



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년01월07일
(11) 등록번호 10-2200308
(24) 등록일자 2021년01월04일

(51) 국제특허분류(Int. Cl.)
C12Q 1/6883 (2018.01) G01N 33/563 (2006.01)
(52) CPC특허분류
C12Q 1/6883 (2018.05)
G01N 33/563 (2019.05)
(21) 출원번호 10-2020-0081048
(22) 출원일자 2020년07월01일
심사청구일자 2020년07월01일
(56) 선행기술조사문헌
Journal of Maternal-Fetal and Neonatal
Medicine (2017) 30(13):1588-1595
(뒷면에 계속)

(73) 특허권자
이화여자대학교 산학협력단
서울특별시 서대문구 이화여대길 52 (대현동, 이
화여자대학교)
(72) 발명자
김영주
서울특별시 양천구 목동서로 38, 120동 902호(목
동, 목동신시가지아파트1단지)
유영아
서울특별시 광진구 독성로 510, 401호(자양동)
(뒷면에 계속)
(74) 대리인
유미특허법인

전체 청구항 수 : 총 19 항

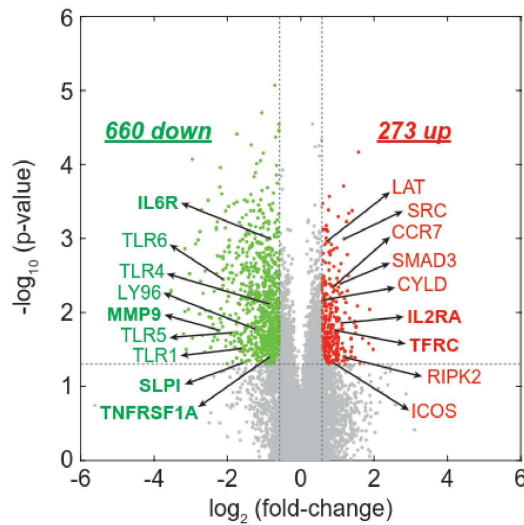
심사관 : 김승범

(54) 발명의 명칭 조산 예측용 조성물 및 이를 이용한 조산 예측 방법

(57) 요약

본 발명은 조산을 예측하여 조산의 예방, 지연, 또는 치료 등을 통해 조산으로 인한 산모 또는 태아의 위험성을 줄이기 위한 것으로, 더욱 자세하게는 조산 위험성을 예측하기 위한 마커를 이용한 조산 예측용 조성물, 조산 예측용 키트, 또는 조산 예측방법 등을 제공하는 것이다.

대표도 - 도1



(52) CPC특허분류

C12Q 2600/158 (2013.01)

G01N 2800/368 (2013.01)

(72) 발명자

안사리 **아부자**

경기도 고양시 일산동구 위시티로 51-14, A동 102호(식사동)

김수민

인천광역시 계양구 계산새로33번길 26, 103동 408호(계산동, 태산아파트)

박선화

서울특별시 양천구 목동서로 155, 105동 1902호(목동)

(56) 선행기술조사문헌

심지은 등, "OB16 : Identification of biomarkers for predicting preterm birth using maternal circulating RNA transcripts based on RNA sequencing", 2019 대한산부인과학회 제105차 학술대회 (2019.09.20.)

KR102099813 B1*

W02020088596 A1

KR101872457 B1

JPW02016024627 A1

JP2019518225 A

이화여자대학교, '조산과 태아손상 조기 진단용 바이오마커 및 맞춤형 조산방지 약물 치료법 개발', 질환극복기술개발사업 최종보고서 (2019.07.발행, 2020.05.16. NDSL DB 구축)*

*는 심사관에 의하여 인용된 문헌

이 발명을 지원한 국가연구개발사업

과제고유번호 1465030172

과제번호 HI18C0378010020

부처명 보건복지부

과제관리(전문)기관명 한국보건산업진흥원

연구사업명 공익적질병극복연구지원사업(R&D)

연구과제명 미생물군유전체와 대사체를 이용한 조산 예측 바이오마커 개발

기 여 율 1/1

과제수행기관명 이화여자대학교 산학협력단

연구기간 2020.01.01 ~ 2020.12.31

공지예외적용 : 있음

명세서

청구범위

청구항 1

생물학적 시료 내에서 CYLD 유전자의 발현량을 측정할 수 있는 제제를 포함하는, 조산 예측용 조성물.

청구항 2

제1항에 있어서, 상기 생물학적 시료는 혈액인, 조성물.

청구항 3

제1항에 있어서, 상기 생물학적 시료는 전혈, 혈장, 및 혈청으로 이루어지는 군에서 선택된 1종 이상인, 조성물.

청구항 4

제1항에 있어서, 상기 생물학적 시료는 산모로부터 분리된 것인, 조성물.

청구항 5

제1항에 있어서, 상기 제제는 상기 유전자의 mRNA의 양을 측정하는 제제, 상기 유전자로부터 발현된 단백질의 양을 측정하는 제제, 또는 이들 둘 모두인, 조성물.

청구항 6

제5항에 있어서, 상기 유전자의 mRNA의 양을 측정하는 제제는, 상기 유전자의 mRNA 또는 cDNA에 특이적으로 결합하는 프라이머 또는 프로브를 포함하는 것인, 조성물.

청구항 7

제5항에 있어서, 상기 유전자로부터 발현된 단백질의 양을 측정하는 제제는, 상기 유전자에 의해 코딩되는 단백질에 특이적으로 결합하는 항체 또는 압타머인 것인, 조성물.

청구항 8

삭제

청구항 9

제1항에 있어서, 상기 조성물은 RIPK2, TRAF1, CCR7, GNA2, GNA15, GRK3, LY96, TLR1, TLR4, TLR5, TLR6, TLR10, LRRK2, RIPK, TNFRSF1A, TNFRSF10C, MAP3K5, MAP2K6, RPS6KA1, RPS6KA5, FCGR2A, PLD1, VAV3, RAC2, WAX, ITGA1, ITGA4, ITGB4, 및 IGF1R로 이루어지는 군에서 선택된 1종 이상의 발현량을 측정할 수 있는 제제를 추가로 포함하는 것인, 조성물.

청구항 10

삭제

청구항 11

삭제

청구항 12

제1항에 있어서, 상기 제제는 CYLD 유전자의 발현량 증가를 측정할 수 있는 것을 포함하는 것인, 조성물.

청구항 13

제1항 내지 제7항, 제9항, 및 제12항 중 어느 한 항에 따른 조성물을 포함하는, 조산 예측용 키트.

청구항 14

분리된 생물학적 시료 내에서 면역 관련 유전자로서 CYLD 유전자의 발현량을 측정하는 단계를 포함하는, 조산 예측을 위한 정보 제공 방법.

청구항 15

제14항에 있어서, 상기 발현량을 측정하는 단계는 상기 유전자의 mRNA 수준을 측정하는 것인, 방법.

청구항 16

제14항에 있어서, 상기 측정된 발현량을 정상 산모와 비교하는 단계를 추가로 포함하는, 방법.

청구항 17

삭제

청구항 18

제14항에 있어서, 상기 면역 관련 유전자는 RIPK2, TFRC, CD4, CD8a, CD8b, CD3d, CD3g, CD3e, LAT, NECTIN2, RAB1B, CD40, ICOS, TRAF1, CCR7, GNA2, GNA15, GRK3, LY96, TLR1, TLR4, TLR5, TLR6, TLR10, LRRK2, RIPK, TNFRSF1A, TNFRSF10C, MAP3K5, MAP2K6, RPS6KA1, RPS6KA5, FCGR2A, PLD1, VAV3, RAC2, WAX, ITGA1, ITGA4, ITGB4, 및 IGF1R로 이루어지는 군에서 선택된 1종 이상을 추가로 포함하는 것인, 방법.

청구항 19

삭제

청구항 20

삭제

청구항 21

제16항에 있어서, 상기 비교된 발현량이 정상 산모보다 높을 경우, 조산 위험성이 있는 것으로 예측하는 단계를 추가로 포함하는, 방법.

청구항 22

제21항에 있어서, 상기 면역 관련 유전자는 RIPK2, TFRC, CD4, CD8a, CD8b, CD3d, CD3g, CD3e, LAT, NECTIN2, RAB1B, CD40, ICOS, TRAF1, CCR7, GNA2, GNA15, 및 GRK3 로 이루어지는 군에서 선택된 1종 이상을 추가로 포함하는 것인, 방법.

청구항 23

삭제

청구항 24

제16항에 있어서, 상기 면역 관련 유전자는 LY96, TLR1, TLR4, TLR5, TLR6, TLR10, LRRK2, RIPK, TNFRSF1A, TNFRSF10C, MAP3K5, MAP2K6, RPS6KA1, RPS6KA5, FCGR2A, PLD1, VAV3, RAC2, WAX, ITGA1, ITGA4, ITGB4, 및 IGF1R 로 이루어지는 군에서 선택된 1종 이상을 추가로 포함하는 것이고,

상기 LY96, TLR1, TLR4, TLR5, TLR6, TLR10, LRRK2, RIPK, TNFRSF1A, TNFRSF10C, MAP3K5, MAP2K6, RPS6KA1, RPS6KA5, FCGR2A, PLD1, VAV3, RAC2, WAX, ITGA1, ITGA4, ITGB4, 및 IGF1R 로 이루어지는 군에서 선택된 1종 이상의 비교된 발현량이 정상 산모보다 낮을 경우, 조산 위험성이 있는 것으로 예측하는 단계를 추가로 포함하는 것인, 방법.

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

시험 대상의 생물학적 시료를 입력받는 입력부;

상기 시료 중 CYLD 유전자의 mRNA 발현량을 정량하는 정량부;

상기 발현량을 정상 산모의 수치값과 비교하는 연산부; 및

상기 비교 결과에 따라 상기 시험 대상의 조산 위험 여부를 나타내는 결과를 표시하는 표시부를 포함하는, 조산 예측용 시스템.

청구항 29

제28항에 있어서, 상기 표시부는 상기 연산부에서 비교된 상기 CYLD 유전자의 mRNA 발현량이 정상 산모의 수치값보다 높을 경우, 상기 시험 대상이 조산 위험성이 있는 것으로 표시하는 것인, 시스템.

발명의 설명

기술 분야

[0001] 본 발명은 조산 예측용 조성물, 조산 예측용 키트, 조산 예측 방법, 또는 조산 예측을 위한 정보 제공 방법에 관한 것이다.

배경 기술

[0002] 조산 (Preterm birth; PTB)은 37주 미만에서 출산하는 것을 말하는데, 최근 아시아 지역 107개 국의 통계에 의하면 약 10.6%의 빈도로 발생하고 있다. 우리나라의 경우, 2006년에는 조산율이 약 4.8%로 나타났으나, 2016년에는 7.2%로 약 2배 가까이 증가하여, 조산율이 증가하고 있는 추세이다. 조산은 단기적으로는 신생아 및 5세 이하 아동의 사망률이 증가하는 위험성이 있고, 신경발달 등의 문제가 발생할 확률이 높아지며, 장기적으로는 2형 당뇨, 비만, 고혈압 등의 유병율이 증가하는 위험성이 있다.

