CBS, GCLC, GCLM, Gluta (GSR), GlnRS (QARS1), GGT7, Perox (PRDX1), PLOD2, RPAP1, RPL9, MITF, CD44v6, CDK1, FZD9, GAD2, PPP2R5A, non-p (b-catenin), B-catenin, SALL4, SOX2, TFCEP2L1, TFEB, ICAM1, TRAF2, TRAF6, IL15RA, AFAP1, CARD16, CD11c, CD73, CD3Z, EBI3 (IL27B), EMX1, DNMT3L, DPH2, EGR2, FYB1, GADD45B, HCFC1R1, KCTD14, MTCH1, OCT2, PCDHB9, PPIL2, RFX7, SLC15A3, TNFAIP8, ANPEP, BDN, EOGT, FOXA1, KIFC2, KIR3DL1, NOTUM, USP2, CCDC6, CK5(KRT5), CK20 (KRT20), GATA3, BHLHE40, C11orf95, REX7, COX17, NTSE, ITGAX, KIAA1671, NT5E, POU2F2, PLEK, POSTN, PPP1R35, THBS2 및 INFAIP8를 포함한다.
- [0028] 본 발명에서 고형암 환자의 선행화학요법 내성 진단에 적합한 글루타티온(Glutathione) 대사체의 조합은, 제1 조합(GLS, IL 15RA, AFAP1, FOXA1) 제2 조합(GLS, PSAT1, MITF, ICAM1, TRAF2, TRAF6, IL 15RA, AFAP1, HCFC1R1, PPIL2, GlnRS (QARS1), Gluta(GSR), KIFC2, KIR3DL 1, non-p (b-catenin)), 제3 조합(GLS, PSAT1, CBS, CD44v6, FOXA1, GCLC, KIFC2, RPAP1, B-catenin, CK5, GATA3), 제4 조합(GLS, PSAT1, CBS, ICAM1, TRAF6, IL15RA, AFAP1, CARD16, CD11c, EGR2, FYB1, GADD45B, HCFC1R1, MTCH1, OCT2, PCDHB9, RFX7, SLC15A3, ANPEP, BDNF, CD3-2, CD44v6, EMX1, EOGT, FZD9, GAD2, GCLC, GCLM, KIFC2, non-p (b-catenin), NOTUM, PLOD2, RPL9, SALL4, TFCEP2L1, TFEB, USP2, GGT7, PPP2R5A, GATA3), 제5 조합(GLS, PSAT1, CBS, IL15RA, TRAF6, RPAP1, B-catenin, AFAP1, KIFC2, ICAM1, TRAF2, PLOD2, TFEB) 및 제6 조합(CBS, EGR2, MTCH1, SALL4, B-catenin), 제7 조합 (CBS, GCLC, SOX2, CK5, CK20) 제8 조합 (CBS, FOXA1, GCLC, RPAP1, B-catenin) 제9 조합 (GLS, CBS, TRAF2, PPIL2, KIFC2], 제10 조합 (CBS, B-catenin, KIFC2, TFEB, MTCH1, SOX2, TRAF2, PPP2R5A, USP2, TRAF6, CK20)을 포함한다 (도 9c 및 도 9d 참조).
- [0029] 본 발명의 바람직한 일 실시예에 따르면, 고형암 환자에서 선행화학요법에 대한 내성이 없는 대조군에 비해 선행 화학요법에 대한 내성이 있는 환자군에서는 GLS1, SLC15A3, TNFAIP8의 발현량이 높게 나타날 수 있다.
- [0030] 본 발명의 바람직한 일 실시예에 따르면, 고형암 환자에서 선행화학요법에 대한 내성이 없는 대조군에 비해 선행 화학요법에 대한 내성이 있는 환자군에서는 EOGT, CD11c, KIFC2, CK5, CK20, CARD16, CD73, DNMT3L, FYB1, HCFC1R1, MTCH1, PFX7의 발현이 감소할 수 있다.
- [0031] 본 발명의 상기 대사체 바이오마커는 웨스턴 블랏, 효소-면역화학 검출법(ELISA), 면역조직화학 염색법, 면역침강 또는 면역형광법 등 글루타티온 발현량 및 활성을 측정하는 알려진 방법에 의하여 검출할 수 있으며, 이를 머신러닝 기법으로 분석하여 발현 정도를 분석할 수 있다.
- [0032] 본 발명에서는 대사체 바이오마커를 검출하기 위해 상기 생물학적 시료를 전처리할 수 있다. 예를 들어, 여과, 증류, 추출, 분리, 농축, 방해 성분의 불활성화, 시약의 첨가 등을 포함할 수 있다.
- [0033] 본 발명에서, "정량 장치"는 생물학적 시료 내에 특정 대사체의 존재 여부뿐 아니라 이의 상대적 또는 절대적 양에 대한 정량적인 수치 정보를 제공하는 장치를 의미한다. 구체적으로는 상기 정량 장치는 크로마토그래피 (chromatography), 질량분석기 (mass spectroscopy; MS), 또는 핵자기 공명 분광 분석기 (nuclear magnetic resonance; NMR)이다.
- [0034] 본 발명에서, "크로마토그래피"는 고성능 액체 크로마토그래피 (High Performance Liquid Chromatography; HPLC), 액체-고체 크로마토그래피 (Liquid-Solid Chromatography; LSC), 종이크로마토그래피(Paper

Chromatography; PC), 박층 크로마토그래피 (Thin-Layer Chromatography; TLC), 기체-고체 크로마토그래피 (Gas-Solid Chromatography; GSC), 액체-액체 크로마토그래피 (Liquid-Liquid Chromatography; LLC), 포말 크로마토그래피 (Foam Chromatography; FC), 유화 크로마토그래피 (Emulsion Chromatography; EC), 기체-액체 크로마토그래피 (Gas-Liquid Chromatography; GLC), 이온 크로마토그래피 (Ion Chromatography; IC), 겔 여과 크로마토그래피 (Gel Filtration Chromatography; GFC) 또는 겔 투과 크로마토그래피 (Gel Permeation Chromatography; GPC)를 포함하나, 이에 제한되지 않고 당업계에서 통상적으로 사용되는 모든 정량용 크로마토그래피를 사용할 수 있다.

[0035] 본 명세서에서 "질량 분석 (MS)"은 시료의 화학적 조성을 분석하기 위해 대상 물질의 질량을 측정하는 과정을 의미한다. 질량 분석은 시료에 존재하는 대상 물질의 이온화를 통해 하전된 분자나 분자조각을 생성하고 질량 대 전하비(m/z) 및 기체상 이온의 존재 비를 측정하여 질량에 대한 정보를 제공한다. 이러한 질량 분석 장치는 예를 들어 MALDI-TOF (Matrix-Assisted Laser Desorption/Ionization Time of Flight), SELDI-TOF (Surface Enhanced Laser Desorption/Ionization Time of Flight), ESI-TOF (Electrospray ionization Time of Flight), 액체 크로마토그래피-질량분석기 (liquid chromatographyMass Spectrometry; LC-MS) 또는 LC-MS/MS (liquid chromatography-Mass Spectrometry/Mass Spectrometry)를 포함하나, 이에 제한되는 것은 아니다.

[0036] 본 발명에서의 용어 "진단"은 특정 질병 또는 질환에 대한 한 객체의 감수성(susceptibility)을 판정하는 것, 한 객체가 특정 질병 또는 질환을 현재 가지고 있는지 여부를 판정하는 것, 특정 질병 또는 질환에 걸린 한 객체의 예후(prognosis)를 판정하는 것, 또는 테라메트릭스(therametrics), 예컨대, 치료 효능에 대한 정보를 제공하기 위하여 객체의 상태를 모니터링하는 것을 포함한다.

[0037] 본 발명에서, 상기 고형암은 방광암(Bladder cancer), 대장암(Colon cancer), 위암(Stomach cancer), 폐암(Lung cancer), 폐선암(Lung adenocarcinoma), 유방침습성유관암종(Breast invasive ductal carcinoma), 결장선암(Colon adenocarcinoma), 전립선선암(Prostate adenocarcinoma), 방광요로피암종(Bladder urothelial carcinoma), 폐편평세포암종(Lung squamous cell carcinoma), 피부흑색종(Cutaneous melanoma), 원발부위불명암(Cancer of unknown primary), 췌장도관선암종(Pancreatic adenocarcinoma), 교모세포종(Glioblastoma multiforme), 대장선암(Colorectal adenocarcinoma), 고등급장액성난소암(High grade serous ovarian cancer), 위선암(Stomach adenocarcinoma), 신세포암종(Renal clear cell carcinoma), 식도암(Esophageal adenocarcinoma), 고환암(Testicular cancer) 및 간내담관암(Intrahepatic cholangiocarcinoma)을 포함하나, 이에 제한되는 것은 아니다.

[0038] 본 발명에서 '환자'는 통상 인간을 포함할 뿐 아니라 다른 동물, 예를 들어 다른 영장류, 설치류, 개, 고양이, 말, 양, 돼지 등을 포함할 수 있다. 본 발명의 '환자'는 고형암으로 판정되거나, 의심되는 인간을 제외한 대상을 포함한다.

[0040] 본 발명에서, 상기 화학요법 내성은 선행화학요법으로 이용되는 항암제(특히, 백금제제(백금 착화합물계 항암제))에 대한 내성을 의미하나, 이에 제한되는 것은 아니다. 본 발명에서, 상기 백금제제에는 헵타플라틴, 네다플라틴, 보플라틴 등 포함되며, 바람직하게는 시스플라틴을 의미한다.

[0042] 이하, 본 발명의 이해를 돕기 위하여 실시예를 들어 상세하게 설명하기로 한다.

실시예 1

[0043] 실험방법

[0044] 1-1. 데이터 및 코드 유효성

[0045] 서울아산병원(AMC) 코호트 환자의 모든 원시 및 처리된 전사체 데이터는 임상 주석과 함께 NCBI Gene Expression Omnibus에 기탁되었으며 GEO 시리즈 등록 번호 GSE212810을 통해 접근할 수 있다. 3개의 외부 코호트에 대한 전사체 데이터 세트 및 NAC에 대한 병리학적 반응은 등록 번호(accession number) GSE48277(MDA MVAC), GSE69795(MDA DDMVAC) 및 GSE87304(NAC 메타데이터)로 Gene Expression Omnibus 데이터베이스에서 검색되었다. "NAC 메타데이터" 코호트의 병리학적 반응 기록은 Peter C. Black 박사 (University of British Columbia, Canada)로부터 제공받았다. AMC 검증 및 외부 코호트의 GSEA 분석에 사용된 유전자 세트에 대한 자체

한 정보는 하기 표 1에 나타나있다. 유전자 네트워크, 생체 기능 및 표준 경로와 관련하여 전사체 데이터 세트의 기능 분석에서 GO 카테고리 및 해당 유전자의 자세한 목록은 하기 표 2에 표시된다.

[0046] AMC 코호트의 디지털 병리학 분석을 위해 종양 상피와 간질을 구별하는 머신러닝 분류기를 개발하기 위한 원본 모델은 하기 표 2에 표시된다. 관련 임상 주석과 함께 종양/간질 분류기를 적용한 디지털 병리학에 의한 종양 및 간질 구획의 각 단백질에 대한 발현 데이터 역시 하기 표 2에 표시된다. 또한 연구에 사용된 소스 코드와 구성된 종양/간질 분류기는 <https://doi.org/10.5281/zenodo.7021255>에서 확인할 수 있다.

표 1

항체		
명칭	SOURCE	식별자(Identifier)
ANPEP (CD13)	Santa cruz	sc13536
BDNF	Invitrogen	PA595183
BICDL1 (CCDC64)	Invitrogen	PA566367
CBS	Origene	TA308071
CD11c	NOVUS	NBP2-44599
CD247 (CD3-zeta)	Santa cruz	sc1239
CD44v6	Abcam	ab78960
CDK1	Abcam	ab131450
CK5/6	DAKO	M7237
CK20	DAKO	M7019
EMX1	Santa cruz	sc398115
EOGT	Invitrogen	PA553990
FOXA1	Santa cruz	sc101058
FZD9	Invitrogen	PA527119
GAD2	Santa cruz	sc377145
GATA3	Cell marque	390M-16
GCLC	Abcam	ab190685
GCLM	Abcam	ab81445
GGT7	Invitrogen	PA553035
GLS1	Abcam	ab156876
GSR	Abcam	ab128933
ICAM1	Invitrogen	MA5407
IL15RA	NSJ Bioreagents	R32936
IL27B (EBI3)	Invitrogen	MA524805
KIFC2	NOVUS	NBP2-17061
KIR3DL1	Invitrogen	PA5102443
MITF	Invitrogen	PA5-82074
Non-phosphorylated β -Catenin	Cell signaling	4270S
PLOD2	Protein tech	21214-1-AP
PPP2R5A	Santa cruz	sc271151
PRDX1	Abcam	ab15571
PSAT1	NOVUS	NBP1-32920
QARS1(GlnRS)	Santa cruz	sc271078
RPAP1	Invitrogen	PA559703
RPL9	Santa cruz	sc100828
SALL4	Biocare	CM384A,C
SOX2	Abcam	ab92494
TFCP2L1-middle	Aviva	OAAB09683
TRAF2	Santa cruz	sc7346
TRAF6	Santa cruz	sc8409
β -Catenin	Santa cruz	sc7199
올리고뉴클레오타이드		
Human ANPEP_F	CGGGCCACGGCGATT	서열번호 1
Human ANPEP_R	GGAGTGGGTAGGGTGTGCATAA	서열번호 2
Human BDNF_F	CTACGAGACCAAGTGAATCCC	서열번호 3
Human BDNF_R	AATCGCCAGCCAATTCTCTTT	서열번호 4
Human BICDL1_F	AGCTGCTCACAAACCGATTCA	서열번호 5

Human BICDL1_R	GTACATTCTTGCCGCGCATC	서열번호 6
Human B2MG_F	GAGGGCTGGCAACTTAGAGG	서열번호 7
Human B2MG_R	ACAAGCTTTGAGTGCAAGAGA	서열번호 8
Human CBS_F	CCCCCTGGCTCACTACGA	서열번호 9
Human CBS_R	CACTGAAGCCACCAGCATGT	서열번호 10
Human CD11c_F	CCGACCATATCTGCCAGGAC	서열번호 11
Human CD11c_R	GCCCTTCAGGGTGAAATCCA	서열번호 12
Human CD247_F	GCCAGAACCAGCTCTATAACGA	서열번호 13
Human CD247_R	GGCCACGTCTCTTGTCCTAA	서열번호 14
Human CD44_F	CTGCCGCTTTGCAGGTGTA	서열번호 15
Human CD44_R	CATTGTGGGCAAGGTGCTATT	서열번호 16
Human CDK1_F	AAACTACAGGTCAAGTGGTAGCC	서열번호 17
Human CDK1_R	TCCTGCATAAGCACATCCTGA	서열번호 18
Human CK14_F	TGAGCCGATTCTGAACGAG	서열번호 19
Human CK14_R	GATGACTGCGATCCAGAGGA	서열번호 20
Human CK20_F	GGACGACACCCAGCGTTTAT	서열번호 21
Human CK20_R	CGCTCCCATAGTTCACCGTG	서열번호 22
Human CK5_F	CCAAGGTTGATGACTGATGG	서열번호 23
Human CK5_R	TGTCAGAGACATGCGTCTGC	서열번호 24
Human CTNNB1_F	CATCTACACAGTTTGATGTGCT	서열번호 25
Human CTNNB1_R	GCAGTTTGTTCAGTTCAGGGA	서열번호 26
Human EMX1_F	CACGAAGCAGGCCAATGG	서열번호 27
Human EMX1_R	CTCTGCCCTCGTGGGTTTGT	서열번호 28
Human FOXA1_F	GCAATACTCGCCTTACGGCT	서열번호 29
Human FOXA1_R	TACACACCTTGGTAGTACGCC	서열번호 30
Human FZD9_F	TGCGAGAACCCCGAGAAGT	서열번호 31
Human FZD9_R	GGGACCAGAACACCTCGAC	서열번호 32
Human GAD2_F	ATTGGGAATTGGCAGACCAAC	서열번호 33
Human GAD2_R	TTGAAGTATCTAGGATGCCCTGTG	서열번호 34
Human GATA3_F	GCCCCCTCATTAAGCCCAAG	서열번호 35
Human GATA3_R	TTGTGGTGGTCTGACAGTTCG	서열번호 36
Human GCLC_F	GGAGGAAACCAAGCGCCAT	서열번호 37
Human GCLC_R	CTTGACGGCGTGGTAGATGT	서열번호 38
Human GCLM_F	TGTCTTGGAATGCACTGTATCTC	서열번호 39
Human GCLM_R	CCCAGTAAGGCTGTAAATGCTC	서열번호 40
Human GGT7_F	CTACAGTACAGCCAGGCAGG	서열번호 41
Human GGT7_R	GAAGTGAAGACAAGGCCCA	서열번호 42
Human GLS1_F	GCATTCCTGTGGCATGTATGACTT	서열번호 43
Human GLS1_R	CCCCCAGCAACTCCAGATTT	서열번호 44
Human GPX1_F	CCAGTTTGGGCATCAGGAGA	서열번호 45
Human GPX1_R	AGCATGAAGTTGGGCTCGAA	서열번호 46
Human GPX2_F	GACTTCACCCAGCTCAACGA	서열번호 47
Human GPX2_R	ATGCTCGTTCTGCCCATTC	서열번호 48
Human GPX4_F	CAGTGAGGCAAGACCGAAGT	서열번호 49
Human GPX4_R	CGGTGTCCAACTTGGTGAAG	서열번호 50
Human GSR_F	TTCCAGAATACCAACGTCAAAGG	서열번호 51
Human GSR_R	GTTTTCGGCCAGCAGCTATTG	서열번호 52
Human HS3ST6_F	ATCTCCGACTACGCCAGAC	서열번호 53
Human HS3ST6_R	CGTGACGACCCGTTTCAGG	서열번호 54
Human ICAM1_F	CGGCTGACGTGTGCAGTAATA	서열번호 55
Human ICAM1_R	GGCGCCGAAAAGCTGTA	서열번호 56
Human IL15RA_F	CCCAGCTCAAACAACACAGC	서열번호 57
Human IL15RA_R	AGGTAGCATGCCAGGAGAGA	서열번호 58
Human IL27B_F	ATCCGTTACAAGCGTCAGGG	서열번호 59
Human IL27B_R	TCCCGTAGTCTGTGAGGTC	서열번호 60
Human KIFC2_F	CGCCTTTTACTCGTTGCTCA	서열번호 61
Human KIFC2_R	CTGTCAACAGTGAGGGGACC	서열번호 62
Human KIR3DL1_F	CCAGGTCCCCTGGTGAAATC	서열번호 63

