(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2013-0088449
(43) 공개일자 2013년08월08일

(51) 국제특허분류(Int. Cl.)
G01N 33/68 (2006.01) *C07K 14/02* (2006.01)
C12Q 1/68 (2006.01)
(21) 출원번호 10-2012-0009688
(22) 출원일자 2012년01월31일
심사청구일자 2012년01월31일

(71) 출원인
울산대학교 산학협력단
울산 남구 무거2동 산29
(72) 발명자
오흥범
서울특별시 송파구 풍납동 388-1 서울아산병원 진단검사의학과
이정훈
서울특별시 서초구 방배동 593-64 명지해드는터 601
이종은
경기도 고양시 일산서구 일산동 후곡마을 1403-1403
(74) 대리인
특허법인태백

전체 청구항 수 : 총 8 항

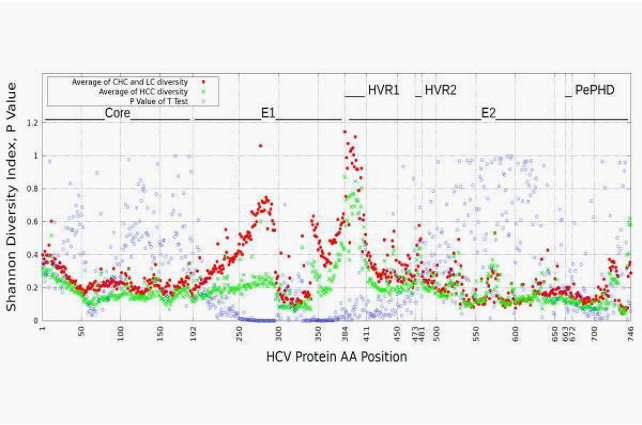
(54) 발명의 명칭 **헤파티티스 C 바이러스로 인한 간암 진단용 마커 및 진단방법**

(57) 요 약

본 발명은 HCV로 인한 간암 진단용 마커 조성물 및 HCV로 인한 간암 진단 방법에 관한 것으로, 보다 구체적으로는 HCV(Hepatitis C Virus)에 감염된 검체로부터 HCV RNA를 검출하여 염기 서열을 시퀀싱하는 단계; 상기 시퀀싱 단계로부터 얻은 염기 서열 정보를 아미노산 서열 정보로 변환하는 단계; 및 서열번호 1의 아미노산 서열로 기재되는 HCV(Hepatitis C Virus)의 레퍼런스 아미노산 서열의 282번째 아미노산; 283번째 아미노산; 285번째 아미노산; 287번째 아미노산; 288번째 아미노산; 290번째 아미노산; 291번째 아미노산; 292번째 아미노산; 293번째 아미노산; 331번째 아미노산; 332번째 아미노산; 338번째 아미노산; 339번째 아미노산; 339번째 아미노산; 및 341번째 아미노산으로 이루어진 군에서 선택된 하나 또는 둘 이상의 아미노산과 대응하는 상기 변환된 아미노산 서열 정보를 비교하는 단계를 포함하는 HCV로 인한 간암 진단에 필요한 정보를 제공하는 방법에 관한 것이다.

본 발명에 따르면, HCV의 아미노산 서열 중 14개 위치의 서열 분석 및 이의 다양성 분석을 통해 만성 C형 간염 환자와 간세포암 환자를 높은 정확도로 구별할 수 있고, 이에 따라 간세포암의 발병 여부를 간편하게 진단할 수 있다.

대 표 도 - 도20



이 발명을 지원한 국가연구개발사업

과제고유번호	2011-0207
부처명	교육과학기술부
연구사업명	미래원천기술개발사업
연구과제명	HTS 나노박막센서를 이용한 바이오센싱 및 의학연구응용[1]
주관기관	울산대학교 산학협력단
연구기간	2011.03.01 ~ 2012.02.29

특허청구의 범위

청구항 1

서열번호 1의 아미노산 서열로 기재되는 HCV(Hepatitis C Virus)의 아미노산 서열 내에서,

282번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

283번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

285번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

287번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

288번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

290번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

291번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

292번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

293번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

331번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

332번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

338번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

339번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드; 및

341번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드; 로 이루어진 군에서 선택된 하나 또는 둘 이상의 조합을 포함하는 HCV로 인한 간암 진단용 마커 조성물.

청구항 2

HCV(Hepatitis C Virus)에 감염된 검체로부터 HCV RNA를 검출하여 염기 서열을 시퀀싱하는 단계;

상기 시퀀싱 단계로부터 얻은 염기 서열 정보를 아미노산 서열 정보로 변환하는 단계; 및

서열번호 1의 아미노산 서열로 기재되는 HCV(Hepatitis C Virus)의 레퍼런스 아미노산 서열의 282번째 아미노산, 283번째 아미노산, 285번째 아미노산, 287번째 아미노산, 288번째 아미노산, 290번째 아미노산, 291번째 아미노산, 292번째 아미노산, 293번째 아미노산, 331번째 아미노산, 332번째 아미노산, 338번째 아미노산, 339번째 아미노산, 339번째 아미노산 및 341번째 아미노산으로 이루어진 군에서 선택된 하나 또는 둘 이상의 아미노산과 대응하는 상기 변환된 아미노산 서열 정보를 비교하는 단계;를 포함하는 HCV로 인한 간암 진단에 필요한 정보를 제공하는 방법.

청구항 3

제2항에 있어서,

상기 HCV(Hepatitis C Virus)는 1b 유전자형인 것인 HCV로 인한 간암 진단에 필요한 정보를 제공하는 방법.

청구항 4

제2항에 있어서,

상기 시퀀싱하는 단계는 에멀전 기반 클론 PCR(emulsion vased clonal PCR) 방법에 의해 시퀀싱하는 것인 HCV로 인한 간암 진단에 필요한 정보를 제공하는 방법.

청구항 5

제2항에 있어서,

상기 시퀀싱하는 단계에서의 HCV(Hepatitis C Virus)의 염기서열은 HCV의 구조단백질인 E1 영역인 것인 HCV로 인한 간암 진단에 필요한 정보를 제공하는 방법.

청구항 6

제2항에 있어서,

상기 시퀀싱하는 단계는

서열번호 3 또는 서열번호 5의 핵산서열을 갖는 정방향 프라이머 및 서열번호 4 또는 서열번호 6의 핵산서열을 갖는 역방향 프라이머로 이루어진 프라이머세트; 및

서열번호 7 또는 서열번호 9의 핵산서열을 갖는 정방향 프라이머 및 서열번호 8 또는 서열번호 10의 핵산서열을 갖는 역방향 프라이머로 이루어진 프라이머 세트; 를 사용하여 수행되는 것인 HCV로 인한 간암 진단에 필요한 정보를 제공하는 방법.

청구항 7

제2항에 있어서,

상기 HCV(Hepatitis C Virus)는 다양한 변이 형태로 존재하는 것인 HCV로 인한 간암 진단에 필요한 정보를 제공하는 방법.

청구항 8

제2항에 있어서,

상기 비교하는 단계 후, 대응하는 위치에서 다양한 변이 형태로 존재하는 HCV의 아미노산 서열 다양성을 분석하는 단계를 추가로 포함하는 HCV로 인한 간암 진단에 필요한 정보를 제공하는 방법.

명 세 서

기 술 분 야

[0001] 본 발명은 헤파티티스 C 바이러스로 인한 간암 진단용 마커 및 이를 통해 헤파티티스 C 바이러스로 인한 간암을 진단하는 방법에 관한 것이다.

배 경 기 술

[0002] 간세포암은 대부분 진행되고 수술할 수 없는 상태에서 검출되는 경우가 많아서 그 예후가 나쁘다. 수술해서 환치할 수 있는 경우는 간세포암의 크기가 작고 국한되어 있을 때이다. 이러한 이유로 위험이 높은 집단에서는 효과적으로 스크리닝하여 간세포암을 조기에 발견하는 것이 중요하다. 그러나 현재까지는 a-fetoprotein (AFP) 외에 일반적으로 받아들여지는 혈액검사는 없고, 따라서 간경변증이 있는 환자들에서는 6개월마다 초음파와 AFP 검사의 시행이 권장되고 있다. 그러나 AFP의 진단적 정확도도 낮아 만족스럽지 못해 간세포암에 대한 바이오마커의 요구가 매우 크다. AFP는 바이러스성 간세포암에 대해 비바이러스성 간세포암보다 양성예측도가 낮다. 또한 AFP의 특이도는 cut-off 수치에 따라 특이도가 76~96% 정도이고, 민감도는 간세포암이 절제 가능한 3cm 이하일 경우 약 25%이고 3cm 이상일 경우 약 50%이다. 그 외에 PIVKA-II라고도 불리는 des-r-carboxyprothrombin

(DCP)도 간세포암의 표지자로 사용된다. 전체 간세포암 환자의 50~60%에서 DCP가 증가하나 조기 간세포암 중에서는 15~30%만 해당된다. 이 검사의 민감도는 48~62%이고 특이도는 81~92%로 간세포암의 표지자로서 DCP의 진단적 가치는 AFP와 비슷하다. 그 외에 a-l-fucosidase, r-glutamyl transferase, glypican-3, squamous cell carcinoma antigen 등이 있으나 민감도와 특이도가 만족할만한 표지자는 아직 없는 실정이다.

[0003] C형 간염 바이러스(hepatitis C virus, HCV)는 외피를 가지는 RNA 바이러스로서, 숙주세포에서 오류교정 능력이 없는 RNA 의존성 RNA 중합효소에 의해 복제되기 때문에 다양한 변이형(quasispecies)을 만들어낸다. HCV 변이형(quasispecies)의 다양성의 정도는 숙주의 HCV에 대한 면역학적 압력에 의해 결정되는 것으로 알려져 있다. 면역반응이 강하여 심한 임상증상을 보이는 만성 C형 간염(chronic hepatitis C, CHC) 환자에서는 그렇지 않은 환자에 비해 HCV 변이형(quasispecies)의 다양성이 더 높다.

[0004] HCV에 감염된 환자의 약 80%가 만성 간염 상태로 이행하며, 이 중 약 20%의 환자에서 20여년의 기간을 거치면서 간경변증으로 진행하게 된다. 그리고 간경변증이 발생하게 되면 간세포암(hepatocellular carcinoma, HCC)의 발생위험도는 연간 약 5%에 달한다. 이러한 HCV 감염의 자연경과 동안 HCV에 대한 숙주의 면역반응은 감소하므로 만성 C형 간염에서 간세포암으로 진행함에 따라 HCV 변이형의 다양성은 낮아질 것으로 예상된다.

발명의 내용

해결하려는 과제

[0005] 따라서 본 발명은 상기 HCV의 변이 정도를 통해 간세포암으로의 진행 여부를 판별할 수 있는 헤파티티스 C 바이러스로 인한 간암 진단용 마커 및 이를 통해 헤파티티스 C 바이러스로 인한 간암을 진단하는 방법을 제공하고자 한다.

과제의 해결 수단

[0006] 상기 과제의 해결을 위하여, 본 발명은 서열번호 1의 아미노산 서열로 기재되는 HCV(Hepatitis C Virus)의 아미노산 서열 내에서,

[0007] 282번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0008] 283번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0009] 285번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0010] 287번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0011] 288번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0012] 290번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0013] 291번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0014] 292번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0015] 293번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0016] 331번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0017] 332번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0018] 338번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0019] 339번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드; 및

[0020] 341번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드; 로 이루어진 군에서 선택된 하나 또는 둘 이상의 조합을 포함하는 HCV로 인한 간암 진단용 마커 조성물을 제공한다.

[0021] 또한 본 발명은 HCV(Hepatitis C Virus)에 감염된 검체로부터 HCV RNA를 검출하여 염기 서열을 시퀀싱하는 단계; 상기 시퀀싱 단계로부터 얻은 염기 서열 정보를 아미노산 서열 정보로 변환하는 단계; 및 서열번호 1의 아미노산 서열로 기재되는 HCV(Hepatitis C Virus)의 레퍼런스 아미노산 서열의 282번째 아미노산, 283번째 아미노산, 285번째 아미노산, 287번째 아미노산, 288번째 아미노산, 290번째 아미노산, 291번째 아미노산, 292번

째 아미노산, 293번째 아미노산, 331번째 아미노산, 332번째 아미노산, 338번째 아미노산, 339번째 아미노산, 339번째 아미노산 및 341번째 아미노산으로 이루어진 군에서 선택된 하나 또는 둘 이상의 아미노산과 대응하는 상기 변환된 아미노산 서열 정보를 비교하는 단계를 포함하는 HCV로 인한 간암 진단에 필요한 정보를 제공하는 방법을 제공한다.

발명의 효과

[0022] 본 발명에 따르면, HCV의 아미노산 서열 중 14개 위치의 서열 분석 및 이의 다양성 분석을 통해 만성 C형 간염 환자와 간세포암 환자를 높은 정확도로 구별할 수 있고, 이에 따라 간세포암의 발병 여부를 간편하게 진단할 수 있다.

도면의 간단한 설명

[0023] 도 1은 Roche GS FLX를 사용하여 UDPS를 시행한 과정에 대한 요약도이다.

도 2는 각 검체에서 UDPS로 나온 read의 개수와 기준주에 정렬되는read 의 개수에 대한 그래프이다.

도 3은 각 검체별 평균 read 의 길이를 나타낸 그래프이다.

도 4는 게놈상 각 위치별로 read가 분포하는 범위에 대한 그래프이다.

도 5는 간경변증이 없는 만성 C형 간염 환자들에 대한 HCV 염기서열의 각 위치에서 구한 Shannon diversity index이다.

도 6는 간세포암 환자들에 대한 HCV 염기서열의 각 위치에서 구한 Shannon diversity index이다.

도 7은 간경변증이 없는 만성 C형 간염 환자군과 HCV로 인한 간세포암이 발생한 환자군에서 HCV 변이형의 다양성에 차이가 있는지 확인하기 위해 HCV 염기서열의 각 위치에서 T 검정을 시행한 결과 그래프이다.

도 8은 간경변증이 없는 만성 C형 간염 환자들에 대한 HCV 아미노산 서열의 각 위치에서 구한 Shannon diversity index이다.

도 9는 간세포암 환자들에 대한 HCV 아미노산 서열의 각 위치에서 구한 Shannon diversity index이다.

도 10은 간경변증이 없는 만성 C형 간염 환자군과 HCV로 인한 간세포암이 발생한 환자군에서 HCV quasispecies 아미노산 서열의 다양성에 차이가 있는지 확인하기 위해 HCV 아미노산 서열의 각 위치에서 T 검정을 시행한 결과 그래프이다.

도 11은 간경변증이 없는 만성 C형 간염으로부터 HCV로 인해 발생한 간세포암을 구별하는 통계적으로 유의한 염기서열상의 각 위치별 receiver operating characteristic (ROC) 커브를 나타낸다.

도 12는 간경변증이 없는 만성 C형 간염으로부터 HCV로 인해 발생한 간세포암을 구별하는 통계적으로 유의한 아미노산 서열상의 각 위치별 receiver operating characteristic (ROC) 커브를 나타낸다.

도 13은 간경변증이 있는 만성 C형 간염 환자들에 대한 HCV 염기서열의 각 위치에서 구한 Shannon diversity index이다.

도 14는 간경변증이 있는 만성 C형 간염 환자들에 대한 HCV 아미노산 서열의 각 위치에서 구한 Shannon diversity index이다.

도 15는 간경변증이 있는 만성 C형 간염 환자군과 간경변증이 없는 만성 C형 간염 환자군에 대해 HCV 염기서열상의 각 위치에서 T 검정을 시행한 결과 그래프이다.

도 16은 간경변증이 있는 만성 C형 간염 환자군과 HCV로 인한 간세포암이 발생한 환자군에 대해 HCV 염기서열상의 각 위치에서 T 검정을 시행한 결과 그래프이다.

도 17은 만성 C형 간염 환자군과 간경변증이 없는 만성 C형 간염 환자군에 대해 HCV 아미노산 서열상의 각 위치에서 T 검정을 시행한 결과 그래프이다.

도 18은 간경변증이 있는 만성 C형 간염 환자군과 HCV로 인한 간세포암이 발생한 환자군에 대해 HCV 아미노산 서열상의 각 위치에서 T 검정을 시행한 결과 그래프이다.

도 19는 간경변증의 유무에 관계 없이 만성 C형 간염 환자들과 HCV로 인한 간세포암이 발생 환자들 사이에 염기

서열 다양성에 차이가 존재하는지 여부에 관한 T 검정 결과이다.

도 20은 간경변증의 유무에 관계 없이 만성 C형 간염 환자들과 HCV로 인한 간세포암이 발생 환자들 사이에 아미노산 서열 다양성에 차이가 존재하는지 여부에 관한 T 검정 결과이다.

발명을 실시하기 위한 구체적인 내용

[0024] 본 발명은 서열번호 1의 아미노산 서열로 기재되는 HCV(Hepatitis C Virus)의 아미노산 서열 내에서,

[0025] 282번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0026] 283번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0027] 285번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0028] 287번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0029] 288번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0030] 290번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0031] 291번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0032] 292번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0033] 293번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0034] 331번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0035] 332번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0036] 338번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드;

[0037] 339번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드; 및

[0038] 341번째 아미노산을 포함하는 20 내지 100개의 연속적인 폴리펩티드; 로 이루어진 군에서 선택된 하나 또는 둘 이상의 조합을 포함하는 HCV로 인한 간암 진단용 마커 조성물을 제공한다.

[0039] HCV는 숙주에서 다양한 변이형(quasispecies)으로 존재하는 특징이 있는데, 변이형은 일반적으로 전체 게놈의 9% 이내에서 염기서열 차이를 보이는 집단으로 정의한다. HCV가 숙주에서 변이형으로 존재하는 것은 바이러스 RNA 중합효소가 오류교정(proofreading)능력이 없어서 바이러스 복제 과정 중에 오류가 생기기 쉽기 때문이다. 이렇게 생성되는 변이주들 중 숙주의 면역반응에 대한 적응도가 높은 변이주들이 변이형을 구성하여 변이형 집단 전체가 하나의 단위로서 면역회피 기능을 수행한다. HCV 면역회피 변이는 프로테오좀 공정, MHC 결합, T 세포의 인지 기능에 영향을 미쳐 T 세포의 공격을 피할 수 있게 된다.

[0040] 본 발명의 발명자들은 HCV로 인한 만성 C형 간염 환자의 HCV 유전자서열과 HCV로 인한 간암 환자의 HCV 유전자 서열을 분석하여, 간암이 발생한 환자들에게서는 만성 C형 간암 환자들에 비해 HCV의 E1 부위의 다양성이 유의하게 감소한다는 것을 확인하였고, 보다 구체적으로 이러한 다양성이 감소하는 특정 HCV 아미노산 위치를 확인하고 이를 포함하는 간암 진단용 단백질성 마커들을 발굴하였다.

[0041] 본 발명은 또한 HCV(Hepatitis C Virus)에 감염된 검체로부터 HCV RNA를 검출하여 염기 서열을 시퀀싱하는 단계; 상기 시퀀싱 단계로부터 얻은 염기 서열 정보를 아미노산 서열 정보로 변환하는 단계; 및 서열번호 1의 아미노산 서열로 기재되는 HCV(Hepatitis C Virus)의 레퍼런스 아미노산 서열의 282번째 아미노산, 283번째 아미노산, 285번째 아미노산, 287번째 아미노산, 288번째 아미노산, 290번째 아미노산, 291번째 아미노산, 292번째 아미노산, 293번째 아미노산, 331번째 아미노산, 332번째 아미노산, 338번째 아미노산, 339번째 아미노산, 339번째 아미노산 및 341번째 아미노산으로 이루어진 군에서 선택된 하나 또는 둘 이상의 아미노산과 대응하는 상기 변환된 아미노산 서열 정보를 비교하는 단계를 포함하는 HCV로 인한 간암 진단에 필요한 정보를 제공하는 방법을 제공한다.

[0042] 환자에서 동정되는 HCV 유전자형은 전 세계적으로 지역적인 차이를 보이는데, 한국의 경우는 1b형(약 50%)과 2a형(약 40%)이 주종을 이루며, HCV 유전자형 1b 감염되면 다른 유전자형에 비해서 간세포암이 발생할 확률이 더

높다고 알려져 있다. 본 발명의 한 구체예에서, 상기 HCV(Hepatitis C Virus)는 1b 유전자형일 수 있으나, 이에 제한되는 것은 아니다.

[0043] 본 발명의 한 구체예에서, 상기 HCV RNA의 검출 대상 시료는 HCV(Hepatitis C Virus)에 감염된 검체의 혈액, 간세포 또는 간조직일 수 있으나 이에 제한되는 것은 아니며, 상기 '검체'로는 인간을 포함하는 포유동물일 수 있으나, 이들 예에 한정되는 것은 아니다.

[0044] 상기 검체로부터 RNA를 추출하는 방법은 특별히 한정되지 않으며, 당업계에 공지된 기술 또는 시판되고 있는 핵산 추출용 키트를 사용할 수 있다. 예를 들면, 페놀/클로로포름 추출법 및 프로테아제 K 처리방법을 사용할 수 있다.

