



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0074555
(43) 공개일자 2020년06월25일

(51) 국제특허분류(Int. Cl.)
C12Q 1/6886 (2018.01) G01N 33/574 (2006.01)
G16H 50/20 (2018.01) G16H 50/30 (2018.01)
(52) CPC특허분류
C12Q 1/6886 (2018.05)
G01N 33/57407 (2013.01)
(21) 출원번호 10-2018-0163080
(22) 출원일자 2018년12월17일
심사청구일자 2018년12월17일

(71) 출원인
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
함원식
서울특별시 강남구 일원로 127, 110동 302호(일원동, 가람아파트)
박지수
서울특별시 송파구 양재대로 1218, 257동 901호(방이동, 올림픽선수기자촌아파트)
(74) 대리인
이재영

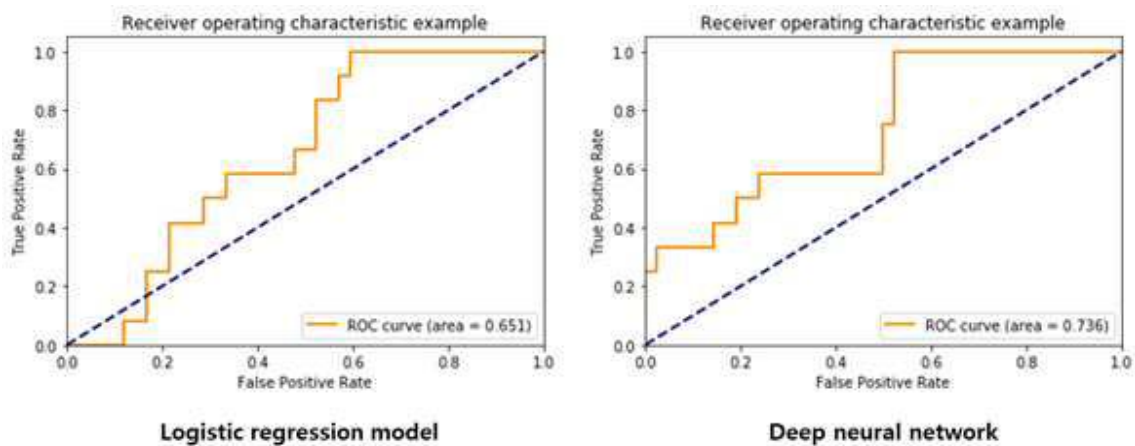
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 암의 진단을 위한 정보를 제공하는 방법

(57) 요약

본 발명은 암의 진단을 위한 정보를 제공하는 방법 및 장치에 관한 것으로, 본 발명의 상기 방법 및 장치를 이용하는 경우 암과 관련성이 있는 유전자 또는 그에 의해 암호화되는 단백질을 독립변수로 하여 분석한 경우에 비하여, 그 정확도 및 AUC(Area under the curve) 값을 현저하게 증가시킬 수 있다. 나아가, 이를 통해 진단 당시에 암의 악성 여부를 예측함으로써, 환자 개인에 특화되어 있는 치료방법을 효과적으로 제공할 수 있다.

대표도 - 도5



(52) CPC특허분류

G01N 33/57438 (2019.01)
G16H 50/20 (2018.01)
G16H 50/30 (2018.01)
C12Q 2600/136 (2013.01)
C12Q 2600/158 (2013.01)
G01N 2333/4703 (2013.01)
G01N 2333/9108 (2013.01)
G01N 2500/00 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	HI17C1095
부처명	보건복지부
연구관리전문기관	한국보건산업진흥원
연구사업명	질환극복기술개발사업
연구과제명	T1 병기 신세포암에서의 개별 맞춤 치료를 위한 예후 바이오마커로서의 FOXC2와 CLIP4의
유효성 평가 및 고위험군 예측 모형의 개발	
기 여 율	1/1
주관기관	연세대학교 산학협력단
연구기간	2018.01.01 ~ 2018.12.31

명세서

청구범위

청구항 1

목적하는 개체로부터 분리된 생물학적 시료에서, *FOXC2*(Forkhead box protein C2), *PBRM1*(Protein polybromo-1) 및 *BAP1*(BRCA1 associated protein 1)으로 구성된 군으로부터 선택되는 2종 이상의 유전자 또는 이에 의해 암호화되는 단백질이 존재하는 수준을 측정하는 단계; 및

상기 유전자 또는 상기 단백질이 존재하는 수준의 측정된 값을 미리 설정된 방법에 적용하여 위험 점수를 산출하는 단계;를 포함하는 암의 진단을 위한 정보를 제공하는 방법.

청구항 2

제 1항에 있어서,

상기 암은 악성 비뇨기암인 것인, 암의 진단을 위한 정보를 제공하는 방법.

청구항 3

제 2항에 있어서,

상기 비뇨기암은 방광암, 전립선암, 요관암 및 신장암으로 구성된 군으로부터 선택되는 적어도 하나인 것인, 암의 진단을 위한 정보를 제공하는 방법.

청구항 4

제 3항에 있어서,

상기 신장암은 투명세포형 신세포암(Clear cell renal cell carcinoma; ccRCC)인 것인, 암의 진단을 위한 정보를 제공하는 방법.

청구항 5

제 1항에 있어서,

상기 미리 설정된 방법은 하기 식 1 내지 4로 구성된 군으로부터 선택되는 적어도 하나의 식을 통해 위험 점수를 산출하는 것인, 암의 진단을 위한 정보를 제공하는 방법:

[식 1]

$$\text{위험 점수} = 1/[1 + \exp(-(A1 - A2 \times (PBRM1_norm) - A3 \times (BAP1_norm)))]$$

[식 2]

$$\text{위험 점수} = 1/[1 + \exp(-(B1 - B2 \times (PBRM1_norm) - B3 \times (FOX2_norm)))]$$

[식 3]

$$\text{위험 점수} = 1/[1 + \exp(-(C1 - C2 \times (BAP1_norm) - C3 \times (FOX2_norm)))]$$

[식 4]

$$\text{위험 점수} = 1 / [1 + \exp(-(D1 - D2 \times (FOX2_norm) - D3 \times (PBRM1_norm) - D4 \times (BAP1_norm)))]$$

상기 식 1 내지 4에서, *PBRM1_norm*은 (*PBRM1*의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준)/0.06, *BAP1_norm*은 (*BAP1*의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준)/0.005, *FOX2_norm*은 (*FOX2*의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준)/0.096이고,

상기 식 1에서 *A1*은 -1.5 내지 -0.9, *A2*는 0.7 내지 1.1, *A3*는 0.3 내지 0.7이고, 상기 식 2에서 *B1*은 -1.5 내지 -0.9, *B2*는 0.6 내지 1.0, *B3*는 0.1 내지 0.5이며, 상기 식 3에서 *C1*은 -1.5 내지 -0.9, *C2*는 0.3 내지

0.7, C3는 0.3 내지 0.7이고, 상기 식 4에서 D1은 -1.5 내지 -0.9, D2는 0.1 내지 0.5, D3는 0.6 내지 1.0, D4는 0.2 내지 0.8이다.

청구항 6

제 1항에 있어서,

상기 위험 점수가 0.2 내지 1인 경우 암으로 분류하는 단계를 더 포함하는 것인, 암의 진단을 위한 정보를 제공하는 방법.

청구항 7

제 1항에 있어서,

상기 미리 설정된 방법은 딥 러닝(Deep learning) 분석 방법인 것인, 암의 진단을 위한 정보를 제공하는 방법.

청구항 8

제 7항에 있어서,

상기 딥 러닝 분석 방법은 딥뉴럴네트워크(Deep neural networks; DNN), 딥컨볼루션뉴럴네트워크(Deep convolutional neural networks), 딥신뢰뉴럴네트워크(Deep belief networks; DBN) 및 순환뉴럴네트워크(Recurrent neural networks; RNN)로 구성된 군으로부터 선택되는 적어도 하나를 사용하여 학습된 것인, 암의 진단을 위한 정보를 제공하는 방법.

청구항 9

목적하는 개체로부터 분리된 생물학적 시료에서, *FOXO2*, *PBRM1* 및 *BAP1*으로 구성된 군으로부터 선택되는 2종 이상의 유전자 또는 이에 의해 암호화되는 단백질이 존재하는 수준이 측정된 값을 입력하는 입력부;

상기 입력부에 입력된 값을 미리 설정된 방법에 적용하여 위험 점수를 산출하는 산출부; 및

상기 산출부로부터 전송 받은 값을 표시하는 표시부를 포함하는, 암의 진단을 위한 정보 제공 장치.

청구항 10

제 9항에 있어서,

상기 암은 악성 비뇨기암인 것인, 암의 진단을 위한 정보 제공 장치.

청구항 11

제 10항에 있어서,

상기 비뇨기암은 방광암, 전립선암, 요관암 및 신장암으로 구성된 군으로부터 선택되는 적어도 하나인 것인, 암의 진단을 위한 정보 제공 장치.

청구항 12

제 11항에 있어서,

상기 신장암은 투명세포형 신세포암인 것인, 암의 진단을 위한 정보 제공 장치.

청구항 13

제 9항에 있어서,

상기 미리 설정된 방법은 하기 식 1 내지 4로 구성된 군으로부터 선택되는 적어도 하나의 식을 통해 위험 점수를 산출하는 것인, 암의 진단을 위한 정보 제공 장치:

[식 1]

$$\text{위험 점수} = 1/[1 + \exp(-(A1 - A2 \times (PBRM1_norm) - A3 \times (BAP1_norm)))]$$

[식 2]

$$\text{위험 점수} = 1/[1 + \exp(-(B1 - B2 \times (PBRM1_norm) - B3 \times (FOXC2_norm)))]$$

[식 3]

$$\text{위험 점수} = 1/[1 + \exp(-(C1 - C2 \times (BAP1_norm) - C3 \times (FOXC2_norm)))]$$

[식 4]

$$\text{위험 점수} = 1/[1 + \exp(-(D1 - D2 \times (FOXC2_norm) - D3 \times (PBRM1_norm) - D4 \times (BAP1_norm)))]$$

상기 식 1 내지 4에서, PBRM1_norm은 (PBRM1의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준)/0.06, BAP1_norm은 (BAP1의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준)/0.005, FOXC2_norm은 (FOXC2의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준)/0.096이고,

상기 식 1에서 A1은 -1.5 내지 -0.9, A2는 0.7 내지 1.1, A3는 0.3 내지 0.7이고, 상기 식 2에서 B1은 -1.5 내지 -0.9, B2는 0.6 내지 1.0, B3는 0.1 내지 0.5이며, 상기 식 3에서 C1은 -1.5 내지 -0.9, C2는 0.3 내지 0.7, C3는 0.3 내지 0.7이고, 상기 식 4에서 D1은 -1.5 내지 -0.9, D2는 0.1 내지 0.5, D3는 0.6 내지 1.0, D4는 0.2 내지 0.8이다.

청구항 14

제 9항에 있어서,

상기 표시부에, 상기 산출부로부터 전송된 위험 점수가 0.2 내지 1인 경우, 암으로 표시되는 것인, 암의 진단을 위한 정보 제공 장치.

청구항 15

암 개체로부터 분리된 생물학적 시료에, 후보물질을 처리하는 단계;

상기 후보물질이 처리된 생물학적 시료에서 FOXC2, PBRM1 및 BAP1으로 구성된 군으로부터 선택되는 2종 이상의 유전자 또는 이에 의해 암호화되는 단백질이 존재하는 수준을 측정하는 단계; 및

상기 유전자 또는 상기 단백질이 존재하는 수준의 측정된 값을 미리 설정된 방법에 적용하여 위험 점수를 산출하는 단계;를 포함하는, 암 치료제의 스크리닝 방법.

청구항 16

제 16항에 있어서,

상기 후보물질을 처리하기 이전의 위험 점수에 비하여, 후보물질을 처리한 이후의 위험 점수가 낮은 경우, 처리된 후보물질을 암 치료제로 선별하는 단계를 더 포함하는 것인, 암 치료제의 스크리닝 방법.

청구항 17

제 15항에 있어서,

상기 미리 설정된 방법은 하기 식 1 내지 4로 구성된 군으로부터 선택되는 적어도 하나의 식을 통해 위험 점수를 산출하는 것인, 암 치료제의 스크리닝 방법:

[식 1]

$$\text{위험 점수} = 1/[1 + \exp(-(A1 - A2 \times (PBRM1_norm) - A3 \times (BAP1_norm)))]$$

[식 2]

$$\text{위험 점수} = 1/[1 + \exp(-(B1 - B2 \times (PBRM1_norm) - B3 \times (FOXC2_norm)))]$$

[식 3]

$$\text{위험 점수} = 1/[1 + \exp(-(C1 - C2 \times (BAP1_norm) - C3 \times (FOXC2_norm)))]$$

[식 4]

위험 점수 = $1 / [1 + \exp(-(D1 - D2 \times (FOXC2_norm) - D3 \times (PBRM1_norm) - D4 \times (BAP1_norm)))]$

상기 식 1 내지 4에서, PBRM1_norm은 (PBRM1의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준)/0.06, BAP1_norm은 (BAP1의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준)/0.005, FOXC2_norm은 (FOXC2의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준)/0.096이고,

상기 식 1에서 A1은 -1.5 내지 -0.9, A2는 0.7 내지 1.1, A3는 0.3 내지 0.7이고, 상기 식 2에서 B1은 -1.5 내지 -0.9, B2는 0.6 내지 1.0, B3는 0.1 내지 0.5이며, 상기 식 3에서 C1은 -1.5 내지 -0.9, C2는 0.3 내지 0.7, C3는 0.3 내지 0.7이고, 상기 식 4에서 D1은 -1.5 내지 -0.9, D2는 0.1 내지 0.5, D3는 0.6 내지 1.0, D4는 0.2 내지 0.8이다.

청구항 18

제 15항에 있어서,

상기 암은 악성 비뇨기암인 것인, 암 치료제의 스크리닝 방법.

청구항 19

제 18항에 있어서,

상기 비뇨기암은 방광암, 전립선암, 요관암 및 신장암으로 구성된 군으로부터 선택되는 적어도 하나인 것인, 악성 비뇨기암 치료제의 스크리닝 방법.

청구항 20

제 19항에 있어서,

상기 신장암은 투명세포형 신세포암인 것인, 악성 비뇨기암 치료제의 스크리닝 방법.

발명의 설명

기술 분야

[0001] 본 발명은 암의 진단을 위한 정보를 제공하는 방법에 관한 것이다.

배경 기술

[0002] 종양(Tumor)은 비정상적인 세포의 과잉으로 인하여 발생하는 비제어적이고, 무질서한 세포 증식의 산물로서, 상기 종양이 파괴적인 증식성, 침윤 및 전이성을 가지게 되면 악성 종양(Malignant tumor), 즉 악성 암으로 분류하게 된다. 암의 종류는 현재까지 밝혀진 것만 해도 수십 종에 이르며, 주로 발병 조직의 위치에 따라 구분된다.

[0003] 20년 전까지만 하여도, 비뇨기암 환자는 모든 비뇨기 질환의 20 내지 30%에 불과하였으나, 최근 비뇨기암의 발생률이 해마다 증가되어 그 비율이 50%에 해당하게 되었다. 비뇨기암 환자 중에서, 신세포암(Renal cell carcinoma; RCC) 환자는 전 세계적으로 매년 270,000명에 이르고 있으며, 매년 약 120,000명이 이로 인해 사망하고 있다. 이와 같은 암은 측면 통증, 측면 종양 및 혈뇨를 포함하는 고전적인 증상을 나타내며, 대부분의 경우 CT와 초음파 이미징의 발전에 의해 증상 없이 우연히 진단되는 것을 특징으로 한다. 신세포암의 유형으로 투명세포형 신세포암(Clear cell renal cell carcinoma; 이하 'ccRCC'라 함), 유두 신세포암, 난염성 신세포암 등이 존재하며, 이와 같은 유형들은 모두 다른 예후와 치료에 대한 적응성을 나타낸다.

[0004] 위와 같은 신세포암의 유형 중에서 가장 일반적인 유형인 ccRCC는 신세포암 중에서 약 70%를 차지한다. 초기에는 ccRCC가 진단되는 경우, 임상적 임상 가이드라인에 의해 외과적 절제, 냉동 수술, 방사선 절제술 및 추적 검사와 같은 다양한 임상적 치료 방법을 선택할 수 있었다. 그러나, 국소화된 ccRCC의 경우, 절제술이 수행된 30%의 환자에서 재발이 일어나고, 15%의 환자에서 잠재적으로 사망에 까지 이르게 할 수 있는 전이성을 나타낼 수 있음이 밝혀졌다.

[0005] 이에 따라 악성 ccRCC 여부를 가려내고, 환자 개인에 적절한 임상적 치료 방법의 전략 결정 및 보조 요법에 관한 상담을 용이하게 하기 위하여, 임상 결과의 정확한 예측 인자를 확인하는 것이 필수적이다. 다양한 종양학적 결과와 관련하여, 예측 인자들이 보고되어 있으며 이러한 예측 인자들을 이용한 Mayo 클리닉 스코어,

UCLA(University California Los Angeles) 통합 시스템, MSKCC(Memorial sloan kettering cancer center) 노모그램 등과 같은 ccRCC를 예측할 수 있는 모델의 연구가 진행되고 있다.

- [0006] 그러나, 이러한 모델에 사용되는 일부 요인은 시간에 의존적인 요인(종양 성장 및 방사선 이미지의 변화 등)이나, 판단자 개인에 의해 편차가 발생할 수 있는 요인(종양의 등급, 종양 괴사 정도 및 환자의 성과 상태 등)이기 때문에 빠른 측정 및 일관된 결과를 얻기에는 한계점이 존재하였다. 또한, 상기 모델의 경우 ccRCC에서 발생하는 종양생물학을 완전히 설명하지 못한다는 한계점이 존재하여 비뇨기암, 특히 악성 ccRCC를 선별해 내어 이에 특이적인 치료법을 선택할 수 있는 방법에 대한 연구는 현재까지 미비한 실정이다.

발명의 내용

해결하려는 과제

- [0007] 본 발명의 일 목적은 암 진단에 관한 정보를 제공하는 방법을 제공하는 것이다.
- [0008] 본 발명의 다른 목적은 암 진단에 관한 정보 제공 장치를 제공한다.
- [0009] 본 발명의 또 다른 목적은 암 치료제 스크리닝 방법을 제공한다.
- [0010] 그러나 본 발명이 이루고자 하는 기술적 과제는 이상에서 언급한 과제에 제한되지 않으며, 언급되지 않은 또 다른 과제들은 아래의 기재로부터 당 업계에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

- [0011] 본 발명자들은 전이, 재발 또는 암에 의한 사망률 높은 수준으로 나타나는 악성 투명세포형 신세포암(Clear cell renal cell carcinoma; ccRCC)에서, *FOXC2*(Forkhead box protein C2), *PBRM1*(Protein polybromo-1) 및 *BAP1*(BRCA1 associated protein 1) 유전자 및 이에 의해 암호화되는 단백질의 발현이, 악성이 아닌 투명세포형 신세포암에 비하여 감소되어 있는 것을 통해 악성 투명세포형 신세포암을 진단할 수 있음을 발견하여 본 발명을 완성하게 되었다.
- [0013] 본 발명의 일 구현 예에서는 암 진단에 관한 정보를 제공하는 방법을 제공한다.
- [0015] 본 발명의 상기 정보를 제공하는 방법은 목적하는 개체로부터 분리된 생물학적 시료에서, *FOXC2*, *PBRM1* 및 *BAP1*으로 구성된 군으로부터 선택되는 2종 이상의 유전자 또는 이에 의해 암호화되는 단백질이 존재하는 수준을 측정하는 단계; 및 상기 유전자 또는 상기 단백질이 존재하는 수준의 측정된 값을 미리 설정된 방법에 적용하여 위험 점수를 산출하는 단계;를 포함한다.
- [0016] 본 발명의 상기 *FOXC2* 유전자 또는 이에 의해 암호화되는 단백질은, *FOXC2* 유전자는 출생 전 배아의 기관과 조직의 형성에 중요한 역할을 하는 전사인자 단백질인 *FOXC2* 단백질을 암호화하는 것으로서, 유방암의 경우 상기 *FOXC2* 유전자에 의해 암호화되는 단백질의 발현이 증가되는 경우 유방암의 전이를 유도할 수 있다는 보고가 존재한다. 본 발명의 목적상 상기 *FOXC2* 유전자 또는 이에 의해 암호화되는 단백질은 비뇨기암에서 유방암에서와 달리, 세포 사이클이 잘못되는 경우를 체크하는 역할을 수행하는 것으로, 이와 같은 유전자 또는 이에 의해 암호화되는 단백질의 발현이 감소되었을 때 비뇨기암의 전이, 재발 및/또는 암으로 인한 사망률이 높아지는 것일 수 있다. 상기 *FOXC2* 유전자는 서열번호 1로 표시되는 염기 서열을 포함하는 것일 수 있고, *FOXC2* 단백질은 서열번호 2로 표시되는 아미노산 서열로 이루어진 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0017] 본 발명의 상기 *PBRM1* 유전자 또는 이에 의해 암호화되는 단백질은, ATP-의존성 크로마틴-리모델링 복합체의 서브유닛을 암호화하는 유전자로서, 이와 같은 유전자에 의해 암호화된 단백질은 핵 호르몬 수용체에 의한 리간드 의존성 전화 활성화에 필요한 복합체의 필수 성분에 해당한다. 상기 *PBRM1* 유전자는 서열번호 3으로 표시되는 염기 서열을 포함하는 것일 수 있고, *PBRM1* 단백질은 서열번호 4로 표시되는 아미노산 서열로 이루어진 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0018] 본 발명의 상기 *BAP1* 유전자 또는 이에 의해 암호화되는 단백질은, 유비퀴틴 카르복실-말단 가수분해 효소 BAP1(ubiquitin carboxyl-terminal hydrolase BAP1)을 암호화하는 유전자로서, 상기 효소는 탈유비퀴틴화 효소로서 작용하여 다양한 세포 과정에 관여하는 많은 단백질의 기능을 조절하는데 도움을 주는 역할을 한다. 상기

BAP1 유전자는 서열번호 5로 표시되는 염기 서열을 포함하는 것일 수 있고, *BAP1* 단백질은 서열번호 6으로 표시되는 아미노산 서열로 이루어진 것일 수 있으나, 이에 제한되는 것은 아니다.

