



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0142334  
(43) 공개일자 2021년11월25일

- (51) 국제특허분류(Int. Cl.)  
C12Q 1/6886 (2018.01) G01N 33/574 (2006.01)  
(52) CPC특허분류  
C12Q 1/6886 (2018.05)  
G01N 33/57484 (2013.01)  
(21) 출원번호 10-2020-0059032  
(22) 출원일자 2020년05월18일  
심사청구일자 2020년05월18일

- (71) 출원인  
고려대학교 산학협력단  
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)  
(72) 발명자  
김열홍  
서울특별시 광진구 광장로1다길 56, 다동 201호(광장동, 위커향파크빌)  
최정윤  
서울특별시 성북구 종암로9길 71, 205동 102호(종암동, 종암2차 아이파크)  
(뒷면에 계속)  
(74) 대리인  
특허법인 무한

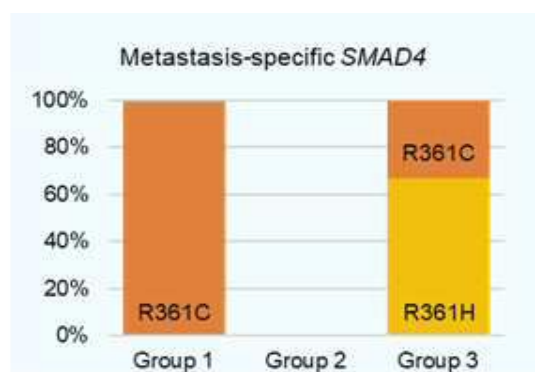
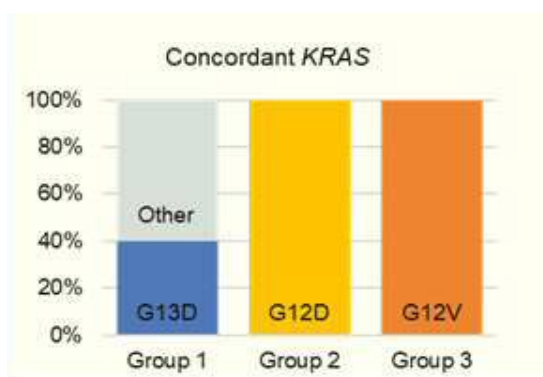
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 대장암 예후 예측용 마커 및 이의 용도

(57) 요약

본 발명은 환자의 원발암 시료 및 전이암 시료로부터 KRAS G12V, SMAD4 유전자 변이 여부를 확인하여 대장암의 예후 예측에 필요한 정보를 제공하는 방법 및 대장암의 전이 부위 예측에 필요한 정보를 제공하는 방법에 관한 것이다.

대표도 - 도2d



(52) CPC특허분류

C12Q 2600/118 (2013.01)

C12Q 2600/156 (2013.01)

C12Q 2600/158 (2013.01)

(72) 발명자

**장원진**

서울특별시 성북구 고려대로 73, 고려대학교 안암  
병원 혈액종양내과(안암동5가)

**박영수**

서울특별시 서초구 남부순환로337가길 52(서초동,  
서초신원아침도시2차)

## 명세서

### 청구범위

#### 청구항 1

- a) 대장암 환자로부터 분리된 원발암 시료 및 전이암 시료로부터 KRAS G12V 변이 여부를 확인하는 단계; 및
- b) 상기 원발암 시료 및 전이암 시료에서 KRAS G12V 변이가 검출된 경우, 대장암 예후가 좋지 않은 것으로 판정하는 단계; 를 포함하는, 대장암의 예후 예측에 필요한 정보를 제공하는 방법.

#### 청구항 2

제1항에 있어서,

상기 단계 a) 이후에, a') 상기 전이암 시료로부터 SMAD4 변이 여부를 확인하는 단계를 더 포함하는, 대장암의 예후 예측에 필요한 정보를 제공하는 방법.

#### 청구항 3

제1항에 있어서,

상기 단계 a)는, 상기 KRAS G12V 변이를 특이적으로 검출할 수 있는 검출제를 사용하여 수행되는, 대장암의 예후 예측에 필요한 정보를 제공하는 방법.

#### 청구항 4

제2항에 있어서,

상기 단계 a')는, 상기 SMAD4 변이를 특이적으로 검출할 수 있는 검출제를 사용하여 수행되는, 대장암의 예후 예측에 필요한 정보를 제공하는 방법.

#### 청구항 5

제3항 또는 제4항에 있어서,

상기 검출제는, 상기 KRAS G12V, SMAD4 변이 유전자에 특이적으로 결합하는 프라이머, 프로브 또는 앵타머인, 대장암의 예후 예측에 필요한 정보를 제공하는 방법.

#### 청구항 6

제3항 또는 제4항에 있어서,

상기 검출제는, 상기 KRAS G12V, SMAD4 변이 유전자에 의해 코딩되는 단백질에 특이적으로 결합하는 항체 또는 앵타머인, 대장암의 예후 예측에 필요한 정보를 제공하는 방법.

#### 청구항 7

제1항 또는 제2항에 있어서,

상기 KRAS G12V 변이 또는 SMAD4 변이 여부는, 직접적 시퀀싱(direct sequencing), 차세대 시퀀싱(next

generation sequencing), 표적 엑솜 시퀀싱(targeted exome sequencing), 시퀀싱 read depth 방법 및 전체 유전체 염기서열 정렬(whole genome sequence assembly)로 이루어진 군으로부터 선택되는 방법으로 확인되는, 대장암의 예후 예측에 필요한 정보를 제공하는 방법.

#### 청구항 8

제2항에 있어서,

상기 SMAD4 변이는 SMAD4 R361H인, 대장암의 예후 예측에 필요한 정보를 제공하는 방법.

#### 청구항 9

a) 대장암 환자로부터 분리된 원발암 시료 및 전이암 시료로부터 KRAS G12V 변이 여부를 확인하는 단계;

b) 상기 전이암 시료로부터 SMAD4 변이 여부를 확인하는 단계; 및

c) 상기 원발암 시료 및 전이암 시료에서 KRAS G12V 변이가 검출되고, 상기 전이암 시료에서 SMAD4 변이가 검출된 경우, 대장암이 파행성 전이(disseminated metastases) 패턴을 갖는 것으로 판정하는 단계; 를 포함하는, 대장암의 전이 부위 예측에 필요한 정보를 제공하는 방법.

#### 청구항 10

제9항에 있어서,

상기 파행성 전이 패턴은, 간, 폐, 및 림프절, 뼈, 흉막, 복막, 위, 공장, 췌장, 부신, 비장, 요관, 방광, 전립선, 난소 및 뇌로 이루어진 군으로부터 선택되는 하나 이상의 장기로 전이되는 것인, 대장암의 전이 부위 예측에 필요한 정보를 제공하는 방법.

### 발명의 설명

#### 기술 분야

[0001] 본 발명은 대장암 예후 예측용 마커 및 이의 용도에 관한 것이다.