Human KIR3DL1_R	CGCTGTTGGCTGTTCTGTTC	서열번호 64
Human MITF_F	AGGGAGCTCACAGCGTGTAT	서열번호 65
Human MITF_R	AGGTCTTGGCTGCAGTTCTC	서열번호 66
Human PLOD2_F	CATGGACACAGGATAATGGCTG	서열번호 67
Human PLOD2_R	AGGGGTTGGTTGCTCAATAAAAA	서열번호 68
Human PPP2R5A_F	TGCCAATTATGTTTGCCAGTTT	서열번호 69
Human PPP2R5A_R	TCCATTAGGGTTTTTCAGCACATT	서열번호 70
Human PRDX1_F	CATTCTTTTGGTATCAGACCCG	서열번호 71
Human PRDX1_R	CCCTGAACGAGATGCCTTCAT	서열번호 72
Human PSAT1_F	GGCCAGTTCAGTGTGTCC	서열번호 73
Human PSAT1_R	GCTCCTGTCACCACATAGTCA	서열번호 74
Human QARS1_F	CGGCGTCTCTCCTTCCTTGT	서열번호 75
Human QARS1_R	TACTCAAGGGCAGCGCTTAGC	서열번호 76
Human RPAP1_F	TCCTGGGAGCAGGTTGTTTG	서열번호 77
Human RPAP1_R	AGACCTTCAGAAGCCCCAGA	서열번호 78
Human RPL9_F	TGCTCACTTCCCCATCAACG	서열번호 79
Human RPL9_R	GCTTGCTGAATCAAAGCCG	서열번호 80
Human SALL4_F	TGCGGAGTCTGTGGTGTACCTA	서열번호 81
Human SALL4_R	GATTCACCGCCACCTTGG	서열번호 82
Human SIRT6_F	GCAGTCTTCCAGTGTGGTGT	서열번호 83
Human SIRT6_R	GATAGAGCCGTTGATCCGGG	서열번호 84
Human SOX2_F	AACCAGCGCATGGACAGTTA	서열번호 85
Human SOX2_R	GACTTGACCACCGAACCCAT	서열번호 86
Human TFCP2L1_F	GCTCTTCAACGCCATCAAA	서열번호 87
Human TFCP2L1_R	CAGGGGCACTCGATTCTG	서열번호 88
Human TRAF2_F	GGAGGCATCCACCTACGATG	서열번호 89
Human TRAF2_R	GGGAGAAGATGGCGGGTATG	서열번호 90
Human TRAF6_F	CTGCTTGATGTCATTACGAGAA	서열번호 91
Human TRAF6_R	TGCAGGCTTTCAGAACCTA	서열번호 92

표 2

생물학적 샘플		
방광암 환자 샘플	This study	IRB; AMC (2020-0064)
중요한 상업적 분석		
SMARTer Stranded Total RNA-Seq Kit v2	Takara	634411
Taqman Reverse Transcription Reagents	Applied Biosystems	N8080234
ultraView Universal DAB detection kit	Roche	760-500
기탁된 데이터		
AMC 코호트용 전사체 데이터 세트	This paper	GSE212810
MDA MVAC 코호트용 전사체 데이터 세트	Choi et al., 2014	GSE48277
MDA DDMVAC 코호트	McConkey et al., 2016	GSE69795
NAC 메타데이터 코호트용 전사체 데이터 세트	Seiler et al., 2017	GSE87304
소프트웨어 및 알고리즘		
GraphPad Prism 7.0	GraphPad Software	N/A
MetaCore	Clarivate Analytics	N/A
GSEA	Broad Institute	N/A
ExDEGA	Ebiogen Inc.	N/A
QuPath	Bankhead et al., 2017	N/A
Manuscript analysis workflow	This study	https://doi.org/10.5281/zenodo.7021255
기타		
BenchMark XT Automated IHC/ISH slide staining system	Roche	N/A
Vectra 3.0 Automated Quantitative Pathology Imaging System	Perkin-Elmer	CLS142568

Clinical annotations	This study	Table S10 and S11
Cryostat	Leica	RM2125 RTS

1-2. 환자 샘플 수집 및 코호트 조직

본 연구는 AMC 기관 검토 위원회(2020-0064)의 승인을 받았다. 본 발명에는 2011년 8월부터 2017년 8월 사이에 AMC에서 TUR을 받은 후 NAC, 이후 이용 가능한 FFPE 종양 조직을 이용한 치료 수술을 받은 63명의 연속 MIBC(cT2-4aNOMO) 환자가 포함되었다. 모든 환자는 수술 표본을 사용하여 병리학적 반응을 평가할 수 있었기 때문에 검증 코호트(validation cohort)로 지정되었다. 63명의 환자 중 "불완전한 TUR"로 기록되고 병리학적 반응에서 현저한 대조를 보인 20명의 환자가 발견 코호트(discovery cohort)로 선택되었다. NAC, 수술 TUR 기록, 종양 진행 및 생존 시간을 포함한 환자의 임상 정보는 전자 의료 기록에서 얻었다.

본 발명의 모든 병리학적 자료는 비뇨기과 병리학자(Y.M.C.)에 의해 검토되었으며 '2016년 세계보건기구 종양 분류'에 따라 재평가되었고, 'American Joint Committee on Cancer Staging System, 8th edition'에 따라 단계가 결정되었다. NAC 반응은 근처 또는 부분 방광절제 표본의 병리학적 검사에서 침윤성 종양이 없거나 림프절 전이(yoN0)가 없는 상태에서 ypT1 이하로 단계가 하향되는 것으로 정의되었다. 병리학적 반응의 결여는 고유근층 이상을 침범한 잔여 종양 또는 림프절 전이($\geq$ ypT2 및/또는 yoN0)로 정의되었다.

1-3. Lase capture microdissection (LCMD) 및 RNA 분리

LCMD는 제조사가 설명한대로, Leica LMD6500 레이저 포획 해부 현미경(Leica Microsystems, Deerfield, IL, USA)을 사용하여 헤마톡실린-에오신 염색 조직 슬라이드에서 수행되었다. 피사성 종양 부위를 피하면서 간질 및 염증 세포의 오염을 피하기 위해 요로상피암종 부위를 종양 세포 주변에 최대한 가깝게 꼼꼼하게 미세 절개했다. LCMD 샘플의 총 RNA는 RNeasy FFPE 키트(73504, QIAGEN, Valencia, CA, USA)를 사용하여 분리되었다. RNA 품질은 RNA 6000 Nano Chip(G2939BA, Agilent Technologies, Amstelveen, The Holland)을 사용하여 Agilent 2100 bioanalyzer로 평가하고 ND-2000 Spectrophotometer(Thermo Fisher Scientific, Waltham, MA, USA)를 사용하여 RNA 정량을 수행했다.

1-4. 라이브러리 준비 및 시퀀싱

분리된 RNA로부터 라이브러리 구축은 제조업체의 지침에 따라 QuantSeq 3' mRNA-Seq Library Prep Kit(Lexogen, Inc., Austria)를 사용하여 수행되었다. 각각의 총 RNA를 준비하고, 5' 말단에 Illumina 호환 서열을 포함하는 올리고-dT 프라이머를 RNA에 혼성화하고 역전사를 수행했다. RNA 주형이 분해된 후, 5' 말단에 Illumina 호환 링커 서열이 포함된 무작위 프라이머에 의해 두 번째 가닥 합성이 시작되었다. 모든 반응 성분을 제거하기 위해 이중 가닥 라이브러리를 자기 비드를 사용하여 정제했다. 클러스터 생성에 필요한 전체 어댑터 시퀀스를 추가하기 위해 라이브러리가 증폭되었다. 완성된 라이브러리는 PCR 구성 요소로부터 정제되었다. 높은 처리량 시퀀싱은 NextSeq 500(Illumina, Inc., San Diego, CA, USA)을 사용하여 싱글-엔드 75 시퀀싱(single-end 75 sequencing)으로 수행되었다.

1-5. 전사체 데이터 세트 분석

QuantSeq 3' mRNA-Seq 판독은 Bowtie2를 사용하여 정렬되었다. Bowtie2 인덱스는 게놈 조립 서열 또는 게놈과 전사체 정렬을 위한 대표적인 전사체 서열로부터 생성되었다. 정렬 파일은 전사체 조립, 존재비 추정 및 차등 유전자 발현 검출에 사용되었다. 차별적으로 발현된 유전자는 Bedtools의 적용 범위를 사용하여 고유하고 다중 정렬의 개수를 기반으로 결정되었다. RC(Read Count) 데이터는 Bioconductor를 사용하여 R EdgeR 패키지(R development Core Team, 2020)를 사용하여 Quantile 정규화 방법을 기반으로 처리되었다. 유전자 분류는 DAVID(david.abcc.ncifcrf.gov/) 및 Medline 데이터베이스(www.ncbi.nlm.nih.gov/)에서 수행한 검색을 기반으로 했다. 데이터 마이닝(Data mining) 및 그래픽 시각화는 ExDEGA(Ebiogen Inc., 대한민국 서울)를 사용하

여 수행되었다.

1-6. MetaCore 및 GSEA 전사체 분석

기능적 전사체 분석은 기본 설정으로 MetaCore(Clarivate Analytics, Philadelphia, PA, USA) 또는 GSEA(Broad Institute, Cambridge, MA, USA) 소프트웨어를 사용하여 수행되었으며, 이는 유전자 네트워크, 생체 기능 및 표준 경로에 대한 핵심 분석과 유전자 세트의 강화를 각각 제시한다. MetaCore 분석에서는 $p < 0.05$ 인 1.5배 상향 또는 하향 조절이 중요한 변화에 대한 컷오프 값으로 정의되었다. MetaCore 분석의 생물학적 과정과 경로에 대한 통계값과 중요한 유전자에 대한 자세한 내용은 표 3에 표시되어 있다. GSEA에 사용된 유전자 세트는 출판된 문헌을 기반으로 한 분류자 또는 여러 라이브러리 소스를 사용하여 선별된 기능적 유전자 세트(C2) 데이터베이스에서 얻었다. FDR(false discovery rate) < 0.25 를 기준으로 유의미한 차이가 결정되었다.

표 3

Maps	$-\log_{10}$ (p-value)	Network objects
세포주기_세포주기 조절에서 Nek의 역할	5.618	Tubulin alpha, Nek8, Histone H3, NEK9, Histone H1, MAD2a, Ran
HCC로 이어지는 HCV 의존성 세포질 신호 전달	3.783	ERK1/2, ERK2 (MAPK1), PKC, PKR, eIF2S1
STK3/4(Hippo) 경로의 조절 및 YAP/TAZ 기능의 긍정적인 조절	3.673	ERK1/2, Nek8, RASSF6, G-protein alpha-q/11, HIPK2, ERK1 (MAPK3), MASK
면역반응_염증반응의 가스트린	3.422	ERK2 (MAPK1), G-protein alpha-q/11, G-protein alpha-q, ERK1 (MAPK3), TRAF6, PKC-delta, MAP2K5 (MEK5)
면역반응_BAFF에 의한 신호전달	3.316	ERK1/2, ERK2 (MAPK1), MAPKAPK2, TRIM2, ERK1 (MAPK3), PKC-delta
DNA 손상_G2/M 체크포인트의 ATM/ATR 조절: 세포질 신호전달	3.316	ERK2 (MAPK1), MAPKAPK2, Histone H3, 14-3-3, PP2A regulatory, ATM
면역반응_NF- κ B를 통한 리소포스파티드 산 신호전달	3.270	ERK1/2, TRIP6, TRAF6, PKC, PKC-delta, ICAM1
면역 반응_MAPK를 통한 IFN-알파/베타 신호 전달	3.233	ERK1/2, Filamin B (TABP), MAPKAPK2, ISG15, PRMT1, PKC-delta, PKR
세포사멸과 생존_스트레스 유발 세포사멸에서 PKR의 역할	3.225	ERK1/2, PPP2R5A, PP2A regulatory, TRAF6, PKR, eIF2S1
위점막 분화에 대한 가스트린	3.066	ERK1/2, TFF1, G-protein alpha-q/11, G-protein alpha-q, PKC

1-7. 조직 마이크로어레이(TMA) 구성 및 면역조직화학(IHC) 분석

TMA 블록은 조직 마이크로어레이(Beecher Instruments, Silver Spring, MD, USA)를 사용하여 NAC 이전에 수확된 TUR 표본의 FFPE 조직 블록으로 구성되었다. 종양 내 이질성을 극복하기 위해 각 종양의 서로 다른 영역에서 직경 0.6mm의 대표 코어 3개를 포함했다. 종양이 다양한 조직학적 등급 및 단계의 영역을 포함하는 경우, TMA는 가장 높은 등급 및 단계의 종양 영역에서 생성되었다.

IHC 염색은 자동 염색 시스템(BenchMark XT, Ventana Medical Systems, Tucson, AZ, USA) 및 ultraView Universal DAB 검출 키트(Ventana Medical Systems)를 사용하여 TMA 구조물에서 수행되었다. 핵은 헤마톡실린으로 대비염색되었다. 자세한 항체 정보 및 염색 조건은 상기 표 1 및 표 2에 표시되어 있다.

1-8. IHC 염색 이미지의 전산 분석(Computational analysis)

IHC 슬라이드는 Pannoramic 250 Flash 슬라이드 스캐너(3D HISTECH, 부다페스트, 헝가리)를 사용하여 픽셀당 0.22mm의 해상도로 X20 배율로 스캔하고 생체 이미지 분석을 위한 오픈 소스 소프트웨어 플랫폼인 QuPath를 기반으로 컴퓨터 분석을 실시했다.

[0074] 상피 및 간질 구획의 발현 강도를 개별적으로 평가하기 위해 본 발명자들은 3,3'-디아미노벤지딘(DAB)으로 염색되고 헤마톡실린으로 대조염색된 면역조직화학적 이미지에서 종양성 상피 세포와 간질을 구별하는 기계 학습 분류기를 개발했다. 먼저, 훈련, 테스트 및 검증 세트(발견 코호트의 CBS 및 CD73, 이 연구에서 분석되지 않은 고급 BC 코호트의 CD11c 및 CDK1)를 구성하기 위해 대표적인 IHC 슬라이드 이미지를 선택했다. QuPath를 사용하여 이미지를 열고 종양성 상피 세포와 간질 영역에 주석을 달았다. 주석이 달린 영역은 기본 매개변수를 사용하여 QuPath의 세포 감지 기능에 따라 분할되었다. QuPath 기능은 특정 영역의 세포를 식별하기 위한 것이다. 딥러닝 기반 접근 방식을 사용하여 감지된 세포를 분류하기 위해 본 발명자들은 IHC 슬라이드의 세포 중심을 중심으로 224 x 224 픽셀 크기의 패치를 추출하는 QuPath 스크립트를 작성했다. 그 결과, "종양"(96,149개 패치) 또는 "간질"(127,772개 패치)로 표시된 224,101개 패치를 얻었다. 머신러닝용 Python 라이브러리 패키지인 Scikit-learn의 StratifiedShuffleSplit 방법을 사용하여 패치를 "종양"과 "간질"의 비율은 보존하면서, 훈련 세트(40%), 검증 세트(30%), 테스트 세트(30%)로 무작위로 분할했다. 본 발명자들은 오픈 소스 기계 학습 프레임워크인 PyTorch에서 제공하는 사전 훈련된 모델을 훈련하여 모델을 구축했다. 사전 훈련된 매개변수를 시작점으로 하는 이미지 분류를 위한 최첨단 컨벌루션 신경망 모델인 ResNet34를 사용하여 정확도 0.96의 모델을 얻었다.