[0045] RNA 추출용 키트는 Qiagen, Inc, 및 Stratagene에서 구입할 수 있다. 상기에서 RNA를 추출하여 사용하는 경우에는 역전사를 통해 cDNA를 제조하여 사용할 수 있다.

[0046] 상기 시퀀싱 단계에서의 서열분석은 당업계에 공지된 방법을 모두 이용할 수 있으며, 구체적으로는 이에 제한되는 것은 아니나, 자동염기서열분석기를 사용하거나, 파이로시퀀싱(pyrosequencing), PCR-RELP법(restriction fragment length polymorphism), PCRSSCP법(single strand conformation polymorphism), PCR-SSO법(specific sequence oligonucleotide), PCR-SSO법과 도트 하이브리드화법을 조합한 ASO(allele specific oligonucleotide) 하이브리드화법, TaqMan-PCR법, MALDI-TOF/MS법, RCA법(rolling circle amplification), HRM(high resolution melting)법, 프라이머 신장법, 서던 블롯 하이브리드화법 및 도트 하이브리드화법 등의 공지의 방법 중 선택된 어느 하나 이상을 사용할 수 있다.

[0047] 본 발명의 일실시예에서, 상기 시퀀싱하는 단계는 에멀전 기반 클론 PCR(emulsion vased clonal PCR) 방법에 의해 수행될 수 있다. 본 발명의 시퀀싱은 예를 들어, Roche 사의 게놈 Sequencer FLX(GS FLX) 시스템을 이용할 수 있다. 상기 시스템은 현재 사용 가능한 시퀀싱 기술 중에 염기 서열을 가장 길게 분석 할 수 있는 것으로서, 에멀전 기반 클론 PCR 방법에 의해 시퀀싱을 수행한다. Roche GS FLX 기술을 미생물의 동일 부위가 증폭된 조각들을 염기서열 분석하여 다수의 유전적 변이형들을 분석하는데 사용하는 것을 ultra-deep pyrosequencing (UDPS)라고 하는데, 이는 바이러스의 변이형 성상을 파악하는데 매우 유용하다. NGS(next generation sequencing) 기술을 이용한 UDPS 방법을 HCV 감염 환자들에서 HCV의 변이형 분석에 적용하면 수천 개의 클론을 분석하는 효과를 낼 수 있기 때문에 기존 방법들에 비해 변이형의 다양성에 대하여 정확하고 민감하게 분석할 수 있다. 상기 시퀀싱의 수행 방법은 Roche 사의 프로토콜에 따른다.

[0048] 본 발명의 또 다른 구체예에서, 상기 시퀀싱하는 단계에서의 HCV의 염기서열은 HCV의 구조단백질인 E1 영역일 수 있다. HCV의 유전체는 9.5 kb 길이의 단일 가닥 positive sense RNA로서 1개의 해독틀(open reading frame, ORF)에 의해 3,011개의 아미노산으로 이루어지는 다기능단백질(polyprotein)로 해독된다. 해독된 다기능단백질은 숙주의 단백분해효소 및 바이러스 유래의 단백분해효소에 의해 3개의 구조단백질인 core, E1, E2와 7개의 비구조단백질인 p7, NS2, NS3, NS4A, NS4B, NS5A, NS5B로 잘려지는데, 이 중 E1 부위를 분석하여 간암 진단에 유용하게 사용할 수 있다.

[0049] 본 발명의 또 다른 구체예에서, 상기 시퀀싱하는 단계는 서열번호 3 또는 서열번호 5의 핵산서열을 갖는 정방향 프라이머 및 서열번호 4 또는 서열번호 6의 핵산서열을 갖는 역방향 프라이머로 이루어진 프라이머세트; 및 서열번호 7 또는 서열번호 9의 핵산서열을 갖는 정방향 프라이머 및 서열번호 8 또는 서열번호 10의 핵산서열을 갖는 역방향 프라이머로 이루어진 프라이머 세트; 를 사용하여 수행할 수 있다.

[0050] 보다 구체적으로, E1 부위를 분석하는 경우, 서열번호 7 또는 서열번호 9의 핵산서열을 갖는 정방향 프라이머 및 서열번호 8 또는 서열번호 10의 핵산서열을 갖는 역방향 프라이머로 이루어진 프라이머 세트를 사용하여 수행할 수 있으며, 보다 더 구체적으로는, 서열번호 7의 핵산서열을 갖는 정방향 프라이머 및 서열번호 8의 핵산서열을 갖는 역방향 프라이머로 이루어진 프라이머 세트를 사용하여 1차 PCR을 수행하고, 서열번호 9의 핵산서열을 갖는 정방향 프라이머 및 서열번호 10의 핵산서열을 갖는 역방향 프라이머로 이루어진 프라이머 세트를 사용하여 2차 PCR을 수행할 수 있다.

[0051] 이후, 상기 방법은 상기 시퀀싱 단계로부터 얻은 염기 서열 정보를 아미노산 서열 정보로 변환하는 단계를 포함한다.

[0052] 상기 시퀀싱 장치(예를 들어, Roche 사의 게놈 Sequencer FLX(GS FLX) 시스템으로부터 시권싱이 완료가 되면,

이로부터 얻은 염기 서열 정보는 open reading frame(ORF)에 따라 아미노산 서열 정보로 변환하게 된다. 이후, 서열번호 1의 아미노산 서열로 기재되는 HCV(Hepatitis C Virus)의 레퍼런스 아미노산 서열의 282번째 아미노산; 283번째 아미노산; 285번째 아미노산; 287번째 아미노산; 288번째 아미노산; 290번째 아미노산; 291번째 아미노산; 292번째 아미노산; 293번째 아미노산; 331번째 아미노산; 332번째 아미노산; 338번째 아미노산; 339번째 아미노산; 339번째 아미노산; 및 341번째 아미노산으로 이루어진 군에서 선택된 하나 또는 둘 이상의 아미노산과 대응하는 상기 변환된 아미노산 서열 정보를 비교하는 단계가 수행된다.

[0053] 서열번호 1의 아미노산 서열로 기재되는 HCV의 레퍼런스 아미노산 서열은 검출된 HCV RNA의 전체 또는 일부 단백질의 아미노산 서열과 비교 대상이 되는 표준이 되는 아미노산 서열이며, 이는 서열번호 1의 아미노산 서열로 기재된다. 서열번호 1의 아미노산 서열의 282번째 아미노산; 283번째 아미노산; 285번째 아미노산; 287번째 아미노산; 288번째 아미노산; 290번째 아미노산; 291번째 아미노산; 292번째 아미노산; 293번째 아미노산; 331번째 아미노산; 332번째 아미노산; 338번째 아미노산; 339번째 아미노산; 339번째 아미노산; 및 341번째 아미노산으로 이루어진 군에서 선택된 하나 또는 둘 이상의 아미노산과 대응하는 상기 변환된 아미노산 서열 정보를 비교하여 서로 상이한 아미노산으로 나타나는 경우, 이는 아미노산 변이가 존재하는 것으로 판단되며, 이를 바탕으로 다양성을 분석하여, 다양성의 정도가 감소될수록 간암으로 진단할 수 있다.

[0054] 상기 다양성은 변이의 다양성을 의미하며, 이는 다양도지수(diversity index), 보다 구체적으로 Shannon Diversity Index로 나타낼 수 있다. 따라서 본 발명의 한 구체예에서, 상기 다양성 분석은 다양성 지수로 계량화하는 단계를 포함할 수 있다.

[0055] 보다 구체적으로, 본 발명의 HCV(Hepatitis C Virus)는 감염된 검체 내에서 다양한 변이 형태로 존재할 수 있다. 상기 다양한 변이 형태로 존재하는 HCV의 아미노산 서열 정보를 바탕으로 분석하여, 대응하는 각 위치에서의 다양성을 분석하고 다양성 지수로 계량화하여, 다양성의 정도가 감소될수록 검체를 간암이라 진단할 수 있다. 즉, 감염된 검체 내에 존재하는 HCV(Hepatitis C Virus)의 변이 형태가 다양하지 않을수록, 간암으로 진단할 수 있다.

[0056] 보다 구체적으로, 다양한 변이 형태로 존재하는 HCV의 아미노산 서열 정보를 바탕으로, 서열번호 1의 아미노산 서열의 282번째 아미노산; 283번째 아미노산; 285번째 아미노산; 287번째 아미노산; 288번째 아미노산; 290번째 아미노산; 291번째 아미노산; 292번째 아미노산; 293번째 아미노산; 331번째 아미노산; 332번째 아미노산; 338번째 아미노산; 339번째 아미노산; 339번째 아미노산; 및 341번째 아미노산으로 이루어진 군에서 선택된 하나 또는 둘 이상의 아미노산 위치와 대응하는 위치에서 분석된 다양성 지수가 낮게 나타날수록 간암으로 진단할 수 있다.

[0057] 이하, 본 발명의 이해를 돕기 위하여 실시예를 들어 상세하게 설명하기로 한다. 다만 하기의 실시예는 본 발명의 내용을 예시하는 것일 뿐 본 발명의 범위가 하기 실시예에 한정되는 것은 아니다. 본 발명의 실시예는 당업계에서 평균적인 지식을 가진 자에게 본 발명을 보다 완전하게 설명하기 위해 제공되는 것이다.

[0058] **<실험예>**

[0059] 하기의 실험예들은 본 발명에 따른 각각의 실시예에 공통적으로 적용되는 실험예를 제공하기 위한 것이다.

[0060] **1. 환자 검체**

[0061] 총 49명의 HCV 감염자의 검체를 사용하였다. 이 중 26명 (검체번호 1-26)은 만성 C형 간염 환자였고, 23명 (검체번호 27-49)은 HCV로 인한 간세포암이 발생한 환자였다. 26명의 만성 C형 간염 환자들 중 20명 (검체번호 1-20)은 간경변증이 없고, 6명 (검체번호 21-26)은 간경변증이 있었다. 모든 환자는 HCV 유전자형 1b형에 감염되었고, 모든 검체의 HCV RNA정량값은 1.0×10^4 IU/mL 이상이었다. 두 환자군 검체의 특성을 표 1에 요약하였다.

[0062] HCV RNA 정량 검사는 Abbott m2000rt (Abbott Molecular Inc., Abbott Park, IL, USA)를 이용하여 Abbott RealTime HCV Quantification Kit (Abbott Molecular Inc.)로 시행하였다. Abbott m2000sp System (Abbott Molecular Inc.)을 이용하여 혈청 200 μL에서 HCV RNA를 추출한 후, 이 HCV RNA에 50 μL의 반응혼합물을 혼합하고 반응 플레이트에 옮긴 후 Abbott m2000rt로 증폭과정을 시행하였다. 검출 민감도는 12 IU/mL이었다.

표 1

no	sex	age	HCV infection diag date	LC	HCC diag date	test date	RIIA IU/mL	AST	ALT	albumin	platelet count	bilirubin	PT (%)	IIR	Cr
1	F	54	2007.01	II		20110304	5.0x10(6)	26	21	4	216	0.9	98.9	1.01	0.96
2	F	72	1999	II		20110304	2.5x10(6)	96	87	3.8	90	1.2	104.3	0.98	0.8
3	M	44	2000	II		20110304	2.1x10(6)	42	100	4.4	272	0.7	97.2	1.02	0.9
4	M	70	2007.01	II		20110304	2.6x10(6)	44	41	4.2	200	1	108	0.96	0.85
5	F	71	2008.06	II		20110217	1.9x10(6)	49	34	3.8	147	1.1	IIA	IIA	0.9
6	F	66	2006	II		20110217	2.5x10(6)	24	13	2.6	96	1	71	1.2	0.8
7	M	83	2008.01	II		20110217	1.9x10(6)	44	22	3.7	106	0.9	97.3	1.02	0.9
8	M	59	2010.07	II		20110221	1.0x10(6)	22	25	3.5	89	0.5	86.6	1.08	12.4
9	M	72	2007.07	II		20110221	2.8x10(5)	34	41	4	184	0.8	98.9	1.01	0.9
10	F	57	2007.11	II		20110222	1.5x10(6)	38	44	4.5	201	1	92.5	1.05	0.8
11	F	57	2008.11	II		20110304	1.1x10(6)	96	124	4.3	210	0.8	86.6	1.08	0.8
12	F	82	2011.01	II		20110315	1.2x10(5)	161	224	3.9	268	0.9	110	0.96	0.6
13	M	35	2011.03	II		20110318	3.0x10(6)	43	68	4.3	245	0.8	110	0.96	0.8
14	F	42	2007.12	II		20110322	1.1x10(6)	36	63	4.3	207	0.8	95.6	1.03	0.7
15	F	49	2007.05	II		20110322	2.2x10(6)	38	19	4.1	194	0.8	110	0.96	0.8
16	F	66	2010.09	II		20110304	3.3x10(5)	42	39	4.3	206	1	108	0.97	0.6
17	F	52	2007.11	II		20110304	1.1x10(5)	33	22	4.1	303	0.7	156.9	0.79	0.8
18	M	72	2009.12	II		20110304	4.4x10(5)	95	92	3.8	134	0.8	80.1	1.13	1
19	M	42	2010.07	II		20110315	9.1x10(5)	98	191	4.2	136	0.8	106.1	0.97	0.8
20	F	58	2006.03	II		20110225	7.1x10(6)	38	35	4.4	204	1	116.2	0.93	0.7
21	M	73	1996	Y		20110304	2.3x10(6)	61	79	3.8	85	2.2	69	1.22	0.9
22	M	63	2001	Y		20110308	2.1x10(6)	291	209	3.5	55	1.6	88	1.07	0.9
23	M	63	2008	Y		20110308	1.8x10(6)	225	159	3.6	50	20.3	77.6	1.15	1
24	M	68	2005.06	Y		20110308	1.5x10(6)	51	31	3.4	65	1.4	89.5	1.07	0.8
25	M	43	IIA	Y		20110308	3.8x10(6)	29	19	4.4	64	1.3	94	1.04	0.7
26	M	50	2011.03	Y		20110324	6.3x10(5)	226	258	4.1	74	1.1	86.6	1.08	0.9
27	M	66	1991	Y	2007.05	20110304	3.2x10(6)	28	23	4.1	148	1.4	71.7	1.18	1
28	F	69	1999	Y	2006	20110304	1.3x10(7)	46	70	2.9	46	0.6	110	0.96	3.9
29	M	64	2008.01	Y	2008.01	20110304	2.4x10(6)	36	40	IIA	IIA	IIA	IIA	IIA	IIA
30	F	56	2007	Y	2010.01	20110315	4.1x10(5)	67	64	3.8	50	2.9	62.4	1.31	0.6
31	F	68	2000	Y	2008.06	20110315	2.7x10(5)	75	62	2.8	89	2.1	52.9	1.46	0.6
32	M	67	1997	Y	2007.04	20110317	1.3x10(6)	92	82	2.9	103	2.2	88	1.08	0.9
33	M	49	2010.01	Y	2009.12	20110322	4.3x10(5)	94	86	3.3	67	1.3	71	1.2	1
34	M	83	1997	Y	2009.05	20110125	5.0x10(5)	71	17	2.7	100	1	95.5	2.02	0.9
35	M	64	2009	Y	2010.07	20110125	2.6x10(6)	64	57	3.3	70	2.3	63.8	1.29	0.7
36	M	70	1993	Y	2009.09	20110225	1.1x10(6)	79	41	3.5	119	0.9	77.7	1.15	0.9
37	M	73	2008	Y	2010.11	20110304	1.0x10(6)	202	285	2.9	76	2.6	69	1.22	0.7
38	M	64	1994	Y	2009.10	20110315	2.9x10(6)	32	30	4	171	0.9	83.9	1.1	0.8
39	F	54	2005.03	Y	2010.03	20110315	1.3x10(5)	137	24	2.3	26	2.2	50.8	1.5	0.7
40	F	60	2002	Y	2010.11	20110304	1.8x10(6)	50	62	3.7	130	1.3	81.3	1.12	0.7
41	F	58	2009.08	Y	2010.11	20110310	3.6x10(5)	73	58	3.4	49	1.3	65.3	1.27	0.7
42	F	71	2007	Y	2010.09	20110118	1.3x10(5)	69	36	2.8	67	1.1	61.7	1.32	0.8
43	M	68	2002	Y	2007.12	20110118	9.3x10(5)	137	114	2.9	84	0.9	74.2	1.18	0.7
44	F	72	1980	Y	2009.01	20110201	2.5x10(5)	75	52	3.5	167	0.7	103.8	0.99	1.4
45	F	81	2010.08	Y	2010.08	20110208	5.2x10(5)	42	26	2.7	121	1	84.5	1.1	0.6
46	M	53	2006	Y	2010.06	20110215	2.6x10(4)	106	239	3.2	98	2.6	114.1	0.93	0.7
47	M	74	2006	Y	2006.09	20110215	1.8x10(5)	80	64	2.8	74	2.7	76.5	1.16	0.9
48	M	65	2008.11	Y	2008.11	20110217	9.5x10(5)	91	110	3.1	63	1	63.8	1.29	0.6
49	M	68	2009.02	Y	2009.02	20110225	3.5x10(5)	88	65	2.5	97	1.5	70	1.21	1

[0063]

[0064] **2. PCR과 UDPS**

[0065] 가. PCR

[0066] HCV RNA는 QIAamp MinElute Virus Spin Kit (Qiagen Inc., Valencia, CA, USA)를 이용하여 제조사의 지침서에 따라 추출하였다. 추출된 HCV RNA는 cDNA로 만들어진 후 2회에 걸쳐 증폭되었다. 증폭은 HCV의 구조단백질인 코어, E1, E2의 약 2kb가 포함되도록 하였다. 중합효소연쇄반응 (polymerase chain reaction; PCR)을 위하여 사용한 시발체(primer)의 염기순서는 표 2와 같다.

표 2

[0067]

Region	name	sequence (5'→3')

core	1st	B1xL.1-AP1	GGGCGACACTCCACCATAG(서열번호 3)
		B1xR.1-AP1	GCCGGCRAARAGYAGCAYC(서열번호 4)
	2nd	B1xL.1-AP2	ACTCCCCTGTGAGGAACTAY(서열번호 5)
		B1xR.1-AP3	TGGGATCCGGAGYARCTG(서열번호 6)
E1/E2	1st	B1yL.1-AP1	TTCTGYTCCGCYATGTAYG(서열번호 7)
		B1yR.1-AP1	CACAAGRAAGGAGAGRANRC(서열번호 8)
	2nd	B1yL.1-AP3	GAACTGGTCRCCYACARCRG(서열번호 9)
		B1yR.1-AP2	GGACCACCARGTTCTCYARG(서열번호 10)

[0068] 1차 PCR은 HCV cDNA 5μL와 중합효소연쇄반응액 20μL로 시행하였다. 중합효소연쇄반응액으로는 Advantage 2 PCR kit (Clontech, Inc, CA, USA)를 사용하였으며, 여기에는 1 x PCR buffer (40 mM Tricine-KOH (pH 8.7), 15 mM KOAc, 3.5mM $Mg(OAc)_2$, 3.75 ug/mL BSA, Tween 20 0.005%), 25 mM dNTP, 12.5 μM primer, 1 x Advantage® 2 Polymerase Mix가 포함되었다. PCR은 95° C에서 10분 동안 전변성 반응 후, 변성반응 95° C에서 30초, 결합반응 55.8° C에서 1분, 연장반응 72° C에서 1분을 25회 반복하고 72° C에서10분간 추가 연장반응을 시행하여 4° C에서 보관하였다.

[0069] 2차 PCR은 1차 PCR 산물 5μL와 중합효소연쇄반응액 45μL로 시행하였다. 중합효소연쇄반응액의 조성물은 1차 증폭과 같았고, PCR은 95° C에서 10분 동안 전변성 반응 후, 변성반응 95° C에서 30초, 결합반응 55.8° C에서 1분, 연장반응 72° C에서 1분을 35회 반복하고 72° C에서 10분간 추가 연장반응을 시행하여 4° C에서 보관하였다.

[0070] PCR 증폭산물을 AMPure beads (Beckman Coulter Inc., Brea, CA, USA)를 이용하여 정화한 후, 이것으로 library preparation, emulsion PCR, sequencing의 순서로 NGS를 이용한 UDPS를 시행하였다.

[0071] 나. 라이브러리 제조

[0072] 제조사에서 제공하는 GS FLX Titanium General Library Preparation 지침서에 따라 라이브러리를 제조하였다. 이 때 각 검체에 서로 다른 multiplex identifiers (MID)를 붙였다. MID는 서로를 구별할 수 있는 10 bp의 올리고뉴클레오티드로써 검체별로 서로 다른 MID를 붙인 후 한꺼번에 염기서열분석하면 소프트웨어에서 MID를 인식하여 각 검체의 구별이 가능하게 한다. 사용한 MID의 염기서열은 표 3과 같다.