[0019] 본 발명의 상기 정보를 제공하는 방법에서 상기 측정의 대상이 되는 유전자 또는 이에 의해 암호화되는 단백질은 *FOXC2*, *PBRM1* 및 *BAP1*으로 구성된 군으로부터 선택되는 2종 이상을 포함하는 것일 수 있고, 바람직하게는 *FOXC2*와, *PBRM1* 또는 *BAP1* 중 어느 하나 이상을 포함하는 것일 수 있고, 더욱 바람직하게는 상기 *FOXC2*, *PBRM1* 및 *BAP1*를 모두 포함하는 것일 수 있으나, 이에 제한되는 것은 아니다. 이와 같이 상기 *FOXC2*, *PBRM1* 및 *BAP1*를 모두 포함하는 경우에는 각각 개개의 유전자 또는 이에 의해 암호화되는 단백질을 측정하는 경우에 비하여 예측 정확도를 현저하게 높일 수 있다.

[0020] 본 발명의 상기 생물학적 시료는 개체로부터 얻어지거나 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), 혈장(plasma) 및 혈청(serum)을 포함하는) 혈액, 객담(sputum), 눈물(tears), 점액(mucus), 세비액(nasal washes), 비강 흡인물(nasal aspirate), 호흡(breath), 소변(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.

[0021] 본 발명의 상기 유전자가 존재하는 수준의 측정은 프라이머를 이용하여 상기 유전자를 증폭시키거나, 형광 등으로 표시된 프로브와 상보적으로 결합하는 등의 통상의 방법을 통해 확인될 수 있는 것이고, 예를 들면, 중합효소연쇄반응(polymerase chain reaction), 형광상관분석법(fluorescence correlation spectroscopy), 마이크로 어레이(microarray) 및 칩어세이(chip-assay)으로 이루어진 군으로부터 선택되는 적어도 하나일 수 있으나, 이에 제한되는 것은 아니다. 상기 프라이머 또는 프로브는 상기 유전자의 염기 서열을 참조하여 당해 기술 분야에서 통상의 지식을 가진 자가 공지된 방법에 따라 쉽게 제작할 수 있다.

[0022] 본 발명의 상기 단백질이 존재하는 수준의 측정은 상기 단백질에 특이적인 항체 또는 앵타머를 이용하여 통상의 방법을 통해 확인될 수 있는 것이고, 예를 들면 웨스턴 블롯 분석(Western blot assay), ELISA(Enzyme linked immunosorbent assay), 방사선면역분석(RIA: Radioimmunoassay), 방사 면역 확산법(Radioimmunodiffusion), 오우크테로니(Ouchterlony) 면역 확산법, 로케트 면역전기영동(Rocket immunoelectrophoresis), 면역조직화학염색법(Immunohistochemical staining), 면역침전 분석법(Immunoprecipitation Assay), 보체 고정 분석법(Complement Fixation Assay), 면역형광법(Immunofluorescence), 면역크로마토그래피법(Immunochromatography), FACS(Fluorescenceactivated cell sorter analysis) 및 단백질 칩 분석법(protein chip technology assay)으로 이루어진 군으로부터 선택되는 적어도 하나일 수 있으나, 이에 제한되는 것은 아니다. 상기 항체 또는 프라이머는 상기 단백질의 아미노산 서열을 참조하여 당해 기술 분야에서 통상의 지식을 가진 자가 공지된 방법에 따라 쉽게 제작할 수 있다.

[0023] 본 발명의 상기 암은 바람직하게는 악성 비뇨기암(Urological cancer)일 수 있으나, 이에 제한되는 것은 아니다.

[0024] 본 발명의 상기 악성 암이란, 동시 전이(Synchronous metastasis), 재발(Recurrence) 및 암으로 인한 사망률(Cancer-specific death) 증가 중 적어도 하나가 발생된 암을 의미하며, 상기 동시 전이는 일차성(Primary) 암 진단 시 또는 진단으로부터 3개월 이내에 전이된 것을 의미한다.

[0025] 본 발명의 상기 비뇨기암은 비뇨기의 암으로서, 방광암, 전립선암, 요관암 및 신장암으로 구성된 군으로부터 선택되는 적어도 하나인 것일 수 있고, 바람직하게는 신장암일 수 있으며, 더욱 바람직하게는 투명세포형 신세포암(Clear cell renal cell carcinoma; ccRCC)일 수 있으나, 이에 제한되는 것은 아니다.

[0026] 본 발명의 상기 신장암은 조직학적 분류에 따른 세포형에 따라 투명세포형 신세포암과 비 투명신세포암(Non-clear renal cell carcinoma)로 분류된다. 상기 투명세포형 신세포암은 전체 신장암의 70-80%를 차지하며, 암 세포 주위의 혈관이 풍부하고 다른 장기로의 전이를 잘 일으키는 특징을 갖는다. 반면, 비 투명신세포암의 경우, 암 세포 주위의 혈관이 풍부하지 않으나, 젊은 연령 층에서 다발성으로 발생된다는 특징을 갖는다. 본 발명의 상기 방법은 조직 검사를 받기 이전 또는 조직 검사와 함께 상기 암의 조직학적 분류에 따른 종류를 구별

해 낼 수 있고, 이를 통해 진단 당시 암 환자에게 최적의 치료 방법을 제공할 수 있다는 장점을 갖는다.

[0027] 본 발명의 상기 투명세포형 신세포암은 종양의 크기가 7cm 이하인 것일 수 있으나, 이에 제한되는 것은 아니다. 상기 투명세포형 신세포암의 종양 크기가 7cm 보다 작은 초기 단계에서는 비 수술적 치료 방법, 예를 들면 동결 절제(Cryoablation), 고주파 열치료(Radiofrequency ablation)등이 적절하게 사용될 수 있다. 따라서, 본 발명의 목적상 상기와 같은 비 수술적 치료 방법에 의해서 충분한 치료 효과를 볼 수 없는 15%의 암, 특히 악성 비뇨기암 환자를 선별해 낼 수 있다.

[0028] 본 발명의 상기 미리 설정된 방법은 하기 식 1 내지 4로 구성된 군으로부터 선택되는 적어도 하나의 식을 통해 위험 점수를 산출하는 것일 수 있다.

[0030] [식 1]

[0031] 위험 점수 = $1/[1 + \exp(-(A1 - A2 \times (PBRM1_norm) - A3 \times (BAP1_norm)))]$

[0033] [식 2]

[0034] 위험 점수 = $1/[1 + \exp(-(B1 - B2 \times (PBRM1_norm) - B3 \times (FOXC2_norm)))]$

[0036] [식 3]

[0037] 위험 점수 = $1/[1 + \exp(-(C1 - C2 \times (BAP1_norm) - C3 \times (FOXC2_norm)))]$

[0039] [식 4]

[0040] 위험 점수 = $1 / [1 + \exp(-(D1 - D2 \times (FOXC2_norm) - D3 \times (PBRM1_norm) - D4 \times (BAP1_norm)))]$

[0042] 본 발명의 상기 식 1 내지 4에서, $PBRM1_norm$ 은 ($PBRM1$ 의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준)/0.06, $BAP1_norm$ 은 ($BAP1$ 의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준)/0.005, $FOXC2_norm$ 은 ($FOXC2$ 의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준)/0.096일 수 있다.

[0043] 본 발명의 일 실시예에서, 상기 유전자의 발현 수준은 상기 $FOXC2$, $PBRM1$ 및 $BAP1$ 의 유전자를 상기 유전자들 각각에 특이적인 프라이머 쌍을 이용하여 qRT-PCR을 수행한 뒤, 항존 유전자, 예를 들면 GAPDH, α -tubulin 또는 β -actin 등의 발현량에 대한 상대적인 비율 값으로 나타낼 수 있다. 또한, 본 발명의 다른 실시예에서, 상기 단백질의 발현 수준은 $FOXC2$, $PBRM1$ 및 $BAP1$ 단백질 각각에 특이적인 항체를 이용하여 웨스턴 블롯 등을 수행한 뒤, 항존 유전자, 예를 들면 GAPDH, α -tubulin 또는 β -actin 등의 발현량에 대한 상대적인 비율 값으로 나타낼 수 있다.

[0044] 본 발명의 상기 식 1에서, 상기 A1은 -1.5 내지 -0.9일 수 있고, 바람직하게는 -1.3 내지 -1.0, 더욱 바람직하게는 -1.2 내지 -1.05일 수 있으나, 이에 제한되는 것은 아니다. 또한, 상기 A2는 0.7 내지 1.1일 수 있고, 바람직하게는 0.8 내지 1.0, 더욱 바람직하게는 0.9 내지 1.0일 수 있으나, 이에 제한되는 것은 아니다. 또한, 상기 A3는 0.3 내지 0.7일 수 있고, 바람직하게는 0.4 내지 0.6일 수 있으며, 더욱 바람직하게는 0.5 내지 0.6일 수 있으나, 이에 제한되는 것은 아니다. 본 발명의 일 실시예에서는 상기 [식 1]에서 A1이 -1.097, A2가 0.963 및 A3이 0.524인 경우 $PBRM1$ 및 $BAP1$ 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준을 이용하여 암, 특히 악성 투명세포형 신세포암을 효과적으로 진단할 수 있음을 확인하였다.

[0045] 본 발명의 상기 식 2에서, 상기 B1은 -1.5 내지 -0.9일 수 있고, 바람직하게는 -1.3 내지 -1.0, 더욱 바람직하게는 -1.2 내지 -1.05일 수 있으나, 이에 제한되는 것은 아니다. 또한, 상기 B2는 0.6 내지 1.0일 수 있고, 바람직하게는 0.7 내지 0.9일 수 있고, 더욱 바람직하게는 0.8 내지 0.9일 수 있으나, 이에 제한되는 것은 아니다. 또한, 상기 B3는 0.1 내지 0.5일 수 있고, 바람직하게는 0.2 내지 0.4일 수 있으며, 더욱 바람직하게는 0.3 내지 0.4일 수 있으나, 이에 제한되는 것은 아니다. 본 발명의 일 실시예에서는 상기 [식 2]에서 B1이 -1.100, B2가 0.894, B3이 0.340인 경우, $PBRM1$ 및 $FOXC2$ 유전자 또는 이에 의해 암호화되는 단백질의 발현 수

준을 이용하여 암, 특히 악성 투명세포형 신세포암을 효과적으로 진단할 수 있음을 확인하였다.

- [0046] 본 발명의 상기 식 3에서, 상기 C1은 -1.5 내지 -0.9일 수 있고, 바람직하게는 -1.3 내지 -1.0, 더욱 바람직하게는 -1.2 내지 -1.05일 수 있으나, 이에 제한되는 것은 아니다. 또한, 상기 C2는 0.3 내지 0.7일 수 있고, 바람직하게는 0.4 내지 0.6일 수 있으며, 더욱 바람직하게는 0.5 내지 0.6일 수 있으나, 이에 제한되는 것은 아니다. 또한, 상기 C3는 0.3 내지 0.7일 수 있고, 바람직하게는 0.4 내지 0.6일 수 있으며, 더욱 바람직하게는 0.5 내지 0.6일 수 있으나, 이에 제한되는 것은 아니다. 본 발명의 일 실시예에서는 상기 [식 3]에서 C1이 -1.11, C2가 0.533, C3이 0.562인 경우 *BAP1* 및 *FOX2* 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준을 이용하여 암, 특히 악성 투명세포형 신세포암을 효과적으로 진단할 수 있음을 확인하였다.
- [0047] 본 발명의 상기 식 4에서, 상기 D1은 -1.5 내지 -0.9일 수 있고, 바람직하게는 -1.0 내지 -1.3, 더욱 바람직하게는 -1.05 내지 -1.2일 수 있으나, 이에 제한되는 것은 아니다. 또한, 상기 D2는 0.1 내지 0.5일 수 있고, 바람직하게는 0.2 내지 0.4일 수 있으며, 더욱 바람직하게는 0.3 내지 0.4일 수 있으나, 이에 제한되는 것은 아니다. 또한, 상기 D3는 0.6 내지 1.0일 수 있고, 바람직하게는 0.7 내지 0.9일 수 있고, 더욱 바람직하게는 0.8 내지 0.9일 수 있으나, 이에 제한되는 것은 아니다. 또한, 상기 D4는 0.2 내지 0.8일 수 있고, 바람직하게는 0.3 내지 0.7일 수 있으며, 더욱 바람직하게는 0.4 내지 0.6일 수 있으나, 이에 제한되는 것은 아니다. 본 발명의 또 다른 실시예에서는 상기 [식 4]에서 D1이 -1.082, D2가 0.336, D3이 0.887, D4가 0.521인 경우 *FOX2*, *PBRM1* 및 *BAP1* 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준을 이용하여 암, 특히 악성 투명세포형 신세포암을 진단함에 있어, 상기 [식 1] 내지 [식 3]에 비하여 더욱 효과적으로 진단할 수 있음을 확인하였다.
- [0048] 본 발명의 상기 A 내지 D의 각 계수 값이 하한 값 미만에 해당하거나, 상한 값을 초과하는 경우에는 이와 같은 식을 통해 도출된 위험 점수를 사용하여 암, 특히 악성 투명세포형 신세포암을 진단하는데 만족할만한 정확도 값을 얻지 못할 수 있다.
- [0049] 본 발명의 상기 위험 점수는 판단하고자 하는 개체가 암에 해당할 확률(P)를 의미하는 것으로서, 이를 통해 암을 진단해낼 수 있다.
- [0050] 본 발명의 상기 방법은 상기 위험 점수가 0.2 이상인 경우, 바람직하게는 0.2 내지 1, 보다 바람직하게 0.5 내지 1, 또는 0.7 내지 1인 경우 암으로 분류하는 단계를 더 포함할 수 있다. 바람직하게는 상기 위험 점수가 0.2 내지 1인 경우, 악성 비뇨기암으로 진단할 수 있다. 즉, 본 발명에서 상기 위험 점수가 0.2 이상이면 암이 악성, 즉 전이, 재발 또는 암으로 인한 사망률 증가의 확률이 높을 것으로 예측할 수 있다.
- [0051] 본 발명의 상기 미리 설정된 방법은 딥 러닝 분석(Deep learning analysis) 일 수 있다.
- [0052] 본 발명의 상기 딥 러닝 분석은 딥뉴럴네트워크(Deep neural networks; DNN), 딥컨볼루션뉴럴네트워크(Deep convolutional neural networks), 딥신뢰뉴럴네트워크(Deep belief networks; DBN) 및 순환뉴럴네트워크(Recurrent neural networks; RNN)로 구성된 군으로부터 선택되는 적어도 하나를 사용하여 복수의 데이터를 바탕으로 판단을 수행할 수 있는 시스템 또는 네트워크일 수 있고, 바람직하게는 딥뉴럴네트워크일 수 있으나, 이에 제한되는 것은 아니다.
- [0053] 본 발명의 상기 딥뉴럴네트워크는 하나 이상의 컴퓨터 내에 하나 이상의 층(Layer)을 구축하여 복수의 데이터를 바탕으로 판단을 수행하는 것으로서, 예를 들어, 딥뉴럴네트워크는 컨볼루션 풀링 층(Convolutional Pooling Layer), 로컬 접속 층(Locally-connected layer) 및 완전 연결 층(Fully-connected layer)을 포함하는 층들의 세트로 구현될 수 있다. 컨볼루션 풀링 층 또는 로컬 접속 층은 영상 내 특징들을 추출하도록 구성 될 수 있다. 완전 연결 층은 영상의 특징 간의 상관 관계를 결정할 수 있다. 본 발명의 목적상 상기 딥뉴럴네트워크는 첫번째 층이 시그모이드(Sigmoid)로, 두번째 층이 정류된 선형 유닛(Rectified linear unit)으로 설정된 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0055] 본 발명의 다른 구현 예에서는 암 진단에 관한 정보 제공 장치를 제공한다.
- [0056] 도 3을 참조하면, 본 발명의 상기 암 진단에 관한 정보 제공 장치는 버스(B)를 통해 연결되는 프로세서(110), 메모리(120), 입력부(130), 출력부(140), 스토리지(150) 및 네트워크 인터페이스(160)가 포함된다.
- [0057] 본 발명의 상기 암은 바람직하게는 악성 비뇨기암(Urological cancer)일 수 있으나, 이에 제한되는 것은 아니다.

- [0058] 본 발명의 상기 악성 암이란, 동시 전이(Synchronous metastasis), 재발(Recurrence) 및 암으로 인한 사망률(Cancer-specific death) 증가 중 적어도 하나가 발생된 암을 의미하며, 상기 동시 전이는 일차성(Primary) 암 진단 시 또는 진단으로부터 3개월 이내에 전이된 것을 의미한다.
- [0059] 본 발명의 상기 비뇨기암은 비뇨기의 암으로서, 방광암, 전립선암, 요관암 및 신장암으로 구성된 군으로부터 선택되는 적어도 하나인 것일 수 있고, 바람직하게는 신장암일 수 있으며, 더욱 바람직하게는 투명세포형 신세포암(Clear cell renal cell carcinoma; ccRCC)일 수 있으나, 이에 제한되는 것은 아니다.
- [0060] 본 발명의 상기 신장암은 조직학적 분류에 따른 세포형에 따라 투명세포형 신세포암과 비 투명신세포암(Non-clear renal cell carcinoma)로 분류된다. 상기 투명세포형 신세포암은 전체 신장암의 70-80%를 차지하며, 암 세포 주위의 혈관이 풍부하고 다른 장기로의 전이를 잘 일으키는 특징을 갖는다. 반면, 비 투명신세포암의 경우, 암 세포 주위의 혈관이 풍부하지 않으나, 젊은 연령 층에서 다발성으로 발생된다는 특징을 갖는다. 본 발명의 상기 방법은 조직 검사를 받기 이전 또는 조직 검사와 함께 상기 암의 조직학적 분류에 따른 종류를 구별해 낼 수 있고, 이를 통해 진단 당시 암 환자에게 최적의 치료 방법을 제공할 수 있다는 장점을 갖는다.
- [0061] 본 발명의 상기 투명세포형 신세포암은 종양의 크기가 7cm 이하인 것일 수 있으나, 이에 제한되는 것은 아니다. 상기 투명세포형 신세포암의 종양 크기가 7cm 보다 작은 초기 단계에서는 비 수술적 치료 방법, 예를 들면 동결 절제(Cryoablation), 고주파 열치료(Radiofrequency ablation)등이 적절하게 사용될 수 있다. 따라서, 본 발명의 목적상 상기와 같은 비 수술적 치료 방법에 의해서 충분한 치료 효과를 볼 수 없는 15%의 암 환자를 선별해 낼 수 있다.
- [0063] 이하, 본 발명의 상기 장치의 각 구성에 대해 자세히 설명한다.
- [0065] **프로세서(110)**
- [0066] 본 발명의 상기 프로세서(110)는 미리 설정된 방법에 의해 위험 점수를 산출하는 산출부(111)를 포함한다. 상기 프로세서(110)는 중앙처리장치 (CPU)로 구현되거나 또는 메모리(120) 및/또는 스토리지(150)에 저장된 명령어들에 대한 처리를 실행하는 반도체 장치로 구현될 수 있다.
- [0067] 본 발명의 상기 산출부(111)에 미리 설정된 방법은 하기 식 1 내지 4로 구성된 군으로부터 선택되는 적어도 하나의 식을 통해 위험 점수를 산출하는 것일 수 있다.
- [0069] [식 1]
- [0070] 위험 점수 = $1/[1 + \exp(-(A1 - A2 \times (PBRM1_norm) - A3 \times (BAP1_norm)))]$
- [0072] [식 2]
- [0073] 위험 점수 = $1/[1 + \exp(-(B1 - B2 \times (PBRM1_norm) - B3 \times (FOX2_norm)))]$
- [0075] [식 3]
- [0076] 위험 점수 = $1/[1 + \exp(-(C1 - C2 \times (BAP1_norm) - C3 \times (FOX2_norm)))]$
- [0078] [식 4]
- [0079] 위험 점수 = $1/[1 + \exp(-(D1 - D2 \times (FOX2_norm) - D3 \times (PBRM1_norm) - D4 \times (BAP1_norm)))]$
- [0081] 본 발명의 상기 식 1 내지 4에서, PBRM1_norm은 (PBRM1의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준)/0.06, BAP1_norm은 (BAP1의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준)/0.005, FOX2_norm은

(*FOXC2*의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준)/0.096일 수 있다.

