



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0081114
(43) 공개일자 2024년06월07일

(51) 국제특허분류(Int. Cl.)
G16B 25/10 (2019.01) C12Q 1/6883 (2018.01)
G16H 50/50 (2018.01)
(52) CPC특허분류
G16B 25/10 (2019.02)
C12Q 1/6883 (2022.01)
(21) 출원번호 10-2022-0165060
(22) 출원일자 2022년11월30일
심사청구일자 2022년11월30일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
윤정
서울시 구로구 구로동로 148 고대 구로병원 진단검사의학과
김정안
경기도 안양시 동안구 부림로 80, 609동 705호
(74) 대리인
특허법인(유)화우

전체 청구항 수 : 총 32 항

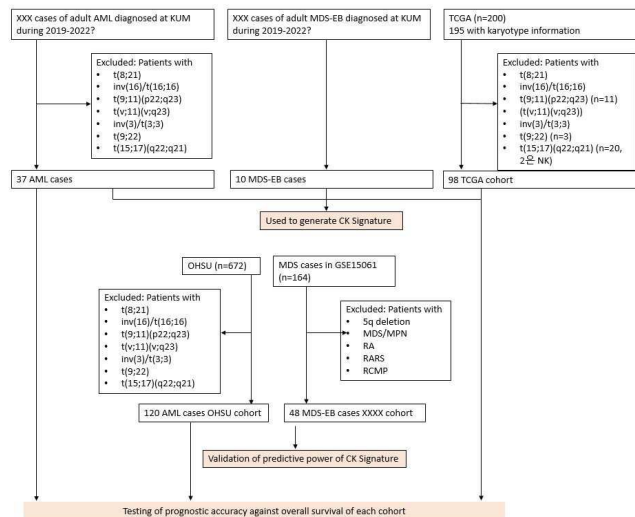
(54) 발명의 명칭 유전자 발현 기반의 급성골수성백혈병 및 모세포증가 골수이형성증후군의 복잡핵형 예측 모델

(57) 요약

본 발명은 복잡핵형 특이적 유전자를 규명한 것으로, 복잡핵형 특이적 유전자의 발현 수준을 측정하는 체제를 포함하는 복잡핵형 예측용 조성물 및 복잡핵형 예측을 위한 정보제공방법에 관한 것이다.

본 발명의 복잡핵형 예측용 조성물 및 복잡핵형 예측을 위한 정보제공방법은 세포배양 과정없이 신속하고 정확하게 복잡핵형 급성골수성백혈병 및 모세포증가 골수이형성증후군을 조기 예측할 수 있고, 결과적으로 급성골수성백혈병 및 모세포증가 골수이형성증후군의 예후를 예측하고 복잡핵형 환자에게 적합한 치료방식을 적용할 수 있다.

대표도 - 도1



(52) CPC특허분류

C12Q 1/6886 (2022.01)

G16H 50/50 (2018.01)

C12Q 2537/165 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711154986
과제번호	2022R1G1A1007629
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)
연구과제명	복잡핵형 급성골수성백혈병의 신호전달 조절 이상 분석 및 치료 target 발굴
기 여 율	1/1
과제수행기관명	고려대학교
연구기간	2022.03.01 ~ 2023.02.28

명세서

청구범위

청구항 1

- (a) 생물학적 시료로부터 SLC18A2, FAM13B 및 LINC01138 유전자의 발현 수준을 측정하는 단계;
- (b) 상기 (a)단계에서 측정된 상기 유전자의 발현 수준에 회귀계수(weight)값을 곱하여 합산하여 결정값을 얻는 단계; 및
- (c) 상기 (b)단계에서 얻은 결정값을 기반으로 복잡핵형 예측하는 단계를 포함하는, 복잡핵형 예측을 위한 정보제공방법.

청구항 2

제1항에 있어서,

상기 (a)단계는 생물학적 시료로부터 PBX3, RPS14, CCL23, SOD1 및 TSPAN5 유전자의 발현 수준을 측정하는 단계를 추가로 포함하는 것인, 복잡핵형 예측을 위한 정보제공방법.

청구항 3

제1항에 있어서,

상기 회귀계수(weight)값은 SLC18A2의 경우 20 내지 25이거나, FAM13B의 경우, -85 내지 -75이거나, LINC01138의 경우, -35 내지 -30 인, 복잡핵형 예측을 위한 정보제공방법.

청구항 4

제2항에 있어서,

상기 회귀계수(weight)값은 PBX3의 경우 -2 내지 1이거나, RPS14의 경우, -6 내지 -2이거나, CCL23의 경우, -3 내지 -1이거나, SOD1의 경우 0.5 내지 5이거나, TSPAN5의 경우 0.5 내지 2인, 복잡핵형 예측을 위한 정보제공방법.

청구항 5

제1항에 있어서,

상기 결정값은 상기 (a)단계에서 측정된 각각의 유전자 발현 수준에 회귀계수(weight)값을 곱하여 합산하여 얻어진 결정값에 보정계수 값을 합한 것인, 복잡핵형 예측을 위한 정보제공방법.

청구항 6

제5항에 있어서,

상기 보정계수 값은 15 내지 50인 것인, 복잡핵형 예측을 위한 정보제공방법.

청구항 7

제1항에 있어서,

상기 (b)단계의 결정값은 하기 수학적 식 1로 표현되는 것인, 복잡핵형 예측을 위한 정보제공방법.

[수학적 식 1]

$$45.23 + (SLC18A2 \times 23.48) + (FAM13B \times -81.50) + (LINC01138 \times -34.03)$$

청구항 8

제2항에 있어서,

상기 (b)단계의 결정값은 하기 수학적 식 2 또는 3으로 표현되는 것인, 복잡핵형 예측을 위한 정보제공방법.

[수학적 식 2]

$$17.8670 + (PBX3 \times -1.0679) + (RPS14 \times -2.8674) + (CCL23 \times -1.3630) + (SOD1 \times 0.9259) + (TSPAN5 \times 1.0268) + (SLC18A2 \times 0.4810) + (FAM13B \times 1.2650) + (LINC01138 \times -3.0967)$$

[수학적 식 3]

$$48.3971 + (PBX3 \times -0.2931) + (RPS14 \times -5.2680) + (CCL23 \times -2.5812) + (SOD1 \times 4.5024) + (TSPAN5 \times 1.4217) + (SLC18A2 \times -0.5298) + (FAM13B \times -4.9628) + (LINC01138 \times -6.9592)$$

청구항 9

제1항에 있어서,

상기 생물학적 시료는 급성골수성백혈병(acute myeloid leukemia) 또는 모세포증가 골수이형성증후군(myelodysplastic syndrome with excessice blasts) 환자로부터 분리된 것인, 복잡핵형 예측을 위한 정보제공방법.

청구항 10

제1항 내지 제8항 중 어느 한 항에 있어서,

상기 결정값이 0.5 이상인 경우에는 복잡핵형으로 예측하는 것인, 복잡핵형 예측을 위한 정보제공방법.

청구항 11

(a) 생물학적 시료로부터 SLC18A2, FAM13B 및 LINC01138 유전자의 발현 수준을 측정하는 단계;

(b) 상기 (a)단계에서 측정된 상기 유전자의 발현 수준에 회귀계수(weight)값을 곱하여 합산하여 결정값을 얻는 단계; 및

(c) 상기 (b)단계에서 얻은 결정값을 기반으로 복잡핵형을 예측하는 단계를 포함하는, 급성골수성백혈병(acute myeloid leukemia) 또는 모세포증가 골수이형성증후군(myelodysplastic syndrome with excessice blasts) 환자의 예후 예측을 위한 정보제공방법.

청구항 12

제11항에 있어서,

상기 (a)단계는 생물학적 시료로부터 PBX3, RPS14, CCL23, SOD1 및 TSPAN5 유전자의 발현 수준을 측정하는 단계를 추가로 포함하는 것인, 급성골수성백혈병(acute myeloid leukemia) 또는 모세포증가 골수이형성증후군

(myelodysplastic syndrome with excessice blasts) 환자의 예후 예측을 위한 정보제공방법.

청구항 13

제11항에 있어서,

상기 회귀계수(weight)값은 SLC18A2의 경우 20 내지 25이거나, FAM13B의 경우, -85 내지 -75이거나, LINC01138의 경우, -35 내지 -30 인, 급성골수성백혈병(acute myeloid leukemia) 또는 모세포증가 골수이형성증후군(myelodysplastic syndrome with excessice blasts) 환자의 예후 예측을 위한 정보제공방법.

청구항 14

제12항에 있어서,

상기 회귀계수(weight)값은 PBX3의 경우 -2 내지 1이거나, RPS14의 경우, -6 내지 -2이거나, CCL23의 경우, -3 내지 -1이거나, SOD1의 경우 0.5 내지 5이거나, TSPAN5의 경우 0.5 내지 2인, 급성골수성백혈병(acute myeloid leukemia) 또는 모세포증가 골수이형성증후군(myelodysplastic syndrome with excessice blasts) 환자의 예후 예측을 위한 정보제공방법.

