



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2021년07월19일

(11) 등록번호 10-2279107

(24) 등록일자 2021년07월13일

(51) 국제특허분류(Int. Cl.)
G16H 50/20 (2018.01) **G01N 30/72** (2006.01)
G01N 30/88 (2006.01)
 (52) CPC특허분류
G16H 50/20 (2018.01)
G01N 30/72 (2013.01)
 (21) 출원번호 10-2019-0114969
 (22) 출원일자 2019년09월18일
 심사청구일자 2019년09월18일
 (65) 공개번호 10-2021-0033347
 (43) 공개일자 2021년03월26일
 (56) 선행기술조사문헌
 KR1020190065267 A*
 *는 심사관에 의하여 인용된 문헌

(73) 특허권자
연세대학교 산학협력단
 서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
 (72) 발명자
강창무
 서울특별시 강남구 삼성로111길 8(삼성동, 삼성동
 힐스테이트2단지아파트)
박민수
 경기도 안양시 만안구 삼성산길 8, 108동 1302호
 (석수동, 석수역푸르지오아파트)
이상국
 서울특별시 강서구 마곡서1로 111-11, 508동 103
 호(마곡동, 마곡엠밸리5단지)
 (74) 대리인
리엔목특허법인

전체 청구항 수 : 총 8 항

심사관 : 김재현

(54) 발명의 명칭 **폐장암을 진단하기 위한 정보를 제공하는 방법****(57) 요약**

개체의 폐장암을 진단하기 위한 정보를 제공하는 방법 및 폐장암을 진단하는 키트에 관한 것이다. 일 양상에 따른 개체의 폐장암을 진단하기 위한 정보를 제공하는 방법 및 이를 이용한 키트에 따르면, 정상인 혹은 다른 질환을 앓는 환자로부터 일정 대사체 만으로 높은 정도의 특이도 및 진단도로 경제적으로 폐장암을 효과적으로 진단할 수 있어 폐장암 환자 맞춤형 치료가 가능하다.

(52) CPC특허분류

G01N 2030/8813 (2013.01)

G01N 2800/00 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1465028959
부처명	보건복지부
과제관리(전문)기관명	한국보건산업진흥원
연구사업명	침단의료기술개발
연구과제명	암 특이 마커 단백질의 혈액 내 비정상 분비기전에 대한 새로운 분석기술 개발과 임
상적 활용	
기 여 율	1/1
과제수행기관명	연세대학교 산학협력단
연구기간	2019.01.01 ~ 2019.12.31

명세서

청구범위

청구항 1

하기 단계를 포함하는 개체의 체지방을 진단하기 위한 정보를 제공하는 방법으로서,

개체의 생물학적 시료에서 70개의 대사체 시료를 획득하는 단계;

수득된 대사체 시료에서 70개의 대사체 수준을 측정하는 단계;

측정된 70개의 대사체 수준에서 각각 대사체의 평균 값을 제하고 남은 값을 표준 편차 값으로 나누고 정규화 하여 x값을 도출하는 단계;

상기 x값을 각각 하기 표 2의 지정된 대사체 상수(coefficient)와 곱하여 각각 70개의 대사체 상수의 합을 구하여 회귀계수인 -10.4432713170571를 더한 a값을 도출하는 단계; 및

상기 도출된 a 값을 하기 수학적식에 대입하여 개체가 체지방 진단 예측 확률(y)을 계산하는 단계를 포함하며,

상기 대사체는 Alanine (Ala), Asparagine (Asn), Aspartate (Asp), Citrulline (Cit), Glutamate (Glu), Histidine (His), Isoleucine (Ile), Leucine (Leu), Lysine (Lys), Methionine (Met), Ornithine (Orn), Phenylalanine (Phe), Proline (Pro), Threonine (Thr), Tryptophan (Trp), Tyrosine (Tyr), Valine (Val), Creatinine, Serotonin, Carnitine (C0), Propionyl-L-carnitine (C3), Malonyl-L-carnitine/Hydroxybutyryl-L-carnitine (C3-DC (C4-OH)), Hydroxypropionyl-L-carnitine (C3-OH), Propenyl-L-carnitine (C3:1), Butenyl-L-carnitine (C4:1), Methylglutaryl-L-carnitine (C5-M-DC), Pimelyl-L-carnitine (C7-DC), Octanoyl-L-carnitine (C8), Decadienyl-L-carnitine (C10:2), Dodecanedioyl-L-carnitine (C12-DC), Tetradecanoyl-L-carnitine (C14), Hydroxyhexadecanoyl-L-carnitine (C16-OH), Hydroxyhexadecenoyl-L-carnitine (C16:1-OH), lysoPhosphatidylcholine acyl C16:0 (lysoPC a C16:0), lysoPhosphatidylcholine acyl C18:0 (lysoPC a C18:0), lysoPhosphatidylcholine acyl C18:2 (lysoPC a C18:2), Phosphatidylcholine diacyl C26:0 (PC aa C26:0), Phosphatidylcholine diacyl C28:1 (PC aa C28:1), Phosphatidylcholine diacyl C30:2 (PC aa C30:2), Phosphatidylcholine diacyl C40:4 (PC aa C40:4), Phosphatidylcholine diacyl C42:0 (PC aa C42:0), Phosphatidylcholine diacyl C42:2 (PC aa C42:2), Phosphatidylcholine acyl-alkyl C30:0 (PC ae C30:0), Phosphatidylcholine acyl-alkyl C30:2 (PC ae C30:2), Phosphatidylcholine acyl-alkyl C32:1 (PC ae C32:1), Phosphatidylcholine acyl-alkyl C38:2 (PC ae C38:2), Phosphatidylcholine acyl-alkyl C38:4 (PC ae C38:4), Phosphatidylcholine acyl-alkyl C40:1 (PC ae C40:1), Phosphatidylcholine acyl-alkyl C40:4 (PC ae C40:4), Phosphatidylcholine acyl-alkyl C40:5 (PC ae C40:5), Phosphatidylcholine acyl-alkyl C42:3 (PC ae C42:3), Phosphatidylcholine acyl-alkyl C42:4 (PC ae C42:4), Phosphatidylcholine acyl-alkyl C42:5 (PC ae C42:5), Phosphatidylcholine acyl-alkyl C44:4 (PC ae C44:4), Phosphatidylcholine acyl-alkyl C44:5 (PC ae C44:5), Phosphatidylcholine acyl-alkyl C44:6 (PC ae C44:6), Hydroxysphingomyeline C14:1 (SM (OH) C14:1), Hydroxysphingomyeline C16:1 (SM (OH) C16:1), Hydroxysphingomyeline C22:1 (SM (OH) C22:1), Hydroxysphingomyeline C22:2 (SM (OH) C22:2), Hydroxysphingomyeline C24:1 (SM (OH) C24:1), Sphingomyeline C16:0 (SM C16:0), Sphingomyeline C16:1 (SM C16:1), Sphingomyeline C18:0 (SM C18:0), Sphingomyeline C18:1 (SM C18:1), Sphingomyeline C20:2 (SM C20:2), Sphingomyeline C24:0 (SM C24:0), Sphingomyeline C24:1 (SM C24:1), Sphingomyeline C26:0 (SM C26:0), 및 Sphingomyeline C26:1 (SM C26:1)인 것인 방법.