- [0082] 본 발명의 일 실시예에서, 상기 유전자의 발현 수준은 상기 *FOXC2*, *PBRM1* 및 *BAP1*의 유전자를 상기 유전자들 각각에 특이적인 프라이머 쌍을 이용하여 qRT-PCR을 수행한 뒤, 항존 유전자, 예를 들면 *GAPDH*, α -tubulin 또는 β -actin 등의 발현량에 대한 상대적인 비율 값으로 나타낼 수 있다. 또한, 본 발명의 다른 실시예에서, 상기 단백질의 발현 수준은 *FOXC2*, *PBRM1* 및 *BAP1* 단백질 각각에 특이적인 항체를 이용하여 웨스턴 블롯 등을 수행한 뒤, 항존 유전자, 예를 들면 *GAPDH*, α -tubulin 또는 β -actin 등의 발현량에 대한 상대적인 비율 값으로 나타낼 수 있다.
- [0083] 본 발명의 상기 식 1에서, 상기 A1은 -1.5 내지 -0.9일 수 있고, 바람직하게는 -1.3 내지 -1.0, 더욱 바람직하게는 -1.2 내지 -1.05일 수 있으나, 이에 제한되는 것은 아니다. 또한, 상기 A2는 0.7 내지 1.1일 수 있고, 바람직하게는 0.8 내지 1.0, 더욱 바람직하게는 0.9 내지 1.0일 수 있으나, 이에 제한되는 것은 아니다. 또한, 상기 A3는 0.3 내지 0.7일 수 있고, 바람직하게는 0.4 내지 0.6일 수 있으며, 더욱 바람직하게는 0.5 내지 0.6일 수 있으나, 이에 제한되는 것은 아니다. 본 발명의 일 실시예에서는 상기 [식 1]에서 A1이 -1.097, A2가 0.963 및 A3이 0.524인 경우 *PBRM1* 및 *BAP1* 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준을 이용하여 암, 특히 악성 투명세포형 신세포암을 효과적으로 진단할 수 있음을 확인하였다.
- [0084] 본 발명의 상기 식 2에서, 상기 B1은 -1.5 내지 -0.9일 수 있고, 바람직하게는 -1.3 내지 -1.0, 더욱 바람직하게는 -1.2 내지 -1.05일 수 있으나, 이에 제한되는 것은 아니다. 또한, 상기 B2는 0.6 내지 1.0일 수 있고, 바람직하게는 0.7 내지 0.9일 수 있고, 더욱 바람직하게는 0.8 내지 0.9일 수 있으나, 이에 제한되는 것은 아니다. 또한, 상기 B3는 0.1 내지 0.5일 수 있고, 바람직하게는 0.2 내지 0.4일 수 있으며, 더욱 바람직하게는 0.3 내지 0.4일 수 있으나, 이에 제한되는 것은 아니다. 본 발명의 일 실시예에서는 상기 [식 2]에서 B1이 -1.100, B2가 0.894, B3이 0.340인 경우, *PBRM1* 및 *FOXC2* 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준을 이용하여 암, 특히 악성 투명세포형 신세포암을 효과적으로 진단할 수 있음을 확인하였다.
- [0085] 본 발명의 상기 식 3에서, 상기 C1은 -1.5 내지 -0.9일 수 있고, 바람직하게는 -1.3 내지 -1.0, 더욱 바람직하게는 -1.2 내지 -1.05일 수 있으나, 이에 제한되는 것은 아니다. 또한, 상기 C2는 0.3 내지 0.7일 수 있고, 바람직하게는 0.4 내지 0.6일 수 있으며, 더욱 바람직하게는 0.5 내지 0.6일 수 있으나, 이에 제한되는 것은 아니다. 또한, 상기 C3는 0.3 내지 0.7일 수 있고, 바람직하게는 0.4 내지 0.6일 수 있으며, 더욱 바람직하게는 0.5 내지 0.6일 수 있으나, 이에 제한되는 것은 아니다. 본 발명의 일 실시예에서는 상기 [식 3]에서 C1이 -1.11, C2가 0.533, C3이 0.562인 경우 *BAP1* 및 *FOXC2* 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준을 이용하여 암, 특히 악성 투명세포형 신세포암을 효과적으로 진단할 수 있음을 확인하였다.
- [0086] 본 발명의 상기 식 4에서, 상기 D1은 -1.5 내지 -0.9일 수 있고, 바람직하게는 -1.0 내지 -1.3, 더욱 바람직하게는 -1.05 내지 -1.2일 수 있으나, 이에 제한되는 것은 아니다. 또한, 상기 D2는 0.1 내지 0.5일 수 있고, 바람직하게는 0.2 내지 0.4일 수 있으며, 더욱 바람직하게는 0.3 내지 0.4일 수 있으나, 이에 제한되는 것은 아니다. 또한, 상기 D3는 0.6 내지 1.0일 수 있고, 바람직하게는 0.7 내지 0.9일 수 있고, 더욱 바람직하게는 0.8 내지 0.9일 수 있으나, 이에 제한되는 것은 아니다. 또한, 상기 D4는 0.2 내지 0.8일 수 있고, 바람직하게는 0.3 내지 0.7일 수 있으며, 더욱 바람직하게는 0.4 내지 0.6일 수 있으나, 이에 제한되는 것은 아니다. 본 발명의 또 다른 실시예에서는 상기 [식 4]에서 D1이 -1.082, D2가 0.336, D3이 0.887, D4가 0.521인 경우 *FOXC2*, *PBRM1* 및 *BAP1* 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준을 이용하여 암, 특히 악성 투명세포형 신세포암을 진단함에 있어, 상기 [식 1] 내지 [식 3]에 비하여 더욱 효과적으로 진단할 수 있음을 확인하였다.
- [0087] 본 발명의 상기 A 내지 D의 각 계수 값이 하한 값 미만에 해당하거나, 상한 값을 초과하는 경우에는 이와 같은 식을 통해 도출된 위험 점수를 사용하여 암, 특히 악성 투명세포형 신세포암을 진단하는데 만족할만한 정확도 값을 얻지 못할 수 있다.
- [0088] 본 발명의 상기 위험 점수는 판단하고자 하는 개체가 암, 특히 악성 암에 해당할 확률(P)를 의미하는 것으로서, 이를 통해 암을 진단해낼 수 있다.
- [0089] 본 발명의 상기 위험 점수는 상기 식 1 내지 4를 통해, 판단하고자 하는 개체가 암에 속할 확률(P)를 의미하는 것으로서, 이를 통해 암을 진단해낼 수 있다.
- [0090] 본 발명의 상기 산출부는 상기 위험 점수가 0.2 이상인 경우, 바람직하게는 0.2 내지 1, 보다 바람직하게 0.5 내지 1, 또는 0.7 내지 1인 경우 암으로 분류하는 단계를 더 포함할 수 있다. 바람직하게는 상기 위험 점수가 0.2 내지 1인 경우, 악성 비뇨기암으로 진단할 수 있다. 즉, 본 발명에서 상기 위험 점수가 0.2 이상이면 암이

악성, 즉 전이, 재발 또는 암으로 인한 사망률 증가의 확률이 높을 것으로 예측될 수 있다.

[0091] 본 발명의 산출부(111)에 미리 설정된 방법은 딥 러닝 분석 방법일 수 있다.

[0092] 본 발명의 상기 딥러닝 분석 방법은 딥뉴럴네트워크, 딥컨볼루션뉴럴네트워크, 딥신뢰뉴럴네트워크 및 순환뉴럴네트워크로 구성된 군으로부터 선택되는 적어도 하나를 사용하여 복수의 데이터를 바탕으로 판단을 수행할 수 있는 시스템 또는 네트워크일 수 있고, 바람직하게는 딥뉴럴네트워크일 수 있으나, 이에 제한되는 것은 아니다.

[0093] 본 발명의 상기 딥뉴럴네트워크는 하나 이상의 컴퓨터 내에 하나 이상의 층(Layer)을 구축하여 복수의 데이터를 바탕으로 판단을 수행하는 것으로서, 예를 들어, 심층신 경망은 컨볼루션 풀링 층, 로컬 접속 층 및 완전 연결 층을 포함하는 층들의 세트로 구현될 수 있다. 컨볼루션 풀링 층 또는 로컬 접속 층은 영상 내 특징들을 추출하도록 구성 될 수 있다. 완전 연결 층은 영상의 특징 간의 상관 관계를 결정할 수 있다. 본 발명의 목적상 상기 딥뉴럴네트워크는 첫번째 층이 시그모이드로, 두번째 층이 정류된 선형 유닛으로 설정된 것일 수 있으나, 이에 제한되는 것은 아니다.

[0094] **메모리(120)**

[0095] 본 발명의 상기 메모리(120)는 상기 프로세서(110)의 동작을 위한 프로그램을 저장할 수 있고, 입/출력되는 데이터들을 임시 저장할 수도 있다. 이러한 메모리(120)는 ROM(Read Only Memory)(121) 및 RAM(Random Access Memory)(122) 등을 포함할 수 있다.

[0096] **입력부(130)**

[0097] 본 발명의 입력부(130)는 목적하는 개체로부터 분리된 생물학적 시료에서, *FOXC2*, *PBRM1* 및 *BAP1*으로 구성된 군 으로부터 선택되는 2종 이상의 유전자 또는 이에 의해 암호화되는 단백질이 존재하는 수준이 측정된 값이 사용자의 조작에 의해 입력됨으로써 데이터가 발생하는 것일 수 있다. 상기 입력부(130)는 키패드, 터치패드, 또는 조그 스위치 등의 사용자 인터페이스로 구현될 수 있으나, 이에 제한되는 것은 아니다.

[0098] 본 발명의 상기 *FOXC2* 유전자 또는 이에 의해 암호화되는 단백질은, *FOXC2* 유전자는 출생 전 배아의 기관과 조직의 형성에 중요한 역할을 하는 전사인자 단백질인 *FOXC2* 단백질(FKHL14(forkhead-related protein FKHL14), 전사인자 FKH-14, MFH1(mesenchyme fork head protein 1))을 암호화하는 것으로서, 유방암의 경우 상기 *FOXC2* 유전자에 의해 암호화되는 단백질의 발현이 증가되는 경우 유방암의 전이를 유도할 수 있다는 보고가 존재한다. 본 발명의 목적상 상기 *FOXC2* 유전자 또는 이에 의해 암호화되는 단백질은 비노기암에서 유방암에서와 달리, 세포 사이클이 잘못되는 경우를 체크하는 역할을 수행하는 것으로, 이와 같은 유전자 또는 이에 의해 암호화되는 단백질의 발현이 감소되었을 때 비노기암의 전이, 재발 및/또는 암으로 인한 사망률이 높아지는 것일 수 있다.

[0099] 본 발명의 상기 *PBRM1* 유전자 또는 이에 의해 암호화되는 단백질은, ATP-의존성 크로마틴-리모델링 복합체의 서브유닛을 암호화하는 유전자로서, 이와 같은 유전자에 의해 암호화된 단백질은 핵 호르몬 수용체에 의한 리간드의 의존성 전화 활성화에 필요한 복합체의 필수 성분에 해당한다.

[0100] 본 발명의 상기 *BAP1* 유전자 또는 이에 의해 암호화되는 단백질은, 유비퀴틴 카르복실-말단 가수분해 효소 BAP1(ubiquitin carboxyl-terminal hydrolase BAP1)을 암호화하는 유전자로서, 상기 효소는 탈유비퀴틴화 효소로서 작용하여 다양한 세포 과정에 관여하는 많은 단백질의 기능을 조절하는데 도움을 주는 역할을 한다.

[0101] 본 발명의 상기 장치에서 상기 측정의 대상이 되는 유전자 또는 이에 의해 암호화되는 단백질은 *FOXC2*, *PBRM1* 및 *BAP1*으로 구성된 군으로부터 선택되는 2종 이상을 포함하는 것일 수 있고, 바람직하게는 *FOXC2*와, *PBRM1* 또는 *BAP1* 중 어느 하나 이상을 포함하는 것일 수 있고, 더욱 바람직하게는 상기 *FOXC2*, *PBRM1* 및 *BAP1*을 모두 포함하는 것일 수 있으나, 이에 제한되는 것은 아니다. 이와 같이 상기 *FOXC2*, *PBRM1* 및 *BAP1*을 모두 포함하는 경우에는 각각 개개의 유전자 또는 이에 의해 암호화되는 단백질을 측정하는 경우에 비하여 예측 정확도를 현저하게 높일 수 있다.

[0102] 본 발명의 상기 유전자가 존재하는 수준의 측정은 프라이머를 이용하여 상기 유전자를 증폭시키거나, 형광 등으로 표지된 프로브와 상보적으로 결합하는 등의 통상의 방법을 통해 확인될 수 있는 것이고, 예를 들면, 중합효소연쇄반응(polymerase chain reaction), 형광상관분석법(fluorescence correlation spectroscopy), 마이크로어레이(microarray) 및 칩어세이(chip-assay)으로 이루어진 군으로부터 선택되는 적어도 하나일 수 있으나, 이에 제한되는 것은 아니다. 상기 프라이머 또는 프로브는 상기 유전자의 염기 서열을 참조하여 당해 기술 분야에서 통상의 지식을 가진 자가 공지된 방법에 따라 쉽게 제작할 수 있다.

[0103] 본 발명의 상기 단백질이 존재하는 수준의 측정은 상기 단백질에 특이적인 항체 또는 앵타머를 이용하여 통상의 방법을 통해 확인될 수 있는 것이고, 예를 들면 웨스턴 블롯 분석(Western blot assay), ELISA(Enzyme linked immunosorbent assay), 방사선면역분석 (RIA: Radioimmunoassay), 방사 면역 확산법(Radioimmunodiffusion), 오우크테로니(Ouchterlony) 면역 확산법, 로케트 면역전기영동(Rocket immunoelectrophoresis), 면역조직화학 염색법(Immunohistochemical staining), 면역 침전 분석법(Immunoprecipitation Assay), 보체 고정 분석법(Complement Fixation Assay), 면역형광법 (Immunofluorescence), 면역크로마토그래피법(Immunochromatography), FACS(Fluorescenceactivated cell sorter analysis) 및 단백질 칩 분석법(protein chip technology assay)으로 이루어진 군으로부터 선택되는 적어도 하나일 수 있으나, 이에 제한되는 것은 아니다. 상기 항체 또는 프라이머는 상기 단백질의 아미노산 서열을 참조하여 당해 기술 분야에서 통상의 지식을 가진 자가 공지된 방법에 따라 쉽게 제작할 수 있다.

[0104] 표시부(140)

[0105] 본 발명의 상기 표시부(140)는 상기 산출부로부터 전송 받은 값에 따라 발생하는 데이터를 표시한다. 상기 산출부에서 위험 점수가 0.2 이상인 경우, 바람직하게는 0.2 내지 1, 보다 바람직하게 0.5 내지 1, 또는 0.7 내지 1 인 경우에 해당하여 암으로 분류된 경우, 상기 표시부에 암으로 표시될 수 있다. 바람직하게는 상기 위험 점수가 0.2 이상인 경우 양성 비노기암으로 표시될 수 있으나, 이에 제한되는 것은 아니다.

[0106] 본 발명의 상기 표시부(140)는 디스플레이 및 스피커 등의 출력수단으로 구성될 수 있으나, 이에 제한되는 것은 아니다.

[0107] 스토리지(150)

[0108] 본 발명의 상기 스토리지(150)는 상기 메모리(120)와 같이 다양한 종류의 휘발성 또는 비휘발성 저장매체를 포함할 수 있다. 상기 스토리지(150)는 인터넷(Internet)상에서 메모리(120)의 저장 기능을 수행하는 웹 스토리지(Web storage)로 구현될 수 있으나, 이에 제한되는 것은 아니다.

[0109] 네트워크 인터페이스(160)

[0110] 본 발명의 상기 네트워크 인터페이스(160)는 네트워크를 통해 다른 단말기와 유무선 통신을 수행할 수 있게 한다. 상기 네트워크 인터페이스(160)는 무선 인터넷, 이동 통신, 및 근거리 통신 등의 통신기술 중 어느 하나 이상을 사용할 수 있으나, 이에 제한되는 것은 아니다.

[0112] 본 발명의 또 다른 구현 예에서는 암 치료제 스크리닝 방법을 제공한다.

[0113] 본 발명의 상기 스크리닝 방법은 암 개체로부터 분리된 생물학적 시료에, 후보물질을 처리하는 단계; 상기 후보물질이 처리된 생물학적 시료에서 *FOX2*, *PBRM1* 및 *BAP1*으로 구성된 군으로부터 선택되는 2종 이상의 유전자 또는 이에 의해 암호화되는 단백질이 존재하는 수준을 측정하는 단계; 및 상기 유전자 또는 상기 단백질이 존재하는 수준의 측정된 값을 미리 설정된 방법에 적용하여 위험 점수를 산출하는 단계;를 포함한다.

[0114] 본 발명의 상기 스크리닝 방법은 상기 후보물질을 처리하기 이전의 위험 점수에 비하여, 후보물질을 처리한 이후의 위험 점수가 낮은 경우, 처리된 후보물질을 암 치료제로 선별하는 단계가 더 포함될 수 있다.

[0115] 본 발명의 상기 스크리닝 방법에서, 암, 악성 암, 비노기암, 투명세포형 신세포암, *FOX2*, *PBRM1* 및 *BAP1* 유전자 또는 이에 의해 암호화되는 단백질, 신장암, 측정 방법, 미리 설정된 방법, 생물학적 시료 및 위험 점수 등은 본 발명의 상기 진단 방법에서 기재한 바와 동일하여, 본 명세서의 과도한 복잡성을 피하기 위하여 생략한다.

[0116] 본 발명의 상기 후보물질은 암을 치료하기 위한 활성이 존재하는지 여부에 대해 테스트하기 위한 약제를 의미하며, 단백질, 올리고 펩타이드, 유기 분자, 다당류, 폴리뉴클레오타이드 및 광범위한 화합물 등의 임의의 분자를 포함한다. 이러한 후보물질은 천연물질뿐만 아니라, 합성 물질도 모두 포함하는 것일 수 있다.

[0118] [서열 목록]

[0119] 서열번호 1: *FOX2* 유전자

[0120] 1 gaaacttttc ccaatcccta aaagggaactt ggcctctttt tctgggctca gcggggcagc

[0121] 61 cgctcggacc cggcgcgct gacctcggg gctgccgatt cgctgggggc ttggagagcc

[0122] 121 tcttgcgccc ctctcgcgc gggccgaggg tccacctggg tcccaggcc gcggcgtctc

[0123] 181 cgctgggtcc gcggccgccc gcctgccccg gctgccgccc ccgggtcctg gagccagcga

[0124] 241 ggagcggggc cggcgctgcg cttgcccggg gcgcgccctc caggatgccg atccgcccgg

[0125] 301 tccgctgaaa gcgcgcgccc ctgctcggcc cagcgcgcgc cgcccgcgca cctcgcgcc

[0126] 361 ggaggctgcc aggagcccgg ggccgcccct ccgctcccc tctctcccc ctctggctct

[0127] 421 ctgcgctct ctgcctctca gggccccct cgctccccg gccgcagtcc gtgcgcgagg

[0128] 481 gcgcggcgca gccgtctcgg aagcagcatg caggcgcgct actcgtgtc cgacccaac

[0129] 541 gccctgggag tgggtgcccta cctgagcgag cagaattact accgggctgc gggcagctac

[0130] 601 ggcggcatgg ccagccccat gggcgtctat tccggccacc cggagcagta cagcgcgggg

[0131] 661 atgggcccgt cctacgcgcc ctaccaccac caccagccc cggcgcctaa ggacctggtg

[0132] 721 aagccgccct acagctacat cgcgtctat accatggcca tccagaacgc gcccgagaag

[0133] 781 aagatcacct tgaacggcat ctaccagttc atcatggacc gtttccctt ctaccgggag

[0134] 841 aacaagcagg gctggcagaa cagcatccgc cacaacctt cgctcaacga gtgcttcgtc

[0135] 901 aaggtgcccc gcgacgacaa gaagcccggc aagggcagtt actggaccct ggacccggac

[0136] 961 tctacaaca tgttcagaa cggcagcttc ctgcggcgcc ggcggcgctt caaaaagaag

[0137] 1021 gacgtgtcca aggagaagga ggagcgggcc cacctcaagg agccgcccc ggcggcgctc

[0138] 1081 aaggcgcccc cggccacccc ccacctagcg gacgccccca aggaggccga gaagaaggtg

[0139] 1141 gtgatcaaga gcgagggcgc gtccccggcg ctgccggtca tcaccaaggt ggagacgtg

[0140] 1201 agccccgaga gcgcgctgca gggcagcccc cgcagcgccg cctccacgcc cgccggtcc

[0141] 1261 cccgacggct cgctgccgga gcaccacgcc gcggcgcccc acgggctgcc tggcttcagc

[0142] 1321 gtggagaaca tcatgacct gcgaacgtcg ccgcccggcg gagagctgag cccggggggc

[0143] 1381 ggacgcgcgg gcctggtggt gccgccgctg gcgctgccct acgccgcgc gccgcccgc

[0144] 1441 gcctacggcc agccgtgcgc tcagggcctg gaggccgggg ccgcccgggg ctaccagtgc

[0145] 1501 agcatgcgag cgatgacct gtacaccggg gccgagcggc cggcgcatat gtgcgtccg

[0146] 1561 cccgccctgg acgaggccct ctcgaccac ccgagcggcc ccagctgcc cctgagcgt

[0147] 1621 ctcaacctcg ccgcccggca ggaggcgcg ctcccgcca cgggccacca ccaccagcac

[0148] 1681 caggccacc accaccgca ggcgccgccc ccccgcccgg ctcccagcc ccagccgacg

[0149] 1741 ccgagcccc gggccgccc gccgcaggcg gcctctggt atctcaacca cagcggggac

[0150] 1801 ctgaaccacc tccccggca cagttcgcg gccagcagc aaactttccc caacgtgcgg

[0151] 1861 gagatgttca actcccacc gctggggatt gagaactga cctcggggga gtcccaggtg

[0152] 1921 agtggcaatg ccagtgcca gctgccctac agatccacgc cgcctctcta tcgccacga

[0153] 1981 gcccctact cctacgactg cagaaatac tgacgtgtcc cgggacctcc cctccccgc

[0154] 2041 ccgctccggc ttcgcttccc agccccgacc caaccagaca attaaggggc tgcagagacg

[0155] 2101 caaaaaagaa aaaaacatg tccaccaacc ttttctcaga cccgggagca gagagcgggc

[0156] 2161 acgctagccc ccagccgtct gtgaagagcg caggtaactt taattcgccg ccccgtttct
 [0157] 2221 gggatcccag gaaaccctc caaaggacg cagcccaaca aaatgagtat tgatcttaaa
 [0158] 2281 atccccctcc cctaccagga cggtctgtct gtgctcgacc tgagctttca aaagttaagt
 [0159] 2341 tatggaccaa atcccatagc gagcccctag tgactttctg taggggtccc cataggtgta
 [0160] 2401 tgggggtctc tatagataat atatgtgctg tgtgtaattt taaatttctc caaccgtgct
 [0161] 2461 gtacaaatgt gtggatttgt aatcaggcta ttttgttgtt gttgttgtt ttcagagcca
 [0162] 2521 ttaatataat atttaaagtt gagttcactg gataagtttt tcactttgcc caaccatttc
 [0163] 2581 taactgccaa attgaattca agaaaccgat gtgggttttg tttcctgtac aattatgaga
 [0164] 2641 tataattctt tttccattg taggtctttt acaaaacaag aaaataattt atttttttgt
 [0165] 2701 tgggtgataa agaagtcaag tatctgatac tttttatita caaagtgtga tggttttgta
 [0166] 2761 tagtaggttc caccctgagt attcctaaaa gaaaaaaaaa aaaaaagctt aaaaactcta
 [0167] 2821 acttcatctg tgtttgtctt acgtggctct aatcgttgta cttaccttaa aataaaccca
 [0168] 2881 tgttgttttt tctgcccata

