



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0116931  
(43) 공개일자 2022년08월23일

(51) 국제특허분류(Int. Cl.)  
C12Q 1/6886 (2018.01)

(52) CPC특허분류  
C12Q 1/6886 (2022.01)  
C12Q 2600/112 (2013.01)

(21) 출원번호 10-2021-0020401  
(22) 출원일자 2021년02월16일  
심사청구일자 2021년02월16일

(71) 출원인  
연세대학교 원주산학협력단  
강원도 원주시 흥업면 연세대길 1  
연세대학교 산학협력단  
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)

(72) 발명자  
이혜영  
강원도 원주시 흥업면 연세대길 1 미래관 205호  
황다솜  
강원도 원주시 흥업면 연세대길 1 미래관 216호

(74) 대리인  
구현서

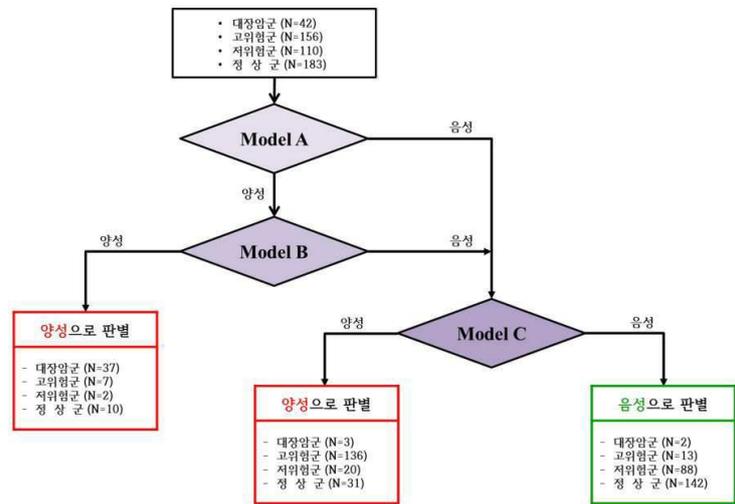
전체 청구항 수 : 총 16 항

(54) 발명의 명칭 **대장암 및 대장용종 선별 방법 및 그 응용**

(57) 요약

본 발명은 a)샘플을 중합효소 연쇄반응을 통하여 프라이머 및 프로브를 이용하여 IL1B, LTF, TNF, ITIH4, CXCL11, MAPK6, GK, 및 CD146 유전자의 상대적 발현량을 측정하고, b) 상기 a)에서 대장암 양성인 샘플을 대상으로 중합효소 연쇄반응을 통하여 프라이머 및 프로브를 이용하여 CES1, IL1B, TNF, ITIH4, PTGS2, CXCL11, MAPK6, GK, 및 CD146 유전자의 상대적 발현량을 측정하여 양성인 경우에는 대장암 양성으로 판단하고, c)상기 a) 및 b)에서 음성으로 판단받은 샘플을 중합효소 연쇄반응을 통하여 프라이머 및 프로브를 이용하여 IL1B, LTF, TNF, ITIH4, CXCL11, 및 MAPK6 유전자의 상대적 발현량을 측정하여 양성인 경우에는 대장암 양성으로 판단하는 것을 특징으로 하는 대장암 및 대장용종 선별 방법 및 그 방법에 사용되는 키트에 관한 것이다.

대표도 - 도11



(52) CPC특허분류  
C12Q 2600/158 (2013.01)

---

## 명세서

### 청구범위

#### 청구항 1

IL1B, LTF, TNF, ITIH4, CXCL11, MAPK6, GK, 및 CD146 유전자의 상대적 발현량을 측정하는데 관여하는 프라이머 또는 프로브를 유효성분으로 포함하는 대장암군과 대장암 고위험군, 저위험군 및 정상군을 구분할 수 있는 조성물.

#### 청구항 2

제1항에 있어서,

상기 프라이머 및 프로브는 서열번호 7 내지 9, 서열번호 13 내지 15, 서열번호 16 내지 18, 서열번호 34 내지 36, 서열번호 40 내지 42, 서열번호 43 내지 45, 서열번호 52 내지 54에 기재된 서열로 이루어진 것을 특징으로 하는 조성물.

#### 청구항 3

제1항에 있어서,

상기 정상군은 대장내시경 검사의 결과를 통해 대장 내 병변이 없는 경우를, 저위험군은 저위험 선종이 3개 미만 존재하는 경우를, 고위험군은 대장내시경 검사를 통하여 저위험 선종이 3개 이상, 고위험 선종이 1개 이상, 및 상피내암인 경우를 포함하는 것을 특징으로 하는 조성물.

#### 청구항 4

CES1, IL1B, TNF, ITIH4, PTGS2, CXCL11, MAPK6, GK, 및 CD146 유전자의 상대적 발현량을 측정하는데 관여하는 프라이머 또는 프로브를 유효성분으로 포함하는 대장암군과 대장암 고위험군, 저위험군 및 정상군을 구분할 수 있는 조성물.

#### 청구항 5

제4항에 있어서,

상기 프라이머 및 프로브는 서열번호 4 내지 6, 서열번호 7 내지 9, 서열번호 16 내지 18, 서열번호 22 내지 24, 서열번호 31 내지 33, 서열번호 34 내지 36, 서열번호 40 내지 42, 서열번호 43 내지 45, 및 서열번호 52 내지 54에 기재된 서열로 이루어진 것을 특징으로 하는 조성물.

#### 청구항 6

IL1B, LTF, TNF, ITIH4, CXCL11, 및 MAPK6 유전자의 상대적 발현량을 측정하는데 관여하는 프라이머 또는 프로브를 유효성분으로 포함하는 대장암 고위험군과 저위험군 및 정상군을 구분할 수 있는 조성물.

#### 청구항 7

제6항에 있어서,

상기 프라이머 및 프로브는 서열번호 7 내지 9, 서열번호 13 내지 15, 서열번호 16 내지 18, 서열번호 22 내지 24, 서열번호 34 내지 36, 및 서열번호 40 내지 42에 기재된 서열로 이루어진 것을 특징으로 하는 조성물.

#### 청구항 8

a) 샘플을 중합효소 연쇄반응을 통하여 프라이머 및 프로브를 이용하여 IL1B, LTF, TNF, ITIH4, CXCL11, MAPK6, GK, 및 CD146 유전자의 상대적 발현량을 측정하고,

b) 상기 a)에서 대장암 양성인 샘플을 대상으로 중합효소 연쇄반응을 통하여 프라이머 및 프로브를 이용하여 CES1, IL1B, TNF, ITIH4, PTGS2, CXCL11, MAPK6, GK, 및 CD146 유전자의 상대적 발현량을 측정하여 양성인

경우에는 대장암 양성으로 판단하고,

c)상기 a) 및 b)에서 음성으로 판단받은 샘플을 중합효소 연쇄반응을 통하여 프라이머 및 프로브를 이용하여 IL1B, LTF, TNF, ITIH4, CXCL11, 및 MAPK6 유전자의 상대적 발현량을 측정하여 양성인 경우에는 대장암 양성으로 판단하는 것을 특징으로 하는 대장암 및 대장용종 선별 방법.

#### 청구항 9

제8항에 있어서,

상기 a)에 사용된 프라이머 및 프로브는 서열번호 7 내지 9, 서열번호 13 내지 15, 서열번호 16 내지 18, 서열번호 34 내지 36, 서열번호 40 내지 42, 서열번호 43 내지 45, 서열번호 52 내지 54에 기재된 서열로 이루어진 것을 특징으로 하는 대장암 및 대장용종 선별 방법.

#### 청구항 10

제8항에 있어서,

상기 b)에 사용된 프라이머 및 프로브는 서열번호 4 내지 6, 서열번호 7 내지 9, 서열번호 16 내지 18, 서열번호 22 내지 24, 서열번호 31 내지 33, 서열번호 34 내지 36, 서열번호 40 내지 42, 서열번호 43 내지 45, 및 서열번호 52 내지 54에 기재된 서열로 이루어진 것을 특징으로 대장암 및 대장용종 선별 방법.

#### 청구항 11

제8항에 있어서,

상기 c)에 사용된 프라이머 및 프로브는 서열번호 7 내지 9, 서열번호 13 내지 15, 서열번호 16 내지 18, 서열번호 22 내지 24, 서열번호 34 내지 36, 및 서열번호 40 내지 42에 기재된 서열로 이루어진 것을 특징으로 대장암 및 대장용종 선별 방법.

#### 청구항 12

a) IL1B, LTF, TNF, ITIH4, CXCL11, MAPK6, GK, 및 CD146 유전자의 상대적 발현량을 측정하는데 관여하는 프라이머 또는 프로브 세트,

b) CES1, IL1B, TNF, ITIH4, PTGS2, CXCL11, MAPK6, GK, 및 CD146 유전자의 상대적 발현량을 측정하는데 관여하는 프라이머 또는 프로브 세트, 및

c) IL1B, LTF, TNF, ITIH4, CXCL11, 및 MAPK6 유전자의 상대적 발현량을 측정하는데 관여하는 프라이머 또는 프로브 세트를 포함하는 대장암 및 대장용종 선별 키트.

#### 청구항 13

제12항에 있어서,

상기 a)의 프라이머 및 프로브 세트는 서열번호 7 내지 9, 서열번호 13 내지 15, 서열번호 16 내지 18, 서열번호 34 내지 36, 서열번호 40 내지 42, 서열번호 43 내지 45, 서열번호 52 내지 54에 기재된 서열로 이루어진 것을 특징으로 하는 대장암 및 대장용종 선별 키트.

#### 청구항 14

제12항에 있어서,

상기 b)의 프라이머 및 프로브 세트는 서열번호 4 내지 6, 서열번호 7 내지 9, 서열번호 16 내지 18, 서열번호 22 내지 24, 서열번호 31 내지 33, 서열번호 34 내지 36, 서열번호 40 내지 42, 서열번호 43 내지 45, 및 서열번호 52 내지 54에 기재된 서열로 이루어진 것을 특징으로 하는 대장암 및 대장용종 선별 키트.

#### 청구항 15

제12항에 있어서,

상기 c)의 프라이머 및 프로브 세트는 서열번호 7 내지 9, 서열번호 13 내지 15, 서열번호 16 내지 18, 서열번호

호 22 내지 24, 서열번호 34 내지 36, 및 서열번호 40 내지 42에 기재된 서열로 이루어진 것을 특징으로 하는 대장암 및 대장용종 선별 키트.

**청구항 16**

제12항에 있어서,

상기 키트는 중합효소, dNTP(dATP, dCTP, dGTP 및 dTTP) 혼합물, 완충액(buffer solution) 및 증류수를 더욱 포함하는 것을 특징으로 하는 대장암 및 대장용종 선별 키트.

**발명의 설명**

**기술 분야**

[0001] 본 발명은 혈액을 이용한 대장암 및 대장용종 선별을 위한 혈액기반 분자진단 방법 및 그 방법에 사용되는 키트에 관한 발명이다.

**배경 기술**

[0003] 대장암이란 대장을 구성하는 결장과 직장에 생기는 악성종양이며 대부분은 대장 점막세포에서 발생하는 선암으로 선종성 용종에서 유래한다고 알려져 있다.

[0004] 2016년 국립암센터의 암등록통계에 의하면 대장암은 전체 암 발생의 12.3%로 두번째로 가장 많이 발병하는 암이고, 전체 암사망자 중 11.1%로 3위를 차지하는 주요 암이다.

[0005] 대장암의 5년 상대 생존율은 암 진행 정도에 따라 현저하게 낮아져 SEER (The Surveillance, Epidemiology and End Results) 병기 분류에 의한 국한된 병기인 경우 5년 상대생존율이 95.3%에 달하지만 원격전이가 나타난 병기에서는 5년 생존율이 20.0%로 급격하게 감소한다. 하지만 초기 단계에는 증상이 나타나지 않기 때문에 정기적인 검진을 통한 조기발견이 중요하다. 또한 대부분의 대장암은 장기간의 선종-암화과정 (Adenoma-carcinoma Sequence)을 거쳐 발생하게 되므로 대장암을 예방하기 위해서는 선종 단계에서의 발견과 제거가 중요하다.

[0006] 현재 우리나라에서 국가적으로 시행되는 암 검진 프로그램은 50세 이상의 남녀를 대상으로 1년의 주기마다 분변 잠혈반응검사를 실시하고, 분변잠혈반응검사 결과 이상 소견이 있는 경우 대장 내시경 검사, 또는 대장 이중 조영 검사를 진행한다.

