

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일

2020년 8월 27일 (27.08.2020)



(10) 국제공개번호

WO 2020/171573 A1

(51) 국제특허분류:

G16B 30/10 (2019.01)

G16B 40/00 (2019.01)

G16B 50/00 (2019.01)

G16H 50/50 (2018.01)

(21) 국제출원번호:

PCT/KR2020/002359

(22) 국제출원일:

2020년 2월 19일 (19.02.2020)

(25) 출원언어:

한국어

(26) 공개언어:

한국어

(30) 우선권정보:

10-2019-0019315 2019년 2월 19일 (19.02.2019) KR

(71) 출원인: 주식회사 녹색자지늄 (GREEN CROSS GENOME CORPORATION) [KR/KR]; 16924 경기도 용인시 기흥구 이현로30번길 107, Gyeonggi-do (KR). 국립암센터 (NATIONAL CANCER CENTER) [KR/KR]; 10408 경기도 고양시 일산동구 일산로 323, Gyeonggi-do (KR). 재단법인 아산사회복지재단 (THE ASAN FOUNDATION) [KR/KR]; 05505 서울시 송파구 올림픽로43길 88, Seoul (KR). 울산대학교 산학협력단 (UNIVERSITY OF ULSAN FOUNDATION FOR INDUSTRY CO-OPERATION) [KR/KR]; 44610 울산시 남구 대학로 93, Ulsan (KR).

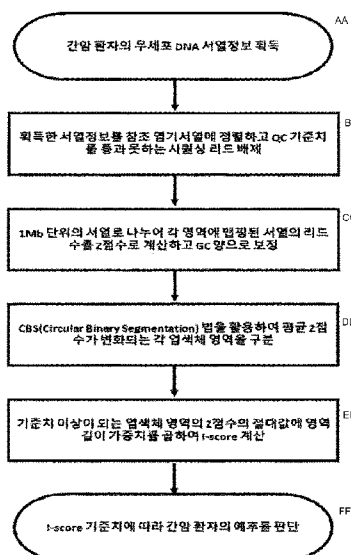
(72) 발명자: 류백렬 (RYOO, Baek-Yeol); 05572 서울시 송파구 올림픽로4길 15 #6-605, Seoul (KR). 박숙련 (PARK, Sook Ryun); 05516 서울시 송파구 한가람로 448 #101-1803, Seoul (KR). 조은해 (CHO, Eun Hae); 16924 경기도 용인시 기흥구 이현로30번길 107, Gyeonggi-do (KR). 이준남 (LEE, Junnam); 16924 경기도 용인시 기흥구 이현로30번길 107, Gyeonggi-do (KR). 전영주 (JEON, Young Joo); 16924 경기도 용인시 기흥구 이현로30번길 107, Gyeonggi-do (KR). 공선영 (KONG, Sun-Young); 10415 경기도 고양시 일산동구 강릉로 195 #815-1005, Gyeonggi-do (KR). 김민경 (KIM, Min Kyung); 14461 경기도 부천시 삼작로 349 #가-104, Gyeonggi-do (KR).

(74) 대리인: 이처영 등 (LEE, Cheo Young et al.); 06210 서울시 강남구 언주로 430 16층, Seoul (KR).

(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD,

(54) Title: BLOOD CELL-FREE DNA-BASED METHOD FOR PREDICTING PROGNOSIS OF LIVER CANCER TREATMENT

(54) 발명의 명칭: 혈중 무세포 DNA 기반 간암 치료 예후예측 방법



(57) Abstract: The present invention relates to a blood cell-free DNA-based method for predicting the prognosis of liver cancer treatment. A method for predicting the prognosis of liver cancer, according to the present invention, uses next generation sequencing (NGS) so as to increase the accuracy of prognosis prediction of a liver cancer patient and also increase the accuracy of prognosis prediction based on a very low concentration cell-free DNA of which detection has been difficult, thereby increasing the commercial utilization thereof. Therefore, the method of the present invention is useful for determining the prognosis of a liver cancer patient.

(57) 요약서: 본 발명은 본 발명은 혈중 무세포 DNA 기반의 간암 치료 예후예측 방법에 관한 것이다. 본 발명에 따른 간암 예후예측 방법은 차세대 염기서열 분석기법(Next Generation Sequencing, NGS)을 이용하여 간암 환자의 예후예측 정확도를 높일 뿐만 아니라 검출하기 어려웠던 매우 낮은 농도의 무세포 DNA에 기반의 예후예측 정확도를 높여서 상업적 활용도를 높일 수 있다. 따라서 본 발명의 방법은 간암 환자의 예후 판단에 유용하다.

AA ... Obtain cell-free DNA sequence information about liver cancer patient
BB ... Align acquired sequence information with reference base sequence and exclude sequencing reads that do not pass QC standard value
CC ... Divide same into sequences of 1Mb unit so as to calculate number of reads of sequences mapped to each region as z-score and correct same to GC amount
DD ... Classify each chromosome region where average z-score is changed, by using CBS (circular binary segmentation) method
EE ... Calculate l-score by multiplying, by region length weight, absolute value of chromosome region z-score that is at least reference value
FF ... Determine prognosis of liver cancer patient according to l-score standard value



SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ,
UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역
내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE,
LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM,
ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유
럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK,
MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML,
MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

명세서

발명의 명칭: 혈중 무세포 DNA 기반 간암 치료 예후예측 방법 기술분야

- [1] 본 발명은 혈중 무세포 DNA 기반의 간암 치료 예후예측 방법에 관한 것으로, 보다 구체적으로는 생체시료에서 무세포 DNA(cell free DNA, cfDNA)를 추출하여, 서열정보를 획득한 다음, 염색체 영역의 정규화 교정 및 회귀분석을 이용한 무세포 DNA 기반의 간암 치료 예후예측 방법에 관한 것이다.

[2]

배경기술

- [3] 원발성 간암은 전 세계적으로 세 번째로 흔한 암 사망 원인이며 그 발생빈도가 점차 증가하고 있는 암이다(Ferlay J et al., Int J Cancer Vol. 136:E359-86, 2015). 2015년 우리나라에서 발생한 214,701건의 암 중 7.3%에 해당하는 15,757건이 간암이었고 이는 전체 암 중 6번째로 많은 수이며 암 사망률 또한 2위의 높은 순위를 보였다. 연령별 간암의 발생 비율은 50대가 27.1%로 가장 많았고 60대, 70대가 각각 26.0%, 23.9%를 차지하였다. 원발성 간암 중 간세포암은 전체 간암의 85~90%를 차지하는 주된 조직학적 아형이다. 간세포암 발병의 주된 원인으로는 B형과 C형 간염바이러스의 감염을 들 수 있다. 또한 간염바이러스 외에도 장기간의 음주와 간경변증 역시 간암의 위험요인으로 알려져 있다. 알코올성 간경변증 환자의 8%, 간경변증 환자의 4%에서 5년 내에 간세포암이 발견되었다는 연구결과가 있으며 간경변증이 심할수록, 연령이 높을수록 간암 발생위험이 높다고 알려져 있다(Fattovich G et al., Gastroenterology, Vol. 127:S35-50, 2004).
- [4] 암은 세포의 유전자 돌연변이가 누적되면서 세포분열이 정상적으로 조절되지 않아 발생한다. 때문에 암세포의 염색체는 결실이나 중복, 전좌와 같은 염색체이상(chromosomal abnormality)이 빈번하게 나타나는 특징이 있다. 특히 염색체이상으로 인한 종양유전자(Oncogene)의 활성화 또는 종양억제유전자(Tumor suppressor gene)의 비활성화가 암의 발생에 큰 영향을 미치는 것으로 알려져 있다. 간암의 경우 1, 7, 8, 17, 20번 염색체의 중복과 4, 8, 13, 16, 17번 염색체의 결실이 간암의 발병과 높은 연관성을 보이는 것으로 알려져 있다(Zhou C, et al., Sci Rep. 2017 Vol. 7(1):10570). 특히 간암 환자의 somatic copy number alteration(SCNA)는 p53 signaling(TP53, CDKN2A), Wnt/ β -catenin pathway(CTNNB1, AXIN1), chromosomal remodeling(ARID1A, ARID1B, ARID2) 관련 유전자들 및 telomerase maintenance와 관련된 TERT유전자에서 빈번하게 나타났다(Ng CKY, et al., Front Med (Lausanne). 2018 Vol. 5:78). 해당 유전자들은 세포 주기와 세포성장 조절과 관련된 유전자들이며 간암의 발생에 있어 해당 유전자들 간에 연관성을 확인한 연구들이 발표된 바

있다(Ju-Seog Lee, Clin Mol Hepatol. 2015 Vol. 21(3): 220-229). 염색체 이상으로 인한 암의 발생 기작에 관한 연구들이 이루어지면서 암의 진단 및 예후판측의 지표로 활용하려는 노력이 계속되고 있다(Parker BC and Zhang W, Chin J Cancer. Vol. 11:594-603. 2013).

- [5] 더 나아가 최근에는 액체생검(Liquid biopsy) 기술을 이용하여 세포의 괴사(necrosis), 세포자살(apoptosis), 분비(secretion)에 의해 혈장 내에 존재하는 무세포 DNA(cfDNA; cell-free DNA)를 이용하여 염색체이상을 검출하려는 연구들이 진행되고 있다. 특히 종양세포에서 유래된 혈중 무세포 DNA는 정상세포에서 나타나지 않는 종양 특이적인 염색체이상 및 돌연변이를 포함하고 있으며, 반감기가 2시간 정도로 짧아서 종양의 현재상태를 반영한다는 장점이 있다. 또한 비침습적이고 반복적으로 채취가 가능하기 때문에 혈중 무세포 DNA는 암의 진단, 모니터링 및 예후 판측 등 암과 관련된 다양한 분야에서 종양 특이적인 생체 표지자로써 각광받고 있다. 최근 분자진단기술이 발전하면서 Digital Karyotyping, PARE 분석 등을 통해 암 환자의 혈중 무세포 DNA에서 종양 특이적인 염색체이상을 검출이 가능하다는 연구와 함께 이를 임상적으로 확인한 연구결과들이 발표된 바 있다(Leary RJ et al., Sci Transl Med. Vol. 4, Issue 162. 2012).
- [6] 난소암 환자 10명을 대상으로 한 Faye R. Harris의 연구에 따르면, 환자의 암 조직 DNA에서 확인한 미세결실을 수술 전후에 얻은 ctDNA에서 분석한 바 있다(Harris FR et al., Sci Rep. Vol. 6: 29831. 2016). 그 결과, 수술 전 8명의 환자, 수술 후 8명 중 3명의 재발환자 모두에게서 미세결실을 검출 하였다. 이를 통해 혈중 무세포 DNA의 미세결실 검출이 임상적으로 유의미하며, 종양 특이적인 염색체 이상이 혈중 무세포 DNA에 반영되는 것을 확인하였다.
- [7] 이외에도 Daniel G. Stover는 전이성 TNBC(Triple-Negative Breast Cancer)환자 164명을 대상으로 cfDNA를 통해 조직특이적 CNA를 분석한 바 있다(Stover DG. et al., J Clin Oncol. Vol. 36(6):543-553). 그 결과, NOTCH2, AKT2, AKT3와 같은 특정 유전자의 copy number gain이 전이성 TNBC에서 원발성 TNBC에 비해 높게 나타났으며, 18q11과 19p13 염색체의 중복을 가진 전이성 TNBC 환자의 생존율이 통계적으로 유의하게 낮은 것을 확인한 바 있다.
- [8] 이러한 기술배경하에, 본 발명자들은 혈중 무세포 DNA 기반의 간암 예후예측 방법을 개발하기 위해 예의 노력한 결과, 혈중 무세포 DNA 농도와 염색체 영역의 정규화 교정 및 회귀분석을 수행할 경우, 높은 민감도로 간암 환자의 예후를 예측할 수 있다는 것을 확인하고, 본 발명을 완성하였다.
- [9]
- [10] **발명의 요약**
- [11] 본 발명의 목적은 무세포 DNA(Cell Free DNA, cfDNA) 기반의 간암 예후예측 방법을 제공하는 것이다.
- [12] 본 발명의 다른 목적은 간암 예후를 예측하는 장치를 제공하는 것이다.