청구항 15

제11항에 있어서,

상기 결정값은 상기 (a)단계에서 측정된 각각의 유전자 발현 수준에 회귀계수(weight)값을 곱하여 합산하여 얻어진 결정값에 보정계수 값을 합한 것인, 급성골수성백혈병(acute myeloid leukemia) 또는 모세포증가 골수이형성증후군(myelodysplastic syndrome with excessice blasts) 환자의 예후 예측을 위한 정보제공방법.

청구항 16

제15항에 있어서,

상기 보정계수 값은 15 내지 50인 것인, 급성골수성백혈병(acute myeloid leukemia) 또는 모세포증가 골수이형성증후군(myelodysplastic syndrome with excessice blasts) 환자의 예후 예측을 위한 정보제공방법.

청구항 17

제11항에 있어서,

상기 (b)단계의 결정값은 하기 수학적 식 1로 표현되는 것인, 급성골수성백혈병(acute myeloid leukemia) 또는 모세포증가 골수이형성증후군(myelodysplastic syndrome with excessice blasts) 환자의 예후 예측을 위한 정보 제공방법.

[수학적 식 1]

$$45.23 + (SLC18A2 \times 23.48) + (FAM13B \times -81.50) + (LINC01138 \times -34.03)$$

청구항 18

제12항에 있어서,

상기 (b)단계의 결정값은 하기 수학적 식 2 또는 3으로 표현되는 것인, 급성골수성백혈병(acute myeloid leukemia) 또는 모세포증가 골수이형성증후군(myelodysplastic syndrome with excessice blasts) 환자의 예후 예측을 위

한 정보제공방법.

[수학식 2]

$$17.8670 + (PBX3 \times -1.0679) + (RPS14 \times -2.8674) + (CCL23 \times -1.3630) + (SOD1 \times 0.9259) + (TSPAN5 \times 1.0268) + (SLC18A2 \times 0.4810) + (FAM13B \times 1.2650) + (LINC01138 \times -3.0967)$$

[수학식 3]

$$48.3971 + (PBX3 \times -0.2931) + (RPS14 \times -5.2680) + (CCL23 \times -2.5812) + (SOD1 \times 4.5024) + (TSPAN5 \times 1.4217) + (SLC18A2 \times -0.5298) + (FAM13B \times -4.9628) + (LINC01138 \times -6.9592)$$

청구항 19

제11항 내지 제18항 중 어느 한 항에 있어서,

상기 결정값이 0.5이상인 경우에는 예후가 나쁠 것으로 판단하는 것인, 급성골수성백혈병(acute myeloid leukemia) 또는 모세포증가 골수이형성증후군(myelodysplastic syndrome with excessice blasts) 환자의 예후 예측을 위한 정보제공방법.

청구항 20

SLC18A2, FAM13B 및 LINC01138 유전자의 발현 수준을 측정하는 제제를 포함하는 복잡핵형 예측용 조성물.

청구항 21

제20항에 있어서,

PBX3, RPS14, CCL23, SOD1 및 TSPAN5 유전자의 발현 수준을 측정하는 제제를 추가적으로 포함하는, 복잡핵형 예측용 조성물.

청구항 22

SLC18A2, FAM13B 및 LINC01138 유전자의 발현 수준을 측정하는 제제를 포함하는 급성골수성백혈병(acute myeloid leukemia) 또는 모세포증가 골수이형성증후군(myelodysplastic syndrome with excessice blasts) 환자의 예후 예측용 조성물.

청구항 23

제22항에 있어서,

PBX3, RPS14, CCL23, SOD1 및 TSPAN5 유전자의 발현 수준을 측정하는 제제를 추가적으로 포함하는, 예후 예측용 조성물.

청구항 24

(a) 생물학적 시료로부터 SLC18A2, FAM13B 및 LINC01138 유전자의 발현 수준을 측정하고 정규화하는 검출부;

(b) 상기 검출부에서의 각각의 유전자 발현 수준에 회귀계수(weight)값을 곱하여 합산하여 얻어진 결정값을 계산하는 연산부; 및

(c) 상기 연산부에서 얻어진 결정값으로 복잡핵형을 예측하여 출력하는 출력부를 포함하는, 복잡핵형 예측용 기

기.

청구항 25

제24항에 있어서,

상기 검출부에서 PBX3, RPS14, CCL23, SOD1 및 TSPAN5 유전자의 발현 수준을 추가로 더 측정하는 것인, 복잡핵형 예측용 기기.

청구항 26

제24항에 있어서,

상기 회귀계수(weight)값은 SLC18A2의 경우 20 내지 25이거나, FAM13B의 경우, -85 내지 -75이거나, LINC01138의 경우, -35 내지 -30인, 복잡핵형 예측용 기기.

청구항 27

제25항에 있어서,

상기 회귀계수(weight)값은 PBX3의 경우 -2 내지 1이거나, RPS14의 경우, -6 내지 -2이거나, CCL23의 경우, -3 내지 -1이거나, SOD1의 경우 0.5 내지 5이거나, TSPAN5의 경우 0.5 내지 2인 복잡핵형 예측용 기기.

청구항 28

제24항에 있어서,

상기 결정값은 상기 검출부에서의 각각의 유전자 발현 수준에 회귀계수(weight)값을 곱하여 합산하여 얻어진 결정값에 보정계수 값을 합한 것인, 복잡핵형 예측용 기기.

청구항 29

제28항에 있어서,

상기 보정계수 값은 15 내지 50인 것인, 복잡핵형 예측용 기기.

청구항 30

제24항에 있어서,

상기 (b)단계에서의 결정값은 하기 수학적 식 1로 표현되는 것인, 복잡핵형 예측용 기기.

[수학적 식 1]

$$45.23 + (SLC18A2 \times 23.48) + (FAM13B \times -81.50) + (LINC01138 \times -34.03)$$

청구항 31

제25항에 있어서,

상기 (b)단계에서의 결정값은 하기 수학적 식 2 또는 3으로 표현되는 것인, 복잡핵형 예측용 기기.

[수학식 2]

$$17.8670 + (\text{PBX3} \times -1.0679) + (\text{RPS14} \times -2.8674) + (\text{CCL23} \times -1.3630) + (\text{SOD1} \times 0.9259) + (\text{TSPAN5} \times 1.0268) + (\text{SLC18A2} \times 0.4810) + (\text{FAM13B} \times 1.2650) + (\text{LINC01138} \times -3.0967)$$

[수학식 3]

$$48.3971 + (\text{PBX3} \times -0.2931) + (\text{RPS14} \times -5.2680) + (\text{CCL23} \times -2.5812) + (\text{SOD1} \times 4.5024) + (\text{TSPAN5} \times 1.4217) + (\text{SLC18A2} \times -0.5298) + (\text{FAM13B} \times -4.9628) + (\text{LINC01138} \times -6.9592)$$

청구항 32

제24 내지 제31항 중 어느 한 항에 있어서,

상기 결정값이 0.5 이상인 경우에는 복잡핵형으로 예측하는 것인, 복잡핵형 예측용 기기.

발명의 설명

기술 분야

[0001] 본 발명은 유전자 발현 기반의 급성골수성백혈병 및 모세포증가 골수이형성증후군의 복잡핵형 예측 모델에 관한 것이다.

배경 기술

[0003] 복잡핵형(Complex Karyotype, CK)은 WHO가 지정한 재발성 전위 또는 역위, 즉 t(8;21), inv(16), t(16;16), t(9;11), t(v;11)(v;q23.3), t(6;9), inv(3) or t(3;3) 및 BCR-ABL을 가진 AML(ELN 2017) 중 하나가 없는 상태에서 3개 이상의 염색체 이상을 가진 것으로 정의된다. 이러한 염색체 이상은 기원을 알 수 없는 염색체 물질, 마커 또는 고리 염색체, 균일하게 염색된 영역 또는 이중극미(double minutes, dm)를 가진 불균형 전위를 포함한다. 일반적으로, 복잡핵형은 균형잡힌 전좌보다는 염색체의 증가 및 결실을 특징으로 하며, 백혈병 발생에 기여하는 것으로 알려져있다. 현재 WHO 체계는 20% blast 임계값을 기준으로 재발성 유전적 이상이 없는 AML과 MDS-EB를 이분법적으로 시행하고 있다.

[0004] 2017 European LeukemiaNet 유전적 위험 계층화에 따르면, 복잡핵형은 AML 및 MDS-EB 환자 모두에서 위험도 및 나쁜 예후와 관련이 있으며, BCL-2 억제제인 베네토클락스(venetoclax)와 저메틸화제인 데시타빈은 AML/MDS 모두에 대해 우수한 항종양 활성을 나타내는 것으로 알려져 있다. 이러한 결과는 백혈병 분류에 있어서 복잡핵형이 가치가 있음을 시사한다.

