(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0088458
(43) 공개일자 2024년06월20일

(51) 국제특허분류(Int. Cl.)
G16B 40/00 (2019.01) *G16B 20/00* (2019.01)
G16B 25/00 (2019.01) *G16B 45/00* (2019.01)
G16B 50/00 (2019.01)
(52) CPC특허분류
G16B 40/00 (2019.02)
G16B 20/00 (2019.02)
(21) 출원번호 10-2022-0174211
(22) 출원일자 2022년12월13일
심사청구일자 2022년12월13일

(71) 출원인
재단법인 아산사회복지재단
서울특별시 송파구 올림픽로43길 88 (풍납동)
울산대학교 산학협력단
울산광역시 남구 대학로 93(무거동)
(72) 발명자
이건희
서울특별시 송파구 올림픽로43길 88 서울아산병원
이희진
서울특별시 송파구 올림픽로43길 88 서울아산병원
(뒷면에 계속)
(74) 대리인
특허법인더웨이브

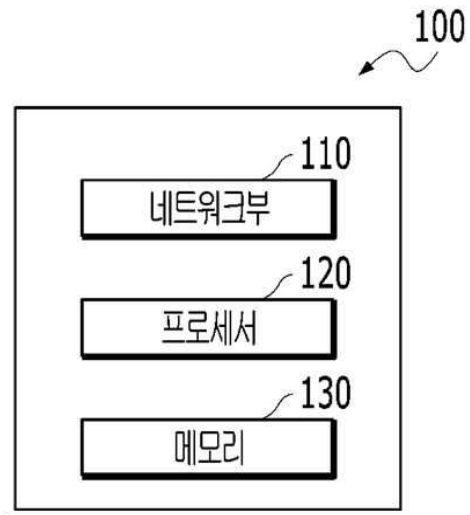
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 **면역 관련 인자를 이용하여 후보 약물 정보를 생성하는 장치 및 방법**

(57) 요 약

면역 관련 인자를 이용하여 후보 약물 정보를 생성하는 장치 및 방법에 관한 것으로, 일 양상에 따르면 면역 아형 및 바이오 마커를 이용하여 후보 약물 정보를 생성하는 장치 및 방법을 제공할 수 있다.

대 표 도 - 도1



(52) CPC특허분류
G16B 25/00 (2019.02)
G16B 45/00 (2019.02)
G16B 50/00 (2019.02)

(72) 발명자
전연진
서울특별시 송파구 올림픽로43길 88 서울아산병원
김경곤
서울특별시 송파구 올림픽로43길 88 서울아산병원
김지선
서울특별시 송파구 올림픽로43길 88 서울아산병원
김상엽
서울특별시 송파구 올림픽로43길 88 서울아산병원
정재호
서울특별시 송파구 올림픽로43길 88 서울아산병원

이 발명을 지원한 국가연구개발사업

과제고유번호	1711171435
과제번호	2022R1F1A1069837
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	기본연구
연구과제명	멀티오믹스를 활용한 삼중음성유방암 Immune subtype 구별 및 후보약물제시 모델 개발
기 여 율	1/1
과제수행기관명	주식회사 네오젠티씨
연구기간	2022.06.01 ~ 2023.02.28

명 세 서

청구범위

청구항 1

적어도 하나의 분석 모델을 포함하는 후보 약물 결정부에 의해, 면역 아형(immune subtype) 정보 및 바이오 마커 정보에 기초하여 후보 약물 정보를 생성하는 단계를 포함하는 컴퓨팅 장치에 의해 수행되는 후보 약물 정보를 생성하는 방법.

청구항 2

청구항 1에 있어서,

상기 후보 약물 정보는 면역 제외 또는 면역 결핍에서 면역 활성으로 면역 아형을 전환하기 위한 바이오 마커의 발현 조절에 관련된 약물 정보를 포함하는 것인, 방법.

청구항 3

청구항 1에 있어서,

상기 면역 아형 정보 및 바이오 마커 정보에 기초하여 후보 약물 정보를 생성하는 단계는,

제1 분석 모델을 이용하여, 약물 처리에 따른 유전자의 발현 조절에 관한 상관계수 값에 기초하여 면역 아형을 전환하기 위한 후보 약물 정보를 생성하는 단계를 포함하는 것인, 방법.

청구항 4

청구항 3에 있어서,

상기 약물 처리에 따른 유전자의 발현 조절에 관한 상관계수 값에 기초하여 면역 아형을 전환하기 위한 후보 약물 정보를 생성하는 단계는,

약물 처리에 따른 유전자의 발현 수준 정보를 비교함으로써 상관계수 값을 획득하는 단계; 및

면역 아형에 따른 유전자의 발현 수준 정보를 비교함으로써 상기 바이오 마커 정보에 대응하는 유전자의 발현 조절 정보를 획득하는 단계;를 포함하는 것인, 방법.

청구항 5

청구항 1에 있어서,

상기 면역 아형 정보 및 바이오 마커 정보에 기초하여 후보 약물 정보를 생성하는 단계는,

제2 분석 모델을 이용하여, GSEA(Gene set enrichment analysis)를 이용하여 면역 아형을 전환하기 위한 후보 약물 정보를 생성하는 단계;를 포함하는 것인, 방법.

청구항 6

청구항 5에 있어서,

상기 GSEA를 이용하여 면역 아형을 전환하기 위한 후보 약물 정보를 생성하는 단계는,

후보 약물과 기준 약물 사이의 유전자 셋(gene set)의 발현 수준 정보를 비교하여 제1 GSEA 스코어를 결정하는 단계; 및

면역 아형에 따른 유전자 셋의 발현 수준 정보를 비교하여 제2 GSEA 스코어를 결정하는 단계;를 포함하는 것인, 방법.

청구항 7

청구항 6에 있어서,

상기 후보 약물과 기준 약물 사이의 유전자 셋의 발현 수준 정보를 비교하여 제1 GSEA 스코어를 결정하는 단계는,

후보 약물 및 기준 약물에 대한 유전자 셋의 발현 수준 정보를 비교하여 제1 통계값을 획득하는 단계;

후보 약물에 대한 유전자 셋의 발현 수준 정보의 제1 평균값 및 기준 약물에 대한 유전자 셋의 발현 수준 정보의 제2 평균값을 획득하는 단계; 및

상기 제1 통계값, 상기 제1 평균값 및 상기 제2 평균값에 기초하여 사전 결정된 식에 따라 상기 제1 GSEA 스코어를 획득하는 단계;를 포함하는 것인, 방법.

청구항 8

청구항 7에 있어서,

상기 제1 통계값, 상기 제1 평균값 및 상기 제2 평균값에 기초하여 사전 결정된 식에 따라 상기 제1 GSEA 스코어를 획득하는 단계는,

상기 제1 GSEA 스코어를 보정하기 위해 검증 데이터베이스를 이용하여 보정값을 결정하는 단계를 포함하고,

상기 보정값은, 상기 제1 비교값과 정상 상태와 암 상태 간의 유전자 발현 수준의 제2 비교값 사이의 관계에 기초하여, 정상 상태 및 암 상태에 대한 유전자 셋의 발현 수준 정보를 비교함으로써 획득되는 제2 통계값에 의해 상기 제1 통계값을 보강 또는 상쇄하는 값인 것인, 방법.

청구항 9

청구항 7에 있어서,

상기 제1 통계값은 P-value 또는 Q-value인 것인, 방법.

청구항 10

청구항 7에 있어서,

상기 사전 결정된 식은 하기의 식 1인 것인, 방법:

[식 1]

$$제1\ GSEA\ 스코어 = i \times -\log(제1\ 통계값 + 보정값)$$

여기서, (제1 평균값-제2 평균값) > 0 인 경우 i = 1이고, (제1 평균값-제2 평균값) < 0 인 경우 i = -1 이다.

청구항 11

청구항 6에 있어서,

면역 아형에 따른 유전자 셋의 발현 수준 정보를 비교하여 제2 GSEA 스코어를 결정하는 단계는,

면역 제외 또는 면역 결핍인 경우에 유전자 셋의 제1 발현 수준 값과 면역 활성인 경우에 유전자 셋의 제2 발현 수준 값의 차를 이용하여 제2 GSEA 스코어를 결정하는 단계;를 포함하는 것인, 방법.

청구항 12

청구항 6에 있어서,

상기 GSEA를 이용하여 면역 아형을 전환하기 위한 후보 약물 정보를 생성하는 단계는,

상기 제1 GSEA 스코어와 제2 GSEA 스코어를 곱한 값을 이용하여 후보 약물 정보를 생성하는 단계;를 포함하는 것인, 방법.

청구항 13

청구항 1에 있어서,

상기 면역 아형 정보는 종양 침윤 림프구(TIL, Tumor Infiltrating Lymphocyte) 비율 및 클린트럽 기준(Klintrup criteria)를 이용하여 면역 활성(immune-inflamed), 면역 제외(immune-excluded), 및 면역 결핍(immune-desert)으로 분류되는 면역 아형의 유형에 관한 정보를 포함하는 것인, 방법.

청구항 14

적어도 하나의 분석 모델을 포함하는 후보 약물 결정부에 의해, 면역 아형(immune subtype) 정보 및 바이오 마커 정보에 기초하여 후보 약물 정보를 생성하는 단계를 포함하는 컴퓨팅 장치에 의해 수행되는 후보 약물 정보를 생성하는 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

청구항 15

적어도 하나의 메모리; 및

적어도 하나의 프로세서;를 포함하고,

상기 프로세서는 분석 모델을 포함하는 후보 약물 결정부에 의해, 면역 아형(immune subtype) 정보 및 바이오 마커 정보에 기초하여 후보 약물 정보를 생성하는 것인, 컴퓨팅 장치.

발명의 설명

기 술 분 야

[0001] 면역 관련 인자를 이용하여 후보 약물 정보를 생성하는 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 최근 종양에서 면역인자의 중요성이 밝혀지면서 면역관련 인자들을 이용한 치료법들이 활발히 개발되고 있다. 예컨대, 면역치료에는 종양관련 항원을 표적으로 하는 백신, 억제 수용체(inhibitory receptor)의 과발현에 의한 T 세포 피로화를 막는 단클론항체, T 림프구의 공동자극 수용체를 활성화하여 T 림프구를 활성화시키는 종양괴사인자(tumor necrosis factor, TNF) 수용체 superfamily의 사용, 면역억제 인자를 표적으로 하는 치료, 자연살해세포(natural kiler cells; NK cells) 또는 종양 특이 T 림프구를 이용한 세포치료법 등이 있다. 최근에는 면역억제 인자를 인지하는 항체(immune checkpoint inhibitor, anti-PD1, anti-CTLA4 antibody)를 이용한 면역치료가 악성 흑색종 등의 종양에서 효과가 있는 것으로 밝혀졌으나, 항체를 기반으로 한 치료는 세포 표면에 위치한 항원을 표적으로 하므로 세포 내에 있는 항원에는 적용이 불가능한 단점이 있다.

[0003] 자연살해세포 또는 종양 특이 T 림프구를 이용한 세포치료법 등을 면역조절 세포치료제라고 하며, 이는 면역세포를 이용하여 체내의 면역반응을 활성화시켜 질병을 치료할 목적으로 사용되는 치료법이다. 면역조절 세포치료제는 주로 암 치료를 적응증으로 개발되고 있으며, 환자에게 직접 면역세포를 투여하여 면역 기능을 활성화함으로써 치료 효과를 얻기 때문에 기존의 암 치료에 이용되는 수술요법, 항암제나 방사선 치료와는 차별화되는 치료 기전 및 효능으로, 향후 바이오 신약의 주요한 부분을 차지할 것으로 예상되는 분야이다.

[0004] 면역조절 세포치료제의 종류로는 크게 림포카인 활성 세포(LAK), 수지상세포, 및 T 세포 기반 치료제가 있으며, T 세포를 기반으로 한 세포치료제로는 대표적으로 환자의 종양 조직에 침투한 T 세포를 증식시켜 제조하는 종양 침윤 림프구(tumor-infiltrating lymphocytes; TIL)와 환자의 T 세포를 분리하여 T 세포 수용체(T cell receptor; TCR)나 키메릭 항원 수용체(chimeric antigen receptor; CAR)의 유전자를 도입하는 T 세포 수용체 발현 T 세포(TCR-modified T cell, TCR-T), 키메릭 항원 수용체 발현 T 세포(CAR-modified T cell, CAR-T)로 크게 구분할 수 있다.

[0005] 종양 침윤 림프구(Tumor-infiltrating lymphocytes; TIL)는 종양 조직에 존재하는 림프구로써 환자의 종양 조직을 채취하여 종양 세포 살상능력이 있는 T 세포들을 분리하여 이를 증식시킨 다음 환자에게 다시 투여하는 것으로, 생검(biopsy)이 가능한 고형암종이 주로 연구되고 있다. TIL은 T 세포를 이용한 면역조절 세포치료제 중 가장 먼저 임상에서의 효능이 연구된 것으로, 미국의 국립암연구소(National Cancer Institute)의 Steven Rosenberg 박사팀을 통해 전이성 흑색종 환자에서 TIL을 투여하여 항암 효과를 관찰한 논문이 1988년 처음으로 보고되었으며, 현재까지 TIL 투여에 의한 심각한 부작용이 보고된 적은 없다.

[0006] TIL과 같은 면역조절 세포치료제는 치료 반응률을 높이기 위해 면역 아형(immune subtype)을 확인하는 것이 중요하다. 또한, 확인된 면역 아형을 이용하여 적합한 약물에 관한 정보를 제공하는 구체적인 수단에 대한 수요가 존재한다.

선행기술문헌

비특허문헌

[0007] (비특허문헌 0001) Rosenberg SA, et al. 2015. Adoptive cell transfer as personalized immunotherapy for human cancer. Science. 2015 Apr 3;348(6230):62-8.

발명의 내용

해결하려는 과제

[0008] 일 양상은 적어도 하나의 분석 모델을 포함하는 후보 약물 결정부에 의해, 면역 아형(immune subtype) 정보 및 바이오 마커 정보에 기초하여 후보 약물 정보를 생성하는 단계를 포함하는 컴퓨팅 장치에 의해 수행되는 후보 약물 정보를 생성하는 방법을 제공하는 것이다.

[0009] 다른 양상은 적어도 하나의 분석 모델을 포함하는 후보 약물 결정부에 의해, 면역 아형(immune subtype) 정보 및 바이오 마커 정보에 기초하여 후보 약물 정보를 생성하는 단계를 포함하는 컴퓨팅 장치에 의해 수행되는 후보 약물 정보를 생성하는 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공하는 것이다.

[0010] 또 다른 양상은 적어도 하나의 메모리; 및 적어도 하나의 프로세서;를 포함하고, 상기 프로세서는 분석 모델을 포함하는 후보 약물 결정부에 의해, 면역 아형(immune subtype) 정보 및 바이오 마커 정보에 기초하여 후보 약물 정보를 생성하는 것인, 컴퓨팅 장치를 제공하는 것이다.

과제의 해결 수단

[0011] 일 양상은 적어도 하나의 분석 모델을 포함하는 후보 약물 결정부에 의해, 면역 아형(immune subtype) 정보 및 바이오 마커 정보에 기초하여 후보 약물 정보를 생성하는 단계를 포함하는 컴퓨팅 장치에 의해 수행되는 후보 약물 정보를 생성하는 방법을 제공한다.

[0012] 일 구체예에 따르면, 상기 후보 약물 정보는 면역 제외 또는 면역 결핍에서 면역 활성으로 면역 아형을 전환하기 위한 바이오 마커의 발현 조절에 관련된 약물 정보를 포함할 수 있다.

[0013] 일 구체예에 따르면, 상기 면역 아형 정보 및 바이오 마커 정보에 기초하여 후보 약물 정보를 생성하는 단계는, 제1 분석 모델을 이용하여, 약물 처리에 따른 유전자의 발현 조절에 관한 상관계수 값에 기초하여 면역 아형을 전환하기 위한 후보 약물 정보를 생성하는 단계를 포함할 수 있다.

