

	(19) 대한민국특허청(KR) (12) 공개특허공보(A)	(11) 공개번호 10-2013-0016736 (43) 공개일자 2013년02월19일
(51) 국제특허분류(Int. Cl.) C12Q 1/68 (2006.01) C12Q 1/02 (2006.01) C12N 15/11 (2006.01) (21) 출원번호 10-2011-0078209 (22) 출원일자 2011년08월05일 심사청구일자 2012년01월18일		(71) 출원인 울산대학교 산학협력단 울산 남구 무거2동 산29 (72) 발명자 이종은 경기도 고양시 일산서구 강선로 164, 1403동 1403호 (일산동, 후곡마을) 오홍범 서울특별시 성동구 성수일로8길 47, 롯데캐슬아파트 106동 1502호 (성수동2가) (74) 대리인 리엔목특허법인

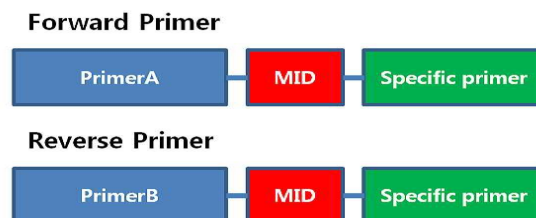
전체 청구항 수 : 총 6 항

(54) 발명의 명칭 **헤파티티스 B 바이러스의 아미노산 변이를 검출하는 키트 및 방법**

(57) 요약

본 발명은 헤파티티스 B 바이러스의 아미노산 변이를 검출하는 키트 및 이를 이용한 헤파티티스 B 바이러스의 아미노산 변이를 검출하는 방법에 관한 것이다. 일 구체예에 따른 헤파티티스 B 바이러스의 아미노산 변이를 검출하는 키트 및 방법에 의하면, 대상 시료로부터 헤파티티스 B 바이러스의 아미노산 변이를 용이하고 정확하게 검출할 수 있으며, 특히, 헤파티티스 B 바이러스의 약제 내성 부위를 분석하는데 효과적이다.

대표도 - 도1



(72) 발명자

유형진

서울특별시 노원구 공릉동 745 동신아파트
103-1302

이한나

경기도 구리시 건원대로 56, 313동 101호 (인창동,
삼보아파트)

고선영

서울특별시 송파구 올림픽로 99, 127동 1703호 (잠
실동, 잠실엘스)

박창욱

서울특별시 서대문구 모래내로17안길 30, 203호 (홍은동, 넥서스빌)

뉴클레오타이드로 이루어진 군으로부터 선택되는 폴리뉴클레오타이드; 및 서열 번호 168의 염기 서열을 갖는 폴리뉴클레오타이드의 순서로 이루어진 정방향 프라이머; 및 서열 번호 2의 염기 서열을 갖는 폴리뉴클레오타이드; 서열 번호 3 내지 서열 번호 153의 염기 서열을 갖는 폴리뉴클레오타이드로 이루어진 군으로부터 선택되는 폴리뉴클레오타이드; 및 서열 번호 169의 염기 서열을 갖는 폴리뉴클레오타이드의 순서로 이루어진 역방향 프라이머를 포함하는 프라이머 세트; 및

서열 번호 1의 염기 서열을 갖는 폴리뉴클레오타이드; 서열 번호 3 내지 서열 번호 153의 염기 서열을 갖는 폴리뉴클레오타이드로 이루어진 군으로부터 선택되는 폴리뉴클레오타이드; 및 서열 번호 170의 염기 서열을 갖는 폴리뉴클레오타이드의 순서로 이루어진 정방향 프라이머; 및 서열 번호 2의 염기 서열을 갖는 폴리뉴클레오타이드; 서열 번호 3 내지 서열 번호 153의 염기 서열을 갖는 폴리뉴클레오타이드로 이루어진 군으로부터 선택되는 폴리뉴클레오타이드; 및 서열 번호 171의 염기 서열을 갖는 폴리뉴클레오타이드의 순서로 이루어진 역방향 프라이머를 포함하는 프라이머 세트; 및

서열 번호 1의 염기 서열을 갖는 폴리뉴클레오타이드; 서열 번호 3 내지 서열 번호 153의 염기 서열을 갖는 폴리뉴클레오타이드로 이루어진 군으로부터 선택되는 폴리뉴클레오타이드; 및 서열 번호 172의 염기 서열을 갖는 폴리뉴클레오타이드의 순서로 이루어진 정방향 프라이머; 및 서열 번호 2의 염기 서열을 갖는 폴리뉴클레오타이드; 서열 번호 3 내지 서열 번호 153의 염기 서열을 갖는 폴리뉴클레오타이드로 이루어진 군으로부터 선택되는 폴리뉴클레오타이드; 및 서열 번호 173의 염기 서열을 갖는 폴리뉴클레오타이드의 순서로 이루어진 역방향 프라이머를 포함하는 프라이머 세트; 및

서열 번호 1의 염기 서열을 갖는 폴리뉴클레오타이드; 서열 번호 3 내지 서열 번호 153의 염기 서열을 갖는 폴리뉴클레오타이드로 이루어진 군으로부터 선택되는 폴리뉴클레오타이드; 및 서열 번호 172의 염기 서열을 갖는 폴리뉴클레오타이드의 순서로 이루어진 정방향 프라이머; 및 서열 번호 2의 염기 서열을 갖는 폴리뉴클레오타이드; 서열 번호 3 내지 서열 번호 153의 염기 서열을 갖는 폴리뉴클레오타이드로 이루어진 군으로부터 선택되는 폴리뉴클레오타이드; 및 서열 번호 174의 염기 서열을 갖는 폴리뉴클레오타이드의 순서로 이루어진 역방향 프라이머를 포함하는 프라이머 세트; 및

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

c) 상기 변환된 아미노산 서열 정보를 레퍼런스 아미노산 서열과 비교하여 서로 상이한 아미노산 서열 정보를 검색하는 단계; 및

d) 상기 검색된 정보를 사용자에게 출력하는 단계를 포함하는 헤파티티스 B 바이러스의 아미노산 변이를 검출하는 방법.

청구항 2

제1항에 있어서, 상기 시퀀싱하는 단계는 에멀전 기반 클론 PCR(emulsion based clonal PCR) 방법에 의해 시퀀싱하는 것인 방법.

청구항 3

제1항에 있어서, 상기 시퀀싱하는 단계에서 헤파티티스 B 바이러스의 염기 서열은 헤파티티스 B 바이러스의 전체 게놈 염기 서열, 헤파티티스 B 바이러스의 역전사 효소 유전자의 염기 서열 및 헤파티티스 B 바이러스의 YMDD 영역을 포함하는 유전자의 염기 서열로 이루어진 군으로부터 선택되는 것인 방법.

청구항 4

제1항에 있어서, 상기 검출 대상 시료 내에 포함된 헤파티티스 B 바이러스는 약제 내성을 갖는 바이러스인 것인 방법.

청구항 5

제1항의 b) 내지 d) 단계의 방법을 컴퓨터에서 실행시키기 위한 프로그램으로 기록한 컴퓨터로 읽을 수 있는 기록매체.

청구항 6

제1항의 프라이머 세트; 및

제5항의 기록매체로 이루어진 헤파티티스 B 바이러스의 아미노산 변이를 검출하기 위한 키트.

명세서

기술 분야

[0001] 본 발명은 헤파티티스 B 바이러스의 아미노산 변이를 검출하는 키트 및 이를 이용한 헤파티티스 B 바이러스의 아미노산 변이를 검출하는 방법에 관한 것이다.

배경 기술

[0002] 헤파티티스 B 바이러스(Hepatitis B virus, HBV)는 게놈의 전체 크기가 약 3,200 bp인 DNA 바이러스로서 한국을 포함하여 전세계적으로 약 3억 5천만 명이 감염되어 있는 것으로 알려져 있다. HBV의 보균자는 만성 B형 간염, 간경화, 간암으로 진행할 위험성이 높다. 현재까지 B형 간염 치료제로 널리 쓰이는 약제는 인터페론(interferon)과 라미부딘(lamivudine)이다. 그러나, 라미부딘 등의 약물 치료는 HBV의 유전자에 약제 내성을 일으키는 돌연변이를 유발할 수 있다. 이러한 약제 내성은 투약 후 매년 10 내지 15%의 비율로 증가하는 것으로 알려져 있으며, 이로 인해 약물의 효과가 저하되고, 치료가 지연됨으로써 간염의 증상이나 간기능의 악화를 초래한다.

[0003] 약제 내성을 일으키는 돌연변이는 주로 역전사 효소(reverse transcriptase)를 코딩하고 있는 RT 도메인에서 일어나는데, 일어나는 부위에 따라 라미부딘, 엔테카비르(entecavir), 아데포비르(adeфовir) 등 HBV 감염 치료제에 대한 내성이 발생하게 된다.

[0004] HBV 감염 환자를 위한 검사로는 혈청학적 및 분자생물학적 검사가 있다. 이중 급성 및 만성 간염의 진단을 위해서는 HBsAg, anti-HBs, HBeAg, anti-HBe와 anti-HBc를 측정하는 혈청학적 검사가 널리 이용된다. 이에 비하여 분자생물학적 기법을 이용한 HBV의 DNA 검출은 HBV 감염의 진단 이외에도 치료 중인 환자의 추적 검사시에 이용되고 있다. 현재 사용되고 있는 분자생물학적 기법에는 실시간 PCR을 이용하여 HBV의 특정 유전자를 단순 정량하는 방법과 라미부딘에 저항성을 나타내는 돌연변이 부위인 YMDD 부위를 포함하는 유전자를 Sanger 등의 방법에 의해 직접 시퀀싱하여 약제 내성 HBV의 존재 여부를 판별하는 방법이 주로 사용되고 있다.

[0005] 실시간 PCR을 이용한 정량 기법은 HBV의 감염 양을 측정하는데 도움을 주지만, 약제 내성을 가진 HBV의 존재를 판별할 수는 없다. 직접 시퀀싱을 이용한 YMDD 부위의 genotyping은 라미부딘에 대한 약제 내성을 가진 HBV의 존재를 확인할 수 있으나, 내성 발생 초기에 변이종이 미량 존재할 경우에는 판별이 불가능하며, 여러 종류의 변이종이 혼재할 경우에는, 우점종만을 판별할 수 있는 등 정확한 판별이 불가능하다.

[0006] 따라서, 종래 기술에 의하더라도 헤파티티스 B 바이러스의 아미노산 변이를 검출하는 키트 및 방법에 대한 새로운 기술이 요구된다.

발명의 내용

해결하려는 과제

[0007] 일 구체에는 헤파티티스 B 바이러스의 아미노산 변이를 검출하는 키트를 제공하는 것이다.

[0008] 다른 구체에는 상기 키트를 이용한 헤파티티스 B 바이러스의 아미노산 변이를 검출하는 방법을 제공하는 것이다.

과제의 해결 수단

[0009] 일 양상은 a) 서열 번호 1의 염기 서열을 갖는 폴리뉴클레오티드; 서열 번호 3 내지 서열 번호 153의 염기 서열을 갖는 폴리뉴클레오티드로 이루어진 군으로부터 선택되는 폴리뉴클레오티드; 및 서열 번호 154의 염기 서열을 갖는 폴리뉴클레오티드의 순서로 이루어진 정방향 프라이머; 및 서열 번호 2의 염기 서열을 갖는 폴리뉴클레오티드; 서열 번호 3 내지 서열 번호 153의 염기 서열을 갖는 폴리뉴클레오티드로 이루어진 군으로부터 선택되는 폴리뉴클레오티드; 및 서열 번호 155의 염기 서열을 갖는 폴리뉴클레오티드의 순서로 이루어진 역방향 프라이머를 포함하는 프라이머 세트;

- [illegible]

