

명세서

청구범위

청구항 1

목적대상의 시료를 수득하는 수득부;

시료의 유전 정보를 분석하는 제 1 연산부;

상기 제 1 연산부의 결과와 목적하는 기준 염색체의 서열을 대조하는 제 2 연산부;

상기 제 2 연산부의 대조결과로부터 염색체 구조변이를 검출하는 검출부; 및

상기 검출부로부터 수득된 염색체 구조변이를 절단점(breaking point)를 중심으로 노드(node) 및 엣지(edge)로 정의하여 출력하는 출력부를 포함하고,

상기 엣지는 절단점 및 오더(order)를 포함하고,

상기 오더는 목적대상의 시료의 유전정보를 증폭한 서열 단편을 정렬했을 때, 매칭 순서에 따라 서로 다른 오더 값으로 표시되고,

상기 오더값에 따라 화살표 표시 여부 또는 화살표 방향이 결정되는 염색체 구조변이 검출 장치.

청구항 2

제 1항에 있어서,

목적대상의 시료는 인간의 혈액임을 특징으로 하는 염색체 구조변이 검출 장치.

청구항 3

제 1항에 있어서,

노드는 염색체의 번호 및 구조변이 위치를 포함하는 염색체 구조변이 검출 장치.

청구항 4

삭제

청구항 5

제 1항에 있어서,

엣지는 1개 내지 3개의 절단점 중 어느 하나 이상을 포함하는 염색체 구조변이 검출 장치.

청구항 6

제 1항에 있어서,

오더는 제 1오더 및 제 2 오더를 포함하는 염색체 구조변이 검출 장치.

청구항 7

제 1항에 있어서,

오더는 0 및 1로 표시되는 염색체 구조변이 검출 장치.

청구항 8

제 6항에 있어서,

제 1 오더 및 제 2 오더의 조합으로 방향성이 생기는 염색체 구조변이 검출 장치.

청구항 9

제 8항에 있어서,

방향성은 양쪽성, 우측, 좌측 및 무방향성 중 어느 하나인 염색체 구조변이 검출 장치.

청구항 10

제 1항에 있어서,

절단점은 절단 리드 및 비정형 리드 쌍 정보를 이용하여 구조 변이의 절단점을 검출하는 염색체 구조변이 검출 장치.

청구항 11

제 1항에 있어서,

염색체 구조변이란 염색체 상에서 유전자가 결실, 삽입, 역위 및 염색체 전위 중 어느 하나인 염색체 구조변이 검출 장치.

청구항 12

삭제

청구항 13

목적대상의 시료를 수득하는 단계;

시료의 유전 정보를 분석하는 단계;

상기 분석하는 단계로부터 수득된 결과와 목적하는 기준 염색체의 서열을 대조하는 단계;

상기 대조하는 단계로부터 수득된 결과로부터 염색체 구조변이를 절단점(breaking point)를 중심으로 노드(node) 및 엣지(edge)로 표현하는 단계를 포함하고,

상기 엣지는 절단점 및 오더(order)를 포함하고,

상기 오더는 목적대상의 시료의 유전정보를 증폭한 서열 단편을 정렬했을 때, 매칭 순서에 따라 서로 다른 오더값으로 표시되고,

상기 오더값에 따라 화살표 표시 여부 또는 화살표 방향이 결정되는 염색체 구조변이에 관한 정보를 제공하는 방법.

발명의 설명

기술 분야

[0001] 본 발명은 염색체 구조변이 검출 장치 및 이의 제조방법에 관한 것이다.

배경 기술

[0002] 인간 유전체 프로젝트가 완료된 후 인간의 DNA 염기서열이 해독되고 이로부터 인간 유전자의 다양한 기능들이 밝혀지고 있다. 특히, 다양한 유전자 변이들이 발견되어 이것들이 인간의 형질의 차이를 일으킬 뿐만 아니라, 특정 질병의 원인으로 작용될 수 있음이 밝혀짐에 따라 인간 유전체 분석 연구는 점점 더 가속화 되어가고

[0003] 있다. 하지만, 인간 유전체에서 발생할 수 있는 방대한 유전적 변이 중 어떠한 변이가 실질적으로 병인이 될 수 있는 것인가를 밝혀내는 데에는 어려움이 있을 수 밖에 없다.