[0170] 서열번호 2: FOXC2 단백질

[0171] 1 mqarysvsdp nalgvvpyls eqnyyraags yggmaspmgv ysghpeqysa gmgrsyapyh
 [0172] 61 hhqpaapkd vlkppysial itmaiqnape kkitlngiyq fimdrfpfyr enkqgwqnsi
 [0173] 121 rhnlslnecf vkvprddkkp gkgsywtldp dsynmfengs flrrrrrfkk kdvskekeer
 [0174] 181 ahlkepppaa skgapatphl adapkeaekk vvikseaasp alpvitkvet lpsedalqgs
 [0175] 241 prsaastpag spdglpehh aaapnlgpgf svenimtlrt sppggelspg agraglvvpp
 [0176] 301 lalpyaaapp aaygqpcaqg leagaaggyq csmramslyt gaerpahmcv ppaldealsd
 [0177] 361 hpsgptspls alnlaagqeg alaathghhq hhghhqpap ppppapqpqp tpqpgaaaaq
 [0178] 421 aaswylnhsg dlhlpghthf aaqqtfpvn remfnshrlg ienstlgesq vsnascqlp
 [0179] 481 yrstpplyrh aapysydcy y

[0181] 서열번호 3: PBRM1 유전자

[0182] 1 gcacctgccg gcagggacaa aggcgcacta cgggtcggaa actccccgag ctctcttttc
 [0183] 61 ccgctcgtcg cggggcagct tcaaagctgt caacgtttcc ctacgtcccc aataccagtg
 [0184] 121 acagcttgcc aacacttggt gtcacatgtg agcctccac atgtattcac tctccattcc
 [0185] 181 agctctgtga ttgaactctg ctcttattga ctagggggca gttgggcagg catgcctcat
 [0186] 241 tcctggaatt gacagtcatt cctaattttt gaataacta acactgaatc agcacttcta
 [0187] 301 aagcccttcc tttcccact ggcttcactt ggctttcaga cataatgagg agactggctt
 [0188] 361 ttcgagggcg tgggtgtgct ctggtaaagc tgaagaagtt ggattccatg ggttccaaga
 [0189] 421 gaagaagagc tacctccct tccagcagtg tcagcgggga ctttgatgat gggcaccatt
 [0190] 481 ctgtgtcaac accaggccca agcaggaaaa ggaggagact ttccaatctt ccaactgtag
 [0191] 541 atcctattgc cgtgtgccat gaactctata ataccatccg agactataag gatgaacagg

[0192] 601 gcagacttct ctgtgagctc ttcattaggg caccaaagcg aagaaatcaa ccagactatt

[0193] 661 atgaagtgggt ttctcagccc attgacttga tgaatcca acagaaacta aaaatggaag

[0194] 721 agtatgatga tgtaatttg ctgactgctg acttccagct tctttttaac aatgcaaagt

[0195] 781 cctattataa gccagattct cctgaatata aagccgcttg caaactctgg gatttgtacc

[0196] 841 ttcgaacaag aaatgagttt gttcagaaag gagaagcaga tgacgaagat gatgatgaag

[0197] 901 atgggcaaga caatcagggc acagtgactg aaggatcttc tccagcttac ttgaaggaga

[0198] 961 tcctggagca gcttcttgaa gccatagttg tagctacaaa tccatcagga cgtctcatta

[0199] 1021 gcgaactttt tcagaaactg ccttctaaag tgcaatatcc agattattat gcaataatta

[0200] 1081 aggagcctat agatctcaag accattgccc agaggataca gaatggaagc tacaaaagta

[0201] 1141 ttcattgcaat ggccaaagat atagatctcc tcgcaaaaaa tgccaaaact tataatgagc

[0202] 1201 ctggctctca agtattcaag gatgcaaatt caattaaaaa aatattttat atgaaaaagg

[0203] 1261 ctgaaattga acatcatgaa atggctaagt caagtcttcg aatgaggact ccatccaact

[0204] 1321 tggctgcagc cagactgaca ggtccttcac acagtaaagg cagccttggg gaagagagaa

[0205] 1381 atcccactag caagtattac cgtaataaaa gagcagtaca aggaggtcgt ttatcagcaa

[0206] 1441 ttacaatggc acttcaatat ggctcagaaa gtgaagaaga tgctgcttta gctgctgcac

[0207] 1501 gctatgaaga gggagagtca gaagcagaaa gcatcacttc ctttatggat gtttcaaac

[0208] 1561 ctttttatca gctttatgac acagttagga gttgtcggaa taaccaaggg cagctaatag

[0209] 1621 ctgaaccttt ttaccatttg ctttcaaaga aaaaatacc tgattattac cagcaaatta

[0210] 1681 aaatgcccac atcactacaa cagatccgaa caaaactgaa gaatcaagaa tatgaaactt

[0211] 1741 tagatcattt ggagtgtgat ctgaatttaa tgtttgaaaa tgccaaacgc tataatgtgc

[0212] 1801 ccaattcagc catctacaag cgagttctaa aattgcagca agttatgcag gcaaagaaga

[0213] 1861 aagagcttgc caggagagac gatatcgagg acggagacag catgatctct tcagccacct

[0214] 1921 ctgatactgg tagtgccaaa agaaaaagga acactcatga cagtgagatg ttgggtctca

[0215] 1981 ggaggctatc cagtaaaaag aacataagaa agcagcgaat gaaaatctta ttcaatgttg

[0216] 2041 ttcttgaagc tcgagagcca ggttcaggca gaagactttg tgacctattt atggttaaac

[0217] 2101 catcaaaaaa ggactatcct gattattata aaatcatctt ggagccaatg gacttgaaaa

[0218] 2161 taattgagca taacatccgc aatgacaaat atgctggtga agagggaatg atagaagaca

[0219] 2221 tgaagctgat gttccggaat gccaggcact ataatgagga gggctcccag gtttataatg

[0220] 2281 atgcacatat cctggagaag ttactcaagg agaaaaggaa agagctgggc ccaatgcctg

[0221] 2341 atgatgatga catggcttct cccaaactca agctgagtag gaagagtggc atttctccta

[0222] 2401 aaaaatcaaa atacatgact ccaatgcagc agaaactaaa tgaggtctat gaagctgtaa

[0223] 2461 agaactatac tgataagagg ggtcgcgcc tcagtcccat atttctgagg ctccctcta

[0224] 2521 gatctgagtt gcctgactac tatctgacta ttaaaaagcc catggacatg gaaaaaatc

[0225] 2581 gaagtcacat gatggccaac aagtaccaag atattgactc tatggttgag gactttgtca

[0226] 2641 tgatgtttaa taatgcctgt acatacaatg agccggagtc ttgatctac aaagatgctc

[0227] 2701 ttgttctaca caaagtcctg cttgaaacac gcagagacct ggaggagat gaggactctc

[0228] 2761 atgtcccaaa tgtgactttg ctgattcaag agcttatcca caatcttttt gtgtcagtca

[0229] 2821 tgagtcacatca ggatgatgag ggaagatgct acagcgattc tttagcagaa attcctgctg

[0230] 2881 tggatcccaa ctttcctaac aaaccacccc ttacatttga cataattagg aagaatgttg

[0231] 2941 aaaataatcg ctaccgtcgg cttgatttat ttcaagagca tatgtttgaa gtattggaac

[0232] 3001 gagcaagaag gatgaatcgg acagattcag aaatatatga agatgcagta gaacttcagc

[0233] 3061 agttttttat taaaattcgt gatgaactct gcaaaaatgg agagattctt ctttcaccgg

[0234] 3121 cactcagctia taccacaaaa catttgcata atgatgtgga gaaagagaga aaggaaaaat

[0235] 3181 tgccaaaaga aatagaggaa gataaactaa aacgagaaga agaaaaaaga gaagctgaaa

[0236] 3241 agagtgaaga ttccctctggt gctgcaggcc tctcaggctt acatcgcaaca tacagccagg

[0237] 3301 actgtagctt taaaaacagc atgtaccatg ttggagatta cgtctatgtg gaacctgcag

[0238] 3361 aggccaacct acaaccacat atcgtctgta ttgaaagact gtgggaggat tcagctggtg

[0239] 3421 aaaaatggtt gtatggctgt tggttttacc gaccaaatga aacattccac ctggctacac

[0240] 3481 gaaaatttct agaaaaagaa gtttttaaga gtgactatta caacaaagt ccagttagta

[0241] 3541 aaattctagg caagtgtgtg gtcattgttg tcaaggaata ctttaagtta tgcccagaaa

[0242] 3601 acttccgaga tgaggatgtt tttgtctgtg aatcacggta ttctgccaaa accaaatctt

[0243] 3661 ttaagaaaat taaactgtgg accatgccca tcagctcagt caggtttgtc cctcgggatg

[0244] 3721 tgcctctgcc tgtggttcgc gtggcctctg tatttgcaaa tgcagataaa ggtgatgatg

[0245] 3781 agaagaatac agacaactca gaggacagtc gagctgaaga caattttaac ttggaaaagg

[0246] 3841 aaaaagaaga tgtccctgtg gaaatgtcca atgggtgaacc aggttgccac tactttgagc

[0247] 3901 agctccatta caatgacatg tggctgaagg ttggcgactg tgtcttcac aagtcccatg

[0248] 3961 gcctggtgcg tcctcgtgtg ggcagaattg aaaaagtatg gggttcgagat ggagctgcat

[0249] 4021 atttttatgg ccccatcttc attcaccagc aagaaacaga gcatgagccc acaaaaatgt

[0250] 4081 tctacaaaaa agaagtattt ctgagtaatc tggaagaaac ctgccccatg acatgtattc

[0251] 4141 tcggaaaagt tgctgtgttg tcattcaagg acttcctctc ctgcaggcca actgaaatac

[0252] 4201 cagaaaatga cattctgctt tgtgagagcc gctacaatga gagcgacaag cagatgaaga

[0253] 4261 aattcaaagg attgaagagg ttttactct ctgctaaagt ggtagatgat gaaatttact

[0254] 4321 acttcagaaa accaattgtt cctcagaagg agccatcacc tttgctggaa aagaagatcc

[0255] 4381 agttgctaga agctaaattt gccgagttag aaggtggaga tgatgatatt gaagagatgg

[0256] 4441 gagaagaaga tagtgaggtc attgaacctc cttctctacc tcagcttcag acccccctgg

[0257] 4501 ccagttagct ggacctcatg ccctacacac cccacagtc taccccaaag tctgccaag

[0258] 4561 gcagtgcaaa gaaggaaggc tccaaacgga aaatcaacat gagggtgtac atcctgttca

[0259] 4621 gcagttagat gagggtgtg attaaggccc aacaccaga ctactcttc ggggagctca

[0260] 4681 gccgcctggt ggggacagaa tggagaaatc ttgagacagc caagaaagca gaatatgaag

[0261] 4741 gtgtgatgaa ccaaggagtg gccctatgg tagggactcc agcaccaggt ggaagtccat

[0262] 4801 atggacaaca ggtgggagtt ttggggcctc cagggcagca ggcaccacct ccataatccc

[0263] 4861 gcccacatcc agctggacc cctgtcatc agcagccaac aacacccatg tttgtagctc

[0264] 4921 cccacacaaa gaccagcgg cttcttact cagaggccta cctgaaatac attgaaggac

[0265] 4981 tcagtgcgga gtccaacagc attagcaagt gggatcagac actggcagct cgaagacgcg

[0266] 5041 acgtccattt gtcgaaagaa caggagagcc gcctaccctc tcaactggctg aaaagcaaag

[0267] 5101 gggcccacac caccatggca gatgccctct ggcgccttcg agatttgatg ctccgggaca

[0268] 5161 ccctcaacat tcgccaagca tacaacctag aaaatgttta atcacatcat tacgtttctt

[0269] 5221 ttatatagaa gcataaagag ttgtggatca gtagccattt tagttactgg ggggtggggg

[0270] 5281 aaggaacaaa ggaggataat ttttattgca ttttactgta catcacaagg ccatttttat

[0271] 5341 atacggacac ttttaataag ctatttcaat ttgtttgtta tattaagtig actttatcaa

[0272] 5401 atacacaaag atttttttgc atatgtttcc ttcgttttaa accagtttca taattggttg

[0273] 5461 tataatgtaga ctggagttt tatcttttta ctgttgcca tggaaactgaa accattagag

[0274] 5521 gtttttgctc tggtttggg tttttgttt ctgtgtttg ggttttttta tatatatata

[0275] 5581 taaaagaaca aaatgaaaaa aaacacacac acacaagagt ttacagatta gtttaaatg

[0276] 5641 ataataaat gtgaagttt tcctagtta catcttagag aggggagtat acttgtgtt

[0277] 5701 gtttcatgtg cctgaatgc ttaagccact ttctgcaaaa gctgtttctt acagatgaag

[0278] 5761 tgctttcttt gaaaggtggt tatttaggtt ttagatgtt aatagacaca gcacatttgc

[0279] 5821 tctattaact cagaggctca ctacagaaat atgtaacag tgctgtgcat ctgtctgcag

[0280] 5881 ctaatgtacc tcctggacac caggagggga aaaagcactt tttcaattgt gctgagttag

[0281] 5941 acatctgtga gttagactat ggtgtcagtg atttttgcag aacacgtgca caacctgag

[0282] 6001 gtatgtttaa tctaggcagg tacgtttaag gatattttga tctatttata atgaattcac

[0283] 6061 aatttatgcc tataaatctc agatgattta aaattttaaa cctgttacat tgaataacat

[0284] 6121 tgaagtctgt ctgaagaaa gcattaaggt atgcatggag gtgatttatt tttaaacata

[0285] 6181 acacctaacc taacatgggt aagagagtat ggaactagat atgagctgta taagaagcat

[0286] 6241 aattgtgaac aagtagattg attgccctca tatacaagta tgttttagta ttccttattt

[0287] 6301 ccttattatc agatgtattt tttcttttaa gtttcaatgt tgttataatt ctcaaccaga

[0288] 6361 aatttaatac tttctaaaat attttttaaa tttagcttgt gcttttgaat tacaggagaa

[0289] 6421 gggaatcata atttaataaa acgttacta gaaagacat tacagatccc aaacacttgg

[0290] 6481 gtttgggtgac cctgtcttcc ttatatgacc ctacaataaa catttgaagg cagcatagga

[0291] 6541 tggcagacag taggaacatt gtttcaattg gcggcatgtt ttgaaacct gctttatagt

[0292] 6601 aactgggtga ttgccattgt ggtagagctt ccaactgctgt ttataatctg agagagttaa

[0293] 6661 tctcagagga tgcttttttc cttttaatct gctatgaatc agtaccaga tgtttaatta

[0294] 6721 ctgtacttat taaatcatga gggcaaaaaga gtgtagaatg gaaaaaagtc tcttgtatct

[0295] 6781 agatacttta aatatgggag gccctttaac ttaattgcct ttagtcaacc actggatttg

[0296] 6841 aatttgcac aagtatttta aataatattg aatttaaaaa aatgtattgc agtagtgtgt

[0297] 6901 cagtacctta ttgttaaagt gattcagata aatcttcaat tcctggctat ttgggcaatt

[0298] 6961 gaatcatcat ggactgtata atgcaatcag attattttgt ttctagacat ccttgaatta

[0299] 7021 caccaaagaa catgaaattt agttgtggtt aaattattta tttatttcat gcattcattt

[0300] 7081 tatttccctt aaggtctgga tgagacttct ttggggagcc tctaaaaaa ttttctactg
 [0301] 7141 ggggccacgt gggtcattag aagccagagc tctctccag gtccttccc agtgcctaga
 [0302] 7201 ggtgctatag gaaacataga tccagccagg ggcttcccta aagcagtga gcaccggccc
 [0303] 7261 agggcatcac tagacaggcc ctaattaagt tttttttaa aagcctgtgt atttatatta
 [0304] 7321 gaatcatgtt tttctgtata ttaacttggg ggatatcgtt aatatttagg atataagatt
 [0305] 7381 tgaggtcagc catcttcaaa aaagaaaaaa aaattgactc aagaaagtac aagtaaaacta
 [0306] 7441 tacacctttt tttcataagt tttaggaact gtagtaatgt ggcttagaaa gtataatggc
 [0307] 7501 ctaaatgttt tcaaatgta agttcctgtg gagaagaatt gtttatattg caaacggggg
 [0308] 7561 gactgagggg aacctgtagg tttaaaacag tatgtttgtc agccaactga tttaaaaggc
 [0309] 7621 ctttaactgt ttgggtgtt gtttttttt taagccactc tcccttctct atgaggaaga
 [0310] 7681 attgagaggg gcacctatct ctgtaaaatc cccaaattgg tgttgatgat tttagcttg
 [0311] 7741 aatgttttca tacctgatta aaacttggtt tattctaatt tctgtatcat atcatctgag
 [0312] 7801 gtttacgtgg taactagtct tataacatgt atgtatcttt ttttgttgt tcatctaaag
 [0313] 7861 ctttttaatc caaataaata cagagtttgc aaagtgttt ggattaacca ggaaaaaaa
 [0314] 7921 aaaaaaaaaa

[0316] 서열번호 4: PBRM1 단백질

[0317] 1 mrrlafrgag calvklkkld smgskrrrat spsssvsgdf ddghhsvstp gpsrkrrrls
 [0318] 61 nlptvdpiav chelyntird ykdeqgrllc elfirapkrr nqpdyyevvs qpidlmiqq
 [0319] 121 klkmeeyddv nlltadfqll fnnaksyykp dspeykaack lwdlylrtrn efvqkgeadd
 [0320] 181 edddedgqdn qgtvtegssp aylkeileql leaivvatnp sgrliselfq klpskvqypd
 [0321] 241 yyaiikepid lktiaqriqn gsyksihama kdidllakna ktynepgsqv fkdansikki
 [0322] 301 fymkkaeieh hemaksslrn rtpsnlaaar ltgpshskgs lgeernptsk yyrnkravqg
 [0323] 361 grlsaitmal qygseeseda alaaaryeeg eseaesitsf mdvsnpyfql ydtvrscrnn
 [0324] 421 qgqliaepfy hlskklkypd yyqqikmpis lqirtklkn qeyetldhle cdlnlmfena
 [0325] 481 krynpnsai ykrvklkqv mqakkkelar rddiedgdsm issatsdtgs akrkrnthds
 [0326] 541 emlgrrlss kknirkqrmk ilfnvlear epgsgrrlcd lfmvkpskdd ypdyykiile
 [0327] 601 pmdlkiiehn irndkyagee gmiedmklmf rnarhyneeg sqvyndahil ekllekrke
 [0328] 661 lgplpdddm aspkklksrk sgispkksky mtpmqklne vyeavknytd kgrrrlsaiif
 [0329] 721 lrlpsrselp dyyltikkm dmekirshmm ankyqdidsm vedfvmmfnn actynepesl
 [0330] 781 iykdalvlhk vllttrdle gdedshvpnv tlliqelihn lfvsvmshqd degreysdsl
 [0331] 841 aeipavdpnf pnkppltfdi irknvennry rrldlfqehm fevlerarm nrtdeiyed
 [0332] 901 avelqqffik irdelcknge illspalsyt tkhlhndvek erkeklpkei eedklkreee
 [0333] 961 kreaekseds sgaaglsgh rtysqdcfsk nsmyhvgdyv yvepaealnq phivcierlw
 [0334] 1021 edsagekwly gcwfyrrnet fhlatrkfle kevfksdyyn kvpvskilgk cvvmfvkeyf
 [0335] 1081 klcpnfrde dvfvcersys aktksfkklk lwtmpissvr fvprdvplpv vrvasvfana

