



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0032127
(43) 공개일자 2020년03월25일

(51) 국제특허분류(Int. Cl.)
C12Q 1/6886 (2018.01) C12Q 1/6827 (2018.01)
(52) CPC특허분류
C12Q 1/6886 (2018.05)
C12Q 1/6827 (2018.05)
(21) 출원번호 10-2020-7004066
(22) 출원일자(국제) 2018년07월11일
심사청구일자 없음
(85) 번역문제출일자 2020년02월11일
(86) 국제출원번호 PCT/CA2018/000141
(87) 국제공개번호 WO 2019/010564
국제공개일자 2019년01월17일
(30) 우선권주장
62/531,527 2017년07월12일 미국(US)

(71) 출원인
유니버시티 헬스 네트워크
캐나다 온타리오 엠5취 2썸4 토론토 엘리자베스
스트리트 190 알. 프레이저 엘리엇 빌딩-룸 1
에스-417
시나이 헬스 시스템
캐나다 엠5취 1엑스5 온타리오주 토론토 유니버시티
에비뉴 600
(72) 발명자
드 카르발로 다니엘 디니즈
캐나다 엠4이 3엑스7 온타리오 토론토 드러먼드빌
레인 31
브래트먼 스캇 빅터
캐나다 엠5티 2지4 온타리오 토론토 소호 스퀘어
44
(뒷면에 계속)
(74) 대리인
유미특허법인

전체 청구항 수 : 총 38 항

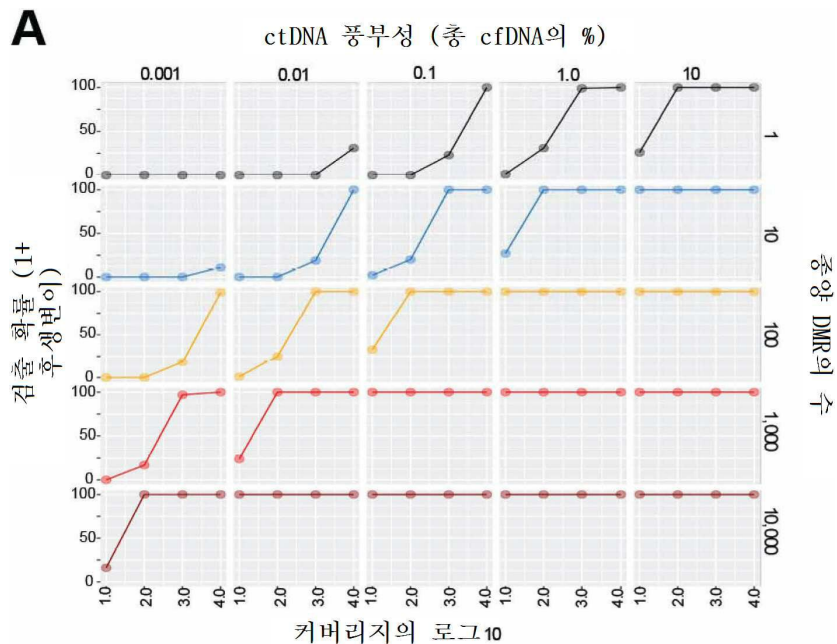
(54) 발명의 명칭 메틸롬 분석을 이용한 암 검출 및 분류

(57) 요약

본 발명은, 대상체로부터 무세포성 DNA 샘플을 제공하는 단계; 메틸화된 무세포성 DNA의 후속적 서열분석을 가능하게 하기 위해 상기 샘플에 대해 라이브러리를 제조하는 단계; 제1 소정량의 충전제 DNA를 상기 샘플에 첨가한 다음 선택적으로 상기 샘플을 변성시키는 단계로서, 상기 충전제 DNA는 적어도 일부가 메틸화된, 단계; 메틸화된

(뒷면에 계속)

대표도 - 도1a



폴리뉴클레오타이드에 대한 선택적인 결합제를 사용하여 메틸화된 무세포성 DNA를 포획하는 단계; 상기 포획된 메틸화된 무세포성 DNA를 서열분석하는 단계; 상기 포획된 메틸화된 무세포성 DNA의 서열을, 건강한 개체 및 구분되는 암 타입 및 서브타입의 암성 개체로부터 유래된 메틸화된 무세포성 대조군 DNA 서열과, 비교하는 단계; 및 상기 포획된 메틸화된 무세포성 DNA와 암성 개체로부터 유래된 메틸화된 무세포성 DNA 서열 중 하나 이상의 서열 간에 통계학적으로 유의한 유사성이 존재하는 경우, 암 세포 유래 DNA의 존재를 동정하는 단계를 포함하는, 대상체에서 암 세포 유래 DNA의 존재를 검출하는 방법을 개시한다.

(52) CPC특허분류

C12Q 2600/154 (2013.01)

(72) 발명자

싱하니아 라자트

캐나다 엠4와이 1에이치5 온타리오 토론토 아파트
806 웰즐리 스트리트 이스트 100

차크라바티 안쿠르 라비나라야나

캐나다 엠4엑스 1쥬7 온타리오 토론토 웰즐리 스트리트 이스트 280 #1414

센 슈 이

캐나다 엘3알 0브이9 온타리오 마크햄 밀리켄 미도우스 드라이브 213

명세서

청구범위

청구항 1

대상체에서 암 세포 유래 DNA의 존재를 검출하는 방법으로서,

상기 방법이:

- (a) 대상체로부터 무세포성 DNA(cell-free DNA) 샘플을 제공하는 단계;
- (b) 메틸화된 무세포성 DNA의 후속적 서열분석을 허용하기 위해, 상기 샘플에 대해 라이브러리를 제조하는 단계;
- (c) 제1 소정량의 충전제 DNA를 상기 샘플에 첨가한 다음 선택적으로 상기 샘플을 변성시키는 단계로서, 상기 충전제 DNA는 적어도 일부가 메틸화된 것인, 단계;
- (d) 메틸화된 폴리뉴클레오타이드에 대한 선택적인 결합제를 사용하여 메틸화된 무세포성 DNA를 포획하는 단계;
- (e) 상기 포획된 메틸화된 무세포성 DNA를 서열분석하는 단계;
- (f) 상기 포획된 메틸화된 무세포성 DNA의 서열과, 건강한 개체 및 암성 개체로부터의 메틸화된 무세포성 대조군 DNA 서열을, 비교하는 단계;
- (g) 상기 포획된 메틸화된 무세포성 DNA의 하나 이상의 서열과 암성 개체로부터 유래된 메틸화된 무세포성 DNA 서열 간에 통계학적으로 유의한 유사성이 존재하는 경우, 암 세포 유래 DNA의 존재를 동정하는 단계를 포함하는, 방법.

청구항 2

제1항에 있어서,

상기 샘플이 대상체의 혈액 또는 혈장으로부터 유래되는, 방법.

청구항 3

제1항 또는 제2항에 있어서,

비교 단계 (f)가 통계적 분류기(statistical classifier)를 이용한 피트(fit)를 토대로 하는, 방법.

청구항 4

제3항에 있어서,

상기 분류기가 기계 학습으로 파생되는(machine learning-derived), 방법.

청구항 5

제4항에 있어서,

상기 분류기가 탄력적 넷 분류기(elastic net classifier), 라소(lasso), 서포트 벡터 머신(support vector machine), 랜덤 포레스트(random forest) 또는 신경 네트워크인, 방법.

청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서,

상기 건강한 개체 및 암성 개체로부터의 메틸화된 무세포성 대조군 DNA 서열이 건강한 개체와 암성 개체 간에 차별적으로 메틸화된 영역 (DMR)의 데이터베이스에 포함된, 방법.

청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서,

상기 건강한 개체 및 암성 개체로부터의 메틸화된 무세포성 대조군 DNA 서열이, 무세포성 DNA로부터 유래된 DNA에서 건강한 개체와 암성 개체 간에 차별적으로 메틸화되는, 메틸화된 무세포성 대조군 DNA 서열로 한정되는, 방법.

청구항 8

제7항에 있어서,

상기 메틸화된 무세포성 대조군 DNA 서열이 혈장 유래 DNA에서 건강한 개체와 암성 개체 간에 차별적으로 메틸화되는, 방법.

청구항 9

제1항 내지 제8항 중 어느 한 항에 있어서,

상기 샘플이 무세포성 DNA를 100 ng, 75 ng, 또는 50 ng 미만으로 포함하는, 방법.

청구항 10

제1항 내지 제9항 중 어느 한 항에 있어서,

상기 제1 소정량의 충전제 DNA가 메틸화된 충전제 DNA를 약 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 또는 100%로 포함하고, 나머지는 비-메틸화된 충전제 DNA이고, 바람직하게 메틸화된 충전제 DNA를 5% 내지 50%, 10% 내지 40%, 또는 15% 내지 30%로 포함하는, 방법.

청구항 11

제1항 내지 제9항 중 어느 한 항에 있어서,

상기 제1 소정량의 충전제 DNA가 20 ng 내지 100 ng, 바람직하게 30 ng 내지 100 ng, 보다 바람직하게 50 ng 내지 100 ng인, 방법.

청구항 12

제1항 내지 제11항 중 어느 한 항에 있어서,

상기 샘플의 무세포성 DNA와 상기 제1 소정량의 충전제 DNA를 합하여 총 DNA의 50 ng 이상, 바람직하게 총 DNA의 100 ng 이상을 차지하는, 방법.

청구항 13

제1항 내지 제12항 중 어느 한 항에 있어서,

상기 충전제 DNA가 50 bp 내지 800 bp 길이이고, 바람직하게 100 bp 내지 600 bp 길이이고, 보다 바람직하게 200 bp 내지 600 bp 길이인, 방법.

청구항 14

제1항 내지 제13항 중 어느 한 항에 있어서,

상기 충전제 DNA가 이중 가닥인, 방법.

청구항 15

제1항 또는 제2항에 있어서,

상기 충전제 DNA가 정크(junk) DNA인, 방법.

청구항 16

제1항 내지 제3항 중 어느 한 항에 있어서,
상기 충전제 DNA가 내인성 또는 외인성 DNA인, 방법.

청구항 17

제16항에 있어서,
상기 충전제 DNA가 비-인간 DNA, 바람직하게 λ DNA인, 방법.

청구항 18

제1항 내지 제17항 중 어느 한 항에 있어서,
상기 충전제 DNA가 인간 DNA에 대한 얼라인먼트(alignment)가 없는, 방법.

청구항 19

제1항 내지 제18항 중 어느 한 항에 있어서,
상기 결합제가 메틸-CpG-결합 도메인을 포함하는 단백질인, 방법.

청구항 20

제1항 내지 제19항 중 어느 한 항에 있어서,
상기 단백질이 MBD2 단백질인, 방법.

청구항 21

제1항 내지 제20항 중 어느 한 항에 있어서,
단계 (d)가 메틸화된 무세포성 DNA를 항체를 사용해 면역침전시키는 것을 포함하는, 방법.

청구항 22

제21항에 있어서,
면역침전을 위해 상기 샘플에 항체를 0.05 μ g 이상, 바람직하게 0.16 μ g 이상으로 첨가하는 것을 포함하는, 방법.

청구항 23

제21항에 있어서,
상기 항체가 5-MeC 항체인, 방법.

청구항 24

제21항에 있어서,
면역침전 반응을 확인하기 위해 단계 (c) 후 제2 소정량의 대조군 DNA를 상기 샘플에 첨가하는 단계를 더 포함하는, 방법.

청구항 25

제1항 내지 제23항 중 어느 한 항에 있어서,
메틸화된 무세포성 DNA의 포획을 확인하기 위해 단계 (c) 후 제2 소정량의 대조군 DNA를 상기 샘플에 첨가하는 단계를 더 포함하는, 방법.

청구항 26

제1항 내지 제25항 중 어느 한 항에 있어서,

암 세포 유래 DNA의 존재를 동정하는 단계가 암 세포의 기원 조직을 동정하는 단계를 더 포함하는, 방법.

청구항 27

제26항에 있어서,

상기 암 세포의 기원 조직을 동정하는 단계가 암 서브타입을 동정하는 단계를 더 포함하는, 방법.

청구항 28

제27항에 있어서,

상기 암 서브타입이 병기(stage), 조직학, 유전자 발현 패턴, 카피수 이상, 재배열 또는 점 돌연변이 상태를 기반으로 암을 구별하는, 방법.

청구항 29

제1항 내지 제28항 중 어느 한 항에 있어서,

단계 (f)에서 비교가 게놈-전체(genome-wide)에서 수행되는, 방법.

청구항 30

제1항 내지 제28항 중 어느 한 항에 있어서,

단계 (f)에서 비교가 게놈-전체에서 특정 조절 영역으로 제한되는, 방법.

청구항 31

제30항에 있어서,

상기 조절 영역이 FANTOM5 인핸서, CpG 아일랜드, CpG 쇼어(shores), CpG 셸프(Shelves) 또는 이들의 조합인, 방법.

청구항 32

제1항 내지 제31항 중 어느 한 항에 있어서,

단계 (f) 및 (g)가 컴퓨터 프로세서에 의해 수행되는, 방법.

청구항 33

암 세포 유래 DNA의 존재를 검출하고 암 서브타입을 동정하는 방법으로서, 상기 방법이:

- 대상체 샘플로부터 메틸화된 무세포성 DNA의 서열분석 데이터를 입수하는 단계;
- 포획한 메틸화된 무세포성 DNA의 서열을, 건강한 개체 및 암성 개체로부터의 메틸화된 무세포성 대조군 DNA 서열과, 비교하는 단계;
- 상기 포획한 메틸화된 무세포성 DNA의 하나 이상의 서열과 암성 개체로부터 유래된 메틸화된 무세포성 DNA 서열 간에 통계학적으로 유의한 유사성이 존재하는 경우, 암 세포 유래 DNA의 존재를 동정하는 단계; 및
- 추가적으로, 단계 c에서 암 세포 유래 DNA가 동정된 경우, 단계 b에서의 비교에 기초하여 암 세포의 기원 조직 및 암 서브타입을 동정하는 단계

를 포함하는, 방법.

청구항 34

암 세포 유래 DNA의 존재를 검출하고 암 서브타입을 동정하기 위한 컴퓨터-구현 방법(computer-implemented method)으로서,

상기 방법이:

- 적어도 하나의 프로세서에서 대상체 샘플로부터 메틸화된 무세포성 DNA의 서열분석 데이터를 입수하는 단계;

b. 적어도 하나의 프로세서에서 포획한 메틸화된 무세포성 DNA의 서열을, 건강한 개체 및 암성 개체로부터의 메틸화된 무세포성 대조군 DNA 서열과, 비교하는 단계;

c. 적어도 하나의 프로세서에서 상기 포획한 메틸화된 무세포성 DNA의 하나 이상의 서열과 암성 개체로부터의 메틸화된 무세포성 DNA 서열 간에 통계학적으로 유의한 유사성이 존재하는 경우, 암 세포 유래 DNA의 존재를 동정하는 단계, 및 추가적으로, 단계 c로부터의 암 세포 기원의 DNA가 동정되는 경우, 단계 b에서의 비교에 기초하여 암 세포의 기원 조직 및 암 서브타입을 동정하는 단계

를 포함하는, 방법.

청구항 35

프로세서 및 상기 프로세서에 연결된 메모리가 구비된 범용 컴퓨터와 연계하여 사용하기 위한 컴퓨터 프로그램 제품으로서,

상기 컴퓨터 프로그램 제품이 암호화된 컴퓨터 메카니즘을 가진 컴퓨터 판독 가능한 저장 매체를 포함하고,

상기 컴퓨터 프로그램 메카니즘이 컴퓨터의 메모리에 로딩되어, 컴퓨터에 의해 제34항의 방법을 수행하게 할 수 있는, 컴퓨터 프로그램 제품.

청구항 36

제35항에 따른 컴퓨터 프로그램 제품을 저장하기 위한 저장된 데이터 구조를 가진 컴퓨터 판독 가능한 매체.

청구항 37

암 세포 유래 DNA의 존재를 검출하고 암 서브타입을 동정하기 위한 장치로서,

상기 장치가:

적어도 하나의 프로세서; 및

적어도 하나의 프로세서와 통신하는 전자 메모리를 포함하고,

상기 전자 메모리는, 상기 적어도 하나의 프로세서에서 실행되는 경우, 상기 적어도 하나의 프로세서가:

a. 대상체 샘플로부터 메틸화된 무세포성 DNA의 서열분석 데이터를 입수하고;

b. 포획한 메틸화된 무세포성 DNA의 서열과 건강한 개체 및 암성 개체로부터 유래된 메틸화된 무세포성 대조군 DNA 서열을 비교하고;

c. 포획한 메틸화된 무세포성 DNA의 하나 이상의 서열과 암성 개체로부터 유래된 메틸화된 무세포성 DNA 서열 간에 통계학적으로 유의한 유사성이 존재하는 경우, 암 세포 유래 DNA의 존재를 동정하고, 암 세포 유래 DNA가 동정되는 경우, 추가적으로, 단계 b에서의 비교에 기초하여 암 세포의 기원 조직 및 암 서브타입을 동정하게 하는, 프로세서-실행가능한 코드(processor-executable code)를 저장한, 장치.

청구항 38

암 세포 유래 DNA의 존재를 검출하고 2 이상의 가능성있는 장기로부터 암 세포가 발생하는 암의 위치를 결정하는 방법으로서,

상기 방법이:

(a) 대상체로부터 무세포성 DNA의 샘플을 제공하는 단계;

(b) 메틸화된 폴리뉴클레오타이드에 대한 선택적인 결합제를 사용하여 상기 샘플로부터 메틸화된 무세포성 DNA를 포획하는 단계;

(c) 상기 포획된 메틸화된 무세포성 DNA를 서열분석하는 단계;

(d) 상기 포획된 메틸화된 무세포성 DNA의 서열 패턴을 2 이상의 대조군 개체 집단(들)의 DNA 서열 패턴과 비교하는 단계로서, 상기 2 이상의 집단은 각각 서로 다른 장기에 암이 위치하는, 단계; 및

(e) 상기 무세포성 DNA와 상기 2 이상의 집단 간의 메틸화 패턴에 대한 통계학적으로 유의한 유사성에 기초하여

암 세포가 기원한 장기를 결정하는 단계
를 포함하는, 방법.

발명의 설명

기술 분야

- [0001] 본 출원은 2017년 7월 12일자에 출원된 미국 가특허 출원 번호 62/531527호에 대한 우선권을 주장하며, 이는 본원에 참조로 포함된다.
- [0002] 본 발명은 암 검출 및 분류, 및 보다 특히 이를 위한 메틸롬 분석의 용도에 관한 것이다.

배경 기술

- [0003] 바이오마커의 공급원으로서 순환 무세포성 DNA(cfDNA)의 사용은 종양학에서 급속도로 탄력을 받고 있다[1]. 바이오마커로서 cfDNA의 DNA 메틸화 맵핑의 용도는 액체 생검 분야에 큰 영향을 미칠 수 있는데, 이는 조직 기원의 동정을 가능하게 할 수 있거나[2], 암 타입 및 서브타입 분류를 가능하게 하고 최소의 침습성 양상으로 암 환자를 계층화시키기 때문이다[3]. 추가로, cfDNA의 게놈-전체 DNA 메틸화 맵핑을 사용하여 질환에 대해 어떠한 방사선촬영 증거도 갖지 않는 초기 단계 암을 갖는 환자에서 순환 종양 DNA(ctDNA)를 검출하는데 중요한 민감성 문제를 극복할 수 있다. 기존의 ctDNA 검출 방법은 돌연변이 서열분석에 기초하고, 부분적으로 종양과 정상 순환 cfDNA를 구분하기 위해 가용한 제한된 수의 재발 돌연변이로 인해 제한된 민감성을 갖는다[4,5]. 한편, 게놈-전체 DNA 메틸화 맵핑은 순환 종양 DNA(ctDNA)를 정상 순환 무세포성 DNA(cfDNA)와 구분하기 위해 사용될 수 있는 다수의 후성적 변화를 이용한다. 예를 들어, 일부 종양 유형, 예를 들어, 뇌질피복 세포종은 임의의 상당한 재발성 체세포 돌연변이 없이 광범위 DNA 메틸화 비정상성을 가질 수 있다[6].
- [0004] 메틸화된 무세포성 DNA를 포획하는 특정 방법은 참조로서 포함된 WO 2017/190215에 기재되어 있다.