[0007] 대장 내시경은 한번에 전체 대장을 검사할 수 있으며 검사 중 발견되는 선종이나 일부 조기 암을 제거할 수 있다. 대장 내시경은 대장 용종을 가장 효과적으로 찾아내고 제거까지 할 수 있는 검사이다. 메타분석을 통해 대장내시경의 시행으로 대장암의 발병률 및 사망률 모두 약 70% 가량 감소함을 확인한 결과가 있다. [Brenner H et al. *Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: systemic review and meta-analysis of randomized controlled trials and observational studies*, BMJ. 2014;348:g2467]

[0008] 대장 내시경 검사는 검사 전 장정결 과정이 필요하며 장 정결 정도가 검사의 정확도와 질에 매우 중요한 영향을 미친다. 합병증으로 장천공이 약 0.09% 빈도로 나타난다. 대규모 인구집단을 대상으로 하는 대장암 선별검사로서는 시행할 수 있는 의사 수의 제한, 검사 시 통증, 전처치의 불편감으로 인해 대상자의 순응도가 떨어질 수 있고 대장벽의 두께가 얇기 때문에 천공이나 출혈 등의 합병증이 비교적 흔히 발생할 수 있다는 문제점이 있다. [Yeol K. *Colorectal Cancer Screening*, Korean J Fam Pract 2013;3:139-146]

[0009] 대장 이중 조영 검사는 직장에 가는 관을 넣고 공기와 고밀도 바륨을 주입하여 대장벽을 바륨으로 코팅하여 대장 내부를 관찰하는 영상법이다. 대부분의 연구에서 대장 이중 조영 검사의 대장암 진단 민감도는 85.0-97.0%로 보고하고 있다. 대장내시경 검사에 비해 검사 중 고통이 비교적 덜하여 수면유도제나 진통제가 필요치 않으며, 전체 대장 상태를 확인할 수 있다는 장점이 있지만 대장 내시경에 비해 상대적으로 낮은 검사의 정확도를 나타내며 정확한 검사를 위해 장을 비우는 것이 필요하고 방사선 노출 위험이 있다는 단점이 있다.

[0010] 분변 잠혈 반응 검사는 대장 내 종괴에서 스며나오는 출혈을 대변에서 감지하여 진단하는 방법이다. 과거에는 주로 guaiac 검사법을 사용하였으나, 1970년대에 면역화학검사법이 개발되어 이용되고 있다. 비침습적인 방법으로 특별한 부작용이 없으며 비교적 시행이 쉽고 비용이 저렴한 장점이 있다. 분변잠혈검사를 시행하지 않은 사

람에 비해 분변잠혈검사를 시행한 사람은 사망률이 10-40% 가량 더 낮게 나타난다. [The IARC Perspective on Colorectal Cancer Screening, 2018]

- [0011] 메타분석에 의하면 면역화학검사의 대장암 및 진생선종에 대한 민감도는 각각 56-74% 와 23-31% 이고, 특이도는 90-95%로 나타났다. [Diagnostic performance of flexible sigmoidoscopy combined with fecal immunochemical test in colorectal cancer screening: meta-analysis and modeling, Eur J Epidemiol. 2017;32(6):481-493]
- [0012] 대장 종양의 출혈은 간헐적인 경우가 많기 때문에 분변잠혈검사를 위한 검체는 3회 연속적인 배변에서 한번씩 세 번 검체를 채취해야 한다. 그러므로 적절한 검체 채취 여부에 따라서 검사의 정확도에 차이가 날 수 있고 [Yeol K. Colorectal Cancer Screening, Korean J Fam Pract 2013;3:139-146]] 대변 검체를 이용함에 있어서 검사 대상자들의 순응도가 낮게 나타날 수 있다.
- [0013] 따라서 대장내시경이나 대장이중조영검사에 비해 검사 중 위험도가 낮고 검체 채취가 용이한 혈액 검체를 이용하여 선별검사로써 유용하도록 민감도와 특이도를 높인 분자진단 검사법의 개발이 필요하다.
- [0014] [선행 특허 문헌]
- [0015] 미국 특허공개번호 20180238893

**발명의 내용**

**해결하려는 과제**

- [0016] 본 발명은 상기의 문제점을 해결하고 상기의 필요성에 의해 안출된 것으로서 본 발명의 목적은 비교적 추출이 용이한 혈액 검체를 기반으로 역전사 정량적 중합반응과 그 결과를 통해 제작한 인공지능 예측 모델을 이용하여 민감도와 특이도가 높은 대장암 및 대장용종의 분자진단 검사법을 개발하기 위한 정보제공방법을 제공하는 것이다.
- [0017] 본 발명의 다른 목적은 비교적 추출이 용이한 혈액 검체를 기반으로 역전사 정량적 중합반응과 그 결과를 통해 제작한 인공지능 예측 모델을 이용하여 민감도와 특이도가 높은 대장암 및 대장용종의 분자진단 검사용 조성물을 제공하는 것이다.

**과제의 해결 수단**

- [0019] 상기 목적을 달성하기 위하여 본 발명은 IL1B, LTF, TNF, ITIH4, CXCL11, MAPK6, GK, 및 CD146 유전자의 상대적 발현량을 측정하는데 관여하는 프라이머 또는 프로브를 유효성분으로 포함하는 대장암군과 대장암 고위험군, 저위험군 및 정상군을 구분할 수 있는 조성물을 제공한다.
- [0020] 본 발명의 일 구현예에 있어서,
- [0021] 상기 프라이머 및 프로브는 서열번호 7 내지 9, 서열번호 13 내지 15, 서열번호 16 내지 18, 서열번호 34 내지 36, 서열번호 40 내지 42, 서열번호 43 내지 45, 서열번호 52 내지 54에 기재된 서열로 이루어진 것이 바람직하나 이에 한정되지 아니한다.
- [0022] 본 발명의 다른 구현예에 있어서,
- [0023] 상기 정상군은 대장내시경 검사의 결과를 통해 대장 내 병변이 없는 경우를, 저위험군은 저위험 선종이 3개미만 존재하는 경우를, 고위험군은 대장내시경 검사를 통하여 저위험 선종이 3개 이상, 고위험 선종이 1개 이상, 및 상피내암인 경우를 포함하는 것이 바람직하나 이에 한정되지 아니한다.
- [0024] 또 본 발명은 CES1, IL1B, TNF, ITIH4, PTGS2, CXCL11, MAPK6, GK, 및 CD146 유전자의 상대적 발현량을 측정하는데 관여하는 프라이머 또는 프로브를 유효성분으로 포함하는 대장암군과 대장암 고위험군, 저위험군 및 정상군을 구분할 수 있는 조성물을 제공한다.
- [0025] 본 발명의 일 구현예에 있어서,
- [0026] 상기 프라이머 및 프로브는 서열번호 4 내지 6, 서열번호 7 내지 9, 서열번호 16 내지 18, 서열번호 22 내지 24, 서열번호 31 내지 33, 서열번호 34 내지 36, 서열번호 40 내지 42, 서열번호 43 내지 45, 및 서열번호 52

내지 54에 기재된 서열로 이루어진 것이 바람직하나 이에 한정되지 아니한다.

- [0027] 또 본 발명은 IL1B, LTF, TNF, ITIH4, CXCL11, 및 MAPK6 유전자의 상대적 발현량을 측정하는데 관여하는 프라이머 또는 프로브를 유효성분으로 포함하는 대장암 고위험군과 저위험군 및 정상군을 구분할 수 있는 조성물을 제공한다.
- [0028] 본 발명의 일 구현예에 있어서,
- [0029] 상기 프라이머 및 프로브는 서열번호 7 내지 9, 서열번호 13 내지 15, 서열번호 16 내지 18, 서열번호 22 내지 24, 서열번호 34 내지 36, 및 서열번호 40 내지 42에 기재된 서열로 이루어진 것이 바람직하나 이에 한정되지 아니한다.
- [0030] 또 본 발명은 a)샘플을 중합효소 연쇄반응을 통하여 프라이머 및 프로브를 이용하여 IL1B, LTF, TNF, ITIH4, CXCL11, MAPK6, GK, 및 CD146 유전자의 상대적 발현량을 측정하고,
- [0031] b) 상기 a)에서 대장암 양성인 나온 샘플을 대상으로 중합효소 연쇄반응을 통하여 프라이머 및 프로브를 이용하여 CES1, IL1B, TNF, ITIH4, PTGS2, CXCL11, MAPK6, GK, 및 CD146 유전자의 상대적 발현량을 측정하여 양성인 경우에는 대장암 양성으로 판단하고,
- [0032] c)상기 a) 및 b)에서 음성으로 판단받은 샘플을 중합효소 연쇄반응을 통하여 프라이머 및 프로브를 이용하여 IL1B, LTF, TNF, ITIH4, CXCL11, 및 MAPK6 유전자의 상대적 발현량을 측정하여 양성인 경우에는 대장암 양성으로 판단하는 것을 특징으로 하는 대장암 및 대장용종 선별 방법을 제공한다.
- [0033] 본 발명의 일 구현예에 있어서, 상기 a)에 사용된 프라이머 및 프로브는 서열번호 7 내지 9, 서열번호 13 내지 15, 서열번호 16 내지 18, 서열번호 34 내지 36, 서열번호 40 내지 42, 서열번호 43 내지 45, 서열번호 52 내지 54에 기재된 서열로 이루어진 것이 바람직하고,
- [0034] 상기 b)에 사용된 프라이머 및 프로브는 서열번호 4 내지 6, 서열번호 7 내지 9, 서열번호 16 내지 18, 서열번호 22 내지 24, 서열번호 31 내지 33, 서열번호 34 내지 36, 서열번호 40 내지 42, 서열번호 43 내지 45, 및 서열번호 52 내지 54에 기재된 서열로 이루어진 것이 바람직하며,
- [0035] 상기 c)에 사용된 프라이머 및 프로브는 서열번호 7 내지 9, 서열번호 13 내지 15, 서열번호 16 내지 18, 서열번호 22 내지 24, 서열번호 34 내지 36, 및 서열번호 40 내지 42에 기재된 서열로 이루어진 것이 바람직하나 이에 한정되지 아니한다.
- [0036] 또한 본 발명은 a) IL1B, LTF, TNF, ITIH4, CXCL11, MAPK6, GK, 및 CD146 유전자의 상대적 발현량을 측정하는데 관여하는 프라이머 또는 프로브 세트,
- [0037] b) CES1, IL1B, TNF, ITIH4, PTGS2, CXCL11, MAPK6, GK, 및 CD146 유전자의 상대적 발현량을 측정하는데 관여하는 프라이머 또는 프로브 세트, 및
- [0038] c) IL1B, LTF, TNF, ITIH4, CXCL11, 및 MAPK6 유전자의 상대적 발현량을 측정하는데 관여하는 프라이머 또는 프로브 세트를 포함하는 대장암 및 대장용종 선별 키트를 제공한다.
- [0039] 본 발명의 일 구현예에 있어서, 상기 a)의 프라이머 및 프로브 세트는 서열번호 7 내지 9, 서열번호 13 내지 15, 서열번호 16 내지 18, 서열번호 34 내지 36, 서열번호 40 내지 42, 서열번호 43 내지 45, 서열번호 52 내지 54에 기재된 서열로 이루어진 것이 바람직하고,
- [0040] 상기 b)의 프라이머 및 프로브 세트는 서열번호 4 내지 6, 서열번호 7 내지 9, 서열번호 16 내지 18, 서열번호 22 내지 24, 서열번호 31 내지 33, 서열번호 34 내지 36, 서열번호 40 내지 42, 서열번호 43 내지 45, 및 서열번호 52 내지 54에 기재된 서열로 이루어진 것이 바람직하며,
- [0041] 상기 c)의 프라이머 및 프로브 세트는 서열번호 7 내지 9, 서열번호 13 내지 15, 서열번호 16 내지 18, 서열번호 22 내지 24, 서열번호 34 내지 36, 및 서열번호 40 내지 42에 기재된 서열로 이루어진 것이 바람직하나 이에 한정되지 아니한다.
- [0042] 본 발명에 사용된 대장암 및 대장용종 선별검사용 혈중 바이오마커의 풀 네임과 약어는 아래와 같다.
- [0043] -C-C motif chemokine receptor 1 (CCR1)
- [0044] -Carboxylesterase 1 (CES1)

