



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2022년05월13일

(11) 등록번호 10-2397822

(24) 등록일자 2022년05월10일

(51) 국제특허분류(Int. Cl.)
G16B 15/00 (2019.01) G16B 30/10 (2019.01)

G16B 40/00 (2019.01)

(52) CPC특허분류

G16B 15/00 (2019.02)

G16B 30/10 (2019.02)

(21) 출원번호 10-2021-0160058

(22) 출원일자 2021년11월19일

심사청구일자 2021년11월19일

(56) 선행기술조사문헌

H. Yang 외, "Cancer Classification based on Chromatin accessibility profiles with deep adversarial learning model", PLOS Computational Biology, 2020.11.09.

Y. Sun 외, "Detect accesible chromatin using ATAC-sequencing, from principlr to applications", Hereditas 156, 2029.08.12.

(73) 특허권자

주식회사 클리노믹스

울산광역시 울주군 언양읍 유니스트길 50, 110동 3층 301-3호

울산과학기술원

울산광역시 울주군 언양읍 유니스트길 50

(72) 발명자

박중화

울산광역시 울주군 언양읍 반천강변길 51, 101동 1710호(현대아파트)

김병철

울산광역시 울주군 삼남읍 도호1길 39-11, 202동 2401호(금아드림팰리스)

(뒷면에 계속)

(74) 대리인

김순웅

전체 청구항 수 : 총 14 항

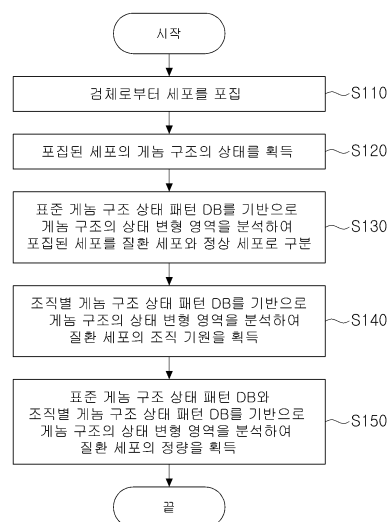
심사관 : 성경아

(54) 발명의 명칭 염색체 구조의 상태 정보를 이용한 세포 분석 장치 및 방법

(57) 요약

염색체 구조의 상태 정보를 이용한 세포 분석 장치 및 방법이 개시된다. 본 발명은, 염색체 구조의 상태 분석을 통해 피검자의 검체로부터 포집된 세포군에 질환 세포가 존재하는지 판별하고, 질환 세포의 조직 기원(tissue origin)과 정량을 예측한다. 본 발명에 따르면, 저렴한 가격에 정확도 높은 질환 세포 구분을 할 수 있고, 기존에 사용하고 있는 세포 염색법보다 쉽고 정확하게 정량 측정을 할 수 있다.

대표도 - 도11



(52) CPC특허분류

G16B 40/00 (2019.02)

(72) 발명자

김창재

울산광역시 울주군 범서읍 대리로 15-12, 101동
904호 (호반베르디움)

안지혜

울산광역시 중구 북부순환도로 680, 1동 904호(복
산동, 새운파레스타운)

엄효진

울산광역시 남구 장생포고래로179번길 4

전하현

울산광역시 남구 와화로 17, 2층(무거동)

김여진

경상남도 양산시 물금읍 가촌서로 11, 102동 501
호(양산물금 한신 더휴 아파트)

이 발명을 지원한 국가연구개발사업

과제고유번호	1425156792
과제번호	P0016195
부처명	중소벤처기업부
과제관리(전문)기관명	한국산업기술진흥원
연구사업명	규제자유특구혁신사업육성(R&D)
연구과제명	지능형 오믹스 빅데이터 기반 질병 예측 및 진단 마커 개발 실증
기 여 율	1/1
과제수행기관명	클리노믹스
연구기간	2021.01.01 ~ 2022.11.30

명세서

청구범위

청구항 1

검체로부터 포집된 세포의 계놈 구조의 상태를 획득하는 단계;

미리 저장되어 있는 표준 계놈 구조 상태 패턴 DB를 기반으로, 계놈 구조의 상태 변형 영역을 분석하여 상기 포집된 세포를 질환 세포와 정상 세포로 구분하는 단계;

미리 저장되어 있는 조직별 계놈 구조 상태 패턴 DB를 기반으로, 계놈 구조의 상태 변형 영역을 분석하여 상기 질환 세포의 조직 기원(tissue origin)을 획득하는 단계; 및

상기 표준 계놈 구조 상태 패턴 DB와 상기 조직별 계놈 구조 상태 패턴 DB를 기반으로, 계놈 구조의 상태 변형 영역을 분석하여 상기 질환 세포의 정량을 획득하는 단계;

를 포함하는 염색체 구조의 상태 정보를 이용한 세포 분석 방법.

청구항 2

제1항에서,

상기 구분 단계는,

계놈 구조의 상태 변형 영역의 서열 수와 피크(peak)를 기반으로, 상기 표준 계놈 구조 상태 패턴 DB에 저장된 계놈 구조의 상태와 상기 포집된 세포의 계놈 구조의 상태를 대비하여, 상기 포집된 세포를 상기 질환 세포와 상기 정상 세포로 구분하는 것으로 이루어지는,

염색체 구조의 상태 정보를 이용한 세포 분석 방법.

청구항 3

제1항에서,

상기 조직 기원 획득 단계는,

계놈 구조의 상태 변형 영역의 피크(peak) 패턴을 기반으로, 상기 조직별 계놈 구조 상태 패턴 DB에 저장된 계놈 구조의 상태와 상기 포집된 세포의 계놈 구조의 상태를 대비하여, 상기 질환 세포의 조직 기원을 획득하는 것으로 이루어지는,

염색체 구조의 상태 정보를 이용한 세포 분석 방법.

청구항 4

제3항에서,

상기 조직 기원 획득 단계는,

상기 조직별 계놈 구조 상태 패턴 DB에 저장된 계놈 구조의 피크 위치와 상기 포집된 세포의 계놈 구조의 피크 위치를 대비하여, 상기 질환 세포의 조직 기원을 획득하는 것으로 이루어지는,

염색체 구조의 상태 정보를 이용한 세포 분석 방법.

청구항 5

제3항에서,

상기 조직 기원 획득 단계는,

상기 조직별 계놈 구조 상태 패턴 DB에 저장된 계놈 구조의 피크 영역과 상기 포집된 세포의 계놈 구조의 피크 영역이 서로 오버랩(overlap)되는 비율을 기반으로, 상기 질환 세포의 조직 기원을 획득하는 것으로

이루어지는,

염색체 구조의 상태 정보를 이용한 세포 분석 방법.

청구항 6

제3항에서,

상기 조직 기원 획득 단계는,

상기 조직별 게놈 구조 상태 패턴 DB에 저장된 게놈 구조의 피크 스코어(peak score)를 기반으로 획득된 매트릭스(matrix)와 상기 포집된 세포의 게놈 구조의 피크 스코어를 기반으로 획득된 매트릭스를 대비하여, 상기 질환 세포의 조직 기원을 획득하는 것으로 이루어지는,

염색체 구조의 상태 정보를 이용한 세포 분석 방법.

청구항 7

제1항에서,

상기 질환 세포 정량 획득 단계는,

상기 표준 게놈 구조 상태 패턴 DB와 상기 조직별 게놈 구조 상태 패턴 DB를 기반으로, 상기 질환 세포의 특이적 게놈 구조의 상태 변형 영역을 대상으로 하여 획득된 서열 수와 상기 정상 세포의 특이적 게놈 구조의 상태 변형 영역을 대상으로 하여 획득된 서열 수를 이용하여, 전체 세포의 수 대비 상기 질환 세포의 수를 계산하여 상기 질환 세포의 정량을 획득하는 것으로 이루어지는,

염색체 구조의 상태 정보를 이용한 세포 분석 방법.

청구항 8

검체로부터 포집된 세포의 게놈 구조의 상태를 획득하는 세포 분석부;

미리 저장되어 있는 표준 게놈 구조 상태 패턴 DB를 기반으로, 게놈 구조의 상태 변형 영역을 분석하여 상기 포집된 세포를 질환 세포와 정상 세포로 구분하는 세포 구분부;

미리 저장되어 있는 조직별 게놈 구조 상태 패턴 DB를 기반으로, 게놈 구조의 상태 변형 영역을 분석하여 상기 질환 세포의 조직 기원(tissue origin)을 획득하는 세포 기원 획득부; 및

상기 표준 게놈 구조 상태 패턴 DB와 상기 조직별 게놈 구조 상태 패턴 DB를 기반으로, 게놈 구조의 상태 변형 영역을 분석하여 상기 질환 세포의 정량을 획득하는 세포 정량 획득부;

를 포함하는 염색체 구조의 상태 정보를 이용한 세포 분석 장치.

청구항 9

제8항에서,

상기 세포 구분부는,

게놈 구조의 상태 변형 영역의 서열 수와 피크(peak)를 기반으로, 상기 표준 게놈 구조 상태 패턴 DB에 저장된 게놈 구조의 상태와 상기 포집된 세포의 게놈 구조의 상태를 대비하여, 상기 포집된 세포를 상기 질환 세포와 상기 정상 세포로 구분하는,

염색체 구조의 상태 정보를 이용한 세포 분석 장치.

청구항 10

제8항에서,

상기 세포 기원 획득부는,

게놈 구조의 상태 변형 영역의 피크(peak) 패턴을 기반으로, 상기 조직별 게놈 구조 상태 패턴 DB에 저장된 게놈 구조와 상기 포집된 세포의 게놈 구조를 대비하여, 상기 질환 세포의 조직 기원을 획득하는,

염색체 구조의 상태 정보를 이용한 세포 분석 장치.

청구항 11

제10항에서,

상기 세포 기원 획득부는,

상기 조직별 게놈 구조 상태 패턴 DB에 저장된 게놈 구조의 피크 위치와 상기 포집된 세포의 게놈 구조의 피크 위치를 대비하여, 상기 질환 세포의 조직 기원을 획득하는,

염색체 구조의 상태 정보를 이용한 세포 분석 장치.

청구항 12

제10항에서,

상기 세포 기원 획득부는,

상기 조직별 게놈 구조 상태 패턴 DB에 저장된 게놈 구조의 피크 영역과 상기 포집된 세포의 게놈 구조의 피크 영역이 서로 오버랩(overlap)되는 비율을 기반으로, 상기 질환 세포의 조직 기원을 획득하는,

염색체 구조의 상태 정보를 이용한 세포 분석 장치.

청구항 13

제10항에서,

상기 세포 기원 획득부는,

상기 조직별 게놈 구조 상태 패턴 DB에 저장된 게놈 구조의 피크 스코어(peak score)를 기반으로 획득된 매트릭스(matrix)와 상기 포집된 세포의 게놈 구조의 피크 스코어를 기반으로 획득된 매트릭스를 대비하여, 상기 질환 세포의 조직 기원을 획득하는,

염색체 구조의 상태 정보를 이용한 세포 분석 장치.

청구항 14

제8항에서,

상기 세포 정량 획득부는,

상기 표준 게놈 구조 상태 패턴 DB와 상기 조직별 게놈 구조 상태 패턴 DB를 기반으로, 상기 질환 세포의 특이적 게놈 구조의 상태 변형 영역을 대상으로 하여 획득된 서열 수와 상기 정상 세포의 특이적 게놈 구조의 상태 변형 영역을 대상으로 하여 획득된 서열 수를 이용하여, 전체 세포의 수 대비 상기 질환 세포의 수를 계산하여 상기 질환 세포의 정량을 획득하는,

염색체 구조의 상태 정보를 이용한 세포 분석 장치.

발명의 설명

기술 분야

[0001]

본 발명은 염색체 구조상의 열리고, 닫힘 등의 상태의 변형 혹은 변화의 정보를 찾아 비교를 통해 질병 등의 예측 및 진단을 하는 세포 분석 장치 및 방법에 관한 것으로, 더욱 상세하게는 피검자의 검체에서 포집된 세포군 중에 질환세포가 있는지 판별하고, 염색체 구조의 열리고 닫힘의 정도에 따른 변형과 변화의 분석을 한다. 또, 이를 통해 질환 세포의 조직 기원(tissue origin)과 정량을 예측하는 장치 및 방법에 관한 것이다.

