



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0122095
(43) 공개일자 2022년09월02일

(51) 국제특허분류(Int. Cl.)
C12Q 1/6806 (2018.01) C12Q 1/6855 (2018.01)
C12Q 1/6869 (2018.01) C40B 40/06 (2006.01)
(52) CPC특허분류
C12Q 1/6806 (2018.05)
C12Q 1/6855 (2018.05)
(21) 출원번호 10-2021-0026287
(22) 출원일자 2021년02월26일
심사청구일자 없음

(71) 출원인
지니너스 주식회사
서울특별시 송파구 정의로 70, 5층(문정동, 케이디유타워)
(72) 발명자
박동현
강원도 춘천시 방송길 70, 105동 804호(온의동, 온의 롯데캐슬 스카이크래스)
정종석
경기도 화성시 동탄대로시범길 134, 1505동 1501호(청계동, 시범 반도유보라 아이비파크4.0)
(74) 대리인
리앤목특허법인

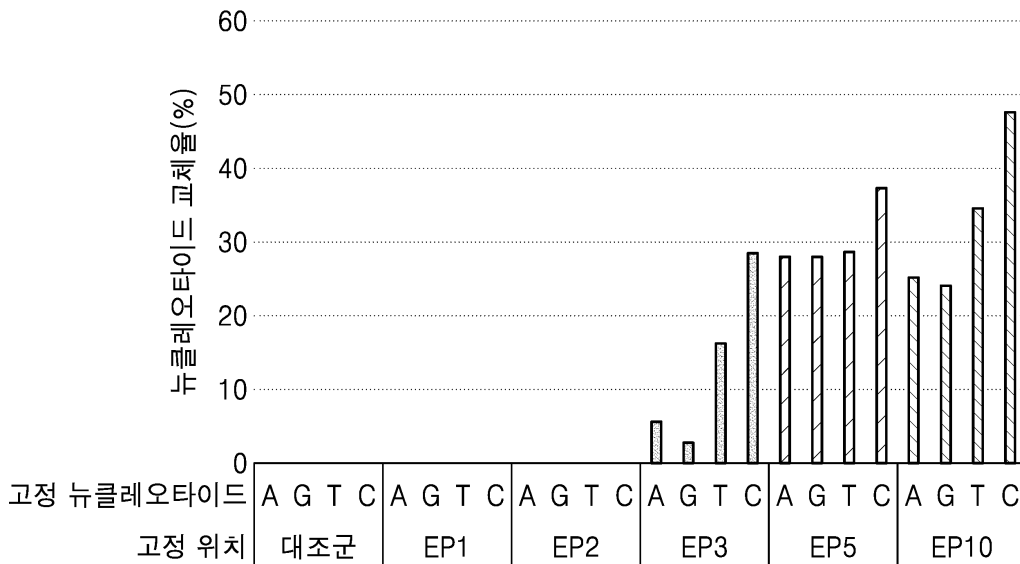
전체 청구항 수 : 총 19 항

(54) 발명의 명칭 분자 바코딩 효율을 향상시키기 위한 조성물 및 이의 용도

(57) 요약

분자 바코딩 효율을 향상시키기 위한 조성물 및 이의 용도에 관한 것으로서, DNA 라이브러리를 제조하는 방법, 핵산 서열 분석을 위한 분자 바코딩 방법 및 DNA 라이브러리 제조용 조성물을 제공한다. 상기 방법 및 조성물에 따르면, DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫번째 뉴클레오타이드 위치를 포함하도록 배열되는 바코드 서열을 이용함으로써, 분자 바코딩 교체 현상을 감소시키고, 핵산 서열 분석의 정확도를 향상시킬 수 있다.

대표도 - 도11



- (52) CPC특허분류
C12Q 1/6869 (2018.05)
C40B 40/06 (2013.01)
- (72) 발명자
신승호
 서울특별시 송파구 송파대로16길 9-1, 101호 (문정동, SD팰리체)
- 김솔이**
 경기도 성남시 수정구 위례순환로 17, 3101동 708호 (창곡동, 위례31단지)

임혜연

경기도 부천시 중동로71번길 39, 108동 106호(송내동, 뉴서울아파트)

박형빈

서울특별시 동작구 사당로16자길 37, 2동 301호(사당동, 그린파크빌)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711131842
과제번호	2018M3C9A6017315
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	포스트게놈신산업육성을위한다부처유전체사업(R&D)(과기정통부)
연구과제명	진행성 대장암 예후 예측 바이오마커 발굴 및 임상유용성 검증
기여율	1/5
과제수행기관명	지니너스 주식회사
연구기간	2021.01.01 ~ 2021.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711137915
과제번호	202011B12-01
부처명	다부처
과제관리(전문)기관명	(재단)범부처전주기료기기연구개발사업단
연구사업명	범부처전주기료기기연구개발사업(R&D)
연구과제명	진립선암의 맞춤형 치료를 위한 전주기 진단용 액체생검 통합 플랫폼 개발
기여율	2/5
과제수행기관명	지니너스 주식회사
연구기간	2021.03.01 ~ 2021.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711128807
과제번호	2017M3A9G5060264
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	바이오.의료기술개발(R&D)
연구과제명	암 조기진단을 위한 세포유리 DNA 분석 기술 개발
기여율	2/5
과제수행기관명	지니너스 주식회사
연구기간	2021.01.01 ~ 2022.06.29

명세서

청구범위

청구항 1

표적 시료에서 추출 및 절편화된 DNA 절편의 양말단에, 바코드 서열을 포함하는 어댑터를 라이게이션하는 단계; 상기 라이게이션된 DNA 절편을 단일 가닥으로 분리하는 단계; 및

상기 어댑터가 라이게이션된 단일 가닥 DNA 절편을 어댑터를 인식하는 프라이머를 사용하여 증합효소연쇄반응을 통해 증폭하는 단계를 포함하는, 핵산 서열 분석을 위한 DNA 라이브러리를 제조하는 방법으로서,

상기 바코드 서열은 1nt 내지 10nt의 길이를 가지고, DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫번째 뉴클레오티드 위치를 포함하도록 배열되는 것인, 핵산 서열 분석을 위한 DNA 라이브러리를 제조하는 방법.

청구항 2

청구항 1에 있어서, 상기 핵산 서열 분석은 차세대 염기서열 분석(next generation sequencing: NGS), 표적 염기서열 분석(targeted sequencing), 표적 딥 염기서열 분석(targeted deep sequencing), 또는 패널 염기서열 분석(panel sequencing)인 것인, 핵산 서열 분석을 위한 DNA 라이브러리를 제조하는 방법.

청구항 3

청구항 1에 있어서, 상기 핵산은 유전체(genome) 또는 그의 절편인 것인, 핵산 서열 분석을 위한 DNA 라이브러리를 제조하는 방법.

청구항 4

청구항 1에 있어서, 상기 어댑터는 40nt 내지 100nt의 길이를 가지고, 1nt 내지 10nt의 길이를 가지는 바코드 서열을 포함하는 것인, 핵산 서열 분석을 위한 DNA 라이브러리를 제조하는 방법.

청구항 5

청구항 1에 있어서, 상기 방법은 복수 개의 DNA 절편을 대상으로 하는 것으로, 복수 개의 어댑터를 포함하고, 상기 어댑터는 DNA 절편의 종류에 따라 상이한 바코드 서열을 포함하는 것인, 핵산 서열 분석을 위한 DNA 라이브러리를 제조하는 방법.

청구항 6

청구항 5에 있어서, 상기 바코드 서열은 DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫번째 서열이 A, T, C, 및 G 중 어느 하나인 것인, 핵산 서열 분석을 위한 DNA 라이브러리를 제조하는 방법.

청구항 7

청구항 5에 있어서, 상기 바코드 서열은 2nt 내지 10nt의 길이를 가지며, DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 두번째 서열이 C, 및 G 중 어느 하나인 것인, 핵산 서열 분석을 위한 DNA 라이브러리를 제조하는 방법.

청구항 8

청구항 5에 있어서, 상기 바코드 서열은 2nt 내지 10nt의 길이를 가지며, DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫번째 서열이 A, T, C, 및 G 중 어느 하나로 고정되고, 및 어댑터 내 두번째 서열이 C, 및 G 중 어느 하나로 고정된 것인, 핵산 서열 분석을 위한 DNA 라이브러리를 제조하는 방법.

청구항 9

표적 시료에서 추출 및 절편화된 DNA 절편의 양말단에, 바코드 서열을 포함하는 어댑터를 라이게이션하는 단계

를 포함하는, 핵산 서열 분석을 위한 분자 바코딩 방법으로서,

상기 바코드 서열은 1nt 내지 10nt의 길이를 가지고, DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫번째 뉴클레오티드 위치를 포함하도록 배열되는 것인, 핵산 서열 분석을 위한 분자 바코딩 방법.

청구항 10

청구항 9에 있어서, 상기 어댑터는 40nt 내지 100nt의 길이를 가지고, 1nt 내지 10nt의 길이를 가지는 바코드 서열을 포함하는 것인, 핵산 서열 분석을 위한 분자 바코딩 방법.

청구항 11

청구항 9에 있어서, 상기 방법은 복수 개의 DNA 절편을 대상으로 하는 것으로, 복수 개의 어댑터를 포함하고, 상기 어댑터는 DNA 절편의 종류에 따라 상이한 바코드 서열을 포함하는 것인, 핵산 서열 분석을 위한 분자 바코딩 방법.

청구항 12

청구항 11에 있어서, 상기 바코드 서열은 DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫번째 서열이 A, T, C, 및 G 중 어느 하나인 것인, 핵산 서열 분석을 위한 분자 바코딩 방법.

청구항 13

청구항 11에 있어서, 상기 바코드 서열은 2nt 내지 10nt의 길이를 가지며, DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 두번째 서열이 C, 및 G 중 어느 하나인 것인, 핵산 서열 분석을 위한 분자 바코딩 방법.

청구항 14

청구항 11에 있어서, 상기 바코드 서열은 2nt 내지 10nt의 길이를 가지며, DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫번째 서열이 A, T, C, 및 G 중 어느 하나로 고정되고, 및 어댑터 내 두번째 서열이 C, 및 G 중 어느 하나로 고정된 것인, 핵산 서열 분석을 위한 분자 바코딩 방법.

청구항 15

청구항 9에 있어서, 상기 어댑터를 라이게이션하는 단계 전, 표적 시료에서 추출 및 절편화된 DNA 절편의 말단을 수선(Repair)하는 단계; 및

상기 말단이 수선된 DNA 절편의 3'말단에 아데닌을 결합시키는 단계를 추가로 포함하는 것인, 핵산 서열 분석을 위한 분자 바코딩 방법.

청구항 16

중합효소연쇄반응을 위한 프라이머에 대한 상보적인 서열 및 바코드 서열을 포함하는 40nt 내지 100nt의 길이를 갖는 복수 개의 어댑터로서,

상기 복수 개의 어댑터 각각은 DNA 절편의 종류에 따라 상이한 바코드 서열을 포함하고,

상기 바코드 서열은 1nt 내지 10nt의 길이를 가지고, DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫번째 뉴클레오티드 위치를 포함하도록 배열된 것인, 핵산 서열 분석을 위한 DNA 라이브러리 제조용 조성물.

청구항 17

청구항 16에 있어서, 상기 바코드 서열은 DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫번째 서열이 A, T, C, 및 G 중 어느 하나인 것인, 핵산 서열 분석을 위한 DNA 라이브러리 제조용 조성물.

청구항 18

청구항 16에 있어서, 상기 바코드 서열은 2nt 내지 10nt의 길이를 가지며, DNA 절편과 라이게이션되는 어댑터의

말단 영역을 기준으로, 어댑터 내 두번째 서열이 C, 및 G 중 어느 하나인 것인, 핵산 서열 분석을 위한 DNA 라이브러리 제조용 조성물.

