

명세서

발명의 명칭: 핵산의 메틸화 차이를 이용한 마커 선별방법, 메틸 또는 탈메틸 마커 및 이 마커를 이용한 진단방법

기술분야

- [1] 본 출원은 2020년 2월 20일에 출원된 대한민국 특허출원 제10-2020-0020974호를 우선권으로 주장하고, 상기 명세서 전체는 본 출원의 참고문헌이다.

[2]

- [3] 본 발명은 핵산의 메틸화 차이를 이용한 마커 선별방법, 탈메틸 마커 및 이 마커를 이용한 진단방법에 관한 것으로, 보다 상세하게는 유리 핵산에서의 메틸화 차이를 이용하는 질환 특이 탈메틸화 마커를 선별하는 새로운 방법, 이 방법에 의해서 선별된 탈메틸화 마커 및 마커의 빈도를 산출하여 암으로 판정하는 메틸화 검출에 의한 새로운 암 진단 방법 및 선별된 cfDNA에서의 암 특이 탈메틸화 마커에 관한 것이다.

배경기술

- [4] 암이란 다양한 원인에 의해 세포의 분열과 사멸 간의 균형이 파괴됨으로써 계속적인 분열과 증식에 의해 발생한 비정상적인 세포의 집단을 의미하며, 종양 또는 신생물이라고도 한다. 일반적으로 장기, 백혈구, 뼈, 림프절 등을 포함한 100 가지 이상의 신체의 여러 부분에 발병하며, 주변조직으로 침윤하는 현상 및 다른 기관으로 이동하는 전이를 통해 심각한 증상으로 발전한다.

[5]

- [6] 의학이 발달한 오늘날에도 인체 암, 특히 대다수를 차지하는 고형암(solid tumor: 혈액암을 제외한 나머지 암)의 경우 5년 생존율은 50% 미만이다. 전체 암환자의 약 3분의 2는 진행된 단계에서 발견되며, 이들 대부분은 진단 후 2년 이내에 사망한다. 이와 같이 저조한 암의 치료효과는 치료법의 문제 뿐만은 아니며, 실제 암을 조기에 진단할 수 있는 방법과 진행된 암을 정확히 진단하고 치료 후 추적 조사하는 것이 용이하지 않기 때문이다.

[7]

- [8] 현재 임상에서 암의 진단은 문진(history taking)과 신체검사, 임상병리검사를 거쳐 일단 의심이 되면 방사선 검사 및 내시경 검사로 진행되며, 최종적으로는 조직 검사로 확인된다. 그러나 현존 임상 검사법으로는 암의 세포수가 10억 개, 암의 직경이 1 cm 이상이 되어야 진단이 가능하다. 이런 경우 이미 암세포는 전이 능력을 갖고 있으며, 실제 절반이상에서 암이 이미 전이되어 있다. 한편, 암이 직간접으로 생산하는 물질을 혈액 내에서 찾는 종양마커(tumor markers)가 암 선별검사(cancer screening)에 이용되는데, 이는 정확도에 한계가 있어서 암이 있을 때도 약 절반까지 정상으로 나타나며, 암이 없을 때도 종종 양성으로

나타나서 혼란을 야기한다. 또한, 암의 치료에 주로 사용되는 항암제의 경우, 암의 용적이 적은 경우에만 그 효과를 나타내는 문제점이 있다.

[9]

[10] 상기한 바와 같이, 암의 진단과 치료가 모두 어려운 것은 정상세포와 다른 점이 많고, 매우 복잡하고 다양하기 때문이다. 암은 제 몇대로 과잉으로 계속 자라며, 사망에서 해방되어 계속 생존하고, 주위 조직을 침범하고 원위 장기로 확산(전이)되어서 인간을 사망하게 한다. 면역기전의 공격이나 항암 치료에도 생존하고, 끊임없이 진화하며 생존에 가장 유리한 세포군(클론)이 선택적으로 증식한다. 암세포는 다수의 유전자의 변이에 의해 발생하는 고도의 생존능력을 가진 생존체이다. 하나의 세포가 암세포로 바뀌고, 임상에서 보는 악성의 암 덩어리로 발전해 나가기 위해서는 다수의 유전자에 변이가 일어나야 한다. 따라서 암을 근원적으로 진단하고 치료하기 위해서는 유전자 수준에서 접근할 필요가 있다.

[11]

[12] 이에, 최근에는 DNA 메틸화 측정을 통하여 암을 진단하는 방법들이 제시되고 있다. DNA 메틸화는 주로 특정 유전자의 프로모터 부위의 CpG 섬(CpG island)의 사이토신(cytosine)에서 일어나고, 그로 인하여 전사인자의 결합이 방해받게 되어 특정 유전자의 발현이 차단(gene silencing)되는 것으로, 이는 생체 내에서 유전자의 단백질 지정 코딩서열(coding sequence)에 돌연변이(mutation)가 없이도 그 유전자의 기능이 소실되는 주요 기전이며, 인체 암에서 다수의 종양 억제 유전자(tumor suppressor genes)의 기능이 소실되는 원인으로 해석되고 있다. 프로모터 CpG 섬의 메틸화가 발암을 직접 유발하는지, 또는 발암에 2차적인 변화인지에 대한 논란이 있으나, 전립선암, 결장암, 자궁암, 유방암 등 다양한 암 세포에서 CpG 섬에서의 이러한 비정상적인 메틸화/탈메틸화가 보고되었다. 따라서, 이는 암의 조기진단, 발암 위험의 예측, 암의 예후 예측, 치료 후 추적 조사, 항암요법에 대한 반응 예측 등 다방면으로 이용될 수 있다. 이를 메틸화 특이 PCR(이하 MSP라고 함), 자동염기분석 또는 bisulfite pyrosequencing 등의 방법으로 검사하여 암의 진단과 스크리닝 등에 이용하려는 시도가 최근 활발하게 이루어지고 있으나, 다수는 소수의 특정 유전자 또는 프로모터 부위의 메틸화를 검출하여 이를 분석하는 방법에 그치고 있으며 (예를 들어, 대한민국 특허 제1557183호, 대한민국 특허 제1191947호). 진단의 효율 및 정확도에는 한계가 있다.

[13]

[14] 특히, 암세포의 게놈에 전반적으로 메틸화 변화가 일어나는데 가장 광범위한 변화는 반복서열에서 일어난다. 게놈의 반복서열은 transposon, retrotransposon, LINE, SINE 등 다양한 종류가 있으며 전체 게놈의 반 이상을 차지할 정도로 큰 비중을 차지하지만 연구가 상대적으로 되지 않았는데, 그 이유는 반복서열은 기능 분석이 어려울 뿐만 아니라 어셈블리가 잘 되지 않아 레퍼런스 서열

(참조표준 서열)에 포함되지 않는 영역이 많아 분석에서 제외되기 쉽기 때문이다. 이런 이유로 반복서열에 대한 메틸레이션 연구는 상대적으로 진행이 많이 되지 않았으며 반복서열에서 많이 일어나는 암과 연관된 메틸화에 대한 의미와 마커 개발은 상대적으로 연구가 활발하지 않았다. 하지만 다양한 게놈분석기술을 활용한 연구에서 암의 진행과 함께 DNA의 저메틸화(hypomethylation)가 광범위하게 일어난다는 다양한 연구결과가 축적되어 있어 (Epigenomics. 2009 December ; 1(2): 239-259, Clin Chem Lab Med. 2012 Oct 1;50(10):1733-42) 반복서열에서의 저메틸화를 암의 진단 마커로 활용할 수 있을 것으로 기대하고 있다.