배경 기술

[0002]

종래에는 혈중 순환 종양 세포(CTC)나 장기의 상피 세포(epithelial cell) 등을 단순하고 특정화된 바이오 마커를 이용하여 확인하고 있다. 그러나, 혈중 순환 종양 세포(CTC)나 장기의 상피 세포(epithelial cell) 등은 암 환자나 염증 및 심장병 환자의 혈액과 소변 등에서 매우 적게 존재하여, 액체생검 분석 디바이스나 키트를 활용

하여 축적(enrichment)해도 정확한 검출이 어려운 문제가 있다.

선행기술문헌

특허문헌

- [0003] (특허문헌 0001) 한국공개특허 제2018-0009763호 (지토비전 게엠베하) 2018. 1. 29. 특허문헌 1은 염색체 이상의 검출 방법으로서, 특허문헌 1에는 반응계내 하이브리드화를 사용하여 염색체 이상, 특히 구조적 염색체 이상을 확인하는 방법에 대한 내용이 개시되어 있다.

발명의 내용

해결하려는 과제

- [0004] 본 발명이 이루고자 하는 기술적 과제는, 염색체 구조의 상태와 그 상태의 패턴들을 분석하여 피검자의 검체로부터 포집된 세포군에 질환 세포가 존재하는지 판별하고, 질환 세포의 조직 기원(tissue origin)과 정량을 예측하는 염색체 구조 정보를 이용한 세포 분석 장치 및 방법을 제공하는 데 있다.

과제의 해결 수단

- [0005] 상기의 기술적 과제를 달성하기 위한 본 발명에 따른 염색체 구조의 상태 정보를 이용한 세포 분석 방법은, 검체로부터 포집된 세포의 게놈 구조의 상태를 획득하는 단계; 미리 저장되어 있는 표준 게놈 구조 상태 패턴 DB를 기반으로, 게놈 구조의 상태 변형 영역을 분석하여 상기 포집된 세포를 질환 세포와 정상 세포로 구분하는 단계; 미리 저장되어 있는 조직별 게놈 구조 상태 패턴 DB를 기반으로, 게놈 구조 상태의 열리고, 닫힘 등의 변형 혹은 변화 영역을 분석하여 상기 질환 세포의 조직 기원(tissue origin)을 획득하는 단계; 및 상기 표준 게놈 구조 상태 패턴 DB와 상기 조직별 게놈 구조 상태 패턴 DB를 기반으로, 게놈 구조 상태의 변형 또는 변화 영역을 분석하여 상기 질환 세포의 정량을 획득하는 단계;를 포함한다. 변형이라 함은 염색체의 저장상태가 정상 등과 비교해서 변형이 일어난 것이고, 그 결과로 상대적으로 생긴 변화를 염색체상의 구조의 상태변화라고 한다.
- [0006] 상기 구분 단계는, 게놈 구조의 상태 변형 영역의 서열 수와 피크(peak)를 기반으로, 상기 표준 게놈 구조 상태 패턴 DB에 저장된 게놈 구조의 상태와 상기 포집된 세포의 게놈 구조의 상태를 대비하여, 상기 포집된 세포를 상기 질환 세포와 상기 정상 세포로 구분하는 것으로 이루어질 수 있다.
- [0007] 상기 조직 기원 획득 단계는, 게놈 구조의 상태 변형 영역의 피크(peak) 패턴을 기반으로, 상기 조직별 게놈 구조 상태 패턴 DB에 저장된 게놈 구조의 상태와 상기 포집된 세포의 게놈 구조의 상태를 대비하여, 상기 질환 세포의 조직 기원을 획득하는 것으로 이루어질 수 있다.
- [0008] 상기 조직 기원 획득 단계는, 상기 조직별 게놈 구조 상태 패턴 DB에 저장된 게놈 구조의 피크 위치와 상기 포집된 세포의 게놈 구조의 피크 위치를 대비하여, 상기 질환 세포의 조직 기원을 획득하는 것으로 이루어질 수 있다.
- [0009] 상기 조직 기원 획득 단계는, 상기 조직별 게놈 구조 상태 패턴 DB에 저장된 게놈 구조의 피크 영역과 상기 포집된 세포의 게놈 구조의 피크 영역이 서로 오버랩(overlap)되는 비율을 기반으로, 상기 질환 세포의 조직 기원을 획득하는 것으로 이루어질 수 있다.
- [0010] 상기 조직 기원 획득 단계는, 상기 조직별 게놈 구조 상태 패턴 DB에 저장된 게놈 구조의 피크 스코어(peak score)를 기반으로 획득된 매트릭스(matrix)와 상기 포집된 세포의 게놈 구조의 피크 스코어를 기반으로 획득된 매트릭스를 대비하여, 상기 질환 세포의 조직 기원을 획득하는 것으로 이루어질 수 있다.
- [0011] 상기 질환 세포 정량 획득 단계는, 상기 표준 게놈 구조 상태 패턴 DB와 상기 조직별 게놈 구조 상태 패턴 DB를 기반으로, 상기 질환 세포의 특이적 게놈 구조의 상태 변형 영역을 대상으로 하여 획득된 서열 수와 상기 정상 세포의 특이적 게놈 구조의 상태 변형 영역을 대상으로 하여 획득된 서열 수를 이용하여, 전체 세포의 수 대비 상기 질환 세포의 수를 계산하여 상기 질환 세포의 정량을 획득하는 것으로 이루어질 수 있다.
- [0013] 상기의 기술적 과제를 달성하기 위한 본 발명에 따른 염색체 구조의 상태 또는 상태변화정보를 이용한 세포 분

석 장치는, 검체로부터 포집된 세포의 게놈 구조의 상태를 획득하는 세포 분석부; 미리 저장되어 있는 표준 게놈 구조 상태 패턴 DB를 기반으로, 게놈 구조의 상태 변형 영역을 분석하여 상기 포집된 세포를 질환 세포와 정상 세포로 구분하는 세포 구분부; 미리 저장되어 있는 조직별 게놈 구조 상태 패턴 DB를 기반으로, 게놈 구조의 상태 변형 영역을 분석하여 상기 질환 세포의 조직 기원(tissue origin)을 획득하는 세포 기원 획득부; 및 상기 표준 게놈 구조 상태 패턴 DB와 상기 조직별 게놈 구조 상태 패턴 DB를 기반으로, 게놈 구조의 상태 변형 영역을 분석하여 상기 질환 세포의 정량을 획득하는 세포 정량 획득부;를 포함한다.

- [0014] 상기 세포 구분부는, 게놈 구조의 상태 변형 영역의 서열 수와 피크(peak)를 기반으로, 상기 표준 게놈 구조 상태 패턴 DB에 저장된 게놈 구조의 상태와 상기 포집된 세포의 게놈 구조의 상태를 대비하여, 상기 포집된 세포를 상기 질환 세포와 상기 정상 세포로 구분할 수 있다.
- [0015] 상기 세포 기원 획득부는, 게놈 구조의 상태 변형 영역의 피크(peak) 패턴을 기반으로, 상기 조직별 게놈 구조 상태 패턴 DB에 저장된 게놈 구조의 상태와 상기 포집된 세포의 게놈 구조의 상태를 대비하여, 상기 질환 세포의 조직 기원을 획득할 수 있다.
- [0016] 상기 세포 기원 획득부는, 상기 조직별 게놈 구조 상태 패턴 DB에 저장된 게놈 구조의 피크 위치와 상기 포집된 세포의 게놈 구조의 피크 위치를 대비하여, 상기 질환 세포의 조직 기원을 획득할 수 있다.
- [0017] 상기 세포 기원 획득부는, 상기 조직별 게놈 구조 상태 패턴 DB에 저장된 게놈 구조의 피크 영역과 상기 포집된 세포의 게놈 구조의 피크 영역이 서로 오버랩(overlap)되는 비율을 기반으로, 상기 질환 세포의 조직 기원을 획득할 수 있다.
- [0018] 상기 세포 기원 획득부는, 상기 조직별 게놈 구조 상태 패턴 DB에 저장된 게놈 구조의 피크 스코어(peak score)를 기반으로 획득된 매트릭스(matrix)와 상기 포집된 세포의 게놈 구조의 피크 스코어를 기반으로 획득된 매트릭스를 대비하여, 상기 질환 세포의 조직 기원을 획득할 수 있다.
- [0019] 상기 세포 정량 획득부는, 상기 표준 게놈 구조 상태 패턴 DB와 상기 조직별 게놈 구조 상태 패턴 DB를 기반으로, 상기 질환 세포의 특이적 게놈 구조의 상태 변형 영역을 대상으로 하여 획득된 서열 수와 상기 정상 세포의 특이적 게놈 구조의 상태 변형 영역을 대상으로 하여 획득된 서열 수를 이용하여, 전체 세포의 수 대비 상기 질환 세포의 수를 계산하여 상기 질환 세포의 정량을 획득할 수 있다.

발명의 효과

- [0020] 본 발명에 따른 염색체 구조의 상태 정보를 이용한 세포 분석 장치 및 방법에 의하면, 염색체 구조의 상태 분석을 통해 피검자의 검체로부터 포집된 세포군에 질환 세포가 존재하는지 판별함으로써, 저렴한 가격에 정확도 높은 구분을 할 수 있다.
- [0021] 그리고, 질환 세포의 조직 기원(tissue origin)과 정량을 예측함으로써, 기존에 사용하고 있는 세포 염색법보다 쉽고 정확하게 정량 측정을 할 수 있다.
- [0022] 또한, 질환 세포 중 순환 종양 세포(CTC)의 게놈 구조의 상태 변형 영역은 암세포와 같은 질환 유래 세포의 유전자 및 외유전자(epigenetic) 변형에 연관이 높은 영역으로, 이를 분석함으로써 다양한 다른 암 분자 마커에 연관 적용할 수 있고, 이 방법론은 심장병 같은 다른 질병에도 같은 원리로 적용할 수 있다.
- [0023] 아울러, 질환 세포의 게놈 구조 기반의 분석을 통하여, 구조와 연관된 질환 유전자 기능 마커, 외유전체(epigenomic) 마커, 돌연변이 마커 등 다중오믹스(multi-omics) 다중 마커 개발과 쉽게 연계할 수 있다.

도면의 간단한 설명

- [0024] 도 1은 본 발명의 바람직한 실시예에 따른 염색체 구조의 상태 정보를 이용한 세포 분석 장치를 설명하기 위한 블록도이다.
- 도 2는 염색체 구조의 상태 정보를 설명하기 위한 그림이다.
- 도 3은 조직 타입에 따른 세포주(cell line)별 염색체 구조의 상태 정보의 차이를 설명하기 위한 그림이다.
- 도 4는 본 발명에 따른 ATAC-seq를 이용한 유크로마틴(euchromatin) 영역의 해독 예시를 설명하기 위한 그림이다.
- 도 5는 본 발명에 따른 게놈 구조의 상태 변형 영역의 서열 수와 피크(peak)를 설명하기 위한 그림이다.

도 6은 본 발명에 따른 표준 백혈구 계놈 구조와 질환 세포의 계놈 구조의 대비 예시를 설명하기 위한 그림이다.

도 7은 본 발명에 따른 조직/질환 특이적 계놈 구조의 패턴 대비 예시를 설명하기 위한 그림이다.

도 8은 본 발명에 따른 계놈 구조의 피크 위치를 이용한 계놈 구조 패턴 대비를 설명하기 위한 도면이다.

도 9는 본 발명의 예시적인 실시예에 따른 계놈 구조의 피크 영역 오버랩 비율을 이용한 계놈 구조 패턴 대비를 설명하기 위한 도면이다.

도 10은 본 발명에 따른 계놈 구조의 피크 스코어를 이용한 계놈 구조 패턴 대비를 설명하기 위한 도면이다.

도 11은 본 발명의 바람직한 실시예에 따른 염색체 구조의 상태 정보를 이용한 세포 분석 방법을 설명하기 위한 흐름도이다.

도 12는 본 발명에 따라 포집 세포의 계놈 구조의 상태 변형 영역을 분석함으로써 그 포집 세포를 정상 세포와 질환 세포로 구분할 수 있음을 실험 데이터로 확인하기 위한 예시적인 실험 단계를 도식화한 흐름도이다.

