



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0000743
(43) 공개일자 2017년01월03일

(51) 국제특허분류(Int. Cl.)
G06F 19/18 (2011.01) G06F 19/22 (2011.01)
(52) CPC특허분류
G06F 19/18 (2013.01)
G06F 19/22 (2013.01)
(21) 출원번호 10-2015-0168829
(22) 출원일자 2015년11월30일
심사청구일자 2015년12월01일
(30) 우선권주장
1020150089448 2015년06월24일 대한민국(KR)

(71) 출원인
사회복지법인 삼성생명공익재단
서울특별시 용산구 이태원로55길 48 (한남동)
(72) 발명자
박웅양
서울특별시 서초구 신반포로33길 66, 102동 207호
(잠원동, 신반포청구아파트)
신현태
서울특별시 성동구 동일로 237, 105동 502호 (송
정동, 서울숲아이파크)
김나영
서울특별시 서초구 신반포로3길 19, 97동 503호
(반포동)
(74) 대리인
리엔목특허법인

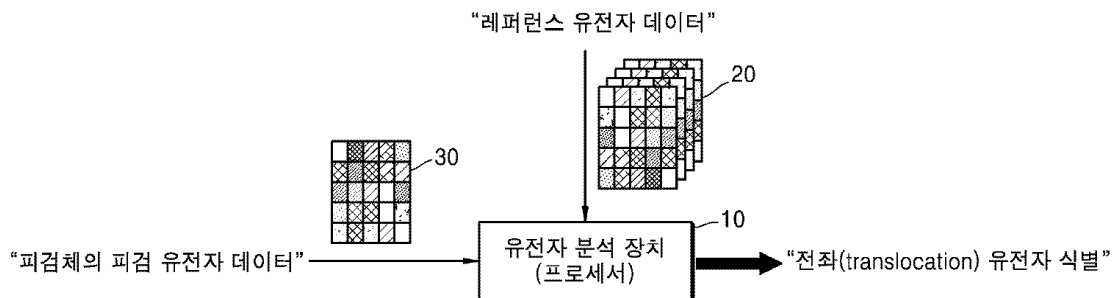
전체 청구항 수 : 총 17 항

(54) 발명의 명칭 유전자의 전좌를 분석하는 방법 및 장치

(57) 요약

유전자를 분석하는 방법 및 장치는, 피검 샘플의 차세대 시퀀싱(NGS) 데이터로부터 리드들에 관한 데이터를 획득하고, 리드들을 이용하여 피검 샘플의 염색체 내 전좌(translocation)의 가능성이 있는 후보 유전자 쌍들을 추출하고, 후보 유전자 쌍들 중에서 전좌 유전자를 식별한다.

대표도



이 발명을 지원한 국가연구개발사업

과제고유번호 HI13C2096

부처명 보건복지부

연구관리전문기관 한국보건산업진흥원

연구사업명 보건의료기술연구개발사업 > 첨단의료기술개발

연구과제명 선도형 암맞춤의료 프로그램의 임상적용

기 여 율 1/1

주관기관 삼성서울병원

연구기간 2013.12.01 ~ 2018.11.30

명세서

청구범위

청구항 1

피검 샘플의 차세대 시퀀싱(NGS) 데이터로부터, 스플릿 리드들(split reads) 및 불일치(discordantly) 정렬된 PE(paired-end) 리드들에 관한 데이터를 획득하는 단계;

상기 스플릿 리드들 및 상기 PE 리드들을 이용하여 상기 피검 샘플의 염색체 내 전좌(translocation)의 가능성이 있는 제 1 후보 유전자 쌍들을 추출하는 단계; 및

상기 스플릿 리드들이 나타내는 브레이크 포인트들(break points) 및 상기 제 1 후보 유전자 쌍들의 퓨전(fusion) 방향에 기초하여, 상기 제 1 후보 유전자 쌍들 중에서 전좌 유전자를 식별하는 단계를 포함하는, 유전자를 분석하는 방법.

청구항 2

제 1 항에 있어서,

상기 식별하는 단계는

상기 추출된 제 1 후보 유전자 쌍들 중에서, 동일한 커버리지에 속한 브레이크 포인트를 갖는 복수의 스플릿 리드들이 정렬된 유전자를 포함하는 제 2 후보 유전자 쌍들을 추출하는 단계를 포함하고,

상기 전좌 유전자는

상기 추출된 제 2 후보 유전자 쌍들로부터 식별되는, 방법.

청구항 3

제 2 항에 있어서,

상기 추출된 제 2 후보 유전자 쌍들에 포함된 상기 유전자는

상기 동일한 커버리지에 속한 상기 브레이크 포인트를 갖는 상기 스플릿 리드들의 개수가 소정 임계값 이상인, 방법.

청구항 4

제 2 항에 있어서,

상기 식별하는 단계는

상기 추출된 제 2 후보 유전자 쌍들 중에서, 서로 다른 유전자들 간의 상기 퓨전 방향이 5' 엔드(end)부터 3' 엔드이거나, 또는 3' 엔드부터 5' 엔드인 제 3 후보 유전자 쌍들을 추출하는 단계를 포함하고,

상기 전좌 유전자는

상기 추출된 제 2 후보 유전자 쌍들로부터 식별되는, 방법.

청구항 5

제 1 항에 있어서,

상기 NGS 데이터는

BAM(binary version of SAM) 포맷 또는 SAM(Sequence Alignment/Map) 포맷의 데이터를 포함하는, 방법.

청구항 6

제 5 항에 있어서,

상기 획득하는 단계는

상기 BAM 포맷 또는 상기 SAM 포맷의 데이터로부터, 상기 스플릿 리드들 및 상기 PE 리드들 각각에 대한 FLAG 및 CIGAR(Compact Idiosyncratic Gapped Alignment Report) 스트링의 데이터를 획득하는, 방법.

청구항 7

제 1 항에 있어서,

상기 NGS 데이터는

상기 피검 샘플에서 표적 유전자들의 염기서열을 식별하기 위한 표적 시퀀싱(targeted sequencing)에 의해 생성되는, 방법.

청구항 8

제 1 항에 있어서,

상기 피검 샘플은

생검 샘플 또는 포르말린-고정 파라핀-내장(Formalin-fixed, paraffin-embedded, FFPE) 샘플인, 방법.

청구항 9

제 1 항 내지 제 8 항 중에 어느 한 항의 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

청구항 10

피검 샘플의 차세대 시퀀싱(NGS) 데이터로부터, 스플릿 리드들(split reads) 및 불일치(discordantly) 정렬된 PE(paired-end) 리드들에 관한 데이터를 획득하는 리드 분석부; 및

상기 스플릿 리드들 및 상기 PE 리드들을 이용하여 상기 피검 샘플의 염색체 내 전좌(translocation)의 가능성이 있는 제 1 후보 유전자 쌍들을 추출하고, 상기 스플릿 리드들이 나타내는 브레이크 포인트들(break points) 및 상기 제 1 후보 유전자 쌍들의 퓨전(fusion) 방향에 기초하여, 상기 제 1 후보 유전자 쌍들 중에서 전좌 유전자를 식별하는 전좌 식별부를 포함하는, 유전자를 분석하는 장치.

청구항 11

제 10 항에 있어서,

상기 전좌 식별부는

상기 추출된 제 1 후보 유전자 쌍들 중에서, 동일한 커버리지에 속한 브레이크 포인트를 갖는 복수의 스플릿 리드들이 정렬된 유전자를 포함하는 제 2 후보 유전자 쌍들을 추출하고,

상기 전좌 유전자는

상기 추출된 제 2 후보 유전자 쌍들로부터 식별되는, 장치.

청구항 12

제 11 항에 있어서,

상기 추출된 제 2 후보 유전자 쌍들에 포함된 상기 유전자는

상기 동일한 커버리지에 속한 상기 브레이크 포인트를 갖는 상기 스플릿 리드들의 개수가 소정 임계값 이상인, 장치.

청구항 13

제 11 항에 있어서,

상기 전좌 식별부는

상기 추출된 제 2 후보 유전자 쌍들 중에서, 서로 다른 유전자들 간의 상기 퓨전 방향이 5' 엔드(end)부터 3' 엔드이거나, 또는 3' 엔드부터 5' 엔드인 제 3 후보 유전자 쌍들을 추출하고,

상기 전좌 유전자는

상기 추출된 제 2 후보 유전자 쌍들로부터 식별되는, 장치.

청구항 14

제 10 항에 있어서,

상기 NGS 데이터는

BAM(binary version of SAM) 포맷 또는 SAM(Sequence Alignment/Map) 포맷의 데이터를 포함하는, 장치.

청구항 15

제 14 항에 있어서,

상기 리드 분석부는

상기 BAM 포맷 또는 상기 SAM 포맷의 데이터로부터, 상기 스플릿 리드들 및 상기 PE 리드들 각각에 대한 FLAG 및 CIGAR(Compact Idiosyncratic Gapped Alignment Report) 스트링의 데이터를 획득하는, 장치.

청구항 16

제 10 항에 있어서,

상기 NGS 데이터는

상기 피검 샘플에서 표적 유전자들의 염기서열을 식별하기 위한 표적 시퀀싱(targeted sequencing)에 의해 생성되는, 장치.

