



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0003692
(43) 공개일자 2021년01월12일

- (51) 국제특허분류(Int. Cl.)
G01N 33/574 (2006.01) G16H 10/40 (2018.01)
G16H 50/70 (2018.01) G16H 70/60 (2018.01)
- (52) CPC특허분류
G01N 33/57419 (2013.01)
G01N 33/57484 (2013.01)
- (21) 출원번호 10-2020-0081406
(22) 출원일자 2020년07월02일
심사청구일자 2020년07월02일
- (30) 우선권주장
1020190079593 2019년07월02일 대한민국(KR)

- (71) 출원인
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
- (72) 발명자
김호근
서울특별시 종로구 사직로8길 24, 4층 406(아파트)호(내수동)
- 장미
경기도 남양주시 진접읍 금강로 1530-14, 110동 503호(진접하우스토리아파트)
(뒷면에 계속)
- (74) 대리인
특허법인충현

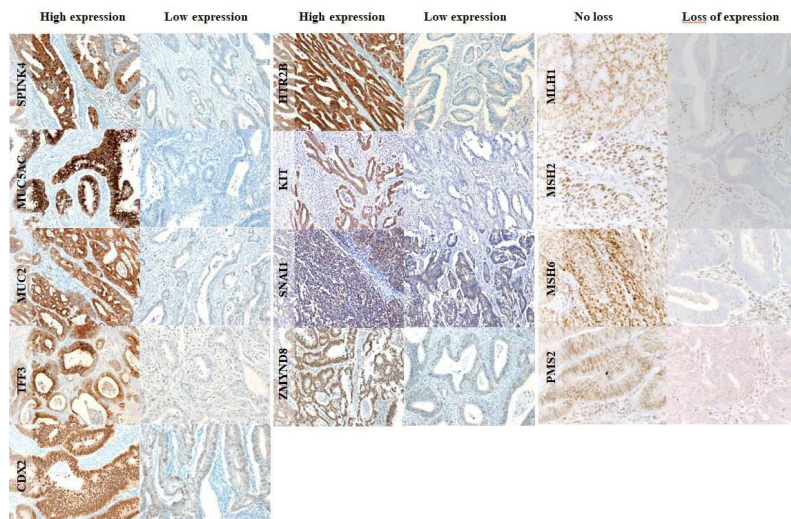
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 면역분석을 이용한 대장암의 분자적 분류방법

(57) 요약

본 발명은 대장암의 예후 예측용 조성물, 대장암의 분자 아형(Molecular subtype) 분류용 조성물 및 이를 이용한 대장암의 분자 아형 분류 방법에 관한 것이다. 본 발명에서 발굴된 마커 단백질들의 발현 수준을 측정함으로써 대장암의 분자 아형을 정확히 분류하고, 분류된 분자 아형을 기반으로 대장암의 재발률, 환자의 전체 생존 기간, 무병 생존률 및 FOLFOX 등의 화학치료제에 대한 치료 반응성을 포함하는 예후를 높은 신뢰도로 예측할 수 있다. 이에, 본 발명은 전사체 발현을 통한 종래의 분류 방법과 달리 섬유아세포 및 기질에 의해 결과에 영향을 받지 않을 뿐 아니라 염색 결과의 육안 확인만으로도 간편하게 분류가 가능하므로, 시간, 비용 및 정확성 측면에서 보다 효율적인 분자 아형 분류 및 예후 예측 방법으로 유용하게 이용될 수 있다.

대표도 - 도1



(52) CPC특허분류

G16H 10/40 (2018.01)

G16H 50/70 (2018.01)

G16H 70/60 (2018.01)

(72) 발명자

김원규

강원도 강릉시 사임당로 679, 사택 313(대전동)

김기열

서울특별시 서대문구 연세로 50-1, 연세대학교 치
과대학 치의학교육연구센터(신촌동)

명세서

청구범위

청구항 1

SNAI1, MUC2, SPINK4, KIT, CDX2, ZMYND8, TFF3, MUC5AC, HTR2B 및 미스매치 복구 단백질(mismatch repair protein)로 구성된 군으로부터 선택되는 하나 이상의 단백질의 발현 수준을 측정하는 제제를 유효성분으로 포함하는 대장암의 예후 예측용 조성물.

청구항 2

제 1 항에 있어서, 상기 미스매치 복구 단백질(mismatch repair protein)은 MLH1, MSH2, MSH6 및 PMS2로 구성된 군으로부터 선택되는 하나 이상의 단백질인 것을 특징으로 하는 조성물.

청구항 3

제 1 항에 있어서, 상기 단백질의 발현 수준을 측정하는 제제는 항체 또는 앵타머인 것을 특징으로 하는 조성물.

청구항 4

제 3 항에 있어서, 상기 단백질의 발현 수준은 면역조직화학 염색(immunohistochemical staining) 방법에 의해 측정하는 것을 특징으로 하는 조성물.

청구항 5

제 1 항에 있어서, 상기 대장암은 2-4기 대장암인 것을 특징으로 하는 조성물.

청구항 6

제 1 항에 있어서, 상기 예후는 재발률, 전체 생존 기간, 무병 생존률 및 화학치료제에 대한 치료 반응성으로 구성된 군으로부터 선택되는 하나 이상의 인자인 것을 특징으로 하는 조성물.

청구항 7

제 6 항에 있어서, 상기 화학치료제는 옥살리플라틴, 5-플루오로우라실(5-FU), 류코보린 및 이들의 조합으로 구성된 군으로부터 선택되는 것을 특징으로 하는 조성물.

청구항 8

제 1 항에 있어서, 상기 조성물은 ZMYND8, MUC5AC 및 이들의 조합으로 구성된 군으로부터 선택되는 단백질에 대한 항체를 유효성분으로 포함하는 것을 특징으로 하는 조성물.

청구항 9

제 8 항에 있어서, 상기 조성물은 SNAI1, MUC2, SPINK4, CDX2, ZMYND8, TFF3, MUC5AC, HTR2B, MLH1, MSH2, MSH6 및 PMS2에 대한 항체를 유효성분으로 포함하는 것을 특징으로 하는 조성물.

청구항 10

제 8 항에 있어서, 상기 조성물은 SNAI1, MUC2, SPINK4, CDX2, ZMYND8, TFF3, MUC5AC, KIT, MLH1, MSH2, MSH6 및 PMS2에 대한 항체를 유효성분으로 포함하는 것을 특징으로 하는 조성물.

청구항 11

다음의 단계를 포함하는 대장암 분자 아형(Molecular subtype)의 분류 방법:

- (a) 대장암 환자로부터 분리된 생물학적 시료로부터 SNAI1, MUC2, SPINK4, CDX2, ZMYND8, TFF3, MUC5AC, KIT, HTR2B, MLH1, MSH2, MSH6 및 PMS2로 구성된 군으로부터 선택되는 하나 이상의 단백질의 발현 수준을 측정하는 단계;
- (b) 단계 (a)에서 측정된 단백질의 발현 수준을 정규화하는 단계;
- (c) 단계 (b)에서 정규화된 단백질의 발현 수준을 이용하여 기준 분자 아형(Reference molecular subtype)에 대한 각 분자 아형의 로그값을 도출하는 단계; 및
- (d) 단계 (c)에서 도출된 각 분자 아형의 로그값을 통해 대장암의 분자 아형을 결정하는 단계.

청구항 12

제 11 항에 있어서, 상기 단계 (a)는 ZMYND8, MUC5AC 또는 이들의 조합을 포함하는 단백질의 발현 수준을 측정함으로써 수행되는 것을 특징으로 하는 방법.

청구항 13

제 12 항에 있어서, 상기 단계 (a)는 SNAI1, MUC2, SPINK4, CDX2, ZMYND8, TFF3, MUC5AC, HTR2B, MLH1, MSH2, MSH6 및 PMS2의 발현 수준을 측정함으로써 수행되는 것을 특징으로 하는 방법.

청구항 14

제 12 항에 있어서, 상기 단계 (a)는 SNAI1, MUC2, SPINK4, CDX2, ZMYND8, TFF3, MUC5AC, KIT, MLH1, MSH2, MSH6 및 PMS2의 발현 수준을 측정함으로써 수행되는 것을 특징으로 하는 방법.

청구항 15

제 11 항에 있어서, 상기 단계 (a)는 면역조직화학 염색(immunohistochemical staining) 방법에 의해 수행되는 것을 특징으로 하는 방법.

청구항 16

제 15 항에 있어서, 상기 단계 (b)는 각 단백질에 대한 면역조직화학 염색 결과에 대한 H-점수를 기반으로 하여 아래와 같이 각 단백질의 발현 수준에 대해 0 또는 1의 값을 부여함으로써 수행되는 것을 특징으로 하는 방법:

TFF3 = 0(H-점수 ≥ 10) 또는 1(H-점수 < 10);

SPINK4 = 0(H-점수<20) 또는 1(H-점수≥20);

KIT = 0(부재) 또는 1(존재);

CDX2 = 0(H-점수≥150) 또는 1(H-점수<150);

ZMYND8 = 0(저) 또는 1(고);

MUC2 = 0(H-점수<10) 또는 1(H-점수≥10);

MUC5AC = 0(H-점수<10) 또는 1(H-점수≥10);

SNAI1 = 0(H-점수<20) 또는 1(H-점수≥20);

HTR2B = 0(H-점수<100) 또는 1(H-점수≥100); 및

MLH1, MSH2, MSH6 및 PMS2 = 0 (초위성체 불안정성 낮음) 또는 1(초위성체 불안정성 높음).

청구항 17

제 16 항에 있어서, 상기 단계 (c)는 상기 단계 (b)에서 정규화된 각 단백질의 발현 수준 값을 하기 식 1-1 내지 식 1-3 또는 식 2-1 내지 식 2-3에 대입함으로써 수행되는 것을 특징으로 하는 방법:

[식 1-1]

$$\ln[p(\text{MS2})/P(\text{MS1})] = -1.33 + (0.34)a + (-2.09)b + (-0.55)c + (0.94)d + (1.51)e + (-0.79)f + (1.89)g + (2.52)h + (2.55)i;$$

[식 1-2]

$$\ln[p(\text{MS3})/P(\text{MS1})] = (-0.4) + (2.05)a + (-2.34)b + (-1.61)c + (1.07)d + (1.8)e + (-0.91)f + (-0.61)g + (-15.8)h + (0.45)i;$$

[식 1-3]

$$\ln[p(\text{MS4})/P(\text{MS1})] = 3.18 + (1.31)a + (-2.39)b + (-2.96)c + (-0.39)d + (0.61)e + (-1.42)f + (-1.22)g + (-19.76)h + (-14.58)i;$$

(a=SNAI1, b=MUC2, c=SPINK4, d=HTR2B, e=CDX2, f=ZMYND8, g=TFF3, h=초위성체 불안정성(MSI) 상태, i=MUC5AC);

[식 2-1]

$$\ln[p(\text{MS2})/P(\text{MS1})] = 286.6 + (-92.2)a + (-5.2)b + (-53.0)c + (-85.3)d + (-80.1)e + (-35.9)f + (-207.7)g + (46.1)h + (-344.4)i;$$

[식 2-2]

$$\ln[p(\text{MS3})/P(\text{MS1})] = 282.0 + (-91.0)a + (-4.5)b + (-52.4)c + (-83.7)d + (-78.9)e + (-32.2)f + (-141.9)g + (43.1)h + (-173.5)i;$$

[식 2-3]

$$\ln[p(\text{MS4})/P(\text{MS1})] = 283.9 + (-92.1)a + (-5.9)b + (-50.9)c + (-83.7)d + (-80.2)e + (-35.2)f + (-143.8)g + (47.7)h + (-366.0)i;$$

(a=TFF3, b=SPINK4, c=KIT, d=CDX2, e=ZMYND8, f=MUC2, g=MUC5AC, h=SNAI1, i=초위성체 불안정성(MSI) 상태).

청구항 18

제 11 항에 있어서, 상기 분자 아형은 CMS1(초위성체 불안정성/면역 활성화형), CMS2(정준(Canonical)형), CMS3(대사조절 장애) 및 CMS4(간엽형)를 포함하는 것을 특징으로 하는 방법.

청구항 19

SNAI1, MUC2, SPINK4, KIT, CDX2, ZMYND8, TFF3, MUC5AC, HTR2B 및 미스매치 복구 단백질(mismatch repair protein)로 구성된 군으로부터 선택되는 하나 이상의 단백질의 발현 수준을 측정하는 체제를 유효성분으로 포함하는 대장암의 분자 아형(Molecular subtype)의 분류용 조성물.

청구항 20

제 19 항에 있어서, 상기 미스매치 복구 단백질(mismatch repair protein)은 MLH1, MSH2, MSH6 및 PMS2로 구성된 군으로부터 선택되는 하나 이상의 단백질을 것을 특징으로 하는 조성물.

발명의 설명

기술 분야

[0001] 본 발명은 항원-항체 반응에 기반한 면역분석, 구체적으로는 면역조직화학 분석을 이용한 대장암의 분자 아형(Molecular Subtype) 분류 방법에 관한 것이다.

배경 기술

[0003] 직장암(Colorectal cancer, CRC)은 전세계에서 세 번째로 흔한 암으로 5년 생존률이 65%이다^{1,2}. 현재, 예후 예측은 대부분 질환의 병적 단계에 기반하여 이루어진다. 보조제 치료를 통해 효과를 거둘 수 있는 환자 그룹을 동정하기 위한 분류전략을 확립하는 것이 시급히 요구된다.

[0004] 3기 직장암에 있어, MOSAIC trial에 기반한 표준적인 치료로서 외과적 수술 후 FOLFOX(옥살리플라틴, 5-플루오로우라실[5-FU] 및 류코보린의 복합제)를 이용한 보조적 화학치료를 시행하는 방법이 있다^{3,4}. 보조제 화학치료에 의해 암의 재발을 30% 감소시킬 수 있으며, 사망률은 22%에서 32%까지 감소시킨다. 옥살리플라틴은 장기간 감각신경병의 병인을 제공하는 부작용을 가지며, 이는 삶의 질 저하로 이어진다⁵. 각각의 환자는 치료제에 다르게 반응하는데, 이는 종양의 이종성(heterogeneity)에 기인한다.

[0005] 최근, 약물에 대한 반응성을 예측하게 해 주는 몇몇 분자적 특징이 알려졌다. 예를 들어, 높은 수준의 초위성체 불안정성(microsatellite instability)(MSI-High)을 가지는 종양은 5-플루오로우라실에 대한 치료 반응성이 떨어질 수 있음이 보고되었다^{6,7}. KRAS/BRAF 축에서의 돌연변이 검출은 전이성 암에서 상피 성장인자 수용체(EGFR)를 타겟으로 한 치료에 대한 저항성을 예측하는 데에 중요하다^{8,9}. 그러나, 현재 이용 가능한 분자 마커는 염색체 불안정성과 같이 제한된 정보만을 제공하여 다양한 의약들에 대한 반응성을 설명하는 데에는 불충분하다.

[0006] 직장암의 분자적 이질성에 대한 몇몇 연구에서 유전자 발현 데이터를 사용하여 질환을 분류하고자 하였다¹⁰⁻¹⁷. 최근, 국제협회에서 종래 보고된 직장암 아형과 제안된 4가지 컨센서스 분자 아형(Consensus Molecular Subtype, CMS) 간의 상호 관계를 분석하였다^{11-15,17,18}. 여기에는 초위성체 불안정성, 점액성, 염증성 아형(CMS1, 14%); WNT 및 MYC 신호의 두드러진 활성화를 보이는 상피성 아형(CMS2, 37%); 대사조절장애를 가지는 상피성 아형(CMS3, 13%); 및 상피-간엽전환(EMT)을 보이는 중간엽-유사 아형, 초위성체 안정성, 혈관신생, 및 TGF- β 과 활성화(CMS4, 23%)를 포함한다. 미분류 표현형(13%)은 혼합된 특성을 보였다. CMS4는 불량한 예후를 가지는 것으로 잘 알려져 있으며 옥살리플라틴을 포함하는 화학치료제에 대한 반응성에서도 유사한 불량 예후를 보인다^{18,19}.

[0007] 이들 분자 아형을 통해 임상병리적인 특성과 질환의 자연적 예후를 예측할 수 있지만, FOLFOX 투여 후 3기 직장암의 예후를 평가하기 위해서는 보다 많은 정보가 필요하다. 특히, 분자 아형에 따른 FOLFOX 보조제 화학치료에 대한 반응성 평가는 3기 직장암의 진행 경과 및 각각의 특정 서브그룹의 동정을 통해 FOLFOX 화학치료의 효

과를 극대화할 수 있는 정보를 제공할 수 있다.

[0008] 본 발명자들은 FOLFOX 보조제 화학치료를 받은 3기 직장암 환자로 구성된 트레이닝 코호트 (코호트 1); 및 I-IV 기 직장암 환자들로 구성된 검증 코호트 (코호트 2)의 2개 코호트를 확립하였다.

[0009] 본 명세서 전체에 걸쳐 다수의 논문 및 특허문헌이 참조되고 그 인용이 표시되어 있다. 인용된 논문 및 특허문헌의 개시 내용은 그 전체로서 본 명세서에 참조로 삽입되어 본 발명이 속하는 기술 분야의 수준 및 본 발명의 내용이 보다 명확하게 설명된다.

선행기술문헌

비특허문헌

[0011] (비특허문헌 0001) 비특허문헌 1. Korean J Gastroenterol Vol. 68 No. 6, 297-302 (2016)

발명의 내용

해결하려는 과제

[0012] 본 발명자들은 대장암 조직의 분자적 특성에 기반하여 대장암의 임상적 특성에 대한 신뢰도 높은 정보를 수득하는 방법을 개발하기 위해 예의 연구 노력하였다. 그 결과, 대장암 조직 내의 상기 나열된 단백질들의 발현 수준을 측정함으로써 대장암의 분자 아형(molecular subtype)을 정확히 분류하고, 분류된 분자 아형을 기반으로 대장암의 재발률, 환자의 전체 생존 기간, 무병 생존률 및 FOLFOX 등의 화학치료제에 대한 치료 반응성을 포함하는 대장암의 예후를 높은 신뢰도로 예측할 수 있음을 발견함으로써, 본 발명을 완성하게 되었다.

[0013] 따라서 본 발명의 목적은 대장암의 예후 예측용 조성물을 제공하는 데 있다.

[0014] 본 발명의 다른 목적은 대장암 분자 아형(Molecular subtype)의 분류용 조성물 및 이를 이용한 분자 아형의 분류 방법을 제공하는 데 있다.

[0016] 본 발명의 다른 목적 및 이점은 하기의 발명의 상세한 설명, 청구범위 및 도면에 의해 보다 명확하게 된다.

과제의 해결 수단

[0018] 본 발명의 일 양태에 따르면, 본 발명은 SNAI1, MUC2, SPINK4, KIT, CDX2, ZMYND8, TFF3, MUC5AC, HTR2B 및 미스매치 복구 단백질(mismatch repair protein)로 구성된 군으로부터 선택되는 하나 이상의 단백질의 발현 수준을 측정하는 제제를 유효성분으로 포함하는 대장암의 예후 예측용 조성물을 제공한다.

[0019] 본 발명자들은 대장암 조직의 분자적 특성에 기반하여 대장암의 임상적 특성에 대한 정확한 정보를 수득하는 방법을 개발하기 위해 예의 연구 노력하였다. 그 결과, 대장암 조직 내에서 상기 나열된 단백질들의 발현 수준을 측정함으로써 대장암의 분자 아형(molecular subtype)을 정확히 분류하고, 분류된 분자 아형을 기반으로 대장암의 재발률, 환자의 전체 생존 기간, 무병 생존률 및 FOLFOX 등의 화학치료제에 대한 치료 반응성을 포함하는, 대장암의 예후를 높은 신뢰도로 예측할 수 있음을 발견하였다.

[0020] 본 명세서에서 용어 “대장암”은 소장과 항문 사이에 위치하는 소화기관인 대장에서 발생한 악성 종양을 의미한다. 대장암은 주로 대장의 상피세포에서 암세포가 발생하게 되며, 암이 발생하는 위치에 따라 결장암과 직장암으로 분류된다. 따라서, 용어 “대장암”은 “결장암” 및 “직장암”을 포괄하는 의미이다.

[0021] 본 발명에서 용어 “분자 아형(molecular subtype)”은 해당 조직의 유전자 및/또는 단백질 발현 프로파일을 기반으로 해당 조직의 질환(예를 들어 암)에 대한 민감성, 예후, 특정 의약에 대한 치료 반응성을 포함하는 고유의 표현형/유전자형에 따라 분류되는 아형(subtype)을 포괄하는 의미이다.

[0022] 종양의 분자 아형 분류는 정확한 진단 뿐 아니라 개개의 암 조직에 대한 생물학적 특성을 예측할 수 있다는 점

에서 매우 중요하다. 대장암은 다른 암에 비해 임상적으로나 형태학적으로 비교적 균일한 특성을 보이면서도 생물학적 특성과 진행 과정이 매우 다양하고 그 발생과정이 분자적으로 뚜렷이 구별되는 유전적, 후성 유전학적 배경을 가지므로, 맞춤형 치료 전략을 조기에 수립하기 위한 정확한 분자 아형의 분류는 매우 중요한 문제이다.

[0023] 본 발명의 구체적인 구현예에 따르면, 대장암의 분자 아형은 예를 들어 Guinney J 등[(Nat Med 21:1350-1356(2015))]이 제시한 기준에 따라 CMS1(초위성체 불안정성/면역 활성화형), CMS2(정준(Canonical)형), CMS3(대사조절 장애) 및 CMS4(간엽형)로 나누어질 수 있다.

[0024] 이에 따라, 본 발명은 대장암의 분자 아형을 분류함으로써 대장암의 예후 또는 치료 방법 결정을 위한 정보를 제공할 수 있다.

[0025] 본 명세서에서 용어 “발현수준”은 유전자에 의해 생성된 유전자 생성물의 정량적인 측정 결과를 의미한다. 예를 들어, 상이하게 발현되는 유전자(DEG) 또는 이로부터 선정된 본 발명의 마커 유전자(단백질)의 발현수준은 DEG의 전사에 의해 생성된 mRNA를 포함하는 전사체의 양 또는 차별 발현 유전자의 DNA의 양을 의미할 수 있고, 또는 번역에 의해 생성된 단백질의 양을 의미할 수 있다. 보다 구체적으로는, 면역 분석 등을 통해 정량화된 단백질의 양을 의미한다.

[0026] 본 명세서에서 “발현수준”은 “발현값”과 동일하게 해석될 수 있다.

[0027] 본 명세서 용어 “예후(prognosis)”는 질병을 진단하여 판단된 분석 시점 이후의 증세 또는 경과에 대한 전망을 포괄하는 의미이다. 대장암 환자에 있어서 예후는 통상적으로 암 발병 또는 외과적 시술 후 일정기간 내의 전이 여부 또는 생존기간을 뜻하며, 예후의 예측은 특히 대장암 환자의 화학치료 여부를 비롯하여 향후 대장암 치료의 방향에 대한 단서를 제시하므로 매우 중요한 임상적 과제이다.

[0028] 본 발명의 구체적인 구현예에 따르면, 상기 예후는 재발률, 전체 생존 기간, 무병 생존률 및 화학치료제에 대한 치료 반응성으로 구성된 군으로부터 선택되는 하나 이상의 인자이다.

[0029] 본 명세서에서 용어 “전체 생존 기간(overall survival, OS)”은 치료 시작후 부터 사망에 이르는 기간을 의미한다. OS는 환자별 편차가 크기 때문에 통상 평균이 아닌 중간값(median)을 기준으로 삼으며, 치료 기간동안 사망하지 않은 환자는 확인된 가장 긴 시간으로 산정해 계산한다.

[0030] 본 명세서에서 용어 “무병 생존율(disease-free survival)”은 전체 환자 군에서 치료 시작 후 5년 동안 재발이나 전이의 증가 없이 생존한 환자들의 비율을 의미한다. 용어 “전이(metastasis)”는 종양이 원발 부위에서 여러 경로를 따라 다른 신체의 부위에 이식되어 그곳에 정착 및 증식하는 상태를 의미한다. 암의 전이여부는 해당 암의 고유의 특성에 의하여 결정될 뿐만 아니라 암의 예후 결정에 있어서 가장 중요한 단서가 되는 사건이므로, 암 환자의 생존과 관련된 가장 중요한 임상정보로 다루어진다.