#### 배경 기술

[0003] 대장암은 2016년 우리나라에서 발생한 암 중 남녀 합산 두 번째로 많은 암이다. 사망률도 매우 높아 폐암, 간암에 이어 3위를 기록하고 있다. 대장암은 근치적 절제술 후 20~50%에서 재발한다. 재발은 국소 재발과 원격전이로 동반되는 광범위한 재발이 많으며 이는 근치적 절제 치료의 한계성을 나타낸다. 전이성 대장암 중 간 전이는 가장 발병률이 높으며 병기가 가장 진행된 상태로 분류되고 대장암 환자의 가장 큰 사망 원인 중 하나이다. 이러한 대장암의 전이와 재발을 추적하기 위한 방법 중 혈청의 암태아성 항원 검사는 비교적 간편하여 대장암의 근치적 절제술 후 추적검사에 중요한 방법으로 쓰이고 있으나, 대장암과 무관한 양성 질환 및 흡연, 음주에 의해서도 양성 반응이 나올 수 있어 진단 활용에 한계가 있다.

[0004] 대장암 환자의 약 15~25%는 진단 당시 이미 간 전이가 있으며, 이 중 80~90%의 경우는 진단 초기에 절제 불가능한 간 전이가 있는 것으로 보고되고 있다. 따라서 대장암 및 전이성 대장암의 조기 진단을 가능케할 신규한 바이오 마커의 개발이 시급하다.

[0005] 바이오 마커는 맞춤형 의료를 앞당길 수 있는 주요 수단이 되고 있으며, DNA, RNA를 기반으로 하는 핵산 기반 바이오 마커와 단백질 및 그 일부분 등을 기반으로 하는 단백질 기반 바이오 마커 등으로 구분된다. 이러한 바이오 마커는 최근 암, 감염성 질환, 심혈관 질환, 뇌졸중, 치매 등 각종 난치병에 대한 조기진단용, 약물반응

진단 및 치료 진단용 등에 적용되고 있다.

## 선행기술문헌

### 특허문헌

[0007] (특허문헌 0001) 대한민국 공개특허 제10-2014-0033617호

## 발명의 내용

### 해결하려는 과제

- [0008] 이러한 배경 하에, 본 발명자들은 대장암 환자의 원발암, 전이암 시료로부터 얻은 유전자 변이 정보와 전이 패턴 및 환자의 생존율을 분석한 결과, 원발암 및 전이암에서의 concordant KRAS G12V 변이와 전이암 특이적 SMAD4 변이가 대장암의 예후와 밀접한 관계가 있음을 확인하여 본 발명을 완성하였다.
- [0009] 따라서, 본 발명의 목적은,
- [0010] a) 대장암 환자로부터 분리된 원발암 시료 및 전이암 시료로부터 KRAS G12V 변이 여부를 확인하는 단계; 및
- [0011] b) 상기 원발암 시료 및 전이암 시료에서 KRAS G12V 변이가 검출된 경우, 대장암 예후가 좋지 않은 것으로 판정하는 단계; 를 포함하는, 대장암의 예후 예측에 필요한 정보를 제공하는 방법을 제공하는 것이다.
- [0012] 본 발명의 다른 목적은,
- [0013] a) 대장암 환자로부터 분리된 원발암 시료 및 전이암 시료로부터 KRAS G12V 변이 여부를 확인하는 단계;
- [0014] b) 상기 전이암 시료로부터 SMAD4 변이 여부를 확인하는 단계; 및
- [0015] c) 상기 원발암 시료 및 전이암 시료에서 KRAS G12V 변이가 검출되고, 상기 전이암 시료에서 SMAD4 변이가 검출된 경우, 대장암이 파행성 전이(disseminated metastases) 패턴을 갖는 것으로 판정하는 단계; 를 포함하는, 대장암의 전이 부위 예측에 필요한 정보를 제공하는 방법을 제공하는 것이다.
- [0016] 그러나, 본 발명이 해결하고자 하는 과제는 이상에서 언급한 과제로 제한되지 않으며, 언급되지 않은 또 다른 과제들은 아래의 기재로부터 해당 기술분야의 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

### 과제의 해결 수단

- [0018] 본 발명의 일 실시예에 따르면, a) 대장암 환자로부터 분리된 원발암 시료 및 전이암 시료로부터 KRAS G12V 변이 여부를 확인하는 단계; 및 b) 상기 원발암 시료 및 전이암 시료에서 KRAS G12V 변이가 검출된 경우, 대장암 예후가 좋지 않은 것으로 판정하는 단계; 를 포함하는, 대장암의 예후 예측에 필요한 정보를 제공하는 방법이 제공된다.
- [0019] 또한, 본 발명의 방법은 상기 단계 a) 이후에, a') 상기 전이암 시료로부터 SMAD4 변이 여부를 확인하는 단계를 더 포함할 수 있다.
- [0020] 본 발명에서, 'KRAS G12V 변이'는 라스 단백질(Ras protein)의 아이소폼인 KRAS의 12번째 아미노산이 Gly에서 Val으로 치환된 변이를 의미한다.
- [0021] 본 발명에서, 'SMAD4 변이'는 SMAD family member 4, Mothers against decapentaplegic homolog 4, DPC4 (Deleted in Pancreatic Cancer-4) 등으로 알려져 있고, 인간에서 18번 염색체 상 q21.2 위치에 12개 엑손으로 구성되는 SMAD4 유전자에 존재하는 변이를 의미한다.
- [0022] 본 발명에서, SMAD4 변이는 SMAD4 유전자의 361번째 아미노산이 Arg에서 His으로 치환된 변이인 'SMAD4 R361H'일 수 있으나, 이에 한정되는 것은 아니다.
- [0023] 본 발명자들이 규명한 바에 따르면, KRAS G12V 변이가 원발암과 전이암에 모두 존재하는 일치성(concordant) 변

이에 해당하고, 전이암 특이적 SMAD4 변이가 존재하는 경우, 대장암 환자의 생존율이 현저하게 낮고, 간과 폐 외에 다른 장기로도 전이되는 파행성 전이(disseminated metastases) 패턴이 나타난다.