[0014] 일 구체예에 따르면, 상기 약물 처리에 따른 유전자의 발현 조절에 관한 상관계수 값에 기초하여 면역 아형을 전환하기 위한 후보 약물 정보를 생성하는 단계는, 약물 처리에 따른 유전자의 발현 수준 정보를 비교함으로써 상관계수 값을 획득하는 단계; 및 면역 아형에 따른 유전자의 발현 수준 정보를 비교함으로써 상기 바이오 마커 정보에 대응하는 유전자의 발현 조절 정보를 획득하는 단계;를 포함할 수 있다.

[0015] 일 구체예에 따르면, 상기 면역 아형 정보 및 바이오 마커 정보에 기초하여 후보 약물 정보를 생성하는 단계는, 제2 분석 모델을 이용하여, GSEA(Gene set enrichment analysis)를 이용하여 면역 아형을 전환하기 위한 후보 약물 정보를 생성하는 단계;를 포함할 수 있다.

[0016] 일 구체예에 따르면, 상기 GSEA를 이용하여 면역 아형을 전환하기 위한 후보 약물 정보를 생성하는 단계는, 후보 약물과 기준 약물 사이의 유전자 셋(gene set)의 발현 수준 정보를 비교하여 제1 GSEA 스코어를 결정하는 단계; 및 면역 아형에 따른 유전자 셋의 발현 수준 정보를 비교하여 제2 GSEA 스코어를 결정하는 단계;를 포함할 수 있다.

[0017] 일 구체예에 따르면, 상기 후보 약물과 기준 약물 사이의 유전자 셋의 발현 수준 정보를 비교하여 제1 GSEA 스코어를 결정하는 단계는, 후보 약물 및 기준 약물에 대한 유전자 셋의 발현 수준 정보를 비교하여 제1 통계값을

획득하는 단계; 후보 약물에 대한 유전자 셋의 발현 수준 정보의 제1 평균값 및 기준 약물에 대한 유전자 셋의 발현 수준 정보의 제2 평균값을 획득하는 단계; 및 상기 제1 통계값, 상기 제1 평균값 및 상기 제2 평균값에 기초하여 사전 결정된 식에 따라 상기 제1 GSEA 스코어를 획득하는 단계;를 포함할 수 있다.

[0018] 일 구체예에 따르면, 상기 제1 통계값, 상기 제1 평균값 및 상기 제2 평균값에 기초하여 사전 결정된 식에 따라 상기 제1 GSEA 스코어를 획득하는 단계는, 상기 제1 GSEA 스코어를 보정하기 위해 검증 데이터베이스를 이용하여 보정값을 결정하는 단계를 포함하고, 상기 보정값은, 상기 제1 비교값과 정상 상태와 암 상태 간의 유전자 발현 수준의 제2 비교값 사이의 관계에 기초하여, 정상 상태 및 암 상태에 대한 유전자 셋의 발현 수준 정보를 비교함으로써 획득되는 제2 통계값에 의해 상기 제1 통계값을 보강 또는 상쇄하는 값일 수 있다.

[0019] 일 구체예에 따르면, 상기 제1 통계값은 P-value 또는 Q-value일 수 있다.

[0020] 일 구체예에 따르면, 상기 사전 결정된 식은 하기의 식 1일 수 있다:

[0021] [식 1]

[0022] $$제1\ GSEA\ 스코어 = i \times -\log(제1\ 통계값 + 보정값)$$

[0023] 여기서, (제1 평균값-제2 평균값) > 0 인 경우 i = 1이고, (제1 평균값-제2 평균값) < 0 인 경우 i = -1 이다

[0024] 일 구체예에 따르면, 면역 아형에 따른 유전자 셋의 발현 수준 정보를 비교하여 제2 GSEA 스코어를 결정하는 단계는, 면역 제외 또는 면역 결핍인 경우에 유전자 셋의 제1 발현 수준 값과 면역 활성인 경우에 유전자 셋의 제2 발현 수준 값의 차를 이용하여 제2 GSEA 스코어를 결정하는 단계;를 포함할 수 있다.

[0025] 일 구체예에 따르면, 상기 GSEA를 이용하여 면역 아형을 전환하기 위한 후보 약물 정보를 생성하는 단계는, 상기 제1 GSEA 스코어와 제2 GSEA 스코어를 곱한 값을 이용하여 후보 약물 정보를 생성하는 단계;를 포함할 수 있다.

[0026] 일 구체예에 따르면, 상기 면역 아형 정보는 종양 침윤 림프구(TIL, Tumor Infiltrating Lymphocyte) 비율 및 클린트럽 기준(Klintrup criteria)를 이용하여 면역 활성(immune-inflamed), 면역 제외(immune-excluded), 및 면역 결핍(immune-desert)으로 분류되는 면역 아형의 유형에 관한 정보를 포함할 수 있다.

[0027] 다른 양상은 적어도 하나의 분석 모델을 포함하는 후보 약물 결정부에 의해, 면역 아형(immune subtype) 정보 및 바이오 마커 정보에 기초하여 후보 약물 정보를 생성하는 단계를 포함하는 컴퓨팅 장치에 의해 수행되는 후보 약물 정보를 생성하는 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공한다.

[0028] 상기 프로그램은 하나 이상의 프로세서에서 실행되는 경우, 후보 약물 정보를 생성하는 방법을 수행하기 위한 동작들을 수행하도록 하며, 상기 동작들은 적어도 하나의 분석 모델을 포함하는 후보 약물 결정부에 의해, 면역 아형(immune subtype) 정보 및 바이오 마커 정보에 기초하여 후보 약물 정보를 생성하는 동작을 포함할 수 있다. 상기 후보 약물 정보는 면역 제외 또는 면역 결핍에서 면역 활성으로 면역 아형을 전환하기 위한 바이오 마커의 발현 조절에 관련된 약물 정보를 포함할 수 있다.

[0029] 또 다른 양상은 적어도 하나의 메모리; 및 적어도 하나의 프로세서;를 포함하고, 상기 프로세서는 분석 모델을 포함하는 후보 약물 결정부에 의해, 면역 아형(immune subtype) 정보 및 바이오 마커 정보에 기초하여 후보 약물 정보를 생성하는 것인, 컴퓨팅 장치를 제공한다. 상기 후보 약물 정보는 면역 제외 또는 면역 결핍에서 면역 활성으로 면역 아형을 전환하기 위한 바이오 마커의 발현 조절에 관련된 약물 정보를 포함할 수 있다.

발명의 효과

[0030] 일 양상에 따르면 면역 아형 및 바이오 마커를 이용하여 후보 약물 정보를 생성하는 장치 및 방법을 제공할 수 있다.

도면의 간단한 설명

[0031] 도 1은 본 개시의 몇몇 실시예에 따라 후보 약물 정보를 생성하기 위한 동작을 수행하는 컴퓨팅 장치의 블록 구성도를 도시한 도면이다.

도 2는 본 개시의 몇몇 실시예에 따른 네트워크 함수를 나타낸 개략도이다.

도 3은 본 개시의 몇몇 실시예에 따른 후보 약물 결정부를 설명하기 위한 도면이다.

도 4는 본 개시의 몇몇 실시예에 따른 면역 아형의 유형을 설명하기 위한 도면이다.

도 5는 본 개시의 몇몇 실시예에 따른 제1 분석 모델을 설명하기 위한 도면이다.

도 6은 본 개시의 몇몇 실시예에 따른 제1 분석 모델의 예시적인 동작을 설명하기 위한 도면이다.

도 7는 본 개시의 몇몇 실시예에 따른 제2 분석 모델을 설명하기 위한 도면이다.

도 8은 본 개시의 몇몇 실시예에 따른 제2 분석 모델의 예시적인 동작을 설명하기 위한 도면이다.

도 9은 본 개시의 몇몇 실시예에 따른 제2 분석 모델의 예시적인 동작을 설명하기 위한 또 다른 도면이다.

도 10은 본 개시의 몇몇 실시예에 따른 후보 약물 정보를 생성하는 방법의 순서도이다.

도 11은 본 개시의 몇몇 실시예에 따른 컴퓨팅 장치의 블록 구성도이다.

발명을 실시하기 위한 구체적인 내용

[0032] 아래에서는 다양한 실시예들이 도면을 참고하여 설명된다. 본 명세서에서, 다양한 설명들이 본 개시의 이해를 제공하기 위해서 제시된다. 그러나, 이러한 실시예들은 이러한 구체적인 설명 없이도 실행될 수 있음이 명백하다.

[0033] 본 명세서에서 사용되는 용어 "컴포넌트", "모듈", "시스템" 등은 컴퓨터-관련 엔티티, 하드웨어, 펌웨어, 소프트웨어, 소프트웨어 및 하드웨어의 조합, 또는 소프트웨어의 실행을 지칭한다. 예를 들어, 컴포넌트는 프로세서 상에서 실행되는 처리과정(procedure), 프로세서, 객체, 실행 스레드, 프로그램, 및/또는 컴퓨터일 수 있지만, 이들로 제한되는 것은 아니다. 예를 들어, 컴퓨팅 장치에서 실행되는 애플리케이션 및 컴퓨팅 장치 모두 컴포넌트일 수 있다. 하나 이상의 컴포넌트는 프로세서 및/또는 실행 스레드 내에 상주할 수 있다. 일 컴포넌트는 하나의 컴퓨터 내에 로컬화 될 수 있다. 일 컴포넌트는 2개 이상의 컴퓨터들 사이에 분배될 수 있다. 또한, 이러한 컴포넌트들은 그 내부에 저장된 다양한 데이터 구조들을 갖는 다양한 컴퓨터 판독가능한 매체로부터 실행할 수 있다. 컴포넌트들은 예를 들어 하나 이상의 데이터 패킷들을 갖는 신호(예를 들면, 로컬 시스템, 분산 시스템에서 다른 컴포넌트와 상호작용하는 하나의 컴포넌트로부터의 데이터 및/또는 신호를 통해 다른 시스템과 인터넷과 같은 네트워크를 통해 전송되는 데이터)에 따라 로컬 및/또는 원격 처리들을 통해 통신할 수 있다.

[0034] 더불어, 용어 "또는"은 배타적 "또는"이 아니라 내포적 "또는"을 의미하는 것으로 의도된다. 즉, 달리 특정되지 않거나 문맥상 명확하지 않은 경우에, "X는 A 또는 B를 이용한다"는 자연적인 내포적 치환 중 하나를 의미하는 것으로 의도된다. 즉, X가 A를 이용하거나; X가 B를 이용하거나; 또는 X가 A 및 B 모두를 이용하는 경우, "X는 A 또는 B를 이용한다"가 이들 경우들 어느 것으로도 적용될 수 있다. 또한, 본 명세서에 사용된 "및/또는"이라는 용어는 열거된 관련 아이템들 중 하나 이상의 아이템의 가능한 모든 조합을 지칭하고 포함하는 것으로 이해되어야 한다.

[0035] 또한, "포함한다" 및/또는 "포함하는"이라는 용어는, 해당 특징 및/또는 구성요소가 존재함을 의미하는 것으로 이해되어야 한다. 다만, "포함한다" 및/또는 "포함하는"이라는 용어는, 하나 이상의 다른 특징, 구성요소 및/또는 이들의 그룹의 존재 또는 추가를 배제하지 않는 것으로 이해되어야 한다. 또한, 달리 특정되지 않거나 단수 형태를 지시하는 것으로 문맥상 명확하지 않은 경우에, 본 명세서와 청구범위에서 단수는 일반적으로 "하나 또는 그 이상"을 의미하는 것으로 해석되어야 한다.

[0036] 그리고, "A 또는 B 중 적어도 하나"이라는 용어는, "A만을 포함하는 경우", "B 만을 포함하는 경우", "A와 B의 구성으로 조합된 경우"를 의미하는 것으로 해석되어야 한다.

[0037] 당업자들은 추가적으로 여기서 개시된 실시예들과 관련되어 설명된 다양한 예시적 논리적 블록들, 구성들, 모듈들, 회로들, 수단들, 로직들, 및 알고리즘 단계들이 전자 하드웨어, 컴퓨터 소프트웨어, 또는 양쪽 모두의 조합들로 구현될 수 있음을 인식해야 한다. 하드웨어 및 소프트웨어의 상호교환성을 명백하게 예시하기 위해, 다양한 예시적 컴포넌트들, 블록들, 구성들, 수단들, 로직들, 모듈들, 회로들, 및 단계들은 그들의 기능성 측면에서 일반적으로 위에서 설명되었다. 그러한 기능성이 하드웨어로 또는 소프트웨어로서 구현되는지 여부는 전반적인 시스템에 부과된 특정 어플리케이션(application) 및 설계 제한들에 달려 있다. 숙련된 기술자들은 각각의 특정 어플리케이션들을 위해 다양한 방법들로 설명된 기능성을 구현할 수 있다. 다만, 그러한 구현의 결정들이 본 개시내용의 영역을 벗어나게 하는 것으로 해석되어서는 안 된다.

[0038] 용어 "바이오 마커(biomarker)"란 약물 투여시 간 조직에서 증가 또는 감소를 보이는 폴리펩타이드 또는 핵산(예컨대, mRNA 등), 지질, 당지질, 당단백질, 당(단당류, 이당류, 올리고당류 등) 등과 같은 유기 생체 분자 등을 포함할 수 있다. 본 개시에서, 바이오 마커란 바람직하게 유전자 또는 단백질을 의미할 수 있다.

[0039] 본 개시에서 상기 "유전자의 발현 수준"은, 바람직하게 유전자가 발현된 mRNA 수준, 즉, mRNA의 양을 의미할 수 있다. 또한, 유전자의 발현 수준은 유전자에 의해 코딩 되는 단백질의 발현 수준을 의미할 수 있다.

[0040] 제시된 실시예들에 대한 설명은 본 개시의 기술 분야에서 통상의 지식을 가진 자가 본 발명을 이용하거나 또는 실시할 수 있도록 제공된다. 이러한 실시예들에 대한 다양한 변형들은 본 개시의 기술 분야에서 통상의 지식을 가진 자에게 명백할 것이다. 여기에 정의된 일반적인 원리들은 본 개시의 범위를 벗어남이 없이 다른 실시예들에 적용될 수 있다. 그리하여, 본 발명은 여기에 제시된 실시예 들로 한정되는 것이 아니다. 본 발명은 여기에 제시된 원리들 및 신규한 특징들과 일관되는 최광의의 범위에서 해석되어야 할 것이다.

[0041] 도 1은 본 개시의 몇몇 실시예에 따라 후보 약물 정보를 생성하기 위한 동작을 수행하는 컴퓨팅 장치의 블록 구성도를 도시한 도면이다.

[0042] 도 1에서 도시되는 바와 같이, 컴퓨팅 장치(100)는 프로세서(110), 메모리(130), 네트워크부(150)를 포함할 수 있다. 도 1에 도시된 컴퓨팅 장치(100)의 구성은 간략화 하여 나타낸 예시일 뿐이다. 본 개시의 몇몇 실시예에서 컴퓨팅 장치(100)는 컴퓨팅 장치(100)의 컴퓨팅 환경을 수행하기 위한 다른 구성들을 포함할 수 있고, 개시된 구성들 중 일부만이 컴퓨팅 장치(100)를 구성할 수도 있다.