도 13은 질환 세포 샘플에서 분리된 세포에 대해서 시퀀싱 데이터를 분석한 결과이고, 도 14는 정상 세포 샘플에서 분리된 세포에 대해서 시퀀싱 데이터를 분석한 결과이다.

도 15는 실험군 샘플 1에서 분리된 세포에 대해서 시퀀싱 데이터를 분석한 결과이고, 도 16은 실험군 샘플 3에서 분리된 세포에 대해서 시퀀싱 데이터를 분석한 결과이다.

도 17은 정상 세포 샘플과 실험군 샘플 3으로부터 각각 분리된 세포의 시퀀싱 데이터에서 검출한 피크의 크기를 동일 염색체의 위치에 대해서 비교한 그래프이다.

도 18은 실험군 샘플 1을 시퀀싱하여 나오는 서열에서 정상 세포 샘플에서 보이지 않는 질환 세포 특이적 영역을 컴퓨터 프로그램으로 분석한 결과 데이터이고, 도 19는 실험군 샘플 2를 시퀀싱하여 나오는 서열에서 정상 세포 샘플에서 보이지 않는 질환 세포 특이적 영역을 컴퓨터 프로그램으로 분석한 결과 데이터이고, 도 20은 실험군 샘플 3을 시퀀싱하여 나오는 서열에서 정상 세포 샘플에서 보이지 않는 질환 세포 특이적 영역을 컴퓨터 프로그램으로 분석한 결과 데이터이다.

발명을 실시하기 위한 구체적인 내용

- [0025] 이하에서 첨부한 도면을 참조하여 본 발명에 따른 염색체 구조의 상태 정보를 이용한 세포 분석 장치 및 방법의 바람직한 실시예에 대해 상세하게 설명한다.
- [0027] 먼저, 도 1 내지 도 3을 참조하여 본 발명의 바람직한 실시예에 따른 염색체 구조의 상태 정보를 이용한 세포 분석 장치에 대하여 설명한다.
- [0028] 도 1은 본 발명의 바람직한 실시예에 따른 염색체 구조의 상태 정보를 이용한 세포 분석 장치를 설명하기 위한 블록도이고, 도 2는 염색체 구조의 상태 정보를 설명하기 위한 그림이며, 도 3은 조직 타입에 따른 세포주(cell line)별 염색체 구조의 상태 정보의 차이를 설명하기 위한 그림이다.
- [0029] 도 1을 참조하면, 본 발명의 바람직한 실시예에 따른 염색체 구조의 상태 정보를 이용한 세포 분석 장치(이하 '세포 분석 장치'라 합니다)(100)는 염색체 구조의 상태 분석을 통해 피검자의 검체로부터 포집된 세포군에 질환 세포(예컨대, 순환 종양 세포, 순환 심혈관 외막 세포, 순환 염증 질환 상피 세포 등)가 존재하는지 판별하고, 질환 세포의 조직 기원(tissue origin)과 정량을 예측한다.
- [0030] 여기서, 염색체 구조(즉, 계놈 구조)의 상태 정보는 도 2에 도시된 바와 같이, 계놈의 서열, 계놈의 영역이 기능적으로 열린 것(open chromatin, euchromatin), 외유전자(epigenetic)들의 종합적 배치, 종류, 패턴 등과 같은 다양한 종류의 정보를 통칭한다. 유크로마틴(open chromatin)은 발현해야 할 유전자들이 많이 분포되어 있으며, 염색질(chromatin)의 밀도가 비교적 낮아 전사활성이 있는 영역이다. 헤테로크로마틴(heterochromatin, closed chromatin)은 염색질의 밀도가 비교적 높고, 전사활성이 낮아 유전자 발현이 억제된 영역이다. 이러한 염색체의 구조 정보를 알아보기 위한 방법으로 Assay for Transposase-Accessible Chromatin using sequencing(ATAC-seq)을 사용한다. ATAC-seq의 경우 염색체의 구조의 열림/닫힘의 상태 차이를 보는 서열해독 방법이며, 염색질이 기능적으로 열린 영역을 해독하여 계놈 상의 모든 유크로마틴 영역을 프로파일링할 수 있다.

- [0031] 그리고, 유크로마틴의 패턴은 도 3에 도시된 바와 같이, 각 세포의 조직 기원에 따라 다르게 나타나는 특징이 있다. 따라서, ATAC-seq 데이터의 분석을 통해 유크로마틴의 패턴을 프로파일링하고, 특정 세포의 조직 기원을 예측 및 분석할 수 있다. 즉, 특정 질환 환자의 혈액에 포함된 질환유래세포(순환 종양세포, 순환 염증질환 상피세포, 순환 심혈관 외막세포 등)에서 ATAC-seq을 분석하여 정상 백혈구 세포 게놈 구조 DB와 비교하여 질환 유래세포 특이적 게놈 구조 변이 영역 판독 후 질환의 유래된 세포를 예측/분석 할 수 있다. 예를 들어, 암환자의 혈액에 포함된 혈중 순환 종양 세포(Circulating Tumor Cell, CTC)를 검출해 염색체 구조 정보의 상태 분석을 통해 순환 종양 세포의 조직 기원(tissue origin)을 찾아낼 수 있다.
- [0033] 그러면, 도 4 내지 도 10을 참조하여 본 발명의 바람직한 실시예에 따른 염색체 구조의 상태 정보를 이용한 세포 분석 장치에 대하여 보다 자세하게 설명한다.
- [0034] 세포 분석 장치(100)는 도 1에 도시된 바와 같이 저장부(110), 세포 포집부(120), 세포 분석부(130), 세포 구분부(140), 세포 기원 획득부(150) 및 세포 정량 획득부(160)를 포함할 수 있다.
- [0036] 저장부(110)는 표준 게놈 구조 패턴 데이터 베이스(DB), 조직별 게놈 구조 패턴 데이터 베이스(DB) 등을 저장하고 있다.
- [0037] 여기서, 표준 게놈 구조 패턴 DB는 정상 세포로 간주할 수 있는 백혈구 세포의 게놈 구조의 상태 정보가 저장되어 있다. 인종별로 백혈구 세포의 게놈 구조 패턴이 상이할 수 있으므로, 표준 게놈 구조 패턴 DB는 인종별로 구축될 수도 있다.
- [0038] 그리고, 조직별 게놈 구조 패턴 DB는 조직별로 혹은 질환별(예컨대, 암 종별 등)로 조직/질환에 대응되는 게놈 구조의 상태 정보가 저장되어 있다.
- [0039] 이때, 표준 게놈 구조 상태 패턴 DB 또는 조직별 게놈 구조 상태 패턴 DB에 저장되어 있는 게놈 구조는, 게놈의 유크로마틴(euchromatin) 영역 구조, 게놈의 헤테로크로마틴(heterochromatin) 영역 구조, 게놈의 크로마틴 교차결합(chromatin cross-link) 영역 구조, 게놈의 단백질 결합 영역 구조, 게놈의 외유전체 영역 구조, 게놈의 일부 복사 수 변이 영역 등이 있을 수 있다. 설명의 편의를 위해 게놈 구조는 게놈의 유크로마틴(euchromatin) 영역 구조인 것으로 가정하고 본 발명에 대해 하기에서 설명한다.
- [0041] 세포 포집부(120)는 피검자의 검체(혈액, 소변 등)로부터 액체생검 디바이스나 키트 등을 통해 세포를 포집한다.
- [0043] 세포 분석부(130)는 검체로부터 포집된 세포의 게놈 구조의 상태 정보를 획득한다.
- [0044] 즉, 세포 분석부(130)는 포집된 세포의 게놈 해독 혹은 검형을 통해 게놈 상의 서열 패턴, 구조 등을 확인할 수 있다.
- [0045] 도 4는 본 발명에 따른 ATAC-seq를 이용한 유크로마틴 영역의 해독 예시를 설명하기 위한 그림이다.
- [0046] 도 4에 도시된 바와 같이, 세포 분석부(130)는 ATAC-seq 실험을 통해 포집된 세포의 게놈을 해독하여 게놈 상의 유크로마틴(open chromatin) 영역을 확인할 수 있다.
- [0048] 세포 구분부(140)는 저장부(110)에 미리 저장되어 있는 표준 게놈 구조 상태 패턴 DB를 기반으로, 게놈 구조의 상태 변형 영역을 분석하여 포집된 세포를 질환 세포와 정상 세포로 구분한다.
- [0049] 도 5는 본 발명에 따른 게놈 구조의 상태 변형 영역의 서열 수와 피크(peak)를 설명하기 위한 그림이고, 도 6은 본 발명에 따른 표준 백혈구 게놈 구조와 질환 세포의 게놈 구조의 대비 예시를 설명하기 위한 그림이다.
- [0050] 즉, 세포 구분부(140)는 도 5에 도시된 바와 같은 게놈 구조의 상태 변형 영역의 서열 수와 피크(peak)를 기반으로, 표준 게놈 구조 패턴 DB에 저장된 게놈 구조와 포집된 세포의 게놈 구조를 대비하여, 포집된 세포를 질환 세포와 정상 세포로 구분할 수 있다.
- [0051] 예컨대, 세포 구분부(140)는 포집된 세포의 특이적 게놈 구조 상태 변형 영역을 분석, 즉 표준 게놈 구조 상태 패턴 DB에 저장된 게놈 구조와 포집된 세포의 게놈 구조를 비교하여 백혈구에서 일반적으로 나타나는 게놈 구조 상태 영역을 제외하여, 질환 세포의 게놈 구조 상태 영역으로 예측되는 후보 영역을 확보할 수 있다. 도 6을 참조하면, Gapdh 유전자의 질환 세포와 정상 세포(백혈구)의 게놈 구조는 차이가 없지만, Grin1 유전자의 게놈 구조는 질환 세포가 정상 세포(백혈구)와 차이가 있음을 확인할 수 있다.
- [0053] 세포 기원 획득부(150)는 저장부(110)에 미리 저장되어 있는 조직별 게놈 구조 패턴 DB를 기반으로, 게놈 구조

변형 영역을 분석하여 세포 구분부(140)를 통해 질환 세포로 구분된 질환 세포의 조직 기원(tissue origin)을 획득한다.