- [0024] 따라서, 본 발명에 따르면, 일치성(concordant) KRAS G12V 변이와 전이암 특이적 SMAD 변이 여부를 이용하여 대장암 환자의 예후를 예측하고, 대장암의 가장 흔한 전이 부위로 알려져 있는 간과 폐 뿐만 아니라 다른 장기로도 전이될 가능성을 미리 예측할 수 있다.
- [0025] 본 발명에서 '예후(prognosis)'는 질병을 진단하여 판단된 장래의 증세 또는 경과에 대한 전망을 말한다. 암 환자에 있어서 예후는 통상적으로 암 재발 또는 외과적 시술 후 일정기간 내의 전이 여부 또는 생존기간을 뜻하며, 예후 예측은 질환 치료제에 대한 환자의 반응, 치료 경과에 대한 예측도 포함된다.
- [0026] 본 발명에서 KRAS G12V 변이와 SMAD 변이를 판단하기 위한 시료는, 구체적으로 대장암 환자의 원발성 대장암 조직 및 전이성 대장암 조직에서 채취하게 된다. 또한, 대장암 조직에서의 genomic DNA 변이를 확인하기 위해 동일한 환자에서 대장암 조직에 대응하는 부위나 대장암 조직 주변의 정상 조직을 추가적으로 채취할 수도 있다. Genomic DNA 변이 분석을 위한 시료는 채취한지 얼마 안된 것(fresh sample) 또는 급속 냉동한 것(frozen sample)이 바람직하나, 포르말린으로 고정하고 파라핀에 포매(FFPE, formalin-fixed paraffin-embedded)한 조직을 시료로 하는 것도 무방하다.
- [0027] 상기 단계 a), 단계 a')는 KRAS G12V 변이를 특이적으로 검출할 수 있는 검출제와 SMAD4 변이를 특이적으로 검출할 수 있는 검출제를 사용하여 수행될 수 있다.
- [0028] 상기 검출제는, 상기 KRAS G12V, SMAD4 변이 유전자에 특이적으로 결합하는 프라이머, 프로브 또는 앵타머를 이용할 수 있다. 이들 프라이머, 프로브, 앵타머는 혼성화 여부를 검출할 수 있도록 형광단(예컨대, 플루오리신(fluorescein), 피코에리트린(phycoerythrin), 로다민, 리사민(lissamine), 그리고 Cy3와 Cy5 (Pharmacia), 발색단, 화학발광단, 자기입자, 방사능동위원소(P32 및 S35)등으로 라벨링된 것일 수 있다. 또한, 프라이머, 프로브 또는 앵타머는 KRAS, SMAD4 유전자와 유전자 주위의 염기서열 정보를 이용하여 공지된 방법에 따라 gDNA 상 적절한 위치와 염기서열의 조성을 선택하여 제작될 수 있다.
- [0029] KRAS G12V, SMAD4 변이 유전자를 검출하기 위한 방법으로는, 역전사효소 중합효소반응, 경쟁적 역전사효소 중합효소반응, 실시간 역전사효소 중합효소반응, RNase 보호 분석법, 노던 블롯(northern blot), DNA 마이크로어레이 칩 등을 예로 들 수 있으나, 이들에 한정되는 것은 아니다.
- [0030] 또한, 상기 KRAS G12V, SMAD4 변이 유전자에 의해 코딩되는 단백질에 특이적으로 결합하는 항체 또는 앵타머 이용할 수도 있다. 상기 항체는 폴리클로날 또는 모노클로날 항체이며, 바람직하게는 모노클로날 항체이고, 검출 가능한 시그널을 발생시키는 라벨을 더 포함할 수 있다. 상기 라벨은 화학물질(예를 들어, 바이오틴), 효소(알칼린 포스파타아제,  $\beta$ -갈락토시다아제, 호스 레디쉬 퍼옥시다아제 및 사이토크롬 P450), 방사능물질(예컨대, C14, I125, P32 및 S35), 형광물질(예를 들어, 플루오레신), 발광물질, 화학발광물질(chemiluminescent) 및 FRET(fluorescence resonance energy transfer)을 이용할 수 있으나, 이에 한정되는 것은 아니다.
- [0031] KRAS G12V, SMAD4 변이 단백질을 검출하는 것은 면역분석(immunoassay), 즉, 항원-항체 반응을 이용하는 방법을 통해 수행될 수 있다. 이러한 면역분석은 종래에 개발된 다양한 정량적 또는 정성적 면역분석 프로토콜에 따라 실시될 수 있다. 상기 면역분석 포맷은 방사능면역분석, 방사능면역침전, 면역침전, 면역조직화학염색, ELISA(enzyme-linked immunosorbent assay), 캡처-ELISA, 억제 또는 경계 분석, 샌드위치 분석, 유세포 분석(flow cytometry), 면역형광염색 및 면역친화성 정제를 포함하지만, 이에 한정되는 것은 아니다.
- [0032] 나아가, 상기 KRAS G12V 변이와 SMAD4 변이는, 시퀀싱에 기반한 방법인 직접적 시퀀싱(direct sequencing), 차세대 시퀀싱(next generation sequencing), 표적 엑솜 시퀀싱(targeted exome sequencing), 시퀀싱 read depth 방법, 전체 유전체 염기서열 정렬(whole genome sequence assembly)로 이루어진 군으로부터 선택되는 방법을 이용하여 확인될 수 있다.
- [0033] 본 발명의 다른 일 실시예에 따르면,
- [0034] a) 대장암 환자로부터 분리된 원발암 시료 및 전이암 시료로부터 KRAS G12V 변이 여부를 확인하는 단계;
- [0035] b) 상기 전이암 시료로부터 SMAD4 변이 여부를 확인하는 단계; 및
- [0036] c) 상기 원발암 시료 및 전이암 시료에서 KRAS G12V 변이가 검출되고, 상기 전이암 시료에서 SMAD4 변이가 검출된 경우, 대장암이 파행성 전이(disseminated metastases) 패턴을 갖는 것으로 판정하는 단계; 를 포함하는, 대

장암의 전이 부위 예측에 필요한 정보를 제공하는 방법이 제공된다.

[0037] 상기 파행성 전이 패턴은, 간, 폐, 및 림프절, 뼈, 흉막, 복막, 위, 공장, 췌장, 부신, 비장, 요관, 방광, 전립선, 난소 및 뇌로 이루어진 군으로부터 선택되는 하나 이상의 장기로 전이되는 것을 의미한다.

[0038] 따라서, 본 발명의 방법을 이용하면, 유전자 변이를 통해 대장암 환자의 원발암 시료 및 전이암 시료 모두에서 KRAS G12V 변이가 검출되고, 전이암 시료에서 SMAD4 변이가 존재하는 경우, 가장 흔한 전이 부위로 알려져 있는 간과 폐뿐만 아니라 다른 장기로도 전이될 가능성을 미리 예측할 수 있다.

## 도면의 간단한 설명

[0040] 도 1a ~ 1d는 원발성 종양(P)과 매칭되는 전이성 종양(M1) 사이의 게놈 프로파일링에 관한 것이다.

도 1a는 TCGA 데이터와 비교하여 원발성 종양과 전이성 종양 사이의 유해한 단백질 변형 유전자 변이(non-synonymous variants) 빈도를 나타낸다.

도 1b는 일치성(concordant), 원발 특이적, 전이암 특이적 변이 개념의 개략도 및 변이 판도(mutational landscape)를 나타낸다.

도 1c는 원발성 종양 및 전이성 종양 사이에서 원발 특이적, 전이암 특이적 또는 일치성으로 식별된 변이의 백분율을 비교한 것이다(유해한 비동일성 변형이 없는 12명의 환자를 제외함).