[15]

[16] 본 명세서 전체에 걸쳐 다수의 논문 및 특허문헌이 참조되고 그 인용이 표시되어 있다. 인용된 논문 및 특허문헌의 개시 내용은 그 전체로서 본 명세서에 참조로 삽입되어 본 발명이 속하는 기술 분야의 수준 및 본 발명의 내용이 보다 명확하게 설명된다.

발명의 상세한 설명

기술적 과제

[17] 이에 본 발명자들은 비침습적인 방법으로 암의 정확한 진단이 가능한 신규한 방법을 개발하기 위하여 연구하던 중, cfDNA를 메틸화 민감성 제한효소로 처리하여 제한효소 타겟 서열중 메틸화 되지 않은 서열을 절단하고 해독(sequencing) 한 후, 해독된 서열에서 일정 길이의 서열정보를 이용하여 각각 분류하는 경우 질환, 특히 암과 같은 질환에 대해서 혈액 내의 cfDNA 종류를 분류할 수 있고, 이를 통해서 질환에 대한 cfDNA 마커로 작용할 수 있음을 확인하여 cfDNA에서의 메틸화와 연관된 암 특이 마커, 특히 암 특이 탈메틸화 마커를 선별하는 방법을 개발함으로써 본 발명을 완성하였다.

[18]

[19] 따라서, 본 발명의 목적은 cfDNA에서의 암 특이 탈메틸화 마커를 선별하는 새로운 방법을 제공하는 것이다.

[20] 또한 본 발명의 다른 목적은 선별된 cfDNA에서의 암 특이 탈메틸화 마커의 빈도를 산출하여 암으로 판정하는 탈메틸화 검출에 의한 새로운 암 진단 방법을 제공하는 것이다.

[21] 본 발명의 또다른 목적은 암 진단에 필요한 정보를 제공하기 위하여, 개체의 혈액에서 분리한 cfDNA의 메틸화 민감성 제한효소 단편의 N-말단의 미리 정해진 길이의 서열 정보를 해독하고 분석하는 방법을 제공하는 것이다.

[22] 본 발명의 다른 목적은 N말단이 메틸화 민감성 제한효소의 인식부위의 점착성 말단 (cohesive end)의 서열이며, 25염기 내지 150염기의 서열로 이루어지며, 본 발명의 방법에 의해서 선별된 cfDNA에서의 암 특이 탈메틸화 마커를 제공하는 것이다.

기술적 해결방법

- [23] 상기와 같은 목적을 달성하기 위하여, 본 발명은 혈액에서 분리한 cfDNA (cell free DNA)에 메틸화 민감성 제한효소 (methylation sensitive restriction enzyme)를 처리하는 단계; 각 단편의 서열을 해독(sequencing)하는 단계; 단편의 N-말단으로부터 미리 정해진 길이의 서열 정보를 획득하는 단계; 각 서열 정보의 빈도를 계수하는 단계; 암 특이 서열 정보를 cfDNA에서의 암 특이 탈메틸화 마커로 선별하는 단계를 포함하는 cfDNA에서의 암 특이 탈메틸화 마커를 선별하는 방법을 제공한다.
- [24] 또한 본 발명의 다른 목적을 달성하기 위하여 본 발명은 선별된 cfDNA에서의 암 특이 탈메틸화 마커의 빈도를 산출하여 암으로 판정하는 메틸화 검출에 의한 새로운 암 진단 방법을 제공한다.
- [25] 본 발명의 또 다른 목적을 달성하기 위하여, 본 발명은 암 진단에 필요한 정보를 제공하기 위하여, 개체의 혈액에서 분리한 cfDNA의 메틸화 민감성 제한효소 단편의 N-말단의 미리 정해진 길이의 서열 정보를 분석하는 방법을 제공한다.
- [26] 본 발명의 다른 목적을 달성하기 위하여, 본 발명은 N말단이 메틸화 민감성 제한효소의 인식부위의 점착성 말단 (cohesive end)의 서열이며, 25염기 내지 150염기의 서열로 이루어지며, 본 발명의 방법에 의해서 선별된 cfDNA에서의 암 특이 탈메틸화 마커를 제공한다.
- [27]
- [28] 다른 정의가 없는 한, 본 명세서에 사용된 모든 기술적 및 과학적 용어는 당업자들에 의해 통상적으로 이해되는 동일한 의미를 가진다. 다음의 참고문헌은 본 발명의 명세서에 사용된 여러 용어들의 일반적인 정의를 갖는 기술(skill)의 하나를 제공한다: Singleton et al., DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY(2th ed. 1994); THE CAMBRIDGE DICTIONARY OF SCIENCE AND TECHNOLOGY(Walkered., 1988); 및 Hale & Marham, THE HARPER COLLINS DICTIONARY OF BIOLOGY.
- [29]
- [30] 이하 본 발명을 상세히 설명한다.
- [31]
- [32] 본 발명은 (a) 혈액에서 분리한 cfDNA (cell free DNA)에 메틸화 민감성 제한효소 (methylation sensitive restriction enzyme)를 처리하는 단계; (b) 각 단편의 서열을 분석(sequencing)하는 단계; (c) 단편의 N-말단으로부터 미리 정해진 길이의 서열 정보를 획득하는 단계; (d) 각 서열 정보의 빈도를 계수하는 단계; 및 (e) 암 특이 서열 정보를 cfDNA에서의 암 특이 탈메틸화 마커로 선별하는 단계를 포함하는 cfDNA에서의 암 특이 탈메틸화 마커를 선별하는 방법에 관한 것이다.
- [33]
- [34] **메틸화(methylation)**

