

(45) 공고일자 2022년 11월 30일  
(11) 등록번호 10-2472027  
(24) 등록일자 2022년 11월 24일

(73) 특허권자  
**일루미나 케임브리지 리미티드**  
 영국 케임브리지 씨비21 6디에프 그렛 애빙턴 19  
 그랜타 파크

(72) 발명자  
**스티머스 프랭크 제이.**  
 미국 캘리포니아주 92122 샌디에이고 일루미나 웨  
 이 5200

**장 판**  
 미국 캘리포니아주 92122 샌디에이고 일루미나 웨  
 이 5200  
 (뒷면에 계속)

(74) 대리인  
**특허법인아주김장리**

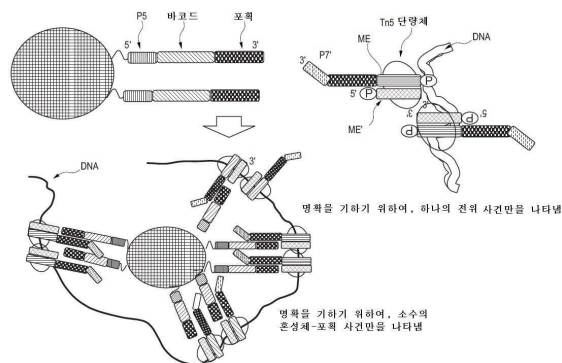
심사관 : 한지혜

(54) 발명의 명칭 근접 보존 전위

(57) 요약

본 명세서에서 제공된 실시형태는 표적 핵산의 바코딩된 DNA 단편의 고정 라이브러리의 제조, 게놈 변이체의 동정, 접근 정보, 페이징 정보 및 표적 핵산의 메틸화 상태의 결정을 위한 방법 및 조성물에 관한 것이다.

대표도 - 도18



(52) CPC특허분류

C12Q 2525/191 (2013.01)

C12Q 2563/179 (2013.01)

C12Q 2565/514 (2013.01)

(72) 발명자

**베틀리 제이슨 리차드**

영국 니어 새프런 월든 에섹스 씨비10 1엑스엘 리틀 체스터포드

**곰리 나일 안토니**

영국 니어 새프런 월든 에섹스 씨비10 1엑스엘 리틀 체스터포드

**건더슨 케빈 엘.**

미국 캘리포니아주 92122 샌디에이고 일루미나 웨이 5200

**플러먼 바우터**

미국 캘리포니아주 92122 샌디에이고 일루미나 웨이 5200

**위어 재클린**

영국 니어 새프런 월든 에섹스 씨비10 1엑스엘 리틀 체스터포드

**아이오아뉴 애브가우스타**

영국 니어 새프런 월든 에섹스 씨비10 1엑스엘 리틀 체스터포드

**젠킨스 가레스**

영국 니어 새프런 월든 에섹스 씨비10 1엑스엘 리틀 체스터포드

**잭슨 로자먼드**

영국 니어 새프런 월든 에섹스 씨비10 1엑스엘 리틀 체스터포드

**모렐 나탈리**

영국 니어 새프런 월든 에섹스 씨비10 1엑스엘 리틀 체스터포드

**포코로크 드미트리 케이.**

미국 캘리포니아주 92122 샌디에이고 일루미나 웨이 5200

**노어버그 스티븐 제이.**

미국 캘리포니아주 92122 샌디에이고 일루미나 웨이 5200

**히 폴리**

미국 캘리포니아주 92122 샌디에이고 일루미나 웨이 5200

**키아 아미탈리**

미국 캘리포니아주 92122 샌디에이고 일루미나 웨이 5200

**고리신 이고르**

미국 위스콘신주 53719 매디슨 스위트 200 리서치파크 블러바드 5602

**판토야 리고**

미국 캘리포니아주 92122 샌디에이고 일루미나 웨이 5200

(30) 우선권주장

62/157,396 2015년05월05일 미국(US)

62/242,880 2015년10월16일 미국(US)

## 명세서

### 청구범위

#### 청구항 1

하기 단계들을 포함하는, 표적 핵산의 바코딩된 DNA 단편의 라이브러리를 제조하는 방법:

- a. 표적 핵산을 복수의 트랜스포좀 복합체(transposome complex)와 접촉시키는 단계로서, 각각의 트랜스포좀 복합체가 트랜스포존(transposon) 및 트랜스포사제(transposase)를 포함하되, 상기 트랜스포존은 전달 가닥 및 비-전달 가닥을 포함하며, 상기 트랜스포좀 복합체의 상기 트랜스포존 중 적어도 하나는 상보성 포획 서열에 혼성화할 수 있는 어댑터 서열을 포함하는, 상기 복수의 트랜스포좀 복합체와 접촉시키는 단계;
- b. 상기 표적 핵산을 복수의 단편으로 단편화시키고, 복수의 전달 가닥을 상기 단편의 적어도 하나의 가닥의 5' 말단에 삽입하는 한편, 상기 표적 핵산의 근접을 유지하는 단계;
- c. 상기 표적 핵산의 상기 복수의 단편을 복수의 고체 지지체와 접촉시키는 단계로서, 상기 복수에서 각각의 상기 고체 지지체는 복수의 고정 올리고뉴클레오타이드를 포함하고, 각각의 상기 올리고뉴클레오타이드는 상보성 포획 서열 및 제1 바코드 서열을 포함하되, 상기 복수의 상기 고체 지지체에서 각각의 고체 지지체로부터의 상기 제1 바코드 서열은 상기 복수의 고체 지지체에서의 다른 고체 지지체로부터의 상기 제1 바코드 서열과 상이한, 상기 복수의 고체 지지체와 접촉시키는 단계;
- d. 상기 바코드 서열 정보를 상기 표적 핵산 단편에 전달함으로써, 이중-가닥 단편의 라이브러리를 생성하는 단계로서, 적어도 하나의 가닥은 상기 제1 바코드로 5'-태그되되, 동일한 표적 핵산의 적어도 두 단편은 동일한 바코드 정보를 수용하는, 상기 라이브러리를 생성하는 단계.

#### 청구항 2

하기 단계들을 포함하는, 표적 핵산 서열의 근접 정보를 결정하는 방법:

- a. 상기 표적 핵산을 복수의 트랜스포좀 복합체와 접촉시키는 단계로서, 각각의 트랜스포좀 복합체는 트랜스포존 및 트랜스포사제를 포함하되, 상기 트랜스포존은 전달 가닥 및 비-전달 가닥을 포함하고, 상기 트랜스포좀 복합체의 상기 트랜스포존 중 적어도 하나는 상보성 포획 서열에 혼성화할 수 있는 어댑터 서열을 포함하는, 상기 복수의 트랜스포좀 복합체와 접촉시키는 단계;
- b. 상기 표적 핵산을 복수의 단편으로 단편화시키고, 복수의 전달 가닥을 삽입하는 한편, 상기 표적 핵산의 근접을 유지하는 단계;
- c. 상기 표적 핵산의 상기 복수의 단편을 복수의 고체 지지체와 접촉시키는 단계로서, 상기 복수에서의 각각의 상기 고체 지지체는 복수의 고정 올리고뉴클레오타이드를 포함하고, 각각의 상기 올리고뉴클레오타이드는 상보성 포획 서열 및 제1 바코드 서열을 포함하되, 상기 복수의 상기 고체 지지체에서 각각의 고체 지지체로부터의 상기 제1 바코드 서열은 상기 복수의 고체 지지체에서의 다른 고체 지지체로부터의 상기 제1 바코드 서열과 상이한, 상기 복수의 고체 지지체와 접촉시키는 단계;
- d. 상기 바코드 서열 정보를 상기 표적 핵산 단편에 전달하는 단계로서, 동일한 표적 핵산 중 적어도 두 단편은 동일한 바코드 정보를 수용하는, 상기 전달하는 단계;
- e. 상기 표적 핵산 단편 서열 및 상기 바코드 서열을 결정하는 단계;
- f. 상기 바코드 서열을 동정함으로써 상기 표적 핵산의 근접 정보를 결정하는 단계.

#### 청구항 3

표적 핵산 서열의 단상형(haplotype) 정보 및 메틸화 상태를 동시에 결정하는 방법으로서:

- a. 상기 표적 핵산을 복수의 트랜스포좀 복합체와 접촉시키는 단계로서, 각각의 트랜스포좀 복합체는 트랜스포존 및 트랜스포사제를 포함하되, 상기 트랜스포존은 전달 가닥 및 비-전달 가닥을 포함하며, 상기 트랜스포좀 복합체의 상기 트랜스포존 중 적어도 하나는 상보성 포획 서열에 혼성화할 수 있는 어댑터 서열을 포함하는, 상

기 복수의 트랜스포존 복합체와 접촉시키는 단계;

b. 상기 표적 핵산을 복수의 단편으로 단편화시키고, 복수의 전달 가닥을 삽입하는 한편, 상기 표적 핵산의 근접을 유지하는 단계;

c. 상기 표적 핵산의 상기 복수의 단편을 복수의 고체 지지체와 접촉시키는 단계로서, 상기 복수에서의 각각의 상기 고체 지지체는 복수의 고정 올리고뉴클레오타이드를 포함하고, 각각의 상기 올리고뉴클레오타이드는 상보성 포획 서열 및 제1 바코드 서열을 포함하되, 상기 복수의 상기 고체 지지체에서 각각의 고체 지지체로부터의 상기 제1 바코드 서열은 상기 복수의 고체 지지체에서의 다른 고체 지지체로부터의 상기 제1 바코드 서열과 상이한, 상기 복수의 고체 지지체와 접촉시키는 단계;

d. 상기 바코드 서열 정보를 상기 표적 핵산 단편에 전달하는 단계로서, 동일한 표적 핵산의 적어도 두 단편은 동일한 바코드 정보를 수용하는, 상기 전달하는 단계;

e. 바코드를 포함하는 상기 표적 핵산 단편에 중아황산염 처리를 실시함으로써 바코드를 포함하는 중아황산염 처리 표적 핵산 단편을 생성하는 단계;

f. 상기 중아황산염 처리 표적 핵산 단편 서열 및 상기 바코드 서열을 결정하는 단계;

g. 상기 바코드 서열을 동정함으로써 상기 표적 핵산의 근접 정보를 결정하는 단계를 포함하되,

상기 서열 정보는 상기 표적 핵산의 메틸화 상태를 나타내고, 상기 근접 정보는 단상형 정보를 나타내는, 표적 핵산 서열의 단상형 정보 및 메틸화를 동시에 결정하는 방법.

#### 청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서, 단일 바코드 서열이 각각 개개의 고체 지지체 상의 상기 복수의 고정 올리고뉴클레오타이드에 존재하는, 방법.

#### 청구항 5

제1항 내지 제3항 중 어느 한 항에 있어서, 상이한 바코드 서열이 각각 개개의 고체 지지체 상의 상기 복수의 고정 올리고뉴클레오타이드에 존재하는, 방법.

#### 청구항 6

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 바코드 서열 정보를 상기 표적 핵산 단편에 전달하는 단계는 결찰에 의하는, 방법.

#### 청구항 7

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 바코드 서열 정보를 상기 표적 핵산 단편에 전달하는 단계는 중합효소 연장에 의하는, 방법.

#### 청구항 8

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 바코드 서열 정보를 상기 표적 핵산 단편에 전달하는 단계는 결찰과 중합효소 연장 둘 다에 의하는, 방법.

#### 청구항 9

제7항에 있어서, 상기 중합효소 연장은 주형으로서 고정된 올리고뉴클레오타이드를 이용하여 DNA 중합효소에 의해 비결찰 트랜스포존 가닥의 3'-말단을 연장시키는 것에 의하는, 방법.

#### 청구항 10

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 어댑터 서열의 적어도 일부는 제2 바코드 서열을 더 포함하는, 방법.

#### 청구항 11

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 트랜스포존 복합체는 다량체이고, 각각의 단량체 단위의 상기

트랜스포존의 상기 어댑터 서열은 상기 동일한 트랜스포좀 복합체에서 다른 단량체 단위와 상이한, 방법.

#### 청구항 12

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 어댑터 서열은 제1 프라이머 결합 서열을 더 포함하는, 방법.

#### 청구항 13

제12항에 있어서, 상기 제1 프라이머 결합 부위는 상기 포획 서열에 대해 또는 상기 포획 서열의 상보체에 대해 서열 상동성이 없는, 방법.

#### 청구항 14

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 고체 지지체 상의 상기 고정 올리고뉴클레오타이드는 제2 프라이머 결합 서열을 더 포함하는, 방법.

#### 청구항 15

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 트랜스포좀 복합체는 다량체이되, 상기 트랜스포좀 단량체 단위는 상기 동일한 트랜스포좀 복합체에서 서로 연결된, 방법.

#### 청구항 16

제15항에 있어서, 트랜스포좀 단량체 단위의 상기 트랜스포사제는 상기 동일한 트랜스포좀 복합체의 다른 트랜스포좀 단량체 단위의 다른 트랜스포사제에 연결된, 방법.

#### 청구항 17

제15항에 있어서, 트랜스포좀 단량체 단위의 상기 트랜스포존은 상기 동일한 트랜스포좀 복합체의 다른 트랜스포좀 단량체 단위의 트랜스포존에 연결된, 방법.

#### 청구항 18

제2항 또는 제3항에 있어서, 표적 핵산 서열의 상기 근접 정보는 단상형 정보를 나타내는, 방법.

#### 청구항 19

제2항 또는 제3항에 있어서, 표적 핵산 서열의 상기 근접 정보는 게놈 변이체를 나타내는, 방법.

#### 청구항 20

제19항에 있어서, 상기 게놈 변이체는 결실, 전좌, 염색체간 유전자 융합, 중복 및 파라로그로 이루어진 군으로부터 선택된, 방법.

#### 청구항 21

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 고체 지지체 상에 고정된 상기 올리고뉴클레오타이드는 부분적 이중 가닥 영역 및 부분적 단일 가닥 영역을 포함하는, 방법.

#### 청구항 22

제21항에 있어서, 상기 올리고뉴클레오타이드의 상기 부분적 단일 가닥 영역은 제2 바코드 서열 및 제2 프라이머 결합 서열을 포함하는, 방법.

#### 청구항 23

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 바코드를 포함하는 표적 핵산 단편은 상기 표적 핵산 단편의 상기 서열을 결정하기 전에 증폭되는, 방법.

#### 청구항 24

제23항에 있어서, 상기 단계 (a) 내지 (d) 및 후속적 증폭은 상기 표적 핵산 단편의 상기 서열을 결정하기 전에

단일 반응 구획에서 수행되는, 방법.

#### 청구항 25

제23항에 있어서, 제3 바코드 서열은 상기 증폭 동안 상기 표적 핵산 단편에 도입된, 방법.

#### 청구항 26

제1항 내지 제3항 중 어느 한 항에 있어서,

복수의 제1 세트의 반응 구획으로부터의 단계 (d)의 상기 바코드를 포함하는 상기 표적 핵산 단편을 상기 바코드를 포함하는 표적 핵산 단편의 풀에 합하는 단계;

상기 바코드를 포함하는 표적 핵산 단편의 풀을 복수의 제2 세트의 반응 구획으로 재분포시키는 단계; 및

서열분석 전에 상기 제2 세트의 반응 구획에서 상기 표적 핵산 단편을 증폭시킴으로써 상기 표적 핵산 단편에 제3 바코드를 도입하는 단계를 더 포함하는, 방법.

#### 청구항 27

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 표적 핵산을 트랜스포좀 복합체와 접촉시키기 전에 상기 표적 핵산을 사전 단편화시키는 단계를 더 포함하는, 방법.

#### 청구항 28

제27항에 있어서, 상기 표적 핵산을 사전 단편화시키는 단계는 음과처리 및 제한 효소로 이루어진 군으로부터 선택된 방법에 의하는, 방법.

#### 청구항 29

태그된 DNA 단편의 고정된 라이브러리의 제조 방법으로서,

(a) 트랜스포좀 복합체가 고정된 복수의 고체 지지체를 제공하는 단계로서, 여기서 상기 트랜스포좀 복합체는 다량체이고, 동일한 트랜스포좀 복합체의 트랜스포좀 단량체 단위는 서로 연결되며, 상기 트랜스포좀 단량체 단위는 제1 폴리뉴클레오타이드에 결합된 트랜스포사제를 포함하고, 상기 제1 폴리뉴클레오타이드는,

(i) 트랜스포존 말단 서열을 포함하는 3' 부분, 및

(ii) 제1 바코드를 포함하는 제1 어댑터를 포함하는, 상기 복수의 고체 지지체를 제공하는 단계

(b) 표적 DNA가 상기 트랜스포좀 복합체에 의해 단편화되고, 상기 제1 폴리뉴클레오타이드의 상기 3' 트랜스포존 말단 서열이 상기 단편의 적어도 하나의 가닥의 5' 말단에 전달되는 조건 하에서 상기 표적 DNA를 상기 복수의 고체 지지체에 적용함으로써; 이중-가닥 단편의 고정된 라이브러리를 생성하는 단계를 포함하되, 적어도 하나의 가닥은 상기 제1 바코드로 5'-태그된, 태그된 DNA 단편의 고정된 라이브러리의 제조 방법.

#### 청구항 30

제29항에 있어서, 트랜스포좀 단량체 단위의 상기 트랜스포사제는 상기 동일한 트랜스포좀 복합체의 다른 트랜스포좀 단량체 단위의 다른 트랜스포사제에 연결된, 태그된 DNA 단편의 고정된 라이브러리의 제조 방법.

#### 청구항 31

제29항에 있어서, 트랜스포좀 단량체 단위의 상기 트랜스포존은 상기 동일한 트랜스포좀 복합체의 다른 트랜스포좀 단량체 단위의 트랜스포존에 연결된, 태그된 DNA 단편의 고정된 라이브러리의 제조 방법.

#### 청구항 32

하기 단계들을 포함하는, 표적 핵산의 메틸화 상태를 결정하기 위한 서열분석 라이브러리의 제조 방법:

a. 상기 표적 핵산을 2 이상의 단편으로 단편화시키는 단계;

b. 제1 공통 어댑터 서열을 상기 표적 핵산의 상기 단편의 5'-말단에 혼입시키는 단계로서, 상기 어댑터 서열은 제1 프라이머 결합 서열 및 친화도 모이어티를 포함하되, 상기 친화도 모이어티는 결합쌍의 하나의 구성원인,

상기 혼입시키는 단계;

c. 상기 표적 핵산 단편을 변형시키는 단계;

d. 고체 지지체 상에 상기 표적 핵산 단편을 고정시키는 단계로서, 상기 고체 지지체는 상기 결합쌍의 또다른 구성원을 포함하고, 상기 표적 핵산의 상기 고정은 상기 결합쌍의 결합에 의하는, 상기 고정시키는 단계;

e. 상기 고정된 표적 핵산 단편에 중아황산염 처리를 실시하는 단계;

f. 제2 공통 어댑터 서열을 상기 중아황산염 처리된 고정 표적 핵산 단편에 혼입시키는 단계로서, 상기 제2 공통 어댑터는 제2 프라이머 결합 부위를 포함하는 단계; 및

g. 제1 프라이머 및 제2 프라이머를 이용하여 고체 지지체 상에 고정된 상기 중아황산염 처리된 표적 핵산 단편을 증폭시킴으로써, 표적 핵산의 상기 메틸화 상태를 결정하기 위해 서열분석 라이브러리를 생성하는 단계.

### 청구항 33

제32항에 있어서, 상기 제1 공통 어댑터 서열은 일측 전위(one-sided transposition)에 의해 상기 표적 핵산의 상기 5'-말단 단편에 혼입된, 표적 핵산의 메틸화 상태를 결정하기 위한 서열분석 라이브러리의 제조 방법.

### 청구항 34

제32항에 있어서, 상기 제1 공통 어댑터 서열은 결합에 의해 상기 표적 핵산의 상기 5'-말단 단편에 혼입된, 표적 핵산의 메틸화 상태를 결정하기 위한 서열분석 라이브러리의 제조 방법.

### 청구항 35

제32항 내지 제34항 중 어느 한 항에 있어서, 상기 제2 공통 어댑터 서열을 상기 중아황산염 처리된 고정 표적 핵산 단편에 혼입시키는 단계는,

(i) 상기 고정된 표적 핵산 단편의 3'-말단이 호모중합체 꼬리를 포함하도록 말단 트랜스퍼라제를 이용하여 상기 고정된 표적 핵산 단편의 상기 3'-말단을 연장시키는 단계;

(ii) 제1 부분 및 제2 부분을 포함하는 올리고뉴클레오타이드를 혼성화시키는 단계로서, 상기 제1 부분은 상기 고정된 표적 핵산 단편의 상기 호모중합체 꼬리에 대해 상보적인 단일 가닥 호모중합체 부분을 포함하고, 상기 제2 부분은 상기 제2 공통 어댑터 서열을 포함하는 이중 가닥 부분을 포함하는, 상기 혼성화시키는 단계;

(iii) 상기 제2 공통 어댑터 서열을 상기 고정된 표적 핵산 단편에 결합시킴으로써, 상기 제2 공통 어댑터 서열을 상기 중아황산염 처리된 고정 표적 핵산 단편에 혼입시키는 단계를 포함하는, 표적 핵산의 메틸화 상태를 결정하기 위한 서열분석 라이브러리의 제조 방법.

### 청구항 36

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 표적 핵산은 단일 세포로부터 유래된, 방법.

### 청구항 37

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 표적 핵산은 단일 세포소기관으로부터 유래된, 방법.

### 청구항 38

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 표적 핵산은 게놈 DNA인, 방법.

### 청구항 39

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 표적 핵산은 다른 핵산에 가교된, 방법.

### 청구항 40

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 표적 핵산은 무 세포 종양 DNA인, 방법.

### 청구항 41

제40항에 있어서, 상기 무 세포 종양 DNA는 태반액으로부터 얻어진, 방법.

#### 청구항 42

제40항에 있어서, 상기 무 세포 종양 DNA는 혈장으로부터 얻어진, 방법.

#### 청구항 43

제42항에 있어서, 상기 혈장은 상기 혈장에 대한 수집 구역을 포함하는 막 분리기를 이용하여 전혈로부터 수집된, 방법.

#### 청구항 44

제43항에 있어서, 상기 혈장에 대한 상기 수집 구역은 고체 지지체 상에 고정된 트랜스포좀 복합체를 포함하는, 방법.

#### 청구항 45

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 표적 핵산은 cDNA인, 방법.

#### 청구항 46

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 표적 핵산은 포말린 고정 파라핀 포매 조직 샘플로부터 유래된, 방법.

#### 청구항 47

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 표적 핵산은 히스톤 보호된 DNA인, 방법.

#### 청구항 48

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 고체 지지체는 비드인, 방법.

#### 청구항 49

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 복수의 고체 지지체는 복수의 비드이되, 상기 복수의 비드는 상이한 크기를 갖는, 방법.

#### 청구항 50

하기 단계들을 포함하는, 표적 핵산의 메틸화를 결정하기 위한 서열분석 라이브러리를 제조하는 방법:

a. 고체 지지체 상에 고정된 트랜스포좀 복합체를 포함하는 복수의 고체 지지체를 제공하는 단계로서, 상기 트랜스포좀 복합체는 트랜스포존 및 트랜스포사제를 포함하되, 상기 트랜스포존은 전달 가닥 및 비-전달 가닥을 포함하고, 상기 전달된 가닥은 하기 (i) 및 (ii)를 포함하며,

(i) 트랜스포사제 인식 서열을 포함하는 3'-말단에서의 제1 부분, 및

(ii) 제1 어댑터 서열 및 결합쌍의 제1 구성원을 포함하는 제1 부분에 대해 5'에 위치한 제2 부분으로서, 상기 결합쌍의 상기 제1 구성원은 상기 고체 지지체 상에서 상기 결합쌍의 제2 구성원에 결합함으로써, 상기 고체 지지체에 상기 트랜스포존을 고정시키며, 상기 제1 어댑터는 제1 프라이머 결합 서열을 포함하는, 제2 부분;

상기 비-전달된 가닥은 하기 (i) 및 (ii)를 포함하는, 상기 복수의 고체 지지체를 제공하는 단계,

(i) 상기 트랜스포사제 인식 서열을 포함하는 5'-말단에서의 제1 부분

(ii) 제2 어댑터 서열을 포함하는 상기 제1 부분에 대해 3'에 위치한 제2 부분으로서, 상기 3'-말단에서 상기 말단의 뉴클레오타이드는 차단되고, 상기 제2 어댑터는 제2 프라이머 결합 서열을 포함하는, 제2 부분;

b. 표적 핵산을 고정된 트랜스포좀 복합체를 포함하는 상기 복수의 고체 지지체와 접촉시키는 단계;

c. 상기 표적 핵산을 복수의 단편으로 단편화시키고, 복수의 전달 가닥을 상기 단편들 중 적어도 하나의 가닥의



상기 5' 말단에 삽입함으로써, 상기 표적 핵산 단편을 상기 고체 지지체에 고정시키는 단계;

d. DNA 중합효소를 이용하여 상기 단편화된 표적 핵산의 상기 3'-말단을 연장시키는 단계;

e. 상기 비-전달된 가닥을 상기 단편화된 표적 핵산의 상기 3'-말단에 결합시키는 단계;

f. 상기 고정된 표적 핵산 단편에 중아황산염 처리를 실시하는 단계;

g. 상기 고정된 표적 핵산 단편의 상기 3'-말단이 호모중합체 꼬리를 포함하도록 DNA 중합효소를 이용함으로써 상기 중아황산염 처리 동안 손상된 상기 고정된 표적 핵산 단편의 상기 3'-말단을 연장시키는 단계;

h. 상기 중아황산염 처리 동안 손상된 상기 고정된 표적 핵산 단편의 상기 3'-말단에 제2 어댑터 서열을 혼입시키는 단계;

i. 제1 프라이머 및 제2 프라이머를 이용하여 고체 지지체 상에 고정된 상기 중아황산염 처리된 표적 핵산 단편을 증폭시킴으로써, 표적 핵산의 상기 메틸화 상태를 결정하기 위한 서열 분석 라이브러리를 생성하는 단계.

#### 청구항 51

하기 단계들을 포함하는, 표적 핵산의 메틸화 상태를 결정하기 위한 서열분석 라이브러리를 제조하는 방법:

a. 상기 표적 핵산을 트랜스포존 복합체와 접촉시키는 단계로서, 상기 트랜스포존 복합체는 트랜스포존 및 트랜스포사제를 포함하되, 상기 트랜스포존은 전달 가닥 및 비-전달 가닥을 포함하고, 상기 전달된 가닥은

(i) 트랜스포사제 인식 서열을 포함하는 3'-말단에서의 제1 부분, 및

(ii) 제1 어댑터 서열 및 결합쌍의 제1 구성원을 포함하는 상기 제1 부분에 대해 5'에 위치한 제2 부분으로서, 상기 결합쌍의 상기 제1 구성원은 상기 결합쌍의 제2 구성원에 결합되는, 제2 부분을 포함하고;

상기 비-전달된 가닥은

(i) 상기 트랜스포사제 인식 서열을 포함하는 상기 5'-말단에서의 제1 부분

(ii) 제2 어댑터 서열을 포함하는 상기 제1 부분에 대해 3'에 위치한 제2 부분으로서, 상기 3'-말단에서 상기 말단의 뉴클레오타이드는 차단되고, 상기 제2 어댑터는 제2 프라이머 결합 서열을 포함하는, 상기 제2 부분을 포함하는, 상기 트랜스포존 복합체와 접촉시키는 단계,

b. 상기 표적 핵산을 복수의 단편으로 단편화시키고, 복수의 전달 가닥을 상기 단편 중 적어도 하나의 가닥의 상기 5' 말단에 삽입함으로써, 상기 표적 핵산 단편을 고체 지지체에 고정시키는 단계;

c. 상기 트랜스포존 말단을 포함하는 표적 핵산 단편을 상기 결합쌍의 제2 구성원을 포함하는 복수의 고체 지지체와 접촉시키는 단계로서, 상기 결합쌍의 상기 제2 구성원에 대한 상기 결합쌍의 상기 제1 구성원의 결합은 상기 고체 지지체에 대해 상기 표적 핵산을 고정시키는, 상기 복수의 고체 지지체와 접촉시키는 단계;

d. DNA 중합효소를 이용하여 상기 단편화된 표적 핵산의 상기 3'-말단을 연장시키는 단계;

e. 상기 단편화된 표적 핵산의 상기 3'-말단에 상기 비-전달된 가닥을 결합시키는 단계;

f. 상기 고정된 표적 핵산 단편에 중아황산염 처리를 실시하는 단계;

g. 상기 고정된 표적 핵산 단편의 상기 3'-말단이 호모중합체 꼬리를 포함하도록 DNA 중합효소를 이용함으로써 상기 중아황산염 처리 동안 손상된 상기 고정된 표적 핵산 단편의 상기 3'-말단을 연장시키는 단계;

h. 상기 중아황산염 처리 동안 손상된 상기 고정된 표적 핵산 단편의 상기 3'-말단에 제2 어댑터 서열을 혼입시키는 단계; 및

i. 제1 프라이머 및 제2 프라이머를 이용하여 고체 지지체 상에 고정된 상기 중아황산염 처리된 표적 핵산 단편을 증폭시킴으로써, 표적 핵산의 상기 메틸화 상태를 결정하기 위한 서열분석 라이브러리를 생성하는 단계.

#### 청구항 52

제50항 또는 제51항에 있어서, 상기 고체 지지체는 비드인, 표적 핵산의 메틸화를 결정하기 위한 서열분석 라이브러리를 제조하는 방법.

### 청구항 53

제50항 또는 제51항에 있어서, 상기 결합쌍의 상기 제1 구성원 및 제2 구성원은 바이오틴 및 스트렙타비딘인, 표적 핵산의 메틸화를 결정하기 위한 서열분석 라이브러리를 제조하는 방법.

### 청구항 54

제50항 또는 제51항에 있어서, 상기 제1 어댑터는 바코드를 더 포함하는, 표적 핵산의 메틸화를 결정하기 위한 서열분석 라이브러리를 제조하는 방법.

### 청구항 55

제50항 또는 제51항에 있어서, 상기 제2 어댑터는 바코드를 더 포함하는, 표적 핵산의 메틸화를 결정하기 위한 서열분석 라이브러리를 제조하는 방법.

### 청구항 56

제50항 또는 제51항에 있어서, 상기 제1 어댑터 및 제2 어댑터는 제1 바코드 및 제2 바코드를 포함하는, 표적 핵산의 메틸화를 결정하기 위한 서열분석 라이브러리를 제조하는 방법.

### 청구항 57

제50항 또는 제51항에 있어서, 상기 중아황산염 처리 동안 손상된 상기 고정된 표적 핵산 단편의 상기 3'-말단을 연장하는 것은 말단 트랜스퍼라제에 의하는, 표적 핵산의 메틸화를 결정하기 위한 서열분석 라이브러리를 제조하는 방법.

### 청구항 58

제50항 또는 제51항에 있어서, 상기 제2 어댑터의 상기 3'-말단에서 상기 말단의 뉴클레오타이드는 다이테옥시 뉴클레오타이드, 인산염 기, 티오인산염 기 및 아지도기로 이루어진 군으로부터 선택된 구성원에 의해 차단되는, 표적 핵산의 메틸화를 결정하기 위한 서열분석 라이브러리를 제조하는 방법.

## 발명의 설명

## 기술 분야

[0001]

### 관련 출원

[0002]

본 출원은 2014년 10월 17일자로 출원된 미국 가출원 특허 제62/065,544호 및 2015년 5월 5일자로 출원된 미국 가출원 특허 제62/157,396호에 대한 우선권을 주장하며, 이들 기초출원은 본 명세서에 그의 전문이 참고로 포함 되어 있다.

[0003]

### 본 발명의 기술분야

[0004]

본 발명의 실시형태는 핵산의 서열분석에 관한 것이다. 특히, 본 명세서에 제공된 방법 및 조성물의 실시형태는 핵산 주형을 제조하는 것 및 그것으로부터 서열 데이터를 얻는 것에 관한 것이다.

## 배경 기술

[0005]

생물학적 샘플에 존재하는 특정 핵산 서열의 검출은 미생물을 동정 및 분류하고, 감염성 질환을 진단하며, 유전자 이상을 검출 및 특성규명하고, 암과 관련된 유전자 변화를 동정하며, 질환에 대한 유전자 감수성을 연구하고, 다양한 치료 유형에 대한 반응을 측정하기 위한 방법으로서 사용되었다. 생물학적 샘플에서 특정 핵산 서열을 검출하기 위한 통상적인 기법은 핵산 서열분석이다.

[0006]

핵산 서열분석 방법은 막심(Maxam)과 길버트(Gilbert)에 의해 사용되는 화학적 분해 방법 및 생어(Sanger)에 의해 사용되는 가닥 신장 방법으로부터 상당히 진화되었다. 오늘날 몇몇 서열분석 방법은 단일 서열분석 실행에서 핵산 모두의 병행 처리를 허용하는 용도이다. 그렇게 해서, 단일 서열분석 실행으로부터 생성된 정보는 엄청날 수 있다.

## 발명의 내용

- [0007] 일 양상에서, 본 명세서에서 표적 핵산의 바코드 DNA 단편의 라이브러리를 제조하는 방법이 기재된다. 상기 방법은 표적 핵산을 복수의 트랜스포존 복합체와 접촉시키는 단계를 포함하며, 각각의 트랜스포좀(transposome) 복합체는 트랜스포존(transposon) 및 트랜스포사제(transposase)를 포함하는데, 이때 트랜스포존은 전달 가닥 및 비전달 가닥을 포함한다. 트랜스포좀 복합체 중 트랜스포존의 적어도 하나는 상보성 포획 서열에 혼성화할 수 있는 어댑터 서열을 포함한다. 표적 핵산은 복수의 단편으로 단편화되고, 단편의 적어도 하나의 단편의 5' 말단에 복수의 전달 가닥을 삽입하는 한편, 표적 핵산의 접근을 유지한다. 표적 핵산의 복수의 단편은 복수의 고체 지지체와 접촉되고, 복수에서 각각의 고체 지지체는 복수의 고정 올리고뉴클레오타이드를 포함하며, 각각의 올리고뉴클레오타이드는 상보성 포획 서열 및 제1 바코드 서열을 포함하되, 복수의 고체 지지체에서의 각각의 고체 지지체로부터의 제1 바코드 서열은 복수의 고체 지지체에서의 다른 고체 지지체로부터의 제1 바코드 서열과 상이하다. 바코드 서열 정보는 표적 핵산 단편에 전달됨으로써, 이중-가닥 단편의 고정 라이브러리를 생성하되, 동일한 표적 핵산의 적어도 두 단편이 동일한 바코드 정보를 수용하도록 적어도 하나의 가닥은 제1 바코드와 5'-태그된다.
- [0008] 일 양상에서, 본 명세서에서 표적 핵산 서열의 접근 정보를 결정하는 방법이 기재된다. 상기 방법은 표적 핵산을 복수의 트랜스포좀 복합체와 접촉시키는 단계를 포함하며, 각각의 트랜스포좀 복합체는 트랜스포존 및 트랜스포사제를 포함하고, 이때 트랜스포존은 전달 가닥 및 비전달 가닥을 포함하며, 트랜스포좀 복합체의 트랜스포존 중 적어도 하나는 상보성 포획 서열에 혼성화할 수 있는 어댑터 서열을 포함한다. 표적 핵산은 복수의 단편으로 단편화되고, 복수의 전달된 가닥은 복수의 단편으로 삽입되는 한편, 표적 핵산의 접근을 유지한다. 표적 핵산의 복수의 단편은 복수의 고체 지지체와 접촉된다. 복수에서 각각의 고체 지지체는 복수의 고정 올리고뉴클레오타이드를 포함하고, 각각의 올리고뉴클레오타이드는 상보성 포획 서열 및 제1 바코드 서열을 포함하되, 복수의 고체 지지체에서 각각의 고체 지지체로부터의 제1 바코드 서열은 복수의 고체 지지체에서 다른 고체 지지체로부터의 제1 바코드 서열과 상이하다. 동일한 표적 핵산 중 적어도 두 단편이 동일한 바코드 정보를 수신하도록 바코드 서열 정보는 표적 핵산 단편에 전달된다. 표적 핵산 단편의 서열 및 바코드 서열이 결정된다. 표적 핵산의 접근 정보는 바코드 서열을 동정함으로써 결정된다. 일부 실시형태에서, 트랜스포좀 복합체의 트랜스포사제는 상보적 포획 서열에 대한 트랜스포존 어댑터 서열의 전위(transposition) 및 후속적 혼성화 후에 제거된다. 일부 실시형태에서, 트랜스포사제는 SDS 처리에 의해 제거된다. 일부 실시형태에서, 트랜스포사제는 프로테아제 처리에 의해 제거된다.
- [0009] 일 양상에서, 본 명세서에서 표적 핵산 서열의 정보 및 메틸화 상태는 표적 핵산 서열의 페이징(phasing) 정보 및 메틸화 상태를 동시에 결정하는 방법이다. 상기 방법은 표적 핵산을 복수의 트랜스포좀 복합체와 접촉시키는 단계를 포함하며, 각각의 트랜스포좀 복합체는 트랜스포존 및 트랜스포사제를 포함하는데, 이때 트랜스포존은 전달 가닥 및 비전달 가닥을 포함하되, 트랜스포좀 복합체의 트랜스포존의 적어도 하나는 상보성 포획 서열에 혼성화할 수 있는 어댑터 서열을 포함한다. 표적 핵산은 복수의 단편으로 단편화되고, 복수의 전달 가닥은 표적 핵산 단편 내로 삽입되는 한편, 표적 핵산의 접근을 유지한다. 표적 핵산의 복수의 단편은 복수의 고체 지지체와 접촉되며, 복수에서 각각의 고체 지지체는 복수의 고정 올리고뉴클레오타이드를 포함하고, 각각의 올리고뉴클레오타이드는 상보성 포획 서열 및 제1 바코드 서열을 포함하되, 복수의 고체 지지체에서의 각각의 고체 지지체로부터의 제1 바코드 서열은 복수의 고체 지지체에서의 다른 고체 지지체로부터의 제1 바코드 서열과 상이하다. 동일한 표적 핵산 중 적어도 두 단편이 동일한 바코드 정보를 수신하도록 바코드 서열 정보는 표적 핵산 단편에 전달된다. 바코드를 포함하는 표적 핵산 단편에 중아황산염 처리가 실시됨으로써, 바코드를 포함하는 중아황산염 처리 표적 핵산 단편을 생성한다. 중아황산염 처리 표적 핵산 단편의 서열 및 바코드 서열이 결정된다. 표적 핵산의 접근 정보는 바코드 서열을 동정함으로써 결정된다.
- [0010] 일 양상에서, 본 명세서에서 태그된 DNA 단편의 고정 라이브러리를 제조하는 방법이 기재된다. 상기 방법은 트랜스포좀 복합체가 고정된 복수의 고체 지지체를 제공함으로써 이중-가닥 단편의 고정된 라이브러리를 생성하는 단계를 포함하는데, 이때 트랜스포좀 복합체는 다량체이고, 동일한 트랜스포좀 복합체의 트랜스포좀 단량체 단위는 서로 연결되되, 상기 트랜스포좀 단량체 단위는 제1 폴리뉴클레오타이드에 결합된 트랜스포사제를 포함하고, 상기 제1 폴리뉴클레오타이드는 (i) 트랜스포존 말단 서열을 포함하는 3' 부분, 및 (ii) 제1 바코드를 포함하는 제1 어댑터를 포함한다. 표적 DNA는 표적 DNA가 트랜스포좀 복합체에 의해 단편화되는 조건 하에서 복수의 고체 지지체에 적용되고, 제1 폴리뉴클레오타이드의 3' 트랜스포존 말단 서열은 단편의 적어도 하나의 가닥의 5' 단부에 전달되고; 적어도 하나의 가닥은 제1 바코드와 5'-태그된다.

