



등록특허 10-2696857



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2024년08월19일
(11) 등록번호 10-2696857
(24) 등록일자 2024년08월14일

(51) 국제특허분류(Int. Cl.)
C12Q 1/6881 (2018.01) *C12Q 1/6869* (2018.01)
C12Q 1/6883 (2018.01) *C12Q 1/6886* (2018.01)
G16B 20/00 (2019.01) *G16B 40/00* (2019.01)
G16B 45/00 (2019.01) *G16H 50/20* (2018.01)

(52) CPC특허분류
C12Q 1/6881 (2018.05)
C12Q 1/6869 (2018.05)

(21) 출원번호 10-2022-7030533(분할)

(22) 출원일자(국제) 2015년07월27일
 심사청구일자 2022년09월30일

(85) 번역문제출일자 2022년09월02일

(65) 공개번호 10-2022-0127359

(43) 공개일자 2022년09월19일

(62) 원출원 특허 10-2017-7004904
 원출원일자(국제) 2015년07월27일
 심사청구일자 2020년07월24일

(86) 국제출원번호 PCT/US2015/042310

(87) 국제공개번호 WO 2016/015058
 국제공개일자 2016년01월28일

(30) 우선권주장
 62/029,178 2014년07월25일 미국(US)
 62/087,619 2014년12월04일 미국(US)

(56) 선행기술조사문헌
 WO2013177086 A1*

*는 심사관에 의하여 인용된 문헌

전체 청구항 수 : 총 17 항

심사관 : 박정웅

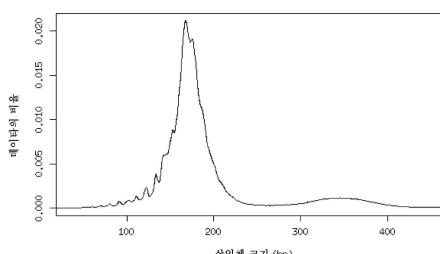
(54) 발명의 명칭 무세포 DNA를 생성하는 조직 및/또는 세포 유형을 결정하는 방법 및 이를 사용하여 질환 또는 장애를 확인하는 방법

(57) 요 약

본 개시내용은 대상체의 생물학적 샘플에서 무세포 DNA ("cfDNA")에 기여하는 하나 이상의 조직 및/또는 세포 유형을 결정하는 방법을 제공한다. 일부 실시양태에서, 본 개시내용은 대상체의 생물학적 샘플에서 cfDNA에 기여하는 하나 이상의 결정된 조직 및/또는 세포 유형의 험수로서 대상체에서 질환 또는 장애를 확인하는 방법을 제

(뒷면에 계속)

대 표 도 - 도2



공한다.

(52) CPC특허분류

C12Q 1/6883 (2022.01)
C12Q 1/6886 (2022.01)
G16B 20/00 (2019.02)
G16B 40/00 (2019.02)
G16B 45/00 (2019.02)
G16H 50/20 (2018.01)
C12Q 2535/122 (2019.08)
C12Q 2537/165 (2013.01)

명세서

청구범위

청구항 1

시험 대상체에서 암의 기원 조직 및/또는 세포 유형을 확인하기 위하여 무세포 DNA(cfDNA)를 생성하는 조직 및/또는 세포 유형을 결정하기 위한 정보를 제공하는 방법이며,

(a) 시험 대상체로부터의 시험 샘플에 존재하는 cfDNA의 적어도 일부분의 서열을 수득하고, 여기서 서열은 종점을 갖고 서열은 계놈 위치에 지도화할 수 있는 것인 단계;

(b) 하나 이상의 계놈 위치에서 샘플 cfDNA 종점의 빈도의 시험 샘플값을 계산하고, 여기서 샘플 cfDNA 종점의 빈도는 계놈 위치에 지도화가능한 서열결정된 분자의 수 및 종점의 수의 함수인 단계;

(c) 시험 샘플값을 제1 참조 샘플로부터 수득된 cfDNA 분자의 서열로부터 수득된 cfDNA 종점의 빈도로부터 계산된 제1 참조값과 비교하고, 여기서 cfDNA 종점의 빈도는 계놈 위치에 지도화가능한 서열결정된 분자의 수 및 종점의 수의 함수인 단계;

여기서, 제1 참조 샘플은 암의 존재 또는 부재를 갖는 것으로 알려져 있는 사람으로부터 수득되는 것이고,

(d) cfDNA 종점의 적어도 일부의 계놈 위치의 함수로서 cfDNA 종점을 생성하는 조직 및/또는 세포 유형의 적어도 일부를 결정하기 위한 정보를 제공하고, 여기서 암의 기원 조직 및/또는 세포 유형이 확인되는 것인 단계를 포함하는, 방법.

청구항 2

제1항에 있어서, 제1 참조 샘플이 암을 갖는 것으로 알려져 있는 사람으로부터 수득되는 것인 방법.

청구항 3

제1항에 있어서, 제1 참조 샘플이 암을 갖지 않는 것으로 알려져 있는 사람으로부터 수득되는 것인 방법.

청구항 4

제2항에 있어서,

샘플 cfDNA 종점의 빈도의 값을 제2 참조 샘플로부터 수득된 cfDNA의 서열로부터 수득된 cfDNA 종점으로부터 계산된 제2 참조값과 비교하고, 여기서 종점의 빈도는 정의된 계놈 위치에 지도화가능한 서열결정된 분자의 수 및 종점의 수의 함수이고, 여기서 제2 참조 샘플은 암을 갖지 않는 것으로 알려져 있는 사람으로부터 수득되는 것인 단계

를 추가로 포함하는 방법.

청구항 5

제4항에 있어서, cfDNA 종점의 적어도 일부의 계놈 위치의 함수로서 cfDNA 종점을 생성하는 조직 및/또는 세포 유형의 적어도 일부를 결정하기 위한 정보를 제공하는 단계가 시험 샘플값이 제1 참조값 또는 제2 참조값과 더 유사한지를 결정하는 것을 포함하는 방법.

청구항 6

제5항에 있어서, 계놈 위치가 전사 인자 결합 부위를 포함하는 것인 방법.

청구항 7

제5항에 있어서, 계놈 위치가 CTCF 결합 부위를 포함하는 것인 방법.

청구항 8

제5항에 있어서, 서열결정된 cfDNA가 조혈 세포 및 비-조혈 세포로부터 유래된 DNA를 포함하고, 여기서 비-조혈 세포의 적어도 일부분이 종양 세포인 방법.

청구항 9

제7항에 있어서, 시험 샘플값 및 참조값이 복수개의 게놈 위치에 대해 계산되는 것인 방법.

청구항 10

제7항에 있어서, cfDNA가 대량 병렬 DNA 서열결정기 상에서 서열결정되는 것인 방법.

청구항 11

제5항에 있어서, 게놈 위치가 각각의 게놈 위치에 대해 복수개의 별개의 게놈 위치 및 별개의 시험 샘플값 및 참조값을 포함하는 것인 방법.

청구항 12

제11항에 있어서, 시험 샘플값 및 참조값을 복수개의 게놈 위치에 대해 비교하는 것인 방법.

청구항 13

제7항에 있어서, 시험 샘플값 및 참조값이 단일 게놈 좌표 해상도 수준에서 계산되는 것인 방법.

청구항 14

제5항에 있어서, 시험 샘플값 및 참조값이 게놈 위치 내의 복수개의 단일 엔기 위치에 대해 단일 게놈 좌표 해상도에서 계산되는 것인 방법.

청구항 15

제13항에 있어서, 시험 샘플값 및 참조값이 게놈 위치 내의 복수개의 단일 엔기 위치에 대해 단일 게놈 좌표 해상도에서 계산되는 것인 방법.

청구항 16

제14항에 있어서, 시험 샘플값 및 참조값이 복수개의 게놈 위치에 대해 비교되는 것인 방법.

청구항 17

제15항에 있어서, 시험 샘플값 및 참조값이 복수개의 게놈 위치에 대해 비교되는 것인 방법.

발명의 설명

기술 분야

[0001]

우선권 주장

[0002]

본원은 마치 본원에 완전히 제시된 것처럼 그 각각의 주제가 전부 본원에 참고로 포함된, 2014년 7월 25일 출원된 미국 특허 가출원 62/029,178, 및 2014년 12월 4일 출원된 미국 특허 가출원 62/087,619를 기초로 한 우선권을 주장한다.

[0003]

미국 정부의 권리에 대한 언급

[0004]

본 발명은 미국 국립 보건원 (NIH)이 부여한 승인 번호 1DP1HG007811 하의 정부 지원으로 이루어졌다. 미국 정부는 본 발명에 대해 특정 권리를 갖는다.

[0005]

기술 분야

[0006]

본 개시내용은 무세포 DNA를 생성하는 하나 이상의 조직 및/또는 세포 유형을 결정하는 방법에 관한 것이다. 일부 실시양태에서, 본 개시내용은 대상체의 생물학적 샘플에서 무세포 DNA와 연관된 하나 이상의 결정된 조직 및/또는 세포 유형의 함수로서 대상체에서 질환 또는 장애를 확인하는 방법을 제공한다.

배경 기술

[0007]

무세포 DNA ("cfDNA")는 순환하는 혈장, 소변 및 인간의 다른 체액에 존재한다. cfDNA는 상대적으로 짧고 (압도적으로 200개 염기쌍 미만) 일반적으로 저농도 (예를 들어, 혈장에서 1-100 ng/mL)인 이중 가닥 DNA 단편을 포함한다. 건강한 개체의 순환하는 혈장에서, cfDNA는 주로 혈액 세포 (즉, 조혈 계통의 정상 세포)의 아폽토시스 (apoptosis)로부터 유래하는 것으로 생각된다. 그러나, 특정 상황에서, 다른 조직은 순환하는 혈장과 같은 체액에서 cfDNA의 조성에 실질적으로 기여할 수 있다.

[0008]

cfDNA는 특정 전문 분야 (예를 들어, 생식 의학, 암 진단 및 이식 의학)에서 사용되었지만, cfDNA에 기초한 기존의 시험은 2개 이상의 세포 집단 (예를 들어, 모계 계놈 대 태아 계놈, 정상 계놈 대 암 계놈, 이식 수혜자 계놈 대 공여자 계놈 등) 사이의 유전자형 (예를 들어, 특정 서열의 1차 서열 또는 카페수 제시)의 차이에 의존한다. 불행하게도, 임의의 제시된 생물학적 샘플에서 발견된 cfDNA 단편의 압도적인 다수가, 기여하는 세포 집단 사이에서 서열이 동일한 계놈의 영역으로부터 유래하기 때문에, 기존의 cfDNA 기반 시험은 그 적용범위가 극히 제한적이다. 또한, 많은 질환 및 장애는 예를 들어 조직 손상 또는 질환 또는 장애와 연관된 염증 과정으로부터 cfDNA를 생성하는 조직 및/또는 세포 유형의 변화를 수반한다. 두 계놈 사이의 특정 서열의 1차 서열 또는 카페수 제시의 차이에 의존하는 기존의 cfDNA 기반 진단 시험은 이러한 변화를 검출할 수 없다. 따라서, 강력한 무생검 진단 방법을 제공하는 cfDNA의 잠재력은 매우 크지만, 매우 다양한 질환 및 장애의 진단에 적용될 수 있는 cfDNA 기반 진단 방법이 여전히 필요하다.

발명의 내용

[0009]

본 개시내용은 대상체의 생물학적 샘플에서 무세포 DNA ("cfDNA")를 생성하는 하나 이상의 조직 및/또는 세포 유형을 결정하는 방법을 제공한다. 일부 실시양태에서, 본 개시내용은 대상체의 생물학적 샘플에서 cfDNA와 연관된 하나 이상의 결정된 조직 및/또는 세포 유형의 함수로서 대상체에서 질환 또는 장애를 확인하는 방법을 제공한다.

[0010]

일부 실시양태에서, 본 개시내용은 대상체에서 무세포 DNA (cfDNA)를 생성하는 조직 및/또는 세포 유형을 결정하는 방법을 제공하고, 이 방법은 대상체로부터의 생물학적 샘플로부터 cfDNA를 단리하고, 단리된 cfDNA는 복수 개의 cfDNA 단편을 포함하는 것인 단계; 복수개의 cfDNA 단편의 적어도 일부분과 연관된 서열을 결정하는 단계; cfDNA 단편 서열의 함수로서 복수개의 cfDNA 단편의 적어도 일부의 cfDNA 단편 종점에 대한 참조 계놈 내의 계놈 위치를 결정하는 단계; 및 cfDNA 단편 종점의 적어도 일부의 계놈 위치의 함수로서 cfDNA 단편을 생성하는 조직 및/또는 세포 유형의 적어도 일부를 결정하는 단계를 포함한다.

[0011]

다른 실시양태에서, 본 개시내용은 대상체에서 질환 또는 장애를 확인하는 방법을 제공하고, 상기 방법은 대상체로부터의 생물학적 샘플로부터 무세포 DNA (cfDNA)를 단리하고, 단리된 cfDNA는 복수개의 cfDNA 단편을 포함하는 것인 단계; 복수개의 cfDNA 단편의 적어도 일부분과 연관된 서열을 결정하는 단계; cfDNA 단편 서열의 함수로서 복수개의 cfDNA 단편의 적어도 일부의 cfDNA 단편 종점에 대한 참조 계놈 내의 계놈 위치를 결정하는 단계; cfDNA 단편 종점의 적어도 일부의 계놈 위치의 함수로서 cfDNA를 생성하는 조직 및/또는 세포 유형의 적어도 일부를 결정하는 단계; 및 cfDNA를 생성하는 결정된 조직 및/또는 세포 유형의 함수로서 질환 또는 장애를 확인하는 단계를 포함한다.

[0012]

다른 실시양태에서, 본 개시내용은 대상체에서 무세포 DNA (cfDNA)를 생성하는 조직 및/또는 세포 유형을 결정하는 방법을 제공하고, 이 방법은 (i) 대상체로부터 생물학적 샘플을 얻고, 생물학적 샘플로부터 cfDNA를 단리하고, cfDNA의 라이브러리 구축 및 대량 병렬 서열결정에 의해 분포 (a), (b) 및/또는 (c)를 측정하여 뉴클레오솜 지도를 작성하는 단계; (ii) 대조군 대상체 또는 알려진 질환을 갖는 대상체로부터 생물학적 샘플을 얻고, 생물학적 샘플로부터 cfDNA를 단리하고, cfDNA의 라이브러리 구축 및 대량 병렬 서열결정에 의해 분포 (a), (b) 및/또는 (c)를 측정하여 뉴클레오솜 지도의 참조 세트를 작성하는 단계; 및 (iii) 생물학적 샘플로부터의 cfDNA로부터 유래된 뉴클레오솜 지도를 뉴클레오솜 지도의 참조 세트와 비교하여 생물학적 샘플로부터 cfDNA를 생성하는 조직 및/또는 세포 유형을 결정하는 단계를 포함하고; 여기서, (a), (b) 및 (c)는 다음과 같다: (a) 인간 계놈 내의 임의의 특정 염기쌍이 cfDNA 단편의 말단에서 나타날 가능성의 분포; (b) 인간 계놈의 염기쌍의 임의의 쌍이 cfDNA 단편의 한 쌍의 말단으로서 나타날 가능성의 분포; 및 (c) 인간 계놈 내의 임의의 특정 염기쌍이 차별적인 뉴클레오솜 점유의 결과로서 cfDNA 단편 내에 나타날 가능성의 분포.

[0013]

또 다른 실시양태에서, 본 개시내용은 대상체에서 cfDNA를 생성하는 조직 및/또는 세포 유형을 결정하는 방법을 제공하고, 이 방법은 (i) 대상체로부터 생물학적 샘플을 얻고, 생물학적 샘플로부터 cfDNA를 단리하고, cfDNA의

라이브러리 구축 및 대량 병렬 서열결정에 의해 분포 (a), (b) 및/또는 (c)를 측정하여 뉴클레오솜 지도를 작성하는 단계; (i) 대조군 대상체 또는 알려진 질환을 갖는 대상체로부터 생물학적 샘플을 얻고, 생물학적 샘플로부터 cfDNA를 단리하고, 미크로코쿠스 뉴클레아제, DNase 또는 트랜스포사제와 같은 효소를 사용한 염색질의 단편화로부터 유래된 DNA의 라이브러리 구축 및 대량 병렬 서열결정에 의해 분포 (a), (b) 및/또는 (c)를 측정하여 뉴클레오솜 지도의 참조 세트를 작성하는 단계; 및 (iii) 생물학적 샘플로부터의 cfDNA로부터 유래된 뉴클레오솜 지도를 뉴클레오솜 지도의 참조 세트와 비교하여 생물학적 샘플로부터 cfDNA를 생성하는 조직 및/또는 세포 유형을 결정하는 단계를 포함하고; 여기서, (a), (b) 및 (c)는 다음과 같다: (a) 인간 계놈 내의 임의의 특정 염기쌍이 서열결정되는 단편의 말단에서 나타날 가능성의 분포; (b) 인간 계놈의 염기쌍의 임의의 쌍이 서열결정되는 단편의 한 쌍의 말단으로서 나타날 가능성의 분포; 및 (c) 인간 계놈 내의 임의의 특정 염기쌍이 차별적인 뉴클레오솜 점유의 결과로서 서열결정되는 단편 내에 나타날 가능성의 분포.

[0014]

다른 실시양태에서, 본 개시내용은 대상체에서 임상 병태를 진단하는 방법을 제공하고, 이 방법은 (i) 대상체로부터 생물학적 샘플을 얻고, 생물학적 샘플로부터 cfDNA를 단리하고, cfDNA의 라이브러리 구축 및 대량 병렬 서열결정에 의해 분포 (a), (b) 및/또는 (c)를 측정하여 뉴클레오솜 지도를 작성하는 단계; (ii) 대조군 대상체 또는 알려진 질환을 갖는 대상체로부터 생물학적 샘플을 얻고, 생물학적 샘플로부터 cfDNA를 단리하고, cfDNA의 라이브러리 구축 및 대량 병렬 서열결정에 의해 분포 (a), (b) 및/또는 (c)를 측정하여 뉴클레오솜 지도의 참조 세트를 작성하는 단계; 및 (iii) 생물학적 샘플로부터의 cfDNA로부터 유래된 뉴클레오솜 지도를 뉴클레오솜 지도의 참조 세트와 비교하여 임상 병태를 결정하는 단계를 포함하고; 여기서, (a), (b) 및 (c)는 다음과 같다: (a) 인간 계놈 내의 임의의 특정 염기쌍이 cfDNA 단편의 말단에서 나타날 가능성의 분포; (b) 인간 계놈의 염기쌍의 임의의 쌍이 cfDNA 단편의 한 쌍의 말단으로서 나타날 가능성의 분포; (c) 인간 계놈 내의 임의의 특정 염기쌍이 차별적인 뉴클레오솜 점유의 결과로서 cfDNA 단편 내에 나타날 가능성의 분포.

[0015]

다른 실시예에서, 본 개시내용은 대상체에서 임상 병태를 진단하는 방법을 제공하고, 이 방법은 (i) 대상체로부터 생물학적 샘플을 얻고, 생물학적 샘플로부터 cfDNA를 단리하고, cfDNA의 라이브러리 구축 및 대량 병렬 서열결정에 의해 분포 (a), (b) 및/또는 (c)를 측정하여 뉴클레오솜 지도를 작성하는 단계; (ii) 대조군 대상체 또는 알려진 질환을 갖는 대상체로부터 생물학적 샘플을 얻고, 생물학적 샘플로부터 cfDNA를 단리하고, 미크로코쿠스 뉴클레아제 (MNase), DNase 또는 트랜스포사제와 같은 효소를 사용한 염색질의 단편화로부터 유래된 DNA의 라이브러리 구축 및 대량 병렬 서열결정에 의해 분포 (a), (b) 및/또는 (c)를 측정하여 뉴클레오솜 지도의 참조 세트를 작성하는 단계; 및 (iii) 생물학적 샘플로부터의 cfDNA로부터 유래된 뉴클레오솜 지도를 뉴클레오솜 지도의 참조 세트와 비교하여 생물학적 샘플로부터의 cfDNA의 기원 조직 조성을 결정하는 단계를 포함하고; 여기서, (a), (b) 및 (c)는 다음과 같다: (a) 인간 계놈 내의 임의의 특정 염기쌍이 서열결정되는 단편의 말단에서 나타날 가능성의 분포; (b) 인간 계놈의 염기쌍의 임의의 쌍이 서열결정되는 단편의 한 쌍의 말단으로서 나타날 가능성의 분포; (c) 인간 계놈 내의 임의의 특정 염기쌍이 차별적인 뉴클레오솜 점유의 결과로서 서열결정되는 단편 내에 나타날 가능성의 분포.

[0016]

이들 및 다른 실시양태는 아래에서 보다 상세하게 설명된다.

도면의 간단한 설명

[0017]

도 1은 작은 계놈 영역에 대해 예시된, cfDNA 단편화 패턴을 뉴클레오솜 점유와 관련시키는 3가지 유형의 정보를 보여준다. 이러한 동일한 유형의 정보는 또한 미크로코쿠스 뉴클레아제 (MNase), DNase 또는 트랜스포사제와 같은 효소를 이용한 염색질의 단편화를 통해 발생할 수 있다. 도 1a는 인간 계놈 내의 임의의 특정 염기쌍이 서열결정되는 단편의 말단 (즉, 단편화 지점)에 나타날 가능성의 분포를 보여주고; 도 1b는 인간 계놈의 염기쌍의 임의의 쌍이 서열결정되는 단편의 한 쌍의 말단 (즉, 개별 분자를 생성하는 단편화 지점의 연속적인 쌍)으로 나타날 가능성의 분포를 보여주고; 도 1c는 인간 계놈 내의 임의의 특정 염기쌍이 상이한 뉴클레오솜 점유의 결과로서 서열결정되는 단편 내에 나타날 가능성 (즉, 상대적 적용범위)의 분포를 보여준다.

도 2는 전형적인 cfDNA 서열결정 라이브러리의 삽입 크기 분포를 보여주고, 여기에는 알 수 없는 수의 건강한 개체로부터의 기증물을 포함하는 인간 혈장으로부터 유래된 한데 모든 cfDNA 샘플 (bulk cfDNA)이 도시된다.

도 3a는 모든 cfDNA 샘플 (혈장), 종양 환자 샘플로부터의 cfDNA (종양), 임신한 여성 개체로부터의 cfDNA (임신), 인간 상이한 인간 세포주의 MNase (세포주) 및 인간 DNA 샷건 (샷건) 서열결정 라이브러리 (Shotgun)에 걸쳐 제1 (chr1) 인간 상염색체에 지도화되는 판독 개시 좌표의 고속 푸리에 변환 (Fast Fourier Transformation, FFT)으로부터의 평균 주기도 (periodogram) 강도를 보여준다.

도 3b는 모든 cfDNA 샘플 (혈장), 종양 환자 샘플로부터의 cfDNA (종양), 임신한 여성 개체로부터의 cfDNA (임신), 인간 상이한 인간 세포주의 MNase (세포주) 및 인간 DNA 샷건 서열결정 라이브러리 (샷건)에 걸쳐 마지막 (chr22) 인간 상염색체에 지도화되는 판독 개시 좌표의 고속 푸리에 변환 (FFT)으로부터의 평균 주기도 강도를 보여준다.

도 4는 모든 상염색체에 걸친 10 킴로염기 쌍 (kbp) 블록에서 196개 염기쌍 (bp)의 주기도에서 강도의 처음 3개의 주성분 (PC)을 도시한다: 도 4a는 PC 2 대 PC 1을 보여주고, 도 4b는 PC 3 대 PC 2를 보여준다.

도 5는 모든 상염색체에 걸친 10 kbp 블록에서 196 bp의 주기도로 측정된 강도의 유클리드 거리 (Euclidean distance)의 계층적 클러스터링 덴도그램 (dendrogram)을 보여준다.

도 6은 모든 상염색체에 걸친 10 kbp 블록에서 181 bp 내지 202 bp의 주기도에서 강도의 처음 3개의 주성분을 보여준다: 도 6a는 PC 2 대 PC 1을 보여주고, 도 6b는 PC 3 대 PC 2를 보여준다.

도 7은 모든 상염색체에 걸친 10 kbp 블록에서 181 bp 내지 202 bp의 주기도에서 측정된 강도의 유클리드 거리의 계층적 클러스터링 덴도그램을 보여준다.

도 8은 cfDNA 데이터세트에 대한 모든 상염색체에 걸친 10 kbp 블록에서 181 bp 내지 202 bp의 주기도에서 강도의 주성분 분석 (10 PC 중 처음 7개)을 보여준다: 도 8a는 PC 2 대 PC 1을 보여주고, 도 8b는 PC 3 대 PC 2를 보여주고, 도 8c는 PC 4 대 PC 3를 보여주고, 도 8d는 PC 5 대 PC 4를 보여주고, 도 8e는 PC 6 대 PC 5를 보여주고, 도 8f는 PC 7 대 PC 6를 보여준다.

도 9는 MNase 데이터세트에 대한 모든 상염색체에 걸친 10 kbp 블록에서 181 bp 내지 202 bp 주기도에서 강도의 주성분 분석을 보여준다: 도 9a는 PC 2 대 PC 1을 보여주고, 도 9b는 PC 3 대 PC 2를 보여주고, 도 9c는 PC 4 대 PC 3을 보여주고, 도 9d는 PC 5 대 PC 4를 보여주고, 도 9e는 PC 6 대 PC 5를 보여준다.

도 10은 모든 합성 cfDNA 및 MNase 데이터세트 혼합물에 걸친 대표적인 인간 상염색체 (chr11)에 대한 평균 주기도 강도를 보여준다.

도 11은 합성 MNase 데이터세트 혼합물에 대한 모든 상염색체에 걸친 10 kbp 블록에서 181 bp 내지 202 bp 주기도에서 강도의 처음 2개의 주성분을 보여준다.

도 12는 합성 cfDNA 데이터세트 혼합물에 대한 모든 상염색체에 걸친 10 kbp 블록에서 181 bp 내지 202 bp 주기도에서 강도의 처음 2개의 주성분을 보여준다.

도 13은 합성 MNase 및 cfDNA 혼합물 데이터세트에 대한 모든 상염색체에 걸친 10 kbp 블록에서 181 bp 내지 202 bp 주기도에서 강도의 유클리드 거리의 계층적 클러스터링 덴도그램을 보여준다.

도 14는 적어도 100M 판독값을 갖는 샘플 세트에 대해 23,666개의 CTCF 결합 부위 주위의 1 kbp 윈도우에서의 판독 개시 밀도를 보여준다.

도 15는 적어도 100M 판독값을 갖는 샘플 세트에 대해 5,644개의 c-Jun 결합 부위 주위의 1 kbp 윈도우에서의 판독 개시 밀도를 보여준다.

도 16은 적어도 100M 판독값을 갖는 샘플 세트에 대한 4,417개의 NF-YB 결합 부위 주위의 1 kbp 윈도우에 대한 판독 개시 밀도를 보여준다.

도 17은 cfDNA 단편을 생성하는 과정의 개략적 개요를 보여준다. 아폽토시스 및/또는 괴사성 세포 사멸로 인해 천연 염색질이 거의 완전히 소화된다. 일반적으로 히스톤 또는 전사 인자와 결합되는 단백질 결합 DNA 단편은 우선적으로 소화에서 살아남아, 네이키드 DNA가 소실되는 동안 순환계로 방출된다. 단편은 프로테이나제 처리 후 말초 혈액 혈장으로부터 회수될 수 있다. 건강한 개체에서, cfDNA는 주로 골수 및 림프 세포 계통에서 유래되지만, 특정 의학적 병태에서는 하나 이상의 추가의 조직으로부터 유래되어 존재할 수 있다.

도 18은 통상적인 서열결정 라이브러리 제조로 관찰된 cfDNA의 단편 길이를 보여준다. 길이는 쌍 형성된 말단 서열결정 판독의 정렬에서 추정된다. 167개 염기쌍 (bp) (녹색 점선)에서 단편 길이의 재현가능한 피크는 크로마토솜과의 결합과 일치한다. 추가의 피크는 뉴클레오티드 코어 상의 DNA의 나선 피치 (helical pitch)에 대응하는 ~10.4 bp 주기도를 증명한다. 라이브러리 제조 동안 효소의 말단-복구는 5' 및 3' 오버행을 제거하고 실제 절단 부위를 손상시킬 수 있다.

도 19는 종래의 라이브러리에서 167 bp 단편 및 인접 계놈 서열의 디뉴클레오티드 조성을 보여준다. BH01 라이

브러리에서 관찰된 디뉴클레오티드 빈도를 시뮬레이션된 단편으로부터 예상된 빈도와 비교하였다 (절단 및 어댑터 라이케이션 선호 둘 모두에 기인한 종점 편향에 대한 매칭).

도 20은 cfDNA 단편에 대한 단일 가닥 라이브러리 제조 프로토콜의 개략도를 보여준다.

도 21은 단일 가닥 서열결정 라이브러리 제조로 관찰된 cfDNA의 단편 길이를 보여준다. 라이브러리 제조 동안 주형 분자에 대한 효소적 말단 복구는 수행되지 않는다. 50-120 bp의 짧은 단편은 종래의 라이브러리에 비해 매우 풍부하다. ~10.4 bp의 주기도가 유지되는 동안, 그의 위상은 ~3 bp만큼 이동한다.

도 22는 단일 가닥 라이브러리에서 167 bp 단편 및 인접 계놈 서열의 디뉴클레오티드 조성을 보여준다. IH02 라이브러리에서 관찰된 디뉴클레오티드 빈도를 다시 종점 편향에 대한 매칭되는 시뮬레이션된 단편으로부터 유도된 예상된 빈도와 비교하였다. BH01과 IH02 사이의 편향의 배경 수준의 명백한 차이는 실제 라이브러리 (데이터 미제시)보다는 시뮬레이션 사이의 차이와 관련된다.

도 23a는 종래의 프로토콜로 제조된 대표적인 cfDNA 서열결정 라이브러리의 젤 영상을 보여준다.

도 23b는 단일 가닥 프로토콜로 제조된 대표적인 cfDNA 서열결정 라이브러리의 젤 영상을 보여준다.

도 24a는 cfDNA 단편의 모노뉴클레오티드 절단 편향을 보여준다.

도 24b는 cfDNA 단편의 디뉴클레오티드 절단 편향을 보여준다.

도 25는 뉴클레오솜 배치의 추정에 대한 개략적인 개요를 보여준다. 윈도우에 완전히 걸치는 단편의 수로부터 120 bp 윈도우 내의 단편 종점의 수를 차감함으로써 염기당 윈도우 보호 스코어 (windowed protection score, WPS)가 계산된다. 높은 WPS 값은 소화로부터 DNA 보호의 증가를 나타내고, 낮은 값은 DNA가 보호되지 않음을 나타낸다. 피크 콜 (call)은 상승된 WPS의 인접 영역을 확인한다.

도 26은 잘 연구된 알파-위성 어레이에서 강하게 위치된 뉴클레오솜을 보여준다. 12번 염색체 상의 동원체 주변 유전자좌에서 긴 단편 (120 bp 윈도우, 120-180 bp 판독) 또는 짧은 단편 (16 bp 윈도우, 35-80 bp 판독) 빈 (bin)에 대해 샘플 CH01의 적용범위, 단편 종점 및 WPS 값이 표시된다. CH01로부터의 뉴클레오솜 콜 (중간, 파란색 상자)은 유전자좌에 걸쳐 규칙적으로 이격된다. 2개의 공개된 연구로부터의 MNase 소화에 기초한 뉴클레오솜 콜 (중간, 보라색 및 검은색 상자)도 표시된다. 이 유전자좌는 주석 달린 알파-위성 어레이와 중첩된다.

도 27은 9번 염색체 상의 DNase I 과민성 부위 (DHS) 주위에 위치하는 추정된 뉴클레오솜 배치를 보여준다. 샘플 CH01의 적용범위, 단편 종점 및 WPS 값은 길고 짧은 단편 빈에 대해 표시된다. 회색으로 강조 표시된 과민성 영역은 긴 단편 빈에서 감소된 적용범위로 표시된다. DHS에 인접한 CH01 (중간, 파란색 상자)의 뉴클레오솜 콜은 전사 인자를 포함하는 조절 단백질에 대한 개재 서열의 접근가능성과 일치하게 전형적인 인접 쌍보다 더 넓게 떨어져 있다. 이러한 단백질과 관련될 수 있는 더 짧은 단편의 적용범위는 DHS에서 증가하고, 이는 몇몇 주석이 첨부된 전사 인자 결합 부위 (도시되지 않음)와 중첩된다. 공개된 2개의 연구로부터 MNase 소화에 기초한 뉴클레오솜 콜은 도 26에 도시된 바와 같다.

도 28은 본 개시내용의 한 실시양태에 따른 피크 콜링 및 스코어링의 개략도를 보여준다.

도 29는 GC 함량에 의한 CH01 피크 밀도를 보여준다.

도 30은 샘플에 의한 인접 피크 사이의 거리의 히스토그램을 보여준다. 거리는 피크 콜로부터 인접 콜까지 측정된다.

도 31은 샘플 사이의 피크 콜의 비교를 보여준다. 각각의 샘플 쌍에 대해, 피크 수가 적은 샘플의 각각의 피크 콜과 다른 샘플의 가장 가까운 피크 콜 사이의 거리가 계산되고, 빈 크기가 1인 히스토그램으로 가시화된다. 음수는 가장 가까운 피크가 상류임을 보여주고, 양수는 가장 가까운 피크가 하류임을 보여준다.

도 32는 샘플 사이의 피크 콜의 비교를 보여준다. 도 32a는 IH01 대 BH01을 보여주고, 도 32b는 IH02 대 BH01 을 보여주고, 도 32c는 IH02 대 IH01을 보여준다.

도 33a는 실제 대 시뮬레이션된 피크에 대한 뉴클레오솜 스코어를 보여준다.

도 33b는 스코어 빈의 함수로서의 스코어 빈 내의 중앙 피크 오프셋 (좌측 y-축) 및 각각의 스코어 빈 내의 피크의 수 (우측 y-축)를 보여준다.

도 34는 샘플과 매칭된 시뮬레이션 사이의 피크 콜의 비교를 보여주고, 도 34a는 BH01 시뮬레이션 대 BH01 실제를 보여주고, 도 34b는 IH01 시뮬레이션 대 IH01 실제를 보여주고, 도 34c는 IH02 시뮬레이션 대 IH01 실제를 보여준다.

도 35는 인접 피크 사이의 거리, 샘플 CH01을 보여준다. 점선으로 표시된 검은 선은 분포 모드 (185bp)를 보여준다.

도 36은 22,626개의 전사 개시 부위 (TSS) 주위의 집계되고 조정된 윈도우 보호 스코어 (WPS; 120 bp 윈도우)를 보여준다. 각각 및 전사 방향을 조정한 후 TSS는 0 위치에 정렬된다. 집계된 WPS는 중심 TSS에 상대적인 각각의 위치에서 TSS당 WPS를 합하여 실제 데이터와 시뮬레이션된 데이터 모두에 대해 표로 작성된다. 플로팅된 값은 실제와 시뮬레이션된 집계된 WPS 사이의 차이를 나타내고, 아래에서 보다 상세히 설명하는 바와 같이 국부 배경으로 추가 조정된다. 보다 높은 WPS 값은 절단으로부터의 우선적인 보호를 보여준다.

도 37은 22,626개의 개시 코돈 주위의 집계된 조정된 WPS를 보여준다.

도 38은 224,910개의 스플라이스 공여자 부위 주위의 집계된 조정된 WPS를 보여준다.

도 39는 224,910개의 스플라이스 수용자 부위 주위의 집계된 조정된 WPS를 보여준다.

도 40은 실제 데이터, 매치된 시뮬레이션 및 그 차이를 비롯하여 CH01의 데이터를 사용하여 다양한 유전자 특징에 대해 집계된 조정된 WPS를 보여준다.

도 41은 A/B 구획 내의 뉴클레오솜 간격을 보여준다. 중첩되지 않는 100 킬로염기 (kb) 빈 (각각 ~500개의 뉴클레오솜 콜 포함)에서 뉴클레오솜 간격의 중간값은 계놈 전체에 걸쳐 계산된다. 100 kb 분해능과 함께 GM12878에 대한 A/B 구획 예측은 간행된 자료로부터 제시된 것이다. 구획 A는 개방된 염색질과 연결되어 있고, 구획 B는 닫힌 염색질과 연관되어 있다.

도 42는 7번 및 11번 염색체 상의 뉴클레오솜 간격 및 A/B 구획을 보여준다. A/B 세분화 (빨간색 및 파란색 막대)는 주로 염색체 G-밴딩 (표의 기호 (ideogram), 회색 막대)을 요약한 것이다. 중앙 뉴클레오솜 간격 (검은 점)은 100 kb 빈에서 계산되고, A/B 세분화 위에 플로팅된다.

도 43은 긴 (상부) 분획 및 짧은 (하부) 분획에서 93,550개의 CTCF 부위에 대한 집계된 조정된 WPS를 보여준다.

도 44는 CTCF 부위에서 짧은 분획 cfDNA에 대한 집계된 조정된 WPS의 확대도를 보여준다. 밝은 빨간색 막대 (및 플롯 내의 해당 음영)는 알려진 52 bp CTCF 결합 모티프의 위치를 나타낸다. 이 막대의 어두운 빨간색 하위 부분은 FIMO 모티프 검색에 사용된 17 bp 모티프의 위치를 나타낸다.

도 45는 클러스터링된 FIMO 예측 CTCF 부위 (순전히 모티프-기반: 518,632개 부위)로부터 유래된 CTCF 부위 주위에서 계산된 -1 내지 +1 뉴클레오솜 간격, ENCODE ChIP-seq 피크 (93,530개 부위)와 중첩되는 이들 예측의 하위세트 및 19개의 세포주에 걸쳐 활성인 것으로 실험적으로 관찰된 추가의 하위세트 (23,723개의 부위)를 보여준다. CTCF 부위의 최소로 염격한 세트는 계놈 전체 평균 (~190 bp)과 거의 동일한 거리로 대개 분리된다. 그러나, 가장 염격한 조건에서 대부분의 CTCF 부위는 활성 CTCF 결합 및 인접 뉴클레오솜의 위치 변경과 일치하게 훨씬 더 넓은 거리 (~260 bp)로 분리된다.

도 46-48은 뉴클레오솜에 인접한 CTCF 점유 재위치를 보여준다: 도 46은 FIMO에 의해 예측된 518,632개의 CTCF 결합 부위에 대한 3개의 가장 가까운 상류 및 3개의 가장 가까운 하류 피크 콜에 대한 피크간 거리를 보여준다. 도 47은 도 46에서와 같이 FIMO에 의해 예측된 518,632개의 CTCF 결합 부위에 대한 3개의 가장 가까운 상류 및 3개의 가장 가까운 하류 피크 콜에 대한 피크간 거리를 보여주지만, 여기서 동일한 세트의 CTCF 부위는 93,530 개의 부위를 남겨두고 ENCODE ChIP-seq 피크와의 중복을 기준으로 필터링되었다. 도 48은 도 47에서와 같이 FIMO에 의해 예측된 93,530개의 CTCF 결합 부위에 대한 3개의 가장 가까운 상류 및 3개의 가장 가까운 하류 피크 콜에 대한 피크간 거리를 보여주지만, 여기서 CTCF 부위 세트는 19개의 세포주에 걸쳐 실험적으로 관찰된 활성 CTCF 부위의 세트와의 중첩에 기초하여 필터링되어 23,732개의 부위가 남게된다.

도 49는 인접 뉴클레오솜이 넓은 간격 (230-270 bp)을 갖는 추정 CTCF 부위의 하위세트에 대해, 긴 (상부) 및 짧은 (하부) 분획 둘 모두가 CTCF 부위의 점점 더 염격한 하위세트에서 위치의 더 강력한 배치 신호를 나타냄을 보여준다. 컬러 선을 정의하는 키에 대해서는 도 45를 참조한다.

도 50-52는 뉴클레오솜에 인접한 CTCF 점유 재위치를 보여준다: 도 50은 518,632개의 부위에 대한 평균 짧은 분획 WPS (상부 패널) 및 평균 긴 분획 WPS (하부 패널)를 보여주고, 각각의 부위에 대한 인접 +1 및 -1 뉴클레오

솜 콜을 분리하는 염기쌍의 수를 나타내는 거리 빈으로 구분된다. 도 51은 도 50의 518,632개 부위에 대한 평균 짧은 분획 WPS (상부 패널) 및 평균 긴 분획 WPS (하부 패널)를 보여주지만, 여기서 동일한 세트의 CTCF 부위는 ENCODE ChIP-seq 피크와의 중첩에 기초하여 필터링되었다. 도 52는 도 51의 부위들에 대한 평균 짧은 분획 WPS (상부 패널) 및 평균 긴 분획 WPS (하부 패널)를 보여주지만, 여기서 동일한 세트의 부위는 19개의 세포 주에 걸쳐 실험적으로 관찰된 활성 CTCF 부위의 세트와의 중첩에 기초하여 추가로 필터링되었다. 도 50에서 컬러 선을 규정하는 키는 도 51 및 도 52에서도 동일하다.

도 53a-h는 짧은 및 긴 cfDNA 단편으로부터의 전사 인자 결합 부위의 풋 프린트 (footprint)을 보여준다. 클러스터링된 FIMO 결합 부위 예측은 ENCODE ChIP-seq 데이터와 교차되어 추가 요인의 세트에 대한 전사 인자 (TF) 결합 부위의 신뢰할 수 있는 세트를 확보하였다. 생성된 TF 결합 부위의 세트에 인접한 영역에 대한 집계된 조정된 WPS가 cfDNA 단편의 긴 및 짧은 분획 둘 모두에 대해 표시된다. 보다 높은 WPS 값은 각각 뉴클레오솜 또는 TF 점유의 보다 높은 가능성을 나타낸다. 도 53a: AP-2; 도 53b: E2F-2; 도 53c: EBOX-TF; 도 53d: IRF; 도 53e: MYC-MAX; 도 53f: PAX5-2; 도 53g: RUNX-AML; 도 53h: YY1.

도 54는 전사 인자 ETS (210,798개 부위)에 대한 집계된 조정된 WPS를 보여준다. 긴 (상단) 및 짧은 (하단) cfDNA 분획 둘 모두로부터 계산된 WPS가 표시된다. 주위의 뉴클레오솜 (긴 분획)의 조직과 결합 부위 자체 (짧은 분획)에서의 TF 보호와 일치하는 신호가 관찰된다. 추가의 TF에 대한 유사한 분석이 도 53a-h에 도시된다.

도 55는 전사 인자 MAFK (32,159개 부위)에 대한 집계된 조정된 WPS를 보여준다. 긴 (상단) 및 짧은 (하단) cfDNA 분획 둘 모두로부터 계산된 WPS가 표시된다. 주위의 뉴클레오솜 (긴 분획)의 조직과 결합 부위 자체 (짧은 분획)에서의 TF 보호와 일치하는 신호가 관찰된다. 추가의 TF에 대한 유사한 분석이 도 53a-h에 도시된다.

도 56은 DNase 과민성 (DHS) 부위에 기초한 무세포 DNA에 기여하는 세포 유형의 혼합물의 추정을 보여준다. 116개의 다양한 생물학적 샘플로부터 DHS 부위에서의 뉴클레오솜 콜의 피크 대 피크 간격의 빈도 분포는 쌍봉 (bimodal) 분포를 보이고, 두 번째 모드는 개재된 전사 인자 결합으로 인해 활성 DHS 부위에서 넓어진 뉴클레오솜 간격에 해당한다 (~ 190 bp \rightarrow 260 bp). 림프계 또는 골수 샘플에서 확인된 DHS 부위는 건강한 개체에서 cfDNA의 주요 공급원으로서 조혈 세포 사멸과 일치하게, 넓은 뉴클레오솜 간격과 함께 DHS 부위의 가장 큰 비율을 갖는다.

도 57은 전사 개시 부위 (TSS) 주위의 조정된 WPS 스코어를 NB-4 (급성 전골수성 백혈병 세포주)에 대해 규정된 5개의 유전자 발현 빈 (5분위수)으로의 분할이 뉴클레오솜의 간격 및 배치의 차이를 어떻게 나타내는지 보여준다. 고도로 발현된 유전자는 전사체 내에서 뉴클레오솜의 강한 위상을 보여준다. TSS의 상류에서, -1 뉴클레오솜은 발현 빈에 걸쳐 잘 위치하지만, -2 및 -3 뉴클레오솜은 중간 내지 고도로 발현되는 유전자에만 잘 위치한다.

도 58은 중간 내지 고도로 발현된 유전자에 대해, 전사 활성 유전자에서 전사 예비 개시 복합체 또는 그의 일부 성분의 풋 프린트와 일치하여 TSS와 -1 뉴클레오솜 사이에 짧은 단편 피크가 관찰됨을 보여준다.

도 59는 전사체 내의 중간값의 뉴클레오솜 거리가 NB-4 세포주 ($p = -0.17$, $n = 19,677$ 개 유전자)에 대해 측정된 유전자 발현과 음의 상관관계가 있음을 보여준다. 거의 내지 전혀 발현되지 않는 유전자는 193 bp의 중간값의 뉴클레오솜 거리를 나타내지만, 발현된 유전자의 경우에는 이 거리는 186-193 bp이다. 이 음의 상관관계는 보다 정확한 중간값 거리를 결정하기 위해 더 많은 뉴클레오솜 콜이 사용될 때 (예를 들어, 적어도 60개의 뉴클레오솜을 필요로 함, $p = -0.50$, $n = 12,344$ 개 유전자), 더 강하다.

도 60은 복수개의 기여를 디콘볼루션(deconvolution)하기 위해 각각의 TSS에서 시작하는 유전자 본체의 처음 10 kb에 대해 긴 단편 WPS에서의 특정 빈도 기여도 (강도)의 풍부성을 정량하기 위해 고속 푸리에 변환 (FFT)을 사용하는 방법을 보여준다. 76개의 세포주에서 RNA 발현과 상기 강도를 상이한 빈도로 갖는 1차 조직 사이의 상관관계에 대한 궤적이 제시된다. 진한 검은색 선으로 표시된 것은 NB-4 세포주이다. 상관관계는 193-199 bp 빈도 범위에서 강도 규모가 가장 강하다.

도 61은 건강한 상태 및 암에서 무세포 DNA에 기여하는 세포 유형의 추정을 보여준다. 상단 패널은 유형별로 분류되고 최상위 순위 (상단 줄)부터 가장 낮은 순위 (하단 줄)까지 나열된 다양한 cfDNA 라이브러리에 대해 193-199 bp 빈도 범위에서 평균 강도를 갖는 76 RNA 발현 데이터세트에 대한 상관관계의 순위를 보여준다. 상관관계 값 및 전체 세포주 또는 조직 명칭이 표 3에 제시되어 있다. 3개의 모든 건강한 샘플 (BH01, IH01 및 IH02; 처음 3개의 컬럼)에 대한 가장 강한 상관관계는 모두 골수뿐만 아니라 림프구 및 골수 세포주와 관련이 있다. 이와 대조적으로, IV기 암 환자로부터 얻은 cfDNA 샘플 (IC15, IC17, IC20, IC35, IC37; 마지막 5개의

컬럼)은 다양한 암 세포주와 최고의 상관관계를 보이고, 예를 들어 IC17 (간세포 암종, HCC)은 HepG2 (간세포 암종 세포주)와, IC35 (유관 암종, DC)는 MCF7 (전이성 유방 선암종 세포주)와 가장 높은 상관관계를 보인다. 암 샘플에서 관찰된 세포주/조직 순위를 3개의 건강한 각각의 샘플과 비교하고 순위 변화의 평균을 구할 때 (하부 패널), 최대 순위 변화는 3개의 건강한 샘플을 서로 비교하고 순위 변경 평균을 구하여 관찰된 것 ('대조') 보다 2배를 초과하여 더 높다. 예를 들어, IC15 (소세포 폐 암종, SCLC)의 경우 SCLC-21H (소세포 폐 암종 세포주)의 순위는 평균 31 위치가 증가하였고, IC20 (편평세포 폐 암종, SCC)의 경우에는 SK-BR-3 (전이성 유방 선암종 세포주)이 평균 21 순위가 상승하였고, IC37 (결장직장 선암종, AC)에 대해서는 HepG2가 24 순위가 증가하였다.

도 62는 적용범위 (도 62a) 또는 대립유전자 균형 (도 62b)에 기초하여, 순환하는 종양 DNA의 높은 부담을 갖는 샘플을 선택하기 위한 이수성 (aneuploidy)의 정량을 보여준다. 도 62a는 이수성이 없다고 가정한 모의 샘플 (적색 도트)과 비교하여 각각의 샘플 (흑색 도트)에 대한 관찰된 서열결정 판독 횟수 대 예상된 서열결정 판독 횟수에 기초하여 계산된 각각의 염색체에 대한 Z 스코어의 합계를 보여준다. 도 62b는 추가의 서열결정을 위해 선택된 샘플의 하위세트에 대해, 염색체당 평가된 48,800개의 공통 SNP 각각에서의 대립유전자 균형을 보여준다.

도 63은 공개된 뉴클레오솜 콜 세트에 대한 피크 콜의 비교를 보여준다. 도 63a는 CA01의 매치된 시뮬레이션을 포함하여 여기서 생성된 콜뿐만 아니라 3개의 공개된 데이터세트 ([Gaffney et al., 2012], [JS Pedersen et al., 2014] 및 [A Schep et al., 2015])에 걸친 뉴클레오솜 피크 콜 사이의 거리를 보여준다. 이전에 공개된 데이터세트는 아마도 빈약한 샘플링 또는 넓은 콜 범위 때문에, 정규의 ~185 bp 뉴클레오솜 거리에서 하나의 정의된 모드를 나타내지 않는다. 이와 대조적으로, cfDNA의 모든 뉴클레오솜 콜은 하나의 잘 정의된 모드를 보여준다. 매치된 시뮬레이션된 데이터세트는 더 짧은 모드 (166 bp) 및 더 넓은 분포를 갖는다. 또한, 콜을 생성하는 데 사용되는 cfDNA 데이터세트의 적용범위가 높을수록 분포 모드로 표시되는 콜의 비율이 높아진다. 도 63b는 도 63a와 동일한 각각의 세트 목록에 대한 뉴클레오솜의 수를 보여준다. cfDNA 뉴클레오솜 콜은 거의 13M의 뉴클레오솜 피크 콜로 가장 포괄적인 콜 세트를 제시한다. 도 63c는 IH01 cfDNA 샘플에서의 각각의 피크 콜과 이전에 공개된 3개의 데이터세트로부터의 가장 가까운 피크 콜 사이의 거리를 보여준다. 도 63d는 IH02 cfDNA 샘플에서의 각각의 피크 콜과 이전에 공개된 3개의 데이터세트로부터의 가장 가까운 피크 콜 사이의 거리를 보여준다. 도 63e는 BH01 cfDNA 샘플에서의 각각의 피크 콜과 이전에 공개된 3개의 데이터세트로부터의 가장 가까운 피크 콜 사이의 거리를 보여준다. 도 63f는 CH01 cfDNA 샘플에서의 각각의 피크 콜과 이전에 공개된 3개의 데이터세트로부터의 가장 가까운 피크 콜 사이의 거리를 보여준다. 도 63g는 CA01 cfDNA 샘플에서의 각각의 피크 콜과 이전에 공개된 3개의 데이터세트로부터의 가장 가까운 피크 콜 사이의 거리를 보여준다. 음수는 가장 가까운 피크가 상류에 존재함을 나타내고, 양수는 가장 가까운 피크가 하류에 존재함을 나타낸다. cfDNA 적용범위가 증가함에 따라, 이전에 공개된 콜의 보다 높은 비율이 결정된 뉴클레오솜 콜에 더 가깝게 나타난다. 최고의 일치는 문헌 [Gaffney et al., PLoS Genet., vol. 8, e1003036 (2012)] 및 [A Schep et al. (2015)]에서 생성된 콜에서 발견되었다. 도 63h는 각각의 피크 콜과 이전에 공개된 3개의 데이터세트로부터의 가장 가까운 피크 콜 사이의 거리를 보여주지만, 이번에는 CA01의 매치된 시뮬레이션에 대한 거리를 보여준다. 가장 가까운 실제 뉴클레오솜 위치는 문헌 [Gaffney et al., PLoS Genet., vol. 8, e1003036 (2012)] 및 [JS Pedersen et al., Genome Research, vol. 24, pp. 454-466 (2014)] 콜에 대한 시뮬레이션에서의 피크로부터 멀어지는 경향이 있다. 문헌 [A Schep et al., (2015)]에 의해 생성된 콜은 시뮬레이션된 콜과의 일부 중첩을 보인다.

