

## 명세서

### 발명의 명칭: 생물학적 시료의 간 조직 유래 여부를 판별하는 방법

#### 기술분야

- [1] 본 발명은 기원을 모르는 생물학적 시료의 간 조직 유래 여부를 판별하는 방법 및 이를 수행하기 위한 간 조직 특이적 DNA 메틸화 마커를 포함하는 조성물에 관한 것이다.

#### 배경기술

- [2] 사람 몸에 있는 세포들은 동일한 유전자 정보를 가지고 있지만, 각 세포들의 기능과 형태는 매우 다양하다. 이는 각 세포마다 특정한 유전자가 발현되고, 이에 따라 세포 분화 과정이 상이하여 최종적으로 세포 표현형의 차이를 보이기 때문이다. 이러한 특정 유전자의 발현에는 DNA 메틸화, 히스톤 변형, 조직 특이적 전사 인자 등 여러 요인들이 관여한다. 특히 DNA 메틸화, 구체적으로 CpG 부위의 메틸화는 세포 특이적 유전자 발현의 필수요소로 세포 또는 조직마다 고유한 DNA 메틸화 특징을 갖는 것으로 알려져 있다. 따라서, DNA 메틸화 특징을 이용하면 세포 또는 조직이 유래한 기원을 용이하게 확인할 수 있다.
- [3] 세포 또는 조직의 기원을 확인하는 것은 질병의 조기 진단을 가능하게 하고, 진단의 정확도를 높일 수 있다. 예를 들어, 생체를 순환하는 혈액 속에는 순환 세포유리 DNA(cell free DNA, cfDNA)가 존재하고, 이 중에는 종양세포에서 흘러나온 순환 종양 DNA(circulating tumor DNA, ctDNA)가 있을 수 있다. 액체 생검으로 ctDNA의 존재를 검출한다고 하더라도 암을 정확히 진단하기 위해서는 해당 ctDNA가 기원한 조직을 추가로 확인하는 과정이 필요하며, 이러한 과정에서 암의 진단은 늦어지게 된다. 그러나 해당 ctDNA를 검출하면서 기원한 조직을 같이 확인할 수 있다면 암의 조기 진단이 가능해지며, 생물학적 시료의 기원을 밝히는 방법은 암뿐만 아니라 다른 질병의 조기 진단에 점점 필요해지고 있다.
- [4] 본 발명자들은 DNA 메틸화 데이터를 기반으로 기원을 모르는 조직 및 세포의 기원을 밝히는 방법을 연구한 결과, 간 조직 특이적으로 메틸화 수준이 높은 마커를 확인하여 본 발명을 완성하였다.

#### 발명의 상세한 설명

##### 기술적 과제

- [5] 본 발명의 목적은 기원이 불분명한 생물학적 시료가 간 조직에서 유래했는지 여부를 판별하는 방법 및 상기 방법을 수행하기 위한 조성물을 제공하는 것이다.

##### 과제 해결 수단

- [6] 상기 목적을 달성하기 위하여, 본 발명의 일 양상은 하기 단계를 포함하는 생물학적 시료의 간 조직 기원 여부를 판별하는 방법을 제공한다:

- [7] (a) 대상체(subject)의 분리된 생물학적 시료에서 DNA를 분리하는 단계; 및
- [8] (b) 상기 분리된 DNA에서 서열번호 1 및 서열번호 2로 표시되는 서열의 CpG 부위의 메틸화 수준을 측정하는 단계.
- [9] 본 발명의 일 구체예에 있어서, 상기 대상체는 인간일 수 있고, 상기 생물학적 시료는 상기 대상체에서 분리된 조직, 조직 단편, 세포, 세포 단편, 혈액, 혈장, 체액, 대변 및 소변 등을 포함하나 이에 제한되는 것은 아니다. 상기 조직, 조직 단편, 세포 및 세포 단편 등은 대상체에서 채취된 혈액, 혈장, 체액, 소변 등에서 분리될 것일 수 있다. 또한, 상기 DNA는 조직, 세포 등에서 분리한 DNA일 수 있고, 혈액, 혈장, 체액 등에 부유하는 세포유리 DNA(cell free DNA, cfDNA) 또는 종양세포에서 흘러나온 순환 종양 DNA(circulating tumor DNA, ctDNA)일 수 있다.
- [10] 본 명세서에 사용된 용어, "메틸화(methylation)"는 DNA를 구성하는 염기에 메틸기(-CH<sub>3</sub>)가 부착되는 것을 의미하며, 바람직하게는 특정 DNA의 특정 CpG 부위의 시토신에서 일어나는 메틸화를 의미한다. 용어, "메틸화 수준"은 특정 DNA 염기서열 내에 존재하는 CpG 부위의 메틸화 상태를 정량적으로 평가한 것을 의미하며, 메틸화 상태는 DNA 염기서열 내의 하나 이상의 CpG 부위에서 5-메틸-시토신의 존재 또는 부존재를 의미한다.
- [11] 본 명세서에 사용된 용어, "CpG 부위"는 시토신(cytosine, C)과 구아닌 (guanine, G)이 인산기(phosphate)로 연결된 서열을 말하며, 프로모터 영역, 단백질 코딩 영역(open reading frame, ORF) 및 터미네이터 영역 등을 포함하는 DNA 서열 내에 존재할 수 있다. CpG 부위의 메틸화는 유전체의 안정성 유지, 유전자의 발현 조절 등에 관여하는 것으로 알려져 있다.
- [12] 본 발명자들은 간 조직 특이적 메틸화 마커를 발굴하기 위해 노력한 결과, 정상 간 조직, 간암 조직 샘플에서는 메틸화 수준이 높고, 혈액을 포함한 기타 조직에서는 메틸화 수준이 낮은 cg12137206 (서열번호 1), cg03792768(서열번호 2) 마커를 발굴하였다. 상기 두 마커의 타겟 메틸화 부위는 각각 서열번호 1 및 서열번호 2로 표시되는 서열에서 61번째 뉴클레오티드에 위치하는 CpG 부위이다. 그러나 상기 타겟 메틸화 부위 이외에 다른 CpG 부위에서도 메틸화가 일어날 수 있으므로 본 발명에서는 상기 타겟 CpG 부위를 포함하여 서열번호 1 및 2로 표시되는 서열에 존재하는 전체 CpG 부위의 메틸화 수준을 측정할 수 있다.
- [13] 본 발명의 일 구체예에 있어서, 상기 (b)는 PCR, 메틸화 특이 PCR(methylation specific PCR), 실시간 메틸화 특이 PCR(real time methylation specific PCR), MethyLight PCR, MehtyLight digital PCR, EpiTYPER, 메틸화 DNA 특이적 결합 단백질을 이용한 PCR, 정량 PCR, DNA 칩, 분자 비콘(molecular beacon), MS-HRM (Methylation-sensitive high resolution melting), 비대칭 PCR (asymmetric PCR), 비대칭 PCR MS-HRMA (asymmetric PCR Methylation-sensitive high resolution melting analysis), 재조합효소-중합효소 증폭법 (Recombinase