- [0054] 도 7은 본 발명에 따른 조직/질환 특이적 게놈 구조의 상태 패턴 대비 예시를 설명하기 위한 그림이다.
- [0055] 즉, 세포 기원 획득부(150)는 게놈 구조의 상태 변형 영역의 피크(peak) 패턴을 기반으로, 조직별 게놈 구조 상태 패턴 DB에 저장된 게놈 구조와 포집된 세포의 게놈 구조 상태를 대비하여, 질환 세포의 조직 기원을 획득할 수 있다.
- [0056] 도 7을 참조하면, 질환 세포 1의 경우 위암의 게놈 구조 상태 변형 영역의 피크 패턴과 유사하므로 질환 세포 1의 조직 기원은 위암(즉, 위)임을 확인할 수 있고, 질환 세포 2의 경우 대장암의 게놈 구조 상태 변형 영역의 피크 패턴과 유사하므로 질환 세포 2의 조직 기원은 대장암(즉, 대장)임을 확인할 수 있다.
- [0057] 보다 자세하게 설명하면, 세포 기원 획득부(150)는 게놈 구조 상태 변형 영역의 피크 패턴을 이용한 유사도 판별을 아래에서 설명하는 세 개의 방법 중에서 하나 이상의 방법을 선택하여 단독으로 또는 조합하여 수행할 수 있다.
- [0058] 첫 번째로, 세포 기원 획득부(150)는 조직별 게놈 구조 상태 패턴 DB에 저장된 게놈 구조의 피크 위치와 포집된 세포의 게놈 구조의 피크 위치를 대비하여, 질환 세포의 조직 기원을 획득할 수 있다.
- [0059] 도 8은 본 발명에 따른 게놈 구조의 피크 위치를 이용한 게놈 구조 패턴 대비를 설명하기 위한 도면이다.
- [0060] 즉, 세포 기원 획득부(150)는 조직/질환 특이적 피크가 포함된 유전자 조절 영역으로 범위를 넓혀서 해당 유전자 영역과 유전자 조절 영역에 질환 세포의 피크가 존재하면 서로 일치하는 것으로 판별할 수 있다.
- [0061] 도 8을 참조하면, 질환 세포 1과 질환 세포 2는 위암 특이적 ABC 유전자의 피크가 포함된 유전자 영역과 유전자 조절 영역에 포함되어 있으므로 위암 특이적 ABC 유전자 피크와 일치하는 것으로 판별하고, 질환 세포 3은 위암 특이적 ABC 유전자의 피크가 포함된 유전자 영역과 유전자 조절 영역에 포함되어 있지 않으므로 위암 특이적 ABC 유전자 피크와 불일치하는 것을 판별한다.
- [0062] 두 번째로, 세포 기원 획득부(150)는 조직별 게놈 구조 상태 패턴 DB에 저장된 게놈 구조의 피크 영역과 포집된 세포의 게놈 구조의 피크 영역이 서로 오버랩(overlap)되는 정도를 이용하여, 질환 세포의 조직 기원을 획득할 수 있다. 예를 들어, 도 9와 같이 게놈 구조의 피크 영역이 서로 오버랩되는 비율에 기반하여 질환 세포의 조직 기원을 획득할 수 있다.
- [0063] 도 9는 본 발명의 예시적인 실시예에 따른 게놈 구조의 피크 영역 오버랩 비율을 이용한 게놈 구조 패턴 대비를 설명하기 위한 도면이다.
- [0064] 예시적인 실시예에서, 세포 기원 획득부(150)는 일반적인 범위 영역의 비교에 사용되는 "reciprocal > 50% overlap" 방법을 사용하여, 샘플 간의 교차되는 영역의 길이가 각 샘플의 피크 영역 길이의 50% 이상일 경우에 두 피크가 일치하는 것으로 판별할 수 있다.
- [0065] 도 9를 참조하면, 질환 세포 1은 위암 특이적 ABC 유전자의 피크 영역과 reciprocal 50% overlap을 충족하므로 위암 특이적 ABC 유전자 피크와 일치하는 것으로 판별하고, 질환 세포 2는 reciprocal 50% overlap을 충족시키지 못하므로 위암 특이적 ABC 유전자 피크와 일치하지 않는 것으로 판별한다.
- [0066] 세 번째로, 세포 기원 획득부(150)는 조직별 게놈 구조 상태 패턴 DB에 저장된 게놈 구조의 피크 스코어(peak score)를 기반으로 획득된 매트릭스(matrix)와 포집된 세포의 게놈 구조의 피크 스코어를 기반으로 획득된 매트릭스를 대비하여, 질환 세포의 조직 기원을 획득할 수 있다.
- [0067] 도 10은 본 발명에 따른 게놈 구조의 피크 스코어를 이용한 게놈 구조 패턴 대비를 설명하기 위한 도면이다.
- [0068] 즉, 세포 기원 획득부(150)는 모든 유전자 영역을 대상으로 피크 스코어의 기준값을 정한 후, 유전자에 해당하는 피크 스코어가 기준값보다 낮으면 Off, 높으면 On으로 매트릭스를 만든다. 그리고, 세포 기원 획득부(150)는 매트릭스를 기준으로 질환 세포의 On/Off값을 비교하여 질환 세포와 유사한 패턴의 조직/질환을 찾을 수 있다.
- [0069] 도 10을 참조하면, 질환 세포 1은 위암 특이적 유전자인 G7에 피크가 발견되었으므로 위암 조직과 일치하는 것으로 판별하고, 질환 세포 2는 조직/질환 특이적 유전자가 없기 때문에 유전자의 조합(A, B, C 모두 On)으로 폐암 조직과 일치하는 것으로 판별한다.

[0071] 세포 정량 획득부(160)는 저장부(110)에 저장되어 있는 표준 게놈 구조 상태 패턴 DB와 조직별 게놈 구조 상태 패턴 DB를 기반으로, 게놈 구조의 상태 변형 영역을 분석하여 질환 세포의 정량을 획득한다.

[0072] 즉, 세포 정량 획득부(160)는 표준 게놈 구조 상태 패턴 DB와 조직별 게놈 구조 상태 패턴 DB를 기반으로, 질환 세포의 특이적 게놈 구조의 상태 변형 영역을 대상으로 하여 획득된 서열 수와 정상 세포의 특이적 게놈 구조의 상태 변형 영역을 대상으로 하여 획득된 서열 수를 이용하여, 전체 세포의 수 대비 질환 세포의 수를 계산하여 질환 세포의 정량을 획득할 수 있다.

[0073] 보다 자세하게 설명하면, 세포 정량 획득부(160)는 표준 게놈 구조 상태 패턴 DB와 조직별 게놈 구조 상태 패턴 DB와 비교하여 정상 세포(백혈구)에는 존재하지 않는 질환 세포 특이적 게놈구조의 상태 변형 영역을 대상으로 해독된 서열의 수(Dr)를 아래의 [수학식 1]을 통해 계산한다.

[0074] [수학식 1]

$$Dr = \sum_{k=1}^n D_k$$

[0075]

$\text{질환세포 특이적 영역1의 정규화 서열 수}(D_1) = \frac{\text{질환세포 특이적 영역1에 정렬된 서열 수} \times \text{서열 길이} \times 10^6}{\text{질환세포 특이적 영역 총 개수} \times \text{질환세포 특이적 영역1의 길이}}$

[0076]

[0077] 여기서, n은 질환세포 특이적 영역의 총 개수를 나타낸다.

[0078] 세포 정량 획득부(160)는 표준 게놈 구조 상태 패턴 DB와 조직별 게놈 구조 상태 패턴 DB와 비교하여 질환 세포에는 존재하지 않는 정상 세포 특이적 게놈구조의 상태 변형 영역을 대상으로 해독된 서열의 수(Cr)를 아래의 [수학식 2]를 통해 계산한다.

[0079] [수학식 2]

$$Cr = \sum_{k=1}^m C_k$$

[0080]

$\text{정상세포 특이적 영역1의 정규화 서열 수}(C_1) = \frac{\text{정상세포 특이적 영역1에 정렬된 서열 수} \times \text{서열 길이} \times 10^6}{\text{정상세포 특이적 영역 총 개수} \times \text{정상세포 특이적 영역1의 길이}}$

[0081]

[0082] 여기서, m은 정상세포 특이적 영역의 총 개수를 나타낸다.

[0083] 이때, 질환 세포 특이적 영역/정상 세포 특이적 영역을 프로파일(profile)은 아래의 과정을 거쳐 수행된다.

[0084] - 먼저 ATAC-seq를 통해 생산된 서열을 human reference genome에 정렬(alignment)한다.

[0085] - 정렬 후에 생성된 BAM(binary alignment map) 파일의 필터링 과정을 거친다(BAM 파일 프로세싱).

[0086] - Peak calling 프로그램을 이용하여 각 샘플의 피크 영역을 프로파일한다.

[0087] - 표준 게놈 구조 상태 패턴 DB와 조직별 게놈 구조 상태 패턴 DB와 비교하여, 질환 세포 특이적 영역/정상 세포 특이적 영역을 찾는다.

[0088] - 질환 세포 특이적 영역/정상 세포 특이적 영역에 해당하는 서열의 수를 계산한다.

[0089] 세포 정량 획득부(160)는 질환 세포 특이적 게놈구조 상태 변형 영역을 대상으로 해독된 서열의 수(Dr)와 정상 세포 특이적 게놈구조 상태 변형 영역을 대상으로 해독된 서열의 수(Cr)를 토대로 아래의 [수학식 4]를 통해 전체 세포의 수 대비 질환 세포의 수(농도)를 계산하여 질환 세포의 정량을 획득할 수 있다.

[0090] 즉, 전체 세포의 수 중 질환 세포의 수의 비율은 아래의 [수학식 3]으로 표현할 수 있다.

[0091] [수학식 3]

[0092]

[0093] 따라서, 질환 세포의 수는 아래의 [수학식 4]를 통해 계산할 수 있다.

[0094] [수학식 4]

[0095]

[0098] 그러면, 도 11을 참조하여 본 발명의 바람직한 실시예에 따른 염색체 구조의 상태 변형 정보를 이용한 세포 분석 방법에 대하여 설명한다

[0099] 도 11은 본 발명의 바람직한 실시예에 따른 염색체 구조의 상태 변형 정보를 이용한 세포 분석 방법을 설명하기 위한 흐름도이다.

[0100] 도 11을 참조하면, 세포 분석 장치(100)는 피검자의 검체로부터 세포를 포집한다(S110).

[0101] 그리고, 세포 분석 장치(100)는 포집된 세포의 계놈 구조의 상태를 획득한다(S120). 즉, 세포 분석 장치(100)는 포집된 세포의 계놈 해독 혹은 검형을 통해 계놈 상의 서열 패턴, 구조 등을 확인할 수 있다.

[0102] 그런 다음, 세포 분석 장치(100)는 표준 계놈 구조 상태 패턴 DB를 기반으로 계놈 구조의 상태 변형 영역을 분석하여 포집된 세포를 질환 세포와 정상 세포로 구분한다(S130). 즉, 세포 분석 장치(100)는 계놈 구조의 상태 변형 영역의 서열 수와 피크(peak)를 기반으로, 표준 계놈 구조 상태 패턴 DB에 저장된 계놈 구조와 포집된 세포의 계놈 구조를 대비하여, 포집된 세포를 질환 세포와 정상 세포로 구분할 수 있다.

[0103] 그리고, 세포 분석 장치(100)는 조직별 계놈 구조 상태 패턴 DB를 기반으로 계놈 구조의 상태 변형 영역을 분석하여 질환 세포의 조직 기원을 획득한다(S140). 즉, 세포 분석 장치(100)는 계놈 구조의 상태 변형 영역의 피크(peak) 패턴을 기반으로, 조직별 계놈 구조 상태 패턴 DB에 저장된 계놈 구조와 포집된 세포의 계놈 구조를 대비하여, 질환 세포의 조직 기원을 획득할 수 있다.

[0104] 보다 자세하게 설명하면, 세포 분석 장치(100)는 계놈 구조의 상태 변형 영역의 피크 패턴을 이용한 유사도 판별을 아래에서 설명하는 세개의 방법 중에서 하나를 선택하여 수행할 수 있다.

[0105] 첫 번째로, 세포 분석 장치(100)는 조직별 계놈 구조 상태 패턴 DB에 저장된 계놈 구조의 피크 위치와 포집된 세포의 계놈 구조의 피크 위치를 대비하여, 질환 세포의 조직 기원을 획득할 수 있다.

[0106] 두 번째로, 세포 분석 장치(100)는 조직별 계놈 구조 상태 패턴 DB에 저장된 계놈 구조의 피크 영역과 포집된 세포의 계놈 구조의 피크 영역이 서로 오버랩(overlap)되는 비율을 기반으로, 질환 세포의 조직 기원을 획득할 수 있다.

[0107] 세 번째로, 세포 분석 장치(100)는 조직별 계놈 구조 상태 패턴 DB에 저장된 계놈 구조의 피크 스코어(peak score)를 기반으로 획득된 매트릭스(matrix)와 포집된 세포의 계놈 구조의 피크 스코어를 기반으로 획득된 매트릭스를 대비하여, 질환 세포의 조직 기원을 획득할 수 있다.

[0108] 이후, 세포 분석 장치(100)는 표준 계놈 구조 상태 패턴 DB와 조직별 계놈 구조 상태 패턴 DB를 기반으로 계놈 구조의 상태 변형 영역을 분석하여 질환 세포의 정량을 획득한다(S150). 즉, 세포 분석 장치(100)는 표준 계놈 구조 상태 패턴 DB와 조직별 계놈 구조 상태 패턴 DB를 기반으로, 질환 세포의 특이적 계놈 구조의 상태 변형 영역을 대상으로 하여 획득된 서열 수와 정상 세포의 특이적 계놈 구조의 상태 변형 영역을 대상으로 하여 획득된 서열 수를 이용하여, 전체 세포의 수 대비 질환 세포의 수를 계산하여 질환 세포의 정량을 획득할 수 있다.

[0110] 이하에서는, 포집 세포의 계놈 구조의 상태 변형 영역을 분석함으로써 질환 세포 양이 적은 경우에도 쉽게 포집 세포가 질환 세포인지 정상 세포인지 구분할 수 있음을 실험 데이터를 통해 다시 한번 설명한다. 도 12는 실험 데이터를 얻기 위해 실시한 실험의 단계를 도식화한 흐름도이다.