[0043] 프로세서(110)는 하나 이상의 코어로 구성될 수 있으며, 컴퓨팅 장치의 중앙 처리 장치(CPU: central processing unit), 범용 그래픽 처리 장치 (GPGPU: general purpose graphics processing unit), 텐서 처리 장치(TPU: tensor processing unit) 등의 데이터 분석 및 처리, 딥러닝을 위한 프로세서를 포함할 수 있다. 프로세서(110)는 면역 아형 및 바이오 마커를 이용하여 후보 약물 정보를 생성할 수 있는 후보 약물 결정부 및 그 컴포넌트들을 구현 및 제공할 수 있다. 프로세서(110)는 메모리(130)에 저장된 컴퓨터 프로그램을 판독하여 본 개시의 몇몇 실시예에 따른 후보 약물 정보를 생성하는 동작을 위한 데이터 변환, 연산, 생성 등을 수행할 수 있다. 또한, 프로세서(110)는 이하에서 설명되는 후보 약물 결정부 및 그 컴포넌트들의 동작들을 수행할 수 있다. 또한, 본 개시의 몇몇 실시예에 따라 프로세서(110)는 후보 약물 결정부의 동작을 수행하기 위해 학습 데이터를 이용하여 신경망의 학습을 위한 연산을 수행할 수 있다. 프로세서(110)는 딥러닝(DL: deep learning)에서 학습을 위한 입력 데이터의 처리, 입력 데이터에서의 피처 추출, 오차 계산, 역전파(backpropagation)를 이용한 신경망의 가중치 업데이트 등의 신경망의 학습을 위한 계산을 수행할 수 있다. 프로세서(110)의 CPU, GPGPU, 및 TPU 중 적어도 하나가 후보 약물 결정부를 구현하기 위한 연산을 처리할 수 있다. 예를 들어, CPU 와 GPGPU가 함께 후보 약물 결정부를 구현하기 위한 연산을 처리할 수 있다. 또한, 본 개시의 몇몇 실시예에서 복수의 컴퓨팅 장치의 프로세서를 함께 사용하여 후보 약물 결정부의 동작을 위한 데이터 변환, 연산, 생성, 네트워크 함수의 학습, 네트워크 함수를 이용한 데이터 분류를 처리할 수 있다. 또한, 본 개시의 몇몇 실시예에 따른 컴퓨팅 장치에서 수행되는 컴퓨터 프로그램은 CPU, GPGPU 또는 TPU 실행가능 프로그램일 수 있다.

[0044] 본 개시의 몇몇 실시예에 따르면, 메모리(130)는 프로세서(110)가 생성하거나 결정한 임의의 형태의 정보 및 네트워크부(150)가 수신한 임의의 형태의 정보를 저장할 수 있다. 예를 들어, 메모리(130)는 프로세서(110)에 의해 후보 약물 결정부의 동작을 수행하는 과정에서 생성되는 데이터들을 저장할 수 있다. 또한 메모리(130)는 프로세서(110)에 의해 후보 약물 정보를 생성하는 방법을 수행하는 과정에서 외부에서 수신되는 데이터들을 저장할 수 있다. 다만, 이에 한정되지 않고, 메모리(130)는 본 개시의 몇몇 실시예에 따른 후보 약물 정보를 생성하는 방법을 수행하기 위한 다양한 정보를 저장할 수 있다.

[0045] 본 개시의 몇몇 실시예에 따르면, 메모리(130)는 플래시 메모리 타입(flash memory type), 하드디스크 타입(hard disk type), 멀티미디어 카드 마이크로 타입(multimedia card micro type), 카드 타입의 메모리(예를 들어 SD 또는 XD 메모리 등), 램(Random Access Memory, RAM), SRAM(Static Random Access Memory), 롬(Read-Only Memory, ROM), EEPROM(Electrically Erasable Programmable Read-Only Memory), PROM(Programmable Read-Only Memory), 자기 메모리, 자기 디스크, 광디스크 중 적어도 하나의 타입의 저장매체를 포함할 수 있다. 컴퓨팅 장치(100)는 인터넷(internet) 상에서 상기 메모리(130)의 저장 기능을 수행하는 웹 스토리지(web storage)와 관련되어 동작할 수도 있다. 전술한 메모리에 대한 기재는 예시일 뿐, 본 개시는 이에 제한되지 않는다.

[0046] 본 개시의 몇몇 실시예에 따른 네트워크부(150)는 임의의 형태의 공지된 유무선 통신 시스템을 사용할 수

있다.

[0047] 네트워크부(150)는 프로세서(110)에 의해 처리된 정보, 사용자 인터페이스 등을 타 단말과의 통신을 통해 송수신할 수 있다. 예를 들어, 네트워크부(150)는 프로세서(110)에 의해 생성된 사용자 인터페이스를 클라이언트(e.g. 사용자 단말)로 제공할 수 있다. 또한, 네트워크부(150)는 클라이언트로 인가된 사용자의 외부 입력을 수신하여 프로세서(110)로 전달할 수 있다. 이때, 프로세서(110)는 네트워크부(150)로부터 전달받은 사용자의 외부 입력을 기초로 사용자 인터페이스를 통해 제공되는 정보의 출력, 수정, 변경, 추가 등의 동작을 처리할 수 있다.

[0048] 구체적으로 예를 들면, 네트워크부(150)는 본 개시의 몇몇 실시예에 따라 후보 약물 결정부의 동작을 수행하기 위한 다양한 정보를 송수신할 수 있다. 예를 들어, 네트워크부(150)는 데이터베이스 상에 저장된 약물 처리 정보 및 유전자의 발현 수전 정보를 수신할 수 있다. 또한, 네트워크부(150)는 이하에서 설명되는 후보 약물 정보를 생성하는 방법을 수행하는 과정에서 생성되는 몇몇의 데이터를 데이터베이스 상에 저장하기 위해 외부로 전송할 수 있다.

[0049] 한편, 본 개시의 몇몇 실시예에 따른 컴퓨팅 장치(100)는 클라이언트와 통신을 통해 정보를 송수신하는 컴퓨팅 시스템으로서 서버를 포함할 수 있다. 이때, 클라이언트는 서버에 엑세스할 수 있는 임의의 형태의 단말일 수 있다. 예를 들어, 서버인 컴퓨팅 장치(100)는 사용자 단말로부터 쿼리를 수신하여 쿼리에 대응하는 단일 정보 처리 결과를 생성할 수 있다. 이 경우에, 서버인 컴퓨팅 장치(100)는 처리 결과를 포함하는 사용자 인터페이스를 사용자 단말로 제공할 수 있다. 이때, 사용자 단말은 서버인 컴퓨팅 장치(100)로부터 수신한 사용자 인터페이스를 출력하고, 사용자와의 상호 작용을 통해 정보를 입력받거나 처리할 수 있다.

[0050] 추가적인 실시예에서, 컴퓨팅 장치(100)는 임의의 서버에서 생성된 데이터 리소스를 전달받아 추가적인 정보 처리를 수행하는 임의의 형태의 단말을 포함할 수도 있다.

[0051] 도 2는 본 개시의 몇몇 실시예에 따라 네트워크 함수를 나타낸 개략도이다.

[0052] 본 명세서에 걸쳐, 연산 모델, 신경망, 네트워크 함수, 뉴럴 네트워크(neural network)는 상호 교환 가능한 의미로 사용될 수 있다. 신경망은 일반적으로 노드라 지칭될 수 있는 상호 연결된 계산 단위들의 집합으로 구성될 수 있다. 이러한 노드들은 뉴런(neuron)들로 지칭될 수도 있다. 신경망은 적어도 하나 이상의 노드들을 포함하여 구성된다. 신경망들을 구성하는 노드(또는 뉴런)들은 하나 이상의 링크에 의해 상호 연결될 수 있다.

[0053] 본 개시의 후보 약물 결정부의 몇몇 컴포넌트는 몇몇 예에서 단순 연산 모델로 구현되거나, 또는 이하에서 설명되는 네트워크 함수를 이용하여 구현될 수 있다.

[0054] 신경망 내에서, 링크를 통해 연결된 하나 이상의 노드들은 상대적으로 입력 노드 및 출력 노드의 관계를 형성할 수 있다. 입력 노드 및 출력 노드의 개념은 상대적인 것으로서, 하나의 노드에 대하여 출력 노드 관계에 있는 임의의 노드는 다른 노드와의 관계에서 입력 노드 관계에 있을 수 있으며, 그 역도 성립할 수 있다. 상술한 바와 같이, 입력 노드 대 출력 노드 관계는 링크를 중심으로 생성될 수 있다. 하나의 입력 노드에 하나 이상의 출력 노드가 링크를 통해 연결될 수 있으며, 그 역도 성립할 수 있다.

[0055] 하나의 링크를 통해 연결된 입력 노드 및 출력 노드 관계에서, 출력 노드의 데이터는 입력 노드에 입력된 데이터에 기초하여 그 값이 결정될 수 있다. 여기서 입력 노드와 출력 노드를 상호 연결하는 링크는 가중치(weight)를 가질 수 있다. 가중치는 가변적일 수 있으며, 신경망이 원하는 기능을 수행하기 위해, 사용자 또는 알고리즘에 의해 가변 될 수 있다. 예를 들어, 하나의 출력 노드에 하나 이상의 입력 노드가 각각의 링크에 의해 상호 연결된 경우, 출력 노드는 상기 출력 노드와 연결된 입력 노드들에 입력된 값들 및 각각의 입력 노드들에 대응하는 링크에 설정된 가중치에 기초하여 출력 노드 값을 결정할 수 있다.

[0056] 상술한 바와 같이, 신경망은 하나 이상의 노드들이 하나 이상의 링크를 통해 상호 연결되어 신경망 내에서 입력 노드 및 출력 노드 관계를 형성한다. 신경망 내에서 노드들과 링크들의 개수 및 노드들과 링크들 사이의 연관관계, 링크들 각각에 부여된 가중치의 값에 따라, 신경망의 특성이 결정될 수 있다. 예를 들어, 동일한 개수의 노드 및 링크들이 존재하고, 링크들의 가중치 값이 상이한 두 신경망이 존재하는 경우, 두 개의 신경망들은 서로 상이한 것으로 인식될 수 있다.

[0057] 신경망은 하나 이상의 노드들의 집합으로 구성될 수 있다. 신경망을 구성하는 노드들의 부분 집합은 레이어(layer)를 구성할 수 있다. 신경망을 구성하는 노드들 중 일부는, 최초 입력 노드로부터의 거리들에 기초하여, 하나의 레이어(layer)를 구성할 수 있다. 예를 들어, 최초 입력 노드로부터 거리가 n인 노드들의 집합은, n 레

이어를 구성할 수 있다. 최초 입력 노드로부터 거리는, 최초 입력 노드로부터 해당 노드까지 도달하기 위해 거쳐야 하는 링크들의 최소 개수에 의해 정의될 수 있다. 그러나, 이러한 레이어의 정의는 설명을 위한 임의적인 것으로서, 신경망 내에서 레이어의 차수는 상술한 것과 상이한 방법으로 정의될 수 있다. 예를 들어, 노드들의 레이어는 최종 출력 노드로부터 거리에 의해 정의될 수도 있다.

[0058] 최초 입력 노드는 신경망 내의 노드들 중 다른 노드들과의 관계에서 링크를 거치지 않고 데이터가 직접 입력되는 하나 이상의 노드들을 의미할 수 있다. 또는, 신경망 네트워크 내에서, 링크를 기준으로 한 노드 간의 관계에 있어서, 링크로 연결된 다른 입력 노드들을 가지지 않는 노드들을 의미할 수 있다. 이와 유사하게, 최종 출력 노드는 신경망 내의 노드들 중 다른 노드들과의 관계에서, 출력 노드를 가지지 않는 하나 이상의 노드들을 의미할 수 있다. 또한, 히든 노드는 최초 입력 노드 및 최후 출력 노드가 아닌 신경망을 구성하는 노드들을 의미할 수 있다.

[0059] 본 개시의 일 실시예에 따른 신경망은 입력 레이어의 노드의 개수가 출력 레이어의 노드의 개수와 동일할 수 있으며, 입력 레이어에서 히든 레이어로 진행됨에 따라 노드의 수가 감소하다가 다시 증가하는 형태의 신경망일 수 있다. 또한, 본 개시의 다른 일 실시예에 따른 신경망은 입력 레이어의 노드의 개수가 출력 레이어의 노드의 개수 보다 적을 수 있으며, 입력 레이어에서 히든 레이어로 진행됨에 따라 노드의 수가 감소하는 형태의 신경망일 수 있다. 또한, 본 개시의 또 다른 일 실시예에 따른 신경망은 입력 레이어의 노드의 개수가 출력 레이어의 노드의 개수보다 많을 수 있으며, 입력 레이어에서 히든 레이어로 진행됨에 따라 노드의 수가 증가하는 형태의 신경망일 수 있다. 본 개시의 또 다른 일 실시예에 따른 신경망은 상술한 신경망들의 조합된 형태의 신경망일 수 있다.

[0060] 딥 뉴럴 네트워크(DNN: deep neural network, 심층신경망)는 입력 레이어와 출력 레이어 외에 복수의 히든 레이어를 포함하는 신경망을 의미할 수 있다. 딥 뉴럴 네트워크를 이용하면 데이터의 잠재적인 구조(latent structures)를 파악할 수 있다. 즉, 사진, 글, 비디오, 음성, 음악의 잠재적인 구조(예를 들어, 어떤 물체가 사진에 있는지, 글의 내용과 감정이 무엇인지, 음성의 내용과 감정이 무엇인지 등)를 파악할 수 있다. 딥 뉴럴 네트워크는 컨볼루션 뉴럴 네트워크(CNN: convolutional neural network), 리커런트 뉴럴 네트워크(RNN: recurrent neural network), 오토 인코더(auto encoder), GAN(Generative Adversarial Networks), 제한 볼츠만 머신(RBM: restricted boltzmann machine), 심층 신뢰 네트워크(DBN: deep belief network), Q 네트워크, U 네트워크, 샴 네트워크, 적대적 생성 네트워크(GAN: Generative Adversarial Network), 트랜스포머(transformer) 등을 포함할 수 있다. 전술한 딥 뉴럴 네트워크의 기재는 예시일 뿐이며 본 개시는 이에 제한되지 않는다.

[0061] 뉴럴 네트워크는 교사 학습(supervised learning), 비교사 학습(unsupervised learning), 반교사학습(semi supervised learning), 또는 강화학습(reinforcement learning) 중 적어도 하나의 방식으로 학습될 수 있다. 뉴럴 네트워크의 학습은 뉴럴 네트워크가 특정한 동작을 수행하기 위한 지식을 뉴럴 네트워크에 적용하는 과정일 수 있다.

[0062] 뉴럴 네트워크는 출력의 오류를 최소화하는 방향으로 학습될 수 있다. 뉴럴 네트워크의 학습에서 반복적으로 학습 데이터를 뉴럴 네트워크에 입력시키고 학습 데이터에 대한 뉴럴 네트워크의 출력과 타겟의 에러를 계산하고, 에러를 줄이기 위한 방향으로 뉴럴 네트워크의 에러를 뉴럴 네트워크의 출력 레이어에서부터 입력 레이어 방향으로 역전파(backpropagation)하여 뉴럴 네트워크의 각 노드의 가중치를 업데이트 하는 과정이다. 교사 학습의 경우 각각의 학습 데이터에 정답이 라벨링 되어있는 학습 데이터를 사용하며(즉, 라벨링된 학습 데이터), 비교사 학습의 경우는 각각의 학습 데이터에 정답이 라벨링되어 있지 않을 수 있다. 즉, 예를 들어 데이터 분류에 관한 교사 학습의 경우의 학습 데이터는 학습 데이터 각각에 카테고리가 라벨링 된 데이터 일 수 있다. 라벨링된 학습 데이터가 뉴럴 네트워크에 입력되고, 뉴럴 네트워크의 출력(카테고리)과 학습 데이터의 라벨을 비교함으로써 오류(error)가 계산될 수 있다. 다른 예로, 데이터 분류에 관한 비교사 학습의 경우 입력인 학습 데이터가 뉴럴 네트워크 출력과 비교됨으로써 오류가 계산될 수 있다. 계산된 오류는 뉴럴 네트워크에서 역방향(즉, 출력 레이어에서 입력 레이어 방향)으로 역전파 되며, 역전파에 따라 뉴럴 네트워크의 각 레이어의 각 노드들의 연결 가중치가 업데이트 될 수 있다. 업데이트 되는 각 노드의 연결 가중치는 학습률(learning rate)에 따라 변화량이 결정될 수 있다. 입력 데이터에 대한 뉴럴 네트워크의 계산과 에러의 역전파는 학습 사이클(epoch)을 구성할 수 있다. 학습률은 뉴럴 네트워크의 학습 사이클의 반복 횟수에 따라 상이하게 적용될 수 있다. 예를 들어, 뉴럴 네트워크의 학습 초기에는 높은 학습률을 사용하여 뉴럴 네트워크가 빠르게 일정 수준의 성능을 확보하도록 하여 효율성을 높이고, 학습 후기에는 낮은 학습률을 사용하여 정확도를 높일 수 있다.

