(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0131959
(43) 공개일자 2017년12월01일

(51) 국제특허분류(Int. Cl.)
C12Q 1/68 (2006.01)
(52) CPC특허분류
C12Q 1/6883 (2013.01)
C12Q 2600/136 (2013.01)
(21) 출원번호 10-2016-0062820
(22) 출원일자 2016년05월23일
심사청구일자 없음

(71) 출원인
울산대학교 산학협력단
울산광역시 남구 대학로 93(무거동)
재단법인 아산사회복지재단
서울특별시 송파구 올림픽로43길 88 (풍납동)
(72) 발명자
오연목
서울특별시 용산구 이촌로 181, 104동 1101호(한강대우아파트)
(74) 대리인
특허법인이룸리온

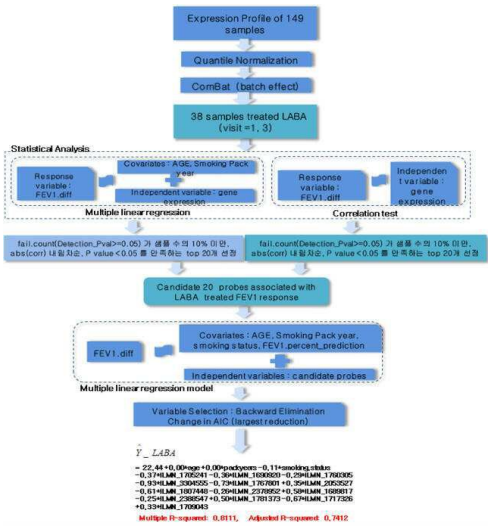
전체 청구항 수 : 총 7 항

(54) 발명의 명칭 **지속성 무스카린 작용제(LAMA)에 대한 약물반응성 예측 방법**

(57) 요 약

본 발명은 수식 1을 이용한 지속성 무스카린 작용제(long-acting muscarinic antagonist, LAMA)에 대한 약물반응성 예측 방법에 관한 것으로서, 구체적으로 만성 폐쇄성 폐질환 환자를 치료하기 위하여 지속성 무스카린 작용제에 대한 약물반응성 예측 방법을 위한 수식을 제공하며, 상기 수식에 의하여 부작용을 줄이고 효과가 큰 약물을 선택할 수 있다.

대 표 도 - 도1



(52) CPC특허분류
C12Q 2600/158 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 HI13C1634
부처명 보건복지부
연구관리전문기관 보건복지부
연구사업명 질환극복기술개발
연구과제명 병원 주도 첨단 생명-정보-공학-임상 융합연구
기 여 율 1/1
주관기관 서울아산병원
연구기간 2013.11.01 ~ 2019.10.31

명 세 서

청구범위

청구항 1

하기 수식 1을 이용한 지속성 무스카린 작용제(long-acting muscarinic antagonist, LAMA)에 대한 약물반응성 예측 방법으로서,

<수식 1>

Y = 0.80 - 0.0040783×1초간 노력성 호기량(FEV1) + 0.24×ILMN_3210917 - 0.21×ILMN_1752027 - 0.22×ILMN_1779492 + 0.30×ILMN_1797107 - 0.24×ILMN_1659761

여기에서, 상기 ILMN_3210917 은 인간 miscRNA(LOC389168)의 발현량;

상기 ILMN_1752027 은 인간 UBE3B (ubiquitin protein ligase E3B)의 전사체 변이 1(transcript variant 1)로서 mRNA의 발현량;

상기 ILMN_1779492 은 인간 SMURF1 (SMAD specifi E3 ubiquitin protein ligase 1)의 전사체 변이 1(transcript variant 1)로서 mRNA의 발현량;

상기 ILMN_1797107 은 인간 SCLT1 (sodium channel and clathrin linker 1)로서 mRNA의 발현량; 및

상기 ILMN_1659761 은 인간 SNX29 (sorting nexin 29)로서 mRNA의 발현량인 것을 특징으로 하는 지속성 무스카린 작용제에 대한 약물반응성 예측 방법.

청구항 2

제 1 항에 있어서,

상기 ILMN_3210917 은 서열번호 1인 프로브;

상기 ILMN_1752027 은 서열번호 2인 프로브;

상기 ILMN_1779492 은 서열번호 3인 프로브;

상기 ILMN_1797107 은 서열번호 4인 프로브; 및

상기 ILMN_1659761 은 서열번호 5인 프로브를 이용하여 각 발현량을 측정하는 것을 특징으로 하는 방법.

청구항 3

제 2 항에 있어서,

상기 측정은 마이크로어레이(microarray), 중합효소 연쇄반응(polymerase chain reaction), 역전사 중합효소 연쇄반응(reverse transcriptase-polymerase chain reaction) 및 실시간 중합효소 연쇄반응(real time-polymerase chain reaction)으로 이루어진 군 중에서 선택되는 어느 하나인 것을 특징으로 하는 방법.

청구항 4

제 1 항에 있어서,

상기 지속성 무스카린 작용제는 스코폴라민(Scopolamine), 아트로핀(Atropine), 하이드록시진(Hydroxyzine), 이프라트로피움(Ipratropium), 트로피카마이드(Tropicamide), 피렌제핀(Pirenzepine), 디펜하이드라민(Diphenhydramine), 디멘하이드리네이트(Dimenhydrinate), 디싸이클로민(Dicyclomine), 플라복세이트(Flavoxate), 옥시부티닌(Oxybutynin), 티오트로피움(Tiotropium), 시클로펜톨레이트(Cyclopentolate), 솔리페나신(Solifenacin), 다리페나신(Darifenacin), 벤즈트로핀(Benzatropine), 메베베린(Mebeverine) 및 프로시클리딘(Procyclidine)으로 이루어진 군에서 선택되는 어느 하나 이상인 것을 포함하는 것을 특징으로 하는 방법.

청구항 5

제 1 항에 있어서,

Y 값이 0.1 이상이면 약물이 효과가 있다는 것이고, Y 값이 0.1 미만이면 약물이 효과가 없다는 것을 특징으로 하는 방법.

청구항 6

제 1 항 내지 제 5 항 중 어느 한 항을 이용하여 만성 폐쇄성 폐질환(COPD) 치료를 위한 약물 선택 결정을 위한 정보의 제공 방법.

청구항 7

제 6 항에 있어서,

상기 만성 폐쇄성 폐질환은 만성 폐쇄성기관기염, 만성 세기관지염, 폐기종으로 이루어진 군으로부터 선택되는 어느 하나 이상인 것을 포함하는 것을 특징으로 하는 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 지속성 무스카린 작용제(LAMA)에 대한 약물반응성 예측 방법으로서, 구체적으로 만성 폐쇄성 폐질환(COPD) 치료제인 지속성 무스카린 작용제(LAMA)에 대한 약물반응성에 대한 수식을 제공하고, 상기 수식에 의해 약물의 효과를 예측할 수 있는 방법에 관한 것이다.

배 경 기 술

[0002] 만성 폐쇄성 폐질환(Chronic Obstructive Pulmonary Disease)은 만성기관지염이나 폐기종에 의한 기도폐쇄를 특징으로 하는 질병상태로서, 비가역적인 기도의 폐쇄를 동반하는 것이 특징이다.

[0003] 만성 폐쇄성 폐질환은 기도폐쇄 여부의 객관적인 측정, 즉 폐기능검사에 의해서만 확인되며, 만성폐쇄성 폐질환에서 폐기능검사 소견의 특징은 노력성폐활량 측정법상 최대 호기류의 감소와 이에 따른 호기의 지연 소견이 수개월 이상의 관찰기간에 저절로 혹은 치료에 의해 현저한 변화가 없으면서 정상화될 수 없다는 점이다. 폐기능검사의 진단기준은 1초간 노력성호기량(FEV1)이 추정 정상치의 80% 미만이면서 1초간 노력성호기량의 노력성폐활량(FVC)에 대한 비(FEV1/FVC)가 70% 미만인 경우이다.