[0004] 차세대 시퀀싱 기술(NGS, Next Generation Sequencing)이 발달함에 따라 개별 인간의 전체 유전체의 염기 서열 해독이 가능하게 되었고, 질병군과 정상군의 염기서열 및 변이 비교 분석 방법을 통하여 질병 특이적 유전자 변

이를 추출하는 것도 가능하게 되었다. 또한, 형질에 관련된 마커를 선별하고 뉴클레오티드 수준에 기존의 변이를 확인하며 표적 뉴클레오티드 교환에 의해 상기 마커의 불변 부위의 위치들에서 하나 또는 그 이상의 뉴클레오티드의 도입으로 선별가능한 마커를 도입하여 고유한 분자 마커들의 생성을 위한 방법을 활용하기도 하였다 (KR 제 10-2011-0094268호 참조). 그러나 다양한 구조 변이 타입을 정확하게 유형화하고 도식화하는 방법이 없어서 이를 대량으로 자료관리 및 새로운 정보 수득이 어려움이 있었다. 이러한 문제를 해결하고자 발명자는 본 발명을 완성하였다.

발명의 내용

해결하려는 과제

[0005] 본 발명은 염색체의 변이 정보를 유형화하고, 이를 대량으로 정보 구축하기 위하여 절단점을 기초로하여 노드 (node) 및 엣지(edge)로 정의하여 표현하는 염색체 구조변이 검출 장치 및 이의 제조방법을 제공하는 것을 목적으로 한다.

[0006] 또한, 본 발명은 염색체 변이 정보의 절단점을 기초로 하여 노드(node) 및 엣지(edge)로 정의하여 출력하는 염색체 구조변이에 관한 정보를 제공하는 것을 목적으로 한다.

과제의 해결 수단

[0007] 일 구체예에서, 목적대상의 시료를 수득하는 수득부; 시료의 유전 정보를 분석하는 제 1 연산부; 상기 제 1 연산부의 결과와 목적하는 기준 염색체의 서열을 대조하는 제 2 연산부; 상기 제 2 연산부의 대조결과로부터 염색체 구조변이를 검출하는 검출부; 및 상기 검출부로부터 수득된 염색체 구조변이를 절단점(breaking point)를 중심으로 노드(node) 및 엣지(edge)로 정의하여 출력하는 출력부를 포함하는 염색체 구조변이 검출 장치를 제공한다. 상기 구체예에서, 목적대상의 시료는 인간의 혈액임을 특징으로 하는 염색체 구조변이 검출 장치를 제공하고, 노드는 염색체의 번호 및 구조변이 위치를 포함하는 염색체 구조변이 검출 장치를 제공하며, 엣지는 절단점 및 오더(order)를 포함하는 염색체 구조변이 검출 장치를 제공하며, 엣지는 1개 내지 3개의 절단점 중 어느 하나 이상을 포함하는 염색체 구조변이 검출 장치를 제공하며, 오더는 제 1오더 및 제 2 오더를 포함하는 염색체 구조변이 검출 장치를 제공하며, 오더는 0 및 1로 표시되는 염색체 구조변이 검출 장치를 제공하며, 제 1 오더 및 제 2 오더의 조합으로 방향성이 생기는 염색체 구조변이 검출 장치를 제공하며, 방향성은 양쪽성, 우측, 좌측 및 무방향성 중 어느 하나인 염색체 구조변이 검출 장치를 제공하며, 절단점은 절단 리드 및 비정형 리드 쌍 정보를 이용하여 구조 변이의 절단점을 검출하는 염색체 구조변이 검출 장치를 제공하며, 염색체 구조 변이란 염색체 상에서 유전자가 결실, 삽입, 역위 및 염색체 전위 중 어느 하나인 염색체 구조변이 검출 장치를 제공한다.

[0008] 일 구체예에서, 목적대상의 시료를 수득하는 수득부; 시료의 유전 정보를 분석하는 제 1 연산부; 상기 제 1 연산부의 결과와 목적하는 기준 염색체의 서열을 대조하는 제 2 연산부; 상기 제 2 연산부의 대조결과로부터 염색체 구조변이를 검출하는 검출부; 및 상기 검출부로부터 수득된 염색체 구조변이를 절단점(breaking point)를 중심으로 노드(node) 및 엣지(edge)로 정의하여 출력하는 출력부를 결합하여 염색체 구조변이 검출 장치를 제조하는 방법을 제공한다.

[0009] 일 구체예에서, 목적대상의 시료를 수득하는 단계; 시료의 유전 정보를 분석하는 단계; 상기 분석하는 단계로부터 수득된 결과와 목적하는 기준 염색체의 서열을 대조하는 단계; 상기 대조하는 단계로부터 수득된 결과로부터 염색체 구조변이를 절단점(breaking point)를 중심으로 노드(node) 및 엣지(edge)로 표현하는 단계를 포함하는 염색체 구조변이에 관한 정보를 제공하는 방법을 제공한다.