- [0045] -Interleukin 1 beta (IL1B)
- [0046] -Integrin alpha 2 (ITGA2)
- [0047] -Lactotransferrin (LTF)
- [0048] -Tumor necrosis factor superfamily 13b (TNF)
- [0049] -Prostaglandin E synthase (PTGES)
- [0050] -Inter-alpha-trypsin inhibitor heavy chain H4 (ITI4)
- [0051] -Taurine upregulated gene1 (TUG1)
- [0052] -Nucleoside diphosphate kinase 1 (NME1)
- [0053] -Prostaglandin-endoperoxide synthase 2 (PTGS2)
- [0054] -C-X-C motif chemokine 11 (CXCL11)
- [0055] -Mitogen-activated protein kinase 6 (MAPK6)
- [0056] -Glycerol kinase (GK)
- [0057] -Cytokeratin 19 (CK19)
- [0058] -Epithelial cell adhesion molecule (EpCAM)
- [0059] -Cluster of differentiation 146 (CD146)
- [0060] 이하 본 발명을 설명한다.
- [0061] 본 발명에서는 혈액 내 해당 바이오 마커들의 상대적 발현양을 나타내기 위해 프라이머 및 프로브 서열을 제공한다.
- [0062] 또한 본 발명은 상기 18개 마커의 발현양을 대입하여 제작한 대장암 및 대장용종 선별검사용 인공지능 예측 모델을 제공한다.
- [0063] 일반적으로 사용되는 전장 RNA (Total RNA)를 분리하는 방법 및 이로부터 cDNA를 합성하는 방법은 공지된 방법을 통해 수행될 수 있으며, 이 과정에 대한 자세한 설명은 Joseph Sambrook 등, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2001); 및 Noonan, K.F. 등에 개시되어 있어 본 발명의 참조로서 삽입될 수 있다.
- [0064] 본 발명의 프라이머는 포스포르아미다이트 고체 지지체 방법, 또는 기타 널리 공지된 방법을 사용하여 화학적으로 합성할 수 있다. 이러한 핵산 서열은 또한 당해 분야에 공지된 많은 수단을 이용하여 변형시킬 수 있다.
- [0065] 이러한 변형의 비제한적인 예로는 메틸화, "캡화", 천연 클레오타이드 하나 이상의 동족체로의 치환, 및 뉴클레오타이드 간의 변형, 예를 들면, 하전되지 않은 결체(예: 메틸 포스포네이트, 포스포트리에스테르, 포스포로아미데이트, 카바메이트 등) 또는 하전된 결체(예: 포스포로티오에이트, 포스포로디티오에이트 등)로의 변형이 있다. 핵산은 하나 이상의 부가적인 공유 결합된 잔기, 예를 들면, 단백질(예: 뉴클레아제, 독소, 항체, 시그널 펩타이드, 리-L-리신 등), 삽입제(예: 아크리딘, 프소랄렌 등), 킬레이트화제(예: 금속, 방사성 금속, 철, 산화성 금속 등), 및 알킬화제를 함유할 수 있다.
- [0066] 본 발명의 핵산 서열은 또한 검출 가능한 시그널을 직접 또는 간접적으로 제공할 수 있는 표지를 이용하여 변형시킬 수 있다. 표지의 예로는 방사성 동위원소, 형광성 분자, 바이오틴 등이 있다.
- [0067] 본 발명의 방법에 있어서, 상기 증폭된 표적 서열(CCR1, 및 GAPDH 유전자 등)은 검출가능한 표지 물질로 표지될 수 있다. 일 구현예에서, 상기 표지 물질은 형광, 인광, 화학발광단 또는 방사성을 발하는 물질일 수 있으나, 이에 제한되지 않는다. 바람직하게는, 상기 표지 물질은 플루오리신(fluorescein), 피코에리트린(phycoerythrin), 로다민, 리사민(lissamine) Cy-5 또는 Cy-3일 수 있다. 표적 서열의 증폭시 프라이머의 5'-말단 및/또는 3' 말단에 Cy-5 또는 Cy-3를 표지하여 RT-PCR을 수행하면 표적 서열이 검출 가능한 형광 표지 물질로 표지될 수 있다.
- [0068] 또한, 방사성 물질을 이용한 표지는 RT-PCR 수행시 32P 또는 35 S 등과 같은 방사성 동위원소를 PCR 반응액에

첨가하면 증폭 산물이 합성되면서 방사성이 증폭 산물에 혼입되어 증폭 산물이 방사성으로 표지될 수 있다. 표적서열을 증폭하기 위해 이용된 하나 이상의 올리고뉴클레오타이드 프라이머 세트를 이용할 수 있다.

- [0069] 표지는 당업계에서 통상적으로 실시되는 다양한 방법, 예컨대, 닉 트랜스레이션 (nick translation) 방법, 무작위 프라이밍 방법 (Multiprime DNA labelling systems booklet, "Amersham"(1989)) 및 카이네이션 방법 (Maxam & Gilbert, Methods in Enzymology, 65:499(1986))을 통해 실시될 수 있다. 표지는 형광, 방사능, 발색 측정, 중량 측정, X-선 회절 또는 흡수, 자기, 효소적 활성, 매스 분석, 결합 친화도, 혼성화 고주파, 나노크리스탈에 의하여 검출할 수 있는 시그널을 제공한다.
- [0070] 본 발명의 한 측면에 따르면, 본 발명에서는 RT-PCR을 통해 mRNA 수준에서 발현수준을 측정하게 된다. 이를 위하여 상기 CCR1, 및 GAPDH 유전자 등에 특이적으로 결합하는 신규한 프라이머 쌍과 형광이 표지된 프로브가 요구되며, 본 발명에서 특정한 염기서열로 특정된 해당 프라이머 및 프로브를 사용할 수 있으나 이에 제한되는 것은 아니며, 이들 유전자에 특이적으로 결합하여 검출가능한 시그널을 제공하여 RT-PCR을 수행할 수 있는 것이라면, 제한 없이 사용될 수 있다. 상기에서 FAM과 Quen(Quencher)는 형광염료를 의미한다.
- [0071] 본 발명에 적용되는 RT-PCR 방법은 당업계에서 통상적으로 사용되는 공지의 과정을 통해 수행될 수 있다.
- [0072] mRNA 발현수준을 측정하는 단계는 통상의 mRNA 발현수준을 측정할 수 있는 방법이면 제한 없이 사용될 수 있으며, 사용한 프로브 표지의 종류에 따라 방사성 측정, 형광 측정 또는 인광 측정을 통해 수행될 수 있으나, 이에 제한되지 않는다.
- [0073] 증폭 산물을 검출하는 방법 중의 하나로서, 형광 측정 방법은 프라이머의 5'-말단에 Cy-5 또는 Cy-3를 표지하여 real-time RT-PCR을 수행하면 표적 서열이 검출 가능한 형광 표지 물질로 표지되며, 이렇게 표지된 형광은 형광 측정기를 이용하여 측정할 수 있다.
- [0074] 또한, 방사성 측정 방법은 RT-PCR 수행 시 32P 또는 35S 등과 같은 방사성 동위원소를 PCR 반응액에 첨가하여 증폭 산물을 표지한 후, 방사성 측정기구, 예를 들면, 가이거 계수기(Geiger counter) 또는 액체섬광계수기(liquid scintillation counter)를 이용하여 방사성을 측정할 수 있다.
- [0075] 본 발명의 바람직한 일구현예에 따르면, 상기 RT-PCR을 통해 증폭된 PCR 산물에 형광이 표지된 프로브가 붙어 특정 파장의 형광을 내게 되고, 증폭과 동시에 PCR 장치의 형광 측정기에서 본 발명의 유전자들의 mRNA 발현 수준을 실시간으로 측정하고, 측정된 값이 계산되어 PC를 통해 시각화 되게 되어 검사자는 쉽게 그 발현 정도를 확인할 수 있다.
- [0076] 본 발명의 다른 측면에 따르면 상기 선별 키트는 역전사 중합효소반응을 수행하기 위해 필요한 필수 요소를 포함하는 것을 특징으로 하는 대장암 및 대장용종 진단용 키트일 수 있다. 역전사 중합효소반응 키트는 상기 본 발명의 유전자에 대한 특이적인 각각의 프라이머 쌍을 포함할 수 있다. 프라이머는 각 마커 유전자의 핵산 서열에 특이적인 서열을 가지는 뉴클레오타이드로서, 약 7 bp 내지 50 bp의 길이, 보다 바람직하게는 약 10 bp 내지 30 bp의 길이일 수 있다.
- [0077] 그 외 역전사 중합효소반응 키트는 테스트 튜브 또는 다른 적절한 컨테이너, 반응 완충액 (pH 및 마그네슘 농도는 다양), 데옥시뉴클레오타이드 (dNTPs), Taq-폴리머라아제 및 역전사효소와 같은 효소, DNase, RNase 억제제, DEPC-수(DEPC-water), 멸균수 등을 포함할 수 있다.
- [0078] 또한, 본 발명의 키트는 최적의 반응 수행 조건을 기재한 사용자 안내서를 추가로 포함할 수 있다.
- [0079] 안내서는 키트 사용법, 예를 들면, 완충액 제조 방법, 제시되는 반응 조건 등을 설명하는 인쇄물이다.
- [0080] 안내서는 팜플렛 또는 진단지 형태의 안내 책자, 키트에 부착된 라벨, 및 키트를 포함하는 패키지의 표면에 설명을 포함할 수 있다. 또한, 안내서는 인터넷과 같이 전기 매체를 통해 공개되거나 제공되는 정보를 포함할 수 있다.
- [0081] 본 발명에서 용어 "대장암 및 대장용종 진단을 위한 정보제공방법"은 진단을 위한 예비적 단계로서 암의 진단을 위하여 필요한 객관적인 기초정보를 제공하는 것이며 의사의 임상학적 판단 또는 소견은 제외된다.
- [0082] 용어 "프라이머"는 짧은 자유 3'말단 수산화기를 가지는 핵산 서열로 상보적인 템플레이트 (template)와 염기쌍을 형성할 수 있고 템플레이트 가닥 복사를 위한 시작 지점으로 기능을 하는 짧은 핵산 서열을 의미한다. 프라이머는 적절한 완충용액 및 온도에서 중합반응 (즉, DNA 중합효소 또는 역전사효소) 을 위한 시약 및 상이한 4 가지 뉴클레오타이드 트리포스페이트의 존재하에서 DNA 합성이 개시할 수 있다. 본 발명의 프라이머는, 각 마커

유전자 특이적인 프라이머로 7개 내지 50개의 뉴클레오타이드 서열을 가진 센스 및 안티센스 핵산이다. 프라이머는 DNA합성의 개시점으로 작용하는 프라이머의 기본 성질을 변화시키지 않는 추가의 특징을 혼입할 수 있다.

- [0083] 용어 "프로브"는 단일쇄 핵산 분자이며, 타겟 핵산 서열에 상보적인 서열을 포함한다.
- [0084] 용어 "실시간 역전사 증합효소 반응 (realtime RT-PCR)"이라 함은 역전사효소를 이용하여 RNA를 상보적인 DNA (cDNA) 로 역전사 시킨 후에 만들어진 cDNA를 주형 (template) 으로 하여 타겟 프라이머와 표지를 포함하는 타겟 프로브를 이용해 타겟을 증폭함과 동시에 증폭된 타겟에 타겟 프로브의 표지에서 발생하는 신호를 정량적으로 검출해 내는 분자생물학적 증합방법이다.
- [0085] 본 발명의 대장암 및 대장용종 진단 또는 예측에는 정보 학습을 통해 대장암 및 대장용종을 진단하도록 할 수 있는 데이터 마이닝 방법이 사용될 수 있으며, 특히 AI 분석을 통해 효과적으로 개선할 수 있다. 따라서 본 발명의 대장암 및 대장용종 진단 또는 예측 방법에는 바람직하게는 대장암 및 대장용종 진단 마커의 상대적 발현량을 측정할 수 있는 방법 및/또는 AI 분석 방법이 사용될 수 있다.
- [0086] 본 발명에 있어 대장암 및 대장용종 예측 모델에 AI 분석을 이용하는 경우, 다양한 해석 가능한 모델을 제한없이 이용할 수 있으며, 선형 회귀, 로지스틱 회귀, 신경망 분석, 의사결정나무, 결정 규칙, 룰셋, 서포트 벡터 머신과 같은 모델을 제한없이 적용가능하고, 본 발명의 바람직한 구현예에서는 특히 로지스틱 회귀 분석, 의사결정 나무, 신경망 분석 및 서포트 벡터 머신을 이용하였다.
- [0087] 한편 본 발명의 예측 모델은 대장암 및 대장용종 진단부, 분류부 및 가중치 부여부를 포함할 수 있으며, 상기 대장암 및 대장용종 진단부는 환자의 해당 질환 관련 유전자 마커의 상대적 발현량 정보 수신부에서 수신된 상대적 발현량 정보를 입력정보로 하고, 상기 대장 관련 질환 분류부는 신경망을 분류기로 하여 대장암 및 대장용종을 분류하는 과정을 수행할 수 있고, 상기 가중치 부여부는 분류 결과에 대하여 가중치를 부여함으로써 대장암 및 대장용종을 선별할 수 있다.
- [0088] 본 발명의 실시예들에 따른 신경망분석은, 하나 이상의 레이어(Layer)를 구축하여 복수의 데이터를 바탕으로 판단을 수행하는 시스템을 의미한다. 예를 들어, 신경망 분석은 입력 층은 유전자 마커의 상대적 발현량 정보를 신경망 분석 모델에 데이터로서 넣어주는 층이며, 출력 층은 입력된 여러 정보를 바탕으로 대장암 및 대장용종 질환 환자의 유무를 판단해 줄 수 있는 결과를 내주는 층이다. 숨김층은 여러가지 판단기준 (유전자 돌연변이 정보)에 대한 가중치를 부여하여 환자 유무를 결정할 수 있는 process를 진행하는 층이다.
- [0089] 본 발명의 실시예에 따른 AI 분석 기법을 이용한 대장암 및 대장용종 예측 방법은 MLP 신경망을 이용하여, 상기 숨김 노드의 수를 가지는 신경망 분석 모형을 추정한다. 또한, 입력변수와 출력변수의 다양한 변수변환을 통하여 구축된 여러 개의 신경망 모형 중 각 모형으로부터 추정된 정확도가 가장 높은 신경망 모형을 피부질환 예측을 위한 최종 신경망 모형으로 결정한다. 상기 AI 분석은 입력 층, 숨김층, 및 출력 층으로 구성될 수 있으며, 상기 신경망 분석 단계를 통한 신경망 분석 모형은 몇 개의 숨김층에 몇 개의 숨김 노드를 가지는 신경망 모형일 수 있다.