는 프라이머 세트에 구성되는 프라이머 세트 또는

- [0018] 서열 번호 1의 염기 서열을 갖는 폴리뉴클레오타이드; 서열 번호 3 내지 서열 번호 153의 염기 서열을 갖는 폴리뉴클레오타이드로 이루어진 군으로부터 선택되는 폴리뉴클레오타이드; 및 서열 번호 172의 염기 서열을 갖는 폴리뉴클레오타이드의 순서로 이루어진 정방향 프라이머; 및 서열 번호 2의 염기 서열을 갖는 폴리뉴클레오타이드; 서열 번호 3 내지 서열 번호 153의 염기 서열을 갖는 폴리뉴클레오타이드로 이루어진 군으로부터 선택되는 폴리뉴클레오타이드; 및 서열 번호 173의 염기 서열을 갖는 폴리뉴클레오타이드의 순서로 이루어진 역방향 프라이머를 포함하는 프라이머 세트; 및
- [0019] 서열 번호 1의 염기 서열을 갖는 폴리뉴클레오타이드; 서열 번호 3 내지 서열 번호 153의 염기 서열을 갖는 폴리뉴클레오타이드로 이루어진 군으로부터 선택되는 폴리뉴클레오타이드; 및 서열 번호 172의 염기 서열을 갖는 폴리뉴클레오타이드의 순서로 이루어진 정방향 프라이머; 및 서열 번호 2의 염기 서열을 갖는 폴리뉴클레오타이드; 서열 번호 3 내지 서열 번호 153의 염기 서열을 갖는 폴리뉴클레오타이드로 이루어진 군으로부터 선택되는 폴리뉴클레오타이드; 및 서열 번호 174의 염기 서열을 갖는 폴리뉴클레오타이드의 순서로 이루어진 역방향 프라이머를 포함하는 프라이머 세트에 구성되는 프라이머 세트를 사용하여 검출 대상 시료 내에 포함된 헤파티티스 B 바이러스의 염기 서열을 시퀀싱하는 단계;
- [0020] b) 상기 시퀀싱 단계로부터 얻은 염기 서열 정보를 아미노산 서열 정보로 변환하는 단계;
- [0021] c) 상기 변환된 아미노산 서열 정보를 레퍼런스 아미노산 서열과 비교하여 서로 상이한 아미노산 서열 정보를 검색하는 단계; 및
- [0022] d) 상기 검색된 정보를 사용자에게 출력하는 단계를 포함하는 헤파티티스 B 바이러스의 아미노산 변이를 검출하는 방법을 제공한다.
- [0023] 상기 헤파티티스 B 바이러스의 아미노산 변이를 검출하는 방법을 각각의 단계별로 상세하게 설명하면 다음과 같다:
- [0024] 먼저, 상기 방법은, 상기 프라이머 세트를 사용하여 검출 대상 시료 내에 포함된 헤파티티스 B 바이러스의 염기 서열을 시퀀싱하는 단계를 포함한다.
- [0025] 용어 “프라이머(primer)”는 적합한 온도에서 적합한 완충액 내에서 적합한 조건(즉, 4종의 다른 뉴클레오타이드 트리포스페이트 및 중합 반응 효소) 하에서 주형-지시 DNA 합성의 개시점으로 작용할 수 있는 단일 가닥의 올리고뉴클레오타이드를 의미한다. 프라이머의 적합한 길이는 다양한 인자, 예를 들어, 온도와 프라이머의 용도에 따라 차이가 있지만 전형적으로 15 내지 30개의 뉴클레오타이드이다. 짧은 프라이머는 주형과 충분히 안정된 혼성화 복합체를 형성하기 위하여 일반적으로 보다 낮은 온도를 요구할 수 있다. 용어 “정방향 프라이머(forward primer)” 및 “역방향 프라이머(reverse primer)”는 중합 효소 연쇄 반응에 의해 증폭되는 주형의 일정한 부위의 3' 말단 및 5' 말단에 각각 결합하는 프라이머를 의미한다. 프라이머의 서열은 주형의 일부 서열과 완전하게 상보적인 서열을 가질 필요는 없으며, 주형과 혼성화되어 프라이머 고유의 작용을 할 수 있는 범위 내에서의 충분한 상보성을 가지면 충분하다. 따라서, 일 구체예에 따른 프라이머 세트는 주형인 표적 핵산 서열에 완벽하게 상보적인 서열을 가질 필요는 없으며, 이 서열에 혼성화되어 프라이머 작용을 할 수 있는 범위 내에서 충분한 상보성을 가지면 충분한 것으로 해석된다. 이러한 프라이머의 디자인은 주형이 되는 표적 핵산의 염기 서열을 참조하여 당업자에 의해 용이하게 실시할 수 있으며, 예를 들어, 프라이머 디자인용 프로그램(예를 들어, PRIMER 3, VectorNTI 프로그램)을 이용하여 할 수 있다. 한편, 일 구체예에 따른 프라이머는 주형의 한 부위에 혼성화 또는 어닐링되어, 이중체 구조를 형성한다. 이러한 이중체 구조를 형성하는 데 적합한 핵산 혼성화의 조건은 Joseph Sambrook, et al., *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.(2001) 및 Haymes, B. D., 등, *Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington, D.C.(1985)에 개시되어 있다.
- [0026] 일 구체예에 따르면, 상기 시퀀싱하는 단계는 에멀전 기반 클론 PCR(emulsion based clonal PCR) 방법에 의해 시퀀싱하는 것일 수 있다. 본 발명의 시퀀싱은 예를 들어, Roche사의 게놈 Sequencer FLX(GS FLX, 454) 시스템을 이용할 수 있다. 상기 시스템은 현재 사용 가능한 시퀀싱 기술 중에 염기 서열을 가장 길게 분석할 수 있는 것으로서, 에멀전 기반 클론 PCR 방법에 의해 시퀀싱을 수행한다. 상기 시퀀싱의 수행 방법은 Roche 사의 프로토콜에 따른다.
- [0027] 상기 프라이머 세트를 구성하는 프라이머는 크게 세부분으로 구분될 수 있다. 서열 번호 1의 염기 서열을 갖는 폴리뉴클레오타이드 및 서열 번호 2의 염기 서열을 갖는 폴리뉴클레오타이드는 Roche 사에서 제공한 염기 서열로,

상기 에멀전 기반 클론 PCR을 수행하기 위해 필수적인 부분이다. 상기 서열 번호 1 또는 서열 번호 2의 염기 서열을 갖는 폴리뉴클레오타이드의 3' 말단에 연결된 서열 번호 3 내지 서열 번호 153의 염기 서열을 갖는 폴리뉴클레오타이드는 Roche사에서 예시로 제공한 일종의 구분자로서, 여러 종류의 시료를 구별할 수 있게 하는 기능을 한다. 상기의 경우, 151 종의 서로 다른 폴리뉴클레오타이드가 연결될 수 있으므로, 최대 151개의 시료의 염기 서열을 분석할 수 있다. 서열 번호 3 내지 서열 번호 153의 염기 서열을 갖는 폴리뉴클레오타이드의 3' 말단에 연결된 서열 번호 154 내지 서열 번호 174의 염기 서열을 갖는 폴리뉴클레오타이드는 헤파티티스 B 바이러스의 게놈 상에 상보적으로 결합할 수 있는 헤파티티스 B 바이러스 특이적 프라이머로서의 기능을 할 수 있다. 이중, 서열 번호 172 내지 서열 번호 174의 염기 서열을 갖는 폴리뉴클레오타이드는 헤파티티스 B 바이러스의 약제 내성 부위(역전사 효소를 코딩하는 유전자(RT), 또는 YMDD 영역을 포함하는 유전자)의 증폭을 위한 프라이머로서 사용될 수 있다. 상기 프라이머 세트에 대한 상세한 설명은 하기 실시예에 나타내었다.

- [0028] 일 구체예에 따르면, 상기 시퀀싱하는 단계에서 헤파티티스 B 바이러스의 염기 서열은 헤파티티스 B 바이러스의 전체 게놈 염기 서열, 헤파티티스 B 바이러스의 역전사 효소 유전자의 염기 서열 및 헤파티티스 B 바이러스의 YMDD 영역을 포함하는 유전자의 염기 서열로 이루어진 군으로부터 선택되는 것일 수 있다.
- [0029] 일 구체예에 따르면, 상기 검출 대상 시료 내에 포함된 헤파티티스 B 바이러스는 약제 내성을 갖는 바이러스인 것일 수 있다.
- [0030] 상기 검출 대상 시료는 헤파티티스 B 바이러스가 감염된 생물학적 시료일 수 있으며, 예를 들어, 혈액, 간세포 또는 간조직일 수 있으나, 이에 한정하지는 않는다. 헤파티티스 B 바이러스가 내성을 가질 수 있는 약제로는 아데포비르(Adefovir), 엔테카비르(Entecavir), 라미부딘(Lamivudine) 또는 텔비부딘(Telbivudine)일 수 있으나, 이에 한정하지는 않는다. 상기 약제에 대해 내성을 갖는 헤파티티스 B 바이러스는 헤파티티스 B 바이러스의 역전사 효소 유전자의 염기 서열 또는 헤파티티스 B 바이러스의 YMDD 영역을 포함하는 염기 서열에 돌연변이가 나타나, 결국 상기 유전자가 발현된 단백질 상의 아미노산 서열에 변이가 일어날 수 있다. 따라서, 상기 아미노산 서열의 변이를 확인함으로써 헤파티티스 B 바이러스의 약제 내성 변이종을 검출할 수 있다.
- [0031] 이후, 상기 방법은, 상기 시퀀싱 단계로부터 얻은 염기 서열 정보를 아미노산 서열 정보로 변환하는 단계를 포함한다.
- [0032] 상기 시퀀싱 장치(예를 들어, Roche사의 게놈 Sequencer FLX(GS FLX, 454) 시스템)로부터 시퀀싱이 완료가 되면, 이로부터 얻은 염기 서열 정보는 상기 장치와 별도로 기록매체에 포함된 프로그램을 통해 컴퓨터에 의해 헤파티티스 B 바이러스의 아미노산 변이를 검출하게 된다. 본 발명은 헤파티티스 B 바이러스의 아미노산 변이를 검출하는 것이 목적이므로, 상기 염기 서열 정보는 open reading frame(ORF)에 따라 아미노산 서열 정보로 변환하게 된다.
- [0033] 이후, 상기 방법은 상기 변환된 아미노산 서열 정보를 레퍼런스 아미노산 서열과 비교하여 서로 상이한 아미노산 서열 정보를 검색하는 단계를 포함한다.
- [0034] 레퍼런스(reference) 아미노산 서열은 검출 대상 시료에 포함된 헤파티티스 B 바이러스 전체 또는 일부 단백질(예를 들어, 역전사 효소)의 아미노산 서열과 비교 대상이 되는 표준이 되는 아미노산 서열로, 예를 들어, 야생형 헤파티티스 B 바이러스의 전체 또는 일부 단백질의 아미노산 서열이 될 수 있다. 본 단계에서, 레퍼런스 아미노산 서열과 상기 변환된 아미노산 서열 정보를 비교하여 서로 상이한 아미노산이 검색되는 경우, 이는 아미노산 변이가 존재하는 것으로 판단하게 된다. 이러한 정보들을 수집하여 마지막으로, 상기 검색된 정보를 사용자에게 출력하게 된다.
- [0035] 다른 양상은 상기 b) 내지 d) 단계의 방법을 컴퓨터에서 실행시키기 위한 프로그램으로 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공한다.
- [0036] 상기 b) 내지 d) 단계의 방법은 컴퓨터에서 실행될 수 있는 프로그램으로 작성 가능하고, 컴퓨터로 읽을 수 있는 기록매체를 이용하여 상기 프로그램을 동작시키는 범용 디지털 컴퓨터에서 구현될 수 있다. 또한, 상기 프로그램은 컴퓨터로 읽을 수 있는 기록매체에 여러 수단을 통하여 기록될 수 있다. 상기 컴퓨터로 읽을 수 있는 기록매체는 자기 저장매체(예를 들면, 롬, 플로피 디스크, 하드 디스크 등), 광학적 판독 매체(예를 들면, 시디롬, 디브이디 등)와 같은 저장매체를 포함하나, 이에 한정하지는 않는다.
- [0037] 또 다른 양상은 상기 a) 단계의 프라이머 세트; 및 상기 기록매체로 이루어진 헤파티티스 B 바이러스의 아미노산 변이를 검출하기 위한 키트를 제공한다.

[0038] 상기 키트는 상기 프라이머 세트와 기록매체를 패키지로 제작하여, 상기 에멀전 기반 클론 PCR을 수행할 수 있는 시퀀싱 장치(예를 들어, Roche사의 게놈 Sequencer FLX(GS FLX, 454) 시스템)를 사용하는 사용자에게 판매될 수 있다.

발명의 효과

[0039] 일 구체예에 따른 헤파티티스 B 바이러스의 아미노산 변이를 검출하는 키트 및 방법에 의하면, 대상 시료로부터 헤파티티스 B 바이러스의 아미노산 변이를 용이하고 정확하게 검출할 수 있으며, 특히, 헤파티티스 B 바이러스의 약제 내성 부위를 분석하는데 효과적이다.

도면의 간단한 설명

[0040] 도 1은 일 구체예에 따른 프라이머의 구성 요소를 나타내는 개략도이다.

도 2는 일 구체예에 따른 키트를 사용하여 나타난 Adefovir, Entecavir, Lamivudine, Telbivudine에 대한 약제 내성도를 나타낸 것이다.

도 3은 일 구체예에 따른 키트의 프로그램에서 상세 보기를 선택한 경우 각 약제 내성의 각 변이주의 개수를 확인한 결과이다.

도 4는 일 구체예에 따른 키트의 프로그램에서 레퍼런스 아미노산 시퀀스의 역전사 효소 단백질의 아미노산 시퀀스를 기준으로 이와 다른 부분이 존재하는 아미노산 시퀀스의 부분에 대해 그 다른 부분의 비율을 표시하여 1%가 넘는 SNP들에 대해서 표시한 결과이다.

도 5는 일 구체예에 따른 키트의 프로그램에서 nuc=n로 설정한 경우를 나타낸 결과이다.

도 6은 일 구체예에 따른 키트의 프로그램에서 nuc=y로 설정한 경우를 나타낸 결과이다.