- [0011] 일 양상에서, 본 명세서에서 표적 핵산의 메틸화 상태를 결정하기 위해 서열분석 라이브러리를 제조하는 방법이 기재된다. 상기 방법은 표적 핵산을 2 이상의 단편으로 단편화시키는 단계를 포함한다. 제1 공통 어댑터 서열은 표적 핵산의 단편의 5'-말단에 혼입되되, 어댑터 서열은 제1 프라이머 결합 서열 및 친화도 모이어티를 포함하고, 친화도 모이어티는 결합쌍 중 하나의 구성원이다. 표적 핵산 단편은 변성된다. 표적 핵산 단편은 고체 지지체 상에 고정되는데, 이때 고체 지지체는 결합쌍의 다른 구성원을 포함하고, 표적 핵산의 고정은 결합쌍의 결합에 의한 것이다. 고정된 표적 핵산 단편에 중아황산염 처리가 실시된다. 제2 공통 어댑터 서열은 중아황산염 처리된 고정 표적 핵산 단편에 혼입되되, 제2 공통 어댑터는 제2 프라이머 결합 부위를 포함한다. 고체 지지체 상에 고정된 중아황산염 처리 표적 핵산 단편은 증폭됨으로써 표적 핵산의 메틸화 상태를 결정하기 위한 서열분석 라이브러리를 생성한다.
- [0012] 일 양상에서, 본 명세서에서 표적 핵산의 메틸화 상태를 결정하기 위한 서열분석 라이브러리를 제조하는 방법이 기재된다. 상기 방법은 트랜스포존 복합체가 고정된 복수의 고체 지지체를 제공하는 단계를 포함한다. 트랜스포존 복합체는 트랜스포존 및 트랜스포사제를 포함하는데, 이때 트랜스포존은 전달 가닥 및 비-전달 가닥을 포함한다. 전달된 가닥은 (i) 트랜스포사제 인식 서열을 포함하는 3'-말단에서 제1 부분, 및 (ii) 제1 어댑터 서열 및 결합쌍의 제1 구성원을 포함하는 제1 부분에 대해 5'에 위치한 제2 부분을 포함한다. 결합쌍의 제1 구성원은 고체 지지체 상에서 결합쌍의 제2 구성원에 결합함으로써, 고체 지지체에 트랜스포존을 고정시킨다. 제1 어댑터는 또한 제1 프라이머 결합 서열을 포함한다. 비-전달된 가닥은 (i) 트랜스포사제 인식 서열을 포함하는 5'-단부에서 제1 부분 및 (ii) 제2 어댑터 서열을 포함하는 제1 부분에 대해 3'에 위치한 제2 부분을 포함하며, 이때 3'-말단에서 말단의 뉴클레오타이드는 차단된다. 제2 어댑터는 또한 제2 프라이머 결합 서열을 포함한다. 표적 핵산은 고정된 트랜스포존 복합체를 포함하는 복수의 고체 지지체와 접촉된다. 표적 핵산은 복수의 단편으로 단편화되고, 복수의 전달 가닥은 단편의 적어도 하나의 가닥의 5' 말단에 삽입됨으로써, 고체 지지체에 표적 핵산 단편을 고정시킨다. 단편화된 표적 핵산의 3'-말단은 DNA 중합효소에 의해 연장된다. 비-전달된 가닥은 단편화된 표적 핵산의 3'-말단에 결합된다. 고정된 표적 핵산 단편에 중아황산염 처리가 실시된다. 고정된 표적 핵산 단편의 3'-말단이 호모중합체 꼬리를 포함하도록 중아황산염 처리 동안 손상된 고정된 표적 핵산 단편의 3'-말단은 DNA 중합효소를 이용함으로써 연장된다. 제2 어댑터 서열은 중아황산염 처리 동안 손상된 고정된 표적 핵산 단편의 3'-말단에 도입된다. 고체 지지체 상에 고정된 중아황산염 처리 표적 핵산 단편은 제1 프라이머 및 제2 프라이머를 이용하여 증폭됨으로써, 표적 핵산의 메틸화 상태를 결정하기 위한 서열분석 라이브러리를 생성한다.
- [0013] 일 양상에서, 본 명세서에서 표적 핵산의 메틸화 상태를 결정하기 위해 서열분석 라이브러리를 제조하는 방법이 개시된다. 상기 방법은 표적 핵산을 트랜스포존 복합체와 접촉시키는 단계를 포함하는데, 이때 트랜스포존 복합체는 트랜스포존 및 트랜스포사제를 포함한다. 트랜스포존은 전달 가닥 및 비-전달 가닥을 포함한다. 전달된 가닥은 (i) 트랜스포사제 인식 서열을 포함하는 3'-말단에서 제1 부분, 및 (ii) 제1 어댑터 서열 및 결합쌍의 제1 구성원을 포함하는 제1 부분에 대해 5'에 위치한 제2 부분을 포함하되, 결합쌍의 제1 구성원은 결합쌍의 제2 구성원에 결합한다. 비-전달된 가닥은 (i) 트랜스포사제 인식 서열을 포함하는 5'-말단에서 제1 부분 및 (ii) 제2 어댑터 서열을 포함하는 제1 부분에 대해 3'에 위치한 제2 부분을 포함하는데, 이때 3'-말단에서 말단의 뉴클레오타이드는 차단되되, 제2 어댑터는 제2 프라이머 결합 서열을 포함한다. 표적 핵산은 복수의 단편으로 단편화되고, 단편의 적어도 하나의 가닥의 5' 말단에 대해 복수의 전달 가닥을 삽입함으로써, 고체 지지체에 표적 핵산 단편을 고정시킨다. 트랜스포존 말단을 포함하는 표적 핵산 단편은 결합쌍의 제2 구성원을 포함하는 복수의 고체 지지체와 접촉되되, 결합쌍의 제2 구성원에 대한 결합쌍의 제1 구성원의 결합은 고체 지지체에 대해 표적 핵산을 고정시킨다. 단편화된 표적 핵산의 3'-말단은 DNA 중합효소에 의해 연장된다. 비-전달된 가닥은 단편화된 표적 핵산의 3'-말단에 결합된다. 고정된 표적 핵산 단편에 중아황산염 처리가 실시된다. 고정된 표적 핵산 단편의 3'-말단이 호모중합체 꼬리를 포함하도록 중아황산염 처리 동안 손상된 고정된 표적 핵산 단편의 3'-말단은 DNA 중합효소를 이용함으로써 연장된다. 제2 어댑터 서열은 중아황산염 처리 동안 손상된 고정된 표적 핵산 단편의 3'-말단에 도입된다. 고체 지지체 상에 고정된 중아황산염 처리된 표적 핵산 단편은 제1 프라이머 및 제2 프라이머를 이용하여 증폭됨으로써, 표적 핵산의 메틸화 상태를 결정하기 위한 서열분석 라이브러리를 생성한다.
- [0014] 일부 실시형태에서, 제2 어댑터의 3'-말단에서 말단의 뉴클레오타이드는 다이데옥시 뉴클레오타이드, 인산염기, 티오인산염기, 및 아지도기로 이루어진 군으로부터 선택된 구성원에 의해 차단된다.
- [0015] 일부 실시형태에서, 친화도 모이어티는 결합쌍의 구성원일 수 있다. 일부 경우에, 변형된 핵산은 결합쌍의 제1 구성원을 포함할 수 있고, 포획 프로브는 결합쌍의 제2 구성원을 포함할 수 있다. 일부 경우에, 포획 프로브는

고체 표면에 고정될 수 있고, 변형된 핵산은 결합쌍의 제1 구성원을 포함할 수 있으며, 포획 프로브는 결합쌍의 제2 구성원을 포함할 수 있다. 이러한 경우에, 결합쌍의 제1 구성원과 제2 구성원의 결합은 고체 표면에 변형된 핵산을 고정시킨다. 결합쌍의 예는 바이오틴-아비딘, 바이오틴-스트렙타비딘, 바이오틴-뉴트라비딘, 리간드-수용체, 호르몬-수용체, 렉틴-당단백질, 올리고뉴클레오타이드-상보성 올리고뉴클레오타이드, 및 항원-항체를 포함하지만, 이들로 제한되지 않는다.

[0016] 일부 실시형태에서, 제1 공통 어댑터 서열은 일측 전위(one-sided transposition)에 의해 표적 핵산의 5'-말단 단편에 혼입된다. 일부 실시형태에서, 제1 공통 어댑터 서열은 결찰에 의해 표적 핵산의 5'-말단 단편에 혼입된다. 일부 실시형태에서, 제2 공통 어댑터 서열을 중아황산염 처리된 고정 표적 핵산 단편 내로 혼입시키는 것은 (i) 호모중합체 꼬리를 포함하도록 말단 트랜스퍼라제를 이용하여 고정된 표적 핵산 단편의 3'-말단을 연장시키는 것; (ii) 단일 가닥 호모중합체 부분을 포함하는 올리고뉴클레오타이드 및 제2 공통 어댑터 서열을 포함하는 이중 가닥 부분을 혼성화시키되, 단일 삽입된 호모중합체 부분은 호모중합체 꼬리에 대해 상보성인 것; 및 (iii) 고정된 표적 핵산 단편에 제2 공통 어댑터 서열을 결찰시킴으로써, 제2 공통 어댑터 서열을 중아황산염 처리된 고정 표적 핵산 단편 내로 혼입시키는 것을 포함한다.

[0017] 일부 실시형태에서, 표적 핵산은 단일 세포로부터 유래된다. 일부 실시형태에서, 표적 핵산은 단일 세포소기관으로부터 유래된다. 일부 실시형태에서, 표적 핵산은 게놈 DNA이다. 일부 실시형태에서, 표적 핵산은 다른 핵산에 가교된다. 일부 실시형태에서, 표적 핵산은 포말린 고정 파라핀 포매(formalin fixed paraffin embedded: FFPE) 샘플로부터 유래된다. 일부 실시형태에서, 표적 핵산은 단백질과 가교된다. 일부 실시형태에서, 표적 핵산은 DNA와 가교된다. 일부 실시형태에서, 표적 핵산은 히스톤 보호된 DNA이다. 일부 실시형태에서, 히스톤은 표적 핵산으로부터 제거된다. 일부 실시형태에서, 표적 핵산은 무 세포 종양 DNA이다. 일부 실시형태에서, 무 세포 종양 DNA는 태반액으로부터 얻어진다. 일부 실시형태에서, 무 세포 종양 DNA는 혈장으로부터 얻어진다. 일부 실시형태에서, 혈장은 혈장에 대한 수집 구역을 포함하는 막 분리기를 이용하여 전혈로부터 수집된다. 일부 실시형태에서, 혈장에 대한 수집 구역은 고체 지지체 상에 고정된 트랜스포좀 복합체를 포함한다. 일부 실시형태에서, 표적 핵산은 cDNA이다. 일부 실시형태에서, 고체 지지체는 비드이다. 일부 실시형태에서, 복수의 고체 지지체는 복수의 비드이고, 복수의 비드는 상이한 크기를 가진다.

[0018] 일부 실시형태에서, 단일 바코드 서열은 각각의 개개 고체 지지체 상에서 복수의 고정 올리고뉴클레오타이드에 존재한다. 일부 실시형태에서, 상이한 바코드 서열은 각각 개개의 고체 지지체 상에서 복수의 고정 올리고뉴클레오타이드에 존재한다. 일부 실시형태에서, 표적 핵산 단편에 대한 바코드 서열 정보의 전달은 결찰에 의한다. 일부 실시형태에서, 표적 핵산 단편에 대한 바코드 서열 정보의 전달은 중합효소 연장에 의한다. 일부 실시형태에서, 표적 핵산 단편에 대한 바코드 서열 정보의 전달은 결찰과 중합효소 연장 둘 다에 의한다. 일부 실시형태에서, 중합효소 연장은 주형으로서 결찰 고정된 올리고뉴클레오타이드를 이용하여 DNA 중합효소에 의해 비결찰 트랜스포존 가닥의 3'-말단을 연장시키는 것에 의한다. 일부 실시형태에서, 어댑터 서열의 적어도 일부는 제2 바코드 서열을 추가로 포함한다.

[0019] 일부 실시형태에서, 트랜스포좀 복합체는 다량체이되, 각각의 단량체 단위의 트랜스포존의 어댑터 서열은 동일한 트랜스포좀 복합체 내 다른 단량체 단위와 상이하다. 일부 실시형태에서, 어댑터 서열은 제1 프라이머 결합 서열을 추가로 포함한다. 일부 실시형태에서, 제1 프라이머 결합 부위는 포획 서열에 또는 포획 서열의 상보체에 대해 서열 상동성을 갖지 않는다. 일부 실시형태에서, 고체 지지체 상에서 고정 올리고뉴클레오타이드는 제2 프라이머 결합 서열을 추가로 포함한다.

[0020] 일부 실시형태에서, 트랜스포좀 복합체는 다량체이고, 트랜스포좀 단량체 단위는 동일한 트랜스포좀 복합체에서 서로 연결된다. 일부 실시형태에서, 트랜스포좀 단량체 단위의 트랜스포사제는 동일한 트랜스포좀 복합체의 트랜스포사제에 연결된다. 일부 실시형태에서, 트랜스포좀 단량체 단위의 트랜스포존은 동일한 트랜스포좀 복합체의 다른 트랜스포좀 단량체 단위의 트랜스포존에 연결된다. 일부 실시형태에서, 트랜스포좀 단량체 단위의 트랜스포사제는 공유 결합에 의해 동일한 트랜스포좀 복합체의 다른 트랜스포좀 단량체 단위의 트랜스포사제에 연결된다. 일부 실시형태에서, 하나의 단량체 단위의 트랜스포사제는 이황화 결합에 의해 동일한 트랜스포좀 복합체의 다른 트랜스포좀 단량체 단위의 트랜스포사제에 연결된다. 일부 실시형태에서, 트랜스포좀 단량체 단위의 트랜스포존은 공유 결합에 의해 동일한 트랜스포좀 복합체의 다른 트랜스포좀 단량체 단위의 트랜스포존에 연결된다.

[0021] 일부 실시형태에서, 표적 핵산 서열의 접근 정보는 단상형 정보를 나타낸다. 일부 실시형태에서, 표적 핵산 서열의 접근 정보는 게놈 변이체를 나타낸다. 일부 실시형태에서, 게놈 변이체는 결실, 전좌, 염색체간 유전자 용



합, 중복 및 파라로그로 이루어진 군으로부터 선택된다. 일부 실시형태에서, 고체 지지체 상에 고정된 올리고뉴클레오타이드는 부분적 이중 가닥 영역 및 부분적 단일 가닥 영역을 포함한다. 일부 실시형태에서, 올리고뉴클레오타이드의 부분적 단일 가닥 영역은 제2 바코드 서열 및 제2 프라이머 결합 서열을 포함한다. 일부 실시형태에서, 바코드를 포함하는 표적 핵산 단편은 표적 핵산 단편의 서열을 결정하기 전에 증폭된다. 일부 실시형태에서, 후속적 증폭은 표적 핵산 단편의 서열을 결정하기 전에 단일 반응 구획에서 수행된다. 일부 실시형태에서, 제3 바코드 서열은 증폭 동안 표적 핵산 단편에 도입된다.

[0022] 일부 실시형태에서, 상기 방법은 복수의 제1 세트의 반응 구획으로부터의 바코드를 포함하는 표적 핵산 단편을 바코드를 포함하는 표적 핵산 단편의 풀에 합하는 단계; 복수의 제2 세트의 반응 구획에 바코드를 포함하는 표적 핵산 단편의 풀을 재분포시키는 단계; 및 서열분석 전에 제2 세트의 반응 구획에서 표적 핵산 단편을 증폭시킴으로써 표적 핵산 단편에서 제3 바코드를 도입하는 단계를 추가로 포함할 수 있다.

[0023] 일부 실시형태에서, 상기 방법은 표적 핵산을 트랜스포좀 복합체와 접촉시키기 전에 표적 핵산을 사전 단편화시키는 단계를 추가로 포함할 수 있다. 일부 실시형태에서, 표적 핵산을 사전 단편화시키는 단계는 음파처리 및 제한 분해로 이루어진 군으로부터 선택된 방법에 의한다.

### 도면의 간단한 설명

[0024] 도 1은 비드 표면에 트랜스포좀을 결합하는 방법의 예의 흐름도를 도시한 도면.  
 도 2는 도 1 방법의 단계를 그림으로 도시한 도면.  
 도 3은 비드 표면 상에서 태그화 과정의 예의 개략적 다이어그램을 도시한 도면.  
 도 4는 도 3의 비드-기반 태그화 과정으로부터의 클러스터 수에 관한 DNA 수율의 예의 데이터 표를 도시한 도면.  
 도 5는 균일한 크기에 대한 도 3의 비드-기반 태그화 과정의 재현성의 다른 예의 데이터 표를 도시한 도면.  
 도 6A 및 도 6B는 도 5의 지표 샘플의 풀 1의 삽입 크기 플롯 및 풀 2의 삽입 크기 플롯을 도시한 도면.  
 도 7은 도 5에 총 판독 수의 재현성 및 기재된 실험에 대해 정렬된 판독 백분율의 막대 그래프를 도시한 도면.  
 도 8A, 도 8B 및 도 8C는 엑솜 풍부 분석에서, 대조군 라이브러리에서 삽입 크기의 플롯, 비드 기반 태그화된 라이브러리에서 삽입 크기의 플롯 및 요약 데이터 표를 각각 도시한 도면.  
 도 9A, 도 9B 및 도 9C는 엑손 풍부 분석에서 각각 dups PF 분석의 막대 그래프, 선택 염기 분석의 막대 그래프, 및 표적 상에서의 PCT 이용 가능한 염기의 막대 그래프를 각각 도시한 도면.  
 도 10은 비드 표면 상에 트랜스포좀 복합체를 형성하는 방법의 예의 흐름도를 도시한 도면.  
 도 11, 도 12 및 도 13은 도 10의 방법의 단계를 그림으로 도시한 도면.  
 도 14는 도 13에 나타난 트랜스포좀 코팅 비드를 이용하는 태그화 과정의 개략적 다이어그램을 도시한 도면.  
 도 15는 고체 지지체 상에서 트랜스포좀을 형성하는 예시적인 도식을 도시한 도면.  
 도 16은 독특한 지표를 지니는 인접하여-연결된 라이브러리를 생성하는 예시적인 도식을 도시한 도면.  
 도 17은 독특한 지표를 지니는 인접하여-연결된 라이브러리를 생성하는 예시적인 도식을 도시한 도면.  
 도 18 및 도 19는 CPT-DNA가 비드 주위를 감싸는 단일 클론 지표 비드 상의 단일 CPT-DNA의 포획을 도시한 도면.  
 도 20은 결찰 및 겔 충전에 의해 태그화된 DNA에 대해 고체 표면 상에 고정된 Y-어댑터를 연결하는 예시적인 도식을 도시한 도면.  
 도 21은 고체 지지체 상에서 고정 올리고뉴클레오타이드에 대한 CPT-DNA의 결찰 동안 이러한 Y-어댑터를 제조하는 예시적인 도식을 도시한 도면.  
 도 22는 크기 배제 크로마토그래피에 의해 인접하여-연결된 라이브러리로부터의 자유 트랜스포좀의 제거를 나타내는 아가로스 겔 전기이동을 도시한 도면.

- 도 23은 특정 DNA 단편의 샷건 서열 라이브러리를 생성하는 예시적 도식을 도시한 도면.
- 도 24는 클론 지표 서열분석 라이브러리로부터의 서열 정보를 어셈블링하는 예시적 도식을 도시한 도면.
- 도 25는 비드 상의 포획 프로브 밀도의 최적화 결과를 도시한 도면.
- 도 26은 분자내 혼성화에 의해 비드 상에서 CPT-DNA의 지표 서열분석 라이브러리의 실행가능성을 시험하는 결과를 도시한 도면.
- 도 27은 클론 지표화의 실행 가능성을 시험하는 결과를 도시한 도면.
- 도 28은 태그화 후 주형 핵산에 대한 관독의 이웃하는 정렬 섬(aligned island) 내의(내에서) 및 또한 사이의(간의) 특정 거리에 대한 서열분석 관독의 빈도를 나타내는 그래프를 도시한 도면.
- 도 29A 및 도 29B는 고체 지지체 상에서 접근 정보를 유도하기 위한 예시적인 접근을 도시한 도면.
- 도 30 및 도 31은 단일 반응 용기(원 포트(one pot))에서의 지표화된 클론 비드 전위 및 전위 결과의 개략도를 도시한 도면.
- 도 32는 5'- 또는 3'-바이오틴일화된 올리고뉴클레오타이드를 이용하여 비드 상에서 클론 트랜스포좀을 생성하는 개략도를 도시한 도면.
- 도 33은 비드 상에서의 트랜스포좀에 대한 라이브러리 크기를 도시한 도면.
- 도 34는 삽입 크기에 대한 트랜스포좀 표면 밀도의 효과를 도시한 도면.
- 도 35는 크기 분포에 대한 입력 DNA의 효과를 도시한 도면.
- 도 36은 비드 기반 및 용액 기반 태그화 반응을 이용하는 섬 크기 및 분포를 도시한 도면.
- 도 37은 각각 독특한 지표를 수신하는 몇몇의 개개 DNA 분자의 클론 지표화를 도시한 도면.
- 도 38은 전혈로부터 혈장을 분리시키기 위한 장치의 다이어그램을 도시한 도면.
- 도 39 및 도 40은 혈장을 분리하기 위한 장치 및 분리된 혈장의 후속적 용도의 다이어그램을 도시한 도면.
- 도 41은 게놈의 특정 영역을 강화함으로써 표적화된 페이지징의 예시적인 도식을 도시한 도면.
- 도 42는 엑손 사이에서 SNP를 이용하는 엑솜 페이지징의 예시적인 도식을 도시한 도면.
- 도 43은 동시 페이지징 및 메틸화 검출의 예시적인 도식을 도시한 도면.
- 도 44는 동시 페이지징 및 메틸화 검출의 대안의 예시적인 도식을 도시한 도면.
- 도 45는 단일 분석에서 다양한 크기의 클론으로 지표화된 비드를 이용하여 다양한 크기의 라이브러리를 생성하는 예시적인 도식을 도시한 도면.
- 도 46은 상이한 길이 척도 라이브러리를 이용하여 유전자 변이체를 결정하는 예시적인 도식을 도시한 도면.
- 도 47a 및 도 47b는 염색체 1에서 60kb 이형 접합적 결실의 검출 결과를 도시한 도면.
- 도 48은 본 출원의 방법을 이용하는 유전자 융합의 검출 결과를 도시한 도면.
- 도 49는 본 발명의 방법을 이용하는 유전자 결실의 검출 결과를 도시한 도면.
- 도 50은 중아황산염 전환 전의 그리고 후의 ME 서열을 도시한 도면.
- 도 51은 중아황산염 전환 효율 최적화 결과를 도시한 도면.
- 도 52는 IVC 플롯에서 중아황산염 전환 후의 결과를 도시한 도면(개개 염기 당 강도 대 주기).
- 도 53은 BSC 후 PCR 후의 지표-연관 라이브러리의 아가로스 겔 전기이동의 이미지를 도시한 도면.
- 도 54는 크기-선택 없이 강화 전에 전체-게놈 지표 연결된 CPT-seq 라이브러리의 생물분석기 추적을 도시한 도면.
- 도 55는 강화 후에 라이브러리의 아가로스 겔 분석을 도시한 도면.

도 56은 염색체에서 HLA 영역에 대한 표적화된 단상형의 적용 결과를 도시한 도면.

도 57은 ME 스와핑(ME swapping)의 일부 가능한 메커니즘을 도시한 도면.

도 58은 ME 스와핑의 일부 가능한 메커니즘을 도시한 도면.

도 59는 Cys으로 치환될 수 있는 예시적인 아미노산 잔기 Asp468, Tyr407, Asp461, Lys459, Ser458, Gly462, Ala466, Met470을 지니는 Tn5 트랜스포사제의 일부를 도시한 도면.

도 60은 시스테인 잔기가 두 단량체 단위 사이에 이황화 결합을 형성할 수 있도록, S458C, K459C 및 A466C의 아미노산 치환을 지니는 Tn5 트랜스포사제의 일부를 도시한 도면.

도 61은 아민 코팅된 나노입자를 이용하여 이량체 트랜스포사제(dTnp)-나노입자(NP) 바이오컨쥬게이트(dTnp-NP)를 제조하고, 이용하는 예시적인 도식을 도시한 도면.

도 62는 아민 코팅된 고체 지지체에 대한 트랜스포좀 이량체의 컨쥬게이션의 예시적인 도식을 도시한 도면.

도 63은 트랜스포존 단부가 연결된 Mu 트랜스포좀 복합체를 도시한 도면.

도 64는 위유전자(pseudogene)의 조립/페이징에 대해 지표화된 관련 판독의 다이어그램 및 더 짧은 단편을 이용하여 위유전자에서 변이체를 동정하는 것의 단점을 도시한 도면.

도 65는 4개의 별개의 실험으로부터의 그리고 스와핑된 지표의 %로서 나타난 지표 교환 플롯을 도시한 도면.

도 66은 Ts-Tn5 적정의 단편 크기의 애질런트 생물분석기(Agilent BioAnalyzer) 분석을 도시한 도면.

도 67은 중아황산염 처리 후 파손된 라이브러리 요소의 회수를 위한 효율적 방법을 이용하여 Epi-CPTSeq 프로토콜의 DNA 수율을 개선시키기 위한 예시적인 도식을 도시한 도면.

도 68A 내지 도 68C는 중아황산염 처리 후에 파손된 라이브러리 요소의 회수를 위해 효율적 방법을 이용하여 Epi-CPTSeq 프로토콜의 DNA 수율을 개선시키기 위한 몇몇 예시적인 도식을 도시한 도면.

도 69는 무작위 프라이머 연장을 이용하는 주형 구조(rescue)에 대한 예시적인 도식을 도시한 도면.

도 70은 중황산나트륨 전환 동안 DNA 라이브러리의 단편화를 도시한 도면. 좌측 패널은 자기 비드 상에서 태그화된 DNA의 일부의 중황산염 전환 동안 단편화를 도시한다. 우측 패널은 CPTSeq 및 Epi-CPTSeq(Me-CPTSeq) 라이브러리의 생물분석기 추적을 나타낸다.

도 71은 TdT 매개된 ssDNA 결합 반응의 예시적인 도식 및 결과를 도시한 도면.

도 72는 중황산나트륨 전환 비드 결합 라이브러리의 TdT 매개 회수의 도식 및 결과를 도시한 도면. 좌측 패널은 TdT 매개 결합 반응을 이용하는 손상된 중아황산염 전환 DNA 라이브러리의 구조 흐름도를 도시한 도면. DNA 라이브러리 구조 실험의 결과를 우측 패널에 나타낸다.

도 73은 메틸-CPTSeq 분석의 결과를 도시한 도면.

도 74는 DNA의 비드 기반 중아황산염 전환의 예시적인 도식을 도시한 도면.

도 75A 내지 도 75B는 중아황산염 전환 효율 최적화 결과를 도시한 도면.

### 발명을 실시하기 위한 구체적인 내용

- [0025] 본 발명의 실시형태는 핵산의 서열분석에 관한 것이다. 특히, 본 명세서에 제공된 방법 및 조성물의 실시형태는 핵산 주형을 제조하는 것 및 그로부터 서열 데이터를 얻는 것에 관한 것이다.
- [0026] 일 양상에서, 본 발명은 태그화된 표적 핵산 라이브러리의 구성을 위해 고체 지지체 상에서 표적 핵산을 태그화(단편화 및 태그)하는 방법에 관한 것이다. 일 실시형태에서, 고체 지지체는 비드이다. 일 실시형태에서, 표적 핵산은 DNA이다.
- [0027] 일 양상에서, 본 발명은 고체-지지체의 방법 및 조성물, 표적 핵산의 접근 정보를 유도할 수 있는 트랜스포사제-기반 방법에 관한 것이다. 일부 실시형태에서, 상기 조성물 및 방법은 조립/페이징 정보를 유도할 수 있다.
- [0028] 일 양상에서, 본 발명은 고체 지지체 상에 인접하여-연결된, 전위된, 표적 핵산을 포획하는 것에 의해 접근 정보를 유도하는 방법 및 조성물에 관한 것이다.



- [0029] 일 양상에서 본 명세서에 개시된 조성물 및 방법은 게놈 변이체의 분석에 관한 것이다. 예시적인 게놈 변이체는 결실, 염색체간 전좌, 중복, 파라로그, 염색체간유전자 융합을 포함하지만, 이들로 제한되지 않는다. 일부 실시형태에서, 본 명세서에 개시된 조성물 및 방법은 게놈 변이체의 페이징 정보를 결정하는 것에 관한 것이다.
- [0030] 일 양상에서, 본 명세서에 개시된 조성물 및 방법은 표적 핵산의 페이징 특정 영역에 관한 것이다. 일 실시형태에서, 표적 핵산은 DNA이다. 일 실시형태에서, 표적 핵산은 게놈 DNA이다. 일부 실시형태에서, 표적 핵산은 RNA이다. 일부 실시형태에서, RNA는 mRNA이다. 일부 실시형태에서, 표적 핵산은 상보성 DNA(cDNA)이다. 일부 실시형태에서, 표적 핵산은 단일 세포로부터 유래된다. 일부 실시형태에서, 표적 핵산은 순환 종양 세포로부터 유래된다. 일부 실시형태에서, 표적 핵산은 무세포 DNA이다. 일부 실시형태에서, 표적 핵산은 무 세포 종양 DNA이다. 일부 실시형태에서, 표적 핵산은 포말된 고정 파라핀 포매 조직 샘플로부터 유래된다. 일부 실시형태에서, 표적 핵산은 가교된 표적 핵산이다. 일부 실시형태에서, 표적 핵산은 단백질에 가교된다. 일부 실시형태에서, 표적 핵산은 핵산에 가교된다. 일부 실시형태에서, 표적 핵산은 히스톤-보호된 DNA이다. 일부 실시형태에서, 히스톤-보호된 DNA는 히스톤에 대한 항체를 이용하여 세포 용해물로부터 침전되고, 히스톤은 제거된다.
- [0031] 일부 양상에서, 지표화된 라이브러리는 임상적으로 지표화된 비드를 이용하여 표적 핵산으로부터 생성된다. 일부 실시형태에서, DNA를 표적화하기 위해 트랜스포사제가 여전히 결합되지만, 태그화된 표적 핵산은 클론으로 지표화된 비드를 이용하여 포획될 수 있다. 일부 실시형태에서, 표적 핵산 내 관심 대상의 특정 영역을 포획하기 위해 특정 포획 프로브가 사용된다. 표적 핵산의 포획 영역은 다양한 엄격함에서 세척되지만, 선택적으로 증폭된 다음 서열분석될 수 있다. 일부 실시형태에서, 포획 프로브는 바이오틴일화될 수 있다. 지표화된 표적 핵산의 특정 영역에 혼성화된 바이오틴일화된 포획 프로브의 복합체는 스트랩타비딘 비드를 이용함으로써 분리될 수 있다. 표적화된 페이징의 예시적인 도식을 도 41에 나타낸다.
- [0032] 일부 양상에서, 본 명세서에 개시된 조성물 및 방법은 페이징 엑손을 사용할 수 있다. 일부 실시형태에서, 엑손, 프로모터는 풍부할 수 있다. 마커, 예를 들어, 엑손 영역 사이의 이형 접합적 SNP는, 특히 엑손 사이의 거리가 클 때 엑손 페이징에 도움을 줄 수 있다. 예시적인 엑손 페이징은 도 42에 나타낸다. 일부 실시형태에서, 지표화된 연결된 판독은 동시에 이웃하는 엑손의 이형 접합적 SNP를 걸칠 수(뒤덮을 수) 없다. 그렇게 해서, 2 이상의 엑손을 페이징하는 것은 도전적이다. 본 명세서에 개시된 조성물 및 방법은 또한 엑손 예를 들어, 페이징 엑손 1 내지 SNP1 및 SNP2 내지 엑손 2 사이에서 이형 접합적 SNP를 강화한다. 그렇게 해서, SNP 1의 사용을 통해, 엑손 1 및 엑손 2는 도 42에 나타낸 바와 같이 페이징될 수 있다.
- [0033] 일 양상에서, 본 명세서에 개시된 조성물 및 방법은 페이징 및 동시 메틸화 검출을 위해 사용될 수 있다. 중아 황산염 전환(BSC)을 통한 메틸화 검출은 BSC 반응이 DNA 상에서 너무 강하기 때문에 도전적이며, DNA를 단편화하고, 접근/페이징 정보를 제거한다. 또한, 본 출원에 개시된 방법은 추가적인 정제 단계가 전통적인 BSC 접근에 필요한 것에 대해 필요하기 때문에 추가적인 이점을 가지며, 이에 의해 수율을 개선시킨다.
- [0034] 일 양상에서, 본 명세서에 개시된 조성물 및 방법은 단일 분석에서 상이한 크기 라이브러리를 제조하는 데 사용될 수 있다. 일부 실시형태에서, 클론으로 지표화된 비드의 상이한 크기는 상이한 크기 라이브러리를 제조하는 데 사용될 수 있다. 도 1은 비드 표면에 트랜스포좀을 결합하는 예시적인 방법(100)의 흐름도를 도시한다. 트랜스포좀은 트랜스포존 올리고뉴클레오타이드, 트랜스포사제 및 고체상 상에 첨가될 수 있는 임의의 화학물질을 이용하여 비드 표면에 결합될 수 있다. 일 예에서, 트랜스포좀은 바이오틴-스트랩타비딘 결합 복합체를 통해 비드 표면에 결합된다. 방법(100)은 다음의 단계들을 포함하지만, 이들로 제한되지 않는다.
- [0035] 일 실시형태에서, 트랜스포존은 서열분석 프라이머 결합 부위를 포함할 수 있다. 서열 결합 부위의 예시적인 서열은 AATGATACGGCGACCGAGATCTACAC(P5 서열) 및 CAAGCAGAAGACGCATACGAGAT(P7 서열)을 포함하지만, 이들로 제한되지 않는다. 일부 실시형태에서, 트랜스포존은 바이오틴일화될 수 있다.
- [0036] 도 1의 단계(110)에서, P5 및 P7 바이오틴일화된 트랜스포존이 생성된다. 트랜스포존은 또한 하나 이상의 지표 서열(독특한 식별자)을 포함할 수 있다. 예시적인 지표 서열은 TAGATCGC, CTCTCTAT, TATCCTCT, AGAGTAGA, GTAAGGAG, ACTGCATA, AAGGAGTA, CTAAGCCT를 포함하지만, 이들로 제한되지 않는다. 다른 예에서, 단지 P5 또는 단지 P7 트랜스포존이 바이오틴일화된다. 또 다른 예에서, 트랜스포존은 단지 모자이크 말단(mosaic end: ME) 서열 또는 ME 서열 + P5 및 P7 서열이 아닌 추가적인 서열을 포함한다. 이 예에서, P5 및 P7 서열은 후속 PCR 증폭 단계에 첨가된다.
- [0037] 도 1의 단계(115)에서, 트랜스포좀이 조립된다. 조립된 트랜스포좀은 P5 및 P7 트랜스포좀의 혼합물이다. P5 및 P7 트랜스포좀의 혼합물은 도 11 및 도 12를 참고로 하여 더 상세하게 기재된다.

- [0038] 도 1의 단계(120)에서, P5/P7 트랜스포좀 혼합물은 비드 표면에 결합된다. 이 예에서, 비드는 스트렙타비딘 코팅된 비드이고, 트랜스포좀은 바이오틴-스트렙타비딘 결합 복합체를 통해 비드 표면에 결합된다. 비드는 다양한 크기를 가질 수 있다. 일 예에서, 비드는 2.8 $\mu$ m 비드일 수 있다. 다른 예에서, 비드는 1 $\mu$ m 비드일 수 있다. 1 $\mu$ m 비드의 현탁액(예를 들어, 1 $\mu$ l)은 트랜스포좀 결합에 대해 용적 당 큰 표면적을 제공한다. 트랜스포좀 결합을 위해 이용 가능한 표면적 때문에, 반응 당 태그화 산물의 수는 증가된다.
- [0039] 도 2는 도 1의 방법(100)의 단계(110, 115 및 120)를 그림으로 나타낸다. 이 예에서, 트랜스포존은 이중가닥으로서 나타낸다. 다른 예에서(도시하지 않음), 다른 구조, 예컨대 헤어핀, 즉, 이중가닥을 형성할 수 있는 자기상보성 영역을 지니는 단일 올리고뉴클레오타이드가 사용될 수 있다.
- [0040] 방법(100)의 단계(110)에서, 복수의 바이오틴일화된 P5 트랜스포존(210a) 및 복수의 P7 트랜스포존(210b)이 생성된다. P5 트랜스포존(210a) 및 P7 트랜스포존(210b)은 바이오틴일화된다.
- [0041] 방법(100)의 단계(115)에서, P5 트랜스포존(210a) 및 P7 트랜스포존(210b)은 트랜스포사제 Tn5(215)와 혼합되어 복수의 조립된 트랜스포좀(220)을 형성한다.
- [0042] 방법(100)의 단계(120)에서, 트랜스포좀(220)은 비드(225)에 결합된다. 비드(225)는 스트렙타비딘 코팅된 비드이다. 트랜스포좀(220)은 바이오틴-스트렙타비딘 결합 복합체를 통해 비드(225)에 결합된다.
- [0043] 일 실시형태에서, 트랜스포좀의 혼합물은 도 10, 11, 12 및 13에 나타내는 바와 같이 비드 표면과 같은 고체 지지체 상에 형성될 수 있다. 이 예에서, P5 및 P7 올리고뉴클레오타이드는 트랜스포좀 복합체의 조립 전에 비드 표면에 처음 결합된다.
- [0044] 도 3은 비드 표면 상에서 예시적인 태그화 과정(300)의 개략적 다이어그램을 도시한다. 과정(300)에서 트랜스포좀(220)이 결합된 도 2의 비드(225)를 나타낸다. DNA 용액(310)은 비드(225)의 현탁액에 첨가된다. DNA(310)가 트랜스포좀(220)과 접촉함에 따라, DNA는 태그화(단편화 및 태그)되고, 트랜스포좀(220)을 통해 비드(225)에 결합된다. 결합 및 태그화된 DNA(310)는 용액(무-비드) 중에서 앰플리콘(315)의 풀을 생성하도록 PCR 증폭될 수 있다. 앰플리콘(315)은 유세포(320)의 표면에 전달될 수 있다. 클러스터 생성 프로토콜(예를 들어, 클러스터 생성을 위해 사용될 수 있는 브릿지 증폭 프로토콜 또는 임의의 다른 증폭 프로토콜)은 유세포(320)의 표면 상에서 복수의 클러스터(325)를 생성하는 데 사용될 수 있다. 클러스터(325)는 태그화된 DNA(310)의 클론 증폭 산물이다. 클러스터(325)는 이제 서열분석 프로토콜에서 다음 단계의 준비가 된다.
- [0045] 다른 실시형태에서, 트랜스포좀은 임의의 고체 표면, 예컨대 마이크로퓨지(microfuge) 관의 벽에 결합될 수 있다.
- [0046] 비드 표면 상에서 트랜스포좀 복합체의 혼합물을 형성하는 다른 실시형태에서, 올리고뉴클레오타이드는 트랜스포좀 조립 전에 비드 표면에 처음 결합된다. 도 10은 비드 표면 상에서 트랜스포좀 복합체를 형성하는 예시적인 방법(1000)의 흐름도를 도시한다. 방법(1000)은 다음의 단계들을 포함하지만, 이들로 제한되지 않는다.
- [0047] 단계(1010)에서, P5 및 P7 올리고뉴클레오타이드는 비드 표면에 결합된다. 일 예에서, P5 및 P7 올리고뉴클레오타이드는 바이오틴일화되고, 비드는 스트렙타비딘 코팅된 비드이다. 이 단계는 또한 도 11의 개략적 다이어그램(1100)에서 그림으로 나타낸다. 이제 도 11에 관해, P5 올리고뉴클레오타이드(1110) 및 P7 올리고뉴클레오타이드(1115)는 비드(1120) 표면에 결합된다. 이 예에서, 단일 P5 올리고뉴클레오타이드(1110) 및 단일 P7 올리고뉴클레오타이드(1115)는 비드(1120) 표면에 결합되지만, 다수의 P5 올리고뉴클레오타이드(1110) 및/또는 P7 올리고뉴클레오타이드(1115)는 복수의 비드(1120)의 표면에 결합될 수 있다. 일 예에서, P5 올리고뉴클레오타이드(1110)는 P5 프라이머 서열, 지표 서열(독특한 식별자), 관독 1 서열분석 프라이머 서열 및 모자이크 말단(ME) 서열을 포함한다. 이 예에서, P7 올리고뉴클레오타이드(1115)는 P7 프라이머 서열, 지표 서열(독특한 식별자), 관독 2 서열분석 프라이머 서열 및 ME 서열을 포함한다. 다른 예에서(도시하지 않음), 지표 서열은 단지 P5 올리고뉴클레오타이드(1110)에서 존재한다. 또 다른 예에서(도시하지 않음), 지표 서열은 단지 P7 올리고뉴클레오타이드(1115)에서 존재한다. 또 다른 예에서(도시하지 않음), 지표 서열은 P5 올리고뉴클레오타이드(1110)와 P7 올리고뉴클레오타이드(1115) 둘 다에 없다.
- [0048] 단계(1015)에서, 상보성 모자이크 말단(ME') 올리고뉴클레오타이드는 비드-결합 P5 및 P7 올리고뉴클레오타이드에 혼성화된다. 이 단계는 또한 도 12의 개략적 다이어그램(1200)에서 그림으로 나타낸다. 이제 도 12에 관해, 상보성 ME 서열(ME')(1125)은 P5 올리고뉴클레오타이드(1110) 및 P7 올리고뉴클레오타이드(1115)에 대한 혼성체이다. 상보성 ME 서열(ME')(1125)(예를 들어, 상보성 ME 서열(ME')(1125a) 및 상보성 ME 서열(ME')(1125b)은

각각 P5 올리고뉴클레오타이드(1110) 및 P7 올리고뉴클레오타이드(1115)에서 ME 서열에 혼성화한다. 상보성 ME 서열(ME')(1125)은 전형적으로 길이로 약 15개의 염기이고, 그의 5' 말단에서 인산화된다.

[0049] 단계(1020)에서, 트랜스포사제 효소는 비드-결합 올리고뉴클레오타이드에 첨가되어 비드-결합된 트랜스포좀 복합체의 혼합물을 형성한다. 이 단계는 또한 도 13의 개략적 다이어그램(1300)에서 그림으로 나타낸다. 이제 도 13에 대해, 트랜스포사제 효소가 첨가되어 복수의 트랜스포좀 복합체(1310)를 형성한다. 이 예에서, 트랜스포좀 복합체(1310)는 트랜스포사제 효소, 2 표면-결합 올리고뉴클레오타이드 서열, 및 그들의 혼성화된 상보성 ME 서열(ME')(1125)을 포함하는 이중가닥 구조이다. 예를 들어, 트랜스포좀 복합체(1310a)는 상보성 ME 서열(ME')(1125)에 혼성화된 P5 올리고뉴클레오타이드(1110) 및 상보성 ME 서열(ME')(1125)(즉, P5:P7)에 혼성화된 P7 올리고뉴클레오타이드(1115)를 포함하고; 트랜스포좀 복합체(1310b)는 상보성 ME 서열(ME')(1125)(즉, P5:P5)에 혼성화된 2개의 P5 올리고뉴클레오타이드(1110)를 포함하며; 그리고 트랜스포좀 복합체(1310c)는 상보성 ME 서열(ME')(1125)(즉, P7:P7)에 혼성화된 2개의 P7 올리고뉴클레오타이드(1115)를 포함한다. P5:P5, P7:P7 및 P5:P7 트랜스포좀 복합체의 비는 각각, 예를 들어, 25:25:50일 수 있다.

[0050] 도 14는 도 13의 트랜스포좀 코팅된 비드(1120)를 이용하는 태그화 공정의 예시적인 도식적 다이어그램(1400)을 나타낸다. 이 예에서, 트랜스포좀 복합체(1310)를 지니는 비드(1120)가 태그화 완충제 중의 DNA 용액(1410)에 첨가될 때, 태그화가 일어나며, DNA는 트랜스포좀(1310)을 통해 비드(1120)의 표면에 연결된다. DNA(1410)의 성공적 태그화는 트랜스포좀(1310) 사이에 복수의 브릿지된 분자(1415)를 초래한다. 브릿지된 분자(1415)의 길이는 비드(1120)의 표면 상에서의 트랜스포좀 복합체(1310)의 밀도에 의존할 수 있다. 일 예에서, 비드(1120)의 표면 상에서의 트랜스포좀 복합체(1310)의 밀도는 도 10의 방법(100)의 단계(1010)에서 비드(1120)의 표면에 결합된 P5 및 P7 올리고뉴클레오타이드의 양을 달리함으로써 조절될 수 있다. 다른 예에서, 비드(1120)의 표면 상에서의 트랜스포좀 복합체(1310)의 밀도는 도 10의 방법(1000)의 단계(1015)에서 P5 및 P7 올리고뉴클레오타이드에 혼성화된 상보성 ME 서열(ME')의 양을 달리함으로써 조절될 수 있다. 또 다른 예에서, 비드(1120)의 표면 상에서의 트랜스포좀 복합체(1310)의 밀도는 도 1의 방법(1000)의 단계(1020)에 첨가된 트랜스포사제 효소의 양을 달리함으로써 조절될 수 있다.