청구항 19

청구항 16에 있어서, 상기 바코드 서열은 2nt 내지 10nt의 길이를 가지며, DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫번째 서열이 A, T, C, 및 G 중 어느 하나로 고정되고, 및 어댑터 내 두번째 서열이 C, 및 G 중 어느 하나로 고정된 것인, 핵산 서열 분석을 위한 DNA 라이브러리 제조용 조성물.

발명의 설명

기술 분야

[0001] 분자 바코딩 효율을 향상시키기 위한 조성물 및 이의 용도에 관한 것이다.

배경 기술

[0002] 다양한 생체 정보는 DNA 서열의 유전자로 표현되고, 개체의 완전한 DNA 서열 정보는 생명 현상을 이해하고 질병과 관련된 정보를 얻을 수 있어 매우 중요하다. DNA 서열 정보의 해독, 즉 게놈 시퀀싱(genome sequencing)의 핵심은 개인차 및 민족적 특성을 파악하거나 유전자 이상과 관련된 질환에서 염색체 이상을 포함한 선천성 원인의 규명과 당뇨병, 고혈압과 같은 복합 질병의 유전자 결함을 찾기 위한 것이다. 또한, 시퀀싱 데이터는 유전자 발현, 유전자 다양성 및 그 상호작용 등의 정보들을 분자 진단과 치료 영역에서 폭넓게 활용할 수 있어 매우 중요하다.

[0003] 게놈 시퀀싱을 위한 방법으로는 차세대 염기서열 분석법(Next Generation Sequencing; NGS)이 2007년 이래로 적용되기 시작하였고, 이러한 NGS의 개발에 따라 전통적인 방법과 비교하여 훨씬 쉽고 저비용으로 분석할 수 있게 되었다. 차세대 염기서열 분석법을 구현하는 차세대 게놈 시퀀서(Next Generation Sequencer)로 대표적인 것으로는 로슈(Roche)/454, 일루미나(Illumina)/Solexa 및 라이프 테크놀로지스(ABI)의 SOLiD 등이 있다. 이러한 차세대 염기서열 분석기기들은 7시간에 8,000만개 이상의 서열 판독이 가능하다. 이러한 기술 발전으로 종래에 막대한 검사 비용으로 인해 연구용으로만 사용되던 차세대 염기서열 분석법을 의료용 임상 검사에서도 활용할 수 있게 되었다.

[0004] 한편, 시퀀싱 기술의 발달로 인해 다양한 종류의 구조적 변이(structure variation)를 발굴하는 다양한 시도가 이루어지고 있으나, 분석 과정에서 상당한 수준의 위양성(false positive)나 위음성(false negative)이 발생하고 있어, 이러한 문제점을 해결하기 위한 다양한 시도가 이루어지고 있다. 이러한 문제점을 해결하기 위한 기술로서, 예를 들어, 분자 바코딩 기술이 널리 활용되고 있으나, 라이게이션 이후 정제 과정에서 제거되지 않은 어댑터가 PCR 증폭 단계에서 프라이머로 작용하여 샘플 정보에 대한 오류를 유발하는 바, 검출의 민감도 및 특이도 측면에서 기술적 한계가 존재한다.

[0005] 오늘날 정밀 의료 기반의 진단/치료 기술 분야가 주목받기 시작하면서, 더욱 정밀하고 정확한 분석의 필요성이 대두되고 있으며, 이에 따라, 분자 바코딩 기술에 대한 연구가 활발하게 진행되고 있으나(한국등록특허 제10-1575457호), 아직은 미비한 실정이다.

발명의 내용

해결하려는 과제

[0006] 일 양상은 핵산 서열 분석을 위한 DNA 라이브러리를 제조하는 방법을 제공하는 것이다.

[0007] 다른 양상은 핵산 서열 분석을 위한 분자 바코딩 방법을 제공하는 것이다.

[0008] 다른 양상은 핵산 서열 분석을 위한 DNA 라이브러리 제조용 조성물을 제공하는 것이다.

[0009] 본 출원의 다른 목적 및 이점은 첨부한 청구범위와 함께 하기의 상세한 설명에 의해 보다 명확해질 것이다. 본 명세서에 기재되지 않은 내용은 본 출원의 기술 분야 또는 유사한 기술 분야 내 숙련된 자이면 충분히 인식하고 유추할 수 있는 것이므로 그 설명을 생략한다.

과제의 해결 수단

- [0010] 본 명세서에서 사용되는 용어는 각 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 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서, 본 명세서에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 명세서 전반에 걸친 내용을 토대로 정의되어야 한다.
- [0011] 각 설명들에서, 어떤 부분이 다른 부분과 연결되어 있다고 할 때, 이는 직접적으로 연결되어 있는 경우 뿐만 아니라, 그 중간에 다른 구성요소를 사이에 두고 유기적으로 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. 본 명세서에서 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.
- [0012] 각 설명은 권리범위를 제한하는 것으로 해석되지 말아야 하며, 해당 기술분야의 통상의 기술자가 용이하게 유추할 수 있는 것은 권리범위에 속하는 것으로 해석되어야 할 것이다.
- [0013] 일 양상은 표적 시료에서 추출 및 절편화된 DNA 절편의 양말단에, 바코드 서열을 포함하는 어댑터를 라이게이션하는 단계; 상기 라이게이션된 DNA 절편을 단일 가닥으로 분리하는 단계; 및 상기 어댑터가 라이게이션된 단일 가닥 DNA 절편을 어댑터를 인식하는 프라이머를 사용하여 증합효소연쇄반응을 통해 증폭하는 단계를 포함하는, 핵산 서열 분석을 위한 DNA 라이브러리를 제조하는 방법으로서, 상기 바코드 서열은 1nt 내지 10nt의 길이를 가지고, DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫번째 뉴클레오티드 위치를 포함하도록 배열되는 것인, 핵산 서열 분석을 위한 DNA 라이브러리를 제조하는 방법을 제공한다.
- [0014] 또한, 다른 양상은 표적 시료에서 추출 및 절편화된 DNA 절편의 양말단에, 바코드 서열을 포함하는 어댑터를 라이게이션하는 단계를 포함하는, 핵산 서열 분석을 위한 분자 바코딩 방법으로서, 상기 바코드 서열은 1nt 내지 10nt의 길이를 가지고, DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫번째 뉴클레오티드 위치를 포함하도록 배열되는 것인, 핵산 서열 분석을 위한 분자 바코딩 방법을 제공한다.
- [0015] 본 명세서에서 사용된 용어, "DNA 라이브러리"는 표적 시료에 대한 핵산 서열 분석을 위한 전 처리의 결과물로서, DNA 샘플 가공의 산물을 지칭한다. 통상적으로, 상기 DNA 라이브러리는 DNA 절편의 양말단에 어댑터 올리고 뉴클레오티드를 라이게이션시키는 단계를 포함하여 제조된 것으로, 이의 증폭된 산물 역시 포함될 수 있다. 상기 DNA 라이브러리는 예를 들어, i) DNA 샘플을 무작위적으로 잘라 DNA 절편을 수득하는 단계 (DNA 절편화); ii) 두 개의 단일 가닥 DNA의 양쪽 말단 부분을 보강하여, 평활 말단(blunt end) 구조를 형성하는 단계 (말단 수선); iii) 이중 가닥 DNA의 3' 말단에 아데닌을 접합시켜 Overhang 구조를 형성하는 단계 (아데닌 접합); iv) 상기 아데닌 접합된 어댑터를 이중 가닥 DNA 말단에 라이게이션시키는 단계 (어댑터 부착); v) 특정 효소를 사용하여, 어댑터에 의해 연결되어 있는 이중 가닥 DNA를 단일 가닥으로 분리하는 단계 (단일 가닥 분리); 및 vi) 증합효소연쇄반응(Polymerase Chain Reaction; PCR)을 통해 어댑터 서열과 부착된 DNA 절편을 증폭시키는 단계 (PCR 증폭)를 포함하는 일련의 과정을 통해 제조된 것일 수 있다. 상기 DNA 라이브러리는 최종 생성물 뿐만 아니라, 상기 언급한 각 단계에서의 중간 생성물까지 포함하는 것으로 해석될 수 있다.
- [0016] 상기 각 단계는 당업계에 알려진 공지된 기술이 비제한적으로 적용될 수 있다. 예를 들어, 상기 어댑터 부착 단계에서 T4 DNA 라이게아제, T7 DNA 라이게아제, 또는 온도 순환시험(temperature cycling)이 가능한 라이게아제가 사용될 수 있고, 또는 상기 PCR 증폭 단계는 4 회 내지 12 회, 4 회 내지 10 회, 4 회 내지 8 회, 4 회 내지 6 회, 6 회 내지 12 회, 6 회 내지 10 회, 6 회 내지 8 회, 8 회 내지 12 회, 8 회 내지 10 회, 또는 10 회 내지 12 회 사이클로 수행될 수 있으며, 그 외, 당업계에 널리 사용되는 기술이 일 실시예에 따른 핵산 서열을 분석하는데 사용될 수 있다.
- [0017] 본 명세서에서 사용된 용어, "분자 바코드(Molecular barcode)"는 표적 시료 내에 존재하는 각각의 DNA 절편에 대해 고유한 뉴클레오티드 서열을 지칭하는 것으로서, 표적 시료 내에 존재하는 DNA 절편을 구분 가능하게 한다. 종래의 분자 바코드 기술은 PCR 증폭 과정에서, 잘못된 바코드 서열 또는 고유 식별자(Unique identifier: UID)를 갖는 어댑터가 후속 PCR 반응에서 프라이머로 작용하는 경우, 잘못된 PCR 복제물을 생성하게 되며(분자 바코딩 교체 현상), 이에 따라, 유전체 분석 과정에서, 위양성의 리드(Read)가 증가될 뿐만 아니라, 정상 대조군 리드의 증가로 인한 양성 검출의 민감도가 감소되어, 높은 Depth를 갖는 유전 정보 분석 결과의 신뢰도를 감소시키는 요인으로 작용하고 있다. 이러한 기술적 배경 하에서, 본 발명자들은 상기 DNA 절편과

라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫번째 뉴클레오티드 위치를 포함하도록 배열되는 바코드 서열, 그리고, 상기 말단 영역 중 특정 위치에 특정 염기를 배열함으로써, 상기 언급한 종래의 문제점을 극복할 수 있음을 확인하고, 이에 기초하여 본 발명을 완성하였다.

[0018] 본 명세서에서 사용된 용어, "표적 시료"는 개체 또는 세포로부터 유래할 수 있다. 상기 개체는 인간, 소, 말, 돼지, 양, 염소, 개, 고양이, 및 설치류를 포함한 포유류일 수 있다. 상기 세포는 개체로부터 유래된 세포 또는 세포주일 수 있다. 또한, 상기 표적 시료는 생물학적 시료일 수 있다. 상기 생물학적 시료는 예를 들어, 혈액, 혈장, 혈청, 소변, 타액, 점막, 분비물, 객담, 대변, 눈물, 또는 이들의 조합으로부터 획득된 것일 수 있다. 상기 생물학적 시료는 다양한 종으로부터 유래하는 진핵세포, 원핵세포, 바이러스, 박테리오파지 등의 시료일 수 있다.