도 7은 일 구체예에 따른 키트의 프로그램에서 Variant analyzer에 의해서 alignment에 성공된 read들만을 대상으로 각 read의 길이의 분포도의 예를 나타낸 것으로, 가로축은 read의 길이를 의미하고 세로축은 각 read의 개수를 의미하며, 가로축의 설명에는 평균 길이를 표시한 것이다.

도 8은 일 구체예에 따른 키트의 프로그램에서 각 read의 alignment score의 분포도의 예를 나타낸 것으로, 가로축은 alignment score를 의미하고, 세로축은 각 read의 개수를 의미한다.

도 9는 일 구체예에 따른 키트의 프로그램에서 HBV 게놈 상 각 read가 alignment된 부분의 분포도의 예를 나타낸 것으로, 가로축은 HBV 게놈 상의 염기의 위치를 의미하고 세로축은 각 위치의 염기를 포함하는 read의 개수를 의미한다.

도 10은 일 구체예에 따른 키트의 프로그램에서 HCC와 연관이 있다고 알려진 부위에 대해서 해당 부분의 변이 여부를 표시한 예로, HBV 게놈상의 염기 시퀀스 부위와 reference 염기, 각 염기를 포함하는 read의 수와 그 분포, 전체 read의 수, 해당 부위의 설명을 표시하였다.

발명을 실시하기 위한 구체적인 내용

[0041] 이하 하나 이상의 구체예를 실시예를 통하여 보다 상세하게 설명한다. 그러나, 이들 실시예는 하나 이상의 구체예를 예시적으로 설명하기 위한 것으로 본 발명의 범위가 이들 실시예에 한정되는 것은 아니다.

[0042] 실시예 1: 샘플 구분을 위한 프라이머 및 태그(MID) 제작

[0043] 한번에 여러 샘플을 동시에 genotyping 하기 위하여, 증폭에 사용되는 표적 특이적 프라이머, 시퀀싱에 필요한 프라이머 A/B 및 샘플 사이에 식별 가능한 태그(multiplex identifier, MID)를 모두 포함하는 프라이머 세트를 제작하였다. 상기 프라이머의 개략도는 도 1과 같다.

[0044] 프라이머 A/B는 Roche 사의 게놈 Sequencer FLX(GS FLX, 454)의 사용시에 제공된 프로토콜에 따라 하기 표 1과 같은 염기 서열을 갖도록 제작하였으며, 프라이머 A를 정방향 프라이머로, 프라이머 B를 역방향 프라이머로 구분하여 사용하였다. 태그(MID)는 총 151종을 제작하였으며, 상기 태그를 사용하면, 한번에 151명의 샘플을 동시에 실험할 수 있다. 151종의 태그에 해당하는 염기 서열을 하기 표 2에 나타내었다.

표 1

[0045]

프라이머 종류	염기 서열
프라이머 A	CCATCTCATCCCTGCGTGTCTCCGACTCAG(서열 번호 1)
프라이머 B	CCTATCCCCTGTGTGCCTTGGCAGTCTCAGTCAG(서열 번호 2)

표 2

[0046]

태그(MID) 종류	염기 서열
MID-1	ACGAGTGCGT(서열 번호 3)
MID-2	ACGCTCGACA(서열 번호 4)
MID-3	AGACGCACTC(서열 번호 5)
MID-4	AGCACTGTAG(서열 번호 6)
MID-5	ATCAGACACG(서열 번호 7)
MID-6	ATATCGCGAG(서열 번호 8)
MID-7	CGTGTCTCTA(서열 번호 9)
MID-8	CTCGCGTGTG(서열 번호 10)
MID-10	TCTCTATGCG(서열 번호 11)
MID-11	TGATACGTCT(서열 번호 12)
MID-13	CATAGTAGTG(서열 번호 13)
MID-14	CGAGAGATAC(서열 번호 14)
MID-15	ATACGACGTA(서열 번호 15)
MID-16	TCACGTACTA(서열 번호 16)
MID-17	CGTCTAGTAC(서열 번호 17)
MID-18	TCTACGTAGC(서열 번호 18)
MID-19	TGTACTACTC(서열 번호 19)
MID-20	ACGACTACAG(서열 번호 20)
MID-21	CGTAGACTAG(서열 번호 21)
MID-22	TACGAGTATG(서열 번호 22)
MID-23	TACTCTCGTG(서열 번호 23)
MID-24	TAGAGACGAG(서열 번호 24)
MID-25	TCGTCGCTCG(서열 번호 25)
MID-26	ACATACGCGT(서열 번호 26)
MID-27	ACGCGAGTAT(서열 번호 27)
MID-28	ACTACTATGT(서열 번호 28)
MID-29	ACTGTACAGT(서열 번호 29)
MID-30	AGACTATACT(서열 번호 30)
MID-31	AGCGTCGTCT(서열 번호 31)
MID-32	AGTACGCTAT(서열 번호 32)
MID-33	ATAGAGTACT(서열 번호 33)
MID-34	CACGCTACGT(서열 번호 34)
MID-35	CAGTAGACGT(서열 번호 35)
MID-36	CGACGTGACT(서열 번호 36)
MID-37	TACACACACT(서열 번호 37)
MID-38	TACACGTGAT(서열 번호 38)
MID-39	TACAGATCGT(서열 번호 39)
MID-40	TACGCTGTCT(서열 번호 40)
MID-41	TAGTGTAGAT(서열 번호 41)
MID-42	TCGATCACGT(서열 번호 42)
MID-43	TCGCACTAGT(서열 번호 43)
MID-44	TCTAGCGACT(서열 번호 44)
MID-45	TCTATACTAT(서열 번호 45)
MID-46	TGACGTATGT(서열 번호 46)
MID-47	TGTGAGTAGT(서열 번호 47)
MID-48	ACAGTATATA(서열 번호 48)
MID-49	ACGCGATCGA(서열 번호 49)

MID-50	ACTAGCAGTA(서열 번호 50)
MID-51	AGCTCACGTA(서열 번호 51)
MID-52	AGTATACATA(서열 번호 52)
MID-53	AGTCGAGAGA(서열 번호 53)
MID-54	AGTGCTACGA(서열 번호 54)
MID-55	CGATCGTATA(서열 번호 55)
MID-56	CGCAGTACGA(서열 번호 56)
MID-57	CGCGTATACA(서열 번호 57)
MID-58	CGTACAGTCA(서열 번호 58)
MID-59	CGTACTCAGA(서열 번호 59)
MID-60	CTACGCTCTA(서열 번호 60)
MID-61	CTATAGCGTA(서열 번호 61)
MID-62	TACGTCATCA(서열 번호 62)
MID-63	TAGTCGCATA(서열 번호 63)
MID-64	ATATATACA(서열 번호 64)
MID-65	TATGCTAGTA(서열 번호 65)
MID-66	CACGCGAGA(서열 번호 66)
MID-67	TCGATAGTGA(서열 번호 67)
MID-68	TCGCTGCGTA(서열 번호 68)
MID-69	TCTGACGTCA(서열 번호 69)
MID-70	TGAGTCAGTA(서열 번호 70)
MID-71	TGTAGTGTGA(서열 번호 71)
MID-72	TGTCACACGA(서열 번호 72)
MID-73	TGTCGTCGCA(서열 번호 73)
MID-74	ACACATACGC(서열 번호 74)
MID-75	ACAGTCGTGC(서열 번호 75)
MID-76	ACATGACGAC(서열 번호 76)
MID-77	ACGACAGCTC(서열 번호 77)
MID-78	ACGTCTCATC(서열 번호 78)
MID-79	ACTCATCTAC(서열 번호 79)
MID-80	ACTCGCGCAC(서열 번호 80)
MID-81	AGAGCGTCAC(서열 번호 81)
MID-82	AGCGACTAGC(서열 번호 82)
MID-83	AGTAGTGATC(서열 번호 83)
MID-84	AGTGACACAC(서열 번호 84)
MID-85	AGTGTATGTC(서열 번호 85)
MID-86	ATAGATAGAC(서열 번호 86)
MID-87	ATATAGTCGC(서열 번호 87)
MID-88	ATCTACTGAC(서열 번호 88)
MID-89	CACGTAGATC(서열 번호 89)
MID-90	CACGTGTCGC(서열 번호 90)
MID-91	CATACTCTAC(서열 번호 91)
MID-92	CGACACTATC(서열 번호 92)
MID-93	CGAGACGCGC(서열 번호 93)
MID-94	CGTATGCGAC(서열 번호 94)
MID-95	CGTCGATCTC(서열 번호 95)
MID-96	CTACGACTGC(서열 번호 96)
MID-97	CTAGTCACTC(서열 번호 97)
MID-98	CTCTACGCTC(서열 번호 98)
MID-99	CTGTACATAC(서열 번호 99)
MID-100	TAGACTGCAC(서열 번호 100)
MID-101	TAGCGCGCGC(서열 번호 101)
MID-102	TAGCTCTATC(서열 번호 102)
MID-103	TATAGACATC(서열 번호 103)
MID-104	TATGATACGC(서열 번호 104)
MID-105	TCACTCATAC(서열 번호 105)
MID-106	TCATCGAGTC(서열 번호 106)

MID-107	TCGAGCTCTC(서열 번호 107)
MID-108	TCGCAGACAC(서열 번호 108)
MID-109	TCTGTCTCGC(서열 번호 109)
MID-110	TGAGTGACGC(서열 번호 110)
MID-111	TGATGTGTAC(서열 번호 111)
MID-112	TGCTATAGAC(서열 번호 112)
MID-113	TGCTCGCTAC(서열 번호 113)
MID-114	ACGTGCAGCG(서열 번호 114)
MID-115	ACTCACAGAG(서열 번호 115)
MID-116	AGACTCAGCG(서열 번호 116)
MID-117	AGAGAGTGTG(서열 번호 117)
MID-118	AGCTATCGCG(서열 번호 118)
MID-119	AGTCTGACTG(서열 번호 119)
MID-120	AGTGAGCTCG(서열 번호 120)
MID-121	ATAGCTCTCG(서열 번호 121)
MID-122	ATCACGTGCG(서열 번호 122)
MID-123	ATCGTAGCAG(서열 번호 123)
MID-124	ATCGTCTGTG(서열 번호 124)
MID-125	ATGTACGATG(서열 번호 125)
MID-126	ATGTGTCTAG(서열 번호 126)
MID-127	CACACGATAG(서열 번호 127)
MID-128	CACTCGCACG(서열 번호 128)
MID-129	CAGACGTCTG(서열 번호 129)
MID-130	CAGTACTGCG(서열 번호 130)
MID-131	CGACAGCGAG(서열 번호 131)
MID-132	CGATCTGTCG(서열 번호 132)
MID-133	CGCGTGCTAG(서열 번호 133)
MID-134	CGCTCGAGTG(서열 번호 134)
MID-135	CGTGATGACG(서열 번호 135)
MID-136	CTATGTACAG(서열 번호 136)
MID-137	CTCGATATAG(서열 번호 137)
MID-138	CTCGCACGCG(서열 번호 138)
MID-139	CTGCGTCACG(서열 번호 139)
MID-140	CTGTGCGTCG(서열 번호 140)
MID-141	TAGCATACTG(서열 번호 141)
MID-142	TATACATGTG(서열 번호 142)
MID-143	TATCACTCAG(서열 번호 143)
MID-144	TATCTGATAG(서열 번호 144)
MID-145	TCGTGACATG(서열 번호 145)
MID-146	TCTGATCGAG(서열 번호 146)
MID-147	TGACATCTCG(서열 번호 147)
MID-148	TGAGCTAGAG(서열 번호 148)
MID-149	TGATAGAGCG(서열 번호 149)
MID-150	TGCGTGTGCG(서열 번호 150)
MID-151	TGCTAGTCAG(서열 번호 151)
MID-152	TGTATCACAG(서열 번호 152)
MID-153	TGTGCGGTG(서열 번호 153)

[0047] 실시예 2: HBV 유전자 전체 염기 서열 분석

[0048] 실시예 2-1: HBV 유전자 전체를 증폭하기 위한 특이적 프라이머 선정 및 증폭 조건 확립

[0049] HBV의 전체 유전자를 증폭하기 위하여 표 3에 나타낸 바와 같이, 18개(9쌍)의 프라이머를 제작하였으며, 실시예 1에서 제작한 표 1 및 표 2의 프라이머 A, 프라이머 B 및 태그(MID)를 조합하여 총 2718개의 프라이머(151×18 = 2718)를 제작하였다.