[0051] 브릿지된 분자(1415)의 길이는 태그화 반응에서 사용된 트랜스포좀 복합체(1310)가 결합된 비드(1120)의 양에 독립적이다. 유사하게, 태그화 반응에서 거의(1410)를 첨가하는 것은 최종 태그화된 산물의 크기를 변경시키지 않지만, 반응 수율에 영향을 미칠 수 있다.

[0052] 일 예에서, 비드(1120)는 상자성 비드이다. 이 예에서, 태그화 반응의 정제는 자기를 지니는 비드(1120)를 고정시키고, 세척함으로써 용이하게 달성된다. 따라서, 태그화 및 후속적 PCR 증폭은 단일 반응 구획("원-포트") 반응으로 수행될 수 있다.

[0053] 일 양상에서, 본 발명은 고체 지지체 상에서 표적 핵산의 접근 정보를 유도할 수 있는 트랜스포사제-기반 방법의 방법 및 조성물에 관한 것이다. 일부 실시형태에서, 조성물 및 방법은 조립/페이징 정보를 유도할 수 있다. 일 실시형태에서, 고체 지지체는 비드이다. 일 실시형태에서, 표적 핵산은 DNA이다. 일 실시형태에서, 표적 핵산은 게놈 DNA이다. 일부 실시형태에서, 표적 핵산은 RNA이다. 일부 실시형태에서, RNA는 mRNA이다. 일부 실시형태에서, 표적 핵산은 상보성 DNA(cDNA)이다.

[0054] 일부 실시형태에서, 트랜스포존은 고체-지지체, 예컨대 비드에 이량체로서 고정된 다음, 트랜스포존에 대한 트랜스포사제의 결합으로 트랜스포좀을 형성할 수 있다.

[0055] 일부 실시형태에서, 고체상 고정된 트랜스포존에 의한 고체상 상에서의 트랜스포좀의 형성 및 트랜스포사제의 첨가와 특히 관련하여, 두 트랜스포존은 고체 지지체에서 서로에 대해 매우 근접하게(바람직하게는 고정된 거리로) 고정될 수 있다. 이 접근에 대한 몇몇 이점이 있다. 첫째로, 두 트랜스포존은 바람직하게는 최적의 링커 길이 및 두 트랜스포존의 배향을 지니도록 항상 동시에 고정되어 트랜스포좀을 효율적으로 형성할 것이다. 둘째로, 트랜스포좀 형성 효율은 트랜스포존 밀도의 기능의 함수가 아닐 것이다. 두 트랜스포존은 항상 그들 사이에 올바른 배향 및 거리를 지니도록 이용되어 트랜스포좀을 형성할 것이다. 셋째로, 표면 상에서 무작위 고정 트랜스포존에 의해, 트랜스포존 사이에 다양한 거리가 생성되고, 따라서 분획만이 최적의 배향 및 거리를 가져서 트랜스포좀을 효율적으로 형성한다. 결과로서, 모든 트랜스포존이 트랜스포좀으로 전환되지는 않으며, 고체상 고정된 비-복합체 트랜스포존이 제공될 것이다. 이들 트랜스포존은 ME-부분이 이중-가닥 DNA이기 때문에 전위에 대한 표적으로서 허용된다. 이는 전위 효율의 감소를 초래하거나 또는 목적으로 하지 않는 부수적 산물을 생성할 수 있었다. 따라서, 트랜스포좀은 고체 지지체 상에서 제조될 수 있는데, 이는 후속적으로 태그화 및 서열분석을 통해 접근 정보를 유도하는 데 사용될 수 있다. 예시적인 도식은 도 15에서 도시한다. 일부 실시형태

에서, 트랜스포존은 화학적 결합보다는 다른 것에 의해 고체 지지체에 고정될 수 있다. 고체 지지체 상에 트랜스포존을 고정하는 예시적인 방법은 스트랩타비딘-바이오틴, 말토스-말토스 결합 단백질, 항원-항체, DNA-DNA 또는 DNA-RNA 혼성화와 같은 친화도 결합을 포함하지만, 이들로 제한되지 않을 수 있다.

- [0056] 일부 실시형태에서, 트랜스포좀은 사전 조립되고, 이어서, 고체-지지체 상에 고정될 수 있다. 일부 실시형태에서, 트랜스포존은 독특한 지표, 바코드 및 증폭 프라이머 결합 부위를 포함한다. 트랜스포사제는 트랜스포존을 포함하는 용액 중에 첨가되어 트랜스포좀 이량체를 형성하는데, 이는 고체 지지체 상에 고정될 수 있다. 일 실시형태에서, 다중 비드 세트가 생성될 수 있는데, 이때 각각의 세트는 고정된 트랜스포존으로부터 유래된 동일한 지표를 가지며, 따라서 지표화된 비드를 생성할 수 있다. 표적 핵산은 도 29A에 나타내는 바와 같은 지표화된 비드의 각각의 세트에 첨가될 수 있다.
- [0057] 일부 실시형태에서, 표적 핵산은 지표화된 비드의 각각의 세트에 첨가될 수 있고, 태그화 및 후속적 PCR 증폭은 별도로 수행될 수 있다.
- [0058] 일부 실시형태에서, 다수의 액적이 하나 이상의 DNA 분자 및 적절한 트랜스포좀을 지니는 단일 비드를 함유하도록 표적 핵산, 지표화된 비드 및 트랜스포좀이 액체에 합쳐질 수 있다.
- [0059] 일부 실시형태에서, 지표화된 비드는 풀링될 수 있고, 표적 핵산은 풀에 첨가되고, 태그화 및 후속적 PCR 증폭이 단일 반응 구획에서 수행될 수 있다("원-포트").
- [0060] 일 양상에서, 본 발명은 고체 지지체 상에 대해 인접하여-연결된, 전위된, 표적 핵산에 의한 접근 정보를 유도하는 방법 및 조성물에 관한 것이다. 일부 실시형태에서, 접근 보존 전위(CPT)는 DNA 상에서 수행되지만, DNA는 무손상(CPT-DNA)을 유지하고, 따라서 인접하여-연결된 라이브러리를 생성한다. 접근 정보는 표적 핵산에서 인접한 주형 핵산 단편의 회합을 유지하기 위해 트랜스포사제의 사용에 의해 보존될 수 있다. CPT-DNA는 독특한 지표 또는 바코드를 갖는 상보성 올리고뉴클레오타이드의 혼성화에 의해 포획될 수 있고, 고체 지지체, 예를 들어 비드 상에 고정될 수 있다(도 29B). 일부 실시형태에서, 고체 지지체 상에 고정된 올리고뉴클레오타이드는 바코드에 추가로 프라이머 결합 부위, 독특한 분자 지표(unique molecular indices: UMI)를 더 포함할 수 있다.
- [0061] 유리하게는, 단편화된 핵산의 물리적 근접성을 유지하기 위한 트랜스포좀의 이러한 사용은 동일한 본래의 분자, 예를 들어 염색체로부터 단편화된 핵산이 고체 지지체 상에 고정된 올리고뉴클레오타이드로부터 동일한 독특한 바코드 및 지표 정보를 수신할 가능성을 증가시킨다. 이는 독특한 바코드를 지니는 인접하여-연결된 서열분석 라이브러리를 초래할 것이다. 인접하여-연결된 서열분석 라이브러리는 인접한 서열 정보를 유도하기 위해 서열 분석될 수 있다.
- [0062] 도 16 및 도 17은 독특한 바코드 또는 지표를 지니는 인접하여-연결된 라이브러리를 생성하는 본 발명의 상기 양상의 예시적인 실시형태의 개략적 도식을 나타낸다. 예시적인 방법은 서열분석 라이브러리를 생성하기 위해 독특한 지표 및 바코드 및 가닥-대체 PCR을 포함하는 고체 지지체 상에서의 고정 올리고뉴클레오타이드와 CPT-DNA의 결합에 영향을 미친다. 일 실시형태에서, 클론 지표 비드는 무작위 또는 특정 프라이머 및 지표와 같은 고정된 DNA 서열에 의해 생성될 수 있다. 인접하여 연결된 라이브러리는 고정 올리고뉴클레오타이드에 대한 혼성화 다음에 결합에 의해 클론-지표화된 비드 상에 포획될 수 있다. 분자내 혼성화 포획이 분자내 혼성화보다 훨씬 더 빠르기 때문에, 인접하여-전위된 라이브러리는 비드 주위를 "감쌀" 것이다. 도 18 및 도 19는 클론 지표 비드 상에서의 CPT-DNA 포획 및 접근 정보의 보존을 도시한다. 가닥-대체 PCR은 개개 분자에 클론 비드 지표 정보를 전달할 수 있다. 따라서, 각각의 인접하여-연결된 라이브러리는 독특하게 지표화될 것이다.
- [0063] 일부 실시형태에서, 고체 지지체 상에 고정된 올리고뉴클레오타이드는, 하나의 가닥이 고체 지지체에 고정되고, 다른 가닥이 고정 가닥에 부분적으로 상보성이 되어서 Y-어댑터를 생성하도록, 부분적으로 이중 가닥 구조를 포함할 수 있다. 일부 실시형태에서, 고체 표면 상에 고정된 Y-어댑터는 결합 및 갭 충전에 의해 인접하여 연결된 태그화된 DNA에 연결되고, 도 20에 나타낸다.
- [0064] 일부 실시형태에서, Y-어댑터는 비드와 같은 고체 지지체 상에서의 프로브/지표에 의한 CPT-DNA의 혼성화 포획을 통해 형성된다. 도 21은 이러한 Y-어댑터 제조의 예시적인 도식을 나타낸다. 이들 Y-어댑터의 사용은 잠재적으로 모든 단편이 서열분석 라이브러리가 될 수 있다는 것을 보장한다. 이는 서열분석 당 적용범위를 증가시킨다.
- [0065] 일부 실시형태에서, 자유 트랜스포좀은 CPT-DNA로부터 분리될 수 있다. 일부 실시형태에서, 자유 트랜스포좀의 분리는 크기 배제 크로마토그래피에 의한다. 일 실시형태에서, 분리는 마이크로스핀(MicroSpin) S-400 HR 칼럼(펜실베이니아주 피츠버그에 소재한 GE 헬스케어 라이프 사이언시즈(GE Healthcare Life Sciences))에 의해 달



성될 수 있다. 도 22는 자유 트랜스포좀으로부터 분리된 CPT-DNA의 아가로스 겔 전기이동을 나타낸다.

[0066] 혼성화를 통해 고체 지지체 상에 인접하여-연결된, 전위된, 표적 핵산을 포획하는 것은 몇몇 독특한 이점을 가진다. 첫째로, 상기 방법은 혼성화에 기반하며, 전위에 기반하지 않는다. 분자내 혼성화 비율 >> 분자간 혼성화 비율. 따라서, 독특하게 지표화된 비드 주위로 감싸기 위한 단일 표적 DNA 분자 상에서의 인접하여 전위된 라이브러리의 기회는 독특하게 지표화된 비드 주위로 감싸기 위한 2 이상의 상이한 단일 표적 DNA 분자를 갖는 것보다 훨씬 더 높다. 둘째로, DNA 전위 및 전위된 DNA의 바코딩은 두 별개의 단계로 일어난다. 셋째로, 비드 상에서 활성 트랜스포좀 조립체 및 고체-표면 상에서의 트랜스포존의 표면 밀도 최적화와 관련된 도전은 회피될 수 있다. 넷째로, 자기-전위 산물은 칼럼 정제에 의해 제거될 수 있다. 다섯째로, 인접하여 연결된, 전위된, DNA는 갭을 함유하기 때문에, DNA는 더 가요성이며, 따라서 비드 상에 트랜스포좀을 고정시키는 방법에 비해 전위 밀도(삽입 크기)에 대한 부담이 덜 하다. 여섯째로, 상기 방법은 조합적 바코드 도식에 의해 사용될 수 있다. 일곱째로, 비드에 지표 올리고를 공유 결합하는 것이 용이하다. 따라서, 지표 교환에 대한 기회가 더 적다. 여덟째로, 태그화 및 후속적 PCR 증폭은 다중복합화될 수 있고, 각각의 지표 서열에 대한 개개 반응을 수행할 필요를 제거하는 단일 반응 구획("원-포트") 반응에서 수행될 수 있다.

[0067] 일부 실시형태에서, 표적 핵산 전체적으로 복수의 독특한 바코드는 전위 동안 삽입될 수 있다. 일부 실시형태에서, 각각의 바코드는 제1 바코드 서열과 제2 바코드 서열을 포함하며, 그 사이에 단편화 부위가 배치된다. 제1 바코드 서열 및 제2 바코드 서열은 서로 짝지어지도록 동정 또는 지정될 수 있다. 제1 바코드가 제2 바코드와 연결되도록 짝짓기는 정보를 줄 수 있다. 유리하게는, 짝지어진 바코드 서열은 주형 핵산의 라이브러리로부터의 서열분석 데이터를 조립하기 위해 사용될 수 있다. 예를 들어, 제1 바코드 서열을 포함하는 제1 주형 핵산 및 제1 바코드 서열과 짝지어진 제2 바코드 서열을 포함하는 제2 주형 핵산을 동정하는 것은 제1 주형 핵산 및 제2 주형 핵산이 표적 핵산의 서열 표현에서 서로 인접한 서열을 나타낸다는 것을 표시한다. 이러한 방법은 기준 게놈의 필요 없이 드 노보(*de novo*)로 표적 핵산의 서열 표현을 조립하기 위해 사용될 수 있다.

[0068] 일 양상에서, 본 발명은 특정 DNA 단편의 샷건 서열 라이브러리를 생성하기 위한 방법 및 조성물에 관한 것이다.

[0069] 일 실시형태에서, 클론 지표 비드는 고정된 올리고뉴클레오타이드 서열에 의해 생성된다: 무작위 또는 특정 프라이머 및 독특한 지표. 표적 핵산은 클론 지표 비드에 첨가된다. 일부 실시형태에서, 표적 핵산은 DNA이다. 일 실시형태에서, 표적 DNA는 변성된다. 표적 DNA는 고체 표면(예를 들어, 비드) 상에 고정된 독특한 지표를 포함하는 프라이머와 혼성화하고 후속적으로 동일한 지표를 지니는 다른 프라이머와 혼성화한다. 비드 상의 프라이머는 DNA를 증폭시킨다. 증폭의 1회 이상의 추가적인 라운드의 증폭이 수행될 수 있다. 일 실시형태에서, 증폭은 3' 무작위 n-량체 서열을 지니는 비드 고정 프라이머를 이용하여 전체 게놈 증폭에 의해 수행될 수 있다. 바람직한 실시형태에서, 무작위 n-량체는 증폭 동안 프라이머-프라이머 상호작용을 방지하기 위해 위상보성(pseudocomplementary) 염기(2-티오티민, 2-아미노 dA, N4-에틸 사이토신 등)를 함유한다(Hoshika, S; Chen, F; Leal, NA; Benner, SA, *Angew. Chem. Int. Ed.* 49(32) 5554-5557 (2010)). 도 23은 특정 DNA 단편의 샷건 서열 라이브러리를 생성하는 예시적인 도식을 나타낸다. 클론 지표 서열분석 라이브러리는 증폭 산물의 라이브러리가 생성될 수 있는 라이브러리일 수 있다. 일 실시형태에서, 이러한 라이브러리는 전위에 의해 생성될 수 있다. 클론 지표 라이브러리의 서열 정보는 가이드로서 지표 정보를 이용하여 인접한 정보를 조립하기 위해 사용될 수 있다. 도 24는 클론 지표 서열분석 라이브러리로부터의 서열 정보를 조립하는 예시적인 도식을 나타낸다.

[0070] 상기 실시형태의 방법은 몇몇 이점을 가진다. 비드 상에서 분자내 증폭은 비드내 증폭보다 훨씬 더 빠르다. 따라서, 비드 상의 산물은 동일한 지표를 가질 것이다. 특정 DNA 단편의 샷건 라이브러리가 생성될 수 있다. 무작위 프라이머는 무작위 위치에서 주형을 증폭시키며, 따라서, 동일한 지표를 지니는 샷건 라이브러리는 특정 분자로부터 생성될 수 있고, 서열 정보는 지표화된 서열을 이용하여 조립될 수 있다. 상기 실시형태의 방법의 상당한 이점은 반응이 단일 반응(원 포트 반응)으로 다중복합화될 수 있고, 다수의 개개 웰을 이용하는 것이 필요하지 않을 것이다. 다수의 지표 클론 비드가 제조될 수 있으며, 따라서 다수의 상이한 단편이 독특하게 표지될 수 있고, 동일한 게놈 영역에 대한 모 대립유전자로 차이가 생길 수 있다. 매우 다수의 지표에 의해, 부(father)의 DNA 복제물 및 모(mother)의 복제물이 동일한 게놈 영역에 대해 동일한 지표를 수용할 기회는 낮다. 상기 방법은 반응 내가 반응 간보다 훨씬 더 빠르다는 사실을 이용하며, 비드는 더 큰 물리적 구획에서 사실상의 분할을 기본적으로 생성한다.

[0071] 본 발명의 모든 상기 양상의 일부 실시형태에서, 상기 방법은 cfDNA 분석에서 무세포 DNA(cfDNA)에 대해 사용될

수 있다. 일부 실시형태에서, cfDNA는 혈장, 태반액으로부터 얻어진다.

- [0072] 일 실시형태에서, 혈장은 막 기반, 침강 보조 혈장 분리기를 이용하여 비회석 전혈로부터 얻을 수 있다(Liu et al. Anal Chem. 2013 Nov 5;85(21):10463-70). 일 실시형태에서, 혈장 분리기의 혈장의 수집 구역은 트랜스포좀을 포함하는 고체 지지체를 포함할 수 있다. 트랜스포좀을 포함하는 고체 지지체는 단리된 혈장으로부터 cfDNA를 포획할 수 있는데, 그것이 전혈로부터 분리되고, cfDNA를 농축시키고/시키거나 DNA를 태그화할 수 있기 때문이다. 일부 실시형태에서, 태그화는 라이브러리 풀의 서열분석 후에 후속적 탈다중복합화를 허용하도록 독 특한 바코드를 추가로 도입할 것이다.
- [0073] 일부 실시형태에서, 분리기의 수집 구역은 PCR 마스터 믹스(프라이머, 뉴클레오타이드, 완충제, 금속) 및 중합 효소를 포함할 수 있다. 일 실시형태에서, 마스터 믹스는, 혈장이 분리기 밖으로 나옴에 따라 재구성되도록 건조 형태일 수 있다. 일부 실시형태에서, 프라이머는 무작위 프라이머이다. 일부 실시형태에서, 프라이머는 특정 유전자에 대한 특정 프라이머일 수 있다. cfDNA PCR 증폭은 분리된 혈장으로부터 직접적으로 라이브러리의 생성을 초래할 것이다.
- [0074] 일부 실시형태에서, 분리기의 수집 구역은 RT-PCR 마스터 믹스(프라이머, 뉴클레오타이드, 완충제, 금속), 역 전사효소 및 중합효소를 포함할 수 있다. 일부 실시형태에서, 프라이머는 무작위 프라이머 또는 올리고 dT 프라이머이다. 일부 실시형태에서, 프라이머는 특정 유전자에 대한 특정 프라이머일 수 있다. 얻어진 cDNA는 서열분석을 위해 사용될 수 있다. 대안적으로, cDNA는 서열 라이브러리 제조를 위해 고체 지지체 상에 고정된 트랜스포좀으로 처리될 수 있다.
- [0075] 일부 실시형태에서, 혈장 분리는 바코드(1D 또는 2D 바코드)를 포함할 수 있다. 일부 실시형태에서, 분리 장치는 혈액 수집 장치를 포함할 수 있다. 이는 혈장 분리기 및 라이브러리분취 장치에 대한 혈액의 직접적 전달을 초래할 것이다. 일부 실시형태에서, 장치는 하류 서열 분석기를 포함할 수 있다. 일부 실시형태에서, 서열 분석기는 일회용 시퀀서이다. 일부 실시형태에서, 시퀀서는 배취에서 서열분석 전에 샘플을 대기시킬 수 있다. 대안적으로, 시퀀서는 무작위 접근 능력을 가질 수 있으며, 여기서, 샘플은 그들의 서열분석 영역으로 전달된다.
- [0076] 일부 실시형태에서, 혈장에 대한 수집 구역은 무세포 DNA가 농축되도록 실리카 기질을 포함할 수 있다.
- [0077] 동시 페이징 및 메틸화 검출
- [0078] 에피 변형(epi modification)으로도 알려져 있는 5-메틸 사이토신(5-Me-C) 및 5-하이드록시메틸 사이토신(5-하이드록시-C)은 세포 대사, 분화 및 암 발생에서 중요한 역할을 한다. 본 출원의 발명자들은 놀랍게도 그리고 예상치 못하게 본 출원의 방법 및 조성물을 이용하여 페이징 및 동시 메틸화 검출이 가능하다는 것을 발견하였다. 본 발명의 방법은 비드 상의 CPT-seq(지표화된 접근 연결 라이브러리)와 DNA 메틸화 검출을 조합하도록 허용한다. 예를 들어, 비드 상에서 생성된 개개 라이브러리는 중아황산염으로 처리되어, 비메틸화된 C로 전환될 수 있지만, 메틸화된 C를 U로 전환시키지 않아서 5-Me-C의 검출을 허용한다. 이형 접합적 SNP를 이용하는 추가적인 페이징 분석을 통해, epi-의약-페이징 차단은 다중 메가염기 범위를 확립할 수 있다.
- [0079] 일부 실시형태에서, 분석되는 DNA의 크기는 약 수백 개의 염기 내지 약 멀티 메가의 염기일 수 있다. 일부 실시형태에서, 분석되는 DNA의 크기는 약 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1300, 1500, 2000, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7,500, 8000, 8500, 9000, 9500, 10,000, 10,500, 11,000, 11,500, 12,000, 12500, 13000, 14000, 14500, 15000, 15500, 16000, 16500, 17000, 17,500, 18,000, 18,500, 19,000, 19,500, 20,000, 20,500, 21,000, 21,500, 22,000, 22,500, 23,000, 23,500, 24,000, 24,500, 25,000, 25,500, 26,000, 26,500, 27,000, 27,500, 28,000, 28,500, 29,500, 30,000, 30,500, 31,000, 31,500, 32,000, 33,000, 34,000, 35,000, 36,000, 37,000, 38,000, 39,000, 40,000, 42,000, 45,000, 50,000, 55,000, 60,000, 65,000, 70,000, 75,000, 80,000, 85,000, 90,000, 95,000, 100,000, 110,000, 120,000, 130,000, 140,000, 150,000, 160,000, 170,000, 180,000, 200,000, 225,000, 250,000, 300,000, 350,000, 400,000, 450,000, 500,000, 550,000, 600,000, 650,000, 700,000, 750,000, 800,000, 850,000, 900,000, 1,000,000, 1,250,000, 1,500,000, 2,000,000, 2,500,000, 3,000,000, 4,000,000, 5,000,000, 6,000,000, 7,000,000, 8,000,000, 9,000,000, 10,000,000, 15,000,000, 20,000,000, 30,000,000, 40,000,000, 50,000,000, 75,000,000, 100,000,000개 이상의 염기일 수 있다.
- [0080] 본 출원의 개시된 방법 및 조성물을 이용하는 페이징과 관련하여 다른 에피-변형 유사 5-하이드록시-C, DNA 산화 산물, DNA 알킬화 산물, 히스톤-포스트 프린팅 등이 또한 분석될 수 있다.

- [0081] 일부 실시형태에서, DNA는 고체-지지체 상에서 지표-연결 라이브러리로 처음 형질전환된다. 본래의 DNA보다 훨씬 더 작은 개개의 지표화된 라이브러리는 단편화 경향이 더 적는데, 개개 라이브러리가 더 작기 때문이다. 지표화된 라이브러리의 소분획이 상실되긴 하지만, 페이징 정보는 지표화된 DNA 분자의 긴 스패(span)를 가로질러 여전히 유지된다. 예를 들어, 전통적인 중아황산염 전환(BSC)에서 100kb 분자가 절반으로 단편화된다면, 접근은 이제 50kb로 제한된다. 본 명세서에 개시된 방법에서, 100kb 라이브러리는 처음 지표화되고, 개개 라이브러리의 분획이 상실된다고 해도, 접근은 여전히 대략 100kb이다(모든 상실된 라이브러리가 DNA 분자의 하나의 단부로부터 유래될 때 가능성 없는 사건을 제외함). 또한, 본 출원에 개시된 방법은 추가적인 이점을 갖는데, 이는 전통적인 중아황산염 전환 접근에 필요한 것과 대조적으로 추가적인 정제 단계가 필요하지 않기 때문이며, 이에 의해 수율을 개선시킨다. 본 출원의 방법에서, 비드는 중아황산염 전환 후에 단순히 세척된다. 추가적으로, DNA는 고체상에 결합되지만, 완충제 교환은 DNA의 최소 상실(지표화된 라이브러리)에 의해 용이하게 수행될 수 있고, 시간에 맞추어서 역할을 감소시킬 수 있다.
- [0082] 동시 페이징 및 메틸화 검출의 예시적인 도식을 도 43에 나타낸다. 작업 흐름은 비드 상에서의 DNA의 태그화, 9-bp 반복부 영역의 갭-충전-결찰, SDS에 의한 Tn5의 제거 및 비드 상에서의 개개 라이브러리의 중아황산염 전환으로 이루어진다. 중아황산염 전환은 이웃하는 상보성 라이브러리가 제어닐링되지 않는다는 것을 보장하는 변성 조건 하에 수행되며, 따라서 중아황산염 전환 효율을 감소시킨다. BCS는 비메틸화된 C를 U로 전환시키며, 메틸화된 C를 전환시키지 않는다.
- [0083] 도 44는 동시 페이징 및 메틸화 검출의 대안의 예시적인 도식을 나타낸다. 전위 후에 서열분석 라이브러리를 제조한 후에, 단일-가닥 주형을 제조하기 위해 갭-충전-결찰 라이브러리의 분획은 분해된다. 단일-가닥 주형은 라이브러리 상실을 감소시키거나 주형이 중아황산염 전환 효율을 개선시킬 수 있는 단일-가닥이기 때문에, 중아황산염 전환을 위해 더 온화한 조건이 필요하다. 일 실시형태에서, 3' 티오-보호된 트랜스포존(Exo 저항성)과 비보호 트랜스포존의 혼합물은 동일한 비드 상에서 사용된다. 효소, 예를 들어, Exo I는 비-티오-보호된 라이브러리를 분해시키기 위해 사용되어, 그들을 단일 가닥 라이브러리로 전환시킬 수 있다. 티오-보호된 트랜스포존: 비-보호된 트랜스포존의 50:50 혼합물을 이용하여, 라이브러리의 50%는 단일-가닥 라이브러리로 전환될 것이고 (50%는 라이브러리의 트랜스포존이 보호된 하나, 그리고 상보체 가닥이 보호되지 않은 하나를 가짐), 25%는 전환되지 않을 것이며(트랜스포존은 둘 다 티오 보호됨), 25%는 둘 다 전환되어 전체 라이브러리를 제거한다. (트랜스포존은 둘 다 보호되지 않음).
- [0084] 고체상, 예컨대 스트랩타비딘 자기 비드에 결합된 DNA의 중아황산염 전환을 수행하는 한 가지 도전은 고온에서 중아황산나트륨에 의한 비드 결합 DNA의 장기간 처리가 DNA와 비드를 둘 다 손상시킨다는 것이다. DNA 손상을 개선시키기 위해, 운반체 DNA(즉, 랬다 DNA)는 중아황산염 처리 전에 반응 혼합물에 첨가된다. 운반체 DNA의 존재 하에서 조차, 출발 DNA의 대략 80%는 상실되는 것으로 추정되었다. 그 결과, CPTSeq 접근 차단은 전통적인 CPTSeq 프로토콜에서의 접근 차단보다 더 소수의 구성원을 가진다.
- [0085] 따라서, Epi-CPTSeq 프로토콜의 DNA 수율을 개선시키기 위해 본 명세서에서 몇몇 전략이 제안된다. 제1 전략은 스트랩타비딘 비드에 트랜스포좀 복합체를 더 밀집하게 모음으로써 라이브러리 삽입 크기를 감소시키는 것에 의존한다. 라이브러리 크기를 감소시킴으로써, 더 적은 비율의 라이브러리 요소는 중아황산염 처리에 의해 분해된다.
- [0086] Epi-CPTSeq 프로토콜의 DNA 수율을 개선시키기 위한 제2 전략은 파손된 라이브러리 요소의 효소적 회수이다. 회수 전략의 목적은 중아황산염 처리 동안 그들의 3' 부분이 분해 및 상실되는 비드 결합 라이브러리 요소에 다시 라이브러리 증폭이 필요한 3' 공통 서열을 첨가하는 것이다. 3' 공통 서열의 첨가 후에, 이들 요소는 이제 PCR 증폭 및 시퀀싱될 수 있다. 도 67 및 도 68은 이 전략의 예시적인 도식을 나타낸다. 이중 가닥 CPTSeq 라이브러리 요소는 변성되고, 중아황산염 전환되었다(상부 패널). 중아황산염 전환 동안, DNA 가닥 중 하나는 손상되어 (중간 패널), 3' 말단 상에서 PCR 공통 서열의 상실을 야기한다. 주형은 PCR 증폭(하부 패널)에 필수적인 3' 공통 서열(녹색)을 복구시킨다. 일 예에서, 3' 인산화된 감쇠기 올리고의 존재 하에 말단 트랜스퍼라제, 서열분석 어댑터 다음에 올리고 dT 신장을 포함하는 서열이 사용된다(도 68A). 간략하게, TdT는 감쇠기 올리고의 올리고 dT 부분으로 어닐링되는 파손된 라이브러리 요소의 3' 말단에 10 내지 15개의 dA의 신장을 첨가한다. 이 DNA 혼성체의 형성은 TdT 반응을 중단시키며, DNA 중합효소에 의한 파손된 라이브러리 요소의 3' 말단의 결과적인 연장을 위한 주형을 제공한다.
- [0087] 대안의 흐름도에서(도 68B), TdT 꼬리 반응은 단일 가닥 올리고 dT 부분 및 5' 인산화된 이중 가닥 서열분석 어댑터 부분을 함유하는 부분적으로 이중 가닥 감쇠기 올리고의 존재 하에 수행된다. TdT 반응의 종결 시, 마지막

첨가된 dA와 5' 인산화된 감쇠기 올리고 사이의 틈은 DNA 리가제(ligase)에 의해 밀봉된다.

- [0088] 기재된 흐름도는 둘 다 미국 특허 공개 제20150087027호에서 최근에 개발되고 기재된 제어 가능한 TdT 꼬리화 반응에 의존한다. 공통 서열분석 어댑터는 또한 MMLV RT의 최근에 도입된 ssDNA 주형 스위칭(stepping) 활성화에 의해 파손된 라이브러리 요소의 3' 말단에 첨가될 수 있다. 요약하면, MMLV RT 및 주형 전환 올리고(TS-올리고)는 손상된 DNA에 첨가된다(도 68C). 이 반응의 제1 단계에서, 역 전사효소는 단일-가닥 DNA 단편의 3' 말단에 소수의 추가적인 뉴클레오타이드를 첨가하며, 올리고(N) 서열을 지니는 이들 염기쌍은 TS-올리고 중 하나의 3' 말단에 제공된다. 이어서, 역 전사효소 주형 전환 활성화는 BSC 파손 라이브러리 요소의 3' 말단에 어닐링된 공통 프라이머 서열을 첨가하여, 공통 서열분석 프라이머를 이용하여 PCR에서 증폭되는 그의 능력을 회복한다.
- [0089] 제3 전략의 부분으로서, 에피센트레 에피게놈 키트(Epicentre's EpiGenome kit) "중아황산염 전환 후" 라이브러리 구성 방법은 중아황산염 전환 동안 3' 말단에서 그들의 공통 서열을 상실한 라이브러리 요소를 구조하기 위해 사용될 수 있다. 도 69에 나타난 바와 같이, 이 라이브러리 구조 방법은 공통 서열 다음에 무작위 서열의 짧은 신장을 지니는 3' 인산화된 올리고를 이용한다. 이들 짧은 무작위 서열은 중아황산염-처리 단일-가닥 DNA에 혼성화하며, 공통 서열은 후속적으로 DNA 중합효소에 의해 파손된 라이브러리 가닥에 대해 복제된다.
- [0090] 도 74는 비드 상에서 중아황산염 서열분석 방법을 개선시키는 제4의 전략을 나타낸다. 포획 태그를 포함하는 제1 공통 서열은 DNA의 5' 말단에 공유 부착된다. 제1 공통 서열은 미국 특허 출원 공개 제20150087027호에 기재된 바와 같은 단일측 전위(그림), 어댑터 결합 또는 말단 트랜스퍼라제(TdT) 어댑터 결합을 포함하는 다양한 방법을 이용하여 DNA에 부착될 수 있다.
- [0091] 다음에, DNA는 변성되고(예를 들어, 고열에서 인큐베이션) 고체 지지체에 결합된다. 바이오틴이 CS1 상에서 포획 태그로서 사용된다면, 예를 들어, DNA는 스트렙타비딘 자기 비드(그림)를 이용하여 결합될 수 있다. 일단 고체 지지체에 결합되면, 완충제 교환이 용이하게 이루어질 수 있다.
- [0092] 다음 단계에서, ssDNA의 중아황산염 전환이 수행된다. 단일 가닥 형태에서, DNA는 중아황산염 전환을 위해 용이하게 접근 가능하여야 하며; 95%까지의 전환 효율은 프로메가사(Promega)의 메틸 에지 BSC 키트(Methyl Edge BSC kit)의 변형 형태를 이용하여 관찰되었다(도 75).
- [0093] 중아황산염 전환 후에, 제2 공통 서열은 고체 지지체에 부착된 ssDNA의 3' 말단에 공유 부착된다. 몇몇 방법은 ssDNA에 올리고를 공유 부착하기 위해 상기 기재되었다. TdT 감쇠기/어댑터 결합 방법을 이용하여, 95% 초과 결합 효율이 달성되었다. 그 결과, 제안된 메틸시크(MethylSeq) 흐름도를 이용하는 최종 라이브러리 수율은 기존의 방법보다 더 커야 한다.
- [0094] 최종 단계에서, PCR을 수행하여 라이브러리를 증폭시키는 것을 수행하고, 고체 지지체로부터 그것을 제거한다. PCR 프라이머는 메틸시크 라이브러리의 말단에 추가적인 공통 서열, 예컨대 서열분석을 첨가하도록 설계될 수 있다.
- [0095] 단일 분석에서 상이한 크기 라이브러리의 제조
- [0096] 게놈 조립의 정확도는 상이한 길이 척도 기법의 사용 여부에 따른다. 예를 들어, -Hi-C(Mb-척도)에 대한 샷건(bp 100) - 교접쌍(matepair)(~3Kb)은 조립체 및 콘택 길이를 후속적으로 개선시키는 모든 방법이다. 도전은 이를 달성하기 위해 다중 분석이 필요하여, 다중층 접근을 번거롭고 비용이 들게 만든다는 점이다. 본 명세서에 개시된 조성물 및 방법은 단일 분석에서 다중 길이 척도를 처리할 수 있다.
- [0097] 일부 실시형태에서, 라이브러리 제조는 차별적으로 크기가 조절된 고체 지지체, 예를 들어, 비드를 이용하여 단일 분석에서 달성될 수 있다. 각각의 비드 크기는 특정 라이브러리 크기 또는 크기의 범위를 생성할 것이고, 비드의 물리적 크기는 라이브러리 크기를 결정한다. 다양한 크기의 비드는 모두 라이브러리에 전달되는 독특한 클론 지표를 가진다. 그렇게 해서, 각각 상이한 라이브러리 척도-길이 척도가 독특하게 지표화된 상이한 크기 라이브러리가 생성된다. 다양한 길이-척도 라이브러리는 동일한 물리적 구획에서 동시에 제조되어, 비용을 감소시키고 전반적인 흐름도를 개선시킨다. 일부 실시형태에서, 각각의 특정 고체 지지체 크기, 예를 들어, 비드 크기는 독특한 지표를 수용한다. 일부 다른 실시형태에서, 동일한 고체 지지체 크기, 예를 들어 비드 크기의 다수 상이한 지표가 또한 제조되고, 따라서 다중 DNA 분자는 크기 범위를 위해 구획화된 지표일 수 있다. 도 45는 단일 분석에서 다양한 크기의 클론으로 지표화된 비드를 이용하여 다양한 크기의 라이브러리를 생성하기 위한 예시적인 도식을 나타낸다.
- [0098] 일부 실시형태에서, 생성된 라이브러리의 크기는 약 50, 75, 100, 150, 200, 250, 300, 350, 400, 500, 600,



700, 800, 900, 1000, 1200, 1300, 1500, 2000, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7,500, 8000, 8500, 9000, 9500, 10,000, 10,500, 11,000, 11,500, 12,000, 12500, 13000, 14000, 14500, 15000, 15500, 16000, 16500, 17000, 17,500, 18,000, 18,500, 19,000, 19,500, 20,000, 20,500, 21,000, 21,500, 22,000, 22,500, 23,000, 23,500, 24,000, 24,500, 25,000, 25,500, 26,000, 26,500, 27,000, 27,500, 28,000, 28,500, 29,500, 30,000, 30,500, 31,000, 31,500, 32,000, 33,000, 34,000, 35,000, 36,000, 37,000, 38,000, 39,000, 40,000, 42,000, 45,000, 50,000, 55,000, 60,000, 65,000, 70,000, 75,000, 80,000, 85,000, 90,000, 95,000, 100,000, 110,000, 120,000, 130,000, 140,000, 150,000, 160,000, 170,000, 180,000, 200,000, 225,000, 250,000, 300,000, 350,000, 400,000, 450,000, 500,000, 550,000, 600,000, 650,000, 700,000, 750,000, 800,000, 850,000, 900,000, 1,000,000, 1,250,000, 1,500,000, 2,000,000, 2,500,000, 3,000,000, 4,000,000, 5,000,000, 6,000,000, 7,000,000, 8,000,000, 9,000,000, 10,000,000, 15,000,000, 20,000,000, 30,000,000, 40,000,000, 50,000,000, 75,000,000, 100,000,000개 이상의 염기이다.

[0099] 일부 실시형태에서, 상기 논의한 다중 길이 척도 라이브러리는 하나의 큰 길이 척도를 갖는 대신에 위유전자의 조립체, 파라로그 등에서 사용될 수 있다. 일부 실시형태에서, 다중 길이 척도 라이브러리는 단일 분석에서 동시에 제조된다. 이점은 적어도 하나의 길이-척도는 위-유전자 또는 유전자만을 지니는(그러나 둘 다는 아님) 독특한 영역을 연결할 것이라는 점이다. 그렇게 해서, 이 길이 척도에 의해 검출된 변이체는 유전자 또는 위-유전자 중 하나에 변이체를 독특하게 부여할 수 있다. 복제수 변이체, 파라로그 등에 대해 동일하게 적용된다. 조립체의 신장은 상이한 길이 척도의 사용이다. 본 명세서에 개시된 방법을 이용하여, 상이한 길이 척도에 대한 개개, 상이한 라이브러리 대신 단일 분석에서 상이한 길이 척도 지표 연결 라이브러리가 생성될 수 있다. 도 46은 상이한 길이 척도 라이브러리를 이용하여 유전자 변이체를 결정하는 예시적인 도식을 나타낸다.

[0100] 게놈 변이체의 분석

[0101] 본 명세서에 개시된 조성물 및 방법은 게놈 변이체의 분석에 관한 것이다. 예시적인 게놈 변이체는 결실, 염색체간 전좌, 중복, 파라로그, 염색체간유전자 융합을 포함하지만, 이들로 제한되지 않는다. 일부 실시형태에서, 본 명세서에 개시된 조성물 및 방법은 게놈 변이체의 페이징 정보를 결정하는 것에 관한 것이다. 이하의 표는 예시적인 염색체간 유전자 융합을 나타낸다.

