



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2024년12월04일

(11) 등록번호 10-2738175

(24) 등록일자 2024년11월29일

- (51) 국제특허분류(Int. Cl.)  
C12Q 1/6869 (2018.01) C12Q 1/6874 (2018.01)
- (52) CPC특허분류  
C12Q 1/6869 (2018.05)  
C12Q 1/6874 (2018.05)
- (21) 출원번호 10-2022-7022673(분할)
- (22) 출원일자(국제) 2015년11월19일  
심사청구일자 2022년08월01일
- (85) 번역문제출일자 2022년07월01일
- (65) 공개번호 10-2022-0100098
- (43) 공개일자 2022년07월14일
- (62) 원출원 특허 10-2021-7027822  
원출원일자(국제) 2015년11월19일  
심사청구일자 2021년09월30일
- (86) 국제출원번호 PCT/US2015/061615
- (87) 국제공개번호 WO 2016/081740  
국제공개일자 2016년05월26일
- (30) 우선권주장  
62/082,883 2014년11월21일 미국(US)
- (56) 선행기술조사문헌  
W02013055995 A2\*  
\*는 심사관에 의하여 인용된 문헌

- (73) 특허권자  
브루커 스페셜 바이오로지, 인크.  
미국 매사추세츠 01821 빌레리카 매닝 로드 40 브루커 코퍼레이션 씨/오
- (72) 발명자  
비켄 조셉 엠  
미국 97405 오리건주 유진 릿지라인 드라이브 46  
카피조브 러스텝  
미국 98109 워싱턴주 시애틀 머서 스트리트 1200  
아파트먼트 212
- (74) 대리인  
김태홍, 김진희

전체 청구항 수 : 총 20 항

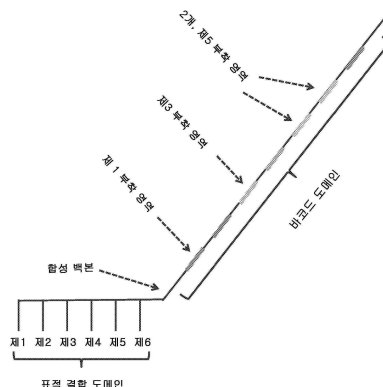
심사관 : 손영희

(54) 발명의 명칭 무효소 및 무증폭 시퀀싱

## (57) 요약

본 발명은 리드 길이가 길고, 오류율이 낮은 무효소, 무증폭, 및 무라이브러리 핵산 시퀀싱을 제공하는 시퀀싱 프로브, 방법, 키트, 및 장치를 제공한다. 표적 결합 도메인 및 바코드 도메인을 포함하고; 상기 표적 결합 도메인은 적어도 4개의 뉴클레오타이드를 포함하고, 표적 핵산에 결합할 수 있고; 상기 바코드 도메인은 합성 백본을 (뒷면에 계속)

## 대표도



포함하고, 상기 바코드 도메인은 적어도 제1 부착 영역을 포함하고, 상기 제1 부착 영역은 제1 상보적인 핵산 분자에 의해 결합될 수 있는 핵산 서열을 포함하고, 상기 제1 부착 영역의 상기 핵산 서열은 상기 표적 결합 도메인의 제1 뉴클레오티드에 의해 결합되는 상기 표적 핵산 중의 제1 뉴클레오티드의 위치 및 아이덴티티를 확인하는 것인 시퀀싱 프로브가 사용된다.

[대표도]

도 1, 도 2

(52) CPC특허분류

*C12Q 2525/113* (2013.01)

*C12Q 2537/149* (2019.08)

*C12Q 2563/179* (2013.01)

---

## 명세서

### 청구범위

#### 청구항 1

표적 결합 도메인 및 적어도 2개의 부착 위치를 포함하는 프로브로서,  
표적 결합 도메인은 적어도 12개의 뉴클레오티드를 포함하고, 표적 핵산에 결합할 수 있으며;  
적어도 2개의 부착 위치는 하나 이상의 바코드 도메인 상에 위치하고,  
각 부착 위치는 적어도 1개의 부착 영역을 포함하고, 부착 영역은 상보적인 핵산 분자에 의해 결합될 수 있는 적어도 1개의 핵산 서열을 포함하고,  
적어도 2개의 부착 위치 각각은 상이한 핵산 서열을 갖고,  
적어도 2개의 부착 위치의 각 위치의 핵산 서열은 표적 결합 도메인에 의해 결합된 표적 핵산의 아이덴티티를 결정하는 것인 프로브.

#### 청구항 2

제1항에 있어서, 하나 이상의 바코드 도메인은 하나 이상의 합성 백본을 포함하는 것인 프로브.

#### 청구항 3

제2항에 있어서, 합성 백본은 단일-가닥 DNA를 포함하는 것인 프로브.

#### 청구항 4

제1항에 있어서, 각 부착 위치는 (a) 동일한 수의 부착 영역, (b) 1개의 부착 영역, 또는 (c) 1개 초과 부착 영역을 갖는 것인 프로브.

#### 청구항 5

제1항에 있어서, 각 부착 위치는 8 내지 20개의 뉴클레오티드를 포함하는 것인 프로브.

#### 청구항 6

제5항에 있어서, 각 부착 위치는 14개의 뉴클레오티드를 포함하는 것인 프로브.

#### 청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서, 제1 부착 위치의 제1 부착 영역에 하이브리드화된 상보적인 핵산 분자를 추가로 포함하는 것인 프로브.

#### 청구항 8

제7항에 있어서, 상보적인 핵산 분자는 검출가능한 표지를 포함하는 것인 프로브.

#### 청구항 9

제7항에 있어서, 상보적인 핵산 분자는 리포터 복합체의 상보적인 핵산 분자인 프로브.

#### 청구항 10

제9항에 있어서, 상보적인 핵산 분자는 1차 핵산 분자에 직접 또는 간접적으로 연결되는 것인 프로브.

#### 청구항 11

제10항에 있어서, 1차 핵산 분자는 적어도 1개, 2개, 3개, 4개 또는 5개의 2차 핵산 분자에 하이브리드화되는 것인 프로브.

## 청구항 12

제11항에 있어서, 각 2차 핵산 분자는 적어도 1개의 검출가능한 표지를 포함하는 것인 프로브.

## 청구항 13

제11항에 있어서, 각 2차 핵산 분자는 적어도 1개의 검출가능한 표지를 각각 포함하는 적어도 1개, 2개, 3개, 4개, 5개, 6개 또는 7개의 3차 핵산 분자에 하이브리드화되는 것인 프로브.

## 청구항 14

제1항 내지 제6항 중 어느 한 항의 프로브를 복수 개 포함하는 프로브의 집단.

## 청구항 15

(1) 제1항 내지 제6항 중 어느 한 항의 프로브를 복수 개 포함하는 제1 프로브의 적어도 1개의 제1 집단을 표적 핵산에 하이브리드화시키는 단계;

(2) 검출가능한 표지를 포함하는 제1 상보적인 핵산 분자 또는 검출가능한 표지를 포함하는 제1 리포터 복합체의 제1 상보적인 핵산 분자를 적어도 2개의 부착 위치 중 제1 부착 위치에 결합시키는 단계;

(3) 결합된 제1 상보적인 핵산 분자의 검출가능한 표지 또는 결합된 제1 리포터 복합체의 검출가능한 표지를 검출하는 단계;

(4) 제1 상보적인 핵산 분자 또는 제1 리포터 복합체의 검출가능한 표지를 제1 부착 위치로부터 결합 해제시키는 단계;

(5) 검출가능한 표지를 포함하는 제2 상보적인 핵산 분자 또는 검출가능한 표지를 포함하는 제2 리포터 복합체의 제2 상보적인 핵산 분자를 적어도 2개의 부착 위치 중 제2 부착 위치에 결합시키는 단계;

(6) 결합된 제2 상보적인 핵산 분자의 검출가능한 표지 또는 결합된 제2 리포터 복합체의 검출가능한 표지를 검출하는 단계; 및

(7) 적어도 2개의 부착 위치의 각 부착 위치가 검출가능한 표지를 포함하는 상보적인 핵산 분자 또는 검출가능한 표지를 포함하는 리포터 복합체의 상보적인 핵산 분자에 의해 결합되고, 결합된 상보적인 핵산 분자의 검출가능한 표지 또는 결합된 리포터 복합체의 검출가능한 표지가 검출될 때까지 단계 (4) 내지 (6)를 반복함으로써, 제1 프로브의 표적 결합 도메인에 의해 하이브리드화된 표적 핵산을 확인하는 단계

를 포함하는, 표적 핵산을 확인하는 방법.

## 청구항 16

제15항에 있어서, 단계 (4) 및 (5)는 순차적으로 또는 동시에 진행되는 것인 방법.

## 청구항 17

제15항에 있어서, 검출가능한 표지를 포함하는 제1 리포터 복합체의 제1 상보적인 핵산 분자는 1차 핵산 분자에 직접 또는 간접적으로 연결되는 것인 방법.

## 청구항 18

제17항에 있어서, 1차 핵산 분자는 적어도 1개, 2개, 3개, 4개 또는 5개의 2차 핵산 분자에 하이브리드화되는 것인 방법.

## 청구항 19

제18항에 있어서, 각 2차 핵산 분자는 적어도 1개의 검출가능한 표지를 포함하는 것인 방법.

## 청구항 20

제18항에 있어서, 각 2차 핵산 분자는 적어도 1개의 검출가능한 표지를 포함하는 적어도 1개, 2개, 3개, 4개, 5개, 6개 또는 7개의 3차 핵산 분자에 하이브리드화되는 것인 방법.



청구항 21

삭제

청구항 22

삭제

청구항 23

삭제

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

청구항 35

삭제

청구항 36

삭제

청구항 37

삭제

청구항 38

삭제

청구항 39

삭제

청구항 40

삭제

청구항 41

삭제

청구항 42

삭제

청구항 43

삭제

청구항 44

삭제

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

삭제

청구항 50

삭제

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

삭제

청구항 55

삭제

청구항 56

삭제

청구항 57

삭제

청구항 58

삭제

청구항 59

삭제

청구항 60

삭제

청구항 61

삭제

청구항 62

삭제

청구항 63

삭제

청구항 64

삭제

청구항 65

삭제

청구항 66

삭제

청구항 67

삭제

청구항 68

삭제

청구항 69

삭제

청구항 70

삭제

청구항 71

삭제

청구항 72

삭제

청구항 73

삭제

청구항 74

삭제

청구항 75

삭제

청구항 76

삭제

청구항 77

삭제

청구항 78

삭제

청구항 79

삭제

청구항 80

삭제

청구항 81

삭제

청구항 82

삭제

청구항 83

삭제

청구항 84

삭제

청구항 85

삭제

청구항 86

삭제

청구항 87

삭제

청구항 88

삭제

청구항 89

삭제

청구항 90

삭제

청구항 91

삭제

청구항 92

삭제

청구항 93

삭제

청구항 94

삭제

청구항 95

삭제

청구항 96

삭제

청구항 97

삭제

청구항 98

삭제

청구항 99

삭제

청구항 100

삭제

청구항 101

삭제

청구항 102

삭제

청구항 103

삭제

청구항 104

삭제

청구항 105

삭제

청구항 106

삭제

청구항 107

삭제

청구항 108

삭제

청구항 109

삭제

청구항 110

삭제

청구항 111

삭제

청구항 112

삭제

청구항 113

삭제

청구항 114

삭제

청구항 115

삭제

청구항 116

삭제

청구항 117

삭제

청구항 118

삭제

청구항 119

삭제

청구항 120

삭제

청구항 121

삭제

청구항 122

삭제

청구항 123

삭제

청구항 124

삭제

청구항 125

삭제

청구항 126

삭제

청구항 127

삭제

청구항 128

삭제

청구항 129

삭제

청구항 130

삭제

청구항 131

삭제

청구항 132

삭제

청구항 133

삭제

청구항 134

삭제

청구항 135

삭제

청구항 136

삭제

청구항 137

삭제

청구항 138

삭제

청구항 139

삭제

청구항 140

삭제

청구항 141

삭제

청구항 142

삭제

청구항 143

삭제

청구항 144

삭제

청구항 145

삭제

청구항 146

삭제

청구항 147

삭제

청구항 148

삭제



청구항 149

삭제

청구항 150

삭제

청구항 151

삭제

청구항 152

삭제

청구항 153

삭제

청구항 154

삭제

청구항 155

삭제

청구항 156

삭제

청구항 157

삭제

청구항 158

삭제

청구항 159

삭제

청구항 160

삭제

청구항 161

삭제

청구항 162

삭제

청구항 163

삭제

청구항 164

삭제

청구항 165

삭제

청구항 166

삭제

청구항 167

삭제

청구항 168

삭제

청구항 169

삭제

청구항 170

삭제

청구항 171

삭제

청구항 172

삭제

청구항 173

삭제

청구항 174

삭제

청구항 175

삭제

청구항 176

삭제

청구항 177

삭제

청구항 178

삭제

청구항 179

삭제

청구항 180

삭제

청구항 181

삭제

청구항 182

삭제

청구항 183

삭제

청구항 184

삭제

청구항 185

삭제

청구항 186

삭제

청구항 187

삭제

청구항 188

삭제

청구항 189

삭제

청구항 190

삭제

청구항 191

삭제

청구항 192

삭제

청구항 193

삭제

청구항 194

삭제

청구항 195

삭제

**발명의 설명**

**기술 분야**

- [0001] 관련 출원에 대한 상호 참조
- [0002] 본 출원은 2014년 11월 21일 출원된 미국 가출원 번호 62/082,883의 이점을 주장한다. 상기 언급된 특허 출원의 내용은 그 전문이 본원에서 참조로 포함된다.
- [0003] 서열 목록
- [0004] 본 출원은 EFS-웹을 통해 ASCII 포맷으로 제출된 서열 목록을 포함하며, 이는 그 전문이 본원에서 참조로 포함된다. 2015년 11월 19일 작성된 상기 ASCII 사본은 NATE-025\_ST25.txt로 명명되고, 그 크기는 20,860 바이트이다.

## 배경 기술

- [0005] 본 발명의 배경기술
- [0006] 현재 다양한 핵산 시퀀싱(서열분석) 방법, 즉, 핵산 분자 내의 뉴클레오타이드의 정확한 순서를 측정하는 방법이 존재한다. 현행 방법은 예컨대, PCR과 같이 효소적으로 및/또는 클로닝하여 핵산을 증폭시키는 것을 필요로 한다. 광 검출 수단에 의해 검출가능한 신호를 생성하는 데 추가의 효소적 중합화가 요구된다. 상기 증폭 및 중합화 단계는 비용이 많이 들고/거나, 시간이 많이 소요된다. 따라서, 당업계에서는 무증폭 및 무효소 핵산 시퀀싱 방법이 요구되고 있다. 본 발명은 이러한 요구를 다룬다.

## 발명의 내용

- [0007] 본 발명의 요약
- [0008] 본 발명은 리드(read) 길이가 길고, 오류율이 낮은 무효소, 무증폭, 및 무라이브러리 핵산 시퀀싱을 제공하는 시퀀싱 프로브, 방법, 키트, 및 장치를 제공한다. 또한, 방법, 키트, 및 장치는 신속하게 샘플링에서부터 응답까지 실행할 수 있는 능력(sample-to-answer capability)을 가진다. 이러한 특징은 특히 임상 환경에서의 시퀀싱에 유용하다.
- [0009] 본원에서는 표적 결합 도메인 및 바코드 도메인을 포함하는 시퀀싱 프로브를 제공한다. 표적 결합 도메인 및 바코드 도메인은 작동가능하게 연결, 예컨대, 공유적으로 연결될 수 있다. 시퀀싱 프로브는 임의적으로 표적 결합 도메인과 바코드 도메인 사이에 스페이서를 포함한다. 스페이서는 예를 들어, (1 내지 100개의 뉴클레오타이드, 예컨대, 2 내지 50개의 뉴클레오타이드로 이루어진) 단일 가닥 또는 이중 가닥 DNA 스페이서와 같은 적절한 기계적 성질을 가진 임의의 중합체일 수 있다. 이중 가닥 DNA 스페이서의 비제한적인 예로는 서열 번호: 25 내지 서열 번호: 29에 의해 커버되는 서열을 포함한다.
- [0010] 표적 결합 도메인은 적어도 4개의 뉴클레오타이드(예컨대, 4, 5, 6, 7, 8, 9, 10, 11, 12개 이상)를 포함하고, 표적 핵산(예컨대, DNA, RNA, 및 PNA)에 결합할 수 있다. 바코드 도메인은 합성 백본을 포함하며, 상기 바코드 도메인은 하나 이상의 부착 영역을 포함하는 적어도 제1 위치를 가진다. 바코드 도메인은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12개 이상의 위치를 가질 수 있고; 각 위치는 하나 이상의(예컨대, 1 내지 50개의) 부착 영역을 가지고; 각 부착 영역은 상보적인 핵산 분자 (RNA 또는 DNA)에 가역적으로 결합할 수 있는 적어도 하나(즉, 1 내지 50개, 예컨대, 10 내지 30개의 핵산 서열(들)의 카피)를 포함한다. 바코드 도메인 중의 특정 위치는 다른 위치보다 더 많은 부착 부위를 가질 수 있고; 대안적으로, 바코드 도메인 중의 각 위치는 동일 개수의 부착 부위를 가진다. 제1 부착 영역의 핵산 서열이 표적 결합 도메인의 제1 뉴클레오타이드가 결합하는 표적 핵산 중의 제1 뉴클레오타이드의 위치 및 아이덴티티(identity)를 결정하는 반면, 제2 부착 영역의 핵산 서열은 표적 결합 도메인의 제2 뉴클레오타이드가 결합하는 표적 핵산 중의 제2 뉴클레오타이드의 위치 및 아이덴티티를 결정한다. 유사하게, 제6 부착 영역의 핵산 서열은 표적 결합 도메인의 제6 뉴클레오타이드가 결합하는 표적 핵산 중의 제6 뉴클레오타이드의 위치 및 아이덴티티를 결정한다. 실시양태에서, 합성 백본은 다당류, 폴리뉴클레오타이드 (예컨대, 단일 또는 이중 가닥 DNA 또는 RNA), 펩티드, 펩티드 핵산, 또는 폴리펩티드를 포함한다. 표적 결합 도메인 중의 뉴클레오타이드의 개수는 바코드 도메인 중의 위치의 개수와 동일하거나, 또는 그보다 크다(예컨대, 1, 2, 3, 4개 이상). 바코드 도메인의 특정 위치의 각 부착 영역은 동일한 핵산 서열의 카피 하나 및/또는 동일한 핵산 서열의 다중 카피를 포함할 수 있다. 그러나, 부착 영역은 심지어 두 부착 영역 모두 같은 유형의 뉴클레오타이드, 예컨대, 아데닌, 티민, 시토신, 구아닌, 우라실, 및 그의 유사체를 확인하는 경우에도, 바코드 도메인의 다른 위치의 부착 영역과 다른 핵산 서열을 포함할 것이다. 부착 영역은 합성 백본 중 변형된 단량체, 예컨대, 변형된 뉴클레오타이드에 연결될 수 있고, 이로써, 백본에 대해 분지를 생성할 수 있다. 부착 영역은 합

성 백본의 폴리뉴클레오티드 서열의 일부일 수 있다. 하나 이상의 부착 영역은 적어도 하나의 플랭킹(flancking) 단일 가닥 폴리뉴클레오티드에 인접해 있을 수 있고, 즉, 부착 영역은 5' 플랭킹 단일 가닥 폴리뉴클레오티드에 및/또는 3' 플랭킹 단일 가닥 폴리뉴클레오티드에 작동가능하게 연결될 수 있다. 1 또는 2개의 플랭킹 단일 가닥 폴리뉴클레오티드를 갖거나 갖지 않은 부착 영역은 검출가능한 표지가 없는 하이브리드화 핵산 분자에 하이브리드화될 수 있다. 검출가능한 표지가 없는 하이브리드화 핵산 분자는 약 4 내지 약 20개의 뉴클레오티드 길이, 예컨대, 12개의 뉴클레오티드 길이일 수 있다.

[0011] 부착 영역은 검출가능한 표지를 포함하는 상보적인 핵산에 의해 결합될 수 있다. 각각의 상보적인 핵산은 검출가능한 표지를 포함할 수 있다.

[0012] 대안적으로, 부착 영역은 (검출가능한 표지를 포함하는) 리포터 복합체의 일부인 상보적인 핵산에 의해 결합될 수 있다. (검출가능한 표지를 포함하거나 리포터 복합체의) 상보적인 핵산은 약 4 내지 약 20개의 뉴클레오티드 길이, 예컨대, 약 8, 10, 12 및 14개 또는 그 초과인 뉴클레오티드 길이일 수 있다. 리포터 복합체에서, 상보적인 핵산은 1차 핵산 분자에 (직접 또는 간접적으로) 연결된다. 상보적인 핵산은 단일 또는 이중 가닥 핵산 링커 (예컨대, 1 내지 100개의 뉴클레오티드를 포함하는 폴리뉴클레오티드)를 통해 1차 핵산 분자에 간접적으로 연결될 수 있다. 1차 핵산은 하나 이상의 (예컨대, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의) 2차 핵산에 하이브리드화된다. 각각의 2차 핵산은 하나 이상의 (예컨대, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의) 3차 핵산에 하이브리드화되고; 3차 핵산은 하나 이상의 검출가능한 표지를 포함한다. 하나의 또는 각각의 2차 핵산은, 1차 핵산 분자에 하이브리드화하지 않고 3차 핵산 분자에 하이브리드화하지 않는 영역("추가 핸들(extra-handle)")을 포함할 수 있고; 이 영역은 4개 이상의(예컨대, 약 6 내지 약 40개, 예컨대, 약 8, 10, 12, 및 14개의) 뉴클레오티드 길이일 수 있다. 1차 핵산 분자에 하이브리드화하지 않고 3차 핵산 분자에 하이브리드화하지 않는 영역은 1차 핵산 분자에 연결되는 상보적인 핵산 분자의 뉴클레오티드 서열을 포함할 수 있다. 상기 영역은 1차 핵산에 하이브리드화하는 상기 영역 단부 원위부에 위치하는 2차 핵산의 단부 인근에 위치할 수 있다. 상보적인 핵산의 뉴클레오티드 서열을 포함하는 "추가 핸들"을 가짐으로써, 리포터 복합체가 시퀀싱 프로브에 결합하는 가능성 및 속도가 크게 증가된다. 본 발명의 임의의 실시양태 또는 측면에서, 리포터 복합체가 "추가 핸들"을 포함하는 경우, 리포터 복합체는 리포터 복합체의 상보적인 핵산을 통해 또는 "추가 핸들"을 통해 시퀀싱 프로브에 하이브리드화할 수 있다. 따라서, 예를 들어, "제1 리포터 복합체의 제1 상보적인 핵산 분자를 ... 제1 부착 영역에 결합시키는 단계"라는 어구는 그의 명백한 의미에 따라 이해될 것이며, 이는 또한 "제1 리포터 복합체의 '추가 핸들'을 ... 제1 영역에 결합시키는 단계"라는 것을 의미하는 것으로도 이해될 것이다.

[0013] 실시양태에서, "바코드 도메인" 및 "합성 백본"이라는 용어는 동의어이다.

[0014] 본원에서는 본 발명의 시퀀싱 프로브를 이용하여 핵산을 시퀀싱하는 방법을 제공한다. 본 방법은 (1) 기재(substrate)에 (예컨대, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 위치에) 고정화된 표적 핵산에 본 발명의 적어도 하나의 시퀀싱 프로브를 하이브리드화시키는 단계; (2) 검출가능한 표지(예컨대, 형광성 표지)를 가지는 제1 상보적인 핵산 분자(RNA 또는 DNA) 또는 검출가능한 표지(예컨대, 형광성 표지)를 포함하는 제1 리포터 복합체의 제1 상보적인 핵산 분자를 제1 부착 영역에 결합시키는 단계; (3) 검출가능한 표지(들)를 검출하는 단계, 및 (4) 고정화된 표적 핵산 중의 제1 뉴클레오티드의 위치 및 아이덴티티를 확인하는 단계를 포함한다. 임의적으로, 고정화된 표적 핵산은 프로브결합 이전에 신장된다. 본 방법은 (5) (1 또는 2개의 플랭킹 단일 가닥 폴리뉴클레오티드를 갖거나 갖지 않은) 제1 부착 영역을, 검출가능한 표지가 없는 제1 하이브리드화 핵산 분자와 접촉시켜 검출가능한 표지를 가지는 제1 상보적인 핵산 분자 또는 검출가능한 표지를 포함하는 제1 리포터 복합체의 제1 상보적인 핵산 분자를 결합 해제시키고(unbinding), 검출가능한 표지가 없는 제1 하이브리드화 핵산을 적어도 제1 부착 영역에 결합시키는 단계; (6) 검출가능한 표지를 가지는 제2 상보적인 핵산 분자 또는 검출가능한 표지를 포함하는 제2 리포터 복합체의 상보적인 핵산 분자를 제2 부착 영역에 결합시키는 단계; (7) 검출가능한 표지(들)를 검출하는 단계; 및 (8) 고정화된 표적 핵산 중의 제2 뉴클레오티드의 위치 및 아이덴티티를 확인하는 단계를 추가로 포함한다. 단계 (5) 내지 (8)은 고정화된 표적 핵산 중의, 표적 결합 도메인에 상응하는 각 뉴클레오티드가 확인될 때까지 반복된다. 단계 (5) 및 (6)은 동시에 또는 순차적으로 진행될 수 있다. (검출가능한 표지 또는 리포터 복합체의 일부를 갖는) 각(예컨대, 제1, 제2, 제3, 제4, 제5, 제6, 제7, 제8, 제9, 제10 또는 그 이상의) 상보적인 핵산 분자는 검출가능한 표지가 없는, 그의 상응하는(즉, 제1, 제2, 제3, 제4, 제5, 제6, 제7, 제8, 제9, 제10 또는 그 이상의) 하이브리드화 핵산 분자와 동일한 핵산 서열을 가진다. 표적 핵산의 제1 위치 및/또는 제2 위치를 제1 및/또는 제2 포획 프로브와 결합시킴으로써 표적 핵산을 기재에 고정화시키고; 각 포획 프로브는 기재에 선택적으로 결합하는 친화성 태그를 포함한다. 제1 및/또는 제2 위치는 표적 핵산의 말단에 또는 그 인근에 위치할 수 있다. 기재는 당업계에 공지된 임의의 고체 지지체, 예컨

대, 코팅된 슬라이드 및 (예컨대, 스트랩트아비딘으로 코팅된) 미세유체 디바이스일 수 있다. 표적 핵산의 말단으로부터 원거리에 위치하는 다른 위치가 기재에 선택적으로 결합될 수 있다. 핵산은 표적 핵산을 연장시키는데 충분한 힘(예컨대, 중력, 수력학적 힘, 전자기력, 유동 연신(flow-stretching), 후퇴 메니스커스(receding meniscus) 기법, 및 이의 조합)을 가함으로써 신장될 수 있다.