표 3

프라이머 종류	프라이머 서열
HBV-S-1(re)-F	TCCAACCTGTCTCGGCTAT(서열 번호 154)
HBV-S-1(re)-R	TCGAAAACATTGCTTGAG(서열 번호 155)
HBV-S-2(re)-F	CTACCAGCACGGGACCAT(서열 번호 156)
HBV-S-2(re)-R	GCCTTGTAAGTTGGCGAGA(서열 번호 157)
HBV-new1(re)-F	AGAATTGTGGGTCTTTGG(서열 번호 158)
HBV-new1(re)-R	CAGAGGTGAAGCGAAGTG(서열 번호 159)
HBV-new2(re)-F	TCTACCGTCCCCTTCTTC(서열 번호 160)
HBV-new2(re)-R	CACAGAATAGCTTGCCTGA(서열 번호 161)
HBV-new3(re)-F	GGTGGGAAGTAATTTGGAA(서열 번호 162)
HBV-new3(re)-R	CCAGCCTTCCACAGAGTAT(서열 번호 163)
HBV-new4(re)-F	GGCCTATATTTTCTGCTG(서열 번호 164)
HBV-new4(re)-R	ACATAGAGGTTCCTTGAGCA(서열 번호 165)
HBV-geneS-F(re)	AACCCTGTTCGACTACTG(서열 번호 166)
HBV-geneS-R(re)	TCGAAAACATTGCTTGAG(서열 번호 167)
HBV-4-F(re)	TTGTTTAAAGACTGGGAGGA(서열 번호 168)
HBV-4-R(re)	AGTTTCGGAAGTGTGAT(서열 번호 169)
HBV-6-F(re)	CTTGACAAAGGCATTAAC(서열 번호 170)
HBV-6-R(re)	CCTCGAGAAGATTGACGAT(서열 번호 171)

상기 제작된 2718개의 프라이머를 사용하여 다음과 같은 조건으로 PCR을 수행하였다: 95℃에서 10분 동안 정치 후, 95℃ 30초, 55℃ 1분, 72℃ 1분의 조건으로 35회 반복한 후, 마지막으로 72℃에서 10분 동안 정치하였다.

실시예 2-2: HBV 전체 유전자 샘플 분석 방법 및 결과

사용된 샘플은 총 15명의 HBV 감염 환자의 혈청(서울 아산병원 소아기내과 환자)이며, 이중 11종은 기존에 직접 시퀀싱 방법(Sanger 방법)에 의해 약제 내성 부위를 genotyping을 하여 변이가 발견된 샘플이고, 4종은 변이가 발견되지 않은 샘플이었다. 상기 혈청들로부터 Qiagen 사의 QIAamp® MinElute® Virus Spin kit를 이용하여 바이러스 DNA를 분리하였다. 각 샘플당 9개의 PCR 산물을 얻었고, 이를 동일한 수의 산물이 들어가도록 혼합하였다. 혼합된 PCR 산물들은 Roche의 게놈 Sequencer FLX(GS FLX, 454)를 이용하여 염기 서열 분석을 실시하였다. PCR 산물들을 변성화시켜 단일 가닥의 DNA로 만든 후, 한 개의 마이크로비드에 한 종류의 단일 가닥 DNA가 결합하도록 한 후, emPCR을 수행하여 단일 가닥 DNA를 증폭하였다. 증폭된 emPCR 산물을 PicoTiter Plate의 웰에 하나씩 들어가게 하고, 여기에 네 종류의 뉴클레오티드를 순차적으로 흘려주어 해당하는 상보적인 뉴클레오티드가 결합될 때 발생하는 빛 에너지를 감지하여 영상을 획득하고, 획득한 영상을 분석하여 시퀀싱 결과를 얻었다. 직접 시퀀싱 방법으로 얻은 결과와 본 발명의 시퀀싱(NGS) 방법으로 얻은 15 명의 환자 혈청의 HBV 변이 분석 결과를 하기 표 4에 나타내었다.

표 4

샘플 종류	직접 시퀀싱에 의한 결과	본 발명의 시퀀싱 방법(NGS)을 이용한 결과
asan-06	L180M, S202G, M204V	L180M(99.1)*, T184A(1.3), S202G(84.4), M204V(97.2), M250V(13.8)
asan-07	L180M, T184L, M204V	L180M(97.3), A181T(1.7), T184L(97.1), M204V(95.8), M204I(1.2)
asan-08	N./D	N/D
asan-09	N./D	N/D
asan-10	N./D	N/D
asan-11	N./D	N/D
21376874	L180M, T184L, M204V	L180M(94.8), T184L(94.8), M204V(94.7)
29614619	L180M, M204I	L180M(29.3), M204I(96.4)
27643510	L180M, T184I, S202G, M204V	V173M(3.1), L180M(99.2), T184I(69.9), T184L(2.0), S202G(97.2), M204V(97.8), M250I(1.0)

18258761	L180M, S202G, M204V	V173M(4.0), L180M(97.9), S202G(98.5), M204V(98.0), M204I(1.1)
21923326	L180M, M204V	V173L(98.7), L180M(98.9), M204V(99.1)
29907157	L180M, T184L, M204V	V173M(1.3), L180M(98.3), A181T(1.4), T184L(93.2), T184F(2.7), M204V(98.1)
27455427	L180M, M204V	V173L(99.4), L180M(99.2), M204V(98.1), M250V(3.9)
36068214	M204I	I169T(1.4), L180M(3.6), L180Q(2.2), A181S(4.6), A181T(5.0), T184I(1.2), S202N(1.1), M204I(97.7), M204V(1.8), M250I(1.3)
26299220	L180M, M204V	V173L(23.8), L180M(42.6), A181V(15.3), A181S(9.4), M204V(41.6), M204I(12.3), N236T(21.6)

[0055] * ()의 숫자는 변이가 전체 집단에서 차지하는 비율 %를 의미한다.

[0056] 두 방법을 비교한 결과 직접 시퀀싱에 의해 판별된 결과는 100% 모두 NGS를 이용한 결과에 포함되어 있었으며, 직접 시퀀싱 방법에 의해 검출하지 못한 변이들도 다수 검출됨을 확인할 수 있었다. 뿐만 아니라, 전체 HBV 집단 중에서 각 변이를 갖는 변이종들의 분포율(%)을 1 % 미만까지 검출할 수 있었다. 약제 내성 부위 이외의 전체 유전자에서도 많은 변이들을 관찰할 수 있었다. 이러한 결과를 통해 HBV의 유전체의 특성 및 기능 연구에 큰 기여를 할 수 있으며, 현재는 밝혀 지지 않은 약제 내성 부위를 밝혀내는 데에도 도움이 될 수 있다. 또한, 이러한 결과는 HBV의 치료제 개발에도 응용될 수 있다.

[0057] 실시예 3: HBV 유전자 약제 내성 부위(RT, YMDD)의 염기 서열 분석

[0058] 실시예 3-1: HBV 유전자 약제 내성 부위 (RT, YMDD) 증폭하기 위한 특이적 프라이머 선정 및 증폭 조건 확립

[0059] HBV의 약제 내성 부위(RT, YMDD) 유전자를 증폭하기 위하여 표 5에 나타난 바와 같이, 3개(4쌍, 역방향 프라이머는 공통으로 사용함)의 프라이머를 제작하였으며, 실시예 1에서 제작한 표 1 및 표 2의 프라이머 A, 프라이머 B 및 태그(MID)를 조합하여 총 453개의 프라이머(151×3 = 2718)를 제작하였다.

표 5

[0060]

프라이머 종류	프라이머 서열
HBV_459F	TATGTTGCCCGTTTGTCTC(서열 번호 172)
HBV_574F	AAAACCTTCGGAMGGAAGT(서열 번호 173)
HBV_1031R	GTAAAAGGGGACGAAAGC(서열 번호 174)

[0061] 상기 제작된 453개의 프라이머를 사용하여 다음과 같은 조건으로 PCR을 수행하였다: 95℃에서 10분 동안 정치 후, 95℃ 30초, 55℃ 1분, 72℃ 1분의 조건으로 35회 반복한 후, 마지막으로 72℃에서 10분 동안 정치하였다.

[0062] 실시예 3-2: HBV 전체 유전자 샘플 분석 방법 및 결과

[0063] 사용된 샘플은 총 49명의 HBV 감염 환자의 혈청(서울 아산병원 소아기내과 환자)이며, 이중 26종은 기존에 직접 시퀀싱 방법(Sanger 방법)에 의해 약제 내성 부위를 genotyping을 하여 변이가 발견된 샘플이고, 7종은 변이가 발견되지 않은 샘플이었으며, 나머지 16 종은 genotyping을 실시하지 않은 샘플이었다. 상기 혈청들로부터 Qiagen 사의 QIAamp® MinElute® Virus Spin kit를 이용하여 바이러스 DNA를 분리하였다. 각 샘플당 9개의 PCR 산물을 얻었고, 이를 동일한 수의 산물이 들어가도록 혼합하였다. 혼합된 PCR 산물들은 Roche의 계놈 Sequencer FLX(GS FLX, 454)를 이용하여 염기 서열 분석을 실시하였다. PCR 산물들을 변성화시켜 단일 가닥의 DNA로 만든 후, 한 개의 마이크로비드에 한 종류의 단일 가닥 DNA가 결합하도록 한 후, emPCR을 수행하여 단일 가닥 DNA를 증폭하였다. 증폭된 emPCR 산물을 PicoTiter Plate의 웰에 하나씩 들어가게 하고, 여기에 네 종류의 뉴클레오티드를 순차적으로 흘려주어 해당하는 상보적인 뉴클레오티드가 결합될 때 발생하는 빛 에너지를 감지하여 영상을 획득하고, 획득한 영상을 분석하여 시퀀싱 결과를 얻었다. 직접 시퀀싱 방법으로 얻은 결과와 본 발명의 시퀀싱(NGS) 방법으로 얻은 49 명의 환자 혈청의 HBV 변이 분석 결과를 하기 표 6에 나타내었다.

표 6

[0064]

샘플 종류	직접 시퀀싱에 의한 결과	본 발명의 시퀀싱 방법(NGS)을 이용한 결과
DS-1	I169T, L180M, T184A, M204V	I169T(68.9), I169M(5.6), V173M(1.3), L180M(97.7), T184A(97.3), M204V(97.1)
DS-2	L180M, 204(I, V)	L180M(39.9), A181T(1.4), M204I(59.0), M204V(39.3)
DS-3	L180M, L184(L, I), S202G, M204V	L180M(93.7), L180I(1.2), A181V(2.5), A181T(1.7), T184L(38.5), A194T(1.6), S202G(51.8), M204V(90.6), M204I(3.5), N236T(2.1), M250I(2.4)
DS-4	L180M, L184(L, I), S202G, M204V	I169V(3.1), V173L(36.9), L180M(97.1), L180I(2.1), A181T(2.4), T184L(63.5), T184A(4.2), A194T(2.4), S202G(23.4), S202N(1.6), M204V(95.0), M204I(3.2), M250I(6.4)
DS-5	L180M, L184I, S202G, M204V	I169T(3.6), V173M(16.1), L180M(98.1), L180I(1.2), A181G(65.5), A181S(2.2), T184L(13.4), T184I(1.9), A194T(3.3), S202G(77.0), S202N(1.4), M204V(98.2), M250L(4.6), M250I(4.4)
DS-6	L180M, L184I, S202G, M204V	I169T(2.1), V173M(9.7), L180M(97.1), L180I(1.6), A181G(70.0), T184L(15.4), T184I(1.6), A194T(2.3), S202G(74.1), M204V(97.2), M204I(1.1), M250L(4.7), M250I(1.7)
DS-7	L180M, M204I	V173M(2.9), L180M(21.4), A181T(3.9), M204I(80.7),
DS-8	L180M, S202G, M204V	V173M(1.2), L180M(97.2), L180I(1.5), A181T(2.8), T184I(2.6), A194T(2.8), S202G(94.3), M204V(93.3), M204I(1.8), M250I(5.5)
DS-9	L180M, S202G, M204V	V173M(1.1), L180M(99.3), T184I(1.7), S202G(98.4), M204V(97.5), M250I(1.2)
DS-10	L180M, S202G, M204V	L180M(99.1), S202G(99.6), M204V(99.1)
DS-11	L180M, S202G, M204V	L180M(98.6), A181T(1.8), T184L(7.1), S202G(86.7), M204V(97.7), M250I(2.0)
DS-12	L180M, S202G, M204V	L180M(99.4), S202G(99.8), M204V(97.5)
DS-13	L180M, S202G, M204V	L180M(98.5), S202G(96.9), M204V(95.1), M204I(1.5), M250I(2.2)
DS-14	L180M, T184A, M204V	V173M(1.6), L180M(98.9), T184A(95.8), T184S(3.9), M204V(99.2)
DS-15	L180M, A181T, T184S, M204I	L180M(71), A181T(12.7), T184S(66.7), 184N(3.7), M204I(87.5), M204V(11.9),
DS-16	L180M, A181V, T184S, M204V	V173M(5.1), L180M(97.3), A181V(1.5), T184S(93.5), T184A(4.6), M204V(97.1),
DS-17	L180M, M204I	L180M(17.5), A181T(13), M204I(40.7),
DS-18	L180M, M204V	L180M(89.9), T184L(5.8), M204V(83.3), M204I(6.4), N236T(8.4),
DS-19	L180M, S202G, M204V	L180M(99.4), A181G(8.3), T184L(2.1), S202G(85.8), M204V(98),
DS-20	L180M, S202G, M204V	173L(2.9), L180M(98.4), A181G(28.6), T184L(3.5), S202G(68.4), M204V(99.6),
DS-21	L180M, S202G, M204V, I	V173M(2.4), L180M(33.2), S202G(26.3), M204I(60.6), M204V(33.3),
DS-22	L180M, T184I, S202G, M204I/V	V173M(4.8), L180M(86.2), T184L(10.8), S202G(69.3), M204V(85.8), M204I(13.6),
DS-23	L180M, T184I, S202G, M204V	V173M(33.1), L180M(99.2), A194T(33.9), S202G(99.9), M204V(99.4),
DS-24	L180M, T184L, M204V	L180M(100), T184L(100), M204V(100),
DS-25	L180M, T184S, F, A, I, M204V	169T(22.3), L180M(99.9), A181V(1.4), T184S(47.2), T184A(23.4), T184F(15.3), T184L(13.6), M204V(99.9),
DS-26	M204I	L180M(1.2), M204I(15.1),
DS-27	Negative	A181T(3.9), A194T(1.9), S202N(3.1), M204I(6.1), M250I(4.4)
DS-28	Negative	
DS-29	Negative	I169L(3.0), A181T(10.4), A194T(1.0), M204I(3.2), M250I(2.5)
DS-30	Negative	
DS-31	Negative	
DS-32	Negative	
DS-33	Negative	250V(1.3),
DS-34		L180M(98.7), T184S(99.3), M204V(99.3),
DS-35		L180M(99.1), S202G(99.2), M204V(98.1),
DS-36		L180M(85.6), A181T(98), M204I(90.2),
DS-37		173L(2.7), L180M(37.2), A181T(43.7), M204V(37),