[0111] 실험에서는 혈액에서 CTC 등의 암세포를 분리할 수 있는 장치(10) (예를 들어 클리노믹스사의 CD-CTC Duo discTM)를 사용하여 실험군 샘플로부터 질환 세포를 분리하였다. 대조군으로는 일반인의 전혈(whole blood)에서 얻

은 PBMC(peripheral blood mononuclear cell)를 정상 세포 샘플(negative control sample)로 준비하였다. 그리고 질환 세포 샘플(positive control sample)로는 암세포주(cancer cell line)를 준비하였다. 암세포주로는 난소암_SK-OV-3(ovarian cancer_SK-OV-3)을 사용하였으며, 일반인의 전혈에서 얻은 PBMC에 암세포주를 개수별로 스파이크(spike)하여 실험군 샘플을 제작하였다. 샘플별로 스파이크된 암세포주의 개수는 표 1과 같다.

표 1

실험군 샘플	암세포주 개수
샘플 1	1개
샘플 2	10개
샘플 3	100개

[0112]

먼저, 정상 세포 샘플(11), 질환 세포 샘플(12), 그리고 실험군 샘플 1 내지 3을 각각 장치(10)에 넣어 샘플별로 세포를 분리한다. 그리고 각 샘플에서 분리된 세포가 담긴 멤브레인(13)을 장치(10)에서 꺼내어 정제(lysis)시킨 후 ATAC-seq 라이브러리(library)를 제작하고, 그 시퀀싱 데이터를 분석하여 샘플들 간에 차이 나는 영역을 선정한다.

[0113]

도 13은 질환 세포 샘플(12)에서 분리된 세포에 대해서 시퀀싱 데이터를 분석한 결과이고, 도 14는 정상 세포 샘플(11)에서 분리된 세포에 대해서 시퀀싱 데이터를 분석한 결과이다. 도 13의 데이터에서는 검출되나 도 14의 데이터에서는 검출되지 않는 서열들을 선정함으로써, 질환 세포 샘플과 정상 세포 샘플 간에 차이 나는 영역을 알 수 있다. 즉, 도 13의 데이터에서는 검출되나 도 14의 데이터에서는 검출되지 않는 서열들이 있는 영역은, 정상 세포 샘플에서는 전사 발현이 일어나지 않은 게놈 영역이고 질환 세포 샘플에서는 전사 발현이 일어난 유크로마틴 영역임을 알 수 있다.

[0114]

이렇게 도 13과 도 14를 비교하여 정상 세포 샘플에서는 검출되지 않지만 질환 세포 샘플에서는 검출되는 서열들을 선정하여 게놈 구조의 상태 변형 영역을 선정할 수 있으며, 질환 세포의 해당 영역에서 피크(도 5를 참조하여 전술한 피크)를 찾을 수 있다. 그리고, 샘플 1 내지 샘플 3에서 해당 피크가 검출되는지를 분석함으로써, 샘플 1 내지 3을 질환 세포와 정상 세포로 구분할 수 있다.

[0115]

도 15는 샘플 1에서 분리된 세포에 대해서 시퀀싱 데이터를 분석한 결과이고, 도 16은 샘플 3에서 분리된 세포에 대해서 시퀀싱 데이터를 분석한 결과이다. 도 15를 참조하면, 정상 세포 샘플(11)과 비교하여 질환 세포 샘플(12)의 시퀀싱 데이터에서만 검출되는 유크로마틴 영역이, PBMC에 난소암_SK-OV-3를 1개(1ea) 스파이크한 샘플 1에서도 잘 검출되는 것을 알 수 있다. 따라서 본 발명에 따르면 실제로 포집된 세포에 소량의 질환 세포가 존재하는 경우에도 게놈 구조의 상태 변형 영역 분석을 통해 그 포집된 세포를 질환세포로 명확하게 구분할 수 있다.

[0116]

또한, 도 15와 도 16을 비교하여 알 수 있는 바와 같이, PBMC에 난소암_SK-OV-3를 100개(100ea) 스파이크한 샘플 3에서는 샘플 1에 비해 해당 유크로마틴 영역에서의 서열 개수(depth)가 더 많이 증가하는 것을 확인할 수 있다.

[0117]

도 17은 정상 세포 샘플(11)에서 분리된 세포의 시퀀싱 데이터(그래프의 아래부분에 도시함)와 샘플 3으로부터 분리된 세포의 시퀀싱 데이터(그래프의 위부분에 도시함)에서 각각 검출한 피크의 크기를 동일 염색체의 위치에서 비교한 그래프이다. 도 17의 그래프에서 x축은 크로마틴 19(chr19)의 염색체의 위치(단위: 메가베이스, MB)를 나타내고 y축은 200개의 염기서열(Bin 200)을 단위로 피크 크기를 나타내었다.

[0118]

이렇게 정상 세포 샘플(11)과 샘플 1 내지 3을 비교하여 실험군 샘플에서만 나타나는 피크를 확인함으로써, 해당 샘플을 질환 세포로 판별할 수 있다. 예를 들어, 도 17에서 7 내지 8 MB에 해당하는 영역(21)에서 샘플 3에 대해서는 피크가 검출되지만 정상 세포 샘플(11)에 대해서는 피크가 검출되지 않으며, 55 내지 56 MB에 해당하는 영역(22)에서도 샘플 3에서만 피크가 검출되므로, 샘플 3을 질환 세포로 판별할 수 있다.

[0119]

이러한 방식으로 각 영역의 서열을 통계처리하여 유의미한 결과를 얻을 수 있다. 도 18 내지 도 20은 전술한 방법으로 샘플 1 내지 3을 각각 시퀀싱하여 나오는 서열을 컴퓨터 프로그램(예를 들어, GEN RICH™)으로 분석한 결과 데이터이다. 도 18 내지 도 20은 각각 샘플 1 내지 샘플 3에 대한 결과 데이터이다. 도 18 내지 도 19의 데이터 테이블에서 첫 번째 컬럼은 크로마틴 번호이고, 네 번째 컬럼(빨간 색 박스로 표시됨)은 유전자 심볼(gene symbol)이다.

[0120]

[0121] 도 18 내지 도 20의 데이터 테이블을 비교하면, 정상 세포 샘플(11)에서와 구별되는 영역의 유전자 심볼이 각 샘플에서 유사한 패턴으로 검출되는 것을 알 수 있다. 도 18을 참조하면, 각 샘플에 대한 데이터는 i) 정상 세포 샘플(11)과 비교하여 질환 세포 샘플(12)에서만 검출되는 서열 영역에 대한 데이터(31)와, ii) 정상 세포 샘플(11)과 비교하여 질환 세포 샘플(12)에서만 검출되는 서열 중에서도 질환 세포의 양에 따라 다르게 검출되는 서열들의 데이터(32)로 구분할 수 있다.

[0122] 즉, 도 18에서 초록색 박스로 표시된 데이터(31) 중에서, 세 번째 컬럼의 피크 데이터와 네 번째 컬럼의 서열수(depth) 데이터의 수치는 도 19와 도 20의 데이터 테이블에서도 동일하게 나타난다. 그리고, 도 18에서 노란색 박스로 표시된 데이터(32) 중에서, 세 번째 컬럼의 피크 데이터와 네 번째 컬럼의 서열수(depth) 데이터의 수치는, 질환 세포의 함유량이 점점 많아지는 도 19에서와 도 20에서 점점 더 크게 나타남을 알 수 있다. 이렇게 도 18 내지 도 20의 데이터를 통해, 질환 세포 샘플을 정상 세포 샘플에서 구별할 수 있으며, 구별되는 서열 영역에서 서열 개수와 리드(read)수를 확인할 수 있다.

[0124] 본 발명은 또한 컴퓨터로 읽을 수 있는 기록 매체에 컴퓨터가 읽을 수 있는 코드로서 구현하는 것이 가능하다. 컴퓨터로 읽을 수 있는 기록 매체는 컴퓨터에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록 장치를 포함한다. 컴퓨터가 읽을 수 있는 기록 매체의 예로는 롬(ROM), 램(RAM), 씨디-롬(CD-ROM), 자기 테이프, 플로피 디스크, 광 데이터 저장장치 등이 있다.

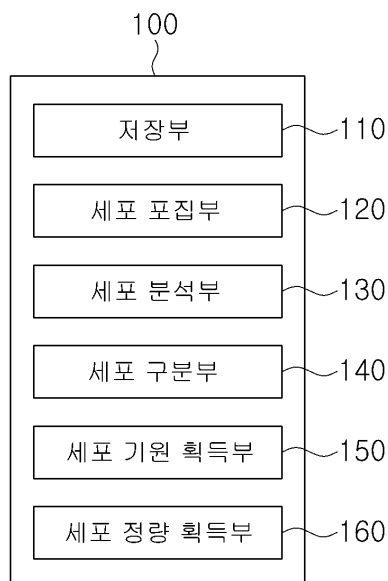
[0125] 이상에서 본 발명의 바람직한 실시예에 대하여 상세하게 설명하였지만 본 발명은 상술한 특징의 바람직한 실시예에 한정되지 아니하며, 다음의 청구범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술분야에서 통상의 지식을 가진 자라면 누구든지 다양한 변형 실시가 가능한 것은 물론이고, 그와 같은 변경은 청구범위 기재의 범위 내에 있게 된다.

부호의 설명

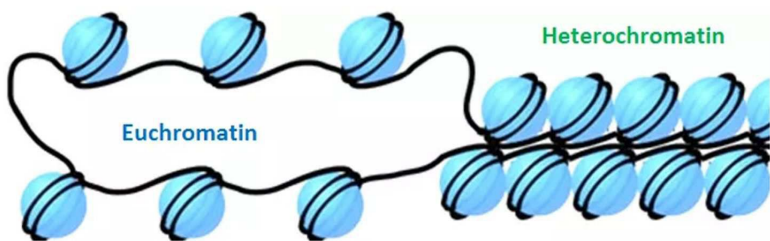
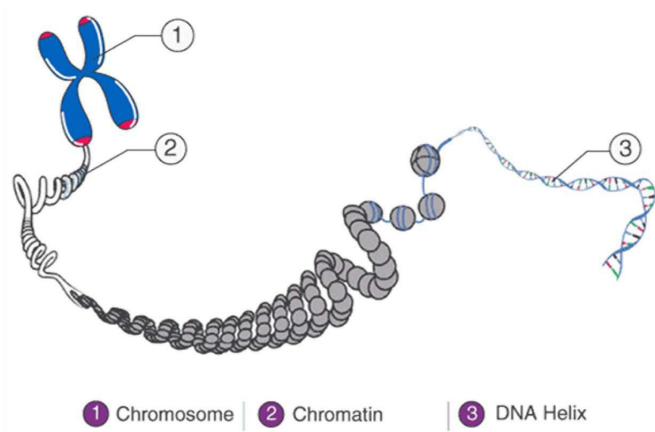
- [0126] 100 : 세포 분석 장치,
 110 : 저장부,
 120 : 세포 포집부,
 130 : 세포 분석부,
 140 : 세포 구분부,
 150 : 세포 기원 획득부,
 160 : 세포 정량 획득부