- [35] 본 발명에서의 정제되거나 정제되지 않은 형태의 어떠한 핵산도 사용될 수 있으며, 타겟 부위(예를 들면, CpG-함유 핵산)를 함유하는 핵산 서열을 함유하고 있거나 함유할 것으로 의심되는 어떠한 핵산도 사용될 수 있다. 차별적으로 메틸화될 수 있는 핵산 부위는 CpG 서열의 C 위치이며 특히 GpG가 밀집해있는 CpG 섬에 메틸화가 많이 이루어진다. 특정 부위에서, CpG 섬의 밀도는 게놈의 다른 부위와 비교하여 10배나 더 높다. CpG 섬은 평균 G*C 비율이 약 60%이며, 반면에 보통의 DNA의 G*C 비율은 평균 40%를 나타낸다. CpG 섬은 전형적으로 약 1~2kb 길이를 가지고, 인간 게놈에는 약 45,000개의 CpG 섬이 존재한다.
- [36] 통상적으로, 시료가 되는 핵산은 DNA이다. 그러나, 핵산 혼합물 또한 사용될 수 있다. 검출될 특이적인 핵산 서열은 큰 분자의 분획일 수 있고, 처음부터 특이 서열이 전체 핵산 서열을 구성하는 분리된 분자 형태로 존재할 수 있다. 상기 핵산 서열은 순수한 형태로 존재하는 핵산일 필요는 없으며, 핵산은 전체 인간 DNA가 포함되어 있는 것과 같이 복잡한 혼합물 내의 적은 분획일 수도 있다. 시료에 포함된 핵산의 메틸화 정도를 측정하는 데 사용되거나, 메틸화된 CpG 섬을 검출하는 데 사용되는 시료에 포함된 핵산은 당업계에 공지된 통상적인 방법으로 추출될 수 있다.
- [37]
- [38] 염기서열 분석 - 시퀀싱(sequencing)
- [39] 시퀀싱 방법은, 예를 들어, 생거(Sanger) 시퀀싱, 고 처리량 시퀀싱, 피로시퀀싱, 합성에 의한 시퀀싱, 단일 분자 시퀀싱, 나노포어 시퀀싱, 반도체 시퀀싱, 라이게이션에 의한 시퀀싱, 혼성화에 의한 시퀀싱, RNA-Seq (일루미나), 디지털 유전자 발현 [헬리코스(Helicos)], 차세대 시퀀싱 (NGS), 합성에 의한 단일 분자 시퀀싱 (SMSS) (헬리코스), 대규모 병렬 시퀀싱, 클로날 단일 분자 어레이 [솔렉사(Solexa)], 샷건 시퀀싱, 이온 토렌트(Ion Torrent), 옥스포드 나노포어, 로슈 제니아(Roche Genia), 맥심-길버트(Maxim-Gilbert) 시퀀싱, 프라이머 워킹; PacBio, SOLiD, 이온 토렌트, 또는 나노포어 플랫폼을 이용하는 시퀀싱을 포함한다. 시퀀싱 반응은 다수의 레인, 다수의 채널, 다수의 웰, 또는 다수의 샘플 세트를 실질적으로 동시에 프로세싱하는 다른 수단일 수 있는 각종 샘플 프로세싱 유닛에서 수행될 수 있다. 샘플 프로세싱 유닛은 또한, 다수의 실행물을 동시에 프로세싱할 수 있게 해주는 다수의 샘플 챔버를 포함할 수 있다.
- [40] 시퀀싱 반응은 그 중 적어도 하나가 질환의 마커를 함유하는 것으로 공지되어 있는 하나 이상의 형태의 핵산 상에서 수행될 수 있다. 시퀀싱 반응은 또한, 샘플에 존재하는 임의의 핵산 단편 상에서 수행될 수 있다.
- [41] 동시 시퀀싱 반응은 멀티플렉스 시퀀싱을 이용하여 수행될 수 있다. 일부 경우에, 무세포 핵산은 적어도 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 50000, 100,000개의 시퀀싱 반응으로 서열 분석될 수 있다. 다른 경우에 무세포 핵산은 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000,

10000, 50000, 100,000개 미만의 시퀀싱 반응으로 서열 분석될 수 있다. 시퀀싱 반응은 순차적으로 또는 동시에 수행될 수 있다. 후속 데이터 분석은 시퀀싱 반응의 전부 또는 일부에 대해 수행될 수 있다. 일부 경우에, 데이터 분석은 적어도 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 50000, 100,000개의 시퀀싱 반응 상에서 수행될 수 있다. 다른 경우에, 데이터 분석은 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 50000, 100,000개 미만의 시퀀싱 반응 상에서 수행될 수 있다. 예시적인 판독물 템스는 유전자 자리 (염기) 당 1000 내지 50000개의 리드(read)이다.

[42]

[43] **샘플**

[44] 샘플은 개체로부터 분리된 임의의 생물학적 샘플일 수 있다. 샘플은 신체 샘플일 수 있다. 샘플은 신체 조직, 예컨대 공지되거나 또는 의심되는 고형 종양, 전혈, 혈청, 혈장, 대변, 백혈구 또는 림프구, 내피 세포, 조직 생검, 뇌척수액, 활액, 림프액, 복수, 간질 액 또는 세포의 유체, 세포 사이 공간 내의 유체 (잇몸 치은구의 유체 포함), 골수, 흉막 삼출액, 뇌척수액, 타액, 점액, 객담, 정액, 땀, 소변을 포함할 수 있다. 샘플은 대상체로부터 원래 분리된 형태일 수 있거나 또는 세포와 같은 성분을 제거 또는 부가하기 위해 또는 또 다른 성분과 비교하여 한 성분을 풍부화시키기 위해 추가로 프로세싱시킬 수 있었다. 샘플은 대상체로부터 분리 또는 수득될 수 있고, 샘플 분석 부위로 수송될 수 있다. 샘플은 원하는 온도, 예를 들어, 실온, 4°C, -20°C, 및/또는 -80°C 하에 보존 및 선적될 수 있다. 샘플은 샘플 분석 부위에서 대상체로부터 분리 또는 수득될 수 있다.

[45] 개체는 인간, 포유동물, 동물, 애완동물 동물, 서비스 동물, 또는 애완동물일 수 있다. 개체는 질환이 있을 수 있다. 개체는 질환 또는 검출 가능한 질환 증상이 없을 수 없다. 개체는 하나 이상의 요법, 예를 들어, 수술, 처치, 투약, 화학요법, 항체, 백신 또는 생물 제제 중 어느 하나 이상으로 치료받은 적이 있을 수 있다. 개체는 차도가 있을 수 있거나 또는 그렇지 않을 수 있다.

[46]

[47] **혈액 샘플의 무세포 핵산**

[48] 혈액 샘플은 게놈 DNA를 함유하는 다양한 양의 핵산을 포함할 수 있다. 예를 들어, 약 33 ng DNA의 샘플은 약 10,000개 (10^4)의 반수체 인간 게놈 DNA를 함유할 수 있고, cfDNA의 경우에는, 약 2천억개 (2×10^{11}) 개별 폴리뉴클레오타이드 분자를 함유할 수 있다. 유사하게, 약 100 ng의 DNA 샘플은 약 30,000개의 반수체 인간 게놈 DNA를 함유할 수 있고, cfDNA의 경우에는, 약 6천억개의 개별 분자를 함유할 수 있다.

[49] 증폭 전의 샘플 내의 무세포 핵산의 예시적인 양은 약 1 fg 내지 약 1 μ g, 예를 들어, 1 pg 내지 200 ng, 1 ng 내지 100 ng, 10 ng 내지 1000 ng의 범위이다. 예를 들어, 그 양은 약 600 ng 이하, 약 500 ng 이하, 약 400 ng 이하, 약 300 ng 이하, 약

200 ng 이하, 약 100 ng 이하, 약 50 ng 이하, 또는 약 20 ng 이하, 또는 약 10ng 이하, 또는 약 5ng 이하, 또는 약 1ng 이하의 무세포 핵산 분자일 수 있다. 그 양은 적어도 1 fg, 적어도 10 fg, 적어도 100 fg, 적어도 1 pg, 적어도 10 pg, 적어도 100 pg, 적어도 1 ng, 적어도 10 ng, 적어도 100 ng, 적어도 150 ng, 또는 적어도 200 ng의 무세포 핵산 분자일 수 있다. 그 양은 1 펩토그램 (fg), 10 fg, 100 fg, 1 피코그램 (pg), 10 pg, 100 pg, 1 ng, 10 ng, 100 ng, 150 ng, 또는 200 ng 이하의 무세포 핵산 분자일 수 있다. 상기 방법은 1 펩토그램 (fg) 내지 200 ng을 수득하는 것을 포함할 수 있다.