청구항 17

제 10 항에 있어서,

상기 피검 샘플은

생검 샘플 또는 포르말린-고정 파라핀-내장(Formalin-fixed, paraffin-embedded, FFPE) 샘플인, 장치.

발명의 설명

기술 분야

[0001] 유전자를 분석하는 방법 및 장치에 관한 것으로서, 특히 전좌(translocation) 유전자에 관한 데이터를 분석하는 방법 및 장치에 관한다.

배경 기술

[0002] 유전체(genome)란 한 생물이 가지는 모든 유전 정보를 말한다. 어느 한 개인의 유전체의 시퀀싱(sequencing)을 위하여, DNA 칩 및 차세대 서열화(Next Generation Sequencing) 기술, 차차세대 서열화(Next Next Generation Sequencing) 기술 등 여러 기술들이 개발되고 있다. 핵산 서열, 단백질 등과 같은 유전 정보들은 분석은 당뇨병, 암과 같은 질병을 발현시키는 유전자를 찾거나, 유전적 다양성과 개체의 발현 특성 간의 상관관계 등을 파악하기 위하여 폭넓게 활용된다. 특히, 개인으로부터 수집된 유전 데이터는 서로 다른 증상이나 질병의 진행과 관련된 개인의 유전적인 특징을 규명하는데 있어서 중요하다. 따라서, 개인의 핵산 서열, 단백질 등과 같은 유전자 데이터는 현재와 미래의 질병 관련 정보를 파악하여 질병을 예방하거나 질병의 초기 단계에서 최적의 치료 방법을 선택할 수 있도록 하는 핵심적인 데이터이다. 최근, 시퀀싱 기술의 발달로 다양한 종류의 구조적 변이(structure variation)를 발굴해내는 시도가 많아졌지만, 여전히 상당한 양의 위양성(false positive)나 위음성(false negative) 결과의 발생은 여전히 바이오인포매틱스(bioinformatics)의 도전적인 요소가 많이 남아있음을 반증한다.

발명의 내용

해결하려는 과제

[0003] 유전자를 분석하는 방법 및 장치를 제공하는데 있다. 본 실시예가 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제들로 한정되지 않으며, 이하의 실시예들로부터 또 다른 기술적 과제들이 유추될 수 있다.

과제의 해결 수단

- [0004] 일 측면에 따르면, 유전자를 분석하는 방법은, 피검 샘플의 차세대 시퀀싱(NGS) 데이터로부터, 스플릿 리드들(split reads) 및 불일치(discordantly) 정렬된 PE(paired-end) 리드들에 관한 데이터를 획득하는 단계; 상기 스플릿 리드들 및 상기 PE 리드들을 이용하여 상기 피검 샘플의 염색체 내 전좌(translocation)의 가능성이 있는 제 1 후보 유전자 쌍들을 추출하는 단계; 및 상기 스플릿 리드들이 나타내는 브레이크 포인트들(break points) 및 상기 제 1 후보 유전자 쌍들의 퓨전(fusion) 방향에 기초하여, 상기 제 1 후보 유전자 쌍들 중에서 전좌 유전자를 식별하는 단계를 포함한다.
- [0005] 또한, 상기 식별하는 단계는 상기 추출된 제 1 후보 유전자 쌍들 중에서, 동일한 커버리지에 속한 브레이크 포인트를 갖는 복수의 스플릿 리드들이 정렬된 유전자를 포함하는 제 2 후보 유전자 쌍들을 추출하는 단계를 포함하고, 상기 전좌 유전자는 상기 추출된 제 2 후보 유전자 쌍들로부터 식별된다.
- [0006] 또한, 상기 추출된 제 2 후보 유전자 쌍들에 포함된 상기 유전자는 상기 동일한 커버리지에 속한 상기 브레이크 포인트를 갖는 상기 스플릿 리드들의 개수가 소정 임계값 이상일 수 있다.
- [0007] 또한, 상기 식별하는 단계는 상기 추출된 제 2 후보 유전자 쌍들 중에서, 서로 다른 유전자들 간의 상기 퓨전 방향이 5' 엔드(end)부터 3' 엔드이거나, 또는 3' 엔드부터 5' 엔드인 제 3 후보 유전자 쌍들을 추출하는 단계를 포함하고, 상기 전좌 유전자는 상기 추출된 제 2 후보 유전자 쌍들로부터 식별된다.
- [0008] 또한, 상기 NGS 데이터는 BAM(binary version of SAM) 포맷 또는 SAM(Sequence Alignment/Map) 포맷의 데이터를 포함한다.
- [0009] 또한, 상기 획득하는 단계는 상기 BAM 포맷 또는 상기 SAM 포맷의 데이터로부터, 상기 스플릿 리드들 및 상기 PE 리드들 각각에 대한 FLAG 및 CIGAR(Compact Idiosyncratic Gapped Alignment Report) 스트링의 데이터를 획득한다.
- [0010] 또한, 상기 NGS 데이터는 상기 피검 샘플에서 표적 유전자들의 염기서열을 식별하기 위한 표적 시퀀싱(targeted sequencing)에 의해 생성된다.
- [0011] 또한, 상기 피검 샘플은 생검 샘플 또는 포르말린-고정 파라핀-내장(Formalin-fixed, paraffin-embedded, FFPE) 샘플이다.
- [0012] 다른 측면에 따르면, 상기 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록 매체를 제공한다.
- [0013] 또 다른 측면에 따르면, 유전자를 분석하는 장치는, 피검 샘플의 차세대 시퀀싱(NGS) 데이터로부터, 스플릿 리드들(split reads) 및 불일치(discordantly) 정렬된 PE(paired-end) 리드들에 관한 데이터를 획득하는 리드 분석부; 및 상기 스플릿 리드들 및 상기 PE 리드들을 이용하여 상기 피검 샘플의 염색체 내 전좌(translocation)의 가능성이 있는 제 1 후보 유전자 쌍들을 추출하고, 상기 스플릿 리드들이 나타내는 브레이크 포인트들(break points) 및 상기 제 1 후보 유전자 쌍들의 퓨전(fusion) 방향에 기초하여, 상기 제 1 후보 유전자 쌍들 중에서 전좌 유전자를 식별하는 전좌 식별부를 포함한다.
- [0014] 또한, 상기 전좌 식별부는 상기 추출된 제 1 후보 유전자 쌍들 중에서, 동일한 커버리지에 속한 브레이크 포인트를 갖는 복수의 스플릿 리드들이 정렬된 유전자를 포함하는 제 2 후보 유전자 쌍들을 추출하고, 상기 전좌 유전자는 상기 추출된 제 2 후보 유전자 쌍들로부터 식별된다.
- [0015] 또한, 상기 추출된 제 2 후보 유전자 쌍들에 포함된 상기 유전자는 상기 동일한 커버리지에 속한 상기 브레이크 포인트를 갖는 상기 스플릿 리드들의 개수가 소정 임계값 이상일 수 있다.
- [0016] 또한, 상기 전좌 식별부는 상기 추출된 제 2 후보 유전자 쌍들 중에서, 서로 다른 유전자들 간의 상기 퓨전 방향이 5' 엔드(end)부터 3' 엔드이거나, 또는 3' 엔드부터 5' 엔드인 제 3 후보 유전자 쌍들을 추출하고, 상기

전좌 유전자는 상기 추출된 제 2 후보 유전자 쌍들로부터 식별된다.

- [0017] 또한, 상기 NGS 데이터는 BAM(binary version of SAM) 포맷 또는 SAM(Sequence Alignment/Map) 포맷의 데이터를 포함한다.
- [0018] 또한, 상기 리드 분석부는 상기 BAM 포맷 또는 상기 SAM 포맷의 데이터로부터, 상기 스플릿 리드들 및 상기 PE 리드들 각각에 대한 FLAG 및 CIGAR(Compact Idiosyncratic Gapped Alignment Report) 스트링의 데이터를 획득한다.
- [0019] 또한, 상기 NGS 데이터는 상기 피검 샘플에서 표적 유전자들의 염기서열을 식별하기 위한 표적 시퀀싱(targeted sequencing)에 의해 생성된다.
- [0020] 또한, 상기 피검 샘플은 생검 샘플 또는 포르말린-고정 파라핀-내장(Formalin-fixed, paraffin-embedded, FFPE) 샘플이다.

발명의 효과

- [0021] 상기된 바에 따르면, 피검체의 피검 샘플로부터 추출된 피검 유전자로부터 전좌 유전자가 존재하는지를 보다 정확하게 분석해 낼 수 있다.