**발명의 효과**

- [0091] 본 발명을 통하여 알 수 있는 바와 같이, 본 발명은 비교적 추출이 용이한 혈액 검체를 이용하여 혈액에서 발현하는 유전자 마커들의 발현 양상을 인공지능 알고리즘에 대입하여 대장암 및 대장용종 선별하는 데에 도움을 줄 수 있다.

**도면의 간단한 설명**

- [0093] 도 1은 본 발명의 요약도,
- 도 2는 실험 및 분석 진행한 그룹별 검체 수,
- 도 3은 유전자 바이오마커 검출을 위해 제작한 프라이머 프로브 염기 서열,
- 도 4는 RT-qPCR 수행을 통해 도출한 그룹별 각 유전자 바이오마커의 상대적 발현양 결과,
- 도 5는 전체 샘플을 이용한 그룹별 t-test 통계분석 결과 (Model A 제작을 위한 바이오마커 선별),

도 6은 Model A에서 양성인 샘플을 이용한 그룹별 t-test 통계분석 결과 (Model B 제작을 위한 바이오마커 선별),

도 7은 Model A와 B 에서 음성인 샘플을 이용한 그룹별 t-test 통계분석 결과 (Model C 제작을 위한 바이오마커 선별),

도 8은 Model A 제작 시, 최종으로 도출된 4가지 인공지능 예측 모델의 정확도, 민감도, 특이도 결과,

도 9는 Model B 제작 시, 최종으로 도출된 4가지 인공지능 예측 모델의 정확도, 민감도, 특이도 결과,

도 10은 Model C 제작 시, 최종으로 도출된 4가지 인공지능 예측 모델의 정확도, 민감도, 특이도 결과, 및

도 11은 Model A, B, C 를 순차적으로 적용한 최종 결과 모식도

### 발명을 실시하기 위한 구체적인 내용

- [0094] 이하, 본 발명의 실시예에 대하여 첨부된 도면을 참조하면서 상세히 설명하기로 한다. 이들 실시예는 오로지 본 발명을 보다 구체적으로 설명하기 위한 것으로, 본 발명의 요지에 따라 본 발명의 범위가 이들 실시예에 의해 제한되지 않는다는 것은 당업계에서 통상의 지식을 가진 자에 있어서 자명할 것이다.
- [0096] 실시예 1 : 재료
- [0097] 2017년부터 2019년까지 신촌 세브란스병원, 강남 세브란스병원, 강북 삼성병원 내 소화기 내과에 내원한 사람들의 혈액 샘플을 이용하였다. 혈액은 Tempus blood tube (Applied Biosystems®)를 이용하여 채혈하여 연세대학교 원주캠퍼스 분자진단학 연구실로 운송되었다.
- [0098] 대장내시경 검사의 결과를 통해
- [0099] - 대장 내 병변이 없는 정상군 183명,
- [0100] -저위험 선종이 3개미만 존재하는 저위험군 110명,
- [0101] -저위험 선종이 3개 이상인 경우, 고위험 선종이 1개 이상인 경우, 상피내암인 경우를 포함하는 고위험군 156명,
- [0102] -대장암군 42명으로 분류되었다.
- [0104] 실시예 2 : 혈액 검체에서 Total RNA 분리
- [0105] Tempus tube로 채혈된 혈액검체로부터 Tempus blood RNA isolation kit (Applied Biosystems®) 를 이용하여 Total RNA를 분리한다.
- [0107] 실시예 3 : 분리된 total RNA로부터 cDNA 제작 및 Real-time PCR 수행
- [0108] i.cDNA 합성
- [0109] 분리된 total RNA 1.5~4.5ug, Random primer (Invitrogen) 0.625 ug, dNTP (Intron), MMLV 역전사 중합효소 (Invitrogen) 500 units 을 첨가하고 최종부피를 50 $\mu$ l 가 되도록 DEPC treated DW를 넣고 잘 섞은 후 합성 반응액을 thermocycler (ABI) 에서 25 $^{\circ}$ C 30분 - 37 $^{\circ}$ C 50분 - 70 $^{\circ}$ C 15분 반응 시켜 cDNA를 합성하였다.
- [0111] ii.qPCR 수행
- [0112] Real time PCR 반응물의 조성은 THUNDERBIRD® Probe qPCR Mix (TOYOBO) 10 $\mu$ l와 Forward / Reverse Primer 10pmole과 probe 또한 10pmole을 넣어주고 합성한 cDNA를 15ug 넣고 최종 부피가 20 $\mu$ l 되도록 Ultra pure water 를 넣어 섞어준다. 각각의 프라이머와 프로브 염기서열은 표 1에 기재 되었다.
- [0113] Real time PCR 반응은 CFX96 (Biorad) 를 이용하였으며 반응 온도 조건은 다음과 같다. 95 $^{\circ}$ C 3분 후 95 $^{\circ}$ C 3초

- 60℃ 30초를 40회 반복하여 수행하였다. Annealing 과정 (60℃ 30초) 이 한번 수행될 때마다 형광을 측정하는 과정을 추가하여, 횟수 별로 증가되는 형광 값을 측정하였다.

표 1

[0114]

Target gene	Primer and TaqMan probe	서열 번호	Primer's and Taqman probe's sequence (5' --> 3')	Direction	PCR product (bp)
CCR1	Forward	1	TCC CTT GGA ACC AGA GAG AAG		75
	Taqman probe	2	FAM-CCG GGA TGG AAA CTC CAA ACA CCA-BHQ1		
	Reverse	3	CAA ACT CTG TGG TCG TGT CAT AG		
CES1	Forward	4	GAC CTG TTC CTG GAC TTG ATA G		82
	Taqman probe	5	FAM-TGG TGT CCC ATC TGT GAT TGT GGC-BHQ1		
	Reverse	6	CTC CAG CAT CTC TGT GGT TC		
IL1B	Forward	7	CAC AGA CCT TCC AGG AGA ATG		82
	Taqman probe	8	FAM-ACC TGA GCA CCT TCT TTC CCT TCA		
	Reverse	9	CCC ATG TGT CGA AGA AGA TAG G		
ITGA2	Forward	10	TAT ACA GGA GCC CTC TGA TGT		110
	Taqman probe	11	FAM-TGC GTG TGG ACA TCA GTC TGG AAA-BHQ1		
	Reverse	12	GAC CTT GGC AGT CTC AGA ATA G		
LTF	Forward	13	AAC GAG AGA TAC TAC GGC TAC A		125
	Taqman probe	14	FAM-CTG GCT GAG AAT GCT GGA GAC GTT-BHQ1		
	Reverse	15	GCC CAT GCC TCA TTG TTA TTT C		
TNF	Forward	16	ACT ACC CAA TAA TTC CTG CTA TTC A		105
	Taqman probe	17	FAM-TGC AAG TTG GAG TTC ATC TCC TTC TTC C-BHQ1	(antisense)	
	Reverse	18	CCA TCC AGT GAT ATT TGT GCA TTT		
PTGES	Forward	19	CCA GTA TTG CAG GAG CGAC		101
	Taqman probe	20	FAM-CCA CCG GAA CGA CAT GGA GAC CAT C-BHQ1		
	Reverse	21	AGT AGA CGA AGC CCA GGA AA		
ITIH4	Forward	22	CAG AAT AAG ACC AAG GCT CAC A		118
	Taqman probe	23	FAM-CCC AGA GCA GCA AGA AAC AGT CCT-BHQ1		
	Reverse	24	GGT CCA CAT CAT AGC GGA TAA T		
TUG	Forward	25	CCC TTA CCT AAC AGC ATC TCA C		110
	Taqman probe	26	FAM-AAG GCT GCA CCA GAT TCC AGA AA-BHQ1		
	Reverse	27	TGA TCT AAG AAT AAG TCG GTC ACA A		
NME1	Forward	28	CAG AAA GGA TTC CGC CTT GT		128
	Taqman probe	29	FAM-AAC ACT ACG TTG ACC TGA AGG ACC G-BHQ1		
	Reverse	30	GGCCCTGAGTGCATGTATTT		
PTGS2	Forward	31	TGT GTT GAC ATC CAG ATC AC		117
	Taqman probe	32	FAM-AGG TTA GAG AAG GCT TCC CAG C-BHQ1	(antisense)	
	Reverse	33	GGA GGA AGG GCT CTA GTA TAA		
CXCL11	Forward	34	TAC AGT TGT TCA AGG CTT CCC		96
	Taqman probe	35	FAM-AGA GGA CGC TGT CTT TGC ATA GGC-BHQ1		
	Reverse	36	GCT TTC TCA ATA TCT GCC ACT TTC		
PPARG	Forward	37	CCC TTC ACT ACT GTT GAC TTC TC		133
	Taqman probe	38	FAM-TCA CAA GAA CAG ATC CAG TGG TTG CA-BHQ1		
	Reverse	39	CTT TGA TTG CAC TTT GGT ACT CTT		
MAPK6	Forward	40	TTA GTC GAG AAG CAC TGG ATT T		97
	Taqman probe	41	FAM-CCC ATG GAT CGG TTA ACA GCA GAA GA-BHQ1		
	Reverse	42	CTC ATG TAA GGA TGG GAG AGT G		
GK	Forward	43	TCA GAT TAA TGC GGA GGA AAG T		99

	Taqman probe	44	FAM-AGC TGT GAT GAA GTC AAT GGG TTG GG-BHQ1	
	Reverse	45	CCA CTT TCT GGA GAT TGA GTT GT	
CK19	Forward	46	GAT GAG CAG GTC CGA GGT TA	96
	Taqman probe	47	FAM-CTG CGG CGC ACC CTT CAG GGT CT-BHQ1	
	Reverse	48	TCT TCC AAG GCA GCT TTC AT	
EpCAM	Forward	49	GCC AGT GTA CTT CAG TTG GTG CAC	82
	Taqman probe	50	FAM-TAC TGT CAT TTG CTC AAA GCT GGC TGC CA-BHQ1	
	Reverse	51	CAT TTC TGC CTT CAT CAC CAA ACA	
CD146	Forward	52	TTC TGA AGT GCG GCC TCT CC	74
	Taqman probe	53	FAM-TCC CAA GGC AAC CTC AGC CAT GTC G-BHQ1	
	Reverse	54	CGC TTC TCC TTG TGG ACA GAA AAC	
GAPDH	Forward	55	CCA TCT TCC AGG AGC GAG ATC C	90
	Taqman probe	56	FAM-TCC ACG ACG TAC TCA GCG CCA GCA-BHQ1	
	Reverse	57	ATG GTG GTG AAG ACG CCA GTG	

[0115] 표 1은 본 발명에 사용된 각각의 프라이머와 프로브 염기서열.

[0117] 실시예 4 : 결과 확인 및 표적 유전자의 상대적 발현양 계산

[0118] Endogenous control로 이용된 GAPDH 유전자의 Cq 값을 이용하여 표적 유전자의 Cq 값을 이용하여 표적 유전자의 상대적 발현양( $2^{-\Delta Cq}$ )을 계산한다.

[0119] [관계식]

[0120] 
$$2^{-\Delta Cq} = 2^{-(\text{표적 유전자 } Cq - \text{GAPDH 유전자 } Cq)}$$

[0122] 실시예 7 : 대장암 및 대장용종 진단 예측 모델 제작

[0123] 대장암 및 대장용종 진단 예측 모델에 대입하기 위한 유전자 바이오마커를 선별하고 선별된 유전자 바이오마커의 상대발현양을 대입하여 대장암 및 대장용종 진단 예측 모델을 제작한다.