[0063] 뉴럴 네트워크는 학습시키기 위한 학습 데이터의 양을 늘리기 위해, 다양한 데이터 증강 방법이 사용될 수 있다. 예를 들어, 2차원 변환인 회전(Rotation), 크기(Scale), 밀림(Shearing), 반사(Reflection), 이동(Translation)을 통해 데이터 증강이 수행될 수 있다. 또한, 노이즈 삽입, 색상, 밝기 변형 등을 활용하여 데이터 증강이 수행될 수 있다.

[0064] 뉴럴 네트워크의 학습에서 일반적으로 학습 데이터는 실제 데이터(즉, 학습된 뉴럴 네트워크를 이용하여 처리하고자 하는 데이터)의 부분집합일 수 있으며, 따라서, 학습 데이터에 대한 오류는 감소하나 실제 데이터에 대해서는 오류가 증가하는 학습 사이클이 존재할 수 있다. 과적합(overfitting)은 이와 같이 학습 데이터에 과하게 학습하여 실제 데이터에 대한 오류가 증가하는 현상이다. 과적합은 머신러닝 알고리즘의 오류를 증가시키는 원인으로 작용할 수 있다. 이러한 과적합을 막기 위하여 다양한 최적화 방법이 사용될 수 있다. 과적합을 막기 위해서는 학습 데이터를 증가시키거나, 레귤라이제이션(regularization), 학습의 과정에서 네트워크의 노드 일부를 비활성화하는 드롭아웃(dropout), 배치 정규화 레이어(batch normalization layer)의 활용 등의 방법이 적용될 수 있다.

[0065] 도 3은 본 개시의 몇몇 실시예에 따른 후보 약물 결정부를 설명하기 위한 도면이다. 도 4는 본 개시의 몇몇 실시예에 따른 면역 아형의 유형을 설명하기 위한 도면이다.

[0066] 본 개시의 몇몇 실시예에서, 후보 약물 결정부(300)는 적어도 하나의 분석 모델을 이용하여 면역 아형(immune subtype) 정보(10) 및 바이오 마커 정보(20)에 기초하여 후보 약물 정보를 생성할 수 있다. 여기서, 후보 약물 정보(30)는 면역 제외(immune-excluded) 또는 면역 결핍(immune-desert)에서 면역 활성(immune-inflamed)으로 면역 아형을 전환하기 위한 바이오 마커의 발현 조절에 관련된 약물 정보를 포함할 수 있다.

[0067] 구체적으로 설명하면, 후보 약물 결정부(300)는 면역 아형 정보(10) 및 바이오 마커 정보(20)를 입력 데이터로 처리함으로써 후보 약물 정보(30)를 출력 데이터로 생성할 수 있다. 여기서 후보 약물 정보(30)는 면역 아형을 전환하기 위해 바이오 마커의 발현 조절에 관여하는 약물에 관한 정보를 포함할 수 있다. 예를 들어, 후보 약물 결정부(300)는 면역 아형 정보(10)를 통해 환자의 면역 아형이 면역 제외임을 식별할 수 있다. 그리고, 후보 약물 결정부(300)는 바이오 마커 정보(30)를 통해 면역 아형의 전환하는데 관여하는 하나 이상의 바이오 마커를 식별할 수 있다. 그리고, 후보 약물 결정부(300)는 하나 이상의 분석 모델을 이용하여 면역 아형의 전환과 관련하여 바이오 마커의 발현을 조절하는 하나 이상의 약물에 대한 후보 약물 정보를 생성할 수 있다. 후보 약물 정보는 면역 아형을 '면역 활성'으로 전환하도록 유전자의 발현을 조절하는 약물에 대한 정보를 포함할 수 있다. 후보 약물 결정부(300)는 면역 아형 정보(10) 및 바이오 마커 정보(20)를 분석 모델을 이용하여 처리함으로써 후보 약물 정보(30)를 생성할 수 있다. 몇몇 예에서, 예시적인 분석 모델의 개수는 2개일 수 있으나, 본 개시는 이에 한정되지 않는다. 분석 모델에 대한 자세한 설명은 도 4 내지 도 7을 참조하여 후술된다.

[0068] 본 개시의 몇몇 실시예에 따라, 면역 아형 정보(10)는 종양 침윤 림프구(TIL, Tumor Infiltrating Lymphocyte) 비율 및 클린트럽 기준(Klintrup criteria)를 이용하여 면역 활성(immune-inflamed), 면역 제외(immune-excluded), 및 면역 결핍(immune-desert)으로 분류되는 면역 아형의 유형에 관한 정보를 포함할 수 있다.

[0069] 면역 아형 정보(10)는 환자가 가지고 있는 면역 아형의 유형에 대한 정보를 포함할 수 있다. 예를 들어, 도 4를 참조하면, 면역 아형은 면역 활성(inflamed), 면역 제외(excluded), 및 면역 결핍(desert)을 포함할 수 있다. 면역 활성은 암 조직에 대한 면역 세포의 침투 정도가 높은 상태일 수 있다. 도 4를 참조하면, 면역 활성으로 추정되는 의료 이미지는 종양 세포와 T-세포가 경계가 없이 섞여 있는 형상을 가질 수 있다. 면역 제외는 경계면(invasive margin)에 면역 세포가 존재하는 상태일 수 있다. 도 4를 참조하면, 면역 제외로 추정되는 의료 이미지는 종양 세포와 T-세포가 섞이지 않고 서로 경계를 이루는 형상을 가질 수 있다. 면역 결핍은 암 조직 내 및 경계에 면역 세포가 거의 없는 상태일 수 있다. 도 4를 참조하면, 면역 결핍으로 추정되는 의료 이미지는 종양 세포 이외에 면역 세포가 거의 존재하지 않는 형상을 가질 수 있다. 전술한 면역 아형에 관한 기재는 예시일 뿐이며, 본 개시는 이에 제한되지 않는다.

[0070] 자세히 설명하면, 면역 아형 정보(10)는 의료 이미지를 분석함으로써 획득될 수 있다. 면역 아형 정보(10)를 획득하는데 사용되는 면역 형상은 다양한 유전체 변이 및 발현(또는 단백질 발현)에 따라 상이한 형상을 가질 수 있다. 이 경우에, 면역 아형 정보(10)는 의료 이미지에 나타나는 면역 형상을 종양 침윤 림프구(TIL, Tumor Infiltrating Lymphocyte) 비율 및 클린트럽 기준(Klintrup criteria)을 이용하여 분류함으로써 획득될 수 있다. TIL(%)은 암세포 주변부에 분포하는 림프구(Lymphocyte)의 정도에 따라 결정될 수 있다. 클린트럽 기준은 암세포들의 경계면에 분포하는 림프구의 정도에 따라 결정될 수 있다. 몇몇 예에서, 면역 아형 정보(10)는 전문가의 육안을 통해 의료 이미지를 분석함으로써 결정될 수 있다. 다른 예로, 면역 아형 정보(10)는 면역 아형을

분류하도록 학습된 모델을 통해 결정될 수 있다. 면역 아형을 분류하도록 학습된 모델은 별도의 장치에 구현되거나, 후보 약물 결정부(300)와 동일한 장치 상에 구현될 수 있다. 다만 이에 한정되지 않고, 면역 아형 정보(10)는 다양한 방식으로 획득될 수 있다.

[0071] 바이오 마커 정보(20)는 두 유형의 면역 아형 간의 유전체 변이 및 발현을 비교함으로써 획득될 수 있다. 예를 들어, 의료 이미지와 관련된 게놈 데이터 또는 프로테오믹스 데이터를 비교 분석함으로써 하나 이상의 바이오 마커가 결정될 수 있다. 예를 들어, '면역 제외'로 분류되는 의료 이미지와 관련된 유전체 발현 정보가 '면역 활성'으로 분류되는 의료 이미지와 관련된 유전체 발현 정보와 비교될 수 있다. 이 경우에, 두 의료 이미지를 비교한 결과, 상대적으로 높게(또는 낮게) 나타나는 유전체 변이 및 발현은 바이오 마커를 결정하는데 사용될 수 있다. 예를 들어, 면역 제외로 분류되는 제1 의료 이미지로부터 식별되는 CORO1A, ACADS, 및 ALDH16A의 유전체 발현 정보가, 면역 활성으로 분류되는 제2 의료 이미지와 관련된 유전체 발현 정보에 비해 차별되는 수치로 식별되는 경우에, CORO1A, ACADS, 및 ALDH16이 바이오 마커로 결정될 수 있다.

[0072] 몇몇 예에서, 바이오 마커 정보(20)는 유전자의 발현 수준 정보에 대한 전문가의 분석을 통해 결정될 수 있다. 다른 예로, 바이오 마커 정보(20)는 면역 형상이 나타나는 의료 이미지 또는 유전자의 발현 수준 정보로부터 바이오 마커를 추정하도록 학습된 모델을 이용하여 획득될 수 있다. 다만 이에 한정되지 않고, 바이오 마커 정보(20)는 다양한 방식으로 획득될 수 있다.

[0073] 본 개시의 몇몇 실시예에 따라, 분석 모델은 제1 분석 모델(310) 및 제2 분석 모델(320)을 포함할 수 있다. 예를 들어, 제1 분석 모델은 약물 처리 이후에 유전자의 발현 변화에 관한 상관계수 값을 이용하여 면역 아형을 전환하기 위한 후보 약물 정보를 생성할 수 있다. 제2 분석 모델은 GSEA(Gene set enrichment analysis)를 통해 면역 아형을 전환하기 위한 후보 약물 정보를 생성할 수 있다. 후보 약물 결정부(300)는 각각의 분석 모델에서 생성된 후보 약물 정보를 통합하여 후보 약물 정보(30)를 생성할 수 있다. 예를 들어, 후보 약물 결정부(300)는 각각의 분석 모델에서 생성된 후보 약물 정보 중에서 공통되는 약물을 후보 약물 정보(30)로 제공할 수 있다. 다만 이에 한정되지 않고, 후보 약물 결정부(300)는 각각의 분석 모델에서 생성된 후보 약물 정보를 이용하여 다양한 방식으로 후보 약물 정보(30)를 생성할 수 있다. 이하에서, 도 5 내지 도 9를 참조하여, 후보 약물 결정부(300)를 구성하는 분석 모델이 설명된다.

[0074] 도 5는 본 개시의 몇몇 실시예에 따른 제1 분석 모델(310)을 설명하기 위한 도면이다. 도 6은 본 개시의 몇몇 실시예에 따른 제1 분석 모델(310)의 예시적인 동작을 설명하기 위한 도면이다.

[0075] 상술한 바와 같이, 제1 분석 모델(310)은 약물 처리 이후에 유전자의 발현 변화에 관한 상관계수 값을 이용하여 면역 아형을 전환하기 위한 후보 약물 정보를 생성할 수 있다. 몇몇 예에서, 도 5를 참조하면, 제1 분석 모델(310)은 제1 서브 모델(311), 제2 서브 모델(312), 및 제1 후보 약물 결정 모듈(313)을 포함할 수 있다. 도 5에서 도시되는 제1 분석 모델(310)의 컴포넌트들은 단지 예로서 도시되며, 일부 컴포넌트가 생략되거나, 추가적인 컴포넌트가 포함될 수 있다.

[0076] 본 개시의 몇몇 실시예에서, 제1 분석 모델(310)은 약물 처리에 따른 유전자의 발현 수준 정보를 비교함으로써 상관계수 값을 획득하는 제1 서브 모델(311)을 포함할 수 있다.

[0077] 구체적으로 설명하면, 제1 서브 모델(311)은 약물 처리와 유전자의 발현 수준과의 관계를 나타내는 상관계수 값을 획득하기 위한 모델일 수 있다. 예를 들어, 제1 서브 모델(311)은 상관계수 값을 획득하기 위해 약물 처리 정보와 유전자의 발현 수준 정보를 매칭한 조합 데이터를 이용할 수 있다. 몇몇 예에서, 제1 서브 모델(311)은 약물(및 투여 시간, 농도 등)과 유전자 간의 상호관계에 대한 정보를 제공할 수 있는 데이터베이스로부터 조합 데이터를 획득할 수 있다. 예를 들어, 제1 서브 모델(311)은 통합된 네트워크 기반의 세포 시그니쳐 라이브러리(Library of Integrated Network-based Cellular Signatures, LINCS)으로부터 조합 데이터를 획득할 수 있다. 예를 들어, 제1 서브 모델(311)은 하버드 의과대학(HMS)에서 제공하는 링크스 프로그램(HMS LINCS, http://lincs.hms.harvard.edu/db/), 미국 국립 보건원(NIH)에서 제공하는 링크스 프로그램(NIH LINCS, www.lincsproject.org), DGIdb(The Drug Gene Interaction Database, www.dgidb.org) 등의 데이터베이스를 사용할 수 있으나, 이에 한정되지 않는다.

[0078] 제1 서브 모델(311)은 다양한 약물, 투여 시간, 농도의 조합으로 구성되는 각각의 약물 처리 정보에 대응하는 유전자의 발현 수준 정보를 이용하여 상관계수 값을 획득할 수 있다. 여기서, 상관계수 값은 약물 처리 정보에 따라 유전자의 발현 수준이 증가하는 경우에 양수 값을 가질 수 있다. 그리고, 상관계수 값은 약물 처리 정보에 따라 유전자의 발현 수준이 감소하는 경우에 음수 값을 가질 수 있다. 도 6을 참조하면, 약물 처리 정보에 대한

유전자의 상관계수 값을 표시한 표가 도시된다.

[0079] 본 개시의 몇몇 실시예에서, 제1 분석 모델(310)은 면역 아형에 따른 유전자의 발현 수준 정보를 비교함으로써 바이오 마커 정보(20)에 대응하는 유전자의 발현 조절 정보를 획득하는 제2 서브 모델(312)을 포함할 수 있다.

[0080] 구체적으로 설명하면, 제2 서브 모델(312)은 면역 아형 간의 유전자의 발현 수준을 비교함으로써, 면역 아형을 전환하기 위한 유전자의 발현 조절 정보를 획득할 수 있다. 몇몇 예에서, 유전자의 발현 조절 정보는 면역 아형을 '면역 제외' 또는 '면역 결핍'에서 '면역 활성'으로 전환하기 위해 바이오 마커와 관련된 발현이 증가되어야 하는지 또는 감소되어야 하는지에 대한 정보를 포함할 수 있다. 제2 서브 모델(312)은 바이오 마커 정보(20)로부터 약물에 의해 발현을 조절할 바이오 마커를 식별할 수 있다. 이 경우에, 제2 서브 모델(312)은'면역 제외' 또는 '면역 결핍'인 경우에 바이오 마커의 발현 수준을 '면역 활성'인 경우에 바이오 마커의 발현 수준을 비교함으로써, 유전자의 발현 조절 정보를 획득할 수 있다. 예를 들어, 바이오 마커의 발현 수준이 '면역 활성'에서 감소되는 경우, 제2 서브 모델(312)은 바이오 마커의 발현 수준을 감소시킬 것을 나타내는 유전자의 발현 조절 정보를 획득할 수 있다. 다른 예로, 바이오 마커의 발현 수준이 '면역 활성'에서 증가되는 경우, 제2 서브 모델(312)은 바이오 마커의 발현 수준을 증가시킬 것을 나타내는 유전자의 발현 조절 정보를 획득할 수 있다. 바이오 마커가 복수개인 경우, 제2 서브 모델(312)은 각각의 바이오 마커에 대한 유전자의 발현 조절 정보를 획득할 수 있다. 다만 이에 한정되지 않고, 제2 서브 모델(312)은 다양한 방식으로 바이오 마커 정보에 대응하는 유전자의 발현 조절 정보를 획득할 수 있다.

[0081] 제2 서브 모델(312)은 각각의 면역 아형에 대한 유전자의 발현 수준 정보를 획득하기 위해 생검 시료로부터 단백질 측정 장치를 이용하여 유전자 또는 단백질의 양을 확인한 데이터를 이용할 수 있다. 예를 들어, 제1 서브 모델(311)은 삼중 음성 유방암(Triple Negative Breast Cancer) 환자를 대상으로 획득된 TNBC 데이터를 이용할 수 있다. 다만, 이에 한정되지 않고, 제2 서브 모델(312)은 다양한 데이터를 이용하여 유전자의 발현 수준 정보를 획득할 수 있다.

