



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0131656
(43) 공개일자 2022년09월29일

(51) 국제특허분류(Int. Cl.)

G01N 33/68 (2006.01)

(52) CPC특허분류

G01N 33/6893 (2013.01)

G01N 2430/00 (2013.01)

(21) 출원번호 10-2021-0036576

(22) 출원일자 2021년03월22일

심사청구일자 2021년03월22일

(71) 출원인

울산대학교 산학협력단

울산광역시 남구 대학로 93(무거동)

(72) 발명자

홍수종

서울특별시 강남구 압구정로29길 71, 24동 902호
(압구정동, 현대아파트)

김경곤

서울특별시 송파구 올림픽로 435, 113동 3401호(신천동, 파크리오)

(뒷면에 계속)

(74) 대리인

특허법인태백

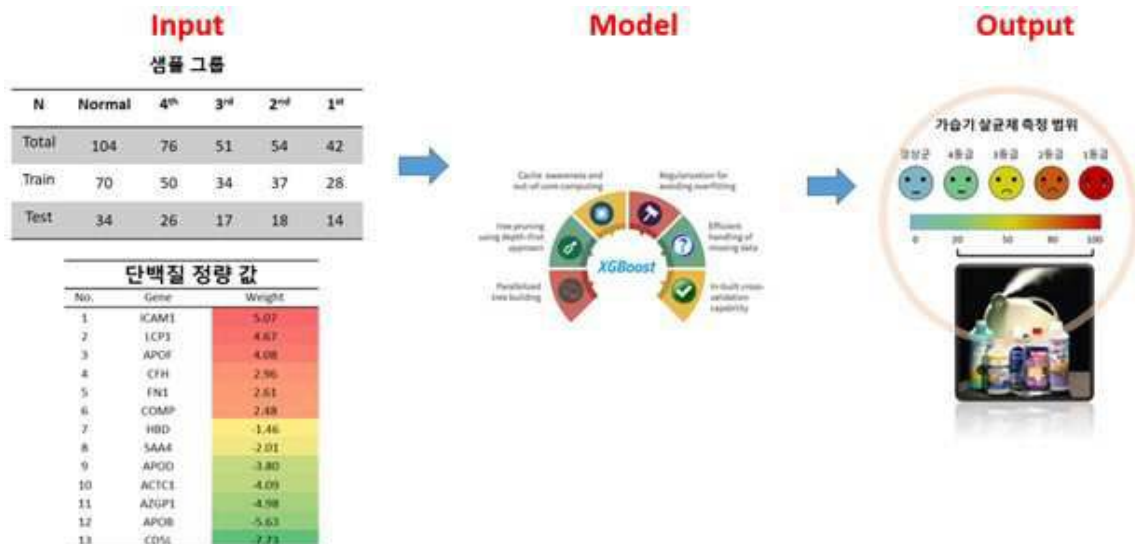
전체 청구항 수 : 총 17 항

(54) 발명의 명칭 가슴기살균제 성분에 대한 노출 여부 확인용 바이오 마커 조성물 및 이의 용도

(57) 요약

본 발명은 가슴기살균제 성분에 대한 노출 여부 확인용 바이오 마커 조성물 및 이의 용도에 관한 것으로, 가슴기살균제에 노출된 환자들에 특이적인 혈장 단백질들을 가슴기살균제 성분에 대한 노출 여부를 확인하는 바이오마커 조성물로 제공하며, 상기 바이오 마커 조성물의 발현 수준을 XGBoost (eXtreme Gradient Boosting) 알고리즘에 입력하여 가슴기살균제 성분에 대한 노출 여부를 확인하는 방법 및 가슴기살균제로 인한 폐손상 여부를 확인하는 방법에 관한 것으로, 상기 바이오마커 조성물을 이용한 알고리즘은 우수한 민감도, 특이도를 나타낼 뿐만 아니라 높은 정확도로 가슴기살균제 비노출군과 가슴기살균제 노출군을 구분하고, 또는 가슴기살균제로 인한 폐손상 환자를 구분하는 방법으로 활용될 수 있다.

대표도 - 도9



(72) 발명자

안희성

서울특별시 영등포구 당산로 95, 104동 202호(당산
동2가, 현대아파트)

엄정훈

서울특별시 강동구 천중로16길 37, 105동 301호(천
호동, 강동리버스타운)

이 발명을 지원한 국가연구개발사업

과제고유번호 20210322

과제번호 NIER-2019-04-02-077

부처명 환경부

과제관리(전문)기관명 국립환경과학원

연구사업명 연구용역사업

연구과제명 가습기살균제 노출 및 건강영향에 관한 생체제표 활용 연구 (II)

기 여 율 1/1

과제수행기관명 울산대학교산학협력단

연구기간 2019.12.19 ~ 2020.10.18

명세서

청구범위

청구항 1

ICAM1(Intercellular adhesion molecule 1), LCP1(Plastin-2), APOB(Apolipoprotein B-100), 및 CD5L(CD5 antigen-like)를 유효성분으로 포함하는 가습기살균제 성분에 대한 노출 여부 확인용 바이오 마커 조성물.

청구항 2

제1항에 있어서, 상기 바이오 마커 조성물은 APOF(Apolipoprotein F), CFH(Complement factor H), FN1(Fibronectin precursor), COMP(Cartilage oligomeric matrix protein), HBD(Hemoglobin subunit delta), SAA4(Serum amyloid A-4 protein), APOD(Apolipoprotein D), ACTC1(Actin, alpha cardiac muscle 1), 및 AZGP1(Zinc-alpha-2-glycoprotein)로 이루어진 군에서 선택된 어느 하나 이상을 포함하는 것을 특징으로 하는 가습기살균제 성분에 대한 노출 여부 확인용 바이오 마커 조성물.

청구항 3

제1항에 있어서, 상기 가습기살균제 성분은 PHMG(polyhexamethylene guanidine), PGH(oligo-[2-(2-ethoxy)-ethoxyethyl]-guanidinium-chlorid)로 이루어진 군에서 선택된 어느 하나 이상인 것을 특징으로 하는 가습기살균제 성분에 대한 노출 여부 확인용 바이오 마커 조성물.

청구항 4

제1항에 있어서, 상기 가습기살균제 성분에 대한 노출 여부는 혈액 시료에서 측정하는 것을 특징으로 하는 바이오 마커 조성물.

청구항 5

가습기살균제 성분에 노출이 의심되는 대상체로부터 분리된 혈액 시료에서 혈장을 분리하는 제1단계;

상기 혈장으로부터 ICAM1(Intercellular adhesion molecule 1), LCP1(Plastin-2), APOB(Apolipoprotein B-100), 및 CD5L(CD5 antigen-like)를 포함하는 단백질들의 발현 수준을 측정하는 제2단계;

상기 단백질들의 발현 수준 및 각 단백질들의 특정 가중치를 알고리즘에 입력하여 스코어를 생성하는 제3단계;

상기 스코어를 미리 결정된 임계값과 비교하는 제4단계; 및

상기 스코어가 미리 결정된 임계값 이상인 경우 대상체가 가습기살균제에 노출된 것으로 판단하는 제5단계를 포함하는 가습기살균제 성분에 대한 노출 여부 확인 방법.

청구항 6

제5항에 있어서, 상기 제2단계에서 PHMG(polyhexamethylene guanidine), PGH(oligo-[2-(2-ethoxy)-ethoxyethyl]-guanidinium-chlorid)로 이루어진 군에서 선택된 어느 하나 이상의 단백질들의 발현 수준을 측정하는 단계를 포함하는 것을 특징으로 하는 가습기살균제 성분에 대한 노출 여부 확인 방법.

청구항 7

제5항 또는 제6항에 있어서, 상기 단백질들의 발현 수준은 웨스턴 블랏(western blotting), ELISA(enzyme linked immunosorbent assay), 방사선면역분석(RIA: radioimmunoassay), 방사 면역 확산법(radial immunodiffusion), 오우크테로니 면역 확산법(Ouchterlony immunodiffusion), 로케트 면역전기영동(rocket immunoelectrophoresis), 면역조직화학염색법(immunohistochemical staining), 면역침전분석법(immunoprecipitation assay), 보체 고정 분석법(complement Fixation Assay), 면역형광법(immunofluorescence), 면역크로마토그래피법(immunochromatography), FACS 분석법(fluorescenceactivated cell sorter analysis), 단백질 칩 분석법(protein chip technology assay) 및 SWATH (Sequential Window Acquisition of all Theoretical Mass Spectra)로 이루어진 군에서 선택된 어느 하나로 측정하는 것을 특징으로

로 하는 가슴기살균제 성분에 대한 노출 여부 확인 방법.

청구항 8

제5항에 있어서, 상기 알고리즘은 XGBoost (eXtreme Gradient Boosting)인 것을 특징으로 하는 가슴기살균제 성분에 대한 노출 여부 확인 방법.

청구항 9

제5항에 또는 제6항에 있어서, 상기 단백질들의 특정 가중치는 ICAM1(Intercellular adhesion molecule 1)의 가중치는 5.07, LCP1(Plastin-2)의 가중치는 4.67, APOF(Apolipoprotein F)의 가중치는 4.08, CFH(Complement factor H)의 가중치는 2.96, FN1(Fibronectin precursor)의 가중치는 2.61, COMP(Cartilage oligomeric matrix protein)의 가중치는 2.48, HBD(Hemoglobin subunit delta)의 가중치는 -1.46, SAA4(Serum amyloid A-4 protein)의 가중치는 -2.01, APOD(Apolipoprotein D)의 가중치는 -3.80, ACTC1(Actin, alpha cardiac muscle 1)의 가중치는 -4.09, AZGP1(Zinc-alpha-2-glycoprotein)의 가중치는 -4.98, APOB(Apolipoprotein B-100)의 가중치는 -5.63, CD5L(CD5 antigen-like)의 가중치는 -7.73인 것을 특징으로 하는 가슴기살균제 성분에 대한 노출 여부 확인 방법.

청구항 10

제5항에 있어서, 상기 임계값은 26.975인 것을 특징으로 하는 가슴기살균제 성분에 대한 노출 여부 확인 방법.

