(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0119710
(43) 공개일자 2017년10월27일

(51) 국제특허분류(Int. Cl.)
*C12Q 1/68* (2006.01)
(52) CPC특허분류
*C12Q 1/6869* (2013.01)
*C12Q 1/6806* (2013.01)
(21) 출원번호 10-2017-7026877
(22) 출원일자(국제) 2016년02월24일
심사청구일자 없음
(85) 번역문제출일자 2017년09월22일
(86) 국제출원번호 PCT/US2016/019382
(87) 국제공개번호 WO 2016/138148
국제공개일자 2016년09월01일
(30) 우선권주장
62/119,996 2015년02월24일 미국(US)
62/146,834 2015년04월13일 미국(US)

(71) 출원인
**10엑스 제노믹스, 인크.**
미국, 캘리포니아 94566, 플레젠튼, 스위트 401, 7068 콜 센터 파크웨이
(72) 발명자
**슈날-레빈 마이클**
미국 94566 캘리포니아주 플레전턴 스위트 401 콜 센터 파크웨이 7068
**자로즈 미르나**
미국 94566 캘리포니아주 플레전턴 스위트 401 콜 센터 파크웨이 7068
(74) 대리인
**유미특허법인**

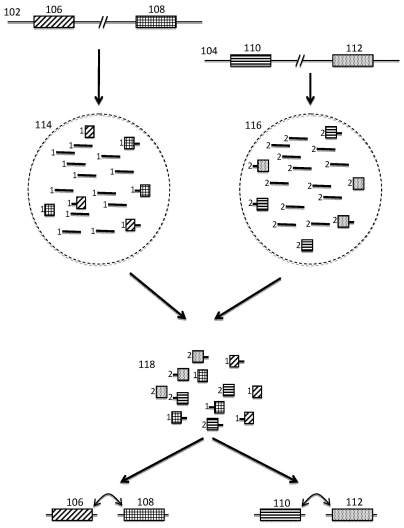
전체 청구항 수 : 총 93 항

(54) 발명의 명칭 **표적화된 핵산 서열 커버리지 방법**

## (57) 요 약

본 발명은 게놈의 표적화된 영역 유래의 서열 정보를 분석하기 위한 방법, 조성물 및 시스템에 관한 것이다. 상기 표적화된 영역은 참조 게놈 서열과 비교하여, 저조하게 특성규명된, 고도로 다형성된, 또는 분지된, 게놈의 영역을 포함할 수 있다.

*대 표 도* - 도1



(52) CPC특허분류
***C12Q 1/6874*** (2013.01)
*C12Q 2535/122* (2013.01)
*C12Q 2563/159* (2013.01)
*C12Q 2565/514* (2013.01)

# *명 세 서*

## *청구범위*

### 청구항 1

게놈의 하나 이상의 선택된 부분을 서열분석하는 방법으로서,

(a) 개시 게놈 재료를 제공하는 단계;

(b) 상기 개시 게놈 재료로부터 개별 핵산 분자를 개별 파티션으로 분배하여, 이로써 각 개별 파티션이 개별 핵산 분자를 함유하도록 하는 단계;

(c) 상기 개별 파티션 내 상기 개별 핵산 분자의 적어도 일부의 선택된 부분을 증폭시켜 증폭산물의 집단을 형성하는 단계;

(d) 상기 증폭산물의 집단을 바코딩하여 상기 증폭산물의 복수의 바코딩된 단편을 형성하는 단계로서, 소정의 개별 파티션 내에서의 단편 각각은 공통 바코드를 포함하고, 이로써 각 단편을 이것이 지향된 상기 개별 핵산 분자와 회합시키는, 상기 형성 단계;

(e) 상기 복수의 단편 유래의 서열 정보를 수득하여, 이로써 게놈의 하나 이상의 선택된 부분을 서열분석하는 단계를 포함하는, 방법.

### 청구항 2

청구항 1에 있어서, 상기 게놈의 상기 하나 이상의 선택된 부분이 상기 게놈의 고도로 다형성된 영역을 포함하는, 방법.

### 청구항 3

청구항 1 또는 2에 있어서, 상기 게놈의 상기 하나 이상의 선택된 부분의 상기 서열분석은 드노보(de-novo) 서열분석인, 방법.

### 청구항 4

청구항 1 내지 3 중 어느 한 항에 있어서, 상기 증폭이 적어도 3.5 메가염기쌍 (Mb)의 영역에 걸쳐 PCR 증폭을 포함하는, 방법.

### 청구항 5

청구항 1 내지 4 중 어느 한 항에 있어서, 상기 증폭이 적어도 3.0 Mb의 영역에 걸쳐 엇갈린(staggered) 다중 프라이머 쌍을 이용하는 PCR 증폭을 포함하는, 방법.

### 청구항 6

청구항 5에 있어서, 상기 다중 프라이머 쌍이 상기 프라이머 서열의 증폭을 예방하기 위해 우라실을 함유하는, 방법.

### 청구항 7

청구항 1 내지 6 중 어느 한 항에 있어서, 상기 수득 단계 (e)는 짧은 판독-길이 서열분석 반응 및 긴 판독-길이 서열분석 반응으로 이루어진 군으로부터 선택된 서열분석 반응을 포함하는, 방법.

### 청구항 8

청구항 7에 있어서, 상기 서열분석 반응이 짧은 판독, 고 정확도 서열분석 반응인, 방법.

### 청구항 9

청구항 1 내지 8 중 어느 한 항에 있어서, 수득 단계 (e)에서 생성된 상기 서열 정보는 이의 본래 개별 핵산의

분자 콘텍스트(molecular context)를 유지하는, 방법.

**청구항 10**

청구항 1 내지 9 중 어느 한 항에 있어서, 상기 수득 단계 (e) 이전에, 상기 복수의 단편은, 하기에 의하여 상기 게놈의 상기 하나 이상의 선택된 부분 중 적어도 일 부분을 포함하는 단편에 대해 추가로 농축되는, 방법:

(i) 상기 게놈의 상기 하나 이상의 선택된 부분에서 또는 상기 부분 근처에서 영역에 상보적인 프로브를 상기 단편에 하이브리드화시켜 프로브-단편 복합체를 형성하는 단계;

(ii) 프로브-단편 복합체를 고형 지지체의 표면 상에 포획하는 단계.

**청구항 11**

청구항 10에 있어서, 상기 고형 지지체는 비드를 포함하는, 방법.

**청구항 12**

청구항 1 내지 11 중 어느 한 항에 있어서, 상기 복수의 단편의 중첩 서열에 기반한, 추론된 콘티그 내 2 이상의 개별 핵산 분자를 연결하는 단계를 추가로 포함하고, 상기 추론된 콘티그는 적어도 10kb의 길이 N50을 포함하는, 방법.

**청구항 13**

청구항 12에 있어서, 상기 추론된 콘티그가 적어도 20 kb의 길이 N50를 포함하는, 방법.

**청구항 14**

청구항 12에 있어서, 상기 추론된 콘티그가 적어도 40 kb의 길이 N50를 포함하는, 방법.

**청구항 15**

청구항 12에 있어서, 상기 추론된 콘티그가 적어도 50 kb의 길이 N50를 포함하는, 방법.

**청구항 16**

청구항 12에 있어서, 상기 추론된 콘티그가 적어도 100 kb의 길이 N50를 포함하는, 방법.

**청구항 17**

청구항 12에 있어서, 상기 추론된 콘티그가 적어도 200 kb의 길이 N50를 포함하는, 방법.

**청구항 18**

청구항 1 내지 17 중 어느 한 항에 있어서, 상기 바코딩된 단편의 상기 바코드가 추가의 서열 분절을 추가로 포함하는, 방법.

**청구항 19**

청구항 18에 있어서, 상기 추가의 서열 분절은 하기로 이루어진 군으로부터 선택된 구성원 중 하나 이상을 포함하는, 방법: 프라이머, 부착 서열, 랜덤 n-량체 올리고뉴클레오티드, 우라실 핵염기를 포함하는 올리고뉴클레오티드.

**청구항 20**

청구항 1 내지 19 중 어느 한 항에 있어서, 상기 바코딩이 적어도 700,000개 바코드의 라이브러리로부터 선택된 바코드를 부착하는 것을 포함하는, 방법.

**청구항 21**

청구항 1 내지 20 중 어느 한 항에 있어서, 상기 개별 파티션 내 상기 증폭산물의 상기 바코딩된 단편이 상기 게놈의 상기 하나 이상의 선택된 부분의 약 100X 내지 5000X 커버리지를 나타내는, 방법.

**청구항 22**

청구항 1 내지 20 중 어느 한 항에 있어서, 상기 개별 파티션 내 상기 증폭산물의 상기 바코딩된 단편이 상기 게놈의 상기 하나 이상의 선택된 부분의 약 200X 내지 1000X 커버리지를 나타내는, 방법.

**청구항 23**

청구항 1 내지 20 중 어느 한 항에 있어서, 상기 개별 파티션 내 상기 증폭산물의 상기 바코딩된 단편이 상기 게놈의 상기 하나 이상의 선택된 부분의 적어도 1000X 커버리지를 나타내는, 방법.

**청구항 24**

청구항 1 내지 20 중 어느 한 항에 있어서, 상기 개별 파티션 내 상기 증폭산물의 상기 바코딩된 단편이 상기 게놈의 상기 하나 이상의 선택된 부분의 적어도 2000X 커버리지를 나타내는, 방법.

**청구항 25**

청구항 1 내지 20 중 어느 한 항에 있어서, 상기 개별 파티션 내 증폭산물의 상기 바코딩된 단편이 상기 게놈의 상기 하나 이상의 선택된 부분의 적어도 5000X 커버리지를 나타내는, 방법.

**청구항 26**

게놈 샘플의 하나 이상의 저조하게 특성규명된 부분 유래의 서열 정보를 수득하는 방법으로서,

(a) 개별 파티션 내 상기 게놈 샘플의 개별 제1 핵산 단편 분자를 제공하는 단계;

(b) 상기 개별 파티션 내 상기 개별 제1 핵산 단편 분자를 단편화(fragmenting)시켜 상기 개별 제1 핵산 단편 분자 각각으로부터 복수의 제2 단편을 형성하는 단계;

(c) 저조하게 특성규명된 상기 복수의 제2 단편의 선택된 영역을 증폭시켜 증폭산물의 집단을 형성하는 단계;

(d) 공통 바코드 서열을 각 개별 파티션 내 상기 증폭산물에 부착시켜, 이로써 상기 증폭산물 각각이, 이것이 함유되는 상기 개별 파티션에 부여가능하도록 하는 단계;

(e) 상기 증폭산물의 서열을 식별하여, 이로써 상기 게놈 샘플의 하나 이상의 저조하게 특성규명된 부분 유래의 서열 정보를 수득하는 단계를 포함하는, 방법.

**청구항 27**

청구항 26에 있어서, 상기 증폭이 적어도 3.5 메가염기쌍 (Mb)의 영역에 걸쳐 PCR 증폭을 포함하는, 방법.

**청구항 28**

청구항 26 또는 27에 있어서, 상기 증폭이 적어도 3.0 Mb의 영역에 걸쳐 엇갈린 다중 프라이머 쌍을 이용하는 PCR 증폭을 포함하는, 방법.

**청구항 29**

청구항 28에 있어서, 상기 다중 프라이머 쌍이 상기 프라이머 서열의 증폭을 예방하기 위해 우라실을 함유하는, 방법.

**청구항 30**

청구항 26 내지 29 중 어느 한 항에 있어서, 상기 식별 단계 (e)는 짧은 판독-길이 서열분석 반응 및 긴 판독-길이 서열분석 반응으로 이루어진 군으로부터 선택된 서열분석 반응을 포함하는, 방법.

**청구항 31**

청구항 30에 있어서, 상기 서열분석 반응이 짧은 판독, 고 정확도 서열분석 반응인, 방법.

**청구항 32**

청구항 26 내지 31 중 어느 한 항에 있어서, 상기 식별 단계 (e)가 상기 증폭산물의 상기 서열의 상기 분자 콘

텍스트를 보존하여, 이로써 상기 식별이 상기 동일한 개별 제1 핵산 단편 분자로부터 유도된 증폭산물을 식별하는 것을 추가로 포함하도록 하는, 방법.

**청구항 33**

청구항 26 내지 32 중 어느 한 항에 있어서, 상기 복수의 제2 단편의 중첩 서열에 기반한, 추론된 콘티그 내 2 이상의 상기 개별 제1 단편 분자를 연결하는 단계를 추가로 포함하고, 상기 추론된 콘티그는 적어도 10kb의 길이 N50을 포함하는, 방법.

**청구항 34**

청구항 33에 있어서, 상기 추론된 콘티그가 적어도 20 kb의 길이 N50를 포함하는, 방법.

**청구항 35**

청구항 33에 있어서, 상기 추론된 콘티그가 적어도 40 kb의 길이 N50를 포함하는, 방법.

**청구항 36**

청구항 33에 있어서, 상기 추론된 콘티그가 적어도 50 kb의 길이 N50를 포함하는, 방법.

**청구항 37**

청구항 33에 있어서, 상기 추론된 콘티그가 적어도 100 kb의 길이 N50를 포함하는, 방법.

**청구항 38**

청구항 33에 있어서, 상기 추론된 콘티그가 적어도 200 kb의 길이 N50를 포함하는, 방법.

**청구항 39**

청구항 26 내지 38 중 어느 한 항에 있어서, 상기 바코드 서열이 추가의 서열 분절을 추가로 포함하는, 방법.

**청구항 40**

청구항 39에 있어서, 상기 추가의 서열 분절은 하기로 이루어진 군으로부터 선택된 구성원 중 하나 이상을 포함하는, 방법: 프라이머, 부착 서열, 랜덤 n-량체 올리고뉴클레오티드, 우라실 핵염기를 포함한 올리고뉴클레오티드.

**청구항 41**

청구항 26 내지 40 중 어느 한 항에 있어서, 상기 부착 단계 (d)가 적어도 700,000 바코드의 라이브러리로부터 선택된 바코드를 부착하는 것을 포함하는, 방법.

**청구항 42**

청구항 26 내지 41 중 어느 한 항에 있어서, 상기 각 개별 파티션 내 상기 게놈 샘플이 단일 세포 유래의 게놈 DNA를 포함하는, 방법.

**청구항 43**

청구항 26 내지 41 중 어느 한 항에 있어서, 각 개별 파티션이 상이한 염색체 유래의 게놈 DNA를 포함하는, 방법.

**청구항 44**

청구항 26 내지 43 중 어느 한 항에 있어서, 상기 개별 파티션이 에멀젼 내 액적을 포함하는, 방법.

**청구항 45**

청구항 26 내지 44 중 어느 한 항에 있어서, 상기 식별 단계 (e) 이전에, 상기 증폭산물이 추가로 증폭되어 이로써 상기 수득한 증폭 생성물이 부분적 또는 완전한 헤어핀 구조를 형성할 수 있도록 하는, 방법.

**청구항 46**

청구항 26 내지 45 중 어느 한 항에 있어서, 상기 개별 파티션 내 상기 바코딩된 증폭산물이 상기 게놈의 상기 하나 이상의 저조하게 특성규명된 부분의 약 100X 내지 5000X 커버리지를 나타내는, 방법.

**청구항 47**

청구항 26 내지 45 중 어느 한 항에 있어서, 상기 개별 파티션 내 상기 바코딩된 증폭산물이 상기 게놈의 상기 하나 이상의 저조하게 특성규명된 부분의 약 200X 내지 1000X 커버리지를 나타내는, 방법.

**청구항 48**

청구항 26 내지 45 중 어느 한 항에 있어서, 상기 개별 파티션 내 상기 바코딩된 증폭산물이 상기 게놈의 상기 하나 이상의 저조하게 특성규명된 부분의 적어도 1000X 커버리지를 나타내는, 방법.

**청구항 49**

청구항 26 내지 45 중 어느 한 항에 있어서, 상기 개별 파티션 내 상기 바코딩된 증폭산물이 상기 게놈의 상기 하나 이상의 저조하게 특성규명된 부분의 적어도 2000X 커버리지를 나타내는, 방법.

**청구항 50**

청구항 26 내지 45 중 어느 한 항에 있어서, 상기 개별 파티션 내 상기 바코딩된 증폭산물이 상기 게놈의 상기 하나 이상의 저조하게 특성규명된 부분의 적어도 5000X 커버리지를 나타내는, 방법.

**청구항 51**

분자 콘텍스트를 유지하는 한편 게놈 샘플의 하나 이상의 부분 유래의 서열 정보를 수득하는 방법으로서,

(a) 개시 게놈 재료를 제공하는 단계;

(b) 상기 개시 게놈 재료로부터 개별 핵산 분자를 개별 파티션으로 분배하여, 이로써 각 개별 파티션이 제1 개별 핵산 분자를 함유하도록 하는 단계;

(c) 상기 게놈의 상기 하나 이상의 선택된 부분의 적어도 일 부분을 포함하는 단편에 대해 농축된 집단을 제공하는 단계;

(d) 각 개별 파티션 내 상기 단편에 공통 바코드 서열을 부착시켜, 상기 단편 각각이 이것이 함유된 상기 개별 파티션에 부여가능하도록 하는 단계;

(e) 상기 단편 유래의 서열 정보를 수득하여, 이로써 분자 콘텍스트를 유지하는 한편 상기 게놈 샘플의 하나 이상의 표적화된 부분을 서열분석하는 단계를 포함하는, 방법.

**청구항 52**

청구항 51에 있어서, 상기 제공 단계 (c)가 상기 게놈의 상기 하나 이상의 선택된 부분 유래의 서열을 함유하는 상기 단편의 적어도 일 부분의 PCR 증폭을 포함하는, 방법.

**청구항 53**

청구항 51 또는 52에 있어서, 상기 게놈의 상기 하나 이상의 선택된 부분이 길이가 적어도 3.0 Mb인 상기 게놈의 인접 영역을 포함하는, 방법.

**청구항 54**

청구항 51 내지 53 중 어느 한 항에 있어서, 상기 수득 단계 (e)는 짧은 판독-길이 서열분석 반응 및 긴 판독-길이 서열분석 반응으로 이루어진 군으로부터 선택된 서열분석 반응을 포함하는, 방법.

**청구항 55**

청구항 54에 있어서, 상기 서열분석 반응이 짧은 판독, 고 정확도 서열분석 반응인, 방법.

**청구항 56**

청구항 51 내지 55 중 어느 한 항에 있어서, 상기 식별 단계 (e)가 상기 단편의 상기 서열의 상기 분자 콘텍스트를 보존하여, 이로써 상기 식별이 상기 동일한 제1 개별 핵산 분자로부터 유도된 단편을 식별하는 것을 추가로 포함하도록 하는, 방법.

**청구항 57**

청구항 51 내지 56 중 어느 한 항에 있어서, 상기 복수의 제2 단편의 중첩 서열에 기반한, 추론된 콘티그 내 2 이상의 상기 개별 제1 단편 분자를 연결하는 단계를 추가로 포함하고, 상기 추론된 콘티그는 적어도 10kb의 길이 N50을 추가로 포함하는, 방법.

**청구항 58**

청구항 57에 있어서, 상기 추론된 콘티그가 적어도 20 kb의 길이 N50를 포함하는, 방법.

**청구항 59**

청구항 57에 있어서, 상기 추론된 콘티그가 적어도 40 kb의 길이 N50를 포함하는, 방법.

**청구항 60**

청구항 57에 있어서, 상기 추론된 콘티그가 적어도 50 kb의 길이 N50를 포함하는, 방법.

**청구항 61**

청구항 57에 있어서, 상기 추론된 콘티그가 적어도 100 kb의 길이 N50를 포함하는, 방법.

**청구항 62**

청구항 57에 있어서, 상기 추론된 콘티그가 적어도 200 kb의 길이 N50를 포함하는, 방법.

**청구항 63**

청구항 51 내지 62 중 어느 한 항에 있어서, 상기 바코드 서열이 추가의 서열 분절을 추가로 포함하는, 방법.

**청구항 64**

청구항 63에 있어서, 상기 추가의 서열 분절은 하기로 이루어진 군으로부터 선택된 구성원 중 하나 이상을 포함하는, 방법: 프라이머, 부착 서열, 랜덤 n-량체 올리고뉴클레오티드, 우라실 핵염기를 포함한 올리고뉴클레오티드.

**청구항 65**

청구항 51 내지 64 중 어느 한 항에 있어서, 상기 부착 단계 (d)가 적어도 700,000 바코드의 라이브러리로부터 선택된 바코드를 부착하는 것을 포함하는, 방법.

**청구항 66**

청구항 51 내지 65 중 어느 한 항에 있어서, 상기 각 개별 파티션 내 상기 게놈 재료가 단일 세포 유래의 게놈 DNA를 포함하는, 방법.

**청구항 67**

청구항 51 내지 65 중 어느 한 항에 있어서, 각 개별 파티션이 상이한 염색체 유래의 게놈 DNA를 포함하는, 방법.

**청구항 68**

청구항 51 내지 67 중 어느 한 항에 있어서, 상기 개별 파티션이 에멀젼 내 액적을 포함하는, 방법.

**청구항 69**

청구항 51 내지 68 중 어느 한 항에 있어서, 상기 수득 단계 (e) 이전에, 상기 단편이 추가로 증폭되어 이로써 상기 수득한 증폭 생성물이 부분적 또는 완전한 헤어핀 구조를 형성할 수 있도록 하는, 방법.

**청구항 70**

청구항 51 내지 69 중 어느 한 항에 있어서, 상기 제공 단계 (c)가 상기 게놈의 상기 하나 이상의 선택된 부분의 적어도 일 부분을 포함한 상기 단편의 PCR 증폭을 포함하는, 방법.

**청구항 71**

청구항 51 내지 70 중 어느 한 항에 있어서, 상기 개별 파티션 내 상기 바코딩된 단편이 상기 게놈의 상기 하나 이상의 선택된 부분의 약 100X 내지 5000X 커버리지를 나타내는, 방법.

**청구항 72**

청구항 51 내지 70 중 어느 한 항에 있어서, 상기 개별 파티션 내 상기 바코딩된 단편이 상기 게놈의 상기 하나 이상의 선택된 부분의 약 200X 내지 1000X 커버리지를 나타내는, 방법.

**청구항 73**

청구항 51 내지 70 중 어느 한 항에 있어서, 상기 개별 파티션 내 상기 증폭산물의 상기 바코딩된 단편이 상기 게놈의 상기 하나 이상의 선택된 부분의 적어도 1X 커버리지를 나타내는, 방법.

**청구항 74**

청구항 51 내지 70 중 어느 한 항에 있어서, 상기 개별 파티션 내 상기 바코딩된 단편이 상기 게놈의 상기 하나 이상의 선택된 부분의 적어도 2000X 커버리지를 나타내는, 방법.

**청구항 75**

청구항 51 내지 70 중 어느 한 항에 있어서, 상기 개별 파티션 내 상기 바코딩된 단편이 상기 게놈의 상기 하나 이상의 선택된 부분의 적어도 5000X 커버리지를 나타내는, 방법.

**청구항 76**

분자 콘텍스트를 유지하는 한편 게놈 샘플의 하나 이상의 부분 유래의 서열 정보를 수득하는 방법으로서,

(a) 개시 게놈 재료를 제공하는 단계;

(b) 상기 개시 게놈 재료로부터 개별 핵산 분자를 개별 파티션으로 분배하여, 이로써 각 개별 파티션이 제1 개별 핵산 분자를 함유하도록 하는 단계;

(c) 상기 게놈의 상기 하나 이상의 선택된 부분의 적어도 일 부분을 포함하는 상기 단편의 서열에 대해 농축된, 상기 개별 파티션 중 적어도 일부 내의 집단을 제공하는 단계;

(d) 각 개별 파티션 내 상기 단편에 공통 바코드 서열을 부착시켜, 상기 단편 각각이 이것이 함유된 상기 개별 파티션에 부여가능하도록 하는 단계;

(e) 상기 게놈의 상기 하나 이상의 선택된 부분의 적어도 일 부분을 포함하는 단편을 함유한 개별 파티션을, 상기 게놈의 상기 하나 이상의 선택된 부분을 포함하는 단편을 함유하지 않은 개별 파티션으로부터 분리시키는 단계;

(f) 상기 게놈의 상기 하나 이상의 선택된 부분 중 적어도 일 부분을 포함하는 상기 단편 유래의 서열 정보를 수득하여, 이로써 분자 콘텍스트를 유지하는 한편 상기 게놈 샘플 중 하나 이상의 표적화된 부분을 서열분석하는 단계를 포함하는, 방법.

**청구항 77**

청구항 76에 있어서, 상기 제공 단계 (c)가 상기 게놈의 상기 하나 이상의 선택된 부분의 적어도 일 부분을 포

함하는 상기 단편의 지향된 PCR 증폭을 포함하여, 상기 게놈의 상기 하나 이상의 선택된 부분의 적어도 일 부분을 포함하는 증폭산물의 집단을 생산하는, 방법.

**청구항 78**

청구항 77에 있어서, 상기 제공 단계 (c)가 상기 증폭산물에 검출가능한 표지를 부착하는 것을 추가로 포함하는, 방법.

**청구항 79**

청구항 78에 있어서, 상기 분리 단계 (e)가, 상기 파티션으로부터 상기 검출가능한 표지에서 신호를 방출하는 상기 파티션을, 상기 신호 없이, 분류하는 것을 포함하는, 방법.

**청구항 80**

청구항 78 또는 79에 있어서, 상기 검출가능한 표지는 형광 분자를 포함하는, 방법.

**청구항 81**

청구항 76 내지 80 중 어느 한 항에 있어서, 상기 수득 단계 (f) 이전에, 상기 개별 파티션이 조합되고, 그리고 상기 단편이 함께 풀링되는, 방법.

**청구항 82**

청구항 76 내지 81 중 어느 한 항에 있어서, 상기 식별 단계 (f)가 상기 단편의 상기 서열의 상기 분자 콘텍스트를 보존하여, 이로써 상기 식별이 상기 동일한 제1 개별 핵산 분자로부터 유도된 단편을 식별하는 것을 추가로 포함하도록 하는, 방법.

**청구항 83**

청구항 76 내지 82 중 어느 한 항에 있어서, 상기 식별 단계 (f)는 짧은 판독-길이 서열분석 반응 및 긴 판독-길이 서열분석 반응으로 이루어진 군으로부터 선택된 서열분석 반응을 포함하는, 방법.

**청구항 84**

청구항 83에 있어서, 상기 서열분석 반응이 짧은 판독, 고 정확도 서열분석 반응인, 방법.

**청구항 85**

청구항 76 내지 84 중 어느 한 항에 있어서, 상기 개별 파티션이 에멀젼 내 액적을 포함하는, 방법.

**청구항 86**

청구항 76 내지 85 중 어느 한 항에 있어서, 상기 개별 파티션 내 상기 바코딩된 단편이 상기 게놈의 상기 하나 이상의 선택된 부분의 약 100X 내지 5000X 커버리지를 나타내는, 방법.

**청구항 87**

청구항 76 내지 85 중 어느 한 항에 있어서, 상기 개별 파티션 내 상기 바코딩된 단편이 상기 게놈의 상기 하나 이상의 선택된 부분의 약 200X 내지 1000X 커버리지를 나타내는, 방법.

**청구항 88**

청구항 76 내지 85 중 어느 한 항에 있어서, 상기 개별 파티션 내 상기 증폭산물의 상기 바코딩된 단편이 상기 게놈의 상기 하나 이상의 선택된 부분의 적어도 1000X 커버리지를 나타내는, 방법.

**청구항 89**

청구항 76 내지 85 중 어느 한 항에 있어서, 상기 개별 파티션 내 상기 바코딩된 단편이 상기 게놈의 상기 하나 이상의 선택된 부분의 적어도 2000X 커버리지를 나타내는, 방법.

**청구항 90**

청구항 76 내지 85 중 어느 한 항에 있어서, 상기 개별 파티션 내 상기 바코딩된 단편이 상기 게놈의 상기 하나 이상의 선택된 부분의 적어도 5000X 커버리지를 나타내는, 방법.

