



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2024년10월25일

(11) 등록번호 10-2722151

(24) 등록일자 2024년10월22일

- (51) 국제특허분류(Int. Cl.)  
G01N 33/574 (2006.01) G16H 50/20 (2018.01)  
G16H 50/50 (2018.01)
- (52) CPC특허분류  
G01N 33/57407 (2013.01)  
G16H 50/20 (2018.01)
- (21) 출원번호 10-2022-0128822
- (22) 출원일자 2022년10월07일  
심사청구일자 2022년10월07일
- (65) 공개번호 10-2024-0048931
- (43) 공개일자 2024년04월16일
- (56) 선행기술조사문헌  
KR1020210015259 A\*  
KR1020210012119 A  
KR1020190089552 A  
KR1020180003079 A  
\*는 심사관에 의하여 인용된 문헌

- (73) 특허권자  
서울대학교병원  
서울특별시 종로구 대학로 101(연건동)
- (72) 발명자  
정창욱  
서울특별시 용산구 이촌로 181, 104동 1602호(이촌동, 한강대우아파트)  
한도현  
서울특별시 도봉구 도봉로110다길 51, 101동 1401호(창동, 태영창동데시앙)  
서준교  
서울특별시 강남구 삼성로111길 8, 201동 1603호(삼성동, 삼성동힐스테이트2단지아파트)
- (74) 대리인  
특허법인리체

전체 청구항 수 : 총 12 항

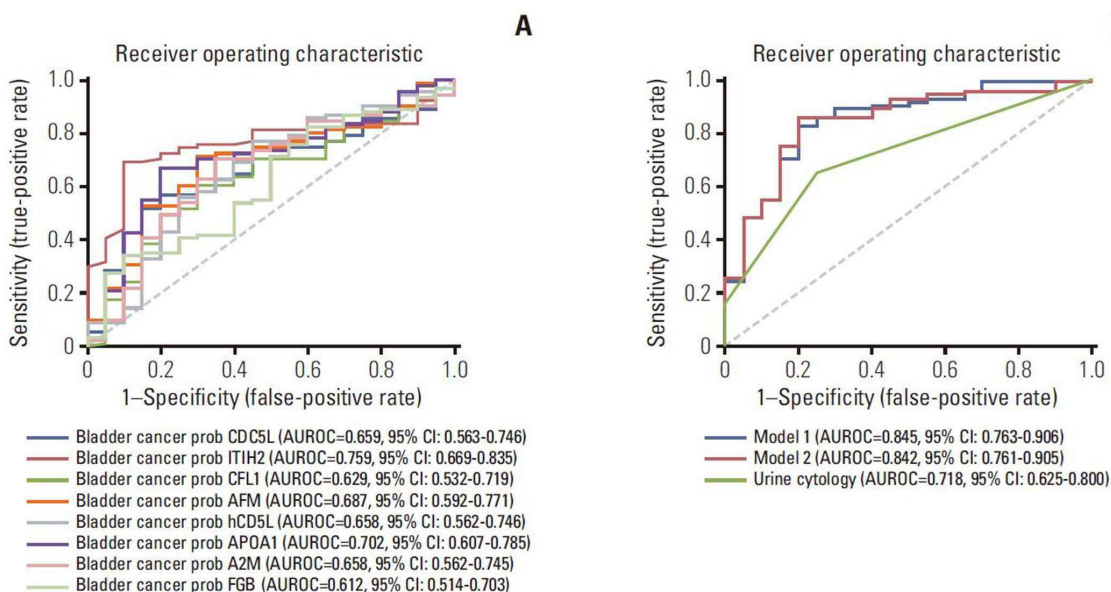
심사관 : 이수진

(54) 발명의 명칭 방광암 진단 시스템 및 방광암 진단에 필요한 정보 제공 방법

(57) 요약

본 발명은 방광암 진단 시스템에 관한 것으로, 보다 상세하게는 진단 대상 개체의 소변 시료 내 CDC5L(cell division cycle 5-like protein) 농도, ITIH2(inter-alpha-trypsin inhibitor heavy chain H2) 농도 및 CFL1(cofilin-1) 농도가 입력되는 입력부; 상기 입력부로부터 수신되는 상기 농도를 방광암 진단 모델에 대입하여 상기 진단 대상 개체가 방광암일 가능성을 계산하는 연산부; 및 상기 연산부에서 계산된 결과를 출력하는 출력부;를 포함함으로써 높은 민감도 및 특이성을 나타내면서도 비침습적인 방법으로 편리하게 방광암을 진단하는데 도움을 줄 수 있다.

대표도 - 도10



(52) CPC특허분류  
**G16H 50/50** (2018.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711027201
과제번호	2015M3A9E2028477
부처명	미래창조과학부
과제관리(전문)기관명	한국연구재단
연구사업명	바이오·의료기술개발
연구과제명	비침습방식으로 재발성 질환(방광암)의 조기 발견을 위한 소변 모니터링 자가 진단
기기용 플랫폼 개발	
기여율	1/1
과제수행기관명	전자부품연구원
연구기간	2015.06.01 ~ 2020.05.31
공지예외적용	: 있음

---

## 명세서

### 청구범위

#### 청구항 1

진단 대상 개체의 소변 시료 내 CDC5L(cell division cycle 5-like protein) 농도, ITIH2(inter-alpha-trypsin inhibitor heavy chain H2) 농도 및 CFL1(cofilin-1) 농도가 입력되는 입력부;

상기 입력부로부터 수신되는 상기 농도를 방광암 진단 모델에 대입하여 상기 진단 대상 개체가 방광암일 가능성을 계산하는 연산부; 및

상기 연산부에서 계산된 결과를 출력하는 출력부;를 포함하는, 방광암 진단 시스템.

#### 청구항 2

청구항 1에 있어서,

상기 방광암 진단 모델은 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도 및 방광암 여부 데이터로부터 생성된 것인, 방광암 진단 시스템.

#### 청구항 3

청구항 1에 있어서,

상기 입력부로부터 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도 및 방광암 여부 데이터를 수신하고, 상기 수신된 데이터로부터 방광암 진단 모델을 생성하는 모델 생성부를 더 포함하는, 방광암 진단 시스템.

#### 청구항 4

청구항 1에 있어서,

상기 입력부에 상기 진단 대상 개체의 AFM(afamin) 농도, CD5L(CD5 antigen-like protein) 농도, APOA1(apolipoprotein A-I) 농도, A2M(alpha-2 macroglobulin) 농도 및 FGB(fibrinogen beta chain) 농도가 더 입력되는, 방광암 진단 시스템.

#### 청구항 5

청구항 4에 있어서,

상기 방광암 진단 모델은 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도, FGB 농도 및 방광암 여부 데이터로부터 생성된 것인, 방광암 진단 시스템.

#### 청구항 6

청구항 4에 있어서,

상기 입력부로부터 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도, FGB 농도 및 방광암 여부 데이터를 수신하고, 상기 수신된 데이터 데이

터로부터 방광암 진단 모델을 생성하는 모델 생성부를 더 포함하는, 방광암 진단 시스템.

**청구항 7**

진단 대상 개체의 소변 시료 내 CDC5L(cell division cycle 5-like protein) 농도, ITIH2(inter-alpha-trypsin inhibitor heavy chain H2) 농도 및 CFL1(cofilin-1) 농도를 방광암 진단 모델에 대입하여, 상기 진단 대상 개체가 방광암일 가능성을 계산하여 제공하는 단계를 포함하는 방광암 진단에 필요한 정보 제공 방법.

**청구항 8**

청구항 7에 있어서,

상기 방광암 진단 모델은 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도 및 방광암 여부 데이터로부터 생성된 것인, 방광암 진단에 필요한 정보 제공 방법.

**청구항 9**

청구항 7에 있어서,

모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도 및 방광암 여부 데이터로부터 방광암 진단 모델을 생성하는 단계를 더 포함하는, 방광암 진단에 필요한 정보 제공 방법.

**청구항 10**

청구항 7에 있어서,

상기 방광암 진단 모델에 상기 진단 대상 개체의 AFM(afamin) 농도, CD5L(CD5 antigen-like protein) 농도, APOA1(apolipoprotein A-I) 농도, A2M(alpha-2 macroglobulin) 농도 및 FGB(fibrinogen beta chain) 농도를 더 대입하는, 방광암 진단에 필요한 정보 제공 방법.

**청구항 11**

청구항 10에 있어서,

상기 방광암 진단 모델은 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도, FGB 농도 및 방광암 여부 데이터로부터 생성된 것인, 방광암 진단에 필요한 정보 제공 방법.

**청구항 12**

청구항 10에 있어서,

모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도, FGB 농도 및 방광암 여부 데이터로부터 방광암 진단 모델을 생성하는 단계를 더 포함하는, 방광암 진단에 필요한 정보 제공 방법.

**발명의 설명**

**기술 분야**

[0001] 본 발명은 진단 대상 개체의 소변 시료 내 단백질 농도를 이용하는 방광암 진단 시스템 및 방광암 진단에 필요한 정보 제공 방법에 관한 것이다.

**배경 기술**

[0003] 방광암은 비뇨기계 영역에서 가장 빈번하게 발생하는 암이다. 방광암의 진단 및 예후를 예측하기 위한 현행 표준으로서 방광경 검사를 기반으로 한다. 방광경 검사는 카테터를 방광에 밀어 넣어 의심되는 조직을 떼어내 검사하는 침습적 방법으로, 방광 점막의 형태학적 특징으로 인해 위음성 결과가 초래될 수 있다.

[0004] 이러한 방광경 기반 검사의 한계를 극복하기 위해 소변 세포검사가 개발되었으나, 소변 세포검사는 낮은 민감도 및 재현성의 한계를 갖는다. 이러한 이유로, 방광암에 대해 편리하고 정확한 진단 바이오마커 개발에 대한 임상적 요구가 여전히 존재한다.

[0005] 최근 수십년 동안, BTA 분석 (Polymedco Inc., NY, USA), NMP22 (Alere, MA, USA)과 같은 수많은 소변 바이오마커가 발견되었으며, 그 중 일부는 FDA 임상 사용 승인을 받았다. 그러나, 진단 성능이 여전히 제한적이며, BTA 분석은 혈뇨 환자에게 위양성 결과가 나타날 위험이 있고, NMP22는 일반 집단에 비해 비침습적 바이오마커가 필요한 고위험 집단에서 극도로 낮은 민감도를 갖는다는 문제점이 있다.

[0006] 따라서, 혈뇨 또는 방광 종괴가 있는 환자와 같이 임상적으로 구분할 수 없는 상황에서 허용 가능한 성능을 가진 바이오마커가 필요한 실정이다.

[0007] 이에, 본 발명자는 비침습적으로 편리하게 검사할 수 있으면서도 높은 민감도를 갖는 방광암 진단용 바이오마커를 완성하였다.

**선행기술문헌**

**특허문헌**

[0009] (특허문헌 0001) 한국공개특허 제10-2018-0024009호

**발명의 내용**

**해결하려는 과제**

[0010] 본 발명은 방광암 진단 시스템을 제공하는 것을 목적으로 한다.

[0011] 본 발명은 방광암 진단에 필요한 정보 제공 방법을 제공하는 것을 목적으로 한다.

**과제의 해결 수단**

[0013] 1. 진단 대상 개체의 소변 시료 내 CDC5L(cell division cycle 5-like protein) 농도, ITIH2(inter-alpha-trypsin inhibitor heavy chain H2) 농도 및 CFL1(cofilin-1) 농도가 입력되는 입력부;

[0014] 상기 입력부로부터 수신되는 상기 농도를 방광암 진단 모델에 대입하여 상기 진단 대상 개체가 방광암일 가능성을 계산하는 연산부; 및

[0015] 상기 연산부에서 계산된 결과를 출력하는 출력부;를 포함하는, 방광암 진단 시스템.

[0016] 2. 위 1에 있어서,

[0017] 상기 방광암 진단 모델은 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도 및 방광암 여부 데이터로부터 생성된 것인, 방광암 진단 시스템.

[0018] 3. 위 1에 있어서,

[0019] 상기 입력부로부터 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도 및 방광암 여부 데이터를 수신하고, 상기 수신된 데이터로부터 방광암 진단 모델을 생성하는 모델 생성부를 더 포함하는, 방광암 진단 시스템.

- [0020] 4. 위 1에 있어서,
- [0021] 상기 입력부에 상기 진단 대상 개체의 AFM(afamin) 농도, CD5L(CD5 antigen-like protein) 농도, APOA1(apolipoprotein A-I) 농도, A2M(alpha-2 macroglobulin) 농도 및 FGB(fibrinogen beta chain) 농도가 더 입력되는, 방광암 진단 시스템.
- [0022] 5. 위 4에 있어서,
- [0023] 상기 방광암 진단 모델은 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도, FGB 농도 및 방광암 여부 데이터로부터 생성된 것인, 방광암 진단 시스템.
- [0024] 6. 위 4에 있어서,
- [0025] 상기 입력부로부터 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도, FGB 농도 및 방광암 여부 데이터를 수신하고, 상기 수신된 데이터로부터 방광암 진단 모델을 생성하는 모델 생성부를 더 포함하는, 방광암 진단 시스템.
- [0026] 7. 진단 대상 개체의 소변 시료 내 CDC5L(cell division cycle 5-like protein) 농도, ITIH2(inter-alpha-trypsin inhibitor heavy chain H2) 농도 및 CFL1(cofilin-1) 농도를 방광암 진단 모델에 대입하여, 상기 진단 대상 개체가 방광암일 가능성을 계산하여 제공하는 단계를 포함하는 방광암 진단에 필요한 정보 제공 방법.
- [0027] 8. 위 7에 있어서,
- [0028] 상기 방광암 진단 모델은 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도 및 방광암 여부 데이터로부터 생성된 것인, 방광암 진단에 필요한 정보 제공 방법.
- [0029] 9. 위 7에 있어서,
- [0030] 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도 및 방광암 여부 데이터로부터 방광암 진단 모델을 생성하는 단계를 더 포함하는, 방광암 진단에 필요한 정보 제공 방법.
- [0031] 10. 위 7에 있어서,
- [0032] 상기 방광암 진단 모델에 상기 진단 대상 개체의 AFM(afamin) 농도, CD5L(CD5 antigen-like protein) 농도, APOA1(apolipoprotein A-I) 농도, A2M(alpha-2 macroglobulin) 농도 및 FGB(fibrinogen beta chain) 농도를 더 대입하는, 방광암 진단에 필요한 정보 제공 방법.
- [0033] 11. 위 10에 있어서,
- [0034] 상기 방광암 진단 모델은 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도, FGB 농도 및 방광암 여부 데이터로부터 생성된 것인, 방광암 진단에 필요한 정보 제공 방법.
- [0035] 12. 위 10에 있어서,
- [0036] 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도, FGB 농도 및 방광암 여부 데이터로부터 방광암 진단 모델을 생성하는 단계를 더 포함하는, 방광암 진단에 필요한 정보 제공 방법.

**발명의 효과**

- [0038] 본 발명의 방광암 진단 시스템 및 방광암 진단에 필요한 정보 제공 방법은 높은 민감도 및 특이성을 나타내며 비침습적인 방법으로 편리하게 방광암을 진단하는 데에 도움을 줄 수 있다. 또한, 본 발명의 시스템 및 정보 제공 방법은 단독으로 혹은 기존의 진단 시스템/방법과 함께 사용하여 대상 개체의 방광암을 진단하는 데에 도움을 줄 수 있다.