표 1

염색체간 유전자 융합

염색체간								
BP1 염색체	BP1 시작	BP1 말단	길이	BP2 염색체	BP2 시작	BP2 말단	길이	보충
염색체2	3930448	3935237	4789	염색체12	124494444	124,500,974	6,530	100
염색체2	11948918	11954246	5328	염색체21	18143482	18,149,316	5,834	104
염색체3	73158440	73162742	4302	염색체17	41379324	41384937	5,613	80
염색체3	73158440	73162742	4302	염색체17	41397704	41,403,894	6,190	132
염색체3	75985674	75999286	13712	염색체20	26198677	26,213,907	15,230	271
염색체3	97545352	97548214	2862	염색체20	51820873	51826137	5,264	54
염색체3	111271573	111275045	3472	염색체8	128529232	128,538,779	9,547	138
염색체5	45603558	45620710	17152	염색체22	39709007	39715736	6,729	64
염색체7	81789161	81794473	5312	염색체10	60900343	60904922	4,579	66
염색체7	100936524	100940377	3853	염색체11	114417947	114425186	7,239	45
염색체7	111050826	111056166	5340	염색체12	108199590	108203183	3,593	55
염색체8	28431852	28436443	4591	염색체17	79758758	79762966	4,208	51
염색체9	12624450	12627475	3025	염색체22	29065133	29067995	2,862	11
염색체9	96381359	96386424	5065	염색체13	19645775	19650874	5,099	29
염색체9	101473886	101479887	6001	염색체22	32575696	32579748	4,052	67
염색체9	109772276	109777707	5431	염색체20	29873930	29878940	5,010	70
염색체9	121298728	121305690	6962	염색체11	11980608	12002075	21,467	178
염색체12	56987510	56991801	4291	염색체15	39995972	39999958	3,986	25
염색체13	55877158	55882393	5235	염색체19	12896441	12904258	7,817	72
염색체13	63618669	63624153	5484	염색체17	21666588	21670266	3,678	115

[0102]

[0103] 표 2는 염색체 1에서의 예시적인 결실을 나타낸다,

표 2

염색체 1에서의 예시적인 결실

결실								
염색체	BP1 시작	BP1 말단		BP2 시작	BP2 말단		크기 (bp)	CPT 신호
염색체1	1219982	1223434	3452	1225577	1229105	3628	2143	있음
염색체1	9591985	9595341	3356	9597227	9600089	2862	1886	있음
염색체1	14432636	14436290	3654	14437570	14441375	3805	1280	있음
염색체1	25154228	25158676	4448	25161514	25165093	3579	2838	있음
염색체1	26465898	26469838	2940	26463789	26468542	4753	3951	있음
염색체1	34985513	34991181	5668	34992539	34996862	4323	1358	있음
염색체1	56827433	56831099	3666	56834909	56838758	3849	3810	없음
염색체1	63702441	63705334	2893	63708226	63712091	3865	2892	??
염색체1	71234176	71237339	3163	71239506	71243649	4143	2167	있음
염색체1	73450444	73454045	3601	73454999	73458655	3656	954	??
염색체1	81400468	81404472	4004	81408850	81413863	5013	4378	있음
염색체1	83465559	83469370	3811	83477194	83482174	4980	7824	있음
염색체1	84513313	84517907	4594	84524680	84528485	3805	6773	있음
염색체1	84708940	84711986	3046	84715625	84719559	3934	3639	있음
염색체1	97657769	97660911	3142	97669287	97672164	2877	8376	없음
염색체1	1.06E+08	1.06E+08	4804	1.06E+08	1.06E+08	3723	6370	있음
염색체1	1.07E+08	1.07E+08	5095	1.07E+08	1.07E+08	6141	887	??
염색체1	1.09E+08	1.09E+08	2999	1.09E+08	1.09E+08	4018	3899	있음
염색체1	1.1E+08	1.1E+08	5069	1.1E+08	1.1E+08	4470	2801	??
염색체1	1.13E+08	1.13E+08	5237	1.13E+08	1.13E+08	4626	3241	있음
염색체1	1.16E+08	1.16E+08	4594	1.16E+08	1.16E+08	4531	3084	??
염색체1	1.45E+08	1.45E+08	5596	1.45E+08	1.45E+08	4721	3724	있음
염색체1	1.59E+08	1.59E+08	4541	1.59E+08	1.59E+08	3981	1856	있음(이중 결실)
염색체1	1.79E+08	1.79E+08	4565	1.79E+08	1.79E+08	5191	2238	있음
염색체1	1.79E+08	1.79E+08	4480	1.79E+08	1.79E+08	3917	2479	있음
염색체1	1.85E+08	1.85E+08	4475	1.85E+08	1.85E+08	3316	6243	있음
염색체1	1.97E+08	1.98E+08	4323	1.98E+08	1.98E+08	3694	2522	없음
염색체1	2.06E+08	2.06E+08	4995	2.06E+08	2.06E+08	5216	3458	있음
염색체1	2.08E+08	2.08E+08	3589	2.08E+08	2.08E+08	4964	2954	있음
염색체1	2.1E+08	2.1E+08	2973	2.1E+08	2.1E+08	2943	7603	있음(이중 결실)
염색체1	2.3E+08	2.3E+08	3828	2.3E+08	2.3E+08	3079	8440	있음
염색체1	2.32E+08	2.32E+08	3531	2.32E+08	2.32E+08	3758	1699	??
염색체1	2.38E+08	2.38E+08	3848	2.38E+08	2.38E+08	3920	2112	있음
염색체1	2.48E+08	2.48E+08	3407	2.48E+08	2.48E+08	3873	6003	있음

[0104]

[0105]

일부 실시형태에서, 표적 핵산은 그것을 트랜스포좀에 노출시키기 전에 단편화될 수 있다. 예시적인 단편화 방법은 음파처리, 기계적 전단 및 제한 분해를 포함하지만, 이들로 제한되지 않는다. 태그화 전에 표적 핵산의 단편화(단편화 및 태그)는 위유전자(예를 들어, CYP2D6)의 조립/페이징에 대해 유리하다. 지표화 연결 관독의 긴 섬(>30kb)은 도 64에 나타내는 바와 같은 위유전자 A 및 A'에 걸칠 것이다. 높은 서열 상동성에 기인하여, 변이체가 유전자 A 및 유전자 A' 속하는지를 결정하는 것에 도전할 것이다. 더 짧은 변이체는 위유전자의 하나의 변이체를 독특한 주변 서열과 연결할 것이다. 이러한 더 짧은 섬은 태그화 전에 표적 핵산을 단편화함으로써 달성될 수 있다.

[0106]

연결된 트랜스포좀

[0107]

일부 실시형태에서, 트랜스포사제는 트랜스포좀 복합체에서 다량체이며, 예를 들어, 그들은 트랜스포좀 복합체에서 이량체, 사량체 등을 형성한다. 본 출원의 발명자는 놀랍게도 그리고 예상치 못하게 다량체 트랜스포좀 복합체에서 단량체 트랜스포사제의 연결 또는 다량체 트랜스포좀 복합체에서 트랜스포좀 단량체의 트랜스포존 말단의 연결이 몇몇 이점을 가진다는 것을 발견하였다. 첫째로, 트랜스포사제 또는 트랜스포존의 연결은 더 안정한 복합체를 야기하며, 더 큰 분획은 활성 상태이다. 둘째로, 더 저농도의 트랜스포좀은 잠재적으로 전위 반응에 의해 단편화에서 사용될 수 있다. 셋째로, 연결은 트랜스포좀 복합체의 모자이크 말단(ME)의 교환을 낮춰서, 바코드 또는 어댑터 분자의 더 적은 혼합을 야기한다. 복합체가 무너지고 재형성된다면 또는 트랜스포좀이 스트랩타비딘/바이오틴에 의해 고체 지지체 상에 고정되는 경우에 ME 말단의 이러한 스와핑은 가능하며, 스트랩타비딘/바이오틴 상호작용은 파손되고, 또는 가능한 오염이 있을 때 재형성될 수 있다. 본 출원의 발명자들은 다량

한 반응 조건 하에서 ME 말단의 상당한 스와핑 또는 교환이 있다는 것을 주목하였다. 일부 실시형태에서, 교환은 15%만큼 높을 수 있다. 교환은 고염 완충제 중에서 두드러지며, 교환은 글루탐산염 완충제 중에서 감소된다. 도 57 및 도 58은 ME 스와핑의 일부 가능한 메커니즘을 나타낸다.

[0108] 일부 실시형태에서, 트랜스포좀 복합체에서 트랜스포사제 서브유닛은 공유 및 비공유 수단에 의해 서로 연결될 수 있다. 일부 실시형태에서, 트랜스포사제 단량체는 트랜스포좀 복합체의 생성 전에(트랜스포존의 첨가 전에) 연결될 수 있다. 일부 실시형태에서, 트랜스포사제 단량체는 트랜스포좀 형성 후에 연결될 수 있다.

[0109] 일부 실시형태에서, 천연 아미노산 잔기는 이황화 결합 형성을 촉진시키기 위해 다량체 계면에서 시스테인(Cys) 아미노산으로 치환될 수 있다. 예를 들어, Tn5 트랜스포사제에서, Asp468, Tyr407, Asp461, Lys459, Ser458, Gly462, Ala466, Met470은 Cys으로 치환되어 단량체 서브유닛 사이에서 이황화 결합을 촉진시킬 수 있으며, 이를 도 59 및 도 60에 나타낸다. Mos-1 트랜스포사제에 대해, 시스테인으로 치환될 수 있는 예시적인 아미노산은 Leu21, Leu32, Ala35, His20, Phe17, Phe36, Ile16, Thr13, Arg12, Gln10, Glu9을 포함하지만, 이들로 제한되지 않으며, 이를 도 61에 나타낸다. 일부 실시형태에서, 시스테인으로 치환되는 아미노산 잔기를 지니는 변형된 트랜스포사제는 말레이미드 또는 피리달다이트올 반응성 기를 이용하여 화학적 가교제를 이용하여 서로 화학적으로 가교될 수 있다. 예시적인 화학적 가교제는 피어스 프로테인 바이올로지사(Pierce Protein Biology)/써모 피셔 사이언티픽사(ThermoFisher Scientific)(미국 뉴욕주에 소재한 그랜드 아일랜드)로부터 상업적으로 입수 가능하다.

[0110] 일부 실시형태에서, 트랜스포좀 다량체 복합체는 고체 지지체에 공유 결합될 수 있다. 예시적인 고체 지지체는 나노입자, 비드, 유세포 표면, 칼럼 매트릭스를 포함하지만, 이들로 제한되지 않는다. 일부 실시형태에서, 고체 표면은 아민기로 코팅될 수 있다. 시스테인으로 치환되는 아미노산 잔기를 지니는 변형된 트랜스포사제는 아민-대-설피하이드릴 가교제(즉, 숙신이미드-4-(N-말레이미도메틸)사이클로헥산-1-카복실레이트(SMCC))를 이용하여 이러한 아민기에 화학적으로 가교될 수 있다. 예시적인 도식을 도 62에 나타낸다. 일부 실시형태에서, 스트랩타 비딘 코팅된 고체 표면에 dTnp를 결합하기 위해 말레이미드-PEG-바이오틴 가교제가 사용될 수 있다.

[0111] 일부 실시형태에서, 트랜스포사제 유전자는 단일 폴리펩타이드에서 다량체 단백질을 발현시키도록 변형될 수 있다. 예를 들어, Tn5 또는 Mos-1 유전자는 단일 폴리펩타이드에서 2개의 Tn5 또는 Mos-1 단백질을 발현시키도록 변형될 수 있다. 유사하게, Mu 트랜스포사제 유전자는 단일 폴리펩타이드에서 4개의 Mu 트랜스포사제 단위를 암호화하도록 변형될 수 있다.

[0112] 일부 실시형태에서, 트랜스포좀 다량체 단위의 트랜스포존 말단은 연결된 트랜스포좀 다량체 복합체를 형성하도록 연결될 수 있다. 트랜스포존 말단을 연결하는 것은 프라이머 부위, 서열분석 프라이머, 증폭 프라이머의 삽입을 허용하거나 또는 임의의 역할 DNA는 표적 DNA를 단편화하는 일 없이 gDNA로 활동할 수 있다. 이러한 기능성의 삽입은 단상형 분석 또는 집합 태깅 분석에서 유리한데, 이때 정보는 무손상 분자로부터 추출될 필요가 있거나 또는 하위-샘플링이 중요하다. 일부 실시형태에서, Mu 트랜스포좀의 트랜스포존 말단은 "루프" Mu 트랜스포사제/트랜스포존 입체배치에 연결될 수 있다. Mu는 사량체이기 때문에, 다양한 입체배치가 가능하지만, R2UJ 및/또는 R1UJ를 R2J 및/또는 R1J와 연결하는 것에 의해 제한되지 않는다. 이들 입체배치에서, R2UJ 및 R1UJ는 각각 R2J 및 R1J와 연결될 수 없다. 도 63은 트랜스포존 말단이 연결된 Mu 트랜스포좀 복합체를 나타낸다. 일부 실시형태에서, Tn5의 트랜스포존 말단 또는 Mos-1 트랜스포좀의 트랜스포존 말단은 연결될 수 있다.

[0113] 본 명세서에 사용된 바와 같은 용어 "트랜스포존"은 시험관내 전위 반응에서 기능성인 트랜스포사제 또는 인테그라제 효소와 복합체를 형성하는 데 필수적인 뉴클레오타이드 서열("트랜스포존 말단 서열")만을 나타내는 이중 가닥 DNA를 의미한다. 트랜스포존은 트랜스포존을 인식하고 이에 결합하는 트랜스포사제 또는 인테그라제와 함께 "복합체" 또는 "시냅스 복합체" 또는 "트랜스포좀 복합체" 또는 "트랜스포좀 조성물"을 형성하는데, 이 복합체는 그와 함께 시험관내 전위 반응에서 인큐베이션되는 표적 DNA 내로 트랜스포존을 삽입 또는 전위시킬 수 있다. 트랜스포존은 "전달된 트랜스포존 서열" 또는 "전달된 가닥" 및 "비전달 트랜스포존 서열" 또는 "비 전달된 가닥"으로 이루어진 2개의 상보성 서열을 나타낸다. 예를 들어, 시험관내 전위 반응에서 활성인 초활성 Tn5 트랜스포사제(예를 들어, EZ-Tn5(상표명) 트랜스포사제, 미국 위스콘신주 메디슨에 소재한 에피센트레 바이오 테크놀로지스(EPICENTRE Biotechnologies))와 복합체를 형성하는 하나의 트랜스포존은 다음과 같은 "전달 트랜스포존 서열"을 나타내는 전달된 가닥을 포함한다:

[0114] 5' AGATGTGTATAAGAGACAG 3'

[0115] 및 다음과 같은 "비전달 트랜스포존 서열"을 나타내는 비-전달된 가닥:



- [0116] 5' CTGTCT CTTATACATCT 3'.
- [0117] 전달된 가닥의 3'-말단은 시험관 내 전위 반응에서 DNA를 표적화하도록 결합 또는 전달된다. 전달된 트랜스포존 말단 서열에 대해 상보성인 트랜스포존 서열을 나타내는 비-전달된 가닥은 시험관 내 전위 반응에서 표적 DNA에 결합 또는 전달되지 않는다. 일부 실시형태에서, 트랜스포존 서열은 다음 중 하나 이상을 포함할 수 있다: 바코드, 어댑터 서열, 태그 서열, 프라이머 결합 서열, 포획 서열, 독특한 분자 식별자(UMI) 서열.
- [0118] 본 명세서에서 사용되는 용어 "어댑터"는 바코드, 프라이머 결합 서열, 포획 서열, 포획 서열에 대해 상보성인 서열, 독특한 분자 식별자(UMI) 서열, 친화도 모이어티, 제한 부위를 포함할 수 있는 핵산 서열을 의미한다.
- [0119] 본 명세서에서 사용되는 용어 "접근 정보"는 공유 정보에 기반한 2 이상의 DNA 단편 사이의 공간적 관계를 지칭한다. 정보의 공유된 양상은 인접, 구획 및 거리 공간적 관계에 대한 것일 수 있다. 이들 관계에 관한 정보는 결국 DNA 단편으로부터 유래된 서열 판독의 계층적 조립 또는 맵핑을 가능하게 한다. 이 접근 정보는 이러한 조립 또는 맵핑의 효율 및 정확성을 개선시키는데, 그들이 개개 서열 판독으로부터 유래된 2 이상의 DNA 단편 사이의 공간적 관계에 관한 것이기 때문에 통상적인 샷건 서열분석과 관련하여 사용된 전통적인 조립 또는 맵핑 방법이 개개 서열 판독의 상대적 계층 기점 또는 좌표를 고려하지 않기 때문이다. 따라서, 본 명세서에 기재된 실시형태에 따르면, 접근 정보를 포획하는 방법은 인접한 공간적 관계를 결정하는 짧은 범위 접근 방법, 구획적 공간 관계를 결정하기 위한 중간-범위 접근 방법 또는 거리 공간 관계를 결정하기 위한 긴 범위 접근 방법에 의해 달성될 수 있다. 이들 방법은 DNA 서열 조립 또는 맵핑의 정확성 및 품질을 가능하게 하며, 임의의 서열분석 방법, 예컨대 상기 기재한 것과 함께 사용될 수 있다.
- [0120] 접근 정보는 그들이 개개 서열 판독으로부터 유래된 2 이상의 DNA 단편 사이의 공간적 관계에 관한 것이기 때문에 개개 서열 판독의 상대적 계층 기점 또는 좌표를 포함한다. 일부 실시형태에서, 접근 정보는 비중첩 서열 판독으로부터의 서열 정보를 포함한다.
- [0121] 일부 실시형태에서, 표적 핵산 서열의 접근 정보는 단상형 정보를 나타낸다. 일부 실시형태에서, 표적 핵산 서열의 접근 정보는 계층 변이체를 나타낸다.
- [0122] 핵산의 단편화와 관련하여 본 명세서에서 사용되는 용어 "표적 핵산의 접근을 유지하는"은 동일한 표적 핵산으로부터 단편의 핵산 서열 순서를 유지하는 것을 의미한다.
- [0123] 본 명세서에서 사용되는 용어 "적어도 일부" 및/또는 이의 문법적 동의어는 전체량의 임의의 분획을 지칭할 수 있다. 예를 들어, "적어도 일부"는 전체량의 적어도 약 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99%, 99.9% 또는 100%를 지칭할 수 있다.
- [0124] 본 명세서에서 사용되는 용어 "약"은  $\pm$  10%를 의미한다.
- [0125] 본 명세서에서 사용되는 용어 "서열분석 판독" 및/또는 이의 문법적 동의어는 중합체 내 단량체 규모를 표시하는 신호를 얻기 위해 수행되는 물리적 또는 화학적 단계의 반복적 공정을 지칭할 수 있다. 신호는 단일 단량체 분해능 또는 더 낮은 분해능에서 단량체 규모를 나타낼 수 있다. 특정 실시형태에서, 단계는 핵산 표적 상에서 개시되고, 핵산 표적에서 염기 규모를 나타내는 신호를 얻기 위해 수행될 수 있다. 상기 과정은 그의 전형적인 완료로 수행될 수 있는데, 완료는 과정으로부터의 신호가 더 이상 표적 염기를 합리적인 수준의 확실성으로 구별할 수 없는 시점까지 보통 정해진다. 원한다면, 완료는 더 빨리 일어날 수 있고, 예를 들어, 일단 목적으로 하는 양의 서열 정보가 얻어졌다. 서열분석 판독은 단일 표적 핵산 분자 상에서 또는 동일한 서열을 갖는 표적 핵산 분자의 집단 상에서 동시에, 또는 상이한 서열을 갖는 표적 핵산 집단 상에서 동시에 수행될 수 있다. 일부 실시형태에서, 서열분석 판독은 단일 신호 획득으로부터 개시된 하나 이상의 표적 핵산 분자로부터 신호가 더 이상 얻어지지 않을 때 종결된다. 예를 들어, 서열분석 판독은 고체상 기질 상에 존재하는 하나 이상의 표적 핵산 분자에 대해 개시될 수 있고, 기질로부터 하나 이상의 표적 핵산 분자의 제거 시 종결된다. 서열분석 실행이 개시될 때 기질 상에 존재한 표적 핵산의 검출을 달리 중단시킴으로써 서열분석이 종결될 수 있다. 예시적인 서열분석 방법은 본 명세서에 전문이 참고로 포함되는 미국 특허 제9,029,103호에 기재되어 있다.
- [0126] 본 명세서에 사용되는 용어 "서열분석 표현" 및/또는 이의 문법적 동의어는 중합체 내 단량체 단위의 규모 및 유형을 나타내는 정보를 지칭할 수 있다. 예를 들어, 정보는 핵산 내 뉴클레오타이드의 규모 및 유형을 나타낼 수 있다. 정보는, 예를 들어, 묘사, 이미지, 전자 매체, 일련의 기호, 일련의 숫자, 일련의 글자, 일련의 색상 등을 포함하는 임의의 다양한 형식일 수 있다. 정보는 단일 단량체 분해능으로 또는 더 낮은 분해능일 수 있다.

예시적인 중합체는 뉴클레오타이드 단위를 갖는 핵산, 예컨대 DNA 또는 RNA이다. 일련의 "A", "T", "G" 및 "C" 글자는 단일 뉴클레오타이드 해상도에서 DNA 분자의 실제 서열과 상관관계가 있을 수 있는 DNA에 대한 잘 알려진 진 서열 표현이다. 다른 예시적인 중합체는 아미노산 단위를 갖는 단백질 및 당류 단위를 갖는 다당류이다.

[0127] **고체 지지체**

[0128] 본 출원 전체적으로, 고체 지지체와 고체 표면은 상호 호환적으로 사용된다. 일부 실시형태에서, 고체 지지체 또는 그의 표면은 비-평면, 예컨대 관 또는 용기의 내면 또는 외면이다. 일부 실시형태에서, 고체 지지체는 마이크로스피어 또는 비드를 포함한다. 본 명세서에서 "마이크로스피어" 또는 "비드" 또는 "입자" 또는 문법적 동의어는 작은 별도의 입자를 의미한다. 적합한 비드 조성물은 플라스틱, 세라믹, 유리, 폴리스타이렌, 메틸스타이렌, 아크릴 중합체, 상자성 물질, 토리아 졸, 흑연 탄소, 이산화티타늄, 라텍스 또는 가교된 텍스트란, 예컨대 세파로스, 셀룰로스, 나일론, 가교된 미셀 및 테플론을 포함하지만, 이들로 제한되지 않을 뿐만 아니라 고체 지지체에 대해 본 명세서에 약술된 임의의 다른 물질이 모두 사용될 수 있다. 피셔 인코포레이티드(Fishers Ind.) 방스 래버러토리즈(Bangs Laboratories)로부터의 "마이크로스피어 검출 가이드"는 도움이 되는 가이드이다. 특정 실시형태에서, 마이크로스피어는 자기 마이크로스피어 또는 비드이다. 일부 실시형태에서, 비드는 코팅된 색일 수 있다. 예를 들어, 텍사스주 오스틴에 소재한 루미넥스사(Luminex)로부터의 마이크로플렉스(MicroPlex)(등록상표) 마이크로스피어가 사용될 수 있다.

[0129] 비드는 구체일 필요는 없으며; 불규칙적 입자가 사용될 수 있다. 대안적으로 또는 추가적으로, 비드는 다공성일 수 있다. 비드 크기는, 일부 실시형태에서 더 작은 또는 더 큰 비드가 사용될 수 있지만, 나노미터로부터, 즉, 직경이 약 10nm 내지 밀리미터, 즉, 1mm의 범위이며, 약 0.2마이크론 내지 약 200마이크론이 바람직하고, 약 0.5 내지 약 5마이크론의 비드가 특히 바람직하다. 일부 실시형태에서, 비드는 직경이 약 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 2.5, 2.8, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150 또는 200 $\mu$ m일 수 있다.

[0130] **트랜스포좀**

[0131] "트랜스포좀"은 통합 인식 부위, 예컨대 트랜스포사제 인식 부위를 포함하는 통합 효소, 예컨대 인테그라제 또는 트랜스포사제 및 핵산을 포함한다. 본 명세서에 제공된 실시형태에서, 트랜스포사제는 전위 반응을 촉매할 수 있는 트랜스포사제 인식 부위와 기능성 복합체를 형성할 수 있다. 트랜스포사제는 트랜스포사제 인식 부위에 결합되고, 때때로 "태그화"로 지칭되는 과정에서 트랜스포사제 인식 부위를 표적 핵산 내로 삽입할 수 있다. 일부 이러한 삽입 사건에서, 트랜스포사제 인식 부위 중 하나의 가닥은 표적 핵산에 전달될 수 있다. 일 예에서, 트랜스포좀은 2개의 서브유닛 및 2개의 비인접 트랜스포존 서열을 포함하는 이량체 트랜스포사제를 포함한다. 다른 예에서, 트랜스포좀은 2개의 서브유닛 및 인접한 트랜스포존 서열을 포함하는 이량체 트랜스포사제를 포함하는 트랜스포사제를 포함한다.

[0132] 일부 실시형태는 초활성 Tn5 트랜스포사제 및 Tn5-형 트랜스포사제 인식 부위(Goryshin and Reznikoff, J. Biol. Chem., 273:7367 (1998)), 또는 R1 및 R2 말단 서열을 포함하는 MuA 트랜스포사제 및 Mu 트랜스포사제 인식 부위(Mizuuchi, K., Cell, 35: 785, 1983; Savilahti, H, et al., EMBO J., 14: 4893, 1995)의 사용을 포함할 수 있다. 초활성 Tn5 트랜스포사제(예를 들어, EZ-Tn5(상표명) 트랜스포사제, 위스콘신주 메디슨에 소재한 에피센트레 바이오테크놀로지)와 복합체를 형성하는 예시적인 트랜스포사제 인식 부위는 다음의 19b 전달된 가닥(때때로 "M" 또는 "ME") 및 비-전달 가닥: 5' AGATGTGTATAAGAGACAG 3', 5' CTGTCT CTTATACACATCT 3'을 각각 포함한다. 당업자에 의해 최적으로 되는 ME 서열이 또한 사용될 수 있다.

[0133] 본 명세서에 제공된 조성물 및 방법의 특정 실시형태와 함께 사용될 수 있는 전위 시스템의 더 많은 예는 스타필로코커스 아우레우스(*Staphylococcus aureus*) Tn552(Colegio et al., J. Bacteriol., 183: 2384-8, 2001; Kirby C et al., Mol. Microbiol., 43: 173-86, 2002), Ty1(문헌[Devine & Boeke, Nucleic Acids Res., 22: 3765-72, 1994] 및 국제 특허 출원 공개 WO 95/23875), 트랜스포존 Tn7(Craig, N L, Science. 271: 1512, 1996; Craig, N L, Review in: Curr Top Microbiol Immunol., 204:27-48, 1996), Tn/O 및 IS10(Kleckner N, et al., Curr Top Microbiol Immunol., 204:49-82, 1996), 마리너(Mariner) 트랜스포사제(Lampe D J, et al., EMBO J., 15: 5470-9, 1996), Tc1(Plasterk R H, Curr. Topics Microbiol. Immunol., 204: 125-43, 1996), P 요소(Gloor, G B, Methods Mol. Biol., 260: 97-114, 2004), Tn3(Ichikawa & Ohtsubo, J Biol. Chem. 265:18829-32, 1990), 박테리아 삽입 서열(Ohtsubo & Sekine, Curr. Top. Microbiol. Immunol. 204: 1-26, 1996), 레트로바이러스(Brown, et al., Proc Natl Acad Sci USA, 86:2525-9, 1989), 및 효모의 레트로트랜스포

존(Boeke & Corces, Annu Rev Microbiol. 43:403-34, 1989)을 포함한다. 더 많은 예는 IS5, Tn10, Tn903, IS911, 잠자는 미녀(Sleeping Beauty), SPIN, hAT, 피기백(PiggyBac), 헤르메스(Hermes), TcBuster, AeBuster1, Tol2 및 트랜스포사제 패밀리 효소의 공학 처리 형태를 포함한다(Zhang et al., (2009) PLoS Genet. 5:e1000689. Epub 2009 Oct 16; Wilson C. et al (2007) J. Microbiol. Methods 71:332-5).

[0134] 본 명세서에 제공된 방법 및 조성물과 함께 사용될 수 있는 인테그라제의 더 많은 예는 레트로바이러스 인테그라제 및 이러한 레트로바이러스 인테그라제, 예컨대 HIV-1, HIV-2, SIV, PFV-1, RSV로부터의 인테그라제에 대한 인테그라제 인식 서열을 포함한다.

#### [0135] 바코드

[0136] 일반적으로, 바코드는 하나 이상의 특정 핵산을 동정하기 위해 사용될 수 있는 하나 이상의 뉴클레오타이드 서열을 포함할 수 있다. 바코드는 인공 서열일 수 있거나, 또는 전위 동안 생성된 천연 유래 서열, 예컨대 이전에 병치된 DNA 단편의 말단에서의 동일한 축적 게놈 DNA 서열(g-암호)일 수 있다. 일부 실시형태에서, 바코드는 표적 핵산 서열에 없고 하나 이상의 표적 핵산 서열을 동정하기 위해 사용될 수 있는 인공 서열이다.

[0137] 바코드는 적어도 약 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개 이상의 연이은 뉴클레오타이드를 포함할 수 있다. 일부 실시형태에서, 바코드는 적어도 약 10, 20, 30, 40, 50, 60, 70, 80, 90, 100개 이상의 연이은 뉴클레오타이드를 포함한다. 일부 실시형태에서, 바코드를 포함하는 핵산 집단에서 바코드의 적어도 일부는 상이하다. 일부 실시형태에서, 바코드의 적어도 약 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%는 상이하다. 더 많은 이러한 실시형태에서, 모든 바코드는 상이하다. 바코드를 포함하는 핵산 집단에서 상이한 바코드의 다양성은 무작위로 생성되거나 또는 비무작위로 생성될 수 있다.

[0138] 일부 실시형태에서, 트랜스포존 서열은 적어도 하나의 바코드를 포함한다. 일부 실시형태에서, 예컨대 트랜스포존은 2개의 비인접 트랜스포존 서열을 포함하며, 제1 트랜스포존 서열은 제1 바코드를 포함하고, 제2 트랜스포존 서열은 제2 바코드를 포함한다. 일부 실시형태에서, 트랜스포존 서열은 제1 바코드 서열 및 제2 바코드 서열을 포함하는 바코드를 포함한다. 앞서 언급한 실시형태의 일부에서, 제1 바코드 서열은 제2 바코드 서열과 짝지어지도록 동정 또는 표기될 수 있다. 예를 들어, 공지된 제1 바코드 서열은 서로 짝지어지는 것으로 알려진 복수의 제1 바코드 서열 및 제2 바코드 서열을 포함하는 참조표를 이용하여 공지된 제2 바코드 서열과 짝지어지는 것으로 알려질 수 있다.

[0139] 다른 예에서, 제1 바코드 서열은 제2 바코드 서열과 동일한 서열을 포함할 수 있다. 다른 예에서, 제1 바코드 서열은 제2 바코드 서열의 역 상보체를 포함할 수 있다. 일부 실시형태에서, 제1 바코드 서열과 제2 바코드 서열은 상이하다. 제1 바코드 서열과 제2 바코드 서열은 2-암호를 포함할 수 있다.

[0140] 본 명세서에 기재된 조성물 및 방법의 일부 실시형태에서, 바코드는 주형 핵산의 제조에서 사용된다. 이해될 바와 같이, 다수의 이용 가능한 바코드는 각각의 주형 핵산 분자가 독특한 식별을 포함하게 한다. 주형 핵산의 혼합물에서 각각의 분자의 독특한 식별은 몇몇 용도에 사용될 수 있다. 예를 들어, 독특하게 식별된 분자는 다중 염색체를 갖는 샘플에서, 게놈에서, 세포에서, 세포 유형에서, 세포 질환 상태에서, 및 종에서, 예를 들어, 단상형 서열분석에서, 모 대립유전자 식별력에서, 메타게놈 서열분석에서, 그리고 게놈의 샘플 서열분석에서, 개개 핵산 분자를 식별하는 데 사용될 수 있다.

[0141] 예시적인 바코드 서열은 TATAGCCT, ATAGAGGC, CCTATCCT, GGCTCTGA, AGGCGAAG, TAATCTTA, CAGGACGT, 및 GTACTGAC를 포함하지만, 이들로 제한되지 않는다.

#### [0142] 프라이머 부위

[0143] 일부 실시형태에서, 트랜스포존 서열은 "서열분석 어댑터" 또는 "서열분석 어댑터 부위", 즉, 프라이머에 혼성화할 수 있는 하나 이상의 부위를 포함하는 영역을 포함할 수 있다. 일부 실시형태에서, 트랜스포존 서열은 증폭, 서열분석 등에 유용한 적어도 제1 프라이머 부위를 포함할 수 있다. 서열 결합 부위의 예시적인 서열은 AATGATACGGCGACCACCGAGATCTACAC(P5 서열) 및 CAAGCAGAAGACGGCATACGAGAT(P7 서열)를 포함하지만, 이들로 제한되지 않는다.

#### [0144] 표적 핵산

[0145] 표적 핵산은 관심 대상의 임의의 핵산을 포함할 수 있다. 표적 핵산은 DNA, RNA, 펩타이드 핵산, 몰폴리노 핵산, 잠금(locked) 핵산, 글리콜 핵산, 트레오스 핵산, 핵산의 혼합 샘플, 배수체 DNA(즉, 식물 DNA), 이들의

혼합물 및 이들의 혼성체를 포함할 수 있다. 바람직한 실시형태에서, 게놈 DNA 또는 이의 증폭된 복제물은 표적 핵산으로서 사용된다. 다른 바람직한 실시형태에서, cDNA, 미토콘드리아 DNA 또는 엽록체 DNA가 사용된다. 일부 실시형태에서, 표적 핵산은 mRNA이다.

[0146] 일부 실시형태에서, 표적 핵산은 단일 세포로부터 또는 단일 세포의 분획으로부터 유래된다. 일부 실시형태에서, 표적 핵산은 단일 세포소기관으로부터 유래된다. 예시적인 단일 세포소기관은 단일 핵, 단일 미토콘드리아 및 단일 리보솜을 포함하지만, 이들로 제한되지 않는다. 일부 실시형태에서, 표적 핵산은 포말린 고정 파라핀 포매(FFPE) 샘플로부터 유래된다. 일부 실시형태에서, 표적 핵산은 가교된 핵산이다. 일부 실시형태에서, 표적 핵산은 단백질과 가교된다. 일부 실시형태에서, 표적 핵산은 가교된 DNA이다. 일부 실시형태에서, 표적 핵산은 히스톤 보호된 DNA이다. 일부 실시형태에서, 히스톤은 표적 핵산으로부터 제거된다. 일부 실시형태에서, 표적 핵산은 뉴클레오솜으로부터 유래된다. 일부 실시형태에서, 표적 핵산은 핵 단백질로부터 제거된 뉴클레오솜으로부터 유래된다.

[0147] 표적 핵산은 임의의 뉴클레오타이드 서열을 포함할 수 있다. 일부 실시형태에서, 표적 핵산은 동종중합체 서열을 포함한다. 표적 핵산은 또한 반복부 서열을 포함할 수 있다. 반복부 서열은, 예를 들어, 2, 5, 10, 20, 30, 40, 50, 100, 250, 500 또는 1000개 이상의 뉴클레오타이드를 포함하는 임의의 다양한 길이일 수 있다. 반복부 서열은, 예를 들어, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 또는 20회 이상을 포함하는 임의의 다양한 횟수로 인접하여 또는 비인접하여 반복될 수 있다.

[0148] 본 명세서에 기재된 일부 실시형태는 단일 표적 핵산을 이용할 수 있다. 다른 실시형태는 복수의 표적 핵산을 이용할 수 있다. 이러한 실시형태에서, 복수의 표적 핵산은 복수의 동일한 표적 핵산, 일부 표적 핵산이 동일한 복수의 상이한 표적 핵산 또는 모든 표적 핵산이 상이한 복수의 표적 핵산을 포함할 수 있다. 복수의 표적 핵산을 이용하는 실시형태는 다중복합 형식으로 수행될 수 있으며, 따라서, 시약은 표적 핵산에, 하나 이상의 챔버 또는 어레이 표면 상에 동시에 전달된다. 일부 실시형태에서, 복수의 표적 핵산은 실질적으로 모든 특정 유기체 게놈을 포함할 수 있다. 복수의 표적 핵산은, 예를 들어, 게놈의 적어도 약 1%, 5%, 10%, 25%, 50%, 75%, 80%, 85%, 90%, 95% 또는 99%를 포함하는 특정 유기체 게놈의 적어도 일부를 포함할 수 있다. 특정 실시형태에서, 일부는 게놈의 약 1%, 5%, 10%, 25%, 50%, 75%, 80%, 85%, 90%, 95% 또는 99% 이하인 상한을 가질 수 있다.

[0149] 표적 핵산은 임의의 공급원으로부터 얻어질 수 있다. 예를 들어, 표적 핵산은 단일 유기체로부터 얻어진 핵산 분자로부터 또는 하나 이상의 유기체를 포함하는 천연 공급원으로부터 얻어진 핵산 분자의 집단으로부터 제조될 수 있다. 핵산 분자의 공급원은 세포소기관, 유기체, 세포, 조직, 기관 또는 유기체를 포함하지만, 이들로 제한되지 않는다. 표적 핵산 분자의 공급원으로서 사용될 수 있는 세포는 원핵세포(박테리아 세포, 예를 들어, 에스케리키아(*Escherichia*), 바실러스(*Bacillus*), 세라티아(*Serratia*), 살모넬라(*Salmonella*), 스탕필로코커스(*Staphylococcus*), 스트렙토코커스(*Streptococcus*), 클로스트리듐(*Clostridium*), 클라미디아(*Chlamydia*), 네이세리아(*Neisseria*), 트레포네마(*Treponema*), 마이코플라즈마(*Mycoplasma*), 보렐리아(*Borrelia*), 레지오넬라(*Legionella*), 슈도모나스(*Pseudomonas*), 마이코박테리움(*Mycobacterium*), 헬리코박터(*Helicobacter*), 에르위니아(*Erwinia*), 아르코박테리움(*Agrobacterium*), 리조비움(*Rhizobium*) 및 스트렙토마이세스 속(*Streptomyces* genera)); 고세균, 예컨대 크렌아케오타, 나노아케오타(nanoarchaeota) 또는 유리아케오타(euryarchaeotia); 또는 진핵생물, 예컨대 진균, (예를 들어, 효모), 식물, 원생동물 및 기타 기생충, 및 동물(곤충(예를 들어, 초파리 종)을 포함, 선충(예를 들어, 예쁜꼬마선충(*Caenorhabditis elegans*)) 및 포유류(예를 들어, 래트, 마우스, 원숭이, 비-인간 영장류 및 인간)일 수 있다. 표적 핵산 및 주형 핵산은 당업계에 잘 공지된 다양한 방법을 이용하여 관심 대상의 특정 서열에 대해 풍부하게 될 수 있다. 이러한 방법의 예는 국제 특허 출원 공개 WO/2012/108864에서 제공되며, 이는 본 명세서에 그의 전문이 참고로 포함된다. 일부 실시형태에서, 핵산은 주형 라이브러리를 제조하는 방법 동안 추가로 풍부하게 될 수 있다. 예를 들어, 핵산은 트랜스포좀의 삽입 전에 트랜스 포좀의 삽입 후에 및/또는 핵산의 증폭 후에 풍부하게 될 수 있다.

[0150] 추가로, 일부 실시형태에서, 표적 핵산 및/또는 주형 핵산은 고도로 정제될 수 있고, 예를 들어, 핵산은 본 명세서에 제공된 방법과 함께 사용하기 전에 오염물질로부터 적어도 약 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 유리될 수 있다. 일부 실시형태에서, 표적 핵산의 품질 및 크기를 유지하는 당업계에 공지된 방법을 사용하는 것이 유리하며, 예를 들어 표적 DNA의 단리 및/또는 직접적 전위는 아가로스 플러그를 이용하여 수행될 수 있다. 전위는 또한 세포에서 직접적으로, 세포 집단, 용해물 및 비 정제 DNA와 함께 수행될 수 있다.

[0151] 일부 실시형태에서, 표적 핵산은 생물학적 샘플 또는 환자 샘플로부터 얻을 수 있다. 본 명세서에서 사용되는 용어 "생물학적 샘플" 또는 "환자 샘플"은 조직 및 체액과 같은 샘플을 포함한다. "체액"은 혈액, 혈청, 혈장,



타액, 뇌척수액, 흉수, 눈물, 유미액 및 정액을 포함하지만, 이들로 제한되지 않는다. 샘플은 "무세포"인 체액을 포함할 수 있다. "무세포 체액"은 약 1%(w/w) 미만의 전세포 물질을 포함한다. 혈장 또는 혈청은 무세포 체액의 예이다. 샘플은 천연 또는 합성 유래의 표본(즉, 무세포가 되도록 이루어진 세포 샘플)을 포함할 수 있다.

[0152] 상기 개시된 방법의 일부 실시형태에서, 표적 핵산은 트랜스포좀에 대한 표적 핵산의 노출 전에 (예를 들어, 음과처리에 의해, 제한 분해에 의해, 다른 기계적 수단에 의해) 단편화될 수 있다.

[0153] 본 명세서에서 사용되는 용어 "혈장"은 혈액 중에서 발견된 무세포 유체를 지칭한다. "혈장"은 당업계에 공지된 방법(예를 들어, 원심분리, 여과 등)에 의해 혈액으로부터 전세포 물질을 제거함으로써 혈액으로부터 얻을 수 있다.

[0154] 달리 특정되지 않는 한, 단수의 용어는 본 출원 전체적으로 "하나 이상"을 의미한다.

[0155] 용어 "예를 들어", "예로서", "예컨대", "포함하다", "포함하는" 또는 이들의 변형이 본 명세서에서 사용될 때, 이들 용어는 제한의 용어가 되는 것으로 여겨져서는 안 될 것이며, "제한되지 않는" 또는 "제한 없이"를 의미하는 것으로 해석될 것이다.

[0156] 다음의 실시예는 예시적인 실시형태를 제공하며, 본 명세서에 제공된 발명을 임의의 방법으로 제한하지 않는다.

[0157] 실시예

[0158] 실시예 1— 비드 기반 태그화 과정으로부터의 DNA 클러스터 수율

[0159] 도 3의 비드 기반 태그화 과정으로부터의 DNA 클러스터 수율을 평가하고 나서, 도 4의 표에 나타낸다. 이 예에서, 태그화 비드(2.8 $\mu$ m 비드)의 동일 배치를 이용하여 50, 250 및 1000ng의 인간 NA12878 DNA를 태그화하였다. NA12878 DNA의 두번째 50ng 분취액을 태그화 비드(전체 반복부; 2.8 $\mu$ m 비드)의 두 번째 배치를 이용하여 태그화하였다. 비드-결합된 태그화된 DNA 샘플을 PCR 증폭시키고 나서, 정제하였다. 각각의 정제된 PCR 산물(비정량화)의 분취액(5.4 $\mu$ l)을 270배 희석시켜 약 50pM의 저장 샘플 용액을 만들었다. 각각의 샘플에 대해, 50pM 저장액을 15, 19, 21 및 24 pM로 희석시켰다. 희석 샘플을 클러스터 생성 및 서열분석을 위해 유세포 상에 부하하였다. 데이터는 샘플 희석(대략 50pM)으로부터 시작해서, 클러스터 수가 동일한 세트의 비드를 이용하여 3가지의 상이한 투입 수준(즉, 50, 250, 및 1000ng)에 대해 100 내지 114%라는 것을 나타낸다. (상이한 배치의 비드를 지니는) 50ng 전체 반복부에 대한 클러스터 수는 81%였다. 상이한 희석(15, 19, 21, 및 24pM)으로 약 10% 내에서 동일한 수의 클러스터를 수득한다. 데이터는 비드가 수율을 크게 제어하고 있고, 수율은 상이한 DNA 투입 및 상이한 반복부에 대해 재현 가능하다는 것을 나타낸다.

[0160] 실시예 2— 비드 기반 태그화 과정의 재현성

[0161] 도 3의 비드 기반 태그화 과정의 재현성을 도 5에 나타낸다. 이 예에서, 50 및 500ng의 투입 NA12878 DNA를 이용하여 태그화된 DNA를 제조하기 위해 "동일한" 트랜스포좀 밀도에서 생성된 지표화된 비드(지표 1 내지 6; 2.8 $\mu$ m 비드)의 6가지 상이한 제조를 사용하였다. 태그화된 DNA를 PCR 증폭 및 정제하였다. 12개의 정제 PCR 산물을 2개의 HiSeq 레인에 대한 6개의 두 혼합물(폴 1 및 폴 2)에 풀링하였다. 각각의 폴은 레인 당 3 내지 50ng 및 3 내지 500ng 샘플을 포함한다. 데이터 표(500)는 각각의 지표 샘플에 대한 중앙값 삽입 크기 및 평균 삽입 크기를 나타낸다.