[0019] 본 명세서에서 사용된 용어, "핵산 서열 분석(nucleic acid sequencing analysis)"은 차세대 핵산 서열분석(next generation sequencing: NGS)인 것일 수 있다. 핵산 서열분석은 염기 서열분석, 서열분석 또는 시퀀싱(sequencing)과 상호 교환적으로 사용되는 것일 수 있다. 상기 NGS는 대규모 병렬 서열분석(massive parallel sequencing) 또는 2세대 서열분석(second-generation sequencing)과 상호 교환적으로 사용되는 것일 수 있다. 상기 NGS는 대량의 단편의 핵산을 동시다발적으로 서열분석하는 기법으로서, 칩(chip) 기반 그리고 중합효소 연쇄 반응 (polymerase chain reaction: PCR) 기반 쌍 말단(paired end) 형식으로 전장 유전체를 조각내고, 상기 조각을 혼성화 반응(hybridization)에 기초하여 초고속으로 서열 분석을 수행하는 것일 수 있다. 상기 NGS는 예를 들면, 454 플랫폼(Roche), GS FLX 티타늄, Illumina MiSeq, Illumina HiSeq, Illumina HiSeq 2500, Illumina Genome Analyzer, Solexa platform, SOLiD System(Applied Biosystems), Ion Proton(Life Technologies), Complete Genomics, Helicos Biosciences Heliscope, Pacific Biosciences의 단일 분자 실시간(SMRT™) 기술, 또는 이들의 조합에 의해 수행되는 것일 수 있다. 상기 핵산 서열분석은 관심 영역만을 분석하기 위한 핵산 서열분석법인 것일 수 있다. 상기 핵산 서열분석은, 예를 들면, NGS 기반의 표적 서열분석(targeted sequencing), 표적 딥 서열분석(targeted deep sequencing), 또는 패널 서열분석(panel sequencing)을 포함하는 것일 수 있다. 여기서, 핵산은 유전체 또는 그의 절편일 수 있다. 본 명세서에서 사용된 용어, "유전체(genome)"는 염색체, 염색질, 또는 유전자의 전체를 총칭하는 용어이다. 상기 유전체 또는 그의 절편은 분리된 DNA일 수 있다. 상기 세포로부터 핵산을 추출 또는 분리하는 방법은 통상의 기술자에게 공지된 방법으로 수행될 수 있다. 여기서, 절편은 유전체를 물리적, 화학적, 또는 효소적으로 절단하는 것을 의미하며, 상기 과정을 통해 다양한 길이(length)를 갖는 리드를 생성하는 것일 수 있다. 본 명세서에서 사용된 용어, "리드(read)"는 핵산 서열 분석에서 생성된 하나 이상의 핵산 절편의 서열 정보를 의미하며, 상기 리드는 약 10bp 내지 약 2000bp, 예를 들어, 약 15bp 내지 약 1500bp, 약 20bp 내지 약 1000bp, 약 20bp 내지 약 500bp, 약 20bp 내지 약 200bp, 약 20bp 내지 약 100bp일 수 있으나, 이에 제한되는 것은 아니다.

[0020] 상기 바코드 서열을 포함하는 어댑터를 라이게이션 하는 단계에서, 상기 바코드 서열은 1nt 내지 10nt의 길이를 가지고, DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫번째 뉴클레오티드 위치를 포함하도록 배열되는 것일 수 있다.

[0021] 본 명세서에서 사용된 용어, "어댑터(Adapter)"는 복수 개의 DNA 절편을 구분하기 위한 바코드 서열을 포함하는 올리고뉴클레오티드를 지칭하는 것으로서, Y자 형태(Forked) 또는 U자 형태(Hairpin) 형태일 수 있다. 예를 들어, 상기 어댑터는 우라실 특이 절제 시약(USER)과 같은 효소를 이용하여, 상기 U 자 모양의 어댑터를 Y 자 형태의 말단을 갖는 것으로 절단된 것일 수 있으나, 이에 제한되는 것은 아니다.

[0022] 또한, 이중 가닥의 DNA 절편의 3' 말단은 아데닌 접합에 의해 Overhang 구조를 지니고 있는 바, 상기 DNA 절편과의 라이게이션을 위하여, 상기 어댑터는 이중 가닥으로 이루어져 있을 뿐만 아니라, 상기 어댑터의 3' 말단에는 티민이 접합되어 있을 수 있다. 여기서, 상기 어댑터 서열의 길이는 40nt 내지 100nt, 40nt 내지 90nt, 40nt 내지 80nt, 40nt 내지 70nt, 40nt 내지 60nt, 40nt 내지 50nt, 50nt 내지 100nt, 50nt 내지 90nt, 50nt 내지 80nt, 50nt 내지 70nt, 50nt 내지 60nt, 60nt 내지 100nt, 60nt 내지 90nt, 60nt 내지 80nt, 60nt 내지 70nt, 70nt 내지 100nt, 70nt 내지 90nt, 70nt 내지 80nt, 80nt 내지 100nt, 또는 80nt 내지 90nt일 수 있으나, 이에 제한되는 것은 아니다.

[0023] 본 명세서에서 사용된 용어, "바코드(Barcode)"는 "인덱스(index)" 또는 "고유 식별자(Unique identifier: UID)"와 상호 교환적으로 사용할 수 있으며, 복수 개의 DNA 절편을 구분하기 위한 1nt 내지 10nt의 길이의 올리고뉴클레오티드를 지칭한다. 상기 바코드 서열은 어댑터 내 포함되어 있으며, 이에 따라, 상보적인 결합을 형성하는 하나의 쌍 또는 조합을 지칭하는 것일 수 있다. 여기서, 상기 바코드 서열의 길이는 1nt 내지 10nt, 1nt

내지 8nt, 1nt 내지 6nt, 1nt 내지 4nt, 1nt 내지 2nt, 3nt 내지 10nt, 3nt 내지 8nt, 3nt 내지 6nt, 1nt 내지 4nt, 5nt 내지 10nt, 5nt 내지 8nt, 5nt 내지 6nt, 7nt 내지 10nt, 또는 7nt 내지 8nt일 수 있으나, 이에 제한되는 것은 아니다.

- [0024] 일 구체예에서, 상기 바코드 서열은 DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫 번째 뉴클레오티드 위치를 포함하도록 배열되는 것일 수 있다. 여기서, 상기 어댑터의 말단 영역 내 첫 번째 뉴클레오티드는 DNA 절편과 라이게이션되는 위치에 존재하는 최말단의 뉴클레오티드로서, i) 표적 DNA 절편의 3' 말단에 존재하는 아데닌과 인접한, 어댑터 서열의 5' 말단의 뉴클레오티드, ii) DNA 절편의 3' 말단에 존재하는 아데닌과 상보적 염기서열(티민)을 포함하는, 어댑터 서열의 3' 말단의 티민과 인접한(연결된) 어댑터 서열의 뉴클레오티드를 지칭하는 것일 수 있다.
- [0025] 일 구체예에서, 상기 방법은 복수 개의 DNA 절편을 대상으로 하는 것으로, 복수 개의 어댑터를 포함하고, 이에 따라, 상기 어댑터는 DNA 절편의 종류에 따라 상이한 바코드 서열을 포함하는 것일 수 있다.
- [0026] 일 구체예에서, 상기 바코드 서열은 DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫 번째 서열이 A, T, C, 및 G 중 어느 하나일 수 있거나, 상기 바코드 서열이 2nt 내지 10nt의 길이를 가지는 경우, DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 두 번째 서열이 C, 및 G 중 어느 하나일 수 있다. 또한, 상기 바코드 서열이 2nt 내지 10nt의 길이를 가지는 경우, DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫 번째 서열이 A, T, C, 및 G 중 어느 하나로 고정되고, 및 어댑터 내 두 번째 서열이 C, 및 G 중 어느 하나로 고정된 것일 수 있다.
- [0027] 일 실시예에 따르면, 상기 언급한 배열을 갖는 바코드 서열 및 이를 포함하는 어댑터 서열은 도 4에 도시한 바와 같이, 잘못된 바코드 서열을 포함하는 어댑터가 PCR 증폭 과정에서 프라이머로 작용하는 경우 DNA 중합효소 I의 교정(Proofreading) 과정(3' → 5' Exonuclease 활성)을 통해, 자체적으로, 잘못된 복제 및 증폭 과정을 중단시켜, 종래의 분자 바코딩 교체 현상을 감소시킬 수 있다.
- [0028] 다른 양상은 중합효소연쇄반응을 위한 프라이머에 대한 상보적 서열 및 바코드 서열을 포함하는 40 내지 100nt의 길이를 갖는 복수 개의 어댑터로서, 상기 복수 개의 어댑터 각각은 DNA 절편의 종류에 따라 상이한 바코드 서열을 포함하고, 상기 바코드 서열은 1nt 내지 10nt의 길이를 가지고, DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫 번째 뉴클레오티드 위치를 포함하도록 배열된 것인, 핵산 서열 분석을 위한 DNA 라이브러리 제조용 조성물을 제공한다.
- [0029] 상기 핵산 서열 분석을 위한 DNA 라이브러리 제조용 조성물은 전술한 DNA 라이브러리를 제조하는 방법 또는 분자 바코딩 방법에 사용되는 기술적 구성을 그대로 포함하거나, 이를 이용하기 때문에, 이 둘 사이에 공통된 내용은 기재를 생략한다.
- [0030] 구체적으로, 본 명세서에서 사용된, "DNA 라이브러리 제조용 조성물"은 DNA 라이브러리 제조용 키트, 또는 DNA 라이브러리 제조를 위한 어댑터 조성물과 상호 교환적으로 사용될 수 있다.
- [0031] 일 구체예에서, 상기 조성물은 핵산 서열 분석을 위한 것으로, 복수 개의 DNA 절편을 대상으로 하는 것일 수 있다. 이에, 상기 조성물은 복수 개의 어댑터를 포함하고, 이에 따라, 상기 어댑터는 DNA 절편의 종류에 따라 상이한 바코드 서열을 포함하는 것일 수 있다.
- [0032] 상기 어댑터는 중합효소연쇄반응을 위한 프라이머에 대한 상보적인 서열 및 바코드 서열을 포함하는 것일 수 있다. 여기서, 상기 프라이머에 대한 상보적인 서열은 PCR 증폭 과정을 개시 또는 수행하기 위한 것으로, 공지 또는 임의의 프라이머 및 이에 상보적인 서열이 비제한적으로 적용될 수 있다.
- [0033] 일 구체예에서, 상기 바코드 서열은, 전술한 바와 같이, DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫 번째 뉴클레오티드 위치를 포함하도록 배열된 것일 수 있다. 또한, 상기 바코드 서열은 상기 바코드 서열은 DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫 번째 서열이 A, T, C, 및 G 중 어느 하나일 수 있거나, 상기 바코드 서열이 2nt 내지 10nt의 길이를 가지는 경우, DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 두 번째 서열이 C, 및 G 중 어느 하나일 수 있다. 또한, 상기 바코드 서열이 2nt 내지 10nt의 길이를 가지는 경우, DNA 절편과 라이게이션되는 어댑터의 말단 영역을 기준으로, 어댑터 내 첫 번째 서열이 A, T, C, 및 G 중 어느 하나로 고정되고, 및 어댑터 내 두 번째 서열이 C, 및 G 중 어느 하나로 고정된 것일 수 있다.
- [0034] 일 실시예에 따르면, 종래의 배열을 갖는 바코드 서열 또는 이를 포함하는 어댑터 서열은 약 6% 내지 50%의 분

자 바코드 교체율을 나타내었던 반면, 일 실시예에 따른 어댑터 서열은 1% 미만의 분자 바코드 교체율, 즉 오류율을 나타냄을 확인하였는 바, 핵산 서열 분석, 예를 들어, 유전 변이 검출 과정에서 변이 검출의 민감도 및 특이도 높일 수 있으므로, 정밀 의료 기반의 진단/치료 기술 분야에 폭넓게 활용될 수 있다.

발명의 효과

[0035] 일 양상에 따른 방법 및 조성물에 따르면, 종래 기술의 문제점으로 지적되어 온 분자 바코딩 교체 현상을 감소시킬 수 있고, 이에 따라, 분자 바코딩 정확도 및 위양상 변이 제거율을 크게 향상시킬 수 있다.

[0036] 일 양상에 따른 방법 및 조성물에 따르면, 유전 변이 검출 과정에서 변이 검출의 민감도 및 특이도 높일 수 있으므로, 계층 시퀀싱에 기반한 치료 및 진단 분야의 정확성을 크게 향상시킬 수 있다.