- [13] 본 발명의 또 다른 목적은 상기 방법으로 간암 예후를 예측하는 프로세서에 의해 실행되도록 구성되는 명령을 포함하는 컴퓨터 판독 가능한 매체를 제공하는 것이다.
- [14] 본 발명의 또 다른 목적은 상기 방법을 포함하는 간암의 예후 판단을 위한 정보의 제공 방법을 제공하는 것이다.
- [15] 상기 목적을 달성하기 위하여, 본 발명은 a) 생체시료에서 분리된 무세포 DNA의 서열정보를 획득하는 단계; b) 상기 서열정보(reads)를 참조집단의 표준 염색체 서열 데이터베이스(reference genome database)에 정렬(alignment)하는 단계; c) 상기 정렬된 서열정보(reads)에 대하여 퀄리티를 확인하여, 기준값(cut-off value) 이상인 서열정보만 선별하는 단계; d) 상기 표준 염색체를 일정 구간(bin)으로 나누고, 상기 선별된 서열정보(reads)에 대하여, 각 구간의 양을 확인하고 정규화하는 단계; e) 참조집단의 정규화된 각 구간(bin)에 매치되는 리드의 평균과 표준편차를 구한 다음, 상기 d) 단계에서 정규화한 값 사이의 Z 점수를 계산하는 단계; f) 상기 Z 점수(Z score)를 이용하여 염색체를 구분하여, I 점수를 계산하는 단계; 및 g) 상기 I 점수(I-score)의 기준값(cut-off value)을 초과할 경우, 간암 예후가 나쁜 것으로 판정하는 단계를 포함하는, 무세포 DNA(cell free DNA, cfDNA) 기반의 간암 예후예측 방법을 제공한다.
- [16] 본 발명은 또한, 생체시료에서 분리된 무세포 DNA의 서열정보를 해독하는 해독부; 해독된 서열을 참조집단의 표준 염색체 서열 데이터베이스에 정렬하는 정렬부; 정렬된 서열정보(reads)에 대하여 기준값(cut-off value) 이상인 샘플의 서열정보만 선별하는 품질관리부; 및 선별된 서열정보(reads)에 대하여, 참조집단 샘플과 비교하여 Z 점수(Z score)를 계산한 다음, 이를 바탕으로 I 점수(I-score)를 도출하여, I 점수가 기준값(cut-off value)을 초과할 경우, 간암 예후가 나쁜 것으로 판정하는 결정부를 포함하는 cfDNA 기반의 간암 예후예측 장치를 제공한다.
- [17] 본 발명은 또한, 컴퓨터 판독 가능한 매체로서, 간암 예후를 예측하는 프로세서에 의해 실행되도록 구성되는 명령을 포함하되, a) 생체시료에서 분리된 무세포 DNA의 서열정보를 획득하는 단계; b) 상기 서열정보(reads)를 참조집단의 표준 염색체 서열 데이터베이스(reference genome database)에 정렬(alignment)하는 단계; c) 상기 정렬된 서열정보(reads)에 대하여 퀄리티를 확인하여, 기준값(cut-off value) 이상인 서열정보만 선별하는 단계; d) 상기 표준 염색체를 일정 구간(bin)으로 나누고, 상기 선별된 서열정보(reads)에 대하여, 각 구간의 양을 확인하고 정규화하는 단계; e) 참조집단의 정규화된 각 구간(bin)에 매치되는 리드의 평균과 표준편차를 구한 다음, 상기 d) 단계에서 정규화한 값 사이의 Z 점수를 계산하는 단계; f) 상기 Z 점수(Z score)를 이용하여 염색체를 구분하여, I 점수를 계산하는 단계; 및 g) 상기 I 점수(I-score)의 기준값(cut-off value)을 초과할 경우, 간암 예후가 나쁜 것으로 판정하는 단계를 포함하는 프로세서에 의해 실행되도록 구성되는 명령을 포함하는 컴퓨터 판독 가능한

매체를 제공한다.

- [18] 본 발명은 또한 상기 방법을 포함하는 간암의 예후 판단을 위한 정보의 제공 방법을 제공한다.

[19]

도면의 간단한 설명

- [20] 도 1은 본 발명의 cfDNA 기반 간암 예후예측을 위한 전체 흐름도이다.
- [21] 도 2는 read data의 QC(퀄리티 관리, quality control) 과정 중, LOESS 알고리즘에 의한 GC 교정 전과 후의 시퀀싱 리드 수의 보정결과를 도식화 한 것이다.
- [22] 도 3은 정상인과 간암 환자의 혈중 무세포 DNA 농도 차이를 확인한 결과이다.
- [23] 도 4는 혈중 무세포 DNA 농도에 따른 간암의 진행 및 생존 여부에 대한 평가 결과이다.
- [24] 도 5는 본 발명의 방법에 따른 간암의 진행 및 생존 여부에 대한 예후 예측 결과이다.
- [25] 도 6은 본 발명의 I 점수를 세분화한 그룹별 간암 환자의 생존 여부에 대한 예후 예측 결과이다.
- [26] 도 7은 본 발명의 I 점수를 세분화한 그룹별 간암의 진행에 대한 예후 예측 결과이다.
- [27] 도 8은 본 발명의 I 점수와 혈중 무세포 DNA 농도의 상관관계를 확인한 결과이다.

[28]

발명의 상세한 설명 및 바람직한 구현예

- [29] 다른 식으로 정의되지 않는 한, 본 명세서에서 사용된 모든 기술적 및 과학적 용어들은 본 발명이 속하는 기술 분야에서 숙련된 전문가에 의해서 통상적으로 이해되는 것과 동일한 의미를 갖는다. 일반적으로 본 명세서에서 사용된 명명법 및 이하에 기술하는 실험 방법은 본 기술 분야에서 잘 알려져 있고 통상적으로 사용되는 것이다.
- [30] 본 발명에서는, 간암 환자 샘플에서 획득한 서열 분석 데이터를 정규화하고, 기준값을 바탕으로 정리한 뒤, 일정 구간(bin)으로 나누어 각 구간(bin) 별 리드 양을 정규화 한 다음, 참조집단 샘플과의 Z 점수(Z score)를 계산하고, 도출된 Z 점수(Z score)를 기반으로 염색체를 다시 나눈 뒤(segmentation), 이를 바탕으로 I 점수(I-score)를 계산하여, I 점수(I-score)가 1637을 초과하면, 나쁜 예후를 나타내고 1637 이하이면 좋은 예후를 나타내는 것으로 판단할 수 있다는 것을 확인하였다. 구체적으로, I 점수의 범위에 따라 간암에 의한 사망 또는 진행에 대한 위험군을 분류하여 확인할 수 있다. 보다 구체적으로, I 점수가 1638 내지 3012인 경우, 중등도 위험군으로 분류되고, I 점수가 3013 내지 7448인 경우 및 7449 내지 13672인 경우 고도 위험군으로 분류되고, I 점수가 13673 내지 28520인 경우 초고도 위험군으로 분류될 수 있다.

- [32] 즉, 본 발명의 일 실시예에서는, 정상인 14명과 간암 환자 151명의 혈액에서 추출한 DNA를 시퀀싱 한 뒤, LOESS 알고리즘을 이용하여 품질을 관리하고, 염색체를 일정 구간(bin)으로 구분하여 각 구간 별 매칭되는 리드 양을 GC 비율로 정규화한 다음, 정상인 샘플에서 각 구간(bin)에 매칭되는 리드의 평균과 표준편차를 구한다음, 상기 정규화한 값과의 Z 점수(Z score)를 계산하고 이를 기반으로 Z 점수(Z score)가 급변하는 염색체 영역을 다시 나눈 뒤(segmentation), 이를 이용하여 I 점수(I-score)를 계산하여, I 점수(I-score)가 1637을 초과할 경우, 간암 환자의 예후가 나쁘다고 판정하는 방법을 개발하였다(도 1)
- [33] 본 발명에서 용어 "리드(reads)"는, 당업계에 알려진 다양한 방법을 이용하여 서열정보를 분석한 하나의 핵산 단편을 의미한다. 따라서, 본 명세서에서 용어 "서열정보" 및 "리드"는 시퀀싱 과정을 통해 서열정보를 수득한 결과물이라는 점에서 동일한 의미를 가진다.
- [34] 본 발명에서 용어 "예후예측"이란, "예후"와 동일한 의미로 사용되는데, 질환의 경과 및 결과를 미리 예측하는 행위를 의미한다. 보다 구체적으로, 예후예측이란 질환의 치료 후 경과를 환자의 생리적 또는 환경적 상태에 따라 달라질 수 있으며, 이러한 환자의 상태를 종합적으로 고려하여 치료 후 병의 경과를 예측하는 모든 행위를 의미하는 것으로 해석될 수 있다.
- [35] 본 발명의 목적상 상기 예후예측은 간암의 치료 후, 질환의 경과를 미리 예상하여 암의 진행, 암의 재발 및/또는 암의 전이의 위험도를 예측하는 행위로 해석될 수 있다. 예를 들어, 용어 "좋은 예후"는 간암 치료 후 환자의 암의 진행, 암의 재발 및/또는 암의 전이의 위험도가 1보다 낮은 값을 나타내어, 간암 환자가 생존할 가능성이 높다는 것을 의미하고, 다른 의미로 "긍정적 예후"로도 표현된다. 용어 "나쁜 예후"는 간암 치료 후 환자의 암의 진행, 암의 재발 및/또는 암의 전이의 위험도가 1보다 높은 값을 나타내어, 간암 환자가 사망할 가능성이 높다는 것을 의미하고, 다른 의미로 "부정적 예후"로도 표현된다.
- [36] 본 발명에서 용어 "위험도"란, 간암의 치료 후, 환자가 암의 진행, 재발 및/또는 암의 전이 등이 나타날 확률에 대한 오즈비, 위험비 등을 의미한다.
- [37]
- [38] 따라서, 본 발명은 일 관점에서,
- [39] a) 생체시료에서 분리된 무세포 DNA의 서열정보를 획득하는 단계;
- [40] b) 상기 서열정보(reads)를 참조집단의 표준 염색체 서열 데이터베이스(reference genome database)에 정렬(alignment)하는 단계;
- [41] c) 상기 정렬된 서열정보(reads)에 대하여 퀄리티를 확인하여, 기준값(cut-off value) 이상인 서열정보만 선별하는 단계;
- [42] d) 상기 표준 염색체를 일정 구간(bin)으로 나누고, 상기 선별된 서열정보(reads)에 대하여, 각 구간의 양을 확인하고 정규화하는 단계;
- [43] e) 참조집단의 정규화된 각 구간(bin)에 매칭되는 리드의 평균과 표준편차를 구한 다음, 상기 d) 단계에서 정규화한 값 사이의 Z 점수를 계산하는 단계;