선행기술문헌

비특허문헌

- [0005] (비특허문헌 0001) Diaz, L.A., Jr. and A. Bardelli, Liquid biopsies: genotyping circulating tumor DNA. J Clin Oncol, 2014. 32(6): p. 579-86.
- (비특허문헌 0002) Lehmann-Werman, R., et al., Identification of tissue-specific cell death using methylation patterns of circulating DNA. Proc Natl Acad Sci U S A, 2016. 113(13): p. E1826-34.
- (비특허문헌 0003) Visvanathan, K., et al., Monitoring of Serum DNA Methylation as an Early Independent Marker of Response and Survival in Metastatic Breast Cancer: TBCRC 005 Prospective Biomarker Study. J Clin Oncol, 2016: p. JCO2015662080.
- (비특허문헌 0004) Newman, A.M., et al., An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med, 2014. 20(5): p. 548-54.
- (비특허문헌 0005) Aravanis, A.M., M. Lee, and R.D. Klausner, Next-Generation Sequencing of Circulating Tumor DNA for Early Cancer Detection. Cell, 2017. 168(4): p. 571-574.
- (비특허문헌 0006) Mack, S.C., et al., Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. Nature, 2014. 506(7489): p. 445-50.
- (비특허문헌 0007) Taiwo, O., et al., Methylome analysis using MeDIP-seq with low DNA concentrations. Nat Protoc, 2012. 7(4): p. 617-36.
- (비특허문헌 0008) Lienhard, M., et al., MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. Bioinformatics, 2014. 30(2): p. 284-6.
- (비특허문헌 0009) Law, C.W., et al., voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol, 2014. 15(2): p. R29.
- (비특허문헌 0010) Chakravarthy, A., et al., Human Papillomavirus Drives Tumor Development

Throughout the Head and Neck: Improved Prognosis Is Associated With an Immune Response Largely Restricted to the Oropharynx. *J Clin Oncol*, 2016. 34(34): p. 4132-4141.

(비특허문헌 0011) Hoadley, K.A., et al., Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*, 2014. 158(4): p. 929-44.

(비특허문헌 0012) Fleischhacker, M. and B. Schmidt, Circulating nucleic acids (CNAs) and cancer--a survey. *Biochim Biophys Acta*, 2007. 1775(1): p. 181-232.

(비특허문헌 0013) Potter, N.T., et al., Validation of a real-time PCR-based qualitative assay for the detection of methylated SEPT9 DNA in human plasma. *Clin Chem*, 2014. 60(9): p. 1183-91.

(비특허문헌 0014) Legendre, C., et al., Whole-genome bisulfite sequencing of cell-free DNA identifies signature associated with metastatic breast cancer. *Clin Epigenetics*, 2015. 7: p. 100.

(비특허문헌 0015) Sun, K., et al., Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. *Proc Natl Acad Sci U S A*, 2015. 112(40): p. E5503-12.

(비특허문헌 0016) Chan, K.C., et al., Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. *Proc Natl Acad Sci U S A*, 2013. 110(47): p. 18761-8.

(비특허문헌 0017) Sharma, S., T.K. Kelly, and P.A. Jones, Epigenetics in cancer. *Carcinogenesis*, 2010. 31(1): p. 27-36.

(비특허문헌 0018) Sturm, D., et al., Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell*, 2012. 22(4): p. 425-37.

(비특허문헌 0019) Hinoue, T., et al., Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res*, 2012. 22(2): p. 271-82.

(비특허문헌 0020) Stirzaker, C., et al., Methylome sequencing in triple-negative breast cancer reveals distinct methylation clusters with prognostic value. *Nat Commun*, 2015. 6: p. 5899.

(비특허문헌 0021) Fang, F., et al., Breast cancer methylomes establish an epigenomic foundation for metastasis. *Sci Transl Med*, 2011. 3(75): p. 75ra25.

(비특허문헌 0022) Laurens van der Maaten, G.H., Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 2008. 9: p. 2579-2605.

(비특허문헌 0023) Kandoth, C., et al., Mutational landscape and significance across 12 major cancer types. *Nature*, 2013. 502(7471): p. 333-9.

(비특허문헌 0024) McGranahan, N., et al., Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci Transl Med*, 2015. 7(283): p. 283ra54.

(비특허문헌 0025) Zauber, P., S. Marotta, and M. Sabbath-Solitare, KRAS gene mutations are more common in colorectal villous adenomas and in situ carcinomas than in carcinomas. *Int J Mol Epidemiol Genet*, 2013. 4(1): p. 1-10.

(비특허문헌 0026) Martincorena, I., et al., Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*, 2015. 348(6237): p. 880-6.

(비특허문헌 0027) Beltran, H., et al., Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. 2016. 22(3): p. 298-305.

발명의 내용

일 측면에서, 본 발명은 대상체에서 암 세포 유래 DNA의 존재를 검출하는 방법을 제공하며, 상기 방법은 대상체

로부터 무세포성 DNA (cell-free DNA) 샘플을 제공하는 단계; 메틸화된 무세포성 DNA의 후속적 서열 분석을 가능하게 하기 위해 상기 샘플에 대해 라이브러리를 제조하는 단계; 제1 소정량의 충전제 DNA를 상기 샘플에 첨가한 다음 선택적으로 상기 샘플을 변성시키는 단계로서, 상기 충전제 DNA는 적어도 일부가 메틸화된, 단계; 메틸화된 폴리뉴클레오타이드에 대한 선택적인 결합제를 사용하여 메틸화된 무세포성 DNA를 포획하는 단계; 상기 포획된 메틸화된 무세포성 DNA를 서열분석하는 단계; 상기 포획된 메틸화된 무세포성 DNA의 서열을, 건강한 개체 및 암성 개체로부터 유래된 메틸화된 무세포성 대조군 DNA 서열과, 비교하는 단계; 상기 포획된 메틸화된 무세포성 DNA와 암성 개체 유래의 메틸화된 무세포성 DNA 서열 중 하나 이상의 서열 간에 통계학적으로 유의한 유사성이 존재하는 경우, 암 세포 유래 DNA의 존재를 동정하는 단계를 포함한다.

[0007] 일 측면에서, 본 발명은 암 세포 유래 DNA의 존재를 검출하고 암 서브타입을 동정하는 방법을 제공하며, 상기 방법은 대상체 샘플로부터 메틸화된 무세포성 DNA의 서열분석 데이터를 입수하는 단계; 포획한 메틸화된 무세포성 DNA의 서열을 건강한 개체 및 암성 개체로부터 유래된 메틸화된 무세포성 대조군 DNA 서열과 비교하는 단계; 상기 포획된 메틸화된 무세포성 DNA와 암성 개체로부터 유래된 메틸화된 무세포성 DNA 서열의 하나 이상의 서열 간에 통계학적으로 유의한 유사성이 존재하는 경우, 암 세포 유래 DNA의 존재를 동정하는 단계, 및 추가적으로, 암 세포 유래 DNA가 동정되는 경우, 상기 비교에 기초하여 암 세포 기원 조직 및 암 서브타입을 동정하는 단계를 포함한다.

[0008] 일 측면에서, 본 발명은, 암 세포 유래 DNA의 존재를 검출하고 암 서브타입을 동정하기 위한 컴퓨터-구현 방법 (computer-implemented method)을 제공하며, 상기 방법은 적어도 하나의 프로세서에서 대상체 샘플로부터 메틸화된 무세포성 DNA의 서열분석 데이터를 입수하는 단계; 적어도 하나의 프로세서에서 포획한 메틸화된 무세포성 DNA의 서열을, 건강한 개체 및 암성 개체로부터 유래된 메틸화된 무세포성 대조군 DNA 서열과, 비교하는 단계; 적어도 하나의 프로세서에서 상기 포획된 메틸화된 무세포성 DNA와 암성 개체로부터 유래된 메틸화된 무세포성 DNA 서열의 하나 이상의 서열 간에 통계학적으로 유의한 유사성이 존재하는 경우, 암 세포 유래 DNA의 존재를 동정하고, 암 세포 유래 DNA가 동정되는 경우, 상기 비교에 기초하여 암 세포 기원 조직 및 암 서브타입을 추가로 동정하는 단계를 포함한다.

[0009] 일 측면에서, 본 발명은, 프로세서 및 상기 프로세서에 연결된 메모리가 구비된 범용 컴퓨터 (general-purpose computer)와 연계하여 사용하기 위한 컴퓨터 프로그램 제품을 제공하며, 상기 컴퓨터 프로그램 제품은 암호화된 컴퓨터 메카니즘 (computer mechanism)을 가진 컴퓨터 판독 가능한 저장 매체를 포함하고, 상기 컴퓨터 프로그램 메카니즘은 상기 컴퓨터의 메모리로 로딩되어, 컴퓨터가 본원에 기재된 방법을 수행할 수 있게 한다.

[0010] 일 측면에서, 본 발명은 본원에 기재된 컴퓨터 프로그램 제품을 저장하기 위한 저장된 데이터 구조를 가진 컴퓨터 판독 가능한 매체를 제공한다.

[0011] 일 측면에서, 본 발명은 암 세포 유래 DNA의 존재를 검출하고 암 서브타입을 동정하기 위한 장치를 제공하며, 상기 장치는 적어도 하나의 프로세서; 및 적어도 하나의 프로세서와 통신하는 전자 메모리를 포함하며, 상기 전자 메모리는, 상기 적어도 하나의 프로세서에서 실행되는 경우 상기 적어도 하나의 프로세서가: 대상체 샘플로부터 메틸화된 무세포성 DNA의 서열분석 데이터를 입수하고; 포획한 메틸화된 무세포성 DNA의 서열을, 건강한 개체 및 암성 개체로부터 유래된 메틸화된 무세포성 대조군 DNA 서열과, 비교하고; 상기 포획한 메틸화된 무세포성 DNA와 암성 개체로부터 유래된 메틸화된 무세포성 DNA 서열의 하나 이상의 서열 간에 통계학적으로 유의한 유사성이 존재하는 경우, 암 세포 유래 DNA의 존재를 동정하고, 암 세포 유래 DNA가 동정되는 경우, 상기 비교에 기초하여 암 세포의 기원 조직 및 암 서브타입을 추가로 동정하게 하는, 프로세서 실행가능한 코드를 저장한다.

[0012] 일 측면에서, 본 발명은, 암 세포 유래 DNA의 존재를 검출하고 이로부터 암 세포가 2 이상의 가능성있는 장기로부터 발생하는 암의 위치를 검출하는 방법을 제공하며, 상기 방법은 대상체로부터 무세포성 DNA 샘플을 제공하는 단계; 메틸화된 폴리뉴클레오타이드에 대한 선택적인 결합제를 사용하여, 상기 샘플로부터 메틸화된 무세포성 DNA를 포획하는 단계; 상기 포획된 메틸화된 무세포성 DNA를 서열분석하는 단계; 상기 포획된 메틸화된 무세포성 DNA의 서열 패턴을 2의 이상의 대조군 개체 집단(들)의 DNA 서열 패턴과 비교하는 단계로서, 상기 2 이상의 집단은 각각 서로 다른 장기에 암이 위치하는, 단계; 및 무세포성 DNA와 상기 2 이상의 집단 간의 메틸화 패턴에 대한 통계학적으로 유의한 유사성에 기초하여, 암 세포가 어느 장기로부터 발생했는지를 결정하는 단계를 포함한다.

도면의 간단한 설명

[0013]

본 발명의 바람직한 구현예에 대한 상기한 및 그외 다른 특성들은 다음과 같이 첨부된 도면을 참조하여 하기의 상세한 설명으로부터 보다 자명해질 것이다:

도 1은 cfDNA의 메틸롬 분석이 낮은 양의 인풋 DNA에서 ctDNA를 집적(enrich)하고 검출하기 위한 고도의 민감성 접근법이다. **a)** ctDNA의 농도(컬럼), 조사되는 DMR의 수(열) 및 서열분석 깊이(x-축)의 함수로서 적어도 하나의 후생변이를 검출할 확률의 컴퓨터 시뮬레이션. **b)** 혈장 cfDNA를 모방하기 위해 단편화된 HCT116 세포주 기원의 1 내지 100 ng의 인풋 DNA에 대한 DNA 메틸화 신호 간의 계층-전체 피어슨 상호관계. 각각의 농도는 2개의 생물학적 레플리케이트를 갖는다. **c)** ENCODE(ENCSR000DFS)로부터 취득된 HCT116(Green Tracks) + PRBS(감소된 표현 바이설과이트 서열분석(Reduced Representation Bisulfite Sequencing)) HCT116 데이터 및 GEO(GSM1465024)로부터 취득된 WGBS(전체-계층 바이설과이트 서열분석(Whole-Genome Bisulfite Sequencing)) HCT116 데이터로부터 상이한 농도의 인풋 DNA 기원의 cfMeDIP-seq로부터 취득된 DNA 메틸화 프로파일. 히트맵(heatmap) (RRBS 트랙)에 대해, 황색은 메틸화된 것을 의미하고 청색은 비메틸화된 것을 의미하며 회색은 적용 범위가 없음을 의미한다. **d-e)** CRC 세포주 HCT116의 다중 골수종(MM) 세포주 MM1.S. cfMeDIP-seq로의 연속 희석은 순수 HCT116 DNA(100% CRC), 순수 MM1.S DNA(100% MM) 및 MM DNA로 희석된 10%, 1%, 0.1%, 0.01%, 및 0.001% CRC DNA에서 수행하였다. 모든 DNA는 혈장 cfDNA로 모방하기 위해 단편화시켰다. 본원 발명자들은 DMR의 관찰된 것 대 예상된(D) 수 및 (E) 상기 DMR 내 DNA 메틸화 신호(RPKM 중) 간의 거의 완벽한 선형 상호관계($r^2=0.99$, $p<0.0001$)를 관찰하였다. **f)** 동일한 연속 희석에서, 공지된 체세포 돌연변이는 백그라운드 서열분석기 및 폴리머라제 오류율 상에서 초심도(>10,000X) 표적화된 서열분석에 의해 1/100 대립형질유전자 분획에서만 검출 가능하다. CRC 세포주에서 각각의 돌연변이의 부위에서 각각의 염기 또는 삽입/결실을 함유하는 판독 분획을 나타낸다. **g)** 2명의 결장직장 암 환자 기원의 환자-유래된 이종이식체(PDX)를 보유한 마우스의 혈장에서 총 cfDNA(인간 + 마우스)의 백분율로서 ctDNA(인간)의 빈도.

도 2는 종양 분류를 가능하게 하는 혈장 cfDNA의 메틸롬 분석을 보여준다. **a)** 암 분류를 위한 기계 학습 분류기 구성의 접근법을 입증하는 도식. **b)** 다중-분류 탄력적 넷 머신 학습 분류기 내 함유된 DMR의 히트맵. 분류기는 건강한 공여자(n=24), 폐암(n=25), 유방암(n=25), 결장직장암(n=23), 급성 골수성 백혈병(AML) (n=28), 및 신경모세포종 다형태(GBM) (n=71) 기원의 혈장 DNA 샘플에 대해 트레이닝하였다. 계층 클러스터링 방법: 워드(Ward). **c)** 모델의 10% 또는 25%에서 동정된 암-유형 관련된 DMR의 tSNE(t-분포 확률적 임베딩(t-Distributed Stochastic Neighbor Embedding))에 의한 2D 가시화. **d)** 혈장 cfDNA 메틸화-기반 다중암 분류기를 위한 수행능 이점. 탄력적 넷 머신 학습 분류기의 50배 생성 후 각각의 암 타입 및 건강한 공여자를 위해 y축 상에 나타낸 수용자 작동자 곡선 이하 면적(auroc).

도 3은 독립적 코호트 상에서 다중암 분류기의 확증을 보여준다. **a)** ROC 곡선은 폐암(LUC) (n=55 LUC 대 n=97 기타), AML(n=35 AML 대 n=117 기타), 및 건강한 공여자(n=62 건강한 공여자 대 n=90 기타)의 코호트 상에서 다중암 분류기의 독립적 확장을 위해 나타낸다. **b)** ROC 곡선은 초기 단계 LUC(n=32 단계 I-II LUC 대 n=97 기타) 및 후기 단계 LUC(n=23 단계 III-IV LUC 대 n=97 기타) 상에 다중암 분류기의 독립적 확증을 위해 나타낸다.

도 4는 종양 서브타입 분류를 가능하게 하는 혈장 cfDNA의 메틸롬 분석을 보여준다. **a)** 암 서브타입 관련 DMR의 tSNE(t-분포 확률적 임베딩)에 의한 2D 가시화. 유방암 서브타입은 특유한 종양 카피수 비정상(HER2 상태) 뿐만 아니라 특유한 유전자 발현 패턴 및 전사 인자 활성(ER 상태)을 갖는 종양을 보유한 환자들을 구분하는 능력을 보여준다. AML 서브타입은 특유한 재배열(FLT3 상태)을 갖는 종양을 보유한 환자들을 구분하는 능력을 보여준다. 신경모세포종 다형태(GBM) 서브타입은 특유한 점 돌연변이(IDH 유전자 돌연변이 상태)를 갖는 종양을 보유한 환자들을 구분하는 능력을 보여준다. 폐암 서브타입은 예후적 및 치료학적 관련성(선암 종 대 편평 암종 대 소세포 암종)을 갖는 특유한 조직학을 사용한 종양을 보유한 환자들을 구분하는 능력을 보여준다. **b)** 유방암 혈장 샘플에서 3개의 유방암 서브타입의 정확한 식별을 가능하게 하는 상부 DMR을 보여주는 히트맵. **c)** AML 환자 혈장 샘플 중에서 FLT3-ITD 상태의 정확한 식별을 가능하게 하는 상부 DMR을 보여주는 히트맵. **d)** 신경모세포종 다형태(GBM) 환자 혈장 샘플에서 IDH 유전자 돌연변이 상태의 정확한 식별을 가능하게 하는 상부 DMR을 보여주는 히트맵. **e)** 폐암 혈장 샘플에서 3개의 폐암 조직학의 정확한 식별을 가능하게 하는 상부 DMR을 보여주는 히트맵.

도 5는 본원에 기재된 바와 같이 하나 이상의 구현예를 가능하게 하기 위한 플랫폼을 제공하기 위해 적합하게 구성된 컴퓨터 장치, 및 관련된 통신 네트워크, 장치, 소프트웨어 및 펌웨어를 보여준다.

도 6은 서열분석 포화 분석 및 품질 관리를 보여준다. **a)** 상기 도면은 혈장 cfDNA를 모방하기 위해 단편화된

HCT116 DNA로부터의 각각의 인풋 농도에 대한 각각의 레플리케이트로부터 cfMeDIP-seq 데이터를 분석하는 바이오컨덕터 패키지(Bioconductor package) MEDIPS로부터의 포화 분석 결과를 보여준다. **b)** 프로토콜은 HCT116 세포주의 4개의 출발 DNA 농도(100, 10, 5, 및 1 ng)의 2개의 레플리케이트에서 시험하였다. 반응의 특이성은 메틸화되고 비메틸화된 스파이크드-인(spiked-in) 에이. 탈리아나(A. thaliana) DNA를 사용하여 계산하였다. 배수 집적 비율은 단편화된 HCT116 DNA(메틸화된 고환-특이적 H2B, TSH2B0 및 비메틸화된 인간 DNA 영역(GAPDH 프로모터)에 대한 프라이머)의 게놈 영역을 사용하여 계산하였다. 수평 점선은 25의 배수-집적 비율 역치를 나타낸다. 오류 막대는 ± 1 s.e.m을 나타낸다. **c)** 서열분석된 샘플의 CpG 집적 스코어는 인풋 대조군과 비교하여 먼 역침전된 샘플로부터의 게놈 영역 내 CpG의 강한 집적을 보여준다. CpG 집적 스코어는 상기 영역의 CpG의 상대적 빈도를 인간 게놈의 CpG의 상대적 빈도로 나누어 수득하였다. 오류 막대는 ± 1 s.e.m을 나타낸다.