청구항 11

가슴기살균제 성분에 노출이 의심되는 대상체로부터 분리된 혈액 시료에서 혈장을 분리하는 제1단계;

상기 혈장에서부터 ICAM1(Intercellular adhesion molecule 1), LCP1(Plastin-2), APOB(Apolipoprotein B-100), 및 CD5L(CD5 antigen-like)의 발현 수준을 측정하는 제2단계;

상기 단백질들의 발현 수준 및 각 단백질들의 특정 가중치를 알고리즘에 입력하여 스코어를 생성하는 제3단계;

상기 스코어를 미리 결정된 임계값과 비교하는 제4단계; 및

상기 스코어가 미리 결정된 임계값 이상인 경우 대상체가 가슴기살균제에 노출된 것으로 판단하는 제5단계를 포함하는 가슴기살균제로 인한 폐손상 여부 확인 방법.

청구항 12

제11항에 있어서, 상기 제2단계에서 PHMG(polyhexamethylene guanidine), PGH(oligo-[2-(2-ethoxy)-ethoxyethyl]-guanidinium-chlorid)로 이루어진 군에서 선택된 어느 하나 이상의 단백질들의 발현 수준을 측정하는 단계를 포함하는 것을 특징으로 하는 가슴기살균제로 인한 폐손상 여부 확인 방법.

청구항 13

제11항 또는 제12항에 있어서, 상기 단백질들의 발현 수준은 웨스턴 블랏(western blotting), ELISA(enzyme linked immunosorbent assay), 방사선면역분석(RIA: radioimmunoassay), 방사 면역 확산법(radial immunodiffusion), 오우크테로니 면역 확산법(Ouchterlony immunodiffusion), 로케트 면역전기영동(rocket immunoelectrophoresis), 면역조직화학염색법(immunohistochemical staining), 면역침전분석법(immunoprecipitation assay), 보체 고정 분석법(complement Fixation Assay), 면역형광법(immunofluorescence), 면역크로마토그래피법(immunochromatography), FACS 분석법(fluorescenceactivated cell sorter analysis), 단백질 칩 분석법(protein chip technology assay) 및 LC-MS를 활용한 DDA (Data-dependent acquisition) 방법 및 SWATH (Sequential Window Acquisition of all Theoretical Mass Spectra)로 이루어진 군에서 선택된 어느 하나로 측정하는 것을 특징으로 하는 가슴기살균제로 인한 폐손상 여부 확인 방법.

청구항 14

제11항에 있어서, 상기 알고리즘은 XGBoost (eXtreme Gradient Boosting), MLP (multi-layer perceptron), SVM (support vector machine), RF (random forest), 및 복수의 모델의 조합으로 구성된 앙상블 모델, 선행 학습된

(pre-trained) 모델 중 적어도 하나의 모델인 것을 특징으로 하는 가슴기살균제로 인한 폐손상 여부 확인 방법.

청구항 15

제11항 또는 제12항에 있어서, 상기 단백질들의 특정 가중치는 ICAM1(Intercellular adhesion molecule 1)의 가중치는 5.07, LCP1(Plastin-2)의 가중치는 4.67, APOF(Apolipoprotein F)의 가중치는 4.08, CFH(Complement factor H)의 가중치는 2.96, FN1(Fibronectin precursor)의 가중치는 2.61, COMP(Cartilage oligomeric matrix protein)의 가중치는 2.48, HBD(Hemoglobin subunit delta)의 가중치는 -1.46, SAA4(Serum amyloid A-4 protein)의 가중치는 -2.01, APOD(Apolipoprotein D)의 가중치는 -3.80, ACTC1(Actin, alpha cardiac muscle 1)의 가중치는 -4.09, AZGP1(Zinc-alpha-2-glycoprotein)의 가중치는 -4.98, APOB(Apolipoprotein B-100)의 가중치는 -5.63, CD5L(CD5 antigen-like)의 가중치는 -7.73인 것을 특징으로 하는 가슴기살균제로 인한 폐손상 여부 확인 방법.

청구항 16

제11항에 있어서, 상기 임계값은 31.065 내지 35.32인 것을 특징으로 하는 가슴기살균제로 인한 폐손상 여부 확인 방법.

청구항 17

제11항에 있어서, 상기 가슴기살균제로 인한 폐손상은 가슴기살균제 노출에 의한 폐손상이 거의 확실하거나 가능성이 높은 환자인 것을 특징으로 하는 가슴기살균제로 인한 폐손상 여부 확인 방법.

발명의 설명

기술 분야

[0001] 본 발명은 가슴기살균제 성분에 대한 노출 여부 확인용 바이오 마커 조성물 및 이의 용도에 관한 것이다.

배경 기술

[0002] PGH(oligo-[2-(2-ethoxy)-ethoxyethyl]-guanidinium-chlorid), PHMG(polyhexamethylene guanidine) 및 PHMB(polyhexamethylene biguanide)는 구아니딘(guanidine) 계열의 살균제로서 항감염, 항균제로 다양하게 사용되어 왔다. 가슴기살균제로 인해 2021.01.29 기준 가슴기살균제총 4,165명의 인명피해가 발생하였고, 이 중 998명이 사망하였다. 피해자 중에서도 특히 PHMG와 PGH 단독 사용자는 2,132명으로 집계되고 있다. 또한, 살균제 사용자 수를 약 800만 명으로 추산하고 있어 실제 피해 규모는 더 광범위할 것으로 보이며, 피해 판정에 대한 과학적 근거가 절실한 실정이다.

[0003] PGH, PHMG 및 PHMB의 살균제 성분들은 뛰어난 살균력을 갖지만, 여러 가지 문제점을 나타낸다는 사실들이 여러 문헌에서 보고된 바 있다. 상기 살균제 성분들은 생분해성이 아니므로 환경오염을 유발하며, 수생 독성이 심각한 것으로 알려져 있는데 10 μ g/L(0.000001%, 0.054 μ M, 0.01 ppm)에서도 매우 높은 독성을 나타낸다. 또한, 흡입 독성이 있어 스프레이로의 사용이 금지되었으며, 규제 없이 자의적, 무차별적으로 응용되어 제품에 사용될 경우 매우 큰 위험을 초래할 수 있다. 2013년 유럽화학국은 PHMB를 "흡입하면 사망, fatal if inhaled" 물질로 규정하고 있다. 상기 성분들은 인지질 이중막을 파괴하여 살균효과를 나타내는 바, 사람 세포의 인지질을 파괴할 가능성이 있으며, 피부 장벽을 파괴할 우려가 있다. 따라서 물티슈처럼 피부에 자주 사용하면 피부염을 유발할 가능성이 크다. 또한, 살균제 성분에 의해 폐 손상을 받았음에도 살균제 노출 시점으로부터 시간이 길어질수록 인과관계를 직접 규명할 수 있는 근거로 인정되는 폐 영상에서의 흔적들이 소실되고, 폐 기능의 회복 등으로 정확한 피해 평가가 어려운 실정이다.

[0004] 이러한 문제점에도 불구하고, 여전히 상기 살균제 성분들은 소독제, 방부제, 탈취제, 물티슈 등의 생활환경 제품에 자의적으로 광범위하게 사용될 가능성이 있고, 특히 제조업체들은 이들 성분들의 사용기준 용량을 지키지 않거나 제품에 성분명과 사용량을 표기하지 않을 가능성이 있어 소비자들은 불안은 커지고 있다. 따라서, 살균제 성분 노출 여부 및 피해를 확인하기 위한 손쉬운 방법의 개발이 필요하며, 혈액 치료에서 간편하게 살균제 성분의 피해 여부를 확인할 수 있는 바이오마커의 개발이 필요한 실정이다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 대한민국 등록특허 제10-15036380000호 (2015.03.11 등록)

발명의 내용

해결하려는 과제

[0006] 상기와 같은 문제점을 해결하기 위해, 본 발명의 목적은 가슴기살균제 성분에 대한 노출 여부를 확인하는 바이오마커 조성물, 상기 바이오 마커 조성물의 발현 수준을 XGBoost (eXtreme Gradient Boosting) 알고리즘에 입력하여 가슴기살균제 성분에 대한 노출 여부를 확인하는 방법 및 가슴기살균제로 인한 폐손상 여부를 확인하는 방법을 제공하는 데 있다.

과제의 해결 수단

[0007] 본 발명은 ICAM1(Intercellular adhesion molecule 1), LCP1(Plastin-2), APOB(Apolipoprotein B-100), 및 CD5L(CD5 antigen-like)를 유효성분으로 포함하는 가슴기살균제 성분에 대한 노출 여부 확인용 바이오 마커 조성물을 제공한다.

[0008] 또한, 본 발명은 가슴기살균제 성분에 노출이 의심되는 대상체로부터 분리된 혈액 시료에서 혈장을 분리하는 제1단계; 상기 혈장으로부터 ICAM1(Intercellular adhesion molecule 1), LCP1(Plastin-2), APOB(Apolipoprotein B-100), 및 CD5L(CD5 antigen-like)를 포함하는 단백질들의 발현 수준을 측정하는 제2단계; 상기 단백질들의 발현 수준 및 각 단백질들의 특정 가중치를 알고리즘에 입력하여 스코어를 생성하는 제3단계; 상기 스코어를 미리 결정된 임계값과 비교하는 제4단계; 및 상기 스코어가 미리 결정된 임계값 이상인 경우 대상체가 가슴기살균제에 노출된 것으로 판단하는 제5단계를 포함하는 가슴기살균제 성분에 대한 노출 여부 확인 방법을 제공한다.

[0009] 또한, 본 발명은 가슴기살균제 성분에 노출이 의심되는 대상체로부터 분리된 혈액 시료에서 혈장을 분리하는 제1단계; 상기 혈장으로부터 ICAM1(Intercellular adhesion molecule 1), LCP1(Plastin-2), APOB(Apolipoprotein B-100), 및 CD5L(CD5 antigen-like)의 발현 수준을 측정하는 제2단계; 상기 단백질들의 발현 수준 및 각 단백질들의 특정 가중치를 알고리즘에 입력하여 스코어를 생성하는 제3단계; 상기 스코어를 미리 결정된 임계값과 비교하는 제4단계; 및 상기 스코어가 미리 결정된 임계값 이상인 경우 대상체가 가슴기살균제에 노출된 것으로 판단하는 제5단계를 포함하는 가슴기살균제로 인한 폐손상 여부 확인 방법을 제공한다.