[수학적식]

$$y = \frac{1}{1 + e^{-a}}$$

[표 2]

대사체	지정된 상수 값 (coefficient)	대사체	지정된 상수 값	대사체	지정된 상수 값
Ala	0.676280223	C14	0.45309149	SM...OH...C24.1	-0.65556429
Asn	-0.685587685589795	C16.OH	0.37169783	SM.C16.0	-0.24091768
Asp	0.0032750441036805	C16.1.OH	0.4530083	SM.C16.1	-0.64612395
Cit	-0.383529257098579	lysoPC.a.C16.0	-0.53395448	SM.C18.0	0.33933332
Glu	0.185851714947175	lysoPC.a.C18.0	-0.45752599	SM.C18.1	-0.19660643
His	-1.73313668962738	lysoPC.a.C18.2	-1.36244982	SM.C20.2	-0.46066381
Ile	-0.665183886269878	PC.aa.C26.0	1.84041225	SM.C24.0	-0.69101574
Leu	-0.251864533741008	PC.aa.C28.1	-0.50069967	SM.C24.1	-0.19607839
Lys	0.41570457	PC.aa.C30.2	1.76119558	SM.C26.0	-0.88196312
Met	-0.49400673	PC.aa.C40.4	0.52173427	SM.C26.1	-0.17657509
Orn	-0.4925206	PC.aa.C42.0	0.28186998	Intercept (회귀계수)	-10.4432713170571
Phe	-0.36327781	PC.aa.C42.2	-0.10879668		
Pro	-0.82326267	PC.ae.C30.0	-0.6777565		
Thr	-0.63381162	PC.ae.C30.2	0.30128848		
Trp	-0.10139695	PC.ae.C32.1	1.90819002		
Tyr	-0.62060258	PC.ae.C38.2	-0.20792852		
Val	0.27740928	PC.ae.C38.4	0.03538989		
Creatinine	-0.40429159	PC.ae.C40.1	-0.16489253		
Serotonin	1.23877177	PC.ae.C40.4	0.00505762		
C0	-0.69693727	PC.ae.C40.5	0.42026536		
C3	-1.04096884	PC.ae.C42.3	0.24801568		
C3.DC...C4.0 H.	0.01366621	PC.ae.C42.4	-0.48295906		
C3.OH	-0.52241362	PC.ae.C42.5	0.03425526		
C3.1	0.6915962	PC.ae.C44.4	-0.62184863		
C4.1	-0.23422694	PC.ae.C44.5	-0.52755386		
C5.M.DC	1.00559787	PC.ae.C44.6	-0.17685685		
C7.DC	0.66469837	SM...OH...C14.1	-0.09675213		
C8	-0.11625818	SM...OH...C16.1	0.08137913		
C10.2	0.16734288	SM...OH...C22.1	-0.76483204		
C12.DC	-0.54757817	SM...OH...C22.2	-0.48884676		

청구항 2

삭제

청구항 3

삭제

청구항 4

삭제

청구항 5

청구항 1에 있어서, 상기 생물학적 시료는 혈액, 혈장, 혈소판, 혈청, 또는 이들의 조합인 것인 방법.

청구항 6

청구항 1에 있어서, 상기 회귀계수는 부분 최소제곱법 및 회소 부분 최소 제곱법으로 이루어진 군으로부터 선택되는 한 가지 이상의 회귀분석 방법으로 회귀 분석을 수행하여 계산하는 것인 방법.

청구항 7

청구항 1에 있어서, 상기 대사체의 수준을 측정하는 단계는 크로마토그래피/질량 분석법을 이용하여 수행되는 것인 방법.

청구항 8

청구항 1, 5 내지 7 중 어느 한 항에 따른 방법을 실행하기 위한 컴퓨터 프로그램을 기록한 컴퓨터로 판독 가능한 기록 매체.

청구항 9

청구항 1, 5 내지 7 중 어느 한 항에 따른 방법에 의해 개체의 췌장암을 진단하기 위한 키트.

청구항 10

청구항 9에 있어서, 상기 키트는 대사체 수준을 측정하는 체제를 포함하고, 상기 대사체는 Ala, Asn, Asp, Cit, Glu, His, Ile, Leu, Lys, Met, Orn, Phe, Pro, Thr, Trp, Tyr, Val, Creatinine, Serotonin, C0, C3, C3-DC (C4-OH), C3-OH, C3:1, C4:1, C5-M-DC, C7-DC, C8, C10:2, C12-DC, C14, C16-OH, C16:1-OH, lysoPC a C16:0, lysoPC a C18:0, lysoPC a C18:2, PC aa C26:0, PC aa C28:1, PC aa C30:2, PC aa C40:4, PC aa C42:0, PC aa C42:2, PC ae C30:0, PC ae C30:2, PC ae C32:1, PC ae C38:2, PC ae C38:4, PC ae C40:1, PC ae C40:4, PC ae C40:5, PC ae C42:3, PC ae C42:4, PC ae C42:5, PC ae C44:4, PC ae C44:5, PC ae C44:6, SM (OH) C14:1, SM (OH) C16:1, SM (OH) C22:1, SM (OH) C22:2, SM (OH) C24:1, SM C16:0, SM C16:1, SM C18:0, SM C18:1, SM C20:2, SM C24:0, SM C24:1, SM C26:0, 및 SM C26:1의 조합인 것인 키트.

청구항 11

Ala, Asn, Asp, Cit, Glu, His, Ile, Leu, Lys, Met, Orn, Phe, Pro, Thr, Trp, Tyr, Val, Creatinine, Serotonin, C0, C3, C3-DC (C4-OH), C3-OH, C3:1, C4:1, C5-M-DC, C7-DC, C8, C10:2, C12-DC, C14, C16-OH, C16:1-OH, lysoPC a C16:0, lysoPC a C18:0, lysoPC a C18:2, PC aa C26:0, PC aa C28:1, PC aa C30:2, PC aa C40:4, PC aa C42:0, PC aa C42:2, PC ae C30:0, PC ae C30:2, PC ae C32:1, PC ae C38:2, PC ae C38:4, PC ae C40:1, PC ae C40:4, PC ae C40:5, PC ae C42:3, PC ae C42:4, PC ae C42:5, PC ae C44:4, PC ae C44:5, PC ae C44:6, SM (OH) C14:1, SM (OH) C16:1, SM (OH) C22:1, SM (OH) C22:2, SM (OH) C24:1, SM C16:0, SM C16:1, SM C18:0, SM C18:1, SM C20:2, SM C24:0, SM C24:1, SM C26:0, 및 SM C26:1의 조합에서 선택된 변수를 수신하는 수신부; 상기 변수에 대한 예측 점수를 산출하는 점수 산출부; 및 상기 예측 점수를 기초로 췌장암 진단 확률을 산출하는 확률 산출부를 포함하는, 청구항 1, 5 내지 7 중 어느 한 항에 따른 방법에 의해 개체의 췌장암을 진단하기 위한 장치.