[0162] 실시예 3— 폴 1의 삽입 크기 및 폴 2의 삽입 크기

[0163] 폴 1의 삽입 크기 및 폴 2의 삽입 크기를 도 5의 지표 샘플의 도 6A(플롯 600) 및 도 6B(플롯 650)에 각각 나타낸다. 데이터는 또한 삽입 크기가 지표화된 비드의 6개 상이한 제조 간에 균일하다는 것을 나타낸다. 비드-기반 태그화는 삽입 크기 및 DNA 수율을 제어하는 메커니즘을 제공한다.

[0164] 실시예 4— 총 관독 수의 재현성

[0165] 도 5에 기재한 실험에 대해 정렬된 총 관독 수 및 관독%의 재현성을 도 7에 나타낸다(막대 그래프 700). 입력 둘 다에서(50ng 및 500ng), 총 관독 수는 동일한 지표화된 비드 제조에 대해 유사하다. 6개의 지표화된 비드 제조 중 넷(지표 1, 2, 3 및 6)은 매우 유사한 수율을 가지며; 나타난 지표화된 비드 제조 4 및 5는 지표 서열에 기인할 수 있는 일부 가변성을 나타내었다.

[0166] 일 적용에서, 비드 기반 태그화 과정은 태그화 단계를 포함하는 엑소 풍부 분석, 예를 들어, 일루미나사(Illumina)의 넥스테라(Nextera)(등록상표) 빠른 포획 강화(Rapid Capture Enrichment) 프로토콜에서 사용될



수 있다. 현재의 엑손 강화 분석(즉, 일루미나의 넥스테라 빠른 포획 풍부 프로토콜)에서, 게놈 DNA를 단편화하기 위해 용액 기반 태그화(넥스테라)를 사용하였다. 이어서, 유전자 특정 프라이머를 사용하여 관심 대상의 특정 유전자 단편을 풀다운시켰다. 2회의 강화 주기를 수행하고, 이어서 풀다운된 단편을 PCR에 의해 강화시키고 나서, 서열분석하였다.

[0167] 엑손 강화 분석에서 비드 기반 태그화 과정의 사용을 평가하기 위해, 25, 50, 100, 150, 200 및 500ng의 입력 DNA를 이용하여 인간 NA12878 DNA를 태그화하였다. 표준 프로토콜에 따라 50ng 입력 DNA로부터 대조군 라이브러리(NA00536)를 제조하였다. 각각의 DNA 입력은 상이한 표지(독특한 식별자)를 가졌다. 표준 방법과 매칭시키기 위해 강화된 중합효소 마스터믹스(EPM)를 이용하는 10회 주기의 PCR을 사용하였으며, 그리고 충분한 양의 단편을 보장하도록 풀다운을 위해 제공하였다. 증폭 프로토콜은 72℃에서 3분, 98℃에서 30초, 다음에 10주기의 98℃에서 10초, 65℃에서 30초, 및 72℃에서 1분이었다. 이어서, 샘플을 10℃에서 유지하였다. 이어서, 샘플을 엑손 강화 풀다운 과정을 통해 처리하고 나서, 서열분석하였다.

[0168] 실시예 5— 엑손 강화 분석에서 대조군 및 비드 기반 태그화된 라이브러리의 삽입 크기

[0169] 도 8A, 8B 및 8C는 엑손 강화 분석에서 대조군 라이브러리에서 삽입 크기의 플롯(800), 비드 기반 태그화된 라이브러리에서 삽입 크기의 플롯(820) 및 요약 데이터 표 840를 각각 나타낸다. 데이터는 비드 기반 태그화 라이브러리가 대조군 라이브러리에 비해서 더 넓은 펼쳐진 삽입 크기를 갖지만, 삽입 크기는 샘플에 대한 DNA 입력과 관계없이 매우 유사하다.

[0170] 실시예 6 — 관독 서열의 품질

[0171] 도 9A, 9B 및 9C는 도 8A, 8B 및 8C의 엑손 강화 분석에서 중복 통과 필터(dups PF) 백분율의 막대 그래프(900), PCT 선택 염기의 막대 그래프(920) 및 표적에 대한 PCT 사용 가능한 염기의 막대 그래프(940)를 각각 나타낸다. 도 9A에 대해, dups PF%는 얼마나 많은 관독이 유세포의 다른 곳에서 중복되는지의 측정이다. 이 수는 모든 클러스터가 데이터를 결과에 유용하게 이끈다는 것을 보장하도록 이상적으로 (본 명세서에서와 같이) 낮을 것이다.

[0172] 도 9B는 PCT 선택 염기를 나타내는데, 이는 강화 과정 동안 강화된 삽입 부위에서 또는 삽입 부위 근처에서 서열 분석한 관독비의 측정이다. 이상적으로, 이 수는 강화 과정의 성공을 반영하기 위해 1에 가까울 것이고, 강화되어서는 안 되는 관독은 과정을 통해 얻어지지 않음을 나타낸다.

[0173] 도 9C는 표적 상의 PCT 이용 가능 염기를 나타내는데, 이는 강화 영역 내에서 관심 대상의 특정 염기에 대해 실제로 서열 분석한 관독비의 측정이다. 이상적으로, 모든 강화된 관독은 강화된 관독 내의 관심 대상 염기에 대한 서열 분석이지만, 태그화의 무작위 특성 및 삽입의 가변적 길이에 기인하여, 관심 대상 영역에 대해 결국 서열 분석되지 않는 관독을 강화할 수 있다.

[0174] 삽입 크기 분포를 최적화하기 위해 두 기법을 사용할 수 있다. 일 예에서, SPRI 세정을 사용하여 너무 작거나 또는 너무 큰 단편을 제거할 수 있다. SPRI 세정은 크기에 기반한 선택적 DNA 침전에 의해 그리고 원한다면 침전 또는 비침전 DNA의 체류 중 하나에 의해 목적으로 하는 크기보다 더 크거나 또는 더 작은 단편을 제거하는 과정이다(즉, 제1 단계는 목적으로 하는 크기보다 더 크고 가용성의 더 작은 단편을 보유하는 DNA만을 침전시키는 것이다). 이어서, 더 작은 단편이 추가로 침전되며, 이때에 (여전히 용액 중의) 원치 않는 매우 작은 단편을 제거하고, 침전된 DNA를 보유하고, 세척한 다음, 재용해시켜 목적으로 하는 크기 범위의 DNA를 제공하였다. 다른 예에서, 비드 표면 상에서의 활성 트랜스포좀의 스페이싱을 사용하여 삽입 크기 분포를 제어할 수 있다. 예를 들어, 비드 표면 상의 겹을 비활성 트랜스포좀(예를 들어, 비활성 트랜스포존을 지니는 트랜스포좀)으로 채울 수 있다.

[0175] 비드 기반 태그화 과정의 접근을 평가하였다. 표 3은 0, 1, 2 또는 3회 관독이 지표를 공유하는 1000bp 창 내에서 생긴다는 것을 나타낸다. 9개의 상이한 지표 트랜스포좀을 지니는 비드를 생성하고 나서, 소량의 인간 DNA를 태그화하기 위해 사용하였다. 관독을 생성하고 나서, 정렬시키고, 동일 지표를 공유한 1000bp 또는 10Kb 창 내에서의 관독 수에 대해 분석하였다. 지표를 우연히 생성할 수 있고, 공유하는 작은 창 내에서의 이것이 일어날 가능성이 얼마나 많은지의 예측을 표 3 및 표 4의 "무작위" 행에서 제공한다. "비드" 행의 수는 지표를 공유하는 1000bp(표 3) 또는 10Kb(표 4) 창의 실제 수를 나타낸다. 표 3 및 표 4에 나타낸 바와 같이, 동일한 지표가 1000bp 또는 10Kb 창 내에서 발견된 실제 횟수는 무작위 경우에 예상된 것보다 상당히 더 크다. "0" 창은 특정 1000bp 창이 그에 대해 지표화된 관독 맵핑을 갖지 않는 모든 횟수를 나타낸다. 매우 소량의 인간 게놈만이 서열이고 대부분의 창은 그들에 대한 관독 정렬을 갖지 않기 때문에 숫자는 본 명세서에서 가장 크다. "1"은

1000bp(또는 10Kb) 창에 대한 단지 하나의 판독 맵의 수이며; "2"는 2회 이상의 판독 수가 1000bp(또는 10KB) 창 등에서 지표를 공유하는 것이다. 이 데이터는 동일한 조각의 DNA(10Kb 이상)가 약 15000회 태그화 사건 중에서 1400 이상의 경우가 적어도 2배 및 5배까지 동일한 비드에 의해 태그화된다는 것을 시사한다. 단편은 지표를 공유하기 때문에, 그들은 우연히 거기에 있을 가능성은 없지만, 동일한 비드로부터 초래된다.

표 3

지표를 공유하는 1000bp 창에서의 판독 수				
	0	1	2	3
비드	25913666	15220	305	7
무작위	25913334	15855	9	0

표 4는 지표를 공유하는 10kb 창 내의 판독 수(5개까지)를 나타낸다.

표 4

지표를 공유하는 10kb 창에서의 판독 수						
	0	1	2	3	4	5
비드	2578669	12683	1267	169	28	3
무작위	2577012	15742	64	1	0	0

실시예 7— CPT-DNA로부터의 자유 트랜스포좀의 분리

전위 후에, CPT-DNA 및 자유 트랜스포좀을 포함하는 반응 혼합물에 세파크릴(Sephacryl) S-400 및 세파크릴 S-200 크기 배제 크로마토그래피를 이용하는 칼럼 크로마토그래피를 실시하고 나서, 도 22에 나타낸다. CPT-DNA를 NCP DNA로서 표시한다.

실시예 8— 비드에 대한 포획 프로브 밀도의 최적화

포획 프로브 A7 및 B7의 밀도를 1 $\mu$ m 비드 상에서 최적화하고 나서, 결과를 도 25에 나타내었다. 레인 1(A7) 및 3(B7)은 더 큰 프로브 밀도를 가졌고, 레인 2 (A7) 및 4(B7)는 1 $\mu$ m 비드 당 추정치 10,000 내지 100,000의 프로브 밀도를 가졌다. 표적 분자에 대한 포획 프로브의 결합 산물을 아가로스겔에서 평가하였다. 비드 당 대략 10,000 내지 100,000의 프로브 밀도는 더 높은 프로브 밀도를 지니는 것보다 더 양호한 결합 효율을 가졌다.

실시예 9— 분자내 혼성화에 의해 비드 상에서 CPT-DNA의 지표화된 서열분석 라이브러리를 제조하는 것의 실현 가능성 시험

비드 상에서 A7 및 B7 포획 서열에 대해 상보성인 A7' 및 B7' 포획 서열을 갖는 트랜스포존을 초활성 Tn5 트랜스포사제와 혼합함으로써 트랜스포좀을 제조하였다. 고분자량 계능 DNA를 트랜스포좀과 혼합하여 CPT-DNA를 생성한다. 별도로, 비드를 고정 올리고뉴클레오타이드: P5-A7, P7-B7 또는 P5-A7+ P7-B7을 이용하여 제조하고, 여기서 P5 및 P7은 프라이머 결합 서열이며, A7 및 B7은 A7' 및 B7' 서열에 대해 각각 상보성인 포획 서열이다. P5-A7 단독, P7-B7 단독, P5-A7+ P7-B7, 또는 P5-A7 및 P7-B7비드의 혼합물을 포함하는 비드를 CPT-DNA로 처리하고 나서, 리가제를 반응 혼합물에 첨가하여 전위된 DNA에 대한 고정 올리고의 혼성화 효율을 결정하였다. 결과를 도 26에 나타낸다. 아가로스 겔 상의 고분자량 밴드에 의해 나타내는 바와 같이 P5-A7 및 P7-B7을 하나의 비드 상에서 함께 고정시킬 때에만 서열분석 라이브러리가 제조된다(레인 4). 결과는 고효율의 분자내 혼성화를 나타내고, 분자내 혼성화에 의해 비드 상에서 CPT-DNA의 지표화된 서열분석 라이브러리를 제조하는 것의 실현 가능성을 증명한다.

실시예 10— 클론 지표화의 실현 가능성 시험.

트랜스포좀의 몇몇 세트를 준비하였다. 일 세트에서, 초활성 Tn5 트랜스포사제를 5' 바이오틴을 지니는 트랜스포존 서열 Tnp1과 혼합하여 트랜스포좀 1을 준비하였다. 다른 세트에서, 5' 바이오틴을 지니는 독특한 지표2를 갖는 Tnp2로 트랜스포좀 2를 준비하였다. 다른 세트에서, 트랜스포좀 3을 제조하기 위해 초활성 Tn5 트랜스포사제를 5' 바이오틴을 지니는 트랜스포존 서열 Tnp3과 혼합한다. 다른 세트에서, 독특한 지표 4 및 5'-바이오틴을 갖는 Tnp4로 트랜스포좀 4를 제조한다. 각각의 트랜스포좀 1 및 2 및 트랜스포좀 3 및 4를 스트렙타비딘 비드와 별도로 혼합하여 비드 세트 1 및 비드 세트 2를 생성한다. 이어서, 비드의 2개 세트를 함께 혼합하고 나서, 게

놈 DNA 및 태그화 완충제와 인큐베이션시켜 게놈 DNA의 태그화를 촉진시킨다. 이어서, 이에 태그화한 서열의 PCR 증폭시킨다. 증폭 DNA를 서열분석하여 지표 서열의 삽입을 분석한다. 태그화가 비드로 국한된다면, 대다수의 단편을 Tnp1/Tnp2 및 Tnp3/Tnp4 지표로 암호화할 것이다. 분자내 혼성화가 있다면, 단편을 Tnp1/Tnp4, Tnp2/Tnp3, Tnp1/Tnp3 및 Tnp2/Tnp4 지표로 암호화할 수 있다. 5 및 10 주기의 PCR 후에 서열분석 결과를 도 27에 나타내었다. 대조군은 모두 4개의 함께 혼합된 트랜스포존을 가지며, 비드 상에 고정시켰다. 결과는 대다수의 서열이 Tnp1/Tnp2 또는 Tnp3/Tnp4 지표를 가진다는 것을 나타내는데, 이는 클론 지표화가 실현 가능하다는 것을 나타낸다. 대조군은 지표 사이에 차이를 나타내지 않는다.

[0187] 실시예 11—단일 반응에서 지표화된 클론 비드 전위

[0188] 96개의 지표화된 트랜스포좀 비드 세트를 준비한다. 5'-말단에서 Tn5 모자이크 말단 서열(ME) 및 지표 서열을 포함하는 올리고뉴클레오타이드를 포함하는 트랜스포존을 혼합함으로써 개개의 지표화된 트랜스포좀을 제조하였다. 개개로 지표화된 트랜스포좀을 스트랩타비딘-바이오틴 상호작용을 통해 비드 상에 고정시켰다. 비드 상의 트랜스포좀을 세척하고 나서, 비드 상에서 모든 96개의 개개로 지표화된 트랜스포좀을 풀링하였다. ME 서열에 대해 상보성이고 지표 서열을 포함하는 올리고뉴클레오타이드를 고정된 올리고뉴클레오타이드로 어닐링하여 독특한 지표를 생성한다. 96개의 클론 지표 트랜스포좀 비드 세트를 합하고 나서, 단일 관에서 넥스테라 태그화 완충제의 존재 하에 고분자량(HMW) 게놈 DNA와 함께 인큐베이션시킨다.

[0189] 비드를 세척하고 나서, 0.1% SDS로 반응 혼합물을 처리함으로써 트랜스포사제를 제거한다. 태그화된 DNA를 지표화한 프라이머를 이용하여 증폭시키고 나서, TrueSeq v3 클러스터 키트를 이용하여 PE HiSeq 유세포 v2로 서열 분석하고, 서열분석 데이터를 분석한다.

[0190] 관독의 클러스터 또는 섬을 관찰한다. 각각의 서열에 대한 관독 사이의 가장 가까운 이웃 거리의 플롯은 본질적으로 주요 피크를 나타낸다(클러스터 내에서부터의(근위) 하나 및 클러스터 사이로부터의 다른 하나(원위)). 개략적 방법 및 결과를 도 30 및 도 31에 나타낸다. 섬 크기는 대략 3 내지 10kb의 범위에 있다. 뒤덮인 염기의 백분율은 대략 5 내지 10%이다. 게놈 DNA의 삽입 크기는 대략 200 내지 300개의 염기이다.

[0191] 실시예 12— 비드 상의 트랜스포좀에 대한 라이브러리 크기

[0192] ME' 서열을 갖는 제1 올리고뉴클레오타이드, ME-바코드-P5/P7 서열을 갖는 제2 올리고뉴클레오타이드와 Tn5 트랜스포사제를 혼합함으로써 트랜스포좀을 용액 중에서 처음 조립한다. 제1 세트에서, ME' 서열을 갖는 제1 올리고뉴클레오타이드는 3'-말단에서 바이오틴일화된다. 제2 경우에, ME-바코드-P5/P7 서열을 갖는 올리고뉴클레오타이드는 5'-말단에서 바이오틴일화된다. 트랜스포좀이 스트랩타비딘 비드 상에서 고정되도록 다양한 농도(10nM, 50nM 및 200nM)의 각각의 얻어진 트랜스포좀 세트에 스트랩타비딘 비드를 첨가한다. 비드를 세척하고 나서 HMW 게놈 DNA를 첨가하고, 태그화를 수행한다. 일부 경우에, 0.1% SDS로 태그화된 DNA를 처리하고 나서, 다른 경우에, 태그화된 DNA를 비처리한다. 5 내지 8회 주기로 태그화된 DNA를 PCR 증폭시키고, 서열분석한다. 도식을 도 32에 나타낸다.

[0193] 도 33에 나타난 바와 같이, SDS의 처리는 증폭 효율 및 서열분석 품질을 개선시킨다. 3'-바이오틴을 지니는 올리고뉴클레오타이드는 트랜스포좀에 대해 더 양호한 라이브러리 크기를 가진다.

[0194] 도 34는 삽입 크기에 대한 트랜스포좀 표면 밀도 효과를 나타낸다. 5'-바이오틴을 지니는 트랜스포좀은 더 작은 크기의 라이브러리 및 더 많은 자기-삽입 부산물을 나타낸다.

[0195] 실시예 13— 입력 DNA의 적정

[0196] 다양한 양의 표적 HMW DNA를 50mM Tn5:트랜스포존 밀도로 클론에 의해 지표화된 비드에 첨가하고 나서, 37°C에서 15 또는 60분 동안 또는 실온에서 60분 동안 인큐베이션시켰다. 트랜스포좀은 3'-바이오틴을 지니는 올리고뉴클레오타이드를 포함하였다. 태그화를 수행하고 나서, 반응 혼합물을 0.1% SDS로 처리하고, PCR 증폭시켰다. 증폭 DNA를 서열분석하였다. 도 35는 크기 분포에 대한 입력 DNA 효과를 나타낸다. 10pg의 입력 DNA와의 반응은 최소 신호를 나타내었다. 크기 분포 패턴은 20, 40 및 200pg의 범위에 있는 DNA 입력에 대해 유사하였다.

[0197] 실시예 14— 용액 기반 및 비드 기반 방법을 이용하는 섬 크기 및 분포

[0198] 용액 기반 및 비드 기반 방법을 이용하여 섬 크기 및 분포를 비교한다. 용액 기반 접근에서, 트랜스포존에서 독특한 지표를 각각 지니는 96개의 트랜스포좀을 96웰 플레이트에서 조립한다. HMW 게놈 DNA를 첨가하고 나서, 태그화 반응을 수행한다. 반응 산물을 0.1% SDS로 처리하고 나서, PCR 증폭시킨다. 증폭 산물을 서열분석하였다.

- [0199] 비드 기반 접근에서, 트랜스포존에서 독특한 지표를 각각 지니는 96 트랜스포존을 96웰 플레이트에서 조립한다. 올리고뉴클레오타이드는 3'-말단 바이오틴을 포함하였다. 스트랩타비딘 비드를 각각의 96 웰 플레이트에 첨가하고 나서, 트랜스포존이 스트랩타비딘 비드 상에 고정되도록 인큐베이션시킨다. 비드를 개개로 세척하고 나서, 풀링하고, HMW 게놈 DNA를 첨가하고 나서, 태그화 반응을 단일 반응 용기에서 수행한다(원 포트). 반응 산물을 0.1% SDS로 처리하고 나서, PCR 증폭시켰다. 증폭 산물을 서열분석하였다.
- [0200] 음성 대조군에서, 각각 독특한 지표를 지니는 모두 96개의 트랜스포존 서열을 처음에 함께 혼합한다. 올리고뉴클레오타이드는 3'-말단 바이오틴을 포함하였다. 개개로 혼합된 지표화된 트랜스포존으로부터 트랜스포존을 제조한다. 스트랩타비딘 비드를 혼합물에 첨가한다. HMW 게놈 DNA를 첨가하고 나서, 태그화 반응을 수행한다. 반응 산물을 0.1% SDS로 처리하고 나서, PCR 증폭시켰다. 증폭 산물을 서열분석하였다.
- [0201] 섬 내 판독 수를 섬 크기에 대해 플롯팅한다. 도 36에 나타난 결과는 용액 기반 방법과 유사하게 원포트 클론 지표 비드를 이용하여 섬(근위 판독)을 관찰한다는 것을 나타낸다. 트랜스포존 형성 전에 지표화된 트랜스포존을 혼합했을 때, 섬(근위 판독)은 관찰되지 않았다. 트랜스포존 형성 전에 트랜스포존의 혼합은 비드 당 상이한 지표/트랜스포존을 지니는 비드(즉, 클론이 아님)를 제공한다.
- [0202] 실시예 15— CPT-seq에 의한 구조적 변이체 분석
- [0203] 60kb 이형 접합적 결실의 검출
- [0204] fastq 파일로서 서열분석 데이터를 추출하고, 각각의 바코드에 대한 개개 fastq 파일을 생성하기 위해 역다중화 과정을 조사한다. CPT 서열분석으로부터의 fastq 파일을 그들의 지표에 따라 역다중화시키고, 중복을 제거한 기준 게놈에 대해 정렬한다. 염색체를 5kb/1kb 창에 의해 스캔하고, 이때 스캔 창 내의 임의의 판독을 나타내는 지표 수를 기록한다. 이형 접합적 결실 영역에 대해 통계학적으로 DNA의 절반의 양만이 그의 이웃하는 영역에 비해서 라이브러리 생성을 위해 이용 가능하며, 따라서, 지표 수는 그의 이웃으로서도 거의 절반이어야 한다. 9216 지표화된 CPT 서열분석 데이터로부터의 5kb 창을 스캔함으로써 NA12878 chr1 60kb 이형 접합적 결실을 도 47a 및 도 47b에 나타낸다.
- [0205] 유전자 융합의 결실
- [0206] CPT 서열분석으로부터의 fastq 파일을 그들의 지표에 따라 역다중화시키고, 중복이 제거된 기준 게놈에 대해 정렬시킨다. 염색체를 2kb 창에서 스캔한다. 각각의 2kb 창은 36864 벡터인데, 이때 각각의 요소는 독특한 지표로부터의 다수 판독이 이 2kb 창에서 발견된 방법을 기록한다. 게놈에 걸친 모든 2kb 창 쌍(X,Y)에 대해, 가중치가 부여된-자카드(Jaccard) 지표를 계산한다. 이 지표는 샘플 내의 (X,Y) 사이의 사실상의 거리를 나타낸다. 해당 지표를 도 48에서 나타난 히트맵으로서 나타내며, 각각의 데이터 지점은 2kb 스캔 창의 쌍을 나타내고; 상부 좌측 정사각형은 영역1로부터 X,Y 둘 다에 대한 것이며, 하부 우측은 영역2로부터 X,Y 둘 다에 대한 것이고, 상부 우측은 영역2를 가로질러 영역1로부터의 X,Y에 대한 것이다. 유전자 융합 신호는 이 경우에 중간에서 수평선으로서 나타낸다.
- [0207] 결실 검출
- [0208] CPT 서열분석으로부터의 fastq 파일을 그들의 지표에 따라 역다중화시키고, 중복을 제거한 기준 게놈에 대해 정렬시킨다. 염색체를 1kb 창에서 스캔한다. 도 49는 유전자 결실의 검출 결과를 나타낸다.
- [0209] 실시예 16— 페이지징 및 메틸화 검출
- [0210] 중아황산염 전환 효율 최적화
- [0211] 비드 상에서 지표 연결 CPT-Seq 라이브러리에 대해 ME(모자이크 요소 영역) 및 gDNA 영역에서 전환을 평가하였다. 프로메가사의 메틸 에지 중아황산염 전환 시스템을 최적화시켜 효율을 개선시켰다.



조건	DNA	비드	BSC 처리
1	10ng	없음	60°C에서 1시간 / 0.3M NaOH
2		있음	60°C에서 1시간 / 0.3M NaOH
3			60°C에서 1시간 / 1M NaOH
4			65°C에서 1시간 / 0.3M NaOH

[0212]

[0213] ME 서열을 분석하여 중아황산염 전환 처리 효율을 결정하고, 도 50에 나타내었다. 비드에 부착된 지표-연결 라이브러리의 95% 중아황산염 전환(BSC). 중아황산염 조건 > 혹독한 중아황산염 처리 사이에서 관찰된 유사한 PCR 수율은 라이브러리를 분해하는 것으로 나타나지 않았고, 도 51에 나타내었다. 비드 상에서 지표 연결된 라이브러리의 대략 95% BSC를 관찰하였다. BSC를 개선시키는 것으로 조사된 변수(C→U)는 온도 및 NaOH 농도(변성)였다. 60°C 및 1M NaOH 또는 °C 및 0.3M NaOH를 제대로 수행하였다.

[0214] 비드 라이브러리 상에서 서열분석 BSC가 CPT-seq로 전환된 후의 예상 서열분석 판독 구조를 관찰하였다. 염기 기질%를 도 52에서 IVC 플롯에 의해 나타내었다.

[0215] 도 53은 중아황산염 전환 후의 PCR 후에 지표-연결된 라이브러리의 아가로스 겔 전기이동 이미지를 나타낸다. 200 내지 500bp 라이브러리의 예상 크기 범위를 관찰하였다. DNA가 없는 반응은 지표-연결 라이브러리를 수득하지 않는다.

[0216] 실시예 17— 표적화 페이지

[0217] 전체 게놈 지표화된 연결 CPT-seq 라이브러리를 강화시켰다. 도 54는 전체 게놈의 생물분석기 흔적이 크기 선택 없이 강화 전에 CPT-seq 라이브러리를 지표화 연결하였다. 도 55는 강화 후 라이브러리의 아가로스 겔 분석을 나타낸다.

[0218] HLA 영역에 대한 강화 통계학을 이하에 나타낸다:

샘플 ID:	C3
샘플명:	HLA 프로브
패딩 크기:	150
표적화된 기준의 총 길이:	5062748
총 PF 판독 :	2516
Q30% :	94.90%
총 정렬 판독:	2498
정렬 판독%:	99.40%
표적화된 정렬 판독:	840
판독 강화:	30.80%
중복 쌍 판독%:	12.70%
단편 길이 중앙값:	195

[0219]

[0220] 도 56은 염색체에서 HLA 영역에 대한 표적화된 단상형 적용의 결과를 나타낸다. 좌측에 전체 게놈 지표 연결된 판독 라이브러리 강화의 도시. 각각의 작은 막대는 지표화된 짧은 라이브러리를 나타낸다. 게놈 척도에 대한 판독의 근접성("섬" 특징) 때문에, 지표화된 라이브러리의 클러스터는 단일 비드 상에서 클론에 의해 지표화된 영

역인 "섬"이다. 표적화 영역에서 라이브러리의 강화(WO 2012108864 A1 핵산의 선택적 강화 참조)를 우측에 나타낸다. 판독을 HLA 영역에 대해 강화한다. 추가적으로, 판독을 지표에 의해 분류하고, 계층에 대해 정렬할 때, 그들은 다시 "섬" 구조를 나타내는데, 이는 접근 정보가 지표화 연결 판독으로부터 유지된다는 것을 나타낸다.

[0221] 실시예 18— 지표 교환

[0222] 트랜스포좀 복합체의 모자이크 말단(ME) 교환을 평가하기 위해, 상이한 지표를 지니는 비드를 준비하였다. 혼합 후에, 라이브러리를 서열분석함으로써 그리고 각각의 라이브러리에 대한 지표를 보고함으로써 지표 교환을 결정하였다. "스와핑" %를  $(D4+D5+E3+E5+f4)/(모두\ 96의\ 합)$ 으로서 계산하고 나서, 도 65에 나타낸다.

[0223] 실시예 19— 스트렙타비딘 비드에 트랜스포좀 복합체를 더 밀집하게 모으는 것에 의한 라이브러리 삽입 크기의 감소

[0224] 스트렙타비딘 자기 비드를 1x, 6x 및 12x 농도의 TsTn5 트랜스포좀 복합체와 함께 부하하였다. 각각의 비드 유형에 대해 Epi-CPT seq 프로토콜을 수행하였다. 최종 PCR 산물을 분석을 위해 애질런트 바이오어날라이저(Agilent BioAnalyzer) 상에 부하하고 나서, 도면에 나타낸다. Epi-CPT seq 라이브러리 단편은 더 작고, 더 많은 TsTn5를 비드 상에 부하할 때 수율이 더 높다.

[0225] 실시예 20— 중황산나트륨 전환 동안 DNA 라이브러리의 단편화.

[0226] 중아황산염 전환 후에, DNA는 손상되었고, PCR 증폭에 필요한 공통 서열(CS2)의 상실을 초래한다. DNA 단편 CPTSeq 및 Epi-CPTSeq(Me-CPTSeq) 라이브러리를 생물분석기에 의해 분석하였다. 중아황산염 전환 동안의 DNA 손상에 기인하여, Epi-CPTSeq 라이브러리는 도 70에 나타낸 바와 같이 CPTSeq 라이브러리에 비해 5-배 더 낮은 수율 및 더 작은 라이브러리 크기 분포를 가진다.

[0227] 실시예 21—TdT 매개 ssDNA 결찰 반응

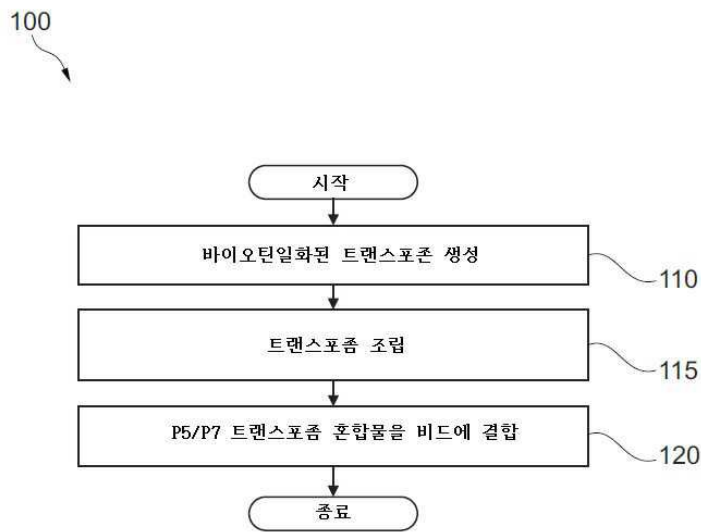
[0228] 말단 트랜스퍼라제(Terminal transferase: TdT) 매개 결찰에 의한 DNA 최종 회수의 실현 가능성을 시험하였다. 간략하게, 5 pmol의 ssDNA 주형을 TdT(10/50U)와 함께 인큐베이션시키고, 감쇠기/어댑터 이중가닥(0/15/25pmol) 및 DNA 리가제(0/10U)를 37°C에서 15분 동안 인큐베이션시켰다. 연장/결찰의 DNA 산물을 TBE-유레아 겔 상에서 분석하고 나서, 결과를 도 71에 나타낸다. 모든 반응 성분의 첨가는 어댑터 분자의 거의 완전한 결찰을 초래하였다(레인 5 내지 8).

[0229] 말단 트랜스퍼라제(TdT) 매개 결찰에 의한 DNA 최소 회수의 실현 가능성을 중황산나트륨 전환 비드 결합 라이브러리에 대해 시험하고 나서, 도 72에 나타낸다. 간략하게, DNA를 비드 상에서 태그화하고 나서(처음 2개 레인), 프로메가사의 메틸에지 중황산염 전환 키트(레인 3 및 4)로 처리하고, DNA 구조 프로토콜을 실시하였다(레인 5 및 6). 구조 반응 후에 DNA 라이브러리의 수율 및 크기의 분명한 증가가 있다. 또한 자가 삽입 트랜스포존(SI)의 존재비 증가가 있는데, 이는 어댑터 분자의 효율적인 결찰을 나타낸다.

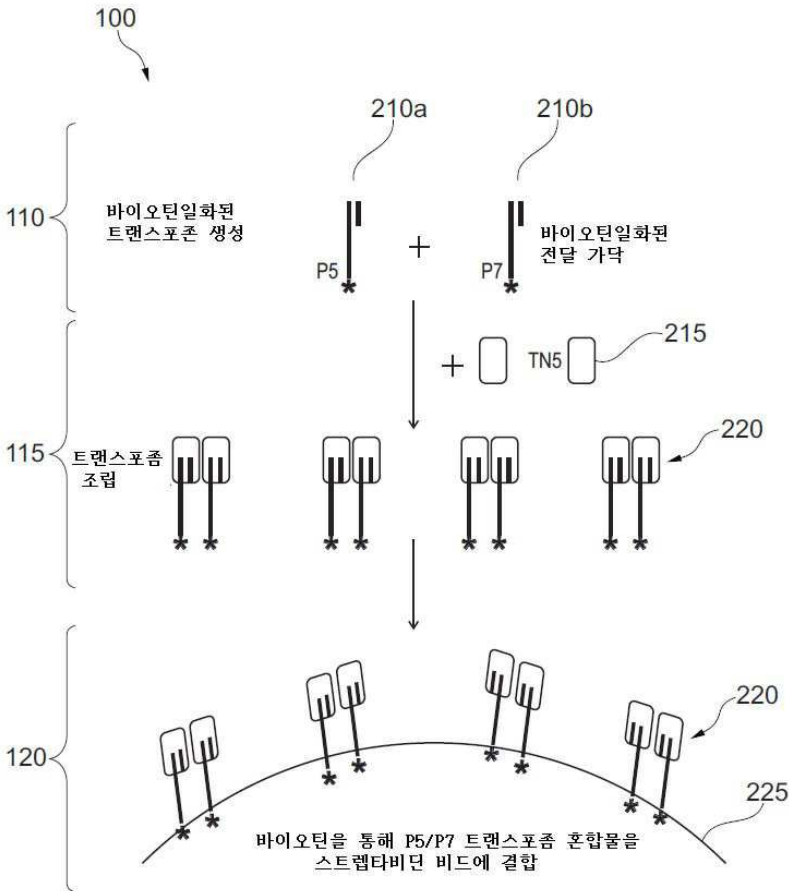
[0230] 메틸-CPTSeq 분석 결과를 도 73에 제시한다.

도면

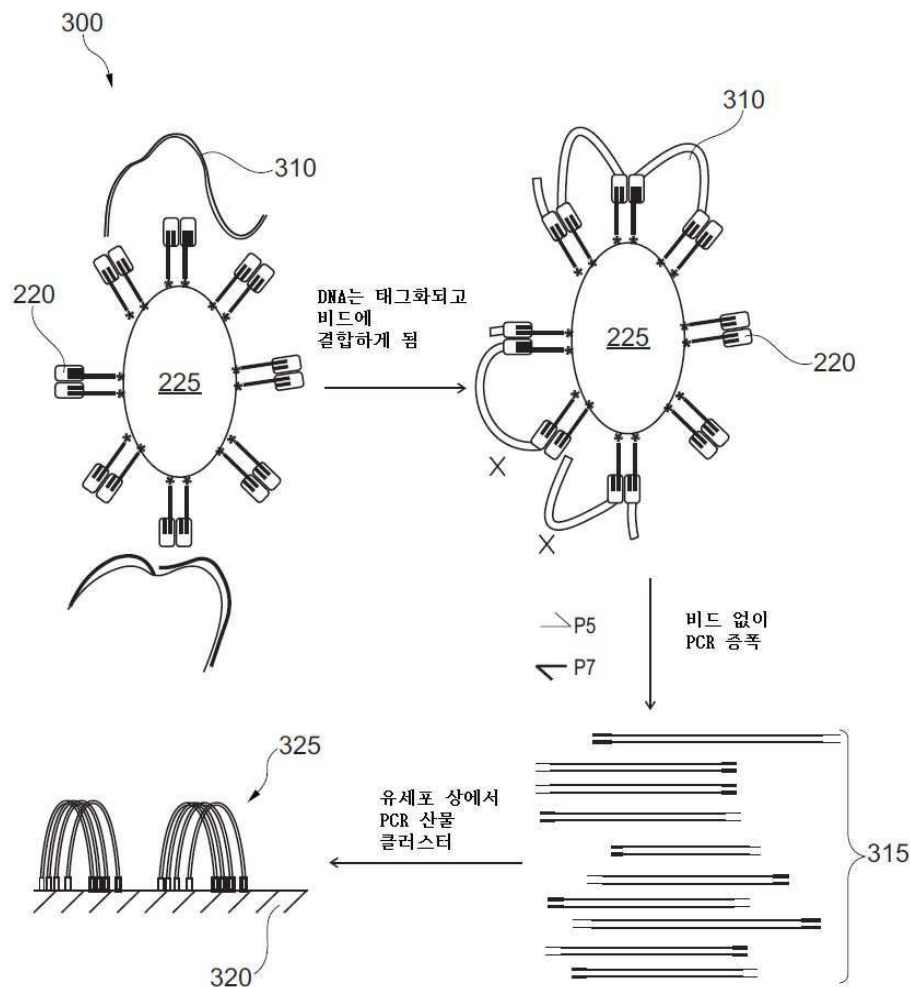
도면1



도면2



도면3



도면4

400

샘플	동일 용적의 PCR 산물을 지니는 클러스터의 번호							
클러스터에 사용한 pH	15		19		21		24	
	K/m m2	%	K/m m2	%	K/m m2	%	K/m m2	%
50 ng	466	100	541	100	618	100	668	100
250 ng	504	108	598	110	703	114	740	111
1000 ng	512	109	602	111	670	108	716	107
*50 ng 전체 반복부	377	81	455	84	516	83	555	83

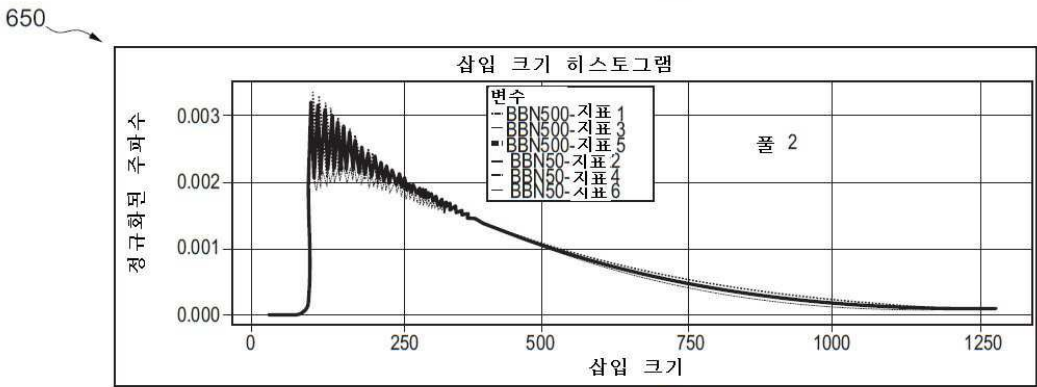
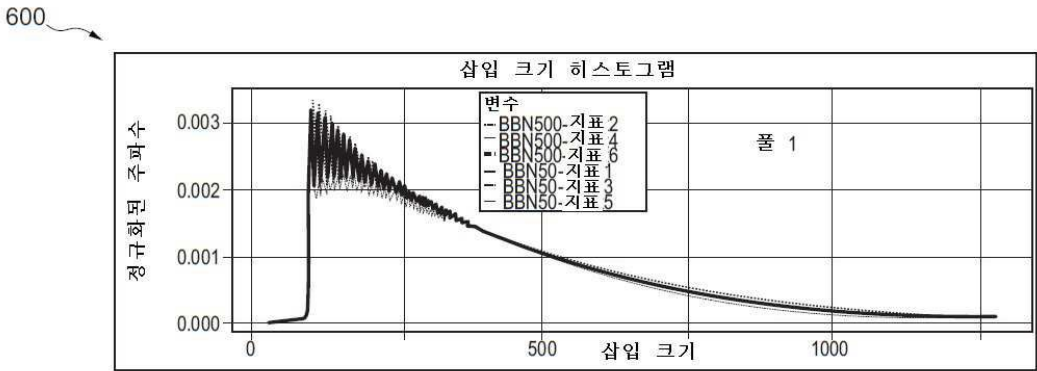


도면5

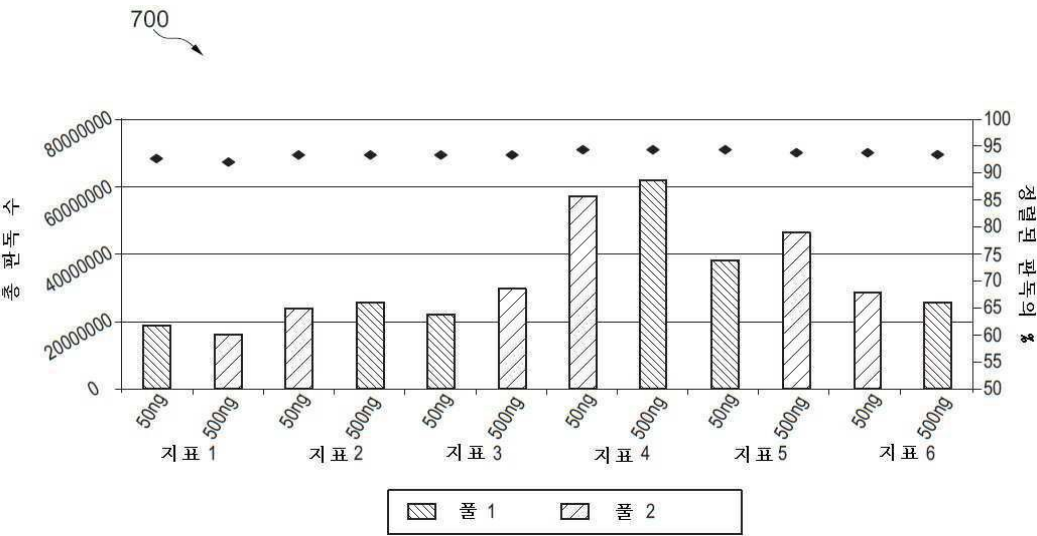
500

	중앙값 삼입 크기	평균 삼입 크기
BBN50-지 표1	341	409.78
BBN500-지 표1	344	415.89
BBN50-지 표2	337	405.97
BBN500-지 표2	346	419.91
BBN50-지 표3	330	400.35
BBN500-지 표3	327	398.65
BBN50-지 표4	321	388.71
BBN500-지 표4	328	400.75
BBN50-지 표5	327	397.62
BBN500-지 표5	328	399.88
BBN50-지 표6	324	389.35
BBN500-지 표6	331	401.50