표 3

[0073]

no	MID	no	MID
1	TCTCTATGCG	27	ATACGACGTA
2	CATAGTAGTG	28	TCACGTACTA
3	CGAGAGATAC	29	TCTACGTAGC
4	CGTCTAGTAC	30	CATAGTAGTG
5	ACGAGTGCGT	31	CGAGAGATAC
6	CTCGCGTGTC	32	ATACGACGTA
7	TCTCTATGCG	33	TCACGTACTA
8	TGATACGTCT	34	CGTCTAGTAC
9	CATAGTAGTG	35	TCTACGTAGC
10	CGAGAGATAC	36	TGTACTACTC
11	ATACGACGTA	37	ACGACTACAG
12	TCACGTACTA	38	CGTAGACTAG
13	CGTCTAGTAC	39	TACGAGTATG
14	TCTACGTAGC	40	TCTCTATGCG
15	TGTACTACTC	41	TGATACGTCT
16	CGTAGACTAG	42	CATAGTAGTG
17	TACGAGTATG	43	CGAGAGATAC
18	ACACGACGAC	44	ATACGACGTA
19	ACACGTAGTA	45	TCACGTACTA
20	CTCGCGTGTC	46	CGTCTAGTAC
21	TGATACGTCT	47	TCTACGTAGC

22	TGTACTACTC	48	ACGAGTGCGT
23	ACGACTACAG	49	ACGCTCGACA
24	CGTAGACTAG		
25	TACGAGTATG		
26	ACACTACTCG		

[0074] DNA 조각에 adaptor를 붙이기 위한 과정은 다음과 같았다. 10 x polishing buffer 5㎕, BSA 5㎕, ATP 5㎕, dNTP Mix 2㎕, T4 PNK 5㎕, T4 DNA 폴리머라아제의 mix를 준비하고 여기에 PCR 증폭산물 1㎍을 섞어 12ºC에서 15분간 반응시키고 이어서 25ºC에서 15분간 반응시켰다. 이것을 하나의 MinElute column (Quiagen)으로 정화하였다. MID adaptor를 붙이기 위해 상기의 DNA 10㎕에 2 x Ligase Buffer 20㎕, MID Adaptors 5㎕, Ligase 5㎕의 mix를 섞어 25ºC에서 15분간 반응시켰다. 이것을 MinElute column (Qiagen Inc.)으로 정화하였다.

[0075] Strepavidin이 코팅된 비드에 library를 붙이기 위한 과정은 다음과 같았다. 2x Library Binding Buffer 100㎕로 Library Immobilization Bead 50㎕를 두 번 씻었다. 2x Library Binding Buffer 25㎕에 Library Immobilization Bead 50㎕를 풀고 여기에 MID adaptor를 붙인 DNA 25㎕를 첨가하였다. 실온에서 20분 동안 튜브 교반기에 놓아두어 MID adaptor를 붙인 DNA library가 비드에 붙게 하였다. Library wash buffer로 씻고 펠릿이 생긴 후 library wash buffer를 제거하였다. Molecular Biology Grade (MBG) water 40㎕, 10 x Fill-In Polyerase Buffer 5㎕, dNTP Mix 2㎕, Fill-In Polymerase 3㎕의 mix를 library가 붙은 비드 펠릿에 첨가하여 37° C에서 20분간 반응시켰다. Library wash buffer 100㎕로 두 번 씻고 펠릿이 생긴 후 library wash buffer를 제거하였다.

[0076] 단일 가닥 DNA library를 골라내기 위한 과정은 다음과 같았다. PBI 500㎕와 3M sodium Acetate pH 5.2 10㎕를 섞어 Neutralization Solution을 준비하고, 9.875 mL의 MBG Water와 10N NaOH 0.125 mL를 섞어 Melt Solution을 준비하였다. 씻은 library 비드에 Melt Solution 50㎕를 첨가한 후 보텍스하고 비드로 펠릿을 만들어 단일 가닥 DNA library가 들어있는 상청액을 Neutralization Solution에 옮겼다. 이것을 하나의 MinElute column (Qiagen Inc.)으로 정화하였다. RiboGreen (Invitrogen, Eugene, OR, USA) 방법으로 DNA library를 정량하고, library의 농도를 계산한 후, library의 농도를 2 x 10^6 molecules/㎕로 희석하였다. 각 검체에서 20㎕의 부피를 취해 pooling하여 이후의 emulsion PCR과 sequencing 단계를 시행하였다.

[0077] 다. 에멀전 PCR

[0078] 제조사에서 제공하는 GS FLX Titanium emPCR 지침서에 따라 에멀전 PCR을 하여 단일 가닥 template DNA library를 클론성으로 증폭하였다. 이 때 SV emPCR kit를 사용하였다. MBG Wager 640㎕, 5x Amplification Mix 190㎕, Amplification Primer 55㎕, emPCR Enzyme Mix 50㎕, PPiase 2㎕를 이용하여 Live and Mock Amplification Mixes를 준비하여 5초 동안 보텍스하고, 다른 튜브에 MBG water 4 mL에 5x Mock Amp Mix 1 mL를 희석하여 보텍스하였다.

[0079] DNA library를 포집 비드에 붙이기 위한 과정은 다음과 같았다. DNA 포집 비드 한 튜브를 보텍스하였다. 이 때 하나의 에멀전 튜브를 비드 부유액 80㎕씩 4개의 튜브에 분주하였다. 10초 동안 spin한 후 상청액을 버리고 1x 포집 비드 Wash Buffer TW 1 mL로 두 번 씻었다. 포집 비드 튜브에 DNA library를 넣고 5초 동안 보텍스하였다. 포집 비드 튜브에 넣는 DNA library의 양을 계산하여, 비드 당 붙게 하고자 하는 분자 수는 첫 검사에서는 0.5로, 두 번째 검사에서는 0.8로 하였다. 이렇게 DNA library가 붙은 DNA 포집 비드에 live Amp Mix 215㎕를 첨가하고 다음의 에멀전화(Emulsification) 과정을 수행하였다.

[0080] Emulsification은 다음과 같이 하였다. TissueLyser라는 기계에서 25Hz 2분 동안 에멀전 오일을 섞고, 1x Mock Amp Mix 290㎕를 넣고 2-3번 뒤집어 섞은 후, TissueLyser에서 25Hz 5분 동안 섞었다. 여기에 준비된 captured library 255㎕를 첨가하여 2-3번 뒤집어 섞은 후, TissueLyser에서 12Hz 5분 동안 섞었다. Emulsification을 통해 각 비드를 오일 상 중 물방울 안에 격리시켰고 이 안에서 증폭 반응이 일어날 수 있게 하였다.

[0081] 증폭은 다음과 같이 하였다. emPCR amplification mix를 96-well thermocycler plate중 10개의 well에 well 당 100㎕씩 분주하였다. Plate를 밀봉하여 thermocycler에 넣고 PCR을 수행하였다. PCR은 94ºC에서 4분 동안 전

변성 후 94ºC에서 30초, 58ºC에서 4.5분, 68ºC에서 30초를 50회 반복하고, 10ºC에서 보관하였다.

[0082] 에멀전 PCR 반응 후 에멀전 내에 있는 비드를 다음과 같은 과정으로 모았다. 10 mL 주사기의 끝에 16 게이지 바늘을 직접 연결하여 각 well에서 emulsion을 뽑아낸 후 각 well에 이소프로판올 100㎕을 첨가하여 같은 주사기로 이소프로판올 행굼액을 뽑아내었다. 주사기의 바늘을 제거하고 blue filter를 연결하여 주사기를 5초 동안 보텍스한 후, 주사기의 내용물을 filter를 통해 waste jar에 버렸다. 비드를 이소프로판올로 3번 씻은 후 1x Enhancing Fluid TW로 3번 씻었다. 다음의 과정을 두 번 수행하여 filter에서 비드를 모았다. 주사기에 공기 3 mL를 빨아들이고 1x Enhancing Fluid TW를 주사기의 0.5 mL까지 당기어 5초 동안 보텍스하였다. 즉시 주사기를 뒤집어 공기 3 mL를 빨아들여 filter에서 재부유된 비드가 떨어져 나오게 하였다. 주사기로부터 filter를 제거하고 1.7 mL 튜브에 비드 부유액을 담았다.

[0083] 이렇게 비드를 모은 후 양쪽에 adaptor가 제대로 붙어서 클론성으로 증폭된 비드를 분리하였는데 그 방법은 다음과 같았다. 비드를 10초 동안 spin하여 상청액을 버렸다. 비드 튜브 당 Melt Solution 1 mL을 첨가한 후 보텍스하고 2분 동안 둔 후 비드를 펠릿하고 상청액을 버림으로써 dsDNA로 증폭된 산물을 ssDNA 비드로 만들었다. 비드 펠릿에 1x Annealing Buffer TW 500㎕를 첨가한 후 원심분리하는 방법으로 두 번 씻어 준 후, 각 튜브에 1x annealing Buffer TW 30㎕, Enrichment Primer 6㎕를 첨가하였다. 이것을 65ºC 발열블록에 5분간 둔 후, 즉시 얼음에 옮겨 2분 동안 두었다. 1x Enhancing Fluid TW 500㎕를 첨가하고 보텍스한 후, 원심분리를 하여 비드를 펠릿으로 만들고 상청액을 버렸다. 1x Enhancing Fluid TW 500㎕로 원심분리하여 비드 펠릿을 두 번 씻어주고, 상청액을 제거한 후에 1x Enhancing Fluid TW 800㎕를 첨가하여 비드 펠릿을 풀어주고 보텍스하였다.

[0084] 제대로 증폭된 비드와 결합하는 비드인 Enrichment Bead를 1분 동안 보텍스하여 잘 풀은 후, Magnetic Particle Concentrator (MPC)에서 펠릿과 상청액을 제거하였다. MPC를 이용하여 1x Enhancing Fluid TW 500㎕로 펠릿 상태의 비드를 두 번 씻고 상청액을 제거한 후에 MPC에서 튜브를 분리하여 1x Enhancing Fluid TW 320㎕를 첨가하고 보텍스하였다. 씻은 Enrichment Bead 20㎕를 증폭된 DNA 비드가 있는 각 튜브에 첨가한 후 보텍스하고, 실온에서 5분 동안 rotation 하였다. 튜브를 MPC에 놓고 펠릿이 될 때까지 약 5분 기다린 후 튜브의 뚜껑을 닫아 마그넷을 여러 번 뒤집어 상청액을 탁해지게 하고, 상청액을 제거하였다. 상청액에 보이는 비드가 없을 때까지 1x Enhancing Fluid TW 500㎕로 비드를 씻어주었다. MPC에서 튜브를 분리하여 비드 펠릿을 Melt Solution 700㎕로 풀은 후, 5초 동안 보텍스하고 Enrichment Bead가 펠릿이 되도록 튜브를 다시 MPC에 두었다. Enriched DNA 비드가 포함된 상청액을 튜브에서 다른 1.7 mL microfuge 튜브로 옮겼다. 비드를 10초 동안 spin하여 상청액을 버리고, Melt Solution을 중화시키기 위해 비드 펠릿을 1x Annealing Buffer TW 500㎕로 세 번 씻은 후, 비드 펠릿을 1x Annealing Buffer TW 200㎕로 풀었다.

[0085] Sequencing Primer는 다음과 같이 붙였다. Sequencing Primer를 튜브 당 6㎕를 넣은 후 보텍스하였다. 튜브를 65ºC 발열블록에 5분간 둔 후, 즉시 얼음에 옮겨 2분 동안 두었다. 1x Annealing Buffer TW 500㎕를 첨가하고, 비드를 펠릿으로 만든 후 상청액을 제거하였다. 비드 펠릿을 1x Annealing Buffer TW 500㎕로 두 번 씻은 후, 비드 펠릿에 1x Annealing Buffer TW 100 ㎕를 첨가하여 풀었다. [% Bead Enrichment = enriched bead의 수 / 총 들어간 비드 x 100]의 공식에 따라 % Bead Enrichment를 구하였다. 첫 번째 검사에서는 1.00, 두 번째 검사에서는 2.00의 값이 나왔다. 비드 3㎕를 덜어 Multisizer 3 Coulter Counter (Beckman Coulter Inc.)로 비드를 세었다.

[0086] 라. Pyrosequencing

[0087] 제조사에서 제공하는 GS FLX Titanium Sequencing 지침서에 따라 GS FLX 454 Genome Sequencer를 사용하여 pyrosequencing 방법으로 염기서열분석을 하였다. 우선 70 mm x 75 mm Picotiter Plate (PTP)를 준비하였다. 미리 차갑게 한 Titanium Bead Buffer 200 mL에 Titanium Supplement CB 1.2 mL와 Apyrase 34㎕를 차례로 첨가하여 Bead buffer 2 (BB2)를 준비하였다. PTP ID 번호를 기록하고, PTP device를 BB2에 담그고 실온에서 10분 이상 두었다. Bead Loading Gasket과 Dead Deposition Device (BDD)를 Sparkleen 용액으로 씻었다.

[0088] BB2 1 mL로 Packing Bead를 씻고 9300g x 5분으로 원심분리하는 과정을 세 번 한 후 BB2 550㎕로 풀어주고 얼음에 둠으로써 의 Packing Bead를 준비하였다. PicoTiter Plate의 1/4 구역을 쓰므로 DNA library 비드는 790,000개가 필요하였는데, 그에 따른 DNA library 비드 부유액 부피는 790,000을 enrichment bead 수로 나누어 계산하였다. enrichment bead 수는 첫 번째 검사에서 100개와 두 번째 검사에서 200개였고, 따라서 계산된 DNA library 비드 부유액 부피는 첫 번째 검사에서 100㎕와 두 번째 검사에서 200㎕였다. 1.7 mL 튜브에 그만

큼을 넣은 후 Control DNA Bead 10㎕를 첨가하여 1분간 spin하였다. Target volume은 30㎕이므로 그것을 초과하는 만큼의 상청액을 제거하였다. BB2 1570㎕, Polymerase Cofactor 150㎕, DNA Pol 300㎕로 DNA Bead Incubation Mix (DBIM)을 준비하여 320㎕를 DNA 비드 튜브에 넣은 후, 교반기에서 실온 15분 동안 반응시켰다. Enzyme과 PPiase Bead는 다음과 같이 준비하였다. Enzyme과 PPiase Bead를 보텍스하고, MPC에서 펠릿으로 만들어 상청액을 제거한 다음 BB2 1 mL로 3번 씻었다. Enzyme에는 BB2 1000㎕를 첨가하고 PPiase Bead에는 500㎕를 첨가한 후 보텍스하였다. BB2 3250㎕와 Enzyme Bead 550㎕를 섞어 Bead Layer 1을 만들고, BB2 2500㎕와 Enzyme Bead 1300㎕를 섞어 Bead Layer 3을 만들고, BB2 3340㎕와 PPiase Bead 460㎕를 섞어 Bead Layer 4를 만들어 보텍스하였다. Bead Layer 2는 DNA bead가 있는 튜브에 Packing Bead 100㎕, BB2 210㎕를 넣어 보텍스한 후, 교반기에 실온 5분 반응시킴으로써 준비하였다.

[0089] 트레이에 담겨 있는 PTP Device를 제거하고 남은 BB2를 버린 후 BDD에 device와 Bead Loading Gasket을 끼웠다. 그 후 PTP Device에 Bead Layer를 다음과 같은 순서로 loading하였다. Bead Layer 1이 들은 튜브를 5초 동안 보텍스하고 660㎕를 load한 후, PTP device를 1620g x 5분 spin하였다. Layer 1의 상청액을 가능한 많이 제거하고, 교반기에서 Bead Layer 2를 꺼내 660㎕를 load 한 후, PTP device를 1620g x 10분 spin하였다. Layer 2의 상청액을 제거하고, Bead Layer 3이 들은 튜브를 보텍스하고 660㎕를 load한 후, PTP device를 1620g x 5분 spin하였다. Layer 3의 상청액을 가능한 많이 제거하고, Bead Layer 4가 들은 튜브를 보텍스하고 660㎕를 load한 후, PTP device를 1620g x 5분 spin하였다.

[0090] 키트에 들어있는 티타늄 버퍼 CV 4개의 병에 6.6 mL의 티타늄 상청액 CB와 1000㎕의 DTT를 차례로 첨가한 후, Reagent Cassette에 티타늄 버퍼 CB와 Sequencing Reagent tray를 놓았다. Inhibitor TW reagent 45㎕에 5㎕의 PPiase reagent를 희석하였다. Tween 튜브에 Sodium Chlorite 1 tablet을 첨가하고, “inhibitor TW” 튜브에 희석한 PPiase 13.2㎕를 첨가하고, “Buffer for Apyrase” 튜브에 Apyrase 260㎕를 첨가하고 “buffer for dATP” 튜브에 dATP 3000㎕를 첨가하여 Sequencing Reagent tray를 준비하였다. 이것을 20번 뒤집어 섞고 용해되지 않은 입자들이 있는지 확인하였다. 모든 시약병과 튜브의 뚜껑을 열고 기계에 Reagent cassette를 장착한 후, Sipper manifold를 아래로 내리고 외부의 fluidics door를 닫았다. Cartridge seal을 설치하고 BDD에서 PTP device를 꺼내 PTP frame 안으로 밀어 넣어 PTP frame을 닫았다. Run Wizard 창에서 Start 버튼을 눌러 sequencing run을 시작하였다.

[0091] Roche GS FLX를 사용하여 UDPS를 시행한 과정을 도 1에 요약하였다.

[0092] **3. UDPS 염기서열 정보의 분석**

[0093] UDPS를 시행하여 얻어진 각 read를 정렬하기 위하여 HCV 기준주를 구성하였다. 기준주의 염기서열은 Los Alamos National Laboratory 사이트(http://hcv.lanl.gov)에 있는 HCV 염기서열 정보 중 HCV 전체 게놈의 염기서열이 밝혀진 유전자형 1b의 염기서열 정보 388개를 정렬하여 만들었다. 특정 염기서열 위치에서 동일한 염기가 70% 이상이면 그 염기를 기준주의 염기로 하였고, 그렇지 않은 경우에는 기준주의 염기를 “N”으로 하였다. 마찬가지로 특정 염기서열 위치에서 70% 이상의 염기가 blank이면 기준주의 염기서열에서는 해당 위치를 삭제하였다. HCV 기준주의 염기서열 중 단백질을 만드는 부분의 염기서열을 아미노산 서열로 치환하여 기준주의 아미노산 서열을 만들었다. Los Alamos National Laboratory 사이트에서 다운받은 388개의 염기서열로부터 구성된 HCV 기준주의 염기서열을 서열목록 2에, HCV 기준주의 아미노산 서열을 서열목록 1에 첨부하였다.

[0094] Roche에서 제공하는 소프트웨어인 Amplicon Variant Analyzer (AVA)를 이용하여 각 read를 HCV 기준주에 정렬하였다. HCV의 변이형(quasispecies)의 다양성을 조사하기 위해 HCV 염기서열을 분석하는 전용 프로그램을 개발하였다.

[0095] 전용 프로그램으로는 먼저 각 환자마다 HCV 변이형의 양상을 분석하기 위해 각 검체의 HCV 염기서열의 각 위치에서 Shannon diversity index를 구하였다. Shannon diversity index를 구하는 식은 다음과 같다.

[0096] [식 1]

[0097] $$H' = -\sum_{i-1}^{S} p_i \ln p_i$$

[0098] S: “A”, “C”, “G”, “T”의 네 가지

[0099] Pi: 해당 위치에서의 i번째 염기가 차지하는 비율

[0100] Pi = i번째 염기의 수 / 전체 염기의 수

[0101] 간경변증이 없는 만성 C형 간염 환자군과 HCV로 인한 간세포암이 발생한 환자군에서 HCV 변이형의 다양성에 차이가 있는지 확인하기 위해 Shannon diversity index의 HCV 염기서열의 각 위치에서 T 검정을 시행하였다. 군별 검체의 수가 다르고 분산이 같은 경우이므로 T value와 자유도 d.f 는 다음과 같이 계산된다.

[0102] $$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{X_1X_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

[0103] $$S_{X_1X_2} = \sqrt{\frac{(n_1 - 1)S_{X_1}^2 + (n_2 - 1)S_{X_2}^2}{n_1 + n_2 - 2}}.$$

[0104] d.f = $n_1 + n_2 - 2$

[0105] X_1: 간경변증이 없는 만성 C형 간염 환자들의 Shannon diversity index의 평균

[0106] X_2: HCV로 인한 간세포암이 발생한 환자들의 Shannon diversity index의 평균

[0107] n_1: 간경변증이 없는 만성 C형 간염 환자들의 수

[0108] n_2: HCV로 인한 간세포암이 발생한 환자들의 수

[0109] S_{x1}^2: 간경변증이 없는 만성 C형 간염 환자들의 Shannon diversity index의 분산

[0110] S_{x2}^2: HCV로 인한 간세포암이 발생한 환자들의 Shannon diversity index의 분산

[0111] S_{x1x2}: 두 군의 공분산 추정치의 제곱근

[0112] HCV 염기서열을 분석하는 것과 마찬가지의 방법으로 HCV 아미노산 서열을 분석하였다. 즉, HCV 염기서열을 아미노산 서열로 변환한 후 각 환자마다 HCV 변이형의 양상을 분석하기 위해 각 검체에서 HCV 아미노산 서열의 각 위치에서 Shannon diversity index를 구하였다. 이 때, Shannon diversity index를 구하는 식에서 S는 20가지의 아미노산 종류가 된다. 그 후 간경변증이 없는 만성 C형 간염 환자군과 HCV로 인한 간세포암이 발생한 환자군에서 차이가 있는지 확인하기 위해 HCV 아미노산 서열의 각 위치에서 T 검정을 시행하였다.