[0031] 본 명세서에서 용어 “치료 반응성(responsiveness)”은 치료제가 환자에게 치료적 유효량으로 투여되었을 때 생체 내에서 질환의 증상의 진행의 억제, 경감, 제거와 같은 작용을 하는 정도를 의미한다.

[0032] 본 발명의 구체적인 구현예에 따르면, 상기 미스매치 복구 단백질(mismatch repair protein)은 MLH1, MSH2, MSH6 및 PMS2로 구성된 군으로부터 선택되는 하나 이상의 단백질이다.

[0033] 본 발명의 구체적인 구현예에 따르면, 상기 단백질의 발현 수준을 측정하는 제제는 항체 또는 앵타머이다.

[0034] 본 명세서에서 용어 “항체”는 당해 기술분야에 공지된 용어로서 항원성 부위에 대하여 지시되는 특이적인 면역 글로불린을 의미한다. 본 발명에서의 항체는 본 발명의 상기 나열된 마커 단백질들에 대해 특이적으로 결합하는 항체를 의미하며, 당해 기술분야의 통상적인 방법에 따라 항체를 제조할 수 있다. 상기 항체의 형태는 폴리클로날 항체 또는 모노클로날 항체를 포함하며, 바람직하게는 모노클로날 항체이다. 상기 항체는 2개의 전체 길이의 경쇄 및 2 개의 전체 길이의 중쇄를 갖는 완전한 형태를 의미한다. 또한, 상기 항체는 인간화 항체 등의 특수 항체도 포함된다.

[0035] 본 발명에 따르면, 상기 나열한 본 발명의 마커 단백질들을 항원-항체 반응을 이용한 면역분석(immunoassay) 방법에 따라 검출하여 대장암의 분자아형 분류 및 이를 통한 예후 예측에 이용할 수 있다. 이러한 면역분석은 종래에 개발된 다양한 면역분석 또는 면역염색 프로토콜에 따라 실시될 수 있다.

[0036] 예를 들어, 본 발명의 방법이 방사능면역분석 방법에 따라 실시되는 경우, 방사능동위원소(예컨대, C^{14} , I^{125} , P^{32} 및 S^{35})로 표지된 항체가 이용될 수 있다. 본 발명의 항체는 당업계에서 통상적으로 실시되는 방법들, 예를 들

어, 융합 방법(Kohler and Milstein, *European Journal of Immunology*, 6:511-519 (1976)), 재조합 DNA 방법(미국 특허 제4,816,567호) 또는 파아지 항체 라이브러리 방법(Clackson et al, *Nature*, 352:624-628(1991) 및 Marks et al, *J. Mol. Biol.*, 222:58, 1-597(1991))에 의해 제조될 수 있다. 항체 제조에 대한 일반적인 과정은 Harlow, E. and Lane, D., *Using Antibodies: A Laboratory Manual*, Cold Spring Harbor Press, New York, 1999; 및 Zola, H., *Monoclonal Antibodies: A Manual of Techniques*, CRC Press, Inc., Boca Raton, Florida, 1984에 상세하게 기재되어 있다.

- [0037] 본 발명은 항체 대신 GCN5 단백질에 특이적으로 결합하는 앵타머를 이용할 수도 있다. 본 명세서에서 용어 “앵타머”는 특정 표적물질에 높은 친화력과 특이성으로 결합하는 단일 줄기의(single-stranded) 핵산(RNA 또는 DNA) 분자 또는 펩타이드 분자를 의미한다. 앵타머의 일반적인 내용은 Hoppe-Seyler F, Butz K "Peptide aptamers: powerful new tools for molecular medicine". *J Mol Med*. 78(8):426-30(2000); Cohen BA, Colas P, Brent R . "An artificial cell-cycle inhibitor isolated from a combinatorial library". *Proc Natl Acad Sci USA*. 95(24):14272-7(1998)에 상세하게 개시되어 있다.
- [0038] 보다 구체적으로는, 상기 단백질의 발현 수준은 면역조직화학 염색(immunohistochemical staining) 방법에 의해 측정된다.
- [0039] 면역조직화학 염색은 대장암 조직을 채취 및 고정한 후, 당업계에서 널리 공지된 방법으로 파라핀 포매 블록을 제조하고, 이들을 일정 두께의 절편으로 만들어 유리 슬라이드에 붙인 후, 본 발명의 마커 단백질에 특이적인 상기의 항체와 공지의 방법에 의하여 반응시킴으로써 수행될 수 있다. 미반응 항체를 세척하고 검출라벨로 표지하여 현미경 상에서 항체의 표지 여부를 판독함으로써 대상 조직 내에서의 단백질 발현수준에 대한 정보를 얻을 수 있다.
- [0040] 본 발명의 가장 큰 특징 중 하나는 대장암의 분자 아형 분류를 위해 바이오마커에 대한 mRNA 발현 측정을 수행 하던 종래 기술과 달리, 면역조직화학 염색을 통해 간단한 육안 확인으로 발현(염색)정도를 등급화함으로써 시간과 노력과 비용을 크게 절감하면서도 정확한 분류 결과를 얻을 수 있다는 점이다.
- [0042] 본 발명의 구체적인 구현예에 따르면, 상기 대장암은 2-4기 대장암이며, 보다 구체적으로는 3기 대장암이다.
- [0043] 보다 구체적으로는, 상기 화학치료제는 옥살리플라틴, 5-플루오로우라실(5-FU), 류코보린 및 이들의 조합으로 구성된 군으로부터 선택된다.
- [0045] 본 발명의 구체적인 구현예에 따르면, 본 발명의 조성물은 ZMYND8, MUC5AC 및 이들의 조합으로 구성된 군으로부터 선택되는 단백질에 대한 항체를 유효성분으로 포함한다.
- [0046] 보다 구체적으로는, 본 발명의 조성물은 SNAI1, MUC2, SPINK4, CDX2, ZMYND8, TFF3, MUC5AC, HTR2B, MLH1, MSH2, MSH6 및 PMS2에 대한 항체를 유효성분으로 포함한다.
- [0047] 보다 구체적으로는, 본 발명의 조성물은 SNAI1, MUC2, SPINK4, CDX2, ZMYND8, TFF3, MUC5AC, KIT, MLH1, MSH2, MSH6 및 PMS2에 대한 항체를 유효성분으로 포함한다.
- [0049] 본 발명의 다른 양태에 따르면, 본 발명은 다음의 단계를 포함하는 대장암 분자 아형(Molecular subtype)의 분류 방법을 제공한다:
- [0050] (a) 대장암 환자로부터 분리된 생물학적 시료로부터 SNAI1, MUC2, SPINK4, CDX2, ZMYND8, TFF3, MUC5AC, KIT, HTR2B, MLH1, MSH2, MSH6 및 PMS2로 구성된 군으로부터 선택되는 하나 이상의 단백질의 발현 수준을 측정하는 단계;
- [0051] (b) 단계 (a)에서 측정된 단백질의 발현 수준을 정규화하는 단계;
- [0052] (c) 단계 (b)에서 정규화된 단백질의 발현 수준을 이용하여 기준 분자 아형(Reference molecular subtype)에 대한 각 분자 아형의 로그값을 도출하는 단계; 및
- [0053] (d) 단계 (c)에서 도출된 각 분자 아형의 로그값을 통해 대장암의 분자 아형을 결정하는 단계.

- [0054] 본 발명에서 이용되는 마커 단백질, 분자 아형의 의미 및 면역조직화학 염색 등을 이용한 단백질의 발현 수준의 측정 방법에 대해서는 이미 상술하였으므로, 과도한 중복을 피하기 위해 그 기재를 생략한다.
- [0055] 본 발명에서 용어 “생물학적 시료”는 대장암 환자로부터 얻어지는, 대장암 조직 또는 대장암 세포를 포함하고 있는 모든 시료로서, 조직, 기관, 세포 또는 세포 배양액을 포함하나, 이에 제한되지 않는다.
- [0056] 본 발명의 구체적인 구현예에 따르면, 상기 단계 (a)는 ZMYND8, MUC5AC 또는 이들의 조합을 포함하는 단백질의 발현 수준을 측정함으로써 수행된다.
- [0057] 보다 구체적으로는, 상기 단계 (a)는 SNAI1, MUC2, SPINK4, CDX2, ZMYND8, TFF3, MUC5AC, HTR2B, MLH1, MSH2, MSH6 및 PMS2의 발현 수준을 측정함으로써 수행된다.
- [0058] 보다 구체적으로는, 상기 단계 (a)는 SNAI1, MUC2, SPINK4, CDX2, ZMYND8, TFF3, MUC5AC, KIT, MLH1, MSH2, MSH6 및 PMS2의 발현 수준을 측정함으로써 수행된다.
- [0059] 본 발명의 구체적인 구현예에 따르면, 상기 단계 (a)는 면역조직화학 염색(immunohistochemical staining) 방법에 의해 수행된다.
- [0060] 본 발명에 따르면, 단계 (a)를 통해 각 마커 단백질에 대한 면역조직화학 염색 결과가 수득되면, 이를 통해 나타난 단백질 발현 수준을 특정 정량값으로 정규화하여 분자 아형 분류에 적용한다. 본 명세서에서 용어 “정규화(normalization)”는 데이터 군 내에서 특정 데이터가 가지는 위치를 파악하기 위해 전체 구간을 일정 크기로 나눈 뒤 해당 구간에서의 위치를 정량값으로 나타내는 것을 의미한다. 구체적으로는, “정규화”는 일정 크기의 구간 내 각 데이터들의 위치에 대한 정량값을 기반으로 각 데이터를 0 또는 1로 변환함으로써 데이터 분포를 조정하는 과정을 의미한다.
- [0061] 보다 더 구체적으로는, 단백질 발현량의 정규화는 각 단백질에 대한 면역조직화학 염색 결과에 대한 H-점수를 기반으로 하여 아래와 같이 각 단백질의 발현 수준에 대해 0 또는 1의 값을 부여함으로써 수행된다:
- [0062] TFF3 = 0(H-점수 \geq 10) 또는 1(H-점수<10);
- [0063] SPINK4 = 0(H-점수<20) 또는 1(H-점수 \geq 20);
- [0064] KIT = 0(부재) 또는 1(존재);
- [0065] CDX2 = 0(H-점수 \geq 150) 또는 1(H-점수<150);
- [0066] ZMYND8 = 0(저) 또는 1(고);
- [0067] MUC2 = 0(H-점수<10) 또는 1(H-점수 \geq 10);
- [0068] MUC5AC = 0(H-점수<10) 또는 1(H-점수 \geq 10);
- [0069] SNAI1 = 0(H-점수<20) 또는 1(H-점수 \geq 20);
- [0070] HTR2B = 0(H-점수<100) 또는 1(H-점수 \geq 100); 및
- [0071] MLH1, MSH2, MSH6 및 PMS2 = 0 (초위성체 불안정성 낮음) 또는 1(초위성체 불안정성 높음).
- [0072] 본 명세서에서 용어 “H-점수(Histo-score)”는 시료 내 특정 단백질의 발현 수준을 정규화하기 위해 면역조직화학 염색 강도(intensity)를 반정량적(semiquantitative)으로 표현한 값을 의미한다. H-점수 방법은 IHC 해석을 위해 널리 쓰이는 방법으로 당업계에 잘 알려져 있다. 예를 들어, 상이한 염색 강도의 세포 비율에 따라 0-300 까지의 연속적인 H-점수를 각 케이스에 할당한다. 핵 또는 세포질 염색에 대해 0(염색 안됨), 1+(높은 배율에서 약한 염색 보임), 2+(중간 염색), 3+(강한 염색)의 4개의 점수를 매기고, 상이한 염색 강도를 보이는 세포의 비율은 시각적 평가에 기반하여 다음 식으로 계산된 값을 통해 H-점수를 결정할 수 있다: 1 x (1+ 세포의 %) + 2 x (2+ 세포의 %) + 3 x (3+ 세포의 %).
- [0073] 보다 더 구체적으로는, 상기 단계 (c)는 상기 단계 (b)에서 정규화된 각 단백질의 발현 수준 값을 하기 식 1-1 내지 식 1-3 또는 식 2-1 내지 식 2-3에 대입함으로써 수행된다:
- [0074] [식 1-1]
- [0075] $\ln[p(\text{MS2})/P(\text{MS1})] = -1.33 + (0.34)a + (-2.09)b + (-0.55)c + (0.94)d + (1.51)e + (-0.79)f + (1.89)g + (2.52)h + (2.55)i$;

- [0076] [식 1-2]
- [0077] $\ln[p(\text{MS3})/P(\text{MS1})] = (-0.4) + (2.05)a + (-2.34)b + (-1.61)c + (1.07)d + (1.8)e + (-0.91)f + (-0.61)g + (-15.8)h + (0.45)i$;
- [0078] [식 1-3]
- [0079] $\ln[p(\text{MS4})/P(\text{MS1})] = 3.18 + (1.31)a + (-2.39)b + (-2.96)c + (-0.39)d + (0.61)e + (-1.42)f + (-1.22)g + (-19.76)h + (-14.58)i$;
- [0080] (a=SNAI1, b=MUC2, c=SPINK4, d=HTR2B, e=CDX2, f=ZMYND8, g=TFF3, h=초위성체 불안정성(MSI) 상태, i=MUC5AC);
- [0081] [식 2-1]
- [0082] $\ln[p(\text{MS2})/P(\text{MS1})] = 286.6 + (-92.2)a + (-5.2)b + (-53.0)c + (-85.3)d + (-80.1)e + (-35.9)f + (-207.7)g + (46.1)h + (-344.4)i$;
- [0083] [식 2-2]
- [0084] $\ln[p(\text{MS3})/P(\text{MS1})] = 282.0 + (-91.0)a + (-4.5)b + (-52.4)c + (-83.7)d + (-78.9)e + (-32.2)f + (-141.9)g + (43.1)h + (-173.5)i$;
- [0085] [식 2-3]
- [0086] $\ln[p(\text{MS4})/P(\text{MS1})] = 283.9 + (-92.1)a + (-5.9)b + (-50.9)c + (-83.7)d + (-80.2)e + (-35.2)f + (-143.8)g + (47.7)h + (-366.0)i$;
- [0087] (a=TFF3, b=SPINK4, c=KIT, d=CDX2, e=ZMYND8, f=MUC2, g=MUC5AC, h=SNAI1, i=초위성체 불안정성(MSI) 상태).

발명의 효과

- [0089] 본 발명의 특징 및 이점을 요약하면 다음과 같다:
- [0090] (a) 본 발명은 대장암의 예후 예측용 조성물, 대장암의 분자 아형(Molecular subtype) 분류용 조성물 및 이를 이용한 대장암의 분자 아형 분류 방법을 제공한다.
- [0091] (b) 본 발명에서 발굴된 마커 단백질들의 발현 수준을 측정함으로써 대장암의 분자 아형을 정확히 분류하고, 분류된 분자 아형을 기반으로 대장암의 재발률, 환자의 전체 생존 기간, 무병 생존률 및 FOLFOX 등의 화학치료제에 대한 치료 반응성을 포함하는 예후를 높은 신뢰도로 예측할 수 있다.
- [0092] (c) 본 발명은 전사체 발현을 통한 종래의 분류 방법과 달리 섬유아세포 및 기질에 의해 결과에 영향을 받지 않을 뿐 아니라 염색 결과의 육안 확인만으로도 간편하게 분류가 가능하므로, 시간, 비용 및 정확성 측면에서 보다 효율적인 분자 아형 분류 및 예후 예측 방법으로 유용하게 이용될 수 있다.

도면의 간단한 설명

- [0094] 도 1은 선정된 마커들의 면역 프로파일을 보여주는 그림이다. 고발현 및 저발현 여부는 각 마커 특이적인 H-점수의 컷오프 값을 이용하여 결정하였다.
- 도 2는 연구 설계에 대한 모식도를 나타낸 그림이다. 코호트 1에서의 총 101개 직장암 조직 표본을 평가하여 분자 분류 및 IHC 분자 아형 분류자를 결정하였으며, 코호트 2를 IHC 분자 아형 분류자를 이용하여 확인하였다.
- 도 3은 대장암 조직의 mRNA 발현 프로파일에 대한 자율(Unsupervised) 계층적 군집 분석결과를 보여주는 그림이다. 자율 분류에 의해 4가지 특징적인 분자 아형이 동정되었다. 붉은색 및 초록색은 치료 중간값의 위 또는 아래임을 각각 나타낸다. 유전자 발현 프로파일에 기반하여, 35개의 정상 대장조직(붉은색) 및 101개 대장암 조직(파란색)을 완전히 분리하였다.
- 도 4는 콘센서스 군집 기반 NMF를 이용하여 분류한 4가지 분자 아형을 나타낸 그림이다. 도 4a는 종양의 1,764개 DEG를 비종양성 조직과 비교한 콘센서스 군집인 NMF 1형 군집을 나타낸 그림이다. 도 4b는 다른 종양 그룹

과 비교하여 하나의 그룹에 특이적인 1,538개 DEG를 가지는 NMF 2형 군집을 나타낸 그림이다. 도 4c 및 4d는 CRC assigner-30에서 추출한 30개 유전자의 mRNA 발현 패턴을 NMF 1형 군집(도 4c) 및 NMF 2형 군집(도 4d)에 따라 분석한 결과를 보여준다. NMF 1형 군집은 NMF 2형 군집에 비해 종양 아형 간 보다 구별되는 유전자 발현 패턴을 보였다.

도 5는 분자 아형에 따른 조직학적 특성을 보여주는 그림이다. 도 5a는 NMF-기반 분류 및 CMS 분류에 따른 각 아형의 주조직 특성을 나타낸 그림이다. 도 5b는 결합조직증식성(desmoplastic), 솔리드, 톱니모양, 유두모양 및 복합 관상 패턴의 다섯 가지 조직학적 특성에 대한 대표적인 이미지들을 보여준다.

도 6은 각각의 네 그룹에서 비종양성 대장 점막 조직 및/또는 다른 그룹에 비해 상이하게 발현되는 mRNA 동정 결과를 보여준다. 도 6a는 특이적으로 발현이 증가한 총 292개의 유전자 마커를 보여주는 벤 다이어그램이다. 도 6b는 이들 292개 DEG의 mRNA 발현 프로파일에 대한 열지도를 나타낸다. 도 6c는 각 그룹에서 선정된 10개의 유전자들의 mRNA 발현 프로파일을 보여준다.

도 7은 IHC 분류자의 정확성 및 AIC를 보여주는 그림으로, 각각 CMS(도 7a) 및 NMF-기반 그룹(도 7b) 예측에 있어서의 정확도 및 AIC를 나타낸다.

도 8은 예측된 아형에 따른 조직학적 특성을 보여주는 그림이다. 도 8a는 NMF 기반 분류 및 CMS 분류에 따른 각 아형의 주조직 특성을 나타낸다. 도 8b는 결합조직증식성(desmoplastic), 솔리드, 톱니모양, 유두모양 및 복합 관상 패턴의 다섯 가지 조직학적 특성에 대한 대표적인 이미지들을 보여준다.

도 9는 코호트 1에서의 전사체 분석 및 IHC 분류자에 의해 분류된 분자 아형의 생존 분석을 비교한 결과이다. 도 9a는 mRNA 발현에 의해 분류된 CMS의 생존 분석이다(n=80). 도 9b는 IHC에 의해 예측된 CMS의 생존분석이다(n=80). 도 9c는 (C) mRNA 발현에 의해 분류된 NMF-기반 그룹의 생존분석이다(n=101). 도 9d는 IHC 분류자에 의해 분류된 NMF-기반 그룹의 생존분석이다(n=101).

도 10은 검증 코호트의 예측 아형에 대한 생존 분석 결과를 나타낸다. 도 10a는 검증 코호트의 예측 CMS에 대한 생존 분석 결과이고, 도 10b는 검증 코호트의 NMF-기반 그룹에 대한 생존 분석 결과이다.

도 11은 FOLFOX 보조제 치료를 받은 3기 직장암 환자에 대해 각 예측된 CMS 별로 생존 분석을 수행하고(도 11a, n=165), 2개 등급(CMS 1 및 2 vs CMS 3 및 4)으로 나누어 다시 생존분석을 하였다(도 11b, n=165).

도 12는 종양 중심과 침습 가장자리에서 각각 아형이 상이하게 예측됨을 나타낸 그림이다.

도 13은 케이스 1 - 43의 면역 프로파일을 보여주는 그림이다.

발명을 실시하기 위한 구체적인 내용

[0095] 이하, 실시예를 통하여 본 발명을 더욱 상세히 설명하고자 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.

[0097] 실시예

[0098] 실험방법

[0099] 임상 치료

[0100] 연구 코호트는 병리 기록으로부터 검색된 후향적 코호트를 이용하였다. 트레이닝 코호트인 환자 코호트 1은 FOLFOX 화학치료를 받은 101명의 3기 직장암 환자(국소적 림프절 전이가 일어났으나 원격 전이는 일어나지 않음)로 구성되었다. 검증 코호트인 코호트 2는 1기에서 4기까지의 직장암 환자 401명으로 구성되었다.

[0101] 선정된 케이스들을 WHO 기준 4판(2010년)에 따라 검토하고 AJCC (American Joint Committee on Cancer) TNM 분류 시스템 7판에 따라 단계화하였다. 인구 데이터 및 임상병리적 특성은 의료 기록으로부터 수집하였다.

[0102] 코호트 1의 포르말린-고정 파라핀-포매된 (FFPE) 종양 조직 및 신선하게 동결된 조직을 2006년 1월부터 2012년 12월까지 세브란스 병원에서 외과적 절제술을 받은 대장직장 선암 환자로부터 획득하였다.

[0103] 코호트 2의 포르말린-고정 파라핀-포매된 (FFPE) 종양 조직을 2010년 1월부터 2012년 12월까지 세브란스 병원에

서 외과적 절제술을 받은 대장직장 선암 환자로부터 수득하였다. 조직 표본은 연세대학교 의과대학 병리학교실의 기록 보관소 및 과학기술부 한국과학재단 산하의 국립연구소재 은행 프로그램의 간암 표본 은행에서 수득하였다.

[0105] RNA 제작

[0106] 코호트 1의 101명의 직장암 환자 조직 및 35개의 상응하는 비종양성 대장 점막성 동결 조직 시료를 저온유지 장치를 이용하여 현미해부하고 분별하여 종양 함량을 향상시켰다. RNA 분리를 위한 절개 이전에, 슬라이드를 헤마톡실린-에오신으로 염색하여 >70%의 종양 세포를 가지는 시료 단면을 선택하고, 종양세포 함량이 70% 미만인 시료는 추가적으로 절개하여 종양 세포 수를 증가시켰다. 이러한 과정을 통해 본 발명에서 분석된 모든 시료는 >70%의 종양 세포를 포함함을 확인할 수 있었다. TRIzol 시약(Invitrogen Life Technologies, Carlsbad, CA, USA)을 이용하여 제조자의 프로토콜에 따라 총 RNA를 추출하였다. DNase 분해 및 다른 제거 과정 후에, RNA 시료를 정량한 뒤 사용 시까지 -80℃에서 보관하였다. 품질 관리를 위해, RNA 순도 및 완전성은 변성 젤 전기영동과 A260/280 비율 측정을 통해 평가하였다. 모든 시료는 2100 Bioanalyzer(Agilent Technologies, Santa Clara, CA, USA)를 이용하여 분석하였다. 모든 시료에서, RIN(RNA integrity number) 점수는 >9.5 였다.

[0108] 유전자 발현 분석

[0109] DNA 마이크로어레이 혼성화를 위해, 총 RNA와 동일한 양을 혼합함으로써 RNA를 추렵하고, 1.5 µg 총 RNA를 시작으로 바이오틴-표지된 cRNA 타겟을 합성하였다. 이중가닥 cDNA 합성은 Illumina® TotalPrep RNA 증폭 키트(Illumina, San Diego, CA, USA)를 이용하여 수행하고, 바이오틴-UTP-표지된 안티센스 RNA는 Ambion MEGAscript 키트(Ambion Life Technologies, Carlsbad, CA, USA)를 이용하여 인 비트로에서 전사하였다. 모든 표지 과정은 제조자의 프로토콜을 이용하여 수행하였다. 마이크로어레이 실험은 31, 332로 표시된 유전자를 나타내는 47, 231 프로브를 포함하는 HumanHT-12 v4 Sentrix 발현 BeadChip(Illumina)에서 수행하였다. Illumina Bead Station 500× 매뉴얼에 따라 표지된 cRNA를 비드 칩에 혼성화하고, 세척한 뒤 스캐닝하였다.