[0003] 조산의 발생기전은 아직 명확히 밝혀져 있지 않으며, 자연적 진통과 흔히 연관된 위험인자는 생식기계 감염, 다태 임신, 2,3 삼분기 출혈, 이전의 조산기왕력 등이 있다. 또한, 조산의 약 75%는 조기진통과 양막 파수, 그리고 이와 관련된 자궁경관 무력증과 융모막염 등을 수반하여 발생하게 된다.

[0004] 조산아의 생존율을 향상시키기 위한 최소한의 임신 주수는 27주, 출생체중은 0.9 kg이고, 출생 신생아의 이환과 사망은 출생체중보다는 일차적으로 임신 주수, 즉 성숙도에 영향을 받는 것으로 알려져 있다. 따라서 이른 주수에 조산의 징후가 왔을 때, 적절한 치료를 통하여 분만 주수를 지연시켜 신생아를 성숙도를 높이는 것은 산모와 신생아의 건강과 삶의 질과 비용에 중대한 관건이 된다.

[0005] 이처럼 조산은 건강과 삶에 중요한 영향을 미치고 있으나, 아직 조산의 정확한 진단이나 예측을 위한 특이성 있는 방법이 없는 실정이다.

발명의 내용

해결하려는 과제

[0006] 본 발명의 일 예는 비침습적인 방법을 이용하여 조산의 예측 방법을 제공하고자 한다. 구체적으로, 본 발명은 모체의 혈액에서 얻어진 RNA의 발현 양상 차이를 토대로 조산과 상관관계를 가질 것으로 예측되는 유전자를 기능에 따라 분류하여 후보유전자를 발굴하였고, 이 결과를 토대로 조산군과 정상분만군의 혈액에서 후보 유전자의 발현량 비교를 통하여 임상적 유의성이 검증된 유전자를 이용하여 조산을 예측할 수 있는 방법을 제공하고자 한다. 구체적인 내용을 살펴보면, 정상군과 조산군에서 발현의 유의한 차이를 보인 유전자들은 림프구 활성화(lymphocyte activation), 세포부착(cell adhesion) 그리고 혈관신생(angiogenesis)의 발현경로와 관련이 있었고, 이는 T 세포의 기능과 관련이 있는 것으로 조사되었다. 조산 환자 100명의 혈액 성분을 분석한 결과 림프구(Lymphocyte)의 수치가 유의하게 증가되었으며, 이러한 결과를 토대로 발현 양상이 유의한 유전자 중 T 세포의 기능과 관련된 유전자를 선별하여 조산의 예측 마커로 개발되었다.

과제의 해결 수단

[0007] 본 발명은 정상분만군과 조산군을 대상으로 혈액의 RNA를 추출하고 RNA sequencing (RNAseq)법을 이용하여 두 그룹에서 유의한 발현량의 차이가 있는 유전자를 확인하여 조산 예측을 위한 후보 유전자군을 선별하였다. 또한, 후보유전자 군의 발현경로(pathway)를 확인하고, 임상샘플에서 발현량을 검증하여 최종적인 후보 유전자를 선별하였다.

[0008] 본 발명의 일 예는 생물학적 시료 내에서 면역 관련 유전자의 발현량을 측정할 수 있는 체계를 포함하는, 조산 예측용 조성물 또는 조산 진단용 조성물에 관한 것이다.

[0009] 본 발명의 또 다른 일 예는, 본 발명에 따른 조산 예측용 조성물을 포함하는, 조산 예측용 키트 또는 조산 진단용 키트에 관한 것이다.

[0010] 본 발명의 또 다른 일 예는, 생물학적 시료 내에서 면역 관련 유전자의 발현량을 측정하는 단계를 포함하는, 조산 예측 방법, 조산 예측을 위한 정보 제공 방법, 조산 진단 방법, 또는 조산 진단을 위한 정보 제공 방법에 관한 것이다.

[0011] 본 발명의 또 다른 일 예는, (1) 발현량이 증가된 마커 유전자에 대한 안티센스 폴리뉴클레오타이드 서열, siRNA, 또는 이들 둘 모두; (2) 발현량이 감소된 마커 유전자의 코딩 서열 및 상기 서열의 전사를 제어하는 단리된 핵산; 또는 (3) 상기 (1) 및 (2) 를 포함하는, 조산의 조산의 예방, 지연, 치료, 또는 위험성 감소용 약학적 조성물에 관한 것이다.

[0012] 본 발명의 또 다른 일 예는, 시험 대상 산모의 생물학적 시료 내에서 T-세포의 수를 측정하는 단계를 포함하는, 조산 예측 방법, 조산 예측을 위한 정보 제공 방법, 조산 진단 방법, 또는 조산 진단을 위한 정보 제공 방법에 관한 것이다. 상기 방법은, 상기 측정하는 단계 이후에, 상기 측정된 T-세포의 수가 정상 산모보다 높을 경우, 상기 시험 대상 산모는 조산의 위험성을 가지는 것으로 예측하는 단계를 추가로 포함할 수 있다.

[0014] 이하, 본 발명을 더욱 자세히 설명하고자 한다.

[0015] 본 명세서에서 용어 "조산", "premature birth", "preterm birth", "PTB"등은 동일한 의미를 갖고, 37주 미만에서 태아의 출산을 의미한다.

[0016] 본 명세서에서 용어 "혈액"은 시험 대상, 또는 시험 개체, 예를 들어 산모로부터 분리 또는 채취된 혈액 시료를 의미한다. 상기 혈액은 전혈, 또는 혈액 세포등을 포함할 수 있으며, 예를 들어 혈장 또는 혈청을 포함할 수 있으나, 이에 제한되는 것은 아니다.

[0017] 본 명세서에서 용어 "또는"은 명확하게 내용이 특정되지 않는 한, "및/또는"을 포함하는 의미로 사용된다.

[0018] 본 명세서에서 용어 "유전자(gene)"는 폴리펩티드 생산과 관련된 DNA 부분이며, 엑손, 인트론과 더불어 유전자 산물의 전사 또는 번역, 전사 또는 번역의 조절과 관련된 코딩 영역 전후의 영역을 포함할 수 있다. 본 명세서에서 유전자가 그의 이름으로 특정될 때 (예를 들면, 도 1, 도 2b, 도 2c, 도 3 등에 나타난 유전자 이름), 유전자는 해당 유전자의 자연적으로 발생하는 변이체 또는 돌연변이체를 포함할 수 있다. 예를 들어, CYLD (CYLD lysine 63 deubiquitinase) 유전자의 서열은 GenBank Accession No. NG_012061.1 혹은 AJ250014.1, TFRC (transferrin receptor) 유전자의 서열은 GenBank Accession No. NG_046395.1 혹은 NM_003234.4, RIPK2 (receptor interacting serine/threonine kinase 2) 유전자의 서열은 GenBank Accession No. NG_033016.2 혹은 NM_003821.6 에 기재되어 있으며, 본 명세서에서 "CYLD 유전자", "TFRC 유전자", 또는 "RIPK2 유전자" 등은 GenBank에 나타난 cDNA와 적어도 80% 이상, 85% 이상, 90% 이상, 95% 이상, 96% 이상, 97% 이상, 98% 이

상, 99% 이상, 99.5% 이상, 또는 99.9% 이상의 상동성을 갖는 변이체를 포함한다.