[0010] 본 발명에서, “절단점”이란 염색체 내에서 유전자가 삽입, 결실, 전좌, 역위 등 유전 변이로 인하여 정상인의 염색체와 달리 절단이 발생한 지점을 말하며, 리드 맵핑을 한 후에 절단 리드 및 비정형 리드 쌍 정보를 이용하여 구조 변이의 절단지점을 찾을 수 있다.

- [0011] 본 발명에서, “노드(node)”란 절단점의 염색체 번호 및 구조변이 위치에 관한 정보를 포함하는 것으로 다수의 절단점을 포함할 수 있다.
- [0012] 본 발명에서, “엣지(edge)”이란 절단점간의 관계를 지칭하는 말로, 컨티그(contig), 오더(order) 또는 링크(link)로 표현되기 하는데, 절단점 간의 방향성을 결정한다.
- [0013] 본 발명에서, “오더(order)”이란 절단점 간의 방향성을 구체적인 값으로 표현한 것으로 하나의 엣지가 염색체 상 서로 다른 위치에 표시되고, 상호 배타적으로 매칭과 미매칭(소프트 클리핑:soft clipped)상태가 나타날 경우 매칭-미매칭 순서 상태를 오더 1으로 표시하고, 미매칭-매칭 순서 상태를 오더 0로 표시한다(도 8 참조).
- [0014] 본 발명에서, “염색체 구조변이”이란 유전자가 이에 한정하지는 않지만 삽입, 결실, 전좌, 역위 등 유전적인 변동으로 인하여 정상인의 염색체와 유전 정보가 달라지는 것을 말한다.
- [0015] 본 발명의 “그래프 작성”은 절단점과 엣지의 오더값에 따라서 작성되고, 절단점과 절단점을 오더값 1에서 오더값 0으로 연결한다(도 9 및 도 1 내지 7 참조).
- [0016] 본 발명의 “리더(read)”란 목적하는 시료의 유전정보를 증폭한 서열 단편으로, 임의적으로 리더 A, 리더 B와 같이 구분하여 표현할 수 있다.

발명의 효과

- [0017] 본 발명의 장치 및 정보 제공방법을 통하여, 목적하는 시료의 유전정보의 변이를 보다 정확하게 정의하여 자료를 구축함으로써, 변이에 관한 정보를 유형화할 수 있고, 대량 자료를 용이하게 구축할 수 있다. 또한, 절단점간의 관계를 용이하게 파악할 수 있고, 종래의 발견되지 못한 절단점을 검출할 수 있을 뿐만 아니라, 이의 패턴에 관한 정보를 제공할 수 있다.

도면의 간단한 설명

- [0018] 도 1은 절단점에 신규한 서열이 삽입된 경우 매칭 상태와 절단점 및 엣지를 이용하여 그래프로 표현한 그림이다.
- 도2는 절단점에 서열이 결실된 경우 매칭 상태와 절단점 및 엣지를 이용하여 그래프로 표현한 그림이다.
- 도3는 절단점에 신규한 서열이 거꾸로 삽입된 경우 매칭상태와 절단점 및 엣지를 이용하여 그래프로 표현한 그림이다.
- 도4는 절단점에 동일한 염색체 내에서 서열이 이동하는 경우(Internal Translocation) 매칭상태와 절단점 및 엣지를 이용하여 그래프로 표현한 그림이다.
- 도5 내지 7은 절단점에 상이한 염색체간에 서열이 이동하는 경우(Cross Translocation) 매칭상태와 절단점 및 엣지를 이용하여 그래프로 표현한 그림이다.
- 도 8은 절단점과 매칭오더와 관계를 보여주는 그림이다.
- 도 9는 절단점과 매칭오더를 통하여 작성된 그래프의 일 구체예이다.

발명을 실시하기 위한 구체적인 내용

[0019] 이하, 본 발명의 구성요소와 기술적 특징을 다음의 실시예들을 통하여 보다 상세하게 설명하고자 한다. 그러나 하기 실시예들은 본 발명의 내용을 예시하는 것일 뿐 발명의 범위가 실시예에 의해 한정되는 것은 아니다.