[0075] 추론을 위해 각 IHC 슬라이드에서 고유 ID 번호로 명명된 셀 패치를 추출했다. 패치를 분류하기 위해 딥러닝 모델을 적용하고, 세포 ID와 분류 결과로 구성된 테이블을 "종양" 또는 "간질"로 내보내도록 Python 스크립트를 작성했다. 테이블을 로드하고 추론 결과를 시각화하기 위해 QuPath 스크립트가 작성되었다. 종양성 상피 세포 및 간질의 주석을 포함한 모델 구축은 병리학자(S.Y.Y.)에 의해 수행되었지만 분류 결과는 다른 병리학자(J.A.J.)에 의해 검증되었다. 필요한 경우 잘못 분류된 영역을 적절하게 수정했다.

[0076] 단백질 발현 수준을 정량화하기 위해 QuPath에서 측정값 'Cell:DAB OD 평균'으로 표시되는 각 세포의 DAB 염색 강도를 내보냈다. TMA 코어의 경우 종양 상피와 간질의 단백질 발현 수준은 각각 "종양"과 "간질"로 분류된 세포의 DAB 강도를 평균하여 별도로 계산되었다. 종양 당 3개의 TMA 코어가 있었기 때문에 이후에 3개의 코어의 평균 결과를 얻어 종양 발현 수준을 나타냈다(도 5a).

[0078] 1-9. 임상 MIBC 환자 코호트 분석

[0079] AMC 코호트의 MIBC 환자에 대한 임상병리학적 분석은 SPSS 버전 21.0을 사용하여 수행되었다. 모든 연속형 데이터는 Student's t-test를 이용하여 비교하였고, 범주형 데이터는 Pearson's Chi-square, Fisher's Exact, Kruskal-Wallis test를 이용하여 비교하였다. 범주형 IHC 발현 데이터는 휴리스틱 방법으로 결정된 컷오프 값을 기준으로 표시된 바이오마커의 높은 발현 대 낮은 발현으로 분류되었다. NAC 반응의 단변량 분석을 수행하여 각 단백질 발현 및 임상병리학적 매개변수의 임상적 중요성을 결정했다. 단변량 분석에서 통계적으로 유의미한 변수는 이항 로지스틱 회귀분석을 적용한 다변량 분석으로 분석되었다. 독립 변수는 역방향 단계적 선택을 통해 선택되었으며, p-값 < 0.05는 통계적으로 유의한 것으로 간주되었다.

[0081] 1-10. 정량화 및 통계 분석

[0082] 본 발명에서 모든 정량적 데이터는 평균 $\pm$ 평균의 표준 오차(SEM)로 표시된다. 3개 이상의 그룹에 대한 통계적 유의성은 Bonferroni 사후 테스트를 통한 일원 분산 분석(ANOVA)을 사용하여 결정되었다. 두 그룹 간의 통계적 유의성은 짝을 이루지 않은 Student's t-test를 사용하여 분석되었다. GraphPad Prism 7.0 (GraphPad Software, La Jolla, CA, USA) 또는 SPSS 버전 21.0 소프트웨어를 사용하여 통계 분석을 수행했다. *p < 0.05, **p < 0.01, ***p < 0.001에 대해 중요성(Significance)이 가정되었다.

[0084] 1-11. 추가 리소스

[0085] 사용된 테스트, 샘플 크기 및 정확한 p-값을 포함한 원시 데이터 및 통계 분석의 모든 정량화는 상기 표 1 및 표 2에 표시된다. RNA-seq 분석 및 관련 임상 주석에 사용되는 환자 사례의 명명법과 함께 종양 및 간질 구획에 대한 원시 디지털 병리학 데이터가 제공된다. 디지털 병리학 및 종양/기질 분류기 구성에 대한 소스 코드는 <https://doi.org/10.5281/zenodo.7021255>에서 확인할 수 있다. DOI(Digital object identifier)는 주요 리소스 테이블에 나열되어 있다.

실시예 2

결과

2-1. MIBC NAC 반응을 기반으로 한 전사체 프로파일링

경요도 절제술(TUR)과 NAC를 후속 근치적/부분 방광절제술로 시행한 63명의 MIBC(cT2-4aN0M0) 환자의 샘플을 후향적으로 수집했다(표 S1). 해당 사례 중 20명의 환자는 TURBT 수술 기록에 "불완전한 TURBT"가 기록되어 AMC 발견 코호트에 배정되었으며, 이는 TUR 후에도 잔류 종양이 남아 있음을 나타내, NAC 이전에 TUR에 의한 완전한 종양 제거 가능성을 배제하고 NAC에 대한 치료 반응의 평가를 가능하게 하였다. 발견 코호트는 NAC 반응에 현저한 차이가 있는 두 그룹으로 분류되었다: 근치/부분 방광절제술 표본에서 잔여 침습성 종양이 없는 NAC 반응환자(R) 8명, 및 잔여 MIBC(4.9 ± 2.5 cm)가 큰 NAC 무반응 환자(NR) 12명. NAC 반응의 간질이 없는 종양 특이적 분자 특성을 조사하기 위해 우리는 16명의 환자($n = 8/\text{그룹}$, NR 및 R)을 발견 코호트에서 추출한 후 게놈 전체 전사체 프로파일을 비교했다(도 2a). 주성분 분석(PCA)에 따르면 전체 전사체는 NR 그룹과 R 그룹 간에 거의 차이가 나타나지 않았다(도 2b). $|\text{fold change}| > 1.5$ 및 $p < 0.05$ 의 컷오프 값을 사용하여 우리는 571개의 차별적으로 발현된 유전자(DEG)를 식별했으며, R 그룹에 비해 NR 그룹에서는 381개가 상향 조절되고 190개가 하향 조절되었다(도 2c).

본 발명자들은 NR 및 R 그룹의 전사체가 MIBC 분자 하위 유형을 정의하는 발표된 분자 분류와 상관관계가 있는지 여부를 조사하기 위해 유전자 세트 농축 분석(GSEA)을 수행했다. 이전 연구 결과와 일치하여 NAC 민감성 종양은 기저 MIBC 하위 유형에서 상승한 p63 경로와 마찬가지로 기저 및 기저 편평상피(Ba/Sq) 분류자(도 2D)가 풍부했다. 그러나 NAC 내성 종양은 발표된 MIBC 분자 분류자(classifier)와 일치하지 않았다.

따라서 AMC 발견 코호트의 종양 전사체는 MetaCore 및 GSEA 분석 방법을 사용하여 유전자 네트워크, 생체 기능 및 표준 경로에 대해 추가로 특성화되었다. 유전자 존재론(Gene Ontology, GO) 분석에 따르면 NAC NR 및 R 그룹의 종양은 Hippo 및 YAP/TAZ 기능, 세포 주기, 염증 및 면역 반응 경로(도 2e) 및 여러 대사 과정에 관련된 유전자의 차등적 발현을 나타냈다. MetaCore 유전자 네트워크 분석에 따르면 대사 과정과 관련된 유전자는 NR 그룹 종양에서 풍부하고 면역 과정은 R 그룹 종양에서 풍부해졌다(도 2f).

일관되게 GSEA는 Valine_leucine_isoleucine(VLI) 아미노산(AA) 분해($\text{NES} = 1.72$; $\text{FDR} = 0.143$), 시토크롬 P450(CYP) 약물 대사($\text{NES} = 1.53$; $\text{FDR} = 0.183$)를 포함한 세포 대사 관련 유전자 세트 및 GSH 대사($\text{NES} = 1.54$; $\text{FDR} = 0.166$)는 NR 그룹 종양의 전사체가 상당히 풍부했다(도 2g). 대조적으로, R 그룹 종양이 풍부한 유전자 세트의 대부분은 사이토카인($\text{NES} = -2.19$; $\text{FDR} = 0.000$) 및 케모카인($\text{NES} = -1.660$; $\text{FDR} = 0.051$) 신호 전달 및 자연 살해 세포 매개 세포 독성($\text{NES} = -1723$; $\text{FDR} = 0.036$)을 포함하여 면역 반응에 관여하는 것들(도 2g)이었다.

2-2. NAC 내성 MIBC에서 GSH 관련 대사 과정이 상향 조절 확인

본 발명자들은 상기의 결과를 추가로 확인하기 위해, 63명의 MIBC 환자(cT2-4aN0M0)를 검증 코호트로 사용했다. 실험에서 모든 환자는 근치적 또는 부분적 수술을 받았으므로 병리학적 반응을 평가할 수 있었다. NAC($\leq \text{ypT1}$ 및 ypN0)에 대한 병리학적 반응은 37명의 환자(57.7%)에서 나타났다. 검증 코호트의 63개 사례 중 FFPE TURBT 표본에서 종양의 LCMD를 수행하여 39명의 환자(NR, $n = 18$, R, $n = 21$)의 전사체 데이터 세트를 수집했다. 발견 코호트(discovery cohort)를 특성화하기 위해 30개의 유전자 세트를 사용하여 GSEA를 수행했을 때, 검증 코호트(validation cohort)의 NR 종양은 CYP 또는 GSH 대사와 관련된 대사 과정 및 엽산 생합성과 관련된 유전자 세트가 강력하게 풍부해졌다. GSH를 생성하여 산화환원 상태를 조절하는 역할(도 3a). 발견 코호트(discovery cohort)를 특성화하기 위해 30개의 유전자 세트를 사용하여 GSEA를 수행했을 때, 검증 코호트(validation cohort)의 NR 종양은 CYP 또는 GSH 대사와 관련된 대사 과정뿐만 아니라 GSH를 생성하여 산화환원 상태 조절에 중요한 역할을 하는 엽산 생합성과 관련된 유전자 세트가 강력하게 나타났다(도 3a). 검증 코호트의 R 그룹의 종양은 사이토카인 및 항원 제시 유전자 세트에 대해 특성화되었다. 종양 침윤 림프구(TIL)는 AMC 검증 코호트의 TURBT 종양 표본에서 계산되었으며, TIL은 NR 그룹 종양에 비해 R 그룹 종양에서 더 높았음을 나타낸다(도 3b).

중요한 것은 MetaCore 분석에 따르면 NR 그룹 종양에서 핵인자 적혈구-2 관련 인자-2(NRF2, nuclear factor

erythroid-2 related factor-2)에 의해 매개되는 항산화 경로가 크게 상향조절되었음을 나타낸다(도 3c). NRF2는 GSH 대사 경로를 포함하여 항산화 반응에 관여하는 유전자의 주요 조절자다. 또한, 검증 코호트의 NAC 내성 종양은 무기 화합물의 해독과 관련된 세포 대사 과정 및 유전자 네트워크의 하위 집합으로 강력하게 표현되었다(도 3d 및 e). 이는 MIBC의 NAC 반응을 결정하는 데 있어 NRF2 매개 항산화, 대사 및 해독 경로에 대한 중요한 역할을 나타낸다.

[0097] 그 후, 발견 코호트에서 NAC 반응과 차등적으로 연관된 30개의 유전자 세트는 NAC에 대한 병리학적 반응 기록을 포함하는 NAC 이전 TURBT 표본의 3개의 독립적으로, 공개적으로 이용 가능한 유전자 발현 프로파일링 데이터 세트에 적용되었다(도 3f). 외부 연구 코호트(external institute cohorts)에는 다음이 포함된다: i) 3상 임상 시험의 맥락에서 MVAC로 균일하게 치료받은 23명의 환자로 구성된 "MDA MVAC 코호트", ii) DDMVAC 및 베바시주맙의 제2상 시험에서 38명의 환자로 구성된 "MDA DDMVAC 코호트", 및 iii) 다양한 NAC 요법으로 치료를 받은 7개 기관의 299명의 환자로 구성된 "NAC 메타데이터 코호트". AMC 검증 코호트의 결과와 일관되게, CYP 약물/이물질 대사 및 GSH 대사 유전자 세트는 "MDA MVAC" 및 "MDA DDMVAC" 코호트 모두의 NR 그룹 종양에서 상당히 풍부했다(도 3g 내지 3i), 이는 MIBC 수술 전 화학요법 민감도에서 GSH 대사의 중요성을 추가로 뒷받침한다. 마찬가지로, "NAC 메타데이터 코호트"의 R 그룹 종양에서 여러 면역 반응 과정이 상당히 강화되었다(그림 3i).

[0098] 종합적으로, 본 실험결과에 따르면, 4개의 독립적인 코호트의 MIBC 환자 종양에 대한 전사체 프로파일링은 GSH 관련 대사 반응이 NAC 반응 MIBC 아형의 생물학적 특징을 나타낼 수 있고, 면역 반응 경로가 NAC 민감성 MIBC 아형의 생물학적 특징을 나타낼 수 있음을 보여주었다.

[0100] 2-3. MIBC NAC 반응 예측 바이오마커

[0101] NAC 반응의 뚜렷한 분자 특징을 정의하는 잠재적인 바이오마커를 식별하기 위해 본 발명자들은 불완전한 TURBT를 특징으로 하는 AMC 발견 코호트의 전사체 데이터 세트를 사용하여 유전자 네트워크(MetaCore) 및 최첨단(GSEA) 분석을 추가로 탐색했다. MetaCore 분석에 따르면 면역 신장 및 종결, 펩타이드 생합성 과정 및 대사 과정과 관련된 유전자와의 네트워크를 통해 여러 징크 핑거 단백질(ZNF) 계열 유전자가 NR 종양에서 상향 조절되는 것으로 나타났다. 마찬가지로 GSEA 최첨단 분석에서는 GSH 대사 유전자 세트를 클러스터링하여 입증된 것처럼, 여러 GSH S-트랜스퍼라제(GST) 및 CYP 유전자가 NR 종양에서 고도로 발현되는 것으로 나타났다(도 4a). NAC에 민감한 MIBC는 면역 세포 경로의 좌표 활성화와 사이토카인 및 케모카인의 상향 조절을 나타냈다.

[0102] 본 발명자들은, 이후 LCMD 종양 표본을 사용하여 확립된 cDNA 라이브러리의 정량적 PCR(qPCR) 분석을 통해 이러한 추정 바이오마커의 발현을 검증했다. NR 종양은 R 종양과 비교하여 다음을 나타내는 유전자의 뚜렷한 발현을 나타냈다: i) GSH 대사, ii) 대사 과정, iii) WNT/MYC 경로, iv) 신경 표지자(neural markers), v) 면역 반응, vi) 세포 부착, 및 vii) 줄기성 특징 (도 4b). 이러한 경로 중에서 GSH 구축 AA 대사(GLS1 및 CBS), 합성(GCLM 및 GCLC) 및 활용(GSR, GGT7, PRDX1, GPX1 및 GPX4)이 상당히 강화되어 NAC 내성 MIBC 바이오마커로서의 잠재적인 용도를 시사하였다 (도 4c). 대조적으로, R군 종양에서는 면역 반응과 관련된 유전자(TRAF6 및 IL15RA)가 활성화되었다.

[0103]

[0104] 2-4. NAC 내성 MIBC(NAC-resistant MIBCs)에서 GSH 관련 단백질의 상향 조절

[0105] 전사체 수준에서 검증된 이러한 바이오마커의 잠재적인 임상적 중요성을 조사하기 위해 본 발명자들은 사전 NAC TURBT 표본에서 생성된 조직 마이크로어레이(TMA) 구축물에 대한 면역조직화학(IHC)을 사용하여 바이오마커의 단백질 수준을 조사했다. 처음에 AMC 발견 코호트는 의료병리학자의 시각적 평가를 통해 반정량적으로 검사되었다. 상기 전사체 발현 결과와 일치하게, GSH 관련 및 면역 반응 단백질은 NR과 R TURBT 표본 사이에 차등 발현을 나타냈다(도 4d). 영향을 받은 단백질 중 GLS1, GSR 및 GCLM 단백질은 NR 종양에서 크게 증가한 반면 CD11c 단백질은 감소했다(도 4e).

[0106] 또한, 본 발명자들은 단백질 수준을 객관적으로 측정하기 위해 AMC 검증 코호트의 IHC 슬라이드를 디지털화하고 전산 병리학 소프트웨어로 분석했다. 본 발명자들은 종양 세포와 간질 구획 간에 단백질 발현 수준이 다른지 여부를 확인하고자 하였다. 이를 위해, 딥 러닝 모델을 구축하여 IHC 슬라이드 이미지를 종양 상피 세포 또는 간질 구획으로 분할하고 발색체(3,3'-디아미노벤지딘, DAB)의 평균 강도를 각 구획에 대해 별도로 정량화했다(도 5a).

[0107] 중요한 것은 AMC 검증 코호트의 NAC 저항성 종양에서도 GLS1 단백질의 높은 발현이 관찰되었으며, 이는 MIBC NAC 반응을 정의하는 데 GLS1의 임상적 중요성을 뒷받침한다(도 5b 및 도 5c). 전사체 프로파일링 결과와 일치하여 NAC에 민감한 종양은 면역 반응(CD11c) 및 세포 접착(EOGT 및 KIFC2)과 관련된 단백질의 높은 발현을 나타냈다(도 5b 및 5c). NAC-민감성(sensitive) 그룹의 간질(stroma)에서 CD11c 및 CD247/CD3Z 단백질의 발현이 감소했다. 특히, NAC-저항성(resistant) 그룹의 간질(stroma)은 BCs의 유지 및 진행에 기여하는 SOX2 단백질의 상당한 상향 조절을 나타냈다.