**도면의 간단한 설명**

- [0040] 도 1은 본 발명의 소변 단백질 바이오마커 개발의 전체 작업흐름을 나타낸 것이다.  
 도 2a 내지 도 2d는 발견 단계에서 무표지 정량화의 결과를 나타낸 것으로, 도 2a는 무표지 정량화의 단백질체학

작업흐름, 도 2b는 소변 및 엑소솜에서 식별 및 정량화 결과, 도 2c는 볼케이노 플롯(volcano plot), 도 2d는 PCA(principal component analysis) 결과를 나타낸다.

도 3은 NTA에 의해 측정된 분리된 엑소솜의 유체역학적 크기 분포 프로파일을 나타낸 것이다.

도 4a 및 도 4b는 각각 발견 코호트(discovery cohort)에서 엑소솜(도 4a) 및 소변(도 4b) DEPs (differentially expressed proteins)의 열지도를 나타낸 것이다.

도 5a 내지 도 5c는 검증 단계에서 데이터 비의존적 수집의 결과를 나타낸 것으로, 도 5a는 데이터 비의존적 수집 작업흐름, 도 5b는 데이터 비의존적 수집 방식을 사용한 검증 프로세스의 흐름도, 도 5c는 발견 코호트와 검증 코호트 사이의 단백질 폴드 체인지(대조군/요로상피암(Urothelial Carcinoma) 환자)의 상관 관계를 나타낸다.

도 6a 내지 도 6c는 검증 코호트(verification cohort)에서의 단백질체 분석 결과로, 도 6a는 소변 및 엑소솜 단백질의 식별 및 정량화 결과, 도 6b는 볼케이노 플롯, 도 6c는 PCA 결과를 나타낸다.

도 7a 및 도 7b는 각각 검증 코호트(verification cohort)에서 소변 엑소솜(도 7a) 및 소변(도 7b) DEPs (differentially expressed proteins)의 열지도를 나타낸 것이다.

도 8a는 ELISA(enzyme-linked immunosorbent assay) 연구에서 단백질 발현 데이터의 왜곡도(skewness)를 나타낸 것이다. \* 표시 도면은 왜곡된 데이터를 나타낸다.

도 8b는 원 데이터에서 왜곡된 단백질(ITIH2, AFM, hCD5L, APOA1 및 FGB) 발현에 대한 자연 로그 변환 후 왜곡도(skewness)를 나타낸다.

도 9는 ELISA 연구에서 T 단계 간의 단백질 발현 수준에 대한 박스플롯을 나타낸다. (ns: not significant, \*:  $1.00e-02 < p \leq 5.00e-02$ , \*\*:  $1.00e-03 < p \leq 1.00e-02$ , \*\*\*:  $1.00e-04 < p \leq 1.00e-03$ , \*\*\*\*:  $p \leq 1.00e-04$ )

도 10은 각 후보 단백질 (A) 및 개발된 다중 바이오마커 모델 (B)에 의한 방광암 진단을 위한 AUROC를 나타낸 것이다 (선택된 단백질에 대한 모델 1 및 모든 단백질 기반 모델에 대한 모델 2).

도 11은 소변 세포검사 및 본 발명에서 개발된 다중 소변 바이오마커 모델의 조합의 AUROC를 나타낸다.

### 발명을 실시하기 위한 구체적인 내용

- [0041] 이하, 본 발명을 상세히 설명한다.
- [0043] 본 발명은 방광암 진단 시스템을 제공한다.
- [0044] 본 발명의 방광암 진단 시스템은 진단 대상 개체의 소변 시료 내 CDC5L(cell division cycle 5-like protein) 농도, ITIH2(inter-alpha-trypsin inhibitor heavy chain H2) 농도 및 CFL1(cofilin-1) 농도가 입력되는 입력부; 상기 입력부로부터 수신되는 상기 농도를 방광암 진단 모델에 대입하여 상기 진단 대상 개체가 방광암일 가능성을 계산하는 연산부; 및 상기 연산부에서 계산된 결과를 출력하는 출력부;를 포함한다.
- [0045] 상기 입력부는 상기 진단 대상 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도 및 CFL1 농도가 입력되는 것으로, 상기 단백질들의 농도는 진단 대상 개체 또는 다른 개체에 의해 직접 입력되거나 별도의 농도 획득 수단으로부터 단백질들의 농도 데이터를 수신 받아 입력될 수 있다. 상기 농도 획득 수단은 기 측정된 상기 단백질들의 농도 데이터가 저장되거나 별도로 구비된 단백질 농도 측정부에 의해 측정된 농도 데이터가 저장될 수 있다.
- [0046] 상기 진단 대상 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도 및 CFL1 농도는 진단 대상 개체의 소변 시료 내에서 각 단백질의 발현량을 의미한다. 상기 단백질들의 농도는 예를 들면, ELISA(enzyme-linked immunosorbent assay) 등에 의해 측정된 것일 수 있다.
- [0047] 상기 연산부는 상기 입력부로부터 수신된 진단 대상 개체의 소변 시료 내 단백질 농도를 기반으로 상기 연산부에 기 저장된 방광암 진단 모델 또는 별도로 구비된 모델 생성부에 의해 생성된 방광암 진단 모델을 적용하여 상기 진단 대상 개체가 방광암일 가능성을 계산한다.
- [0048] 상기 방광암 진단 모델은 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도 및 방광암 여부 데이터로부터 생성된 것일 수 있다.

- [0049] 상기 방광암 진단 시스템은 상기 입력부로부터 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도 및 방광암 여부 데이터를 수신하고, 상기 수신된 데이터로부터 방광암 진단 모델을 생성하는 모델 생성부를 더 포함할 수 있다.
- [0050] 상기 모델 생성 대상 집단은 상기 방광암 진단 시스템에 이용되는 방광암 진단 모델 생성에 활용되는 데이터를 제공하는 집단이다.
- [0051] 상기 방광암 진단 모델은 교사 학습(supervised learning)에 의해 생성된 것일 수 있다. 구체적으로, 상기 방광암 진단 모델은 상기 모델 생성 대상 집단의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도 및 방광암 여부를 교사 데이터로서 이용하고, 입력을 각 개체의 상기 CDC5L 농도, ITIH2 농도 및 CFL1 농도로 하고, 출력을 각 개체의 방광암 여부로 하여 기계학습에 의해 생성된 것일 수 있다.
- [0052] 상기 방광암 진단 모델을 생성하기 위해 상기 단백질 ITIH2의 농도를 로그 변환(logarithmic transformation)된 형태로 이용할 수 있다.
- [0053] 방광암 진단 모델 생성에는 공지된 통계 툴, 모델, 알고리즘 등이 활용될 수 있다. 예를 들면 Python, 로지스틱 회귀(예컨대, 다변수 로지스틱 회귀), 머신 러닝 분류 알고리즘 등이 활용될 수 있다.
- [0054] 상기 출력부는 상기 연산부에서 계산된 결과인 상기 진단 대상 개체가 방광암일 가능성이 출력된다. 상기 계산된 결과는 예를 들면 0 내지 1, 0% 내지 100% 등의 값으로 출력되며, 상기 출력된 결과가 클수록 상기 진단 대상 개체는 방광암일 가능성이 높은 것으로 판단할 수 있다.
- [0055] 상기 출력부는 사용자(예컨대, 의사, 환자 등)에게 상기 방광암일 가능성을 제공할 수 있는 형태라면 제한되지 않고 이용 가능하다. 예를 들면, 상기 계산된 결과를 화면이 구비된 형태의 출력부에 의해 시각적으로 출력되거나, 스피커가 구비된 형태의 출력부에 의해 청각적으로 출력될 수 있다.
- [0056] 방광암 여부를 확인하고자 하는 진단 대상 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도 및 CFL1 농도를 본 발명의 방광암 진단 시스템의 입력부에 입력하면, 연산부에서 기 저장된 또는 모델 생성부에 의해 생성된 방광암 진단 모델에 의해 상기 진단 대상 개체가 방광암일 가능성이 계산되고, 연산부에서 계산된 결과가 출력부에 출력된다. 상기 출력된 결과가 클수록 상기 진단 대상 개체는 방광암일 가능성이 높은 것으로 판단할 수 있다.
- [0058] 본 발명의 방광암 진단 시스템의 입력부에 상기 진단 대상 개체의 AFM(afamin) 농도, CD5L(CD5 antigen-like protein) 농도, APOA1(apolipoprotein A-1) 농도, A2M(alpha-2 macroglobulin) 농도 및 FGB(fibrinogen beta chain) 농도가 더 입력될 수 있다.
- [0059] 즉, 본 발명의 방광암 진단 시스템은 진단 대상 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도 및 FGB 농도가 입력되는 입력부; 상기 입력부로부터 수신되는 상기 농도를 방광암 진단 모델에 대입하여 상기 진단 대상 개체가 방광암일 가능성을 계산하는 연산부; 및 상기 연산부에서 계산된 결과를 출력하는 출력부;를 포함할 수 있다.
- [0060] 상기 입력부에 입력되는 상기 진단 대상 개체의 소변 시료 내 단백질들의 농도는 전술한 바와 같이 진단 대상 개체 또는 다른 개체에 의해 직접 입력되거나 별도의 농도 획득 수단으로부터 단백질들의 농도 데이터를 수신 받아 입력될 수 있다. 상기 농도 획득 수단, 상기 단백질 농도의 의미 및 측정 방법의 예시는 전술한 바와 같다.
- [0061] 상기 연산부는 전술한 바와 같이 상기 입력부로부터 수신된 단백질 농도를 기반으로 진단 대상 개체가 방광암일 가능성을 계산한다.
- [0062] 상기 방광암 진단 모델은 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도, FGB 농도 및 방광암 여부 데이터로부터 생성된 것일 수 있다.
- [0063] 상기 방광암 진단 시스템은 상기 입력부로부터 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도, FGB 농도 및 방광암 여부 데이터를 수신하고, 상기 수신된 데이터로부터 방광암 진단 모델을 생성하는 모델 생성부를 더 포함할 수 있다.
- [0064] 상기 방광암 진단 모델은 교사 학습(supervised learning)에 의해 생성된 것일 수 있다. 구체적으로, 상기 방광암 진단 모델은 상기 모델 생성 대상 집단의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도, FGB 농도 및 방광암 여부를 교사 데이터로서 이용하고, 입력을 각 개체의 상기 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도 및 FGB 농도로 하고, 출력을

각 개체의 방광암 여부로 하여 기계학습에 의해 생성된 것일 수 있다.

- [0065] 상기 방광암 진단 모델을 생성하기 위해 상기 단백질 ITIH2, AFM, CD5L, APOA1 및 FGB 중 적어도 하나의 농도를 로그 변환(logarithmic transformation)된 형태로 이용할 수 있다.
- [0066] 방광암 진단 모델 생성에는 공지된 통계 틀, 모델, 알고리즘 등이 활용될 수 있다. 예를 들면 Python, 로지스틱 회귀(예컨대, 다변수 로지스틱 회귀), 머신 러닝 분류 알고리즘 등이 활용될 수 있다.
- [0067] 상기 출력부는 전술한 바와 같다.
- [0068] 방광암 여부를 확인하고자 하는 진단 대상 개체의 기 측정된 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도 및 FGB 농도를 본 발명의 방광암 진단 시스템의 입력부에 입력하면, 연산부에서 기 저장된 또는 모델 생성부에 의해 생성된 방광암 진단 모델에 의해 상기 진단 대상 개체가 방광암일 가능성이 계산되고, 연산부에서 계산된 결과가 출력부에 출력된다. 상기 출력된 결과가 클수록 상기 진단 대상 개체는 방광암일 가능성이 높은 것으로 판단할 수 있다.
- [0069] 본 발명의 방광암 진단 시스템은 높은 민감도 및 특이성을 나타내며 비침습적인 방법으로 편리하게 방광암을 진단하는 데에 도움을 줄 수 있다. 본 발명의 방광암 진단 시스템은 단독으로 혹은 기존의 요세포 검사와 함께 사용하여 대상 개체의 방광암을 진단하는 데에 도움을 줄 수 있다.
- [0071] 또한, 본 발명은 방광암 진단에 필요한 정보 제공 방법을 제공한다.
- [0072] 본 발명의 방광암 진단에 필요한 정보 제공 방법은 진단 대상 개체의 소변 시료 내 CDC5L(cell division cycle 5-like protein) 농도, ITIH2(inter-alpha-trypsin inhibitor heavy chain H2) 농도 및 CFL1(cofilin-1) 농도를 방광암 진단 모델에 대입하여, 상기 진단 대상 개체가 방광암일 가능성을 계산하여 제공하는 단계를 포함한다.
- [0073] 상기 방광암 진단에 필요한 정보 제공 방법은 상기 진단 대상 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도 및 CFL1 농도를 측정하는 단계를 더 포함할 수 있다. 상기 단백질들의 농도는 진단 대상 개체의 소변 시료 내에서 각 단백질의 발현량을 의미하고, 이는 예를 들면, ELISA(enzyme-linked immunosorbent assay) 등에 의해 측정된 것일 수 있다.
- [0074] 상기 진단 대상 개체가 방광암일 가능성을 계산하여 제공하는 단계에서는 기 생성된 방광암 진단 모델 또는 별도의 모델 생성 단계에 의해 생성된 방광암 진단 모델을 이용하여 상기 진단 대상 개체가 방광암일 가능성이 계산된다.
- [0075] 상기 방광암 진단 모델은 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도 및 방광암 여부 데이터로부터 생성된 것일 수 있다.
- [0076] 상기 방광암 진단에 필요한 정보 제공 방법은 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도 및 방광암 여부 데이터로부터 방광암 진단 모델을 생성하는 단계를 더 포함할 수 있다.
- [0077] 상기 모델 생성 대상 집단은 상기 방광암 진단에 필요한 정보 제공 방법에 이용되는 방광암 진단 모델 생성에 활용되는 데이터를 제공하는 집단이다.
- [0078] 상기 방광암 진단 모델을 생성하는 단계는 교사 학습(supervised learning)에 의해 수행될 수 있다. 구체적으로, 상기 방광암 진단 모델을 생성하는 단계는 상기 모델 생성 대상 집단의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도 및 방광암 여부를 교사 데이터로서 이용하고, 입력을 각 개체의 상기 CDC5L 농도, ITIH2 농도 및 CFL1 농도로 하고, 출력을 각 개체의 방광암 여부로 하여 기계학습에 의해 방광암 진단 모델을 생성하는 것일 수 있다.
- [0079] 상기 방광암 진단 모델을 생성하기 위해 상기 단백질 ITIH2의 농도를 로그 변환(logarithmic transformation)된 형태로 이용할 수 있다.
- [0080] 방광암 진단 모델 생성에는 공지된 통계 틀, 모델, 알고리즘 등이 활용될 수 있다. 예를 들면 Python, 로지스틱 회귀(예컨대, 다변수 로지스틱 회귀), 머신 러닝 분류 알고리즘 등이 활용될 수 있다.
- [0081] 상기 계산된 진단 대상 개체가 방광암일 가능성은 예를 들면 0 내지 1, 0% 내지 100% 등의 값일 수 있고, 상기 계산된 결과가 클수록 상기 진단 대상 개체는 방광암일 가능성이 높은 것으로 판단할 수 있다.
- [0082] 방광암 여부를 확인하고자 하는 진단 대상 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도 및 CFL1 농도를 기 생

성된 또는 방광암 진단 모델을 생성하는 단계에 의해 생성된 방광암 진단 모델에 대입하면 상기 진단 대상 개체가 방광암일 가능성을 계산할 수 있다. 상기 계산된 결과가 클수록 상기 진단 대상 개체는 방광암일 가능성이 높은 것으로 판단할 수 있다.