[0082] 본 개시의 몇몇 실시예에 따라, 제1 후보 약물 결정 모듈(313)은 상관계수 값 및 유전자의 발현 조절 정보를 이용하여 후보 약물 정보를 생성할 수 있다. 구체적으로 설명하면, 제1 후보 약물 결정 모듈(313)은 유전자의 발현 조절 정보에 대응하는 바이오 마커의 발현 조절을 유도하는 약물 처리에 관한 정보를 후보 약물 정보로 생성할 수 있다. 예를 들어, 도 6을 참조하면, 유전자의 발현 조절 정보가 바이오 마커인 'NAT1'의 발현을 증가시킬 것을 나타내는 경우에, 제1 후보 약물 결정 모듈(313)은 양의 상관관계 값을 나타내는 약물 처리 정보 'Drug_1'를 이용하여 후보 약물 정보를 생성할 수 있다. 다른 예로, 제1 후보 약물 결정 모듈(313)은 복수개의 바이오 마커의 발현을 조절하는 하나 이상의 약물 처리 정보를 식별할 수 있다. 이 경우에, 제1 후보 약물 결정 모듈(313)은 하나 이상의 약물 처리 정보를 이용하여 후보 약물 정보를 생성할 수 있다. 다만 이에 한정되지 않고, 제1 후보 약물 결정 모듈(313)은 다양한 방식으로 후보 약물 정보를 생성할 수 있다.

[0083] 도 7는 본 개시의 몇몇 실시예에 따른 제2 분석 모델(320)을 설명하기 위한 도면이다. 도 8은 본 개시의 몇몇 실시예에 따른 제2 분석 모델(320)의 동작을 설명하기 위한 도면이다. 도 9은 본 개시의 몇몇 실시예에 따른 제2 분석 모델(320)의 동작을 설명하기 위한 또 다른 도면이다.

[0084] 몇몇 예에서, 제2 분석 모델(320)은 GSEA(Gene set enrichment analysis)를 통해 면역 아형을 전환하기 위한 후보 약물 정보를 생성할 수 있다. GSEA(Gene set enrichment analysis)는 마이크로어레이(microarray) 정량데이터 분석에서 제안된 방법론으로, 개개의 유전자의 정량적 변화가 아닌 기능 또는 집단 단위의 유전자 셋((gene set)의 정량적 변화를 측정함으로써 실험군, 대조군 간의 유의한 차이를 보이는 생물학적 변화를 평가하는 분석 방법론일 수 있다. 몇몇 예에서, 도 7를 참조하면, 제2 분석 모델(310)은 제1 GSEA 모델(321), 제2 GSEA 모델(322), 및 제2 후보 약물 결정 모듈(323)을 포함할 수 있다. 도 7에서 도시되는 제2 분석 모델(320)의 컴포넌트들은 단지 예로서 도시되며, 일부 컴포넌트가 생략되거나, 추가적인 컴포넌트가 포함될 수 있다.

[0085] 본 개시의 몇몇 제2 분석 모델(320)은 후보 약물과 기준 약물 사이의 유전자 셋(gene set)의 발현 수준 정보를 비교함으로써 제1 GSEA 스코어를 결정하는 제1 GSEA 모델(321)을 포함할 수 있다.

[0086] 구체적으로 설명하면, 제1 GSEA 모델(321)은 후보 약물과 기준 약물 사이의 유전자 셋의 발현 수준과의 관계를 나타내는 제1 GSEA 스코어를 획득하기 위한 모델일 수 있다. 예를 들어, 제1 GSEA 모델(321)은 제1 GSEA 스코어를 획득하기 위해 약물 처리 정보와 유전자(또는 유전자 셋)의 발현 수준 정보를 매칭한 조합 데이터를 이용할 수 있다. 몇몇 예에서, 제1 서브 모델(311)과 유사하게, 제1 GSEA 모델(321)은 약물과 유전자 간의 상호관계에 대한 정보를 제공할 수 있는 데이터베이스로부터 조합 데이터를 획득할 수 있다. 예를 들어, 제1 GSEA 모델

(321)은 통합된 네트워크 기반의 세포 시그니쳐 라이브러리(Library of Integrated Network-based Cellular Signatures, LINCS)으로부터 조합 데이터를 획득할 수 있다. 예를 들어, 제1 GSEA 모델(321)은 하버드 의과대학(HMS)에서 제공하는 링크스 프로그램(HMS LINCS, http://lincs.hms.harvard.edu/db/), 미국 국립 보건원(NIH)에서 제공하는 링크스 프로그램(NIH LINCS, www.lincsproject.org), DGIdb(The Drug Gene Interaction Database, www.dgidb.org) 등의 데이터베이스를 사용할 수 있으나, 이에 한정되지 않는다.

[0087] 본 개시의 몇몇 실시예에서, 제1 GSEA 모델(321)은 후보 약물 및 기준 약물에 대한 유전자 셋의 발현 수준 정보를 비교함으로써 제1 통계값을 획득할 수 있다. 여기서 제1 통계값은 P-value 또는 Q-value일 수 있다. 또한, 제1 GSEA 모델(321)은 후보 약물에 대한 유전자 셋의 발현 수준 정보의 제1 평균값을 획득할 수 있다. 또한, 제1 GSEA 모델(321)은 기준 약물에 대한 유전자 셋의 발현 수준 정보의 제2 평균값을 획득할 수 있다. 그리고, 제1 GSEA 모델(321)은 제1 통계값, 제1 평균값 및 제2 평균값에 기초하여 사전 결정된 식에 따라 제1 GSEA 스코어를 획득할 수 있다. 이 경우에, 사전 결정된 식에서, 제1 평균값과 제2 평균값의 제1 비교값은 후보 약물에 따른 유전자 셋의 발현 수준의 증감에 대응할 수 있다.

[0088] 도 8을 참조하면, LINCS 데이터를 이용하여 획득된 예시적인 제1 평균값(도 8의 (a) 참조, 제2 평균값(도 8의 (b) 참조), 및 제1 통계값(도 8의 (c)이 도시된다. 도 8에서 도시되는 값들은 LINCS 데이터로부터 획득된 약 1800개의 약물 반응에 대한 어레이 데이터를 SSGSEA(single sample gene enrichment analysis)를 진행한 데이터를 이용하여 획득되었다. 그리고, 여기서 기준 약물은 DMSO(Dimethyl Sulfoxide)이 선택되었다.

[0089] 본 개시의 몇몇 실시예에서, 사전 결정된 식은 하기의 식 (1)일 수 있다.

[0090] [식 (1)]

[0091] $$제1\ GSEA\ 스코어 = i \times -\log(제1\ 통계값 + 보정값)$$

[0092] 여기서, (제1평균값-제2 평균값) > 0 인 경우 i = 1이고, 그리고 (제1 평균값-제2 평균값) < 0 인 경우 i = -1일 수 있다.

[0093] 식 (1)을 참조하여 제1 GSEA 스코어를 구하는 예시적인 실시예가 설명된다. 제1 통계값은 후보 약물에 의해 발생하는 유전자 셋의 발현 수준의 변화량을 나타낼 수 있다. 몇몇 예에서, 제1 통계값은 LINCS 데이터로부터 획득된 후보 약물 및 기준 약물 간의 유전자 셋의 발현 수준을 비교함으로써 획득될 수 있다. 예를 들어, 후보 약물이 유전자 셋의 발현 수준이 크게 증가시키거나 또는 크게 감소시키는 경우에, 제1 통계값은 비교적 큰 값을 가질 수 있다. 다른 예로, 후보 약물이 유전자 셋의 발현 수준이 크게 변화시키지 않는 경우에, 제1 통계값은 비교적 작은 값을 가질 수 있다. 몇몇 예에서, 제1 통계값은 후보 약물 및 기준 약물 간의 유전자 셋의 발현 수준 정보를 이용하여 계산되는 P-value 또는 Q-value일 수 있다. 몇몇 예에서, 제1 통계값은 스케일링을 위해 -log(제1통계값)으로 제1 GSEA 스코어를 구하는데 사용될 수 있다. 다만 이에 한정되지 않고, 제1 통계값은 다양한 방식으로 획득될 수 있다.

[0094] 제1 GSEA 스코어가 후보 약물에 의해 유전자 셋의 발현 수준이 증가하는지 또는 감소하는지 나타내기 위해, 제1 평균값 및 제2 평균값이 계산될 수 있다. 여기서 제1 평균값은 후보 약물에 의한 유전자 셋의 발현 수준의 평균값일 수 있다. 그리고, 제2 평균값은 기준 약물에 의한 유전자 셋의 발현 수준의 평균값일 수 있다. 이 경우에, 제1 평균값 및 제2 평균값 사이의 제1 비교값은 후보 약물에 따른 유전자 셋의 발현 수준의 증감에 대응할 수 있다. 예를 들어, 제1 비교값은 후보 약물에 의한 유전자 셋의 발현 수준의 제1 평균값과 기준 약물에 의한 유전자 셋의 발현 수준의 제2 평균값의 차일 수 있다. 이 경우에, 양수인 제1 비교값은 후보 약물에 의해 유전자 셋의 발현 수준이 증가함을 나타낼 수 있다. 또한, 음수인 제1 비교값은 후보 약물에 의해 유전자 셋의 발현 수준이 감소함을 나타낼 수 있다.

[0095] 식 (1)을 참조하면, 제1 비교값이 양수인 경우, 제1 GSEA 스코어가 후보 약물에 의해 유전자 셋의 발현이 증가됨을 나타내도록, 'i = 1'로 결정될 수 있다. 이 경우에, 식 (1)를 통해 획득된 제1 GSEA 스코어는 후보 약물 처리에 의해 유전자 셋의 발현이 증가함을 나타내도록 양의 값을 가질 수 있다. 다른 예로, 제1 비교값이 음수인 경우, 제1 GSEA 스코어가 후보 약물에 의해 유전자 셋의 발현이 감소됨을 나타내도록, 'i = -1'로 결정될 수 있다. 이 경우에, 식 (1)를 통해 획득된 제1 GSEA 스코어는 후보 약물 처리에 의해 유전자 셋의 발현이 감소함을 나타내도록 음의 값을 가질 수 있다. 다만 이에 한정되지 않고, 제1 GSEA 모델(321)은 다양한 방식으로 제1 GSEA 스코어를 결정할 수 있다.

[0096] 본 개시의 몇몇 실시예에서, 제1 GSEA 모델(321)은 제1 GSEA 스코어를 보정(또는 검증)하기 위해 검증 데이터베

이스를 이용하여 보정값을 획득할 수 있다. 여기서, 보정값은, 제1 비교값과 정상 상태 및 암 상태 간의 유전자 발현 수준의 제2 비교값의 관계에 기초하여 정상 상태 및 암 상태에 대한 유전자 셋의 발현 수준 정보를 비교함으로써 획득되는 제2 통계값에 의해 제1 통계값을 보강 또는 상쇄하는 값일 수 있다.

[0097] 구체적으로 설명하면, 제1 GSEA 스코어는 제1 통계값을 획득하는데 사용된 데이터베이스와 상이한 검증 데이터베이스에 의해 결정되는 보정값에 의해 추가 검증될 수 있다. 보정값을 획득하는데 사용되는 검증 데이터베이스는 하나 이상의 공공 데이터(public data)를 포함할 수 있다. 예를 들어, 제1 통계값이 LINCS 데이터를 이용하여 획득되는 경우에, 검증 데이터베이스는 LINCS 데이터와 상이한 TCGA 데이터베이스일 수 있다. 다만 이에 한정되지 않고, 다양한 검증 데이터베이스가 사용될 수 있다.

[0098] 보정값은 제1 통계값을 보강 또는 상쇄함으로써 제1 GSEA 스코어를 보정하기 위해 제1 통계값과 연산 가능한 값일 수 있다. 예를 들어, 보정값은 제1 통계값을 획득하기 위해 사용된 데이터베이스와 상이한 검증 데이터베이스를 이용하여 획득되지만, 제1 통계값과 동일하거나 유사한 방식으로 획득될 수 있다. 예를 들어, 제1 통계값이 P-value 또는 Q-value인 경우에, 제2 통계값 또한 P-value 또는 Q-value일 수 있다.

[0099] 제1 GSEA 모델(321)은 보정값을 획득하기 위해 검증 데이터베이스로부터 획득되는 정상 상태 및 암 상태에 대한 유전자 셋의 발현 수준 정보를 비교함으로써 제2 통계값을 획득할 수 있다.

표 1

[0100]

	제2비교값	제2통계값	제1평균값		제2평균값		제1비교값	보정값
유전자 셋	정상 상태 - 암 상태 (i)	SSGSEA (P-value)	후보 약물	평균값	기준 약물	평균값	제1 평균값 - 제2 평균값 (j)	제2 통계값 x (i x j)
Gene Set1	0.4 (+)	7.43	Drug 1	0.1	DMSO	0.6	-0.5 (-)	-7.43
Gene Set2	-0.5 (-)	0.6	Drug 1	0.5	DMSO	0.1	0.4 (+)	-0.6
Gene Set3	0.3 (+)	3.4	Drug 1	0.3	DMSO	0.2	0.1 (+)	3.4
Gene Set4	0.5 (+)	0.3	Drug 1	0.3	DMSO	0.6	-0.3 (-)	-0.3
Gene Set5	-0.3 (-)	0.1	Drug 1	0.2	DMSO	0.4	-0.2 (-)	0.1

[0102] 상기 표 1을 참조하면, 검증 데이터베이스로부터 획득되는 정상 상태 및 암 상태의 유전자 발현 수준을 비교함으로써 획득되는 제2 통계값의 예시들이 도시된다. 여기서 제2 통계값은 검증 데이터베이스인 TCGA 데이터베이스로부터 획득된 데이터에 대하여 SSGSEA를 진행하여 획득되었다. 제2 통계값은 제1 통계값과 유사하게 유전자 발현 수준의 변화량과 비례하는 값을 가질 수 있다. 예를 들어, 정상 상태의 유전자 셋의 발현 수준이 암 상태에 비해 크게 증가되거나 감소되는 경우에, 제2 통계값은 비교적 큰 값을 가질 수 있다. 다른 예로, 정상 상태와 암 상태 간의 유전자 셋의 발현 수준이 크게 변하지 않는 경우에, 제2 통계값은 비교적 작은 값을 가질 수 있다. 몇몇 예에서, 제1 통계값과 제2 통계값이 상호 연산 가능하도록, 제2 통계값은 제1 통계값과 동일하게 P-value 또는 Q-value일 수 있다.

[0103] 추가적으로, 검증 데이터베이스로부터 획득되는 정상 상태 및 암 상태의 유전자 발현 수준은 제2 비교값을 획득하는데 사용될 수 있다. 제2 비교값은 제1 비교값과 유사하게 유전자 셋의 발현 수준의 증감에 대응할 수 있다. 예를 들어, 제2 비교값은 정상 상태의 유전자 셋의 발현 수준과 암 상태의 유전자 셋의 발현 수준의 차일 수 있다. 이 경우에, 양수인 제2 비교값은 정상 상태에서 유전자 셋의 발현 수준이 더 증가함을 나타낼 수 있다. 또한, 음수인 제2 비교값은 정상 상태에서 유전자 셋의 발현 수준이 감소함을 나타낼 수 있다.