[0020] 본원에 기재된 다양한 구체예가 도면을 참조로 기재된다. 하기 설명에서, 본 발명의 완전한 이해를 위해서, 다양한 특이적 상세사항, 예컨대, 특이적 형태, 조성물, 및 공정 등이 기재되어 있다. 그러나, 특정의 구체예는 이들 특이적 상세 사항 중 하나 이상 없이, 또는 다른 공지된 방법 및 형태와 함께 실행될 수 있다. 다른 예에서, 공지된 공정 및 제조 기술은 본 발명을 불필요하게 모호하게 하지 않게 하기 위해서, 특정의 상세사항으로 기재되지 않는다. "한 가지 구체예" 또는 "구체예"에 대한 본 명세서 전체를 통한 참조는 구체예와 결부되어 기재된 특별한 특징, 형태, 조성 또는 특성이 본 발명의 하나 이상의 구체예에 포함됨을 의미한다. 따라서, 본 명세서 전체에 걸친 다양한 위치에서 표현 "한 가지 구체예에서" 또는 "구체예"의 상황은 반드시 본 발명의 동일한 구체예를 나타내지는 않는다. 추가로, 특별한 특징, 형태, 조성, 또는 특성은 하나 이상의 구체예에서 어떠한 적합한 방법으로 조합될 수 있다.

[0021] 염색체 내에서 유전변이가 일어난 경우에, 이를 절단점 및 매칭 오더값인 0 및 1로 유전변이의 삽입, 결실, 역위, 내부 전좌 및 외부 전좌를 그래프로 표현할 수 있다.

[0022] 염색체 상에 신규한 서열이 삽입된 경우 1개의 절단점이 발생하고, 기준 서열과 목적하는 서열인 리드를 기준 서열에 대응시켜서 매칭되는 상태에 따라서, 0 또는 1의 값을 줄 수 있으며, 이를 기초로 도 1의 우측 그래프와 같이 표현할 수 있다.

[0023] 염색체 상에서 서열이 결실된 경우 2개의 절단점이 발생하고, 2개의 절단점에서는 기준 서열과 시료의 리드 서열이 매칭이 되지 않게 되며 이에 따라서 0 또는 1의 값이 부여되고, 이를 2개의 절단점과 매칭상태의 값을 도 2의 우측 그래프와 같이 표현할 수 있다.

[0024] 염색체 상에서 신규한 서열이 거꾸로 삽입된 경우, 2개의 절단점이 발생되고, 절단점간의 서열에 대하여 매칭은 되지만 2개의 절단점 외부에서는 매칭이 안되는 리드와 2개의 절단점 외부에서는 매칭이 되지만 절단점 간에는 매칭이 안되는 리드가 존재하게 되며, 이를 도 3의 우측과 같이 표현할 수 있고, 이 경우 2개의 엷지를 가지게 된다.

[0025] 동일한 염색체 상에서 서열이 이동된 경우 3개의 절단점이 발생하고, 2개의 엷지를 가지게 된다(도 4 참조).

[0026]

[0027] 상이한 염색체상에서 서열이 이동된 경우 2개의 절단점이 발생하고, 2개의 엷지를 가지게 된다(도 5 내지 도 6 참조).

[0028] 일 구체예에서, 유전자가 삽입된 경우 도 1의 우측 그림과 같이, 절단점 및 이를 연결한 선으로 표현되고, 하기와 같이 전산상 DB에 저장될 수 있다

표 1

노드	BP_att1	BP_att2
BP1	Chr1	123456

[0029]

표 2

[0030]

엣지	BP1	오더 1
리드A	BP1	1
리드B	BP1	0

[0031] 이때,

[0032] BP1은 노드상의 임의의 절단점이고,

[0033] BP_att1 및 BP_att1는 염색체 번호 및 구조변이 위치에 관한 정보이며,

[0034] 리드 A 및 B는 목적하는 시료의 유전정보를 증폭한 서열 단편이며,

[0035] 오더값 1은 절단점 간의 방향성을 구체적인 값으로, 예를 들어 값이 1인 경우 매칭-미매칭의 순서이고, 오더값 0은 미매칭-매칭의 순서이다.

[0036] 일 구체예에서, 결실은 도 2의 우측 그림과 같이, 절단점 2개 및 이를 연결한 선으로 표현되고, 하기와 같이 전산상 DB에 저장될 수 있다

표 3

[0037]

노드	BP_att1	BP_att2
BP1	Chr1	123456
BP2	Chr1	4567890

표 4

[0038]

엣지	BP1	BP2	오더 1	오더2
리드A	BP1	BP2	1	0

[0039] 이때,

[0040] BP1 및 BP2는 노드상의 임의의 절단점이고,

[0041] BP_att1 및 BP_att1는 염색체 번호 및 구조변이 위치에 관한 정보이며,

[0042] 리드 A는 목적하는 시료의 유전정보를 증폭한 서열 단편이며,

[0043] 오더값 1은 절단점 간의 방향성을 구체적인 값으로, 예를 들어 값이 1인 경우 매칭-미매칭의 순서이고, 오더값 0은 미매칭-매칭의 순서이다.