- [44] f) 상기 Z 점수(Z score)를 이용하여 염색체를 구분하여, I 점수를 계산하는 단계; 및
- [45] g) 상기 I 점수(I-score)의 기준값(cut-off value)을 초과할 경우, 간암 예후가 나쁜 것으로 판정하는 단계를 포함하는 무세포 DNA(cell free DNA, cfDNA) 기반의 간암 예후예측 방법에 관한 것이다.
- [46]
- [47] 본 발명에 있어서,
- [48] 상기 a) 단계는
- [49] (a-i) 채취된 무세포 DNA에서 염색 방법(salting-out method), 컬럼크로마토그래피 방법(column chromatography method), 또는 비드 방법(beads method)을 사용하여 단백질을, 지방, 및 기타 잔여물을 제거하고 정제된 핵산을 획득하는 단계;
- [50] (a-ii) 상기 정제된 핵산에 대하여, 싱글-엔드 시퀀싱(single-end sequencing) 또는 페어-엔드 시퀀싱(pair-end sequencing) 라이브러리(library)를 제작하는 단계;
- [51] (a-iii) 상기 제작된 라이브러리를 차세대 유전자서열검사기(next-generation sequencer)에 반응시키는 단계; 및
- [52] (a-iv) 상기 차세대 유전자서열검사기에서 핵산의 서열정보(reads)를 획득하는 단계를 포함하는 방법으로 수행되는 것을 특징으로 할 수 있다.
- [53] 상기 (a-i) 및 상기 (a-ii) 단계 사이에, 상기 (a-i) 단계에서 정제된 핵산을, 효소적 절단, 분쇄 또는 하이드로셰어방법(hydroshear method)으로 무작위 단편화(random fragmentation)하여 싱글-엔드 시퀀싱 또는 페어-엔드 시퀀싱 라이브러리를 제작하는 단계를 추가로 포함하는 방법으로 수행할 수 있다.
- [54] 본 발명에 있어서, 상기 a) 단계의 서열정보를 획득하는 단계는 분리된 무세포 DNA를 1백만 내지 1억 리드 깊이로 전장 유전체 시퀀싱을 통해 획득하는 것을 특징으로 할 수 있다.
- [55]
- [56] 본 발명에서 용어 "참조집단"은 표준 염기서열 데이터베이스와 같이 비교할 수 있는 기준(reference) 집단으로, 현재 특정 질환 또는 병증이 없는 사람의 집단을 의미한다. 본 발명에 있어서, 상기 참조집단의 표준 염색체 서열 데이터베이스에서 표준 염기서열은 NCBI 등의 공공보건기관에 등록되어 있는 참조 염색체일 수 있다.
- [57] 본 발명에 있어서, 상기 차세대 유전자서열 검사기(next-generation sequencer)는 이에 제한되지는 않으나, 일루미나 컴퍼니의 하이섹(Hiseq) 시스템, 일루미나 컴퍼니의 마이섹(Miseq) 시스템, 일루미나 컴퍼니의 게놈 분석기(GA) 시스템, 로슈 컴퍼니(Roche Company)의 454 FLX, 어플라이드 바이오시스템즈 컴퍼니의 SOLiD 시스템, 라이프 테크놀로지 컴퍼니의 이온토허트 시스템일 수 있다.
- [58] 본 발명에 있어서, 상기 정렬단계는 이에 제한되지는 않으나, BWA 알고리즘 및 Hg19 서열을 이용하여 수행되는 것일 수 있다.

- [59] 본 발명에 있어서, 상기 BWA 알고리즘은 BWA-ALN, BWA-SW 또는 Bowtie2 등이 포함될 수 있으나 이에 한정되는 것은 아니다.
- [60] 본 발명에 있어서, 상기 c) 단계에서 상기 정렬된 서열정보에 대하여 퀄리티를 확인하는 것은, 정렬 일치도 점수(Mapping Quality Score) 지표를 이용하여 실제 시퀀싱 리드가 참조 염색체 서열과 얼마나 일치하는지를 확인하는 것을 의미한다.
- [61] 본 발명에 있어서, 상기 c) 단계는
- [62] (c-i) 각 정렬된 핵산서열의 영역을 특정하는 단계; 및
- [63] (c-ii) 상기 영역 내에서 정렬 일치도 점수(mapping quality score)와 GC 비율의 기준값을 만족하는 서열을 선별하는 단계; 를 포함하여 수행되는 것을 특징으로 할 수 있다.
- [64]
- [65] 본 발명에 있어서, 상기 (c-i) 단계의 핵산서열의 영역을 특정하는 단계에서, 핵산서열의 영역은 이에 제한되는 않으나, 20kb~1MB일 수 있다.
- [66] 본 발명에 있어서, 상기 (c-ii) 단계에서, 상기 기준값은 상기 정렬 일치도 점수(mapping quality score)는 원하는 기준에 따라 달라질 수 있으나, 구체적으로는 15 내지 70, 보다 구체적으로는 30 내지 65, 가장 구체적으로는 60일 수 있다. 상기 (c-ii) 단계에서, 상기 GC 비율이 원하는 기준에 따라 비율이 달라질 수 있으나, 구체적으로는 20 내지 70%, 보다 구체적으로는 30 내지 60% 인 것을 특징으로 할 수 있다.
- [67] 본 발명에 있어서, 상기 c) 단계는 염색체의 중심체 또는 말단체의 데이터를 제외하고 수행되는 것을 특징으로 할 수 있다.
- [68] 본 발명에서 용어 “중심체”는 각 염색체 장완(q arm)의 시작점으로부터 1Mb 내외인 것을 특징으로 할 수 있으나, 이에 한정되는 것은 아니다.
- [69] 본 발명에서 용어 “말단체”는 각 염색체 단완(p arm)의 시작점으로부터 1 Mb 내외 이내 또는 장완(q arm)의 종료점으로부터 1 Mb 이내인 것을 특징으로 할 수 있으나, 이에 한정되는 것은 아니다.
- [70]
- [71] 본 발명에 있어서, 상기 d) 단계는
- [72] (d-i) 표준 염색체를 일정구간(bin)으로 나누는 단계;
- [73] (d-ii) 상기 구간별 정렬된 리드 개수 및 리드들의 GC양을 산출하는 단계;
- [74] (d-iii) 상기 리드 개수 및 GC양을 바탕으로 회귀분석을 실시하여 회귀계수를 산출하는 단계; 및
- [75] (d-iv) 상기 회귀계수를 이용하여 리드 개수를 정규화하는 단계를 포함하여 수행되는 것을 특징으로 할 수 있다.
- [76] 본 발명에 있어서, (d-i)에서의 일정구간(bin)은, 구체적으로는 100 kb 내지 2000 kb 일 수 있다.
- [77] 본 발명에 있어서, 상기 (d-i) 단계의 핵산서열의 영역을 특정하는 단계에서,

일정구간(bin)은 이에 제한되는 않으나, 100 kb 내지 2MB, 구체적으로 500kb 내지 1500 kb, 보다 구체적으로는 600kb 내지 1600 kb, 보다 더 구체적으로 800kb 내지 1200 kb, 가장 구체적으로 900 kb 내지 1100 kb 일 수 있다.

[78] 본 발명에 있어서, 상기 (iii) 단계의 회귀분석은 회귀계수를 산출할 수 있는 회귀분석 방법이면 모두 이용가능하나, 구체적으로는 LOESS 분석인 것을 특징으로 할 수 있으나, 이에 한정되는 것은 아니다.

[79]

[80] 본 발명에 있어서, 상기 e) 단계의 Z 점수(Z score)를 계산하는 단계는 특정 영역(bin)별 시퀀싱 리드 값을 표준화하는 것을 특징으로 할 수 있으며, 구체적으로는 하기의 수식 1로 계산하는 것을 특징으로 할 수 있다.

[81]

$$\text{수식 1: Z 점수} = \frac{\text{생체시료의 서열정보 샘플의 리드값} - \text{참조 집단의 평균 서열정보 리드값}}{\text{참조 집단의 평균 서열정보 리드값에 대한 표준편차}}$$

[82]

[83] 본 발명에 있어서, 상기 (f) 단계는

[84] (f-i) 각 구간별 Z 점수(Z score)를 기반으로 CBS 방법(Circular Binary segmentation method)으로 염색체 영역을 구분하는 단계;

[85] (f-ii) 상기 구분된 구역의 Z 점수(Z score)의 평균 절대값이 기준값 이상인 지역의 염색체 길이(size)를 구하는 단계; 및

[86] (f-iii) 하기 수식 2로 I 점수(I-score)를 계산하는 단계:

[87] 수식 2

$$: I = \sum_{j \text{ from all segmented above absolute mean Z score value 2}} |MeanZ_j| * Size_j$$

[88]

[89] 본 발명에 있어서, 상기 Z 점수(Z score)의 평균 절대값의 기준값은 1-2이고, 보다 구체적으로는 2인 것을 특징으로 할 수 있다.

[90] 본 발명에서 CBS 알고리즘은 상기 단계에서 계산된 Z 점수의 변화가 발생하는 지점을 검출하는 방법을 의미한다.

[91] 즉, 염색체의 Z 점수의 변화가 시작되는 임의의 지점을 i, 끝나는 임의의 지점을 j, 전체 영역 길이를 N, r을 각 핵산 서열(특정 bin 구간)의 bin 값, s를 bin 값들의 표준 편차라고 가정하면 $1 \leq i < j \leq N$ 의 조건 하에서, 아래의 식을 만족한다.