**청구항 91**

분자 콘텍스트를 유지하는 한편 게놈 샘플의 하나 이상의 부분 유래의 서열 정보를 수득하는 방법으로서,

(a) 개시 게놈 재료를 제공하는 단계;

(b) 개별 핵산 분자를 상기 게놈 재료로부터 분리시켜 분리된 개별 핵산 분자를 형성하는 단계;

(c) 상기 분리된 개별 핵산 분자 유래의 상기 게놈의 상기 하나 이상의 선택된 부분 중 적어도 일 부분을 포함하는 단편에 대해 농축된 집단을 제공하는 단계로서, 적어도 복수의 상기 단편은 이들이 유도된 상기 개별 핵산 분자에 부여가능한, 상기 제공 단계;

(d) 상기 단편 유래의 서열 정보를 수득하여, 이로써 분자 콘텍스트를 유지하는 한편 상기 게놈 샘플의 하나 이상의 표적화된 부분을 서열분석하는 단계를 포함하는, 방법.

**청구항 92**

청구항 91에 있어서, 상기 분리 단계 (b)가 하나 이상의 개별 핵산 분자를 개별 파티션에 분배하는 것을 포함하는, 방법.

**청구항 93**

청구항 91 또는 92에 있어서, 상기 제공 단계 (c)가 상기 게놈의 상기 하나 이상의 선택된 부분의 적어도 일 부분을 포함한 상기 핵산 분자의 지향된 증폭을 포함하여, 상기 게놈의 상기 하나 이상의 선택된 부분의 적어도 일 부분을 포함한 증폭산물의 집단을 생산하는, 방법.

## *발명의 설명*

### *기 술 분 야*

[0001] **관련 출원에 대한 교차참조**

[0002] 본 출원은 2015년 4월 13일 월요일 출원된 미국 가특허원 제62/146,834호, 및 2015년 2월 24일 화요일 출원된 미국 가특허원 일련번호 제62/119,996호의 이점을 청구하며, 이는, 모든 목적을 위해 이의 전문이 참고로 본원에 포함되어 있다.

### *배 경 기 술*

[0003] 서열분석 기술에서 유의미한 진전에도 불구하고, 인간 게놈의 약 5-10%는 계속해서 조립되지 않고, 맵핑되지 않고, 저조하게 특성규명된다. 참조 조립체는 일반적으로 다중-메가염기 이색성 갭으로서 이들 누락 영역에 주석을 단다. 게놈의 상기 누락 부분은 일반적으로 사용된 서열분석 기술을 이용하여 정확한 특성규명에 대해 계속해서 내성인 구조적 특징을 포함한다. 전체 게놈의 드노보 서열분석은 경제적으로 실행가능하지 않고, 따라서 대규모로 게놈 분석의 이점을 유지하는 한편 게놈 서열분석과 관련된 비용을 감소시키기 위한 요구가 남아 있다.

### *발명의 내용*

[0004] **본 발명의 요약**

[0005] 따라서, 본 개시내용은 상기 선택된 영역의 드노보 서열 조립체를 허용하기 위해, 및 일부 양태에서, 고 처리율 및 고 정확도로 상기 게놈의 잔류 영역의 재서열분석과 드노보 커버리지의 조합을 허용하기 위해 게놈의 선택된 영역의 표적화된 커버리지의 제공 방법, 시스템 및 조성물을 제공한다.

[0006] 일부 양태에서, 본 개시내용은 하기 단계를 포함하는, 게놈의 하나 이상의 선택된 부분의 서열분석 방법을 제공

한다: (a) 개시 게놈 재료를 제공하는 단계; (b) 개별 핵산 분자를 상기 개시 게놈 재료로부터 개별 파티션으로 분배하여, 이로써 각각의 개별 파티션이 개별 핵산 분자를 함유하도록 하는 단계; (c) 상기 개별 파티션 내 상기 개별 핵산 분자의 적어도 일부의 선택된 부분을 증폭시켜 증폭산물의 집단을 형성하는 단계; (d) 증폭산물의 상기 집단을 바코딩하여 상기 증폭산물의 복수의 바코딩된 단편을 형성하는 단계로서, 소정의 개별 파티션 내에서의 단편 각각은 공통 바코드를 포함하고, 이로써 이것이 유도된 상기 개별 핵산 분자와 각 단편을 회합시키는, 상기 형성 단계; (e) 상기 복수의 단편 유래의 서열 정보를 수득하여 이로써 게놈의 하나 이상의 선택된 부분을 서열분석하는 단계.

[0007] 추가 구현예에서, 그리고 상기 중 임의의 것에 따라, 상기 게놈의 상기 하나 이상의 선택된 부분은 상기 게놈의 고도로 다형성된 영역을 포함한다. 또 다른 추가 구현예에서, 게놈의 하나 이상의 선택된 부분의 서열분석은 드노보 서열분석이다.

[0008] 또 다른 추가 구현예에서 그리고 상기 중 임의의 것에 따라, 증폭은 적어도 3.5 메가염기쌍(Mb)의 영역에 걸쳐 PCR 증폭을 포함한다. 또 다른 추가 구현예에서, 증폭은 적어도 3.0 Mb의 영역에 걸쳐 엇갈린 다중 프라이머 쌍을 이용하는 PCR 증폭을 포함한다.

[0009] 일부 구현예에서, 그리고 상기에 따라, 상기 서열분석 반응은 짧은 판독, 고 정확도 서열분석 반응이다. 추가 구현예에서, 수득 단계에서 생성된 서열 정보는 이의 본래 개별 핵산의 분자 콘텍스트를 유지한다.

[0010] 특정 구현예에서, 그리고 상기 중 임의의 것에 따라, 수득 단계 전, 복수의 단편은 추가로, 하기에 의하여 게놈의 하나 이상의 선택된 부분 중 적어도 일 부분을 포함하는 단편에 대해 농축된다: (i) 상기 게놈의 상기 하나 이상의 선택된 부분에서 또는 부분 근처에서 영역에 상보적인 프로브를 상기 단편에 하이브리드화시켜 프로브-단편 복합체를 형성하는 단계; (ii) 프로브-단편 복합체를 고형 지지체의 표면 상에 포획하는 단계.

[0011] 일부 구현예에서, 그리고 상기 중 임의의 것에 따라, 개별 파티션 내 증폭산물의 바코딩된 단편은 게놈의 하나 이상의 선택된 부분의 약 100X-5000X 커버리지를 나타낸다. 추가 구현예에서, 개별 파티션 내 증폭산물의 바코딩된 단편은 게놈의 하나 이상의 선택된 부분의 약 200X-1000X 커버리지를 나타낸다. 추가의 추가 구현예에서, 개별 파티션 내 증폭산물의 바코딩된 단편은 게놈의 하나 이상의 선택된 부분의 적어도 1000X 커버리지를 나타낸다. 또 다른 추가 구현예에서, 개별 파티션 내 증폭산물의 바코딩된 단편은 게놈의 하나 이상의 선택된 부분의 적어도 2000X 또는 5000X 커버리지를 나타낸다.

[0012] 추가 양태에서, 본 개시내용은 하기 단계를 포함하는, 게놈 샘플의 하나 이상의 저조하게 특성규명된 부분 유래의 서열 정보의 수득 방법을 제공한다: (a) 개별 파티션에서 상기 게놈 샘플의 개별 제1 핵산 단편 분자를 제공하는 단계; (b) 상기 개별 파티션 내에서 상기 개별 제1 핵산 단편 분자를 단편화시켜 각각의 상기 개별 제1 핵산 단편 분자로부터 복수의 제2 단편을 형성하는 단계; (c) 저조하게 특성규명된 상기 복수의 제2 단편의 선택된 영역을 증폭시켜 증폭산물의 집단을 형성하는 단계; (d) 각 개별 파티션 내 상기 증폭산물에 공통 바코드 서열을 부착시켜, 이로써 상기 증폭산물 각각이 이것이 함유된 상기 개별 파티션에 부여가능하도록 하는 단계; (e) 상기 증폭산물의 서열을 식별하여, 이로써 상기 게놈 샘플의 하나 이상의 저조하게 특성규명된 부분 유래의 서열 정보를 수득하는 단계.

[0013] 특정 구현예에서 그리고 상기 중 임의의 것에 따라, 증폭은 적어도 3.5 메가염기쌍(Mb)의 영역에 걸쳐 PCR 증폭을 포함한다. 추가 구현예에서, 증폭은 적어도 3.0 Mb의 영역에 걸쳐 엇갈린 다중 프라이머 쌍을 이용하는 PCR 증폭을 포함한다. 또 다른 추가 구현예에서, 다중 프라이머 쌍은 프라이머 서열의 증폭을 방지하기 위해 우라실을 함유한다.

[0014] 일부 구현예에서, 그리고 상기 중 임의의 것에 따라, 식별 단계는 증폭산물의 서열의 분자 콘텍스트를 보존하여, 이로써 상기 식별은 동일한 개별 제1 핵산 단편 분자로부터 유도된 증폭산물을 식별하도록 하는 것을 추가로 포함한다. 추가 구현예에서, 방법은 추가로 복수의 제2 단편의 중첩 서열에 기반하여 추론된 콘티그에서 2 이상의 개별 제1 단편 분자의 연결을 포함하고, 여기서 추론된 콘티그는 적어도 10kb의 길이 N50을 포함한다.

[0015] 일부 구현예에서, 그리고 상기 중 임의의 것에 따라, 바코드 서열은 추가의 서열 분절을 추가로 포함한다. 추가 구현예에서, 추가의 서열 분절은 하기로 이루어진 군으로부터 선택된 구성원 중 하나 이상을 포함한다: 프라이머, 부착 서열, 랜덤 n-량체 올리고뉴클레오티드, 우라실 핵염기를 포함한 올리고뉴클레오티드. 또 다른 추가 구현예에서, 바코드는 적어도 700,000 바코드의 라이브러리로부터 선택된다.

[0016] 일부 구현예에서, 그리고 상기 중 임의의 것에 따라, 각 개별 파티션 내에서 게놈 샘플은 단세포 유래의 게놈

DNA를 포함한다. 추가 구현예에서, 각 개별 파티션은 상이한 염색체 유래의 게놈 DNA를 포함한다.

[0017] 일부 구현예에서, 그리고 상기 중 임의의 것에 따라, 개별 파티션은 에멀젼 내 액적을 포함한다.

[0018] 일부 구현예에서, 그리고 상기 중 임의의 것에 따라, 개별 파티션 내에서 바코딩된 증폭산물은 게놈의 하나 이상의 저조하게 특성규명된 부분의 약 1000X-5000X 커버리지를 나타낸다.

[0019] 추가 양태에서, 본 출원은 하기 단계를 포함하는, 분자 콘텍스트를 유지하는 한편 게놈 샘플의 하나 이상의 부분 유래의 서열 정보를 수득하는 방법을 제공한다: (a) 개시 게놈 재료를 제공하는 단계; (b) 개별 핵산 분자를 상기 개시 게놈 재료로부터 개별 파티션으로 분배하여, 이로써 각 개별 파티션이 제1 개별 핵산 분자를 함유하도록 하는 단계; (c) 게놈의 하나 이상의 선택된 부분 중 적어도 일 부분을 포함하는 단편에 대해 농축된 집단을 제공하는 단계; (d) 각 개별 파티션 내 상기 증폭산물에 공통 바코드 서열을 부착시켜, 이로써 상기 단편 각각이 이것이 함유된 상기 개별 파티션에 부여가능하도록 하는 단계; (e) 상기 단편 유래의 서열 정보를 수득하여, 이로써 분자 콘텍스트를 유지하는 한편 상기 게놈 샘플 중 하나 이상의 표적화된 부분을 서열분석하는 단계.

[0020] 또 다른 추가 양태에서, 본 개시내용은 하기 단계를 포함하는, 분자 콘텍스트를 유지하는 한편 게놈 샘플의 하나 이상의 부분 유래의 서열 정보를 수득하는 방법을 제공한다: (a) 개시 게놈 재료를 제공하는 단계; (b) 개별 핵산 분자를 상기 개시 게놈 재료로부터 개별 파티션으로 분배하여, 이로써 각 개별 파티션이 제1 개별 핵산 분자를 함유하도록 하는 단계; (c) 게놈의 하나 이상의 선택된 부분 중 적어도 일 부분을 포함하는 단편의 서열에 대해 농축된 개별 파티션 중 적어도 일부 내의 집단을 제공하는 단계; (d) 각 개별 파티션 내의 단편에 공통 바코드 서열을 부착하여, 상기 단편의 각각이 이것이 함유되어 있는 개별 파티션에 부여가능하도록 하는 단계; (e) 상기 게놈의 하나 이상의 선택된 부분을 포함하는 단편을 함유하지 않는 개별 파티션으로부터, 상기 게놈의 하나 이상의 선택된 부분의 적어도 일 부분을 포함하는 단편을 함유하는 개별 파티션을 분리하는 단계; (f) 상기 게놈의 하나 이상의 선택된 부분 중 적어도 일 부분을 포함하는 상기 단편 유래의 서열 정보를 수득하여, 이로써 분자 콘텍스트를 유지하는 한편 상기 게놈 샘플 중 하나 이상의 표적화된 부분을 서열분석하는 단계.

[0021] 추가 구현예에서 그리고 상기 중 임의의 것에 따라, 게놈의 하나 이상의 선택된 부분의 적어도 일 부분을 포함하는 단편의 서열이 농축된 집단의 제공 단계는 게놈의 하나 이상의 선택된 부분의 적어도 일 부분을 포함한 증폭산물의 집단을 생산하기 위해 게놈의 하나 이상의 선택된 부분의 적어도 일 부분을 포함하는 단편의 지향된 PCR 증폭을 포함한다. 또 다른 추가 구현예에서, 상기 제공 단계는 추가로, 일부 구현예에서 형광 분자를 포함할 수 있는, 증폭산물에 검출가능한 표지 부착을 포함한다. 또 다른 추가의 구현예에서, 게놈의 하나 이상의 선택된 부분의 적어도 일 부분을 포함하는 단편을 함유한 개별 파티션을 상기 게놈의 상기 하나 이상의 선택된 부분을 포함하는 단편을 함유하지 않은 개별 파티션으로부터 분리시키는 단계는, 그와 같은 신호 없이 상기 파티션으로부터 상기 검출가능한 표지에서 신호를 방출하는 상기 파티션 분류를 포함한다.

[0022] 일부 구현예에서 그리고 상기 중 임의의 것에 따라, 단편 유래의 서열 정보의 수득 이전에, 개별 파티션은 조합되고 함께 풀링된다. 추가 구현예에서, 상기 단편 유래의 서열 정보를 수득하는 상기 단계는 상기 단편의 상기 서열의 상기 분자 콘텍스트를 보존하여 이로써 상기 식별이 상기 동일한 제1 개별 핵산 분자로부터 유도된 단편을 식별하도록 하는 방식으로 수행된다. 또 다른 추가 구현예에서, 서열 정보의 상기 수득은 하기로 이루어진 군으로부터 선택된 서열분석 반응을 포함한다: 짧은 판독-길이 서열분석 반응 및 긴 판독-길이 서열분석 반응. 또 다른 추가 구현예에서, 서열분석 반응은 짧은 판독, 고 정확도 서열분석 반응이다.

[0023] 일부 구현예에서, 그리고 상기 중 임의의 것에 따라, 개별 파티션은 에멀젼 내 액적을 포함한다. 추가 구현예에서, 개별 파티션 내에서 바코딩된 단편은 게놈의 하나 이상의 선택된 부분의 약 100X-5000X 커버리지를 나타낸다. 또 다른 추가 구현예에서, 개별 파티션 내에서 바코딩된 단편은 게놈의 하나 이상의 선택된 부분의 약 200X-1000X 커버리지를 나타낸다. 또 다른 추가의 추가 구현예에서, 개별 파티션 내 증폭산물의 바코딩된 단편은 게놈의 하나 이상의 선택된 부분의 적어도 1000X 커버리지를 나타낸다. 또 다른 추가 구현예에서, 개별 파티션 내에서 바코딩된 단편은 게놈의 하나 이상의 선택된 부분의 적어도 2000X 또는 5000X 커버리지를 나타낸다.

[0024] 일부 양태에서, 그리고 상기 중 임의의 것에 따라, 본 개시내용은 하기 단계를 포함하는, 분자 콘텍스트를 유지하는 한편 게놈 샘플 중 하나 이상의 부분 유래의 서열 정보를 수득하는 방법을 제공한다: (a) 게놈 재료를 제공하는 단계; (b) 개별 핵산 분자를 상기 게놈 재료로부터 분리시켜 분리된 개별 핵산 분자를 형성하는 단계; (c) 상기 분리된 개별 핵산 분자 유래의, 게놈의 하나 이상의 선택된 부분 중 적어도 일 부분을 포함하는 단편에 대해 농축된 집단을 제공하는 단계. 특정 구현예에서, 분리 단계는 하나 이상의 핵산 분자가 다른 하나 이상

의 핵산 분자로부터 상대 단리에서 분류 및 가공되도록 허용하는 임의의 방법을 이용하여 달성된다. 일부 구현예에서, 분리는 기재상의 상이한 분획으로 또는 상이한 파티션으로 물리적 분리이다. 추가 구현예에서, 적어도 복수의 단편은 이들이 유도된 개별 핵산 분자에 부여가능하다. 상기 부여성은 특정한 개별 핵산 분자로 유래한 경우 특정한 단편의 표시를 허용하는 임의의 방법을 이용하여 수득된다. 특정 예시적 구현예에서, 상기 부여성은 단편의 바코딩에 의해 수득된다. 추가 양태에서, 서열 정보는 단편으로부터 수득되고, 이로써 분자 콘텍스트를 유지하는 한편 게놈 샘플의 하나 이상의 표적화된 부분을 서열분석한다.

## 도면의 간단한 설명

[0025] 도 1은 종래의 공정 대 실시예 공정 및 본원에 기재된 시스템을 사용하는 표적화된 게놈 영역의 식별 및 분석의 도식적 예시를 제공한다.

도 2는 본원에서 기재된 공정 및 시스템을 이용한 표적화된 게놈 영역의 식별 및 분석의 도식적 예시를 제공한다.

도 3은 본원에 개시된 방법 및 조성물을 사용하여 서열 정보를 검출하기 위한 검정을 수행하는 전형적인 작업 흐름을 예시한다.

도 4는 핵산 샘플과 비드를 조합하고 핵산 및 비드를 개별 액적으로 파티셔닝하기 위한 공정의 도식적 예시를 제공한다.

도 5는 염색체 핵산 단편의 바코딩 및 증폭을 위한 공정의 도식적 예시를 제공한다.

도 6a 및 6b는 이의 유래 공급원 핵산 분자에 서열 데이터 부여에서 핵산 단편의 바코딩의 사용의 도식적 예시를 제공한다.

도 7은 본 발명의 일 구현예의 도식적 예시를 제공한다.

도 8은 본 발명의 일 구현예의 도식적 예시를 제공한다.

도 9는 주형(NTC)를 함유하지 않은 것과 비교된 주형로 수행된 증폭 반응을 비교하는 실험 유래의 데이터를 보여준다.

도 10은 어닐링 온도의 범위를 거쳐 수행된 증폭 반응 유래의 데이터를 보여준다.

## 발명을 실시하기 위한 구체적인 내용

[0026] 본 발명의 실시는, 유기 화학, 중합체 기술, 분자 생물학 (재조합 기술 포함), 세포 생물학, 생화학 및 면역학의 종래의 기술 및 기재내용을, 다르게 명시되지 않으면, 채택할 수 있다. 상기 종래의 기술은 폴리머 어레이 합성, 하이브리드화, 결찰, 파아지 디스플레이, 및 표지를 이용한 하이브리드화의 검출을 포함한다. 적절한 기술의 구체적 예시는 하기 실시예를 참조하였을 수 있다. 그러나, 다른 동등한 종래의 절차는, 물론, 또한 사용될 수 있다. 상기 기존의 기술 및 기재내용은 하기에서 발견될 수 있다: standard laboratory manuals such as *Genome Analysis: A Laboratory Manual Series* (Vols. I-IV), *Using Antibodies: A Laboratory Manual, Cells: A Laboratory Manual, PCR Primer: A Laboratory Manual, and Molecular Cloning: A Laboratory Manual* (all from Cold Spring Harbor Laboratory Press), Stryer, L. (1995) *Biochemistry* (4th Ed.) Freeman, New York, Gait, "*Oligonucleotide Synthesis: A Practical Approach*" 1984, IRL Press, London, Nelson and Cox (2000), *Lehninger, Principles of Biochemistry* 3rd Ed., W. H. Freeman Pub., New York, N.Y. and Berg et al. (2002) *Biochemistry,* 5th Ed., W. H. Freeman Pub., New York, N.Y. (이의 전체 내용은 모든 목적을 위하여 그 전체로 본원에 참고로 포함되어 있음).

[0027] 주목할 만한 것은 본원 명세서 및 첨부된 청구범위에서 사용되는 단수 형태는 문맥상 명확하게 다르게 지시하지 않는 한 복수 대상을 포함한다는 것이다. 따라서, 예를 들어, "폴리머라아제"에 대한 언급은 일 제제 또는 상기 제제의 혼합물을 지칭하고, "방법"에 대한 언급은 당업자에게 공지된 동등 단계 및 방법 등에 대한 언급을 포함한다.

[0028] 달리 정의되지 않는 한, 본원에서 사용된 전체 기술 및 과학적 용어들은 본 발명이 속하는 당해 분야의 숙련가에 의해 통상적으로 이해되는 바와 같이 동일한 의미를 갖는다. 본원에 언급된 모든 공보는, 예를 들면, 현재 기재된 본 발명과 함께 사용될 수도 있고, 공보에 기술된 장치, 조성물, 제형, 및 방법론을 개시 및 기술할 목

적으로 본원에 참고로 포함된다.

[0029] 값의 범위가 제공되는 경우, 각각의 그 사이의 값은, 콘텍스트를 명백히 달리 나타내지 않는 한 하한의 단위의 1/10까지, 상기 범위의 상한과 하한 사이 및 상기 언급된 범위에서 임의의 기타 언급된 또는 그 사이의 값이 본 발명 내에 포함되는 것으로 이해된다. 이들 보다 작은 범위의 상한치와 하한치는 독립적으로 보다 작은 범위에 포함될 수 있고 또한 상기 진술된 범위에서 임의의 특정 배제된 제한 값으로서 본 발명에서 포괄된다. 언급된 범위가 한계의 한쪽 또는 양쪽을 포함하는 경우, 상기 포함된 한계의 한쪽 또는 양쪽을 배제한 범위가 본 발명에서 또한 포함된다.

[0030] 이하의 설명에서, 많은 구체적인 세부사항이 본 발명에 대한 보다 철저한 이해를 제공하기 위해 제시된다. 그러나, 해당 기술분야의 통상의 기술자에게 본 발명이 이러한 구체적인 세부사항들 중 하나 이상 없이도 실시될 수 있다는 것이 분명할 것이다. 다른 사례에서, 당해 분야의 숙련가에 잘 알려진 공지된 특징 및 절차는 본 발명의 불명확화를 피하기 위해 기재되지 않는다.

[0031] 본원에서 사용된 바와 같이, 용어 "포함하는"은 상기 조성물 및 방법이 기타를 배제하지 않으면서 인용된 요소들을 포함하는 것을 의미하는 것이다. "본질적으로 ~으로 이루어진"이 조성물 및 방법을 정의하기 위해 사용될 경우, 상기 조성물 또는 방법에 임의의 필수적인 유의성을 갖는 다른 요소들을 배제함을 뜻한다. "으로 이루어진"은 청구된 조성물 및 실질적인 방법 단계에 대하여 다른 성분의 미량 원소 초과의 배제를 의미할 것이다. 이들 이행 용어들 각각에 의해 정의된 구현예는 본 발명의 범위 내에 있다. 따라서, 방법 및 조성물이 (포함하는) 추가의 단계 및 성분 또는 대안적으로 (으로 본질적으로 이루어진) 유의성 없는 포함 단계 및 조성물 또는 대안적으로, (으로 이루어진) 언급된 방법 단계 또는 조성물만의 의도를 포함할 수 있도록 의도된다.

[0032] 모든 숫자 표시, 예를 들면, 범위를 포함한, pH, 온도, 시간, 농도, 및 분자량은 0.1의 증분만큼 ( + ) 또는 ( - ) 다양한 근사치이다. 항상 명백하게 언급되지 않아도 모든 숫자 표시가 용어 "약" 뒤에 온다는 것이 이해되어야 한다. 용어 "약"은 또한 "X"의 소수의 증분에 더하여 정확한 값 "X" 예컨대 "X + 0.1" 또는 "X - 0.1"을 포함한다. 항상 명백하게 언급되지 않아도, 본원에서 기재된 시약이 단지 예시적이라는 것 및 상기의 등가물이 당해 기술에 공지되어 있다는 것이 또한 이해되어야 한다.

[0033] *I. 개관*

[0034] 본 개시내용은 유전 물질의 특성규명에 유용한 방법, 조성물 및 시스템을 제공한다. 특히, 본원에서 기재된 방법, 조성물 및 시스템은 추가의 불필요한 서열 정보가 게놈의 상기 선택된 부분으로부터 수득될 수 있는 정도로 게놈의 선택된 부분의 증가된 및 불필요한 커버리지를 제공한다. 특이적 사례에서, 상기 추가의 서열 정보는 게놈의 상기 선택된 부분의 드노보 서열분석을 허용하기 위해 충분한 정보를 제공한다.

[0035] 일반적으로, 본원에서 기재된 방법, 조성물, 및 시스템은 게놈의 선택된 영역의 유전적 특성규명을 제공한다. 상기 유전적 특성규명은 게놈의 선택된 영역의 드노보 서열분석을 허용하기 위해 충분히 심도있다. 상기 드노보 서열분석은 저조하게 특성규명된, 고도로 다형성된, 및/또는 참조 서열에서 갈라진 게놈의 영역에 대하여 특히 유용하다. 인정될 바와 같이, 인간 게놈의 유의미한 백분율(예를 들어 Altemose et al., *PLOS* Computational Biology, May 15, 2014, Vol. 10, Issue 5에 따라, 적어도 5-10%)는 계속해서 조립되지 않고, 맵핑되지 않고, 저조하게 특성규명된다. 참조 조립체는 일반적으로, 주로 동원체에 근접하여, 그리고 단부동원체성 염색체의 짧은 아암 상에 발견되는, 다중-메가염기 이색성 갭으로서 이들 누락 영역에 주석에 주석을 단다. 게놈의 상기 누락 부분은 일반적으로 사용된 서열분석 기술을 이용하여 정확한 특성규명에 대해 계속해서 내성인 구조적 특징을 포함한다. 정확한 특성규명에 내성인 예시적인 영역은 밀접한 상동성 위유전자(예를 들어 SMN1/2 CYP2D6)를 갖는 영역, 비제한적으로 트랜스포존(예컨대 SINEs, LINEs)을 포함한, 게놈 전반에 걸쳐 실질적인 반복된 서열을 갖는 영역, 및 특히 참조 서열이 불량한 안내로서 작용하는 엄청난 변동을 갖는 영역(예컨대 인간 백혈구 항원(HLA) 착물에 대하여 유전자를 암호화하는 영역)을 포함한다. 본원에서 기재된 방법, 조성물, 및 시스템은 관심 영역의 선택적 증폭을 분자 콘텍스트를 유지하기 위한 능력과 조합하여, 이로써 일반적으로 저조하게 특성규명된 게놈 영역의 드노보 서열분석을 허용할 뿐만 아니라 임의로, 보다 긴 게놈 내에 보다 긴 범위의 분자 콘텍스트를 제공한다.