- Polymerase Amplification), LAMP법(Loop-Mediated Isothermal Amplification), Eclipse probe, 차세대 염기서열분석 패널(NGS panel), 파이로시퀀싱 및 바이셀파이트 시퀀싱으로 이루어진 군에서 선택되는 방법으로 수행될 수 있다.
- [14] 예를 들어, 상기 메틸화 수준은 마이크로어레이에 의해 식별될 수 있고, 상기 마이크로어레이는 고상표면에 고정화된 프로브를 이용할 수 있다. 상기 프로브는 상기 CpG 부위를 포함하는 10 내지 100개의 연속 뉴클레오티드 서열에 상보적인 서열을 포함할 수 있다.
- [15] 본 발명의 일 구체예에 있어서, 상기 방법은 (c) 상기 (b) 단계 이후에, 상기 메틸화 수준을 정상 대조군의 메틸화 수준과 비교하는 단계;를 추가로 포함할 수 있다. 예를 들어 생물학적 시료의 메틸화 수준을 정상 대조군의 메틸화 수준과 비교하여 메틸화 수준이 높은 경우 상기 생물학적 시료가 간 조직으로부터 유래한 것으로 판별할 수 있고, 메틸화 수준이 낮거나 유사한 경우 상기 생물학적 시료는 간을 제외한 기타 조직에서 유래한 것으로 판별할 수 있다.
- [16]
- [17] 본 발명의 다른 양상은 하기 단계를 포함하는 생물학적 시료에서 간 조직 유래 DNA를 검출하는 방법을 제공한다:
- [18] (a) 대상체의 분리된 생물학적 시료에서 DNA를 분리하는 단계; 및
- [19] (b) 상기 분리된 DNA에서 서열번호 1 및 서열번호 2로 표시되는 서열의 CpG 부위의 메틸화 수준을 측정하는 단계.
- [20] 상기 방법은 생물학적 시료의 간 조직 기원 여부를 판별하는 방법과 동일한 서열의 메틸화 수준을 측정하는 기술을 사용하므로 두 방법 사이에 중복되는 내용은 명세서의 과도한 기재를 피하기 위하여 생략한다.
- [21] 본 발명에서, 상기 간 조직 유래 DNA 검출 방법은 기존의 간암 진단 방법과 병용하여 사용될 수 있다.
- [22]
- [23] 본 발명의 또 다른 양상은 서열번호 1 및 서열번호 2로 표시되는 서열의 메틸화 수준을 측정할 수 있는 제제를 포함하는 생물학적 시료의 간 조직 기원 여부 판별용 조성물을 제공한다.
- [24] 본 발명의 일 구체예에 있어서, 상기 CpG 부위는 서열번호 1 및 서열번호 2로 표시되는 서열에서 61번째 뉴클레오티드에 위치하는 CpG 부위를 포함하는 1개 내지 복수개가 될 수 있다.
- [25] 본 발명의 일 구체예에 있어서, 상기 메틸화 수준을 측정할 수 있는 제제는 서열번호 1 및 서열번호 2로 표시되는 서열의 CpG 부위에 결합하는 프라이머, 프로브 또는 안티센스 핵산일 수 있으며, 상기 프라이머, 프로브 또는 안티센스 핵산은 혼성화 어레이 요소(hybridizable array element)로 이용될 수 있고 기체(substrate) 상에 고정화될 수 있다.
- [26] 또한, 상기 서열번호 1 또는 2로 표시되는 서열은 게놈 DNA일 수 있고, 술폰산(bisulfite)이 처리되어 비메틸화된 시토신이 우라실로 전환된 서열일 수

있다.