- [0019] 본 명세서에서 용어 "증가", 또는 "감소"는 비교 대조군, 예를 들면, 확립된 표준 대조군 (예를 들면, 정상 분만을 한 산모의 생물학적 시료에서 나타난 마커 유전자의 mRNA의 평균 수준)으로부터 양(quantity)의 변화를 의미한다.
- [0020] 예를 들어, 상기 증가는 대조군 값의 10% 이상, 15% 이상, 20% 이상, 25% 이상, 30% 이상, 35% 이상, 40% 이상, 45% 이상, 또는 50% 이상의 발현량의 증가, 또는 대조군 값의 1.5배 이상, 2배 이상, 2.5배 이상, 3배 이상, 3.5배 이상, 4배 이상, 또는 5배 이상의 발현량 증가일 수 있으나, 통상의 기술자라면 절대적인 증가량이 특정되지 않더라도, 대조군 값 대비 유의하게 증가된 변화를 명확하게 이해할 수 있을 것이다.
- [0021] 예를 들어, 상기 감소는 대조군 값의 90% 이하, 85% 이하, 80% 이하, 75% 이하, 70% 이하, 65% 이하, 60% 이하, 55% 이하, 또는 50% 이하의 발현량의 감소, 또는 대조군 값의 0.5배 이하, 0.4배 이하, 0.3배 이하, 0.2배 이하, 또는 0.1배 이하의 발현량 감소일 수 있으나, 통상의 기술자라면 절대적인 감소량이 특정되지 않더라도, 대조군 값 대비 유의하게 감소된 변화를 명확하게 이해할 수 있을 것이다.
- [0022] 상기 증가 또는 감소를 나타내는 용어, 예를 들어 “더”, “더 높은”, “유의하게 높은”, “통계적으로 유의하게 높은”, “덜”, “더 낮은”, “유의하게 낮은”, “통계적으로 유의하게 높은” 등의 용어는 상기 기술된 바와 같이 동일한 의미로 사용된다. 대조적으로, 용어 “유의하지 않게”, “동등하게”, “실질적으로 동일”, “실질적으로 변화 없는” 등은 표준 대조군의 $\pm 10\%$ 이내, $\pm 5\%$ 이내, $\pm 3\%$ 이내, 또는 $\pm 2\%$ 이내 등, 표준 대조군 값으로부터 변화가 거의 없거나 유의하지 않은 변화를 보이는 것을 의미한다.
- [0023] 본 발명의 일 예는 생물학적 시료 내에서 면역 관련 유전자의 발현량을 측정할 수 있는 체제를 포함하는, 조산 예측용 조성물 또는 조산 진단용 조성물에 관한 것이다. 상기 생물학적 시료는 혈액일 수 있으며, 예를 들어 전혈, 혈장, 및 혈청으로 이루어지는 군에서 선택된 1종 이상인 것일 수 있다. 상기 생물학적 시료는 산모로부터 분리된 것일 수 있으며, 예를 들어 조산이 의심되는 산모로부터 분리된 것일 수 있다.
- [0024] 상기 면역 관련 유전자는 T-세포 활성화 관련 유전자, 후천성 면역 반응 (adaptive immune response) 관련 유전자, 및 선천성 면역 반응 (innate immune response) 관련 유전자로 이루어지는 군에서 선택된 1종 이상인 것일 수 있다. 예를 들어, 상기 면역 관련 유전자는 CYLD, TFRC, RIPK2, CD4, CD8a, CD8b, CD3d, CD3g, CD3e, LAT, NECTIN2, RAB1B, CD40, ICOS, TRAF1, CCR7, GNA2, GNA15, GRK3, LY96, TLR1, TLR4, TLR5, TLR6, TLR10, LRRK2, RIPK, TNFRSF1A, TNFRSF10C, MAP3K5, MAP2K6, RPS6KA1, RPS6KA5, FCGR2A, PLD1, VAV3, RAC2, WAX, ITGA1, ITGA4, ITGB4, 및 IGF1R 로 이루어지는 군에서 선택된 1종 이상인 것일 수 있다.
- [0025] 상기 체제는 유전자의 mRNA의 양을 측정하는 체제, 상기 유전자로부터 발현된 단백질의 양을 측정하는 체제, 또는 이들 둘 모두인 것일 수 있다. 상기 유전자의 mRNA의 양을 측정하는 체제는, 상기 유전자의 mRNA 또는 cDNA에 특이적으로 결합하는 프라이머, 프로브, 또는 이들의 조합을 포함하는 것일 수 있다. 상기 유전자로부터 발현된 단백질의 양을 측정하는 체제는, 상기 유전자에 의해 코딩되는 단백질에 특이적으로 결합하는 항체, 압타머, 또는 이들의 조합을 포함하는 것일 수 있다.
- [0026] 상기 체제는 CYLD, TFRC, RIPK2, CD4, CD8a, CD8b, CD3d, CD3g, CD3e, LAT, NECTIN2, RAB1B, CD40, ICOS, TRAF1, CCR7, GNA2, GNA15, 및 GRK3 로 이루어지는 군에서 선택된 1종 이상의 발현량 증가를 측정할 수 있는 것을 포함하거나, LY96, TLR1, TLR4, TLR5, TLR6, TLR10, LRRK2, RIPK, TNFRSF1A, TNFRSF10C, MAP3K5, MAP2K6, RPS6KA1, RPS6KA5, FCGR2A, PLD1, VAV3, RAC2, WAX, ITGA1, ITGA4, ITGB4, 및 IGF1R 로 이루어지는 군에서 선택된 1종 이상의 발현량 감소를 측정할 수 있는 것을 포함하거나, 또는 이들 둘 모두를 포함하는 것일 수 있다.
- [0027] 본 발명의 또 다른 일 예는, 본 발명에 따른 조산 예측용 조성물을 포함하는, 조산 예측용 키트 또는 조산 진단용 키트에 관한 것이다. 상기 키트는 본 발명에 따른 조산 예측의 마커 유전자 (예를 들어, CYLD, TFRC, 및/또는 RIPK2)의 mRNA 수준을 평가하여 조산을 예측하는 것일 수 있다. 상기 mRNA 수준을 평가하기 위한 키트는 마커 유전자 서열 또는 마커 유전자가 코딩하는 하나 이상의 세그먼트와 혼성화 가능한 올리고뉴클레오타이드 프로브를 포함할 수 있다. 상기 올리고뉴클레오타이드 프로브는 검출 가능한 모이어티로 표지될 수 있다. 상기 키트는 PCR, 특히 RT-PCR에 의한 마커 유전자 DNA 또는 mRNA의 하나 이상의 세그먼트 증폭에 사용될 수 있는 두 개 이상의 올리고뉴클레오타이드 프로브를 포함할 수 있다. 예를 들어, 상기 키트는 상기 프로브 및/또는 프라이머의 복수의 세트를 포함할 수 있으며, 하나 이상의 마커 유전자 mRNA가 시험되고 정량화될 수 있다. 또한, 상기 키트는 기준 유전자의 mRNA 수준을 정량하기 위해 사용될 수 있는 상기 프로브 및/또는 프라이머를 추가로 포함할 수 있다. 상기 키트는 표준 대조군을 포함할 수 있으며, 상기 표준 대조군은 생물학적 시료 내 하나 이상의 마

커 유전자의 mRNA의 평균값을 나타낸다.

- [0028] 본 명세서에서 용어 "표준 대조군(standard control)"은 조산이 아닌 정상 분만을 한 건강한 임신 여성에서 분리된 생물학적 시료 중에 존재하는, 본 발명의 일 예에 따른 조산에 대한 마커 유전자의 mRNA 양 또는 농도를 의미한다. 즉, 상기 표준 대조군은 조산을 하지 않은 "정상 산모" 일 수 있다. 표준 대조군은 시료 내에 존재하는 마커 유전자 mRNA 양을 비교하기 위한 기준이며, 표준 대조군으로서 제공되는 시료는 통상적으로 정의된 바와 같이 정상 임신 기간을 거쳐 태아를 분만한 평균의 건강한 산모의 생물학적 시료 내 마커 mRNA의 평균 양을 제공한다. 표준 대조군 값은 개체의 나이, 임신 기간, 인종 등에 의존하여 다양할 수 있으며, 마커 유전자의 mRNA 수준이 또 다른 기준 유전자의 mRNA 수준에 대해 정상적인 비율을 가지는지 여부 등에 따라 설정될 수 있다. 또한, 조산이 아닌 정상 분만을 한 산모를 정의하기 위해 사용되는 용어인 "평균 (average)" 은, 정상 분만을 한 산모의 생물학적 시료의 특성, 예를 들어 본 발명의 일 예에 따른 마커 유전자의 mRNA 발현량을 의미한다.
- [0029] 상기 "기준 유전자(reference gene)"는 생물학적 시료에서 일관되게 발현되고, 실질적으로 변화가 없는 수준으로 알려진 "하우스키핑(housekeeping)" 유전자일 수 있다. 예를 들어, 혈액 시료에 대한 이러한 "하우스키핑" 유전자의 예는 *GAPDH* (glyceraldehyde 3-phosphate dehydrogenase), *SDHA* (succinate dehydrogenase), *HPRT1* (hypoxanthine phosphoribosyl transferase 1), *HBS1L* (HBS1-like protein), *AHSP* (alpha haemoglobin stabilising protein), *ACTB* (beta-actin), *RNA18S5* (RNA, 18S ribosomal 5), *FCGR3A* (the Fc fragment of IgG, low affinity IIIa, receptor), *FCGR3B* (the Fc fragment of IgG, low affinity IIIa, receptor), *B2M* (beta-2-microglobulin), *HUWE1* (HECT, UBA and WWE domain containing 1, E3 ubiquitin protein ligase), *TPT1* (tumor protein, translationally-controlled 1), *MYL12B* (myosin, light chain 12B, regulatory), 및 *SKP1* (S-phase kinase-associated protein 1) 로 이루어지는 군에서 선택된 1종 이상을 포함할 수 있다.
- [0030] 본 명세서에서 용어 "양(amount)", 또는 "발현량"은 특정 폴리뉴클레오티드, 예를 들면, 생물학적 시료 내에 존재하는 마커 유전자 mRNA의 정량값(quantity)을 의미한다. 이러한 정량값은 절대적인 용어, 예를 들어 시료 중 폴리뉴클레오티드의 총 정량, 또는 상대적인 용어, 예를 들어 마커 유전자 mRNA 수준과 기준 유전자 mRNA 수준 사이의 비율일 수 있으며, 일례로 생물학적 시료 내 폴리뉴클레오티드의 농도로 표현될 수 있다.
- [0031] 본 발명의 또 다른 일 예는, 생물학적 시료 내에서 면역 관련 유전자의 발현량을 측정하는 단계를 포함하는, 조산 예측 방법, 조산 예측을 위한 정보 제공 방법, 조산 진단 방법, 또는 조산 진단을 위한 정보 제공 방법에 관한 것이다. 상기 발현량을 측정하는 단계는 상기 유전자의 mRNA 수준을 측정하는 것일 수 있다. 상기 면역 관련 유전자는 전술한 바와 같다.
- [0032] 상기 조산 예측 방법, 조산 예측을 위한 정보 제공 방법, 조산 진단 방법, 또는 조산 진단을 위한 정보 제공 방법은, 상기 측정된 발현량을 정상 산모 (표준 대조군)와 비교하는 단계를 추가로 포함하는 것일 수 있다. 또한, 상기 비교된 발현량이 정상 산모보다 높거나 낮을 경우, 상기 생물학적 시료가 분리된 산모는 조산 위험성이 있는 것으로 예측하는 단계를 추가로 포함하는 것일 수 있다.
- [0033] 상기 비교된 발현량이 정상 산모보다 높을 때, 상기 유전자는 *CYLD*, *TFRC*, *RIPK2*, *CD4*, *CD8a*, *CD8b*, *CD3d*, *CD3g*, *CD3e*, *LAT*, *NECTIN2*, *RAB1B*, *CD40*, *ICOS*, *TRAF1*, *CCR7*, *GNA2*, *GNA15*, 및 *GRK3* 로 이루어지는 군에서 선택된 1종 이상인 것일 수 있다. 예를 들어, 상기 예측하는 단계는, *CYLD*, *TFRC*, *RIPK2*, *CD4*, *CD8a*, *CD8b*, *CD3d*, *CD3g*, *CD3e*, *LAT*, *NECTIN2*, *RAB1B*, *CD40*, *ICOS*, *TRAF1*, *CCR7*, *GNA2*, *GNA15*, 및 *GRK3* 로 이루어지는 군에서 선택된 1종 이상의 유전자 (예컨대, mRNA)의 발현량이 표준 대조군(정상 산모)보다 높은 경우, 상기 생물학적 시료가 분리된 산모는 조산 위험성이 있는 것으로 예측하는 것일 수 있다.
- [0034] 상기 비교된 발현량이 정상 산모보다 낮을 때, 상기 유전자는 *LY96*, *TLR1*, *TLR4*, *TLR5*, *TLR6*, *TLR10*, *LRRK2*, *RIPK*, *TNFRSF1A*, *TNFRSF10C*, *MAP3K5*, *MAP2K6*, *RPS6KA1*, *RPS6KA5*, *FCGR2A*, *PLD1*, *VAV3*, *RAC2*, *WAX*, *ITGA1*, *ITGA4*, *ITGB4*, 및 *IGF1R* 로 이루어지는 군에서 선택된 1종 이상인 것일 수 있다. 예를 들어, 상기 예측하는 단계는, *LY96*, *TLR1*, *TLR4*, *TLR5*, *TLR6*, *TLR10*, *LRRK2*, *RIPK*, *TNFRSF1A*, *TNFRSF10C*, *MAP3K5*, *MAP2K6*, *RPS6KA1*, *RPS6KA5*, *FCGR2A*, *PLD1*, *VAV3*, *RAC2*, *WAX*, *ITGA1*, *ITGA4*, *ITGB4*, 및 *IGF1R* 로 이루어지는 군에서 선택된 1종 이상의 유전자 (예컨대, mRNA)의 발현량이 표준 대조군(정상 산모)보다 낮은 경우, 상기 생물학적 시료가 분리된 산모는 조산 위험성이 있는 것으로 예측하는 것일 수 있다.
- [0035] 본 발명의 일 예에 따라 조산 위험성이 예측 또는 진단된 산모는, 조산을 예방, 지연, 치료, 또는 조산의 위험

성을 감소하기 위해 적절한 처치가 처리될 수 있으며, 예를 들어 코르티코스테로이드 처방 또는 투여, 자궁수축 억제제 투여, 또는 경피 글리세릴 트리니트레이트 처치 등이 처리될 수 있다.