[0036] 특이적 사례에서, 본원에서 기재된 방법은 게놈의 선택된 영역이 서열분석 이전에 임의로 증폭되는 단계를 포함한다. (비제한적으로 PCR 증폭을 포함한) 당해 기술에서 공지된 방법을 이용하여 일반적으로 수행되는, 상기 증폭은 게놈의 선택된 영역의 적어도 1X, 10X, 20X, 50X, 100X, 200X, 500X, 1000X, 1500X, 2000X, 5000X, 또는 10000X 커버리지를 제공하고, 이로써 상기 선택된 영역의 드노보 서열분석을 허용하기 위해 핵산의 양을 제공한

다. 추가 구현예에서, 증폭은 게놈의 선택된 영역의 적어도 1X-20X, 50X-100X, 200X-1000X, 1500X-5000X, 5000X-10,000X, 1000X-10000X, 1500X-9000X, 2000X-8000X, 2500X-7000X, 3000X-6500X, 3500X-6000X, 4000X-5500X 커버리지를 제공한다.

[0037] 증폭은 게놈의 선택된 영역 내에서 또는 근처에서 서열에 상보적인 프라이머의 확장을 통해 일반적으로 수행된다. 일부 경우에서, 관심 영역에 걸쳐 타일링하도록 설계되는 프라이머의 라이브러리가 사용되고 - 환언하면, 프라이머의 라이브러리는 게놈의 선택된 영역을 따라 특정한 거리에서 영역을 증폭시키도록 설계된다. 일부 사례에서, 선택적 증폭은 게놈의 선택된 영역을 따라 매 10, 15, 20, 25, 50, 100, 200, 250, 500, 750, 1000, 또는 10000 염기에 상보적인 프라이머를 이용한다. 또 다른 추가 예에서, 프라이머의 타일링된 라이브러리는 거리의 혼합을 포획하도록 설계되거나 - 상기 혼합은 거리의 랜덤 혼합일 수 있거나 선택된 영역의 특정한 부분 또는 백분율이 상이한 프라이머 쌍에 의해 증폭되도록 지능적으로 설계된다.

[0038] 일반적으로, 본원에서 기재된 방법 및 시스템은 게놈의 선택된 영역의 서열의 결정을 제공함으로써 표적화된 게놈 서열분석을 완수하고, 상기 서열분석 정보는 극도로 낮은 서열분석 오차율 및 짧은 판독 서열분석 기술의 고처리율의 이점을 갖는 방법을 이용하여 수득된다.

[0039] 핵산의 서열분석은 서열 판독의 분자 콘텍스트 또는 서열 판독의 부분을 보존하는 방식으로 전형적으로 수행된다. 상기로 다중 서열 판독 또는 서열 판독의 다중 부분이 핵산의 단일 유래 분자에 부여될 수 있다는 의미이다. '에 부여하는'으로 서열 판독이 핵산의 이의 특정한 유래 분자의 염기의 선형 서열 내에 해당하는 경우 식별될 수 있다는 의미이고 - 환언하면, 단편 1 및 2가 유래 핵산 분자 A로부터 생성되면, 서열분석은 단편 1, 2, 3 및 4 유래의 서열 판독이 이의 분자 콘텍스트를 유지하는 방식으로 수행되고, 그리고 단편 1 및 2가 유래 분자 A로부터 유도되는 것이 용이하게 확인된다.

[0040] 핵산의 상기 단일 분자가 임의의 다양한 길이일 수 있는 반면, 바람직한 양태에서, 긴 범위 분자 콘텍스트의 보존을 허용하는 상대적으로 긴 분자일 것이다. 특히, 단일 유래 분자는 바람직하게는 전형적인 짧은 판독 서열 길이보다 실질적으로 더 길고, 예를 들면, 200 염기보다 더 길고, 종종 적어도 1000 염기 이상, 5000 염기 이상, 10,000 염기 이상, 20,000 염기 이상, 30,000 염기 이상, 40,000 염기 이상, 50,000 염기 이상, 60,000 염기 이상, 70,000 염기 이상, 80,000 염기 이상, 90,000 염기 이상, 또는 100,000 염기 이상이고, 일부 경우에서 최대 1 메가염기 이상이다.

[0041] 일반적으로, 도 1에 나타낸 바와 같이, 본원에서 기재된 방법 및 시스템은, 분자 콘텍스트를 유지하는 한편, 핵산, 특히 게놈의 선택된 영역으로부터 핵산을 특성규명하기 위해 사용될 수 있다. 나타낸 바와 같이, 2개의 개별 개별 핵산 102 및 104는 예시되고, 각각은 수많은 관심 영역, 예를 들면, 핵산 102에서 영역 106 및 108, 및 핵산 104에서 영역 110 및 112를 갖는다. 각 핵산내 관심 영역은 동일한 핵산 분자 내에서 연결되지만 (예를 들면, 이로부터 유래하지만), 일부 경우에서 이들 영역은 서로, 예를 들면, 1 kb 초과 떨어져서, 5 kb 초과 떨어져서, 10 kb 초과 떨어져서, 20 kb 초과 떨어져서, 30 kb 초과 떨어져서, 40 kb 초과 떨어져서, 50 kb 초과 떨어져서, 및 일부 경우에서, 100 kb 초과만큼 많이 떨어져서 상대적으로 분리될 수 있다. 관심 영역은 일반적으로 게놈의 개별 및 분리된 부분이고 - 일부 경우에서, 상기 영역은 저조하게 특성규명된 영역이다. 관심 영역은 또한 개별 유전자, 유전자 그룹, 엑손을 표시할 수 있다. 나타낸 바와 같이, 각 핵산 102 및 104는 분리된다. 도 1에 예시된 바와 같이, 각 핵산은 이의 자체 파티션 114 및 116, 각각으로 분리되지만; 인정될 바와 같이, 본원에서 기재된 방법은 상기 파티션의 용도에 제한되지 않고 핵산 분자의 임의의 분리 방법은 사용될 수 있고 그 다음 상기 분리된 핵산 분자는 본원에서 개시된 임의의 방법에 따라 추가로 가공될 수 있다. 본원에서 다른 곳에 언급된 바와 같이, 도 1 내 파티션 예컨대 114 및 116은, 많은 경우에서, 유중수 에멀젼에서 수성 액적이다. 각 액적 내에서, 각 단편의 부분은, *예를 들면*, 동일한 분자로부터 유래된 경우, 상기 단편의 본래 분자 콘텍스트를 보존하는 방식으로 복제된다. 상기 분자 콘텍스트는 이것이 유도된 본래 핵산 분자에 단편의 부여성을 허용하는 임의의 방법을 이용하여 보존될 수 있다. 도 1에 나타낸 바와 같이, 이것이 달성되는 한 방법은 유래 단편이 파티셔닝된 액적을 대표하는, 바코드 서열, *예를 들면*, 예시된 대로 바코드 서열 "1" 또는 "2"의 각 복제된 단편에서 포함을 통한다. 전체의 게놈 서열분석 적용에 대하여, 각각의 유래 핵산 102 및 104로부터 전체 범위 서열 정보를 서열분석 및 재조립하기 위해, 모든 복제된 단편 및 이의 관련된 바코드를 간단히 풀링할 수 있다. 그러나, 많은 경우에서, 게놈의 과학적으로 관련된 부분에 더 큰 집중을 제공하기 위해, 그리고 게놈의 덜 관련된 또는 무관한 부분에 관한 서열분석 수행의 시간 및 비용 최소화하기 위해, 전체 게놈의 과학적 표적화된 부분을 단지 분석하는 것이 더욱 바람직하다. 분자 콘텍스트 보존에 일조하는 다른 서열분석 방법은 단일 분자 서열분석 공정, 예컨대 Pacific Biosciences로부터 이용가능한 SMRT 서열분석, 및, 예를 들면, Oxford Nanopore에 의해 기재된 나노포어 서열분석, 그리고 Illumina, Inc.로부터 이용가능한 Truseq SLR 공정을 포함

한다.

[0042] 상기에 따라서, 바코딩 단계에 더하여, 선택적 증폭의 하나 이상의 단계가 있을 수 있고, 이로써 핵산 102 또는 104가 선택된 관심 게놈 영역을 함유하면, 상기 영역으로부터 증폭산물은 각각의 파티션 114 및 116에서 단편의 더 큰 백분율을 형성할 것이다. 비록 일부 구현예에서 증폭 단계가 또한 바코드의 부착에 이어서 발생할 수 있어도, 상기 증폭 단계는 본원에서 기재된 방법에 따라서 바코드의 부착 이전에 또는 이와 동시에 일반적으로 발생할 것이다.

[0043] 라이브러리 118 내에 풀링된 단편이, 예를 들면, 바코드 정보의 체류를 통해, 이의 본래 분자 콘텍스트를 유지하기 때문에, 이들은 포매된 (때때로, 긴 범위) 연결 정보, 예를 들면, 각각의 조립된 관심 영역 106:108과 110:112 사이로서 추론된 연결로 이의 본래 분자 콘텍스트으로 재조립될 수 있다. 예로써, 게놈의 2개의 이질적인 표적화된 부분, 예를 들면, 2 이상의 엑손 사이 직접적인 분자 연결을 식별할 수 있고, 그 직접적인 분자 연결은 구조적 변동 및 다른 게놈 특징을 식별하기 위해 사용될 수 있다. 선택적 증폭이 게놈의 선택된 영역의 부분을 함유한 핵산 단편의 양을 증가시키기 위해 이용되는 상황에 대하여, 그 다음 분자 콘텍스트를 식별하기 위한 능력은 또한, 종종 상기 영역의 드노보 조합을 허용하는 커버리지의 깊이에서, 게놈의 상기 선택된 영역을 서열분석하기 위한 방식을 제공한다.

[0044] 특정 상황에서, 본원에서 기재된 서열분석 방법은 게놈의 더 긴 범위를 거쳐 더 낮은 수준 연결된 판독과 선택된 영역의 깊은 커버리지의 조합을 포함한다. 인정될 바와 같이, 드노보와 재서열분석의 상기 조합이 전체 게놈 및/또는 게놈의 큰 부분을 서열분석하기 위한 효율적인 방식을 제공한다. 본원에서 기재된 선택적 증폭 방법을 통해 저조하게 특성규명된 및/또는 고도로 다형성된 영역의 표적화된 커버리지는 일정 커버리지 수준에서 이러한 영역의 드노보 서열 조립체에 필요한 핵산 물질의 양을 제공하고, 게놈의 다른 영역에 대해 연결된 게놈 서열분석은, 이들의 분자 콘텍스트의 보존을 통하여 함께 연결된 개별 영역에 대한 서열 정보를 제공함으로써 게놈의 나머지의 고 처리율 서열분석을 허용한다. 동일한 서열분석 플랫폼 및 서열분석 라이브러리가 커버리지의 양쪽 유형에 대하여 사용될 수 있기 때문에, 본원에서 기재된 방법 및 조성물은 드노보와 연결된 판독 서열분석의 조합 허용을 본질적으로 잘 받아들인다. 본원에서 기재된 방법에 따라서 서열분석되는 핵산 및/또는 핵산 단편의 집단은 드노보 서열분석용 게놈 영역 및 재서열분석용 게놈 영역 둘 모두로부터 서열을 함유하고 - 드노보 서열분석용 관심 영역을 커버하는 핵산의 비율은 본원에서 추가로 상세히 기재된 표적화된 증폭 방법 때문에 게놈의 다른 영역을 커버하는 핵산보다 더 높다. 본원에서 기재된 방법이 조립 동안에 단계 정보를 유지시키기 때문에, 상기 방법은 일배체형의 드노보 조합에 대하여 추가로 잘 받아들인다.

[0045] 게놈의 선택된 영역 유래의 서열 정보를 수득하기 위한 능력 제공에 더하여, 본원에서 기재된 방법 및 시스템은, 모든 목적 및 특히 게놈 재료의 특성규명에 관한 모든 쓰여진 설명, 도면 및 실시예로 그 전체가 참고로 본원에서 편입되는, 미국 특허 출원 번호 14/752,589 및 14/752,602에 기재된 바와 같이, 비제한적으로 일배체형 단계화, 구조적 변동의 식별, 및 복제수 변동의 식별을 포함한, 게놈 재료의 다른 특성규명을 또한 제공할 수 있다.

[0046] 본 출원에 기술된 방법 및 시스템에 따라 핵산을 가공하고 서열분석하는 방법은 또한 추가로 상세히 하기에 기술된다: 모든 목적 및 특히 가공 핵산 그리고 게놈 재료의 서열분석 및 다른 특성규명에 관한 모든 쓰여진 설명, 도면 및 실시예로 그 전체가 참고로 본원에서 편입되는 미국 특허 출원 번호 14/316,383; 14/316,398; 14/316,416; 14/316,431; 14/316,447; 및 14/316,463.

[0047] 일반적으로, 본 발명의 방법은, 본원에서 추가로 상세히 논의된 본 발명의 방법의 도식적 개요를 제공하는, 도 2에서 예시된 대로 단계를 포함한다. 인정될 바와 같이, 도 2에서 개괄된 방법은 필요에 따라 및 본원에서 기재된 바와 같이 변경 또는 변형될 수 있는 예시적인 구현예이다.

[0048] 도 2에 나타낸 바와 같이, 본원에서 기재된 방법은 대부분의 예에서 표적화된 관심 영역을 함유한 샘플 핵산이, 예를 들어 파티션(201)으로 분리되는 단계를 포함할 것이다. 일반적으로, 관심 게놈 영역으로부터 핵산을 함유한 각 파티션은 큰 비율이 선택된 게놈 영역(202)으로부터 서열을 함유할 단편의 집단을 생산하기 위해 표적화된 농축을 경험할 것이다. 단편의 본래 분자 콘텍스트의 임의의 다른 부여 방법이 사용될 수 있어도, 이들이 함유되는 파티션에 특이적인 단편의 보통 바코딩에 의해, 상기 단편은 그 다음 단편(203)의 본래 분자 콘텍스트를 보존하기 위한 방식으로 추가로 단편화 또는 복제된다. 각 파티션은 일부 예에서 1 초과 핵산을 포함할 수 있고, 일부 사례에서 몇 백 핵산 분자를 함유할 것이고 - 다중 핵산이 파티션 내에 있는 상황에서, 게놈의 임의의 특정한 유전자좌는 일반적으로 바코딩 이전에 단일 개별 핵산에 의해 나타낼 것이다. 단계 203의 바코딩된 단편은 당해 기술에 공지된 임의의 방법을 이용하여 생성될 수 있고 - 일부 예에서, 올리고뉴클레오티드는 상이

한 파티션 내 샘플이다. 상기 올리고뉴클레오티드는 샘플의 다수의 상이한 영역을 무작위로 프라이밍하도록 의도된 랜덤 서열일 수 있거나, 이들은 샘플의 표적화 영역의 업스트림을 프라이밍하도록 표적화된 특이적 프라이머 서열을 포함할 수 있다. 추가 예에서, 이들 올리고뉴클레오티드는 또한 바코드 서열을 함유하고, 이로써 복제 공정은 또한 본래 샘플 핵산의 수득한 복제 단편을 바코딩한다. 상기 바코드는, 개별 핵산 분자의 분절을 증폭시키는 증폭 방법 동안 바코드 서열의 부가 뿐만 아니라, 방법 예컨대 하기에 기재된 것을 포함한, 트랜스포존을 이용하는 본래 개별 핵산 분자에 바코드의 삽입을 포함한, 당해 기술에 공지된 임의의 방법을 이용하여 부가될 수 있다: Amini et al., Nature Genetics 46: 1343-1349 (2014) (advance online publication on October 29, 2014). 샘플의 증폭 및 바코딩에서 이들 바코드 올리고뉴클레오티드의 특히 품격있는 사용 공정은 모든 목적 및 특히 가공 핵산 그리고 게놈 재료의 서열분석 및 다른 특성규명에 관한 모든 쓰여진 설명, 도면 및 실시예로 그 전체가 참고로 본원에서 편입되는 미국 특허 출원 번호 USSNs 14/316,383; 14/316,398; 14/316,416; 14/316,431; 14/316,447; 및 14/316,463에서 상세히 기재된다. 또한 연장 반응 시약, 예를 들면, DNA 폴리머라아제, 뉴클레오시드 트리포스페이트, 파티션 내 또한 함유된 동시-인자(예를 들면, $Mg^{2+}$ 또는 $Mn^{2+}$ 등)는 주형으로서 샘플을 사용하여 프라이머 서열을 연장시킴으로써 프라이머가 어닐링되는 주형의 가닥에 대한 상보적 단편을 생산하며, 당해 상보적 서열은 올리고뉴클레오티드 및 이의 연관된 바코드 서열을 포함한다. 샘플의 상이한 부분으로의 다중 프라이머의 어닐링 및 확대는 샘플의 중첩성 상보적 단편의 거대한 풀을 초래할 수 있고, 각각은 그것이 생성된 파티션을 나타내는 자체 바코드 서열을 포함한다. 일부 경우에, 이러한 상보적 단편은 자체가 다시 바코드 서열을 포함하는 보체의 보체를 생산하기 위해 파티션에 존재하는 올리고뉴클레오티드에 의해 프라이밍되는 주형으로서 사용될 수 있다. 추가 예시에서, 이 복제 공정은 제1 보체가 복제되는 경우, 그것이 이의 말단 또는 부근에 두 개의 상보적 서열을 생성하여 반복 사본을 추가로 생성하기 위한 기초가 되는 분자의 능력을 감소시키는 헤어핀 구조 또는 부분적 헤어핀 구조의 형성을 가능하게 하도록 구성된다.

[0049] 도 2에서 예시된 방법으로 돌아가서, 파티션-특이적 바코드가 복제된 단편에 부착되면, 바코딩된 단편은 그 다음 풀링된다(204). 풀링된 단편은 그 다음 서열분석되고(205) 단편의 서열은 이의 유래 분자 콘텍스트(206)에 부여되어, 이로써 표적화된 관심 영역은 모두 그 유래 분자 콘텍스트으로 식별되고 또한 연결된다. 본원에서 기재된 방법 및 시스템의 이점은 표적화된 게놈 영역에 대하여 단편의 농축 이전에 복제된 단편의 파티션- 또는 샘플-특이적 바코드의 부착이 그들 표적화된 영역의 본래 분자 콘텍스트를 보존하여, 그들의 본래 파티션 및 따라서 그들의 본래 샘플 핵산 분자에 부여되도록 허용한다는 점이다.

[0050] 상기 작업흐름에 더하여, 표적화된 게놈 영역은, 칩-기반 및 용액-기반 포획 방법 둘 모두를 포함하는 방법을 이용하여, 추가 분석, 특히 서열분석을 위하여 추가로 농축, 단리 또는 분리, 즉, "붕괴(pulled down)"될 수 있다. 상기 방법은 관심 게놈 영역 또는 관심 게놈 영역 근처 또는 인접한 영역에 상보적인 프로브를 이용한다. 예를 들어, 혼성(또는 칩-기반)포획에서, 함께 합쳐져서 관심 영역을 커버하는 서열을 가진 포획 프로브(보통 단일가닥 올리고뉴클레오티드)를 함유한 마이크로어레이는 표면에 고정된다. 게놈 DNA는 단편화되고 추가로 가공 예컨대 무딘 말단을 생산하기 위한 말단-치유 및/또는 추가의 특징 예컨대 보편적 프라이밍 서열의 부가를 경험할 수 있다. 이들 단편은 마이크로어레이 상에서 프로브에 하이브리드화된다. 미하이브리드화된 단편은 세정제거되고 요망된 단편은 서열분석 또는 다른 분석을 위하여 표면상에서 용출되거나 달리 가공되고, 따라서 표면상에 남아있는 단편의 집단은 표적화된 관심 영역(예를 들면, 포획 프로브에 함유된 것들에 상보적인 서열을 포함한 영역)을 함유한 단편에 대해 농축된다. 단편의 농축된 집단은 추가로 당해 기술에 공지된 임의의 증폭 기술을 이용하여 증폭될 수 있다. 상기 표적화된 붕괴 농축 방법에 대하여 예시적인 방법은, 모든 쓰여진 설명, 도면 및 실시예를 포함하여, 모든 목적 및 특히 표적화된 붕괴 농축 방법 및 서열분석 방법에 관련된 모든 교시로 그 전체가 참고로 본원에서 편입되는, 2015년 10월 29일 출원된, 미국 특허 출원 번호 14/927,297에 기재된다.

[0051] 상기 언급된 바와 같이, 본원에 기재된 방법 및 시스템은 보다 긴 핵산의 짧은 서열 판독을 위한 개별적 분자 콘텍스트를 제공한다. 상기 개별 분자 콘텍스트는 본래 개별 핵산에 더 짧은 서열 판독의 부여성을 허용하는 임의의 방법 또는 조성물에 의해 제공될 수 있다. 본원에 사용된 바와 같이, 개별적 분자 콘텍스트는, 예를 들면, 서열 판독 자체 내에 포함되지 않는 인접 또는 근위 서열에 관련되는 특이적 서열 판독을 초과하는 서열 콘텍스트를 의미하고, 그 자체로, 전형적으로, 그들이 짧은 서열 판독, 예를 들면, 쌍형성된 판독에 대해 약 150 염기 또는 약 300 염기의 판독에 전체로 또는 부분적으로 포함되지 않을 정도일 것이다. 특히 바람직한 양태에서, 방법 및 시스템은 짧은 서열 판독을 위한 긴 범위 서열 콘텍스트를 제공한다. 이러한 긴 범위 콘텍스트는 소정의 서열 판독과 서로 1 kb 초과, 5 kb 초과, 10 kb 초과, 15 kb 초과, 20 kb 초과, 30 kb 초과, 40 kb 초과, 50 kb 초과, 60 kb 초과, 70 kb 초과, 80 kb 초과, 90 kb 초과 또는 심지어 100 kb 초과 또는 그 이상의 거리 내

에 있는 서열 판독의 관계 또는 연결을 포함한다. 인정될 바와 같이, 긴 범위 개별 분자 콘텍스트를 제공함으로써, 또한 그 개별 분자 콘텍스트 내에서 변이체의 단계화 정보를 유도할 수 있고, 예를 들면, 특정한 긴 분자상의 변이체는, 통상적으로 단계화된 정의에 의해 될 것이다.

[0052] 더 긴 범위 개별 분자 콘텍스트를 제공함으로써, 본 발명의 방법 및 시스템은 또한 훨씬 더 긴 추론된 분자 콘텍스트(또한 본원에서 일명 "긴 가상 단일 분자 판독")를 제공한다. 본원에서 기재된 바와 같이, 서열 환경은 전체 게놈 서열의 상이한 (일반적으로 킬로베이스 규모상의) 범위를 거쳐 단편의 연결 제공 또는 맵핑을 포함할 수 있다. 이러한 방법은, 짧은 서열 판독의 개별적인 더 긴 분자 또는 연결된 분자의 콘티그으로의 매핑 뿐만 아니라, 예를 들면, 개별 분자의 인접 결정 서열을 갖는 보다 긴 개별 분자의 많은 부분의 긴 범위 서열분석을 포함할 수 있고, 여기서, 상기 결정 서열은 1 kb 초과, 5 kb 초과, 10 kb 초과, 15 kb 초과, 20 kb 초과, 30 kb 초과, 40 kb 초과, 50 kb 초과, 60 kb 초과, 70 kb 초과, 80 kb 초과, 90 kb 초과, 또는 심지어 100 kb 초과이다. 서열 콘텍스트의 경우와 같이, 짧은 서열의 긴 핵산, 예를 들면, 개별적인 긴 핵산 분자 또는 연결된 핵산 분자 또는 콘티그의 수집 둘 모두로의 부여는 고 수준의 서열 콘텍스트를 제공하기 위해 긴 핵산 연신에 대한 짧은 서열의 매핑 뿐만 아니라 이러한 긴 핵산을 통해 짧은 서열로부터 조립된 서열을 제공함을 모두 포함할 수 있다.

[0053] 또한, 긴 개별 분자와 관련된 긴 범위 서열 콘텍스트를 사용할 수 있는 반면, 이러한 긴 범위 서열 콘텍스트를 갖는 것은 또한 훨씬 더 긴 범위 서열 콘텍스트를 추정할 수 있도록 한다. 일례로서, 상기한 긴 범위 분자 콘텍스트를 제공함으로써, 상이한 본래 분자로부터의 긴 서열 중에서중첩 변이체 부분, 예를 들면, 위상 변이체, 전좌된 서열 등을 식별하여 이들 분자 사이의 추론된 연결을 가능하게 할 수 있다. 상기 추론된 연결기 또는 분자 콘텍스트는 본 명세서에서 일명 "추론된 콘티그"이다. 위상 서열의 콘텍스트에서 논의된 일부 경우에, 추론된 콘티그는 일반적으로, 예를 들면, 중첩 위상 변이체에 의해, 개별 유래 분자보다 실질적으로 더 긴 길이의 위상 콘티그를 추정할 수 있는 위상 서열을 나타낼 수 있다. 이러한 위상 콘티그는 본원에서 “단계 블록”으로서 칭명된다.

[0054] 보다 긴 단일 분자 판독(예를 들어, 상기 검토된 “보다 긴 실제적 단일 분자 판독”)로 개시함으로써, 다르게 짧은 판독 서열분석 기술 또는 위상 서열분석에의 다른 접근법을 사용하여 달성가능한 것보다 긴 추론된 콘티그 또는 단계 블록을 유도할 수 있다. 예를 들어, 예컨대 문헌[미국 특허 출원 제2013-0157870호]를 참조한다. 특히, 본원에 기술된 방법 및 시스템을 사용하여, 적어도 약 10kb, 적어도 약 20kb, 적어도 약 50kb의 N50(기술된 N50 수보다 큰 블록 길이의 총합은 전체 블록 길이의 총합의 50%임)을 갖는 추론된 콘티그를 수득할 수 있다. 바람직한 양태에서, 적어도 약 100kb, 적어도 약 150kb, 적어도 약 200kb, 및 많은 경우에, 적어도 약 250kb, 적어도 약 300 kb, 적어도 약 350 kb, 적어도 약 400 kb, 및 일부 경우에, 적어도 약 500 kb 이상의 N50을 갖는 추론된 콘티그 또는 상 블록 길이가 획득된다. 추가의 기타 경우에서, 200 kb 초과, 300 kb 초과, 400 kb 초과, 500 kb 초과, 1 Mb 초과, 또는 심지어 2 Mb 초과의 최대 상 블록 길이가 수득될 수 있다.

[0055] 한 양태에서, 및 본원에서 상기 및 하기 기재된 임의의 포획 방법과 함께, 본원에서 기재된 방법 및 시스템은 본원에서 기재된 임의의 방법에 따라 추가 가공을 위하여 샘플 핵산의 분리를 제공한다. 상기 분리는 이들이 분리되는 다른 핵산으로부터 상대 단리에서 추가 가공 및 반응을 핵산이 경험하게 하는 임의의 형태일 수 있다. 분리는 모든 다른 핵산으로부터, 또는 그 다음 핵산의 다른 그룹으로부터 분리되는, 2 이상의 핵산의 그룹으로 각각 분리된 단일 핵산에 관한 것일 수 있다. 일부 예시적 구현예에서, 상기 분리는 샘플 핵산 또는 이의 단편의 개별 분획 또는 파티션(본원에서 파티션으로 상호교환적으로 칭명됨)으로의 분획화, 증착 또는 파티셔닝을 포함하고, 여기서, 각 파티션은 다른 파티션의 내용물로부터 이의 내용물의 분리를 유지시킨다. 부여성의 고유 식별자 또는 다른 수단(일부 예에서, 바코드)은, 그 정보가 유도되는 샘플 핵산에, 특징, 예를 들면, 핵산 서열 정보의 추후 부여성을 허용하기 위해 분리된 핵산에 이전에, 이후에 또는 동반하여 전달될 수 있다. 핵산이 분획 또는 파티션으로 분리되는 특정 예시적 구현예에서, 식별자는 특정한 분획 내에 포함될 수 있거나 이에, 및 특히 파티션 속에 본래 침착될 수 있는 인접 샘플 핵산의 상대적으로 긴 연신부에 도입될 수 있다.