DS-38		L180M(39.3), A181V(59.6), S202G(12.5), M204V(37.8), M204I(2), N236T(2.1),
DS-39		L180M(99.1), M204I(99.4),
DS-40		V173M(1.4), L180M(98.8), T184I(2.4), T184A(1.5), S202G(97.7), M204V(98.2),
DS-41		V173M(3.5), L180M(97.6), T184F(97.6), A194T(1.2), 202N(1.3), M204V(97.4),
DS-42		
DS-43		173L(97.2), L180M(96.3), 181S(2.9), T184S(93.9), T184I(2.1), M204I(99.1),
DS-44		L180M(93.6), A181T(4.3), T184L(90.9), 184M(2.3), M204V(91.2),
DS-45		173L(67.5), L180M(66.6), A181T(3.9), 184N(1), M204V(64.2), M204I(31.6),
DS-46		169T(6.8), V173M(1.4), L180M(98.9), T184A(75.2), , L(23.6), M204V(98.6),
DS-47		L180M(96.2), T184S(33.7), S202G(20.1), M204V(83.7), M204I(12.6),
DS-48		A181T(1.6), M204I(1.2),
DS-49		L180M(12.5), A194T(3), S202G(9.5), M204V(8.9), M204I(1.4),

[0065] * ()의 숫자는 변이가 전체 집단에서 차지하는 비율 %를 의미한다.

[0066] 직접 시퀀싱 방법으로 변이가 발견되었던 26종의 샘플 중 3종의 샘플이 본 발명의 시퀀싱(NGS) 결과의 결과가 상이했으나, 직접 시퀀싱 방법으로 다시 실험한 결과 혼재되어 있는 HBV 종을 선정하는 과정에서의 주관적인 판단으로 인한 잘못된 선정이었음이 밝혀졌다. 또한 직접 시퀀싱 방법은 여러 종이 섞여 있는 전체 HBV 바이러스 풀 중 20% 이하로 존재하는 HBV 변이종의 경우 판정이 불가하여, 약제 내성인 HBV가 없다고(상기 표에서 negative) 판정된 샘플의 경우에도 미량의 약제 내성인 HBV가 존재함을 확인할 수 있었다.

[0067] 상기의 결과로 볼 때, 본 발명의 시퀀싱 방법(NGS)을 이용한 HBV 검출 방법은 HBV의 약제 내성 변이종의 분포를 1% 이내까지 판별해 낼 수 있음을 확인할 수 있었다.

[0068] 실시예 4: NGS를 이용한 약제 내성 HBV 검사 방법의 결과 분석 소프트웨어

[0069] 변이 분석 모듈

[0070] 변이 분석 모듈은 Python으로 작성되었고 변이 분석 모듈 본체와 시퀀스 라이브러리로 이루어져 있다. 이 모듈은 fasta 포맷으로 된 시퀀스 파일을 읽어 변이를 분석하고 분석된 변이를 여러 분석 결과 파일들에 저장하는 기능을 수행한다. 설정 파일을 이용하여 분석의 종류와 결과 파일들의 이름 등을 지정한다.

[0071] 변이 분석 모듈의 실행 방법

[0072] Usage: position.py [conf] [in]

[0073] where:

[0074] conf: position configuration file

[0075] in: input read file in fasta format

[0076] 변이 분석 모듈을 실행하기 위해서는 fasta 포맷으로 된 입력 시퀀스 파일이 필요하고 각 설정을 지정하는 설정 파일이 필요하다.

[0077] 변이 분석 모듈의 설정

[0078] 설정 파일에서 설정 가능한 각 설정 값들은 다음과 같다.

[0079] *reference_name*

- [0080] 레퍼런스 시퀀스의 이름을 지정한다. 입력 파일에서 이 이름을 가지는 시퀀스는 입력 시퀀스로 인식되지 않고 레퍼런스 시퀀스로 인식된다. `reference_name` 을 "first" 로 지정할 경우 "first"의 이름을 가지는 시퀀스가 레퍼런스 시퀀스로 인식되는 것이 아니라 입력 파일의 첫 번째 시퀀스가 이름에 상관 없이 레퍼런스 시퀀스로 인식된다.
- [0081] ***reference_length***
- [0082] 레퍼런스 시퀀스의 길이를 지정한다. 불러들인 레퍼런스 시퀀스의 실제 길이와 비교하여 서로 다를 경우 오류를 출력한다. 레퍼런스 시퀀스의 오류 검출 목적으로 지정한다.
- [0083] ***position_alias***
- [0084] 변이를 분석할 염기 위치를 기억하기 쉽도록 각각 이름을 붙인다. `name:position` 의 형식으로 지정한다. `position`의 위치를 `name` 으로 부른다. 여러 개의 `position_alias` 옵션을 지정하는 것이 가능하다.
- [0085] ***position***
- [0086] 변이를 분석할 위치의 목록을 지정한다. 분석할 위치들은 `reference` 시퀀스의 염기의 위치를 나타낸다. 위치의 목록에는 `position_alias`에서 정의한 `alias name` 이 사용된다. 한 `position` 옵션에 나열한 각 `position` 들은 haplotype 분석에 사용된다. 한 `conf` 파일 내에 여러 개의 `position` 옵션을 지정하는 것이 가능하다.
- [0087] ***window_from***
- [0088] 분석 결과를 표시할 때 한 `position`에 대해서 5' 방향으로 얼마나 많은 개수의 염기를 표시할 지 그 개수를 정수로 지정한다.
- [0089] ***window_to***
- [0090] 분석 결과를 표시할 때 한 `position`에 대해서 3' 방향으로 얼마나 많은 개수의 염기를 표시할 지 그 개수를 정수로 지정한다.
- [0091] ***output_name_prefix***
- [0092] 시퀀스를 출력할 때 시퀀스의 이름 앞에 붙는 문자열이다. `fasta` 형식의 시퀀스의 경우 ">"를 사용한다.
- [0093] ***nucleotide_blank***
- [0094] 입력 시퀀스에 나타난 염기들 중 이 문자열에 포함되는 염기가 존재할 경우 해당 염기를 `blank`로 인식한다. `blank`는 해당 위치에 염기 정보가 없음을 표시하며 해당 위치에 염기가 있으나 어떤 염기인지 알 수 없다는 의미의 "N"과는 의미가 다르다.
- [0095] ***output_reference_include***
- [0096] 출력 파일에 `reference` 파일도 입력 파일로 간주하여 출력할 것인지를 "yes" 또는 "no"로 결정한다. "yes"로 지정된 경우 출력 파일에 `reference` 파일에 대한 분석 결과도 출력된다.
- [0097] ***output_format***
- [0098] 출력 파일의 형식을 지정한다. "machine"으로 지정할 경우 변이 분석 결과 표시 모듈이 불러들이는 형식으로 출력한다. "machine"으로 지정되지 않았을 경우 사람이 확인할 수 있는 텍스트 형식으로 출력한다.
- [0099] ***output_remove_indel***
- [0100] 출력 파일에서 `insertion` 과 `deletion` 을 제거하고 출력할 것인지를 결정한다. "yes" 로 지정할 경우 `insertion` 과 `deletion` 이 제거되어 출력된다.
- [0101] ***output_position_file***
- [0102] 분석 결과 출력 파일 이름을 지정한다. 이 이름은 온전한 파일의 이름이 아니라 입력 파일의 이름 뒤에 붙이는 접미 문자열이다.
- [0103] ***output_stat***
- [0104] 분석 결과 출력 시 통계 파일도 출력할 것인지를 결정한다. "yes" 로 지정하면 통계 파일들도 출력된다.

- [0105] *output_stat_coverage_file*
- [0106] 통계 파일 출력이 지정된 경우 coverage 파일의 접미 문자열이다.
- [0107] *output_stat_length_file*
- [0108] 통계 파일 출력이 지정된 경우 length 파일의 접미 문자열이다.
- [0109] *output_stat_variation_file*
- [0110] 통계 파일 출력이 지정된 경우 variation 파일의 접미 문자열이다.
- [0111] **변이 분석 모듈의 각 함수 설명**
- [0112] 변이 분석 모듈은 position.py 파일에 저장되어 있고 다음은 이 모듈에 사용되는 각 함수에 대한 설명이다.
- [0113] *Align2Reference (index)*
- [0114] index로 지정된 alignment 시퀀스 상의 위치를 reference 시퀀스 상의 위치로 변경하여 되돌린다.
- [0115] *Reference2Align (index)*
- [0116] index로 지정된 reference 시퀀스 상의 위치를 alignment 시퀀스 상의 위치로 변경하여 되돌린다.
- [0117] *NucleotideDensity (position_align, sequence_align)*
- [0118] sequence_align 시퀀스에서 position_align 으로 지정된 alignment 시퀀스 상의 위치에 대하여 염기의 밀도를 구한다. 염기의 밀도는 position_align 을 기준으로 5' 방향으로 window_from 만큼, 3'방향으로 window_to 만큼의 범위에 대해 뉴클레오티드_blank 옵션에서 지정된 염기가 아닌 염기의 수로 정의한다.
- [0119] *Reference2 Align For Sequence (position_reference, sequence_align)*
- [0120] position_reference 로 지정된 reference 시퀀스 상의 위치를 sequence_align 시퀀스를 참고하여 alignment 시퀀스 상의 위치로 변경한다. HBV의 circular 게놈의 특성으로 인하여 한 개의 reference 시퀀스 상의 위치는 여러 개의 alignment 시퀀스 상의 위치를 가질 수 있고 이 후보 위치들 중 sequence_align시퀀스에서 가장 염기 밀도가 높은 위치를 선택하여 되돌린다.
- [0121] *stat_init ()*
- [0122] 각종 통계 자료를 초기화한다.
- [0123] *stat_write (file_in)*
- [0124] 지금까지 축적된 통계 자료를 file_in 에 지정된 파일에 출력한다. length, coverage, score, variation 에 해당하는 통계 자료를 file_in의 뒤에 output_stat_length_file, output_stat_coverage_file, output_stat_score_file, output_stat_variation_file 의 접미어를 붙인 파일에 각각 저장한다.
- [0125] *stat_add (name, seq)*
- [0126] 통계 기능이 활성화된 경우 name 과 seq로 지정된 시퀀스로부터 길이와 alignment score, 게놈상에서 read가 존재하는 부위 등의 통계 자료를 추출하여 기존 통계 자료에 추가한다. 통계 자료는 read 들의 개수, coverage, length, variation count 등을 말한다.
- [0127] *FillReverse ()*
- [0128] 계산 도중 reference 위치를 align 위치로 변경하고 align 위치를 reference 위치로 변경해야 할 필요성이 많이 있으므로 FillReverse 함수에서는 모든 시퀀스에 대해서 align 위치를 reference 위치로 변경할 수 있도록 미리 계산해 놓는다. 계산 결과는 g_list_position_align[] 리스트에 저장된다.
- [0129] *ReadConfigFile (file_conf)*
- [0130] file_conf에 지정된 파일로부터 설정사항을 불러들이는 함수이다. 설정 파일은 한 줄에 하나의 설정이 가능하고 "name=value"의 쌍으로 한 줄을 구성한다. 설정 파일의 빈 줄과 첫 칸에 "#"으로 표시된 줄은 무시된다.