[0113] HCV로 인한 간세포암이 발생한 환자들의 결과가 간경화로 인한 것인지 여부를 판단하기 위해 6명의 간경변증이 있는 만성 C형 간염 환자들의 HCV 변이형의 다양성을 간경변증이 없는 만성 C형 간염 환자군과 HCV로 인한 간세포암이 발생한 환자군과 비교하였다. 앞서와 마찬가지로 HCV 염기서열의 각 위치에서 T 검정을 시행하였다.

[0114] T검정 결과 통계적으로 유의한 다양성의 차이가 있는지 조사하기 위해 bonferroni correction 을 적용하였다. 염기서열 다양성 조사에서는 조사 대상 염기서열의 길이인 2238을 적용하였고 아미노산 서열 다양성 조사에서는 조사 대상 아미노산 서열의 길이인 746을 적용하였다. 적용 후 p value 가 0.05보다 작은 경우를 통계적으로 유의한 것으로 간주하였다.

[0115] 통계적으로 유의한 다양성의 차이를 보이는 아미노산 서열 상의 위치에 대해서 각 위치마다 ROC curve를 구하였다. 각 ROC curve에서 90% 이상의 민감도를 보이는 위치에서 특이도를 구하고 AUC를 계산하였다.

[0116] **<실시예 1> PCR과 UDPS 결과 분석**

[0117] 49개 검체에서 UDPS로 총 1,822,601개의 read 결과가 나왔고, 한 검체 당 read의 개수는 평균 37,196 (4,996-94,180)개였다. AVA 소프트웨어를 사용하여 기준주에 맞추어 정렬하였을 때 정렬되는 read는 한 검체 당 평균 35,311개로 전체 read의 평균 94%였다. 각 검체에서 UDPS로 나온 read의 개수와 기준주에 정렬되는read 의 개수를 도 2에 표시하였다. 각 검체별 평균 read의 길이를 도 3에 표시하였다. 게놈상의 각 위치별 read의 coverage를 도 4에 표시하였다.

[0118] **<실시예 2> 간경변증이 없는 만성 C형 간염 환자들과 HCV로 인한 간세포암이 발생한 환자들에서 HCV 변이형(quasispecies)의 염기서열 다양성 검토**

[0119] 간경변증이 없는 만성 C형 간염 환자들에서 HCV 변이형의 양상을 분석하기 위해 전용 프로그램으로 각 검체에서 HCV 염기서열의 각 위치에서 구한 Shannon diversity index를 도 5에 나타내었다. Shannon diversity index 값이 클수록 해당 HCV 염기 위치에서 변이형의 다양성이 높은 것이다.

[0120] 마찬가지로 HCV로 인한 간세포암이 발생한 환자들에서 전용 프로그램으로 각 검체에서 HCV 염기서열의 각 위치에서 구한 Shannon diversity index를 도 6에 나타내었다.

[0121] **<실시예 3> 간경변증이 없는 만성 C형 간염 환자들과 HCV로 인한 간세포암이 발생한 환자들의 HCV 변이형(quasispecies)의 염기서열 다양성 비교**

[0122] 간경변증이 없는 만성 C형 간염 환자군과 HCV로 인한 간세포암이 발생한 환자군에서 HCV 변이형의 다양성에 차이가 있는지 확인하기 위해 HCV 염기서열의 각 위치에서 T 검정을 시행한 결과를 도 7에 표시하였다. T 검정의 p value 값이 E1 부위에서 유의하게 낮아 간경변증이 없는 만성 C형 간염 환자군에 비하여 HCV로 인한 간세포암이 발생한 환자군에서 E1 부위의 HCV 염기서열의 다양성이 유의적으로 감소한 것으로 확인되었다.

[0123] **<실시예 4> 간경변증이 없는 만성 C형 간염 환자들과 HCV로 인한 간세포암이 발생한 환자들에서 HCV 변이형(quasispecies)의 아미노산 서열 다양성 검토**

[0124] 간경변증이 없는 만성 C형 간염 환자들에서 HCV 변이형의 아미노산 서열의 양상을 분석하기 위해 전용 프로그램으로 각 검체에서 HCV 아미노산 서열의 각 위치에서 구한 Shannon diversity index를 도 8에 나타내었다.

[0125] 마찬가지로 HCV로 인한 간세포암이 발생한 환자들에서 전용 프로그램으로 각 검체에서 HCV 아미노산 서열의 각 위치에서 구한 Shannon diversity index를 도 9에 나타내었다.

[0126] **<실시예 5> 간경변증이 없는 만성 C형 간염 환자들과 HCV로 인한 간세포암이 발생 환자들의 HCV 변이형(quasispecies)의 아미노산 서열 다양성 비교**

[0127] 간경변증이 없는 만성 C형 간염 환자군과 HCV로 인한 간세포암이 발생한 환자군에서 HCV quasispecies 아미노산 서열의 다양성에 차이가 있는지 확인하기 위해 HCV 아미노산 서열의 각 위치에서 T 검정을 시행한 결과를 도 10에 표시하였다. 염기서열의 다양성 차이를 비교했을 때와 마찬가지로 T 검정의 p value 값이 E1 부위에서 유의하게 낮아 간경변증이 없는 만성 C형 간염 환자군에 비하여 HCV로 인한 간세포암이 발생한 환자군에서 HCV 아미노산 서열의 다양성이 유의하게 감소한 것을 알 수 있었다.

[0128] **<실시예 6> 간경변증이 없는 만성 C형 간염 환자들과 HCV로 인한 간세포암이 발생 환자들 사이에 통계적으로 유의한 다양성의 차이를 보이는 염기서열 상의 위치와 아미노산 서열 상의 위치 분석**

[0129] 간경변증이 없는 만성 C형 간염 환자들과 HCV로 인한 간세포암이 발생 환자들 사이에 염기 서열 다양성 차이가 존재하는지 T 검정을 실행하였다. 각 위치에서의 T검정 결과에 대해 유의수준 0.05에서 통계적으로 유의한 염기서열상의 위치와 그 결과를 표 4 에 표시하였다. Bonferroni correction 후 모두 11개의 위치에서 유의한 결과

를 얻었다. 각 위치에서 간경변증이 없는 만성 C형 간염으로부터 HCV로 인해 발생한 간세포암을 구별하는 receiver operating characteristic (ROC) curve를 도 11에 표시하였다.

표 4

[0130]

N pos	CHC without LC		HCC		t value	p value	
	average	var	average	var		raw	corrected
1320	0.140	0.0035	0.048	0.0006	6.89	2.37E-08	5.30.E-05
1324	0.121	0.0016	0.056	0.0012	5.77	9.32E-07	2.09.E-03
1327	0.127	0.0022	0.060	0.0012	5.37	3.41E-06	7.62.E-03
1205	0.463	0.0286	0.186	0.0334	5.13	7.36E-06	1.65.E-02
1381	0.318	0.0276	0.121	0.0081	4.93	1.40E-05	3.14.E-02
1211	0.518	0.0569	0.205	0.0323	4.89	1.58E-05	3.54.E-02
1392	0.331	0.0397	0.114	0.0052	4.87	1.73E-05	3.86.E-02
1344	0.098	0.0019	0.044	0.0008	4.85	1.82E-05	4.06.E-02
1200	0.471	0.0390	0.176	0.0401	4.85	1.83E-05	4.10.E-02
1188	0.479	0.0549	0.174	0.0317	4.84	1.87E-05	4.19.E-02
1388	0.272	0.0284	0.084	0.0060	4.82	2.01E-05	4.50.E-02

[0131] 같은 방법으로 간경변증이 없는 만성 C형 간염 환자들과 HCV로 인한 간세포암이 발생 환자들 사이에 아미노산 서열 다양성 차이가 존재하는지 T 검정을 실행하였다. 각 위치에서의 T검정 결과에 대해 유의수준 0.05에서 통계적으로 유의한 아미노산 서열상의 위치와 그 결과를 표 5 에 표시하였다. Bonferroni correction 후 모두 11개의 위치에서 유의한 결과를 얻었다. 각 위치에서 간경변증이 없는 만성 C형 간염으로부터 HCV로 인해 발생한 간세포암을 구별하는 receiver operating characteristic (ROC) curve를 도 12에 표시하였다.

표 5

[0132]

AA pos	CHC without LC		HCC		t value	p value	
	average	var	average	var		raw	corrected
332	0.377	0.0361	0.116	0.0120	5.61	1.54E-06	1.15E-03
331	0.178	0.0050	0.081	0.0020	5.42	2.91E-06	2.17E-03
355	0.470	0.0732	0.178	0.0154	4.64	3.50E-05	2.61E-02
357	0.423	0.0442	0.194	0.0113	4.58	4.22E-05	3.15E-02
290	0.615	0.0679	0.225	0.0862	4.57	4.46E-05	3.33E-02
339	0.155	0.0043	0.081	0.0015	4.55	4.71E-05	3.51E-02
350	0.473	0.0789	0.174	0.0181	4.54	4.78E-05	3.57E-02
287	0.678	0.0967	0.264	0.0821	4.54	4.81E-05	3.59E-02
354	0.428	0.0700	0.152	0.0137	4.53	4.96E-05	3.70E-02
349	0.651	0.1609	0.248	0.0188	4.53	5.04E-05	3.76E-02
340	0.210	0.0050	0.130	0.0020	4.51	5.36E-05	4.00E-02

[0133] **<실시예 7> 간경변증이 있는 만성 C형 간염 환자들의 HCV 변이형(quasispecies)의 염기서열 다양성과 아미노산 서열 다양성 검토**

[0134] HCV로 인한 간세포암이 발생한 환자들의 결과가 간경화로 인한 것인지 여부를 판단하기 위해 6명의 간경변증이 있는 만성 C형 간염 환자들의 HCV 변이형의 양상을 분석하였다. 전용 프로그램으로 각 검체에서 HCV 염기서열의 각 위치에서 구한 Shannon diversity index를 도 13에 표시하였다. 또한 6명의 간경변증이 있는 만성 C형 간염 환자들의 HCV 변이형의 아미노산 서열의 양상을 분석하기 위해 전용 프로그램으로 각 검체에서 HCV 아미노산 서열의 각 위치에서 구한 Shannon diversity index를 도 14에 표시하였다.

[0135] **<실시예 8> 간경변증이 있는 만성 C형 간염 환자들의 HCV 변이형(quasispecies)의 염기서열 다양성과 아미노산 서열 다양성을 간경변증이 없는 만성 C형 간염 환자들과 HCV로 인한 간세포암이 발생 환자들과 비교**

[0136] 간경변증이 있는 만성 C형 간염 환자군의 HCV 변이형의 염기서열 다양성이 간경변증이 없는 만성 C형 간염 환자군과 HCV로 인한 간세포암이 발생한 환자군과 비교하여 차이가 있는지 확인하였다. 간경변증이 있는 만성 C형

간염 환자군과 간경변증이 없는 만성 C형 간염 환자군에 대해 HCV 염기서열상의 각 위치에서 T 검정을 시행하고 그 결과를 도 15에 표시하였다. 간경변증이 있는 만성 C형 간염 환자군과 HCV로 인한 간세포암이 발생한 환자군에 대해 HCV 염기서열상의 각 위치에서 T 검정을 시행하고 그 결과를 도 16에 표시하였다.

[0137] 마찬가지로 간경변증이 있는 만성 C형 간염 환자군의 HCV 변이형의 아미노산 서열 다양성이 다른 두 군과 비교하여 차이가 있는지 확인하였다. 간경변증이 있는 만성 C형 간염 환자군과 간경변증이 없는 만성 C형 간염 환자군에 대해 HCV 아미노산 서열상의 각 위치에서 T 검정을 시행하고 그 결과를 도 17에 표시하였다. 간경변증이 있는 만성 C형 간염 환자군과 HCV로 인한 간세포암이 발생한 환자군에 대해 HCV 아미노산 서열상의 각 위치에서 T 검정을 시행하고 그 결과를 도 18에 표시하였다.

[0138] **<실시예 9> 만성 C형 간염 환자들의 HCV 변이형(quasispecies)의 염기서열 다양성과 아미노산 서열다양성을 HCV로 인한 간세포암 발생 환자들과 비교**

[0139] 간경변증의 유무에 관계 없이 만성 C형 간염 환자들과 HCV로 인한 간세포암이 발생한 환자들 사이에 염기 서열 다양성에 차이가 존재하는지 T 검정을 시행하여 두 군을 비교하였다. 그 결과를 도 19에 표시하였다. 염기 서열상의 각 위치에서의 T검정 결과에 대해 유의수준 0.05에서 통계적으로 유의한 염기서열상의 위치와 그 결과를 표 6 에 표시하였다. Bonferroni correction 후 모두 8개의 위치에서 유의한 결과를 얻었다.

표 6

[0140]

N pos	CHC		HCC		t value	p value	
	average	var	average	var		raw	corrected
1320	0.131	0.0034	0.048	0.0006	6.40	6.75E-08	1.51E-04
1324	0.117	0.0017	0.056	0.0012	5.57	1.18E-06	2.64E-03
1344	0.098	0.0017	0.044	0.0008	5.25	3.65E-06	8.17E-03
1327	0.121	0.0020	0.060	0.0012	5.22	3.94E-06	8.82E-03
1205	0.477	0.0448	0.186	0.0334	5.12	5.57E-06	1.25E-02
1188	0.494	0.0625	0.174	0.0317	5.10	6.04E-06	1.35E-02
1200	0.504	0.0602	0.176	0.0401	5.08	6.36E-06	1.42E-02
1213	0.392	0.0355	0.135	0.0277	5.02	7.81E-06	1.75E-02

[0141] 마찬가지로 간경변증의 유무에 관계 없이 만성 C형 간염 환자들과 HCV로 인한 간세포암이 발생 환자들 사이에 아미노산 서열 다양성에 차이가 존재하는지 T 검정을 시행하여 두 군을 비교하였다. 그 결과를 도 20에 표시하였다. 아미노산 서열상의 각 위치에서의 T검정 결과에 대해 유의수준 0.05에서 통계적으로 유의한 염기서열상의 위치와 그 결과를 표 7 에 표시하였다. Bonferroni correction 후 모두 14개의 위치에서 유의한 결과를 얻었다.

표 7

[0142]

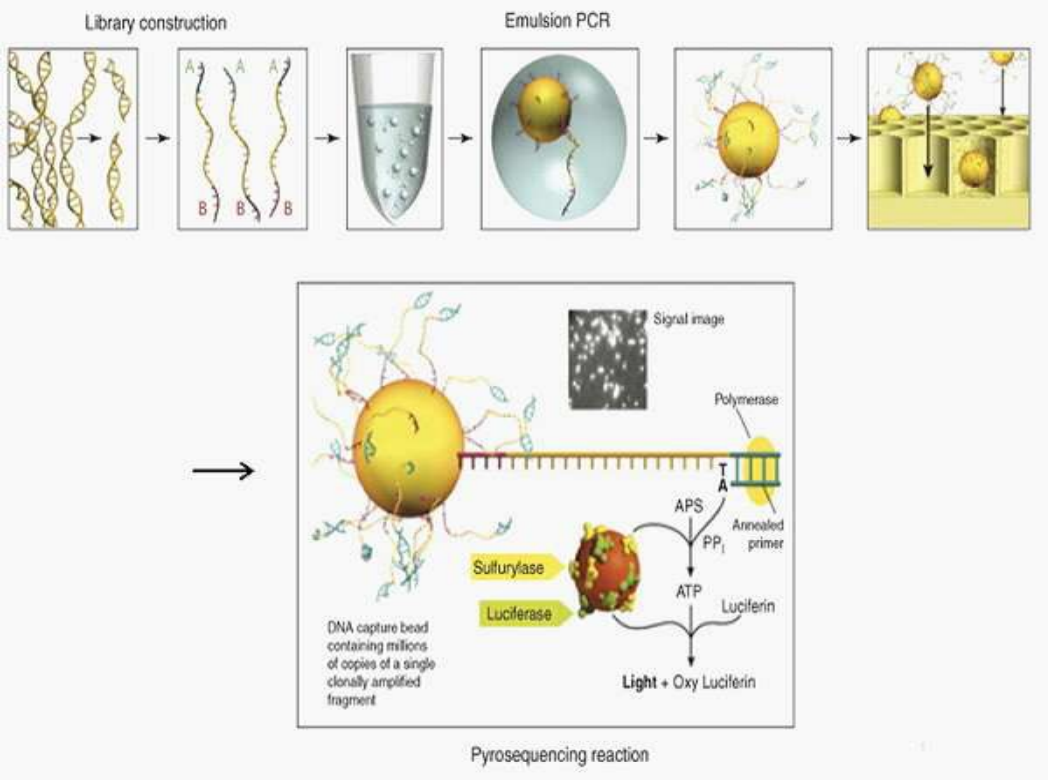
N pos	CHC		HCC		t value	p value	
	average	var	average	var		raw	corrected
331	0.170	0.004	0.081	0.0020	5.381	2.30E-06	1.71E-03
339	0.157	0.004	0.081	0.0015	5.088	6.25E-06	4.66E-03
287	0.726	0.141	0.264	0.0821	4.795	1.68E-05	1.25E-02
290	0.676	0.127	0.225	0.0862	4.793	1.69E-05	1.26E-02
332	0.334	0.038	0.116	0.0120	4.748	1.96E-05	1.47E-02
338	0.147	0.004	0.076	0.0016	4.673	2.51E-05	1.88E-02
291	0.706	0.142	0.254	0.0851	4.649	2.72E-05	2.03E-02
285	0.734	0.172	0.251	0.0858	4.645	2.76E-05	2.06E-02
292	0.605	0.098	0.224	0.0677	4.591	3.30E-05	2.46E-02
293	0.548	0.082	0.220	0.0470	4.479	4.77E-05	3.56E-02
283	0.716	0.175	0.247	0.0891	4.456	5.14E-05	3.84E-02
282	0.669	0.162	0.220	0.0851	4.418	5.82E-05	4.34E-02
341	0.132	0.004	0.068	0.0013	4.417	5.84E-05	4.36E-02
288	0.656	0.142	0.234	0.0792	4.399	6.21E-05	4.63E-02

[0143] 본 발명에 따르면, NGS 기술을 이용하여 수천 개의 클론을 분석함으로써 간세포암이 발생한 환자들에서는 만성 C형 간염 환자들에 비해 HCV E1 부위의 다양성이 유의하게 감소해 있음을 확인할 수 있었고, 보다 구체적으로 두 군을 구별하는 ROC curve 분석에서 AUC가 0.8 이상인 아미노산 부위를 14개 확인할 수 있었다. 상기 아미노산 부위의 분석을 통해 만성 C형 간염 환자와 간세포암 환자를 높은 정확도로 구별할 수 있고, 이에 따라 간세포암의 발병 여부를 간편하게 진단할 수 있다.

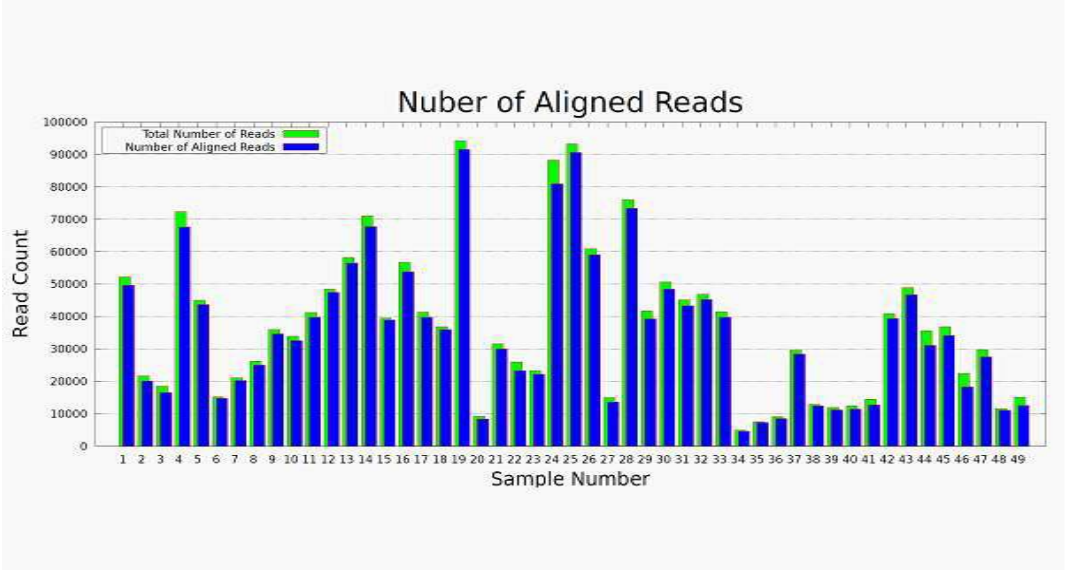
[0144] 이상으로 본 발명의 특정한 부분을 상세히 기술하였는 바, 당업계의 통상의 지식을 가진 자에게 있어서, 이러한 구체적 기술은 단지 바람직한 실시예일 뿐이며, 이에 의해 본 발명의 범위가 제한되는 것이 아닌 점은 명백할 것이다. 따라서, 본 발명의 실질적인 범위는 첨부된 청구항들과 그것들의 등가물에 의하여 정의된다고 할 것이다.