- [27] 상기 기체는 적합한 견고성 또는 반-견고성 지지체로서, 예컨대, 막, 필터, 칩, 슬라이드, 웨이퍼, 파이버, 자기성 비드 또는 비자기성 비드, 겔, 튜빙, 플레이트, 고분자, 미소입자 및 모세관을 포함할 수 있다. 상기 혼성화 어레이 요소는 화학적 결합 방법, UV와 같은 공유 결합 방법 또는 링커(예: 에틸렌 글리콜 올리고머 및 디아민)를 통해 기체 상에 고정될 수 있다.
- [28] 본 발명의 일 구체예에 있어서, 생물학적 시료로부터 분리된 DNA(시료 DNA)는 혼성화 어레이에 적용되어 어레이 요소와 혼성화될 수 있으며, 혼성화 조건은 다양하게 변경될 수 있고, 혼성화 정도의 검출 및 분석은 당업계에 알려진 기술에 따라 다양하게 실시될 수 있다. 또한, 상기 시료 DNA 및/또는 프라이머, 프로브 또는 안티센스 핵산은 혼성화 여부를 확인케 하는 신호를 제공하기 위해 표지(labeling)될 수 있고, 올리고뉴클레오타이드에 연결될 수 있다.
- [29] 상기 표지는 형광단(예컨대, 플루오리신(fluorescein), 피코에리트린(phycoerythrin), 로다민, 리사민(lissamine), Cy3 및 Cy5(Pharmacia), 발색단, 화학발광단, 자기입자, 방사성동위원소(P32 및 S35), 효소(알칼린 포스파타아제 또는 호스래디쉬 퍼옥시다아제), 조인자, 효소에 대한 기질, 중금속(예를 들어, 금), 항체, 스트렙타비딘, 바이오틴, 디곡시게닌(digoxigenin)과 킬레이팅기와 같은 특정 결합 파트너를 갖는 헵텐을 포함할 수 있으나, 이에 한정되는 것은 아니다.
- [30] 본 발명에서, 상기 프라이머, 프로브 또는 안티센스 핵산과 시료 DNA의 혼성화는 반응 온도, 혼성화 및 세척 시간, 완충액 성분 및 이들의 pH 및 이온세기, 뉴클레오타이드의 길이, 뉴클레오타이드 서열, GC 서열의 양 등의 다양한 인자에 의존한다. 상기 혼성화를 위한 상세한 조건은 Joseph Sambrook, et al., Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.(2001); 및 M.L.M. Anderson, Nucleic Acid Hybridization, Springer-Verlag New York Inc. N.Y.(1999)을 참조할 수 있다.
- [31] 상기 혼성화 반응 이후에, 혼성화 반응을 통하여 나오는 혼성화 시그널을 검출할 수 있다. 예를 들어, 상기 프로브가 효소에 의해 표지된 경우, 상기 효소의 기질을 혼성화 반응 결과물과 반응시켜 혼성화 여부를 확인할 수 있다.
- [32] 상기 효소 및 기질은 퍼옥시다아제(예컨대, 호스래디쉬 퍼옥시다아제)와 클로로나프톨, 아미노에틸카바졸, 디아미노벤지딘, D-루시페린, 루시게닌(비스-N-메틸아크리디늄 니트레이트), 레소루핀 벤질 에테르, 루미놀, 암플렉스 레드 시약(10-아세틸-3,7-디하이드록시페녹사진), HYR (p-phenylenediamine-HCl 및 pyrocatechol), TMB (tetramethylbenzidine), ABTS (2,2'-Azine-di[3-ethylbenzthiazoline sulfonate]), o-페닐렌디아민(OPD) 및 나프톨/파이로닌; 알칼린 포스파타아제와 브로모클로로인돌일 포스페이트(BCIP), 니트로 블루 테트라졸리움(NBT), 나프톨-AS-B1-포스페이트(naphthol-AS-B1-phosphate) 및 ECF 기질; 글루코오스 옥시다아제와

t-NBT(nitroblue tetrazolium) 및 m-PMS(phenazine methosulfate)가 사용될 수 있다.

[33]

[34] 본 발명에서는 DNA 메틸화를 기반으로 간 조직 특이적인 마커를 발굴하였다. 구체적으로 혈액 내의 다양한 세포 타입, 여러 장기 샘플의 메틸화 분석을 진행하여 간 조직을 제외한 다른 장기에서는 메틸화가 낮고, 간 조직 특이적으로 메틸화가 높은 마커를 선별하고, 기계학습 기법으로 선별한 마커의 효용을 확인했다. 선별된 마커는 독립적 코호트인 정상 간 조직에서 간암까지 이르는 모든 샘플에서 메틸화 수준이 높았고, 혈액을 포함하여 다른 조직에서는 메틸화가 낮았으며, 2개 마커만으로도 99% 이상의 정확도로 간 특이성을 갖는 것을 검증하였다. 해당 마커의 유전자 발현 또한 간 특이적으로 높은 경향성을 확인했다. 간 특이적 메틸화 마커를 통해 조직의 기원을 확인할 수 있는 주요 마커로 활용될 수 있음을 규명했고, 향후 간 관련 질환의 모니터링 및 조기 검출의 정확도를 향상시킬 수 있는 지표로 활용하고자 한다.