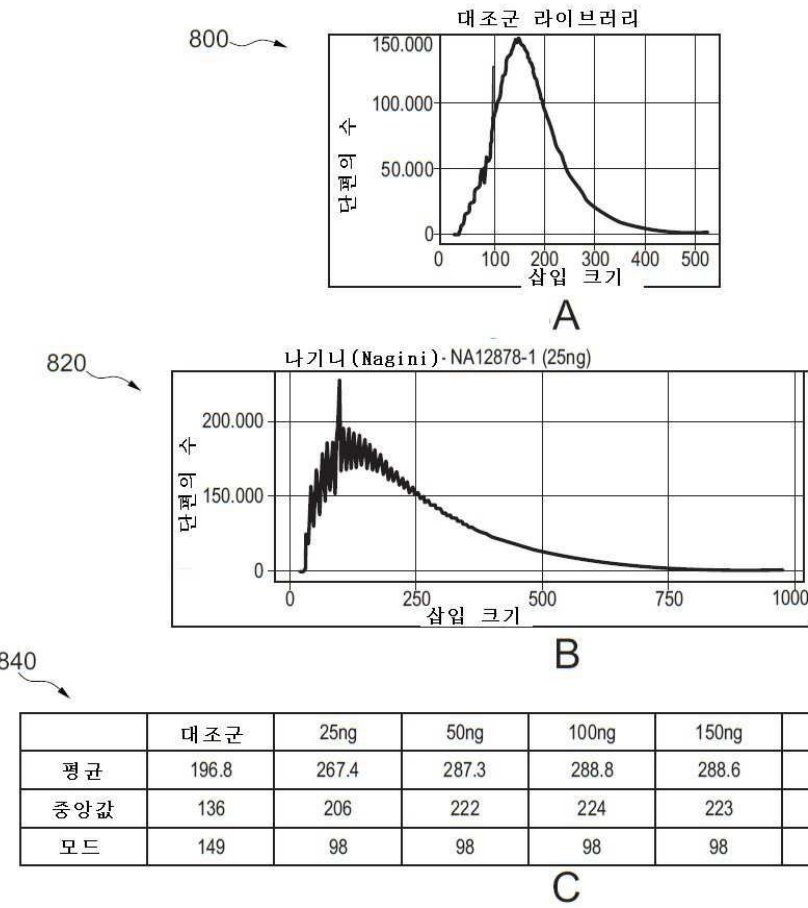
도면6



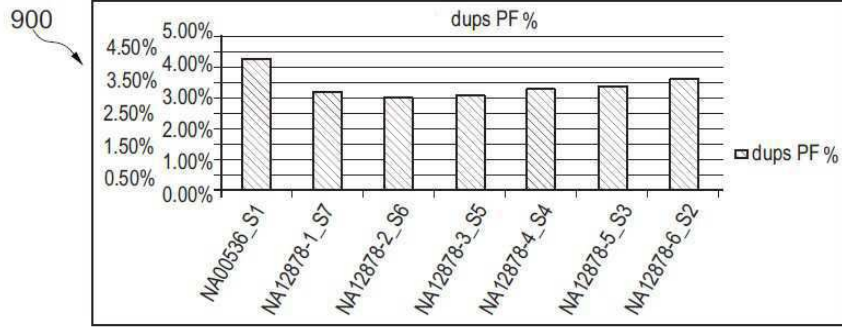
도면7



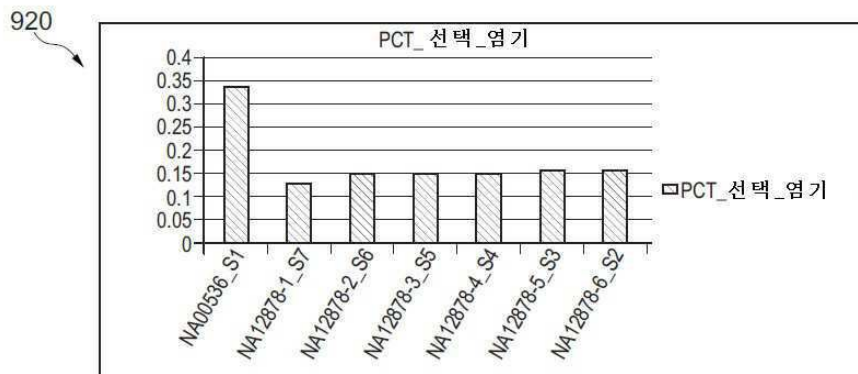
도면8



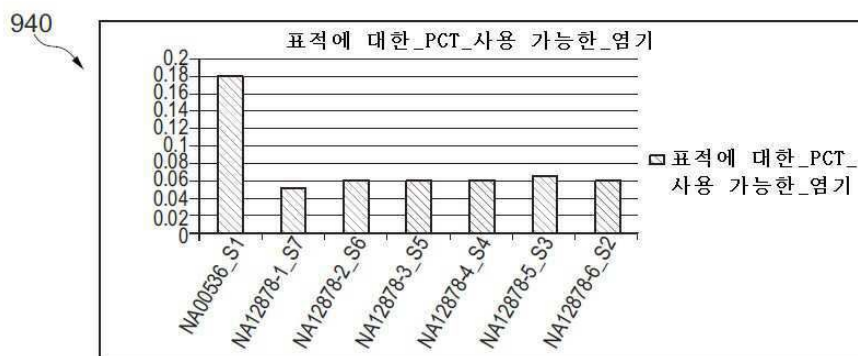
도면9



A

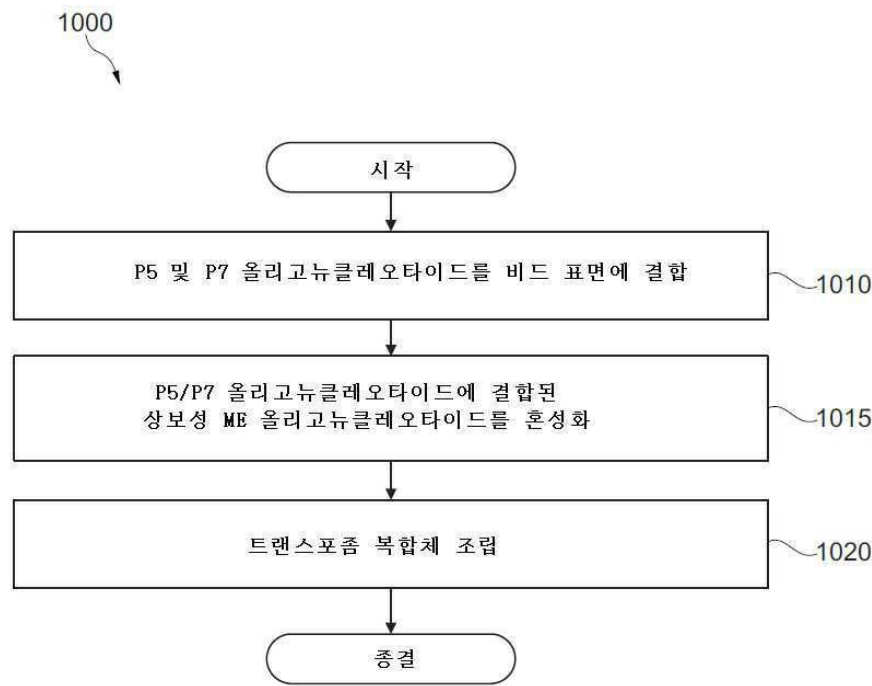


B

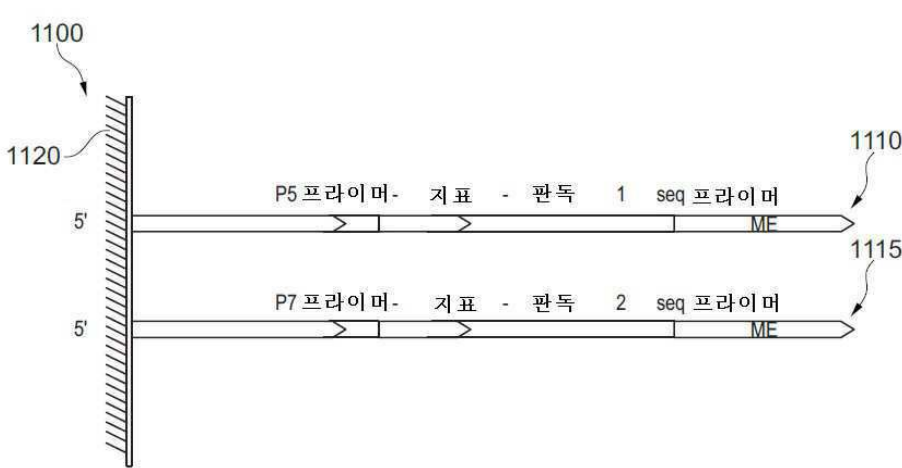


C

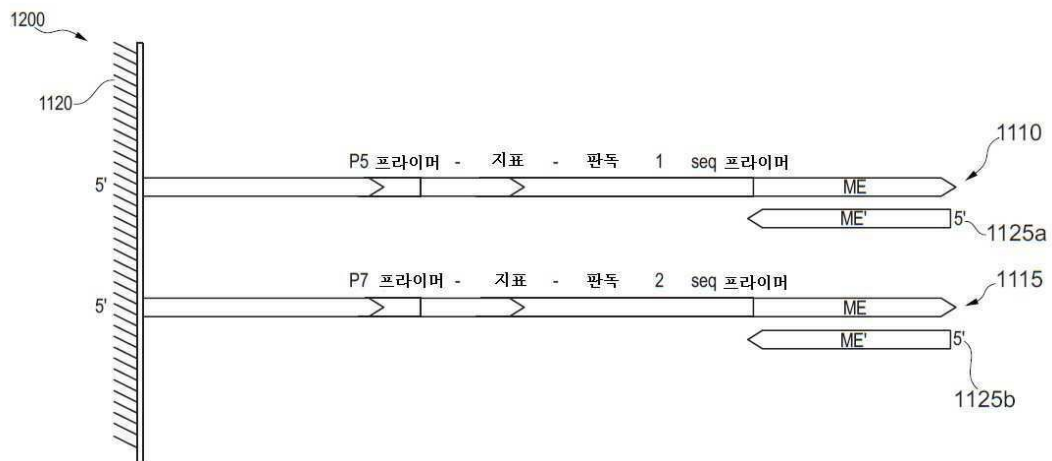
도면10



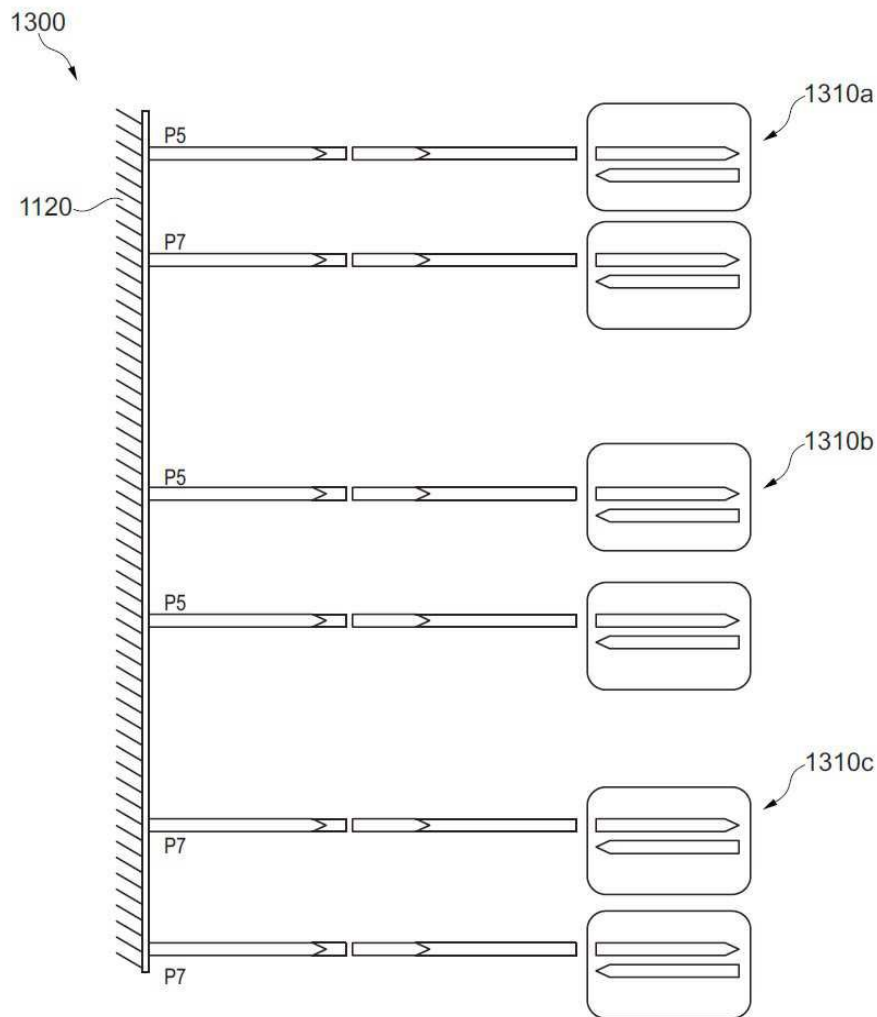
도면11



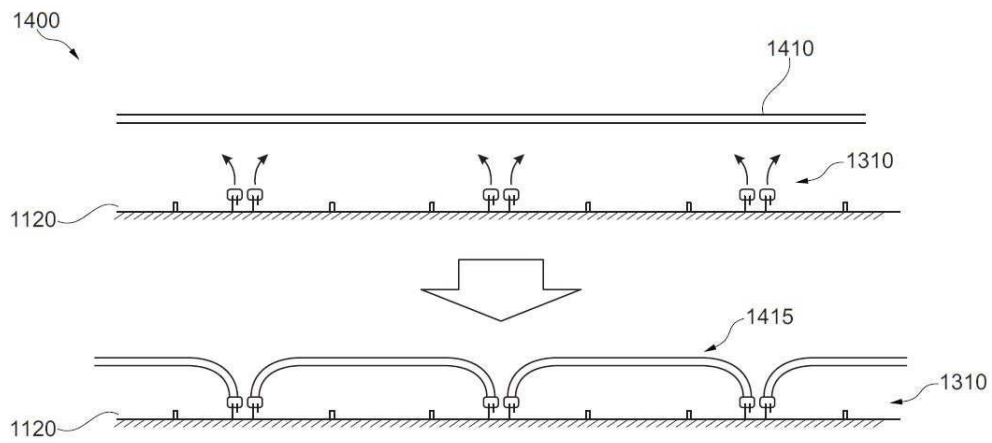
도면12



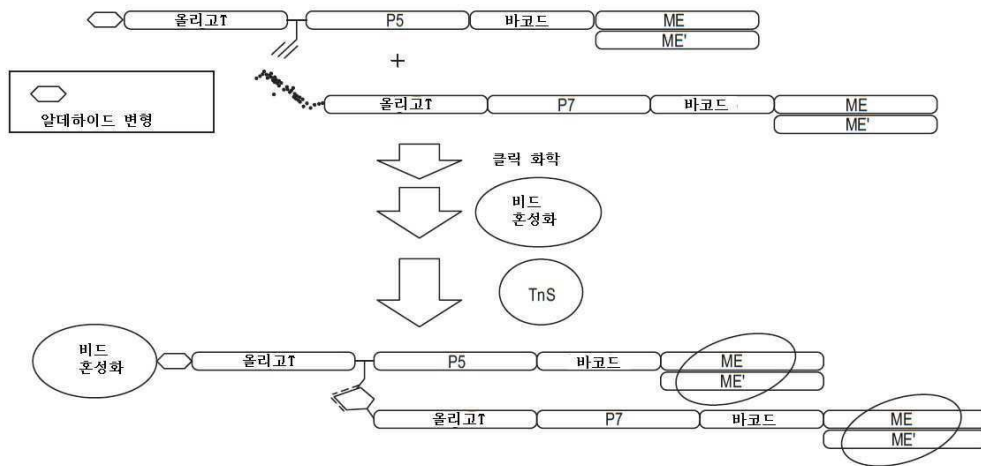
도면13



도면14

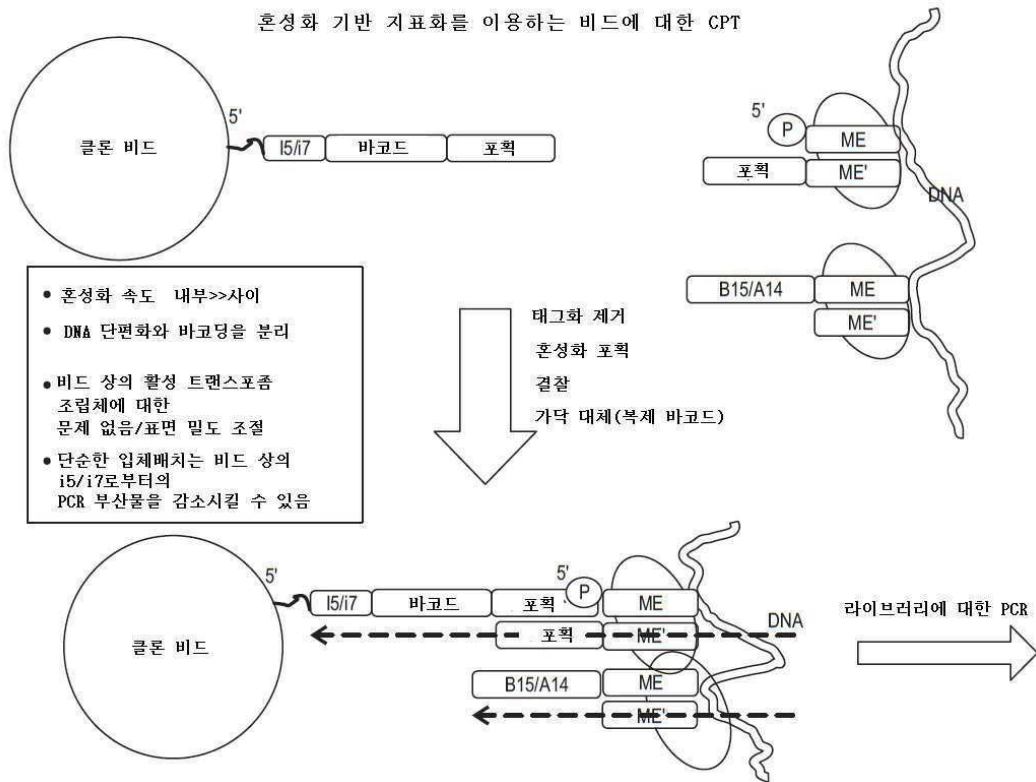


도면15

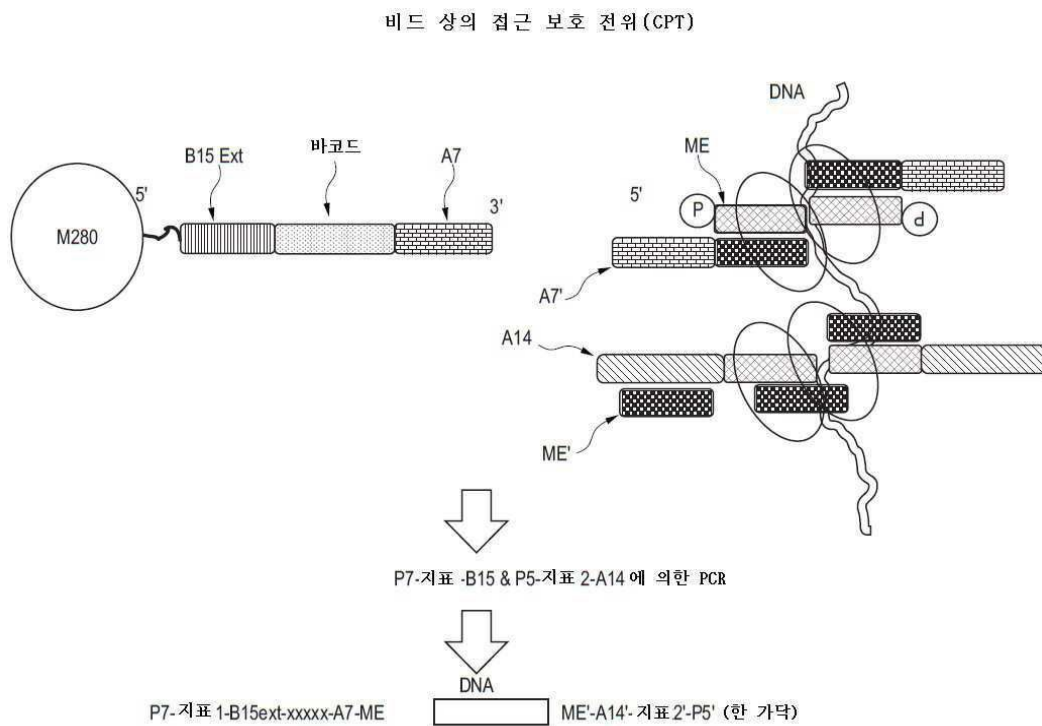




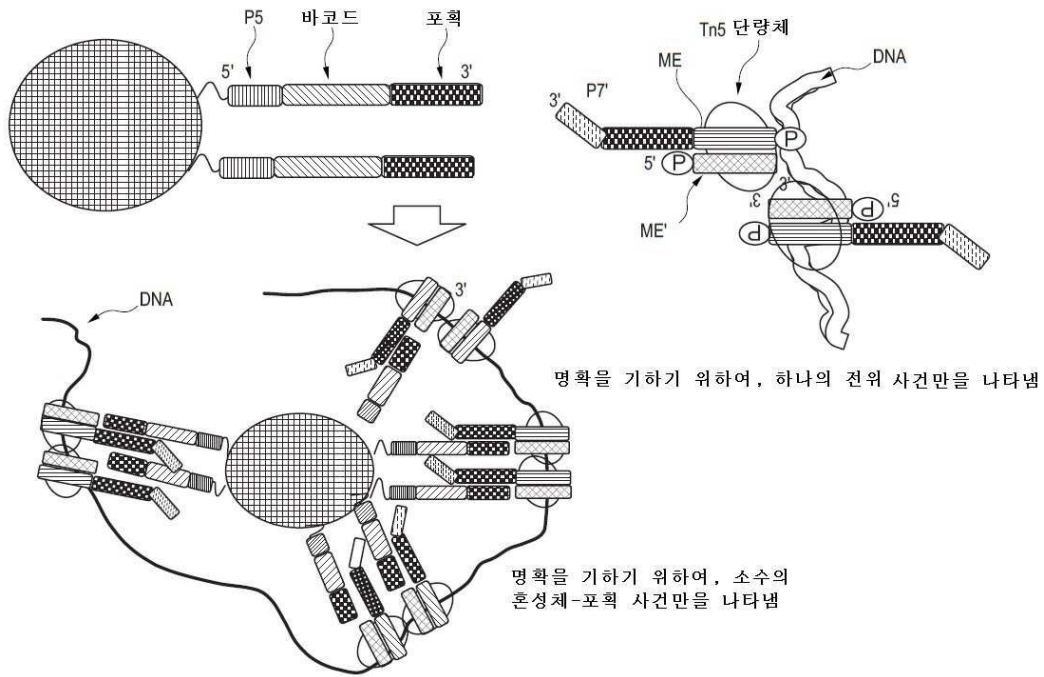
도면16



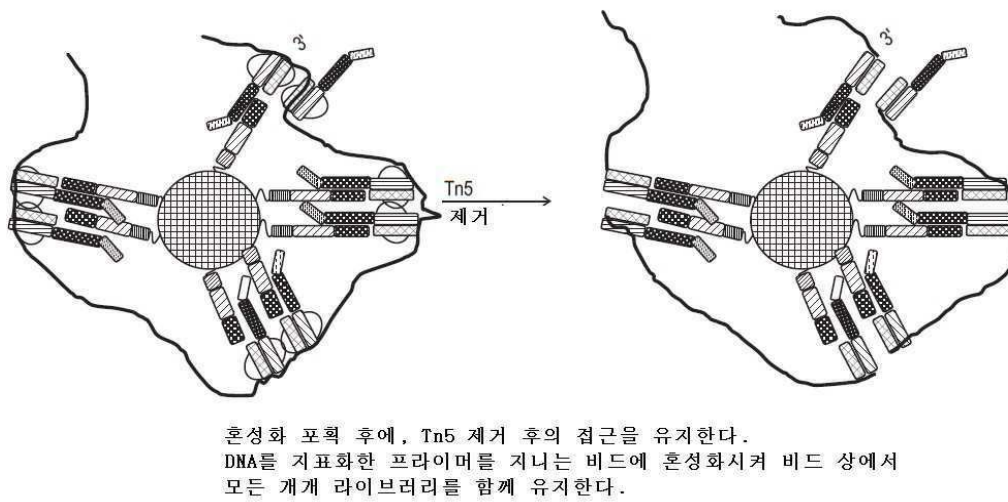
도면17



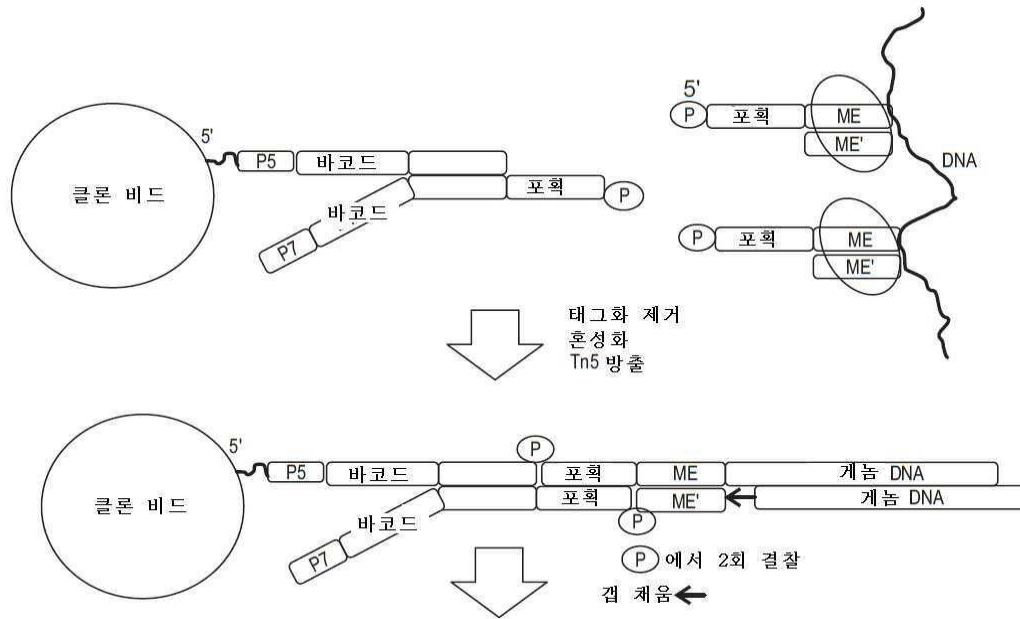
도면18



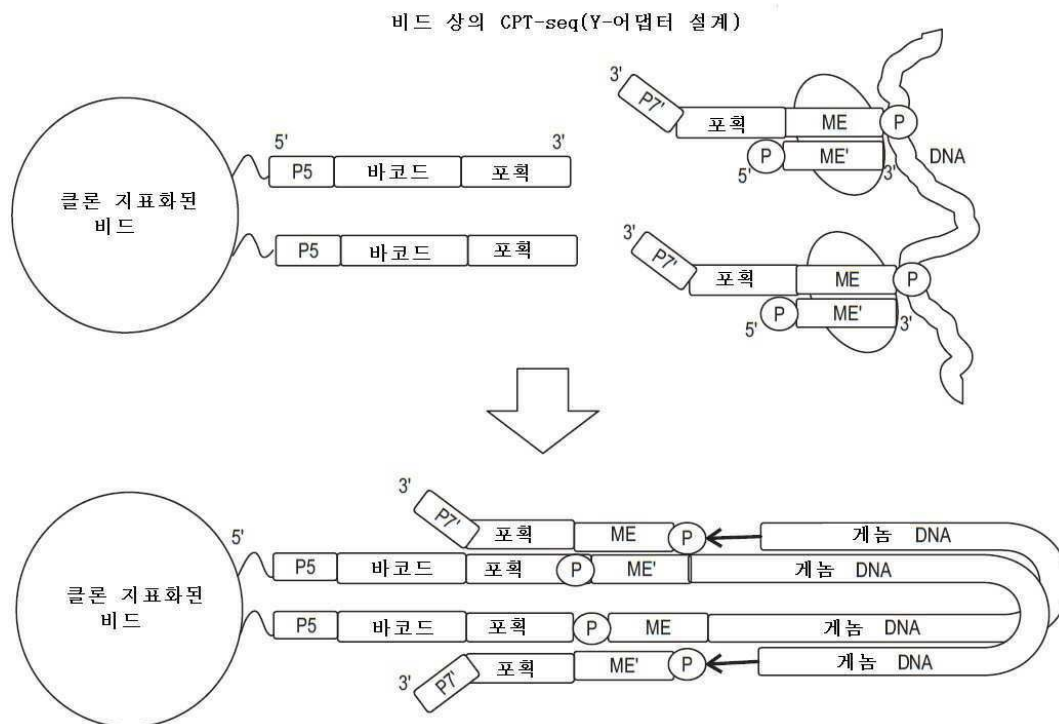
도면19



도면20

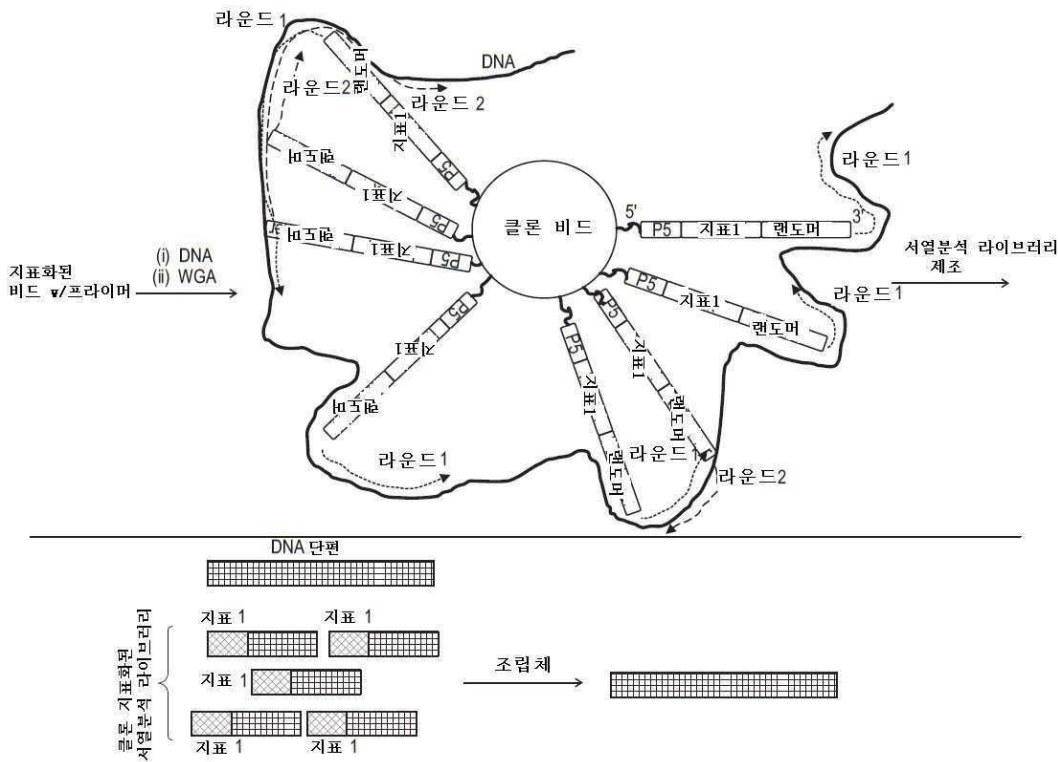


도면21



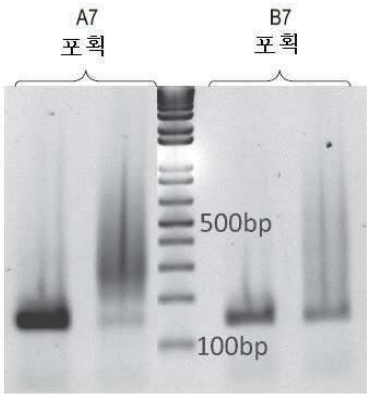


도면24



도면25

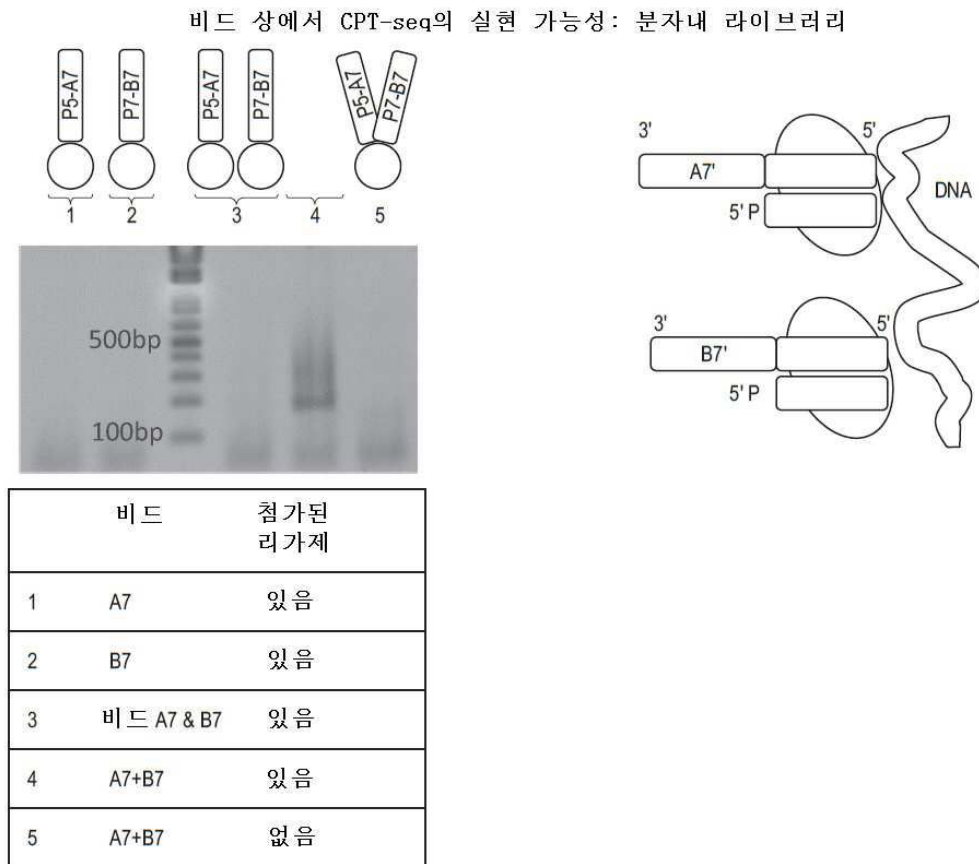
비드 상의 포획 밀도의 최적화



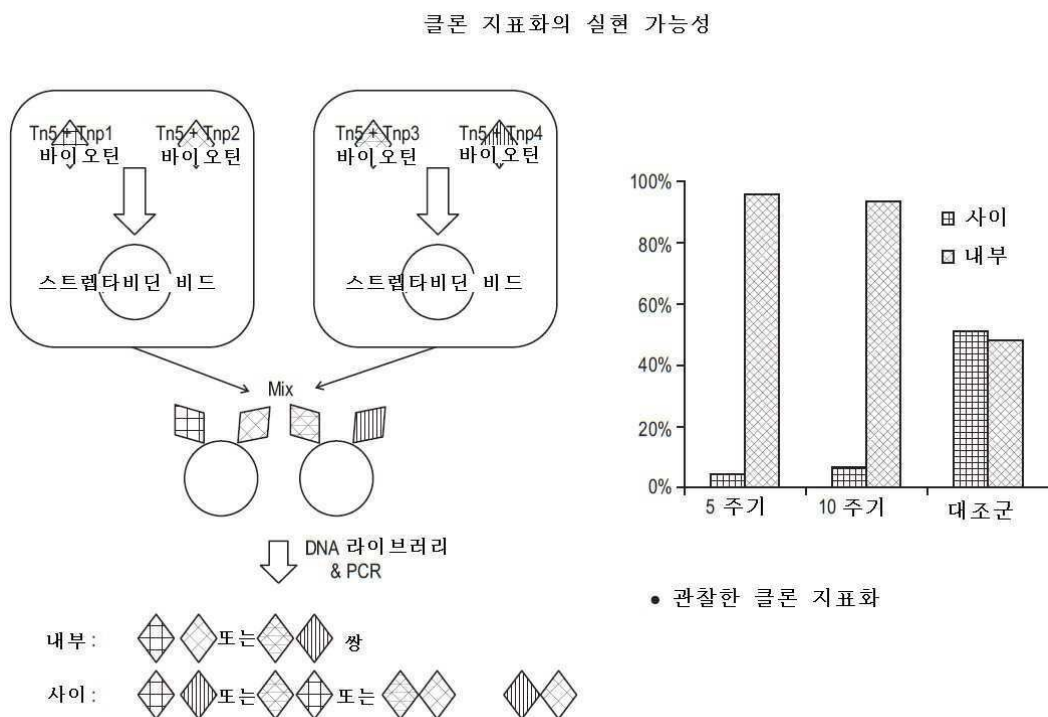
프로브 밀도 (회석)	결찰
1 1x	없음
2 10x	있음
3 1x	없음
4 10x	있음



