



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2012년01월17일
 (11) 등록번호 10-1106014
 (24) 등록일자 2012년01월09일

(51) Int. Cl.
C12N 15/12 (2006.01) *C12N 15/10* (2006.01)
C12Q 1/68 (2006.01)
 (21) 출원번호 10-2009-0006101
 (22) 출원일자 2009년01월23일
 심사청구일자 2009년01월23일
 (65) 공개번호 10-2010-0064276
 (43) 공개일자 2010년06월14일
 (30) 우선권주장
 1020080122469 2008년12월04일 대한민국(KR)
 (56) 선행기술조사문헌
 W02003042409 A2*
 *는 심사관에 의하여 인용된 문헌

(73) 특허권자
차의과학대학교 산학협력단
 경기도 포천시 해룡로 120, 차의과학대학교내 (동교동, 포천중문의과대학)
 (72) 발명자
곽규범
 서울특별시 서초구 사임당로 130, 1동 305호 (서초동, 신동아아파트)
편정아
 경기도 안양시 동안구 관양2동 1493-51번지 창신마이빌 402호
(뒷면에 계속)
 (74) 대리인
오국진

전체 청구항 수 : 총 6 항

심사관 : 이현지

(54) 1번 염색체로부터 유래된 단일염기다형을 포함하는 폴리뉴클레오티드, 이를 포함하는 마이크로어레이 및 진단키트, 및 이를 이용한 분석방법

(57) 요약

본 발명은 1번 염색체상의 유전자, 특히, ZMPSTE24, GPR37L1, LMX1A, SRGAP2, FAM5C, EDEM3 유전자로부터 유래된 단일염기다형(Single-Nucleotide Polymorphism; SNP)을 포함하는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 제공하며, 상기 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드는 장형 위암(Intestinal type of gastric cancer)의 임상적 진단의 표지 인자로서, 사용될 수 있다.

(72) 발명자

이광재

경기도 성남시 분당구 정자동 정든마을 한진아파트
808동 201호

김진

서울특별시 양천구 목5동 목동트윈빌 D동 311호

조성원

경기도 성남시 분당구 수내동 양지마을 금호 1단지
아파트 101동 101호

이 발명을 지원한 국가연구개발사업

과제고유번호 A010383

부처명 보건복지가족부

연구관리전문기관

연구사업명 미래보건기술개발사업

연구과제명 간 및 소화기질환 유전체 연구센터

기여율

주관기관 아주대학교의료원

연구기간 2007년 11월 1일 ~ 2009년 3월 31일

특허청구의 범위

청구항 1

삭제

청구항 2

삭제

청구항 3

삭제

청구항 4

삭제

청구항 5

삭제

청구항 6

서열번호 6의 DNA 서열로 구성된 폴리뉴클레오티드에서, 501번째 염기(다형성 부위)가 G이고, 상기 501번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (f);

서열번호 7의 DNA 서열로 구성된 폴리뉴클레오티드에서, 262번째 염기(다형성 부위)가 G이고, 상기 262번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (g); 및

서열번호 8의 DNA 서열로 구성된 폴리뉴클레오티드에서, 2187번째 염기(다형성 부위)가 G이고, 상기 2187번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (h)

로 이루어진 군으로부터 하나 이상 선택된 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 포함하는, 장형 위암 진단용 키트.

청구항 7

장형 위암 진단에 필요한 정보를 제공하기 위하여, (i) 검체로부터 핵산 시료를 얻는 단계; 및 (ii) 상기 핵산 시료로부터, 서열번호 6 내지 8의 폴리뉴클레오티드로부터 하나 이상 선택된 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드의 다형성 부위의 염기 서열을 분석하는 단계 (단, 상기 다형성 부위는 서열번호 6의 DNA 서열의 경우 501번째 염기, 서열번호 7의 DNA 서열의 경우 262번째 염기, 서열번호 8의 DNA 서열의 경우 2187번째 염기를 각각 의미한다)를 포함하는, 검체 중의 단일염기다형을 검출하는 방법.

청구항 8

제7항에 있어서, 상기 다형성 부위의 염기 서열을 분석하는 단계가 서열번호 6 내지 8의 폴리뉴클레오티드로 이루어진 군으로부터 선택된 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드 서열로서 다형성 부위를 포함하는 서열(단, 상기 다형성 부위는 서열번호 6의 DNA 서열의 경우 501번째 염기, 서열번호 7의 DNA 서열의 경우 262번째 염기, 서열번호 8의 DNA 서열의 경우 2187번째 염기를 각각 의미한다)을 주형으로 하는 프라이머 또는 프로브를 이용하여 수행되는 것을 특징으로 하는, 검체 중의 단일염기다형을 검출하는 방법.