[0056] 본원에 기술된 방법 내 이용된 샘플 핵산은 전형적으로, 예를 들면, 전체 염색체, 엑솜, 또는 다른 대형 게놈 부분과 같은, 분석될 전체 샘플 핵산의 다수의 중첩 부분을 나타낼 수 있다. 이러한 샘플 핵산은 전체 게놈, 개별 염색체, 엑솜, 증폭산물, 또는 목적하는 임의의 다양한 상이한 핵산을 포함할 수 있다. 샘플 핵산은 전형적으로 핵산이 파티션에서 인접 핵산 분자의 비교적 긴 단편 또는 연신부로 존재하도록 파티셔닝된다. 전형적으로, 샘플 핵산의 이러한 단편은 1 kb 초과, 5 kb 초과, 10 kb 초과, 15 kb 초과, 20 kb 초과, 30 kb 초과, 40 kb 초과, 50 kb 초과, 60 kb 초과, 70 kb 초과, 80 kb 초과, 90 kb 초과 또는 심지어 100 kb 초과일

수 있고, 이는 상기 기술된 보다 긴 범위의 분자 콘텍스트를 허용한다.

[0057] 샘플 핵산은 또한 전형적으로 소정의 파티션이 게놈 좌위의 두 개의 중첩 단편을 포함할 매우 낮은 가능성을 갖는 수준으로 파티셔닝된다. 이는 전형적으로 샘플 핵산을 파티셔닝 공정 동안 낮은 유입량 및/또는 농도로 제공함으로써 달성된다. 그 결과, 바람직한 경우에, 소정의 파티션은 개시 샘플 핵산의 다수의 길지만 비중첩성 단편을 포함할 수 있다. 이어서, 다른 파티션 중의 샘플 핵산은 고유 식별자와 회합되고, 여기서, 임의의 소정의 파티션에 대하여, 내부에 함유된 핵산은 동일한 고유 식별자를 포함하지만, 상이한 파티션은 상이한 고유 식별자를 포함할 수 있다. 더욱이, 파티셔닝 단계는 샘플 성분을 극소 용적의 파티션 또는 액적으로 할당하므로, 위에서 설정된 바와 같은 바람직한 할당을 달성하기 위하여, 보다 큰 용적의 공정, 예를 들면, 튜브, 또는 다중웰 플레이트의 웰 속에 요구될 수 있는 바와 같은, 샘플의 실질적인 희석을 수행할 필요가 없다. 또한, 본원에 기재된 시스템이 이러한 높은 수준의 바코드 다양성을 사용하기 때문에, 상기 제공된 바와 같이, 높은 수의 게놈 등가물 중에서 다양한 바코드를 할당할 수 있다. 특히, 상기한 다중웰 플레이트(참조: 예를 들면, 미국 출원 공보 번호 2013-0079231 및 2013-0157870, 이의 전체 개시 내용은 그 전체가 본원에 참조로 인용됨)는 전형적으로 단지 백개 내지 수 백개의 상이한 바코드 서열로 작동하고, 바코드를 상이한 세포/핵산으로 부여시킬 수 있도록 하기 위해 이들 샘플의 제한적인 희석 공정을 채택한다. 이와 같이, 그들은 일반적으로 100개보다 훨씬 적은 세포로 작동하고, 이는 1:10 정도의 게놈:(바코드 형태)의 비, 확실히 충분히 1:100 이상의 비를 전형적으로 제공할 것이다. 본원에 기재된 시스템은, 한편으로, 높은 수준의 바코드 다양성, 예를 들면, 10,000, 100,000, 500,000개 등을 초과하는 다양한 바코드 형태가 1:50 이하, 1:100 이하, 1:1000 이하 정도의 게놈:(바코드 형태) 비 또는 훨씬 더 작은 비에서 작동할 수 있는 반면, 또한 게놈당 훨씬 개선된 바코드 다양성을 제공하면서 높은 수의 게놈(예: 검정당 100개 초과 게놈, 검정당 500개 초과 게놈, 검정당 1000개 게놈 또는 그 이상의 정도)의 부하를 가능하게 한다.

[0058] 흔히, 샘플은 파티셔닝 단계 전에 비드에 박리가능하게 부착된 올리고뉴클레오티드 태그의 세트와 조합된다. 핵산을 바코딩하는 방법은 본 기술분야에 공지되어 있고, 본원에 기술되어 있다. 일부 예에서, 방법은, 모든 목적 및 특히 핵산에 바코드 또는 다른 올리고뉴클레오티드 태그의 부착에 관련된 모든 교시로 그 전체가 참고로 본원에서 편입되는, 하기에서 기재된 바와 같이 이용된다: Amini et al, 2014, *Nature* Genetics, Advance Online Publication. 추가 예시에서, 올리고뉴클레오티드는 적어도 제1 및 제2 영역을 포함할 수 있다. 제1 영역은, 소정의 파티션 내의 올리고뉴클레오티드 사이에서와 같이, 실질적으로 동일한 바코드 서열일 수 있지만, 상이한 파티션 사이에서와 같이, 상이한 바코드 서열일 수 있고, 대부분의 경우, 상이한 바코드 서열인 바코드 영역일 수 있다. 제2 영역은 파티션 내의 샘플 내 핵산을 프라이밍하는데 사용될 수 있는 N-량체(특정 서열을 표적화하도록 설계된 랜덤 N-량체 또는 N-량체)일 수 있다. 일부 경우에, N-량체가 특정 서열을 표적화하도록 설계된 경우, 이는 특정 염색체(예: 염색체 1, 13, 18 또는 21), 또는 염색체의 영역, 예를 들면, 엑솜 또는 다른 표적화 영역을 표적화하도록 설계될 수 있다. 일부 경우에, N-량체는 특정 유전자 또는 유전 영역, 예를 들면, 질환 또는 장애(예: 암)와 관련된유전자 또는 영역을 표적화하도록 설계될 수 있다. 파티션 내에서, 증폭 반응은 핵산의 길이에 따라 상이한 장소에서 핵산 샘플을 프라이밍하기 위해 제2 N-량체를 사용하여 수행될 수 있다. 증폭의 결과로서, 각 파티션은 동일하거나 거의 동일한 바코드에 부착되고, 각 파티션에서 핵산의 중첩성 작은 단편을 나타낼 수 있는 핵산의 증폭된 제품을 함유할 수 있다. 바-코드는 핵산의 세트가 동일 파티션으로부터 기원되고, 따라서 잠재적으로 또한 핵산의 동일한 가닥으로부터 기원된다는 것을 의미하는 마커로서 기능할 수 있다. 증폭 후, 핵산은 서열분석 알고리즘을 사용하여 풀링되고, 서열분석되고, 정렬될 수 있다. 짧은 서열 판독이 그들의 관련 바코드 서열에 의해 정렬되고, 샘플 핵산의 단일 긴 단편에 부여되기 때문에, 그 서열 상의 식별된 변이체 모두는 단일 기원 단편 및 단일 기원 염색체로 부여될 수 있다. 또한, 복수의 긴 단편에서 다수의 동일 위치 변이체를 정렬함으로써, 그 염색체 기여를 추가로 특성규명할 수 있다. 따라서, 특정한 유전적 변이체의 단계화에 관한 결론은 그 다음, 긴 범위의 게놈 서열 - 예를 들어, 게놈의 저조하게 특성규명된 영역의 연신부를 거쳐 서열 정보의 식별을 거쳐 분석할 수 있음에 따라, 도출될 수 있다. 이러한 정보는 또한 일반적으로 동일한 핵산 가닥 상 또는 상이한 핵산 가닥 상에 존재하는 유전적 변이체의 특정 세트인 일배체형을 식별하는데 유용할 수 있다. 사본 수 변형도 또한 이러한 방식으로 식별될 수 있다.

[0059] 기재된 방법 및 시스템은 현재의 핵산 서열분석 기술 및 이들의 관련된 샘플 제조 방법에 유의미한 이점을 제공한다. 샘플 제조 및 서열분석 방법은 또한 주로 샘플 중의 대부분의 성분을 식별하고 특성규명하는 경향이 있고, 소수 성분, 예를 들면, 게놈 또는 물질의 저조하게 특성규명되거나 고도로 다형성된 영역 유래의 일 염색체에 의해, 또는 추출된 샘플에서 전체 DNA의 작은 비율을 구성하는 하나 또는 소수의 세포, 또는 혈류에서 순환하는 단편화된 종양 세포 DNA 분자에 의해, 기여된 유전 물질을 식별하고 특성규명되도록 설계되지 않는다. 본원에서 기재된 방법은 이들 소수 성분로부터 유전 물질을 증가시키는 선택적 증폭 방법을 포함하고, 상기 유

전 물질의 분자 콘텍스트를 유지하는 능력은 추가로 이들 성분의 유전적 특성규명을 제공한다. 기재된 방법 및 시스템은 또한 더 큰 샘플 내에 존재하는 집단을 검출하는 상당한 이점을 제공한다. 이와 같이, 이들은 일배체형 및 사본 수 변동 평가에 특히 유용하고 - 본원에서 개시된 방법은 샘플 제조 동안 도입된 바이어스 때문에 핵산 표적의 집단에서 저조하게 특성규명되거나 저조하게 나타낸 게놈의 영역에 대해 서열 정보 제공에 또한 유용하다.

[0060] 본원에 개시된 바코딩 기술의 사용은 유전 마커의 소정의 세트에 개별 분자 콘텍스트를 제공하는, 즉(단일 마커와 대조적으로) 유전 마커의 소정의 세트를 개별 샘플 핵산 분자로 부여하는 고유한 능력을 부여하고, 변이체 배위된 조립체를 통해 및/또는 특정 염색체에 복수의 단일 핵산 분자 중에서 광범위하거나 심지어 더 긴 범위의 추론된 개별 분자 콘텍스트를 제공한다. 이러한 유전 마커는 특이적 유전자좌, 예를 들면, SNP와 같은 변이체를 포함할 수 있거나, 그들은 짧은 서열을 포함할 수 있다. 또한, 바코딩의 사용은, 예를 들면, 혈류 중 순환성 종양 DNA의 검출 및 특성규명을 위한 샘플로부터 추출된 전체 핵산 집단의 소수 성분 및 주요 성분을 구별하는 능력을 촉진시키는 추가의 이점을 부여하고, 또한 임의의 증폭 단계 동안 증폭 바이어스를 감소시키거나 제거한다. 또한, 미세 유체 포맷의 구현은 DNA의 매우 작은 샘플 용적 및 낮은 유입량으로 작동하는 능력 뿐만 아니라 게놈 전체 태깅을 촉진시키기 위해 다수의 샘플 파티션(액적)을 신속하게 처리하는 능력을 부여한다.

[0061] 이미 기재된 바와 같이, 본원에 기재된 방법 및 시스템의 이점은, 그들이 보편적으로 이용가능한 짧은 판독 서열분석 기술의 사용을 통해 목적 결과를 달성할 수 있다는 것이다. 이러한 기술은 충분히 특성규명되고 매우 효과적인 프로토콜 및 시약 시스템과 함께 쉽게 이용가능하고 연구 단체에 널리 분산되는 이점을 갖는다. 이들 짧은 판독 서열분석 기술은, 예를 들면, Illumina, inc. (GAIIx, NextSeq, MiSeq, HiSeq, X10), Ion Torrent division of Thermo-Fisher (Ion Proton and Ion PGM), 파이로서열분석 방법, 뿐만 아니라 기타로부터 이용가능한 것을 포함한다.

[0062] 특별한 장점은, 본원에 기재된 방법 및 시스템이 이러한 짧은 판독 서열분석 기술을 이용하고, 이들의 관련된 낮은 오차율 및 고처리율으로 그렇게 수행한다는 점이다. 특히, 본원에 기재된 방법 및 시스템은 상기한 바와 같이, 그러나 짝짓기 쌍 확장을 제외하고, 1000 bp 미만, 500 bp 미만, 300 bp 미만, 200 bp 미만, 150 bp 미만 또는 훨씬 짧은 개별 목적 분자 판독 길이; 및 상기 개별 분자 판독 길이에 대해 5% 미만, 1% 미만, 0.5% 미만, 0.1% 미만, 0.05% 미만, 0.01% 미만, 0.005% 미만, 또는 심지어 0.001% 미만의 서열분석 오차율을 사용하여 개별 분자 판독 길이 또는 콘텍스트를 달성한다.

[0063] *II. 작업 흐름 개요*

[0064] 개시내용에 기재된 방법 및 시스템은 상이한 그룹 또는 상이한 영역으로 핵산 분리를 제공하여 이로써 분리된 핵산이 하나 이상의 다른 핵산으로부터 상대 단리에서 추가 가공 및/또는 반응을 경험할 수 있다. 특정 예시적 경우에서 상기 분리는 개별 샘플(예: 핵산)을 개별 파티션으로 증착 또는 파티셔닝을 포함하고, 여기서 각 파티션은 다른 파티션의 내용물로부터 이의 자체 내용물의 분리를 유지시킨다. 본원에 사용된 바와 같이, 파티션은 다양한 상이한 형태, 예를 들면, 웰, 튜브, 마이크로 또는 나노웰, 관통 홀 등을 포함할 수 있는 용기 또는 용기를 의미한다. 그러나, 바람직한 양태에서, 상기 파티션은 유체 스트림(fluid stream) 내에서 유동가능하다. 이러한 용기는, 예를 들면, 내부 유체 중심 또는 코어를 둘러싸는 외부 장벽을 갖는 마이크로캡슐또는 마이크로-소포로 구성될 수 있거나, 그들은 이의 매트릭스 내부에 물질을 비말 동반하고/하거나 보유할 수 있는 다공성 매트릭스일 수 있다. 바람직한 양태에서, 그러나, 이러한 파티션은 비수성 연속 상, 예를 들면, 오일 상 내에 수성 유체의 액적을 포함할 수 있다. 다양한 상이한 용기는, 예를 들면, 하기에 기술되어 있다: 미국 특허 출원 번호 13/966,150 (2013년 8월 13일 출원됨). 유사하게, 비-수성 또는 인접한 오일 상 속에 안정한 액적을 생성하기 위한 유액 시스템은 예를 들면, 발표된 미국 특허 출원 공보 제2010-0105112호에 상세히 기술되어 있다. 특정 경우에, 미세 유체 채널 네트워크는 본원에 기재된 파티션을 생성하기에 특히 적합하다. 이러한 미세유동성 장치의 예는 2015년 4월 9일 목요일자로 출원된 미국 특허원 제14/682,952호에 상세히 기술된 것들을 포함하며, 이의 전체 개시내용은 모든 목적을 위해 이의 전문이 본원에 참고로 포함된다. 대체 기전은 또한 세포의 수성 혼합물이 비수성 유체에서 압출되는 다공성 막을 포함하는 개별 세포의 파티셔닝에 사용될 수 있다. 그러한 시스템들은 일반적으로 예를 들면, Nanomi, Inc.에서 구입할 수 있다.

[0065] 에멀젼 중 액적을 이용하는 방법에서, 샘플 재료, 예를 들면, 핵산의 개별 파티션으로의 파티셔닝은 일반적으로 수성 샘플 함유 스트림을, 또한 수성 액적이 유동 스트림 파티셔닝 유체 내에서 생성되도록 파티셔닝 유체, 예를 들면, 불소화 오일의 비수성 스트림이 유동하는 접합부로 유동시킴으로써 달성될 수 있고, 여기서 상기 액적은 샘플 재료를 포함한다. 하기 기술된 바와 같이, 파티션, 예를 들어, 액적은 또한 전형적으로 공동-파티셔닝

된 바코드 올리고뉴클레오티드를 포함한다. 임의의 특별한 파티션 내의 샘플 재료의 상대적인 양은, 예를 들면, 수성 스트림 중의 샘플의 농도, 수성 스트림 및/또는 비수성 스트림의 유동 속도 등을 포함하는 시스템의 다양한 상이한 파라미터를 조절함으로써 조정될 수 있다. 본원에 기재된 파티션은 종종 매우 적은 용적을 가짐을 특징으로 한다. 예를 들면, 액적 기반 파티션의 경우에, 액적은 1000 pL 미만, 900 pL 미만, 800 pL 미만, 700 pL 미만, 600 pL 미만, 500 pL 미만, 400pL 미만, 300 pL 미만, 200 pL 미만, 100pL 미만, 50 pL 미만, 20 pL 미만, 10 pL 미만 또는 심지어 1 pL 미만인 전체 용적을 가질 수 있다. 비드와 공동-파티셔닝되는 경우, 파티션 내의 샘플 유체 용적은 상기한 용적의 90% 미만, 상기한 용적의 80% 미만, 70% 미만, 60% 미만, 50% 미만, 40% 미만, 30% 미만, 20% 미만 또는 심지어 10% 미만일 수 있다는 것이 이해될 것이다. 일부 경우에, 낮은 반응 용적 파티션의 사용은 매우 소량의 개시 시약, 예를 들면, 유입 핵산으로 반응을 수행하는데 특히 유리하다. 적은 투입물 핵산을 사용하여 샘플을 분석하는 방법 및 시스템은 미국 특허 출원 번호 14/752,589 및 14/752,602에 나타나 있으며, 이의 전체 내용은 이의 전문이 참고로 본원에 포함된다.

[0066] 일단 샘플이 그들의 각 파티션에 도입되면, 본원에서 기재된 방법 및 시스템에 따라서, 파티션내 샘플 핵산은 일반적으로 선택적 증폭 처리되어, 이로써 드노보 서열분석을 허용하기 위해 표적화된 커버리지에 대하여 관심 있는 게놈의 영역은 게놈의 다른 영역에 비교로 더 높은 비율로 존재한다 (비록, 인정될 바와 같이, 게놈의 상기 다른 영역이, 드노보 커버리지에 대하여 관심 영역이 아니어서, 보다 덜한 정도로, 또한 증폭될 수 있어도). 특정 구현예에서, 관심 게놈 영역은 게놈의 상기 선택된 영역의 적어도 1X, 2X, 5X, 10X, 20X, 30X, 40X 또는 50X 커버리지를 제공하기 위해 증폭된다. 추가 구현예에서, 파티션내 모든 핵산은 증폭되지만, 선택된 게놈 영역은 적어도 1-5, 2-10, 3-15, 4-20, 5-25, 6-30, 7-35, 8-40, 9-45, 또는 10-50 배 더 많은 증폭산물이 게놈의 다른 부분으로부터 보다 상기 선택된 게놈 영역으로부터 생산되는 표적화된 방식으로 증폭된다.

[0067] 게놈의 선택된 영역의 선택적 증폭과 동시에 또는 이에 이어서, 파티션 내 핵산(또는 이의 단편)은 고유 식별자와 함께 제공되어 이로써, 상기 핵산의 특성규명시 이들이 그들의 각 기원으로부터 유도된 바와 같이 부여될 수 있다. 따라서, 샘플 핵산은 전형적으로 고유 식별자와 공동-파티셔닝된다. 일부 예시적 구현예에서, 상기 고유 식별자는 바코드 서열이다. 명료함을 위해, 본원에서 논의의 대부분은 바코드 서열을 포함한 식별자에 관련하지만, 인정될 바와 같이, 서열 판독을 위하여 분자 콘텍스트를 유지하기 위해 사용될 수 있는 임의의 고유 식별자는 본원에서 기재된 방법에 따라 사용될 수 있다. 일부 바람직한 양태에서, 고유 식별자는 핵산 샘플에 부착될 수 있는 핵산 바코드 서열을 포함하는 올리고뉴클레오티드 형태로 제공된다. 올리고뉴클레오티드는 분배되어 특정한 파티션 내 올리고뉴클레오티드 사이에서와 같이, 이에 함유된 핵산 바코드 서열은 동일하지만, 상이한 파티션 사이에서와 같이, 올리고뉴클레오티드는 상이한 바코드 서열일 수 있거나, 바람직하게는 이를 갖는다. 바람직한 양태에서, 둘 이상의 상이한 바코드 서열이 존재할 수 있지만, 일부 국면에서, 하나의 핵산 바코드 서열만이 소정의 파티션과 관련될 것이다.

[0068] 핵산 바코드 서열은 전형적으로 올리고뉴클레오티드의 서열내 6 내지 약 20개 이상의 뉴클레오티드를 포함할 것이다. 이러한 뉴클레오티드는, 즉, 인접 뉴클레오티드의 단일 연신부로 완전히 연속적일 수 있거나, 그들은 하나 이상의 뉴클레오티드에 의해 분리되는 둘 이상의 별도의 서열로 분리될 수 있다. 전형적으로, 분리된 서열은 길이가 전형적으로 약 4 내지 약 16개 뉴클레오티드일 수 있다.

[0069] 공동-파티셔닝된 올리고뉴클레오티드는 또한 전형적으로 파티셔닝된 핵산의 처리에 유용한 다른 기능적 서열을 포함한다. 이러한 서열은, 예를 들면, 관련된 바코드 서열을 부착하고 프라이머를 서열분석하면서 파티션 내의 개별 핵산으로부터 게놈 DNA를 증폭시키기 위한 표적화된 또는 랜덤/범용 증폭 프라이머 서열, 예를 들면, 서열의 존재를 식별하거나 바코딩 핵산을 풀 다운(pulling down)시키기 위한 하이브리드화 또는 프로빙 서열, 또는 임의의 다수의 다른 잠재적인 기능적 서열을 포함한다. 또한, 샘플 재료와 함께 올리고뉴클레오티드 및 연관된 바코드 및 다른 기능성 서열의 공동-파티셔닝은 예를 들어, 하기에 기술되며, 이의 전체 개시내용은 이의 전문이 본원에 참고로 포함된다: 미국 특허 출원 번호 14/316,383; 14/316,398; 14/316,416; 14/316,431; 14/316,447; 및 14/316,463.

[0070] 간단히, 하나의 예시적 공정에서, 각각 비드에 박리가능하게 부착된 상기한 올리고뉴클레오티드의 다수를 포함할 수 있는 비드가 제공되고, 여기서, 특정 비드에 부착된 올리고뉴클레오티드는 모두 동일한 핵산 바코드 서열을 포함할 수 있지만, 다양한 바코드 서열의 다수가 사용된 비드 집단에 걸쳐 나타날 수 있다. 전형적으로, 비드의 집단은 적어도 1000개의 상이한 바코드 서열, 적어도 10,000개의 상이한 바코드 서열, 적어도 100,000개의 상이한 바코드 서열, 또는 일부 경우에, 적어도 1,000,000개의 상이한 바코드 서열을 포함할 수 있는 다양한 바코드 서열 라이브러리를 제공할 수 있다. 추가로, 각 비드는 부착된 다수의 올리고뉴클레오티드 분자가 전형적으로 제공될 수 있다. 특히, 개별 비드 상에 바코드 서열을 포함하는 올리고뉴클레오티드의 분자의 수는 적어도

약 10,000개 올리고뉴클레오티드, 적어도 100,000개 올리고뉴클레오티드 분자, 적어도 1,000,000개 올리고뉴클레오티드 분자, 적어도 100,000,000개 올리고뉴클레오티드 분자, 및 일부 경우에, 적어도 10억개 올리고뉴클레오티드 분자일 수 있다.

[0071] 올리고뉴클레오티드는 특정 자극을 비드에 적용시 비드로부터 박리가능할 수 있다. 일부 경우에, 자극은, 예를 들면, 올리고뉴클레오티드를 박리시킬 수 있는 광-불안정한 연결의 절단을 통한 광-자극일 수 있다. 일부 경우에, 비드 환경의 온도 상승이 연결의 절단 또는 비드로부터 올리고뉴클레오티드의 다른 박리를 유도할 수 있는 경우, 열 자극이 사용될 수 있다. 일부 경우에, 비드에 대한 올리고뉴클레오티드의 연결을 절단하거나 다르게는 비드로부터 올리고뉴클레오티드의 박리를 유도할 수 있는 화학적 자극이 사용될 수 있다.

[0072] 본원에 기재된 방법 및 시스템에 따라서, 부착된 올리고뉴클레오티드를 포함하는 비드는 단일 비드 및 단일 샘플이 개별 파티션 내에 함유되도록 개별 샘플과 공동-파티셔닝될 수 있다. 일부 경우에, 단일의 비드 파티션이 바람직한 경우, 유체의 상대적인 유동 속도를 조절함으로서 평균적으로, 파티션이 파티션 당 1 미만의 비드를 함유하도록 하여 점유된 이들 파티션이 주로 단일로 점유되도록 하는 것이 바람직할 수 있다. 마찬가지로, 높은 비율의 파티션이 점유되어, 예를 들면, 점유되지 않은 파티션의 적은 비율만을 허용함을 제공하기 위해 유동 속도를 조절하는 것을 원할 수 있다. 바람직한 양태에서, 유동 및 채널 건축양식은 바람직한 수의 단일로 점유된 파티션, 특정 수준 이하의 점유되지 않은 파티션 및 특정 수준 이하의 다수의 점유된 파티션을 보증하도록 조절한다.

[0073] 도 3은 바코딩에 이어, 특히, 사본 수 변형 또는 일배체형 검정을 위해 특히 사용하기 위한 샘플 핵산을 서열분석하기 위한 예시적 방법을 예시한다. 먼저, 핵산을 포함하는 샘플은 공급원(300)으로부터 수득할 수 있고, 바코딩 비드의 세트도 또한 수득될 수 있다(310). 비드는 바람직하게는 하나 이상의 바코드 서열 뿐만 아니라 프라이머, 예를 들면, 랜덤 N-량체 또는 다른 프라이머를 함유하는 올리고뉴클레오티드에 연결된다. 바람직하게는, 바코드 서열은 바코딩 비드로부터, 예를 들면, 바코드와 비드 사이의 연결의 절단을 통해 또는 바코드를 박리시키는 기초 비드의 분해를 통해 또는 둘의 조합으로 박리가능하다. 예를 들면, 특정 바람직한 양태에서, 바코딩 비드는 바코드 서열을 박리시키는 환원제와 같은 제제에 의해 분해 또는 용해될 수 있다. 이 예에서, 핵산(305)을 포함하는 적은 양의 샘플, 바코딩 비드(315), 그리고 임의로, 다른 시약, 예를 들면, 환원제(320)를 조합하고, 파티셔닝에 적용한다. 예로써, 이러한 파티셔닝은 액적 생성 시스템, 예를 들면, 미세 유체 장치(325)에 성분을 도입함을 포함할 수 있다. 미세 유체 장치, 325의 보조로, 유중수 에멀젼, 330이 형성되고, 여기서, 상기 에멀젼은 샘플 핵산, 305, 환원제, 320 및 바코딩 비드, 315를 함유하는 수성 액적을 함유한다. 환원제는 바코딩 비드를 용해시키거나 분해하여 액적, 335 내의 비드로부터 바코드 및 랜덤 N-량체를 포함하는 올리고뉴클레오티드를 박리시킬 수 있다. 이어서, 랜덤 N-량체는 샘플 핵산의 상이한 영역을 프라이밍하여 증폭 후 샘플의 증폭된 사본을 유도할 수 있고, 여기서 각 사본은 바코드 서열, 340로 태깅된다. 바람직하게는, 각 액적은 동일한 바코드 서열 및 상이한 랜덤 N-량체 서열을 함유하는 올리고뉴클레오티드의 세트를 함유한다. 후속적으로, 에멀젼은 파괴되고, 345, 추가의 서열(예: 특정 서열분석 방법, 추가의 바코드 등에 보조하는 서열)을, 예를 들면, 증폭 방법, 350을 통해 첨가할 수 있다(예: PCR). 이어서, 서열분석을 수행할 수 있고, 355, 알고리즘을 서열분석 데이터를 해석하기 위해 적용한다, 360. 서열분석 알고리즘은 일반적으로, 예를 들면, 서열분석 판독을 정렬시키고/시키거나 특정 서열 판독이 속하는 샘플을 식별하기 위해 바코드의 분석을 실행할 수 있다. 또한, 및 본원에서 기재된 바와 같이, 이들 알고리즘은 그들의 유래 분자 콘텍스트에 사본의 서열을 부여하기 위해 또한 추가로 사용될 수 있다.