- [50] 무세포 핵산은 세포 내에 함유되지 않거나 또는 달리 세포와 결합되지 않는 핵산, 또는 다시 말해서, 무손상 세포를 제거한 후에도 샘플 내에 남아있는 핵산이다. 무세포 핵산은 DNA, RNA, 및 그의 혼성체를 포함하며, 이는 게놈 DNA, 미토콘드리아 DNA, siRNA, miRNA, 순환 RNA (cRNA), tRNA, rRNA, 작은 핵소체 RNA (snoRNA), Piwi-상호 작용성 RNA (piRNA), 긴 비-코딩 RNA (긴 ncRNA), 또는 이들 중 임의의 것의 단편을 포함한다. 무세포 핵산은 이중 가닥, 단일 가닥, 또는 그의 혼성체일 수 있다. 무세포 핵산은 분비 또는 세포 사멸 프로세스, 예를 들어, 세포성 괴사 및 아포토시스를 통해 체액 내로 방출될 수 있다. 일부 무세포 핵산은 암세포, 예를 들어, 순환 종양 DNA (ctDNA)로부터 체액 내로 방출된다. 다른 것은 건강한 세포로부터 방출된다. 일부 실시양태에서, 무세포 핵산은 종양 세포에 의해 생산된다. 일부 실시양태에서, 무세포 핵산은 종양 세포와 비-종양 세포의 혼합물에 의해 생산된다.
- [51] 무세포 핵산은 예를 들어 약 100 내지 500개의 뉴클레오티드의 길이 분포를 나타내며, 110 내지 약 230개의 뉴클레오티드의 분자가 이러한 분자의 약 90%를 차지하고, 240 내지 440개의 뉴클레오타이드 범위의 제2 마이너 피크를 수반한다.
- [52] 무세포 핵산은 분획화 또는 분할 단계를 통해 체액으로부터 분리될 수 있으며, 여기서 용액에서 발견된 바와 같은 무세포 핵산은 무손상 세포 및 체액의 다른 비-가용성 성분으로부터 분리된다. 분할은 원심분리 또는 여과와 같은 기술을 포함할 수 있다. 또 다른 한편으로, 체액 중의 세포는 용해될 수 있고, 무세포와 세포성 핵산은 함께 프로세싱될 수 있다. 일반적으로, 완충액의 부가 및 세척 단계 후, 핵산은 알콜로 침전될 수 있다. 추가의 정화 단계는, 예컨대 실리카 기반 칼럼을 사용하여 오염물질 또는 염을 제거할 수 있다. 중아황산염 시퀀싱, 혼성화, 및/또는 라이게이션을 위한 비-특이적 벌크 캐리어 핵산, 예컨대 Cot-1 DNA, DNA 또는 단백질이 반응 내내 부가되어, 본 절차의 특정 측면, 예컨대 수율을 최적화할 수 있다.
- [53] 이러한 프로세싱 후, 샘플은 이중 가닥 DNA, 단일 가닥 DNA 및 단일 가닥 RNA를 포함한 다양한 형태의 핵산을 포함할 수 있다. 일부 실시양태에서, 단일 가닥 DNA 및 RNA는 이중 가닥 형태로 전환될 수 있으므로, 이들은 후속 프로세싱 및 분석 단계에 포함된다.

- [54]
- [55] 본 발명의 하나의 양태에서 cfDNA는 인간 게놈 DNA에서 유래될 수 있고, 인간과 공생하거나 인간에게 감염된 인간 이외의 세포, 세균, 균 또는 바이러스의 DNA에서 유래될 수도 있다.
- [56]
- [57] 본 발명의 하나의 양태에서 cfDNA에서의 암 특이 탈메틸화 마커를 선별하는 방법은 다음과 같은 단계를 포함할 수 있다:
- [58] (a) 혈액에서 분리한 cfDNA (cell free DNA)에 메틸화 민감성 제한효소 (methylation sensitive restriction enzyme)를 처리하는 단계;
- [59] (b) 각 단편의 서열을 분석(sequencing)하는 단계;
- [60] (c) 단편의 N-말단으로부터 미리 정해진 길이의 서열 정보를 획득하는 단계;
- [61] (d) 각 서열 정보의 빈도를 계수하는 단계; 및
- [62] (e) 암 특이 서열 정보를 cfDNA에서의 암 특이 탈메틸화 마커로 선별하는 단계.
- [63]
- [64] (a) 단계는 혈액에서 분리한 cfDNA (cell free DNA)에 메틸화 민감성 제한효소 (methylation sensitive restriction enzyme)를 처리하는 단계이다.
- [65] cfDNA는 개체에서 분리한다. 바람직하게는 cfDNA는 혈장에서 분리될 수 있다. 분리 방법은 제한효소의 처리 및 서열분석 (sequencing)에 적절한 순도가 얻어질 수 있는 당업계에 공지된 통상적인 DNA 분리 방법에 의할 수 있다.
- [66] 본 발명의 하나의 양태에서 메틸화 민감성 제한효소는 AatII, AclI, AgeI, Aor13H I, AscI, AsiSI, AvaI, BsaHI, BsiEI, BsiWI, BspDI, BsrFI, BssHII, BstBI, ClaI, Cpo I, EagI, FseI, HaeII, HhaI, HinP1I, HpaII(또는 HapII), HpyCH4IV, Hpy99I, KasI, MluI, NarI, NgoMIV, NotI, PaeR7I, PluTI, PvuI, RsrII, SacII, SalI, SgrAI 또는 TspMI이다. 바람직하게는 본 발명의 메틸화 민감성 제한효소는 i) 메틸화되지 않은 타겟 영역을 선택적으로 절단하며, ii) 절단된 말단이 (blunt end가 아닌) cohesive end를 만들어 상보적인 cohesive end를 가진 adapter의 접합효율을 높힐 수 있어 양질의 라이브러리를 만들 수 있는 특징을 가질 수 있다.
- [67] 본 발명의 하나의 양태에서 메틸화 민감성 제한효소는 CpG methylation에 대한 선택적 절단이 가능한 효소, 즉, demethylation 된 CpG를 포함한 제한효소 인식 부위를 특이적으로 절단할 수 있는 효소인 것이 바람직하다. 다만, 제한효소 인식 부위가 게놈 내에 얼마나 존재하느냐에 따라 실제로 전체 게놈에 대한 커버리지(coverage)와 서열분석(sequencing)을 해야 하는 분석 비용이 다를 수 있으므로 목적에 맞게 적절한 제한효소를 선택할 수 있다.
- [68]
- [69]

[표1]

효소	인식부위	서열 번호	효소	인식부위	서열 번호
AatII	GACGT↓C	1	HhaI	GCG↓C	20
AclI	AA↓CGTT	2	HinPII	G↓CGC	21
AgeI	A↓CCGGT	3	HpaII	C↓CGG	22
Aor13H I	T↓CCGGA	4	HpyCH4IV	A↓CGT	23
AscI	GG↓CGCGCC	5	Hpy99I	CGWCG↓	24
AsiSI	GCGAT↓CGC	6	KasI	G↓GCGCC	25
AvaI	C↓YCGRG	7	MluI	A↓CGCGT	26
BsaHI	GR↓CGYC	8	NarI	GG↓CGCC	27
BsiEI	CGRY↓CG	9	NgoMIV	G↓CCGGC	28
BsiWI	C↓GTACG	10	NotI	GC↓GGCCGC	29
BspDI	AT↓CGAT	11	PaeR7I	C↓TCGAG	30
BsrFI	R↓CCGGY	12	PluTI	GGCGC↓C	31
BssHII	G↓CGCGC	13	PvuI	CGAT↓CG	32
BstBI	TT↓CGAA	14	RsrII	CG↓GWCCG	33
ClaI	AT↓CGAT	15	SacII	CCGC↓GG	34
Cpo I	CG↓GWCCG	16	SalI	G↓TCGAC	35
EagI	C↓GGCCG	17	SgrAI	CR↓CCGGYG	36
FseI	GGCCGG↓CC	18	TspMI	C↓CCGGG	37
HaeII	RGCGC↓Y	19			

[70] (↓는 절단부위를 나타냄)

[71]

[72] (b) 단계는 각 단편의 서열을 분석(sequencing)하는 단계이다.

