

청구범위

- [청구항 1] a) 생체시료에서 핵산을 추출하여 서열정보를 획득하는 단계;
 b) 획득한 서열정보(reads)를 표준 염색체 서열 데이터베이스(reference genome database)에 정렬(alignment)하는 단계;
 c) 상기 정렬된 서열정보(reads) 기반의 핵산단편(fragments)을 이용한 벡터화된 데이터를 생성하는 단계;
 d) 생성된 상기 벡터화된 데이터를 학습된 인공지능 모델에 입력하여 분석하고, 기준값(cut-off value)과 비교하여 염색체 이상 유무를 판정하는 단계를 포함하는 인공지능 기반 염색체 이상 검출방법.
- [청구항 2] 제1항에 있어서, 상기 a) 단계는 다음의 단계를 포함하는 방법으로 수행되는 것을 특징으로 하는 인공지능 기반 염색체 이상 검출방법:
 (a-i) 혈액, 정액, 질 세포, 모발, 타액, 소변, 구강세포, 태반세포 또는 태아세포를 포함하는 양수, 조직세포 및 이의 혼합물에서 핵산을 수득하는 단계;
 (a-ii) 채취된 핵산에서 솔팅-아웃 방법(salting-out method), 컬럼 크로마토그래피 방법(column chromatography method) 또는 비드 방법(beads method)을 사용하여 단백질, 지방, 및 기타 잔여물을 제거하고 정제된 핵산을 수득하는 단계;
 (a-iii) 정제된 핵산 또는 효소적 절단, 분쇄, 수압 절단 방법(hydroshear method)으로 무작위 단편화(random fragmentation)된 핵산에 대하여, 싱글 엔드 시퀀싱(single-end sequencing) 또는 페어 엔드 시퀀싱(pair-end sequencing) 라이브러리(library)를 제작하는 단계;
 (a-iv) 제작된 라이브러리를 차세대 유전자서열검사기(next-generation sequencer)에 반응시키는 단계; 및
 (a-v) 차세대 유전자서열검사기에서 핵산의 서열정보(reads)를 획득하는 단계.
- [청구항 3] 제1항에 있어서, 상기 c) 단계의 벡터화된 데이터는 그랜드 캐년 플롯(Grand Canyon plot, GC plot) 또는 지노믹 캐슬 월 플롯(Genomic Castle Wall plot, GCW plot)인 것을 특징으로 하는 인공지능 기반 염색체 이상 검출 방법.
- [청구항 4] 제3항에 있어서, 상기 GC plot은 정렬된 핵산단편의 염색체 구간 별 분포를 구간 별 수(count) 또는 핵산단편(fragment) 사이의 거리를 계산하여 벡터화된 데이터로 생성하는 것을 특징으로 하는 인공지능 기반 염색체 이상 검출 방법.
- [청구항 5] 제4항에 있어서, 상기 염색체 구간 별 분포를 핵산단편의 수로 계산하는 것은 하기의 단계를 포함하여 수행하는 것을 특징으로 하는 인공지능 기반 염색체 이상 검출 방법:

- i) 염색체를 일정구간(bin)으로 구분하는 단계;
 ii) 각 구간에 정렬된 핵산단편의 수를 결정하는 단계;
 iii) 각 구간에 결정된 핵산단편 수를 샘플의 전체 핵산단편 수로 나누어 정규화(normalization)하는 단계; 및
 iv) 각 구간의 순서를 X 축 값으로 하고, 상기 iii) 단계에서 계산한 정규화 값을 Y축 값으로 하여 GC plot을 생성하는 단계.
- [청구항 6] 제4항에 있어서, 상기 염색체 구간 별 분포를 핵산단편 사이의 거리로 계산하는 것은 하기의 단계를 포함하여 수행하는 것을 특징으로 하는 인공지능 기반 염색체 이상 검출 방법:
 i) 염색체를 일정구간(bin)으로 구분하는 단계;
 ii) 각 구간에 정렬된 핵산단편 사이의 거리를 계산하는 단계;
 iii) 각 구간별로 계산된 거리값을 기반으로 각 구간의 거리의 대표값(RepFD)을 결정하는 단계;
 iv) 상기 iii) 단계에서 계산된 대표값을 전체 핵산단편 사이의 거리 값의 대표값으로 나누어 정규화(normalization)하는 단계; 및
 iv) 각 구간의 순서를 X 축 값으로 하고, 상기 iv) 단계에서 계산한 정규화 값을 Y축 값으로 하여 GC plot을 생성하는 단계.
- [청구항 7] 제6항에 있어서, 상기 대표값은 핵산단편 사이의 거리의 합, 차, 곱, 평균, 중앙값, 분위수, 최소값, 최대값, 분산, 표준편차, 중앙값 절대편차, 변동계수, 이들의 역수값 및 이들의 조합으로 구성된 군에서 선택되는 하나 이상인 것을 특징으로 하는 인공지능 기반 염색체 이상 검출방법.
- [청구항 8] 제3항에 있어서, 상기 GCW plot은 정렬된 핵산단편 사이의 거리를 염색체 단위로 계산하고, 정상 염색체와 이수성을 판별하고자 하는 염색체의 거리 값을 교차 배열하여 그래프로 생성하는 것을 특징으로 하는 인공지능 기반 염색체 이상 검출방법.
- [청구항 9] 제8항에 있어서, 상기 GCW plot은 하기의 단계를 포함하는 방법으로 수행되는 것을 특징으로 하는 인공지능 기반 염색체 이상 검출방법:
 i) 각 염색체 별로 정렬된 핵산단편 사이의 거리를 계산하는 단계;
 ii) 상기 i) 단계에서 계산된 거리값을 기반으로 각 구간의 거리의 대표값을 결정하는 단계;
 iii) 상기 ii) 단계에서 결정된 염색체별 대표값을 표준화(standardization)하는 단계;
 iv) 이수성을 판별할 염색체를 제외한 염색체들 중 일부를 비교염색체군으로 선발하는 단계; 및
 iv) 비교염색체군에 속하는 염색체들과 이수성을 판별할 염색체를 X축에 따라 순차적으로 교차 배열하고, 각 염색체의 표준화된 계산값을 Y값으로 하는 GCW plot을 생성하는 단계.
- [청구항 10] 제9항에 있어서, 상기 iii) 단계의 대표값을 표준화하는 단계는 하기의

단계를 포함하는 방법으로 수행되는 것을 특징으로 하는 인공지능 기반
염색체 이상 검출방법:

- 1) 염색체 이수성이 검출되지 않은 정상 샘플로 구성된 표준 집단을
설정하는 단계;
- 2) 상기 표준 집단에서 관찰되는 염색체별 RepFD 값의
평균(Reference_Mean)과 표준편차(Reference_Standard_Deviation)를
계산하는 단계;
- 3) 상기 2) 단계에서 계산한 평균과 표준편차를 하기 수식 1에 대입하여 Z
표준화하는 단계;

수식 1:

$$Z_{chr} = ((RepFD_{chr} - Reference_Mean_{chr}) / Reference_Standard_Deviation_{chr}) + 5$$

- 4) 상기 3) 단계에서 계산한 염색체별 Zchr 값을 이수성을 판별하고자
하는 염색체의 Zchr 값으로 나누는 단계.

[청구항 11] 제9항에 있어서, 상기 대표값은 핵산단편 사이의 거리의 합, 차, 곱, 평균,
중앙값, 분위수, 최소값, 최대값, 분산, 표준편차, 중앙값 절대편차,
변동계수, 이들의 역수값 및 이들의 조합으로 구성된 군에서 선택되는
하나 이상의 값인 것을 특징으로 하는 인공지능 기반 염색체 이상
검출방법.

