



**(19) 대한민국특허청(KR)**  
**(12) 등록특허공보(B1)**

(45) 공고일자 2019년02월07일  
 (11) 등록번호 10-1945093  
 (24) 등록일자 2019년01월28일

(51) 국제특허분류(Int. Cl.)  
 G16B 20/00 (2019.01) G16B 30/00 (2019.01)  
 (52) CPC특허분류  
 G16B 20/00 (2019.02)  
 G16B 30/00 (2019.02)  
 (21) 출원번호 10-2016-7036845  
 (22) 출원일자(국제) 2015년06월01일  
 심사청구일자 2017년06월09일  
 (85) 번역문제출일자 2016년12월29일  
 (65) 공개번호 10-2017-0077063  
 (43) 공개일자 2017년07월05일  
 (86) 국제출원번호 PCT/US2015/033497  
 (87) 국제공개번호 WO 2015/184439  
 국제공개일자 2015년12월03일  
 (30) 우선권주장  
 62/005,766 2014년05월30일 미국(US)  
 (56) 선행기술조사문헌  
 KR1020130105296 A\*  
 \*는 심사관에 의하여 인용된 문헌

(73) 특허권자  
 난토믹스, 엘엘씨  
 미국 캘리포니아 컬버 시티 제퍼슨 블리바드 9920  
 (우:90232)  
 난트 홀딩스 아이피, 엘엘씨  
 미국 90232 캘리포니아 컬버 시티 제퍼슨 블러바  
 드 9920  
 (72) 발명자  
 산본, 존 자카리  
 미국, 캘리포니아 95060, 산타 크루즈, 쿠파 스트  
 리트 101  
 바스크, 찰스 조셉  
 미국, 캘리포니아 95060, 산타 크루즈, 쿠파 스트  
 리트 101  
 (뒷면에 계속)  
 (74) 대리인  
 한인열

전체 청구항 수 : 총 20 항

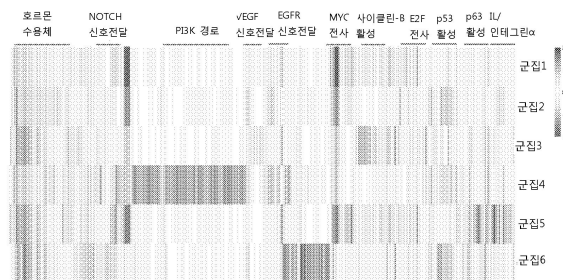
심사관 : 김지윤

(54) 발명의 명칭 **다중 종양 및 생식세포 유전자 엑솜에 걸친 분자 프로파일의 종합 분석을 위한 시스템 및 방법**

**(57) 요약**

오믹스 환자 데이터는, 환자 및 질환 특이적인 돌연변이들을 확인하도록 종양 및 매칭된 정상 조직의 서열들 또는 다른 목적물들을 사용하고, 변이된 유전자들의 발현 수준들 확인하는 전사체학 데이터 및 병든 조직의 특이적인 경로 특징들을 확인하도록 획득된 오믹스 데이터를 근거로 하는 경로 분석을 사용하여, 분석된다. 가장 명확하게, 많은 서로 다른 종양들은 공유된 경로 특징들을 가지고, 따라서 종양의 경로 특징의 확인은 종양 분석이 해부학적 종양 유형만을 기초로 할 때 보통 고려되지 않는 효과적인 치료 조건들을 지시할 수 있다.

**대표도**



(72) 발명자

**벤즈, 스테펜 카를**

미국, 캘리포니아 95060, 산타 크루즈, 쿠퍼 스트리트 101

**라비자드 샤루즈**

미국 캘리포니아 컬버 시티 제퍼슨 블러바드 9920  
(우:90232)

**순-시용, 패트릭**

미국 캘리포니아 컬버 시티 제퍼슨 블러바드 9920  
(우:90232)

## 명세서

### 청구범위

#### 청구항 1

상이한 해부학적 종양들로 진단되는 다수의 환자들의 다수의 종양 세포들을 분자적 서명에 기초하여 분류하는 방법으로서,

각각의 데이터 집합이 종양 및 매칭된 정상 세포들로부터 얻은 게놈학 정보를 나타내는 다수의 데이터 집합들을, 다수의 환자들로부터 분석 엔진에 의해 수신하는 단계;

다수의 환자들을 위한 전사체학 정보를 분석 엔진에 의해 수신하는 단계;

적어도 두 명의 환자들의 종양 세포들 중에서 공유된 경로 특징들을 게놈학 정보 및 전사체학 정보를 사용하여 분석 엔진에 의해 확인하는 단계;

상기 공유된 경로 특징들을 근거로 하여, 해부학적 종양 유형과는 독립적으로 할당되는 분자적 서명을 다수의 종양 세포들에 할당하도록 분석 엔진을 사용하는 단계;

분석 엔진에 의해, 종양 세포들을 분자적 서명에 기초하여 다수의 군집들 (clusters)로 분류하되, 군집 중 적어도 하나는 적어도 하나의 분자적 서명을 공유하며 해부학적으로 관련되지 않은 구별된 종양 세포들을 포함하는 단계; 및

다수의 군집들을 사용하여 환자 기록을 생성하거나 갱신하는 단계:

를 포함하는, 방법.

#### 청구항 2

제 1항에 있어서,

상기 다수의 데이터 집합들은 BAMBAM 포맷, SAMBAM 포맷, FASTQ 포맷 또는 FASTA 포맷으로 있는, 방법.

#### 청구항 3

제 1항에 있어서,

상기 다수의 데이터 집합들은 BAMBAM 디프 목적물들인, 방법.

#### 청구항 4

제 1항에 있어서,

상기 다수의 데이터 집합들은 돌연변이 정보, 사본수 정보, 삽입 정보, 결실 정보, 방향성 정보 및/또는 중단점 정보를 포함하는, 방법.

#### 청구항 5

제 1항에 있어서,

상기 다수의 환자들 중 적어도 세 명은 서로 다른 종양들로 진단되는, 방법.

#### 청구항 6

제 1항에 있어서,

상기 다수의 환자들 중 적어도 다섯 명은 서로 다른 종양들로 진단되는, 방법.

#### 청구항 7

제 1항에 있어서,

상기 게놈학 정보는 전체 게놈 서열결정 정보인, 방법.

**청구항 8**

제 1항에 있어서,

상기 게놈학 정보는 엑솜 서열결정 정보인, 방법.

**청구항 9**

제 1항에 있어서,

상기 전사체학 정보는 전사 수준에 관한 정보를 포함하는, 방법.

**청구항 10**

제 1항에 있어서,

상기 전사체학 정보는 RNA 서열에 관한 정보를 포함하는, 방법.

**청구항 11**

제 1항에 있어서,

상기 전사체학 정보는 종양 세포들로부터 얻은 게놈학 정보에 있는 모든 엑솜들의 적어도 50%를 포괄하는, 방법.

**청구항 12**

제 1항에 있어서,

상기 전사체학 정보는 종양 세포들로부터 얻은 게놈학 정보에 있는 모든 엑솜들의 적어도 80%를 포괄하는, 방법.

**청구항 13**

제 1항에 있어서,

상기 공유된 경로 특징들은 기본 구성적으로 활성화된 경로, 기능적으로 손상된 경로 및 조절이 곤란한 경로로 구성되는 그룹으로부터 선택되는, 방법.

**청구항 14**

제 1항에 있어서,

상기 공유된 경로 특징들은 경로에서 변이된 비-기능적인 단백질, 변이된 조절이 곤란한 단백질, 과다발현된 단백질 또는 과소발현된 단백질을 특징으로 하는, 방법.