도 1d는 주요 유전자의 상이한 단백질 변화 및 TCGA 데이터와의 비교에 기초한 빈도를 나타낸다(X: 변이 수, Y: 위치). 상이한 단백질 변화는 상이한 변이체의 위치를 나타낸다.

도 2a ~ 2d는 게놈 그룹화 및 생존을 분석에 관한 것이다.

도 2a는 PC1, PC2 및 PC3 클러스터링을 사용하여 3개의 게놈 그룹을 결정하는 것을 나타낸다(그룹 1 (117, 74.0 %, 빨간색), 그룹 2 (24, 15.2 %, 녹색) 및 그룹 3 (17, 10.8 %, 검정색)).

도 2b는 3개 그룹의 전체 생존(Overall survival, OS) 데이터를 나타낸다(전체 log-rank P=0.028).

도 2c는 그룹 간, 일치성, 원발 특이적, 전이암 특이적 유전자 간 빈도를 나타낸다. 상위 20개의 유전자를 일치성(빨간색), 원발 특이적(녹색), 전이암 특이적(파란색)으로 나누어 나타냈고, X축은 유전자, Y는 각 그룹에서 변이를 가진 환자의 백분율을 나타낸다.

도 2d는 그룹 간 일치성 KRAS 및 전이암 특이적 SMAD4 변이의 비율을 나타낸다.

도 3은 짝을 이룬 유전자(일치성, 원발 특이적, 전이암 특이적)와 전이 부위(간만, 폐만, 간과 폐 및 다른 장기)의 상관 관계를 나타내는 것으로, (A)는 카이 제곱 검정 결과이고(진한 파란색은  $P < 0.05$ 를 의미), (B)는 랜덤 포레스트 모델을 이용한 분석 결과이다.

## 발명을 실시하기 위한 구체적인 내용

[0041] 이하에서, 첨부된 도면을 참조하여 실시예들을 상세하게 설명한다. 그러나, 실시예들에는 다양한 변경이 가해질 수 있어서 특허출원의 권리 범위가 이러한 실시예들에 의해 제한되거나 한정되는 것은 아니다. 실시예들에 대한 모든 변경, 균등물 내지 대체물이 권리 범위에 포함되는 것으로 이해되어야 한다.

[0042] 실시예에서 사용한 용어는 단지 설명을 목적으로 사용된 것으로, 한정하려는 의도로 해석되어서는 안된다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 명세서 상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0043] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

- [0044] 또한, 첨부 도면을 참조하여 설명함에 있어, 도면 부호에 관계없이 동일한 구성 요소는 동일한 참조부호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다. 실시예를 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 실시예의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다.
- [0045] 또한, 실시 예의 구성 요소를 설명하는 데 있어서, 제 1, 제 2, A, B, (a), (b) 등의 용어를 사용할 수 있다. 이러한 용어는 그 구성 요소를 다른 구성 요소와 구별하기 위한 것일 뿐, 그 용어에 의해 해당 구성 요소의 본질이나 차례 또는 순서 등이 한정되지 않는다. 어떤 구성 요소가 다른 구성요소에 "연결", "결합" 또는 "접속"된다고 기재된 경우, 그 구성 요소는 그 다른 구성요소에 직접적으로 연결되거나 접속될 수 있지만, 각 구성 요소 사이에 또 다른 구성 요소가 "연결", "결합" 또는 "접속"될 수도 있다고 이해되어야 할 것이다.
- [0046] 어느 하나의 실시 예에 포함된 구성요소와, 공통적인 기능을 포함하는 구성요소는, 다른 실시 예에서 동일한 명칭을 사용하여 설명하기로 한다. 반대되는 기재가 없는 이상, 어느 하나의 실시 예에 기재한 설명은 다른 실시 예에도 적용될 수 있으며, 중복되는 범위에서 구체적인 설명은 생략하기로 한다.
- [0047] <실험 방법>
- [0048] 1. 환자 및 종양 샘플의 선정
- [0049] 고려대학교 안암 병원(서울)에서 2005 년 9월부터 2018년 8월 사이에 진행성 또는 전이성(IV기) 대장암(CRC) 환자를 대상으로 연구를 실시했다. 연구 포함 기준은 전이암 절제술(metastasectomy)이었다. 종양 조직학, 의학적 동반 질환, 이전 치료, 선행 치료 또는 수행 상태에 대한 제외는 없었다. 게놈 프로파일링은 Ion Torrent Proton system과 함께 Ion AmpliSeq Cancer Hotspot Panel (ICP)을 이용하여 수행되었다. 이 연구는 고려 대학교 안암 병원 임상 시험 심사위원회 (2017AN0070)의 승인을 받았다.
- [0051] 2. 조직 genomic DNA 추출
- [0052] 제조사의 프로토콜에 따라 QIAamp DNA FFPE 조직 키트 (Qiagen, Valencia, CA)를 사용하여 포르말린 고정 파라핀 포매 (FFPE) 조직으로부터 조직의 genomic DNA (gDNA)를 추출하고 50 µL 부피로 용출 시켰다. 추출된 gDNA의 순도는 1% 아가로스겔 전기 영동으로 평가하였고, DNA 농도는 Agilent High Sensitivity DNA Kit(Agilent Technologies, Santa Clara, CA, USA)를 사용하여 Qubit 2.0 Fluorometer에서 정량화 하였다.
- [0054] 3. 차세대 시퀀싱(Next Generation Sequencing)
- [0055] 전체적으로, 최대 10 ng의 gDNA를 FFPE 샘플로부터 추출하고, 각 샘플에 대해 바코드를 사용하여 20 사이클에 걸쳐 Ion AmpliSeq Library Kit 2.0(Life Technologies, Carlsbad, CA)을 사용하여 증폭시켰다. 라이브러리 농도는 QuantStudio Real-Time PCR System (Thermo Fisher Scientific)(Thermo Fisher Scientific)을 사용하여 평가하였다. 희석된 각 라이브러리 (100 pM)를 OneTouch™ Instrument(Life Technologies)를 사용하여 에멀전 PCR로 증폭시키고 제조사의 지침에 따라 Ion PI Hi-Q OT2 200 키트를 사용하여 OneTouch™ES Instrument(Life Technologies)로 농축하였다. 마지막으로, Ion PI Hi-Q Sequencing 200 Kit(Life Technologies)를 사용하여 Ion Proton instrument(Life Technologies)에서 시퀀싱을 수행했다. 바코드 샘플을 Ion PI Chip v3에 로딩하고, 시퀀싱 리드(read) 매핑(mapping) 및 변이 검출(variant calling)은 Ion Torrent Suite v5.0.4.0에서 수행되었다.
- [0057] 4. 변이 주석(Variant annotation)
- [0058] 데이터 분석은 Torrent Suite Software V.5.0 (Life Technologies, Carlsbad, CA, USA)을 사용하여 수행하였다. hg19 인간 기준 게놈에 정렬시킨 후, Variant Caller 플러그인을 적용하고, Ion Reporter suite (Life Technologies)을 사용하여 다형성 변이를 필터링했다. 또한, 1% 미만의 변이 빈도로 존재하는 모든 뉴클레오타이드 변이는 마스킹하였고, 이는 passenger mutation이 검출되고 일치된 샘플이 여과됨에 따라 이전 연구에서 사용된 5%의 값과 비교하여 상대적으로 낮은 빈도로 존재하는 변이를 식별하는데 유용하였다. 검출된 모든 변이체는 Integrative Genomics Viewer(IGV V.2.1, Broad Institute, MA, USA, MA)를 이용하여 수동 검토되었다. 그 다음, 5000Exomes, dbSNP, Canonical RefSeq Transcripts, ClinVar, DGV, DrugBank, Gene Ontology, COSMIC, OMIM, Pfam, PhyloP Scores, RefGene Functional Canonical Transcripts Scores, RefSeq GeneModel