발명을 실시하기 위한 구체적인 내용

- [0014] 하기의 기재에서, 다수의 특이적 세부사항은 본 발명에 대한 완전한 이해를 제공하기 위해 제공된다. 그러나, 본 발명은 이들 특이적 세부사항 없이 실시될 수 있는 것으로 이해된다.
- [0015] DNA 메틸화 프로파일은 세포-타입에 특이적이며, 암에 걸렸을 때 교란된다. 극소량의 순환 무세포성 DNA(cfDNA)를 메틸롬 분석하기 위해 고안된 강건하고 민감한 방법을 사용하여, 본원 발명자들은 서로 간에, 그리고 건강한 개체로부터 복수의 종양 타입들을 구별하는 수천개의 차등적으로 메틸화된 영역(Differentially Methylated Region, DMR)을 동정하였다. cfDNA의 메틸롬 분석은 고도로 민감하며, 초기 단계 환자에서 순환 종양 DNA(ctDNA)를 검출하는데 적합하다. cfDNA 메틸롬을 이용한 기계-학습 유래 분류기(machine-learning derived classifier)는 교차 검증에 기초하여 5가지 암 타입 환자 및 건강한 공여자로부터 유래된 196개의 혈장 샘플을 정확하게 분류할 수 있었다. 독립적 확증에서, 혈장 cfDNA에서 동정된 동일한 DMR을 사용하여, 분류기는 초기 및 후기 단계의 폐암뿐 아니라 AML, 폐암 및 건강한 공여자를 올바르게 분류할 수 있었다. 따라서, cfDNA의 메틸롬 분석은 ctDNA의 비침습성 초기 단계 검출을 위해 사용될 수 있고 암 타입을 확실하게 분류할 수 있다.
- [0016] 일 측면에서, 본 발명은 대상체에서 암 세포 유래 DNA의 존재를 검출하는 방법을 제공하며, 상기 방법은 대상체로부터 무세포성 DNA의 샘플을 제공하는 단계; 샘플에 대해 라이브러리를 제조하여 메틸화된 무세포성 DNA의 후속적 서열분석을 가능하게 하는 단계; 제1 소정량의 충전제 DNA를 샘플에 첨가하고(상기 충전제 DNA 중 적어도 일부는 메틸화된), 이어서 선택적으로, 상기 샘플을 변성시키는 단계; 메틸화된 폴리뉴클레오타이드에 대한 선택적인 결합제를 사용하여 메틸화된 무세포성 DNA를 포획하는 단계; 포획된 메틸화된 무세포성 DNA를 서열분석하는 단계; 상기 포획된 메틸화된 무세포성 DNA의 서열을, 건강한 개체 및 암성 개체로부터 유래된 메틸화된 무세포성 대조군 DNA와, 비교하는 단계; 상기 포획된 메틸화된 무세포성 DNA의 하나 이상의 서열과 암성 개체로부터 유래된 메틸화된 무세포성 DNA 서열 간에 통계학적으로 유의한 유사성이 존재하는 경우, 암 세포 유래 DNA의 존재를 동정하는 단계를 포함한다.
- [0017] 2016년 5월 3일로 출원된 공동-출원의 출원 미국 가특허 출원 제62/331,070호 및 2017년 5월 3일자로 출원된 국제 특허 출원 번호 PCT/CA2017/000108은 메틸화된 무세포성 DNA를 포획하기 위한 방법을 개시하고 있으며, 이들 문헌은 원용에 의해 본 명세서에 포함된다.
- [0018] 암은 통상적으로 조직 기원 - 예를 들어, 결장직장암, 유방암, 폐암 등에 의해 통상적으로 분류되어 왔다. 임상 종양학의 현행 실무에서는, 다양한 분자, 발생 및 기능성 강화(functional underpinning)에 의해 암의 서브타입을 구분할 수 있는 것이 날로 중요해지고 있다. 치료학적 결정은 흔히 암의 정확한 서브타입에 의존하며, 임상 의는 치료요법 개시 전 서브타입을 반드시 파악하여야 할 수 있다. 치료학적 결정에 영향을 미칠 수 있는 암 서브타입 분류의 예로는, (비-제한적으로) 병기(예를 들어, 수술로 치료하는 초기 폐암 대 화학요법으로 치료하는 말기 폐암), 조직학(예를 들어, 폐암에서 소세포 암종 대 선암종 대 편평 세포 암종), 유전자 발현 패턴 또는 전사 인자 활성(예를 들어, 유방암에서 ER 상태), 카피수 이상(예를 들어, 유방암에서 HER2 상태), 특이적 재배열(예를 들어, AML에서 FLT3), 특이적 유전자 점 돌연변이 상태(예를 들어, IDH 유전자 점 돌연변이), 및 DNA 메틸화 패턴(예를 들어, 뇌암에서 MGMT 유전자 프로모터 메틸화) 등이 있다.
- [0019] 본원에 기재된 방법은 광범위한 암에 적용될 수 있으며, 이는 부신암, 항문암, 담관암, 방광암, 골암, 뇌/cns 종양, 유방암, 캐슬먼 질환, 자궁경부암, 결장/직장암, 자궁내막암, 식도암, 유잉 계열의 종양(ewing family of tumors), 안암, 담낭암, 위장 카르시노이드 종양, 위장 기질 종양(gist), 임신 영양막 질환, 호지킨 질환, 카포시 육종, 신장암, 후두 및 하인두암, 백혈병(급성 림프구, 급성 골수, 만성 림프구, 만성 골수, 만성 골단핵구), 간암, 폐암(비-소세포, 소 세포, 폐 카르시노이드 종양), 림프종, 피부의 림프종, 악성 중피종, 다중 골수종, 골이형성 증후군, 비강 및 부비동 암, 비인두암, 신경모세포종, 비-호지킨 림프종, 구강 및 구강인두암, 골

육종, 난소암, 음경암, 뇌하수체 종양, 전립선암, 망막모세포종, 횡문근육종, 침샘암, 육종-성인 연조직암, 피부암(기저 및 편평세포, 흑색종, 메르켈 세포), 소장암, 위암, 고환암, 흉선암, 갑상선암, 자궁 육종, 질암, 외음부암, 발텐스트롬 마크로글로불린혈증, 빌름스 종양을 포함하지만 이들로 제한되는 것은 아니다.

- [0020] 폴리머라제 연쇄 반응(PCR) 및 후속적인 생거 서열분석과 같은 다양한 서열분석 기술들이 당업자에게 공지되어 있다. 또한, 고 성능 서열분석 (high-throughput sequencing)으로도 공지된 차세대 서열분석(NGS) 기술 역시 이용가능하며, 이는 다음과 같은 다양한 서열분석 기술을 포함한다: 일루미나(Solexa) 서열분석, Roche 454 서열분석, 이온 토렌트(Ion torrent): 양성자/PGM 서열분석, SOLiD 서열분석. NGS는 기존의 생거 서열분석보다 더 신속하고 저렴하게 DNA 및 RNA의 서열분석을 가능하게 한다. 일부 구현예에서, 상기 서열분석은 짧은 판독 서열 분석을 위해 최적화된다.
- [0021] 본원에 사용된 바와 같은 용어 "대상체"는 동물계의 임의의 구성원, 바람직하게 인간, 가장 바람직하게는 전립선 암을 앓고 있거나, 앓았거나, 걸린 것으로 의심되는 인간을 지칭한다.
- [0022] 메틸화된 무세포성 DNA는 혈류에서 자유롭게 순환하는 DNA이고, DNA의 다양한 공지된 영역에서 메틸화된다. 샘플, 예를 들어, 혈장 샘플은 메틸화된 무세포성 DNA를 분석하기 위해 취해질 수 있다. 따라서, 일부 구현예에서, 샘플은 대상체의 혈액 또는 혈장이다.
- [0023] 본원에 사용된 바와 같이, "라이브러리 제조"는 리스트 말단-복구, A-테일링, 어댑터 라이게이션, 또는 DNA의 후속적 서열분석을 가능하게 하기 위해 무세포성 DNA에 대해 수행되는 임의의 다른 제조를 포함한다.
- [0024] 본원에 사용된 바와 같은 "충전제 DNA"는 비암호화 DNA(noncoding DNA)일 수 있거나, 앰플리콘(amplicons)들로 이루어질 수 있다.
- [0025] DNA 샘플은, 예를 들어, 충분한 열을 사용하여 변성될 수 있다.
- [0026] 일부 구현예에서, 비교 단계는 통계적 분류기를 이용한 피트(fit)를 토대로 한다. DNA 메틸화 데이터를 이용하는 통계적 분류기는 특정 질환 상태, 예를 들어, 암 타입 또는 서브타입과 같은 특정 질환 상태로 지정하기 위해 사용될 수 있다. 암 타입 또는 서브타입 분류 목적을 위해, 분류기는 통계학적 모델 내 하나 이상의 DNA 메틸화 변수(즉, 특성)로 이루어지고, 통계학적 모델의 아웃풋은 특유의 질환 상태를 구분하기 위해 하나 이상의 역치 값을 갖는다. 통계적 분류기에서 사용되는 특정 특성(들) 및 역치 값(들)은 암 타입 또는 서브타입의 사전 지식으로부터, 가장 유익한 정보를 제공할 가능성이 있는 특성들의 사전 지식으로부터, 기계 학습으로부터, 또는 이들 접근법의 2개 이상의 조합으로부터 유래할 수 있다.
- [0027] 일부 구현예에서, 분류기는 기계 학습-유래된다. 바람직하게, 분류기는 탄력적 넷 분류기(elastic net classifier), 라소(lasso), 서포트 벡터 머신(support vector machine), 랜덤 포레스트(random forest) 또는 신경 네트워크이다.
- [0028] 분석되는 게놈 공간은 게놈-전체일 수 있거나, 바람직하게 조절 영역(즉, FANTOM5 인핸서, CpG 아일랜드, CpG 쇼어(shores) 및 CpG 셀프)으로 제한된다.
- [0029] 바람직하게는, 수득한 스파이크-인 메틸화된 DNA(spike-in methylated DNA)의 백분율은 폴다운 효율 변동을 제어하기 위한 공변량으로서 포함된다.
- [0030] 다중 암 타입 (또는 서브타입)을 상호 구분할 수 있는 분류기의 경우, 상기 분류기는 바람직하게 관심 대상의 각각의 유형 (또는 서브타입)의 쌍별 비교로부터 차등적으로 메틸화된 영역으로 이루어진다.
- [0031] 일부 구현예에서, 건강한 개체 및 암성 개체로부터의 메틸화된 무세포성 대조군 DNA 서열은 건강한 개체와 암성 개체 간에 차등적으로 메틸화된 영역(DMR)의 데이터베이스에 포함된다.
- [0032] 일부 구현예에서, 건강한 개체 및 암성 개체로부터 유래된 메틸화된 무세포성 대조군 DNA 서열은, 예를 들어, 혈청, 뇌척수액, 뇨변, 가래, 흉수액, 복수, 눈물, 땀, 자궁경부액(pap smear fluid), 내시경 브러싱액(endoscopy brushings fluid) 등과 같은 체액 기원, 바람직하게는 혈장 기원의 무세포성 DNA로부터 유래된 DNA에서 건강한 개체 및 암성 개체 간에서와 같이 차등적으로 메틸화된 메틸화된 무세포성 대조군 DNA 서열로 제한된다.
- [0033] 일부 구현예에서, 샘플은 무세포성 DNA를 100 ng, 75 ng, 또는 50 ng 미만으로 포함한다.
- [0034] 일부 구현예에서, 제1 소정량의 충전제 DNA는 메틸화된 충전제 DNA를 약 5%, 10%, 15%, 20%, 30%, 40%, 50%,

60%, 70%, 80%, 90%, 또는 100%로 포함하며, 나머지는 비-메틸화된 충전제 DNA이고, 바람직하게 메틸화된 충전제 DNA를 5% 내지 50%, 10% 내지 40%, 또는 15% 내지 30%로 포함한다.

- [0035] 일부 구현예에서, 제1 소정량의 충전제 DNA는 20 ng 내지 100 ng, 바람직하게 30 ng 내지 100 ng, 보다 바람직하게 50 ng 내지 100 ng이다.
- [0036] 일부 구현예에서, 샘플로부터 무세포성 DNA 및 제1 소정량의 충전제 DNA는 함께 총 DNA의 적어도 50 ng, 바람직하게 총 DNA의 적어도 100 ng을 차지한다.
- [0037] 일부 구현예에서, 충전제 DNA는 50 bp 내지 800 bp 길이고, 바람직하게 100 bp 내지 600 bp 길이며, 보다 바람직하게 200 bp 내지 600 bp 길이다.
- [0038] 일부 구현예에서, 충전제 DNA는 이중가닥이다. 충전제 DNA는 이중가닥이다. 예를 들어, 충전제 DNA는 정크(junk) DNA일 수 있다. 충전제 DNA는 또한 내인성 또는 외인성 DNA일 수 있다. 예를 들어, 충전제 DNA는 비-인간 DNA이고, 바람직한 구현예에서, λDNA이다. 본원에 사용된 바와 같은 "λDNA"는 엔테로박테리아(*Enterobacteria*) 파아지 λDNA이다. 일부 구현예에서, 충전제 DNA는 인간 DNA에 대한 얼라인먼트(alignment)를 가지지 않는다.
- [0039] 일부 구현예에서, 결합제는 메틸-CpG-결합 도메인을 포함하는 단백질이다. 하나의 상기 예시적 단백질은 MBD2 단백질이다. 본원에 사용된 바와 같은 "메틸-CpG-결합 도메인(MBD)"은 단백질 및 효소의 특정 도메인을 지칭하며, 이는 대략 70개 잔기 길이며, 하나 이상의 대칭적으로 메틸화된 CpG를 함유하는 DNA에 결합한다. MeCP2, MBD1, MBD2, MBD4 및 BAZ2의 MBD는 DNA에의 결합을 매개하고, MeCP2, MBD1 및 MBD2의 경우에는 우선적으로 메틸화된 CpG에의 결합을 매개한다. 인간 단백질 MECP2, MBD1, MBD2, MBD3 및 MBD4는 메틸-CpG-결합 도메인(MBD) 각각의 존재에 의해 관련있는 핵 단백질 패밀리를 포함한다. MBD3을 제외한 이들 단백질은 각각 메틸화된 DNA에 특이적으로 결합할 수 있다.
- [0040] 다른 구현예에서, 결합제는 항체이며, 메틸화된 무세포성 DNA를 포획하는 것은 항체를 이용하여 메틸화된 무세포성 DNA를 면역침전시키는 것을 포함한다. 본원에 사용된 바와 같은 "면역침전"은 상기 특정 항원에 특이적으로 결합하는 항체를 사용하여 용액으로부터 항원(예를 들어, 폴리펩타이드 및 뉴클레오타이드)을 침전시키는 기술을 의미한다. 이러한 과정은 샘플로부터 특정 단백질 또는 DNA를 분리 및 농축하기 위해 사용될 수 있으며, 항체는 이 과정에서 일부 시점에 고체 기질에 커플링하여야 한다. 고체 기질은, 예를 들어, 자기 비드와 같은 비드를 포함한다. 다른 유형의 비드 및 고체 기질들도 당업계에 공지되어 있다.
- [0041] 하나의 예시적 항체는 5-MeC 항체이다. 면역침전 과정을 위해, 일부 구현예에서, 적어도 0.05 μg의 항체를 샘플에 첨가하고; 보다 바람직한 구현예에서, 적어도 0.16 μg의 항체를 샘플에 첨가한다. 면역침전 반응을 확인하기 위해, 일부 구현예에서, 본원에 기재된 방법은 제2 소정량의 대조군 DNA를 샘플에 첨가하는 단계를 추가로 포함한다.
- [0042] 일부 구현예에서, 상기 방법은 면역침전 반응을 확인하기 위해 제2 소정량의 대조군 DNA를 샘플에 첨가하는 단계를 추가로 포함한다.
- [0043] 본원에 사용된 바와 같은 "대조군"은 양성 대조군 및 음성 대조군 둘 다, 또는 적어도 양성 대조군을 포함할 수 있다.
- [0044] 일부 구현예에서, 상기 방법은 메틸화된 무세포성 DNA의 포획을 확인하기 위해 제2 소정량의 대조군 DNA를 샘플에 첨가하는 단계를 추가로 포함한다.
- [0045] 일부 구현예에서, 암 세포 유래 DNA의 존재를 동정하는 단계는 암 세포의 기원 조직을 동정하는 것을 추가로 포함한다.
- [0046] 일부 경우에, 종양 조직 샘플링은 해결 과제이거나 상당한 위험을 안고 있으며, 이 경우에 종양 조직 샘플링 없이 암을 진단하고/하거나 서브타입 분류하는 것이 적합할 수 있다. 예를 들어, 폐 종양 조직 샘플링은 중막내시경 검사, 흉강 절제술 또는 경피 바늘 생검과 같은 침습성 절차가 필요할 수 있고; 이들 절차는 입원, 흉관, 기계적 환기, 항생제 또는 다른 의학적 개입을 요할 수 있다. 일부 개체는 의학적 동반이환 때문에 또는 선호도로 인해 종양 조직 샘플링을 위해 필요한 침습성 절차를 진행하지 않을 수 있다. 일부 경우에, 종양 조직 채취를 위한 실제 절차는 의심되는 암 서브타입에 의존할 수 있다. 다른 경우에, 암 서브타입은 동일한 개체 내에서 시간 경과에 따라 진화할 수 있어; 침습성 종양 조직 샘플링 절차를 이용한 일련의 평가는 종종 비실용적이고 환자가 잘 받아들이는 것은 아니다. 따라서, 혈액 검사를 통한 비-침습적인 암 서브타입 분류가 임상적 종양학 실

무에서 많은 유리한 이용성을 가질 수 있다.