[0111] mRNA 유전자 발현 데이터 수득 및 통계적 분석

[0112] Illumina Genome Studio v2011.1 [유전자 발현 모듈 v1.9.0] 소프트웨어를 이용하여 미가공 데이터를 추출하고, 퀀타일(quantile) 정규화를 이용하여 발현 강도를 정규화하였다²⁰. 이러한 정규화된 강도에 기반하여, 종래에 보고된 통합적인 통계 방법을 통해 비종양성 대장 점막 조직과 대장 종양에서 상이하게 발현되는 유전자들을 결정하였다²¹. 요약하면, 스튜던트 *t*-검정 및 로그2-중양값-비율 검정의 두 가지 독립적인 검정을 수행하였다. 시료의 임의 순열(random-permutation)을 통해 수득한 유전자의 평균값은 다르지 않을 것이라는 가정의 경험적 분포를 이용하여 각 검정에서의 조정된 P-값을 계산하였다. 두 가지 검정에서 얻은 P-값을 조합하여 Stouffer 방법으로 전체적인 P-값을 계산하였으며²², 자율 계층적 군집화를 위해 마이크로어레이 미가공 데이터로부터 |배수 변화|>2 및 P-값<0.05를 만족하는 4,823개 DEG를 선정하였다. NMF 콘센서스 군집 분석을 위하여, 마이크로어레이 미가공 데이터로부터 |배수 변화|>2 및 P-값<0.01을 만족하는 1,764개의 DEG를 선정하였고, |배수 변화|>2 및 P-값<0.01을 만족하는 1,538개의 DEG를 자율 계층적 군집화에 의해 나누어진 종양 그룹으로부터 선정하였다. 유전자 발현 분석을 위해 |배수 변화|>1.5 및 P-값<0.01의 조건을 적용하여 CRC assigner-30에 속하는 DEG를 얻었으며, 이들은 PAM[최근접 축소 중심(nearest shrunken centroids)-기반 방법]²³으로 계산된 높은 점수를 가지는 30개 유전자의 그룹이다. 마지막으로, GO 생물학적 과정 및 KEGG(Kyoto Encyclopedia of Genes and Genomes) 경로를 동정하기 위해 DAVID 소프트웨어²⁴를 이용하여 상이하게 발현되는 유전자에 대한 증폭 분석을 수행하였다.

[0114] 분자 아형의 결정 및 공지된 콘센서스 분자 아형과의 비교

[0115] 분자 아형의 분류를 위해, 직장암 조직의 유전자 발현 프로파일을 이용한 계층적 군집과 음수 미포함 행렬 분해(NMF)²⁵⁻²⁷를 적용하였다. 요약하면, NMF 콘센서스 군집화 방법은 유전자-발현 매트릭스 A를 $A \sim WH$ 의 양성 입력

값의 두 개의 매트릭스에 대한 출력값으로 재분류한다. 매트릭스 W 는 $N \times k$ 의 크기를 가지고 매트릭스 H 는 $k \times M$ 의 크기를 가지며, k 는 M 보다 훨씬 작다. W 컬럼은 ‘메타 유전자’를 정의하며, w_{ij} 는 메타 유전자 j 에서 유전자 I 의 계수를 의미한다. 매트릭스 H 의 컬럼은 상응하는 샘플에서의 메타 유전자 발현 패턴을 나타내어 h_{ij} 는 샘플 j 에서 메타 유전자 I 의 발현 수준을 나타낸다. $A \sim WH$ 의 인수분해를 통해, 매트릭스 H 는 군집 내에 어떤 샘플이 포함될지를 결정하는 데에 이용될 수 있다: 샘플 j 는 h_{ij} 가 컬럼 j 에서 가장 큰 입력값인 경우 군집 I 에 위치시켰다²⁵. 피어슨 계수 상관 알고리즘에 기반한 계층적 군집에 대한 완전 연결(complete linkage)을 이용하여 상이하게 발현되는 유전자들을 선정하였고, 이를 MATLAB 소프트웨어를 이용하여 NMF에 이용하였다. 분자아형 분류는 코페네틱(cophenetic) 상관계수 값($coph = 0.8961$)^{27, 28}에 기반하여 결정된 최적 군집 수에 따라 콘센서스 군집-기반 NMF를 통해 수행하였다. 본 발명자들은 1,764개 DEG의 동일한 시료를 이용하여 NMF 군집 분석을 반복적으로 수행한 결과, 1형 군집에 의해 나누어진 4개 그룹이 재현이 가능한 것으로 확인되었다(93% 일치). 다른 방법으로 본 발명의 시료에 대한 아형 분석을 수행하기 위해, 유클리드(Euclidean) 거리 및 WGCNA(weighted gene co-expression network analysis)²⁹를 Ward 최소 분산법과 함께 적용하였다. WGCNA를 이용하여 본 발명의 마이크로어레이 데이터의 품질을 확인하였으며, 유클리드 거리에 따른 Ward 최소 분산법을 이용하여 101개의 대장 종양을 클러스터링하였다. 본 발명의 분자 아형과 종래에 보고된 CMS를 비교하기 위해, 랜덤 포레스트 분류자와 단일-샘플 예측기’(SSP) 분류자를 포함하는 CMS 분류자¹⁸를 이용하였다. 나아가, ‘CMS 분류자’를 이용하여 본 발명의 분자 아형과 CMS 아형에 대한 예후 예측 값을 비교 분석하였다.

[0117] 초위성체 불안정성 분석

[0118] 초위성체 불안정성 분석을 위해, 분자 시험과 면역조직화학 염색 시험을 모두 수행하였다(표 1). 분자 분석을 위해 BAT26, BAT25, D5S346, D2S123 및 D17S250 마커로 구성된 NCI 패널을 시험하였다. PCR 산물을 모세관 전기영동으로 분석하였다. 결과 해석을 위해, 하나보다 많은 좌위에서의 불안정성을 MSI-H로, 단일 좌위에서의 불안정성을 낮은 MSI(MSI-L)로, 어떠한 좌위에서도 불안정성이 나타나지 않는 경우를 MSS로 정의하였다. 면역조직화학 염색 시험을 위해 MLH1, MSH2, MSH6 및 PMS2 항체를 사용하였다. 최소 하나의 단백질에서의 발현 손실이 발생할 경우 MSI-High로 해석하였다.

[0120] KRAS 및 BRAF 돌연변이 분석

[0121] KRAS 돌연변이 분석은 CE-IVD Marked PyroMark KRAS 키트(QIAGEN, Hilden, Germany)를 이용하여 파이로시퀀싱을 통해 제조자의 프로토콜(Therascreen KRAS Pyro Kit Handbook, version 1, July 2011)에 따라 코호트 1에 대해 수행하였다. 각 샘플에 대해, 코돈 12 및 13에서의 돌연변이 분석을 위해 10 ng의 지놈 DNA를 사용하고, 코돈 61의 돌연변이 분석을 위해 또 다른 10 ng의 지놈 DNA를 사용하였다. BRAF V600E 돌연변이를 검출하기 위해 종래에 보고된 바 대로 파이로시퀀싱을 적용하였다³⁰.

[0123] 조직 마이크로어레이 구축

[0124] H&E 염색된 종양 슬라이드에서, 대표적인 부위를 선정하고 상응하는 지점을 파라핀 블록 표면에 표시하였다. 조직 마이크로어레이 매뉴얼에 따라, 선정된 부위에 구멍을 뚫고 3-mm 조직 코어를 6×5 수용 블록에 위치시켰다. Two 두 개의 종양 중심 조직 코어와 침습 변연(margin)을 보이는 한 개의 종양 주변 조직 코어를 추출하였다. 100개 케이스의 전이 림프절로부터 조직 코어를 추출하였다. 각 조직 코어에 임상병리적 데이터를 포함하는 데이터베이스와 연결된 고유의 조직 마이크로어레이(TMA) 위치 번호를 부여하였다.

[0126] IHC 분류자를 위한 면역조직화학

[0127] 파라핀-포매된 조직 블록을 4-mm 단면으로 절개하였다. Ventana XT 자동 염색기(Ventana, Tucson, AZ, USA)를 이용하여 면역조직화학 분석을 하였다. 하기 표 1의 항체를 이용하여 직장암에 대한 면역조직화학 염색을 수행하였다(표 1).

표 1

[0128] 분자적 분류에 기반하여 선정된 IHC 마커들의 정보

| 타겟 CMS | 타겟 NMF 그룹 | 마커 | 포함 | | 제외 | IHC 마커 최종 선정 여부 |
|----------------------------|-----------|--------|-------------|------------|----------------------|-----------------|
| | | | 특이적 mRNA 발현 | 종래 연구에서 제안 | | |
| 3 ¹² | 1 | SPINK4 | ○* | ○ | | ○ |
| 3 ^{17, 31-33} | 1 | MUC2 | ○ | ○ | | ○ |
| 3 | 1 | DEFA5 | ◎** | | 유의하지 않은 IHC 발현 | |
| 1, 3 | 1, 2 | MUC5AC | ◎ | | | ○ |
| 1 | 2 | FAM81A | ◎ | | 유의하지 않은 IHC 발현 | |
| 1 ^{17, 31} | 2 | TFF3 | ○ | ○ | | ○ |
| 1, 4 ^{13, 31, 32} | 2, 3 | CDX2 | ○ | ○ | | ○ |
| 1 ^{17, 18} | 2 | MLH1 | | ○ | | ○ |
| 1 ^{17, 18} | 2 | MSH2 | | ○ | | ○ |
| 1 ^{17, 18} | 2 | MSH6 | | ○ | | ○ |
| 1 ^{17, 18} | 2 | PMS2 | | ○ | | ○ |
| 1 | 2 | MOCOS | ◎ | | IHC 염색 실패 | |
| 4 | 3 | HOPX | ◎ | | 유의하지 않은 IHC 발현 | |
| 4 ^{13, 32} | 3 | HTR2B | ○ | ○ | | ○ |
| 4 | 3 | MYL9 | ◎ | | 유의하지 않은 IHC 발현 | |
| 4 | 3 | TAGLN | ◎ | | 유의하지 않은 IHC 발현 | |
| 4 ^{17, 32} | 3 | ZEB1 | | ○ | 유의하지 않은 IHC 발현 | |
| 4 ¹⁷ | 3 | SNAI1 | ○ | ○ | | ○ |
| 4 ³⁴ | 3 | NOX4 | ○ | ○ | 유의하지 않은 IHC 발현 | |
| 4 ³⁵ | 3 | PDGFRA | ○ | ○ | 유의하지 않은 IHC 발현 | |
| 4 ³⁵ | 3 | PDGFRB | ○ | ○ | 유의하지 않은 IHC 발현 | |
| 4 ³⁵ | 3 | PDGFC | ○ | ○ | 유의하지 않은 IHC 발현 | |
| 4 ³⁵ | 3 | KIT | ○ | ○ | | ○ |
| 4 ³² | 3 | KRT | | ○ | 유의하지 않은 IHC 발현 | |
| 4 | 3 | PRRX1 | ◎ | | Failure of IHC stain | |
| 4 | 3 | SPARC | ◎ | | Failure of IHC stain | |
| 4 | 3 | FRMD6 | ○ | ○ | Failure of IHC stain | |
| 2 | 4 | DSC3 | ◎ | | 유의하지 않은 IHC 발현 | |
| 2 | 4 | ZMYND8 | ◎ | | | ○ |
| 2 | 4 | ACSL6 | ◎ | | 유의하지 않은 IHC 발현 | |
| 2 | 4 | TSPAN6 | ◎ | | 유의하지 않은 IHC 발현 | |
| 2 ^{17, 18} | 4 | CTNNB1 | | ○ | 유의하지 않은 IHC 발현 | |

*○: >1.0 배의 그룹 특이적 발현 **◎: >1.5 배의 그룹 특이적 발현 및 p-값 <0.05, MSI 상태는 MLH1, MSH2, MSH6, 및 PMS2로 분석함, 1차 항체 정보: SPINK4(HPA007286, Sigma-Aldrich, St. Louis, MO, USA, 1/100), MUC2(MRQ-18, Cell marque Rocklin, CA, USA, 1/100), DEFA5(HPA015775, Sigma-Aldrich, St. Louis, MO, USA, 1/400), MUC5AC(MRQ-19, Cell marque Rocklin, CA, USA, 1/100), TFF3(15C6, Merck Millipore, Darmstadt, Germany, 1/100), FAM81A(HPA001847, Sigma-Aldrich, St. Louis, MO, USA, 1/100), MLH1(554073, BD Biosciences, San Jose, CA, USA, 1/50), MSH2(556349, BD Biosciences, San Jose, CA, USA, 1/200), MSH6(44, Cell Marque Rocklin, CA, USA, 1/100), PMS2(MRQ-28, Cell Marque Rocklin, CA, USA, 1/40), HOPX(SC-30216, Santa Cruz, Dallas, TX, USA, 1/100), HTR2B(HPA012867, Sigma-Aldrich, St. Louis, MO, USA, 1/400), TAGLN(HPA019467, Sigma-Aldrich, St. Louis, MO, USA, 1/25), ZEB1(HPA027524, Sigma-Aldrich, St. Louis, MO, USA, 1/500), SNAIL(LS-C176686-100, LifeSpan Biosciences, Seattle, WA, USA, 1/100), NOX4(ab133303, Abcam, Cambridge, United Kingdom, 1/100), PDGFRA(ab65258, Abcam, Cambridge, United Kingdom, 1/100), PDGFRB(ab69506, Abcam, Cambridge, United Kingdom, 1/400), PDGFC(ab200401, Abcam, Cambridge, United Kingdom, 1/200), CKIT(ab5505, Abcam, Cambridge, United Kingdom, 1/400), KRT[CK(AE1/AE3), M3515, Dako, Glostrup, Denmark, 1/600], CDX2(EPR2764Y, Cell Marque Rocklin, CA, USA, 1/400), DSC3(61093, PROGEN Biotechnik, Heidelberg, Germany, 1/50), ZMYND8(HPA020949, Sigma-Aldrich, St. Louis, MO, USA, 1/800), ACSL6(HPA040470, Sigma-Aldrich, St. Louis, MO, USA, 1/50), TSPAN6(HPA004109, Sigma-Aldrich, St. Louis, MO, USA, 1/100), CTNNB1(610153, BD Biosciences, San Jose, CA, USA, 1/200), CD3(IS503 LOT 20017497, Dako, Glostrup, Denmark, prediluted), CD8(C8/144B, Dako, Glostrup, Denmark, prediluted), PRRX1 (NBP2-13816; Novus Biologicals, Centennial, CO, USA), LUM(HPA001522, Sigma-Aldrich, St. Louis, MO, USA), MOCOS (ab150852, Abcam, Cambridge, United Kingdom), RPL8(HPA045095, Sigma-Aldrich, St. Louis, MO, USA), SPARC(33-5500, Invitrogen, Carlsbad, CA, USA), VCAN(MABT16, Merck Millipore, Darmstadt, Germany), FRMD6(CSB-PA822270ESR2HU, cusabio, Houston, TX, USA)

[0130] 분자 아형 간 상이한 mRNA 발현을 보이는 마커 중에서, IHC에 사용 가능한 13개 마커가 후보 마커로 포함되었다. 분자 아형 간 상이한 발현 양상을 보여 CMS 분류용 마커로서 종래에 제안되었던 추가적인 19개 마커도 포함되었다. MLH1, MSH2, MSH6 및 PMS2를 사용하여 MSI 상태를 평가하였다. 총 32개 유전자를 선정하고 IHC를 수행하였다. 이들 중 염색 실패 또는 불충분한 IHC 발현으로 인해 19개 마커는 제외되었다. 최종적으로, IHC 아형 분류자를 확립하는 데에 SPINK4, MUC2, MUC5AC, TFF3, CDX2, HTR2B, KIT, SNAI1, ZMYND8 및 MSI 상태를 선정하였다(도 1).

[0131] IHC 해석을 위해 널리 쓰이는 H-점수 방법을 적용하였다. 이 방법은 상이한 염색 강도의 세포 비율에 근거하여 0-300까지의 연속적인 IHC H-점수를 각 케이스에 할당한다. 핵 또는 세포질 염색은 4개의 카테고리로 점수를 매겼다: ‘염색 안됨’은 0, ‘높은 배율에서 약한 염색 보임’은 1+, ‘중간 염색’은 2+, ‘강한 염색’은 3+. 상이한 염색 강도를 보이는 세포의 비율은 시각적 평가에 기반하여 다음의 식으로 계산된 값을 통해 결정하였다: $1 \times (1+ \text{세포의 } \%) + 2 \times (2+ \text{세포의 } \%) + 3 \times (3+ \text{세포의 } \%)$. 이후 염색 결과의 해석은 항체 특이적 컷-오프 값에 따라 단백질의 고발현 또는 저발현으로 분류되었다. 각 항체의 컷오프 값은 다음과 같다: SPINK4: 20, MUC2: 10, MUC5AC: 10, FAM81A: 100, TFF3: 10, HTR2B: 100, SNAI1: 20, CDX2: 150, ACSL6: 100, PDGFRB: 100, HOPX: 200, 및 ZEB1: 70. 컷오프 값보다 높은 H-점수 값은 고발현으로 간주하였고, 컷오프 값보다 낮은 값은 저발현으로 간주하였다. ZMYND8에 대해서, 정상 점막에 비해 염색 강도가 낮을 경우 저발현으로 간주하였으며 그렇지 않은 경우 고발현으로 간주하였다. KIT, DEFA5, DSC3, TAGLN, NOX4, TSPAN6, PDGFRA 및 PDGFC에 대하여 단백질의 여하한 발현도 양성(positive)으로 분류하고, 나머지는 음성(negative)으로 간주하였다. ZMYND8의 경우, 강도가 정상 선(gland)에 비해 약하면 저발현으로, 동일하거나 강하면 고발현으로 간주하였다. CTNNB1(β -카테닌)의 경우, 10% 이상 핵에서 발현하는 경우 양성으로, 그렇지 않은 경우 음성으로 분류하였다.

[0133] 컷-오프 중앙값을 이용하여 측정된 면역점수

[0134] CD3 양성 또는 CD8 양성 T 림프구 밀도를 측정하기 위해, 모든 염색된 TMA 슬라이드를 VENTANA iScan HT 슬라이드 스캐너(Ventana Medical Systems)를 이용하여 $\times 200$ 배율로 스캐닝하였다. 디지털 슬라이드를 VENTANA Virtuoso Digital Pathology Image Analysis software Version 5.3 (Ventana Medical Systems, Inc, Sunnyvale, CA)으로 분석하였다. ImageJ 플랫폼(<http://rsb.info.nih.gov/ij/index.html>)의 Cell Counter plug-in을 이용하여 $\times 200$ 역장에서 무작위로 선정된 3개 지점을 카운팅함으로써 염색된 세포 수를 정량화(세포 수/mm²)하였다. 중앙 중심 및 침습 가장자리의 CD3 또는 CD8 양성 림프구 수를 따로 측정하였다. 중앙 중심에서 카운팅한 CD3+ T 세포 (CD3 TC), 침습 가장자리에서 카운팅한 CD3+ T (CD3 IM), CD8+ T 세포 count at the 중앙 중심에서 카운팅한 CD8+ T 세포(CD8 TC), 그리고 침습 가장자리에서 카운팅한 CD8+ T 세포 (CD8 IM)를 전

체 코호트 케이스의 평균 값에 따라 높음(high) 또는 낮음(low)으로 분류하였다. 4개의 면역 파라미터의 평균 백분위수를 낮음(low), 중간(intermediate), 높음(high)으로 분류하였다(표2)^{36,37}

표 2

면역점수 계산의 기준

| CD3 TC | CD3 IM | CD8 TC | CD8 IM | 4개 마커의 평균 |
|--------|--------|--------|--------|-----------|
| High | High | High | High | 높은 면역점수 |
| High | High | High | Low | |
| High | High | Low | High | |
| High | Low | High | High | |
| Low | High | High | High | |
| High | High | Low | Low | 중간 면역점수 |
| High | Low | High | Low | |
| High | Low | Low | High | |
| Low | High | High | Low | |
| Low | High | Low | High | |
| Low | Low | High | High | 낮은 면역점수 |
| High | Low | Low | Low | |
| Low | High | Low | Low | |
| Low | Low | High | Low | |
| Low | Low | Low | High | |
| Low | Low | Low | Low | |

TC: 종양 중심, IM: 침습 가장자리

조직학적 파라미터 해석

가. 종양 기질 비율(TSP)

종래 보고된 방법으로, 종양 침습의 가장 깊은 지점의 단면을 H&E-염색하여 TSP의 측정을 수행하였다³⁸. ×4 대물렌즈를 이용하여 각 슬라이드에서 가장 침습적인 종양 부위를 선정하였다. ×10 대물렌즈를 이용하여 종양 및 기질 조직을 모두 사용할 수 있는 샘플의 일부를 선정하였다. 종양 세포는 이미지 필드의 모든 경계(북-동-남-서)에서 존재해야 한다. 괴사하거나 무신이 분포된 부위는 필드에서 제외되었다. 종양 비율에 대한 기질 비율이 TSP≥50%이면 높은 TSP로, 그렇지 않으면 낮은 TSP로 간주하였다.

나. Klintrup-M€akinen (KM) 등급

일반적인 염증세포의 침투는 종래에 보고된 바와 같이 KM 등급을 이용하여 평가하였다³⁹. 요약하면, 침습 가장 자리에서의 일반적인 염증 세포 침윤은 4-포인트 스케일(0: 증가 없음, 1: 염증 세포에서 조금 또는 불균등한 증가, 2: 침습 가장자리 밴드를 형성하는 현저한 염증 반응, 3: 암세포 섬(island)을 파괴하면서 침습 테두리 부분에서의 불그스름한 컵 모양의 침윤이 있음)을 이용한 반-정량적 방법으로 H&E(Haematoxylin and Eosin) 염색된 슬라이드에서 조사하였다. 통계적 분석을 위해 환자들을 순차적으로 저(low) 등급 (0/1) 또는 고(high) 등급 (2/3)으로 분류하였다.

다. 글래스고 미세환경 점수(Glasgow Microenvironment Score, GMS)

GMS는 종래 보고된 방법에 따라 TSP 및 KM 등급을 조합하여 평가하였다⁴⁰. 구체적인 배점 시스템은 다음과 같다. GMS 0: 강한 KM 등급과 높거나 낮은 TSP, GMS 1: 낮은 KM 등급과 높거나 낮은 TSP, GMS 2: 낮은 KM 등급과 높은 TSP.