[0109] **2-5. NAC 반응 예측을 위한 임상병리학적 데이터와 바이오마커 발현 데이터 간의 연관성**

[0110] 다양한 바이오마커의 임상병리학적 특징과 발현 수준 중 단변량 분석 결과, NAC 저항성은 TURBT 동안 종양의 불완전한 절제($p = 0.002$), 높은 수준의 GLS1 단백질($p = 0.002$), 종양 세포에서 CD11c의 낮은 발현($p = 0.035$)과 관련이 있었다(도 5d). 간질(stroma) 구획에서 SOX2의 높은 발현은 단변량 분석에서 NAC 무응답과도 관련이 있었다($p = 0.036$). 다변량 분석에서 종양 세포에서 GLS1의 높은 발현($p = 0.004$)과 CD11c의 낮은 발현($p = 0.010$)은 NAC 저항성 예측을 위한 독립적인 예후 인자로 남아 있었다(도 5d). 특히, 높은 GLS1 발현은 GSR 단백질의 종양 상향 조절과 유의한 상관관계가 있었고($p = 0.038$), 이는 GSH 역학과 관련된 경로의 조정된 활성화를 뒷받침한다(도 5e).

[0111] 종합적으로, 이러한 결과는 GSH 대사 단백질, 특히 GLS1의 발현이 NAC 반응에 의한 MIBC 아형의 분자 특징을 정의하는 데 임상적으로 중요하다는 것을 입증했다.

[0113] **2-6. 다중-코호트 MIBC 환자의 전사체 데이터셋에 대한 딥러닝 기반 분석**

[0114] 상기 실시예에서 구축한 근침윤성(MIBC) 방광암 환자 시료에 대한 multicohort datasets에 공통적으로 적용될 수 있는 neoadjuvant chemotherapy (NAC) 항암치료 반응성 예측용 바이오마커들을 발굴하기 위하여, 1) 본원의 AMC validation cohort와 본 발명에서 사용한 3가지 외부 코호트들(MDA_MVAC, MDA_DDMVAC, NAC_metadata)의 전사체(transcriptome) datasets, 2) 본 연구에서 발굴한 NAC response 예측용 GSH metabolism과 면역 반응 등의 신규 바이오마커 유전자군(genesets)과 기존 연구에서 MIBC 분자 분류(molecular classification)로 개발된 15가지의 molecular classifiers 들을(도 6aa, 6ab) 모두 사용하여, deep learning based 교차 검증 분석 모델링과 NAC 반응성 구분에 유익적으로 작용하는 molecular classifier들을 최적화하였다.

[0115] 본원과 외부 코호트들의 dataset 차이점을 보정하기 최적의 scaler와 각 molecular classifier 내 최적의 유전자 10개 미만으로 선정하기 위한 사전 작업을 실시하였다. 우선 교차 검증을 위해 4개의 데이터셋 모두에 존재하는 유전자 16,999개를 선택하였다. 9가지 scaling algorithms(log2, log2_minmax, standard, minmax, max_abs, robust, power, quantile, rankgauss) 중에서 4개의 환자 코호트 datasets의 pairwise T-test 검사시 p-value가 0.01 이상으로 기무가설(두 분포의 평균이 같다)을 유지 가능하고, Gaussian 분포를 따르는 scaler인 standard, power, rankgauss scaler를 선택하였다. Scaler 별 classifier 별 random forest 로 변수 최적화하기 위하여, 1) classifier 내 gene(변수)수가 10개 이하일 때 그대로 유지, 2) classifier 내 gene(변수)수가 10개 초과일 때 random forest의 모델의 주요변수 top 10이 선택, 3) 각 random forest 모델의 성능인 Accuracy, Recall, Precision, F1, MCC, R2의 합을 scaler의 퍼포먼스 스코어로 합산하였다. 그 결과 최종적으로 power scaler가 최적의 scaler로 선택되었으며, 각 16가지 molecular classifiers 내 최적의 top10 유전자 군들을 최적화하였다.

[0116] 교차 검증 모델링을 위하여, 사전 처리 작업에서 결정된 power scaling으로 표준화한 각 4개 환자 코호트 데이터셋 중 하나는 학습, 하나는 검증, 나머지 둘은 테스트로 활용해서 12 set의 교차검증 수행하였다. 이때 logistic regression, decision tree, random tree 통계 기법에 따라 molecular classifiers 마다 dataset pair 중 첫번째 데이터로 모델을 만들고, 검증과 테스트 작업을 수행하였다. 교차 검증되었다고 선택한 기준은 1) 동일 dataset 내 NAC response (R)과 non-response (NR) 그룹 간 차이의 유효성, 2) 각 코호트들의 R 그룹 간의 유의성, 3) 각 코호트들의 NR 그룹 간의 유의성을 기준으로 선정하였으며, 이러한 기준으로 총 612번($12 \times 3 \times 17 = 612$ 번)의 교차 검증을 수행하고, 위와 같은 과정을 본 연구팀이 발굴한 신규 AMC union classifier 에도 적용하여 총 1224번의 검증 분석을 수행하였다.

[0117] 이러한 작업 프로세스로 deep learning 교차 검증을 실시한 결과, Meta_Datasets (NAC_metadata)을

모델링하여, MDA 코호트에서 검증한 조건에서는 BC luminal-16 genes, BC p53 pathway, and BC p53-like pathway molecular classifiers 들이 logistic regression 모델링에서 NAC NR과 R 그룹을 유의적으로 구분할 수 있었으며, 본원 코호트 (AMC validation cohort)을 검증한 경우, logistic regression 모델에서는 BC p63 pathway가 유의적인 classifier로 선정되어 최적화되었으며, decision tree 모델링에서는 본 연구팀에서 구축한 AMC Up-regulated molecular classifier가 NAC NR과 R 그룹을 유의적으로 구분하는 것으로 확인되었다 (도 6b). 특히 BC p53-like pathway의 경우, AMC 코호트의 경우에도 NAC NR과 R 그룹 간에도 차이가 보였으며, 다른 3가지 외부 코호트의 경우 모두 유효하다는 것을 확인하였다 (도 6c).

2-7. MIBC 환자의 NAC 반응을 예측하기 위한 예상 바이오마커 식별

본 실시예에서는 Deep learning 기반 멀티 코호트 전사체 분석 모델링 분석법에서 NAC response 예측에 유의적인 기능을 하는 최적화된 molecular classifier 내에서 바이오마커 중에서 본 발명에서 구축한 digital pathology 분석법으로 활용하여 임상적 유의성 검증과 유효성을 평가하고자 하였다. 이를 위하여, 전사체 분석 모델링에서 가장 통계유의성이 높았던 NAC_metadata 코호트를 선정하고, 이를 AMC validation cohort 전사체 dataset들과 동일한 모델링 방법으로 교차 검증을 실시하였다. 그 결과 logistic regression 모델링에서는 BC p63 pathway가 NAC response를 구분하는 molecular classifier로 선정되었으며, Rando forest 모델링에서는 Luminal_ref.EUA_cancer genome classifier가 선정되고, decision tree 모델링에서는 BC luminal과 AMC Up-regulated molecular classifiers 들이 NAC NR과 R 그룹을 유의적으로 구분할 수 있는 molecular classifier라는 것을 확인하였다 (도 7a). Logistic regression 모델링에서 NAC NR과 R 두 그룹을 구분하는 최적의 molecular classifiers에는 ITGAX, AFAP1, KCTD14, RFX7, DPH2, SLC15A3, PPIL2, CARD16, FYB, GADD45B, DNMT3L 등의 핵심 유전자들을 포함하고 있다. 또한 가장 우수한 유의성을 보였던 decision tree 모델링의 경우 (도 7b 및 도 7c), molecular classifiers 들은 PCDHB9, HCFC1R1, NT5E, TNFAIP8, POU2F2 핵심 유전자들을 포함하고 있다 (도 7c 및 도 7d).

전사체 관련 deep learning 모델링 분석에서 도출된 핵심 유전자들의 임상적 유의성을 검증하기 위하여, 본 발명에서 개발한 tumor/stroma classifier을 적용한 digital pathology 분석법을 활용하여 NAC NR과 R 그룹 환자들의 방광암 치료에서 해당 유전자들의 단백질 발현 차이를 평가하였다. 그 결과, AMC discovery cohort에서 CARD16, CD73, DNMT3L, FYBI, HCFC1R1, MTCH1, RFX7, SLC15A3, TNFAIP8 단백질이 발현이 NAC 반응성에 따른 두 그룹에서 발현 차이를 보였으며(도 8a), 특히 CARD16, DNMT3L, HCFC1R1, RFX7, USP2 단백질들의 발현 차이는 AMC validation cohort 환자들의 임상 치료에서 NAC NR과 R 두 그룹 간의 유의적인 차이를 보였다 (도 8b). 따라서 근침윤성 방광암 임상 치료의 멀티 코호트 전사체들을 deep learning 모델링을 통하여, NAC 반응성을 예측할 수 있는 분석 모델링 기법과 이를 결정하는 핵심 유전자들을 개발하였으며, 디지털 병리 분석을 통하여 이들 신규 바이오 마커들의 임상적 유의성을 단백질 변화 수준에서 검증하였다.

2-8. MIBC 환자의 NAC 반응 예측을 위한 딥러닝 기반 디지털 병리학적 분석

본 연구 개발에서 방광암 NAC 반응성 예측을 위한 바이오 마커들은 1) GSH 세포 대사 genesets, 2) 면역 반응 관련 genesets, 3) deep learning 기반 전사체 교차 검증 모델링에서 최적화된 molecular classifiers 들을 포함하여 총 61 종의 유전자들을 선정하였으며, 본원에서 NAC 반응성이 확인 가능한 63명 환자들의 방광암 치료에 대한 디지털 병리 분석 datasets들을 확보하였으며, 본 발명에서 디지털 병리 분석에 사용한 단백질 마커 61종은 다음과 같다 (GLS1, PSAT1, CBS, GCLC, GCLM, Gluta (GSR), GlnRS (QARS1), GGT7, Perox (PRDX1), PLOD2, RPAP1, RPL9, MITF, CD44v6, CDK1, FZD9, GAD2, PPP2R5A, non-p (b-catenin), B-catenin, SALL4, SOX2, TFCP2L1, TFEB, ICAM1, TRAF2, TRAF6, IL15RA, AFAP1, CARD16, CD11c, CD73, CD3Z, , EBI3 (IL27B), EMX1, DNMT3L, DPH2, EGR2, FYBI, GADD45B, HCFC1R1, KCTD14, MTCH1, OCT2, PCDHB9, PPIL2, RFX7, SLC15A3, TNFAIP8, ANPEP, BDN, EOGT, FOXA1, KIFC2, KIR3DL1, NOTUM, USP2, CCDC6, CK5(KRT5), CK20 (KRT20), GATA3).

근침윤성 (MIBC) 방광암 환자 치료에 대한 디지털 병리 datasets들을 활용하여, NAC 반응성을 예측할 수 있는 면역 조직 염색을 위한 최적의 단백질 조합을 발굴하는 것이 본 연구 개발 기술의 최종 목표이다. 대규모 환자 임상 치료와 대량의 항체를 활용한 디지털 병리 dataset을 활용하기 위하여, 앞의 전사체 교차 검증 모델링 연구 사용한 유사한 deep learning 분석 방법을 적용하였다. 외부 코호트의 면역 염색 결과들이 없으므로, 디지털

병리 분석 결과의 교차 검증을 위하여, 본 연구 개발에서 사용한 NAC 치료 환자군 (cT2-4aNOMO) 코호트와 함께, 본원에서 first line으로 NAC 치료를 받은 환자 (Non-cT2-4aNOMO) 84명이 포함된 독립적인 코호트 (그룹4)를 구축하고, 이들 코호트의 방광암 임상 시료에 대하여 동일한 방법으로 디지털 병리 분석을 수행하였다.

[0127] 디지털 병리 dataset의 deep learning modeling에 전체 수행 과정은 다음과 같다. 우선 tumor compartment 결과들의 경우 모든 값들이 존재하였으나, stroma compartment 결과들의 경우, 두 환자 관측값의 손실이 있어서 분석에서 제외하였다. 아래 추가로 기술한 데이터 기반 주요 바이오마커 추출 프로세스에 따라 총 10개 gene pool (tumor 5개, stroma 5개)을 추출하였다. 최적 모델을 발굴하기 위하여, 추출된 gene Pool 별로 1) 유효 마커 수를 2개 부터 5개까지 반복, 2) 베이스 모델을 decision Tree와 logistic Regression으로 반복, 3) 학습 데이터 (NAC 치료 환자군; 그룹 1+2)와 first-line 검증 환자군 (그룹4)을 모두 합쳐 랜덤하게 3-fold 교차 검증을 수행하면서 유효 마커 수만큼의 최적의 마커 선택, 4) 변수 선택 과정은 forward, backward 모두 수행하는 작업을 진행하였다. 여기서 사용한 forward process는 유효 마커 수가 될 때까지 선택 가능한 마커 풀 중에서 이미 선택된 마커들과 함께 최적의 효과를 낼 수 있는 마커 1개를 순차적으로 추가하는 방법이며, backward process는 전체 마커 풀 중에서 유효 마커 수가 되기까지 최적의 효과를 낼수 있게 마커 1개씩을 순차적으로 제거하는 방법이다. 이런 분석 방법을 반복하면서, 도 9a에서 제시된 조건을 만족하는 새로운 모델이 도출되는 경우 최적 모델로 선정하였다. 전사체 교차 검증 모델링과 달리, 디지털 병리 모델링의 경우, 1) 전체 측정 바이오 마커 수가 적고 (61개), 2) 학습데이터(55건)과 검증데이터 (36건)은 환자 수가 적고, 3) 학습과 검증 데이터에 포함되어 있는 환자군의 임상학적 특성도 다르다는 한계점을 가지고 있어, 데이터 기반 주요 바이오마커 추출 프로세스를 적용하였다. 이를 위하여, 전체 변수를 모두 넣은 dataset pool에서부터 순차적으로 1개씩 빼고 모델링 했을 때 이전 모델보다 성능이 나아지면 해당 변수에 높은 순위를, 그렇지 않다면 낮은 순위를 주는 방법인 Recursive Feature Elimination with Cross Validation (RFECV) 분석 방법을 선택하였다 (도 9b). RFECV 과정의 교차 검증으로 변별력이 강화된 변수 랭킹을 만들 수 있었으며, 반복적인 모델링 과정에서 사용할 모델링 방법은 decision Tree(tree)와 logistic Regression(reg)을 사용하였다.

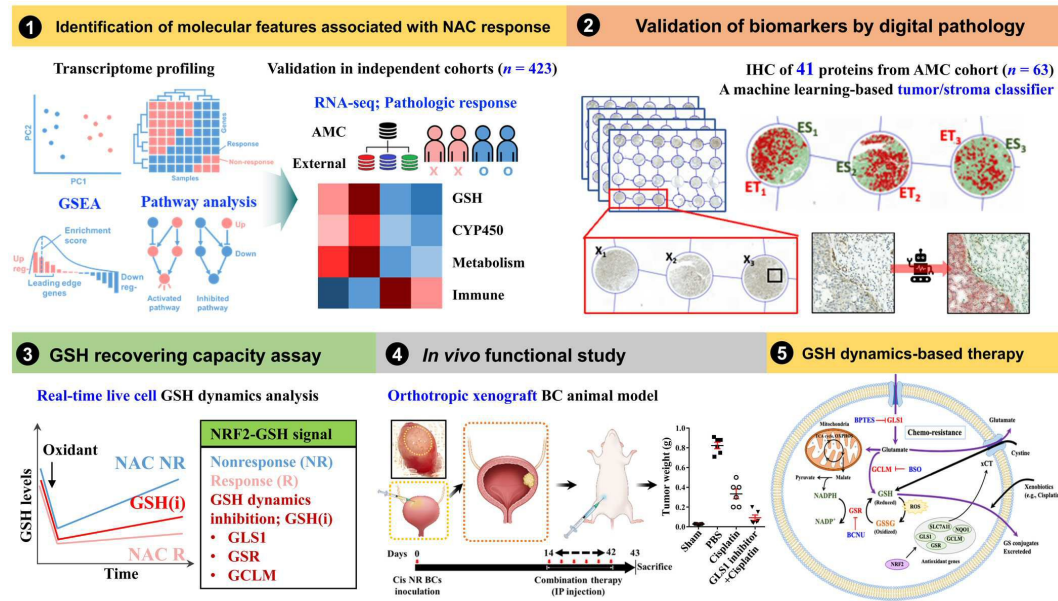
[0128] Digital pathology 관측 값 자체를 사용한 data type (raw)과 표준화를 진행한 (std) datasets을 생성하여 분석에 사용하였다. Mean_tumor_vals을 제외하고는 두가지 데이터 타입별로 decision tree와 logistic regression 모델링을 통해 Recursive Feature Selection (순차적 변수 선택) 방법 중 forward, backward별로 2개 ~ 5개의 최적 유효 마커를 추출한 뒤, 해당 베이스 모델로 학습 데이터에 대해 모델링을 수행하였다. Mean_tumor_vals에 세팅의 경우 raw 데이터뿐 아니라 std 데이터에서도 동일 작업을 수행하였다. 각 tumor and stroma compartment 결과, data type, 모델링 기법마다 NAC response NR과 R 두 그룹을 구분할 수 있는 단백질 그룹들을 선정하였으며 (도 9c 및 9d), 최종적으로 도출된 2개의 모델 중 tumor_tree_vals (raw data type, decision Tree 모델)의 gene Pool에서 최적 모델로 선정되었다 (도 9e). Tumor 와 stroma compartment에 대한 디지털 병리 dataset에서 최적의 모델로 선정된 각 decision tree 모델의 구체적인 결과와 핵심 유전자들은 도 9f 및 도 9g에 표시되어 있다. 따라서, 본 발명에서는 대규모 근침윤성 방광암 임상 시료에 대한 고속 자동화 digital pathology 분석과 함께, tumor와 stroma에서 발현되는 대량의 단백질들의 발현량에 대한 digital pathology 결과를 deep learning based 모델링을 통하여 항암 치료 반응성을 예측할 수 있는 최적의 단백질들의 조합군들을(도 9c 및 9d) 개발하였다.