[0005] AML-CK 및 MDS-CK 환자는 신규 AML 및 원발성 MDS의 각각 20% 및 30%를 차지하므로, 치료 실패 및 재발에 취약한 이러한 유해 위험군의 큰 하위 집합을 구성한다. 따라서, 복잡핵형의 분자적 특징에 대한 연구는 종양 성장에 관한 병인 과정에 대한 통찰력을 제공하고, 결과적으로 효과적인 치료법을 제공할 수 있다.

[0006] AML 및 MDS-EB의 유전적 특성화는 핵형 분석에 의해 수행되었다. 핵형 분석법은 변형에 대한 범의학적 관점을 제공하기는 하나, 소요 시간이 길고, 실패하기 쉬운 세포 배양 단계, 제한적인 분석능 및 주관적인 시각적 해석 등의 문제점을 가지고 있다. 특히, 복잡핵형은 종종 표준의 세포유전학적 기술을 사용할 경우, 부분적 또는 전체적으로 해석될 수 없는 수많은 염색체 이상을 포함한다.

[0007] RNA-Seq 분석은 전사체 발현을 정량화하고, 새로운 융합 전사체를 발견하고, 심지어 골수 악성 종양의 발병 기전과 관련이 있는 체세포 돌연변이를 검출하는 능력을 포함하는 잠재적인 임상 이점을 제공한다. 이러한 RNA-seq의 장점과 함께, 최근 몇 년간 시퀀싱 비용이 감소함에 따라 RNA-seq는 핵형 분석을 포함한 현재 테스트 알고리즘을 대체할 수 있을 것으로 기대된다. 최근 한 연구에서는 RNA-seq 데이터에서 융합 유전자, 변이체, 순차 중복(tandem duplications) 및 유전자 발현의 검출을 통합하는 AML 진단을 위한 단일 생물정보학 파이프라인을 제안하였다.

[0008] 이러한 배경하에, 본 발명은 RNA-seq 기반으로 복잡핵형을 분석할 수 있는지 및 임상적 연관성을 확인할 수 있는지를 평가하기 위해, 비-복잡핵형 하위그룹과 구별되는 복잡핵형과 연관된 특이적 예측 유전자군을 발굴하고, 피험자의 전체 생존율에 대한 모델의 예측 성능을 평가하였다.

선행기술문헌

특허문헌

[0010] (특허문헌 0001) 유럽 공개특허 제2458014호

발명의 내용

해결하려는 과제

- [0011] 본 발명은 전술한 문제 및 이와 연관된 다른 문제를 해결하는 것을 목적으로 한다.
- [0012] 본 발명의 일 예시적 목적은 유전자 발현 기반의 급성골수성백혈병 및 모세포증가 골수이형성증후군의 복잡핵형 예측 모델을 제공하는 것이다.
- [0013] 본 발명의 다른 예시적 목적은 복잡핵형 예측용 조성물을 제공하는 것이다.
- [0014] 본 발명의 또 다른 예시적 목적은 복잡핵형 예측을 위한 정보제공방법을 제공하는 것이다.
- [0016] 본 명세서에 개시된 발명의 기술적 사상에 따라 이루고자 하는 기술적 과제는 이상에서 언급한 문제점을 해결하기 위한 과제로 제한되지 않으며, 언급되지 않은 또 다른 과제는 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

- [0018] 이를 구체적으로 설명하면 다음과 같다. 한편, 본 출원에서 개시된 각각의 설명 및 실시형태는 각각의 다른 설명 및 실시 형태에도 적용될 수 있다. 즉, 본 출원에서 개시된 다양한 요소들의 모든 조합이 본 출원의 범주에 속한다. 또한, 하기 기술된 구체적인 서술에 의하여 본 출원의 범주가 제한된다고 볼 수 없다.
- [0020] 상기 목적을 달성하기 위한 일 양태로서, 본 발명은 복잡핵형 예측을 위한 정보제공방법을 제공한다.
- [0021] 상기 복잡핵형 예측을 위한 정보제공방법은 (a) 생물학적 시료로부터 SLC18A2, FAM13B 및 LINC01138 유전자의 발현 수준을 측정하는 단계; (b) 상기 (a)단계에서 측정된 상기 유전자의 발현 수준에 회귀계수(weight)값을 곱하여 합산하여 결정값을 얻는 단계; 및 (c) 상기 (b)단계에서 얻은 결정값을 기반으로 복잡핵형을 예측하는 단계를 포함한다.
- [0022] 본 발명에서 용어 "복잡핵형(complex karyotype)"은 염색체의 구조적 또는 수적 이상으로, 세 가지 이상의 염색체 이상을 동반하는 경우를 의미하며, 대부분의 복잡핵형은 염색체 소실, 결손과 같은 이상을 포함하고 있으며, 5번, 7번 염색체 또는 두 염색체 모두의 이상을 동반하는 경우가 많은 것으로 보고되어 있다. 특히, 복잡핵형은 급성골수성백혈병 또는 골수이형성증후군의 나쁜 예후 및 BCL2 억제제인 베네토클락스(venetoclax)와 같은 특정 치료제에 의한 치료반응과 연관이 있는 것으로 알려져 있다. 상기 복잡핵형은 구체적으로 급성골수성백혈병 또는 모세포증가 골수이형성증후군에서의 복잡핵형을 의미할 수 있으나 이에 제한되는 것은 아니다.
- [0023] 상기 생물학적 시료는 개체로부터 분리되거나 개체로부터 유래된 임의의 물질, 생물학적 체액, 조직 또는 세포를 의미하는 것으로, 예를 들면, 전혈(whole blood), 백혈구(leukocytes), 말초혈액 단핵 세포(peripheral blood mononuclear cells), 백혈구 연층(buffy coat), 골수(bone marrow), 혈장(plasma), 혈청(serum), 객담(sputum), 눈물(tears), 점액(mucus), 세비액(nasal washes), 비강 흡인물(nasal aspirate), 호흡(breath), 소

변(urine), 정액(semen), 침(saliva), 복강 세척액(peritoneal washings), 골반 내 유체액(pelvic fluids), 낭종액(cystic fluid), 뇌척수막 액(meningeal fluid), 양수(amniotic fluid), 선액(glandular fluid), 췌장액(pancreatic fluid), 림프액(lymph fluid), 흉수(pleural fluid), 유두 흡인물(nipple aspirate), 기관지 흡인물(bronchial aspirate), 활액(synovial fluid), 관절 흡인물(joint aspirate), 기관 분비물(organ secretions), 세포(cell), 세포 추출물(cell extract) 또는 뇌척수액(cerebrospinal fluid)을 포함할 수 있으나, 이에 제한되는 것은 아니다.