도면의 간단한 설명

- [0022] 도 1은 일 실시예에 따른 유전자 분석 장치를 설명하기 위한 도면이다.
- 도 2는 일 실시예에 따른 유전자 분석 장치의 하드웨어 구성들을 도시한 블록도이다.
- 도 3은 일 실시예에 따른 PE 리드들을 설명하기 위한 도면이다.
- 도 4는 일 실시예에 따라 불일치 정렬된 PE 리드들을 설명하기 위한 도면이다.
- 도 5는 일 실시예에 따른 스플릿 리드들을 설명하기 위한 도면이다.
- 도 6은 일 실시예에 따른 피검체의 생검 샘플로부터 획득된 리드들을 레퍼런스 유전자 데이터와 비교한 IGV(Integrative Genomics Viewer) 스크린샷을 설명하기 위한 도면이다.
- 도 7은 일 실시예에 따른 피검체의 FFPE 샘플로부터 획득된 리드들을 레퍼런스 유전자 데이터와 비교한 IGV 스크린샷을 설명하기 위한 도면이다.
- 도 8은 일 실시예에 따라 전좌 식별부에서 후보 유전자 쌍들을 추출하여 전좌 유전자를 식별하는 방법의 흐름도이다.
- 도 9는 일 실시예에 따라 스플릿 리드들의 브레이크 포인트들을 이용하여 제 2 후보 유전자 쌍들을 추출하는 것을 설명하기 위한 도면이다.
- 도 10은 일 실시예에 따라 퓨전 방향의 적절성을 이용하여 제 3 후보 유전자 쌍들을 추출하는 것을 설명하기 위한 도면이다.
- 도 11은 일 실시예에 따라 EML4(echinoderm microtubule-associated protein-like 4) 및 ALK(anaplastic lymphoma kinase)의 전좌 유전자를 식별한 결과를 설명하기 위한 도면이다.
- 도 12는 일 실시예에 따라 유전자를 분석하는 방법의 흐름도이다.
- 도 13은 일 실시예에 따른 컴퓨팅 장치의 하드웨어 구성들을 도시한 블록도이다.

발명을 실시하기 위한 구체적인 내용

- [0023] 본 실시예들에서 사용되는 용어는 본 실시예들에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 당 기술분야에 종사하는 기술자의 의도 또는 관례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 임의로 선정된 용어도 있으며, 이 경우 해당 실시예의 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서, 본 실시예들에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 실시예들의 전반에 걸친 내용을 토대로 정의되어야 한다.
- [0024] 실시예들에 대한 설명들에서, 어떤 부분이 다른 부분과 연결되어 있다고 할 때, 이는 직접적으로 연결되어 있는

경우뿐 아니라, 그 중간에 다른 구성요소를 사이에 두고 전기적으로 연결되어 있는 경우도 포함한다. 또한 어떤 부분이 어떤 구성요소를 포함한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 실시예들에 기재된 “...부”, “...모듈”의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다.

- [0025] 본 실시예들에서 사용되는 “구성된다” 또는 “포함한다” 등의 용어는 명세서 상에 기재된 여러 구성 요소들, 또는 여러 단계들을 반드시 모두 포함하는 것으로 해석되지 않아야 하며, 그 중 일부 구성 요소들 또는 일부 단계들은 포함되지 않을 수도 있고, 또는 추가적인 구성 요소 또는 단계들을 더 포함할 수 있는 것으로 해석되어야 한다.
- [0026] 하기 실시예들에 대한 설명은 권리범위를 제한하는 것으로 해석되지 말아야 하며, 해당 기술분야의 당업자가 용이하게 유추할 수 있는 것은 실시예들의 권리범위에 속하는 것으로 해석되어야 할 것이다. 이하 첨부된 도면들을 참조하면서 오로지 예시를 위한 실시예들을 상세히 설명하기로 한다.
- [0027] 도 1은 일 실시예에 따른 유전자 분석 장치를 설명하기 위한 도면이다.
- [0028] 도 1을 참고하면, 유전자 분석 장치(10)는 정상인 집단의 레퍼런스 유전자 데이터(20) 및 피검체의 피검 생물학적 샘플로부터 획득된 피검 유전자 데이터(30)를 이용하여, 피검 샘플의 피검 유전자에 전좌(translocation) 유전자가 존재하는지 여부를 식별할 수 있다.
- [0029] 유전자 분석 장치(10)에서 수신하는 피검 유전자 데이터(30)는, 차세대 시퀀싱(next generation sequencing, NGS)에 의해 획득된 NGS 데이터일 수 있고, NGS 데이터는 BAM(binary version of SAM) 포맷 또는 SAM(Sequence Alignment/Map) 포맷의 유전자 데이터를 포함할 수 있다. BAM 포맷 또는 SAM 포맷은 보통 짧은 리드들(short reads)에 관한 데이터를 서술하는 포맷으로 이용될 수 있다. BAM 포맷 또는 SAM 포맷의 파일에는 리드(read)의 시작 포인트, 리드의 방향(direction), 매핑 퀄리티, 얼라인먼트의 차수를 나타내는 FLAG, CIGAR(Compact Idiosyncratic Gapped Alignment Report) 스트링 등에 관한 텍스트 데이터가 포함될 수 있다. 여기서, FLAG는 1차(primary) 얼라인먼트-1차 얼라인먼트 쌍, 1차 얼라인먼트-2차(secondary) 얼라인먼트 쌍, 2차 얼라인먼트-1차 얼라인먼트 쌍 또는 2차 얼라인먼트-2차 얼라인먼트 쌍 중 어느 얼라인먼트 쌍인지를 구분하기 위한 식별자일 수 있다. 다양한 얼라인먼트 쌍을 생성함으로써 다양한 서포팅 리드들(supporting reads)을 확보할 수 있다.
- [0030] 레퍼런스 유전자 데이터(20)는, NCBI(National Center for Biotechnology Information), Gene Expression Omnibus (GEO) 등과 같은 당해 기술분야에서 이미 공지된 데이터베이스(DB)로부터 획득되거나, 또는 피검체의 피검 유전자들을 분석하기 위하여 모집된 사람들의 생물학적 샘플들로부터 획득된 것일 수 있다.
- [0031] 한편, 레퍼런스 유전자 데이터(20)에 포함된 레퍼런스 유전자들 또는 피검 유전자 데이터(30)에 포함된 피검 유전자들은, 생검 조직, 포르말린-고정 조직 또는 파라핀-내장(Formalin-fixed, paraffin-embedded) 조직으로부터 획득된 것일 수 있다.
- [0032] 전좌는, 염색체의 일부분에 절단이 일어나고, 그 절단된 단편이 같은 염색체 내의 다른 부분 또는 다른 염색체에 결합된 현상을 의미하는 것으로써, 결국 염색체의 구조적 변이(structure variation)를 의미한다.
- [0033] 유전자 분석 장치(10)는 정상인 집단으로부터 획득된 레퍼런스 유전자 데이터(20) 대비 피검체의 피검 샘플로부터 획득된 피검 유전자 데이터(30)에 전좌 유전자가 존재하는지 여부를 판단할 수 있다. 여기서, 유전자 분석 장치(10)에 의해 분석되는 유전자는 DNA(deoxyribonucleic acid), RNA(ribonucleic acid) 등과 같은 핵산을 의미할 수 있다.
- [0034] 본 실시예들에서, 정상인 집단은 특정 질병, 예를 들어 암, 종양 등이 발견되지 않은 일반 사람들로 구성된 집단을 의미하고, 피검체는 암, 종양 등과 같은 특정 질병이 발견된 환자를 의미할 수 있다. 한편, 본 실시예들에서 정상인 집단, 피검체는 인간이 아닌, 다른 동물들에 해당될 수도 있다.
- [0035] 유전자 분석 장치(10)는 유전자 데이터들(20 및 30)을 분석하여 전좌 유전자를 식별하기 위한 다양한 명령어들, 다양한 알고리즘들을 수행하는 데이터 프로세싱의 기능을 갖는 적어도 하나의 프로세서로 구현될 수 있다.
- [0036] 도 2는 일 실시예에 따른 유전자 분석 장치의 하드웨어 구성들을 도시한 블록도이다.
- [0037] 도 2를 참고하면, 유전자 분석 장치(10)는 리드 분석부(110) 및 전좌 식별부(120)를 포함할 수 있다. 한편, 도 2에 도시된 유전자 분석 장치(10)는 본 실시예의 특징이 흐려지는 것을 방지하기 위하여 본 실시예에 관련된 구

성요소들만이 도시되어 있을 뿐이므로, 유전자 분석 장치(10)는 도 2에 도시된 구성요소들 외에 다른 범용적인 구성요소들이 더 포함될 수 있다.