[0004] 지금까지 만성 폐쇄성 폐질환의 경우 염증성 질환의 치료를 위해 항염증 작용이나 기관지확장 효과를 갖는 치료제를 이용하여 왔다. 그런 기존의 상당수 치료제들은 많은 부작용으로 사용시 주의를 요하고 있다. 대표적인 치료제는 글루코콜티코이드(glucocorticoid), 루코트리엔 모디파이어(leukotriene modifiers), 데오필린(theophylline) 등이 있다. 글루코콜티코이드(glucocorticoid)는 효과 면에서는 강력하나 선택적으로 작용하는 것이 아니라 모든 면역반응과 항염증반응을 억제하기 때문에 경우에 따라서 필요한 면역반응까지 억제하는 문제점이 있으며, 약물부작용이 문제가 되어 흡입 치료를 시행하고 있다. 루코트리엔 모디파이어 (leukotriene modifiers)는 부작용은 적지만 효과 면에서 한계가 있어 단독으로 사용 시에는 천식을 조절할 수 없어, 대부분 보조적으로 사용하고 있는 문제점이 있다. 데오필린(theophylline)은 효과 면에서도 뛰어나지 못하며 부작용의 우려가 있는 문제점이 있다. 코르티코스테로이드(corticosteroid) 제제는 뛰어난 치료 효과를 나타내지만 장기적으로 사용할 경우 용량과 사용시간에 비례하여 부신 억제, 골밀도 감소, 성장 장애, 눈과 피부의 합병증, 콜라겐의 합성 증가 등을 유발하는 것으로 알려져 있다. 살메테롤(salmeterol)과 포르메테롤(formeterol)과 같은 지속성 베타-2 길항제 (beta-2 agonist)는 발작에 대해 예방 효과를 나타내지만 경우에 따라서는 환자를 사망케 할 수도 있다고 경고된 바 있다. 이와 같은 다양한 부작용들로 인해 종래의 염증성 질환의 치료제들은 그 사용에 있어 신중한 고려가 필요하며, 효과가 뛰어나고 부작용이 적은 치료제의 개발이 절실히 요구되고 있다.

[0005] 따라서, 본 발명자들은 COPD 환자에게 지속성 무스카린 작용제를 사용하는 경우 약물에 대한 반응을 예측하여 약물의 부작용을 줄이고 효과가 뛰어난 약물을 선별하여 치료할 수 있도록 약물반응성 예측을 위한 방법을 개발하여 본 발명을 완성하였다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 1. 미국등록특허 제6797472호

(특허문헌 0002) 2. 한국등록특허 제0502523호

(특허문헌 0003) 3. 한국공개특허 제20117028551호

비특허문헌

[0007] (비특허문헌 0001) 1. The Korean Journal of Medicine, 2007, 73(2); 159-168 (2007.08)

발명의 내용

해결하려는 과제

[0008] 본 발명의 목적은 지속성 무스카린 작용제(long-acting muscarinic antagonist, LAMA)에 대한 약물반응성 예측 방법을 제공하는 것이다.

[0009] 본 발명의 또 다른 목적은 만성 폐쇄성 폐질환을 치료하기 위한 약물 선택 결정을 위한 정보의 제공 방법을 제공하는 것이다.

과제의 해결 수단

[0010] 상기의 목적을 달성하기 위하여, 본 발명은 하기 수식 1을 이용한 지속성 무스카린 작용제에 대한 약물반응성 예측 방법을 제공한다.

[0011] 본 발명의 수식 1은 다음과 같다.

[0012] Y = 0.80 - 0.0040783×1초간 노력성 호기량(FEV1) + 0.24×ILMN_3210917 - 0.21×ILMN_1752027 - 0.22×ILMN_1779492 + 0.30×ILMN_1797107 - 0.24×ILMN_1659761

[0013] 상기 수식 1에서 결정계수(Multiple R-squared)는 0.8457이고, 수정된 결정계수(Adjusted R-squared)는 0.8114이다.

[0015] 본 발명에서 용어, "1초간 노력성 호기량(FEV1, forced expiratory volume at 1 second)은 만성 폐쇄성 폐질환 환자에서 폐기능검사의 진단기준에 사용되는 판단기준으로 활용되고 있다. 폐기능검사의 진단기준은 1초간 노력성호기량(FEV1)이 추정 정상치의 80% 미만이면서 1초간 노력성호기량의 노력성폐활량(FVC)에 대한 비(FEV1/FVC)가 70% 미만인 경우이다.

[0016] 본 발명의 일실시예에 있어서, 수식 1에서 상기 ILMN_3210917 은 인간 miscRNA(LOC389168)의 발현량; 상기 ILMN_1752027 은 인간 UBE3B (ubiquitin protein ligase E3B)의 전사체 변이 1(transcript variant 1)로서 mRNA의 발현량; 상기 ILMN_1779492 은 인간 SMURF1 (SMAD specifi E3 ubiquitin protein ligase 1)의 전사체 변이 1(transcript variant 1)로서 mRNA의 발현량; 상기 ILMN_1797107 은 인간 SCLT1 (sodium channel and clathrin linker 1)로서 mRNA의 발현량; 및 상기 ILMN_1659761 은 인간 SNX29 (sorting nexin 29)로서 mRNA의 발현량이다.

[0017] 본 발명의 일실시예에 있어서, 수식 1에서, 상기 ILMN_3210917 은 서열번호 1인 프로브; 상기 ILMN_1752027 은 서열번호 2인 프로브; 상기 ILMN_1779492 은 서열번호 3인 프로브; 상기 ILMN_1797107 은 서열번호 4인 프로브; 및 상기 ILMN_1659761 은 서열번호 5인 프로브를 이용하여 각 발현량을 측정하는 것일 수 있다.

[0018] 본 발명의 일실시예에 있어서, 프로브를 이용한 각 발현량의 측정은 마이크로어레이(microarray), 중합효소 연쇄반응(polymerase chain reaction), 역전사 중합효소 연쇄반응(reverse transcriptase-polymerase chain

reaction) 및 실시간 중합효소 연쇄반응(real time-polymerase chain reaction)으로 이루어진 군 중에서 선택되는 어느 하나인 것일 수 있다.

[0019] 본 발명의 일실시예에 있어서, 프로브를 이용한 각 발현량의 측정을 위한 생물학적 시료는 개체로부터 유래된 신선한 또는 보존된 기관 또는 조직 시료 또는 생검(biopsy)와 같은 고체 조직(solid tissue); 혈액 또는 혈액 구성성분; 양수(amniotic fluid), 복수(peritoneal fluid), 또는 세포간질액(interstitial fluid)와 같은 체액(bodily fluid); 세포, 또는 이들의 조합일 수 있다. 상기 시료는 보존제, 항응고제, 버퍼, 고정제(fixatives), 영양성분(nutrients), 항생제 등과 같은 생물학적 물질과 자연적으로 혼합되어 있지 않은 화합물을 포함할 수 있다. 상기 생물학적 시료는 예를 들어 폐 조직일 수 있다. 폐 조직은 외과 수술에 의해 얻어지거나, 물질적 또는 레이저 등을 통한 광학적 적출에 의해 얻어질 수 있다.