- [0084] 본 발명의 방광암 진단에 필요한 정보 제공 방법은 진단 대상 개체의 소변 시료 내 CDC5L(cell division cycle 5-like protein) 농도, ITIH2(inter-alpha-trypsin inhibitor heavy chain H2) 농도, CFL1(cofilin-1) 농도, AFM(afamin) 농도, CD5L(CD5 antigen-like protein) 농도, APOA1(apolipoprotein A-I) 농도, A2M(alpha-2 macroglobulin) 농도 및 FGB(fibrinogen beta chain) 농도를 방광암 진단 모델에 대입하여, 상기 진단 대상 개체가 방광암일 가능성을 계산하여 제공하는 단계를 포함할 수 있다.
- [0085] 상기 방광암 진단에 필요한 정보 제공 방법은 상기 진단 대상 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도 및 FGB 농도를 측정하는 단계를 더 포함할 수 있다. 상기 단백질 농도의 의미 및 측정 방법의 예시는 전술한 바와 같다.
- [0086] 상기 진단 대상 개체가 방광암일 가능성을 계산하여 제공하는 단계에서는 전술한 바와 같이 기 생성된 방광암 진단 모델 또는 별도의 모델 생성 단계에 의해 생성된 방광암 진단 모델을 이용하여 상기 진단 대상 개체가 방광암일 가능성이 계산된다
- [0087] 상기 방광암 진단 모델은 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도, FGB 농도 및 방광암 여부 데이터로부터 생성된 것일 수 있다.
- [0088] 상기 방광암 진단에 필요한 정보 제공 방법은 모델 생성 대상 집단의 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도, FGB 농도 및 방광암 여부 데이터로부터 방광암 진단 모델을 생성하는 단계를 더 포함할 수 있다.
- [0089] 상기 방광암 진단 모델을 생성하는 단계는 교사 학습(supervised learning)에 의해 수행될 수 있다. 구체적으로, 상기 방광암 진단 모델을 생성하는 단계는 상기 모델 생성 집단의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도, FGB 농도 및 방광암 여부를 교사 데이터로서 이용하고, 입력을 각 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도 및 FGB 농도로 하고, 출력을 각 개체의 방광암 여부로 하여 기계학습에 의해 방광암 진단 모델을 생성하는 것일 수 있다.
- [0090] 상기 방광암 진단 모델을 생성하기 위해 상기 단백질 ITIH2, AFM, CD5L, APOA1 및 FGB 중 적어도 하나의 농도를 로그 변환(logarithmic transformation)된 형태로 이용할 수 있다.
- [0091] 방광암 진단 모델 생성에는 공지된 통계 툴, 모델, 알고리즘 등이 활용될 수 있다. 예를 들면 Python, 로지스틱 회귀(예컨대, 다변수 로지스틱 회귀), 머신 러닝 분류 알고리즘 등이 활용될 수 있다.
- [0092] 상기 계산된 진단 대상 개체가 방광암일 가능성은 예를 들면 0 내지 1, 0% 내지 100% 등의 값일 수 있고, 상기 계산된 결과가 클수록 상기 진단 대상 개체는 방광암일 가능성이 높은 것으로 판단할 수 있다.
- [0093] 방광암 여부를 확인하고자 하는 진단 대상 개체의 소변 시료 내 CDC5L 농도, ITIH2 농도, CFL1 농도, AFM 농도, CD5L 농도, APOA1 농도, A2M 농도 및 FGB 농도를 기 생성된 또는 방광암 진단 모델을 생성하는 단계에 의해 생성된 방광암 진단 모델에 대입하면 상기 진단 대상 개체가 방광암일 가능성을 계산할 수 있다. 상기 계산된 결과가 클수록 상기 진단 대상 개체는 방광암일 가능성이 높은 것으로 판단할 수 있다.
- [0094] 본 발명의 방광암 진단에 필요한 정보 제공 방법은 높은 민감도 및 특이성을 나타내며 비침습적인 방법으로 편리하게 방광암을 진단하는 데에 도움을 줄 수 있다. 본 발명의 방광암 진단에 필요한 정보 제공 방법은 단독으로 혹은 기존의 요세포 검사와 함께 사용하여 대상 개체의 방광암을 진단하는 데에 도움을 줄 수 있다.
- [0096] 이하, 실시예를 통하여 본 발명을 더욱 상세히 설명하고자 한다. 이들 실시예는 오로지 본 발명을 보다 구체적으로 설명하기 위한 것으로서, 이에 의해 본 발명이 한정되는 것은 아니다.
- [0098] **실시예**
- [0100] **재료 및 방법**
- [0101] **윤리 승인 및 사전 동의**
- [0102] 본 연구는 서울대학교 병원 기관 심의회의 승인 (No: 1801-015-912)을 받았다. 모든 피험자 및 보호자로부터 사

전 동의를 얻었다. 모든 실험을 관련 지침 및 규정에 따라 수행하였다.

[0104] **환자 선택 및 전체 연구 흐름**

[0105] 바이오마커 발견 및 검증을 위해, 2016년 10월부터 2017년 8월까지 방광 종양의 경요도 절제술 (TURB)을 받은 환자로부터 전향적으로 수집된 소변 샘플을 사용하였다. 소변 단백질 및 엑소좀 분석을 위한 프로토콜을 처음에 설정하는데 3명의 환자의 소변 샘플을 사용하였다. 동일한 기간 동안 수술 전에 공여자 신장절제술을 받은 신장 공여자 중에서 건강한 대조군을 선택하였다.

[0106] 발견 단계에서 12명의 피험자 (방광암 6명 및 대조군 6명), 검증 단계에서 24명의 피험자 (방광암 18명 및 대조군 6명)의 소변 표본을 사용하였다. 타당성 검증 단계에서, 효소결합면역흡착측정법 (ELISA)을 통해, 전향적, 바이오 표본 연계 코호트, SUPER-UC-TURB (Seoul National University Prospectively Enrolled Registry for Urothelial Cancer)로부터 120명의 환자의 소변 샘플을 사용하였다. 이러한 타당성 검증 코호트에서, 비뇨기과 병리학자에 의해 판정된 25명의 환자는 양성(benign) 질환이고, 95명의 환자는 요로상피세포 암종이었다. 수술 1-2일 전 모든 소변 샘플을 전향적으로 수집하여 (각각 20 ml) -195°C의 액체 질소에서 보관하였다. 연구 개요를 도 1에 나타내었다.

[0108] **소변 및 소변 엑소좀 준비**

[0109] 소변 단백질을 추출을 위해, 소변을 10분 동안 원심분리하여 세포 파편을 제거하였다. 시작 부피는 2.5 ml이었다. 4°C, 3000 × g, 15분 동안 원심분리한 후, 소변 상층액을 Amicon Ultra 원심 필터 장치 (3 kDa MWC0, Millipore)를 사용하여 14,000 × g에서 ~100 μl의 부피로 농축하였다. Bradford 분석 (Bio-Rad protein assay kit, Bio-Rad)을 사용하여 단백질을 농도를 측정하였다. 발견 단계에서 무표지 정량화를 위해, 5배 부피의 얼음처럼 차가운 아세톤을 첨가하여 50 μg의 소변 단백질을 침전시켰다. 검증 단계에서 데이터 의존적 수집 (data-independent acquisition, DIA)을 위해, 개별 샘플 당 100 μg의 소변 단백질을 분해하였다. 침전된 단백질을 SDS 함유 변성 완충액 (4% SDS, 1 mM TCEP 및 0.1 M Tris-HCl pH 7.4)에 용해시켰다. 95°C에서 가열한 후, 필터-보조 샘플 전처리(filter-aided sample preparation, FASP) 방법을 사용하여 변성된 단백질을 분해하였다. 37°C에서 트립신 (효소 대 단백질 비율 1:100)으로 밤새 단백질 분해를 수행하였다.

[0110] 소변 엑소좀 분석을 위해, 17,000 × g에서 15분 동안 원심분리하여 이물질을 제거한 후, 2.5 ml의 소변 샘플에서 total Exosome Isolation Reagent (Invitrogen)를 사용하여 엑소좀을 분리하였다. 제조사의 지침에 따라 발견 및 검증 단계의 모든 샘플을 준비하였다. 분리된 엑소좀은 SDS 함유 완충액을 사용하여 엑소좀 전체 단백질을 추출하기 위해 사용되었다. 초음파 처리 및 95°C에서 가열하여 변성한 후, BCA 환원제 호환 키트 (Thermo)를 사용하여 단백질 농도를 측정하였다. 발견 및 검증 단계에서, -20°C에서 밤새 5배 부피의 차가운 아세톤으로 25 μg의 엑소좀을 침전시켰다.

[0111] SDS-함유 변성 완충액에 침전된 단백질을 용해시킨 후, 분해를 수행하였다. 앞서 설명한 바와 같이 homemade C18-StagetTips를 사용하여 분해된 모든 샘플을 탈염하였다. 탈염된 펩티드 샘플을 진공 원심분리기를 사용하여 건조시키고, LC-MS/MS 분석까지 -80°C에서 보관하였다.

[0113] **액체 크로마토그래피-탠덤 질량 분석법 (LC-MS/MS)**

[0114] 준비된 소변 펩티드를 사용하여 액체 크로마토그래피-탠덤 질량 분석을 수행하였다. 데이터 의존적 수집(data-dependent acquisition, DDA) 및 DIA 방법을 포함한 모든 LC-MS/MS 분석은 Q-Exactive Plus 질량 분광계 (Thermo)에 결합된 Ultimate 3000 UHPLC 시스템 (Dionex, Sunnyvale, CA, USA)을 사용하여 수행되었다. 300 nl/분, 7 내지 32% 아세토니트릴, 120분 구배로 트랩 컬럼 (300 μm × 5 mm) 및 분석 컬럼 (75 μm × 50 cm)의 2개의 컬럼 시스템에서 펩티드 샘플을 분리하였다. 컬럼 히터를 사용하여 컬럼 온도를 60°C로 유지하였다. 나노-전기분무를 통해 컬럼 용리액을 Q-Exactive Plus로 전달하였다. 무표지 정량화를 위한 DDA 방법에서, 70,000의 해상도, m/z 200에서 측량 스캔 (350 내지 1650 m/z)을 획득하였다. 1.2 m/z의 분리창으로 전구체 이온을 선택하기 위해 top-20 방법을 사용하였다. m/z 200에서 17,500의 해상도를 가진 30의 HCD-정규화된 충돌 에너지에서 MS/MS 스펙트럼을 획득하였다. 전체 및 MS/MS 스캔에 대한 최대 이온 주입 시간은 각각 20 및 100 ms였다. HRM DIA 방법은 400 내지 1,220 m/z 까지 35,000 해상도에서 측량 스캔으로 이루어진다(AGC 표적 3 × 10<sup>6</sup> 또는 60 ms 주입 시간). 그 다음, 3e6의 자동 이득 조절 표적 및 자동 주입 시간으로 35,000의 해상도에서 19개의 DIA 창을 얻었다.

[0116] **무표지 정량화를 위한 데이터 처리**

- [0117] 모든 MS raw 파일을 MaxQuant (버전 1.5.3.1)에서 처리하였다. MS/MS 스펙트럼은 Andromeda 검색 엔진을 사용하여 Human Uniprot 단백질 서열 데이터베이스 (2014년 12월, 88,657개 항목)에 대해 검색되었다. 총 단백질 수준 분석을 위해 6 ppm 전구체 이온 허용 오차를 사용하여 1차 검색을 수행하였다. MS/MS 이온 허용 오차를 20 ppm으로 설정하였다. 시스테인 카르바미도-메틸화(Cysteine carbamido-methylation)는 고정된 변형으로 설정되었다. 단백질의 N-아세틸화 및 메티오닌의 산화는 가변 변형으로 설정되었다. 효소 특이성은 완전한 트립신 분해로 설정되었다. 최소 길이가 6개의 아미노산이고 최대 2개의 누락된 절단을 갖는 펩티드가 고려되었다. false discovery rate(FDR)는 펩티드, 단백질 및 변형 수준에서 1%로 설정되었다. 샘플 간 정량화 이벤트 수를 최대화하기 위해 MaxQuant 플랫폼에서 'Match between Runs' 옵션을 활성화하였다. 무표지 정량화를 위해, 강도 기반 절대 정량화 (iBAQ) 알고리즘이 MaxQuant 플랫폼의 일부로 사용되었다. 간단히 말해서, MaxQuantdp 의해 계산된 iBAQ 값은 raw 강도를 이론적 펩티드 수로 나눈 값이다. 따라서, iBAQ 값은 단백질 몰량에 비례하였다.
- [0119] **데이터 비의존적 분석을 위한 데이터 처리**
- [0120] 스펙트럼 라이브러리를 생성하기 위해, 소변 샘플로 12개의 DDA 측정을 수행하였다. DDA 스펙트럼은 MaxQuant를 사용하여 Uniprot Human 데이터베이스 (2014년 12월, 88,657개 항목) 및 iRT 표준 펩티드 서열에 대해 검색되었다. Spectronaut 10의 스펙트럼 라이브러리 생성 기능을 사용하여 스펙트럼 라이브러리가 생성되었다. 개별 샘플의 DIA 데이터는 Spectronaut 10 (Biognosys, Schlieren, Switzerland)으로 분석되었다. 먼저, Spectronaut와 함께 제공되는 GTRMS 변환기 틀을 사용하여 DIA raw 파일을 htrm 형식으로 변환하였다. FDR은 mProphet 접근법으로 추정되었고, 펩티드 전구체 및 단백질 수준에서 1%로 설정되었다. 소프트웨어를 사용하여 단백질을 추론하고 정량화 정보를 p 값 < 0.01 기준을 사용해 단백질 수준에서 획득하여 후속 분석에 사용하였다.
- [0122] **단백체에 대한 통계 및 생물정보 분석**
- [0123] 발견, 검증 및 타당성 검증 단계에서 단백질 분석에 대한 개략적인 흐름을 도 1에 나타내었다. 두 단계는 추출된 엑소솜 및 소변 단백질을 사용하여 LC-MS/MS를 통해 수행되었다. 무표지 정량화를 위해 60% 이상의 환자에서 유의하게 차별적으로 발현된 단백질 (DEPs) 및 DIA 분석을 위해 70% 이상의 환자에서 유의하게 차별적으로 발현된 단백질을 (DEPs)을 선택하였다. 페어와이즈(pairwise) 비교를 위해, 순열 기반 FDR < 5%로 2개 샘플의 Student's t-test를 수행하였다.
- [0124] 엑소솜 분석을 위해, Exocarta 데이터베이스 (<http://www.exocarta.org>)에서 알려진 엑소솜 단백질을 선택하고 농축 분석을 수행하였다. 무표지 정량화 및 농축 분석이 있는 DIA DEPs를 결합하여, 일반적인 생물학적 프로세스로 일반적인 상향 및 하향 조절된 단백질을 선택하였다. 비클러스터형 열지도 분석을 사용하여 발현 패턴을 매칭한 후, 엑소솜 및 소변 단백질 중 방광암 바이오마커 후보를 선택하였다.
- [0126] **바이오마커 검증을 위한 소변 ELISA**
- [0127] 후보 바이오마커는 암성(cancerous, 혹은 악성(malignant)) 및 양성(benign) 소변 데이터 간의 유의한 차이를 반영하는 5개의 점수 시스템 합계로 선택되었다. 첫 번째 및 두 번째 점수는 암성 및 양성 환자 소변 간의 LC-MS/MS 기반 DIA 연구 (테스트 및 반복 테스트)의 fold change를 기반으로 하였다. 세 번째 점수는 방광암 진단을 위한 각 단백질의 수신자 조작 특성 곡선(AUROC)의 곡선 아래 면적을 기반으로 하였다; AUROC > 0.95 (10 점), > 0.9 (8점), > 0.85 (6점), > 0.8 (4점) 및 > 0.75 (2점). 반복 DIA 연구에서 다변수 로지스틱 회귀 모델을 사용하여 네 번째 및 다섯 번째 점수를 할당하였다. 5개의 점수 시스템에서, ELISA 연구를 위해 총 8개의 후보 단백질을 선택하였다. 8개의 후보 단백질은 다음과 같다: alpha-2 macroglobulin (A2M), cofilin-1 (CFL1), apolipoprotein A-I (APOA1), inter-alpha-trypsin inhibitor heavy chain H2 (ITI2), afamin (AFM), fibrinogen beta chain (FGB), cell division cycle 5-like protein (CDC5L), 및 CD5 antigen-like protein (CD5L).
- [0128] 그리고 상기 단백질의 수준은 각각 하기의 상용 ELISA를 사용하여 제조사의 지침에 따라 환자 소변 샘플에서 분석되었다: A2M (Magnetic Luminescence Assays, Catalogue No. LXSAMH, R&D Systems Inc.), CFL1 (Catalogue No. MBS2886911, MyBioSource Inc.), FGB (ProcartaPlex™ Multiplex Immunoassay, Catalogue No. MAN0016941, Thermo Fisher Scientific Inc.), CDC5L (Catalogue No. MBS7227993, MyBioSource Inc.) 및 CD5L (Catalogue No. ELH-CD5L, RayBiotech).
- [0129] 각 단백질을 평가하기 전에 정제된 표준을 사용하여 교정 곡선을 준비하였다. 단백질 발현이 고도로 왜곡된 경우 (+2 이상 또는 -2 미만의 왜곡도), 자연 로그 [ln] 변환을 사용하여 데이터를 정규화하였다. 양성 및 음성 샘플 간 및 T 단계 (양성, Tis, Ta, T1 및 > T2)간 각 단백질 발현 수준을 비교하였다.