[0015] 본원에서는 본 발명의 한 프로브 집단 또는 본 발명의 복수 개의 프로브 집단을 이용하여 핵산을 시퀀싱하는 방법을 제공한다. 본 방법은 (1) 기재에 고정화된 표적 핵산에 (본 발명의) 제1 시퀀싱 프로브 집단을 하이브리드화시키는 단계(여기서, 제1 집단 중의 각 시퀀싱 프로브는 거의 동일한 조건하에서, 예컨대, 거의 동일한 수준의 카오토평제, 온도, 염 농도, pH, 및 수력학적 힘 조건하에서 고정화된 표적 핵산으로부터 탈하이브리드화된다); (2) 각각 검출가능한 표지를 가지는 복수 개의 제1 상보적인 핵산 분자, 또는 각 복합체가 검출가능한 표지를 포함하는 복수 개의 제1 리포터 복합체의 복수 개의 제1 상보적인 핵산 분자를 제1 집단 중의 각 시퀀싱 프로브 중의 제1 부착 영역에 결합시키는 단계; (3) 검출가능한 표지(들)를 검출하는 단계; (4) 제1 집단 중의 시퀀싱 프로브에 의해 하이브리드화된 고정화된 표적 핵산 중의 복수 개의 제1 뉴클레오타이드의 위치 및 아이덴티티를 확인하는 단계; (5) 제1 집단의 각각의 시퀀싱 프로브의 각각의 제1 부착 영역을, 검출가능한 표지가 없는 복수 개의 제1 하이브리드화 핵산 분자와 접촉시켜 검출가능한 표지를 가지는 또는 리포터 복합체의 제1 상보적인 핵산 분자를 결합 해제시키고, 각각의 제1 부착 영역에 검출가능한 표지가 없는 제1 하이브리드화 핵산 분자를 결합시키는 단계; (6) 각각 검출가능한 표지를 가지는 복수 개의 제2 상보적인 핵산 분자, 또는 각 복합체가 검출가능한 표지를 포함하는 복수 개의 제2 리포터 복합체의 복수 개의 제2 상보적인 핵산 분자를 제1 집단 중의 각 시퀀싱 프로브 중의 제2 부착 영역에 결합시키는 단계; (7) 검출가능한 표지(들)를 검출하는 단계; 및 (8) 제1 집단 중의 시퀀싱 프로브에 의해 하이브리드화된 고정화된 표적 핵산 중의 복수 개의 제2 뉴클레오타이드의 위치 및 아이덴티티를 확인하는 단계를 포함한다. 단계 (9)에서, 단계 (5) 내지 (8)은 고정화된 표적 핵산 중의 및 제1 집단 중의 각 시퀀싱 프로브의 표적 결합 도메인에 상응하는 각 뉴클레오타이드가 확인될 때까지 반복된다. 단계 (5) 및 (6)은 동시에 또는 순차적으로 진행될 수 있다. 이로써, 제1 시퀀싱 프로브 집단 중의 시퀀싱 프로브의 표적 결합 도메인에 의해 하이브리드화된 고정화된 표적 핵산의 영역에 대한 뉴클레오타이드의 선형 순서가 확인된다.

[0016] 실시양태에서, 복수 개의 프로브 집단(즉, 하나 초과)의 집단)이 사용되는 경우, 본 방법은 (10) 제1 집단의 각 시퀀싱 프로브를 핵산으로부터 탈하이브리드화시키는 단계; (11) 제1 집단의 각각의 탈하이브리드화된 시퀀싱 프로브를 제거하는 단계; (12) 본 발명의 적어도 제2 시퀀싱 프로브 집단을 하이브리드화시키는 단계로서, 제2 집단의 각 시퀀싱 프로브를 거의 동일한 조건하에서 고정화된 표적 핵산으로부터 탈하이브리드화시키고 제1 집단의 시퀀싱 프로브와 상이한 조건하에서 고정화된 표적 핵산으로부터 탈하이브리드화시키는 것인 단계; (13) 각각 검출가능한 표지를 가지는 복수 개의 제1 상보적인 핵산 분자, 또는 각 복합체가 검출가능한 표지를 포함하는 복수 개의 제1 리포터 복합체의 복수 개의 제1 상보적인 핵산 분자를 제2 집단 중의 각 시퀀싱 프로브 중의 제1 부착 영역에 결합시키는 단계; (14) 검출가능한 표지(들)를 검출하는 단계; (15) 제2 집단 중의 시퀀싱 프로브에 의해 하이브리드화된 고정화된 표적 핵산 중의 복수 개의 제1 뉴클레오타이드의 위치 및 아이덴티티를 확인하는 단계; (16) 제2 집단의 각각의 시퀀싱 프로브의 각각의 제1 부착 영역을, 검출가능한 표지가 없는 복수 개의 제1 하이브리드화 핵산 분자와 접촉시켜 (검출가능한 표지를 가지거나 또는 리포터 복합체로부터의) 제1 상보적인 핵산 분자를 결합 해제시키고, 각각의 제1 부착 영역에 검출가능한 표지가 없는 제1 하이브리드화 핵산 분자를 결합시키는 단계; (17) 각각 검출가능한 표지를 가지는 복수 개의 제2 상보적인 핵산 분자, 또는 각 복합체가 검출가능한 표지를 포함하는 복수 개의 제2 리포터 복합체의 복수 개의 제2 상보적인 핵산 분자를 제2 집단 중의 각 시퀀싱 프로브 중의 제2 부착 영역에 결합시키는 단계; (18) 검출가능한 표지(들)를 검출하는 단계; (19) 제2 집단 중의 시퀀싱 프로브에 의해 하이브리드화된 고정화된 표적 핵산 중의 복수 개의 제2 뉴클레오타이드의 위치 및 아이덴티티를 확인하는 단계; 및 (20) 제2 시퀀싱 프로브 집단 중 시퀀싱 프로브의 표적 결합 도메인에 의해 하이브리드화된 고정화된 표적 핵산의 영역에 대한 뉴클레오타이드의 선형 순서가 확인될 때까지 단계 (16) 내지 (19)를 반복하는 단계를 추가로 포함한다. 단계 (16) 및 (17)은 동시에 또는 순차적으로 진행될 수 있다.

[0017] 제2 집단의 각 시퀀싱 프로브를, 제1 집단 중의 시퀀싱 프로브를 표적 핵산으로부터 탈하이브리드화시키기 위한 평균 조건과는 다른 조건(예컨대, 더 높은 온도, 더 높은 수준의 카오토평제, 더 높은 염 농도, 더 높은 유속, 및 상이한 pH)에서 고정화된 표적 핵산으로부터 탈하이브리드화될 수 있다.

[0018] 그러나, 2개 초과)의 프로브 집단이 사용되는 경우, 이때 순차적인 두 집단 중의 프로브는 상이한 조건에서 탈하이브리드화될 수 있고, 비순차적인 집단 중의 프로브는 유사한 조건에서 탈하이브리드화될 수 있다. 한 예로서,



제1 집단 및 제3 집단 중의 프로브는 유사한 조건에서 탈하이브리드화될 수 있다. 실시양태에서, 순차적인 프로브 집단은 엄격성이 점점 더 증가하는 조건에서 (예컨대, 더 높은 수준의 카오트로픽제, 염 농도, 및 온도) 탈하이브리드화된다. 미세유체 디바이스의 경우, 한 예로서 온도를 사용하여, 제1 프로브 집단이 제1 온도에서는 하이브리드화된 상태 그대로 유지될 수 있지만, 제1 온도보다 높은 제2 온도에서는 탈하이브리드화될 수 있다. 제2 프로브 집단은 제2 온도에서는 하이브리드화된 상태 그대로 유지될 수 있지만, 제2 온도보다 높은 제3 온도에서는 탈하이브리드화될 수 있다. 상기 예에서, 초기 프로브 집단에 대한 표적 핵산 위로 유동하는 (본 발명에 의해 요구되는 시약을 포함하는) 용액은 후속 프로브 집단에 대한 표적 핵산 위로 유동하는 용액보다 더 낮은 온도에 있다.

[0019] 일부 실시양태에서, 프로브 집단은 사용된 후, 프로브 집단은 표적 핵산으로부터 탈하이브리드화되고, 동일한 프로브 집단의 새 분취물이 사용된다. 예를 들어, 제1 프로브 집단이 하이브리드화되고, 검출되고, 탈하이브리드화된 후, 제1 프로브 집단의 후속 분취물이 하이브리드화된다. 대안적으로, 한 예로서, 제1 프로브 집단은 탈하이브리드화되고 제2 프로브 집단으로 대체될 수 있고; 일단 제2 집단이 검출되고 탈하이브리드화되고 나면, 제1 프로브 집단의 후속 분취물이 표적 핵산에 하이브리드화된다. 따라서, 후속 집단 중의 프로브가 앞서(미리) 시퀀싱된 표적 핵산의 영역에 하이브리드화될 수 있거나(이로써, 중복 및/또는 확인 서열 정보를 획득함), 또는 후속 집단 중의 프로브가 앞서 시퀀싱되지 않은 표적 핵산의 영역에 하이브리드화될 수 있다(이로써, 새로운 서열 정보를 획득함). 따라서, 선행 리드가 (어느 이유에서든) 만족스럽지 못할 때, 및/또는 시퀀싱 리드로부터 생성된 정렬의 정확도를 개선시키기 위해 프로브 집단을 재분취할 수 있다.

[0020] 유사한 조건하에서 하이브리드화 및 탈하이브리드화하는 프로브는 그의 표적 결합 도메인의 길이, GC 함량, 또는 반복 염기 빈도 및 이의 조합이 유사할 수 있다. Tm과 올리고뉴클레오타이드의 길이 사이의 관계는 예를 들어, 문헌[Sugimoto *et al.*, *Biochemistry*, 34, 11211-6]에 교시되어 있다.

[0021] 2개 초과인 프로브 집단이 사용되는 경우, 제1 및 제2 시퀀싱 프로브 집단에 대해 기술된 바와 같은 단계는 추가 프로브 집단(예컨대, 10 내지 100 내지 1,000개의 집단)으로 반복된다. 사용되는 프로브 집단의 개수는 표적 핵산의 크기, 각 집단내 독특한 프로브의 개수, 원하는 시퀀싱 프로브 사이의 중복 정도, 및 관심 영역에 대한 프로브의 강화 정도를 포함하나, 이에 제한되지 않는, 다양한 인자에 의존할 것이다.

[0022] 프로브 집단은 표적 핵산 중 관심의 대상이 되는 특정 영역, 예컨대, 돌연변이(예컨대, 점 돌연변이) 또는 SNP 대립유전자를 함유하는 영역에 대한 추가의 시퀀싱 프로브를 함유할 수 있다. 프로브 집단은 표적 핵산 중 더 적은 관심의 대상이 되는 특정 영역에 대한 더 적은 수의 시퀀싱 프로브를 함유할 수 있다.

[0023] 시퀀싱 프로브 집단은 별개의 더 작은 시퀀싱 프로브 풀(pool)로 구획화될 수 있다. 구획화는 시퀀싱 프로브 중의 표적 결합 도메인의 융점 예측치에, 및/또는 시퀀싱 프로브 중의 표적 결합 도메인의 서열 모티프에 기초할 수 있다. 구획화는 경험적으로 도출된 규칙에 기초할 수 있다. 상이한 시퀀싱 프로브 풀을 예컨대, 온도, 염 농도, 및/또는 완충제 함량에 기초하여, 상이한 반응 조건을 사용하여 표적 핵산과 반응시킬 수 있다. 구획화는 균일한 커버리지로 표적 핵산을 커버하도록 수행될 수 있다. 구획화는 공지된 커버리지 프로파일로 표적 핵산을 커버하도록 수행될 수 있다.

[0024] 시퀀싱 프로브 집단 중 표적 결합 도메인의 길이는 표적 핵산의 특정 영역에서 프로브의 커버리지를 증가시키기 위해 감소될 수 있다. 시퀀싱 프로브 집단 중 표적 결합 도메인의 길이는 표적 핵산의 특정 영역에서 프로브의 커버리지를 감소시키기 위해 예컨대, 시퀀싱 장치의 해상도 한계 이상으로 증가될 수 있다.

[0025] 대안적으로 또는 추가로, 집단 중 시퀀싱 프로브의 농도는 표적 핵산의 특정 영역에서 프로브의 커버리지를 증가시키기 위해 증가될 수 있다. 시퀀싱 프로브의 농도는 표적 핵산의 특정 영역에서 프로브의 커버리지를 감소시키기 위해 예컨대, 시퀀싱 장치의 해상도 한계 이상으로 감소될 수 있다.

[0026] 핵산을 시퀀싱하는 방법은 고정화된 표적 핵산의 각 영역에 대한 각각의 확인된 선형 순서의 뉴클레오타이드를 조립하여, 고정화된 표적 핵산에 대한 서열을 확인하는 단계를 추가로 포함한다. 조립 단계는, 각각의 확인된 선형 순서의 뉴클레오타이드를 배열하여 핵산 서열을 수득하도록 마이크로프로세서에 지시하는 실행가능한 프로그램이 그 안에 저장되어 있는 비일시적 컴퓨터 판독가능한 저장 매체를 사용한다. 조립은 "실시간"으로, 즉, 모든 데이터가 수집된 이후보다는 시퀀싱 프로브로부터 데이터가 수집되는 동안에 이루어질 수 있다.

[0027] 즉, 시퀀싱되는 표적 핵산은 약 4 내지 1,000,000개의 뉴클레오타이드 길이일 수 있다. 표적은 1,000,000개 초과인 뉴클레오타이드 길이인, 전체의, 무손상(intact) 염색체 또는 이의 단편을 포함할 수 있다.

- [0028] 본원에서는 본 발명의 방법을 수행하기 위한 장치를 제공한다.
- [0029] 본원에서는 본 발명의 시퀀싱 프로브를 포함하는, 본 발명의 방법을 수행하기 위한 키트를 제공한다. 실시양태에서, 본 키트는 포획 프로브를 통해 핵산을 고정화시킬 수 있는 기재, 본 발명의 복수 개의 시퀀싱 프로브, 적어도 하나의 포획 프로브, 검출가능한 표지를 가지는 적어도 하나의 상보적인 핵산 분자, 검출가능한 표지가 없는 적어도 하나의 상보적인 핵산 분자, 및 사용 설명서를 포함한다. 실시양태에서, 본 키트는 약 또는 적어도 4,096개의 독특한 시퀀싱 프로브를 포함한다. 4,096개는 각각의 가능한 육량체 조합을 포함하는데 필요한 독특한 프로브(즉, 바코드 도메인에 각각 6개의 부착 영역을 가지는 프로브)의 최소 개수이다. 여기서, 6개의 위치에 대하여 4개의 뉴클레오티드 옵션이 존재하는 바, "4,096":  $4^6$ 이 달성되는 것이다. 바코드 도메인에 4개의 부착 영역을 가지는 프로브 세트의 경우에는 단지 256개(즉,  $4^4$ 개)의 독특한 프로브가 요구될 것이다. 그의 표적 결합 도메인 중 8개의 뉴클레오티드를 가지는 프로브 세트의 경우에는,  $4^8$ 개(즉, 65,536개)의 독특한 프로브가 요구될 것이다. 그의 표적 결합 도메인 중 10개의 뉴클레오티드를 가지는 프로브 세트의 경우에는,  $4^{10}$ 개(즉, 1,048,576개)의 독특한 프로브가 요구될 것이다.
- [0030] 실시양태에서, 본 키트는 약 또는 적어도 24개의, 검출가능한 표지를 가지는 별개의 상보적인 핵산 분자 및 약 또는 적어도 24개의, 검출가능한 표지가 없는 별개의 하이브리드화 핵산 분자를 포함한다. 비제한적인 예로서, 상보적인 핵산은 서열 번호: 1 내지 24 중 하나의 서열을 가지는 부착 영역에 결합할 수 있다. 바코드 도메인에 포함될 수 있는 추가의 예시적인 서열은 서열 번호: 42 내지 서열 번호: 81에 열거되어 있다. 실제로, 뉴클레오티드 서열은 제한되지 않으며; 바람직하게는, 공지된 뉴클레오티드 서열과의 실질적인 상동성(예컨대, 50% 내지 99.9%)은 없으며; 이는 상보적인 핵산과 표적 핵산의 바람직하지 못한 하이브리드화를 피하는 데 도움을 준다.
- [0031] 상기 측면 및 실시양태 중 임의의 것은 임의의 다른 측면 및 실시양태와 조합될 수 있다.
- [0032] 달리 정의되지 않는 한, 본원에서 사용된 모든 기술 용어 및 과학 용어는 본 발명이 속하는 당업계의 숙련자가 통상 이해하는 것과 동일한 의미를 가진다. 본 명세서에서, 단수 형태는 또한 문맥상 달리 명백하게 명시되지 않는 한, 복수 형태를 포함하고; 예로서, "하나"("a," "an") 및 "그"라는 용어는 단수 또는 복수의 것으로 이해되고, "또는"이라는 용어는 포괄적인 것으로 이해하여야 한다. 예로서, "한 요소"란, 하나 이상의 요소를 의미한다. 본 명세서 전역에 걸쳐, "포함하는"이라는 단어, 또는 예컨대, "포함하다" 또는 "포함하는"이라는 파생어는 임의의 다른 요소, 정수 또는 단계, 또는 요소들, 정수들 또는 단계들의 군을 배제하는 것이 아니라, 언급된 요소, 정수 또는 단계, 또는 요소들, 정수들 또는 단계들의 군을 포함하는 것을 암시하는 것으로 이해해야 한다. 약이란, 언급된 값의 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05%, 또는 0.01% 내의 값인 것으로 이해할 것이다. 문맥상 달리 분명하게 되지 않는 한, 본원에서 제공하는 모든 수치 값은 "약"이라는 수식어로 한정된다.
- [0033] 본원에 기술된 것과 유사하거나, 또는 등가인 방법 및 물질이 본 발명의 실행 또는 시험에서 사용될 수는 있지만, 적합한 방법 및 물질을 하기에 기술한다. 본원에서 언급된 모든 공개 문헌, 특허 출원, 특허 및 다른 참고 문헌은 그 전문이 참조로 포함된다. 본원에서 인용된 참고 문헌이 청구하는 본 발명의 선행 기술인 것으로 인정되는 것은 아니다. 의견 상충하는 경우, 정의를 포함하는 본 명세서를 통해 조정이 이루어질 것이다. 추가로, 물질, 방법 및 예는 단지 예시적인 것이며, 제한하는 것으로 의도되지 않는다. 본 발명의 다른 특징 및 이점은 하기의 상세한 설명 및 특허청구범위로부터 자명해질 것이다.

### 도면의 간단한 설명

- [0034] 본 특허 또는 출원 파일은 칼라로 작성된 적어도 하나의 도면을 포함한다. 칼라 도면을 포함하는 본 특허 또는 특허 출원 공개 사본은 요청 및 필요한 수수료 납부시 관청으로부터 제공받게 될 것이다.
- 상기 특징 및 추가 특징은 첨부된 도면과 함께 해석될 때 하기 상세한 설명으로부터 더욱 명확하게 이해될 것이다.
- 도 1 내지 도 5는 본 발명의 예시적인 시퀀싱 프로브의 개략도를 보여주는 것이다.
- 도 6a 내지 도 6d는 본 발명의 시퀀싱 프로브의 변이체를 보여주는 개략도이다.
- 도 7은 본 발명의 시퀀싱 프로브의 표적 결합 도메인의 개략도를 보여주는 것이며; 상기 도메인은 범용 염기를 포함하는 0, 2, 또는 4개의 뉴클레오티드를 포함한다.



도 8a 내지 도 8e는 본 발명의 시퀀싱 방법의 단계를 도시한 것이다.

도 9a는 본 발명의 시퀀싱 방법의 초기 단계를 보여주는 것이다.

도 9b는 검출가능한 표지를 포함하는 리포터 복합체의 개략도를 보여주는 것이다.

도 9c는 각각이 검출가능한 표지를 포함하는 복수 개의 리포터 복합체를 보여주는 것이다.

도 9d 내지 9g는 도 9a에서 시작된 시퀀싱 방법의 추가 단계를 보여주는 것이다.

도 10은 도 9d 및 9e에 제시된 단계의 대안적 예시 및 그로부터 수득된 예시적인 데이터를 보여주는 것이다. 제시된 시퀀싱 프로브의 단편은 서열 번호: 82의 서열을 가진다.

도 11은 도 10에 제시된 방법의 변형을 도시한 것이다. 제시된 시퀀싱 프로브의 단편은 마찬가지로 서열 번호: 82의 서열을 가진다.

도 12는 본 발명의 방법을 도시한 것이다.

도 13은 본 발명의 시퀀싱 방법에서 요구되는 단계를 다른 시퀀싱 방법에 요구되는 단계와 비교하는 것이다.

도 14 및 도 15는 본 발명에 의해 수득가능한 실행 평가 측정을 예시한 것이다.

도 16은 본 발명과 다양한 다른 시퀀싱 방법/장치에 대한 시퀀싱 속도, 리드 개수, 및 임상적 유용성을 비교한 것이다.

도 17은 본 발명의 시퀀싱 방법의 낮은 원시 오류율을 입증하는 것이다. 제시된 주형 서열은 서열 번호: 83의 서열을 가진다.

도 18은 본 발명으로부터 수득가능한 시퀀싱 데이터를 다른 시퀀싱 방법과 비교한 것이다.

도 19는 본 발명의 시퀀싱 방법의 단일 염기 특이성을 입증하는 것이다. (상단에서부터 하단으로) 제시된 주형 및 프로브 서열은 서열 번호: 84 내지 서열 번호: 88의 서열을 가진다.

도 20a는 본 발명의 리포터 복합체의 다양한 디자인을 보여주는 것이다.

도 20b는 도 20a에 제시된 리포터 복합체로부터 수득된 형광 계수를 보여주는 것이다.

도 20c는 본 발명의 리포터 복합체를 구성하기 위한 예시적인 레시피를 보여주는 것이다.

도 21a는 "추가 핸들"을 포함하는 리포터 복합체의 디자인을 보여주는 것이다.

도 21b는 "추가 핸들"을 가지는 리포터 복합체로부터 수득된 형광 계수를 보여주는 것이다.

도 22a 및 도 22b는 본 발명의 리포터 복합체의 2가지 예시적인 디자인의 하이브리드화 동역학적 성질을 보여주는 것이다.

도 23은 도 8에서부터 도 12까지 제시된 것과 다른 방법에서 사용된 본 발명의 시퀀싱 프로브의 개략도를 보여주는 것이다.

도 24는 본 발명에서 유용한 소모품 시퀀싱 카드의 개략도를 보여주는 것이다.

도 25는 실시예 3에 기술된 바와 같이, 10 mer의 미스매치 검출을 보여주는 것이다. (상단에서부터 하단으로) 제시된 뉴클레오티드는 서열 번호: 89 내지 서열 번호: 99의 서열을 가진다.

도 26은 실시예 3에 기술된 바와 같이, 표적 결합 도메인의 크기에 따른 하이브리드화 능력을 보여주는 것이다. 리포터 농도가 매우 높고, 사전 정제가 수행되지 않았기 때문에 배경이 높다. (상단에서부터 하단으로) 제시된 뉴클레오티드는 서열 번호: 100 내지 서열 번호: 104의 서열을 가진다.

도 27은 단일 스코트 대 전장의 리포터 사이의 비교를 보여주는 것이다. 단일 스코트에 대한 결과는 하이브리드화 속도가 전장의 바코드에 대한 것보다 1,000x 더 크다는 것을 나타낸다(조건 100 nM 표적, 30분 하이브리드화).

### 발명을 실시하기 위한 구체적인 내용

본 발명의 상세한 설명

본 발명은 리드 길이가 길고, 오류율이 낮은 무효소, 무증폭, 및 무라이브러리 핵산 시퀀싱을 제공하는 시퀀싱

[0035]

[0036]

프로브, 방법, 키트, 및 장치를 제공한다.