발명을 실시하기 위한 구체적인 내용

[0018]

발명의 상세한 설명

[0019]

본 개시내용은 대상체의 생물학적 샘플에서 무세포 DNA를 생성하는 하나 이상의 조직 및/또는 세포 유형을 결정하는 방법을 제공한다. 일부 실시양태에서, 본 개시내용은 대상체로부터의 생물학적 샘플에서 cfDNA와 연관된 하나 이상의 결정된 조직 및/또는 세포 유형의 함수로서 대상체에서 질환 또는 장애를 확인하는 방법을 제공한다.

[0020]

본 개시내용은 상이한 세포 유형 또는 조직으로부터 유래하는 cfDNA 분자가 다음에 대해 상이하다는 예측에 기초한다: (a) 인간 계놈 내의 임의의 특정 염기쌍이 cfDNA 단편의 말단 (단편화 지점)에 나타날 가능성의 분포; (b) 인간 계놈의 염기쌍의 임의의 쌍이 cfDNA 단편의 한 상의 말단 (즉, 개별적인 cfDNA 분자를 생성하는 연속적인 단편화 점의 쌍)으로 나타날 가능성의 분포; (c) 인간 계놈 내의 임의의 특정 염기쌍이 뉴클레오솜 점유율

차이의 결과로서 cfDNA 단편 (즉, 상대적 적용범위)에 나타날 가능성의 분포. 이것들은 아래에서 분포 (a), (b) 및 (c)로 언급되거나, 또는 집합적으로 "뉴클레오솜 의존성 절단 확률 지도", "절단 접근가능성 지도" 또는 "뉴클레오솜 지도"로 불린다 (도 1). 참고로, 뉴클레오솜 지도는 미크로코쿠스 뉴클레아제 (MNase), DNase, 또는 트랜스포사제와 같은 효소를 사용한 염색질의 단편화에 의해 유래된 단편의 서열결정을 통해 측정되거나, 또는 뉴클레오솜 또는 크로마토솜 (chromatosome)의 경계에서 또는 그 사이에서 게놈 DNA를 우선적으로 단편화하는 동등한 절차를 통해 측정될 수 있다.

[0021] 건강한 개체에서, cfDNA는 혈액 세포, 즉 조혈 계통 세포의 아폽토시스로부터 암도적으로 유래한다. 이들 세포가 프로그램된 아폽토시스를 겪으면, 그의 게놈 DNA는 절단되어 순환계로 방출되고, 뉴클레아제에 의해 계속 분해된다. cfDNA의 길이 분포는 뉴클레오솜 주위에 감겨진 DNA의 나선형 피치에 대응하는 약 10.5 염기쌍 (bp)의 주기로 진동하고, 링커-회합된 모노뉴클레오솜과 회합된 DNA의 길이에 대응하는 167 bp 주위에 뚜렷한 피크를 갖는다 (도 2). 이 증거는 cfDNA와 뉴클레오솜과의 회합이 순환계에서 완전하고 빠른 분해로부터 cfDNA를 보호한다는 것의 가설을 도출하였다. 또 다른 가능성은 길이 분포가 단순히, 뉴클레오솜 배치에 직접적인 영향을 받는 아폽토시스 자체 동안 DNA 절단 패턴으로부터 발생한다는 것이다. 어쨌든, cfDNA의 길이 분포는 cfDNA를 생성하는 단편화 과정이 뉴클레오솜 배치에 영향을 받는다는 명확한 증거를 제시한다.

[0022] 일부 실시양태에서, 본 개시내용은 라이브러리 구축 및 체액으로부터 유래된 cfDNA 또는 미크로코쿠스 뉴클레아제 (MNase), DNase, 또는 트랜스포사제와 같은 효소로 염색질을 단편화하거나 뉴클레오솜 또는 크로마토솜의 경계 또는 그 사이에서 게놈 DNA를 우선적으로 단편화하는 동등한 절차로부터 유래된 DNA의 대규모 병렬 서열결정에 의한 분포 (a), (b) 및/또는 (c)의 측정으로서 뉴클레오솜 지도를 규정한다. 아래에서 설명되는 바와 같이, 이러한 분포는 예를 들어 인접 윈도우에서의 또는 대안적으로, 전사 인자 결합 부위, 유전자 모델 특정부 (예를 들어, 전사 개시 부위 또는 유전자 본체), 위상적으로 회합된 도메인, 조직 발현 데이터 또는 뉴클레오솜 배치의 다른 상관물에 의해 정의된 게놈의 불연속적인 하위세트에서의 주기도를 정량하는 것과 같은 게놈의 다양한 하위세트 내에서의 뉴클레오솜 위치의 주기적 신호를 모으거나 요약하기 위해 "변환"될 수 있다. 또한, 이것은 조직 특이적 데이터에 의해 정의될 수 있다. 예를 들어, 조직 특이적 DNase I 과민성 부위 부근에서 신호를 모으거나 요약할 수 있다.

[0023] 본 개시내용은 혈장-함유 cfDNA 단편으로부터 추정된 생체 내 뉴클레오솜 보호에 대한 치밀한 게놈 전체에 걸친 지도를 제공한다. 건강한 개체의 cfDNA로부터 유래된 CH01 지도는 지도화할 수 있는 인간 참조 게놈의 대다수에 걸쳐있는 뉴클레오솜 보호의 약 13M의 균일하게 이격된 국부 최대치를 포함한다. 피크의 수가 CH01에서 본 질적으로 포화되지만, 다른 품질 기준은 서열결정 깊이의 함수로 계속되었다 (도 33a-b). 따라서, 본 연구 및 다른 연구 ('CA01', 145억 (G) 단편; 700배 적용범위; 13.0M 피크)에 대해 본 발명자들이 지금까지 수행한 거의 모든 cfDNA 서열결정을 기반으로 하는, 동일한 방법에 의해 추가의 게놈 전체에 걸친 뉴클레오솜 지도가 구축되었다. 이 지도가 훨씬 더 균일한 간격 및 더 높게 지지된 피크 콜을 보였음에도 불구하고 (도 33a-b, 63a-h), 본 발명자들은 이것이 건강한 개체와 그렇지 않은 개체 모두로부터의 cfDNA를 기반으로 한다고 주의를 촉구한다 (표 1, 5).

[0024] 본원에 개시된 뉴클레오솜 보호의 치밀한 게놈 전체에 걸친 지도는 사람 참조 게놈의 지도화가능한 부분의 포화에 접근하고, 뉴클레오솜 배치 또는 보호의 인간 게놈 전체에 걸친 지도를 작성하기 위한 이전의 노력보다 상당히 더 균일하고 예측된 뉴클레오솜 반복 길이와 일치하는 피크 대 피크 간격을 제시한다 (도 63a-h). 거의 모든 이전의 노력과는 대조적으로, 본원에서 관찰된 단편은 내인성 생리적 과정에 의해 생성되고, 따라서 시험관내 미크로코쿠스 뉴클레아제 소화와 관련된 기술적 변이에 적용될 가능성이 작다. 이 참조 지도에서 고려되는 cfDNA를 일으키는 세포 유형은 필연적으로 이질적이다 (예를 들어, 건강한 개체의 립프구와 골수양 세포 유형의 혼합물). 그럼에도 불구하고, 지도의 상대적 완전성은 인간 세포에서의 뉴클레오솜 배치 및 간격, 및 후성 유전 조절, 전사 산물 및 핵 구조와 뉴클레오솜의 상호 작용을 지시하는 과정에 대한 더 깊은 이해를 촉진할 수 있다.

대상체의 생물학적 샘플에서 cfDNA의 공급원(들)을 결정하는 방법

[0025] 일반적으로 상기 논의된 바와 같이, 및 하기 실시예에서 보다 구체적으로 입증되는 바와 같이, 본 발명의 기술은 대상체의 생물학적 샘플에서 cfDNA에 기여하는 조직(들) 및/또는 세포 유형(들)을 결정 (예를 들어, 예측)하기 위해 사용될 수 있다.

[0026] 따라서, 일부 실시양태에서, 본 개시내용은 대상체에서 무세포 DNA (cfDNA)를 생성하는 조직 및/또는 세포 유형을 결정하는 방법을 제공하고, 이 방법은 대상체로부터의 생물학적 샘플로부터 cfDNA를 단리하고, 단리된 cfDNA

는 복수개의 cfDNA 단편을 포함하는 것인 단계; 복수개의 cfDNA 단편의 적어도 일부분과 연관된 서열을 결정하는 단계; cfDNA 단편 서열의 함수로서 복수개의 cfDNA 단편의 적어도 일부의 cfDNA 단편 종점에 대한 참조 게놈 내의 게놈 위치를 결정하는 단계; 및 cfDNA 단편 종점의 적어도 일부의 게놈 위치의 함수로서 cfDNA 단편을 생성하는 조직 및/또는 세포 유형의 적어도 일부를 결정하는 단계를 포함한다.

[0028] 일부 실시양태에서, 생물학적 샘플은 전혈, 말초 혈액 혈장, 소변 또는 뇌 척수액을 포함하거나, 이로 본질적으로 이루어지거나, 또는 이로 이루어진다.

[0029] 일부 실시양태에서, cfDNA 단편을 생성하는 조직 및/또는 세포 유형의 적어도 일부를 결정하는 단계는 cfDNA 단편 종점의 적어도 일부의 게놈 위치, 또는 그의 분포의 수학적 변환을 하나 이상의 참조 지도와 비교하는 것을 포함한다. 본원에서 사용된 용어 "참조 지도"는 cfDNA 서열이 정렬된 게놈 (예를 들어, 참조 게놈) 내의 좌표의 함수로서 대상체의 생물학적 샘플에서 cfDNA의 특징과 상관되거나 비교될 수 있는 임의의 유형 또는 형태의 데이터를 나타낸다. 참조 지도는 임의의 적합한 수단에 의해 대상체의 생물학적 샘플에서 cfDNA의 특징과 상관되거나 비교될 수 있다. 비제한적인 예를 들어, 상관관계 또는 비교는 직접적으로 또는 참조 게놈 내의 원도우에 걸친 그의 분포에 대한 수학적 변환을 수행한 후, 수치 또는 참조 지도에 의해 참조 게놈의 등가 좌표에 대해 정의된 임의의 다른 상태의 측면에서 대상체의 생물학적 샘플에서 cfDNA 종점의 빈도를 분석하여 달성될 수 있다. 또 다른 비제한적인 예에서, 상관관계 또는 비교는 결정된 뉴클레오솜 간격(들)에 비추어 대상체의 생물학적 샘플의 cfDNA에 기초하여 결정된 뉴클레오솜 간격(들), 또는 참조 지도에서 뉴클레오솜 간격(들)과 상관관계가 있는 또 다른 특성을 분석하여 달성될 수 있다.

[0030] 참조 지도(들)은 예를 들어 게놈 정보의 공개 데이터베이스, 공개된 데이터, 또는 공통 특징 (예를 들어, 질환 상태)을 각각 가질 수 있는 참조 대상체의 특정 집단에 대해 생성된 데이터를 포함하는 임의의 적합한 데이터 공급원으로부터 공급되거나 유래될 수 있다. 일부 실시양태에서, 참조 지도는 DNase I 과민성 데이터세트를 포함한다. 일부 실시양태에서, 참조 지도는 RNA 발현 데이터세트를 포함한다. 일부 실시양태에서, 참조 지도는 염색체 입체형태 지도를 포함한다. 일부 실시양태에서, 참조 지도는 염색질 접근가능성 지도를 포함한다. 일부 실시양태에서, 참조 지도는 질환 또는 장애와 연관된 적어도 하나의 조직 또는 세포 유형으로부터 생성된 데이터를 포함한다. 일부 실시양태에서, 참조 지도는 조직 또는 세포 유형에서 뉴클레오솜 및/또는 크로마토솜의 위치를 포함한다. 일부 실시양태에서, 참조 지도는 외인성 뉴클레아제 (예를 들어, 미크로코쿠스 뉴클레아제)로 염색질을 소화시키는 것을 포함하는 절차에 의해 작성된다. 일부 실시양태에서, 참조 지도는 전위 (transposition) 기반 방법 (예를 들어, ATAC-seq)에 의해 결정된 염색질 접근가능성 데이터를 포함한다. 일부 실시양태에서, 참조 지도는 조직 또는 세포 유형에 대한 DNA 결합 및/또는 DNA 점유 단백질의 위치와 연관된 데이터를 포함한다. 일부 실시양태에 있어서, DNA 결합 및/또는 DNA 점유 단백질은 전사 인자이다. 일부 실시양태에서, 위치는 가교된 DNA-단백질 복합체의 염색질 면역침전을 포함하는 절차에 의해 결정된다. 일부 실시양태에서, 위치는 조직 또는 세포 유형과 연관된 DNA를 뉴클레아제 (예를 들어, DNase-I)로 처리하는 것을 포함하는 절차에 의해 결정된다. 일부 실시양태에서, 참조 지도는 알려진 질환을 가진 하나 이상의 개체로부터의 생물학적 샘플로부터의 cfDNA 단편의 서열결정에 의해 작성된다. 일부 실시양태에서, 참조 지도가 그로부터 작성되는 상기 생물학적 샘플은 인간 세포 또는 조직이 이종이식된 동물로부터 수집된다.

[0031] 일부 실시양태에서, 참조 지도는 조직 또는 세포 유형에 대한 DNA 결합 또는 DNA 점유 단백질의 위치에 대응하는 생물학적 특징을 포함한다. 일부 실시양태에서, 참조 지도는 하나 이상의 유전자의 정량적 RNA 발현에 대응하는 생물학적 특징을 포함한다. 일부 실시양태에서, 참조 지도는 하나 이상의 히스톤 마크의 존재 또는 부재에 대응하는 생물학적 특징을 포함한다. 일부 실시양태에서, 참조 지도는 뉴클레아제 절단에 대한 과민성에 대응하는 생물학적 특징을 포함한다.

[0032] cfDNA 단편 종점의 적어도 일부의 게놈 위치를 하나 이상의 참조 지도와 비교하는 단계는 다양한 방법으로 수행될 수 있다. 일부 실시양태에서, 생물학적 샘플로부터 생성된 cfDNA 데이터 (예를 들어, cfDNA 단편의 게놈 위치, 그의 종점, 그의 종점의 빈도 및/또는 그의 분포로부터 추정된 뉴클레오솜 간격(들))는 하나 초과의 참조 지도에 비교된다. 이러한 실시양태에서, 생물학적 샘플에서의 cfDNA 데이터와 가장 높은 상관관계를 갖는 참조 지도와 관련된 조직 또는 세포 유형은 기여하는 것으로 간주된다. 비제한적인 예를 들어, cfDNA 데이터가 유사한 cfDNA 종점 목록 및 그의 참조 게놈 내의 위치를 포함하면, cfDNA 종점 및 그의 참조 게놈 내의 위치의 가장 유사한 목록을 갖는 참조 지도(들)는 기여하는 것으로 간주될 수 있다. 또 다른 비제한적인 예로서, 생물학적 샘플로부터의 cfDNA 단편 종점의 분포의 수학적 변환과 가장 높은 상관관계 (또는 건강한 대상체로부터의 cfDNA에 비해 증가된 상관관계)를 갖는 참조 지도(들)는 기여하는 것으로 간주될 수 있다. 참조 지도에 대응하는 조

직 유형 및/또는 세포 유형은 생물학적 샘플에서 단리된 cfDNA의 잠재적인 공급원으로 간주된다.

[0033] 일부 실시양태에서, cfDNA 단편을 생성하는 조직 및/또는 세포 유형의 적어도 일부를 결정하는 단계는 cfDNA 단편 종점의 적어도 일부의 계놈 위치의 분포에 대해 수학적 변환을 수행하는 것을 포함한다. 본 발명의 기술과 관련하여 사용하기에 적합한 수학적 변환의 비제한적인 하나의 예는 고속 푸리에 변환 ("FFT")과 같은 푸리에 변환이다.

[0034] 일부 실시양태에 있어서, 상기 방법은 참조 계놈의 적어도 일부의 좌표 각각에 대한 스코어를 결정하는 단계를 추가로 포함하고, 여기서 스코어는 적어도 복수개의 cfDNA 단편 종점 및 그의 계놈 위치의 함수로서 결정되고, 관찰된 cfDNA 단편을 생성하는 조직 및/또는 세포 유형의 적어도 일부를 결정하는 단계는 스코어를 하나 이상의 참조 지도와 비교하는 단계를 포함한다. 스코어는 참조 계놈의 좌표에 상대적 또는 절대적 값을 할당하기 위해 사용될 수 있는 임의의 기준 (예를 들어, 숫자 랭킹 또는 확률)일 수 있다. 예를 들어, 스코어는 좌표가 cfDNA 단편 종점의 위치를 나타내는 확률 또는 좌표가 뉴클레오솜 또는 단백질 결합에 의해 뉴클레아제 절단으로부터 우선적으로 보호되는 계놈의 위치를 나타낼 확률과 같은 확률로 이루어지거나 이들 확률에 관련될 수 있다. 또 다른 예로서, 스코어는 계놈의 특정 영역 내에서 cfDNA 단편 종점 분포의 수학적 변환에 의해 결정되는 바와 같이, 상기 특정 영역에서 뉴클레오솜 간격과 관련될 수 있다. 이러한 스코어는 해당 특정 좌표와 관련된 절대적 또는 상대적 이벤트 (예를 들어, cfDNA 단편 종점의 수)를 계수하거나 또는 영역 또는 계놈 좌표에서 상기 계수의 값에 대한 수학적 변환을 수행하는 것을 포함한 임의의 적합한 방법에 의해 좌표에 할당될 수 있다. 일부 실시양태에서, 좌표에 대한 스코어는 좌표가 cfDNA 단편 종점의 위치일 확률과 관련된다. 다른 실시양태에서, 좌표에 대한 스코어는 좌표가 뉴클레오솜 또는 단백질 결합에 의해 뉴클레아제 절단으로부터 우선적으로 보호되는 계놈의 위치를 나타낼 확률과 관련된다. 일부 실시양태에서, 스코어는 좌표의 계놈 영역에서 뉴클레오솜 간격과 관련된다.

[0035] 본원에서 설명되는 방법에서 언급된 조직(들) 및/또는 세포 유형(들)은 cfDNA를 생성하는 임의의 조직 또는 세포 유형일 수 있다. 일부 실시양태에서, 조직 또는 세포 유형은 질환 또는 장애를 갖는 대상체로부터의 1차 조직이다. 일부 실시양태에서, 질환 또는 장애는 암, 정상 임신, 임신 합병증 (예를 들어, 이수성 임신), 심근경색, 염증성 장 질환, 전신성 자가면역 질환, 국부 자가면역 질환, 거부 반응을 수반하는 동종이식, 거부 반응을 수반하지 않는 동종이식, 출중, 및 국부 조직 손상으로 이루어진 군으로부터 선택된다.

[0036] 일부 실시양태에서, 조직 또는 세포 유형은 건강한 대상체로부터의 1차 조직이다.

[0037] 일부 실시양태에서, 조직 또는 세포 유형은 불멸화된 세포주이다.

[0038] 일부 실시양태에서, 조직 또는 세포 유형은 종양으로부터의 생검이다.

[0039] 일부 실시양태에서, 참조 지도는 적어도 하나의 참조 대상체로부터 얻은 샘플로부터 얻은 서열 데이터에 기초한다. 일부 실시양태에서, 이 서열 데이터는 예를 들어 알려진 질환을 갖는 대상체(들)로부터의 cfDNA를 서열 결정하여 참조 지도가 작성되는 경우, 참조 계놈 내의 cfDNA 단편 종점의 위치를 규정한다. 다른 실시양태에서, 참조 지도가 기초로 하는 상기 서열 데이터는 DNase I 과민성 부위 데이터세트, RNA 발현 데이터세트, 염색체 입체형태 지도, 또는 염색질 접근가능성 지도, 또는 미크로코糗스 뉴클레아제의 염색질 소화에 의해 작성된 뉴클레오솜 배치 지도 중 어느 하나 이상을 포함할 수 있다.

[0040] 일부 실시양태에서, 참조 대상체는 건강하다. 일부 실시양태에서, 참조 대상체는 임의적으로 암, 정상 임신, 임신 합병증 (예를 들어, 이수성 임신), 심근경색, 염증성 장 질환, 전신성 자가면역 질환, 국부 자가면역 질환, 거부 반응을 수반하는 동종이식, 거부 반응을 수반하지 않는 동종이식, 출중 및 국부 조직 손상으로 이루어진 군으로부터 선택되는 질환 또는 장애를 갖는다.

[0041] 일부 실시양태에서, 참조 지도는 조직 또는 세포 유형과 연관된 참조 계놈의 좌표의 적어도 일부분에 대한 스코어를 포함한다. 일부 실시양태에서, 참조 지도는 스코어의 푸리에 변환과 같은 스코어의 수학적 변환을 포함한다. 일부 실시양태에서, 스코어는 조직 또는 세포 유형에 대한 참조 계놈 좌표의 주석에 기초한다. 일부 실시양태에서, 스코어는 뉴클레오솜 및/또는 크로마토솜의 위치에 기초한다. 일부 실시양태에서, 스코어는 전사 개시 부위 및/또는 전사 종결 부위에 기초한다. 일부 실시양태에서, 스코어는 적어도 하나의 전사 인자의 예측된 결합 부위에 기초한다. 일부 실시양태에서, 스코어는 예측된 뉴클레아제 과민성 부위에 기초한다. 일부 실시양태에서, 스코어는 예측된 뉴클레오솜 간격에 기초한다.

[0042] 일부 실시양태에서, 스코어는 적어도 하나의 직교 (orthogonal) 생물학적 특징과 연관된다. 일부 실시양태에서, 직교 생물학적 특징은 고도로 발현된 유전자와 관련된다. 일부 실시양태에서, 직교 생물학적 특

정은 저 발현 유전자와 관련된다.

[0043] 일부 실시양태에서, 복수개의 스코어 중 적어도 일부는 역치 (최소) 값 초과의 값을 갖는다. 그러한 실시양태에서, 역치 (최소) 값 미만의 스코어는 스코어를 참조 지도와 비교하는 단계로부터 배제된다. 일부 실시양태에서, 역치 값은 cfDNA를 생성하는 조직(들) 및/또는 세포 유형(들)을 결정하기 전에 결정된다. 다른 실시양태에서, 역치 값은 cfDNA를 생성하는 조직(들) 및/또는 세포 유형(들)을 결정한 후에 결정된다.

[0044] 일부 실시양태에서, cfDNA 단편 종점의 적어도 일부의 복수개의 계놈 위치의 함수로서 cfDNA를 생성하는 조직 및/또는 세포 유형을 결정하는 단계는 샘플의 cfDNA 단편 종점의 적어도 일부의 계놈 위치 분포의 수학적 변환을 하나 이상의 참조 지도의 하나 이상의 특징과 비교하는 것을 포함한다. 이 목적으로 적합한 수학적 변환의 비제한적인 하나의 예는 고속 푸리에 변환 ("FFT")과 같은 푸리에 변환이다.

[0045] 본원에서 설명되는 임의의 실시양태에서, 상기 방법은 단리된 cfDNA를 생성하는 결정된 조직 및/또는 세포 유형의 목록을 포함하는 보고서를 작성하는 단계를 추가로 포함할 수 있다. 보고서는 임의적으로 샘플 및/또는 대상체, 생물학적 샘플의 유형, 생물학적 샘플이 대상체로부터 획득된 날짜, cfDNA 단리 단계가 수행된 날짜 및/또는 생물학적 샘플로부터 단리된 임의의 cfDNA를 생성하지 않을 가능성이 있는 조직(들) 및/또는 세포 유형(들)에 대한 임의의 다른 정보를 추가로 포함할 수 있다.

[0046] 일부 실시양태에서, 보고서는 비제한적인 예를 들어, 대상체로부터 추가의 진단 시험의 제안, 치료 요법 개시 제안, 기준의 치료 요법의 변경 제안 및/또는 기준의 치료 요법의 연기 또는 중지 제안을 포함하는 권고된 치료 프로토콜을 추가로 포함한다.

대상체에서 질환 또는 장애를 확인하는 방법

[0048] 일반적으로 상기 논의되고, 하기 실시예에서 보다 구체적으로 입증되는 바와 같이, 본 발명의 기술은 적어도 부분적으로는 대상체의 생물학적 샘플에서 cfDNA에 기여하는 조직 및/또는 세포 유형에 기초로 하여, 질환 또는 장애, 또는 질환 또는 장애의 부재를 결정 (예를 들어, 예측)하기 위해 사용될 수 있다.

[0049] 따라서, 일부 실시양태에서, 본 개시내용은 대상체에서 질환 또는 장애를 확인하는 방법을 제공하고, 상기 방법은 대상체로부터의 생물학적 샘플로부터 무세포 DNA (cfDNA)를 단리하고, 단리된 cfDNA는 복수개의 cfDNA 단편을 포함하는 것인 단계; 복수개의 cfDNA 단편의 적어도 일부분과 연관된 서열을 결정하는 단계; cfDNA 단편 서열의 함수로서 복수개의 cfDNA 단편의 적어도 일부의 cfDNA 단편 종점에 대한 참조 계놈 내의 계놈 위치를 결정하는 단계; cfDNA 단편 종점의 적어도 일부의 계놈 위치의 함수로서 cfDNA를 생성하는 조직 및/또는 세포 유형의 적어도 일부를 결정하는 단계; 및 상기 질환 또는 장애를 cfDNA를 생성하는 결정된 조직 및/또는 세포 유형의 함수로서 확인하는 단계를 포함한다.

[0050] 일부 실시양태에서, 생물학적 샘플은 전혈, 말초 혈액 혈장, 소변 또는 뇌 척수액을 포함하거나, 이로 본질적으로 이루어지거나, 또는 이로 이루어진다.

[0051] 일부 실시양태에서, cfDNA를 생성하는 조직 및/또는 세포 유형을 결정하는 단계는 cfDNA 단편 종점의 적어도 일부의 계놈 위치 또는 그의 분포의 수학적 변환을 하나 이상의 참조 지도에 비교하는 것을 포함한다. 이들 실시양태와 관련하여 사용되는 용어 "참조 지도"는 대상체의 생물학적 샘플에서 cfDNA를 생성하는 조직(들) 및/또는 세포 유형(들)을 결정하는 방법과 관련하여 상기 설명한 것과 동일한 의미를 가질 수 있다. 일부 실시양태에서, 참조 지도는 DNase I 과민성 부위 데이터세트, RNA 발현 데이터세트, 염색체 입체형태 지도, 염색질 접근가능성 지도, 적어도 하나의 참조 대상체로부터 얻은 샘플로부터 생성된 서열 데이터, 질환 또는 장애와 연관된 적어도 하나의 조직에 대응하는 효소 매개된 단편화 데이터, 및/또는 조직 또는 세포 유형 내의 뉴클레오솜 및/또는 크로마토솜의 위치 중 어느 하나 이상을 포함할 수 있다. 일부 실시양태에서, 참조 지도는 알려진 질환을 갖는 하나 이상의 개체로부터의 생물학적 샘플로부터의 cfDNA 단편의 서열결정에 의해 작성된다. 일부 실시양태에서, 참조 지도가 그로부터 작성되는 상기 생물학적 샘플은 인간 세포 또는 조직이 이식된 동물로부터 수집된다.

[0052] 일부 실시양태에서, 참조 지도는 외인성 뉴클레아제 (예를 들어, 미크로코쿠스 뉴클레아제)로 염색질을 소화시켜 작성된다. 일부 실시양태에서, 참조 지도는 전위 기반 방법 (예를 들어, ATAC-seq)에 의해 결정된 염색질 접근가능성 데이터를 포함한다. 일부 실시양태에서, 참조 지도는 조직 또는 세포 유형에 대한 DNA 결합 및/또는 DNA 점유 단백질의 위치와 연관된 데이터를 포함한다. 일부 실시양태에서, DNA 결합 및/또는 DNA 점유 단백질은 전사 인자이다. 일부 실시양태에서, 위치는 가교된 DNA-단백질 복합체의 염색질 면역침전에 의해 결정된다. 일부 실시양태에서, 위치는 조직 또는 세포 유형과 연관된 DNA를 뉴클레아제 (예를 들어, DNase-I)로 처리

하여 결정된다.

[0053] 일부 실시양태에서, 참조 지도는 조직 또는 세포 유형에 대한 DNA 결합 또는 DNA 접유 단백질의 위치에 대응하는 생물학적 특징을 포함한다. 일부 실시양태에서, 참조 지도는 하나 이상의 유전자의 정량적 발현에 대응하는 생물학적 특징을 포함한다. 일부 실시양태에서, 참조 지도는 하나 이상의 히스톤 마크의 존재 또는 부재에 대응하는 생물학적 특징을 포함한다. 일부 실시양태에서, 참조 지도는 뉴클레아제 절단에 대한 과민성에 대응하는 생물학적 특징을 포함한다.

[0054] 일부 실시양태에서, cfDNA를 생성하는 조직 및/또는 세포 유형을 결정하는 단계는 복수개의 cfDNA 단편 종점의 적어도 일부의 계놈 위치의 분포에 대해 수학적 변환을 수행하는 것을 포함한다. 일부 실시양태에서, 수학적 변환은 푸리에 변환을 포함한다.

[0055] 일부 실시양태에서, 상기 방법은 참조 계놈의 적어도 일부의 좌표 각각에 대한 스코어를 결정하는 단계를 추가로 포함하고, 여기서 스코어는 적어도 복수개의 cfDNA 단편 종점 및 그의 계놈 위치의 함수로서 결정되고, 관찰된 cfDNA 단편을 생성하는 조직 및/또는 세포 유형의 적어도 일부를 결정하는 단계는 스코어를 하나 이상의 참조 지도와 비교하는 단계를 포함한다. 스코어는 참조 계놈의 좌표에 상대적 또는 절대적 값을 할당하기 위해 사용될 수 있는 임의의 기준 (예를 들어, 숫자 랭킹 또는 확률)일 수 있다. 예를 들어, 스코어는 좌표가 cfDNA 단편 종점의 위치를 나타낼 확률 또는 좌표가 뉴클레오솜 또는 단백질 결합에 의해 뉴클레아제 절단으로부터 우선적으로 보호되는 계놈의 위치를 나타낼 확률과 같은 확률로 이루어지거나 이를 확률과 관련될 수 있다. 또 다른 예로서, 스코어는 계놈의 특정 영역 내에서 cfDNA 단편 종점 분포의 수학적 변환에 의해 결정된 바와 같이 상기 특정 영역에서 뉴클레오솜 간격과 관련될 수 있다. 이러한 스코어는 예를 들어 해당 특정 좌표와 관련된 절대적 또는 상대적 이벤트 (예를 들어, cfDNA 단편 종점의 수)를 계수하거나 또는 영역 또는 계놈 좌표에서 상기 계수의 값에 대한 수학적 변환을 수행하는 것을 포함한 임의의 적합한 방법에 의해 좌표에 할당될 수 있다. 일부 실시양태에서, 좌표에 대한 스코어는 좌표가 cfDNA 단편 종점의 위치일 확률과 관련된다. 다른 실시양태에서, 좌표에 대한 스코어는 좌표가 뉴클레오솜 또는 단백질 결합에 의해 뉴클레아제 절단으로부터 우선적으로 보호되는 계놈의 위치를 나타낼 확률과 관련된다. 일부 실시양태에서, 스코어는 좌표의 계놈 영역에서 뉴클레오솜 간격과 관련된다.

[0056] 이들 실시양태와 관련하여 사용되는 용어 "스코어"는 대상체의 생물학적 샘플에서 cfDNA를 생성하는 조직(들) 및/또는 세포 유형(들)을 결정하는 방법과 관련하여 상기 설명된 것과 동일한 의미를 가질 수 있다. 일부 실시양태에서, 좌표에 대한 스코어는 좌표가 cfDNA 단편 종점의 위치일 확률과 관련된다. 다른 실시양태에서, 좌표에 대한 스코어는 좌표가 뉴클레오솜 또는 단백질 결합에 의해 뉴클레아제 절단으로부터 우선적으로 보호되는 계놈의 위치를 나타낼 확률과 관련된다. 일부 실시양태에서, 스코어는 좌표의 계놈 영역에서 뉴클레오솜 간격과 관련된다.

[0057] 일부 실시양태에서, 참조 지도를 작성하기 위해 사용된 조직 또는 세포 유형은 질환 또는 장애를 갖는 대상체로부터의 1차 조직이다. 일부 실시양태에서, 질환 또는 장애는 암, 정상 임신, 임신 합병증 (예를 들어, 이수성 임신), 심근경색, 전신성 자가면역 질환, 국부 자가면역 질환, 염증성 장 질환, 거부 반응을 수반하는 동종이식, 거부 반응을 수반하지 않는 동종이식, 출중, 및 국부 조직 손상으로 이루어진 군으로부터 선택된다.

[0058] 일부 실시양태에서, 조직 또는 세포 유형은 건강한 대상체로부터의 1차 조직이다.

[0059] 일부 실시양태에서, 조직 또는 세포 유형은 불멸화된 세포주이다.

[0060] 일부 실시양태에서, 조직 또는 세포 유형은 종양으로부터의 생검이다.

[0061] 일부 실시양태에서, 참조 지도는 적어도 하나의 참조 대상체로부터 얻은 샘플로부터 수득된 서열 데이터에 기초한다. 일부 실시양태에서, 이 서열 데이터는 예를 들어, 알려진 질환을 갖는 대상체(들)로부터의 cfDNA의 서열 결정에 의해 참조 지도가 작성되는 경우, 참조 계놈 내에서의 cfDNA 단편 종점의 위치를 정의한다. 다른 실시양태에서, 참조 지도가 그를 기초로 하는 상기 서열 데이터는 DNase I 과민성 부위 데이터세트, RNA 발현 데이터세트, 염색체 입체형태 지도, 또는 염색질 접근가능성 지도, 또는 미크로코쿠스 뉴클레아제에 의한 소화에 의해 작성된 뉴클레오솜 배치 지도 중 어느 하나 이상을 포함할 수 있다. 일부 실시양태에서, 참조 대상체는 건강하다. 일부 실시양태에서, 참조 대상체는 질환 또는 장애를 갖는다. 일부 실시양태에서, 질환 또는 장애는 암, 정상 임신, 임신 합병증 (예를 들어, 이수성 임신), 심근경색, 전신성 자가면역 질환, 염증성 장 질환, 국부 자가면역 질환, 거부 반응을 수반하는 동종이식, 거부 반응을 수반하지 않는 동종이식, 출중, 및 국부 조직 손상으로 이루어진 군으로부터 선택된다.

[0062] 일부 실시양태에서, 참조 지도는 조직 또는 세포 유형과 연관된 참조 게놈의 적어도 일부분에 대해, cfDNA 단편 종점 확률 또는 상기 확률과 상관되는 양을 포함한다. 일부 실시양태에서, 참조 지도는 cfDNA 단편 종점 확률의 수학적 변환 또는 그러한 확률과 상관되는 양을 포함한다.

[0063] 일부 실시양태에서, 참조 지도는 조직 또는 세포 유형과 연관된 참조 게놈의 좌표의 적어도 일부분에 대한 스코어를 포함한다. 일부 실시양태에서, 참조 지도는 스코어의 푸리에 변환과 같은 스코어의 수학적 변환을 포함한다. 일부 실시양태에서, 스코어는 조직 또는 세포 유형에 대한 참조 게놈 좌표의 주석에 기초한다. 일부 실시양태에서, 스코어는 뉴클레오솜 및/또는 크로마토솜의 위치에 기초한다. 일부 실시양태에서, 스코어는 전사 개시 부위 및/또는 전사 종결 부위에 기초한다. 일부 실시양태에서, 스코어는 적어도 하나의 전사 인자의 예측된 결합 부위에 기초한다. 일부 실시양태에서, 스코어는 예측된 뉴클레아제 과민성 부위에 기초한다.

[0064] 일부 실시양태에서, 스코어는 적어도 하나의 직교 생물학적 특징과 관련된다. 일부 실시양태에서, 직교 생물학적 특징은 고도로 발현된 유전자와 관련된다. 일부 실시양태에서, 직교 생물학적 특징은 저 발현 유전자와 관련된다.

[0065] 일부 실시양태에서, 복수개의 스코어 중 적어도 일부는 각각 역치 값 초과의 스코어를 갖는다. 그러한 실시양태에서, 역치 (최소) 값 미만의 스코어는 스코어를 참조 지도와 비교하는 단계로부터 배제된다. 일부 실시양태에서, 역치 값은 cfDNA를 생성하는 조직(들) 및/또는 세포 유형(들)을 결정하기 전에 결정된다. 다른 실시양태에서, 역치 값은 cfDNA를 생성하는 조직(들) 및/또는 세포 유형(들)을 결정한 후에 결정된다.

[0066] 일부 실시양태에서, cfDNA 단편 종점의 적어도 일부의 복수개의 게놈 위치의 함수로서 cfDNA를 생성하는 조직 및/또는 세포 유형을 결정하는 단계는 하나 이상의 참조 지도의 하나 이상의 특징을 갖는 샘플의 cfDNA 단편 종점의 적어도 일부의 게놈 위치 분포의 수학적 변환을 포함한다.

[0067] 일부 실시양태에서, 이 수학적 변환은 푸리에 변환을 포함한다.

[0068] 일부 실시양태에서, 참조 지도는 질환 또는 장애와 연관된 적어도 하나의 조직에 대응하는 효소 매개된 단편화 데이터를 포함한다.

[0069] 일부 실시양태에서, 참조 게놈은 인간과 관련된다.

[0070] 본 발명의 한 측면에서, 본원에서 설명되는 방법은 체액 중의 cfDNA의 분석으로부터의 악성 종양의 검출, 모니터링 및 조직(들) 및/또는 기원 세포 유형(들)의 평가에 사용된다. 악성 종양 환자에서 순환 혈장과 같은 체액 내의 cfDNA의 일부 부분이 종양으로부터 유래될 수 있음이 현재 잘 입증되었다. 본원에서 설명되는 방법은 잠재적으로 상기 종양 유래 부분을 검출하고 정량하기 위해 사용할 수 있다. 또한, 뉴클레오솜 점유 지도는 세포 유형 특이적이기 때문에, 본원에서 설명되는 방법은 잠재적으로 악성 종양의 조직(들) 및/또는 기원 세포 유형(들)을 결정하기 위해 사용될 수 있다. 또한, 상기 언급한 바와 같이, 암에서 순환 혈장 cfDNA의 농도가 크게 증가하고 종양 자체로부터의 기여도에 불균형이 있을 수 있다는 것이 관찰되었다. 이것은 다른 조직 (예를 들어, 간질, 면역계)이 암 동안 순환 혈장 cfDNA에 기여할 가능성이 있음을 시사한다. cfDNA에 대한 상기 다른 조직으로부터의 기여도가 제시된 다른 유형의 암에 대한 환자 사이에서 일관성이 있는 정도로, 위에서 설명된 방법을 사용하여 암세포 그 자체가 아닌 상기 다른 조직으로부터의 신호를 기초로 하여 암 검출, 모니터링 및/또는 조직(들) 및/또는 기원 세포 유형(들) 평가를 실시할 수 있다.

[0071] 본 발명의 또 다른 측면에서, 본원에서 설명되는 방법은 체액 중의 cfDNA의 분석으로부터 조직 손상의 검출, 모니터링 및 기원 조직(들) 및/또는 세포 유형(들) 평가에 사용된다. 많은 병리학적 과정이 손상된 조직으로부터 유래하는 순환 혈장과 같은 체액에서 cfDNA의 일부를 생성할 것으로 예상된다. 여기에서 설명되는 방법은 관련 조직 및/또는 기원 세포 유형의 확인을 포함하여 조직 손상에서 유래된 cfDNA를 검출하고 정량하기 위해 잠재적으로 사용될 수 있다. 이것은 심근경색 (심장 조직의 급성 손상), 자가면역 질환 (다양한 조직의 만성 손상) 및 급성 또는 만성 조직 손상을 수반하는 많은 다른 병리학적 과정의 진단 및/또는 모니터링을 가능하게 할 수 있다.

[0072] 본 발명의 또 다른 측면에서, 본원에서 설명되는 방법은 임신 중 cfDNA의 태아 분획을 평가하고/하거나 염색체 또는 다른 유전적 이상의 검출을 향상시키기 위해 사용된다. 위에 설명된 뉴클레오솜 지도와 커플링된, 모체 혈장 내 DNA 단편의 비교적 피상적인 서열결정은 남녀 태아 임신 모두에서 태아 분획의 비용 효과적이고 신속한 평가를 가능하게 할 수 있다. 또한, 모체 또는 태아 게놈에서 유래했을 가능성에 대해 개별 서열결정 판독에 비균등 확률을 할당할 수 있으므로, 상기 방법은 모체 체액에서 cfDNA의 분석을 통해 염색체 이상 (예를 들어,

삼염색체성)을 검출하는 시험의 성능을 향상시킬 수 있다.

[0073] 본 발명의 또 다른 측면에서, 본원에서 설명되는 방법은 cfDNA에 대한 이식 (자가 또는 동종이식)의 기여도를 정량하기 위해 사용된다. 급성 동종이식 거부의 조기 및 비침습적 검출을 위한 현재의 방법은 혈장 내 DNA를 서열결정하고 공여자 게놈으로부터 유래된 단편의 증가된 농도를 확인하는 것을 수반한다. 이 방법은 예를 들어 5-10%의 공여자 분획을 검출하기 위해 상기 단편 풀의 상대적으로 심층적인 서열결정에 의존한다. 기증된 기관의 뉴클레오솜 지도를 기반으로 하는 방법은 보다 괴상적인 서열결정을 사용하여 유사한 평가를 또는 동등한 양의 서열결정을 사용하여 보다 민감한 평가를 가능하게 할 수 있다. 암과 유사하게, 이식 자체 이외의 다른 세포 유형이 이식 거부 동안 cfDNA 조성에 기여하는 것도 가능하다. 이러한 다른 조직으로부터 cfDNA에 대한 기여가 이식 거부 동안 환자 사이에 일치하는 정도로, 위에서 설명되는 방법으로 이식 공여자 세포 자체가 아닌 상기 다른 조직으로부터의 신호를 기초로 하여 이식 거부를 모니터링할 수 있다.

본 개시내용의 추가의 실시양태.

[0075] 본 개시내용은 또한 알려진 질환 또는 장애를 갖는 대상체로부터 작성된 뉴클레오솜 참조 지도(들)를 사용하여 질환 또는 장애를 진단하는 방법을 제공한다. 일부 실시양태에서, 상기 방법은 (1) 뉴클레오솜 지도의 참조 세트를 작성하고, 여기서 각각의 뉴클레오솜 지도는 정의된 임상 상태 (예를 들어, 정상, 임신, 암 유형 A, 암 유형 B 등)를 갖는 개체(들)의 체액으로부터의 cfDNA 및/또는 특정 조직 및/또는 세포 유형의 염색질의 소화로부터 유래된 DNA로부터 유래되는 것인 단계; (2) 그의 cfDNA로부터 유래된 뉴클레오솜 지도를 뉴클레오솜 지도의 참조 세트와 비교하여 개체(들)의 체액으로부터의 cfDNA의 임상적 상태 및/또는 조직/기원 세포 유형의 조성을 예측하는 것을 포함한다.

[0076] 단계 1: 뉴클레오솜 지도의 참조 세트를 작성하고, 뉴클레오솜 배치로부터의 신호를 모으거나 요약한다.

[0077] 뉴클레오솜 지도를 작성하기 위한 바람직한 방법은 DNA 정제, 라이브러리 구축 (어댑터 라이케이션 및 가능하게는 PCR 증폭에 의한) 및 체액으로부터의 cfDNA의 대량 병렬 서열결정을 포함한다. 참조 점으로서 또는 변형의 주성분을 확인하기 위해 본 발명의 측면에서 유용한, 뉴클레오솜 지도의 대안적인 공급원은 미크로코쿠스 뉴클레아제 (MNase)에 의한 염색질의 소화, DNase 처리, ATAC-Seq 또는 분포 (a), (b) 또는 (c)에서 뉴클레오솜 배치에 대한 정보가 수집되는 다른 관련 방법으로부터 유래된 DNA이다. 이들 분포 (a), (b) 및 (c)에 대한 설명은 상기 단락 번호 [0020]에서 제공되어 있고, 도 1에 그래프로 도시된다.

[0078] 원칙적으로, 이러한 라이브러리의 고도 심층 서열결정은 게놈의 특정 좌표에서 cfDNA에 기여하는 집계된 세포 유형에서 뉴클레오솜 점유를 정량하기 위해 사용될 수 있지만, 이것은 현재 매우 비싸다. 그러나, 뉴클레오솜 점유 패턴과 관련된 신호는 게놈의 연속 또는 불연속 영역에 걸쳐 요약되거나 집계될 수 있다. 예를 들어, 본원에서 제시되는 실시예 1 및 2에서, 서열결정 판독 개시 부위가 지도화되는 참조 인간 게놈 내의 부위의 분포, 즉 분포 (a)는 10 킬로염기 쌍 (kbp)의 인접 윈도우에서 푸리에 변환되고, 뉴클레오솜 점유와 연관된 빈도 범위에 대한 강도의 정량이 이어진다. 이것은 뉴클레오솜이 각각의 10 kbp 윈도우 내에 구조적 배치를 나타내는 정도를 효과적으로 요약한다. 본 명세서에서 제시되는 실시예 3에서, 본 발명자들은 TFBS가 TF에 의해 결합될 때 종종 뉴클레오솜에 의해 바로 측면에 위치하는, 특정 전사 인자 (TF)의 전사 인자 결합 부위 (TFBS)의 바로 근처에서, 서열결정 판독 개시 부위가 지도화되는 참조 인간 게놈 내의 부위의 분포, 즉 분포 (a)를 정량한다. 이것은 cfDNA에 기여하는 세포 유형(들)에서 TF 활성의 결과로서 뉴클레오솜 배치를 효과적으로 요약한다. 중요하게는, 뉴클레오솜 점유 신호가 의미있게 요약될 수 있는 많은 관련 방법이 있다. 이것은 다른 게놈 웬드마크, 예컨대 DNaseI 과민성 부위, 전사 개시 부위, 위상 도메인, 다른 후성학적 마크 또는 다른 데이터세트에서 상관된 거동 (예를 들어, 유전자 발현 등)에 의해 규정되는 모든 이러한 부위의 하위세트 주위에서 분포 (a), (b), 및/또는 (c)로부터 신호의 집계를 포함한다. 서열결정 비용이 계속 떨어짐에 따라, 알려지지 않은 cfDNA 샘플과 비교하기 위해 알려진 질환과 관련된 cfDNA 샘플로부터 작성된 것을 포함하여 뉴클레오솜 점유 지도를 참조 지도로서, 즉 신호를 집계하지 않고 직접 사용하는 것이 가능할 것이다. 일부 실시양태에서, 뉴클레오솜 점유의 참조 지도가 그로부터 작성되는 상기 생물학적 샘플은 인간 세포 또는 조직이 이종이식된 동물로부터 수집된다. 이것의 이점은 조혈 계통과 함께 관심 세포/조직에서 유래된 cfDNA의 혼합물을 나타내는 것과 달리, 인간 게놈에 지도화된 서열결정된 cfDNA 단편이 이종이식된 세포 또는 조직으로부터 독점적으로 유래될 것이라는 것이다.

[0079] 단계 2: 하나 이상의 새로운 개체/샘플의 cfDNA-유래 뉴클레오솜 지도를 직접 또는 각각의 지도의 수학적 변환 후에 뉴클레오솜 지도의 참조 세트와 비교하는 것에 기초하여 병상(들), 임상 상태(들) 및/또는 조직/기원 세포 유형 조성을 예측함.

[0080] 일단 뉴클레오솜 지도의 참조 세트를 작성하면, 추가의 뉴클레오솜 지도(들)를 참조 세트와 비교하기 위한 다양한 통계적 신호 처리 방법이 존재한다. 실시예 1 & 2에서, 본 발명자들은 먼저 다양한 샘플 세트의 계놈을 따라 10 kbp 윈도우 내에서 긴 범위의 뉴클레오솜 배열을 요약한 다음, 샘플을 클러스터링하거나 (실시예 1) 혼합물 비율을 추정하기 위해 (실시예 2) 주성분 분석 (PCA)을 수행한다. 본 발명자들은 이를 실시예에서 사용된 모든 세포주 샘플의 모든 cfDNA 샘플 및 조직/기원 세포 유형의 임상적 상태를 알지만, 샘플 중 어느 하나는 원칙적으로 "알려지지 않은" 것일 수 있고, PCA 분석에서 그의 거동은 모든 다른 뉴클레오솜 지도에 비교한 PCA 분석에서의 그의 거동에 기초하여 임상 상태의 존재/부재 또는 그의 조직/기원 세포 유형을 예측하기 위해 사용된다.

[0081] 미지 샘플은 반드시 1:1 방식으로 참조 세트의 1+ 멤버에 정확하게 매치될 필요는 없다. 오히려, 각각에 대한 그의 유사성을 정량할 수 있거나 (실시예 1), 또는 그의 뉴클레오솜 지도를 참조 세트로부터 2+ 샘플의 비-균일한 혼합물로 모델링할 수 있다 (실시예 2).

[0082] 각각의 샘플에서 cfDNA의 조직/기원 세포 유형 조성은 본 발명의 방법의 성공을 위해 예측되거나 궁극적으로 알려져 있을 필요는 없다. 오히려, 본원에서 설명되는 방법은 특정 병상 또는 임상 상태의 측면에서 cfDNA의 조직/기원 세포 유형 조성의 일관성에 의존한다. 그러나, 염색질 소화로부터 유래된 DNA를 분석하고 이를 뉴클레오솜 지도에 추가하여 많은 조직 및/또는 세포 유형의 뉴클레오솜 지도를 직접 조사함으로써, 알려지지 않은 cfDNA 유래 샘플에 기여하는 조직(들) 및/또는 세포 유형(들)을 추정할 수 있다.