[0131] **ELISA 결과의 데이터 분석**

[0132] 연속 변수는 중앙값, ± 표준 편차 (사분위수 범위)로 설명하였고, 범주형 변수는 빈도 (백분율)로 설명하였다. 두 AUROC의 통계적 유의성은 De Long의 비모수 접근법(De Long's non-parametric approach)을 따랐다. 두 샘플의 비교는 Student's t-test를 사용하여 수행되었고, 통계적 유의성은  $p < 0.05$ 로 설정되었으며, 보고된 모든 p 값은 양면(2-sided)이었다. 모든 분석은 Python 3.9.0 및 SciPy 의존 패키지를 기반으로 한 통계 분석을 사용하여 수행하였다. 방광암 검출을 위한 각 단백질의 진단 성능 분석을 위해 각 단백질의 AUROC를 사용하였다. 선택된 단백질의 조합을 사용하여 진단 모델을 개발하기 위하여, 역소거법(backward elimination methods) 여하에 관계없이 로지스틱 회귀(logistic regression) 및 머신 러닝 분류 알고리즘(machine learning classification algorithms)을 사용하였다. 개발된 모델의 AUROC를 소변 세포검사의 AUROC와 비교하였다. 진양성률 (민감도)을 최대화하고 위양성률을 최소화하여 컷 오프(cut-off) 값을 선택하였다. 따라서, Youden 지수를 사용하였다. 최적화된 임계값을 사용하기 위해, 각 모델에 대해 양성 예측 값 (PPV), 음성 예측값 (NPV), 민감도 및 특이성을 계산하였다. 소변 세포검사 및 각 개발된 모델의 조합의 AUROC를 계산하였다.

[0134] **결과**

[0135] **환자 특성**

[0136] 개발 및 검증 단계에서, 암 그룹의 환자는 대조군 그룹의 환자보다 나이가 많았다. 검증 단계에서 세포학적 이상 소견 (비정형 및 악성 세포)은 양성 그룹(benign group)보다 암 그룹(cancer group)에서 유의하게 더 높았다 ( $p$  값  $< 0.01$ ). 타당성 검증 단계에서, 나이는 양성 및 음성 샘플 간에 통계적으로 다르지 않았다. 음성 샘플의 백분율은 남성에서 더 높았고, 세포학적 이상 소견과 양의 상관 관계를 가졌다. 등록된 모든 환자 특성을 표 1에 나타내었다.

[0137] [표 1]

	Discovery <sup>a)</sup>			Verification <sup>b)</sup>			Validation		
	Control	Cancer	p-value	Control	Cancer	p-value	Benign	Cancer	p-value
No. of participants	6	6		6	18		25	95	
Age (yr)	54.5 (45.5-56.0)	74.5 (66.0-80.0)	< 0.01 <sup>c)</sup>	57.0 (49.5-59.3)	73.0 (68.0-80.8)	0.01 <sup>c)</sup>	66.0 (55.0-79.0)	72.0 (64.0-77.0)	0.21 <sup>e)</sup>
Female sex	3 (50.0)	2 (33.3)	1.00 <sup>d)</sup>	3 (50.0)	2 (11.1)	0.08 <sup>d)</sup>	11 (44.0)	17 (17.9)	< 0.01 <sup>d)</sup>
Body mass index (kg/m <sup>2</sup> )	24.8 (22.5-26.7)	22.7 (22.6-24.6)	0.55 <sup>e)</sup>	22.1 (21.1-22.3)	25.0 (21.0-25.7)	0.50 <sup>e)</sup>	22.7 (21.2-25.4)	24.5 (22.0-26.7)	0.11 <sup>e)</sup>
Diabetes	0	1 (16.7)	1.00 <sup>d)</sup>	0	1 (5.6)	1.00 <sup>d)</sup>	7 (28.0)	18 (18.9)	0.32 <sup>d)</sup>
Hypertension	1 (16.7)	3 (50.0)	0.55 <sup>d)</sup>	0	7 (38.9)	0.13 <sup>d)</sup>	9 (36.0)	44 (46.3)	0.56 <sup>d)</sup>
Urinalysis abnormality									
RBC	1 (16.7)	2 (33.3)	1.00 <sup>d)</sup>	0	9 (50.0)	1.00 <sup>d)</sup>	21 (84.0)	72 (75.8)	0.38 <sup>e)</sup>
WBC	1 (16.7)	1 (16.7)	0.55 <sup>d)</sup>	1 (16.7)	5 (27.8)	1.00 <sup>d)</sup>	6 (24.0)	30 (31.6)	0.46 <sup>e)</sup>
Urine cytology			0.14 <sup>e)</sup>			0.01 <sup>e)</sup>			< 0.01 <sup>e)</sup>
Benign cellular change	6 (100)	3 (50.0)	N/A	6 (100)	5 (27.8)	N/A	20 (0.0)	38 (40.0)	N/A
Atypical cell	0	2 (33.3)		0	7 (38.9)		5 (20.0)	29 (30.5)	
Malignant cell	0	1 (16.7)		0	6 (33.3)		0	28 (29.5)	
T category			N/A			N/A			N/A
Benign	N/A	0		N/A	0		25 (100)	0	
Tis	N/A	0		N/A	4 (22.2)		N/A	20 (26.3)	
Ta	N/A	3 (50.0)		N/A	6 (33.3)		N/A	35 (36.8)	
T1	N/A	3 (50.0)		N/A	3 (16.7)		N/A	20 (21.1)	
≥ T2	N/A	0		N/A	5 (27.8)		N/A	20 (21.1)	
Concomitant carcinoma <i>in situ</i>	N/A	1 (16.7)	N/A	N/A	5 (27.8)	N/A	N/A	24 (25.3)	N/A
Grade			N/A			N/A			N/A
Low grade	N/A	3 (50.0)		N/A	0		N/A	20 (21.1)	
High grade	N/A	3 (50.0)		N/A	18 (100)		N/A	75 (78.9)	

Values are presented as median (range) or number (%). Control, kidney donor; N/A, not available; RBC, red blood cell; WBC, white blood cell. <sup>a)</sup>Label-free quantification, <sup>b)</sup>Data-independent acquisition. <sup>c)</sup>Student t test, <sup>d)</sup>Fisher exact test, <sup>e)</sup>Pearson's chi-square test.

[0138]

[0140] **바이오마커 발견 - 무표지 정량화**

[0141]

소변 및 엑소좀 바이오마커 후보를 식별하기 위해, 방광암 소변 샘플 (n=12)의 발견 코호트에서 무표지 정량 단백질 분석을 수행하였다 (도 2a). Nanosight를 사용하여 검증된 침전 기반 총 엑소좀 분리 키트(precipitation-based Total Exosomes Isolation kit)를 사용하여 소변 엑소좀을 분리하였다(도 3). 품질 평가 후, 대조군 그룹에서 두 명의 환자 (공여자 2, 6) 및 방광암 그룹에서 한명의 환자(TURBT 3)로부터 얻은 3개의 샘플은 엑소좀 추출에 실패하였다. 식별된 단백질의 평균 수는 엑소좀의 경우 606.2 ± 138.4이고, 소변 단백질의 경우 1021.7 ± 172.1이었다(도 2b). 페어와이즈 비교 분석(Pairwise comparison analysis)은 엑소좀 및 소변 샘플에서 각각 대조군과 방광암 그룹 간에 132 및 248개의 DEPs(permutation-based FDR < 5%)가 있음을 보여주었다 (도 2c). 요로상피암에 대해 67개의 상향조절된 엑소좀 단백질 및 65개의 하향조절된 엑소좀 단백질 뿐만 아니라 137개의 상향조절된 소변 단백질 및 111개의 하향조절된 소변 단백질을 확인하였다. 엑소좀 및 소변 단백질의 unsupervised hierarchy clustering형 열지도에서 발견 코호트의 암 샘플 및 엑소좀 간의 단백질 발현 차이를

시각화하였다 (도 4a 및 도 4b). 엑소솜 및 소변 단백질체의 주성분 분석 (PCA)은 샘플 그룹의 명확한 분리를 보여주었다 (도 2d).

[0143] **바이오마커 검증 - DIA**

[0144] 발견 코호트에서 DEP를 검증하기 위해 데이터-비의존적 수집 전략(DIA)을 채택하였다. 둘 다 높은 데이터 완전성을 달성할 수 있기 때문이다(도 5a). 먼저 DIA 분석을 위해 스펙트럼 라이브러리를 구축하였다. 약 1209개 단백질의 DIA 라이브러리는 풀화(poolled)된 소변으로부터 계산적으로 병합되었다. 발견 단계에서 확인된 DEP 중 소변 단백질에 대한 190개의 후보 및 엑소솜에 대한 104개의 후보를 스펙트럼 데이터로부터 선택하였다 (도 5b). 개인 샘플 분석을 위해 24명의 참가자로부터 소변 및 엑소솜 단백질의 DIA 측정을 수행하였다 (도 5a). 저품질 정량화로 인해 세명의 환자 (TURBT 5, 14, 15)의 엑소솜 데이터를 제외하였다. 추출된 단백질의 평균 수는 엑소솜의 경우  $796.6 \pm 141.0$ 이고, 소변 단백질의 경우  $1279.4 \pm 159.7$ 이었다 (도 6a). 엑소솜 및 소변 단백질에 대한 PCA 플롯은 도 6b 나타내었다. 엑소솜 및 소변 단백질의 volcano plot과 unsupervised hierarchy clustering형 히트맵으로부터, 검증 코호트에서 암 샘플과 엑소솜 간의 단백질 발현 차이를 시각화하였다 (도 6c, 도 7a 및 도 7b).

[0145] 발견 및 검증 코호트를 비교하면, 44%(109/248 소변 DEP)와 38%(50/132 엑소솜 DEP)는 각각 FDR의 5%의 유의 수준에서 일치하였다(도 5b). 다음으로, 발견 및 검증 단계에서 요로상피암 관련 단백질 변화의 일관성을 평가하였다. 엑소솜 샘플에서 48개의 단백질이 암 상태에 따라 충분히 일관되게 달랐다. 이 단백질 중 요로상피암에서 28개 단백질은 상향조절되고, 20개 단백질은 하향조절되었다. 소변 샘플에서 암 상태에 따라 109개 단백질 중 83개의 단백질(76%)이 충분히 일관되게 달랐다. 그 중 27개는 요로상피암에서 충분히 증가되었고, 대조군에서는 56개가 충분히 증가되었다.

[0146] 또한, 대조군과 요로상피암 간에 단백질 수준의 양적 변화는 발견 및 검증 코호트에서 매우 일관되었다. 단백질의 대조군/요로상피암 fold change는 엑소솜 및 소변의 비교에 대해 각각  $r = 0.945$  및  $r = 0.92$ 에서 Pearson's correlation coefficients와 높은 상관 관계가 있었다(도 5c).

[0148] **후보 바이오마커의 선택**

[0149] 발견 및 검증 단계에서 일부 표본의 분리의 한계로 인해, 바이오마커 후보 선택을 위하여 소변 단백질을 사용하였다. 27개의 상향 조절된 소변 단백질 및 28개의 상향조절된 엑소솜 단백질로부터 9개의 단백질이 암환자의 두 샘플에서 풍부하였다. 본 발명자들은 잠재적인 바이오마커로 상향 조절된 단백질을 선택하였다; 그러나, 4개의 단백질은 사용가능한 ELISA 항체가 없어서 제거되었다. 발견 세트에서, Fibrinogen beta chain(FGB)은 7배 상향 조절된 반면, A2M, CD5L, fibrinogen gamma chain(FGG), complement factor H (CFH) 및 Rho GDP dissociation inhibitor beta (ARHGDIB)는 5배 상향 조절되었다. 검증 세트에서, A2M은 6배 상향 조절되었고, FGB 및 FGG는 5배 상향 조절되었으며, APOA1, 보체 C3(C3), CFH 및 아포지질단백질 C-III은 4배 상향 조절되었다. AUROC의 경우, A2M은 10점인 반면, 아파민(AFM), FGB, FGG, C3, CFH, 단백질 S isoform (PROS1), 아포지질단백질 M, 헤파린 보조인자 2(SERPIND1) 및 플라스미노젠 은 8점이었다. 다변수 회귀 모델링에 의해, 첫 번째 모델에서는 A2M, CFL1, APOA1, CDC5L 및 CD5L이 선택되었고, 두 번째 모델에서는 A2M, CFL1, ITIH2 및 AFM이 선택되었다(표 2 및 표 3). 마지막으로, ELISA 연구를 위해 높은 점수의 8개 단백질(A2M, CFL1, APOA1, ITIH2, AFM, FGB, CDC5L, CD5L)을 선택하였다(표 2). 8개의 선택된 단백질 중 3개 단백질(APOA1, ITIH2, FGB)은 엑소솜 샘플에서도 풍부하였다.