및 Named 변이체 데이터베이스를 사용하여 공통 데이터베이스 중의 변이 존재 여부를 평가하였다. 병원성 변이는 COSMIC 및 ClinVar 데이터베이스에서 "유사병원성(likely pathogenic)" 또는 "병원성(pathogenic)"으로 주석이 달렸다. 병원성 변이의 기능적 중요성을 고려하여, 본 연구는 단백질 변형 유전자 변이가 해로운 영향을 미칠 것으로 예상하였다. ICP 패널을 사용한 표적 시퀀싱은 표적당 평균 96.63%의 샘플로, 샘플 당 약 100Mb를 생성하였고, 모든 샘플의 서열은 3,900×의 평균 딥스(depth)를 달성하였다.

## 5. 연구 설계 및 정의

본 연구는 원발성 종양과 전이성 종양 간 유전자 변이를 비교함으로써 일치(concordant) 또는 불일치(discordant) 게놈 프로파일을 확인하였다. 본 발명자들은 원발성 종양 및 일치된 전이성 종양 모두에 존재하는 것을 일치성 변이(concordant variant)로 정의하였다. 원발-특이적 변이는 원발성 종양에만 존재하는 것으로 정의하였고, 전이-특이적 변이는 전이성 종양에만 존재하는 것으로 정의하였다. 이러한 게놈 불일치(discrepancy) 패턴을 사용하여 생존과 관련된 게놈 그룹을 분류했습니다. 전체 생존(Overall survival, OS)은 전이 절제술(metastectomy)에서 사망에 이르기까지의 기간으로 정의하였다. 무병 생존(Disease-free survival, DFS)은 전이 절제술에서 재발 및/또는 진행까지의 기간으로 정의하였다.

## 6. 통계 분석

모든 통계적 분석은 R v. 3.1.2 소프트웨어 (<https://www.r-project.org/>) 및 SPSS v. 22.0 (SPSS, Inc., Chicago, IL, USA)을 사용하여 수행되었다. 기술 통계량은 범주형 변수의 빈도 분포와 연속 변수의 수단으로 요약되었다. 맥네마 검정(McNemar test)은 원발성 및 일치된 전이성 종양의 유전자 빈도를 분석하는 데 사용되었다. PCA(Principal Component Analysis) 및 partitioning around medoids (PAM) 클러스터링 알고리즘을 사용하여 게놈 그룹을 분류하였다. 생존율은 Kaplan-Meier법과 로그 순위 검정으로 분석하였다. 콕스 비례 위험 모델을 이용한 단변량 및 다변량 분석을 수행하여 예후 인자를 확인하였다. 피셔 정확 검정(Fisher's exact test)을 사용하여 명목 변수(Nominal variable)를 비교했다. 카이 제곱 검정과, 랜덤 포레스트 플롯으로 머신러닝 알고리즘을 이용하여 게놈 패턴과 전이 부위 사이의 상관 관계를 식별하고, 평균 비교를 위해 맨-휘트니 검정(Mann-Whitney test)을 이용하였다. 전체 변이의 유무에 따라 개별 환자의 여러 표본의 계통 발생을 재구성하기 위해 완전 연결법(complete linkage method)을 사용했다. P-value<0.05는 통계적으로 유의한 것으로 간주되었다.

### 실시예 1. 원발성 및 일치된 전이성 종양의 유전자 변이 비교

#### 1-1. 분석 대상 환자 모집단과 유전자 변이 데이터 셋

원발성 대장암과 전이성 대장암의 게놈 프로파일을 비교하기 위해 158명의 환자로부터 335개의 수술 샘플을 수집했다. 표본은 158쌍의 원발성 종양(P)과 일치된 전이성 종양(M1) 및 2차 전이가 일어난 19쌍의 종양(M2)으로 구성되었으며, 각 쌍은 동일한 환자로부터 파생되었다. 158쌍의 P와 M1 종양을 기준으로 했을 때 전이 절제술의 평균 연령은 61세였고, 이 그룹은 96명(60.8%)의 남성으로 구성되었다. 초기 단계는 6명이 1기암(3.8%), 24명이 2기암(15.2%), 49명이 3기암(31.0%), 79명이 4기암(50.0%)이었다. 전이 유형은 원발성 종양 조직과 전이 종양 조직이 6개월 내에 채취되었을 때(n = 71, 44.9%) 동시성 전이(synchronous metastasis)로 정의하고, 차이가 6개월 이상일 때(n = 87, 55.1%) 이시성 전이(metachronous metastasis)로 정의했다. 원발성 종양의 가장 흔한 부위는 직장(n = 80, 50.6%), 왼쪽 결장(n = 54, 34.2%), 오른쪽 결장(n = 24, 15.2%) 순서로 조사되었다. 가장 흔한 전이 부위는 간(n = 95, 60.1%), 폐(n = 56, 35.4%), 기타(n = 7.4%) 순이었다.

316개의 샘플에서 50개의 유전자를 선별한 결과(P와 M1 종양 158쌍) 744종의 변이가 나타났다. 이러한 변이 대부분은 이전에 COSMIC 또는 ClinVar 데이터베이스에서 보고된 바가 있다(744개 중 412개, 55.4%). 이 중에서, 412개(75.2%) 중 310개는 유사 병원성 또는 병원성 변이일 가능성이 높은 것으로 나타났다. 이들 310개 중에서, 미스센스 및 넌센스 변이를 포함하는 282개(91.0%)의 단백질 변형 유전자 변이(non-synonymous variant)를 분석한 결과, 이는 해로운 영향을 미칠 것으로 예상되었고, 본 연구는 이들을 유해한 단백질 변형 유전자 변이로 정의했다.