[0083] 본원에서 설명되는 임의의 실시양태에서, 상기 방법은 질환 또는 장애를 확인하는 언급을 포함하는 보고서를 작성하는 단계를 추가로 포함할 수 있다. 일부 실시양태에서, 보고서는 단리된 cfDNA를 생성하는 결정된 조직 및/또는 세포 유형의 목록을 추가로 포함할 수 있다. 일부 실시양태에서, 보고서는 대상체와 관련될 것 같지 않은 질환 및/또는 장애의 목록을 추가로 포함한다. 보고서는 임의적으로 샘플 및/또는 대상체, 생물학적 샘플의 유형, 생물학적 샘플이 대상체로부터 획득된 날짜, cfDNA 단리 단계가 수행된 날짜 및/또는 생물학적 샘플로부터 단리된 임의의 cfDNA를 생성하지 않을 가능성이 있는 조직(들) 및/또는 세포 유형(들)에 대한 임의의 다른 정보를 추가로 포함할 수 있다.

[0084] 일부 실시양태에서, 보고서는 비체한적인 예를 들어, 대상체로부터 추가의 진단 시험의 제안, 치료 요법 개시 제안, 기존의 치료 요법의 변경 제안 및/또는 기존의 치료 요법의 연기 또는 중지 제안을 포함하는 권고된 치료 프로토콜을 추가로 포함한다.

[0085] 실시예

[0086] 실시예 1. 무세포 DNA 뉴클레오솜 지도의 주성분 분석

[0087] 뉴클레오솜 배치와 관련된 신호의 존재를 평가하기 위해 cfDNA 추출 및 MNase 소화 실험으로부터 유래된 서열결정 데이터에서 판독 개시 위치의 분포를 조사하였다. 이를 위해, 한데 모은 cfDNA 샘플 (알려지지 않은 수의 건강한 개체로부터의 기여를 포함하는 인간 혈장; bulk.cfDNA), 한 명의 건강한 남성 대조군 개체 (MC2.cfDNA)로부터의 cfDNA 샘플, 두개내 종양 환자로부터의 4개의 cfDNA 샘플 (tumor.2349, tumor.2350, tumor.2351, tumor.2353), 5개의 상이한 인간 세포주로부터의 6개의 MNase 소화 실험 (Hap1.MNase, HeLa.MNase, HEK.MNase, NA12878.MNase, HeLaS3, MCF.7) 및 상이한 임신 여성 개체로부터의 7개의 cfDNA 샘플 (gmlmatplas, gm2matplas, imlmatplas, fgs002, fgs003, fgs004, fgs005)을 분석하고, 여성 림프아구성 세포주로부터 추출된 DNA (NA12878)의 일반 샷건 서열결정 데이터세트와 대조하였다. 한데 모은 cfDNA 샘플 (26%, bulk.cfDNA_part)의 및 한 명의 건강한 남성 대조군 개체 (18%, MC2.cfDNA_part)의 하위세트가 또한 서열결정 깊이의 효과를 조사하기 위해 별개의 샘플로 포함되었다.

[0088] 판독 개시 좌표가 추출되고, 방법 섹션에서 설명된 바와 같이 고속 푸리에 변환 (FFT)을 사용하여 주기도가 생성되었다. 이 분석은 판독 개시 부위의 분포에서 얼마나 많은 비균일성이 특정 빈도/주기도의 신호에 의해 설명될 수 있는지를 결정한다. 본 발명자들은 하나의 뉴클레오솜 (147 bp) 및 뉴클레오솜 링커 서열 (10-80 bp)의 추가의 서열 주위를 감싸는 DNA의 길이 범위를 포함하는 120-250 bp의 범위에 초점을 맞추었다. 도 3은 인간 1번 염색체 및 인간 22번 염색체의 모든 블록에 걸친 각각의 빈도에 대한 평균 강도를 보여준다. cfDNA 샘플뿐만 아니라 MNase 소화 실험은 200 bp 주기도 아래에서 명확한 피크를 보임을 알 수 있다. 이러한 피크는 인간 샷건 데이터에서 관찰되지 않는다. 이러한 분석은 cfDNA의 단편 경계의 분포에 대한 뉴클레오솜 배치의 주요 효과와 일치한다.

[0089] 샘플 사이의 정확한 피크 빈도의 변화가 또한 관찰되었다. 이것은 아마도 각각의 세포 유형에서 링커 서열 길

이가 상이한 분포의 결과일 수 있다. 피크가 뉴클레오솜 결합 DNA + 링커 서열의 패턴으로부터 유래한다는 것은 피크 주변의 측면이 대칭이 아니고 피크보다 낮은 빈도에 비해 피크보다 높은 빈도에 대한 강도가 더 낮다는 관찰에 의해 지지된다. 이것은 도 3에 제시된 것과 유사한 플롯이 cfDNA 및 MNase 서열결정 데이터의 품질 관리를 수행하기 위해 사용될 수 있음을 제시한다. 규칙적인 (샷건) DNA를 사용한 cfDNA 및 MNase의 무작위 단편화 또는 오염은 주기도에서 이러한 특징적인 강도 패턴의 회색 또는 극단적인 경우 완전한 제거를 야기할 것이다.

[0090]

아래에서, 데이터는 181 bp 내지 202 bp의 빈도 범위에 대해 결정된 모든 강도뿐만 아니라 196 bp의 주기에서 측정된 강도를 기초로 하여 분석되었다. 더 넓은 범위의 링커 길이가 포착되기 때문에, 더 높은 해상도를 제공하기 위해 더 넓은 범위 범위가 선택되었다. 이들 강도는 순전히 계산상의 이유로 주로 선택되었지만; 관련된 실시양태에서 상이한 빈도 범위가 사용될 수 있다. 도 4 및 5는 인간 상염색체의 전체 길이를 배열하는 연속적이며 중첩되지 않는 10 kbp 블록에 걸쳐 196 bp에서 주기도 강도의 시각화를 탐색한다 (자세한 내용은 방법 참조). 도 4는 데이터의 주성분 분석 (PCA) 및 처음 3개의 성분에 걸친 투영도를 보여준다. 주성분 1 (PC1) (28.1%의 분산)은 도 3에 제시된 강도 크기의 차이를 포착하고, 따라서 MNase 및 cfDNA 샘플을 게놈 샷건 데이터로부터 분리한다. 이와 대조적으로, PC2 (9.7%의 분산)는 MNase와 cfDNA 샘플 사이의 차이점을 포착한다. PC3 (6.4% 분산)은 개별 샘플 사이의 차이를 포착한다. 도 5는 강도 벡터의 유클리드 거리에 기초한, 상기 데이터의 계층적 클러스터링 덴도그램을 보여준다. 본 발명자들은 데이터가 다른 실험실에서 다른 실험 프로토콜에 따라 생성되었지만, 2개의 HeLa S3 실험이 PCA 및 덴도그램에서 빠르게 클러스터링됨에 주목한다. "정상적인" cfDNA 샘플, 종양 cfDNA 샘플 및 세포주 MNase 샘플군도 클러스터링되었다. 특히, 동일한 종양 유형 (다형성 교모세포종)에서 유래된 3개의 종양 샘플은 다른 종양 유형에서 유래된 tumor.2351 샘플과는 별개로 클러스터링되는 것으로 보인다 (표 1 참조). GM1 및 IM1 샘플은 임신한 여성으로부터 얻은 다른 cfDNA 샘플과 별개로 클러스터링된다. 이것은 이들 샘플에서 피크 미만의 빈도에 대해 관찰된 보다 높은 강도와 일치한다 (즉, 도 3에서 더 뚜렷한 좌측 어깨). 이것은 두 샘플 세트 사이의 cfDNA 제조의 미묘한 차이 또는 관리되지 않은 생물학적 차이 (예를 들어, 수태 기간)를 나타낼 수 있다.

[0091]

도 6 및 도 7은 동등하지만, 181 bp 내지 202 bp의 빈도 범위에 기초한 분석 결과를 보여준다. 이들 플롯을 비교할 때, 결과는 더 넓은 빈도 범위에서 크게 안정적이지만, 추가의 빈도는 더 정밀한 규모의 분석에서 감도를 개선할 수 있다. 세포 유형 기원 특이적 패턴을 더 탐색하기 위해, cfDNA 및 MNase 데이터세트는 이 빈도 범위에 대한 강도의 PCA를 사용하여 별개로 분석되었다. 다음의 분석 세트에서, 도 3에서 더 뚜렷한 좌측 어깨를 나타낸 임신한 여성으로부터의 5개의 cfDNA 샘플은 제외되었다. 도 8은 cfDNA 데이터의 처음 7개의 주성분을 나타내고, 도 9는 6개의 MNase 데이터세트에 대한 6개 모두의 주성분을 나타낸다. 관련 샘플의 클러스터링이 존재하지만, 나머지 샘플로부터 각각의 샘플을 분리하기 위해 상당한 변형 (생물학적 및 기술적 변형)이 또한 존재한다. 예를 들어, bulk.cfDNA 및 bulk.cfDNA_part는 물론 MC2.cfDNA 및 MC2.cfDNA_part의 분리로부터 알 수 있듯이, 서열결정 깊이의 효과가 관찰되었다. 이 기술 교란 요인을 수정하기 위해 판독 샘플링을 사용할 수 있다.

[0092]

본 실시예의 일부의 주요 관찰은 다음을 포함한다:

[0093]

1) cfDNA 서열결정 데이터에서 판독 개시 좌표는 뉴클레오솜 배치의 강력한 신호를 포착한다.

[0094]

2) 연속적인 10 kbp 윈도우와 같은 게놈의 하위세트에 걸쳐 집계된 뉴클레오솜 배치의 신호의 차이는 샘플의 기원과 상관관계가 있다.

[0095]

실시예 2 - 뉴클레오솜 지도의 혼합물 비율 추정

[0096]

실시예 1에서, 공개 데이터베이스로부터 생성되거나 다운로드된 샘플의 기본 클러스터링이 연구되었다. 분석은 이 데이터세트의 판독 개시 좌표가 (2천만 개의 서열에서 10억 개 초과의 서열까지 얻은 서열결정 깊이의 범위에 걸쳐) 뉴클레오솜 배치의 강력한 신호를 포착하고 샘플 기원이 이 신호와 상호관련됨을 보여준다. 이 방법의 목표를 위해, 알려진 세포 유형의 혼합물을 확인하고 이 신호로부터 각각의 세포 유형의 기여도를 어느 정도 정량할 수 있는 것이 또한 유용할 것이다. 이 목적을 위해, 본 실시예는 두 샘플의 합성 혼합물 (즉, 서열 판독에 기초한)을 조사하였다. 본 발명자들은 2개의 MNase 데이터세트 (MCF.7 및 NA12878.MNase) 및 2개의 cfDNA 데이터세트 (tumor.2349 및 bulk.cfDNA)에 대해 5:95, 10:90, 15:85, 20:80, 30:70, 40:60, 50:50, 60:40, 30:70, 80:20, 90:10 및 95:5의 비율로 서열결정 판독을 혼합하였다. 합성 MNase 혼합물 데이터세트는 두 세트의 1억9천6백9십만 개의 정렬된 판독 (각각 기원 샘플 중 하나로부터)에서 추출되었고, 합성 cfDNA 혼합물 데이터세트는 두 세트의 1억8천백십만 개의 정렬된 판독 (각각 기원 샘플 중 하나로부터)으로부터

추출되었다.

[0097] 도 10은 이들 합성 혼합물에 대한 것을 제외하고 도 3과 동등한, 11번 염색체에 대한 평균 강도를 보여준다. 도 10으로부터, 다른 샘플 기여가 전역 빈도 강도 패턴의 이동을 어떻게 유발하는지 알 수 있다. 이 신호는 합성 혼합물 비율을 추정하기 위해 이용될 수 있다. 도 11은 MNase 데이터세트 혼합물에 대한 처음 2개의 주성분을 보여주고, 도 12는 cfDNA 데이터세트 혼합물의 처음 2개의 주성분을 보여준다. 두 경우 모두에서, 처음 PC는 혼합된 데이터세트의 조성을 직접 포착한다. 따라서, 적절한 참조 세트가 주어지고 예를 들어 회귀 모델을 사용하여 빈도 강도 데이터의 변환으로부터 2개 또는 가능하게는 더 많은 세포 유형에 대한 혼합물 비율이 어떻게 추정될 수 있는지 직접 생각할 수 있다. 도 13은 유사한 샘플 비율 및 cfDNA 및 MNase 샘플의 분리로부터 유래된 혼합물 샘플의 전반적인 유사성을 확인하는, 두 데이터세트의 텐도그램을 보여준다.

[0098] 본 실시예의 핵심 관찰 중 하나는 미지 샘플에 대한 다양한 샘플 유형 (cfDNA 또는 세포/조직 유형)의 혼합물 비율이 뉴클레오솜 점유 패턴의 모델링에 의해 추정될 수 있다는 것이다.

[0099] 실시예 3: cfDNA 서열결정 데이터를 사용한, 전사 인자 결합 부위에 대한 뉴클레오솜 점유의 추정

[0100] 이전의 실시예는 게놈을 연속적이고 중첩되지 않는 10 kbp 윈도우로 분할함으로써 뉴클레오솜 배치의 신호가 얻어질 수 있음을 입증하지만, 직교 방법이 또한 절단 접근가능성 지도를 작성하기 위해 사용될 수 있고, 윈도우 크기와 경계를 기초로 하여 인공물의 영향이 더 작을 수 있다. 본 실시예에서 약간 상세히 조사된 그러한 방법 중 하나는 전사 인자 (TF) 결합 부위 주변의 판독 개시의 관찰된 주기도를 통한 뉴클레오솜 배치의 추정이다.

[0101] 국부 뉴클레오솜 배치가 근처의 TF 점유에 의해 영향을 받는다는 것이 잘 확립되어있다. 염색질의 국부 리모델링에 대한 영향 및 인접 뉴클레오솜의 안정적인 배치에 대한 효과는 TF 세트에 걸쳐 균일하지 않고, 제시된 TF의 점유는 우선적으로 결합 부위의 5' 또는 3'에 위치하는 뉴클레오솜 배치에 국부적인 영향을 미칠 수 있고, 특정 세포 유형에서 더 크거나 작은 게놈 거리에 대해 이어질 수 있다. 또한, 본원의 개시내용의 목적을 위해서 중요하게는, 특정 세포에서 생체 내에서 점유되는 TF 결합 부위의 세트는 조직과 세포 유형 사이에서 다양하고, 따라서 관심 조직 또는 세포 유형에 대한 TF 결합 부위 점유 지도를 확인할 수 있고 하나 이상의 TF에 대해 상기 과정을 반복한다면, 하나 이상의 세포 유형 또는 조직 특이적 TF 결합 부위 점유 프로파일의 농축 또는 고갈을 확인하여 cfDNA의 집단에 기여하는 세포 유형 및 조직의 혼합물의 성분을 식별할 수 있다.

[0102] 상기 아이디어를 입증하기 위해, TF 결합 부위 근처의 판독 개시를 사용하여 우선적인 국부 뉴클레오솜 배치를 반영하는 절단 편향을 시각적으로 확인하였다. ChIP-seq 전사 인자 (TF) 피크는 [Encyclopedia of DNA Elements] ("ENCODE") 프로젝트 (미국 국립 인간 게놈 연구원 (National Human Genome Research Institute), 미국 국립 보건원, 미국 메릴랜드주 베데스다)로부터 얻었다. 이들 피크의 게놈 간격은 광범위하기 때문에 (평균 200 내지 400 bp), 이 간격 내의 활성 결합 부위는 보수적인 p -값 컷오프 (1×10^{-5} , 상세한 내용은 방법 참조)과 각각의 결합 모티프에 대해 게놈의 정보를 스캐닝하여 식별하였다. 예측된 TF 결합 부위의 상기 2개의 독립적으로 유도된 세트의 교차점은 하류 분석으로 이어졌다.

[0103] 500 bp의 각각의 후보 TF 결합 부위 내의 각각의 위치에서 판독 개시의 수는 적어도 1억 개의 서열을 갖는 샘플에서 계산되었다. 각각의 샘플 내에서, 모든 판독 개시는 각각의 위치에서 합산되어, TF 인식 서열의 길이에 따라 TF당 샘플당 총 1,014 내지 1,019개의 위치를 산출하였다.

[0104] 도 14는 결합 부위 자체 주위를 중심으로, 다양한 상이한 샘플에서 인간 게놈 내의 24,666개의 CTCF 결합 부위 주위의 판독 개시의 분포를 보여준다. CTCF는 절연 (insulator) 결합 단백질이고, 전사 억제에 중요한 역할을 수행한다. 이전의 연구는 CTCF 결합 부위가, 적어도 20개의 뉴클레오솜이 185 bp의 대략적인 간격으로 제시된 결합 부위 주위에 대칭적이고 규칙적으로 이격되도록 위치하는 국부 뉴클레오솜 배치에 고정된다고 제시한다. 도 14의 거의 모든 샘플에 공통적인 하나의 현저한 특징은 결합 부위의 상류 및 하류 둘 모두에 위치하는 뉴클레오솜 배치의 명확한 주기도이고, 이것은 생체 내에서 CTCF 결합의 국부적 및 대칭적 효과가 다양한 cfDNA 및 MNase-소화 샘플에서 반복적으로 나타남을 시사한다. 흥미롭게도, 상류 및 하류 피크의 주기는 샘플의 세트에 걸쳐 일정하지 않고; MNase 소화 샘플은 결합 부위에 비해 피크의 약간 더 넓은 간격을 보이고, 이것은 피크의 강도뿐만 아니라 그들의 기간의 유용성을 제시한다.

[0105] 도 15는 5,644개의 c-Jun 결합 부위 주위의 판독 개시 분포를 보여준다. 이 도면에서 익숙한 주기도가 다시 여러 샘플에 대해 시각적으로 식별될 수 있지만, 효과는 일정하지 않다. 흥미롭게도, MNase 소화 샘플 중 3개 (Hap1.MNase, HEK.MNase 및 NA12878.MNase)는 훨씬 평평한 분포를 갖고, 이것은 c-Jun 결합 부위가 이들 세포에서 많이 점유되지 않음을 나타내거나 또는 국부 염색질 리모델링에 대한 c-Jun 결합의 그 효과가 이들 세포

유형에서 덜 현저함을 나타낼 수 있다. 기저 메카니즘에 관계없이, 판독 개시의 국부적인 이웃에서의 편향이 TF로부터 TF까지 및 샘플 유형 사이에서 상이하다는 관찰은 cfDNA 샘플 내의 기원 조직 조성을 서로 연관시키거나 디콘볼루션하기 위한 뉴클레오솜 점유의 판독 개시 기반 추정에 대한 잠재적 역할을 강화한다.

[0106] 도 16은 4,417개의 NF-YB 결합 부위 주위의 판독 개시의 분포를 보여준다. 이 TF 결합 부위 부근의 시작 부위 분포는 대칭성에서 벗어나는 것을 보여주고, 여기에서 cfDNA 샘플의 약간의 상향 편향에 의해 입증되는 바와 같이 하류 효과 (각각의 플롯 내에서 오른쪽으로)가 상류 효과보다 더 강한 것으로 보인다. 또한, MNase 소화 샘플과 cfDNA 샘플 사이의 차이점은 주목할 만하고, 전자는 평균적으로 피크가 식별하기 어려운 평평한 프로파일을 보여주지만, 후자는 더 명확하게 식별 가능한 주기도 및 더 많은 확인 가능한 피크 모두를 갖는다.

[0107] 실시예 1 내지 3의 방법

[0108] 임상 및 대조 샘플

[0109] 일상적인 임신 후기 산전 관리 동안 임신 여성 fgs002, fgs003, fgs004 및 fgs005로부터 전혈을 채혈하고, EDTA (BD)가 들어있는 배큐테이너 (Vacutainer) 튜브에 잠시 보관하였다. 임신 여성 IM1, GM1, GM2로부터의 전혈을 임신 18, 13 및 10주차에 각각 얻은 다음, EDTA가 들어있는 배큐테이너 튜브 (BD)에 잠시 보관하였다. 신경아교종 환자 2349, 2350, 2351 및 2353으로부터의 전혈을 뇌 수술 과정의 일부로 수집하고, EDTA가 들어있는 배큐테이너 튜브 (BD)에서 3시간 미만 동안 보관하였다. 건강한 성인 남성인 남성 대조군 (Male Control) 2 (MC2)로부터의 전혈을 EDTA가 들어있는 배큐테이너 튜브 (BD)에서 수집하였다. 4 내지 10 mL의 혈액이 각각의 개체를 위해 이용 가능하였다. 혈장을 4°C에서 10분 동안 1,000 x g에서 원심분리하여 전혈로부터 분리한 후, 상청액을 모아 4°C에서 15분 동안 2,000 x g에서 다시 원심분리하였다. 정제된 혈장을 사용할 때까지 -80°C에서 1 mL 분취액으로 보관하였다.

[0110] 미지의 수의 건강한 개체로부터의 기여를 포함하는 벌크 인간 혈장을 스템셀 테크놀로지스 (STEMCELL Technologies, 캐나다 브리티ッシュ 콜럼비아 밴쿠버)로부터 수득하고, 사용시까지 -80°C에서 2 mL 분취액으로 보관하였다.

[0111] 혈장 샘플 처리

[0112] 동결된 혈장 분취액을 사용 직전에 벤치-탑 (bench-top)에서 해동시켰다. 순환 cfDNA는 제조사의 프로토콜에 따라 QiaAMP 순환 핵산 (Circulating Nucleic Acids) 키트 (퀴아젠 (Qiagen), 네덜란드 벤로)를 사용하여 각각의 혈장 샘플 2 mL로부터 정제하였다. DNA를 Qubit 형광 측정기 (인비트로겐 (Invitrogen), 미국 캘리포니아주 칼스바드) 및 인간 Alu 서열을 표적으로 하는 주문형 qPCR 검정으로 정량하였다.

[0113] MNase 소화

[0114] 각각의 세포주 (GM12878, HeLa S3, HEK, Hap1)의 약 5천만 개의 세포를 표준 방법을 사용하여 성장시켰다. 성장 배지를 흡인하고, 세포를 PBS로 세척하였다. 세포를 트립신 처리하고, 2x 부피의 CSS 배지로 중화시킨 다음, 1,300 rpm에서 5분 동안 4°C에서 원심분리하여 원추형 튜브에서 펠렛화하였다. 세포 펠렛을 1X 프로테아제 억제제 각테일이 첨가된 12 mL의 냉동 PBS에 재현탁하고, 계수한 후, 4°C에서 1,300 rpm에서 5분 동안 원심분리하여 펠렛화하였다. 세포 펠렛을 RSB 완충제 (10 mM 트리스(Tris)-HCl, 10 mM NaCl, 3 mM MgCl₂, 0.5 mM 스페르미딘, 0.02% NP-40, 1X 프로테아제 억제제 각테일)에 mL당 300만 개의 세포의 농도로 재현탁하고, 조심스럽게 뒤집으면서 10분 동안 얼음 위에서 인큐베이션하였다. 핵을 4°C에서 1,300 rpm에서 5분 동안 원심분리하여 펠렛화하였다. 펠렛화된 핵은 mL당 15 M의 최종 농도로 NSB 완충제 (25% 글리세롤, 5 mM의 MgAc₂, 5 mM HEPES, 0.08 mM EDTA, 0.5 mM 스페르미딘, 1 mM의 DTT, 1X 프로테아제 억제제 각테일) 내에 재현탁하였다. 핵을 4°C에서 1,300 rpm에서 5분 동안 원심분리하여 다시 펠렛화하고, mL당 30 M의 최종 농도로 MN 완충액 (500 mM 트리스-HCl, 10 mM NaCl, 3 mM MgCl₂, 1 mM CaCl₂, 1X 프로테아제 억제제 각테일) 내에 재현탁하였다. 핵을 200 μL 분취액으로 나누고, 37°C에서 5분 동안 4U의 미크로코쿠스 뉴클레아제 (워싱턴 바이오케미컬 코포레이션 (Worthington Biochemical Corp.), 미국 뉴저지주 레이크우드)로 소화시켰다. 85 μL의 MNSTOP 완충제 (500 mM NaCl, 50 mM EDTA, 0.07% NP-40, 1X 프로테아제 억제제)를 첨가하여 반응을 냉장에서 켄칭한 후, 조심스럽게 뒤집으면서 4°C에서 90분 동안 인큐베이션하였다. DNA를 페놀:클로로포름:이소아밀 알콜 추출을 사용하여 정제하였다. 모노뉴클레오솜 단편은 표준 방법을 사용하여 2% 아가로스 겔 전기영동에 의해 크기를 선택하고 나노드롭 (Nanodrop) 분광광도계 (씨모 피셔 사이언티픽 잉크. (Thermo Fisher Scientific Inc.), 미국 매사추세츠주 월담)로 정량하였다.

[0115] 서열결정 라이브러리 제조

독점 소유의 일련의 말단 복구, 라이제이션 및 증폭 반응을 포함하는 ThruPLEX-FD 또는 ThruPLEX DNA-seq 48D 키트 (루비콘 게노믹스 (Rubicon Genomics), 미국 미시건주 앤 아버)를 사용하여 모든 샘플에 대한 바코드 처리된 서열결정 라이브러리를 제조하였다. 3.0 내지 10.0 ng의 DNA를 모든 임상 샘플 라이브러리에 대한 투입물로 사용하였다. 2개의 별크 혈장 cfDNA 라이브러리는 각각의 라이브러리에 30 ng의 투입물을 사용하여 구축되었고; 각각의 라이브러리는 별도로 바코드 처리되었다. MC2로부터의 두 라이브러리는 각각의 라이브러리에 2 ng의 투입물을 사용하여 구축되었고; 각각의 라이브러리는 별도로 바코드 처리되었다. 각각의 MNase 소화 세포주에 대한 라이브러리는 20 ng의 크기 선택된 투입 DNA를 사용하여 구축되었다. 과다 증폭을 피하기 위해 모든 샘플의 라이브러리 증폭을 실시간 PCR로 모니터링하였다.

[0117] 서열결정

9 bp의 인덱스 판독값을 갖는 쌍을 형성한 말단의 101 bp 판독을 사용하여 모든 라이브러리를 HiSeq 2000 기기 (일루미나 인크. (Illumina, Inc.), 미국 캘리포니아주 샌 디에고)에서 서열결정하였다. 한 레인의 서열결정은 한데 모은 샘플 fgs002, fgs003, fgs004 및 fgs005에 대해 수행되어, 샘플당 총 약 4.5×10^7 의 판독 쌍을 생성하였다. 샘플 IM1, GM1 및 GM2는 여러 레인에 걸쳐 서열결정되어, 각각 1.2×10^9 , 8.4×10^8 , 및 7.6×10^7 판독 쌍을 생성하였다. 한 레인의 서열결정은 샘플 2349, 2350, 2351 및 2353 각각에 대해 수행되어, 샘플당 약 2.0×10^8 의 판독 쌍을 생성하였다. 한 레인의 서열결정은 4개의 세포주 MNase-소화된 라이브러리 각각에 대해 수행하여, 라이브러리당 약 2.0×10^8 의 판독 쌍을 생성하였다. 4개 레인의 서열결정은 2개의 복제 MC2 라이브러리 중 하나에 대해, 3개의 레인은 2개의 복제 별크 혈장 라이브러리 중 하나에 대해 수행되어, 각각 라이브러리 당 총 10.6×10^9 및 7.8×10^8 의 판독 쌍을 생성하였다.

[0119] cfDNA 서열결정 데이터의 처리

cfDNA 및 MNase 라이브러리 둘 모두에 대한 DNA 삽입체 크기는 짧은 경향이 있고 (데이터의 대부분이 80 bp 내지 240 bp임); 따라서 일부 분자의 판독 말단에서 어댑터 서열이 예상되었다. 판독 말단부에서 시작하는 어댑터 서열은 트리밍되었고, 짧은 원래 분자에 대한 쌍 형성 말단 ("PE") 데이터의 정방향 및 역방향 판독은 단일 판독 ("SR")으로 붕괴되었고; 적어도 11 bp 판독과 중복되는 PE 판독은 SR로 붕괴되었다. 30 bp보다 짧거나 품질 스코어가 10 미만인 5개 초과의 염기를 나타내는 SR은 폐기되었다. 나머지 PE 및 SR 데이터는 신속 정렬 도구 (BWA-ALN 또는 BWA-MEM)를 사용하여 인간 참조 게놈 (GRCh37, 1000G 릴리스 v2)에 정렬되었다. 생성되는 SAM (Sequence Alignment/Map) 형식은 SAMtools를 사용하여 분류된 BAM (Binary Sequence Alignment/Map 형식)으로 변환되었다.

[0121] 추가의 공개적으로 이용가능한 데이터

HeLa-S3 MNase (등록번호 SRR633612, SRR633613) 및 MCF-7 MNase 실험 (등록번호 SRR999659-SRR999662)의 공개적으로 이용가능한 PE 데이터를 상기한 바와 같이 다운로드하고 처리하였다.

일루미나 캠브리지 엘티디. (Illumina Cambridge Ltd., 영국 에섹스)에 의해 생성된 CEPH 가계 146 개체 NA12878의 공개적으로 이용가능한 게놈 샷건 서열결정 데이터를 유럽 뉴클레오티드 아카이브 (ENA 등록번호 ERR174324-ERR174329)로부터 얻었다. 이 데이터는 일루미나 HiSeq 플랫폼에서 2x101 bp 판독으로 PE 서열을 결정하였고, 라이브러리는 서열결정 전에 더 긴 삽입체 크기에 대해 선택되었다. 따라서, 판독 종료시에 어댑터 서열이 예상되지 않았고, 따라서 이 데이터는 BWA-MEM을 사용하여 직접 정렬되었다.

[0124] 판독 최종 정보 추출

PE 데이터는 서열결정 라이브러리 제조에 사용되는 DNA 분자의 2개의 물리적 말단에 대한 정보를 제공한다. 이 정보는 BAM 파일의 SAMtools 어플리케이션 프로그래밍 인터페이스 (API)를 사용하여 추출되었다. 두 판독이 모두 동일한 염색체에 정렬되고 판독이 반대 배향인 PE 데이터의 외부 정렬 좌표 둘 모두가 사용되었다. 트리밍되지 않은 SR 데이터의 경우, 하나의 판독 말단만이 원래의 DNA 분자의 물리적인 말단에 대한 정보를 제공한다. 판독이 참조 게놈의 플러스 가닥에 정렬되면, 가장 왼쪽의 좌표가 사용되었다. 판독이 역 가닥에 정렬된 경우, 그의 가장 오른쪽의 좌표가 대신 사용된다. PE 데이터가 어댑터 트리밍에 의해 단일 판독 데이터로 전환되는 경우, 양 말단 좌표가 고려되었다. SR 서열결정 실험에서 적어도 5개의 어댑터 염기가 트리밍된 경우, 양쪽 말

단 좌표가 모두 고려되었다.

[0126] 인간 참조 서열 는의 모든 상염색체 (1 내지 22번 염색체)에 대해, 10,000개 염기의 원도우 (블록)에서 모든 위치에서의 판독 말단의 수 및 적용범위를 추출하였다. 블록에서 정렬된 판독이 없으면, 해당 특정 샘플에 대해 블록이 비어있는 것으로 간주하였다.

평탄한 주기도

[0128] 판독 개시 및 적용범위의 비율은 각각의 샘플의 각각의 비어 있지 않은 블록에 대해 계산되었다. 적용범위가 0이면, 비율은 0으로 설정하였다. 이 비율을 사용하여 1/500 염기 내지 1/100 염기의 빈도로 고속 푸리에 변환(FFT, R 통계 프로그래밍 환경의 spec.pgram)을 사용하여 각각의 블록의 주기도를 계산하였다. 임의적으로, 데이터를 평탄하게 하고 (3 bp 다니엘 스무더 (Daniell smoother); 최종값에 1/2의 가중치를 주는 이동 평균) 추세를 제거 (예를 들어, 계열의 평균을 빼고 선형 경향을 제거)하는 파라미터가 사용되었다. 각각의 블록에 대해 120-250 bp의 빈도 범위에 대한 강도가 저장되었다.

평균 염색체 농도

[0130] 샘플 세트에 대해, 모든 샘플에 걸쳐 비어 있지 않은 블록이 확인되었다. 특정 빈도에 대한 강도는 각각의 상염색체에 대한 각각의 샘플의 모든 블록에서 평균화되었다.

주성분 분석 및 텐도그램

[0132] 샘플에 걸쳐 비어 있지 않은 블록이 수집되었다. 주성분 분석 (PCA; R 통계 프로그래밍 환경의 prcomp)은 데이터의 차원수를 줄이고 2차원 공간에 도시하기 위해 사용되었다. PCA는 데이터의 가장 큰 변화를 포착하는 차원을 확인하고 직교 차원을 구축하여, 데이터의 변화량이 감소하는 것을 설명한다.

[0133] 샘플 강도 사이의 쌍별 유클리드 거리를 계산하고, 텐도그램 (R 통계적 프로그래밍 환경의 stats 라이브러리)으로서 가시화하였다.

전사 인자 결합 부위 예측

[0135] 복수개의 세포 유형에 걸쳐 생성된 ChIP-seq 데이터의 분석을 통해 얻은 추정 전사 인자 결합 부위는 ENCODE 프로젝트로부터 수득되었다.

[0136] 후보 전사 인자 결합 부위의 독립적인 세트는 MEME 소프트웨어 패키지 (버전 4.10.0_1)로부터 fimo 프로그램을 사용하여 인간 참조 게놈 (GRCh37, 1000G 릴리스 v2)을 스캐닝하여 얻었다. 스캔은 "--verbosity 1 --thresh 1e-5" 옵션을 사용하여 JASPAR_CORE_2014_vertebrates 데이터베이스로부터 얻은 위치 가중치 매트릭스를 사용하여 수행하였다. 사용된 전사 인자 모티프 식별자는 MA0139.1, MA0502.1 및 MA0489.1이었다.

[0137] 예측된 부위의 두 세트 모두로부터의 염색체 좌표를 베드툴 (bedtools) v2.17.0과 교차시켰다. 플롯에서 임의의 비대칭성을 보존하기 위해, "+" 가닥 상에서 예측된 결합 부위만이 사용되었다. 판독 개시는 예측된 결합 부위의 어느 한 말단에서 500 bp 이내인 경우 각각의 샘플에 대해 집계되었고, 모든 이러한 부위에 걸쳐 위치별로 샘플 내에서 합산되었다. 이 분석에는 적어도 총 1억 개의 판독이 있는 샘플만 사용되었다.

실시예 4: cfDNA로부터 기원 정상/건강한 조직(들)의 결정

[0139] 단일 개체의 cfDNA에서 관찰된 단편화 패턴이 심지어 기여하는 세포 유형 사이에 유전자형 차이가 없을 때에도, 이를 단편을 생성하는 세포의, 및 따라서 cfDNA 분자 집단의 기원 조직(들)의 계놈 조직화의 증거를 포함할 수 있는지 평가하기 위해, cfDNA는 이를 생성하는 과정을 보다 잘 이해하기 위해 심층적으로 서열결정되었다. 생성되는 데이터는 다른 연구자들에 의한 이전 연구를 기반으로 하는 뉴클레오솜 점유의 계놈 전체에 대한 지도를 구축하기 위해 사용되었지만, 상당히 더 포괄적이다. 짧은 단편을 회수하기 위해 라이브러리 제조 프로토콜을 최적화함으로써, CTCF와 같은 전사 인자 (TF)의 생체 내 점유가 cfDNA에 의해 직접적으로 추적된다는 것을 발견하였다. 마지막으로, 건강한 개체에서 cfDNA 서열결정에 의해 밝혀진 바와 같이, 조절 요소 및 유전자 본체의 뉴클레오솜 간격이 림프구성 및 골수성 세포주에서 DNase 과민성 및 유전자 발현과 가장 밀접한 상관관계가 있음이 밝혀졌다.

cfDNA 단편은 크로마토솜에 대응하고 실질적인 DNA 손상을 포함한다

[0141] 통상적인 서열결정 라이브러리는 미지의 수의 건강한 개체 ("BH01")로부터 한데 모은 혈장 또는 단일 개체 ("IH01")로부터의 혈장으로부터 정제된 cfDNA 단편에 대한 말단 복구 및 어댑터 라이게이션에 의해 제조하였다

(도 17; 留 1):

<丑 1>

혈장 샘플에 대한 서열결정 통계.

샘플 명칭	라이브러리 유형	판독	서열결정된 단편	정렬	정렬 Q30	적용범위	추정 중복 %	35-80bp	120- 180bp
BH01	DSP	2x101	1489569204	97.20%	88.85%	96.32	6.00%	0.65%	57.64%
IH01	DSP	2x101	1572050374	98.58%	90.60%	104.92	21.00%	0.77%	47.83%
IH02	SSP	2x50, 43/42	779794090	93.19%	75.27%	30.08	20.05%	21.83%	44.00%
CH01	--	--	3841413668	96.95%	86.81%	231.32	14.99%	5.00%	50.85%

SSP, 단일 가닥 라이브러리 제조 프로토콜. DSP, 이중 가닥 라이브러리 제조 프로토콜.

각각의 샘플에 대해, 서열결정된 단편의 총수, 판독 길이, 지도화 품질 역치를 갖거나 갖지 않는 참조 물질에 정렬하는 상기 단편의 백분율, 평균 적용범위, 중복율 및 2개의 길이 빈에서 서열결정된 단편의 비율을 포함하는 서열결정 관련 통계를 표로 만들었다. 단편 길이는 쌍 형성 말단 판독의 정렬로부터 추정하였다. 판독 길이가 짧기 때문에, 전체 단편을 판독한 것으로 가정하여 적용범위를 계산하였다. 추정되는 중복 단편의 수는 단편 종점을 기초로 하고, 이것은 고도로 정형화된 절단의 존재 하에 진정한 중복율을 과대평가할 수 있다. SSP, 단일 가닥 라이브러리 제조 프로토콜, DSP, 이중 가닥 라이브러리 제조 프로토콜.

라이브러리 BH01 및 IH01을 각각 96배 및 105배 적용범위 (1.5G 및 1.6G 단편)로 서열결정하였다. 쌍을 형성한 말단 판독의 정렬로부터 추정된 단편 길이 분포는 ~167 bp (크로마토솜과 관련된 DNA의 길이와 일치함) 및 100-160 bp 길이 범위에서 ~10.4 bp 주기에서 우세한 피크를 갖는다 (도 18). 이러한 분포는 cfDNA 단편이 단백질과의 회합에 의해, 이 경우 뉴클레아제 코어 입자 및 링커 히스톤에 의해 단백질에 의해 세포 사멸 전 및 사멸 후 둘 모두의 뉴클레아제 절단으로부터 우선적으로 보호되는 모델과 일치하지만, 어느 정도의 추가의 낉 형성 (nicking) 또는 절단이 뉴클레오솜-결합 DNA의 나선 피치와 관련하여 발생한다. 이 모델을 추가로 뒷받침하는 것은 MNase 유래 뉴클레오솜 관련 단편 (예를 들어, 이분 염색체에서 A/T 디뉴클레오티드에 대한 편향)에 대한 초기 연구의 주요 특징을 재현하고 뉴클레오솜 코어 입자는 크로 토솜에 대해 대칭으로 위치한다는 개념을 지지하는 상기 167 bp 단편의 디뉴클레오티드 조성이다 (도 19).

이러한 cfDNA 온톨로지 (ontology) 모델에 대한 예측은 광범위한 DNA 손상, 예를 들어 단일 가닥 뉘뿐만 아니라 5' 및 3' 오버행이다. 전통적인 라이브러리 제조 과정 동안, 뉘 형성된 가닥은 증폭되지 않고, 오버행은 말단 복구에 의해 평활해지고, 전체 cfDNA의 실질적인 비율을 나타낼 수 있는 짧은 이중 가닥 DNA (dsDNA) 분자는 단지 불량하게 회수될 수 있다. 이를 해결하기 위해, 추가의 건강한 개체 ('IH02')로부터 유래된 혈장 함유 cfDNA로부터의 단일 가닥 서열결정 라이브러리를, 광범위한 DNA 손상 및 뉘클레오솜 주위의 뉘클레아제 절단이 보고된 간소지 (Gansauge) 등의 고대 DNA의 연구에서 채택된 프로토콜을 사용하여 제조하였다. 간단히 설명하면, cfDNA는 변성되었고, 비오틴-접합된 단일 가닥 어댑터가 생성 단편에 라이게이션되었다. 이어서, 단편을 스트렙타비딘 비드에 고정시키면서, 라이게이션된 단편을 제2 가닥 합성, 말단 복구 및 제2 어댑터의 라이게이션에 적용하였다. 마지막으로, 샘플 인덱스를 또한 부가하면서, 최소 PCR 증폭을 수행하여 어댑터-보유 분자를 풍부하게 하였다 (도 20; 표 2).

<丑 2>

단일 가닥 서열결정 라이브러리의 제조에 사용되는 합성 올리고.

울리고 명칭	서열 (5'-3')	주
CL9	GTGACTGGAGTTCAGACGTGTGCTTCCGATCT	HPLC 정제
어댑터 2.1	CGACGCTCTTCCGATC/ddT/	HPLC 정제
어댑터 2.2	/5Phos/AGATCGGAAGAGCGTCGTAGGGAAAGAG*T*T*A	HPLC 정제
CL78	/5Phos/AGATCGGAAG/ISpC3/ISpC3/ISpC3/ISpC3/ISpC3/ISpC3/ISpC3/ISpC3/ISpC3/ISpC3/ISpC3/ISpC3/ISpC3/ISpC3/ISpC3/3BioTEG/	이 중 HPLC 정제

IHO2의 경우, 생성된 라이브리리를 30배 적용범위 (779M 단편)로 서열결정하였다. 단편 길이 분포는 다시 크로

마토솜에 대응하는 ~167 bp에서 우세한 피크를 나타내지만, 통상적인 라이브러리 제조와 비교할 때 더 짧은 단편에서 상당히 농축되었다 (도 21, 22, 23a-b, 24a-b). 모든 라이브러리가 ~10.4 bp의 주기도를 나타내지만, 단편 크기는 두 방법에 대해 3 bp만큼 차이나고, 이것은 그의 진정한 종점이 단일 가닥 라이브러리에서 보다 정확하게 표시되는 손상된 또는 비-플러시 (non-flush) 투입 문자와 일치한다.

[0153] 심층적인 cfDNA 서열결정을 기초로 한 생체 내 뉴클레오솜 보호의 계놈 전체에 걸친 지도

cfDNA에 기여하는 조직(들)에서 인간 계놈에 걸친 뉴클레오솜의 현저한 국부 위치가 정렬된 단편 종점의 분포 또는 그의 수학적 변환을 하나 이상의 참조 지도와 비교하여 추정될 수 있는지 평가하기 위해, 윈도우 보호 스코어 ("WPS")가 개발되었다. 구체적으로, cfDNA 단편 종점은 뉴클레오솜 자체에서 고갈되면서, 뉴클레오솜 경계에 인접하여 클러스터링되어야 하는 것으로 예상되었다. 이를 정량하기 위해, WPS가 개발되었고, 이것은 제시된 계놈 좌표를 중심으로 한 120 bp 윈도우에 완전히 걸치는 DNA 단편의 수로부터 동일한 윈도우 내의 종점이 있는 단편의 수를 차감한 값을 나타낸다 (도 25). 의도한 바와 같이, WPS의 값은 시험관 내 방법 또는 고대 DNA를 사용한 다른 그룹에 의해 지도화된 바와 같이, 강하게 위치된 어레이 내의 뉴클레오솜의 위치와 상관관계가 있다 (도 26). 다른 부위에서, WPS는 DNase I 과민성 (DHS) 부위 (예를 들어, 원위의 조절 요소의 측면에 존재하는 뉴클레오솜의 재위치와 일치)과 같은 계놈 특징과 상관관계가 있다 (도 27).

[0155] 휴리스틱 (heuristic) 알고리즘을 BH01, IH01 및 IH02 데이터세트의 계놈 전체에 걸친 WPS에 적용하여, 각각 뉴클레오솜 보호의 12.6M, 11.9M 및 9.7M 국부 최대치를 확인하였다 (도 25-31). 각각의 샘플에서, 인접 피크 사이의 거리 분포 방식은 낮은 편차로 185 bp이었고 (도 30), 이것은 인간 또는 마우스 세포에서의 뉴클레오솜 반복 길이의 이전 분석과 일반적으로 일치하였다.

[0156] 피크 콜의 위치가 샘플에 걸쳐 유사한지 결정하기 위해, 각각의 다른 샘플 내의 가장 가까운 피크에 대한 샘플 내의 각각의 피크의 계놈 거리가 계산되었다. 높은 일치성이 관찰되었다 (도 31; 도 32a-c). BH01 피크 콜로부터 가장 가까운 이웃 IH01 피크 콜까지의 중앙 (절대) 거리는 전체적으로 23 bp이었지만, 가장 높은 스코어의 피크에서는 10 bp 미만이었다 (도 33a-b).

[0157] 뉴클레아제 특이성에 의해 또는 라이브러리 제조 동안 도입된 편향이 뉴클레오솜 보호 신호에 인위적으로 기여할 수 있기 때문에, 단편 종점을 또한 시뮬레이션하여, 각각의 샘플의 깊이, 크기 분포 및 말단 디뉴클레오티드 빈도를 매칭시켰다. 이어서, 계놈 전체에 걸친 WPS를 계산하고, 각각 BH01, IH01 및 IH02와 매치하는 시뮬레이션 데이터세트에 대해, 동일한 휴리스틱에 의해 10.3 M, 10.2 M 및 8.0 M은 국부 최대치로 언급되었다. 시뮬레이션된 데이터세트로부터의 피크는 실제 데이터세트의 피크보다 더 낮은 스코어와 관련되었다 (도 33a-b). 또한, 실제 데이터세트로부터 언급된 비교적 재현가능한 피크의 위치 (도 31, 도 32a-c)는 시뮬레이션된 데이터세트로부터 언급된 피크의 위치와 잘 정렬되지 않았다 (도 31; 도 34a-c).

[0158] 계놈 전체에 걸친 뉴클레오솜 지도의 정확성 및 완전성을 개선하기 위해, BH01, IH01 및 IH02로부터의 cfDNA 서열결정 데이터를 모으고, 조합된 231배 적용범위 ('CH01'; 3.8B 단편; 표 1)에 대해 재분석하였다. WPS가 계산되었고, 상기 조합된 샘플에 대해 12.9 M 피크가 언급되었다. 이 피크 콜 세트는 보다 높은 스코어와 관련되었고, 피크 수의 측면에서 포화에 이르렀다 (도 33a-b). 500 bp 미만인 모든 피크 대 피크 거리를 고려하면 (도 35), CH01 피크 세트는 인간 참조 계놈의 2.53 기가염기 (Gb)에 걸쳐 있다.

[0159] 뉴클레오솜은 유전자 조절의 랜드마크, 예를 들어 전사 개시 부위 및 엑손-인트론 경계와 관련하여 잘 위치하는 것으로 알려져 있다. 그 이해와 일치하게, 유사한 배치가 전사, 번역 및 스플라이싱의 랜드마크와 관련하여 이 데이터에서도 관찰되었다 (도 36-40). 뉴클레오솜 간격과 전사 활성 및 염색질 표시 사이의 상관관계에 대한 파괴 관찰치를 토대로, 림프아구성 세포주에서 긴 범위 상호 작용 (제자리 Hi-C)을 기초로 하여 구획 A (개방 염색질이 풍부함) 또는 구획 B (닫힌 염색질이 풍부함)에 할당된 100 킬로염기 (kb) 윈도우 내의 피크 대 피크 간격 중간값을 조사하였다. 구획 A의 뉴클레오솜은 구획 B의 뉴클레오솜보다 더 춤춤한 간격을 보였고 (중간값 187 bp (A) 대 190 bp (B)), 특정 하위구획 사이에 추가의 차이가 존재하였다 (도 41). 염색체의 길이에 따라, 알파 위성의 어레이에 걸친 강한 배치 (171 bp 단량체 길이; 도 42; 도 26)에 의해 유도되는, 중앙 뉴클레오솜 간격이 동원체 주변 영역에서 급격하게 감소하는 것을 제외하고는, 일반적인 패턴이 보이지 않았다.

[0160] 짧은 cfDNA 단편은 CTCF 및 다른 전사 인자를 직접적으로 추적한다

[0161] DNase I 절단 패턴에 대한 이전의 연구는 2개의 우세한 단편 클래스, 즉 뉴클레오솜 사이의 절단과 관련된 더 긴 단편, 및 전사 인자 결합 부위 (TFBS)에 인접한 절단과 관련된 더 짧은 단편을 확인하였다. 생체 내에서 유도된 cfDNA 단편이 또한 뉴클레아제 절단에 대한 2개의 민감도 클래스에서 유래했는지 평가하기 위해, 서열 판

독 (CH01)을 추정된 단편 길이에 기초하여 분할하고, WPS를 긴 단편 (120–180 bp; 120 bp 윈도우; 사실상 뉴클레오솜 콜링 (calling)에 대해 상기 설명한 WPS와 동일) 또는 짧은 단편 (35–80 bp; 16 bp 윈도우)을 별개로 재계산하였다 (도 26–27). 본 발명자들의 데이터에서 활발하게 결합된 부위에 대해 놓축된 잘 정의된 TFBS의 세트를 얻기 위해 클러스터링된 FIMO 예측은 각각의 TF에 대해 ENCODE (TfbsClusteredV3)의 ChIP-seq 피크의 통합 세트와 교차되었다.

[0162] 긴 분획 WPS는 CTCF 결합 부위 부근에서 뉴클레오솜의 강한 조직화를 지지한다 (도 43). 그러나, 짧은 분획 WPS에서 강한 신호가 관찰되고, 이것은 CTCF 결합 부위 자체와 일치한다 (도 44–45). CTCF 결합 부위는 이들이 생체 내에서 결합된다는 가정에 기초하여 계층화되었다 (모든 FIMO 예측 대 ENCODE ChIP-seq와 교차하는 하위세트 대 19개의 세포주에 걸쳐 사용되는 것으로 보이는 것과 교차하는 추가의 하위세트). 실험적으로 잘 지지된 CTCF 부위는 CTCF 결합시 그의 재위치화 (~190 bp → ~260 bp, 도 45–48)과 일치하게, 긴 분획 WPS에 기초하여 인접 -1과 +1 뉴클레오솜 사이에 실질적으로 더 넓은 간격을 나타낸다. 또한, 실험적으로 잘 지지된 CTCF 부위는 CTCF 결합 부위 자체보다 짧은 분획 WPS에 대해 훨씬 강한 신호를 나타낸다 (도 49–52).

[0163] FIMO 예측 및 ENCODE ChIP-seq 데이터 둘 모두가 이용가능한 추가의 TF에 대해 유사한 분석을 수행하였다 (도 53a–h). 이들 TF 중의 많은 것, 예컨대 ETS 및 MAFK의 경우 (도 54–55), 짧은 분획 풋프린트가 관찰되었고, 긴 분획 WPS에서 주기적 신호가 수반되었다. 이것은 결합된 TFBS를 둘러싼 뉴클레오솜의 강력한 배치와 일치한다. 전체적으로, 이들 데이터는 단일 가닥 프로토콜 (도 18, 도 21)에 의해 현저히 양호하게 회복된 짧은 cfDNA 단편이 CTCF 등을 포함하는 DNA 결합 전사 인자의 생체 내 점유를 직접적으로 추적한다는 견해를 지지한다.