- [0036] 본 발명의 또 다른 일 예는, (1) 발현량이 증가된 마커 유전자에 대한 안티센스 폴리뉴클레오타이드 서열, siRNA, 또는 이들 둘 모두; (2) 발현량이 감소된 마커 유전자의 코딩 서열 및 상기 서열의 전사를 제어하는 단리된 핵산; 또는 (3) 상기 (1) 및 (2) 를 포함하는, 조산의 조산의 예방, 지연, 치료, 또는 위험성 감소용 약학적 조성물에 관한 것이다.
- [0037] 구체적으로, 조산을 치료 또는 예방하기 위해, 표준 대조군 값으로부터 벗어난 발현량을 보이는 마커 유전자의 mRNA 수준을 인위적으로 조절할 수 있다. 예를 들어, 조산 마커 유전자가 표준 대조군 값보다 증가된 것으로 측정된 경우, 상기 마커 유전자의 mRNA 수준을 감소시키기 위한 처치가 처리될 수 있으며, 일 예로 발현량이 증가된 마커 유전자에 대한 안티센스 폴리뉴클레오타이드 서열, siRNA, 또는 이들 둘 모두가 산모에 투여될 수 있다. 또는, 예를 들어, 조산 마커 유전자가 표준 대조군 값보다 감소된 것으로 측정된 경우, 상기 마커 유전자의 mRNA 수준을 증가시키기 위한 처치가 처리될 수 있으며, 예를 들어 발현량이 감소된 마커 유전자의 코딩 서열을 포함하고, 서열의 전사를 제어하는 단리된 핵산, 일 예로 발현 카세트가 산모에 투여될 수 있다.
- [0038] 상기 발현량이 증가된 마커 유전자는 CYLD, TFRC, RIPK2, CD4, CD8a, CD8b, CD3d, CD3g, CD3e, LAT, NECTIN2, RAB1B, CD40, ICOS, TRAF1, CCR7, GNA2, GNA15, 및 GRK3 로 이루어진 군으로부터 선택된 1종 이상일 수 있다.
- [0039] 상기 발현량이 감소된 마커 유전자는 LY96, TLR1, TLR4, TLR5, TLR6, TLR10, LRRK2, RIPK, TNFRSF1A, TNFRSF10C, MAP3K5, MAP2K6, RPS6KA1, RPS6KA5, FCGR2A, PLD1, VAV3, RAC2, WAX, ITGA1, ITGA4, ITGB4, 및 IGF1R 로 이루어진 군으로부터 선택된 1종 이상일 수 있다.
- [0040] 본 명세서에서 용어 "치료"는 관련된 상태, 증상, 징후, 또는 재발의 완화, 저감, 제거, 예방, 지연, 또는 위험성 감소를 가져오는 행위를 포함할 수 있다.
- [0041] 본 발명의 일 예에 따른 조산의 예방, 지연, 치료, 또는 위험성 감소용 약학적 조성물은 조산이 예측되는 산모에게 투여될 수 있으며, 예를 들어 이전의 임신 중 조산 경험이 있는 산모, 자궁경부의 길이가 짧아진 산모, 자궁경부의 모양 변형이 나타난 산모, 생식계 감염 또는 염증 증상을 보이는 산모, 분만 전 출혈을 보이는 산모, 다태 임신 산모, 또는 임신 37주 전 자궁 수축을 보이는 산모 등에 투여될 수 있다.
- [0042] 본 발명의 또 다른 일 예는, 시험 대상 산모의 생물학적 시료 내에서 면역 세포, 예를 들어 백혈구 또는 림프구 세포의 수를 측정하는 단계를 포함하는, 조산 예측 방법, 조산 예측을 위한 정보 제공 방법, 조산 진단 방법, 또는 조산 진단을 위한 정보 제공 방법에 관한 것이다. 본원 실시예에서 조산 산모에서 면역과 관련된 경로, 예를 들어 T-세포 활성화 경로 또는 T-세포에서의 케모카인 신호전달이 정상 산모 대비 증가된 것을 알 수 있었고, 이는 조산 산모에서 면역 세포, 예를 들어 백혈구, 림프구, 또는 T-세포의 수가 정상 산모 대비 증가하는 것을 시사하였으며, 실제 조산 산모 군에서 분리된 생물학적 시료를 분석해본 결과, 조산 군에서 백혈구 및 림프구 수가 정상 산모 군과 통계적으로 유의한 차이를 보임을 확인하였다.
- [0043] 상기 방법은, 상기 측정하는 단계 이후에, 상기 측정된 면역 세포의 수가 정상 산모보다 높을 경우, 상기 시험 대상 산모는 조산의 위험성을 가지는 것으로 예측하는 단계를 추가로 포함할 수 있다.
- [0044] 본 발명의 또 다른 일 예는, 조산을 예측 또는 진단하기 위한 장치 또는 시스템에 관한 것이다. 구체적으로, 상기 기술된 조산 예측 방법, 조산 예측을 위한 정보 제공 방법, 조산 진단 방법, 또는 조산 진단을 위한 정보 제공 방법 단계의 전부 또는 일부를 수행할 수 있는 장치 또는 시스템에 관한 것이다. 상기 장치 또는 시스템은 본 발명의 일 예에 따른 조산 예측용 조성물, 또는 조산 예측용 키트 등을 사용하여 상기 단계의 전부 또는 일부를 수행하는 것일 수 있다.
- [0045] 예를 들어, 상기 장치 또는 시스템은 시험 대상의 생물학적 시료, 예를 들어 시험 대상 산모에서 분리된 혈액 시료를 얻고, 상기 시료 중 본 발명에 따른 마커 유전자 mRNA 양 또는 농도를 정량하고, 상기 양 또는 농도를 표준 대조군 (정상 산모)의 수치값과 비교하고, 상기 비교 결과에 따라 상기 시험 대상의 조산 위험 존재여부를 나타내는 결과를 표시하는 것일 수 있다. 이에, 상기 장치 또는 시스템은 시험 대상의 생물학적 시료를 입력받는 입력부; 상기 시료 중 본 발명에 따른 마커 유전자 mRNA 양 또는 농도를 정량하는 정량부; 상기 양 또는 농도를 표준 대조군 (정상 산모)의 수치값과 비교하는 연산부; 및 상기 비교 결과에 따라 상기 시험 대상의 조산 위험 존재여부를 나타내는 결과를 표시하는 표시부를 포함하는 것일 수 있다.
- [0046] 상기 마커 유전자는 전술한 바와 같으며, 예를 들어 CYLD, TFRC, RIPK2, CD4, CD8a, CD8b, CD3d, CD3g, CD3e,

LAT, NECTIN2, RAB1B, CD40, ICOS, TRAF1, CCR7, GNA2, GNA15, GRK3, LY96, TLR1, TLR4, TLR5, TLR6, TLR10, LRRK2, RIPK, TNFRSF1A, TNFRSF10C, MAP3K5, MAP2K6, RPS6KA1, RPS6KA5, FCGR2A, PLD1, VAV3, RAC2, WAX, ITGA1, ITGA4, ITGB4, 및 IGF1R로 이루어지는 군에서 선택된 1종 이상인 것일 수 있다.

[0047] 상기 마커 유전자가 CYLD, TFRC, RIPK2, CD4, CD8a, CD8b, CD3d, CD3g, CD3e, LAT, NECTIN2, RAB1B, CD40, ICOS, TRAF1, CCR7, GNA2, GNA15, 및 GRK3 로 이루어진 군으로부터 선택된 1종 이상일 때, 상기 연산부에서 비교된 상기 마커 유전자의 mRNA 발현량이 정상 산모의 수치값보다 높을 경우, 상기 표시부는 상기 시험 대상이 조산 위험성이 있는 것으로 표시하는 것일 수 있다.

[0048] 상기 마커 유전자가 LY96, TLR1, TLR4, TLR5, TLR6, TLR10, LRRK2, RIPK, TNFRSF1A, TNFRSF10C, MAP3K5, MAP2K6, RPS6KA1, RPS6KA5, FCGR2A, PLD1, VAV3, RAC2, WAX, ITGA1, ITGA4, ITGB4, 및 IGF1R 로 이루어진 군 으로부터 선택된 1종 이상일 때, 상기 연산부에서 비교된 상기 마커 유전자의 mRNA 발현량이 정상 산모의 수치 값보다 낮을 경우, 상기 표시부는 상기 시험 대상이 조산 위험성이 있는 것으로 표시하는 것일 수 있다.

발명의 효과

[0049] 산모의 주기적인 혈액 시료에서 본 발명의 일 예에 따른 표지유전자(marker gene)의 발현을 real-time PCR 등으로 분석하고, 혈액 검사 결과와 조합하여 분석할 경우 조산을 조기에 예측할 수 있고, 혈액 검사는 통상적으로 수행하고 있으므로 본 발명의 일 예에 따른 조산 예측 방법을 추가적으로 수행하여도 검사 비용의 증가가 거의 없어 시장성도 매우 밝다고 할 수 있다. 따라서, 본 발명은 과거의 조산 예측 기술에 비해 우수한 새로운 조산 예측법을 제공할 수 있으며, 본 발명에 따라 비침습적이면서도 정확도 높은 조산 예측이 가능할 것으로 기대된다.

도면의 간단한 설명

[0050] 도 1은 본 발명에 따른 조산군에서 정상군과 상이한 발현량을 나타낸 993개의 DEG (differentially expressed genes)를 나타낸 도면이다.

도 2a는 DAVID 소프트웨어를 사용하여 933개의 DEG에서 gene ontology 분석을 진행한 결과를 나타낸 도면이다.

도 2b는 조산 군에서 하향 조절된 DEG 간의 상호작용을 설명하는 네트워크 모델을 나타낸 도면이다.

도 2c는 조산 군에서 상향 조절된 DEG 간의 상호작용을 설명하는 네트워크 모델을 나타낸 도면이다.

도 3은 T-세포 활성화와 관련된 과정에 관여하는 상향 조절된 유전자 사이의 상호 작용을 설명하는 통합 네트워크 모델을 나타낸 도면이다.

도 4a는 조산 군에서 혈액의 림프구 수가 정상 군 대비 유의적으로 증가한 것을 나타낸 도면이다.