[0037] 시퀀싱 프로브

[0038] 본 발명은 표적 결합 도메인 및 바코드 도메인을 포함하는 시퀀싱 프로브에 관한 것이다. 본 발명의 시퀀싱 프로브의 비제한적인 예는 도 1 내지 6에 제시되어 있다.

[0039] 도 1은 본 발명의 시퀀싱 프로브의 개략도를 보여주는 것이다. 본 예시적인 시퀀싱 프로브는, 각각이 (하나 이상의 부착 영역을 포함하는) 바코드 도메인 중의 위치에 상응하는 6개의 뉴클레오타이드의 표적 결합 도메인을 가진다. 제1 부착 영역이 제시되어 있는데; 이는 표적 결합 도메인 중의 제1 뉴클레오타이드가 결합하는 표적 핵산의 뉴클레오타이드에 상응한다. 바코드 도메인 상의 제3 위치가 제시되어 있다. 2개의 부착 영역을 포함하는 제5 위치가 제시되어 있다. 바코드 도메인 상의 각 위치는 다중 부착 영역을 가질 수 있다. 예를 들어, 한 위치는 1 내지 50개의 부착 영역을 가질 수 있다. (위치 1 내지 4 및 6에 대한 위치 5에 제시된 바와 같이) 바코드 도메인 중의 특정 위치는 다른 위치보다 더 많은 부착 영역을 가질 수 있고; 대안적으로, 바코드 도메인 중의 각 위치는 같은 개수의 부착 영역을 가진다(예컨대, 도 2, 3, 5 및 6 참조). 비록 도시되어 있지는 않지만, 각 부착 영역은 상보적인 핵산 분자(RNA 또는 DNA)에 가역적으로 결합할 수 있는 적어도 하나의(즉, 1 내지 50개, 예컨대, 10 내지 30개) 핵산 서열(들) 카피를 포함한다. 도 1에서, 부착 영역은 바코드 도메인을 구성하는 선형 폴리뉴클레오타이드 분자에 통합되어 있다.

[0040] 도 2는 본 발명의 시퀀싱 프로브의 개략도를 보여주는 것이다. 본 예시적인 시퀀싱 프로브는, 각각이 바코드 도메인 중의 부착 영역에 상응하는 6개의 뉴클레오타이드의 표적 결합 도메인을 가진다. 제1 부착 영역이 제시되어 있는데; 이는 표적 결합 도메인 중의 제1 뉴클레오타이드가 결합하는 표적 핵산의 뉴클레오타이드에 상응한다. 바코드 도메인의 일부 및 2개의 제4 부착 영역을 포함하는, 바코드 도메인 상의 제4 위치가 동그라미로 표시되어 있다. 2개의 제6 부착 영역이 제시되어 있다. 여기서, 각 위치는 2개의 부착 영역을 가지지만; 그러나, 바코드 도메인 상의 각 위치는 1개의 부착 영역 또는 다수의 부착 영역, 예컨대, 2 내지 50개의 부착 영역을 가질 수 있다. 비록 도시되어 있지는 않지만, 각 부착 영역은 상보적인 핵산 분자(RNA 또는 DNA)에 가역적으로 결합할 수 있는 적어도 하나의(즉, 1 내지 50개, 예컨대, 10 내지 30개) 핵산 서열(들) 카피를 포함한다. 도 2에서, 바코드 도메인은 부착 영역이 연결된 선형 폴리뉴클레오타이드 분자이고; 부착 영역은 폴리뉴클레오타이드 분자에 통합되어 있지 않다.

[0041] 도 3은 본 발명의 시퀀싱 프로브의 또 다른 개략도를 보여주는 것이다. 본 예시적인 시퀀싱 프로브는 4개의 뉴클레오타이드의 표적 결합 도메인을 가지는데, 이들 4개의 뉴클레오타이드는 바코드 도메인 중의 4개의 위치에 상응하는 것이다. 각 위치는 3개의 연결된 부착 영역을 갖는 것으로 도시되어 있다.

[0042] 도 4는 본 발명의 시퀀싱 프로브의 추가의 또 다른 개략도를 보여주는 것이다. 본 예시적인 시퀀싱 프로브는 10개의 뉴클레오타이드의 표적 결합 도메인을 가진다. 그러나, 오직 제1의 6개의 뉴클레오타이드만이 바코드 도메인 중의 6개의 위치에 상응한다. (" $n_1$  내지  $n_4$ "로 표시된) 제6 내지 제10 뉴클레오타이드는 표적 결합 도메인의 길이를 증가시켜 프로브가 하이브리드화하여 표적 핵산에 하이브리드화된 상태 그대로 유지될 수 있는 가능성에 영향을 주기 위해 첨가된 것이다. 실시양태에서, " $n$ " 뉴클레오타이드는 바코드 도메인 중의 위치에 상응하는 뉴클레오타이드 앞에 올 수 있다. 실시양태에서, " $n$ " 뉴클레오타이드는 바코드 도메인 중의 위치에 상응하는 뉴클레오타이드 뒤에 올 수 있다. 도 4에서, 4개의 " $n$ " 뉴클레오타이드가 도시되어 있지만; 그러나, 표적 결합 도메인은 4개 초과 및 " $n$ " 뉴클레오타이드를 포함할 수 있다. " $n$ " 뉴클레오타이드는 4개의 정규 염기 중 임의의 것과 염기쌍을 형성할 수 있는 범용 염기(예컨대, 이노신, 2'-데옥시이노신(히포크산틴 데옥시뉴클레오타이드) 유도체, 니트로인돌, 니트로아줄 유사체, 및 소수성 방향족 비수소 결합 염기)를 가질 수 있다.

[0043] 본 발명의 또 다른 시퀀싱 프로브가 도 5에 도시되어 있다. 여기서, " $n$ " 뉴클레오타이드가 바코드 도메인 중의 위치에 상응하는 뉴클레오타이드 앞에 그리고 뒤에 있다. 도시된 예시적인 시퀀싱 프로브는 10개의 뉴클레오타이드의 표적 결합 도메인을 가진다. 그러나, 표적 결합 도메인 중 제3 내지 제8 뉴클레오타이드만이 바코드 도메인 중의 6개의 위치(제1 내지 제6)에 상응한다. (" $n_1$  내지  $n_4$ "로 표시된) 제1, 제2, 제9, 및 제10 뉴클레오타이드는 표적 결합 도메인의 길이를 증가시키기 위해 첨가된 것이다. 도 5에서, 4개의 " $n$ " 뉴클레오타이드가 도시되어 있지만; 그러나, 표적 결합 도메인은 4개 초과 또는 4개 미만의 " $n$ " 뉴클레오타이드를 포함할 수 있다.

[0044] 도 6a 내지 도 6d는 도 1의 시퀀싱 프로브의 변이체를 보여주는 것이다. 도 6a에서, 표적 결합 도메인 중의 뉴클레오타이드의 선형 순서 및 바코드 도메인 중의 부착 영역의 선형 순서는 (도면과 관련하여) 좌측에서 우측으로 진행된다. 도 6b에서, 표적 결합 도메인 중의 뉴클레오타이드의 선형 순서 및 바코드 도메인 중의 부착 영역의 선

형 순서는 (도면과 관련하여) 우측에서 좌측으로 진행된다. 도 6c에서, 표적 결합 도메인 중의 뉴클레오티드의 선형 순서는 바코드 도메인 중의 부착 영역의 선형 순서에 대해 반대이다. 본 발명의 임의의 프로브에서, 표적 결합 도메인 중의 각 뉴클레오티드가 바코드 도메인 중의 부착 도메인 또는 부착 도메인들에 상응하도록 프로브가 디자인되는 한, 표적 결합 도메인의 뉴클레오티드 및 바코드 도메인 중의 부착 영역의 엄격한 순서는 결여되어 있을 수 있고; 엄격한 순서의 결여는 도 6d에 도시되어 있다. 본 발명의 임의의 프로브(예컨대, 도 1 내지 5에 예시된 것)는 도 6에 도시된 바와 같은 뉴클레오티드 및 부착 영역의 배치를 가질 수 있다.

[0045] 표적 결합 도메인은 적어도 4개의 뉴클레오티드, 예컨대, 적어도 4, 5, 6, 7, 8, 9, 10, 11, 12개 이상의 뉴클레오티드를 가질 수 있다. 표적 결합 도메인은 바람직하게는 폴리뉴클레오티드이다. 표적 결합 도메인은 표적 핵산에 결합할 수 있다.

[0046] 프로브는 합성 백본에 작동가능하게 연결된 표적 결합 도메인의 다중 카피를 포함할 수 있다.

[0047] 프로브는 하이브리드화 및 탈하이브리드화의 가능성 및 그의 발생 속도를 제어하도록 디자인될 수 있다. 일반적으로, 프로브의  $T_m$ 이 낮을수록, 프로브가 표적 핵산에 및/또는 표적 핵산으로부터 탈하이브리드화하게 되는 속도는 더 빨라지고, 가능성은 더 커질 것이다. 따라서,  $T_m$ 이 더 낮은 프로브를 사용하는 것이 표적 핵산에 결합하는 프로브의 개수를 감소시키게 될 것이다.

[0048] 표적 결합 도메인의 길이가 부분적으로는 프로브 하이브리드화 가능성 및 표적 핵산에 하이브리드화된 상태 유지에 영향을 미친다. 일반적으로, 표적 결합 도메인 길이가 길수록(뉴클레오티드 개수가 더 많을수록), 표적 뉴클레오티드 중에 상보적인 서열이 존재할 가능성은 더 작아진다. 반대로, 표적 결합 도메인 길이가 짧을수록, 표적 뉴클레오티드 중에 상보적인 서열이 존재할 가능성은 더 커진다. 예를 들어, 6-mer 서열이 표적 핵산 중에 위치할 가능성은 1/4096인데 반해, 4-mer 서열이 표적 핵산 중에 위치할 가능성은 1/256이다. 결과적으로, 길이가 더 긴 프로브 집합물과 비교하였을 때, 가능하게는 길이가 더 짧은 프로브 집합물이 주어진 핵산 스트레치에 대하여 더 많은 위치에서 결합하게 될 것이다.

[0049] 도 7은 10-mer 표적 결합 도메인을 보여주는 것이다. 일부 실시양태에서, 표적 결합 도메인은 4개의 정규 뉴클레오티드(A, G, C, 및 T) 중 임의의 것과 염기쌍을 형성하는 ("U<sub>b</sub>"로 식별 표시되어 있는) 4개의 범용 염기를 포함한다. 실시양태에서, 표적 결합 도메인은 1 내지 6개의 (예컨대, 2 및 4개의) 범용 염기를 포함한다. 표적 결합 도메인은 범용 뉴클레오티드를 포함하지 않을 수 있다. 도 7에는 표적 결합 도메인 중 6개의 특이적인 뉴클레오티드를 가지는 "완전한(complete)" 프로브 집단은 4096개의 독특한 프로브를 요구하게 될 것이며, 10개의 특이적인 뉴클레오티드를 가지는 "완전한" 프로브 집단은 ~100만개의 독특한 프로브를 요구하게 될 것이라고 제시되어 있다.

[0050] 상황에 따라서는 주어진 핵산 스트레치 중의 리드 개수를 증가시켜 표적 핵산 또는 표적 핵산의 부분, 특히, 특별한 관심의 대상이 되는 부분의 커버리지를 강화시키기 위해서는, 예컨대, 돌연변이 또는 SNP 대립유전자를 검출할 경우에는 길이가 더 짧은 표적 결합 도메인을 가지는 프로브를 가지는 것이 바람직하다.

[0051] 그러나, 영역 중 너무 많은 프로브는 이들 검출가능한 표지의 중복을 일으킴으로써 가까이 있는 두 프로브의 해상도를 방해하는 경우가 존재하는 바, 이에 표적 핵산에 결합하는 프로브 개수는 더 적은 것이 바람직할 수 있다. 이는 하기와 같이 설명된다. 한 뉴클레오티드의 길이가 0.34 nm라고 가정하고, 시퀀싱 장치의 측면 (x-y) 공간 해상도가 약 200 nm라고 가정할 때, 시퀀싱 장치의 해상도 한계는 약 588개의 염기쌍(즉, 1개의 뉴클레오티드/0.34 nm x 200 nm)이다. 다시 말해, 두 프로브가 서로 약 588개의 염기쌍 범위 이내일 때, 상기 언급된 시퀀싱 장치는 표적 핵산에 하이브리드화된 두 프로브로부터 신호를 분석(분해)하지 못할 수 있다. 따라서, 두 프로브는 시퀀싱 장치의 해상도에 따라 이들 검출가능한 표지가 별개의 "스팟"으로 분석될 수 있기 이전에 대략 600 bp만큼 이격될 필요가 있을 것이다. 따라서, 최적으로 이격되어 있을 때, 600 bp의 표적 핵산당 단일 프로브가 존재하여야 한다. 표적 핵산의 분석가능한 영역 내부에 하이브리드화하는 프로브의 개수를 모니터링, 제한 및 잠재적으로는 데콘볼루션하고, 이에 따라 프로브 집단을 디자인하는 데 (예컨대, 형광 강도 값 및 파장 의존비를 이용하는) 다양한 소프트웨어 접근법이 사용될 수 있다. 또한, 더 많은 이산 신호를 제공하는 검출가능한 표지(예컨대, 형광 표지)가 선택될 수 있다. 추가로, 문헌 (예컨대, Small and Parthasarthy: "Superresolution localization methods." *Annu. Rev. Phys. Chem.*, 2014; 65:107-25)의 방법은 구조화된 조명 및 시퀀싱 현미경의 해상도 한계를 최대 수십 나노미터까지 감소시키는 다양한 초해상도 접근법을 기술한다. 더 높은 해상도의 시퀀싱 장치를 사용하면 길이가 더 짧은 표적 결합 도메인을 갖는 프로브를 사용할 수 있게 된다.

- [0052] 상기 언급한 바와 같이, 프로브의 T<sub>m</sub>을 디자인하는 것이 표적 핵산에 하이브리드화되는 프로브의 개수에 영향을 줄 수 있다. 대안적으로 또는 추가로, 집단 중 시퀀싱 프로브의 농도는 표적 핵산의 특정 영역 중의 프로브의 커버리지를 증가시키기 위해 증가될 수 있다. 시퀀싱 프로브의 농도는 표적 핵산의 특정 영역 중의 프로브의 커버리지를 감소시키기 위해, 예컨대, 시퀀싱 장치의 해상도 한계 이상으로 감소될 수 있다.
- [0053] "표적 핵산"이라는 용어는 그의 서열이 본 발명의 프로브, 방법, 및 장치에 의해 결정되는 핵산 분자(DNA, RNA, 또는 PNA)를 의미한다. 일반적으로, "표적 핵산", "핵산 분자", "핵산 서열", "핵산", "핵산 단편", "올리고뉴클레오타이드" 및 "폴리뉴클레오타이드"는 상호교환적으로 사용되고, 이는 데옥시리보뉴클레오타이드 또는 리보뉴클레오타이드, 또는 이의 유사체인, 다양한 길이일 수 있는 중합체 형태의 뉴클레오타이드를 포함하나, 이에 제한되지 않는 것으로 의도된다. 핵산의 비제한적인 예로는 유전자, 유전자 단편, 엑손, 인트론, 유전자간 DNA(제한 없이, 이염색질 DNA 포함), 메신저 RNA(mRNA), 전달 RNA, 리보솜 RNA, 리보자임, 짧은 간섭 RNA(siRNA), 비코딩 RNA(ncRNA: non-coding RNA), cDNA, 재조합 폴리뉴클레오타이드, 분지형 폴리뉴클레오타이드, 플라스미드, 벡터, 서열의 단리된 DNA, 서열의 단리된 RNA, 핵산 프로브, 및 프라이머를 포함한다.
- [0054] 본 방법은 샘플, 예컨대, 유기체로부터의 샘플로부터 수득된 핵산 분자를, 바람직하게는, 전환(또는 증폭) 단계 없이, 직접 시퀀싱한다. 한 예로서, RNA 기반 시퀀싱의 경우, 본 방법은 서열 수득 이전에 RNA 분자를 DNA 분자로(즉, cDNA의 합성을 통해) 전환시킬 필요가 없다. 증폭 또는 전환이 필요하지 않는 바, 본 발명에서 시퀀싱되는 핵산은, 핵산이 샘플 중에 존재할 때, 또는 핵산이 샘플로부터 수득되었을 때, 핵산 중에 존재하는 임의의 독특한 염기 및/또는 후생적 마커를 보유하게 될 것이다. 이러한 독특한 염기 및/또는 후생적 마커는 당업계에 공지된 시퀀싱 방법에서는 손실된다.
- [0055] 표적 핵산은 임의의 샘플 또는 핵산 공급원으로부터, 예컨대, 임의의 세포, 조직, 또는 유기체, 시험관내, 화학적 합성기 등으로부터 수득될 수 있다. 표적 핵산은 당업계에 알려져 있는 임의의 방법에 의해 수득될 수 있다. 실시양태에서, 핵산은 임상적 피험체의 혈액 샘플로부터 수득된다. 핵산은 당업계에 널리 공지된 방법 및 키트를 사용하여 공급원 또는 샘플로부터 추출, 단리, 또는 정제될 수 있다.
- [0056] 표적 핵산을 포함하는 핵산 분자는 당업계에 공지된 임의 수단에 의해 단편화될 수 있다. 바람직하게, 단편화는 효소적 또는 기계적 수단에 의해 수행된다. 기계적 수단은 초음파 처리 또는 물리적 전단일 수 있다. 효소적 수단은 뉴클레아제(예컨대, 데옥시리보뉴클레아제 I(DN아제 I)) 또는 하나 이상의 제한 엔도뉴클레아제로 분해함으로써 수행될 수 있다.
- [0057] 표적 핵산을 포함하는 핵산 분자가 무손상 염색체일 때, 염색체를 단편화하는 것을 막는 단계를 취하여야 한다.
- [0058] 표적 핵산은 당업계에 널리 공지되어 있는 바와 같이, 변형된 뉴클레오타이드를 포함하는, 천연 또는 비천연 뉴클레오타이드를 포함할 수 있다.
- [0059] (표적 결합 도메인, 바코드 도메인, 및 임의의 선택적인 도메인을 포함하는) 본 발명의 프로브의 총 길이는 약 20 나노미터 내지 약 50 나노미터일 수 있다. 프로브의 백본은 약 120개의 뉴클레오타이드를 포함하는 폴리뉴클레오타이드 분자일 수 있다.
- [0060] 바코드 도메인은 합성 백본을 포함한다. 합성 백본 및 표적 결합 도메인은 작동가능하게 연결되어 있고, 예컨대, 공유적으로 부착되어 있거나, 또는 링커를 통해 부착되어 있다. 합성 백본은 임의의 물질, 예컨대, 당류, 폴리뉴클레오타이드, 중합체, 플라스틱, 섬유, 펩티드, 펩티드 핵산, 또는 폴리펩티드를 포함할 수 있다. 바람직하게, 합성 백본은 강성이다. 실시양태에서, 백본은 6개의 DNA 이중 나선으로 이루어진 "DNA 오리가미"를 포함한다(예컨대, 문헌 [Lin *et al*, "Submicrometre geometrically encoded fluorescent barcodes self-assembled from DNA," *Nature Chemistry*; 2012 Oct; 4(10): 832-9] 참조). 바코드는 DNA 오리가미 타일로 제조될 수 있다(문헌 [Jungmann *et al*, "Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT", *Nature Methods*, Vol. 11, No. 3, 2014]).
- [0061] 바코드 도메인은 복수 개의 위치, 예컨대, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 위치를 포함한다. 위치 개수는 표적 결합 도메인 중의 뉴클레오타이드의 개수보다 적거나, 그와 같거나, 또는 그보다 많을 수 있다. 백본 도메인 중의 위치 개수보다 표적 결합 도메인 중 추가의 뉴클레오타이드, 예컨대, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상 뉴클레오타이드를 포함하는 것이 바람직할 수 있다. 상기 기술된 바와 같이, 4개 이상의 위치에 대한 충분한 공간이 있는 한, 바코드 도메인 길이에 대해서는 제한되지 않는다.
- [0062] 바코드 도메인 중의 각 위치는 표적 결합 도메인 중의 뉴클레오타이드에 상응하고, 따라서, 표적 핵산 중의 뉴클



레오티드에 상응한다. 예로서, 바코드 도메인 중의 제1 위치는 표적 결합 도메인 중의 제1 뉴클레오티드에 상응하고, 바코드 도메인 중의 제6 위치는 표적 결합 도메인 중의 제6 뉴클레오티드에 상응한다.

[0063] 바코드 도메인 중의 각 위치는 적어도 하나의 부착 영역, 예컨대, 1 내지 50개, 또는 그 초과와 부착 영역을 포함한다. 바코드 도메인 중의 특정 위치는 다른 위치보다 더 많은 부착 영역을 포함할 수 있고(예컨대, 제1 위치는 3개의 부착 영역을 가질 수 있는 반면, 제2 위치는 2개의 부착 위치를 가질 수 있고); 대안으로, 바코드 도메인 중의 각 위치는 같은 개수의 부착 영역을 가진다. 각 부착 영역은 상보적인 핵산 분자(예컨대, DNA 또는 RNA)가 가역적으로 결합할 수 있는 적어도 하나의(즉, 1 내지 50개, 예컨대, 10 내지 30개) 핵산 서열(들) 카피를 포함한다. 예에서, 제1 부착 영역 중의 핵산 서열이, 표적 결합 도메인의 제1 뉴클레오티드가 결합하는 표적 핵산 중의 제1 뉴클레오티드의 위치 및 아이덴티티를 결정한다. 각 부착 영역은 부착 영역이 합성 백본으로부터 분지되도록 합성 백본 중 변형된 단량체(예컨대, 변형된 뉴클레오티드)에 연결될 수 있다. 실시양태에서, 부착 영역은 폴리뉴클레오티드 백본에 통합되어 있고; 다시 말해, 백본은 단일 폴리뉴클레오티드이고, 부착 영역은 단일 폴리뉴클레오티드 서열의 일부이다. 실시양태에서, "바코드 도메인" 및 "합성 백본"이라는 용어는 동의어이다.

[0064] 부착 영역 중의 핵산 서열이 시퀀싱 프로브의 표적 결합 도메인 중의 뉴클레오티드가 결합하는 표적 핵산 중의 뉴클레오티드의 위치 및 아이덴티티를 확인시켜 준다. 프로브에서, 각 부착 영역은 독특한 전체 서열을 가질 것이다. 실제로, 바코드 도메인 상의 각 위치는, 즉, 아데닌, 티민/우라실, 시토신, 및 구아닌 중 하나에 특이적인 4개의 뉴클레오티드 중 하나를 코딩하는 핵산 서열을 포함하는 부착 영역을 가질 수 있다. 또한, 제1 위치의 (및 예를 들어, 시토신을 코딩하는) 부착 영역은 제2 위치의 (및 예를 들어, 시토신을 코딩하는) 부착 영역과는 다른 핵산 서열을 포함할 것이다. 따라서, 티민을 코딩하는 제1 위치의 부착 영역 중의 핵산 서열에는, 표적 결합 도메인의 제1 뉴클레오티드에 상응하는 표적 핵산 중의 아데닌을 동정하는 상보적인 핵산 분자의 결합이 존재하지 않을 것이다. 또한, 제2 위치의 부착 영역에는, 표적 결합 도메인의 제1 뉴클레오티드에 상응하는 표적 핵산 중의 아데닌을 동정하는 상보적인 핵산 분자의 결합이 존재하지 않을 것이다.

[0065] 바코드 도메인 상의 각 위치는 하나 이상의(최대 50개, 바람직하게 10 내지 30개) 부착 영역을 포함할 수 있고; 따라서, 각 부착 영역은 하나 이상의(최대 50개, 바람직하게 10 내지 30개) 상보적인 핵산 분자에 결합할 수 있다. 예로서, 도 1의 프로브는 2개의 부착 영역을 포함하는 제5 위치를 가지고, 도 2의 프로브는 6개의 부착 영역을 가지는 제2 위치를 가진다. 실시양태에서, 한 위치에서의 부착 영역의 핵산 서열은 동일하고; 따라서, 이들 부착 영역에 결합하는 상보적인 핵산 분자는 동일하다. 대안적 실시양태에서, 한 위치에서의 부착 영역의 핵산 서열은 동일하지 않고, 따라서, 이들 부착 영역에 결합하는 상보적인 핵산 분자는 동일하지 않고, 예컨대, 각각은 상이한 핵산 서열 및/또는 검출가능한 표지를 포함한다. 따라서, 대안적 실시양태에서, 함께 부착 영역에 부착된 동일하지 않은 핵산 분자들(예컨대, 이들의 검출가능한 표지)의 조합은 표적 핵산 중의 뉴클레오티드를 확인하기 위한 코드를 제공한다.

[0066] 하기 표 1은 단지 예시적인 목적으로, 그의 바코드 도메인 중 최대 6개의 위치를 가지는 시퀀싱 프로브에 대한 부착 영역 및 그에 결합하는 상보적인 핵산 상의 검출가능한 표지에 대한 예시적인 서열을 제공한다.