도면

도면1



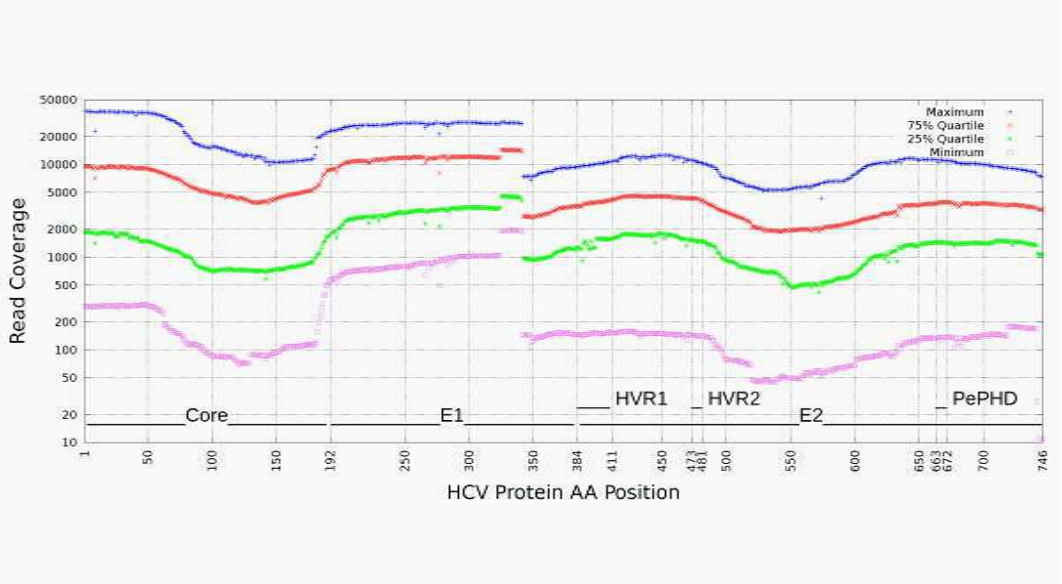
도면2


Nuber of Aligned Reads
Total Number of Reads
Number of Aligned Reads
Read Count
Sample Number

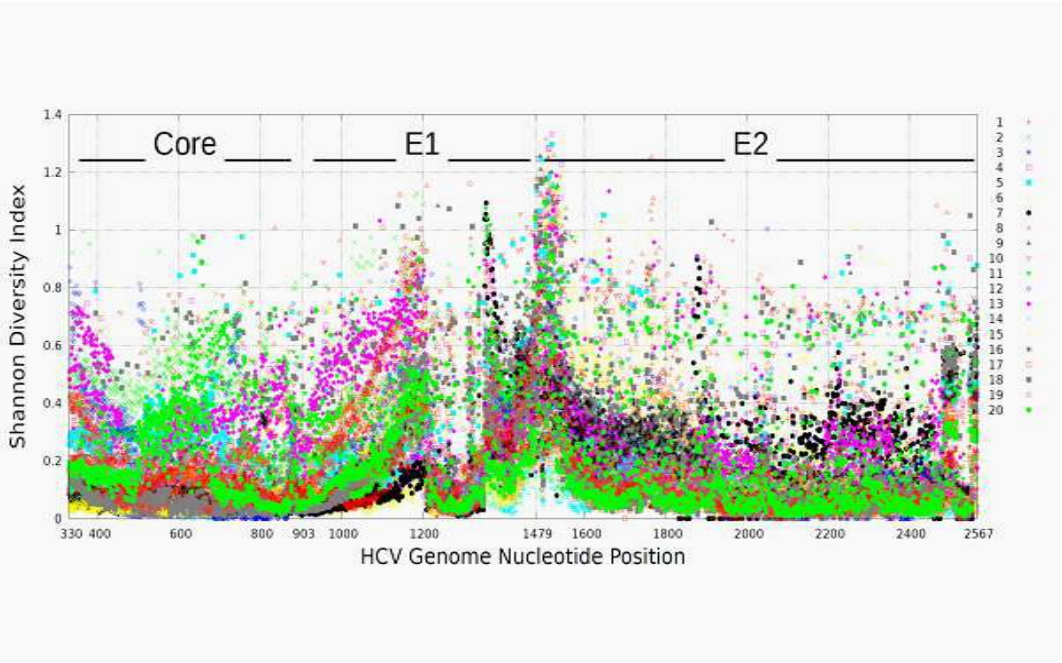
도면3


Average Read Length
Average Read Length
Average Read Length
Sample Number

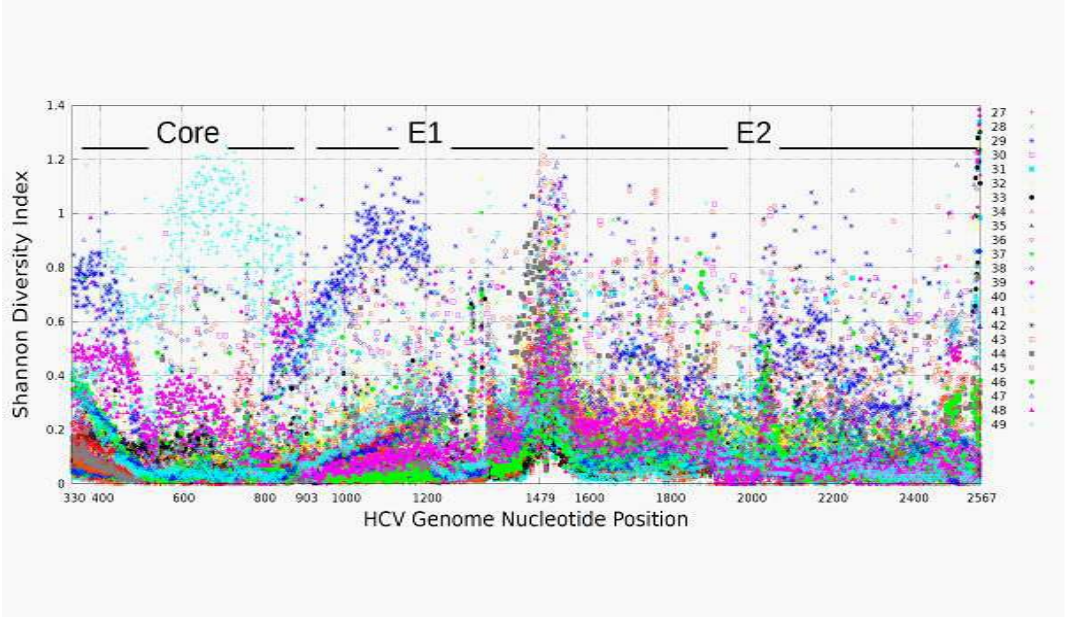
도면4


Maximum
75% Quartile
25% Quartile
Minimum
Read Coverage
50000
20000
10000
5000
2000
1000
500
200
100
50
20
10
Core
E1
HVR1
HVR2
E2
PePHD
1
50
100
150
192
250
300
350
384
411
450
473
481
500
550
600
650
663
672
700
746
HCV Protein AA Position

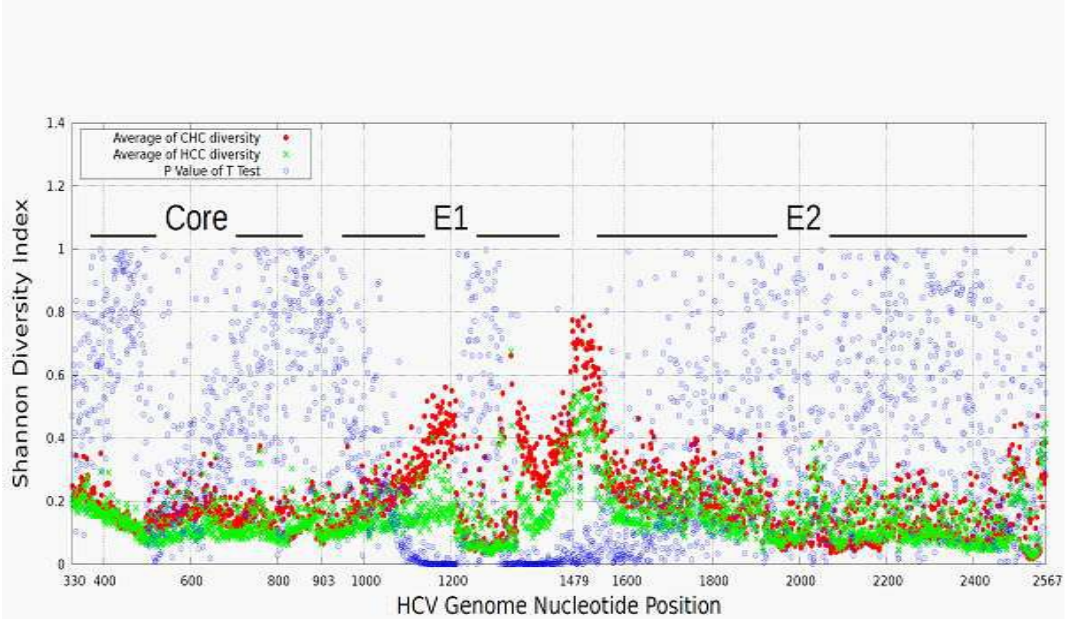
도면5


Core
E1
E2
Shannon Diversity Index
1.4
1.2
1
0.8
0.6
0.4
0.2
0
330
400
600
800
903
1000
1200
1479
1600
1800
2000
2200
2400
2567
HCV Genome Nucleotide Position
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

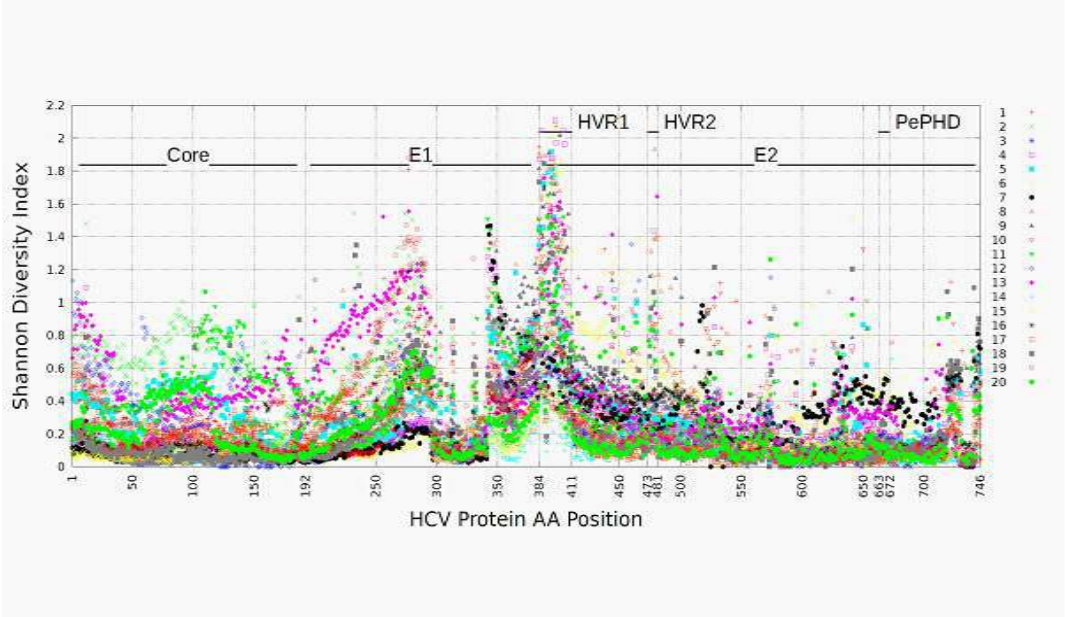
도면6


Core
E1
E2
Shannon Diversity Index
HCV Genome Nucleotide Position
330 400 600 800 903 1000 1200 1479 1600 1800 2000 2200 2400 2567

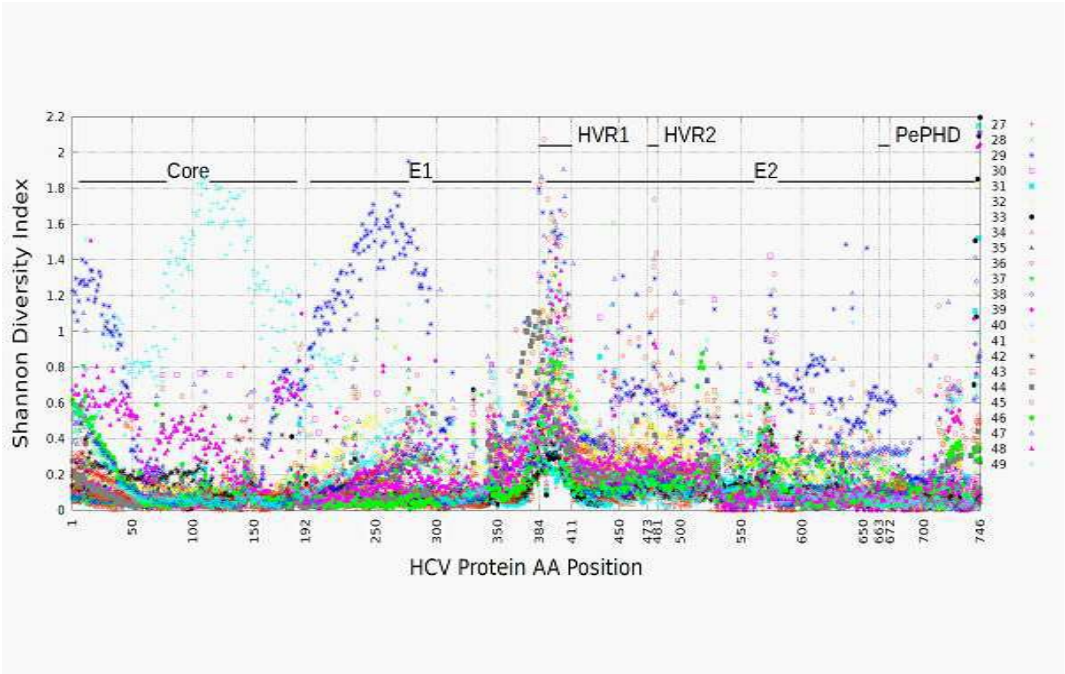
도면7


Average of CHC diversity
Average of HCC diversity
P Value of T Test
Core
E1
E2
Shannon Diversity Index
HCV Genome Nucleotide Position
330 400 600 800 903 1000 1200 1479 1600 1800 2000 2200 2400 2567

