

명세서

발명의 명칭: 간암 재발 예측용 DNA 메틸화 마커 및 이의 용도

기술분야

- [1] 본 발명은 간암의 예후에 따라 계놈 DNA의 메틸화 수준이 상이한 점을 이용한 간암의 재발 예측용 DNA 메틸화 마커 및 이의 용도에 관한 것으로, 구체적으로는 간암 환자의 치료 후 정상 간 조직에서 간암의 재발 여부를 예측하기 위한 DNA 메틸화 마커에 관한 것이다.

배경기술

- [2] 간암(liver cancer)은 크게 간세포 자체로부터 발생한 원발성 간암(간세포암; hepatocellular carcinoma, hepatoma)과 다른 조직의 암이 간으로 전이되어 온 전이성 간암으로 구분할 수 있다. 간암의 약 90% 이상은 원발성 간암이고, 흔히 간암이라 함은 원발성 간암을 지칭하는 것으로 이해된다. 통계청이 2018년 9월에 발표한 '2017 사망원인통계'에 따르면 2017년 인구 10만명 기준 암으로 사망한 사람은 153.9명으로 전년 대비 1.0명(0.6%) 증가했다. 암 사망률은 폐암(35.1명), 간암(20.9명), 대장암(17.1명), 위암(15.7명), 췌장암(11.3명) 순으로 높았다.
- [3] 다른 암과 마찬가지로 간암의 치료도 크게 근치적 치료와 보존적 치료가 있다. 여기서 근치란 암 자체를 완전히 없애는 치료를 의미하며 간 절제 수술, 간이식, 고주파 열치료 등이 있다. 또한 이러한 치료가 불가능한 경우 경동맥 화학 색전술, 방사선 치료, 항암치료를 하게 된다. 그러나 이러한 치료로는 완치가 거의 어렵기 때문에 조기 발견을 통해 근치적 치료를 하는 것이 검진의 목적이다. 간절제 수술의 경우 약 50%의 환자에서 암 재발을 경험하게 되는데, 그 이유는 절제되고 남아있는 간도 정상적인 간이 아닌 간염에 노출되어 있기 때문이다.
- [4] 간암의 예후(prognosis)는 간암이 진단된 이후에 간암의 완치가가능성, 치료 후 재발가능성, 환자의 생존가능성 등 간암에 따른 환자의 각종 상태를 예측하는 것을 의미하며, 이는 질병의 심각성, 진단 시점, 치료경과 등의 다양한 조건에 의존하여 달라진다. 간암은 그 예후에 따라 적합하게 다양한 치료방법이 적용되어야만 효율적인 치료가 가능하다. 예를 들어, 예후가 양호할 것으로 예측되는 환자들에 대해서는 공격적인 화학치료나 수술, 방사선 치료 등 환자에게 심각한 부작용을 야기할 가능성이 있는 위험한 치료방법은 지양할 필요가 있고, 상대적으로 온건하고 보존적이며 안전한 치료방법을 선택하여야 한다. 반대로 예후가 불량할 것으로 예측되는 환자들에 대해서는 화학치료나 수술, 방사선 치료와 같은 치료방법을 적극적으로 동원하여 생존의 기간이나 확률을 높이려고 시도하여야 한다.
- [5] 그러나 간암 환자의 예후를 정확하게 예측하는 것은 종래의 기술로는 매우 어렵다. 예후의 정확한 예측을 위해서는 환자들을 위험군별로 분류하는

분석방법이 필요한데, 현재까지는 간암의 예후를 정확하게 예측할 수 있는 수단 없이 진단시의 임상병리학적인 간암의 단계와 1차적 외과 치료에만 의존하여 예후를 판단하고 있는 실정이다.

- [6] 본 발명자들은 이러한 문제를 해결하고자 노력한 결과, 간암 조직의 DNA 메틸화 데이터를 근거로 간암 환자의 예후를 예측할 수 있는 방법을 마련하였다.

발명의 상세한 설명

기술적 과제

- [7] 본 발명의 목적은 DNA 메틸화 마커를 포함하는 간암의 재발 예측용 조성물과 이를 이용한 간암의 재발 가능성을 예측하는 방법을 제공하는 것이다.