표 1

표 1:

표적 결합 도메인 중의 뉴클레오티드 /바코드 도메인 중의 위치	뉴클레오티드	부착 영역 중의 핵산 서열 (5' → 3')	상보적인 핵산의 검출가능한 표지	서열 번호
1	A	ATACATCTAG	GFP	1
1	G	GATCTACATA	RFP	2
1	C	TTAGGTAAAG	CFP	3
1	U/T	TCTTCATTAC	YFP	4
2	A	ATGAATCTAC	GFP	5
2	G	TCAATGTATG	RFP	6
2	C	AATTGAGTAC	CFP	7
2	U/T	ATGTTAATGG	YFP	8
3	A	AATTAGGATG	GFP	9
3	G	ATAATGGATC	RFP	10
3	C	TAATAAGGTG	CFP	11
3	U/T	TAGTTAGAGC	YFP	12
4	A	ATAGAGAAGG	GFP	13
4	G	TTGATGATAC	RFP	14
4	C	ATAGTGATTC	CFP	15
4	U/T	TATAACGATG	YFP	16
5	A	TTAAGTTTAG	GFP	17
5	G	ATACGTTATG	RFP	18
5	C	TGTACTATAG	CFP	19
5	U/T	TTAACAAAGTG	YFP	20
6	A	AACTATGTAC	GFP	21
6	G	TAACATGAC	RFP	22
6	C	ACTAATGTTT	CFP	23
6	U/T	TCATTGAATG	YFP	24

[0067]

[0068]

표 1에서 알 수 있는 바와 같이, 제1 부착 영역의 핵산 서열은 서열 번호: 1 내지 서열 번호: 4 중 하나일 수 있고, 제2 부착 영역의 핵산 서열은 서열 번호: 5 내지 서열 번호: 8 중 하나일 수 있다. 표적 핵산 중의 제1 뉴클레오티드가 아데닌일 때, 제1 부착 영역의 핵산 서열은 서열 번호: 1의 서열을 가지게 될 것이며; 표적 핵산 중의 제2 뉴클레오티드가 아데닌일 때, 제2 부착 영역의 핵산 서열은 서열 번호: 5의 서열을 가지게 될 것이다.

[0069]

실시양태에서, 상보적인 핵산 분자에는 검출가능한 표지가 결합할 수 있다. 대안적 실시양태에서, 상보적인 핵산은 검출가능한 표지를 포함하는 리포터 복합체와 회합된다.

[0070]

상보적인 핵산의 뉴클레오티드 서열은 제한되지 않으며; 바람직하게, 이는 공지된 뉴클레오티드 서열과 실질적인 상동성(예컨대, 50% 내지 99.9%)을 가지지 않고; 이는 상보적인 핵산 및 표적 핵산의 바람직하지 못한 하이브리드화를 막는 데 도움이 된다.

[0071]

본 발명에서 유용한 리포터 복합체의 예는 도 9b에 도시되어 있다. 본 예에서, 상보적인 핵산은 1차 핵산 분자에 연결되고, 이는 차례로 복수 개의 2차 핵산 분자에 하이브리드화되고, 이들 각각은 차례로, 하나 이상의 검출가능한 표지가 그에 부착되어 있는 복수 개의 3차 핵산 분자에 하이브리드화된다.

[0072]

실시양태에서, 1차 핵산 분자는 약 90개의 뉴클레오티드를 포함할 수 있다. 2차 핵산 분자는 약 87개의 뉴클레오티드를 포함할 수 있다. 3차 핵산 분자는 약 15개의 뉴클레오티드를 포함할 수 있다.

[0073]

도 9c는 예시적인 리포터 복합체 집단을 보여주는 것이다. 프로브의 부착 영역 1에 하이브리드화하는 4개의 복합체가 도 9c의 좌측 상단 패널에 포함되어 있다. 프로브의 표적 결합 도메인의 뉴클레오티드 위치 1에 존재할 수 있는 각각의 가능한 뉴클레오티드에 대하여 한 유형의 리포터 복합체가 존재한다. 여기서, 본 발명의 시퀀싱 방법을 수행하는 동안, 프로브의 리포터 도메인의 위치 1이 "청색의" 검출가능한 표지를 가지는 리포터 복합체에 의해 결합된다면, 이때, 표적 결합 도메인 중의 제1 뉴클레오티드는 아데닌인 것으로 확인된다. 대안적으로, 위치 1이 "녹색의" 검출가능한 표지를 가지는 리포터 복합체에 의해 결합된다면, 이때, 표적 결합 도메인 중의 제1 뉴클레오티드는 티민인 것으로 확인된다.

[0074]

리포터 복합체는 다양한 디자인을 가질 수 있다. 예를 들어, 1차 핵산 분자는 적어도 하나의(예컨대, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의) 2차 핵산 분자에 하이브리드화될 수 있다. 각 2차 핵산 분자는 적어도 하나의(예컨대, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의) 3차 핵산 분자에 하이브리드화될 수 있다. 예시적인 리포터 복합체는 도 20a에 도시되어 있다. 여기서, "4x3" 리포터 복합체는, 각각이 (각각 검출가능한 표지를 포함하는) 3개의 3차 핵산 분자에 하이브리드화되는 4개의 2차 핵산 분자에 하이브리드화되는 (상보적인 핵산 분자에 연결된) 하나의 1차 핵산 분자를 가진다. 본 도면에서, 복합체의 각 상보적인 핵산은 12개의 뉴클레오티드 길이("12

개의 염기")이지만; 그러나, 상보적인 핵산의 길이에는 제한이 없으며, 12개 미만 또는 12개 초과 뉴클레오티드 길이일 수 있다. 우측 하단의 복합체는 그의 상보적인 핵산과 그의 1차 핵산 분자 사이에 스페이서 영역을 포함한다. 스페이서는 20 내지 40개의 뉴클레오티드 길이인 것으로 확인되지만; 그러나, 스페이서의 길이에는 제한이 없으며, 20개의 뉴클레오티드보다 짧거나, 또는 40개의 뉴클레오티드보다 길 수 있다.

[0075] 도 20b는 도 20a에 도시된 4개의 예시적인 리포터 복합체로부터 수득된 가변적인 평균(형광) 계수를 보여주는 것이다. 도 20b에서, 10 pM의 비오틴화된 표적 주형을 스트렙타비딘으로 코팅된 플로우 셀(flow-cell) 표면 상에 부착시키고, 10 nM의 리포터 복합체를 플로우 셀 상에 유동시키고; 1분 동안 인큐베이션시킨 후, 플로우 셀을 세척하고, 플로우 셀을 영상화하고, 형광 피치를 계수하였다.

[0076] 실시양태에서, 리포터 복합체는 "미리 구성(구축)된다". 즉, 복합체 중의 각 폴리뉴클레오티드는 복합체를 프로브와 접촉시키기 이전에 하이브리드화된다. 5개의 예시적인 리포터 복합체를 미리 구성하기 위한 예시적인 레시피가 도 20c에 도시되어 있다.

[0077] 도 21a는, 2차 핵산 분자가 3차 핵산 분자에 하이브리드화되지 않고 1차 핵산 분자로부터 원거리에 위치하는 "추가 핸들"을 가지는 대안적 리포터 복합체를 보여주는 것이다. 본 도면에서, 각 "추가 핸들"은 12개의 뉴클레오티드 길이("12 mer")이지만; 그러나, 그의 길이는 제한되지 않으며, 12개 미만 또는 12개 초과 뉴클레오티드 길이일 수 있다. 실시양태에서, "추가 핸들"은 각각 상보적인 핵산의 뉴클레오티드 서열을 포함하고; 따라서, 리포터 복합체가 "추가 핸들"을 포함할 때, 리포터 복합체는 리포터 복합체의 상보적인 핵산을 통해 또는 "추가 핸들"을 통해 시퀀싱 프로브에 하이브리드화할 수 있다. 따라서, 리포터 복합체가 시퀀싱 프로브에 결합할 가능성은 증가된다. "추가 핸들" 디자인은 또한 하이브리드화 동역학적 성질을 개선시킬 수 있다. 이론으로 제한하지 않으면서, "추가 핸들"은 본질적으로 리포터 복합체의 상보적인 핵산의 유효 농도를 증가시킨다.

[0078] 도 21b는 도 20b에 기술된 절차를 사용하여 "추가 핸들"을 가지는 5개의 예시적인 리포터 복합체로부터 수득된 가변적인 평균(형광) 계수를 보여주는 것이다.

[0079] 도 22a 및 22b는 두 예시적인 리포터 복합체에 대한 하이브리드화 동역학적 성질 및 형광 강도를 보여주는 것이다. 약 5분까지, 총 계수는 안정 수준에 이르기 시작하고, 이는 첨가된 대부분의 리포터 복합체가 이용가능한 표적을 발견하였다는 것을 시사하는 것이다.

[0080] 검출가능한 모이어티, 표지 또는 리포터가 예컨대, 형광 모이어티, 비색 모이어티 등과 같은 검출가능한 모이어티의 직접적인 또는 간접적인 부착을 비롯한 다양한 방식으로 상보적인 핵산에 또는 3차 핵산 분자에 결합될 수 있다. 당업자는 핵산을 표지하는 것에 관한 참조 문헌을 참조할 수 있다. 형광 모이어티의 예로는 황색 형광 단백질(YFP), 녹색 형광 단백질(GFP), 시안 형광 단백질(CFP), 적색 형광 단백질(RFP), 움벨리페론, 플루오레세인, 플루오레세인 이소티오시아네이트, 로다민, 디클로로트리아지닐아민 플루오레세인, 시아닌, 단실클로라이드, 피코시아닌, 피코에리트린 등을 포함하나, 이에 제한되지 않는다. 형광성 표지 및 그의 뉴클레오티드 및/또는 올리고뉴클레오티드에의 부착은 문헌 [Haugland, Handbook of Fluorescent Probes and Research Chemicals, Ninth Edition (Molecular Probes, Inc., Eugene, 2002)]; [Keller and Manak, DNA Probes, 2nd Edition (Stockton Press, New York, 1993)]; [Eckstein, editor, Oligonucleotides and Analogues: A Practical Approach (IRL Press, Oxford, 1991)]; 및 [Wetmur, Critical Reviews in Biochemistry and Molecular Biology, 26:227-259 (1991)]을 비롯한 다수의 리뷰에 기술되어 있다. 본 발명에 적용가능한 특정 방법은 하기 참고 문헌 샘플에 개시되어 있다: 미국 특허 번호 4,757,141; 5,151,507; 및 5,091,519. 한 측면에서는, 하나 이상의 형광 염료가 예컨대, 미국 특허 번호 5,188,934(4,7-디클로로플루오레세인 염료); 5,366,860(스펙트럼으로 분석가능한 로다민 염료); 5,847,162(4,7-디클로로로다민 염료); 4,318,846(에테르 치환된 플루오레세인 염료); 5,800,996(에너지 전달 염료); Lee 등의 5,066,580(크산틴 염료); 5,688,648(에너지 전달 염료) 등에 개시되어 있는 바와 같이, 표지된 표적 서열에 대한 표지로서 사용된다. 표지는 또한 하기 특허 및 특허 공개: 미국 특허 번호 6,322,901; 6,576,291; 6,423,551; 6,251,303; 6,319,426; 6,426,513; 6,444,143; 5,990,479; 6,207,392; 2002/0045045; 및 2003/0017264에 개시되어 있는 바와 같이, 양자점을 이용하여 수행될 수 있다. 본원에서 사용되는 바, "형광 표지"라는 용어는 하나 이상의 분자의 형광 흡수 및/또는 방출 특성을 통해 정보를 전달하는 신호전달 모이어티를 포함한다. 상기 형광 특성으로는 형광 강도, 형광 수명, 방출 스펙트럼 특징, 에너지 전달 등을 포함한다.

[0081] 뉴클레오티드 및/또는 올리고뉴클레오티드 서열 내로 쉽게 도입되는 상업적으로 이용가능한 형광 뉴클레오티드 유사체로는 Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy5-dUTP(아머샴 바이오사이언시스(AmershamBiosciences: 미국 뉴저지주 피츠카타웨이)), 플루오레세인-12-dUTP, 테트라메틸로다민-6-dUTP, 텍사스 레드(TEXAS RED)<sup>TM</sup>-5-dUTP,

캐스케이드 블루(CASCADE BLUE)<sup>TM</sup>-7-dUTP, BODIPY TMFL-14-dUTP, BODIPY TMR-14-dUTP, BODIPY TMTR-14-dUTP, 로다민 그린<sup>TM</sup>-5-dUTP, 오레곤 그린(OREGON GREENR)<sup>TM</sup> 488-5-dUTP, 텍사스 레드<sup>TM</sup>-12-dUTP, BODIPY TM 630/650-14-dUTP, BODIPY TM 650/665-14-dUTP, 알렉사 플루오르(ALEXA FLUOR)<sup>TM</sup> 488-5-dUTP, 알렉사 플루오르<sup>TM</sup> 532-5-dUTP, 알렉사 플루오르<sup>TM</sup> 568-5-dUTP, 알렉사 플루오르<sup>TM</sup> 594-5-dUTP, 알렉사 플루오르<sup>TM</sup> 546-14-dUTP, 플루오레세인-12-UTP, 테트라메틸로다민-6-UTP, 텍사스 레드<sup>TM</sup>-5-UTP, mCherry, 캐스케이드 블루<sup>TM</sup>-7-UTP, BODIPY TM FL-14-UTP, BODIPY TMR-14-UTP, BODIPY TM TR-14-UTP, 로다민 그린<sup>TM</sup>-5-UTP, 알렉사 플루오르<sup>TM</sup> 488-5-UTP, 텍사스 플루오르<sup>TM</sup> 546-14-UTP (몰레큘러 프로브즈 인크.(Molecular Probes, Inc.: 미국 오레곤주 유진)) 등을 포함하나, 이에 제한되지 않는다. 대안적으로, 상기 형광단 및 본원에서 언급된 것은 예를 들어, 포스포로아미다이트 또는 NHS 화학법을 이용하여 올리고뉴클레오티드 합성 동안 첨가될 수 있다. 다른 형광단을 가지는 뉴클레오티드의 통상적인 합성에 관한 프로토콜은 당업계에 공지되어 있다(문헌 [Henegariu *et al.* (2000) Nature Biotechnol. 18:345] 참조). 2-아미노퓨린은 합성 동안 올리고뉴클레오티드 서열에 직접 도입될 수 있는 형광 염기이다. 핵산은 또한 선택적으로 삽입성 염료, 예컨대, DAPI, YOYO-1, 에티디움 브로마이드, 시아닌 염료 (예컨대, SYBR 그린) 등으로 염색될 수 있다.

[0082] 합성 후 부착을 위해 이용가능한 다른 형광단으로는 알렉사 플루오르<sup>TM</sup> 350, 알렉사 플루오르<sup>TM</sup> 405, 알렉사 플루오르<sup>TM</sup> 430, 알렉사 플루오르<sup>TM</sup> 532, 알렉사 플루오르<sup>TM</sup> 546, 알렉사 플루오르<sup>TM</sup> 568, 알렉사 플루오르<sup>TM</sup> 594, 알렉사 플루오르<sup>TM</sup> 647, BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, 캐스케이드 블루, 캐스케이드 옐로우(Cascade Yellow), 단실, 리사민 로다민 B, 마리나 블루(Marina Blue), 오레곤 그린 488, 오레곤 그린 514, 파시픽 블루(Pacific Blue), 파시픽 오렌지(Pacific Orange), 로다민 6G, 로다민 그린, 로다민 레드, 테트라메틸 로다민, 텍사스 레드(몰레큘러 프로브즈 인크.(미국 오레곤주 유진)로부터 이용가능), Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7(아머삼 바이오사이언시스: 미국 뉴저지주 피츠카타웨이) 등을 포함하나, 이에 제한되지 않는다. PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-텍사스 레드, APC-Cy7, PE-알렉사 염료(610, 647, 680), APC-알렉사 염료 등을 포함하나, 이에 제한되지 않는, FRET 탠덤 형광단 또한 사용될 수 있다.

[0083] 형광 표지된 뉴클레오티드 및/또는 올리고뉴클레오티드 서열로부터 신호를 증강시키는 데 금속성 은 또는 금 입자가 사용될 수 있다(문헌 [Lakowicz *et al.* (2003) BioTechniques 34:62]).

[0084] 올리고뉴클레오티드 서열에 적합한 다른 표지로는 플루오레세인(FAM, FITC), 디곡시게닌, 디니트로페놀(DNP), 단실, 비오틴, 브로모데옥시우리딘(BrdU), 헥사히스티딘(6xHis), 포스포-아미노산(예컨대, P-tyr, P-ser, P-thr) 등을 포함할 수 있다. 한 실시양태에서, 하기 합텐/항체 쌍: 비오틴/a-비오틴, 디곡시게닌/a-디곡시게닌, 디니트로페놀 (DNP)/a-DNP, 5-카복시플루오레세인(FAM)/a-FAM이 검출을 위해 사용되며, 여기서, 각 항체는 검출 가능한 표지로 유도체화된다.

[0085] 본원에 기술된 검출가능한 표지는 스펙트럼으로 분석가능한 것이다. 복수 개의 형광성 표지와 관련하여 "스펙트럼으로 분석가능하다"는 것은 표지의 형광 방출 밴드가 충분히 상이하고, 즉, 충분히 비중복성을 띠며, 이로써, 미국 특허 번호 4,230,558; 4,811,218 등에서, 또는 문헌 [Wheless *et al.*, pgs. 21-76, in Flow Cytometry: Instrumentation and Data Analysis (Academic Press, New York, 1985)]에서 기술된 시스템에 의해 예시된 바와 같이, 각 표지가 부착되는 분자 태그는 예컨대, 대역 통과 필터 및 광전자 증배관 등을 이용하는 광검출 시스템에 의해 각 표지에 의해 생성된 형광 신호에 기초하여 구별될 수 있다는 것을 의미한다. 한 측면에서, 스펙트럼으로 분석가능한 유기 염료, 예컨대, 플루오레세인, 로다민 등은 파장 방출 최대치가 적어도 20 nm만큼 이격되어 있고, 또 다른 측면에서는, 적어도 40 nm만큼 이격되어 있다는 것을 의미한다. 또 다른 측면에서, 킬레이팅된 란탄족 화합물, 양자점 등이 스펙트럼으로 분석가능하다는 것은 파장 방출 최대치가 적어도 10 nm만큼 이격되어 있고, 추가 측면에서, 적어도 15 nm만큼 이격되어 있다는 것을 의미한다.

## [0086] 시퀀싱 방법

[0087] 본 발명은 본 발명의 시퀀싱 프로브를 이용하여 핵산을 시퀀싱하는 방법에 관한 것이다. 방법의 예는 도 8 내지 12에 도시되어 있다.

[0088] 본 방법은 본 발명의 적어도 하나의 시퀀싱 프로브를 기재에 (예컨대, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 위치에) 고정화된 표적 핵산에 가역적으로 하이브리드화시키는 단계를 포함한다.

[0089] 기재는 당업계에 공지된 임의의 고체 지지체, 예컨대, 표적 핵산을 고정화시킬 수 있는 코팅된 슬라이드 및 미



세유체 디바이스일 수 있다. 특정 실시양태에서, 기재는 표면, 막, 비드, 다공성 물질, 전극 또는 어레이이다. 표적 핵산은 당업자에게 자명한 임의의 기재 상에 고정화될 수 있다.

- [0090] 실시양태에서, 표적 핵산은 표적 핵산의 부분에 상보적인 도메인을 포함하는 포획 프로브에 의해 결합된다. 상기 부분은 표적 핵산의 단부일 수 있거나, 또는 단부 쪽에 있지 않을 수도 있다.
- [0091] 예시적인 유용한 기재로는 리간드, 항원, 탄수화물, 핵산, 수용체, 렉틴 및 항체로 이루어진 군으로부터 선택되는 결합 모이어티를 포함하는 것을 포함한다. 포획 프로브는 기재의 결합 모이어티와 결합할 수 있는 결합 모이어티를 포함한다. 반응성 모이어티를 포함하는 것으로서 예시적인 유용한 기재로는 에폭시, 알데히드, 금, 히드라지드, 술폰히드릴, NHS-에스테르, 아민, 티올, 카복실레이트, 말레이미드, 히드록시메틸 포스포네, 이미도에스테르, 이소시아네이트, 히드록실, 펜타플루오로페닐-에스테르, 소랄렌, 피리딜 디설피드 또는 비닐 술폰, 폴리 에틸렌 글리콜 (PEG), 히드로겔, 또는 그의 혼합물을 포함하는 표면을 포함하나, 이에 제한되지 않는다. 상기 표면은 상업적 공급처로부터 입수할 수 있거나, 또는 표준 기법에 따라 제조될 수 있다. 반응성 모이어티를 포함하는 것으로서 예시적인 유용한 기재로는 옵티어레이(OptArray)-DNA NHS 그룹(엑셀러8(Acceler8)), 넥스테리온 슬라이드 AL(Nexterion Slide AL)(스코트(Schott)) 및 넥스테리온 슬라이드 E(스코트)을 포함하나, 이에 제한되지 않는다.
- [0092] 실시양태에서, 포획 프로브의 결합 모이어티는 비오틴이고, 기재는 아비딘(예컨대, 스트렙트아비딘)을 포함한다. 아비딘을 포함하는 유용한 기재는 상업적으로 이용가능한 것으로서, TB0200(엑셀러8(Acceler8)), SAD6, SAD20, SAD100, SAD500, SAD2000(크산테크(Xantec)), 슈퍼아비딘(SuperAvidin)(어레이-이트(Array-It)), 스트렙트아비딘 슬라이드(카탈로그 #MPC 000, 제노포어(Xenopore)) 및 스트렙트아비딘엔슬라이드(STREPTAVIDINnslide)(카탈로그 #439003, 그레이너 바이오-원(Greiner Bio-one))을 포함한다.
- [0093] 실시양태에서, 포획 프로브의 결합 모이어티는 아비딘(예컨대, 스트렙트아비딘)이고, 기재는 비오틴을 포함한다. 상업적으로 이용가능한 것으로서, 비오틴을 포함하는 유용한 기재로는 옵티어레이(Optarray)-비오틴(엑셀러8), BD6, BD20, BD100, BD500 및 BD2000(크산테크)을 포함하나, 이에 제한되지 않는다.
- [0094] 실시양태에서, 포획 프로브의 결합 모이어티는 광활성화에 의해 기재에 결합할 수 있는 반응성 모이어티를 포함할 수 있다. 기재는 광반응성 모이어티를 포함할 수 있거나, 또는 나노리포터의 제1 부분은 광반응성 모이어티를 포함할 수 있다. 광반응성 모이어티의 일부 예로는 아릴 아지드, 예컨대, N((2-피리딜디티오)에틸)-4-아지도 살리실아미드; 플루오르화된 아릴 아지드, 예컨대, 4-아지도-2,3,5,6-테트라플루오로벤조산; 벤조페논계 시약, 예컨대, 4-벤조일벤조산의 숙신이미딜 에스테르; 및 5-브로모-데옥시우리딘을 포함한다.
- [0095] 실시양태에서, 포획 프로브의 결합 모이어티는 당업자에게 자명한 다른 결합 쌍을 통해 기재에 고정화될 수 있다.
- [0096] 기재에의 결합 후, 표적 핵산은 표적 핵산을 연장시키는 데 충분한 힘(예컨대, 중력, 수력학적 힘, 전자기력 "일렉트로스트레칭", 유동 연신, 후퇴 메니스커스 기법, 및 이의 조합)을 가함으로써 신장될 수 있다.
- [0097] 표적 핵산은 표적 핵산의 제2 부분에 상보적인 도메인을 포함하는 제2 포획 프로브에 의해 결합될 수 있다. 상기 부분은 표적 핵산의 단부일 수 있거나, 또는 단부 쪽에 있지 않을 수도 있다. 제2 포획 프로브의 결합은 표적 핵산의 신장 이후에 또는 신장 동안 이루어질 수 있거나, 또는 신장되지 않은 표적 핵산에 대하여 이루어질 수 있다. 제2 포획 프로브는 상기 기술된 바와 같이 결합을 가질 수 있다.
- [0098] 포획 프로브는 검출가능한 표지, 즉, 기준(fiducial) 스폿을 포함할 수 있거나, 또는 그와 회합될 수 있다.
- [0099] 포획 프로브는 샘플로부터 표적 핵산을 단리시킬 수 있다. 여기서, 포획 프로브는 표적 핵산을 포함하는 샘플에 첨가된다. 포획 프로브는 표적 핵산의 영역에 상보적인 포획 프로브의 영역을 통해 표적 핵산에 결합한다. 표적 핵산이 포획 프로브의 결합 모이어티에 결합하는 모이어티를 포함하는 기재와 접촉할 때, 핵산은 기재 상에 고정화된다.
- [0100] 사용자가 고도로 단편화된 샘플로부터 가능한 한 많은 표적 핵산 분자를 확실하게 "포획"할 수 있도록 하기 위해, 각각이 표적 핵산의 상이한 영역에 상보적인 것인 복수 개의 포획 프로브를 포함하는 것이 도움이 된다. 예를 들어, 3개의 포획 프로브 풀이 존재할 수 있으며, 여기서, 제1 풀은 그의 5' 단부 인근의 표적 핵산의 영역에 상보적이고, 제2 풀은 표적 핵산의 중단 부분의 영역에 상보적이고, 제3 풀은 그의 3' 단부 인근의 표적에 상보적이다. 이는 표적 핵산당 "n개의 관심 영역"으로 일반화될 수 있다. 예에서, 단편화된 표적 핵산의 각 개의 풀은 비오틴 태그를 포함하거나, 또는 그에 결합한 포획 프로브에 결합하였다. 입력 샘플 중 1/n(여기서,

$n$  = 표적 핵산 중 별개의 영역의 개수)이 각 풀 챔버에 대해 단리된다. 포획 프로브는 관심 표적 핵산에 결합한다. 이어서, 표적 핵산은 기재에 부착된 아비딘 분자에 포획 프로브의 비오틴을 통해 고정화된다. 임의적으로, 표적 핵산은 예컨대, 유동 또는 정전기력을 통해 연신된다. 전체  $n$ 개의 풀은 동시에 연신 및 결합될 수 있거나, 또는 완전히 연신된 분자의 개수를 최대화시키기 위해, (가장 5' 쪽의 영역을 포획하는) 풀 1이 먼저 연신되고 결합된 후; 이어서, (표적 영역의 중간 부분을 포획하는) 풀 2가 이어서 연신되고 결합되고; 마지막으로, 풀 3이 연신되고 결합될 수 있다.