도면의 간단한 설명

- [0037] 도 1은 종래 분자 바코딩 기술의 문제점을 개략적으로 나타낸 도이다.
- 도 2는 분자 바코딩 교체 현상을 유도하기 위한 실험 과정 및 이에 따른 결과물을 개략적으로 나타낸 도이다.
- 도 3은 종래 분자 바코딩 기술이 적용된 PCR 증폭 과정에서, 분자 바코딩의 교체 비율을 정제 횟수 및 어댑터의 농도에 따라 확인한 결과이다.
- 도 4는 일 구체예에서 따른 어댑터 및 이를 이용한 PCR 증폭 과정에서, 분자 바코딩 교체 현상을 감소시키는 일련의 과정을 개략적으로 나타낸 도이다.
- 도 5는 일 구체예에 따른 바코드 서열을 어댑터 말단 영역으로 그 위치를 고정된 어댑터를 이용한 PCR 증폭 과정에서, 분자 바코딩 교체 현상 수준을 평가하기 위한 실험 과정을 개략적으로 나타낸 도이다.
- 도 6은 일 구체예에 따른 바코드 서열의 위치가 어댑터 말단 영역으로 고정된 어댑터를 이용한 PCR 증폭 과정에서, 분자 바코딩 교체 수준을 확인한 결과이다.
- 도 7은 일 구체예에 따른 바코드 서열의 위치가 어댑터 말단 영역으로 고정된 어댑터를 이용한 PCR 증폭 과정에서, 바코드 서열 내 특정 위치가 아데닌(A)으로 고정된 경우, 분자 바코딩 교체 수준을 확인한 결과이다.
- 도 8은 일 구체예에 따른 바코드 서열의 위치가 어댑터 말단 영역으로 고정된 어댑터를 이용한 PCR 증폭 과정에서, 바코드 서열 내 특정 위치가 티민(T)으로 고정된 경우, 분자 바코딩 교체 수준을 확인한 결과이다.
- 도 9는 일 구체예에 따른 바코드 서열의 위치가 어댑터 말단 영역으로 고정된 어댑터를 이용한 PCR 증폭 과정에서, 바코드 서열 내 특정 위치가 시토신(C)으로 고정된 경우, 분자 바코딩 교체 수준을 확인한 결과이다.
- 도 10은 일 구체예에 따른 바코드 서열의 위치가 어댑터 말단 영역으로 고정된 어댑터를 이용한 PCR 증폭 과정에서, 바코드 서열 내 특정 위치가 구아닌(G)으로 고정된 경우, 분자 바코딩 교체 수준을 확인한 결과이다.
- 도 11은 일 구체예에 따른 바코드 서열의 위치가 어댑터 말단 영역으로 고정된 어댑터를 이용한 PCR 증폭 과정에서, 바코드 서열 내 특정 위치 및 이의 염기서열이 고정된 경우, 특정 위치 및 고정된 염기서열에 따른 분자 바코딩 교체 수준을 확인한 것이다.
- 도 12는 일 구체예에 따른 바코드 서열의 위치가 어댑터 말단 영역으로 고정된 어댑터를 이용한 PCR 증폭 과정에서, 어댑터 말단 첫번째 염기서열(EP1)의 고정에 따른 분자 바코딩 교체 수준을 확인한 결과이다.
- 도 13은 일 구체예에 따른 바코드 서열의 위치가 어댑터 말단 영역으로 고정된 어댑터를 이용한 PCR 증폭 과정에서, 바코딩 서열에 따른 효능을 확인한 결과로서, 어댑터 말단 두번째 염기서열(EP2)의 고정에 따른 분자 바코딩 교체 수준을 확인한 결과이다.

발명을 실시하기 위한 구체적인 내용

[0038] 이하, 본 발명의 이해를 돕기 위하여 바람직한 실시예를 제시한다. 그러나 하기의 실시예는 본 발명을 보다 쉽게 이해하기 위하여 제공되는 것일 뿐, 하기 실시예에 의해 본 발명의 내용이 한정되는 것은 아니다.

[0039] [실시예]

[0040] 실시예 1. 분자 바코딩 교체에 따른 오류 가능성의 확인

[0041] 본 실시예에서는 PCR 증폭 과정에서 발생할 수 있는, 분자 바코딩 교체 현상에 따른 오류 가능성을 확인하고자 하였다. 통상의 유전체 분석 기술은 무작위적인 DNA의 절편화 및 높은 Depth를 갖는 시퀀싱 과정을 필요로 하며, 이에 따라, 다량의 DNA 절편을 증폭 및 선별하는 과정에서 유효한 유전 정보를 상실할 우려가 존재한다. 따라서, 당업계에서는 특정 바코드 서열을 포함하는 어댑터를 DNA 절편에 라이게이션시키는 과정을 포함하는 분자 바코딩 기술을 통해, 이러한 문제점으로부터 유발되는 오류를 최소화시키고 있다. 그러나, 이러한 종래 분자 바코딩 기술 역시, 도 1에 도시한 바와 같이, PCR 증폭 과정에서, 잘못된 바코드 서열을 포함하는 어댑터가 후속 PCR 반응에서 프라이머로 작용하여 잘못된 PCR 복제물을 생성하게 되어, 검출의 민감도 및 특이도를 감소시키는바 (분자 바코딩 교체 현상), 기술적 한계가 존재하는 실정이다.

[0042] 본 실시예에서는 도 2의 A에 도시한 바와 같이, 이러한 종래 기술의 문제점을 실험적으로 확인하고자 하였다. 우선, 인간 참조 유전체 정보를 활용하여 5개의 유전자 KRAS, IDH1, BRCA1, ALK, ERBB2의 염기서열 일부를 확보한 후, 상기 서열의 말단에 'CTTC'서열을 추가하였다. 이후, 각 염기서열에 일루미나(Illumina) 시퀀서 활용 가능한 라이브러리 형태의 어댑터 서열을 추가하면서, 상기 어댑터 서열 내 바코드 서열 위치에 특정 시퀀스 (AGTC)를 고정된 형태로, 어댑터가 부착된 DNA를 합성하였다. 상기 각 염기서열에 대한 정보는 하기 표 1에 나타낸 바와 같다.

표 1

유전자	Sequence (5' → 3')	UID
KRAS	CTTCATCCTGAGAAGGGAGAAACACAGTCTGGATTATTACAGTGCACCTTTACTTCAAAAAAGGTGTATATAACAATC AACAAACAAAAATTCATTTAAAAATGGGCAAGGACTTGAAGACATTGTCTCTGCTCCAAAGACTTC	AGTC
IDH1	CTTCAATGGCTTCTCTGAAGACCGTGCCACCCAGAATATTTTCGTATGGTGCCATTTGGTGATTCCACATTTGTTTCAAC TTGAACCTCTCAACCTCTTCTCATCAGGAGTGATAGTGGCACATTTGACGCCAACATTATGCTTCCTTC	AGTC
BRCA1	CTTCTTCTGGCTTCTCCCTGTCTCACATTTCTTCCATTGCATTATACCCAGCAGTATCAGTAGTATGAGCAGCAGCTGGA CTCTGGGCAGATTCTGCAACTTTCAACTTTCAATTGGGGAACCTTCAATGCAGAGGTTGAAGATGGCTTC	AGTC
ALK	CTTCACTGATGGAGGAGGCTTGGCCAGCAAAGCAGTAGTTGGGGTGTAGTCGGTCATGATGGTCGAGGTGCGGAGCTTG CTCAGCTTGTACTCAGGGCTCTGCAGTCCATCTGCATGGCTTGCAGCTCCTGGTGCTCCGGCGGCTTC	AGTC
ERBB2	CTTCGTACGTGCTCATCGCTCACAAACCAAGTGAGGCAGGTCCCAGTGCAGAGGCTGCGGATTGTGCGGAGGCACCCAGCT CTTTGAGGACAACATATGCCCTGGCCGTGCTAGACAATGGAGACCCGCTGAACAATACCACCCTGTCTTC	AGTC

[0044] 한편, 상기 언급된 잘못된 바코드 서열을 포함하는 어댑터가 PCR 증폭 과정에서 작용하는 과정을 모사하기 위하여, 표적 서열 내 인공 변이 서열이 도입된 염기서열을 추가로 합성하였으며, 본 실시예에서는 이를 specific sequence synthesis fragment 또는 spike-in 절편으로 지칭하였다. 이후, 도 2의 B에 도시한 바와 같이, 50ng의 Input DNA, 1.8x SPRI의 라이게이션 이후 정제 조건이 적용된 PCR 증폭 과정을 포함하는 일련의 실험을 수행하였으며, 이를 통해, 증폭된 전체 절편 중 spike-in 절편의 비율을 산출하였다. 구체적으로, 상기 spike-in 절편의 비율은 정제 횟수(1회 또는 2회), 및 어댑터의 농도(50ng, 또는 5ng)에 따라 비교하였다.

[0045] 도 3은 종래 분자 바코딩 기술이 적용된 PCR 증폭 과정에서, 분자 바코딩의 교체 비율을 정제 횟수 및 어댑터의 농도에 따라 확인한 것이다.

[0046] 그 결과, 도 3에 나타낸 바와 같이, spike-in 절편은 실험 조건에 따라, 전체 분자 바코드에서 약 6% 내지 50%를 차지하였으며, 특히, 정제 횟수가 감소할 수록, 및/또는 바코드 서열을 포함하는 어댑터의 농도가 증가할수록 spike-in 절편의 비율이 높아지는 경향을 보여주었다.

[0047] 이러한 실험 결과는, 종래의 분자 바코딩 기술은 바코드 서열을 포함하는 어댑터가 정제 과정에서 완전하게 제거되지 않고, pre-PCR 시 프라이머로 작용할 경우, 분자 바코딩 교체 현상이 발생할 수 있으므로, 유전체 분석 결과의 오류를 유발할 수 있음을 나타내는 것이다.

[0048] **실시예 2. 바코드 서열의 위치에 의한 분자 바코딩 교체 감소 효과 확인**

[0049] 본 실시예에서는 바코드 서열을 어댑터 말단 영역으로 그 위치를 고정된 어댑터 구조를 디자인하였으며, 이를 통해, 상기 실시예 1에서 언급한 종래 기술의 문제점을 해소할 수 있는지 여부를 확인하고자 하였다. 구체적으로, 도 4에 도시한 바와 같이, 본 실시예에서의 어댑터 서열이 적용된 PCR 증폭 과정에서는, 잘못된 바코드 서열을 포함하는 어댑터가 PCR 증폭 과정에 프라이머로 작용하는 경우, 3' → 5' Exonuclease 활성에 의해 복제가 중단되는 DNA 중합효소 I의 교정(Proofreading) 과정을 통해, 종래의 분자 바코딩 교체 현상을 감소시킬 수 있을지 여부에 대하여 확인하고자 하였다.

[0050] 이를 위하여, 도 5의 A에 도시한 바와 같이, 우선, DNA 절편과 라이게이션이 진행되는, 어댑터의 말단 영역에 바코드 서열이 위치하도록 어댑터를 디자인하였고, 보다 구체적으로, DNA 절편과 라이게이션이 진행되는, 어댑터 최말단 염기 위치를 end point 1(EP1), 상기 EP1을 기준으로 2, 3, 5, 또는 10번째 위치를 각각 EP2, EP3, EP5, 또는 EP10으로 설정한 뒤, 이들 영역을 각각 A, T, G 또는 C 로 고정된 어댑터를 제조하였다. 또한, 바코드 서열 내 특정 서열로 고정된 위치의 해당 서열이 다른 염기로 변경(치환)된 specific sequence synthesis fragment 또는 spike-in 어댑터를 합성하였다. 이후, 도 5의 B에 도시한 바와 같이, 10ng의 Plasmid DNA 일부, 1.0 uM의 Index adaptor working concentration, 0.1uM의 Spike-in working concentration의 조건이 적용된 PCR 증폭 과정을 포함하는 일련의 실험을 수행하였으며, imageJ 프로그램을 사용하여 시퀀싱 피크를 확인한 뒤, 분자 바코드 교체율을 하기의 식 1에 따라 산출하였다. 한편, 본 실험에서 대조군은 변경된 서열이 도입된 spike-in 어댑터가 적용되지 않은 군으로 설정하였다.