- [0024] 상기 "개체"는 복잡핵형을 예측하기 위한 대상이 되는 인간을 포함한 포유동물일 수 있고, 구체적으로는 복잡핵형과 백혈병 또는 골수이형성증후군 환자일 수 있고, 보다 바람직하게는 급성골수성백혈병 또는 모세포증가 골수이형성증후군 환자일 수 있으나, 이에 제한되는 것은 아니다.
- [0025] 본 발명에서 용어 "급성골수성백혈병(Acute Myeloid Leukemia, AML)"은 급성백혈병의 일종으로, 급성백혈병의 약 65%를 차지하며, 발병 후 수일에서 수주 내 사망에 이를 수 있어 신속하고 정확한 진단과 치료가 중요한 질환으로 알려져 있다. 급성골수성 백혈병에서 가장 흔하게 관찰되는 염색체 이상은 복잡핵형으로, 급성골수성 백혈병의 매우 나쁜 예후와 연관이 있고, 표준항암치료시 관해 도달 확률이 50%이며, 5년 생존율이 10~20%에 불과한 것으로 보고되어 있다.
- [0026] 본 발명에서 용어 "모세포증가 골수이형성증후군"은 모세포의 증가가 관찰되는 골수이형성증후군으로, 약 25~33%가 급성백혈병으로 진행되며, 생존율도 9~16개월로 예후가 매우 나쁜 것으로 알려져 있다.
- [0027] 발현 수준을 측정하는 특정 유전자들은 복잡핵형-특이적인 발현 차이를 보이는 유전자, 즉 3개 미만의 관련없는 염색체 이상을 가진 급성골수성백혈병 환자 및 모세포증가 골수이형성증후군 환자와 비교하여 복잡핵형의 환자에서 발현 차이를 나타내는 유전자로서, 구체적으로 본 발명에서는 SLC18A2, FAM13B, LINC01138, PBX3, RPS14, CCL23, SOD1 또는 TSPAN5 유전자일 수 있다.
- [0028] 상기 유전자의 발현 수준은 당분야에 공지된 방법으로 측정될 수 있으며, 예를 들면 노던 블로팅, PCR, RT-PCR, 실시간 PCR 또는 마이크로어레이 분석 등을 이용하여 수행될 수 있으나 이에 제한되는 것은 아니다.
- [0029] 본 발명의 일 예시로서, 상기 유전자의 발현 수준 측정은 급성골수성백혈병 또는 모세포증가 골수이형성증후군 환자 조직으로부터 추출한 RNA를 이용하여 당 분야에 공지된 방법인 일반적인 RNA-seq 또는 마이크로어레이를 수행하여, 이로부터 얻은 발현 데이터를 분석하는 방식으로 이루어질 수 있다.
- [0030] 상기 (a)단계는 생물학적 시료로부터 PBX3, RPS14, CCL23, SOD1 및 TSPAN5 유전자의 발현 수준을 측정하는 단계를 추가로 포함할 수 있다.
- [0031] 상기 (b)단계에서 특정 유전자의 발현 수준에 곱해지는 회귀계수(weight)값은 유전자별로 상이하며, 구체적으로 SLC18A2의 경우 15 내지 30이거나, FAM13B의 경우, -100 내지 -60이거나, LINC01138의 경우, -50 내지 -10일 수 있고, PBX3의 경우 -5 내지 3이거나, RPS14의 경우, -7 내지 -0.5이거나, CCL23의 경우, -5 내지 -0.5이거나, SOD1의 경우 0.3 내지 6이거나, TSPAN5의 경우 0.5 내지 3의 유리수일 수 있다.
- [0032] 보다 바람직하게는 SLC18A2의 경우 20 내지 25이거나, FAM13B의 경우, -85 내지 -75이거나, LINC01138의 경우, -35 내지 -30이거나, PBX3의 경우 -2 내지 1이거나, RPS14의 경우, -6 내지 -2이거나, CCL23의 경우, -3 내지 -1이거나, SOD1의 경우 0.5 내지 5이거나, TSPAN5의 경우 0.5 내지 2의 유리수일 수 있으나, 이에 제한되는 것은 아니다.
- [0033] 본 발명의 일 예시로서, 회귀계수 값은 각각 SLC18A2의 경우, 23.48, FAM13B의 경우, -81.50, LINC01138의 경우, -34.03일 수 있고, PBX3의 경우 -1.0679 또는 -0.2931이거나, RPS14의 경우 -2.8674 또는 -5.2680이거나, CCL23의 경우, -1.3630 또는 -2.5812이거나, SOD1의 경우 0.9259 또는 4.5024이거나, TSPAN5의 경우 1.0268 또는 1.4217일 수 있다.
- [0034] 상기 (b)단계에서의 '결정값'은 상기 (a)단계에서 측정된 각각의 유전자 발현 수준에 회귀계수(weight)값을 곱하여 합산하여 얻어진 결정값에 보정계수 값을 합하여 구할 수 있는 것으로, 상기 보정계수 값은 10 내지 60의 유리수일 수 있고, 보다 바람직하게는 15 내지 50의 유리수일 수 있으나, 이에 제한되는 것은 아니다.
- [0035] 본 발명의 일 예시로서, 보정계수 값은 45.23, 17.8670 또는 48.3971일 수 있다.
- [0036] 상기 (b) 단계의 결정값은 하기 수학적 식 1 내지 3 중 어느 하나의 수학적 식으로 표현되는 것일 수 있고, 구체적으로는 (a)단계에서 측정된 상기 유전자의 발현 수준을 하기 수학적 식 1 내지 3 중 어느 하나의 수학적 식의 각각의 유

전자에 대입하여 복잡핵형을 예측하는 것일 수 있다.

- [0037] [수학식 1]
- [0038] $45.23 + (SLC18A2 \times 23.48) + (FAM13B \times -81.50) + (LINC01138 \times -34.03)$
- [0039] [수학식 2]
- [0040] $17.8670 + (PBX3 \times -1.0679) + (RPS14 \times -2.8674) + (CCL23 \times -1.3630) + (SOD1 \times 0.9259) + (TSPAN5 \times 1.0268) + (SLC18A2 \times 0.4810) + (FAM13B \times 1.2650) + (LINC01138 \times -3.0967)$
- [0041] [수학식 3]
- [0042] $48.3971 + (PBX3 \times -0.2931) + (RPS14 \times -5.2680) + (CCL23 \times -2.5812) + (SOD1 \times 4.5024) + (TSPAN5 \times 1.4217) + (SLC18A2 \times -0.5298) + (FAM13B \times -4.9628) + (LINC01138 \times -6.9592)$
- [0043] 상기 (b) 단계의 결정값, 즉 수학식 1 내지 3 중 어느 하나의 수학식으로부터 도출된 결과가 0.5 이상인 경우에는 복잡핵형으로 예측하는 것일 수 있다.
- [0045] 상기 목적을 달성하기 위한 다른 양태로서, 본 발명은 복잡핵형을 예측하는 단계를 포함하는, 급성골수성백혈병 또는 모세포증가 골수이형성증후군 환자의 예후 예측을 위한 정보제공방법을 제공한다.
- [0046] 상기 급성골수성백혈병 또는 모세포증가 골수이형성증후군 환자의 예후 예측을 위한 정보제공방법은 (a) 생물학적 시료로부터 SLC18A2, FAM13B 및 LINC01138 유전자의 발현 수준을 측정하는 단계; (b) 상기 (a) 단계에서 측정된 상기 유전자의 발현 수준에 회귀계수(weight)값을 곱하여 합산하여 결정값을 얻는 단계; 및 (c) 상기 (b) 단계에서 얻은 결정값을 기반으로 복잡핵형을 예측하는 단계를 포함한다.
- [0047] 상기 (a) 단계는 생물학적 시료로부터 PBX3, RPS14, CCL23, SOD1 및 TSPAN5 유전자의 발현 수준을 측정하는 단계를 추가로 포함할 수 있다.
- [0048] 용어 '복잡핵형', '급성골수성백혈병', '모세포증가 골수이형성증후군', '회귀계수값', '결정값'은 전술한 바와 같다.
- [0049] 상기 (b) 단계의 결정값, 즉 전술한 수학식 1 내지 3 중 어느 하나의 수학식으로부터 도출된 결과가 0.5 이상인 경우에는 예후가 나쁠 것으로 판단하는 것일 수 있다.
- [0051] 상기 목적을 달성하기 위한 다른 양태로서, 본 발명은 SLC18A2, FAM13B 및 LINC01138 유전자의 발현 수준을 측정하는 제제를 포함하는 복잡핵형 예측용 조성물을 제공한다.
- [0052] 상기 복잡핵형 예측용 조성물은 PBX3, RPS14, CCL23, SOD1 및 TSPAN5 유전자의 발현 수준을 측정하는 제제를 추가적으로 포함할 수 있다.
- [0053] 상기 복잡핵형 예측용 조성물은 복잡핵형-특이적인 발현 차이, 즉 3개 미만의 관련없는 염색체 이상을 가진 급성골수성백혈병 환자 및 모세포증가 골수이형성증후군 환자와 비교하여 복잡핵형의 환자에서 발현 차이를 나타내는 유전자의 발현 수준을 측정하는 제제를 포함할 수 있다.
- [0054] 본 발명의 일 예시로서, 상기 복잡핵형 예측용 조성물은 SLC18A2, FAM13B, LINC01138, PBX3, RPS14, CCL23, SOD1 또는 TSPAN5 유전자의 발현 수준을 측정하는 제제를 포함할 수 있다.
- [0055] 본 발명의 일 예시로서, 상기 복잡핵형 예측용 조성물은 SLC18A2, FAM13B 및 LINC01138 유전자 중 어느 하나 이상의 유전자의 발현 수준을 측정하는 제제를 포함할 수 있다.
- [0056] 본 발명의 일 예시로서, 상기 복잡핵형 예측용 조성물은 PBX3, RPS14, CCL23, SOD1 및 TSPAN5 중 어느 하나 이상의 유전자의 발현 수준을 측정하는 제제를 포함할 수 있다.
- [0057] 상기 유전자의 발현 수준을 측정하는 제제는 각 유전자의 발현 수준 또는 발현량을 측정할 수 있는 물질이라면 그 종류에는 제한되지 않으나, 일 예시로 각 유전자에 특이적으로 결합하는 프라이머 또는 프로브일 수 있다.
- [0058] 본 발명에서 용어 "프라이머"는 표적 유전자 서열을 인지하는 단편으로서, 정방향 및 역방향의 프라이머 쌍을 포함하나, 바람직하게는, 특이성 및 민감성을 가지는 분석 결과를 제공하는 프라이머 쌍이다. 프라이머의 핵산

서열이 시료 내 존재하는 비-표적 서열과 불일치하는 서열이어서, 상보적인 프라이머 결합 부위를 함유하는 표적 유전자 서열만 증폭하고 비특이적 증폭을 유발하지 않는 프라이머일 때, 높은 특이성이 부여될 수 있다.