- [0101] 필요한 별개의 포획 프로브의 개수는 표적 핵산 단편 크기와 역의 관계를 가진다. 다시 말해, 고도로 단편화된 표적 핵산의 경우, 더 많은 포획 프로브가 요구될 것이다. 고도로 단편화되고, 분해된 표적 핵산을 가지는 샘플 유형의 경우(예컨대, 포르말린 고정된 파라핀 포매된 조직), 포획 프로브의 다중 풀을 포함하는 것이 유용할 수 있다. 다른 한편으로는, 예컨대, 시험관내에서 수득된 단리된 핵산과 같이, 긴 표적 핵산 단편을 가진 샘플의 경우, 5' 단부의 단일 포획 프로브가 충분한 것일 수 있다.
- [0102] 두 포획 프로브 사이, 또는 한 포획 프로브 뒤에서부터 표적 핵산의 말단 앞까지의 표적 핵산의 영역을 본원에서는 "갭"으로 지칭한다. 갭은 본 발명의 시퀀싱 프로브가 결합하는 데 이용가능한 표적 핵산의 부분이다. 최소 갭은 표적 결합 도메인 길이(예컨대, 4 내지 10개의 뉴클레오타이드)이고, 최대 갭은 전체 염색체 대다수이다.
- [0103] 고정화된 표적 핵산은 도 12에 도시되어 있다. 여기서, 2개의 포획 프로브는 "5' 포획 프로브" 및 "3' 포획 프로브"로서 확인된다.
- [0104] 도 8a는 표적 핵산에 결합된 시퀀싱 프로브의 개략도를 보여주는 것이다. 여기서, 표적 핵산은 티미딘(T)을 가진다. 검출가능한 표지를 포함하는 상보적인 핵산 또는 리포터 복합체의 제1 풀은 상단에 제시되어 있고, 풀의 각 구성원은 상이한 검출가능한 표지(예컨대, 티미딘은 녹색 신호로 확인된다) 및 상이한 뉴클레오타이드 서열을 가진다. 표적 결합 도메인 중의 제1 뉴클레오타이드는 표적 핵산 중의 T에 결합한다. 프로브의 제1 부착 영역은 프로브의 표적 결합 도메인 중의 제1 뉴클레오타이드가 티미딘에 결합한다는 것을 명시하는 하나 이상의 뉴클레오타이드 서열(들)을 포함한다. 따라서, 티미딘에 상보적인 핵산만이 바코드 도메인의 제1 위치에 결합한다. 도시된 바와 같이, 검출가능한 표지 또는 검출가능한 표지를 포함하는 리포팅된 복합체를 포함하는 티미딘을 코딩하는 제1 상보적인 핵산은 프로브의 바코드 도메인의 제1 위치 중의 부착 영역에 결합한다.
- [0105] 상보적인 핵산 또는 리포터 복합체 풀의 개수는 바코드 도메인 중의 위치의 개수와 동일하다. 따라서, 바코드 도메인이 6개의 위치를 가지는 경우, 프로브 상에서 6개의 풀이 사이클링될 것이다.
- [0106] 대안적으로, 표적 핵산과 프로브의 접촉 이전에, 프로브는 그의 제1 위치에서 검출가능한 표지를 포함하는 상보적인 핵산 또는 리포터 복합체에 하이브리드화될 수 있다. 따라서, 그의 표적 핵산과 접촉하였을 때, 프로브는 그의 제1 위치로부터 검출가능한 신호를 방출할 수 있고, 바코드 도메인 상의 제1 위치에 대한 상보적인 핵산 또는 리포터 복합체의 제1 풀을 제공할 필요가 없다.
- [0107] 도 8b는 도 8a에 제시된 방법을 계속 진행하는 것이다. 여기서, 바코드 도메인의 제1 위치 중의 부착 영역에 결합한, 티미딘에 대한 제1 상보적인 핵산(또는 리포터 복합체)은 검출가능한 표지가 없는, 티미딘에 대한 제1 하이브리드화 핵산으로 대체되었다. 검출가능한 표지가 없는, 티미딘에 대한 제1 하이브리드화 핵산은 검출가능한 표지를 포함하는 앞서 결합된 상보적인 핵산, 또는 앞서 결합된 리포터 복합체를 대체한다. 이로써, 바코드 도메인의 위치 1은 더 이상 검출가능한 신호를 방출하지 않게 된다.
- [0108] 실시양태에서, 검출가능한 표지를 포함하는 상보적인 핵산 또는 리포터 복합체는 부착 영역으로부터 제거될 수는 있지만, 검출가능한 표지가 없는 하이브리드화 핵산으로 대체될 수는 없다. 이는 예를 들어, 카오토로픽제를 첨가함으로써, 온도를 증가시킴으로써, 염 농도에 변화를 줌으로써, pH를 조정함으로써 및/또는 수력학적 힘을 가함으로써 이루어질 수 있다. 본 실시양태에서, 더 적은 수의 시약(즉, 검출가능한 표지가 없는 하이브리드화 핵산)이 요구된다.
- [0109] 도 8c는 청구하는 본 발명의 방법을 계속 진행하는 것이다. 여기서, 표적 핵산은 그의 티미딘(T) 다음으로 시티딘(C)을 가진다. 상보적인 핵산 또는 리포터 복합체의 제2 풀은 상단에 제시되어 있고, 풀의 각 구성원은 상이한 검출가능한 표지 및 상이한 뉴클레오타이드 서열을 가진다. 또한, 제1 풀의 상보적인 핵산 또는 리포터 복합체의 상보적인 핵산에 대한 뉴클레오타이드 서열은 제2 풀의 것에 대한 뉴클레오타이드 서열과 다르다. 그러나, 염기 특이적인 검출가능한 표지는 상보적인 핵산의 풀에 공통된 것이고, 예컨대, 티미딘은 녹색 신호에 의해 확인된다. 여기서, 표적 결합 도메인 중의 제2 뉴클레오타이드는 표적 핵산 중의 C에 결합한다. 프로브의 제2 부착 영역은 프로브의 표적 결합 도메인 중의 제2 뉴클레오타이드가 시티딘에 결합한다는 것을 명시하는 뉴클레오타이드 서열

을 가진다. 따라서, 오직 시터딘에 대한, 제2 폴로부터의 검출가능한 표지를 포함하는 상보적인 핵산 또는 리포터 복합체가 바코드 도메인의 제2 위치에 결합한다. 도시된 바와 같이, 시터딘을 코딩하는 제2 상보적인 핵산 또는 리포터 복합체는 프로브의 바코드 도메인의 제2 위치에 결합한다.

[0110] 실시양태에서, 도 8c에 도시된 단계가 도 8b에 도시된 단계에 후속된다. 여기서, 일단 (도 8a의) 상보적인 핵산 또는 리포터 복합체의 제1 풀이 (도 8b의) 검출가능한 표지가 없는 제1 하이브리드화 핵산으로 대체되고 나면, 이때, (도 8c에 도시된 바와 같은) 상보적인 핵산 또는 리포터 복합체의 제2 풀이 제공된다. 대안적으로, 도 8c에 제시된 단계가 도 8b에 제시된 단계가 동시에 이루어진다. 여기서, (도 8b의) 검출가능한 표지가 없는 제1 하이브리드화 핵산은 (도 8c에 도시된 바와 같은) 상보적인 핵산 또는 리포터 복합체의 제2 풀이 동시에 제공된다.

[0111] 도 8d는 도 8c에 도시된 방법을 계속 진행하는 것이다. 여기서, 바코드 도메인 상의 제1부터 제5 위치가 검출가능한 표지를 포함하는 상보적인 핵산 또는 리포터 복합체에 의해 결합되었고, 검출가능한 표지가 없는 하이브리드화 핵산으로 대체되었다. 바코드 도메인의 제6 위치가 현재 검출가능한 표지를 포함하는 상보적인 핵산 또는 리포터 복합체에 의해 결합되고 있고, 표적 결합 도메인 중의 제6 위치는 구아닌(G)에 결합하는 것으로 확인된다.

[0112] 상기 언급된 바와 같이, 검출가능한 표지를 포함하는 상보적인 핵산 또는 리포터 복합체는 부착 영역으로부터 제거될 수는 있지만, 검출가능한 표지가 없는 하이브리드화 핵산으로 대체될 수는 없다.

[0113] 필요한 경우, 검출가능한 표지 교환 속도는 검출가능한 표지 교환 속도를 가속화시키는 소형 단일 가닥 올리고 뉴클레오티드를 도입함으로써 가속화될 수 있다(예컨대, "토-홀드(Toe-Hold)" 프로브; 예컨대, 문헌 [Seeling *et al.*, "Catalyzed Relaxation of a Metastable DNA Fuel"; *J. Am. Chem. Soc.* 2006, 128(37), pp12211-12220] 참조).

[0114] 바코드 도메인 상의 마지막 위치(도 8d에서 제6 위치) 상의 상보적인 핵산 또는 리포터 복합체를 대체할 수 있지만; 그러나, 이는 시퀀싱 프로브가 또 다른 시퀀싱 프로브로 대체되는 때에는 필요하지 않을 수도 있다. 실제로, 도 8d의 시퀀싱 프로브는 이제 탈하이브리드화될 수 있고, 표적 핵산으로부터 제거될 수 있으며, 도 8e에 도시된 바와 같이, 임의의 상보적인 핵산에 의해 아직 결합되지 않은 제2(중복 또는 비중복) 시퀀싱 프로브로 대체될 수 있다. 도 8e의 프로브는 제2 프로브 집단에 포함된 것일 있다.

[0115] 도 8a 내지 8e와 유사하게, 도 9a 및 9d 내지 9g는 본 발명의 방법의 단계를 보여주는 것이지만; 그러나, 도 9a 및 9d 내지 9g는 (검출가능한 표지를 포함하는) 리포터 복합체가 시퀀싱 프로브의 부착 영역에 결합한다는 것으로 명확하게 도시하는 것이다. 도 9d 및 9e는 리포터 복합체에 하이브리드화된 프로브로부터 방출된 형광 신호를 보여주는 것이다. 도 9d 및 9e는 표적 핵산이 "T-A"의 서열을 가진다는 것을 보여주는 것이다.

[0116] 도 10은 도 9d 및 9e에 제시된 단계를 요약해 놓은 것이다. 도면 상단에는 예시적인 프로브의 뉴클레오티드 서열이 제시되어 있고, 프로브의 중요한 도메인이 식별표시되어 있다. 프로브는 그의 표적 결합 도메인과 그의 바코드 도메인 사이에 임의적인 이중 가닥 DNA 스페이서를 포함한다. 바코드 도메인은 순서대로 "Flank 1" 부분, "AR-1" 부분, "AR-1/Flank 2" 부분, "AR-2" 부분, 및 "AR-2/Flank 3" 부분을 포함한다. 단계 1에서, "AR-1 Detect"는 프로브의 "AR-1" 및 "AR-1/Flank 2" 부분에 하이브리드화된다. "AR-1 Detect"는 제1 위치 티미딘을 코딩하는 검출가능한 표지를 포함하는 상보적인 핵산 또는 리포터 복합체에 상응한다. 따라서, 단계 1은 도 9d에 상응하는 것이다. 단계 2에서, "Lack 1"은 프로브의 "Flank 1" 및 "AR-1" 부분에 하이브리드화된다. "Lack 1"은 (도 9e에서 검은색 막대로 도시된, 제1 부착 영역을 커버하는) 프로브의 제1 부착 영역에 특이적인, 검출가능한 표지가 없는, 하이브리드화 핵산에 상응하는 것이다. 리포터 복합체 또는 상보적인 핵산에 대해 5'인, "Flank 1" 위치에 하이브리드화함으로써, 하이브리드화 핵산은 프로브로부터 리포터 복합체/상보적인 핵산을 더욱 효율적으로 대체한다. "Flank" 부분은 또한 "토-홀드"로도 알려져 있다. 단계 3에서, "AR-2 Detect"는 프로브의 "AR-2" 및 "AR-2/Flank 3" 부분에 하이브리드화된다. "AR-2 Detect"는 제2 위치 구아닌을 코딩하는 검출가능한 표지를 포함하는 상보적인 핵산 또는 리포터 복합체에 상응한다. 따라서, 단계 3은 도 9e에 상응하는 것이다. 본 실시양태에서, 검출가능한 표지가 없는 하이브리드화 핵산 및 검출가능한 표지를 포함하는 상보적인 핵산/리포터 복합체가 순차적으로 제공된다.

[0117] 대안적으로, 검출가능한 표지가 없는 하이브리드화 핵산 및 검출가능한 표지를 포함하는 상보적인 핵산/리포터 복합체가 동시에 제공된다. 상기 대안적 실시양태는 도 11에 도시되어 있다. 단계 2에서, "Lack 1"(검출가능한 표지가 없는 하이브리드화 핵산)이 "AR-2 Detect"(제2 위치 구아닌을 코딩하는 리포터 복합체)와 함께

제공된다. 상기 대안적 실시양태는 두 단계를 하나로 조합하는 바, 이에 도 10에 도시된 실시양태보다 시간상 더욱 효과적일 수 있다.

[0118] 도 12는 본 발명의 방법을 도시한 것이다. 여기서, 표적 핵산은 두 위치에서 포획되고, 고정화됨으로써, 프로브가 결합할 수 있는 "갭"이 생성된다. 제1 프로브 집단이 표적 핵산 상에 하이브리드화되고, 검출가능한 표지가 검출된다. 제2 프로브 집단, 제3 프로브 집단, 내지 100개 초과 프로브 집단을 사용하여 초기 단계를 반복한다. 약 100개의 프로브 집단을 사용함으로써 표적 핵산 중 각 뉴클레오타이드에 대한 약 5X 커버리지를 제공한다. 도 12는 한 관측 시계(FOV: Field of View)로부터 신호를 검출하는 데 소요되는 시간에 기초한 실시간의 추정된 속도를 제공하는 것이다.

[0119] 표적 핵산의 길이와 함께 프로브 분포도 검출가능한 신호를 분석하는 데 중요하다. 상기 논의된 바와 같이, 두 검출가능한 표지에 대한 해상도 한계는 약 600개의 뉴클레오타이드이다. 바람직하게, 프로브 집단 중 각 시퀀싱 프로브는 서로 600개의 뉴클레오타이드보다 가깝게는 결합하지 않을 것이다. 상기 논의된 바와 같이, 600개의 뉴클레오타이드가 전형적인 시퀀싱 장치의 해상도 한계이다. 이러한 경우, 시퀀싱 프로브는 단일 리드를 제공할 것이며; 이는 도 12에서 제일 왼쪽에 있는 해상 제한 스폿에 도시되어 있다.

[0120] 무작위로, 그러나, 부분적으로는 표적 결합 도메인의 길이, 프로브의  $T_m$ , 및 적용된 프로브의 농도에 의존하여, 집단 내의 두 상이한 시퀀싱 프로브는 서로 600개의 뉴클레오타이드 이내에서 결합할 수 있다. 이러한 경우, 비정돈된 다중 리드가 단일 해상 제한 스폿으로부터 방출될 것이며; 이는 도 12에서 제2의 해상 제한 스폿에 도시되어 있다.

[0121] 대안적으로 또는 추가로, 집단 중 시퀀싱 프로브의 농도는 표적 핵산의 특이적인 영역에서 프로브의 커버리지를 감소시키기 위해 예컨대, 시퀀싱 장치의 해상도 한계 이상으로 감소될 수 있고, 이로써, 해상 제한 스폿으로부터 단일 리드가 생성될 수 있다.

[0122] 도 23은 도 8 내지 12에서 사용된 것과 다른 시퀀싱 프로브의 개략도를 보여주는 것이다. 여기서, 바코드 도메인 상의 각 위치는 검출가능한 표지를 포함하는 상보적인 핵산에 의해, 또는 리포터 복합체에 의해 결합된다. 따라서, 본 예에서, 6개의 뉴클레오타이드 서열은 순차적인 상보적인 핵산의 대체를 필요로 하지 않고 관독될 수 있다. 상기 시퀀싱 프로브를 사용하면, 상기 기술된 방법의 많은 단계가 생략되는 바, 서열 정보를 얻는 데 소요되는 시간은 단축될 수 있다. 그러나, 상기 프로브는 비중복성인 검출가능한 표지로부터 이익을 얻게 될 것이며, 예컨대, 형광단은 빛의 비중복 파장에 의해 여기되거나, 또는 형광단은 빛의 비중복 파장을 방출한다.

[0123] 본 방법은 고정화된 표적 핵산의 각 영역에 대한 각각의 확인된 선형 순서의 뉴클레오타이드를 조립하여 고정화된 표적 핵산에 대한 서열을 확인하는 단계를 추가로 포함한다. 조립 단계는 실행가능 프로그램이 그 안에 저장된 비밀시적 컴퓨터 관독가능한 저장 매체를 사용한다. 프로그램은 표적 핵산의 각 영역에 대한 각각의 확인된 선형 순서의 뉴클레오타이드를 배열하여 핵산 서열을 수득하도록 마이크로프로세서에 지시한다. 조립은 "실시간"으로, 즉, 모든 데이터가 수집된 이후라기보다는 시퀀싱 프로브로부터 데이터가 수집되는 동안에 이루어질 수 있다.

[0124] 상기 측면 및 실시양태 중 임의의 것은 본원 요약 및/또는 상세한 설명 섹션에 개시된 바와 같이 임의의 다른 측면 또는 실시양태와 조합될 수 있다.

[0125] 정의:

[0126] 특정 예시적인 실시양태에서, 본원에서 사용되는 바, 본원에서 사용되는 바, "어닐링" 및 "하이브리드화"라는 용어는 상호교환적으로 사용되며, 이는 안정적인 이중체를 형성한다는 것을 의미한다. 한 측면에서, 안정적인 이중체란, 이중체 구조가 예컨대, 이중체 가닥의  $T_m$ 보다 약 5°C 낮거나, 또는 약 5°C 높은 온도 및 낮은 염 농도, 예컨대, 0.2 M 미만, 또는 0.1 M 미만, 또는 당업자에게 공지된 염 농도와 같은 조건하에서의 엄격한 세척에 의해 파괴되지 않는다는 것을 의미한다. 이중체와 관련하여 사용될 때, "완벽하게 매칭된"이라는 용어는 이중체를 구성하는 폴리뉴클레오타이드 및/또는 올리고뉴클레오타이드 가닥이 서로 이중 가닥 구조를 형성함으로써 각 가닥의 모든 뉴클레오타이드가 나머지 다른 가닥 중의 뉴클레오타이드와 왓슨-크릭(Watson-Crick) 염기쌍을 형성한다는 것을 의미한다. "이중체"라는 용어는 사용될 수 있는 2-아미노퓨린 염기, PNA 등과 뉴클레오시드 유사체, 예컨대, 테옥시이노신, 뉴클레오시드의 쌍 형성을 포함하나, 이에 제한되지 않는다. 두 올리고뉴클레오타이드 사이의 이중체에서 "미스매치"란, 이중체에서 뉴클레오타이드 쌍이 왓슨-크릭 결합을 이루지 못한다는 것을 의미한다.

[0127] 본원에서 사용되는 바, "하이브리드화 조건"이라는 용어는 전형적으로 약 1 M 미만, 더욱 일반적으로, 약 500



mM 미만, 및 더욱더 일반적으로 약 200 mM 미만인 염 농도를 포함할 것이다. 하이브리드화 온도는 5℃만큼 낮을 수 있지만, 전형적으로는 22℃ 초과, 더욱 전형적으로는 약 30℃, 및 대개는 약 37℃ 초과일 수 있다. 하이브리드화는 일반적으로 엄격한 조건하에서, 예컨대, 프로브가 그의 표적 서브서열에 특이적으로 하이브리드화할 수 있는 조건하에서 수행된다. 엄격한 조건은 서열 의존적이며, 다른 환경에서는 상이하다. 단편의 길이가 길수록, 특이적인 하이브리드화를 위해 더 높은 하이브리드화 온도가 요구될 수 있다. 염기 조성 및 상보적인 가닥의 길이, 유기 용매의 존재 및 염기 미스매칭 정도를 비롯한 다른 인자가 하이브리드화의 엄격성에 영향을 줄 수 있는 바, 어느 하나의 단독의 절대 척도보다는 파라미터 조합이 더욱 중요하다.

[0128] 일반적으로, 엄격한 조건은 정의된 이온 강도 및 pH에서 특이적인 서열에 대한 T<sub>m</sub>보다 약 5℃ 낮은 것으로 선택된다. 예시적인 엄격한 조건은 적어도 25℃의 온도 및 pH 7.0 내지 8.3에서 적어도 0.01 M 내지 1 M 이하의 Na 이온 농도(또는 다른 염)인 염 농도를 포함한다. 예를 들어, 25-30℃의 온도 및 5X SSPE(750 mM NaCl, 50 mM Na 포스페이트, 5 mM EDTA, pH 7.4)인 조건이 대립유전자에 특이적인 프로브 하이브리드화에 적합하다. 엄격한 조건에 대해서는 예를 들어, 문헌 [Sambrook, Fritsche and Maniatis, "Molecular Cloning A Laboratory Manual, 2nd Ed." Cold Spring Harbor Press (1989)] 및 [Anderson Nucleic Acid Hybridization, 1st Ed., BIOS Scientific Publishers Limited (1999)]를 참조할 수 있다. 본원에서 사용되는 바, "특이적으로 ~에 하이브리드화하는" 또는 "~에 특이적으로 하이브리드화하는"이라는 용어 또는 유사 용어는 엄격한 조건하에서 분자가 실질적으로 특정 뉴클레오타이드 서열 또는 서열들에 결합, 이중체 형성, 또는 하이브리드화한다는 것을 의미한다.

[0129] 프로브의 특정 위치와 회합된 검출가능한 표지는 1회 또는 다회에 걸쳐 "판독"될 수 있고(예컨대, 그의 형광이 검출될 수 있고); "판독"은 "베이스콜(basecall)"이라는 용어와 동의어일 수 있다. 다중 리드는 정확도를 향상시킨다. 단일의 원래 표적 분자로부터 도출된, 연속된 서열 스트레치 정보가 검출될 때, 표적 핵산 서열이 "판독되는 것이고"; 전형적으로, 이는 (하기 정의되는 바와 같이) 다중 통과 컨센서스를 통해 생성된다. 본원에서 사용되는 바, "커버리지" 또는 "커버리지의 심도"란, 표적 영역이 (별개의 리드를 통하여) 시퀀싱되고, 참조 서열에 대하여 정렬되는 횟수를 지칭한다. 리드 커버리지는 특이적인 참조 표적 서열에 대해 맵핑되는 리드의 총 개수이고; 염기 커버리지는 특이적인 게놈 위치에서 이루어지는 베이스콜의 총수이다.

[0130] 본원에서 사용되는 바, "hybe 및 seq 사이클"은 특정 프로브 또는 프로브 집단 상의 각 부착 영역을 검출하는데 필요한 모든 단계를 지칭한다. 예를 들어, 표적 핵산 상의 6개의 위치를 검출할 수 있는 프로브의 경우, 한 "hybe 및 seq 사이클"은 적어도, 프로브를 표적 핵산에 하이브리드화하고, 프로브의 바코드 도메인 상의 6개의 각 위치의 부착 영역에 상보적인 핵산/리포터 복합체를 하이브리드화하고, 6개의 각 위치와 회합된 검출가능한 표지를 검출하는 것을 포함할 것이다.

[0131] "k-mer 프로브"라는 용어는 본 발명의 프로브와 동의어이다.

[0132] 별개의 리드로부터 2개 이상의 서열을 정렬할 때, 중복 부분을 조합하여 단일 컨센서스 서열을 생성할 수 있다. 중복 부분이 같은 염기(정렬의 단일 칼럼)를 가지는 위치에서, 상기 염기는 컨센서스가 된다. 다양한 법칙을 사용하여 중복 서열 중에 불일치가 존재하는 위치에 대한 컨센서스를 생성할 수 있다. 간단한 다수의 법칙은 컨센서스로서 칼럼 내에서 가장 공통된 염기를 사용한다. "다중 통과 컨센서스"란, 단일 표적 분자로부터의 모든 별개의 프로브 판독치의 정렬이다. 적용된 프로브 집단/폴의 총 사이클 횟수에 의존하여, 상이한 수준의 반복 또는 중복을 이용하여 단일 표적 분자 내의 각 염기 위치를 질의할 수 있고; 일반적으로, 반복은 베이스콜의 신뢰 수준을 증가시킨다.