발명의 효과

[0010] 본 발명에 따르면, 가슴기살균제에 노출된 환자들에 특이적인 혈장 단백질들을 가슴기살균제 성분에 대한 노출 여부를 확인하는 바이오마커 조성물로 제공하며, 상기 바이오 마커 조성물의 발현 수준을 XGBoost (eXtreme Gradient Boosting) 알고리즘에 입력하여 가슴기살균제 성분에 대한 노출 여부를 확인하는 방법 및 가슴기살균제로 인한 폐손상 여부를 확인하는 방법에 관한 것으로, 상기 바이오마커 조성물을 이용한 알고리즘은 우수한 민감도, 특이도를 나타낼 뿐만 아니라 높은 정확도로 가슴기살균제 비노출군과 가슴기살균제 노출군을 구분하고, 또는 가슴기살균제로 인한 폐손상 환자를 구분하는 방법으로 활용될 수 있다.

도면의 간단한 설명

[0011] 도 1은 혈장 단백질의 양과 피크 범위(peak area)를 분석한 결과이다.

도 2는 가슴기살균제 비노출 정상군 및 가슴기살균제 노출 폐손상군(1단계 내지 4단계)의 혈장 단백질을 정성 분석한 결과이다.

도 3은 가슴기살균제 비노출 정상군 및 가슴기살균제 노출 폐손상군(1단계 내지 4단계)의 혈장 단백질의 정보를 로그 스케일(Log scale)로 분석한 결과이다.

도 4는 가슴기살균제 비노출 정상군 및 가슴기살균제 노출 폐손상군(1단계 내지 4단계)의 혈장 단백질의 정량적 및 정성적 특징을 주성분 분석(Principal component analysis)으로 분석한 결과이다.

도 5는 가슴기살균제 비노출 정상군 및 가슴기살균제 노출 폐손상군(1단계 내지 4단계)의 혈장 단백질의 정량적 및 정성적 특징을 t-SNE(t-Stochastic Neighbor Embedding) 분석으로 분석한 결과이다.

도 6은 가슴기살균제 비노출 정상군과 가슴기살균제 노출 폐손상군(1단계 내지 4단계)을 구분할 수 있는 모델을 구축하기 위해 XGBoost 기법을 통해 특징 추출(Feature selection)을 수행한 결과이다.

도 7은 상기 도 6의 결과를 토대로 상위 29개의 단백질 종류 및 계수를 선정한 결과이다.

도 8은 가슴기살균제 비노출 정상군과 가슴기살균제 노출 폐손상군(1단계 내지 4단계)을 구분할 수 있는 모델을 구축하기 위해 평균제곱근 편차(Root Mean Square Deviation, RMSD)값의 변화 그래프를 작성하고 13개의 단백질을 선별한 결과이다.

도 9는 가슴기살균제 비노출 정상군과 가슴기살균제 노출 폐손상군(1단계 내지 4단계)을 구분할 수 있는 13개의 단백질 패널을 포함하는 XGBoost 분석 모델을 구성한 결과이다.

도 10은 상기 도 9의 13개의 단백질 패널을 포함하는 XGBoost 분석 모델을 이용하여 가슴기살균제 비노출 정상군과 가슴기살균제 노출 폐손상군(1단계 내지 4단계)을 구분한 결과이다.

도 11은 19개의 단백질 패널을 포함하는 XGBoost 분석 모델을 이용하여 가슴기살균제 비노출 정상군과 가슴기살균제 노출 폐손상군(1단계 내지 4단계)을 구분한 결과이다.

도 12는 상기 도 9의 13개의 단백질 패널을 포함하는 XGBoost 분석 모델을 이용하여 가슴기살균제 비노출군과 노출 폐손상군(1, 2, 3 및 4 단계)을 구분할 수 있는지를 검증한 결과이다.

도 13은 상기 도 9의 13개의 단백질 패널을 포함하는 XGBoost 분석 모델을 이용하여 가슴기살균제 비노출군과 가슴기살균제로 인한 폐손상군(1, 2단계)을 구분할 수 있는지 검증한 결과이다.

도 14는 상기 도 9의 13개의 단백질 패널을 포함하는 XGBoost 분석 모델을 이용하여 가슴기살균제 비노출군과 가슴기살균제로 인한 폐손상이 임상적으로 거의 확실한 1단계를 구분할 수 있는지 검증한 결과이다.

발명을 실시하기 위한 구체적인 내용

- [0012] 본 명세서에서 사용되는 용어는 본 발명에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 당 분야에 종사하는 기술자의 의도 또는 판례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 발명의 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서 본 발명에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 발명의 전반에 걸친 내용을 토대로 정의되어야 한다.
- [0013] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0014] 수치 범위는 상기 범위에 정의된 수치를 포함한다. 본 명세서에 걸쳐 주어진 모든 최대의 수치 제한은 낮은 수치 제한이 명확히 쓰여져 있는 것처럼 모든 더 낮은 수치 제한을 포함한다. 본 명세서에 걸쳐 주어진 모든 최소의 수치 제한은 더 높은 수치 제한이 명확히 쓰여져 있는 것처럼 모든 더 높은 수치 제한을 포함한다. 본 명세서에 걸쳐 주어진 모든 수치 제한은 더 좁은 수치 제한이 명확히 쓰여져 있는 것처럼, 더 넓은 수치 범위 내의 더 좋은 모든 수치 범위를 포함할 것이다.
- [0015] 이하, 본 발명을 보다 상세하게 설명한다.
- [0016] 본 발명은 ICAM1(Intercellular adhesion molecule 1), LCP1(Plastin-2), APOB(Apolipoprotein B-100), 및 CD5L(CD5 antigen-like)를 유효성분으로 포함하는 가슴기살균제 성분에 대한 노출 여부 확인용 바이오 마커 조성물을 제공한다.
- [0017] 상기 바이오 마커 조성물은 APOF(Apolipoprotein F), CFH(Complement factor H), FN1(Fibronectin precursor), COMP(Cartilage oligomeric matrix protein), HBD(Hemoglobin subunit delta), SAA4(Serum amyloid A-4 protein), APOD(Apolipoprotein D), ACTC1(Actin, alpha cardiac muscle 1), 및 AZGP1(Zinc-alpha-2-

glycoprotein)로 이루어진 군에서 선택된 어느 하나 이상을 포함할 수 있다.

- [0018] 상기 가슴기살균제 성분은 PHMG(polyhexamethylene guanidine), PGH(oligo-[2-(2-ethoxy)-ethoxyethyl]-guanidinium-chlorid)로 이루어진 군에서 선택된 어느 하나 이상이다.
- [0019] 상기 가슴기살균제 성분에 대한 노출 여부는 혈액 시료에서 측정한다.
- [0020] 또한, 본 발명은 가슴기살균제 성분에 노출이 의심되는 대상체로부터 분리된 혈액 시료에서 혈장을 분리하는 제1단계; 상기 혈장으로부터 ICAM1(Intercellular adhesion molecule 1), LCP1(Plastin-2), APOB(Apolipoprotein B-100), 및 CD5L(CD5 antigen-like)를 포함하는 단백질들의 발현 수준을 측정하는 제2단계; 상기 단백질들의 발현 수준 및 각 단백질들의 특정 가중치를 알고리즘에 입력하여 스코어를 생성하는 제3단계; 상기 스코어를 미리 결정된 임계값과 비교하는 제4단계; 및 상기 스코어가 미리 결정된 임계값 이상인 경우 대상체가 가슴기살균제에 노출된 것으로 판단하는 제5단계를 포함하는 가슴기살균제 성분에 대한 노출 여부 확인 방법을 제공한다.
- [0021] 또한, 본 발명은 가슴기살균제 성분에 노출이 의심되는 대상체로부터 분리된 혈액 시료에서 혈장을 분리하는 제1단계; 상기 혈장으로부터 ICAM1(Intercellular adhesion molecule 1), LCP1(Plastin-2), APOB(Apolipoprotein B-100), 및 CD5L(CD5 antigen-like)의 발현 수준을 측정하는 제2단계; 상기 단백질들의 발현 수준 및 각 단백질들의 특정 가중치를 알고리즘에 입력하여 스코어를 생성하는 제3단계; 상기 스코어를 미리 결정된 임계값과 비교하는 제4단계; 및 상기 스코어가 미리 결정된 임계값 이상인 경우 대상체가 가슴기살균제에 노출된 것으로 판단하는 제5단계를 포함하는 가슴기살균제로 인한 폐손상 여부 확인 방법을 제공한다.
- [0022] 상기 제2단계에서 PHMG(polyhexamethylene guanidine), PGH(oligo-[2-(2-ethoxy)-ethoxyethyl]-guanidinium-chlorid)로 이루어진 군에서 선택된 어느 하나 이상의 단백질들의 발현 수준을 측정하는 단계를 포함한다.
- [0023] 상기 단백질들의 발현 수준은 당 업계에 통상적으로 사용되는 방법으로, 구체적으로 웨스턴 블랏(western blotting), ELISA(enzyme linked immunosorbent assay), 방사선면역분석(RIA: radioimmunoassay), 방사 면역 확산법(radial immunodiffusion), 오우크테로니 면역 확산법(Ouchterlony immunodiffusion), 로케트 면역전기영동(rocket immunoelectrophoresis), 면역조직화학염색법(immunohistochemical staining), 면역침전분석법(immunoprecipitation assay), 보체 고정 분석법(complement Fixation Assay), 면역형광법(immunofluorescence), 면역크로마토그래피법(immunochromatography), FACS 분석법(fluorescenceactivated cell sorter analysis), 단백질 칩 분석법(protein chip technology assay) 및 LC-MS를 활용한 DDA (Data-dependent acquisition) 방법 및 SWATH (Sequential Window Acquisition of all Theoretical Mass Spectra)로 측정할 수 있으며, 보다 바람직하게는 SWATH (Sequential Window Acquisition of all Theoretical Mass Spectra)로 측정할 수 있다.
- [0024] 상기 알고리즘은 XGBoost (eXtreme Gradient Boosting), MLP (multi-layer perceptron), SVM (support vector machine), RF (random forest), 및 복수의 모델의 조합으로 구성된 앙상블 모델, 선행 학습된 (pre-trained) 모델 중 적어도 하나의 모델일 수 있다.
- [0025] 상기 단백질들의 특정 가중치는 ICAM1(Intercellular adhesion molecule 1)의 가중치는 5.07, LCP1(Plastin-2)의 가중치는 4.67, APOF(Apolipoprotein F)의 가중치는 4.08, CFH(Complement factor H)의 가중치는 2.96, FN1(Fibronectin precursor)의 가중치는 2.61, COMP(Cartilage oligomeric matrix protein)의 가중치는 2.48, HBD(Hemoglobin subunit delta)의 가중치는 -1.46, SAA4(Serum amyloid A-4 protein)의 가중치는 -2.01, APOD(Apolipoprotein D)의 가중치는 -3.80, ACTC1(Actin, alpha cardiac muscle 1)의 가중치는 -4.09, AZGP1(Zinc-alpha-2-glycoprotein)의 가중치는 -4.98, APOB(Apolipoprotein B-100)의 가중치는 -5.63, CD5L(CD5 antigen-like)의 가중치는 -7.73이다.
- [0026] 상기 가슴기살균제 성분에 대한 노출 여부 확인 방법에서 상기 임계값은 26.975이며, 혈장 단백질의 수준을 알고리즘에 입력하여 나타난 스코어가 상기 임계값을 초과하면, 상기 대상체는 가슴기살균제 성분에 노출된 것으로 판단할 수 있다.
- [0027] 상기 가슴기살균제로 인한 폐손상 여부 확인 방법에서 상기 임계값은 31.065 내지 35.32이며, 혈장 단백질의 수준을 알고리즘에 입력하여 나타난 스코어가 상기 임계값을 초과하면, 상기 대상체는 가슴기살균제 노출에 의한 폐손상 발생이 거의 확실하거나 가능성이 높은 환자로 판단할 수 있다.
- [0028] 이하, 본 발명의 이해를 돕기 위하여 실시예를 들어 상세하게 설명하기로 한다. 다만 하기의 실시예는 본 발명