과제 해결 수단

- [8] 상기 목적을 달성하기 위하여, 본 발명의 일 양상은 서열번호 1 내지 4로 표시되는 DNA 메틸화 마커로 이루어진 군에서 선택되는 하나 이상의 DNA 메틸화 마커를 포함하는 간암의 재발 예측용 조성물을 제공한다.
- [9] 본 발명에서, 상기 서열번호 1 내지 4로 표시되는 DNA 메틸화 마커는 각각 CG 유전자좌 기호(Illumina® Infinium HumanMethylation 850K 비드 어레이 CpG ID)로 표시되는 DNA 메틸화 마커인 cg21325760, cg10544510, cg06702718 및 cg15997204를 의미한다.
- [10] 본 명세서에 사용된 용어, "간암(liver cancer)"은 간 조직에 발생하는 암을 총칭하는 용어로, 간세포 자체에서 발생한 원발성 간암 (간세포암; hepatocellular carcinoma, hepatoma)과 다른 조직의 암이 간으로 전이되어 온 전이성 간암으로 구분할 수 있다. 간암 치료에는 간 절제술, 간 이식, 고주파열 치료술, 경동맥 화학 색전술 등 다양한 치료법이 사용되고 있으나, 5년 생존율이 33.6%(2011년 내지 2015년 평균; 보건복지부 통계) 밖에 되지 않고, 재발률 또한 70%에 이른다. 다른 암종에 비해 간암은 5년 생존율은 낮고, 재발률은 상당히 높기 때문에 재발 여부를 예측할 수 있는 방법이 필요하다.
- [11] 본 명세서에 사용된 용어, "DNA 메틸화(DNA methylation)"는 게놈 DNA에서 시토신 염기의 C5-위치에 메틸기가 공유결합한 것을 의미한다. 메틸화 수준은 예를 들면 모든 게놈 영역 및 일부 비-게놈 영역 내의 DNA 염기서열에 존재하는 메틸화의 양을 의미하며, 본 발명에서는 서열번호 1 내지 4로 표시되는 DNA 메틸화 마커의 메틸화 정도를 의미한다. 상기 서열번호 1 내지 4의 DNA 메틸화 마커에서 메틸화는 서열 전체에 걸쳐 일어날 수 있다.
- [12] 본 발명에서, 상기 간암의 재발 예측용 조성물은 상기 DNA 메틸화 마커 중에서 1개 이상, 2개 이상, 3개 이상 또는 4개를 포함하는 조성물일 수 있다. 구체적으로 상기 간암의 재발 예측용 조성물은 DNA 메틸화 마커 cg21325760, cg10544510, cg06702718 및 cg15997204를 각각 한 개씩 포함할 수 있고, 두 개의 조합인 cg21325760 및 cg10544510; cg21325760 및 cg06702718; cg21325760 및 cg15997204; cg10544510 및 cg06702718; cg10544510 및 cg15997204; 또는

cg06702718 및 cg15997204 조합을 포함할 수 있다. 또한, 상기 간암의 재발 예측용 조성물은 세 개의 DNA 메틸화 마커를 포함할 수 있으며, 예를 들어 cg21325760, cg10544510 및 cg06702718; cg21325760, cg10544510 및 cg15997204; cg21325760 cg06702718 및 cg15997204; 또는 cg10544510, cg06702718 및 cg15997204의 조합을 포함할 수 있다. 아울러, 상기 간암의 재발 예측용 조성물은 DNA 메틸화 마커 cg21325760, cg10544510, cg06702718 및 cg15997204 모두를 포함할 수 있다.

- [13] 본 발명에서, 상기 DNA 메틸화 마커 cg15997204는 MYT1L 유전자에 존재하는 마커로서 MYT1L (Myelin Transcription Factor 1 Like) 유전자는 C-C-H-C 징크 핑거 단백질에 속하는 전사 인자를 코딩한다. TCGA(The Cancer Genome Atlas)에 공개된 데이터에 따르면 다양한 암에서 MYT1L이 발현되나, 이 유전자의 메틸화 수준과 간암의 재발 여부와의 상관 관계에 대해서는 아직까지 명확히 알려진 바가 없다.
- [14] 본 발명에서, 상기 DNA 메틸화 마커 cg21325760은 MAGEL2 (MAGE(melanoma-associated antigen) family member L2) 유전자에 존재하고, TCGA 데이터에 따르면 정소암에서 많이 발현되나, 이 유전자의 메틸화 수준과 간암의 재발 여부와의 상관 관계에 대해서는 아직까지 명확히 알려진 바가 없다.
- [15] 본 발명의 일 구체예에서, 본 발명자들은 간암 조직에서 게놈 DNA를 분리한 후 Illumina® Infinium HumanMethylation 850K 비드 어레이로 메틸화 수준을 확인하였다. 그 결과, 메틸화 수준에 따라 간암 조직이 두 그룹으로 분류되는 것을 확인하였고, 상기 두 그룹이 간암의 재발률에 있어 현저한 차이가 있음을 알 수 있었다. 이후 두 그룹에서 메틸화 수준이 현저히 차이나는 DNA 메틸화 마커를 랜덤 포레스트 등의 기계학습 기법으로 확인하여 상기 DNA 메틸화 마커를 발굴하였다(표 3 참조). 또한, 하나의 DNA 메틸화 마커보다 4개의 DNA 메틸화 마커를 같이 사용하는 경우, 간암의 재발 예측의 정확도가 1에 수렴하는 것을 확인하여 본 발명을 완성하였다(도 5 및 표 2).
- [16] 본 발명에서, 상기 간암은 간세포 자체로부터 발생한 간세포암 또는 다른 조직의 암이 간으로 전이되어 온 전이성 간암일 수 있으나, 바람직하게는 간세포암일 수 있다. 간세포암(hepatocellular carcinoma)은 간에서 발생하는 악성 종양의 약 90%를 차지하며, 우리나라와 일본, 동남아시아, 중국 등지에서 많이 발생한다. 대부분 간경변증이 있는 상태에서 발생하나, 일부는 만성 B형 또는 C형 간염으로 인해 발생하기도 한다.
- [17] 본 명세서에 사용된 용어, "재발(relapse)"은 간암 환자에서 간암을 치료한 후 간에서 다시 비정상적인 세포 증식, 즉, 간암이 발생하는 것을 의미한다. 상기 간암 환자의 치료 방법은 간 절제술, 고주파열 치료술, 경동맥 화학 색전술, 경피적 에탄올 주입술 등이 사용될 수 있으나, 이에 제한되지 않는다.
- [18] 본 발명에서, 상기 간암의 예후 예측용 조성물은 DNA 메틸화 마커의 메틸화 수준 확인에 필요한 시약 등을 추가로 포함하는 간암의 재발 예측용 키트 형태로