[0133] "원시 정확도"는 염기를 정확하게 확인할 수 있는 시스템 고유 능력에 관한 척도이다. 원시 정확도는 시퀀싱 기술에 의존한다. "컨센서스 정확도"는 추가 리드 및 통계학적 검정력을 사용하여 염기를 정확하게 확인할 수 있는 시스템의 능력에 관한 척도이다. "특이성"이란, 1회 실행당 총 리드 중, 의도되는 표적으로 맵핑되는 리드의 비율(%)을 지칭한다. "균일성"이란, 표적 영역 간의 서열 커버리지의 가변성을 지칭하고; 높은 균일성은 낮은 가변성과 상관 관계를 가진다. 상기 특징은 모든 표적화된 영역 간에 걸쳐 ≥20%의 평균 커버리지 심도로 커버되는 표적화된 영역의 분율로서 통상 기록된다. 확률론적 오류(즉, 고유한 시퀀싱 화학 오류)는 같은 표적 핵산의 "다중 통과" 시퀀싱으로 쉽게 보정될 수 있고; 충분한 회차에 걸쳐 통과가 이루어졌다면, 실질적으로 '완벽한 컨센서스' 또는 '오류가 없는' 시퀀싱이 달성될 수 있다.

[0134] 방법을 실행하고/거나, 결과를 기록할 수 있는 임의의 장치를 사용하여 본원에 기술된 방법은 실행될 수 있고/거나, 결과는 기록될 수 있다. 사용될 수 있는 장치의 예로는 모든 유형의 컴퓨터를 비롯한, 전자 컴퓨터 장치를 포함할 수 있지만, 이에 제한되지 않는다. 컴퓨터에서 본원에 기술된 방법을 실행하고/거나, 기록할 때, 컴퓨터가 본 방법의 단계를 수행하도록 설정하는 데 사용될 수 있는 컴퓨터 프로그램은 컴퓨터 프로그램을 포함할

수 있는 임의의 컴퓨터 판독가능한 매체에 포함될 수 있다. 사용될 수 있는 컴퓨터 판독가능한 매체의 예로는 디스켓, CD-ROM, DVD, ROM, RAM, 비일시적 컴퓨터 판독가능한 매체, 및 다른 메모리 및 컴퓨터 저장 장치를 포함하나, 이에 제한되지 않는다. 컴퓨터가 본 방법의 단계를 수행하고/거나, 서열 정보를 조립하고/거나, 결과를 기록하도록 설정하는 데 사용될 수 있는 컴퓨터 프로그램 또한 전자 네트워크 상에, 예를 들어, 인터넷, 인트라넷, 또는 다른 네트워크 상에 제공될 수 있다.

[0135] "소모품 시퀀싱 카드(Consumable Sequencing Card)"(도 24)는 당업계에 공지된 형광 영상화 장치로 도입될 수 있다. 다수의 다양한 피처를 가진 임의의 형광 현미경은 본 시퀀싱 판독을 수행할 수 있다. 예를 들어: 광역장 램프, 레이저, LED, 다광자, 공초점 또는 내부 전반사 조명여기 및/또는 검출에 사용될 수 있다. 필터 기반 또는 격자 기반 스펙트럼 분석(하나 이상의 스펙트럼으로 분석되는 방출 파장)을 포함하는 카메라(단일 또는 다중) 및/또는 광전자 증배관(단일 또는 다중)이 형광 현미경의 방출 검출 채널 상에서 가능하다. 표준 컴퓨터는 소모품 시퀀싱 카드, 상기 카드를 통해 이동하는 시약, 및 형광 현미경에 의해 검출, 둘 모두를 제어할 수 있다.

[0136] 시퀀싱 데이터는 임의의 개수의 표준 차세대 시퀀싱 어셈블러에 의해 분석될 수 있다(예컨대, 문헌 [Wajid and Serpedin, "Review of general algorithmic features for genome assemblers for next generation sequencers" *Genomics, proteomics & bioinformatics*, 10 (2), 58-73, 2012] 참조). 현미경의 단일 회절 제한 영역 내에서 획득된 시퀀싱 데이터는 "국소적으로 조립되어" 회절 스폿 내의 다중 리드로부터 컨센서스 서열을 생성한다. 이어서, 다중의 회절 스폿 조립된 리드는 함께 맵핑되어 표적화된 전체 유전자 세트, 또는 전체 게놈(들)의 새(de-novo) 조립체를 나타내는 연속된 서열을 생성한다.

[0137] 본 발명과 관련된 추가의 교시는 하기: U.S. 8,148,512, U.S. 7,473,767, U.S. 7,919,237, U.S. 7,941,279, U.S. 8,415,102, U.S. 8,492,094, U.S. 8,519,115, U.S. 2009/0220978, U.S. 2009/0299640, U.S. 2010/0015607, U.S. 2010/0261026, U.S. 2011/0086774, U.S. 2011/0145176, U.S. 2011/0201515, U.S. 2011/0229888, U.S. 2013/0004482, U.S. 2013/0017971, U.S. 2013/0178372, U.S. 2013/0230851, U.S. 2013/0337444, U.S. 2013/0345161, U.S. 2014/0005067, U.S. 2014/0017688, U.S. 2014/0037620, U.S. 2014/0087959, U.S. 2014/0154681, 및 U.S. 2014/0162251(상기 문헌은 각각 그 전문이 본원에서 참조로 포함된다) 중 하나 이상의 것에 기술되어 있다.

[0138] 실시예

[0139] 실시예 1: 본 발명의 표적 핵산을 시퀀싱하는 방법은 신속한 방법이다.

[0140] 도 8 내지 12에 도시된 바와 같은, 본 발명의 방법의 단계에 대한 타이밍은 하기 기술된다.

[0141] 본 발명은 최소의 샘플 제조를 필요로 한다. 예를 들어, 도 13에 도시된 바와 같이, 샘플 중 핵산은 2시간 이하의 제조 시간 이후 판독이 시작될 수 있고; 이는 각각 약 12시간 또는 9시간의 제조 시간이 소요되는 이온 토렌트(Ion Torrent)(암플리Seq(AmpliSeq)<sup>TM</sup>) 또는 일루미나(Illumina)(트루사이트(TruSight)) 시퀀싱에 소요되는 시간보다 유의적으로 짧은 시간이다.

[0142] 예시적인 실행에 대한 계산 결과는 도 14에 도시되어 있고, 사이클링 시간에 대한 계산은 도 15에 도시되어 있다.

[0143] 프로브 집단을 고정화된 표적 핵산에 결합시키는 데 약 60초가 소요된다. 본 반응은 합성 백본 상의 다중의 표적 결합 도메인 카피를 사용함으로써 가속화될 수 있다. 미세유체 제어식 유체 교환 디바이스를 사용하여 비결합 프로브를 세척하여 제거하는 데 약 1/2초가 소요된다.

[0144] (검출가능한 표지를 포함하는) 상보적인 핵산의 제1 풀을 첨가하고, 그를 바코드 도메인의 제1 위치의 부착 영역에 결합시키는 데 약 15초가 소요된다.

[0145] 각 관측 시계(FOV)를, 각 색상이 단일 염기를 나타내는 4개의 상이한 색상에 대해 영상화한다. 5' 포획 프로브 또는 3' 포획 프로브(또는 그 둘 모두) 상에 배치된 기준 스폿은 두 위치 사이에 (갭이 있는 표적 핵산의 존재와 관련된) 한 줄로 오직 상기의 광학 바코드만을 판독하는 데 도움이 될 수 있다. 시퀀싱 프로세스에서 연속 단계시 동일한 영상 정렬을 생성하기 위해서 기준 스폿을 각 관측 시계에 추가로 첨가할 수 있다. 단일 FOV에서 4개의 영상 모두를 얻을 수 있고, 이어서, 광학 판독 장치를 새 FOV로 이동시킬 수 있거나, 한 색상으로 모든 FOV를 촬영한 후, 제2 색상으로 재영상화할 수 있다. 단일 FOV는 약 1/2초 내에 판독될 수 있다. 새 FOV로 이동시키는 데 약 1/2초가 소요된다. 그러므로, "n"개의 FOV를 판독하는 데 소요되는 시간은 "n" x 1 sec이다).

- [0146] 검출가능한 표지를 가지는 상보적인 핵산은 열을 가하거나, 또는 과량의 검출가능한 표지가 없는 상보적인 핵산으로 세척함으로써 바코드 도메인의 제1 위치로부터 제거된다. 필요한 경우, 검출가능한 표지 교환 속도는 검출가능한 표지 교환 속도를 가속화시키는 소형 단일 가닥 올리고뉴클레오타이드를 도입함으로써 가속화될 수 있다 (예컨대, "토-홀드" 프로브; 예컨대, 문헌 [Seeling *et al.*, "Catalyzed Relaxation of a Metastable DNA Fuel"; *J. Am. Chem. Soc.* 2006, 128(37), pp 12211-12220] 참조). 이동을 계속 진행하기 이전에 검출가능한 표지를 가지는 상보적인 핵산이 모두 제거되었는지 확인하기 위해 FOV를 재영상화할 수 있다. 이는 약 15초가 소요된다. 상기 단계는 배경 신호 수준에 도달할 때까지 반복될 수 있다.
- [0147] 상기 단계는 프로브의 바코드 도메인 중 나머지 위치에 대하여 반복된다.
- [0148] 관독에 소요되는 총 시간은  $m(\text{염기 리드}) \times (15 \text{ sec} + n \text{ FOVs} \times 1 \text{ sec} + 15 \text{ sec})$ 이다. 예를 들어, 바코드 도메인 중 위치의 개수가 6 및 20개의 FOV일 때, 관독에 소요되는 시간은  $6 \times (30 + 20 + 15)$  또는 390초이다.
- [0149] 제1 프로브 집단은 탈하이브리드화된다. 이는 약 60초가 소요된다.
- [0150] 상기 단계는 프로브의 제2 및 후속 집단에 대하여 반복된다. 시퀀싱 프로브 집단이 용점( $T_m$ )에 의해 조직될 때, 각 프로브 집단은 각 염기가 필요한 심도를 확실하게 커버할 수 있도록 하기 위해(이는 오류율에 의해 유도된다) 다중 하이브리드화를 필요로 할 것이다. 또한, 실행 동안 하이브리드화 리드 분석하기 위해, 전체 서열이 실제로 측정되기 이전에 잘 시퀀싱된 각각의 개별 유전자를 인식할 수 있다. 그러므로, 특정의 원하는 오류 빈도(또는 커버리지)를 충족시킬 때까지 사이클링은 반복될 수 있다.
- [0151] 일부 갭이 있는 핵산 결합 밀도 추정치와 함께 상기 기술된 타이밍을 사용함으로써 본 발명의 나노스트링(NSTG)-넥스트 제너레이션 시퀀서(Nanostring (NSTG)-Next Generation Sequencer)의 처리량을 추정할 수 있다.
- [0152] 시퀀서의 순 처리량은
- [0153]  $\text{분획 염기 점유율} \times \text{<갭-길이>} \times \text{FOV당 갭의 개수} \times \text{광학 바코드당 염기의 개수} / [60 \text{ sec}(\text{프로브의 표적 핵산에의 하이브리드화}) + 0.5 \text{ sec}(\text{세척}) + m: \text{바코드 도메인 중 위치} \times (15 \text{ sec}(\text{상보적인 핵산 결합}) + n_{\text{fovs}} \times 1 + 15 \text{ sec}(\text{상보적인 핵산 결합 해제})) + 60 \text{ sec}(\text{프로브의 표적 핵산에의 탈하이브리드화})]$ .
- [0154] 그러므로, 예에서, 단일의 갭이 있는 핵산에 대한 총 "사이클"(도 10에 도시된 방법으로부터 함께 추가):
- [0155]  $60 \text{ sec}(\text{프로브의 표적 핵산에의 하이브리드화}) + 0.5 \text{ sec}(\text{세척}) + m\text{-염기} \times (15 \text{ sec}(\text{상보적인 핵산 결합}) + n_{\text{FOVs}} \times 1 + 15 \text{ sec}(\text{상보적인 핵산 결합 해제})) + 60 \text{ sec}(\text{프로브의 표적 핵산에의 탈하이브리드화})$ .  $m = 6$ ,  $n_{\text{FOVs}} = 20$ 을 사용하면, 수율 시간 =  $60 + 0.5 + 390 + 60 = 510.5 \text{ sec}$ 이다.
- [0156] 갭이 있는 핵산 영역의 점유율 1%, 갭당 염기 4,000개, 및 FOV당 갭이 있는 핵산 단편 5,000개 및  $m=6$  및  $n_{\text{FOVs}}=20$ (상기 기술된 바와 같음)이라고 가정할 때, 순 처리량은
- [0157]  $0.01 \times 4,000 \times 5,000 \times 20 = 4,000,000$ 개의 6 염기 리드/ $510.5 \text{ secs} = 47,012.73$ 개의 염기/sec이다.
- [0158] 그러므로, 본 예에서, 24시간의 연속 측정당 순 처리량 =  $4.062 \text{ 기가베이스(Gb)/일(day)}$ 이다. 대안적 추정치는 최대 12 Gb/일이다. 도 12를 참조할 수 있다.
- [0159] 도 14에 도시된 바와 같이, 100개의 상이한 표적 핵산("100플렉스")을 시퀀싱하는 데 필요한 실행 시간은 약 4.6시간이고; 1,000개의 상이한 표적 핵산("1,000플렉스")을 시퀀싱하는 데 필요한 실행 시간은 약 16시간이다.
- [0160] 도 16은 본 발명과 다양한 다른 시퀀싱 방법/장치에 대해 시퀀싱 속도, 리드 개수, 및 임상적 유용성을 비교한 것이다.
- [0161] **실시예 2: 본 발명의 방법은 오류율이 낮다**
- [0162] 도 17은 말단 위치가 누락되었을 때, 본 발명의 원시 오류율이 약 2.1%라는 것을 보여준다.
- [0163] 청구하는 발명의 경우, 시퀀싱과 관련된 오류율은 완전히 매칭된  $(m+n)$ -mer 및 단일 염기 미스매치  $(m-1+n)$ -mer 사이의 자유 에너지 차이에 관한 것이다.  $m+n$ 인 합산은 표적 결합 도메인 중 뉴클레오타이드의 개수이고,  $m$ 은 바코드 도메인 중 위치의 개수를 나타낸다. 하이브리드화 선택도에 관해서는 방정식을 사용하여 추정할 수 있다 (문헌 [Owczarzy, R. (2005), *Biophys. Chem.*, 117:207-215 and Integrated DNA Technologies website: at the World Wide Web ([www.idtdna.com/analyzer/Applications/Instructions/Default.aspx?AnalyzerDefinitions=true#MismatchMeltTemp](http://www.idtdna.com/analyzer/Applications/Instructions/Default.aspx?AnalyzerDefinitions=true#MismatchMeltTemp)):

$$\theta = 1 - \left( \frac{K_a([가닥\ 2] - [가닥\ 1]) - 1}{2K_a[가닥\ 2]} + \sqrt{\frac{K_a^2([가닥\ 1] - [가닥\ 2])^2 + 2K_a([가닥\ 1] + [가닥\ 2]) + 1}{2K_a[가닥\ 2]}} \right)$$

[0164]

[0165] 여기서,  $K_a$ 는 예측 열역학적 파라미터로부터 구한 회합 평형 상수이다,

$$K_a = e^{\exp\left(\frac{-(\Delta H^\circ - T\Delta S^\circ)}{RT}\right)}$$

[0166]

[0167] 켄타는 정확한 상보체와, 명시된 하이브리드화 온도에서 표적에 어닐링되는 것으로 예상되는 단일 염기 미스매치 서열의 결합률(%)을 나타낸다. T는 하이브리드화 온도(켈빈 온도)이고,  $\Delta H^\circ$  (엔탈피) 및  $\Delta S^\circ$  (엔트로피)는 서열로부터 계산된 용융 파라미터 및 공개된 최근린 열역학적 파라미터이고, R은 이상 기체 상수( $1.987 \text{ cal} \cdot \text{K}^{-1} \cdot \text{mole}^{-1}$ )이고, [가닥1/2]는 올리고뉴클레오타이드의 몰 농도이고, 상수 -273.15는 온도를 켈빈 온도에서 섭씨 온도로 전환시킨다. 가장 정확한 가장 가까운 인접 파라미터는 DNA/DNA 염기쌍(문헌 [Allawi, H., SantaLucia, J. *Biochemistry*, 36, 10581] 참조), RNA/DNA 염기쌍(문헌 [Sugimoto *et al.*, *Biochemistry*, 34, 11211-6] 참조), RNA/RNA 염기쌍(문헌 [Xia, T. *et al.*, *Biochemistry*, 37, 14719] 참조)에 대한 공개 문헌으로부터 입수하였다.

[0168] NSTG-시퀀서로부터 예측된 대략적인 오류율의 추정치의 예는 하기와 같다. (m + n)=8'mer인 경우, 하기 8-mer 바코드 및 그의 단일 염기 미스매치를 고려하라.

5'ATCGTACG3'  
(시퀀싱을 위한 영역)  
3'TAGCATGC5'  
(완벽하게 매칭되는 시퀀싱 광학 바코드)  
3'TAGTATGC5'  
(단일 염기 미스매치(G-T) 쌍 형성을 갖는 시퀀싱 광학 바코드)

[0169]

[0170] 상기 방정식에 기초하여 IDT 계산기를 사용하였을 때, 하기와 같이 얻을 수 있다:

[0171] 17.4°C(완벽하게 매칭되는 경우의  $T_m$ )에서, (50%/0.3%)은  $T_m$ 에서 상기 서열에 하이브리드화된 정확한 광학 바코드/부정확한 바코드의 비가 될 것이며, 상기 서열에 대하여 추정된 오류율은 0.6%가 될 것이다.

[0172] GC 함량이 매우 높은 시퀀싱 계산은 하기와 같이 얻을 수 있다:

5'CGCGGGCC3'  
(시퀀싱을 위한 영역)  
3'GCGGCCGG5'  
(완벽하게 매칭되는 시퀀싱 광학 바코드)  
3'GCGGACGG5'  
(단일 염기 미스매치(G-A)의 잘못된 쌍 형성을 갖는 시퀀싱 광학 바코드)

[0173]

[0174] 41.9°C(완벽하게 매칭되는 경우의  $T_m$ )에서, (50%/0.4%)은  $T_m$ 에서 상기 서열에 하이브리드화된 정확한 광학 바코드/부정확한 바코드의 비가 될 것이며, 상기 서열에 대하여 추정된 오류율은 0.8%가 될 것이다.

[0175] 다수의 8-mer 쌍을 조사한 결과, 오류율은 0.2% 내지 1% 범위로 분포하는 것으로 나타났다. 상기 계산이 사용된 조건과 일치하지는 않지만, 상기 계산은 본 발명의 방법은 다른 단일 분자 시퀀싱 기술, 예컨대, 오류율이 유의적일 수 있는 (>> 10%) 파시픽 바이오사이언시스(Pacific Biosciences) 및 옥스포드 나노포어 테크놀로지스(Oxford Nanopore Technologies)와 비교하였을 때, 고유 오류율이 비교적 낮다는 것을 시사한다.

[0176] 도 18은 본 발명의 원시 정확도가 다른 시퀀싱 방법보다 더 높다는 것을 입증하는 것이다. 따라서, 본 발명은 다른 시퀀싱 방법에서 요구되는 통과 횟수보다 더 적은 통과 이후에 단일 표적으로부터 컨센서스 서열을 제공한다. 추가로, (예를 들어) PacBio 시퀀싱 방법은 70회 통과 이후에도 해당 컨센서스를 달성할 수 없지만, 본 발명은 30회 이상의 통과 이후 "완벽한 컨센서스"/"오류가 없는" 시퀀싱(즉, 99.9999%/Q60)을 얻을 수 있다.

[0177] 실시예 3: 본 발명은 단일 염기쌍 분석능을 가진다.

[0178] 도 19는 본 발명이 낮은 오류율(특이적인 뉴클레오타이드 치환에 의존하여 0% 내지 1.5% 범위)로 단일 염기 분석능을 가진다는 것을 보여주는 것이다.

[0179] 일반 나노스트링 유전자 발현 결합 기술을 사용함으로써 바코드와 하이브리드화되고, 카트리지 표면에 고정화된



표적 RNA를 사용하여 추가 실험을 수행하였다(예컨대, 문헌 [Direct multiplexed measurement of gene expression with color-coded probe pairs"; *Nature Biotechnology*, 26, 317 - 325 (2008)] 참조). 표적 결합 도메인 길이가 상이하고, 완벽하게 매칭되는 바코드(완벽한 10-mer 매치 서열에 연결된 YGBYGR-2um 광학 바코드)의, RNA-표적에 하이브리드화할 수 있는 능력을 측정하였다(도 26). 표적 결합 도메인의 길이가 길수록 더 높은 계수를 제공한다. 또한, 배경보다 높은 서열을 등록하는 데에는 10-mer 표적 결합 도메인이 충분한 것으로 나타났다. 각각의 개별 단일 염기가 변경된 매치를 대안적 광학 바코드를 이용하여 합성하였다. 정확한 광학 바코드 대 부정확한 광학 바코드의 비를 계수하였다(도 24 및 25).

- [0180] 실제 서열에서 SNP를 검출할 수 있는 10 mer의 능력은 배경보다 >15000 계수인 반면, 부정확한 서열은 배경보다 많아야 > 400이다. 정확한 프로브의 존재하에서, 오류율은 실제 서열의 <3%가 될 것으로 예상된다. 상기 데이터는 (본질적으로) 좋지 않은 경우의 시나리오인 것에 주목하기 바란다. 10-염기쌍 하이브리드화 서열만이 6.6 킬로베이스 광학 바코드 리포터(Gen2 스타일)에 부착된다. 어떤 특이적인 조건 최적화도 수행하지 않았다. 그러나, 본 데이터를 통해 나노스트링 차세대 시퀀싱 접근법이 서열의 단일 염기쌍을 분석할 수 있다는 것이 밝혀졌다.
- [0181] 상기 연구에서 사용된 재료 및 방법에 관한 상세한 설명은 하기와 같다:
- [0182] 하이브리드화 프로토콜 프로브 B + 코드세트
- [0183] · 25 ul의 엘리먼트(194 코드세트)를 취한다.
- [0184] · 5 ul 프로브 B+ 상보적인 서열을 표적(100 uM)에 첨가한다.
- [0185] · 15 ul Hyb 완충제(14.56 X SSPE 0.18% 트윈 20)
- [0186] SSPE(150 mM NaCl,  $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$  10 mM,  $\text{Na}_2\text{EDTA}$  10 mM)를 첨가한다.
- [0187] · 10 min 동안 얼음 상에서 인큐베이션시킨다.
- [0188] · 150 ul G 비드(40ul G 비드(10 mg/ml) + 110 ul 5x SSPE 0.1% 트윈 20)를 첨가한다.
- [0189] · RT에서 10 min 동안 인큐베이션시킨다.
- [0190] · 자석 수집기를 사용하여 0.1SSPE 0.1% 트윈 20으로 3회에 걸쳐 세척한다.
- [0191] · 45C에서 10 min 동안 100 ul 0.1x SSPE 중에서 용출시킨다.
- [0192] 표적 하이브리드화 프로토콜(750 mM NaCl)
- [0193] · 상기 용출된 샘플 20 ul를 취한다.
- [0194] · 10 ul hyb 완충제를 첨가한다.
- [0195] · 1 ul 표적(100 nM 비오틀화된 RNA)을 첨가한다.
- [0196] · 30 min 동안 얼음 상에서 인큐베이션시킨다.
- [0197] 15 ul를 취하여 20 min 동안 스트렙타비딘 슬라이드에 결합시키고, G 후크를 이용하여 유동 연신시키고, n계 수기를 이용하여 계수한다.
- [0198] 재료
- [0199] 엘리먼트 194 코드세트
- [0200] IDT로부터 구입한 올리고
- [0201] SSPE(150 mM NaCl,  $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$  10 mM,  $\text{Na}_2\text{EDTA}$  10 mM)
- [0202] Hyb 완충제(14.56 X SSPE 0.18% 트윈 20)

표 2

표 2: 12, 11, ..., 8 mers 에 대한 프로브 B 서열 (서열 번호: 30 내지 서열 번호: 34)

GBRYBG	5	GACTGTACCCACGCGATGACGTTTCGTCAAGAGTCGCATAATCT	3
YRBYRG	5	AGACTGTACCACAAGAATCCCTGCTAGCTGAAGGAGGGTCAAAC	3
YGBYGR	5	GAGACTGTACCCTACGTATATATCCAAGTGGTTATGTCCGACGGC	3
GBRYGB	5	TGAGACTGTACCACCCCTCCAAACGCATTCTTATTGGCAAATGGAA	3
RYGBRG	5	CTGAGACTGTACCCGGGAATCGGCATTTCGCATTCTTAGGATCTAAA	3

[0203]

표 3

표 3: 표적 서열 (굵은체 표시; 서열 번호: 35)

RNA	5	CAATGTGAGTCTCTTGGT <b>TACAGTCTCAGTT</b> AGTCACTCCCTAAG\ Bio	3
		TEG\	

[0204]

표 4

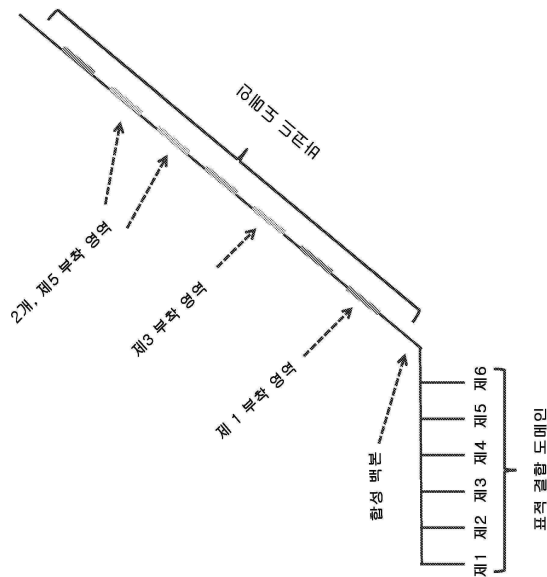
표 4: 10mer 미스매치에 대한 프로브 B 서열 (굵은체 표시; 서열 번호: 36 내지 서열 번호: 41)