의 내용을 예시하는 것일 뿐 본 발명의 범위가 하기 실시예에 한정되는 것은 아니다. 본 발명의 실시예는 당업계에서 평균적인 지식을 가진 자에게 본 발명을 보다 완전하게 설명하기 위해 제공되는 것이다.

[0029] 실시예 1. 혈액 시료 준비

[0030] 가슴기살균제 성분에 대한 노출 여부 판별 및 노출에 의한 폐손상 단계 구분을 위한 바이오 마커를 선별하기 위해, 가슴기살균제에 노출된 환자군 및 가슴기살균제에 노출되지 않은 비노출 정상군을 모집하고, 이들의 혈장 단백질을 분석하였다. 실험에 참여한 대상자는 총 327명으로 각 그룹별 대상군의 수, 평균 나이 및 성별은 하기 표 1과 같다.

표 1

그룹	대상자 수	평균 나이 (범위)	성별 (남 : 여)
정상군	104	7.88 (7-8)	55 : 49
1단계	42	8.83 (5-16)	16 : 26
2단계	54	10.44 (6-17)	30 : 24
3단계	51	9.08 (4-18)	25 : 26
4단계	76	11.14 (4-18)	27 : 49
총	327	-	-

[0032] 가슴기살균제 비노출 정상군의 혈장 샘플은 서울아산병원에서 2015년 ~ 2016년에 수행한 “전국 소아표본 대상 호흡기 알레르기 질환 검사 참고치 조사 연구”에서 제공받았다. 가슴기살균제에 노출된 환자들의 피해 정도는 심사에 따라 1단계 ~ 4단계(1단계: 질병의 원인이 가슴기살균제일 가능성이 거의 확실, 2단계: 질병의 원인이 가슴기살균제일 가능성이 높음, 3단계: 질병의 원인이 가슴기살균제일 가능성이 낮음, 4단계: 질병의 원인이 가슴기살균제일 가능성이 거의 없음)로 나누어지며, 1단계 ~ 4단계의 혈장 샘플은 서울아산병원 가슴기살균제 보건센터에서 제공받았고, 모두 가슴기살균제 성분 중에서 PGH(oligo-[2-(2-ethoxy)-ethoxyethyl]-guanidinium-chlorid) 또는 PHMG(polyhexamethylene guanidine)에 50% 이상 노출된 대상자로 선정되었다.

[0033] 실시예 2. 정량적 혈장 단백질체 분석

[0034] 선별된 혈장 샘플로부터 가슴기살균제 노출 여부 및 노출 단계와 관련이 있는 바이오 마커를 선별하기 위해, 혈장 단백질을 정량적으로 분석하였다. 총 327명의 혈장 샘플로부터 혈장 단백질을 정량적으로 분석하기 위해서 사중극자 비행시간 (Quadrupole Time of Flight) 질량분석기 (Mass spectrometry) 기반의 정량적 혈장 단백질체 분석기법인 SWATH (Sequential Window Acquisition of all Theoretical Mass Spectra)를 이용하였다. 혼합된 시료를 이용하여 DPHL: A pan-human protein mass spectrometry library (참조 논문: <https://pubmed.ncbi.nlm.nih.gov/32795611/>)(스펙트럼 라이브러리)이용하여, 40L의 적은 양의 혈액 시료를 절단된 펩티드로 시료전처리하여 LC/MS SWATH 분석을 통해 혈장 단백질을 정량분석 하였다.

[0035] 237개의 혈장 샘플들을 SWATH 분석에 적용하기 전에 재현성있는 분석을 위해 고성능 액체 크로마토그래피 (High-performance liquid chromatography, HPLC)에 연결한 고농도 혈장 단백질 제거 컬럼 MARS14 컬럼(100 x 4.6 mm; Agilent Technology, Palo Alto, CA, USA)을 이용하여 저농도 단백질을 분리하고 이를 자동화된 액체 분주 시스템 (Bravo Automated Liquid Handling Platform - Agilent사)를 이용하여 펩티드를 제조하는 전처리 과정을 수행하였다. 펩티드 제조 후 NanoDrop으로 정량한 값을 기반으로 서로 다른 327개의 시료를 동일량으로 SWATH-MS 분석하였다. 자세히 살펴보면, 40 L의 혈장을 MARS14 컬럼에 주입하여 상위 14 개의 풍부한 단백질 (albumin, IgA, IgG, IgM, α1-antitrypsin, α1-acid glycoprotein, apolipoprotein A1, apolipoprotein A2, complement C3, transferrin, α2-macroglobulin, transthyretin, haptoglobin and fibrinogen)을 제거하였습니다. 이를 위해 혼합물을 Agilent사의 buffer A (부품번호: 5185-5987)로 4 배 희석하고 Shimadzu HPLC LC20AT 시스템의 MARS14 컬럼에 로드 했습니다. 결합되지 않은 저농도 혈장 단백질의 분획을 자동화된 액체 분주 시스템로 자동화된 filter aided sample preparation (FASP)를 수행하는 데 사용되었습니다. Multiscreen Vacuum Manifold™ (Millipore, Billerica, MA, USA)를 사용하여 샘플 준비 중에 진공을 공급했습니다. 간단히 말해서, 단백질 샘플은 96 well 플레이트 (AcroPrep Advance Filter Plates, 350 μL, Omega 30K MWCO)에 로드되었습니다. 단백질은 100 μL의 5mM TCEP (37° C, 30분)로 환원되고, 100 μL의 50mM IAA (25° C, 30분, 어둠 속에서)로 알킬화되고, 6g의 트립신으로 37도에서 18시간동안 분해됩니다. 이후 포름산을 최종 농도 0.3 %로 첨가하여 샘플의 pH를 2-3으로 낮추어 분해를 중단시켰다. 이어서 펩티드 혼합물을 Sep Pak C-18 카트리지 (Waters, USA)로 탈염하고 콜드 트랩 (CentriVap Cold Traps, LABCONCO)으로 동결 건조하고 사용할 때까지 영

하 80도에서 보관했습니다.