**청구항 15**

제 1항에 있어서,

상기 전사체학 정보는 변이된 유전자에 의해 인코딩된 단백질의 감소된 기능 또는 기능의 부재를 추론하도록 확인하는 단계에 사용되는, 방법.

**청구항 16**

제 1항에 있어서,

상기 확인하는 단계는 PARADIGM을 사용하여 수행되는, 방법.

**청구항 17**

제 1항에 있어서,

상기 분자적 서명은 하나 이상의 경로 요소들에 대한 정보를 포함하는, 방법.

**청구항 18**

제 17항에 있어서,

상기 분자적 서명의 정보는 약물 확인 및 하나 이상의 경로 요소들과 상호작용의 유형을 포함하는, 방법.

**청구항 19**

제 1항에 있어서,

상기 환자 기록은 종양 세포들의 군집을 근거로 하는 치료 추천을 포함하는, 방법.

**청구항 20**

제 19항에 있어서,

첫 번째 종양을 가진 첫 번째 환자를 위한 치료 추천은 두 번째 구별된 종양을 가진 두 번째 환자에 대하여 알려진 치료 양식에 근거하되, 첫 번째 환자의 종양세포 및 두 번째 환자의 종양세포는 동일한 군집으로 분류되는, 방법.

**발명의 설명**

**기술 분야**

[0001] 본 발명의 분야는 특히 다수의 환자 및 종양 시료들로부터 얻은 많은 종양 및 생식계열 엑솜들이 교차하는 분자적 프로파일들의 분석에 관한 것인 바, 전산적인 오믹스 (omics)이다.

**배경 기술**

[0002] 배경의 설명은 본 발명을 이해하는 데 유용할 수 있는 정보를 포함한다. 이것은 본 명세서에서 제공된 임의의 정보가 선행기술이거나 현재까지 청구된 발명에 적절한 점 또는 특정하게 또는 함축적으로 참조된 임의의 출판물이 선행기술인 점을 인정하는 것은 아니다.

[0003] 임상 분야가 분자적 식견을 치료법들로 유도하는 수단으로서 제한된 수의 돌연변이들에게로 표적화되는 게놈적 검정법들에 친숙한 한편, 더욱 해석적이고 추측이 아닌 확률적 분석을 전달하는 힘이 치료제들에 관한 전통적인 조직-대-조직 해부학적 과제, 또는 수백 가지의 DNA 돌연변이들이 암의 유도인자들이라는 사전 가정들에 치우치지 않는 치료 결정들을 안내하는 데 전적으로 필요하다. 따라서, 대부분의 의사들은 오늘날 보조를 맞추기가 점차 어려워지는 신속하게 발전하는 과학의 홍수로 도전을 받는다. 이러한 개인화 의학의 시대에, 종양의 성장을 추진하는 특이적인 단백질 표적들에게로 표적화되는 800가지 약물들이 개발 중이다. 이러한 지식적 과부하는 암과 같이 복잡한 인생-위협 질환들의 결정에서 유의한 결론들을 가질 수 있다.

[0004] 오늘날 암종학자들이 변경된 단백질들에게로 표적화되는 약물들의 치료 선택을 안내하는 데 가장 널리 사용되는 접근법은 500가지 이하의 "작용가능한" 유전자들의 패널들을 배치한 종양 시료들에서 유전자 DNA 돌연변이들을 확인하는 것이다. 이러한 작용가능한 유전자들은 전형적으로 다양한 암들의 대규모 연구들로부터 확인된다 (예로, Nature Genetics 45, 1127-1133, 2013 참조). 본 명세서에서 모든 출판물들 및 출원들은 각각의 개별적 출판물 또는 특허 출원이 특정하게 및 개별적으로 참고문헌으로 통합되도록 지시되는 바와 동일한 정도로 참고문헌으로 통합된다. 통합된 참고문헌에서 용어의 정의 또는 용도가 본 명세서에서 제공된 해당 용어의 정의와 부합하지 않거나 상반되는 경우에, 본 명세서에서 제공된 용어의 정의가 적용되고 참고문헌에서의 해당 용어의 정의는 적용되지 않는다.

[0005] 불행하게도, 치료 결정들을 내리도록 종양 시료들의 유전형 결정에 대한 최근의 의존성은 변이된 DNA의 확인이 일상적으로 선택될 치료법에 의해 표적화되는 단백질 경로들에서 변경으로 하류 ("DNA로부터 단백질 발현에 이르기까지") 해독된다는 가정을 주로 근거로 하며, 이들 확인된 DNA 돌연변이들은 이에 따라 임상적으로 작동가능한 것으로 주장된다. 그러나, 종양 게놈들에서 유전적 돌연변이들의 베타적인 분석이 변이된 유전자들이 결국 전사되는지 여부, 게놈에서 변화들이 변이체들 또는 질환-유도인자들인지 여부, 및/또는 이러한 돌연변이들의 기능적 맥락이 무엇인지, 그리고 이러한 돌연변이에 의해 침범된 세포에 보상적 기작들이 존재하는지 여부는 고

려하지 못하고 있다.

[0006] 따라서, 상기 단점들을 무시하는 선별된 돌연변이들의 분석은 아마도 다양한 위양성, 위음성 및 다음 순서로 환자의 치료를 잘못 안내할 수 있는 부적절한 결과들을 유도할 것이다. 따라서, 분자적 프로파일들의 해석적 분석을 위한 개선된 시스템들 및 방법들을 위한 필요성이 여전히 존재한다.

### 발명의 내용

#### 해결하려는 과제

[0007] 본 출원은 2014년 5월 30일자로 제출된, 일련번호 제 62/005766호를 가지는 미국 가특허출원에 관한 우선권의 이익을 주장하고, 이는 본 명세서에서 참고문헌으로 통합되어 있다.

[0008] 본 발명의 주제 문제는 공유된 경로 특징들이 다양한 구별된 종양 시료들로부터 획득되는 오믹스 분석의 시스템들 및 방법들에 관한 것이다. 더욱 바람직하게, 오믹스 분석은 환자 및 종양 특이적 변화들을 확인하도록 종양 및 매칭된 정상 조직의 분석을 포함하고, 이는 또한 전사체학 데이터를 사용하여 정련된다. 이러한 분석을 근거로 하여, 다음으로 전형적으로는 해부학적 종양 유형과는 독립적이지만 공유된 경로 특징들의 분자적 서명 특징을 고려하는 치료 추천이 준비된다.

[0009] 본 발명의 주제 문제의 한 가지 관점에서, 본 발명자들은 종양 세포를 위한 분자적 서명을 확인하는 방법으로서, 각각의 다수의 환자들로부터 다수의 데이터 집합들을 수신하는 데 분석 엔진을 사용하는 단계를 포함하고, 다수의 환자들 중 적어도 두 명 (또는 적어도 세 명 또는 적어도 다섯 명)이 서로 다른 종양들로 진단되며, 각각의 데이터 집합은 종양 및 매칭된 정상 세포들로부터 얻은 게놈학 정보를 나타내는, 방법을 고려한다. 또 다른 단계에서, 분석 엔진은 적어도 두 명의 환자들에 대한 전사체학 정보를 수신하고, 또한 또 다른 단계에서 분석 엔진은 게놈학 정보 및 전사체학 정보를 사용하여 적어도 두 명의 환자들의 종양 세포들 중에서 공유된 경로 특징들을 확인한다.

#### 과제의 해결 수단

[0010] 본 발명의 주제 문제에 제한되지 않으면서, 일반적으로 데이터 집합들은 BAMBAM 포맷, SAMBAM 포맷, FASTQ 포맷 또는 FASTA 포맷으로 있는 것이 고려되고, 전형적으로 데이터 집합들은 BAMBAM 디프 목적물들인 것이 바람직하다. 따라서, 더 고려된 관점들에서, 데이터 집합들은 바람직하게 돌연변이 정보, 사본수 정보, 삽입 정보, 결실 정보, 방향성 정보 및/또는 중단점 정보를 포함할 것이다.