[0020] 본 발명의 일실시예에 있어서, 상기 개체는 포유동물일 수 있다. 상기 포유동물은 영장류일 수 있다. 상기 포유동물은 사람, 마우스, 소, 돼지, 말, 양, 개, 고양이, 또는 그 조합일 수 있다. 상기 개체는 COPD에 걸린 것으로 의심되는 개체일 수 있다. 상기 개체는 흡연자일 수 있다. 상기 흡연자는 담배를 피우는 행위를 하거나, 간접적으로 담배 연기에 노출된 개체를 포함한다. 상기 흡연자는 흡연 강도 하위 3분위인 개체일 수 있다. 예를 들어, 흡연 강도 하위 3분위는 35 갑년(pack-year; PY), 35 갑년 미만, 1 갑년 내지 35 갑년, 10 갑년 내지 35 갑년, 20 갑년 내지 35 갑년, 또는 30 갑년 내지 35갑년일 수 있다. 갑년은 1 년 동안 하루 한 갑씩 담배를 피웠을 경우를 기준으로 하는 담배 소비량을 말하고, 1 일 당 흡연량 및 흡연한 연수로 산출할 수 있다.

[0022] 본 발명의 일실시예에 있어서, 지속성 무스카린 작용제는 스코폴라민(Scopolamine), 아트로핀(Atropine), 하이드록시진(Hydroxyzine), 이프라트로피움(Ipratropium), 트로피카마이드(Tropicamide), 피렌제핀(Pirenzepine), 디펜하이드라민(Diphenhydramine), 디멘하이드리네이트(Dimenhydrinate), 디싸이클로민(Dicyclomine), 플라복세이트(Flavoxate), 옥시부티닌(Oxybutynin), 티오트로피움(Tiotropium), 시클로펜톨레이트(Cyclopentolate), 솔리페나신(Solifenacin), 다리페나신(Darifenacin), 벤즈트로핀(Benzatropine), 메베베린(Mebeverine) 및 프로시클리딘(Procyclidine)으로 이루어진 군에서 선택되는 어느 하나 이상인 것일 수 있다.

[0023] 본 발명의 일실시예에 있어서, 상기 수식 1에서 Y 값이 0.1 이상이면 약물이 효과가 있다는 것이고, 값이 0.1 미만이면 약물이 효과가 없다는 것일 수 있다.

[0024]

[0025] 또한, 본 발명은 만성 폐쇄성 폐질환의 치료를 위한 약물 선택 결정을 위한 정보의 제공 방법을 제공한다.

[0026] 본 발명의 일실시예에 있어서, 만성 폐쇄성 폐질환은 만성 폐쇄성기관기염, 만성 세기관지염, 폐기종으로 이루어진 군으로부터 선택되는 어느 하나 이상인 것일 수 있다.

[0027] 본 발명에서 용어, "만성 폐쇄성 폐질환"이란, 유해한 입자나 가스의 흡입에 의해 폐에 비정상적인 염증 반응이 일어나면서 이로 인해 점차 기류 제한이 진행되어 폐 기능이 저하되고 호흡곤란을 유발하게 되는 호흡기 질환을 말한다. 만성 폐쇄성 폐질환의 주된 증상은 만성적인 기침 또는 만성적인 가래(객담)을 비롯하여 호흡곤란 등이 있으며, 베타항진제, 항콜린제, 메틸잔틴계 등의 기관지 확장제 또는 부신피질 호르몬 흡입제 등이 대표적인 치료제로 사용된다. 본 발명에서 상기 만성폐쇄성 폐질환은 바람직하게는 만성 기관지염(chronic bronchitis) 또는 폐기종(emphysema)일 수 있으나, 이에 제한되지 않는다.

[0028] 본 발명에서 용어, "만성 기관지염(chronic bronchitis)"이란, 2년 연속, 1년에 3개월 이상 가래가 있고 기침이 지속되는 질환을 의미한다. 흡연, 대기오염, 직업적 노출 등의 자극에 의한 기관지 손상이 원인으로 추정되고 있으며, 주요 증상으로는 만성 기침, 가래, 운동시 호흡곤란 등이 있다. 또한, 만성폐쇄성 폐질환의 특징인 급성 악화가 있을 수 있으며, 이때에는 수 시간에서 수 일 사이 호흡곤란이 빠르게 악화되고 가래의 양이 늘어나거나 가래의 성상이 점액성에서 화농성으로 변하면서 진한 노란색이나 푸르스름한 색을 띄게 되고 점도가 높아져 뱉어내기 힘들어진다.

[0029] 본 발명에서 용어, "폐기종(emphysema)"이란, 종말 세기관지(terminal bronchiole) 원위부 공기공간(airspace)의 파괴로 인하여 비정상적이며 영구적인 말초 기도 및 폐포의 확장상태를 의미한다. 유해 입자와 가스의 흡입에 의하여 발생하며, 임상적으로 가장 의미있는 위험인자는 흡연으로 알려져 있다. 주요 증상으로는 만성적인 기침과 가래, 호흡곤란 등이 있다.

발명의 효과

[0031] 본 발명에 의하면 만성 폐쇄성 폐질환 환자에서 지속성 무스카린 작용제(LAMA)에 대한 약물반응성을 예측하여 약물에 대한 부작용을 줄이고 효과를 올릴 수 있어 만성 폐쇄성 폐질환을 치료하는데 유용하게 이용될 수 있다.

도면의 간단한 설명

[0032] 도 1은 전체적인 분석 워크플로우를 다이아그램(diagram)으로 나타낸 것이다.

도 2는 수행 시기별 처리된 샘플 수를 나타낸다.

도 3는 각 internal control에 관한 것이다.

도 4는 각 internal control 별로 실험 성공시 기대되는 수치(expected value)를 정리한 것이다.

도 5는 방문회수 1, 3회에서의 LAMA를 처리한 샘플을 나타낸 것이다.

도 6은 로그 변환 후, quantile normalization 전과 후를 비교한 density plot 을 나타낸다.

도 7은 샘플 그룹 내의 반복성을 확인하기 위하여 Pearson's correlation coefficient를 계산한 것이고, 높은 양의 상관관계로 갈수록 노란색으로 표시하였다.

도 8은 집단효과(batch effect)를 보정 전과 후 데이터의 양상으로 나타낸 것이다.

도 9는 약물 처리 전의 발현 데이터인 1, 3 회차의 약물별로 샘플 수를 확인한 것이다.

도 10은 계층적 군집분석(hierarchical clustering analysis)을 통해 적외선열지도(heatmap)으로 표현한 것이다. 오른쪽 상단의 범례에서 빨간색은 호전정도가 없음을 나타내며, 파란색은 호전정도가 있음을 나타낸다.

도 11은 상관성 분석과 다중선형회귀분석에서 유의성이 있는 결과에 대한 리스트를 선별한 것이다.

도 12는 LAMA 약물처치 전의 샘플 대상으로 폐기능 호전정도와 연관성이 있는 1047 프로브에 대하여 biological process 에 해당하는 GO gene set 중 가장 유의한 상위 10개의 gene set을 나타낸다.

도 13은 molecular function에 해당하는 GO gene set 중 유의했던 상위 10개의 gene set을 나타낸다.

도 14는 cellular component에 해당하는 GO gene set 중 유의했던 상위 10개의 gene set을 나타낸다.

도 15는 LAMA 예측모델에 대한 모형 진단을 나타낸 것이다.

도 16은 LAMA 샘플 set의 폐기능 호전정도를 예측하는 모델을 나타낸 것이다.

발명을 실시하기 위한 구체적인 내용

[0033] 이하 본 발명을 실시예에 의하여 더욱 상세하게 설명한다. 하기 실시예는 단지 본 발명을 보다 구체적으로 설명하기 위한 것으로, 본 발명의 범위가 이들 실시예에 국한되지 않는다는 것은 당업계에서 통상의 지식을 가진 자에게 있어서 자명할 것이다.

[0035] **<실시예 1> 전체적인 분석 워크플로우(workflow)**

[0036] 만성 폐쇄성 폐질환(Chronic Obstructive Pulmonary Disease, COPD)의 약물 치료제로 알려져 있는 지속성 무스카린 작용제(long-acting muscarinic antagonist, 이하 "LAMA"라고 함)에 대한 폐기능 호전도와 연관된 후보 유전자를 선별하기 위하여 약물 처치 전의 샘플에 대해 Illumina Human Expression HT12 v4 칩을 이용하여 발현 프로파일(expression profile)을 만들었다. 전체적인 분석 워크플로우는 다이아그램(diagram)으로 나타내었다(도 1).