[0044] 일 구체예에서, 역위는 도 3의 우측 그림과 같이, 절단점 2개 및 이를 연결한 선으로 표현되고, 하기와 같이 전산상 DB에 저장될 수 있다

표 5

[0045]

노드	BP_att1	BP_att2
BP1	Chr1	123456
BP2	Chr1	4567890

표 6

[0046]

엣지	BP1	BP2	오더 1	오더2
리드A	BP1	BP2	1	1

리드B	BP1	BP2	0	0
-----	-----	-----	---	---

- [0047] 이때,
- [0048] BP1 및 BP2는 노드상의 임의의 절단점이고,
- [0049] BP_att1 및 BP_att1는 염색체 번호 및 구조변이 위치에 관한 정보이며,
- [0050] 리드 A 및 B는 목적하는 시료의 유전정보를 증폭한 서열 단편이며,
- [0051] 오더값 1은 절단점 간의 방향성을 구체적인 값으로, 예를 들어 값이 1인 경우 매칭-미매칭의 순서이고, 오더값 0은 미매칭-매칭의 순서이다.

[0052] 일 구체예에서, 동일한 염색체 내에서 서열이 이동하는 경우 도 4의 우측 상단 그림과 같이, 절단점 3개 및 이를 연결한 선으로 표현되고, 하기와 같이 전산상 DB에 저장될 수 있다

표 7

노드	BP_att1	BP_att2
BP1	Chr1	111111
BP2	Chr1	200000
BP3	Chr1	200100

표 8

엣지	BP1	BP2	오더1	오더2
리드A	BP1	BP2	1	0
리드B	BP1	BP3	0	1

- [0055] 이때,
- [0056] BP1, BP2 및 BP3는 노드상의 임의의 절단점이고,
- [0057] BP_att1 및 BP_att1는 염색체 번호 및 구조변이 위치에 관한 정보이며,
- [0058] 리드 A 및 B는 목적하는 시료의 유전정보를 증폭한 서열 단편이며,
- [0059] 오더값 1은 절단점 간의 방향성을 구체적인 값으로, 예를 들어 값이 1인 경우 매칭-미매칭의 순서이고, 오더값 0은 미매칭-매칭의 순서이다.

[0060] 일 구체예에서, 동일한 염색체 내에서 서열이 역위로 이동하는 경우 도 4의 우측 하단 그림과 같이, 절단점 3개 및 이를 연결한 선으로 표현되고, 하기와 같이 전산상 DB에 저장될 수 있다

표 9

노드	BP_att1	BP_att2
BP1	Chr1	111111
BP2	Chr1	200000
BP3	Chr1	200100

표 10

엣지	BP1	BP2	오더 1	오더2
리드A	BP1	BP3	1	1
리드B	BP1	BP2	0	0

[0063] 이때,
 [0064] BP1, BP2 및 BP3는 노드상의 임의의 절단점이고,
 [0065] BP_att1 및 BP_att1는 염색체 번호 및 구조변이 위치에 관한 정보이며,
 [0066] 리드 A 및 B는 목적하는 시료의 유전정보를 증폭한 서열 단편이며,
 [0067] 오더값 1은 절단점 간의 방향성을 구체적인 값으로, 예를 들어 값이 1인 경우 매칭-미매칭의 순서이고, 오더값 0은 미매칭-매칭의 순서이다.

[0068] 일 구체예에서, 상이한 염색체 내에서 서열이 이동하는 경우 도 5의 우측 그림과 같이, 절단점 2개 및 이를 연결한 선으로 표현되고, 하기와 같이 전산상 DB에 저장될 수 있다

표 11

노드	BP_att1	BP_att2
BP1	Chr1	123456
BP2	Chr2	456789

표 12

엣지	BP1	BP2	오더 1	오더2
리드A	BP1	BP2	1	1
리드B	BP1	BP2	0	0

[0071] 이때,
 [0072] BP1 및 BP2는 노드상의 임의의 절단점이고,
 [0073] BP_att1 및 BP_att1는 염색체 번호 및 구조변이 위치에 관한 정보이며,
 [0074] 리드 A 및 B는 목적하는 시료의 유전정보를 증폭한 서열 단편이며,
 [0075] 오더값 1은 절단점 간의 방향성을 구체적인 값으로, 예를 들어 값이 1인 경우 매칭-미매칭의 순서이고, 오더값 0은 미매칭-매칭의 순서이다.