발명의 설명

기술 분야

[0001] 개체의 췌장암을 진단하기 위한 정보를 제공하는 방법 및 췌장암을 진단하기 위한 키트 등에 관한 것이다.

배경 기술

[0002] 췌장암은 위장관에서 발생하는 가장 치명적인 암 중 하나로, 다른 암에 비해 치료 성적이 극히 불량하여 생존기간이 14개월에 불과하다. 2008년 국가 암 등록 통계에 의하면 췌장암은 전체 암 중 중 발생분율 9위(10만명당 8.7명 발생)를 차지한 반면, 사망분율 5위를 차지하고 있다. 우리나라 전체 췌장암 환자의 5년 생존율은 평균 7.6%로서, 중앙 의학의 지속적인 발전에 힘입어 전체 암 환자의 생존율은 꾸준히 증가추세를 보임에도 불구하고, 다른 암과 다르게 췌장암의 생존율은 지난 20여 년간 거의 향상되지 않았다. 2030년 췌장암이 암 관련 사망의 두 번째 원인이 될 것으로 예측되고 있다.

[0003] 따라서, 상기 개시한 바와 같은 사망 위험도가 높은 것은 대부분 췌장암이 진행된 후에 발견되기 때문이며, 발

견 당시 수술 절제가 가능한 경우가 20% 이내이고, 육안으로 보기에 완전히 절제되었다 하더라도 미세 전이에 의해 재발율이 높고 생존율 향상이 적으며, 항암제 및 방사선 치료에 대한 반응이 낮기 때문이다. 따라서 생존율을 향상시킬 수 있는 가장 중요한 방법은 증상이 없거나 비특이적일 때 조기 발견하여 수술하는 것이다. 그러나 췌장은 후복막에 다른 장기들에 둘러 싸여져 있고, 초기에 증상이 거의 없어 조기 진단이 매우 어렵다. 이에 췌장암을 우수한 민감도와 특이도로 간편하고 경제적으로 진단하기 위한 췌장암 진단에 대한 연구가 필요한 실정이다.

발명의 내용

해결하려는 과제

- [0004] 일 양상은 개체의 췌장암을 진단하기 위한 정보를 제공하는 방법에 관한 것이다.
- [0005] 다른 양상은 개체의 췌장암을 진단하기 위한 키트에 관한 것이다.
- [0006] 다른 양상은 상기 방법을 실행하기 위한 컴퓨터 프로그램을 기록한 컴퓨터로 판독 가능한 기록 매체에 관한 것이다.
- [0007] 또 다른 양상은 개체의 췌장암을 진단하기 위한 키트에 관한 것이다.
- [0008] 또 다른 양상은 Ala, Asn, Asp, Cit, Glu, His, Ile, Leu, Lys, Met, Orn, Phe, Pro, Thr, Trp, Tyr, Val, Creatinine, Serotonin, C0, C3, C3-DC (C4-OH), C3-OH, C3:1, C4:1, C5-M-DC, C7-DC, C8, C10:2, C12-DC, C14, C16-OH, C16:1-OH, lysoPC a C16:0, lysoPC a C18:0, lysoPC a C18:2, PC aa C26:0, PC aa C28:1, PC aa C30:2, PC aa C40:4, PC aa C42:0, PC aa C42:2, PC ae C30:0, PC ae C30:2, PC ae C32:1, PC ae C38:2, PC ae C38:4, PC ae C40:1, PC ae C40:4, PC ae C40:5, PC ae C42:3, PC ae C42:4, PC ae C42:5, PC ae C44:4, PC ae C44:5, PC ae C44:6, SM (OH) C14:1, SM (OH) C16:1, SM (OH) C22:1, SM (OH) C22:2, SM (OH) C24:1, SM C16:0, SM C16:1, SM C18:0, SM C18:1, SM C20:2, SM C24:0, SM C24:1, SM C26:0, SM C26:1, 또는 이의 조합에서 선택된 변수를 수신하는 수신부; 상기 변수에 대한 예측 점수를 산출하는 점수 산출부; 및 상기 예측 점수를 기초로 췌장암 진단 확률을 산출하는 확률 산출부를 포함하는 췌장암 진단장치에 관한 것이다.

과제의 해결 수단

- [0009] 본 발명은 개체로부터 분리된 혈액 시료에서 췌장암 환자만을 높은 정도의 특이도 및 정확도로 진단할 수 있는 방법을 발견하기 위한 연구 결과, 대사체만으로 췌장암 환자를 우수하게 진단할 수 있는 모델을 확립하고 이를 검증함으로써 개체의 췌장암을 진단하기 위한 정보를 제공하는 방법을 발명하였다.
- [0010] 일 양상은 개체의 췌장암을 진단하기 위한 정보를 제공하는 방법을 제공한다.
- [0011] 본 명세서에서 "췌장암"은 가장 흔하고 중요한 악성종양인 췌관 관세포암(pancreatic ductal adenocarcinoma)만을 포함하는 것이며, 또한 1기, 2기 등과 같이 분류되는 암의 진행단계에 따른 췌장암을 포함할 수 있다.
- [0012] 명세서에서 용어 "대사체"는 생체 기원의 시료로부터 수득한 대사물질을 말하며 바람직하게, 상기 대사체를 수득할 수 있는 생체 기원의 시료는 전혈, 혈장, 혈청, 혈소판일 수 있고, 더욱 바람직하게는 혈장일 수 있다. 상기 대사체는 대사 및 대사 과정에 의해 생산된 물질 또는 생물학적 효소 및 분자에 의한 화학적 대사작용으로 발생한 물질 등을 포함할 수 있다.
- [0013] 상기 방법은 개체의 생물학적 시료에서 n개의 대사체 시료를 수득하는 단계; 수득된 대사체 시료에서 n개의 대사체 수준을 측정하는 단계; 측정된 n개의 대사체 수준을 정규화 하여 x값을 도출하는 단계; 상기 x값을 각각 지정된 대사체 상수와 곱하여 n개의 대사체 상수의 합에 회귀계수를 더한 a값을 도출하는 단계; 및 상기 도출된 a 값을 하기 수학적식에 대입하여 개체가 췌장암 진단 예측 확률(y)을 계산하는 단계를 포함하는 방법을 포함할 수 있다.
- [0014] [수학적식]
- [0015]
$$y = \frac{1}{1 + e^{-a}}$$
- [0016] 상기 방법에 있어서, n은 1 내지 200일 수 있으나, 구체적으로 n은 70일 수 있다. 본 명세서에서는 구체적인 실시예를 통하여 70개의 대사체만으로 90%이상의 높은 정확도, 민감도 및 특이도로 효과적인 췌장암 진단이 가능

한 것을 확인하였다.