- [0047] 이에, 일부 구현예에서, 암 세포 기원 조직을 동정하는 단계는 암 서브타입을 동정하는 것을 추가로 포함한다. 바람직하게는, 암 서브타입은 하기를 기준으로 암을 구분한다: 병기(예를 들어, 수술로 치료하는 초기 폐암 대 화학치료요법으로 치료하는 말기 폐암), 조직학(예를 들어, 폐암에서 소세포 암종 대 선암종 대 편평 세포 암종), 유전자 발현 패턴 또는 전사 인자 활성화(예를 들어, 유방암에서 ER 상태), 카피수 이상(예를 들어, 유방암에서 HER2 상태), 특이적 재배열(예를 들어, AML에서 FLT3), 특이적 유전자 점 돌연변이 상태(예를 들어, IDH 유전자 점 돌연변이), 및 DNA 메틸화 패턴(예를 들어, 뇌암에서 MGMT 유전자 프로모터 메틸화).
- [0048] 일부 구현예에서, 단계 (f)에서 비교는 게놈-전체에서 수행된다.
- [0049] 다른 구현예에서, 단계 (f)에서의 비교는 게놈-전체로부터 특이적 조절 영역, 예를 들어, 비-제한적으로 FANTOM5 인핸서, CpG 아일랜드, CpG 쇼어, CpG 셸프, 또는 이들의 임의의 조합으로 제한된다.
- [0050] 일부 구현예에서, 특정 단계는 컴퓨터 프로세서에 의해 수행된다.
- [0051] 일 측면에서, 본 발명은 암 세포 유래 DNA의 존재를 검출하고 암 서브타입을 동정하는 방법을 제공하며, 상기 방법은 대상체 샘플 기원의 메틸화된 무세포성 DNA의 서열분석 데이터를 입수하는 단계; 포획한 메틸화된 무세포성 DNA의 서열을 건강한 개체 및 암성 개체로부터 유래된 메틸화된 무세포성 대조군 DNA 서열과 비교하는 단계; 포획한 메틸화된 무세포성 DNA의 하나 이상의 서열과 암성 개체로부터 유래된 메틸화된 무세포성 DNA 서열 간에 통계학적으로 유의한 유사성이 존재하는 경우, 암 세포 유래 DNA의 존재를 동정하는 단계; 및 암 세포 기원의 DNA가 동정되는 경우, 상기 비교 단계에 기초하여 암 세포 기원 조직 및 암 서브타입을 추가로 동정하는 단계를 포함한다
- [0052] 일 측면에서, 본 발명은 암 세포 유래 DNA의 존재를 검출하고 2 이상의 가능성있는 장기로부터 암 세포가 유래된 암의 위치를 결정하는 방법을 제공하며, 상기 방법은 대상체로부터 무세포성 DNA의 샘플을 제공하는 단계; 상기 샘플로부터 메틸화된 폴리뉴클레오타이드에 대한 선택적인 결합제를 사용하여 메틸화된 무세포성 DNA를 포획하는 단계; 상기 포획된 메틸화된 무세포성 DNA를 서열분석하는 단계; 상기 포획된 메틸화된 무세포성 DNA의 서열 패턴을 2 이상의 대조군 개체 집단(들)의 DNA 서열 패턴과 비교하는 단계로서, 상기 2 이상의 집단은 각각 서로 다른 장기에서 국소 암을 가진, 단계; 무세포성 DNA와 상기 2 이상의 집단 간에 메틸화 패턴에 대한 통계학적으로 유의한 유사성에 기초하여 어느 장기로부터 암 세포가 발생하는지를 결정하는 단계를 포함한다.
- [0053] 본 시스템 및 방법은 다양한 구현예로 구현될 수 있다. 적합하게 구성된 컴퓨터 장치, 및 관련된 통신 네트워크, 장치, 소프트웨어 및 펌웨어는 상기된 바와 같이 하나 이상의 구현예를 가능하게 하기 위한 플랫폼을 제공할 수 있다. 예를 들어, 도 5는 저장 유닛(104) 및 랜덤 액세스 메모리(106)에 연결된 중앙 처리 유닛("CPU") (102)을 포함할 수 있는 범용 컴퓨터 장치(100)를 보여준다. CPU(102)는 작동 시스템(101), 응용 프로그램(103), 및 데이터(123)를 처리할 수 있다. 작동 시스템(101), 응용 프로그램(103), 및 데이터(123)는 요구될 수 있는 바와 같이 저장 유닛(104)에 저장될 수 있고 메모리(106)에 로딩될 수 있다. 컴퓨터 장치(100)는 그래픽 처리 유닛(GPU) (122)을 추가로 포함할 수 있고 이는 작동적으로 CPU(102) 및 메모리(106)에 연결되어 있어 CPU(102)로부터 집중 이미지 처리 계산을 오프로딩하고 이들 계산을 CPU(102)와 병행하여 전개한다. 작동자(107)는 비디오 인터페이스(105)에 의해 연결된 비디오 디스플레이(108)를 사용하여 컴퓨터 장치(100), 및 키보드(115), 마우스(112)와 같은 다양한 인풋/아웃풋 장치, 및 I/O 인터페이스(109)에 의해 연결된 디스크 드라이브 또는 고체 상태 드라이브(114)와 상호작용할 수 있다. 공지된 방식으로, 마우스(112)는 비디오 디스플레이(108)에서 커서의 이동을 제어하고 마우스 버튼과 함께 비디오 디스플레이(108)에서 나타나는 다양한 그래픽 사용자 인터페이스(GUI) 제어를 작동시키도록 구성될 수 있다. 디스크 드라이브 또는 고체 상태 드라이브(114)는 컴퓨터 판독 가능한 매체(116)를 수용하도록 구성될 수 있다. 컴퓨터 장치(100)는 네트워크 인터페이스(111)를 통한 네트워크의 일부를 형성하여 상기 컴퓨터 장치(100)는 다른 적합하게 구성된 데이터 처리 시스템(미도시됨)과 통신하도록 할 수 있다. 하나 이상의 상이한 유형의 센서(135)는 다양한 공급원으로부터 인풋을 수용하기 위해 사용될 수 있다.
- [0054] 본 시스템 및 방법은 데스크탑 컴퓨터, 랩탑 컴퓨터, 태블릿 컴퓨터 또는 무선 핸드헬드를 포함하는 실제로 임의의 방식의 컴퓨터 장치 상에서 수행될 수 있다. 본 시스템 및 방법은 또한 본 발명에 따른 방법에서 하나 이상의 컴퓨터 장치가 다양한 공정 단계의 각각을 구현하도록 할 수 있는 컴퓨터 프로그램 코드를 포함하는 컴퓨터-판독 가능하고/사용 가능한 매체로서 구현될 수 있다. 전체 작동을 수행하는 컴퓨터 장치를 초과하는 경우에, 컴퓨터 장치는 다양한 작동 단계를 분배하도록 네트워크화된다. 컴퓨터-판독 가능한 매체 또는 컴퓨터

사용 가능한 매체라는 용어는 프로그램 코드의 하나 이상의 임의의 유형의 물리적 구현예를 포함하는 것으로 이해된다. 특히, 컴퓨터-판독 가능/사용 가능한 매체는 컴퓨터 계산 장치, 예를 들어, 컴퓨터 및/또는 저장 시스템과 관련된 메모리의 분할된 하나 이상의 데이터 저장 상에 하나 이상의 휴대용 저장 제품(예를 들어, 광 디스크, 자기 디스크, 테이프 등) 상에 구현된 프로그램 코드를 포함할 수 있다.

[0055] 일 측면에서, 본 발명은 암 세포 유래 DNA의 존재를 검출하고 암 서브타입을 동정하기 위한 컴퓨터-구현 방법을 제공하며, 상기 방법은 적어도 하나의 프로세서에서 대상체 샘플로부터 메틸화된 무세포성 DNA의 서열분석 데이터를 입수하는 단계; 적어도 하나의 프로세서에서 상기 포획된 메틸화된 무세포성 DNA의 서열을 건강한 개체 및 암성 개체로부터 유래된 메틸화된 무세포성 대조군 DNA 서열과 비교하는 단계; 적어도 하나의 프로세서에서 상기 포획된 메틸화된 무세포성 DNA와 암성 개체 기원의 메틸화된 무세포성 DNA 서열의 하나 이상의 서열 간에 통계학적으로 유의한 유사성이 존재하는 경우, 암 세포 유래 DNA의 존재를 동정하는 단계 및 암 세포 유래 DNA가 동정되는 경우, 상기 비교 단계에 기초하여 암 세포 기원 조직 및 암 서브타입을 추가로 동정하는 단계를 포함한다.

[0056] 일 측면에서, 본 발명은 프로세서 및 상기 프로세서에 연결된 메모리가 구비된 범용 컴퓨터와 연계하여 사용하기 위한 컴퓨터 프로그램 제품을 제공하며, 상기 컴퓨터 프로그램 제품은 그 위에 암호화된 컴퓨터 메커니즘을 갖는 컴퓨터 판독 저장 매체를 포함하며, 여기서, 상기 컴퓨터 프로그램 메커니즘은 상기 컴퓨터의 메모리로 로딩될 수 있고 컴퓨터가 본원에 기재된 방법을 수행하도록 한다.

[0057] 일 측면에서, 본 발명은, 본원에 기재된 컴퓨터 프로그램 제품을 저장하기 위한 데이터 구조를 그 위에 저장한 컴퓨터 판독 가능한 매체를 제공한다.

[0058] 일 측면에서, 본 발명은 암 세포 유래 DNA의 존재를 검출하고 암 서브타입을 동정하기 위한 장치를 제공하며, 상기 장치는 적어도 하나의 프로세서; 및 상기 적어도 하나의 프로세서와 통신하는 전자 메모리를 포함하며, 상기 전자 메모리는 상기 적어도 하나의 프로세서에서 실행되는 경우 상기 적어도 하나의 프로세서가 대상체 샘플로부터 메틸화된 무세포성 DNA의 서열분석 데이터를 수용하도록 하고; 상기 포획된 메틸화된 무세포성 DNA의 서열을 건강한 개체 및 암성 개체로부터 유래된 메틸화된 무세포성 대조군 DNA 서열과 비교하도록 하고; 상기 포획된 메틸화된 무세포성 DNA와 암성 개체 기원의 메틸화된 무세포성 DNA 서열의 하나 이상의 서열 간에 통계학적으로 유의한 유사성이 존재하는 경우, 암 세포 유래 DNA의 존재를 동정하도록 하고; 이로부터 기원하는 암 세포 유래 DNA가 동정되는 경우, 상기 비교 단계를 기반으로 암 세포 기원 조직 및 암 서브타입을 추가로 동정하도록 하는 프로세서 실행가능한 코드를 저장한다.

[0059] 본원에 사용된 바와 같은, "프로세서"는, 예를 들어, 임의의 유형의 범용 마이크로프로세서 또는 마이크로컨트롤러(예를 들어, IntelTM x86, PowerPCTM, ARMTM 프로세서 등), 디지털 신호 처리(DSP) 프로세서, 집적 회로, 필드 프로그램가능한 게이트 어레이(FPGA) 또는 이의 임의의 조합체와 같은 임의의 유형의 프로세서일 수 있다.

[0060] 본원에 사용된 바와 같은 "메모리"는, 예를 들어, 랜덤-액세스 메모리(RAM), 판독 전용 메모리(ROM), 콤팩트 디스크 판독 전용 메모리(CDROM), 전기-광학 메모리, 자기-광학 메모리, 소거 가능 프로그램 가능한 판독 전용 메모리(EPROM) 및 전기 소거 가능 프로그램 가능한 판독 전용 메모리(EPROM), 및 전기적으로 소거 가능 프로그램 가능한 판독 전용 메모리(EEPROM) 등과 같이 내부적으로 또는 외부적으로 위치한 임의의 유형의 컴퓨터 메모리의 적합한 조합체를 포함할 수 있다. 메모리(102) 부분은 통상의 파일 시스템을 사용하여 구성될 수 있고, 장치의 전체 작동을 제어하는 작동 시스템에 의해 제어되고 관리될 수 있다.

[0061] 본원에 사용된 바와 같은 "컴퓨터 판독 가능한 저장 매체" (기계-판독 가능한 매체, 프로세서-판독 가능한 매체 또는 이에 구현된 컴퓨터-판독 가능한 매체를 갖는 컴퓨터 사용 가능한 매체로도 언급됨)는 컴퓨터 또는 기계에 의해 판독 가능한 포맷으로 데이터를 저장할 수 있는 매체이다. 기계-판독 가능한 매체는 디스켓, 콤팩트 디스크 판독 전용 메모리(CD-ROM), 메모리 장치(휘발성 또는 비휘발성), 또는 유사 저장 메커니즘을 포함하는, 자기, 광학, 또는 전기 저장 매체를 포함하는 임의의 적합한 유형의 비일시적 매체일 수 있다. 컴퓨터 판독 가능한 저장 매체는 다양한 세트의 지침, 코드 서열, 구성 정보 또는 실행되는 경우 프로세서가 본원의 개시내용의 구현예에 따른 방법에서 단계를 수행하도록 하는 기타 데이터를 함유할 수 있다. 통상의 기술자는 기재된 수행을 수행하기 위해 필요한 다른 지침 및 작동이 또한 컴퓨터 판독 가능한 저장 매체 상에 저장될 수 있음을 이해할 것이다. 컴퓨터 판독 가능한 저장 매체 상에 저장된 지침은 프로세서 또는 다른 적합한 처리 장치에 의해 실행될 수 있고, 기재된 작업을 수행하기 위해 회로와 인터페이스 할 수 있다.

[0062] 본원에 사용된 바와 같은 "데이터 구조"는 이것이 효율적으로 사용될 수 있도록 컴퓨터에 데이터를 구성하는 특

정 방식이다. 데이터 구조는 하나 이상의 특정 발체 데이터 유형(ADT)을 실행할 수 있고 이는 상기 데이터 구조 상에서 실행될 수 있는 작동 및 이들 작동의 계산 복잡성을 특징한다. 이에 비해, 데이터 구조는 ADT에 의해 제공된 사양의 구체적인 구현이다.

[0063] 본 발명의 이점은 하기의 실시예에 의해 추가로 설명된다. 실시예 및 본원에 제시된 이들의 특정 세부사항은 단지 설명을 위해 제공되고 본 발명의 청구항에 대한 제한으로서 해석되지 말아야 한다.

[0064] **실시예**

[0065] **방법 및 재료**

[0066] *공여자 모집 및 샘플 획득*

[0067] CRC, 유방암, 및 GBM 샘플은 기관(University Health Network BioBank)으로부터 입수하였고; AML 샘플은 기관(University Health Network Leukemia BioBank)으로부터 입수하였고; 마지막으로, 건강한 대조군은 캐나다 토론토 내 마운트 시나이 병원(MSH))에서 가족 의학 센터를 통해 모집하였다. 환자 동의와 함께 수거된 모든 샘플은 캐나다 토론토에 위치한 기관(University Health Network 및 Mount Sinai Hospital)의 윤리 위원회로부터의 기관 승인을 받아 입수하였다.

[0068] *표본 처리 - cfDNA*

[0069] EDTA 및 ACD 혈장 샘플은 기관(BioBank)으로부터 그리고 캐나다 토론토 내 마운트 시나이 병원(MSH) 내 가족 의학 센터로부터 획득하였다. 모든 샘플은 사용때 까지 -80℃에서 또는 기상 액체 질소에서 저장하였다. 무세포성 DNA는 QIAamp 순환 핵산 키트(Qiagen)를 사용하여 0.5-3.5 ml의 혈장으로부터 추출하였다. 추출된 DNA는 사용 전에 Qubit를 통해 정량하였다.

[0070] *표본 처리 - PDX cfDNA*

[0071] 연구소(Research Ethics Board at University Health Network)에 의해 승인된 바와 같이 기관(University Health Network Biobank)으로부터의 환자 동의와 함께 수득한 인간 결장직장 종양 조직을 콜라게나제 A를 사용하여 단일 세포로 분해하였다. 단일 세포는 4-6주령 NOD/SCID 수컷 마우스에 피하 주사하였다. 마우스는 심장 천공에 의한 혈액 수거 전 CO2 흡입시켜 안락사시키고 EDTA 튜브에 저장하였다. 수거된 혈액 샘플로부터, 혈장을 분리하고 -80℃에서 저장하였다. 무세포성 DNA는 QIAamp 순환 핵산 키트(Qiagen)를 사용하여 0.3-0.7 ml의 혈장으로부터 추출하였다. 모든 동물 연구는 위원회(Animal Care Committee at University Health Network)에 의해 승인된 윤리 규제를 준수하여 수행하였다.

[0072] *cfMeDIP-seq*

[0073] cfMeDIP-seq 프로토콜의 개략도는 WO2017/190215에 제시되어 있다. cfMeDIP 전, DNA 샘플은 카파 하이퍼 프렙 키트(Kapa Hyper Prep Kit) (Kapa Biosystems)를 사용하여 라이브러리화 제조하였다. 제조업자 프로토콜에 일부 수정을 가하여 다음과 같이 실시하였다. 간략하게, 대상 DNA를 0.2 mL PCR 튜브에 첨가하고 말단-복구 및 A-테일링을 실시하였다. 어댑터 연결은 NEBNext 어댑터(제조원: NEBNext Multiplex Oligos for Illumina kit, New England Biolabs)를 사용하여 0.181 μM의 최종 농도에서 수행하고, 20분 동안 20℃에서 인큐베이션한 다음 AMPure XP 비드로 정제하였다. 용출된 라이브러리는 USER 효소(New England Biolabs Canada)에 의해 절단한 다음 MeDIP 전에 Qiagen MinElute PCR 정제 키트를 사용하여 정제하였다.

[0074] 제조된 라이브러리를 풀링된 메틸화된/비메틸화된 λPCR 생성물과 최종 DNA 양 100 ng으로 조합하고, 일부 수정하여 Taiwo et al. 2012[7]의 프로토콜에 따라 MeDIP에 투입하였다. 간략하게, MeDIP의 경우, Diagenode MagMeDIP 키트(Cat# C02010021)를 제조업자의 프로토콜에 일부 수정을 가하여 사용하였다. 0.3 ng의 대조군 메틸화된 및 0.3 ng의 대조군 비-메틸화된 아라비도시스 탈리아나(*A. thaliana*) DNA, 충전제 DNA(DNA 총 양 [cfDNA + 충전제 + 대조군]이 100 ng이 되게 첨가) 및 완충액을 어댑터 라이게이션된 DNA가 든 PCR 튜브에 첨가한 후, 샘플을 10분 동안 95℃로 가열하고 이어서 즉시 얼음 수조에 10분간 두었다. 각각의 샘플은 2개의 0.2 mL PCR 튜브로 분할하였다: 10% 인풋 대조군을 위한 하나 및 면역침전에 적용될 샘플에 대한 다른 하나. MagMeDIP 키트에 포함된 5-mC 모노클로날 항체 33D3(Cat#C15200081)을 희석된 항체 혼합물을 제조하기 전에 1:15로 희석하고 샘플에 첨가하였다. 세척한 자기 비드(제조업자의 지침에 따라)를 17시간 동안 4℃에서 인큐베이션하기 전에 첨가하였다. 샘플은 Diagenode iPure 키트를 사용하여 정제하였으며, 50 μl의 완충액 C로 용출시켰다. 반응의 성공(QC1)은 다음 단계로 진행하기 전에 비-메틸화된 스파이크드-인 DNA <1%의 % 회수 및 >99% 반응의 % 특이성(1-[스�파이크드-인 메틸화된 대조군 DNA의 회수에 대한 스파이크드-인 비메틸화된 대조군 DNA의

회수]에 의해 계산된 바와 같이)이 보장되도록, 스파이크드-인 아라비둡시스 탈리아나 DNA의 존재를 검출하기 위해 qPCR을 통해 검증하였다. 각각의 라이브러리를 증폭시키기 위한 최적의 사이클 수는 qPCR의 사용을 통해 결정하였고, 이후에 샘플은 카파 HiFi 핫스타트 마스터믹스(KAPA HiFi Hotstart Mastermix)를 사용하여 증폭시키고 NEBNext 멀티플렉스 올리고스를 0.3 μ M의 최종 농도로 첨가하였다. 라이브러리를 증폭시키기 위해 사용되는 PCR 세팅은 다음과 같았다: 3분 동안 95°C에서 활성화, 이어서 98°C에서 20초 동안, 65°C에서 15초 동안 및 72°C에서 30초 동안 미리 결정된 사이클의 적용 및 72°C에서 1분 동안 최종 연장. 증폭된 라이브러리는, MinElute PCR 정제 컬럼에 이어서 임의의 어댑터 이량체를 제거하기 위해 3% Nusieve GTG 아가로스 겔과 함께 선택된 겔 크기를 사용하여 정제하였다. 서열분석을 수행하기 전에, 메틸화된 인간 DNA 영역(고환-특이적 H2B, TSH2B) 및 비메틸화된 인간 DNA 영역(GAPDH 프로모터)의 배수 농화를, 무세포성 DNA를 모방하기 위해 전단된 HCT116 세포주 DNA(ATCC로부터 수득된 세포주, 마이크로플라스마 부재)로부터 생성된 MeDIP-seq 및 cfMeDIP-seq 라이브러리에 대해 결정하였다. 최종 라이브러리는 UHN Princess Margaret Genomic Centre에서 Illumina HiSeq 2000로 서열분석하기 전에 BioAnalyzer 분석에 투입하였다.

[0075] 점 돌연변이 검출을 위한 초-심층 표적화된 서열분석

[0076] 본원 발명자들은 기관(Princess Margaret Cancer Centre)에서 초기 단계 임상 시험에서 입회 전 생성된 종양 조직 분자 프로파일링 데이터와 일치하는 환자로부터의 ~20 mL의 혈장(4-5x 10 mL EDTA 혈액 튜브)으로부터 무세포성 DNA를 단리하기 위해 QIAgen 순환 핵산 키트를 사용하였다. DNA는 PureGene Gentra 키트를 사용하여 세포주(CRC 및 MM 세포주의 희석)로부터 추출하고, Covaris 초음파분쇄기를 사용하여 ~180 bp로 단편화시키고, 보다 큰 크기의 단편은 무세포성 DNA의 단편 크기를 모방하기 위해 Ampure 비드를 사용하여 배제하였다. DNA 서열분석 라이브러리는 NEXTflex-96 DNA 바코드 어댑터(Bio Scientific, Austin, TX) 어댑터를 활용하는 카파 하이퍼프랩 키트(KAPA Hyper Prep Kit) (Kapa Biosystems, Wilmington, MA)를 사용하여 83 ng의 단편화된 DNA로부터 작제하였다. 공지된 돌연변이를 함유하는 DNA 단편을 단리하기 위해, 본원 발명자들은 일루미나 TruSeq 앰플리콘 암 패널(Illumina TruSeq Amplicon Cancer Panel)을 사용한 임상 연구소에 의해 시험된 48개 유전자로부터의 돌연변이 핫스팟을 표적화하는 비오틀화된 DNA 포획 프로브(xGen Lockdown Custom Probes Mini Pool, Integrated DNA Technologies, Coralville, IA)를 디자인하였다. 바코드화된 라이브러리를 풀링하고 이어서 제조업자의 지침에 따라 커스텀 하이브리드 포획 라이브러리(IDT xGEN Lockdown protocol version 2.1)를 적용하였다. 이들 단편은 일루미나(Illumina) HiSeq 2000 장비를 사용하여 >10,000X 판독 커버리지까지 서열분석하였다. 수득한 판독은 bwa-mem을 사용하여 정렬하고 돌연변이는 samtools 및 muTect 버전 1.1.4를 사용하여 검출하였다.