[0336] 1141 dkgddekntd nsedsraedn fnlekekedv pvemsngepg chyfeqlhyn dmwlkvgdv

[0337] 1201 fikshglvrp rvgriekvww rdgaayfygp ifihpeeteh eptkmfykke vflsnleetc

[0338] 1261 pmtcilgkca vlsfkdfisc rpteipendi llcesrynes dkqmkkfkgf krfslsakvv

[0339] 1321 ddeiyyfrkp ivpqkepspl lekkiqllea kfaeleggdd dieemgeeds evieppslpq

[0340] 1381 lqtplaseld lmpytpqqst pksakgsakk egskrkinms gyilfssemr avikaqhpdy

[0341] 1441 sfgelsrlvg tewrnletak kaeyegvmnq gvapmvgtpa pggspygqqv gvlpgpgqqa

[0342] 1501 pppypgphpa gppviqqptt pmfvapppkt qrllhseayl kyieglasae nsiskwdqtl

[0343] 1561 aarrdrvhlv keqesrlpsh wlkskahttt madalwrlrd lmlrdtlnir qaynlenv

[0345] 서열번호 5: BAP1 유전자

[0346] 1 ggggtgacgcg actgggcccc ttgtctgtgt gtgggactga ggggccccgg gggcggtggg

[0347] 61 ggctccccgt gggggcagcg gtggggaggg agggcctgga catggcgctg aggggccgcc

[0348] 121 ccgcgggaag atgaataagg gctggctgga gctggagagc gaccagggcc tcttcaccct

[0349] 181 gctcgtggaa gatttcggtg tcaagggggt gcaagtggag gagatctacg accttcagag

[0350] 241 caaatgtcag ggcctgtat atggatttat ctctctgttc aaatggatcg aagagcgccg

[0351] 301 gtcccgcgca aaggtctcta ccttggtgga tgatactcc gtgattgatg atgatatgt

[0352] 361 gaataacatg ttctttgccc accagctgat acccaactct tgtgcaactc atgccttgc

[0353] 421 gagcgtgctc ctgaactgca gcagcgtgga cctgggaccc accctgagtc gcatgaagga

[0354] 481 cttcaccaag ggtttcagcc ctgagagcaa aggatatgag attggcaatg ccccgaggt

[0355] 541 ggccaaggcc cataatagcc atgccaggcc cgagccacgc cactccctg agaagcagaa

[0356] 601 tggccttagt gcagtgcgga ccatggaggc gtccacttt gtcagctatg tgcctatcac

[0357] 661 aggccggctc tttgagctgg atgggctgaa ggtctacccc attgaccatg ggccctgggg

[0358] 721 ggaggacgag gaggggacag acaaggcccc gcgggtcatc atggagcgta tcggcctcgc

[0359] 781 cactgcaggg gagccctacc acgacatccg ctccaacctg atggcagtgg tgcccgaccg

[0360] 841 caggatcaag tatgaggcca ggctgcatgt gctgaagggt aaccgtcaga cagtactaga

[0361] 901 ggctctgcag cagctgataa gagtaacaca gccagagctg attcagaccc acaagtctca

[0362] 961 agagtacacag ctgcctgagg agtccaagtc agccagcaac aagtccccgc tgggtgctgga

[0363] 1021 agcaaacagg gcccctgcag cctctgaggg caaccacaca gatggtgcag aggaggcggc

[0364] 1081 tggttcatgc gcacaagccc catccacag cctcccaac aaaccaagc tagtggtgaa

[0365] 1141 gcctccaggc agcagcctca atggggttca cccaacccc actccattg tccagcggt

[0366] 1201 gccggccttt ctagacaatc acaattatgc caagtcccc atgcaggagg aagaagacct

[0367] 1261 ggccggcagg gtgggccgca gccgagttcc agtccgcca cccagcagt actcagatga

[0368] 1321 tgaggatgac tatgaggatg acgaggagga tgacgtgcag aacaccaact ctgcccttag

[0369] 1381 gtataagggg aagggaacag ggaagccagg ggcattgagc ggttctgctg atgggcaact

[0370] 1441 gtcagtgtg cagcccaaca ccatcaacgt cttggctgag aagctcaaag agtccagaa

[0371] 1501 ggacctctca attcctctgt ccatcaagac tagcagcggg gctgggagtc cggctgtggc

[0372] 1561 agtgtccaca cactcgcagc cctcaccac cccagcaat gagagtacag acacggcctc

[0373] 1621 tgagatcggc agtgccttca actcgccact gcgctcgctt atccgctcag ccaaccgcag

[0374] 1681 gcggccctcc agccctgtca cctcccacat ctccaagggtg ctttttgag aggatgacag

[0375] 1741 cctgtctgctg gttgactgca tacgtacaa ccgtgtctgtc cgtgatctgg gtcctgtcat

[0376] 1801 cagcacaggc ctgctgcacc tggctgagga tgggtgtctg agtcccctgg cgctgacaga

[0377] 1861 ggggtgggaag ggttctctgc cctccatcag accaatccaa ggagccagg ggtccagcag

[0378] 1921 cccagtggag aaggaggtcg tggaagccac ggacagcaga gagaagacgg ggatggtag

[0379] 1981 gcctggcgag cccttgagtg gggagaaata ctacccaag gagctgctgg cactgctgaa

[0380] 2041 gtgtgtggag gctgagattg caaactatga ggcgtgcctc aaggaggagg tagagaagag

[0381] 2101 gaagaagttc aagattgatg accagagaag gaccacaac tacgatgagt tcatctgcac

[0382] 2161 ctttatctcc atgtggctc aggaaggcat gctggccaac ctatggagc agaactctc

[0383] 2221 cgtgcggcgg cgccaagggg tcagcatcgg cggctccac aagcagcga agcctgaccg

[0384] 2281 gcggaacgc tctgcccc tacaaggcaa gcgccagtga ggactgctgg ccctgactct

[0385] 2341 gcagccact cttgccgtgt ggccctcacc agggctcttc cctgccccac tcccccttt

[0386] 2401 cccagtatta ctgaatagtc ccagctggag agtccaggcc ctgggaatgg gaggaaccag

[0387] 2461 gccacattcc ttccatcgtg ccctgaggcc tgacacggca gatcagcccc atagtgtca

[0388] 2521 ggaggcagca tctggagttg gggcacagcg aggtactgca gcttctcca cagccggtg

[0389] 2581 tggagcagca ggacctggcc cttctgcctg ggcagcagaa tatatattt acctatcaga

[0390] 2641 gacatctatt tttctgggt ccaacccaac atgccacat gttgacataa gttctacct

[0391] 2701 gactatgctt tctctctag gagctgtcct ggtgggcca ggtccttga tcatgccacg

[0392] 2761 gtccaacta cagggtccta gctggggcc tgggtgggct cttgggtctg ggccctgctg

[0393] 2821 ctctagcccc agccaccagc ctgtccctgt tgtaaggaag ccaggtcttc tctcttcatt

[0394] 2881 cctcttagga gagtgcacaa ctcagggacc cagcactggg ctgggttggg agtaggggtg

[0395] 2941 cccagtgggg ttgggtgag caggctgctg ggatccatg gcctgagcag agcatgtggg

[0396] 3001 aactgttcag tggcctgtga actgtcttcc ttgttctagc caggctgttc aagactgtc

[0397] 3061 tccatagcaa ggttctaggg ctcttcgctt tcagtgttgt ggccctagct atgggcctaa

[0398] 3121 attgggctct aggtctctgt ccctggcgct tgaggctcag aagagcctct gtccagcccc

[0399] 3181 tcagtattac catgtctccc tctcaggggt agcagagaca gggttgctta taggaagctg

[0400] 3241 gcaccactca gctcttctg ctactccagt ttcctcagcc tctgcaaggc actcagggtg

[0401] 3301 ggggacagca ggatcaagac aaccgttgg agccccgtg ttccagagga cctgatgcca

[0402] 3361 aggggtaatg ggcccagcag tgcctctgga gccagggccc caacacagcc ccatggcctc

[0403] 3421 tgccagatgg ctttgaaaaa ggtgatccaa gcaggcccc ttaictgtac atagtactg

[0404] 3481 agtgggggt gctggcaagt gtggcagct cctctgggt gagcacagct tgaccctct

[0405] 3541 agccctgtg aatactggat caatgaatga ataaaactct cctaagaatc tcctgagaaa

[0407] 서열번호 6: BAP1 단백질

[0408] 1 mnkgwleles dpglftllve dfgvgqvqe eiyclqskcq gpyvgfiflf kwieerrsr
 [0409] 61 kvstlvddts vidddivnm ffahqlipns cathallsvl lncssvdlgp tlrmdkftk
 [0410] 121 gfspekgya ignapelaka hnsharpepr hlpekqngls avrtmeafhf vsyvpitgrl
 [0411] 181 feldgklvyp idhgpgwede ewtdkarrvi meriglatag epyhdirfnl mavvpdrrik
 [0412] 241 yearlhvkv nrtvlealq qlirvtqpel iqthksqesq lpeesksasn ksplvleanr
 [0413] 301 apaasegnht dgaeeagasc aqapshsppn kpklvvkppg sslngvhpnp tpivqrlpaf
 [0414] 361 ldnhnyaksp mqeeldlaag vgrsrpvrp pqqsddedd yeddeeddvq ntnsalrykg
 [0415] 421 kgtgkpgals gsadgqlsvl qpntinvlae klkesqkds iplsiktssg agspavavpt
 [0416] 481 hsqpsptpsn estdtaseig safnsplrsp irsanptrps spvtshiskv lfgeddslr
 [0417] 541 vdcirynrav rdlgpvistg llhlaedgvl splalteggk gsspsirpiq gsqgssspve
 [0418] 601 kevveatdsr ektgmvrpge plsgekyspk ellallkcve aeianyeacl keevekrkkf
 [0419] 661 kiddqrthn ydefictfis mlaqegmlan lveqnisvrr rpgvsigrh kqrkpdrrkr
 [0420] 721 srpykakrq

발명의 효과

[0421] 본 발명의 상기 방법 및 장치를 이용하는 경우 암과 관련성이 있는 유전자 또는 그에 의해 암호화되는 단백질을 독립변수로 하여 분석한 경우에 비하여, 그 정확도 및 AUC(Area under the curve) 값을 현저하게 증가시킬 수 있다. 나아가, 이를 통해 진단 당시에 암의 악성 여부를 예측함으로써, 환자 개개인에 특화되어 있는 치료방법을 효과적으로 제공할 수 있다.

도면의 간단한 설명

[0422] 도 1 및 도 2는 본 발명의 일 실시예에 따른 암의 진단을 위한 정보를 제공하는 방법의 각 단계를 나타낸 것이다.

도 3은 본 발명의 일 실시예에 따른 암의 진단을 위한 정보 제공 장치의 모식도를 나타낸 것이다.

도 4는 본 발명의 일 실시예에 따른 면역조직화학염색 결과를 나타낸 것이다.

도 5는 본 발명의 일 실시예에 따른 유전자 발현 수준 값을 이용한 로지스틱 회귀 분석(Logistic regression analysis) 및 딥 러닝(Deep learning) 분석 방법에 대한 정확도(Accuracy) 및 AUC(Area under the curve) 값을 그래프로 나타낸 것이다.

도 6은 본 발명의 일 실시예에 따른 발현 수준 값을 이용한 로지스틱 회귀 분석 및 딥 러닝 분석 방법에 대한 정확도 및 AUC 값을 그래프로 나타낸 것이다.

발명을 실시하기 위한 구체적인 내용

[0423] 이하, 실시예를 통하여 본 발명을 더욱 상세히 설명하고자 한다. 이들 실시예는 오로지 본 발명을 보다 구체적으로 설명하기 위한 것으로서, 본 발명의 요지에 따라 본 발명의 범위가 이들 실시예에 의해 제한되지 않는다는 것은 당업계에서 통상의 지식을 가진 자에게 있어서 자명할 것이다.

[0425] 실시예

[0427] 하기 실시예는 연세대학교 헬스 시스템(프로젝트 번호: 4-2013-0742)의 기관 검토위원회에서 검토 및 승인되었다. 또한, 인간을 대상으로 한 연구에서 수행된 모든 절차는 헬싱키 선언의 원칙에 따라 수행되었다.

[0429] [준비에 1] 신세포암 환자 시료 및 데이터 수집

[0430] 연세대학교 의료원에서 2008년 1월부터 2014년 12월까지 근치적 신장절제술(Radical nephrectomy) 또는 신부분 절제술(Partial nephrectomy)을 받은 7cm 이하의 종양 크기를 갖는 투명세포형 신세포암(Clear cell renal cell carcinoma; 이하 'ccRCC'라 함) 환자 1,069명의 시료를 수집하였다. 시료는 절제술로만 치료를 받았으며, 종양 조직이 FFPE(formalin-fixed paraffin-embedded)으로 사용 가능한 경우에만 선별될 수 있도록 하였다. 재발 위험에 대한 영향, ccRCC의 대표성 저해 등을 배제하기 위해, 수술 전 보조요법(Neoadjuvant) 또는 보조제(Adjuvant)를 사용한 경우의 환자를 제외하였고, 폰힙펠 린도우병(Von Hippel-Lindau disease)이나 동시 또는 이원성의 양측성 신장암의 병력이 있는 환자를 제외하는 조건으로 악성 ccRCC 환자 시료를 선별하였다.

[0431] 나아가, 데이터 수집을 위해 영상이 존재하지 않거나, 수술 후 6개월 이내에 재발한 경우와, 단백질 또는 RNA 시료가 불충분하거나 실험에 적합하지 않은 경우의 환자를 제외하였다. 선별된 환자 시료의 데이터로부터 연령, 성별, 신장, 체중, 체질량 지수, 종양 크기 및 전이 유무에 대한 임상적 특징을 수집하였다. 악성 암의 경우, 동시 전이, 재발 또는 암 특이적 사망 증가를 나타내는 종양으로 정의하였으며, 동시 전이의 경우에는 일차성 비뇨기암 진단 시 또는 그 진단으로부터 3개월 이내에 전이된 것으로 정의하였다. 이렇게 선별된 초기 단계 ccRCC 177명 환자의 정보는 하기 표 1과 같다.

[0432] 또한, 연세대학교 의과대학 병리학교실(서울)로부터 ccRCC 환자의 FFPE(Formalin-fixed paraffin-embedded) 섹션을 얻었다. 비뇨기과 병리학자가 상기 FFPE 섹션에 해당하는 환자의 조직을 검토 및 분리하였다. 또한, 헤마톡실린(Haematoxylin) 및 에오신(eosin)에 의해 염색된 조직을 현미경을 이용하여 비뇨기과 병리학자가 비 종양 부분(Non-tumor elements)을 확인하였다.

표 1

[0434]

구분		값
성별	남성	127 (71.8%)
	여성	50 (28.2%)
나이		58.5 ± 11.7
BMI(kg/m ²)		24.8 ± 3.6
근치적 수술, 명(%)		94 (53.1%)
종양 크기(cm)		4.1 ± 1.6
펼먼 등급 (Fuhrman grade)	1	8 (4.5%)
	2	79 (44.6%)
	3	81 (45.8%)
	4	9 (5.1%)
침습, 명(%)		11 (6.2%)
양성 임파절(Positive nodal status), 명 (%)		1 (1.0%)
동시 전이, 명(%)		19 (10.7%)
재발, 명(%)		23 (13.0%)
암 특이적 사망, 명(%)		30 (16.9%)
상기 값은 평균±표준편차(SD) 또는, 피시험자의 수로 표시되는 것임		
침습: (신장 주위(perinephric)/부비강 지방(sinus fat)/혈관(vascular))		
BMI: Body mass index		

[0435]

[0436] [실시에 1] 악성 ccRCC에서 유전자 발현 수준 변화 분석

[0437] 상기 준비예의 FFPD 섹션으로부터 추출된 시료에, 트리졸® 시약(Ambion, Life technologies)를 처리하여 전체 RNA(Total RNA)를 얻었다. 그런 다음, iNtRon Maxime RT PreMix(Intronbio. Cat No. 25081)를 사용하여 제조사가 제공하는 프로토콜에 따라, 1μg의 전체 RNA를 cDNA로 합성하였다.

[0438] 10μl 부피의 튜브에 5μl의 SYBR® Green master PCR mix(Thermo Fisher, Cat No. A25742, USA), 10 pmol에 해당하는 1μl의 센스 및 안티센스 프라이머(표 3)와 상기 cDNA 1μl를 넣고, ABI StepOnePlus RT-PCR 시스템(Applied Biosystems)을 사용하여 qRT-PCR(Quantitative Real-Time polymerase chain reaction)을

수행하였다. 여기서, 95℃에서 10분간 초기 변성, 95℃에서 15초간 변성, 58℃에서 60초간 어닐링, 72℃에서 60초간 연장되도록 하는 조건을 45 사이클 반복하였고, 최종 신장을 72℃에서 5분간 수행하였다. 2-ΔΔCT를 이용하여 GAPDH의 유전자 발현량을 통해 각각의 프라이머에 대한 유전자 발현을 상대적으로 정량화 하였다.

표 2

[0440]

유전자	서열번호	구분	서열
FOXC2	서열번호 7	센스	5'- GAT CAC CTT GAA CGG CAT CT - 3'
	서열번호 8	안티센스	5'- ACC TTG ACG AAG CAC TCG TT - 3'
CLIP4	서열번호 9	센스	5'- GCA TCA TGC CAG GAA ATT CT - 3'
	서열번호 10	안티센스	5'- TTT GTT GGA CCT GAG GAA CC - 3'
PBRM1	서열번호 11	센스	5'- TGA TGG CCA ACA AGT ACC AA - 3'
	서열번호 12	안티센스	5'- AGA TCA AAG ACT CCG GCT CA - 3'
BAP1	서열번호 13	센스	5'- GCC TGA GGA GTC CAA GTC AG - 3'
	서열번호 14	안티센스	5'- CTG GAG GCT TCA CCA CTA GC - 3'
SETD2	서열번호 15	센스	5'- TCA CAA GGC AGA CTC AGT GG - 3'
	서열번호 16	안티센스	5'- CTG CTG TCT TGG GCT TTT TC - 3'
KDM5C	서열번호 17	센스	5'- GTC ATT TGC AAC CCC TGA GT - 3'
	서열번호 18	안티센스	5'- AAT GGG ATG AGG GGT AAA GG - 3'
GAPDH	서열번호 19	센스	5'- CAG CCT CAA GAT CAT CAG CA - 3'
	서열번호 20	안티센스	5'- GGT GCT AAG CAG TTG GTG GT - 3'

각 유전자에 대한 발현 수준 변화 값을 이용하여 단변량 분석 및 로지스틱 회귀분석을 통해 통계 분석을 수행하였다. 단변량 분석에서 연속 변수를 비교하는데, t-검정(t-test)을 사용하였다. 상기 단변량 분석 결과에서 임상 결과와 유의한 상관관계가 있을 것으로 확인된 위험 요소를 모두 포함하도록 하여 로지스틱 회귀 분석(Logistic regression analysis)을 수행하여, 그 결과를 하기 표 3 및 4에 나타내었다. 여기서, 분석 결과 값은 연속 변수에 대한 평균 ± 표준편차(SD) 및 범주형 변수에 대한 백분율로 표시될 수 있도록 하였다.

표 3

[0444]

위험 요소		악성 ccRCC(40명)	악성이 아닌 ccRCC(137명)	P value	
성별	남	26 (65.0%)	101 (73.7%)	0.281	
	여	14 (35.0%)	36 (26.3%)		
나이		58.0 ± 11.3	58.7 ± 11.8	0.755	
BMI(kg/m ²)		25.2 ± 4.2	24.8 ± 3.4	0.520	
근치적 수술, 명(%)		29 (72.5%)	65 (47.4%)	0.005	0.023
종양 크기(cm)		4.4 ± 1.5	4.1 ± 1.7	0.241	
폐문 등급	1	1 (2.5%)	7 (5.1%)	0.023	
	2	12 (30.0%)	67 (48.9%)		
	3	22 (55.0%)	59 (43.1%)		
	4	5 (12.5%)	4 (2.9%)		
	3-4, vs 1-2	27 (67.5%)	63 (46.0%)	0.017	0.049
침습, 명(%)		1 (2.5%)	10 (7.3%)	0.269	
양성 임파절		1 (2.5%)	0 (0.0%)	0.063	

상기 표 3에서 보는 바와 같이, 위험 요소 중 성별, 나이, BMI, 종양의 크기, 양성 임파절의 경우에는 악성 ccRCC와 악성이 아닌 ccRCC 간의 유의미한 차이가 존재하지 않았다. 그러나, 단변량 분석 및 다변량 분석에서 근치적 수술의 경우, 악성 ccRCC와 악성이 아닌 ccRCC 간의 유의미한 차이점이 존재하였다. 또한, 폐문 등급(3

내지 4)의 경우에도 악성이 아닌 경우에 비하여, 악성 ccRCC에서 유의미한 차이점이 존재하는 것을 확인하였다.

표 4

위험 요소	유전자 발현 수준 변화		P value	
	악성 ccRCC(40명)	악성이 아닌 ccRCC(137명)	단변량	다변량
<i>FOXC2</i>	0.0033 ± 0.0057	0.0074 ± 0.0169	0.018	0.034
<i>CLIP4</i>	0.0042 ± 0.0055	0.0088 ± 0.0169	0.008	0.499
<i>PBRM1</i>	0.0008 ± 0.0014	0.0051 ± 0.0124	<0.001	0.036
<i>SETD2</i>	0.0003 ± 0.0003	0.0004 ± 0.0007	0.092	
<i>BAP1</i>	0.0001 ± 0.0001	0.0002 ± 0.0005	0.004	0.015
<i>KDM5C</i>	0.0016 ± 0.0015	0.0030 ± 0.0094	0.350	

상기 표 4에서 보는 바와 같이, 단변량 분석을 통해 악성이 아닌 ccRCC에 비하여, *FOXC2*, *CLIP4*, *PBRM1* 및 *BAP1* 유전자 각각의 경우, 악성 ccRCC에서 낮은 수준으로 발현되는 것을 확인하였다. 반면, 다변량 분석에서 *FOXC2*, *PBRM1* 및 *BAP1*만이 독립적으로 악성 ccRCC와 유의미한 상관관계(P value<0.05)가 있음으로 확인되었다.

본 발명의 상기 결과를 통해, 상기 유전자들의 발현 수준을 이용하는 경우 악성 ccRCC를 예측할 수 있음을 알 수 있다. 특히, *FOXC2* 유전자의 경우 세포 사이클이 잘못되는 경우를 체크하는 역할을 하는 것이므로, 이와 같은 유전자의 발현이 감소되었을 때, ccRCC가 악성 예후를 나타내도록 하는 것임을 예측할 수 있다.

[실시예 2] 악성 ccRCC에서 단백질 발현 수준 확인

상기 준비예 1의 FFPE 시료를 4μm 두께로 절단하고, 절단된 시료를 Superfrost Plus 현미경 슬라이드 (Thermo Fisher Scientific)에 올려놓은 뒤, EZ Prep (Ventana)을 이용하여 75℃에서 8분 동안 탈파라핀화 하였다. 그런 다음, 상기 슬라이드를 세포 컨티셔닝 용액(높은 pH CC1 standard)에 넣고, 100℃에서 60 분간 반응시켜 항원 회수를 수행하였다. 이후, 상기 슬라이드를 *FOXC2*, *BAF1* 및 *BAP1*에 특이적인 항체와 함께 37℃에서 32분 동안 배양하고, 반응 완충액(pH 7.6)으로 세척한 뒤에 2차 항체(유니버설 HRP 멀티머)를 넣고 37℃에서 8분 동안 추가로 배양하였다.

배양이 완료된 상기 슬라이드에 기질(DAB 및 H₂O₂)을 넣고 8분동안 상온에서 배양한 뒤, 카운터 염색 시약인 헤마톡실린(Hematoxylin) II 및 블루잉 용액(Bluing reagent)으로 염색하고, 추가적으로 세척하였다. 광학 현광현미경을 이용하여 세척이 완료된 슬라이드를 100배 내지 400배의 배율로 염색 강도를 측정하여, 그 결과를 도 1에 나타내었다.

도 4에서 보는 바와 같이, 악성이 아닌 ccRCC 조직에 비하여, 악성 ccRCC 조직에서 *FOXC2*, *PBRM1* 및 *BAP1* 단백질의 발현 수준이 현저하게 감소되어 있는 것을 확인하였다.

[실시예 3] *FOXC2*, *PBRM1* 및 *BAP1*의 발현을 이용한 로지스틱 회귀 및 딥러닝 모델 설정

[3-1] 로지스틱 회귀 분석 모델 설정

FOXC2, *PBRM1* 및 *BAP1*의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준에 대한 조합에 따른 로지스틱 회귀분석 모델을 테스트하기 위하여, 하기 식 1 내지 4를 도출하였다. 여기서, 위험 점수는 악성 ccRCC에 해당하는지 여부를 확인할 수 있는 것으로 정의하였고, *PBRM1*_norm은 (*PBRM1*의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준)/0.06, *BAP1*_norm은 (*BAP1*의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준)/0.005, *FOXC2*_norm은 (*FOXC2*의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준)/0.096에 해당하도록 하였다.