[0104] 몇몇 예에서, 보정값은 제1 GSEA 스코어를 보정하기 위해 제2 통계값만큼 제1 통계값을 보강 또는 상쇄하는 값일 수 있다. 구체적으로 설명하면, TGGA 데이터베이스에 의해 획득되는 정상 상태 및 암 상태의 유전자 발현 수준의 변화 경향이 LINCS 데이터에 의해 획득되는 후보 약물에 의한 특정 유전자 발현 수준의 변화 경향과 상보적인 경우에(예를 들어, 정상 상태의 유전자 발현 수준이 암 상태의 유전자 발현 수준 보다 크고, 후보 약물이 유전자 발현 수준을 증가시키는 경우), 보정값은 제1 통계값을 보강하는 방향(예를 들어, 제1 통계값과 제2 통

계값이 합산됨)으로 결정될 수 있다. 또한, TGGA 데이터베이스에 의해 획득되는 정상 상태 및 암 상태의 유전자 발현 수준의 변화 경향이 LINCS 데이터에 의해 획득되는 후보 약물에 의한 특정 유전자 발현 수준의 변화 경향과 상충하는 경우에(예를 들어, 정상 상태의 유전자 발현 수준이 암 상태의 유전자 발현 수준 보다 크고, 후보 약물이 유전자 발현 수준을 감소시키는 경우), 보정값은 제1 통계값을 상쇄하는 방향(예를 들어, 제1 통계값에 제2 통계값이 감산됨)으로 결정될 수 있다.

[0105] 보정값이 제1 통계값을 보강 또는 상쇄하는지 여부는 제1 비교값과 제2 비교값의 관계에 의해 결정될 수 있다. 구체적으로 설명하면, LINCS 데이터에 기초한 제1 비교값은 후보 약물에 의한 유전자 셋의 발현 수준의 증감에 대응하고, TCGA 데이터베이스에 기초한 제2 비교값은 정상 상태(또는 암 상태)의 유전자 셋의 발현 수준의 증감에 대응할 수 있다. 이 경우에, 이하에서 설명하는 바와 같이, 제1 비교값과 제2 비교값의 관계를 이용함으로써 2개 이상의 상이한 데이터가 제1 GSEA 스코어를 결정하고 검증하는데 사용될 수 있다.

[0106] 구체적으로 예를 들면, 표 1을 참조하면, 유전자 셋 'GeneSet1'에 대한 정상 상태 및 암 상태 간의 유전자 발현 수준의 제2 비교값은 정상 상태 및 암 상태 간의 유전자 발현 수준의 차인 '0.4'일 수 있다. 이 경우에, 유전자 셋 'GeneSet_1'에 대한 제2 비교값은 정상 상태의 유전자 발현 수준이 암 상태 보다 더 높음을 나타낼 수 있다. 또한, 유전자 셋 'GeneSet1'에 대한 후보 약물 'Drug1'에 대한 유전자 셋의 발현 수준 정보의 제1 평균값(0.1)과, 기준 약물 'DMSO'에 대한 유전자 셋의 발현 수준 정보의 제2 평균값(0.6)의 제1 비교값은 두 값의 차인 '-0.5'일 수 있다. 따라서, 정상 상태인 경우에 유전자 셋 'GeneSet1'의 유전자 발현 수준은 암 상태에 비해 더 높지만, 후보 약물 'Drug1'은 유전자 셋 'GeneSet1'의 유전자 발현 수준을 감소시키므로, 제1 비교값과 제2 비교값의 관계는 서로 상충하는 관계일 수 있다. 따라서, 보정값은 제1 통계값을 상쇄하도록 '-제2통계값'으로 결정될 수 있다.

[0107] 다른 예로, 유전자 셋 'GeneSet3'에 대한 정상 상태 및 암 상태 간의 유전자 발현 수준의 제2 비교값은 정상 상태 및 암 상태 간의 유전자 발현 수준의 차인 '0.3'일 수 있다. 이 경우에, 유전자 셋 'GeneSet3'에 대한 제2 비교값은 정상 상태의 유전자 발현 수준이 암 상태 보다 더 높음을 나타낼 수 있다. 또한, 유전자 셋 'GeneSet3'에 대한 후보 약물 'Drug1'에 대한 유전자 셋의 발현 수준 정보의 제1 평균값(0.3) 및 기준 약물 'DMSO'에 대한 유전자 셋의 발현 수준 정보의 제2 평균값(0.2)의 제1 비교값은 두 값의 차인 '0.1'일 수 있다. 따라서, 유전자 셋 'GeneSet3'의 유전자 발현 수준은 암 상태에 비해 정상 상태일 때 더 높고, 후보 약물 'Drug1'은 유전자 셋 'GeneSet_1'의 유전자 발현 수준을 증가시키므로, 제1 비교값과 제2 비교값의 관계는 서로 상보적인 관계일 수 있다. 따라서, 보정값은 제1 통계값을 보강하도록 '+제2통계값'으로 결정될 수 있다. 다만 이에 한정되지 않고, 보정값은 다양한 방식으로 결정될 수 있다.

[0108] 본 개시의 몇몇 실시예에서, 제2 분석 모델(320)은 면역 아형에 따른 유전자 셋의 발현 수준 정보를 비교함으로써 제2 GSEA 스코어를 결정하는 제2 GSEA 모델(322)을 포함할 수 있다.

[0109] 구체적으로 설명하면, 제2 GSEA 모델(322)은 '면역 제외' 또는 '면역 결핍'인 경우에 유전자 셋의 발현 수준을 '면역 활성'인 경우에 유전자 셋의 발현 수준과 비교함으로써, 제2 GSEA 스코어를 결정할 수 있다. 몇몇 예에서, 제2 서브 모델(312)은 '면역 제외' 또는 '면역 결핍'인 경우에 유전자 셋의 제1 발현 수준 값과'면역 활성'인 경우에 유전자 셋의 제2 발현 수준 값의 차를 이용하여 제2 GSEA 스코어로 결정할 있다 (예를 들어, 제2 GSEA 스코어 = 제1 발현 수준 값 - 제2 발현 수준 값). 이 경우에, 양의 값을 가지는 제2 GSEA 스코어는 '면역 활성'인 경우에 발현이 감소함을 나타낼 수 있다. 또한, 음의 값을 가지는 제2 GSEA 스코어는 '면역 활성'인 경우에 발현이 증가함을 나타낼 수 있다. 다만 이에 한정되지 않고, 제2 GSEA 모델(322)은 다양한 방식으로 제2 GSEA 스코어를 결정할 수 있다.

[0110] 제2 GSEA 모델(322)은 각각의 면역 아형에 대한 유전자(또는 유전자 셋)의 발현 수준 정보를 획득하기 위해 생검 시료로부터 단백질 측정 장치를 이용하여 유전자 또는 단백질의 양을 확인한 데이터를 이용할 수 있다. 예를 들어, 제2 서브 모델(312)과 유사하게, 제2 GSEA 모델(322)은 삼중 음성 유방암(Triple Negative Breast Cancer) 환자를 대상으로 획득된 TNBC 데이터를 이용할 수 있다. 다만, 이에 한정되지 않고, 제2 서브 모델(312)은 다양한 데이터를 이용하여 유전자의 발현 수준 정보를 획득할 수 있다.

[0111] 본 개시의 몇몇 실시예에서, 제2 분석 모델(322)은 제1 GSEA 스코어와 제2 GSEA 스코어를 곱한 값을 이용하여 후보 약물 정보를 생성할 수 있다.

[0112] 구체적으로 설명하면, 제2 분석 모델(322)의 제2 후보 약물 결정 모듈(323)은 제1 GSEA 스코어 및 제2 GSEA 스코어를 이용하여 후보 약물 정보를 생성할 수 있다. 상술한 바와 같이, 양의 값을 가지는 제1 GSEA 스코어는 약

물 처리에 의해 유전자 셋의 발현이 증가함을 나타낼 수 있다. 또한, 음의 값을 가지는 제1 GSEA 스코어는 약물 처리에 의해 유전자 셋의 발현이 감소함을 나타낼 수 있다. 양의 값을 가지는 제2 GSEA 스코어는 '면역 활성'인 경우에 발현이 감소함을 나타낼 수 있다. 또한, 음의 값을 가지는 제2 GSEA 스코어는 '면역 활성'인 경우에 발현이 증가함을 나타낼 수 있다. 도 9를 참조하면, 제1 GSEA 스코어가 Y 좌표로 표시되고, 제2 GSEA 스코어가 X 좌표로 표시된 그래프가 도시된다. 제1 GSEA 스코어 및 제2 GSEA 스코어의 성질을 고려하면, 1 사분면 및 3 사분면에 분포되는 점들이 후보 약물 정보에 적합할 수 있다. 또한, 1 사분면 또는 3 사분면의 경우, 제1 GSEA 스코어 및 제2 GSEA 스코어는 동일한 부호를 가지므로, 제1 GSEA 스코어와 제2 GSEA 스코어의 곱은 양의 값을 가질 수 있다. 또한, 2 사분면 또는 4 사분면의 경우, 제1 GSEA 스코어 및 제2 GSEA 스코어는 서로 다른 부호를 가지므로, 제1 GSEA 스코어와 제2 GSEA 스코어의 곱은 음의 값을 가질 수 있다. 따라서, 제2 후보 약물 결정 모듈(323)은 제1 GSEA 스코어와 제2 GSEA 스코어의 곱이 큰 값을 가지는 약물 처리 정보를 이용하여 후보 약물 정보를 생성할 수 있다. 다만 이에 한정되지 않고, 제2 분석 모델(322)은 다양한 방식으로 제1 GSEA 스코어 및 제2 GSEA 스코어를 이용하여 후보 약물 정보를 생성할 수 있다.

[0113] 도 10은 본 개시의 몇몇 실시예에 따른 후보 약물 정보를 생성하는 방법의 순서도이다.

[0114] 본 개시의 몇몇 실시예에 따라, 후보 약물 정보를 생성하는 방법은 적어도 하나의 분석 모델을 포함하는 후보 약물 결정부에 의해, 면역 아형(immune subtype) 정보 및 바이오 마커 정보에 기초하여 후보 약물 정보를 생성하는 단계(s100)를 포함할 수 있다. 여기서, 상기 후보 약물 정보는 면역 제외 또는 면역 결핍에서 면역 활성으로 면역 아형을 전환하기 위한 바이오 마커의 발현 조절에 관련된 약물 정보를 포함할 수 있다.

[0115] 전술한 후보 약물 정보를 생성하는 방법의 단계는 단지 설명을 위해 제시된 것이며, 일부 단계가 생략되거나 별도의 단계가 추가될 수 있다. 또한, 전술한 단계들은 임의의 순서에 따라 수행될 수 있다.

[0116] 도 11은 본 개시의 몇몇 실시예에 따른 컴퓨팅 장치의 블록 구성도이다.

[0117] 도 11은 본 개시의 몇몇 실시예들이 구현될 수 있는 예시적인 컴퓨팅 환경에 대한 간략하고 일반적인 개략도를 도시한다.

[0118] 본 개시가 일반적으로 하나 이상의 컴퓨터 상에서 실행될 수 있는 컴퓨터 실행가능 명령어와 관련하여 전술되었지만, 당업자라면 본 개시가 기타 프로그램 모듈들과 결합되어 및/또는 하드웨어와 소프트웨어의 조합으로써 구현될 수 있다는 것을 잘 알 것이다.

[0119] 일반적으로, 프로그램 모듈은 특정의 태스크를 수행하거나 특정의 추상 데이터 유형을 구현하는 루틴, 프로그램, 컴포넌트, 데이터 구조, 기타 등등을 포함한다. 또한, 당업자라면 본 개시의 방법이 단일-프로세서 또는 멀티프로세서 컴퓨터 시스템, 미니컴퓨터, 메인프레임 컴퓨터는 물론 퍼스널 컴퓨터, 핸드 헬드 컴퓨팅 장치, 마이크로프로세서-기반 또는 프로그램가능 가전 제품, 기타 등등(이들 각각은 하나 이상의 연관된 장치와 연결되어 동작할 수 있음)을 비롯한 다른 컴퓨터 시스템 구성으로 실시될 수 있다는 것을 잘 알 것이다.

[0120] 본 개시의 설명된 실시예들은 또한 어떤 태스크들이 통신 네트워크를 통해 연결되어 있는 원격 처리 장치들에 의해 수행되는 분산 컴퓨팅 환경에서 실시될 수 있다. 분산 컴퓨팅 환경에서, 프로그램 모듈은 로컬 및 원격 메모리 저장 장치 둘 다에 위치할 수 있다.

[0121] 컴퓨터는 통상적으로 다양한 컴퓨터 판독가능 매체를 포함한다. 컴퓨터에 의해 액세스 가능한 매체는 그 어떤 것이든지 컴퓨터 판독가능 매체가 될 수 있다. 컴퓨터 판독가능 매체는 휘발성 및 비휘발성 매체, 일시적(transitory) 및 비일시적(non-transitory) 매체, 이동식 및 비-이동식 매체를 포함한다. 제한이 아닌 예로서, 컴퓨터 판독가능 매체는 컴퓨터 판독가능 저장 매체 및 컴퓨터 판독가능 전송 매체를 포함할 수 있다. 컴퓨터 판독가능 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보를 저장하는 임의의 방법 또는 기술로 구현되는 휘발성 및 비휘발성 매체, 일시적 및 비-일시적 매체, 이동식 및 비이동식 매체를 포함한다. 컴퓨터 저장 매체는 RAM, ROM, EEPROM, 플래시 메모리 또는 기타 메모리 기술, CD-ROM, DVD(digital video disk) 또는 기타 광 디스크 저장 장치, 자기 카세트, 자기 테이프, 자기 디스크 저장 장치 또는 기타 자기 저장 장치, 또는 컴퓨터에 의해 액세스될 수 있고 원하는 정보를 저장하는 데 사용될 수 있는 임의의 기타 매체를 포함하지만, 이에 한정되지 않는다.

[0122] 컴퓨터 판독가능 전송 매체는 통상적으로 기타 전송 메커니즘(transport mechanism)과 같은 피변조 데이터 신호(modulated data signal)에 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터 등을 구현하고 모든 정보 전달 매체를 포함한다. 피변조 데이터 신호라는 용어는 신호 내에 정보를 인코딩하도록 그 신호의 특성들 중 하나 이상을 설정 또는 변경시킨 신호를 의미한다. 제한이 아닌 예로서, 컴퓨터 판독가능 전송 매

체는 유선 네트워크 또는 직접 배선 접속(direct-wired connection)과 같은 유선 매체, 그리고 음향, RF, 적외선, 기타 무선 매체와 같은 무선 매체를 포함한다. 상술된 매체들 중 임의의 것의 조합도 역시 컴퓨터 판독가능 전송 매체의 범위 안에 포함되는 것으로 한다.

[0123] 컴퓨터(1102)를 포함하는 본 개시의 여러가지 측면들을 구현하는 예시 적인 환경(1100)이 나타내어져 있으며, 컴퓨터(1102)는 처리 장치(1104), 시스템 메모리(1106) 및 시스템 버스(1108)를 포함한다. 시스템 버스(1108)는 시스템 메모리(1106)(이에 한정되지 않음)를 비롯한 시스템 컴포넌트들을 처리 장치(1104)에 연결시킨다. 처리 장치(1104)는 다양한 상용 프로세서들 중 임의의 프로세서일 수 있다. 듀얼 프로세서 및 기타 멀티프로세서 아키텍처도 역시 처리 장치(1104)로서 이용될 수 있다.

[0124] 시스템 버스(1108)는 메모리 버스, 주변장치 버스, 및 다양한 상용 버스 아키텍처 중 임의의 것을 사용하는 로컬 버스에 추가적으로 상호 연결될 수 있는 몇 가지 유형의 버스 구조 중 임의의 것일 수 있다. 시스템 메모리(1106)는 판독 전용 메모리(ROM)(1110) 및 랜덤 액세스 메모리(RAM)(1112)를 포함한다. 기본 입/출력 시스템(BIOS)은 ROM, EPROM, EEPROM 등의 비휘발성 메모리(1110)에 저장되며, 이 BIOS는 시동 중과 같은 때에 컴퓨터(1102) 내의 구성요소들 간에 정보를 전송하는 일을 돕는 기본적인 루틴을 포함한다. RAM(1112)은 또한 데이터를 캐싱하기 위한 정적 RAM 등의 고속 RAM을 포함할 수 있다.