[0051] [식 1]

$$\text{분자 바코드 교체율 (\%)} = \frac{\text{Spike-in 어댑터에 고정된 DNA 절편의 양}}{\text{라이게이션된 어댑터에 고정된 DNA 절편의 양}}$$

[0052] [0053] 도 6은 일 구체예에 따른 바코드 서열의 위치가 어댑터 말단 영역으로 고정된 어댑터를 이용한 PCR 증폭 과정에서, 분자 바코딩 교체 수준을 확인한 결과이다. 그 결과, 도 6에 나타낸 바와 같이, EP1 영역에서, 본래 디자인한 염기서열이외, 다른 3 종류의 염기서열 검출 수준은 매우 낮음을 확인하였다(A=0.058).

[0054] 또한, 표 2 및 도 7 내지 11은 일 구체예에 따른 바코드 서열의 위치가 어댑터 말단 영역으로 고정된 어댑터를 이용한 PCR 증폭 과정에서, 바코드 서열 내 특정 위치 및 이의 염기서열이 고정된 경우, 특정 위치 및 고정된 염기서열에 따른 분자 바코딩 교체 수준을 확인한 것이다. 참고로, 도 6 내지 도 11에서, A에 대한 피크는 붉은 실선은 G에 대한 피크는 실선, T에 대한 피크는 이점 쇄선, C에 대한 피크는 점선으로 표기하였다.

표 2

[0055]

뉴클레오타이드 타입			검출량			
고정 위치	라이게이션 adapter	spike-in adapter	라이게이션 adapter (A)	spike-in adapter (B)	교체율 (%) (B/A*100)	교체율 평균 (%)
Control			0.058		0	0
			> LOD		0	0
			> LOD		0	0
			> LOD		0	0
EP1	A	G	0.06		0%	0%
	A	T	0.054		0%	
	A	C	0.082		0%	
	G	A	0.067		0%	
	G	T	0.076		0%	
	G	C	0.074		0%	
	T	G	0.097		0%	
	T	A	0.081		0%	
	T	C	0.087		0%	
	C	G	0.083		0%	
	C	A	0.087		0%	
	C	T	0.079		0%	

EP2	A	G	0.061		0%	
	A	T	0.081		0%	
	A	C	0.074		0%	0%
	G	A	0.061		0%	
	G	T	0.035		0%	
	G	C	0.051		0%	0%
	T	G	0.099		0%	
	T	A	0.096		0%	
	T	C	0.092		0%	0%
	C	G	0.082		0%	
	C	A	0.066		0%	
	C	T	0.108		0%	0%
	EP3	A	G	0.183		0%
A		T	0.142	0.024	17%	
A		C	0.116		0%	6%
G		A	0.196	0.016	8%	
G		T	0.2		0%	
G		C	0.251		0%	3%
T		G	0.129	0.016	12%	
T		A	0.067	0.024	36%	
T		C	0.156		0%	16%
C		G	0.102	0.008	8%	
C		A	0.093	0.041	44%	
C		T	0.092	0.031	34%	29%
EP5		A	G	0.086	0.02	23%
	A	T	0.092	0.021	23%	
	A	C	0.1	0.038	38%	28%
	G	A	0.138	0.042	30%	
	G	T	0.145	0.049	34%	
	G	C	0.167	0.037	22%	29%
	T	G	0.123	0.038	31%	
	T	A	0.123	0.045	37%	
	T	C	0.124	0.023	19%	29%
	C	G	0.074	0.032	43%	
	C	A	0.088	0.024	27%	
	C	T	0.077	0.032	42%	37%
	EP10	A	G	0.072	0.024	33%
A		T	0.079	0.018	23%	
A		C	0.125	0.024	19%	25%
G		A	0.107	0.028	26%	
G		T	0.105	0.017	16%	
G		C	0.121	0.036	30%	24%
T		G	0.09	0.042	47%	
T		A	0.099	0.028	28%	
T		C	0.08	0.023	29%	35%
C		G	0.065	0.028	43%	
C		A	0.065	0.028	43%	
C		T	0.056	0.032	57%	48%

[0056] 그 결과, 표 2 및 도 7 내지 11에 나타난 바와 같이, EP1 및 EP2 위치에서는 분자 바코드 교체 현상이 검출되지 않았던 반면, DNA 절편과 라이게이션이 진행되는 어댑터의 말단 영역에서 멀어질수록 분자 바코드 교체율이 높아지는 경향을 보여주었다. 특히, 이러한 경향은 EP3 위치에서부터 뚜렷하게 나타났으며, EP3 이후 위치에서, 시토신(C)이 도입될 경우 상대적으로 분자 바코드 교체율은 더욱 증가하였다.

[0057] 이러한 실험 결과는 바코드 서열이 어댑터의 말단 영역, 예를 들어, EP1 및 EP2 위치를 포함하는 영역에 존재하

는 경우, PCR 증폭 과정에서 발생하는 분자 바코딩 교체 현상을 감소시킬 수 있음을 나타내는 것이다.

실시예 3. 염기서열의 고정에 의한 분자 바코딩 교체 감소 효과 확인

본 실시예에서는, 실시예 2의 실험 결과에 기초하여, 바코드 서열을 DNA 절편과 라이게이션이 진행되는, 어댑터 말단 영역으로 그 위치를 고정한 어댑터 구조에서, EP1 또는 EP2 위치 염기서열의 고정이 분자 바코딩 교체 현상에 미치는 영향을 확인하고자 하였다.

3-1. 어댑터 말단 첫번째 위치(EP1)

EP1 영역을 A, C, G, 또는 T로 각각 고정하고, EP2 영역은 C 또는 G로 설정하여 디자인된 바코드 서열 및 이를 포함하 어댑터 서열을 합성 및 제조하였고, 상기 실시예 2와 동일한 과정으로 실험을 수행한 뒤, 바코드 서열이 변화하지 않은 어댑터의 비율, 즉, 분자 바코딩 교체가 발생하지 않은 비율을 산출하였다.

표 3은 바코드 서열 내 EP1 위치가 아데닌(A)으로 고정된 경우, 그 결과를 나타낸 것이다.

표 3

		EP2	EP1	A(%)
		Read Pos 1	Read Pos 2	
control	Index 1	C	A	99.62%
	Index 2	G	A	
	Spike-in 1	-	-	
	Spike-in 2	-	-	
1	Index 1	C	A	99.53%
	Index 2	G	A	
	Spike-in 1	C	C	
	Spike-in 2	G	C	
2	Index 1	C	A	99.44%
	Index 2	G	A	
	Spike-in 1	C	T	
	Spike-in 2	G	T	
3	Index 1	C	A	98.28%
	Index 2	G	A	
	Spike-in 1	C	G	
	Spike-in 2	G	G	

표 4는 바코드 서열 내 EP1 위치가 시토신(C)으로 고정된 경우, 그 결과를 나타낸 것이다.

표 4

		EP2	EP1	C(%)
		Read Pos 1	Read Pos 2	
control	Index 1	C	C	99.68%
	Index 2	G	C	
	Spike-in 1	-	-	
	Spike-in 2	-	-	
1	Index 1	C	C	99.14%
	Index 2	G	C	
	Spike-in 1	C	T	
	Spike-in 2	G	T	

2	Index 1	C	C	99.57%
	Index 2	G	C	
	Spike-in 1	C	G	
	Spike-in 2	G	G	
3	Index 1	C	C	99.66%
	Index 2	G	C	
	Spike-in 1	C	A	
	Spike-in 2	G	A	

[0067] 표 5는 바코드 서열 내 EP1 영역이 구아닌(G)으로 고정된 경우, 그 결과를 나타낸 것이다.

표 5

		EP2	EP1	G(%)
		Read Pos 1	Read Pos 2	
control	Index 1	C	G	99.64%
	Index 2	G	G	
	Spike-in 1	-	-	
	Spike-in 2	-	-	
1	Index 1	C	G	99.67%
	Index 2	G	G	
	Spike-in 1	C	C	
	Spike-in 2	G	C	
2	Index 1	C	G	99.53%
	Index 2	G	G	
	Spike-in 1	C	T	
	Spike-in 2	G	T	
3	Index 1	C	G	99.50%
	Index 2	G	G	
	Spike-in 1	C	A	
	Spike-in 2	G	A	

[0069] 표 6은 바코드 서열 내 EP1 위치가 티민(T)으로 고정된 경우, 그 결과를 나타낸 것이다.

표 6

		EP2	EP1	T(%)
		Read Pos 1	Read Pos 2	
control	Index 1	C	T	99.56%
	Index 2	G	T	
	Spike-in 1	-	-	
	Spike-in 2	-	-	
1	Index 1	C	T	99.43%
	Index 2	G	T	
	Spike-in 1	C	C	
	Spike-in 2	G	C	
2	Index 1	C	T	99.50%
	Index 2	G	T	
	Spike-in 1	C	G	
	Spike-in 2	G	G	
3	Index 1	C	T	99.53%

Index 2	G	T
Spike-in 1	C	A
Spike-in 2	G	A

[0071] 도 12는 일 구체예에 따른 바코드 서열의 위치가 어댑터 말단 영역으로 고정된 어댑터를 이용한 PCR 증폭 과정에서, 어댑터 말단 첫번째 염기서열의 고정에 따른 분자 바코딩 교체 수준을 확인한 것이다. 그 결과, 도 12에 나타난 바와 같이, EP1 영역에 고정된 염기서열의 종류와 관계없이 바코드 분자 바코딩 교체율, 즉, 오류율이 1% 미만임을 확인하였다.

[0072] 3-2. 어댑터 말단 두번째 위치(EP2)

[0073] EP2 영역을 A, C, G, 또는 T로 각각 고정하고, EP1 영역은 C, T, G, 또는 A로 설정하여 디자인된 바코드 서열 및 이를 포함하 어댑터 서열을 합성 및 제조하였고, 상기 실시예 2와 동일한 과정으로 실험을 수행한 뒤, 바코드 서열이 변화하지 않은 어댑터의 비율, 분자 바코딩 교체가 발생하지 않은 비율을 산출하였다.

[0074] 표 7 및 표 8은 바코드 서열 내 EP2 위치가 시토신(C)으로 고정된 경우, 그 결과를 나타낸 것이다.

표 7

[0075]

		EP2	EP1	C(%)
		Read Pos 1	Read Pos 2	
control	Index 1	C	C	99.787%
	Index 2	C	T	
	Index 3	C	G	
	Index 4	C	A	
	Spike-in 1	-	-	
	Spike-in 2	-	-	
	Spike-in 3	-	-	
	Spike-in 4	-	-	
1	Index 1	C	C	99.044%
	Index 2	C	T	
	Index 3	C	G	
	Index 4	C	A	
	Spike-in 1	G	C	
	Spike-in 2	G	T	
	Spike-in 3	G	G	
	Spike-in 4	G	A	

표 8

[0076]

		EP2	EP1	C(%)
		Read Pos 1	Read Pos 2	
control	Index 1	C	C	99.167%
	Index 2	C	T	
	Index 3	C	G	
	Index 4	C	A	
	Spike-in 1	-	-	
	Spike-in 2	-	-	
	Spike-in 3	-	-	
	Spike-in 4	-	-	

1	Index 1	C	C	98.963%
	Index 2	C	T	
	Index 3	C	G	
	Index 4	C	A	
	Spike-in 1	G	C	
	Spike-in 2	G	T	
	Spike-in 3	G	G	
	Spike-in 4	G	A	

[0077] 표 9 및 표 10은 바코드 서열 내 EP2 위치가 구아닌(G)으로 고정된 경우, 그 결과를 나타낸 것이다.