[0038] **<실시예 2> 샘플 확보 및 샘플 품질관리(QC)**

[0039] 아산병원으로부터 만성 폐쇄성 폐질환 환자의 혈액에서 뽑은 RNA 샘플을 제공받아 RNA QC기준을 통과하는 총

149개에 대해 Illumina Human Expression HT12 칩 실험이 수행되었다. 세부 실험 디자인은 1명의 환자별 4번의 실험이 이뤄졌고, 회차가 1, 3일 경우는 약물처치 전 데이터이며, 회차가 2, 4일 경우는 약물 처치 후 데이터이다. 또한, 한 번의 약물 처치 후 이전 약물 효과가 없어지는 일정 기간이 지난 후, 개인별 다른 약물 처치하는 무작위추출(ramdomization)방법을 이용하였다. 도 2는 수행 시기별 처리된 샘플 수를 나타내며 총 8차례의 칩 실험이 수행되었다.

[0041] RNA 샘플 QC 기준은 도 3과 같으며, 추출된 total RNA를 Nanodrop(ND-1000) 기계를 이용하여 농도 및 A260/280, A260/230 값을 측정하고, Agilent bioanalyzer를 이용하여 28S rRNA/ 18S rRNA 값을 측정하여 total RNA의 깨진 정도를 판단하며 이 값을 기준으로 실험 진행 가능한 샘플을 선별하였다.

[0042] QC를 통과한 149개 샘플은 Illumina's standard protocol에 따라 실험 결과를 생산하였고, 각 실험 단계별 문제가 없었는지는 도 3a 내지 도 3e의 5가지 internal control을 통해 데이터 양상을 확인하였다(도 3f에서는 labeling 과 background 데이터를 추가하였음). 각 internal control별로 실험 성공시 기대되는 수치(expected value)를 정리한 것은 도 4에 나타내었다.

[0044] **<실시예 3> COPD 관련 약물반응성 연관된 유전체 발현 분석**

[0045] **1. 기초데이터(Raw data) 추출**

[0046] 총 여덟 차례에 걸쳐서 149개 샘플에 대한 실험결과로 얻었고, 각 샘플별 칩의 이미지파일(*.idat) 파일을 GenomeStudio의 Expression module을 이용하여 불러들여 샘플별, 프로브(probe)별 발현량을 대변하는 "avg_signal"값과 해당 signal이 background signal을 비교하여 유의한 signal인지를 판단하기 위한 "Detection_Pval"값을 추출하였다. 상기에 따라, probe의 "Detection_Pval"가 0.05보다 작은 경우(즉, Detection_Pval<0.05)에는 신뢰성 있는 signal이라고 판단하였다. 약물이 처리된 샘플별(X 축)로 "Detection_Pval <0.05"를 만족하는 프로브의 수(Y 축)를 파악해 보았고, 평균적으로 13,400여 개의 유전자 발현 값이 검출되었다(도 5).

[0048] **2. 데이터의 전처리**

[0049] 이미지(image) 데이터를 분석하고, 통계 분석을 용이하게 하기 위해 데이터를 로그 변환(log transformation) 및 체계적 편차(systematic bias)를 줄이기 위하여 정상화(normalization)를 진행하였다.

[0050] 로그 변환의 이유는 원본 signal 값의 범위가 크기 때문이다. Signal값에 "Log based 2"를 취해 변환을 하게 되면 범위가 줄어들게 되어 분석에 용이하며, 원본 signal값의 정보에 손실이 없으며, 아웃라이어(outliers)에도 덜 민감하게 된다. 또한, 마이크로어레이 분석의 대부분은 정규분포(normal distribution)를 가정하게 되는데, 로그 변환을 하게 되면 점근적으로 정규분포에 가까워지게 되어, 통계적인 방법들을 적용하기가 좋으며, 배수변환(fold change)을 설명하기에도 용이하다.

[0051] 실험과정 중에 발생되는 체계적 편차(예를 들어, 시약, 작업자, 스캐너, 실험환경 등)를 보정하기 위해 Quantile normalization 방법을 적용하였고, 도 6는 로그 변환 후, quantile normalization 전과 후를 비교한 density plot을 나타내고 있다. 샘플 그룹 내의 반복성을 확인하기 위하여 Pearson's correlation coefficient를 계산하였고, 상관관계가 많을수록 노란색으로 표시되었다(도 7).

[0052] 샘플 간의 유사 정도를 파악하기 위해 MDS(multidimensional scaling plot)을 통하여 확인해 보았고, 이 결과 샘플의 실험 집단간의 편차가 존재하여 ComBat 알고리즘을 통하여 집단효과(batch effect)를 보정한 것이다(ComBat, http://www.bu.edu/jlab/wp-assets/ComBat/Abstract.html). 집단효과에 대한 보정 전 후의 데이터는 도 8에 나타내었다.

[0054] **3. 약물반응성과 연관된 후보유전자 발굴**

[0055] 만성 폐쇄성 폐질환 환자를 대상으로 LAMA 약물처치 전과 후의 실험데이터를 축적하였고, 예비 테스트로 LAMA 약물 처리 전(회차=1, 3)의 데이터를 이용하여 약물 반응성과 연관된 후보 유전자를 조사하였고, 도 9은 약물

처리 전 발현 데이터인 "회차=1, 3" 에 대해서 약물별로 샘플 수를 확인한 내용이다.

[0056] 각 약물별 유전자 발현량과의 상관분석에서 상관계수의 절대값이 0.5 이상인 프로브를 대상으로 아래와 같이 계층적 군집분석(hierarchical clustering analysis)을 통해 적외선열지도(heatmap)으로 표현하여 샘플들간의 발현량이 폐기능 호전 여부에 따라 그룹화가 되는지 확인 하였다(도 10). 도 10에서 오른쪽 상단의 범례에서 빨간색은 호전정도가 없음을 나타내며, 파란색은 호전정도가 있음을 나타낸다.

[0057] 또한, 샘플의 나이 및 흡연기간을 고려하여 폐기능 호전정도와 연관성이 있는 유전자를 선별하기 위해 독립변수를 유전자, 나이, 흡연기간으로 두고, 종속변수를 "FEV1.diff" 값으로 둔 다중선형회귀분석을 추가로 진행하였다. 상관성 분석에서 유의하면서 다중선형회귀분석에서 유전자에 대한 회귀계수의 유의성 검사에서 유의한 리스트를 선별하였다(유의수준 0.05).

[0058] LAMA 약물처치 전의 폐기능 호전정도와 연관이 있었던 프로브들은 1047개였고(도 11), 이 유전자들에 대해 DAVID functional annotation tool을 이용하여 Public classification category인 GO(Gene Ontology), INTERPRO, PIR_SUPERFAMILY, SMART와 같은 단백질 도메인 분류, BBID, BIOCATA, KEGG와 같은 pathway 분류, OMIM 등 질병 정보를 이용하여 유전자가 관여하는 생리적, 분자생물학적 기능별로 분류하였다. EASE score를 이용한 GSEA(Gene Set Enrichment Analysis)을 통해 통계적으로 유의한 gene set을 최종적으로 검출하였다.

[0060] LAMA 약물처치 전의 샘플 대상으로 폐기능 호전정도와 연관성이 있는 1047 프로브들에 대해 biological process에 해당하는 GO gene set 중 가장 유의했던 term은 "GO:0008284~positive regulation of cell proliferation"이었고, 해당 gene set에 관여하는 유전자는 EDN3, BMP10, FGFR4, NRP1, FGFR3, TP63, KITLG, IL15, RPA1, ZFP91, HLX, FGF2, SPN, MATK, FLT1, GAS1, AGER, CCL15, MYCN, FOXP2, PTHLH, HIPK1, DLX5, BNC1, TGIF1, ADAM17 이었고, 주로 catabolic process에 해당하는 gene set이 나타났다(도 12).