[92]

$$\text{수식 6: } S_i = r_1 + r_2 + \dots + r_i$$

$$\text{수식 7: } S_j = r_1 + r_2 + \dots + r_j$$

$$\text{수식 8: } S_{ij} = S_j - S_i = \sum_{n=i+1}^j r_n$$

$$\text{수식 9: } T_{ij} = \left(\frac{S_{ij}}{j-i} - \frac{S_{j-i} - S_{ij}}{N-j-i} \right) / \left(s \sqrt{\frac{1}{j-i} + \frac{1}{N-j-i}} \right)$$

$$\text{수식 10: } (i_c, j_c) = \arg \max |T_{ij}|$$

- [93] 여기서 (i_c, j_c) 는 Z 점수 변화가 실제 일어난 위치를 의미하며, max는 최대값, arg는 편각을 의미한다.
- [94] 본 발명에 있어서 상기 I 점수의 기준값은 1637인 것을 특징으로 할 수 있다.
- [95]
- [96] 본 발명에 있어서, 상기 방법은 분리된 무세포 DNA의 농도를 측정하여, 무세포 DNA의 농도가 기준값을 초과하면 나쁜 예후인 것으로 판단하는 단계를 추가로 포함하는 것을 특징으로 할 수 있다.
- [97] 본 발명에 있어서, 상기 분리된 무세포 DNA 농도의 기준값은 0.71ng/ μ l인 것을 특징으로 할 수 있다.
- [98]
- [99] 본 발명에 있어서, 상기 방법은 상기 I 점수가 1638 내지 3012인 경우, 중등도 위험군으로 분류하고, 3013 내지 13672인 경우, 고도 위험군으로 분류하고, 13673 내지 28520인 경우 초고도 위험군으로 분류하는 단계를 추가로 포함하는 것을 특징으로 할 수 있다.
- [100]
- [101] 본 발명은 다른 관점에서, 생체시료에서 분리된 무세포 DNA의 서열정보를 해독하는 해독부; 해독된 서열을 참조집단의 표준 염색체 서열 데이터베이스에 정렬하는 정렬부; 정렬된 서열정보(reads)에 대하여 기준값(cut-off value) 이상인 샘플의 서열정보만 선별하는 품질관리부; 및 선별된 서열정보(reads)에 대하여, 참조집단 샘플과 비교하여 Z 점수(Z score)를 계산한 다음, 이를 바탕으로 I 점수(I-score)를 도출하여, I 점수가 기준값을 초과할 경우, 나쁜 예후로 판정하는 결정부를 포함하는 cfDNA 기반의 간암 예후예측 장치에 관한 것이다.
- [102] 본 발명에 있어서, 상기 I 점수의 기준값은 1637인 것을 특징으로 할 수 있다.
- [103]
- [104] 본 발명에 있어서, 상기 장치는 분리된 무세포 DNA의 농도를 측정하여, 무세포 DNA의 농도가 기준값을 초과하면 나쁜 예후인 것으로 판단하는 농도 기반 예후 결정부를 추가로 포함하는 것을 특징으로 할 수 있다.
- [105] 본 발명에 있어서, 상기 분리된 무세포 DNA 농도의 기준값은 0.71ng/ μ l인 것을 특징으로 할 수 있다.
- [106]
- [107] 본 발명은 또 다른 관점에서, 간암 예후를 예측하는 프로세서에 의해 실행되도록 구성되는 명령을 포함하되, a) 생체시료에서 분리된 무세포 DNA의 서열정보를 획득하는 단계; b) 상기 서열정보(reads)를 참조집단의 표준 염색체 서열 데이터베이스(reference genome database)에 정렬(alignment)하는 단계; c) 상기 정렬된 서열정보(reads)에 대하여 퀄리티를 확인하여, 기준값(cut-off value) 이상인 서열정보만 선별하는 단계; d) 상기 표준 염색체를 일정 구간(bin)으로 나누고, 상기 선별된 서열정보(reads)에 대하여, 각 구간의 양을 확인하고 정규화하는 단계; e) 참조집단의 정규화된 각 구간(bin)에 매치되는 리드의

평균과 표준편차를 구한 다음, 상기 d) 단계에서 정규화한 값 사이의 Z 점수를 계산하는 단계; f) 상기 Z 점수(Z score)를 이용하여 염색체를 구분하여, I 점수를 계산하는 단계; 및 g) 상기 I 점수(I-score)의 기준값을 초과할 경우, 나쁜 예후인 것으로 판정하는 단계를 포함하는 프로세서에 의해 실행되도록 구성되는 명령을 포함하는 컴퓨터 판독 가능한 매체에 관한 것이다.

[108] 본 발명에 있어서, 상기 I 점수의 기준값은 1637인 것을 특징으로 할 수 있다.

[109]

[110] 본 발명에 있어서, 상기 프로세서는 분리된 무세포 DNA의 농도를 측정하여, 분리된 무세포 DNA의 농도가 기준값을 초과하면 나쁜 예후인 것으로 판단하는 단계를 추가로 포함하는 것을 특징으로 할 수 있다.

[111] 본 발명에 있어서, 상기 분리된 무세포 DNA 농도의 기준값은 0.71ng/ μ l인 것을 특징으로 할 수 있다.

[112]

[113] 본 발명은 또 다른 관점에서 상기 방법을 포함하는 간암의 예후 판단을 위한 정보의 제공 방법에 관한 것이다.

[114]

[115] 본 발명에 있어서, 상기 간암은 간에서 발생하는 모든 종류의 암이면 제한없으며, 보다 구체적으로는 간세포 암종 (섬유증관성 변형이 있거나 없는 간세포 암종), 담관암종 (간내 쓸개관 암종) 및 혼합 간세포 담관암종을 포함하나, 이에 한정되는 것은 아니다.

[116]

[117] 본 발명의 용어 “예후(prognosis)”는 암의 진행, 암의 재발 및/또는 암의 전이 가능성의 예측을 의미한다. 본 발명의 상기 예측 방법은 임의의 특정환자에 대한 가장 적절한 치료 양식을 선택하는 것으로 임상적으로 치료 결정을 내리기 위해 사용될 수 있다. 본 발명의 상기 예측 방법은 환자의 암의 진행, 암의 재발 및/또는 암의 전이가 발생할 가능성이 높은지를 판단하는 것에 대한 진단 및/또는 진단을 보조하는 가치있는 도구이다.

[118]

[119] 실시예

[120] 이하, 실시예를 통하여 본 발명을 더욱 상세히 설명하고자 한다. 이들 실시예는 오로지 본 발명을 예시하기 위한 것으로서, 본 발명의 범위가 이들 실시예에 의해 제한되는 것으로 해석되지는 않는 것은 당업계에서 통상의 지식을 가진 자에게 있어서 자명할 것이다.

[121]

[122] 실시예 1. 간암 환자 및 정상인에서 I-score 계산

[123] 151명 간암 환자의 혈장검체와 14명의 정상인 혈장검체에서 무세포 DNA를 추출하고 전장 염색체에 대한 라이브러리를 제조하였다. 무세포 DNA의 추출은 다음과 같은 순서로 진행하였다. 1) EDTA Tube채혈 후 4시간 이내에 1600g에서

10분, 3000g에서 10분 순차적으로 원심분리하여 상층액(혈장)을 분리; 2) 분리한 혈장 중 1.5ml을 사용하여 QIAamp Circulating Nucleic Acid Kit로 무세포 DNA 추출; 3) 최종 추출된 무세포DNA를 Qubit 2.0 Fluorometer에 반응시키고 농도(ng/ul) 측정; 라이브러리의 제조는 일루미나의 Truseq nano kit를 기반으로 진행하였으며 총 5ng의 무세포 DNA를 반응에 사용하였다. 연구에 참여한 151명 간암환자의 정보는 표 1과 같다.

[124]

[125] [표1]

151명 간암 환자의 임상 정보

Characteristics	N = 151
Age, years	57 (52–63)
Sex	
Male	137 (90.7%)
Female	14 (9.3%)
ECOG performance status	
0	52 (34.4%)
1	97 (64.2%)
2	2 (1.3%)
Etiology	
Hepatitis B	134 (88.7%)
Hepatitis C	4 (2.6%)
Alcohol	7 (4.6%)
Others	6 (4.0%)
Child-Pugh class	
A	140 (92.7%)
B	11 (7.3%)
BCLC stage	
B	5 (3.3%)
C	146 (96.7%)
Macrovascular invasion	
Yes	63 (41.7%)
No	88 (58.3%)
No. of extrahepatic spread organ sites	
0	16 (10.6%)
1	78 (51.7%)
2	41 (27.2%)
≥3	16 (10.6%)
Sites of extrahepatic spread	
Lymph node	64 (42.4%)
Lung	77 (51.0%)
Bone	32 (21.2%)
Peritoneum	23 (15.2%)
Adrenal gland	13 (8.6%)
Others?	
AFP (ng/mL)	
<20	41 (27.1%)
20–200	32 (21.2%)
>200	77 (51.0%)
Not available	1 (0.7%)
Platelet count ($\times 10^3/\text{mm}^3$)	122.0 (85.0–165.0)
Prothrombin time (INR)	1.08 (1.02–1.16)
Albumin (g/dL)	3.7 (3.4–4.0)
Total bilirubin (mg/dL)	0.7 (0.5–1.0)

[126]	AST (IU/L)	39 (28–58)
	ALT (IU/L)	26 (18–39)
	Previous therapy	
	No	10 (6.6%)
	Yes	141 (93.4%)
	Surgical resection	69 (45.7%)
	RFA	37 (24.5%)
	TACE	118 (78.1%)
	Radiotherapy	79 (52.3%)
	Liver transplantation	12 (7.9%)
Data are the median (Interquartile range) or number (%) unless otherwise indicated. ECOG, Eastern Cooperative Oncology Group; BCLC, Barcelona Clinic Liver Cancer; AFP, alpha fetoprotein; INR, international normalized ratio; AST, aspartate aminotransferase; ALT, alanine aminotransferase; RFA, radiofrequency ablation; TACE, transcatheter arterial chemoembolization.		

[127]

[128] 완성된 라이브러리는 NextSeq 장비에서 염기서열 분석을 수행하였으며, 샘플당 평균 10 million read(1 million read - 100 million read)의 서열정보 데이터를 생산하였다.

[129] 차세대염기서열분석(NGS) 장비에서 Bcl 파일(염기서열정보 포함)을 fastq 형식으로 변환한 다음, fastq 파일을 BWA-mem 알고리즘을 사용하여 참조염색체 Hg19서열 기준으로 라이브러리 서열을 정렬하였다. 정렬 일치도 점수(Mapping quality score)가 60을 만족하는 것을 확인하였다.

[130] GC양에 따라 각 염색체 좌위 구간(bin)의 시퀀싱 리드 수의 분포가 편향되는 것을 확인했고(도 2), 회귀분석을 사용하여 염색체별 GC 비율에 따라 정렬된 라이브러리 서열의 숫자를 교정하였다.