- [0462] [식 1]
- [0463] 위험 점수 = $1/[1 + \exp(-(-1.097 - 0.963 \times (\text{PBRM1_norm}) - 0.524 \times (\text{BAP1_norm})))]$
- [0465] [식 2]
- [0466] 위험 점수 = $1/[1 + \exp(-(-1.100 - 0.894 \times (\text{PBRM1_norm}) - 0.340 \times (\text{FOXC2_norm})))]$
- [0468] [식 3]
- [0469] 위험 점수 = $1/[1 + \exp(-(-1.114 - 0.533 \times (\text{BAP1_norm}) - 0.562 \times (\text{FOXC2_norm})))]$
- [0471] [식 4]
- [0472] 위험 점수 = $1 / [1 + \exp(-(-1.082 - 0.336 \times (\text{FOXC2_norm}) - 0.887 \times (\text{PBRM1_norm}) - 0.521 \times (\text{BAP1_norm})))]$
- [0474] 상기 식 1의 AUC(Area under the curve)는 0.714, 정확도(Accuracy)는 0.537이었고, 컷오프 값은 0.247으로 확인되었고, 상기 식 2의 AUC는 0.532, 정확도는 0.463이었고, 컷오프 값은 0.248로 확인되었으며, 상기 식 3의 AUC는 0.597, 정확도는 0.519이었고, 컷오프 값은 0.244로 확인되었다. 또한, 상기 식 4의 AUC는 0.555, 정확도는 0.651이었고, 컷오프 값은 0.249195로 확인되었다.
- [0476] [3-2] 딥 러닝 모델 설정
- [0477] 상기 [3-1]에서 사용된 *FOXC2*, *PBRM1* 및 *BAP1*의 정규화된 값들을 각각 아래의 식 5를 이용하여 0 내지 1의 변수 범위에 해당하는 일반화 수치(Normalized value; z)로 변환하였다.
- [0479] [식 5]
- [0480] $\{Z = [x - \min(x)]/[\max(x) - \min(x)]\}$
- [0482] 상기 [식 5]에서, x는 유전자 또는 이에 의해 암호화되는 단백질이 존재하는 수준의 값을 의미한다.
- [0483] 177명의 ccRCC 환자 데이터 세트를 무작위로 두개의 독립적인 교육 그룹 및 검증 그룹으로 나누어 내부 검증 테스트를 수행하였다. 상기 교육 그룹은 전체 데이터 세트의 약 70%(악성 ccRCC 28명 및 악성이 아닌 ccRCC 95명)로 구성될 수 있도록 하여, 분석 모델을 도출하는데 사용하였다. 상기 검증 그룹은 상기 교육 그룹의 나머지 약 30%(악성 ccRCC 12명 및 악성이 아닌 ccRCC 42명)로 구성될 수 있도록 하여, 도출된 분석 모델의 성능을 평가하는데 사용하였다.
- [0484] TensorFlow (버전 1.8.0) 백엔드가있는 Python 라이브러리 Keras (버전 2.2.0)를 사용하는 DNN(Deep neural network)를 선택하여 상기 분석 모델이 심층 학습될 수 있도록 하였다. 데이터 관리 및 전처리를 위해, scikit-learn 라이브러리 (<http://scikitlearn.org/>)를 사용하였고, 상기 분석 모델에서 발생하는 오버피팅(Overfitting) 문제를 해소하기 위해, 30% 드롭아웃 속도의 2개의 층(Layer) DNN 네트워크를 사용하였다. 상기 2층 DNN 네트워크에서, 첫번째 층을 시그모이드(Sigmoid)로, 두번째 층을 정류된 선형 유닛(Rectified linear unit)으로 설정 하였다.
- [0485] 상기 분석 모델은 2진 교차 엔트로피 손실 함수가 있는 Adam 최적화 알고리즘을 사용하여 최적화 될 수 있도록 하였으며, 500 에포치(Epoch) 및 30 배치 크기를 선택하였다.

[0487] [3-3] 로지스틱 회귀 모델과 딥 러닝 모델을 이용한 분석 비교

[0488] 상기 [3-1] 및 [3-2]에서 사용된 악성 ccRCC를 예측하기 위해 설정된 로지스틱 회귀 모델과 딥 러닝 모델의 성능을 비교분석 하여, 그 결과를 하기 표 5와, 도 5 및 6에 나타내었다.

표 5

	로지스틱 회귀 모델	딥 러닝 모델
정확도	0.555	0.537
AUC	0.651	0.736

[0491] 상기 표 5와, 도 5 및 6에서 보는 바와 같이, *FOXC2*, *PBRM1* 및 *BAP1*의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준에 대한 조합에 따른 로지스틱 회귀분석 모델과, 딥 러닝 모델 모두에서 정확도 및 AUC 값이 높은 것을 확인하였다.

[0492] 상기 결과를 통해 본 발명에 따른 *FOXC2*, *PBRM1* 및 *BAP1*의 유전자 또는 이에 의해 암호화되는 단백질의 발현 수준에 대한 조합을 위험 요소로 포함하는 로지스틱 회귀분석 모델뿐만 아니라, 딥 러닝 모델 모두를 이용하는 경우, 악성 ccRCC를 매우 효과적으로 예측할 수 있음을 알 수 있다.

[0493]

[0494] 이상으로 본 발명의 특정한 부분을 상세히 기술하였는 바, 당업계의 통상의 지식을 가진 자에게 있어서 이러한 구체적인 기술은 단지 바람직한 구현 예일 뿐이며, 이에 본 발명의 범위가 제한되는 것이 아닌 점은 명백하다. 따라서, 본 발명의 실질적인 범위는 첨부된 청구항과 그의 등가물에 의하여 정의된다고 할 것이다.

부호의 설명

[0495] 110: 프로세서

111: 산출부

120: 메모리

121: ROM

122: RAM

130: 입력부

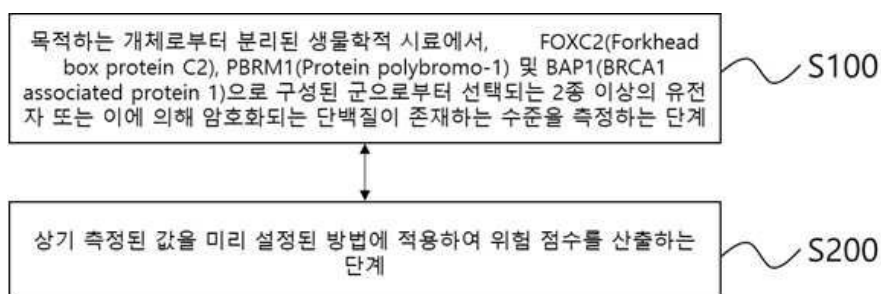
140: 출력부

150: 스토리지

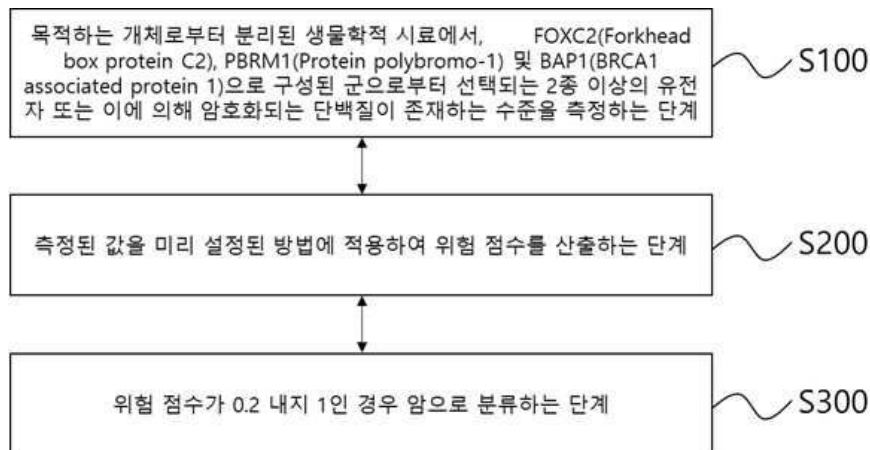
160: 네트워크인터페이스

도면

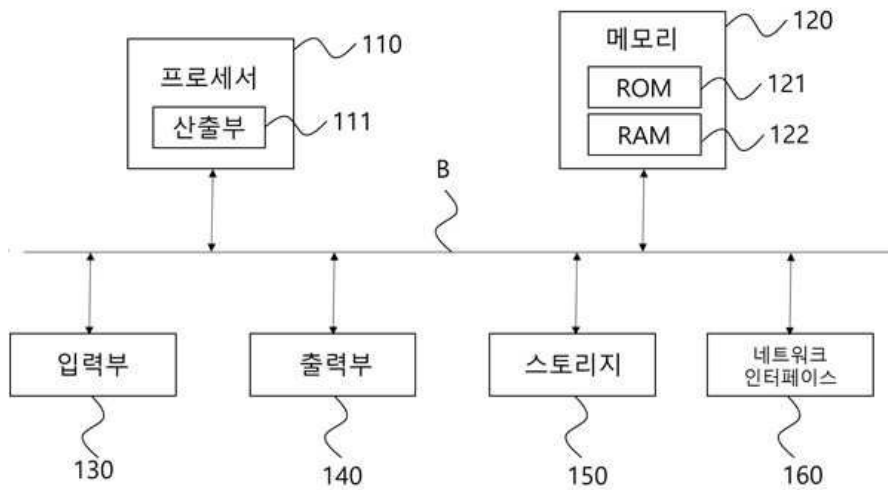
도면1



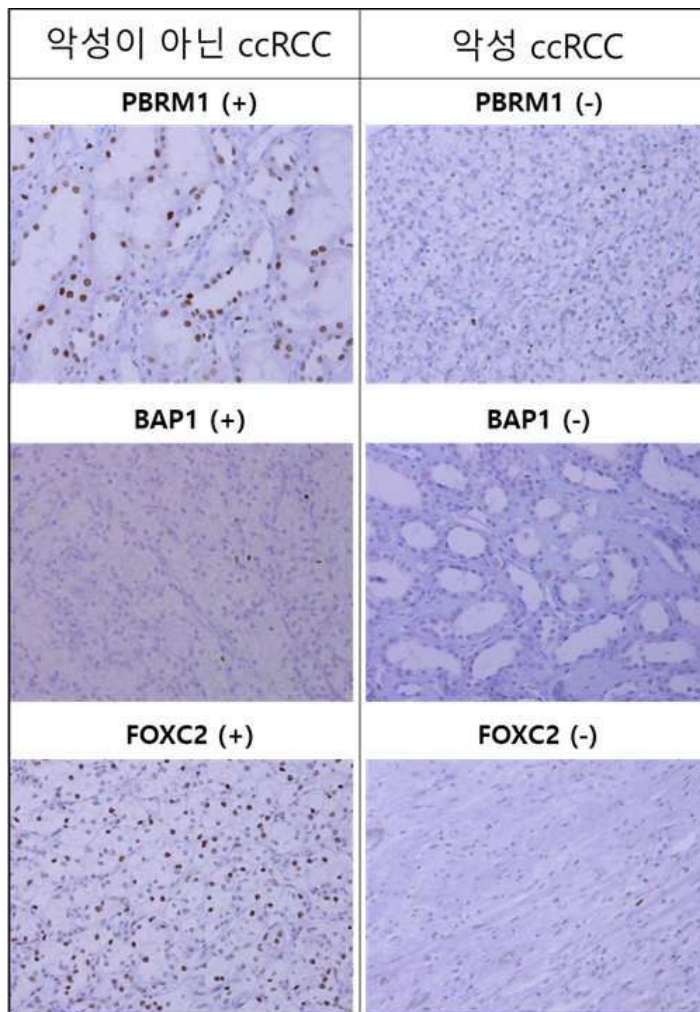
도면2



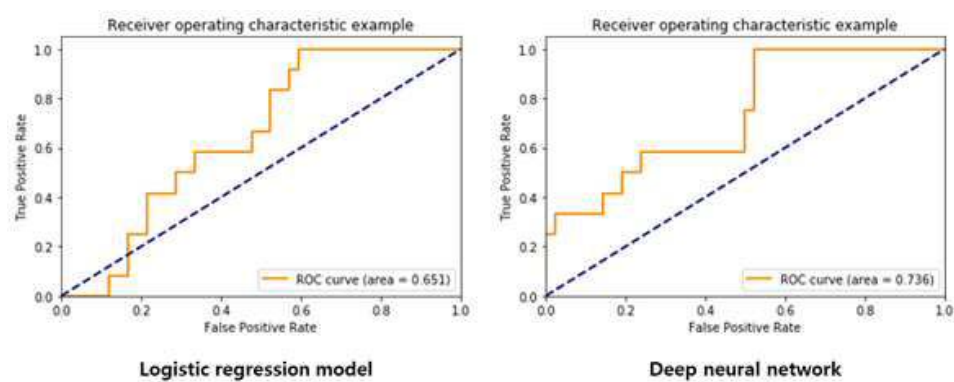
도면3



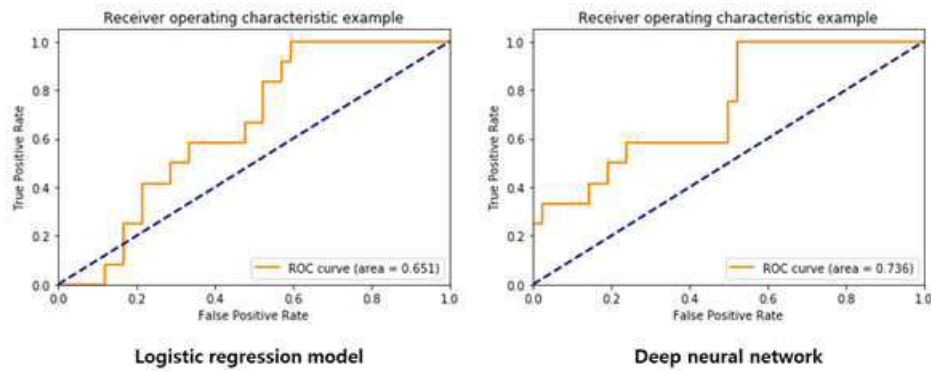
도면4



도면5



도면6



서열 목록

- <110> Industry-Academic Cooperation Foundation, Yonsei University
- <120> A method of providing information for diagnosing cancer
- <130> PDPB187344
- <160> 20
- <170> KoPatent In 3.0
- <210> 1
- <211> 2900
- <212> DNA
- <213> Homo sapiens
- <400> 1

```

gaaacttttc ccaateccta aaagggaactt ggctctttt tctgggtca gcggggcagc      60
cgctcggacc ccggcgcgct gacctcggg gctgccgatt cgctgggggc ttggagagcc      120
tcctgcgcc ctctcgcgc gggccgaggg tccacctggg tcccaggcc gcggcgtctc      180

cgctgggtcc gcggccgccc gcctccccgc gctgccccg ccgggtcctg gagccagcga      240
ggagcggggc cggcgtcgc cttccccggg gcgcgccctc caggatgccg atccgcccg      300
tccgtgaaa gcgcgcgcc ctgctcgcc cgagcgccgc gcccgcgca cctcgcgcc      360
ggaggctgcc aggagcccg ggccggccct ccgctcccc tctctcccc ctctggctct      420
ctcgcgtct ctctctca gggccccct cgtcccccg gccgcagtcc gtgcgcgagg      480
gcgcggcgca gccgtctcg aagcagcatg caggcgcgct actcgtgtc cgacccaac      540
gccctgggag tgggtgcccta cctgagcgag cagaattact accgggctgc ggccagctac      600

ggcggcatgg ccagcccat gggcgtctat tccggccacc cggagcagta cagcggggg      660
atgggccgct cctacgcgcc ctaccaccac caccagcccg cggcgccctaa ggacctggtg      720
aagccgccct acagctacat cgcgtctatc accatggcca tccagaacgc gcccgagaag      780

```

aagatcacct tgaacggcat ctaccagttc atcatggacc gcttccccctt ctaccgggag	840
aacaagcagg gctggcagaa cagcatccgc cacaacctct cgctcaacga gtgcttcgtc	900
aaggtgcccc gcgacgacaa gaagcccggc aagggcagtt actggaccct ggacccggac	960
tcctacaaca tgttcgagaa cggcagcttc ctgcggcgcc ggcgggcgctt caaaaagaag	1020
gacgtgtcca aggagaagga ggagcgggcc cacctcaagg agccgcccc ggcggcgtcc	1080
aagggcgccc cggccacccc ccacctagcg gacgccccca aggaggccga gaagaaggtg	1140
gtgatcaaga gcgaggcggc gtccccggcg ctgccgttca tcaccaaggt ggagacgtg	1200
agccccgaga gcgcgttga gggcagcccc cgagcgcgcg cctccacgcc cgccggctcc	1260
cccgacggct cgctgccgga gcaccacgcc gcggcgccca acgggctgcc tggcttcagc	1320
gtggagaaca tcatgacct gcgaacgtcg ccgccggcg gagagctgag cccggggggc	1380
ggacgcgcgg gcctggtggt gccgccgtg gcgtgcctt acgccgccgc gccgccgcc	1440
gcctacggcc agccgtgcgc tcaggccctg gaggccgggg ccgccggggg ctaccagtgc	1500
agcatgcgag cgatgagcct gtacaccggg gccgagcggc cggcgacat gtgcgtccc	1560
cccgccctgg acgaggccct ctgggaccac ccgagcggcc ccacgtcgcc cctgagcgt	1620
ctcaacctcg ccgccggcca ggaggcgcg ctgccgccca cgggccacca ccaccagcac	1680
cacggccacc accaccgca ggccgcccg ccccgccgg ctccccagcc ccagccgacg	1740
ccgcagccc gggccgcgc ggccgagcg gcctcctggt atctcaacca cagcggggac	1800
ctgaaccacc tccccggcca cagttcgcg gccagcagc aaactttccc caacgtgcgg	1860
gagatgttca actcccacc gctggggatt gagaactcga ccctcgggga gtcccaggtg	1920
agtggcaatg ccagctgcca gctgccctac agatccacgc cgctctctta tcgccacga	1980
gccccctact cctacactg cagaaatac tgacgtgtcc cgggacctcc cctccccggc	2040
ccgtccggc ttgcgttccc agccccgacc caaccagaca attaaggggc tgcagagacg	2100
caaaaaagaa aaaaaacatg tccaccaacc ttttctcaga cccgggagca gagagcgggc	2160
acgctagccc ccagccgtct gtgaagagcg caggtaactt taattcgccg ccccgtttct	2220
gggatcccag gaaaccctc caaagggacg cagcccaaca aaatgagtat tgatcttaa	2280
atccccctcc cctaccagga cggctgtgct gtgctcgacc tgagctttca aaagttaagt	2340
tatggaccaa atcccatagc gagccctag tgactttctg taggggtccc cataggtgta	2400
tgggggtctc tatagataat atatgtgctg tgtgtaattt taaattttct caaccgtgct	2460
gtacaaatgt gtggatttgt aatcaggcta tttgttgtt gttgttgtt ttcagagcca	2520
ttaatataat atttaaagtt gagtccactg gataagtttt tcattctgcc caaccatttc	2580
taactgccaa attgaattca agaaaccgat gtgggttttg tttcctgtac aattatgaga	2640

tataattctt ttcccatg taggtctttt acaaaacaag aaaataattt attttttgt 2700

tggtggataa agaagtcaag tatctgatac tttttattta caaagtgtga tggttttgta 2760

tagtaggttc caccctgagt attcctaaaa gaaaaaaaaa aaaaaagctt aaaaactcta 2820

acttcactcg tgtttgtctt acgtggcttt aatcgttgta cttaccttaa aataaaccca 2880

tgttgttttt tctgcccaaa 2900

<210> 2

<211> 501

<212> PRT

<213> Homo sapiens

<400> 2

Met Gln Ala Arg Tyr Ser Val Ser Asp Pro Asn Ala Leu Gly Val Val

1 5 10 15

Pro Tyr Leu Ser Glu Gln Asn Tyr Tyr Arg Ala Ala Gly Ser Tyr Gly

20 25 30

Gly Met Ala Ser Pro Met Gly Val Tyr Ser Gly His Pro Glu Gln Tyr

35 40 45

Ser Ala Gly Met Gly Arg Ser Tyr Ala Pro Tyr His His His Gln Pro

50 55 60

Ala Ala Pro Lys Asp Leu Val Lys Pro Pro Tyr Ser Tyr Ile Ala Leu

65 70 75 80

Ile Thr Met Ala Ile Gln Asn Ala Pro Glu Lys Lys Ile Thr Leu Asn

85 90 95

Gly Ile Tyr Gln Phe Ile Met Asp Arg Phe Pro Phe Tyr Arg Glu Asn

100 105 110

Lys Gln Gly Trp Gln Asn Ser Ile Arg His Asn Leu Ser Leu Asn Glu

115 120 125

Cys Phe Val Lys Val Pro Arg Asp Asp Lys Lys Pro Gly Lys Gly Ser

130 135 140

Tyr Trp Thr Leu Asp Pro Asp Ser Tyr Asn Met Phe Glu Asn Gly Ser

145 150 155 160

Phe Leu Arg Arg Arg Arg Phe Lys Lys Lys Asp Val Ser Lys Glu

165 170 175
 Lys Glu Glu Arg Ala His Leu Lys Glu Pro Pro Pro Ala Ala Ser Lys
 180 185 190
 Gly Ala Pro Ala Thr Pro His Leu Ala Asp Ala Pro Lys Glu Ala Glu
 195 200 205
 Lys Lys Val Val Ile Lys Ser Glu Ala Ala Ser Pro Ala Leu Pro Val
 210 215 220
 Ile Thr Lys Val Glu Thr Leu Ser Pro Glu Ser Ala Leu Gln Gly Ser

 225 230 235 240
 Pro Arg Ser Ala Ala Ser Thr Pro Ala Gly Ser Pro Asp Gly Ser Leu
 245 250 255
 Pro Glu His His Ala Ala Ala Pro Asn Gly Leu Pro Gly Phe Ser Val
 260 265 270
 Glu Asn Ile Met Thr Leu Arg Thr Ser Pro Pro Gly Gly Glu Leu Ser
 275 280 285
 Pro Gly Ala Gly Arg Ala Gly Leu Val Val Pro Pro Leu Ala Leu Pro
 290 295 300