- [0059] 본 발명에서 용어 "프로브"는 시료 내의 검출하고자 하는 표적 물질과 결합하여 특이적으로 시료 내의 표적 물질의 존재를 확인할 수 있는 물질을 의미한다. 프로브의 종류는 당업계에서 통상적으로 사용되는 물질로서 제한되지 않으며, 예를 들면 PNA(peptide nucleic acid), LNA(locked nucleic acid), 펩타이드, 폴리펩타이드, 단백질, RNA 또는 DNA일 수 있다. 또한, 상기 프로브는 바이오 물질로서 생물에서 유래되거나 이와 유사한 것 또는 생체 외에서 제조된 것을 포함하는 것으로, 예를 들어, 효소, 단백질, 항체, 미생물, 동식물 세포 및 기관, 신경세포, DNA, 및 RNA일 수 있으며, DNA는 cDNA, 게놈 DNA, 올리고뉴클레오타이드를 포함하며, RNA는 게놈 RNA, mRNA, 올리고뉴클레오타이드를 포함하며, 단백질의 예로는 항체, 항원, 효소, 펩타이드 등을 포함할 수 있다.
- [0060] 본 발명의 일 예시로서, 상기 유전자의 발현 수준을 전술한 수학적 식 1 내지 3 중 어느 하나의 수학적 식의 각각의 유전자에 대입하여 도출된 결과가 0.5 이상인 경우에는 복잡핵형으로 예측할 수 있다.
- [0061] 상기 목적을 달성하기 위한 다른 양태로서, 본 발명은 복잡핵형 예측용 기기를 제공한다.
- [0062] 상기 복잡핵형 예측용 기기는 (a) 생물학적 시료로부터 SLC18A2, FAM13B 및 LINC01138 유전자의 발현 수준을 측정하고 정규화하는 검출부; (b) 상기 검출부에서의 각각의 유전자 발현 수준에 회귀계수(weight)값을 곱하여 합산하여 얻어진 결정값을 계산하는 연산부; 및 (c) 상기 연산부에서 얻어진 결정값으로 복잡핵형을 예측하여 출력하는 출력부를 포함한다.
- [0063] 상기 검출부는 PBX3, RPS14, CCL23, SOD1 및 TSPAN5 유전자의 발현 수준을 추가로 더 측정하는 것일 수 있다.
- [0064] 용어 '복잡핵형', '생물학적 시료', '유전자 발현 수준 측정', '회귀계수값', '결정값'은 전술한 바와 같다.
- [0065] 상기 복잡핵형 예측용 기기는 혈액, 침, 소변, 골수 등 인체에서 생성된 물질을 기반으로 체외에서 질환 진단이 가능한 장비를 의미하며, 예를 들어 검출부; 연산부; 및 출력부 등을 포함하며, 상기 물질로부터 유전자의 발현 수준을 측정하거나 분석 등을 수행할 수 있는 형태라면 제한이 없다.
- [0066] 상기 복잡핵형 예측용 기기는 복잡핵형을 예측하기 위한 대상이 되는 개체의 복잡핵형을 예측할 수 있을 뿐만 아니라, 이를 바탕으로, 개체의 예후 또는 치료반응을 예측하는 데에 사용될 수 있다.

발명의 효과

- [0068] 본 발명의 복잡핵형 예측용 조성물 및 복잡핵형 예측을 위한 정보제공방법은 세포배양 과정없이 신속하고 정확하게 복잡핵형 급성골수성백혈병 및 모세포증가 골수이형성증후군을 조기 예측할 수 있고, 결과적으로 급성골수성백혈병 및 모세포증가 골수이형성증후군의 예후를 예측하고 복잡핵형 환자에게 적합한 치료방식을 적용할 수 있다.

도면의 간단한 설명

- [0070] 도 1은 복잡핵형 예측 모델 개발을 위한 연구 설계 개요를 나타낸 것으로, 환자 코호트 선별 및 필터링을 포함한 모델 개발, 검증 및 임상 결과 예측 단계를 나타낸다.
- 도 2는 복잡핵형(CK) 그룹의 비지도 계층적 클러스터 분석에 관한 것으로, 각 코호트에서 두 그룹의 별개의 유전자 발현 패턴을 보여주는 KUMC(A) 및 TCGA AML 코호트(B)의 히트맵이다.
- 도 3은 KUMC 및 TCGA 코호트(A), Beat AML 코호트(B) 및 GSE15061 코호트(C)에 대한 로지스틱 회귀 모델 및 곡선 아래 면적(AUC)을 사용하여 복잡핵형을 예측하기 위한 ROC(수신기 작동 특성) 곡선 그래프이다.
- 도 4A 내지 4C는 기존 핵형 분석 기반으로, KUMC(A), TCGA AML(B) 및 Beat AML 코호트(C)에서의 복잡핵형에 따른 전체 생존율의 Kaplan-Meier 플롯을 표시한 것이고, 도 4D 내지 4F는 복잡핵형 시그니처 예측 기반으로 KUMC(D), TCGA AML(E) 및 Beat AML 코호트(F)에서의 Kaplan-Meier 플롯을 표시한 것으로, 코호트 내 각 특징의 유무를 가진 환자는 각각 주황색과 녹색 선으로 표시하였다.

발명을 실시하기 위한 구체적인 내용

[0071] 이하, 본 발명을 하기 실시예를 통하여 보다 상세하게 설명한다. 그러나 이들 실시예는 본 발명을 예시적으로 설명하기 위한 것으로 본 발명의 범위가 이들 실시예만으로 한정되는 것은 아니다.

[0073] 실시예 1. 연구 설계 및 RNA 추출**[0074] (1) 샘플 선정 및 코호트 데이터 분석**

[0075] 연구 설계 개요는 도 1에 표시하였다. AML 또는 MDS-EB로 진단된 총 47명의 성인이 연구에 포함되었으며, t(8;21)(q22;q22.1), inv(16)(p13.1;q22) 또는 t(16;16)(p13.1;q22), t(15;17)(q22;q11-12)을 포함하는 WHO-지정 균형 이상(balanced abnormalities), KMT2A 및 inv(3)(q21.3;q26.2) 또는 t(3;3)(q21.3;q26.2)와 관련된 모든 전좌를 가진 AML 환자는 제외되었다. 복잡핵형(complex karyotype, CK)은 AML과 MDS-EB 모두에서 3개 이상의 관련 없는 염색체 이상으로 정의되었다. 연구는 고려대학교 구로병원 기관심사위원회(2021GR0247)의 승인을 받았으며 헬싱키 선언에 따라 수행되었다.

[0076] 복잡핵형 특이 유전자군(CK signature, CKS)을 설정하기 위해 고대구로병원(KUMC) 코호트와 TCGA(The Cancer Genome Atlas) AML 코호트로 구성된 탐색 코호트가 사용되었다. TCGA AML 코호트 클리닉 데이터를 사용하여, (CK 분류에 사용할 수 있는 데이터에 대해) 상기 WHO-지정 균형 이상(32건)을 제외하고, 범주에 따라 CK를 정의하여 “세포유전학적 이상 유형”을 검증하였다. 파생된 CKS를 추가로 평가하기 위하여, Beat AML 코호트의 발현 데이터를 분석하였다. CK 분류에 사용할 수 있는 RNA-seq 데이터 및 염색체 분석으로 진단된 골수(BM) 흡인(aspirate) 샘플(161건) 중 상기 배제 기준이 의심되지 않는 샘플을 분석하였다(120건). TCGA AML 및 Beat AML 코호트 모두에 대해, 염색체 분석 결과 및 mRNA 발현 데이터를 포함한 임상 정보는 각각 National Cancer Institute Genomic Data Commons(GDC) 및 cBioportal에서 검색되었다.

[0077] MDS-EB에 적용 가능한 CKS는 MDS 코호트(164건)의 MDS-EB 환자(48건)의 진단 골수(BM) 샘플의 유전자 발현 프로필을 사용하여 결정되었다(Gene Expression Omnibus(GEO) accessions GSE15061).

[0078] (2) 연구 샘플에 대한 임상적 특징 및 복잡핵형 분석

[0079] 연구에 포함된 총 47명의 AML 또는 MDS-EB(AML/MDS-EB) 환자에 대한 주요 임상적 및 혈액학적 특징은 표 1(KUMC 코호트의 환자 특성)에 나타내었다. 3개 이상의 관련 없는 염색체 이상으로 정의되는, CK는 AML/MDS 환자 16명(AML 10명 및 MDS-EB 6명)에서 검출되었으며, 염색체 이상의 평균값은 15개 였다. 핵형 복잡성 정도는 CK가 있는 AML과 CK가 있는 MDS-EB 간에 유의한 차이가 없었다($P = 0.3855$).