### 발명의 효과

[35] 본 발명의 생물학적 시료의 간 조직 유래 여부를 판별하는 방법 및 이를 수행하기 위한 조성물은 간 조직 특이적 DNA 메틸화 마커를 이용하며, 상기 간 조직 특이적 마커는 간 조직을 제외한 다른 조직에서는 메틸화 수준이 낮고, 정상 간 조직 및 간암 조직에서는 메틸화 수준이 높으므로 생물학적 시료가 간 조직에서 유래했는지 여부를 우수한 정확도로 판별할 수 있다.

### 도면의 간단한 설명

[36] 도 1은 본 발명의 간 조직 특이적 마커 발굴 과정을 개략적으로 나타낸다.

[37] 도 2는 발굴한 간 조직 특이적 마커의 변수 중요도 플랏(variable importance plot, VarImp Plot)을 나타낸다.

[38] 도 3은 발굴한 간 조직 특이적 마커의 수에 따른 성능을 확인한 결과를 나타낸다.

[39] 도 4에서 A는 최종 선별한 cg12137206, cg03792768 마커의 메틸화 수준을 정상 간 조직(Liver N) 및 간암 조직(Liver T)에서 확인한 결과, B는 간암 샘플의 정상 간 조직(LIHC\_N)과 간암 조직(LIHC\_T), 간을 제외한 기타 정상 조직(Pan\_N) 및 암 조직(Pan\_T)에서 확인한 결과를 나타낸다.

[40] 도 5는 cg12137206 마커의 메틸화 수준을 다양한 정상 조직(A)과 암 조직(B)에서 확인한 결과를 나타낸다.

[41] 도 6은 cg03792768 마커의 메틸화 수준을 다양한 정상 조직(A)과 암 조직(B)에서 확인한 결과를 나타낸다.

[42] 도 7은 cg12137206 및 cg03792768 마커의 메틸화 수준을 주요 정상 조직과 암 조직에서 확인한 결과를 나타낸다: 방광 (bladder, BL), 유방(breast, BR), 자궁 경부 (cervix, CE), 대장(colon, CO), 식도 (esophagus, ES), 교모세포종 (glioblastoma, GB), 두부 및 경부 (head and neck, HN), 신장(kidney, KI), 간(liver,

LI), 폐(lung, LU), 이자 (pancreas, PA), 부신경절종 (paraganglioma, PC), 전립선 (prostate, PR), 직장 (rectum, RE), 육종 (sarcoma, SA), 피부 (skin, SK), 위 (stomach, ST), 흉선 (thymus, TH), 자궁 (uterine, UC) 및 혈액(blood, B).

#### 발명의 실시를 위한 형태

- [43] 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.
- [44]
- [45] **실시예 1: 간 조직 특이적 메틸화 마커 발굴**
- [46] 암유전체지도(The Cancer Genome Atlas; 이하, TCGA로 기재함)로부터 다양한 암종 및 정상 조직에 대한 DNA 메틸화 데이터를 다운로드하였다. TCGA에서 다운로드한 데이터로부터 간암 조직 샘플(n=379)에서 메틸화된 CpG 사이트를 선별하고, 이 중에서 다른 조직의 암종 샘플(n=7,260)에서는 비메틸화된 CpG 사이트를 추가로 선별하였다. 이때 기준은 50% 이상의 샘플에서 메틸화된 경우는 메틸화, 90% 이상의 샘플에서 비메틸화된 경우는 비메틸화로 분류하였다.
- [47] 이후 상기 추가 선별된 CpG 사이트로부터 간암을 제외한 암종 샘플(pan-tumor)의 90%에서 비메틸화된 CpG 사이트, 간을 제외한 정상 조직(pan-normal)의 90%에서 비메틸화된 CpG 사이트를 추가로 선별하였다.
- [48] 도 1에 본 발명의 간 조직 특이적 마커의 발굴 과정을 개략적으로 도시하였다.
- [49]
- [50] **실시예 2: 간 조직 특이적 메틸화 마커의 성능 검증**
- [51] 랜덤 포레스트 방법으로 상기 실시예 1에서 발굴한 간 조직 특이적 마커의 변수 중요도 플랏(variable importance plot, VarImp Plot)을 작성하고, 이를 도 2에 도시하였다. 도 2에서 X축의 MeanDecreaseAccuracy는 간 조직/기타 조직의 구분에 있어 각 마커가 정확도 개선에 기여하는 정도, Y축의 MeanDecreaseGini는 간 조직/기타 조직의 구분에 있어 각 마커가 불순도 개선에 기여하는 정도를 나타낸다. 즉, 값이 클수록 그 마커가 간 조직과 기타 조직을 명확히 구분할 수 있음을 의미한다.
- [52] 다음으로 실시예 1에서 사용한 암 조직의 메틸화 데이터를 무작위로 훈련 데이터와 검증 데이터로 구분하여 간 조직 특이적 마커를 테스트하였다. 그 결과, 아래 표 1에 기재된 바와 같이 발굴한 마커들이 간 조직과 기타 조직을 높은 효율과 정확도로 구분하는 것을 확인할 수 있었다. 간 조직 특이적 마커의 성능은 정확도(accuracy) 0.9988, 민감도(sensitivity) 0.9884, 특이도(specificity) 0.9994, 곡선 아래 면적(area under the curve, AUC) 0.9999로 나타났다.