[73] 본 발명의 하나의 양태에서 서열을 해독하는 것은 당업계에서 공지된 서열 해독 방법에 의해서 수행된다. 서열 해독은 메틸화 민감성 제한효소에 의해서 절단된 단편 또는 절단되지 않은 단편 각각의 서열을 해독한다. 서열 해독은 대량의 단편, 바람직하게는 적어도 10000개이상, 적어도 20000개이상, 적어도 30000개이상, 적어도 40000개이상, 적어도 50000개이상, 적어도 100000개이상, 적어도 1000000개 이상의 단편을 해독하므로 이에 적합한 해독 방법이 바람직하다.

- [74] 서열해독은 당업계에 공지된 서열해독 법이 사용될 수 있으나, 각 단편의 서열을 충분한 수량으로 해독하기 위하여 대량의 서열해독이 가능한 방법이라면 제한없이 사용될 수 있다. 예를 들어 차세대 시퀀싱 방법(NGS, Next generation sequencing)이 사용되는 경우 대량의 서열을 18시간 내에 적은 비용으로 해독할 수 있다는 장점이 있으며, 충분한 양의 서열을 읽는 경우 정확도가 아주 높으며 해독된 데이터를 정성, 정량적으로 분석이 가능하다.
- [75] 서열 해독을 위해서는 바람직하게는 메틸화 민감성 제한효소에 의해 절단된 DNA 단편만을 해독할 수 있도록 적절한 어댑터(adapter)를 부착할 수 있다. 시료 내의 DNA는 메틸화 상태에 따라서 메틸화 민감성 제한효소에 의해서 절단되거나 또는 절단되지 않을 수 있다. 예를 들어, 정상인의 cfDNA에서는 메틸화되어있지만, 암화되며 탈메틸화된 암 DNA를 검출하는 경우에 탈메틸화되어 절단된 단편만을 해독할 수 있다면 아주 낮은 비율로 섞인 cfDNA를 검출하기에 용이해진다. 따라서 메틸화 민감성 제한효소에 의해 절단되어 생성된 cohesive end에 상보적인 구조를 가진 어댑터를 사용하면, 절단된 단편으로만 라이브러리가 만들어지기 때문에 해독단계에서 선별적으로 암에 의한 탈메틸된 단편을 해독할 수 있게 된다.
- [76]
- [77] (c) 단계는 단편의 5'말단으로부터 미리 정해진 길이의 서열 정보를 수득하는 단계이다.
- [78] 본 발명의 하나의 양태에서 '미리 정해진 길이'는 서열 해독된 각 단편에서 5'말단으로 부터의 염기 또는 염기쌍의 길이를 나타내며, 바람직하게는 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150 염기일 수 있다. 한편 미리 정해진 길이는 선별 또는 분석 대상의 암, 시료의 종류 등에 따라서 25미만의 자연수, 150초과의 자연수 중에 하나일 수 있다. 더 바람직하게는 미리 정해진 길이는 30, 60, 80 또는 90 염기일 수 있다. 또한, 본 발명의 하나의 양태에서 '미리 정해진 길이' 내지 1000의 임의의 자연수일 수 있다.
- [79]
- [80] (d) 단계는 각 서열 정보의 빈도를 계수하는 단계이다.
- [81] 수득된 서열 정보에서 제한효소 절단에 의해 생긴 cohesive end 서열 (HpaII의 경우 CGG) 로 5'말단이 시작하는 각 서열 정보의 빈도를 계수한다. 즉, 한 개 샘플 해독으로 얻은 모든 서열에서 정해진 한 가지 길이 (예를 들어 30)의 서열의 종류(30nt 일 경우 이론적으로 4^{30} 가지 종류의 서열이 가능함) 를 계수하고, 각

[89] 또한, 본 발명은 개체에서 분리한 혈액에서 cfDNA를 분리하는 단계; 분리한 cfDNA (cell free DNA)에 메틸화 민감성 제한효소를 처리하는 단계; 각 단편의 서열을 해독하는 단계; 단편의 N-말단으로부터 미리 정해진 길이의 서열 정보를 획득하는 단계; 각 서열 정보의 빈도를 계수하는 단계; cfDNA에서의 암특이 탈메틸화 마커의 빈도를 산출하여 암으로 판정하는 단계를 포함하는 암 진단 방법에 관한 것이다.

- [90]
- [91] 본 발명의 하나의 양태에서 개체는 암 진단이 필요한 환자이다.
- [92] 본 발명의 하나의 양태에서 미리 정해진 길이는 cfDNA에서의 암특이 탈메틸화 마커와 동일한 길이이다.
- [93] 본 발명의 하나의 양태에서 cfDNA에서의 암특이 탈메틸화 마커는 1 내지 50개, 바람직하게는 3 내지 40개, 더 바람직하게는 5 내지 30개로 이루어진 마커 세트이다.
- [94]
- [95] 또한, 본 발명은 암 진단에 필요한 정보를 제공하기 위하여, 개체의 혈액에서 분리한 cfDNA의 메틸화 민감성 제한효소 단편의 N-말단의 미리 정해진 길이의 서열 정보를 해독하고 분석하는 방법에 관한 것이다.
- [96]
- [97] 또한, 본 발명은 N말단이 메틸화 민감성 제한효소의 인식부위의 점착성 말단의 서열 (예를 들어, CGG의 서열)이며, 연속하여 25염기 내지 150염기의 서열로 이루어지며 (바람직하게는 30염기, 35염기, 40염기, 45염기 또는 50염기)의 염기서열로 이루어지며, 본 발명의 방법에 의해서 선별된 cfDNA에서의 암 특이 탈메틸화 마커를 제공하며, 본 문단에서 점착성 말단의 서열은 ACGTC(서열번호 39), ATCG(서열번호 40), ATCGC(서열번호 41), CCGGA(서열번호 42), CCGGC(서열번호 43), CCGGCC(서열번호 44), CCGGG(서열번호 45), CCGGT(서열번호 46), CCGGY(서열번호 47), CCGGYG(서열번호 48), CG(서열번호 49), CGAA(서열번호 50), CGAT(서열번호 51), CGC(서열번호 52), CGCC(서열번호 53), CGCGC(서열번호 54), CGCGCC(서열번호 55), CGCGT(서열번호 56), CGG(서열번호 57), CGT(서열번호 58), CGTT(서열번호 59), CGWCG(서열번호 60), CGYC(서열번호 61), GCGCC(서열번호 62), GCGCY(서열번호 63), GCGG(서열번호 64), GGCCG(서열번호 65), GGCCGC(서열번호 66), GTACG(서열번호 67), GWCCG(서열번호 68), RYCG(서열번호 69), TCGAC(서열번호 70), TCGAG(서열번호 71) 및 YCGRG(서열번호 72)로 이루어진 군에서 선택될 수 있다. 이 때, 염기의 표시는 표준적인 표기법에 따르며, 예를 들어, A는 아데닌, C는 사이토신, T는 티민, G는 구아닌, Y는 C 또는 T, W는 A 또는 T, R을 A 또는 G를 나타낸다.
- [98]
- [99] 본 발명의 하나의 양태에서 암은 이에 제한되는 것은 아니나 자궁경부암, 폐암, 췌장암, 비소세포성폐암, 간암, 결장암, 골암, 피부암, 두부 또는 경부암, 피부 또는 안구내 흑색종, 자궁암, 난소암, 직장암, 위암, 항문부근암, 결장암, 유방암, 나팔관암종, 자궁내막암종, 질암종, 음문암종, 식도암, 소장암, 내분비선암, 갑상선암, 부갑상선암, 부신암, 연조직 육종, 요도암, 음경암, 전립선암, 방광암, 신장암 또는 수뇨관 암일 수 있다.