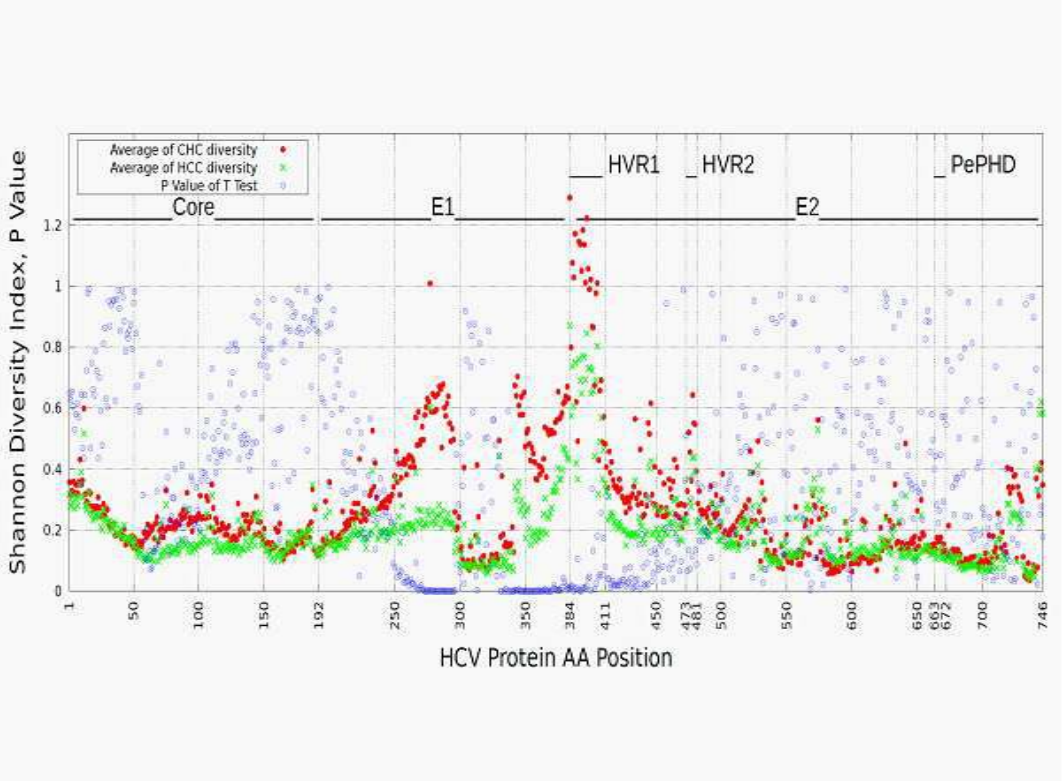
도면8


Core
E1
HVR1
HVR2
E2
PePHD
Shannon Diversity Index
HCV Protein AA Position

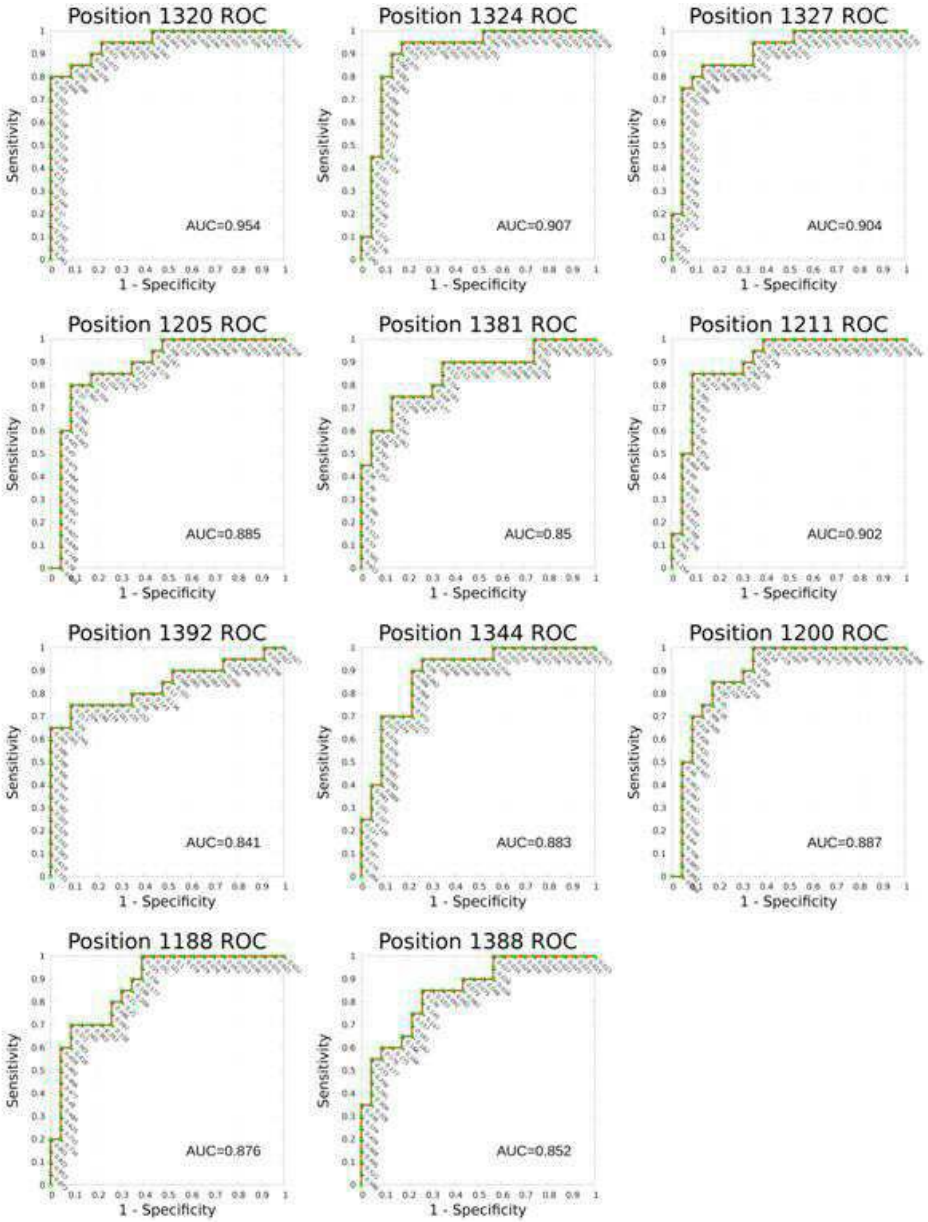
도면9


Core
E1
HVR1
HVR2
E2
PePHD
Shannon Diversity Index
HCV Protein AA Position

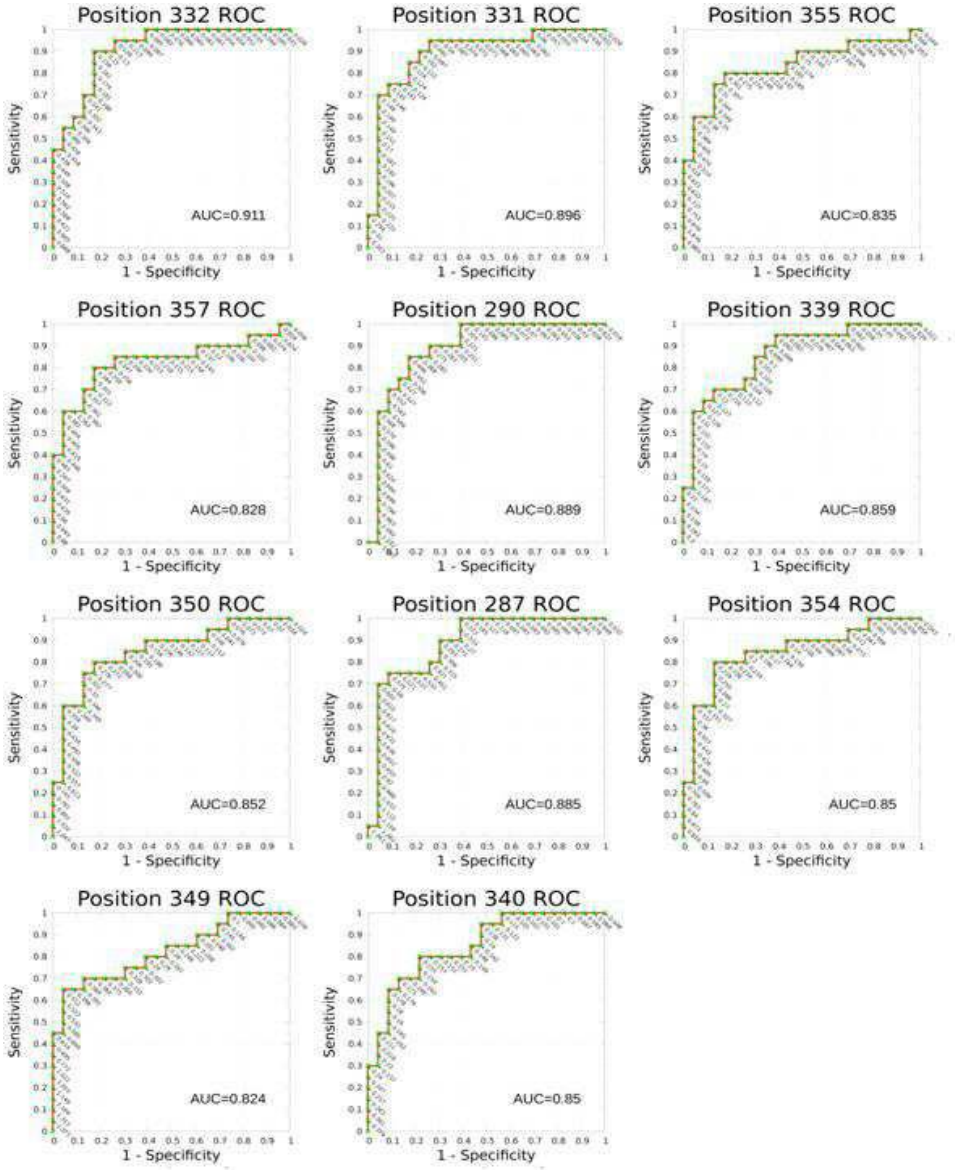
도면10


Average of CHC diversity
Average of HCC diversity
P Value of T Test
Core
E1
HVR1
HVR2
E2
PePHD
Shannon Diversity Index, P Value
HCV Protein AA Position

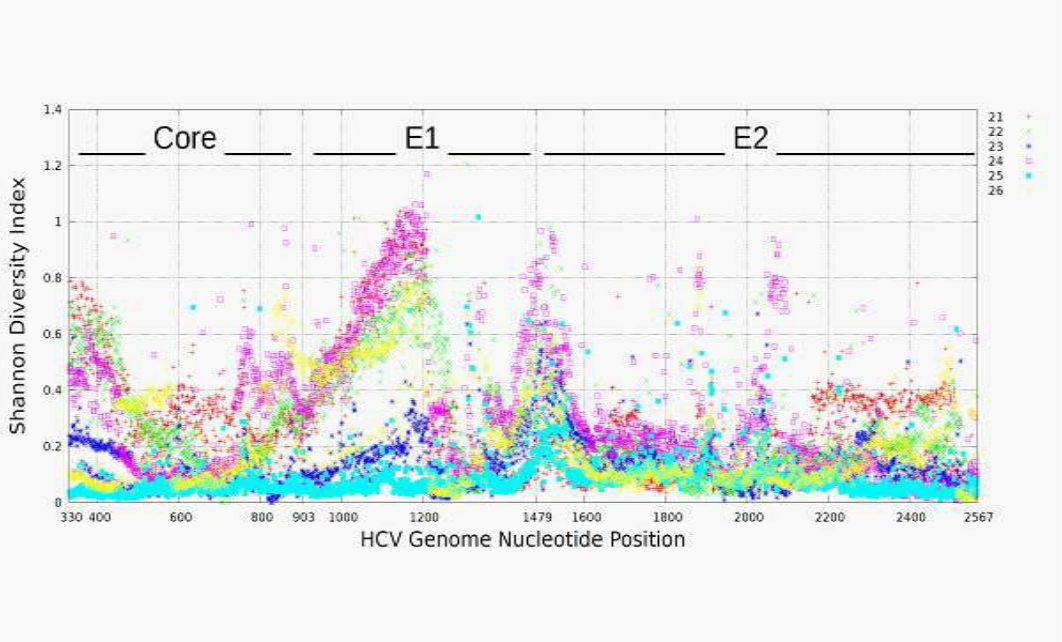
도면11


Position 1320 ROC
AUC=0.954
Position 1324 ROC
AUC=0.907
Position 1327 ROC
AUC=0.904
Position 1205 ROC
AUC=0.885
Position 1381 ROC
AUC=0.85
Position 1211 ROC
AUC=0.902
Position 1392 ROC
AUC=0.841
Position 1344 ROC
AUC=0.883
Position 1200 ROC
AUC=0.887
Position 1188 ROC
AUC=0.876
Position 1388 ROC
AUC=0.852
Sensitivity
1 - Specificity

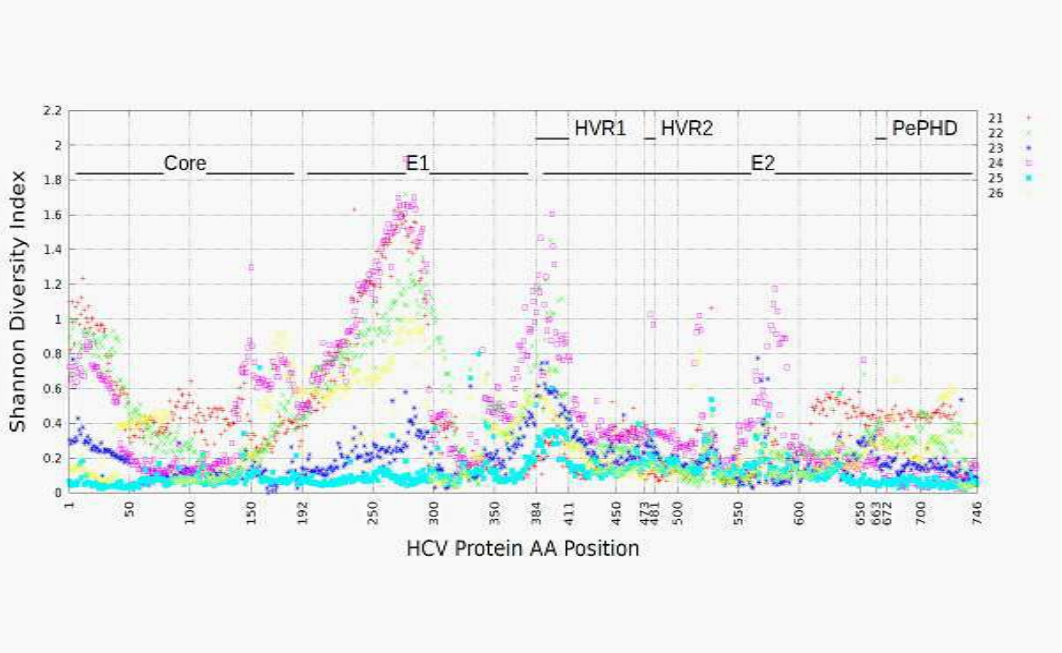
도면12


Position 332 ROC
AUC=0.911
Position 331 ROC
AUC=0.896
Position 355 ROC
AUC=0.835
Position 357 ROC
AUC=0.828
Position 290 ROC
AUC=0.889
Position 339 ROC
AUC=0.859
Position 350 ROC
AUC=0.852
Position 287 ROC
AUC=0.885
Position 354 ROC
AUC=0.85
Position 349 ROC
AUC=0.824
Position 340 ROC
AUC=0.85
Sensitivity
1 - Specificity

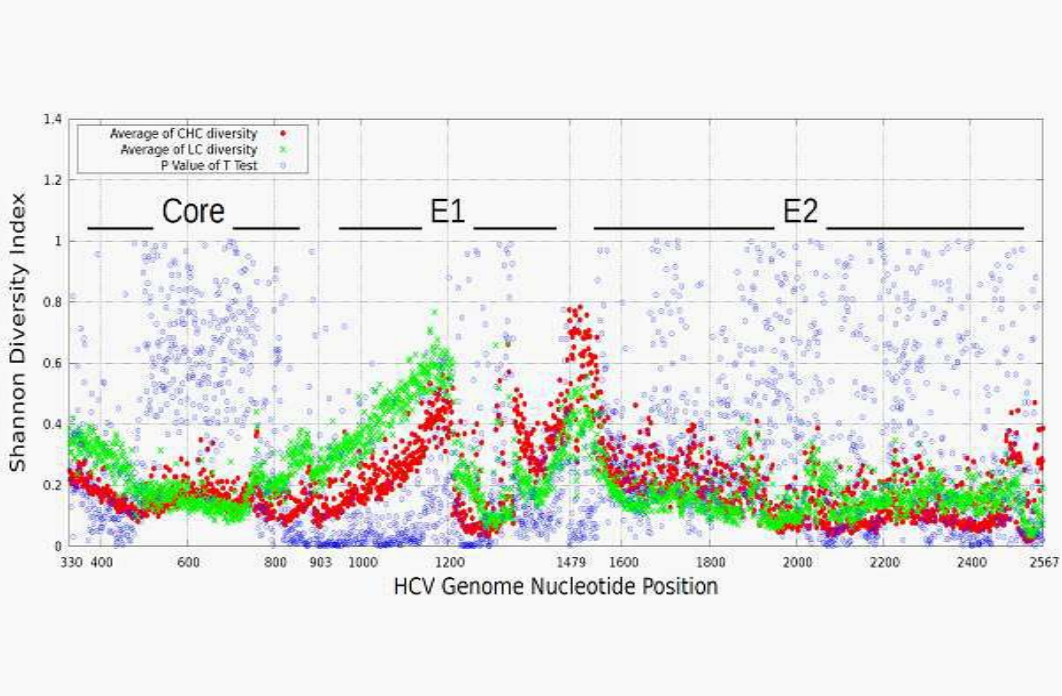
도면13


Core
E1
E2
Shannon Diversity Index
HCV Genome Nucleotide Position
21
22
23
24
25
26

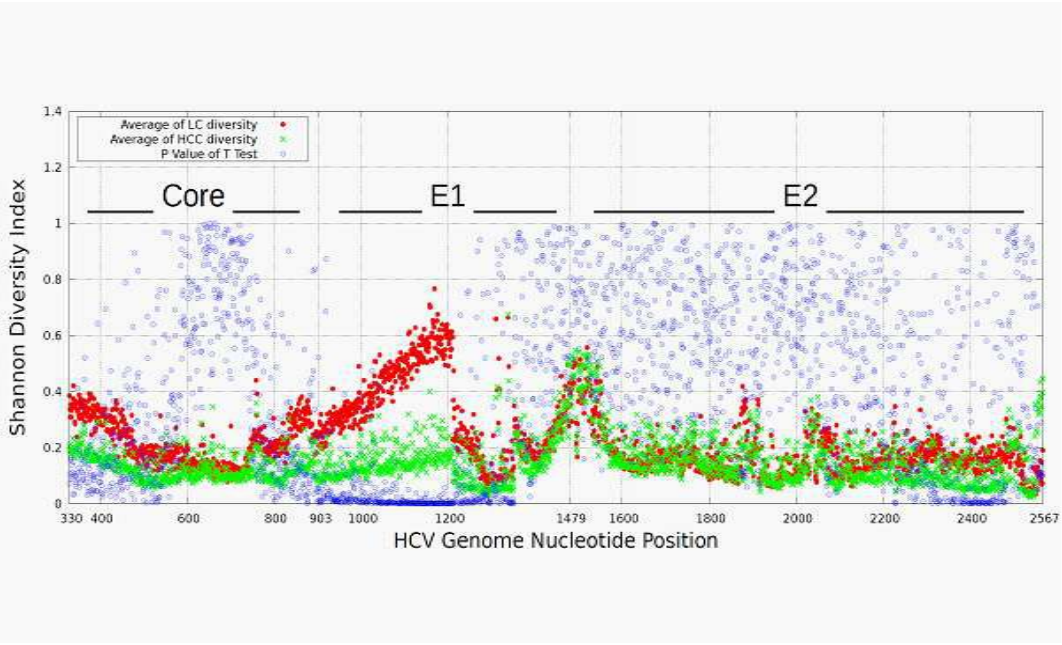
도면14


HVR1
HVR2
PePHD
Core
E1
E2
Shannon Diversity Index
HCV Protein AA Position
21
22
23
24
25
26

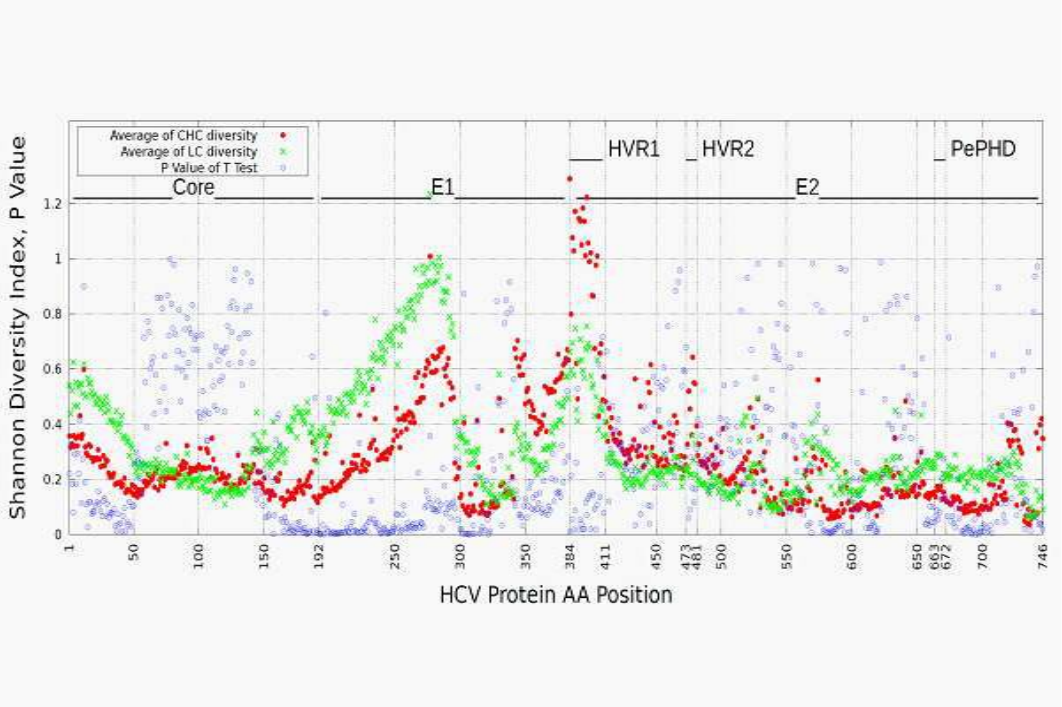
도면15


Average of CHC diversity
Average of LC diversity
P Value of T Test
Core
E1
E2
Shannon Diversity Index
HCV Genome Nucleotide Position
0
0.2
0.4
0.6
0.8
1
1.2
1.4
330
400
600
800
903
1000
1200
1479
1600
1800
2000
2200
2400
2567

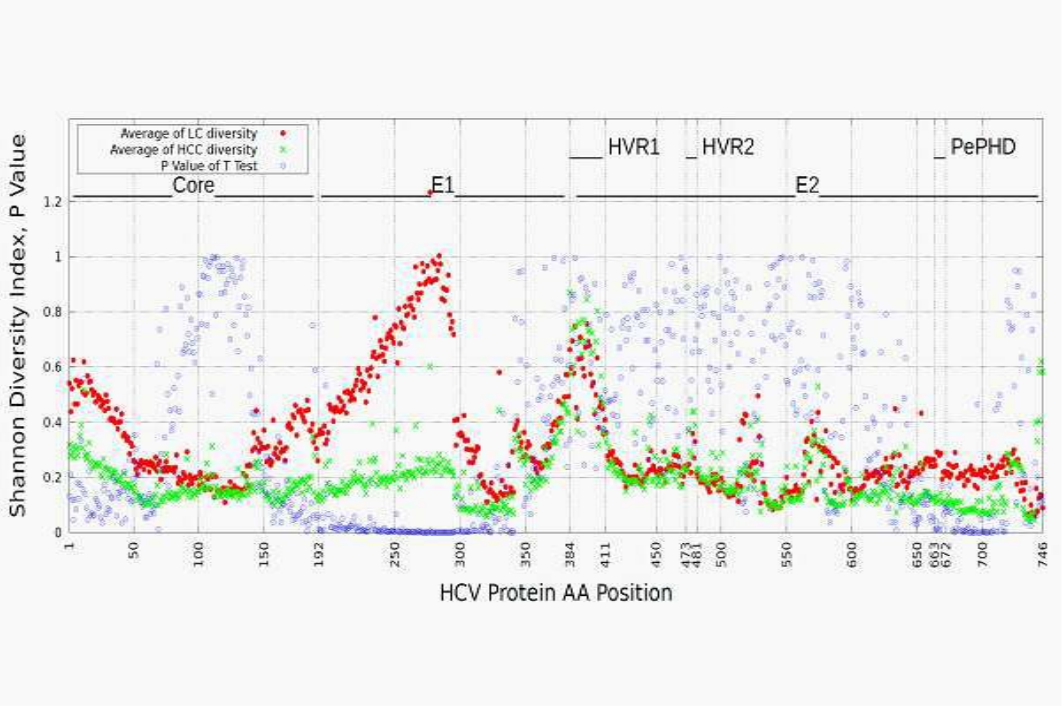
도면16


Average of LC diversity
Average of HCC diversity
P Value of T Test
Core
E1
E2
Shannon Diversity Index
HCV Genome Nucleotide Position
0
0.2
0.4
0.6
0.8
1
1.2
1.4
330
400
600
800
903
1000
1200
1479
1600
1800
2000
2200
2400
2567

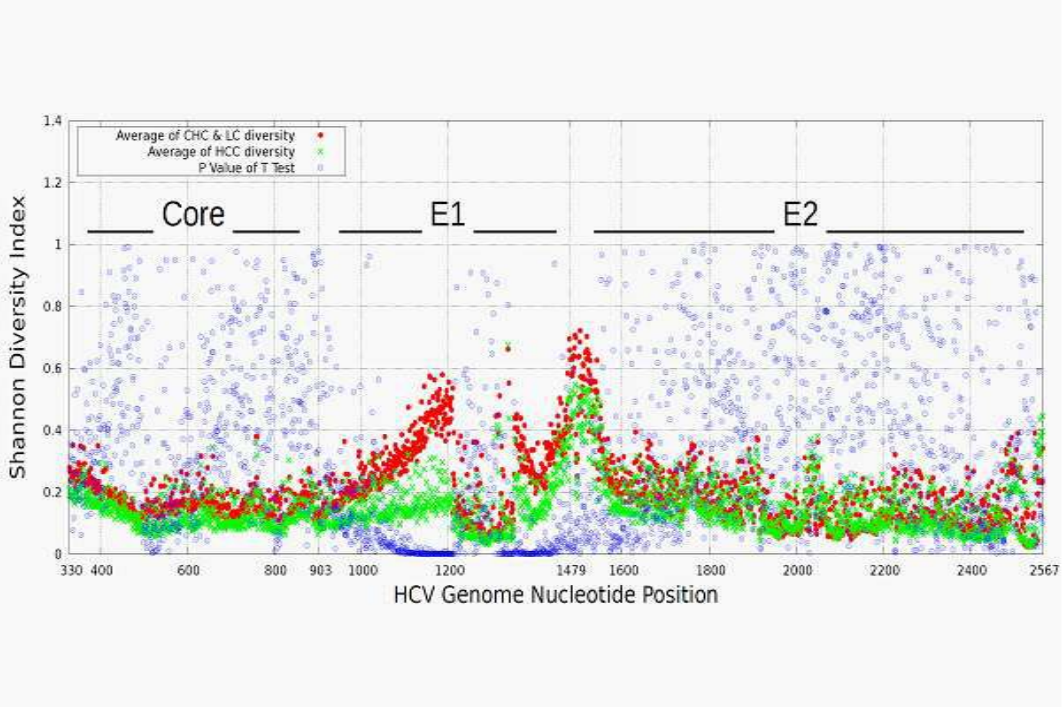
도면17


Average of CHC diversity
Average of LC diversity
P Value of T Test
Core
E1
HVR1
HVR2
E2
PePHD
Shannon Diversity Index, P Value
HCV Protein AA Position