- [0036] 액체 크로마토 그래피-탠덤 질량 분석법 (LC-MS/MS) 분석 전에 건조된 펩티드 샘플을 버퍼 A (0.1 % formic acid in HPLC water)으로 녹이고 총 펩티드 농도를 UV/Vis 분광 광도계 (NanoDrop One, Thermo Fisher Scientific)로 흡광도를 측정했습니다. 파장 280 nm, 샘플 유형 옵션이 "1 Abs = 1 mg / mL"로 설정되었습니다. 그런 다음 iRT-Kit (Biognosys AG, Schlieren, Switzerland)에서 제공하는 iRT 표준을 1/10 부피로 샘플에 추가하여 공급 업체 지침 (참고논문: <https://pubmed.ncbi.nlm.nih.gov/22577012/>)에 따라 머무름 시간을 보정했습니다. 각 샘플의 총량은 40g이며 40L에 용해됩니다. 주입된 4 L의 샘플은 다음 LC-MS/MS 설정과 함께 SCIEX TripleTOF 5600+ 시스템 질량 분석기를 사용하여 분석되었습니다. LC 분리를 위해 nanoLC 425 (Eksigent, Dublin, CA, USA)를 Eksigent 마이크로 트랩 C18컬럼 (ChromXP C18CL, 5 μ m, 120 A)과 분석 컬럼으로서 Eksigent 컬럼 (C18-CL, 0.3 x 150 mm, 입자 크기 3 μ m, 기공 크기 120 A)을 함께 컬럼 온도가 40 C로 유지되도록 하였습니다. 샘플을 버퍼 A를 사용하여 10 μ L/min의 유속으로 트랩 컬럼에 로드 했습니다. 10분 후, 펩티드 혼합물은 버퍼 A와 버퍼 B (0.1 % formic acid in HPLC acetonitrile)을 사용하여 57 분에 의해 5 μ L/분으로 유속으로 농도구배를 통한 분리를 수행하였습니다. 버퍼 B를 38분 동안 3-25% 변경하고, 5분 동안 25-32% 변경하고, 2분 동안 32-80% 변경하고, 3분 동안, 80%로 유지하고, 3분 동안 80-3% 변경하고, 8분 동안 3%로 유지하였다. 개별 샘플의 경우 모든 질량 분석 실행은 SCIEX 기술 노트에 따라 100개의 가변 창을 사용하여 모든 이론적 질량 스펙트럼 (SWATH) 모드의 순차 창 수집에서 작동되었습니다. SWATH 매개 변수는 다음과 같이 설정되었습니다: Lower m/z limit 400; m/z 상한선 1250; 윈도우 오버랩 (Da) 1.0; CES는 작은 창에 대해 5, 큰 창에 대해 8, 가장 큰 창에 대해 10으로 설정되었습니다. MS2 스펙트럼은 고감도 모드에서 2.5ms 동안 100 - 1500m/z 범위에서 수집되었으며 총 사이클 시간은 2.8초였습니다. 다른 MS 파라미터는 다음과 같이 설정되었습니다: 이온 소스 가스 1 (GS1) 15 psi; 이온 소스 가스 2 (GS2) 20 psi; 커튼 가스 (CUR) 30 psi; 온도 (TEM) 250 ° C; 이온 스프레이 전압 부동 (ISVF) 5500V로 설정하였습니다.
- [0037] 도 1에 나타난 바와 같이, 단백질의 양과 피크 범위(peak area)가 327개의 각 시료 별 제조된 단백질의 양 (왼쪽의 Y축; 바그래프)과 혈장에서 MARS14컬럼으로 제거된 저 농도 단백질의 UV 피크 넓이 값 (오른쪽 y축, 꺾은 선 그래프)로 표현하였다. 모든 시료에서 40ug이상의 펩티드를 제조하였고, 이는 SWATH LC-MS분석이 가능한 정도의 펩티드 양이 생산되었음을 확인하였다.
- [0038] 펩티드로 제조된 혈장 단백질 시료 327개를 분석함에 있어서 분석 간 Batch별 bias를 최소화하기 위해 5개의 그룹들을 무작위로 혼합하여 순차적으로 SWATH LC-MS 분석을 14일간에 걸쳐 수행하였다. 327개의 시료에 대해 SWATH LC-MS 분석을 수행하여 327개의 스펙트럼(spectrum) 결과를 획득하였다. 해당 결과 스펙트럼(spectrum)들을 DIA-NN software와 Pan-human proteome library를 이용하여 정성적으로 단백질 분석을 수행하였으며 327개 시료로부터 총 1,296개의 단백질을 정성분석할 수 있었다. 도 2에 나타난 바와 같이, 5개 그룹에 대해 단백질 정성 분석 결과, 비노출 대조군에서는 313개, 1단계 노출군에서는 304개, 2단계, 3단계 및 4단계 노출군에서는 각각 296개, 305개, 306개의 단백질들이 평균적으로 재현성있게 동정되었다.
- [0039] 도 3에 나타난 바와 같이, 5개 그룹 총 327개의 시료별로 정량된 단백질들의 양 정보를 로그 스케일(Log scale)로 분석한 결과 단백질의 발현양의 분포는 중간(median)값을 중심으로 모두 재현성을 보이면서 분석된 것으로 확인되었다.
- [0040] 또한, 5개 그룹의 327개의 혈장 시료들에서 80% 이상 검출되면서 MARS-14 high abundant protein depletion에 관련되지 않은 단백질 총 226개 단백질들의 정성, 정량 정보를 이용하여 비지도 학습(Unsupervised learning) 분석법 중의 하나인 주성분 분석 (Principal component analysis)을 수행한 결과, 도 4와 같이 가슴기살균제 비노출군과 노출 폐손상군(1단계 내지 4단계)은 혈장 단백질체의 정량적, 정성적 특징을 통해 통계적으로 차이가 나는 것을 확인할 수 있었다.
- [0041] 다른 비지도 학습(Unsupervised learning) 분석법 중의 하나인 t-SNE(t-Stochastic Neighbor Embedding)?분석을 수행한 결과, 도 5와 같이 주성분 분석의 결과와 동일하게 가슴기살균제 비노출군은 노출 폐손상군과 구별되고 있음을 확인할 수 있었다. 노출 폐손상군 중에서 4개의 단계군 중 1단계군은 가슴기살균제 비노출군과 가장 큰 distance를 보이고 있어 단백질체의 정량적, 정성적인 특징의 차이가 크다는 것을 확인할 수 있었다. PCA와 t-SNE 분석 결과로부터 5개 그룹의 단백질체 정량, 정성 정보를 이용하면 1차적으로 가슴기살균제 비노출군과 노출군은 구분이 될 수 있음을 확인하였다.
- [0042] 실시예 4. 혈장 단백질 데이터를 이용한 모델 구축 및 검증

[0043] 상기 실시예 3에서 수집된 비노출 대조군 및 노출 폐손상군(1, 2, 3, 및 4 단계)의 단백질의 정량 정보를 이용하여 가슴기살균제 노출 폐손상 단계 또는 노출 여부를 구별할 수 있는 모델을 구축하였다. 이를 위해 의사 결정 나무 구조 (Determination tree structure)를 이용하여 단백질 정량정보를 변수로 하고 특징적인 변수 (Feature)를 선택하는 단계에서 모델 자체에 변수 선택이 포함된 임베디드 방법 (Embedded method)을 적용하였다. 무작위 숲 알고리즘 (Random forest algorithme)의 진화된 형태인 XGBoost (eXtreme Gradient Boosting)이라는 기법은 의사 결정 구조에서 가중치를 경사하강법으로 최적화하여 모델을 구성하는 최신 기법이다. XGBoost 기법은 뛰어난 예측 성능을 가지고 있으며 분류와 회귀영역에서 우수한 예측 성능을 보인다. 또한 분석 수행 시간이 짧으며 과적합 규제 기능으로 내구성이 우수한 성능을 보이며 자체 내장된 교차 검증 기능과 결손값을 자체적으로 처리할 수 있다. XGBoost 기법을 이용하여 모델을 구성하였다. 통계 모델에서 각 그룹별 예측 점수를 혈장 수집 시점, 나이, 가슴기살균제 노출 시점, 노출 물질 및 노출량, 가족력 등의 임상 정보와 연관 분석을 통해 임상적 연관성이 있는지를 확인하였다.

[0044] 비노출 대조군과 노출 폐손상군 1, 2, 3, 4 단계로부터 추출된 단백질의 정량 정보를 이용하여 가슴기살균제 노출 폐손상 단계 또는 노출 여부를 구별할 수 있는 모델을 구축하기 위해 1차적으로 XGBoost 기법을 통해 특징 추출 (Feature selection)을 수행하였다. 도 6에 나타난 바와 같이, 217개의 단백질 중 29개의 단백질들이 5개의 그룹에서 차이를 보이는 것으로 확인되었으며 30번째 단백질부터는 모델에 기여하는 계수 (Coefficient)가 0으로 나타나 모델에 기여하지 못하는 것으로 나타났다. 또한, 217개의 단백질 중에서 XGBoost 기법으로 선정된 29개의 단백질의 종류와 그 계수는 도 7과 같이 선별되었다.

[0045] 2차적으로 상기 선별된 29개의 단백질들을 이용하여 10배 교차 검증 (10 fold cross validation)을 10,000회 반복하면서 training 오류를 줄였습니다. 사용한 파라미터는 최대 깊이는 7이고 학습율은 0.7로 하였다. 계수가 큰 순서대로 XGBoost 알고리즘 중 진입 (Forward) 기법으로 29개 단백질들 중 13개의 단백질들을 선별하였다. 도 8과 같이, 13개의 단백질들이 조합이 되었을 때 통계모델이 구별할 수 있는 평균제곱근 편차 (Root Mean Square Deviation, RMSD)값의 변화 그래프가 변곡점 (Inflection point)를 구성하였다. 이는 13개의 조합으로 구성된 통계 모델이 5개의 그룹을 통계적으로 유의적으로 구분할 수 있는 최소화된 조합임을 의미한다.

[0046] 상기 최종적으로 선정된 13개의 단백질 패널을 이용하여 전체 시료를 그룹별로 3 등분을 하고 이중 2/3의 정보를 이용하여 모델을 구성하는 훈련 세트 (training set)로 사용하고 나머지 1/3의 정보는 모델의 완성도를 검증하는 검증 세트 (test set)로 사용하여 모델을 구성하고 XGBoost 기법으로 가장 우수한 변별력을 보이는 모델도 도 9와 같이 구성하였다. 단백질 정량값은 표 2와 같으며, 총 327개의 샘플 그룹은 표 3과 같이 훈련 세트 (training set) 또는 검증 세트 (test set)로 분류하였다. 총 327개의 시료로부터 훈련 세트 (training set)로 218개를, 검증 세트 (test set)로 109개의 시료를 사용하였으며 각 그룹별로 1/3과 2/3에 해당되는 시료에 대한 비율을 유지하기 XGBoost를 이용한 모델을 구성하고 검증하였다.

표 2

No.	단백질	가중치
1	ICAM1(Intercellular adhesion molecule 1)	5.07
2	LCP1(Plastin-2)	4.67
3	APOF(Apolipoprotein F)	4.08
4	CFH(Complement factor H)	2.96
5	FN1(Fibronectin precursor)	2.61
6	COMP(Cartilage oligomeric matrix protein)	2.48
7	HBD(Hemoglobin subunit delta)	-1.46
8	SAA4(Serum amyloid A-4 protein)	-2.01
9	APOD(Apolipoprotein D)	-3.80
10	ACTC1(Actin, alpha cardiac muscle 1)	-4.09
11	AZGP1(Zinc-alpha-2-glycoprotein)	-4.98
12	APOB(Apolipoprotein B-100)	-5.63
13	CD5L(CD5 antigen-like)	-7.73

표 3

	대조군	4단계	3단계	2단계	1단계	총 샘플
훈련 세트	70	50	34	36	28	218

검증 세트	34	26	17	18	14	109
총 샘플	104	76	51	54	42	327

[0049] 도 10에 나타난 바와 같이, 분석 결과 13개의 단백질 패널로 구성된 모델은 가슴기살균제 비노출 대조군 (평균 예측 점수 7.48)과 단계별 노출군 4 그룹들 (평균 예측 점수 42.94, 47.87, 55.30 및 74.54)들을 구분할 수 있는 것으로 나타났다. 가슴기살균제 노출 폐손상 4단계 그룹과 3단계 그룹 및 2단계 그룹의 평균 예측 점수들은 차이가 나는 것으로 나타났다.

[0050] 또한, 하기 표 4와 같이 19개의 단백질(PROZ, ABI3BP, ICAM1, AZGP1, CD5L, APOB, ACTC1, LCP1, APOD, CFH, PTGDS, FN1, APOF, SERPING1, VWF, THBS1, HBD, COMP, SAA4)들로 구성된 모델을 이용한 5개 그룹을 구분한 결과, 도 11에 나타난 바와 같이, 19개 단백질 기반의 모델의 구분능 성능은 13개 단백질 기반의 모델 구분능보다는 저하된 것으로 나타났다. 가슴기살균제 노출 폐손상 4단계 그룹과 3단계 그룹 및 2단계 그룹의 최소-최대값의 차이가 거의 유사하며 상호 교차되는 영역이 보다 많아 구분능이 떨어지는 것으로 나타났다. 이 결과는 보다 많은 단백질들이 모델에 투입된다고 해서 모델의 성능이 개선되는 것이 아님을 보여주는 결과이다.