[0011] 게놈학 정보의 측면에서, 이러한 정보는 전체 게놈 서열결정 정보 또는 엑솜 서열결정 정보일 수 있고, 전사체학 정보는 전사 수준에 관한 정보 및/또는 서열 정보를 포함하는 것으로 고려된다. 가장 전형적으로, 전사체학 정보는 종양 세포들로부터 얻은 게놈학 정보에 있는 모든 엑솜들의 적어도 50% (또는 적어도 80%)를 포괄할 것이다. 또한, 전사체학 정보는 변이된 유전자에 의해 인코딩된 단백질의 감소된 기능 또는 기능의 부재를 추론하도록 확인하는 단계에 사용되는 것으로 고려된다.

[0012] 따라서, 본 발명자들은 공유된 경로 특징들이 기본적으로 활성화된 경로, 기능적으로 손상된 경로 및 조절이 곤란한 경로를 포함할 것이고/거나, 공유된 경로 특징들이 경로에서 변이된 비-기능적인 단백질, 변이된 조절이 곤란한 단백질, 과다발현된 단백질 또는 과소발현된 단백질을 특징으로 할 수 있다. 보다 더 바람직한 관점들에서, 확인하는 단계는 PARADIGM 또는 다른 경로-중심적 분석 방법을 사용하여 수행된다.

[0013] 추가적으로, 분자적 서명은 하나 이상의 경로 요소들에 대한 정보, 및 특히 약물 확인 및 하나 이상의 경로 요소들과의 상호작용의 유형을 포함하는 것으로 고려된다. 따라서, 환자 기록은 또한 종양 세포들의 분자적 서명을 근거로 하는 치료 추천을 포함할 수 있는 것으로 이해되어야 한다 (예로, 첫 번째 종양을 가진 첫 번째 환자를 위한 치료 추천은 두 번째 구별된 종양을 가진 두 번째 환자와 공유된 경로 특징들을 근거로 할 수 있다).

#### 도면의 간단한 설명

[0014] 본 발명의 주제 문제의 다양한 목적들, 특징들, 관점들 및 장점들은 바람직한 구현예들의 다음의 상세한 설명으로부터, 동일한 숫자들이 동일한 구성성분들을 나타내는 첨부된 도면들과 함께 더 자명해질 것이다.

도 1은 선택된 종양들을 위한 "작용가능한 유전자들"의 빈도 분포를 도시하는 그래프를 나타낸 것이다.

도 2는 선택된 종양들을 위한 변이된 게놈 DNA의 RNA 발현 수준을 상관시키는 그래프를 나타낸 것이다.

도 3은 선택된 종양들에서 선택된 종양원 유전자들을 위한 기본 구성성분을 도시하는 대표적인 그래프를 나타낸 것이다.

도 4는 게놈 재배열들의 함수로서 생존 시간들을 도시하는 대표적인 그래프를 나타낸 것이다.

도 5는 선택된 종양들을 위한 대표적인 중단점 분석을 도시하는 차트를 나타낸 것이다.

도 6은 과다- 또는 과소-활성화된 중심 경로들을 근거로 하는 경로 활성 군집들을 도시하는 그래프를 나타낸 것이다.

도 7은 다양한 종양 유형들을 통해 군집화된 경로 활성들을 도시하는 그래프를 나타낸 것이다.

도 8은 다양한 종양 유형들을 위한 돌연변이 분포를 도시하는 대표적인 그래프를 나타낸 것이다.

**발명을 실시하기 위한 구체적인 내용**

[0015] 다음의 상세한 설명은 본 발명을 이해하는 데 유용할 수 있는 정보를 포함한다. 이것은 본 명세서에서 제공된 임의의 정보가 선행기술이거나 현재까지 청구된 발명에 적절한 점 또는 특정하게 또는 함축적으로 참조된 임의의 출판물이 선행기술인 점을 인정하는 것은 아니다.

[0016] 본 발명의 주제 문제는 다양한 종양들의 개선된 오믹스 분석을 위한 장치, 시스템들 및 방법들을 제공한다. 보다 상세하게, 본 발명자들은 오믹스 데이터 분석이 게놈에서 먼저 환자 및 종양에 적절한 변화들을, 전형적으로 종양 및 매칭된 정상 시료들의 비교에 의해 확인함으로써 유의하게 개선될 수 있는 점을 발견하였다. 이러한 차이점들이 규명되는 경우라면, 동일한 환자의 추가적인 전사체학 데이터는 변화된 서열들이 종양에서 발견되는지 여부를 확인하는 데 사용된다. 이와 같이 획득된 환자 데이터는 다음으로 종양의 경로 특징들, 상세하게 다양한 기타 유형들의 종양들과 공유된 경로 특징들을 확인하도록 경로 분석에 착수된다. 바로 이해되어야 할 바와 같이, 공유된 경로 특징들은 달리 확인되지 않았던 해부학적으로 관련되지 않는 종양들로부터 나온 하나 이상의 치료 양식들을 사용하는 치료를 알려주도록 채용될 수 있다. 서로 다른 견지로부터 바라보면, 서로 다른 종양 유형들은 해부학적 종양 유형과는 상관없이 경로 특징들을 공유하고, 각각의 분자적 서명들을 가진 공유된 경로 특징들의 지식은 특정한 종양 유형의 경우에 평가되지 않았던 약물 치료 전략들을 확인할 수 있다.

[0017] 결론적으로, 본 발명의 주제 문제의 한 가지 관점에서, 본 발명자들은 종양 세포를 위한 분자적 서명, 특히 세포 신호전달 경로의 분자적 서명을 확인하는 방법을 고려한다. 가장 전형적으로, 확인 및 분석은 전적으로 통합되고, 클라우드-기초하고, 슈퍼컴퓨터-추진되는, 게놈학 및 전사체학 분석 엔진을 사용하여 수행된다. 컴퓨터로 안내되는 임의의 언어는, 서버들, 인터페이스들, 시스템들, 데이터베이스들, 에이전트들, 피어들, 엔진들, 컨트롤러들 또는 임의의 유형들의 개별적으로 또는 종합적으로 작동하는 전산처리 장치들을 포함하는 전산처리 장치들의 임의의 적합한 조합을 포함하도록 해석되는 점이 주목되어야 한다. 전산처리 장치들은 접근가능한 비-일과성 컴퓨터 해독가능한 저장 매체 (예로, 하드 드라이브, 고형상 드라이브, RAM, 플래쉬, ROM 등) 상에 저장된 소프트웨어 지침들을 실행하도록 구조화된 프로세서를 포함하는 것으로 이해되어야 한다. 소프트웨어 지침들은 바람직하게 개시된 기기의 측면에서 하기에 논의된 바와 같이 역할들, 책무들 또는 기타 기능성을 제공하도록 전산처리 장치를 구조화한다. 특히 바람직한 구현예들에서, 다양한 서버들, 시스템들, 데이터베이스들 또는 인터페이스들은 가능하게 HTTP, HTTPS, AES, 공공-개인 핵심 교환기술, 웹 서비스 API들, 기지의 금융적 트랜스작용 프로토콜들 또는 기타 전자적 정보 교환 방법들을 기초로 하는, 표준화된 프로토콜들 또는 알고리즘들을 사용하여 데이터를 교환한다. 데이터 교환들은 바람직하게 패킷-스위치 전산망, 인터넷, LAN, WAN, VPN 또는 다른 유형의 패킷 스위치 전산망을 거쳐서 시행된다.