구현될 수 있다.

- [19]
- [20] 본 발명의 일 양상은 하기 단계를 포함하는 간암의 재발 예측 방법을 제공한다:
- [21] (a) 분석 대상 간 조직에서 게놈 DNA를 분리하는 단계;
- [22] (b) 서열번호 1 내지 4로 표시되는 DNA 메틸화 마커로 이루어진 군에서 선택되는 하나 이상의 DNA 메틸화 마커의 메틸화 수준을 검출하는 단계; 및
- [23] (c) 상기 검출된 DNA 메틸화 마커의 메틸화 수준을 정상 간 조직에서 검출한 DNA 메틸화 마커의 메틸화 수준과 비교하는 단계.
- [24] 본 발명에서, 상기 서열번호 1 내지 4로 표시되는 DNA 메틸화 마커는 각각 CG 유전자좌 기호(Illumina® Infinium HumanMethylation 850K 비드 어레이 CpG ID)로 표시되는 DNA 메틸화 마커인 cg21325760, cg10544510, cg06702718 및 cg15997204를 의미한다.
- [25] 본 발명의 일 구체예에서, 상기 분석 대상 간 조직은 분석 대상 개체에서 분리된 간암 조직, 간세포암 조직, 전이성 간암 또는 간암이 전이된 부위의 전이암 조직일 수 있으며, 바람직하게는 간세포암 조직일 수 있다. 상기 분석 대상 간 조직을 분석 대상 개체에서 분리하는 방법은 본 발명이 속하는 기술분야에 알려져 있는 경피생검, 경정맥 생검 등의 방법을 사용할 수 있다.
- [26] 본 발명의 일 구체예에서, 상기 간암의 예후 예측 방법은 단계 (b)에서 최소 2개 이상의 DNA 메틸화 마커의 메틸화를 확인하는 방법일 수 있다. 본 발명자들은 간암의 재발 여부를 예측할 수 있는 4개의 DNA 메틸화 마커(프로브)를 발굴하였고(표 2 참조), 마커의 수에 따른 간암 재발 예측의 성능을 평가하였다. 평가 결과, 마커가 2개 이상인 경우 정확도, 민감도 등의 성능이 0.9에 수렴하고 마커의 수가 3개 이상인 경우에는 간암 재발 예측 성능이 1에 수렴하는 것을 확인하였다(도 5).
- [27] 이 결과로부터 본 발명의 DNA 메틸화 마커는 단일로 사용되어도 간암 예후 예측 성능이 우수하나, 복수개 사용하는 경우에는 그 성능이 현저히 우수한 것을 알 수 있다. 따라서, 본 발명의 DNA 메틸화 마커는 단일 또는 복수개가 조합되어 간암 재발 예측 용도로 사용될 수 있다.
- [28] 구체적으로, 본 발명의 DNA 메틸화 마커는 cg21325760, cg10544510, cg06702718 및 cg15997204가 각각 단일의 마커로 사용될 수 있고, 두 개 마커의 조합 (cg21325760 및 cg10544510; cg21325760 및 cg06702718; cg21325760 및 cg15997204; cg10544510 및 cg06702718; cg10544510 및 cg15997204; 또는 cg06702718 및 cg15997204) 또는 세 개 마커의 조합(cg21325760, cg10544510 및 cg06702718; cg21325760, cg10544510 및 cg15997204; cg21325760 cg06702718 및 cg15997204; 또는 cg10544510, cg06702718 및 cg15997204)으로 사용될 수 있다. 또한, 상기 네 개 마커 모두를 조합하여 사용할 수 있다.
- [29] 본 발명의 일 구체예에서, 상기 간암의 예후 예측 방법은 단계 (c) 이후 분석 대상 샘플에서 확인한 DNA 메틸화 마커의 메틸화 수준이 정상 간 조직의 메틸화