- [0038] 리드 분석부(110)는 앞서 도 1에서 설명된 피검 유전자 데이터(30)에 포함된, 피검 샘플의 차세대 시퀀싱(NGS) 데이터로부터, 스플릿 리드들(split reads) 및 불일치 정렬된(discordantly aligned) PE(paired-end) 리드들에 관한 데이터를 획득한다.
- [0039] 피검 유전자 데이터(30)에 포함된 NGS 데이터는, BAM 포맷 또는 SAM 포맷의 데이터로서, 리드 분석부(110)는 BAM 포맷 또는 SAM 포맷의 데이터로부터, 스플릿 리드들 및 PE 리드들 각각에 대한 리드(read)의 시작 포인트, 리드의 방향(direction), 매핑 퀄리티, 얼라인먼트의 차수를 나타내는 FLAG, CIGAR(Compact Idiosyncratic Gapped Alignment Report) 스트링 등에 관한 텍스트 데이터들을 획득할 수 있다.
- [0040] 일반적으로, 피검 샘플로부터 피검 유전자의 염기서열을 분석하기 위한 시퀀싱 기술로서, WGS(whole genome sequencing), WES(whole exome sequencing) 등과 같은 NGS 기술들이 알려져 있다. 다만, 본 실시예에 따른 NGS 데이터는 피검 샘플에서 전체 게놈이 아닌, 일부의 표적 유전자들(target genes)의 염기서열을 식별하기 위한 표적 시퀀싱(targeted sequencing)에 의해 생성될 수 있다.
- [0041] 한편, 피검 샘플은 피검체로부터 획득된 생검 샘플, 포르말린-고정 파라핀-내장(Formalin-fixed, paraffin-embedded, FFPE) 샘플일 수 있다.
- [0042] 전좌 식별부(120)는 스플릿 리드들 및 불일치 정렬된 PE 리드들을 이용하여 피검 샘플의 염색체 내 전좌(translocation)의 가능성이 있는 제 1 후보 유전자 쌍들을 추출한다. 스플릿 리드들로 시퀀싱된 유전자, 또는 불일치 정렬된 PE 리드들로 시퀀싱된 유전자는, 유전자 일부의 염기서열이 레퍼런스 유전자(정상인의 유전자)와 차이가 있을 가능성이 높은 후보로 간주될 수 있다.
- [0043] 전좌 식별부(120)는 스플릿 리드들이 나타내는 브레이크 포인트들(break points) 및 제 1 후보 유전자 쌍들의 퓨전(fusion) 방향에 기초하여, 제 1 후보 유전자 쌍들 중에서 전좌 유전자를 식별한다.
- [0044] 보다 상세하게 설명하면, 전좌 식별부(120)는 제 1 후보 유전자 쌍들 중에서, 동일한 커버리지에 속한 브레이크 포인트를 갖는 복수의 스플릿 리드들이 정렬된 유전자를 포함하는 제 2 후보 유전자 쌍들을 추출할 수 있다. 즉, 전좌 유전자는, 제 1 후보 유전자 쌍들보다 좁은 범위로 압축된 제 2 후보 유전자 쌍들로부터 식별될 수 있다. 여기서, 제 2 후보 유전자 쌍들에 포함된 유전자는, 동일한 커버리지에 속한 브레이크 포인트를 갖는 스플릿 리드들의 개수가 소정 임계값 이상일 수 있다. 커버리지는, 시퀀싱 오차를 고려하여 동일한 브레이크 포인트라 간주될 수 있는 브레이크 포인트의 오차 범위를 의미한다. 예를 들어, 소정 임계값이 3개인 경우, 동일한 커버리지 내에 속한 브레이크 포인트를 갖는 스플릿 리드들의 개수가 3개 이상인 유전자가 제 2 후보 유전자 쌍들에 포함될 수 있다. 다만, 소정 임계값은 다양하게 바뀔 수 있다.
- [0045] 나아가서, 전좌 식별부(120)는 제 2 후보 유전자 쌍들 중에서, 서로 다른 유전자들 간의 퓨전 방향이 5' 엔드(end)부터 3' 엔드이거나, 또는 3' 엔드부터 5' 엔드인 제 3 후보 유전자 쌍들을 추출할 수 있다. 즉, 전좌 유전자는, 제 2 후보 유전자 쌍들보다 더 좁은 범위로 압축된 제 3 후보 유전자 쌍들로부터 식별될 수 있다. 예를 들어, 유전자 A의 3' 엔드와 유전자 B의 3' 엔드가 결합된 퓨전 유전자의 경우에는, 생물학적 발현의 기능을 갖지 않는 무의미한 퓨전 유전자일 수 있다. 따라서, 전좌 식별부(120)는 퓨전 방향을 고려하여 제 2 후보 유전자 쌍들 중에서 퓨전 방향이 부적절한 유전자 쌍들을 필터링하여 제 3 후보 유전자 쌍들을 추출한다.
- [0046] 전좌 식별부(120)는, 최종적으로, 제 3 후보 유전자 쌍들에 포함된 유전자 쌍은 전좌 유전자인 것으로 판단할 수 있다.
- [0047] 도 3은 일 실시예에 따른 PE 리드들을 설명하기 위한 도면이다.
- [0048] 도 3을 참고하면, PE의 시퀀싱은 피검 샘플의 피검 유전자를 양쪽 엔드들에서 각각 시퀀싱하는 것을 의미한다. 본 실시예들에서는 표적 시퀀싱을 수행하는 것으로 앞서 설명되었으므로, 시퀀싱될 피검 샘플의 핵산(DNA, RNA 등)(300)은 500bp(base pair)의 크기인 것으로 가정할 수 있다. 리드 크기가 100bp인 것으로 설정된 경우, PE 리드들(310 및 320)은 핵산(300)의 양쪽 엔드들로부터 시퀀싱되어 생성될 수 있다. 리드 크기가 핵산(300)의 크기보다 작으므로, 핵산(300)의 나머지 부분들에 대해서는 별도의 리드들이 생성되지 않을 수 있다. 한편, 본 실시예에 따른 PE의 시퀀싱은 엑손(305)뿐만 아니라, 인트론도 함께 시퀀싱하여 PE 리드들(310 및 320)을 획득할 수 있다. PE의 시퀀싱을 이용하는 이유에 대해서는 도 4를 참조하여 보다 상세하게 설명하도록 한다.

- [0049] 도 4는 일 실시예에 따라 불일치 정렬된 PE 리드들을 설명하기 위한 도면이다.
- [0050] 전좌 유전자는 동일 염색체 내의 서로 다른 유전자가 결합되거나 또는 다른 염색체들 내의 서로 다른 유전자가 결합된(퓨전된) 것일 수 있다. 결국, 전좌 유전자의 5' 엔드 쪽의 유전자 염기서열과 3' 엔드 쪽의 유전자 염기서열은 각각 서로 다른 유전자들로부터 유래된 것이다. 따라서, PE 리드들(410 및 420)의 염기서열은, 그 시퀀싱 위치들에 대응되는 정상인의 레퍼런스 유전자의 염기서열과는 차이가 명백할 것이다.
- [0051] 도 4를 참고하면, 정상인의 레퍼런스 유전자의 염기서열을 기준으로 PE 리드(410)가 염색체 2 내에 존재하는 어느 유전자(401)에 매핑되고 PE 리드(420)가 염색체 3 내에 존재하는 다른 유전자(402)에 매핑된 경우, PE 리드들(410 및 420) 각각은 피검 샘플에 존재하는 퓨전 유전자(전좌 유전자)로부터 기인되었을 가능성이 있다는 점을 유추할 수 있다. 이와 같은 리드들은 불일치 정렬된 PE 리드들(410 및 420)이라 정의될 수 있다. 이와 같은 불일치 정렬된 PE 리드들(410 및 420)에 매핑된 유전자 쌍(401 및 402)은 앞서 설명된 제 1 후보 유전자 쌍들에 포함될 수 있다.
- [0052] 도 5는 일 실시예에 따른 스플릿 리드들을 설명하기 위한 도면이다.
- [0053] 도 5를 참고하면, 스플릿 리드(510)는 레퍼런스 유전자의 염기서열과 일부만 일치하고 나머지 일부와는 일치하지 않는 염기서열을 갖는 리드를 의미한다. 피검 샘플의 피검 유전자에 대응되는 어느 리드가, 대응되는 레퍼런스 유전자의 염기서열과 일치하는 경우에는, 그 피검 유전자의 염기서열에는 구조적 변이가 없는 것으로 간주될 수 있다. 하지만, 스플릿 리드(510)와 같이, 레퍼런스 유전자의 염기서열과 일부만 일치하고 나머지 일부는 일치하지 않는다면, 피검 유전자의 염기서열은 레퍼런스 유전자의 염기서열과 차이가 있다는 점을 유추할 수 있다. 따라서, 이와 같은 스플릿 리드(510)에 매핑된 유전자들(501 및 502)은 앞서 설명된 제 1 후보 유전자 쌍들에 포함될 수 있다.
- [0054] 예를 들어, 스플릿 리드(511)의 CIGAR 스트링이 75M25S인 경우, 스플릿 리드(511)는 75개의 염기서열만 유전자 A(501)와 일치하고 나머지 25개의 염기서열은 유전자 A(501)와 일치하지 않는 리드이다. 또한, 스플릿 리드(512)의 CIGAR 스트링이 80M20S인 경우, 스플릿 리드(512)는 80개의 염기서열만 유전자 A(501)와 일치하고 나머지 20개의 염기서열은 유전자 A(501)와 일치하지 않는 리드이다. 리드 분석부(도 1의 110)는 BAM 포맷 또는 SAM 포맷의 데이터로부터 이와 같은 스플릿 리드들의 데이터를 획득할 수 있다.
- [0055] 도 6은 일 실시예에 따른 피검체의 생검 샘플로부터 획득된 리드들을 레퍼런스 유전자 데이터와 비교한 IGV(Integrative Genomics Viewer) 스크린샷을 설명하기 위한 도면이다.
- [0056] 도 6을 참고하면, IGV 스크린샷(600)에서, 레퍼런스 유전자(예를 들어, ALK(anaplastic lymphoma kinase) 유전자)의 염기서열과 일치하는 리드들(610)은 회색 컬러로 표시된다. 하지만, 레퍼런스 유전자의 염기서열과 일치하지 않는 리드들(620)에 대해서는 회색이 아닌 다양한 컬러들로 표시된다. 즉, 다양한 컬러들로 표시되는 리드들(620)은 예를 들어, 불일치 정렬된 PE 리드들 또는 스플릿 리드들에 해당될 가능성이 높다. 따라서, 리드 분석부(도 1의 110)는 이와 같이 레퍼런스 유전자와 다른 염기서열을 갖는 리드들(620)에 대한 데이터를 획득한다.
- [0057] 도 7은 일 실시예에 따른 피검체의 FFPE 샘플로부터 획득된 리드들을 레퍼런스 유전자 데이터와 비교한 IGV 스크린샷을 설명하기 위한 도면이다.
- [0058] 도 7에 도시된 IGV 스크린샷(700)에서 레퍼런스 유전자(예를 들어, ALK 유전자)의 염색체 위치는 도 6에 도시된 IGV 스크린샷(600)에서의 염색체 위치와 유사하다.
- [0059] 하지만, 앞서 설명된 도 6과 달리, IGV 스크린샷(700)은 IGV 스크린샷(도 6의 600)보다 컬러풀하다. 이는 레퍼런스 유전자의 염기서열과 일치하지 않는 리드들(예를 들어, PE 리드들, 스플릿 리드들)이 도 6의 경우보다 많다는 것을 의미한다. 그 이유로서, 도 7의 경우에서 레퍼런스 유전자는 FFPE 샘플로부터 획득된 것이기 때문이다. 생검 샘플의 수명은 짧으므로, FFPE는 생검 샘플의 생화학적 특성을 오랜 기간 유지시키기 위해 필수적인 처리이다. 생검 샘플과 달리 FFPE 샘플은 FFPE 처리에 의한 화학적 변이, 구조적 변이가 발생되므로, 일치하지 않는 리드들이 도 6의 경우보다 많아질 수 밖에 없다. 그 결과, 유전자 분석 결과에 위양성(false positive) 또는 위음성(false negative)의 판단들이 다수 포함될 수 있다. 그러나, 본 실시예들에 따르면, 레퍼런스 유전자, 피검 샘플의 피검 유전자가 생검 샘플로부터 획득된 것이거나 또는 FFPE 샘플로부터 획득된 것일지라도, 위양성(false positive) 또는 위음성(false negative)의 판단들을 제거하거나 또는 줄일 수 있다. 이하에서 보다 계속적으로 설명하도록 한다.