도면

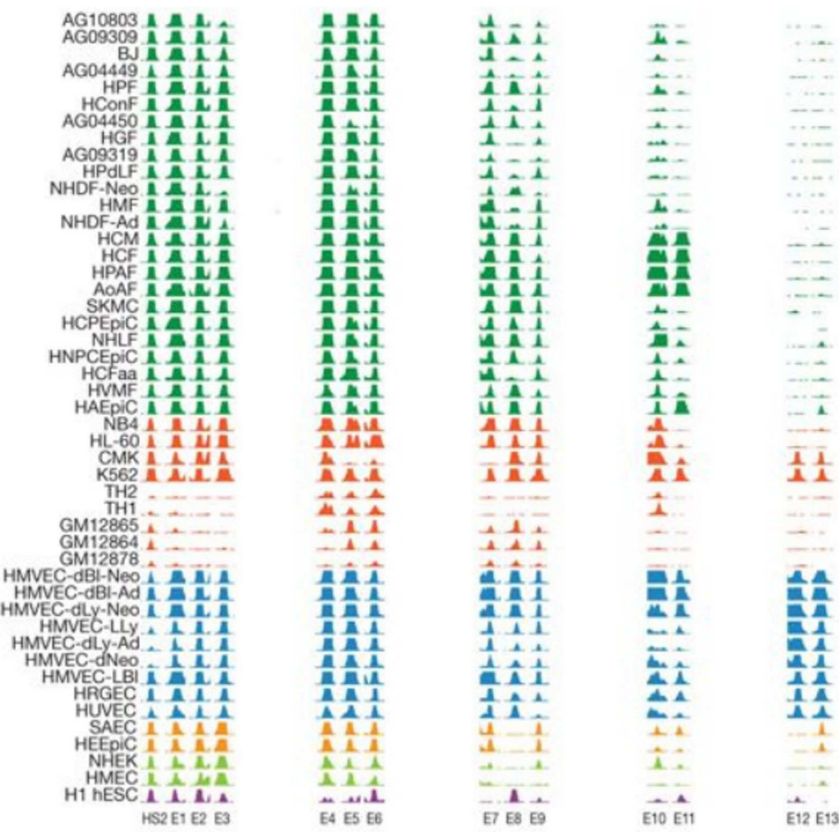
도면1



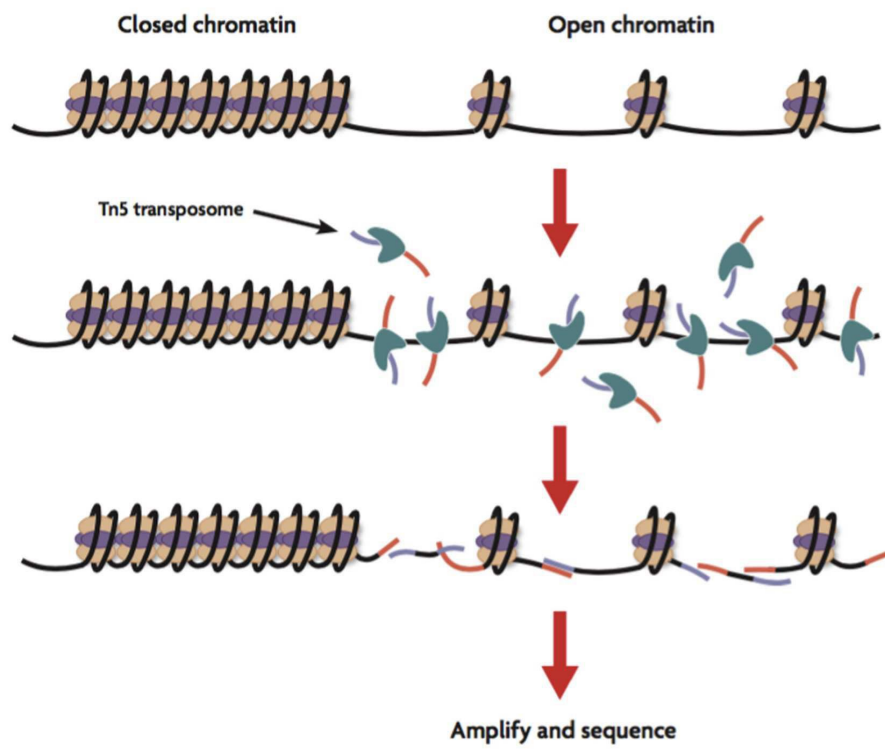
도면2



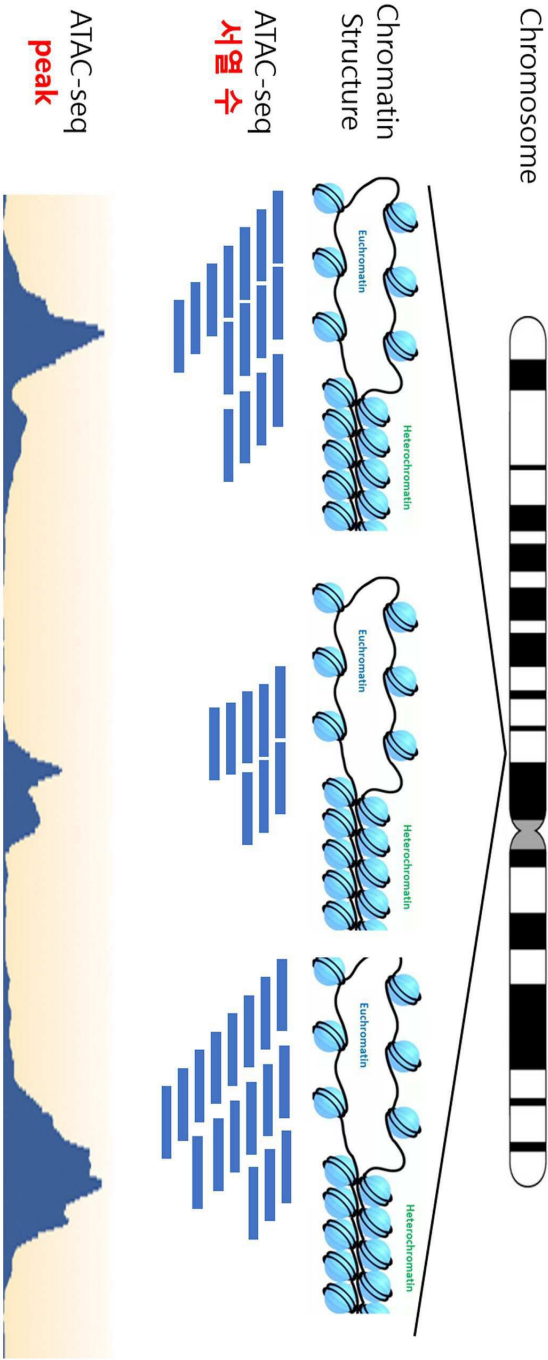
도면3



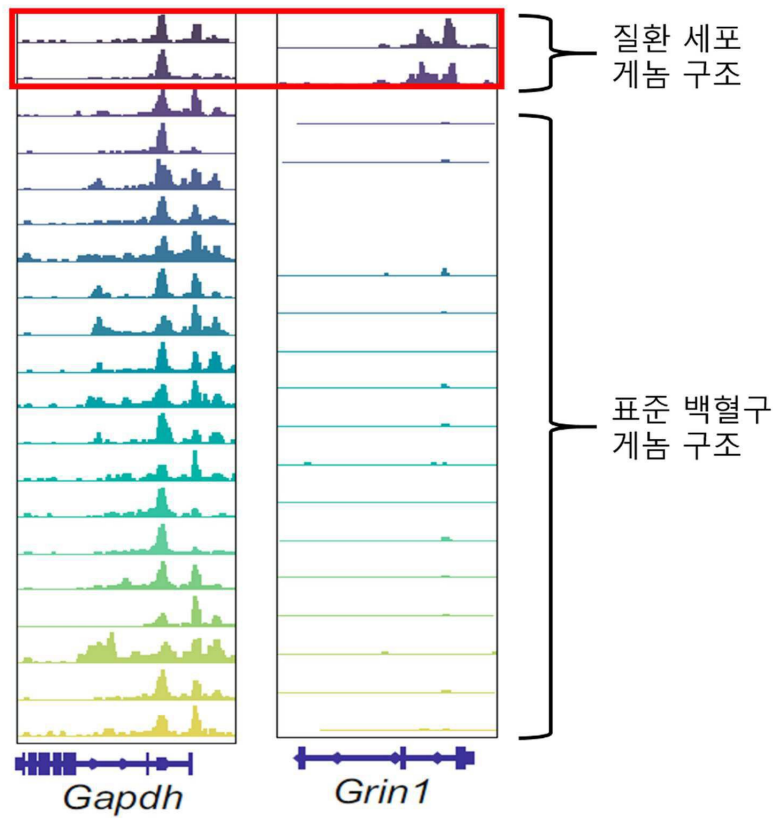
도면4



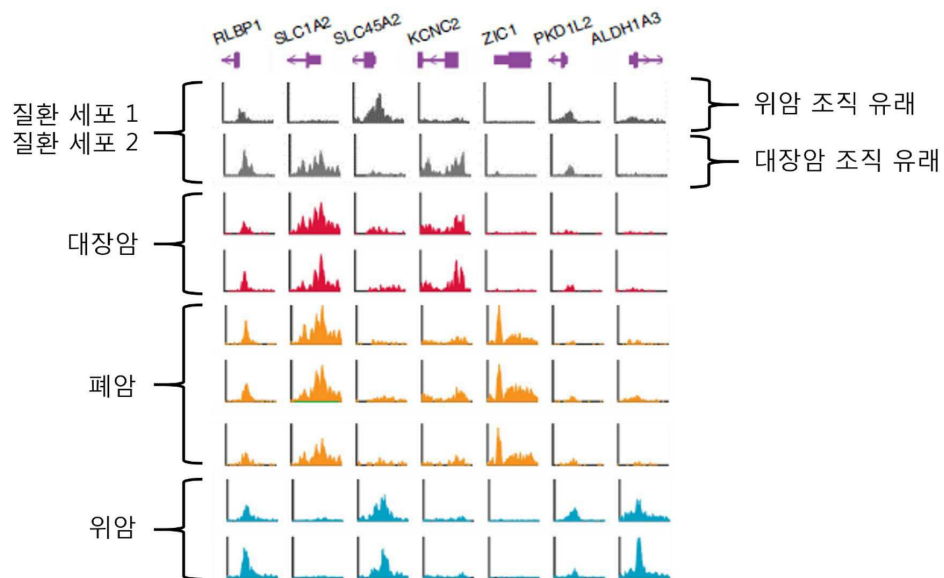
도면5



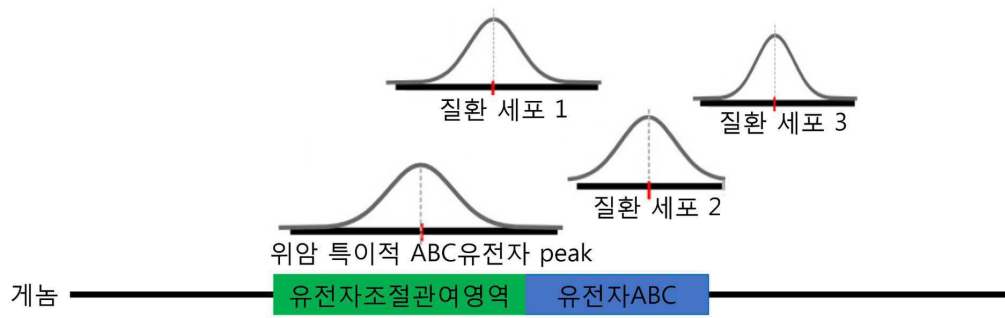
도면6



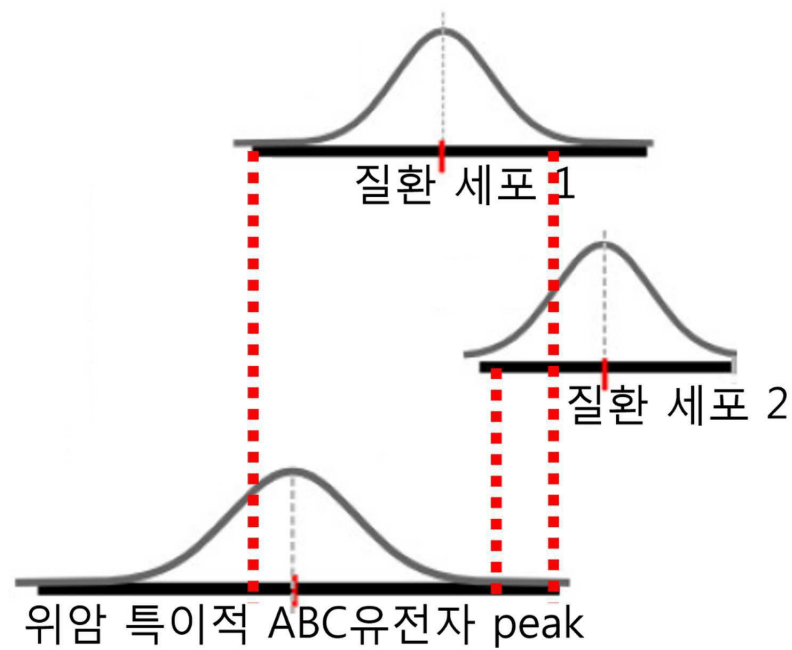
도면7



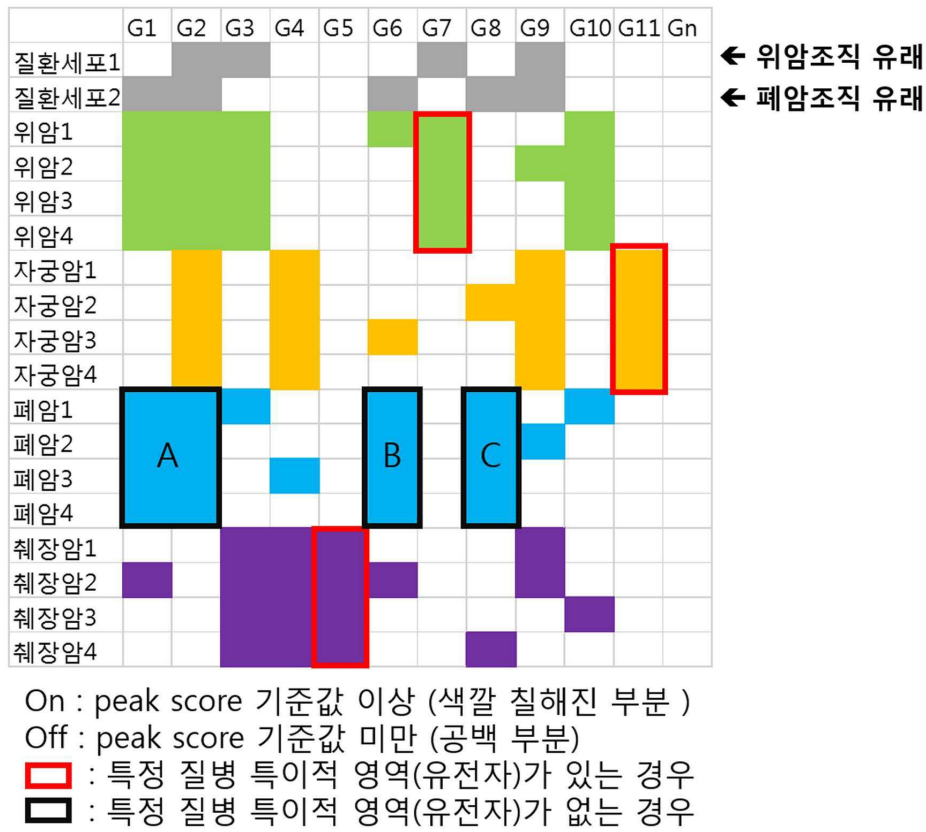
도면8



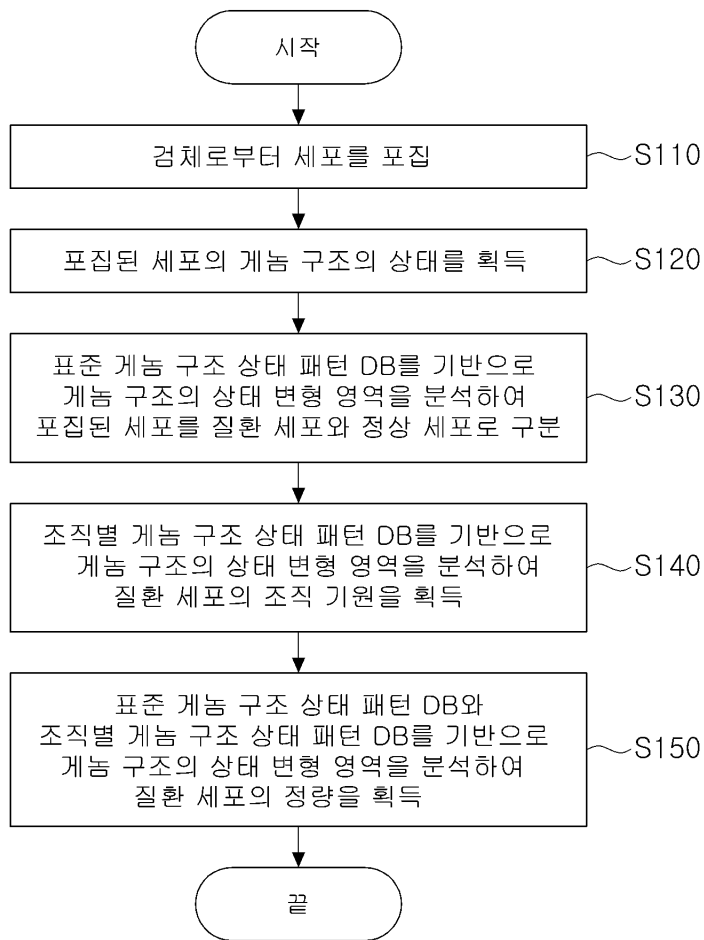
도면9



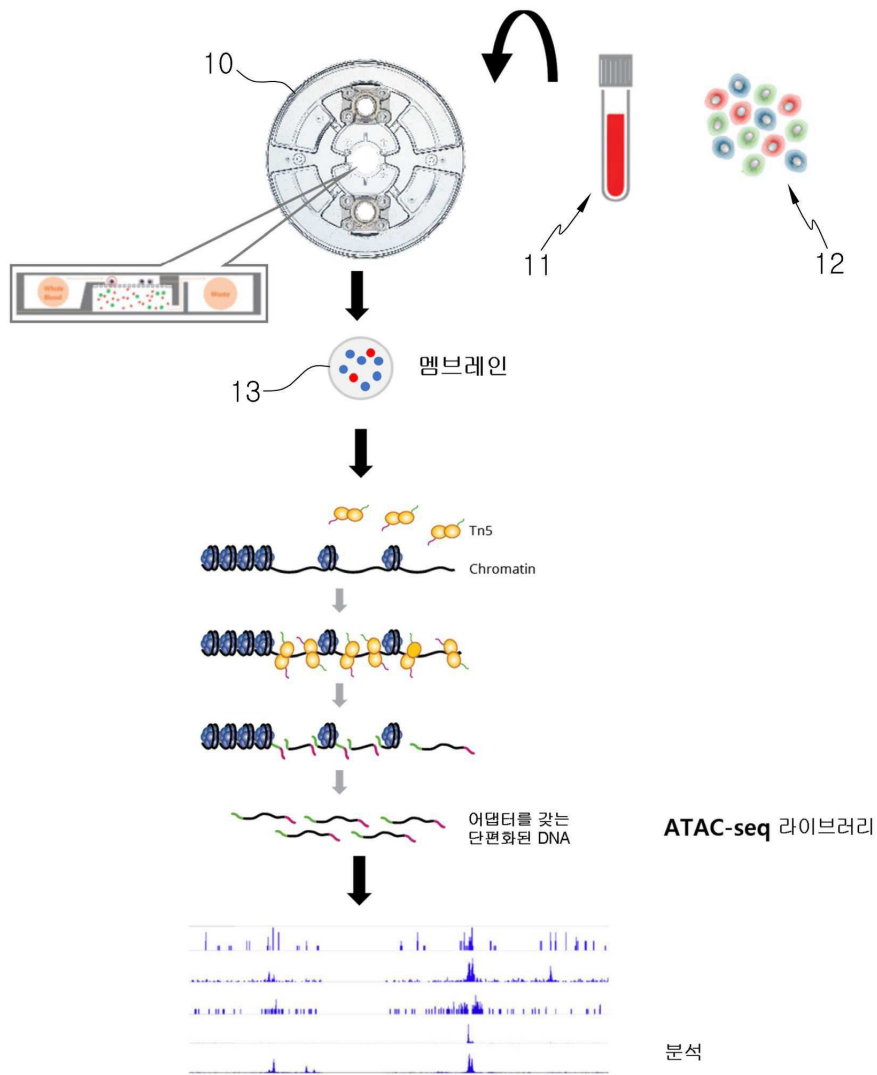
도면10



도면11



도면12



도면 13

Positive control_Ovarian cancer cell(SK-OV-3)

[illegible]

도면14

Negative control_PBMC

```