[0070] 1-2. 원발성 종양 및 일차된 전이성 종양의 게놈 프로파일링

[0071] 상기 실시예 1-1에서 얻은 원발성 종양(P)과 전이성 종양(M1) 사이의 기본 게놈 프로파일링은 도 1에 나타났다. 구체적으로, 원발성 종양과 전이성 종양 사이의 유해한 단백질 변형 유전자 변이 빈도에서 TP53, KRAS, APC 및 PIK3CA는 원발성 대장암과 그 전이암에서 가장 많았다(도 1a). 원발성 종양과 전이성 종양 간 돌연변이 빈도에는 전반적인 통계적 차이가 없었다. 그러나, KRAS, APC, KIT, FBXW7, SMAD4, PTEN, CDKN2A의 경우, 원발성 종양에서의 돌연변이 빈도가 전이성 종양보다 높았다. 특히, APC 돌연변이는 원발성 종양과 전이성 종양 간 차이가 가장 높았으나, 통계적으로 유의하지는 않았다(32.3% vs 26.6%, McNemar test  $p=0.057$ ). TCGA (Cancer Genome Atlas)의 보고와 비교하였을 때, 전체 돌연변이 빈도수는 대장암(원발암) 샘플의 돌연변이 빈도수가 비슷한 것을 알 수 있다. 그러나 연구에서는 TCGA에 비해 APC 돌연변이 발생 빈도가 낮았다(원발성 32.3%, 전이성 27.2% vs TCGA 73.2%).

[0072] 종양 내 이질성을 식별하기 위해, 원발암과 전이암의 돌연변이 일치와 불일치를 확인하였다(도 1b). 유전적 변이를 일치성(concordant), 원발암 특이적(primary-specific), 전이암 특이적(metastasis-specific) 변이로 분류하고, 158명의 환자에 대한 유전자 변이의 전체적인 변이 판도(mutational landscape)를 확인하였다. 변이 판도로부터, 각 환자와 각 유전자에 따른 변이의 분포를 확인하였다(도 1b, mutational landscape). 모든 유해한 단백질 변형 유전자 변이 중에서, 평균 60%가 일치하고 평균 40%가 원발암과 전이암 사이에서 불일치였고, 원발암 특이적 변이는 24%, 전이암 특이적 변이의 경우 16%로 나타났다(도 1c). 총 일치 변이 수는 538개(원발암 종양과 전이암 종양 모두에 존재하여 269개의 두배)이며, 불일치 변이는 359개였다. 환자 당 평균 일치 변이 수는 3.4개였고 불일치 변형은 2.3개였다. 불일치 변이는 원발암(215/359, 60%)에서 더 빈번하게 발생하였고, 전이암(144/359, 40%)은 그렇지 않았다.

[0073] 도 1 d는 대장암 환자에서 관찰된 주요 유전자의 단백질 변화에 기초한 빈도를 보여준다. TP53, KRAS, APC, NRAS는 원발암, 전이암 공통 돌연변이의 비율이 컸지만, 불일치 돌연변이의 비율은 PIK3CA, SMAD4, BRAF, PTEN의 유전자에서 큰 것으로 나타났다. 원발성 종양에서 단백질 변화의 빈도는 TCGA 데이터와 비교하였을 때, KRAS, PIK3CA, NRAS, BRAF는 TCGA 데이터와 유사한 빈도수를 보였다. TCGA와 일치하는 단백질 변화는 KRAS G12D, PIK3CA E545K/G, NRAS Q61L/R, BRAF V600E이었다. TP53, APC, SMAD4, PTEN은 TCGA 데이터와 차이를 보였다. 공통된 단백질 변화는 TP53 P152R, APC R876\*, SMAD4 A118V, PTEN W111\*이었다.

[0075] 실시예 2. 게놈 그룹화 및 생존 분석

[0076] 상기 실시예 1에서 확인한 게놈 프로파일을 이용하여 그룹화 및 생존 분석을 수행하였다. PCA는 변이 유형과 변종 위치의 수를 포함하여 가능한 상관 변수 집합을 선형적으로 상관되지 않는 변수 집합(주요 구성요소, PC)으로 변환하는 데 사용되었다. PAM 클러스터링 알고리즘에는 세 가지 개별 PC(PC1, PC2, PC3)가 사용되었다. Silhouette method를 사용하여 환자를 분류한 결과, 최적의 게놈 집단은 3개의 그룹으로 확인되었다(도 2a). 3개의 게놈 그룹 간 생존율을 비교한 결과(Overall log-rank,  $p=0.008$ ), 도 2b에 나타난 바와 같이 그룹 3의 생존율은 그룹 1 ( $p=0.002$ )보다 상당히 짧았다. 이는 다변량 콕스 회귀 모델(multivariate Cox regression model)을 이용하여 전이 절제술, 연령, 성별, 초기 기수, 전이 유형, 원발암의 부위, 전이 부위 및 전이암의 지름을 포함한 다른 변수에 대해 조정한 후에도 유지되었다(전체  $p=0.009$ , 표 2).

표 1

[0077]

| 변수            | 단변량 분석             |         | 다변량 분석             |         |
|---------------|--------------------|---------|--------------------|---------|
|               | 조정된 HR<br>(95% CI) | p-value | 조정된 HR<br>(95% CI) | p-value |
| Adjusted 연령   | 1.04 (0.99-1.10)   | 0.117   | -                  | -       |
| 성별            |                    |         |                    |         |
| 남성            | 1 (reference)      |         | -                  |         |
| 여성            | 3.29 (1.11-9.71)   | 0.031   | -                  | -       |
| 초기 기수         |                    | 0.569   |                    |         |
| Stage I or II | 1 (reference)      |         | -                  |         |
| Stage III     | 0.64 (0.17-2.39)   | 0.504   | -                  |         |
| Stage IV      | 0.53 (0.16-1.74)   | 0.295   | -                  | -       |
| 전이 유형         |                    |         |                    |         |

|                   |                   |       |                   |       |
|-------------------|-------------------|-------|-------------------|-------|
| 동시성(Synchronous)  | 1 (reference)     |       | -                 |       |
| 이시성(Metachronous) | 1.49 (0.53-4.20)  | 0.450 | -                 | -     |
| <b>원발암 부위</b>     |                   | 0.396 |                   |       |
| 직장                | 1 (reference)     |       | -                 |       |
| 왼쪽 결장             | 0.85 (0.25-2.87)  | 0.797 | -                 |       |
| 오른쪽 결장            | 2.06 (0.58-7.33)  | 0.263 | -                 | -     |
| <b>전이 부위</b>      |                   |       |                   |       |
| 간                 | 1 (reference)     |       | -                 |       |
| 폐 및 기타 조직         | 1.45 (0.52-4.04)  | 0.473 | -                 | -     |
| <b>전이암의 지름</b>    | 1.01 (1.00-1.02)  | 0.016 | 1.01 (1.01-1.02)  | 0.002 |
| <b>계놈 그룹</b>      |                   | 0.019 |                   | 0.009 |
| 그룹 1              | 1 (reference)     |       | 1 (reference)     |       |
| 그룹 2              | 2.38 (0.60-9.56)  | 0.220 | 3.61 (0.81-16.02) | 0.092 |
| 그룹 3              | 5.16 (1.65-16.14) | 0.005 | 6.98 (2.01-24.20) | 0.002 |