수준보다 낮은 경우 간암의 재발 가능성이 높은 것으로 판별하는 단계를 추가로 포함할 수 있다.

[30] 본 발명에서, DNA 메틸화 마커의 메틸화 수준을 확인하는 방법은 당업계에 알려진 PCR, 메틸화 특이 PCR(methylation specific PCR), 실시간 메틸화 특이 PCR(real time methylation specific PCR), 메틸화 DNA 특이적 결합 단백질을 이용한 PCR, 정량 PCR, DNA 칩, 파이로시퀀싱, 상업적으로 판매되는 칩 등으로 수행될 수 있으나, 이에 제한되는 것은 아니다.

[31] 본 발명자들은 간암이 재발한 환자(MG2)에서는 상기 DNA 메틸화 마커의 메틸화 수준이 정상 간 조직(N)보다 현저히 낮은 것을 확인하였는바(도 3, 7 및 8), 상기 DNA 메틸화 마커는 간암의 재발 예측 용도로 유용하게 사용될 수 있다.

발명의 효과

[32] 본 발명의 간암 재발 예측용 DNA 메틸화 마커를 사용하면 간암의 재발 여부를 용이하게 예측할 수 있고, 간암 재발 위험이 높은 환자에게 보다 개인화된 치료법을 제공하고 불필요한 과다 치료를 피하기 위한 임상 정보를 제공할 수 있다.

도면의 간단한 설명

[33] 도 1은 간세포암 샘플(A; n=140) 및 정상 간 샘플(B; n=96)에서 분리한 DNA의 메틸화 수준을 확인한 결과를 보여준다.

[34] 도 2는 간세포암 샘플(n=140)을 예후에 따라 메틸화 그룹 1(MG1; n=81) 및 메틸화 그룹 2(MG2; n=59)로 분류한 후 DNA의 메틸화 패턴에 따라 각 샘플을 컨센서스 클러스터링 방법으로 분류한 결과를 보여준다.

[35] 도 3은 메틸화 그룹 1(MG1) 및 메틸화 그룹 2(MG2)의 무-재발 생존기간(recurrence free survival, RFS)을 확인한 결과를 보여준다

[36] 도 4는 메틸화 그룹 1(MG1) 및 메틸화 그룹 2(MG2)에서 변수 중요도 플롯(variable importance plot, varImpPlot)을 확인한 후(A), 메틸화 그룹 분류에 관여하는 상위 50개 프로브를 선별하는 과정을 보여준다(B).

[37] 도 5는 메틸화 그룹 분류에 관여하는 프로브의 개수에 따른 간암 예후 예측 효율을 AUC(area under the curve) 값으로 확인한 결과를 보여준다.

[38] 도 6은 정상 간 샘플(N), 메틸화 그룹 1(MG1) 및 메틸화 그룹 2(MG2)에서 메틸화 그룹 분류에 관여하는 프로브의 메틸화 수준을 확인한 결과를 보여준다.

[39] 도 7은 메틸화 그룹 분류에 관여하는 프로브로 자체적으로 생산한 간암 검증 코호트 데이터를 분류한 결과를 나타낸다.

[40] 도 8은 메틸화 그룹 분류에 관여하는 프로브로 TCGA(the cancer genome atlas)에 공개된 간암 데이터를 분류한 결과를 나타낸다.

발명의 실시를 위한 형태

[41] 이하 하나 이상의 구체예를 실시예를 통하여 보다 상세하게 설명한다. 그러나, 이들 실시예는 하나 이상의 구체예를 예시적으로 설명하기 위한 것으로 본

발명의 범위가 이들 실시예에 한정되는 것은 아니다.