IHC 분류자의 구축

[0147] 다중 회귀모델에 있어 세 개의 독립적인 이분형 로지스틱 회귀 모델을 가동함으로써 다항 로지스틱 모델을 4개의 가능한 결과물에 대한 통계적 방법으로 적용하였다. CMS1 또는 NMF1를 "기준"으로 선정하고 결과물을 기준 결과물에 대해 개별적으로 회귀시켰다.

$$\ln\left(\frac{P(CMS2)}{P(CMS1)}\right) = \beta_{20} + \beta_{21}(TFF3) + \beta_{22}(SPINK4) + \beta_{23}(KIT) + \beta_{24}(CDX2) + \beta_{25}(ZMYND8) + \beta_{26}(MUC2) + \beta_{27}(MUC5AC) + \beta_{28}(MSIstatus) + \beta_{29}(Snail_nuclear)$$

$$\ln\left(\frac{P(CMS3)}{P(CMS1)}\right) = \beta_{30} + \beta_{31}(TFF3) + \beta_{32}(SPINK4) + \beta_{33}(KIT) + \beta_{34}(CDX2) + \beta_{35}(ZMYND8) + \beta_{36}(MUC2) + \beta_{37}(MUC5AC) + \beta_{38}(MSIstatus) + \beta_{39}(Snail_nuclear)$$

$$\ln\left(\frac{P(CMS4)}{P(CMS1)}\right) = \beta_{40} + \beta_{41}(TFF3) + \beta_{42}(SPINK4) + \beta_{43}(KIT) + \beta_{44}(CDX2) + \beta_{45}(ZMYND8) + \beta_{46}(MUC2) + \beta_{47}(MUC5AC) + \beta_{48}(MSIstatus) + \beta_{49}(Snail_nuclear)$$

[0148]

[0149] CMS 분류를 위해, CMS 1을 기준으로 선정하고 3개의 식을 수립하였다.

$$[0150] \ln[p(CMS2)/P(CMS1)] = 286.6 + (-92.2)a + (-5.2)b + (-53.0)c + (-85.3)d + (-80.1)e + (-35.9)f + (-207.7)g + (46.1)h + (-344.4)i$$

$$[0151] \ln[p(CMS3)/P(CMS1)] = 282.0 + (-91.0)a + (-4.5)b + (-52.4)c + (-83.7)d + (-78.9)e + (-32.2)f + (-141.9)g + (43.1)h + (-173.5)i;$$

$$[0152] \ln[p(CMS4)/P(CMS1)] = 283.9 + (-92.1)a + (-5.9)b + (-50.9)c + (-83.7)d + (-80.2)e + (-35.2)f + (-143.8)g + (47.7)h + (-366.0)i;$$

a=TFF3 (0:H-점수 ≥ 10, 1:H-점수 < 10), b=SPINK4 (0:H-점수 < 20, 1:H-점수 ≥ 20), c=KIT(0:부재, 1:존재), d=CDX2(0:H-점수 ≥ 150, 1:H-점수 < 150), e=ZMYND8(0:저(low), 1:고(high)), f=MUC2(0:H-점수 < 10, 1:H-점수 ≥ 10), g=MUC5AC(0:H-점수 < 10, 1:H-점수 ≥ 10), h=SNAI1(0:H-점수 < 20, 1:H-점수 ≥ 20), i=MSI 상태 (0:MSS/MSI-low, 1:MSI-High).

[0153] NMF 분류에서, NMF1를 기준으로 선정하였다.

$$[0154] \ln[p(NMF2)/P(NMF1)] = -1.33 + (0.34)a + (-2.09)b + (-0.55)c + (0.94)d + (1.51)e + (-0.79)f + (1.89)g + (2.52)h + (2.55)i;$$

$$[0155] \ln[p(NMF3)/P(NMF1)] = (-0.4) + (2.05)a + (-2.34)b + (-1.61)c + (1.07)d + (1.8)e + (-0.91)f + (-0.61)g + (-15.8)h + (0.45)i;$$

$$[0156] \ln[p(NMF4)/P(NMF1)] = 3.18 + (1.31)a + (-2.39)b + (-2.96)c + (-0.39)d + (0.61)e + (-1.42)f + (-1.22)g + (-19.76)h + (-14.58)i;$$

$$[0157] a=SNAI1 (0:H-점수 < 20, 1:H-점수 ≥ 20), b=MUC2(0:H-점수 < 10, 1:H-점수 ≥ 10), c=SPINK4 (0:H-점수 < 20, 1:H-점수 ≥ 20), d=HTR2B(0:H-점수 < 100, 1:H-점수 ≥ 100), e=CDX2(0:H-점수 ≥ 150, 1:H-점수 < 150), f=ZMYND8(0:low, 1:high), g=TFF3 (0:H-점수 ≥ 10, 1:H-점수 < 10), h=MSI 상태(0:MSS/MSI-low, 1:MSI-High), i=MUC5AC(0:H-점수 < 10, 1:H-점수 ≥ 10). 위의 식으로부터 CMS 및 NMF 분류에 대한 예측 가능성을 계산하였다.$$

[0159] 통계적 분석

[0160] 돌연변이, 단백질 발현 및 유전자 카피 수와 임상적 특성 간의 관련성을 분석하기 위해 피어슨 카이 제곱(Chi square) 및 피셔 정확 검정(exact test)을 이용한 통계적 분석을 수행하였다. 각 그룹의 생존율은 Kaplan-Meier 방법을 이용하여 측정하였다. 생존 곡선의 동등성은 로그-순위(log-rank) 통계를 이용하여 시험하였다. 일변량 및 다변량 생존 분석을 위해 Cox 비례위험모델을 이용하였다. 일변량 생존 분석에서 발견되는, 통계적 유의성을 가지는 이들 변수는 다변량 생존 분석에서 평가하였다. 모든 결과 값은 양측 검정 p-값 < 0.05인 경우 통계적으로 유의한 것으로 간주하였다. 통계적 계산은 통계 패키지 SPSS 21.0(IBM Corp., Armonk, NY)를 이용하여 수행하였다.

[0162] 윤리 규정

[0163] 본 연구는 세브란스 병원 생명윤리위원회의 승인 하에 윤리 가이드라인 및 헬싱키 선언을 준수하면서 수행되었으며, 분리된 인체 유래 시료는 환자 동의 하에 기부되었다.

[0165] 실험결과

[0166] 직장암 환자의 임상병리적 특성

[0167] 트레이닝 코호트인 환자 코호트 1은 FOLFOX 투여를 받은 101명의 3기 직장암 환자로 구성되고, 검증 코호트인 코호트 2는 1-4기의 직장암 환자 401명으로 구성되었다. 이들의 임상병리적 특성은 표 3에 정리되었다. 코호트 1이 코호트 2에 비해 더 진행된 단계의 암환자가 많았기 때문에, 코호트 1의 직장암이 종양 크기가 더 크고 보다 침습적인 침윤 패턴을 보였다. 평균 환자 나이는 코호트 1이 코호트 2보다 어렸다.

[0168] 계약적인 실험 설계는 도 2에 나타내었다. 코호트 1을 이용하여 mRNA 유전자 발현 프로파일을 분석하고 직장암을 각 분자 아형으로 분류하였다. 이후 아형 특이적 유전자 마커를 동정한 뒤 IHC 분류자 구축에 사용할 마커를 선정하였다. IHC 분류자 모델을 만든 뒤, 코호트 2를 IHC 분류자 모델에 적용하여 임상병리적 특성을 평가하였다.

표 3

[0169] 코호트 1 및 코호트 2 직장암 환자의 임상병리적 특성

| 카테고리 | 변수 | 코호트 1 (n=101) | | 코호트 2 (n=401) | | p-값 |
|---------------|---------|---------------|--------|---------------|--------|-------|
| 나이(년) | | 58.98±11.17 | | 62.21±11.92 | | 0.014 |
| 성별 | 남성 | 51 | 50.50% | 243 | 60.60% | 0.071 |
| | 여성 | 50 | 49.50% | 158 | 39.40% | |
| 위치 | 오른쪽 | 36 | 35.64% | 186 | 46.38% | 0.057 |
| | 왼쪽 | 65 | 64.36% | 215 | 53.62% | |
| Preop CEA수준 | ≤5ng/ml | 73 | 72.28% | 276 | 68.83% | 0.547 |
| | >5ng/ml | 28 | 27.72% | 125 | 31.17% | |
| 육안 형태 | 돌출형 | 77 | 76.24% | 298 | 74.31% | 0.798 |
| | 비돌출형 | 24 | 23.76% | 103 | 25.69% | |
| 크기 (cm) | | 5.11±1.96 | | 4.68±2.27 | | 0.06 |
| 분화 | 양호 | 5 | 4.95% | 13 | 3.24% | 0.581 |
| | 중간 | 89 | 88.12% | 355 | 88.53% | |
| | 불량 | 7 | 6.93% | 33 | 8.23% | |
| 뮤신 형성 | 없음 | 72 | 71.29% | 303 | 75.56% | 0.4 |
| | 국소적 | 25 | 24.75% | 76 | 18.95% | |
| | 광범위 | 4 | 3.96% | 22 | 5.49% | |
| Crohn 유사 림프반응 | 저밀도 | 68 | 67.33% | 272 | 67.83% | 0.906 |
| | 고밀도 | 33 | 32.67% | 129 | 32.17% | |
| 침습 패턴 | 확장 | 15 | 14.85% | 112 | 27.93% | 0.007 |
| | 침습적 | 86 | 85.15% | 289 | 72.07% | |

[0171] mRNA 발현 프로파일의 자율 군집 분석에 의한 4가지 분자 아형의 동정

[0172] 31,332개 유전자에 대한 47,323개 프로브를 포함하는 Human HT-12 v4 Expression BeadChip (Illumina, Inc., San Diego, CA, USA)을 이용하여 101개의 직장암 및 35개 비종양성 대장점막 조직시료에 대해 마이크로어레이 분석을 함으로써 코호트 1의 mRNA 발현 프로파일을 분석하였다. 그 결과, 대장암에서 비종양성 대장 점막과 비교하여 상이하게 발현되는 유전자(DEG)로서 |배수 변화|>2 및 P-값<0.05을 만족하는 4,823개 유전자를 동정하였다. 대장암-특이적 유전자 발현에 대한 추가적 분석을 위하여, 4,823개 DEGI에 대한 자율 계층적 군집 분석을 수행한 결과 직장암 시료가 비종양성 대장 점막 조직과 구별되는 4개의 개별적인 클러스터로 그루핑됨을 발견하

였다(도 3).

[0174] 유전자 발현-기반 분자 아형은 음수 미포함 행렬 분해(NMF)에 의해 4개의 그룹으로 분류된다

[0175] 코호트 1의 101개 대장직장 종양을 유전자 프로파일에 따라 분류하기 위해, $|\text{배수 변화}| > 2$ 및 $P\text{-값} < 0.01$ 의 조건을 만족하는 1,764개의 종양-특이적 DEG를 이용하여 NMF 콘센서스 클러스터링을 수행하였다. NMF는 비-음수적 요소를 포함하여 데이터 매트릭스에 초점을 맞추는 음수 미포함 행렬 분해 알고리즘이다. 일치 정도가 높은 클러스터(코페네틱 계수(ρ_k)가 1.0에 근접)를 동정하기 위해, 분해 알고리즘에 다양한 조건($k=2, 3, 4, 5, 6$ 및 7)을 적용하였다.

[0176] 이후 $k=2$ 내지 7에서의 콘센서스 플롯과 코페네틱 계수를 획득하여 일치 정도가 높은(high consensus) k 값을 선정하였다. 1,764개 DEG에 대한 콘센서스 분석을 이용하여, $k=4$ 클러스터에 대한 양호한 콘센서스를 획득함으로써($\text{coph}=0.8961$) 101개 대장직장 종양의 DEG로부터 밝혀진 4개의 기능적으로 구별되는 특성을 가지는 4개의 콘센서스 클러스터(1형)가 존재함을 입증하였다; 그룹 1 ($n=22, 21.8\%$), 그룹 2 ($n=21, 20.8\%$), 그룹 3 ($n=16, 15.8\%$) 및 그룹 4 ($n=42, 41.6\%$)(도 4a). 콘센서스 매트릭스를 통해 그룹 1 및 4는 특성화된 클러스터로 나타난 반면(뚜렷한 콘센서스 군집), 그룹 2 및 3은 구별이 덜 뚜렷했다.

[0177] 본 발명자들은 다른 선정된 유전자 세트를 이용한 NMF 콘센서스 클러스터링도 수행하였다; 종양 그룹 간 상이하게 발현되는 1,538개 DEG를 $|\text{배수 변화}| > 2$ 및 $P\text{-값} < 0.01$ 로 계층적 군집화 방법을 이용하여 나누었다. NMF 콘센서스 분석 결과 높은 코페네틱 계수($\text{coph}=0.8612$)를 가지는 $k = 4$ 클러스터(2형)가 존재하였다; 그룹 1($n=20, 19.8\%$), 그룹 2($n=29, 28.7\%$), 그룹 3($n=24, 23.8\%$) 및 그룹 4($n=28, 27.7\%$)(도 4b). 1,764 또는 1,538개 DEG를 이용하여 NMF를 통해 나누어진 각 그룹의 시료는 유사하였다(77.2% 일치).

[0178] 상이한 종양 아형의 유전자 발현 특성을 추가적으로 조사하기 위해, 1형 및 2형 군집으로부터 $|\text{배수 변화}| > 1.5$ 및 $P\text{-값} < 0.01$ 의 조건 하에 각각 얻어진 4개의 아형에서 CRC assigner-30¹⁷으로 알려진 대장암 분류 시스템을 포함하는 30개 유전자의 발현 패턴을 분석하였다(도 4c 및 4d). 1형 군집에서 얻어진 4개의 종양 그룹에서, 5개 유전자의 발현이 특정 종양 그룹에서 정상 그룹에 비해 증가함을 발견하였다(도 4c의 좌측 패널). 나아가, 특정 종양 그룹에서 18개 유전자의 발현이 나머지 종양 그룹에 비해 증가하였음을 발견하였다(도 4c의 우측 패널). 2형 군집에서 획득한 4개의 종양 그룹에서는, 2개 및 17개 유전자가 특정 종양 그룹에서 정상 그룹 또는 나머지 종양 그룹에 비해 각각 발현이 증가하였다(도 4d). 종합하면, 1형 군집(비종양성 조직 대비 종양 조직에서의 1,764개 DEG를 이용한 NMF)은 2형 군집(자율 계층적 군집에 의해 나누어진 종양 그룹에서의 1,538개 DEG를 이용한 NMF)에 비해 보다 명확한 유전자 발현 패턴을 보였다. 따라서, 본 발명자들은 이후의 유전자 발현 분석 대상으로 1형 군집을 선정하였다.

[0180] NMF 기반 아형과 CMS 간의 관계

[0181] 다음으로, 101개 직장암 시료에 대한 CMS 분류를 위해 'CMS 분류자' (스크립트와 코드는 <https://github.com/Sage-Bionetworks/crcsc>에서 검색 가능)를 사용하였다^{17,18} 101개 시료 중에서 80개 종양이 각 CMS에 대응하고, 나머지 21개의 표지되지 않은 시료(비-콘센서스 시료)는 4개의 CMS 그룹 내에서 어떠한 일관된 패턴도 보이지 않았다. 본 발명자들은 CMS 그루핑과 본 발명의 아형을 비교한 결과, 101개 암 중 63개가 4개의 아형 중 하나와 밀접한 밀접한 관련성이 있음을 발견하였다(표 4).

표 4

[0183] NMF-기반 그룹과 CMS 아형 간의 관계

| | NMF-기반 그룹 | | | | 계 |
|--|-----------|------|------|------|---|
| | 그룹 2 | 그룹 4 | 그룹 1 | 그룹 3 | |

| | | | | | | |
|-----|------|-----------|-----------|----------|----------|-----|
| CMS | CMS1 | 10(66.6%) | 0(%) | 0(0%) | 0(0%) | 10 |
| | CMS2 | 4(26.7%) | 31(79.5%) | 3(18.8%) | 0(0%) | 38 |
| | CMS3 | 1(6.7%) | 0(0%) | 12(75%) | 0(0%) | 13 |
| | CMS4 | 0(0%) | 8(20.5%) | 1(6.2%) | 10(100%) | 19 |
| | 불명확 | 6 | 3 | 6 | 6 | 21 |
| | 계 | 21 | 42 | 22 | 16 | 101 |

[0185] 특히 그룹 2의 21개 시료 중 10개는 CMS1(초위성체 불안정성 면역 및 과돌연변이, 미세부수체 불안정, 강한 면역 활성화의 특성을 가짐)에 포함되었으며, CMS1으로 분류된 10개 시료 모두는 그룹 2로도 분류되었다. 유사하게, 그룹 4의 42개 시료 중 31개는 CMS2(정준이며 WNT 및 MYC 신호 활성화의 특징을 가짐)로 분류된 반면, CMS2의 38개 시료 중 31개는 그룹 4에서도 발견되었다. 나아가 그룹 1 시료 22개 중 12개가 CMS3(대사조절 장애의 특성을 가짐)으로 분류되었고, 13개의 CMS3 케이스 중 12개가 그룹 1에 속함을 발견하였다. 마지막으로, 그룹 3에 속하는 16개 시료 중 10개가 CMS4(간염이며 현저한 TNF 활성화, 기질 침습 및 혈관신생의 특징을 가짐)로 분류된 반면, CMS4 시료 19개 중 10개가 그룹 3에 속함을 발견하였다. 21개의 비-콘센서스 케이스를 제외시키자, CMS 분류 및 본 발명의 분자 분류 간 79%의 상관성(80 케이스 중 63개)이 나타났다 .

[0187] 분자 아형에 따른 직장암의 임상병리적 특성

[0188] 본 발명자들은 불명확한 케이스를 제외하고 CMS에 따른 몇몇 임상병리적 및 분자적 특징을 비교한 결과(표 5), CMS 간에 임상병리적 파라미터가 유의한 차이를 보임을 발견하였다. 특히 종양 위치($P = 0.004$), 뮤신 형성($P < 0.001$), Crohn-유사 반응($P = 0.029$), 침습 패턴($P = 0.012$), TSP ($P = 0.041$), GMS($P = 0.017$), 면역점수($P < 0.001$) 및 초위성체 불안정성의 비율(MSI-High) ($P < 0.001$)에서의 차이가 두드러졌다. CMS1에서 종양 위치는 대부분 오른쪽인 반면, 다른 아형에서는 왼쪽에 위치한 종양이 주로 나타났다. 점액성 선암(세포외 뮤신 형성의 $> 50\%$)은 CMS3 및 CMS4에서 관찰되었다. CMS1에서, Crohn-유사 반응(60%), 높은 면역점수 및 확장형 침습 패턴(50%)이 다른 타입에서보다 빈번하게 발견되었다. 높은 TSP, 낮은 KM 등급 및 GMS2는 CMS4의 특징으로 나타났다. NMF-기반 그룹 간에도 몇몇 파라미터에서의 유의한 차이점이 발견되었는데, 특히 종양 위치($P=0.001$), 뮤신 형성($P < 0.001$), 종양 기질 비율(TSP)($P=0.018$), Klintrap-Makinen (KM) 등급($P=0.018$), 글래스고 미세환경 점수(GMS)($P=0.012$), 면역점수($P=0.002$), KRAS 돌연변이 발생율($P=0.01$) 및 초위성체 불안정성의 비율(MSI-High)($P=0.001$)에서 유의한 차이가 나타났다. 나아가, 그룹 3(62.5%) 및 4(85.7%)의 암의 $>50\%$ 은 좌측에 자리한 종양이었으나, 반대로 그룹2 (57.1%)의 암의 $>50\%$ 은 오른쪽에 위치하였다. 조직학적으로, 대부분의 종양은 어느 정도 분화된 선암의 형태로서, 4개의 점액성 선암($>50\%$ 이 세포외 뮤신 형성)이 존재하였으며 모두 그룹 1에 속했다. 세포외 뮤신 생성과 관련하여, 그룹 3(87.5%) 및 4 (92.9%)의 대부분의 종양은 세포외 뮤신을 생산하지 않은 반면, 그룹 1 및 2의 $>50\%$ 는 세포외 뮤신을 생산하였다. 그룹 3의 암의 절반 가량은 높은 종양 기질 비율(TSP)을 나타낸 반면, 다른 그룹들은 대부분 낮은 TSP를 보였다. 그룹 2의 거의 모든 케이스(95.24%)에서 높은 KM 등급을 보였고, 그룹 2 및 그룹 3에서 높은 면역점수가 나타났다. 그룹 3 종양의 약 30%는 GMS2, 약한 KM 등급 및 높은 TSP의 특성을 가졌다. 그룹 1(63.6%) 및 그룹 3(56.3%)에서 빈번한 KRAS 돌연변이를 발견하였으며, 그룹 2(28.6%) 및 4(26.2%)에서는 돌연변이 빈도가 보다 낮았다. 나아가, 그룹 2(28.6%) 및 그룹 1(9%)에서 MSI-High를 관찰하였으나 그룹 3 및 4에서는 관찰되지 않았다. MSI-High인 대장암에서 Crohn-유사 림프반응은 일반적인 특성이므로, Väyrynen-Mäkinen 기준에 의해 평가되었다. 그러나, 이러한 기준에 의할 때 상이한 아형 간 유의한 차이가 관찰되지 않았다.

표 5

[0189] CMS에 따른 코호트 1의 임상병리적 특성

| 카테고리 | 변수 | CMS 1 | | CMS 2 | | CMS 3 | | CMS 4 | | p-값 |
|--------|----|--------------|--------|---------------|--------|-------------|--------|--------------|--------|-------|
| | | n=10, 9.9% | | n=38, 37.6% | | n=13, 12.9% | | n=19, 18.8% | | |
| 나이 (년) | | 51.1 ± 15.15 | | 59.08 ± 11.33 | | 61 ± 9.69 | | 60.53 ± 9.82 | | 0.337 |
| 성별 | 남성 | 4 | 40.00% | 22 | 57.90% | 9 | 69.20% | 9 | 47.40% | 0.467 |
| | 여성 | 6 | 60.00% | 16 | 42.10% | 4 | 30.80% | 10 | 52.60% | |

| | | | | | | | | | | |
|-------------------------|-------------|-----------|---------|-----------|---------|-----------|---------|-----------|---------|--------|
| 위치 | 오른쪽 | 7 | 70.00% | 6 | 15.80% | 6 | 46.20% | 8 | 42.10% | 0.004 |
| | 왼쪽 | 3 | 30.00% | 32 | 84.20% | 7 | 53.80% | 11 | 57.90% | |
| Preop CEA 수준 | ≤5ng/ml | 10 | 100.00% | 26 | 68.40% | 11 | 84.60% | 15 | 78.90% | 0.189 |
| | >5ng/ml | 0 | 0.00% | 12 | 31.60% | 2 | 15.40% | 4 | 21.10% | |
| 육안 형태 | 돌출형 | 10 | 100.00% | 27 | 71.10% | 11 | 84.60% | 14 | 73.70% | 0.24 |
| | 비돌출형 | 0 | 0.00% | 11 | 28.90% | 2 | 15.40% | 5 | 26.30% | |
| 크기 (cm) | | 6.36±2.42 | | 4.88±1.66 | | 4.93±2.40 | | 4.62±1.32 | | 0.239 |
| 분화 | 양호 | 0 | 0.00% | 3 | 7.89% | 1 | 7.69% | 1 | 5.26% | 0.146 |
| | 중간 | 7 | 70.00% | 34 | 89.47% | 12 | 92.31% | 16 | 84.21% | |
| | 불량 | 3 | 30.00% | 1 | 2.63% | 0 | 0.00% | 2 | 10.53% | |
| 무신 형성 | 없음 | 4 | 40.00% | 34 | 89.50% | 6 | 46.20% | 17 | 89.50% | <0.001 |
| | 국소적 | 6 | 60.00% | 4 | 10.50% | 6 | 46.20% | 1 | 5.30% | |
| | 광범위(>50%) | 0 | 0.00% | 0 | 0.00% | 1 | 7.70% | 1 | 5.30% | |
| Crohn 유사 림프 반응 | 저밀도 그룹 | 4 | 40.00% | 31 | 81.60% | 7 | 53.80% | 11 | 57.90% | 0.029 |
| | 고밀도 그룹 | 6 | 60.00% | 7 | 18.40% | 6 | 46.20% | 8 | 42.10% | |
| 침습 패턴 | 확장 | 5 | 50.00% | 5 | 13.20% | 0 | 0.00% | 2 | 10.50% | 0.012 |
| | 침습적 | 5 | 50.00% | 33 | 86.80% | 13 | 100.00% | 17 | 89.50% | |
| 림프 관 침 윤 | 부재 | 4 | 40.00% | 17 | 44.70% | 6 | 46.20% | 10 | 52.60% | 0.938 |
| | 존재 | 6 | 60.00% | 21 | 55.30% | 7 | 53.80% | 9 | 47.40% | |
| 종양 발아 | 부재 | 3 | 30.00% | 6 | 15.79% | 2 | 15.38% | 1 | 5.26% | 0.543 |
| | 낮음 | 6 | 60.00% | 21 | 55.26% | 6 | 46.15% | 11 | 57.89% | |
| | 높음 | 1 | 10.00% | 11 | 28.95% | 5 | 38.46% | 7 | 36.84% | |
| TSP | 낮음 | 10 | 100.00% | 32 | 84.21% | 10 | 76.92% | 11 | 57.89% | 0.041 |
| | 높음 | 0 | 0.00% | 6 | 15.79% | 3 | 23.08% | 8 | 42.11% | |
| KM 등 급 | 낮음 | 0 | 0.00% | 13 | 34.21% | 3 | 23.08% | 8 | 42.11% | 0.078 |
| | 높음 | 10 | 100.00% | 25 | 65.79% | 10 | 76.92% | 11 | 57.89% | |
| GMS | GMS 0 | 10 | 100.00% | 25 | 65.79% | 10 | 76.92% | 11 | 57.89% | 0.017 |
| | GMS1 | 0 | 0.00% | 10 | 26.32% | 2 | 15.38% | 1 | 5.26% | |
| | GMS2 | 0 | 0.00% | 3 | 7.89% | 1 | 7.69% | 7 | 36.84% | |
| 면역 점수 | 저 | 2 | 20.00% | 23 | 60.53% | 6 | 46.15% | 10 | 52.63% | <0.001 |
| | 중간 | 0 | 0.00% | 14 | 36.84% | 5 | 38.46% | 5 | 26.32% | |
| | 고 | 8 | 80.00% | 1 | 2.63% | 2 | 15.38% | 4 | 21.05% | |
| BRAF | 야생형 | 9 | 90.00% | 38 | 100.00% | 13 | 100.00% | 19 | 100.00% | 0.125 |
| | 돌연변이 | 1 | 10.00% | 0 | 0.00% | 0 | 0.00% | 0 | 0.00% | |
| KRAS | 야생형 | 8 | 80.00% | 27 | 71.10% | 5 | 38.50% | 13 | 68.40% | 0.149 |
| | 돌연변이 | 2 | 20.00% | 1 | 28.90% | 8 | 61.50% | 6 | 31.60% | |
| MSI 상태 | MSS/MSI-low | 4 | 40.00% | 38 | 100.00% | 12 | 92.31% | 19 | 100.00% | <0.001 |
| | MSI-High | 6 | 60.00% | 0 | 0.00% | 1 | 7.69% | 0 | 0.00% | |