- [0017] 상기 방법에 있어서, 측정된 대사체의 수준인 x 값은 사용된 대사체들의 상호 간 측척을 동등하게 비교하기 위하여 평균 값을 제하고, 이를 표준 편차 값으로 나누어 정규화한 값을 사용할 수 있으며, 이는 각 대사체의 평균과의 차이와 표준편차의 보정을 통해서 정규화한 된 것을 의미할 수 있다.
- [0018] 상기 방법에 있어서, 상기 회귀 계수는 부분 최소제곱법 및 회소 부분 최소 제곱법으로 이루어진 군으로부터 선택되는 한 가지 이상의 회귀분석 방법으로 회귀 분석을 수행하여 계산하는 것일 수 있다.
- [0019] 상기 방법에 있어서 대사체는 상기 방법에 있어서 대사체는 Alanine (Ala), Asparagine (Asn), Aspartate (Asp), Citrulline (Cit), Glutamate (Glu), Histidine (His), Isoleucine (Ile), Leucine (Leu), Lysine (Lys), Methionine (Met), Ornithine (Orn), Phenylalanine (Phe), Proline (Pro), Threonine (Thr), Tryptophan (Trp), Tyrosine (Tyr), Valine (Val), Creatinine, Serotonin, Carnitine (C0), Propionyl-L-carnitine (C3), Malonyl-L-carnitine/Hydroxybutyryl-L-carnitine (C3-DC (C4-OH)), Hydroxypropionyl-L-carnitine (C3-OH), Propenyl-L-carnitine (C3:1), Butenyl-L-carnitine (C4:1), Methylglutaryl-L-carnitine (C5-M-DC), Pimelyl-L-carnitine (C7-DC), Octanoyl-L-carnitine (C8), Decadienyl-L-carnitine (C10:2), Dodecanedioyl-L-carnitine (C12-DC), Tetradecanoyl-L-carnitine (C14), Hydroxyhexadecanoyl-L-carnitine (C16-OH), Hydroxyhexadecenoyl-L-carnitine (C16:1-OH), lysoPhosphatidylcholine acyl C16:0 (lysoPC a C16:0), lysoPhosphatidylcholine acyl C18:0 (lysoPC a C18:0), lysoPhosphatidylcholine acyl C18:2 (lysoPC a C18:2), Phosphatidylcholine diacyl C26:0 (PC aa C26:0), Phosphatidylcholine diacyl C28:1 (PC aa C28:1), Phosphatidylcholine diacyl C30:2 (PC aa C30:2), Phosphatidylcholine diacyl C40:4 (PC aa C40:4), Phosphatidylcholine diacyl C42:0 (PC aa C42:0), Phosphatidylcholine diacyl C42:2 (PC aa C42:2), Phosphatidylcholine acyl-alkyl C30:0 (PC ae C30:0), Phosphatidylcholine acyl-alkyl C30:2 (PC ae C30:2), Phosphatidylcholine acyl-alkyl C32:1 (PC ae C32:1), Phosphatidylcholine acyl-alkyl C38:2 (PC ae C38:2), Phosphatidylcholine acyl-alkyl C38:4 (PC ae C38:4), Phosphatidylcholine acyl-alkyl C40:1 (PC ae C40:1), Phosphatidylcholine acyl-alkyl C40:4 (PC ae C40:4), Phosphatidylcholine acyl-alkyl C40:5 (PC ae C40:5), Phosphatidylcholine acyl-alkyl C42:3 (PC ae C42:3), Phosphatidylcholine acyl-alkyl C42:4 (PC ae C42:4), Phosphatidylcholine acyl-alkyl C42:5 (PC ae C42:5), Phosphatidylcholine acyl-alkyl C44:4 (PC ae C44:4), Phosphatidylcholine acyl-alkyl C44:5 (PC ae C44:5), Phosphatidylcholine acyl-alkyl C44:6 (PC ae C44:6), Hydroxysphingomyeline C14:1 (SM (OH) C14:1), Hydroxysphingomyeline C16:1 (SM (OH) C16:1), Hydroxysphingomyeline C22:1 (SM (OH) C22:1), Hydroxysphingomyeline C22:2 (SM (OH) C22:2), Hydroxysphingomyeline C24:1 (SM (OH) C24:1), Sphingomyeline C16:0 (SM C16:0), Sphingomyeline C16:1 (SM C16:1), Sphingomyeline C18:0 (SM C18:0), Sphingomyeline C18:1 (SM C18:1), Sphingomyeline C20:2 (SM C20:2), Sphingomyeline C24:0 (SM C24:0), Sphingomyeline C24:1 (SM C24:1), Sphingomyeline C26:0 (SM C26:0), Sphingomyeline C26:1 (SM C26:1) 또는 이의 조합일 수 있다.
- [0020] 상기 방법에 있어서, 생물학적 시료는 혈액, 혈장, 혈소판, 혈청, 또는 이들의 조합일 수 있다.
- [0021] 본 명세서에서 용어 "수준 측정"이란 체장암을 진단하기 위하여 생물학적 시료에서의 대사체 발현 수준을 확인하는 과정으로, 본 발명의 일 구현예에 따르면 상기 기재된 70개이상의 대사체의 수준을 확인할 수 있다. 또한 상기 방법에 있어서, 상기 대사체의 수준을 측정하는 단계는 크로마토그래피/질량 분석법을 이용하여 수행되는 것일 수 있으나, 이에 한정되는 것은 아니다.
- [0022] 본 명세서에서 용어 "진단"은 병리 상태의 존재 또는 특징을 확인하는 것을 의미한다. 본 발명의 목적상, 체장암의 발병 여부를 확인하는 것이다.
- [0023] 다른 양상은 상기 방법을 실행하기 위한 컴퓨터 프로그램을 기록한 컴퓨터로 판독 가능한 기록 매체를 제공한다.
- [0024] 또 다른 양상은 개체의 체장암을 진단하기 위한 키트를 제공한다.
- [0025] 또 다른 양상은 Ala, Asn, Asp, Cit, Glu, His, Ile, Leu, Lys, Met, Orn, Phe, Pro, Thr, Trp, Tyr, Val, Creatinine, Serotonin, C0, C3, C3-DC (C4-OH), C3-OH, C3:1, C4:1, C5-M-DC, C7-DC, C8, C10:2, C12-DC, C14, C16-OH, C16:1-OH, lysoPC a C16:0, lysoPC a C18:0, lysoPC a C18:2, PC aa C26:0, PC aa C28:1, PC aa C30:2, PC aa C40:4, PC aa C42:0, PC aa C42:2, PC ae C30:0, PC ae C30:2, PC ae C32:1, PC ae C38:2, PC

ae C38:4, PC ae C40:1, PC ae C40:4, PC ae C40:5, PC ae C42:3, PC ae C42:4, PC ae C42:5, PC ae C44:4, PC ae C44:5, PC ae C44:6, SM (OH) C14:1, SM (OH) C16:1, SM (OH) C22:1, SM (OH) C22:2, SM (OH) C24:1, SM C16:0, SM C16:1, SM C18:0, SM C18:1, SM C20:2, SM C24:0, SM C24:1, SM C26:0, SM C26:1, 또는 이의 조합에서 선택된 변수를 수신하는 수신부; 상기 변수에 대한 예측 점수를 산출하는 점수 산출부; 및 상기 예측 점수를 기초로 췌장암 진단 확률을 산출하는 확률 산출부를 포함하는 췌장암 진단장치를 제공한다.