[0125] 컴퓨터(1102)는 또한 내장형 하드 디스크 드라이브(HDD)(1114)(예를 들어, EIDE, SATA)-이 내장형 하드 디스크 드라이브(1114)는 또한 적당한 섀시(도시 생략) 내에서 외장형 용도로 구성될 수 있음-, 자기 플로피 디스크 드라이브(FDD)(1116)(예를 들어, 이동식 디스켓(1118)으로부터 판독을 하거나 그에 기록을 하기 위한 것임), 및 광 디스크 드라이브(1120)(예를 들어, CD-ROM 디스크(1122)를 판독하거나 DVD 등의 기타 고용량 광 매체로부터 판독을 하거나 그에 기록을 하기 위한 것임)를 포함한다. 하드 디스크 드라이브(1114), 자기 디스크 드라이브(1116) 및 광 디스크 드라이브(1120)는 각각 하드 디스크 드라이브 인터페이스(1124), 자기 디스크 드라이브 인터페이스(1126) 및 광 드라이브 인터페이스(1128)에 의해 시스템 버스(1108)에 연결될 수 있다. 외장형 드라이브 구현을 위한 인터페이스(1124)는 USB(Universal Serial Bus) 및 IEEE 1394 인터페이스 기술 중 적어도 하나 또는 그 둘 다를 포함한다.

[0126] 이들 드라이브 및 그와 연관된 컴퓨터 판독가능 매체는 데이터, 데이터 구조, 컴퓨터 실행가능 명령어, 기타 등등의 비 휘발성 저장을 제공한다. 컴퓨터(1102)의 경우, 드라이브 및 매체는 임의의 데이터를 적당한 디지털 형식으로 저장하는 것에 대응한다. 상기에서의 컴퓨터 판독가능 매체에 대한 설명이 HDD, 이동식 자기 디스크, 및 CD 또는 DVD 등의 이동식 광 매체를 언급하고 있지만, 당업자라면 집 드라이브(zip drive), 자기 카세트, 플래쉬 메모리 카드, 카트리지, 기타 등등의 컴퓨터에 의해 판독가능한 다른 유형의 매체도 역시 예시 적인 운영 환경에서 사용될 수 있으며 또 임의의 이러한 매체가 본 개시의 방법들을 수행하기 위한 컴퓨터 실행가능 명령어를 포함할 수 있다는 것을 잘 알 것이다.

[0127] 운영 체제(1130), 하나 이상의 애플리케이션 프로그램(1132), 기타 프로그램 모듈(1134) 및 프로그램 데이터(1136)를 비롯한 다수의 프로그램 모듈이 드라이브 및 RAM(1112)에 저장될 수 있다. 운영 체제, 애플리케이션, 모듈 및/또는 데이터의 전부 또는 그 일부분이 또한 RAM(1112)에 캐싱될 수 있다. 본 개시가 여러가지 상업적으로 이용가능한 운영 체제 또는 운영 체제들의 조합에서 구현될 수 있다는 것을 잘 알 것이다.

[0128] 사용자는 하나 이상의 유선/무선 입력 장치, 예를 들어, 키보드(1138) 및 마우스(1140) 등의 포인팅 장치를 통해 컴퓨터(1102)에 명령 및 정보를 입력할 수 있다. 기타 입력 장치(도시 생략)로는 마이크, IR 리모콘, 조이스틱, 게임 패드, 스타일러스 펜, 터치 스크린, 기타 등등이 있을 수 있다. 이들 및 기타 입력 장치가 종종 시스템 버스(1108)에 연결되어 있는 입력 장치 인터페이스(1142)를 통해 처리 장치(1104)에 연결되지만, 병렬 포트, IEEE 1394 직렬 포트, 게임 포트, USB 포트, IR 인터페이스, 기타 등등의 기타 인터페이스에 의해 연결될 수 있다.

[0129] 모니터(1144) 또는 다른 유형의 디스플레이 장치도 역시 비디오 어댑터(1146) 등의 인터페이스를 통해 시스템 버스(1108)에 연결된다. 모니터(1144)에 부가하여, 컴퓨터는 일반적으로 스피커, 프린터, 기타 등등의 기타 주변 출력 장치(도시 생략)를 포함한다.

[0130] 컴퓨터(1102)는 유선 및/또는 무선 통신을 통한 원격 컴퓨터(들)(1148) 등의 하나 이상의 원격 컴퓨터로의 논리적 연결을 사용하여 네트워크화 된 환경에서 동작할 수 있다. 원격 컴퓨터(들)(1148)는 워크스테이션, 컴퓨팅 디바이스 컴퓨터, 라우터, 퍼스널 컴퓨터, 휴대용 컴퓨터, 마이크로프로세서-기반 오락 기기, 피어 장치 또는 기타 통상의 네트워크 노드일 수 있으며, 일반적으로 컴퓨터(1102)에 대해 기술된 구성요소들 중 다수 또는 그

전부를 포함하지만, 간략함을 위해, 메모리 저장 장치(1150)만이 도시되어 있다. 도시되어 있는 논리적 연결은 근거리 통신망(LAN)(1152) 및/또는 더 큰 네트워크, 예를 들어, 원거리 통신망(WAN)(1154)에의 유선/무선 연결을 포함한다. 이러한 LAN 및 WAN 네트워킹 환경은 사무실 및 회사에서 일반적인 것이며, 인트라넷 등의 전사적 컴퓨터 네트워크(enterprise-wide computer network)를 용이하게 해주며, 이들 모두는 전세계 컴퓨터 네트워크, 예를 들어, 인터넷에 연결될 수 있다.

[0131] LAN 네트워킹 환경에서 사용될 때, 컴퓨터(1102)는 유선 및/또는 무선 통신 네트워크 인터페이스 또는 어댑터(1156)를 통해 로컬 네트워크(1152)에 연결된다. 어댑터(1156)는 LAN(1152)에의 유선 또는 무선 통신을 용이하게 해줄 수 있으며, 이 LAN(1152)은 또한 무선 어댑터(1156)와 통신하기 위해 그에 설치되어 있는 무선 액세스 포인트를 포함하고 있다. WAN 네트워킹 환경에서 사용될 때, 컴퓨터(1102)는 모뎀(1158)을 포함할 수 있거나, WAN(1154) 상의 통신 컴퓨팅 디바이스에 연결되거나, 또는 인터넷을 통하는 등, WAN(1154)을 통해 통신을 설정하는 기타 수단을 갖는다. 내장형 또는 외장형 및 유선 또는 무선 장치일 수 있는 모뎀(1158)은 직렬 포트 인터페이스(1142)를 통해 시스템 버스(1108)에 연결된다. 네트워크화 된 환경에서, 컴퓨터(1102)에 대해 설명된 프로그램 모듈들 또는 그의 일부분이 원격 메모리/저장 장치(1150)에 저장될 수 있다. 도시된 네트워크 연결이 예시 적인 것이며 컴퓨터들 사이에 통신 링크를 설정하는 기타 수단이 사용될 수 있다는 것을 잘 알 것이다.

[0132] 컴퓨터(1102)는 무선 통신으로 배치되어 동작하는 임의의 무선 장치 또는 개체, 예를 들어, 프린터, 스캐너, 데스크톱 및/또는 휴대용 컴퓨터, PDA(portable data assistant), 통신 위성, 무선 검출가능 태그와 연관된 임의의 장비 또는 장소, 및 전화와 통신을 하는 동작을 한다. 이것은 적어도 Wi-Fi 및 블루투스 무선 기술을 포함한다. 따라서, 통신은 종래의 네트워크에서와 같이 미리 정의된 구조이거나 단순하게 적어도 2개의 장치 사이의 애드혹 통신(ad hoc communication)일 수 있다.

[0133] Wi-Fi(Wireless Fidelity)는 유선 없이도 인터넷 등으로의 연결을 가능하게 해준다. Wi-Fi는 이러한 장치, 예를 들어, 컴퓨터가 실내에서 및 실외에서, 즉 기지국의 통화권 내의 아무 곳에서나 데이터를 전송 및 수신할 수 있게 해주는 셀 전화와 같은 무선 기술이다. Wi-Fi 네트워크는 안전하고 신뢰성 있으며 고속인 무선 연결을 제공하기 위해 IEEE 802.11(a,b,g, 기타)이라고 하는 무선 기술을 사용한다. 컴퓨터를 서로에, 인터넷에 및 유선 네트워크(IEEE 802.3 또는 이더넷을 사용함)에 연결시키기 위해 Wi-Fi가 사용될 수 있다. Wi-Fi 네트워크는 비인가 2.4 및 5 GHz 무선 대역에서, 예를 들어, 11Mbps(802.11a) 또는 54 Mbps(802.11b) 데이터 레이트로 동작하거나, 양 대역(듀얼 대역)을 포함하는 제품에서 동작할 수 있다.

[0134] 본 개시의 기술 분야에서 통상의 지식을 가진 자는 정보 및 신호들이 임의의 다양한 상이한 기술들 및 기법들을 이용하여 표현될 수 있다는 것을 이해할 것이다. 예를 들어, 위의 설명에서 참조될 수 있는 데이터, 지시들, 명령들, 정보, 신호들, 비트들, 심볼들 및 칩들은 전압들, 전류들, 전자기파들, 자기장들 또는 입자들, 광학장들 또는 입자들, 또는 이들의 임의의 결합에 의해 표현될 수 있다.

[0135] 본 개시의 기술 분야에서 통상의 지식을 가진 자는 여기에 개시된 실시예들과 관련하여 설명된 다양한 예시 적인 논리 블록들, 모듈들, 프로세서들, 수단들, 회로들 및 알고리즘 단계들이 전자 하드웨어, (편의를 위해, 여기에서 "소프트웨어"로 지칭되는) 다양한 형태들의 프로그램 또는 설계 코드 또는 이들 모두의 결합에 의해 구현될 수 있다는 것을 이해할 것이다. 하드웨어 및 소프트웨어의 이러한 상호 호환성을 명확하게 설명하기 위해, 다양한 예시 적인 컴포넌트들, 블록들, 모듈들, 회로들 및 단계들이 이들의 기능과 관련하여 위에서 일반적으로 설명되었다. 이러한 기능이 하드웨어 또는 소프트웨어로서 구현되는지 여부는 특정한 애플리케이션 및 전체 시스템에 대하여 부과되는 설계 제약들에 따라 좌우된다. 본 개시의 기술 분야에서 통상의 지식을 가진 자는 각각의 특정한 애플리케이션에 대하여 다양한 방식들로 설명된 기능을 구현할 수 있으나, 이러한 구현 결정들은 본 개시의 범위를 벗어나는 것으로 해석되어서는 안 될 것이다.

[0136] 여기서 제시된 다양한 실시예들은 방법, 장치, 또는 표준 프로그래밍 및/또는 엔지니어링 기술을 사용한 제조 물품(article)으로 구현될 수 있다. 용어 "제조 물품"은 임의의 컴퓨터-판독가능 장치로부터 액세스 가능한 컴퓨터 프로그램 또는 매체(media)를 포함한다. 예를 들어, 컴퓨터-판독가능 매체는 자기 저장 장치(예를 들면, 하드 디스크, 플로피 디스크, 자기 스트립, 등), 광학 디스크(예를 들면, CD, DVD, 등), 스마트 카드, 및 플래쉬 메모리 장치(예를 들면, EEPROM, 카드, 스틱, 키 드라이브, 등)를 포함하지만, 이들로 제한되는 것은 아니다. 또한, 여기서 제시되는 다양한 저장 매체는 정보를 저장하기 위한 하나 이상의 장치 및/또는 다른 기계-판독가능한 매체를 포함한다.

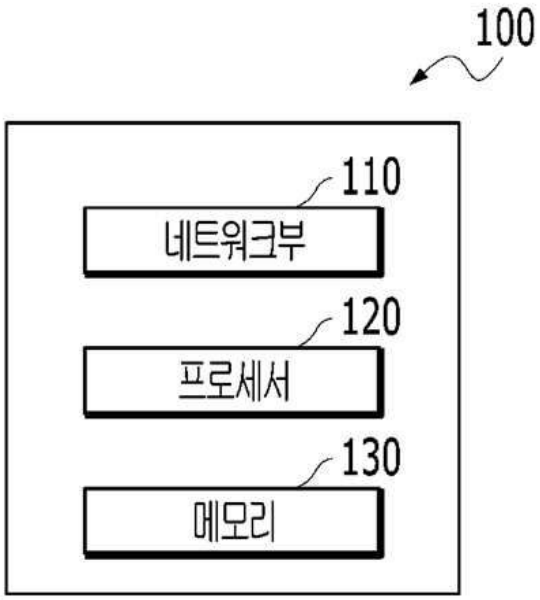
[0137] 제시된 프로세스들에 있는 단계들의 특정한 순서 또는 계층 구조는 예시 적인 접근들의 일례임을 이해하도록 한다. 설계 우선순위들에 기반하여, 본 개시의 범위 내에서 프로세스들에 있는 단계들의 특정한 순서 또는 계층

구조가 재배열될 수 있다는 것을 이해하도록 한다. 첨부된 방법 청구항들은 샘플 순서로 다양한 단계들의 엘리먼트들을 제공하지만 제시된 특정한 순서 또는 계층 구조에 한정되는 것을 의미하지는 않는다.

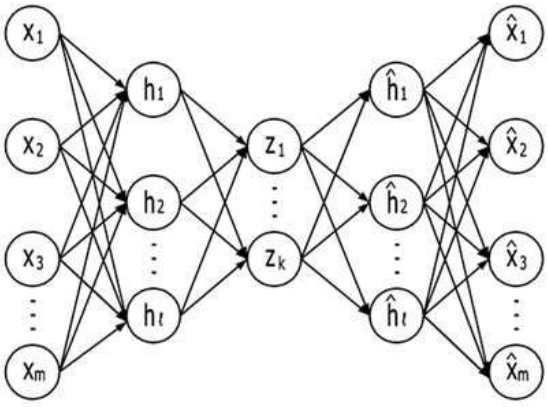
[0138] 제시된 실시예들에 대한 설명은 임의의 본 개시의 기술 분야에서 통상의 지식을 가진 자가 본 개시를 이용하거나 또는 실시할 수 있도록 제공된다. 이러한 실시예들에 대한 다양한 변형들은 본 개시의 기술 분야에서 통상의 지식을 가진 자에게 명백할 것이다. 여기에 정의된 일반적인 원리들은 본 개시의 범위를 벗어남이 없이 다른 실시예들에 적용될 수 있다. 그리하여, 본 개시는 여기에 제시된 실시예들로 한정되는 것이 아니라, 여기에 제시된 원리들 및 신규한 특징들과 일관되는 최광의의 범위에서 해석되어야 할 것이다.