[0130] 본 연구에서 개발한 deep learning 기반 디지털 병리 모델링 분석법의 활용성을 검증하기 위하여, first-line 방광암 환자군으로 구성된 독립적인 코호트를 활용하여 교차 검증 분석을 진행한 결과, tumor compartment 데이터 기반 모델링의 경우 학습과 검증 환자군에서 모두 NAC 반응성에 대한 분포가 유의적으로 차이가 나는 것을 확인하였다 (도 10의 A). 하지만 stroma compartment 데이터 기반 모델링의 경우에는 first-line 환자군으로 구성된 검증 코호트에서 NAC response에 대한 구분이 명확하지 않은 것을 확인하였으며 (도 10의 B), 이러한 원인으로 변수 선택에서부터 학습 및 검증 데이터 모두에서 유의성이 높은 변수 도출 빈도가 적은 것이 원인이라고 판단된다.

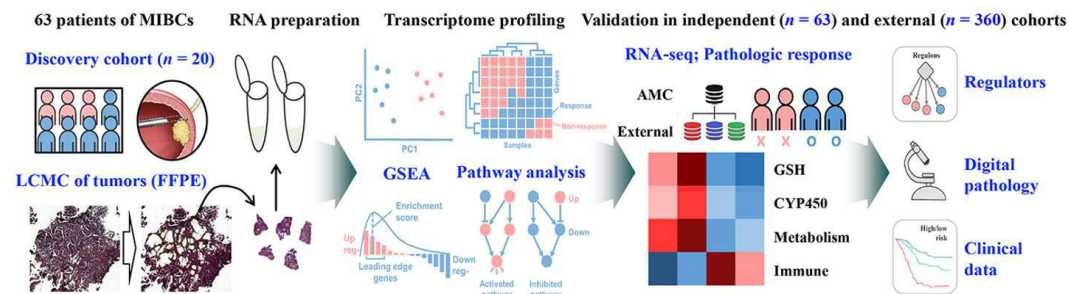
[0132] 상기의 실시예는 본 발명의 내용을 예시하는 것일 뿐 본 발명의 범위가 상기 실시예에 한정되는 것은 아니다. 본 발명의 실시예는 당업계에서 평균적인 지식을 가진 자에게 본 발명을 보다 완전하게 설명하기 위해 제공되는 것이다.

도면

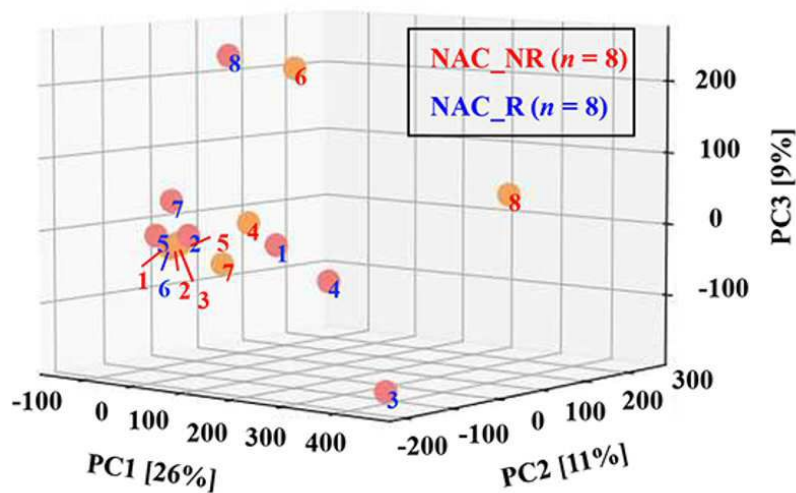
도면1



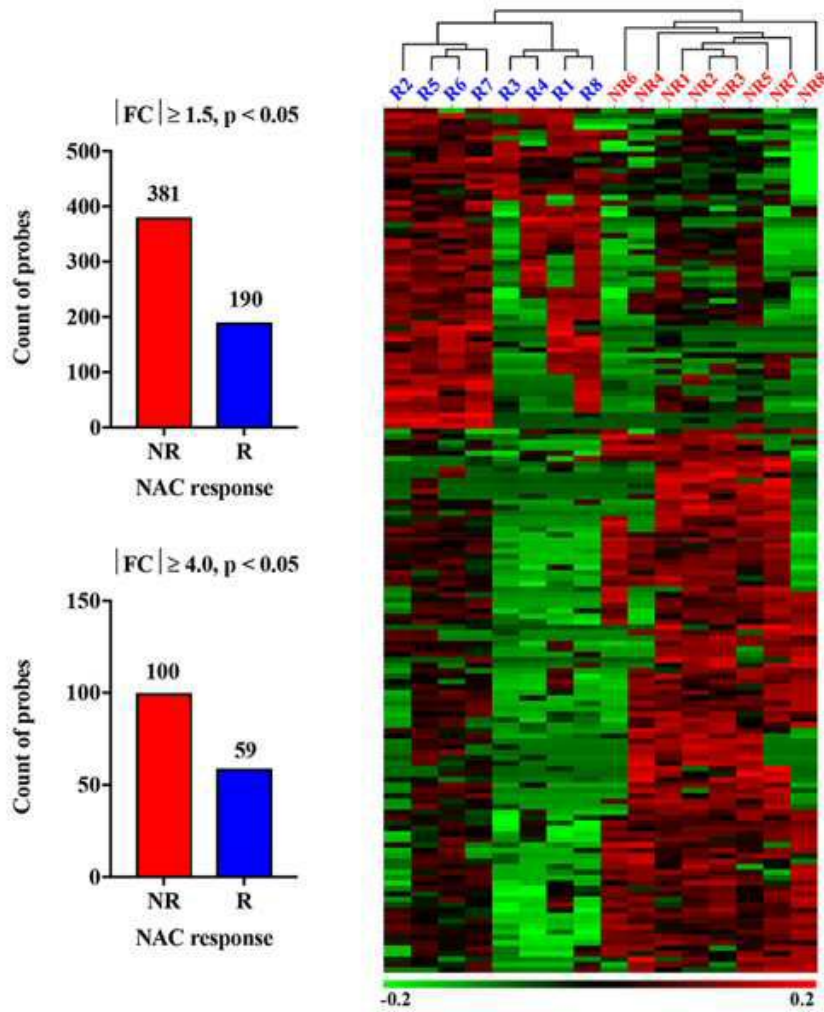
도면2a



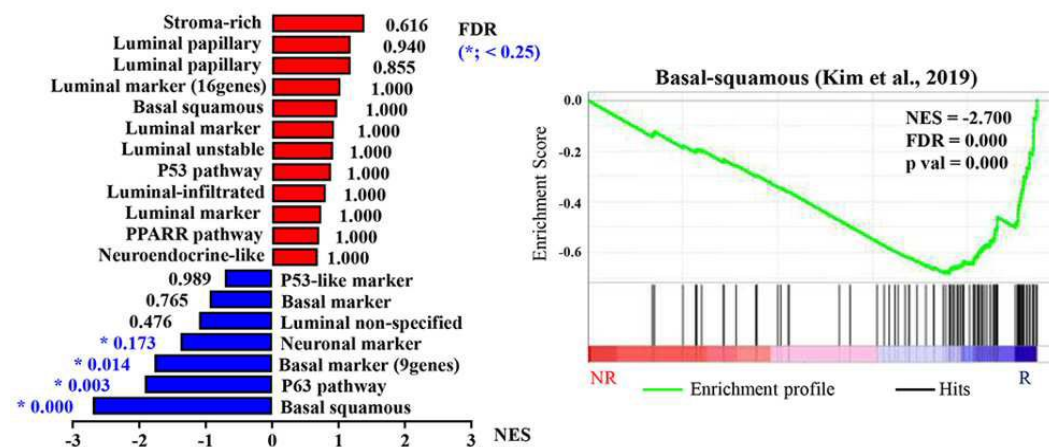
도면2b



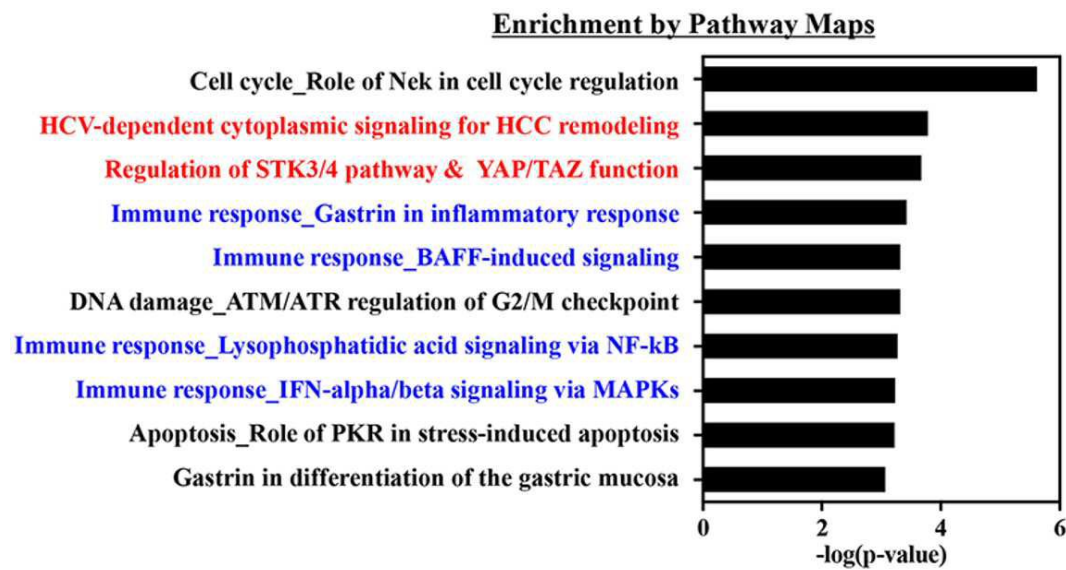
도면2c



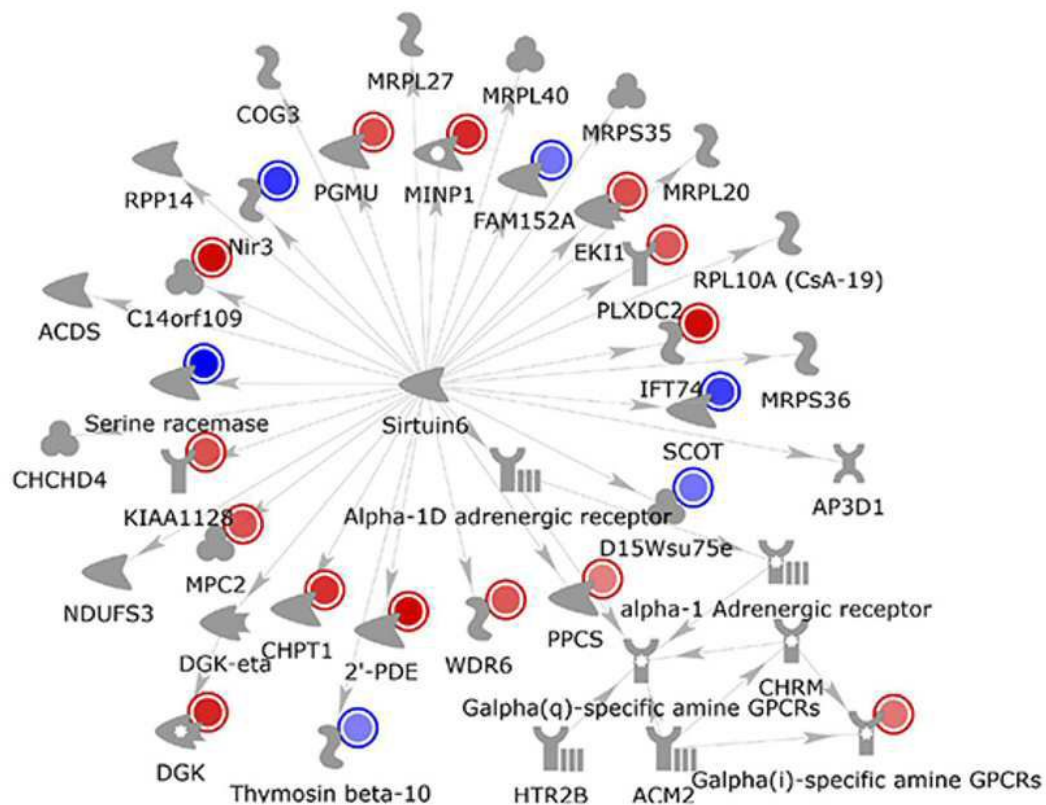
도면2d



도면2e

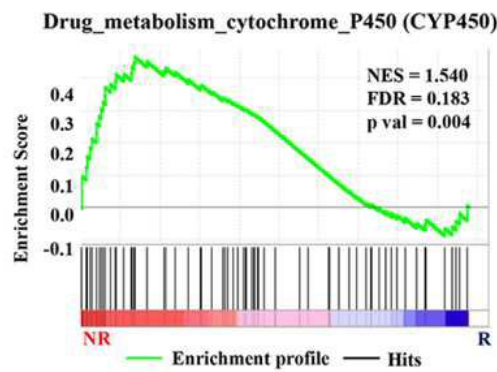
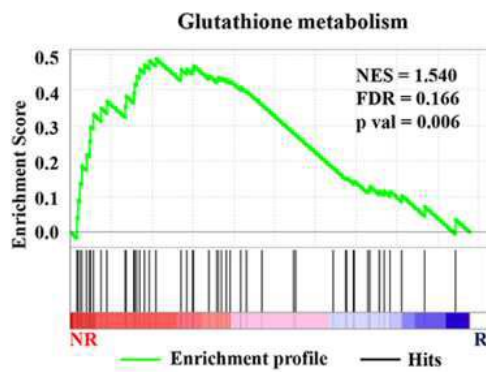
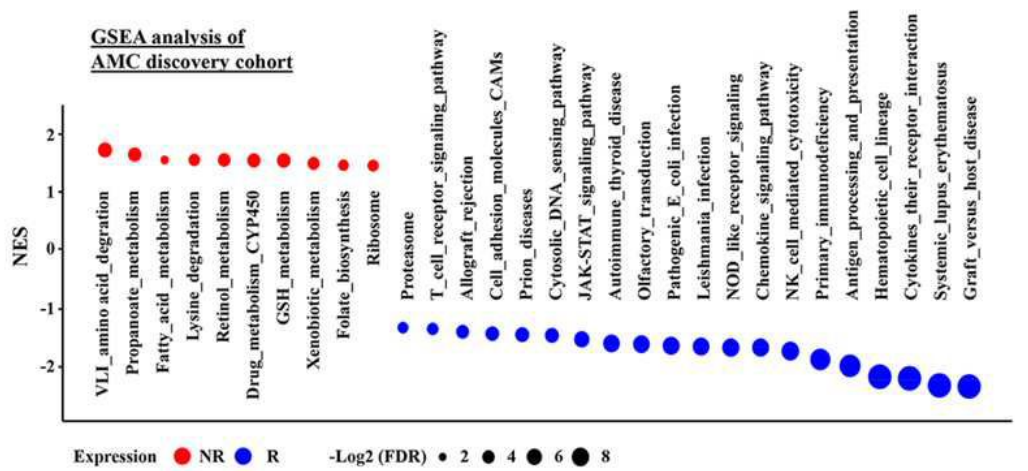


도면 $2f$

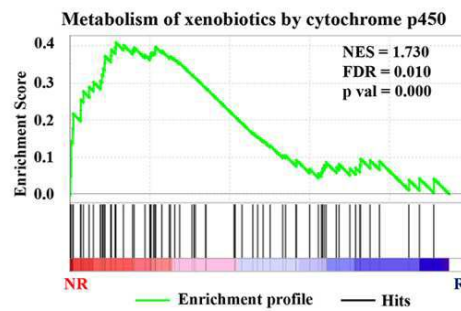
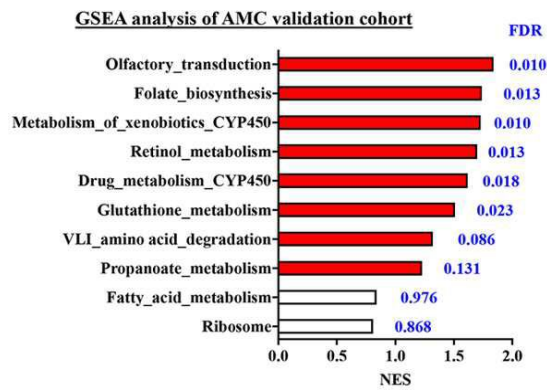