[청구항 12] 제1항에 있어서, 상기 (d) 단계의 인공지능 모델은 염색체 상태가 정상인
벡터화된 데이터와 염색체 이상이 있는 벡터화된 데이터를 구별할 수
있도록 학습하는 것을 특징으로 하는 인공지능 기반 염색체 이상
검출방법.

[청구항 13] 제12항에 있어서, 상기 인공지능 모델은 합성곱 신경망(convolutional
neural network, CNN), 심층 신경망(Deep Neural Network, DNN), 순환
신경망(Recurrent Neural Network, RNN) 및 오토 인코더(autoencoder)로
구성된 군에서 선택되는 것을 특징으로 하는 인공지능 기반 염색체 이상
검출방법.

[청구항 14] 제12항에 있어서, 상기 인공지능 모델이 CNN이고, binary classification 을
학습할 경우, 손실함수는 하기 수식 2로 표시되며, 상기 인공지능 모델이
CNN이고, Multi-class classification을 학습할 경우, 손실함수는 하기 수식
3으로 표시되는 것을 특징으로 하는 인공지능 기반 염색체 이상
검출방법:

수식 2: Binary classification

$$\text{loss}(\text{model}(x), y) = -\frac{1}{n} \left[\sum_{i=1}^n (y_i \log(\text{model}(x_i)) + (1 - y_i) \log(1 - \text{model}(x_i))) \right]$$

$\text{model}(x_i) = i$ 번째 input에 인공지능 model output

$y =$ 실제 label 값

$n =$ input data 수

수식 3: Multi-class classification

$$\text{loss}(\text{model}(x), y) = -\frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^c (y_{ij} \log(\text{model}(x_i)_j)) \right)$$

$\text{model}(x_i)_j = i$ 번째 input에 j 번째 인공지능 model output

$y =$ 실제 label 값

$n =$ input data 수

$c =$ class 수

- [청구항 15] 제1항에 있어서, 상기 d) 단계의 인공지능 모델이 입력된 벡터화된 데이터를 분석하여 출력하는 결과값은 DPI(Deep Probability Index)값인 것을 특징으로 하는 인공지능 기반 염색체 이상 검출방법.
- [청구항 16] 제1항에 있어서, 상기 d) 단계의 기준값은 0.5이며, 0.5 이상일 경우, 염색체 이상이 있는 것으로 판정하는 것을 특징으로 하는 인공지능 기반 염색체 이상 검출방법.
- [청구항 17] 생체시료에서 핵산을 추출하여 서열정보를 해독하는 해독부; 해독된 서열을 표준 염색체 서열 데이터베이스에 정렬하는 정렬부; 및 정렬된 서열 기반의 핵산단편을 이용한 벡터화된 데이터를 생성하는 데이터 생성부; 생성된 벡터화된 데이터를 학습된 인공지능 모델에 입력하여 분석하고, 기준값과 비교하여 염색체 이상 유무를 판정하는 염색체 이상 판정부를 포함하는 인공지능 기반 염색체 이상 검출 장치.
- [청구항 18] 컴퓨터 판독 가능한 저장 매체로서, 염색체 이상을 검출하는 프로세서에 의해 실행되도록 구성되는 명령을 포함하되,
 a) 생체시료에서 핵산을 추출하여 서열정보를 획득하는 단계;
 b) 획득한 서열정보(reads)를 표준 염색체 서열 데이터베이스(reference genome database)에 정렬(alignment)하는 단계;
 c) 상기 정렬된 서열정보(reads) 기반의 핵산단편을 이용한 벡터화된

데이터를 생성하는단계;

d) 생성된 상기 벡터화된 데이터를 학습된 인공지능 모델에 입력하여 분석하고, 기준값(cut-off value)을 비교하여 염색체 이상 유무를 판정하는 단계를 통하여, 염색체 이상을 검출하는 프로세서에 의해 실행되도록 구성되는 명령을 포함하는 컴퓨터 판독 가능한 저장 매체.