표 1

[0080]

	KUMC (n=47) 평균 ± SD
성별 (%)	
남성	29 (61.7)
여성	18 (38.3)
나이 (세)	65.5 ± 12.8
Hb (g/dL)	8.1 ± 1.2
WBC 수 ($\times 10^9/L$)	26.1 ± 53.0
혈소판 수 ($\times 10^9/L$)	67.9 ± 50.1
BM 아세포 (%)	42.1 ± 27.6
진단	
재발성 유전적 이상이 있는 AML	17 (36.2)
돌연변이 NPM1을 가진 AML	8 (17.0)
CEBPA의 이대립유전자 돌연변이가 있는 AML	3 (6.4)
돌연변이 RUNX1을 가진 AML	6 (12.8)
AML-MRC	11 (23.4)
AML, NOS	9 (19.1)
MDS-EB	10 (21.3)

염색체 이상 수	5.2 ± 8.2
핵형	
복잡핵형	16 (34.0)
비-복잡핵형	31 (66.0)
NA	-
전체 생존율 (개월)	10.8 ± 9.3

(3) RNA 추출, 라이브러리 제작 및 RNA-seq

BM 샘플은 RNA-seq을 수행하여 진단하였다. RNA는 QIAamp RNA Blood Mini Kit(Qiagen, Venlo, Netherlands)를 사용하여 추출하였고, SuperScript VILO cDNA 합성 키트(Thermo Fisher Scientific)를 사용하여 상보적 DNA(cDNA)로 역전사시켰다. RNA-seq 라이브러리는 TruSeq Standard Total RNA LT Sample Prep Kit(Illumina, Inc., San Diego, CA, USA)를 사용하여 제작하였고 NovaSeq Sequencing 시스템(Illumina, Inc.)에서 시퀀싱되었다. 시퀀싱된 리드는 HISAT2(버전 2.1.0)를 사용하여 참조 유전체 GrCh37(hg19)에 정렬되었고, 정렬된 리드는 StringTie(버전.2.1.3b)를 사용하여 평가하였다.

실시예 2. 복잡핵형 예측을 위한 유전자 발현 분석

(1) 유전자 발현 분석

KUMC 및 TCGA AML 코호트의 RNA-seq 데이터를 이용하여 DESeq2 패키지(버전 1.32.0)를 사용하여 각 코호트 내에서 CK가 있는 환자와 CK가 없는 환자간의 차등 유전자 발현 분석을 수행하였다. 조정된 p-값 < 0.05 및 절대 log2 배수 변화(fold change) > 0.5의 임계값을 사용하여 두 코호트에서 공통적으로 차등적으로 발현된 유전자를 선택하고(도 2), ConsensusPathDB를 사용하여 과다-표현 분석을 수행하였다. 분석을 위해서는 KEGG(Kyoto Encyclopedia of Genes and Genomes), Reactome 및 Wikipathways를 포함한 신호전달경로 데이터 세트가 선택되었다.

시그니처 개발을 위해, 다운스트림 분석에는 FPKM(매핑된 백만 조각당 엑손 킬로베이스당 조각 단위)이 사용되었다. 평균 FPKM 값이 0인 유전자를 제외하고 FPKM 값은 $\log_2(\text{FPKM}+1)$ 값으로 변환하였다. 분위수(quantile) 정규화는 preprocessCore 패키지(버전 1.58.0)를 사용하여 수행하였다. 시그니처 개발을 위한 유전자의 하위 집합을 선택하기 위하여, glmnet 패키지(버전 4.1-4)를 사용한 LASSO 회귀가 적용되었으며, 검증에는 로지스틱 회귀 모델에 맞게 캐릿 패키지(버전 6.0-92)의 LOOCV(leave-one-out cross validation)를 사용하였다. 다음과 같이 KUMC 및 TCGA AML 코호트에서 CK를 예측하기 위해 계산된 가중(weighted) 조합된 유전자 발현(즉, CK 점수) 중, 8개 유전자의 하위 집합(subset)이 선택되었다:

복잡핵형 특이 유전자군(CKS) 점수 = $17.8670 + (\text{PBX3} \times -1.0679) + (\text{RPS14} \times -2.8674) + (\text{CCL23} \times -1.3630) + (\text{SOD1} \times 0.9259) + (\text{TSPAN5} \times 1.0268) + (\text{SLC18A2} \times 0.4810) + (\text{FAM13B} \times 1.2650) + (\text{LINC01138} \times -3.0967)$.

glmnet 패키지의 '예측' 기능을 사용하여 환자를 임계값 0.5로 'CKS' 또는 '비-CKS'로 분류하였다.

(2) 복잡핵형 및 비-복잡핵형 그룹간 차등 유전자 발현 프로파일 분석

탐색적 코호트에서, CK 정의된 KUMC 코호트의 47명 환자 및 TCGA AML 코호트의 98명의 환자는 CK 및 비-CK 그룹간에 차등 유전자 발현 프로파일링을 수행하였다. 두 코호트에서 공통으로 총 80개의 차등 발현 유전자가 선택되었다. CK 그룹의 비지도(unsupervised) 계층적 클러스터 분석은 두 코호트에서 구별되는 유전자 발현 패턴을 보여주었다(도 2).

CK 예측을 위한 CKS 유전자 세트를 설정하기 위하여, 다변량 로지스틱 회귀 모델에 맞는 LASSO 회귀를 사용하여 상기 80개의 차등 발현 유전자 중에서 유전자의 하위 집합을 선택하였다. 이 모델에는 이전에 AML(PBX3, CCL23) 또는 MDS(RPS14, FAM13B)와 연관된 것으로 설명된 유전자와 신규 유전자가 포함되었다. LOOCV는 탐색적 코호트에서 CK 환자 예측에 대해 전체 정확도 95.17%로 0.9617의 AUC값을 추정하였다.

(3) 통계

모든 통계 분석 및 시각화는 R 소프트웨어(버전 3.4.3)를 사용하여 수행되었다. LOOCV를 이용한 모델 검증을 위

하여, ROC(수신기 작동 특성) 곡선과 AUC(곡선 아래 면적)를 분석하였다. 환자 특성 분석을 위해, 두 그룹 간의 비교는 연속 변수에 대한 Kruskal-Wallis 순위-합 검정과 범주형 변수에 대한 Fisher의 정확 검정을 사용하여 수행되었다. 생존은 survminer(버전 3.4.3) 패키지를 사용하여 Kaplan-Meier 방법, 로그 순위 검정 및 Cox 비례 위험 회귀에 의해 추정되었다.

실시예 3. 복잡핵형 예측을 위한 복잡핵형 시그니처 검증

(1) Beat AML 및 GSE15061 코호트를 이용한 복잡핵형 시그니처(CKS) 유전자 세트 검증

CKS 유전자 세트는 사용 가능한 RNA-seq 데이터가 있는 Beat AML 코호트와 마이크로어레이 데이터가 있는 GSE15061 코호트를 사용하여 검증하였다. Beat AML 코호트의 경우, CPM(매핑된 백만 리드 당 카운트 단위) 값을 사용하고 KUMC 및 TCGA 코호트와 동일한 정규화 방법을 사용하여 유전자 발현을 정량화하였다. 다른 단위가 적용됨에 따라, CKS는 다음과 같이 Beat AML 코호트에서 추정된 회귀 계수에 의해 가중치가 부여된 위의 8개 유전자를 사용하여 계산하였다:

복잡핵형 특이 유전자군(CKS) 점수 = $48.3971 + (PBX3 \times -0.2931) + (RPS14 \times -5.2680) + (CCL23 \times -2.5812) + (SOD1 \times 4.5024) + (TSPAN5 \times 1.4217) + (SLC18A2 \times -0.5298) + (FAM13B \times -4.9628) + (LINC01138 \times -6.9592)$.

GSE15061의 경우, CKS에서 생성된 미가공 Affymetrix CEL 파일은 상기 설명한 회귀 방법을 MDS-EB 케이스(n=48)에 적용하고, 모든 CKS 유전자에 대한 분석을 제한함으로써 식별되었다. FAM13B, LINC01138 및 SLC18A2 유전자를 사용하는 GSE15061 코호트에서 다음과 같은 새로운 방정식이 생성되었다:

복잡핵형 특이 유전자군(CKS) 점수 = $45.23 + (SLC18A2 \times 23.48) + (FAM13B \times -81.50) + (LINC01138 \times -34.03)$.

두 코호트 모두에서, 임계값 0.5를 사용하여 환자는 'CKS' 또는 '비-CKS'로 분류되었다.

(2) 복잡핵형 시그니처 유전자 세트를 이용한 복잡핵형 예측

복잡핵형 시그니처(CKS) 유전자 세트의 유용성을 평가하기 위해, Beat AML 코호트 및 GSE15061 코호트에 대해 CKS 유전자 세트를 적용하고 다변량 로지스틱 회귀 모델을 구축하였다. GSE15061 코호트에는 MDS-EB 환자 중 적은 CK 그룹만 포함되었으며(n=8/48, 16.67%), CKS 유전자 세트를 다변량 로지스틱 회귀 모델에 적용하는 것은 통계적으로 제한되었다. 따라서 GSE15061 코호트의 CKS 점수에 기여하는, 8개의 CKS 유전자 세트 중에 3개 유전자(FAM13B, LINC01138, SLC18A2)를 재선택하였다. LOOCV는 Beat AML 및 GSE15061 코호트의 AUC 값을 각각 0.9522 및 0.834로 추정하였다(도 3B 및 3C). CK 시그니처 점수는 Beat AML 코호트에서 93.33%(n=112/120)의 정확도로, GSE15061 코호트에서 93.75%(n=45/48)의 정확도로 CK를 예측하여, 다양한 유전자 발현 프로파일링 기술(즉, RNA-seq 및 마이크로어레이)에서 가능한 견고성(robustness)을 보여주었다.