Network : IFT74, SCOT, Nir3, Serine racemase, C14orf109

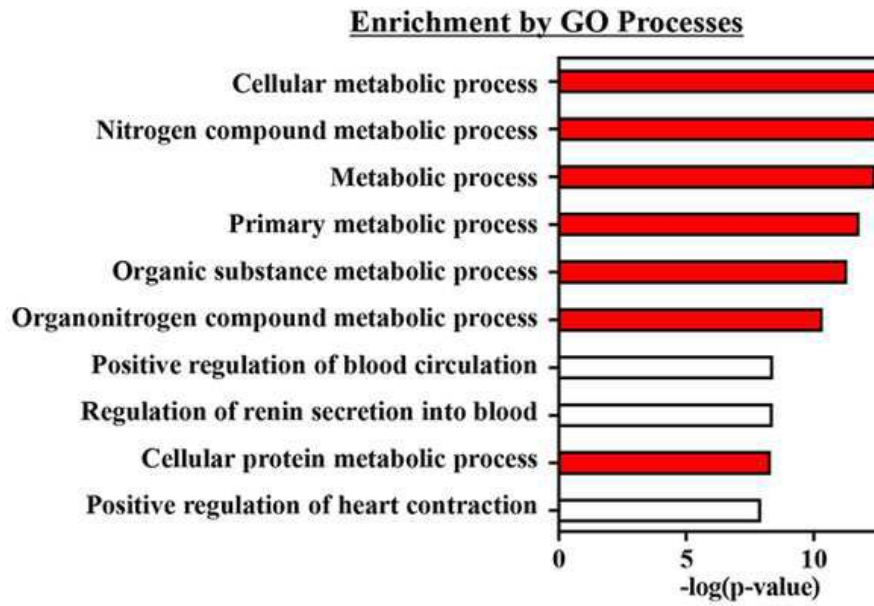
도면2g



도면3a

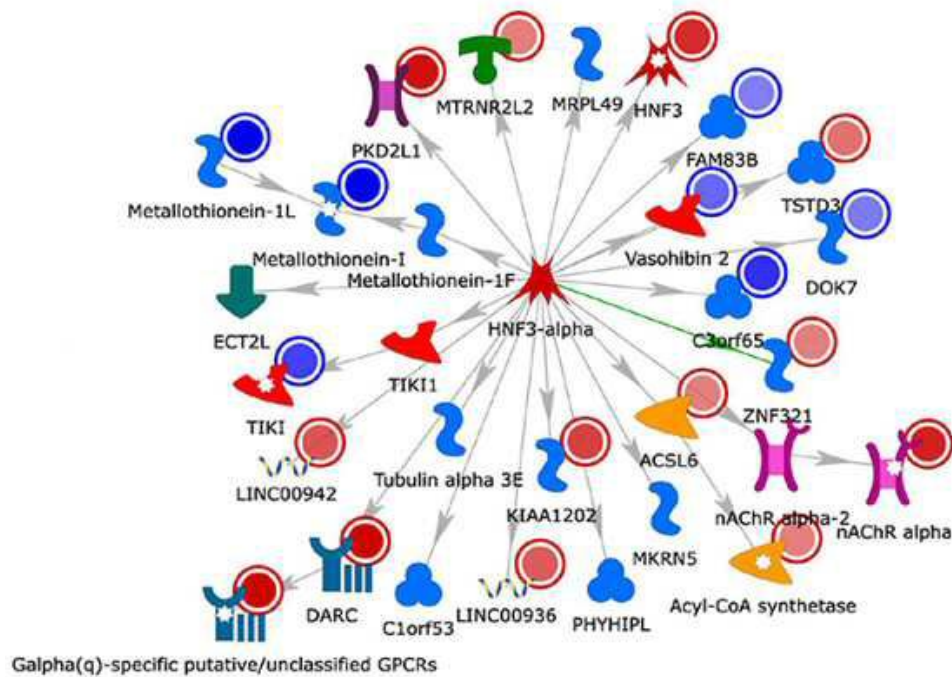


도면3d



도면3e

**Network : ACSL6, Vasohibin 2,
KIAA1202, PKD2L1, DOK7**



도면3f

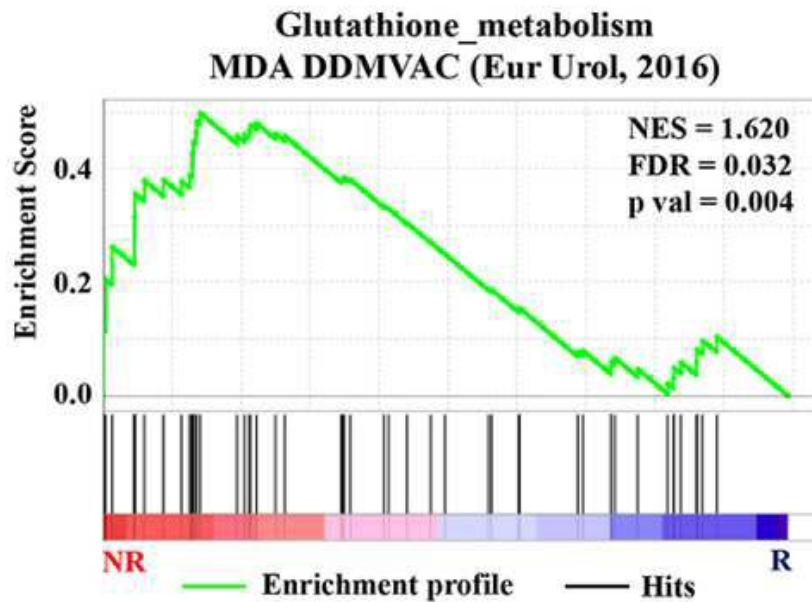
Dataset	Publication	Data accession	MIBC samples		Number of significant genesets		ES plots
			Nonresponder (NR)	Responder (R)	Enrichment in NR	Enrichment in R	
ASAN (VAL)	This study	GSE212810	18	21	8 (26.6 %)	4 (13.3 %)	Fig. 2A, S2B,C
MDA MVAC	Cancer Cell, 2014	GSE48277	17	6	4 (13.3 %)	5 (16.6 %)	Table S3
MDA DDMVAC	Eur. Urol. 2016	GSE69795	20	18	17 (56.6 %)	2 (6.6 %)	Table S4
NAC metadata	Eur. Urol. 2017	GSE87304	163	136	1 (3.3 %)	13 (43.3 %)	Table S5

도면3g

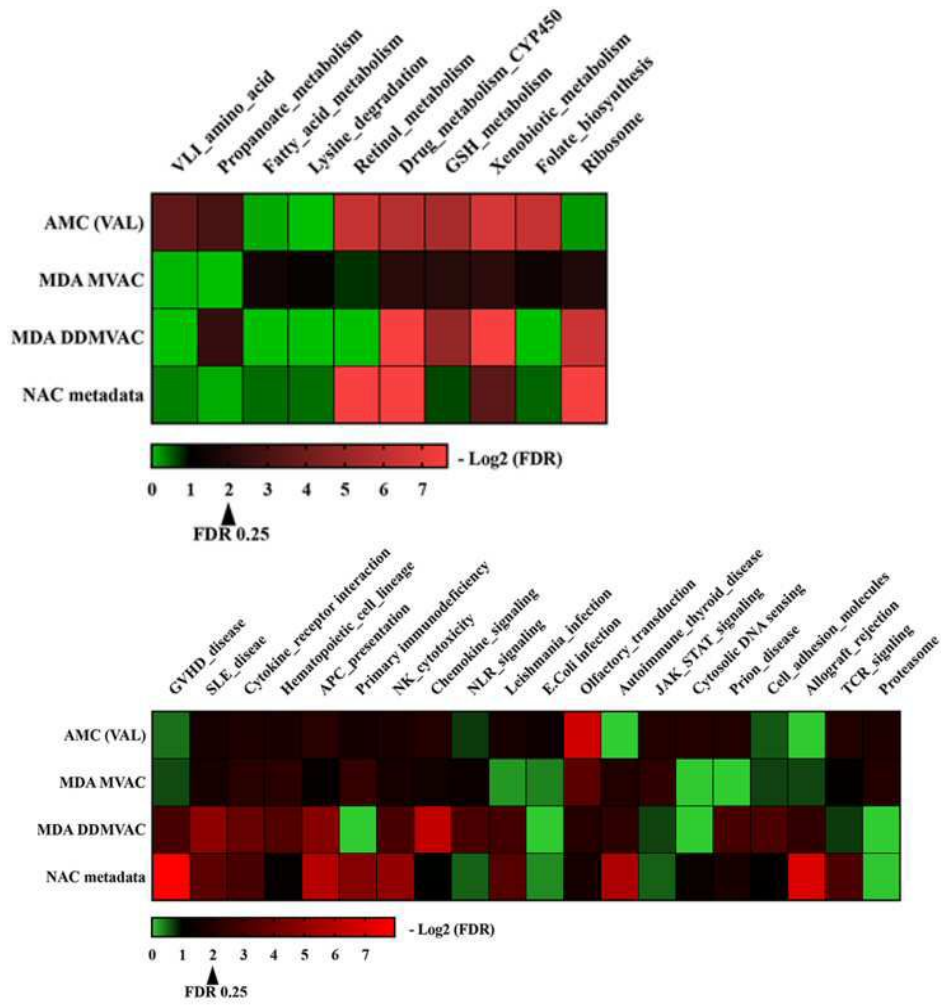
Enriched in NR group (MDA MVAC cohort)

	GS (Standard name)_KEGG	GS (System#)	SIZE	ES	NES	NOM p-val	FDR q-val
1	Olfactory_transduction	M14091	381	0.36	1.59	0.000	0.091
2	Drug_metabolism_cytochrome_P450 (CYP450)	M9257	72	0.39	1.41	0.036	0.238
3	Glutathione_metabolism	M1840	50	0.38	1.28	0.121	0.247
4	Metabolism_of_xenobiotics_by_cytochrome_P450	M16794	70	0.35	1.26	0.104	0.226

도면3h

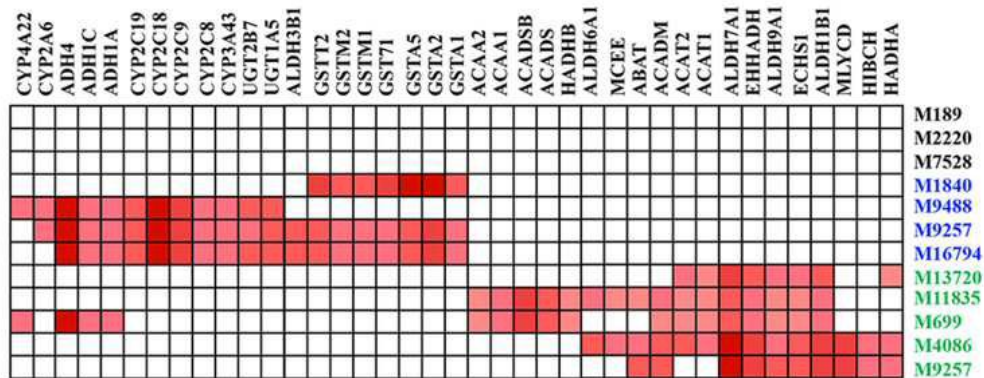
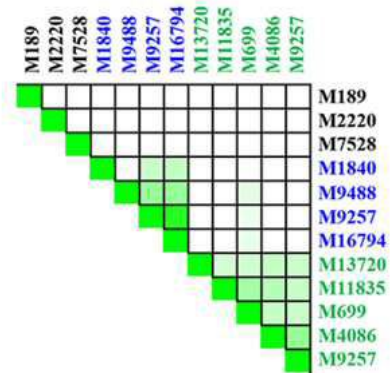


도면3i

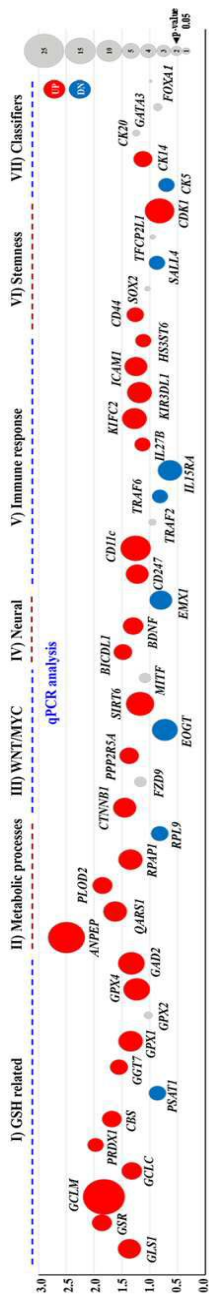


도면4a

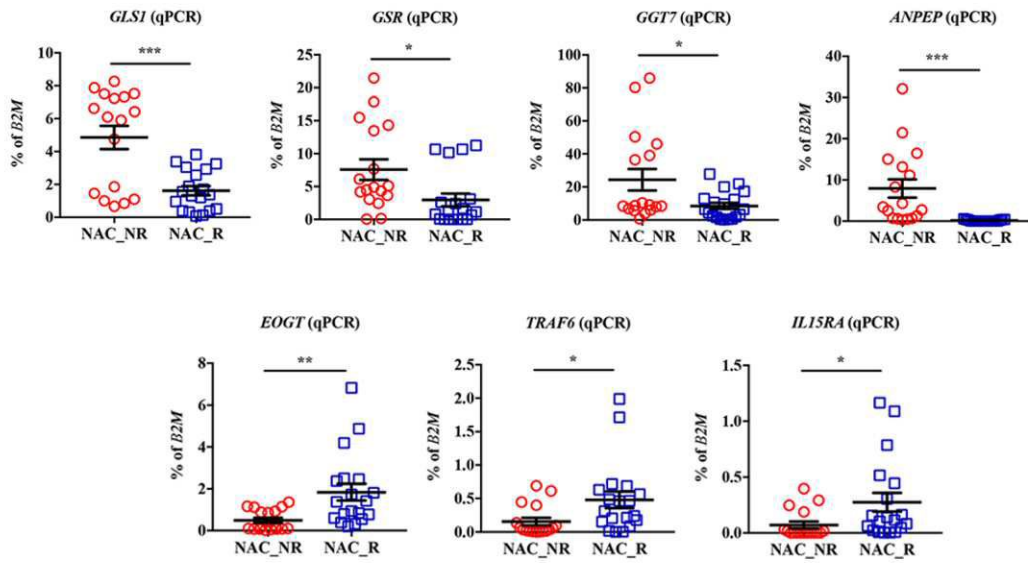
GS #	GS (Standard name)_KEGG
M189	Ribosome
M2220	Folate_biosynthesis
M4086	Propanoate_metabolism
M1840	Glutathione_metabolism
M9488	Retinol_metabolism
M9257	Drug_metabolism_cytochrome_P450
M16794	Metabolism_of_xenobiotics_by_cytochrome_P450
M13720	Lysine_degradation
M11835	Valine_leucine_and_isoleucine_degradation
M699	Fatty_acid_metabolism
M7528	Pantothenate_and_coa_biosynthesis
M2668	Beta_alanine_metabolism