[0076] 일 구체예에서, 상이한 염색체 내에서 서열이 이동하는 경우 도 6의 우측 그림과 같이, 절단점 2개 및 이를 연결한 선으로 표현되고, 하기와 같이 전산상 DB에 저장될 수 있다

표 13

노드	BP_att1	BP_att2
BP1	Chr1	123456
BP2	Chr2	456789

표 14

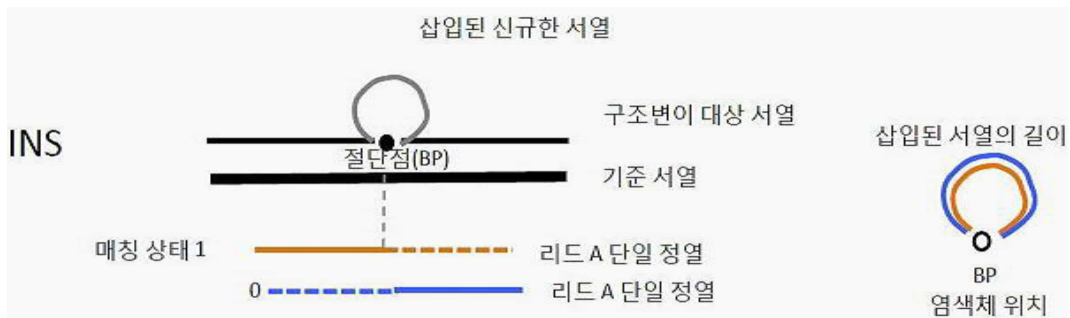
엣지	BP1	BP2	오더 1	오더2
리드A	BP1	BP2	1	0
리드B	BP1	BP2	0	1

[0079] 이때,

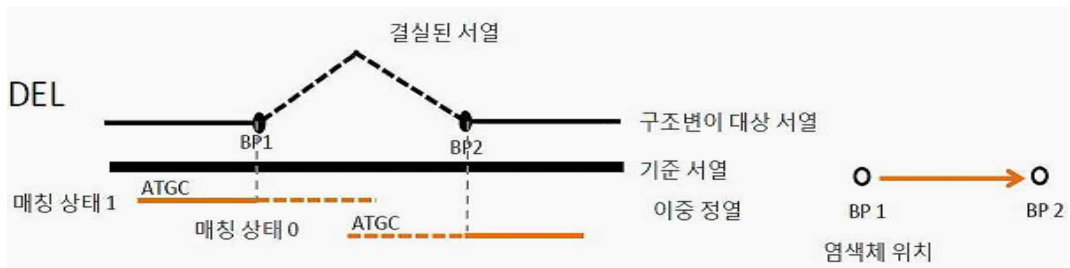
- [0080] BP1 및 BP2는 노드상의 임의의 절단점이고,
- [0081] BP_att1 및 BP_att1는 염색체 번호 및 구조변이 위치에 관한 정보이며,
- [0082] 리드 A 및 B는 목적하는 시료의 유전정보를 증폭한 서열 단편이며,
- [0083] 오더값 1은 절단점 간의 방향성을 구체적인 값으로, 예를 들어 값이 1인 경우 매칭-미매칭의 순서이고, 오더값 0은 미매칭-매칭의 순서이다.