[131] 이후 하기 수식 1로 Z 점수(Z score)를 계산하였다:

[132] 수식 1: Z 점수 =
$$\frac{\text{생체시료의 서열정보 샘플의 리드값} - \text{참조 집단의 평균 서열정보 리드값}}{\text{참조 집단의 평균 서열정보 리드값에 대한 표준편차}}$$

[133]

[134] I-score를 계산하기 위해, 계산된 bin별 Z score를 데이터로 사용해, CBS 알고리즘으로, 염색체를 분할(Segmentation)하는 과정이 선행되었다.

[135] 평균 Z score 값이 절대값 2 이상인 분할 지역의 평균 Z score와 염색체 길이를 곱한 뒤, 이 값들의 합으로 각 샘플의 I-score를 구하였고 I-score 값이 1637을 넘어가는 샘플은 혈액 내 무세포 DNA의 양이 증가하고 Sorafenib 치료에 대한 예후가 좋지 않은 샘플로 판단하였다. I-score는 하기의 수식 2로 계산하였으며, 백분위에 따른 I-score의 값은 표 2와 같다.

[136] 수식 2

$$I = \sum_{j \text{ from all segmented above absolute mean Z score value 2}} |MeanZ_j| * Size_j$$

[137]

[138] [표2]

간암 151명의 백분위에 따른 I-score 분포

백분위(간암코호트)	I-score
0~12.50%	256~611
12.51%~25.00%	612~762
25.01%~37.50%	763~1003
37.51%~50.00%	1004~1637
50.01%~62.50%	1638~3012
62.51%~75.00%	3013~7448
75.01%~87.50%	7449~13672
87.51%~100.00%	13673~28520

[139]

[140] 실시예 2. 혈중 무세포 DNA 농도(ng/μl)가 간암의 진행 및 생존에 미치는 영향 확인

[141] 총 151명 간암 환자의 혈장에서 추출한 무세포 DNA 농도 값 분포는 최소 0.13 ng/ul에서 최대 15.00 ng/ul 였으며, 중간값(Median)은 0.71 ng/ul였다. 정상인 14명의 무세포 DNA 농도 값 분포는 최소 0.28 ng/ul에서 최대 0.54 ng/ul 였으며, 중간값(Median)은 0.34 ng/ul였다. 두 군의 차이에 대한 검정은 Mann-Whitney Test로 진행하였으며, 그 결과 유의미하게 차이가 있는 것($p<0.0001$)을 확인하였다(도 3).

[142] 혈중 무세포 DNA 농도는 간암 환자 151명의 예후(생존 여부와 무증약기간; Overall Survival and Time To Progression)에도 영향을 주었다. 151명의 혈중 무세포 DNA 농도의 중간값인 0.71 ng/ul을 기준으로 하여 해당 값을 초과하였을 때의 생존 여부(Overall Survival)와 무증약기간(Time To Progression)에 대한 위험도를 평가하였다. 151 명의 간암 환자는 모두 400mg의 sorafenib을 1일당 2회씩 복용하였으며, 항암치료 반응에 대한 평가는 RECIST 가이드라인 Version 1.1에 의거하여 6-8주 단위로 진행하였다.

[143] 분석 결과, 무세포 DNA 농도가 0.71 ng/ul을 초과하였을 때 무증약기간에 대한 위험비(Hazard Ratio, HR)가 1.71(95% CI, 1.20-2.44; log-rank $p=0.002$)이었으며, 생존 여부에 대한 위험비(Hazard Ratio, HR)가 3.50(95% CI, 2.36-5.20; log-rank $p<0.0001$)이었다. 이를 근거로 무세포 DNA의 혈중 농도 증가가 암의 진행 및 사망의 위험을 증가시키는 것을 확인하였다(도 4).

[144]

[145] 실시예 3. I-score가 간암의 진행 및 생존에 미치는 영향 확인

[146] 총 151명 간암 환자의 I-score는 최소 256에서 최대 28520로 분포하였으며,

중간값은 1637이었다. 정상인 14명은 Somatic CNA가 발견되지 않아서 I-score의 값이 모두 0에 해당하였다. I-score의 중간값인 1637을 기준으로 하여 을 평가하였다. 151 명의 간암 환자는 모두 400mg의 sorafenib을 1일당 2회씩 복용하였으며, 항암치료 반응에 대한 평가는 RECIST 가이드라인 Version 1.1에 의거하여 6-8주 단위로 진행하였다.

[147] 분석 결과, I-score가 1637을 초과하였을 때 질병의 무증약기간에 대한 위험비(Hazard Ratio, HR)가 2.09(95% CI, 1.46-3.00; log-rank $p < 0.0001$)이었으며, 생존 여부에 대한 위험비(Hazard Ratio, HR)가 3.35(95% CI, 2.24-5.01; log-rank $p < 0.0001$)이었다 (도 5).

[148] I-score를 8분위수로 세분화 하였을 때 생존 여부에 대한 위험비는 5분위수(1638~3012)가 2.97(95% CI, 1.28-6.90; $p = 0.01$), 6분위수(3013~7448)가 4.99(95% CI, 2.19-11.41; $p = 0.0001$), 7분위수(7449~13672)가 4.52(95% CI, 2.01-10.18; $p = 0.0003$), 8분위수(13673~28520)가 7.72(95% CI, 3.31-18.02; $p < 0.0001$)로 점차적으로 증가하는 경향을 나타냈다 (도 6).

[149] 이와 같은 경향은 무증약기간에 대한 위험비에서도 유사하게 나타나는데, 5분위수가 2.43(95% CI, 1.21-4.86; $p = 0.01$), 6분위수가 2.73(95% CI, 1.36-5.48; $p = 0.0047$), 7분위수가 2.26(95% CI, 1.09-4.70; $p = 0.0294$), 8분위수가 3.08(95% CI, 1.50-6.35; $p = 0.0022$)로 I-score가 증가할수록 암의 진행에 대한 위험이 증가하였다 (도 7).

[150] 이를 근거로 I-score의 증가가 암의 진행 및 사망의 위험을 증가시키는 것을 확인하였다.

[151]

[152] 실시예 4. 무세포 DNA 농도와 I-score간 상관관계 확인

[153] 혈중 무세포 DNA 농도와 I-score 모두 간암의 진행 및 생존에 영향을 주는 것으로 분석되어 두 변수간 상관관계를 파악하기 위하여 Spearman Correlation Analysis를 실시하였다.

[154] 그 결과, $R^2 = 0.24$, $p < 0.0001$ 로 분석되어 정비례의 상관관계가 있는 것을 확인할 수 있었다(도 8).

[155]

[156] 이상으로 본 발명 내용의 특정한 부분을 상세히 기술하였는 바, 당업계의 통상의 지식을 가진 자에게 있어서 이러한 구체적 기술은 단지 바람직한 실시 양태일 뿐이며, 이에 의해 본 발명의 범위가 제한되는 것이 아닌 점은 명백할 것이다. 따라서, 본 발명의 실질적인 범위는 첨부된 청구항들과 그것들의 등가물에 의하여 정의된다고 할 것이다.

[157]

산업상 이용가능성

[158] 본 발명에 따른 간암 예후예측 방법은 차세대 염기서열 분석기법(Next

Generation Sequencing, NGS)을 이용하여 간암 환자의 예후예측 정확도를 높일 뿐만 아니라 검출하기 어려웠던 매우 낮은 농도의 무세포 DNA에 기반의 예후예측 정확도를 높여서 상업적 활용도를 높일 수 있다. 따라서 본 발명의 방법은 간암 환자의 예후 판단에 유용하다.

청구범위

- [청구항 1] 다음의 단계를 포함하는 무세포 DNA(cell free DNA, cfDNA) 기반의 간암 예후예측 방법:
- a) 생체시료에서 분리된 무세포 DNA의 서열정보를 획득하는 단계;
 - b) 상기 서열정보(reads)를 참조집단의 표준 염색체 서열 데이터베이스(reference genome database)에 정렬(alignment)하는 단계;
 - c) 상기 정렬된 서열정보(reads)에 대하여 퀄리티를 확인하여, 기준값(cut-off value) 이상인 서열정보만 선별하는 단계;
 - d) 상기 표준 염색체를 일정 구간(bin)으로 나누고, 상기 선별된 서열정보(reads)에 대하여, 각 구간의 양을 확인하고 정규화하는 단계;
 - e) 참조집단의 정규화된 각 구간(bin)에 매치되는 리드의 평균과 표준편차를 구한 다음, 상기 d) 단계에서 정규화한 값 사이의 Z 점수를 계산하는 단계;
 - f) 상기 Z 점수(Z score)를 이용하여 염색체를 구분하여, I 점수를 계산하는 단계; 및
 - g) 상기 I 점수(I-score)의 기준값(cut-off value)을 초과할 경우, 간암 예후가 나쁜 것으로 판정하는 단계.
- [청구항 2] 제1항에 있어서, 상기 a) 단계는 다음의 단계를 포함하는 방법으로 수행되는 것을 특징으로 하는 cfDNA 기반의 간암 예후예측 방법:
- (a-i) 채취된 무세포 DNA에서 염색 방법(salting-out method), 컬럼크로마토그래피 방법(column chromatography method), 또는 비드 방법(beads method)을 사용하여 단백질, 지방, 및 기타 잔여물을 제거하고 정제된 핵산을 수득하는 단계;
 - (a-ii) 상기 정제된 핵산에 대하여, 싱글-엔드 시퀀싱(single-end sequencing) 또는 페어-엔드 시퀀싱(pair-end sequencing) 라이브러리(library)를 제작하는 단계;
 - (a-iii) 상기 제작된 라이브러리를 차세대 유전자서열검사기(next-generation sequencer)에 반응시키는 단계; 및
 - (a-iv) 상기 차세대 유전자서열검사기에서 핵산의 서열정보(reads)를 획득하는 단계.
- [청구항 3] 제2항에 있어서,
- 상기 (a-i) 및 상기 (a-ii) 단계 사이에, 상기 (a-i) 단계에서 정제된 핵산을, 효소적 절단, 분쇄 또는 하이드로셰어방법(hydroshear method)으로 무작위 단편화(random fragmentation)하여 싱글-엔드 시퀀싱 또는 페어-엔드 시퀀싱 라이브러리를 제작하는 단계를 추가로 포함하는 방법으로 수행되는 것을 특징으로 하는 cfDNA 기반의 간암 예후예측 방법.
- [청구항 4] 제1항에 있어서, 상기 a) 단계의 서열정보를 획득하는 단계는 분리된