[0124] 대장암 및 대장용종 진단 예측 모델에 대입하기 위한 유전자 바이오마커를 선별하기 위해서는 SPSS 통계분석 패키지를 이용하였고 선별된 유전자 바이오마커의 상대발현양을 대입하여 대장암 및 대장용종 진단 예측 모델을 제작하기 위한 통계분석은 R 패키지를 사용하였다.

[0125] 대장암 및 대장용종 진단 예측 모델의 제작은 Decision tree (DT), Logistic regression (LR), Neural network (NN), Support vector machine (SVM) 에 의해 수행되었으나 이에 한정되지 않는다.

[0126] 인공지능 예측 모델은 전체 샘플 결과의 일부로 구성된 Training set 결과를 대입하여 제작된다. Training set 에 포함되지 않은 샘플들로 Validation set 를 구성한 후 Validation set 의 결과를 대입하여 Training set 로 제작한 모델의 정확도를 확인한다. 이 때 정확도는 예측 모델의 예측이 얼마나 정확한 지를 의미한다.

[0127] 총 4가지 종류의 모델 (DT, NR, NN, SVM) 을 제작하였고, Training set와 Validation set 를 이용한 결과 도출을 1000회 반복한다. 1000회 동안 만들어지는 모델 중 Validation set 로 확인한 정확도가 가장 높은 모델이 최종 결과로 나타난다. 최종적으로 만들어진 4가지 종류의 모델 중에서 모든 샘플이 포함된 Total set (총 491개) 의 민감도 또는 특이도가 가장 높은 종류의 모델을 선정하였다.

[0129] i. 대장암군과 나머지 그룹 (고위험군/저위험군/정상군)을 구분할 수 있는 인공지능 예측 모델 A 제작

[0130] -t-test 통계 분석을 통하여 각 그룹을 구분하는 데에 통계학적 유의성이 나타나는 유전자 마커들을 확인한다. 이에 해당하는 총 8개 유전자 마커들의 상대적 발현양을 대입하여 모델 A를 제작한다.

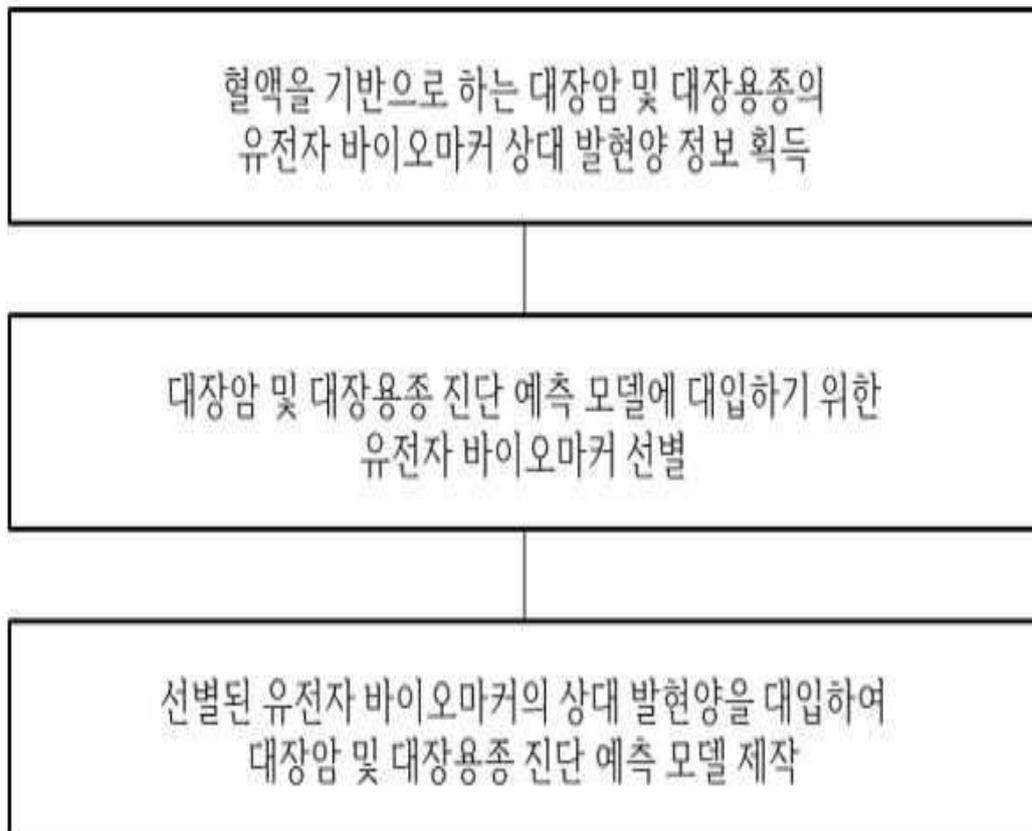
[0131] -이에 따른 결과로 민감도 92.9%, 특이도 65.0%로 대장암군과 고위험군/저위험군/정상군을 구별하는 SVM 모델이

선정되었다.

- [0133] ii. 대장암군과 나머지 그룹 (고위험군/저위험군/정상군) 을 구분할 수 있는 인공지능 예측 모델 B 제작
- [0134] -모델 A는 대장암군을 구별하는 민감도는 높지만 나머지 그룹의 40% 까지 대장암군으로 판별하므로 모델 A에서 양성인 나온 샘플들을 이용해 다시 한번 대장암군과 나머지 그룹을 구별할 수 있는 모델을 제작해야한다.
- [0135] -모델 A에서 양성인 나온 샘플들을 대상으로 t-test 통계 분석을 통하여 각 그룹을 구분하는 데에 통계학적 유의성이 나타나는 유전자 마커들을 확인한다. 이에 해당하는 총 9개 유전자 마커들의 상대적 발현양을 대입하여 모델 B를 제작한다.
- [0136] -이에 따른 결과로 민감도 94.9%, 특이도 87.9% 로 대장암군과 고위험군/저위험군/정상군을 구별하는 SVM 모델이 선정되었다.
- [0137] iii. 고위험군과 저위험군/정상군을 구분할 수 있는 인공지능 예측 모델 C 제작
- [0138] -모델 A와 B에서 음성이 나온 샘플들에는 대장내시경이 필요한 고위험군이 포함되어 있으므로 고위험군과 저위험군/정상군을 구분하는 모델을 제작해야한다.
- [0139] -모델 A와 B에서 음성이 나온 샘플들을 대상으로 t-test 통계 분석을 통하여 각 그룹을 구분하는 데에 통계학적 유의성이 나타나는 유전자 마커들을 확인한다. 이에 해당하는 총 6개 유전자 마커들의 상대적 발현양을 대입하여 모델 C를 제작한다.
- [0140] -이에 따른 결과로 민감도 91.3%, 특이도 81.9% 로 고위험군과 저위험군/정상군을 구별하는 SVM 모델이 선정되었다.
- [0142] iv. 인공지능 예측 모델 A - B - C를 순차적으로 적용한 전체적인 결과
- [0143] -i, ii, iii에서 제작한 인공지능 예측 모델 A - B - C에 순차적으로 전체 샘플의 유전자 마커 상대적 발현양을 대입한 결과, 95.2%의 대장암군과 91.7%의 고위험군이 양성으로 나타났으며 78.5%의 저위험군과 정상군이 음성으로 나타났다.
- [0144] 상기 모델 A, B 및 C에 사용된 바이오마커는 하기와 같다.
- [0145] 모델 A- IL1B, LTF, TNF, ITIH4, CXCL11, MAPK6, GK, CD146 (8개),
- [0146] 모델 B- CES1, IL1B, TNF, ITIH4, PTGS2, CXCL11, MAPK6, GK, CD146 (9개) 및
- [0147] 모델 C- IL1B, LTF, TNF, ITIH4, CXCL11, MAPK6 (6개)