- [0060] 도 8은 일 실시예에 따라 전좌 식별부에서 후보 유전자 쌍들을 추출하여 전좌 유전자를 식별하는 방법의 흐름도이다.
- [0061] 801 단계에서, 전좌 식별부(120)는, 리드 분석부(110)에 의해 획득된 스플릿 리드들 및 불일치 정렬된 PE 리드들에 관한 데이터를 이용하여, 전좌의 가능성이 있는 제 1 후보 유전자 쌍들을 추출한다. 예를 들어, 전좌 식별부(120)는, 앞서 도 6 또는 도 7에서 설명된 IGV 스크린샷(600 또는 700)에서 레퍼런스 유전자의 염기서열과 일치하는 리드들에 관한 데이터를 이용하여, 스플릿 리드들에 매핑된 유전자들, 불일치 정렬된 PE 리드들에 매핑된 유전자들을 다양하게 조합하여 제 1 후보 유전자 쌍들을 추출할 수 있다.
- [0062] 802 단계에서, 전좌 식별부(120)는, 동일한 커버리지에 속한 브레이크 포인트를 갖는 복수의 스플릿 리드들이 정렬된 유전자를 포함하는 제 2 후보 유전자 쌍들을 추출한다. 앞서 설명된 바와 같이, 동일한 커버리지에 속한 브레이크 포인트를 갖는 스플릿 리드들의 개수가 소정 임계값 이상인 경우, 그 스플릿 리드들에 매핑된 유전자는 전좌 유전자의 실제 브레이크 포인트를 갖고 있을 가능성이 보다 높은 것으로 간주될 수 있다. 따라서, 전좌 식별부(120)는, 제 1 후보 유전자 쌍들 중에서, 동일한 커버리지에 속한 브레이크 포인트를 갖는 복수의 스플릿 리드들이 정렬된 유전자를, 제 2 후보 유전자 쌍들로 선별한다. 즉, 제 2 후보 유전자 쌍들에 포함된 유전자들은 제 1 후보 유전자 쌍들에 포함된 유전자들보다, 전좌 유전자일 가능성이 높은 유전자들일 수 있다.
- [0063] 803 단계에서, 전좌 식별부(120)는, 서로 다른 유전자들 간의 퓨전 방향이 5' 엔드부터 3' 엔드이거나, 또는 3' 엔드부터 5' 엔드인 제 3 후보 유전자 쌍들을 추출한다. 제 2 후보 유전자 쌍들에 포함된, 퓨전 유전자(전좌 유전자)인 것으로 예상되는 유전자 쌍들이라 할지라도, 퓨전 방향이 부적절한 경우에는 전좌 유전자에 해당되지 않는다. 따라서, 서로 다른 유전자들의 퓨전 방향이 적절한지, 즉 전좌 식별부(120)는, 제 2 후보 유전자 쌍들에 포함된 유전자 쌍들의 퓨전 방향이 5' 엔드부터 3' 엔드의 방향 또는 3' 엔드부터 5' 엔드의 방향으로 적절하게 결합되었는지를 판단하고, 퓨전 방향이 적절한 유전자 쌍들을 제 3 후보 유전자 쌍들로서 필터링한다. 즉, 제 3 후보 유전자 쌍들에 포함된 유전자들은 제 2 후보 유전자 쌍들에 포함된 유전자들보다, 전좌 유전자일 가능성이 보다 높은 유전자들일 수 있다.
- [0064] 804 단계에서, 전좌 식별부(120)는, 제 3 후보 유전자 쌍들이 추출된 경우, 제 3 후보 유전자 쌍들에 포함된 유전자 쌍들은 전좌 유전자에 해당되는 것으로 식별한다.
- [0065] 도 9는 일 실시예에 따라 스플릿 리드들의 브레이크 포인트들을 이용하여 제 2 후보 유전자 쌍들을 추출하는 것을 설명하기 위한 도면이다.
- [0066] 도 9를 참고하면, 피검 샘플의 어느 유전자(유전자 X)(900)에는 복수의 스플릿 리드들(910)이 매핑될 수 있다. 이 스플릿 리드들(910)이 매핑된 유전자 X(900)는 제 1 후보 유전자 쌍들에 포함될 수 있다. 스플릿 리드들(910)에 대해서는 각각 브레이크 포인트(920)에 대한 데이터가 함께 매핑될 수 있다. 스플릿 리드들(910)은 시퀀싱 에러, 유전자 삽입, 유전자 결실 등과 같은 다양한 원인들로 인해 존재할 수 있으므로, 유전자 X(900)에 다수의 스플릿 리드들(910)이 매핑되어 있다 할지라도, 유전자 X(900)가 바로 전좌 유전자의 일부에 해당되는 것으로 판단할 수 없다.
- [0067] 하지만, 유전자 X(900)에 매핑된 스플릿 리드들(910) 중 동일한 커버리지에 속한 브레이크 포인트(940)를 갖는 다수의 스플릿 리드들(930)이 매핑된 경우에는, 유전자 X(900)에 전좌 유전자의 브레이크 포인트가 존재할 가능성이 높은 것으로 간주될 수 있다. 따라서, 동일한 커버리지에 속한 브레이크 포인트(940)를 갖는 스플릿 리드들(930)의 개수가 소정 임계값 이상인 경우, 유전자 X(900)는 전좌 유전자의 일부에 해당될 가능성이 높은 것으로 식별된다. 즉, 유전자 X(900)는 제 2 후보 유전자 쌍들에 포함된 유전자인 것으로 추출될 수 있다.
- [0068] 한편, 실제 브레이크 포인트를 갖는 스플릿 리드들(930)이 나타내는 브레이크 포인트는 시퀀싱 오차 등의 다양한 원인으로 인해, 완벽하게 동일하지 않을 수 있다. 따라서, 전좌 식별부(120)는, 브레이크 포인트가 동일한 수치인지 여부를 판단하기 보다는, 일정 범위(즉, 커버리지) 내에 브레이크 포인트가 존재하는지 여부를 판단하는 것이 바람직할 수 있다.
- [0069] 도 10은 일 실시예에 따라 퓨전 방향의 적절성을 이용하여 제 3 후보 유전자 쌍들을 추출하는 것을 설명하기 위한 도면이다.
- [0070] 도 10을 참고하면, 염색체 2(1001) 상의 유전자 X(1010)와 염색체 3(1002) 상의 유전자 Y(1020)가 제 2 후보 유전자 쌍들에 포함된 경우, 전좌 식별부(120)는 유전자 X(1010) 및 유전자 Y(1020)의 퓨전 방향을 판단할 수 있다.