- 무세포 DNA를 1백만 내지 1억 리드 깊이로 전장 유전체 시퀀싱을 통해 획득하는 것을 특징으로 하는 cfDNA 기반의 간암 예후예측 방법.
- [청구항 5] 제1항에 있어서, 상기 c) 단계는 다음의 단계를 포함하는 방법으로 수행되는 것을 특징으로 하는 cfDNA 기반의 간암 예후예측 방법:
(c-i) 각 정렬된 핵산서열의 영역을 특정하는 단계; 및
(c-ii) 상기 영역 내에서 정렬 일치도 점수(mapping quality score)와 GC 비율의 기준값을 만족하는 서열을 선별하는 단계.
- [청구항 6] 제5항에 있어서, 상기 기준값은, 상기 정렬 일치도 점수(mapping quality score)가 15 내지 70이고, GC 비율은 30 내지 60%인 것을 특징으로 하는 cfDNA 기반의 간암 예후예측 방법.
- [청구항 7] 제5항에 있어서, c) 단계는, 염색체의 중심체 또는 말단체의 데이터를 제외하고 수행되는 것을 특징으로 하는 cfDNA 기반의 간암 예후예측 방법.
- [청구항 8] 제1항에 있어서, 상기 (d) 단계는 다음의 단계를 포함하는 방법으로 수행되는 것을 특징으로 하는 cfDNA 기반의 간암 예후예측 방법:
(d-i) 표준 염색체를 일정구간(bin)으로 나누는 단계;
(d-ii) 상기 구간별 정렬된 리드 개수 및 리드들의 GC양을 산출하는 단계;
(d-iii) 상기 리드 개수 및 GC양을 바탕으로 회귀분석을 실시하여 회귀계수를 산출하는 단계; 및
(d-iv) 상기 회귀계수를 이용하여 리드 개수를 정규화하는 단계.
- [청구항 9] 제8항에 있어서, (d-i)에서의 일정구간(bin)은 100 kb 내지 2 Mb인 것을 특징으로 하는 cfDNA 기반의 간암 예후예측 방법.
- [청구항 10] 제1항에 있어서, 상기 e) 단계는, 하기의 수식 1로 계산하는 것을 특징으로 하는 cfDNA 기반의 간암 예후예측 방법:

$$\text{수식 1: } Z \text{ 점수} = \frac{\text{생체시료의 서열정보 샘플의 리드값} - \text{참조 집단의 평균 서열정보 리드값}}{\text{참조 집단의 평균 서열정보 리드값에 대한 표준편차}}$$

- [청구항 11] 제1항에 있어서, 상기 (f) 단계는 다음의 단계를 포함하는 방법으로 수행되는 것을 특징으로 하는 cfDNA 기반의 간암 예후예측 방법:
(f-i) 각 구간별 Z 점수를 기반으로 CBS(Circular Binary Segmentation) 방법으로 염색체 영역을 구분하는 단계;
(f-ii) 상기 구분된 구역의 Z 점수의 평균 절대값이 기준값 이상인 지역의 염색체길이(size)를 구하는 단계; 및
(f-iii) 하기 수식 2로 I 점수를 계산하는 단계

수식 2

$$I = \sum_j^{\text{from all segmented above absolute mean Z score value 2}} |MeanZ_j| * Size_j$$

- [청구항 12] 제11항에 있어서, 상기 Z 점수의 평균 절대값의 기준값은 1-2인 것을

- 특징으로 하는 순환 종양 DNA 검출 방법.
- [청구항 13] 제1항에 있어서, 상기 I 점수의 기준값은 1637인 것을 특징으로 하는 순환 종양 DNA 검출 방법.
- [청구항 14] 제1항에 있어서,
분리된 무세포 DNA의 농도를 측정하여, 무세포 DNA의 농도가 기준값을 초과하면 나쁜 예후인 것으로 판단하는 단계를 추가로 포함하는 것을 특징으로 하는 cfDNA 기반의 간암 예후예측 방법.
- [청구항 15] 제14항에 있어서,
분리된 무세포 DNA 농도의 기준값은 0.71ng/ μ l인 것을 특징으로 하는 cfDNA 기반의 간암 예후예측 방법.
- [청구항 16] 제1항에 있어서,
상기 I 점수가 1638 내지 3012인 경우, 중등도 위험군으로 분류하고, 3013 내지 13672인 경우, 고도 위험군으로 분류하고, 13673 내지 28520인 경우 초고도 위험군으로 분류하는 단계를 추가로 포함하는 것을 특징으로 하는 cfDNA 기반의 간암 예후예측 방법.
- [청구항 17] 제1항 내지 제16항 중 어느 한 항의 방법으로 간암 예후를 예측하는 단계를 포함하는 간암의 예후 판단을 위한 정보의 제공 방법.
- [청구항 18] 생체시료에서 분리된 무세포 DNA의 서열정보를 해독하는 해독부;
해독된 서열을 참조집단의 표준 염색체 서열 데이터베이스에 정렬하는 정렬부;
정렬된 서열정보(reads)에 대하여 기준값(cut-off value) 이상인 샘플의 서열정보만 선별하는 품질관리부; 및
선별된 서열정보(reads)에 대하여, 참조집단 샘플과 비교하여 Z 점수(Z score)를 계산한 다음, 이를 바탕으로 I 점수(I-score)를 도출하여, I 점수가 기준값(cut-off value)을 초과할 경우, 간암 예후가 나쁜 것으로 판정하는 결정부를 포함하는 cfDNA 기반의 간암 예후예측 장치.
- [청구항 19] 제18항에 있어서, 분리된 무세포 DNA의 농도를 측정하여, 농도가 기준값을 초과하면 나쁜 예후인 것으로 판단하는 농도 기반 예후 결정부를 추가로 포함하는 것을 특징으로 하는 cfDNA 기반의 간암 예후예측 장치.
- [청구항 20] 컴퓨터 판독 가능한 매체로서, 간암 예후를 예측하는 프로세서에 의해 실행되도록 구성되는 명령을 포함하되,
a) 생체시료에서 분리된 무세포 DNA의 서열정보를 획득하는 단계;
b) 상기 서열정보(reads)를 참조집단의 표준 염색체 서열 데이터베이스(reference genome database)에 정렬(alignment)하는 단계;
c) 상기 정렬된 서열정보(reads)에 대하여 퀄리티를 확인하여, 기준값(cut-off value) 이상인 서열정보만 선별하는 단계;
d) 상기 표준 염색체를 일정 구간(bin)으로 나누고, 상기 선별된

서열정보(reads)에 대하여, 각 구간의 양을 확인하고 정규화하는 단계;

e) 참조집단의 정규화된 각 구간(bin)에 매치되는 리드의 평균과 표준편차를 구한 다음, 상기 d) 단계에서 정규화한 값 사이의 Z 점수를 계산하는 단계;

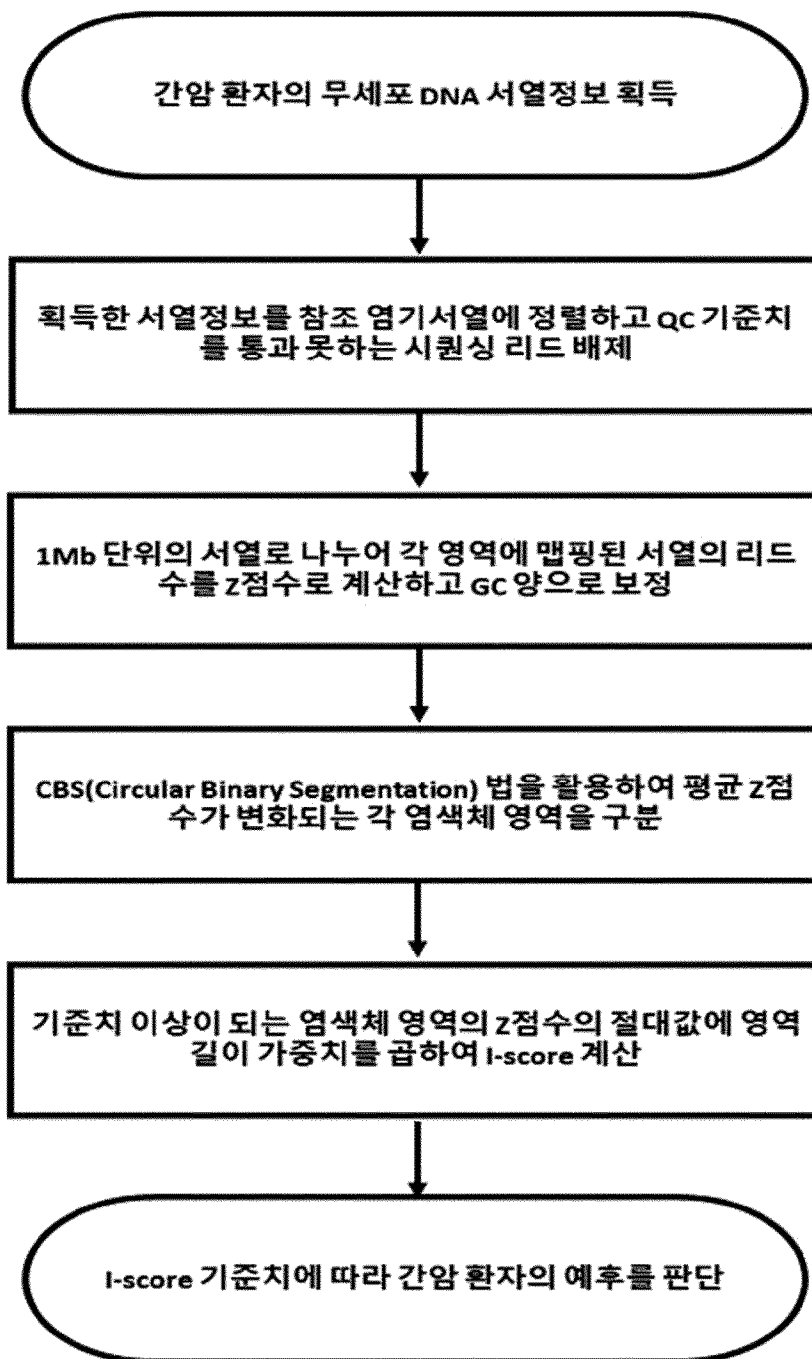
f) 상기 Z 점수(Z score)를 이용하여 염색체를 구분하여, I 점수를 계산하는 단계; 및

g) 상기 I 점수(I-score)의 기준값(cut-off value)을 초과할 경우, 간암 예후가 나쁜 것으로 판정하는 단계.