- [0131] *get_sequence_name_option_list (name)*
- [0132] Variant Analyzer에서 alignment 된 시퀀스 파일의 경우에는 alignment 정보를 시퀀스의 name 에 실어서 출력한다 이 정보를 name의 문자열로부터 해석하여 읽어들인다. name 문자열에는 공백으로 분리된 "name=value" 형식으로 저장되어 있으며 이를 분석하여 리스트를 작성하여 되돌린다.
- [0133] *SequenceValidate (name, seq)*
- [0134] name 의 이름과 seq 의 시퀀스를 갖는 시퀀스가 유효한 시퀀스인지를 검사하는 함수이다. 유효한 시퀀스란 시퀀스의 길이가 reference 시퀀스와 동일하고 name 이 유효한 규칙을 갖고 있는 시퀀스이다. 이 함수에서는 name의 refStart 옵션이 존재할 경우 이에 따라 시퀀스의 시작 부분의 위치를 reference 시퀀스의 refStart 위치부터 시작하도록 조정하여 저장한다.
- [0135] *SequencePosition (name, seq)*
- [0136] name 과 seq 로 지정된 시퀀스로부터 각 위치에 존재하는 염기 시퀀스 정보를 추출한다. 추출해야 하는 위치 정보는 설정 파일의 position 옵션으로 지정한다. 설정 파일의 position 옵션에 나열되어 있는 position 목록의 각 position 들의 주변부에 대해 염기 시퀀스와 아미노산 시퀀스 정보를 추출한다. position들의 주변부의 경계는 window_from 옵션과 window_to 옵션으로 설정한다.
- [0137] *write_position_file (file_in)*
- [0138] SequencePosition() 에서 추출된 변이 정보들을 file_in 에서 지정된 파일 이름의 뒤에 output_position_file 옵션 스트링을 붙인 새로운 파일에 저장한다. 파일의 첫 줄에는 reference position 정보를 출력하고 두번째 줄부터는 각 줄마다 각 시퀀스들의 변이 정보를 출력한다.
- [0139] 시퀀스 처리 라이브러리의 각 함수 설명
- [0140] 시퀀스 처리 라이브러리는 sequence.py 파일에 저장되어 있고 시퀀스 처리에 대한 공통적인 사항을 가지고 있다. 본 라이브러리는 변이 분석 모듈에서 사용된다.
- [0141] *append_sequence_to_file (name, seq, file)*
- [0142] 시퀀스를 fasta 형식의 file 에 추가한다. 시퀀스의 이름은 name 으로 하고 시퀀스의 내용은 seq 로 한다.
- [0143] *align_file (file_in, file_out)*
- [0144] 외부 alignment 프로그램을 사용하여 file_in 파일에 들어 있는 시퀀스들을 align 하여 file_out 파일에 저장한다.
- [0145] *read_file (file_in, callback)*
- [0146] file_in 파일에 저장되어 있는 fasta 형식의 시퀀스 파일을 읽어들인다. fasta 형식에 맞는 시퀀스를 읽었을 경우 한 개의 시퀀스를 읽어들이기 때 마다 callback 함수를 호출한다.
- [0147] *read_qfile (file_in, callback)*
- [0148] file_in 파일에 저장되어 있는 fasta 형식의 quality 파일을 읽어들인다. 형식에 맞는 quality 를 읽었을 경우 한 개의 quality 를 읽어들이기 때 마다 callback 함수를 호출한다.
- [0149] *reverse_complement (seq)*
- [0150] seq로 지정된 시퀀스를 reverse complement 하여 되돌린다. "ACGT" 중 하나는 "TGCA"중 하나로 변경하고 "N", "-", "." 시퀀스는 그대로 유지하고 이외의 시퀀스는 "N"으로 치환한다.
- [0151] *reverse_complement_file (file_in, file_out)*
- [0152] file_in 파일에 저장되어 있는 시퀀스를 읽어 reverse complement 하여 file_out 파일에 저장한다. 각 시퀀스를 reverse complement 하는 방법은 reverse_complement () 함수와 같다.
- [0153] *reverse_complement_file_callback (name, seq)*
- [0154] reverse_complement_file () 에서 사용되는 함수이다. 한 개의 시퀀스를 읽을 때 마다 실행된다. name과 seq

로 지정된 시퀀스를 `reverse_complement()` 함수를 이용하여 처리한다.

[0155] ***translate3 (seq)***

[0156] `seq`의 길이가 3인 경우에만 `seq`에 들어 있는 염기 시퀀스를 아미노산 문자로 변경한다. 변경된 amino acid는 return value에 지정된다. `seq`의 길이가 3이 아닌 경우에는 "*"를 되돌리고 번역이 불가능한 염기 시퀀스인 경우에는 "@"를 되돌린다. 번역 종료에 해당하는 염기 시퀀스인 경우에는 ")"를 되돌린다.

[0157] ***translate (seq)***

[0158] `seq`로 지정된 염기 시퀀스에 해당하는 아미노산 시퀀스로 번역을 실행한다. `translate3()`을 사용하여 각 3개의 염기들을 번역한다.

[0159] **변이 분석 결과 표시 모듈**

[0160] 변이 분석 결과 표시 모듈은 php로 작성되어 있으며 php 파일에 웹 브라우저로 접속하는 것으로 모듈을 실행한다. 본 모듈의 입력으로는 샘플의 이름을 지정하고 표시의 상세도를 지정할 수 있다.

[0161] 변이 분석 결과 표시 모듈의 설정

[0162] 변이 분석 결과 표시 모듈의 설정 값들은 php 파일을 호출할 때 환경 변수로 지정된다. 웹 브라우저에서 php 파일을 호출할 때 URL 인자 목록에 지정하여 설정 가능하다. 지정 가능한 설정 값들은 다음과 같다.

[0163] ***run***

[0164] 변이 분석 결과 파일들이 들어 있는 디렉토리를 지정한다. `run` 문자열에서 지정된 디렉토리에서 변이 분석 결과 파일을 검색한다. `run` 디렉토리는 본 모듈의 php 파일이 있는 디렉토리의 "jobs/" 디렉토리 안에 있는 디렉토리 중 하나이다. 분석 모듈에서 분석이 끝나면 각 분석 결과 파일들을 "jobs/[run]/" 디렉토리에 넣어 보관한다.

[0165] ***sample***

[0166] 변이를 분석할 샘플을 지정한다. 본 표시 모듈은 `run`으로 지정된 디렉토리에서 `sample`에 해당하는 분석 샘플을 불러들여 결과를 표시한다.

[0167] ***diff***

[0168] 각 read에 있는 각 아미노산과 염기의 정보를 있는 있는 그대로 보여줄 것인지 아니면 reference와 같으면 "."으로 표시할 것인지를 결정한다. "y"인 경우 reference와 같으면 "."으로 표시하고 그렇지 않으면 각 아미노산과 염기의 정보를 보여준다.

[0169] ***nuc***

[0170] 아미노산 정보 이외에 염기 정보를 보여줄 것인지를 지정한다. "y"인 경우 염기의 정보까지 보여주고 "y"가 아닌 경우 염기의 정보를 보여주지 않는다.

[0171] ***partial***

[0172] 약제 내성 부위 전체를 모두 포함하는 read만을 분석에 포함시킬 것인지 아니면 한 부위라도 포함되는 read까지 분석에 포함시킬 것인지를 지정한다. "y"인 경우 일부라도 약제 내성 부위를 포함하는 read를 분석에 포함시킨다.

[0173] 변이 분석 결과 표시 화면의 각 부분 해설

[0174] ***Drug Resistance Rates***

[0175] Drug Resistance Rates는 내성의 정도를 표시한다. 전체 read 수를 N, 내성을 갖는 read 수를 C라고 하면 Drug Resistance Rate는 C/N으로 계산된다. 대표적인 4개의 약제인 Adefovir, Entecavir, Lamivudine, Telbivudine에 대해 각각의 내성도를 표시하고 이를 그래프로 나타낸다. 이의 예는 도 2에 나타내었다.

[0176] 다음 약제 내성 변이 조합에서 언급하는 위치는 모두 역전사 효소의 아미노산 시퀀스상의 위치이고 위치 다음에 붙은 알파벳은 해당 위치에 존재하는 아미노산이다. 한 위치에 아미노산이 여러 개 존재하여 "/"로 구분된 경우 해당 위치에 여러 개의 아미노산 중 한 개를 의미한다.

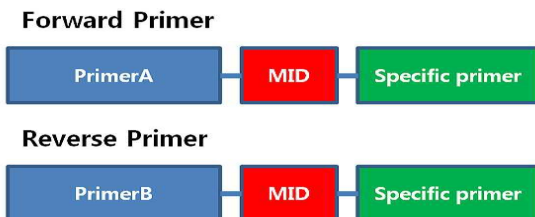
- [0177] Adefovir에 대한 약제 내성은 181(V/T) 또는 236T가 존재하는 경우 내성이 존재하는 것으로 판단한다.
- [0178] Entecavir에 대한 약제 내성은 204(I/V)가 존재하고 169I, 184T, 202S, 250M 중 하나 이상 존재하는 경우와 180M이 존재하고 169I, 184T, 202S, 250M 중 하나 이상 존재하는 경우 내성이 존재하는 것으로 판단한다.
- [0179] Lamivudine에 대한 약제 내성은 204(I/V)가 존재하는 경우 내성이 존재하는 것으로 판단한다.
- [0180] Telbivudine에 대한 약제 내성은 204I가 존재하는 경우이거나 204V와 180M이 존재하는 경우 내성이 존재하는 것으로 판단한다.
- [0181] 도 3은 상세 보기를 선택한 경우 각 약제 내성의 각 변이주의 개수를 확인할 수 있는 예의 그림이다.
- [0182] **SNPs**
- [0183] Reference 아미노산 시퀀스의 reverse transcriptase 단백질의 아미노산 시퀀스를 기준으로 이와 다른 부분이 존재하는 아미노산 시퀀스의 부분에 대해 그 다른 부분의 비율을 표시한다. 아미노산 시퀀스상의 특정 위치를 포함하는 read를 "covering read"라 하고 reference 시퀀스와 다른 아미노산을 갖는 read를 mutant read라고 한다. Variation의 위치는 "X#Y"로 표시하고 X는 reference 시퀀스의 아미노산, #는 reference 시퀀스상의 아미노산의 위치, Y는 변경된 아미노산을 의미한다. SNP의 비율은 mutant read의 수를 covering read로 나눈 값으로 한다. 이 SNP 비율이 1%가 넘는 SNP들에 대해서만 결과 화면에 표시한다. 도 4는 상기 SNP 결과의 예이다.
- [0184] **Strands**
- [0185] HBV의 reverse transcriptase 단백질의 아미노산의 시퀀스 중 내성과 관련이 있다고 알려진 아미노산 부위에 대해서 reference 아미노산을 기준으로 이와 다른 부분이 존재하는 아미노산 시퀀스들에 대해서 각 시퀀스들을 이를 빈도순으로 정렬하여 표시한다. 내성과 관련이 있는 부위에 어떤 아미노산 변이가 있는지를 표시한다. 각 read마다 read의 개수와 각 read의 비율을 표시한다. 각 read의 비율을 전체 read의 수에 대한 비율이다. 전체 read의 수는 내성 부위를 모두 포함하는 read의 수이다. 상기 예를 도 5에 나타내었다. 상기 예는 nuc=n를 설정한 경우이다.
- [0186] 각 아미노산 시퀀스의 차이뿐만 아니라 각 염기의 시퀀스를 볼 수 있다. 상기 예를 도 6에 나타내었다. 다음은 nuc=y를 지정한 경우이다.
- [0187] **Read Length Distribution, After Alignment**
- [0188] Variant analyzer에 의해서 alignment에 성공된 read들만을 대상으로 각 read의 길이의 분포도를 나타낸다. 가로축은 read의 길이를 의미하고 세로축은 각 read의 개수를 의미한다. 가로축의 설명에는 평균 길이를 표시한다. 상기 그래프의 예를 도 7에 나타내었다.
- [0189] **Alignment Score Distribution**
- [0190] 각 read의 alignment score의 분포도를 나타낸다. 시퀀스의 길이를 L, reference 시퀀스와 일치하는 부분의 염기의 개수를 M, insertion에 의해서 불일치된 부분의 염기의 개수를 I, deletion에 의해서 불일치된 부분의 염기의 개수를 D라고 하면 alignment score는 $100 * (M - 0.2 * (I + D)) / L$ 으로 계산된다. 가로축은 alignment score를 의미하고 세로축은 각 read의 개수를 의미한다. 상기 그래프의 예를 도 8에 나타내었다.
- [0191] **Alignment Position Distribution**
- [0192] HBV 게놈 상에서 각 read가 alignment된 부분의 분포도를 나타낸다. 가로축은 HBV 게놈상의 염기의 위치를 의미하고 세로축은 각 위치의 염기를 포함하는 read의 개수를 의미한다. 상기 그래프의 예를 도 9에 나타내었다.
- [0193] **HCC Associated Mutations**
- [0194] HCC와 연관이 있다고 알려진 부위에 대해서 해당 부분의 변이 여부를 표시한다. HBV 게놈상의 염기 시퀀스 부위와 reference 염기, 각 염기를 포함하는 read의 수와 그 분포, 전체 read의 수, 해당 부위의 설명을 표시한다. 상기 표의 예를 도 10에 나타내었다.
- [0195] **변이 분석 결과 표시 모듈의 각 파일 설명**

- [0196] *haplotype_view_with_snps.php*
- [0197] SNPs 정보와 Strands의 정보를 보여준다.
- [0198] *sample_view_ava.php*
- [0199] 한 샘플에 대한 모든 분석 결과를 보여준다. 다른 php 파일들을 호출하는 기능을 갖는다.
- [0200] *chart_length_ava.php*
- [0201] 각 read들의 길이의 분포를 그림으로 표시한다.
- [0202] *chart_score_ava.php*
- [0203] 각 read들의 alignment score 의 분포를 그림으로 표시한다.
- [0204] *chart_coverage_ava.php*
- [0205] 각 read들의 HBV 게놈상의 coverage 의 분포를 그림으로 표시한다.
- [0206] *mutations.php*
- [0207] HCC 연관 변이주를 표시한다.