TSP:종양 기질 비율, KM 등급: Klintrap-Makinen 등급, GMS: Glasgow 미세환경 점수, MSS: 초위성체 안정성, MSI: 초위성체 불안정성

표 6

NMF 분자 아형에 따른 임상병리적 특성

| 카테고리 | 변수 | 그룹 1 | 그룹 2 | 그룹 3 | 그룹 4 | p-값 |
|-----------|----|-------------|-------------|-------------|-------------|-------|
| | | n=22, 21.8% | n=21, 20.8% | n=16, 15.8% | n=42, 41.6% | |
| 나이 (년) | | 57.95±13.26 | 58.09±13.67 | 62±9.04 | 58.81±9.42 | 0.802 |

| | | | | | | | | | | |
|-------------------------|-------------|-----------|---------|------------|--------|------------|---------|-----------|---------|--------|
| 성별 | 남성 | 14 | 63.60% | 8 | 38.10% | 7 | 43.80% | 22 | 52.40% | 0.376 |
| | 여성 | 8 | 36.40% | 13 | 61.90% | 9 | 56.20% | 20 | 47.60% | |
| 위치 | 오른쪽 | 12 | 54.50% | 12 | 57.10% | 6 | 37.50% | 6 | 14.30% | <0.001 |
| | 왼쪽 | 10 | 45.50% | 9 | 42.90% | 10 | 62.50% | 36 | 85.70% | |
| PreOP CEA 수 준 | ≤5ng/ml | 18 | 81.80% | 17 | 81.00% | 10 | 62.50% | 28 | 66.70% | 0.367 |
| | >5ng/ml | 4 | 18.20% | 4 | 19.00% | 6 | 37.50% | 14 | 33.30% | |
| 육안 형태 | 돌출형 | 20 | 90.91% | 18 | 85.71% | 11 | 68.75% | 28 | 66.67% | 0.102 |
| | 비돌출형 | 2 | 9.09% | 3 | 14.29% | 5 | 31.25% | 14 | 33.33% | |
| 크기 (cm) | | 5.5±2.535 | | 5.79±2.197 | | 5.01±1.224 | | 4.6±1.609 | | 0.171 |
| 분화 | 양호 | 2 | 9.09% | 0 | 0.00% | 1 | 6.25% | 2 | 4.76% | 0.507 |
| | 중간 | 20 | 90.91% | 18 | 85.71% | 14 | 87.50% | 37 | 88.10% | |
| | 불량 | 0 | 0.00% | 3 | 14.29% | 1 | 6.25% | 3 | 7.14% | |
| 무신 형성 | 없음 | 10 | 45.50% | 9 | 42.80% | 14 | 87.50% | 39 | 92.90% | <0.001 |
| | 국소적 | 8 | 36.40% | 12 | 57.20% | 2 | 12.50% | 3 | 7.10% | |
| | 광범위(>50%) | 4 | 18.20% | 0 | 0.00% | 0 | 0.00% | 0 | 0.00% | |
| Crohn 유사 림프 반응 | 저밀도 그룹 | 14 | 63.60% | 13 | 61.90% | 8 | 50.00% | 33 | 78.60% | 0.157 |
| | 고밀도 그룹 | 8 | 36.40% | 8 | 38.10% | 8 | 50.00% | 9 | 21.40% | |
| 침습 패턴 | 확장 | 2 | 9.10% | 7 | 33.30% | 2 | 12.50% | 4 | 9.50% | 0.098 |
| | 침습적 | 20 | 90.90% | 14 | 66.70% | 14 | 87.50% | 38 | 90.50% | |
| 림프관 침윤 | 부재 | 12 | 54.55% | 10 | 47.62% | 8 | 50.00% | 18 | 42.86% | 0.847 |
| | 존재 | 10 | 45.45% | 11 | 52.38% | 8 | 50.00% | 24 | 57.14% | |
| 종양 발아 | 부재 | 5 | 22.73% | 3 | 14.29% | 1 | 6.25% | 5 | 11.90% | 0.901 |
| | 낮은 | 11 | 50.00% | 11 | 52.38% | 10 | 62.50% | 23 | 54.76% | |
| | 높음 | 6 | 27.27% | 7 | 33.33% | 5 | 31.25% | 14 | 33.33% | |
| TSP | 저 | 18 | 81.82% | 19 | 90.48% | 8 | 50.00% | 36 | 85.71% | 0.018 |
| | 고 | 4 | 18.18% | 2 | 9.52% | 8 | 50.00% | 6 | 14.29% | |
| KM 등 급 | 저 | 6 | 27.27% | 1 | 4.76% | 6 | 37.50% | 18 | 42.86% | 0.018 |
| | 고 | 16 | 72.73% | 20 | 95.24% | 10 | 62.50% | 24 | 57.14% | |
| GMS | GMS 0 | 16 | 72.73% | 20 | 95.24% | 10 | 62.50% | 24 | 57.14% | 0.012 |
| | GMS1 | 4 | 18.18% | 1 | 4.76% | 1 | 6.25% | 12 | 28.57% | |
| | GMS2 | 2 | 9.09% | 0 | 0.00% | 5 | 31.25% | 6 | 14.29% | |
| 면역 점수 | 저 | 14 | 63.64% | 7 | 33.33% | 5 | 31.25% | 27 | 64.29% | 0.002 |
| | 중간 | 6 | 27.27% | 5 | 23.81% | 4 | 25.00% | 13 | 30.95% | |
| | 고 | 2 | 9.09% | 9 | 42.86% | 7 | 43.75% | 2 | 4.76% | |
| BRAF | 야생형 | 22 | 100.00% | 20 | 95.20% | 16 | 100.00% | 42 | 100.00% | 0.366 |
| | 돌연변이 | 0 | 0.00% | 1 | 4.80% | 0 | 0.00% | 0 | 0.00% | |
| KRAS | 야생형 | 8 | 36.40% | 15 | 71.40% | 7 | 43.80% | 31 | 73.80% | 0.01 |
| | 돌연변이 | 14 | 63.60% | 6 | 28.60% | 9 | 56.30% | 11 | 26.20% | |
| MSI 상 태 | MSS/MSI-low | 20 | 90.91% | 15 | 71.43% | 16 | 100.00% | 42 | 100.00% | 0.001 |
| | MSI-High | 2 | 9.09% | 6 | 28.57% | 0 | 0.00% | 0 | 0.00% | |

TSP:종양 기질 비율, KM 등급: Klintrap-Makinen 등급, GMS: Glasgow 미세환경 점수, MSS: 초위성체 안정성, MSI: 초위성체 불안정성

[0193]

복합 관상, 유두모양, 톱니모양, 솔리드 및 결합조직증식성(desmoplastic)의 다섯가지 조직학적 특성이 동정되었다. 일반적인 조직학적 특성은 분자 아형에 따라 동정하였다(도 5). 복합 관상 패턴은 모든 아형에서 빈번하게 나타났다. 복합 관상 패턴 외에, 그룹 1 및 CMS 3에서 유두모양 및 톱니모양 특성이 일반적으로 발견된 반면, 그룹 2 및 CMS 1에서는 솔리드 특성이 특징적이었다. 그룹 3 및 CMS4에서 결합조직을 형성하는 특성이 광범위하게 관찰되었으며, 그룹 4 및 CMS2에서는 복합 관상 형태 다음으로 유두 모양과 결합조직증식성 특성이 관찰되었다.

[0195] 아형 특이적 마커의 동정

[0196] 본 발명자들은 4개 그룹에서 대장 점막 조직 및/또는 다른 그룹과 비교하여 상이하게 발현되는 유전자군을 동정하기 위해 유전자 발현 패턴을 추가적으로 분석하였다. 각 그룹에서 상이하게 발현되는 유전자(그룹 1에서 194개 유전자, 그룹 2에서 736개 유전자, 그룹 3에서 493개 유전자, 그룹 4에서 634개 유전자)를 동정하였다. 나아가, 275개 유전자, 372개 유전자, 558개 유전자 및 184개 유전자가 각 그룹에서 다른 그룹에 비해 상이하게 발현되었다. 다음으로 정상 조직이나 다른 분자 그룹에 비해 상이하게 조절되는 유전자를 선발함으로써 각 그룹에서 특징적인 유전자를 선정하였으며, 그 결과 292개 유전자를 잠재적인 그룹 마커로 동정하였다(도 6a 및 6b). 마지막으로 이들 그룹 내에서 정상 조직 및 다른 종양 그룹에 비해 1.5배를 초과하여 과발현되는 10개의 유전자를 선정하였다. 이들 유전자 중에서, 그룹 1의 DEFA5, DUOX2, KLK12 및 ALDOB, 그룹 2의 CCDC58, MOCOS 및 FAM81A, 그룹 3의 HOPX, TAGLN, GREM1, THBS4, COL3A1, PRRX1, RAB31, MYL9, CTSK 및 SPARC, 그룹 4의 DSC3 및 CAB39L가 특히 발현이 증가하였다(>2.0배)(도 6c).

[0198] 분자 아형예측을 위한 IHC 아형 분류자의 확립

[0199] 본 발명자들은 별도로 분자아형을 예측할 수 있는 IHC 분류자 모델을 확립하고자 하였다. IHC 분류자 설계를 위해, NMF-기반 그룹에 대한 101개 케이스의 IHC 발현 데이터를, CMS에 대한 80개 케이스의 IHC 발현 데이터를 사용하고 명확하지 않은 케이스는 제외하였다. 모든 가능한 IHC 분류자의 정확도 및 AIC 값은 도 7에 표시하였다. 식별력 측면에서 높은 정확도 값을 가지면 더 좋은 모델로 간주하였으며, AIC 값도 식별력이 우수한 모델의 기준이 되었다. 정확도는 조합된 마커의 수가 증가할수록 개선되었다. 10개의 마커를 사용한 경우 평균 정확도가 0.79 임에 반하여, 두 가지 마커만이 조합되어 사용된 경우 평균 정확도는 약 0.55였다. 평균 AIC의 경우 반대로, 조합된 마커 수가 증가할수록 AIC는 감소하였다. 높은 정확도와 낮은 AIC 값을 가지는 IHC 분류자를 선정하였다(표 7). CMS를 예측하기 위해 MUC2, CDX2, ZMYND8, SPINK4, KIT, MUC5AC, Snail, TFF3 및 MSI 상태의 9개 인자를 선정하였다. 정확도 값은 0.825(95% CI = 0.737-0.825)이고 AIC 값은 134.771 (95% CI=134.507-160.549)이다. 4개의 CMS 중 예측 가능성이 가장 높은 CMS 그룹을 각 케이스의 CMS로 선정하였다. 그러나, 급내(intraclass) 상관관계수 (ICC)를 이용하여 분자 분류자와 IHC 분류자의 예측 가능성을 비교한 결과, 평균 ICC 는 0.18(P = 0.219)이었다.

[0200] NMF-기반 그룹에 대해 다음의 9개 인자를 선정하였다: SNAIL1, MUC2, SPINK4, HTR2B, CDX2, ZMYND8, TFF3, MUC5AC 및 MSI 상태. 이는 다른 마커의 조합에 비하여 정확도(0.702, 95% 신뢰구간[CI]=0.643-0.702)는 높았고 AIC 값(226.576, 95% CI = 226.576 - 237.358)은 낮았다.

표 7

[0202] 아형 예측을 위한 IHC 분류자

| | 예측 그룹 | 인자 수 | 결합된 인자 | 정확도(95% CI) | AIC (95% CI) |
|-------------------------------------|-----------|------|---|------------------------|------------------------------|
| IHC 분류자 | CMS | 9 | MUC2, CDX2, ZMYND8, MSI, SPINK4, KIT, MUC5AC, Snail, TFF3 | 0.825 (0.737-0.825) | 134.771 (134.507-160.549) |
| | NMF-기반 그룹 | 9 | Snail, MUC2, SPINK4, HTR2B, CDX2, ZMYND8, TFF3, MSI, MUC5AC | 0.702 (0.643-0.702) | 226.576 (226.576-237.358) |
| AIC: Akaike's Information Criterion | | | | | |

[0204] IHC 분류자로 예측된 아형과 전사체에 의한 분자 아형의 임상병리적 특성의 유사성

[0205] 검증 코호트에서 예측된 CMS는 임상병리적 특성에 있어 트레이닝 코호트와 유사한 패턴을 보였다(표 8). 401개 케이스 중 18.0%는 CMS1(72 케이스)로, 42.9%는 CMS2(172 케이스)로, 19.2%는 CMS3(77 케이스)로, 20.0%는 CMS4(80 케이스)로 분류되었다.

[0206] 코호트 1에서처럼 유의한 차이를 보이는 인자는 다음과 같다: 종양 위치($P < 0.001$), 림프관 침윤($P < 0.001$), Crohn-유사 반응($P < 0.001$), 침습 패턴($P = 0.011$), TSP($P \leq 0.001$), GMS ($P < 0.001$), 및 면역점수($P = 0.001$). 종양 위치는 CMS1에서 주로 오른쪽인 반면 CMS2 및 CMS4에서는 좌측 종양이 우세하였다. CMS1에서, 높은 면역점수(47.22%) 및 확장 침습 패턴(40.28%)과 함께, Crohn-유사 반응(63.89%)이 다른 타입에 비해 빈번하게 나타났다. 코호트 1에서처럼 CMS4에서도 높은 TSP와 낮은 KM 등급이 관찰되었다. 추가적인 인자들은 각 CMS 별로 주목할만한 차이를 보였다. 종양 크기는 CMS1에서 크고 CMS4($P < 0.001$)에서는 작았다. 종양 분화에 있어서, 제대로 분화되지 못한 선암이 CMS1에 많이 포함된 반면($P < 0.001$), 림프관 침윤 및 종양 발아는 CMS4에서 다수 발생하였다. 코호트 2에서는 점액성 선암(50% 이상의 세포외 림프관 침윤)이 CMS1에서 주로 관찰된 반면, 코호트 1에서는 CMS3 및 CMS4에서 주로 관찰되었다. MSI-High는 CMS1에서 많이 존재하였고(76.39%), 혼합 MSI 아형으로 알려진 CMS3에도 소량(3.9%) 존재하였다.

표 8

[0207] IHC 분류자에 의해 예측된 각 CMS들의 임상병리학적 특성

| 카테고리 | 변수 | CMS1 | | CMS2 | | CMS3 | | CMS4 | | p-값 |
|-----------------|-------------|-------------|--------|--------------|---------|-------------|--------|-------------|---------|--------|
| | | n=72, 18.0% | | n=172, 42.9% | | n=77, 19.2% | | n=80, 20.0% | | |
| 나이 | | 59.44±12.76 | | 62.69±10.76 | | 63.26±12.56 | | 61.66±12.25 | | 0.202 |
| 성별 | 남성 | 46 | 63.89% | 108 | 62.79% | 44 | 57.14% | 45 | 56.25% | 0.636 |
| | 여성 | 26 | 36.11% | 64 | 37.21% | 33 | 42.86% | 35 | 43.75% | |
| 위치 | 오른쪽 | 58 | 80.56% | 49 | 28.49% | 41 | 53.25% | 38 | 47.50% | <0.001 |
| | 왼쪽 | 14 | 19.44% | 123 | 71.51% | 36 | 46.75% | 42 | 52.50% | |
| Preop CEA 수준 | ≤5ng/ml | 54 | 75.00% | 114 | 66.28% | 57 | 74.03% | 51 | 63.75% | 0.298 |
| | >5ng/ml | 18 | 25.00% | 58 | 33.72% | 20 | 25.97% | 29 | 36.25% | |
| 육안 형태 | 돌출형 | 58 | 80.56% | 123 | 71.51% | 59 | 76.62% | 58 | 72.50% | 0.479 |
| | 비돌출형 | 14 | 19.44% | 49 | 28.49% | 18 | 23.38% | 22 | 27.50% | |
| 크기 | | 5.93±2.74 | | 4.04±1.83 | | 4.55±2.09 | | 5.06±2.41 | | <0.001 |
| 분화 | 양호 | 1 | 1.39% | 4 | 2.33% | 3 | 3.90% | 1 | 1.25% | <0.001 |
| | 중간 | 50 | 69.44% | 163 | 94.77% | 73 | 94.81% | 73 | 91.25% | |
| | 불량 | 21 | 29.17% | 5 | 2.91% | 1 | 1.30% | 6 | 7.50% | |
| 림프관 형성 | 없음 | 32 | 44.44% | 163 | 94.77% | 43 | 55.84% | 65 | 81.25% | <0.001 |
| | 국소적 | 27 | 37.50% | 7 | 4.07% | 28 | 36.36% | 14 | 17.50% | |
| | 광범위(>50%) | 13 | 18.06% | 2 | 1.16% | 6 | 7.79% | 1 | 1.25% | |
| Crohn 유사 림프관 반응 | 저밀도 그룹 | 26 | 36.11% | 127 | 73.84% | 59 | 76.62% | 60 | 75.00% | <0.001 |
| | 고밀도 그룹 | 46 | 63.89% | 45 | 26.16% | 18 | 23.38% | 20 | 25.00% | |
| 침습 패턴 | 확장 | 29 | 40.28% | 36 | 20.93% | 20 | 25.97% | 27 | 33.75% | 0.011 |
| | 침습적 | 43 | 59.72% | 136 | 79.07% | 57 | 74.03% | 53 | 66.25% | |
| 림프관 침윤 | 부재 | 65 | 90.28% | 139 | 80.81% | 71 | 92.21% | 59 | 73.75% | 0.004 |
| | 존재 | 7 | 9.72% | 33 | 19.19% | 6 | 7.79% | 21 | 26.25% | |
| 종양 발아 | 부재 | 61 | 84.72% | 120 | 69.77% | 61 | 79.22% | 44 | 55.00% | <0.001 |
| | 저 | 11 | 15.28% | 37 | 21.51% | 14 | 18.18% | 27 | 33.75% | |
| | 고 | 0 | 0.00% | 15 | 8.72% | 2 | 2.60% | 9 | 11.25% | |
| TSP | 저 | 69 | 95.83% | 138 | 80.23% | 56 | 72.73% | 50 | 62.50% | <0.001 |
| | 고 | 3 | 4.17% | 34 | 19.77% | 21 | 27.27% | 30 | 37.50% | |
| KM 등급 | 저 | 38 | 52.78% | 114 | 66.28% | 47 | 61.04% | 51 | 63.75% | 0.257 |
| | 고 | 34 | 47.22% | 58 | 33.72% | 30 | 38.96% | 29 | 36.25% | |
| GMS | GMS 0 | 34 | 47.22% | 58 | 33.72% | 30 | 38.96% | 29 | 36.25% | <0.001 |
| | GMS1 | 35 | 48.61% | 83 | 48.26% | 29 | 37.66% | 26 | 32.50% | |
| | GMS2 | 3 | 4.17% | 31 | 18.02% | 18 | 23.38% | 25 | 31.25% | |
| 면역점수 | 저 | 23 | 31.94% | 99 | 57.89% | 46 | 59.74% | 50 | 63.29% | 0.001 |
| | 중간 | 15 | 20.83% | 30 | 17.54% | 11 | 14.29% | 15 | 18.99% | |
| | 고 | 34 | 47.22% | 42 | 24.56% | 20 | 25.97% | 14 | 17.72% | |
| MSI 상태 | MSS/MSI-low | 17 | 23.61% | 172 | 100.00% | 74 | 96.10% | 80 | 100.00% | <0.001 |
| | MSI-High | 55 | 76.39% | 0 | 0.00% | 3 | 3.90% | 0 | 0.00% | |

TSP: 종양 기질 비율, KM 등급: Klintrap-Makinen 등급, GMS: Glasgow 미세환경점수, MSS: 초위성체 안정성, MSI: 초위성체 불안정성

TSP: 종양 기질 비율, KM 등급: Klintrap-Makinen 등급, GMS: Glasgow 미세환경점수, MSS: 초위성체 안정성, MSI: 초위성체 불안정성

[0209] 검증 코호트에서 IHC 분류자를 이용해 예측된 NMF-기반 그룹 역시 트레이닝 코호트와 유사한 임상병리적 특성을 보였다. 401개 케이스 중 21.7%는 그룹 1(87 케이스)로, 23.9%는 그룹 2(96 케이스)로, 14.2%는 그룹 3(57 케이스)으로, 40.1%는 그룹 4(161 케이스)로 각각 분류되었다.

[0210] 그룹 1은 점액성 암, 외성장성 생장 패턴과 같이 코호트 1과 유사한 혼합된 MSI 특성을 보였다. MSI가 높고, 오른쪽에 위치한, 크기가 큰 종양으로, 잘 분화되지 않은 세포의 비율이 높고 팽창하는 생장 패턴을 가졌다. 또한, 그룹 2에서는 예상대로 높은 KM 등급이 발견되었다. 그룹 3에서, 종양은 대부분 왼쪽에 위치하고 드물게 세포외 뭉신을 생성하였다. TSP와 GMS2가 두드러지게 높았다. 그룹 4에서 대부분의 종양은 좌측에 위치했으며, 크기는 더 작고, 세포외 뭉신 형성이 드물었으며, TSP는 상대적으로 높았다.

[0211] 조직학적 특성은 분석 결과 종래의 분자 아형과 유사한 패턴을 보였다(도 8). CMS1에서는, 복합 관상 패턴 뿐 아니라 견고한 외형이 특징적이었으며, CMS2에서는 유두 모양의 결합조직 형성이 관찰되었다. CMS 3에서는 유두 모양의 톱니 형태가 흔하게 발견되는 반면, CMS 4에서는 관상 형태 다음으로 결합조직 형태가 관찰되었다.

[0213] NMF-기반 그룹 및 IHC 분류자에 의한 CMS 간의 관련성

[0214] 코호트 1의 CMS 및 NMF-기반 그룹을 비교한 결과, 그룹 2의 66.6%는 CMS 1로 분류되었고, 그룹 4의 79.5%는 CMS 2로 분류되었으며, 그룹 1의 75%는 CMS 3로 분류되었고, 그룹 3은 100%가 CMS 4로 분류되었다. 코호트 2에서, CMS 및 MF-기반 그룹 간 유사한 상관관계가 있는지를 분석하였다(표 9). 코호트 2의 401 케이스 중 그룹 2는 CMS 1과(55.2%), 그룹 4는 CMS 2와(90.1%), 그룹 1은 CMS 3과(57.5%), 그리고 그룹 3은 CMS 4와(77.2%) 주로 연관되어 있는 것으로 나타났는데, 이는 코호트 1에서 관찰된 패턴과 유사한 것이었다.