도면

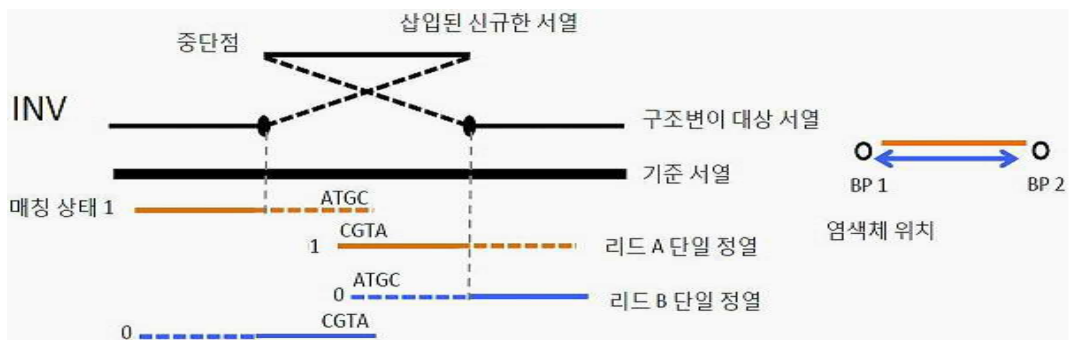
도면1



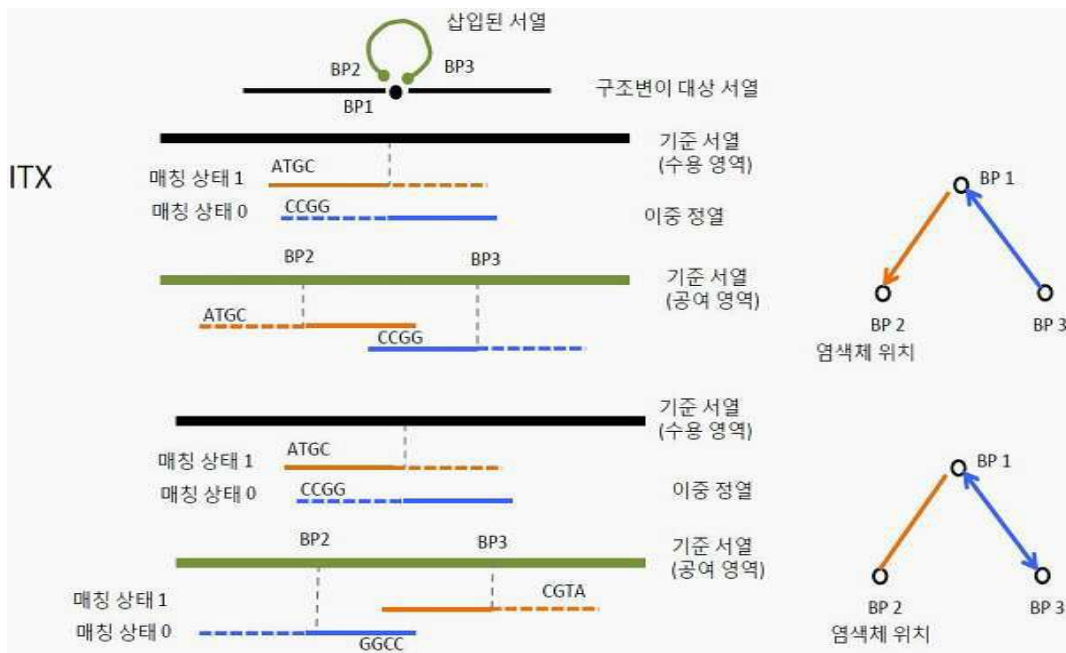
도면2



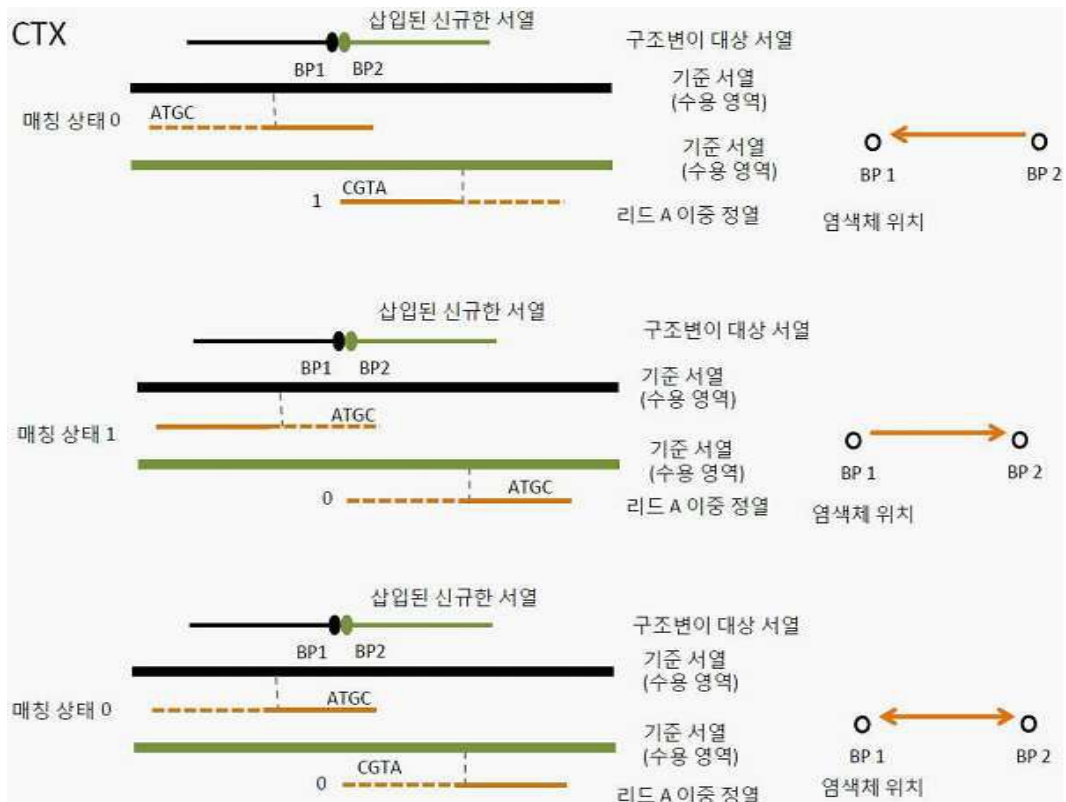
도면3