[0074] 인정될 바와 같이, 바코드 서열 340을 이용한 태깅 이전에 또는 이와 동시에, 샘플은 게놈의 선택된 영역의 표적화된 커버리지를 제공하기 위해 본원에서 기재된 임의의 방법에 따라 증폭될 수 있다. 상기 표적화된 커버리지는 게놈의 다른 영역으로부터 증폭산물에 비교된 경우 게놈의 상기 선택된 영역을 함유한 파티션에서 핵산(또는 이의 부분)의 서열을 나타내는 증폭산물의 더 큰 집단을 일반적으로 초래한다. 그 결과, 게놈의 다른 영역으로부터 보다 게놈의 선택된 영역으로부터 파티션내 바코드 서열 340을 함유한 다수의 증폭된 사본일 것이다.

[0075] 일부 구현예에서 그리고 상기 중 임의의 것에 따라, 상이한 증폭 프로토콜은 단편에 바코드 서열을 부착시키기 위해 사용된 프로토콜보다 게놈의 선택된 영역의 부분을 함유한 단편의 증폭을 선호하기 위해 사용된다. 일 비제한적 예시에서, 표적화된 PCR 프라이머를 이용한 선택적 증폭은 표준 PCR 증폭 열 사이클링 조건하에서 수행되고, 반면에 바코드의 부착용 증폭은 온도에서 급격한 저하 그 다음 온도 증가의 저속 램프로 수행되어 랜덤 N-량체의 프라이밍 및 확장을 허용한다.

[0076] 상기 나타낸 바와 같이, 단일 점유도는 대부분의 바람직한 상태일 수 있지만, 다수의 점유된 파티션, 또는 점유되지 않은 파티션이 흔히 존재할 수 있음은 인식될 것이다. 샘플 및 바코드 올리고뉴클레오티드를 포함하는 비드를 공동-파티셔닝하기 위한 미세 유체 채널 구조의 예는 도 4에 개략적으로 도시된다. 도시된 바와 같이, 채널 분절 402, 404, 406, 408 및 410이 채널 접합부 412에서 유체 연통으로 제공된다. 개별 샘플 414를 포함하는 수성 스트림은 채널 접합부 412를 향해 채널 분절 402를 통해 유동한다. 본원의 다른 곳에서 기재된 바와 같이, 이들 샘플은 파티셔닝 공정 전에 수성 유체 내에 현탁될 수있다.

[0077] 동시에, 바코드 운반 비드(416)를 포함하는 수성 스트림은 채널 접합부(412)를 향해 채널 분절(404)를 통해 유동한다. 비수성 파티셔닝 유체는 양태 채널(406 및 408) 각각으로부터 채널 접합부(412)로 도입되고, 조합된 스트림은 유출 채널(410)로 유동한다. 채널 접합부(412) 내에서, 채널 분절(402 및 404)로부터의 두 개의 조합된 수성 스트림이 조합되고, 공동-파티셔닝된 샘플(414) 및 비드(416)를 포함하는 액적(418)으로 파티셔닝된다. 이미 언급한 바와 같이, 채널 접합부(412)에서 조합하는 유체 각각의 유동 특성을 조절할 뿐만 아니라 채널 접합부의 기하학을 조절함으로써, 생성되는 파티션(418) 내에서 비드, 샘플 또는 둘 다의 목적하는 점유도 수준을 달성하기 위해 조합 및 파티셔닝을 최적화할 수 있다.

[0078] 알 수 있는 바와 같이, 예를 들면, 화학적 자극, 핵산 확대, 전사 및/또는 증폭 시약, 예를 들면, 폴리머라아제, 역전사 효소, 뉴클레오시드 트리포스페이트 또는 NTP 유사체, 프라이머 서열 및 추가의 공동 인자, 예를 들면, 이러한 반응에 사용된 2가 금속 이온, 결찰 반응 시약, 예를 들면, 리가아제 효소 및 결찰 서열, 염료, 표지, 또는 다른 태깅 시약을 포함하는 다수의 다른 시약은 샘플 및 비드와 함께 공동-파티셔닝될 수 있다. 프라이머 서열은 게놈의 선택된 영역의 증폭에 관련된 랜덤 프라이머 서열 또는 표적화된 PCR 프라이머 또는 이들의 조합을 포함할 수 있다.

[0079] 공동-파티셔닝되면, 비드 상에 배치된 올리고뉴클레오티드는 파티셔닝된 샘플을 바코딩하고 증폭시키기 위해 사용될 수 있다. 샘플의 증폭 및 바코딩에서 이들 바코드 올리고뉴클레오티드의 특히 품격있는 사용 공정은 모든 목적 및 특히 가공 핵산 그리고 게놈 재료의 서열분석 및 다른 특성규명에 관한 모든 쓰여진 설명, 도면 및 실시예로 그 전체가 참고로 본원에서 편입되는 미국 특허 출원 번호 USSNs 14/316,383; 14/316,398; 14/316,416; 14/316,431; 14/316,447; 및 14/316,463에서 상세히 기재된다. 간단히, 하나의 국면에서, 올리고뉴클레오티드는 샘플과 함께 공동-파티셔닝되고, 샘플을 갖는 파티션에서 이들의 비드로부터 박리되는 비드 상에 존재한다. 올리고뉴클레오티드는, 전형적으로 바코드 서열과 함께 이의 5' 말단에 프라이머 서열을 포함한다. 프라이머 서열은 무작위이거나 구조화될 수 있다. 랜덤 프라이머 서열은 일반적으로 샘플의 다수의 상이한 영역을 무작위로 프라이밍하도록 의도된다. 구조화된 프라이머 서열은 샘플의 특정 표적화된 영역의 업스트림을 프라이밍하기 위해 표적화된 정의된 서열 뿐만 아니라, 비제한적으로 특정 염기의 백분율 (예컨대 GC N-량체의 백분율)을 함유한 프라이머, 부분적으로 또는 전체적으로 퇴화 서열을 함유한 프라이머, 및/또는 본원에서 임의의 설명에 따라 부분적으로 랜덤 및 부분적으로 구조화된 서열을 함유한 프라이머를 포함한, 부분적으로 정의된 구조의 일부 종류를 갖는 프라이머를 포함할 수 있다. 인정될 바와 같이, 랜덤 및 구조화된 프라이머의 상기 유형의 임의의 하나 이상은 임의의 조합으로 올리고뉴클레오티드에 포함될 수 있다.

[0080] 박리되면, 올리고뉴클레오티드의 프라이머 부분은 샘플의 상보적 영역으로 어닐링할 수 있다. 또한 연장 반응 시약, 예를 들면, DNA 폴리머라아제, 뉴클레오시드 트리포스페이트, 또한 샘플 및 비드로 공동-파티셔닝된 동시-인자(예를 들면, $Mg^{2+}$ 또는 $Mn^{2+}$ 등)는 주형으로서 샘플을 사용하여 프라이머 서열을 연장시킴으로써 프라이머가 어닐링되는 주형의 가닥에 대한 상보적 단편을 생산하며, 당해 상보적 서열은 올리고뉴클레오티드 및 이의 연관된 바코드 서열을 포함한다. 샘플의 상이한 부분으로의 다중 프라이머의 어닐링 및 확대는 샘플의 중첩성 상보적 단편의 거대한 풀을 초래할 수 있고, 각각은 그것이 생성된 파티션을 나타내는 자체 바코드 서열을 포함한다. 일부 경우에, 이러한 상보적 단편은 자체가 다시 바코드 서열을 포함하는 보체의 보체를 생성하기 위해 파티션에 존재하는 올리고뉴클레오티드에 의해 프라이밍되는 주형으로서 사용될 수 있다. 일부 경우에, 이 복제 공정은 제1 보체가 복제되는 경우, 그것이 이의 말단 또는 부근에 두 개의 상보적 서열을 생성하여 반복 사본을 추가로 생성하기 위한 기초가 되는 분자의 능력을 감소시키는 헤어핀 구조 또는 부분적 헤어핀 구조의 형성을 가능하게 하도록 구성된다. 이것의 한 예의 개략적 예시는 도 5에 도시되어 있다.

[0081] 도면이 도시하는 바와 같이, 바코드 서열을 포함하는 올리고뉴클레오티드는, 예를 들면, 샘플 핵산(504)과 함께 에멀젼 중의 액적(502)으로 공동-파티셔닝된다. 본원에서 다른 곳에 언급된 바와 같이, 올리고뉴클레오티드 508은 샘플 핵산 504와 공동-파티셔닝되는 비드 506상에 제공될 수 있고, 상기 올리고뉴클레오티드는, 패널 A에 나타낸 바와 같이, 비드 506으로부터 바람직하게는 방출가능하다. 올리고뉴클레오티드 508은, 하나 이상의 작용성

서열, 예를 들면, 서열 510, 514 및 516에 더하여, 바코드 서열 512를 포함한다. 예를 들어, 올리고뉴클레오티드(508)는 바코드 서열(512), 뿐만 아니라 소정의 서열분석 시스템을 위한 부착 또는 고정화 서열로서 기능할 수 있는 서열(510), 예를 들면, Illumina Hiseq 또는 Miseq 시스템의 유동 세포에서 부착용으로 사용되는 P5 서열을 포함하는 것으로 도시된다. 도시된 바와 같이, 올리고뉴클레오티드는 또한 샘플 핵산(504)의 부분의 프라이밍 복제를 위한 랜덤 또는 표적화 N-량체를 포함할 수 있는 프라이머 서열(516)을 포함한다. 또한, 올리고뉴클레오티드(508) 내에 서열분석 시스템에서 합성 반응에 의해 폴리머라아제 매개된, 주형 지시된 서열분석을 프라이밍하는데 사용되는 서열분석 프라이밍 영역, 예를 들면, “read1” 또는 R1 프라이밍 영역을 제공할 수 있는 서열(514)이 포함된다. 많은 경우에, 바코드 서열(512), 고정화 서열(510) 및 R1 서열(514)은 소정의 비드에 부착된 모든 올리고뉴클레오티드에 공통일 수 있다. 프라이머 서열 516은 랜덤 N-량체 프라이머에 대해 가변적일 수 있거나, 특정의 표적화된 적용을 위한 특정한 비드에서 올리고뉴클레오티드에 대해 공통일 수 있다.

[0082] 프라이머 서열(516)의 존재에 기초하여, 올리고뉴클레오티드는 폴리머라아제 효소 및 또한 비드(506) 및 샘플 핵산(504)과 공동-부분화된 다른 확대 시약을 사용하여 올리고뉴클레오티드(508 및 508a)의 확대를 가능하게 하는 패널 B에 도시된 바와 같은 샘플 핵산을 프라이밍할 수 있다. 패널 C에 도시된 바와 같이, 랜덤 N-량체 프라이머의 경우, 샘플 핵산(504)의 복수의 상이한 영역에 어닐링하는 올리고뉴클레오티드의 확대 후, 핵산의 다중 중첩 보체 또는 단편, 예를 들면, 단편(518 및 520)이 생성된다. 샘플 핵산의 부분에 상보적인 서열 부분, 예를 들면, 서열(522 및 524)을 포함하지만, 이들 작제물은 본원에서 일반적으로 부착된 바코드 서열을 갖는 샘플 핵산(504)의 단편을 포함하는 것으로 칭명된다. 알 수 있는 바와 같이, 상기한 바와 같은 주형 서열의 복제 부분은 종종 본원에서 그 주형 서열의 ”단편”으로 칭명된다. 그러나, 전술한 내용에도 불구하고, 용어 “단편”은 주형 서열의 부분을 제공하는 다른 기전, 예를 들면, 소정의 서열 분자의 실제 단편화에 의해, 예를 들면, 효소, 화학적 또는 기계적 간편화를 통해 생성된 것들을 포함하여 기원 핵산 서열, 예를 들면, 주형 또는 샘플 핵산의 일부의 임의의 대표를 포함한다. 그러나, 바람직한 양태에서, 주형 또는 샘플 핵산 서열의 단편은 기본 서열 또는 이의 보체의 복제 부분을 나타낼 것이다.

[0083] 바코딩된 핵산 단편은 그 다음, 예를 들면, 서열분석을 통해 특성규명 처리될 수 있거나, 이들은 패널 D에 나타낸 바와 같이 공정에서 추가로 증폭될 수 있다. 예를 들어, 비드 506으로부터 또한 방출된, 추가의 올리고뉴클레오티드, 예를 들면, 올리고뉴클레오티드 508b는 단편 518 및 520을 프라이밍할 수 있다. 특히, 다시, (많은 경우에, 소정의 파티션 중의 다른 랜덤 N-량체, 예를 들면, 프라이머 서열(516)과 상이할 것인) 올리고뉴클레오티드(508b) 중의 랜덤 N-량체 프라이머(516b)의 존재에 기초하여, 올리고뉴클레오티드는 단편(518)으로 어닐링하고, 샘플 핵산 서열의 일부의 복제본을 포함하는 서열(528)을 포함하는 단편(518)의 적어도 일부에 대한 보체(526)을 생성하기 위해 확대된다. 올리고뉴클레오티드 (508b)의 연장은 이것이 단편 (518)의 올리고뉴클레오티드 부분 (508)을 통해 복제될 때까지 지속된다. 본원의 다른 곳에서 언급된 바와 같고, 패널 D에 도시된 바와 같이, 올리고뉴클레오티드는 목적하는 지점에서, 예를 들면, 단편(518) 내에 포함되는 올리고뉴클레오티드(508)의 서열(516 및 514)을 통한 복제 후, 폴리머라아제에 의한 복제의 중지를 유발하도록 구성될 수 있다. 본원에 기재된 바와 같이, 이는, 예를 들면, 사용된 폴리머라아제 효소에 의해 처리될 수 없는 상이한 뉴클레오티드 및/또는 뉴클레오티드 유사체의 도입을 포함하는 상이한 방법에 의해 달성될 수 있다. 예를 들면, 이는 서열 영역 (512) 내에서 우라실 함유 뉴클레오티드를 혼입시켜 비-우라실 내성 폴리머라아제를 방지함으로써 당해 영역의 복제를 중지시킬 수 있다. 그 결과, 한쪽 말단에 바코드 서열(512), 부착 서열(510), R1 프라이머 영역(514), 및 랜덤 N-량체 서열(516b)을 포함하는 전장 올리고뉴클레오티드(508b)를 포함하는 단편(526)이 생성된다. 서열의 다른 말단에는 제1 올리고뉴클레오티드 (508)의 랜덤 N-량체에 대한 상보체 (516'), 및 서열 (514')로서 나타낸 R1 서열의 모두 또는 일부에 대한 상보체가 포함될 것이다. 그러면, R1 서열 (514) 및 그것의 상보체 (514')는 함께 혼성하여 부분적 헤어핀 구조 (528)를 형성할 수 있다. 인식되는 바와 같이, 랜덤 N-량체는 상이한 올리고뉴클레오티드 중에서 상이하므로, 이들 서열 및 이들의 상보체는 랜덤 N-량체 (516)에 대해 상보적인, 헤어핀 형성, 예를 들면, 서열 (516')에 관여하는 것으로 기대되지 않을 수 있으며, 랜덤 N-량체 서열(516b)에 대해 상보적인 것으로 예측되지 않을 수 있다. 이는 N-량체가 소정의 파티션 내의 올리고뉴클레오티드 사이에 공통인 다른 적용, 예를 들면, 표적화 프라이머에 대한 경우는 아닐 것이다. 이러한 부분 헤어핀 구조를 형성함으로써, 추가의 복제로부터 샘플 서열의 제1 수준 복제물의 제거를 가능하게 하고, 예를 들면, 사본의 반복 사본을 방지한다. 부분 헤어핀 구조는 또한 생성된 단편, 예를 들면, 단편(526)의 후속 가공에 유용한 구조를 제공한다.

[0084] 이어서, 복수의 상이한 파티션으로부터의 모든 단편은 본원에 기재된 바와 같은 고처리율 서열분석기 상의 서열분석을 위해 풀링될 수 있다. 각 단편이 본래의 이의 파티션에 대해 암호화되기 때문에, 그 단편의 서열은 바코드의 존재에 기초하여 이의 기원으로 다시 부여될 수 있다. 이는 도 6a에 도식적으로 예시된다. 일례로 도시된

바와 같이, 제1 공급원(600)(예: 개별 염색체, 핵산의 가닥 등)으로부터 유도된 핵산(604) 및 상이한 염색체(602) 또는 핵산의 가닥으로부터 유도된 핵산(606)은 각각 상기한 바와 같은 바코드 올리고뉴클레오티드의 자체 세트와 함께 파티셔닝된다.

[0085] 각 파티션 내에서, 각 핵산(604 및 606)은 이어서 별도로 제1 단편(들)의 제2 단편의 중첩 세트, 예를 들면, 제2 단편 세트(608 및 610)를 제공하기 위해 처리한다. 이 처리는 또한 특정 제1 단편으로부터 유래된 제2 단편 각각에 대해 동일한 바코드 서열을 갖는제2 단편을 제공한다. 도시된 바와 같이, 제2 단편 세트(608)의 바코드 서열은 “1”로 표시되는 반면, 단편 세트(610)의 바코드 서열은 “2”로 표시된다. 바코드의 다양한 라이브러리는 상이한 단편 세트의 다수를 차별적으로 바코딩하는데 사용될 수 있다. 그러나, 상이한 제1 단편으로부터 모든 제2 단편 세트가 상이한 바코드 서열로 바코딩될 필요는 없다. 사실상, 다수 경우에서, 복수의 상이한 제1 단편은 동일한 바코드 서열을 포함하기 위해 동시에 처리될 수있다. 다양한 바코드 라이브러리는 본원의 다른 곳에서 상세히 기재된다.

[0086] 예를 들면, 단편 세트 608 및 610으로부터 바코딩된 단편은 그 다음, 예를 들어, Illumina 또는 Ion Torrent division of Thermo Fisher, Inc.로부터 이용가능한 합성 기술에 의한 서열을 이용하여 서열분석에 대하여 풀링될 수 있다. 일단 서열분석되면, 풀링된 단편 612로부터 서열 판독은, 포함된 바코드에 적어도 부분적으로 기반된, 및 임의로, 및 바람직하게는, 단편 자체의 서열에 부분적으로 기반된, 예를 들면, 응집된 판독 614 및 616에 나타낸 바와 같이, 그들의 각 단편 세트에 부여될 수 있다. 이후에, 각각의 단편 세트에 대한 기여된 서열 판독물을 조립하여 각각의 샘플 단편에 대한 조립된 서열, 예를 들면, 서열 (618) 및 (620)을 제공하며, 이는 최종적으로 이들 각각의 본래 염색체 또는 공급원 핵산 분자 (600 및 602)에 다시 부여될 수 있다. 게놈 서열을 조립하기 위한 방법 및 시스템은 예를 들면, 하기에 기술되어 있으며, 이의 전체 개시내용은 이의 전문이 본원에 참고로 포함된다: 미국 특허 출원 번호 14/752,773 (2015년 6월 26일 출원됨).

[0087] 일부 구현예에서 및 도 6b에서 예시된 바와 같이, 프라이머 세트 613은 단편 세트 608 또는 610을 함유한 파티션과 함께 포함된다. 프라이머 세트 613은 추가 구현예에서 게놈의 선택된 영역에 관련하고, 이로써 바코드 서열(바코드 608에 대하여 "1" 및 610에 대하여 "2")의 제공 이전에, 이와 동시에 또는 이에 이어서, 단편 세트 608 및 610은 게놈의 선택된 영역이 게놈의 다른 영역에 대해 추가의 정도로 커버되도록 증폭된다. 도 6b에서 묘사된 예시적인 구현예에서, 단편 세트 608은 프라이머 세트 613이 관련되는 게놈의 선택된 영역으로부터 서열을 함유하지만, 단편 세트 610은 게놈의 상기 선택된 영역으로부터 서열을 함유하지 않는다. 이와 같이, 세트 610으로부터 보다 세트 608로부터 단편의 커버리지 (예를 들면, 더 많은 사본)가 증가될 것이다. 따라서, 풀링된 단편 612는 표적화된 방식으로 증폭되었던 단편을 함유하는 바코딩된 단편을 함유하여, 단편 세트 610 ("2" 바코딩된 단편)으로부터 보다 단편 세트 608 ("1" 바코딩된 단편)으로부터 서열 판독의 더 큰 비율을 허용한다. 또한, 바코드 때문에, 세트 608로부터 서열 판독의 더 큰 비율은, 풀링된 세트 612에서 단편의 나머지처럼, (도 6a에서 보여진) 그들의 각 본래 공급원 핵산 분자 600 및 602에 역으로 부여될 수 있다.

[0088] *III. 핵산 서열분석에 방법 및 시스템의 적용*

[0089] 본원에서 기재된 방법, 조성물, 및 시스템은 핵산 서열분석 기술에서 사용에 특히 잘 받아들인다. 상기 서열분석 기술은, 짧은-판독 및 긴-판독 서열분석 기술을 포함하여, 당해 기술에서 공지된 임의의 기술을 포함할 수 있다. 특정 양태에서, 본원에서 기재된 방법, 조성물 및 시스템은 짧은 판독, 고 정확도 서열분석 기술에서 사용된다.

[0090] 본원에서 기재된 방법, 조성물, 및 시스템은 저조하게 특성규명된, 고도로 다형성된, 및/또는 참조 서열로부터 분지된 게놈의 영역의 유전적 특성규명을 허용한다. 특히, 본원에서 기재된 방법, 조성물 및 시스템은 추가의 불필요한 서열 정보가 게놈의 상기 선택된 부분으로부터 수득될 수 있는 정도로 게놈의 선택된 부분의 증가된 및 불필요한 커버리지를 제공한다. 특정 경우에서, 상기 추가의 서열 정보 (예컨대, 증가된 범위의 게놈의 표적화 영역)는 게놈의 상기 선택된 부분의 드노보 서열분석을 허용하기 위해 충분한 정보를 제공한다. 상기 드노보 서열분석은 저조하게 특성규명된, 고도로 다형성된, 및/또는 참조 서열에서 갈라진 게놈의 영역에 대하여 특히 유용하다. 인정될 바와 같이, 인간 게놈의 유의미한 백분율(예를 들어 Altemose et al., *PLOS* Computational Biology, May 15, 2014, Vol. 10, Issue 5에 따라, 적어도 5-10%)는 계속해서 조립되지 않고, 맵핑되지 않고, 저조하게 특성규명된다. 참조 조립체는 일반적으로, 주로 동원체에 근접하여, 그리고 단부동원체성 염색체의 짧은 아암 상에 발견되는, 다중-메가염기 이색성 갭으로서 이들 누락 영역에 주석에 주석을 단다. 게놈의 상기 누락 부분은 일반적으로 사용된 서열분석 기술을 이용하여 정확한 특성규명에 대해 계속해서 내성인 구조적 특징을 포함한다. 정확한 특성규명에 내성인 추가의 예시적인 영역은 밀접한 상동성 위유전자(예를 들어 SMN1/2

Cyp2d6)를 갖는 영역, 비제한적으로 트랜스포존(예컨대 SINEs, LINEs)을 포함한, 게놈 전반에 걸쳐 실질적인 반복된 서열을 갖는 영역, 뿐만 아니라 참조 서열이 불량한 가이드로서 작용하는 엄청난 변동을 갖는 영역(예컨대 인간 백혈구 항원(HLA) 착물에 대하여 유전자를 암호화하는 영역)을 비제한적으로 포함한다. 본원에서 기재된 방법, 조성물, 및 시스템은 관심 영역의 선택적 증폭을 분자 콘텍스트를 유지하기 위한 능력과 조합하여, 이로써 일반적으로 저조하게 특성규명된 게놈 영역의 드노보 서열분석을 허용한다.

[0091] 특이적 사례에서, 본원에서 기재된 방법은 게놈의 선택된 영역이 서열분석 이전에 임의로 증폭되는 단계를 포함한다. (비제한적으로 PCR 증폭을 포함한) 당해 기술에서 공지된 방법을 이용하여 일반적으로 수행되는, 상기 증폭은 게놈의 선택된 영역의 적어도 1X, 2X, 3X, 4X, 5X, 6X, 7X, 8X, 9X, 10X, 11X, 12X, 13X, 14X, 15X, 16X, 17X, 18X, 19X, 또는 20X 커버리지를 제공하고, 이로써 상기 선택된 영역의 드노보 서열분석을 허용하기 위해 핵산의 양을 제공한다. 추가 구현예에서, 증폭은 게놈의 선택된 영역의 적어도 1X-30X, 2X-25X, 3X-20X, 4X-15X, 또는 5X-10X 커버리지를 제공한다.

[0092] 증폭은 게놈의 선택된 영역 내에서 또는 근처에서 서열에 상보적인 프라이머의 확장을 통해 일반적으로 수행된다. 일부 경우에서, 관심 영역에 걸쳐 타일링하도록 설계되는 프라이머의 라이브러리가 사용되고 - 환언하면, 프라이머의 라이브러리는 게놈의 선택된 영역을 따라 특정한 거리에서 영역을 증폭시키도록 설계된다. 일부 사례에서, 선택적 증폭은 게놈의 선택된 영역을 따라 매 10, 15, 20, 25, 50, 100, 200, 250, 500, 750, 1000, 또는 10000 염기에 상보적인 프라이머를 이용한다. 또 다른 추가 예에서, 프라이머의 타일링된 라이브러리는 거리의 혼합을 포획하도록 설계되거나 - 상기 혼합은 거리의 랜덤 혼합일 수 있거나 선택된 영역의 특정한 부분 또는 백분율이 상이한 프라이머 쌍에 의해 증폭되도록 지능적으로 설계된다. 추가 구현예에서, 프라이머 쌍은 각 쌍이 게놈의 선택된 부분의 임의의 인접 영역의 약 1-5%, 2-10%, 3-15%, 4-20%, 5-25%, 6-30%, 7-35%, 8-40%, 9-45%, 또는 10-50%를 증폭시키도록 설계된다.

[0093] 특정 구현예에서 그리고 상기 중 임의의 것에 따라, 증폭은 적어도 3 메가염기쌍(Mb)인 게놈 영역에 걸쳐 발생한다. 추가 구현예에서, 본원에서 기재된 임의의 방법에 따라서 임의로 증폭되는 게놈의 선택된 영역은 적어도 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 또는 10 Mb이다. 또 다른 추가의 구현예에서, 게놈의 선택된 영역은 그 길이가 약 2-20, 3-18, 4-16, 5-14, 6-12, 또는 7-10 Mb이다. 상기 논의된 바와 같이, 증폭은 이들 영역의 말단에서 또는 말단 근처에서 서열에 상보적인 단일 프라이머 쌍을 이용하여 이들 영역에 걸쳐 발생할 수 있다. 다른 구현예에서, 증폭은 영역의 길이를 거쳐 타일링되는 프라이머 쌍의 라이브러리로 수행되어, 이로써 규칙적 분절, 랜덤 분절, 또는 영역을 따라 상이한 분절 거리의 일부 조합이, 상기 설명에 따른 커버리지의 정도로, 증폭된다.

[0094] 일부 구현예에서, 게놈의 선택된 영역의 선택적 증폭에서 사용된 프라이머는 프라이머 자체가 증폭되지 않도록 우라실을 함유한다.

[0095] 일반적으로, 본원에서 기재된 방법 및 시스템은 게놈의 선택된 영역의 서열의 결정을 제공함으로써 표적화된 게놈 서열분석을 완수하고, 상기 서열분석 정보는 일반적으로 극도로 낮은 서열분석 오차율 및 짧은 판독 서열분석 기술의 고 처리율의 이점을 갖는 방법을 이용하여 수득된다. 이미 기재된 바와 같이, 본원에 기재된 방법 및 시스템의 이점은, 그들이 보편적으로 이용가능한 짧은 판독 서열분석 기술의 사용을 통해 목적 결과를 달성할 수 있다는 것이다. 이러한 기술은 충분히 특성규명되고 매우 효과적인 프로토콜 및 시약 시스템과 함께 쉽게 이용가능하고 연구 단체에 널리 분산되는 이점을 갖는다. 이들 짧은 판독 서열분석 기술은, 예를 들면, Illumina, inc. (GAIIx, NextSeq, MiSeq, HiSeq, X10), Ion Torrent division of Thermo-Fisher (Ion Proton and Ion PGM), 파이로서열분석 방법, 뿐만 아니라 기타로부터 이용가능한 것을 포함한다.