도면

도면1



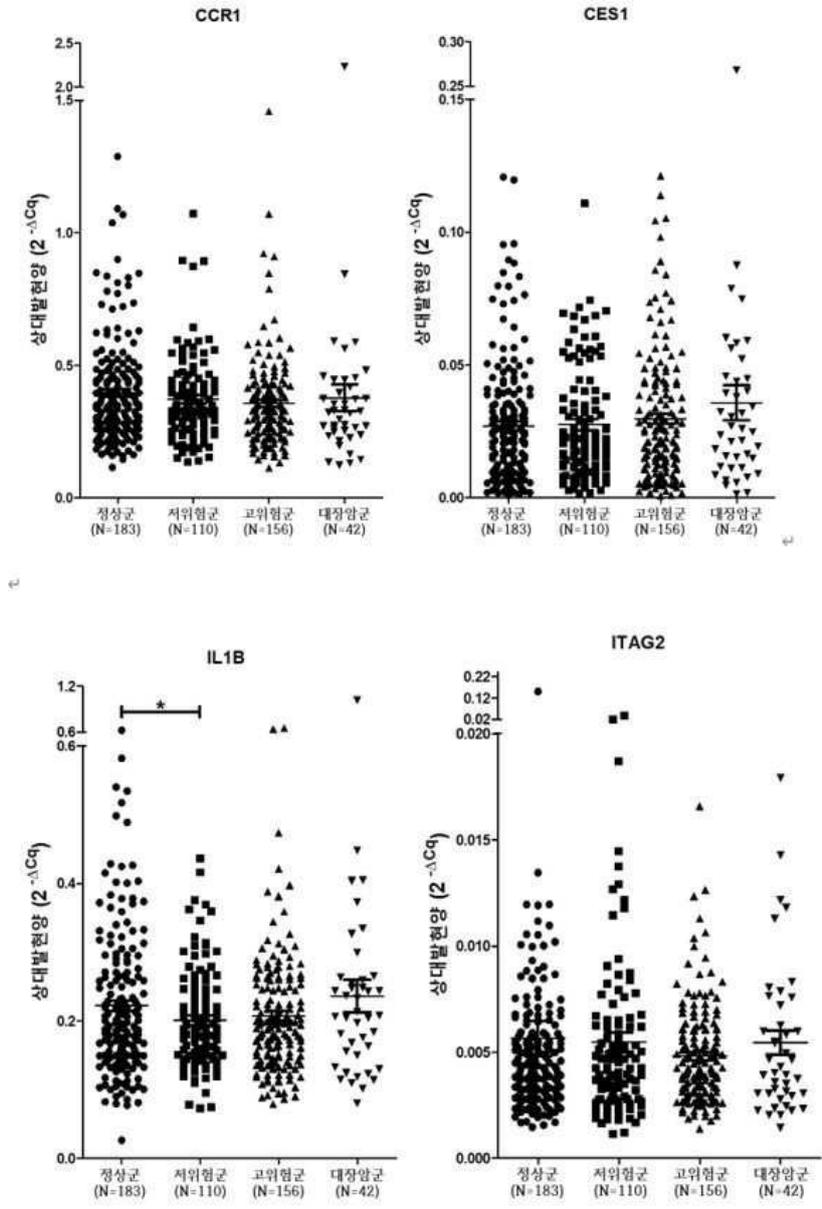
도면2

그룹	검체 수 (명)
대장암군	42
고위험군	156
저위험군	110
정상군	183
총합	491

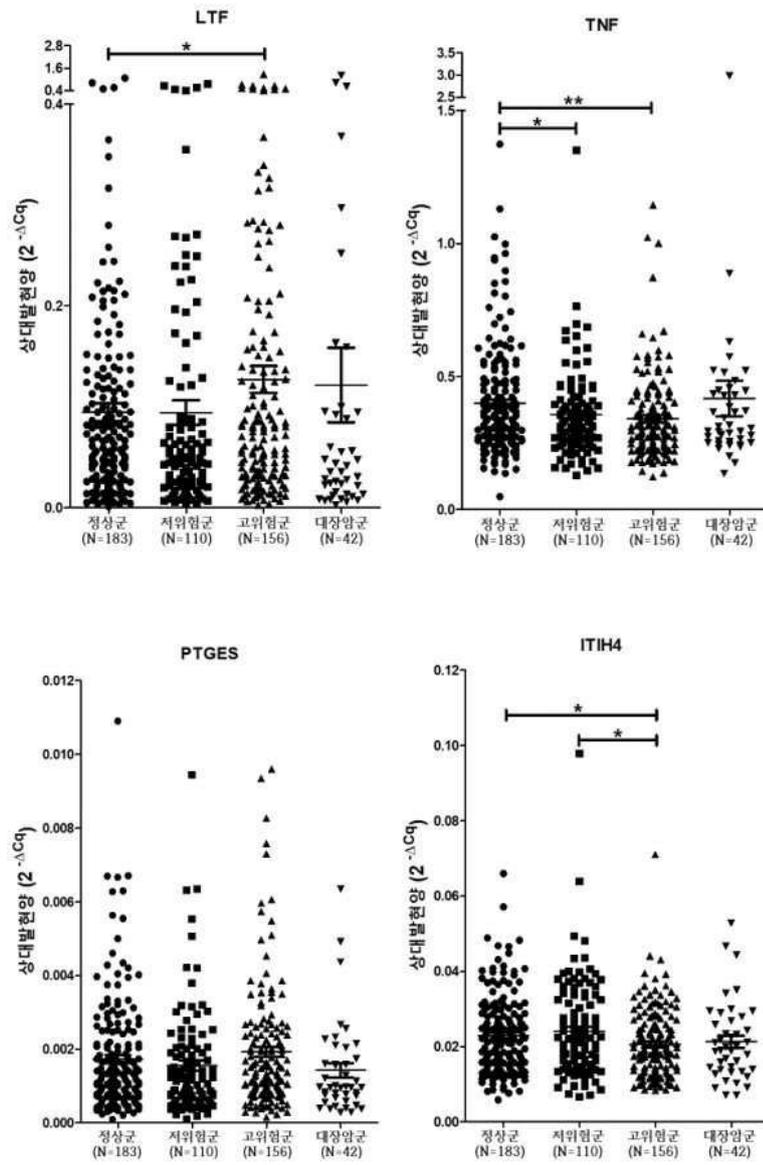
도면3

Target gene	Primer and TaqMan probe	Primer's and Taqman probe's sequence (5' → 3')
CCR1	Forward	TCCCTTGGAAACCAGAGAGAAG
	Taqman	FAM-CCGGGATGGAACTCCAAACACCA-BHQ1
	Reverse	CAAACCTGTGGTCGTGCATAG
CES1	Forward	GACCTGTTCCTGGACTTGATAG
	Taqman	FAM-TGGGTGCCATCTGTGATTGTGGC-BHQ1
	Reverse	CTCCAGCATCTCTGTGGTTC
IL1B	Forward	CACAGACCTTCCAGGAGAATG
	Taqman	FAM-ACCTGAGCACCTTCTTCCCTTCA
	Reverse	CCCATGTGTGGAAGAAGATAGG
ITGA2	Forward	TATACAGGAGCCCTCTGATGT
	Taqman	FAM-TGCGTGTGGACATCAGTCTGGAAA-BHQ1
	Reverse	GACCTTGGCAGTCTCAGAATAG
LTF	Forward	AACGAGAGATACTACGGCTACA
	Taqman	FAM-CTGGCTGAGAATGCTGGAGACGTT-BHQ1
	Reverse	GCCCATGCCCTATTGTATTTC
TNF	Forward	ACTACCCAATAATTCCTGCTATTCA
	Taqman	FAM-TGCAAGTTGGAGTTTCACTCCTTCTTCC-BHQ1
	Reverse	CCATCCAGTGATATTGTGCATTT
PTGES	Forward	CCAGTATTGCAGGAGCGAC
	Taqman	FAM-CCACCGGAACGACATGGAGACCATC-BHQ1
	Reverse	AGTAGACGAAGCCAGGAAA
ITIH4	Forward	CAGAATAAGACCAAGGCTCACA
	Taqman	FAM-CCGAGAGCAGCAAGAACAGTCT-BHQ1
	Reverse	GGTCCACATCATAGCGGATAAT
TUG	Forward	CCCTTACCTAACAGCATCTCAC
	Taqman	FAM-AAGGCTGCACCAGATTCCAGAAA-BHQ1
	Reverse	TGATCTAAGAATAAGTCGGTCACAA
NME1	Forward	CAGAAAGGATTCCGCCTTGT
	Taqman	FAM-AACACTACGTTGACCTGAAGGACCG-BHQ1
	Reverse	GGCCCTGAGTGCATGATTT
PTGS2	Forward	TGT GTT GAC ATC CAG ATC AC
	Taqman	FAM-AGG TTA GAG AAG GCT TCC CAG C-BHQ1
	Reverse	GGA GGA AGG GCT CTA GTA TAA
CXCL11	Forward	TACAGTTGTTCAAGGCTTCCC
	Taqman	FAM-AGAGGACGCTGTCTTTGCATAGGC-BHQ1
	Reverse	GCTTCTCAATATCTGCCACTTTC
PPARG	Forward	CCCTTCACTACTGTGACTTCTC
	Taqman	FAM-TCACAAGAACAGATCCAGTGGTTGCA-BHQ1
	Reverse	CTTTGATTGCACCTTGGTACTCTT
MAPK6	Forward	TTAGTCGAGAAGCACTGGATT
	Taqman	FAM-CCCATGGATCGGTTAACAGCAGAAGA-BHQ1
	Reverse	CTCATGTAAGGATGGGAGAGTG
GK	Forward	TCAGATTAATGCGGAGGAAAGT
	Taqman	FAM-AGCTGTGATGAAGTCAATGGGTTGGG-BHQ1
	Reverse	CCACTTCTGGAGATTGAGTTGT
CK19	Forward	GATGAGCAGGTCGAGGTTA
	Taqman	FAM-CTGCGGCGCACCTTCAGGGTCT-BHQ1
	Reverse	TCTTCCAAGGCAGCTTTCAT
EpCAM	Forward	GCCAGTGTACTTCAGTTGGTGAC
	Taqman	FAM-TACTGTCAATTGCTCAAAGCTGGTGCCA-BHQ1
	Reverse	CATTTCTGCCTTCATCACCACA
CD146	Forward	TTCTGAAAGTGGGCTCTCC
	Taqman	FAM-TCCCAAGGCAACCTCAGCCATGTCG-BHQ1
	Reverse	CGCTTCTCCTTGTGGACAGAAAAC
GAPDH	Forward	CCATCTCCAGGAGCGAGATCC
	Taqman	FAM-TCCACGACGTACTCAGCGCCAGCA-BHQ1
	Reverse	ATGGTGGTGAAGACGCCAGTG