[0018] 특히 바람직한 방법들에서, 분석 엔진은 각각의 다수의 환자들로부터 얻은 다수의 데이터 집합들로서, 다수의 환자들 중 적어도 두 명은 서로 다른 종양들로 진단되고, 각각의 데이터 집합은 종양 및 매칭된 정상 세포들로부터 얻은 게놈학 정보를 나타내는, 다수의 데이터 집합들을 수신한다. 추가 단계에서, 분석 엔진은 적어도 두 명의 환자들에 대한 전사체학 정보를 수신하고, 적어도 두 명의 환자들의 종양 세포들 중에서 게놈학 정보 및 전사체학 정보를 사용하여 공유된 경로 특징들을 확인한다 (당연하게, 공유된 경로 특징들이 단일한 환자 시료에 대해서만 역시 확인될 수 있는 한편, 다른 종양들의 경로 특징들이 경로 데이터베이스로부터 획득될 수 있는 점은 주목되어야 한다). 또한 또 다른 단계에서, 분석 엔진은 다음으로 공유된 경로 특징들을 근거로 하여 종양 세포들에게로 분자적 서명을 할당하는 데 사용되고, 여기에서 분자적 서명은 해부학적 종양 유형과는 독립적으로 (예로, 불가지론적 방식으로) 할당된다. 또한 추가 단계에서, 환자 기록은 분자적 서명을 사용하여 생성되거나 갱신될 수 있다.

[0019] 다수의 환자들로부터 얻은 데이터 집합들의 측면에서, 데이터 집합들의 유형은 상당히 다양해질 수 있고, 수많은 유형들의 데이터 집합들이 본 명세서에서의 용도에 적합한 것으로 판단된다. 따라서, 데이터 집합들은 미프로세싱된 또는 프로세싱된 데이터 집합들을 포함할 수 있고, 대표적인 데이터 집합들은 BAMBAM 포맷, SAMBAM 포맷, FASTQ 포맷 또는 FASTA 포맷을 가지는 것들을 포함한다. 그러나, 특히 데이터 집합들은 BAMBAM 포맷으로 또는 BAMBAM 디프 목적물들로서 제공되는 것이 바람직하다 (예로, 미국 특허 제 US2012/0059670 A1호 및 제 US2012/0066001 A1호 참조). 따라서, 그리고 또 다른 견지로부터 바라보면, 데이터 집합들이 환자 및 종양 특이적 정보를 획득하도록 동일한 환자의 종양 및 매칭된 정상 시료를 반영하는 점이 주목되어야 한다. 이에 따라, 종양을 유발하지 않은 유전적 생식계열 변경들 (예로, 침묵 돌연변이, SNP 등)이 배제될 수 있다. 당연히, 종양 시료가 초기 종양으로부터, 치료의 개시 시점의 종양으로부터, 재발성 종양 또는 전이 부위 등으로부터 나올 수 있는 점을 인식하여야 한다. 대부분의 경우들에서, 환자의 매칭된 정상 시료는 혈액, 또는 종양과 동일한 조직 유형으로부터 얻은 병들지 않은 조직일 수 있다.

[0020] 데이터 집합들이 데이터 집합을 생성하는 장치 (예로, 서열결정기, qPCR 기계 등)로부터 나오거나 데이터 집합들을 저장하는 데이터베이스로부터 제공될 수 있는 점도 역시 주목되어야 한다. 예를 들면, 적합한 데이터 집합들은 BAM 서버 (예로, 미국 특허 제 US2012/0059670 A1호 및 제 US2012/0066001 A1호에 기술된 바와 같음) 및/또는 경로 분석 엔진 (예로, 국제특허출원 제 W02011/139345 A2호 및 제 W02013/062505 A1호에 기술된 바와 같음)으로부터 유래될 수 있다. 이러한 점은 상세하게 종양 및 매칭된 정상 시료로부터 얻은 데이터 집합들이 환자로부터 유래되지 않는 곳에서 사실이다. 이에 따라, 적어도 일정의 데이터 집합들은 독립적으로 저장되고 제공될 수 있고, 분석은 환자의 종양 및 매칭된 정상 시료로부터 얻은 데이터 집합들 및 환자로부터 유래되지 않은 이전에 저장된 종양 및 매칭된 정상 시료를 사용하여 새로이 획득된 환자 시료 상에서 (예로, 환자 조직 시료를 획득하고 한 주 이내에) 수행될 수 있다.

[0021] 추가로 데이터 집합들의 측면에서, 모든 종양들로부터 얻은 데이터 집합들이 추가 전환 및/또는 프로세싱이 없이 준비된 비교를 허용하는 포맷인 점이 주목된다. 이에 따라, 데이터 집합들은 바람직하게 종양 및 환자에 특이적인 돌연변이 정보, 메틸화 상태 정보, 사본수 정보, 삽입/결실 정보, 방향성 정보 및/또는 중단점 정보를 포함할 것이다. 또한 추가로 데이터 집합은 적어도 전체 게놈의 일부, 및 가장 전형적으로 게놈 전부를 나타내는 것으로 고려된다. 따라서, 데이터 집합들은 바람직하게 전체 게놈 (또는 전체 게놈의 적어도 50%, 또는 적어도 70%, 또는 적어도 90%)을 포괄하는 게놈 전부의 서열결정으로부터 준비된다. 대안적으로, 엑솜 서열결정도 역시 고려되고, 대부분의 경우들에서 전체 엑솜의 적어도 50%, 보다 전형적으로 적어도 70%, 가장 전형적으로 적어도 90%가 서열결정되는 것이 고려된다.

[0022] 또한, 데이터 집합들의 기원의 측면에서 수많은 비-환자 종양 데이터가 사용되는 것으로 이해되어야 한다. 따라서, 환자 데이터 집합이 아닌 데이터 집합들은 공유된 경로 특징들을 확인하도록 적어도 두 가지의 서로 다른 종양들, 더욱 바람직하게 적어도 세 가지, 또는 적어도 다섯 가지의 서로 다른 종양 유형들로부터 유래할 것으로 고려된다. 서로 다른 종양 유형들의 데이터 집합들은 이러한 시료들이 입수가능한 (예로, 병원, 임상 시험, 역학적 연구 등으로부터) 서로 다른 환자 시료들로부터 획득될 수 있고/거나 이전에 입수된 분석들 또는 데이터로부터 제공될 수 있다. 예를 들면, TCGA는 본 명세서에서의 용도에 적합한 데이터 집합들을 준비하는 데 유용한 잘-특성화된 오픈믹스 정보의 좋은 시료를 제공하고, 하기 표 1은 본 분석에 사용된 데이터를 예시적으로 도시하고 있다.