[0061] LAMA set에서 molecular function에 해당하는 GO gene set 중 유의했던 term으로는 acid-amino acid ligase activity, calcium ion binding, small conjugating protein ligase activity, ligase activity, forming carbon-nitrogen bonds, ubiquitin-protein ligase activity 등 protein ligase activity 에 관련된 gene set이 유의하였다. 관련된 유전자 리스트는 SUMO2, UBE2E3, UBE2Z, TTLL10, UBE3B, CBL, UBE4B, FBXO2, FBXO25, RNF168, SMURF1, AMFR, BCOR, FBXO9, UBE2E1, ASNS 이었다(도 13).

[0062] LAMA set에서 세포의 구성요소에 해당하는 GO gene set 중 유의했던 term으로는 extracellular region, extracelluar matrix 지역에 enriched gene들이 존재하는 것으로 보였다(도 14).

[0064] **4. LAMA 예측모델에 대한 모형 진단**

[0065] LAMA 약물에 따라서 폐기능 호전정도와 상관석 분석 및 다중회귀분석에서 유의했던 프로브들, 나이, 흡연여부, 흡연기간, FEV1% 예측치 등의 주요 임상정보를 이용하여 각 약물의 반응성을 예측하는 모델을 시험적으로 구축하였다.

[0066] LAMA 약물 처치 전의 샘플을 대상으로 폐기능 호전정도와 연관이 있었던 1047개 유의한 프로브의 전체 샘플 중 fail count(Detection_Pval≥0.05)가 10% 미만이면서 상관계수의 절대값이 높은 순으로 상위 10개의 프로브를 선정하였다.

[0067] 각 약물 처치 전의 샘플 대상으로 뽑은 후보 프로브들과 임상정보(나이, 흡연여부, 흡연기간, FEV1% 예측치)들을 설명변수로 넣고, FEV1.diff값을 종속변수로 두어 다중선형회귀분석을 진행하여 최적의 모델을 구성하였다. 변수선택 방법은 backward elimination 방법으로 AIC(Akaike Information Criterion)을 기반으로 편차가 높은 변수를 제거하는 방향으로 진행하였다.

[0068] 예측모델의 잔차의 선형성과 정규성 가정을 만족하는지, 영향치가 존재하였는지를 모형진단을 진행하였고, 각 모델별 잔차의 정규성은 만족하나 영향치(outlier)의 경계에 존재하는 샘플을 포함하고 있다.

[0070] LAMA 샘플 set의 폐기능호전 정도를 예측하는 모델은 도 15와 같다. 해당 모델의 설명력은 81%이며, 입력변수로 FEV1 % 예측값과 5개의 유의한 프로브가 설명변수로 선택되었다. 표 1은 입력변수로 선택되어진 5개 프로브에

대한 유전자리스트를 정리한 것이다.

표 1

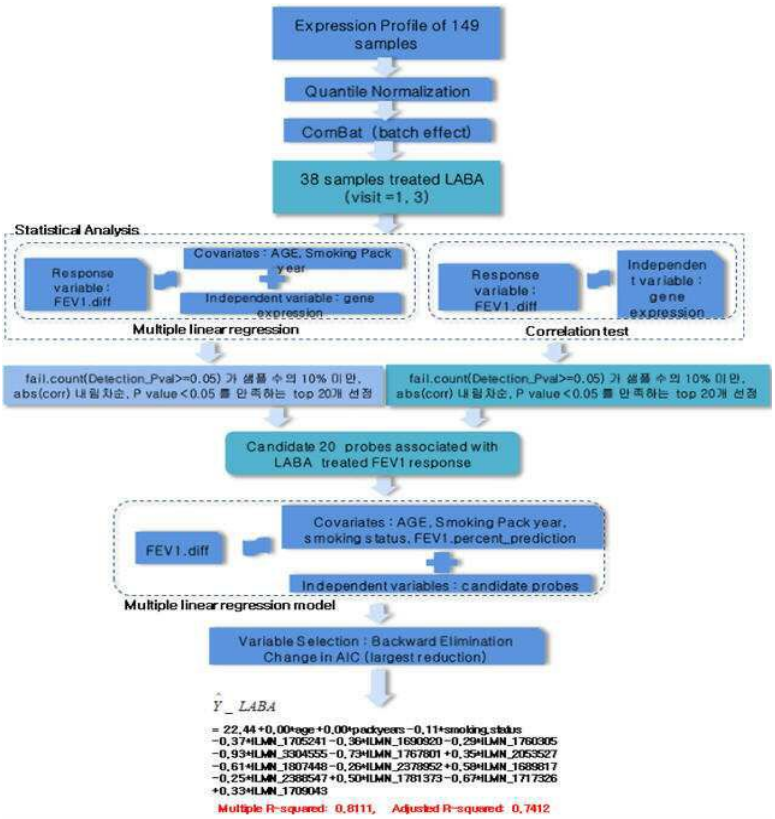
[0072]

ProbeID	Gene	ENTREZ GENE_ID	CHR	PROBE_SEQUENCE
ILMN_3210917	LOC389168	389168		CGCTCTGCTAGAAGCTGGAGGGTGCTAGGCCCCTGTGTAGTAGTGCATAG (서열번호 1)
ILMN_1752027	UBE3B	89910	12	TCCCATCCTAATTTTTATCACCTGAAGGTTGGAACCAGTGAGGGACTGGG (서열번호 2)
ILMN_1779492	SMURF1	57154	7	CCGGGGCGACCGGAAAAGGGCTCTCTCAAGTTCTGAAAAGAGAATCTGCC (서열번호 3)
ILMN_1797107	SCLT1	132320	4	AAGGCGTCTAAGTCAGGCAGAAGAGAGAGCTGCTTCAGCTTCCCAGCAGC (서열번호 4)
ILMN_1659761	SNX29	92017	16	CCCATGATCGGCTCCCAGTGCACCCCCTTAAGGGTAAGCAGGCCACATAT (서열번호 5)

[0074] 표 1에서 ILMN_3210917 은 인간 miscRNA(LOC389168)이고, ILMN_1752027 은 인간 UBE3B (ubiquitin protein ligase E3B)의 전사체 변이 1(transcript variant 1)로서 mRNA이며, ILMN_1779492 은 인간 SMURF1 (SMAD specifi E3 ubiquitin protein ligase 1)의 전사체 변이 1(transcript variant 1)로서 mRNA이고, ILMN_1797107 은 인간 SCLT1 (sodium channel and clathrin linker 1)로서 mRNA이며, ILMN_1659761 은 인간 SNX29 (sorting nexin 29)로서 mRNA 이다.