[100]

[101] 암 질환 진단을 위한 분석

[102] 본 진단 방법은 특정 대상체에서 병태, 특히 질환의 존재를 진단하거나, 병태의 특징을 규명하거나 (예를 들어, 암의 병기를 결정하거나 또는 암의 이질성을 결정한다), 병태의 치료에 대한 반응을 모니터링하거나, 병태 또는 병태의 후속 과정의 발생 위험을 예측하기 위해 사용될 수 있다. 본 기재 내용은 또한, 특별한 치료 요법의 효능을 결정하는 데 유용할 수 있다. 또 다른 예에서, 특징의 치료 요법은 시간 경과에 따른 암의 유전적 프로파일과 상관이 있을 수 있다. 이러한 상관 관계는 요법을 선택하는 데 유용할 수 있다. 부가적으로, 치료 후에 암에 차도가 있는 것으로 관찰되는 경우, 본 진단 방법은 잔여 질환 또는 질환의 재발을 모니터링하기 위해 사용될 수 있다.

[103] 유전적 데이터는 또한, 특이적 형태의 암의 특징을 규명하기 위해 사용될 수 있다. 암은 종종, 조성과 병기 둘 다에 있어서 이질적이다. 유전적 프로파일 데이터는 특이적 하위 유형의 암을 진단 또는 치료하는 데 중요할 수 있는, 그러한 특이적 하위 유형의 암의 특징 규명을 허용할 수 있다. 이러한 정보는 또한, 특이적 유형의 암의 예후에 관한 대상체 또는 실무자 단서를 제공할 수 있으며, 대상체 또는 실무자가 질환의 진행에 따라 치료 옵션을 채택하도록 허용할 수 있다. 일부 암은 보다 공격적이고 유전적으로 불안정하도록 진행될 수 있다. 다른 암은 양성, 비활성 또는 휴면 상태로 남아있을 수 있다. 본 기재 내용의 시스템 및 방법은 질환 진행을 결정하는데 유용할 수 있다.

[104]

[105] 마커 및 패널

[106] 본 발명은 진단 또는 예측 마커로서 각 마커를 개별적으로 사용하거나, 몇몇 마커를 조합하여 패널 디스플레이 형태로 하여 사용할 수 있고, 몇몇의 마커는 전체적인 패턴 또는 메틸화된 부위의 목록을 통하여 신뢰성 및 효율성을 향상시키는 것을 확인할 수 있다. 본 발명에서 확인된 마커는 개별적으로, 또는 조합된 마커 세트로서 사용될 수 있다. 마커들은 함께 메틸화된 마커의 수 및 그 중요도에 따라 순위를 매길 수 있고, 가중치를 둘 수 있으며, 질환으로 발전할 가능성의 수준을 선정할 수 있다. 이러한 알고리즘은 본 발명에 속한다.

[107]

[108] 기질

[109] 타겟 핵산 부위는 고체 지지체(기질)에 고정된 알려진 프로브와 하이브리다이제이션될 수 있다.

[110] 여기서, "기질"은 물질, 구조, 표면 또는 재료, 비생물학적이고, 합성되고, 무생물, 평면, 구형 또는 특이적 결합, 평편한 표면의 물질을 포함하는 혼합물 수단으로, 하이브리다이제이션 또는 효소 인식 부위 또는 대다수의 다른 인식 부위 또는 표면, 구조 또는 재료로 구성된 수많은 다른 분자 종을 넘어서는 수많은 다른 인식 부위를 포함할 수 있다. 상기 기질은 예를 들면, 반도체,

(유기)합성 메탈, 합성 반도체, 인슐레이터 및 도판트; 금속, 합금, 원소, 화합물 및 미네랄; 합성되고, 분해되며, 에칭되고, 리소그래프되며, 프린트되고 마이크로패브리케이션된 슬라이드, 장치, 구조 및 표면; 산업적, 폴리머, 플라스틱, 멤브레인, 실리콘, 실리콘에이트, 유리, 금속 및 세라믹; 나무, 종이, 카드보드, 먼, 울, 천, 직조 및 비직조 섬유, 재료 및 패브릭일 수 있으나, 이에 한정되는 것은 아니다.

- [111] 몇몇 형태의 멤브레인은 당해 분야에서 핵산 서열에 대하여 부착력을 가진다고 알려져 있다. 이러한 멤브레인의 특이적이고 비제한적인 예로 니트로셀룰로오스 또는 폴리비닐클로라이드, 디아조티즈드(diazotized) 페이퍼 및 상품명 GENESCREEN, 상품명 ZETAPROBE 및 상품명 NYTRAN 등의 상업적으로 사용되는 멤브레인과 같이 유전자 발현 검출용 멤브레인을 들 수 있다. 비드, 글래스, 웨이퍼 및 금속 기질도 포함된다. 이러한 목적물에 핵산을 부착시키는 방법은 당해 분야에서 잘 알려져 있다. 이와 다르게, 액체 상에서도 스크리닝을 수행할 수 있다.

발명의 효과

- [112] 따라서, 본 발명의 방법은 cfDNA에서의 암 특이 탈메틸화 마커를 선별할 수 있으며, 선별된 마커는 암의 진단, 치료 요법에 대한 모니터링, 암 환자의 예후에 필요한 정보를 제공할 수 있어 항암 치료에 유용하게 이용될 수 있다.