표 4

No.	Gene	Weight
1	PROZ (Vitamin K-dependent protein Z)	7.80
2	ABI3BP (Target of Nesh-SH3)	7.51
3	ICAM1 (Intercellular adhesion molecule 1)	6.67
4	LCP1 (Plastin-2)	4.75
5	CFH (Complement factor H)	4.46
6	FN1 (Fibronectin)	3.50
7	APOF (Apolipoprotein F)	3.29
8	SERPING1 (Plasma protease C1 inhibitor)	2.79
9	VWF (von Willebrand factor)	2.50
10	THBS1 (Thrombospondin-1)	2.06
11	COMP (Cartilage oligomeric matrix protein)	1.56
12	SAA4 (Serum amyloid A-4 protein)	-0.71
13	HBD (Hemoglobin subunit delta)	-1.63
14	PTGDS (Prostaglandin-H2 D-isomerase)	-4.23
15	APOD (Apolipoprotein D)	-4.54
16	ACTC1 (Actin, alpha cardiac muscle 1)	-5.38
17	APOB (Apolipoprotein B-100)	-5.74
18	CD5L (CD5 antigen-like)	-5.82
19	AZGP1 (Zinc-alpha-2-glycoprotein)	-6.04

[0052] 상기 13개의 단백질 패널로 구성된 모델을 이용하여 가슴기살균제 비노출군과 노출 폐손상군(1, 2, 3 및 4 단계)을 구분할 수 있는지를 검증하였다. 도 12에 나타난 바와 같이, 특이도를 98.08% 고정할 경우 모델의 예측 점수의 임계값을 26.975로 하여 민감도가 90.13%인 것으로 나타났다. 양성예측도 즉 노출군을 노출군이라고 정확하게 예측하는 정확도는 99.01%였으며 음성 예측도 즉 비노출군을 비노출군이라고 정확하게 예측하는 정확도는 82.58%로 나타났다. 임상적 유용성을 나타내는 수신동작특성(ROC, receiver operating characteristics)의 곡면하면적 (Area under curve, AUC)은 0.9755로 매우 우수하게 나타났다.

[0053] 상기 결과는 가슴기살균제에 노출된 그룹을 구분함에 있어서 13개의 단백질들로 구성된 모델을 이용하는 경우, 실제 가슴기살균제에 노출된 그룹을 폐질환 단계와 상관없이 99.01%의 정확도를 가지고 가슴기살균제에 노출되지 않은 그룹과 구분할 수 있음을 의미한다. 따라서 가슴기살균제를 실제로 썼음에도, 구매 영수증이나 관련 사진 등의 직접적인 근거를 확보하고 있지 못한 신청자들의 경우 단백질 검사를 통해 26.975점 이상일 경우 가슴기살균제 사용의 근거로 사용할 수도 있음을 시사한다.

[0054] 또한, 상기 13개의 단백질 패널로 구성된 모델을 이용하여 가슴기살균제 비노출군과 가슴기살균제로 인한 폐손상군(1, 2단계)를 구분할 수 있는지 검증하였다. 도 13에 나타난 바와 같이, 특이도를 99.04% 고정하면서 모델의 예측 점수의 임계값을 31.065로 정하는 경우 모델의 민감도는 90.625%로 나타났다. 그리고 양성예측도는 98.86%였으며 음성 예측도는 91.96%로 나타났다. 수신 동작 특성(ROC, receiver operating characteristics)의

곡면하면적 (Area under curve, AUC)은 0.9734로 매우 우수하게 나타났다.

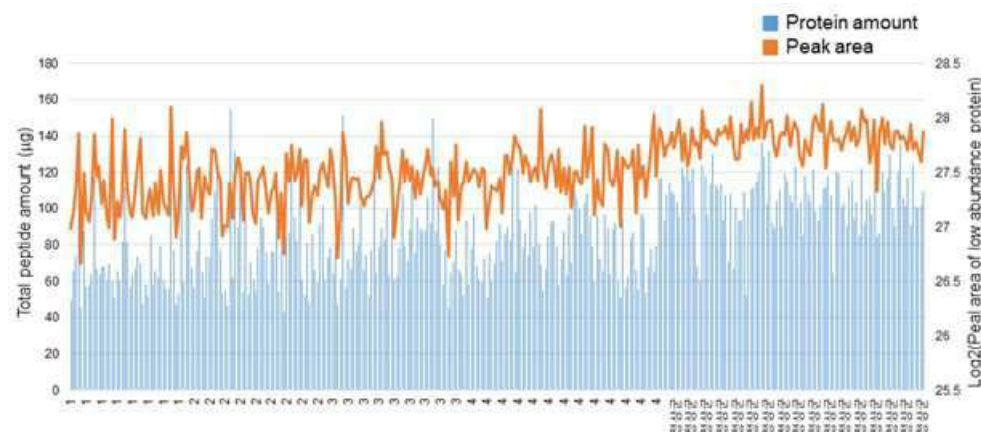
[0055] 상기 결과는 13개의 단백질들로 구성된 모델은 가슴기살균제로 인한 폐손상군(1, 2단계)을 98.86%의 정확도로 가슴기살균제에 노출되지 않은 그룹과 구분할 수 있음을 의미하고 있다. 총 96명의 1, 2단계 환자들 중 90명을 1, 2단계 그룹으로 정확하게 분류할 수 있다. 그리고 비노출군을 비노출군이라고 정확하게 예측하는 음성예측도의 정확도는 91.96%로 총 104명의 비노출군 중 2명의 노출군으로 잘못 판별할 수 있는 가능성이 8%가 존재하여 개선된 성능을 보이는 것으로 확인하였다.

[0056] 또한, 상기 13개의 단백질 패널로 구성된 모델을 이용하여 가슴기살균제 비노출군과 가슴기살균제로 인한 폐손상이 임상적으로 거의 확실한 1단계 노출군을 구분할 수 있는지 검증하였다. 도 14에 나타난 바와 같이, 모델의 예측 점수의 임계값을 35.32로 정하는 경우 특이도 100%, 민감도 100%로 나타났다. 그리고 양성예측도와 음성예측도 모두 각각 100%로 나타났다. 1단계 노출군 42명은 비노출군 104명으로부터 100% 구분이 되므로 수신 동작 특성의 곡면하면적은 1.0로 나타났다. XGBoost로 선별한 13개의 혈장 단백질들로 구성된 예측 모델은 1단계를 가슴기살균제에 노출되지 않은 그룹으로부터 100%의 정확도를 가지고 구분할 수 있다. 총 42명의 1단계 가슴기살균제에 노출된 환자들 중 42명을 1단계로 정확하게 분류할 수 있음을 의미한다. 이는 1단계 노출군이 임상적으로 노출로 인한 폐질환 등 임상적 지표가 명확하기 때문에 단백질이라는 생체 분자를 기반으로한 분류를 할 수 있는 모델에서 가장 잘 적용되는 것임을 입증한다.

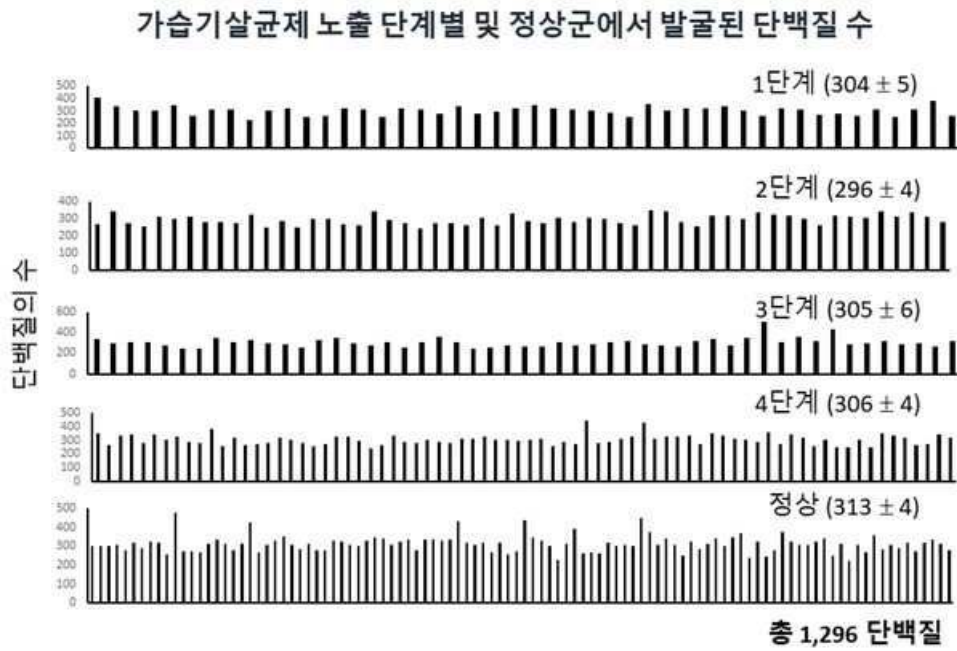
[0057] 이상으로 본 발명 내용의 특징한 부분을 상세히 기술하였는 바, 당업계의 통상의 지식을 가진 자에게 있어서, 이러한 구체적 기술은 단지 바람직한 실시양태일 뿐이며, 이에 의해 본 발명의 범위가 제한되는 것이 아닌 점은 명백하다. 즉, 본 발명의 실질적인 범위는 첨부된 청구항들과 그것들의 등가물에 의하여 정의된다.