표 9

[0215] NMF 기반 그룹과 CMS 아형 간의 관계

| | | 예측 NMF | | | | 계 |
|-----------|------|-----------|------------|-----------|-----------|-----|
| | | 그룹 2 | 그룹 4 | 그룹 1 | 그룹 3 | |
| 예측 CMS | CMS1 | 53(55.2%) | 0(0%) | 19(21.8%) | 0(0%) | 72 |
| | CMS2 | 12(12.5%) | 145(90.1%) | 11(12.6%) | 4(7.0%) | 172 |
| | CMS3 | 18(18.8%) | 0(0%) | 50(57.5%) | 9(15.8%) | 77 |
| | CMS4 | 13(13.5%) | 16(9.9%) | 7(8.0%) | 44(77.2%) | 80 |
| | 계 | 96 | 161 | 87 | 57 | 401 |

[0217] 환자 생존 분석

[0218] 코호트 1의 평균 추적관찰 기간은 105.85월(15-159 개월)이었다. 추적관찰 기간 동안, 13명의 환자(12.9%)는 재발하였고, 8명(7.9%)은 대장암-관련 원인으로 사망하였다. 코호트 2의 평균 추적관찰 기간은 83.58월(1-110개월)이었으며, 46명의 환자(11.5%)가 재발하고 76명 환자(19.0%)가 사망하였다.

[0219] FOLFOX 치료를 받은 3기 직장암 환자들로 구성된 코호트 1에서 분자 아형 및 IHC 분류자에 의해 예측된 아형에 따라 생존 분석을 비교하였다(도 9). 분자 아형 및 예측된 CMS 간의 예후를 비교한 결과, 유사한 경향성이 관찰되었다. 분자 아형 및 예측된 아형 모두에서, CMS1은 CMS 3 및 CMS 4와 비교하여 상대적으로 양호한 무병생존율을 보였다. 분자적 분석과 IHC 분류자에 의해 분류된 NMF-기반 그룹들 간의 비교 결과, 유사하게도 그룹 1의 전반적인 생존율이 좋지 못함을 알 수 있었다.

[0220] I-IV기 직장암 환자들로 구성된 검증 코호트의 생존율을 분석하였다. CMS4는 CMS1, CMS2 및 CMS3에 비해 불량한 DFS를 보였다(각각 $P=0.004$, $P=0.040$ 및 $P=0.001$, 도 10a). CMS4는 CMS1보다 OS가 불량한 경향을 보였다. NMF-기반 그룹 3은 그룹 1, 그룹 2 및 그룹 4에 비해 통계적으로 유의한 정도로 DFS가 불량하였다(각각 $P=0.001$, $P=0.001$ 및 $P=0.012$, 도 10b). 그룹 3은 그룹 2에 비해 통계적으로 유의한 정도로 DFS가 나빴

다($P=0.010$).

[0221] 이후 FOLFOX 보조제 화학치료를 받은 3기 직장암 환자의 생존률을 분석하였다. 전체 코호트에서 FOLFOX 보조제 화학치료를 받은 3기 직장암 환자는 165명에 존재하였다. Kaplan-Meier 곡선을 통해 CMS 3 및 CMS 4가 전체 코호트에서 불량한 예후를 보임을 알 수 있었다(도 11).

[0222] OS에 대한 일변량 Cox-비례위험 모델에서(표 10), 고령, 수술전 CEA 수준의 증가, 비-외장성 육안 형태, 큰 크기, 잦은 종양 발아, 높은 TSP, 낮은 면역점수, GMS2, CMS4, 및 NMF 그룹 3은 불량한 예후와 관련되어 있었다. 이들 인자 중 고령, 수술전 CEA 수준의 증가, 비-외장성 육안 형태 및 높은 TSP는 다변량 분석에서 독립적인 불량예후 인자였다. DFS에 대한 일변량 Cox-비례위험 모델에서(표 11), MSS, 수술 전 CEA 수준의 증가, 비-외장성 육안 형태, 큰 크기, 잦은 종양 발아, 높은 TSP, 낮은 면역점수, 낮은 KM 등급, GMS2, 높은 기수(stage), CMS4, 및 NMF 그룹 3은 불량한 예후와 관련이 있었다. 이들 인자 중, CMS4, MSS, 수술 전 CEA 수준의 증가, 큰 크기와 낮은 면역점수 및 높은 기수(stage)는 다변량 분석에서 각각 독립적인 불량 예후 인자였다.

표 10

[0223] OS에 대한 일변량 및 다변량 Cox-비례위험 모델

| 변수 | 카테고리 | 일변량 | | | 다변량 | | |
|-----------|---------------------------------|-------|--------------|--------|-------|-------------|-------|
| | | HR | 95% CI | p-값 | HR | 95% CI | p-값 |
| 나이 (년) | 연속 변수 | 1.041 | 1.020-1.062 | <0.001 | 1.036 | 1.015-1.057 | 0.001 |
| MSI | MSI-H vs. MSS(Ref.) | 0.772 | 0.353-1.690 | 0.517 | | | |
| PreOP CEA | >5ng/ml vs. ≤5ng/ml(Ref.) | 2.112 | 1.364-3.268 | <0.001 | 1.777 | 1.130-2.794 | 0.013 |
| 육안 형태 | 비-외장성 vs. 돌출형(Ref.) | 2.238 | 1.445-3.467 | <0.001 | 2.152 | 1.366-3.389 | 0.001 |
| 크기 | 연속 변수 | 1.111 | 1.011-1.220 | 0.029 | | | |
| 종양 발아 | intermediate vs. low(Ref.) | 0.567 | 0.332-0.966 | 0.037 | | | |
| | high vs. low(Ref.) | 0.546 | 0.268-1.115 | 0.097 | | | |
| TSP | high vs. low(Ref.) | 2.914 | 1.825-4.653 | <0.001 | 2.019 | 1.231-3.311 | 0.005 |
| 면역 점수 | intermediate vs. high(Ref.) | 1.333 | 0.618-2.0877 | 0.463 | | | |
| | low vs. high(Ref.) | 2.228 | 1.220-4.070 | 0.009 | | | |
| KM 등급 | high(G3-4) vs. low (G1-2)(Ref.) | 0.504 | 0.316-0.805 | 0.004 | | | |
| GMS | GMS1 vs. GMS0(Ref.) | 1.212 | 0.702-2.093 | 0.491 | | | |
| | GMS2 vs. GMS0(Ref.) | 2.637 | 1.527-4.554 | 0.001 | | | |
| | 3/4기 vs 1/2기(Ref.) | 1.371 | 0.890-2.113 | 0.152 | | | |
| CMS | CMS1 vs. CMS4(Ref.) | 0.508 | 0.771-4.064 | 0.603 | | | |
| | CMS2 vs. CMS4(Ref.) | 0.563 | 0.265-2.629 | 0.032 | | | |
| | CMS3 vs. CMS4(Ref.) | 0.708 | 1.101-6.504 | 0.285 | | | |
| NMF | NMF1 vs. NMF4(Ref.) | 1.217 | 0.702-2.108 | 0.484 | | | |
| | NMF2 vs. NMF4(Ref.) | 0.667 | 0.346-1.285 | 0.226 | | | |
| | NMF3 vs. NMF4(Ref.) | 1.922 | 1.076-3.432 | 0.027 | | | |

표 11

[0225] DFS에 대한 일변량 및 다변량 Cox-비례위험 모델

| 변수 | 카테고리 | 일변량 | | | 다변량 | | |
|-----------|---------------------------|-------|-------------|--------|-------|-------------|-------|
| | | HR | 95% CI | p-값 | HR | 95% CI | p-값 |
| 나이 (년) | 연속 변수 | 1.011 | 0.988-1.034 | 0.35 | | | |
| MSI | MSI-H vs. MSS(Ref.) | 0.109 | 0.015-0.787 | 0.028 | 0.096 | 0.011-0.880 | 0.038 |
| PreOP CEA | >5ng/ml vs. ≤5ng/ml(Ref.) | 3.015 | 1.806-5.035 | <0.001 | 2.01 | 1.179-3.428 | 0.010 |

| | | | | | | | |
|-------|---------------------------------|-------|--------------|--------|-------|-------------|-------|
| 육안 형태 | 비-외장성 vs. 돌출형(Ref.) | 2.741 | 1.642-4.576 | <0.001 | | | |
| 크기 | 연속 변수 | 1.113 | 1.002-1.235 | 0.045 | 1.185 | 1.044-1.346 | 0.009 |
| 종양 발아 | intermediate vs. low(Ref.) | 0.926 | 0.493-1.740 | 0.811 | | | |
| | high vs. low(Ref.) | 2.397 | 1.275-4.506 | 0.007 | | | |
| TSP | high vs. low(Ref.) | 3.239 | 1.944-5.399 | <0.001 | | | |
| 면역점수 | intermediate vs. high(Ref.) | 1.613 | 0.492-5.284 | 0.43 | | | |
| | low vs. high(Ref.) | 4.986 | 1.983-12.539 | 0.001 | 3.151 | 1.241-8.000 | 0.016 |
| KM 등급 | high(G3-4) vs. low (G1-2)(Ref.) | 0.483 | 0.275-0.849 | 0.011 | | | |
| GMS | GMS1 vs. GMS0(Ref.) | 1.516 | 0.746-3.082 | 0.25 | | | |
| | GMS2 vs. GMS0(Ref.) | 4.149 | 2.093-8.226 | <0.001 | | | |
| 단계 | 3/4기 vs 1/2기(Ref.) | 2.891 | 1.628-5.132 | <0.001 | 2.401 | 1.275-4.523 | 0.007 |
| CMS | CMS1 vs. CMS4(Ref.) | 0.242 | 0.092-0.640 | 0.004 | 0.937 | 0.311-2.827 | 0.908 |
| | CMS2 vs. CMS4(Ref.) | 0.480 | 0.272-0.847 | 0.011 | 0.517 | 0.277-0.964 | 0.038 |
| | CMS3 vs. CMS4(Ref.) | 0.265 | 0.107-0.653 | 0.004 | 0.314 | 0.122-0.804 | 0.016 |
| NMF | NMF1 vs. NMF4(Ref.) | 0.866 | 0.428-1.752 | 0.688 | | | |
| | NMF2 vs. NMF4(Ref.) | 0.433 | 0.178-1.053 | 0.065 | | | |
| | NMF3 vs. NMF4(Ref.) | 2.389 | 1.281-4.453 | 0.006 | | | |

[0227] 종양 중심과 침습 가장자리에서는 아형이 상이하게 예측된다.

[0228] 종양 중심, 종양 침습 가장자리 및 전이 림프절에서 IHC 결과를 해석하였다. 침습 가장자리의 종양 세포는 정상세포와 암 세포 간 경계로부터 2mm 이내에 존재하는 종양 군집을 간주하였다. 몇몇 케이스는 종양 중심과 종양 침습 가장자리에서 상이한 IHC 발현 패턴을 보였다. 종양 중심과 침습 가장자리 모두의 CMS를 평가하는 데 이용할 수 있는 케이스는 408개가 있었는데, 종양 중심에서 CMS1로 분류된 86개의 케이스 중 하나가 CMS2로, 두 개 케이스가 CMS3로, 그리고 5개 케이스가 침습 가장자리에서는 CMS4로 분류가 바뀌었다. 종양 중심에서는 184개 케이스가 CMS2로 관찰되었으며 이중 4개 케이스가 침습 가장자리에서 CMS4로 변경되었다. 종양 중심에서 CMS3로 분류된 케이스는 66개였으며, 침습 가장자리에서는 이중 12개가 CMS2로, 1개가 CMS1으로 바뀌었다. 72개의 CMS4 케이스 중에서는 6개가 CMS2로 바뀌었다(도 12).

[0229] 56개 케이스를 통해 종양 중심, 침습 가장자리 및 전이 병소에서의 아형 분류를 평가하였다. CMS1가 11개 케이스, CMS2가 31개 케이스, CMS3가 7개 케이스였으며, 10 케이스는 종양 중심, 침습 가장자리 및 전이 병소의 아형이 상이했다(표 12). 1-43 케이스의 면역 프로파일은 도 13에 나타내었다. 케이스는 종양 중심에서 CMS 1로 예측되었으나, 침습 가장자리 및 전이 병소에서는 CMS 2로 바뀌었다. SPINK4, MUC2 및 MUC5AC는 종양 중심에서 고발현되었으나, 침습 가장자리 및 전이 병소에서는 이들의 발현이 검출되지 않았다. CDX2의 낮은 발현은 CMS1 및 CMS4의 특징이다. 종양 중심에서 낮게 나타나는 CDX2의 발현량은 침습 가장자리 및 전이 병소에서 높아졌다.

표 12

[0230] 종양 중심, 침습 가장자리 및 전이 병소에서 상이하게 예측되는 CMS

| Case no. | 위치 | 예측된 CMS | SPINK4 | MUC2 | MUC5AC | TFF3 | KIT | Snail | ZMYND8 | CDX2 | MSI 상태 |
|----------|---------|---------|--------|------|--------|------|-----|-------|--------|------|--------|
| 1-57 | 종양 중심 | CMS1 | | 0 | | | | 0 | | 0 | 0 |
| | 침습 가장자리 | CMS1 | | 0 | | | | 0 | | 0 | 0 |
| | 전이 병소 | CMS3 | | | | 0 | | | | 0 | 0 |
| 1-71 | 종양 중심 | CMS1 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| | 침습 가장자리 | CMS1 | 0 | | 0 | 0 | 0 | 0 | | | |
| | 전이 병소 | CMS2 | | | | | | 0 | | | |
| 1-43 | 종양 중심 | CMS1 | 0 | 0 | 0 | 0 | | | | | |

| | | | | | | | | | | | |
|---------------------------------|----------|------|---|---|---|---|--|---|---|---|---|
| | 침습 가장 자리 | CMS2 | | | | 0 | | 0 | | 0 | |
| | 전이 병소 | CMS2 | | | | 0 | | | | 0 | |
| 1-47 | 종양 중심 | CMS1 | 0 | 0 | | 0 | | 0 | | | 0 |
| | 침습 가장 자리 | CMS3 | 0 | 0 | | 0 | | 0 | | 0 | 0 |
| | 전이 병소 | CMS1 | | 0 | | 0 | | | | 0 | 0 |
| 1-83 | 종양 중심 | CMS1 | | | 0 | | | 0 | | | |
| | 침습 가장 자리 | CMS4 | 0 | | 0 | 0 | | 0 | | | |
| | 전이 병소 | CMS1 | 0 | | 0 | | | | | | |
| 1-41 | 종양 중심 | CMS2 | | | | 0 | | | 0 | 0 | |
| | 침습 가장 자리 | CMS2 | | | | 0 | | | 0 | 0 | |
| | 전이 병소 | CMS4 | | | 0 | 0 | | 0 | 0 | 0 | |
| 1-32 | 종양 중심 | CMS3 | 0 | 0 | | 0 | | | | 0 | |
| | 침습 가장 자리 | CMS2 | | | | 0 | | | | 0 | |
| | 전이 병소 | CMS2 | | | | 0 | | | | 0 | |
| 1-94 | 종양 중심 | CMS3 | 0 | 0 | | 0 | | | 0 | | |
| | 침습 가장 자리 | CMS2 | 0 | | | 0 | | | 0 | 0 | |
| | 전이 병소 | CMS2 | 0 | | | 0 | | | 0 | 0 | |
| 1-84 | 종양 중심 | CMS3 | 0 | 0 | | 0 | | | 0 | | |
| | 침습 가장 자리 | CMS2 | 0 | | | 0 | | | 0 | 0 | |
| | 전이 병소 | CMS3 | 0 | 0 | | 0 | | | | 0 | |
| 1-51 | 종양 중심 | CMS3 | | 0 | | | | | | 0 | |
| | 침습 가장 자리 | CMS3 | | 0 | 0 | 0 | | | | 0 | |
| | 전이 병소 | CMS1 | | 0 | 0 | | | | | 0 | |
| 빈칸: 저발현, 채워진 칸: 고발현 또는 MSI-High | | | | | | | | | | | |

- [0232] 이상으로 본 발명의 특정한 부분을 상세히 기술하였는 바, 당업계의 통상의 지식을 가진 자에게 있어서 이러한 구체적인 기술은 단지 바람직한 구현예일 뿐이며, 이에 본 발명의 범위가 제한되는 것이 아닌 점은 명백하다. 따라서, 본 발명의 실질적인 범위는 첨부된 청구항과 그의 등가물에 의하여 정의된다고 할 것이다.
- [0234] **참고문헌**
- [0235] 1. Siegel RL, Miller KD, Jemal A. 2018. CA Cancer J Clin 2018;68:7-30.
- [0236] 2. Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, et al. 2016. CA Cancer J Clin 2016;66:271-89.
- [0237] 3. Andre T, Boni C, Mounedji-Boudiaf L, Navarro M, Tabernero J, Hickish T, et al. N Engl J Med 2004;350:2343-51.
- [0238] 4. Benson AB, 3rd. Semin Oncol 2005;32:S74-7.
- [0239] 5. Stefansson M, Nygren P. Acta Oncol 2016;55:1227-35.
- [0240] 6. Des Guetz G, Schischmanoff O, Nicolas P, Perret GY, Morere JF, Uzzan B. A systematic review with meta-analysis. Eur J Cancer 2009;45:1890-6.
- [0241] 7. Sinicrope FA, Sargent DJ. Curr Opin Oncol 2009;21:369-73.
- [0242] 8. De Roock W, Claes B, Bernasconi D, De Schutter J, Biesmans B, Fountzilas G, et al. Lancet Oncol

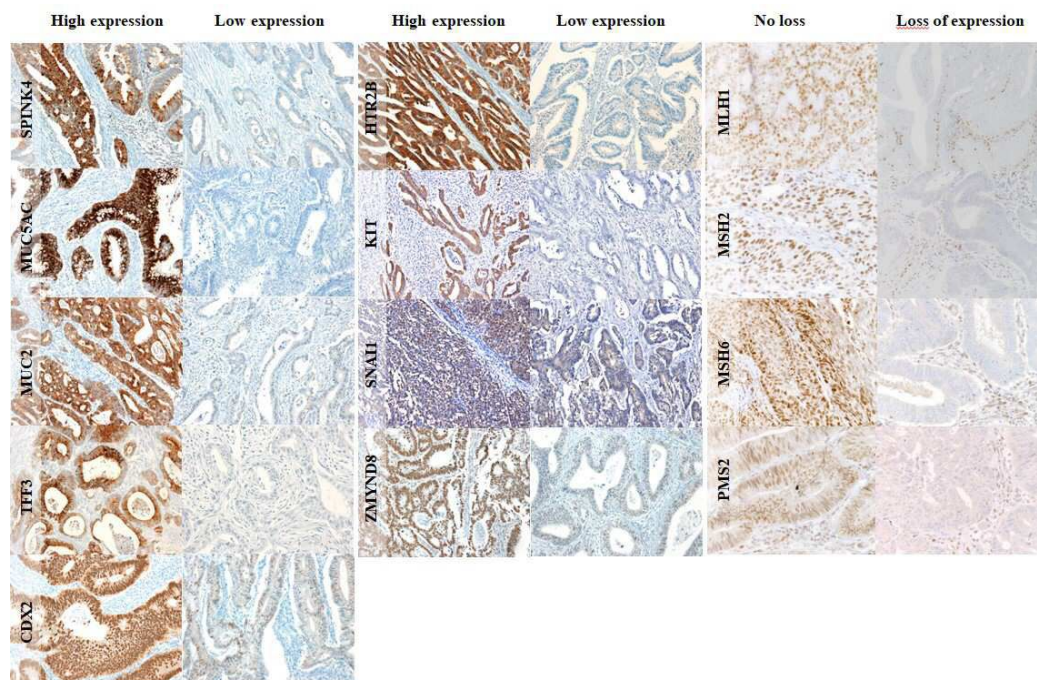
2010;11:753-62.

- [0243] 9. Lievre A, Bachet JB, Le Corre D, Boige V, Landi B, Emile JF, et al. *Cancer Res* 2006;66:3992-5.
- [0244] 10. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012;487:330-7.
- [0245] 11. Roepman P, Schlicker A, Tabernero J, Majewski I, Tian S, Moreno V, et al. *Int J Cancer* 2014;134:552-62.
- [0246] 12. Budinska E, Popovici V, Tejpar S, D'Ario G, Lapique N, Sikora KO, et al. *J Pathol* 2013;231:63-76.
- [0247] 13. De Sousa EMF, Wang X, Jansen M, Fessler E, Trinh A, de Rooij LP, et al. *Nat Med* 2013;19:614-8.
- [0248] 14. Marisa L, de Reynies A, Duval A, Selves J, Gaub MP, Vescovo L, et al. *PLoS Med* 2013;10:e1001453.
- [0249] 15. Schlicker A, Beran G, Chresta CM, McWalter G, Pritchard A, Weston S, et al. *BMC Med Genomics* 2012;5:66.
- [0250] 16. Perez-Villamil B, Romera-Lopez A, Hernandez-Prieto S, Lopez-Campos G, Calles A, Lopez-Asenjo JA, et al. *BMC Cancer* 2012;12:260.
- [0251] 17. Sadanandam A, Lyssiotis CA, Homicsko K, Collisson EA, Gibb WJ, Wullschlegel S, et al. *Nat Med* 2013;19:619-25.
- [0252] 18. Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Soneson C, et al. *Nat Med* 2015;21:1350-6.
- [0253] 19. Song N, Pogue-Geile KL, Gavin PG, Yothers G, Kim SR, Johnson NL, et al. *JAMA Oncol* 2016;2:1162-9.
- [0254] 20. Bolstad BM, Irizarry RA, Astrand M, Speed TP. *Bioinformatics* 2003;19:185-93.
- [0255] 21. Lee HJ, Suk JE, Patrick C, Bae EJ, Cho JH, Rho S, et al. *J Biol Chem* 2010;285:9262-72.
- [0256] 22. Hwang D, Rust AG, Ramsey S, Smith JJ, Leslie DM, Weston AD, et al. *Proc Natl Acad Sci U S A* 2005;102:17296-301.
- [0257] 23. Tibshirani R, Hastie T, Narasimhan B, Chu G. *Proc Natl Acad Sci U S A* 2002;99:6567-72.
- [0258] 24. Huang da W, Sherman BT, Lempicki RA. *Nat Protoc* 2009;4:44-57.
- [0259] 25. Brunet JP, Tamayo P, Golub TR, Mesirov JP. *Proc Natl Acad Sci U S A* 2004;101:4164-9.
- [0260] 26. Gao Y, Church G. *Bioinformatics* 2005;21:3970-5.
- [0261] 27. Kim H, Park H. *Bioinformatics* 2007;23:1495-502.
- [0262] 28. Gaujoux R, Seoighe C. *BMC Bioinformatics* 2010;11:367.
- [0263] 29. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.
- [0264] 30. Chen D, Qi W, Zhang P, Zhang Y, Liu Y, Guan H, et al. *Pathol Res Pract* 2018;214:303-7.
- [0265] 31. Isella C, Brundu F, Bellomo SE, Galimi F, Zanella E, Porporato R, et al. *Nat Commun* 2017;8:15107.
- [0266] 32. Trinh A, Trumpi K, De Sousa EMF, Wang X, de Jong JH, Fessler E, et al. *Clin Cancer Res* 2017;23:387-98.
- [0267] 33. Pilati C, Taieb J, Balogoun R, Marisa L, de Reynies A, Laurent-Puig P. *Ann Oncol* 2017;28:1032-5.
- [0268] 34. Wang WH, Xie TY, Xie GL, Ren ZL, Li JM. *Genes (Basel)* 2018;9.
- [0269] 35. Ubink I, Elias SG, Moelans CB, Lacle MM, van Grevenstein WMU, van Diest PJ, et al. *J Natl Cancer Inst* 2017;109.
- [0270] 36. Taube JM, Galon J, Sholl LM, Rodig SJ, Cottrell TR, Giraldo NA, et al. *Mod Pathol* 2018;31:214-34.