 Tyr Ala Ala Ala Pro Pro Ala Ala Tyr Gly Gln Pro Cys Ala Gln Gly
 305 310 315 320
 Leu Glu Ala Gly Ala Ala Gly Gly Tyr Gln Cys Ser Met Arg Ala Met
 325 330 335
 Ser Leu Tyr Thr Gly Ala Glu Arg Pro Ala His Met Cys Val Pro Pro
 340 345 350
 Ala Leu Asp Glu Ala Leu Ser Asp His Pro Ser Gly Pro Thr Ser Pro
 355 360 365
 Leu Ser Ala Leu Asn Leu Ala Ala Gly Gln Glu Gly Ala Leu Ala Ala

 370 375 380
 Thr Gly His His His Gln His His Gly His His His Pro Gln Ala Pro
 385 390 395 400
 Pro Pro Pro Pro Ala Pro Gln Pro Gln Pro Thr Pro Gln Pro Gly Ala
 405 410 415
 Ala Ala Ala Gln Ala Ala Ser Trp Tyr Leu Asn His Ser Gly Asp Leu

420	425	430	
Asn His Leu Pro Gly His Thr Phe Ala Ala Gln Gln Gln Thr Phe Pro			
435	440	445	
Asn Val Arg Glu Met Phe Asn Ser His Arg Leu Gly Ile Glu Asn Ser			
450	455	460	
Thr Leu Gly Glu Ser Gln Val Ser Gly Asn Ala Ser Cys Gln Leu Pro			
465	470	475	480
Tyr Arg Ser Thr Pro Pro Leu Tyr Arg His Ala Ala Pro Tyr Ser Tyr			
485	490	495	
Asp Cys Thr Lys Tyr			
500			
<210>	3		
<211>	7930		
<212>	DNA		
<213>	Homo sapiens		
<400>	3		
gcacctgccg gcagggacaa aggcgcacta cgggtcggaa actccccgcg ctctcttttc		60	
ccgtctgtcg cggggcagct tcaaagctgt caacgtttcc ctccagtcgcc aataaccagtg		120	
acagcttgcc aacacttggg gtcacatgtg agcctccac atgtattcac tctccattcc		180	
agctctgtga ttgaactctg ctcttattga ctagggggca gttgggcagg catgcctcat		240	
tccttgaatt gacagtcatt cctaattttt gaataatac aactgaatc agcacttcta		300	
aagcccttcc ttctccact ggcttcactt ggctttcaga cataatgagg agactggctt		360	
ttcgaggcgc tgggtgtgct ctggtaaagc tgaagaagtt ggattccatg ggttccaaga		420	
gaagaagagc tacctccct tccagcagtg tcagcgggga ctttgatgat gggcaccatt		480	
ctgtgtcaac accaggccca agcaggaaaa ggaggagact ttccaatctt ccaactgtag		540	
atcctattgc cgtgtgcat gaactctata ataccatccg agactataag gatgaacagg		600	
gcagacttct ctgtgagctc ttcataggg caccaaagcg aagaaatcaa ccagactatt		660	
atgaagtggg ttctcagccc attgacttga tgaaaatcca acagaaacta aaaatggaag		720	
agtatgatga tgtaatttg ctgactgtg acttcagct tcttttaac aatgcaaagt		780	
cctattataa gccagattct cctgaatata aagccgttg caaactctgg gatttgtacc		840	
ttcgaacaag aaatgagttt gttcagaaag gagaagcaga tgacgaagat gatgatgaag		900	

atgggcaaga caatcagggc acagtgactg aaggatcttc tccagcttac ttgaaggaga	960
tcctggagca gcttcttgaa gccatagttg tagctacaaa tccatcagga cgtctcatta	1020
gcgaactttt tcagaaactg ccttctaaag tgcaatatcc agattattat gcaataatta	1080
aggagcctat agatctcaag accattgccc agaggataca gaatggaagc tacaaaagta	1140
ttcatgcaat ggccaaagat atagatctcc tcgcaaaaaa tgccaaaact tataatgagc	1200
ctggctctca agtattcaag gatgcaaatt caattaaaaa aatatattat atgaaaaagg	1260
ctgaaattga acatcatgaa atggctaagt caagtcttcg aatgaggact ccatccaact	1320
tggtctcagc cagactgaca ggtccttcac acagtaaagg cagccttggt gaagagagaa	1380
atcccactag caagtattac cgtaataaaa gagcagtaca aggaggctcg ttatcagcaa	1440
ttacaatggc acttcaatat ggctcagaaa gtgaagaaga tgctgcttta gctgctgcac	1500
gctatgaaga gggagagtc gaagcagaaa gcatcacttc ctttatggat gtttcaaac	1560
ctttttatca gctttatgac acagttagga gttgtcggaa taaccaaggg cagctaatag	1620
ctgaaccttt ttaccatttg ctttcaaaga aaaaatacc tgattattac cagcaaatta	1680
aatgcccac atcactacaa cagatccgaa caaaactgaa gaatcaagaa tatgaaactt	1740
tagatcattt ggagtgtgat ctgaatttaa tgtttgaaaa tgccaaacgc tataatgtgc	1800
ccaattcagc catctacaag cgagttctaa aattgcagca agttatgcag gcaaagaaga	1860
aagagcttgc caggagagac gatatcgagg acggagacag catgatctct tcagccacct	1920
ctgatactgg tagtgccaaa agaaaaagga acactcatga cagtgagatg ttgggtctca	1980
ggaggtatc cagtaaaaag aacataagaa agcagcgaat gaaaatctta ttcaatgttg	2040
ttcttgaagc tcgagagcca ggttcaggca gaagactttg tgacctattt atggttaaac	2100
catcaaaaaa ggactatcct gattattata aaatcatctt ggagccaatg gacttgaaaa	2160
taattgagca taacatccgc aatgacaaat atgctggtga agagggaatg atagaagaca	2220
tgaagctgat gticcggat gccaggcact ataatgagga gggctcccag gtttataatg	2280
atgcacatat cctggagaag ttactcaagg agaaaaggaa agagctgggc cactgcctg	2340
atgatgatga catggcttct cccaaactca agctgagtag gaagagtggc atttctccta	2400
aaaaatcaaa atacatgact ccaatgcagc agaaactaaa tgaggtctat gaagctgtaa	2460
agaactatac tgataagagg ggtcgccgcc tcagtgccat atttctgagg ctccctcta	2520
gatctgagtt gcctgactac tatctgacta ttaaaaagcc catggacatg gaaaaaatc	2580
gaagtcacat gatggccaac aagtaccaag atattgactc tatggttgag gactttgtca	2640
tgatgtttaa taatgcctgt acatacaatg agccggagtc tttgatctac aaagatgctc	2700
ttgttctaca caaagtctg cttgaaacac gcagagacct ggagggatg gaggactctc	2760

atgtcccaaa tgtgactttg ctgattcaag agcttatcca caatcttttt gtgtcagtca	2820
tgagtcacatca ggatgatgag ggaagatgct acagcgattc tttagcagaa attcctgctg	2880
tggatcccaa ctttctaac aaaccacccc ttacatttga cataattagg aagaatgttg	2940
aaaataatcg ctaccgtcgg ctgtatttat ttcaagagca tatgtttgaa gtattggaac	3000
gagcaagaag gatgaatcgg acagattcag aaatatatga agatgcagta gaacttcagc	3060
agttttttat taaaattcgt gatgaactct gcaaaaatgg agagattctt ctttcaccgg	3120
cactcagcta taccacaaaa catttgcata atgatgtgga gaaagagaga aaggaaaaat	3180
tgccaaaaga aatagaggaa gataaactaa aacgagaaga agaaaaaaga gaagctgaaa	3240
agagtgaaga ttctcttggt gctgcaggcc tctcaggctt acatcgaca tacagccagg	3300
actgtagctt taaaaacagc atgtaccatg ttggagatta cgtctatgtg gaacctgcag	3360
aggccaacct acaaccacat atcgtctgta ttgaaagact gtgggaggat tcagctggtg	3420
aaaaatgggt gtatggctgt tggttttacc gaccaaata aacattccac ctggctacac	3480
gaaaatttct agaaaaagaa gtttttaaga gtgactatta caacaaagt ccagtttagta	3540
aaattctagg caagtgtgtg gtcatgtttg tcaaggaata ctttaagtta tgcccagaaa	3600
acttccgaga tgaggatgtt ttgtctgtg aatcacggtt ttctgccaaa accaaatctt	3660
ttaagaaaat taaactgtgg accatgcca tcagctcagt caggtttgct cctcgggatg	3720
tgctctgcc tgtggttcgc gtggcctctg tatttgcaaa tgcagataaa ggtgatgatg	3780
agaagaatac agacaactca gaggacagtc gagctgaaga caattttaac ttggaaaagg	3840
aaaaagaaga tgtccctgtg gaaatgtcca atggtgaacc aggttgccac tactttgagc	3900
agctccatta caatgacatg tggctgaagg ttggcgactg tgtcttcac aagtcccatg	3960
gcctggtgcg tctctgtgtg ggcagaattg aaaaagtatg ggttcgagat ggagctgcat	4020
atttttatgg ccccatcttc attcacccag aagaaacaga gcatgagccc acaaaaatgt	4080
tctacaaaa agaagtattt ctgagtaatc tggaagaaac ctgccccatg acatgtattc	4140
tcggaaagtg tgcgtgtgtg tcattcaagg acttcctctc ctgcaggcca actgaaatac	4200
cagaaaaatga cattctgctt tgtgagagcc gctacaatga gagcgacaag cagatgaaga	4260
aattcaaagg attgaagagg ttttactct ctgctaaagt ggtagatgat gaaatttact	4320
acttcagaaa accaattgtt cctcagaagg agccatcacc tttgctggaa aagaagatcc	4380
agttgctaga agctaaattt gccgagttag aagggtggaga tgatgatatt gaagagatgg	4440
gagaagaaga tagtgaggtc attgaacctc cttctctacc tcagcttcag acccccctgg	4500
ccagtgagct ggacctcatg ccctacacac ccccacagtc taccccaaag tctgccaaag	4560
gcagtgcaaa gaaggaaggc tccaaacgga aaatcaacat gagtggctac atcctgttca	4620

gcagtgagat gagggctgtg attaaggccc aacacccaga ctactctttc ggggagctca	4680
gccgcctggg ggggacagaa tggagaaatc ttgagacagc caagaaagca gaatatgaag	4740
gtgtgatgaa ccaaggagtg gccctatgg tagggactcc agcaccaggt ggaagtccat	4800
atggacaaca ggiggagatt ttggggcctc cagggcagca ggcaccacct ccataccccg	4860
gcccacatcc agctggacct cctgtcatac agcagccaac aacacccatg ttgttagctc	4920
ccccacaaa gaccagcggg cttcttcact cagaggccta cctgaaatac attgaaggac	4980
tcagtgcgga gtccaacagc attagcaagt gggatcagac actggcagct cgaagacgcg	5040
acgtccattt gtcgaaagaa caggagagcc gcctaccctc tcactggctg aaaagcaaag	5100
gggcccacac caccatggca gatgccctct ggcgcttcg agatttgatg ctccgggaca	5160
ccctcaacat tcgccaagca tacaacctag aaaatgttta atcacatcat tacgtttctt	5220
ttatatagaa gcataaagag ttgtggatca gttagcattt tagttactgg ggggtggggg	5280
aaggaacaaa ggaggataat ttttattgca ttttactgta catcacaagg ccatttttat	5340
atacggacac ttttaataag ctatttcaat ttgtttgtta tattaagttg actttatcaa	5400
atacacaag atttttttgc atatgtttcc ttctgtttaaa accagtttca taattggttg	5460
tatatgtaga ctggagattt tatcttttta ctgtttgcca tggaactgaa accattagag	5520
gtttttgtct tggtttgggg tttttgtttt ctgtgttttg ggttttttta tatatatata	5580
taaaagaaca aaatgaaaaa aaacacacac acacaagagt ttacagatta gtttaaatg	5640
ataatgaaat gtgaagtttg tcttagttta catcttagag aggggagtat acttgtttt	5700
gtttcatgtg cctgaatata ttaagccact ttctgcaaaa gctgtttctt acagatgaag	5760
tgctttcttt gaaagggtgt tatttaggtt ttatagtttt aatagacaca gcacatttgc	5820
tctattaact cagaggctca ctacagaaat atgtaatcag tgctgtgcat ctgtctgcag	5880
ctaattgacc tctggacac caggagggga aaaagcactt tttcaattgt gctgagttag	5940
acatctgtga gtagactat ggtgtcagtg atttttgcag aacacgtgca caacctgag	6000
gtatgtttaa tctaggcagg tacgtttaag gatattttga tctatttata atgaattcac	6060
aatttatgcc tataaatttc agatgattta aaattttaaa cctgttacat tgaaaaacat	6120
tgaagttcgt ctgaagaaa gcattaaggt atgcatggag gtgatttatt tttaaacata	6180
acacctaacc taacatgggt aagagagtat ggaactagat atgagctgta taagaagcat	6240
aattgtgaac aagtagattg attgccttca tatacaagta tgttttagta ttccttattt	6300
ccttattatc agatgtattt tttcttttaa gtttcaatgt tgttataatt ctcaaccaga	6360

aatttaatac ttictaaaa atttttttaa tttagcttgt gcttttgaat tacaggagaa 6420
 gggaatcata atttaataaa acgcttacta gaaagacat tacagatccc aaacacttgg 6480
 gtttggtgac cctgtcttct ttatatgacc ctacaataaa catttgaagg cagcatagga 6540
 tggcagacag taggaacatt gtttcacttg gcggcatgtt tttgaaacct gctttatagt 6600
 aactgggtga ttgccattgt ggtagagctt ccactgctgt ttataatctg agagagttaa 6660
 tctcagagga tgcttttttc cttttaatct gctatgaatc agtaccaga tgtttaatta 6720
 ctgtacttat taaatcatga gggcaaaaaga gtgtagaatg gaaaaaagtc tcttgtatct 6780

agatacttta aatatgggag gccctttaac ttaattgcct ttagtcaacc actggatttg 6840
 aatttgcac aagtatttta aataatattg aatttaaaaa aatgtattgc agtagtgtgt 6900
 cagtaccta ttgttaaagt gagtacata aatcttcaat tcctggctat ttgggcaatt 6960
 gaatcatcat ggactgtata atgcaatcag attattttgt ttctagacat ccttgaatta 7020
 caccaaagaa catgaaattt agtttgtgtt aaattattta tttatttcat gcattcattt 7080
 tatttccctt aaggtctgga tgagacttct ttggggagcc tctaaaaaaa tttttactg 7140
 ggggccacgt gggtcattag aagccagagc tctctccag gtccttccc agtgcctaga 7200

ggtgctatag gaaacataga tccagccagg ggcttccta aagcagtga gcaccggccc 7260
 agggcatcac tagacaggcc ctaattaagt tttttttaa aagcctgtgt atttatttta 7320
 gaatcatgtt tttctgtata ttaacttggg ggatatcgtt aatathtagg atataagatt 7380
 tgaggtcagc catcttcaaa aaagaaaaaa aaattgactc aagaaagtac aagtaacta 7440
 tacacctttt tttcataagt tttaggaact gtagtaatgt ggcttagaaa gtataatggc 7500
 ctaaagtgtt tcaaaatgta agttcctgtg gagaagaatt gtttatattg caaacggggg 7560
 gactgagggg aacctgtagg tttaaaacag tatgtttgtc agccaactga tttaaaaggc 7620

ctttaactgt ttgtgtgtt gttttttttt taagccactc tccccttct atgaggaaga 7680
 attgagaggg gcacctattt ctgtaaaatc cccaaattgg tgttgatgat tttagacttg 7740
 aatgttttca tacctgatta aaacttgggt tatttctaatt tctgtatcat atcatctgag 7800
 gtttacgtgg taactagtct tataacatgt atgtatcttt ttttgttgt tcatctaaag 7860
 ctttttaate caaataaata cagagtttgc aaagtgttt ggattaacca ggaaaaaaa 7920
 aaaaaaaaaa 7930

<210> 4

<211> 1618

<212> PRT

<213> Homo sapiens

<400> 4

Met Arg Arg Leu Ala Phe Arg Gly Ala Gly Cys Ala Leu Val Lys Leu

1 5 10 15

Lys Lys Leu Asp Ser Met Gly Ser Lys Arg Arg Arg Ala Thr Ser Pro

20 25 30

Ser Ser Ser Val Ser Gly Asp Phe Asp Asp Gly His His Ser Val Ser

35 40 45

Thr Pro Gly Pro Ser Arg Lys Arg Arg Arg Leu Ser Asn Leu Pro Thr

50 55 60

Val Asp Pro Ile Ala Val Cys His Glu Leu Tyr Asn Thr Ile Arg Asp

65 70 75 80

Tyr Lys Asp Glu Gln Gly Arg Leu Leu Cys Glu Leu Phe Ile Arg Ala

85 90 95

Pro Lys Arg Arg Asn Gln Pro Asp Tyr Tyr Glu Val Val Ser Gln Pro

100 105 110

Ile Asp Leu Met Lys Ile Gln Gln Lys Leu Lys Met Glu Glu Tyr Asp

115 120 125

Asp Val Asn Leu Leu Thr Ala Asp Phe Gln Leu Leu Phe Asn Asn Ala

130 135 140

Lys Ser Tyr Tyr Lys Pro Asp Ser Pro Glu Tyr Lys Ala Ala Cys Lys

145 150 155 160

Leu Trp Asp Leu Tyr Leu Arg Thr Arg Asn Glu Phe Val Gln Lys Gly

165 170 175

Glu Ala Asp Asp Glu Asp Asp Asp Glu Asp Gly Gln Asp Asn Gln Gly

180 185 190

Thr Val Thr Glu Gly Ser Ser Pro Ala Tyr Leu Lys Glu Ile Leu Glu

195 200 205

Gln Leu Leu Glu Ala Ile Val Val Ala Thr Asn Pro Ser Gly Arg Leu

210 215 220

Ile Ser Glu Leu Phe Gln Lys Leu Pro Ser Lys Val Gln Tyr Pro Asp

225 230 235 240

Tyr Tyr Ala Ile Ile Lys Glu Pro Ile Asp Leu Lys Thr Ile Ala Gln

245 250 255
 Arg Ile Gln Asn Gly Ser Tyr Lys Ser Ile His Ala Met Ala Lys Asp
 260 265 270
 Ile Asp Leu Leu Ala Lys Asn Ala Lys Thr Tyr Asn Glu Pro Gly Ser

 275 280 285
 Gln Val Phe Lys Asp Ala Asn Ser Ile Lys Lys Ile Phe Tyr Met Lys
 290 295 300
 Lys Ala Glu Ile Glu His His Glu Met Ala Lys Ser Ser Leu Arg Met
 305 310 315 320
 Arg Thr Pro Ser Asn Leu Ala Ala Ala Arg Leu Thr Gly Pro Ser His
 325 330 335
 Ser Lys Gly Ser Leu Gly Glu Glu Arg Asn Pro Thr Ser Lys Tyr Tyr
 340 345 350

 Arg Asn Lys Arg Ala Val Gln Gly Gly Arg Leu Ser Ala Ile Thr Met
 355 360 365
 Ala Leu Gln Tyr Gly Ser Glu Ser Glu Glu Asp Ala Ala Leu Ala Ala
 370 375 380
 Ala Arg Tyr Glu Glu Gly Glu Ser Glu Ala Glu Ser Ile Thr Ser Phe
 385 390 395 400
 Met Asp Val Ser Asn Pro Phe Tyr Gln Leu Tyr Asp Thr Val Arg Ser
 405 410 415
 Cys Arg Asn Asn Gln Gly Gln Leu Ile Ala Glu Pro Phe Tyr His Leu

 420 425 430
 Pro Ser Lys Lys Lys Tyr Pro Asp Tyr Tyr Gln Gln Ile Lys Met Pro
 435 440 445
 Ile Ser Leu Gln Gln Ile Arg Thr Lys Leu Lys Asn Gln Glu Tyr Glu
 450 455 460
 Thr Leu Asp His Leu Glu Cys Asp Leu Asn Leu Met Phe Glu Asn Ala
 465 470 475 480
 Lys Arg Tyr Asn Val Pro Asn Ser Ala Ile Tyr Lys Arg Val Leu Lys
 485 490 495

Leu Gln Gln Val Met Gln Ala Lys Lys Lys Glu Leu Ala Arg Arg Asp
 500 505 510
 Asp Ile Glu Asp Gly Asp Ser Met Ile Ser Ser Ala Thr Ser Asp Thr
 515 520 525
 Gly Ser Ala Lys Arg Lys Arg Asn Thr His Asp Ser Glu Met Leu Gly
 530 535 540
 Leu Arg Arg Leu Ser Ser Lys Lys Asn Ile Arg Lys Gln Arg Met Lys
 545 550 555 560
 Ile Leu Phe Asn Val Val Leu Glu Ala Arg Glu Pro Gly Ser Gly Arg
 565 570 575
 Arg Leu Cys Asp Leu Phe Met Val Lys Pro Ser Lys Lys Asp Tyr Pro
 580 585 590
 Asp Tyr Tyr Lys Ile Ile Leu Glu Pro Met Asp Leu Lys Ile Ile Glu
 595 600 605
 His Asn Ile Arg Asn Asp Lys Tyr Ala Gly Glu Glu Gly Met Ile Glu
 610 615 620
 Asp Met Lys Leu Met Phe Arg Asn Ala Arg His Tyr Asn Glu Glu Gly
 625 630 635 640
 Ser Gln Val Tyr Asn Asp Ala His Ile Leu Glu Lys Leu Leu Lys Glu
 645 650 655
 Lys Arg Lys Glu Leu Gly Pro Leu Pro Asp Asp Asp Asp Met Ala Ser
 660 665 670
 Pro Lys Leu Lys Leu Ser Arg Lys Ser Gly Ile Ser Pro Lys Lys Ser
 675 680 685
 Lys Tyr Met Thr Pro Met Gln Gln Lys Leu Asn Glu Val Tyr Glu Ala
 690 695 700
 Val Lys Asn Tyr Thr Asp Lys Arg Gly Arg Arg Leu Ser Ala Ile Phe
 705 710 715 720
 Leu Arg Leu Pro Ser Arg Ser Glu Leu Pro Asp Tyr Tyr Leu Thr Ile
 725 730 735
 Lys Lys Pro Met Asp Met Glu Lys Ile Arg Ser His Met Met Ala Asn
 740 745 750

Lys Tyr Gln Asp Ile Asp Ser Met Val Glu Asp Phe Val Met Met Phe
 755 760 765
 Asn Asn Ala Cys Thr Tyr Asn Glu Pro Glu Ser Leu Ile Tyr Lys Asp
 770 775 780