[0078] 계놈 그룹에 의한 하위군 분석에서 전체 생존(Overall survival)은 그룹 1의 환자(위험비율, 6.98; 95% 신뢰 구간 [CI], 2.01 - 24.20;  $p=0.002$ )보다 그룹 3의 환자가 훨씬 짧았다. 단, 무병 생존(DFS)(Overall log-rank  $p=0.252$ )에는 큰 차이가 없었다. 각 그룹의 특성을 파악하기 위해 주요 유전자의 빈도와 특정 단백질 변화를 확인했다(도 2c). 빈도수 분석에서 생존이 양호한 그룹 1은 전체적인 일치성(concordant), 원발암 특이적, 전이암 특이적인 유전자의 빈도가 가장 낮았다. 그룹 2는 그룹 1과 그룹 3 사이의 중간 빈도수를 나타냈으나, 원발암 및 전이암 모두에서 변이가 발견된 TP53은 그룹 2에서 가장 높은 돌연변이 빈도수를 보였다. 생존율이 낮은 그룹 3은 원발암, 전이암 모두에서(concordant) KRAS(100%)와 PIK3CA(41.2%)의 변이 빈도가 높은 것으로 나타났고, 5개의 전이암 특이적 유전자(TP53, APC, SMAD4, CTNNB1, PTEN)의 변이 빈도가 다른 그룹에 비해 높았다.

[0079] 생존이 낮은 그룹 3의 특성을 조사하기 위해 그룹 3의 특성 유전자의 단백질 변화 비율을 다른 그룹과 비교하였다(도 2d). 그룹 3에서 다른 그룹과 가장 차이를 나타낸 일치성(concordant) KRAS 돌연변이의 경우, 예후가 좋은 그룹 1에서는 G13D의 비율이 가장 높았고, 예후가 좋지 않은 그룹 3은 100% G12V 변이를 가졌다. 또한, 그룹 2는 G12D가 100%로 나타나 그룹 3과는 분명히 구별되었다. TP53, APC, PIK3CA와 같은 다른 주요 유전자의 변이 비율도 비교하였으나, 그룹 1이나 그룹 3에서 고유한 단백질 변화가 관찰되지는 않았고, 공통적으로 분포되어 있었다.

[0080] 전이암 특이적 유전자의 경우 그룹 3에서는 6개의 유전자, TP53, APC, PIK3CA, SMAD4, CTNNB1 및 PTEN가 확인되었고, 해당 유전자 중 SMAD4가 그룹 3과 다른 그룹을 구별하는 전이유전자로 분석되었다(Fisher's exact test  $p=0.007$ ). 그리고 전이암 특이적 SMAD4의 경우 예후가 좋지 않은 그룹 3에서 R361H 변이가 가장 높은 비율을 차지했다(도 2d).

### [0082] 실시예 3. 유전자 및 전이 부위의 상관관계 분석

[0083] 특정한 두 유전자가 전이 부위에 영향을 미치는지 여부를 확인하기 위해 chi-squared test와 random forest model을 사용하여 분석하였다. 환자의 전체 추적 기간 동안 쌍을 이루는 유전자(일치성(concordant), 원발암 특이적, 전이암 특이적)와 전체 전이 부위(liver only, lungs only, both liver and lungs, others) 사이의 상관관계를 분석하였다(도 3). 원발암 및 전이암 모두에서 변이가 발생하는 유전자인 KRAS는 chi-squared test( $p=0.014$ )로 확인한 결과, 간과 폐 이외에 다른 부위인 림프절, 뼈, 흉막, 복막, 위, 공장, 췌장, 부신, 비장, 요관, 방광, 전립선, 난소 및 뇌 전이에서도 유의하게 연관성이 있는 것으로 나타났다. Random forest model에서는 KRAS, APC, TP53, KIT, SMAD4의 순서로 전이 부위에서 각 유전자의 상호작용으로 인한 조건부 효과를 고려할 때 가장 큰 영향을 미치는 상위 5개 유전자로 확인되었다. 원발암 특이적 유전자에서 BRAF는 chi-squared test ( $p=0.002$ )에서 폐 전이성과 유의미한 상관관계를 보였다. Random forest model에서는 5개의 유전자, TP53, BRAF, RET, KRAS, PTEN가 확인되었다. 전이암 특이적 유전자와 관련하여 TP53과 PTEN은 간 및 폐 전이(각각  $p=0.004$ 와  $0.016$ )와 유의미한 상관관계를 보였다. Random forest model에서는 TP53, PTEN, KRAS, PIK3CA, ATM 등이 가장 많은 영향을 주는 5개 유전자로 확인되었다.

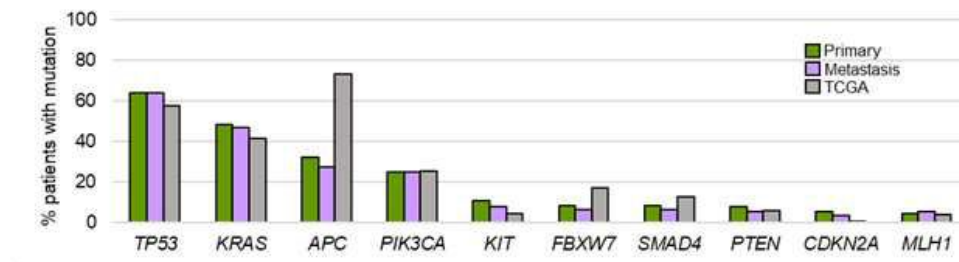
[0084] 이상과 같이 실시예들이 비록 한정된 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기를 기초로 다양한 기술적 수정 및 변형을 적용할 수 있다. 예를 들어, 설명된 기술들이 설명된 방법과 다

른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

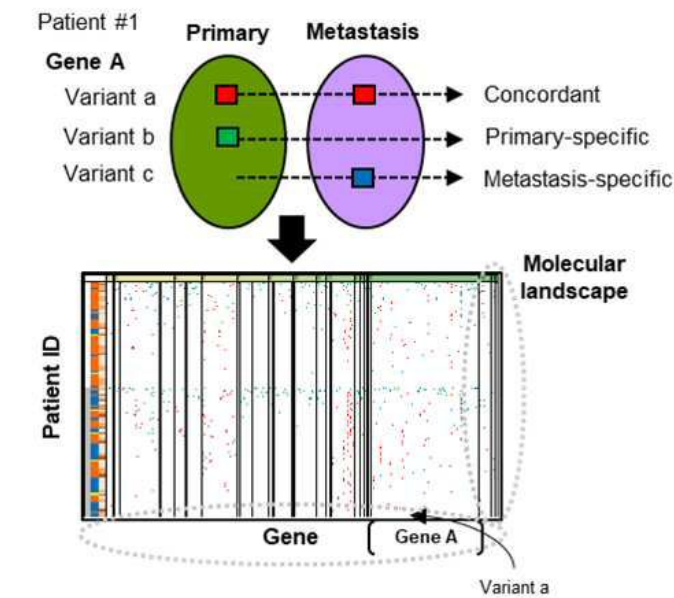
[0085] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 청구범위의 범위에 속한다.