[0208] 한편, 상술한 상기 실시예 5는 컴퓨터에서 실행될 수 있는 프로그램으로 작성 가능하고, 컴퓨터로 읽을 수 있는 기록매체를 이용하여 상기 프로그램을 동작시키는 범용 디지털 컴퓨터에서 구현될 수 있다. 또한, 상술한 본 발명의 실시예에서 사용된 데이터의 구조는 컴퓨터로 읽을 수 있는 기록매체에 여러 수단을 통하여 기록될 수 있다. 상기 컴퓨터로 읽을 수 있는 기록매체는 마그네틱 저장매체(예를 들면, 롬, 플로피 디스크, 하드 디스크 등), 광학적 판독 매체(예를 들면, 시디롬, 디브이디 등)와 같은 저장매체를 포함한다.

도면

도면1



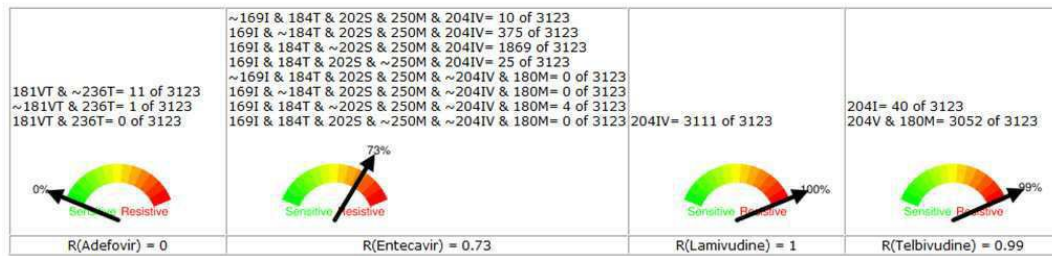
도면2

Drug Resistance Rates



도면3

Drug Resistance Rates



도면4

SNPs (greater than 1 percent)

Variation	Percent	# mutant / covering reads
L180M	4.9%	1061 / 21777
A181T	9.7%	2097 / 21606
A194T	1.1%	212 / 19812
S202N	1.1%	211 / 19003
M204I	12.4%	2357 / 19060
M204V	4.9%	942 / 19060
M250I	9.9%	1625 / 16462

도면5

Strands

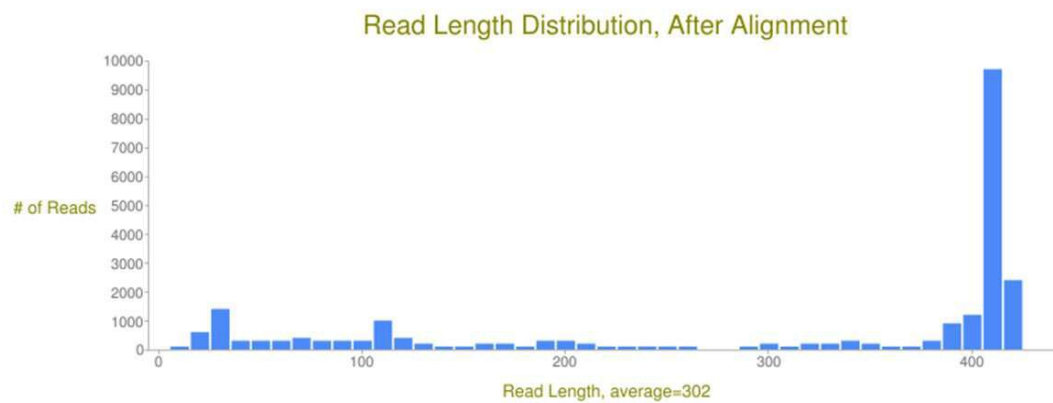
169	173	180	181	184	194	202	204	236	250	Count	percent
I	V	L	A	T	A	S	M	N	M	total 16247	Reference
.	.	.	.	.	.	.	.	.	.	9042	55.65
.	.	.	.	.	.	.	.	.	I	1141	7.02
.	.	.	T	.	.	.	.	.	.	1010	6.22
.	.	.	.	.	.	.	I	.	.	1004	6.18
.	.	M	.	.	.	.	V	.	.	410	2.52
.	.	.	.	.	.	.	V	.	.	195	1.2
.	.	.	.	.	.	.	I	.	.	195	1.2
.	.	M	.	.	.	.	I	.	.	177	1.09
.	.	.	T	.	.	.	.	.	.	105	0.65
.	.	.	.	.	.	.	.	.	.	104	0.64
.	.	.	.	.	.	.	.	.	.	91	0.56
.	.	.	.	.	.	.	.	.	.	84	0.52
.	.	.	.	.	.	.	.	.	.	79	0.49
.	.	.	.	.	.	.	I	.	I	72	0.44
.	.	.	.	.	.	.	.	.	.	66	0.41
.	.	.	.	.	.	.	S	.	.	64	0.39
.	.	.	.	.	.	N	.	.	.	59	0.36

도면6

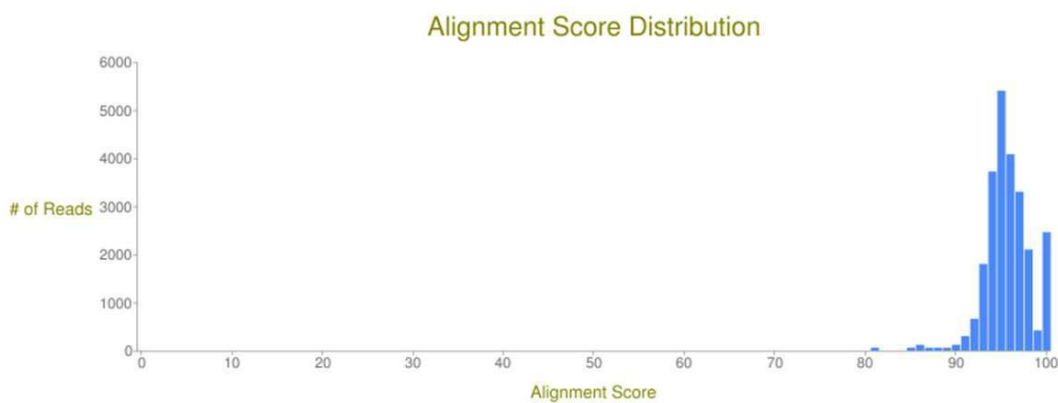
Strands

169	173	180	181	184	194	202	204	236	250	Count	percent
I	V	L	A	T	A	S	M	N	M	total 12703	Reference
ATT	GTG	CTG	GCT	ACT	GCT	AGC	ATG	AAC	ATG		
.	.	M	G	.	.	G	V	.	.	1262	9.93
...	...	ATG	GGT	...	...	GGT	GTG	...	...		
.	.	M	.	I	.	G	V	.	.	515	4.05
...	...	ATG	...	ATT	...	GGT	GTG	...	...		
.	.	M	.	.	.	G	V	.	.	394	3.1
...	...	ATG	...	...	...	GGT	GTG	...	...		
.	.	M	.	L	.	.	V	.	.	281	2.21
...	...	ATG	...	CTT	...	AGT	GTG	...	...		
.	M	M	.	.	.	G	V	.	.	93	0.73
...	ATG	ATG	...	...	...	GGT	GTG	...	...		
.	.	M	.	M	.	.	V	.	.	68	0.54
...	...	ATG	...	ATG	...	AGT	GTT	...	...		
.	M	M	G	.	.	G	V	.	.	58	0.46
...	ATG	ATG	GGT	...	...	GGT	GTG	...	...		
.	.	M	G	.	.	G	V	.	stop	50	0.39
...	...	ATG	GGT	...	...	GGT	GTG	...	TAG		
.	.	M	G	I	.	G	V	.	.	31	0.24
...	...	ATG	GGT	ATT	...	GGT	GTG	...	...		

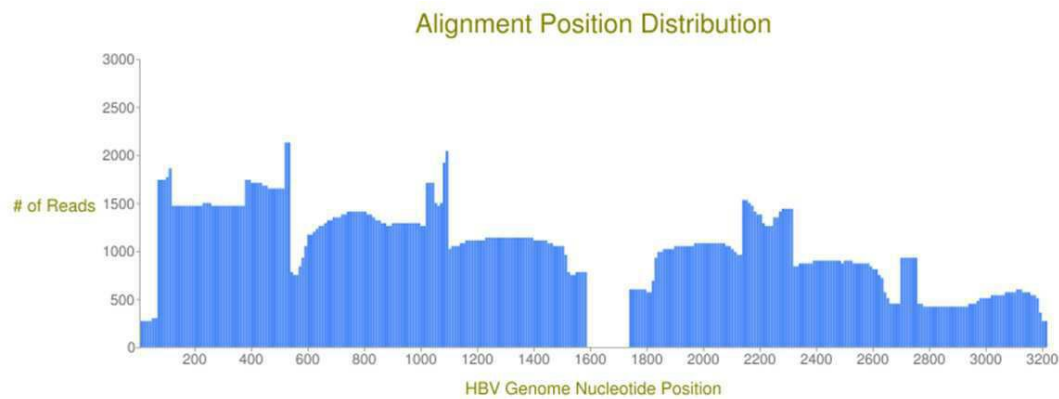
도면7



도면8



도면9



도면10

HCC Associated Mutations

position	ref	A		C		G		T		total	note
		count	percent	count	percent	count	percent	count	percent		
1653	C	0	0	0	0	0	0	0	0	0	enhancer II
1753	T	0	0	0	0	0	0	438	100	438	basal core promotor
1762	A	348	79	0	0	0	0	90	21	438	basal core promotor
1764	G	104	24	0	0	325	76	0	0	429	basal core promotor
2848	A	478	100	0	0	1	0	0	0	479	pre-S1 start codon 1
2849	T	0	0	0	0	0	0	479	100	479	pre-S1 start codon 2
2850	G	0	0	0	0	479	100	0	0	479	pre-S1 start codon 3
3205	A	377	99	0	0	1	0	2	1	380	pre-S2 start codon 1
3206	T	0	0	0	0	0	0	380	100	380	pre-S2 start codon 2
3207	G	0	0	0	0	379	100	0	0	379	pre-S2 start codon 3

서열 목록

- <110> Foundation for Industry Cooperation, University of Ulsan
- <120> Kit and method for detecting amino acid variation of Hepatitis B virus
- <130> PN094148
- <160> 174
- <170> Kopatent In 1.71
- <210> 1
- <211> 30
- <212> DNA
- <213> Artificial Sequence
- <220><223> primer A
- <400> 1
- ccatctcacc cctgcgtgtc tccgactcag
- <210> 2
- <211> 34

<212> DNA

<213> Artificial Sequence

<220><223> primer B

<400>

> 2

cctatccct gtgtgccttg gcagtctcag tcag 34

<210> 3

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> MID-1

<400> 3

acgagtgcgt 10

<210> 4

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> MID-2

<400> 4

acgctcgaca 10

<210> 5

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> MID-3

<400> 5

agacgcactc 10

<210> 6

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> MID-4

<400> 6

agcactgtag 10

<210>	7	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-5	
<400>	7	
atcagacacg		10
<210>	8	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-6	
<400>	8	
atatcgcgag		10
<210>	9	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-7	
<400>	9	
cgtgtctcta		10
<210>	10	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-8	
<400>	10	
ctcgcggtgc		10
<210>	11	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-10	

<400>	11	
tctctatgcg		10
<210>	12	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-11	
<400>	12	
tgatacgtct		10
<210>	13	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-13	
<400>	13	
catagtagtg		10
<210>	14	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-14	
<400>	14	
cgagagatac		10
<210>	15	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-15	
<400>	15	
atacgacgta		10
<210>	16	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	

<220><223>	MID-16	
<400>	16	
tcacgtacta		10
<210>	17	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-17	
<400>	17	
cgtctagtagc		10
<210>	18	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-18	
<400>	18	
tctacgtagc		10
<210>	19	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-19	
<400>	19	
tgtactactc		10
<210>	20	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-20	
<400>	20	
acgactacag		10
<210>	21	
<211>	10	

<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-21	
<400>	21	
cgtagactag		10
<210>	22	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-22	
<400>	22	
tacgagtatg		10
<210>	23	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-23	
<400>	23	
tactctcgtg		10
<210>	24	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-24	
<400>	24	
tagagacgag		10
<210>	25	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-25	
<400>	25	
tcgtcgctcg		10