도 4b는 조산 군의 혈액 샘플에서 T-세포 마커 유전자 (CD4, CD8a/b 및 CD3d/g/e)의 mRNA 발현 수준이 증가한 것을 나타낸 도면이다.

도 4c는 조산 군에서 CYLD, TFRC, 및 RIPK2의 mRNA 발현량이 상향된 것을 확인한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0051] 이하, 본 발명을 하기의 실시예에 의하여 더욱 상세히 설명한다. 그러나 이들 실시예는 본 발명을 예시하기 위한 것일 뿐이며, 본 발명의 범위가 이들 실시예에 의하여 한정되는 것은 아니다.

[0053] 실험 재료 및 실험 방법

[0054] 실시예 1. 샘플 수집 (Sample collection)

[0055] 2014년부터 2019년도에 이화여자대학교의료원 목동병원 산부인과에 입원한 환자 중 88 명 (RNAseq 분석 5명 포함)의 조산 환자 (Preterm birth; PTB)와, 118명 (RNAseq 분석 5명 포함)의 정상 환자 (Full-term birth; FT B)를 연구 대상으로 하였다. 본 연구는 이화여자대학교병원의 기관 검토위원회에 의해 승인되었다 (EUMC 2018-07-007-010). 실험은 승인된 가이드라인에 따라 수행되었으며, 모든 피실험자로부터 사전 동의를 얻었다.

[0057] 실시예 2. RNA 추출 (RNA extraction)

[0058] 본 연구에 사용된 환자의 혈액 (전혈 샘플)은 조산 혹은 분만 증상이 있어 본원에 입원하였을 때 채취하여 -80

℃ 초저온 냉동고에 보관하였으며, RNA 시퀀싱 을 위하여 total RNA는 TRIzol RNA Isolation Reagent (Life technologies, Carlsbad, CA)를 이용하여 분리하였다. 추출된 RNA 정량적 분석과 정성적 분석을 위하여 NanoDrop ND-1000을 이용하여 흡광도를 측정하였으며, RNA integrity number (RIN)을 측정하기 위하여 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA)과 Agilent RNA 6000 Nano Kit(Agilent Technologies, Santa Clara, CA)를 이용하여 RIN 값이 8 이상임을 확인하고 실험에 이용하였다.

[0059] 본 연구에 사용된 모든 환자의 임상 데이터를 표 1에 나타내었다. 조산군과 정상군 두 집단에 있어서 임신주수 (gestational age), 백혈구 수치(white blood cell), 임파구 수치(lymphocyte) 값이 통계적으로 유의한 차이를 보였다.

표 1

Items		Full term birth	Preterm birth	P-value*
Maternal age		33.4 ±0.6	32.3 ±0.6	0.224
Body weight (kg)		59.5 ±2.6	58.1 ±3.1	0.746
Body weight at birth(kg)		71.4±2.4	64.0±1.6	0.010
Maternal height (cm)		162.7 ±0.9	161.3 ±0.9	0.301
Pre-pregnancy BMI		22.35±0.9	21.82±1.0	0.719
BMI at delivery		26.98±0.8	24.81±0.6	0.037
Parity		1.6± 0.1	1.7 ±0.2	0.580
Cervix length (cm)		2.5 ±0.2	2.3 ±0.2	0.477
Gestational Age		39.0±0.2	30.1±0.9	P<0.001
Blood profiles	WBC	9.2 ±0.4	12.5 ±0.7	p<0.001
	Platelets	218.6 ±10.7	226.2 ±9.2	0.590
	Neutrophil	72.8 ±1.1	77.5 ±1.1	0.004
	Lymphocyte	18.7 ±1.0	15.2 ±0.9	0.008
	Monocyte	6.8 ±0.3	5.9 ±0.3	0.032
	Eosinophil	1.1 ±0.2	0.9 ±0.2	0.355
	Basophil	0.3 ±0.0	0.2 ±0.0	0.005
	CRP	17.6 ±17.2	1.4 ±0.4	0.050
Neonatal Outcomes	Birth weight (g)	3238.1 ±66.6	1776.8 ±155.0	p<0.001
	APGAR score 1min	9.7 ±0.1	6.6 ±0.6	p<0.001
	APGAR score 5min	9.9 ±0.0	8.1 ±0.5	0.001

[0061] 실시예 3. Stranded mRNA 라이브러리 구축 (Stranded mRNA library construction)

[0062] RNA sequencing을 위한 mRNA libraries의 제작을 위하여 TruSeq Stranded Total RNA Sample Preparation Kit (Illumina, San Diego, CA)를 통해 수행하였다. Poly-A가 포함된 mRNA를 정제하기 위해 poly-T oligo가 부착된 oligo (dT) magnetic beads를 사용하였으며, 정제 후에 mRNA를 고농도의 2가 양이온을 이용하여 작은 조각으로 잘라내었다. 이후 잘라진 RNA 조각들은 random primers와 SuperScript II reverse transcriptase (Invitrogen, Carlsbad, CA)를 사용하여 first-strand cDNA를 복제하였고, RNase H와 DNA polymerase I를 사용하여 두 번째 cDNA가닥을 합성하였다. 말단이 손상되어 있는 합성된 cDNA 조각들은 말단 복구과정을 거쳤으며, 이 과정을 거치면서 말단 부분에 하나의 A염기가 첨가되었고, 그 후 합성된 cDNA를 ligation하였다.

[0063] 정제된 product들은 PCR 과정을 거쳐 최종 cDNA library로 만들어졌고, cDNA를 크기와 정성 분석은 Agilent DNA 1000 Kit (part # 5067-1504)를 2100 BioAnalyzer를 이용하였으며, Illumina HiSeq 2500 platform을 사용하여 sequencing 하였다. Raw data를 얻기 위하여 HiSeq control software (ver. 2.2.58)을 이용하여 이미지를 분석하였고, standard Illumina pipeline (CASAVA version 1.8.2 and RTA version 1.18.64)을 이용하였다.

[0064] 각각의 라이브러리에 대해, Agilent DNA 키트 (Agilent Technologies, Santa Clara, CA)를 사용하는 바이오 분석기에 의해 대략 200 bp의 크기임이 확인되었고, 라이브러리의 정량은 CFX96 시스템 (BioRad, Hercules, CA, USA)을 사용하는 real-time PCR에 의해 측정되었다. 각 라이브러리의 시퀀싱은 Illumina NextSeq500에서 수행되었고, cDNA 라이브러리의 클러스터는 TruSeq 플로우 셀 (flow cell)에서 생성되었으며, TruSeq 200 사이클 SBS 키트 (Illumina, San Diego, CA, USA)를 사용하여 76-bp 양 방향 읽기 (76-bp paired end reads) (2 × 76)법으로 시퀀싱되었다.

[0066] **실시예 4. RNA 시퀀싱 데이터 분석 (Analysis of RNA sequencing data)**

[0067] RNA 시퀀싱으로 생성된 서열에 대해, 어댑터 서열 (Truseq universal and indexed adapters)을 Cutadapt 소프트웨어 (ver. 1.2.1)를 사용하여 제거하였다. 리드 결과는 기본 옵션 (default option)의 TopHat 얼라이너 (TopHat aligner ver. 2.1.1)를 사용하여 인간 유전체 서열 중 GRCh38 버전을 참조로 (human reference genome)로 하여 정렬 (align)되었다. 정렬 후, 맵핑된 리드를 HTSeq (ver. 0.6.1)를 사용하여 유전자 특징 (GTF file of GRCh38)에 대해 커프 링크 (ver. 2.2.1)를 사용하여 FPKM법으로 표준화하고 맵핑하여 카운트하여 발현 양상을 비교하였다. 미가공된 원 데이터와 정규화 된 데이터를 GEO (Gene Expression Omnibus) 데이터베이스로 저장했다(GSE148402).

[0069] **실시예 5. DEG 분석 및 후보유전자 발굴 (Identification of DEGs)**

[0070] 5개의 PTB 또는 FTB 샘플 중 적어도 절반 이상 (즉, $n \geq 3$) 에서 0 보다 큰 FPKM 값과 0이 아니면서 1이 넘는 최대값을 발현된 유전자로 식별하였다. 이러한 발현된 유전자에 대해, 리드 카운트는 edgeR 패키지에서 TMM 표준화 방법을 사용하여 정규화되었다. RNAseq 결과는 $\log_2$ 로 치환하여 표준화 하였으며, 통계적 유의성을 검증하였다. quantile 정규화 방법을 사용하여 $\log_2$ (리드 카운트 + 1) 를 정규화하였다. 각 유전자에 대해, 조산군 (PTB) 와 정상군(FTB) 에서 T- 통계값 및 $\log_2$ -fold-changes를 기반으로 통계검증을 실시하였다.

[0071] 10 개의 표본을 300 번 무작위로 배치하여, T- 통계값 및 $\log_2$ -fold-changes의 경험적 귀무분포 (null distribution)를 추정했다. 추정된 경험적 분포를 사용하여, 각 유전자에 대한 두 테스트 (two tests)에 대해 조정된 p-value 를 계산 한 다음, 이러한 p-value 를 Stouffer 방법으로 검증하였다.

[0072] 마지막으로, p-value < 0.05, t-test p-values < 0.05, 및 absolute $\log_2$ -fold-changes > 0.58 (1.5-fold) 를 모두 만족하는 것을 DEG로 식별하였다.

[0073] 이전의 유전자 발현 데이터 세트의 분석을 위해, GEO 데이터베이스 (GSE46510 및 GSE96083)에서 처리된 데이터를 수집하고, 본 데이터에 동일한 통합 통계 방법을 적용하였다. GSE46510 및 GSE96083의 $\log_2$ -fold-changes 분포는 본 데이터보다 작은 분산을 나타냈다.

[0074] 따라서 GSE46510 및 GSE96083의 경우, $\log_2$ -fold-changes 의 경험적 귀무 분포 (empirical null distributions)의 10번째 및 90번째 백분위수 (percentiles)의 평균, 및 2.5번째 및 97.5번째 백분위수 (percentiles)의 평균을 각각 컷오프 (cutoffs)로 사용하였다 (GSE46510의 경우 0.34, GSE96083의 경우 0.50).

[0076] **실시예 6. GOBP enrichment 분석 (GOBP enrichment analysis)**

[0077] DEG가 나타내는 발현경로를 식별하기 위해, 발현이 증가하거나 감소한 유전자 (up- or down regulated genes)에 대하여 DAVID 소프트웨어를 사용하여 KEGG 기반 경로 분석 및 Gene ontology 분석을 수행하였으며 통계적으로 유의한 (p < 0.05) 경로만 선정하였다.