[0026] 상기 키트는 대사체 수준을 측정하는 제제를 포함하고, 상기 대사체는 Ala, Asn, Asp, Cit, Glu, His, Ile, Leu, Lys, Met, Orn, Phe, Pro, Thr, Trp, Tyr, Val, Creatinine, Serotonin, CO, C3, C3-DC (C4-OH), C3-OH, C3:1, C4:1, C5-M-DC, C7-DC, C8, C10:2, C12-DC, C14, C16-OH, C16:1-OH, lysoPC a C16:0, lysoPC a C18:0, lysoPC a C18:2, PC aa C26:0, PC aa C28:1, PC aa C30:2, PC aa C40:4, PC aa C42:0, PC aa C42:2, PC ae C30:0, PC ae C30:2, PC ae C32:1, PC ae C38:2, PC ae C38:4, PC ae C40:1, PC ae C40:4, PC ae C40:5, PC ae C42:3, PC ae C42:4, PC ae C42:5, PC ae C44:4, PC ae C44:5, PC ae C44:6, SM (OH) C14:1, SM (OH) C16:1, SM (OH) C22:1, SM (OH) C22:2, SM (OH) C24:1, SM C16:0, SM C16:1, SM C18:0, SM C18:1, SM C20:2, SM C24:0, SM C24:1, SM C26:0, SM C26:1, 또는 이의 조합일 수 있다.

[0027] 상기 유도체의 대사체, 대사체의 수준, 이의 측정은 전술한 바와 같다.

[0028] 상기 키트는 췌장암 진단에 필요한 시료를 더 포함할 수 있다.

[0029] 일 구체예에 있어서, 상기 키트는 크로마토그래피 키트일 수 있으나, 이에 제한되는 것은 아니다.

[0030] 상기 키트는 상기 대사체 수준을 측정하기 위한 제제, 장치, 및 알고리즘이 내장된 컴퓨터를 포함할 수 있고, 상기 알고리즘을 통해 상기 대사체의 수준 측정 결과를 췌장암의 진단과 연관시키는 것인, 키트에 관한 것일 수 있다.

[0031] 상기 알고리즘은 췌장암 진단 예측 확률(y)을 계산하는 단계를 포함하는 췌장암 진단 예측 확률 모델에 의할 수 있다.

[0032] 구체적으로, 상기 키트는 대사체, 예를 들어 Ala, Asn, Asp, Cit, Glu, His, Ile, Leu, Lys, Met, Orn, Phe, Pro, Thr, Trp, Tyr, Val, Creatinine, Serotonin, CO, C3, C3-DC (C4-OH), C3-OH, C3:1, C4:1, C5-M-DC, C7-DC, C8, C10:2, C12-DC, C14, C16-OH, C16:1-OH, lysoPC a C16:0, lysoPC a C18:0, lysoPC a C18:2, PC aa C26:0, PC aa C28:1, PC aa C30:2, PC aa C40:4, PC aa C42:0, PC aa C42:2, PC ae C30:0, PC ae C30:2, PC ae C32:1, PC ae C38:2, PC ae C38:4, PC ae C40:1, PC ae C40:4, PC ae C40:5, PC ae C42:3, PC ae C42:4, PC ae C42:5, PC ae C44:4, PC ae C44:5, PC ae C44:6, SM (OH) C14:1, SM (OH) C16:1, SM (OH) C22:1, SM (OH) C22:2, SM (OH) C24:1, SM C16:0, SM C16:1, SM C18:0, SM C18:1, SM C20:2, SM C24:0, SM C24:1, SM C26:0, SM C26:1, 또는 이의 조합에서 선택된 변수를 수신하는 수신부; 상기 변수에 대한 예측 점수를 산출하는 점수 산출부; 및 상기 예측 점수를 기초로 췌장암 진단 확률을 산출하는 확률 산출부를 포함한다.

[0033] 상기 장치는 상기 수신부, 점수 산출부, 및 확률 산출부 중 적어도 하나와 연결된 출력부를 더 포함할 수 있다.

[0034] 상기 출력부는 상기 변수를 입력할 수 있는 모델을 출력하는 것일 수 있다.

[0035] 상기 장치는 웹 페이지, 어플리케이션 등을 구동하기 위한 장치로서, 예를 들어 컴퓨팅 디바이스, 모바일 디바이스, 서버 등을 포함 할 수 있다. 상기 장치는 프로세서, 저장부, 메모리, 입력부 및 출력부의 구성요소를 포함할 수 있으며, 수신부, 점수 산출부 및 확률 산출부는 장치의 구성요소들을 통해 구현될 수 있다. 서버로 구현되는 경우, 췌장암 진단 예측 장치는 산출된 값들을 출력부를 갖는 다른 디바이스로 전송하도록 구동될 수도 있다.

[0036] 구체적으로, 상기 수신부는 대사체, 예를 들어 Ala, Asn, Asp, Cit, Glu, His, Ile, Leu, Lys, Met, Orn, Phe, Pro, Thr, Trp, Tyr, Val, Creatinine, Serotonin, CO, C3, C3-DC (C4-OH), C3-OH, C3:1, C4:1, C5-M-DC, C7-DC, C8, C10:2, C12-DC, C14, C16-OH, C16:1-OH, lysoPC a C16:0, lysoPC a C18:0, lysoPC a C18:2, PC aa C26:0, PC aa C28:1, PC aa C30:2, PC aa C40:4, PC aa C42:0, PC aa C42:2, PC ae C30:0, PC ae C30:2, PC ae C32:1, PC ae C38:2, PC ae C38:4, PC ae C40:1, PC ae C40:4, PC ae C40:5, PC ae C42:3, PC ae C42:4, PC ae C42:5, PC ae C44:4, PC ae C44:5, PC ae C44:6, SM (OH) C14:1, SM (OH) C16:1, SM (OH) C22:1, SM (OH) C22:2, SM (OH) C24:1, SM C16:0, SM C16:1, SM C18:0, SM C18:1, SM C20:2, SM C24:0, SM C24:1, SM C26:0, SM C26:1, 또는 이의 조합 수준 각각에 대한 변수를 수신한다.