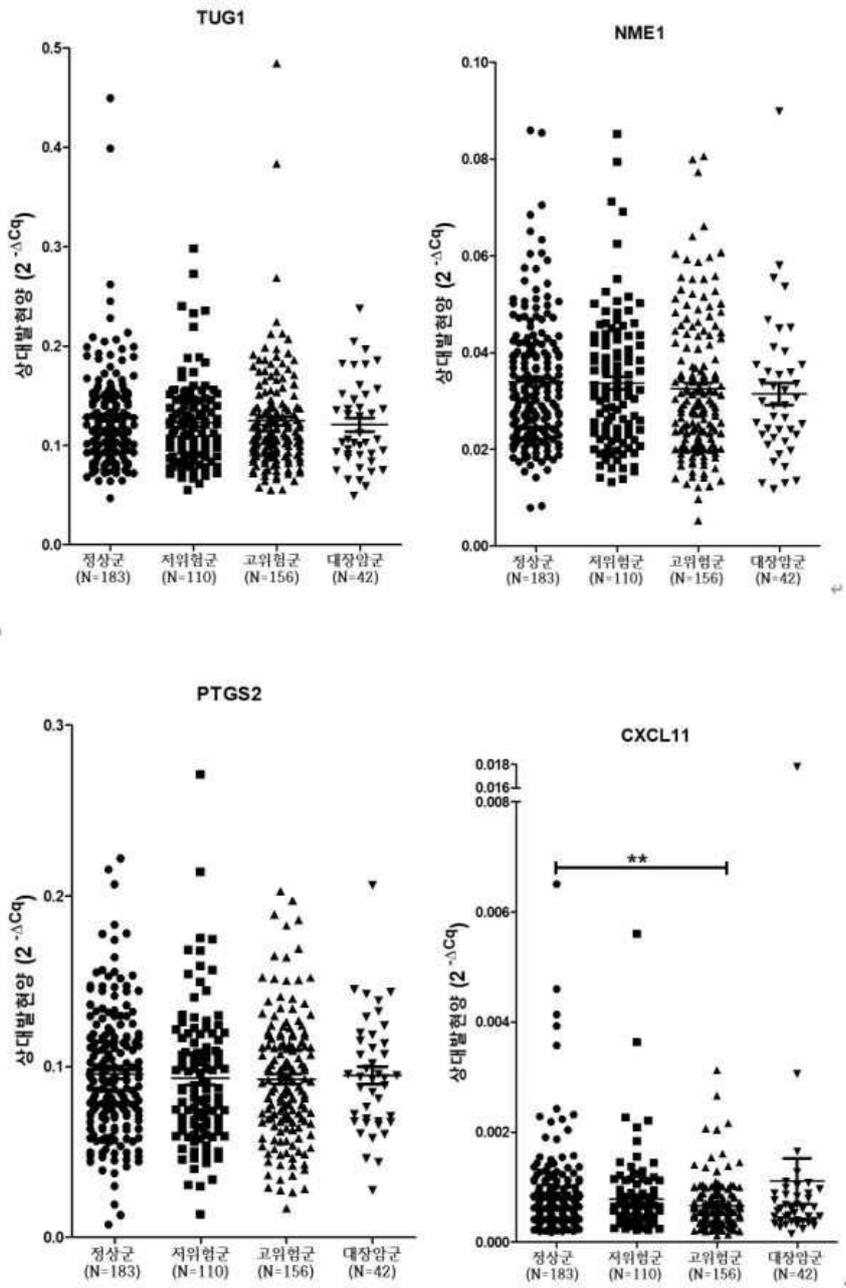
도면4a



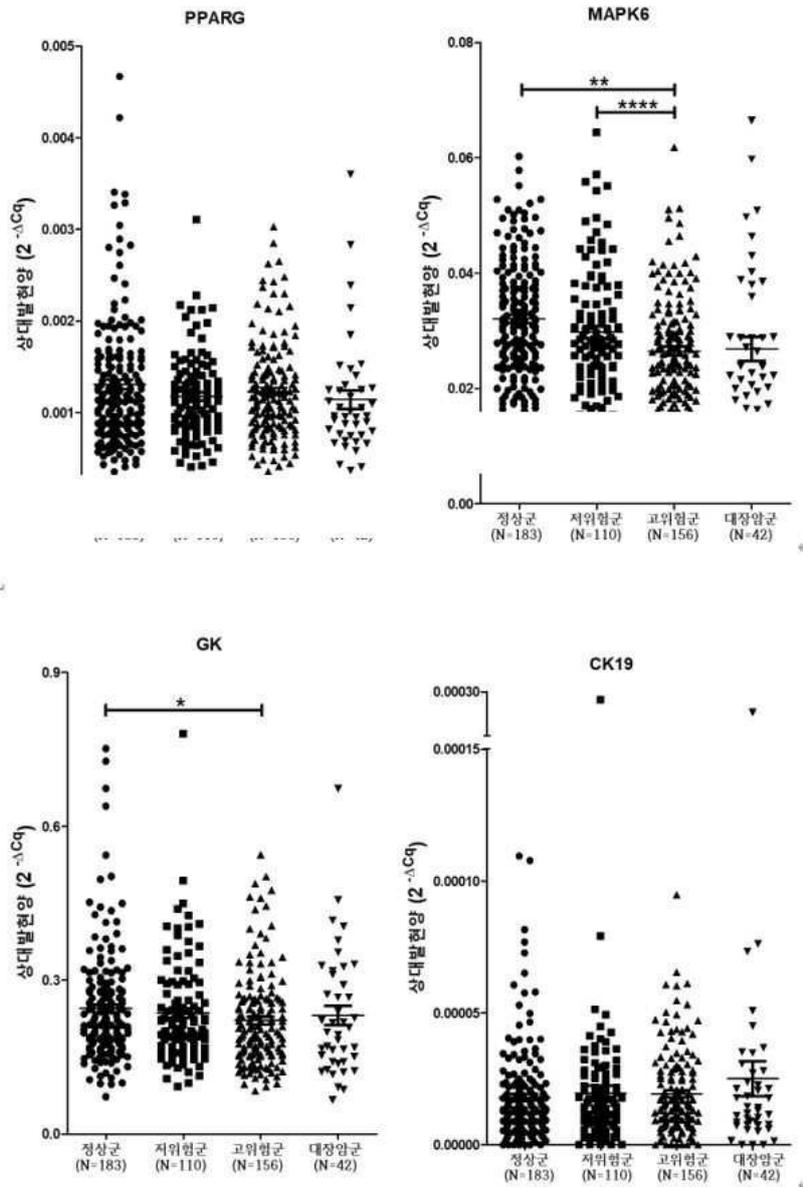
도면4b



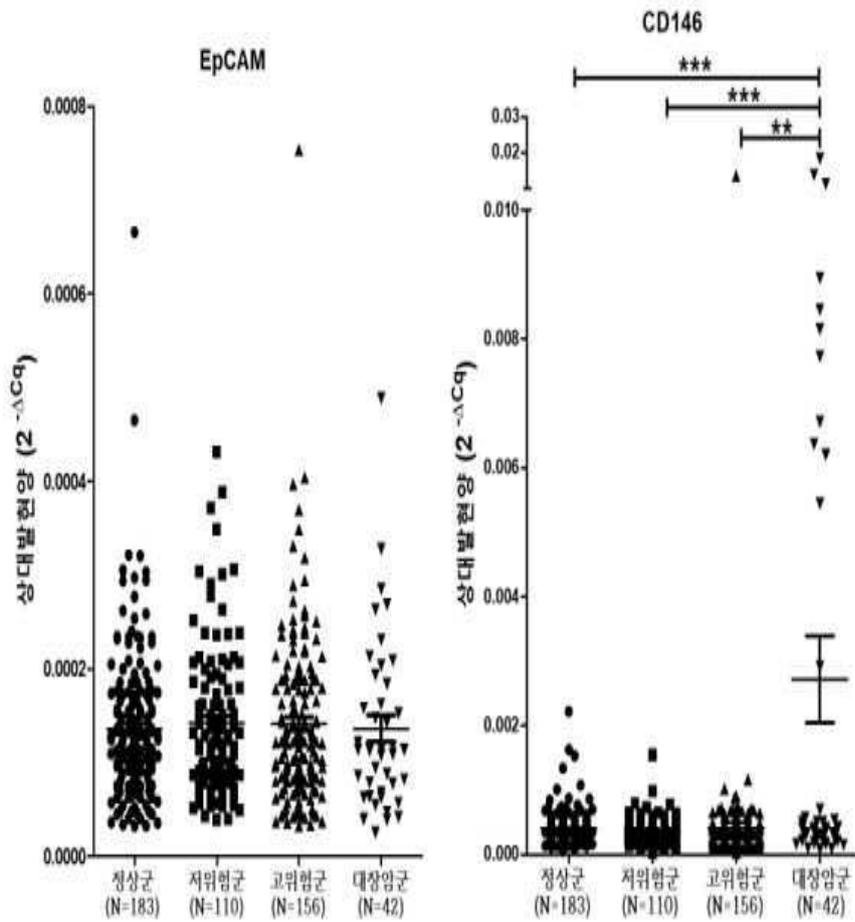
도면4c



도면4d



도면4e



도면5

Markers	대장암군 (N=42)		고위험군 (N=156)		저위험군 (N=110)		정상군 (N=183)		대장암군 vs 고위험군		대장암군 vs 저위험군		대장암군 vs 정상군		고위험군 vs 저위험군		고위험군 vs 정상군		저위험군 vs 정상군	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	p-value	FC	p-value	FC	p-value	FC	p-value	FC	p-value	FC	p-value	FC
CCRI	0.37784	0.32613	0.35688	0.17633	0.37237	0.16025	0.39675	0.19512	0.578	1.059	0.890	1.015	0.624	0.952	0.465	0.958	0.051	0.900	0.270	0.939
CES1	0.03576	0.04274	0.02973	0.02470	0.02754	0.02146	0.02706	0.02357	0.240	1.203	0.119	1.298	0.072	1.321	0.454	1.079	0.311	1.099	0.862	1.018
IL1B	0.23631	0.15212	0.20724	0.08828	0.20140	0.07330	0.22221	0.10486	0.113	1.140	0.160	1.173	0.475	1.063	0.570	1.029	0.154	0.933	0.047	0.906
ITGA2	0.00545	0.00362	0.00481	0.00238	0.00547	0.00474	0.00566	0.01109	0.283	1.133	0.981	0.996	0.904	0.963	0.181	0.880	0.350	0.850	0.865	0.967
LTF	0.12167	0.23865	0.12711	0.16649	0.09423	0.13039	0.09466	0.12391	0.865	0.957	0.367	1.291	0.480	1.285	0.085	1.349	0.046	1.343	0.978	0.995
TNF	0.41759	0.43010	0.34212	0.15879	0.35638	0.15982	0.39882	0.20436	0.270	1.221	0.201	1.172	0.675	1.047	0.473	0.960	0.004	0.858	0.049	0.894
PTGES	0.00143	0.00127	0.00192	0.00170	0.00156	0.00145	0.00173	0.00155	0.078	0.741	0.611	0.917	0.243	0.826	0.065	1.237	0.265	1.114	0.349	0.901
ITIH4	0.02128	0.01039	0.02055	0.00916	0.02401	0.01305	0.02287	0.01016	0.657	1.035	0.227	0.887	0.365	0.931	0.018	0.856	0.030	0.899	0.434	1.050
TUG1	0.12088	0.04447	0.12468	0.05273	0.12311	0.04350	0.12804	0.04879	0.669	0.970	0.780	0.982	0.385	0.944	0.797	1.013	0.543	0.974	0.384	0.961
NME1	0.03144	0.01477	0.03259	0.01442	0.03375	0.01372	0.03402	0.01316	0.649	0.965	0.366	0.932	0.265	0.924	0.511	0.966	0.342	0.958	0.868	0.992
PTGS2	0.09492	0.03385	0.09268	0.03637	0.09330	0.03973	0.09565	0.03742	0.719	1.024	0.815	1.017	0.908	0.992	0.895	0.993	0.461	0.969	0.611	0.975
CXCL11	0.00111	0.00268	0.00066	0.00044	0.00078	0.00068	0.00085	0.00079	0.291	1.670	0.438	1.420	0.532	1.312	0.089	0.851	0.008	0.786	0.473	0.923
PPARG	0.00114	0.00064	0.00123	0.00053	0.00118	0.00046	0.00131	0.00070	0.363	0.929	0.714	0.971	0.159	0.873	0.392	1.045	0.250	0.940	0.052	0.899
MAPK6	0.02690	0.01359	0.02649	0.00957	0.02981	0.01086	0.03208	0.01022	0.856	1.015	0.170	0.902	0.024	0.838	0.009	0.889	<0.0001	0.826	0.073	0.929
GK	0.23132	0.12002	0.22152	0.09169	0.23599	0.09934	0.24489	0.10727	0.625	1.044	0.807	0.980	0.470	0.945	0.222	0.939	0.033	0.905	0.480	0.964
CK19	0.00003	0.00004	0.00002	0.00002	0.00002	0.00003	0.00002	0.00002	0.407	1.295	0.382	1.273	0.313	1.385	0.903	0.983	0.490	1.069	0.560	1.088
EpCAM	0.00014	0.00009	0.00014	0.00009	0.00014	0.00008	0.00014	0.00008	0.740	0.963	0.727	0.963	0.976	1.003	0.999	1.000	0.537	1.042	0.550	1.042
CD146	0.00272	0.00438	0.00041	0.00107	0.00035	0.00020	0.00040	0.00026	0.002	6.606	0.001	7.697	0.001	6.723	0.571	1.165	0.930	1.018	0.079	0.874

도면6

Markers	대상암군 (N=39)		고위험군 (N=55)		저위험군 (N=28)		정상군 (N=74)		대상암군 vs 고위험군		대상암군 vs 저위험군		대상암군 vs 정상군		고위험군 vs 저위험군		고위험군 vs 정상군		저위험군 vs 정상군	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	p-value	FC	p-value	FC	p-value	FC	p-value	FC	p-value	FC	p-value	FC
CCR1	0.384332	0.337758	0.365106	0.294299	0.420632	0.200240	0.460876	0.219156	0.732	1.05	0.616	0.91	0.148	0.83	0.247	0.87	<b>0.013</b>	0.79	0.403	0.91
CESI	0.038722	0.043198	0.031382	0.025722	0.029125	0.025201	0.031137	0.026918	0.544	1.14	0.472	1.23	0.489	1.15	0.705	1.08	0.958	1.01	0.733	0.94
IL1B	0.240095	0.156602	0.210983	0.103503	0.203129	0.082167	0.255617	0.130171	0.281	1.14	0.259	1.18	0.576	0.94	0.728	1.04	<b>0.032</b>	0.83	0.018	0.79
ITGA2	0.005602	0.003691	0.004751	0.002930	0.007516	0.007333	0.007338	0.017189	0.216	1.18	0.165	0.75	0.535	0.76	0.064	0.63	0.272	0.65	0.958	1.02
LTF	0.127899	0.246484	0.198215	0.235640	0.182506	0.198198	0.129699	0.177356	0.165	0.65	0.664	0.84	0.964	0.99	0.382	1.30	0.074	1.53	0.576	1.18
TNF	0.427662	0.444545	0.371052	0.209292	0.440145	0.235645	0.489425	0.243146	0.411	1.15	0.893	0.97	0.341	0.87	0.177	0.84	<b>0.004</b>	0.76	0.359	0.90
PTGS	0.001453	0.001303	0.001797	0.001344	0.001732	0.001602	0.001695	0.001534	0.219	0.81	0.435	0.84	0.403	0.86	0.847	1.04	0.697	1.06	0.915	1.02
ITIH4	0.022024	0.010388	0.019566	0.008121	0.027996	0.017714	0.025039	0.011082	0.201	1.13	0.088	0.79	0.163	0.88	<b>0.023</b>	0.70	<b>0.002</b>	0.78	0.315	1.12
TUG1	0.123522	0.044878	0.120805	0.053304	0.138704	0.065355	0.130413	0.043962	0.796	1.02	0.263	0.89	0.433	0.95	0.184	0.87	0.265	0.93	0.539	1.06
NME1	0.031674	0.015238	0.033191	0.016247	0.031966	0.016891	0.036208	0.014743	0.648	0.95	0.941	0.99	0.127	0.87	0.749	1.04	0.273	0.92	0.216	0.88
PTGS2	0.095248	0.034775	0.087679	0.031116	0.092990	0.036173	0.103869	0.041866	0.271	1.09	0.797	1.02	0.273	0.92	0.489	0.94	<b>0.017</b>	0.84	0.228	0.90
CXCL11	0.001141	0.002785	0.000746	0.000555	0.001081	0.001001	0.001109	0.001071	0.308	1.53	0.914	1.06	0.931	1.03	0.053	0.69	<b>0.014</b>	0.67	0.904	0.97
PPARG	0.001117	0.000597	0.001268	0.000565	0.001200	0.000466	0.001360	0.000806	0.216	0.88	0.540	0.93	0.099	0.82	0.587	1.06	0.466	0.93	0.324	0.88
MAPK6	0.027416	0.013903	0.025243	0.012265	0.033036	0.014670	0.031834	0.011616	0.425	1.09	0.116	0.83	0.076	0.86	<b>0.012</b>	0.76	<b>0.002</b>	0.79	0.699	1.04
GK	0.237335	0.122507	0.219543	0.112196	0.260422	0.137165	0.277353	0.136938	0.468	1.08	0.472	0.91	0.129	0.86	0.150	0.84	<b>0.012</b>	0.79	0.579	0.94
CK19	0.000025	0.000045	0.000020	0.000014	0.000015	0.000012	0.000022	0.000024	0.400	1.28	0.257	1.66	0.646	1.13	0.160	1.30	0.458	0.89	0.052	0.68
EpCAM	0.000140	0.000094	0.000130	0.000059	0.000146	0.000100	0.000150	0.000080	0.551	1.08	0.816	0.96	0.561	0.93	0.445	0.89	0.119	0.87	0.825	0.97
CD146	0.002894	0.004506	0.000314	0.000176	0.000336	0.000261	0.000408	0.000295	<b>0.001</b>	9.21	<b>0.001</b>	8.61	<b>0.001</b>	7.10	0.650	0.93	<b>0.039</b>	0.77	0.264	0.82