10memis2A	GAGACAGTACCCTGGTCTAGGTATCTAATTCGTGGGTCGGGTACT
10memis2C	GAGACCGTACCGCTCATTTTGAACATACGATTGCGATTACGGAAA
10memis2G	GAGACGGTACCTTAAAGCTATCCACGAATGTCAAAAATGTGGTTT
10memis1G	GAGAGTGTACCCAATGCTTGCAATATGTATCCTGATCGTGCGTGC
10memis1A	GAGAATGTACCCTCATACCAATGTAAAGTATAGTTAACGCCCTGT
10memis1T	GAGATTGTACCCTACATATATAGGAAAAGGGAAGGTAGAAGAGCT

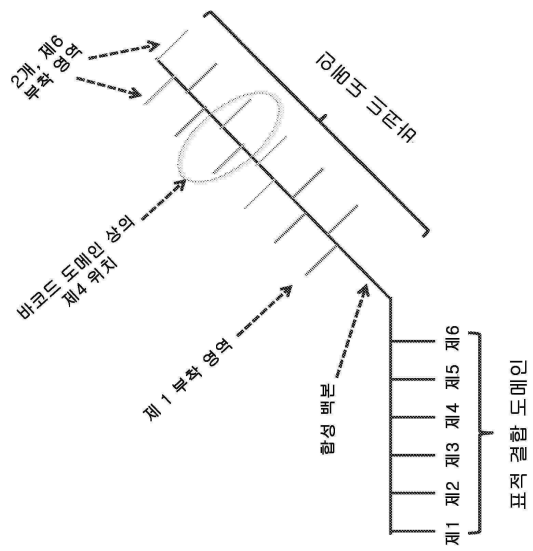
[0205]

도면

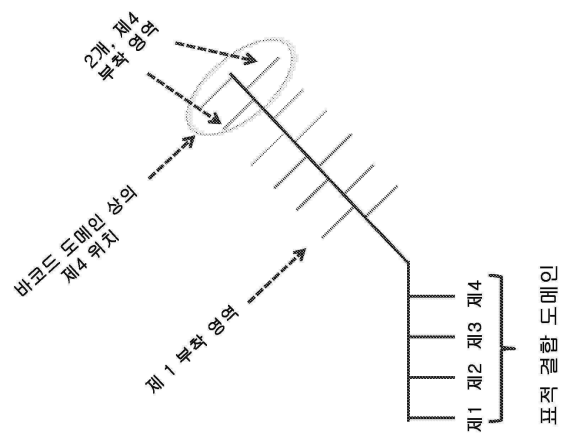
도면1



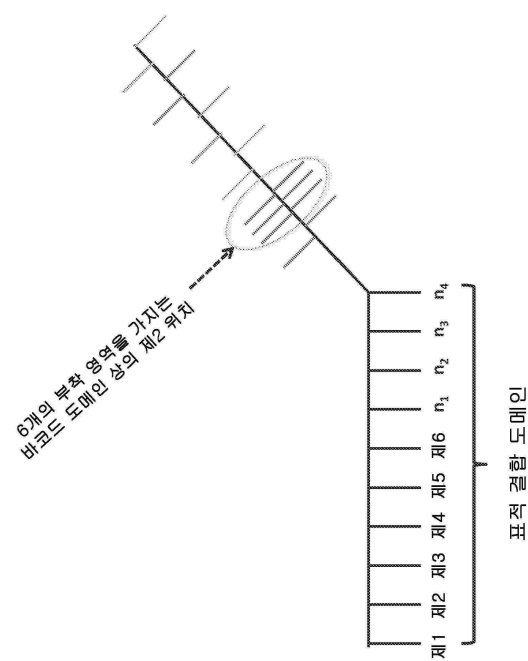
도면2



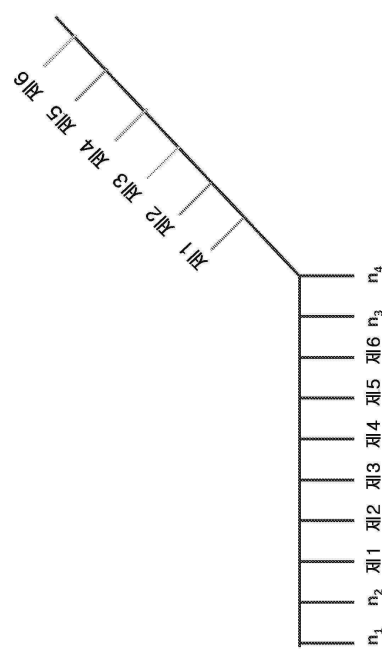
도면3



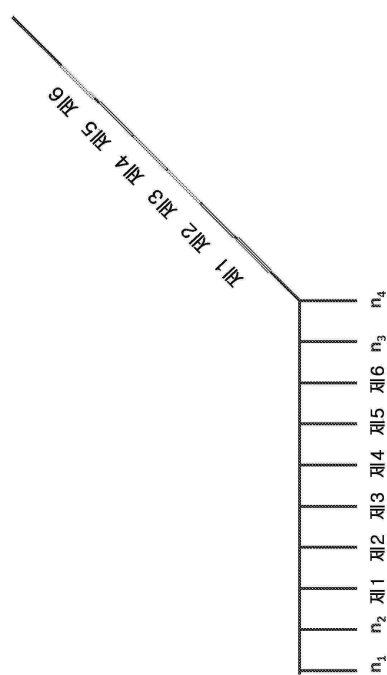
도면4



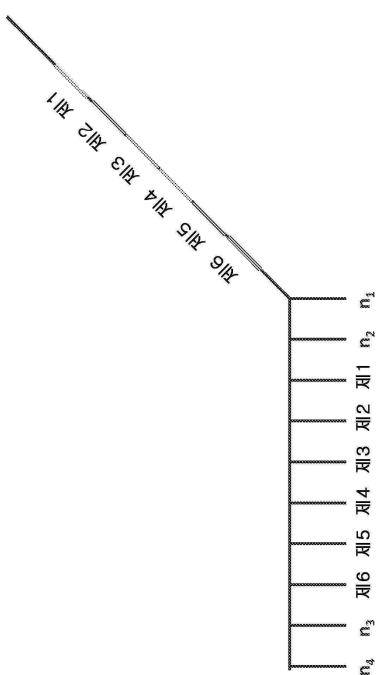
도면5



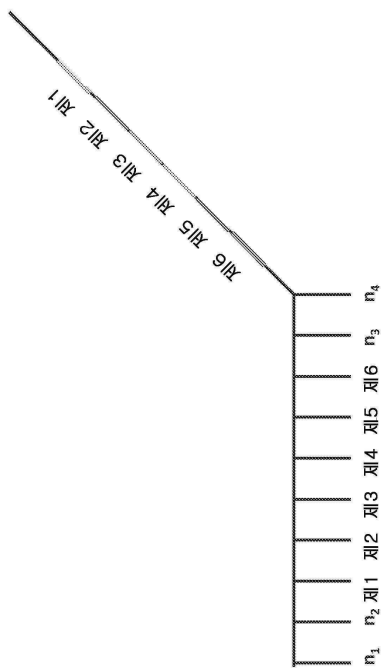
도면6a



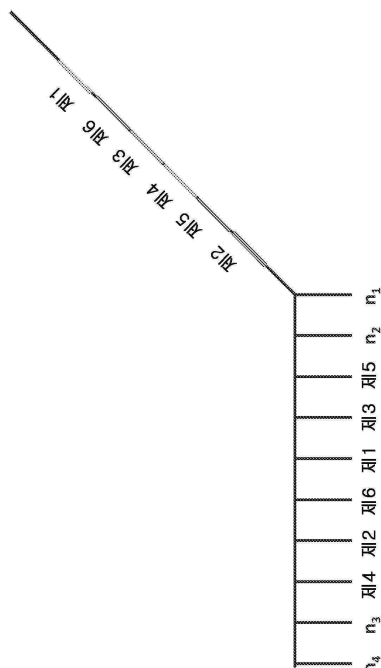
도면6b



도면6c

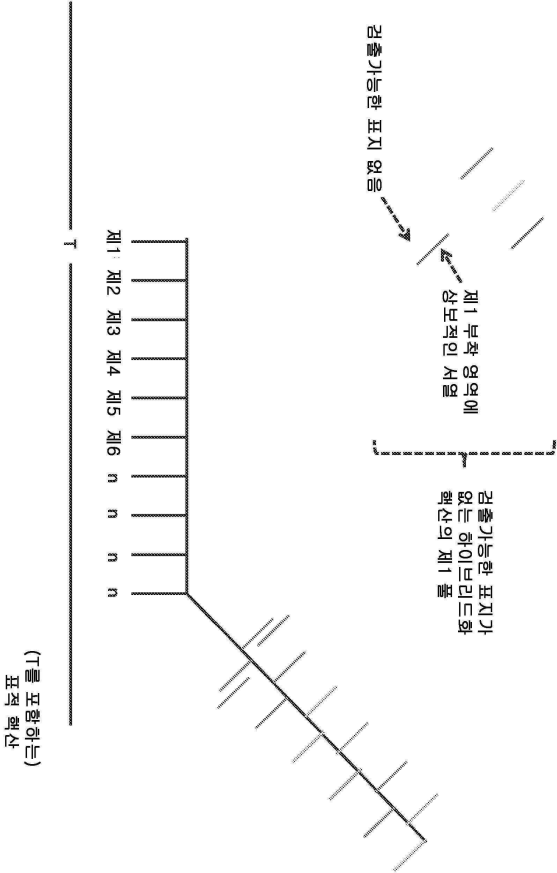


도면6d



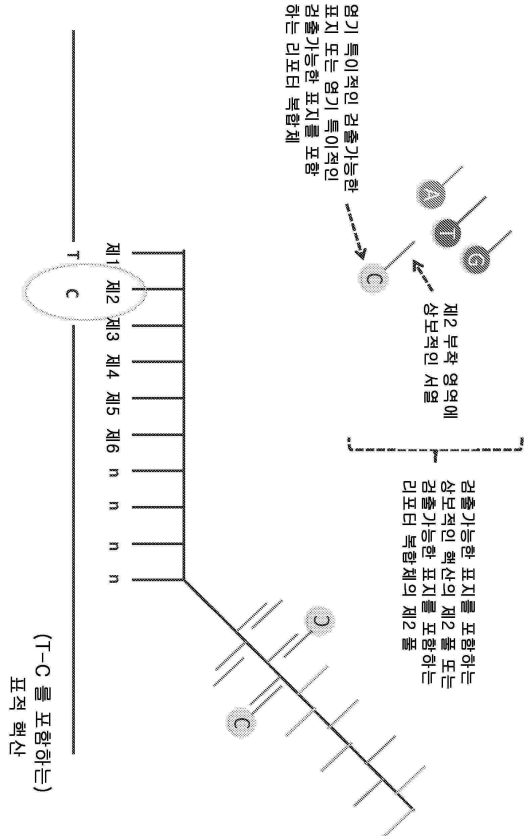


도면8b

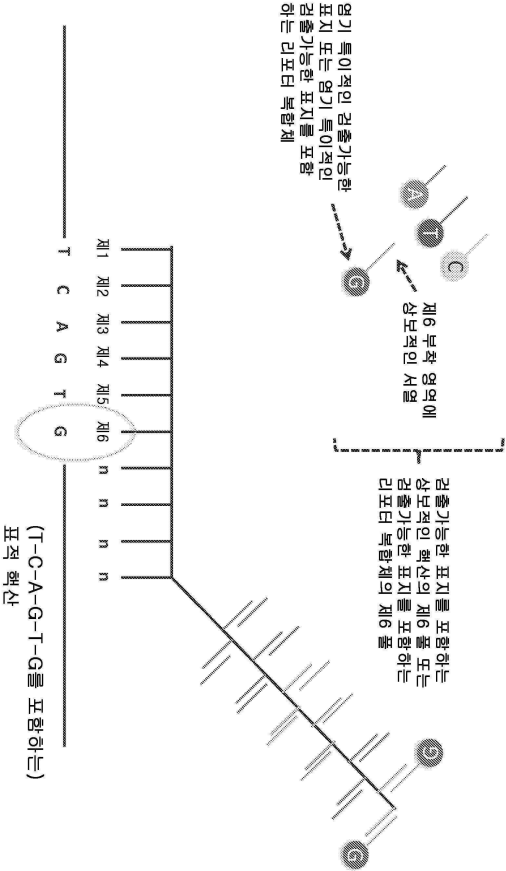




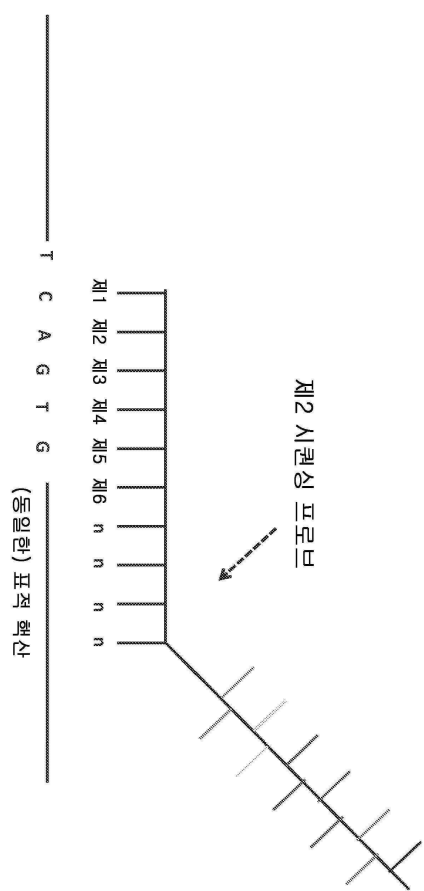
도면8c



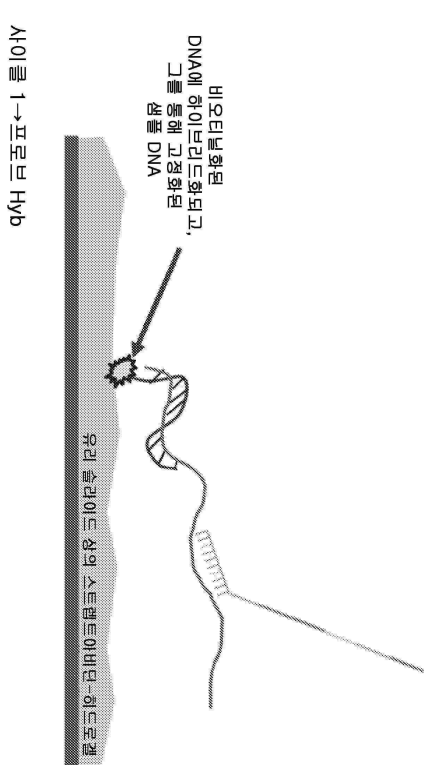
도면8d



도면8e

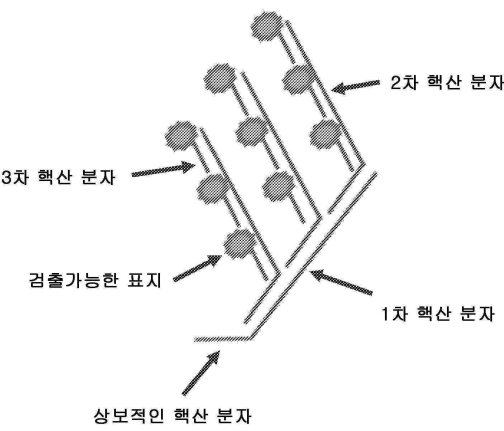


도면9a



도면9b

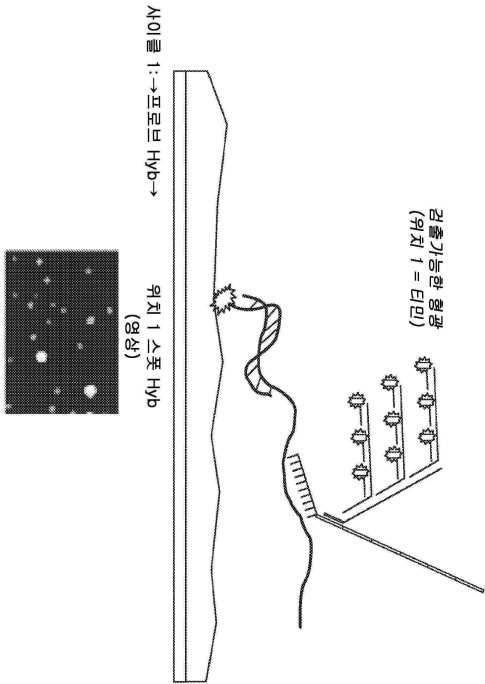
검출가능한 표지를 포함하는 리포터 복합체



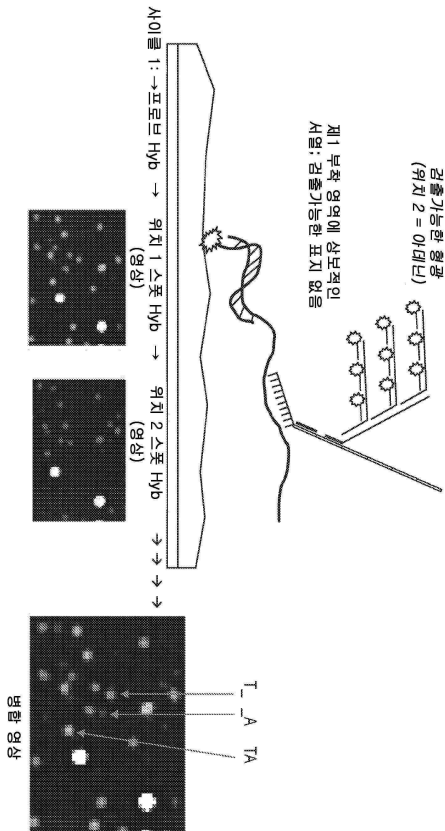
도면9c

<div>부착 영역-1 리포터 복합체</div> <div><div><div>위치</div><div>1_아데닌 (AR-1_A)</div></div><div><div>위치</div><div>1_아데닌 (AR-1_C)</div></div><div><div>위치</div><div>1_구아닌 (AR-1_G)</div></div><div><div>위치</div><div>1_티민 (AR-1_T)</div></div></div>	<div>부착 영역-2 리포터 복합체</div> <div><div><div>위치</div><div>2_아데닌 (AR-2_A)</div></div><div><div>위치</div><div>2_아데닌 (AR-2_C)</div></div><div><div>위치</div><div>2_구아닌 (AR-2_G)</div></div><div><div>위치</div><div>2_티민 (AR-2_T)</div></div></div>
<div>부착 영역-3 리포터 복합체</div> <div><div><div>위치</div><div>3_아데닌 (AR-3_A)</div></div><div><div>위치</div><div>3_아데닌 (AR-3_C)</div></div><div><div>위치</div><div>3_구아닌 (AR-3_G)</div></div><div><div>위치</div><div>3_티민 (AR-3_T)</div></div></div>	<div>부착 영역-4 리포터 복합체</div> <div><div><div>위치</div><div>4_아데닌 (AR-4_A)</div></div><div><div>위치</div><div>4_아데닌 (AR-4_C)</div></div><div><div>위치</div><div>4_구아닌 (AR-4_G)</div></div><div><div>위치</div><div>4_티민 (AR-4_T)</div></div></div>
<div>부착 영역-5 리포터 복합체</div> <div><div><div>위치</div><div>5_아데닌 (AR-5_A)</div></div><div><div>위치</div><div>5_아데닌 (AR-5_C)</div></div><div><div>위치</div><div>5_구아닌 (AR-5_G)</div></div><div><div>위치</div><div>5_티민 (AR-5_T)</div></div></div>	<div>부착 영역-6 리포터 복합체</div> <div><div><div>위치</div><div>6_아데닌 (AR-6_A)</div></div><div><div>위치</div><div>6_아데닌 (AR-6_C)</div></div><div><div>위치</div><div>6_구아닌 (AR-6_G)</div></div><div><div>위치</div><div>6_티민 (AR-6_T)</div></div></div>