[0079] **실시예 7. 네트워크 모델의 재건 (Reconstruction of network models)**

[0080] KEGG 경로 데이터베이스 및 종래 문헌에서 얻은 활성 또는 억제 정보에 따라 네트워크 모델의 노드가 정렬되었다. 발현 양상을 분석한 결과를 기반으로 유전자의 up-regulated 된 집단에서는 T 세포의 활성 (lymphocyte activation and cytokine signaling)과 관련된 유전자를 선택하였으며, down-regulated 된 집단에서는 면역과 관련된 유전자들을 선별하였다 (innate immune response, TLR signaling pathway, cytokine secretion, and granulocyte migration/chemotaxis).

[0082] **실시예 8. qRT-PCR analysis (후보 유전자 검증)**

[0083] 후보 유전자의 검증은 PRISM 7000 sequence detection system (Applied BioSystems, Foster City, CA, USA)을 이용하여 qRT-PCR법으로 확인하였다. qRT-PCR 분석을 위해, SuperScript III 역전사 효소 (Invitrogen, Carlsbad, CA, USA) 및 RNasin (Promega, Madison, WI, USA)를 사용하여 20 μ L 반응물에서 1 μ g의 RNA를 cDNA로 역전사시켰다. 이어서, cDNA, 각각의 유전자에 대한 200 nM 프라이머, SYBR Premix EX Taq (Takara Bio, Shiga, Japan) 및 ROX 기준 염료 (ROX reference dye) (Takara Bio)를 포함하는 20 μ L 반응 혼합물에서 PRISM 7000 서열 검출 시스템 (Applied BioSystems, Foster City, CA, USA)을 사용하여 qRT-PCR을 수행 하였다. PCR 조건은 95 $^{\circ}$ C 10분, 95 $^{\circ}$ C 15초, 및 62 $^{\circ}$ C 1분으로 40 사이클을 수행하였으며, 마지막 사이클에서는 95 $^{\circ}$ C 15초, 및 62 $^{\circ}$ C 20초로 수행하였다. 대상 유전자의 정량은 $\Delta\Delta CT$ 법 ($\Delta\Delta CT$ method)을 사용하였으며, 항존유전자 (house keeping gene)는 glyceraldehyde-3-phosphate dehydrogenase를 사용하였다. qRT-PCR에 사용된 프라이머

리스트를 표 2에 나타내었다.

표 2

Gene name	Primer category	Primer sequence (5' → 3')	서열번호
CYLD	Forward	GCA ACC TCA TGC AGT TCT	1
	Reverse	AAA CCT TGA CCA CGA CCT	2
TFRC	Forward	GAC GCG CTA GTG TTC TTC T	3
	Reverse	AAC CGG GTA TAT GAC AAT GG	4
SRC	Forward	ACT ATG AGT CTA GGA CGG AG	5
	Reverse	CTG TGT TGT TGA CAA TCT GG	6
RIPK2	Forward	TTT GGG AAT TTG CAA TGA GC	7
	Reverse	AAG GAG GAG TCA TAT TGT GC	8
SMAD3	Forward	GGG GTT GGA CTT TCC TTC	9
	Reverse	CAG CAG AAG TTT GGG TTT C	10
GAPDH	Forward	TGG GCT ACA CTG AGC ACC AG	11
	Reverse	AAG TGG TCG TTG AGG GCA AT	12

CYLD: CYLD lysine 63 deubiquitinase

TFRC: transferrin receptor,

SRC: SRC proto-oncogene, non-receptor tyrosine kinase,

RIPK2: receptor interacting serine/threonine kinase 2,

SMAD3: SMAD family member 3,

GAPDH: glyceraldehyde-3-phosphate dehydrogenase.

연구 결과

실시예 9. 조산군과 정상군의 바이오마커 후보로서의 DEG (DEGs between PTB and FTB groups as biomarker candidates)

코호트에서 분만과 관련된 임상 징후가 있는 대상체로부터 전혈 샘플을 수집 하였다 : 1) 후보유전자 개발을 위한 코호트는 5 명의 PTB 환자 및 5 명의 FTB 대조군이 포함되었고, 2) 검증 코호트에 (validation cohort)는 83 명의 PTB 환자 및 113 명의 FTB 대조군이 포함되었다. 임신 연령, 백혈구 및 림프구를 포함한 임상 파라미터는 FTB 대조군보다 PTB 환자에서 유의하게 ($P < 0.05$) 높았다. discovery 및 검증 코호트 사이의 임상 파라미터의 비교는 표 1에 요약되어있다.

바이오 마커 후보를 확인하기 위해, 발굴 코호트로부터 수집된 전혈 샘플의 mRNA 서열 분석을 수행하였다. 서열 분석 데이터에서, 15,662 개의 유전자가 PTB 또는 FTB 그룹 중 절반 이상 ($n \geq 3$)에서 발현되었다. 이러한 15,662 개의 유전자 중에서, 본 발명자들은 통합 통계적 테스트 방법을 사용하여 PTB와 FTB 그룹 사이에서 933 DEG (differentially expressed genes) (273 개의 상향 조절된 유전자 및 660 개의 하향 조절된 유전자)를 확인하였다 (도 1).

실시예 10. 조산의 바이오 마커 후보가 나타내는 세포 과정 (Cellular processes represented by the biomarker candidates for PTB)

DAVID 소프트웨어를 사용하여 273 개의 상향 조절 및 660 개의 하향 조절 된 유전자에 대한 생물학적 Gene ontology 분석 (enrichment analysis of GOBPs)을 수행하여, 한국 PTB 코호트에서 933 개의 DEG가 나타내는 발현 경로를 조사하였다.

DAVID 소프트웨어를 사용하여 933개의 DEG에서 gene ontology 분석을 진행한 결과, 발현이 증가된 유전자는 주로 세포 증식 (세포주기, 세포 증식, 세포사멸 신호 전달, ER 스트레스에 대한 반응, RAS 신호, 및 번역/RNA 국소화), 후천성 면역 반응 (림프구 세포 활성화, 사이토카인에 대한 반응, 및 사이토카인 신호 전달), 세포 부착, 혈관 형성 및 탄수화물 대사 (당화 과정) 등과 연관되었다 (도 2a의 위쪽). 반면에, 발현이 감소된 유전자는 세포 내 이입 (endocytosis), 선천성 면역 반응 (TLR 신호 전달, 과립구 화학주성/이주 (chemotaxis/migration), 사이토카인 분비), 및 지질 대사 (지질 대사 과정; 도 2a의 하단) 등과 관련된 과정과

연관되었다.

- [0100] 흥미롭게도, 후천성 면역 반응 및 T-세포 활성화는 한국 PTB 환자에서 발현이 증가되는 반면, 선천성 면역 반응은 발현이 감소 조절되었다. 이러한 불일치를 분자 수준에서 분류하기 위해, 면역 반응과 관련된 과정에 연관된 DEG 간의 상호작용을 나타내는 네트워크 모델을 재구성했다. 네트워크 모델은 1) TLR (LY96, TLR1/4/5/6/10, LRRK2, 및 RIPK) 시그널링, 2) TNF (TNFRSF1A/10C, MAP3K5, MAP2K6, 및 RPS6KA1/5) 시그널링, 3) Fc 감마 수용체 매개 식세포 작용 (FCGR2A, PLD1, VAV3, RAC2 및 WAX) 및 4) 인테그린(integrin) (ITGA1/4) 및 성장인자 (IGF1R)-매개 세포골격 재구성 경로는 PTB에서 발현이 감소되었다 (도 2b).
- [0101] 대조적으로, 1) T-세포 활성화 (LAT, NECTIN2, RAB1B, TFRC, CYLD 및 RIPK2) 및 공동-자극 (co-stimulation) (CD40, ICOS 및 TRAF1) 경로, 및 2) T-세포에서의 케모카인 신호전달 (CCR7, GNA2/15, 및 GRK3) 은 PTB에서 발현이 증가되었다 (도 2c, 빨간색 점 (red nodes)). 이러한 데이터는 선천성 면역 세포가 감소하는 동안, PTB에서 T-세포의 수가 FTB 대비 증가하였음을 시사하였다.
- [0103] **실시예 11. 유전자 발현 프로파일의 통합에 의해 확인된 조산과 관련된 경로 (Core PTB-associated pathways identified through integration of gene expression profiles)**
- [0104] 신뢰할 수 있는 바이오 마커를 규명하기 위해, 본 발명자들은 호주 코호트의 전혈 샘플 및 혼합 인종 집단의 코호트의 상기 2 개의 유전자 발현 프로파일과 데이터를 통합하였다; 그 외의 프로파일은 비-혈액 샘플 (태반 및 조직생검) 또는 분만 전 (출생 4 일 후) 채집되지 않은 전혈 샘플로부터 생성되었기 때문에, 오해의 소지가 있는 편향으로 인해 제외되었다. 한국의 PTB 코호트에 사용된 것과 동일한 방법을 사용하여, 호주와 혼합 인종 코호트에서 각각 733 개 (279 개 상향 및 454 개 하향 조절) 및 1712 개 (971 개 상향 및 741 개 하향 조절) DEG를 각각 확인했다. 확인된 DEG는 종래에 알려진 바 없는 조산에 대한 신규한 마커였다.
- [0105] 도 2a에 의하면, 신규 마커는 유전자 수준에서 유의한 중복이 없고, pathway 수준에서 강한 중복이 존재하였으므로, 중복되지 않은 DEG는 동일한 경로에 관련된 것이라는 가정을 할 수 있었다. 먼저 두 개의 이전 데이터 세트에서 상향 및 하향 조절된 유전자에 대한 GOBP의 농축 분석 (enrichment analysis of GOBPs)을 수행한 후, 세 데이터 세트 모두에서 공유되는 GOBP를 검색했다. DEG에 의해 강화된 GOBP 중에서, 3 개의 모든 데이터 세트에서 상향 조절된 유전자에 의해 “T-세포 (림프구) 활성화 (T-cell activation)” 만이 일관되게 강화되었다. PTB와 관련된 핵심 경로로 T-세포 활성화에 초점을 맞추고, 3 개의 데이터 세트 모두에서 T-세포 활성화 관련된 과정 (T-세포/림프구 활성화 및 사이토카인에 대한 반응)에 관여하는 상향 조절된 유전자 사이의 상호 작용을 설명하는 통합 네트워크 모델을 재구성했다 (도 3).
- [0106] 한국인을 대상으로 한 본 사례에서 후천성 면역반응 (특히 T-cell 활성화)은 up-regulated 에서 중요하였으며, 선천적 면역반응은 down-regulated 에서 중요한 것으로 나타났다. 이는 T-세포 활성화 (LAT, NECTIN2, RAB1B, TFRC, CYLD 및 RIPK2) 및 공동-자극(co-stimulation) (CD40, ICOS 및 TRAF1) 경로 및 T-세포 (CCR7, GNA2/15, GRK3)는 조산군에서 up-regulated 되었다 (도 3의 빨간색 표시). 이러한 결과는 선천성 면역 세포의 양이 감소되는 동안 정상 분만군에 비해 조산 분만군에서 T-세포의 양이 증가됨을 시사한다. 따라서 본 발명에서는 T 세포의 활성화와 관련(T-cell receptor (TCR), chemokine, CD40 stimulation, transferrin, and peptidoglycan signaling)된 유전자를 후보 유전자로 선택하였고, 그 중 CYLD, TFRC, RIPK2 세 개의 유전자를 선별하여 검증을 실시하였다.
- [0107] 통합 네트워크 모델은 TCR 활성화, 케모카인, CD40 자극, 트랜스페린 (transferrin) 및 펩티도글리칸 신호와 같은 T-세포 활성화를 유도하는 몇 가지 신호 경로를 나타냈다. TCR 및 케모카인 신호전달 경로는 잘 알려져 있기 때문에, 새롭게 식별된 다른 신호 전달 경로에 초점을 맞추었다. 네트워크 모델은 상향 조절된 유전자 중 CYLD, TFRC 및 RIPK2가 각각 CD40 자극, 트랜스페린 및 펩티도글리칸 신호 전달 경로를 나타냄을 보여 주었다.
- [0109] **실시예 12. T-세포 활성화의 중요 조산-연관 경로와 관련된 바이오마커 후보 (Biomarker candidates involved in the core PTB-associated pathway of T-cell activation)**
- [0110] PTB에서 중요한 T-세포 활성화 경로를 검증하기 위해, 먼저 검증 코호트 (83 PTB 및 113 FTB)에서 fluorescence activated cell sorting (FACS)를 사용하여 PTB 환자의 전혈 샘플에서 6 가지 유형의 면역 세포 (lymphocyte, platelet, neutrophil, eosinophil, basophil, and monocyte) 의 양을 조사했다. 면역 세포 중에서, 혈액에서 림프구 수만이 FTB에서와 비교하여 PTB에서 유의적으로 ($P < 0.05$) 증가하였다 (도 4a).
- [0111] 이에 상응하여, PTB의 혈액 샘플에서 T-세포 마커 유전자 (CD4, CD8a/b 및 CD3d/g/e)의 mRNA 발현 수준이 증가하였다 (도 4b). 즉, mRNA 수준에서의 T 세포 표지 유전자에 대하여 분석한 결과에서도 조산군에서 증가되고 있