도면의 간단한 설명

- [113] 도 1a는 폐암 환자 시료군과 정상인 시료군에서 HpaII로 처리하여 분석한 결과의 하나의 예이며, 도 1b는 이를 도식화한 것이다.
- [114] 도 2a는 유방암 환자 시료군과 정상인 시료군에서 SacII로 처리하여 효소 절단 부위에 매핑된 read의 수를 표준화(z-score)하여 두 개의 시료군 간에 통계적으로 유의하게 차이가 나는 영역을 표현한 것이다. 도 2b는 도 2a 과정에서 추출된 유방암 특이적 마커를 이용하여 기계학습 모델을 만들어 유방암 예측 확률값 (0.0~1.0)을 계산하고 정상군과 유방암군간의 확률값 차이를 도식화한 것이다. 도 2c는 모델 학습을 20회 반복하여 각 테스트 샘플들의 평균 확률값을 통해 ROC(Receiver Operator Characteristic) curve를 도식화하고 AUC(Area Under Curve : 0.0~1.0)값을 표현하였다.
- [115] 도 3a는 유방암 환자 시료군과 정상인 시료군에서 HpaII로 처리하여 분석한 결과의 하나의 예이며, 도 3b는 이를 도식화한 것이다.
- [116] 도 4a는 폐암 환자 시료군과 정상인 시료군에서 SacII로 처리하여 효소 절단 부위에 매핑된 read의 수를 표준화(z-score)하여 두 개의 시료군 간에 통계적으로 유의하게 차이가 나는 영역을 표현한 것이다. 도 4b는 도 4a 과정에서 추출된 폐암 특이적 마커를 이용하여 기계학습 모델을 만들어 폐암 예측 확률값 (0.0~1.0)을 계산하고 정상군과 폐암군간의 확률값 차이를 도식화한 것이다. 도 4c는 모델 학습을 20회 반복하여 각 테스트 샘플들의 평균 확률값을 통해

ROC(Receiver Operator Characteristic) curve를 도식화하고 AUC(Area Under Curve : 0.0~1.0)값을 표현하였다. 도 4d는 폐암의 각 병기(stage)로 분류하여 ROC curve를 도식화 하였다.

발명의 실시를 위한 형태

- [117] 이하, 본 발명의 이해를 돕기 위하여 바람직한 실시예를 제시한다. 그러나 하기의 실시예는 본 발명을 보다 쉽게 이해하기 위하여 제공되는 것일 뿐, 이에 의해 본 발명의 내용이 한정되는 것은 아니다.

[118]

[119] 실험방법

[120] 1. 혈액에서 cfDNA의 분리

- [121] 유방암 환자(n=102), 폐암 환자(n=75) 및 건강한 사람(n=139)을 대상으로 하여 cfDNA 전용 채혈 튜브에 채혈을 하였다. 분리된 혈액은 각각 2000rpm 20분 원심분리하여 혈장(plasma)을 분리하였다. 분리된 혈장 (상층액)을 1.5ml 튜브에 옮긴 후 16000rpm에서 10분간 원심분리하였다. 이 후 Chemagen cfDNA prep 장비를 이용하여 제조사의 지침에 따라서 cfDNA를 분리하였다.

[122]

[123] 2. Library 제작

- [124] 분리된 cfDNA의 말단을 blunt end로 만든 후 A tailing을 유도하였고, 여기에 p7 adapter와 cfDNA의 ligation을 유도하였다. 제한효소인 HpaII를 처리하여 탈메틸화된(demethylated) CCGG부위를 절단하였다. 이 때, 메틸화된 CpG의 경우 HpaII에 의해서 절단되지 않게 된다. 여기에 HpaII sticky end를 가지고 있는 p5 adapter를 첨가하고, HpaII에 의해서 절단된 cfDNA 단편과 ligation을 하였고, 이를 분석 라이브러리로 하였다.

[125]

[126] 3. 서열 정보의 분석

- [127] 각 샘플에 대한 분석 라이브러리를 NGS로 하여 라이브러리에 포함된 각 서열을 서열 정보를 수득하였다. 각 샘플의 해독된 서열에서 제한효소인식서열 (HpaII의 경우 CGG, SacII의 경우 GC)로 시작하는 서열을 선발하였고, 선발된 서열들을 5'부터 일정길이 (ex. 30, 60, 80 등)의 서열 정보를 얻어 정해진 길이별로 서열을 분류하였다. 각 샘플에서 분류된 서열들의 빈도를 계수하고 샘플간 비교를 위해 정규화하였다. 각 서열후보에 대해 정상샘플군 대비 암 샘플군에서 유의하게 높게 해독된 서열들을 확인하였으며, 이 때 해당 서열은 암 샘플군에서 특이적으로 탈메틸화되어 있는 것이므로 탈메틸화 마커(HpaII의 경우에 해당; 메틸화 부위를 절단하는 제한효소의 경우에는 메틸화 마커)로 선별될 수 있었다. 정해진 마커에 대해 각 샘플별로 평균값 (DHM score)를 구하고, 암샘플과 정상샘플을 구분할 수 있는 DHM score의 기준값을 정한다. 이 DHM score는 시료의 판정에 이용하였다.

[128]

[129] **4. 시료의 검사 및 암발생 여부 판별**

[130] 미지의 시료를 상기 기술한 실험방법 및 분석방법에 따라 해독과 분석을 실행하였다. 상기 서열정보의 분석 (3번 항목)에서 선발된 마커에 해당하는 값의 평균인 DHM score를 구하고, 정해진 DHM 기준값 이상일 경우 암이 있는 것으로 판정하였다.

[131]

[132] **실시예 1: 유방암 cfDNA에 대한 암 특이 탈메틸화 마커의 선별과 이를 이용한 판정 정확도 테스트**

[133] 34개 유방암 샘플군과 53개의 정상 샘플에서 분리한 cfDNA를 메틸화 민감성 제한효소중 한가지인 HpaII로 절단하고 상기 기술한 실험방법에 따라 해독, 분석하였다. 해독된 서열중 CGG로 시작하는 서열의 앞부분 80nt의 서열을 마커 후보로 하여 계수하고 비교하였다. 유방암에서의 각 마커에 해당하는 값들의 평균이 5이상이면서 정상군의 평균값보다 10배 이상의 값을 가지는 173개를 마커로 정하였고, 각 샘플에서 이들 173개의 값의 평균인 DHM score를 구하였다.

[134] 그 결과, 도 1에서 보듯이 유방암과 정상 샘플들의 각 마커에 대한 정규화된 점수를 기록한 테이블을 만들었으며, 점수를 쉽게 보기 위해 높은 수를 빨강계, 낮은 수를 파랑계 표현하는 heatmap으로 표현하였다. 도 1a는 173개 마커중 상위 일부를 나타낸 것인데, 선발된 마커에서, 유방암 샘플에서는 세 개 샘플을 제외한 31개 샘플이 일정 수치 이상의 값을 가지는 반면 정상 샘플에서는 모든 샘플이 0에 가까운 값을 가지는 것을 볼 수 있었다.

[135] 도 1b는 도 1a에서 표현된 각 샘플의 173개 마커 값들의 평균인 DHM score를 막대그래프로 표현한 것이다. DHM score 1을 기준으로 유방암과 정상샘플이 명확히 구분이 됨을 볼 수 있다.

[136] 이와 같이 유방암군과 정상군의 DHM score를 비교한 결과 DHM score를 1을 기준으로 할 경우 34개 유방암 샘플 중 31개를 유방암으로 판정하고 정상 53개 중 53개 모두를 정상으로 판정할 수 있었다. 정확도를 산출한 결과, 민감도는 91.2%, 특이도는 100% 이었다 (도 1a 및 도 1b 참조).

[137]

[138] **실시예 2: SacII로 처리한 유방암 cfDNA에 대한 암 특이 탈메틸화 마커의 선별과 이를 이용한 판정 정확도 테스트**

[139] 102개 유방암 샘플군과 139개의 정상 샘플에서 분리한 cfDNA를 메틸화 민감성 제한효소인 SacII로 절단하고 상기 기술한 실험방법에 따라 해독, 분석하였다. 해독된 서열중 GC로 시작하는 서열의 앞부분 80nt의 서열을 마커 후보로 하여 계수하고 비교하였다.