[53]

[표1]

|        | 유형    | 기타 조직 | 간 조직 |
|--------|-------|-------|------|
| 훈련 데이터 | 기타 조직 | 6345  | 3    |
|        | 간 조직  | 7     | 337  |
| 검증 데이터 | 기타 조직 | 1586  | 1    |
|        | 간 조직  | 1     | 85   |

- [54] 또한, 마커의 수에 따른 성능을 확인하기 위하여 마커의 수를 다르게 하여 5번씩 테스트하고, 평균값을 확인하였다. 그 결과, 표 2 및 도 3에 나타난 바와 같이 2개 이상의 마커를 사용한 경우 성능이 0.999에 수렴하는 것을 알 수 있었다.

[55] [표2]

|                   | 1개       | 2개           | 3개           | 4개           | 5개           | 6개           |
|-------------------|----------|--------------|--------------|--------------|--------------|--------------|
| Accuracy          | 0.9916   | 0.9976       | 0.99736      | 0.99784      | 0.99796      | 0.9982       |
| Kappa             | 0.9142   | 0.97526      | 0.97264      | 0.97762      | 0.97886      | 0.98142      |
| Sensitivity       | 0.923036 | 0.96968<br>6 | 0.96503<br>4 | 0.96963      | 0.97195<br>4 | 0.97671<br>8 |
| Specificity       | 0.995338 | 0.99911<br>8 | 0.99911<br>8 | 0.99937      | 0.99937      | 0.99936<br>2 |
| Pos Pred Value    | 0.915048 | 0.98376<br>4 | 0.98376<br>4 | 0.98817<br>2 | 0.98831<br>8 | 0.98842<br>8 |
| Neg Pred Value    | 0.995844 | 0.99836<br>4 | 0.99811<br>2 | 0.99836<br>4 | 0.99849      | 0.99874<br>2 |
| Balanced Accuracy | 0.959188 | 0.98440<br>2 | 0.98207<br>8 | 0.9845       | 0.98566<br>4 | 0.98804<br>8 |

- [56] 또한, 변수 중요도 플랏에서 중요도가 높은 것으로 나온 2개 마커 (cg12137206, cg03792768)의 성능을 추가로 검증하였다. 그 결과, 아래 표 3에 기재된 바와 같이 상기 마커가 간 조직과 기타 조직을 명확하게 구분하는 것을 확인할 수 있었다. 정확도 0.9982, 민감도 0.9647, 특이도 1, 곡선 아래 면적(AUC) 0.9939로 나타났다.