뉴클레오솜 간격 패턴은 cfDNA 기원 조직에 대한 정보를 제공한다

[0165] cfDNA 서열결정을 통해 측정된 생체 내 뉴클레오솜 보호가 건강한 개체에서 cfDNA에 기여하는 세포 유형을 추정하기 위해 사용될 수 있는지 결정하기 위해, 116개의 다양한 생물학적 샘플에서 정의된 DHS 부위 내의 뉴클레오솜 콜의 피크 대 피크 간격을 조사하였다. 확대된 간격은 조절 요소에서 (예를 들어, DHS 부위에서 일화적으로 (anecdotally) (도 27) 또는 결합된 CTCF 부위에서 전반적으로 (도 45)) -1 및 +1 뉴클레오솜 사이에서 이전에 관찰되었다. 결합된 CTCF 부위와 유사하게, DHS 부위의 하위세트 내에서 뉴클레오솜 쌍에 대해 실질적으로 더 넓은 간격이 관찰되었고, 타당하게 이것은 cfDNA를 생성하는 세포 유형(들)에서 결합하는 개별 전사 인자 결합에 의해 뉴클레오솜이 재위치되는 부위에 대응한다 (~190 bp → ~260 bp, 도 56). 실제로, 확장된 뉴클레오솜 간격 (~260 bp)의 비율은 어떤 세포 유형의 DHS 부위가 사용되는지에 따라 상당히 다르다. 그러나, 이 비율이 가장 높은 모든 세포 유형은 램프 또는 골수 기원이다 (예를 들어, 도 56에서 CD3_CB-DS17706 등). 이것은 건강한 개체에서 cfDNA의 주요 공급원으로서 조혈 세포 사멸과 일치한다.

[0166] 다음으로, 전사 개시 부위 부근에서 뉴클레오솜 보호 신호가 재조사되었다 (도 36). 신호가 램프 계통 세포주인 NB-4에서의 유전자 발현에 기초하여 계층화된 경우, TSS와 관련하여 뉴클레오솜 보호의 위치 또는 강도의 강한 차이가 고도로 발현된 유전자 대 낮게 발현된 유전자에서 관찰되었다 (도 57). 또한, 짧은 분획 WPS는, 그의 강도가 또한 발현 수준과 강한 상관관계가 있는 TSS의 바로 상류에 명확한 풋프린트를 보인다 (도 58). 이것은 전사 활성 유전자에서 전사 예비 개시 복합체 또는 그의 일부 성분의 풋프린트를 반영한다.

[0167] 이들 데이터는 cfDNA 단편화 패턴이 실제로 cfDNA를 생성하는 조직(들) 또는 세포 유형(들)을 추정하기 위해 사용될 수 있는 신호를 포함함을 입증한다.

[0168] 그러나, 문제는 계놈 전체에 걸친 cfDNA 라이브러리에서 비교적 적은 수의 판독이 DHS 부위 및 전사 개시 부위와 직접 중복된다는 것이다.

[0169] 뉴클레오솜 간격은 세포 유형 사이에, 및 염색질 상태 및 유전자 발현의 함수로서 상이하다. 일반적으로, 개방 염색질 및 전사는 보다 짧은 뉴클레오솜 반복 길이와 관련되고, 이것은 구획 A 대 B의 상기 실시예의 분석과 일치한다 (도 41). 이 실시예의 피크 콜 데이터는 또한 유전자 본체에 걸친 뉴클레오솜 간격과 이들의 발현 수준 사이의 상관관계를 나타내며, 보다 촘촘한 간격은 보다 높은 발현과 연관된다 (도 59; $p = -0.17$; $n = 19,677$ 유전자). 상관관계는 인접 영역 (상류 10 kb $p = -0.08$; 하류 10 kb $p = -0.01$)에 비해 유전자 본체 자체에 대해서 가장 높다. 분석이 적어도 60개의 뉴클레오솜 콜에 걸친 유전자 본체에 제한된다면, 보다 촘촘한 뉴클레오솜 간격은 유전자 발현과 훨씬 더 강력한 상관관계가 있다 ($p = -0.50$; $n = 12,344$ 유전자).

[0170] 유전자 본체 또는 다른 도메인에 걸친 뉴클레오솜 간격과 같은 신호를 이용하는 한 가지 이점은 cfDNA 단편의 훨씬 더 많은 비율이 정를 제공할 것이라는 것이다. 또 다른 잠재적인 이점은 cfDNA에 기여하는 여러 세포 유형에서 발생하는 신호의 혼합물을 검출할 수 있다는 것이다. 이것을 시험하기 위해, 처음 10 kb의 유전자 본체

에 걸친 긴 단편 WPS에 대해 및 유전자 단위 기준으로 추가의 수학적 변환, 즉 고속 푸리에 변환 (FFT)을 수행하였다. FFT 신호의 강도는 특정 빈도 범위에서 유전자 발현과 상관관계가 있었고, 양성 상관관계에서는 177-180 bp에서 가장 높았고 음성 상관관계에서는 ~199 bp에서 최소화되었다 (도 60). 인간 세포주 및 1차 조직에 대한 76개의 발현 데이터세트에 대한 상기 분석을 수행할 때, 가장 강력한 상관관계는 조혈 계통에서 관찰되었다 (도 60). 예를 들어, 3개의 건강한 샘플 (BH01, IH01, IH02) 각각에 대해 193-199 bp 빈도 범위에서 평균 강도를 갖는 가장 높은 순위의 음성 상관관계는 모두 림프 세포주, 골수 세포주 또는 골수 조직에서 나타났다 (도 61; 표 3).

[0171] <표 3>

[0172] 유전자 발현 데이터세트와 WPS FFT 강도의 상관관계.

R명칭	범주	유형	설명	상관관계												순위 차이	
				BH01	IH01	IH02	IC15	IC20	IC17	IC37	IC35	건강한	IC15	IC20	IC17	IC37	
A.431	피부	피부암 (편평 세포)	유모피 암종 세포주	-0.298	-0.188	-0.149	-0.200	-0.140	-0.176	-0.195	-0.178	2	3	-9	-9	-12	-21
A549	폐	폐암종	폐암종 세포주	-0.289	-0.185	-0.144	-0.202	-0.139	-0.172	-0.188	-0.170	3	-14	-12	-8	-2	-13
저방 조직	1차 조직	저방 조직	1차 조직	-0.270	-0.169	-0.137	-0.169	-0.121	-0.153	-0.166	-0.148	1	12	5	0	14	12
부신	1차 조직	부신	1차 조직	-0.257	-0.158	-0.131	-0.173	-0.118	-0.145	-0.161	-0.138	-2	-11	-5	1	5	8
AN3.CA	유방/여성 생식	자궁암	정이성 자궁내막 선암종 세포주	-0.303	-0.194	-0.157	-0.213	-0.147	-0.183	-0.195	-0.171	4	-16	-13	-15	-8	-2
행장	1차 조직	행장	1차 조직	-0.287	-0.185	-0.137	-0.168	-0.118	-0.148	-0.171	-0.152	6	24	20	23	8	9
BEWO	기타	자궁암	정이성 유모마 암종 세포주	-0.284	-0.184	-0.147	-0.193	-0.139	-0.173	-0.194	-0.173	-5	3	-12	-15	-19	-27
골수	1차 조직	골수	1차 조직	-0.343	-0.230	-0.165	-0.192	-0.142	-0.167	-0.193	-0.165	2	40	9	30	16	28
CACO-2	복부	정장 선암종	정장 선암종 세포주	-0.281	-0.177	-0.137	-0.192	-0.128	-0.169	-0.184	-0.164	5	-5	-5	-14	-10	-9
CAPAN-2	복부	췌장 선암종	췌장 선암종 세포주	-0.291	-0.187	-0.145	-0.202	-0.136	-0.176	-0.195	-0.175	3	-12	-2	-18	-19	-25
데뇌피질	1차 조직	데뇌피질	1차 조직	-0.225	-0.136	-0.120	-0.168	-0.108	-0.134	-0.142	-0.125	-1	-9	-3	0	0	0
결장	1차 조직	결장	1차 조직	-0.261	-0.162	-0.124	-0.164	-0.111	-0.145	-0.168	-0.146	7	6	6	6	-7	1
다수다	림프성	임진 비黠 (Burkitt) 림프종	임진 비黠 림프종 세포주	-0.321	-0.206	-0.153	-0.185	-0.133	-0.185	-0.188	-0.180	4	17	19	18	13	24
실이지장	1차 조직	실이지장	1차 조직	-0.261	-0.164	-0.122	-0.159	-0.109	-0.144	-0.166	-0.144	10	10	10	7	-4	7
EFO.21	유방/여성 생식	난소암	정이성 난소암 정장 선암종 세포주	-0.287	-0.186	-0.149	-0.201	-0.140	-0.176	-0.186	-0.169	-7	-9	-14	-20	-1	-6
자궁내막	1차 조직	자궁내막	1차 조직	-0.257	-0.158	-0.132	-0.178	-0.119	-0.151	-0.166	-0.151	-3	-11	-4	-8	-3	-12
식도	1차 조직	식도	1차 조직	-0.237	-0.147	-0.124	-0.156	-0.116	-0.141	-0.158	-0.145	-3	1	-7	0	0	-7
나팔관	1차 조직	나팔관	1차 조직	-0.247	-0.157	-0.120	-0.171	-0.114	-0.145	-0.161	-0.145	4	-13	-2	-3	3	-2
담낭	1차 조직	담낭	1차 조직	-0.249	-0.156	-0.119	-0.153	-0.103	-0.136	-0.154	-0.141	4	4	4	3	4	1
HaCat	피부	각질세포 세포주	각질세포 세포주	-0.290	-0.186	-0.149	-0.194	-0.142	-0.173	-0.193	-0.173	-5	7	-16	-6	-8	-17
HDLIM-2	림프성	호지킨 (Hodgkin) 림프종	호지킨 림프종 세포주	-0.316	-0.200	-0.154	-0.201	-0.136	-0.173	-0.195	-0.171	1	6	11	1	-5	-5
심근	1차 조직	심근	1차 조직	-0.246	-0.149	-0.126	-0.168	-0.113	-0.141	-0.155	-0.140	-3	-3	0	3	2	
HEK-293	기타	신장 부신 경구체 세포주	비이 신장 부신 경구체 세포주 이모노바이러스 임5에 의해 형질전환형	-0.292	-0.187	-0.150	-0.209	-0.159	-0.172	-0.189	-0.188	-4	-17	-4	-2	3	0
HEL	골수성	적색 혈액	적색 혈액 세포주 (호지킨 병에 대한 치료 후 재발된 AML M6)	-0.324	-0.205	-0.161	-0.210	-0.140	-0.172	-0.194	-0.166	-1	-5	4	12	5	14
HeLa	유방/여성 생식	자궁경부암	자궁경부 암종 선암종 세포주	-0.296	-0.186	-0.149	-0.203	-0.139	-0.172	-0.190	-0.171	1	-10	-5	-3	1	-6
Hep_G2	복부	간세포 암종	간세포 암종 세포주	-0.294	-0.186	-0.152	-0.202	-0.145	-0.186	-0.196	-0.167	4	-6	-18	-24	-17	2
HL-60	골수성	전립수성 백혈병	전립수성 백혈병 세포주	-0.332	-0.208	-0.161	-0.202	-0.137	-0.171	-0.197	-0.170	2	8	18	18	-1	11
HMC-1	골수성	비만세포성 백혈병	비만세포성 백혈병 세포주	-0.337	-0.228	-0.165	-0.212	-0.149	-0.181	-0.199	-0.180	0	-1	-2	3	0	-2
K-562	림프성	백혈병	간세포 암종 세포주 (CML) 세포주	-0.317	-0.202	-0.158	-0.211	-0.143	-0.178	-0.195	-0.166	-3	-9	-5	-6	-1	13
Karpas.707	림프성	다발성 골수종	다발성 골수종 세포주	-0.325	-0.210	-0.155	-0.195	-0.136	-0.167	-0.188	-0.164	4	20	18	22	21	22
신장	1차 조직	신장	1차 조직	-0.245	-0.160	-0.130	-0.168	-0.119	-0.153	-0.171	-0.147	-7	4	-12	-19	-21	-6
간	1차 조직	간	1차 조직	-0.248	-0.148	-0.122	-0.150	-0.110	-0.150	-0.164	-0.138	1	4	-1	-13	-4	3
폐	1차 조직	폐	1차 조직	-0.264	-0.170	-0.133	-0.170	-0.121	-0.148	-0.167	-0.149	3	4	0	7	3	6

[0173]

R명칭	범주	유형	설명	상관관계												순위 차이							
				BH01	II01	II02	IC15	IC20	IC17	IC37	IC35	건강한	IC15	IC20	IC17	IC37	IC35	14	22	25	17	24	-19
림프절	1차 조작	림프절	1차 조작	-0.308	-0.195	-0.148	-0.182	-0.126	-0.155	-0.181	-0.156	7	24	17	25	14	22						
MCF7	유방/여성 생식	유방암	유방암	-0.298	-0.195	-0.154	-0.207	-0.145	-0.183	-0.196	-0.181	-3	-9	-12	-18	-11	-19						
MOLT.4	림프성	백혈병 (ALL)	유방암 유방암 성인암 세포주	-0.323	-0.204	-0.163	-0.212	-0.144	-0.177	-0.197	-0.173	-3	-7	-2	-1	-5	-1						
NB.4	골수성	질환수성 백혈병	골수성 질환수성 백혈병 (ALL) 세포주	-0.348	-0.228	-0.172	-0.211	-0.148	-0.182	-0.202	-0.171	0	4	3	5	2	13						
NTERA.2	비뇨기/남성 생식	비뇨기암	비뇨기암	-0.269	-0.170	-0.137	-0.193	-0.117	-0.157	-0.169	-0.153	-2	-8	16	-2	5	0						
남소	1차 조작	남소	1차 조작	-0.266	-0.162	-0.135	-0.181	-0.120	-0.152	-0.166	-0.151	1	-7	2	-2	6	-1						
췌장	1차 조작	췌장	1차 조작	-0.250	-0.159	-0.132	-0.170	-0.116	-0.150	-0.166	-0.150	-5	-5	1	-6	-5	-7						
PC.3	비뇨기/남성 생식	전립선암	전립선암 유방암 분화면 전립선 성인암 세포주	-0.295	-0.190	-0.151	-0.204	-0.138	-0.174	-0.188	-0.173	-3	-10	2	-6	8	-12						
췌관	1차 조작	췌관	1차 조작	-0.266	-0.166	-0.134	-0.168	-0.126	-0.151	-0.168	-0.150	3	10	-7	1	9	6						
전립선	1차 조작	전립선	1차 조작	-0.248	-0.161	-0.133	-0.175	-0.123	-0.150	-0.165	-0.151	-8	-10	-11	-8	1	-12						
직장	1차 조작	직장	1차 조작	-0.255	-0.154	-0.117	-0.159	-0.102	-0.136	-0.161	-0.142	6	0	5	4	-2	0						
REH	림프성	백혈병 (ALL)	B-경구세포 백혈병 세포주 (ALL, 1차, 2차) 관	-0.330	-0.216	-0.165	-0.214	-0.150	-0.182	-0.204	-0.174	-2	-5	-5	-2	-4	1						
RH.30	육종	평균근육종	평균근육종 평균근육종 세포주	-0.280	-0.165	-0.137	-0.194	-0.125	-0.158	-0.175	-0.156	2	-14	-3	-7	-7	-7						
RPML.8226	림프성	다발성 골수종	다발성 골수종 세포주	-0.322	-0.207	-0.155	-0.198	-0.138	-0.169	-0.190	-0.164	1	16	11	19	14	22						
RT4	비뇨기/남성 생식	방광암	방광암 방광암 세포주	-0.282	-0.168	-0.145	-0.192	-0.136	-0.170	-0.191	-0.171	-5	-1	-12	-16	-19	-25						
타액선	1차 조작	침설	1차 조작	-0.262	-0.166	-0.138	-0.177	-0.128	-0.154	-0.172	-0.155	-7	2	-9	-2	-5	-5						
SCLC.21H	폐	소세포 폐암종	소세포 폐암종 세포주	-0.259	-0.160	-0.138	-0.201	-0.123	-0.157	-0.172	-0.146	-11	-31	-5	-12	-10	8						
SH.SY5Y	뇌	신경모세포종	신이성 신경모세포종, 신경상피종 세포주 SK-N-SH와 글로벌 하위세포주	-0.271	-0.170	-0.137	-0.201	-0.124	-0.157	-0.170	-0.151	2	-25	2	-5	1	6						
SiHa	유방/여성 생식	자궁경부암	자궁경부 경부세포 암 세포주 봉합전 1-2 카페의 HPV16	-0.288	-0.180	-0.148	-0.201	-0.139	-0.176	-0.193	-0.175	-2	-7	-15	-19	-11	-27						
SK.BR.3	유방/여성 생식	유방암	유방암 유방암 세포주	-0.288	-0.176	-0.148	-0.195	-0.140	-0.176	-0.191	-0.169	-3	-4	-21	-22	-12	-11						
SK.MEL.30	1차 조작	흑색종	청이성 흑색종 세포주	-0.301	-0.187	-0.154	-0.208	-0.141	-0.174	-0.193	-0.171	-2	-12	-8	-3	-1	-6						
골격근	1차 조작	골격근	골격근 1차 조작	-0.261	-0.166	-0.134	-0.179	-0.125	-0.150	-0.164	-0.145	-1	-7	-7	0	9	11						
피부	피부	피부	1차 조작	-0.259	-0.166	-0.134	-0.168	-0.127	-0.148	-0.167	-0.151	-4	8	-14	5	-1	-4						
소장	1차 조작	소장	1차 조작	-0.260	-0.164	-0.121	-0.158	-0.107	-0.141	-0.166	-0.142	9	10	11	9	9	7						
췌장근	1차 조작	췌장근	1차 조작	-0.259	-0.158	-0.127	-0.169	-0.113	-0.144	-0.161	-0.149	2	-6	3	4	4	-5						
비장	1차 조작	비장	1차 조작	-0.308	-0.202	-0.148	-0.180	-0.130	-0.155	-0.177	-0.154	7	27	15	25	20	25						
위	1차 조작	위	1차 조작	-0.264	-0.170	-0.131	-0.170	-0.117	-0.149	-0.169	-0.151	6	3	9	6	0	2						
고환	1차 조작	고환	1차 조작	-0.215	-0.142	-0.109	-0.147	-0.093	-0.126	-0.133	-0.123	0	0	0	0	0	0						
THP.1	골수성	단핵구성 백혈병	단핵구성 백혈병 (AML) 세포주	-0.338	-0.218	-0.168	-0.206	-0.149	-0.182	-0.204	-0.176	-1	8	-1	1	-3	0						

R명칭	범주	유형	설명	상관관계										순위 차이			
				BH01	IH01	IH02	IC15	IC20	IC17	IC37	IC35	건강한	IC15	IC20	IC17	IC37	IC35
감상선	1차 조직	감상선	1차 조직	-0.261	-0.158	-0.136	-0.178	-0.121	-0.153	-0.170	-0.161	-2	-7	-2	-6	-5	-19
TIME	기타	미세혈관 내피세포주	밀로미아제- 블랙풀 인간 마우스 혈관 내피세포 (한국 모음)	-0.296	-0.180	-0.147	-0.198	-0.134	-0.170	-0.186	-0.170	5	-3	3	-1	3	-11
편도선	1차 조직	편도선	1차 조직	-0.282	-0.179	-0.141	-0.169	-0.125	-0.147	-0.173	-0.152	-1	20	8	23	4	9
U.138_MG	뇌	교모세포주	교모세포주	-0.288	-0.177	-0.144	-0.191	-0.128	-0.162	-0.177	-0.161	1	8	7	0	2	2
U.2_O5	육종	글육종	글육종 세포주	-0.275	-0.175	-0.139	-0.192	-0.134	-0.159	-0.170	-0.160	-2	0	-11	-3	6	-3
U.2497	육종	육종	악성 성유조직구종 세포주	-0.290	-0.181	-0.148	-0.195	-0.129	-0.164	-0.180	-0.165	2	1	5	3	4	0
U.251_MG	뇌	교모세포종	교모세포종 세포주	-0.292	-0.178	-0.140	-0.197	-0.125	-0.160	-0.177	-0.165	9	-6	11	4	4	-4
U.266.70	림프성	다발성 골수종	다발성 골수종 세포주 (1984, IL-6-이온성)	-0.320	-0.207	-0.157	-0.202	-0.135	-0.170	-0.191	-0.165	-1	4	19	15	12	17
U.266.84	림프성	다발성 골수종	다발성 골수종 세포주 (1984, 세포형 내 분화형)	-0.326	-0.212	-0.162	-0.207	-0.139	-0.175	-0.194	-0.169	-1	2	11	8	10	14
U.698	림프성	B-세포 림프종	B-세포 림프종 세포주 (림프모구성 림프육종)	-0.328	-0.212	-0.159	-0.203	-0.137	-0.170	-0.194	-0.166	2	5	18	20	6	20
U.87_MG	뇌	교모세포종, 성상세포종	교모세포종, 성상세포종 세포주	-0.285	-0.175	-0.143	-0.192	-0.127	-0.160	-0.174	-0.162	1	0	2	-2	2	-4
U.937	골수성	골수단핵구성 조직구성 림프종	골수단핵구성 조직구성 림프종 세포주	-0.346	-0.224	-0.167	-0.201	-0.146	-0.180	-0.199	-0.173	1	18	3	5	2	6
방광	1차 조직	방광	1차 조직	-0.260	-0.158	-0.130	-0.165	-0.118	-0.146	-0.164	-0.150	3	5	-2	1	3	-6
WM.115	폐부	흑색종	악성 흑색종 세포주	-0.284	-0.175	-0.144	-0.193	-0.130	-0.160	-0.178	-0.157	-1	-4	-4	-3	-3	2

[0175]

[0176]

휴먼 프로테인 아틀라스 (Human Protein Atlas)에 의해 44개의 인간 세포주 및 32개의 1차 조직에서 19,378개의 Ensembl 유전자 식별자에 대해 측정된 FPKM 발현 값을 갖는 전사 개시 부위 하류의 처음 10 kb에서 193-199 bp 빈도에 대한 평균 FFT (고속 푸리에 변환) 강도 사이의 상관관계 값. 표 3에는 프로테인 아틀라스에서 제공되는 각각의 발현 샘플에 대한 간략한 설명 및 IH01, IH02 및 BH01 샘플에 대한 순위 변환 및 순위 차이가 포함되어 있다.

[0177]

실시예 5: cfDNA로부터 비-건강한 기원 조직의 결정

[0178]

건강하지 않은 상태의 추가의 기여 조직이 추정될 수 있는지 시험하기 위해, 5명의 후기 암 환자로부터 얻은 cfDNA 샘플의 서열을 결정하였다. 이들 샘플에서 뉴클레오솜 간격 패턴은 종종 환자의 암의 해부학적 기원과 일치하는 비-조혈 조직 또는 세포주와 가장 강한 상관관계가 있는 cfDNA에 대한 추가의 기여를 나타낸다.

[0179]

암 환자의 cfDNA에서 뉴클레오솜 간격은 비-조혈 기여를 확인한다.

[0180]

건강하지 않은 상태에서 순환하는 cfDNA에 기여하는 비-조혈 계통의 시그너처가 검출될 수 있는지 결정하기 위해, 다양한 IV기 암으로 임상적으로 진단된 개체로부터의 44개의 혈장 샘플을 cfDNA로부터 제조된 단일 가닥 라이브리의 가벼운 서열결정으로 스크리닝하였다 (표 4, 중간값 2.2배 적용범위):

[0181]

<표 4>

[0182] 암 패널 (cancer panel)을 위한 임상 진단 및 cfDNA 수율.

샘플 ID	임상 Dx	병기	cfDNA 수율 (ng/ml)	환자 성별
IC01 †	신장암 (이행 세포)	IV	242	F
IC02	난소암 (비규정됨)	IV	22.5	F
IC03	피부암 (흑색종)	IV	12.0	M
IC04	유방암 (침습성/침윤성 유관)	IV	12.6	F
IC05	폐암 (선암종)	IV	5.4	M
IC06	폐암 (중피종)	IV	11.4	M
IC07 †	위암 (비규정됨)	IV	52.2	M
IC08	자궁암 (비규정됨)	IV	15.0	F
IC09	난소암 (장액성 종양)	IV	8.4	F
IC10	폐암 (선암종)	IV	11.4	F
IC11	결장직장암 (비규정됨)	IV	11.4	M
IC12	유방암 (침습성/침윤성 소엽)	IV	12.0	F
IC13	전립선암 (비규정됨)	IV	12.3	M
IC14	두경부암 (비규정됨)	IV	27.0	M
IC15 §	폐암 (소세포)	IV	22.5	M
IC16	방광암 (비규정됨)	IV	14.1	M
IC17 §	간암 (간세포 암종)	IV	39.0	M
IC18	신장암 (투명세포)	IV	10.5	F
IC19	고환암 (정상피종)	IV	9.6	M
IC20 §	폐암 (편평세포 암종)	IV	21.9	M
IC21	췌장암 (췌관 선암종)	IV	35.4	M
IC22	폐암 (선암종)	IV	11.4	F
IC23	간암 (간세포 암종)	IV	17.1	M
IC24	췌장암 (췌관 선암종)	IV	37.2	M
IC25	췌장암 (췌관 선암종)	IV	27.9	M
IC26	전립선암 (선암종)	IV	24.6	M
IC27	자궁암 (비규정됨)	IV	19.2	F
IC28	폐암 (편평세포 암종)	IV	33.3	M

[0183]

IC29	두경부암 (비규정됨)	IV	14.4	M
IC30	식도암 (비규정됨)	IV	10.5	M
IC31 †	난소암 (비규정됨)	IV	334.8	F
IC32	폐암 (소세포)	IV	9.6	F
IC33	결장직장암 (선암종)	IV	13.8	M
IC34	유방암 (침습성/침윤성 소엽)	IV	33.6	F
IC35 §	유방암 (유관 상피내 암종)	IV	16.2	F
IC36	간암 (비규정됨)	IV	26.4	M
IC37 §	결장직장암 (선암종)	IV	15.9	F
IC38	방광암 (비규정됨)	IV	6.6	M
IC39	신장암 (비규정됨)	IV	39.0	M
IC40	전립선암 (선암종)	IV	13.8	M
IC41	고환암 (정상피종)	IV	16.5	M
IC42	폐암 (선암종)	IV	11.4	F
IC43	피부암 (흑색종)	IV	21.9	F
IC44	식도암 (비규정됨)	IV	25.8	F
IC45 †	결장직장암 (선암종)	IV	3.0	M
IC46 **	유방암 (유관 상피내 암종)	IV	36.6	F
IC47	췌장암 (췌관 선암종)	IV	19.2	F
IC48 **	유방암 (침습성/침윤성 소엽)	IV	13.8	F

[0184]

§ : 샘플은 추가의 서열결정을 위해 선택되었다.

[0186]

**: 단지 0.5 ml의 혈장만이 상기 샘플에 이용될 수 있었다.

[0187]

† : 샘플은 QC에 실패하였고, 향후 분석에 사용되지 않았다.

[0188]

표 4는 각각의 개체로부터의 1.0 ml의 혈장으로부터의 총 cfDNA 수율 및 관련된 임상 공변량과 함께, 높은 종양 부하의 증거에 대해 혈장-함유 cfDNA가 스크리닝된 48명의 환자에 대한 임상 및 조직학적 진단을 보여준다. 이들 48개 중에서, 44개는 QC를 통과하였고, 충분한 재료를 가지고 있었다. 이들 44개 중에서, 5개를 심층 서열 결정을 위해 선택하였다. cfDNA 수율은 Qubit 형광 측정기 2.0 (라이프 테크놀로지스 (Life Technologies))에 의해 결정되었다.

[0189]

이들 샘플은 실시예 4의 IH02와 동일한 프로토콜로, 동일한 배치에서 다수 제조되었다. IV기 암으로 임상 진단된 52명의 개체 (표 4)에 대한 인간 말초 혈액 혈장은 컨버센트 바이오 (Conversant Bio) 또는 플라스마랩 인터내셔널 (PlasmaLab International, 미국 워싱턴주 에버렛)로부터 얻고, 사용시까지 -80°C에서 0.5 ml 또는 1 ml 분취액으로 보관하였다. 전신 홍반성 루푸스로 임상 진단된 4명의 개체에 대한 인간 말초 혈액 혈장은 컨버센트 바이오로부터 얻고, 사용시까지 -80°C에서 0.5 ml 분취액으로 보관하였다. 냉동 혈장 분취액은 사용 직전에 벤치-탑에서 해동하였다. 제조사의 프로토콜에 따라 QiaAMP 순환 핵산 키트 (퀘이아젠)를 사용하여 2 ml의 각각의 혈장 샘플로부터 순환 무세포 DNA를 정제하였다. DNA를 Qubit 형광 측정기 (인비트로젠)로 정량하였다. 샘플의 하위세트에서 cfDNA 수율을 확인하기 위해, 정제된 DNA를 다중카피의 인간 Alu 서열을 표적으로 하는 맞춤 qPCR 검정으로 추가로 정량하였고; 두 추정치는 일치하는 것으로 판명되었다.

[0190]

매치된 종양 유전자형이 이용가능하지 않았기 때문에, 각각의 샘플은 높은 비율의 종양-유도된 cfDNA를 포함할 가능성이 있는 하위세트를 확인하기 위해 다음과 같은 이수성의 2개의 측정 기준으로 스코어링되었다. 첫째, 각각의 염색체로부터 유도된 판독의 예측된 비율로부터의 편차 (도 62a); 둘째, 공통적인 단일 뉴클레오티드 대형성의 패널에 대한 염색체당 대립유전자형 균형 프로파일 (도 62b). 이러한 측정 기준을 기초로 하여, 5명의 개체 (소세포 폐암, 편평세포 폐암, 결장직장 선암종, 간세포 암종 및 유관 상피내 암종 유방암이 있는)로부터 유래된 단일 가닥 라이브러리를 실시예 4의 IH02의 것과 유사한 정도로 서열결정하였다 (표 5; 평균 30배 적용

범위):

<표 5>

[0192] CA01 세트에 포함된 추가의 샘플에 대한 서열결정 통계

샘플 명칭	라이브리 유형	판독	서열결정된 단편	정렬	정렬 Q30	적용범위	추정 증분 %	35-80bp	120- 180bp
IH03	SSP	2x39	53292855	92.66%	72.37%	2.29	15.46%	11.05%	52.34%
IP01 †	DSP	2x101, 2x102	1214536629	97.22%	86.38%	76.11	0.55%	0.08%	62.77%
IP02 †	DSP	2x101, 2x102	855040273	97.16%	87.72%	52.46	0.83%	0.07%	68.10%
IA01	SSP	2x39	53934607	87.42%	68.30%	2.02	22.70%	15.20%	49.77%
IA02	SSP	2x39	42496222	95.42%	76.61%	1.95	4.74%	12.28%	59.00%
IA03	SSP	2x39	51278489	93.12%	71.33%	2.05	25.68%	14.27%	52.57%
IA04	SSP	2x39	50768476	90.30%	70.51%	2.14	7.83%	17.80%	36.76%
IA05	DSP	2x101	194985271	98.80%	90.61%	11.09	12.05%	2.24%	71.67%
IA06	DSP	2x101	171670054	98.90%	90.88%	9.90	5.41%	1.93%	71.26%
IA07	DSP	2x101	208609489	98.67%	90.34%	11.69	11.45%	2.59%	74.84%
IA08	DSP	2x101	193729556	98.81%	90.70%	10.84	11.96%	2.58%	76.24%
IC02	SSP	2x39	57913605	95.07%	75.57%	2.59	5.40%	12.98%	60.00%
IC03	SSP	2x39	63862631	95.78%	75.66%	2.79	8.32%	13.25%	62.20%
IC04	SSP	2x39	55239248	95.47%	76.26%	2.57	8.28%	10.98%	58.48%
IC05	SSP	2x39	39623850	89.80%	69.92%	1.60	9.24%	14.63%	50.33%
IC06	SSP	2x39	59679981	95.57%	74.90%	2.11	3.93%	24.30%	41.46%
IC08	SSP	2x39	46933688	94.38%	74.21%	1.92	5.92%	16.04%	45.25%
IC09	SSP	2x42	59639583	91.22%	71.15%	2.13	6.69%	21.39%	43.50%
IC10	SSP	2x42	53994406	93.73%	73.40%	1.83	2.00%	27.08%	37.62%
IC11	SSP	2x42	59225460	93.25%	72.51%	2.15	5.26%	21.30%	43.33%
IC12	SSP	2x42	57884742	93.52%	74.33%	2.34	2.66%	18.28%	46.58%
IC13	SSP	2x42	71946779	92.94%	72.47%	2.52	2.18%	23.51%	43.97%
IC14	SSP	2x42	61649203	94.54%	73.47%	2.20	3.23%	22.26%	43.37%
IC15	SSP	2x50, 43/42	908512803	95.49%	76.83%	29.77	10.66%	25.42%	38.47%
IC16	SSP	2x42	62739733	92.81%	72.85%	2.47	2.77%	17.71%	48.04%
IC17	SSP	2x50, 2x39	1072374044	96.02%	76.42%	42.08	12.16%	17.08%	50.02%
IC18	SSP	2x39	59976914	87.91%	68.67%	2.24	4.39%	18.85%	44.44%
IC19	SSP	2x39	51447149	89.38%	69.39%	2.02	8.24%	17.30%	46.33%
IC20	SSP	2x50, 2x39	640838540	96.30%	79.11%	23.38	12.43%	25.72%	39.87%
IC21	SSP	2x39	53000679	94.64%	74.57%	1.79	37.39%	29.89%	43.81%
IC22	SSP	2x39	58102606	94.08%	74.08%	2.51	6.24%	13.65%	58.41%
IC23	SSP	2x39	65859970	95.67%	75.67%	2.94	5.34%	11.09%	60.85%
IC24	SSP	43/42	66344431	94.63%	74.46%	2.48	2.00%	22.46%	46.31%
IC25	SSP	43/42	75066833	93.75%	73.66%	2.86	2.24%	21.30%	46.19%
IC26	SSP	43/42	79180860	92.59%	72.32%	2.97	2.93%	22.34%	40.42%
IC27	SSP	43/42	78037377	88.81%	67.04%	2.20	1.50%	31.31%	30.59%
IC28	SSP	43/42	61402081	95.24%	75.74%	2.60	2.46%	18.71%	46.44%
IC29	SSP	2x39	49989522	94.46%	73.36%	1.75	3.03%	25.82%	36.23%
IC30	SSP	2x39	58439504	93.52%	71.19%	1.75	17.35%	29.58%	30.47%
IC32	SSP	43/42	78233981	87.86%	66.80%	2.25	1.79%	30.12%	31.20%
IC33	SSP	43/42	62196185	87.26%	65.71%	1.93	1.93%	27.44%	36.92%
IC34	SSP	43/42	63572169	95.42%	76.74%	2.53	2.35%	19.64%	48.55%

[0193]

샘플 명칭	라이브러리 유형	판독	서열결정된 단편	정렬	정렬 Q30	적용범위	추정 중복 %	35-80bp	120- 180bp
IC35	SSP	43/42	618554393	85.47%	65.90%	18.22	5.23%	28.18%	35.24%
IC36	SSP	43/42	54402943	94.62%	74.73%	2.21	3.32%	17.02%	52.42%
IC37	SSP	2x50, 43/42	1175553677	93.00%	74.46%	38.22	10.15%	28.47%	35.11%
IC38	SSP	43/42	47981963	89.35%	69.45%	1.78	6.47%	18.59%	43.03%
IC39	SSP	43/42	61968854	95.29%	75.57%	2.62	2.54%	14.42%	57.28%
IC40	SSP	2x39	53228209	93.54%	71.69%	1.81	8.85%	24.88%	34.95%
IC41	SSP	43/42	78081655	87.11%	65.25%	2.26	1.61%	27.94%	35.21%
IC42	SSP	2x39	53017317	93.59%	74.33%	2.02	10.74%	19.04%	44.12%
IC43	SSP	43/42	76395478	88.41%	67.21%	2.40	1.56%	26.68%	37.76%
IC44	SSP	43/42	61354307	95.15%	74.88%	2.45	4.34%	19.10%	46.39%
IC46	SSP	2x39	60123123	94.51%	72.23%	2.13	10.37%	15.46%	50.93%
IC47	SSP	2x39	59438172	95.58%	73.84%	2.07	9.33%	21.67%	43.34%
IC48	SSP	43/42	55704417	91.35%	72.79%	2.01	13.87%	22.56%	38.68%
IC49	DSP	2x101	170489015	99.02%	90.53%	11.19	5.93%	2.41%	59.93%
IC50	DSP	2x101	203828224	98.72%	90.28%	10.82	2.83%	4.81%	66.23%
IC51	DSP	2x101	200454421	98.63%	90.53%	11.77	9.50%	2.58%	67.04%
IC52	DSP	2x101	186975845	98.97%	91.25%	11.37	2.57%	0.83%	68.96%

[0194]

SSP, 단일 가닥 라이브러리 제조 프로토콜. DSP, 이중 가닥 라이브러리 제조 프로토콜. †샘플은 이전에 공개되었다 (J.O. Kitzman et al., *Science Translational Medicine* (2012)).

[0196]

표 5는 각각의 샘플에 대해 서열결정된 단편의 총수, 판독 길이, 지도화 품질 역치를 갖거나 갖지 않는 참조 물질에 정렬하는 상기 단편의 백분율, 평균 적용범위, 중복율 및 2개의 길이 빈에서 서열결정된 단편의 비율을 포함하는 서열결정 관련 통계를 표로 제시한다. 단편 길이는 쌍 형성 말단 판독의 정렬로부터 추정하였다. 판독 길이가 짧기 때문에, 전체 단편을 판독한 것으로 가정하여 적용범위를 계산하였다. 추정되는 중복 단편의 수는 단편 종점율을 기초로 하고, 이것은 고도로 정형화된 절단의 존재 하에 진정한 중복율을 과대평가할 수 있다.

[0197]

상기한 바와 같이, FFT는 유전자 본체에 걸친 긴 단편 WPS 값에 대해 수행되었고, 193-199 bp 빈도 범위에서의 평균 강도를 인간 세포주 및 1차 조직에 대한 동일한 76개의 발현 데이터세트에 대해 관찰시켰다. 실시에 4의 건강한 개체로부터의 3개의 샘플 (상위 10개 전부 및 상위 20개 거의 전부의 상관관계가 림프성 또는 골수성 계통에 대한 것임)과는 대조적으로, 가장 높은 순위의 세포주 또는 조직의 많은 수가, 일부 경우에 암 종류와 정렬되는 비-조혈 계통을 나타낸다 (도 61; 표 3). 예를 들어, 환자가 간세포 암종이 존재하는 IC17의 경우, 최상위 상관관계는 간세포 암종 세포주인 HepG2에 대한 것이었다. 환자가 유관 상피내 암종 유방암이 존재하는 IC35의 경우, 최상위 상관관계는 전이성 유방 선암종 세포주인 MCF7에 대한 것이었다. 다른 경우에, 상관관계 순위에서 가장 큰 변화를 보이는 세포주 또는 1차 조직은 암 종류와 정렬되었다. 예를 들어, 환자가 소세포 폐암이 존재하는 IC15의 경우, 상관관계 순위 (-31)의 가장 큰 변화는 소세포 폐암 세포주 (SCLC-21H)에 대한 것이었다. IC20 (폐 편평세포 암종) 및 IC35 (결장직장 선암종)의 경우, 상관관계 계수의 측면에서 림프/골수 세포주를 대체하는 많은 비-조혈 암 세포주가 있었지만, 이들의 특정 암 유형에 대한 정렬은 덜 분명하였다. 이들 암의 특정 분자 프로파일은 76개의 발현 데이터세트 사이에서 잘 표시되지 않았을 가능성이 있다 (예를 들어, 이들 중 어느 것도 폐 편평세포 암종이 아니고; CACO-2는 결장직장 선암종으로부터 유래된 세포주이지만, 고도로 이질적인 것으로 알려져 있다).

[0198]

생물학적 샘플로부터 유래된 cfDNA에 기여하는 다양한 세포 유형 및/또는 조직의 비율을 평가하기 위해 탐욕적인 (greedy) 반복 방법이 사용되었다. 먼저, 그의 참조 지도 (여기서 76개의 RNA 발현 데이터세트에 의해 정의됨)가, 제시된 cfDNA 샘플에 대해 유전자 본체에 걸쳐 WPS 긴 단편 값의 193-199 bp 빈도에서 평균 FFT 강도와 가장 높은 상관관계를 갖는 세포 유형 또는 조직이 확인되었다. 다음으로, 가장 높은 상관관계를 갖는 세포 유형 또는 조직뿐만 아니라 참조 지도의 전체 세트로부터의 각각의 나머지 다른 세포 유형 또는 조직을 포함하는 일련의 "2개의 조직" 선형 혼합 모델이 피팅되었다. 후자의 세트에서, 가장 높은 계수를 갖는 세포 유형 또는 조직은 계수가 절차가 종결되고 그 마지막 조직 또는 세포 유형이 포함되지 않는 1% 미만이 아니라면 기여 인자로 유지되었다. 이 절차는 혼합 모델에 의해 1% 미만으로 기여할 것으로 추정되는 새로 추가된 조직을 기초로 하여 종료될 때까지 "3-조직", "4-조직" 등으로 반복되었다. 혼합 모델은 다음 형식을 취한다:

[0199]

```
argmax_{a,b,c,...} cor(평균_FFT강도_193-199, a*log2Exp조직1 + b*log2조직2 + c*log2조직3 + ... + (1-a-b-c-...)*log2Exp조직N).
```

[0200]

예를 들어, 진행 간세포 암종 환자로부터 유래된 cfDNA 샘플인 IC17의 경우, 상기 절차는 Hep_G2 (28.6%),

HMC.1 (14.3%), REH (14.0%), MCF7 (12.6%), AN3.CA (10.7%), THP.1 (7.4%), NB.4 (5.5%), U.266.84 (4.5%), 및 U.937 (2.4%)를 포함하는 9개의 기여 세포 유형을 예측하였다. 건강한 개체의 혼합물에 대응하는 cfDNA 샘플인 BH01의 경우, 상기 절차는 골수 (30.0%), NB.4 (19.6%), HMC.1 (13.9%), U.937 (13.4%), U.266.84 (12.5%), Karpas.707 (6.5%), 및 REH (4.2%)를 포함하는 7개의 기여 세포 유형 또는 조직을 예측하였다. 주목 할 것은, 암 환자로부터 유래된 샘플인 IC17의 경우, 예측된 가장 높은 기여 비율은 이 cfDNA가 유래한 환자에 존재하는 암 유형 (Hep_G2 및 간세포 암종)과 밀접하게 연관된 세포주에 대응한다는 것이다. 이와 대조적으로, BH01의 경우, 이 방법은 건강한 개체에서 혈장 cfDNA의 주요 공급원인, 주로 조혈과 연관된 조직 또는 세포 유형에만 대응하는 기여를 예측한다.

[0201] 실시예 6: 실시예 4-5의 일반적인 방법

[0202] 샘플

미지의 수의 건강한 개체로부터의 기여를 포함하는 벌크 인간 말초 혈액 혈장을 스템셀 테크놀로지스 (캐나다 브리티쉬 콜럼비아 밴쿠버)로부터 수득하고, 사용할 때까지 -80°C에서 2 ml 분취액으로 보관하였다. 익명의 건강한 공여자의 개별 인간 말초 혈액 혈장을 컨버센트 바이오 (미국 앨라배마주 헌츠빌)에서 얻고, 사용할 때까지 -80°C에서 0.5 ml 분취액으로 보관하였다.

[0204] 임신한 여성 IPO1 및 IPO2로부터의 전혈은 임신 18주 및 13주에 각각 얻고, 이전에 설명한 바와 같이 처리하였다.⁴¹

[0205] 컨버센트 바이오 또는 플라스마랩 인터내셔널 (미국 워싱턴주 에버렛)로부터 IV기 암 (보충 표 4)으로 임상 진단된 52명의 개체에 대한 인간 말초 혈액 혈장을 수득하고, 사용할 때까지 -80°C에서 0.5 ml 또는 1 ml 분취액으로 보관하였다. 전신 홍반성 루푸스로 임상 진단된 4명의 개체에 대한 인간 말초 혈액 혈장을 컨버센트 바이오로부터 얻고, 사용할 때까지 -80°C에서 0.5 ml 분취액으로 보관하였다.

[0206] 혈장 샘플 처리

[0207] 동결된 혈장 분취액을 사용 직전에 벤치-탑에서 해동하였다. 제조사의 프로토콜에 따라 QiaAMP 순환 핵산 키트 (퀘이安康)를 사용하여 2 ml의 각각의 혈장 샘플로부터 순환 무세포 DNA를 정제하였다. DNA를 Qubit 형광 측정기 (인비트로겐)로 정량하였다. 샘플의 하위세트에서 cfDNA 수율을 확인하기 위해, 정제된 DNA를 다중카피 인간 Alu 서열을 표적으로 하는 맞춤 qPCR 검정으로 추가로 정량하였고; 두 평가는 일치하는 것으로 밝혀졌다.

[0208] 이중 가닥 서열결정 라이브러리의 제조

[0209] 바코드 처리된 서열결정 라이브러리는 독점 소유의 일련의 말단 복구, 라이게이션 및 증폭 반응을 포함하는 ThruPLEX-FD 또는 ThruPLEX DNA-seq 48D 키트 (루비콘 지노믹스)로 제조하였다. 0.5 ng 내지 30.0 ng의 cfDNA를 모든 임상 샘플 라이브러리의 투입물로 사용하였다. 모든 샘플에 대한 라이브러리 증폭은 과다 증폭을 피하기 위해 실시간 PCR에 의해 모니터링되었고, 전형적으로 4-6 사이클 후에 종결되었다.

[0210] 단일 가닥 서열결정 라이브러리의 제조

[0211] 어댑터 2는 4.5 μ l TE (pH 8), 0.5 μ l 1M NaCl, 10 μ l의 500 μ M 올리고 어댑터 2.1 및 10 μ l의 500 μ M 올리고 어댑터 2.2를 조합하고 95°C에서 10초 동안 인큐베이션하고 0.1°C/s의 속도로 14°C까지 온도를 낮추어 제조하였다. 정제된 cfDNA 단편은 2x CircLigase II 완충제 (에피센터 (Epicentre)), 5 mM MnCl₂, 및 1U FastAP 알칼리성 포스파타제 (씨모 피셔)를 20 μ l 반응 부피로 0.5-10 ng의 단편과 조합하고, 30분 동안 37°C에서 인큐베이션함으로써 텔인산화하였다. 이어서, 단편을 95°C에서 3분 동안 가열함으로써 변성시키고, 즉시 냉조로 옮겼다. 상기 반응물에 40 μ l의 총 부피를 위해 비오틴-접합된 어댑터 올리고 CL78 (5 pmol), 20% PEG-6000 (w/v) 및 200U CircLigase II (에피센터)를 보충하고, 60°C에서 회전시키면서 인큐베이션하고, 95°C로 3분 동안 가열한 후, 냉조에 넣었다. 각각의 샘플에 대해, 20 μ l MyOne C1 비드 (라이프 테크놀로지스)를 비드 결합 완충제 (BBB) (10 mM 트리스-HCl [pH 8], 1 M NaCl, 1 mM EDTA [pH 8], 0.05% Tween-20 및 0.5% SDS)로 2회 세척하고, 250 μ l BBB에 재현탁하였다. 어댑터-라이게이션된 단편을 실온에서 60분 동안 회전시켜 비드에 결합시켰다. 비드를 자성 선반 (magnetic rack)에 모으고, 상청액을 버렸다. 비드를 500 μ l 세척 완충제 A (WBA) (10 mM 트리스-HCl [pH 8], 1 mM EDTA [pH 8], 0.05% Tween-20, 100 mM NaCl, 0.5% SDS)로 1회, 500 μ l 세척 완충제 B (WBB) (10 mM 트리스-HCl [pH 8], 1 mM EDTA [pH 8], 0.05% Tween-20, 100 mM NaCl)로 1회 세척하였다. 비드를 50 μ l의 반응 부피에서 1X 등온 증폭 완충제 (NEB), 2.5 μ M 올리고 CL9, 250 μ M (각각의) dNTP 및 24U Bst 2.0 DNA 폴리머라제 (NEB)와 조합하고, 조심스럽게 진탕하면서 온도를 1°C/분으로 15°C로부터 37°C

로 상승시켜 완만하게 인큐베이션하고, 37°C에서 10분 동안 유지하였다. 자성 선반에 수집한 후, 비드를 200 μ l WBA로 1회 세척하고, 200 μ l의 염격한 세척 버퍼 (SWB) (0.1X SSC, 0.1% SDS)에 재현탁하고, 3분 동안 45°C에서 인큐베이션하였다. 비드를 다시 모아 200 μ l WBB로 1회 세척하였다. 이어서, 비드를 1X CutSmart 완충제 (NEB), 0.025% Tween-20, 100 μ M (각각의) dNTP 및 5U T4 DNA 폴리미라제 (NEB)와 조합하고, 실온에서 30분 동안 조심스럽게 진탕하면서 인큐베이션하였다. 비드를 상기한 바와 같이 각각의 WBA, SWB 및 WBB로 1회 세척하였다. 이어서, 비드를 1X CutSmart 완충제 (NEB), 5% PEG-6000, 0.025% Tween-20, 2 μ M 이중 가닥 어댑터 2 및 10U T4 DNA 리가제 (NEB)와 혼합하고, 실온에서 2시간 동안 조심스럽게 진탕하면서 인큐베이션하였다. 상기한 바와 같이 비드를 각각의 WBA, SWB 및 WBB로 1회 세척하고, 25 μ l TET 완충제 (10 mM 트리스-HCl [pH 8], 1 mM EDTA [pH 8], 0.05% Tween-20)에 재현탁하였다. 제2 가닥은 95°C로 가열하여 비드로부터 용리되었고, 자성 선반에 비드를 모으고, 상청액을 새튜브로 옮겼다. 모든 샘플에 대한 라이브러리 증폭은 과다증폭을 방지하기 위해 실시간 PCR로 모니터링하였고, 라이브러리당 평균 4 내지 6 사이클을 필요로 하였다.

[0212] 서열결정

[0213] 모든 라이브러리는 HiSeq 2000 또는 NextSeq 500 기기 (일루미나)에서 서열결정하였다.

[0214] 1차 서열결정 데이터 처리

[0215] 바코드 처리된 쌍 형성 말단 (PE) 일루미나 서열결정 데이터를 분할하여 바코드 서열에서 하나의 치환을 허용하였다. 판독 길이가 같거나 더 짧은 판독은 컨센서스로 언급되고, 어댑터는 트리밍되었다. 남아있는 컨센서스 단일 말단 판독 (SR) 및 개별 PE 판독은 BWA v0.7.10에 실행된 ALN 알고리즘을 사용하여 인간 참조 게놈 서열 (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence/로부터 다운로드되는 GRCh37, 1000 Genomes 2 단계 기술 참조)에 정렬되었다. PE 판독은 판독 쌍의 모호한 배치를 해결하거나 또는 하나의 배치된 판독 말단의 위치 주위에서 보다 민감한 정렬 단계에 의해 누락된 정렬을 구제하기 위해 BWA SAMPE로 추가로 처리되었다. 정렬된 SR 및 PE 데이터는 SAMtools API를 사용하여 분류된 BAM 형식으로 직접 전환되었다. 샘플의 BAM 파일은 레인 및 서열결정 실행에 걸쳐 병합되었다.