음을 확인하였다.

[0112] 마지막으로, 무작위로 추출된 30 명의 PTB 환자와 30 명의 FTB 대조군의 전혈 샘플을 사용하여, 통합 네트워크 분석에 따라 T- 세포 활성화 관련 신호 경로를 나타내는 CYLD, TFRC, 및 RIPK2의 발현이 증가함을 확인했다. qRT-PCR 분석은 FTB에서와 비교하여 PTB에서 3 개의 대표적인 유전자의 유의미한 ($P < 0.01$) 상향 조절을 확인하였다 (도 4c). 즉, 조산군 및 정상 분만군에서 무작위로 각각 30명씩 추출하여 qRT-PCR 법을 분석한 결과에서도 조산군에서 유의하게 증가됨을 알 수 있어, 본 발명에 따른 CYLD, TFRC, RIPK2 유전자는 조산을 예측하는 유력한 표지유전자로서의 유용성이 확인되었다.

[0113] 종합적으로, 이러한 데이터는 핵심 T-세포 활성화 경로를 나타내는 CYLD, TFRC 및 RIPK2가 PTB의 진단을 위한 신뢰할 수 있는 바이오 마커임을 시사한다.

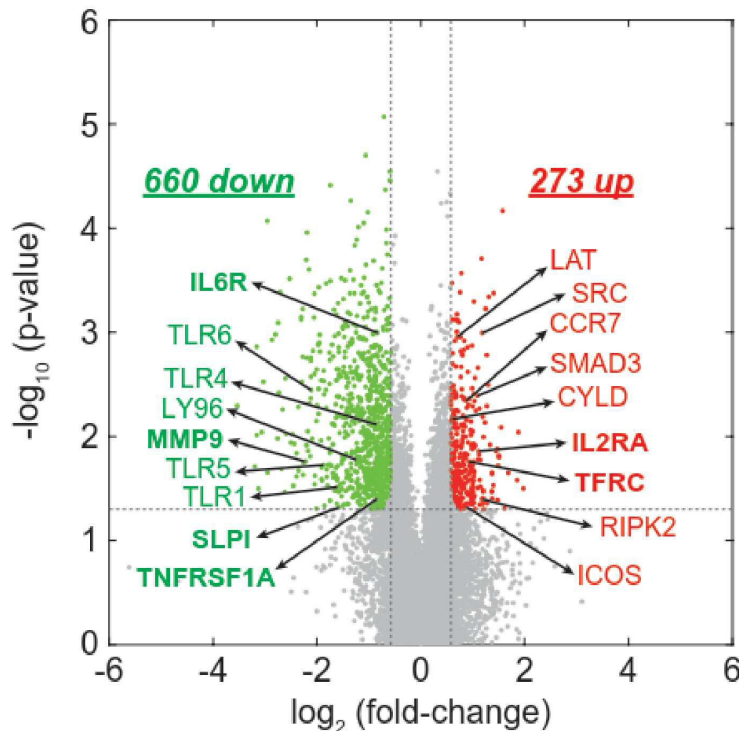
[0115] **실시예 13. Discussion**

[0116] PTB-연관 발현 변화의 mRNA 특징은 다차원적이며, 이러한 이질성 문제를 해결하기 위하여, 3 개의 서로 다른 데이터 집합에서 pathway-수준 통합의 PTB-연관 분자적 특징을 사용하여 DEGs에 의해 지속적으로 강화된 핵심 세포 경로를 찾았다. 이러한 접근법은 PTB에서 3 개의 데이터 세트에 의해 뒷받침되는 핵심 경로로서, 상향 조절된 T-세포 활성화-관련 경로를 확인하였고, DEG 중 핵심 T-세포 활성화 경로의 3 개의 대표적인 조절자 (CYLD, TFRC 및 RIPK2)를 PTB에 대한 전혈 바이오 마커의 후보로 규명하였으며, 이는 다른 인종 집단에서도 유효성을 가질 수 있다.

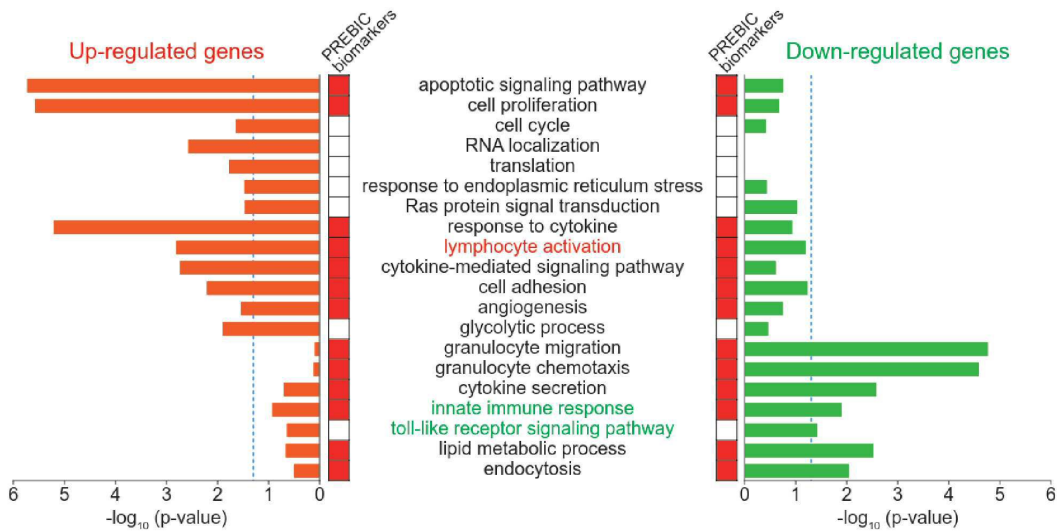
[0117] T-세포 활성화를 대표하는 CYLD, TFRC, 및 RIPK2를 PTB와 관련된 핵심 세포 경로를 나타내는 바이오 마커로 확인하였다. 이들 중 TFRC는 태반의 세포 철분의 흡수를 매개하는 것으로 알려져 있으며, 임신 중 철 결핍은 PTB의 위험을 증가시킨다. 태반 조직의 유전자 발현 분석은 TFRC가 PTB 환자의 태반에서 상향 조절되었음을 보여준다. CYLD 및 RIPK2는 T-세포 활성화에서 중요한 역할을 한다. 그러나 CYLD, RIPK2 와 PTB 간의 상관관계에 대해 알려진 바 없다. RIPK2의 동형 단백질인 RIPK1은 용모양막염 환자의 전혈 샘플에서 발현이 증가된다.