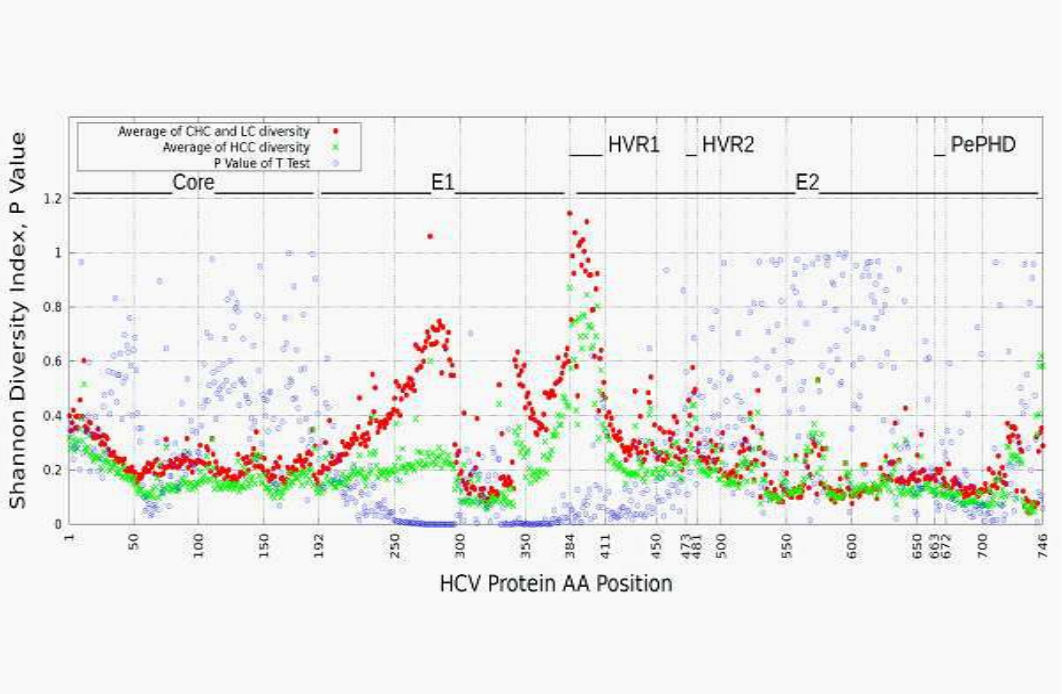
도면18


Average of LC diversity
Average of HCC diversity
P Value of T Test
Core
E1
HVR1
HVR2
E2
PePHD
Shannon Diversity Index, P Value
HCV Protein AA Position

도면19



도면20



서 열 목 록

<110> University of Ulsan Foundation For Industry Cooperation

<120> Markers and Method for Diagnosing Liver Cancer Caused Hapatatis C Virus

<130> DP-2011-0838

<160> 10

<170> KopatentIn 2.0

<210> 1

<211> 3010

<212> PRT

<213> Hepatitis C Virus

<400> 1

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn
1 5 10 15

Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly
20 25 30

Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala
35 40 45

Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro
50 55 60

Ile Pro Lys Ala Arg Xaa Pro Glu Gly Arg Xaa Trp Ala Gln Pro Gly
65 70 75 80

Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Xaa Gly Trp Ala Gly Trp
85 90 95

Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro

100 105 110

Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys
115 120 125

Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu
130 135 140

Gly Gly Ala Ala Arg Ala Leu Ala His Gly Val Arg Val Leu Glu Asp
145 150 155 160

Gly Val Asn Tyr Ala Thr Gly Asn Xaa Pro Gly Cys Ser Phe Ser Ile
165 170 175

Phe Leu Leu Ala Leu Leu Ser Cys Leu Thr Ile Pro Ala Ser Ala Tyr

180 185 190

Glu Val Arg Asn Val Ser Gly Val Tyr His Val Thr Asn Asp Cys Ser
195 200 205
Asn Xaa Ser Ile Val Tyr Glu Ala Ala Asp Met Ile Met Xaa Thr Pro
210 215 220
Gly Cys Val Pro Cys Val Arg Glu Xaa Asn Ser Ser Arg Cys Trp Val
225 230 235 240
Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Xaa Ser Xaa Pro Thr Thr

245 250 255
Thr Ile Arg Arg Xaa Val Xaa Leu Leu Val Gly Ala Ala Ala Phe Cys
260 265 270
Ser Ala Met Tyr Val Gly Asp Leu Cys Gly Ser Val Phe Leu Val Ser
275 280 285
Gln Leu Phe Thr Phe Ser Pro Arg Arg His Glu Thr Val Gln Asp Cys
290 295 300
Asn Cys Ser Ile Tyr Pro Gly His Val Xaa Gly Xaa Arg Met Ala Trp
305 310 315 320

Asp Met Met Met Asn Trp Ser Pro Thr Xaa Ala Leu Val Val Ser Gln
325 330 335
Leu Leu Arg Ile Pro Gln Ala Val Val Xaa Met Val Ala Gly Ala His
340 345 350
Trp Gly Val Leu Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp
355 360 365
Ala Lys Val Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly Xaa
370 375 380
Thr Xaa Xaa Xaa Gly Gly Xaa Xaa Xaa Xaa Xaa Thr Xaa Xaa Xaa Xaa

385 390 395 400
Xaa Xaa Xaa Xaa Xaa Gly Xaa Xaa Gln Xaa Ile Gln Leu Xaa Asn Thr
405 410 415
Asn Gly Ser Trp His Ile Asn Arg Thr Ala Leu Asn Cys Asn Asp Ser
420 425 430
Leu Xaa Thr Gly Phe Xaa Ala Ala Leu Phe Tyr Xaa His Xaa Phe Asn
435 440 445

Xaa Ser Gly Cys Pro Glu Arg Met Ala Ser Cys Arg Xaa Xaa Asp Xaa
450 455 460

Phe Xaa Gln Gly Trp Gly Pro Ile Thr Tyr Ala Glu Pro Xaa Xaa Xaa
465 470 475 480
Asp Gln Arg Pro Tyr Cys Trp His Tyr Ala Pro Xaa Pro Xaa Gly Xaa
485 490 495
Val Pro Ala Ser Xaa Val Cys Gly Pro Val Tyr Cys Phe Thr Pro Ser
500 505 510
Pro Val Val Val Gly Thr Thr Asp Arg Xaa Gly Xaa Pro Thr Tyr Xaa
515 520 525
Trp Gly Glu Asn Glu Thr Asp Val Leu Leu Leu Asn Asn Thr Arg Pro

530 535 540
Pro Xaa Gly Asn Trp Phe Gly Cys Thr Trp Met Asn Xaa Thr Gly Phe
545 550 555 560
Thr Lys Thr Xaa Gly Gly Pro Pro Cys Asn Ile Gly Gly Val Gly Xaa
565 570 575
Asn Thr Leu Xaa Cys Pro Thr Asp Cys Phe Arg Lys His Pro Glu Ala
580 585 590
Thr Tyr Thr Lys Xaa Gly Ser Gly Pro Trp Leu Thr Pro Arg Cys Xaa
595 600 605

Val Asp Tyr Pro Tyr Arg Leu Trp His Tyr Pro Cys Thr Val Asn Phe
610 615 620
Thr Ile Phe Lys Val Arg Met Tyr Val Gly Gly Val Glu His Arg Leu
625 630 635 640
Xaa Ala Ala Cys Xaa Trp Thr Arg Gly Glu Arg Cys Xaa Leu Glu Asp
645 650 655
Arg Asp Arg Ser Glu Leu Ser Pro Leu Leu Leu Ser Thr Thr Glu Trp
660 665 670
Gln Xaa Leu Pro Cys Ser Phe Thr Thr Leu Pro Ala Leu Ser Thr Gly

675 680 685
Leu Ile His Leu His Gln Asn Ile Val Asp Val Gln Tyr Leu Tyr Gly

690 695 700

Ile Gly Ser Ala Val Val Ser Phe Ala Ile Lys Trp Glu Tyr Val Xaa

705 710 715 720

Leu Leu Phe Leu Leu Leu Ala Asp Ala Arg Val Xaa Ala Cys Leu Trp

725 730 735

Met Met Leu Leu Ile Ala Gln Ala Glu Ala Ala Leu Glu Asn Xaa Val

740 745 750

Val Leu Asn Ala Ala Ser Val Ala Gly Ala His Gly Ile Leu Ser Phe

755 760 765

Leu Val Phe Phe Cys Ala Ala Trp Tyr Ile Lys Gly Arg Leu Val Pro

770 775 780

Gly Ala Ala Tyr Ala Xaa Tyr Gly Val Trp Pro Leu Leu Leu Leu Leu

785 790 795 800

Leu Ala Leu Pro Pro Arg Ala Tyr Ala Met Asp Arg Glu Met Ala Ala

805 810 815

Ser Cys Gly Gly Ala Val Xaa Xaa Gly Leu Xaa Leu Leu Thr Leu Ser

820 825 830

Pro His Tyr Lys Val Phe Leu Ala Arg Leu Ile Trp Trp Xaa Gln Tyr

835 840 845

Xaa Ile Thr Arg Ala Glu Ala Xaa Leu Gln Val Trp Xaa Pro Pro Leu

850 855 860

Asn Val Arg Gly Gly Arg Asp Ala Ile Ile Leu Leu Thr Xaa Ala Val

865 870 875 880

His Pro Glu Leu Ile Phe Asp Ile Thr Lys Xaa Xaa Leu Ala Ile Leu

885 890 895

Gly Pro Leu Met Val Leu Gln Ala Gly Ile Thr Arg Val Pro Tyr Xaa

900 905 910

Val Arg Ala Xaa Gly Leu Ile Arg Ala Cys Met Xaa Val Arg Lys Val

915 920 925

Ala Gly Gly His Tyr Val Gln Met Ala Xaa Met Lys Leu Ala Ala Leu

930 935 940

Thr Gly Thr Tyr Val Tyr Asp His Leu Thr Pro Leu Xaa Asp Trp Ala

945 950 955 960

His Ala Gly Leu Arg Asp Leu Ala Val Ala Val Glu Pro Val Val Phe

965 970 975

Ser Asp Met Glu Thr Lys Ile Ile Thr Trp Gly Ala Asp Thr Ala Ala

980 985 990

Cys Gly Asp Ile Ile Xaa Gly Leu Pro Val Ser Ala Arg Arg Gly Arg

995 1000 1005

Glu Ile Leu Leu Gly Pro Ala Xaa Ser Xaa Xaa Gly Gln Gly Trp Arg

1010 1015 1020

Leu Leu Ala Pro Ile Thr Ala Tyr Xaa Gln Gln Thr Arg Gly Leu Leu

1025 1030 1035 1040

Gly Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu

1045 1050 1055

Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr

1060 1065 1070

Cys Xaa Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys

1075 1080 1085

Thr Leu Ala Gly Pro Xaa Gly Pro Ile Thr Gln Met Tyr Thr Asn Val

1090 1095 1100

Asp Gln Asp Leu Val Gly Trp Xaa Ala Pro Pro Gly Ala Arg Ser Xaa

1105 1110 1115 1120

Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Xaa His

1125 1130 1135

Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu

1140 1145 1150

Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro

1155 1160 1165

Leu Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val

1170 1175 1180

Cys Thr Arg Gly Val Ala Lys Ala Val Asp Xaa Xaa Pro Val Glu Ser

1185 1190 1195 1200

Met Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro
1205 1210 1215
Pro Ala Val Pro Gln Thr Phe Gln Val Ala His Leu His Ala Pro Thr
1220 1225 1230
Gly Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly
1235 1240 1245
Tyr Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe

1250 1255 1260
Gly Ala Tyr Met Ser Lys Ala His Gly Xaa Asp Pro Asn Ile Arg Thr
1265 1270 1275 1280
Gly Val Arg Thr Ile Thr Thr Gly Ala Pro Ile Thr Tyr Ser Thr Tyr
1285 1290 1295
Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile
1300 1305 1310
Ile Ile Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu Gly
1315 1320 1325

Ile Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val
1330 1335 1340
Val Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro Xaa Pro
1345 1350 1355 1360
Xaa Ile Glu Glu Val Ala Leu Ser Asn Xaa Gly Glu Ile Pro Phe Tyr
1365 1370 1375
Gly Lys Ala Ile Pro Xaa Glu Xaa Ile Lys Gly Gly Arg Xaa Leu Xaa
1380 1385 1390
Phe Cys His Ser Lys Lys Xaa Cys Xaa Glu Leu Ala Ala Lys Leu Ser

1395 1400 1405
Gly Leu Gly Leu Xaa Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser
1410 1415 1420
Val Ile Pro Thr Ser Gly Asp Val Val Val Val Ala Thr Asp Ala Leu
1425 1430 1435 1440
Met Thr Gly Xaa Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Xaa Thr
1445 1450 1455

Cys Val Thr Gln Thr Val Xaa Phe Ser Leu Asp Pro Thr Phe Thr Ile
1460 1465 1470

Glu Thr Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg
1475 1480 1485
Gly Arg Thr Gly Arg Gly Arg Xaa Gly Ile Tyr Arg Phe Val Thr Pro
1490 1495 1500
Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys
1505 1510 1515 1520
Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser
1525 1530 1535
Val Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln

1540 1545 1550
Asp His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile
1555 1560 1565
Asp Ala His Phe Xaa Ser Gln Thr Xaa Gln Ala Gly Asp Asn Phe Pro
1570 1575 1580
Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro
1585 1590 1595 1600
Pro Pro Ser Trp Xaa Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro
1605 1610 1615

Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln
1620 1625 1630
Xaa Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys
1635 1640 1645
Met Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly
1650 1655 1660
Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Thr Thr Gly Ser Val
1665 1670 1675 1680
Val Ile Val Gly Arg Ile Ile Leu Ser Gly Xaa Pro Ala Xaa Ile Pro

1685 1690 1695
Asp Arg Glu Val Leu Tyr Xaa Glu Phe Asp Glu Met Glu Glu Cys Ala

1700 1705 1710

Ser His Leu Pro Tyr Ile Glu Gln Gly Met Gln Leu Ala Glu Gln Phe

1715 1720 1725

Lys Gln Lys Ala Leu Gly Leu Leu Gln Thr Ala Thr Lys Gln Ala Glu

1730 1735 1740

Ala Ala Ala Pro Val Val Glu Ser Lys Trp Arg Ala Leu Glu Xaa Phe

1745 1750 1755 1760

Trp Ala Lys His Met Trp Xaa Phe Ile Ser Gly Ile Gln Tyr Leu Ala

1765 1770 1775

Gly Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala

1780 1785 1790

Phe Thr Ala Ser Ile Thr Ser Pro Leu Thr Thr Gln Xaa Thr Leu Leu

1795 1800 1805

Phe Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Pro Pro Ser

1810 1815 1820

Ala Ala Ser Ala Phe Val Gly Ala Gly Xaa Ala Gly Ala Ala Val Gly

1825 1830 1835 1840

Ser Ile Gly Leu Gly Lys Val Leu Val Asp Ile Xaa Ala Gly Tyr Gly

1845 1850 1855

Ala Gly Val Ala Gly Ala Leu Val Ala Phe Lys Val Met Ser Gly Glu

1860 1865 1870

Met Pro Ser Thr Glu Asp Leu Val Asn Xaa Leu Pro Ala Ile Leu Ser

1875 1880 1885

Pro Gly Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg

1890 1895 1900

Xaa Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile

1905 1910 1915 1920

Ala Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro

1925 1930 1935

Glu Ser Asp Ala Ala Ala Arg Val Thr Gln Ile Leu Ser Ser Leu Thr

1940 1945 1950

Ile Thr Gln Leu Xaa Lys Arg Leu His Gln Trp Xaa Xaa Glu Asp Cys

1955 1960 1965

Ser Thr Pro Cys Ser Gly Ser Trp Leu Arg Asp Val Trp Asp Trp Ile

1970 1975 1980

Cys Thr Val Leu Thr Xaa Phe Lys Thr Trp Leu Gln Ser Lys Leu Leu

1985 1990 1995 2000

Pro Arg Xaa Pro Gly Val Pro Phe Xaa Ser Xaa Gln Arg Gly Tyr Lys

2005 2010 2015

Gly Val Trp Arg Gly Asp Gly Ile Met Xaa Thr Thr Cys Pro Cys Gly

2020 2025 2030

Ala Xaa Ile Thr Gly His Val Lys Asn Gly Ser Met Arg Ile Val Gly

2035 2040 2045

Pro Xaa Thr Cys Ser Asn Thr Trp His Gly Thr Phe Pro Ile Asn Ala

2050 2055 2060

Tyr Thr Thr Gly Pro Cys Thr Pro Ser Pro Ala Pro Asn Tyr Ser Arg

2065 2070 2075 2080

Ala Leu Trp Arg Val Ala Ala Glu Glu Tyr Val Glu Val Thr Arg Val

2085 2090 2095

Gly Asp Phe His Tyr Val Thr Gly Met Thr Thr Asp Asn Val Xaa Cys

2100 2105 2110

Pro Cys Gln Val Pro Ala Pro Glu Phe Phe Thr Xaa Val Asp Gly Val

2115 2120 2125

Arg Xaa His Arg Tyr Ala Pro Ala Cys Lys Pro Leu Leu Arg Xaa Glu

2130 2135 2140

Val Thr Phe Gln Val Gly Leu Asn Gln Tyr Leu Val Gly Ser Gln Leu

2145 2150 2155 2160

Pro Xaa Glu Pro Glu Pro Asp Val Xaa Val Leu Thr Ser Met Leu Thr

2165 2170 2175

Asp Pro Ser His Xaa Thr Ala Glu Thr Ala Lys Arg Arg Leu Ala Arg

2180 2185 2190

Gly Ser Pro Pro Ser Leu Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala

2195 2200 2205

Pro Ser Leu Lys Ala Thr Cys Thr Thr Xaa His Asp Ser Pro Asp Ala
2210 2215 2220
Asp Leu Ile Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn
2225 2230 2235 2240
Ile Thr Arg Val Glu Ser Glu Xaa Lys Val Val Ile Leu Asp Ser Phe
2245 2250 2255
Xaa Pro Leu Arg Ala Glu Glu Asp Glu Arg Glu Val Ser Val Pro Ala

2260 2265 2270
Glu Ile Leu Arg Lys Ser Arg Xaa Phe Pro Xaa Ala Met Pro Ile Trp
2275 2280 2285
Ala Arg Pro Asp Tyr Asn Pro Pro Leu Xaa Glu Ser Trp Lys Asp Pro
2290 2295 2300
Asp Tyr Val Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Thr Lys
2305 2310 2315 2320
Ala Pro Pro Ile Pro Pro Pro Arg Xaa Lys Arg Thr Val Val Leu Thr
2325 2330 2335

Xaa Ser Thr Val Ser Ser Ala Leu Ala Glu Leu Ala Thr Lys Thr Phe
2340 2345 2350
Gly Ser Ser Xaa Ser Ser Ala Val Asp Ser Gly Thr Ala Thr Ala Pro
2355 2360 2365
Pro Asp Gln Xaa Ser Asp Asp Gly Asp Xaa Gly Ser Asp Val Glu Ser
2370 2375 2380
Tyr Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu
2385 2390 2395 2400
Ser Asp Gly Ser Trp Ser Thr Val Ser Glu Glu Ala Ser Glu Asp Val

2405 2410 2415
Val Cys Cys Ser Met Ser Tyr Thr Trp Thr Gly Ala Leu Ile Thr Pro
2420 2425 2430
Cys Ala Ala Glu Glu Ser Lys Leu Pro Ile Xaa Ala Xaa Ser Asn Ser
2435 2440 2445
Leu Leu Arg His His Asn Met Val Tyr Ala Thr Thr Ser Arg Ser Ala
2450 2455 2460

Ser Xaa Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp
2465 2470 2475 2480

Asp His Tyr Arg Asp Val Leu Lys Glu Met Lys Ala Lys Ala Ser Thr
2485 2490 2495
Val Lys Ala Lys Leu Leu Ser Val Glu Glu Ala Cys Lys Leu Thr Pro
2500 2505 2510
Pro His Ser Ala Xaa Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg
2515 2520 2525
Asn Leu Ser Ser Lys Ala Xaa Asn His Ile Arg Ser Val Trp Lys Asp
2530 2535 2540
Leu Leu Glu Asp Thr Xaa Thr Pro Ile Asp Thr Thr Ile Met Ala Lys