표 1

조직	아형	N	연령	성별		종양 등급							평균 생존 (개월)
				M	F	G1	G2	G3	G4	GX	GB	?	
유방 엽	ER-	10	59.3 (58.9)	0	10	0	0	C	C	0	0	10	20.04
	ER+	148	62.0 (61.9)	0	148	0	0	C	C	0	0	148	147.48
폐 편평세포		366	68.5 (67.8)	267	99	0	0	C	C	0	0	366	46.78
직장		96	67.7 (66.0)	58	48	0	0	C	C	0	0	96	51.96
유방 관	ER-	225	55.3 (56.7)	0	225	0	0	C	C	0	0	225	100.70
	ER+	310	59.5 (59.2)	9	507	0	0	C	C	0	0	310	113.82
교모세포종		354	60.3 (60.1)	222	132	0	0	C	C	0	0	354	13.94
위장	MSI	88	69.0 (68.6)	46	42	3	30	55	C	0	0	0	26.47
	MSS	160	67.0 (65.9)	105	55	3	54	99	C	4	0	0	72.23
AML		4	50.1 (50.2)	3	1	0	0	C	C	0	0	4	47.05
저등급 교아종		138	41.1 (42.8)	72	66	0	66	70	C	0	0	0	78.21
두경부	HPV-	246	62.4 (62.7)	160	70	26	159	55	C	6	0	0	47.97
	HPV+	56	59.4 (59.6)	46	8	1	31	19	2	3	0	0	52.31
방광		110	68.3 (67.1)	86	32	0	0	C	C	0	0	110	19.50
자궁		299	63.8 (64.0)	0	299	0	0	C	C	0	0	299	NA
전립선		178	61.3 (60.8)	178	0	0	0	C	C	0	0	178	NA
폐 샘		369	68.9 (68.7)	171	198	0	0	C	C	0	0	369	40.60
결장	MSS	144	68.2 (68.6)	82	62	0	0	C	C	0	0	144	NA
	MSI	76	73.1 (69.0)	38	48	0	0	C	C	0	0	76	NA
갑상샘		419	46.9 (47.0)	104	315	0	0	C	C	0	0	419	NA
신장 투명세포		325	60.4 (60.4)	209	116	6	142	128	47	2	0	0	90.48
신장 비염색		50	48.3 (50.5)	28	21	0	0	C	C	0	0	50	NA
난소		336	59.5 (60.4)	0	336	3	37	285	C	8	1	2	43.86
흑색종		301	56.8 (57.3)	192	109	0	0	C	C	0	0	301	96.40
췌장		4	?	?	?	0	0	C	C	0	0	4	NA
전체		5052											

[0023]

[0024]

TCGA 데이터를 참조하여, 서로 다른 종양 유형들은 복수의 유전자들에서 복수의 돌연변이들을 가지는 것으로 추가로 관찰되었다. 이와 같이, 개별적인 약물화 가능한 표적들의 단순한 표적화가 대부분의 상황들에서 생존가능한 조건인 것은 자명하다. 따라서, 도 1은 통상적인 단일한 분자적 진단학의 단점을 예시적으로 도시하고, 다양한 종양 유형들이 잠재적으로 작용가능한 유전자들의 각각의 수적인 분포로 보여진다. 도 1로부터 바로 자명해질 바와 같이, 거의 모든 종양들에는 단지 단일한 돌연변이가 아닌 다수의 작용가능한 유전자들이 존재한다. 따라서, 종양의 분석 및 치료는 하나 이상의 변화된 유전자의 고려를 요구하는 것으로 이해되어야 한다. 또한, 이에 따라 변이된 유전자들 모두가 발견되는 것은 아니고, 이로써 도 2에 예시적으로 도시된 바와 같이 작용가능한 또는 약물화 가능한 단백질 표적들을 유도하거나 유도하지 않을 수 있는 점이 이전에는 이해되지 못하였다.

[0025]

도 2로부터 얻을 수 있는 바와 같이, 소정의 종양들에서 선별된 돌연변이들은 발견되지 않거나 약하게만 발견되었다 (예로, RNA 내로 전사됨, 아래 박스 참조). 결론적으로, 이러한 변이체 단백질들을 표적화하는 약제학적 개입 (예로, 교모세포종에서 BRAF V600 표적화)이 유의한 방식으로 종양에 영향을 미치는 것으로 기대되지 않는다. 역으로, 소정의 기타 변이된 단백질들은 그들의 매우 높은 발현율로 인해 (예로, 흑색종에서 BRAF V600 표적화에 의해) 매력적인 표적을 제공할 것이다. 따라서, 동일한 변이된 단백질이 일정 암들 또는 환자들에서 적합한 표적이고 다른 곳에서는 전적으로 부적합한 표적일 수 있는 것으로 판단되어야 한다. 다른 견지로부터 바라보면, 전사체학 데이터의 고려가 없는 게놈학 정보는 치료 결정들을 안내하는 데 필요한 세부사항이 부족할 것이다.

[0026]

상세하게 바람직한 관점들에서, 전사 정보는 종양 세포들로부터 얻은 게놈학 정보에서 모든 엑솜의 적어도 50%, 또는 적어도 70%, 또는 적어도 80%, 또는 적어도 90%를 포괄하도록 획득된다. 따라서, 종양 세포 또는 조직의 전사체들도 역시 그들의 정량에 대해 (또한 선택적으로 RNA 편집화 및/또는 RNA 스프라이싱을 확인하는 서열 정보에 대해서도) 분석될 수 있는 것으로 고려된다. 이러한 분석은 전형적으로 심사 중인 2015년 5월 15일에 제출된 미국 가특허출원 일련번호 제 62/162530호에 더 자세히 기술된 바와 같이 정의되는 유제인 한계 수치들을 포함할 수 있다.

[0027]

전사체학 데이터의 부족한 고려에 추가하여, 세포 신호전달 망 내의 돌연변이의 기능적 효과는 본 명세서에서 대부분의 기지의 시스템들 및 방법들에서, 특히 다수의 돌연변이들이 종양과 연관된 다수의 유전자들에 존재하

는 점에서 이해되지 않았다. 이러한 단점을 극복하도록, 본 발명자들은 세포 신호전달 경로들의 분석에서 환자 및 종양 특이적 돌연변이 정보 및 연관된 발현 수준들을 사용하고, 이로써 경로 기능이 억제되는 것에서 경로 유용성 및 보상에 관한 정보를 획득한다. 따라서, 전사체학 정보가 바람직하게 변이된 유전자에 의해 인코딩되는 단백질의 감소된 기능 또는 기능의 부재를, 특정한 경로에 미치는 해당 영향력으로 추론하는 데 사용되는 점이 주목된다.

[0028] 다양한 경로 분석적 도구들이 해당 기술분야에서 알려져 있는 한편, 본 발명자들은 특히 경로들이 확률적 경로 모델로서 표현되는 역학적 경로 맵들의 사용을 고려한다. 예를 들면, 경로 분석들은 데이터 집합들 및 전사체학 정보를 이로써 특이적 종양의 특정한 경로들 유용성에 도달하도록 사용하는, 국제특허출원 제 WO 2011/139345호, 제 WO 20 13/062505호, 제 WO 2014/059036호 또는 제 WO 2014/193982호에 기술된 바와 같이 PARADIGM을 사용하여 수행될 수 있다. 바로 이해될 바와 같이, 채용된 바와 같이 구별되는 종양들을 가지는 다수의 환자들로부터 얻은 다수의 데이터 집합들에서, 분석 엔진은 종양 세포들의 분자적 서명으로 각각의 종양 특정한 경로 특징들을 확인할 수 있을 것이다. 예를 들면, 분석 엔진은 다수의 종양 유형들 중에서 공유된 경로 특징들을 확인할 수 있고, 이러한 공유된 특징들로는 기본 구성적으로 활성화된 경로, 기능적으로 손상된 경로 및 조절이 곤란한 경로를 포함할 수 있다. 이러한 공유된 경로는 다양한 인자들로 인해 특성화될 수 있고, 특정한 경로 특징을 가져오는 대표적인 인자들은 경로 등에서 변이된 비-기능적인 단백질, 변이된 조절이 곤란한 단백질, 과다발현된 단백질 또는 과소발현된 단백질을 포함한다. 당연히, 적어도 일정 경로 특징들이 미리 결정되어 데이터베이스에 저장될 수 있거나, 적어도 일정 경로 특징들이 새로이 (*de novo*) 역시 결정될 수 있는 점이 주목되어야 한다. 따라서, 새로운 환자 데이터는 데이터베이스로부터 미리 획득된 데이터와 대비될 수 있는 것으로 인식되어야 한다.