[0096] 특별한 장점은, 본원에 기재된 방법 및 시스템이 이러한 짧은 판독 서열분석 기술을 이용하고, 이들의 관련된 낮은 오차율로 그렇게 수행한다는 점이다. 특히, 본원에 기재된 방법 및 시스템은 상기한 바와 같이, 그러나 짝짓기 쌍(mate pair) 확장을 제외하고, 1000 bp 미만, 500 bp 미만, 300 bp 미만, 200 bp 미만, 150 bp 미만 또는 훨씬 짧은 개별 목적 분자 판독 길이; 및 상기 개별 분자 판독 길이에 대해 5% 미만, 1% 미만, 0.5% 미만, 0.1% 미만, 0.05% 미만, 0.01% 미만, 0.005% 미만, 또는 심지어 0.001% 미만의 서열분석 오차율을 사용하여 개별 분자 판독 길이 또는 콘텍스트를 달성한다.

[0097] 본 출원에 기술된 방법 및 시스템에 따라 핵산을 가공하고 서열분석하는 방법은 또한 추가로 상세히 하기에 기술된다: 모든 목적 및 특히 가공 핵산 그리고 게놈 재료의 서열분석 및 다른 특성규명에 관한 모든 쓰여진 설명, 도면 및 실시예로 그 전체가 참고로 본원에서 편입되는 USSN 14/316,383; 14/316,398; 14/316,416; 14/316,431; 14/316,447; 및 14/316,463.

[0098] 사용된 서열분석 플랫폼에 관계 없이, 일반적으로, 그리고 본원에 기술된 방법 중 임의의 것에 따라, 핵산의 서열분석은 서열 판독의 분자 콘텍스트 또는 서열 판독의 부분을 보존하는 방식으로 전형적으로 수행된다. 상기로 다중 서열 판독 또는 서열 판독의 다중 부분이 핵산의 단일 유래 분자에 부여될 수 있다는 의미이다. '에 부여하는'으로 서열 판독이 핵산의 이의 특정한 유래 분자의 염기의 선형 서열 내에 해당하는 경우 식별될 수 있다는 의미이고 - 환언하면, 도 7을 참고하여, 단편 703, 704, 705 및 706이 본래 핵산 분자 701 및 702로부터 생성되면, 서열분석은 단편 703, 704, 705 및 706 유래의 서열 판독이 이의 분자 콘텍스트를 유지하는 방식으로 수행되고, 그리고 단편 703 및 704가 본래 분자 701로부터 유도되고, 단편 705 및 706이 본래 분자 702로부터 유도된다 (심지어 전체 단편이 서열분석 반응에 대해 함께 풀링된다고 하더라도)는 것이 용이하게 확인된다. 또한, 서열분석은 유래 분자가 식별될 뿐만 아니라, 그 선형 분자를 따라 각 단편의 상대 위치인 정도로 일반적으로 수행되고 - 예를 들면, 단편 703이 유래 핵산 701의 선형 서열을 따라 단편 704로부터 "업스트림"인 것이 결정될 수 있다. 일반적으로, 분자 콘텍스트는 임의의 식별자 또는 하나 이상의 단편을 다른 단편과의 임의의 다른 구별 방법의 사용을 통해 유지된다. 일반적으로, 상기 식별자는 그룹 또는 개별 단위체로 분리된 단편상에서 사용된다. 일부 예에서, 인정될 바와 같이, 분자의 임의의 다른 분리 방법이 사용될 수 있어도, 상기 분리는 개별 파티션으로의 분리이다. 또 다른 추가 예에서, 사용된 식별자는 바코드이고, 선형 위치는 바코딩 뿐만 아니라 중첩 단편으로부터 서열 판독의 알고리즘성 조합 모두를 통해 계측된다. 명료함을 위해 본원에서 논의의 대부분이 파티션으로의 분리 및/또는 바코딩에 관한 것이어도, 본래 핵산 분자의 임의의 분리 방법 및 단편의 임의의 식별 또는 달리 부여화 방법이 본원에서 기재된 방법 및 시스템에서 사용중임이 인정될 것이다.

[0099] 숙지된 바와 같이, 핵산의 단일 본래 분자가 임의의 다양한 길이일 수 있는 반면, 바람직한 양태에서, 긴 범위 분자 콘텍스트의 보존을 허용하는 상대적으로 긴 분자일 것이다. 특히, 단일 유래 분자는 바람직하게는 전형적인 짧은 판독 서열 길이보다 실질적으로 더 길고, 예를 들면, 200 염기보다 더 길고, 종종 적어도 1000 염기 이상, 5000 염기 이상, 10,000 염기 이상, 20,000 염기 이상, 30,000 염기 이상, 40,000 염기 이상, 50,000 염기 이상, 60,000 염기 이상, 70,000 염기 이상, 80,000 염기 이상, 90,000 염기 이상, 또는 100,000 염기 이상이고, 일부 경우에서 1 메가염기 이상이다.

[0100] 특정 상황에서, 본원에서 기재된 서열분석 방법은 게놈의 더 긴 범위를 거쳐 더 낮은 수준 연결된 판독과 선택된 영역의 깊은 커버리지의 조합을 포함한다. 인정될 바와 같이, 드노보와 재서열분석의 상기 조합이 전체 게놈 및/또는 게놈의 큰 부분을 서열분석하기 위한 효율적인 방식을 제공한다. 본원에서 기재된 선택적 증폭 방법을 통해 저조하게 특성규명된 및/또는 고도로 다형성된 영역의 표적화된 커버리지는 드노보 서열 조립체에 필요한 핵산 물질의 양을 제공하고, 반면에 게놈의 다른 영역에 대해 연결된 게놈 서열분석은 게놈의 나머지의 고 처리율 서열분석을 유지한다. 동일한 서열분석 플랫폼이 커버리지의 양쪽 유형에 대하여 사용될 수 있기 때문에, 본원에서 기재된 방법 및 조성물은 드노보와 연결된 판독 서열분석의 이러한 조합 허용을 본질적으로 잘 받아들인다. 본원에서 기재된 방법에 따라서 서열분석되는 핵산 및/또는 핵산 단편의 집단은 드노보 서열분석용 게놈 영역 및 재서열분석용 게놈 영역 둘 모두로부터 서열을 함유하고 - 드노보 서열분석용 관심 영역을 커버하는 핵산의 비율은 본원에서 추가로 상세히 기재된 표적화된 증폭 방법 때문에 게놈의 다른 영역을 커버하는 핵산보다 더 높다.

[0101] 일반적으로, 도 1에 나타낸 바와 같이, 본원에서 기재된 방법 및 시스템은, 분자 콘텍스트를 유지하는 한편, 핵산, 특히 게놈의 선택된 영역으로부터 핵산을 특성규명하기 위해 사용될 수 있다. 나타낸 바와 같이, 2개의 개별 개별 핵산 102 및 104는 예시되고, 각각은 수많은 관심 영역, 예를 들면, 핵산 102에서 영역 106 및 108, 및 핵산 104에서 영역 110 및 112를 갖는다. 각 핵산내 관심 영역은 동일한 핵산 분자 내에서 연결되지만 (예를 들면, 이로부터 유래하지만), 일부 경우에서 이들 영역은 서로, 예를 들면, 1 kb 초과 떨어져서, 5 kb 초과 떨어져서, 10 kb 초과 떨어져서, 20 kb 초과 떨어져서, 30 kb 초과 떨어져서, 40 kb 초과 떨어져서, 50 kb 초과 떨어져서, 및 일부 경우에서, 100 kb 만큼 많이 떨어져서 상대적으로 분리될 수 있다. 관심 영역은 일반적으로 게놈의 개별 및 분리된 부분이고 - 일부 경우에서, 상기 영역은 저조하게 특성규명된 영역이다. 관심 영역은 또한 개별 유전자, 유전자 그룹, 엑손을 표시할 수 있다. 나타낸 바와 같이, 각 핵산 102 및 104는 각각 이의 고유 파티션 114 및 116으로 분리된다. 본원에 언급된 바와 같이, 이러한 파티션은, 많은 경우에서, 유중수 에멀젼에서 수성 액적이다. 각 액적 내에서, 각 단편의 부분은, *예를 들면*, 동일한 분자로부터 유래된 경우, 상기 단편의 본래 분자 콘텍스트를 보존하는 방식으로 복제된다. 나타낸 바와 같이, 이것이 달성되는 한 방법은 유래 단편이 파티셔닝된 액적을 대표하는, 바코드 서열, *예를 들면*, 예시된 대로 바코드 서열 "1" 또는 "2"의 각 복제된 단편에서의 포함을 통한 것이다. 전체의 게놈 서열분석 적용에 대하여, 각각의 유래 핵산 102 및 104로부터 전체 범위 서열 정보를 서열분석 및 재조립하기 위해, 모든 복제된 단편 및 이의 관련된 바코드를 간단히 풀링

할 수 있다. 그러나, 많은 경우에서, 게놈의 과학적으로 관련된 부분에 더 큰 집중을 제공하기 위해, 그리고 게놈의 덜 관련된 또는 무관한 부분에 관한 서열분석 수행의 시간 및 비용 최소화하기 위해, 전체 게놈의 과학적 표적화된 부분을 단지 분석하는 것이 더욱 바람직하다.

[0102] 상기에 따라서, 바코딩 단계에 더하여, 선택적 증폭의 하나 이상의 단계가 있을 수 있고, 이로써 핵산 102 또는 104가 선택된 관심 게놈 영역을 함유하면, 상기 영역으로부터 증폭산물은 각각의 파티션 114 및 116에서 단편의 더 큰 백분율을 형성할 것이다. 비록 일부 구현예에서 증폭 단계가 또한 바코드의 부착에 이어서 발생할 수 있어도, 상기 증폭 단계는 본원에서 기재된 방법에 따라서 바코드의 부착 이전에 또는 이와 동시에 일반적으로 발생할 것이다.

[0103] 라이브러리 118 내에 풀링된 단편이, 예를 들면, 바코드 정보의 체류를 통해, 이의 본래 분자 콘텍스트를 유지하기 때문에, 이들은 포매된 (때때로, 긴 범위) 연결 정보, 예를 들면, 각각의 조립된 관심 영역 106:108과 110:112 사이로서 추론된 연결로 이의 본래 분자 콘텍스트으로 재조립될 수 있다. 예로써, 게놈의 2개의 이질적인 표적화된 부분, 예를 들면, 2 이상의 엑손 사이 직접적인 분자 연결을 식별할 수 있고, 그 직접적인 분자 연결은 구조적 변동 및 다른 게놈 특징을 식별하기 위해 사용될 수 있다. 선택적 증폭이 게놈의 선택된 영역의 부분을 함유한 핵산 단편의 양을 증가시키기 위해 이용되는 상황에 대하여, 그 다음 분자 콘텍스트를 식별하기 위한 능력은 또한, 종종 상기 영역의 드노보 조합을 허용하는 깊이에서, 게놈의 상기 선택된 영역을 서열분석하기 위한 방식을 제공한다.

[0104] 일반적으로, 본 발명의 방법은, 본원에서 추가로 상세히 논의된 본 발명의 방법의 도식적 개요를 제공하는, 도 2에서 예시된 바와 같은 단계를 포함한다. 인정될 바와 같이, 도 2에서 개괄된 방법은 필요에 따라 및 본원에서 기재된 바와 같이 변경 또는 변형될 수 있는 예시적인 구현예이다.

[0105] 도 2에 나타낸 바와 같이, 본원에서 기재된 방법은 대부분의 예에서 표적화된 관심 영역을 함유한 샘플 핵산이, 분리되는 (201) 것을 포함할 것이다. 일반적으로, 관심 게놈 영역으로부터 핵산을 함유한 각 파티션은 큰 비율이 선택된 게놈 영역(202)으로부터 서열을 함유할 단편의 집단을 생산하기 위해 표적화된 농축을 경험할 것이다. 일반적으로, 이들이 함유되는 파티션에 특이적인 단편의 바코딩에 의해, 상기 단편은 그 다음 단편(203)의 본래 분자 콘텍스트를 보존하기 위한 방식으로 추가로 단편화 또는 복제된다. 각 파티션은 일부 예에서 1 초과 핵산을 포함할 수 있고, 일부 사례에서 수백개의 핵산 분자를 함유할 것이고 - 다중 핵산이 파티션 내에 있는 상황에서, 게놈의 임의의 특정한 유전자좌는 일반적으로 바코딩 이전에 단일 개별 핵산에 의해 나타낼 것이다. 단계 203의 바코딩된 단편은 당해 기술에 공지된 임의의 방법을 이용하여 생성될 수 있고 - 일부 예에서, 올리고뉴클레오티드는 상이한 파티션 내 샘플이다. 상기 올리고뉴클레오티드는 샘플의 다수의 상이한 영역을 무작위로 프라이밍하도록 의도된 랜덤 서열일 수 있거나, 이들은 샘플의 표적화 영역의 업스트림을 프라이밍하도록 표적화된 특이적 프라이머 서열을 포함할 수 있다. 추가 예에서, 이들 올리고뉴클레오티드는 또한 바코드 서열을 함유하고, 이로써 복제 공정은 또한 본래 샘플 핵산의 수득한 복제 단편을 바코딩한다. 샘플의 증폭 및 바코딩에서 이들 바코드 올리고뉴클레오티드의 특히 품격있는 사용 공정은 모든 목적 및 특히 가공 핵산 그리고 게놈 재료의 서열분석 및 다른 특성규명에 관한 모든 쓰여진 설명, 도면 및 실시예로 그 전체가 참고로 본원에서 편입되는 미국 특허 출원 번호 14/316,383; 14/316,398; 14/316,416; 14/316,431; 14/316,447; 및 14/316,463에서 상세히 기재된다. 또한 연장 반응 시약, 예를 들면, DNA 폴리머라아제, 뉴클레오시드 트리포스페이트, 파티션 내 또한 함유된 동시-인자(예를 들면, $Mg^{2+}$ 또는 $Mn^{2+}$ 등)는 주형으로서 샘플을 사용하여 프라이머 서열을 연장시킴으로써 프라이머가 어닐링되는 주형의 가닥에 대한 상보적 단편을 생산하며, 당해 상보적 서열은 올리고뉴클레오티드 및 이의 연관된 바코드 서열을 포함한다. 샘플의 상이한 부분으로의 다중 프라이머의 어닐링 및 확대는 샘플의 중첩성 상보적 단편의 거대한 풀을 초래할 수 있고, 각각은 그것이 생성된 파티션을 나타내는 자체 바코드 서열을 포함한다. 일부 경우에, 이러한 상보적 단편은 자체가 다시 바코드 서열을 포함하는 보체의 보체를 생성하기 위해 파티션에 존재하는 올리고뉴클레오티드에 의해 프라이밍되는 주형으로서 사용될 수 있다. 추가 예시에서, 이 복제 공정은 제1 보체가 복제되는 경우, 그것이 이의 말단 또는 부근에 두 개의 상보적 서열을 생성하여 반복 사본을 추가로 생성하기 위한 기초가 되는 분자의 능력을 감소시키는 헤어핀 구조 또는 부분적 헤어핀 구조의 형성을 가능하게 하도록 구성된다.

[0106] 도 2에서 예시된 방법으로 돌아가서, 파티션-특이적 바코드가 복제된 단편에 부착되면, 바코딩된 단편은 그 다음 풀링된다(204). 풀링된 단편은 그 다음 서열분석되고(205) 단편의 서열은 이의 유래 분자 콘텍스트(206)에 부여하여, 이로써 표적화된 관심 영역은 모두 그 유래 분자 콘텍스트으로 식별되고 또한 연결된다. 본원에서 기재된 방법 및 시스템의 이점은 표적화된 게놈 영역에 대하여 단편의 농축 이전에 복제된 단편의 파티션- 또는

샘플-특이적 바코드의 부착이 그들 표적화된 영역의 본래 분자 콘텍스트를 보존하여, 그들의 본래 파티션 및 따라서 그들의 유래 샘플 핵산에 부여되도록 허용한다는 점이다.

[0107] 상기 작업흐름에 더하여, 표적화된 게놈 영역은, 칩-기반 및 용액-기반 포획 방법 둘 모두를 포함하는 방법을 이용하여, 추가 분석, 특히 서열분석을 위하여 추가로 농축, 단리 또는 분리, 즉, "붕괴(pulled down)"될 수 있다. 상기 방법은 관심 게놈 영역 또는 관심 게놈 영역 근처 또는 인접한 영역에 상보적인 프로브를 이용한다. 예를 들어, 혼성(또는 칩-기반)포획에서, 함께 합쳐져서 관심 영역을 커버하는 서열을 가진 포획 프로브(보통 단일가닥 올리고뉴클레오티드)를 함유한 마이크로어레이는 표면에 고정된다. 게놈 DNA는 단편화되고 추가로 가공 예컨대 무딘 말단을 생산하기 위한 말단-치유 및/또는 추가의 특징 예컨대 보편적 프라이밍 서열의 부가를 경험할 수 있다. 이들 단편은 마이크로어레이 상에서 프로브에 하이브리드화된다. 미하이브리드화된 단편은 세정제거되고 요망된 단편은 서열분석 또는 다른 분석을 위하여 표면상에서 용출되거나 달리 가공되고, 따라서 표면상에 남아있는 단편의 집단은 표적화된 관심 영역(예를 들면, 포획 프로브에 함유된 것들에 상보적인 서열을 포함한 영역)을 함유한 단편에 대해 농축된다. 단편의 농축된 집단은 추가로 당해 기술에 공지된 임의의 증폭 기술을 이용하여 증폭될 수 있다. 상기 표적화된 붕괴 농축 방법에 대하여 예시적인 방법은, 모든 쓰여진 설명, 도면 및 실시예를 포함하여, 모든 목적 및 특히 표적화된 붕괴 농축 방법 및 서열분석 방법에 관련된 모든 교시로 그 전체가 참고로 본원에서 편입되는, 2015년 10월 29일 출원된, USSN 14/927,297에 기재된다.

[0108] 일부 양태에서, 게놈의 선택된 영역의 커버리지용 방법은 선택된 영역으로부터 핵산 분자 및/또는 이의 단편을 함유한 개별 파티션이 추가 가공을 위하여 자체 분류되는 방법을 포함한다. 인정될 바와 같이, 개별 파티션의 상기 분류는 본원에서 기재된 관심 게놈 영역의 선택적 증폭 및/또는 표적화된 붕괴의 다른 방법과 임의의 조합으로, 특히 상기 기재된 작업 흐름의 단계와 임의의 조합으로 발생할 수 있다.

[0109] 일반적으로, 개별 파티션의 상기 분류 방법은 게놈의 하나 이상의 선택된 부분의 적어도 일 부분을 함유한 파티션이 게놈의 상기 부분으로부터 임의의 서열을 함유하지 않은 파티션으로부터 분리되는 단계를 포함한다. 이러한 방법은 게놈의 부분 유래의 서열을 함유하는 개별 파티션 내에 게놈의 하나 이상의 선택된 부분 중 적어도 일 부분을 포함하는 단편의 서열에 대해 증폭된 집단을 제공하는 단계들을 포함한다. 상기 농축은 집단을 생산하기 위해 게놈의 하나 이상의 선택된 부분의 적어도 일 부분을 포함하는 개별 파티션 내에서 단편의 지향된 PCR 증폭의 사용을 통해 일반적으로 달성된다. 상기 지향된 PCR 증폭은 따라서 게놈의 하나 이상의 선택된 부분의 적어도 일 부분을 포함한 증폭산물을 생산한다. 특정 구현예에서, 이들 증폭산물은, 일부 비제한적인 구현예에서 형광 분자를 포함할 수 있는, 검출가능한 표지에 부착된다. 일반적으로, 상기 부착이 발생하여 이로써 게놈의 하나 이상의 선택된 부분을 함유한 단편으로부터 생성된 상기 증폭산물이 검출가능한 표지에 부착된다. 일부 구현예에서, 검출가능한 표지의 부착은 게놈의 하나 이상의 선택된 부분의 선택적 증폭 동안 발생한다. 상기 검출가능한 표지는 추가 구현예에서 비제한적으로 형광 표지, 전기화학 표지, 자기 비드, 및 나노입자를 포함할 수 있다. 검출가능한 표지의 부착은 본 분야에 공지된 방법들을 사용하여 달성될 수 있다. 또 다른 추가의 구현예에서, 상기 게놈의 상기 하나 이상의 선택된 부분의 적어도 일 부분을 포함하는 단편을 함유한 개별 파티션은 상기 파티션 내 증폭산물에 부착된 검출가능한 표지로부터 방출된 신호 상에 기반하여 분류된다.

[0110] 추가 구현예에서, 상기 서열을 함유하지 않는 것들로부터 게놈의 선택된 부분을 함유하는 개별 파티션을 분류하는 단계는 하기 단계를 포함한다: (a) 개시 게놈 재료를 제공하는 단계; (b) 개별 핵산 분자를 상기 개시 게놈 재료로부터 개별 파티션으로 분배하여, 이로써 각 개별 파티션이 제1 개별 핵산 분자를 함유하도록 하는 단계; (c) 게놈의 하나 이상의 선택된 부분 중 적어도 일 부분을 포함하는 단편의 서열에 대해 농축된 개별 파티션 중 적어도 일부 내의 집단을 제공하는 단계; (d) 각 개별 파티션 내 상기 단편에 공통 바코드 서열을 부착시켜, 이로써 상기 단편 각각이 이것이 함유된 상기 개별 파티션에 부여가능하도록 하는 단계; (e) 상기 게놈의 하나 이상의 선택된 부분을 포함하는 단편을 함유하지 않는 개별 파티션으로부터, 상기 게놈의 하나 이상의 선택된 부분의 적어도 일 부분을 포함하는 단편을 함유하는 개별 파티션을 분리하는 단계; (f) 상기 게놈의 하나 이상의 선택된 부분 중 적어도 일 부분을 포함하는 상기 단편 유래의 서열 정보를 수득하여, 이로써 분자 콘텍스트를 유지하는 한편 상기 게놈 샘플 중 하나 이상의 표적화된 부분을 서열분석하는 단계.

[0111] 추가 구현예에서 그리고 상기 중 임의의 것에 따라, 단편 유래의 서열 정보의 수득 이전에, 개별 파티션은 조합되고 함께 풀링된다. 추가 구현예에서, 상기 단편 유래의 서열 정보를 수득하는 상기 단계는 상기 단편의 상기 서열의 상기 분자 콘텍스트를 보존하여 이로써 상기 식별이 상기 동일한 제1 개별 핵산 분자로부터 유도된 단편을 식별하는 방식으로 수행된다. 또 다른 추가 구현예에서, 서열 정보의 상기 수득은 하기로 이루어진 군으로부터 선택된 서열분석 반응을 포함한다: 짧은 판독-길이 서열분석 반응 및 긴 판독-길이 서열분석 반응. 또 다른

추가 구현예에서, 서열분석 반응은 짧은 판독, 고 정확도 서열분석 반응이다.

[0112] 또 다른 추가 구현예에서, 그리고 상기 중 임의의 것에 따라, 개별 파티션은 에멀젼 내 액적을 포함한다. 추가 구현예에서, 개별 파티션 내에서 바코딩된 단편은 게놈의 하나 이상의 선택된 부분의 약 1X-10X 커버리지를 나타낸다. 또 다른 추가 구현예에서, 개별 파티션 내에서 바코딩된 단편은 게놈의 하나 이상의 선택된 부분의 약 2X-5X 커버리지를 나타낸다. 또 다른 추가의 추가 구현예에서, 개별 파티션 내 증폭산물의 바코딩된 단편은 게놈의 하나 이상의 선택된 부분의 적어도 1X 커버리지를 나타낸다. 또 다른 추가 구현예에서, 개별 파티션 내에서 바코딩된 단편은 게놈의 하나 이상의 선택된 부분의 적어도 2X 또는 5X 커버리지를 나타낸다.

[0113] 게놈의 선택된 영역 유래의 서열 정보를 수득하기 위한 능력 제공에 더하여, 본원에서 기재된 방법 및 시스템은, 모든 목적 및 특히 게놈 재료의 특성규명에 관한 모든 쓰여진 설명, 도면 및 실시예로 그 전체가 참고로 본원에서 편입되는, 미국 특허 출원 번호 14/752,589 및 14/752,602에 기재된 바와 같이, 비제한적으로 일배체형 단계화, 구조적 변동의 식별, 및 복제수 변동의 식별을 포함한, 게놈 재료의 다른 특성규명을 또한 제공할 수 있다.

[0114] 상기 언급된 바와 같이, 본원에 기재된 방법 및 시스템은 보다 긴 핵산의 짧은 서열 판독을 위한 개별적 분자 콘텍스트를 제공한다. 본원에 사용된 바와 같이, 개별적 분자 콘텍스트는, 예를 들면, 서열 판독 자체 내에 포함되지 않는인접 또는 근위 서열에 관련되는 특이적 서열 판독을 초과하는 서열 콘텍스트를 의미하고, 그 자체로, 전형적으로, 그들이 짧은 서열 판독, 예를 들면, 쌍형성된 판독에 대해 약 150 염기 또는 약 300 염기의 판독에 전체로 또는 부분적으로 포함되지 않을 정도일 것이다. 특히 바람직한 양태에서, 방법 및 시스템은 짧은 서열 판독을 위한 긴 범위 서열 콘텍스트를 제공한다. 이러한 긴 범위 콘텍스트는 소정의 서열 판독과 서로 1 kb 초과, 5 kb 초과, 10 kb 초과, 15 kb 초과, 20 kb 초과, 30 kb 초과, 40 kb 초과, 50 kb 초과, 60 kb 초과, 70 kb 초과, 80 kb 초과, 90 kb 초과 또는 심지어 100 kb 초과 또는 그 이상의 거리 내에 있는 서열 판독의 관계 또는 연결을 포함한다. 인정될 바와 같이, 긴 범위 개별 분자 콘텍스트를 제공함으로써, 또한 그 개별 분자 콘텍스트 내에서 변이체의 단계화 정보를 유도할 수 있고, 예를 들면, 특정한 긴 분자상의 변이체는, 통상적으로 단계화된 정의에 의해 될 것이다.

[0115] 더 긴 범위 개별 분자 콘텍스트를 제공함으로써, 본 발명의 방법 및 시스템은 또한 훨씬 더 긴 추론된 분자 콘텍스트(또한 본원에서 일명 "긴 가상 단일 분자 판독")을 제공한다. 본원에서 기재된 바와 같이, 서열 환경은 전체 게놈 서열의 상이한 (일반적으로 킬로베이스 규모상의) 범위를 거쳐 단편의 연결 제공 또는 맵핑을 포함할 수 있다. 이러한 방법은, 짧은 서열 판독의 개별적인 더 긴 분자 또는 연결된 분자의 콘티그으로의 매핑 뿐만 아니라, 예를 들면,개별 분자의 인접 결정 서열을 갖는 보다 긴 개별 분자의 많은 부분의 긴 범위 서열분석을 포함할 수 있고, 여기서, 상기 결정 서열은 1 kb 초과, 5 kb 초과, 10 kb 초과, 15 kb 초과, 20 kb 초과, 30 kb 초과, 40 kb 초과, 50 kb 초과, 60 kb 초과, 70 kb 초과, 80 kb 초과, 90 kb 초과, 또는 심지어 100 kb 초과이다. 서열 콘텍스트의 경우와 같이, 짧은 서열의 긴 핵산, 예를 들면, 개별적인 긴 핵산 분자 또는 연결된 핵산 분자 또는 콘티그의 수집 둘 모두로의 부여는 고 수준의 서열 콘텍스트를 제공하기 위해 긴 핵산 연신에 대한 짧은 서열의 매핑 뿐만 아니라 이러한 긴 핵산을 통해 짧은 서열로부터 조립된 서열을 제공함을 모두 포함할 수 있다.