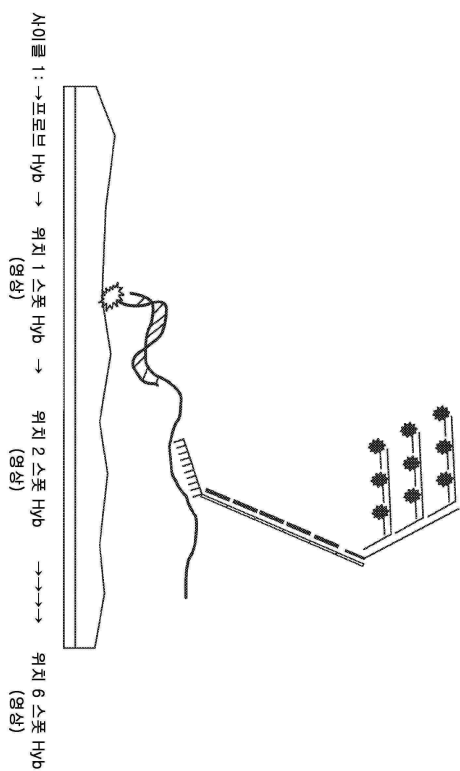
도면9d



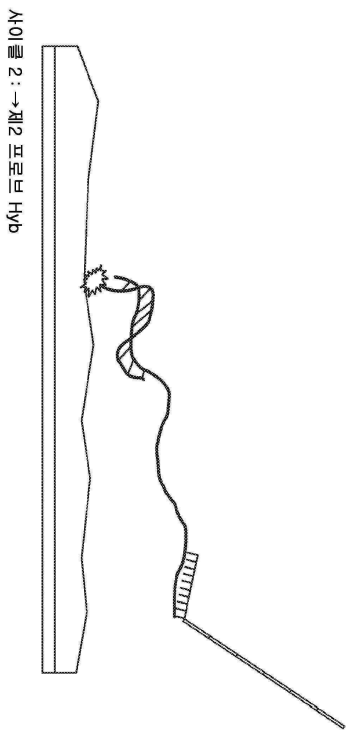
도면9e



도면9f

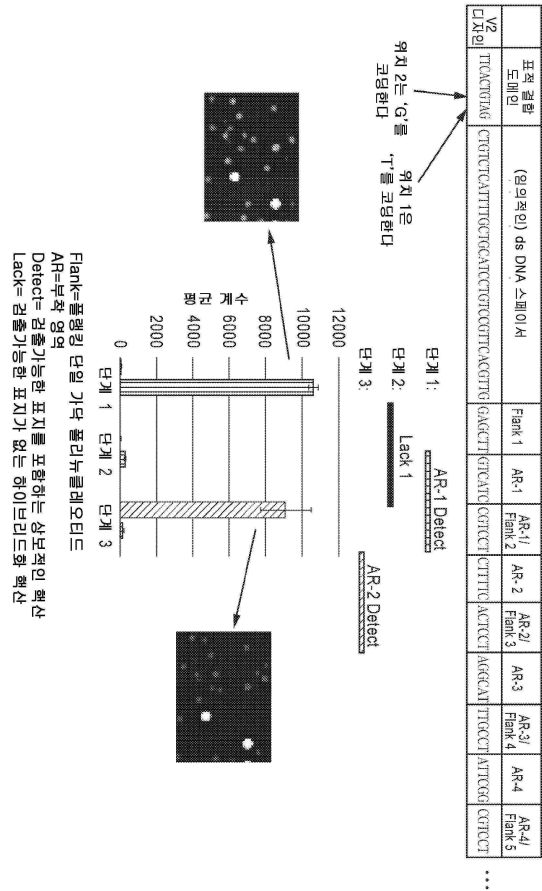


도면9g

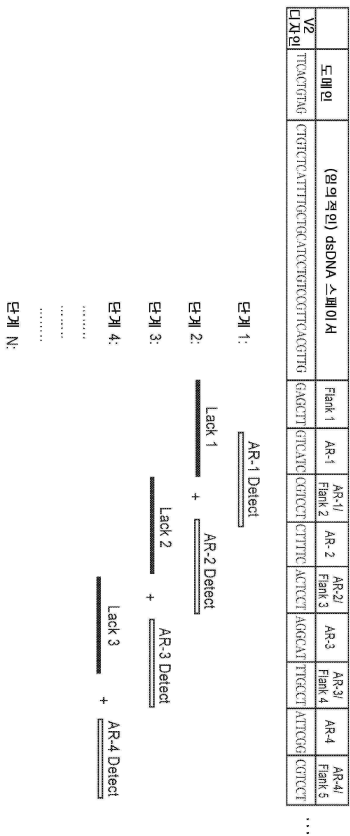




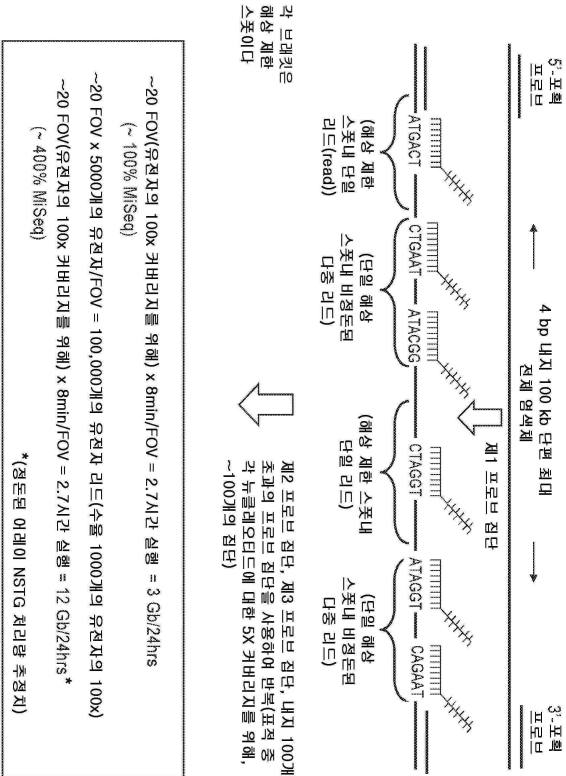
도면10



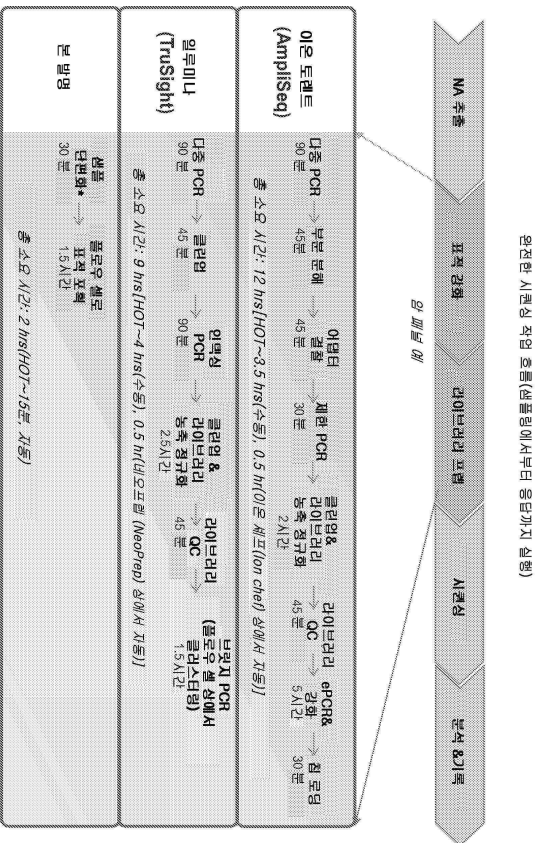
도면11



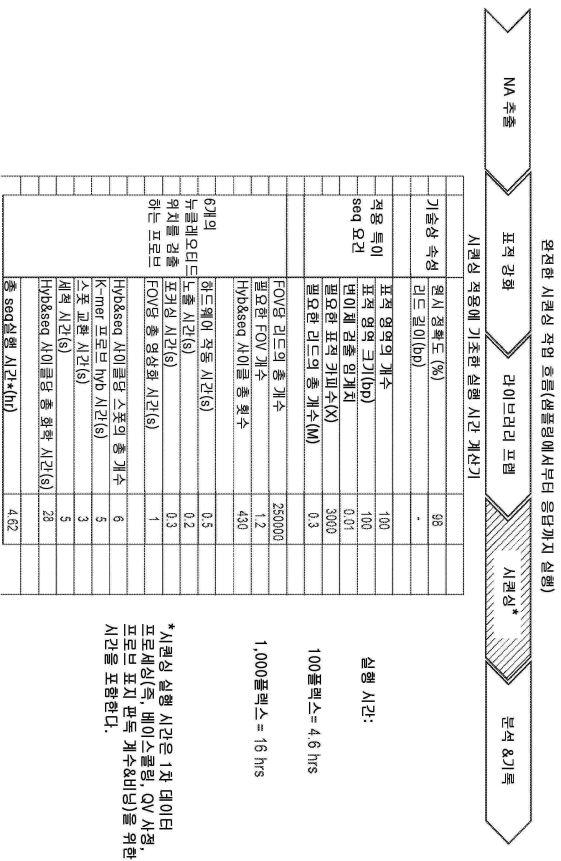
도면12



도면 13



도면14



도면15

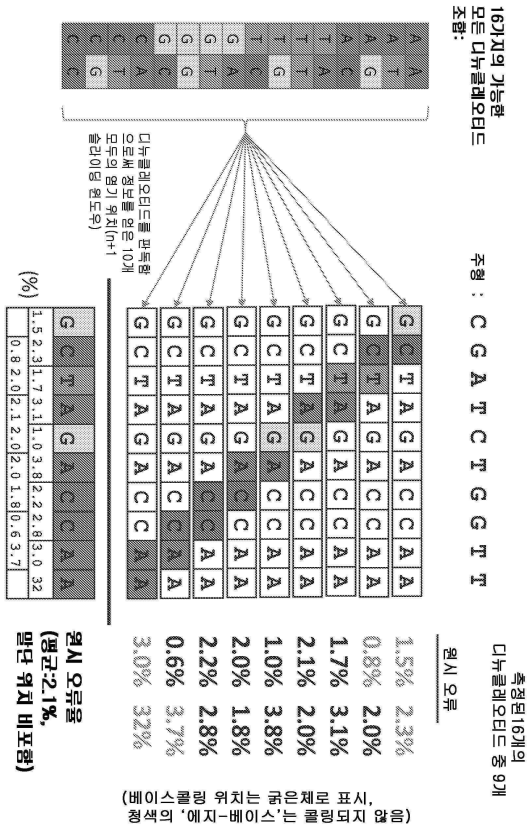
프로브 점유도에 기초한 사이클링 계산기

플로우 셀 상의 단일 분자 표적의 총 개수	30000	
'갭이 있는' 표적의 평균 길이(bp)	100	
Seq 사이클당 바코드 점유율(%)	0.75	
플로우 셀 상의 표적 중 >95%가 적어도 1회 '히트'하는 데 필요한 바코드 사이클 횟수	3	
위치당 염료 점유율(%)	0.8	
염료 위치의 총 개수	6	
Seq 사이클당 1bp 코딩된 리드의 총 개수	6	35
Seq 사이클당 2bp 코딩된 리드의 총 개수	23	346
Seq 사이클당 3bp 코딩된 리드의 총 개수	92	1843
Seq 사이클당 4bp 코딩된 리드의 총 개수	369	5530
Seq 사이클당 5bp 코딩된 리드의 총 개수	1475	8847
Seq 사이클당 6bp 코딩된 리드의 총 개수	5898	5898
Seq 사이클당 '이용가능한' 리드의 총 개수		22499
~3x커버리지로 주어진 단일 분자 표적을 커버하는 데 필요한 바코드 사이클 횟수	143	
~5x커버리지로 주어진 단일 분자 표적을 커버하는 데 필요한 바코드 사이클 횟수	238	
~3x커버리지로 플로우 셀 상의 모든 단일 분자 표적을 커버하는 데 필요한 바코드 사이클 횟수	430	
~5x커버리지로 플로우 셀 상의 모든 단일 분자 표적을 커버하는 데 필요한 바코드 사이클 횟수	713	

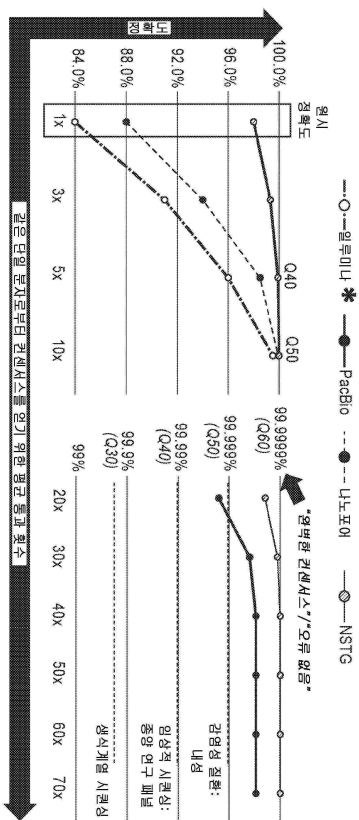
도면16

기기배이스				시간(hrs)	리드 개수	임상적 유용성
HiSeq 2500	900Gb	11일	30억	가능 서열 양 서열		
Next Seq	120GB	27hrs	400M	엑스 트랜스크립롬		
프로톤 PI	10Gb	2-4	70M	엑스 트랜스크립롬		
본 발명	3 나지 12 Gb	3	0.5M	표적화(Dx, cDx) + 다중 표본적용 + 샘플링에서 부담 응답까지, 긴 리드		
MiSeq	3.5Gb	24hrs	25M	트랜스크립롬 표적화(Dx)		
PGM 318	최대 2 Gb	7.3/4.4	4M	표적화		
PacBio RS II	2.2Gb/일	3 hrs	50k	나지 가능, 긴 리드, 높은 오류율		
PGM 314	0.1Gb	3.7/2.3	400k	표적화		

도면17



도면 18

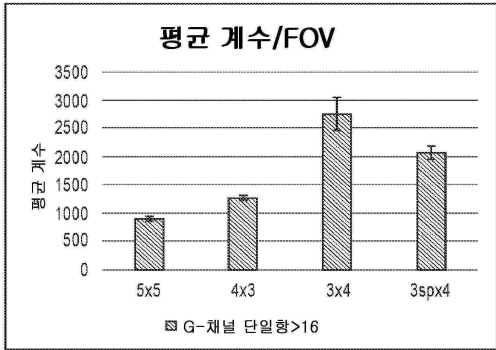








도면20b



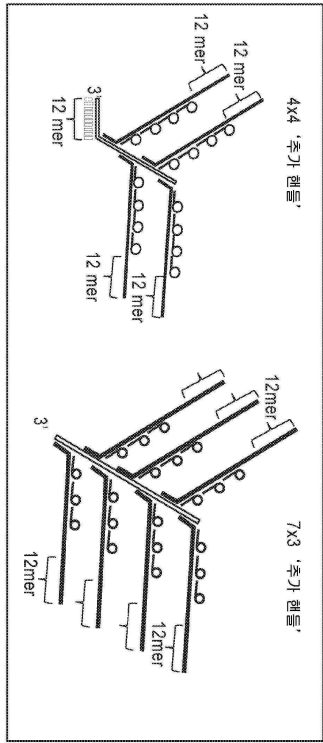
도면20c

검출가능한 표지를 포함하는 리포터 복합체를  
구성하기 위한 예시적인 레시피

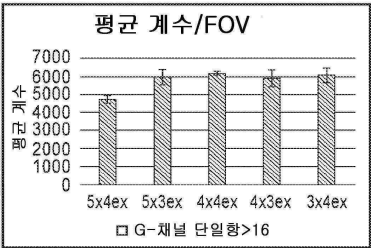
부피 (uL)	1차 핵산 분자 (10uM)	2 차 핵산 분자 (10uM)	3 차 핵산 분자 (100uM)	H2O
5x4	1	4.5	2.25	92.25
5x3	1	4.5	1.8	92.7
4x4	1.28	4.5	2.25	91.97
4x3	1.28	4.5	1.8	92.42
3x4	1.8	4.5	2.25	91.45

표에서의 색상은 좌측에 제시된 복합체의 다이어그램  
색상에 상응한다.

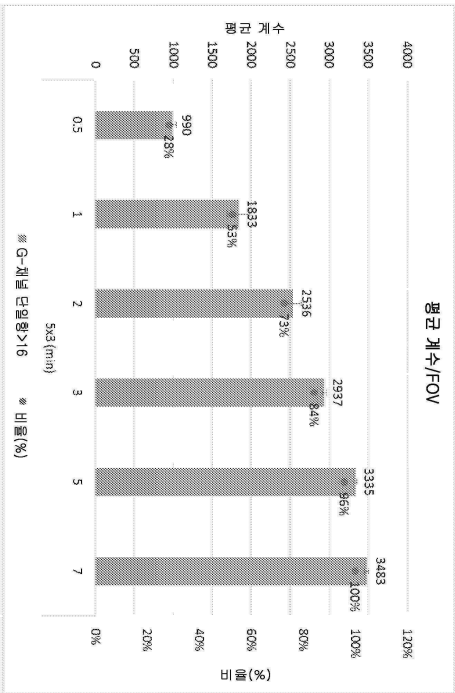
도면21a



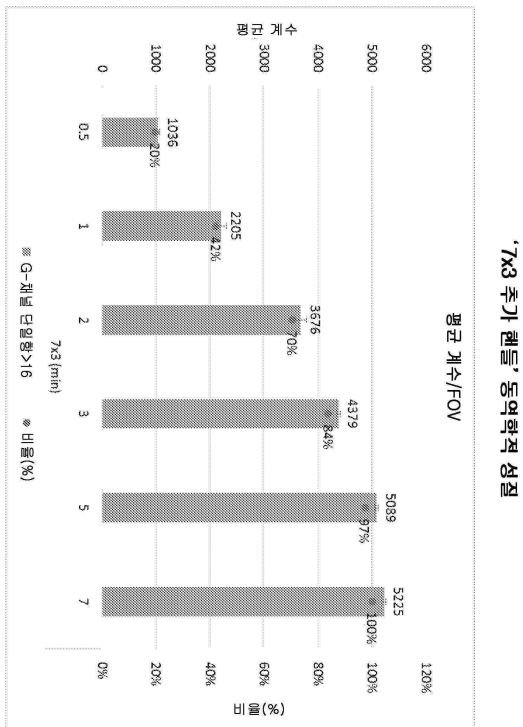
도면21b



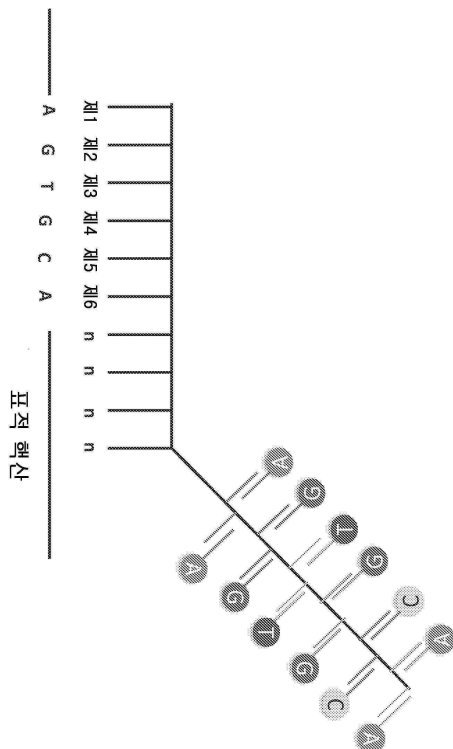
도면22a



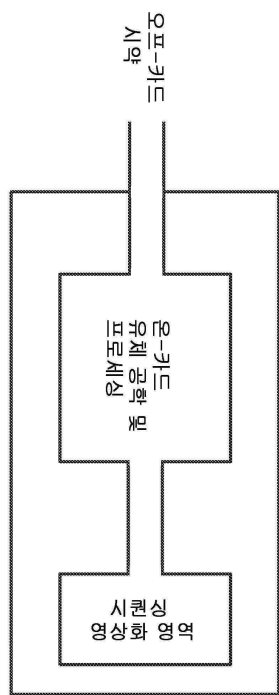
도면 22b



도면23



도면24



도면25

코드	서열	
YGBYGR	GAGACTGTAC - 오리지널	
BGYRGB	GAGACAGTAC	
YBRYGB	GAGACCGTAC	
RGBYGR	GAGACGGTAC	
YRBGYR	GAGAGTGTAC	
GRYGRB	GAGAATGTAC	
RGYBRG	GAGATTGTAC	

코드	계수	서열
YGBYGR	16074	GAGACTGTAC
RGYBRG	1132	GAGATTGTAC
YBRYGB	998	GAGACCGTAC
YRBGYR	784	GAGAGTGTAC
YBRYBG	712	
YBGRBY	699	
YGRYBG	692	
YBRGBR	674	
배경 임계치	717	

도면26

DNA	서열	코드	계수
12 mer	CTGAGACTGTAC	RYGBRG	80845
11 mer	TGAGACTGTAC	GBRYGB	24474
10 mer	GAGACTGTAC	YGBYGR	18339
9 mer	AGACTGTAC	YRBVRG	1667
"12 mer"의 거울상	CATGTCAGAGTC	GRBGYR	1473
"11 mer"의 거울상	CATGTCAGAGT	BGYRBBG	1384
8 mer	GACTGTAC	GBRYVBG	900
		배경 인계치	622

도면27

전장의 리포터  
20 pM 전장 (6개의 스폿) 33개의 계수/fov

단일 스폿 리포터

단일 스폿 [fM]	계수/FOV
20,000	포화
2,000	12000
200	308
<b>20</b>	<b>40</b>
2	11

20pM NS 리포터  $\approx$  20fM 단일 스폿

서열 목록

SEQUENCE LISTING

<110> NanoString Technologies, Inc.

BEECHEM, Joseph

KHAFIZOV, Rustem

<120> ENZYME- AND AMPLIFICATION-FREE SEQUENCING

<130> NATE-025/001  
 <150> US 62/082,883  
 <151> 2014-11-21  
 <160> 104  
 <170> PatentIn version 3.5  
 <210> 1  
 <211> 10  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 1  
 atacatctag 10  
 <210> 2  
 <211> 10  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220><223> Synthetic Polynucleotide  
 <400> 2  
 gatctacata 10  
 <210> 3  
 <211> 10  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 3  
 ttaggtaaag 10  
 <210> 4  
 <211> 10  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 4  
 tcttcattac 10  
 <210>



> 5

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 5

atgaatctac 10

<210> 6

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 6

tcaatgtatg 10

<210> 7

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 7

aattgagtac 10

<210> 8

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 8

atgttaatgg 10

<210> 9

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 9

aattaggatg	10
<210> 10	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 10	
ataatggatc	10
<210> 11	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 11	
taataaggtg	10
<210> 12	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 12	
tagttagagc	10
<210> 13	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 13	
atagagaagg	10
<210> 14	
<211> 10	
<212> DNA	
<213> Artificial Sequence	

<220><223> Synthetic Polynucleotide	
<400> 14	
ttgatgatac	10
<210> 15	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 15	
atagtgattc	10
<210> 16	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 16	
tataacgatg	10
<210> 17	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 17	
ttaagtttag	10
<210> 18	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 18	
atacgttatg	10
<210> 19	
<211> 10	
<212> DNA	

<213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 19  
 tgtactatag 10

<210> 20  
 <211> 10  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 20  
 ttaacaagtg 10

<210> 21  
 <211> 10  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 21  
 aactatgtac 10

<210> 22  
 <211> 10  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 22  
 taactatgac 10

<210> 23  
 <211> 10  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 23  
 actaatgttc 10

<210> 24

<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 24	
tcattgaatg	10
<210> 25	
<211> 14	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 25	
ctgtctcacc tctt	14
<210> 26	
<211> 26	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 26	
ctgtctcacc tcttgctgca tctgtg	26
<210> 27	
<211> 38	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 27	
ctgtctcacc tcttgctgca tctgtcggg tcacgttg	38
<210> 28	
<211> 36	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 28	

ctgtctcatc ttgctgcatc ctgtcggttc acgttg 36

<210> 29

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 29

ctgtctcatt ttgctgcatc ctgtcggttc acgttg 36

<210> 30

<211> 43

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 30

gactgtaccc acgcgatgac gttcgtcaag agtcgcataa tct 43

<210> 31

<211> 44

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 31

agactgtacc acaagaatcc ctgctagctg aaggagggtc aaac 44

<210> 32

<211> 45

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 32

gagactgtac cctacgtata tatccaagtg gttatgtccg acggc 45

<210> 33

<211> 46

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide  
 <400> 33  
 tgagactgta ccaccctcc aaacgcattc ttattggcaa atggaa 46  
 <210> 34  
 <211> 47  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 34  
 ctgagactgt acccggaat cggcatttcg cattcttagg atctaaa 47  
 <210> 35  
 <211> 44  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 35  
 caatgtgagt ctcttggtac agtctcagtt agtcactccc taag 44  
 <210> 36  
 <211> 45  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 36  
 gagacagtac cctggcttag gtatctaatt cgtgggtcgg gtact 45  
 <210> 37  
 <211> 45  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 37  
 gagaccgtac cgctcatttt gaacatacga ttgcgattac ggaaa 45  
 <210> 38  
 <211> 45



<212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 38  
 gagacgttac cttaaagcta tccacgaatg tcaaaaatgt ggttt 45  
 <210> 39  
 <211> 45  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 39  
 gagagtgtac ccaatgcttg cagtatgtat cctgacgtg cgtgc 45  
 <210> 40  
 <211> 45  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 40  
 gagaatgtac cctcatacca atgtaaagta tagttaacgc cctgt 45  
 <210> 41  
 <211> 45  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 41  
 gagattgtac cctacatata taggaaaagg gaaggtagaa gagct 45  
 <210> 42  
 <211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 42  
 cacgaacgtc ag 12  
 <210> 43

<211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 43  
 catcgcatgc ct 12

<210> 44  
 <211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 44  
 gtcattctct ac 12

<210> 45  
 <211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 45  
 gtcattcgct ac 12

<210> 46  
 <211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 46  
 gtcattgact ac 12

<210> 47  
 <211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 47

gtcatcttct ac	12
<210> 48	
<211> 12	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 48	
gtcatcacct ac	12
<210> 49	
<211> 12	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 49	
gtcatcactc ac	12
<210> 50	
<211> 12	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 50	
gtcatcttcg ac	12
<210> 51	
<211> 12	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 51	
gtcatcaact ac	12
<210> 52	
<211> 12	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	

<400> 52	
gtcatccgta ac	12
<210> 53	
<211> 12	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 53	
gtcatccgaa ac	12
<210> 54	
<211> 12	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 54	
gtcatcacia ac	12
<210> 55	
<211> 12	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 55	
gtcatcttgc ac	12
<210> 56	
<211> 12	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 56	
gtcatcttgc ct	12
<210> 57	
<211> 12	
<212> DNA	

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 57

gtcatccgtc ct 12

<210> 58

<211> 12

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 58

cttttcacct ct 12

<210> 59

<211> 12

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 59

cttttcctct ct 12

<210> 60

<211> 12

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 60

cttttcgact ct 12

<210> 61

<211> 12

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 61

cttttctgct ct 12

<210> 62

<211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 62  
 cttttctgta ct 12  
 <210> 63  
 <211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 63  
 cttttctgtg ct 12  
 <210> 64  
 <211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 64  
 cttttctgtc ct 12  
  
 <210> 65  
 <211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 65  
 cttttcactc ct 12  
 <210> 66  
 <211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 66  
 cttttcgttc ct 12

<210> 67  
 <211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 67  
 cttttcgtac ct 12

<210> 68  
 <211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 68  
 cttttccgtc ct 12

<210> 69  
 <211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 69  
 cttttctgac ct 12

<210> 70  
 <211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 70  
 aggcattgcga tg 12

<210> 71  
 <211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide

<400> 71	
aggcattgtg ct	12
<210> 72	
<211> 12	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 72	
aggcattgct ct	12
<210> 73	
<211> 12	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 73	
aggcatttct ac	12
<210> 74	
<211> 12	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 74	
aggcatacct ac	12
<210> 75	
<211> 12	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 75	
aggcatttgc ac	12
<210> 76	
<211> 12	
<212> DNA	
<213> Artificial Sequence	



<220><223> Synthetic Polynucleotide  
 <400> 76  
 aggcacgcgc ct 12

<210> 77  
 <211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 77  
 tcctgtcggc tc 12

<210> 78  
 <211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 78  
 gttcaatgct ct 12

<210> 79  
 <211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 79  
 attcggcgc ct 12

<210> 80  
 <211> 12  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic Polynucleotide  
 <400> 80  
 gatgcctgct ct 12

<210> 81  
 <211> 12

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 81

tttgcttgct ct 12

<210> 82

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 82

ttcactgtag ctgtctcatt ttgctgcac ctgtccgttc acgttggagc ttgtcatcgc 60

tcctcttttc actcctaggc atttgcctat tcggcgtcct 100

<210> 83

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 83

cgatctgggtt 10

<210> 84

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 84

cgatctgggtt 10

<210> 85

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 85

gctagaccaa	10
<210> 86	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 86	
gctggaccaa	10
<210> 87	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 87	
gctcgaccaa	10
<210> 88	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 88	
gcttgaccaa	10
<210> 89	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 89	
gagactgtac	10
<210> 90	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	

<400> 90	
gagacagtac	10
<210> 91	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 91	
gagaccgtac	10
<210> 92	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 92	
gagacggtac	10
<210> 93	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 93	
gagagtgtac	10
<210> 94	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 94	
gagaatgtac	10
<210> 95	
<211> 10	
<212> DNA	

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 95

gagattgtac 10

<210> 96

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 96

gagactgtac 10

<210> 97

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 97

gagattgtac 10

<210> 98

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 98

gagaccgtac 10

<210> 99

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic Polynucleotide

<400> 99

gagagtgtac 10

<210> 100

<211> 12	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 100	
ctgagactgt ac	12
<210> 101	
<211> 11	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 101	
tgagactgta c	11
<210> 102	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 102	
gagactgtac	10
<210> 103	
<211> 12	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 103	
catgtcagag tc	12
<210> 104	
<211> 11	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic Polynucleotide	
<400> 104	
catgtcagag t	11