<210>	26	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-26	
<400>	26	
acatacgcgt		10
<210>	27	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-27	
<400>	27	
acgcgagtat		10
<210>	28	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-28	
<400>	28	
actactatgt		10
<210>	29	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-29	
<400>	29	
actgtacagt		10
<210>	30	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-30	
<400>	30	

agactatact	10
<210> 31	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-31	
<400> 31	
agcgtcgtct	10
<210> 32	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-32	
<400> 32	
agtacgttat	10
<210> 33	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-33	
<400> 33	
atagagtact	10
<210> 34	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-34	
<400> 34	
cacgctacgt	10
<210> 35	
<211> 10	
<212> DNA	
<213> Artificial Sequence	

<220><223>	MID-35	
<400>	35	
cagtagacgt		10
<210>	36	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-36	
<400>	36	
cgacgtgact		10
<210>	37	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-37	
<400>	37	
tacacacact		10
<210>	38	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-38	
<400>	38	
tacacgtgat		10
<210>	39	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-39	
<400>	39	
tacagatcgt		10
<210>	40	
<211>	10	
<212>	DNA	

<213> Artificial Sequence

<220><223> MID-40

<400> 40

tacgctgtct

10

<210> 41

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> MID-41

<400> 41

tagtgtagat

10

<210> 42

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> MID-42

<400> 42

tcgatcacgt

10

<210> 43

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> MID-43

<400> 43

tcgcactagt

10

<210> 44

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> MID-44

<400> 44

tctagcgact

10

<210> 45

<211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-45
 <400> 45
 tctatactat 10
 <210> 46
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-46
 <400> 46
 tgacgtatgt 10

 <210> 47
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-47
 <400> 47
 tgtgagtagt 10
 <210> 48
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-48
 <400> 48
 acagtatata 10
 <210> 49
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-49
 <400> 49

acgcgatcga	10
<210> 50	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-50	
<400> 50	
actagcagta	10
<210> 51	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-51	
<400> 51	
agctcacgta	10
<210> 52	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-52	
<400> 52	
agtatacata	10
<210> 53	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-53	
<400> 53	
agtcgagaga	10
<210> 54	
<211> 10	
<212> DNA	
<213> Artificial Sequence	

<220><223>	MID-54	
<400>	54	
agtgctacga		10
<210>	55	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-55	
<400>	55	
cgatcgtata		10
<210>	56	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-56	
<400>	56	
cgcagtacga		10
<210>	57	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-57	
<400>	57	
cgcgtataca		10
<210>	58	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-58	
<400>	58	
cgtacagtca		10
<210>	59	
<211>	10	

<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-59	
<400>	59	
	cgtactcaga	10
<210>	60	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-60	
<400>	60	
	ctacgctcta	10
<210>	61	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-61	
<400>	61	
	ctatagcgta	10
<210>	62	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-62	
<400>	62	
	tagtcatca	10
<210>	63	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-63	
<400>	63	
	tagtcgcata	10
<210>	64	

<211> 9
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-64
 <400> 64
 atatatata 9

<210> 65
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-65
 <400> 65
 tatgctagta 10

<210> 66
 <211> 9
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-66
 <400> 66
 cacgcgaga 9

<210> 67
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-67
 <400> 67
 tcgatagtga 10

<210> 68
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-68
 <400> 68

tcgctgcgta	10
<210> 69	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-69	
<400> 69	
tctgacgtca	10
<210> 70	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-70	
<400> 70	
tgagtcagta	10
<210> 71	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-71	
<400> 71	
tgtagtgtga	10
<210> 72	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-72	
<400> 72	
tgtcacacga	10
<210> 73	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-73	

<400>	73	
tgtcgtcgca		10
<210>	74	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-74	
<400>	74	
acacatacgc		10
<210>	75	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-75	
<400>	75	
acagtcgtgc		10
<210>	76	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-76	
<400>	76	
acatgacgac		10
<210>	77	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-77	
<400>	77	
acgacagctc		10
<210>	78	
<211>	10	
<212>	DNA	

<213>	Artificial Sequence	
<220><223>	MID-78	
<400>	78	
	acgtctcatc	10
<210>	79	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-79	
<400>	79	
	actcatctac	10
<210>	80	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-80	
<400>	80	
	actcgcgcac	10
<210>	81	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-81	
<400>	81	
	agagcgtcac	10
<210>	82	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-82	
<400>	82	
	agcgactagc	10
<210>	83	

<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-83	
<400>	83	
	agtagtgatc	10
<210>	84	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-84	
<400>	84	
	agtgacacac	10
<210>	85	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-85	
<400>	85	
	agtgtatgtc	10
<210>	86	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-86	
<400>	86	
	atagatagac	10
<210>	87	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-87	
<400>	87	
	atatagtcgc	10

<210> 88
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-88
 <400> 88
 atctactgac 10

<210> 89
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-89
 <400> 89
 cacgtagatc 10

<210> 90
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-90
 <400> 90
 cacgtgtcgc 10

<210> 91
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-91
 <400> 91
 cataactctac 10

<210> 92
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-92

<400>	92	
cgacactatc		10
<210>	93	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-93	
<400>	93	
cgagacgcgc		10
<210>	94	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-94	
<400>	94	
cgtatgcgac		10
<210>	95	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-95	
<400>	95	
cgtcgatctc		10
<210>	96	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-96	
<400>	96	
ctacgactgc		10
<210>	97	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	

<220><223>	MID-97	
<400>	97	
ctagtcactc		10
<210>	98	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-98	
<400>	98	
ctctacgctc		10
<210>	99	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-99	
<400>	99	
ctgtacatac		10
<210>	100	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-100	
<400>	100	
tagactgcac		10
<210>	101	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-101	
<400>	101	
tagcgcgcg		10
<210>	102	
<211>	10	

<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-102	
<400>	102	
	tagctctatc	10
<210>	103	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-103	
<400>	103	
	tatagacatc	10
<210>	104	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-104	
<400>	104	
	tatgatacgc	10
<210>	105	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-105	
<400>	105	
	tcactcatac	10
<210>	106	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-106	
<400>	106	
	tcatcgagtc	10

<210>	107	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-107	
<400>	107	
tcgagctctc		10
<210>	108	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-108	
<400>	108	
tcgcagacac		10
<210>	109	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-109	
<400>	109	
tctgtctcgc		10
<210>	110	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-110	
<400>	110	
tgagtgcgc		10
<210>	111	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-111	
<400>	111	

tgatgtgtac	10
<210> 112	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-112	
<400> 112	
tgctatagac	10
<210> 113	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-113	
<400> 113	
tgctcgctac	10
<210> 114	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-114	
<400> 114	
acgtgcagcg	10
<210> 115	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-115	
<400> 115	
actcacagag	10
<210> 116	
<211> 10	
<212> DNA	
<213> Artificial Sequence	

<220><223>	MID-116	
<400>	116	
agactcagcg		10
<210>	117	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-117	
<400>	117	
agagagtgtg		10
<210>	118	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-118	
<400>	118	
agctatcgcg		10
<210>	119	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-119	
<400>	119	
agtctgactg		10
<210>	120	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-120	
<400>	120	
agtgagctcg		10
<210>	121	
<211>	10	
<212>	DNA	

<213> Artificial Sequence

<220><223> MID-121

<400> 121

atagctctcg 10

<210> 122

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> MID-122

<400> 122

atcacgtgcg 10

<210> 123

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> MID-123

<400> 123

atcgtagcag 10

<210> 124

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> MID-124

<400> 124

atcgtctgtg 10

<210> 125

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> MID-125

<400> 125

atgtacgatg 10

<210> 126

<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-126	
<400>	126	
atgtgtctag		10
<210>	127	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-127	
<400>	127	
cacacgatag		10
<210>	128	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-128	
<400>	128	
cactcgcacg		10
<210>	129	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-129	
<400>	129	
cagacgtctg		10
<210>	130	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-130	
<400>	130	

cagtactgcg	10
<210> 131	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-131	
<400> 131	
cgacagcgag	10
<210> 132	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-132	
<400> 132	
cgatctgtcg	10
<210> 133	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-133	
<400> 133	
cgcgtgctag	10
<210> 134	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-134	
<400> 134	
cgctcgagtg	10
<210> 135	
<211> 10	
<212> DNA	
<213> Artificial Sequence	

<220><223>	MID-135	
<400>	135	
cgtgatgacg		10
<210>	136	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-136	
<400>	136	
ctatgtacag		10
<210>	137	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-137	
<400>	137	
ctcgatatag		10
<210>	138	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-138	
<400>	138	
ctcgcacgcg		10
<210>	139	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-139	
<400>	139	
ctgcgtcacg		10
<210>	140	
<211>	10	

<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-140	
<400>	140	
	ctgtgcgtcg	10
<210>	141	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-141	
<400>	141	
	tagcatactg	10
<210>	142	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-142	
<400>	142	
	tatacatgtg	10
<210>	143	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-143	
<400>	143	
	tatcactcag	10
<210>	144	
<211>	10	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	MID-144	
<400>	144	
	tatctgatag	10
<210>	145	

<211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-145
 <400> 145
 tcgtgacatg 10

<210> 146
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-146
 <400> 146
 tctgatcgag 10

<210> 147
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-147
 <400> 147
 tgacatctcg 10

<210> 148
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-148
 <400> 148
 tgagctagag 10

<210> 149
 <211> 10
 <212> DNA
 <213> Artificial Sequence
 <220><223> MID-149
 <400> 149

tgatagagcg	10
<210> 150	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-150	
<400> 150	
tcggtgtgctg	10
<210> 151	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-151	
<400> 151	
tgctagtcag	10
<210> 152	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-152	
<400> 152	
tgtatcacag	10
<210> 153	
<211> 10	
<212> DNA	
<213> Artificial Sequence	
<220><223> MID-153	
<400> 153	
tgtgcgcgtg	10
<210> 154	
<211> 19	
<212> DNA	
<213> Artificial Sequence	

<220><223> forward primer for amplifying HBV genome (HBV-S-1(re)-F)

<400> 154

tccaacttgt cctggctat 19

<210> 155

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> reverse primer for amplifying HBV genome (HBV-S-1(re)-R)

<400> 155

ttcgaaaaca ttgcttgag 19

<210> 156

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> forward primer for amplifying HBV genome (HBV-S-2(re)-F)

<400> 156

ctaccagcac gggacat 18

<210> 157

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> reverse primer for amplifying HBV genome (HBV-S-2(re)-R)

<400> 157

gccttgtaag ttggcgaga 19

<210> 158

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> forward primer for amplifying HBV genome (HBV-new1(re)-F)

<400> 158

agaattgtgg gtcttttgg 19

<210> 159

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> reverse primer for amplifying HBV genome (HBV-new1(re)-R)

<400> 159

cagaggtgaa gcgaagtg 18

<210> 160

<211> 18

<212> DNA

<213> Artificial Sequence

<220><223> forward primer for amplifying HBV genome (HBV-new2(re)-F)

<400> 160

tctaccgtcc ccttcttc 18

<210> 161

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> reverse primer for amplifying HBV genome (HBV-new2(re)-F)

<400> 161

cacagaatag cttgcctga 19

<210> 162

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> forward primer for amplifying HBV genome (HBV-new3(re)-F)

<400> 162

ggtgggaagt aatttgaa 19

<210> 163

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> reverse primer for amplifying HBV genome (HBV-new3(re)-R)

<400> 163
ccagccttcc acagagtat 19

<210> 164
<211> 19
<212> DNA
<213> Artificial Sequence

<220><223> forward primer for amplifying HBV genome (HBV-new4(re)-F)

<400> 164
ggcctatatt ttctgctg 19

<210> 165
<211> 20
<212> DNA
<213> Artificial Sequence

<220><223> reverse primer for amplifying HBV genome (HBV-new4(re)-R)

<400> 165
acatagaggt tccttgagca 20

<210> 166
<211> 19
<212> DNA
<213> Artificial Sequence

<220><223> forward primer for amplifying HBV genome (HBV-geneS-F(re))

<400> 166
aaccctgttc cgactactg 19

<210> 167
<211> 19
<212> DNA
<213> Artificial Sequence

<220><223> reverse primer for amplifying HBV genome (HBV-geneS-R(re))

<400> 167
ttcgaaaaca ttgcttgag 19

<210> 168
<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> forward primer for amplifying HBV genome (HBV-4-F(re))

<400> 168

ttgttttaag actgggagga 20

<210> 169

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> reverse primer for amplifying HBV genome (HBV-4-R(re))

<400> 169

agtttccgga agtgttgat 19

<210> 170

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> forward primer for amplifying HBV genome (HBV-6-F(re))

<400> 170

cttgacaaa ggcattaaac 20

<210> 171

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> reverse primer for amplifying HBV genome (HBV-6-R(re))

<400> 171

cctcgagaag attgacgat 19

<210> 172

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> forward primer for amplifying HBV drug resistant site (HBV_459F)

<400> 172

tatgttgccc gttgtcctc 20

<210> 173

<211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> forward primer for amplifying HBV drug resistant site (HBV_574F)
 <400> 173
 aaaaccttcg gamggaaact 20
 <210> 174
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220><223> reverse primer for amplifying HBV drug resistant site (HBV_1031R)
 <400> 174
 gtaaaagggg cagcaaagc 19