도면

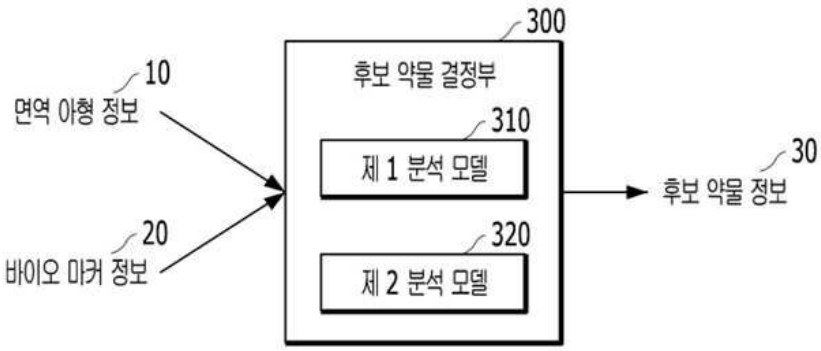
도면1



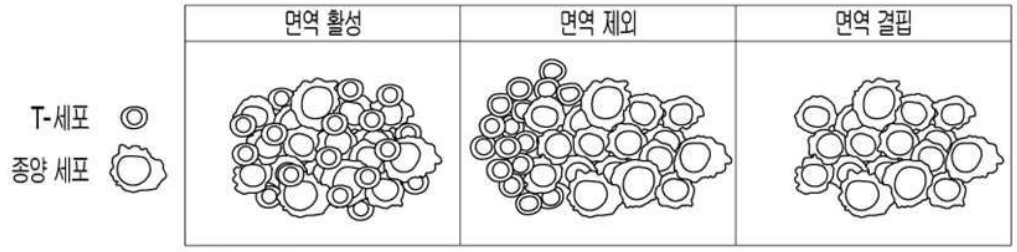
도면2



도면3


300
후보 약물 결정부
10
면역 아형 정보
310
제 1 분석 모델
30
후보 약물 정보
20
바이오 마커 정보
320
제 2 분석 모델

도면4


면역 활성
면역 제외
면역 결핍
T-세포
종양 세포

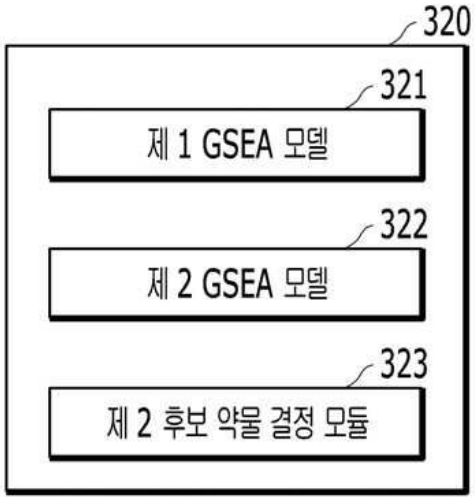
도면5


310
311
제 1 서브 모델
312
제 2 서브 모델
313
제 1 후보 약물 결정 모듈

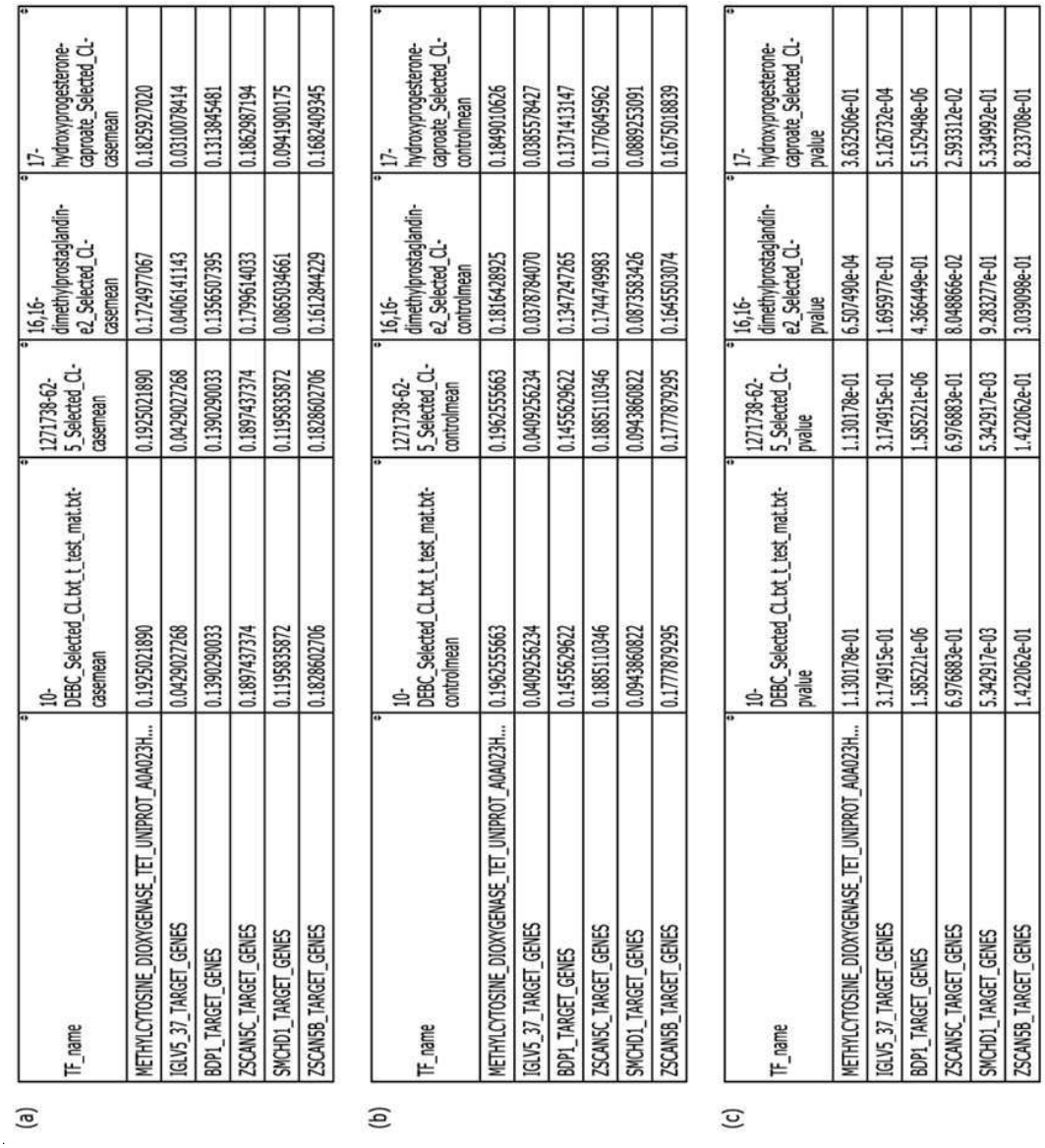
도면6

	Drug_1	Drug_2	Drug_3	Drug_4	Drug_5	Drug_6
A2M	-0.03873	0.28607	0.28607	0.05587	0.28607	0.05587
NAT1	0.00331	-0.12111	-0.12111	-0.18896	-0.12111	-0.18896
NAT2	-0.12195	-0.02230	-0.02230	-0.01975	-0.02230	-0.01975
SERPINA3	0.04685	-0.06873	-0.06873	0.03917	-0.06873	0.03917
AADAC	0.18844	0.19165	0.19165	0.07437	0.19165	0.07437
AAMP	-0.02062	0.13111	0.13111	-0.02977	0.13111	-0.02977
AANAT	0.06465	0.12731	0.12731	0.17195	0.12731	0.17195
AARS1	0.09062	0.10196	0.10196	0.36474	0.10196	0.36474
ABAT	-0.00334	-0.06402	-0.06402	0.02799	-0.06402	0.02799
ABCA1	-0.13162	-0.09097	-0.09097	0.00683	-0.09097	0.00683
ABCA2	-0.07438	-0.03323	-0.03323	0.05044	-0.03323	0.05044

도면7



도면8



(a)

TF_name	10-DEBC_Selected_CL.txt_t_test_mat.txt-casemean	1271738-62-5_Selected_CL-casemean	16,16-dimethylprostaglandin-e2_Selected_CL-casemean	17-hydroxyprogesterone-caproate_Selected_CL-casemean
METHYLCYTOSINE_DIOXYGENASE_TET_UNIPROT_A0A023H...	0.1925021890	0.1925021890	0.1724977067	0.1825927020
IGLV5_37_TARGET_GENES	0.0429027268	0.0429027268	0.0406141143	0.0310078414
BDP1_TARGET_GENES	0.1390290033	0.1390290033	0.1356507395	0.1313845481
ZSCAN5C_TARGET_GENES	0.1897437374	0.1897437374	0.1799614033	0.1862987194
SMCHD1_TARGET_GENES	0.1195835872	0.1195835872	0.0865034661	0.0941900175
ZSCAN5B_TARGET_GENES	0.1828602706	0.1828602706	0.1612844229	0.1682409345

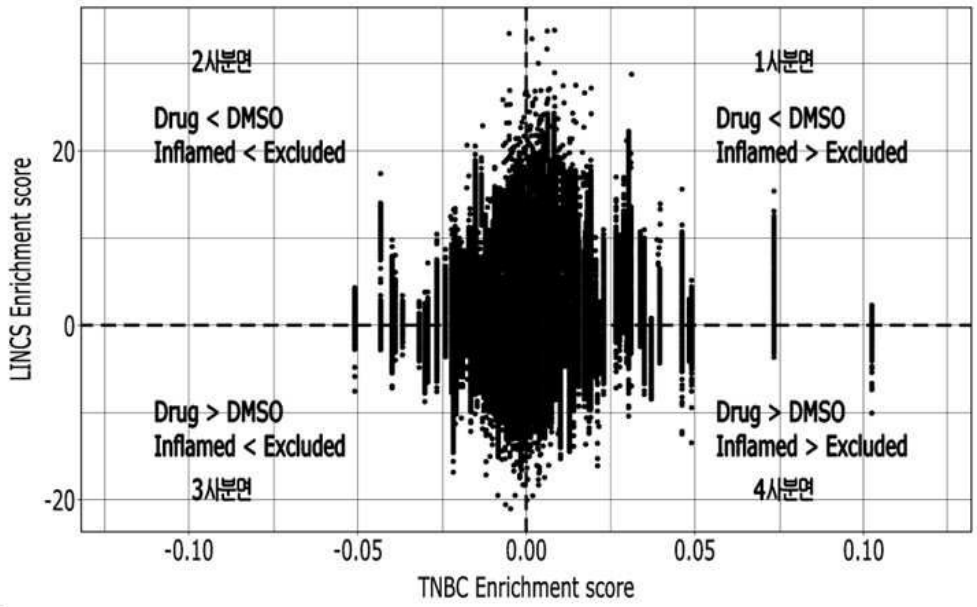
(b)

TF_name	10-DEBC_Selected_CL.txt_t_test_mat.txt-controlmean	1271738-62-5_Selected_CL-controlmean	16,16-dimethylprostaglandin-e2_Selected_CL-controlmean	17-hydroxyprogesterone-caproate_Selected_CL-controlmean
METHYLCYTOSINE_DIOXYGENASE_TET_UNIPROT_A0A023H...	0.1962555663	0.1962555663	0.1816428925	0.1849010626
IGLV5_37_TARGET_GENES	0.0409256234	0.0409256234	0.0378784070	0.0385578427
BDP1_TARGET_GENES	0.1455629622	0.1455629622	0.1347247265	0.1371413147
ZSCAN5C_TARGET_GENES	0.1885110346	0.1885110346	0.1744749983	0.1776045962
SMCHD1_TARGET_GENES	0.0943860822	0.0943860822	0.0873583426	0.0889253091
ZSCAN5B_TARGET_GENES	0.1777879295	0.1777879295	0.1645503074	0.1675018839

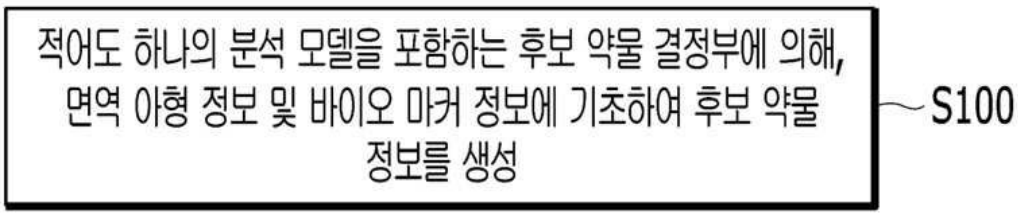
(c)

TF_name	10-DEBC_Selected_CL.txt_t_test_mat.txt-pvalue	1271738-62-5_Selected_CL-pvalue	16,16-dimethylprostaglandin-e2_Selected_CL-pvalue	17-hydroxyprogesterone-caproate_Selected_CL-pvalue
METHYLCYTOSINE_DIOXYGENASE_TET_UNIPROT_A0A023H...	1.130178e-01	1.130178e-01	6.507490e-04	3.632506e-01
IGLV5_37_TARGET_GENES	3.174915e-01	3.174915e-01	1.695977e-01	5.126732e-04
BDP1_TARGET_GENES	1.585221e-06	1.585221e-06	4.366449e-01	5.152948e-06
ZSCAN5C_TARGET_GENES	6.976883e-01	6.976883e-01	8.048866e-02	2.593312e-02
SMCHD1_TARGET_GENES	5.342917e-03	5.342917e-03	9.283277e-01	5.334992e-01
ZSCAN5B_TARGET_GENES	1.422062e-01	1.422062e-01	3.039098e-01	8.233708e-01

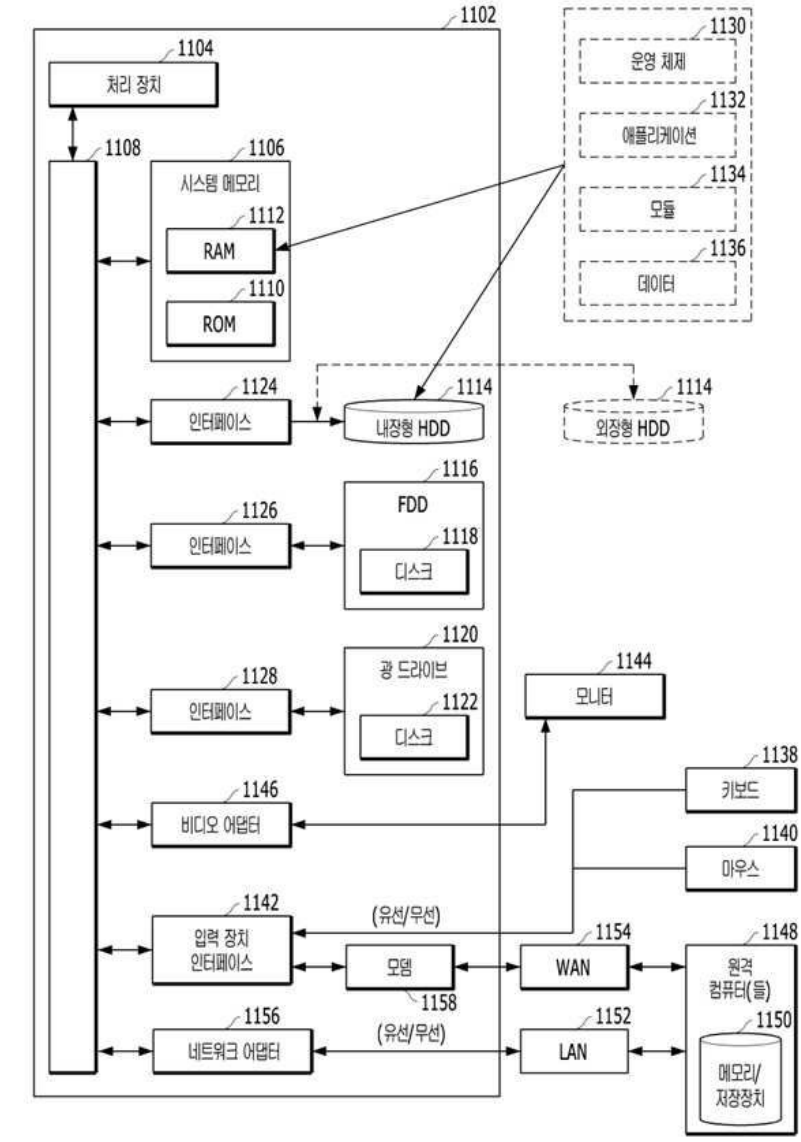
도면9


2사분면
Drug < DMSO
Inflamed < Excluded
1사분면
Drug < DMSO
Inflamed > Excluded
Drug > DMSO
Inflamed < Excluded
3사분면
Drug > DMSO
Inflamed > Excluded
4사분면
LINCS Enrichment score
20
0
-20
-0.10
-0.05
0.00
0.05
0.10
TNBC Enrichment score

도면10


적어도 하나의 분석 모델을 포함하는 후보 약물 결정부에 의해, 면역 아형 정보 및 바이오 마커 정보에 기초하여 후보 약물 정보를 생성
S100

도면11


1102
1104
처리 장치
1108
1106
시스템 메모리
1112
RAM
1110
ROM
1130
운영 체제
1132
애플리케이션
1134
모듈
1136
데이터
1124
인터페이스
1114
내장형 HDD
1114
외장형 HDD
1116
FDD
1126
인터페이스
1118
디스크
1120
광 드라이브
1128
인터페이스
1122
디스크
1144
모니터
1138
키보드
1146
비디오 어댑터
1140
마우스
1142
입력 장치 인터페이스
(유선/무선)
모뎀
1158
1154
WAN
1148
원격 컴퓨터(들)
1156
네트워크 어댑터
(유선/무선)
1152
LAN
1150
메모리/저장장치