[0216] 품질 관리는 FastQC (v0.11.2)를 사용하여 수행하고, 라이브러리 복잡성 평가 (Picard tools v1.113)를 얻고, 어댑터 이량체의 비율, 추정된 라이브러리 삽입체 크기의 분석, 외부 판독 종료시의 뉴클레오티드 및 디뉴클레오티드 빈도를 결정할 뿐만 아니라, 각각의 라이브러리의 지도화 품질 분포를 조사하였다.

[0217] 시뮬레이션된 판독 데이터세트

[0218] 정렬된 서열결정 데이터는 인간 참조 물질 (GRC37h)의 모든 주요 염색체에 대해 시뮬레이션하였다 (45 bp보다 짧으면 SR, 45 bp보다 짧지 않으면 PE). 이 목적을 위해, 디뉴클레오티드 빈도는 두 판독 말단 및 두 가닥 방향 모두에 대해 실제 데이터로부터 결정되었다. 디뉴클레오티드 빈도는 또한 두 가닥 모두에서 참조 게놈에 대해 기록되었다. 또한, 실제 데이터의 삽입체 크기 분포는 1-500 bp 범위에 대해 추출되었다. 판독은 주요 참조 염색체의 서열을 통해 반복하여 시뮬레이션되었다. 각각의 단계에서 (즉, 원하는 적용범위에 따라 각각의 위치에서 1회 이상), (1) 가닥을 무작위로 선택하고, (2) 개시 디뉴클레오티드가 고려되는지의 여부를 무작위로 결정하기 위해 참조 서열의 빈도에 대한 실제 데이터의 디뉴클레오티드 빈도의 비율을 사용하고, (3) 제공된 삽입체 크기 분포로부터 삽입체 크기가 샘플 조사되고, (4) 생성된 정렬이 보고되는지의 여부를 무작위로 결정하기 위해 말단 디뉴클레오티드의 빈도 비율이 사용된다. 시뮬레이션된 적용범위는 PCR 복제를 제거 후 원 데이터의 적용범위와 일치하였다.

[0219] 적용범위, 판독 개시 및 윈도우 보호 스코어

[0220] 본 개시내용의 데이터는 서열결정 라이브러리 제조에 사용되는 DNA 문자의 2개의 물리적 말단에 대한 정보를 제공한다. 본 발명자들은 SAMtools 어플리케이션 프로그래밍 인터페이스 (API)를 사용하여 BAM 파일로부터 상기 정보를 추출한다. 판독이 시작될 때, 본 발명자들은 두 판독이 동일한 염색체에 정렬되고 판독이 반대 방향을 갖는 PE 데이터의 외부 정렬 좌표를 이용한다. PE 데이터가 어댑터 트리밍을 통해 단일 판독 데이터로 전환된 경우, 본 발명자들은 SR 정렬의 양 말단 좌표를 판독 개부로 간주한다. 적용범위를 위해, 본 발명자들은 이들 말단 위치를 포함하여 2개의 (추정된) 문자 말단 사이의 모든 위치를 고려한다. 본 발명자들은 윈도우 크기 k의 윈도우 보호 스코어 (WPS)를 윈도우에 걸친 문자 수로부터 윈도우에 포함된 임의의 염기에서 시작하는 문자 수를 차감한 값으로 규정한다. 본 발명자들은 결정된 WPS를 윈도우 중앙에 지정한다. 35-80 bp 범위 (짧은 분획)의 문자의 경우, 본 발명자들은 16의 윈도우 크기를 사용하고, 120-180 bp (긴 분획)의 문자의 경우 본 발명자들은 120의 윈도우 크기를 사용한다.

[0221] 뉴클레오솜 피크 콜링

[0222] 뉴클레오솜 보호의 국부 최대치는 본 발명자들이 0의 실행 중간값 (1 kb 윈도우)으로 국부 조정하고 사비츠키-골레이 (Savitzky-Golay) 필터 (윈도우 크기 21, 2차 다항식)를 사용하여 평坦하게 한 긴 분획 WPS로부터 콜링된다. 이어서, WPS 트랙을 0 위의 영역으로 세분화한다 (5개까지의 연속적인 위치를 0 아래에 허용함). 생성되는 영역이 50-150 bp 길이인 경우, 본 발명자들은 해당 영역의 중간값을 확인하고, 중간값 초과의 최대 합계의 연속 윈도우를 검색한다. 본 발명자들은 이 윈도우의 시작, 끝 및 중앙 좌표를 보고한다. 피크 대 피크 거리 등은 중앙 좌표로부터 계산된다. 콜의 스코어는 윈도우의 최대값과 그 영역에 이웃하는 2개의 인접 WPS 최소값의 평균 사이의 거리로서 결정된다. 확인된 지역이 150-450 bp 길이이면, 본 발명자들은 동일한 중간값 초과의 연속 접근 방식을 적용하지만, 크기가 50-150 bp인 윈도우만 보고한다. 150-450 bp 영역으로부터 유래된 여러 윈도우의 스코어 계산을 위해, 본 발명자들은 영역 내의 이웃하는 최소값을 0으로 가정한다. 본 발명자들은 50 bp보다 짧고 450 bp보다 긴 영역을 폐기한다.

[0223] 167 bp 단편의 디뉴클레오티드 조성

[0224] 단편 크기 분포의 우세한 피크에 대응하는, 정확히 167 bp의 추정 길이를 갖는 단편을 샘플 내에서 여과하여 복제물을 제거하였다. 디뉴클레오티드 빈도는 하나의 단편 종점의 50 bp 상류에서 시작하여 다른 종점의 50 bp 하류에서 끝나는, 각각의 위치에서 슬라이딩 (sliding) 2 bp 윈도우 및 참조 대립유전자를 사용하여 가닥 인식 (strand-aware) 방식으로 계산되었다. 각각의 위치에서 관찰된 디뉴클레오티드 빈도를 라이브러리 특이적 방식으로 계산된 동일한 절단 편향을 반영하는 시뮬레이션된 판독 세트로부터 결정된 예상된 디뉴클레오티드 빈도와 비교하였다 (상세한 내용은 상기 참조).

[0225] 전사 인자 결합 부위를 둘러싼 WPS 프로파일 및 게놈 특징

[0226] 분석은 컴퓨터에 의해 예측된 전사 인자 결합 부위의 세트를 정의하는 클러스터링된 FIMO (모티프-기반) 간격의 초기 세트로 시작하였다. 클러스터링된 전사 인자의 하위세트 (AP-2-2, AP-2, CTCF_Core-2, E2F-2, EBF1, Ebox-CACCTG, Ebox, ESR1, ETS, IRF-2, IRF-3, IRF, MAFK, MEF2A-2, MEF2A, MYC-MAX, PAX5-2, RUNX2, RUNX-AML, STAF-2, TCF-LEF, YY1)에 대해, 부위의 세트는 실험 데이터를 기초로 하여 활발하게 결합된 전사 인자 결합 부위의 보다 확실한 세트로 정제하였다. 이를 위해, 공개적으로 이용 가능한 ENCODE 데이터 (UCSC로부터 다운로드된 TfbsClusteredV3 세트)로부터 ChIP-seq 실험에 의해 정의된 피크와 겹치는 예측된 결합 부위만 유지되었다.

[0227] 이들 부위를 둘러싸는 윈도우 보호 스코어는 CH01 샘플 및 대응하는 시뮬레이션 둘 모두에 대해 추출되었다. 각각의 부위/특징에 대한 보호 스코어는 각각의 결합 부위의 시작 좌표와 비교하여 각각의 위치에서 계산되고, 합계되었다. CTCF 결합 부위의 플롯은 x-축 상의 0 좌표가 CTCF의 알려진 52 bp 결합 풋프린트의 중앙에 위치하도록 이동하였다. 이어서, 5 kb의 추출된 WPS 신호의 첫 번째 및 마지막 500 bp (우세하게 편평하고 평균 오프셋을 나타냄)의 평균을 원래 신호로부터 차감한다. 긴 단편 신호에 대해서만, 200 bp 윈도우를 사용하여 슬라이딩 윈도우 평균을 계산하고, 원래 신호로부터 차감하였다. 마지막으로, 시뮬레이션을 위한 수정된 WPS 프로파일을 CH01에 대한 수정된 WPS 프로파일로부터 빼내어, 단편 길이 및 라이제이션 편향의 결과인 신호를 수정하였다. 이 최종 프로파일을 플로팅하고, "조정된 WPS"로 칭하였다.

[0228] 전사 개시 부위, 전사 종결 부위, 개시 코돈, 스플라이스 공여자 및 스플라이스 수용자 부위와 같은 게놈 특징은 Ensembl Build 버전 75로부터 입수하였다. 이러한 특징을 둘러싼 조정된 WPS를 계산하고, 상기한 바와 같이 전사 인자 결합 부위에 대해 플로팅하였다.

[0229] CTCF 결합 부위 주위의 뉴클레오솜 간격 및 대응하는 WPS의 분석

[0230] 이 분석에 사용된 CTCF 부위는 먼저 CTCF 결합 부위의 클러스터링된 FIMO 예측 (모티프를 통해 컴퓨터에 의해 예측됨)을 포함하였다. 본 발명자들은 이어서 다음과 같은 상기 세트의 2개의 추가의 하위세트를 생성하였다: 1) ENCODE TfbsClusteredV3 (상기 참조)를 통해 이용 가능한 CTCF ChIP-seq 피크 세트와의 교차, 및 2) 19개의 조직에 걸쳐 활성을 갖는 것으로 실험적으로 관찰된 CTCF 부위의 세트와의 교차.

[0231] 결합 부위의 어느 한쪽에 있는 10개의 뉴클레오솜의 위치를 각각의 부위에 대해 추출하였다. 본 발명자들은 각각의 부위 세트에 대해 뉴클레오솜간 거리 분포를 얻기 위해 모든 인접 뉴클레오솜 사이의 거리를 계산하였다. -1 내지 +1의 뉴클레오솜 간격의 분포는 상당히 변하여, 특히 230-270 bp 범위에서 더 큰 간격으로 이동하였다. 이것은 실제 활성을 보이는 CTCF 부위가 주로 -1과 +1 뉴클레오솜 사이의 더 넓은 간격으로 이동하고, 따라서

길고 짧은 판독 분획 둘 모두에 대한 WPS의 차이가 명백할 수 있음을 시사한다. 따라서, CTCF 부위의 중심에 대한 각각의 위치에서의 짧은 및 긴 단편의 평균 WPS를 추가로 계산하였다. 뉴클레오솜 간격의 효과를 조사하기 위해, 상기 평균은 160 bp 미만, 160-200, 200-230, 230-270, 270-420, 420-460 bp 및 420 bp 초과의 -1 내지 +1 뉴클레오솜 간격의 빈 내에서 취하였다. 이들 간격은 보다 확실하게 활성을 보이는 부위에 대해 지배적인 피크 및 230-270 bp에서 나타나는 피크와 같이 관심 간격을 포착한다.

[0232] DNase I 과민성 부위 (DHS)의 분석

[0233] 모라노 (Maurano) 등 (Science, vol. 337(6099), pp. 1190-95 (2012); 2012년 2 월 13일에 마지막으로 수정된 "all_fdr_0.05_hot" 파일)에 의한 BED 포맷의 349개의 1차 조직 및 세포주 샘플에 대한 DHS 피크는 유니버시티 오브 워싱턴 (University of Washington) Encode 데이터베이스에서 다운로드하였다. 이들 피크 세트 중 233개를 포함하는 태아 조직으로부터 유래된 샘플은 아마도 각각의 조직 샘플 내에 복수개의 세포 유형이 불균등하게 제시되기 때문에, 이들이 조직 유형 내에서 일관성 없이 행동할 때 분석으로부터 제거되었다. 다양한 세포 계통을 나타내는 116개의 샘플을 분석을 위해 보관하였다. 특정 세트의 각각의 DHS 피크의 중간 지점에 대해, CH01 콜세트 (callset)의 가장 가까운 상류 및 하류 콜을 확인하고, 두 콜의 중심 사이의 개놈 거리를 계산하였다. 모든 상기 거리의 분포는 0 내지 500 bp의 거리에 대해 계산된 평탄한 밀도 (smoothed density) 추정을 사용하여 각각의 DHS 피크 콜세트에 대해 가시화되었다.

[0234] 유전자 발현 분석

[0235] 본 연구에서는 휴먼 프로테인 아틀라스 ("ma.csv" 파일)에 의해 44개의 인간 세포주 및 32개의 1차 조직에서 20,344개의 Ensembl 유전자 식별자에 대해 측정된 FPKM 발현 값을 사용하였다. 조직에 걸친 분석을 위해, 3 미만의 비-제로 발현 값을 갖는 유전자는 배제되었다 (19,378개의 유전자가 이 필터를 통과함). 발현 데이터세트에는 FPKM 값에 대해 소수점 아래 첫째 자리 정밀도가 제시되었다. 따라서, 0 표현 값 (0.0)은 0과 0.05 미만의 값 사이의 표현을 나타낸다. 특별히 언급하지 않는 한, 최소 표현 값은 표현 값의 \log_2 -전환 전에 0.04 FPKM으로 설정되었다.

[0236] 평탄한 주기도 및 궤적의 평탄화

[0237] 긴 분획 WPS는 1/500 염기 내지 1/100 염기 사이의 빈도로 고속 푸리에 변환 (FFT, R 통계적 프로그래밍 환경에서의 spec.pgram)을 사용하여 개놈 영역의 주기도를 계산하기 위해 사용되었다. 데이터를 평탄화하고 (3 bp 단위 평탄화제, 최종값에 1/2의 가중치를 주는 이동 평균) 추세를 제거 (예를 들어, 계열의 평균을 빼고 선형 경향을 제거)하는 파라미터가 임의적으로 추가로 사용된다.

[0238] 표시된 경우, R 통계적 프로그래밍 환경에서 실행된 재귀적 시계열 필터를 사용하여 궤적으로부터 고빈도 변동을 제거하였다. 24개의 필터 빈도 ($1/\text{seq}(5, 100, 4)$)가 사용되었고, 초기 값으로서 궤적의 처음 24개의 값이 사용되었다. 생성되는 궤적의 24-값 이동에 대한 조정은 궤적의 마지막 24개의 값을 반복함으로써 이루어졌다.

[0239] FFT 강도와 발현 값의 상관관계

[0240] 120-280 bp 범위에 대한 유전자 발현과 관련하여 평탄한 주기도 (FFT)로부터 결정된 강도 값을 분석하였다. 유전자 발현 값과 주요 뉴클레오솜간 거리 피크 주위의 FFT 강도 사이의 S-형 피어슨 (S-shaped Pearson) 상관관계가 관찰되었다. 193-199 bp 범위에서 명백한 음의 상관관계가 관찰되었다. 그 결과, 평균한 상기 빈도 범위에서의 강도는 \log_2 -전환된 발현 값과 상관관계를 보였다.

[0241] 추가의 실시예

[0242] 실시예 7. 대상체로부터의 생물학적 샘플로부터 무세포 DNA (cfDNA)를 단리하고, 단리된 cfDNA는 복수개의 cfDNA 단편을 포함하는 것인 단계;

[0243] 복수개의 cfDNA 단편의 적어도 일부분과 연관된 서열을 결정하는 단계;

[0244] cfDNA 단편 서열의 함수로서 복수개의 cfDNA 단편의 적어도 일부의 cfDNA 단편 종점에 대한 참조 개놈 내의 개놈 위치를 결정하는 단계; 및

[0245] cfDNA 단편 종점의 적어도 일부의 개놈 위치의 함수로서 cfDNA 단편을 생성하는 조직 및/또는 세포 유형의 적어도 일부를 결정하는 단계;

[0246] 를 포함하는, 대상체에서 cfDNA를 생성하는 조직 및/또는 세포 유형을 결정하는 방법.

[0247] 실시예 8. 실시예 7에 있어서, cfDNA 단편을 생성하는 조직 및/또는 세포 유형의 적어도 일부를 결정하는 단계가 cfDNA 단편 종점의 적어도 일부의 개놈 위치를 하나 이상의 참조 지도와 비교하는 것을 포함하는 것인 방법.

[0248] 실시예 9. 실시예 7 또는 실시예 8에 있어서, cfDNA 단편을 생성하는 조직 및/또는 세포 유형의 적어도 일부를 결정하는 단계가 cfDNA 단편 종점의 적어도 일부의 개놈 위치의 분포에 대해 수학적 변환을 수행하는 것을 포함하는 것인 방법.

[0249] 실시예 10. 실시예 9에 있어서, 수학적 변환이 푸리에 변환을 포함하는 것인 방법.

[0250] 실시예 11. 실시예 7 내지 실시예 10 중 어느 한 실시예에 있어서, 참조 개놈의 적어도 일부 좌표 각각에 대한 스코어를 결정하는 단계를 추가로 포함하고, 여기서 스코어는 적어도 복수개의 cfDNA 단편 종점 및 그의 개놈 위치의 함수로서 결정되고, 관찰된 cfDNA 단편을 생성하는 조직 및/또는 세포 유형의 적어도 일부를 결정하는 단계는 스코어를 하나 이상의 참조 지도와 비교하는 단계를 포함하는 것인 방법.

[0251] 실시예 12. 실시예 11에 있어서, 좌표에 대한 스코어가, 그 좌표가 cfDNA 단편 종점의 위치일 확률을 나타내거나 또는 그와 관련되는 것인 방법.

[0252] 실시예 13. 실시예 8 내지 실시예 12 중 어느 한 실시예에 있어서, 참조 지도가 적어도 하나의 세포 유형 또는 조직으로부터 생성된 DNase I 과민성 부위 데이터세트를 포함하는 것인 방법.

[0253] 실시예 14. 실시예 8 내지 실시예 13 중 어느 한 실시예에 있어서, 참조 지도가 적어도 하나의 세포 유형 또는 조직으로부터 생성된 RNA 발현 데이터세트를 포함하는 것인 방법.

[0254] 실시예 15. 실시예 8 내지 실시예 14 중 어느 한 실시예에 있어서, 참조 지도가 인간 조직 또는 세포가 이종이식된 동물로부터의 cfDNA로부터 작성된 것인 방법.

[0255] 실시예 16. 실시예 8 내지 실시예 15 중 어느 한 실시예에 있어서, 참조 지도가 적어도 하나의 세포 유형 또는 조직으로부터 작성된 염색체 입체형태 지도를 포함하는 것인 방법.

[0256] 실시예 17. 실시예 8 내지 실시예 16 중 어느 한 실시예에 있어서, 참조 지도가 적어도 하나의 세포 유형 또는 조직으로부터 작성된 염색질 접근가능성 지도를 포함하는 것인 방법.

[0257] 실시예 18. 실시예 8 내지 실시예 17 중 어느 한 실시예에 있어서, 참조 지도가 적어도 하나의 참조 대상체로부터 수득된 샘플로부터 수득된 서열 데이터를 포함하는 것인 방법.

[0258] 실시예 19. 실시예 8 내지 실시예 18 중 어느 한 실시예에 있어서, 참조 지도가 질환 또는 장애와 연관된 적어도 하나의 세포 유형 또는 조직에 대응하는 것인 방법.

[0259] 실시예 20. 실시예 8 내지 실시예 19 중 어느 한 실시예에 있어서, 참조 지도가 조직 또는 세포 유형 내의 뉴클레오솜 및/또는 크로마토솜의 위치 또는 간격을 포함하는 것인 방법.

[0260] 실시예 21. 실시예 8 내지 실시예 20 중 어느 한 실시예에 있어서, 참조 지도가 적어도 하나의 세포 유형 또는 조직으로부터 수득된 염색질을 외인성 뉴클레아제 (예를 들어, 미크로코쿠스 뉴클레아제)로 소화시켜 작성되는 것인 방법.

[0261] 실시예 22. 실시예 8 내지 실시예 21 중 어느 한 실시예에 있어서, 참조 지도가 적어도 하나의 세포 유형 또는 조직으로부터의 전위 기반 방법 (예를 들어, ATAC-seq)에 의해 결정된 염색질 접근가능성 데이터를 포함하는 것인 방법.

[0262] 실시예 23. 실시예 8 내지 실시예 22 중 어느 한 실시예에 있어서, 참조 지도가 조직 또는 세포 유형에 대한 DNA 결합 및/또는 DNA 점유 단백질의 위치와 연관된 데이터를 포함하는 것인 방법.

[0263] 실시예 24. 실시예 23에 있어서, DNA 결합 및/또는 DNA 점유 단백질이 전사 인자인 방법.

[0264] 실시예 25. 실시예 23 또는 실시예 24에 있어서, 위치가 가교된 DNA-단백질 복합체의 염색질 면역침전에 의해 결정되는 것인 방법.

[0265] 실시예 26. 실시예 23 또는 실시예 24에 있어서, 위치가 조직 또는 세포 유형과 연관된 DNA를 뉴클레아제 (예를 들어, DNase-I)로 처리하여 결정되는 것인 방법.

[0266] 실시예 27. 실시예 8 내지 실시예 26 중 어느 한 실시예에 있어서, 참조 지도가 뉴클레오솜, 크로마토솜, 또는 조직 또는 세포 유형 내의 다른 DNA 결합 또는 DNA 점유 단백질의 위치 또는 간격과 관련된 생물학적 특징을 포

함하는 것인 방법.

[0267] 실시예 28. 실시예 27에 있어서, 생물학적 특징이 하나 이상의 유전자의 정량적 발현인 방법.

[0268] 실시예 29. 실시예 27 또는 실시예 28에 있어서, 생물학적 특징이 하나 이상의 히스톤 마크의 존재 또는 부재인 방법.

[0269] 실시예 30. 실시예 27 내지 실시예 29 중 어느 한 실시예에 있어서, 생물학적 특징이 뉴클레아제 절단에 대해 과민성인 방법.

[0270] 실시예 31. 실시예 8 내지 실시예 30 중 어느 한 실시예에 있어서, 참조 지도를 작성하기 위해 사용된 조직 또는 세포 유형이 질환 또는 장애를 갖는 대상체로부터의 1차 조직인 방법.

[0271] 실시예 32. 실시예 31에 있어서, 질환 또는 장애가 암, 정상 임신, 임신 합병증 (예를 들어, 이수성 임신), 심근경색, 염증성 장 질환, 전신성 자가면역 질환, 국부 자가면역 질환, 거부 반응을 수반하는 동종이식, 거부 반응을 수반하지 않는 동종이식, 출중, 및 국부 조직 손상으로 이루어진 군으로부터 선택되는 것인 방법.

[0272] 실시예 33. 실시예 8 내지 실시예 30 중 어느 한 실시예에 있어서, 참조 지도를 작성하기 위해 사용된 조직 또는 세포 유형이 건강한 대상체로부터의 1차 조직인 방법.

[0273] 실시예 34. 실시예 8 내지 실시예 30 중 어느 한 실시예에 있어서, 참조 지도를 작성하기 위해 사용된 조직 또는 세포 유형이 불멸화된 세포주인 방법.

[0274] 실시예 35. 실시예 8 내지 실시예 30 중 어느 한 실시예에 있어서, 참조 지도를 작성하기 위해 사용된 조직 또는 세포 유형이 종양으로부터의 생검인 방법.

[0275] 실시예 36. 실시예 18에 있어서, 서열 데이터가 cfDNA 단편 종점의 위치를 포함하는 것인 방법.

[0276] 실시예 37. 실시예 36에 있어서, 참조 대상체가 건강한 대상체인 방법.

[0277] 실시예 38. 실시예 36에 있어서, 참조 대상체가 질환 또는 장애를 갖는 것인 방법.

[0278] 실시예 39. 실시예 38에 있어서, 질환 또는 장애가 암, 정상 임신, 임신 합병증 (예를 들어, 이수성 임신), 심근경색, 염증성 장 질환, 전신성 자가면역 질환, 국부 자가면역 질환, 거부 반응을 수반하는 동종이식, 거부 반응을 수반하지 않는 동종이식, 출중, 및 국부 조직 손상으로 이루어진 군으로부터 선택되는 것인 방법.

[0279] 실시예 40. 실시예 19 내지 실시예 39 중 어느 한 실시예에 있어서, 참조 지도가 조직 또는 세포 유형과 연관된 참조 계놈의 좌표의 적어도 일부분에 대한 참조 스코어를 포함하는 것인 방법.

[0280] 실시예 41. 실시예 40에 있어서, 참조 지도가 스코어의 수학적 변환을 포함하는 것인 방법.

[0281] 실시예 42. 실시예 40에 있어서, 스코어가 조직 또는 세포 유형에 대한 모든 참조 계놈 좌표의 하위세트를 나타내는 것인 방법.

[0282] 실시예 43. 실시예 42에 있어서, 하위세트가 뉴클레오솜 및/또는 크로마토솜의 위치 또는 간격과 연관되는 것인 방법.

[0283] 실시예 44. 실시예 42 또는 실시예 43에 있어서, 하위세트가 전사 개시 부위 및/또는 전사 종결 부위와 연관되는 것인 방법.

[0284] 실시예 45. 실시예 42 내지 실시예 44 중 어느 한 실시예에 있어서, 하위세트가 적어도 하나의 전사 인자의 결합 부위와 연관되는 것인 방법.

[0285] 실시예 46. 실시예 42 내지 실시예 45 중 어느 한 실시예에 있어서, 하위세트가 뉴클레아제 과민성 부위와 연관되는 것인 방법.

[0286] 실시예 47. 실시예 40 내지 실시예 46 중 어느 한 실시예에 있어서, 하위세트가 적어도 하나의 직교 생물학적 특징과 추가로 연관되는 것인 방법.

[0287] 실시예 48. 실시예 47에 있어서, 직교 생물학적 특징이 고발현 유전자와 연관되는 것인 방법.

[0288] 실시예 49. 실시예 47에 있어서, 직교 생물학적 특징이 저발현 유전자와 연관되는 것인 방법.

[0289] 실시예 50. 실시예 41 내지 실시예 49 중 어느 한 실시예에 있어서, 수학적 변환이 푸리에 변환을 포함하는 것

인 방법.

[0290] 실시예 51. 실시예 11 내지 실시예 50 중 어느 한 실시예에 있어서, 적어도 복수개의 스코어의 하위세트가 역치 값 초과의 스코어를 갖는 것인 방법.

[0291] 실시예 52. 실시예 7 내지 실시예 51 중 어느 한 실시예에 있어서, cfDNA 단편 종점의 적어도 일부의 복수개의 게놈 위치의 함수로서 cfDNA를 생성하는 조직 및/또는 세포 유형을 결정하는 단계가 cfDNA 단편 종점의 적어도 일부의 복수개의 게놈 위치의 푸리에 변환 또는 그의 수학적 변환을 참조 지도와 비교하는 것을 포함하는 것인 방법.

[0292] 실시예 53. 실시예 7 내지 실시예 52 중 어느 한 실시예에 있어서, 단리된 cfDNA를 생성하는 결정된 조직 및/또는 세포 유형의 목록을 포함하는 보고서를 작성하는 단계를 추가로 포함하는 방법.

[0293] 실시예 54. 대상체로부터의 생물학적 샘플로부터 무세포 DNA (cfDNA)를 단리하고, 단리된 cfDNA는 복수개의 cfDNA 단편을 포함하는 것인 단계;

[0294] 복수개의 cfDNA 단편의 적어도 일부분과 연관된 서열을 결정하는 단계;

[0295] cfDNA 단편 서열의 함수로서 복수개의 cfDNA 단편의 적어도 일부의 cfDNA 단편 종점에 대한 참조 게놈 내의 게놈 위치를 결정하는 단계;

[0296] cfDNA 단편 종점의 적어도 일부의 게놈 위치의 함수로서 cfDNA를 생성하는 조직 및/또는 세포 유형의 적어도 일부를 결정하는 단계; 및

[0297] cfDNA를 생성하는 결정된 조직 및/또는 세포 유형의 함수로서 질환 또는 장애를 확인하는 단계

[0298] 를 포함하는, 대상체에서 질환 또는 장애를 확인하는 방법.

[0299] 실시예 55. 실시예 54에 있어서, cfDNA를 생성하는 조직 및/또는 세포 유형을 결정하는 단계가 cfDNA 단편 종점의 적어도 일부의 게놈 위치를 하나 이상의 참조 지도와 비교하는 것을 포함하는 것인 방법.

[0300] 실시예 56. 실시예 54 또는 실시예 55에 있어서, cfDNA를 생성하는 조직 및/또는 세포 유형을 결정하는 단계가 복수개의 cfDNA 단편 종점의 적어도 일부의 게놈 위치의 분포에 대해 수학적 변환을 수행하는 것을 포함하는 것인 방법.

[0301] 실시예 57. 실시예 56에 있어서, 수학적 변환이 푸리에 변환을 포함하는 것인 방법.

[0302] 실시예 58. 실시예 54 내지 실시예 57 중 어느 한 실시예에 있어서, 참조 게놈의 적어도 일부의 좌표 각각에 대한 스코어를 결정하는 단계를 추가로 포함하고, 여기서 스코어는 적어도 복수개의 cfDNA 단편 종점 및 그의 게놈 위치의 함수로서 결정되고, 관찰된 cfDNA 단편을 생성하는 조직 및/또는 세포 유형의 적어도 일부를 결정하는 단계는 스코어를 하나 이상의 참조 지도와 비교하는 것을 포함하는 것인 방법.

[0303] 실시예 59. 실시예 58에 있어서, 좌표에 대한 스코어가, 그 좌표가 cfDNA 단편 종점의 위치일 확률을 나타내거나 또는 그와 관련되는 것인 방법.

[0304] 실시예 60. 실시예 55 내지 실시예 59 중 어느 한 실시예에 있어서, 참조 지도가 적어도 하나의 참조 대상체로부터 수득된 샘플로부터 수득되고 질환 또는 장애와 연관된 적어도 하나의 세포 유형 또는 조직에 대응하는 DNase I 과민성 부위 데이터세트, RNA 발현 데이터세트, 발현 데이터, 염색체 입체형태 지도, 염색질 접근가능성 지도, 염색질 단편화 지도 또는 서열 데이터, 및/또는 조직 또는 세포 유형 내의 뉴클레오솜 및/또는 크로마토솜의 위치 또는 간격을 포함하는 것인 방법.

[0305] 실시예 61. 실시예 55 내지 실시예 60 중 어느 한 실시예에 있어서, 참조 지도가 적어도 하나의 세포 유형 또는 조직으로부터의 염색질을 외인성 뉴클레아제 (미크로코쿠스 뉴클레아제)로 소화시켜 작성되는 것인 방법.

[0306] 실시예 62. 실시예 60 또는 실시예 61에 있어서, 참조 지도가 적어도 하나의 세포 유형 또는 조직으로부터의 핵 또는 염색질에 전위 기반 방법 (예를 들어, ATAC-seq)을 적용하여 결정된 염색질 접근가능성 데이터를 포함하는 것인 방법.

[0307] 실시예 63. 실시예 55 내지 실시예 62 중 어느 한 실시예에 있어서, 참조 지도가 조직 또는 세포 유형에 대한 DNA 결합 및/또는 DNA 점유 단백질의 위치와 연관된 데이터를 포함하는 것인 방법.

[0308] 실시예 64. 실시예 63에 있어서, DNA 결합 및/또는 DNA 점유 단백질이 전사 인자인 방법.

[0309] 실시예 65. 실시예 63 또는 실시예 64에 있어서, 위치가 가교된 DNA-단백질 복합체의 염색질 면역침전을 적어도 하나의 세포 유형 또는 조직에 적용하여 결정되는 것인 방법.

[0310] 실시예 66. 실시예 63 또는 실시예 64에 있어서, 위치가 조직 또는 세포 유형과 연관된 DNA를 뉴클레아제 (예를 들어, DNase-I)로 처리하여 결정되는 것인 방법.

[0311] 실시예 67. 실시예 54 내지 실시예 66 중 어느 한 실시예에 있어서, 참조 지도가 뉴클레오솜, 크로마토솜, 또는 조직 또는 세포 유형 내의 다른 DNA 결합 또는 DNA 점유 단백질의 위치 또는 간격과 관련된 생물학적 특징을 포함하는 것인 방법.

[0312] 실시예 68. 실시예 67에 있어서, 생물학적 특징이 하나 이상의 유전자의 정량적 발현인 방법.

[0313] 실시예 69. 실시예 67 또는 실시예 68에 있어서, 생물학적 특징이 하나 이상의 히스톤 마크의 존재 또는 부재인 방법.

[0314] 실시예 70. 실시예 67 내지 실시예 69 중 어느 한 실시예에 있어서, 생물학적 특징이 뉴클레아제 절단에 대해 과민성인 방법.

[0315] 실시예 71. 실시예 55 내지 실시예 70 중 어느 한 실시예에 있어서, 참조 지도를 작성하기 위해 사용된 조직 또는 세포 유형이 질환 또는 장애를 갖는 대상체로부터의 1차 조직인 방법.

[0316] 실시예 72. 실시예 71에 있어서, 질환 또는 장애가 암, 정상 임신, 임신 합병증 (예를 들어, 이수성 임신), 심근경색, 염증성 장 질환, 전신성 자가면역 질환, 국부 자가면역 질환, 거부 반응을 수반하는 동종이식, 거부 반응을 수반하지 않는 동종이식, 출증, 및 국부 조직 손상으로 이루어진 군으로부터 선택되는 것인 방법.

[0317] 실시예 73. 실시예 55 내지 실시예 70 중 어느 한 실시예에 있어서, 참조 지도를 작성하기 위해 사용된 조직 또는 세포 유형이 건강한 대상체로부터의 1차 조직인 방법.

[0318] 실시예 74. 실시예 55 내지 실시예 70 중 어느 한 실시예에 있어서, 참조 지도를 작성하기 위해 사용된 조직 또는 세포 유형이 불멸화된 세포주인 방법.

[0319] 실시예 75. 실시예 55 내지 실시예 70 중 어느 한 실시예에 있어서, 참조 지도를 작성하기 위해 사용된 조직 또는 세포 유형이 종양으로부터의 생검인 방법.

[0320] 실시예 76. 실시예 60에 있어서, 적어도 하나의 참조 대상체로부터 수득된 샘플로부터 수득된 서열 데이터가 cfDNA 단편 종점 확률의 위치를 포함하는 것인 방법.

[0321] 실시예 77. 실시예 76에 있어서, 참조 대상체가 건강한 대상체인 방법.

[0322] 실시예 78. 실시예 76에 있어서, 참조 대상체가 질환 또는 장애를 갖는 것인 방법.

[0323] 실시예 79. 실시예 78에 있어서, 질환 또는 장애가 암, 정상 임신, 임신 합병증 (예를 들어, 이수성 임신), 심근경색, 염증성 장 질환, 전신성 자가면역 질환, 국부 자가면역 질환, 거부 반응을 수반하는 동종이식, 거부 반응을 수반하지 않는 동종이식, 출증, 및 국부 조직 손상으로 이루어진 군으로부터 선택되는 것인 방법.

[0324] 실시예 80. 실시예 60 내지 실시예 79 중 어느 한 실시예에 있어서, 참조 지도가 조직 또는 세포 유형과 연관된 참조 게놈의 적어도 일부분에 대한 cfDNA 단편 종점 확률을 포함하는 것인 방법.

[0325] 실시예 81. 실시예 80에 있어서, 참조 지도가 cfDNA 단편 종점 확률의 수학적 변환을 포함하는 것인 방법.

[0326] 실시예 82. 실시예 80에 있어서, cfDNA 단편 종점 확률이 조직 또는 세포 유형에 대한 모든 참조 게놈 좌표의 하위세트를 나타내는 것인 방법.

[0327] 실시예 83. 실시예 82에 있어서, 하위세트가 뉴클레오솜 및/또는 크로마토솜의 위치 또는 간격과 연관되는 것인 방법.

[0328] 실시예 84. 실시예 82 또는 실시예 83에 있어서, 하위세트가 전사 개시 부위 및/또는 전사 종결 부위와 연관되는 것인 방법.

[0329] 실시예 85. 실시예 82 내지 실시예 84 중 어느 한 실시예에 있어서, 하위세트가 적어도 하나의 전사 인자의 결합 부위와 연관되는 것인 방법.

[0330] 실시예 86. 실시예 82 내지 실시예 85 중 어느 한 실시예에 있어서, 하위세트가 뉴클레아제 과민성 부위와 연

관되는 것인 방법.

[0331] 실시예 87. 실시예 82 내지 실시예 86 중 어느 한 실시예에 있어서, 하위세트가 적어도 하나의 직교 생물학적 특징과 추가로 연관되는 것인 방법.

[0332] 실시예 88. 실시예 87에 있어서, 직교 생물학적 특징이 고발현 유전자와 연관되는 것인 방법.

[0333] 실시예 89. 실시예 87에 있어서, 직교 생물학적 특징이 저발현 유전자와 연관되는 것인 방법.

[0334] 실시예 90. 실시예 81 내지 실시예 89 중 어느 한 실시예에 있어서, 수학적 변환이 푸리에 변환을 포함하는 것인 방법.

[0335] 실시예 91. 실시예 58 내지 실시예 90 중 어느 한 실시예에 있어서, 적어도 복수개의 cfDNA 단편 종점 스코어의 하위세트가 각각 역치 값 초과의 스코어를 갖는 것인 방법.

[0336] 실시예 92. 실시예 54 내지 실시예 91 중 어느 한 실시예에 있어서, cfDNA의 조직(들) 및/또는 세포 유형(들)을 cfDNA 단편 종점의 적어도 일부의 복수개의 게놈 위치의 함수로서 결정하는 단계가 cfDNA 단편 종점의 적어도 일부의 복수개의 게놈 위치의 푸리에 변환 또는 그의 수학적 변환을 참조 지도와 비교하는 것을 포함하는 것인 방법.

[0337] 실시예 93. 실시예 54 내지 실시예 92 중 어느 한 실시예에 있어서, 참조 지도가 질환 또는 장애와 연관된 적어도 하나의 조직에 대응하는 DNA 또는 염색질 단편화 데이터를 포함하는 것인 방법.

[0338] 실시예 94. 실시예 54 내지 실시예 93 중 어느 한 실시예에 있어서, 참조 게놈이 인간과 연관되는 것인 방법.

[0339] 실시예 95. 실시예 54 내지 실시예 94 중 어느 한 실시예에 있어서, 질환 또는 장애를 확인하는 언급을 포함하는 보고서를 작성하는 것을 추가로 포함하는 방법.

[0340] 실시예 96. 실시예 95에 있어서, 보고서가 단리된 cfDNA의 결정된 조직(들) 및/또는 세포 유형(들)의 목록을 추가로 포함하는 것인 방법.

[0341] 실시예 97. 실시예 7 내지 실시예 96 중 어느 한 실시예에 있어서, 생물학적 샘플이 전혈, 말초 혈액 혈장, 소변 또는 뇌 척수액을 포함하거나, 이로 본질적으로 이루어지거나, 또는 이로 이루어지는 것인 방법.

[0342] 실시예 98. (i) 대상체로부터 생물학적 샘플을 얻고, 생물학적 샘플로부터 무세포 DNA (cfDNA)를 단리하고, cfDNA의 라이브러리 구축 및 대량 병렬 서열결정에 의해 분포 (a), (b) 및/또는 (c)를 측정하여 뉴클레오솜 지도를 작성하는 단계;

[0343] (ii) 대조군 대상체 또는 알려진 질환을 갖는 대상체로부터 생물학적 샘플을 얻고, 생물학적 샘플로부터 cfDNA를 단리하고, cfDNA의 라이브러리 구축 및 대량 병렬 서열결정에 의해 분포 (a), (b) 및/또는 (c)를 측정하여 뉴클레오솜 지도의 참조 세트를 작성하는 단계; 및

[0344] (iii) cfDNA로부터 유래된 뉴클레오솜 지도를 뉴클레오솜 지도의 참조 세트와 비교하여 cfDNA를 생성하는 조직 및/또는 세포 유형을 결정하는 단계

[0345] 를 포함하고; 여기서

[0346] (a)는 인간 게놈 내의 임의의 특정 염기쌍이 cfDNA 단편의 말단에서 나타날 가능성의 분포이고;

[0347] (b)는 인간 게놈의 염기쌍의 임의의 쌍이 cfDNA 단편의 한 쌍의 말단으로서 나타날 가능성의 분포이고;

[0348] (c)는 인간 게놈 내의 임의의 특정 염기쌍이 차별적인 뉴클레오솜 점유의 결과로서 cfDNA 단편 내에 나타날 가능성의 분포인,

[0349] 대상체에서 cfDNA를 생성하는 조직 및/또는 세포 유형을 결정하는 방법.

[0350] 실시예 99. (i) 대상체로부터 생물학적 샘플을 얻고, 생물학적 샘플로부터 cfDNA를 단리하고, cfDNA의 라이브러리 구축 및 대량 병렬 서열결정에 의해 분포 (a), (b) 및/또는 (c)를 측정하여 뉴클레오솜 지도를 작성하는 단계;

[0351] (ii) 대조군 대상체 또는 알려진 질환을 갖는 대상체로부터 생물학적 샘플을 얻고, 생물학적 샘플로부터 cfDNA를 단리하고, 미크로코쿠스 뉴클레아제 (MNase), DNase 처리 또는 ATAC-Seq으로 염색질을 소화시켜 유래된 DNA의 라이브러리 구축 및 대량 병렬 서열결정에 의해 분포 (a), (b) 및/또는 (c)를 측정하여 뉴클레오솜 지도의

참조 세트를 작성하는 단계; 및

[0352] (iii) cfDNA로부터 유래된 뉴클레오솜 지도를 뉴클레오솜 지도의 참조 세트와 비교하여 cfDNA를 생성하는 조직 및/또는 세포 유형을 결정하는 단계

[0353] 를 포함하고; 여기서

[0354] (a)는 인간 게놈 내의 임의의 특정 염기쌍이 서열결정된 단편의 말단에서 나타날 가능성의 분포이고;

[0355] (b)는 인간 게놈의 염기쌍의 임의의 쌍이 서열결정된 단편의 한 쌍의 말단으로서 나타날 가능성의 분포이고;

[0356] (c)는 인간 게놈 내의 임의의 특정 염기쌍이 차별적인 뉴클레오솜 점유의 결과로서 서열결정된 단편 내에 나타날 가능성의 분포인,

[0357] 대상체에서 무세포 DNA를 생성하는 조직 및/또는 세포 유형을 결정하는 방법.

[0358] 실시예 100. (i) 대상체로부터 생물학적 샘플을 얻고, 생물학적 샘플로부터 cfDNA를 단리하고, cfDNA의 라이브러리 구축 및 대량 병렬 서열결정에 의해 분포 (a), (b) 및/또는 (c)를 측정하여 뉴클레오솜 지도를 작성하는 단계;

[0359] (ii) 대조군 대상체 또는 알려진 질환을 갖는 대상체로부터 생물학적 샘플을 얻고, 생물학적 샘플로부터 cfDNA를 단리하고, cfDNA의 라이브러리 구축 및 대량 병렬 서열결정에 의해 분포 (a), (b) 및/또는 (c)를 측정하여 뉴클레오솜 지도의 참조 세트를 작성하는 단계; 및

[0360] (iii) cfDNA로부터 유래된 뉴클레오솜 지도를 뉴클레오솜 지도의 참조 세트와 비교하여 임상 병태를 결정하는 단계

[0361] 를 포함하고; 여기서

[0362] (a)는 인간 게놈 내의 임의의 특정 염기쌍이 cfDNA 단편의 말단에서 나타날 가능성의 분포이고;

[0363] (b)는 인간 게놈의 염기쌍의 임의의 쌍이 cfDNA 단편의 한 쌍의 말단으로서 나타날 가능성의 분포이고;

[0364] (c)는 인간 게놈 내의 임의의 특정 염기쌍이 차별적인 뉴클레오솜 점유의 결과로서 cfDNA 단편 내에 나타날 가능성의 분포인,

[0365] 대상체에서 임상 병태를 진단하는 방법.

[0366] 실시예 101.

[0367] (i) 대상체로부터 생물학적 샘플을 얻고, 생물학적 샘플로부터 cfDNA를 단리하고, cfDNA의 라이브러리 구축 및 대량 병렬 서열결정에 의해 분포 (a), (b) 및/또는 (c)를 측정하여 뉴클레오솜 지도를 작성하는 단계;

[0368] (ii) 대조군 대상체 또는 알려진 질환을 갖는 대상체로부터 생물학적 샘플을 얻고, 생물학적 샘플로부터 cfDNA를 단리하고, 미크로코쿠스 뉴클레아제 (MNase), DNase 처리 또는 ATAC-Seq으로 염색질을 소화시켜 유래된 DNA의 라이브러리 구축 및 대량 병렬 서열결정에 의해 분포 (a), (b) 및/또는 (c)를 측정하여 뉴클레오솜 지도의 참조 세트를 작성하는 단계; 및

[0369] (iii) cfDNA로부터 유래된 뉴클레오솜 지도를 뉴클레오솜 지도의 참조 세트와 비교하여 cfDNA의 기원 조직 조성을 결정하는 단계

[0370] 를 포함하고; 여기서

[0371] (a)는 인간 게놈 내의 임의의 특정 염기쌍이 서열결정된 단편의 말단에서 나타날 가능성의 분포이고;

[0372] (b)는 인간 게놈의 염기쌍의 임의의 쌍이 서열결정된 단편의 한 쌍의 말단으로서 나타날 가능성의 분포이고;

[0373] (c)는 인간 게놈 내의 임의의 특정 염기쌍이 차별적인 뉴클레오솜 점유의 결과로서 서열결정된 단편 내에 나타날 가능성의 분포인,

[0374] 대상체에서 임상 병태를 진단하는 방법.

[0375] 실시예 102. 실시예 98 내지 실시예 101 중 어느 한 실시예에 있어서, 뉴클레오솜 지도가

[0376] 생물학적 샘플로부터 단리된 cfDNA를 정제하는 단계;

[0377] 어댑터 라이게이션 및 임의적으로 PCR 증폭에 의해 라이브러리를 구축하는 단계; 및

[0378] 생성된 라이브러리를 서열결정하는 단계

[0379] 예 의해 작성되는 것인 방법.

[0380] 실시예 103. 실시예 98 내지 실시예 101 중 어느 한 실시예에 있어서, 뉴클레오솜 지도의 참조 세트가

[0381] 대조군 대상체로부터의 생물학적 샘플로부터 단리된 cfDNA를 정제하는 단계;

[0382] 어댑터 라이게이션 및 임의적으로 PCR 증폭에 의해 라이브러리를 구축하는 단계; 및

[0383] 생성된 라이브러리를 서열결정하는 단계

[0384] 예 의해 작성되는 것인 방법.

[0385] 실시예 104. 실시예 98 내지 실시예 101 중 어느 한 실시예에 있어서, 분포 (a), (b) 또는 (c), 또는 이들 분포 중 하나의 수학적 변환이 인접 윈도우에서 푸리에 변환에 적용된 후, 뉴클레오솜이 각각의 인접 윈도우 내에서 구조화된 배치를 나타내는 정도를 요약하기 위해 뉴클레오솜 점유와 연관된 빈도 범위에 대한 강도의 정량화가 실시되는 것인 방법.

[0386] 실시예 105. 실시예 98 내지 실시예 101 중 어느 한 실시예에 있어서, 분포 (a), (b) 또는 (c), 또는 이들 분포 중 하나의 수학적 변환에서, cfDNA에 기여하는 세포 유형(들)에서 전사 인자 (TF) 활성의 결과로서 뉴클레오솜 배치를 요약하기 위해 전사 인자 결합 부위 (TFBS)가 TF에 의해 결합될 때 종종 뉴클레오솜이 바로 측면에 위치하는 특정 TF의 TFBS의 바로 근처에서 서열결정 판독 개시 부위가 지도화되는 참조 인간 게놈 내의 부위의 분포를 정량하는 방법.

[0387] 실시예 106. 실시예 98 내지 실시예 101 중 어느 한 실시예에 있어서, 뉴클레오솜 점유 신호가, 다른 게놈 랜드마크, 예컨대 DNaseI 과민성 부위, 전사 개시 부위, 위상 도메인, 다른 후성학적 마크 또는 다른 데이터세트에서 상관된 거동 (예를 들어, 유전자 발현 등)에 의해 규정되는 모든 이러한 부위의 하위세트 주위에서 분포 (a), (b) 및/또는 (c), 또는 이들 분포 중 하나의 수학적 변환으로부터 집계된 신호 중 어느 하나에 따라 요약되는 것인 방법.

[0388] 실시예 107. 실시예 98 내지 실시예 101 중 어느 한 실시예에 있어서, 분포가, 예를 들어 인접 윈도우에서의 또는 대안적으로, 전사 인자 결합 부위, 유전자 모델 특징부 (예를 들어, 전사 개시 부위), 조직 발현 데이터 또는 뉴클레오솜 배치의 다른 상관물에 의해 정의된 게놈의 불연속적인 하위세트에서의 주기도를 정량하는 것과 같은 게놈의 다양한 하위세트 내에서의 뉴클레오솜 배치의 주기적 신호를 집계하거나 요약하기 위해 변환되는 것인 방법.

[0389] 실시예 108. 실시예 98 내지 실시예 101 중 어느 한 실시예에 있어서, 분포가 조직 특이적 데이터, 즉 조직 특이적 DNase I 과민성 부위 부근의 집계된 신호에 의해 정의되는 것인 방법.

[0390] 실시예 109. 실시예 98 내지 실시예 101 중 어느 한 실시예에 있어서, 추가의 뉴클레오솜 지도(들)를 참조 세트와 비교하기 위한 통계적 신호 처리 단계를 추가로 포함하는 방법.

[0391] 실시예 110. 실시예 109에 있어서, 먼저 다양한 샘플 세트의 게놈을 따라 연속적 윈도우 내에서 긴 범위의 뉴클레오솜 배열을 요약한 다음, 샘플을 클러스터링하거나 혼합률 비율을 추정하기 위해 주성분 분석 (PCA)을 수행하는 방법.

[0392] 실시예 111. 실시예 100 또는 실시예 101에 있어서, 임상 병태가 암, 즉 악성 종양인 방법.

[0393] 실시예 112. 실시예 111에 있어서, 생물학적 샘플이, 그 일부 부분이 종양으로부터 유래된 cfDNA를 함유하는 순환 혈장인 방법.

[0394] 실시예 113. 실시예 100 또는 실시예 101에 있어서, 임상 병태가 조직 손상, 심근경색 (심장 조직의 급성 손상), 자가면역 질환 (다양한 조직의 만성 손상), 임신, 염색체 이상 (예를 들어, 삼염색체성) 및 이식 거부로부터 선택되는 것인 방법.

[0395] 실시예 114. 실시예 7 내지 실시예 113 중 어느 한 실시예에 있어서, cfDNA에 기여하는 것으로 결정된 하나 이상의 조직 또는 세포 유형 각각에 대해 비율을 할당하는 것을 추가로 포함하는 방법.

[0396] 실시예 115. 실시예 114에 있어서, 하나 이상의 결정된 조직 또는 세포 유형 각각에 할당된 비율이 적어도 부

분적으로 상관관계 또는 건강한 대상체 또는 대상체들로부터의 cfDNA에 비해 증가된 상관관계의 정도에 기초하는 것인 방법.