[0077] 심층 서열분석에 의한 종양-특이적 특성의 수와 검출 가능성 간의 모델링 관계

[0078] 본원 발명자들은 암-특이적 메틸화된 DMR의 비율이 0.001%, 0.01%, 0.1%, 1%, 및 10%로 설정되고 각각 1, 10, 100, 1000 및 10000개의 독립적 DMR로 이루어진 145,000개 시뮬레이션된 게놈을 생성하였다. 본원 발명자들은 이들 본래의 혼합물로부터 14,500개 이배체 게놈(100 ng의 DNA를 나타내는)을 샘플링하였고 추가로 이들 심층 정도로 서열분석 커버리지를 나타내기 위해 유전자좌 당 10, 100, 1000, 및 10000개의 판독을 샘플링하였다. 이러한 공정은 커버리지, 풍부성 및 특성의 수의 각각의 조합을 위해 100회 반복하였다. 본원 발명자들은 파라미터의 각각의 조합을 위해 적어도 1개의 DMR의 성공적인 검출의 빈도를 추정하고 심층 서열분석 상에서 조건적인 성공적 검출 확률에 대한 특성의 수의 영향을 가시적으로 평가하기 위해 확률 곡선(도 1a)을 플롯팅하였다.

[0079] 조직-특유의 특성의 유도, 다중-조직 분류기의 개발 및 450k 데이터에서 검증

[0080] cfDNA MeDIP 프로파일은 MEDIPS R 패키지[8]를 사용하여 정량하였고, RPKM으로 전환시키고, 이후 로그 2 cpm(counts-per-million)으로 변환시켰다. 후속적으로, 선형 모델은 FANTOM5 인핸서, CpG 아일랜드, CpG 쇼어 및 CpG 셀프에 맵핑된 특성의 매트릭스 상에서 리마-트렌드(limma-trend)[9]를 사용하여 피팅하였고, 회수된 스파이크-인 메틸화된 DNA의 백분율은 풀다운(pulldown) 효율 변화에 대해 제어하기 위한 공변수로서 포함된다. 쌍형성 대조는 각각의 조직 유형의 쌍에 대해 평가하였고 상부 150 및 하부 150 DMR은 탄력적 네트 분류기 트레이닝 및 암-유형 특이성의 확증을 위해 선택되었다. 수행능 지표는 문헌(Chakravarthy et al[10])에서 이전에 사용된 학습적 교차-확증에서 최고 카파 값을 갖는 모델로부터 아웃-오브-폴드 콜(out-of-fold call)에 대한 다수 부류의 투표에 의해 유래하였다.

[0081] 분류 정확도의 평가를 위한 기계 학습 분석

[0082] 디스커버리 코호트에 대한 모델 트레이닝 및 평가

- [0083] 높은 컴퓨터 계산 비용 없이 중양 분류에서 cfMeDIP 데이터의 수행능을 평가하기 위해, 본원 발명자들은 초기 세트의 가능한 후보물 특성을 CpG 아일랜드, 쇼어, 셀프 및 FANTOM5 인핸서(이에 의해 "조절 특성"으로 표지된)를 포괄하는 윈도우로 감소시켜 196개 샘플 및 505,027개 특성들의 매트릭스를 획득하였다. 이어서, 본원 발명자들은 카렛(caret) R 패키지를 디스커버리 코호트 데이터를 80% 내지 20% 방식으로 50개 독립적 트레이닝 및 시험 세트로 분할하기 위해 사용하였다(도 2a). 디스커버리 코호트에 걸친 부류 비율을 유지하면서 분할을 수행하였다. 이어서, 본원 발명자들은 각각의 부류 대 다른 부류에 대한 리마-트렌드를 사용하여 트레이닝 데이터 분할에 대한 중간 정도의 t-통계(150개 과메틸화된, 150개 저메틸화된)에 의해 상위 300개 DMR을 선택하였다. 이항 GLMnet는 이어서 수행능 지표로서 코헨(Cohen) 카파를 사용하여 혼합 파라미터(알파, 값 = 0, 0.2, 0.5, 0.8 및 1) 및 페널티(람다, 값 = 0.01 증분으로 0 - 0.05)의 값을 최적화하기 위해 10-배-교차-확증(CV)의 3 반복의 사용과 함께 이들 DMR(최대 300 DMR x 7 다른 부류 = 2100 특성)을 사용하여 트레이닝하였다. 각각의 트레이닝 세트에 대해, 이것은 6 1개-부류 대-다른-부류 이항 분류기의 수집물을 산출하였다.
- [0084] 본원 발명자들은 이어서 AUROC(수용자 작동 특징 곡선 이하 면적)를 사용하여 보류(held-out) 시험 세트 상에서 분류 수행능을 추정하였다. 이들 추정치는 상기 보류 시험 세트 샘플이 DRM 사전 선택 또는 GLMet 트레이닝 및 튜닝을 위해 사용되지 않기 때문에 분류의 비편향된 측정을 나타낸다. 50개 독립적 트레이닝 및 시험 세트는 또한 트레이닝-세트 편향으로 인한 낙관적 추정의 최소화를 가능하게 하였다.
- [0085] *확증 코호트에 대한 모델 평가*
- [0086] 각각의 확증 코호트 cfMeDIP 샘플에 대해, 본원 발명자들은 디스커버리 코호트 내 50개 상이한 트레이닝 세트 상에서 트레이닝된 AML, LUC 및 정상 1개-대-모든 이항 분류기에 대한 부류 확률을 추정하였다. 50개 모델로부터의 확률은 평균화하여 AUROC 추정을 위해 이어서 사용되는 단일 스코어를 생성하였다. 본원 발명자들은 또한 초기(단계 I 및 II) 또는 후기 단계 LUC 샘플(단계 III 및 IV)이 1개-대-모든 분류기에 대해 제외된 경우, 질환 단계가 AUROC를 추정함에 의해 수행능에 영향을 미치는지를 평가하였다.
- [0087] **결과 및 논의**
- [0088] 본원 발명자들은 생물정보학적으로 0.001%로부터 10%까지 ctDNA의 상이한 비율을 갖는 혼합물을 시뮬레이션하였다(도 1a, 컬럼 측면). 본원 발명자들은 또한 정상 cfDNA와 비교하여 ctDNA가 1, 10, 100, 1000, 또는 10000 DMR을 갖는 경우의 시나리오를 시뮬레이션하였다(도 1a, 열 측면). 이어서, 판독은 각각의 유전자좌 (10X, 100X, 1000X, 및 10000X)에서 심층의 다양한 서열분석에서 샘플링하였다(도 1a, x-축). 본원 발명자들은 암 ctDNA의 낮은 풍부성 및 얇은 커버리지에서도 DMR의 수가 증가하기 때문에 적어도 하나의 암-특이적 이벤트의 검출 확률이 증가함을 밝혔다.
- [0089] 더욱이, 기관(Cancer Genome Atlas (TCGA))로부터의 범-암 데이터는 실제로 모든 종양 유형에 걸친 종양과 정상 조직 간에 다수의 DMR을 보여주었다[11]. 따라서, 이러한 발견은 ctDNA로부터 암-특이적 DNA 메틸화 변화를 성공적으로 회복한 검정이 낮은 서열분석 관련 비용으로 악성 질환을 검출하고, 분류하고 모니터링하기 위해 매우 민감한 도구로서 작용할 수 있음을 강조한다.
- [0090] 그러나, 혈장 cfDNA에서 DNA 메틸화의 게놈-전체 맵핑은 순환계에서 DNA의 매우 낮은 양 및 단편화로 인해 도전 과제이다[12]. 결과로서, cfDNA의 메틸화 프로파일링에서 이전의 노력은 주로 결장직장 암 스크리닝을 위한 FDA 승인된 SEPT9 메틸화 검정과 같은, 유전자좌 특이적 PCR-기반 검정[2,3]으로 제한되었다[13]. 최근 노력은 단편화된 cfDNA의 전체-게놈 바이셀파이트-서열분석을 수행하기 위해 만들어졌지만[14-16], CpG의 낮은 게놈-전체 풍부성은 서열분석으로부터 가용한 유용한 메틸화-관련된 정보의 양을 감소시킬 가능성이 있다. 따라서, 혈장 DNA 상에서 WGBS를 사용한 주요 문제는 바이셀파이트 전환과 연관된 고비용, 낮은 효율 및 DNA 손실이다. 한편, 메틸화 경향이 있는 CpG-풍부 특성을 위해 선택적으로 집적시키는 방법은 판독당 가용한 유용한 정보 양을 최대화하고 DNA 손실을 감소시킬 가능성이 있다.
- [0091] *cfDNA 메틸화 맵핑을 위해 적합한 게놈-전체 방법*
- [0092] 본원 발명자들은 무세포성 DNA를 사용한 게놈-전체 DNA 메틸화 맵핑을 수행하기 위해 cfMeDIP-seq(메틸화된 무세포성 DNA 면역침전 및 고처리량 서열분석)로 호칭되는 새로운 방법을 개발하였다. 본원에 기재된 cfMeDIP-seq 방법은 본원 발명자들의 경험에서 100 ng의 인풋 DNA 아래에서 매우 강한 기존의 낮은 인풋 MeDIP-seq 프로토콜의 변형을 통해 개발하였다[7]. 그러나, 다수의 혈장 샘플은 100 ng 훨씬 미만의 DNA를 산출한다. 상기 도전 과제를 극복하기 위해, 본원 발명자들은 출발 DNA의 양을 100 ng으로 인위적으로 부풀리기 위해 외인성 λDNA(충전제 DNA)를 어댑터-연결된 cfDNA 라이브러리에 첨가하였다. 이것은 항체에 의한 비-특이적 결합의 양을 최소화

하고 또한 플라스틱웨어로의 결합으로 인해 상실된 DNA 양을 최소화한다. 충전제 DNA는 어댑터-연결된 cfDNA 라이브러리 크기가 유사한 앰플리콘으로 이루어져 있고 상이한 CpG 밀도에서 비메틸화되고 시험관내 메틸화된 DNA로 구성되었다. 상기 충전제 DNA의 첨가는 또한 상이한 환자들이 상이한 양의 cfDNA를 산출하기 때문에 실제 용도로서 작용하여 인풋 DNA 양의 100 ng으로의 정상화를 가능하게 한다. 이것은 다운스트림 프로토콜이 가용한 cfDNA의 양과는 상관 없이 모든 샘플에 대해 정확하게 동일하게 남아있도록 보장한다.

[0093] 본원 발명자들은 먼저 cfDNA에서 관찰된 것과 유사한 단편 크기로 전단된, 인간 결장직장 세포주 HCT116으로부터의 DNA를 사용하여 cfMeDIP-seq 프로토콜을 검증하였다. HCT116은 공개된 DNA 메틸화 데이터의 가용성 때문에 선택되었다. 본원 발명자들은 동시에 100 ng의 전단된 세포주 DNA를 사용한 골드 표준 MeDIP-seq 프로토콜[7] 및 10 ng, 5 ng 및 1 ng의 동일하게 전단된 세포주 DNA를 사용한 cfMeDIP-seq 프로토콜을 수행하였다. 이것은 2개의 생물학적 레플리케이트에서 수행하였다. 모든 조건에 대해, 본원 발명자들은 반응의 99% 초과 특이성(1-[스파이크드-인 메틸화된 대조군 DNA의 회수에 대한 스팅크드-인 비메틸화된 대조군 DNA의 회수]), 및 비메틸화된 영역에 대한 공지된 메틸화된 영역의 매우 높은 집적(각각, TSH2B0 및 GAPDH)을 획득하였다(도 6b).

[0094] 라이브러리를 라이브러리 당 약 3000만 내지 7000만 관독으로 포화(도 6a)까지 서열분석하였다(보충 표 1). 미가공 관독은 인간 게놈 및 λ 게놈 둘 다에 정렬하고 실제로 λ 게놈으로 정렬되지 않는 것으로 밝혀졌다(보충 표 1). 따라서, 충전제 DNA로서 외인성 λ DNA의 첨가는 서열분석 데이터의 생성물 방해하지 않았다. 최종적으로, 본원 발명자들은 면역침전 단계를 위한 품질 관리 척도로서 CpG 집적 스코어를 계산한다[8]. 모든 라이브러리는 CpG에 대해 유사한 집적을 보여주었고 예상된 바와 같이 상기 인풋 대조군은 어떠한 집적을 보여주지 않았고(도 6c), 이는 심지어 극히 낮은 인풋(1ng)에서도 본원 발명의 면역침전을 확장한다.

[0095] 상이한 인풋 DNA 수준을 비교하는 게놈-전체 상호관계 추정치는 MeDIP-seq(100 ng) 및 cfMeDIP-seq(10, 5, 및 1 ng) 방법 둘 다 매우 강력함을 보여주고, 임의의 2개의 생물학적 레플리케이트 간에 피어슨 상호관계는 적어도 0.94이다(도 1b). 상기 분석은 또한 5 및 10 ng의 인풋 DNA에서 cfMeDIP-seq가 100 ng에서 통상의 MeDIP-seq(적어도 0.9의 쌍형성 피어슨 상호관계)에 의해 획득된 메틸화 프로파일을 강하게 개괄할 수 있음을 입증한다(도 1b). 1 ng의 인풋 DNA에서 cfMeDIP-seq의 수행능은 100 ng에서 MeDIP-seq와 비교하여 감소하지만 >0.7에서 강한 피어슨 상호관계를 여전히 보여준다(도 1b). 본원 발명자들은 또한 골드 표준 RRBS(감소된 대표 바이셀 파이트 서열 분석) 및 WGBS(전체-게놈 바이셀피트 서열 분석)를 사용하여 HCT116의 DNA 메틸화 프로파일을 개괄함을 관찰하였다(도 1c). 함께, 본원 발명의 데이터는 cfMeDIP-seq가 순환 cfDNA와 같이 단편화되고 낮은 인풋 DNA 물질의 게놈-전체 메틸화 맵핑을 위해 강력한 프로토콜임을 시사한다.

[0096] cfMeDIP-seq는 종양-유래된 ctDNA의 검출을 위해 높은-민감성을 나타낸다

[0097] cfMeDIP-seq 프로토콜의 민감성을 평가하기 위해, 본원 발명자들은 결장직장 암(CRC) HCT116 세포주 DNA의 다중 골수종(MM) MM1.S 세포주 DNA로의 연속 희석을 수행하였고, 둘 다 cfDNA 크기를 모방하도록 전단되었다. 본원 발명자들은 CRC DNA를 100%, 10%, 1%, 0.1%, 0.01%, 0.001%로부터 0%로 희석하였고 이들 희석 각각에 대해 cfMeDIP-seq를 수행하였다. 본원 발명자들은 또한 동일한 샘플에서 3개의 점 돌연변이의 검출을 위해 초-심층(10,000X 메디안 커버리지) 표적화된 서열분석을 수행하였다. 5% 거짓 디스커버리 레이트(FDR) 역치를 사용한 순수 MM DNA에 대해 각각의 CRC 희석 점에서 동정된 DMR의 관찰된 수는 거의 완벽하게 선형이고($r^2=0.99$, $p<0.0001$) 상기 DMR의 예상된 수는 0.001% 희석으로 하강된 희석 인자 (도 1d)를 기준으로 한다. 더욱이, 이들 DMR 내에서 DNA 메틸화 신호는 또한 관찰된 것 대 예상된 신호 간에 거의 완벽한 선형성($r^2=0.99$, $p<0.0001$)을 보여준다(도 1e; 보충 표 2B). 이에 비해, 1% 희석을 초과하여, 초-심층 표적화된 서열분석은 CRC-특이적 변이체와 PCR 또는 서열분석-오류로 인한 거짓의 변이체 간을 용이하게 구분할 수 없다(도 1f; 보충 표 2A). 따라서, cfMeDIP-seq는 암-유래된 DNA의 검출을 위해 우수한 민감성을 나타내고, 표준 프로토콜을 사용한 초-심층 표적화된 서열분석에 의해 변이체 검출 수행능을 초과한다.

[0098] 암 DNA는 흔히 CpG-풍부 영역에서 과메틸화된다[17]. cfMeDIP-seq는 메틸화된 CpG-풍부 서열을 특이적으로 표적화하기 때문에, 본원 발명자들은 ctDNA가 면역침전 과정 동안에 우선적으로 집적될 수 있음을 추론하였다. 이를 시험하기 위해, 본원 발명자들은 2명의 결장직장 암 환자로부터 환자-유래된 이중이식체(PDX)를 생성하였고 마우스 혈장을 수거하였다. 종양-유래된 인간 cfDNA는 총 cfDNA 풀 내에서 1% 미만의 빈도로 그리고 면역침전 후 2배 보다 큰 풍부성으로 존재하였다(도 1g; 보충 표 3). 이들 결과는 ctDNA의 편향된 서열 분석을 통해 cfMeDIP 과정이 ctDNA 검출 민감성을 추가로 증가시킬 수 있음을 시사한다.