- [0071] 염색체 2(1001) 상 및 염색체(1002)에서 퓨전 유전자들(1030 및 1040)은 유전자 X(1010)의 3' 엔드와 유전자 Y(1020)의 5' 엔드와 결합되었으므로, 전좌 식별부(120)는 퓨전 유전자(1030)의 퓨전 방향은 적절하다고 판단할 수 있다.
- [0072] 제 2 후보 유전자 쌍들 중에서, 이와 같이 퓨전 방향이 적절한 퓨전 유전자들(1030 및 1040)은 제 3 후보 유전자 쌍들에 해당되고, 결국 전좌 식별부(120)는 제 3 후보 유전자 쌍들에 포함된 유전자 쌍들은 전좌 유전자인 것으로 판단한다.
- [0073] 앞서 FFPE 샘플의 경우(도 7), 다수의 위양성(false positive)의 리드들이 추출된다 할지라도, 위와 같이 브레이크 포인트의 판단 및 퓨전 방향의 판단을 통해 위양성 판단을 제거하거나 줄이면서 실제 전좌 유전자를 식별해 낼 수 있다.
- [0074] 도 11은 일 실시예에 따라 EML4(echinoderm microtubule-associated protein-like 4) 및 ALK의 전좌 유전자를 식별한 결과를 설명하기 위한 도면이다.
- [0075] 도 11을 참고하면, 앞서 설명된 전좌 유전자의 분석 방법들을 통해 식별된 EML4-ALK의 전좌 유전자에 대한 데이터가 도시되어 있다. 우측의 IGV 스크린샷(1101)에는 EML4에 매핑된 리드들이 표시되어 있고, 좌측의 IGV 스크린샷(1102)에는 ALK에 매핑된 리드들이 표시되어 있다. EML4에 매핑된 리드들은 42536701~42559688의 브레이크 포인트 커버리지에서 스플릿되었고, ALK에 매핑된 리드들은 29415639~29446500의 브레이크 포인트 커버리지에서 스플릿되었음을 알 수 있다. 그리고, EML4-ALK의 전좌 유전자를 식별하기 위한 서포팅 리드들은 39개가 이용되었다. 도 11은 본 실시예들에서 설명된 유전자 분석 방법을 실제 환자의 피검 샘플에 적용하여 전좌 유전자의 식별 결과를 검증한 시뮬레이션 결과일 뿐이므로, 본 실시예들은 도 11에 의해 제한되지 않는다.
- [0076] 도 12는 일 실시예에 따라 유전자를 분석하는 방법의 흐름도이다. 도 12를 참고하면, 유전자 분석 방법은 앞선 도면들에서 설명된 유전자 분석 장치(10)에서 시계열적으로 처리되는 단계들을 포함한다. 따라서, 이하 생략된 내용이라 하더라도 앞선 도면들에서 설명되었던 내용들은 도 12의 유전자 분석 방법에도 적용될 수 있다.
- [0077] 1201 단계에서, 리드 분석부(110)는 피검 샘플의 차세대 시퀀싱(NGS) 데이터로부터, 스플릿 리드들 및 불일치 정렬된 PE 리드들에 관한 데이터를 획득한다.
- [0078] 1202 단계에서, 전좌 식별부(120)는 스플릿 리드들 및 불일치 정렬된 PE 리드들을 이용하여 피검 샘플의 염색체 내 전좌의 가능성이 있는 제 1 후보 유전자 쌍들을 추출한다.
- [0079] 1203 단계에서, 전좌 식별부(120)는 스플릿 리드들이 나타내는 브레이크 포인트들 및 제 1 후보 유전자 쌍들의 퓨전 방향에 기초하여, 제 1 후보 유전자 쌍들 중에서 전좌 유전자를 식별한다.
- [0080] 도 13은 일 실시예에 따른 컴퓨팅 장치의 하드웨어 구성들을 도시한 블록도이다.
- [0081] 도 13을 참고하면, 컴퓨팅 장치(1)는 유전자 분석 장치(프로세서)(10), 데이터 인터페이스(11) 및 메모리(12)를 포함한다. 한편, 도 13에 도시된 컴퓨팅 장치(1)는 본 실시예의 특징이 흐려지는 것을 방지하기 위하여 본 실시예에 관련된 구성요소들만이 도시되어 있을 뿐이므로, 도 13에 도시된 구성요소들 외에 다른 범용적인 구성요소들이 더 포함될 수 있다.
- [0082] 데이터 인터페이스(11)는 앞서 도 1에서 설명된, 정상인 집단의 레퍼런스 유전자 데이터(20) 및 피검체의 피검 유전자 데이터(30)를 수신한다. 즉, 데이터 인터페이스(11)는 컴퓨팅 장치(1)가 외부의 다른 디바이스들과 통신하기 위한 유/무선 네트워크 인터페이스의 하드웨어로 구현될 수 있다. 데이터 인터페이스(11)는 수신된 유전자 데이터(20 및 30)를 유전자 분석 장치(프로세서)(10)로 전송한다.
- [0083] 데이터 인터페이스(11)는 피검체의 피검 유전자를 시퀀싱하기 위한 외부의 차세대 시퀀싱 장치, 마이크로어레이 등으로부터 피검체의 피검 유전자 데이터(30)를 수신할 수 있다.
- [0084] 메모리(12)는 컴퓨팅 장치(1) 내에서 처리될 데이터들 및 처리가 완료된 결과들을 저장하기 위한 하드웨어로서, RAM(random access memory), ROM(read only memory) 등의 메모리 칩들 또는 HDD(hard disk drive), SSD(solid state drive) 등의 스토리지를 포함한다. 즉, 메모리(12)는 데이터 인터페이스(11)에 의해 수신된 유전자 데이터(20 및 30)를 저장할 수 있고, 유전자 분석 장치(프로세서)(10)에 의해 처리된 제 1 내지 제 3 후보 유전자 쌍들에 관한 데이터, 식별된 전좌 유전자에 대한 데이터 등도 저장할 수 있다.
- [0085] 유전자 분석 장치(프로세서)(10)는 하나 이상의 프로세싱 유닛들로 구현된 모듈로서, 다수의 논리 게이트들의 어레이를 갖는 마이크로프로세서와 이 마이크로프로세서에서 실행될 수 있는 프로그램이 저장된 메모리 모듈의

조합으로 구현될 수도 있다. 유전자 분석 장치(프로세서)(10)는 응용 프로그램의 모듈 형태로 구현될 수도 있다. 유전자 분석 장치(프로세서)(10)는 앞서 도 1 내지 도 12에서 설명된 유전자 분석을 처리하는 하드웨어 장치이다.

[0086] 유전자 분석 장치(프로세서)(10)에 의해 식별된 전좌 유전자에 대한 정보는 데이터 인터페이스(11)를 통해 외부의 다른 디바이스, 예를 들어 디스플레이 디바이스, 다른 컴퓨팅 장치 등으로 전송되거나, 또는 외부 네트워크, 예를 들어 인터넷, 공개 데이터베이스(DB) 서버 상으로 전송될 수 있다.

[0087] 앞서 설명된 본 실시예들에 따르면, 피검체(예를 들어, 암 환자)의 암 조직으로부터 전좌 유전자를 검출할 수 있다. 나아가서, 피검체로부터 획득된 암 조직(피검 샘플)의 유전자들(피검 유전자들)이 FFPE 처리에 의하여 화학적으로 약간 손상된다 할지라도, 전좌 유전자를 정확하게 판단할 수 있다.

[0088] 본 실시예들에 따른 장치는 프로세서, 프로그램 데이터를 저장하고 실행하는 메모리, 디스크 드라이브와 같은 영구 저장부(permanent storage), 외부 장치와 통신하는 통신 포트, 터치 패널, 키(key), 버튼 등과 같은 사용자 인터페이스 장치 등을 포함할 수 있다. 소프트웨어 모듈 또는 알고리즘으로 구현되는 방법들은 상기 프로세서 상에서 실행 가능한 컴퓨터가 읽을 수 있는 코드들 또는 프로그램 명령들로서 컴퓨터가 읽을 수 있는 기록 매체 상에 저장될 수 있다. 여기서 컴퓨터가 읽을 수 있는 기록 매체로 마그네틱 저장 매체(예컨대, ROM(read-only memory), RAM(random-access memory), 플로피 디스크, 하드 디스크 등) 및 광학적 판독 매체(예컨대, 시디롬(CD-ROM), 디브이디(DVD: Digital Versatile Disc)) 등이 있다. 컴퓨터가 읽을 수 있는 기록 매체는 네트워크로 연결된 컴퓨터 시스템들에 분산되어, 분산 방식으로 컴퓨터가 판독 가능한 코드가 저장되고 실행될 수 있다. 매체는 컴퓨터에 의해 판독가능하며, 메모리에 저장되고, 프로세서에서 실행될 수 있다.

[0089] 본 실시예는 기능적인 블록 구성들 및 다양한 처리 단계들로 나타내어질 수 있다. 이러한 기능 블록들은 특정 기능들을 실행하는 다양한 개수의 하드웨어 또는/및 소프트웨어 구성들로 구현될 수 있다. 예를 들어, 실시 예는 하나 이상의 마이크로프로세서들의 제어 또는 다른 제어 장치들에 의해서 다양한 기능들을 실행할 수 있는, 메모리, 프로세싱, 로직(logic), 룩 업 테이블(look-up table) 등과 같은 직접 회로 구성들을 채용할 수 있다. 구성 요소들이 소프트웨어 프로그래밍 또는 소프트웨어 요소들로 실행될 수 있는 것과 유사하게, 본 실시예는 데이터 구조, 프로세스들, 루틴들 또는 다른 프로그래밍 구성들의 조합으로 구현되는 다양한 알고리즘을 포함하여, C, C++, 자바(Java), 어셈블리(assembler) 등과 같은 프로그래밍 또는 스크립팅 언어로 구현될 수 있다. 기능적인 측면들은 하나 이상의 프로세서들에서 실행되는 알고리즘으로 구현될 수 있다. 또한, 본 실시예는 전자적인 환경 설정, 신호 처리, 및/또는 데이터 처리 등을 위하여 종래 기술을 채용할 수 있다. “매커니즘”, “요소”, “수단”, “구성”과 같은 용어는 넓게 사용될 수 있으며, 기계적이고 물리적인 구성들로서 한정되는 것은 아니다. 상기 용어는 프로세서 등과 연계하여 소프트웨어의 일련의 처리들(routines)의 의미를 포함할 수 있다.