168999061 168999071 168999081 168999091 168999101 168999111 168999121 168999131 168999141 168999151
```

도면 15

PBMC + Ovarian cancer cell_1ea

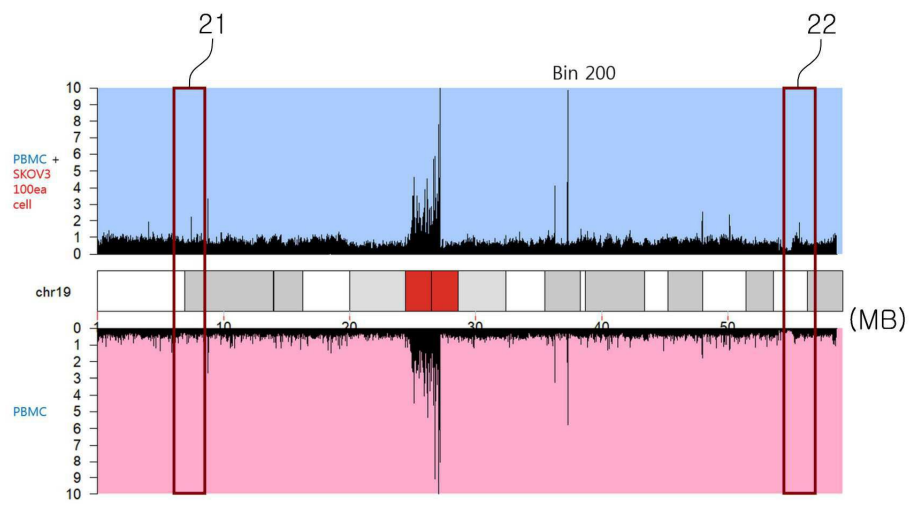
0051 1689990061 168999071 168999081 168999091 168999101 168999111 168999121 168999131 168999141 168999151
TGAATCTTAGTCTGAAACTCTCTGATTGCAAGACTATTATTTCTTTGAATCATAGCTCTTGTTGCCAGACCTAAATTCCTCACTTTTCCCCGAGAGTGTTCTGGCCCCA
TGAAT attgcaagactattatTTCTTTgaatcatagctcttggtgtccagaccttaaattctcacttttccccgagagtgttctggcccc
TGAATCTTAGTCTGAAA
tgaaattcttagctgaaactctctgattgcaagactattatTTCTTTgaatcatagctcttggtgtccagaccttaaattctcacttttccccgag

도면 16

PBMC + Ovarian cancer cell_100ea

[illegible]