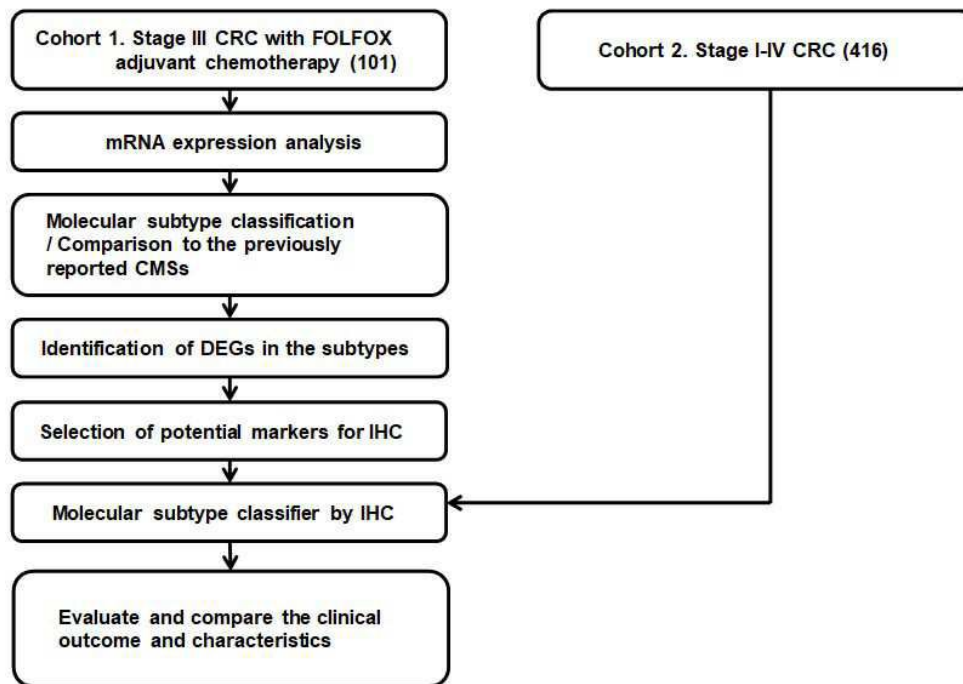
- [0271] 37. Kwak Y, Koh J, Kim DW, Kang SB, Kim WH, Lee HS. Oncotarget 2016;7:81778-90.
- [0272] 38. Huijbers A, Tollenaar RA, v Pelt GW, Zeestraten EC, Dutton S, McConkey CC, et al. Ann Oncol 2013;24:179-85.
- [0273] 39. Roxburgh CS, Salmond JM, Horgan PG, Oien KA, McMillan DC. Ann Surg 2009;249:788-93.
- [0274] 40. Park JH, McMillan DC, Powell AG, Richards CH, Horgan PG, Edwards J, et al. Clin Cancer Res 2015;21:882-8.

도면

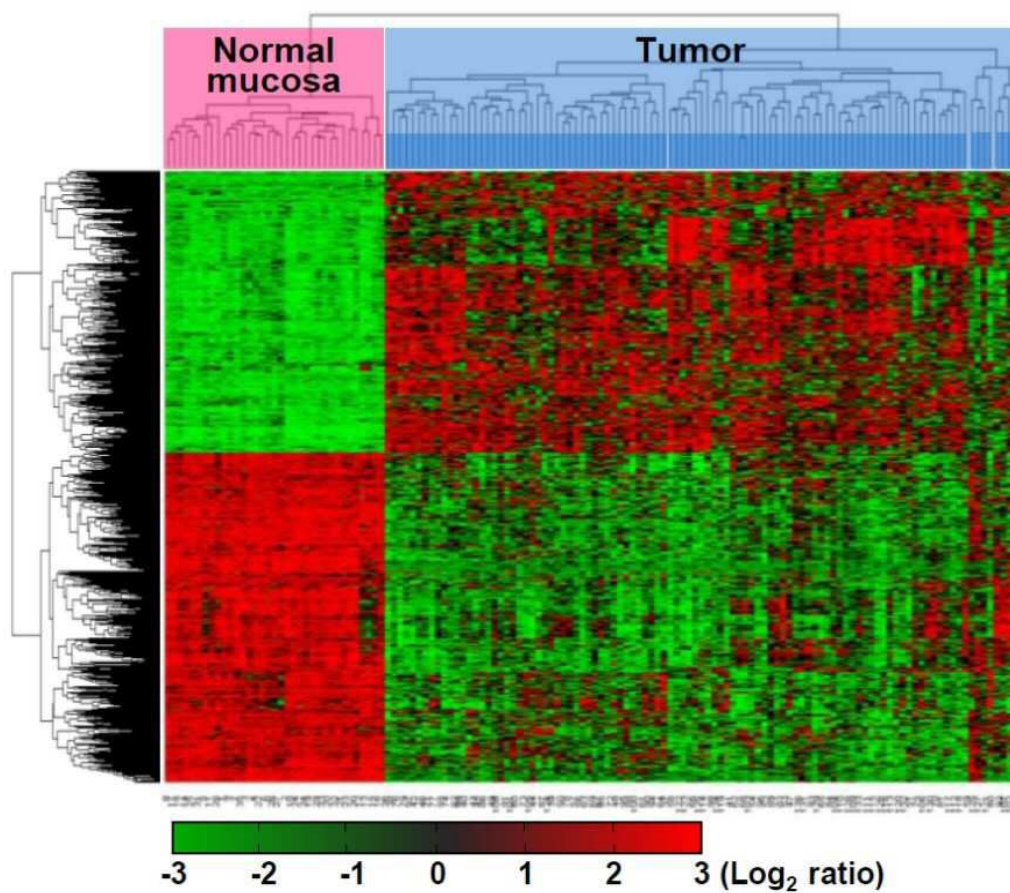
도면1



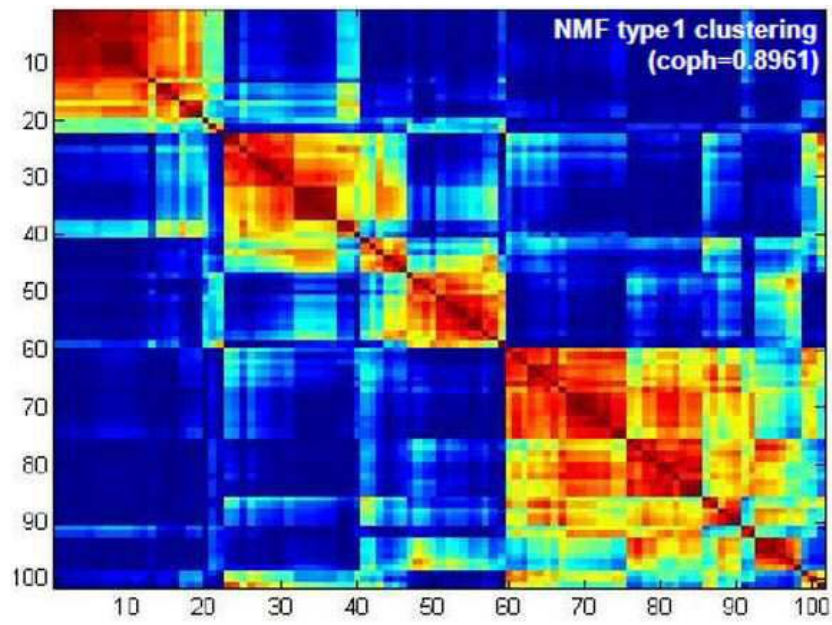
도면2



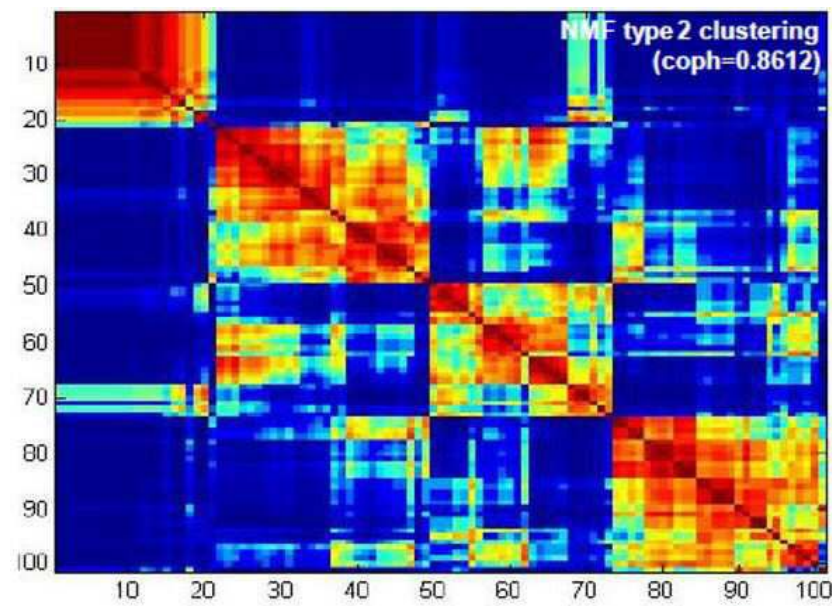
도면3



도면4a



도면4b



도면4c

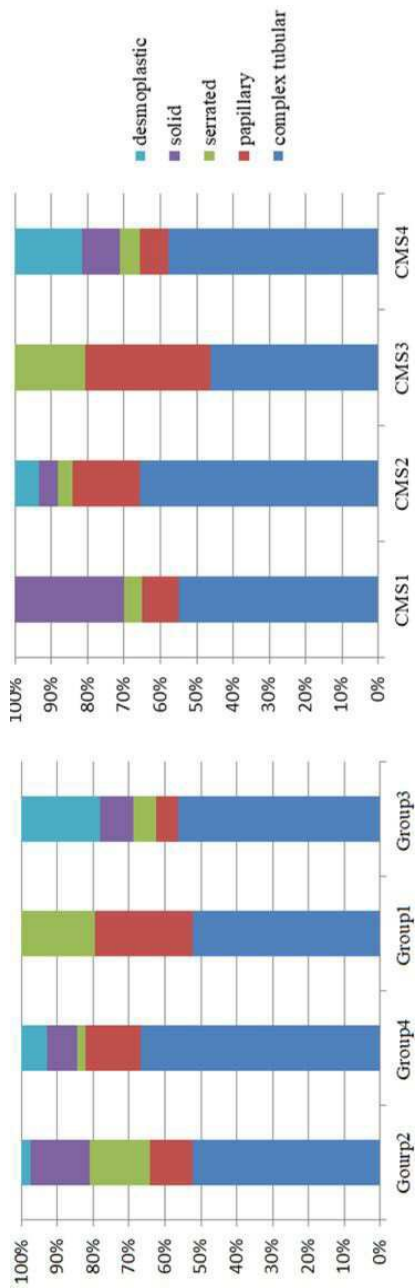
| Tumor/ Normal | | | | | Specific tumor group/ other groups | | | | |
|---------------|----------|----------|----------|----------|------------------------------------|-------------------|-------------------|-------------------|-------------------|
| SYMBOL | Group1/N | Group2/N | Group3/N | Group4/N | SYMBOL | Group1/ others | Group2/ others | Group3/ others | Group4/ others |
| ACSL6 | | | | | ACSL6 | | | | |
| CA1 | | | | | CA1 | | | | |
| CA4 | | | | | CA4 | | | | |
| CEL | | | | | CEL | | | | |
| CFTR | | | | | CFTR | | | | |
| CLCA4 | | | | | CLCA4 | | | | |
| COL10A1 | | | | | COL10A1 | | | | |
| CXCL13 | | | | | CXCL13 | | | | |
| CXCL9 | | | | | CXCL9 | | | | |
| CYP1B1 | | | | | CYP1B1 | | | | |
| EREG | | | | | EREG | | | | |
| FLNA | | | | | FLNA | | | | |
| IDO1 | | | | | IDO1 | | | | |
| KRT23 | | | | | KRT23 | | | | |
| LY6G6D | | | | | LY6G6D | | | | |
| MGP | | | | | MGP | | | | |
| MGP | | | | | MGP | | | | |
| MS4A12 | | | | | MS4A12 | | | | |
| MSRB3 | | | | | MSRB3 | | | | |
| MSRB3 | | | | | MSRB3 | | | | |
| MUC2 | | | | | MUC2 | | | | |
| PCSK1 | | | | | PCSK1 | | | | |
| RARRES3 | | | | | RARRES3 | | | | |
| REG4 | | | | | REG4 | | | | |
| SFRP2 | | | | | SFRP2 | | | | |
| SPINK4 | | | | | SPINK4 | | | | |
| TCN1 | | | | | TCN1 | | | | |
| TFF3 | | | | | TFF3 | | | | |
| ZEB1 | | | | | ZEB1 | | | | |

도면4d

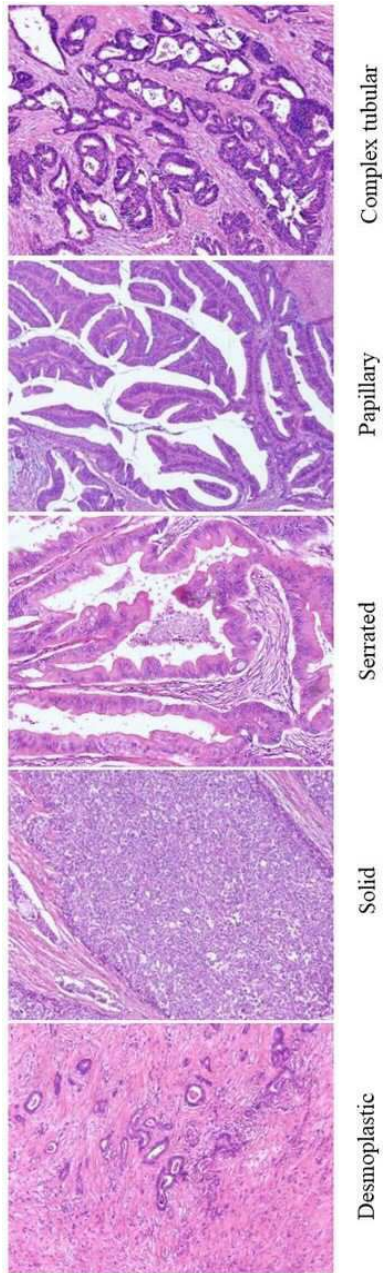
| Tumor/ Normal | | | | | Specific tumor group/ other groups | | | | |
|---------------|----------|----------|----------|----------|------------------------------------|-------------------|-------------------|-------------------|-------------------|
| SYMBOL | Group1/N | Group2/N | Group3/N | Group4/N | SYMBOL | Group1/ others | Group2/ others | Group3/ others | Group4/ others |
| ACSL6 | | | | | ACSL6 | | | | |
| CA1 | | | | | CA1 | | | | |
| CA4 | | | | | CA4 | | | | |
| CEL | | | | | CEL | | | | |
| CFTR | | | | | CFTR | | | | |
| CLCA4 | | | | | CLCA4 | | | | |
| COL10A1 | | | | | COL10A1 | | | | |
| CXCL13 | | | | | CXCL13 | | | | |
| CXCL9 | | | | | CXCL9 | | | | |
| CYP1B1 | | | | | CYP1B1 | | | | |
| EREG | | | | | EREG | | | | |
| FLNA | | | | | FLNA | | | | |
| IDO1 | | | | | IDO1 | | | | |
| KRT23 | | | | | KRT23 | | | | |
| LY6G6D | | | | | LY6G6D | | | | |
| MGP | | | | | MGP | | | | |
| MGP | | | | | MGP | | | | |
| MS4A12 | | | | | MS4A12 | | | | |
| MSRB3 | | | | | MSRB3 | | | | |
| MSRB3 | | | | | MSRB3 | | | | |
| MUC2 | | | | | MUC2 | | | | |
| PCSK1 | | | | | PCSK1 | | | | |
| RARRES3 | | | | | RARRES3 | | | | |
| REG4 | | | | | REG4 | | | | |
| SFRP2 | | | | | SFRP2 | | | | |
| SPINK4 | | | | | SPINK4 | | | | |
| TCN1 | | | | | TCN1 | | | | |
| TFF3 | | | | | TFF3 | | | | |
| ZEB1 | | | | | ZEB1 | | | | |