실시예 4. 복잡핵형 시그니처와 예후의 연관성 확인

(1) 복잡핵형 및 복잡핵형 시그니처의 임상적 영향 확인

임상 결과 예측시 CKS 점수를 사용하여 결정된 'CKS'의 예후 영향을 조사하였다. KUMC, TCGA AML, Beat AML 코호트와 같은 생존 데이터가 있는 3개의 코호트를 분석했다. KUMC 및 TCGA 코호트에서, CKS 모델 개발에 포함된 모든 환자는 생존 분석에도 포함되었고, 이용가능한 생존 데이터가 있는 107명의 환자(CK 그룹 n=89; 비-CK 그룹 n=18)는 Beat AML 코호트에 포함되었다.

KUMC 코호트에서, 임상 매개변수와 관련하여 CK 환자는 백혈구(WBC) 수가 더 높았고(P=0.010; 36.7 vs. $5.4 \times 10^9/L$), BM 아세포 비율은 더 낮았다(P=0.006; 22.3 vs. 45.6%)(표 2). CK 환자의 추정 1년 전체 생존율(OS)은 31.3%로, CK는 더 짧은 OS와 유의하게 관련이 있었으며, CK가 없는 환자의 추정 1년 OS는 45.2%였다(p=0.028). CKS의 생존 분석도 더 짧은 OS와 유의한 연관성을 나타내었다(p = 0.006, 1년 OS 21.4% vs. 48.5%)(도 4A 및 4D). 단변량 Cox 분석은 CK 및 CKS 모두에 대해 전체 생존율(OS)과 강한 연관성을 확인했다(CK: HR=2.1812, P=0.0318; CKS: HR=2.0745, P=0.0491).

그 다음, TCGA AML 및 Beat AML 코호트에서 CK 및 CKS의 임상 영향을 조사하였다. 두 코호트 모두에서, KUMC 코

호트와 마찬가지로 CK 그룹에서 높은 WBC 및 낮은 BM 아세포 백분율도 관찰되었다.

[0111] CK는 TCGA AML 코호트에서 더 짧은 전체 생존율(OS)과 유의하게 연관되었고($P=0.031$; 1년 OS, 48.8% vs. 18.2%), Beat AML 코호트에서 유사한 경향성을 나타내었다($P=0.062$; 1년 OS, 39.3% vs. 16.7%). 반면, CKS는 두 코호트에서 더 짧은 OS와 유의하게 연관되었다(TCGA AML: $P=0.012$; 1년 OS, 48.8% vs. 18.2%; Beat AML: $P=0.008$; 1년 OS, 39.6% vs. 12.5%) (도 4). 두 코호트에서, 단변량 Cox 분석에 기초하여 CK 및 OS와의 강한 연관성이 두 코호트에서 모두 나타났기 때문에(TCGA AML: $HR=2.2569$, $P=0.0357$; Beat AML: $HR=1.8371$, $P=0.0657$), 두 코호트 모두에서 CKS와 더 강한 연관성이 관찰되었다(TCGA AML: $HR=3.2004$, $P=0.0021$; Beat AML: $HR=2.3568$, $P=0.0102$).

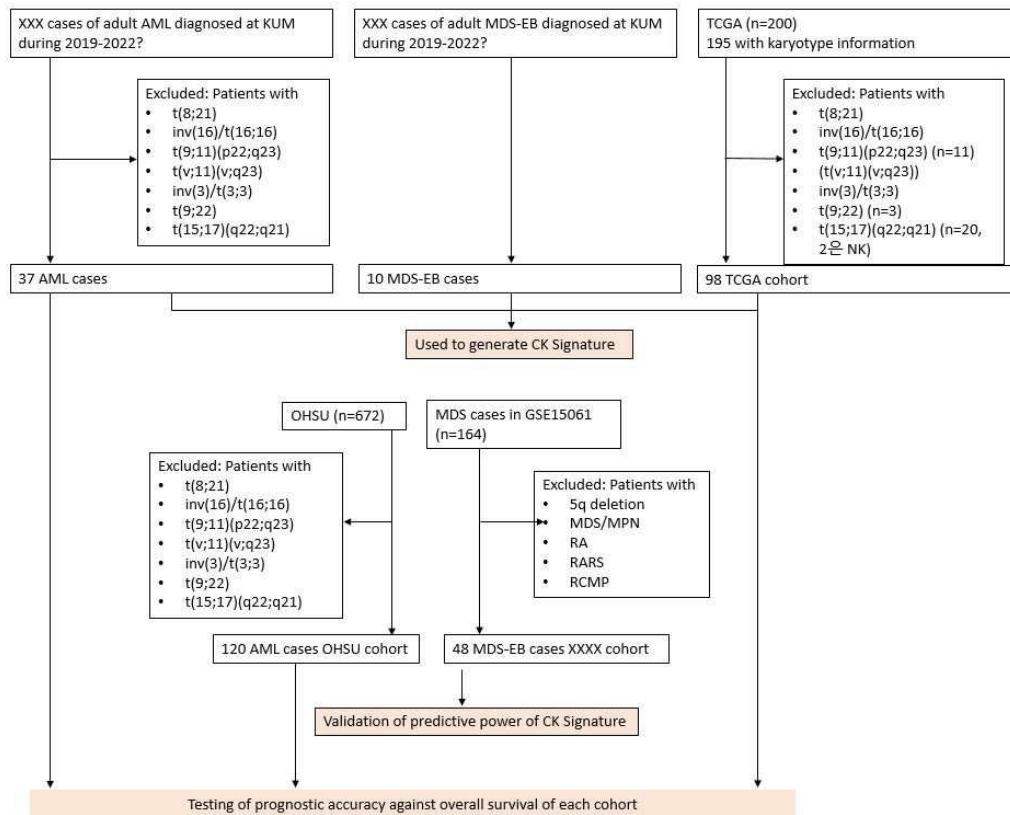
[0112] (2) 다른 모델과의 예후 예측 성능 비교

[0113] 이전 연구에서, 백혈병 줄기 세포 유전자와 관련된 17개 유전자의 발현을 기반으로 한 모델인 LSC17(leukemia stem cell 17) 점수는 불량한 예후를 가진 AML 환자의 정확한 예측을 입증하였다. 또한, AML 예후 점수(APS)는 이전에 AML 환자의 발현 데이터를 사용하여 개발된 것으로, 환자 결과와 밀접하게 관련된 모델을 개발하였다. 상기 2 종류의 이전에 개발된 모델의 목표는 WHO-지정 균형 이상을 가진 AML 환자를 포함하여 모든 AML 환자의 예후를 예측하는 것이었다. 예측 모델의 목적이 CK를 예측하는 것이기는 하나, 본 발명의 모델에서 관찰된 예후 영향을 기반으로, 예측 목적에서 CKS를 LSC17 및 APS와 비교하였다. TCGA AML 및 Beat AML 코호트 내 각 AML 샘플에 대해 LSC17 점수와 APS를 계산하였고, 각 코호트 내에서 중앙값보다 높은 점수를 가진 샘플은 TCGA AML 코호트(LSC17: $HR = 2.0164$, $P=0.0069$, APS: $HR = 1.7173$, $P=0.0360$) 및 Beat AML 코호트(LSC17: $HR = 2.1762$, $P=0.0076$, APS: $HR = 2.2517$, $P=0.0052$)에서 유의하게 나쁜 생존 결과를 가진다는 것을 관찰하였다. 그러나 각 코호트에서 CKS에 대한 위험 비율은 단변량 Cox 분석에서 LSC17 및 APS보다 높았다. 또한 다변량 Cox 분석에서도 다변량 방정식에 공변량으로 포함될 때 상기 두 가지 예후 모델에 비해 높은 위험 비율을 보였다(표 3).