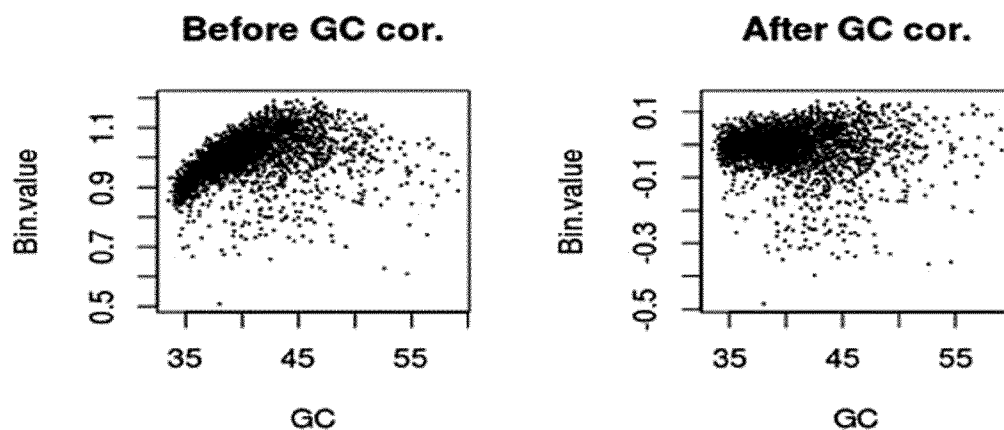
를 포함하는 프로세서에 의해 실행되도록 구성되는 명령을 포함하는 컴퓨터 판독 가능한 매체.

[청구항 21] 제20항에 있어서, 분리된 무세포 DNA의 농도를 측정하여, 무세포 DNA의 농도가 기준값을 초과하면 나쁜 예후인 것으로 판단하는 단계를 추가로 포함하는 것을 특징으로 하는 컴퓨터 판독 가능한 매체.

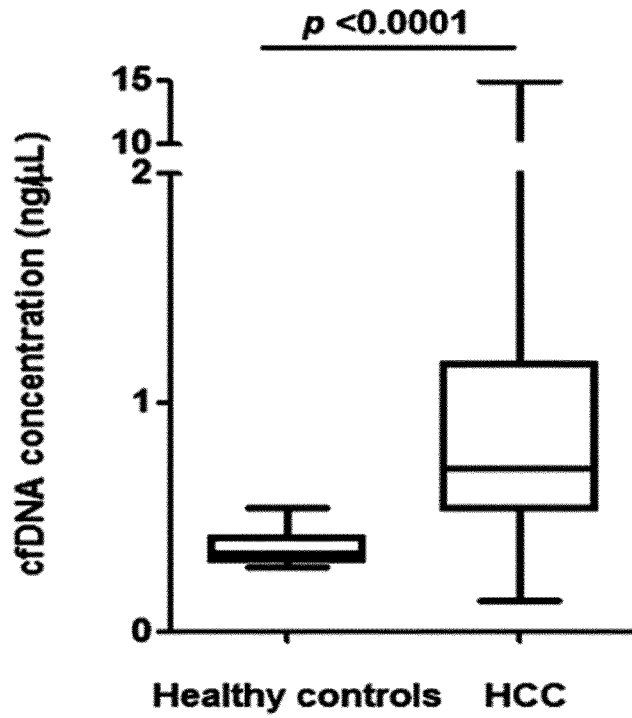
[도1]



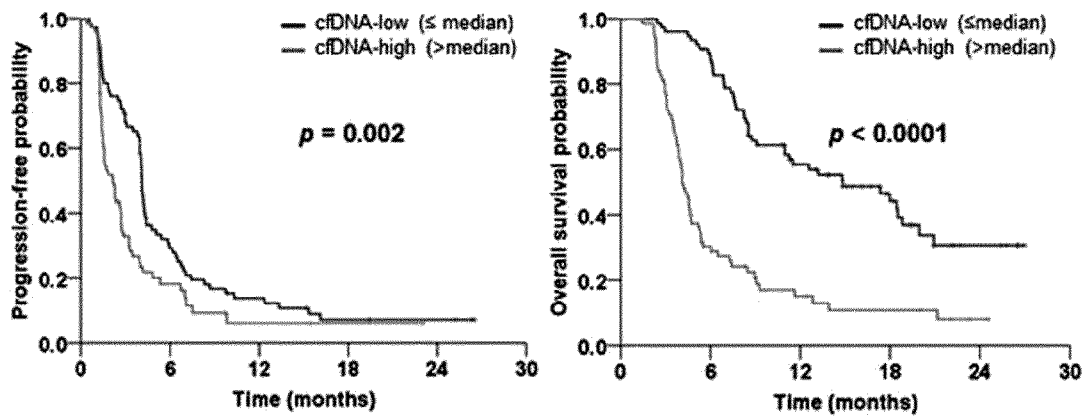
[도2]



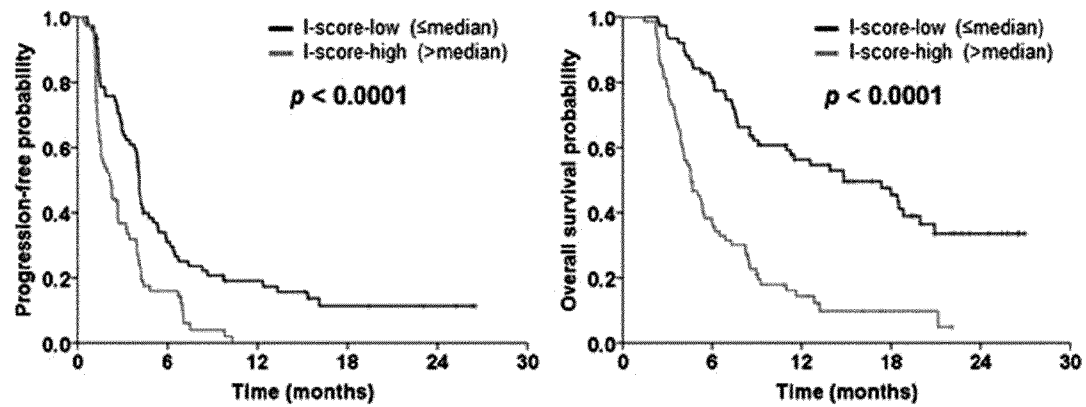
[도3]



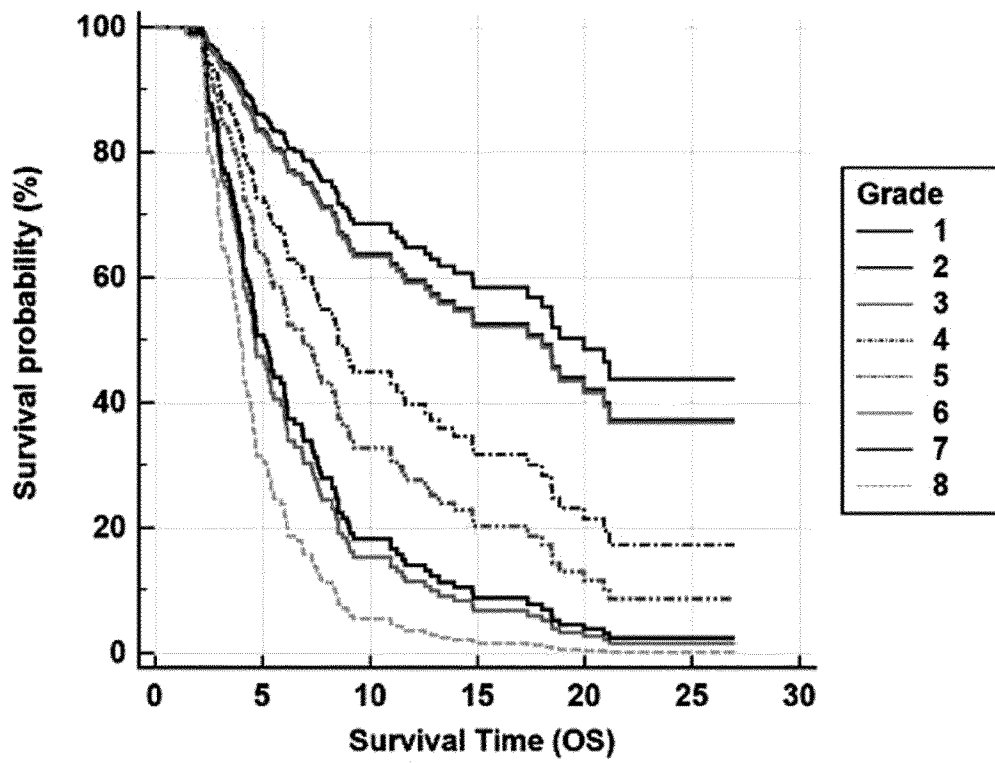
[도4]



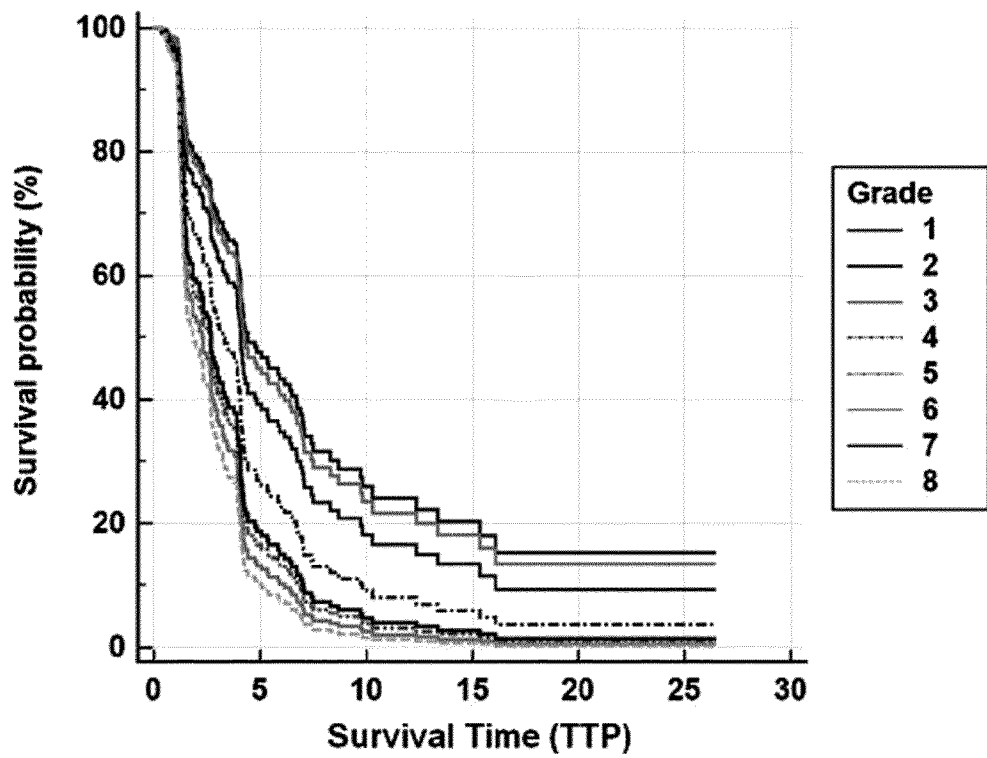
[도5]