표 9

		EP2	EP1	G(%)
		Read Pos 1	Read Pos 2	
control	Index 1	G	C	99.262%
	Index 2	G	T	
	Index 3	G	G	
	Index 4	G	A	
	Spike-in 1	-	-	
	Spike-in 2	-	-	
	Spike-in 3	-	-	
	Spike-in 4	-	-	
1	Index 1	G	C	99.234%
	Index 2	G	T	
	Index 3	G	G	
	Index 4	G	A	
	Spike-in 1	C	C	
	Spike-in 2	C	T	
	Spike-in 3	C	G	
	Spike-in 4	C	A	

표 10

		EP2	EP1	G(%)
		Read Pos 1	Read Pos 2	
control	Index 1	G	C	99.311%
	Index 2	G	T	
	Index 3	G	G	
	Index 4	G	A	
	Spike-in 1	-	-	
	Spike-in 2	-	-	
	Spike-in 3	-	-	
	Spike-in 4	-	-	

1	Index 1	G	C	99.259%
	Index 2	G	T	
	Index 3	G	G	
	Index 4	G	A	
	Spike-in 1	C	C	
	Spike-in 2	C	T	
	Spike-in 3	C	G	
	Spike-in 4	C	A	

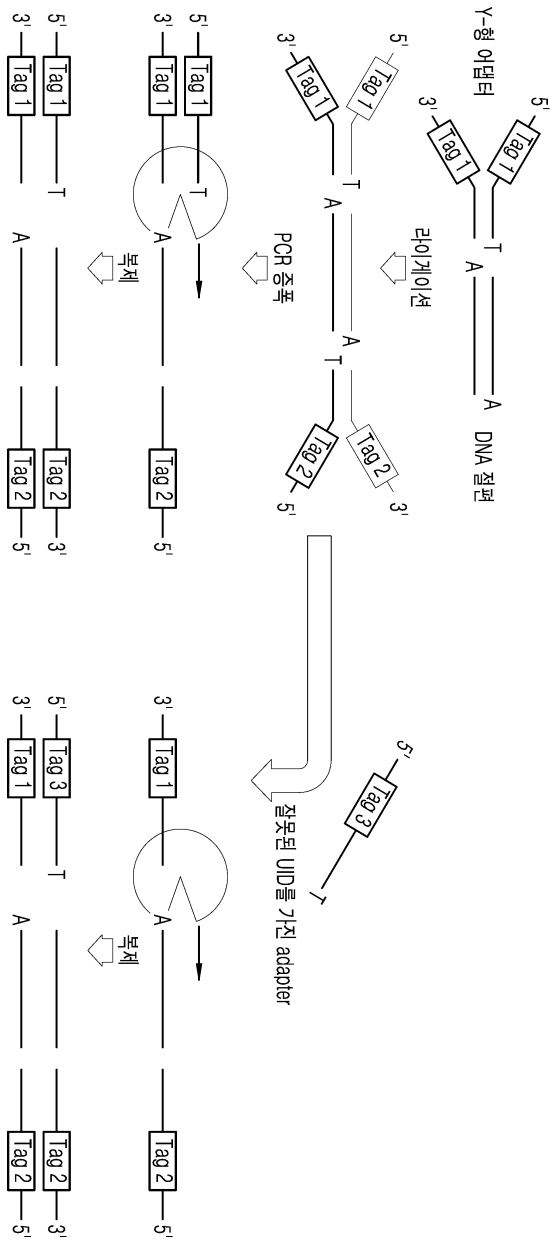
[0080] 도 13는 일 실시예에 따른 바코드 서열의 위치가 어댑터 말단 영역으로 고정된 어댑터를 이용한 PCR 증폭 과정에서, 어댑터 말단 두번째 염기서열의 고정에 따른 분자 바코딩 교체 수준을 확인한 것이다. 그 결과, 도 13에 나타난 바와 같이, EP2 영역에 고정된 염기서열이 C 또는 G인 경우, 분자 바코드 교체율, 즉, 오류율이 1% 미만임을 확인하였다.

[0081] 이러한 실험 결과는 일 실시예에 따른 바코드 서열을 포함하는 어댑터는, 분자 바코딩 현상을 감소시킬 수 있고, 구체적으로, 염기서열 분석을 통한 유전 변이 검출 과정에서 변이 검출의 민감도 및 특이도 향상에 기여할 수 있는바, 정밀 의료 기반의 진단/치료 기술 분야에 폭넓게 활용될 수 있음을 나타내는 것이다.

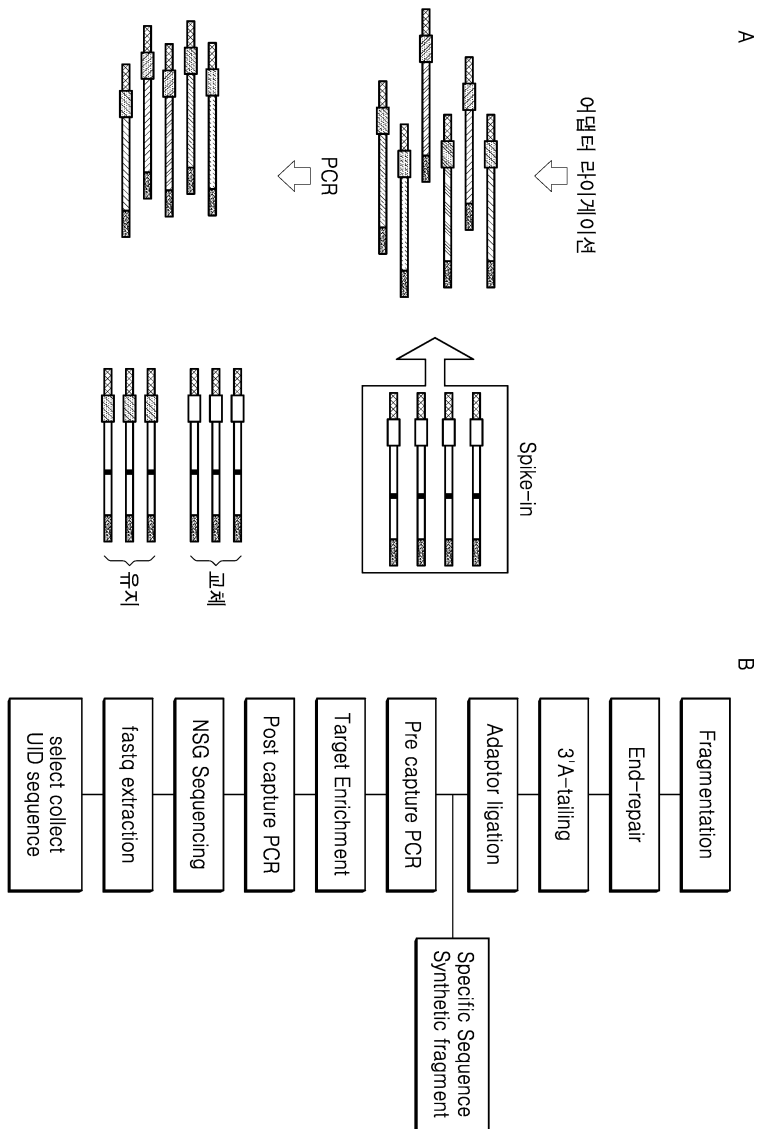
[0082] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다.