[42]

[43] 실험방법

[44] 1. 간세포암 환자 모집

[45] 본 발명과 관련하여 수행된 연구는 서울대학교 병원(Seoul National University of Hospital) 검토위원회의 승인을 받고, 모든 환자로부터 서면 동의서를 받아 진행하였다. 2011년부터 2016년까지 서울대학교 병원 외과에서 간세포암 (hepatocellular carcinoma; 이하, HCC로 기재함) 절제술을 받은 184명의 환자를 모집하고, 수술 후 CT 스캔을 통해 깨끗한 미세 경계(microscopic margins) 및 종양의 완전 절제를 확인하였다. 상기 환자들은 정기적으로 병원을 방문하여 혈청 종양 마커, 알파-태아단백(α -fetoprotein, AFP) 수치, 조영 CT 스캔(contrast CT scans) 또는 자기공명영상(magnetic resonance imaging, MRI)을 포함하는 표준 프로토콜을 기반으로 3개월마다 재발(recurrence) 여부를 모니터링 받았다. 재발 여부는 조기(≤ 1 년) 및 후기(> 1 년) 재발로 분류하였으며, HCC 절제술 후 1년 이내에 재발한 환자는 제외시켰다. 인구통계학적 데이터와 임상병리학적 특징은 병원 차트에서 확인하였고, 상기 환자들은 절제술 이후 사망할 때까지 추적조사를 받았다. 1년 이내에 재발한 환자를 제외한 이유는 조기 재발의 경우 HCC 절제술이 잘못되었거나, 간암 이외의 이유로 재발하는 경우가 많아 암 치료 후 순수히 정상 간에서 새로운 암이 재발하는 데 관여하는 마커를 발굴하기 위한 본 발명의 목적에 부합하지 않기 때문이다.

[46]

[47] 2. 메틸화 데이터 생성

[48] 상기 182명의 HCC 환자 중에서 수술 후 1년 이내에 HCC가 재발한 42명은 분석에서 제외하였다. 남은 140명의 환자에서 수득한 냉동 HCC 샘플로부터 제조사의 지시에 따라 MagListo™ 5M 게놈 DNA 추출 키트(Bioneer)를 사용하여 DNA를 추출하였다. 각 샘플마다 500 ng의 DNA를 사용하여 제조사의 표준 프로토콜에 따라 Illumina iScan System (Illumina, CA, USA)과 함께 Illumina Infinium HumanMethylation EPIC 850K BeadChip (Illumina) 분석을 수행하였다. 상기 Illumina Infinium HumanMethylation EPIC BeadChips은 모든 단일 뉴클레오타이드의 분해능 (resolution)을 위해 85만 개 이상의 CpG 사이트에 대한 DNA 메틸화 값을 포함한다.

[49]

[50] 3. DNA 메틸화 데이터 가공

[51] 상기 2에서 Illumina Infinium HumanMethylation EPIC BeadChips으로 얻은 데이터에 대하여 공개 소프트웨어인 R 플랫폼 (<https://cran.r-project.org/bin/windows/base/>; 버전3.3.2)에서 minfi R 패키지를 사용하여 메틸화 데이터에 대한 사전-가공(pre-processing)을 수행하였다. 구체적으로 HCC 샘플의 원본(raw) IDAT 파일을 가공하고, 배경 보정 및 염료 편차를 정정하였다. 이후 메틸화 및

비메틸화(unmethylated) 프로브의 신호 강도를 측정하여 각 프로브의 DNA 메틸화 값을 0 (비메틸화) 내지 1 (완전 메틸화) 범위의 b-값(b-value)으로 정량화하였다. 다음으로 모든 샘플에 대해 데이터 품질 관리를 수행하였으며, 140개 샘플 모두 품질 관리 기준을 통과했다. 마지막으로, 메틸화 데이터는 배치 효과 문제(batch effect problem)를 줄이기 위해 기능적 정규화 방법(functional normalization method)을 사용하여 정규화시켰다.

[52]

[53] **4. DNA 메틸화 값(methylation values)**

[54] 메틸화 b-값은 통계 분석을 위해 하기 수학적 식 1에 따라 M-값으로 변환시켰다.

[55] [수학적 식 1]

[56] $M\text{-값} = \log((b\text{-값}) / (1 - (b\text{-값})))$

[57] M-값은 컨센서스 클러스터링(consensus clustering)에 사용하고, b-값은 히트맵 시각화 및 박스 플롯 생성에 사용하였다.

[58]

[59] **5. 컨센서스 클러스터링**[60] 컨센서스 클러스터링은 ConsensusClusterPlus (<https://bioconductor.org/packages/release/bioc/vignettes/ConsensusClusterPlus/inst/doc/ConsensusClusterPlus.pdf>)를 사용하여 HCC 소그룹을 확인하기 위해 수행하였다. 입력값(input)으로는 상위 3,000개의 가장 가변적인 프로브(중앙값 절대 편차(median absolute deviation, MAD))(MAD 3K)를 사용하였다.