[0099] 순환 혈장 cfDNA 메틸화 프로파일은 다중 암 타입과 건강한 공여자 간을 구분할 수 있다

- [0100] DNA 메틸화 패턴은 조직-특이적이고 암 환자들을 많은 다른 암 타입 중에서 교모세포종[18], 뇌실막세포종[6], 결장직장[19], 및 유방[20,21]에서의 임상적으로 적절한 질환 서브그룹으로 계층화하기 위해 사용되어 왔다. 본원 발명자들은 cfDNA 연관된 프로파일이 다중 종양 유형에 대해 조직 기원을 동정하기 위해 사용될 수 있는지를 물었다. 이를 위해, 본원 발명자들은 5개 상이한 종양 유형으로부터의 196개 샘플 및 초기 및 후기 단계 종양으로부터의 정상 대조군을 프로파일링하였다. 본원 발명자들은 각각의 쌍형성 비교를 위해 CpG 쇼어, 셀프, 아일랜드 및 FANTOM5 인헨서에 맵핑하는 상위 300개 DMR을 동정하기 위한 선형 모델링을 사용하여 총 2,100 특유한 DMR을 유도하였다(도 2a). 이들 특성들의 메틸화 상태를 기반으로 196개 혈장 샘플의 t-분포된 확률적 임베딩(t-distributed Stochastic Neighbor Embedding) (tSNE)[22]을 기반으로 하는 밀도 클러스터링은 조직 기원 및 종양 유형을 기준으로 샘플의 특유한 클러스터링을 밝혔다(도 2b,c). 이들 특성들을 갖는 탄력적 넷 다중-암 분류기를 사용하여(도 2a), 본원 발명자들은 상이한 종양 유형 간의 고도로 정확한 식별을 관찰하였다(도 2d).
- [0101] *질환 서브타입의 식별*
- [0102] 본원 발명자들은 5개 특유의 경우 - 유전자 발현 패턴(유방암에서 ER 상태), 카피수 비정상(유방암에서 HER2 상태), 재배열(AML에서 FLT3 ITD 상태), 점 돌연변이(GBM에서 IDH 돌연변이), 및 폐암에서 최종 조직학에서 질환 서브타입 간을 식별하는 cfDNA MeDIP 프로파일의 능력을 평가하였다. 각각의 경우에, 선형 모델을 사용하여 처음에 기재된 바와 같은 특성들을 선택하고 랭킹하였다. 각각의 경우에, 계층 클러스터링을 사용하여 샘플의 분류를 평가하였다. 선택된 특성들의 메틸화 상태를 기반으로 하는 t-분포된 확률적 임베딩(tSNE)[22]을 기반으로 하는 밀도 클러스터링은 암 서브타입 분류의 이들 5개의 특유한 예의 각각에 기초하여 샘플의 특유한 클러스터링을 밝혔다.
- [0103] *암의 검출 및 기계 학습을 사용한 암 타입의 분류*
- [0104] 암을 검출하고 추가로 암 타입을 분류하는 cfMeDIP 프로파일의 능력을 왕성하게 평가하기 위해, 본원 발명자들은 이어서 본원 발명의 디스커버리 코호트에 대한 기계 학습 분석 세트를 수행하였다. 가속화된 컴퓨터 계산 분석을 가능하게 하기 위해, 본원 발명자들은 처음에 본원 발명의 cfMeDIP 디스커버리 코호트를 CpG 아일랜드, 쇼어, 셀프 및 FANTOM5 인헨서(n=505,027 윈도우)에 맵핑하는 특성들로 감소시켰다. 본원 발명자들은 이어서 본 발명의 디스커버리 코호트 샘플에 대한 전략을 수행하여 트레이닝-세트 편향을 설명하면서, 수행능의 비편향된 추정치를 유도하였다.
- [0105] 여기서, 본원 발명자들은 디스커버리 코호트를 균형조절된 트레이닝 및 시험 세트(80% 트레이닝 세트, 20% 시험 세트)로 분할하였다. 트레이닝 세트에서 샘플만을 사용하여, 본원 발명자들은 리마-트렌드 시험 통계를 기준으로, 다른 부류들에 대해 각각의 부류(샘플 유형)에 대한 상위 300 DMR을 선택하였고 트레이닝 세트 데이터 상에서 이들 특성들을 사용하는 일련의 1개-대-다른-부류 GLMnet를 트레이닝하였다. 트레이닝 과정은 코헨 카파의 최적화와 함께 알파 및 람다에 대한 값들의 그리드에 걸쳐 3라운드의 10배 교차 검증(CV)으로 이루어져 있다. 다중 라운드의 10-배 CV의 사용은 보다 일반화될 수 있는 모델 튜닝을 위해 추가의 무작위화를 활용하기 위한 열망에 의해 동기 부여되었다.
- [0106] 수행능은 이어서 시험 세트 샘플(DMR 선택 및 후속적 GLMnet 트레이닝/튜닝 단계 동안에 보류된)로부터 유래된 AUROC(수용자 작동 특징적 곡선 이하 면적)을 사용하여 평가하였다. 이러한 공정은 트레이닝-세트 편향의 영향을 완화시키기 위해 디스커버리 코호트를 트레이닝 및 시험 세트로의 50개 상이한 분할과 함께 반복하였다. 이것은 각각 1개-대-다른-부류 비교(총 480개 모델)에 대한 50개 모델의 수거에서 정점에 도달하였다. 이로써, 본원 발명자들은 E50으로서 모델의 상기 수거물을 참조한다.
- [0107] 후속적으로, 본원 발명자들은 추가의 152개 혈장 샘플의 검증 코호트를 생성시킴에 의해 배치에 걸친 수행능을 평가하였다: AML(n=35), 폐암(n=55) 및 건강한 대조군(n=62) 샘플. 각각의 부류에 대해, 본원 발명자들은 E50에서 모델에 의한 부류 확률 아웃풋을 평균화하였고 하나의 부류 대 모든 다른 부류들에 대해 AUROC를 추정하였다(도 3a). 분류기는 AML 대 기타(0.993), LUC 대 기타(0.943) 및 정상 대 기타(1.000)에 대한 높은 AUROC 값을 보여주었다. 이것은 추가로 종양 유형을 정확하게 검출하고 분류하는 기계 학습 접근법과 커플링된 cfMeDIP-seq의 능력을 확인시켜주었다. 최종적으로, 본원 발명자들은 분류기가 후기 단계 샘플(0.934)에서와 같이 초기 단계 샘플(0.950)에서 정확함을 관찰하였고(도 3b), 이는 상기 접근법이 암 초기 검출을 위해서 및 초기 단계 및 후기 단계 둘 다에서 암의 검출을 위해 적용될 수 있음을 시사하였다.
- [0108] *cfMeDIP-seq를 사용한 cfDNA 메틸화 프로파일링의 추가의 이점*
- [0109] 조직 기원을 정확하게 나타내는 cfDNA 메틸화 패턴의 능력은 또한 돌연변이 기반 검정의 한계를 극복하고, 여기

서, 조직 기원에 대한 특이성은 상이한 조직에서 암에 걸친 많은 잠재적 구동자 돌연변이의 재발 특성으로 인해 낮을 수 있다[23]. 돌연변이 기반 검정은 또한 종양의 클론 구조에 의해 비민감성이 될 수 있고, 여기서, 서브클론 구동자는 ctDNA의 낮은 풍부성으로 인해 검출하기가 더욱 어려울 수 있다[24]. 돌연변이 기반 ctDNA 접근법은 또한 양성(positive) 선택의 증거를 나타내는 것으로 관찰되고[25] 보고된[26] 양성(benign) 조직에서 구동자 돌연변이에 의한 잠재적 교란(confounding)에 취약할 수 있다.

[0110] 이를 종합해보면, 지금까지 유래된 암 cfDNA 메틸롬의 최대 수거물을 기반으로 하는 본 발명의 발견은 암의 관리 및 조기 검출에 영향을 미치는 잠재력을 갖는 효율적이고 저렴한 도구로서 cfMeDIP-seq를 확립한다. cfMeDIP-seq의 정확도 및 다재다능은 내성이 전립선 암에서 안드로겐 수용체 억제에 대한 민감성과 같은 후생적 변화와 상호관련된 세팅에서 치료학적 결정을 고지하는데 유용할 수 있다[27]. 조기 진단 및 스크리닝을 위한 잠재적 기회는 특히 스크리닝이 이미 임상적 용도를 보여주었지만 이에 대한 기존의 스크리닝 시험(즉, 낮은 선량의 CT 스캐닝)은 이온화 방사선 노출 및 높은 거짓 양성율과 같은 유의적 한계를 갖는 질환인 폐암에서 명확할 수 있다.

[0111] 결론적으로, 본 발명의 발견은 비-침습성, 저렴한, 민감한, 고도로 정확한 조기 종양 검출, 다중-암 분류 및 암 서브타입 분류를 위한 기초로서 cfDNA 메틸화 프로파일의 용도를 강조한다.

[0112] 본 발명의 바람직한 구현예가 본원에 기재되었지만, 당업자는 발명의 취지 또는 첨부된 청구항의 범위를 벗어나는 것 없이 변화가 여기에 이루어질 수 있다는 것을 이해할 것이다. 하기의 참조 목록에 있는 것들을 포함하는 본원에 기재된 모든 문헌들은 본원에 참조로 포함된다.

[0113] **표 1:** 인간(Hg19) 게놈 및 λ 게놈에 대한 cfDNA를 모방하도록 전단된 HCT116 세포주 DNA의 다양한 출발 인풋을 사용하여 제조된 서열분석된 MeDIP-seq(100 ng Rep1 및 Rep2) 및 cfMeDIP-seq(10 ng, 5 ng 및 1 ng, Rep1 및 Rep2) 라이브러리의 판독 수 및 맵핑 효율. 2개의 생물학적 레플리케이트는 출발 인풋 DNA에 대해 사용되었다. 100 ng 미만의 출발 인풋은, 샘플의 MeDIP 전에 출발 양을 100 ng으로 인위적으로 높이기 위해 외인성 λ DNA를 첨가하였다.

샘플	미가공 판독 #	인간 게놈에 대한 정렬된 판독 # (Hg19)	인간 게놈에 대한 맵핑 효율 (Hg19)	λ 게놈에 대한 정렬된 판독 #	λ 게놈으로의 맵핑 효율
인풋	74,504,053	71,343,168	95.76	12	0.00
100 ng 레플리케이트 1	55,396,238	50,472,273	91.11	0	0.00
100 ng 레플리케이트 2	66,569,209	60,770,277	91.29	1	0.00
10 ng 레플리케이트 1	70,054,607	64,020,441	91.39	0	0.00
10 ng 레플리케이트 2	58,297,539	53,308,777	91.44	0	0.00
5 ng 레플리케이트 1	65,845,430	60,540,743	91.94	1	0.00
5 ng 레플리케이트 2	64,750,879	59,358,412	91.67	0	0.00
1 ng 레플리케이트 1	35,102,361	32,258,451	91.90	0	0.00
1 ng 레플리케이트 2	33,881,118	31,194,711	92.07	0	0.00

[0114] **표 2A:** CRC 세포주 HCT116 DNA의 MM 세포주 MM1.S DNA로의 연속 회석을 사용한 초-심층 표적화된 변이체 서열분석의 평균 커버리지

희석 (CRC DNA의 %)	붕괴되지 않은 판독은 표적 커버리지를 의미함	SSCS (단일 가닥 컨센서스 서열)은 표적 커버리지를 의미함	DCS (듀플렉스 컨센서스 서열)은 표적 커버리지를 의미함
100	155,964	4284	655
10	154,657	4877	654
1	154,419	4890	654
0.1	183,271	5674	887
0.01	238,291	8068	1602
0.001	199,766	7337	1299
0.0001	187,695	6891	1181
0	216,434	7721	1412

[0116]

[0117] 표 2B: CRC 세포주 HCT116 DNA의 MM 세포주 MM1.S DNA로의 연속 희석으로부터 수득한 관찰된 DMR 및 DNA 메틸화 신호

희석 (CRC DNA의 %)	관찰된 DMR의 수	관찰된 DNA 메틸화 신호 (DMR 내 RPKM의 합)
100	111,472	645,683.90
10	1,597	8,775.61
1	692	4,521.60
0.1	12	75.71
0.01	8	79.73
0.001	2	22.42

[0118]

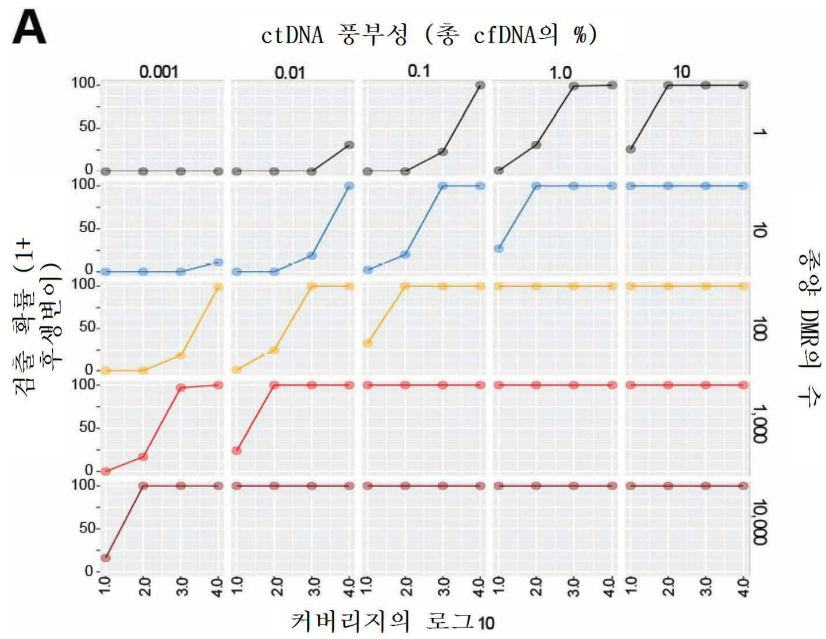
[0119] 표 3: 인간(Hg19) 게놈으로의 정렬 후 PDX 및 인풋 대조군 샘플의 cfMeDIP.seq 라이브러리의 판독 수 및 맵핑 효율

샘플	미가공 판독 #	인간 게놈에 대한 정렬된 판독 # (Hg19)	인간 게놈에 대한 맵핑 효율 (Hg19)
인풋 대조군 1	46,857,633	389,073	0.83
인풋 대조군 2	35,658,454	283,799	0.80
PDX 1	49,997,949	1,080,277	2.16
PDX 2	34,802,767	614,988	1.77

[0120]