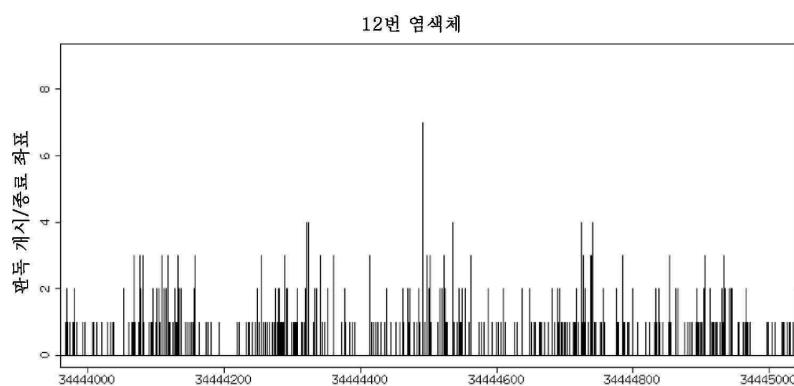
[0397] 실시예 116. 실시예 114 또는 실시예 115에 있어서, 상관관계의 정도가 적어도 부분적으로 생물학적 샘플로부터의 cfDNA 단편 종점의 분포의 수학적 변환과 결정된 조직 또는 세포 유형과 연관된 참조 지도의 비교에 기초하는 것인 방법.

[0398] 실시예 117. 실시예 114 내지 실시예 116 중 어느 한 실시예에 있어서, 하나 이상의 결정된 조직 또는 세포 유형 각각에 할당된 비율이 혼합물 모델에 기초하는 것인 방법.

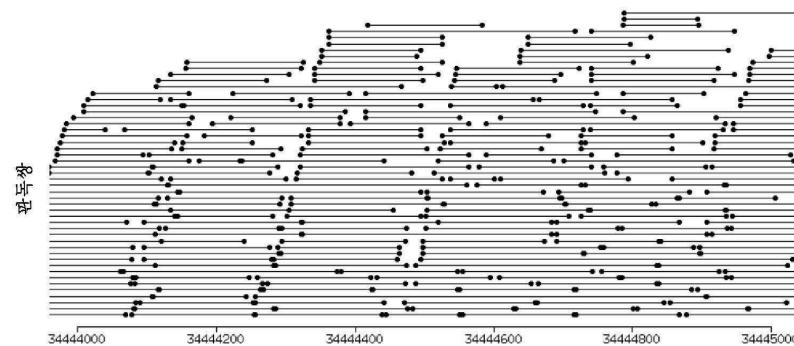
[0399] 상기한 내용으로부터, 본 발명의 구체적인 실시예가 예시를 위해 설명되었지만, 본 발명의 범위를 벗어나지 않는 다양한 변형이 이루어질 수 있음을 이해할 것이다. 따라서, 본 발명은 첨부된 청구범위를 제외하고는 이들로 제한되지 않는다.

도면

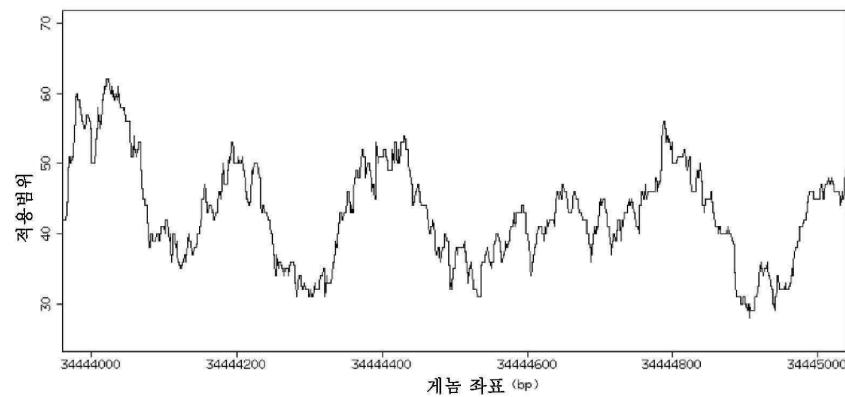
도면 1a



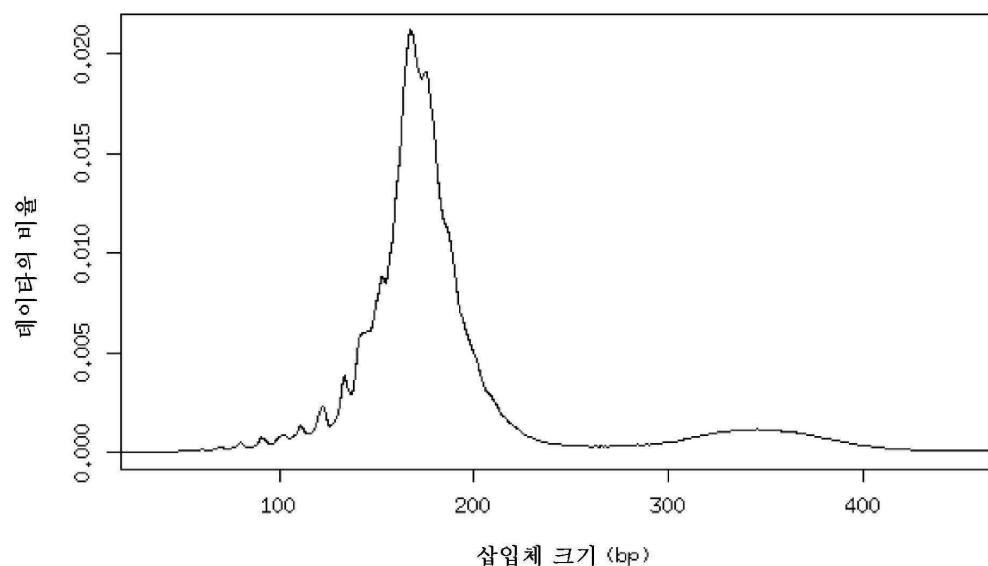
도면 1b



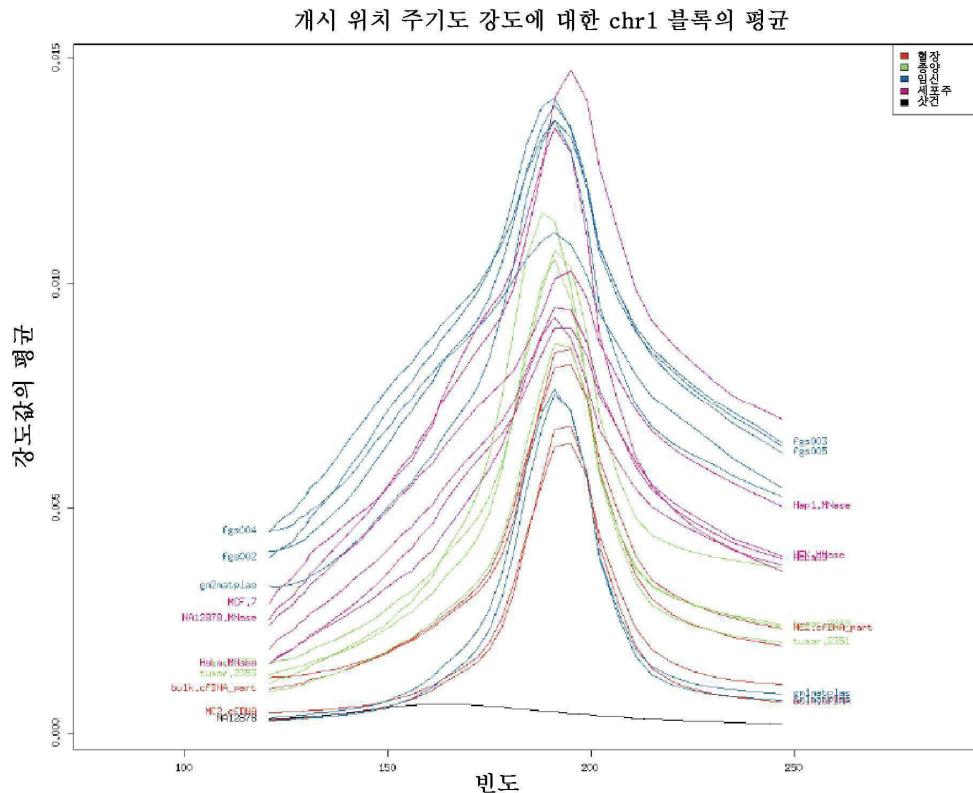
도면1c



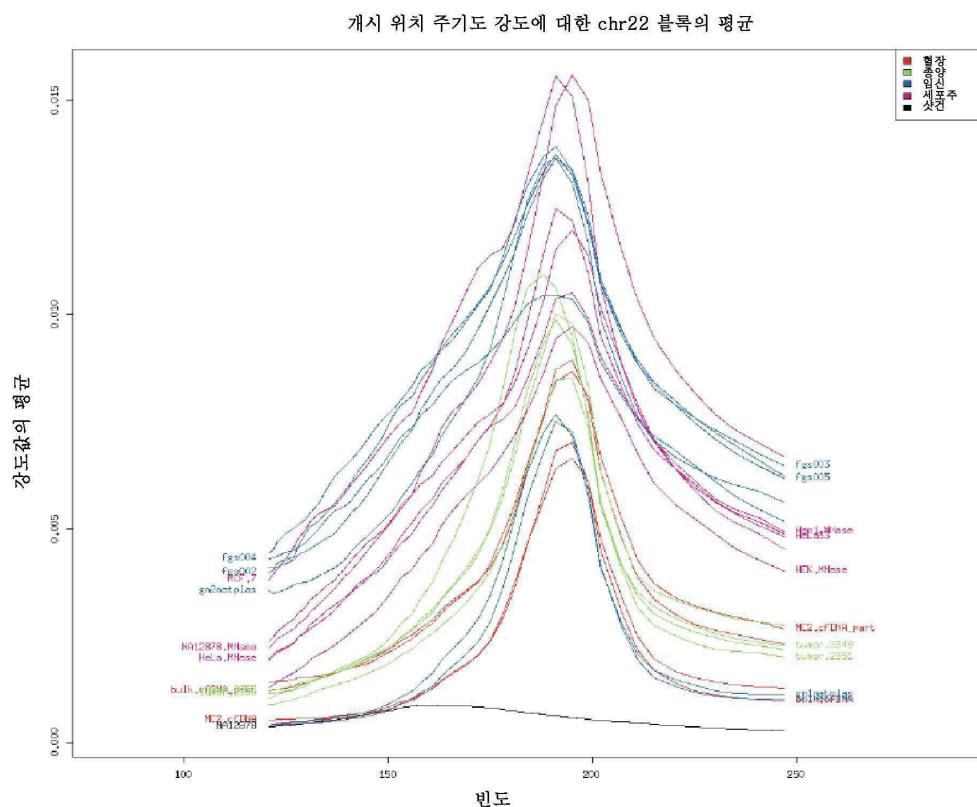
도면2



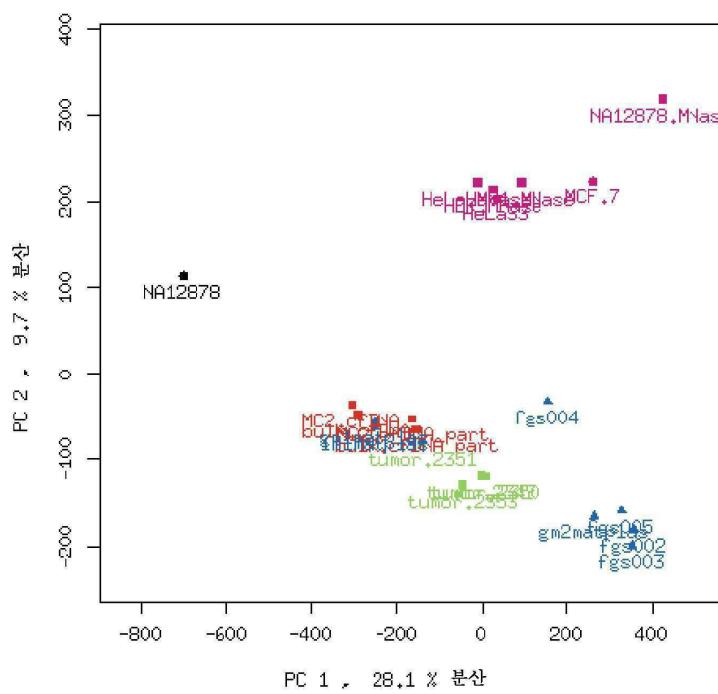
도면3a



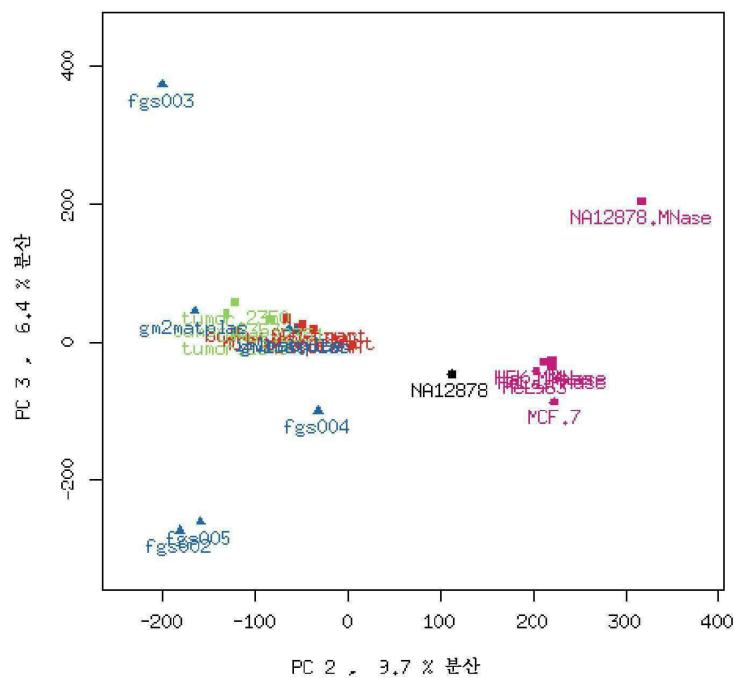
도면3b



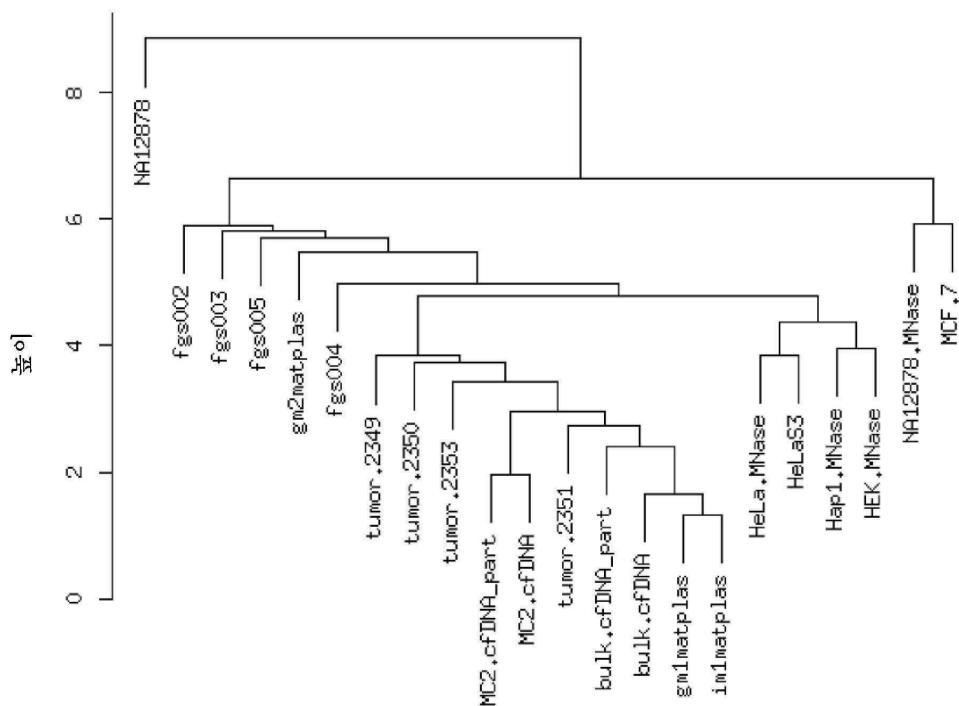
도면4a



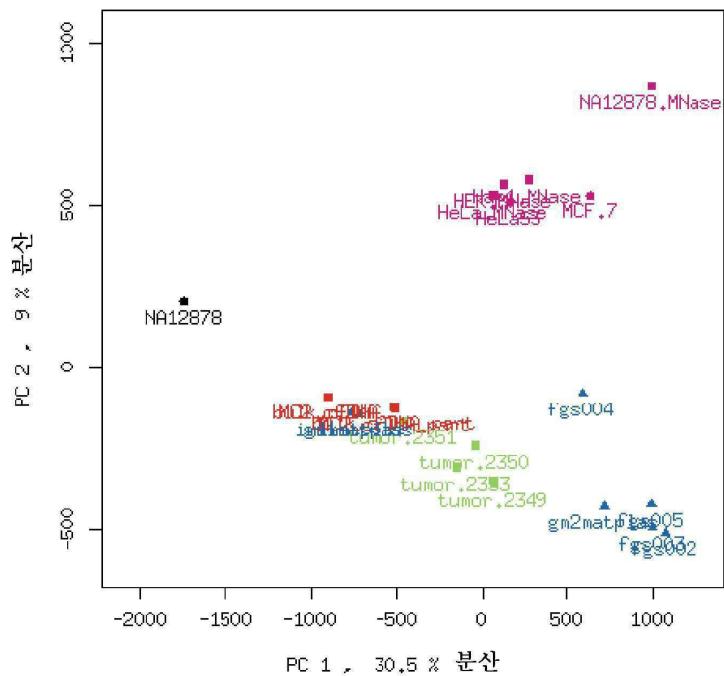
도면4b



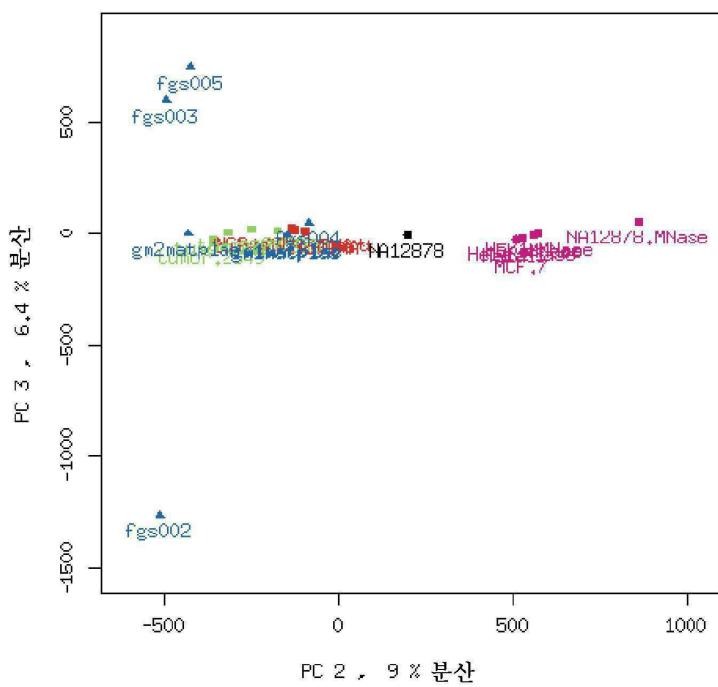
도면5



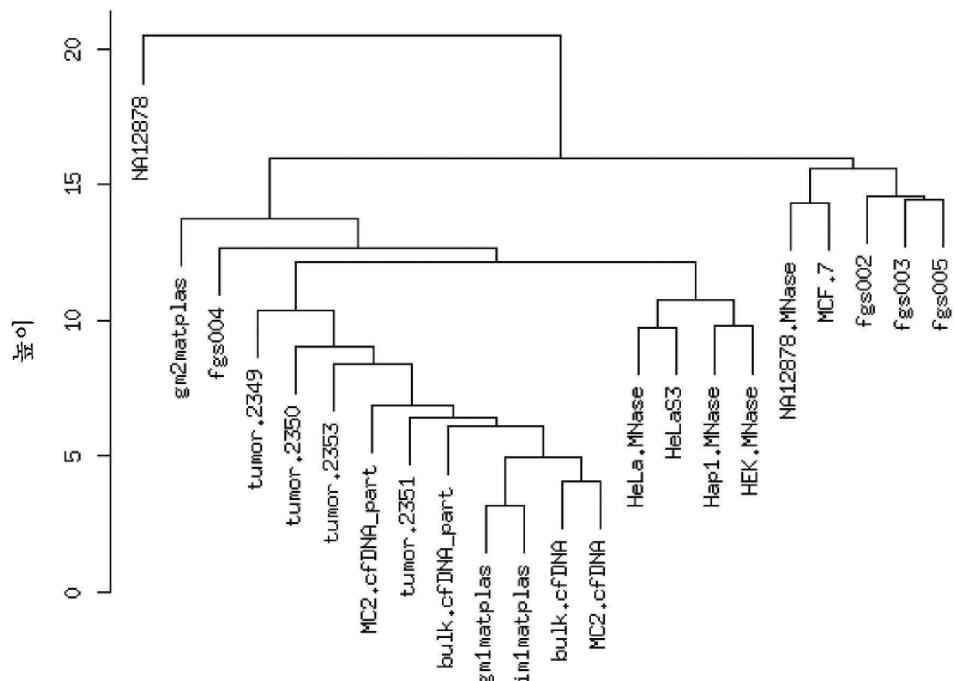
도면6a



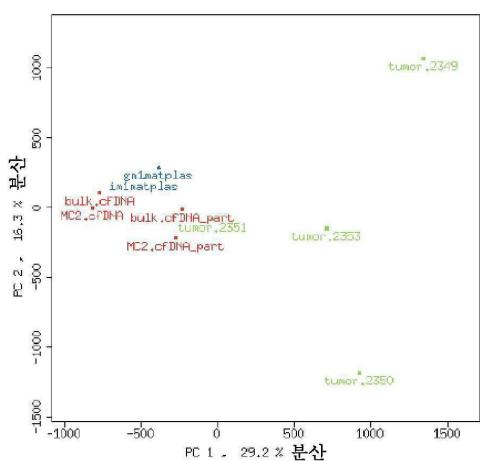
도면6b



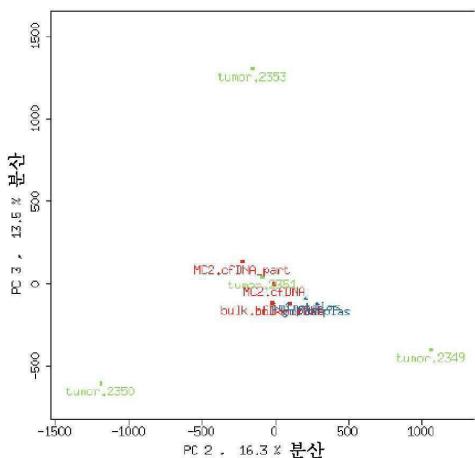
도면7



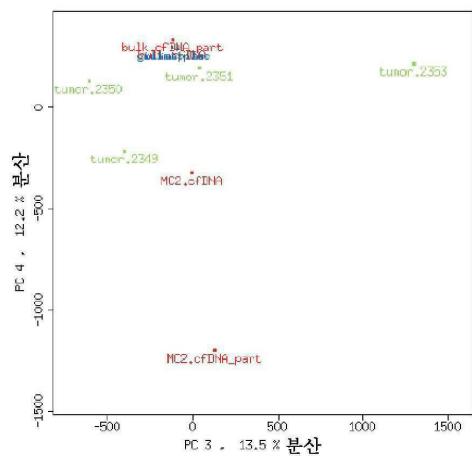
도면8a



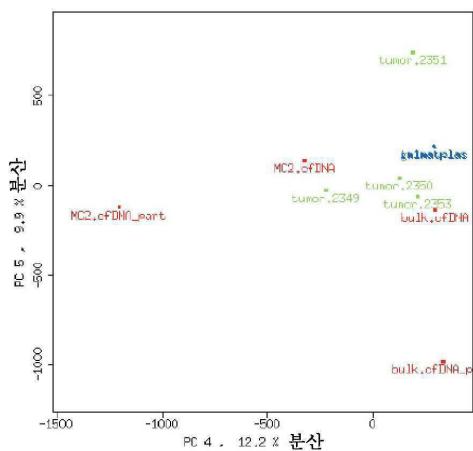
도면8b



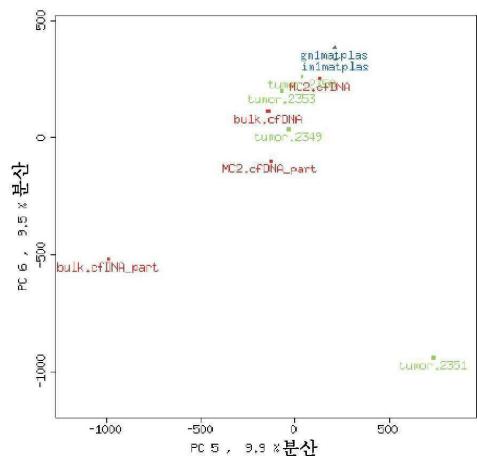
도면8c



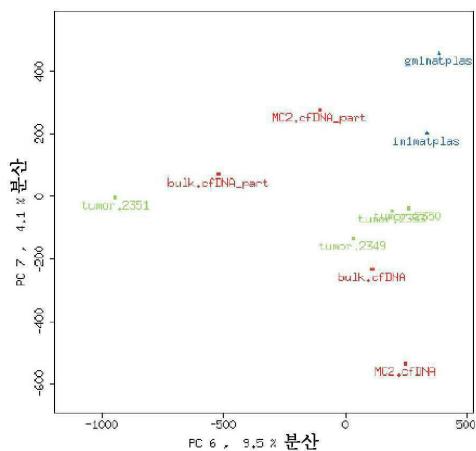
도면8d



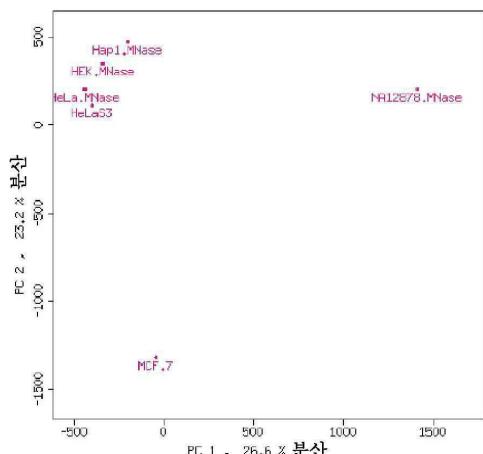
도면8e



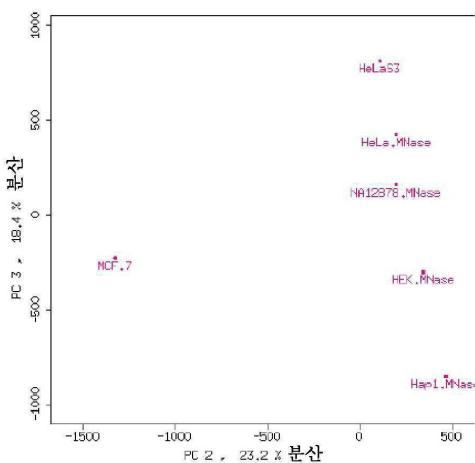
도면8f



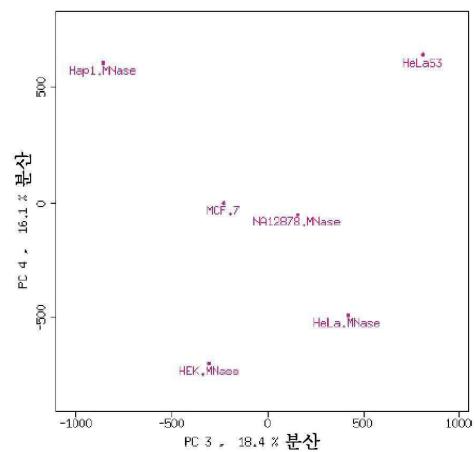
도면9a



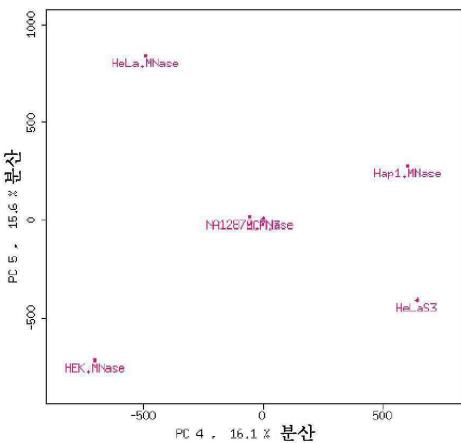
도면9b



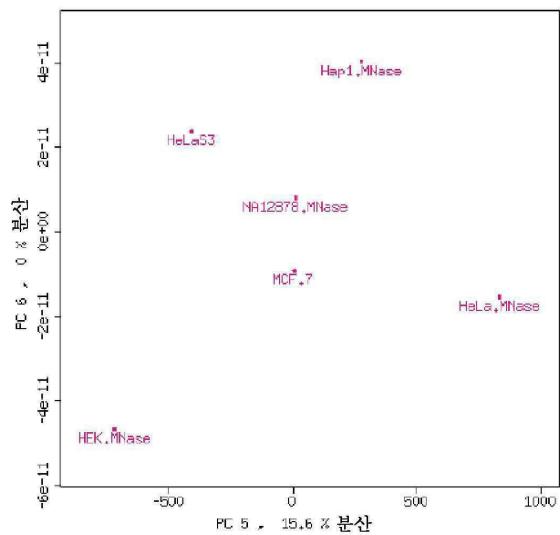
도면9c



도면9d

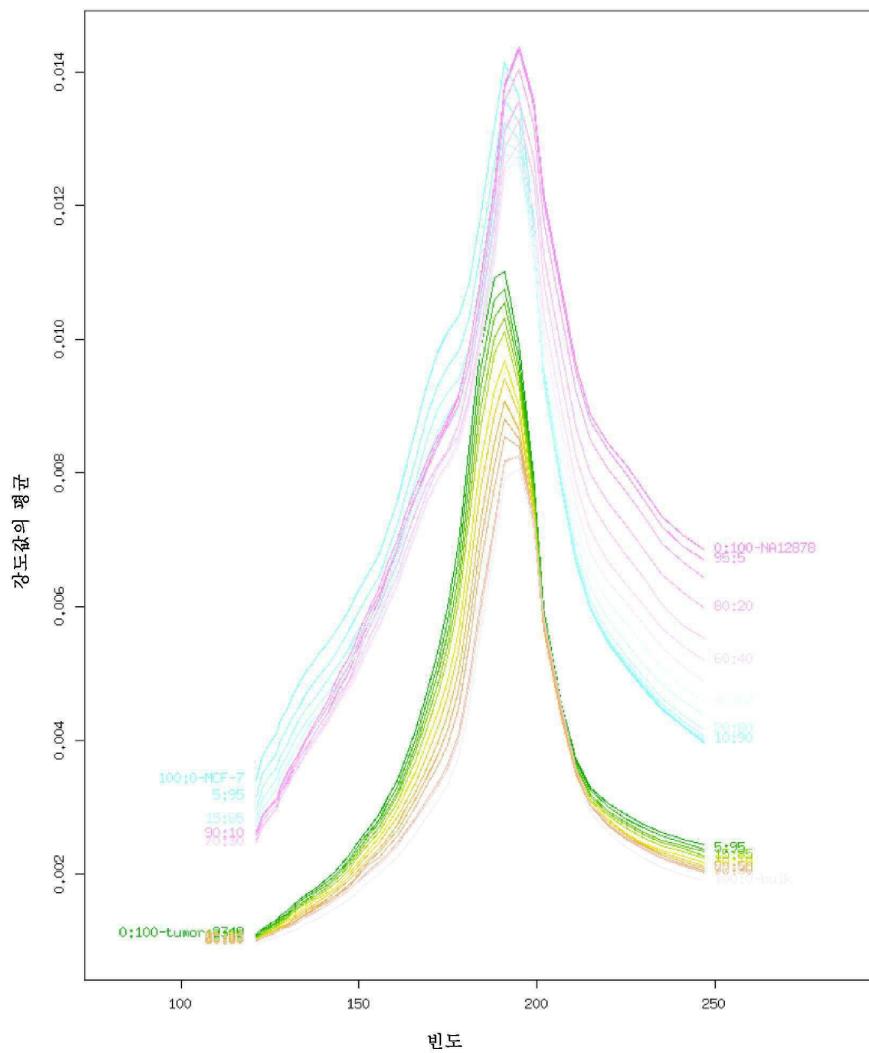


도면9e

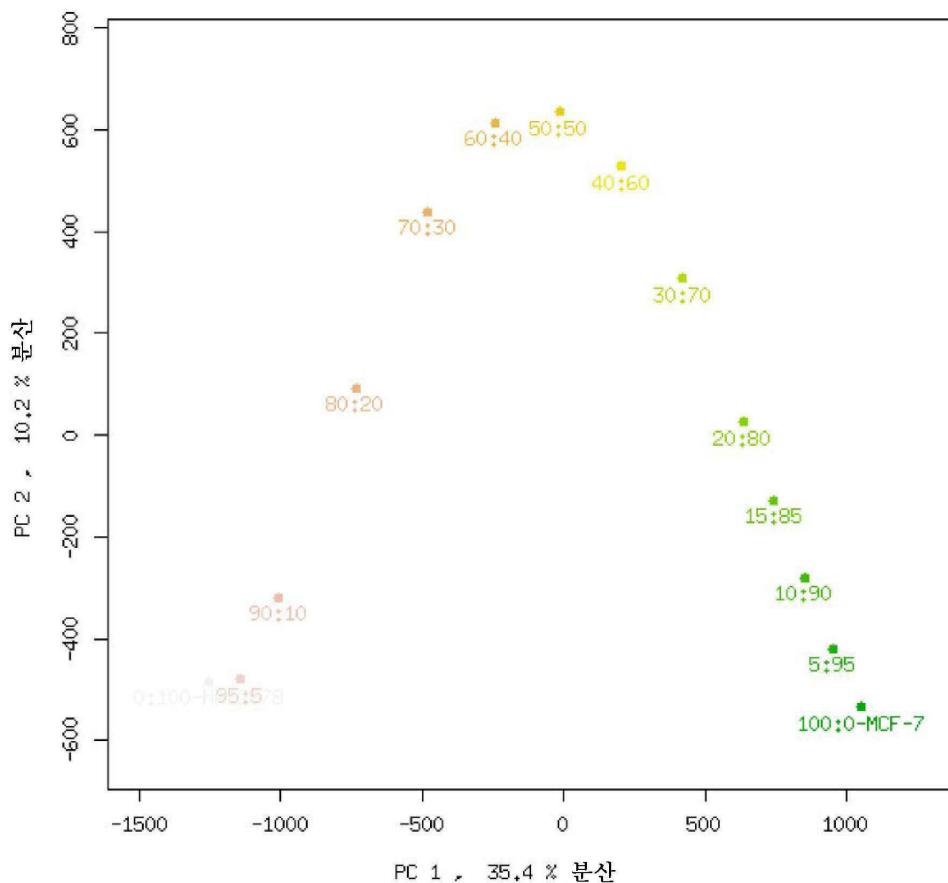


도면10

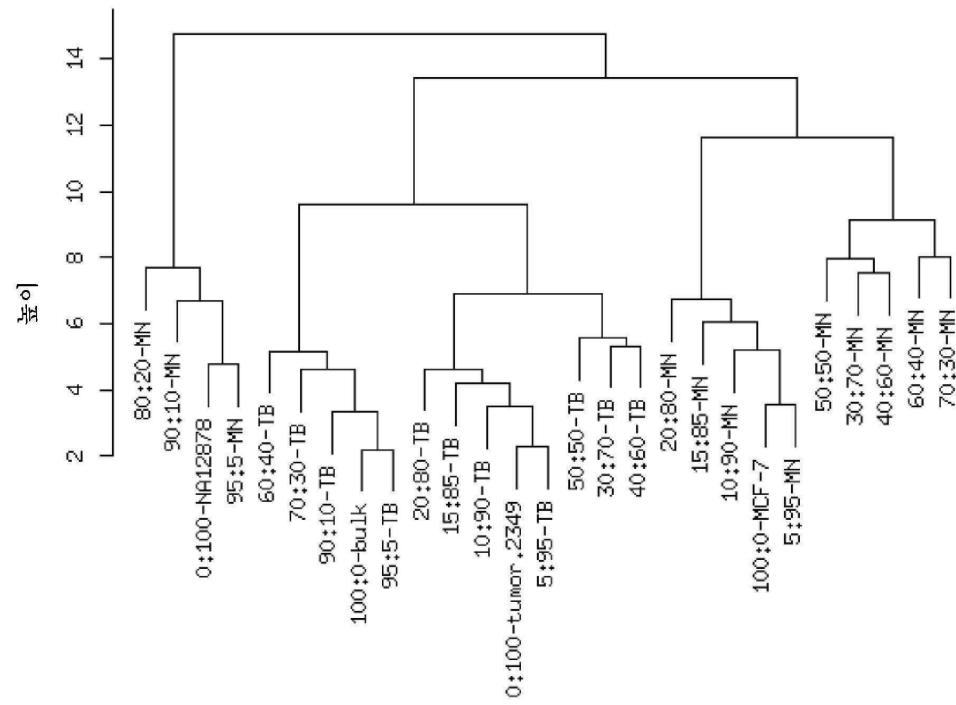
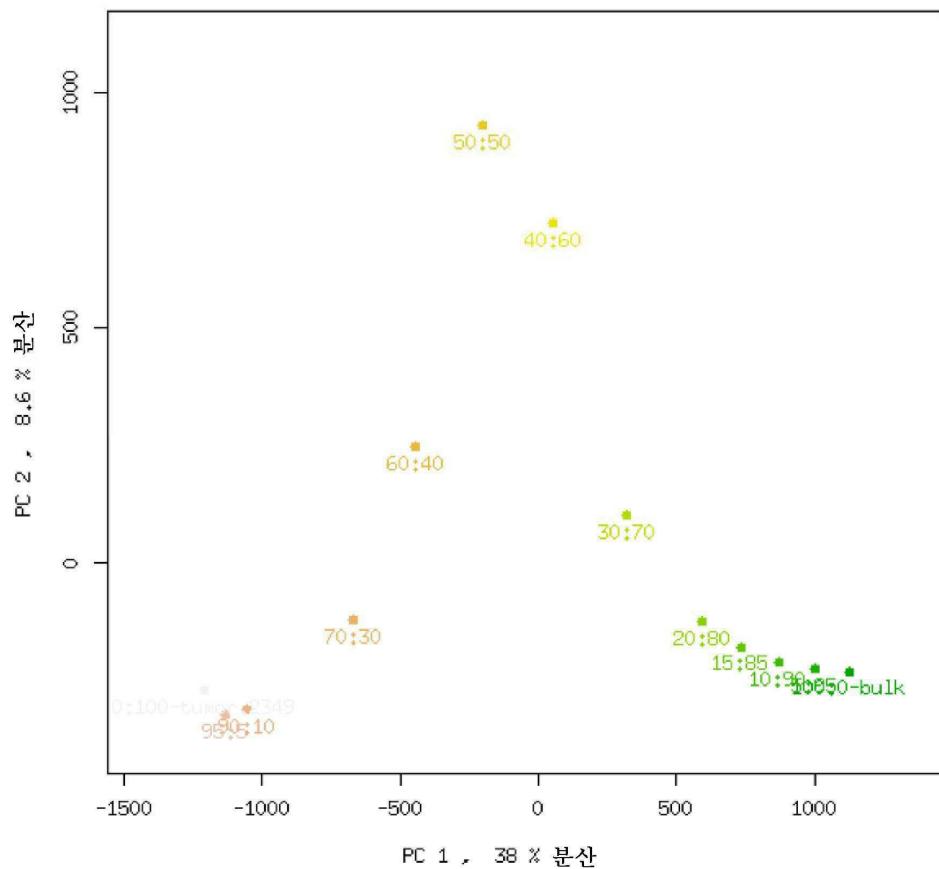
개시 위치 주기도 강도에 대한 chr11 블록의 평균