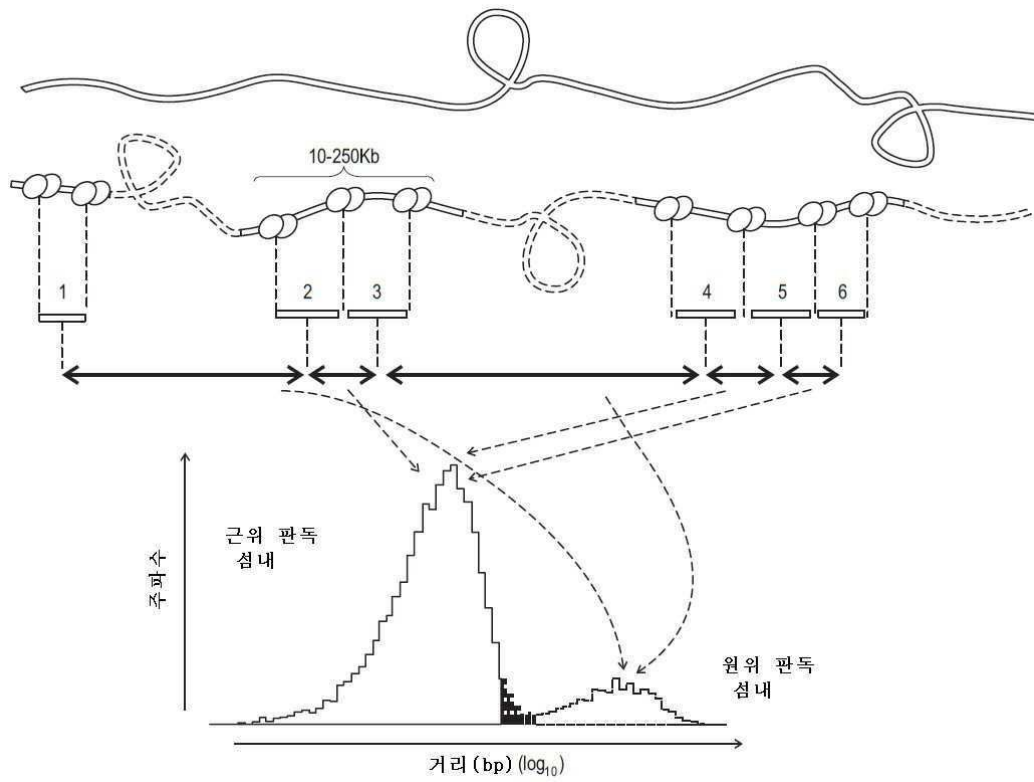
도면26



도면27

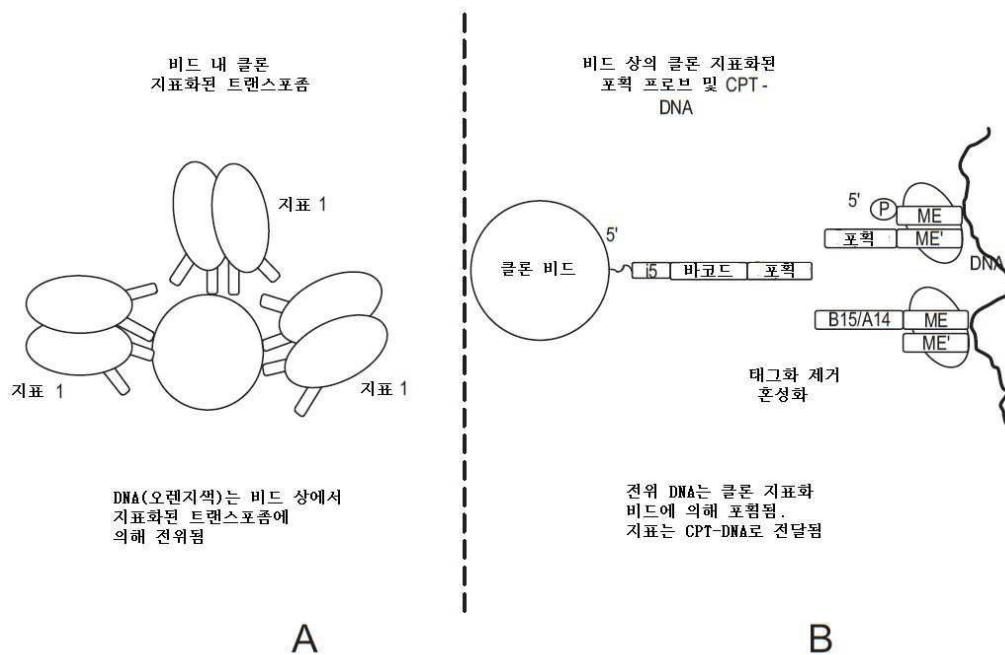


도면28



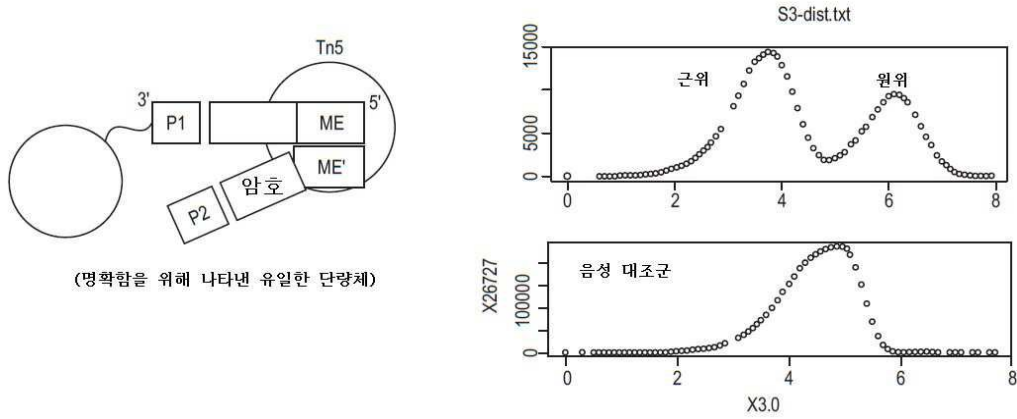
도면29

고체 지지체 상의 접근 정보를 유도하기 위한 접근



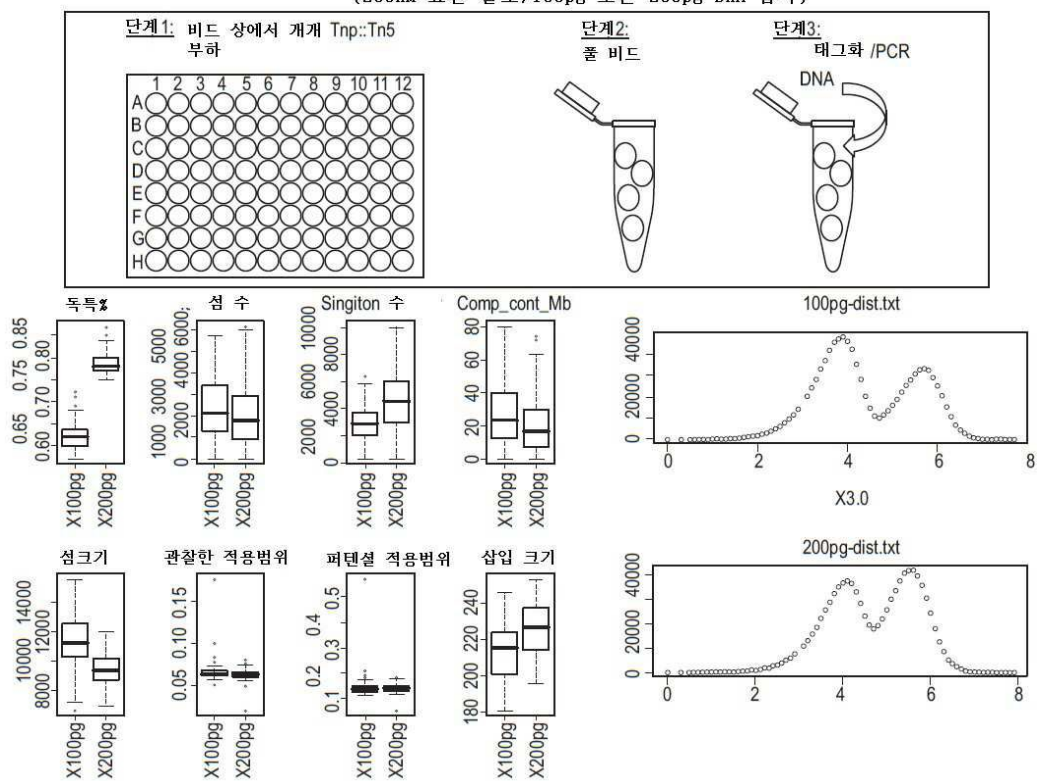
도면30

원포트 96 지표화된 클론 비드 전위

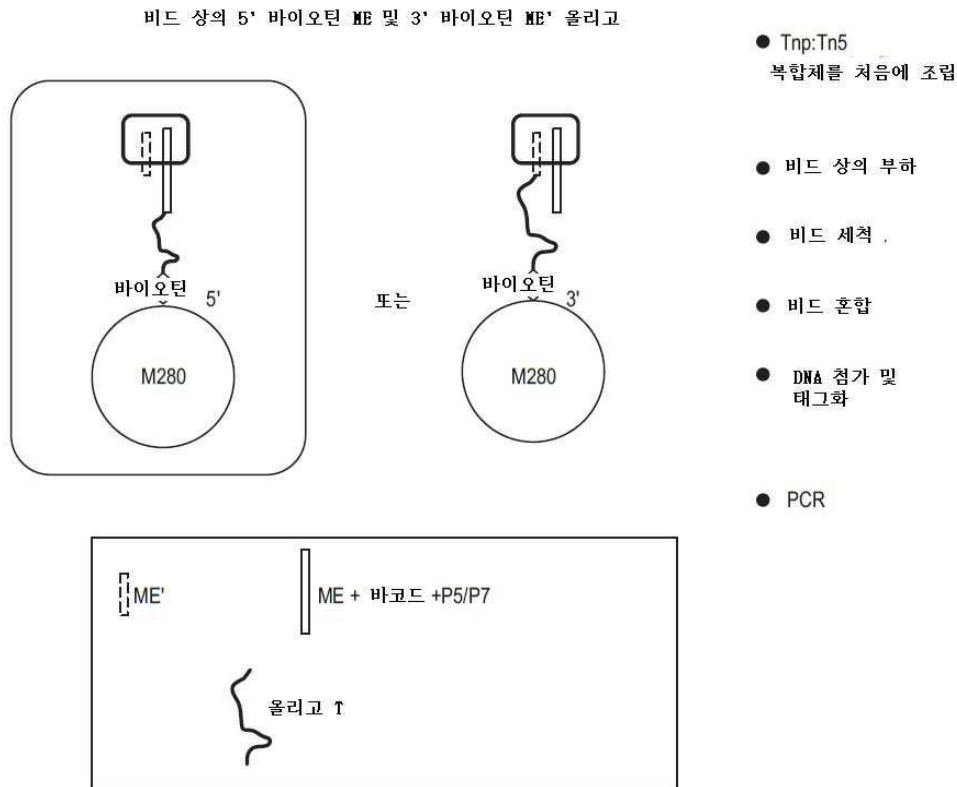


도면31

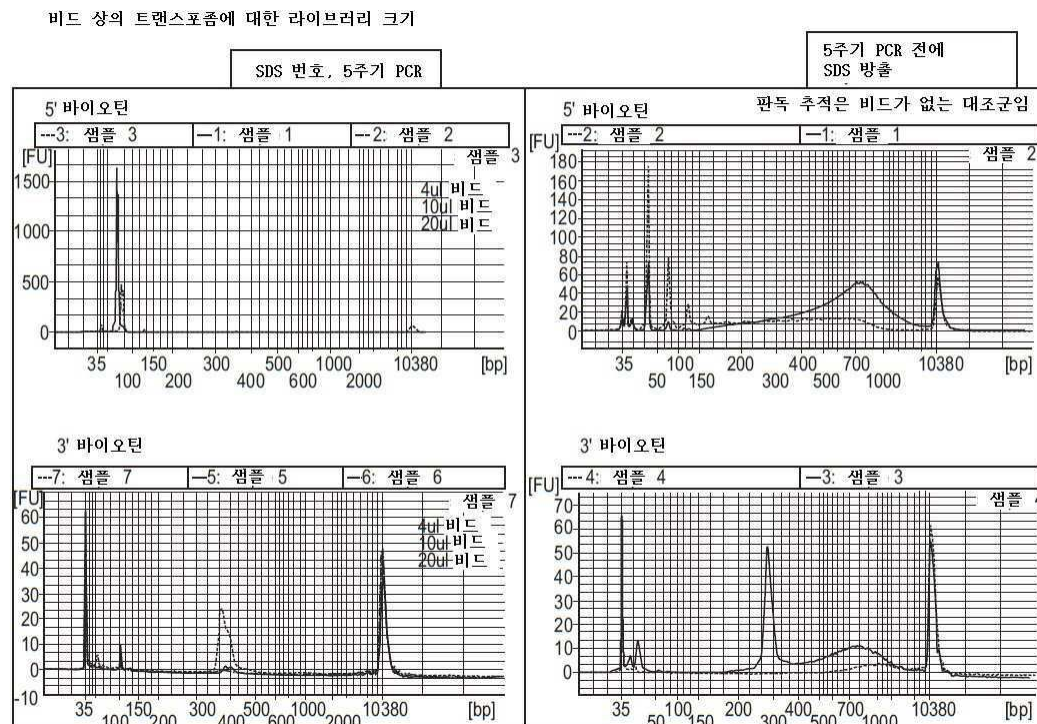
96플렉스 상의 비드 접근(비드 상의 트랜스포좀)  
(200nM 표면 밀도/100pg 또는 200pg DNA 입력)



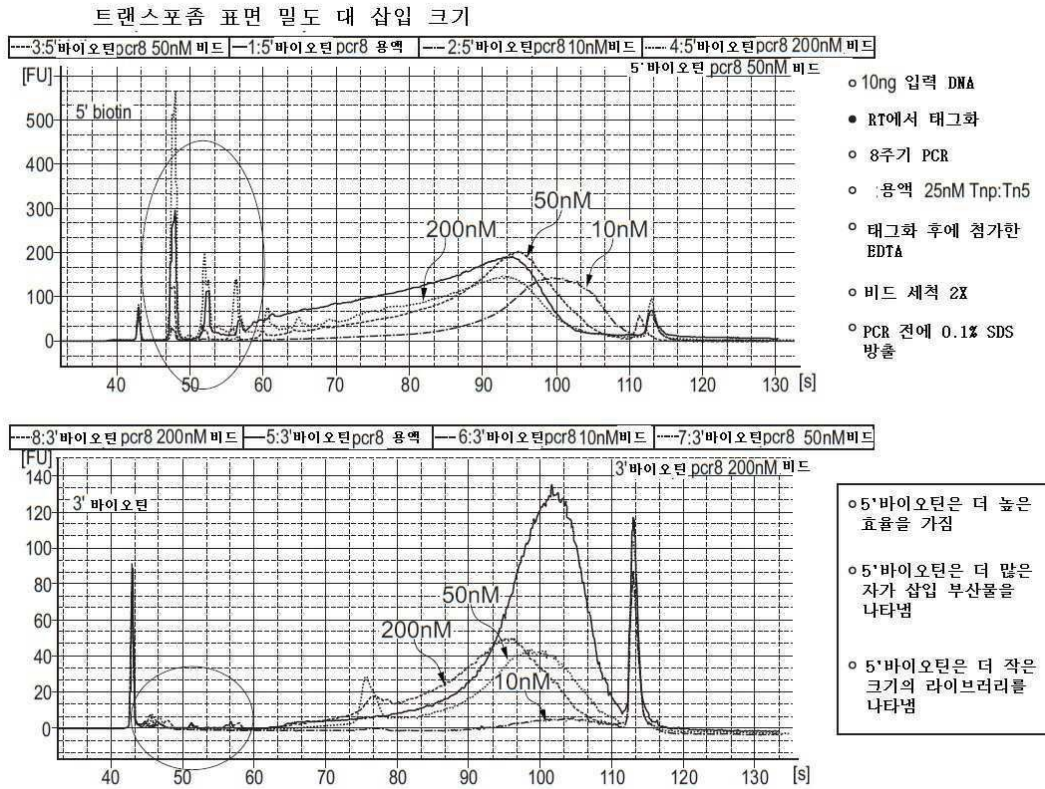
도면32



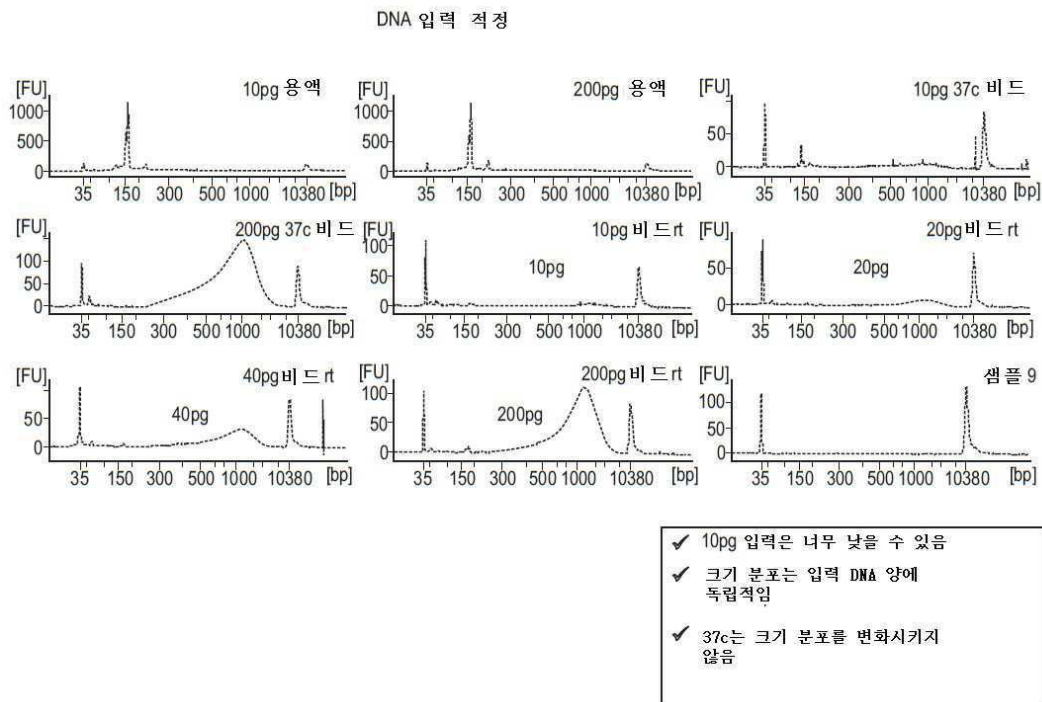
도면33



도면34



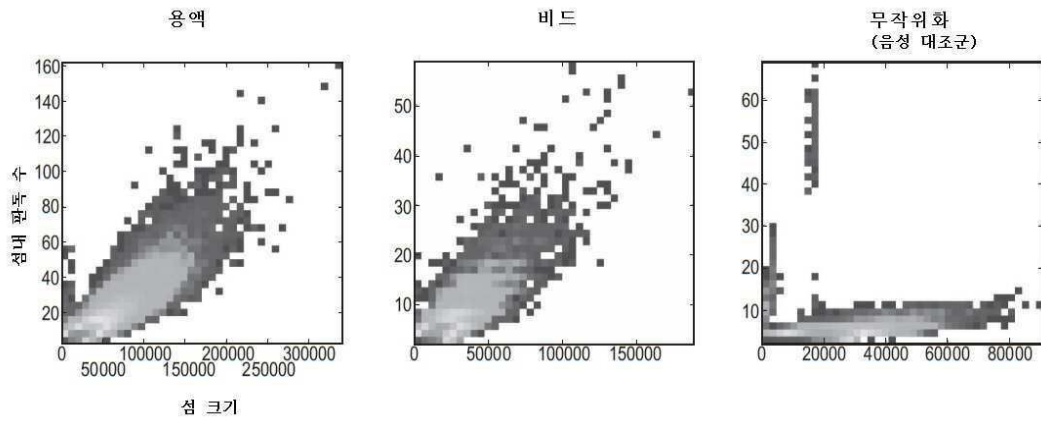
도면35





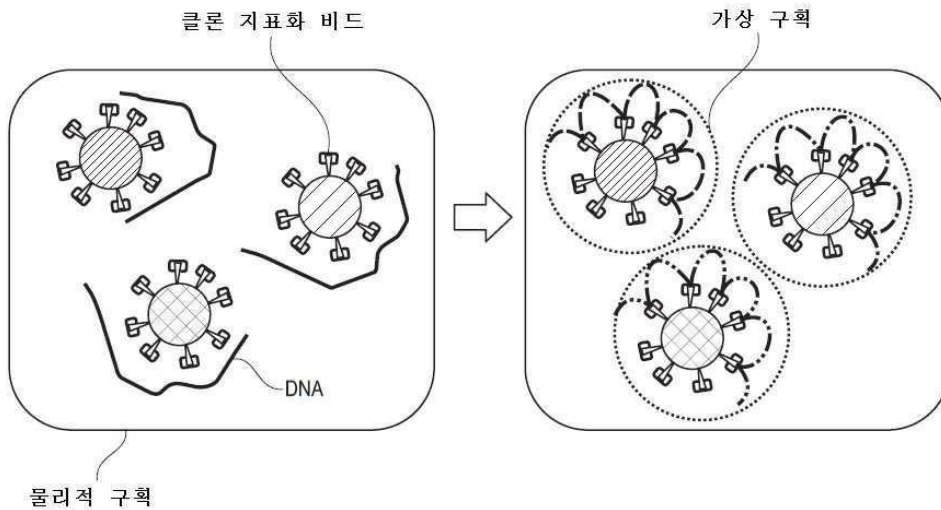
도면36

비드 지표화에 대한 원포트를 위한 점 크기 및 분포



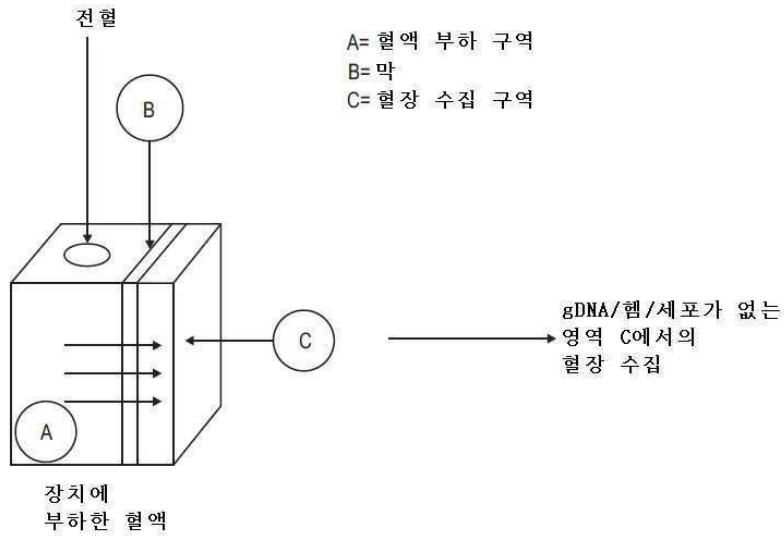
도면37

DNA 분자를 가로지르는 클론 지표화

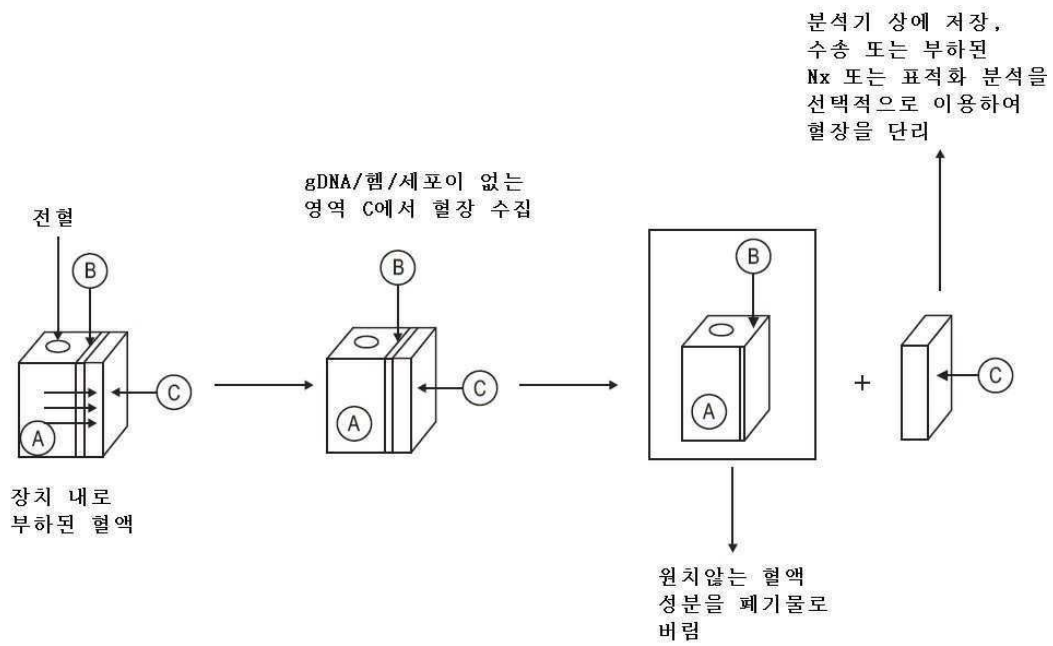


- ▶ 클론 비드에 근접한 DNA는 동일한 지표를 수용한다
- ▶ 분자내>>분자간 지표화
- ▶ 100kb 접근을 포획하기에 충분한 5um 비드 (C=15um, 100kb~34um)

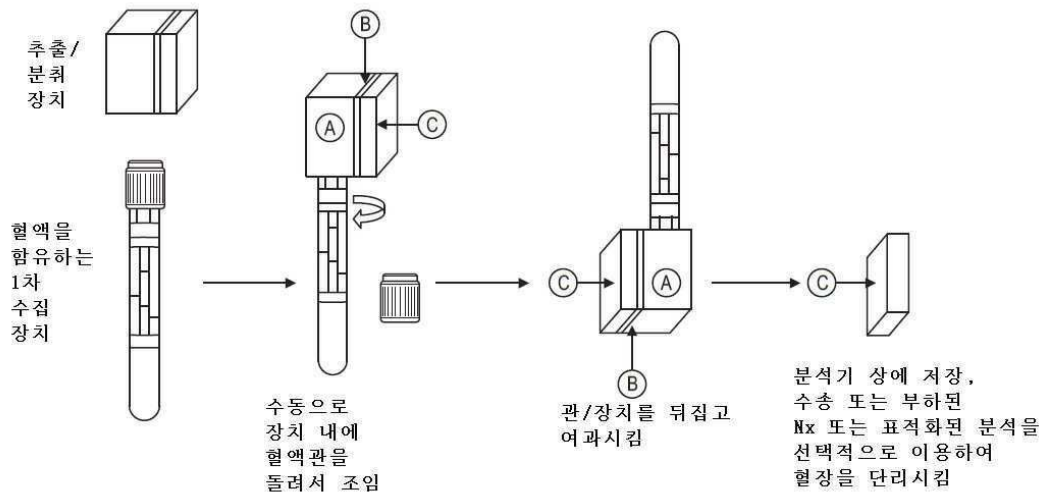
도면38



도면39



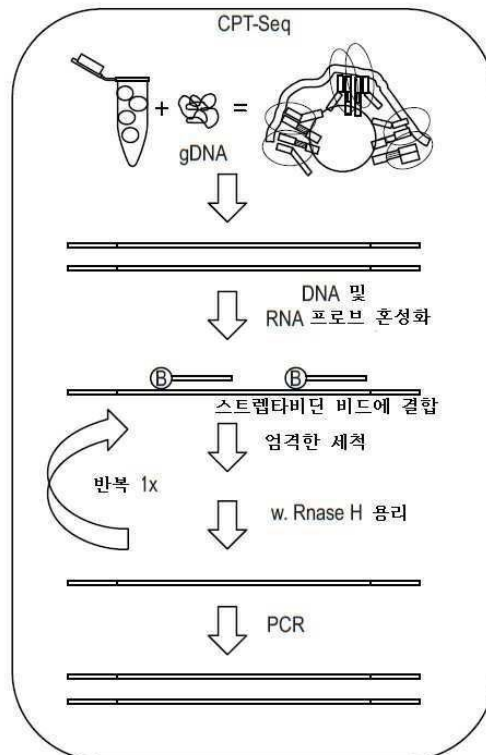
도면40



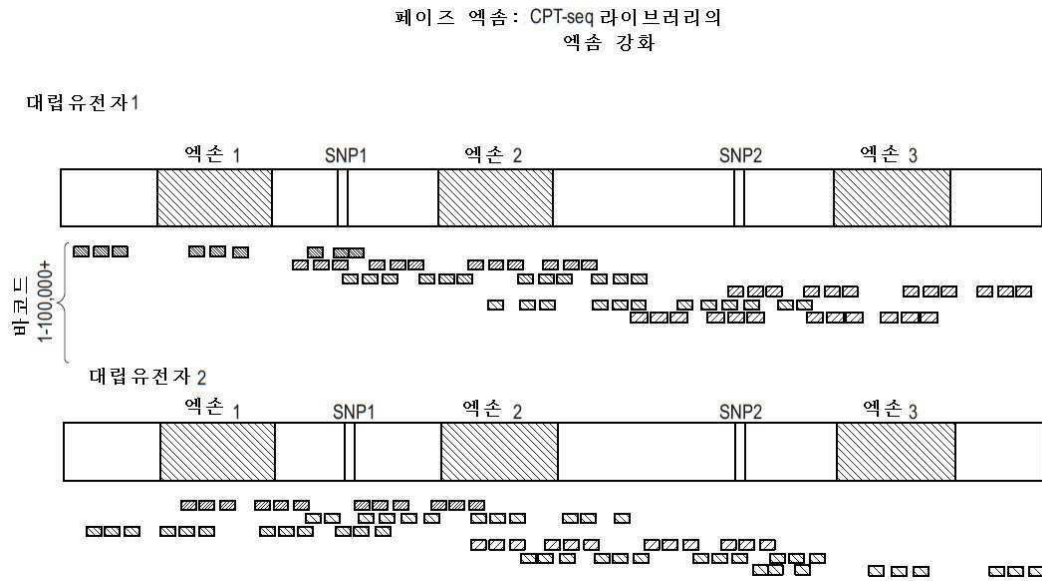
도면41

CPT-Seq 강화

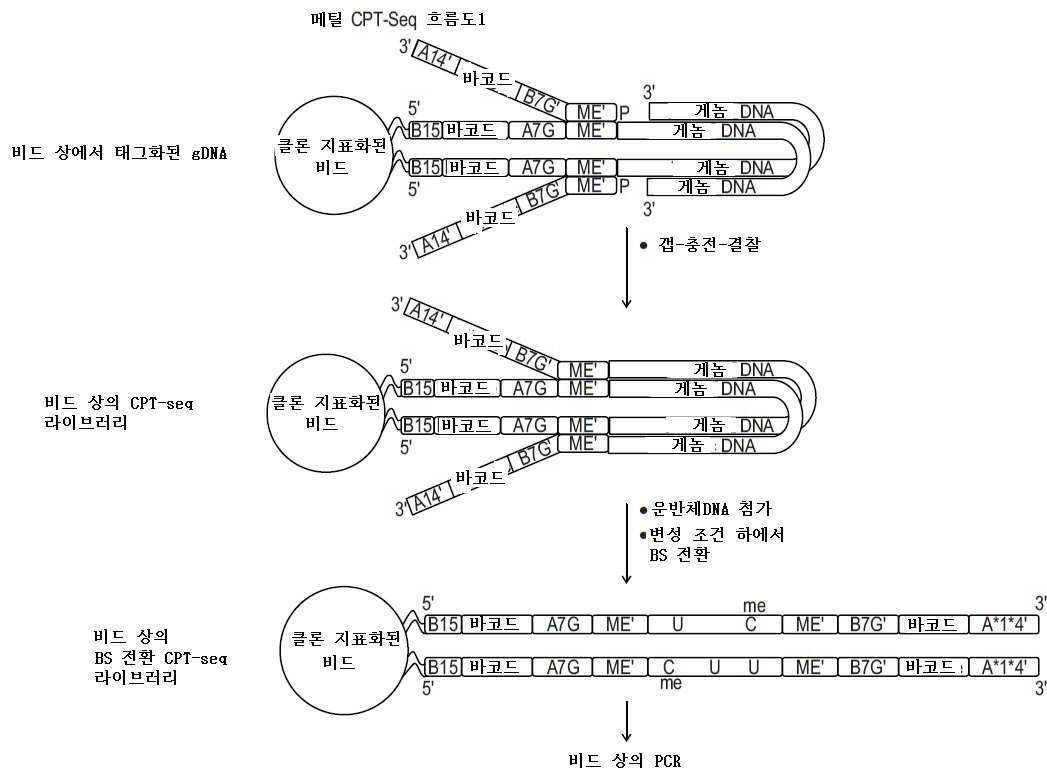
- 전체-게놈 지표화-연결 라이브러리의 표적화된 강화
- 샘플 지표화+양립 가능한 강화



도면42



도면43

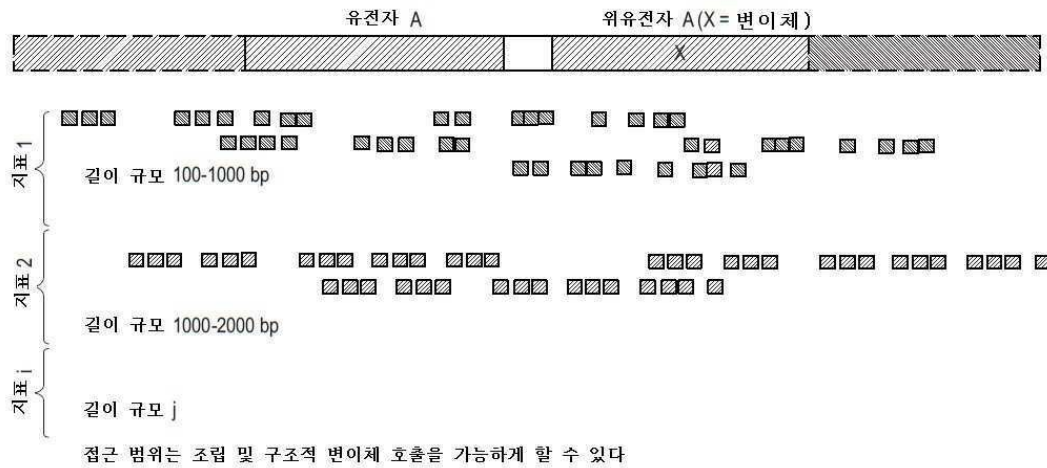






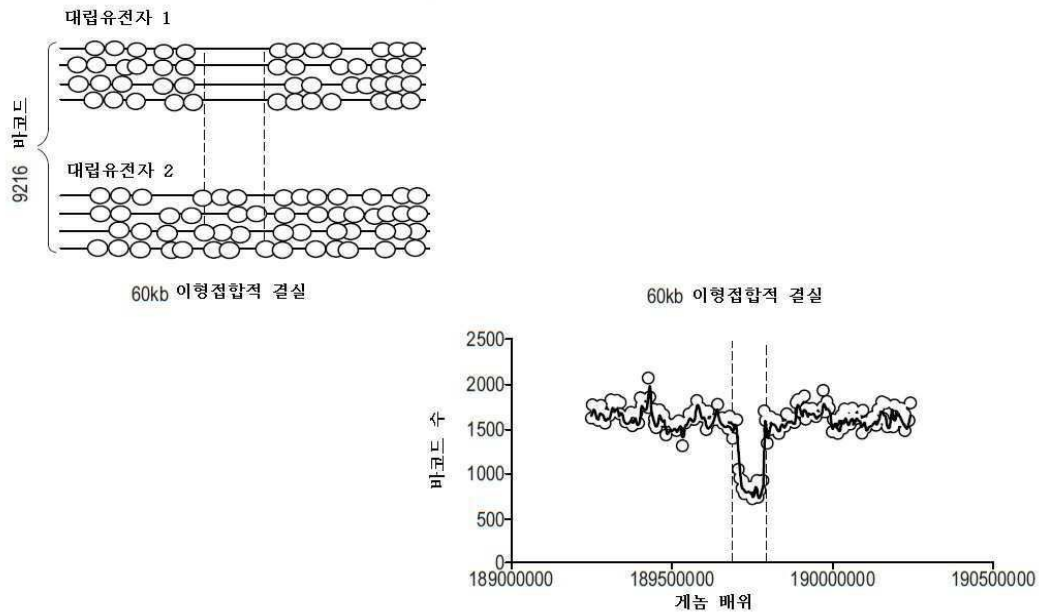
도면46

조립체/SV 특징  $\nabla$ /상이한 길이  
규모 라이브러리의 예



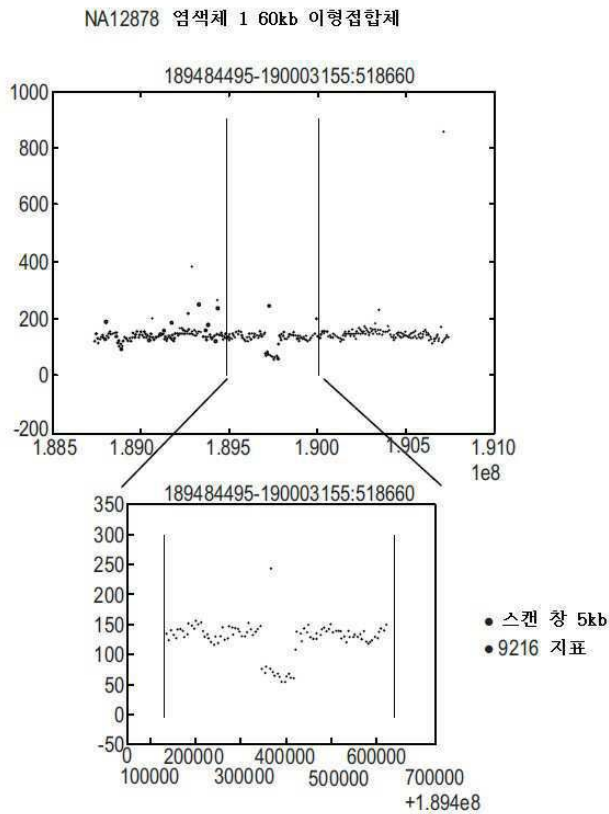
도면47a

CPT-seq를 이용하는 구조적 변이체 분석

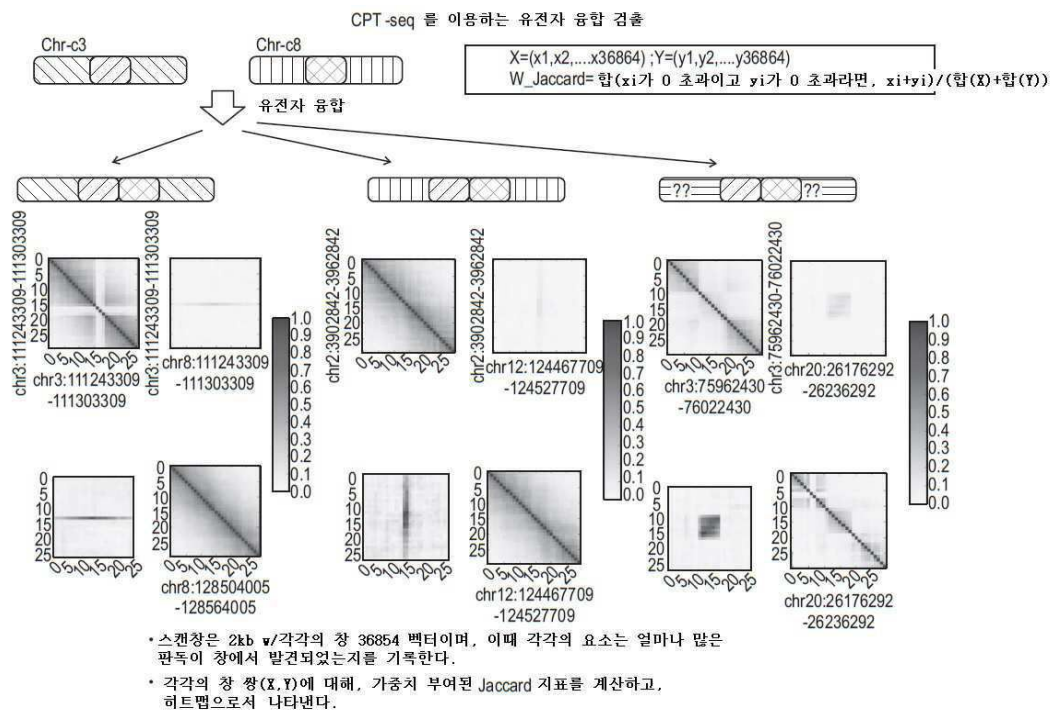


▶ CPT-seq를 이용하는 CHR1 결실에 대한 이형접합적 60kb 결실

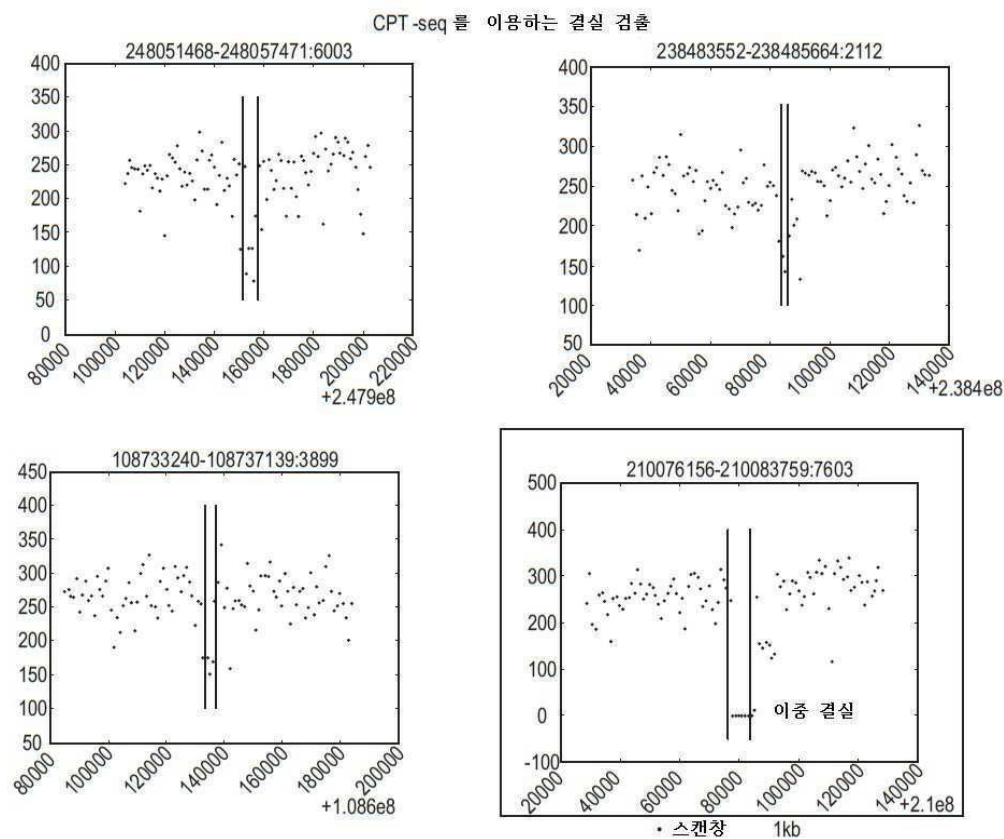
도면47b



도면48



도면49



도면50

중아황산염 전환 효율 최적화

· 전환 효율을 개선시키기 위한 최적화된 프로메가 메탈에지 중아황산염 전환 시스템

조건	DNA	비드	BSC 처리
1	10ng	없음	60°C 에서 1시간 / 0.3M NaOH
2		있음	60°C 에서 1시간 / 0.3M NaOH
3			60°C 에서 1시간 / 1M NaOH
4			65°C 에서 1시간 / 0.3M NaOH

· 중아황산염 전환 처리 효율을 결정하기 위해 분석한 ME 서열

비전환 ME	GATGTGTATAAGAGACAG
BSC ME	AATATATATAAAAAACAA

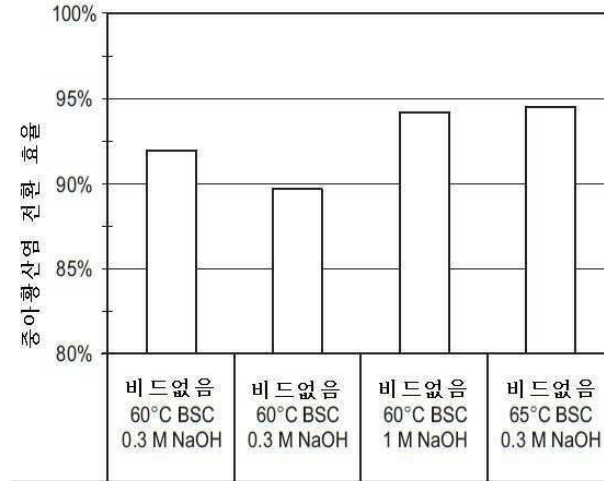


도면51

중아황산염 전환 효율 최적화

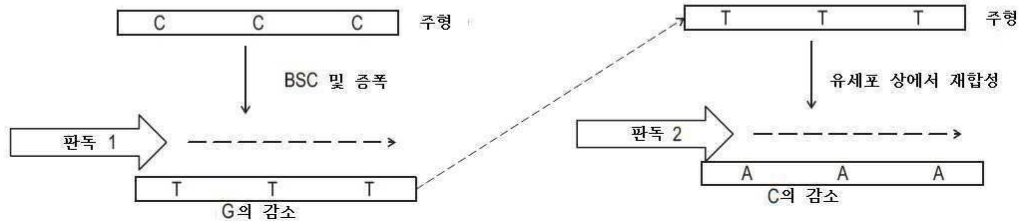
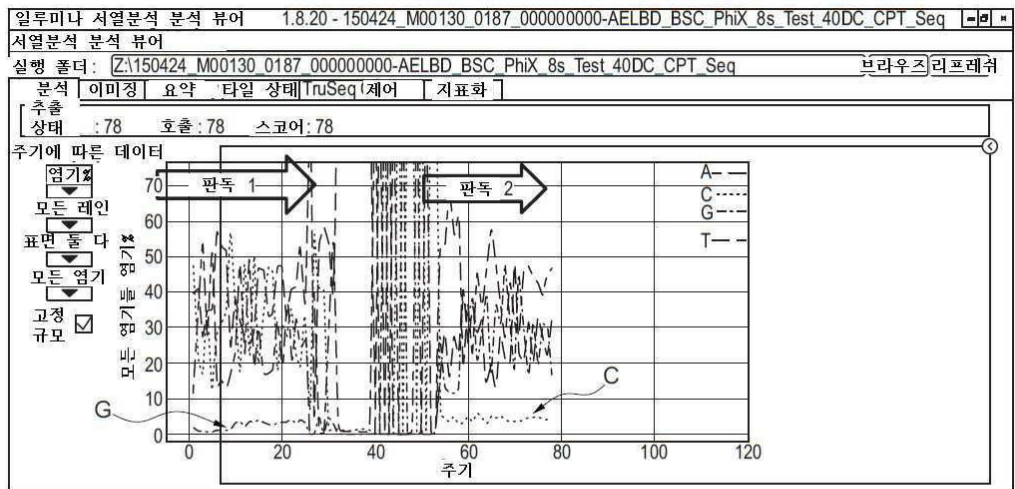
결과

- 비드에 부착된 지표-연결된 라이브러리의 95% 중아황산염 전환(BSC)
- 중아황산염 조건> 획득한 중아황산염 처리 간에 관찰된 유사한 PCR 수율은 라이브러리를 분해시키는 것으로 나타나지 않음.