[0029] 다수의 환자들의 다수의 종양 유형들을 위한 통합된 계층학, 전사체학 및 경로 분석의 다른 유익들 중에서, 다양한 연속적인 분석들은 현재 소정의 분자적 사건들을 달리 관찰가능하지 않은 카테고리들 내로 그룹화하거나 분류할 수 있는 것으로 이해되어야 한다. 예를 들면, 도 3에 도시된 바와 같이 서로 다른 종양들로부터 얻은 다양한 발현된 변이된 종양원 유전자들의 기본적인 구성성분 분석은 이로써 다수의 특이적 돌연변이들을 다수의 서로 다른 종양들과 연관시키도록 수행될 수 있다. 마찬가지로, 서로 다른 종양들을 통한 중단점 분석은 도 4에 예시적으로 도시된 바와 같이 예후적 결과와 연관될 수 있거나, 중단점 빈도 및 분포는 도 5에 예시적으로 도시된 바와 같이 서로 다른 종양들과 연관될 수 있다.

[0030] 가장 명확하게, 도 6에 예시적으로 도시된 바와 같이 계층학 및 전사체학 정보를 근거로 하는 경로 분석은 다양한 서로 다른 종양들과 공통적인 소정의 공유된 분자적 서명을 확인하도록 작용할 수 있다. 따라서, 종양은 특이적인 공유된 경로 특징들에 의해 특성화되는 종양의 부류에 속하는 것으로 분류될 수 있는 점이 인식되어야 한다. 도 6을 추가로 참조하여, 계층학 및 전사체학 정보를 다함께 가진 표 1의 종양들은 해부학적 위치와 독립적인 여섯 가지의 구별된 부류들 내로 분류되었다. 여기에서, 서로 다른 종양들을 위한 공통적인 부류들은 선택된 신호전달 경로들의 활성화 또는 억제에 의해 정의되었고 (예로, myc 전사의 과다-활성화 및 NOTCH 신호전달의 억제), 이는 전적으로 해부학적 종양 유형 (췌장 종양, 유방 관상 종양 등으로서 분류됨)을 근거로 하는 부류와는 독립적이다.

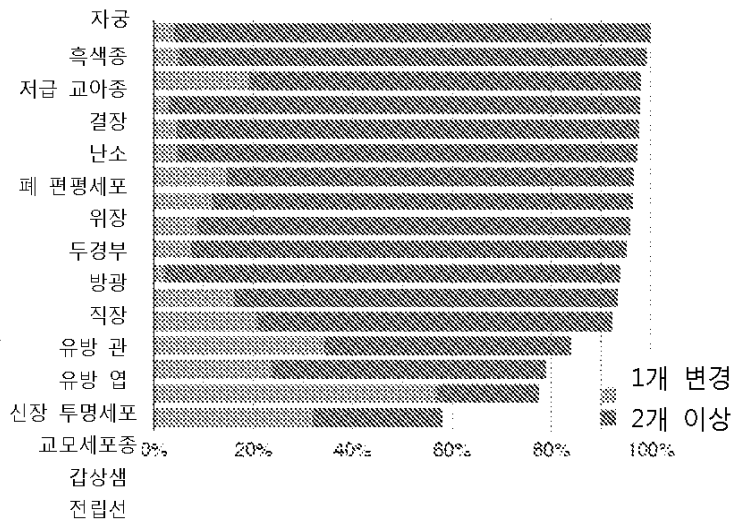
[0031] 도 7은 예시적으로 종양 분류가 도 6의 군집들로서 표현되는 도 6의 연구들의 다른 관점을 도시하고 있다. 도면에서, 전적으로 관련되지 않은 종양들 (예로, 자궁, 직장, 폐, 췌)이 특이적인 분자적 서명들을 가지는 특이적인 신호전달 경로들 특징들에 따라 분류될 수 있는 점은 바로 자명할 것이다. 예를 들면, 분자적 서명은 경로 내의 하나 이상의 경로 요소들에 대한 정보 (예로, Ras, Raf, MEK, Myc)를 포함할 수 있다. 이와 같이, 종양이 하나 이상의 공통적인 분자적 서명들을 가진 공통적인 경로 특징들을 또 다른 관련되지 않은 종양과 공유하는 곳에서, 사실상 종양은 관련되지 않은 종양에 대해 알려진 치료 양식들을 사용하여 치료가 가능할 수 있다. 가장 전형적으로, 분자적 서명 정보는 약물 확인 (예로, Ras가 변이되고 과다발현되는 곳에서, 약물 정보는 적합한 Ras 저해제들을 포함할 수 있음) 및/또는 하나 이상의 경로 요소들과 상호작용의 유형 (예로, Hec1이 변이되고 과다 활성을 가진 곳에서, 약물 정보는 적합한 Hec1/Nek 저해제들을 포함할 수 있음)을 포함할 수 있다. 따라서, 또 다른 견지로부터 바라보면, 환자 종양은 해당 부류가 경로 내에서 공통적인 경로 특징들/분자적 서명들을 공유하는 관련되지 않은 구별된 구성원들 (종양들)을 가지는 것으로서 정의되는 특이적인 부류에 속하는 것으로서 특성화될 수 있다. 이와 같이 확립된 분류를 근거로 하여, 치료 조건들이 기지의 또는 관련되지 않은 구별된 구성원들에게 사용가능한 치료 조건들을 기초로 하여 선택될 수 있다. 치료 조건은 특정한 경로의 변이된 요소들을 표적할 수 있지만, 치료 조건은 변이된 효소가 배치되는 경로에서 결함을 보상하는 또 다른 경로의 변이되지 않은 요소도 역시 표적할 수 있는 것으로 이해되어야 한다.