[0150]

[표 2]

Protein name	Gene name	SCORE		Difference (Log2Foldchange) in 1st SET	"-Log(P-value) in 1st set"	Difference (Log2Foldchange) in 2nd set	"-Log(P-value) in 2nd set"	ROC (AUC value)	SCORE		SCORE 3: AUC	SCORE 4: Multiple regression modeling 1		SCORE 5: Multiple regression modeling 2		Sum of five scores	Rank
		1: 1st fold-change Score	2: 2nd fold-change Score						1: 1st fold-change Score	2: 2nd fold-change Score		1: 1st fold-change Score	2: 2nd fold-change Score				
Alpha-2-macroglobulin	<i>A2M</i>	2.85381613	4.50576	10.11450924	5.511215164	0.986	5	6	10	10	10	10	10	10	40	1	
Cofilin-1	<i>CFIL1</i>	3.299330024	3.70566	2.651602382	1.628830782	0.803	4	2	4	10	10	10	10	10	29	2	
Apolipoprotein A-I	<i>APOA1</i>	2.446607348	3.56486	6.081590897	4.30588865	0.898	4	4	6	10	10	10	10	10	24	3	
Inter-alpha-trypsin inhibitor heavy chain H2	<i>ITIH2</i>	5.726811153	4.43631	4.392752992	2.983399958	0.898	4	3	6	6	10	10	10	10	23	4	
Afamin	<i>AFM</i>	3.269501158	2.15676	6.702358989	2.007927237	0.949	2	2	8	8	10	10	10	10	22	5	
Fibrinogen beta chain	<i>FGB</i>	6.8033265	6.65579	6.005915497	4.744389442	0.933	7	5	8	8	10	10	10	10	19	6	
Cell division cycle 5-like protein	<i>CDC5L</i>	2.363678632	1.90223	5.285292126	1.321161714	0.887	2	1	6	6	10	10	10	10	19	7	
CD5 antigen-like	<i>CDSL</i>	4.739899978	4.57045	2.198924371	1.961528501	0.752	5	2	2	2	10	10	10	10	19	8	
Fibrinogen gamma chain	<i>FGG</i>	3.540535136	5.23772	6.217428389	5.041801739	0.947	5	5	8	8	10	10	10	10	18	9	
Complement C3	<i>C3</i>	3.394467628	4.41276	6.966474391	4.339245433	0.937	4	4	8	8	10	10	10	10	17	10	
Complement factor H	<i>CFH</i>	4.97094105	4.53503	5.505663063	3.60581189	0.928	5	4	8	8	10	10	10	10	16	11	
Protein S isoform 1	<i>PROS1</i>	2.696179958	2.5525	5.034412876	2.863956838	0.919	3	3	8	8	10	10	10	10	13	12	
Apolipoprotein C-III	<i>APOC3</i>	2.572817484	3.59382	4.339920935	3.591881942	0.894	4	4	6	6	10	10	10	10	13	13	
Apolipoprotein M	<i>APOM</i>	1.825323949	1.94726	5.617349202	2.163536325	0.914	2	2	8	8	10	10	10	10	12	14	
Heparin cofactor 2	<i>SERPIND1</i>	2.211264022	1.38109	5.408781564	2.473802496	0.944	1	2	8	8	10	10	10	10	12	15	

[0151]

[0152]

[표 3]

Protein name	Gene name	"-Log(P-value) in 1st set"	Difference (Log2Foldchange) in 1st SET	"-Log(P-value) in 2nd set"	Difference (Log2Foldchange) in 2nd set	ROC (AUC value)	SCORE		SCORE 3: AUC	SCORE 4: Multiple regression modeling 1	SCORE 5: Multiple regression modeling 2	Sum of five scores	Rank
							SCORE 1: 1st fold-change Score	SCORE 2: 2nd fold-change Score					
Rho GDP dissociation inhibitor (GDI) beta, isoform CRA_a	<i>ARHGDI B</i>	5.534562122	4.58498	3.838979607	2.578171797	0.831	5	3	4	-	-	11	16
Complement component C9	<i>C9</i>	2.591737981	3.06952	3.682511307	1.914698593	0.87	3	2	6	-	-	11	17
Proenkephalin-A	<i>PENK</i>	2.116403288	2.5838	4.015143833	2.282791747	0.863	3	2	6	-	-	11	18
Plasminogen	<i>PLG</i>	1.970644635	1.02685	5.211532536	1.474719475	0.91	1	1	8	-	-	11	19
Antithrombin-III	<i>SERPINC1</i>	2.439119399	1.50534	5.012623878	2.459498133	0.875	2	2	6	-	-	10	20
Heat shock protein HSP 90-beta	<i>HSP90AB1</i>	2.558430537	2.66912	2.787216262	2.473571038	0.782	3	2	2	-	-	7	21
Serum amyloid P-component;Pentaxin	<i>APCS</i>	2.559768064	3.78168	1.904305582	1.788756785	0.743	4	2	0	-	-	6	22
Myosin light polypeptide 6	<i>MYL6</i>	4.941881583	3.94346	1.987436676	1.533932082	0.743	4	2	0	-	-	5	23
Protein IGLY3-19	<i>IGLY3-19</i>	2.319897563	2.42852	2.894478155	2.074889668	-	2	2	0	-	-	-	-
Ig gamma-4 chain C region	<i>IGHG4</i>	2.191019057	2.66229	2.733874052	1.530675178	-	3	2	0	-	-	-	-
IGL@ protein	<i>IGL@</i>	3.106141739	4.01097	7.10992347	2.389454719	-	4	2	0	-	-	-	-
IGL@ protein	<i>IGL@</i>	3.106141739	4.01097	6.082814679	1.520416717	-	4	2	0	-	-	-	-

[0153]

[0155]

[0156]

**ELISA를 사용한 바이오마커 타당성 검증**

분석 전에, 자연 로그 변환을 통해 AFM, CD5L, APOA1, ITIH2 및 FGB의 발현 수준을 정규화하였다 (도 8a 및 도 8b). CDC5L, ITIH2, AFM, CFL1, APOA1, A2M, FGB 및 CD5L의 발현은 암성 샘플과 양성 샘플 간에 유의하게 달랐다(표 4). A2M, CFL1, APOA1, AFM, FGB 및 CD5L은 양성 종양과 T1 또는 T2 종양 간에 통계적 차이를 보였다. ITIH2만이 양성 종양 및 Ta 종양 간의 단백질 발현 수준에서 통계적 차이를 보였다. (도 9). 8개의 바이오마커의 각각의 AUROC는 0.629-0.759 범위였고, 소변 세포검사의 AUROC는 0.718이었다 (도 10A).

[0157] [표 4]

	Benign	Cancer	p-value
No. of participants	25	95	
<b>Raw data</b>			
A2M	76,897.4 (4,579.0 to 35,307.0)	159,609.5 (1,429.5 to 224,686.5)	0.045
CFL1	16,766.87 (2,242.5 to 14,108.0)	33,026.01 (3,692.0 to 65,795.5)	0.017
APOA1	6,779,594.0 (175,659.0 to 1,158,217.0)	19,927,277.3 (703,284.5 to 18,851,324.0)	0.061
ITIH2	8.07 (3.42 to 3.68)	7.29 (0.33 to 3.45)	0.875
AFM	12,894.0 (716.0 to 13,301.0)	48,942.7 (1,997.5 to 46,017.0)	0.002
FGB	256,907.1 (6,421.0 to 160,181.0)	673,964.6 (26,786.75 to 731,388.5)	0.039
CDC5L	2.45 (1.98 to 2.74)	2.85 (2.08 to 3.45)	0.034
CD5L	272.0 (88.70 to 189.72)	508.5 (130.90 to 511.93)	0.038
<b>After logarithmic transformation</b>			
ln_ITIH2	1.28 (1.22 to 1.30)	0.11 (-1.11 to 1.24)	< 0.001
ln_AFM	7.85 (6.57 to 9.50)	9.26 (7.60 to 10.74)	0.003
ln_CD5L	5.06 (4.49 to 5.25)	5.57 (4.87 to 6.24)	0.023
ln_APOA1	13.22 (12.08 to 13.96)	15.05 (13.46 to 16.75)	0.001
ln_FGB	10.64 (8.77 to 10.20)	11.6 (10.20 to 13.49)	0.049

A2M, alpha-2 macroglobulin; AFM, afamin; APOA1, apolipoprotein A-I; CD5L, CD5 antigen-like protein; CDC5L, cell division cycle 5-like protein; CFL1, cofilin-1; ELISA, enzyme-linked immunosorbent assay; FGB, fibrinogen beta chain; ITIH2, inter-alpha-trypsin inhibitor heavy chain H2.

[0158]

[0159] 다변수 분석을 사용하여 ITIH2, AFM 및 CFL1이 모델 1로 선택되었으며, 모델 2는 8개의 모든 바이오마커를 입력하여 생성되었다.

[0160] 모델 1은 하기 수학적 식 1과 같이 구성되었다.

[0161] [수학적 식 1]

[0162] 
$$P=1/e^{-\{0.4678*[CDC5L]+0.9898*\ln[ITIH2]+4.839 \times 10^{-5}*[CFL1]+(-0.115)\}+1}$$

[0163] 상기 수학적 식 1에서 P는 방광암을 가질 확률이고, Ln은 자연로그 함수이고, [단백질]은 각 단백질의 소변 내 농도를 나타낸다.

[0164] 모델 2는 하기 수학적 식 2과 같이 구성되었다.

[0165] [수학적 식 2]

[0166] 
$$P=1/e^{-\{0.3328*[CDC5L]+0.9589*\ln[ITIH2]+4.12 \times 10^{-5}*[CFL1]+0.0576*\ln[AFM]+0.0295*\ln[CD5L]+0.1693*\ln[APOA1]+(-8.418) \times 10^{-7}*[A2M]+(-0.0580)*\ln[FGB]\}+1}$$

[0167] 상기 수학적 식 2에서 P는 방광암을 가질 확률이고, Ln은 자연로그 함수이고, [단백질]은 각 단백질의 소변 내 농도를 나타낸다.

[0168] 모델 1의 AUROC는 0.845 (95% CI: 0.764-0.906)이었고, 모델 2의 AUROC는 0.842 (95% CI: 0.761-0.905)이었다. 모델 1 (p 값 = 0.036) 및 모델 2 (p 값 = 0.039)는 소변 세포검사의 AUROC (0.718, 95% CI: 0.625-0.800)와 통계적으로 유의한 차이를 보였다 (도 10B).

[0169] 모델 1 및 모델 2의 최적 컷 오프 값은 각각 0.735 및 0.870이었다. 모델 1의 민감도, 특이성, PPV 및 NPV는 각각 0.880, 0.813, 0.485 및 0.949이었고, 모델 2의 민감도, 특이성, PPV 및 NPV는 각각 0.850, 0.747, 0.425 및 0.958이었다. 소변 세포검사 및 모델 1의 예측 값을 조합하면, 암 예측에 대한 AUROC는 0.851 (95% CI: 0.770-0.912)이었고, 소변 세포 검사 및 모델 2의 조합은 0.827 (95% CI: 0.743-0.893)이었다(도 9).

[0171] 본 발명은 방광암 진단을 위한 다중 소변 바이오마커 개발을 목표로 하였다. LC/MS-MS를 사용하여 후보 바이오마커를 식별하기 위해 소변 단백질 및 엑소좀의 결합된 접근법을 사용하여, 발견 및 검증 단계에서 여러 후보 단백질을 선택하였다. 참조 연구와 비교한 후, 타당성 검증을 위해 가장 유망한 바이오마커 8개를 선택하였다. 마침내, 본 발명자는 소변 세포검사에 비해 더 정확한 2개의 진단 모델을 개발하였다. 소변 세포검사 및 개발된 컷 오프 값 적용 모델을 조합하면, 소변 세포검사와 모델 1의 AUROC는 0.851이었고, 소변 세포검사와 모델 2의 AUROC는 0.827이었다.

[0172] 개발 및 검증 단계에서, 암 환자 및 건강한 공여자의 소변을 사용하는 체계적 접근법과 결합된 차세대 단백질체학 기반의 대량 스크리닝을 사용하여, 암 환자의 소변에서 높은 수준으로 발견되는 몇 가지 후보 단백질을 선택하

였다. 타당성 검증 단계에서, 방광암의 임상적 의심을 위해 TURB(요도 경유 방광 종양 절제술, transurethral resection of bladder tumor)를 받은 환자의 소변 샘플을 사용하였다.

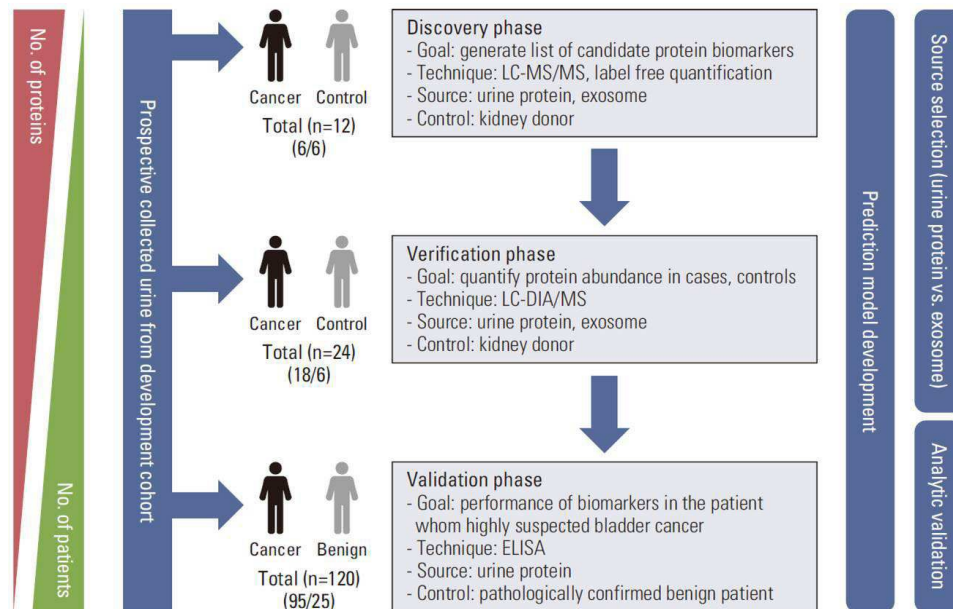
[0173] 여러 바이오마커의 조합은 단일 바이오마커 단독보다 더 높은 정확도를 보여준다. 본 발명에서는, 고처리량 LC-MS/MS 기반 바이오마커 발견 및 검증 단계에서 선택된 8개의 후보 단백질을 사용하여 다중 바이오마커 패널을 개발하였다. 방광암 진단을 위한 각 단백질의 AUROC는 소변 세포검사의 AUROC (0.718)와 유사한, 0.612-0.759이었다. 8개 중 선택된 3개의 바이오마커의 조합을 사용하여 모델 1을 구축하였고, 발견 및 검증 단계에서 선택된 8개의 바이오마커 모두를 사용하여 모델 2를 구축하였다. 모델 1 (0.845) 및 모델 2 (0.842)의 AUROC는 단일 단백질 (0.612-0.759)보다 높았고, 소변 세포검사 (0.718)보다 성능이 우수하였다. 모델 1 및 소변 세포검사의 조합에서, AUROC는 0.851까지 증가하여, 임상 적용에 적합한 성능을 보였다.