도면

도면1

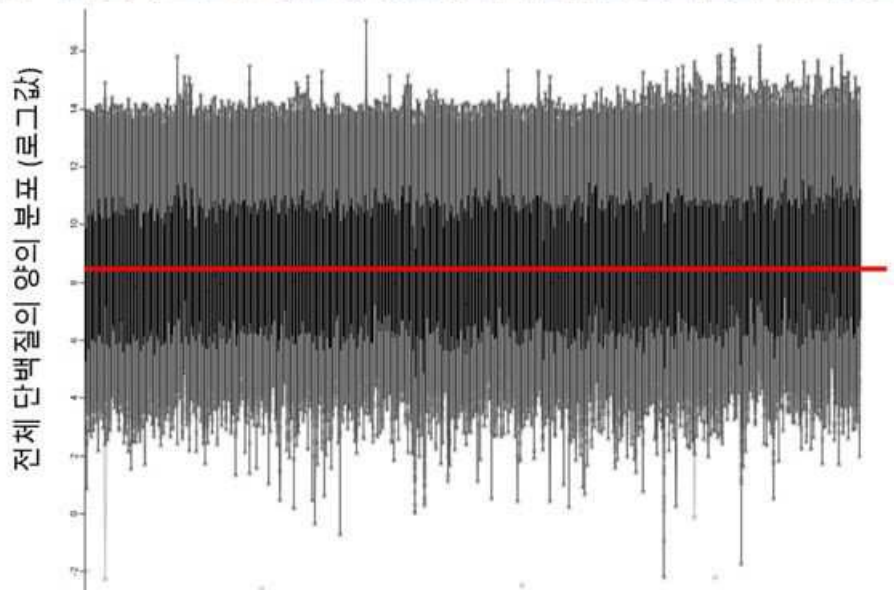


도면2

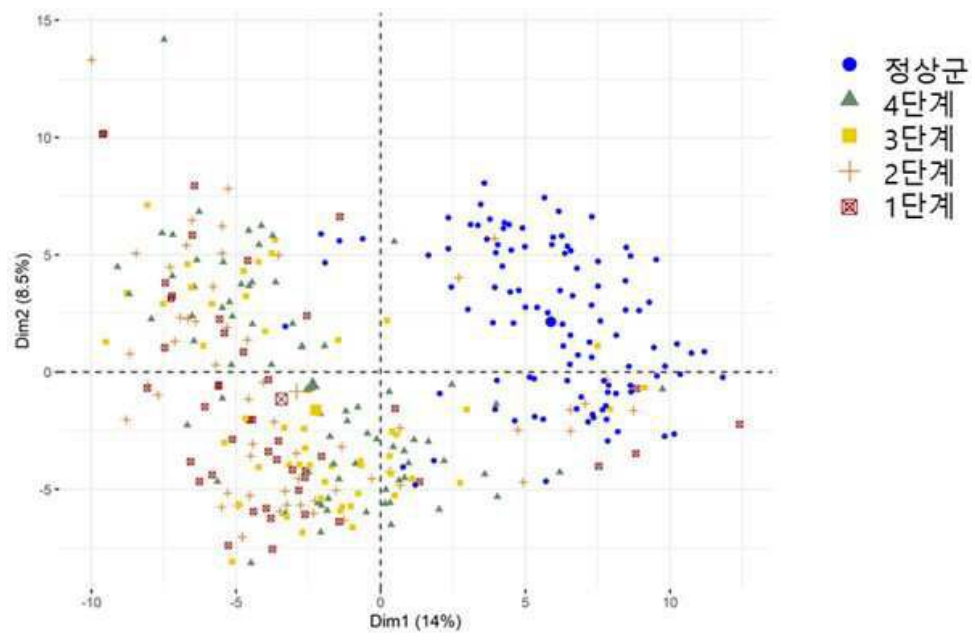


도면3

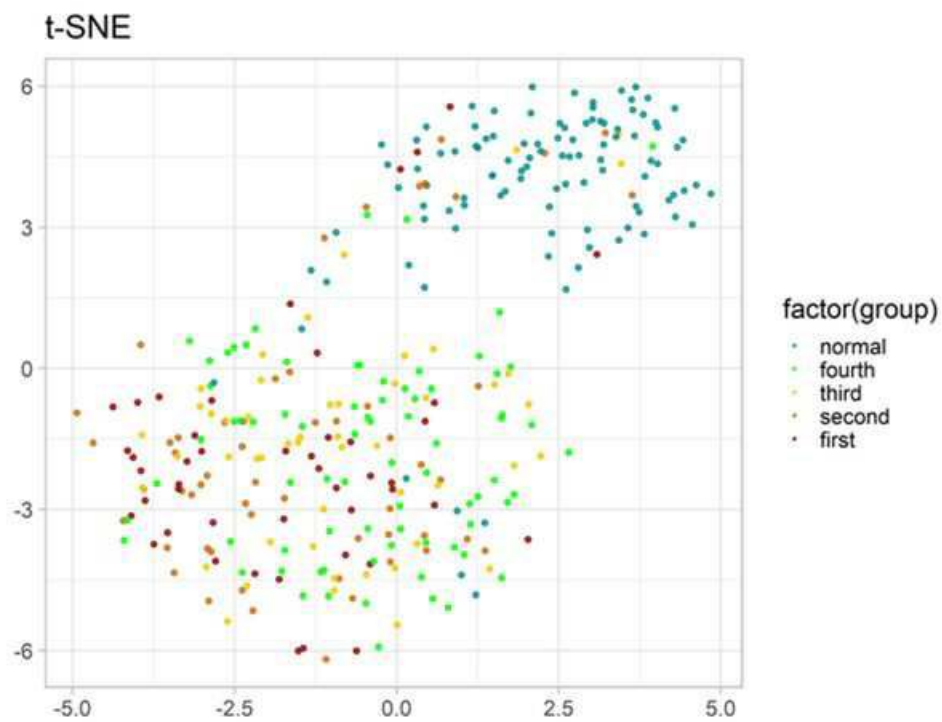
가습기살균제 노출 단계별 및 정상군에서 발굴된 단백질의 발현양 분포



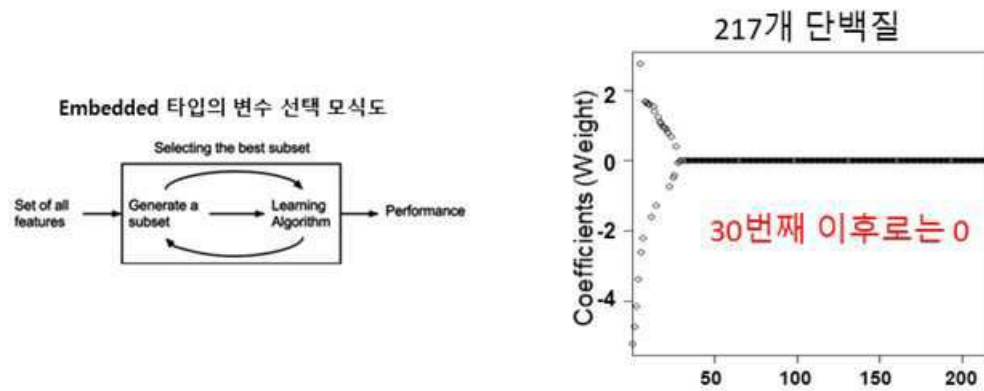
도면4



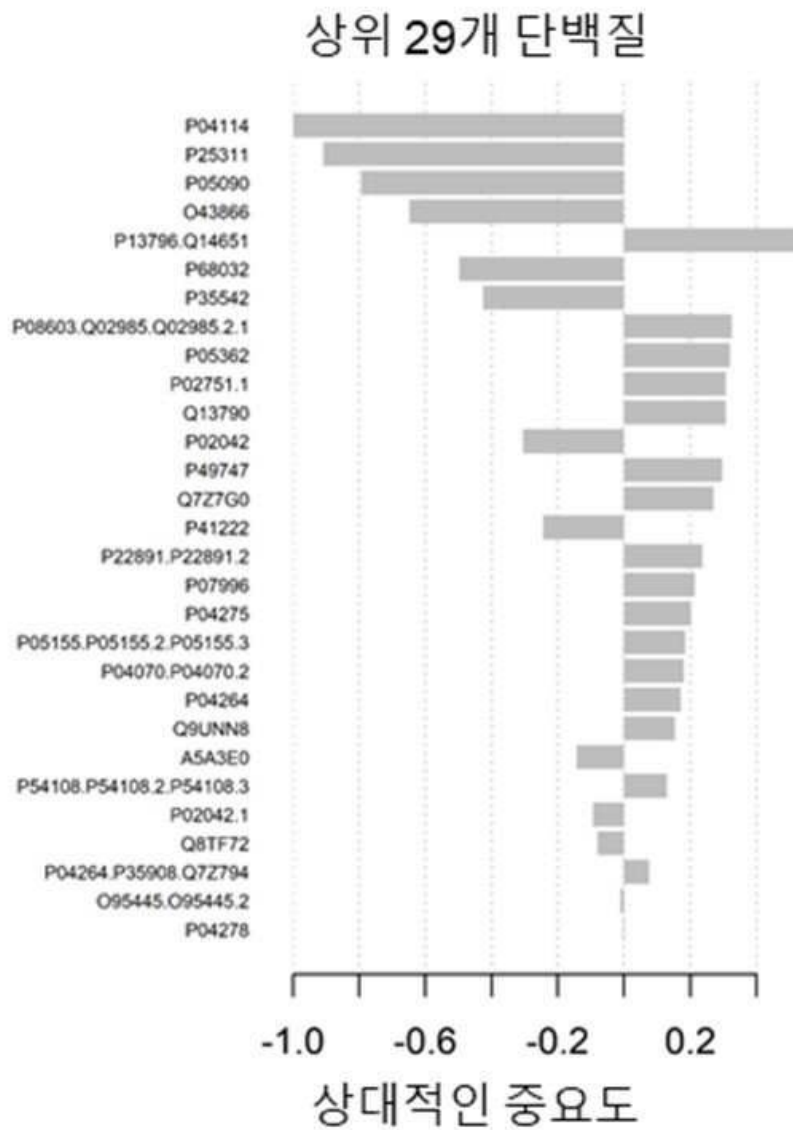
도면5



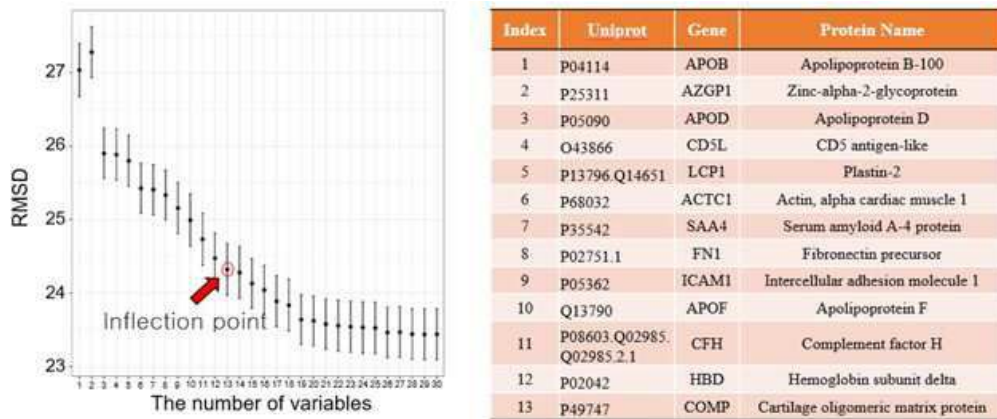
도면6



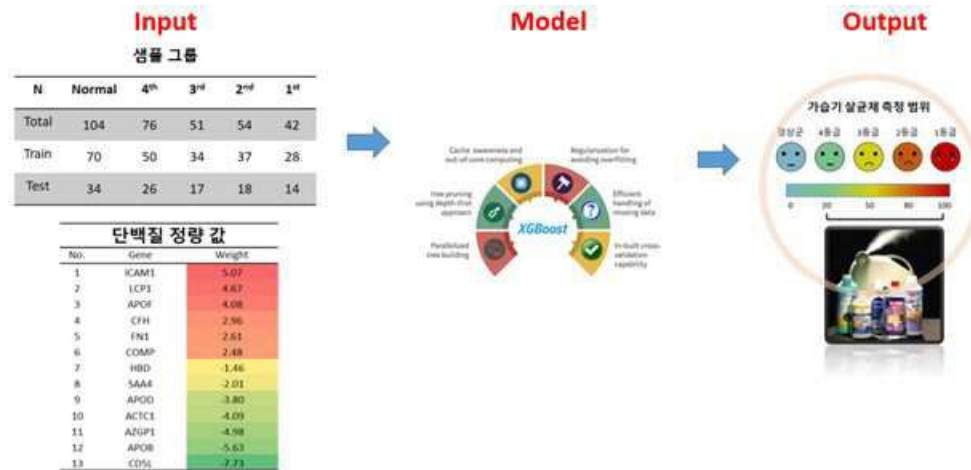
도면7



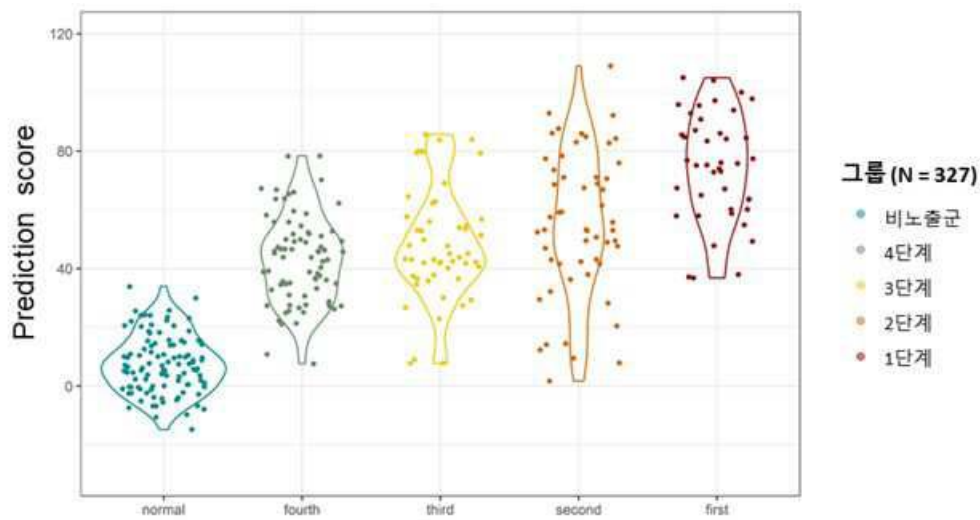
도면8



도면9