[0037] 상기 수신부에 출력부가 연결될 수 있고, 상기 출력부는 변수를 입력할 수 있는 입력 모델을 시각적으로 출력할

수 있다.

- [0038] 상기 점수 산출부는 상기 수신부에서 수신한 변수에 대한 예측 점수를 산출한다. 구체적으로, 점수 산출부는 상기 변수들마다 미리 결정된 값을 매칭시킬 수 있다. 예를 들어, 상기 매칭은 최저점과 최고점을 가지는 점수선의 적어도 일부에 각 변수의 측정값 범위가 매칭된 모델을 사용하여 수행될 수 있다.
- [0039] 상기 확률 산출부는 상기 점수 산출부에서 매칭된 각 변수에 대한 예측 점수를 수신하여 이를 기초로 해당 진단 예측 확률을 산출한다. 구체적으로, 확률 산출부는 상기 변수에 대한 예측 점수를 모두 합산하여 총점을 얻고, 총점과 미리 결정된 생존율을 매칭시킬 수 있다. 예를 들어, 상기 매칭은 해당 환자 진단 확률과 각 변수에 대한 예측 점수를 합산한 총점이 매칭된 모델을 사용하여 수행될 수 있다.
- [0040] 상기 모델은 해당 진단 예측 확률 모델에 의하여 결정될 수 있다.
- [0041] 상기 키트에는 알고리즘이 사용될 수 있는 있으며, 이로 제한하는 것은 아니나, 부분 최소제곱법 및 최소 부분 최소 제곱법으로 이루어진 군으로부터 선택되는 한 가지 이상이 선택될 수 있다.
- [0042] 상기 부분 최소 제곱법(partial least squares: PLS)은 변수들 간의 다양한 관계성을 지닌 원 자료 대신 상관관계가 없는 성분 집합들로 설명 변수들을 줄이고, 이러한 성분들을 가지고 최소 제곱법을 수행하는 방법이다. PLS는 변수들 간의 상관성이 높은 공선적 관계에 있거나 설명 변수의 수가 관측된 대상자의 수보다 더 많으며, 이로 인해 기존의 최소 제곱법을 사용하지 못하거나 표준 오차가 높은 계수를 생성할 경우에 효율성이 높아지는 것으로 알려져 있다. 설명 변수 간에 상관 관계를 갖고 있는 유전체 자료의 분석이나 물리 화학 속성 사이에 대한 관련 연구에서 활용도가 높은 분석 기법이다. PLS는 주성분분석(principal component analysis; PCA)과 유사한 방식으로 성분 집합을 구성한다. 변수들로부터 주성분과 유사한 잠재변수들을 찾는 다음 잠재변수들을 통해 선형 결합하게 되는데, 이 때 잠재변수들은 서로 직교하게 되어 다중공선성(multicollinearity)의 문제로부터 벗어날 수 있어 대사체와 같이 서로 상관도가 높은 변수들에 대해서 효과적으로 추정할 수 있다. PLS가 PCA와 서로 다른 점은 PCA는 설명변수 간의 분산구조를 고려한다면, PLS는 종속변수와 극대화한 상관관계를 갖도록 하는 방향으로 성분을 구성하게 되는 것으로 알려져 있다.
- [0043] PLS는 예측 모델을 개발하는데 중요성을 부여하고 반응 변수를 설명하는데 필요하지 않은 설명 변수들은 제거하지 않는다. 하지만 본 명세서에서 사용하는 최소 부분 최소 제곱법(sparse partial least squares : SPLS)은 반응 변수에 미치는 영향이 미미한 설명 변수들은 제거한 후 일정 수준 이상의 영향을 갖는 변수들만 선택하여 모형에 추가하는 것으로, 이로 인해 모형 수립 시에는 모든 설명 변수들을 필요로 하지만 모형 수립 이후에는 선택된 변수들만 가지고도 예측이 가능하므로 활용도 측면에서 장점을 지닌 것으로 알려져 있다.
- [0044] 본 명세서에서 달리 정의되지 않은 용어들은 본 발명이 속하는 기술분야에서 통상적으로 사용되는 의미를 갖는 것이다.

발명의 효과

- [0045] 일 양상에 따른 개체의 해당 진단을 위한 정보를 제공하는 방법 및 이를 이용한 키트에 따르면, 정상인 혹은 다른 질환을 앓는 환자로부터 얻은 혈액에서의 일정 대사체 수치 만으로 높은 정도의 특이도 및 진단도로 경제적으로 해당 진단을 효과적으로 진단할 수 있어 해당 환자 맞춤형 치료가 가능하다.

발명을 실시하기 위한 구체적인 내용

- [0046] 이하 실시예를 통하여 보다 상세하게 설명한다. 그러나, 이들 실시예는 하나 이상의 구체예를 예시적으로 설명하기 위한 것으로 본 발명의 범위가 이들 실시예에 한정되는 것은 아니다.

[0047] 실시예 1. 해당 진단을 위하여 혈액시료에서 혈액 대사체의 검출 및 수준 측정

- [0048] 정상인, 해당 진단을 앓고 있는 환자 및 기타 다양한 암 및 질환을 앓고 있는 환자를 314명의 혈액 시료를 총 3번에 걸쳐 수집하고, 하기 실험을 진행하였다. 본 실험은 세브란스 병원의 기관 검토의원회로부터 승인받았다. 환자가 앓고 있는 질환의 종류 및 이에 해당되는 환자의 수는 하기 표 1에 나타내었다.

표 1

	1회차 (N=157)		2회차 (N=82)		3회차 (N=75)	
	Group	N	Group	N	Group	N
	해당 진단	57	정상	38	유방암	10

담낭암	20	폐렴	12	간세포암	8
정상	20	폐혈증	32	폐선암	5
췌장낭성종양(SPN&SCN&IPMN&MCN)	20			폐편평상피세포암	5
바터 췌대부암	10			만성 췌장염	29
대장암	10			췌장 신경내분비종양	9
위암	10			감상선 유두암	9
담관암	10				

[0050] 낭종성 양성종양인 장액성 낭성 종양(serous cystadenoma; SCN), 점액성 낭성 종양(mucinous cystic neoplasm;MCN), 췌관 내 유두상 점액 종양(intraductal papillary mucinous neoplasm, IPMN), 고형 가유두상 종양(solid pseudopapillary tumor, SPN)