[0174] 본 발명에서 임상적으로 유용한 바이오마커 개발에 대한 또 다른 고려 사항은 바이오마커의 출처이다. 소변은 쉽게 접근할 수 있는 체액이며 소변에 함유된 다량의 단백질은 바이오마커 발견의 좋은 후보물질로 여겨진다. 엑소솜은 세포 내 엔도솜 과정을 통해 형성되고 인근 체액으로 분비되는 작은 소포이다. 건강한 집단에 비해 암 환자에서 세포 내 분자를 함유하는 소변 엑소솜 수준이 높아진다; 따라서, 암 진단 또는 예후에 대한 임상적 통찰력을 제공할 수 있다. 엑소솜 바이오마커의 단점은 적용 전에 엑소솜을 적격화하고 선택하기 위한 추가 과정을 필요로 한다는 것이다. 적격화된 소변 전처리 방법을 사용하여, 본 발명자는 상용 엑소솜 추출 키트로 소변 단백질 및 엑소솜 수율 간의 차이가 없음을 발견하였다 (도 2b 및 도 5c). 게다가, 바이오마커로 엑소솜을 사용할 때 품질 검사 실패의 위험이 있다 (도 5c). 다른 암과 달리, 방광 세포는 암 세포와 직접적으로 접촉하여 방광에 오래 머무르고 배설되기 때문에 소변에 많은 암 및 주변 환경 유래 단백질이 포함되어 있다. 이러한 이유로, 본 발명자는 다중 바이오마커 패널 구성을 위해 엑소솜이 아닌 소변 단백질을 사용하였다.

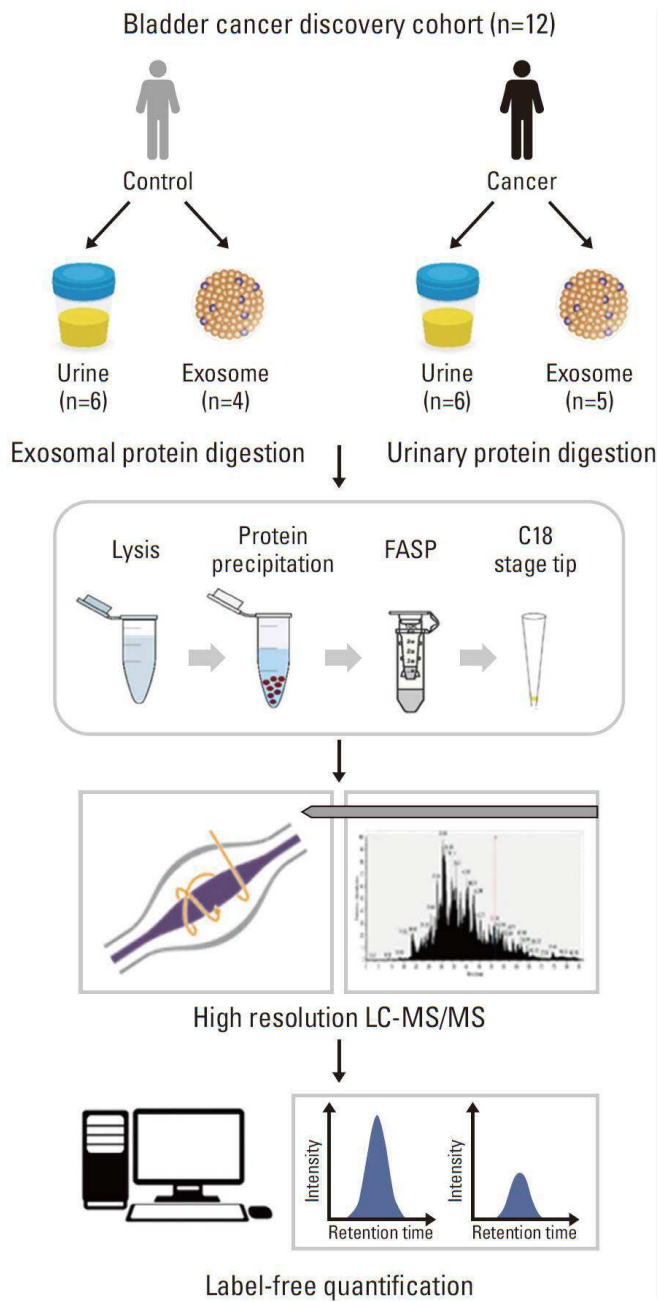
[0176] 본 발명자는 차세대 단백체학을 사용하여 다중 소변 바이오마커 기반 모델을 성공적으로 개발하였다. 다중 소변 바이오마커 기반 패널은 높은 민감도 및 특이성을 보여줄 뿐만 아니라, 소변 세포검사 단독의 예측 능력을 극복하는 AUROC도 보여준다. 소변 세포검사 및 개발된 모델의 조합으로, AUROC가 더욱 증가하였다.

도면

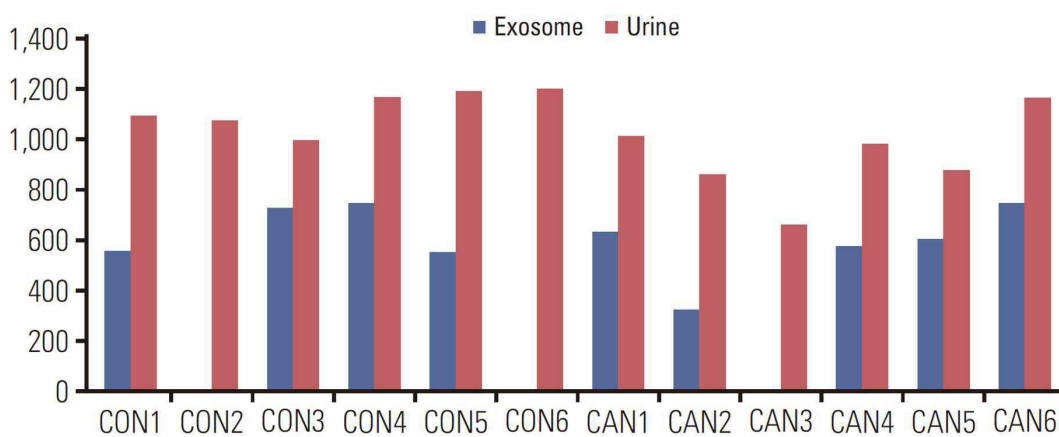
도면1



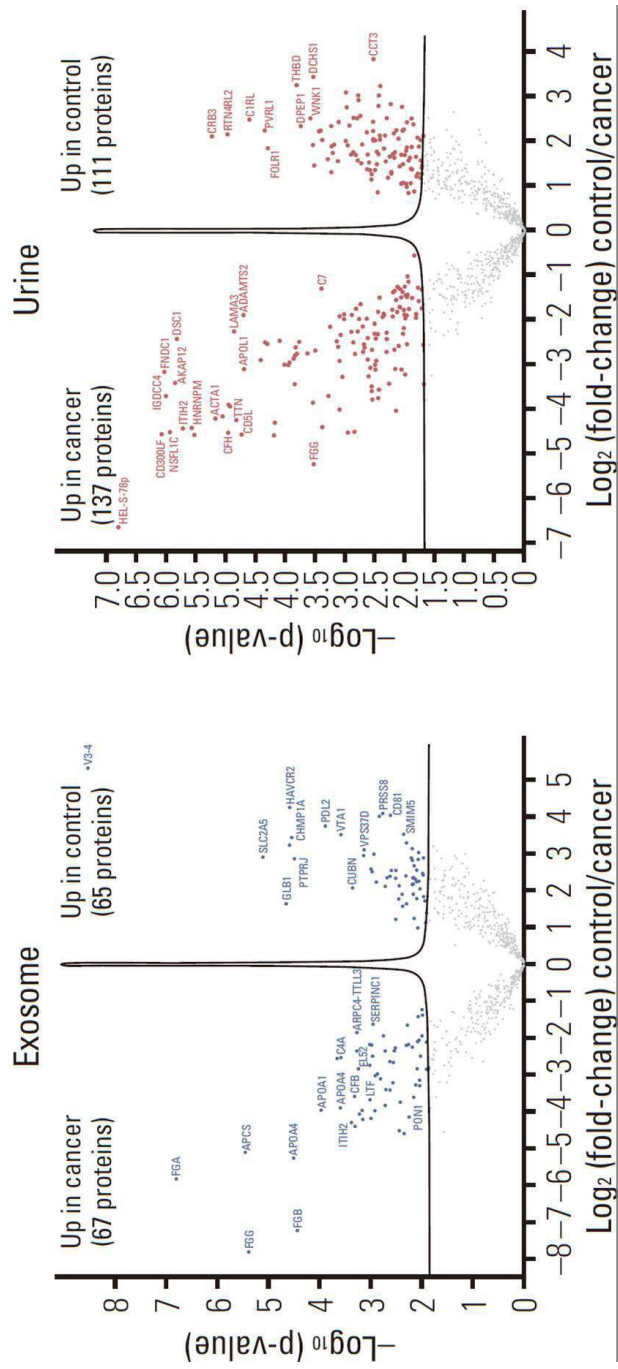
도면2a



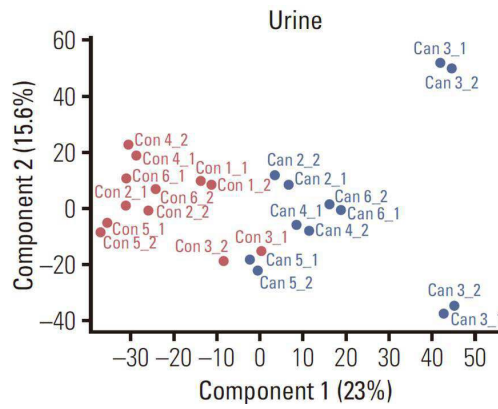
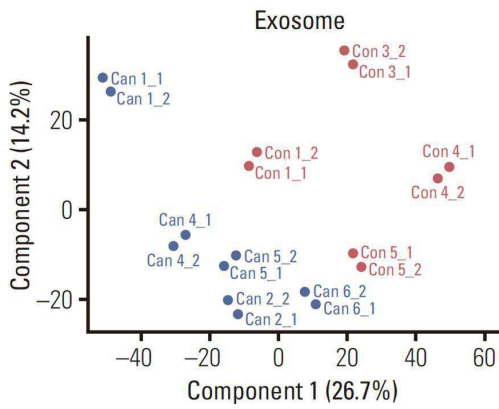
도면2b



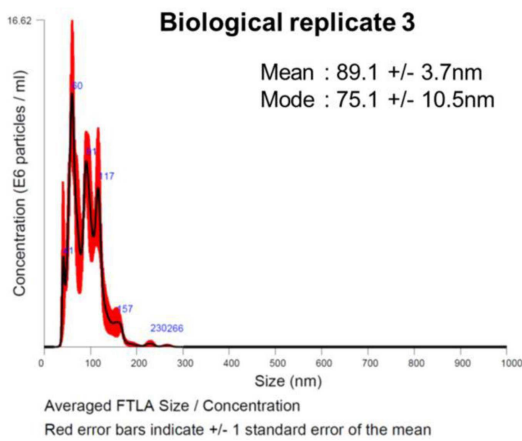
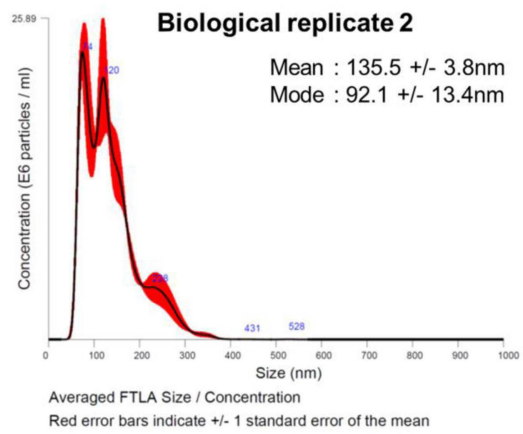
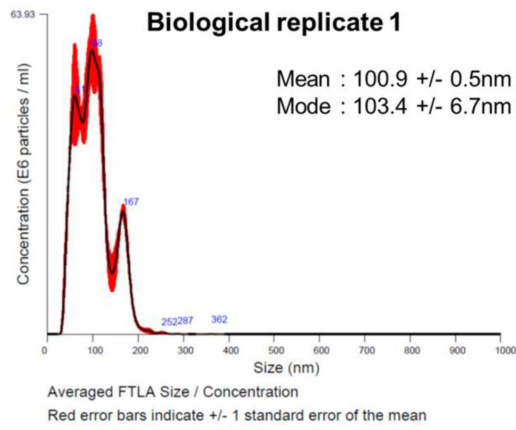
도면2c