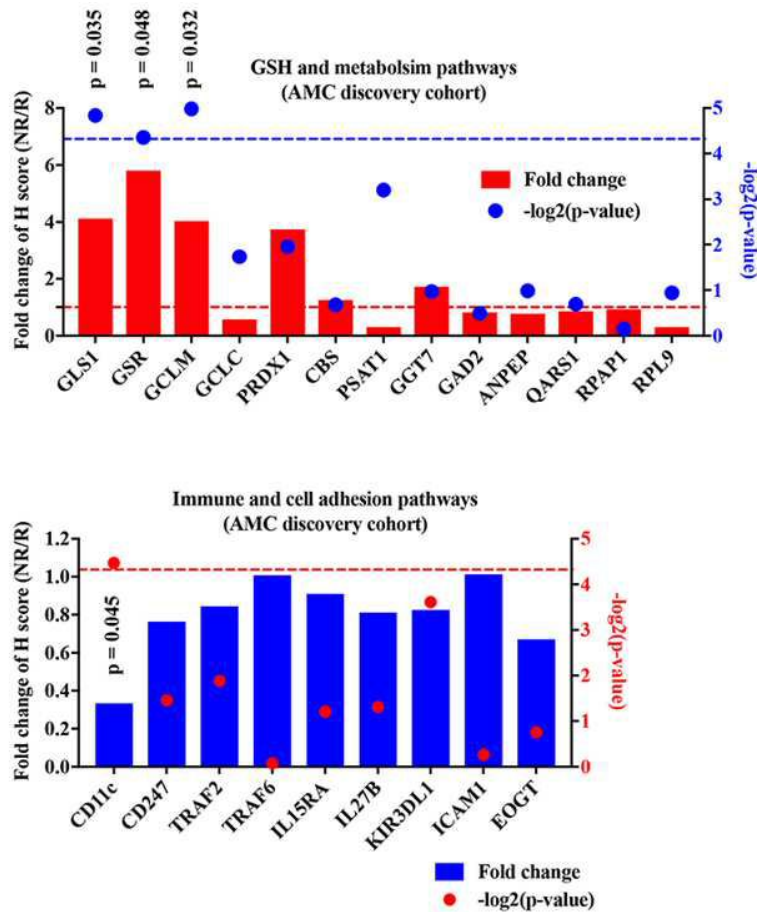
도면4b



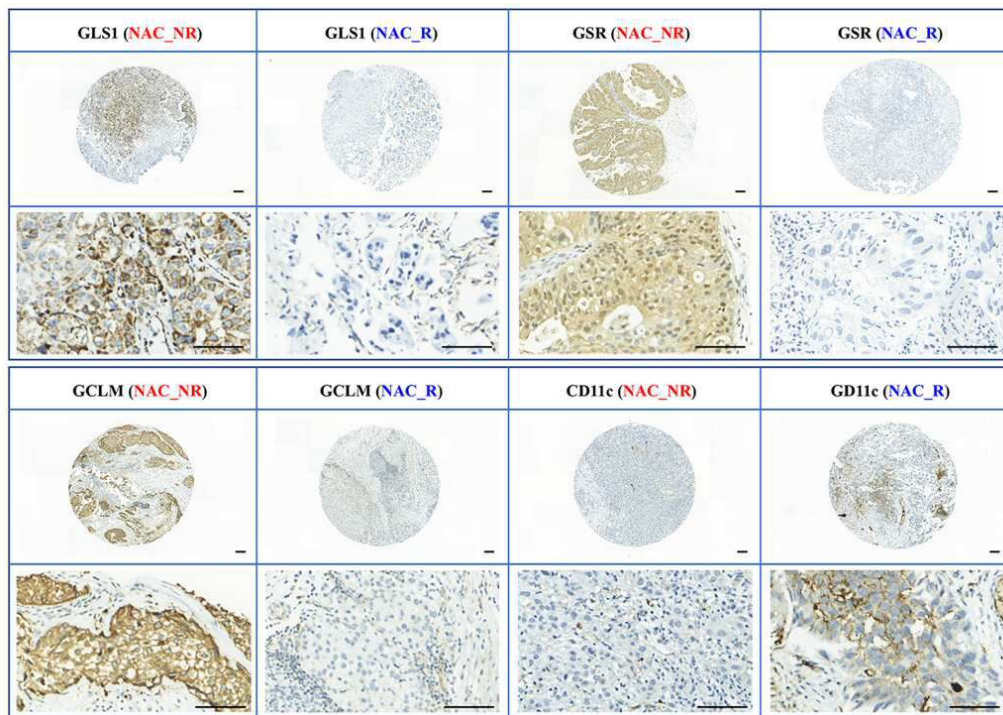
도면4c



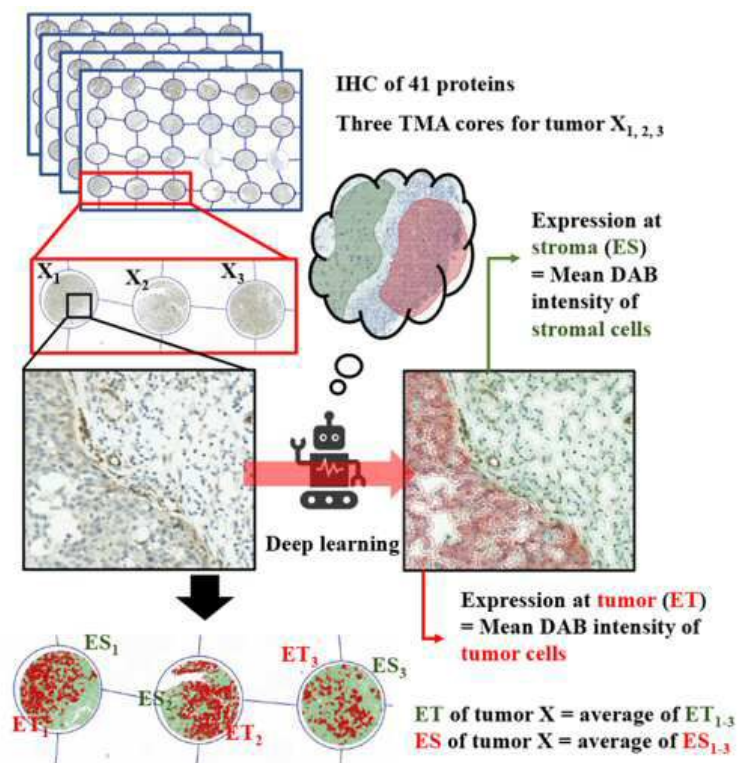
도면4d



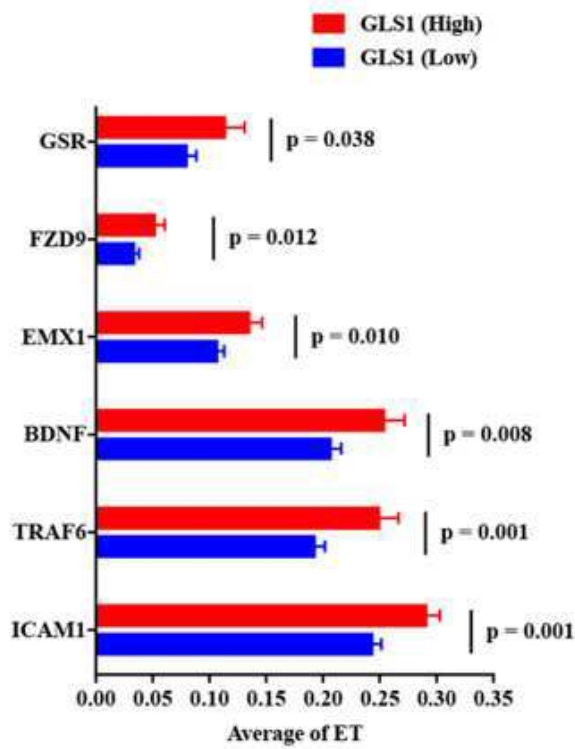
도면4e



도면5a



도면5e



도면6aa

Metacore&GSEA target genes	GSEA_R enrichment_all	GSEA_R enrichment_intersection	BC Basal	BC basal-9gene	BC Luminal	BC Luminal-16gene	BC P53 pathway	BC-P53like
CCDC64	CCL5	CD247	GATA3	CD44	AHR	CD24	ANLN	AHR
CHKB	CCL8	CD3D	HIF1A	CDH3	CD24	CYP2J2	ANXA2	CDKN2A
CYP2C18	CD247	CD3E	IRF3	KRT1	CYP2J2	ERBB2	BUB1	E2F2
DCAKD	CD3E	CD3G	IRF7	KRT14	ERBB2	ERBB3	CA9	FOXM1
EIF4E	CSF2	CSF2	JUN	KRT16	ERBB3	FABP4	CHEK1	HEY2
GAMT	CXCL10	HIST1H2BF	KLF2	KRT5	FABP4	FGFR3	CKAP2	MKL1
MAMDC4	CXCL11	ICAM1	MEOX2	KRT6A	FGFR3	FOXA1	CNN1	MYC
MDP1	CXCL9	IL1B	NCOR1	KRT6B	FOXA1	GATA3	CSR1	MYOCD
NUBP2	FCGR2B	KIR3DL1	NFKB1	KRT6C	GATA3	GPX2	DLGAP5	NFE2L2
PCDH17	GBP1	KIR3DL2	PIA54		GPX2	KRT18	FEN1	RB1
PM20D2	HIST1H2BF	KLRD1	PPARA		HDAC1	KRT19	FHL1	SMAD7
RPAP1	IFIT1		RELA		HTT	KRT20	GRB2	SMARCB1
RPL9	IFNA2		SP1		KRT18	KRT7	IGFBP5	SPDEF
STK36	IL11		SPDEF		KRT19	KRT8	KIAA0101	SRF
TATDN1	IL12B		STAT1		KRT20	PPARG	MAD2L1	TBX2
TMEM63A	IL1B		STAT3		KRT7	XBP1	MAP2K1	TCF3
WFDC8	IL2		TP63		KRT8		M8NL2	TP53
ZNF493	ISG15		TRIM24		MKL1		MCL1	
ZNF592	ITGB2				MYC		MMP23B	
ZNF615	KIR3DL1				NFKB1		MTMR11	
ZNF626	KIR3DL2				PGR		NUPR1	
	KLRD1				PPARA		PDGFRA	
	MX1				PPARG		PIK3R1	
	MX2				SMAD3		PPP1CA	
	OAS2				SMAD7		PRKD1	
	SLA				SMARCA4		PTPRE	
					SPDEF		PTTG1	
					SREBF2		RRM2	
					SRF		SORBS1	
					STAT1		STMN1	
					TP53		TBL1X	
					TRIM24		TFDP1	
					XBP1		VRK1	
							ZMAT3	

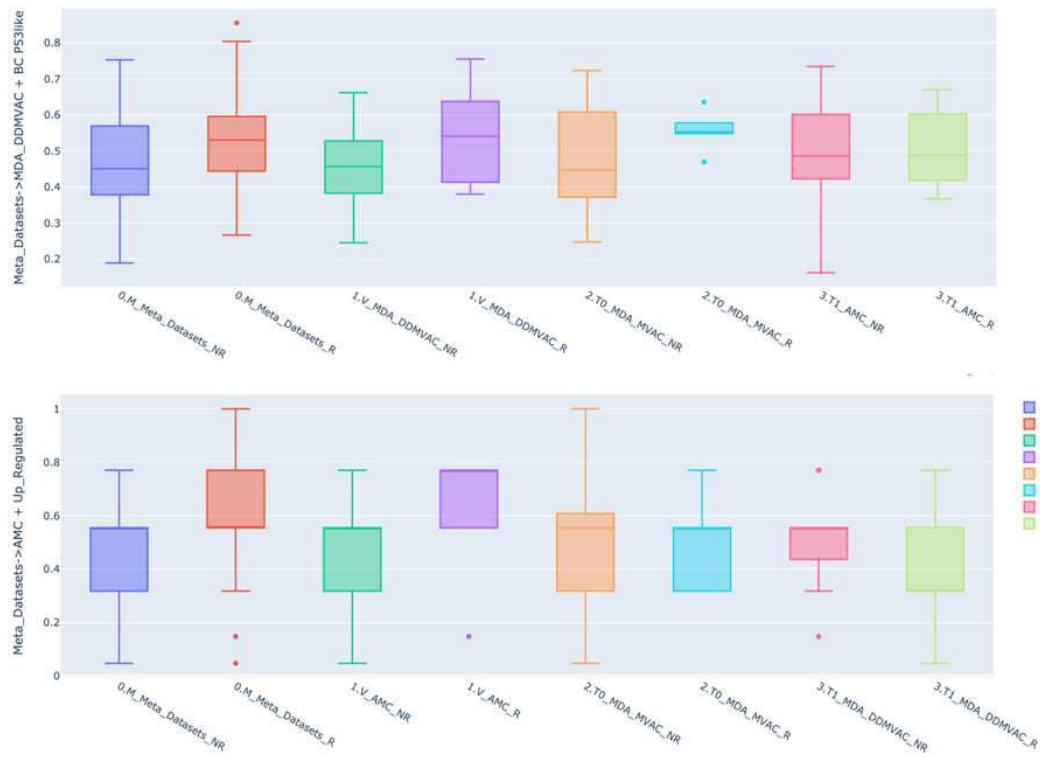
도면6ab

BC P63 pathway	Basal-squamous_refeau-ca	Luminal_refeau-cancer genome	Luminal-inf	Luminal-papillar	Neuronal_refe; GSEA_YES_Genes	Up_Regulated
ADA	AHNAK2	ADH6	ACTC1	ATOH8	ART5	AFAP1
CCNA2	BARX2	AGAP11	ACTG2	BTBD16	BEX1	ATAD38
CCNG1	CDH3	B3GAT1	ADAM33	CHD2	BRSK1	CADM4
CD82	CDSN	BHMT	BLK	CLCA4	CDH12	CARD16
CDK6	CXCL5	CCR7	C7	CRTAC1	CHGB	CD53
CKS2	CXCL9	EMX1	CCL19	CYP3A5	CHRN2	CDK6
F3	DSG3	FMO9P	CHRD1	CYP4F8	CNTFR	DPH2
GFI1	FYB	GABBR2	CILP	ERN2	DLX5	INHBA
IGFBP3	GBP5	GPLD1	COMP	FAM3D	DPF1	ITGAX
IGFBP6	GBP6	HS3ST6	DES	FGFR3	DUSP15	KCTD1
IL18	GNLY	IGDCC3	ELN	FOXA1	ENHO	KCTD14
ITGA3	HMGA2	IGF2	HSPB6	HS6ST3	FSD1	MYLK
KIF23	IDO1	KRT20	IGF1	HSD17B2	GFRA3	NEK6
LYN	IL20RB	MAL	MFAP4	LGALS4	GNG4	NT5E
MYC	KLK11	PKD1L1	MS4A1	PACRG	HOXC10	PCDH89
PI3	KRT6C	PRAP1	MYH11	REG4	HRA5LS	PLEK
PRNP	L1CAM	PSD2	OGN	SLC14A1	HUNK	POSTN
PTPN12	LAMC2	PVALB	OMD	TFF1	JPH3	PPIL2
S100A4	MUC16	SCNN1B	P2RX1	TRIM31	KCNH2	PTPRC
S100A8	PCSK9	SCNN1G	PDLIM3	UCA1	LHX2	RFX7
SERPINB2	PDZK1IP1	SLC1A6	PLA2G2D	UGT1A1	MAGED4B	SHANK3
SERPINE1	PI3	SLC6A4	PRELP		NAT8L	SIX5
SFN	PPP1R14C	SNX31	PRUNE2		NKAIN1	TFAM
SH2B3	SAA1	SPINK1	SFRP4		NOTUM	TMEM132A
TNFRSF10A	SERPINA1	UPK1A	SLC2A4		PEG10	TFAM
	SERPINA3	UPK3A	SMOC2		RIMS4	THBS2
	SERPINB7		SORBS1		SCG2	TMEM132A
	SPRR2A		SPON1		SLC38A3	TNFAIP8
	TGFB1		TCEAL2		SNAP25	
	TGM1		ZEB2		STAC2	
	TMEM45A				TWIST1	
	VSNL1				UCHL1	
	WARS				YBX2	

도면6b

Modeling Algorithm	Data/Modeling, Validation, Test0, Test1	Classifier										
			tr_AUC	val_AUC	ts0_AUC	ts1_AUC	tr_PRAUC	al_PRAUC	o_PRAUC	1_PRAUC	tr_R2	val_R2
Logistic Regression	Meta_Datasets, MDA, MVAAC, MDA, DDMVAAC, AMC	BC Luminal-1 6gene (union)	0.672852	0.794118	0.608333	0.578571	0.59268	0.606944	0.617866	0.354699	0.075575	-0.1361
	Meta_Datasets, MDA, DDMVAAC, MDA, MVAAC, AMC	BC P53 pathway (top10)	0.723752	0.891667	0.705882	0.367857	0.659874	0.8814	0.424206	0.166518	0.146277	0.356669
	Meta_Datasets, MDA, DDMVAAC, MDA, MVAAC, AMC	BC P53 pathway (top10)	0.653102	0.705556	0.539216	0.510714	0.606169	0.706582	0.294639	0.243945	0.058002	0.133152
	Meta_Datasets, MDA, DDMVAAC, MDA, MVAAC, AMC	BC P53like (top10)	0.632607	0.694444	0.696078	0.517857	0.580717	0.707386	0.377783	0.233428	0.052114	0.116252
	Meta_Datasets, AMC, MDA, MVAAC, MDA, DDMVAAC	BC P63 pathway (top10)	0.644188	0.782143	0.72549	0.522222	0.576635	0.494363	0.449074	0.521302	0.050597	-0.50174
Logistic Regression decision tree	Meta_Datasets, AMC, MDA, MVAAC, MDA, DDMVAAC	Up_Regulated (union)	0.720546	0.75	0.5	0.294444	0.626015	0.374547	0.261905	0.38769	0.163793	-0.54565
												-0.56546