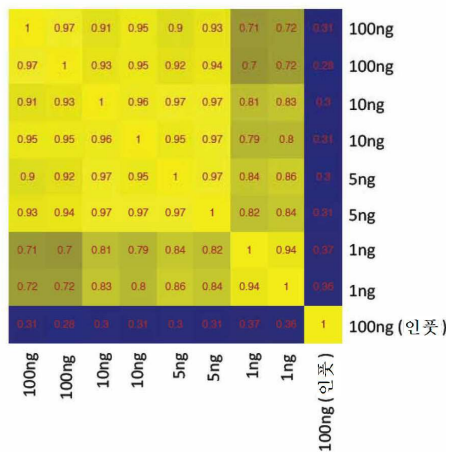
도면

도면1a



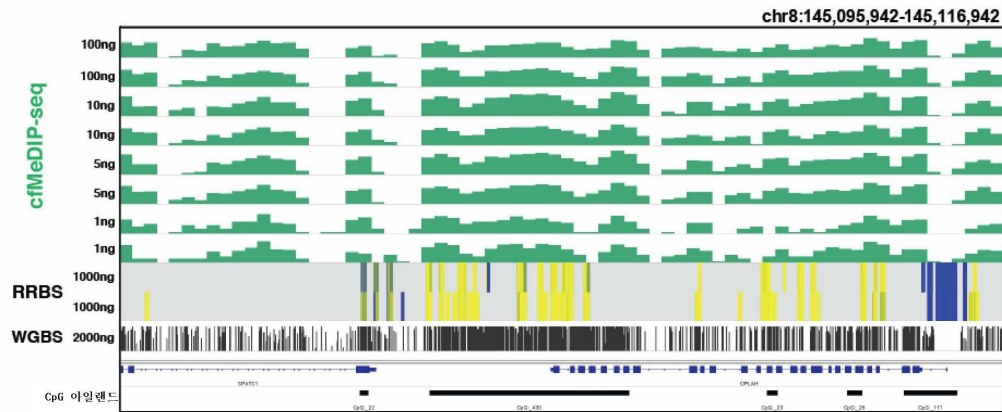
도면1b

B



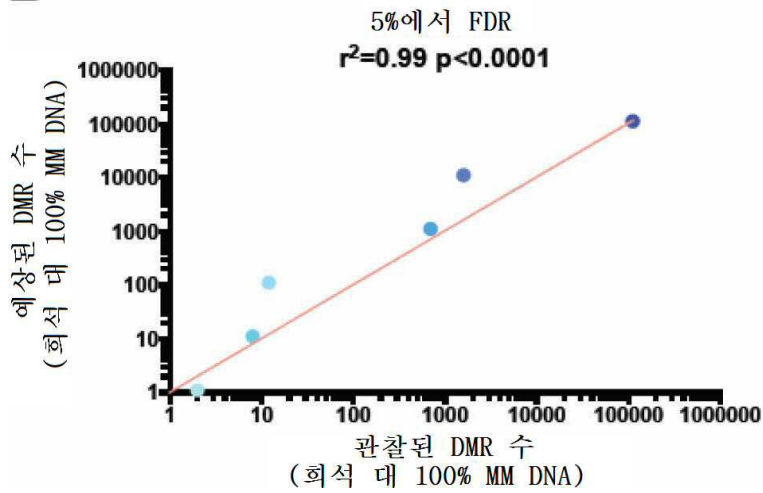
도면1c

C



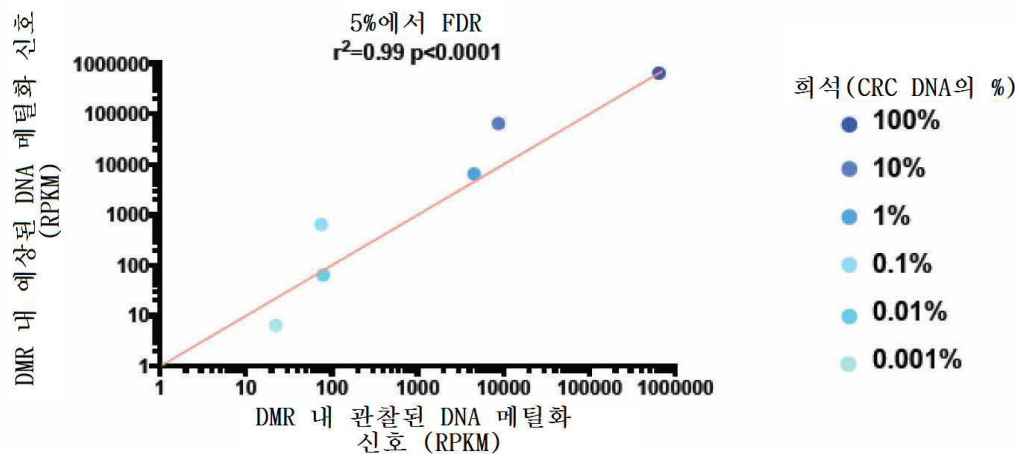
도면1d

D



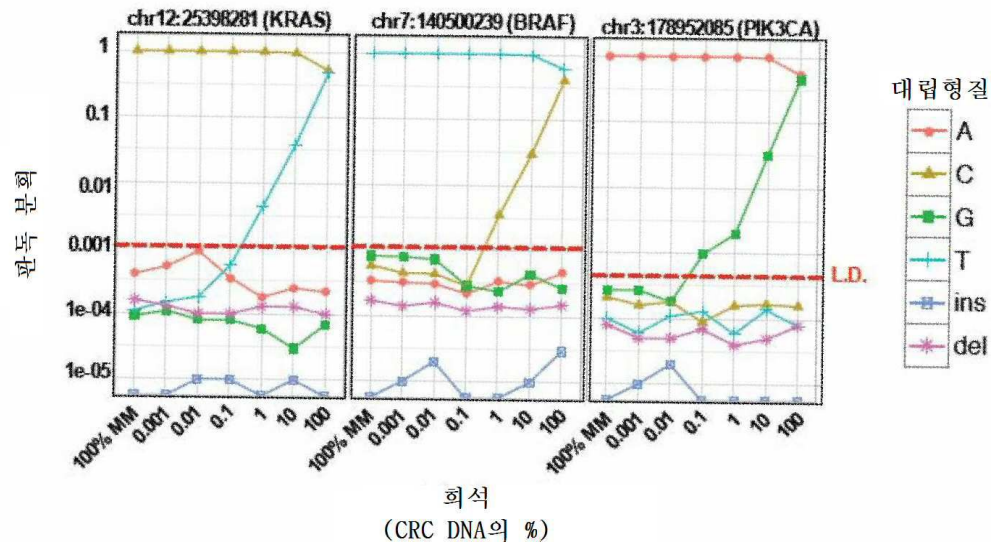
도면1e

E



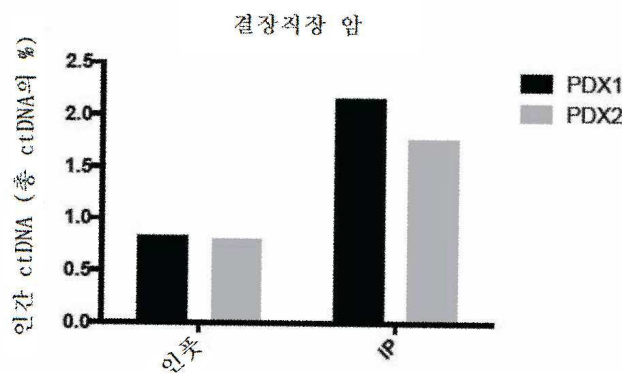
도면1f

F

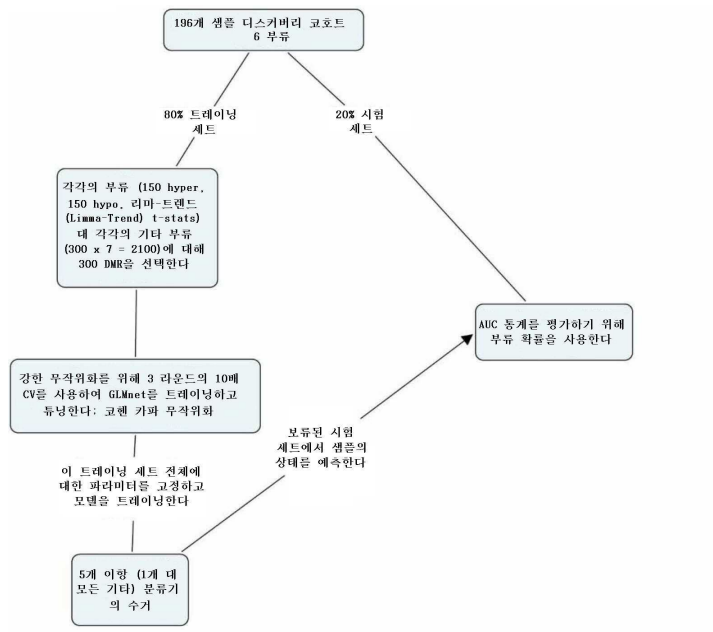


도면1g

G

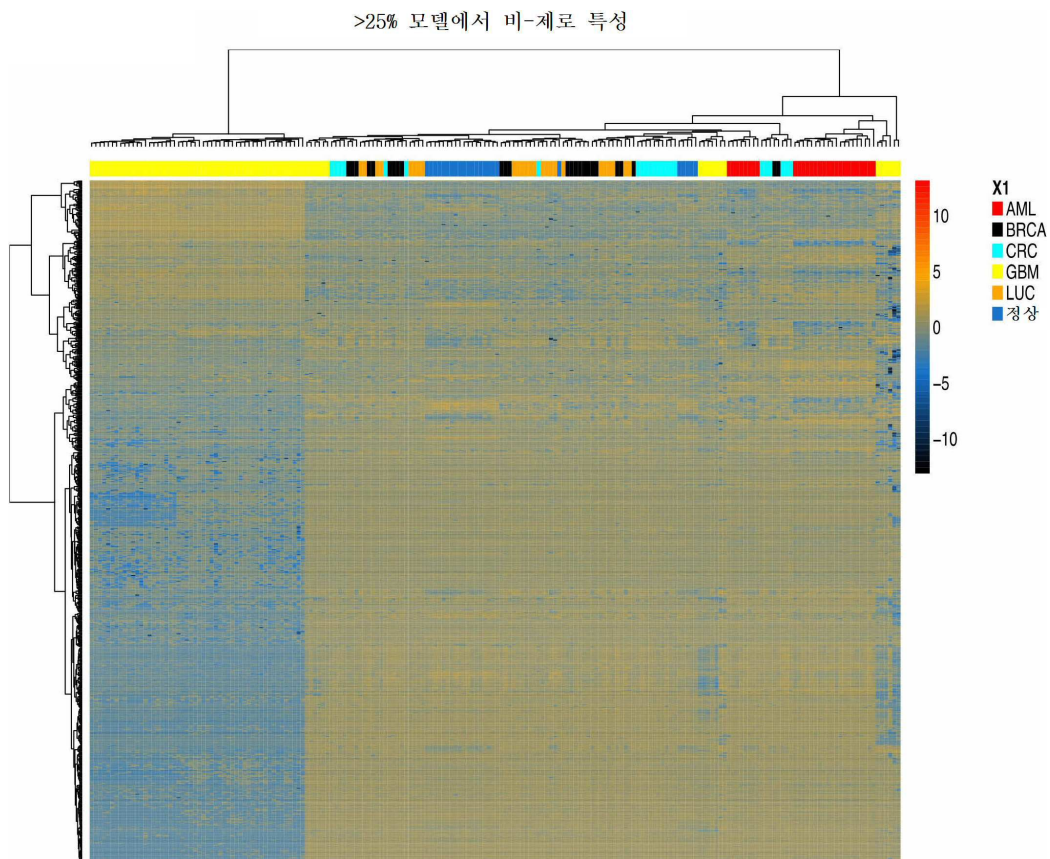


도면2a

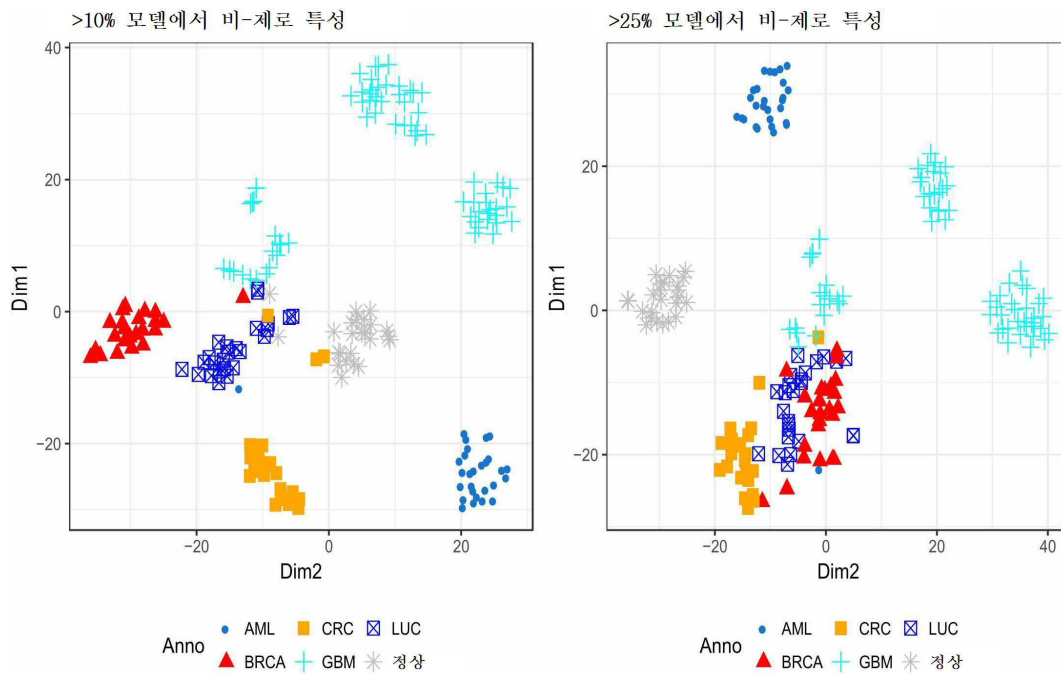


디스커버리 코호트의
트레이닝 및 시험 세트
로의 상이한 부류-균형
분할과 함께 공정을 50회
반복한다

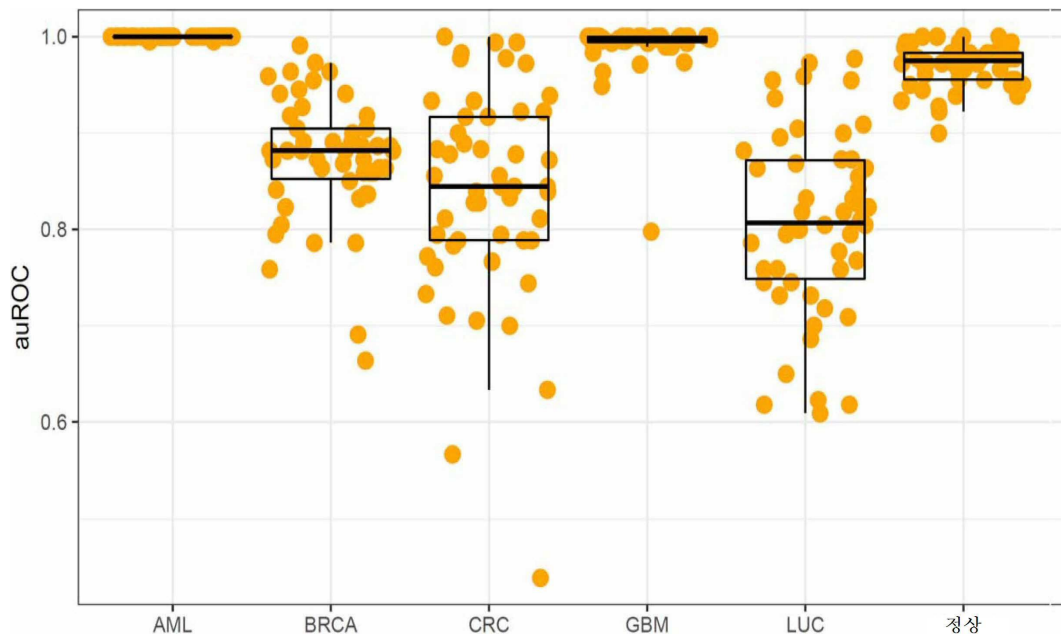
도면2b



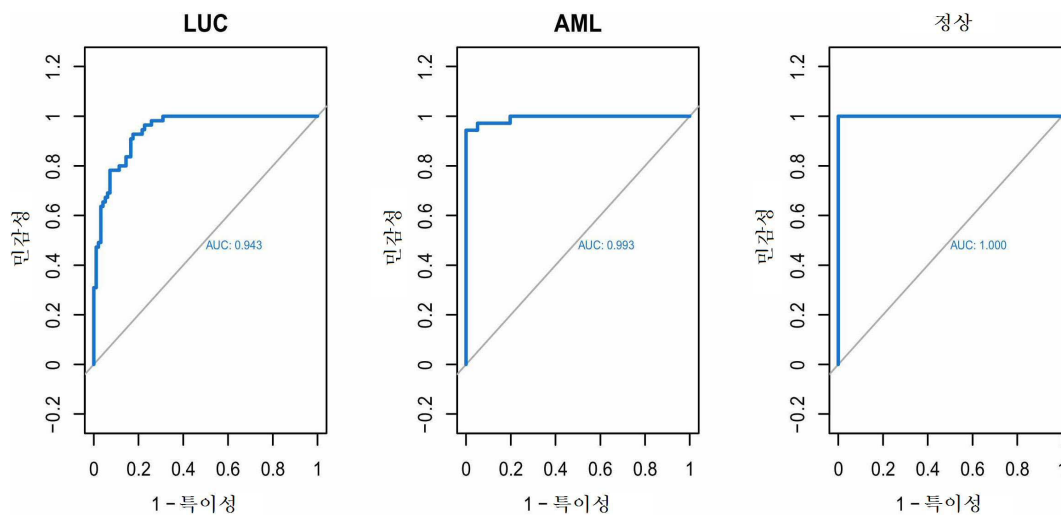
도면2c



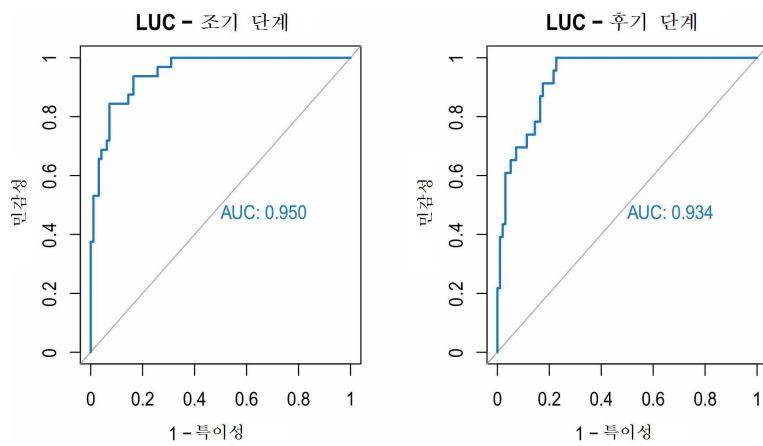
도면2d