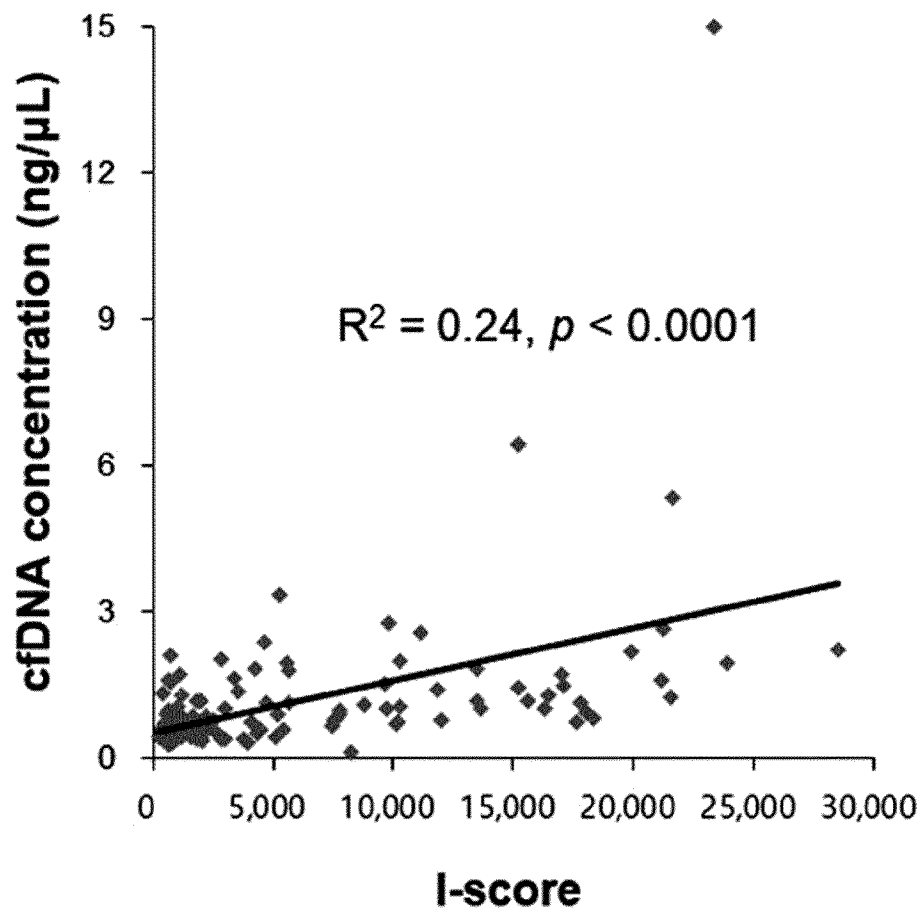
[도6]



[도7]



[도8]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2020/002359

A. CLASSIFICATION OF SUBJECT MATTER

G16B 30/10(2019.01)i, G16B 40/00(2019.01)i, G16B 50/00(2019.01)i, G16H 50/50(2018.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G16B 30/10; C12Q 1/68; G06F 19/22; G06F 19/24; G16B 40/00; G16B 50/00; G16H 50/50

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models: IPC as above

Japanese utility models and applications for utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & Keywords: cell free DNA, reads, normalization, Z-score, I-score, Hepatocellular carcinoma, prognosis

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	XU, Hongtao et al. Non-invasive analysis of genomic copy number variation in patients with hepatocellular carcinoma by next generation DNA sequencing. Journal of Cancer. 2015, vol. 6, no. 3, pages 247-253 See abstract; pages 248-250, 252.	1-7,10,13-21
Y		8-9,11-12
Y	ZHAO, Chen et al. Detection of fetal subchromosomal abnormalities by sequencing circulating cell-free DNA from maternal plasma. Clinical Chemistry. 2015, vol. 61, Article No. 4, pages 608-616 See page 2.	8-9,11-12
A	KR 10-1884909 B1 (THE CHINESE UNIVERSITY OF HONG KONG) 02 August 2018 See the entire document.	1-21
A	WO 2017-212428 A1 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA et al.) 14 December 2017 See the entire document.	1-21
A	KR 10-2016-0013183 A (SEQUENOM, INC.) 03 February 2016 See the entire document.	1-21



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 26 MAY 2020 (26.05.2020)	Date of mailing of the international search report 26 MAY 2020 (26.05.2020)
Name and mailing address of the ISA/KR  Korean Intellectual Property Office Government Complex Daejeon Building 4, 189, Cheongsu-ro, Seo-gu, Daejeon, 35208, Republic of Korea Facsimile No. +82-42-481-8578	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2020/002359

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	OH, Chung Ryul et al. Genome-wide copy number alteration and VEGFA amplification of circulating cell-free DNA as a biomarker in advanced hepatocellular carcinoma patients treated with Sorafenib. BMC Cancer. 01 April 2019(Online publication date), vol. 19, Article No. 292, pages 1-13 See abstract, pages 2-4, 10-11.	1-21

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2020/002359

Patent document cited in search report	Publication date	Patent family member	Publication date
KR 10-1884909 B1	02/08/2018	CN 104662168 A	27/05/2015
		CN 107779506 A	09/03/2018
		EP 2864501 A2	29/04/2015
		EP 2864501 A4	16/03/2016
		EP 2864501 B1	08/08/2018
		EP 3456843 A1	20/03/2019
		JP 2015-527057 A	17/09/2015
		JP 2018-191640 A	06/12/2018
		JP 6371280 B2	08/08/2018
		KR 10-2015-0032708 A	27/03/2015
		KR 10-2018-0088922 A	07/08/2018
		KR 10-2020-0035507 A	03/04/2020
		KR 10-2096611 B1	02/04/2020
		US 2014-0100121 A1	10/04/2014
		US 2018-0202003 A1	19/07/2018
		US 2019-0264291 A1	29/08/2019
		WO 2013-190441 A2	27/12/2013
WO 2017-212428 A1	14/12/2017	CN 110168099 A	23/08/2019
		EP 3464644 A1	10/04/2019
		JP 2019-521673 A	08/08/2019
KR 10-2016-0013183 A	03/02/2016	CN 105555968 A	04/05/2016
		EP 3004383 A2	13/04/2016
		EP 3004383 B1	24/04/2019
		EP 3578670 A1	11/12/2019
		JP 2016-526879 A	08/09/2016
		JP 2018-186825 A	29/11/2018
		JP 6561046 B2	14/08/2019
		US 2016-0224724 A1	04/08/2016
		WO 2014-190286 A2	27/11/2014
		WO 2014-190286 A3	26/02/2015

A. 발명이 속하는 기술분류(국제특허분류(IPC))

G16B 30/10(2019.01)i, G16B 40/00(2019.01)i, G16B 50/00(2019.01)i, G16H 50/50(2018.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

G16B 30/10; C12Q 1/68; G06F 19/22; G06F 19/24; G16B 40/00; G16B 50/00; G16H 50/50

조사된 기술분야에 속하는 최소문헌 이외의 문헌

한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우))

eKOMPASS(특허청 내부 검색시스템) & 키워드: 무세포 DNA(cell free DNA), 서열정보(reads), 정규화(normalization), Z-점수(Z-score), I-점수(I-score), 간암(Hepatocellular carcinoma), 예후(prognosis)

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
X	XU, HONGTAO 등, 'Non-invasive analysis of genomic copy number variation in patients with hepatocellular carcinoma by next generation DNA sequencing', Journal of Cancer, 2015, 6권, 3호, 페이지 247-253 요약; 페이지 248-250, 252 참조.	1-7,10,13-21
Y		8-9,11-12
Y	ZHAO, CHEN 등, 'Detection of fetal subchromosomal abnormalities by sequencing circulating cell-free DNA from maternal plasma', Clinical Chemistry, 2015, 61권, Article No. 4, 페이지 608-616 페이지 2 참조.	8-9,11-12
A	KR 10-1884909 B1 (더 차이니스 유니버시티 오브 홍콩) 2018.08.02 전체 문헌 참조.	1-21
A	WO 2017-212428 A1 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA 등) 2017.12.14 전체 문헌 참조.	1-21
A	KR 10-2016-0013183 A (시쿼넘, 인코포레이티드) 2016.02.03 전체 문헌 참조.	1-21

☒ 추가 문헌이 C(계속)에 기재되어 있습니다.

☒ 대응특허에 관한 별지를 참조하십시오.

* 인용된 문헌의 특별 카테고리:

“A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌

“D” 본 국제출원에서 출원인이 인용한 문헌

“E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후 “X”에 공개된 선출원 또는 특허 문헌

“L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌

“O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌

“P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌

“T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌

“X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다.

“Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다.

“&” 동일한 대응특허문헌에 속하는 문헌

국제조사의 실제 완료일

2020년 05월 26일 (26.05.2020)

국제조사보고서 발송일

2020년 05월 26일 (26.05.2020)

ISA/KR의 명칭 및 우편주소



대한민국 특허청
(35208) 대전광역시 서구 청사로 189,
4동 (둔산동, 정부대전청사)

팩스 번호 +82-42-481-8578

심사관

한인호

전화번호 +82-42-481-3362



C (계속). 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
PX	OH, CHUNG RYUL 등, 'Genome-wide copy number alteration and VEGFA amplification of circulating cell-free DNA as a biomarker in advanced hepatocellular carcinoma patients treated with Sorafenib' , BMC Cancer, 2019.04.01(온라인 공개일), 19권, Article No. 292, 페이지 1-13 요약; 페이지 2-4, 10-11 참조.	1-21

국제조사보고서에서
인용된 특허문헌

공개일

대응특허문헌

공개일

KR 10-1884909 B1

2018/08/02

CN 104662168 A

2015/05/27

CN 107779506 A

2018/03/09

EP 2864501 A2

2015/04/29

EP 2864501 A4

2016/03/16

EP 2864501 B1

2018/08/08

EP 3456843 A1

2019/03/20

JP 2015-527057 A

2015/09/17

JP 2018-191640 A

2018/12/06

JP 6371280 B2

2018/08/08

KR 10-2015-0032708 A

2015/03/27

KR 10-2018-0088922 A

2018/08/07

KR 10-2020-0035507 A

2020/04/03

KR 10-2096611 B1

2020/04/02

US 2014-0100121 A1

2014/04/10

US 2018-0202003 A1

2018/07/19

US 2019-0264291 A1

2019/08/29

WO 2013-190441 A2

2013/12/27

WO 2017-212428 A1

2017/12/14

CN 110168099 A

2019/08/23

EP 3464644 A1

2019/04/10

JP 2019-521673 A

2019/08/08

KR 10-2016-0013183 A

2016/02/03

CN 105555968 A

2016/05/04

EP 3004383 A2

2016/04/13

EP 3004383 B1

2019/04/24

EP 3578670 A1

2019/12/11

JP 2016-526879 A

2016/09/08

JP 2018-186825 A

2018/11/29

JP 6561046 B2

2019/08/14

US 2016-0224724 A1

2016/08/04

WO 2014-190286 A2

2014/11/27

WO 2014-190286 A3

2015/02/26