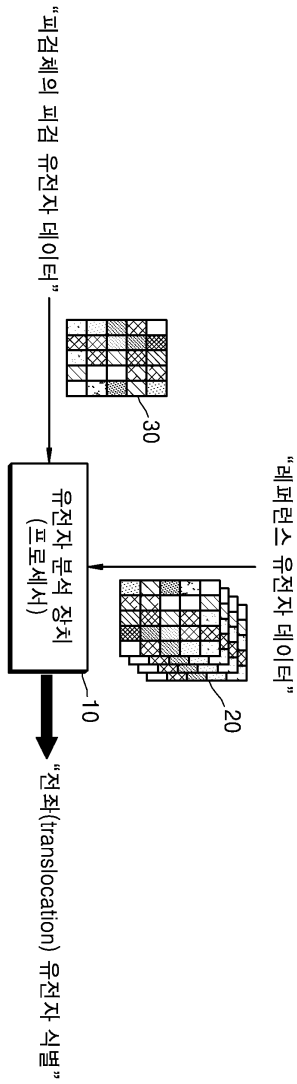
[0090] 본 실시예에서 설명하는 특정 실행들은 예시들로서, 어떠한 방법으로도 기술적 범위를 한정하는 것은 아니다. 명세서의 간결함을 위하여, 종래 전자적인 구성들, 제어 시스템들, 소프트웨어, 상기 시스템들의 다른 기능적인 측면들의 기재는 생략될 수 있다. 또한, 도면에 도시된 구성 요소들 간의 선들의 연결 또는 연결 부재들은 기능적인 연결 및/또는 물리적 또는 회로적 연결들을 예시적으로 나타낸 것으로서, 실제 장치에서는 대체 가능하거나 추가의 다양한 기능적인 연결, 물리적인 연결, 또는 회로 연결들로서 나타내어질 수 있다.

[0091] 본 명세서(특히 특허청구범위에서)에서 “상기”의 용어 및 이와 유사한 지시 용어의 사용은 단수 및 복수 모두에 해당하는 것일 수 있다. 또한, 범위(range)를 기재한 경우 상기 범위에 속하는 개별적인 값을 포함하는 것으로서(이에 반하는 기재가 없다면), 상세한 설명에 상기 범위를 구성하는 각 개별적인 값을 기재한 것과 같다. 마지막으로, 방법을 구성하는 단계들에 대하여 명백하게 순서를 기재하거나 반하는 기재가 없다면, 상기 단계들은 적당한 순서로 행해질 수 있다. 반드시 상기 단계들의 기재 순서에 한정되는 것은 아니다.

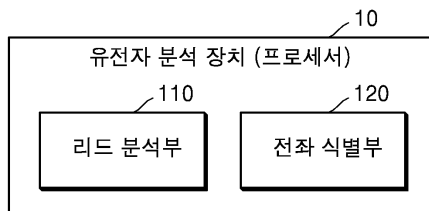
[0092] 이제까지 본 발명에 대하여 그 바람직한 실시예들을 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