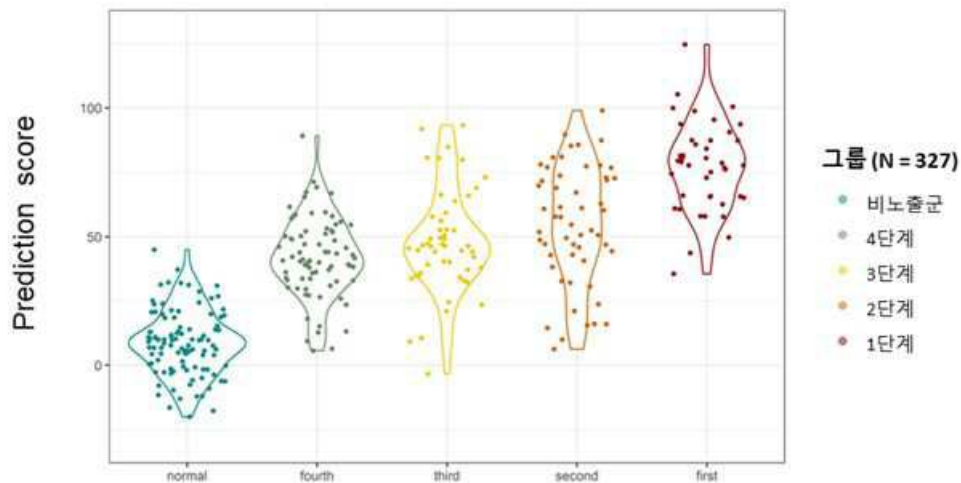


도면10



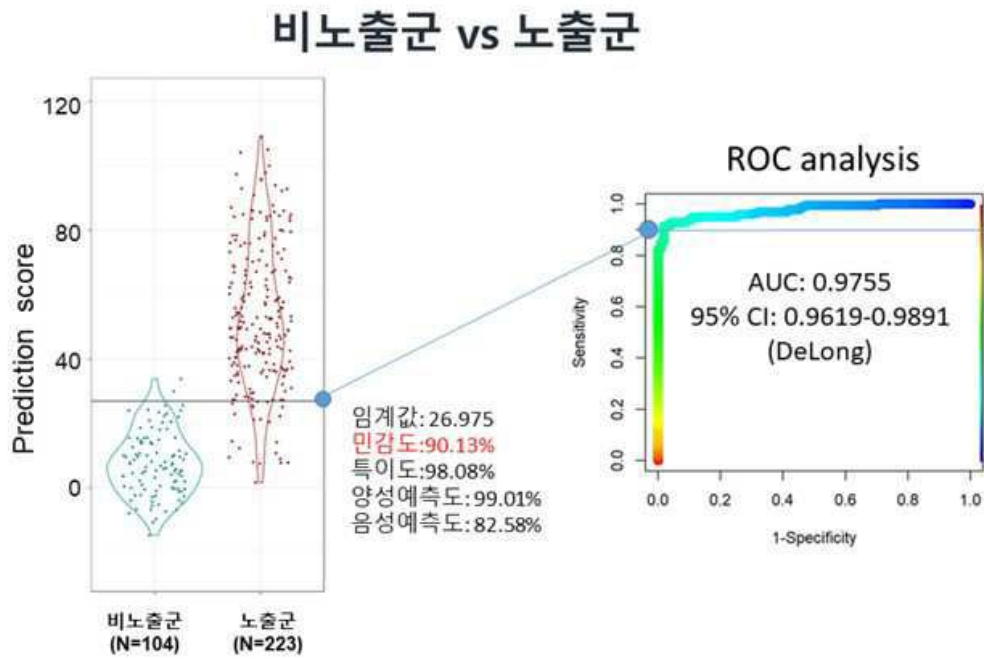
단계	Normal	4	3	2	1
평균	7.48	42.94	47.87	55.30	74.54
표준편차	9.87	14.83	18.84	24.75	18.31
최소값	-14.79	7.55	7.67	1.65	36.82
최대값	33.82	78.41	85.48	108.98	105.02

도면11

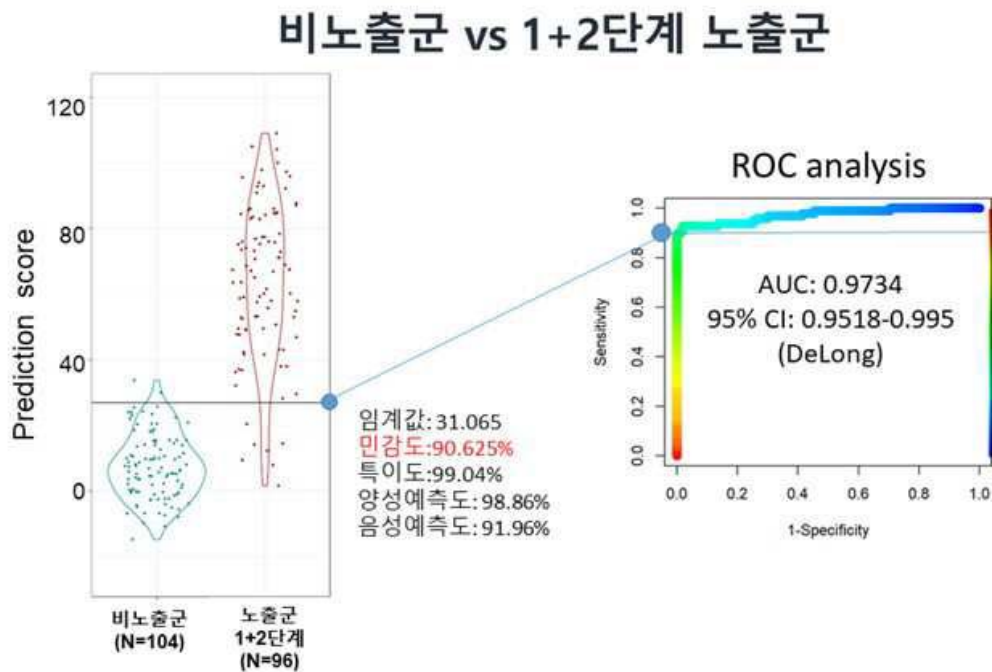


단계	Normal	4	3	2	1
평균	8.62	42.0	47.9	54.8	77.1
표준편차	12.5	15.6	20.1	22.9	17.3
최소값	-20.0	5.64	-3.33	6.31	35.5
최대값	44.9	89.1	93.3	99.0	125.0

도면12



도면13



도면14

비노출군 vs 1단계 노출군