도면

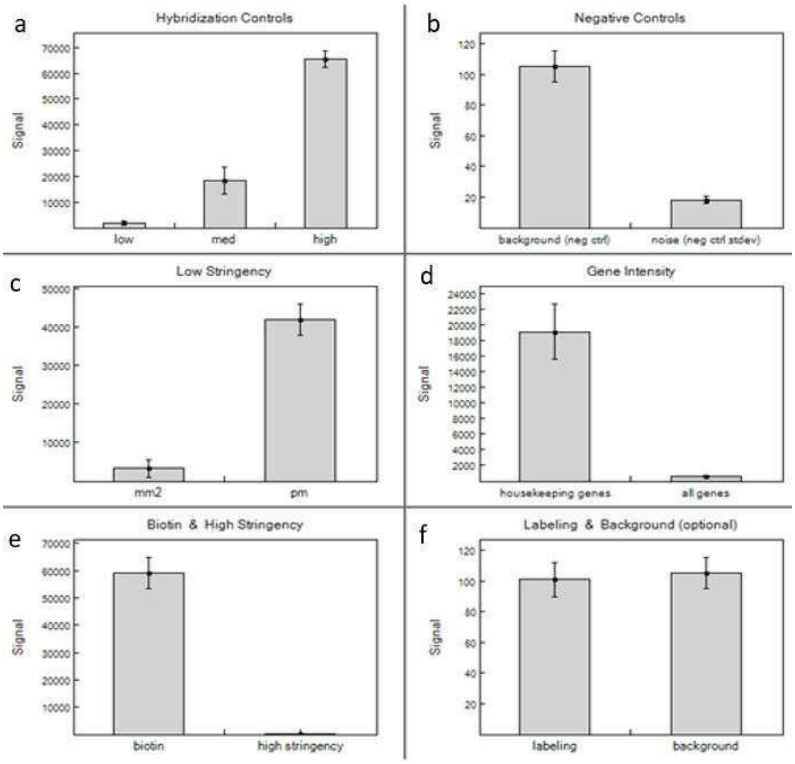
도면1



도면2

실험날짜	분석 샘플 수
140821	3
141127	14
150116	21
150422	12
150513	58
150528	8
150622	16
150702	17
합계	149

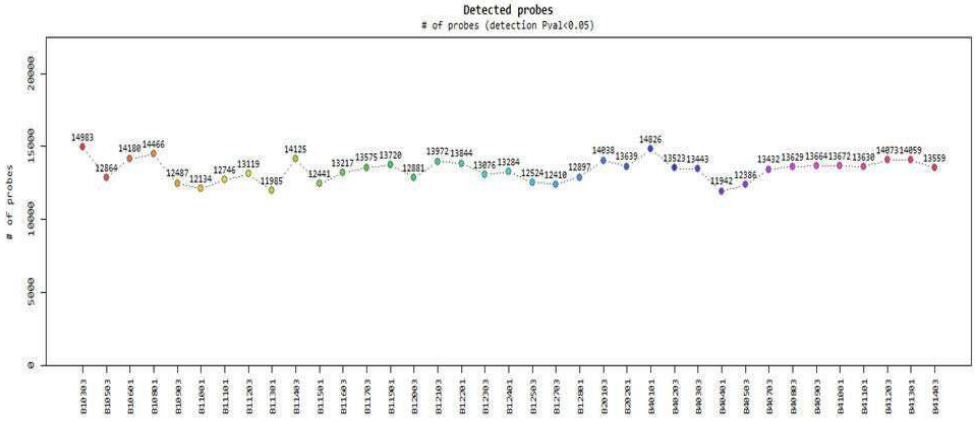
도면3



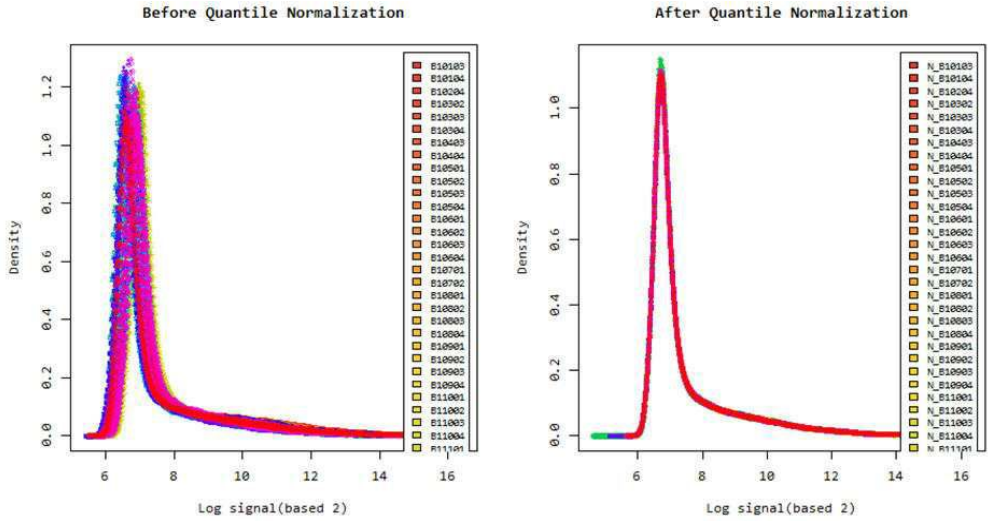
도면4

	Internal control	Expected value
1	Hybridization Controls	High > Medium > Low
2	Low Stringency	PM > MM2
3	Biotin and High Stringency	Biotin is high
4	Negative Controls (Background and Noise)	Low
5	Gene Intensity (House-Keeping and All Genes)	Higher than Background (Housekeeping > All Genes)
6	Labeling and Background	mRNA 실험에는 해당되지 않음

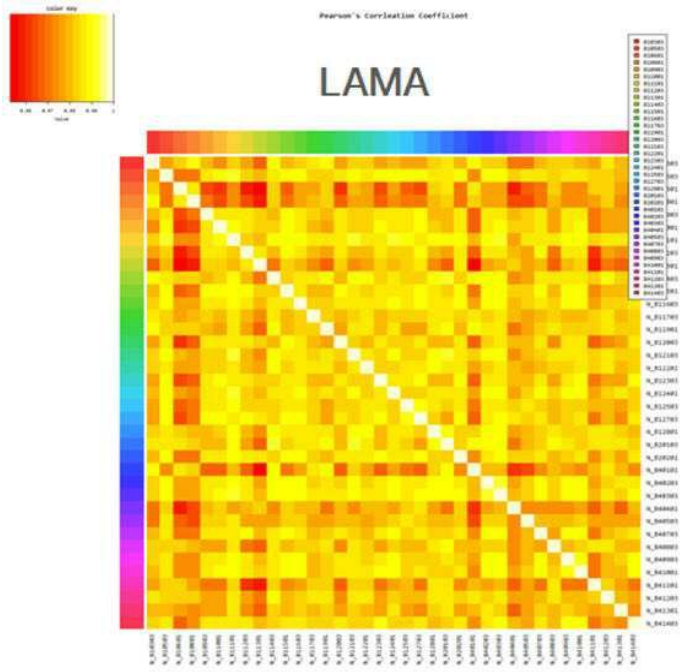
도면5


Detected probes
# of probes (detection Pval<0.05)
# of probes

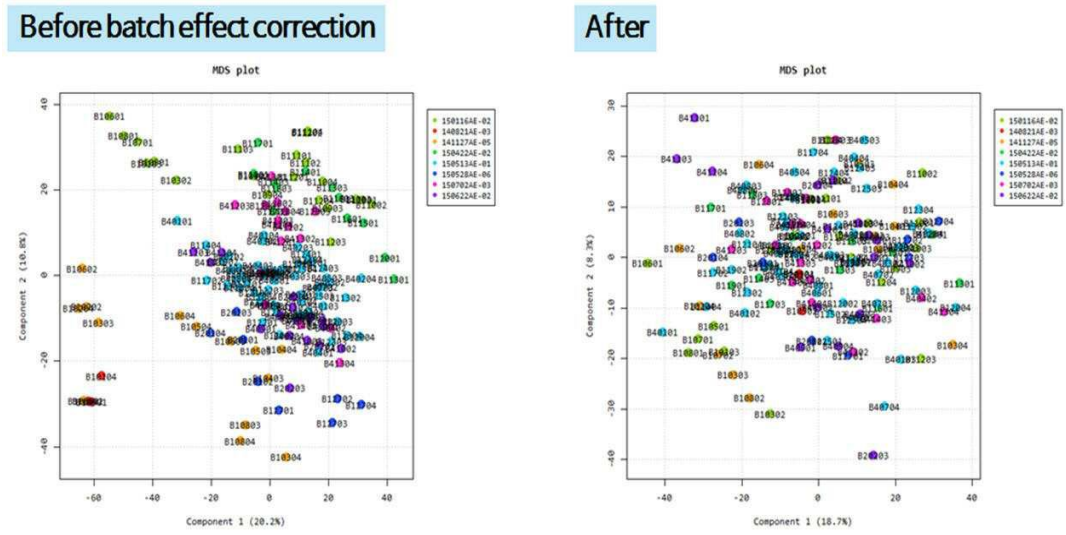
도면6


Before Quantile Normalization
After Quantile Normalization
Density
Log signal(based 2)