 Ala Leu Val Leu His Lys Val Leu Leu Glu Thr Arg Arg Asp Leu Glu
 785 790 795 800
 Gly Asp Glu Asp Ser His Val Pro Asn Val Thr Leu Leu Ile Gln Glu
 805 810 815
 Leu Ile His Asn Leu Phe Val Ser Val Met Ser His Gln Asp Asp Glu
 820 825 830
 Gly Arg Cys Tyr Ser Asp Ser Leu Ala Glu Ile Pro Ala Val Asp Pro
 835 840 845
 Asn Phe Pro Asn Lys Pro Pro Leu Thr Phe Asp Ile Ile Arg Lys Asn

 850 855 860
 Val Glu Asn Asn Arg Tyr Arg Arg Leu Asp Leu Phe Gln Glu His Met
 865 870 875 880
 Phe Glu Val Leu Glu Arg Ala Arg Arg Met Asn Arg Thr Asp Ser Glu
 885 890 895
 Ile Tyr Glu Asp Ala Val Glu Leu Gln Gln Phe Phe Ile Lys Ile Arg
 900 905 910
 Asp Glu Leu Cys Lys Asn Gly Glu Ile Leu Leu Ser Pro Ala Leu Ser
 915 920 925

 Tyr Thr Thr Lys His Leu His Asn Asp Val Glu Lys Glu Arg Lys Glu
 930 935 940
 Lys Leu Pro Lys Glu Ile Glu Glu Asp Lys Leu Lys Arg Glu Glu Glu
 945 950 955 960
 Lys Arg Glu Ala Glu Lys Ser Glu Asp Ser Ser Gly Ala Ala Gly Leu
 965 970 975
 Ser Gly Leu His Arg Thr Tyr Ser Gln Asp Cys Ser Phe Lys Asn Ser
 980 985 990
 Met Tyr His Val Gly Asp Tyr Val Tyr Val Glu Pro Ala Glu Ala Asn

995 1000 1005
 Leu Gln Pro His Ile Val Cys Ile Glu Arg Leu Trp Glu Asp Ser Ala
 1010 1015 1020
 Gly Glu Lys Trp Leu Tyr Gly Cys Trp Phe Tyr Arg Pro Asn Glu Thr
 1025 1030 1035 1040
 Phe His Leu Ala Thr Arg Lys Phe Leu Glu Lys Glu Val Phe Lys Ser
 1045 1050 1055
 Asp Tyr Tyr Asn Lys Val Pro Val Ser Lys Ile Leu Gly Lys Cys Val
 1060 1065 1070

 Val Met Phe Val Lys Glu Tyr Phe Lys Leu Cys Pro Glu Asn Phe Arg
 1075 1080 1085
 Asp Glu Asp Val Phe Val Cys Glu Ser Arg Tyr Ser Ala Lys Thr Lys
 1090 1095 1100
 Ser Phe Lys Lys Ile Lys Leu Trp Thr Met Pro Ile Ser Ser Val Arg
 1105 1110 1115 1120
 Phe Val Pro Arg Asp Val Pro Leu Pro Val Val Arg Val Ala Ser Val
 1125 1130 1135
 Phe Ala Asn Ala Asp Lys Gly Asp Asp Glu Lys Asn Thr Asp Asn Ser

 1140 1145 1150
 Glu Asp Ser Arg Ala Glu Asp Asn Phe Asn Leu Glu Lys Glu Lys Glu
 1155 1160 1165
 Asp Val Pro Val Glu Met Ser Asn Gly Glu Pro Gly Cys His Tyr Phe
 1170 1175 1180
 Glu Gln Leu His Tyr Asn Asp Met Trp Leu Lys Val Gly Asp Cys Val
 1185 1190 1195 1200
 Phe Ile Lys Ser His Gly Leu Val Arg Pro Arg Val Gly Arg Ile Glu
 1205 1210 1215

 Lys Val Trp Val Arg Asp Gly Ala Ala Tyr Phe Tyr Gly Pro Ile Phe
 1220 1225 1230
 Ile His Pro Glu Glu Thr Glu His Glu Pro Thr Lys Met Phe Tyr Lys
 1235 1240 1245
 Lys Glu Val Phe Leu Ser Asn Leu Glu Glu Thr Cys Pro Met Thr Cys

1250 1255 1260
 Ile Leu Gly Lys Cys Ala Val Leu Ser Phe Lys Asp Phe Leu Ser Cys
 1265 1270 1275 1280
 Arg Pro Thr Glu Ile Pro Glu Asn Asp Ile Leu Leu Cys Glu Ser Arg

 1285 1290 1295
 Tyr Asn Glu Ser Asp Lys Gln Met Lys Lys Phe Lys Gly Leu Lys Arg
 1300 1305 1310
 Phe Ser Leu Ser Ala Lys Val Val Asp Asp Glu Ile Tyr Tyr Phe Arg
 1315 1320 1325
 Lys Pro Ile Val Pro Gln Lys Glu Pro Ser Pro Leu Leu Glu Lys Lys
 1330 1335 1340
 Ile Gln Leu Leu Glu Ala Lys Phe Ala Glu Leu Glu Gly Gly Asp Asp
 1345 1350 1355 1360

 Asp Ile Glu Glu Met Gly Glu Glu Asp Ser Glu Val Ile Glu Pro Pro
 1365 1370 1375
 Ser Leu Pro Gln Leu Gln Thr Pro Leu Ala Ser Glu Leu Asp Leu Met
 1380 1385 1390
 Pro Tyr Thr Pro Pro Gln Ser Thr Pro Lys Ser Ala Lys Gly Ser Ala
 1395 1400 1405
 Lys Lys Glu Gly Ser Lys Arg Lys Ile Asn Met Ser Gly Tyr Ile Leu
 1410 1415 1420
 Phe Ser Ser Glu Met Arg Ala Val Ile Lys Ala Gln His Pro Asp Tyr

 1425 1430 1435 1440
 Ser Phe Gly Glu Leu Ser Arg Leu Val Gly Thr Glu Trp Arg Asn Leu
 1445 1450 1455
 Glu Thr Ala Lys Lys Ala Glu Tyr Glu Gly Val Met Asn Gln Gly Val
 1460 1465 1470
 Ala Pro Met Val Gly Thr Pro Ala Pro Gly Gly Ser Pro Tyr Gly Gln
 1475 1480 1485
 Gln Val Gly Val Leu Gly Pro Pro Gly Gln Gln Ala Pro Pro Pro Tyr
 1490 1495 1500

Pro Gly Pro His Pro Ala Gly Pro Pro Val Ile Gln Gln Pro Thr Thr
 1505 1510 1515 1520
 Pro Met Phe Val Ala Pro Pro Pro Lys Thr Gln Arg Leu Leu His Ser
 1525 1530 1535
 Glu Ala Tyr Leu Lys Tyr Ile Glu Gly Leu Ser Ala Glu Ser Asn Ser
 1540 1545 1550
 Ile Ser Lys Trp Asp Gln Thr Leu Ala Ala Arg Arg Arg Asp Val His
 1555 1560 1565
 Leu Ser Lys Glu Gln Glu Ser Arg Leu Pro Ser His Trp Leu Lys Ser
 1570 1575 1580
 Lys Gly Ala His Thr Thr Met Ala Asp Ala Leu Trp Arg Leu Arg Asp
 1585 1590 1595 1600
 Leu Met Leu Arg Asp Thr Leu Asn Ile Arg Gln Ala Tyr Asn Leu Glu
 1605 1610 1615
 Asn Val

<210> 5

<211> 3600

<212> DNA

<213> Homo sapiens

<400> 5

gggtgacgcg actgggcccg ttgtctgtgt gtgggactga ggggccccgg gggcgggtggg 60

ggctcccggt gggggcagcg gtggggaggg agggcctgga catggcgctg aggggccgcc 120

ccgcgggaag atgaataagg gctgctgga gctggagagc gaccaggcc tcttcacct 180

gctcgtggaa gatttcggtg tcaaggggt gcaagtggag gagatctacg accttcagag 240

caaatgtcag ggcctgtat atggatttat ctctctgttc aaatggatcg aagagcgccg 300

gtcccgccga aaggtctcta ccttggtgga tgatactcc gtgattgatg atgatattgt 360

gaataacatg ttctttgccc accagctgat acccaactct tgtgcaactc atgccttget 420

gagcgtgctc ctgaactgca gcagcgtgga cctgggaccc accctgagtc gcatgaagga 480

cttcaccaag ggtttcagcc ctgagagcaa aggatatgcg attggcaatg ccccgagtt 540

ggccaaggcc cataatagcc atgccaggcc cgagccacgc cacctccctg agaagcagaa 600

tggccttagt gcagtgcgga ccatggaggc gttccacttt gtcagctatg tgcctatcac 660

aggccggctc tttagctgg atgggctgaa ggtctacccc attgacatg ggccctgggg	720
ggaggacgag gattggacag acaaggcccg gcgggtcatc atggagcgta tcggcctcgc	780
cactgcaggg gagccctacc acgacatccg cttcaacctg atggcagtgg tgcccgaccg	840
caggatcaag tatgaggcca ggctgcatgt gctgaagggtg aaccgtcaga cagtactaga	900
ggctctgcag cagctgataa gagtaacaca gccagagctg attcagaccc acaagtctca	960
agagtacag ctgcctgagg agtccaagtc agccagcaac aagtccccgc tgggtgctgga	1020
agcaaacagg gccctgcag cctctgaggg caaccacaca gatgggtcag aggaggcggc	1080
tggttcatgc gcacaagccc catccacag ccctcccaac aaaccaagc tagtggtgaa	1140
gcctccaggc agcagcctca atggggttca cccaacccc actccattg tccagcggct	1200
gccggccttt ctagacaatc acaattatgc caagtcccc atgcaggagg aagaagacct	1260
ggcggcaggt gtgggccga gccgagttcc agtccgcca cccagcagt actcagatga	1320
tgaggatgac tatgaggatg acgaggagga tgacgtgcag aacaccaact ctgcccttag	1380
gtataagggg aagggaacag ggaagccagg ggcatcgagc ggttctgctg atgggcaact	1440
gtcagtgtg cagcccaaca ccatcaacgt cttggctgag aagctcaaag agtcccagaa	1500
ggacctctca attcctctgt ccatcaagac tagcagcggg gctgggagtc cggctgtggc	1560
agtcccaca cactgcagc cctcaccac cccagcaat gagagtacag acacggcctc	1620
tgagatggc agtgcttca actcgccact gcgctgcct atccgctcag ccaaccgac	1680
gcggccctcc agccctgtca cctccacat ctccaagtg ctttttgag aggatgacag	1740
cctgtctcgt gttgactgca tacgtacaa cgtgtctgtc cgtgatctgg gtcctgtcat	1800
cagcacaggc ctgctgcacc tggctgagga tgggtgctg agtcccctgg cgctgacaga	1860
gggtgggaag ggttcctcgc cctccatcag accaatccaa ggcagccagg ggtccagcag	1920
cccagtggag aaggaggtcg tggaagccac ggacagcaga gagaagacgg ggatggtgag	1980
gcctggcgag cccttgagtg gggagaaata ctaccaag gagctgctgg cactgctgaa	2040
gtgtgtggag gctgagattg caaactatga ggcgtgcctc aaggaggagg tagagaagag	2100
gaagaagttc aagattgatg accagagaag gaccacaac tacgatgagt tcactctcac	2160
ctttatctcc atgctggctc aggaaggcat gctggccaac ctatggagc agaacatctc	2220
cgctcggcgg cgccaagggg tcagcatcgg ccggtccac aagcagcgga agcctgaccg	2280
gcggaacgc tctcgccct acaaggccaa gcgccagtga ggactgctgg ccctgactct	2340
gcagcccact cttgccgtgt ggccctcacc agggctcttc cctgccccac ttcccccttt	2400
cccagtatta ctgaatagtc ccagctggag agtccaggcc ctgggaatgg gaggaaccag	2460
gccacattcc ttccatcgtg ccctgaggcc tgacacggca gatcagcccc atagtgtca	2520

ggaggcagca tctggagtgt gggcacagcg aggtactgca gcttcctcca cagccggctg 2580
 tggagcagca ggacctggcc cttctgcctg ggcagcagaa tataatattt acctatcaga 2640

gacatctatt tttctgggct ccaacccaac atgccacat gttgacataa gttcctacct 2700
 gactatgctt tctctcctag gagctgtcct ggtgggcca ggtccttgta tcatgccacg 2760
 gtcccaacta cagggtccta gctgggggcc tgggtgggcc ctgggctctg ggccctgctg 2820
 ctctagcccc agccaccagc ctgtccctgt tgaaggaag ccaggctctt tctcttcatt 2880
 cctcttagga gactgccaaa ctgaggacc cagcactggg ctgggttggg agtagggtgt 2940
 cccagtgggg ttgggtgag caggctgctg ggatcccatg gcctgagcag agcatgtggg 3000
 aactgttcag tggcctgtga actgtcttcc ttgttctagc caggctgttc aagactgctc 3060

tccatagcaa ggttctaggg ctcttcgctt tcagtgttgt ggccctagct atgggcctaa 3120
 attgggctct aggtctctgt ccctggcgtt ttaggtctag aagagcctct gtccagcccc 3180
 tcagtattac catgtctccc tctcaggggt agcagagaca ggggttgctta taggaagctg 3240
 gcaccactca gctcttcctg ctactccagt ttctcagcc tctgcaagc actcagggtg 3300
 ggggacagca ggatcaagac aaccgttgg agccctgtg ttccagagga cctgatgcca 3360
 aggggtaatg ggcccagcag tgcctctgga gccagggccc caacacagcc ccatggcctc 3420
 tgccagatgg ctttgaaaaa ggtgatccaa gcaggccctt ttatctgtac atagtactg 3480

agtgggggtt gctggcaagt gtggcagctg cctctgggct gagcacagct tgaccctctt 3540
 agccctgtga aatactgat caatgaatga ataaaactct cctaagaatc tctgagaaa 3600
 3600

<210> 6

<211> 729

<212> PRT

<213> Homo sapiens

<400> 6

Met Asn Lys Gly Trp Leu Glu Leu Glu Ser Asp Pro Gly Leu Phe Thr

1 5 10 15

Leu Leu Val Glu Asp Phe Gly Val Lys Gly Val Gln Val Glu Glu Ile

20 25 30

Tyr Asp Leu Gln Ser Lys Cys Gln Gly Pro Val Tyr Gly Phe Ile Phe

35 40 45

Leu Phe Lys Trp Ile Glu Glu Arg Arg Ser Arg Arg Lys Val Ser Thr

50 55 60
 Leu Val Asp Asp Thr Ser Val Ile Asp Asp Asp Ile Val Asn Asn Met
 65 70 75 80
 Phe Phe Ala His Gln Leu Ile Pro Asn Ser Cys Ala Thr His Ala Leu
 85 90 95
 Leu Ser Val Leu Leu Asn Cys Ser Ser Val Asp Leu Gly Pro Thr Leu

 100 105 110
 Ser Arg Met Lys Asp Phe Thr Lys Gly Phe Ser Pro Glu Ser Lys Gly
 115 120 125
 Tyr Ala Ile Gly Asn Ala Pro Glu Leu Ala Lys Ala His Asn Ser His
 130 135 140
 Ala Arg Pro Glu Pro Arg His Leu Pro Glu Lys Gln Asn Gly Leu Ser
 145 150 155 160
 Ala Val Arg Thr Met Glu Ala Phe His Phe Val Ser Tyr Val Pro Ile
 165 170 175

 Thr Gly Arg Leu Phe Glu Leu Asp Gly Leu Lys Val Tyr Pro Ile Asp
 180 185 190
 His Gly Pro Trp Gly Glu Asp Glu Glu Trp Thr Asp Lys Ala Arg Arg
 195 200 205
 Val Ile Met Glu Arg Ile Gly Leu Ala Thr Ala Gly Glu Pro Tyr His
 210 215 220
 Asp Ile Arg Phe Asn Leu Met Ala Val Val Pro Asp Arg Arg Ile Lys
 225 230 235 240
 Tyr Glu Ala Arg Leu His Val Leu Lys Val Asn Arg Gln Thr Val Leu

 245 250 255
 Glu Ala Leu Gln Gln Leu Ile Arg Val Thr Gln Pro Glu Leu Ile Gln
 260 265 270
 Thr His Lys Ser Gln Glu Ser Gln Leu Pro Glu Glu Ser Lys Ser Ala
 275 280 285
 Ser Asn Lys Ser Pro Leu Val Leu Glu Ala Asn Arg Ala Pro Ala Ala
 290 295 300
 Ser Glu Gly Asn His Thr Asp Gly Ala Glu Glu Ala Ala Gly Ser Cys

305						310						315						320
Ala	Gln	Ala	Pro	Ser	His	Ser	Pro	Pro	Asn	Lys	Pro	Lys	Leu	Val	Val			
					325						330						335	
Lys	Pro	Pro	Gly	Ser	Ser	Leu	Asn	Gly	Val	His	Pro	Asn	Pro	Thr	Pro			
					340						345						350	
Ile	Val	Gln	Arg	Leu	Pro	Ala	Phe	Leu	Asp	Asn	His	Asn	Tyr	Ala	Lys			
					355						360						365	
Ser	Pro	Met	Gln	Glu	Glu	Glu	Asp	Leu	Ala	Ala	Gly	Val	Gly	Arg	Ser			
					370						375						380	
Arg	Val	Pro	Val	Arg	Pro	Pro	Gln	Gln	Tyr	Ser	Asp	Asp	Glu	Asp	Asp			
					385						390						395	400
Tyr	Glu	Asp	Asp	Glu	Glu	Asp	Asp	Val	Gln	Asn	Thr	Asn	Ser	Ala	Leu			
					405						410						415	
Arg	Tyr	Lys	Gly	Lys	Gly	Thr	Gly	Lys	Pro	Gly	Ala	Leu	Ser	Gly	Ser			
					420						425						430	
Ala	Asp	Gly	Gln	Leu	Ser	Val	Leu	Gln	Pro	Asn	Thr	Ile	Asn	Val	Leu			
					435						440						445	
Ala	Glu	Lys	Leu	Lys	Glu	Ser	Gln	Lys	Asp	Leu	Ser	Ile	Pro	Leu	Ser			
					450						455						460	
Ile	Lys	Thr	Ser	Ser	Gly	Ala	Gly	Ser	Pro	Ala	Val	Ala	Val	Pro	Thr			
					465						470						475	480
His	Ser	Gln	Pro	Ser	Pro	Thr	Pro	Ser	Asn	Glu	Ser	Thr	Asp	Thr	Ala			
					485						490						495	
Ser	Glu	Ile	Gly	Ser	Ala	Phe	Asn	Ser	Pro	Leu	Arg	Ser	Pro	Ile	Arg			
					500						505						510	
Ser	Ala	Asn	Pro	Thr	Arg	Pro	Ser	Ser	Pro	Val	Thr	Ser	His	Ile	Ser			
					515						520						525	
Lys	Val	Leu	Phe	Gly	Glu	Asp	Asp	Ser	Leu	Leu	Arg	Val	Asp	Cys	Ile			
					530						535						540	
Arg	Tyr	Asn	Arg	Ala	Val	Arg	Asp	Leu	Gly	Pro	Val	Ile	Ser	Thr	Gly			
					545						550						555	560

Leu Leu His Leu Ala Glu Asp Gly Val Leu Ser Pro Leu Ala Leu Thr
 565 570 575
 Glu Gly Gly Lys Gly Ser Ser Pro Ser Ile Arg Pro Ile Gln Gly Ser
 580 585 590
 Gln Gly Ser Ser Ser Pro Val Glu Lys Glu Val Val Glu Ala Thr Asp
 595 600 605

 Ser Arg Glu Lys Thr Gly Met Val Arg Pro Gly Glu Pro Leu Ser Gly
 610 615 620
 Glu Lys Tyr Ser Pro Lys Glu Leu Leu Ala Leu Leu Lys Cys Val Glu
 625 630 635 640
 Ala Glu Ile Ala Asn Tyr Glu Ala Cys Leu Lys Glu Glu Val Glu Lys
 645 650 655
 Arg Lys Lys Phe Lys Ile Asp Asp Gln Arg Arg Thr His Asn Tyr Asp
 660 665 670
 Glu Phe Ile Cys Thr Phe Ile Ser Met Leu Ala Gln Glu Gly Met Leu

 675 680 685
 Ala Asn Leu Val Glu Gln Asn Ile Ser Val Arg Arg Arg Gln Gly Val
 690 695 700
 Ser Ile Gly Arg Leu His Lys Gln Arg Lys Pro Asp Arg Arg Lys Arg
 705 710 715 720
 Ser Arg Pro Tyr Lys Ala Lys Arg Gln
 725

 <210> 7
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> FOXC2 forward primer
 <400> 7
 gatcaccttg aacggcatct

20

<210> 8
 <211> 20
 <212> DNA

<213> Artificial Sequence

<220><223> FOXC2 reverse primer

<400> 8

accttgacga agcactcgtt 20

<210> 9

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> CLIP4 forward primer

<400> 9

gcatcatgcc aggaaattct 20

<210> 10

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> CLIP4 reverse primer

<400> 10

tttgttgac ctgaggaacc 20

<210> 11

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> PBRM1 forward primer

<400> 11

tgatggccaa caagtaccaa 20

<210> 12

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> PBRM1 reverse primer

<400> 12

agatcaaaga ctccggctca 20

<210> 13

<211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> BAP1 forward primer
 <400> 13
 gcctgaggag tccaagtcag 20
 <210> 14
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> BAP1 reverse primer
 <400> 14
 ctggaggctt caccactagc 20
 <210> 15
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> SETD2 forward primer
 <400> 15
 tcacaaggca gactcagtgg 20
 <210> 16
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> SETD2 reverse primer
 <400> 16
 ctgctgtctt gggctttttc 20
 <210> 17
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> KDM5C forward primer
 <400> 17
 gtcatttgca acccctgagt 20

<210>	18	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	KDM5C reverse primer	
<400>	18	
	aatgggatga ggggtaaagg	20
<210>	19	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	GAPDH forward primer	
<400>	19	
	cagcctcaag atcatcagca	20
<210>	20	
<211>	20	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	GAPDH reverse primer	
<400>	20	
	ggtgctaagc agttggtggt	20