■ Up(Fold change>1.5)
 ■ No change
 ■ Down(Fold change<-1.5)
 P<0.01

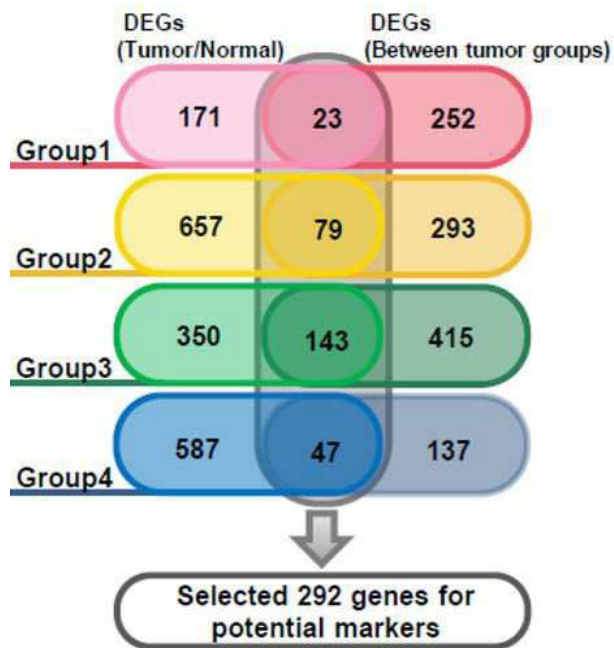
도면5a



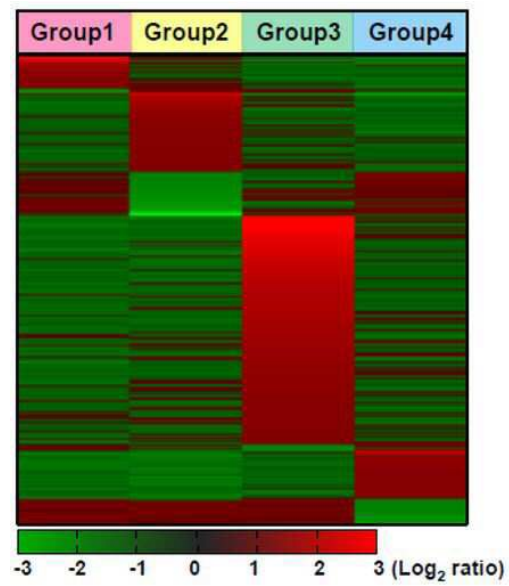
도면5b



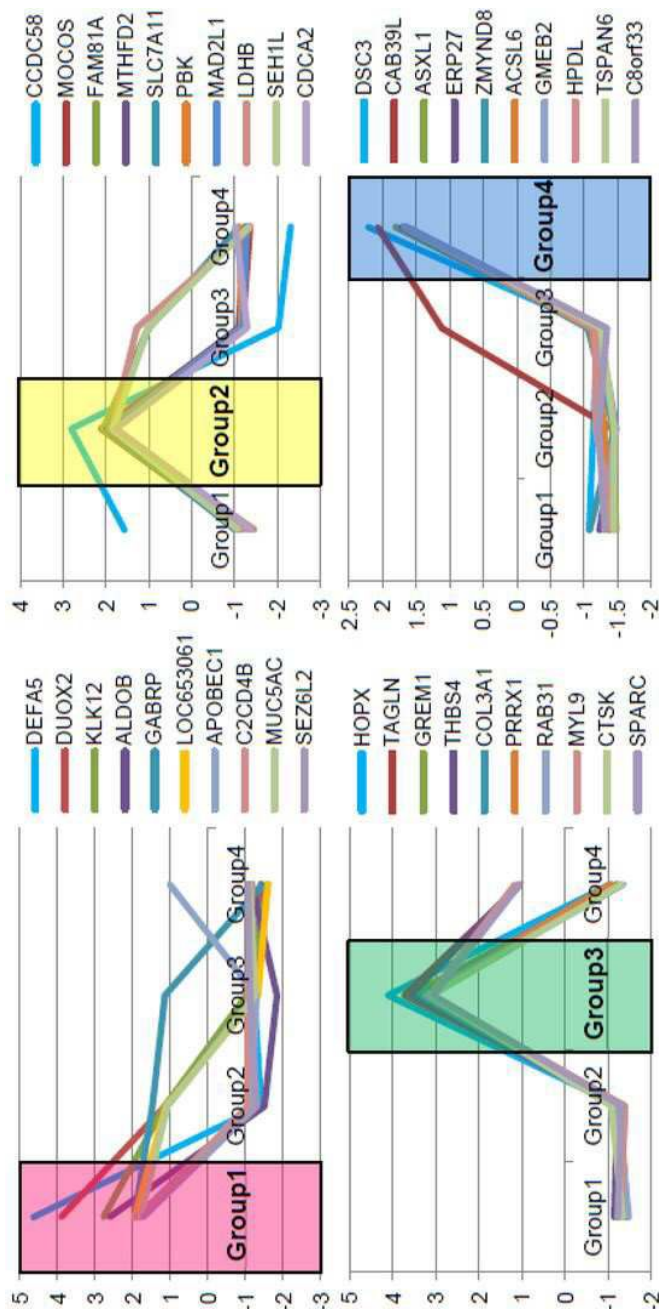
도면6a



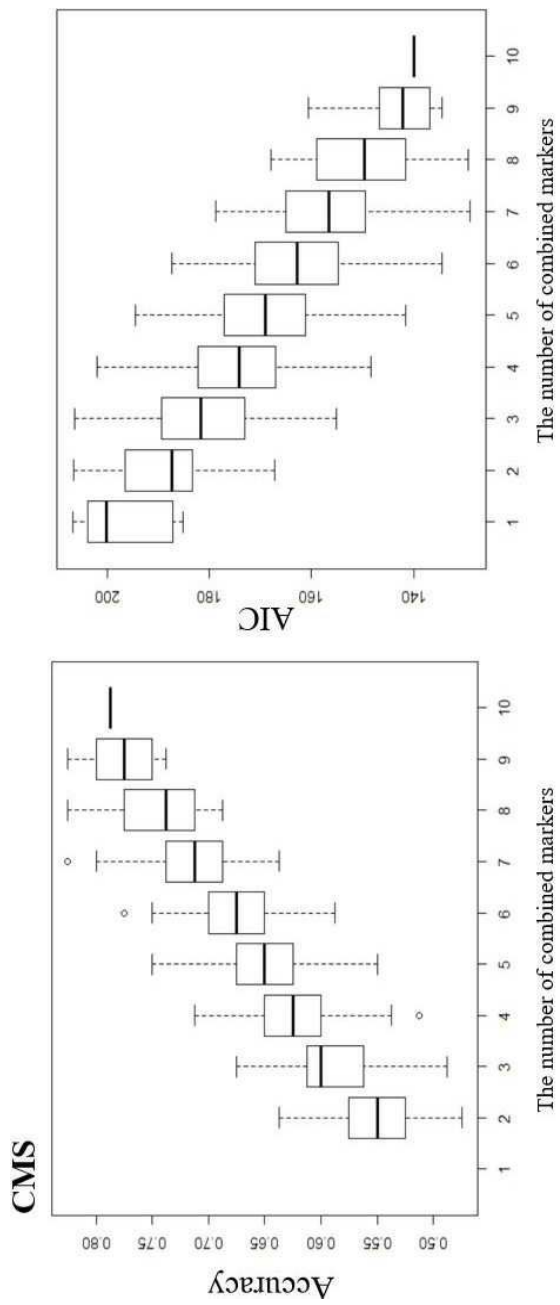
도면6b



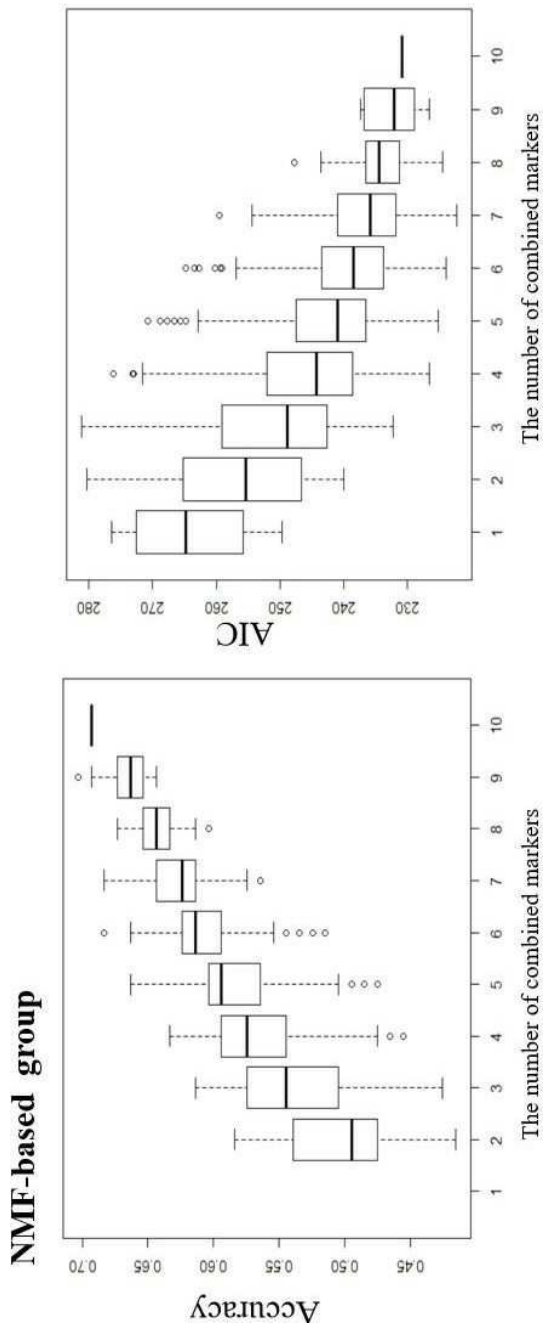
도면6c



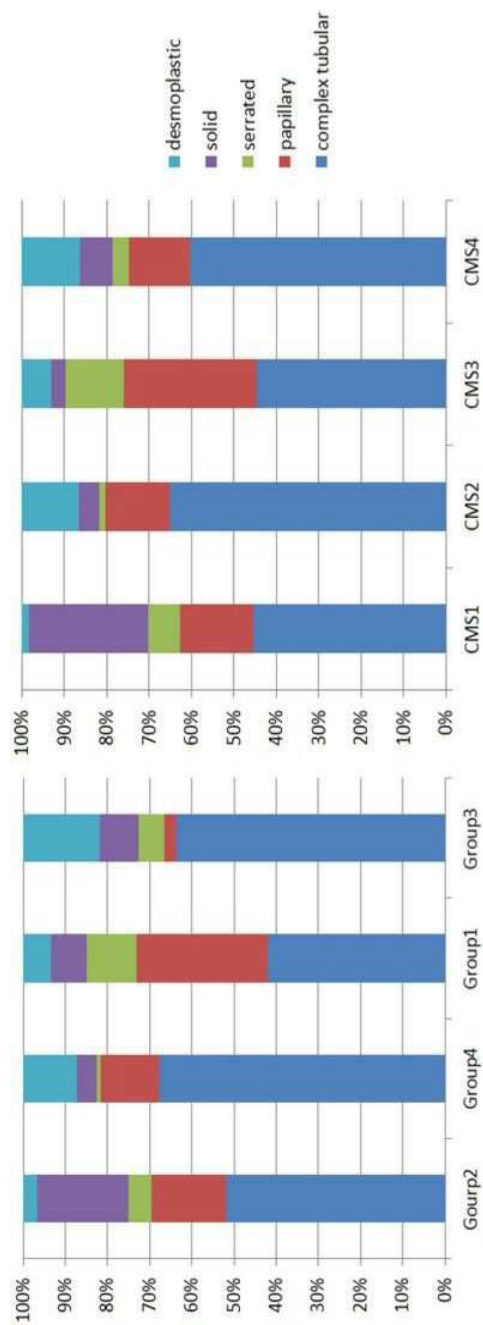
도면7a



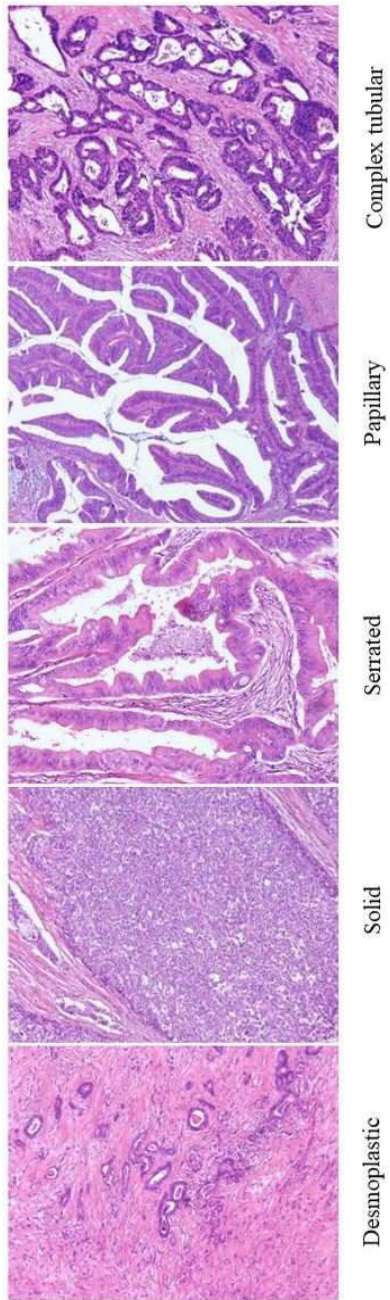
도면7b



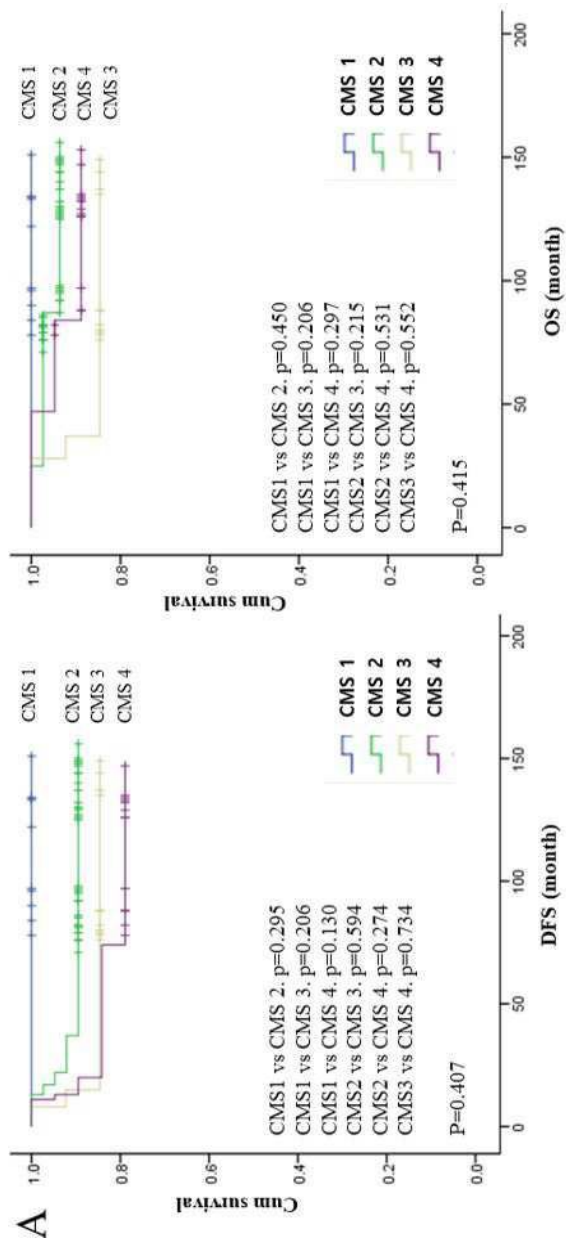
도면8a



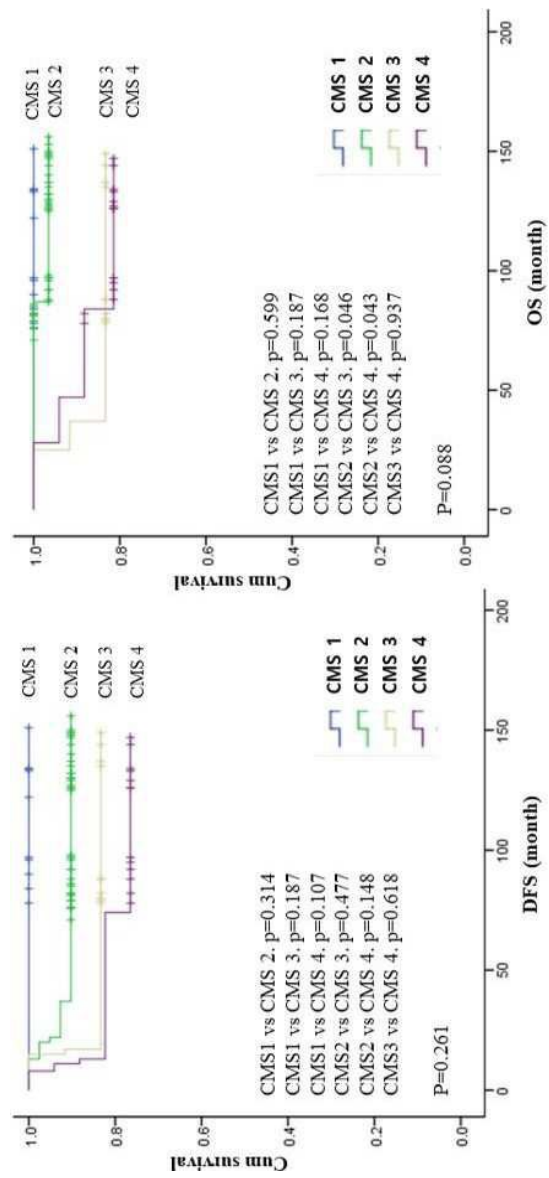
도면8b



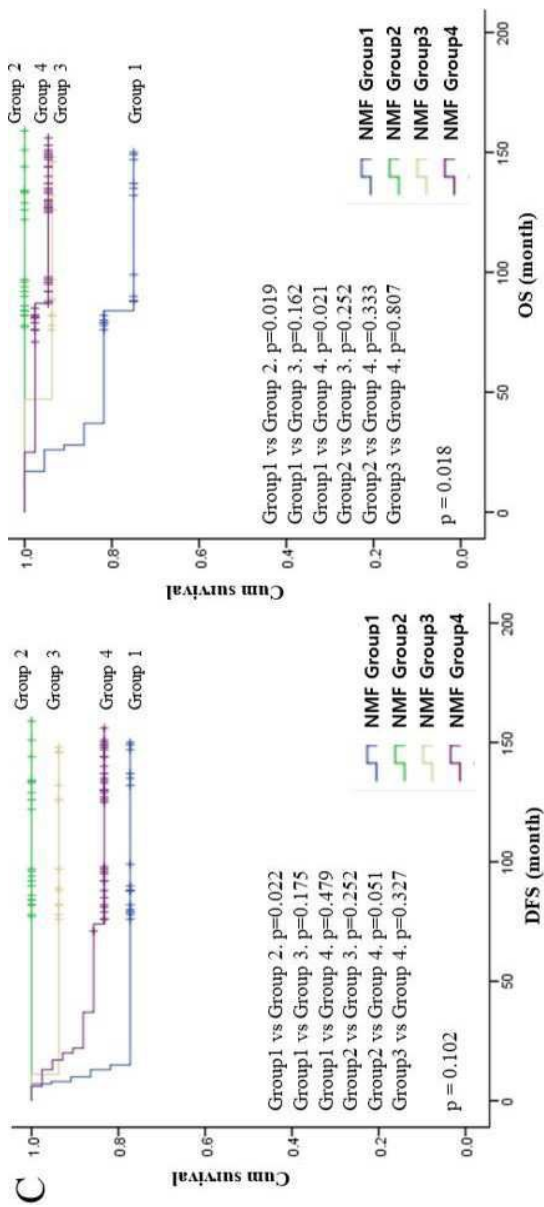
도면9a



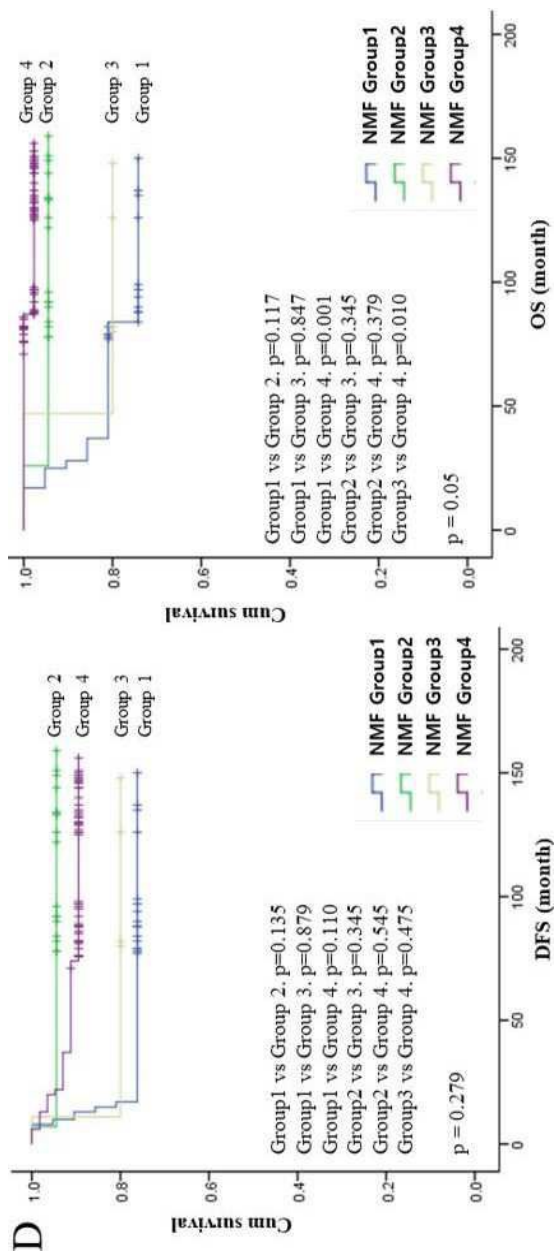
도면9b



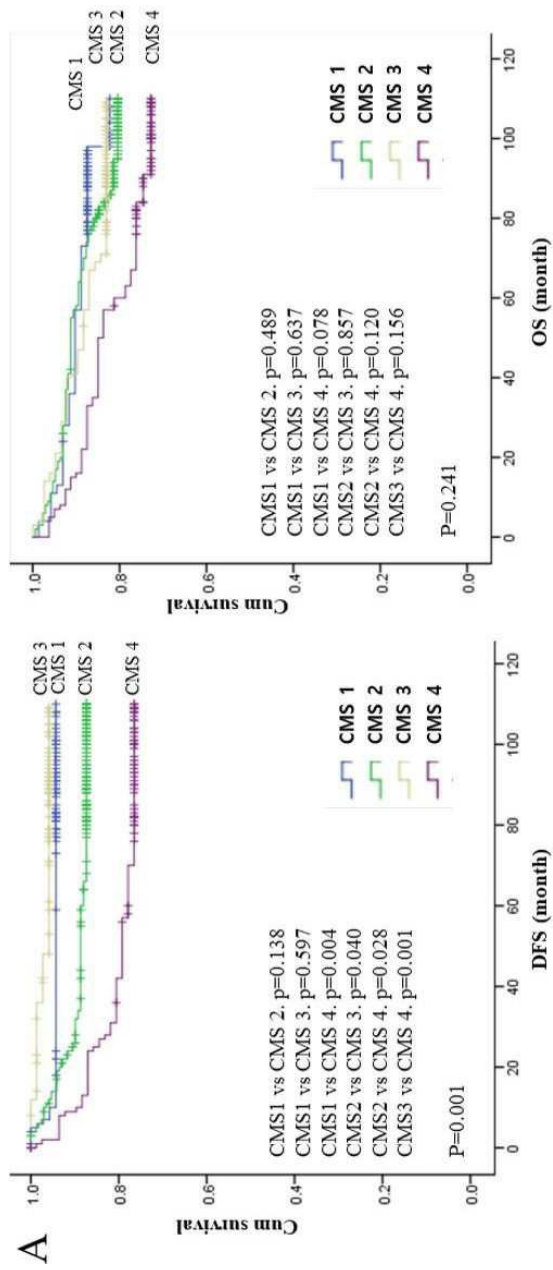
도면9c



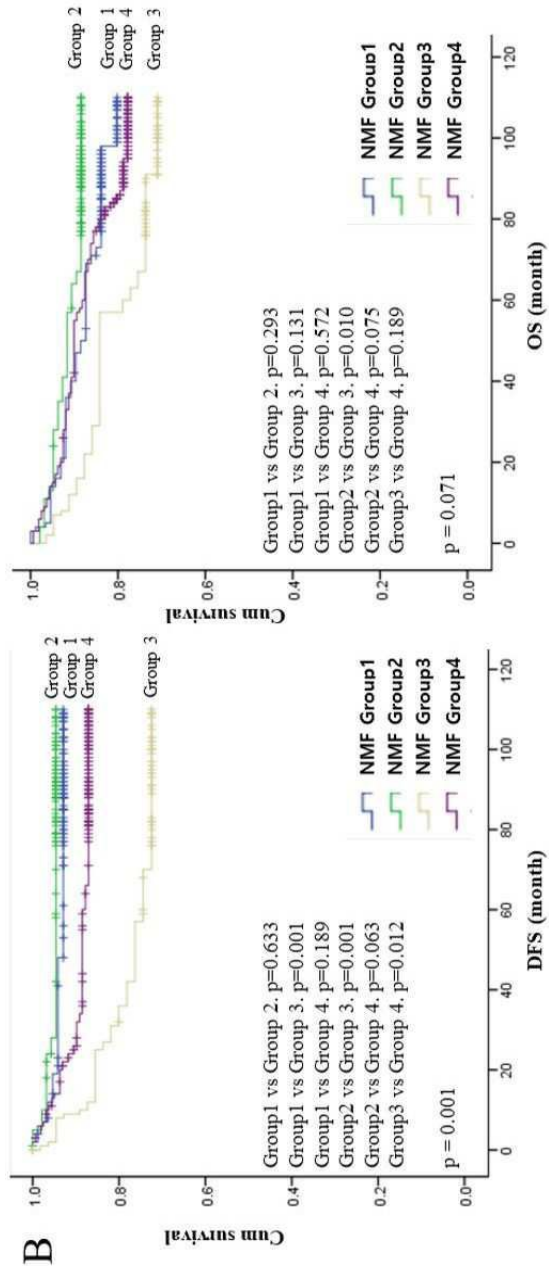
도면9d



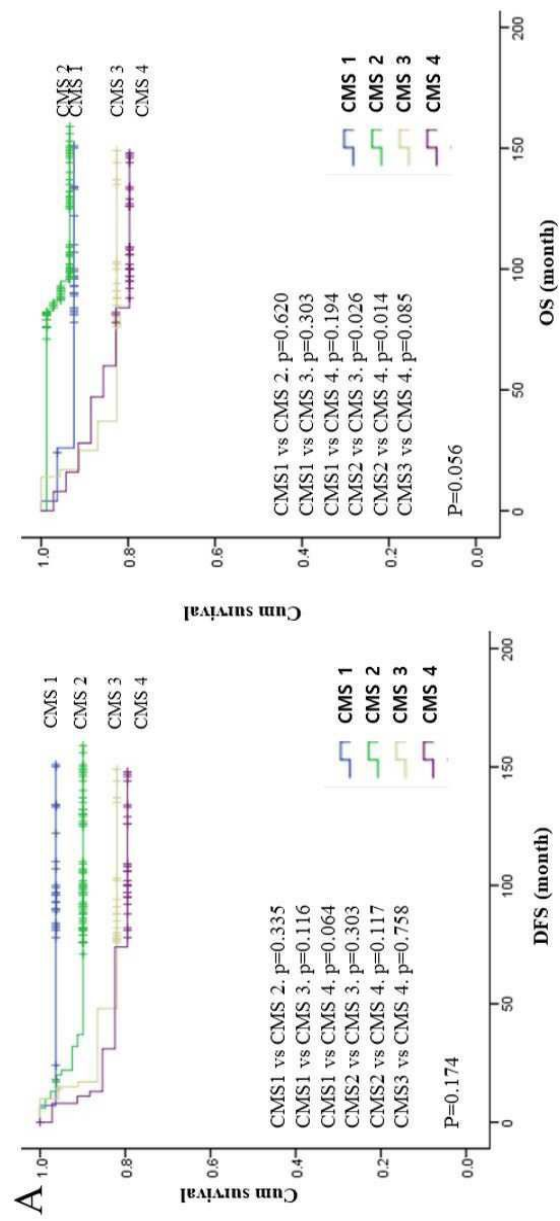
도면10a



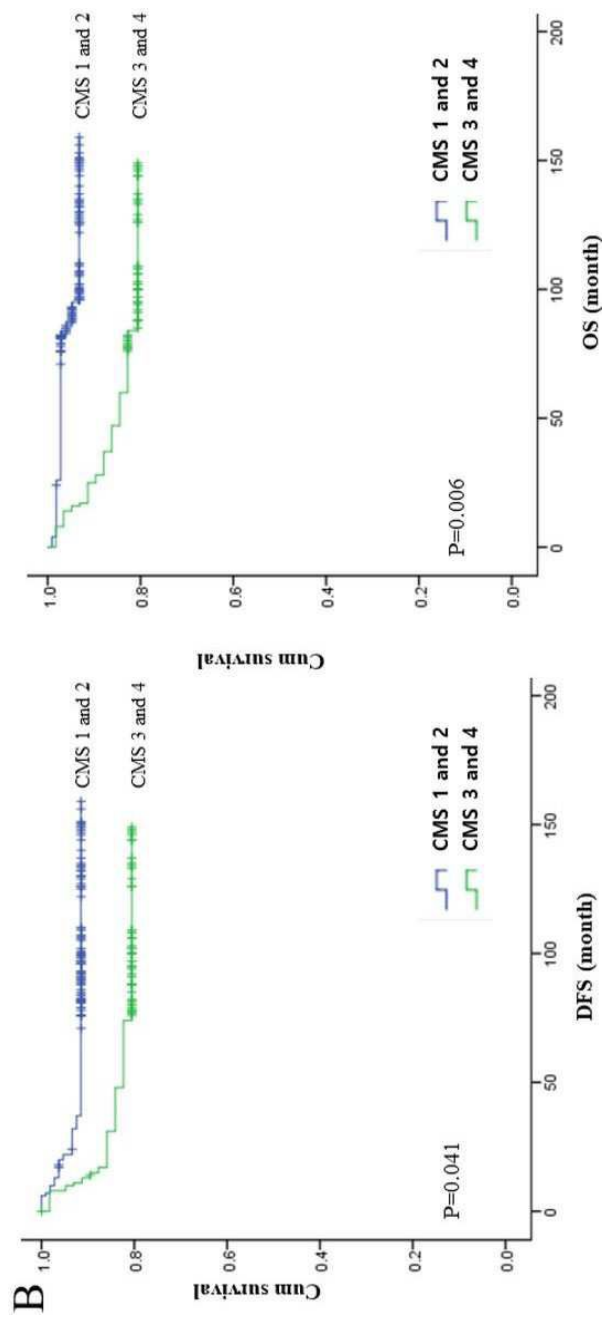
도면10b



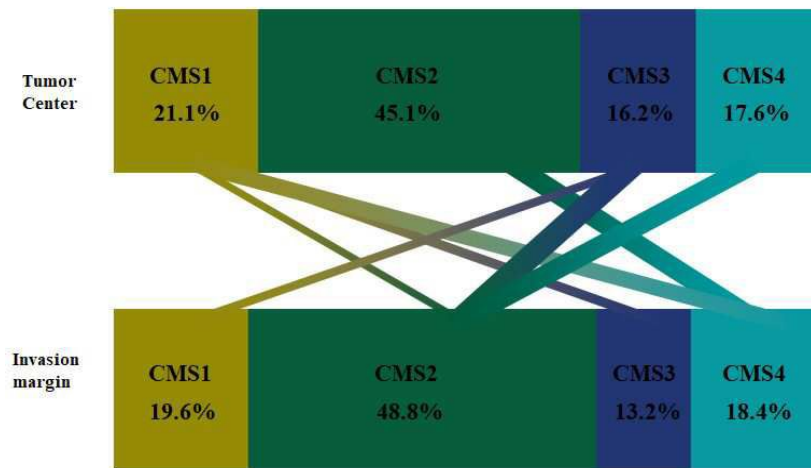
도면11a



도면11b



도면12



도면13