도면

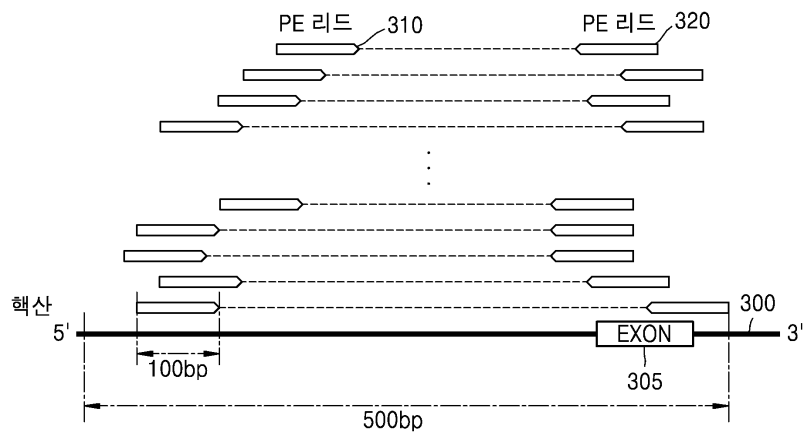
도면1



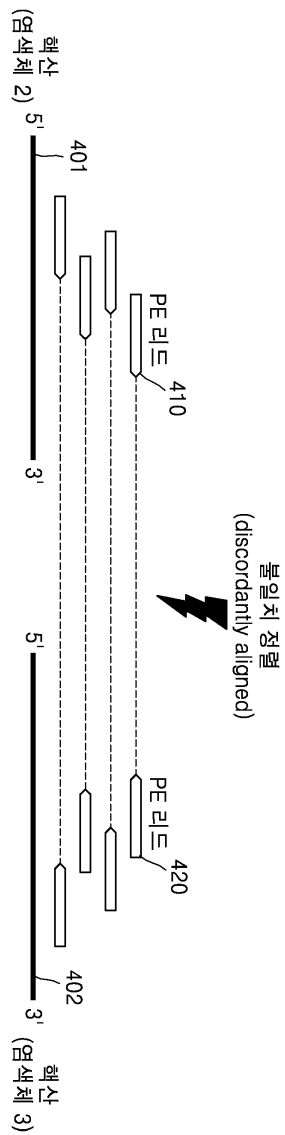
도면2



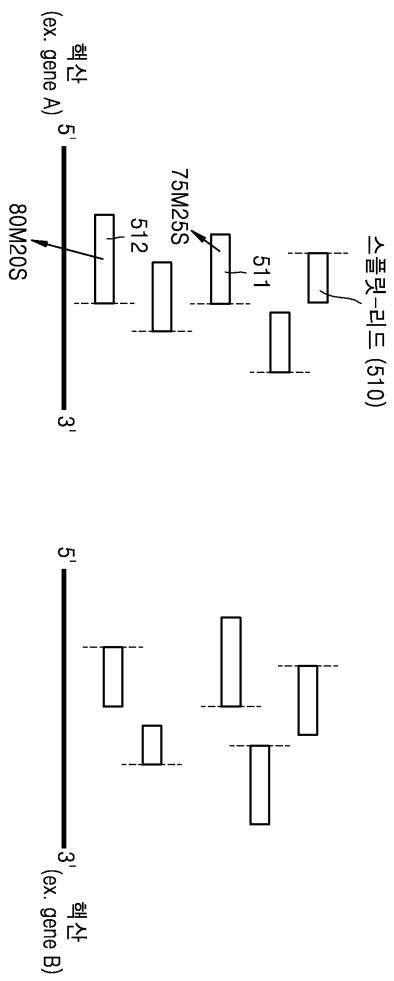
도면3



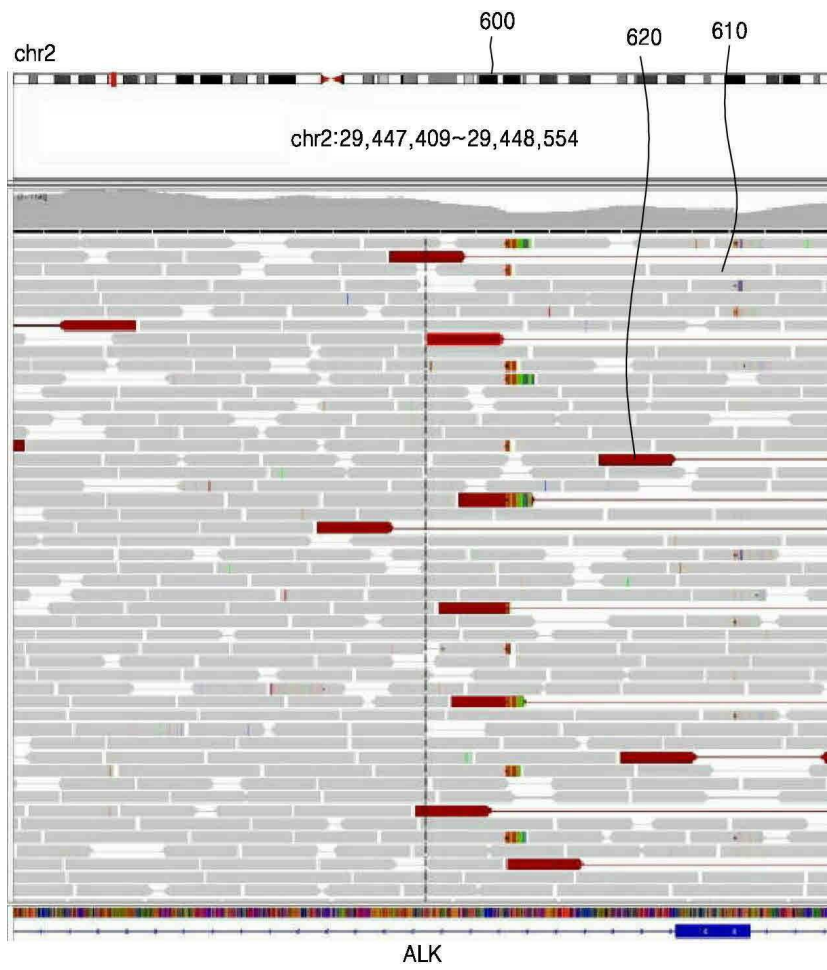
도면4



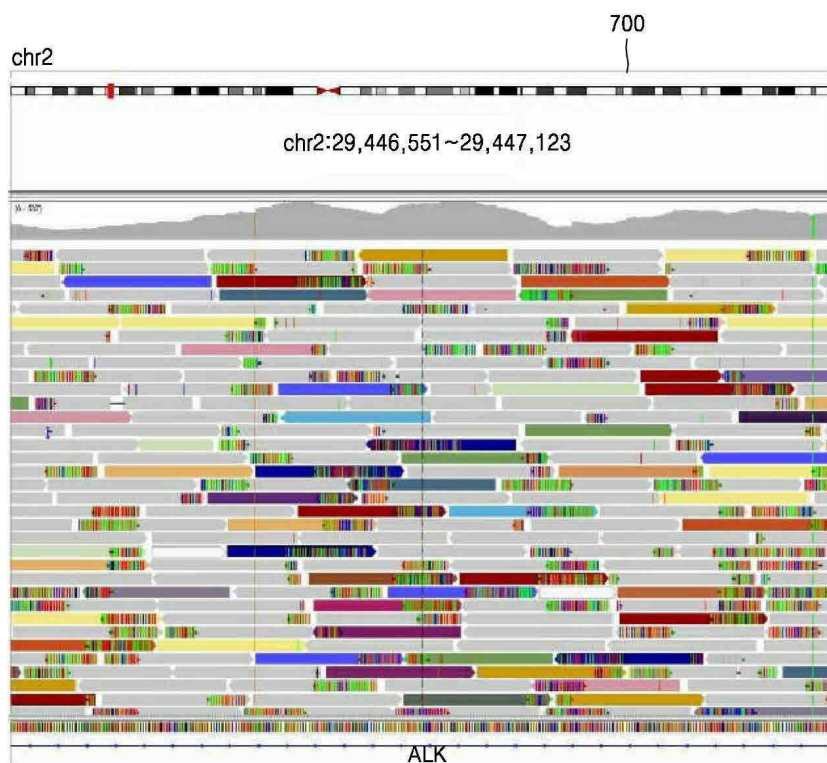
도면5



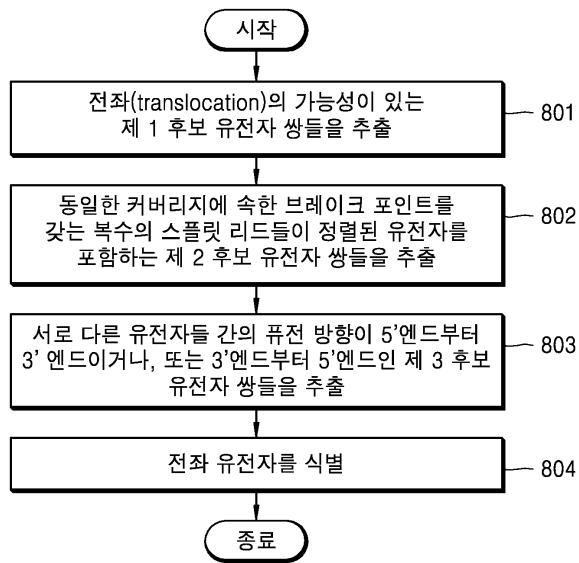
도면6



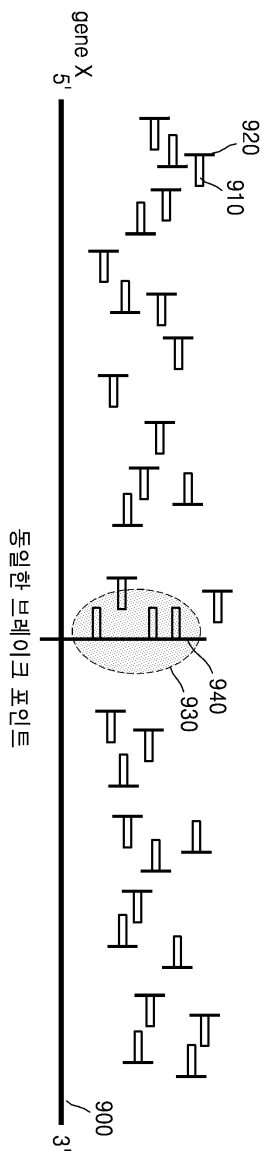
도면7



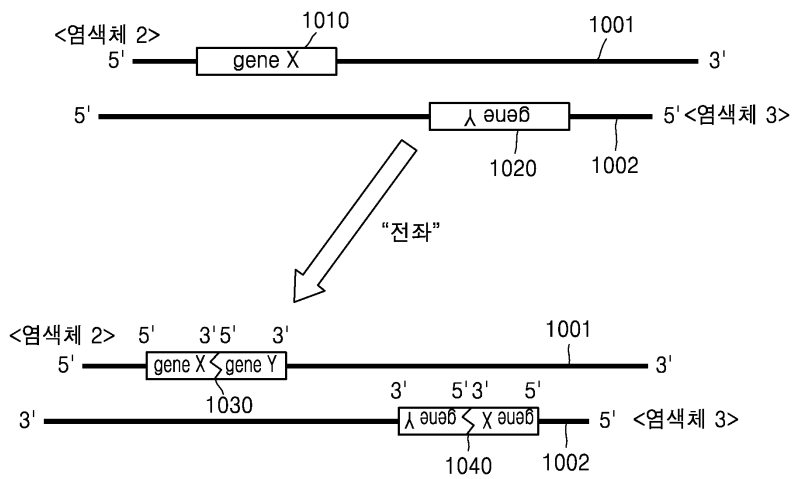
도면8



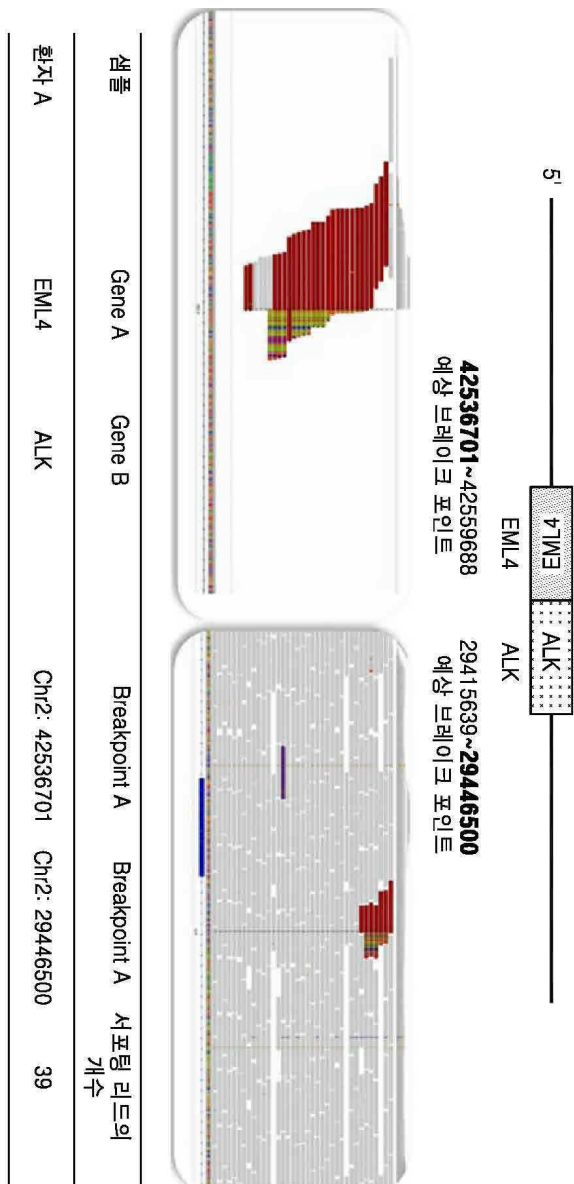
도면9



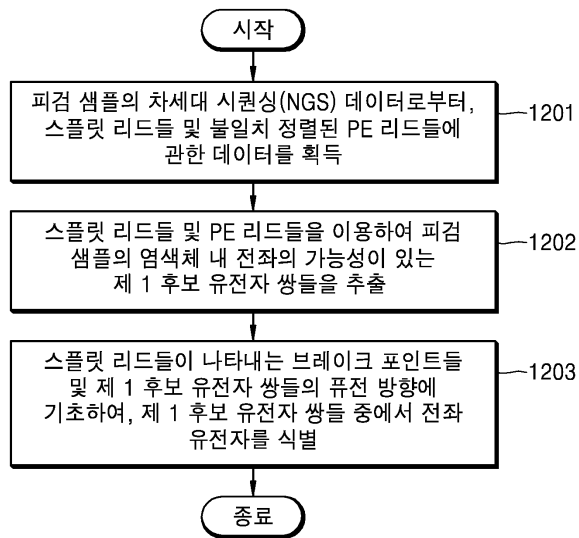
도면10



도면11



도면12



도면13