- [0032] 또 다른 방식의 분류에서, 본 발명자들은 다수의 서로 다른 종양들로부터 얻은 선택된 경로들 및/또는 경로 요소들이 도 8에서 예시적으로 도시된 바와 같이 분석될 수 있는 점을 고려한다. 도면에서, 선택된 경로 요소들 (예로, 종양 억제인자들 및 종양원 유전자들)이 서로 다른 종양들과 대비하여 좌표화되고, 이는 다수의 종양들에 공통적인 공유된 경로 특징들 및 분자적 서명들의 신속한 확인을 제공한다. 예를 들면, KRAS G12 변이체는 자궁, 직장 및 결장 암들과 연관되는 한편, APC는 결장 샘암종 및 직장 암들과 연관된다.
- [0033] 따라서, 본 발명자들은 환자 기록이 전형적으로 종양 세포들의 분자적 서명을 근거로 하는 하나 이상의 치료 추천들을 (다른 관련되지 않은 종양들과 공유된 경로 특징들을 근거로 하는 것과 함께) 포함할 것이라는 점을 고려한다. 다른 말로 하면, 첫 번째 종양을 가진 첫 번째 환자를 위한 치료 추천은 두 번째 구별된 종양을 가진 두 번째 환자와 공유된 경로 특징들을 근거로 할 수 있다.
- [0034] 본 명세서에서 상세한 설명에 및 다음의 청구항들 전체를 통하여 사용된 바, "하나 (a)", "하나 (an)" 및 "그 (the)"의 의미는 달리 문맥이 분명하게 진술하지 않는 경우라면 복수의 대상을 포함한다. 또한, 본 명세서에서 상세한 설명에 사용되는 바, "내에서 (in)"의 의미는 달리 문맥이 분명하게 진술하지 않는 경우라면 "내에서 (in)" 및 "위에서 (on)"를 포함한다. 또한, 본 명세서에서 사용되는 바, 그리고 달리 문맥이 진술하지 않는 경우라면, 용어 "결합된 (coupled to)"은 직접적인 결합 (서로 결합된 두 가지 요소들이 서로 접촉함) 및 간접적인 결합 (적어도 하나의 추가적인 요소가 두 가지 요소들 사이에 위치함) 둘 다를 포함하도록 의도된다. 따라서, 용어들 "결합된 (coupled to)" 및 "결합된 (coupled with)"은 동의어로 사용된다. 또한 본 명세서에 기술된 모든 방법들은 달리 본 명세서에서 표시되지 않거나 달리 문맥에 의해 분명하게 반박되지 않는 경우라면, 임의의 적합한 순서로 수행될 수 있다. 본 명세서에서 소정의 구현예들의 측면에서 제공된 임의의 및 모든 예들 또는 예시적인 언어 (예로, "와 같은")은 단지 본 발명을 더 잘 조명하도록 의도되고, 달리 주장되지 않으면 본 발명의 범위에 관한 제한이 되지는 않는다. 본 명세서에서의 언어는 본 발명의 실행에 필수적인 임의의 청구되지 않은 요소를 가리키는 것으로서 참작되어서는 안된다.
- [0035] 본 명세서에서 이미 기술된 변형들 외에도 더 많은 변형들이 본 발명의 개념들을 벗어나지 않고도 가능하다는 점은 당연히 당업자들에게 자명할 것이다. 본 발명의 주제 문제는 따라서 첨부된 청구항들의 범위를 제외하고 한정되어서는 안된다. 또한, 본 명세서 및 청구항들 둘 다를 해석할 때, 모든 용어들은 문맥과 부합하는 가장 광범위한 가능한 방식으로 해석되어야 한다. 상세하게, 용어들 "포함하다 (comprise)" 및 "포함하는 (comprising)"은 배타적이지 않은 방식으로 요소들, 구성성분들 또는 단계들을 언급하는 것으로서 해석되어야 하고, 언급된 요소들, 구성성분들 또는 단계들은 존재하고, 사용되고, 명확하게 언급되지 않은 다른 요소들, 구성성분들 또는 단계들과 함께 조합될 수 있는 점을 가리킨다. 본 명세서의 청구항들이 A, B, C... 및 N으로 이루어진 그룹으로부터 선택되는 적어도 하나의 물건을 언급하는 곳에서, 내용은 A 더하기 N 또는 B 더하기 N 등이 아닌 그룹으로부터 단지 하나의 요소를 요구하는 것으로서 해석되어야 한다.

도면

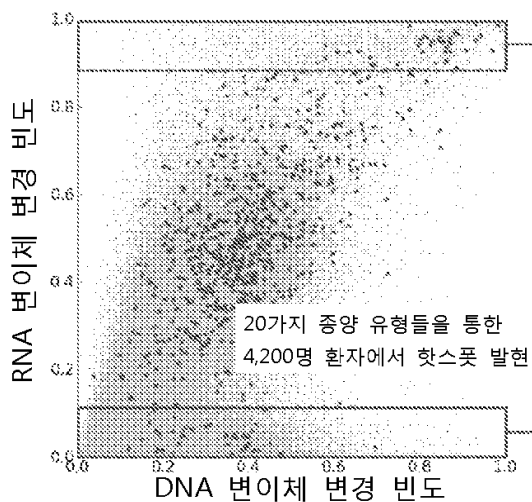
도면1

다수의 "작용가능한" 유전자를 가진 시료



도면2

변이체의 발현



고-발현된 핫스팟

BRAF V600	119개 흑색종 중 18개
AKT1 E17	15개 유방 관 중 3개
PIK3CA E542	10개 유방 엽 중 3개

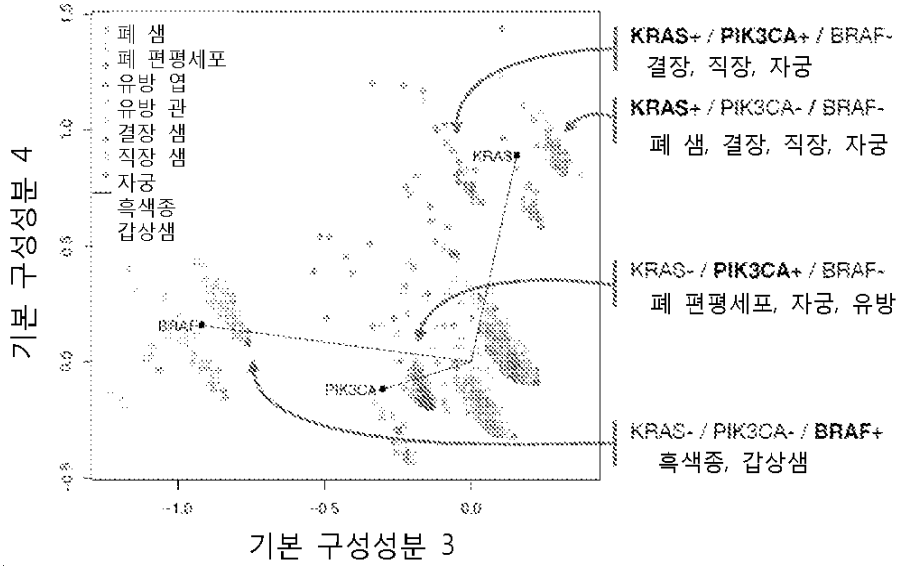
저-발현된 핫스팟

BRAF V600	204개 갑상샘 중 15개
	1개 교모세포종 중 1개
PIK3CA E545	41개 유방 관 중 4개
	7개 방광 중 2개
PIK3CA E542	10개 유방 엽 중 3개
NRAS Q61	31개 갑상샘 중 2개