[61]

[62] **6. 예후 마커(prognostic marker) 선별**

[63] 랜덤 포레스트(Random Forest) 방법은 비선형 효과(nonlinear effects)를 모델링하기 위해 개발된 기계 학습 알고리즘을 기반으로 한다. 본 발명자들은 R 플랫폼의 랜덤 포레스트 R 패키지를 사용하여 좋거나 나쁜 예후를 갖는 HCC 그룹을 구별할 수 있는 프로브를 발굴하였고, HCC 중앙 샘플을 4개의 트레이닝 코호트(cohort)와 1개의 테스트 코호트로 무작위로 나누어 5배 교차 검증(five-fold cross validation)을 수행하였다. 수행 결과, 본 발명자들은 랜덤 포레스트 모델을 구축하였고, 각 그룹의 예상 비율은 혼돈 행렬(confusion matrix)(예측된 샘플 vs 관찰된 샘플)에 의해 예측되었고, 각 모델의 성능은 곡선 아래 면적(area under the curve, AUC), 민감도 및 특이도 값에 기초하여 측정하였다.

[64] 본 발명자들은 지니 인덱스(Gini index), 프로브 불순도(impurity)의 전체적 감소 및 랜덤 포레스트로 측정한 변수 중요도의 도트 차트를 고려하여 각 프로브의 중요도 값을 측정하였다. 그 결과, 본 발명자들은 HCC 샘플을 메틸화 그룹 1 및 2(이하, 각각 MG1 및 MG2로 기재함)로 분류할 수 있는 상위 50개의 프로브를 발굴하였다(도 4). 이후 하나의 프로브와 상위 50개 프로브의 성능 효율을 비교하면서 AUC 값을 고려하여 최대 효율을 지닌 프로브의 최소 개수를

확인하였다.

[65]

[66] **7. 유전자 온톨로지 및 경로 분석 (Gene ontology and pathway analysis)**

[67] 본 발명자들은 innateDB (<http://www.innatedb.com/>)를 사용하여 다양한 경로 검색 및 유전자 온톨로지 분석을 수행하였다. 상기 innateDB에서는 유전자 온톨로지 생물학적 과정(Gene Ontology biological processes), KEGG (Kyoto Encyclopedia of Genes and Genomes) 및 리액툼 리소스(Reactome resources)를 사용하였다. 유의미한 컷오프 기준은 $p\text{-값} < 0.05$ 이었다.

[68]

[69] **9. 통계 분석**

[70] 범주형 변수(categorical variable)의 통계 분석은 피어슨 카이(χ^2) 검정 또는 피셔 정확 검증을 사용하여 수행하였다. 무-재발 생존(relapse-free survival, RFS)의 중간 기간은 카플란-메이어 방법으로 계산하였고, 서로 다른 그룹 간의 비교는 로그 순위법(log-rank tests)으로 수행하였다.

[71]

[72] **실험 결과**

[73] **1. 환자 선별 및 기본 특성**

[74] 182명의 HCC 환자 중에서 수술 후 1년 이내에 HCC가 재발한 42명은 분석에서 제외하였다. 140개 HCC 샘플의 메틸화를 확인한 결과, HCC 샘플은 샘플 간에 메틸화의 변이가 큰 반면, 정상 조직은 메틸화가 거의 일정하게 나타나는 것을 알 수 있었다(도 1).

[75]

[76] **2. MG1과 MG2의 메틸화 프로파일 비교**

[77] 140개 HCC 샘플을 DNA 메틸화 패턴에 따라 분류한 결과, 메틸화 그룹 1($n=81$) 및 그룹 2($n=59$)(이하, 각각 MG1 및 MG2로 기재함)로 나뉘는 것을 확인하였다(도 2의 A). 또한, MG2의 메틸화 프로파일은 비종양성 간조직의 메틸화 프로파일과 가장 상이하였고, MG1의 메틸화 프로파일은 비종양성 간조직의 메틸화 프로파일과 유사하였다(도 2의 B). MG1은 무-재발 생존(RFS) 기간이 긴 반면, MG2는 무-재발 생존이 짧아서 예후가 나쁜 것을 알 수 있었다(도 3)

[78]

이러한 후성적 변화는 신호 전달, MAPK (mitogen-activated protein kinase) 및 G 단백질-결합 수용체 (GPCR) 신호 전달 경로와 관련이 있었다(표 1). 본 발명자들이 상기 그룹의 GO(gene ontology) 시스템에 기초하여 일련의 유전자들을 해석할 때, 상위 50개 프로브는 이화 과정, 신경전달물질 분비, 분자 기능의 조절, Wnt 신호전달 경로, 세포 이동 조절, mRNA 대사 과정, 작은(small) GTPase 매개 신호 전달, 유전자 발현 조절, 세포사멸 과정의 조절과 관련이 있었다.