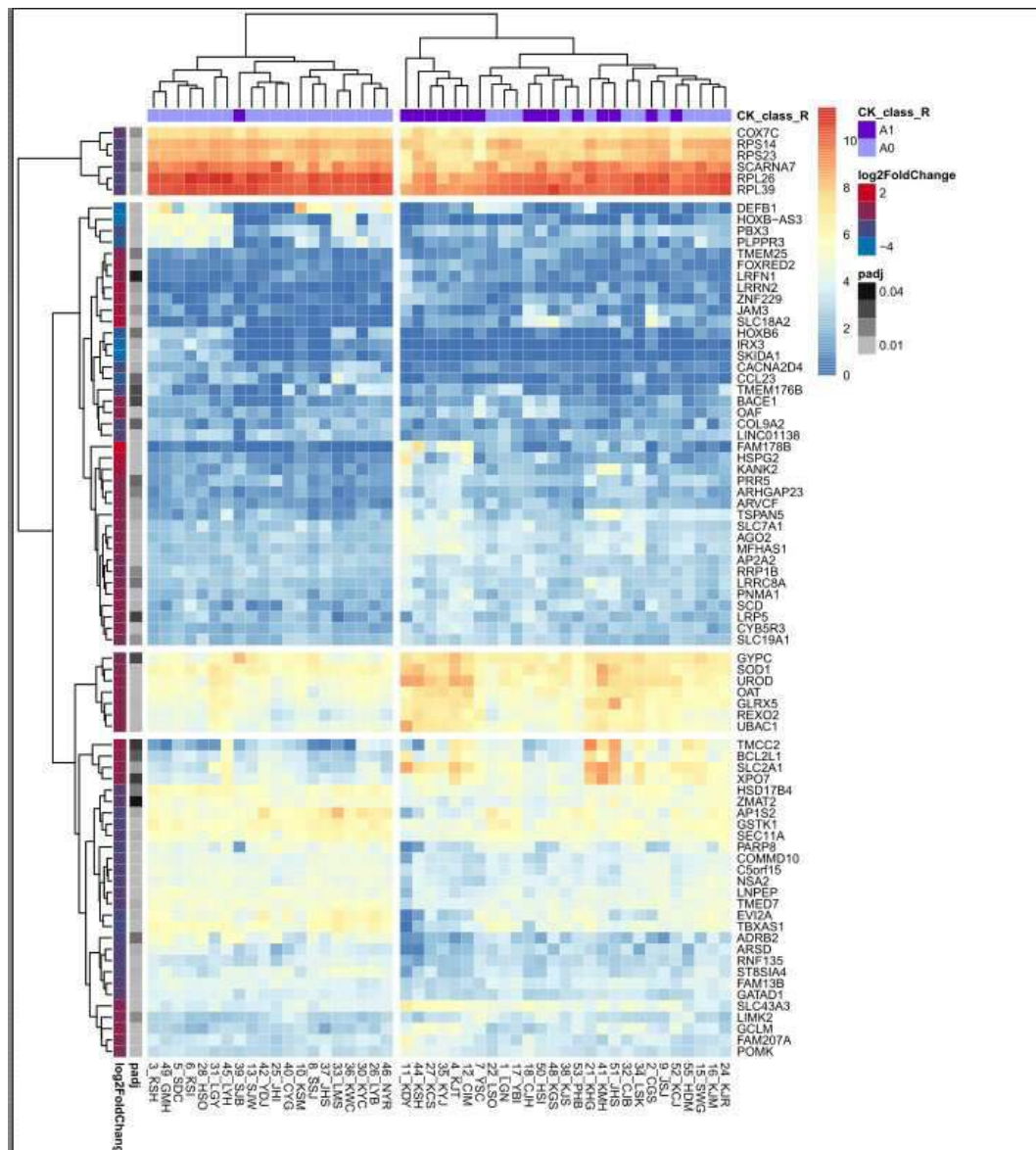
[0115] 이상의 설명으로부터, 본 발명이 속하는 기술분야의 당업자는 본 발명이 그 기술적 사상이나 필수적 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 이와 관련하여, 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적인 것이 아닌 것으로 이해해야만 한다. 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허 청구범위의 의미 및 범위 그리고 그 등가 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

도면

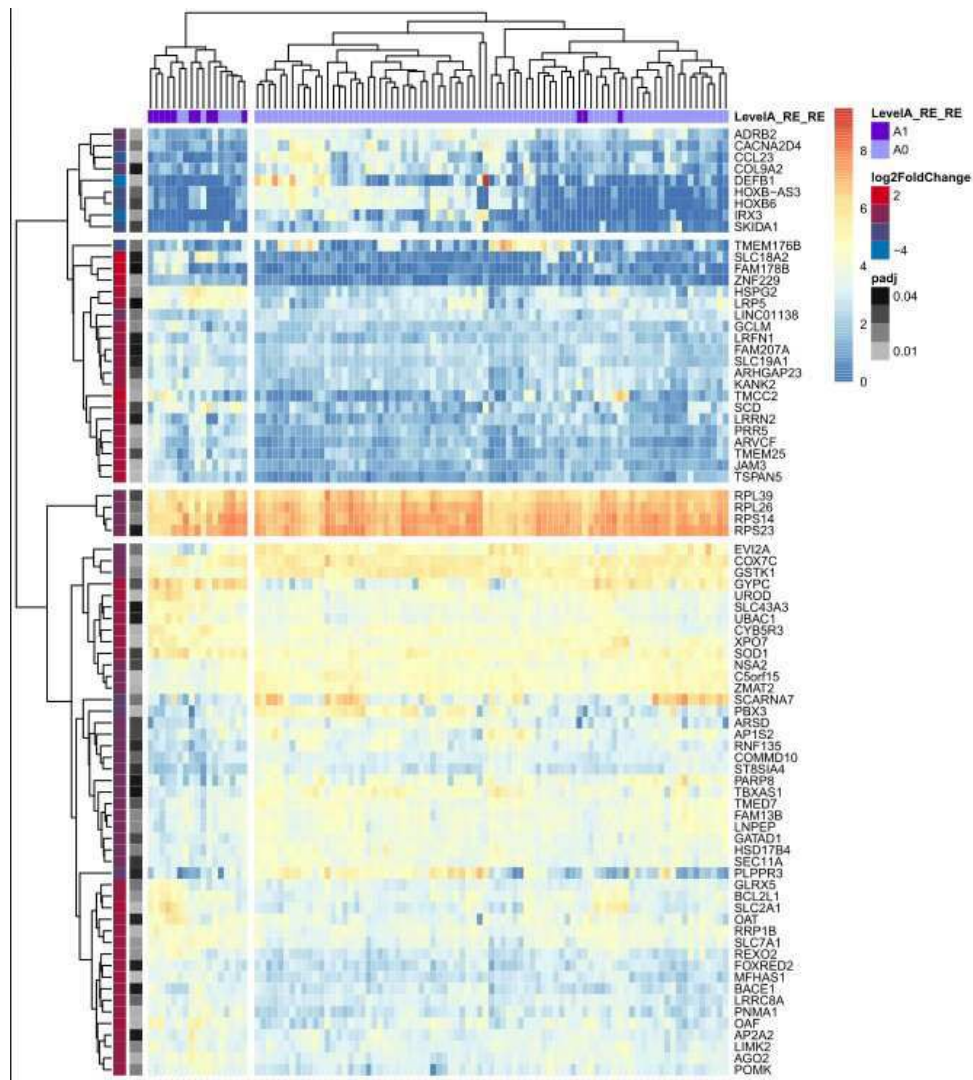
도면1



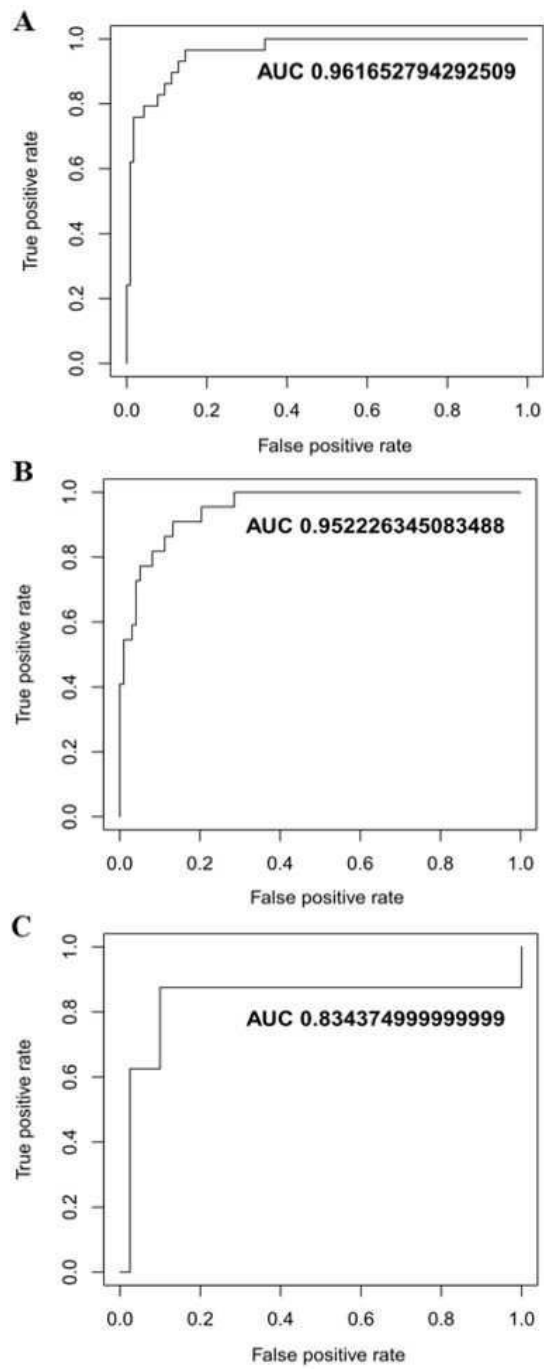
도면2a



도면2b



도면3



도면4