[0051] 상기 표 1에 나타난 바와 같이, 57명의 췌장암 환자와 정상인 및 다른 질환 환자들(257명)의 혈액 시료로부터 대사체를 검출하고, 이를 고성능 액체 크로마토그래피 (HPLC)-탠덤(tandem) 질량 분석기와 Absolute IDQTM p180 키트 (BIOCRATES Life Sciences AG, Innsbruck, Austria)를 사용, 표적화된 대사체학(metabolomics) 접근법을 기반으로 분석하였다. 대사체는 21개의 아미노산, 21개의 생체 아민(biogenic amines), 40개의 아실카르니틴(Cx:y), 93개의 글리세로인지질[14개의 리소포스파티딜콜린 (lyso PCx:y) 및 79개의 포스파티딜콜린 (PC aa x:y 또는 PC ae x:y)], 15개의 스펅고지질 (SMx:y 또는 SM (OH)x:y) 및 1개의 헥소오스이다. Cx:y는 지질 측쇄 배열을 나타내며, 여기서 x는 측쇄에서 탄소 수를 나타내고, y는 불포화 사슬의 수를 나타낸다. 시료 준비 및 분석 절차는 3 가지 품질 관리 물질의 동시 측정과 함께 제조업체의 지침에 따라 수행되었다. 시료를 Agilent 1290 시리즈 HPLC와 결합된 QTRAP 5500 질량 분석기 (SCIEX, Woodlands Central, Singapore)로 분석하였다. 그러나 상기 대사체 중 10개 대사체는 주요한 비율로 검출되지 않아 통계분석에 사용하지 않았고, 최종적으로 총 178개의 대사체를 통계분석에 사용하였으나, 최종적으로 효과적으로 췌장암 진단을 가능케 하는 70종의 대사체를 선정하고, 이를 통한 췌장암 진단을 위한 모델을 정립하였다.

[0052] 실시예 2. 췌장암의 진단을 위한 모델의 확립과 회귀 계수의 계산

[0053] 췌장암을 효과적으로 진단하기 위한 총 70종의 대사체를 활용한 췌장암 진단 모델을 수립하기 위하여 최소 부분 최소 제곱법(sparse partial least squares: SPLS)을 활용하였으며, 이의 회귀계수를 계산하였다.

[0054] 부분 최소 제곱법(partial least squares: PLS, Wold 1966)은 변수들 간의 다양한 관계성을 지닌 원 자료 대신 상관 관계가 없는 성분 집합들로 설명 변수들을 줄이고, 이러한 성분들을 가지고 최소 제곱법을 수행하는 방법으로, 전체 대사체의 회귀계수를 설정하여, 췌장암 환자를 효과적으로 진단하기 위하여 부분 최소 제곱법을 하였고, 최소 부분 최소 제곱법(sparse partial least squares: SPLS)는 반응 변수에 미치는 영향이 미비한 설명 변수들은 제거한 후 일정 수준 이상의 영향을 갖는 변수들만 선택하여 모형에 추가하게 되므로, SPLS를 수행하기 위해서 Chung & Keles (2010)가 개발한 분류 알고리즘을 사용하여, 총 70개 대사체의 회귀 계수를 위해서 아래의 식을 사용하였다.

[0055] 부분 최소 제곱법(partial least squares: PLS, Wold 1966)은 변수들 간의 다양한 관계성을 지닌 원 자료 대신 상관 관계가 없는 성분 집합들로 설명 변수들을 줄이고, 이러한 성분들을 가지고 최소 제곱법을 수행하는 방법으로, 전체 대사체의 회귀계수를 설정하여, 췌장암 환자를 효과적으로 진단하기 위하여 부분 최소 제곱법을 하였고, 최소 부분 최소 제곱법(sparse partial least squares: SPLS)는 반응 변수에 미치는 영향이 미비한 설명 변수들은 제거한 후 일정 수준 이상의 영향을 갖는 변수들만 선택하여 모형에 추가하게 되므로, SPLS를 수행하기 위해서 Chung & Keles (2010)가 개발한 분류 알고리즘을 사용하여, 총 70개 대사체의 회귀 계수($\hat{\beta}$)를 위해서 아래의 식을 사용하였다.

[0056]
$$\hat{Y} = X\hat{\beta}$$

[0057] 여기서 X는 독립변수, 즉 대사체를 의미하며 종속변수(=Y)와 독립변수는 잠재변수 T와 각각의 계수행렬(Q, P)의 곱과 오차항(F, E)의 합으로 다음과 같이 구성하였다.

[0058]
$$Y = TQ^T + F,$$

[0059]
$$X = TP^T + E,$$

[0060] 잠재변수 T 는 독립변수 X 와 K -차원의 방향 벡터 ($1 \leq K \leq \min\{n, p\}$) 들을 갖는 방향벡터(direction vector)들로 구성된 행렬 W 로 아래와 식과 같이 구성된다.

[0061] $T = XW$

[0062] 여기서 n 은 자료의 총 대상자 수이고 p 는 총 변수 수이다. 본 모델에서는 최적의 K 를 찾기 위해서 5-fold 교차 검증(cross validation)을 사용하였다.

[0063] 잠재변수를 사용한 최소제곱 추정량을 $\hat{\beta}^{LC}$ 라고 정의할 수 있으며, $T\hat{\beta}^{LC} = XW\hat{\beta}^{LC} = X\hat{\beta}$ 의 관계성을 이용하여 최종적으로 선형계수는 다음과 같이 추정된다.

[0064]
$$\hat{\beta} = W\hat{\beta}^{LC} \quad (1)$$

[0065] SPLS는 PLS와 다르게 사용되는 독립변수의 개수를 줄이는 차원축소 과정을 거치므로, 본 모델에서는 다음과 같은 최적화 문제를 풀어 모형에 선정할 최적의 변수 개수를 선택하여 최종적인 선형계수를 추정하였다.

[0066]
$$\min_{w, c} \{-\kappa w^T M w + (1 - \kappa)(c - w)^T M (c - w) + \lambda_1 \|c\|_1 + \lambda_2 \|c\|_2\},$$

subject to $w^T w = 1$, where $M = X^T Y Y^T X$. (2)

[0067] 여기서 w 는 방향벡터이며, c 는 방향벡터의 대체(surrogate)벡터로 정의한다. 또한, λ_1 과 λ_2 는 L1-노름(norm)과 L2-노름(norm)을 각각 조절하는 모수(parameter) 값이고, κ 는 회소성을 조절하는 모수이다.

[0068] 아울러, 본 연구에서 사용된 종속변수, 즉 회귀 계수는 연속형 자료가 아닌 이분형 자료이므로, logit 연결 함수를 통해서 일반화하여 추정하여 사용하였다.

[0069] 상기 분류 모형 수립 시 사용된 대사체들은 상호 간의 측척을 동등하게 비교하고자 각 구성물질의 평균 차이와 표준편차의 보정을 통해서 정규화 하여 사용하였다. 아울러, 상기 정규화한 70개의 대사체의 측정값을 지정되어 있는 각각의 대사체 상수와 곱하고, 70개 대사체의 측정값과 대사체 상수와 곱한 값을 모두 더하고, 상기 모델로 구한 회귀 계수를 더한 값(a)을 도출하였다. 이에, 도출된 a값을 하기 수학적식에 대입하여, 개체가 채장암으로 진단될 확률을 구하였다.