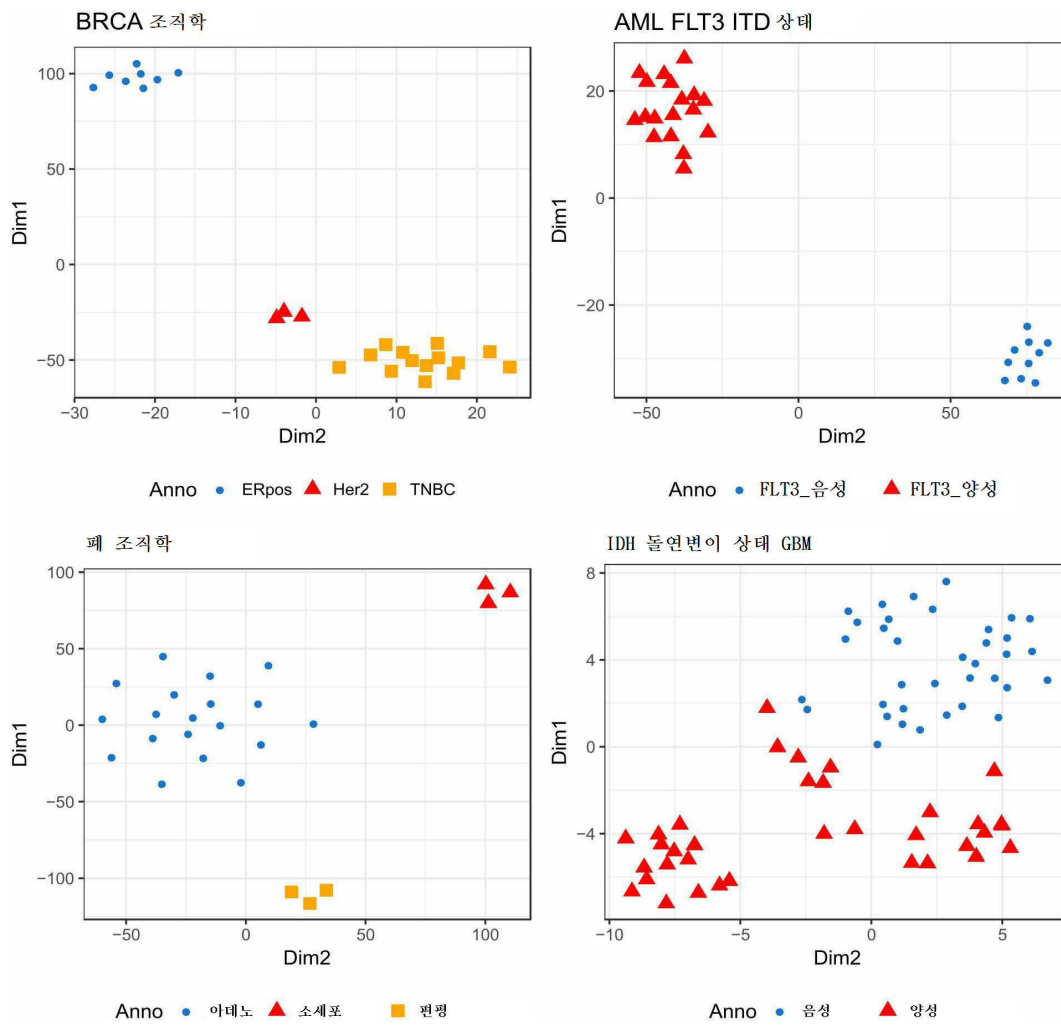
도면3a



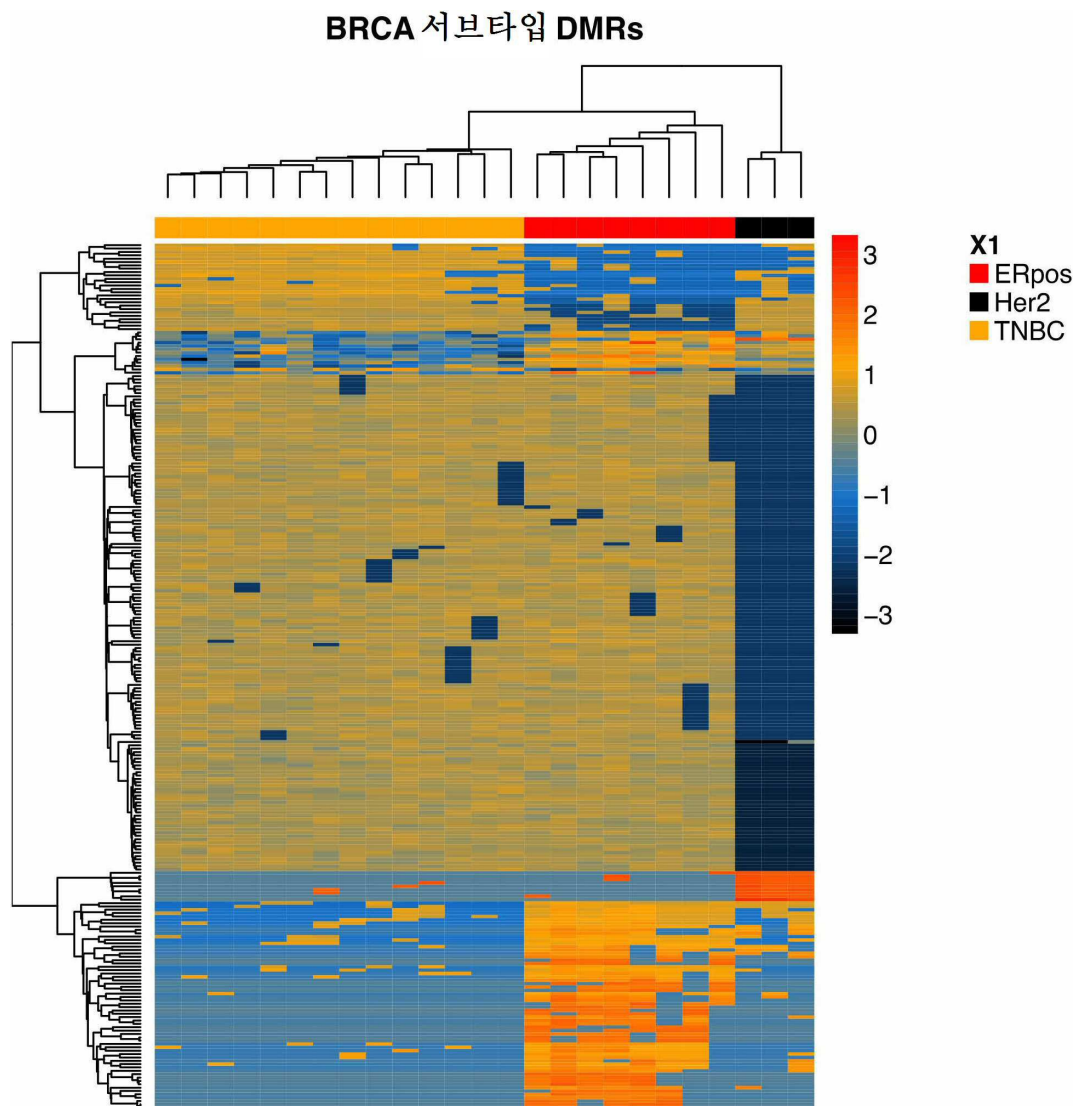
도면3b



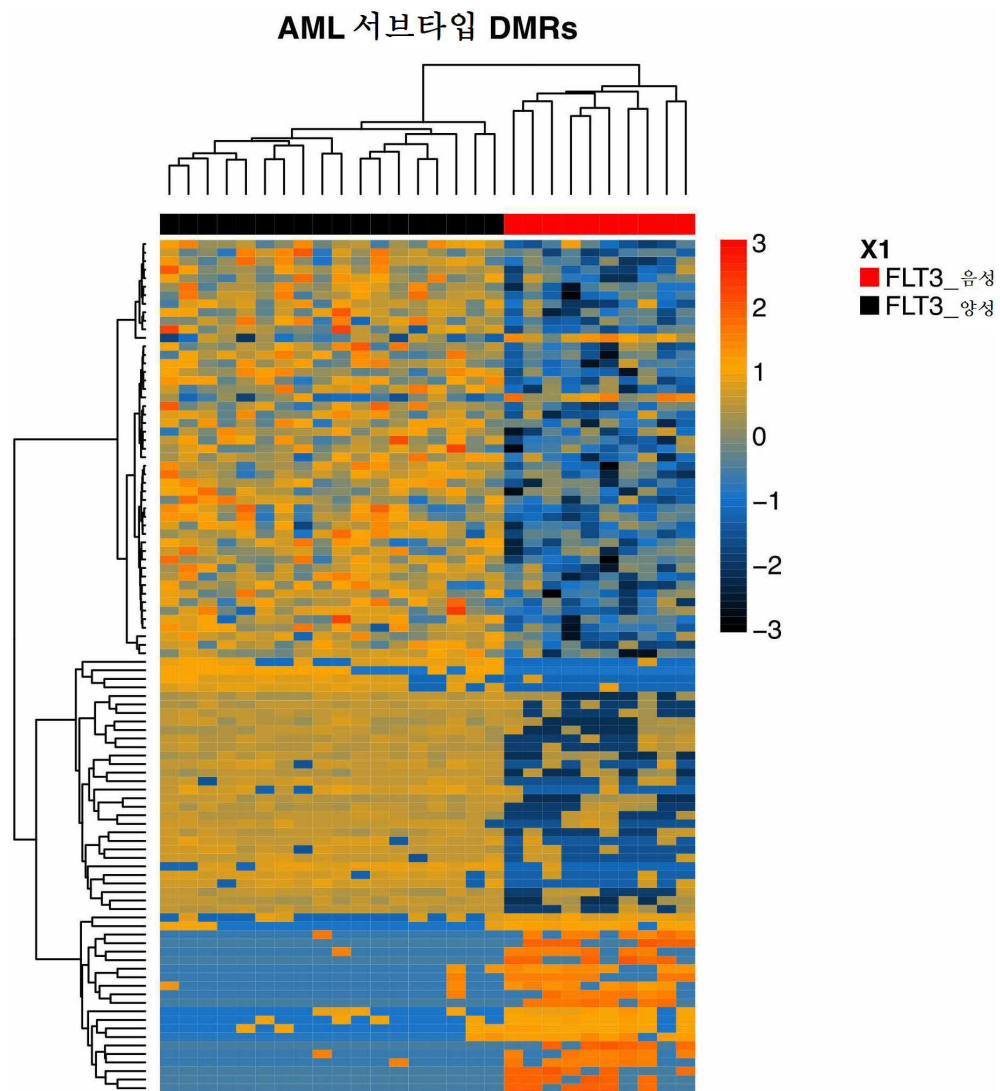
도면4a



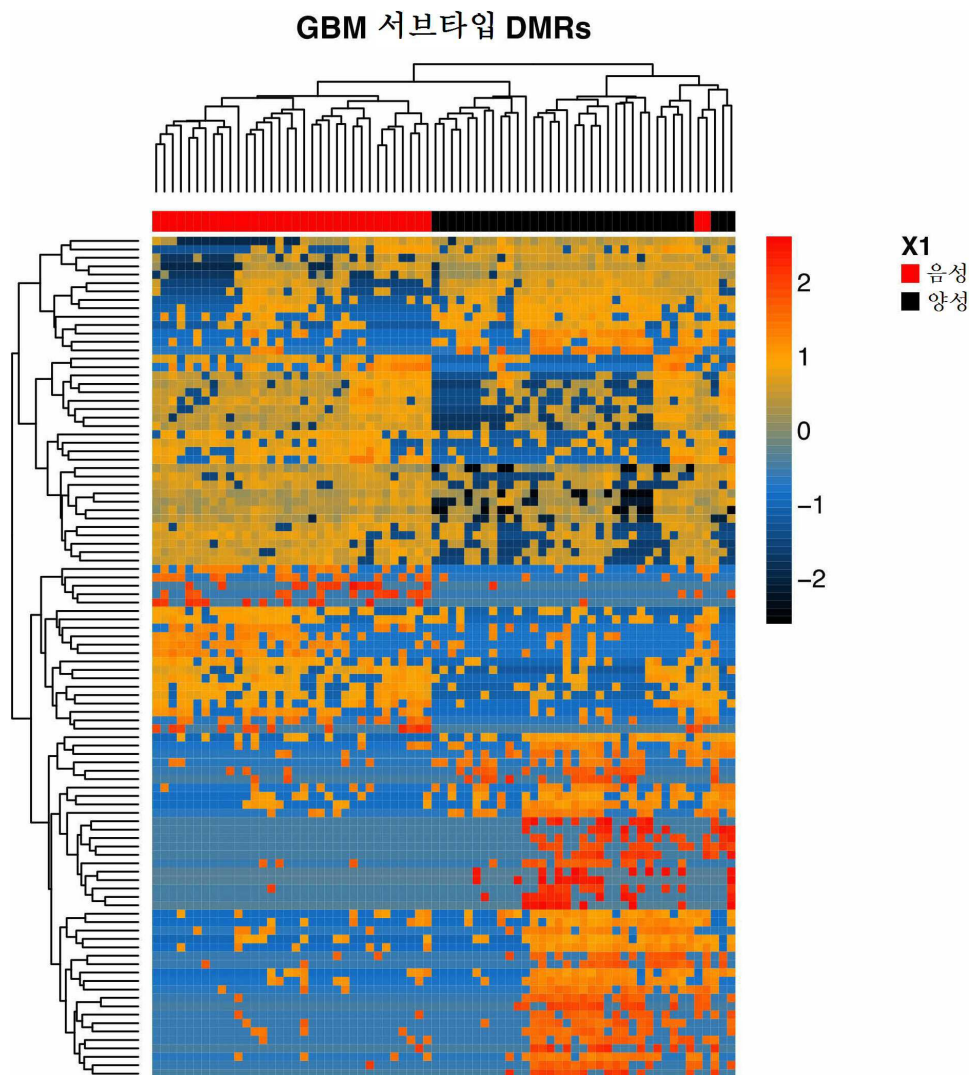
도면4b



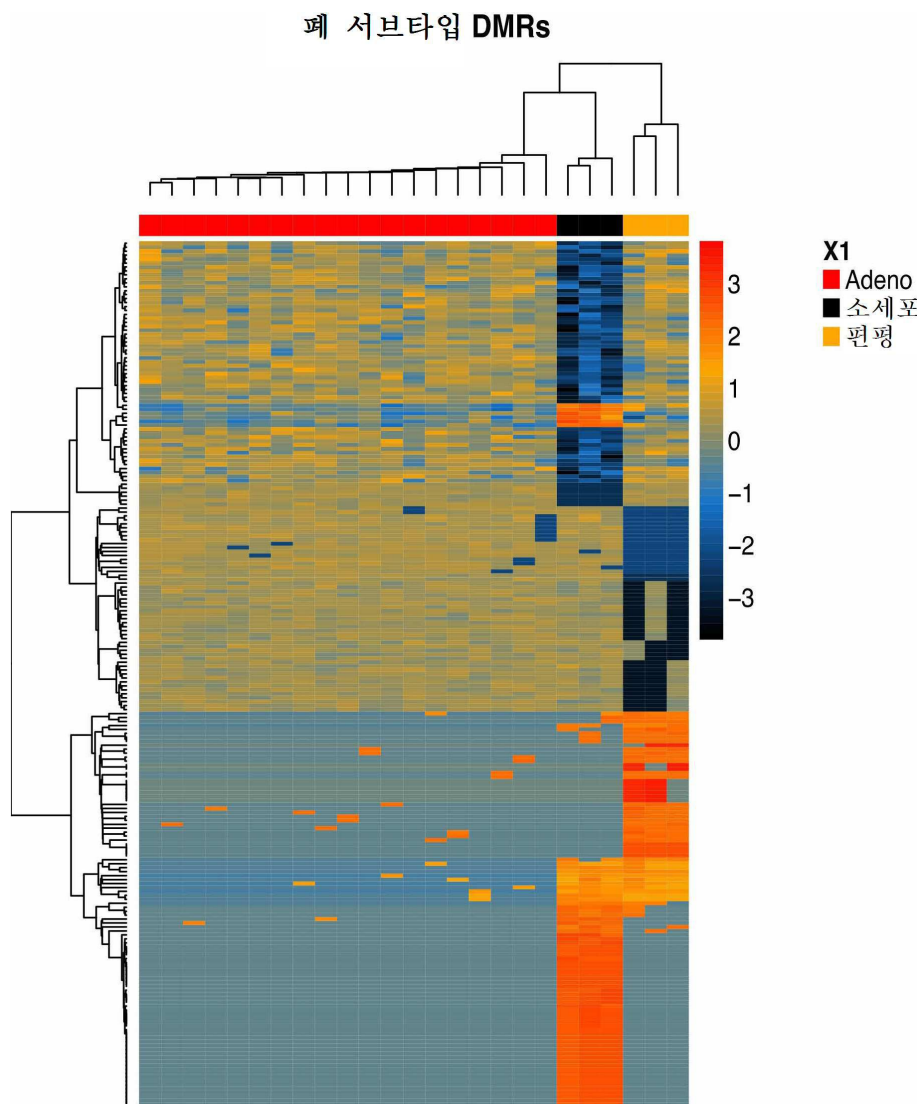
도면4c



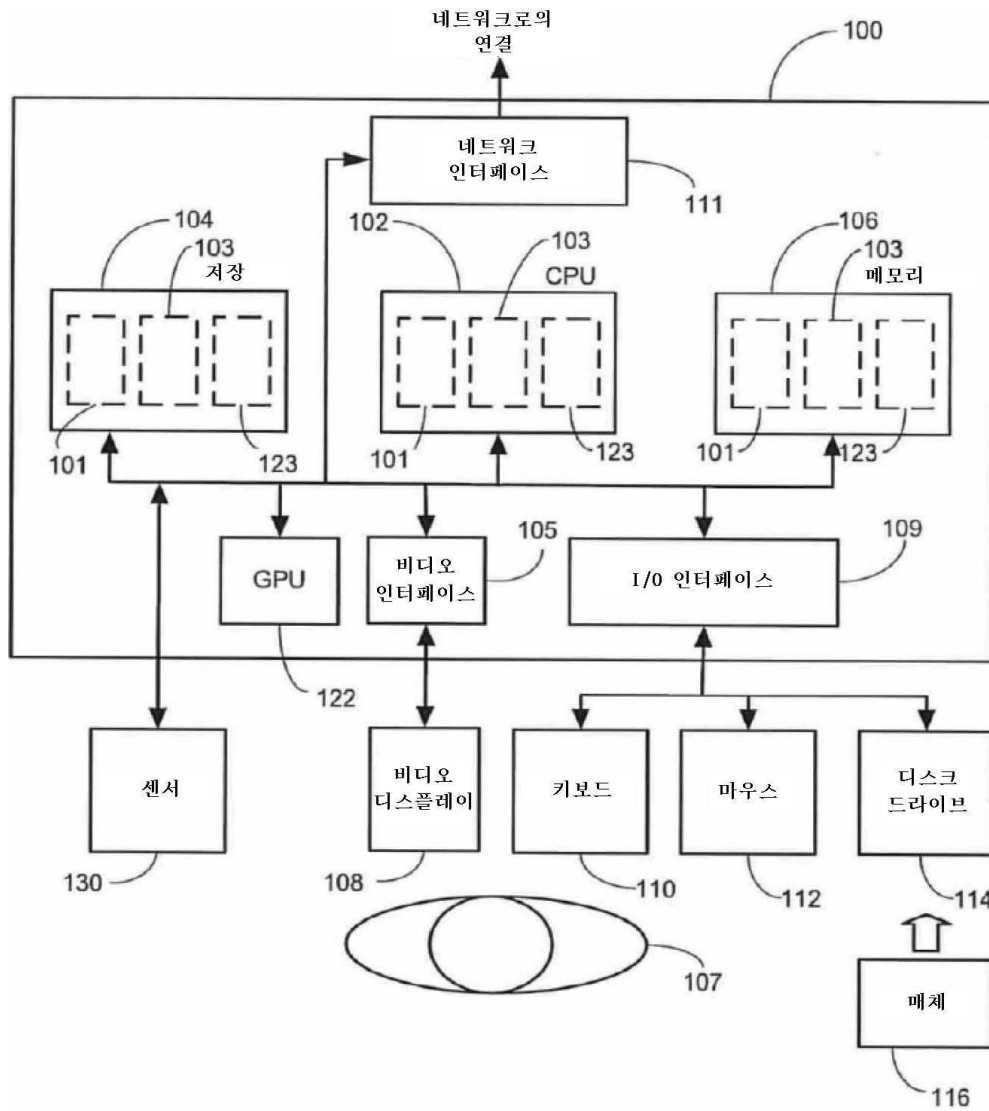
도면4d



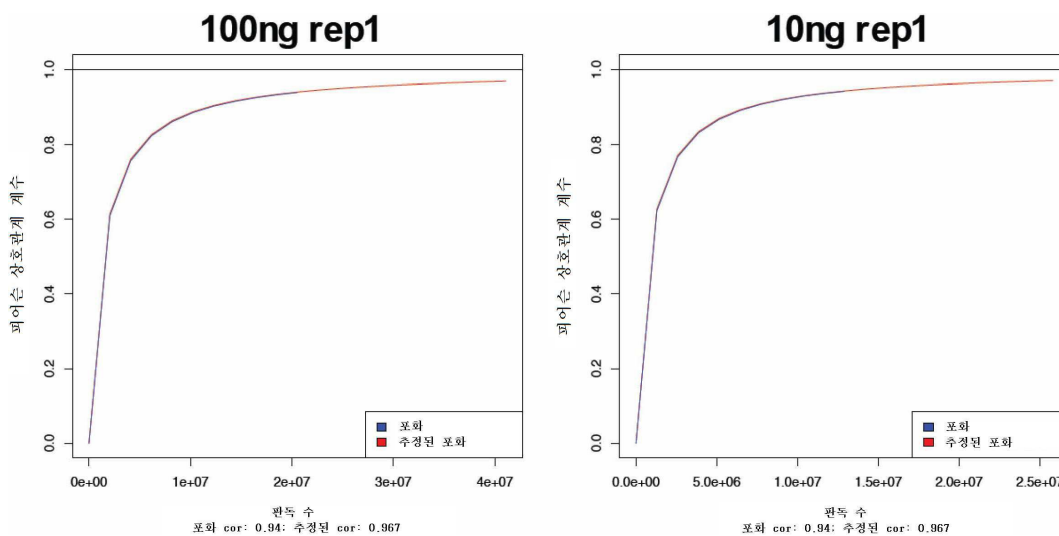
도면4e



도면5

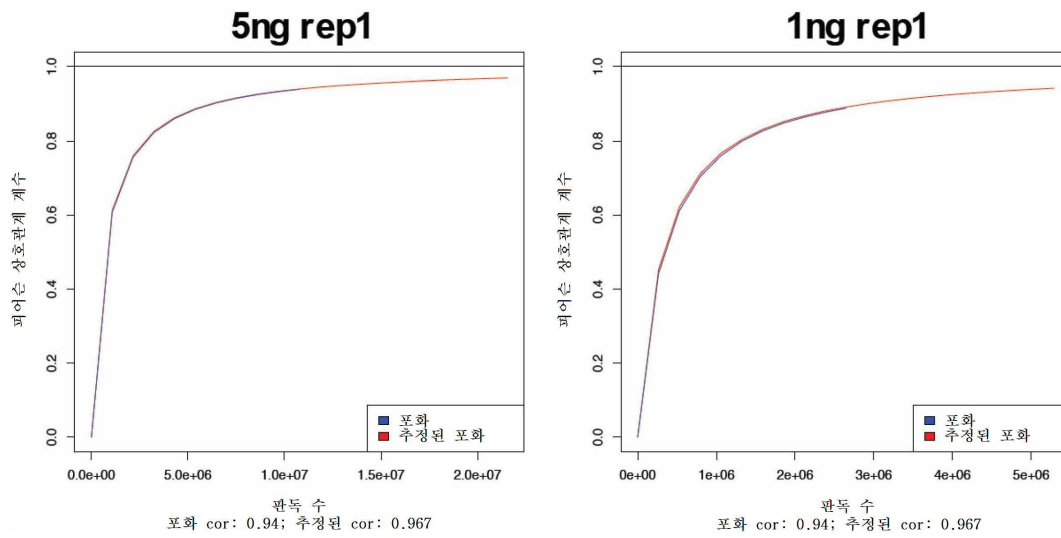


도면6a



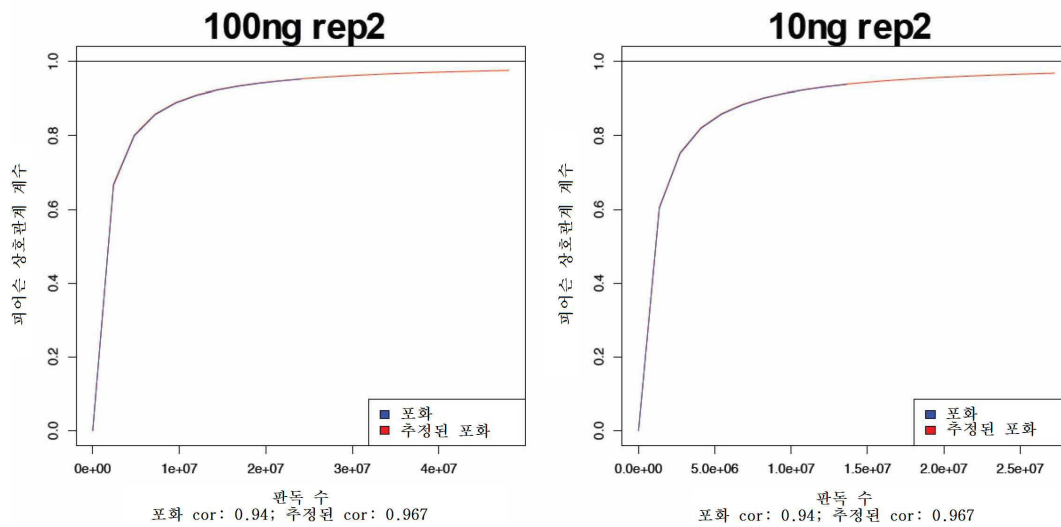
도면6ab

도 6A 계속



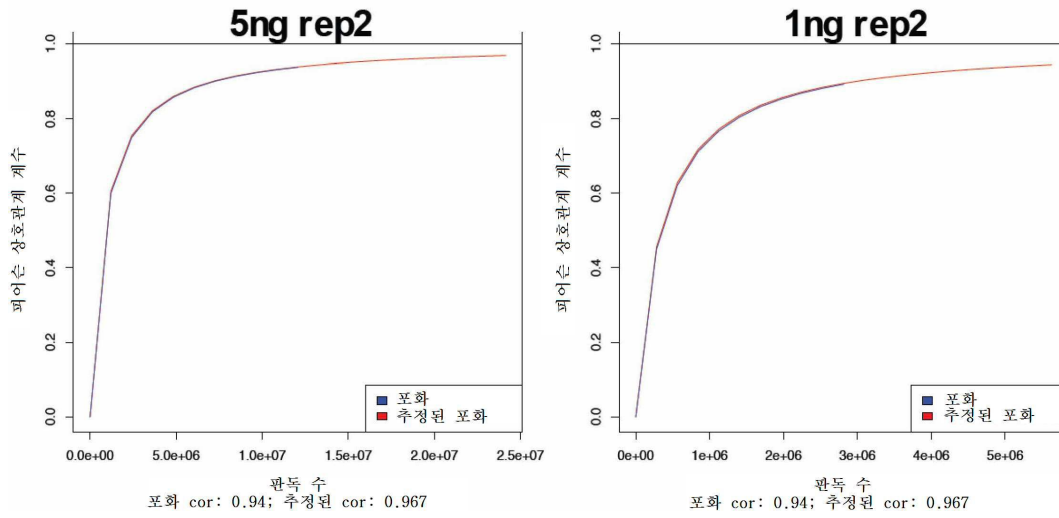
도면6ac

도 6A 계속



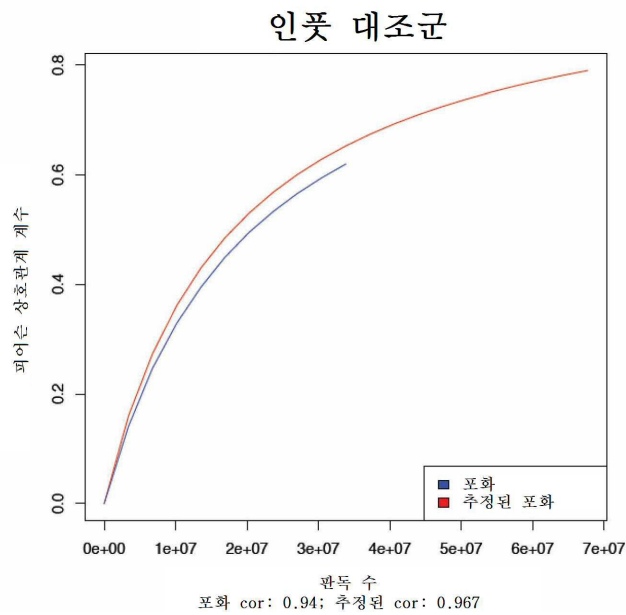
도면6ad

도 6A 계속



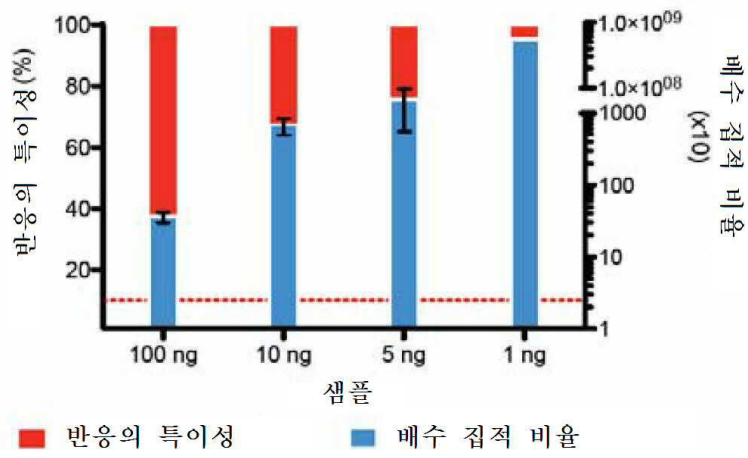
도면6ae

도 6A 계속



도면6b

B



도면6c

C