도면17



도면18

chr	name	chr	start	end	match_start	match_end	chr	start	end	SC_start	SC_end	SC_peak	SC_depth	SC_mq	T_start	T_end	T_peak	T_depth	T_avg_depth	Overlap_start	Overlap_end
chr1	MM_00190702.2 -	MTNRN2L8	10,497,893	10,517,893	10,507,852	10,508,001	chr1	10,507,765	10,509,015	1,167	18,069	14	10,507,852	10,508,001	67	9,863	66	10,507,852	10,508,001		
chr1	MM_016422.4 -	RNF141	10,501,672	10,517,893	10,507,852	10,508,001	chr1	10,507,765	10,509,015	1,167	18,069	14	10,507,852	10,508,001	67	9,863	66	10,507,852	10,508,001		
chr1	MM_016422.4 -	MTNRN2L8	10,497,893	10,517,893	10,508,848	10,509,012	chr1	10,507,765	10,509,015	1,167	18,069	14	10,508,848	10,509,012	96	13,581	82	10,508,848	10,509,012		
chr1	MM_016422.4 -	RNF141	10,501,672	10,517,893	10,508,848	10,509,012	chr1	10,507,765	10,509,015	1,167	18,069	14	10,508,848	10,509,012	96	13,581	82	10,508,848	10,509,012		
chr1	MM_00190702.2 -	MTNRN2L8	10,497,893	10,517,893	10,509,166	10,509,779	chr1	10,509,162	10,510,285	486	113,264	103	10,509,166	10,509,779	482	156,191	254	10,509,166	10,509,779		
chr1	MM_016422.4 -	RNF141	10,501,672	10,517,893	10,509,166	10,509,779	chr1	10,509,162	10,510,285	486	113,264	103	10,509,166	10,509,779	482	156,191	254	10,509,166	10,509,779		
chr1	MM_00190702.2 -	MTNRN2L8	10,497,893	10,517,893	10,509,891	10,510,041	chr1	10,509,162	10,510,285	486	113,264	103	10,509,891	10,510,041	96	11,007	73	10,509,891	10,510,041		
chr1	MM_016422.4 -	RNF141	10,501,672	10,517,893	10,509,891	10,510,041	chr1	10,509,162	10,510,285	486	113,264	103	10,509,891	10,510,041	96	11,007	73	10,509,891	10,510,041		
chr1	MM_00190702.2 -	MTNRN2L8	10,497,893	10,517,893	10,510,296	10,511,589	chr1	10,510,296	10,511,589	106	332,236	810	10,510,296	10,511,589	92	141,974	483	10,510,296	10,511,589		
chr1	MM_016422.4 -	RNF141	10,501,672	10,517,893	10,510,296	10,511,589	chr1	10,510,296	10,511,589	106	332,236	810	10,510,296	10,511,589	92	141,974	483	10,510,296	10,511,589		
chr1	MM_00190702.2 -	MTNRN2L8	10,497,893	10,517,893	10,511,589	10,512,768	chr1	10,511,589	10,512,768	112	73,442	346	10,511,589	10,512,768	112	73,442	346	10,511,589	10,512,768		
chr1	MM_016422.4 -	RNF141	10,501,672	10,517,893	10,511,589	10,512,768	chr1	10,511,589	10,512,768	112	73,442	346	10,511,589	10,512,768	112	73,442	346	10,511,589	10,512,768		
chr2	MM_003522.3 -	KIF5C	148,865,226	148,885,226	148,881,689	148,881,895	chr2	148,881,689	148,881,923	117	19,747	84	148,881,689	148,881,895	420	86,903	106	148,881,689	148,881,895		
chr3	MM_00132759.3 -	FRG2	75,654,190	75,674,190	75,668,845	75,669,661	chr3	75,668,419	75,670,912	834	98,636	40	75,668,845	75,669,661	53	59,829	80	75,668,845	75,669,661		
chr4	MM_00128820.1 -	MTNRN2L2	80,639,999	80,659,999	80,650,054	80,650,203	chr4	80,650,008	80,650,338	112	5,047	15	80,650,054	80,650,203	64	7,901	53	80,650,054	80,650,203		
chr5	MM_00190470.2 -	MTNRN2L2	80,644,651	80,664,651	80,650,054	80,650,203	chr5	80,650,008	80,650,338	112	5,047	15	80,650,054	80,650,203	64	7,901	53	80,650,054	80,650,203		
chr5	MM_00190470.2 -	MTNRN2L2	80,639,999	80,659,999	80,650,471	80,650,599	chr5	80,650,455	80,651,934	1,053	98,681	67	80,650,471	80,650,599	82	4,007	31	80,650,471	80,650,599		
chr5	MM_003439.5 -	MTNRN2L2	80,644,651	80,664,651	80,650,471	80,650,599	chr5	80,650,455	80,651,934	1,053	98,681	67	80,650,471	80,650,599	82	4,007	31	80,650,471	80,650,599		
chr5	MM_00190470.2 -	MTNRN2L2	80,639,999	80,659,999	80,650,768	80,651,716	chr5	80,650,455	80,651,934	1,053	98,681	67	80,650,768	80,651,716	729	126,136	133	80,650,768	80,651,716		
chr5	MM_003439.5 -	MTNRN2L2	80,644,651	80,664,651	80,650,768	80,651,716	chr5	80,650,455	80,651,934	1,053	98,681	67	80,650,768	80,651,716	729	126,136	133	80,650,768	80,651,716		
chrX	MM_00190708.2 -	MTNRN2L10	53,171,391	53,191,391	53,181,609	53,181,746	chrX	53,181,453	53,181,732	245	3,282	10	53,181,609	53,181,746	57	4,917	36	53,181,609	53,181,746		

PBMC + SKOV3 1ea cell

31

32

도면19

PBMC + SKOV3 10ea cell

chr	name	start	end	symbol	start	end	match_start	match_end	chr	SC_start	SC_end	SC_peak	SC_depth	SC_avg_dep	T_start	T_end	T_pos	T_dep	T_avg_dep	Overlap_start	Overlap_end
chr11	NM_001190702.2	-	MTRNR1.8	10,497,893	10,517,893	10,507,854	10,508,000	chr11	10,507,765	10,509,015	1,167	18,069	14	10,507,854	10,508,000	71	9,299	63	10,507,854	10,508,000	
chr11	NM_016422.4	-	RNF-141	10,501,672	10,521,672	10,507,852	10,509,009	chr11	10,507,765	10,509,015	1,167	18,069	14	10,507,854	10,508,000	83	11,959	76	10,507,854	10,508,000	
chr11	NM_001190702.2	-	MTRNR1.8	10,497,893	10,517,893	10,508,852	10,509,009	chr11	10,507,765	10,509,015	1,167	18,069	14	10,508,852	10,509,009	83	11,959	76	10,508,852	10,509,009	
chr11	NM_016422.4	-	RNF-141	10,501,672	10,521,672	10,509,169	10,509,771	chr11	10,507,765	10,510,285	1,167	18,069	103	10,508,852	10,509,771	479	152,693	253	10,508,852	10,509,771	
chr11	NM_001190702.2	-	MTRNR1.8	10,497,893	10,517,893	10,509,169	10,509,771	chr11	10,509,162	10,510,285	486	113,264	103	10,509,169	10,509,771	479	152,693	253	10,509,169	10,509,771	
chr11	NM_016422.4	-	RNF-141	10,501,672	10,521,672	10,509,896	10,510,036	chr11	10,509,162	10,510,285	486	113,264	103	10,509,896	10,510,036	64	9,304	66	10,509,896	10,510,036	
chr11	NM_001190702.2	-	MTRNR1.8	10,497,893	10,517,893	10,509,896	10,510,036	chr11	10,509,162	10,510,285	486	113,264	103	10,509,896	10,510,036	64	9,304	66	10,509,896	10,510,036	
chr11	NM_016422.4	-	RNF-141	10,501,672	10,521,672	10,509,896	10,510,036	chr11	10,509,162	10,510,285	486	113,264	103	10,509,896	10,510,036	64	9,304	66	10,509,896	10,510,036	
chr17	NM_001190452.1	+	MTRNR1.1	22,513,110	22,533,110	22,521,296	22,521,584	chr17	22,521,244	22,521,642	106	323,236	810	22,521,296	22,521,584	92	166,572	576	22,521,296	22,521,584	
chr2	NM_004522.3	+	KIF5C	148,865,226	148,885,226	148,881,689	148,881,895	chr3	148,881,689	148,881,923	117	19,747	84	148,881,684	148,881,895	111	76,045	359	148,881,689	148,881,895	
chr2	NM_001124759.3	+	KIF5C	75,654,190	75,674,190	75,668,879	75,669,643	chr3	75,668,419	75,670,912	834	98,636	40	75,668,879	75,668,879	374	91,802	120	75,668,879	75,668,879	
chr4	NM_001124759.3	+	KIF5C	190,014,366	190,034,366	190,022,062	190,022,660	chr4	190,021,366	190,022,819	1,016	67,981	47	190,022,062	190,022,660	311	56,093	94	190,022,062	190,022,660	
chr5	NM_001190470.2	-	MTRNR1.2	80,639,999	80,659,999	80,650,063	80,650,192	chr5	80,650,008	80,650,338	112	5,047	15	80,650,063	80,650,192	53	6,415	49	80,650,063	80,650,192	
chr5	NM_002439.5	+	MSH3	80,644,651	80,664,651	80,650,833	80,651,711	chr5	80,650,455	80,651,954	1,053	98,681	67	80,650,833	80,651,711	674	127,454	145	80,650,833	80,651,711	
chr5	NM_001190470.2	-	MTRNR1.2	80,639,999	80,659,999	80,650,833	80,651,711	chr5	80,650,455	80,651,954	1,053	98,681	67	80,650,833	80,651,711	674	127,454	145	80,650,833	80,651,711	
chr5	NM_002439.5	+	MSH3	80,644,651	80,664,651	80,650,833	80,651,711	chr5	80,650,455	80,651,954	1,053	98,681	67	80,650,833	80,651,711	674	127,454	145	80,650,833	80,651,711	

도면20

chr	name	start	end	symbol	start	end	match_start	match_end	chr	SC_start	SC_end	SC_peak	SC_depth	SC_avg_dep	T_start	T_end	T_peak	T_depth	T_avg_dep	Overlap_start	Overlap_end
chr1	NM_001190702.2	-	-	MTRNR2L8	10,497,893	10,517,893	10,507,854	10,508,005	chr1	10,507,765	10,509,015	1,167	18,069	14	10,507,854	10,508,005	77	8,157	54	10,507,854	10,508,005
chr1	NM_016422.4	-	-	RNF-41	10,501,672	10,521,672	10,507,854	10,509,009	chr1	10,507,765	10,509,015	1,167	18,069	14	10,507,854	10,508,005	87	8,157	54	10,507,854	10,508,005
chr1	NM_001190702.2	-	-	MTRNR2L8	10,497,893	10,517,893	10,508,852	10,509,009	chr1	10,507,765	10,509,015	1,167	18,069	14	10,508,852	10,509,009	87	11,188	71	10,508,852	10,509,009
chr1	NM_016422.4	-	-	RNF-41	10,501,672	10,521,672	10,508,852	10,509,009	chr1	10,507,765	10,509,015	1,167	18,069	14	10,508,852	10,509,009	87	11,188	71	10,508,852	10,509,009
chr1	NM_001190702.2	-	-	MTRNR2L8	10,497,893	10,517,893	10,509,168	10,509,771	chr1	10,509,162	10,510,265	486	113,264	103	10,509,168	10,509,771	480	129,726	215	10,509,168	10,509,771
chr1	NM_016422.4	-	-	RNF-41	10,501,672	10,521,672	10,509,168	10,509,771	chr1	10,509,162	10,510,265	486	113,264	103	10,509,168	10,509,771	480	129,726	215	10,509,168	10,509,771
chr1	NM_001190702.2	-	-	MTRNR2L8	10,497,893	10,517,893	10,509,894	10,510,037	chr1	10,509,162	10,510,265	486	113,264	103	10,509,894	10,510,037	94	8,596	60	10,509,894	10,510,037
chr1	NM_016422.4	-	-	RNF-41	10,501,672	10,521,672	10,509,894	10,510,037	chr1	10,509,162	10,510,265	486	113,264	103	10,509,894	10,510,037	94	8,596	60	10,509,894	10,510,037
chr1	NM_001190452.1	+	+	MTRNR2L1	22,513,110	22,533,110	22,521,296	22,521,601	chr17	22,521,244	22,521,642	106	323,236	810	22,521,296	22,521,601	176	142,645	466	22,521,296	22,521,601
chr2	NM_004522.3	+	+	KIF5C	148,865,226	148,885,226	148,881,689	148,881,689	chr2	148,881,689	148,881,923	117	19,747	84	148,881,684	148,881,895	102	66,499	314	148,881,689	148,881,895
chr3	NM_001124759.3	+	+	FRG2C	190,014,366	190,034,366	190,022,019	190,022,650	chr4	190,021,366	190,022,819	834	98,636	40	190,022,019	190,022,650	363	83,741	108	190,022,019	190,022,650
chr4	NM_0011286820.1	-	-	FRG2	80,639,999	80,659,999	80,650,062	80,650,201	chr5	80,650,008	80,650,338	112	67,981	15	80,650,062	80,650,201	57	6,235	45	80,650,062	80,650,201
chr5	NM_002439.5	+	+	M5H3	80,644,651	80,664,651	80,650,062	80,650,201	chr5	80,650,008	80,650,338	112	5,047	15	80,650,062	80,650,201	57	6,235	45	80,650,062	80,650,201
chr5	NM_001190470.2	-	-	M5H3	80,639,999	80,659,999	80,650,833	80,651,713	chr5	80,650,455	80,651,934	1,053	98,681	67	80,650,833	80,651,713	674	108,340	123	80,650,833	80,651,713
chr5	NM_002439.5	+	+	M5H3	80,644,651	80,664,651	80,650,833	80,651,713	chr5	80,650,455	80,651,934	1,053	98,681	67	80,650,833	80,651,713	674	108,340	123	80,650,833	80,651,713

PBMC + SKOV3 100ea cell