● 핫스팟 돌연변이 ○ 모든 돌연변이

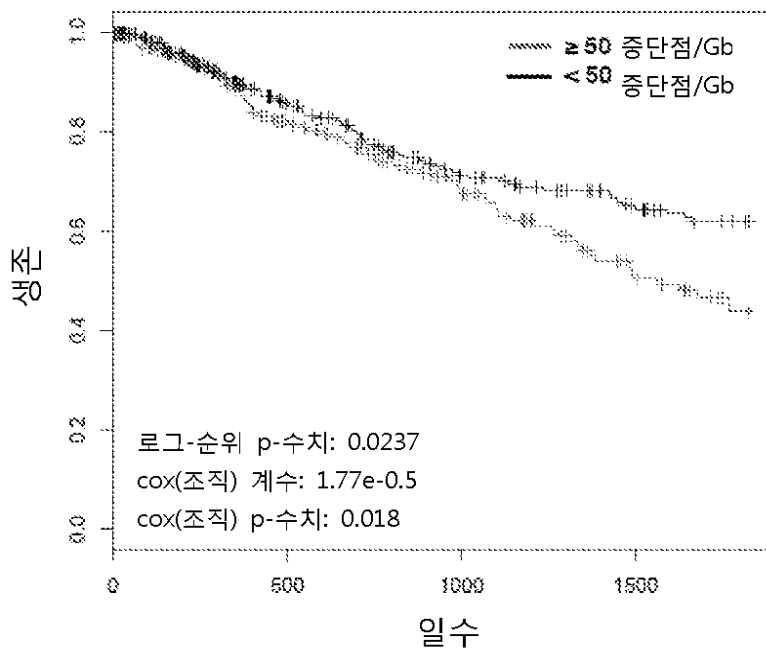
도면3

종양원 유전자 돌연변이들의 기본 구성성분 분석



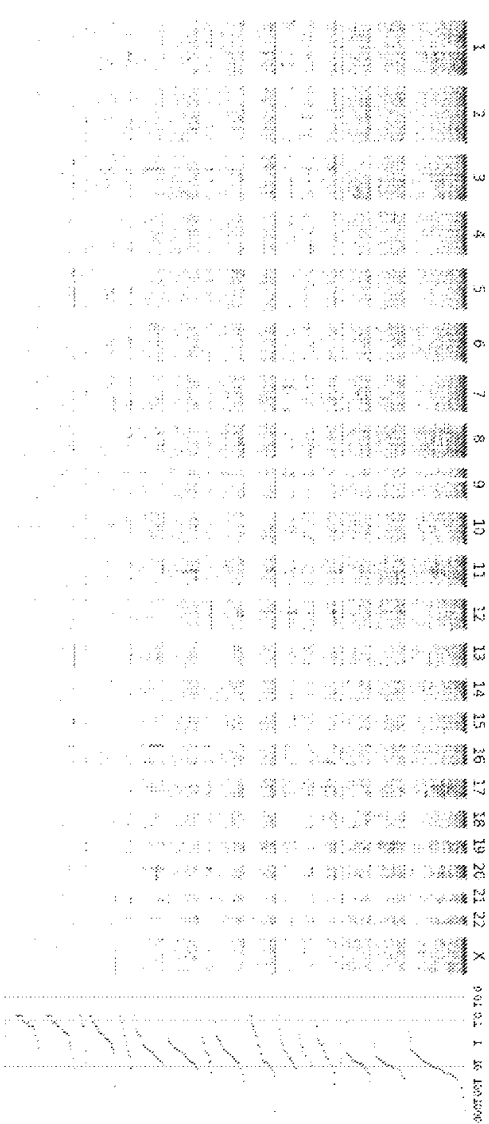
도면4

"초-재배열인자"는 더 나쁜 예후를 가진다

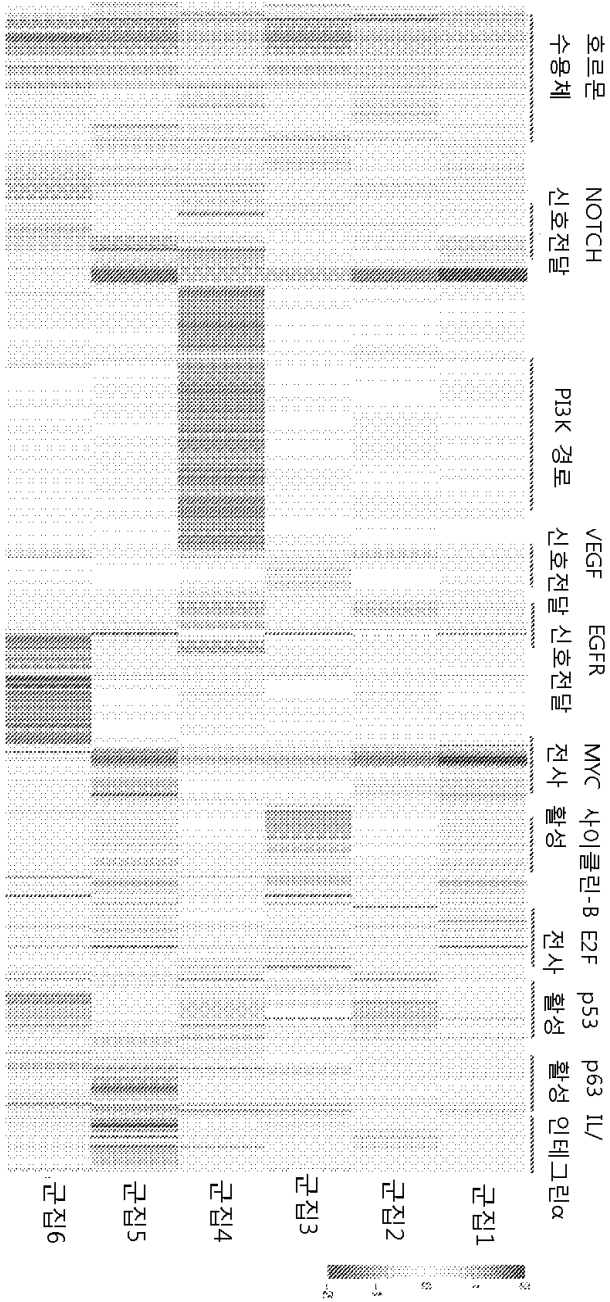


도면5

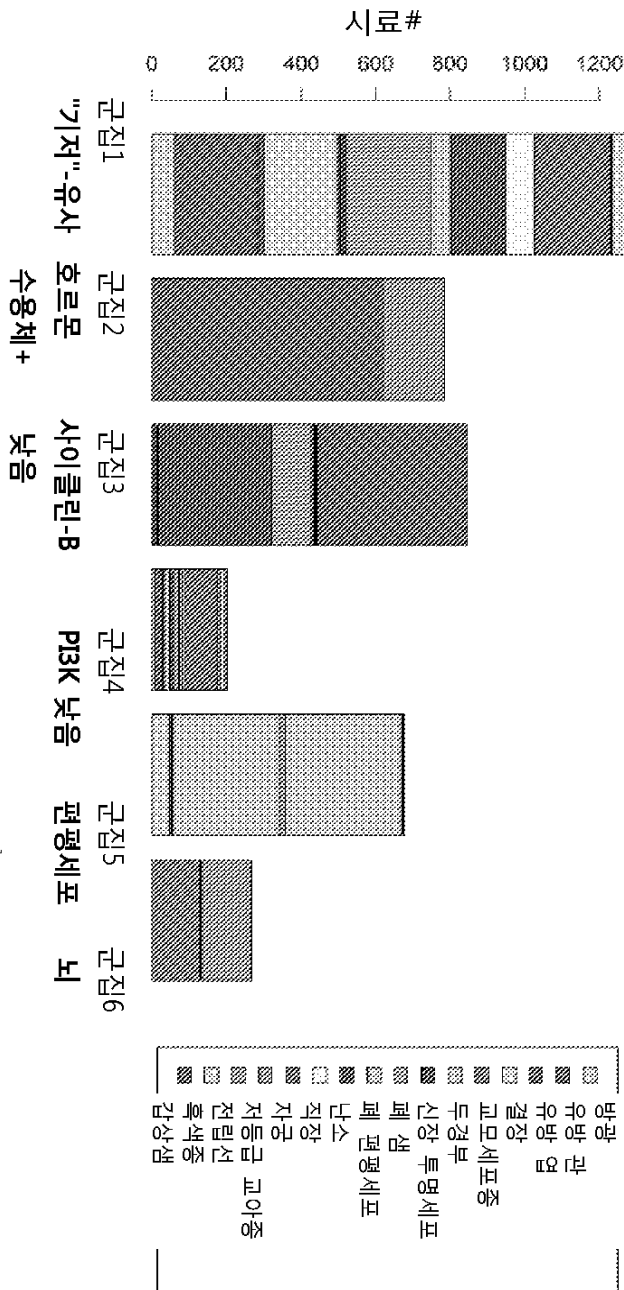
- 유방 관
- 난소
- 폐 편평세포
- 유방 암
- 유정
- 방광
- 고모세포종
- 자궁
- 폐 췌
- 직장
- 두경부
- 결장 암
- 전립선
- 신경 투명세포
- 신경 유두
- 신경 비염색
- 감상색
- AML



도면6



도면7



양방을 통해 군집된 경로 활성



도면8

흑색증  
폐 편평세포  
폐 샘  
방광  
위장  
결장 샘  
직장 샘  
두경부  
난소  
자궁  
교모세포종  
신장 투명세포  
유방 관  
유방 염  
저등급 교아종  
전립선  
신장 비염색  
감상샘  
췌장  
AML