도면7

Markers	고위험군 (N=149)		저위험군 (N=108)		정상군 (N=173)		고위험군 vs 저위험군		고위험군 vs 정상군		저위험군 vs 정상군	
	Mean	SD	Mean	SD	Mean	SD	p-value	FC	p-value	FC	p-value	FC
CCR1	0.35897	0.17949	0.37394	0.16127	0.39738	0.19717	0.492	0.96	0.070	0.90	0.300	0.94
CES1	0.02977	0.02508	0.02774	0.02161	0.02658	0.02286	0.499	1.07	0.234	1.12	0.674	1.04
IL1B	0.21062	0.08823	0.20167	0.07334	0.22347	0.10467	0.39	1.04	0.233	0.94	0.041	0.90
ITGA2	0.00487	0.00236	0.00551	0.00477	0.00572	0.01139	0.204	0.88	0.372	0.85	0.854	0.96
LTF	0.12170	0.16295	0.09564	0.13117	0.09062	0.10269	0.172	1.27	0.046	1.34	0.721	1.06
TNF	0.34669	0.16057	0.35678	0.16044	0.40158	0.20713	0.619	0.97	0.008	0.86	0.043	0.89
PTGES	0.00191	0.00172	0.00157	0.00146	0.00174	0.00157	0.092	1.22	0.350	1.10	0.356	0.90
ITIH4	0.02089	0.00918	0.02413	0.01306	0.02305	0.01032	0.028	0.87	0.050	0.91	0.441	1.05
TUG1	0.12669	0.05292	0.12332	0.04363	0.12987	0.04926	0.588	1.03	0.578	0.98	0.259	0.95
NME1	0.03305	0.01445	0.03386	0.01378	0.03416	0.01337	0.651	0.98	0.475	0.97	0.858	0.99
PTGS2	0.09389	0.03657	0.09344	0.03971	0.09562	0.03728	0.924	1.00	0.676	0.98	0.642	0.98
CXCL11	0.00067	0.00045	0.00078	0.00068	0.00084	0.00079	0.118	0.86	0.017	0.80	0.524	0.93
PPARG	0.00123	0.00053	0.00119	0.00045	0.00132	0.00071	0.486	1.04	0.186	0.93	0.052	0.90
MAPK6	0.02698	0.00931	0.02982	0.01095	0.03251	0.01021	0.026	0.90	<0.0001	0.83	0.038	0.92
GK	0.22561	0.09155	0.23674	0.09998	0.24448	0.10738	0.356	0.95	0.094	0.92	0.547	0.97
CK19	0.00002	0.00002	0.00002	0.00003	0.00002	0.00002	0.947	0.99	0.365	1.09	0.501	1.11
EpCAM	0.00014	0.00009	0.00014	0.00008	0.00014	0.00008	0.924	1.01	0.443	1.05	0.519	1.05
CD146	0.00041	0.00109	0.00035	0.00020	0.00041	0.00027	0.583	1.16	0.946	1.01	0.080	0.87

도면8

1000회 최고	DT			LR			NN			SVM		
	정확도	민감도	특이도	정확도	민감도	특이도	정확도	민감도	특이도	정확도	민감도	특이도
Training set	76.1%	60.6%	89.5%	63.4%	33.3%	89.5%	83.1%	72.7%	92.1%	100.0%	100.0%	100.0%
Validation set	90.0%	77.8%	100.0%	90.0%	77.8%	100.0%	90.0%	100.0%	81.8%	85.0%	66.7%	100.0%
Total set	87.8%	64.3%	90.0%	79.8%	42.9%	83.3%	68.8%	78.6%	67.9%	67.4%	92.9%	65.0%

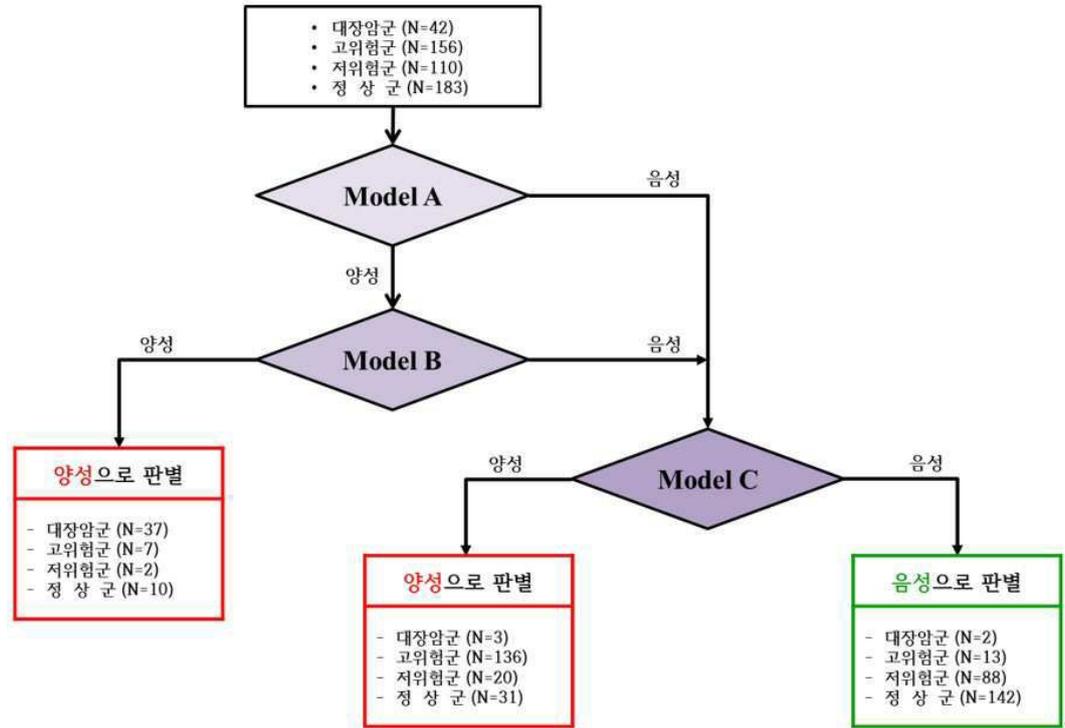
도면9

1000회 최고	DT			LR			NN			SVM		
	정확도	민감도	특이도	정확도	민감도	특이도	정확도	민감도	특이도	정확도	민감도	특이도
Training set	84.3%	83.9%	84.6%	71.4%	48.4%	89.7%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Validation set	89.5%	75.0%	100.0%	84.2%	62.5%	100.0%	94.7%	100.0%	90.9%	84.2%	62.5%	100.0%
Total set	77.0%	82.1%	75.8%	78.6%	51.3%	85.4%	73.0%	100.0%	66.2%	85.7%	92.3%	84.1%

도면10

1000회 최고	DT			LR			NN			SVM		
	정확도	민감도	특이도	정확도	민감도	특이도	정확도	민감도	특이도	정확도	민감도	특이도
Training set	74.8%	68.1%	80.4%	59.5%	47.1%	69.9%	61.5%	39.5%	79.7%	100.0%	100.0%	100.0%
Validation set	70.1%	56.7%	81.1%	76.1%	63.3%	86.5%	77.6%	60.0%	91.9%	73.1%	56.7%	86.5%
Total set	70.9%	65.8%	73.7%	62.8%	50.3%	69.4%	65.8%	43.6%	77.6%	85.1%	91.3%	81.9%

도면11



서열 목록

- <110> UNIVERSITY INDUSTRY FOUNDATION, YONSEI UNIVERSITY WONJU CAMPUS
- <120> A method for sorting colon polyp and colorectal cancer and use of the same}
- <130> P21-0012
- <160> 57
- <170> KoPatentIn 3.0
- <210> 1
- <211> 21
- <212> DNA
- <213> Artificial Sequence
- <220><223> primer
- <400> 1
- tcacctggaa ccagagagaa g 21
- <210> 2
- <211> 24
- <212> DNA
- <213> Artificial Sequence

<220><223> probe

<400> 2  
 ccgggatgga aactccaaac acca 24

<210> 3  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence

<220><223> primer

<400> 3  
 caaactctgt ggtcgtgtca tag 23

<210> 4  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence

<220><223> primer

<400> 4  
 gacctgttcc tggacttgat ag 22

<210> 5  
 <211> 24  
 <212> DNA  
 <213>  
 > Artificial Sequence

<220><223> probe

<400> 5  
 tgggtgccca tctgtgattg tggc 24

<210> 6  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence

<220><223> primer

<400> 6  
 ctccagcatc tctgtggttc 20

<210> 7  
 <211> 21

<212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 7  
 cacagacctt ccaggagaat g 21

<210> 8  
 <211> 24  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> probe  
 <400> 8  
 acctgagcac cttctttccc ttca 24

<210> 9  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 9  
 cccatgtgtc gaagaagata gg 22

<210> 10  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 10  
 tatacaggag ccctctgatg t 21

<210> 11  
 <211> 24  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> probe  
 <400> 11  
 tgcgtgtgga catcagtctg gaaa 24

<210> 12  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 12  
 gaccttggca gtctcagaat ag 22  
 <210> 13  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 13  
 aacgagagat actacggcta ca 22  
  
 <210> 14  
 <211> 24  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> probe  
 <400> 14  
 ctggctgaga atgctggaga cggt 24  
 <210> 15  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 15  
 gcccatgcct cattgttatt tc 22  
 <210> 16  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 16

actaccaat aattcctgct attca	25
<210> 17	
<211> 28	
<212> DNA	
<213> Artificial Sequence	
<220><223> probe	
<400> 17	
tgcaagttgg agttcatctc cttcttcc	28
<210> 18	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 18	
ccatccagtg atatttgatc attt	24
<210> 19	
<211> 19	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 19	
ccagtattgc aggagcgac	19
<210> 20	
<211> 25	
<212> DNA	
<213> Artificial Sequence	
<220><223> probe	
<400> 20	
ccaccggaac gacatggaga ccatc	25
<210> 21	
<211> 20	
<212> DNA	
<213> Artificial Sequence	

<220><223> primer  
 <400> 21  
 agtagacgaa gccaggaaa 20  
 <210> 22  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 22  
 cagaataaga ccaaggctca ca 22  
  
 <210> 23  
 <211> 24  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> probe  
 <400> 23  
 cccagagcag caagaaacag tcct 24  
 <210> 24  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 24  
 ggtccacatc atagcggata at 22  
 <210> 25  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 25  
 cccttaccta acagcatctc ac 22  
  
 <210> 26  
 <211> 23

<212> DNA  
 <213> Artificial Sequence  
 <220><223> probe  
 <400> 26  
 aaggctgcac cagattccag aaa 23  
 <210> 27  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 27  
 tgatctaaga ataagtcggt cacaa 25  
 <210> 28  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 28  
 cagaaaggat tccgccttgt 20  
  
 <210> 29  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> probe  
 <400> 29  
 aacactacgt tgacctgaag gaccg 25  
 <210> 30  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 30  
 ggccctgagt gcatgtattt 20  
 <210> 31

<211> 20  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 31  
 tgtgttgaca tccagatcac 20

<210> 32  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> probe  
 <400> 32  
 aggttagaga aggcttccca gc 22

<210> 33  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 33  
 ggaggaaggg ctctagtata a 21

<210> 34  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 34  
 tacagttggt caaggcttcc c 21

<210> 35  
 <211> 24  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> probe  
 <400> 35

agaggacgct gtctttgcat aggc	24
<210> 36	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 36	
gcttttctcaa tatctgccac tttc	24
<210> 37	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 37	
cccttcacta ctgttgactt ctc	23
<210> 38	
<211> 26	
<212> DNA	
<213> Artificial Sequence	
<220><223> probe	
<400> 38	
tcacaagaac agatccagtg gttgca	26
<210> 39	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	
<400> 39	
ctttgattgc accttggtac tctt	24
<210> 40	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220><223> primer	

<400> 40  
 ttagtcgaga agcactggat tt 22

<210> 41  
 <211> 26  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> probe

<400> 41  
 cccatggatc ggtaacagc agaaga 26  
 <210> 42  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer

<400> 42  
 ctcatgtaag gatgggagag tg 22  
 <210> 43  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer

<400> 43  
 tcagattaat gcggaggaaa gt 22

<210> 44  
 <211> 26  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> probe

<400> 44  
 agctgtgatg aagtcaatgg gttggg 26  
 <210> 45  
 <211> 23  
 <212> DNA

<213> Artificial Sequence  
 <220><223> primer  
 <400> 45  
 ccactttctg gagattgagt tgt 23  
 <210> 46  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 46  
 gatgagcagg tccgaggtta 20  
  
 <210> 47  
 <211> 23  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> probe  
 <400> 47  
 ctgcggcgca cccttcaggg tct 23  
 <210> 48  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 48  
 tcttccaagg cagctttcat 20  
 <210> 49  
 <211> 24  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 49  
 gccagtgtac ttcagttggt gcac 24  
  
 <210> 50

<211> 29  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> probe  
 <400> 50  
 tactgtcatt tgetcaaagc tggctgcca 29  
 <210> 51  
 <211> 24  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 51  
 catttctgcc ttcacacca aaca 24  
 <210> 52  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 52  
 ttctgaagtg cggcctctcc 20  
  
 <210> 53  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> probe  
 <400> 53  
 tccaaggca acctcagcca tgtcg 25  
 <210> 54  
 <211> 24  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> primer  
 <400> 54  
 cgcttctcct tgtggacaga aaac 24

<210> 55  
<211> 22  
<212> DNA  
<213> Artificial Sequence  
<220><223> primer  
<400> 55  
ccatcttcca ggagcgagat cc 22

<210> 56  
<211> 24  
<212> DNA  
<213> Artificial Sequence  
<220><223> probe  
<400> 56  
tccacgacgt actcagcgcc agca 24

<210> 57  
<211> 21  
<212> DNA  
<213> Artificial Sequence  
<220><223> primer  
<400> 57  
atggtggtga agacgccagt g 21