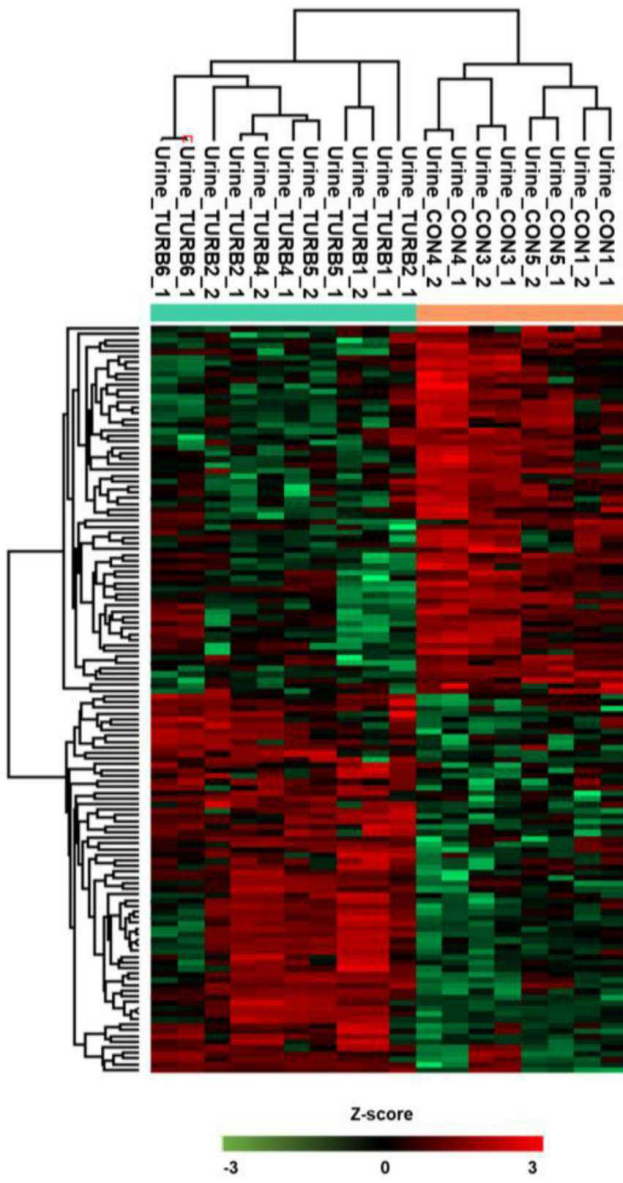
도면2d



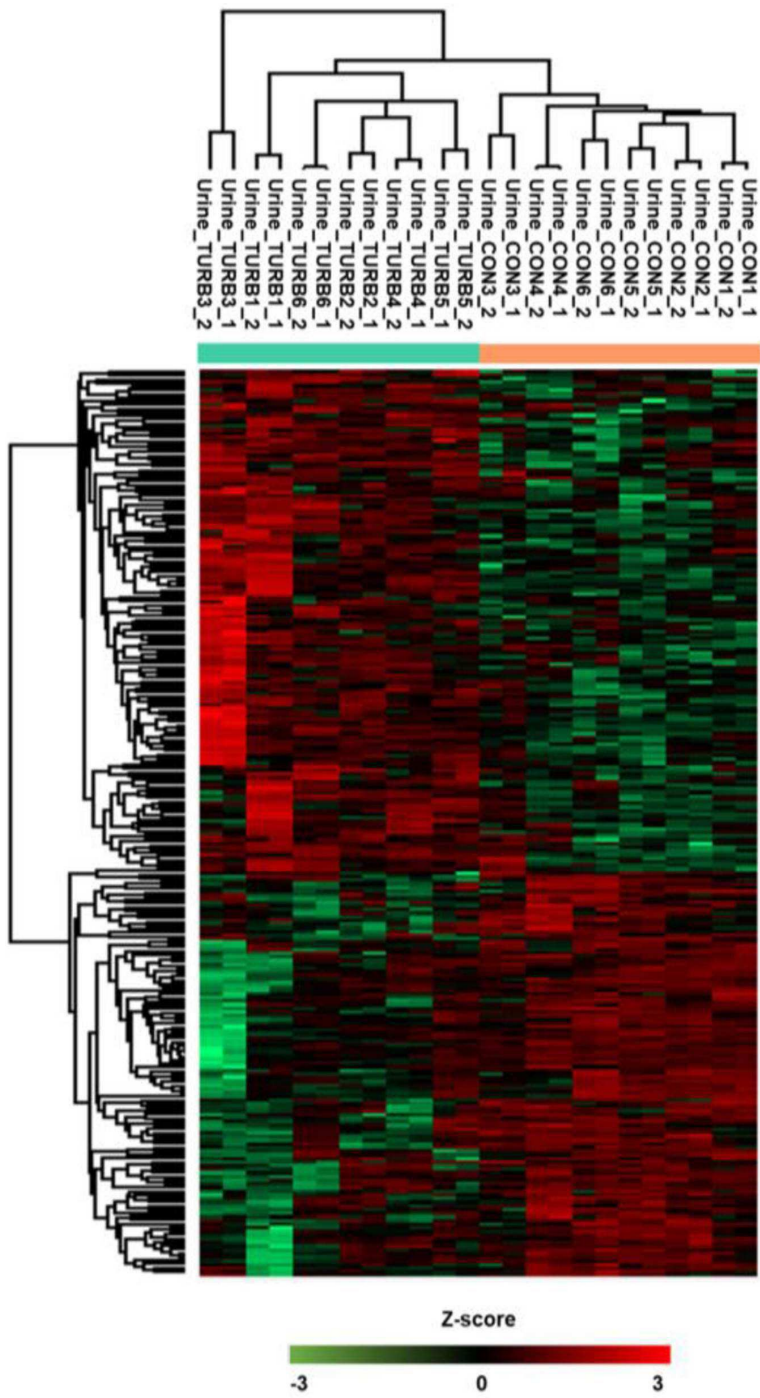
도면3



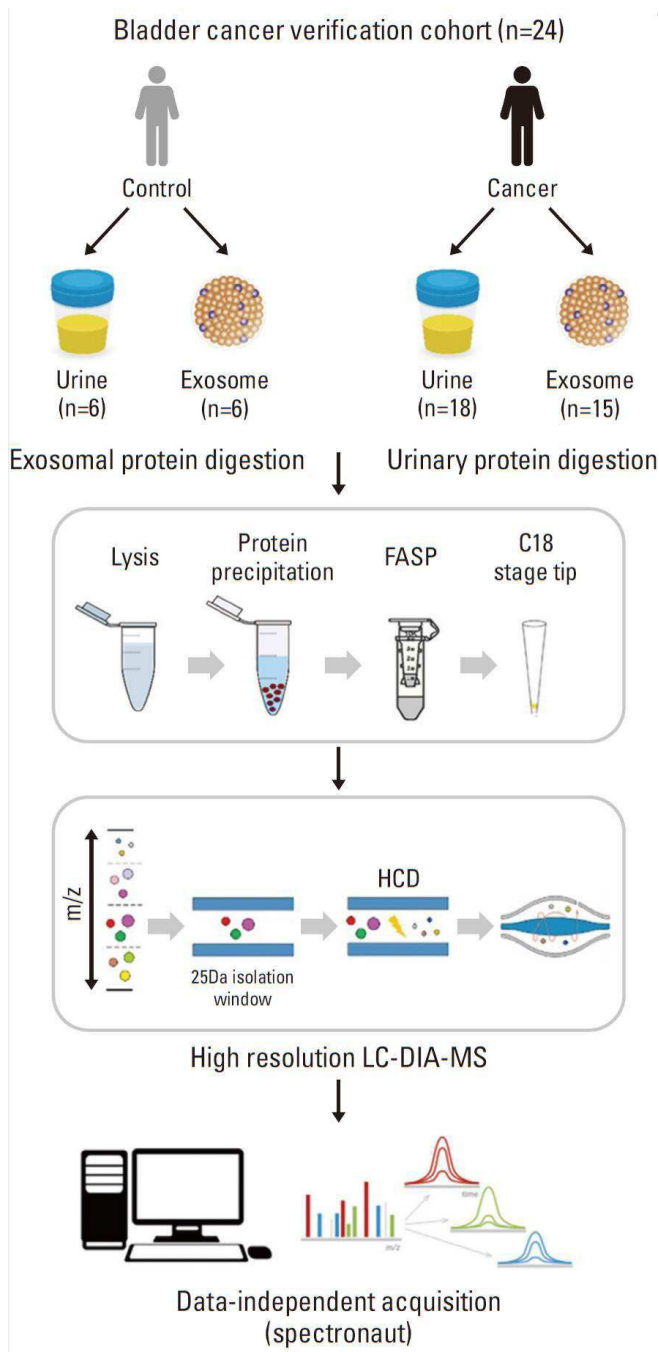
도면4a



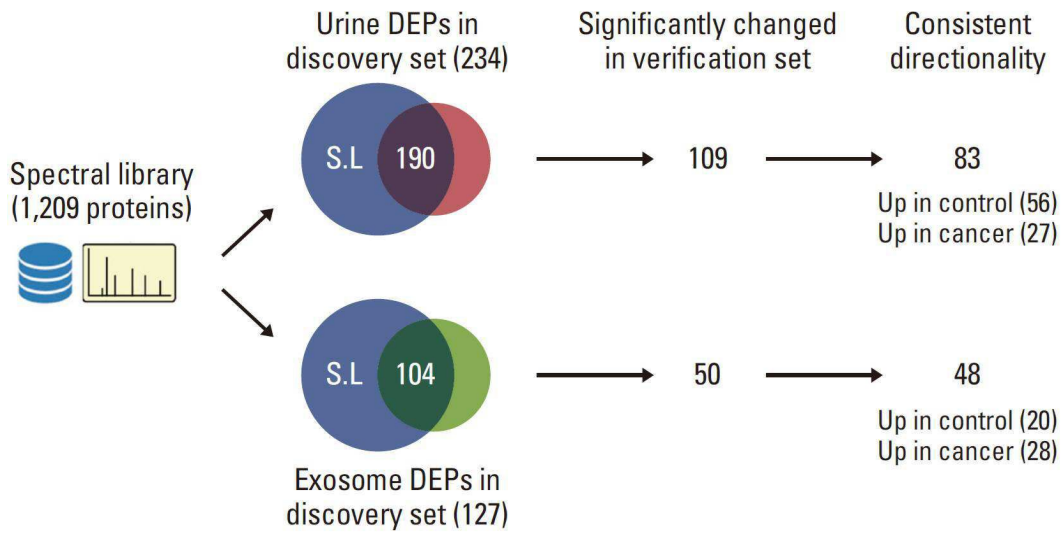
도면4b



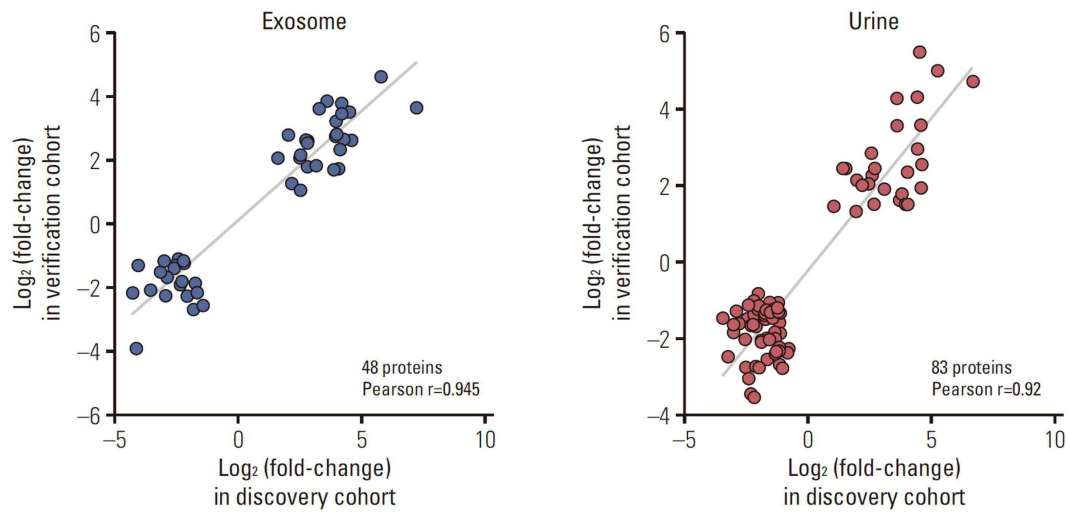
도면5a



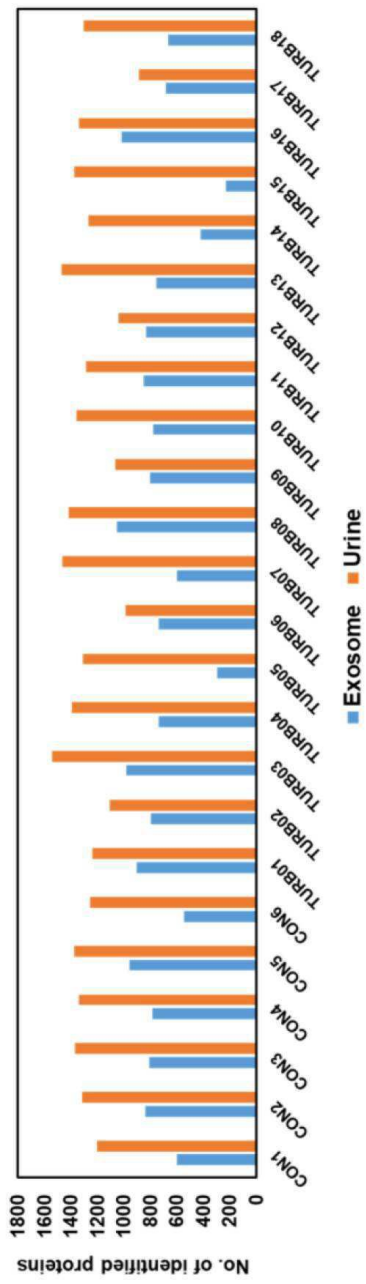
도면5b



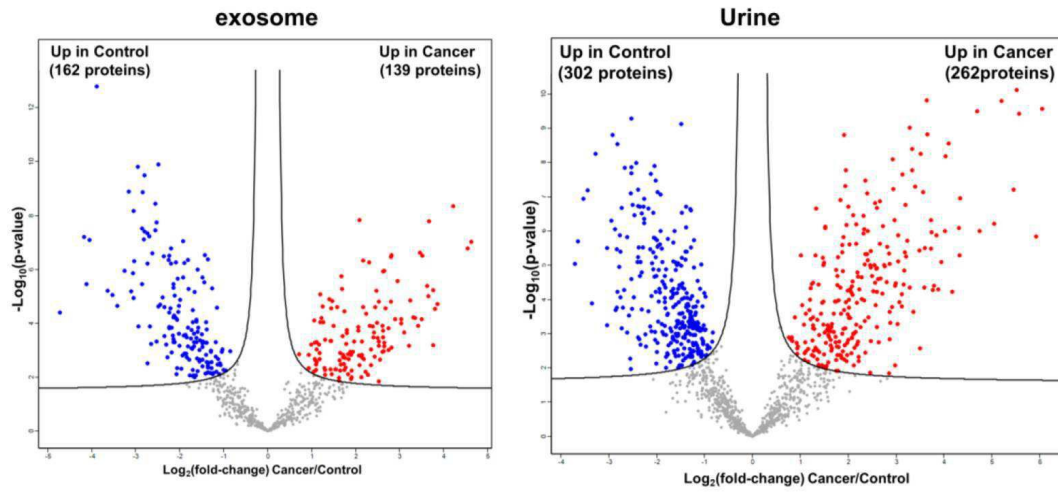
도면5c