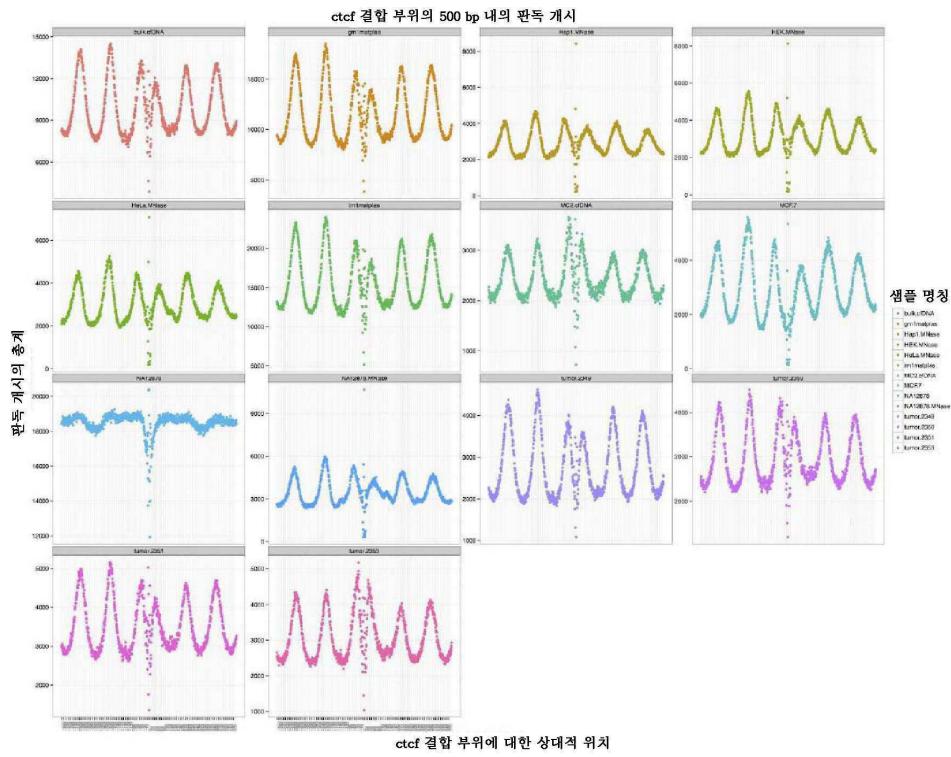
도면11



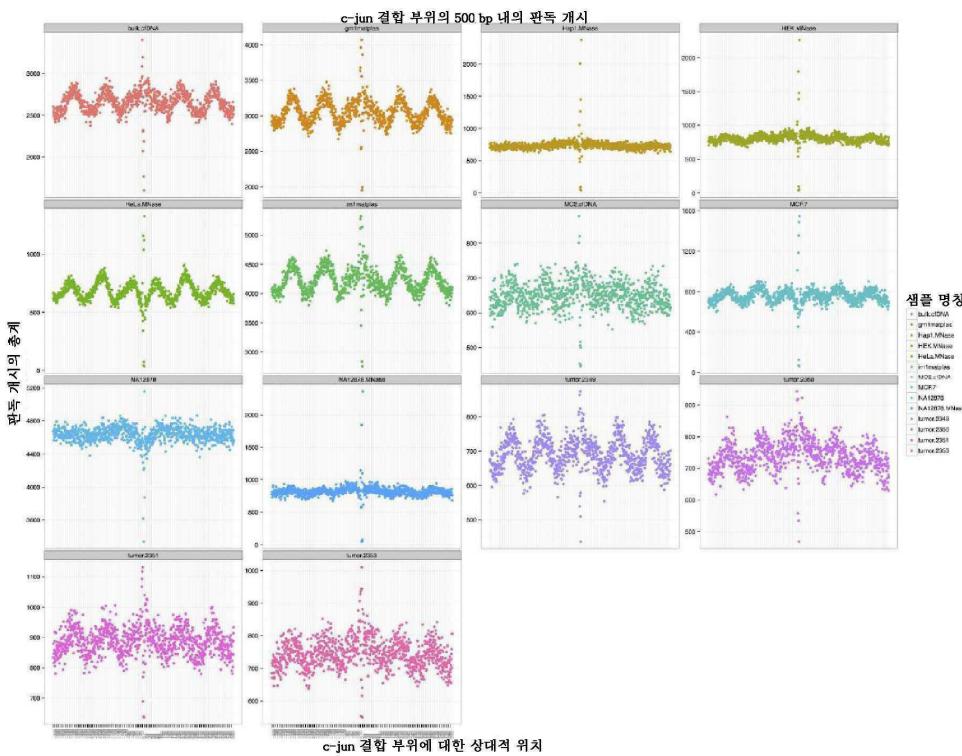
도면12



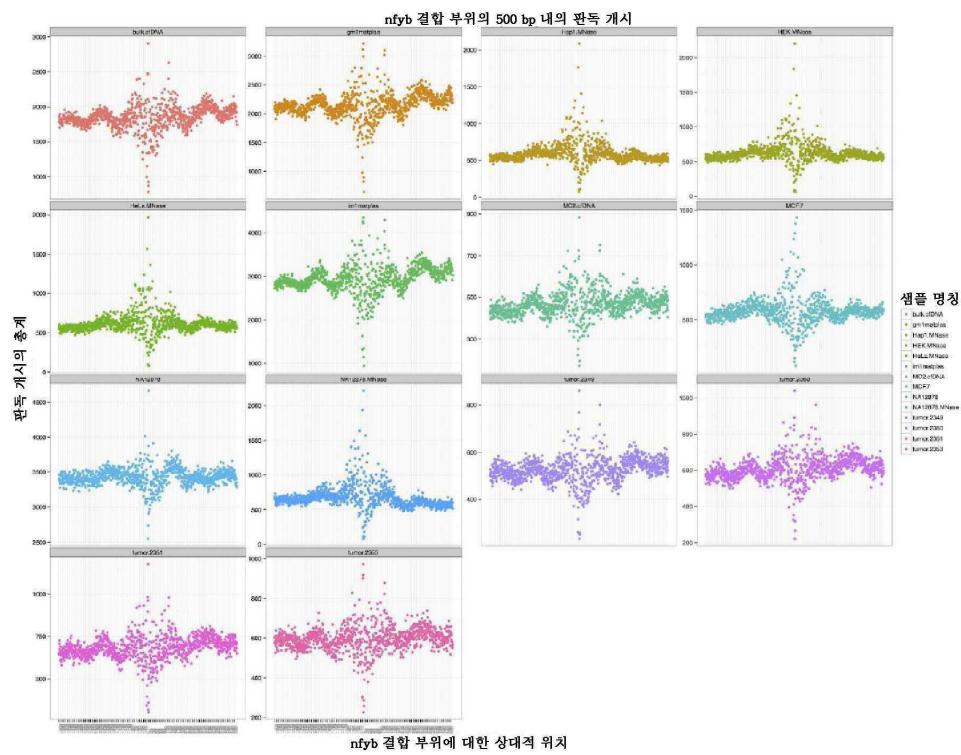
도면14



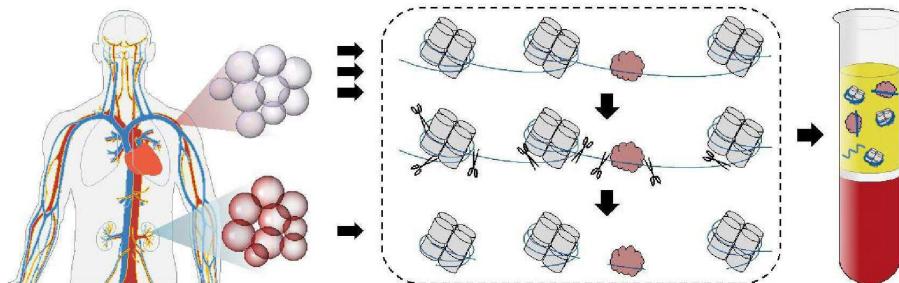
도면15



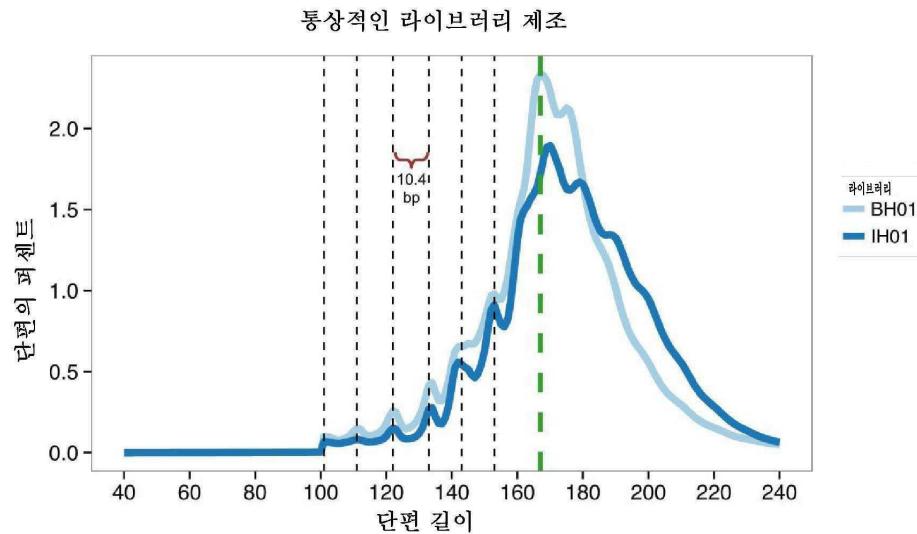
도면16



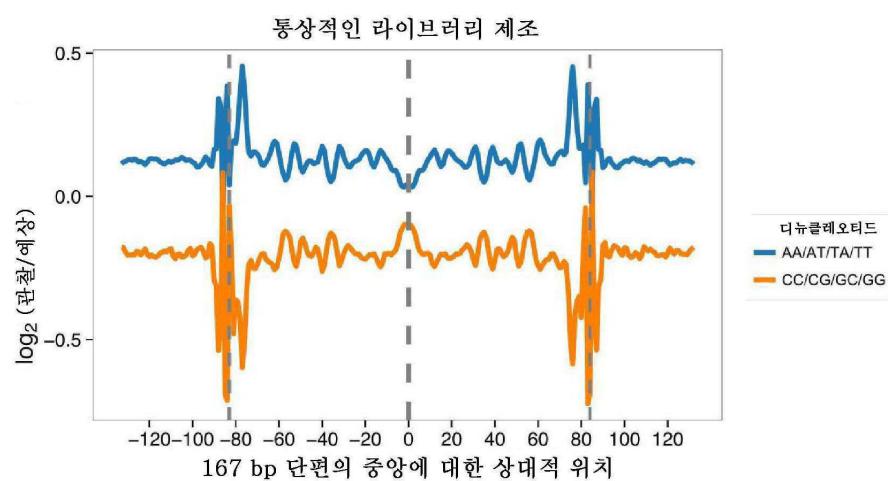
도면17



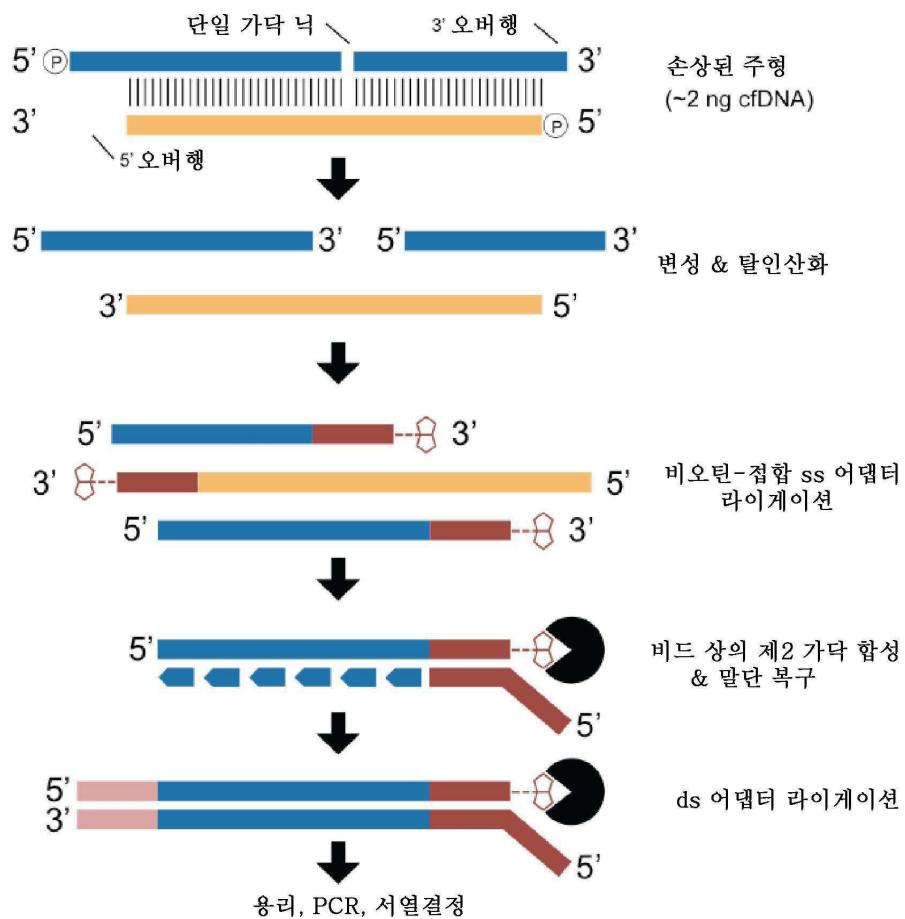
도면18



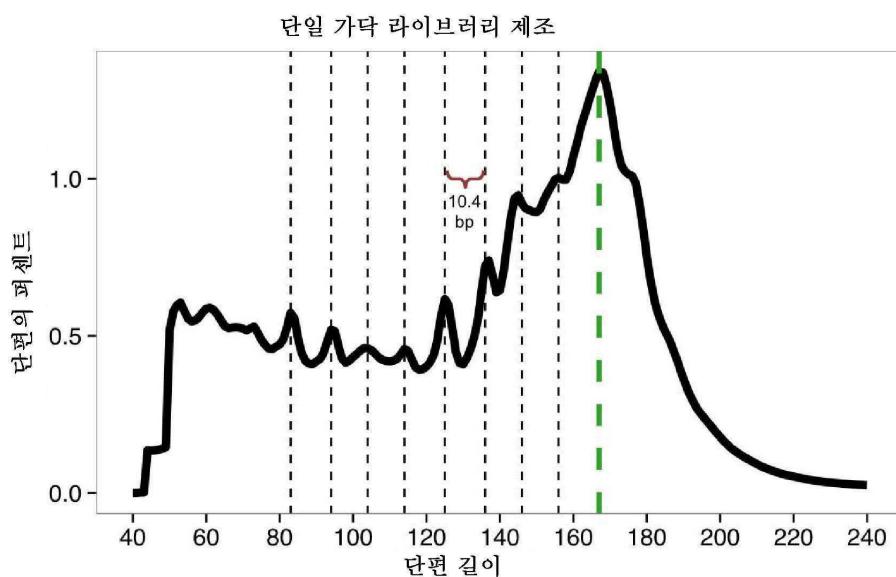
도면19



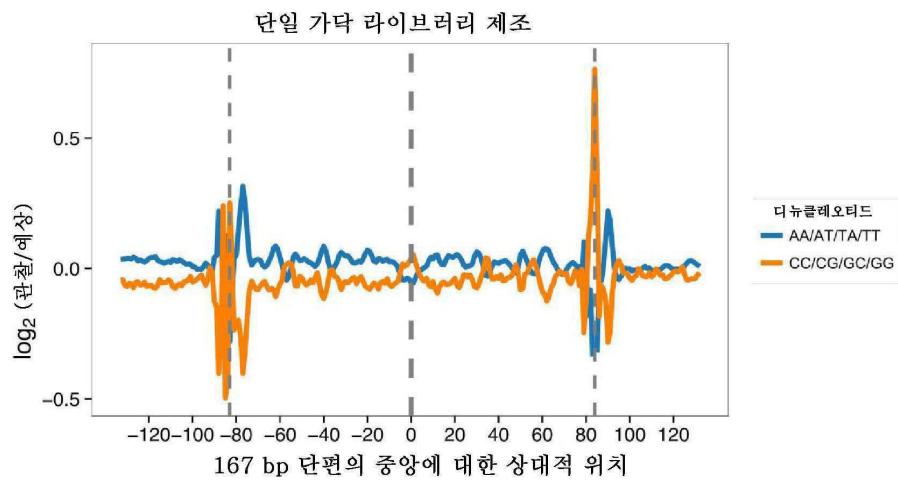
도면20



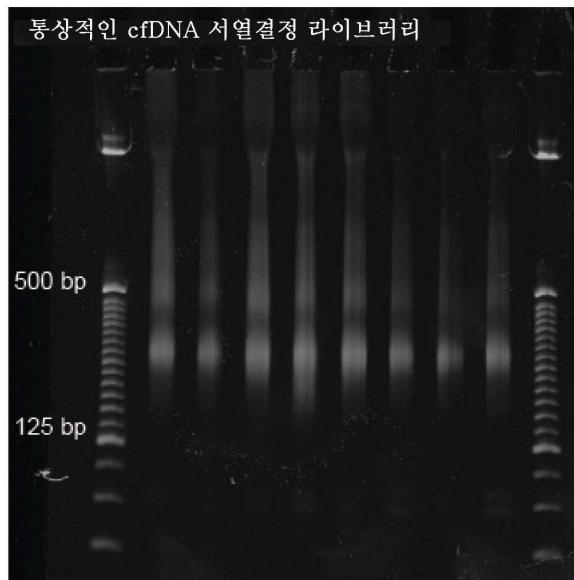
도면21



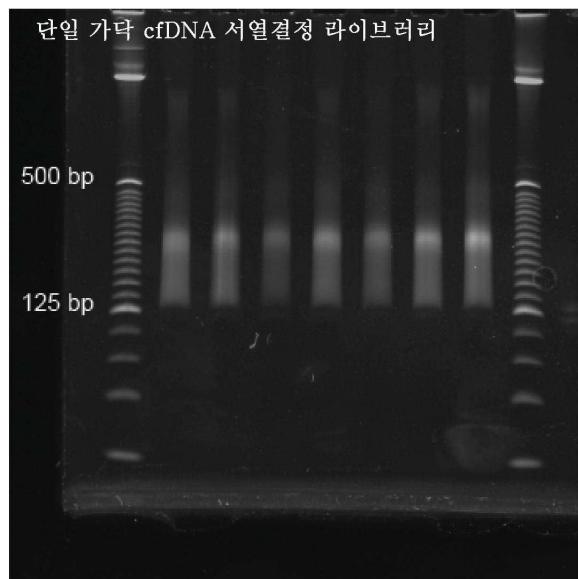
도면22



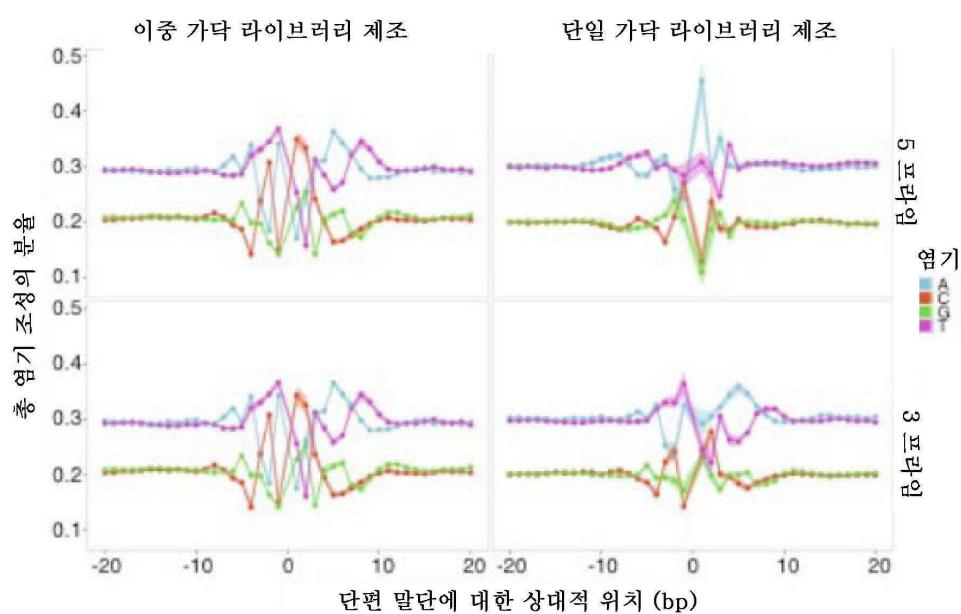
도면23a



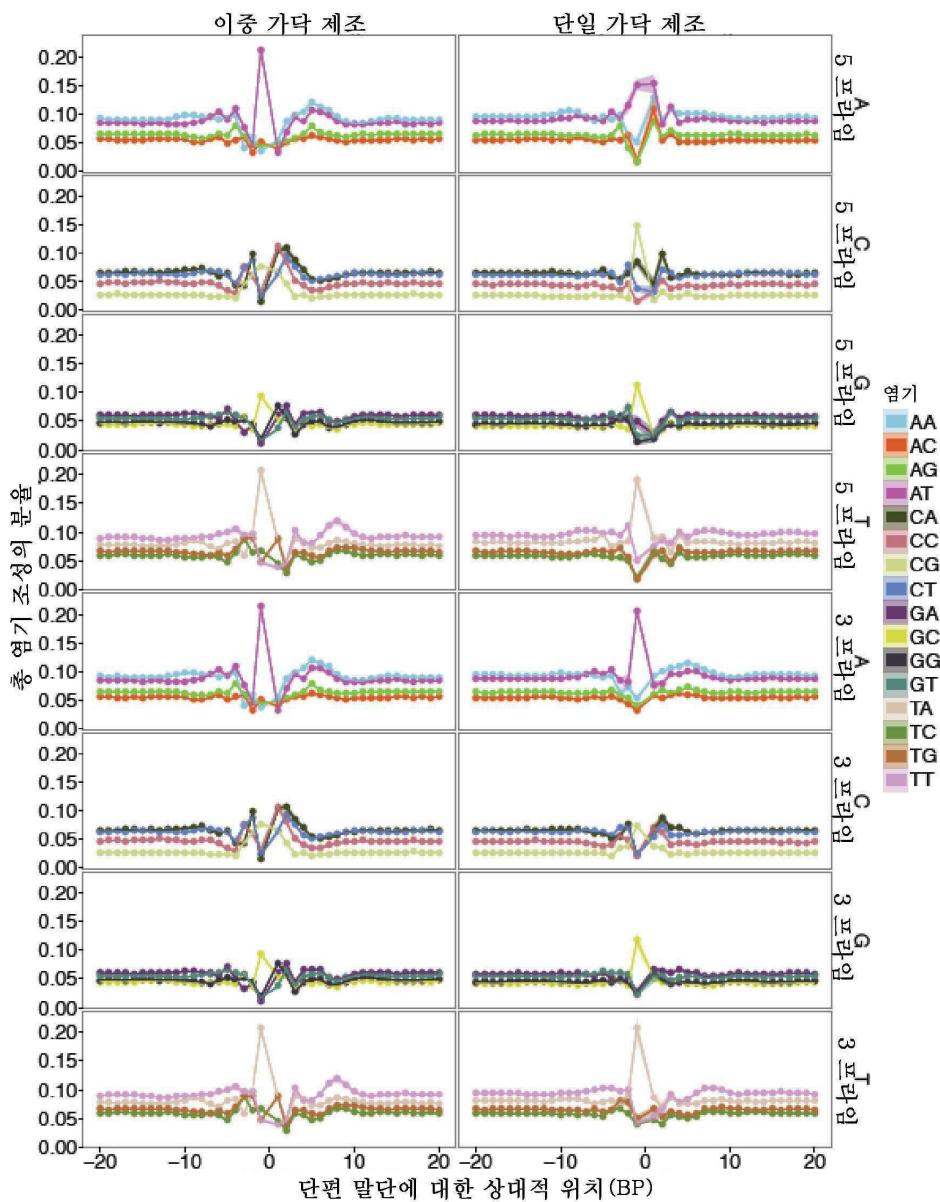
도면23b



도면24a

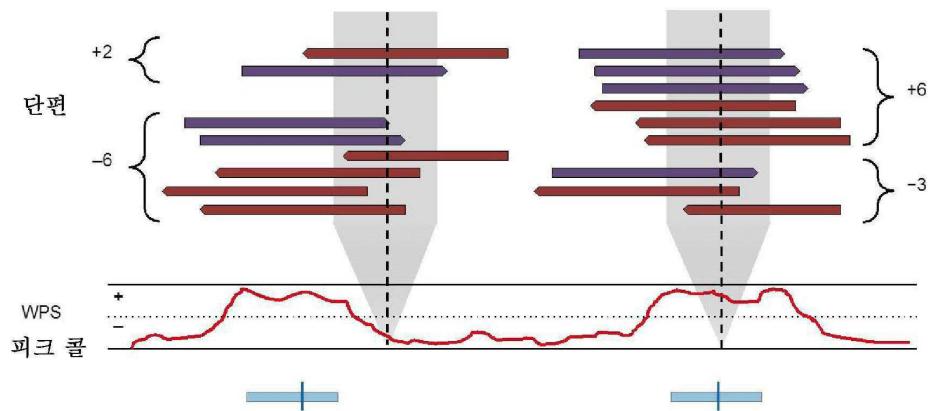


도면24b

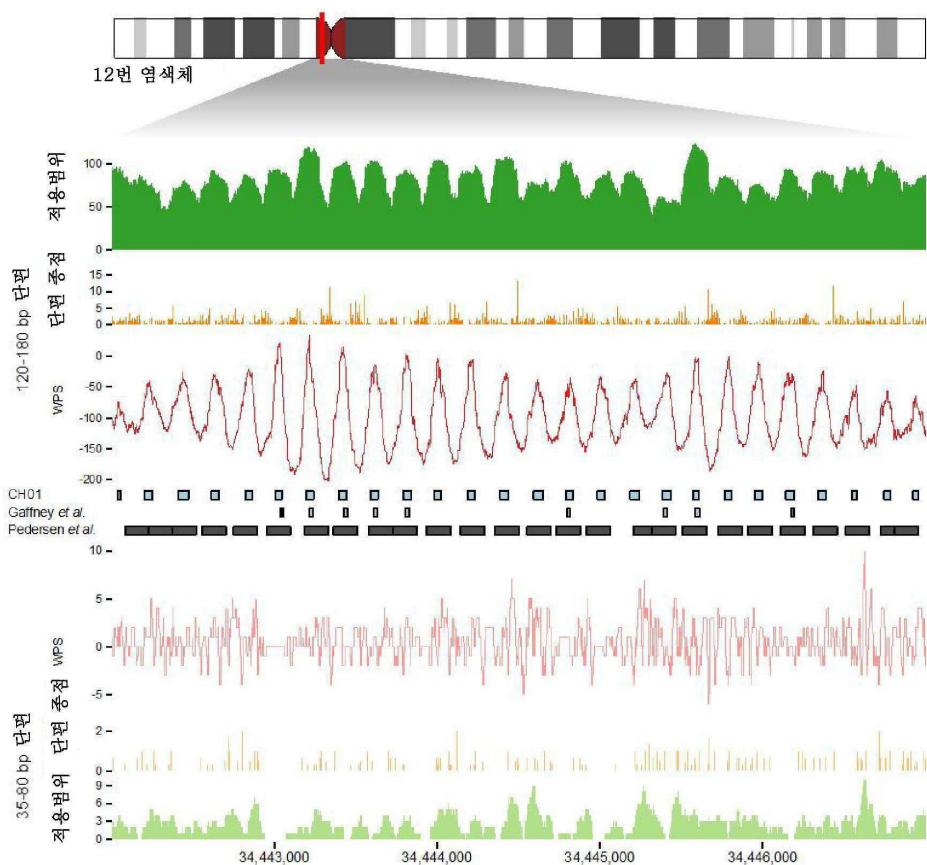


도면25

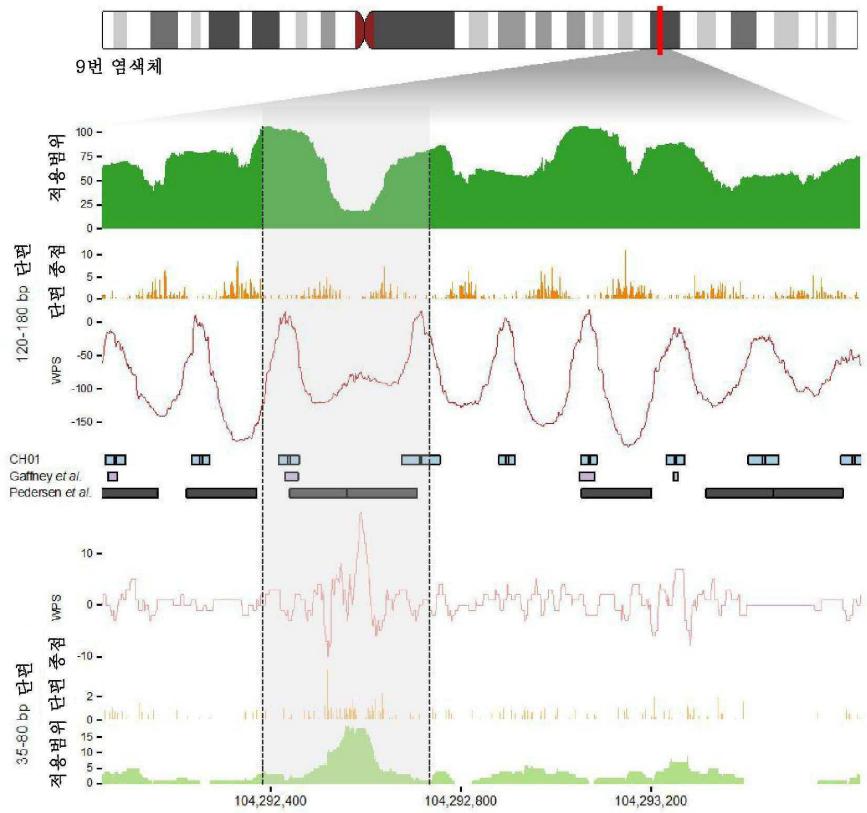
뉴클레오솜 위치



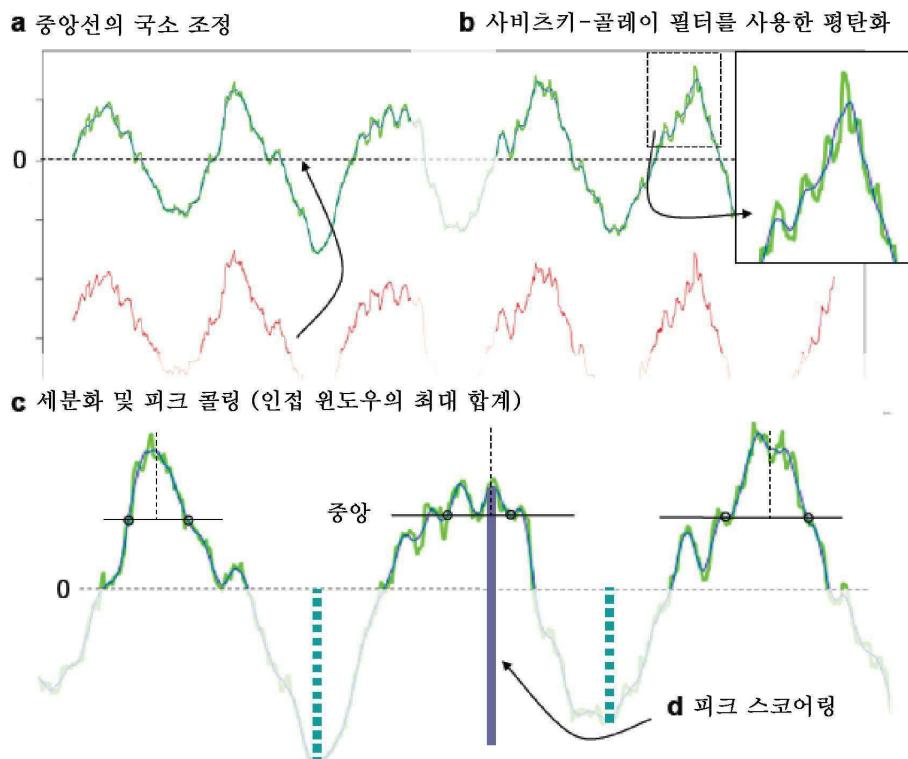
도면26



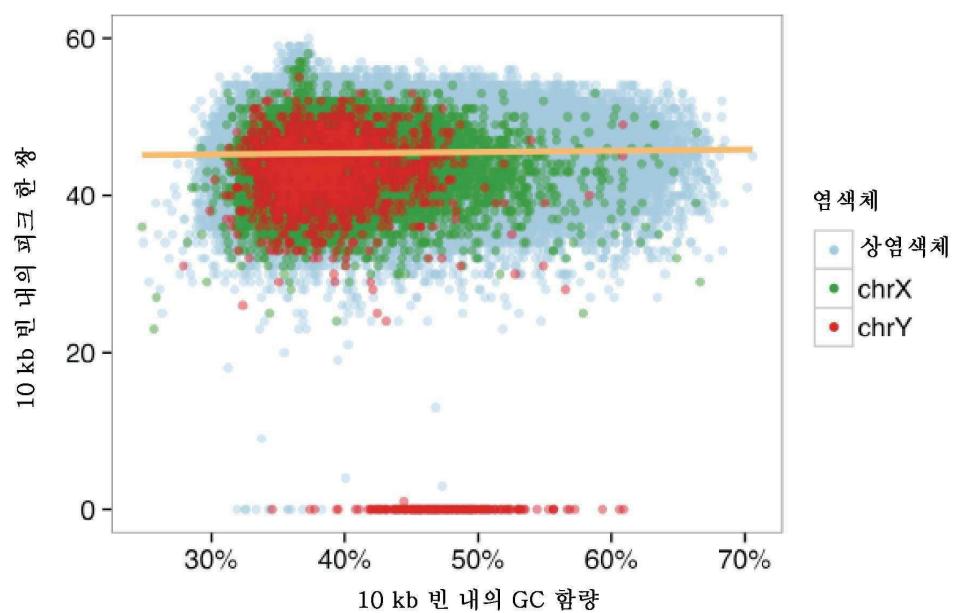
도면27



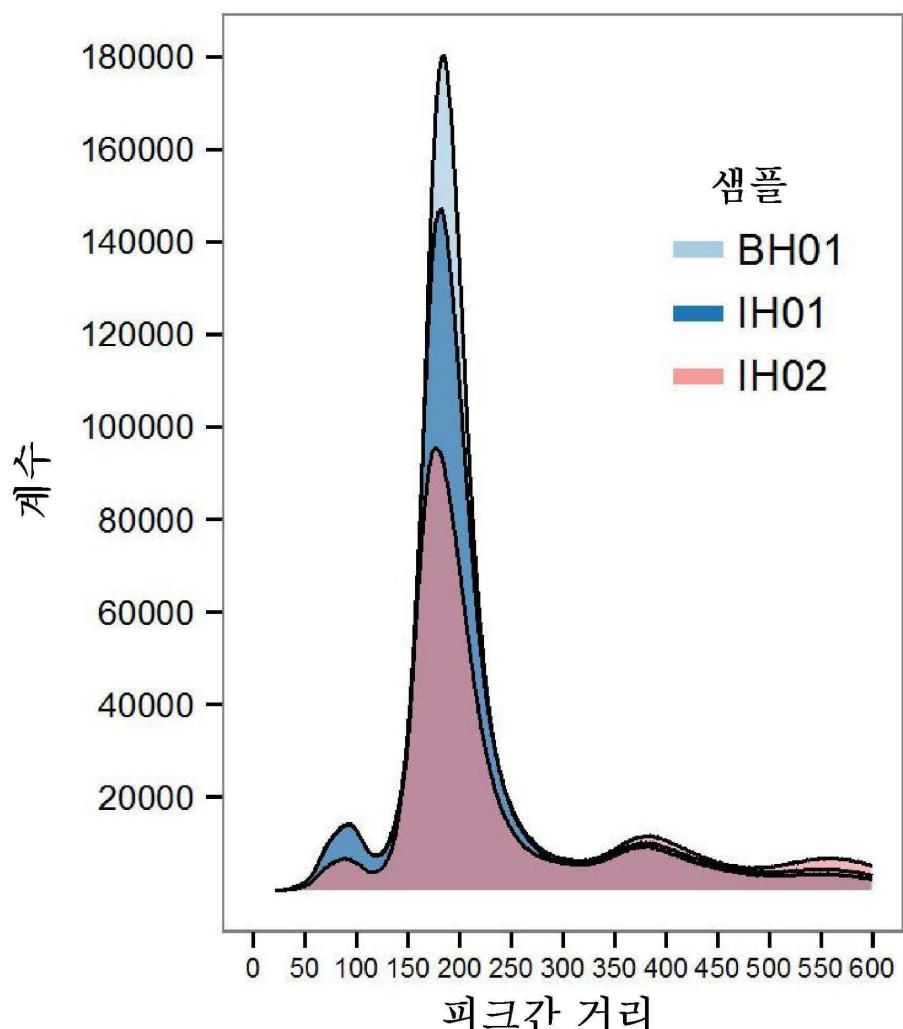
도면28



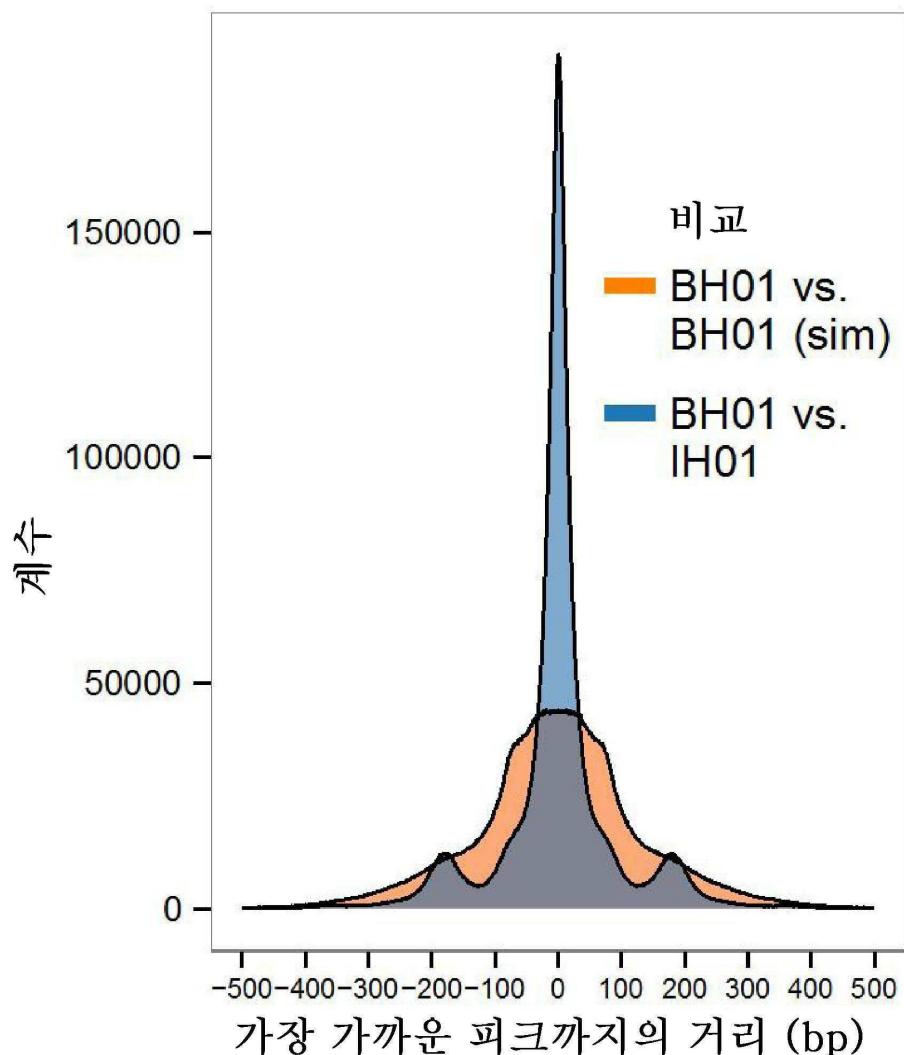
도면29



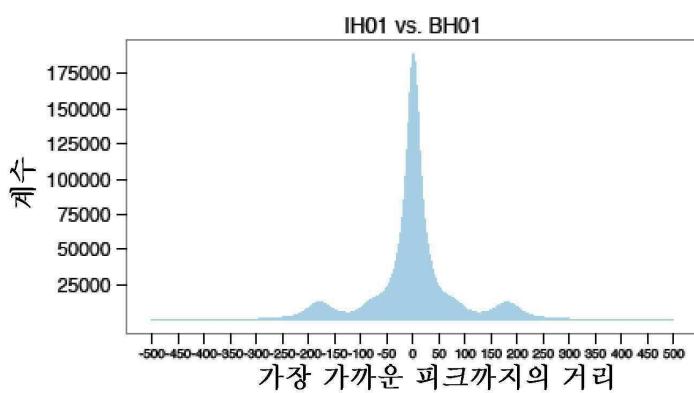
도면30



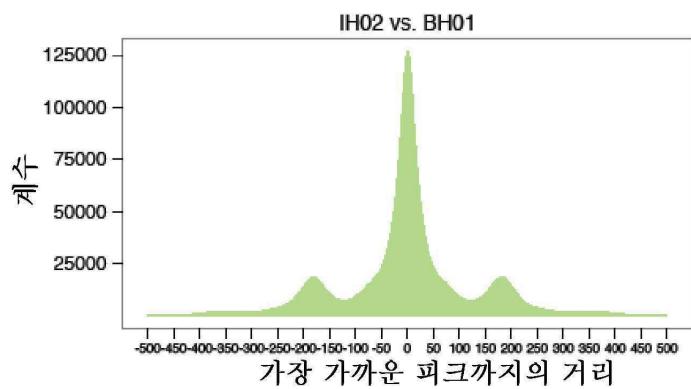
도면31



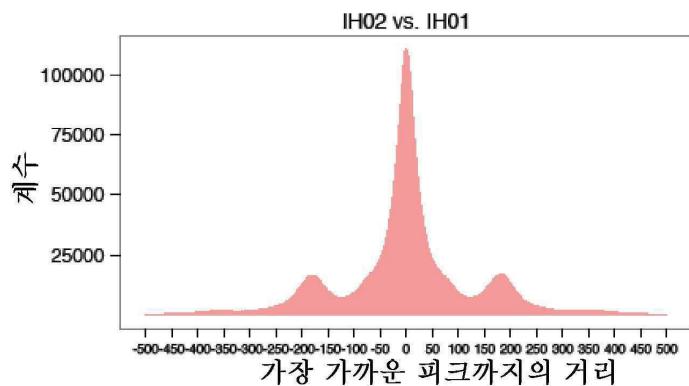
도면32a



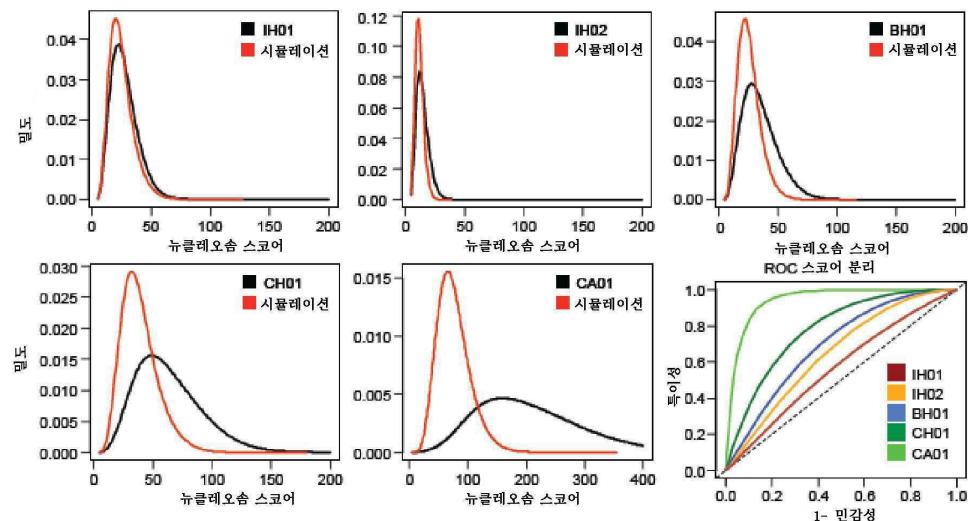
도면32b



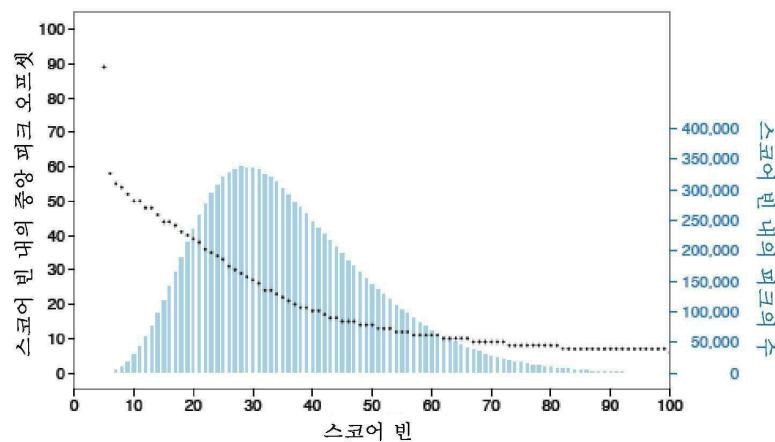
도면32c



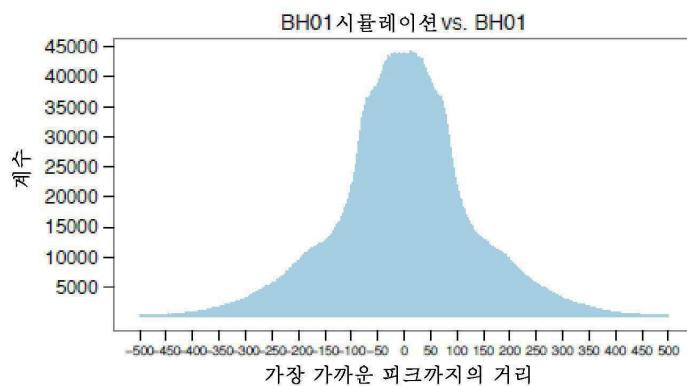
도면33a



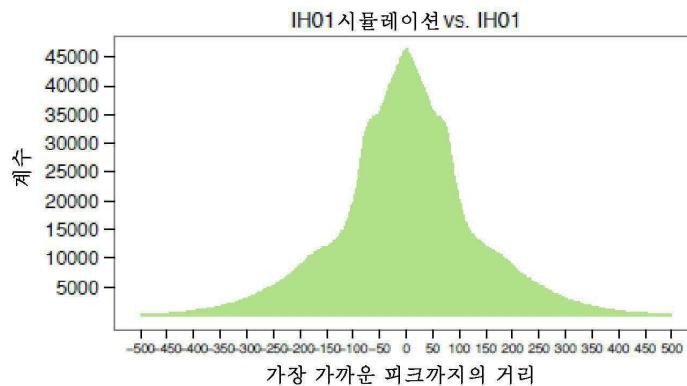
도면33b



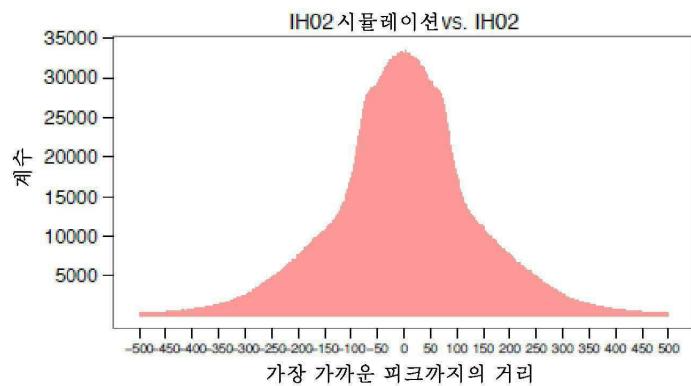
도면34a



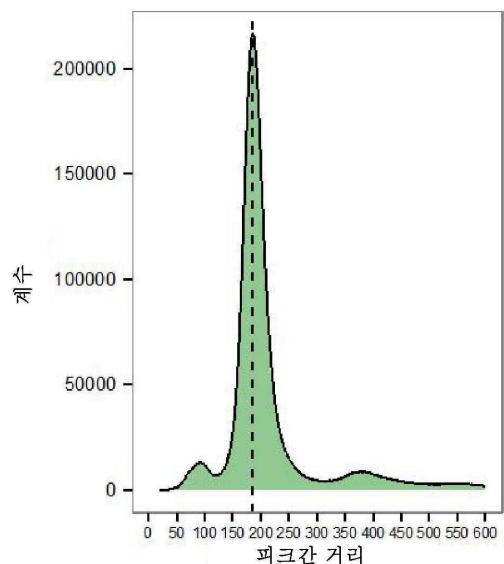
도면34b



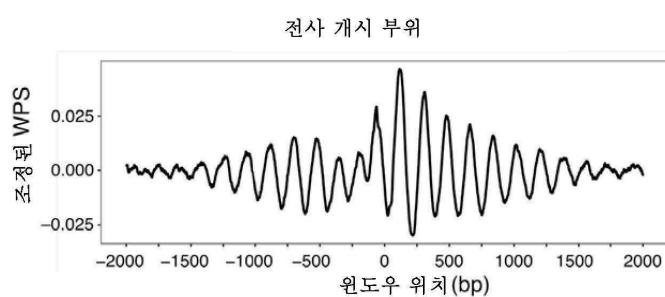
도면34c



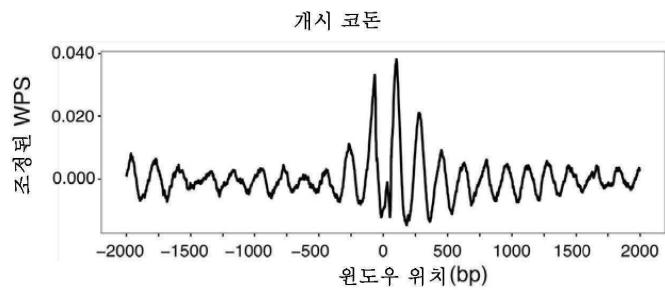
도면35



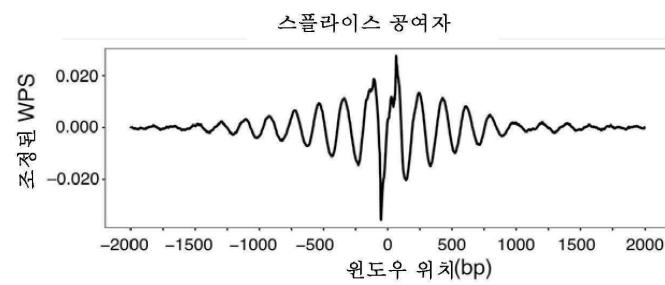
도면36



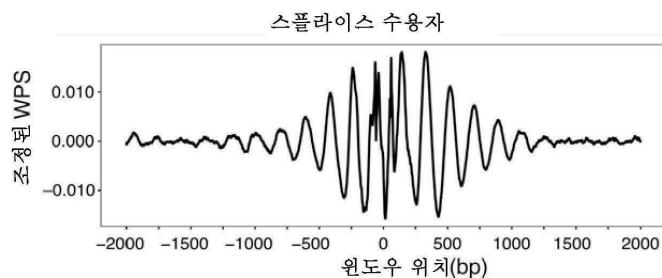
도면37



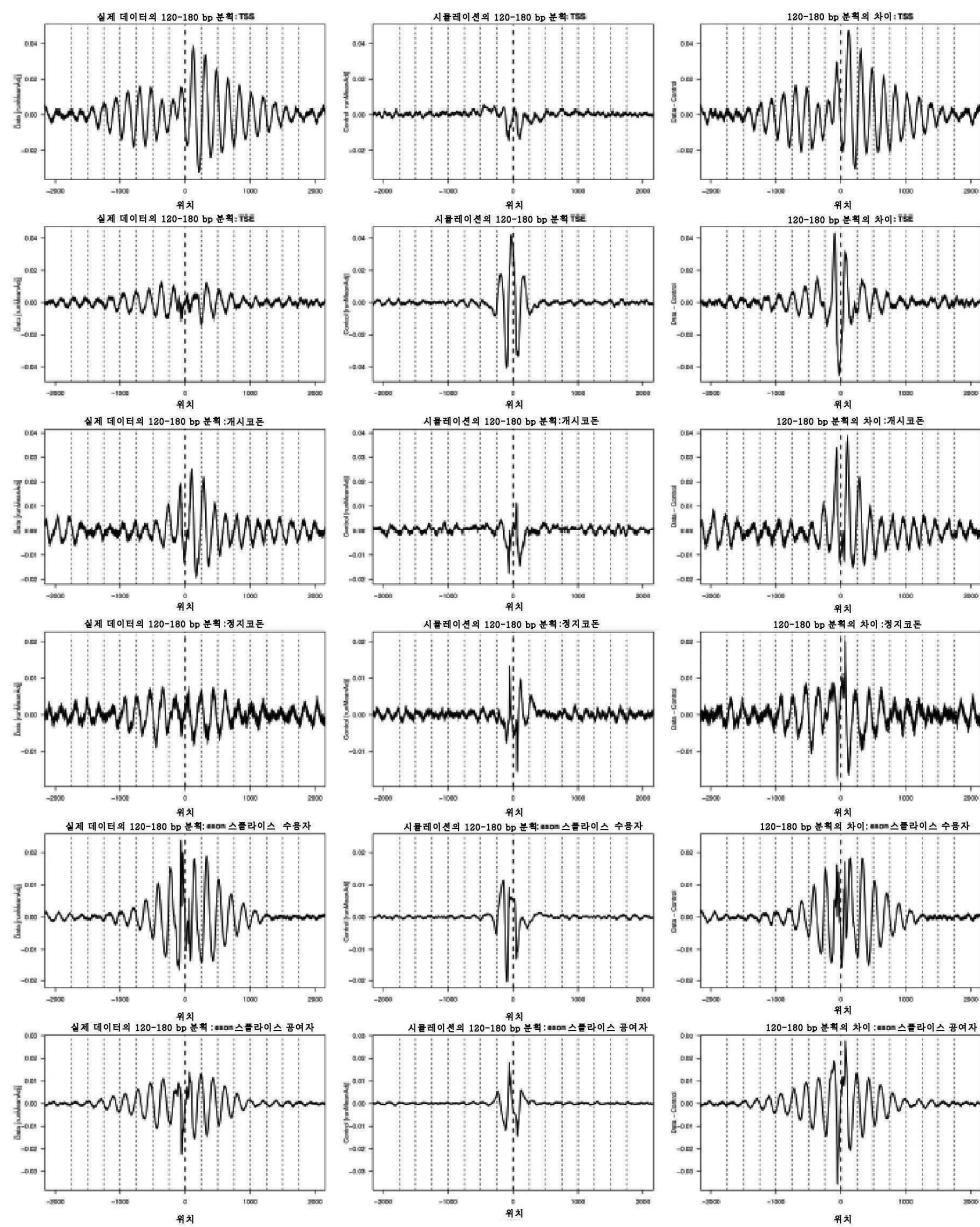
도면38



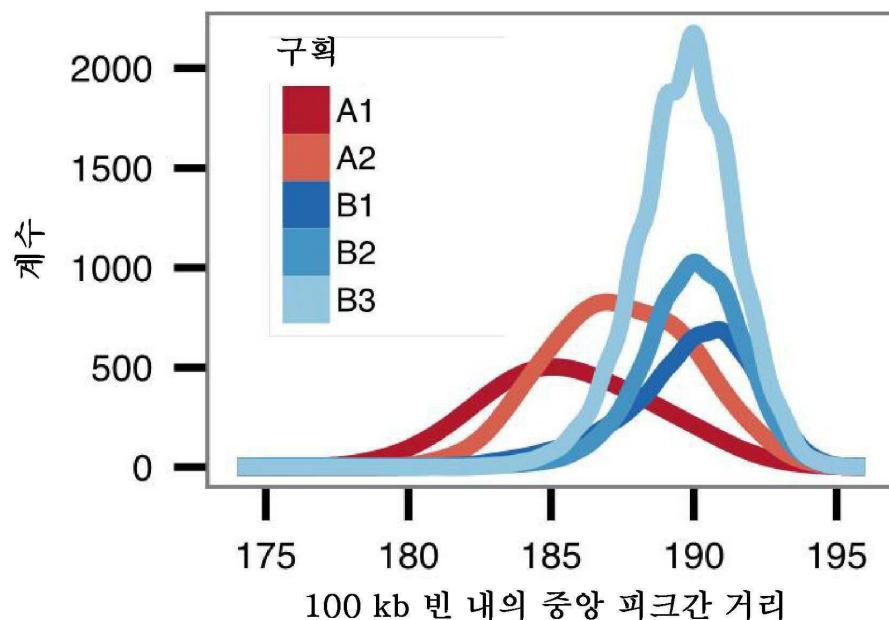
도면39



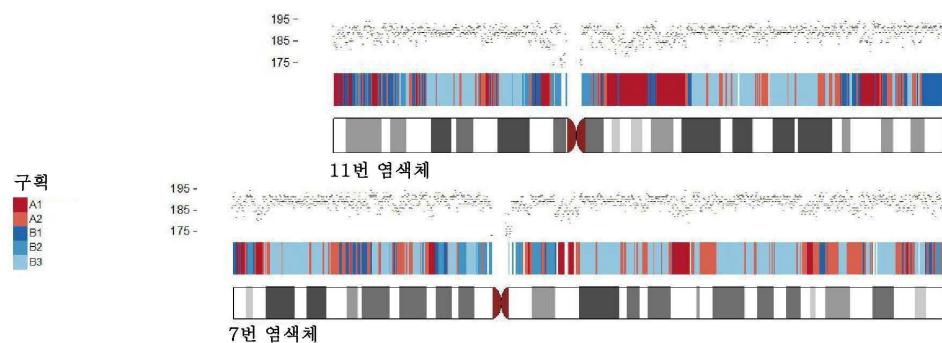
도면40



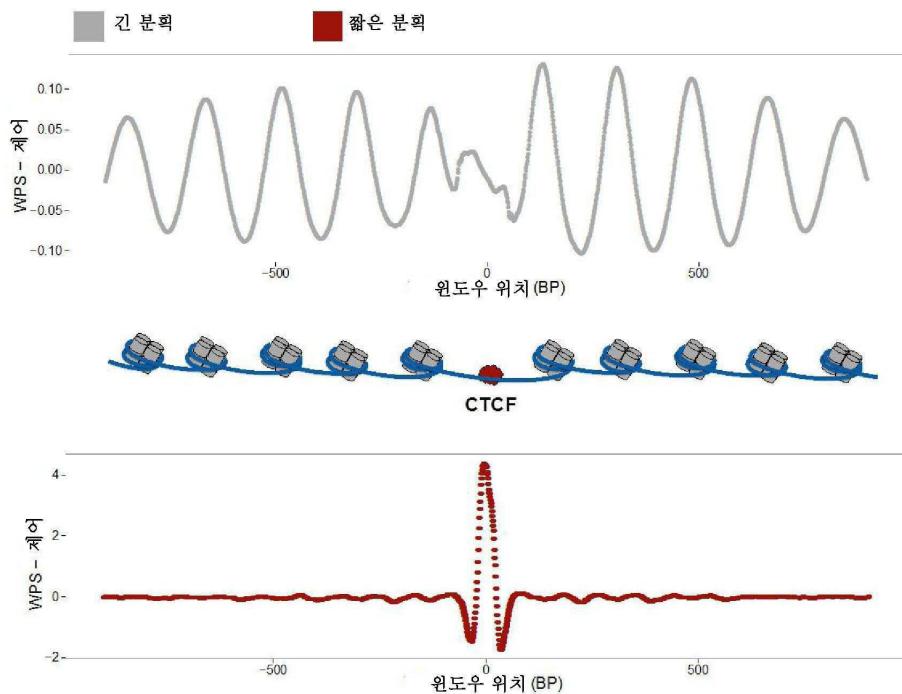
도면41



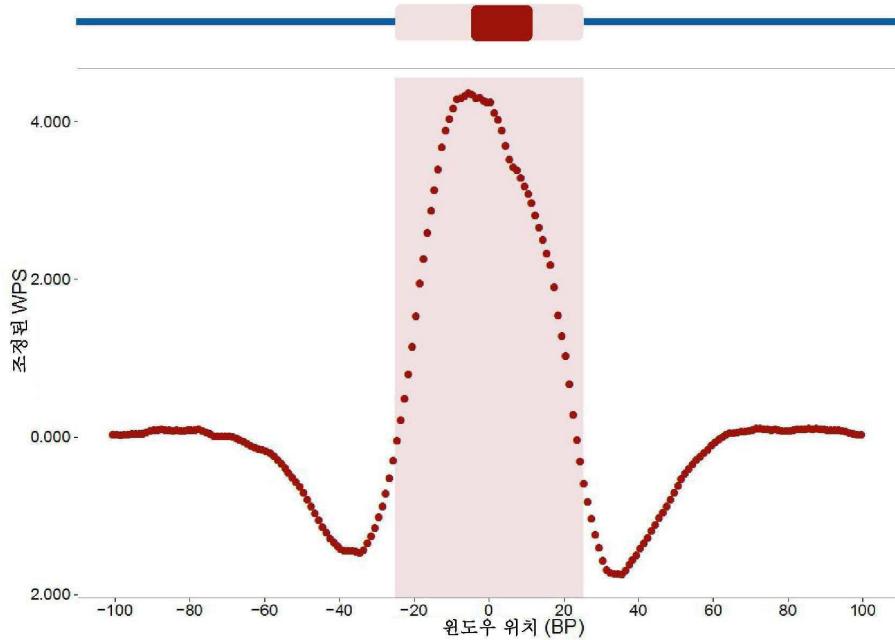
도면42



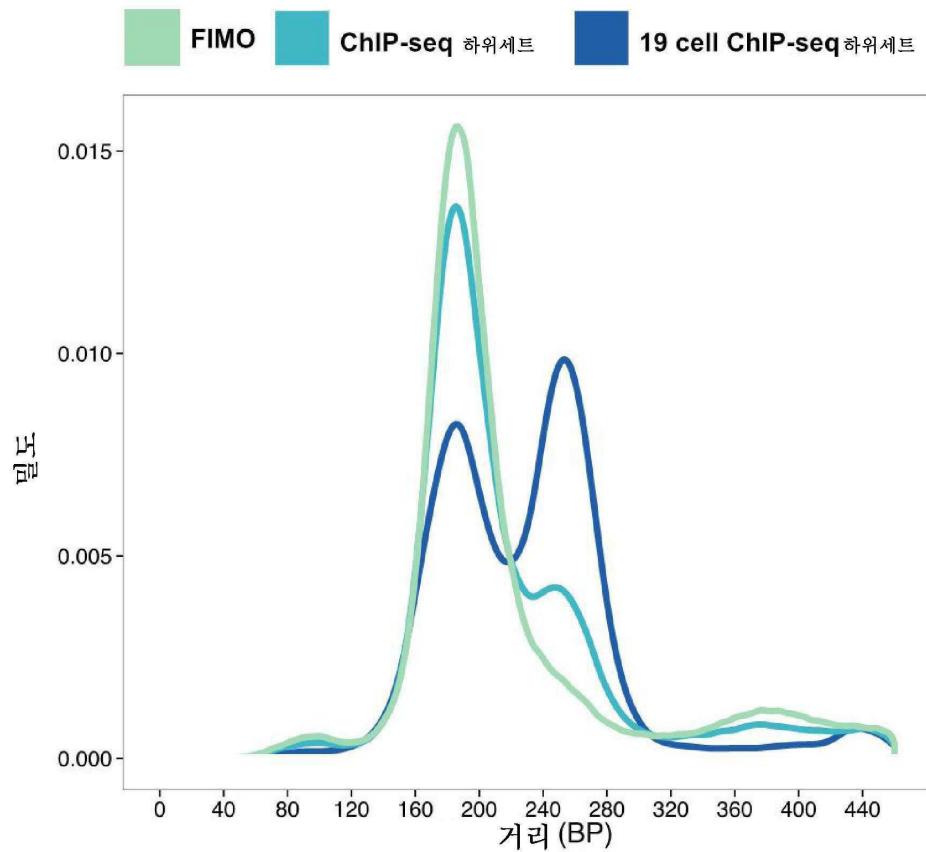
도면43



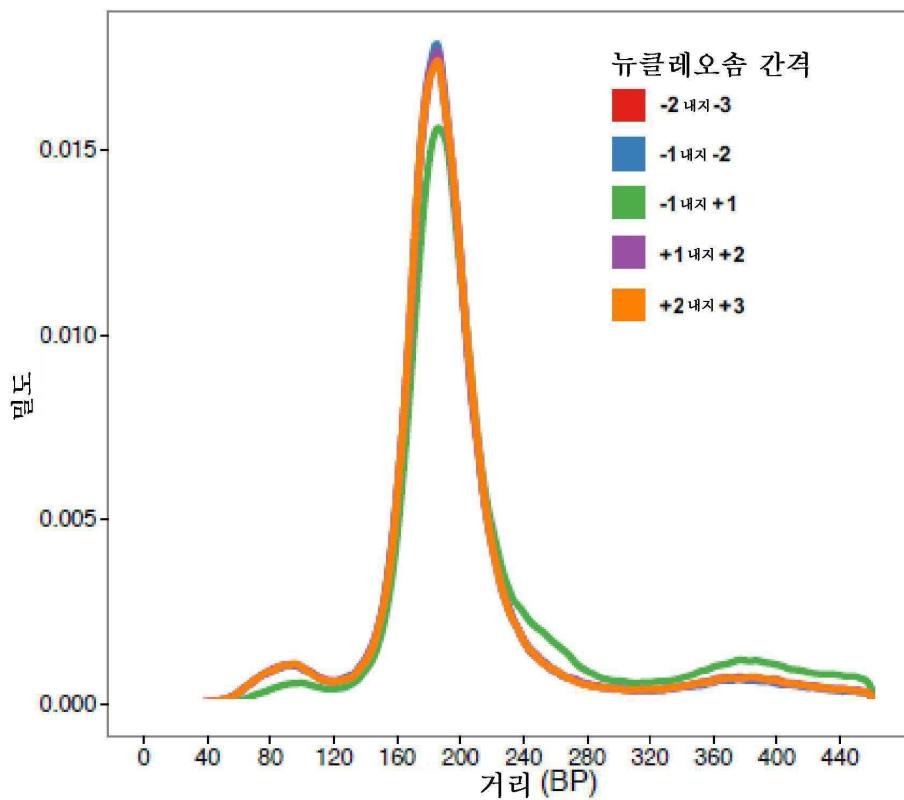
도면44



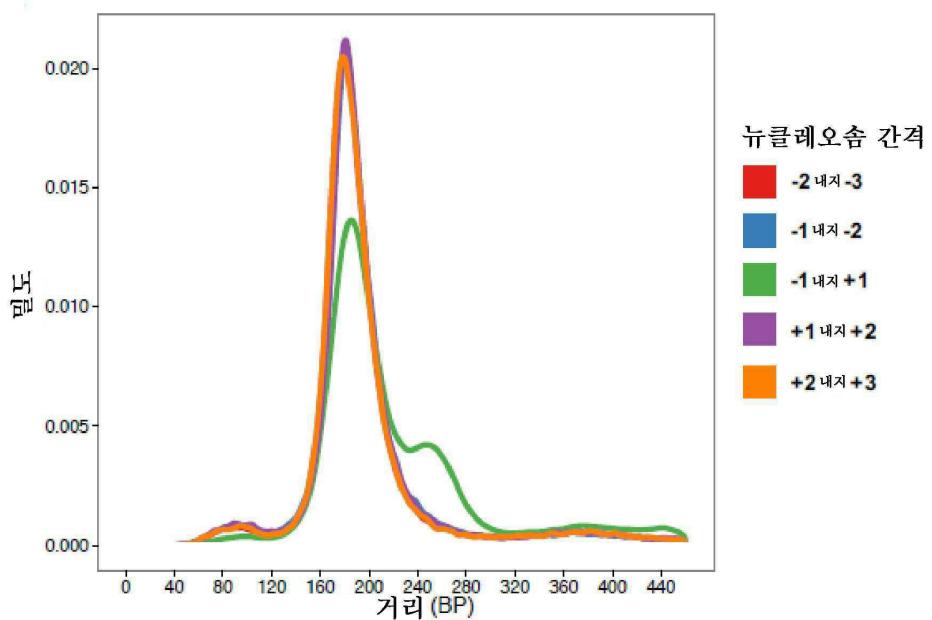
도면45



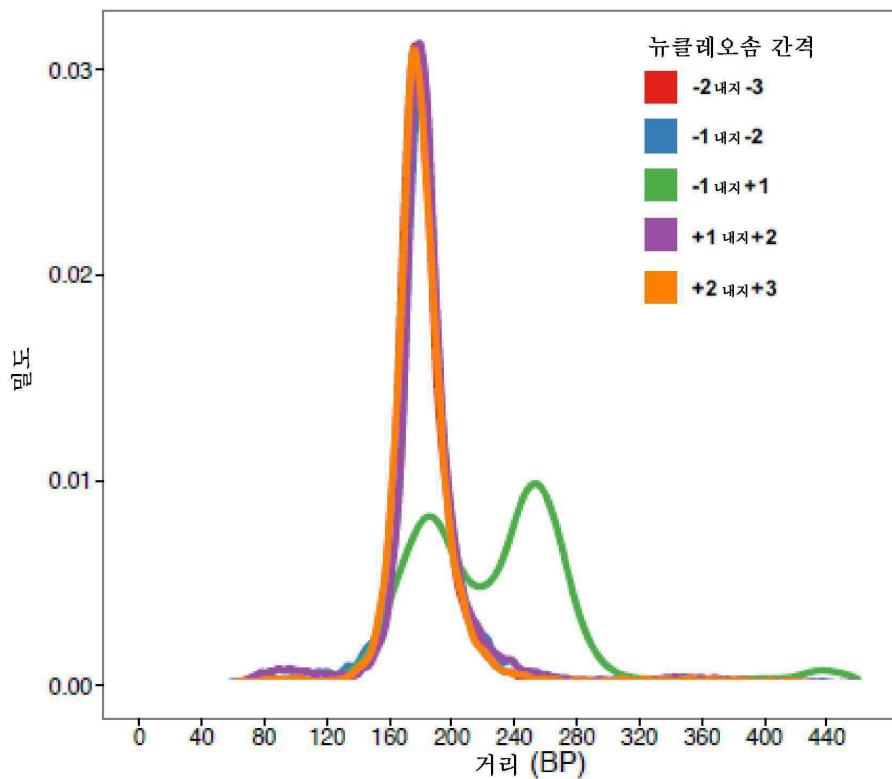
도면46



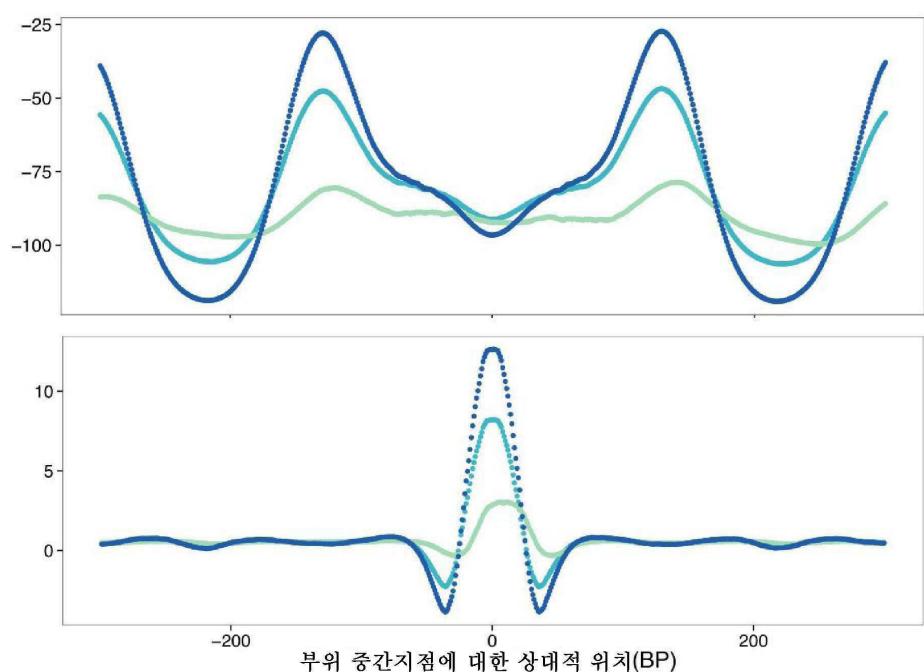
도면47



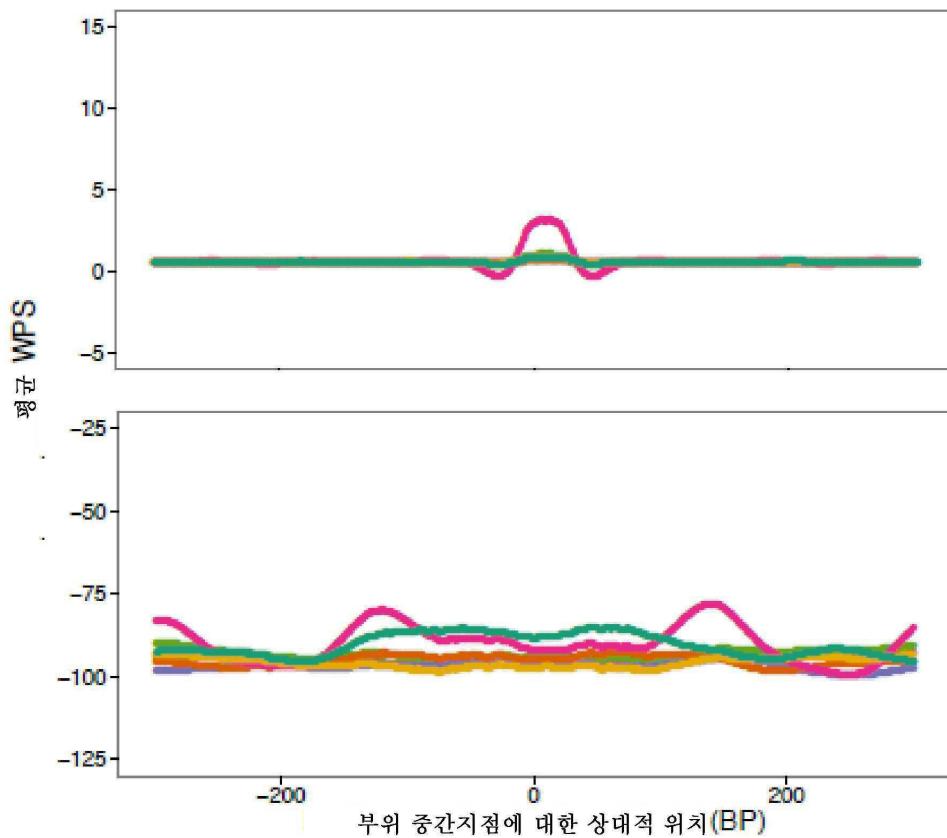
도면48



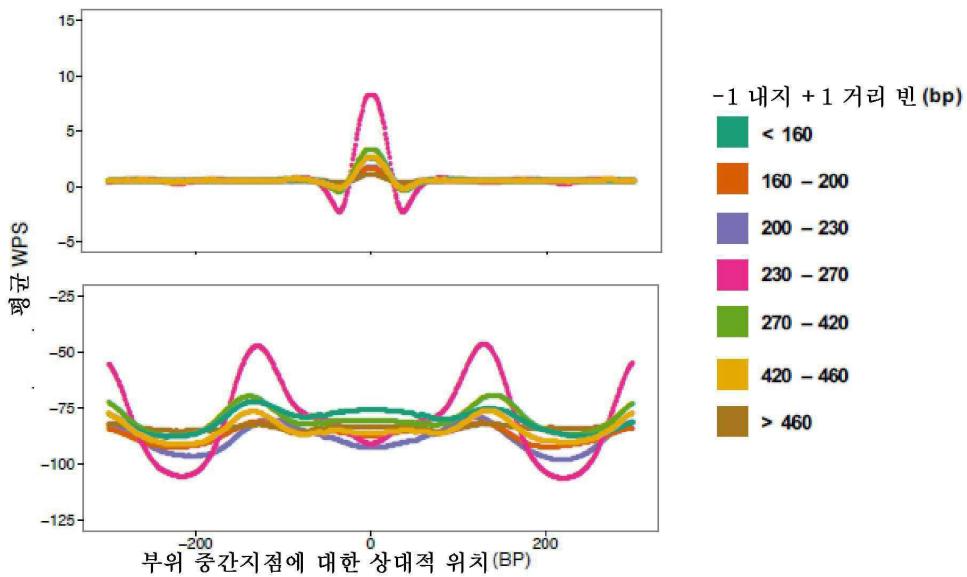
도면49



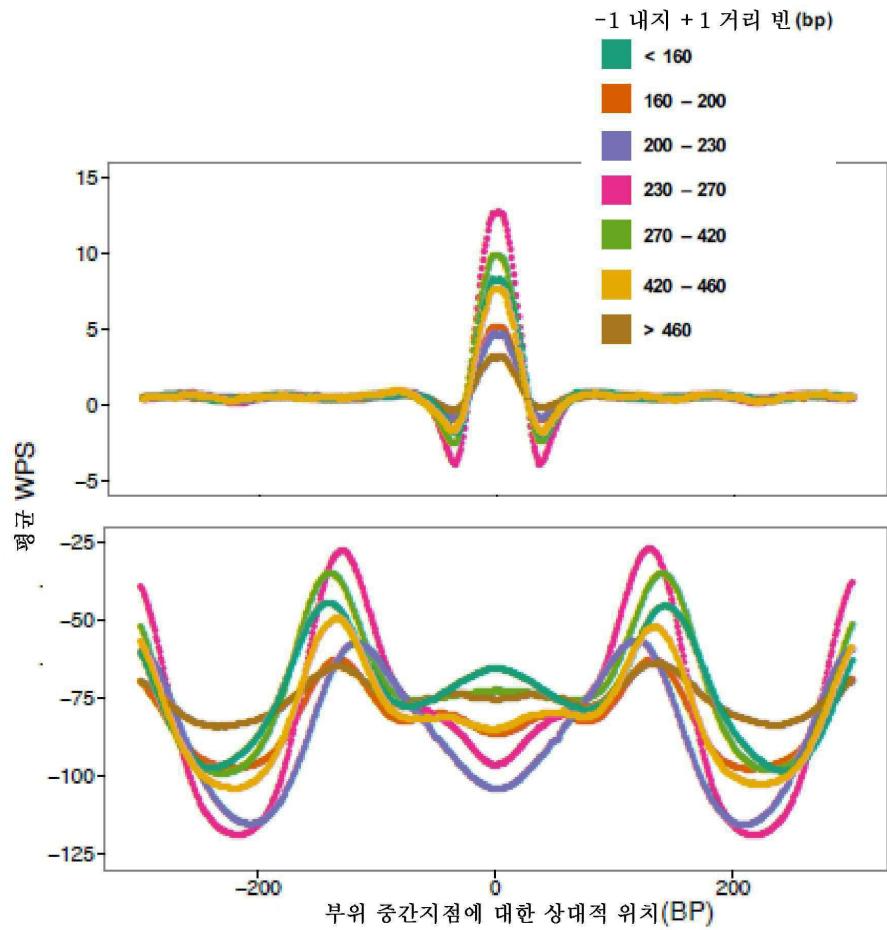
도면50



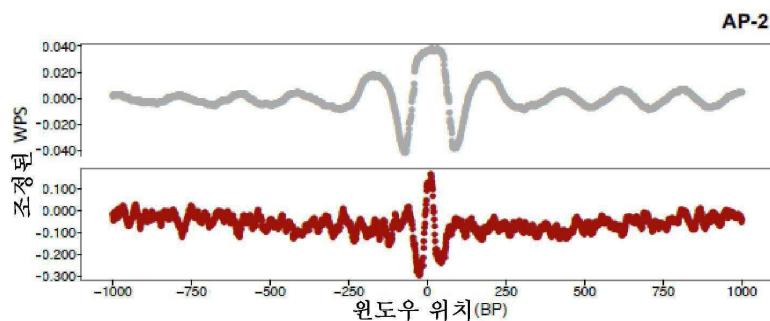
도면51



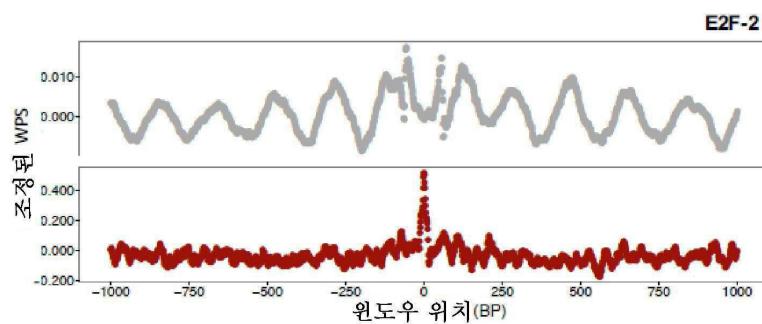
도면52



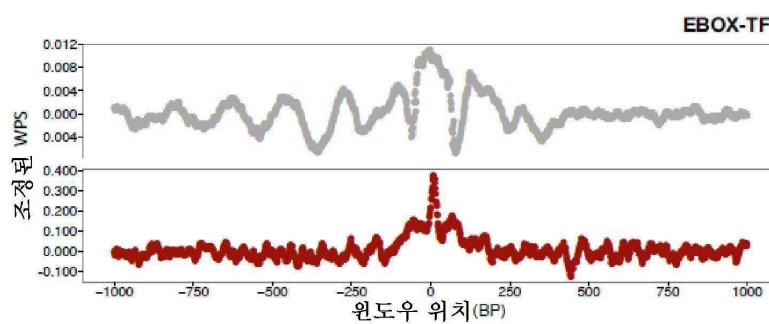
도면53a



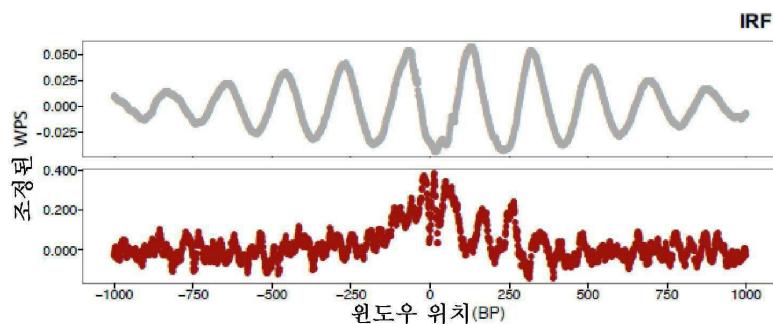
도면53b



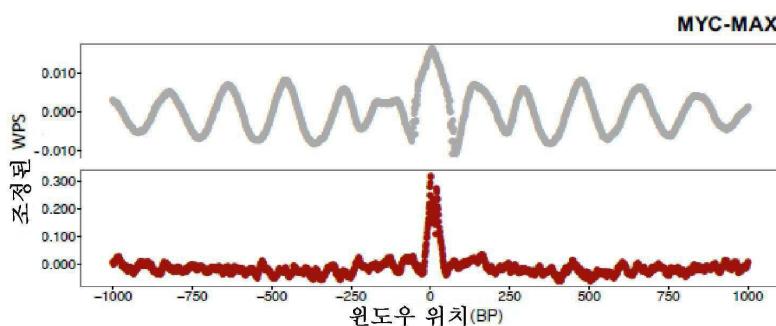
도면53c



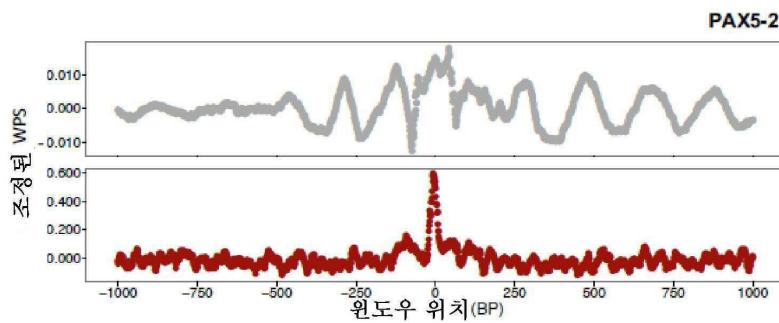
도면53d



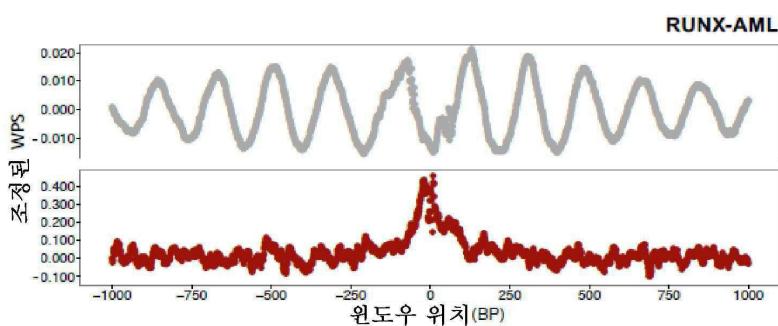
도면53e



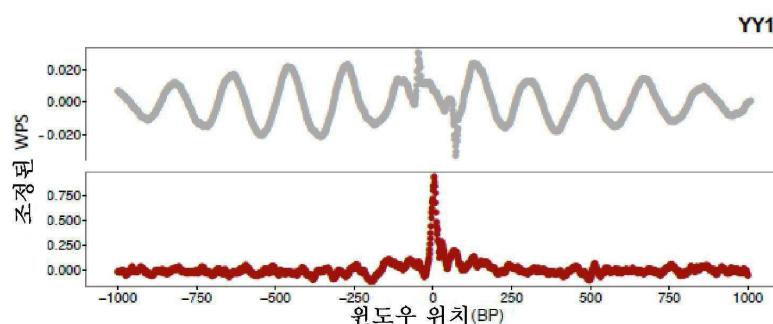
도면53f



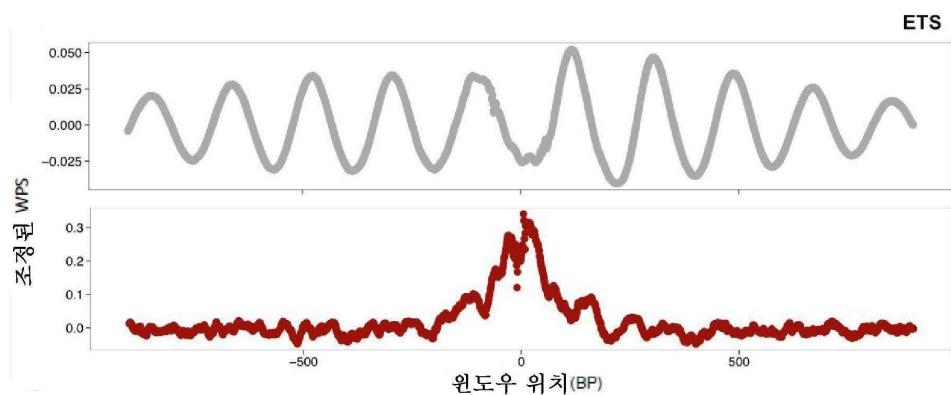