[0116] 또한, 긴 개별 분자와 관련된 긴 범위 서열 콘텍스트를 사용할 수 있는 반면, 이러한 긴 범위 서열 콘텍스트를 갖는 것은 또한 훨씬 더 긴 범위 서열 콘텍스트를 추정할 수 있도록 한다. 일례로서, 상기한 긴 범위 분자 콘텍스트를 제공함으로써, 상이한 유래 분자로부터의 긴 서열 중에서 중첩 변이체 부분, 예를 들면, 위상 변이체, 전좌된 서열 등을 식별하여 이들 분자 사이의 추론된 연결을 가능하게 할 수 있다. 상기 추론된 연결기 또는 분자 콘텍스트는 본 명세서에서 일명 "추론된 콘티그"이다. 위상 서열의 콘텍스트에서 논의된 일부 경우에, 추론된 콘티그는 일반적으로, 예를 들면, 중첩 위상 변이체에 의해, 개별 유래 분자보다 실질적으로 더 긴 길이의 위상 콘티그를 추정할 수 있는 위상 서열을 나타낼 수 있다. 이러한 위상 콘티그는 본원에서 “단계 블록”으로서 칭명된다.

[0117] 보다 긴 단일 분자 판독(예를 들어, 상기 검토된 “보다 긴 실제적 단일 분자 판독”)로 개시함으로써, 다르게 짧은 판독 서열분석 기술 또는 위상 서열분석에의 다른 접근법을 사용하여 달성가능한 것보다 긴 추론된 콘티그 또는 단계 블록을 유도할 수 있다. 예를 들어, 예컨대 문헌[미국 출원번호 제2013-0157870호]를 참조한다. 특히, 본원에 기술된 방법 및 시스템을 사용하여, 적어도 약 10kb, 적어도 약 20kb, 적어도 약 50kb의 N50(기술된 N50 수보다 큰 블록 길이의 총합은 전체 블록 길이의 총합의 50%임)을 갖는 추론된 콘티그를 수득할 수 있다. 바람직한 양태에서, 적어도 약 100kb, 적어도 약 150kb, 적어도 약 200kb, 및 많은 경우에, 적어도 약

250kb, 적어도 약 300 kb, 적어도 약 350 kb, 적어도 약 400 kb, 및 일부 경우에, 적어도 약 500 kb 이상의 N50을 갖는 추론된 콘티그 또는 상 블록 길이가 획득된다. 추가의 기타 경우에서, 200 kb 초과, 300 kb 초과, 400 kb 초과, 500 kb 초과, 1 Mb 초과, 또는 심지어 2 Mb 초과의 최대 상 블록 길이가 수득될 수 있다.

[0118] 일 양태에서, 본원에 상기에, 그리고 하기에 기술된 포획 방법 중 임의의 것과 결합하여, 본원에 기재된 방법 및 시스템은 샘플 핵산 또는 이의 단편의 개별 분획 또는 파티션(본원에서 파티션으로 상호교환적으로 칭명됨)으로의 파티셔닝, 증착 또는 파티션을 제공하고, 여기서, 각 파티션은 다른 파티션의 내용물로부터 이의 내용물의 분리를 유지시킨다. 고유 식별자, 예를 들면, 바코드는 이전에, 이후에 또는 동시에 특별한 파티션 내에 포함된 샘플 핵산, 특히 본래 파티션에 증착될 수 있는 연속 샘플 핵산의 비교적 긴 연신부에 대한 특성, 예를 들면, 핵산 서열 정보의 부여를 허용하기 위해 파티셔닝되거나 파티셔닝된 샘플 핵산을 유지하는 파티션으로 전달될 수 있다.

[0119] 본원에 기술된 방법 내 이용된 샘플 핵산은 전형적으로, 예를 들면, 전체 염색체, 엑솜, 또는 다른 대형 게놈 부분과 같은, 분석될 전체 샘플 핵산의 다수의 중첩 부분을 나타낼 수 있다. 이러한 샘플 핵산은 전체 게놈, 개별 염색체, 엑솜, 증폭산물, 또는 목적하는 임의의 다양한 상이한 핵산을 포함할 수 있다. 샘플 핵산은 전형적으로 핵산이 파티션에서 인접 핵산 분자의 비교적 긴 단편 또는 연신부로 존재하도록 파티셔닝된다. 전형적으로, 샘플 핵산의 이러한 단편은 1 kb 초과, 5 kb 초과, 10 kb 초과, 15 kb 초과, 20 kb 초과, 30 kb 초과, 40 kb 초과, 50 kb 초과, 60 kb 초과, 70 kb 초과, 80 kb 초과, 90 kb 초과 또는 심지어 100 kb 초과일 수 있고, 이는 상기 기술된 보다 긴 범위의 분자 콘텍스트를 허용한다.

[0120] 샘플 핵산은 또한 전형적으로 소정의 파티션이 개시 샘플 핵산의 두 개의 중첩 단편을 포함할 매우 낮은 가능성을 갖는 수준으로 파티셔닝된다. 이는 전형적으로 샘플 핵산을 파티셔닝 공정 동안 낮은 유입량 및/또는 농도로 제공함으로써 달성된다. 그 결과, 바람직한 경우에, 소정의 파티션은 개시 샘플 핵산의 다수의 길지만 비중첩성 단편을 포함할 수 있다. 이어서, 다른 파티션 중의 샘플 핵산은 고유 식별자와 회합되고, 여기서, 임의의 소정의 파티션에 대하여, 내부에 함유된 핵산은 동일한 고유 식별자를 포함하지만, 상이한 파티션은 상이한 고유 식별자를 포함할 수 있다. 더욱이, 파티셔닝 단계는 샘플 성분을 극소 용적의 파티션 또는 액적으로 할당하므로, 위에서 설정된 바와 같은 바람직한 할당을 달성하기 위하여, 보다 큰 용적의 공정, 예를 들면, 튜브, 또는 다중 웰 플레이트의 웰 속에 요구될 수 있는 바와 같은, 샘플의 실질적인 희석을 수행할 필요가 없다. 또한, 본원에 기재된 시스템이 이러한 높은 수준의 바코드 다양성을 사용하기 때문에, 상기 제공된 바와 같이, 높은 수의 게놈 등가물 중에서 다양한 바코드를 할당할 수 있다. 특히, 상기한 다중웰 플레이트(참조: 예를 들면, 미국 출원 공보 번호 2013-0079231 및 2013-0157870, 이의 전체 개시 내용은 그 전체가 본원에 참조로 인용됨)는 전형적으로 단지 백개 내지 수 백개의 상이한 바코드 서열로 작동하고, 바코드를 상이한 세포/핵산으로 부여시킬 수 있도록 하기 위해 이들 샘플의 제한적인 희석 공정을 채택한다. 이와 같이, 그들은 일반적으로 100개보다 훨씬 적은 세포로 작동하고, 이는 1:10 정도의 게놈:(바코드 형태)의 비, 확실히 충분히 1:100 이상의 비를 전형적으로 제공할 것이다. 본원에 기재된 시스템은, 한편으로, 높은 수준의 바코드 다양성, 예를 들면, 10,000, 100,000, 500,000, 600,000, 700,000개 등을 초과하는 다양한 바코드 형태가 1:50 이하, 1:100 이하, 1:1000 이하 정도의 게놈:(바코드 형태) 비 또는 훨씬 더 작은 비에서 작동할 수 있는 반면, 또한 게놈당 훨씬 개선된 바코드 다양성을 제공하면서 높은 수의 게놈(예: 검정당 100개 초과 게놈, 검정당 500개 초과 게놈, 검정당 1000개 게놈 또는 그 이상의 정도)의 부하를 가능하게 한다.

[0121] 흔히, 샘플은 파티셔닝 단계 전에 비드에 박리가능하게 부착된 올리고뉴클레오티드 태그의 세트와 조합된다. 일부 예에서, 증폭 방법은, 일부 예에서 이들이 유도된 전체 본래 핵산 분자의 더 작은 분절(단편)을 함유하는, 수득된 증폭 생성물에 바코드를 부가하기 위해 사용된다. 일부 예에서, 트랜스포존을 이용한 방법은, 모든 목적 및 특히 핵산에 바코드 또는 다른 올리고뉴클레오티드 태그의 부착에 관련된 모든 교시로 그 전체가 참고로 본원에서 편입되는, Amini et al, Nature Genetics 46: 1343-1349 (2014) (advance online publication on October 29, 2014)에서 기재된 바와 같이 이용된다. 추가 예에서, 바코드의 부착 방법은 이중가닥 샘플 핵산을 따라 갭을 생산하기 위해 효소 또는 폴리머라아제 및/또는 침습성 프로브 예컨대 recA의 사용을 포함할 수 있고 - 바코드는 그 다음 상기 갭 속에 삽입될 수 있다.

[0122] 증폭이 핵산 단편을 태깅하기 위해 사용되는 예에서, 올리고뉴클레오티드 태그는 적어도 제1 및 제2 영역을 포함할 수 있다. 제1 영역은, 소정의 파티션 내의 올리고뉴클레오티드 사이에서와 같이, 실질적으로 동일한 바코드 서열일 수 있지만, 상이한 파티션 사이에서와 같이, 상이한 바코드 서열일 수 있고, 대부분의 경우, 상이한 바코드 서열인 바코드 영역일 수 있다. 제2 영역은 파티션 내의 샘플 내 핵산을 프라이밍하는데 사용될 수 있는 N-량체(특정 서열을 표적화하도록 설계된 랜덤 N-량체 또는 N-량체)일 수 있다. 일부 경우에, N-량체가 특정 서

열을 표적화하도록 설계된 경우, 이는 특정 염색체(예: 염색체 1, 13, 18 또는 21), 또는 염색체의 영역, 예를 들면, 엑솜 또는 다른 표적화 영역을 표적화하도록 설계될 수 있다. 상기 검토된 바와 같이, N-량체는 또한, 참조 서열 유래의 저조하게 특성규명되거나 고도로 다형성되거나, 또는 분지되는 경향이 있는 게놈의 선택 영역으로 설계될 수 있다. 일부 경우에, N-량체는 특정 유전자 또는 유전 영역, 예를 들면, 질환 또는 장애(예: 암)와 관련된유전자 또는 영역을 표적화하도록 설계될 수 있다. 파티션 내에서, 증폭 반응은 핵산의 길이에 따라 상이한 장소에서 핵산 샘플을 프라이밍하기 위해 제2 N-량체를 사용하여 수행될 수 있다. 증폭의 결과로서, 각 파티션은 동일하거나 거의 동일한 바코드에 부착되고, 각 파티션에서 핵산의 중첩성 작은 단편을 나타낼 수 있는 핵산의 증폭된 제품을 함유할 수 있다. 바-코드는 핵산의 세트가 동일 파티션으로부터 기원되고, 따라서 잠재적으로 또한 핵산의 동일한 가닥으로부터 기원된다는 것을 의미하는 마커로서 기능할 수 있다. 증폭 후, 핵산은 서열분석 알고리즘을 사용하여 풀링되고, 서열분석되고, 정렬될 수 있다. 짧은 서열 판독이 그들의 관련 바코드 서열에 의해 정렬되고, 샘플 핵산의 단일 긴 단편에 부여되기 때문에, 그 서열 상의 식별된 변이체 모두는 단일 기원 단편 및 단일 기원 염색체로 부여될 수 있다. 또한, 복수의 긴 단편에서 다수의 동일 위치 변이체를 정렬함으로써, 그 염색체 기여를 추가로 특성규명할 수 있다. 따라서, 특정한 유전적 변이체의 단계화에 관한 결론은 그 다음, 긴 범위의 게놈 서열 - 예를 들어, 게놈의 저조하게 특성규명된 영역의 연신부를 거쳐 서열 정보의 식별을 거쳐 분석할 수 있음에 따라, 도출될 수 있다. 이러한 정보는 또한 일반적으로 동일한 핵산 가닥 상 또는 상이한 핵산 가닥 상에 존재하는 유전적 변이체의 특정 세트인 일배체형을 식별하는데 유용할 수 있다. 사본수 변형도 또한 이러한 방식으로 식별될 수 있다.

[0123] 기재된 방법 및 시스템은 현재의 핵산 서열분석 기술 및 이들의 관련된 샘플 제조 방법에 유의미한 이점을 제공한다. 샘플 제조 및 서열분석 방법은 또한 주로 샘플 중의 대부분의 성분을 식별하고 특성규명하는 경향이 있고, 소수 성분, 예를 들면, 게놈 또는 물질의 저조하게 특성규명되거나 고도로 다형성된 영역 유래의 일 염색체에 의해, 또는 추출된 샘플에서 전체 DNA의 작은 비율을 구성하는 하나 또는 소수의 세포, 또는 혈류에서 순환하는 단편화된 종양 세포 DNA 분자에 의해, 기여된 유전 물질을 식별하고 특성규명되도록 설계되지 않는다. 본원에서 기재된 방법은 이들 소수 성분로부터 유전 물질을 증가시키는 선택적 증폭 방법을 포함하고, 상기 유전 물질의 분자 콘텍스트를 유지하는 능력은 추가로 이들 성분의 유전적 특성규명을 제공한다. 기재된 방법 및 시스템은 또한 더 큰 샘플 내에 존재하는 집단을 검출하는 상당한 이점을 제공한다. 이와 같이, 이들은 일배체형 및 사본 수 변동 평가에 특히 유용하고 - 본원에서 개시된 방법은 샘플 제조 동안 도입된 바이어스 때문에 핵산 표적의 집단에서 저조하게 특성규명되거나 저조하게 나타낸 게놈의 영역에 대해 서열 정보 제공에 또한 유용하다.

[0124] 본원에 개시된 바코딩 기술의 사용은 유전 마커의 소정의 세트에 개별 분자 콘텍스트를 제공하는, 즉(단일 마커와 대조적으로) 유전 마커의 소정의 세트를 개별 샘플 핵산 분자로 부여하는 고유한 능력을 부여하고, 변이체 배위된 조립체를 통해 및/또는 특정 염색체에 복수의 단일 핵산 분자 중에서 광범위하거나 심지어 더 긴 범위의 추론된 개별 분자 콘텍스트를 제공한다. 이러한 유전 마커는 특이적 유전자좌, 예를 들면, SNP와 같은 변이체를 포함할 수 있거나, 그들은 짧은 서열을 포함할 수 있다. 또한, 바코딩의 사용은, 예를 들면, 혈류 중 순환성 종양 DNA의 검출 및 특성규명을 위한 샘플로부터 추출된 전체 핵산 집단의 소수 성분 및 주요 성분을 구별하는 능력을 촉진시키는 추가의 이점을 부여하고, 또한 임의의 증폭 단계 동안 증폭 바이어스를 감소시키거나 제거한다. 또한, 미세 유체 포맷의 구현은 DNA의 매우 작은 샘플 용적 및 낮은 유입량으로 작동하는 능력 뿐만 아니라 게놈 전체 태깅을 촉진시키기 위해 다수의 샘플 파티션(액적)을 신속하게 처리하는 능력을 부여한다.

[0125] 상기 언급된 바와 같이, 본원에 기재된 방법 및 시스템은 보다 긴 핵산의 짧은 서열 판독을 위한 개별적 분자 콘텍스트를 제공한다. 본원에 사용된 바와 같이, 개별적 분자 콘텍스트는, 예를 들면, 서열 판독 자체 내에 포함되지 않는인접 또는 근위 서열에 관련되는 특이적 서열 판독을 초과하는 서열 콘텍스트를 의미하고, 그 자체로, 전형적으로, 그들이 짧은 서열 판독, 예를 들면, 쌍형성된 판독에 대해 약 150 염기 또는 약 300 염기의 판독에 전체로 또는 부분적으로 포함되지 않을 정도일 것이다. 특히 바람직한 양태에서, 방법 및 시스템은 짧은 서열 판독을 위한 긴 범위 서열 콘텍스트를 제공한다. 이러한 긴 범위 콘텍스트는 소정의 서열 판독과 서로 1 kb 초과, 5 kb 초과, 10 kb 초과, 15 kb 초과, 20 kb 초과, 30 kb 초과, 40 kb 초과, 50 kb 초과, 60 kb 초과, 70 kb 초과, 80 kb 초과, 90 kb 초과 또는 심지어 100 kb 초과 또는 그 이상의 거리 내에 있는 서열 판독의 관계 또는 연결을 포함한다. 보다 긴 범위 개별 분자 콘텍스트를 제공함으로써, 본 발명의 방법 및 시스템은 또한 더욱 더 긴추정 분자 콘텍스트를 제공한다. 본원에 기재된 서열 콘텍스트는, 예를 들면, 짧은 서열 판독의 개별적인 더 긴 분자 또는 연결된 분자의 콘티그으로의 매핑으로부터 저해상도 콘텍스트 뿐만 아니라, 예를 들면, 개별 분자의 인접 결정 서열을 갖는 보다 긴 개별 분자의 많은 부분의 긴 범위 서열분석으로부터 고해상도 서열 콘텍스트를 포함할 수 있고, 여기서, 상기 결정 서열은 1 kb 초과, 5 kb 초과, 10 kb 초과, 15 kb 초과, 20 kb

초과, 30 kb 초과, 40 kb 초과, 50 kb 초과, 60 kb 초과, 70 kb 초과, 80 kb 초과, 90 kb 초과, 또는 심지어 100 kb 초과이다. 서열 콘텍스트의 경우와 같이, 짧은 서열의 긴 핵산, 예를 들면, 개별적인 긴 핵산 분자 또는 연결된 핵산 분자 또는 콘티그의 수집 둘 모두로의 부여는 고 수준의 서열 콘텍스트를 제공하기 위해 긴 핵산 연신에 대한 짧은 서열의 매핑 뿐만 아니라 이러한 긴 핵산을 통해 짧은 서열로부터 조립된 서열을 제공함을 모두 포함할 수 있다.

[0126] *IV. 샘플*

[0127] 인정될 바와 같이, 본원에서 논의된 방법 및 시스템은 임의의 유형의 게놈 재료로부터 표적화된 서열 정보를 수득하기 위해 사용될 수 있다. 상기 게놈 재료는 환자로부터 뽑은 샘플로부터 수득될 수 있다. 본원에 검토된, 방법 및 시스템 내 사용되는 게놈 재료의 예시적인 샘플 및 유형은 비제한적으로 하기를 포함한다: 폴리뉴클레오티드, 핵산, 올리고뉴클레오티드, 순환 무세포 핵산, 순환성 종양 세포 (CTC), 핵산 단편, 뉴클레오티드, DNA, RNA, 펩티드 폴리뉴클레오티드, 상보적 DNA (cDNA), 이중 가닥 DNA (dsDNA), 단일 가닥 DNA (ssDNA), 플라스미드 DNA, 코스미드 DNA, 염색체 DNA, 게놈 DNA (gDNA), 바이러스 DNA, 박테리아 DNA, mtDNA (미토콘드리아 DNA), 리보솜 RNA, 무세포 DNA, 무세포 태아 DNA (cffDNA), mRNA, rRNA, tRNA, nRNA, siRNA, snRNA, snoRNA, scaRNA, 마이크로RNA, dsRNA, 바이러스 RNA 등. 요약하면, 사용되는 샘플은 특정 가공 요구 사항에 따라 달라질 수 있다.

[0128] 핵산을 포함하는 임의의 물질은 샘플의 공급원일 수 있다. 물질은 유체, 예를 들면, 생물학적 유체일 수 있다. 유체 물질은, 혈액, 제대혈, 타액, 소변, 땀, 혈청, 정액, 질 유체, 위 및 소화액, 척수액, 태반액, 공동 유체, 안구액, 혈청, 모유, 림프액 또는 이의 조합을 포함할 수 있지만, 이에 한정되지 않는다. 물질은 고체, 예를 들면, 생물학적 조직일 수 있다. 물질은 정상적인 건강한 조직, 병든 조직, 또는 건강한 조직 및 병든 조직의 혼합물을 포함할 수있다. 일부 경우에, 물질은 종양을 포함할 수 있다. 종양은 양성(비-암) 또는 악성(암)일 수 있다. 종양의 비제한적인 예는 다음을 포함할 수 있다: 섬유육종, 점액육종, 지방육종, 연골육종, 골육종, 척색종, 혈관육종, 내피육종, 림프관육종, 림프관내피육종, 활막종, 중피종, 유잉(Ewing's), 평활근육종, 횡문근육종, 위장관계 암종, 결장 암종, 췌장암, 유방암, 비뇨생식기계 암종, 난소암, 전립선암, 편평 세포 암종, 기저 세포 암종, 선암종, 한선 암종, 피지선 암종, 유두상 암종, 유두상 선암종, 낭종암, 수질 암종, 기관지 암종, 신장 세포 암종, 간세포암, 담관 암종, 융모막 암종, 정상 피종, 배아종, 윌름 종양, 자궁경부암, 내분비계 암종, 고환 종양, 폐 암종, 소세포 폐 암종, 비소세포 폐 암종, 방광 암종, 상피성 암종, 신경교종, 성상 세포종, 수모 세포종, 두개인두종, 상의세포종, 송과체종, 혈관모 세포종, 청신경종, 핍지교종, 수막종, 흑색종, 신경아세포종, 망막아종 또는 이들의 조합. 물질은 다양한 종류의 기관과 관련될 수 있다. 기관의 비제한적인 예는 뇌, 간, 폐, 신장, 전립선, 난소, 비장, 림프절(편도선 포함), 갑상선, 췌장, 심장, 골격근, 장, 후두, 식도, 위 또는 이들의 조합을 포함할 수 있다. 일부 경우에서, 물질은, 비제한적으로 하기를 포함한, 다양한 세포를 포함할 수 있다: 진핵 세포, 원핵 세포, 진균 세포, 심장 세포, 폐 세포, 신장 세포, 간 세포, 췌장 세포, 생식세포, 줄기세포, 유도 만능 줄기 세포, 위장관 세포, 혈구, 암 세포, 박테리아 세포, 인간 미생물군집 샘플로부터 단리된 박테리아 세포, 등. 일부 경우에, 물질은, 예를 들면, 단일 세포의 내용물 또는 다중 세포의 내용물과 같은 세포의 내용물을 포함할 수 있다. 개별 세포를 분석하기 위한 방법 및 시스템은 예를 들면, 하기에 제공되어 있으며, 이의 전체 개시내용은 이의 전문이 본원에 참고로 포함된다: 미국 특허 출원 번호 14/752,641 (2015년 6월 26일 출원됨).

[0129] 샘플은 다양한 대상체로부터 수득될 수 있다. 대상체는 살아 있는 대상체 또는 죽은 대상체일 수 있다. 대상체의 예는 인간, 포유동물, 비인간 포유동물, 설치류, 양서류, 파충류, 개과, 고양이과, 소과, 염소, 양류, 암탉, 아빈(avines), 마우스, 토끼, 곤충, 민달팽이, 미생물, 세균, 기생충 또는 물고기를 포함할 수 있지만, 이에 한정되지 않는다. 일부 경우에, 대상체는 질환 또는 장애를 갖는, 갖는 것으로 의심되거나 발병시킬 위험이 있는 환자일 수 있다. 일부 경우에, 대상체는 임신부일 수 있다. 일부 경우에, 대상체는 정상적인 건강한 임신부일 수 있다. 일부 경우에, 대상체는 특정 선천성 결함이 있는 아기를 임신할 위험이 있는 임신부일 수 있다.

[0130] 샘플은 당해기술에 공지된 임의의 수단에 의해 대상체로부터 수득될 수 있다. 예를 들면, 샘플은 순환계 접근(예: 주사기 또는 다른 장치를 통해 정맥내 또는 동맥내), 분비된 생물학적 샘플(예: 타액, 객담, 소변, 대변 등)의 수집, 생물학적 샘플(예: 수술 중 샘플, 수술 후 샘플 등)의 외과적(예: 생검) 획득, 면봉 표본 채집(예: 구강 면봉, 인두 면봉) 또는 피펫팅을 통해 대상체로부터 수득될 수 있다.

[0131] 본 발명의 바람직한 구현예가 본원에 도시되고 기재되었지만, 이러한 구현예가 단지 예로써 제공된다는 것이 당업자에게 명백할 것이다. 본 발명으로부터 벗어나지 않고, 당업자에게 수많은 변형, 변화 및 치환이 이제 일어

날 것이다. 이제, 기술분야에서 숙련된 자는, 본 발명의 범위를 벗어나지 않으면서 수 많은 변형예, 변경예, 및 치환예를 떠올릴 것이다. 다음 특허청구범위는 본 발명의 범위를 규정하고, 이들 특허청구범위 내의 방법 및 구조 및 이들의 등가물은 그것으로 포함되는 것이 의도된다.

[0132] **실시예**

[0133] *실시예: TP53 유전자의 표적화된 커버리지*

[0134] TP53 유전자를 표적하는 증폭 반응은 수행되었다. p53으로서 또한 공지된, 종양 단백질 p53, 세포성 종양 항원 p53 (UniProt 명칭), 인단백질 p53, 종양 억제인자 p53, 항원 NY-CO-13, 또는 전환-관련된 단백질 53 (TRP53)은 인간에서 TP53 유전자에 의해 암호화되는 단백질이다. p53 단백질은 다중세포 유기체에서 매우 중요하고, 여기서 세포 주기를 조절하고, 따라서, 암을 예방하는, 종양 억제인자로서 기능한다. 이와 같이, p53은 게놈 돌연변이 예방에 의한 안정성 보호에서 이의 역할 때문에 "게놈의 보호자"로서 기재되었다. 따라서 TP53은 종양 억제인자 유전자로서 분류된다.

[0135] (길이 약 19149 bp인) TP53 유전자를 함유한 게놈의 영역의 표적화된 증폭은 멀티플렉스 반응에서 전체 유전자에 미치는 총 96 프라이머를 이용하여 수행되었다. 프라이머는 약 400 bp 떨어져서 게놈의 상기 영역에 걸쳐 타일링하도록 설계되었다. 증폭 반응은 어닐링 단계용 온도 구배, 14 사이클, 및 DNA의 약 3 ng의 유입량으로 수행되었다. 본 예에서 사용된 열순환 프로토콜은 아래와 같았다:

| | | |
|---|---|---|
| 초기 변성 | 98℃ | 30초 |
| 18 주기 | 98℃ | 10초 |
| | 30-55℃ | 15초 |
| | 72℃ | 15초 |
| 최종 연신 | 72℃ | 2분 |
| 보유 | 4C | |

[0136]

[0137] 이러한 유형의 반응에 대하여 예시적인 작업흐름은 도 8에서 묘사된다. 인정될 바와 같이, 이는 본원에서 기재된 본 발명에 따른 방법의 예시적인 구현예이고 공지된 방법을 이용하여 변경 또는 확장될 수 있다. 도 8에 나타낸 바와 같이, 게놈의 선택된 영역 (이 경우에, TP53 유전자)는 표적 특이적 프라이머, 예컨대 802 및 803으로서 묘사된 것을 이용하여 증폭된다. 또한, 바코드 801이 있는 프라이머는 증폭산물 속에 또한 편입되었고, 이는 본원에서 기재된 바와 같이 특정 구현예에서 후속의 서열 판독 (808)에 분자 콘텍스트를 제공할 수 있다.

[0138] 프라이머 802 및 803은, 본 실험에서 "테일" R1 및 R2를 가졌고, 이는 특정 플랫폼, 예컨대 Illumina 플랫폼 상에서 서열분석에 수득한 증폭산물이 잘 받아들이도록 하였다. SI 프라이머 (806)를 이용한 증폭은 Illumina 플랫폼과 함께 또한 사용되는 샘플 지수를 추가로 제공하였다. 인정될 바와 같이, 다른 서열분석 플랫폼에 유용한 서열은 R1 및 R2 그리고 S1 프라이머 대신 사용될 수 있다.