[140] 각 마커는 IQR(InterQuartile Range) 평균값을 통해 정규화를 하고 이를 Z-score를 통해 표준화를 수행하여 시퀀싱간에 나타날 수 있는 차이를 감소시킨다. 이후 각 마커에 대해서 유방암군과 정상군간에 t-test를 수행하여

p-value가 특정 임계치(예: 10^{-5}) 이하에 해당하는 마커를 선정하고, 선정된 마커를 통해 최종 DHM Score를 계산한다. 최종 Score는 선정된 마커에 대해 샘플별로 해당 값을 단순히 더하여 계산할 수 있고, 로지스틱 회귀분석과 같은 기계학습의 분류모델을 만들어서 예측 확률 값으로 계산할 수 있다.

- [141] 그 결과, 도 2a에서 보듯이 유방암과 정상 샘플들의 각 마커에 대한 정규화/표준화된 값을 기록한 테이블을 만들었으며, 점수를 쉽게 보기 위해 높은 수를 빨강계, 낮은 수를 초록색으로 표현하는 heatmap으로 표현하였다. 도 2b는 선정된 마커를 통해 기계학습 예측 모델을 만들어서 0에서 1까지 확률값으로 결과치를 만들어서 유방암군과 정상군에서 확률값의 분포 차이가 명확하게 나타나는 것을 확인할 수 있었다.
- [142] 도 2c에서는 기계학습 모델 테스트 방법 중 K-Fold Cross Validation을 사용하여 각 Cycle마다 학습과 테스트군을 8:2로 Random 추출을 하고, 이와 같은 동작을 20회 반복하여 하나의 샘플이 20회의 서로 다른 학습 데이터를 통해 결과값이 계산되고, 이를 평균값을 취해 ROC(Receiver Operating Characteristic) curve를 그려 성능을 측정하였다.
- [143] 이와 같이 유방암군과 정상군의 DHM score를 비교한 결과 AUC가 0.9492이고, 특이도 100% 기준에서 민감도는 70.87%를 가짐을 알 수 있었다 (도 2c 참조).
- [144]
- [145] **실시예 3: 폐암 cfDNA에 대한 암 특이 탈메틸화 마커의 선별과 이를 이용한 판정 정확도 테스트**
- [146] 11개 샘플의 폐암군과 53개의 정상샘플군을 30nt의 길이로 비교했을 때, 정상군의 평균값보다 폐암군에서 5배 이상의 값을 가지는 마커는 198개, 10배 이상의 값을 가지는 마커는 157개였다. 각 샘플에서 이들 198개의 값을 모두 더해서 DHM score를 구하였다.
- [147] 그 결과, 도 3a에서 보듯이 폐암과 정상 샘플들의 각 마커에 대한 정규화된 점수를 기록한 테이블을 만들었으며, 점수를 쉽게 보기 위해 높은 수를 빨강계, 낮은 수를 파랑계 표현하는 heatmap으로 표현하였다. 도 3a은 198개 마커중 상위 일부를 나타낸 것이다. 선별된 마커에서, 폐암 샘플에서는 세 개 샘플을 제외한 8개 샘플이 기준치 이상의 값을 가지는 반면 정상 샘플에서는 모든 샘플이 기준치 4보다 낮은 3 이하의 값을 가지는 것을 볼 수 있었다.
- [148] 도 3b는 도 3a에서 표현된 각 샘플의 198개 마커 값들의 평균인 DHM score를 막대그래프로 표현한 것이다. DHM score 4를 기준으로 폐암과 정상샘플이 명확히 구분이 됨을 볼 수 있었다.
- [149] 이와 같이 폐암군과 정상군의 DHM score를 비교한 결과 DHM score를 4을 기준으로 할 경우 11개 폐암 샘플 중 8개를 폐암으로 판정하고 정상 53개 중 53개 모두를 정상으로 판정할 수 있었다. 정확도를 산출한 결과, 민감도는 72.7%, 특이도 100% 이었다 (도 3a 및 도 3b 참조).

- [150]
- [151] 실시예 4: **SacII**로 처리한 폐암 cfDNA에 대한 암 특이 탈메틸화 마커의 선별과 이를 이용한 판정 정확도 테스트
- [152] 75개 폐암 샘플군과 129개의 정상 샘플에서 분리한 cfDNA를 메틸화 민감성 제한효소중 한가지인 **SacII**로 절단하고 상기 기술한 실험법에 따라 해독, 분석하였다. 해독된 서열중 GC로 시작하는 서열의 앞부분 80nt의 서열을 마커 후보로 하여 계수하고 비교하였다.
- [153] 각 마커는 IQR(InterQuartile Range) 평균값을 통해 정규화를 하고 이를 Z-score를 통해 표준화를 수행하여 시퀀싱간에 나타날 수 있는 차이를 감소시킨다. 이후 각 마커에 대해서 폐암군과 정상군간에 t-test를 수행하여 p-value가 특정 임계치(예: 10^{-5}) 이하에 해당하는 마커를 선정하고, 선정된 마커를 통해 최종 DHM Score를 계산한다. 최종 Score는 선정된 마커에 대해 샘플별로 해당 값을 단순히 더하여 계산할 수 있고, 로지스틱 회귀분석과 같은 기계학습의 분류모델을 만들어서 예측 확률 값으로 계산할 수 있다.
- [154] 그 결과, 도 4a에서 보듯이 폐암과 정상 샘플들의 각 마커에 대한 정규화/표준화된 값을 기록한 테이블을 만들었으며, 점수를 쉽게 보기 위해 높은 수를 빨강색, 낮은 수를 초록색으로 표현하는 heatmap으로 표현하였다. 도 4b는 선정된 마커를 통해 기계학습 예측 모델을 만들어서 0에서 1까지 확률값으로 결과치를 만들어서 폐암군과 정상군에서 확률값의 분포 차이가 명확하게 나타나는 것을 확인할 수 있었다.
- [155] 도 4c에서는 기계학습 모델 테스트 방법 중 K-Fold Cross Validation을 사용하여 각 Cycle마다 학습과 테스트군을 8:2로 Random 추출을 하고, 이와 같은 동작을 20회 반복하여 하나의 샘플이 20회의 서로 다른 학습 데이터를 통해 결과값이 계산되고, 이를 평균값을 취해 ROC(Receiver Operating Characteristic) curve를 그려 성능을 측정하였다. 도 4d는 폐암의 각 병기에 따라 분류된 샘플의 정확도를 보여준다.
- [156] 이와 같이 폐암군과 정상군의 DHM score를 비교한 결과 AUC가 0.8837이고, 특이도 100% 기준에서 민감도는 41.67%를 가짐을 알 수 있었다 (도 4c 참조).
- 산업상 이용가능성**
- [157] 이상 살펴본 바와 같이 본 발명의 방법은 cfDNA에서의 암 특이 탈메틸화 마커를 선별할 수 있으며, 선별된 마커는 암의 진단, 치료 요법에 대한 모니터링, 암 환자의 예후에 필요한 정보를 제공할 수 있어 항암 치료에 유용하게 이용될 수 있다.