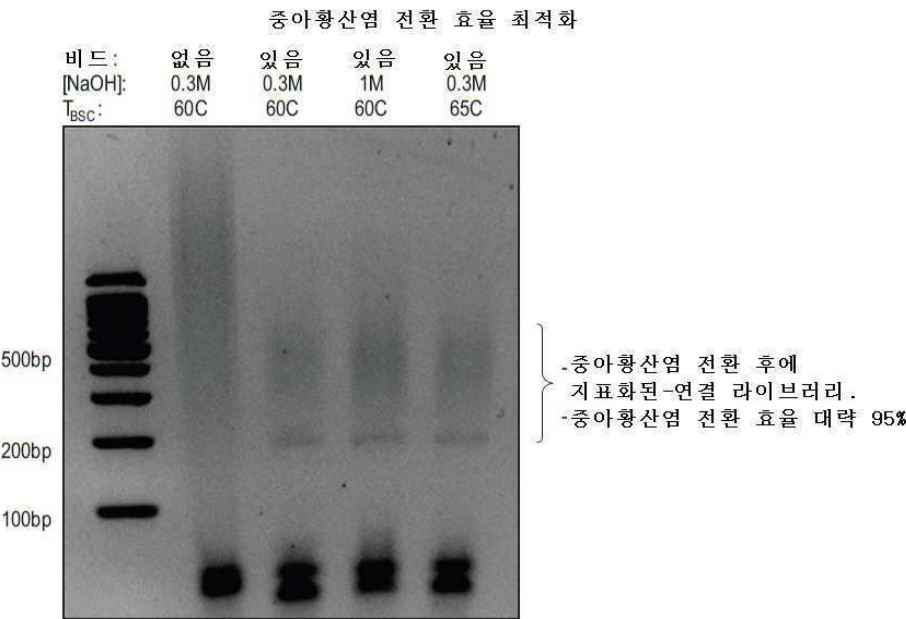


도면52

IVC 플롯(개개 염기 당 강도 대 주기)



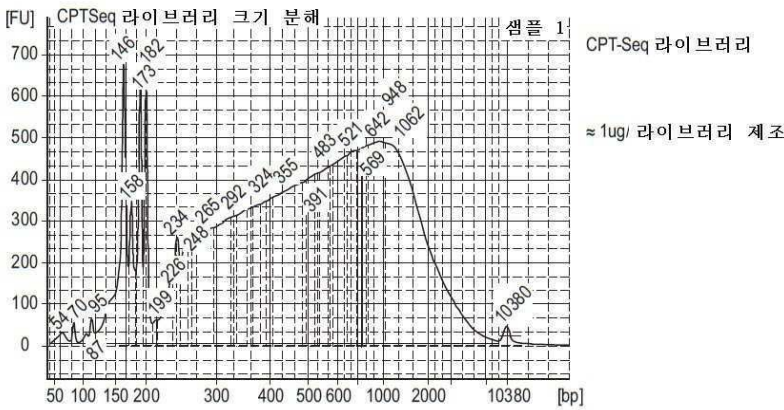
도면53



도면54

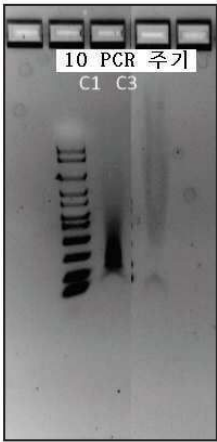
CPT-Seq 라이브러리의 강화

조건	라이브러리	프로브 세트	[프로브] pM/ea	혼성체 용적	혼성체 램프	EHB	용리	세척	SMB
1	1.1ug CPTSeq	DNA Dunn A, B, K	100pM	25uL	95 > 58 (O/N)	형태, Nxt Blkrs	ET1	표준	3x
2	1.1ug CPTSeq	EEX	2.5uL	25uL	95 > 58 (O/N)	형태, Nxt Blkrs	ET1	표준	3x
3	1.1ug CPTSeq	RNA Dunn A, B, K	100pM	25uL	95 > 65 (O/N)	형태, Nxt Blkrs	RNase H	표준	3x



도면55

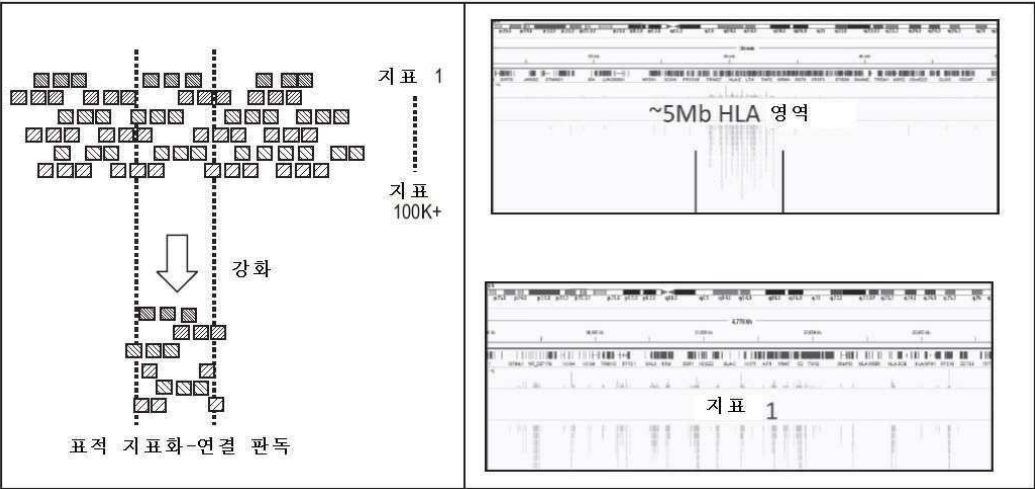
강화 후 CPT-Seq 라이브러리



· 표적화된 지표화-연결 서열분석 라이브러리

도면56

표적 단상형, HLA에 대한 적용





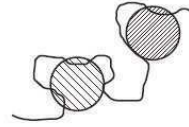
도면57

스와핑을 위한 메커니즘

1-DNA는 2개의 상이하게 지표화된 비드에 결합함

-가능한 해결책:

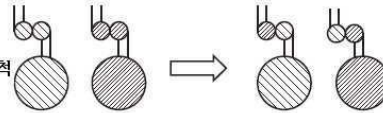
- 비드 사이의 더 많은 공간(첨가 전에 DNA를 더 많이 희석)
- 모형 비드를 첨가하여 복합체-비드를 더 밖으로 이격.



2- 복합체는 부서지고 재형성됨

- 가능한 해결책:

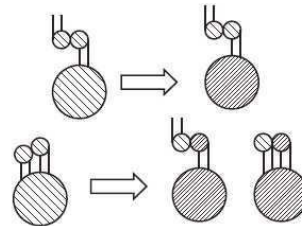
- 혼합 단계들에 의해 더 부드럽게 함
- SA에 의해 결합되지 않은 단량체를 세척
- 가교?
- 복합체를 새로 형성?



3- SA-바이오틴은 파손을 결합하고, 복합체는 다른 비드에 재혼성화함

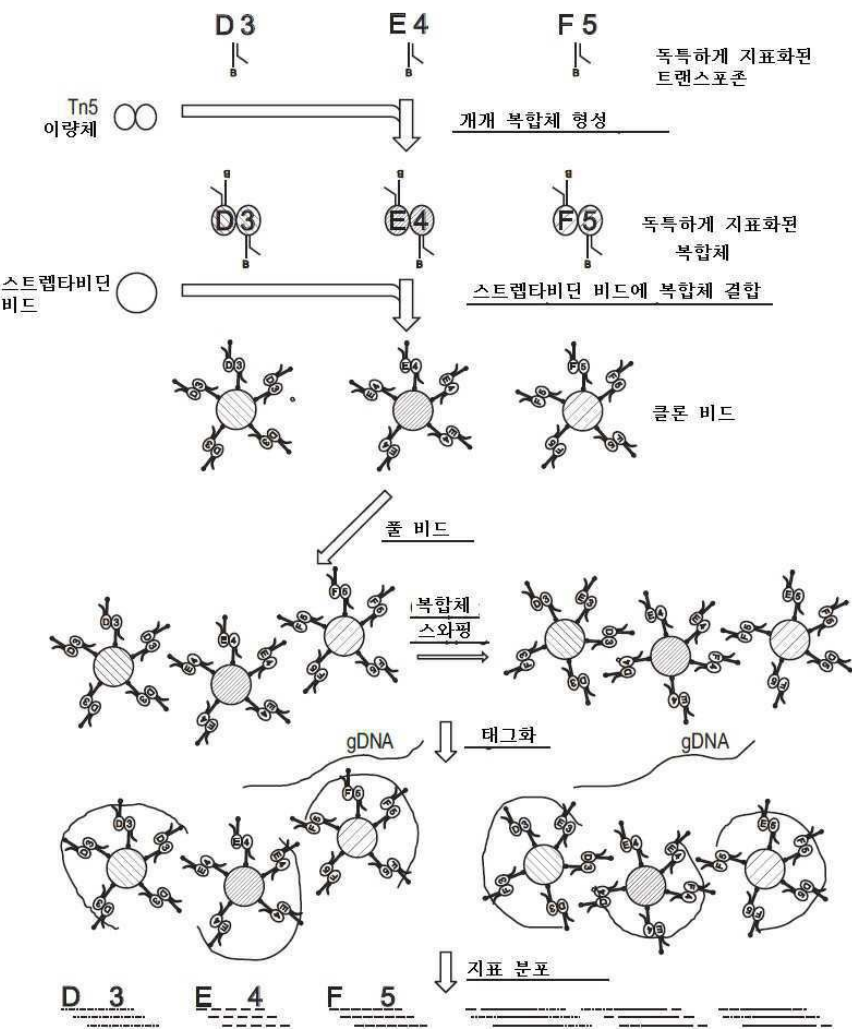
-가능한 해결책:

- 트랩타비딘 비드
- 바이오틴일화된 올리고가 결합할 수 없도록 자유 스트랩타비딘을 차단



4- 오염

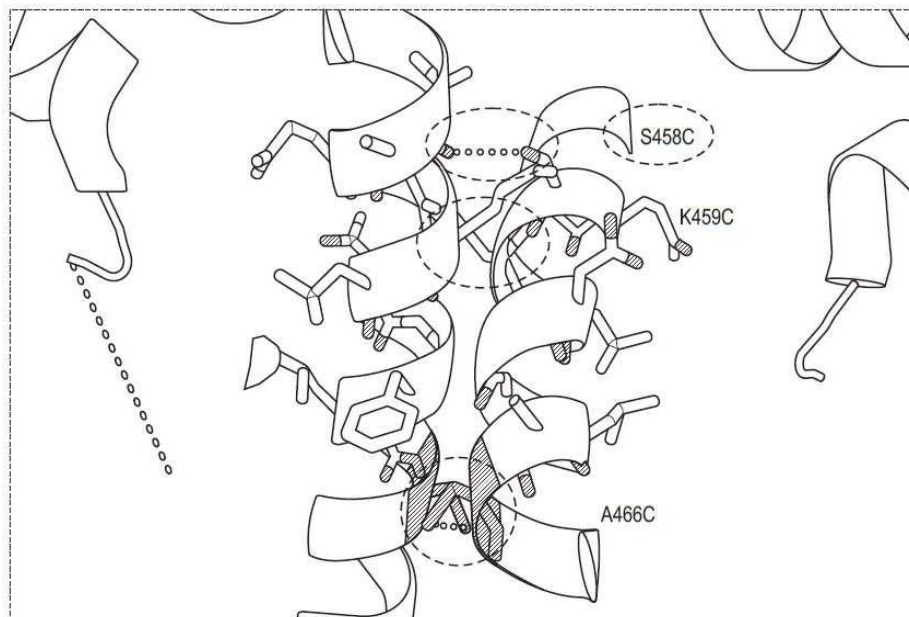
도면58



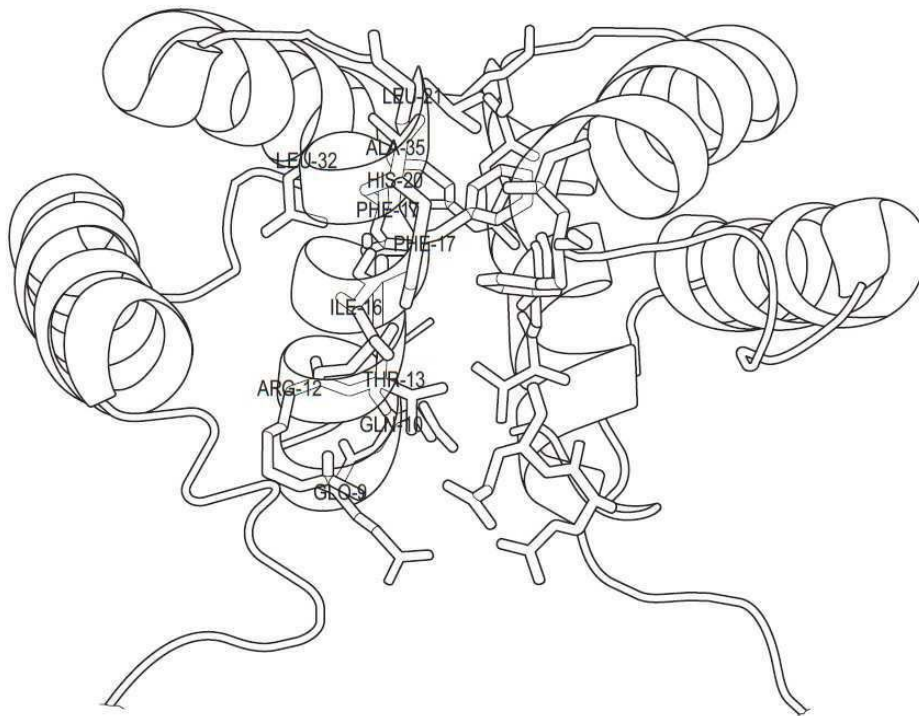
도면59



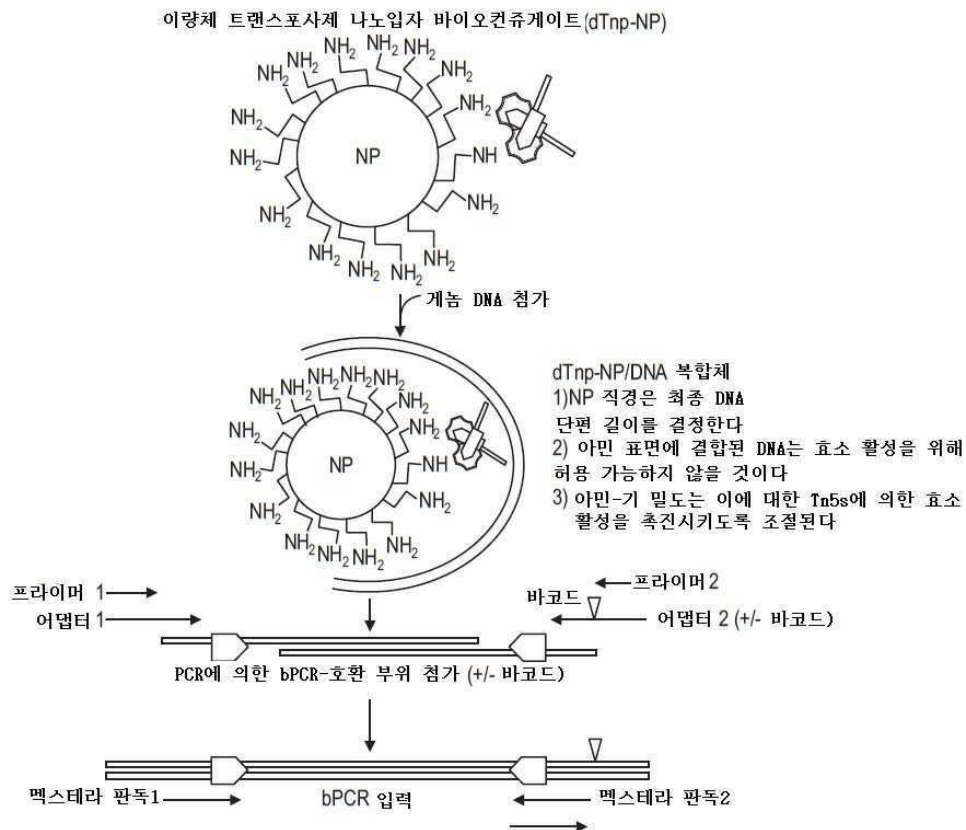
도면60



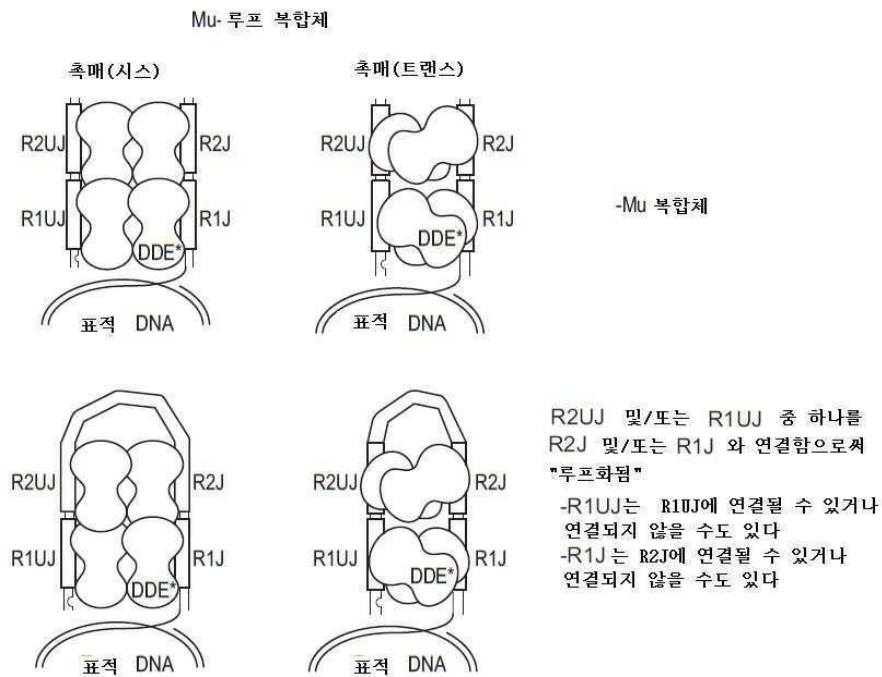
도면61



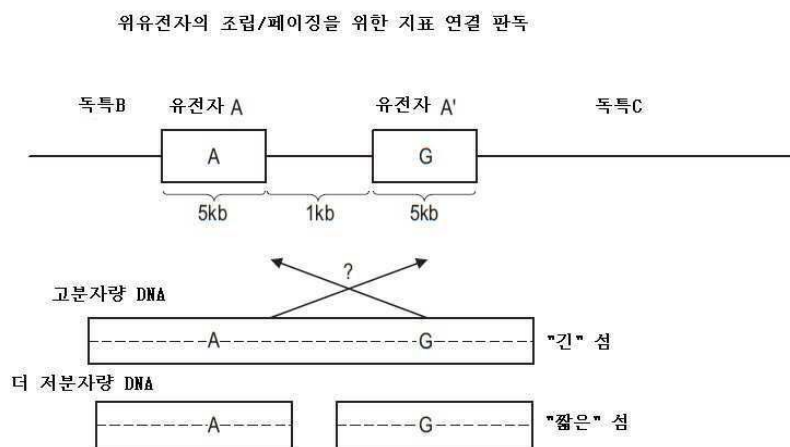
도면62



도면63

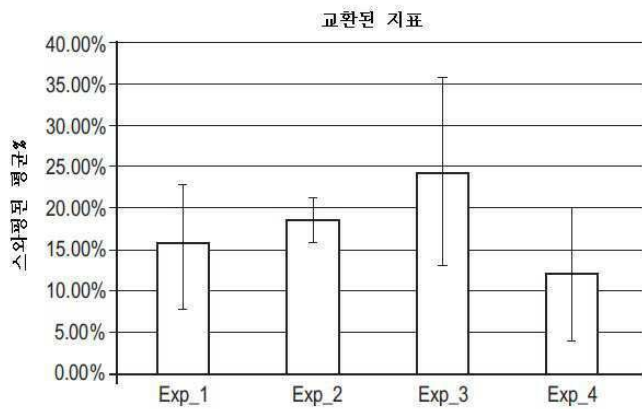


도면64

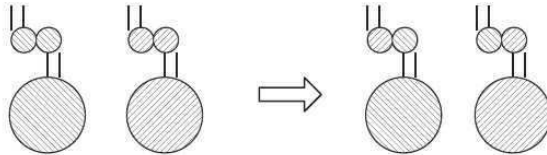


- 예, CYP2D6
  - 지표화된 연결 판독의 긴 섬 (30kb 초과)는 위유전자인 유전자 A 및 유전자 A'에 걸쳐 있다. 높은 서열때문에, 변이체가 유전자 A 또는 유전자 A'에 속하는 것을 결정할 수 없다
  - 더 짧은 섬은 독특한 주변 서열을 지니는 위유전자 중 하나에서 하나의 변이체만을 연결할 것이다
- 이는 CPT-seq 전에 DNA를 전단함으로써 달성되거나 또는 전단 및 비전단 DNA의 조합을 사용할 수 있다

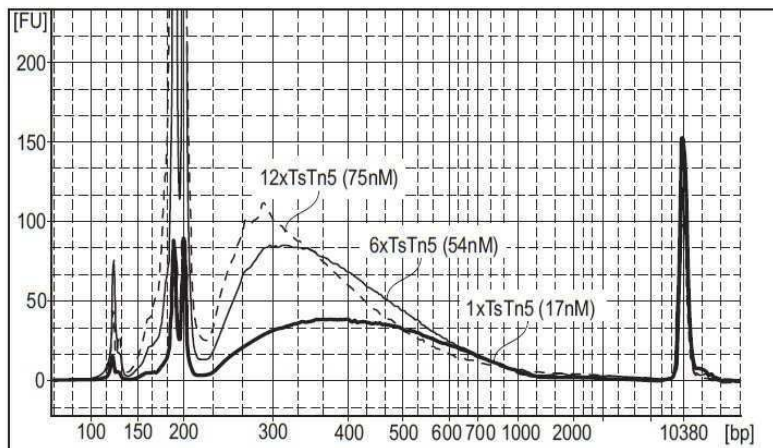
도면65



지표 교환. 상이한 지표를 지니는 비드를 제조하였다. 혼합 후에, 라이브러리를 서열분석하고 각각의 라이브러리에 대한 지표를 보고함으로써 지표 교환을 결정하였다. "스와핑된" %를  $(D4+D5+E3+E5+f4) / (\text{모두 96개의 합})$ 으로서 계산하고 도면에서 보고하였다.

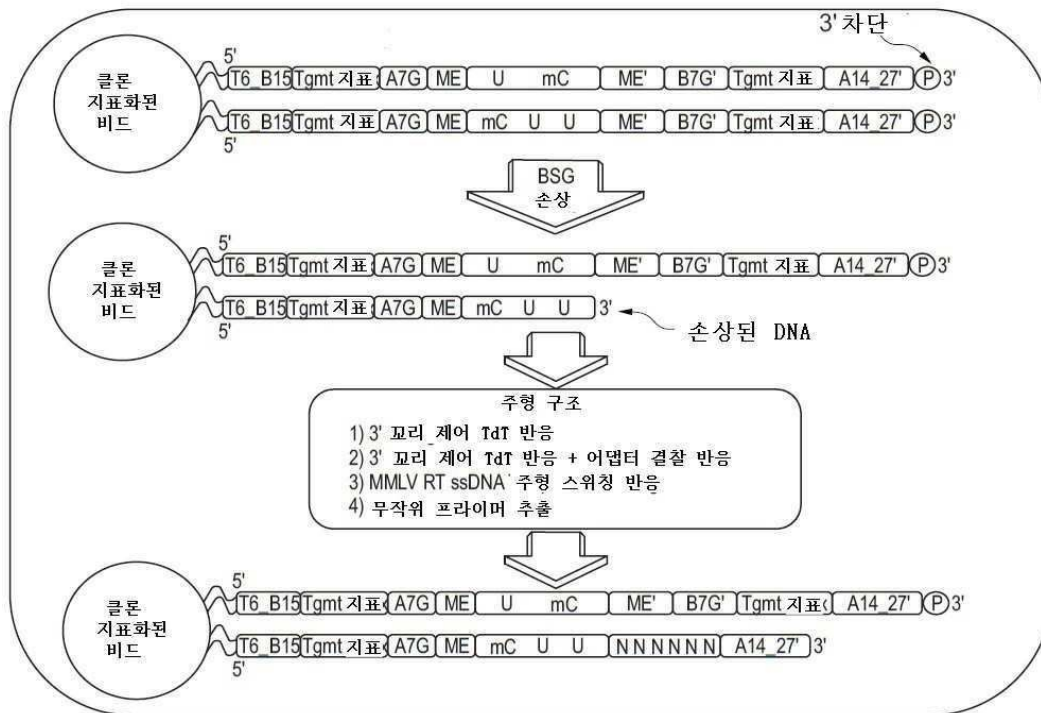


도면66

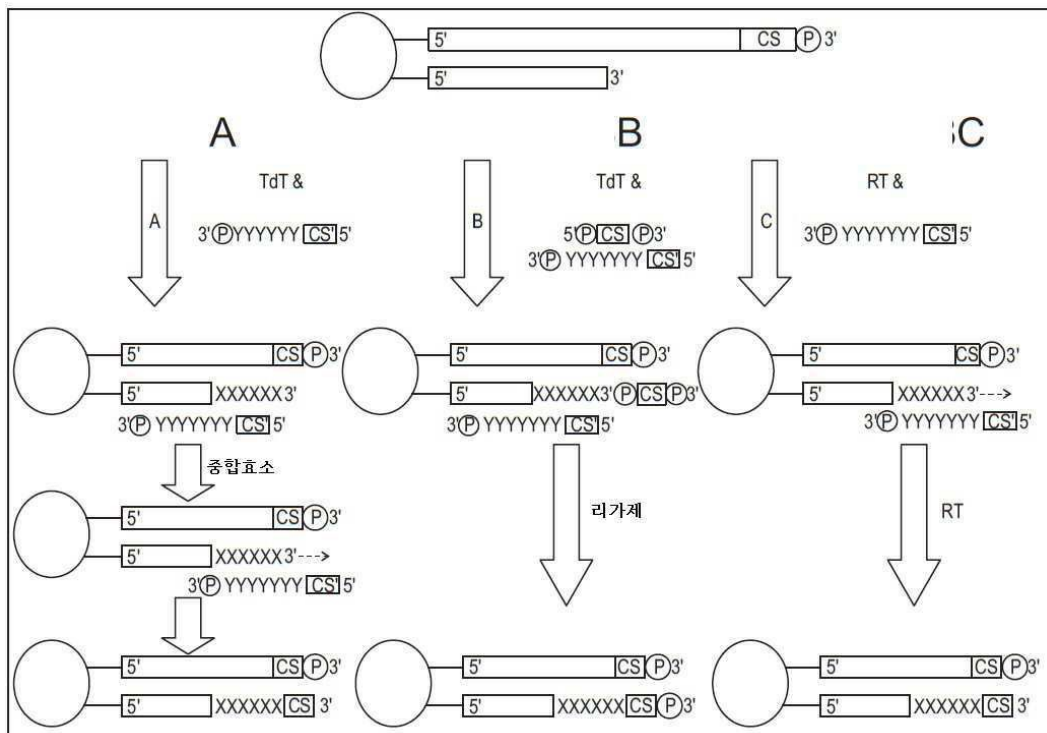




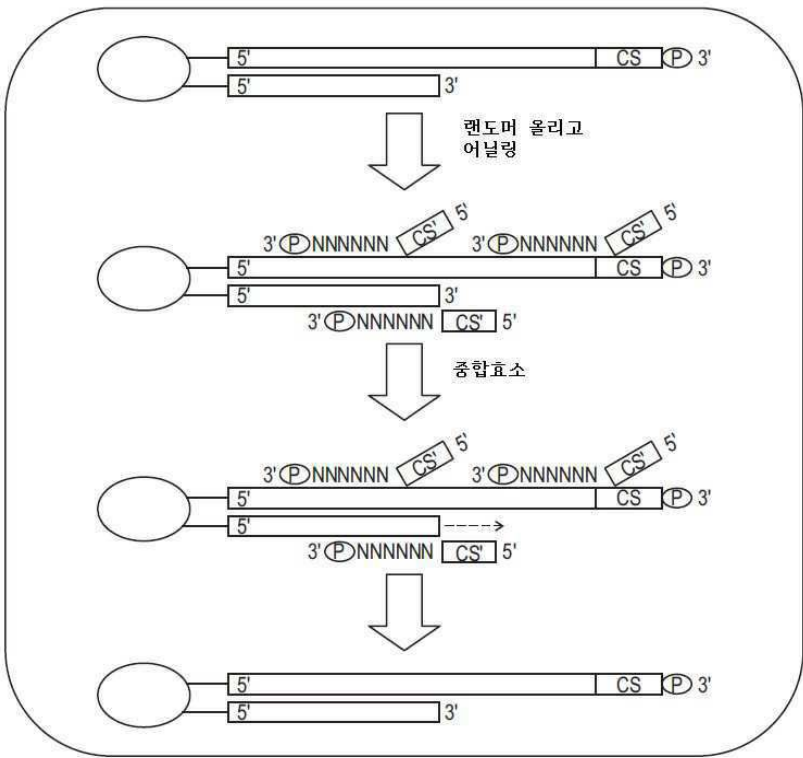
도면67



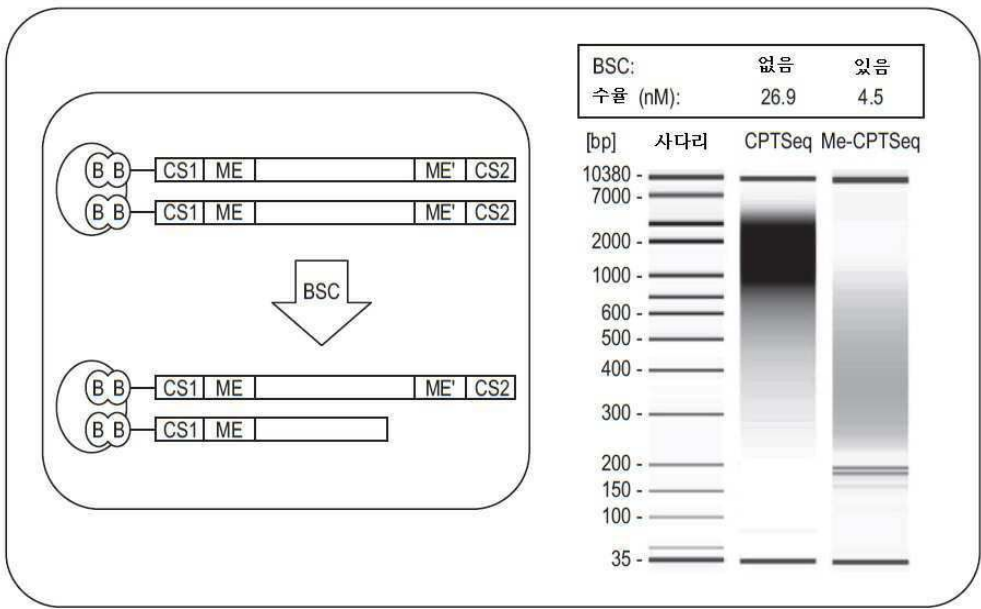
도면68



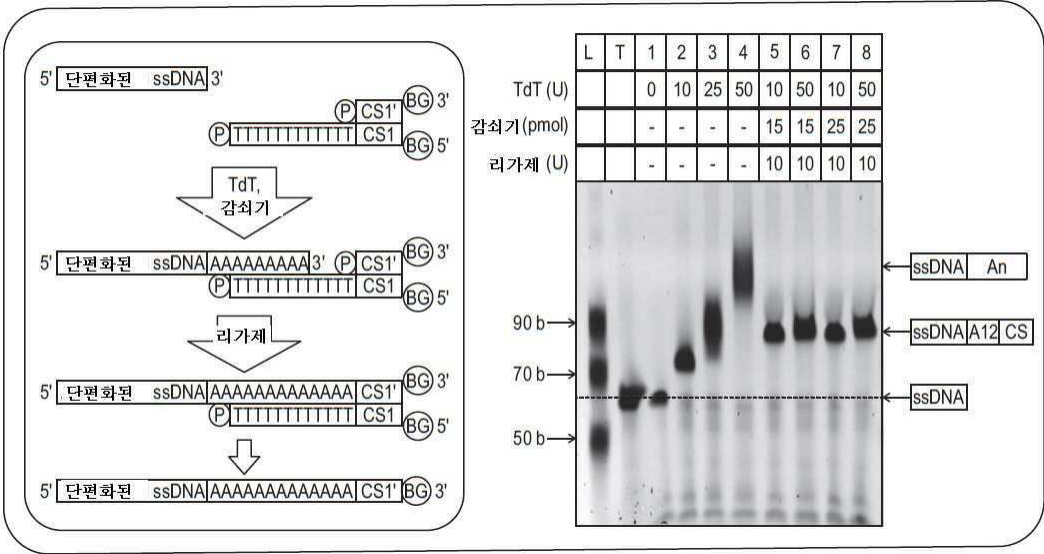
도면69



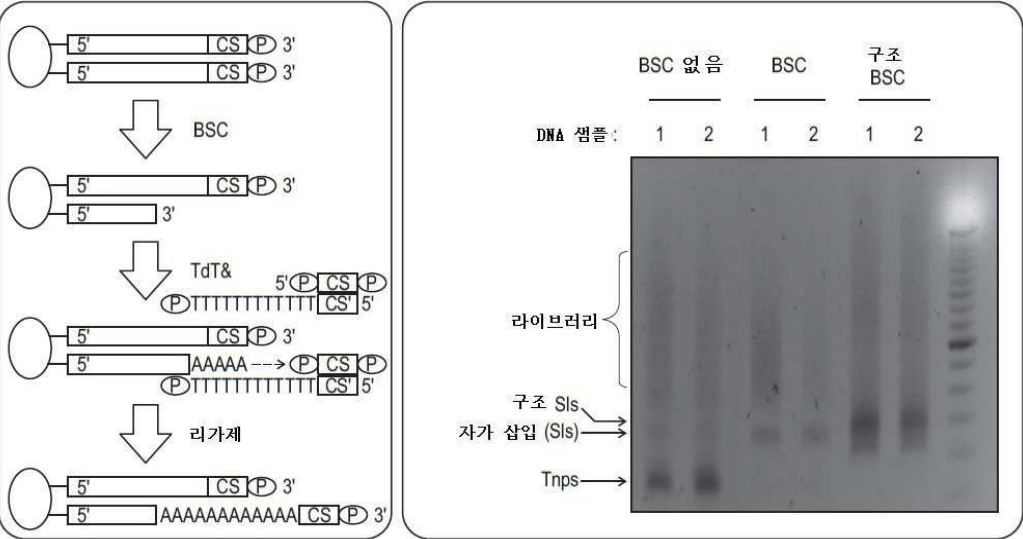
도면70



도면71



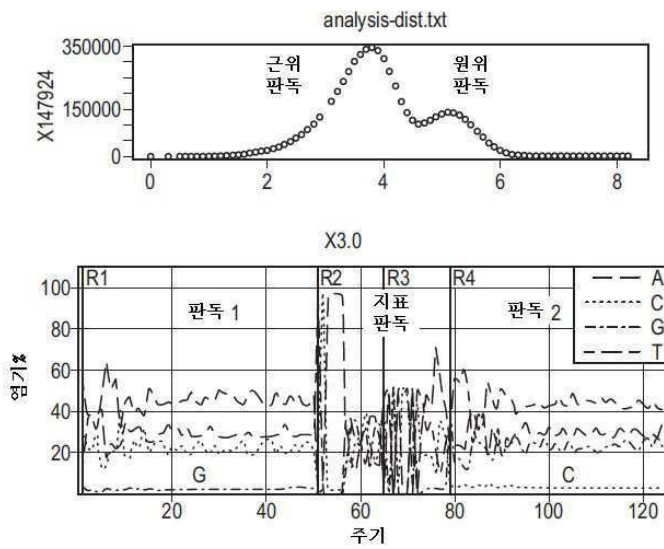
도면72



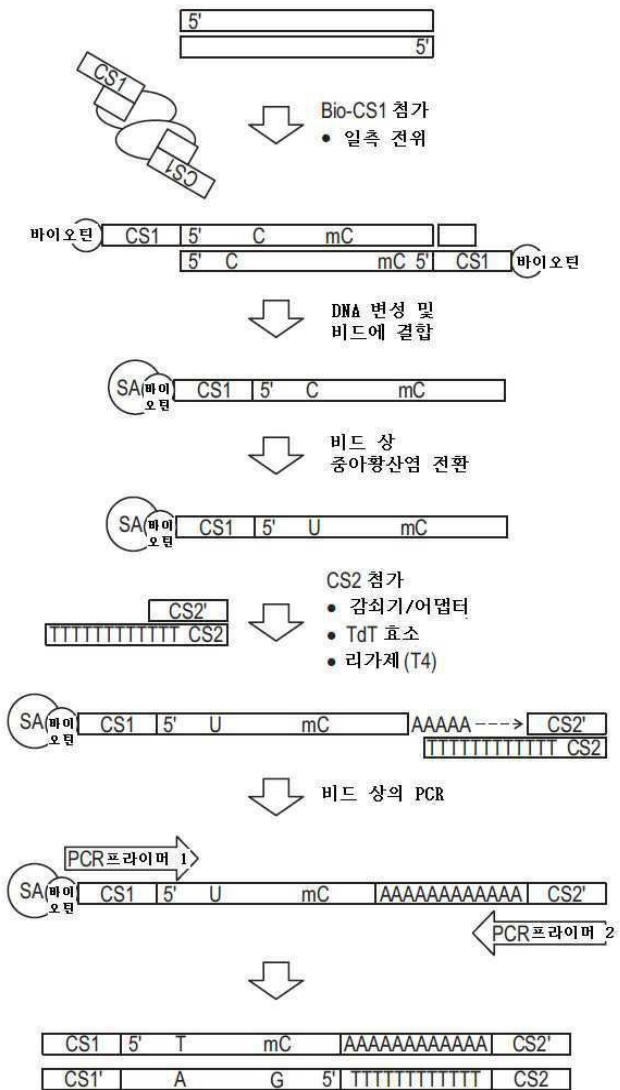
도면73

MiSeq 메틸 CPT-Seq 통계

- 관찰한 페이징 판독 "섬"
  - 평균 섬 크기 17kb
  - 50% 초과와 독특한 판독
  - 88%의 C는 U로 전환됨



## 도면 74



도면75

중아황산염 전환 효율 최적화

전환 효율을 개선시키기 위해 최적화된 프로메가의 메틸에지 중아황산염 전환 시스템

조건	DNA	비드	BSC 처리
1	10 ng	없음	60°C에서 1시간 / 0.3M NaOH
2		없음	60°C에서 1시간 / 0.3M NaOH
3		있음	60°C에서 1시간 / 1M NaOH
4		없음	65°C에서 1시간 / 0.3M NaOH

중아황산염 전환 처리 효율을 결정하기 위해 분석한 ME 서열

비전환 ME	GATGTGTATAAGAGACAG
BSC ME	AATATATATAAAAAACAA

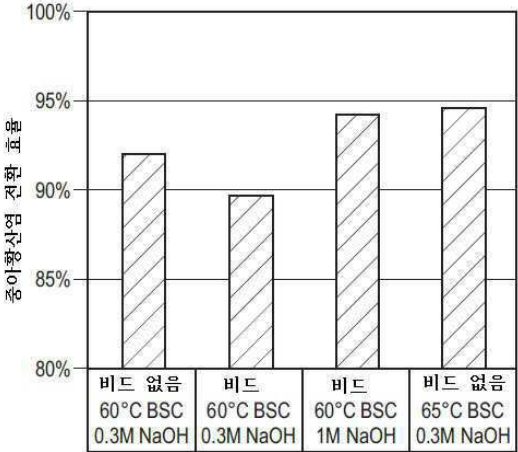
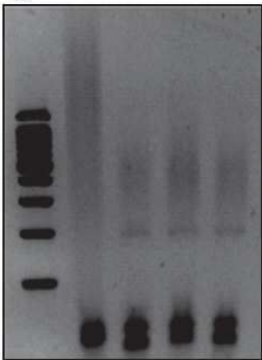
P7	B15'	바코드	'A7G'	ME	게놈 DNA	ME	B7G	바코드	A*1*4*	P5
P7	B15'	바코드	A7G	ME	게놈 DNA	ME	B7G	바코드	A14'	P5'

A

결과

- 1M NaOH 또는 65°C 열처리를 이용하는 95% 중아황산염 전환
- 중아황산염 조건 > 혹독한 중아황산염 처리 사이에 관찰된 유사한 PCR 수율은 라이브러리를 분해한 것으로 나타나지 않음

비드: 없음 있음 있음 있음  
[NaOH]: 0.3M 0.3M 1M 0.3M  
T<sub>asc</sub>: 60C 60C 60C 65C



B