도면7


Pearson's Correlation Coefficient
LAMA

도면8

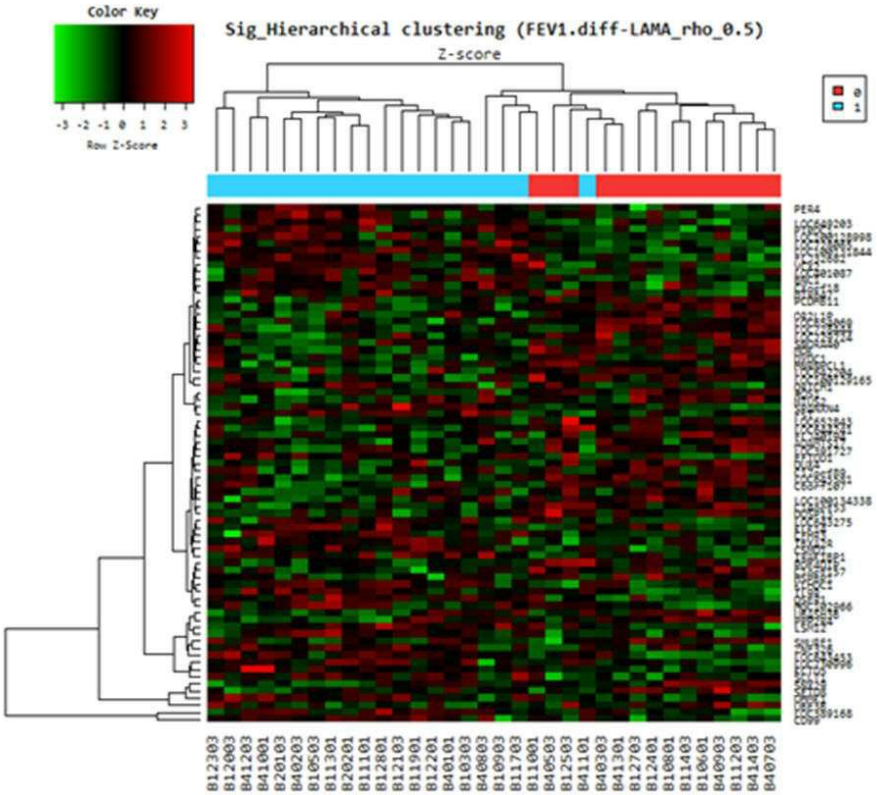

Before batch effect correction
After
MDS plot
MDS plot
Component 1 (20.2%)
Component 2 (10.0%)
Component 1 (18.7%)
Component 2 (8.3%)

도면9

LAMA (34)

Sample.ID	Sentrix.Barcode	Sample.Section	수주번호	연구대상자	방문회수
B10303	3999992036	C	141127AE-05	103	3
B10503	3999992036	G	141127AE-05	105	3
B11001	10011068004	E	150116AE-02	110	1
B11101	10011068016	A	150116AE-02	111	1
B11203	10011068016	G	150116AE-02	112	3
B10903	10011068016	J	150116AE-02	109	3
B10601	10011068022	E	150116AE-02	106	1
B10801	10011068022	G	150116AE-02	108	1
B11301	3998797011	A	150422AE-02	113	1
B11403	3998797011	D	150422AE-02	114	3
B11703	3998797011	I	150422AE-02	117	3
B11901	3998797011	K	150422AE-02	119	1
B12401	3998443019	C	150513AE-01	124	1
B12503	3998443019	H	150513AE-01	125	3
B40101	3998443019	J	150513AE-01	401	1
B40203	3998443046	C	150513AE-01	402	3
B40303	3998443046	G	150513AE-01	403	3
B12003	3998443051	L	150513AE-01	120	3
B12103	3998443076	D	150513AE-01	121	3
B12201	3998443076	F	150513AE-01	122	1
B12303	3998443076	K	150513AE-01	123	3
B40503	3998479109	B	150513AE-01	405	3
B40703	3998479109	G	150513AE-01	407	3
B40803	3998479109	K	150513AE-01	408	3
B20103	3998386098	C	150528AE-06	201	3
B12703	3998386098	G	150528AE-06	127	3
B20201	3998386100	A	150622AE-02	202	1
B40903	3998386100	G	150622AE-02	409	3
B41001	3998386100	I	150622AE-02	410	1
B41101	3998386111	A	150622AE-02	411	1
B12801	3998386112	J	150702AE-03	128	1
B41403	3998386115	G	150702AE-03	414	3
B41203	3998386123	B	150702AE-03	412	3
B41301	3998386123	D	150702AE-03	413	1