도면

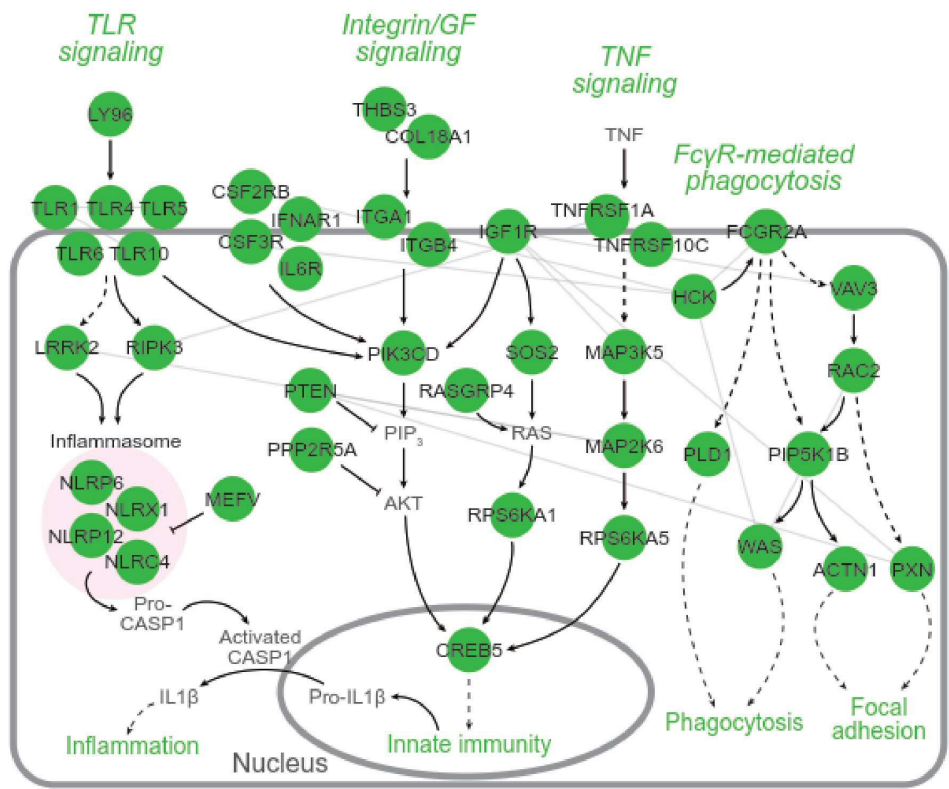
도면1



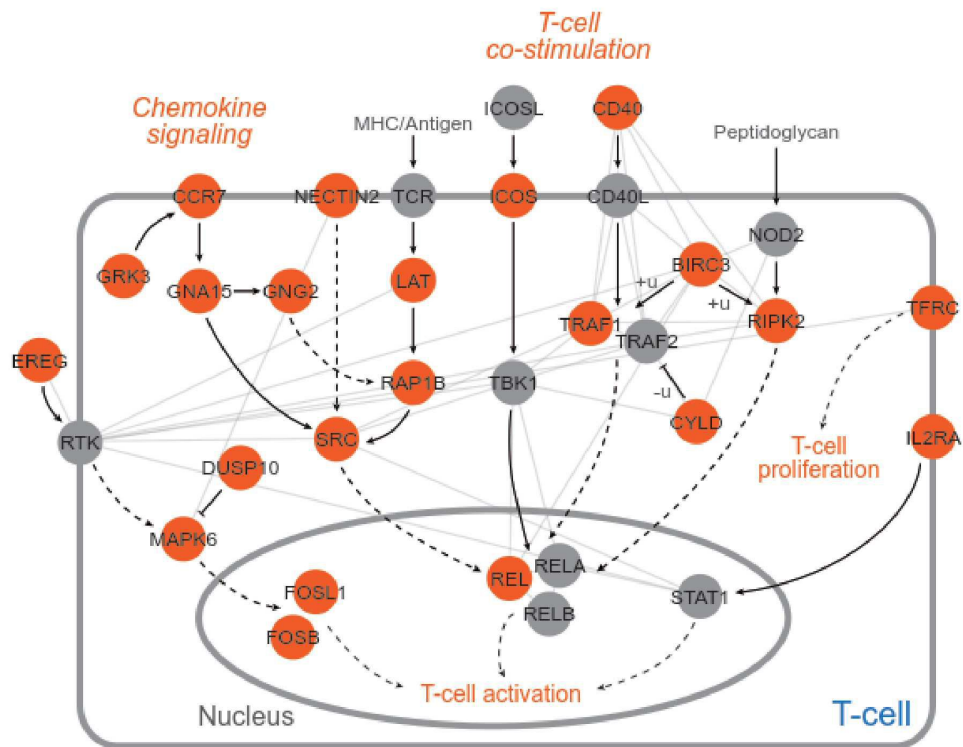
도면2a



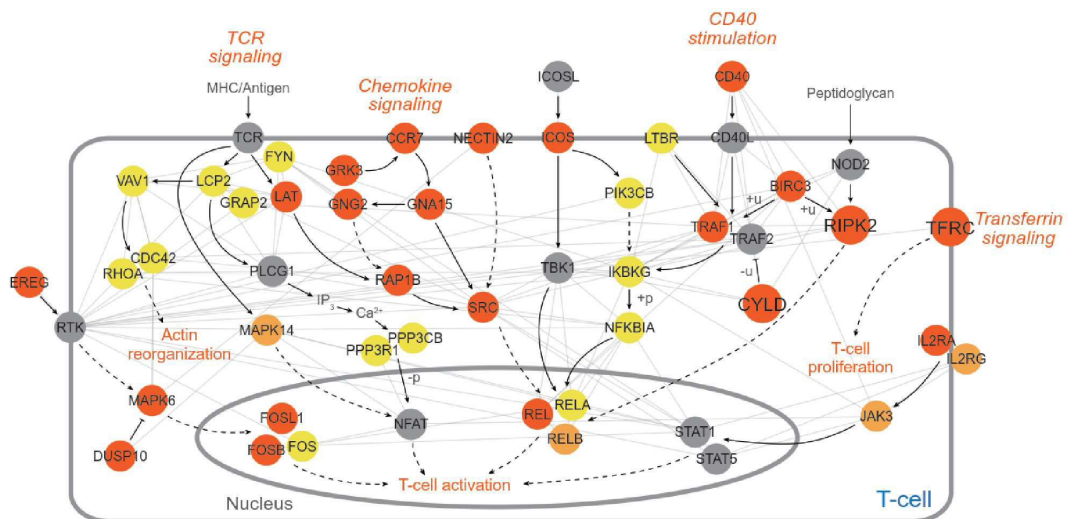
도면2b



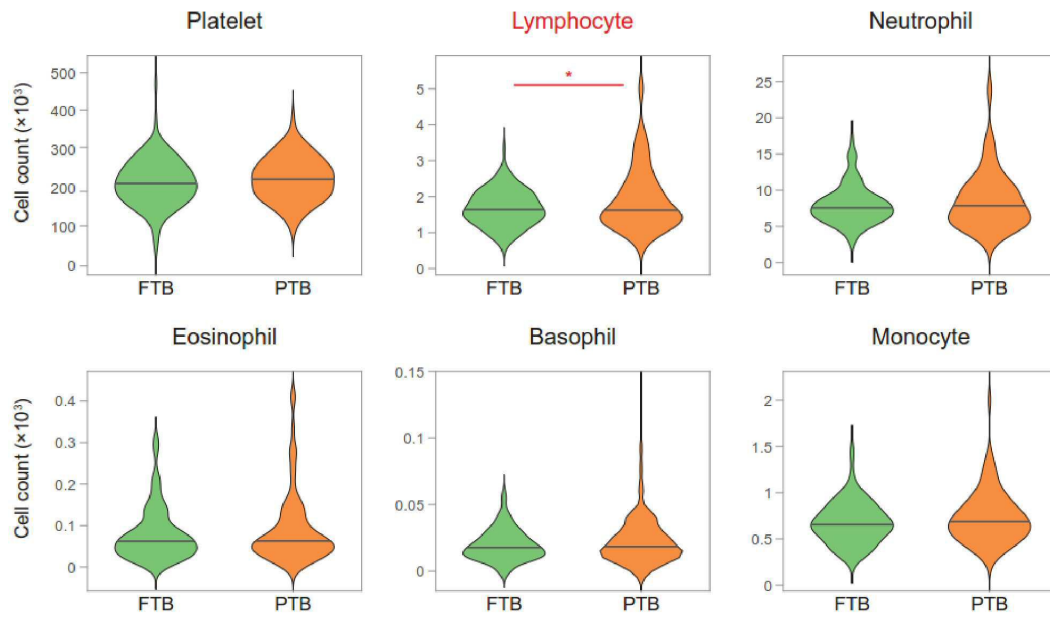
도면2c



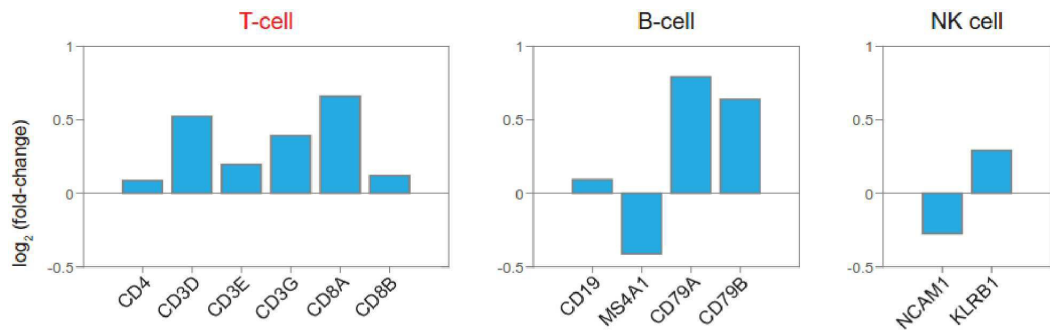
도면3



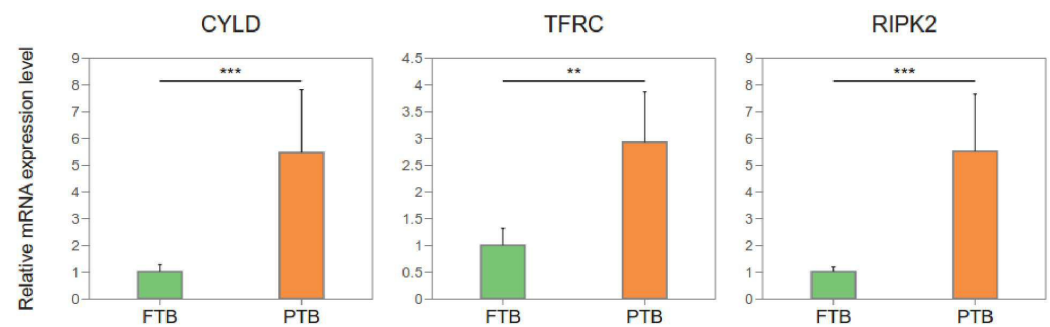
도면4a



도면4b



도면4c



서열 목록

- <110> Ewha University - Industry Collaboration Foundation
- <120> Composition for Predicting Premature Birth and Method for Predicting Premature Birth using the same
- <130> DPP20202430KR
- <160> 12

<170> KoPatent In 3.0
 <210> 1
 <211> 18
 <212> DNA
 <213> Artificial Sequence
 <220><223> CYLD primer FWD
 <400> 1
 gcaacctcat gcagttct 18
 <210> 2
 <211> 18
 <212> DNA
 <213> Artificial Sequence
 <220><223> CYLD primer RVS
 <400> 2
 aaaccttgac cagcacct 18
 <210> 3
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220><223> TFRC primer FWD
 <400> 3
 gacgcgctag tgtttctt 19
 <210> 4
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> TFRC primer RVS
 <400> 4
 aaccgggtat atgacaatgg 20
 <210> 5
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220><223> SRC primer FWD
 <400> 5
 actatgagtc taggacggag 20
 <210> 6
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> SRC primer RVS
 <400> 6
 ctgtgttgtt gacaatctgg 20
 <210> 7
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> RIPK2 primer FWD
 <400> 7
 tttgggaatt tgcaatgagc 20
 <210> 8
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> RIPK2 primer RVS
 <400> 8
 aaggaggagt catattgtgc 20
 <210> 9
 <211> 18
 <212> DNA
 <213> Artificial Sequence
 <220><223> SMAD3 primer FWD
 <400> 9
 ggggttggac tttccttc 18
 <210> 10
 <211> 19
 <212> DNA

<213> Artificial Sequence
 <220><223> SMAD3 primer RVS
 <400> 10

cagcagaagt ttgggtttc 19

<210> 11

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> GAPDH primer FWD

<400> 11

tgggctacac tgagcaccag 20

<210> 12

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> GAPDH primer RVS

<400> 12

aagtggtcgt tgagggaat 20