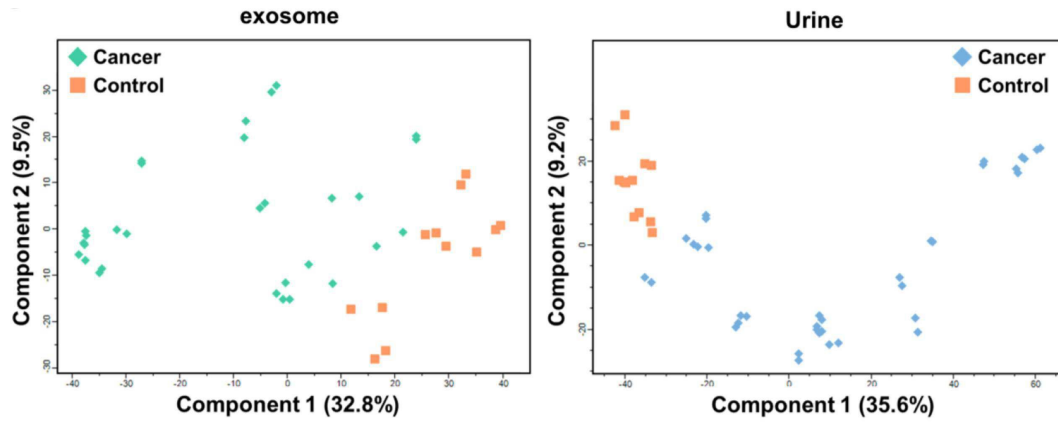
도면6a



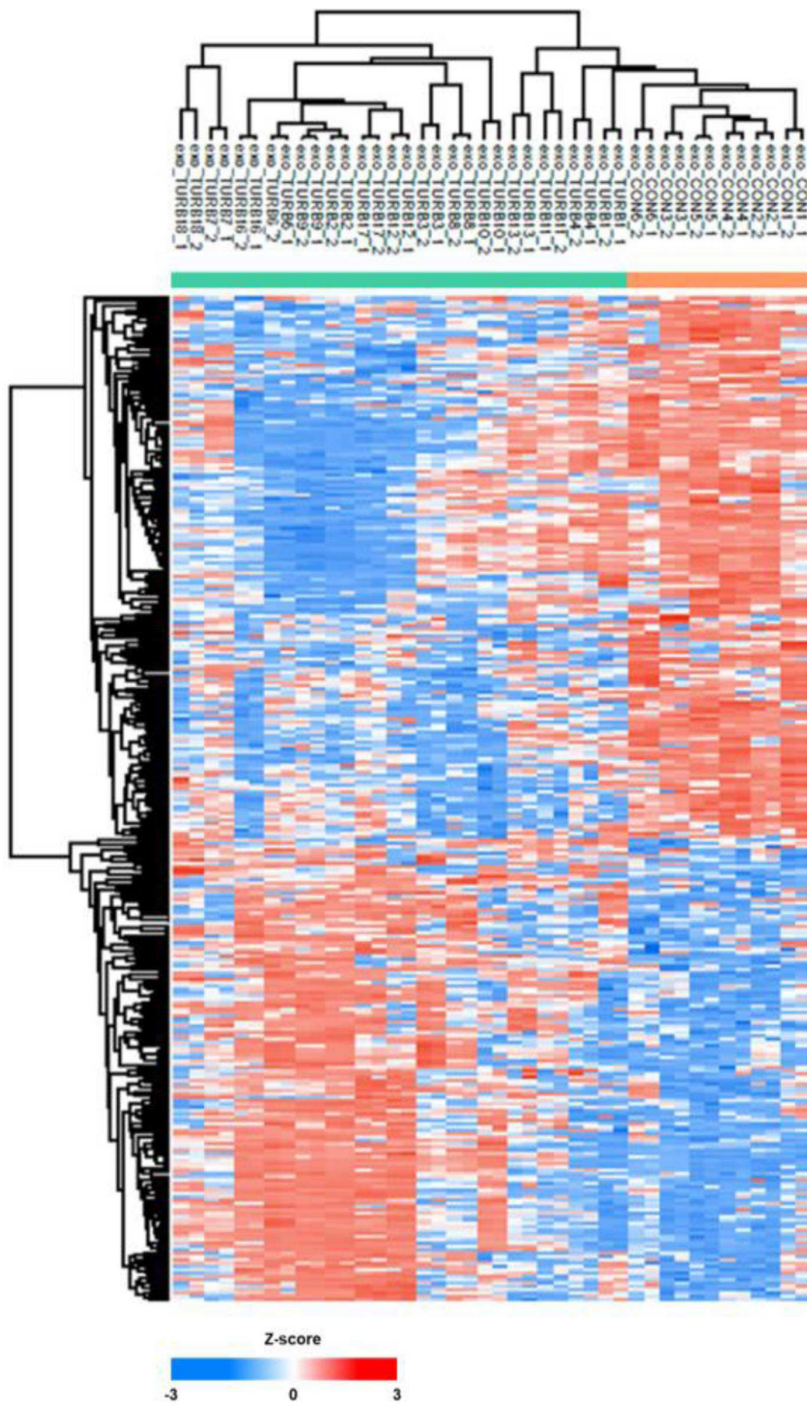
도면6b



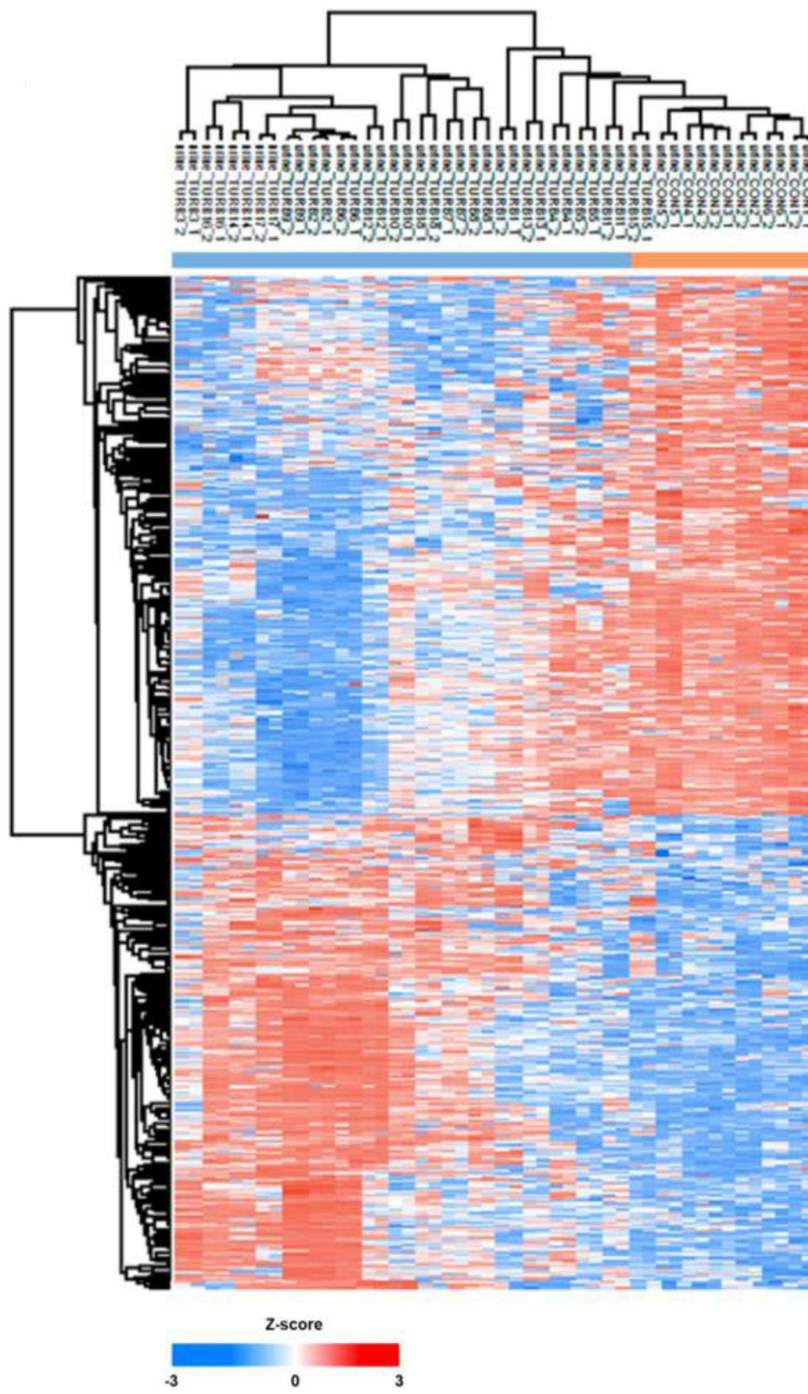
도면6c



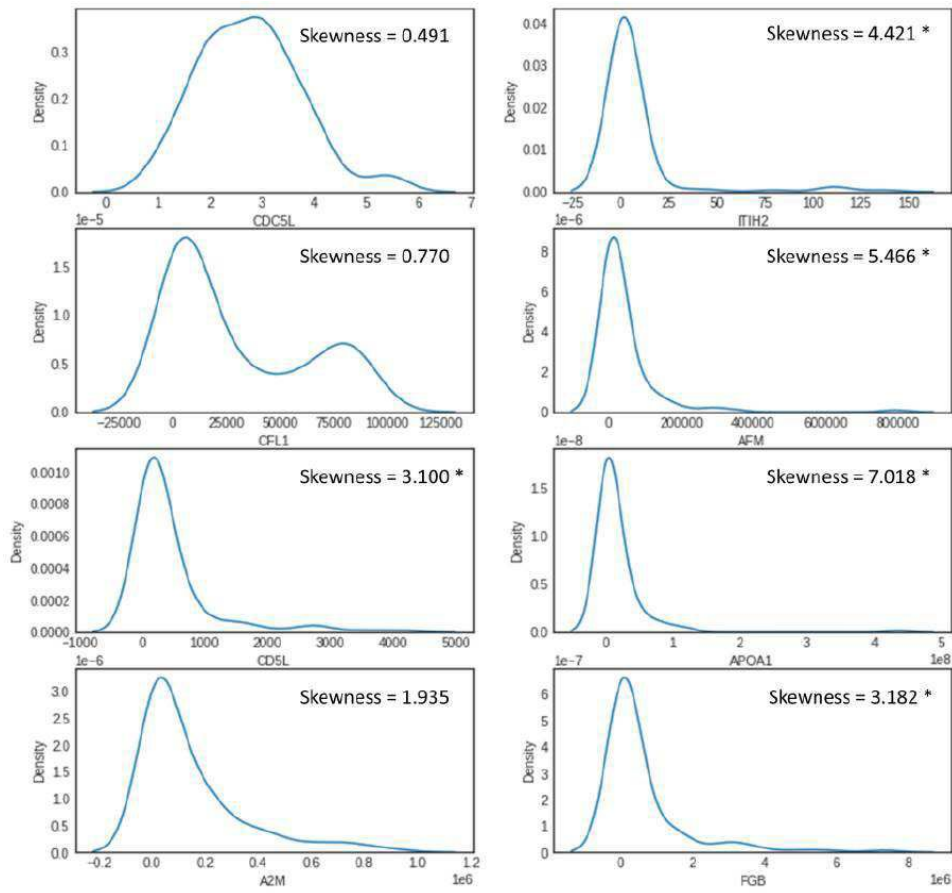
도면7a



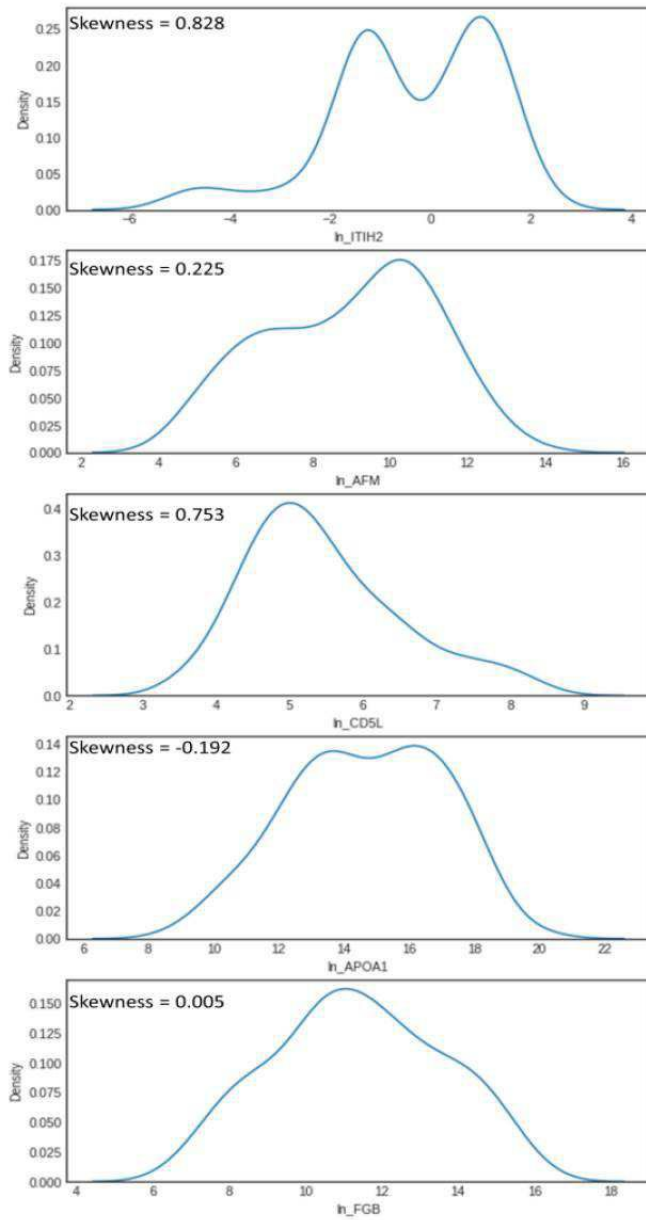
도면7b



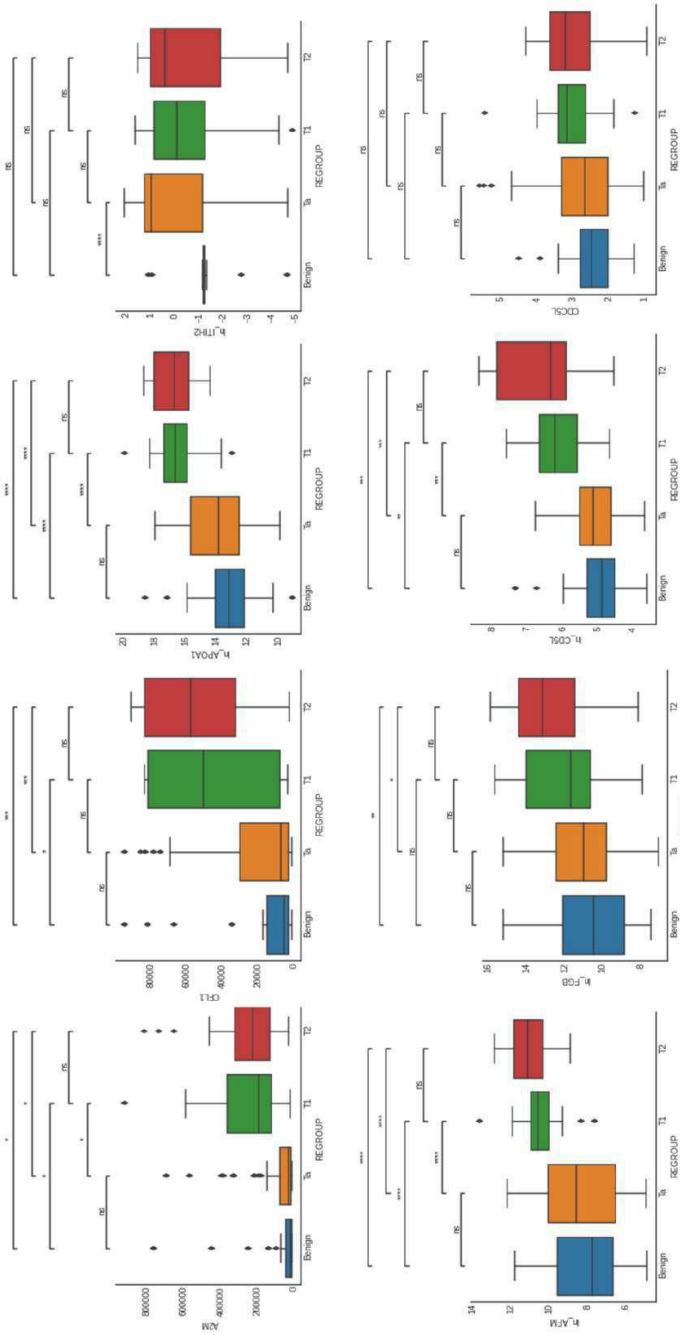
도면8a



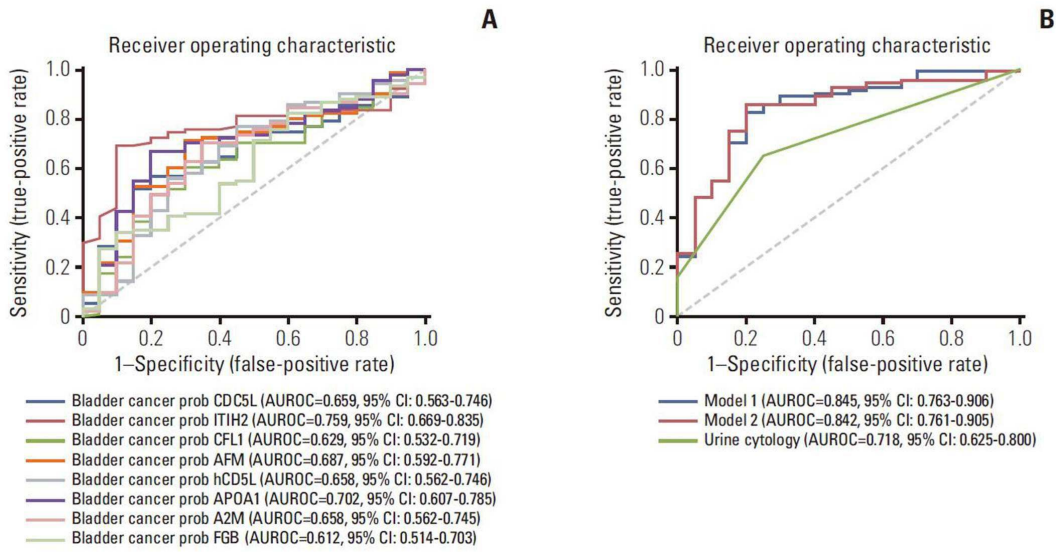
도면 8b



도면9



도면10



도면11