[57]

[표3]

|        | 유형    | 기타 조직 | 간 조직 |
|--------|-------|-------|------|
| 훈련 데이터 | 기타 조직 | 6340  | 8    |
|        | 간 조직  | 11    | 333  |
| 검증 데이터 | 기타 조직 | 1587  | 3    |
|        | 간 조직  | 0     | 82   |

[58]

[59] 상기 결과들을 근거로 cg12137206, cg03792768 마커를 간 조직 특이적 마커로 최종 선별하였다. 아래 표 4 및 표 5에 cg12137206, cg03792768 마커의 정보 및 서열을 기재하였다.

[60]

[표4]

| Probe_ID   | CGRC_ID    | Gene_ID | chr   | start     | end       | CGI                       | CGI_location |
|------------|------------|---------|-------|-----------|-----------|---------------------------|--------------|
| cg12137206 | CGRC_LTO.1 | GPAM    | chr10 | 113943397 | 113943398 | chr10:113943283-113943657 | pCGI         |
| cg03792768 | CGRC_LTO.2 | BDH1    | chr3  | 197281934 | 197281935 | chr3:197281605-197283128  | pCGI         |

[61]

[표5]

| Probe_ID   | 서열                                                                                                                                      |
|------------|-----------------------------------------------------------------------------------------------------------------------------------------|
| cg12137206 | GGGGACACGACTGCCCCAGCAACTTGCAGGAGTCGCACCACCTCCATGCACTTGTCCCGG[CG]CTCCCGGCCCGAGTAGCCTCCGCAGCCCACACCTGCCCTGGCAGTTCGCACCCTAGCAG<br>(서열번호 1) |
| cg03792768 | TTCCCATTGGTTGAGACAGCACCGCCCAGCCAAAGCCCCCTTGTCCTCGCGCGGGTGC[CG]CCTGGACTCCCACCCTGGCCAGTCCCGGGCCCACCACCTCTGGCATCCCCAGCCTGTC<br>(서열번호 2)    |

[62]

- 상기 표에서 [CG]는 각 프로브의 타겟 메틸화 부위를 의미함

[63]

[64] 실시예 3: 최종 선별한 간 조직 특이적 마커의 검증

[65]

최종 선별한 cg12137206, cg03792768 마커의 간 조직, 다른 조직에서의 메틸화 수준을 확인하였다. 그 결과, 도 4의 A에 나타난 바와 같이 정상 간 조직(Liver N)과 간암 조직(Liver T) 모두에서 두 마커가 높은 수준으로 메틸화되는 것을 알

수 있었고, 도 4의 B에 나타난 바와 같이 두 마커 모두 기타 암 조직 및 기타 정상 조직에서는 메틸화 수준이 낮은 것을 알 수 있었다.

[66] 도 5 내지 도 7에 간 조직과 기타 조직에서 확인한 cg12137206, cg03792768 마커의 메틸화 수준을 도시하였고, 표 6 내지 8에는 상기 마커의 교차 검증 결과를 기재하였다.

[67] [표6]

| 조직                               |               | 샘플수 | 정확도_N  | 정확도_T  |
|----------------------------------|---------------|-----|--------|--------|
| 혈액(whole blood, B)               | KNIH          | 400 | 1      | NA     |
|                                  | GEO_WBset     | 107 | 1      | NA     |
| 간 종양 형성<br>(liver tumorigenesis) | GSE48325      | 79  | 1      | NA     |
|                                  | GSE49542      | 59  | 1      | NA     |
| 간세포암(HCC)                        | GSE43091      | 54  | 1      | 0.94   |
|                                  | GSE54503      | 132 | 1      | 0.9851 |
|                                  | GSE56588      | 234 | 1      | 0.9872 |
|                                  | GSE60753      | 66  | 0.9412 | 0.9375 |
|                                  | TCGA_LIHC     | 379 | 1      | 0.9921 |
|                                  | CGRC_HCC      | 307 | 1      | 0.9835 |
|                                  | TCGA_CHO<br>L | 45  | 1      | 0.5278 |
| 다른 조직                            | CGRC_CRC      | 709 | 1      | 1      |
|                                  | CGRC_lung     | 42  | NA     | 1      |