도면

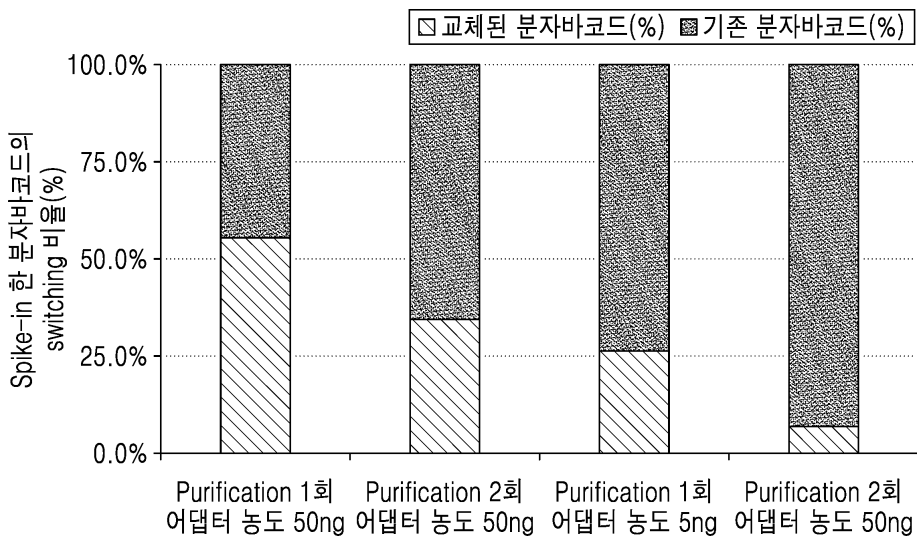
도면1



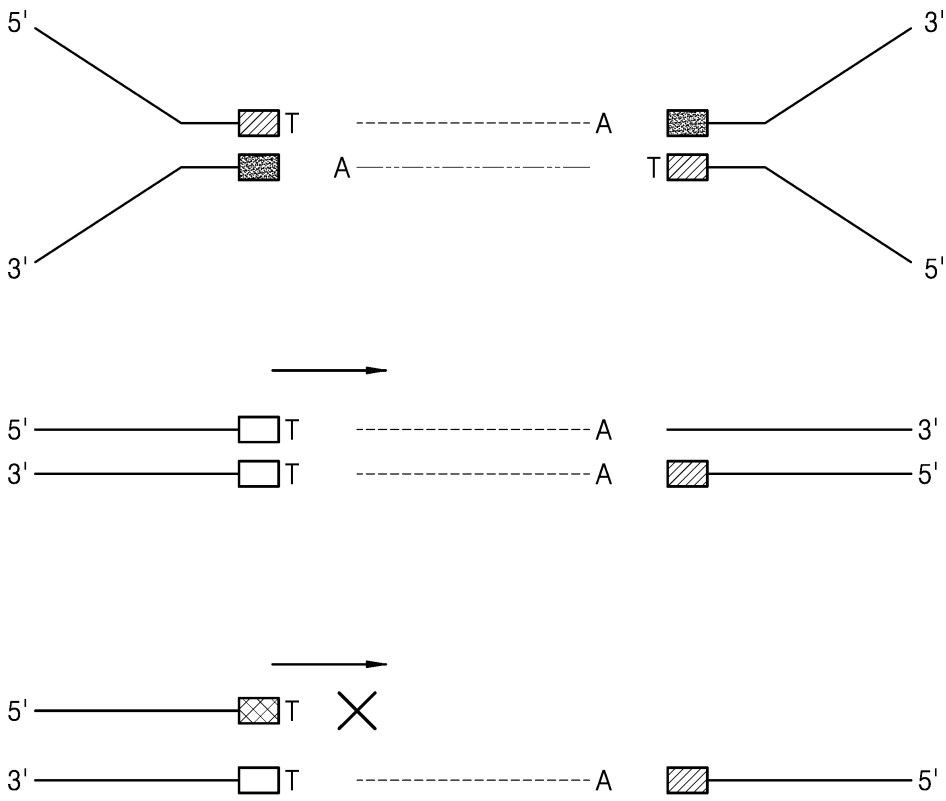
도면2



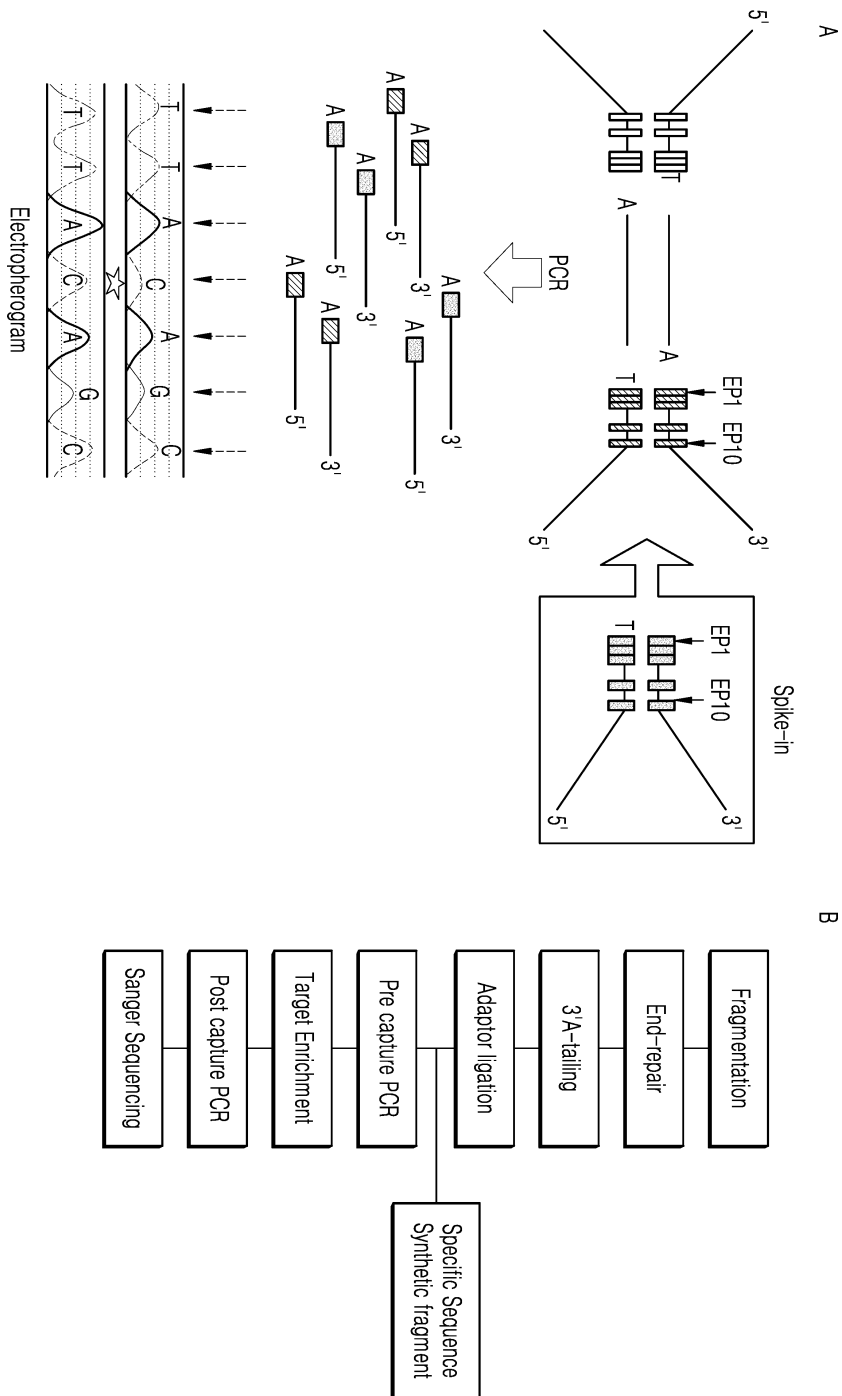
도면3



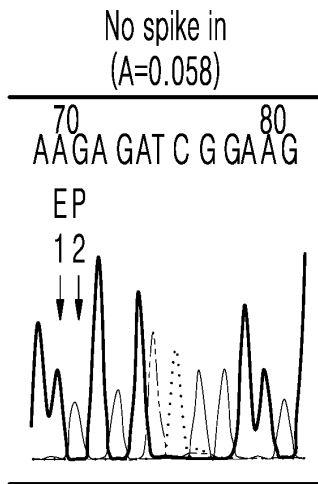
도면4



도면5



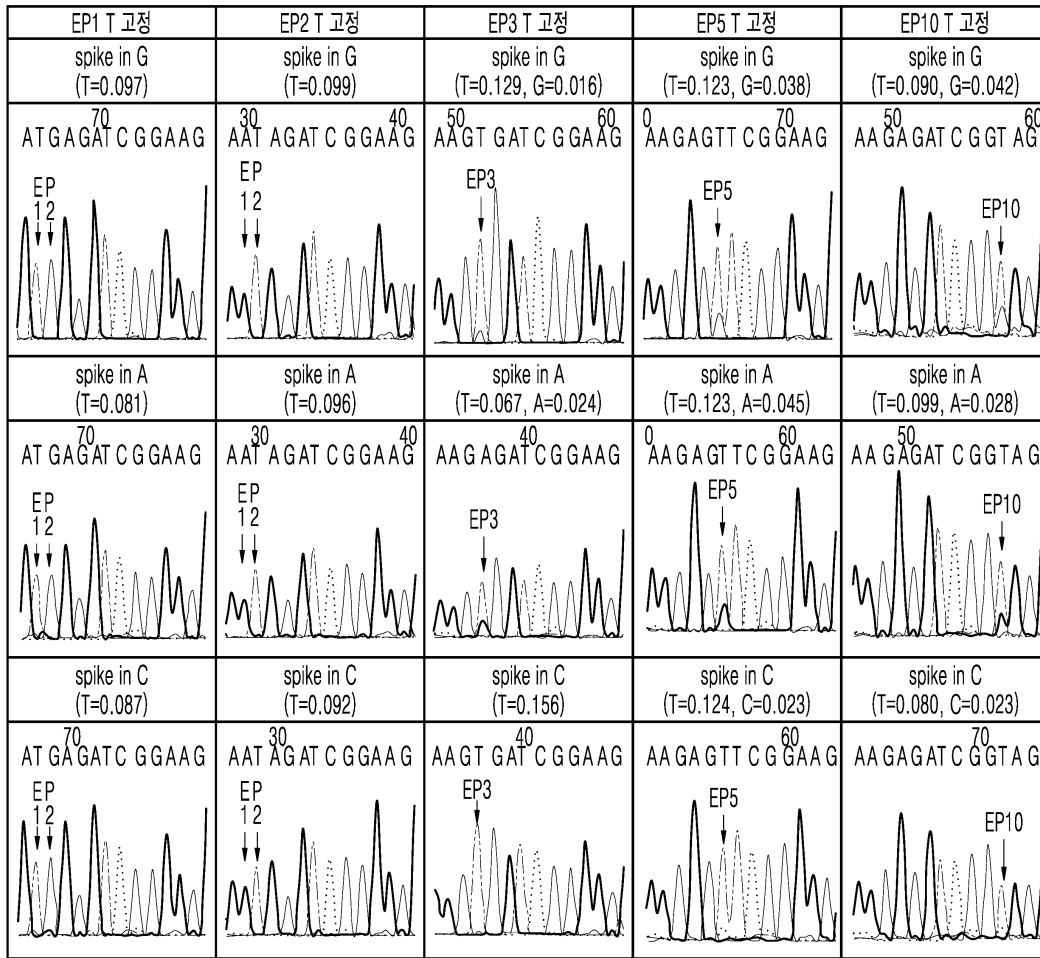
도면6



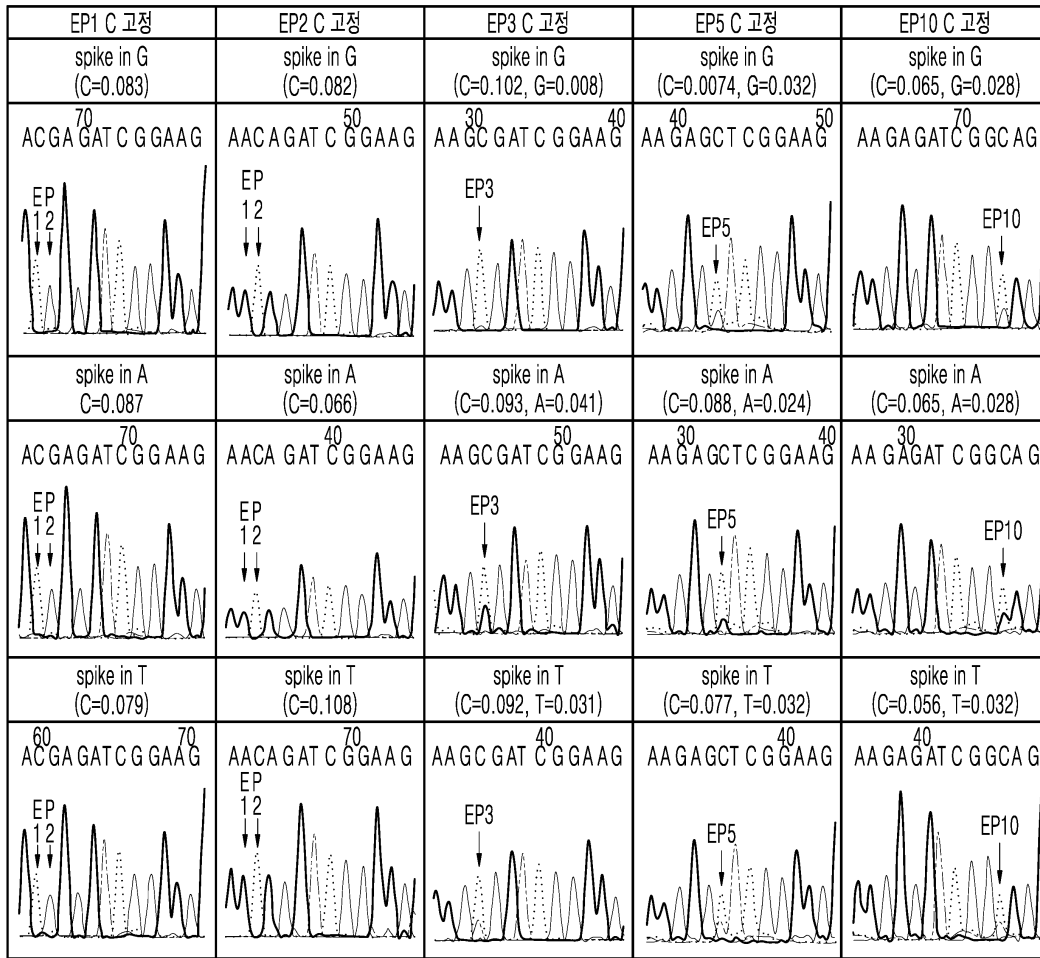
도면7

EP1 A 고정	EP2 A 고정	EP3 A 고정	EP5 A 고정	EP10 A 고정
spike in T (A=0.054)	spike in T (A=0.081)	spike in T (A=0.142, T=0.024)	spike in T (A=0.092, T=0.021)	spike in T (A=0.079, T=0.018)
70 80 AAGA GAT C G GAAG	30 40 AAAAGAT C G GAAG	70 80 AAGA GAT C G GAAG	70 80 AAGAGAT C G GAAG	70 80 AAGAGAT C G GAAG
EP 12	EP 12	EP3	EP 5	EP 10
spike in G (A=0.060)	spike in G (A=0.061)	spike in G (A=0.183)	spike in G (A=0.086, G=0.020)	spike in G (A=0.072, G=0.024)
70 80 AAGAGAT C G GAAG	30 40 AAAAGAT C G GAAG	60 80 AAGAGAT C G GAAG	70 80 AAGAGAT C G GAAG	70 80 AAGAGAT C G GAAG
EP 12	EP 12	EP3	EP 5	EP 10
spike in C (A=0.082)	spike in C (A=0.074)	spike in C (A=0.116)	spike in C (A=0.100, C=0.038)	spike in C (A=0.125, C=0.024)
70 80 AAGA GAT C G GAAG	40 80 AAAAGAT C G GAAG	70 80 AAGAGAT C G GAAG	0 70 AAGAGAT C G GAAG	70 80 AAGAGAT C G GAAG
EP 12	EP 12	EP3	EP5	EP10