[79] [표1]

경로 분석(n: 3226 참조서열 유전자)	소스	유전자의 수	P-값
Neuronal system	REACTOME	54	2.92E-13
Transmission across chemical synapses	REACTOME	36	1.61E-08
NRAGE signals death through JNK	REACTOME	13	3.56E-06
Signaling by Rho GTPases	REACTOME	22	1.25E-05
Extracellular matrix organization	REACTOME	36	3.71E-05
Axon guidance	REACTOME	42	5.75E-05
Pathways in cancer	KEGG	35	0.004137
Calcium signaling pathway	KEGG	22	0.005238
Signal transduction	REACTOME	157	0.006263
Cell adhesion molecules (CAMs)	KEGG	16	0.030841
MAPK signaling pathway	KEGG	25	0.03735
GPCR downstream signaling	REACTOME	75	0.040999
Hemostasis	REACTOME	44	0.044817

[80]

[81] 3. HCC 재발 예측 마커 선별

[82] 변수 중요도 플랏(variable importance plot, varImpPlot)을 이용하여 HCC 샘플을 MG1 및 MG2로 분류할 수 있는 상위 50개의 프로브를 발굴하였다(도 4). 이후 프로브 수에 따른 성능 효율을 비교하면서 AUC(area under the curve) 값을 고려하여 최대 효율을 지닌 프로브의 최소 개수를 선택하였다. 그 결과, 프로브 개수가 3개 이상인 경우 간암 재발 예측의 정확도(accuracy)가 1에 수렴하는 것을 알 수 있었다(도 5 및 표 2). 아래 표 2에 기재된 문자는 각각 정확도(Accu.), 민감도(Sen.) 및 특이도(Spec.)를 의미한다.

[83] [표2]

CGRC_ID	Accu.	Sen.	Spec.	AUC
CGRC_Prog.1	0.8621	0.9167	0.8235	0.9069
CGRC_Prog.2	0.8214	0.8333	0.8125	0.9063
CGRC_Prog.3	0.7857	0.8333	0.7555	0.9010
CGRC_Prog.4	0.7857	0.7500	0.8125	0.9010
CGRC_Prog.1+2	0.8929	0.8333	0.9375.	0.9479
CGRC_Prog.1+3	0.8966	0.9167	0.8824	0.9436
CGRC_Prog.1+4	0.8519	0.8182	0.8750	0.9432
CGRC_Prog.2+3	0.9286	1.0000	0.8750	0.9453
CGRC_Prog.2+4	0.8889	0.8182	0.9375	0.9347
CGRC_Prog.3+4	0.8929	1.0000	0.8125	0.9427
CGRC_Prog.1+2+3	0.9630	0.9091	1.0000	0.9773
CGRC_Prog.1+2+4	0.9286	0.9167	0.9375	0.9740
CGRC_Prog.1+3+4	0.9286	0.9167	0.9375	0.9740
CGRC_Prog.2+3+4	0.8889	0.9091	0.8750	0.9773
CGRC_Prog.1+2+3+4	0.963	0.9091	1.0000	0.9886

[84] 상기 표 2의 결과를 바탕으로 최종적으로 아래 표 3에 기재된 프로브를 선별하였으며, 프로브는 중요도 값이 높은 순서로 기재하였다.

[85] [표3]

Probe_ID	CGRC_ID	Gene_ID	chr	start	end
cg21325760	CGRC_Prog.1	MAGEL 2	chr15	23891342	23891463
cg10544510	CGRC_Prog.2		chr5	3761166	3761287
cg06702718	CGRC_Prog.3		chr12	132921267	132921388
cg15997204	CGRC_Prog.4	MYT1L	chr2	2189602	2189723

[86] 아래 표 4에 표 3에 기재된 프로브의 서열을 기재하였다.

[87] [표4]

서열 번호	Probe_ID	게놈 서열*
1	cg21325760	GGCCTGCCAGAGCGGTGGCTGGGTGGCCAGGACCTGTG GGGCAGGTCGGATGGGCGGCGG[CG]CCTGGCGGATCAG CGGCGGGGCCTGGCGGATCACAGGTGGAGCCTGGCGGA TCACAGGTG
2	cg10544510	CCTTCAGAGAGGAAGAATCTTAAGGCGCGTCCAAGACA AGCGGAGTCGAGGACTCCGTTT[CG]GGGAGGTTTACCT GAAAAGCTGTGTGGAAAAACCCGACTGCGACTTTCAAG CTTGTGCTT
3	cg06702718	CTGCAACATCACCGACAAAGAGCATGAAAATGTGAATG ACAGGACTCTAAAAAGAACGCA[CG]TTTACTCCGAGAG CGGAAACAAGAAGGCGGGTGCGGCCTCACTCCTTCCCG GCCAGGCAC
4	cg15997204	CTTCTGTGAGGGGGACGCAGCCGGGCCTCCAGACGCC CGCGTTCCTTACGTGAGGGGGA[CG]CAGCCGGGCCTTC CGGGCCTTCCCAAGTCCCGCGTTCCTTCTGTGAGGCGGA CGCAGCCG