도면53g



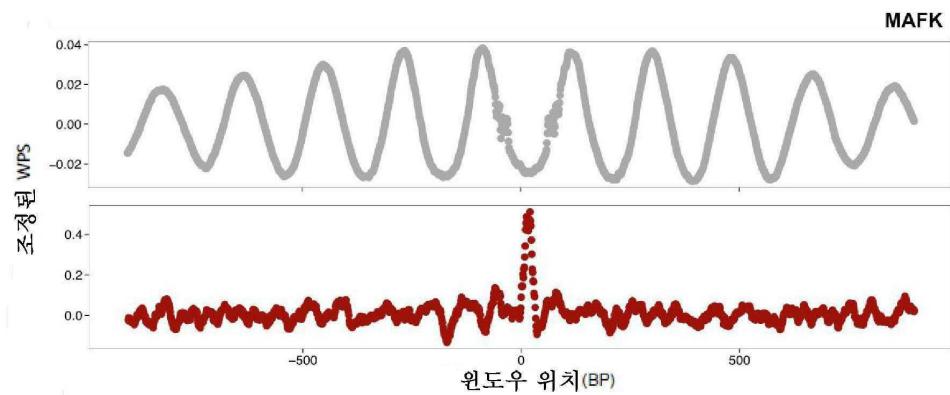
도면53h



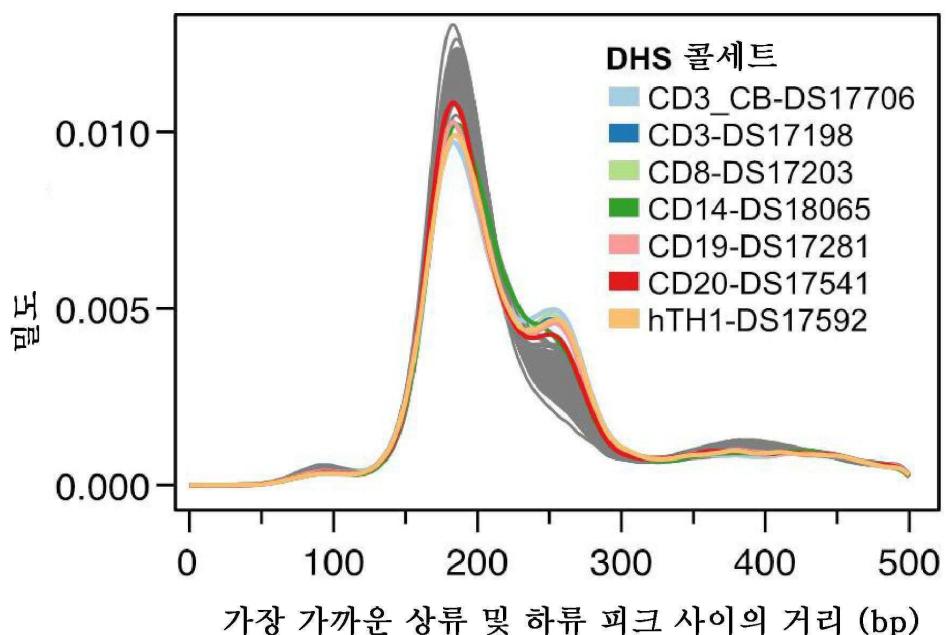
도면54



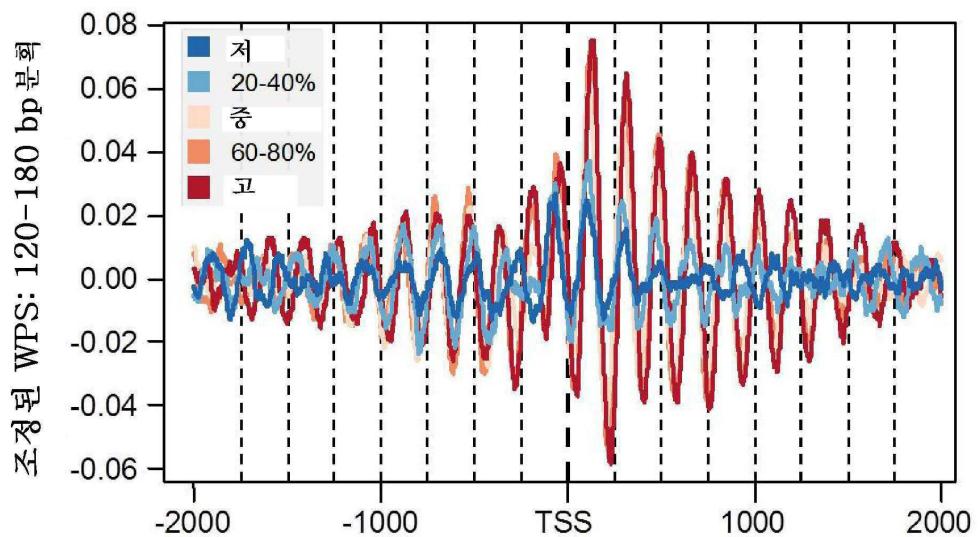
도면55



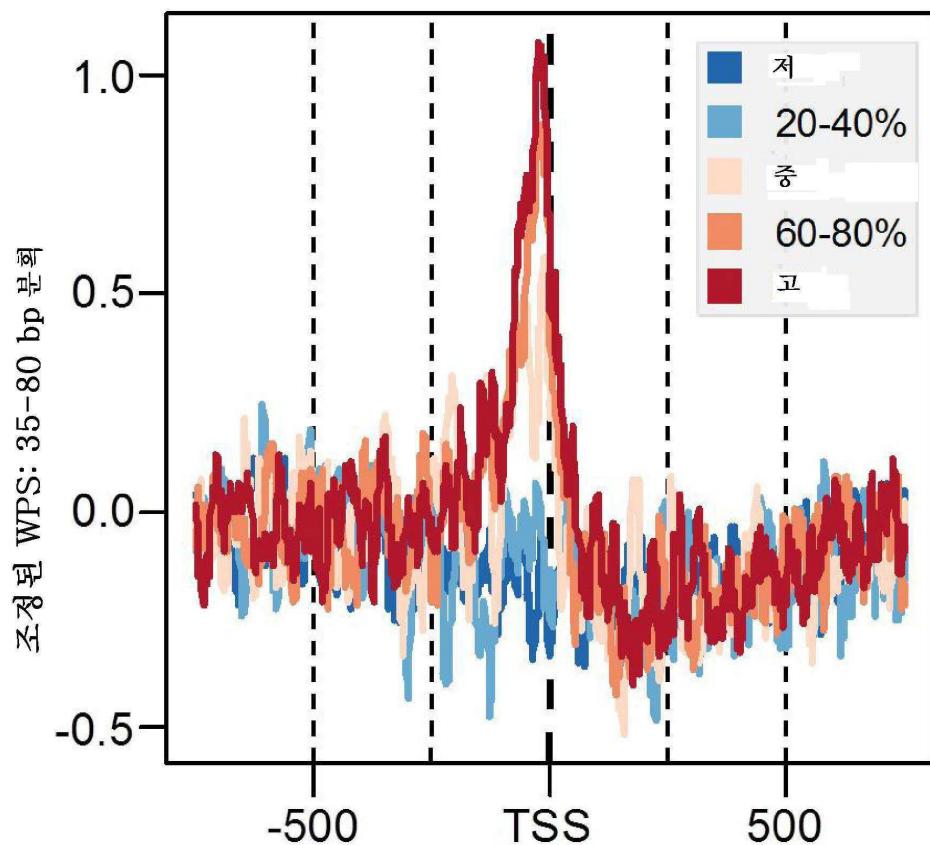
도면56



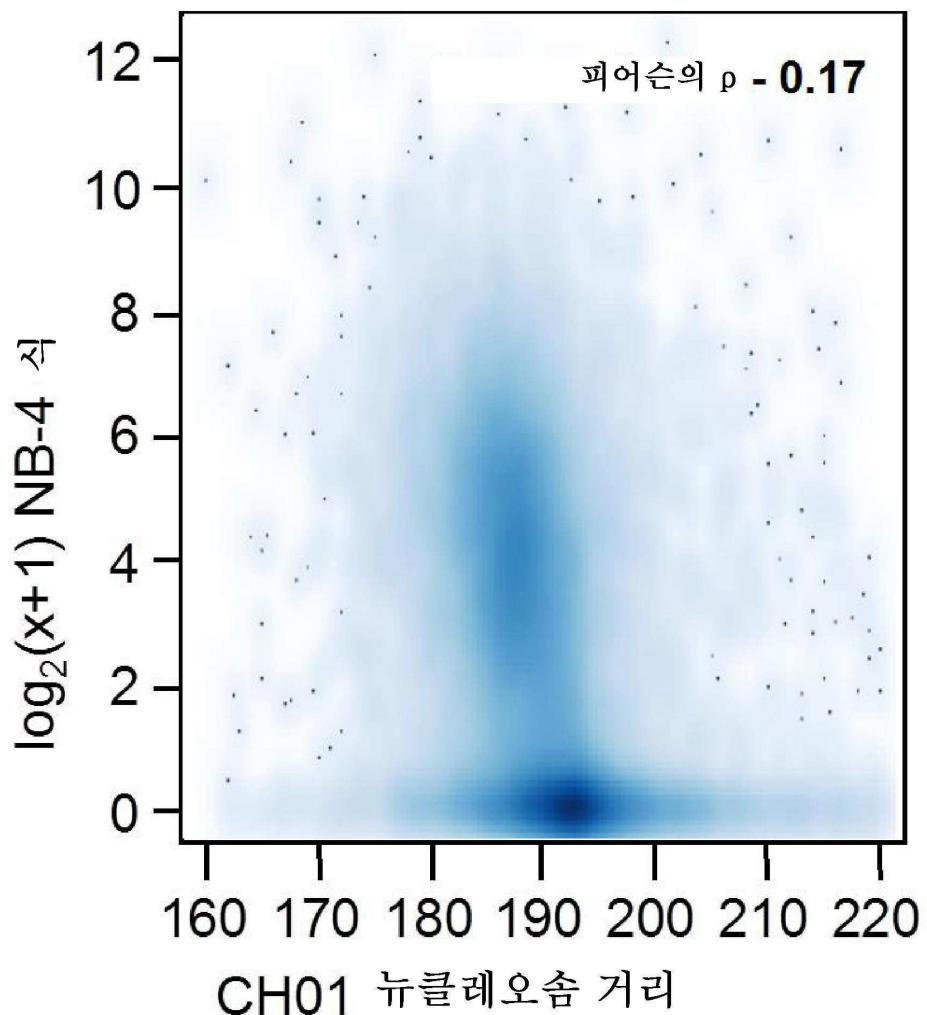
도면57



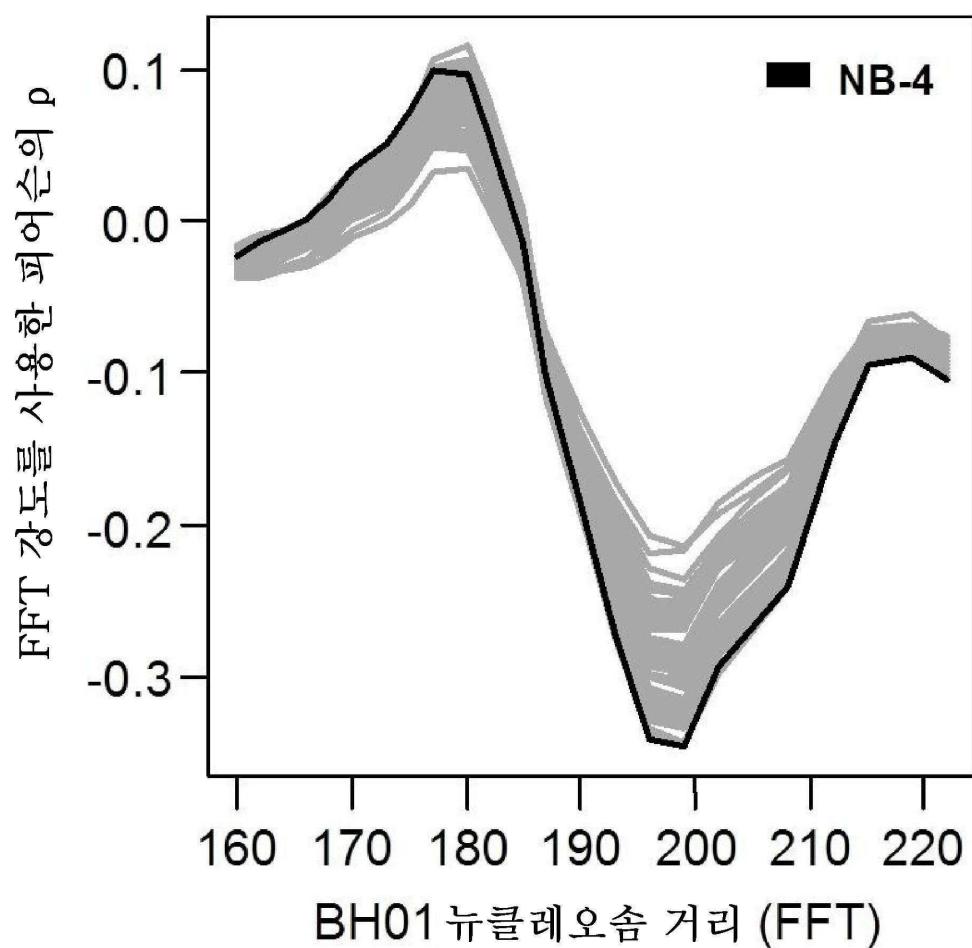
도면58



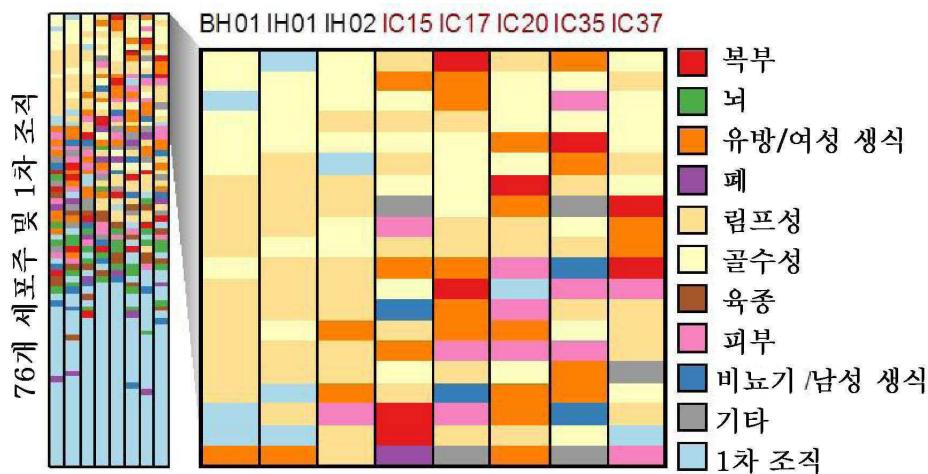
도면59



도면60

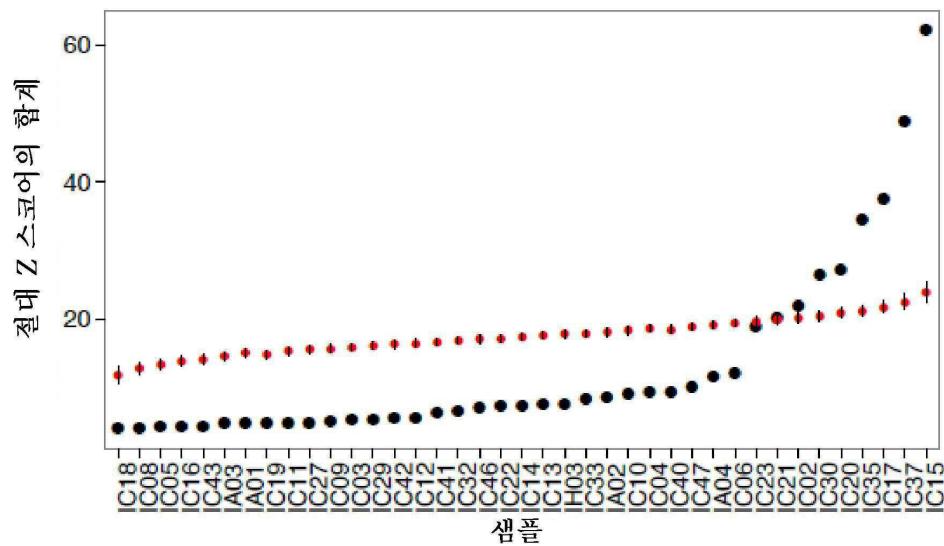


도면61

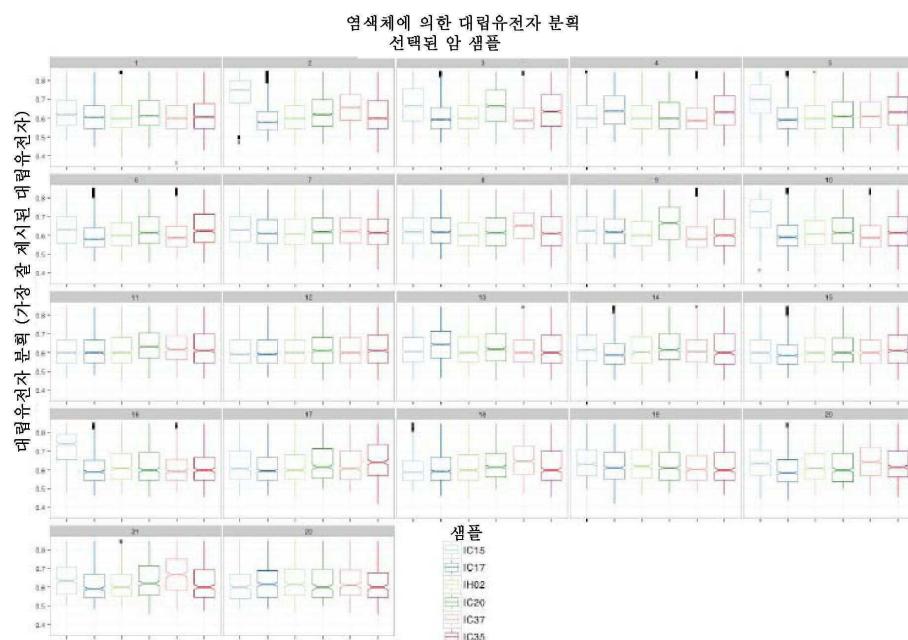


샘플	암	순위 증가한 상위 3개
IC15 ♂ (SCLC)	폐	SCLC-21H (31), SH-SY5Y (25), HEK 293 (17)
IC17 ♂ (HCC)	간	신장 (21), CAPAN-2 (19), BEWO (19)
IC20 ♂ (SCC)	폐	SK-BR-3 (21), Hep G2 (18), HaCaT (18)
IC35 ♀ (DC)	유방	BEWO (27), SiHa (27), CAPAN-2 (25)
IC37 ♀ (AC)	결장직장	Hep G2 (24), SK-BR-3 (22), EFO-21 (20)
대조군	-	십이지장(10), U-251 MG (9), 소장 (9)

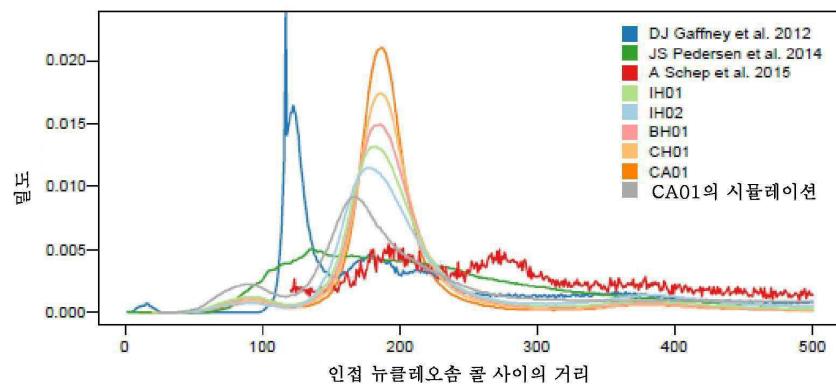
도면62a



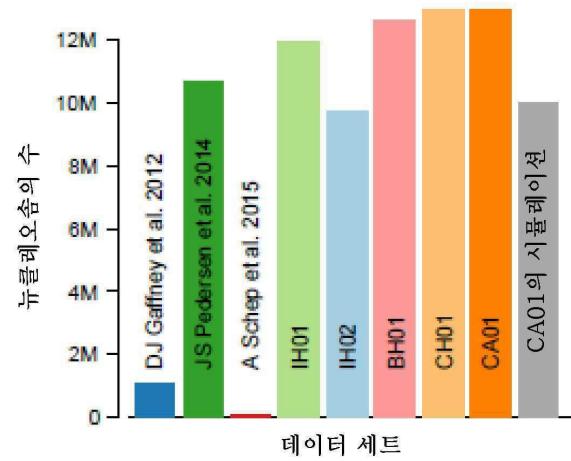
도면62b



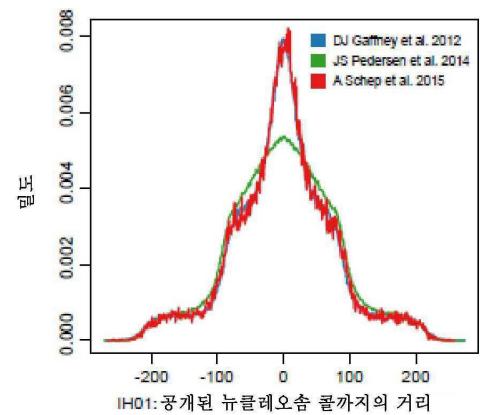
도면63a



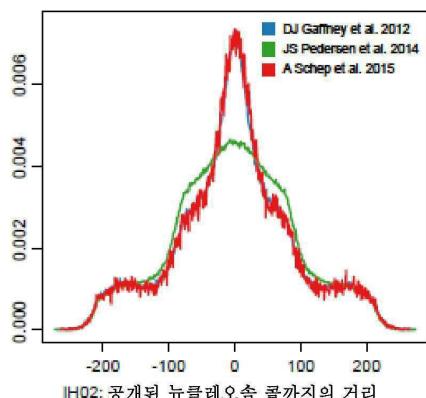
도면63b



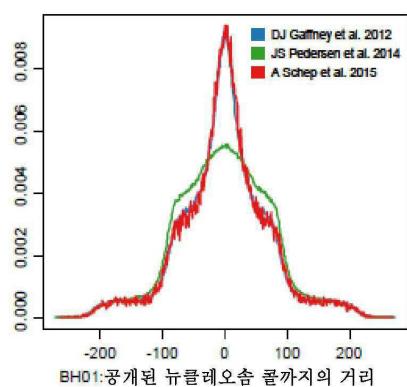
도면63c



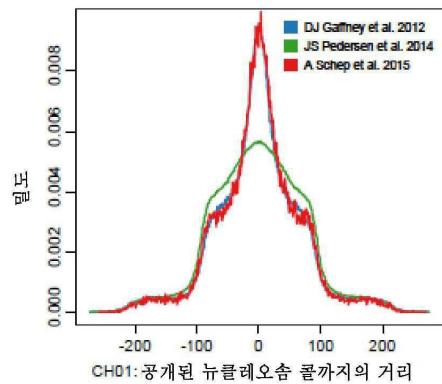
도면63d



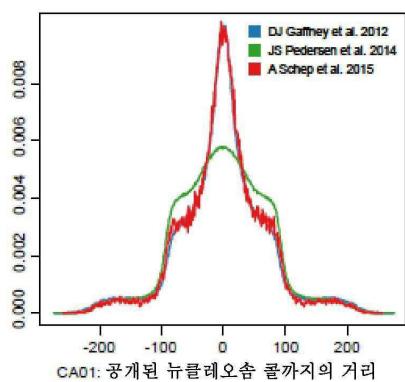
도면63e



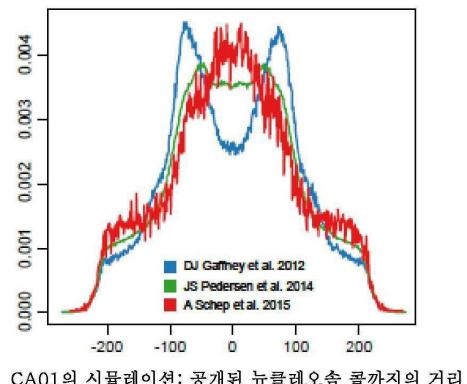
도면63f



도면63g



도면63h



서열목록

SEQUENCE LISTING

<110> SHENDURE, Jay

<120> METHODS OF DETERMINING TISSUES AND/OR CELL TYPES GIVING RISE TO CELL-FREE DNA, AND METHODS OF IDENTIFYING A DISEASE OR DISORDER USING SAME

<130> 72227-8115.W000

<140> PCT/US2015/042310

<141> 2015-07-27

<150> US 62/029,178

<151> 2014-07-25

<150> US 62/087,619

<151> 2014-12-04

<160> 4

<170> PatentIn version 3.5

<210> 1
 <211> 34
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic oligonucleotide: CL9

<400> 1
 gtgactggag ttcagacgtg tgctttccg atct 34

<210> 2
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic oligonucleotide: Adapter 2.1

<220><221> misc_feature
 <222> (16)..(16)

<223> ddT at 3' end

<400> 2

cgacgctt ccgatc 16

<210> 3
 <211> 34
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic oligonucleotide: Adapter 2.2

<220><221> misc_feature

<222>

> (1)..(1)

<223> 5Phos at 5' end

<220><221> misc_feature

<222> (30)..(31)

<223> Phosphorothioate bond

<220><221> misc_feature

<222> (31)..(32)

<223> Phosphorothioate bond

<220><221> misc_feature

<222> (32)..(33)

<223> Phosphorothioate bond

<220><221> misc_feature

<222> (33)..(34)

<223> Phosphorothioate bond

<400> 3

agatcggaag agcgctgtgt agggaaagag tgta

34

<210> 4

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223>

> Synthetic oligonucleotide: CL78

<220><221> misc_feature

<222> (1)..(1)

<223> 5Phos at 5' end

<220><221> misc_feature

<222> (10)..(10)

<223> (iSpC3)10 and 3BioTEG at 3' end

<400> 4

agatcggaag

10