2545 2550 2555 2560
Asn Glu Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala
2565 2570 2575
Arg Leu Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Xaa Met
2580 2585 2590
Ala Leu Tyr Asp Val Val Ser Thr Leu Pro Gln Ala Val Met Gly Ser
2595 2600 2605
Ser Tyr Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val
2610 2615 2620

Asn Ala Trp Lys Ser Lys Lys Xaa Pro Met Gly Phe Xaa Tyr Asp Thr
2625 2630 2635 2640
Arg Cys Phe Asp Ser Thr Val Thr Glu Xaa Asp Ile Arg Val Glu Glu
2645 2650 2655
Ser Xaa Tyr Gln Cys Cys Asp Leu Ala Pro Glu Ala Arg Gln Ala Ile
2660 2665 2670
Arg Ser Leu Thr Glu Arg Leu Tyr Ile Gly Gly Pro Leu Thr Asn Ser
2675 2680 2685
Lys Gly Gln Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu

2690 2695 2700
Thr Thr Ser Cys Gly Asn Thr Leu Thr Cys Tyr Leu Lys Ala Xaa Ala

2705 2710 2715 2720

Ala Cys Arg Ala Ala Lys Leu Gln Asp Cys Thr Met Leu Val Xaa Gly

2725 2730 2735

Asp Asp Leu Val Val Ile Cys Xaa Ser Ala Gly Thr Xaa Glu Xaa Ala

2740 2745 2750

Ala Ser Leu Arg Val Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro

2755 2760 2765

Pro Gly Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Xaa Ile Thr Ser

2770 2775 2780

Cys Ser Ser Asn Val Ser Val Ala His Asp Ala Ser Gly Lys Arg Val

2785 2790 2795 2800

Tyr Tyr Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp

2805 2810 2815

Glu Thr Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile

2820 2825 2830

Met Tyr Ala Pro Thr Xaa Trp Ala Arg Met Ile Leu Met Thr His Phe

2835 2840 2845

Phe Ser Ile Leu Leu Ala Gln Glu Gln Leu Glu Lys Ala Leu Asp Cys

2850 2855 2860

Gln Ile Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Gln

2865 2870 2875 2880

Ile Ile Xaa Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr

2885 2890 2895

Ser Pro Gly Glu Ile Asn Arg Val Ala Ser Cys Leu Arg Lys Leu Gly

2900 2905 2910

Val Pro Pro Leu Arg Val Trp Arg His Arg Ala Arg Ser Val Arg Ala

2915 2920 2925

Lys Leu Leu Ser Gln Gly Gly Arg Ala Ala Thr Cys Gly Xaa Tyr Leu

2930 2935 2940

Phe Asn Trp Ala Val Xaa Thr Lys Leu Lys Leu Thr Pro Xaa Pro Ala

2945 2950 2955 2960

Ala Ser Gln Xaa Asp Leu Ser Xaa Trp Phe Val Ala Gly Tyr Ser Gly

2965 2970 2975

Gly Asp Ile Tyr His Ser Leu Ser Arg Ala Arg Pro Arg Trp Phe Met

2980 2985 2990

Xaa Cys Leu Leu Leu Leu Ser Val Gly Val Gly Ile Tyr Leu Leu Pro

2995 3000 3005

Asn Arg

3010

<210> 2

<211> 9386

<212> DNA

<213> Hepatitis C Virus

<400> 2

ttgggggcga cactccacca tagatcactc ccctgtgagg aactactgtc ttcacgcaga 60

aagcgtctag ccatggcgtt agtatgagtg tcgtgcagcc tccaggaccc cccctcccgg 120

gagagccata gtggtctgcg gaaccggtga gtacaccgga attgccagga cgaccgggtc 180

ctttcttgga tcaacccgct caatgcctgg agatttgggc gtgcccccgc gagactgcta 240

gccgagtagt gttgggtcgc gaaaggcctt gtggtactgc ctgatagggt gcttgcgagt 300

gccccgggag gtctcgtaga ccgtgcacca tgagcacgaa tcctaaacct caaagaaaaa 360

ccaaacgtaa caccaaccgc cgcccacagg acgtcaagtt cccgggcggt ggtcagatcg 420

ttggtggagt ttacctgttg ccgcgcaggg gccccaggtt gggtgtgcgc gcgactagga 480

agacttccga gcggtcgcaa cctcgtggaa ggcgacaacc tatccccaag gctcgccngc 540

ccgagggcag gncctgggct cagcccgggt acccttggcc cctctatggc aatgagggcn 600

tggggtgggc aggatggctc ctgtcacccc gnggctcncg gcctagttgg ggccccacng 660

acccccggcg taggtcgcgt aatttgggta aggtcatcga taccctcaca tgcggcttcg 720

ccgacctcat ggggtacatt ccgctcgtcg gcgccccccct agggggcgct gccagggccc 780

tggcgcatgg cgtccgggtt ctggaggacg gcgtgaacta tgcaacaggg aatntgcccg 840

gttgctcttt ctctatcttc ctcttggctt tgctgtcctg tttgaccatc ccagcttccg 900

cttatgaagt gcgcaacgtg tccggggtgt accatgtcac gaacgactgc tccaacncaa 960

gcattgtgta tgaggcagcg gacatgatca tgcanacccc cgggtgcgtg ccctgcgttc 1020

gggagnncaa ctcctcccgn tgctgggtag cgctcactcc cacgctcgcg gccaggaacn 1080

ccagcntccc cactacgaca atacgacgnc angtcgantt gctcgttggg gcggctgctt 1140

tctgctccgc tatgtacgtg ggggatctct gcggatctgt tttcctngtc tcccagctgt 1200

tcaccttctc gcctcgccgg catgagacag tacaggactg caattgctca atctatcccg 1260

gccacgtanc aggtcancgc atggcttggg atatgatgat gaactggtca cctacancag 1320

ccctagtggt ntcgcagtta ctccggatcc cacaagcngt cgtgganatg gtggcggggg 1380

cccactgggg agtcctggcg ggccttgcct actattccat ggtggggaac tgggctaagg 1440

tnttgattgt gatgctactc tttgccggcg ttgacgggnn naccnacnng ncnggggggn 1500

ngnnngnccn nancaccnnn ngnntnncgn ccntcttnnc nnnngggncg ncncagaana 1560

tccagcttnt aaacaccaac ggcagctggc acatcaacag gactgccctg aactgcaatg 1620

actccctcna nacngggttc nttgccgcgc tgttctacnc acacangttc aacncgtccg 1680

gatgcccaga gcgcatggcc agctgccgcn ccatngacna gttcgntcag gggtggggtc 1740

ccatcactta ngctgagccn nncnnctngg accagaggcc ntattgctgg cactacgcnc 1800

cncnaccgtg nggtatngta cccgcgtcgn aggtgtgtgg tccagtgtat tgcttcaccc 1860

caagccctgt tgtggtgggg acgaccgatc gntncggcgn ccctacgtan anctgggggg 1920

agaatgagac ggacgtgctg ctnctnaaca acacgcggcc gccgcanggc aactggttcg 1980

gctgtacatg gatgaatngc actgggttca ccaagacgtg ngggggcccc ccgtgtaaca 2040

tcgggggggt cggcaanaac accttganct gccccacgga ctgcttccgg aagcaccccg 2100

aggccactta cacnaaatgn ggntcggggc cttggttgac acctaggtgc ntngttgact 2160

acccatacag gctntggcac taccccctgca ctgtcaactt taccatcttc aaggttagga 2220

tgtatgtggg gggngtggag cacaggctca angccgcntg caantggact cgaggagagc 2280

gntgtnactt ggaggacagg gatagatcag agctnagccc gctgctgctg tctacaacag 2340

agtggcagnt actgccctgt tccttcacca ccctaccggc tctgtccact ggnttgatcc 2400

atctccatca gaacatcgtg gacgtgcaat acctgtacgg tatagggtcn gcggttgtct 2460

cctttgcaat caaatgggag tatgtcntgt tgctcttcct nctcctggcg gacgcgcgcg 2520

tctgngcctg cttgtggatg atgctgctga tagcncaggc tgaggccgcc ttagagaacn 2580

tggtggtcct caatgcggcg tccgtggccg gagcgcatgg cattctctcc ttccttgtgt 2640

tcttctgtgc tgcctggtac atcaagggca ggctggtccc tggggcggca tatgctntct 2700

atggcgtatg gccgctgctc ctgctcctgc tggcgttacc accacgagcn tacgccatgg 2760

accgggagat ggctgcatcg tgcggaggcg cggttttnnt aggtctggna ctcttgacct 2820

tgtcaccaca ctanaaagtg ttcctcgcna ggctcatatg gtggttncaa tatttnatca 2880

ccagagccga ggcgcanttg caagtgtggn tccccccnct caacgttcgg gggggccgcg 2940

atgccatcat cctcctcacg tgngcggtcc acccagagct aatctttgac atcaccaaan 3000

tcntgctcgc catactcggt ccgctcatgg tgctccaggc tggcataacn agagtgccgt 3060

acttngtgcg cgctcanggg ctcattcgtg catgcatgnt ggtgcggaaa gtcgctgggg 3120

gncattatgt ccaaatggcn ntcatgaagc tggccgcnct gacnggnacg tacgtttatg 3180

accatcttac tccactgcng gactgggccc acgcgggcct acgagacctt gcggtggcag 3240

tngagcccgt cgtcttctct gacatggaga ccaagatcat cacctggggg gcagacaccg 3300

cggcgtgtgg ggacatcatc tngggnctnc ccgtctccgc ccgaaggggg agggagatac 3360

tnctgggacc ggcnganagt nttgangggc aggggtggcg actcctngcg ccnatcacgg 3420

cctacnccca acagacgcgg ggcctacttg gctgcatcat cacnagcctc acaggccggg 3480

acaagaacca ggtcgagggg gaggttcaag tggtntccac cgcaacacaa tctttcctgg 3540

cgacctgcnt caacggcgtg tgttggactg tctancatgg ngccggctca aagaccctag 3600

ccggcccaaa nggnccaatc acccaaatgt acaccaatgt agaccaggac ctcgtcggct 3660

ggcangcgcc ccccggggcg cgntccntga caccntgcac ctgcggcagc tcggaccttt 3720

acttggtcac gagncatgct gatgtcattc cggtgcgccg gcggggcgac agcaggggga 3780

gcctnctctc ccccaggccc gtctcctact tgaagggctc ttcgggtggt ccactgctct 3840

gcccctcggg gcacgctgtg ggcatcttcc gggctgcngt gtgcacccgg ggggttgcga 3900

aggcggtgga cttnntnccc gttgagtcta tggaaacnac natgcggtcn ccggtcttca 3960

cggacaactc ntccccnccg gccgtaccgc agacattcca agtggcccat ctacacgcnc 4020

ccactggcag cggcaagagc actaaggtgc cggctgcnta tgcagcccaa gggtacaagg 4080

tnctcgtcct gaacccgtcc gttgccgcca ccttaggntt tggggcgtat atgtcnaagg 4140

cacatggtnt cgacccnaac atcagaactg gggtaaggac catcaccacg ggcgcccccа 4200

tcacgtactc cacctanggc aagttccttg ccgacggtgg ttgctcnggg ggcgcctatg 4260

acatcataat atgtgatgag tgccactcaa ctgactcgac taccatcttg ggcatcggca 4320

cagtcctgga ccaagcggag acggctggag cgcgnctcgt cgtgctcgcc accgctacnc 4380

ctccgggatc ggtcaccgtg ccacanccca anatcgagga ggtggccctg tccaacantg 4440

gagagatccc cttctatggc aaagccatcc ccatngagnc catcaagggg gggaggcanc 4500

tcatnttctg ccattccaag aagaantgtg angagctcgc cgcaaagctg tcnggcctcg 4560

gactcaangc tgtagcntat taccggggtc ttgatgtgtc cgtcataccg acnagcggag 4620

acgtcgttgt cgtggcaaca gacgctctaa tgacgggctn naccggcgac tttgactcag 4680

tgatcgactg taanacatgt gtcacccaga cagtcgantt cagcttggac ccnaccttca 4740

ccattgagac gacgaccgtg ccccaagacg cggtgtcgcg ctcgcagcgg cgaggcagga 4800

ctggnagggg caggnnnggc atctacaggt ttgtgactcc aggagaacgg ccctcgggca 4860
tgttcgattc ntcggtcctg tgtgagtgct atgacgcggg ctgtgcttgg tangagctca 4920
cgcccgccga gacctcngtt aggttgcggg cttacctnaa tacaccaggg ttgcccgtct 4980
gccaggacca tctggagttc tgggagagcg tcttcacagg cctcacccac atagatgccc 5040
acttcntgtc ccagactaan caggcaggag acaacttccc ctacctggta gcataccagg 5100
cnacngtgtg cgccagggct caggcnccac ctccatcgtg ggancaaatg tggaagtgtc 5160
tcatacggct naagcctacg ctgcacgggc caacacccct gctgtatagg ctaggagccg 5220

tccaaaanga ggtcaccctc acacacccca taaccaaata catcatggca tgcatgtcgg 5280
ctgacctgga ggtcgtcacn agcacctggg tgctngtngg cggngtcctn gcagctctgg 5340
ccgcgtantg cctgacaaca ggcagcgtgg tcattgtggg caggatcatc ttgtccggga 5400
ngccggcnnt nattcccgac agggaagtnc tctaccngga gttcgatgag atggaagagt 5460
gcgcctcaca cctcccttac atcgaacagg gaatgcagct cgccgagcaa ttcaagcaga 5520
aggcgctcgg gttgctgcaa acagccacca agcaagcgga ggctgctgct cccgtggtgg 5580
agtccaagtg gcgagccctt gagnccttct gggcgaagca catgtggaan ttcatcagcg 5640

ggatacagta cttagcaggc ttgtccactc tgcctgggaa ccccgcnata gcatcactga 5700
tggcattcac agcctcnatc accagcccgc tcaccaccca annnaccctc ctgtttaaca 5760
tcttgggggg atgggtggcn gcccaactcg cncccccccag cgctgcttcn gctttcgtng 5820
gcgccggcat ngcnggngcg gcngttggca gcataggcct tgggaaggtg cttgtggaca 5880
tcntggcggg ntatggngcn ggggtggcng gcgcnctcgt ggcctttaag gtcatgagcg 5940
gcgagatgcc ctccacngag gacctggtca acttnctccc tgccatcctc tcncctggng 6000
ccctggtcgt cggggtcgtg tgcgcagcaa tactgcgtcg gcangtgggc ccnggngagg 6060

gggctgtgca gtggatgaac cggctgatag cgttcgcttc gcggggtaac cacgtctccc 6120
ccacgcacta tgtgcctgag agcgacgcng cagcgcgtgt cactcagatc ctctccagcc 6180
ttaccatcac tcagctgntg aagaggctcc accagtggat naangaggac tgctccacgc 6240
catgctccgg ctcgtggcta agggatgttt gggactggat atgcacggtg ttgactgant 6300
tcaagacctg gctccagtcc aagctcctgc cgcggttncc gggagtcccn ttcntctcat 6360
gncaacgngg gtacaaggga gtctggcggg gngacggcat catgcanacc acctgcccat 6420
gtggagcaca natcaccgga catgtcaaaa acggttccat gaggatcgtt gggcctanna 6480

cctgcagcaa cacgtggcat ggaacattcc ccatcaacgc ntacaccacg ggcccctgca 6540

cnccctcccc ggcgccnaac tattccaggg cgctgtggcg ggtggctgct gaggagtacg 6600

tggaggtnac gcgggtgggg gatttccact acgtgacggg catgaccact gacaacgtaa 6660

antgcccatg ccaggttccg gcccccgaat tcttcacnga ngtggatggg gtgcggntgc 6720

acaggtacgc tccggcgtgc aaacctctcc tacgggganga ggtcacattc caggtcgggc 6780

tcaaccaata cctggtnggg tcacagctcc catgngagcc cgaaccggat gtancagtgc 6840

tcacttccat gctcaccgac ccctcccaca tnacagcaga gacggctaag cgtaggctgg 6900

ccaggggggtc tccccccctcc ttggccagct cttcagctag ccagttgtct gcgccttcnt 6960

tgaaggcgac atgcactacc cntcatgact ccccngacgc tgacctcatc gaggccaacc 7020

tcctgtggcg gcaggagatg ggcggnaaca tcacccgcgt ggagtcagag aanaaggtag 7080

taattctgga ctctttcgan ccgcttcgag cggaggagga tgagagggaa gtatccgttc 7140

cggcggagat cctgcggaaa tccaggaant tccccccnngc gatgcccata tgggcacgcc 7200

cggattacaa ccctccactg ntagagtcct ggaaggaccc ggactacgtc cctccggtgg 7260

tacacgggtg cccattgcca cctaccaagg cccctccaat accacctcca cggagnaaga 7320

ggacggttgt cctgacagan tccacngtgt cttctgcctt ggcggagctn gcnacnaaga 7380

ccttcggcag ctccgnatcg tcggccgtcg acagcggcac ggcgaccgcc cctcctgacc 7440

agncctccga cgacggngac ncaggatccg acgttgagtc gtactcctcc atgccccccc 7500

ttgagggnga gccgggggac cccgatctca gcgacgggtc ttggtctacc gtgagcgagg 7560

aggcnagtga ggacgtcgtc tgctgctcga tgtcctacac atggacaggc gccctgatca 7620

cnccatgcgc ngcggaggaa agcaagctgc ccatcaangc gntgagcaac tcnttgctgc 7680

gncaccacaa catggtctan gccacaacat cccgcagcgc aagccancgg cagaagaagg 7740

tcacctttga cagactgcaa gtcctggacg accactaccg ggacgtgctc aaggagatga 7800

aggcgaaggc gtccacagtt aaggctaaac ttctatcngt agaagaagcc tgcaagctga 7860

cgcccccaca ttcggccana tccaaatttg gctatggggc aaaggacgtc cggaacctat 7920

ccagcaaggc cnttaaccac atccgctccg tgtggaagga cttgctggaa gacactgana 7980

caccaattga caccaccatc atggcaaaaa atgaggtttt ctgcgtccaa ccagagaaag 8040

gaggccgcaa gccagctcgc cttatcgtnt tcccagactt gggngttcgt gtgtgcgaga 8100

anatggccct ttangacgtg gtctccaccc ttcctcaggc cgtgatgggc tcctcatacg 8160

gattccagta ctctcctggn cagcgggtcg agttcctggt gaatgcctgg aaatcaaaga 8220

aannccctat gggcttcnca tatgacaccc gctgttttga ctcaacggtc actgagantg 8280

acatccgtgt tgaggagtca atntaccaat gttgtgactt ggcccccgaa gccagacagg 8340

ccataaggtc gctcacagag cggctttata tcgggggtcc cctgactaat tcaaaagggc 8400

agaactgcgg ttatcgccgg tgccgcgcna gcggcgtgct gacgacnagc tgcggtaata 8460

ccctcacatg ttacttgaag gccnctgcng cctgtcgagc tgcnaagctc caggactgca 8520

cgatgctcgt gnncggagac gaccttgtcg ttatctgtga nagcgcggga acccangagg 8580

angcggcgag cctacgagtc ttcacggagg ctatgactag gtactctgcc ccccccgggg 8640

acccgcccca accagaatac gacttggagn tgataacatc atgctcctcc aatgtgtcgg 8700

tcgcgcacga tgcatctggc aaaagggtgt actacctcac ccgtgacccc accacccccc 8760

tngcncgggc tgcgtgggag acagctagac acactccagt naactcctgg ctaggcaaca 8820

tcatcatgta tgcgcccacc ttntgggcaa ggatgattct gatgactcac ttcttctcca 8880

tccttctagc ncaggagcaa cttgaaaaag ccctngattg tcagatctac ggggcctgtt 8940

actccattga gccacttgac ctacctcaga tcattnancg actccatggt cttagcgcat 9000

tttcactcca tagttactct ccaggtgaga tcaatagggt ggcttcatgc ctcaggaaac 9060

ttggggtacc acccttgcga gtctggagac atcgggccag aagtgtccgc gctaagctac 9120

tgtcccaggg ggggagggcn gccacttgtg gcaantacct cttcaactgg gcagtaanga 9180

ccaagctnaa actcactcca atnccggctg cgtcccagnt ggacttgtcc ngctggttcg 9240

ttgctggtta cagcgggggga gacatatatc acagcctgtc tcgtgcccga ccccgctggt 9300

tcatgtngtg cctactccta ctttcngtag gggtaggcat ctacctgctc cccaaccgat 9360

gaacggggag ctaancactc caggcc 9386

<210> 3

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Oligonucleotide primer

<400> 3

gggcgacact ccaccatag 19

<210> 4

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Oligonucleotide primer

<400> 4

gccggcraar agyagcayc 19

<210> 5

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Oligonucleotide primer

<400> 5

actcccctgt gaggaactay 20

<210> 6

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> Oligonucleotide primer

<400> 6

tgggatccgg agyarctg 18

<210> 7

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Oligonucleotide primer

<400> 7

ttctgytccg cyatgtayg 19

<210> 8

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Oligonucleotide primer

<400> 8

cacaagraag gagagranrc 20

<210> 9

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Oligonucleotide primer

<400> 9

gaactggtcr ccyacarcrg 20

<210> 10

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Oligonucleotide primer

<400>

10

gaactggtcr ccyacarcrg 20