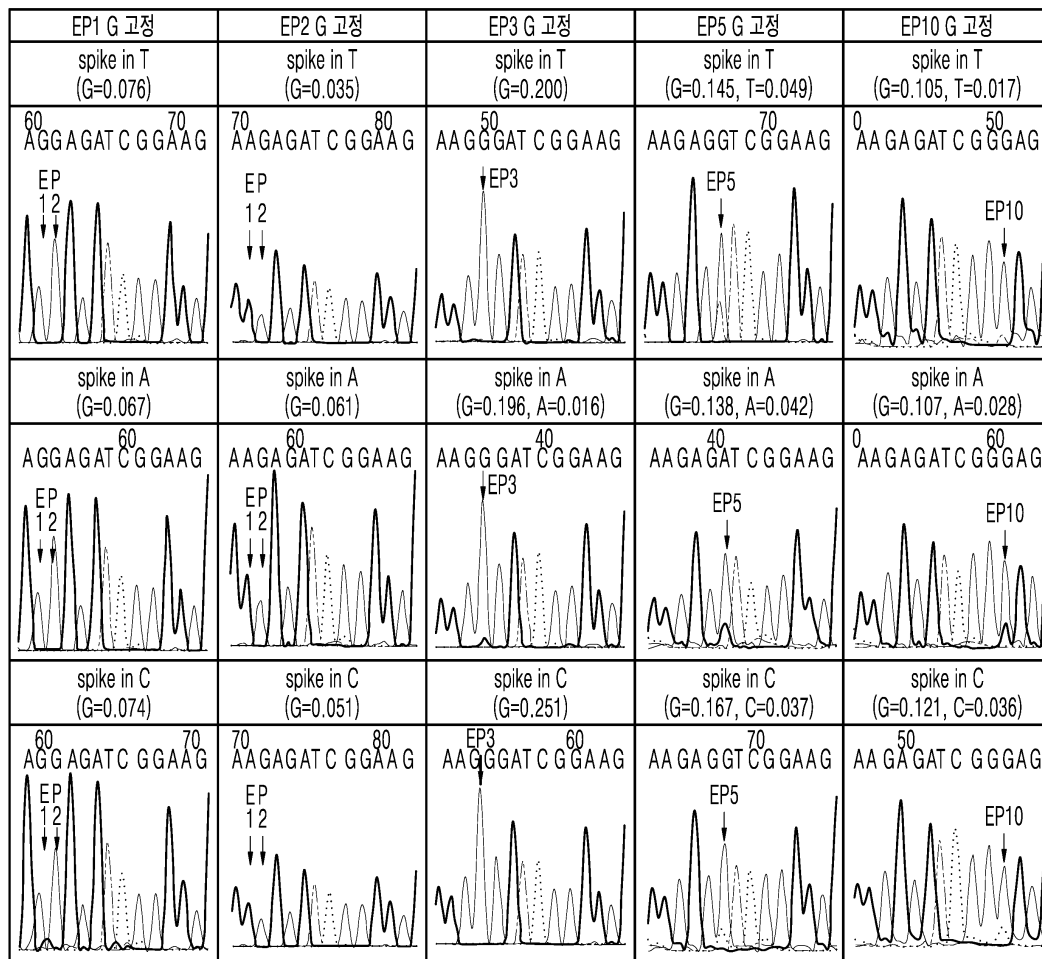
도면8



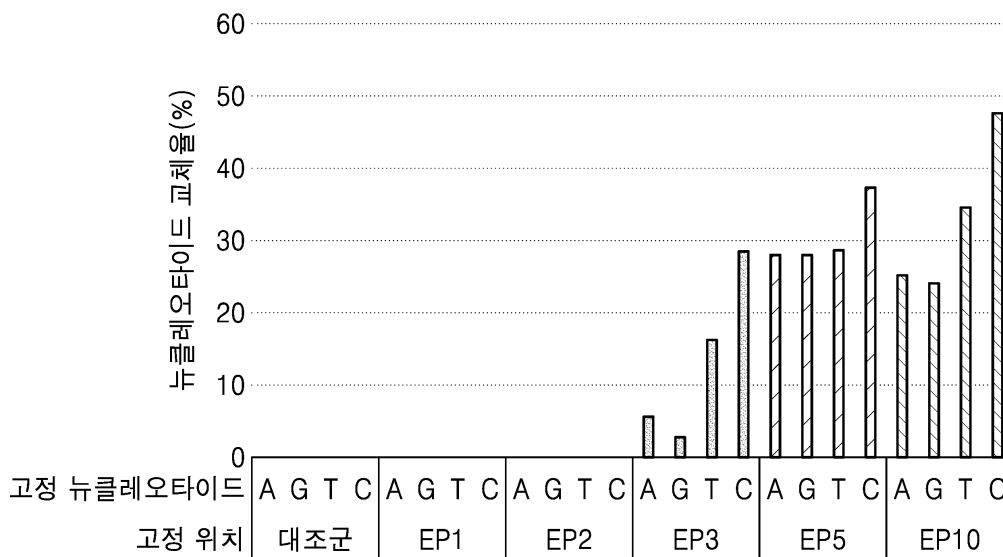
도면9



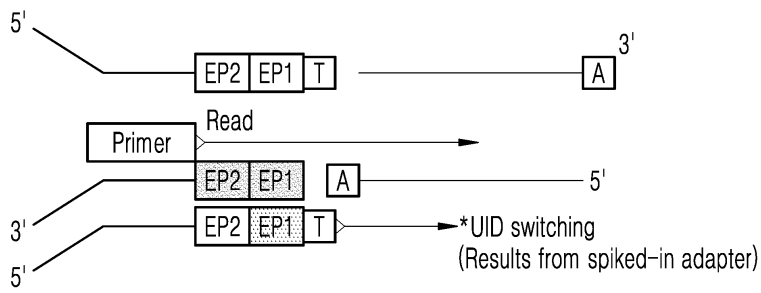
도면10



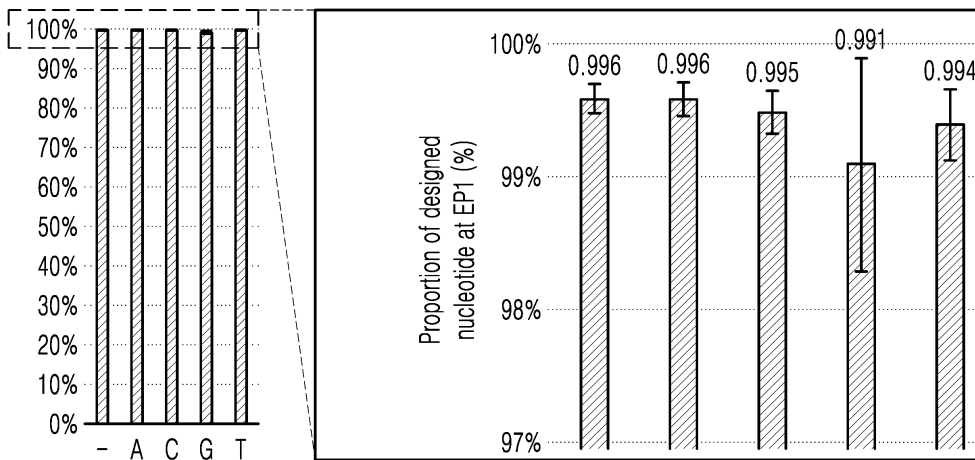
도면11



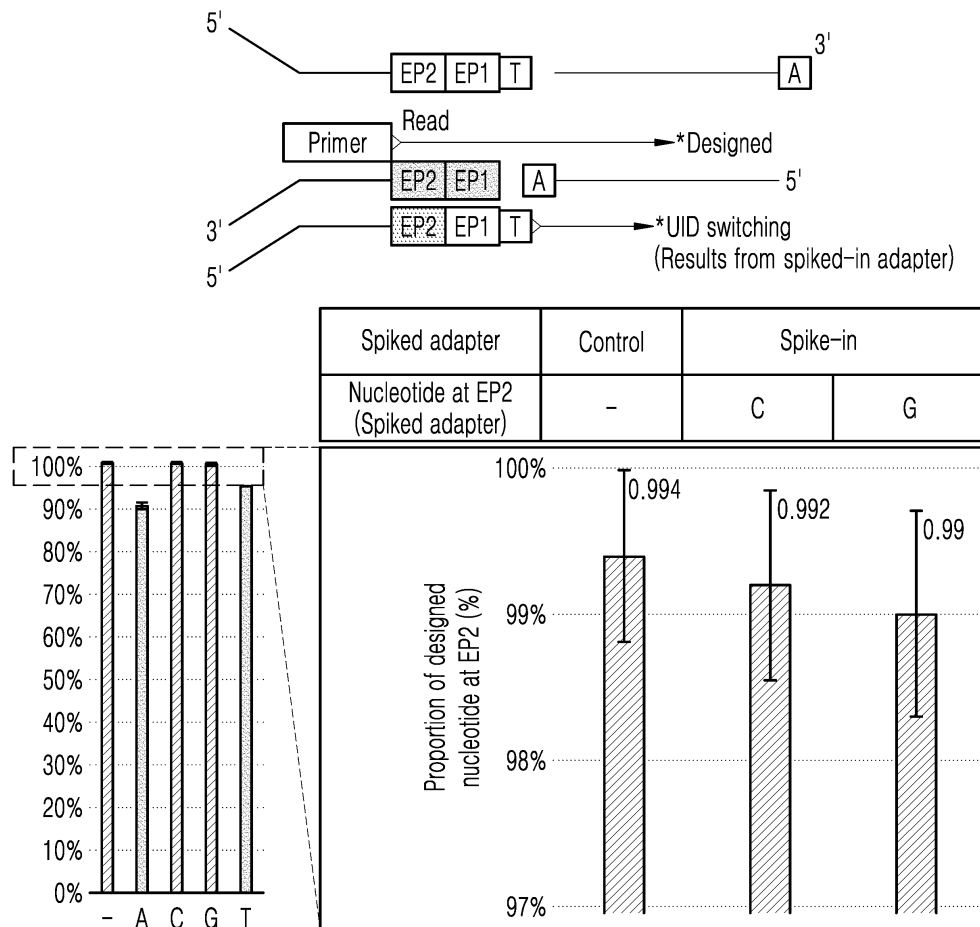
도면12



Spiked adapter	Control	Spike-in			
Nucleotide at EP1 (Spiked adapter)	-	A	C	G	T



도면13



서열 목록

<110> GENINUS Inc.

<120> Composition for improving molecular barcoding efficiency and use thereof

<130> PN137999KR

<160> 6

<170> KoPatent In 3.0

<210> 1

<211> 150

<212> DNA

<213> Artificial Sequence

<220><223> KRAS

<400> 1

cttcacctg agaaggaga aacacagtct ggattattac agtgcacctt ttacttcaa 60

aaagggtgta tatacaactc aacaacaaaa aattcaattt aaaaatgggc aaaggacttg 120

aaaagacatt gttcctgctc caaagacttc 150

<210> 2
 <211> 150
 <212> DNA
 <213> Artificial Sequence
 <220><223> IDH1
 <400> 2

cttcaatggc ttctctgaag accgtgccac ccagaatatt tcgtatggtg ccatttggtg 60
 atttcacat ttgtttcaac ttgaactcct caaccctctt ctcatcagga gtgatagtgg 120
 cacatttgac gccaacatta tgcttccttc 150

<210> 3
 <211> 150
 <212> DNA
 <213> Artificial Sequence
 <220><223> BRCA1
 <400> 3

cttcttctgg ctctcctcg ctcacacttt cttccattgc attatacca gcagtatcag 60

tagtatgagc agcagctgga ctctgggcag attctgcaac tttcaacttt caattgggga 120
 actttcaatg cagaggttga agatggcttc 150

<210> 4
 <211> 150
 <212> DNA
 <213> Artificial Sequence
 <220><223> ALK
 <400> 4

cttcactgat ggaggagtc ttgccagcaa agcagtagtt ggggtttag tcggtcatga 60
 tggtcgaggt gccgagcttg ctcagcttgt actcagggt ctgcagctcc atctgcatgg 120
 cttgcagctc ctggtgcttc cggcggcttc 150

<210> 5

<211> 150
 <212> DNA
 <213> Artificial Sequence

<220><223> ERBB2
<400> 5
cttcgctacg tgctcatcgc tcacaaccaa gtgaggcagg tcccactgca gaggctgcgg 60
attgtgcgag gcacccagct ctttgaggac aactatgccc tggccgtgct agacaatgga 120
gaccgctga acaataccac ccctgtcttc 150
<210> 6
<211> 4
<212> DNA
<213> Artificial Sequence
<220><223> UID
<400> 6
agtc 4