[88] * 게놈 서열에서 [CG]는 해당 CG 마커가 표적하는 메틸화 부위를 의미함

[89]

[90] **3. HCC 재발 예측 마커 검증**

[91] 랜덤 포레스트 방법을 사용하여 표 3에 기재된 프로브에 대하여 메틸화 수준을 확인한 결과, 정상 간 조직(녹색)은 메틸화 수준이 거의 100%에 근접하고 일정한 수준이며, MG1(적색)은 정상 간 조직과 유사한 메틸화 수준을 보인 반면, MG2(회색)는 메틸화 수준이 정상 간 조직에 비해 약 50% 이하인 것을 확인할 수 있었다(도 6; X 축은 선별된 마커, Y 축은 메틸화 수준을 의미하는 b-값).

[92] 또한, 140개 HCC 샘플을 무작위로 훈련 데이터와 검증 데이터로 나누고, 표 3의 프로브로 다시 테스트한 결과, 기존의 MG1은 다시 MG1으로 분류되고, 기존의 MG2 또한 다시 MG2로 분류되는 것을 확인할 수 있었다(표 5).

[93] [표5]

	Type	MG1	MG2
훈련 데이터	MG1	59	6
	MG2	4	44
검증 데이터	MG1	16	1
	MG2	0	10

[94]

[95] 본 발명자가 자체적으로 생산한 간암 코호트 데이터를 상기 표 3의 프로브로 검증하였다. 그 결과, 메틸화 수준에 따라 간암 코호트 데이터가 두 개의 그룹(Pre_MG1 및 Pre_MG2)으로 분류되고, Pre_MG2에서 재발율이 높은 것을 확인할 수 있었다(도 7의 A). 또한, RFS를 확인한 결과, Pre_MG2에 비해 Pre_MG1의 RFS가 유의하게 긴 것을 확인할 수 있었다($p=0.0136$; 도 7의 B). 이 결과는 본 발명의 프로브로 간암의 재발 여부를 성공적으로 예측할 수 있음을 의미한다.

[96] 또한, 표 3의 프로브로 공개 데이터베이스인 TCGA(the cancer genome atlas) 간암 데이터에 대하여 동일하게 검증한 결과, 간암 환자들이 그룹(Pre_MG1 및 Pre_MG2)으로 분류되고(도 8의 A), 이들 그룹은 예후가 상이한 것을 알 수 있었다(도 8의 B).

청구범위

- [청구항 1] 서열번호 1 내지 4로 표시되는 DNA 메틸화 마커로 이루어진 군에서 선택되는 하나 이상의 DNA 메틸화 마커를 포함하는 간암의 재발 예측용 조성물.
- [청구항 2] 제1항에 있어서, 상기 간암의 재발 예측용 조성물은 2개 이상의 DNA 메틸화 마커를 포함하는 것인, 간암의 재발 예측용 조성물.
- [청구항 3] 제1항에 있어서, 상기 간암은 간세포암(hepatocellular carcinoma)인 것인, 간암의 재발 예측용 조성물.
- [청구항 4] 제1항에 있어서, 상기 간암의 재발 예측용 조성물은 간암 환자의 치료 후 정상 간에서 간암의 재발 여부를 예측하는 것인, 간암의 재발 예측용 조성물.
- [청구항 5] (a) 분석 대상 간 조직 시료에서 게놈 DNA를 분리하는 단계;
(b) 서열번호 1 내지 4로 표시되는 DNA 메틸화 마커로 이루어진 군에서 선택되는 하나 이상의 DNA 메틸화 마커의 메틸화 수준을 검출하는 단계; 및
(c) 상기 검출된 DNA 메틸화 마커의 메틸화 수준을 정상 간 조직에서 검출한 메틸화 수준과 비교하는 단계를 포함하는, 간암의 재발 예측 방법.
- [청구항 6] 제5항에 있어서, 상기 간암의 재발 예측 방법은 단계 (b)에서 최소 2개 이상의 DNA 메틸화 마커의 메틸화를 확인하는 것인, 간암의 재발 예측 방법.
- [청구항 7] 제5항에 있어서, 상기 간암의 재발 예측 방법은 단계 (c) 이후 분석 대상 간 조직에서 확인한 DNA 메틸화 마커의 메틸화 수준이 정상 간 조직의 메틸화 수준보다 낮은 경우 간암의 재발 가능성이 높은 것으로 판별하는 단계를 추가로 확인하는 것인, 간암의 재발 예측 방법.