## 도면

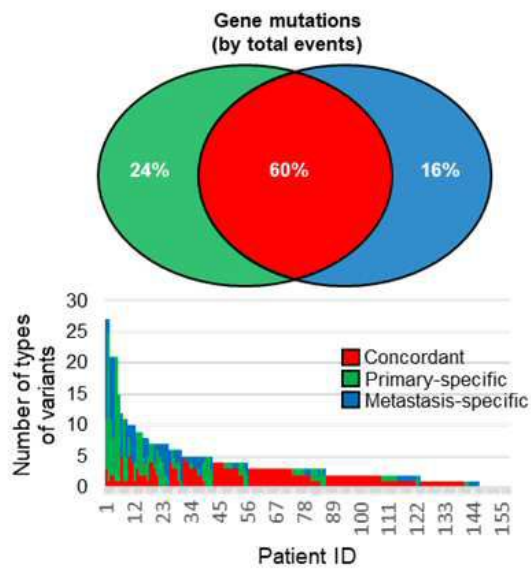
### 도면1a



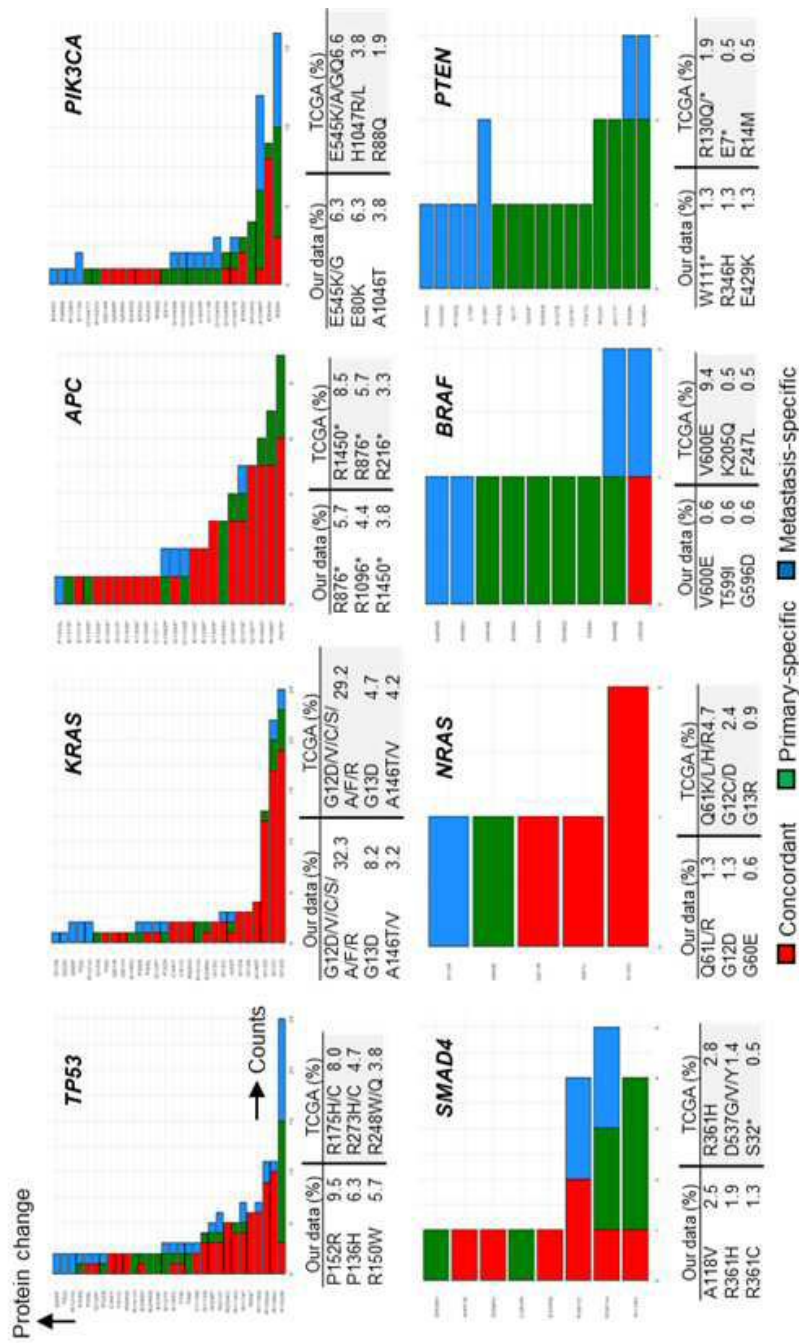
### 도면1b



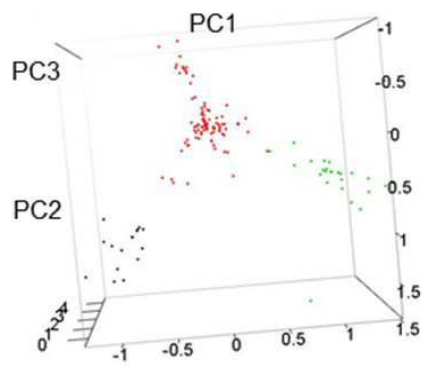
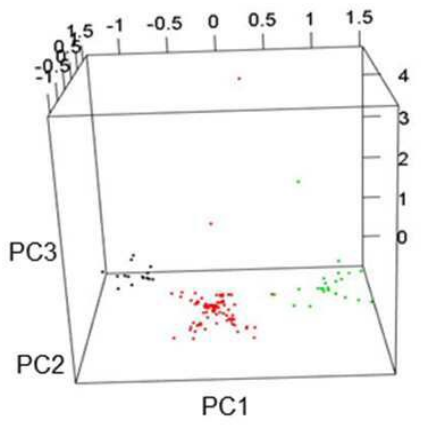
도면1c



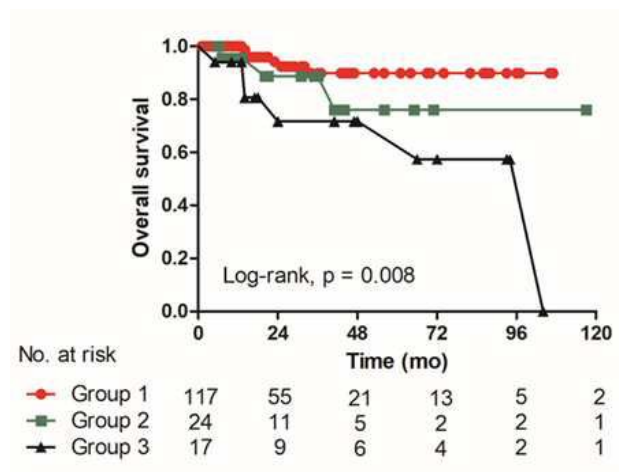
도면1d



도면2a



도면2b



도면2c



도면2d



도면3