[0070] [수학적식]

[0071]
$$y = \frac{1}{1 + e^{-a}}$$

[0072] 아울러, (1)과 (2)의 식을 통해서 선발되어 추정된 상기 70개 대사체에 대한 각각의 지정된 상수값은 하기 표 2에 나타내었다.

표 2

[0073]

대사체	지정된 상수 값	대사체	지정된 상수 값	대사체	지정된 상수 값
Ala	0.676280223	C14	0.45309149	SM. .OH. .C24.1	-0.65556429
Asn	-0.685587685589795	C16.OH	0.37169783	SM.C16.0	-0.24091768
Asp	0.0032750441036605	C16.1.OH	0.4530083	SM.C16.1	-0.64612395
Cit	-0.383529257098579	lysoPC.a.C16.0	-0.53395448	SM.C18.0	0.33933332
Glu	0.185851714947175	lysoPC.a.C18.0	-0.45752599	SM.C18.1	-0.19860643
His	-1.73313668962738	lysoPC.a.C18.2	-1.36244982	SM.C20.2	-0.46066381
Ile	-0.665183886269878	PC.aa.C26.0	1.84041225	SM.C24.0	-0.69101574
Leu	-0.251864533741008	PC.aa.C28.1	-0.50069967	SM.C24.1	-0.19607839
Lys	0.41570457	PC.aa.C30.2	1.76119558	SM.C26.0	-0.88196312
Met	-0.49400673	PC.aa.C40.4	0.52173427	SM.C26.1	-0.17657509

Orn	-0.4925206	PC.aa.C42.0	0.28186993	Intercept	-10.4432713170571
Phe	-0.36327781	PC.aa.C42.2	-0.10879668		
Pro	-0.82328267	PC.ae.C30.0	-0.6777565		
Thr	-0.63381162	PC.ae.C30.2	0.30128848		
Trp	-0.10139695	PC.ae.C32.1	1.90819002		
Tyr	-0.62060258	PC.ae.C38.2	-0.20792852		
Val	0.27740928	PC.ae.C38.4	0.03538989		
Creatinine	-0.40429159	PC.ae.C40.1	-0.16489253		
Serotonin	1.23877177	PC.ae.C40.4	0.00505762		
C0	-0.69693727	PC.ae.C40.5	0.42026535		
C3	-1.04096984	PC.ae.C42.3	0.24801568		
C3.DC..C4.0H.	0.01366621	PC.ae.C42.4	-0.48295906		
C3.OH	-0.52241362	PC.ae.C42.5	0.03425526		
C3.1	0.6915982	PC.ae.C44.4	-0.62184863		
C4.1	-0.23422694	PC.ae.C44.5	-0.52755386		
C5.M.DC	1.00559787	PC.ae.C44.6	-0.17685685		
C7.DC	0.66469837	SM..OH..C14.1	-0.09675213		
C8	-0.11625818	SM..OH..C16.1	0.08137913		
C10.2	0.16734288	SM..OH..C22.1	-0.76483204		
C12.DC	-0.54757817	SM..OH..C22.2	-0.48694676		

[0074] 실시예 3. 췌장암의 진단을 위한 모델의 검증

[0075] 상기 실시예 2에서 확립한 모델을 검증하기 위하여, 분석 도구를 R package, version 3.5.1 (<http://www.R-project.org>)로 하고, 부트스트랩(Bootstrapping) 1,000번을 통한 internal-validation을 수행하였으며, 오차행렬(confusion matrix) 결과를 통해 분류능력을 평가하는 도구로 활용하였다.

[0076] 상기 1000번의 부트스트랩 결과를 확인한 결과, 평균적으로 정확도가 0.98(0.952, 1)이었으며, 평가자가 다른 경우에 우연히 일치할 확률을 제외한 나머지의 확률로 일치성 확인 척도인 kappa 값이 0.933(0.826, 1)이었고, 민감도가 0.959(0.922, 0.988)이었고, 특이도가 0.992(0.973, 1)인 것을 확인하였다(괄호 안은 95% 신뢰구간). 밸런스 된 정확도(Balanced Accuracy)는 0.9484인 것을 확인하였다.

[0077] 따라서, 1000번의 부트스트랩 결과를 통하여, 평균적으로 0.95가 넘는 높은 정확도, 민감도, 특이도가 확인되어 본 확률 확인모델이 효과적인 췌장암 진단이 가능함을 확인하였다.

[0078] 실험예 1. 췌장암의 진단을 가능하게 하는 모델의 검증

[0079] 1.1 178개의 대사체를 통한 모델의 검증

[0080] 정상인, 췌장암을 앓고 있는 환자 및 기타 다양한 암 및 질환을 앓고 있는 환자 314명의 혈액 시료를 총 3번에 걸쳐 수집하고, 이의 대사체 178개를 활용하여, 검증을 수행하였다.

표 3

[0081]

		시료	
		0	1
예측	0	256	1
	1	1	56

[0082] 상기 표 3에서 나타난 바와 같이, 95% 신뢰구간 (0.9772, 0.9992)에서 정확도가 0.9936 이었으며, 평가자가 다른 경우에 우연히 일치할 확률을 제외한 나머지의 확률로 일치성 확인 척도인 kappa 값이 0.9786 이었고, 민감도가 0.9825 (0.922, 0.988)이었고, 특이도가 0.9961이며, 밸런스 된 정확도(Balanced Accuracy)는 0.9893인 것을 확인하였다.

[0083] 따라서, 평균적으로 0.95가 넘는 높은 정확도, 민감도, 특이도가 확인되어 본 확률 확인모델이 효과적으로 췌장암 진단이 가능함을 확인하였다.

[0084] 1.2 70개의 대사체를 통한 모델의 검증

[0085] 정상인, 췌장암을 앓고 있는 환자 및 기타 다양한 암 및 질환을 앓고 있는 환자 314명의 혈액 시료를 총 3번에 걸쳐 수집하고, 이의 대사체 중 70개를 활용하여, 검증을 수행하였다.

표 4

[0086]

		시료	
		0	1
예측	0	253	5
	1	4	52

[0087] 상기 표 3에서 나타난 바와 같이, 95% 신뢰구간 (0.9463, 0.9868)에서 정확도가 0.9713이었으며, 평가자가 다른 경우에 우연히 일치할 확률을 제외한 나머지의 확률로 일치성 확인 척도인 kappa 값이 0.9029 이었고, 민감도가 0.9123이었고, 특이도가 0.9844이며, 밸런스 된 정확도(Balanced Accuracy)는 0. 9484인 것을 확인하였다.

[0088] 따라서, 평균적으로 90%가 넘는 높은 정확도, 민감도, 특이도가 확인되어 본 확률 확인모델이 178개의 전체 대사체가 아닌 70개의 대사체만으로도 높은 정도의 정확도, 민감도 및 특이도가 확인되어 시간 및 경제적으로 우수한 췌장암 진단을 가능케 하는 진단모델임을 확인할 수 있었다.