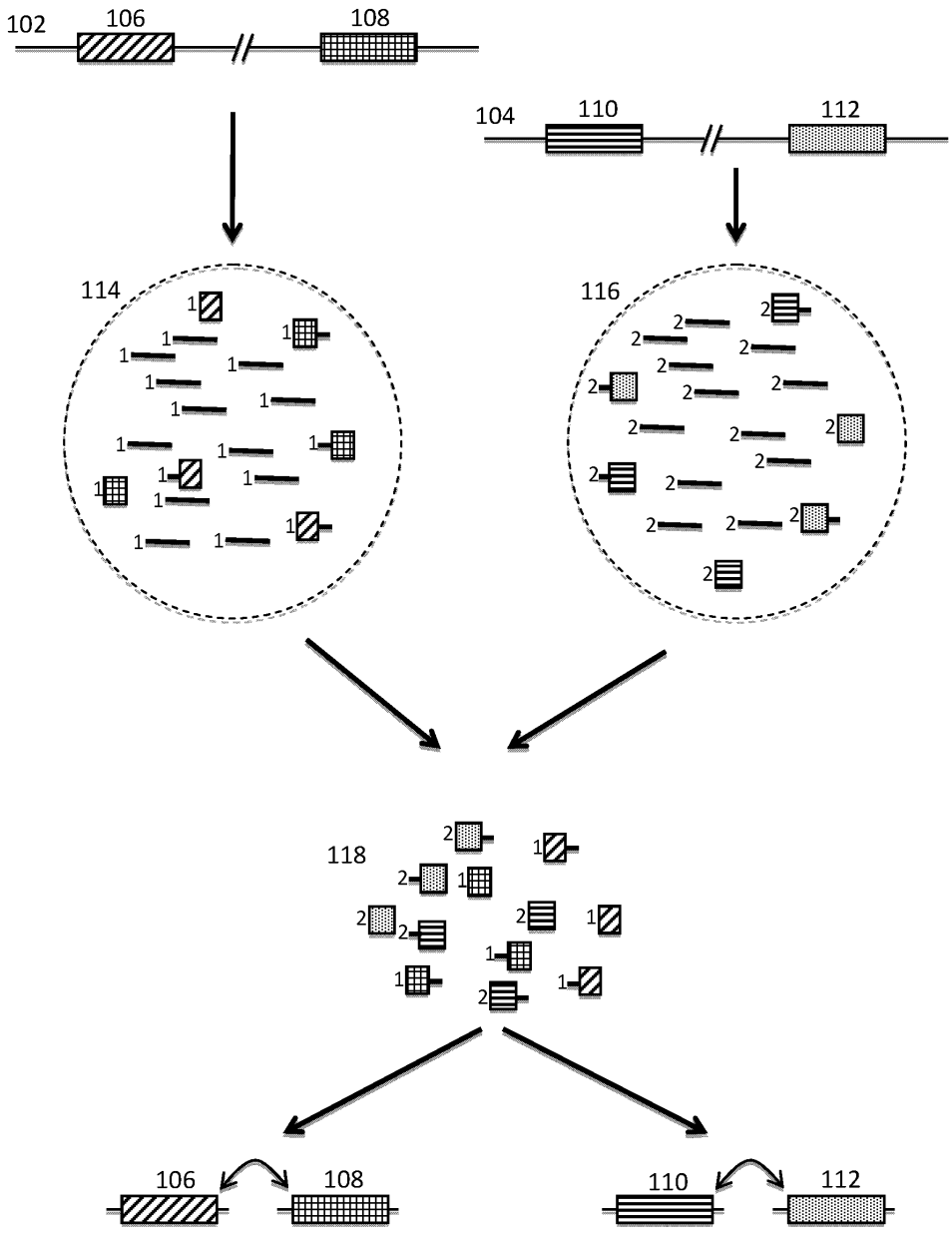
[0139] 도 9는, 무 주형 대조군 (NTC)이 무 생성물을 보여주었던 경우, 증폭 반응이 특이적이었다는 것을 보여준다. 도 10은 온도의 범위를 거쳐 상기-기재된 프로토콜의 결과로서 보여진 배수-농축을 제공한다.

[0140] 본 명세서는 현재-기재된 기술의 예 양태에서 방법론, 시스템 및/또는 구조 및 이의 용도의 완전한 설명을 제공한다. 상기 기술의 다양한 양태가 특수성의 특정 정도로, 또는 하나 이상의 개별 양태과 관련하여, 상기 기재되었어도, 당해 분야의 숙련가는 본원 기술의 취지 또는 범위의 이탈 없이 개시된 양태에 수많은 변경을 할 수 있다. 많은 양태가 현재 기재된 기술의 취지 및 범위로부터 이탈 없이 실시될 수 있기 때문에, 적절한 범주는 이하에서 첨부된 청구항에 있다. 다른 양태는 따라서 고려된다. 더욱이, 명백하게 달리 청구되지 않거나 특정 순서가 청구항 언어에 의해 본질적으로 필요해지지 않는 한, 임의의 작업은 임의의 순서로 수행될 수 있다는 것이 이해되어야 한다. 상기 설명에 포함되고 수반되는 도면들에서 보여진 모든 내용이 특정한 양태의 단지 예시적으로서 해석될 것이고 보여진 구현예로 제한되지 않는다는 의도이다. 콘텍스트으로부터 달리 명백하지 않거나 명확히 언급되지 않는 한, 본원에서 제공된 임의의 농도 값은 혼합물의 특정한 성분의 부가시 또는 부가 이후 발생하는 임의의 전환과 무관하게 혼합물 값 또는 백분율에 관하여 일반적으로 소정의다. 본원에서 이미 명확히

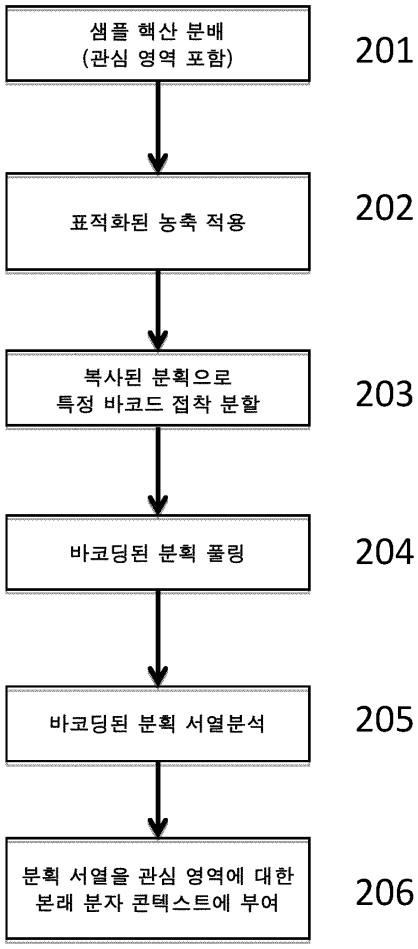
편입되지 않은 정도로, 본 개시내용에서 참조된 모든 공개된 참조 및 특허 문서는 모든 목적으로 그 전체가 참고로 본원에서 편입된다. 상세 또는 구조에서 변화는 하기 청구항에서 정의된 바와 같이 본 기술의 기본 요소로부터 이탈 없이 실시될 수 있다.

## 도면

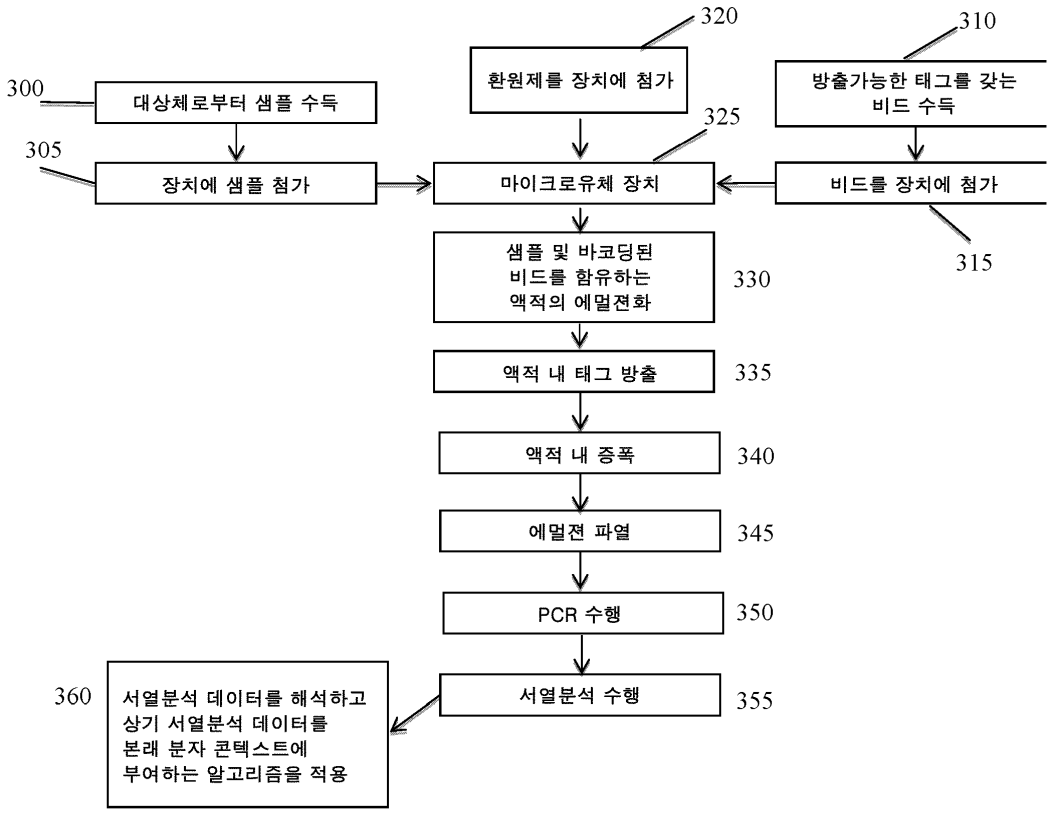
### 도면1



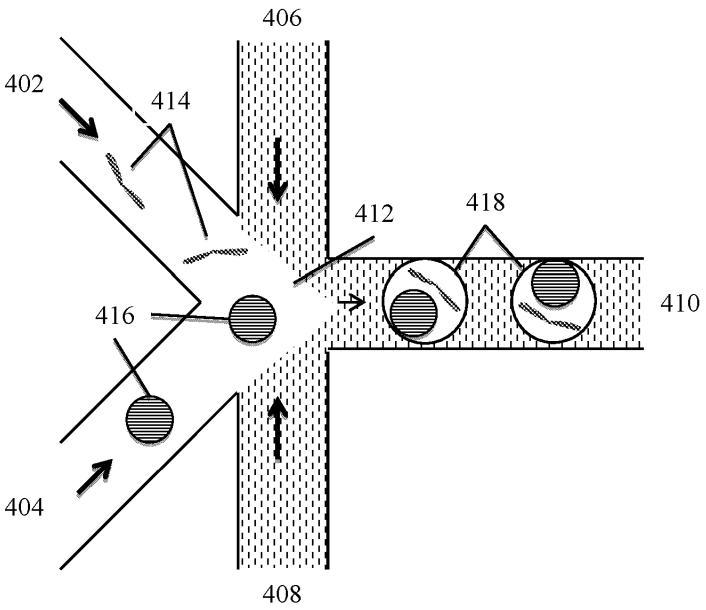
도면2


샘플 핵산 분배
(관심 영역 포함)
201
표적화된 농축 적용
202
복사된 분획으로
특정 바코드 접착 분할
203
바코딩된 분획 풀링
204
바코딩된 분획 서열분석
205
분획 서열을 관심 영역에 대한
본래 분자 콘텍스트에 부여
206

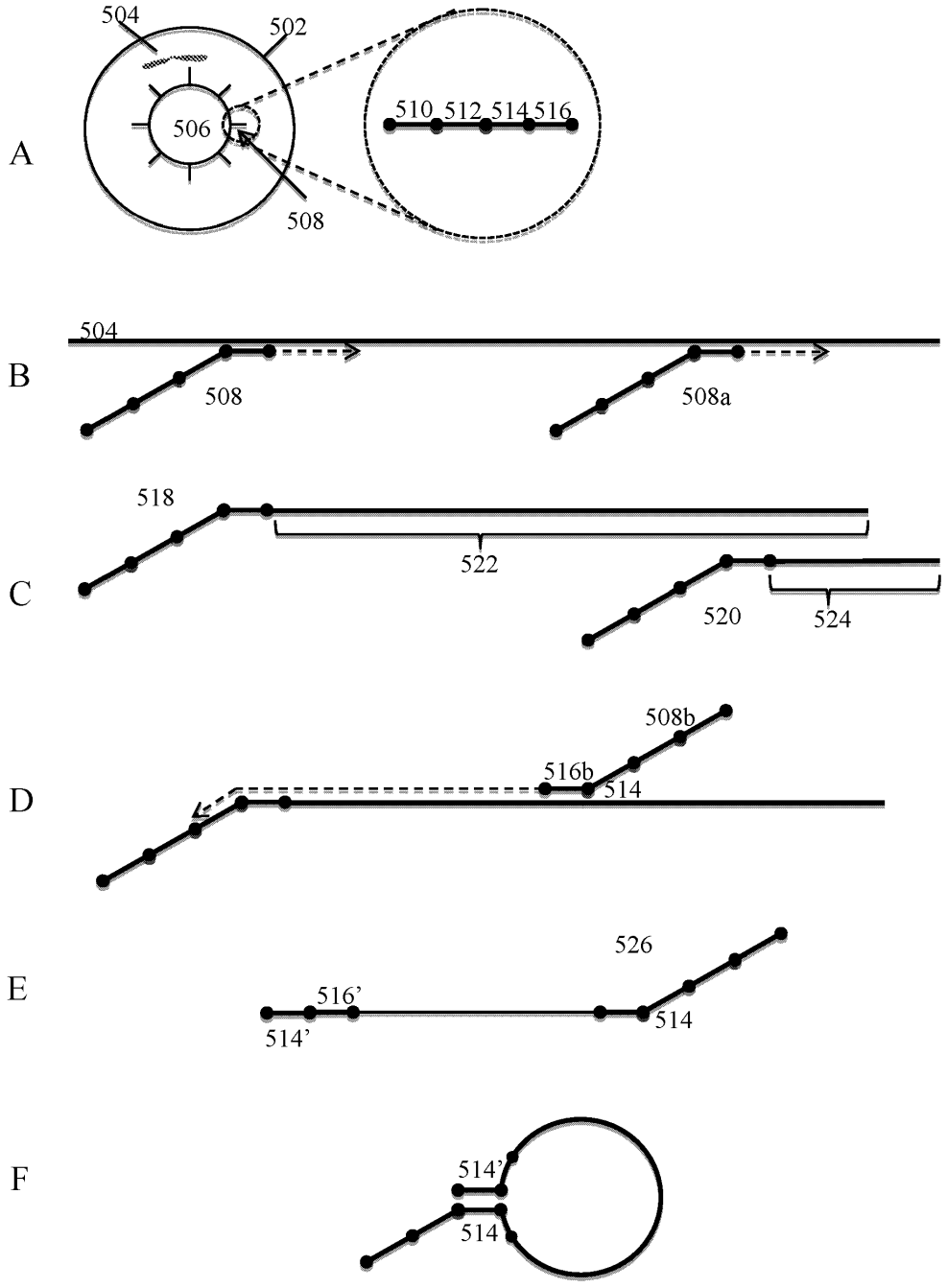
도면3


320
310
환원제를 장치에 첨가
방출가능한 태그를 갖는
비드 수득
300
대상체로부터 샘플 수득
325
305
장치에 샘플 첨가
마이크로유체 장치
비드를 장치에 첨가
315
샘플 및 바코딩된
비드를 함유하는
액적의 에멀젼화
330
액적 내 태그 방출
335
액적 내 증폭
340
에멀젼 파열
345
PCR 수행
350
360
서열분석 데이터를 해석하고
상기 서열분석 데이터를
본래 분자 콘텍스트에
부여하는 알고리즘을 적용
서열분석 수행
355

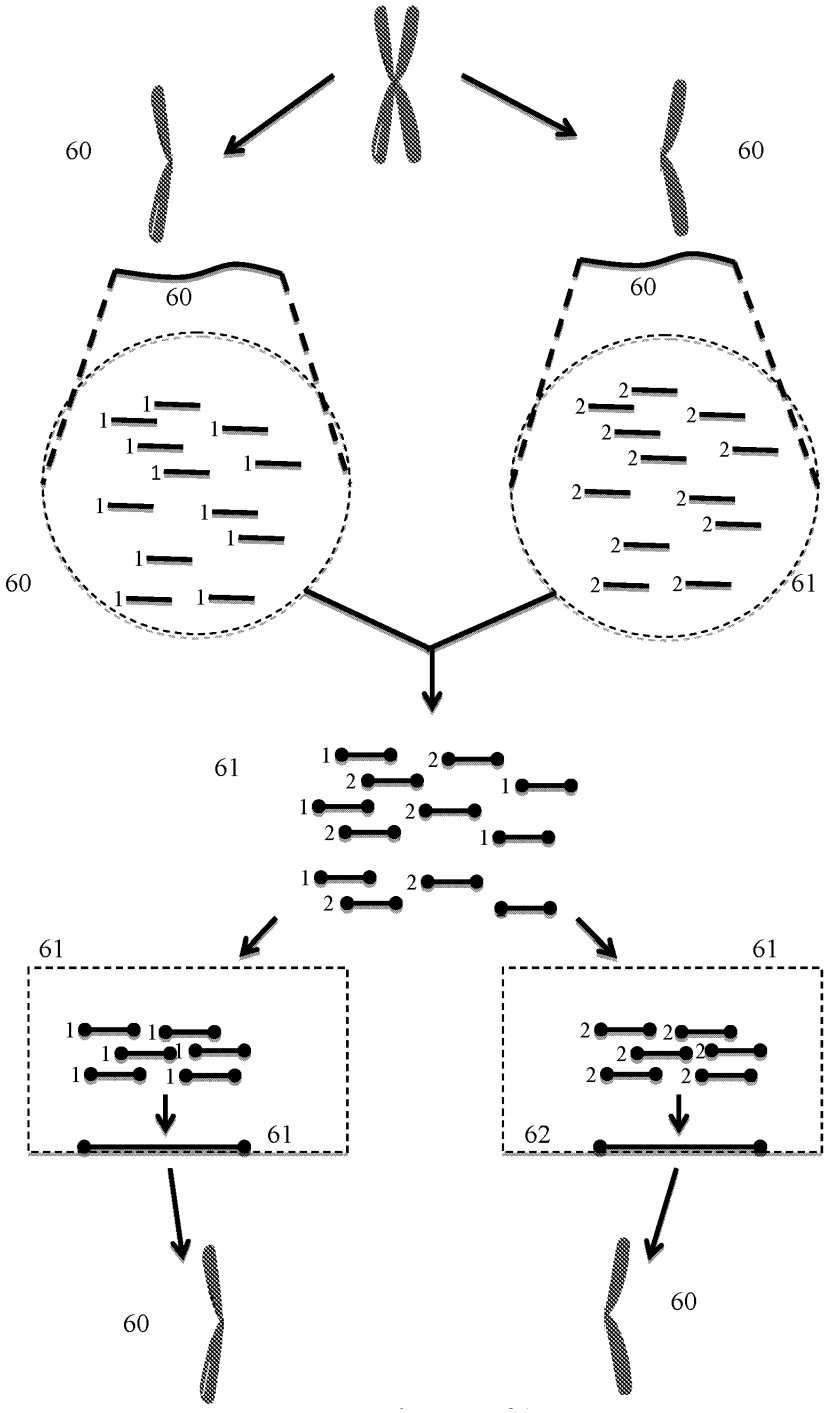
도면4


406
402
414
412
418
410
416
404
408

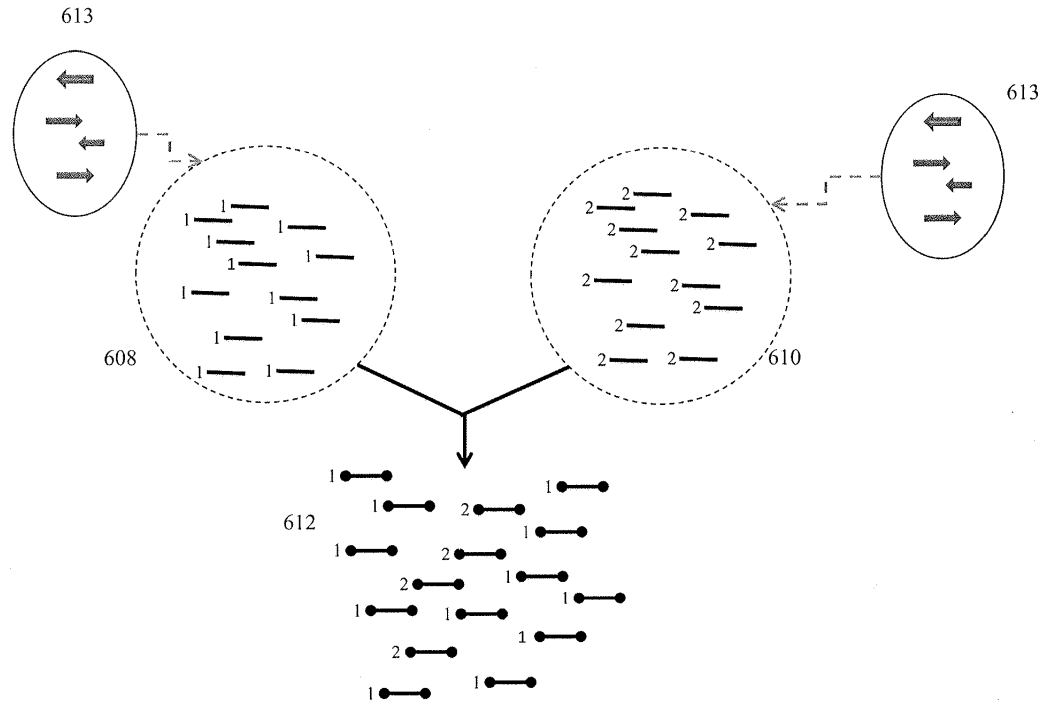
도면5


A
504
502
506
508
510 512 514 516
B
504
508
508a
C
518
522
520
524
D
508b
516b
514
E
526
516'
514'
514
F
514'
514

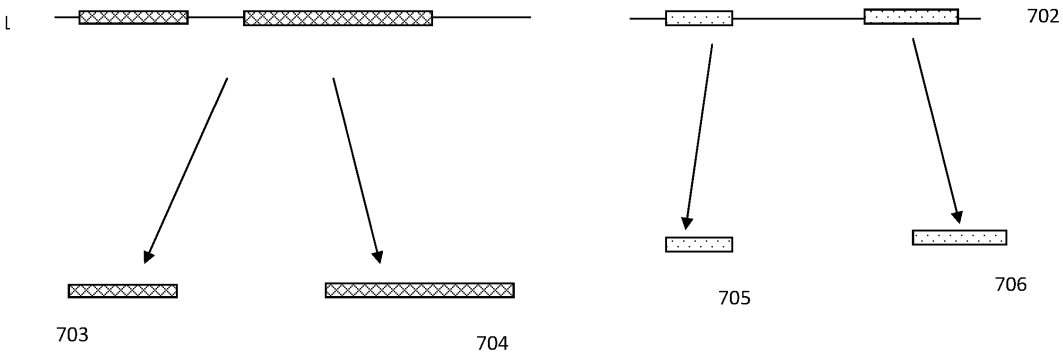
도면6a



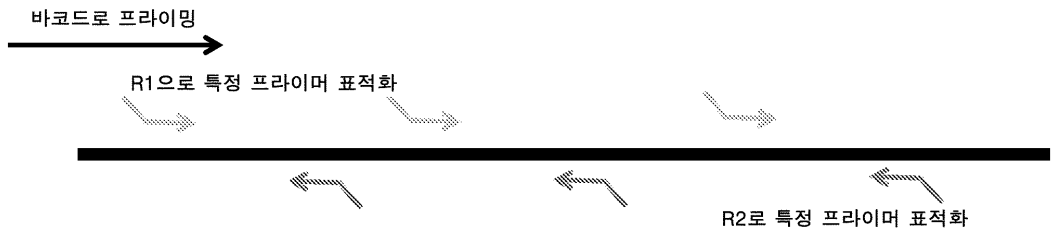
도면6b




도면7



도면8




증폭 반응

↓

세정

↓

SI 프라이머로의 라이브러리 증폭 PCR

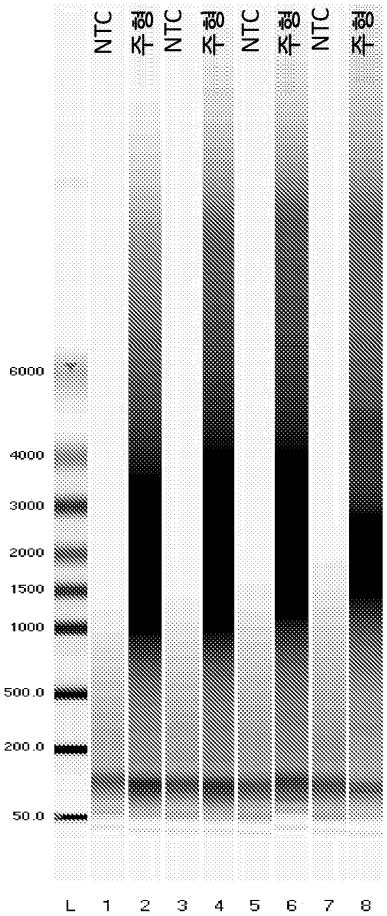
↓

세정

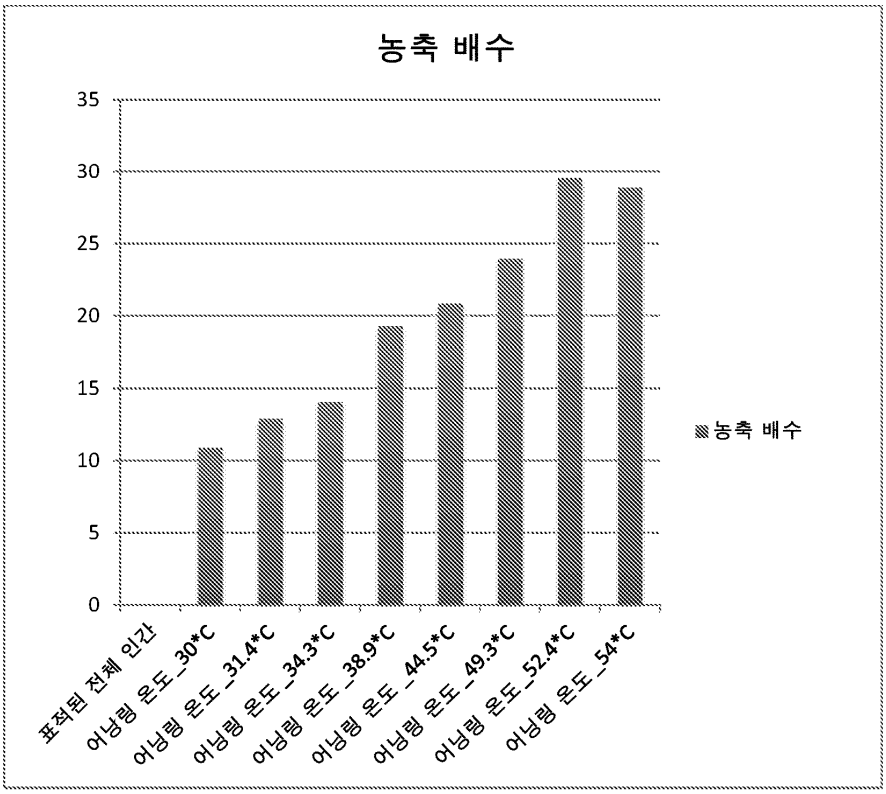
↓

정량화 및 서열분석

도면9



도면10


농축 배수
35
30
25
20
15
10
5
0
표적된 전체 인간
어닝링 온도_30*C
어닝링 온도_31.4*C
어닝링 온도_34.3*C
어닝링 온도_38.9*C
어닝링 온도_44.5*C
어닝링 온도_49.3*C
어닝링 온도_52.4*C
어닝링 온도_54*C
농축 배수