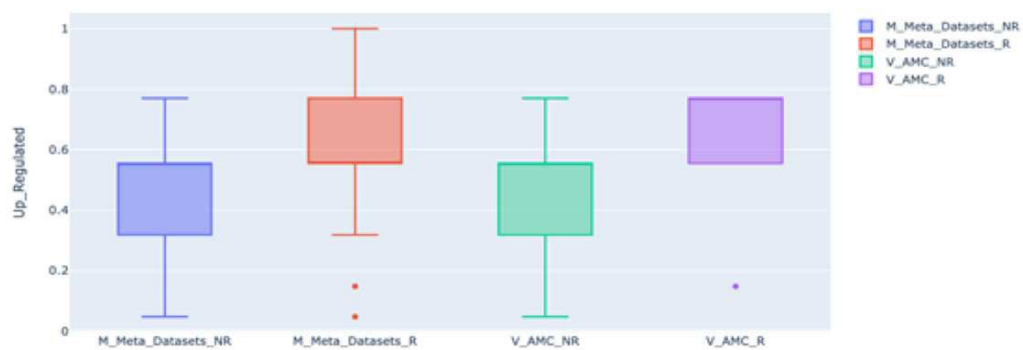
도면6c



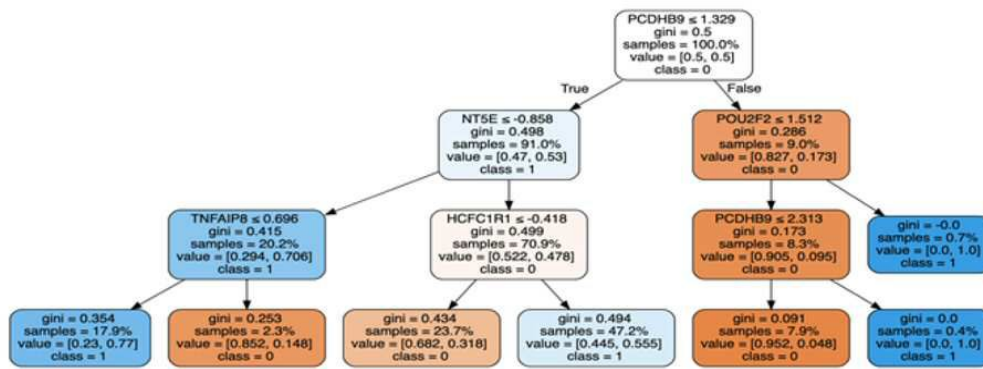
도면7a

Algorithm	Logistic Regression top10_BC_RCS pathway	Decision Tree						Random Forest	
		Union_BC Luminal		Union_Up Regulated		Union_Dn Regulated		top10_Luminal	refraucancer genome
	AUC	PRAUC	R ²	AUC	PRAUC	R ²	AUC	PRAUC	R ²
Classifier									
Performance									
Ranking Metadataets	0.641188	0.782143	0.576625	0.7442	0.661586	0.720546	0.626015	0.867076	0.845271
Validation AMC	0.494383	-0.050597	-0.501742	0.257333	-0.982405	0.374547	0.163793	0.345806	-0.377002

도면 7b



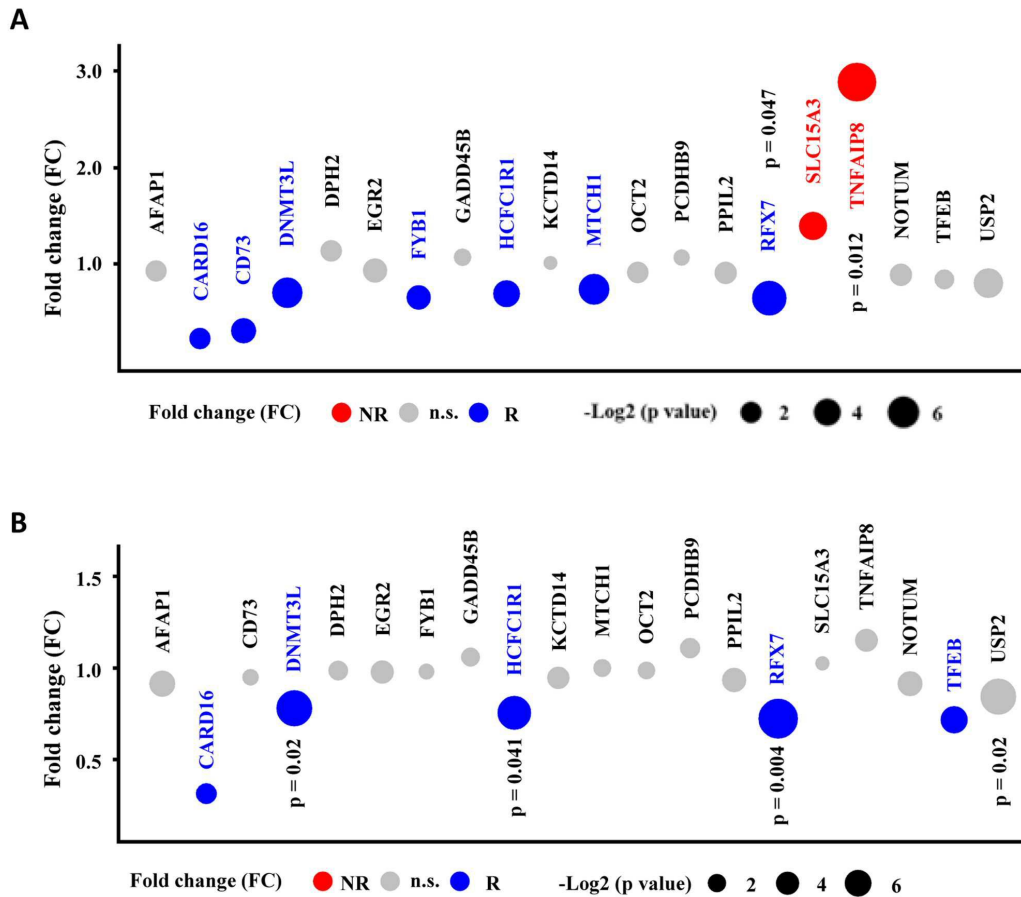
도면7c



도면7d

union_classifier	top10_classifier
AFAP1	PCDHB9
BHLHE40	HCFC1R1
C11orf95	RFX7
COX17	NT5E
HCFC1R1	ITGAX
ITGAX	TNFAIP8
KCTD14	KCTD14
KIAA1671	SLC15A3
NT5E	AFAP1
PCDHB9	POU2F2
PLEK	
POSTN	
POU2F2	
PPP1R35	
RFX7	
SLC15A3	
THBS2	
TNFAIP8	

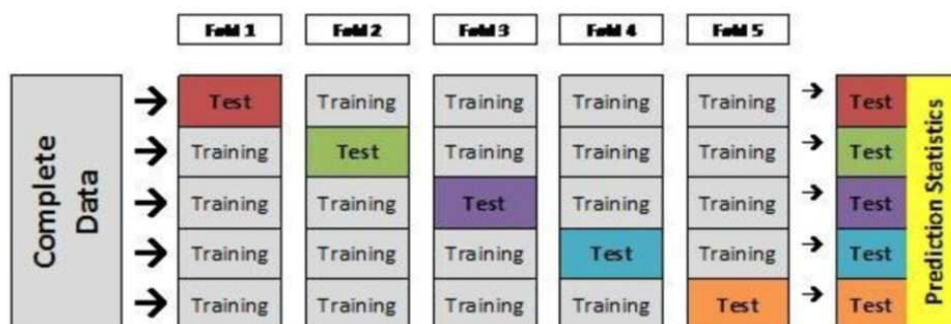
도면8



도면9a

metric name	metric description	condition1	condition2
pvalue_bt看_Rs	두 R 데이터 간 t-test p-value	0.05 이상으로 두 분포 유사	
pvalue_bt看_NRs	두 NR 데이터 간 t-test p-value	0.05 이상으로 두 분포 유사	
pvalue_in_train	학습 데이터 내 R과 NR 간 t-test p-value	0.05 미만으로 두 분포 상이	
pvalue_in_val	검증 데이터 내 R과 NR 간 t-test p-value	0.05 미만으로 두 분포 상이	
아래는 모두 cutoff=0.5 기준으로 이상이면 1, 미만이면 0으로 간주하고 계산, train_fdr과 val_fdr은 참고용으로 계산할 때 조건에 쓰이지 않음			
train_accuracy	전체 학습 데이터 중 정확히 답을 맞춘 비율	0.7 이상	현재 최적 모델보다 클 것
val_accuracy	전체 검증 데이터 중 정확히 답을 맞춘 비율	0.7 이상	현재 최적 모델보다 클 것
f1과 같이 잘 맞추는 척도는 이미 accuracy에서 보고 있으므로, 오류율에 대한 척도로 치환함			
train_fdr	fdr = false discovery rate = false positive / (false positive + true positive) 학습	-	-
val_fdr	검증 데이터 중 얼마나 잘못 R로 판단하고 있는지에 대한 척도	-	-
train_fpr	fpr = false positive rate = false positive / (false positive + true negative) 학습	-	현재 최적 모델보다 작을 것
val_fpr	검증 데이터 중 얼마나 실제 NR 중 잘못 판단하는지에 대한 척도	-	현재 최적 모델보다 작을 것

도면9b



도면9c

Gene Pool 명	gene list	#gene
tumor_tree_vals	[GLS, IL15RA, AFAP1, FOXA1]	4
std_tumor_tree_vals	[GLS, PSAT1, MITF, ICAM1, TRAF2, TRAF6, IL15RA, AFAP1, HCFC1R1, PPIL2, GlnRS (QARS1), Gluta(GSR), KIFC2, KIR3DL1, non-p (b-catenin)]	15
tumor_reg_vals	[GLS, PSAT1, CBS, CD44v6, FOXA1, GCLC, KIFC2, RPAP1, B-catenin, CK5, GATA3]	11
std_tumor_reg_vals	[GLS, PSAT1, CBS, ICAM1, TRAF6, IL15RA, AFAP1, CARD16, CD11c, EGR2, FYB1, GADD45B, HCFC1R1, MTCH1, OCT2, PCDHB9, RFX7, SLC15A3, ANPEP, BDNF, CD3-z, CD44v6, EMX1, EOGT, FZD9, GAD2, GCLC, GCLM, KIFC2, non-p (b-catenin), NOTUM, PLOD2, RPL9, SALL4, TFCP2L1, TFEB, USP2, GGT7, PPP2R5A, GATA3]	40
mean_tumor_vals	[GLS, PSAT1, CBS, IL15RA, TRAF6, RPAP1, B-catenin, AFAP1, KIFC2, ICAM1, TRAF2, PLOD2, TFEB]	13

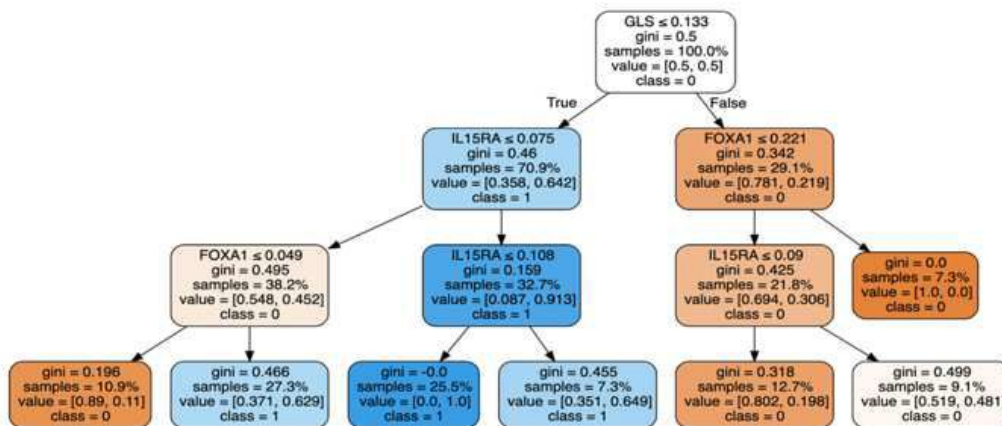
도면9d

Gene Pool 명	gene list	#gene
stroma_tree_vals	[CBS, EGR2, MTCH1, SALL4, B-catenin]	5
std_stroma_tree_vals	[CBS, GCLC, SOX2, CK5, CK20]	5
stroma_reg_vals	[CBS, FOXA1, GCLC, RPAP1, B-catenin]	5
std_stroma_reg_vals	[GLS, CBS, TRAF2, PPIL2, KIFC2]	5
mean_stroma_vals	[CBS, B-catenin, KIFC2, TFEB, MTCH1, SOX2, TRAF2, PPP2R5A, USP2, TRAF6, CK20]	11

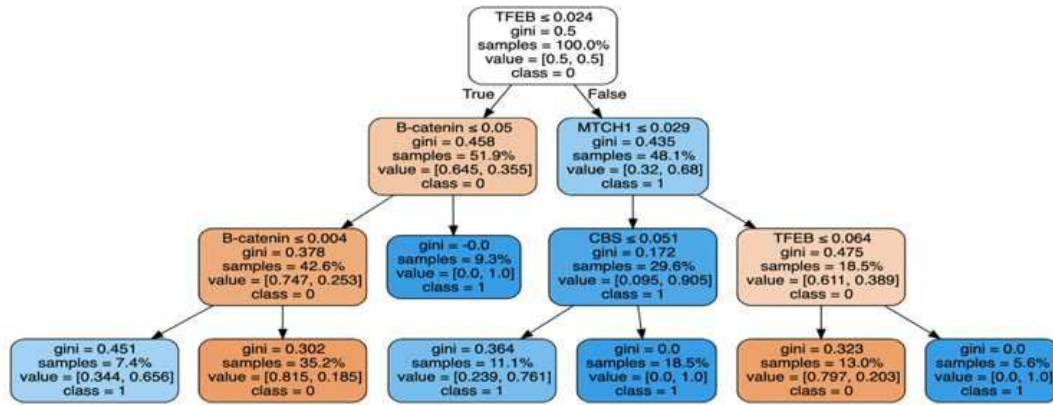
도면9e

	train_accuracy	val_accuracy	train_fdr	val_fdr	train_fpr	val_fpr	pvalue_bt看Rs	pvalue_bt看NR	pvalue_in_train	pvalue_in_val
tumor_tree_vals	0.8	0.75	0.151515	0.416667	0.238095	0.2	0.919528	0.254	3.25E-08	0.006185
std_tumor_reg_vals	0.763636	0.722222	0.137931	0.466667	0.190476	0.28	0.87007	0.248307	3.40E-05	0.033903

도면9f

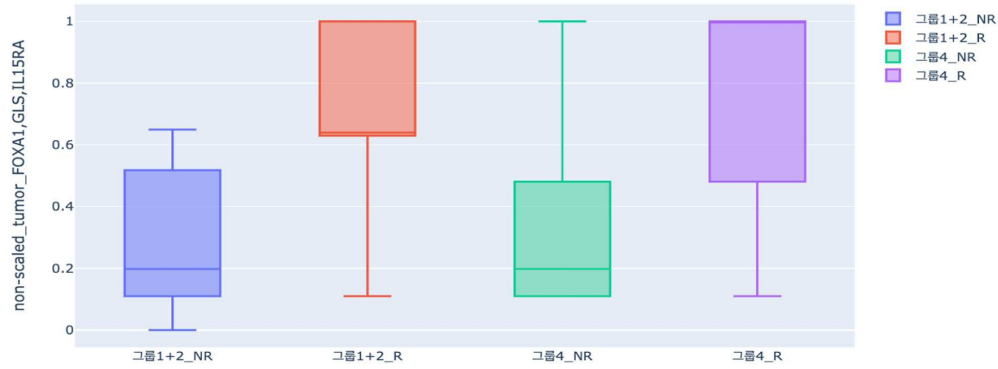


도면9g

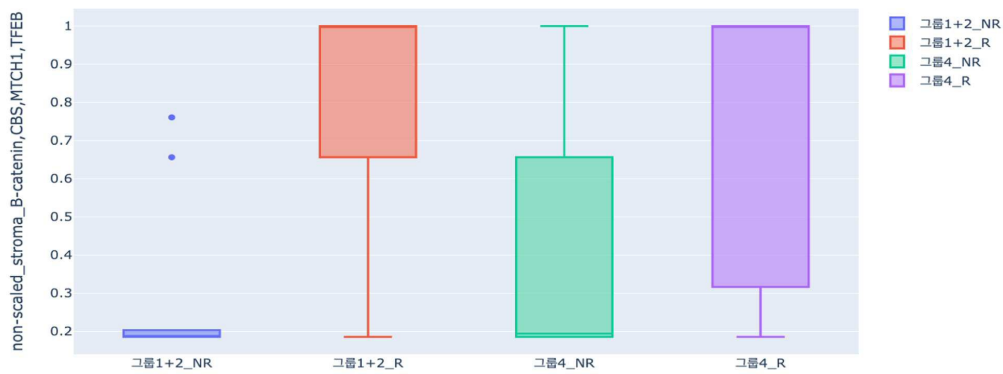


도면10

A



B



C

	non-scaled_tumor_ FOXA1, GLS, IL15RA	non-scaled_stroma_ B-catenin, CBS, MTCH1, TFEB
train_accuracy	0.8	0.833
val_accuracy	0.75	0.714
train_fdr	0.152	0.071
val_fdr	0.417	0.467
train_fpr	0.238	0.095
val_fpr	0.2	0.292
pvalue_btw_Rs	0.92	0.768
pvalue_btw_NRs	0.254	0.087
pvalue_in_train	0	0
pvalue_in_val	0.006	0.002

서 열 목 록 (첨부)



아이콘을 클릭하시면 서열목록 파일이 열립니다.

본 공보 PDF는 첨부파일을 가지고 있습니다. Acrobat Reader PDF뷰어를 제공하지 않는 브라우저(크롬, 파이어폭스, 사파리 등)의 경우 첨부파일 열기가 제한되어 있으므로 Acrobat Reader PDF뷰어 설치 후 공보 PDF를 다운로드 받아 해당 뷰어에서 조회해주시기 바랍니다.