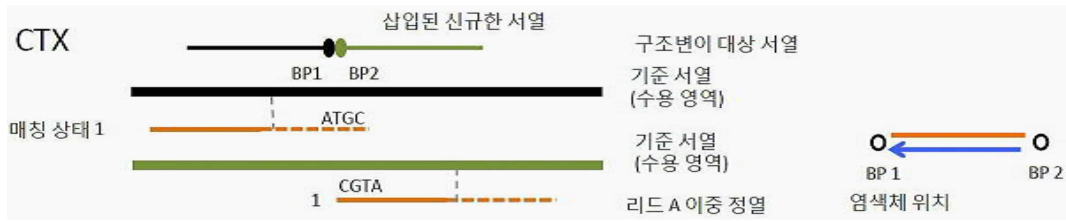
도면4



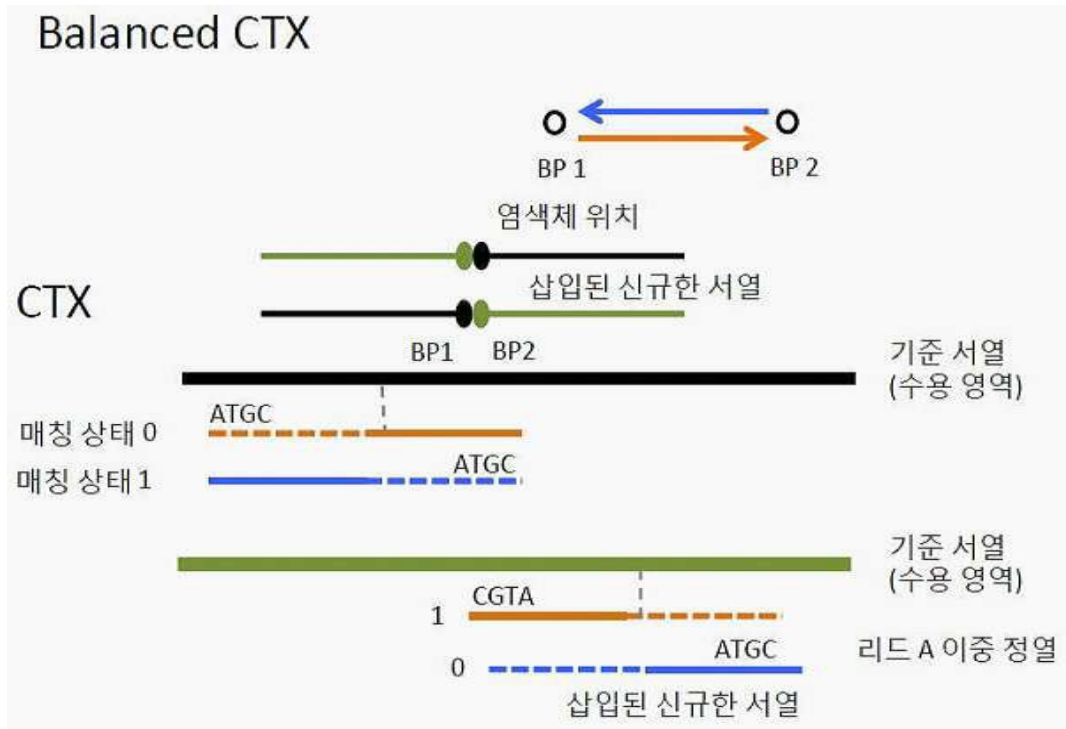
도면5



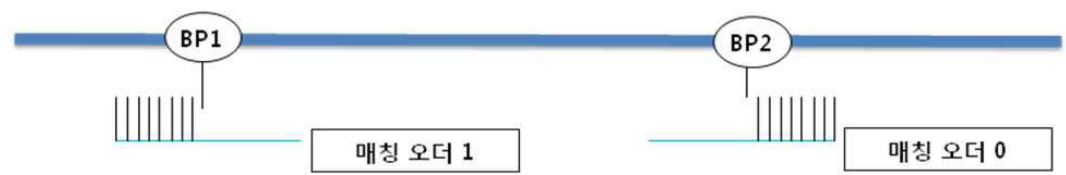
도면6



도면7



도면8



도면9