도면10

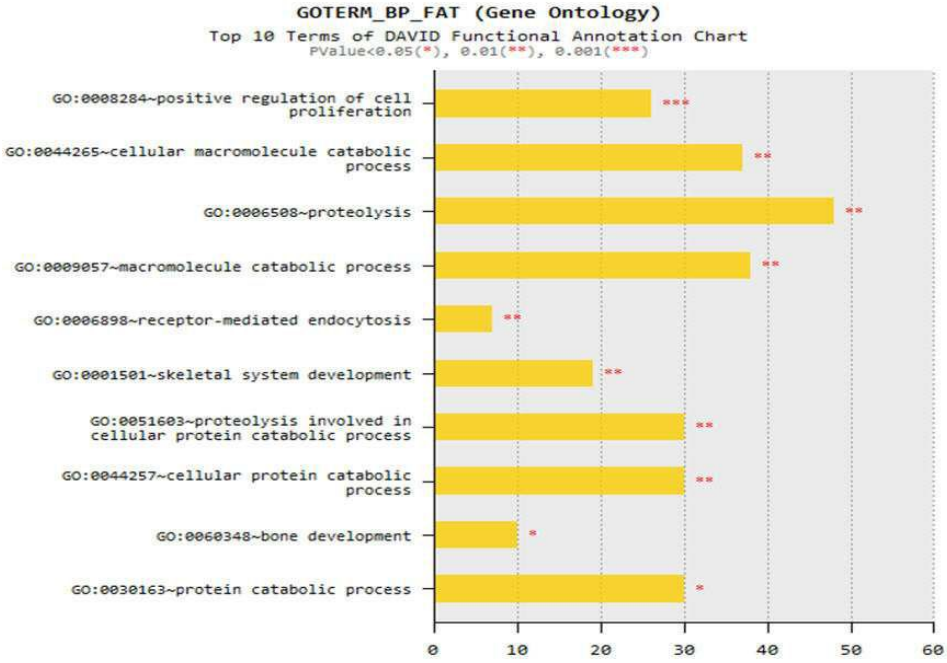


도면11

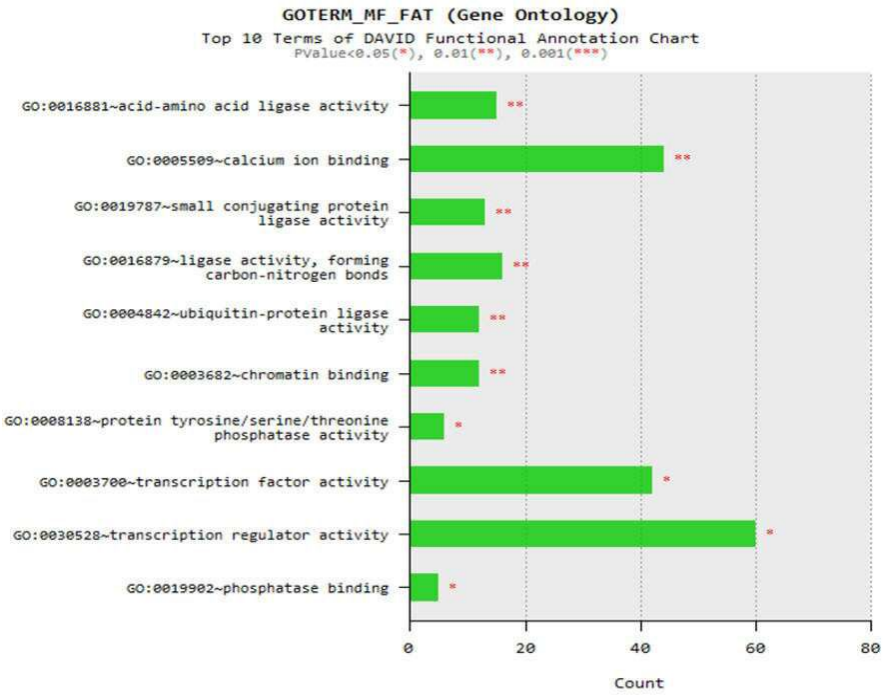
LAMA

Result File	Statistical Test	Significant Result File	# of Sig.probes
data2-FEV1.diff-LAMA.xlsx	lm(FEV1.diff~ExpLv+age+packyears)	data3-FEV1.diff-LAMA.xlsx	1047

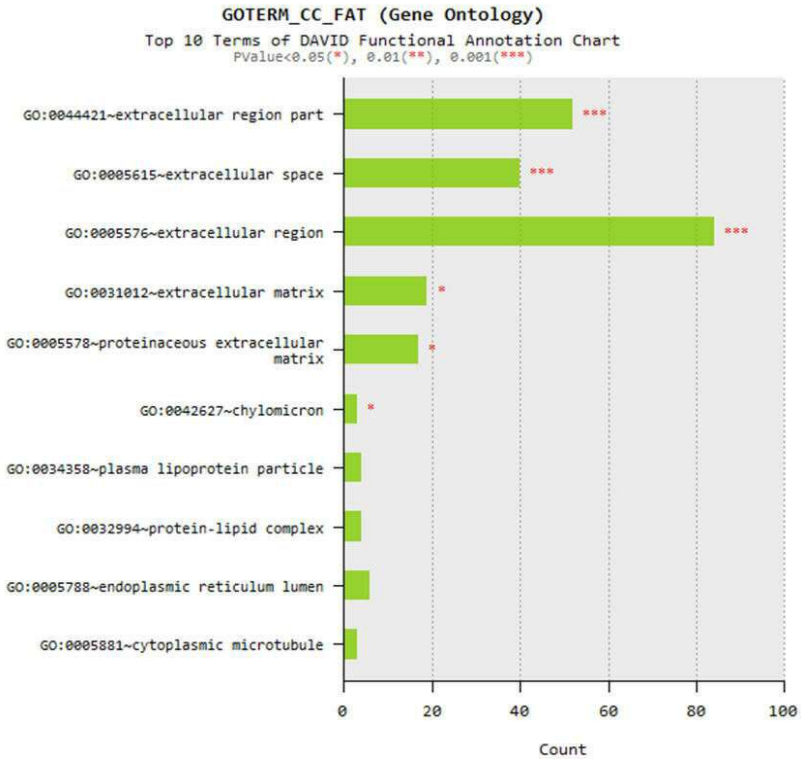
도면12



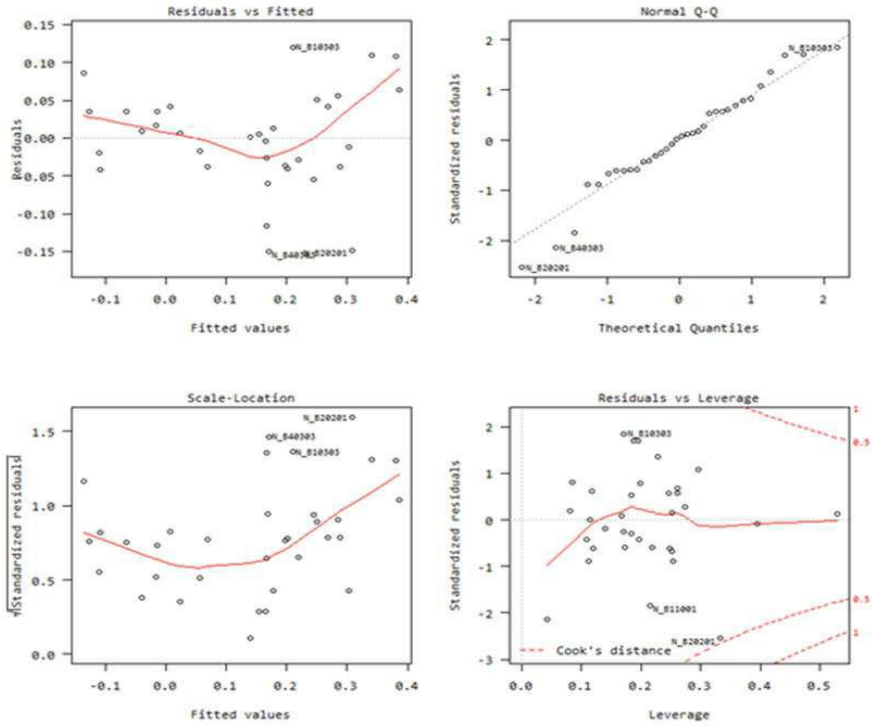
도면13



도면14


GOTERM_CC_FAT (Gene Ontology)
Top 10 Terms of DAVID Functional Annotation Chart
PValue<0.05(*), 0.01(**), 0.001(***)
GO:0044421~extracellular region part
GO:0005615~extracellular space
GO:0005576~extracellular region
GO:0031012~extracellular matrix
GO:0005578~proteinaceous extracellular matrix
GO:0042627~chylomicron
GO:0034358~plasma lipoprotein particle
GO:0032994~protein-lipid complex
GO:0005788~endoplasmic reticulum lumen
GO:0005881~cytoplasmic microtubule
***
***
***
*
*
*
0
20
40
60
80
100
Count

도면15


Residuals vs Fitted
Residuals
Fitted values
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage

도면16

$$\hat{Y}_LAMA = 0.80 - 0.00*FEV1_percent_prediction + 0.24*ILMN_3210917 - 0.21*ILMN_1752027 - 0.22*ILMN_1779492 + 0.30*ILMN_1797107 - 0.24*ILMN_1659761$$

Multiple R-squared: 0.8457, Adjusted R-squared: 0.8114

서 열 목 록

<110> Industry Academic Cooperation Foundation of Ulasn University

Asan Medical Center

<120> Method of predicting for drug-responsibility to long-acting

muscarinic antagonists

<130> PN1603-089

<160> 5

<170> KoPatentIn 3.0

<210> 1

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> artificial sequence

<400> 1

cgctctgcta gaagctggag ggtgctaggc ccctgtgtag tagtgcatag 50

<210> 2

<211> 50

<212> DNA

<

213> Artificial Sequence

<220><223> artificial sequence

<400> 2

tcccatccta atttttatca cctgaaggtt ggaaccagtg agggactggg 50

<210> 3

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> artificial sequence

<400> 3

ccggggcgac cggaaaaggg ctctctcaag ttctgaaaag agaatctgcc 50

<210> 4

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> artificial sequence

<400> 4

aaggcgtcta agtcaggcag aagagagagc tgcttcagct tcccagcagc 50

<210> 5

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> artificial sequence

<400> 5

cgctctgcta gaagctggag ggtgctaggc ccctgtgtag tagtgcatag 50