[68]

[표7]

| 조직                          | TCGA | 샘플수 | 정확도_N | 정확도_T |
|-----------------------------|------|-----|-------|-------|
| 방광 (bladder, BL)            | BLCA | 434 | 1     | 1     |
| 유방 (breast, BR)             | BRCA | 869 | 1     | 1     |
| 자궁 경부 (cervix, CE)          | CESC | 312 | 1     | 1     |
| 대장 (colon, CO)              | COAD | 335 | 1     | 1     |
| 직장 (rectum, RE)             | READ | 106 | 1     | 1     |
| 식도 (esophagus, ES)          | ESCA | 202 | 1     | 1     |
| 교모세포종 (glioblastoma, GB)    | GBM  | 154 | 1     | 1     |
| 두부 및 경부 (head and neck, HN) | HNSC | 580 | 1     | 1     |
| 신장 (kidney, KI)             | KIRC | 480 | 1     | 1     |
| 신장 (kidney, KI)             | KIRP | 321 | 1     | 1     |
| 폐 (lung, LU)                | LUAD | 492 | 1     | 1     |
| 폐 (lung, LU)                | LUSC | 412 | 1     | 1     |
| 이자 (pancreas, PA)           | PAAD | 195 | 1     | 1     |
| 부신경절종 (paraganglioma, PC)   | PCPG | 187 | 1     | 1     |
| 전립선 (prostate, PR)          | PRAD | 549 | 1     | 1     |
| 위 (stomach, ST)             | STAD | 397 | 1     | 1     |
| 육종 (sarcoma, SA)            | SARC | 269 | 1     | 1     |
| 피부 (skin, SK)               | SKCM | 475 | 1     | 1     |
| 흉선 (thymus, TH)             | THCA | 574 | 1     | 1     |
| 흉선 (thymus, TH)             | THYM | 126 | 1     | 1     |
| 자궁 (uterine, UC)            | UCEC | 466 | 1     | 1     |

[69]

[표8]

| 조직             | GEO #    | 샘플수 | 정확도    |
|----------------|----------|-----|--------|
| 방광 (BL)        | GSE52955 | 30  | 1      |
| 유방 (BR)        | GSE52865 | 57  | 1      |
|                | GSE39451 | 20  | 1      |
|                | GSE60185 | 285 | 1      |
| 자궁 경부 (CE)     | GSE46306 | 44  | 1      |
| 대장 (colon, CO) | GSE39958 | 45  | 1      |
|                | GSE42752 | 63  | 1      |
|                | GSE48684 | 147 | 1      |
| 식도 (ES)        | GSE52826 | 12  | 1      |
| 교모세포종 (GB)     | GSE36278 | 142 | 0.9787 |
|                | GSE58298 | 40  | 1      |
|                | GSE60274 | 77  | 0.987  |
| 두부 및 경부 (HN)   | GSE40005 | 24  | 1      |
|                | GSE38266 | 42  | 0.9286 |
| 신장 (KI)        | GSE50874 | 85  | 1      |
|                | GSE61441 | 92  | 1      |
| 폐 (LU)         | GSE39279 | 444 | 1      |
|                | GSE52401 | 244 | 1      |
|                | GSE56044 | 136 | 1      |
| 이자 (PA)        | GSE49149 | 196 | 1      |
| 부신경절종 (PC)     | GSE43293 | 24  | 1      |
| 전립선 (PR)       | GSE47915 | 8   | 1      |
|                | GSE55598 | 48  | 1      |
| 위 (ST)         | GSE34387 | 76  | 0.9867 |
| 흉선 (TH)        | GSE55111 | 11  | 1      |
| 자궁 (UC)        | GSE45187 | 9   | 1      |

[70]

[71]

지금까지의 결과를 통하여, cg12137206, cg03792768 마커는 간 조직에서는 메틸화 수준이 높고, 기타 조직에서는 메틸화 수준이 낮은 것을 알 수 있으며,

따라서 상기 두 마커는 간 조직 특이적 마커로 사용될 수 있다.