청구항 9

제7항에 있어서, 상기 다형성 부위의 염기 서열을 분석하는 단계가 폴리뉴클레오티드 (f) 내지 (h)로 이루어진 군으로부터 선택된 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드가 고정된 마이크로어레이에 상기 핵산 시료를 혼성화시키는 단계 (단, 상기 폴리뉴클레오티드 (f) 내지 (h)는 제6항에서 정의한 바와 같다); 및 얻어진 혼성화 결과를 분석하는 단계를 포함하는 것을 특징으로 하는, 검체 중의 단일염기다형을 검출하는 방법.

청구항 10

제9항에 있어서, 상기 마이크로어레이에 고정되는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드의 DNA 서열의 길이가 각각 10 내지 100 뉴클레오티드인 것을 특징으로 하는, 검체 중의 단일염기다형을 검출하는 방법.

청구항 11

제7항에 있어서, 상기 다형성 부위의 염기 서열을 분석하는 단계가 서열번호 6 내지 8의 DNA 서열로 이루어진 군으로부터 선택된 DNA 서열의 다형성 부위(단, 상기 다형성 부위는 서열번호 6의 DNA 서열의 경우 501번째 염기, 서열번호 7의 DNA 서열의 경우 262번째 염기, 서열번호 8의 DNA 서열의 경우 2187번째 염기를 각각 의미한다)의 유전자형(genotype) 분석을 포함하는 것을 특징으로 하는, 검체 중의 단일염기다형을 검출하는 방법.

명세서

발명의 상세한 설명

기술분야

[0001] 본 발명은 1번 염색체상의 유전자, 특히, ZMPSTE24, GPR37L1, LMX1A, SRGAP2, FAM5C, EDEM3 유전자로부터 유래된 단일염기다형을 포함하는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드; 상기 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드로 이루어진 장형 위암(intestinal type of gastric cancer) 진단을 위한 증폭용 프라이머 또는 장형 위암 진단용 프로브; 상기 프로브를 포함하는 마이크로어레이; 및 상기 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 이용한 장형 위암 진단에 필요한 정보를 제공하기 위한 분석 방법에 관한 것이다.

배경기술

[0002] 인간의 게놈은 모두 22쌍의 상동 염색체와 두 개의 성염색체로 이루어져 있다. 그러나 모든 인간이 각기 다른 외모와 성격을 갖는다. 이는 개개인이 같은 수의 염색체를 가지고 있기는 하지만, 그 염색체로부터 발현되는 유전자들에 의한 표현형이 모두 다르기 때문이다. 개개인의 이러한 유전자 발현의 다양성은 곧 유전자를 가지고 있는 게놈의 유전적 다양성 때문이다. 이러한 유전적 다양성은 많은 연구가 수행되어 왔으며, 최근 인간게놈프로젝트의 완성으로 그 구조가 모두 밝혀졌다.

[0003] 게놈에서의 유전적 다양성을 대표하는 것으로는 Transposable element, STR (short tandem repeats), Microsatellites, STS (sequence tagged site), VNTR (variable number tandem repeat), SNP (single nucleotide polymorphism : 단일염기다형) 등이 알려져 있다. 이중 SNP는 단일염기다형으로서, 인간 게놈 상에서 약 1000개의 뉴클레오티드 중 1개의 빈도로 나타나며 동일한 종의 개체 사이의 단일뉴클레오티드 변이의 형태를 취한다. 이러한 단일염기다형이 나타내는 유전학에서의 의미는 그 위치에 따라 각 개체에 큰 차이를 가져온다. 예를 들면, 단일염기다형이 단백질을 암호화하고 있는 위치에 존재할 경우, 단백질의 구조에 영향을 미칠 수 있게 되어 단백질 기능이 달라질 수 있게 되며, 질병과도 연관될 수 있다. 다른 예로 단일염기다형이 단백질을 암호화하지 않는 다른 유전자상에서 존재할 경우, 즉 프로모터(promoter) 또는 인트론(intron)에 존재할 경우, 각각에 대하여 단백질의 발현 수준에 차이를 가져와 그 단백질의 전체적인 활성이 줄어들 수 있고, 변성적인 인트론 제거 과정(alternative splicing)을 통하여 단백질이 비정상적으로 발현될 수도 있다.

[0004] 단일염기다형들은 질병과 같은 특정 표현형(phenotype)의 지표로 사용될 수 있고, 이를 위해서는 가장 우선적으로 인간 유전체에 존재하는 대량의 단일염기다형을 발굴하고 발굴된 단일염기다형은 다수의 인간 유전체 시료에서 전체 인간집단의 1% 이상 존재하는 유전변이형 마커인지를 조사하는 검정작업을 거치게 된다. 검정된 SNP은 특정 질병 또는 약물반응과 같은 유전적 차이에 기인된 원인을 찾기 위한 연관성 연구에 사용되게 된다. 예를 들면, 특정 단일염기다형이 환자군과 대조군 간의 비교에서 유의적인 빈도의 차이가 관측된다면 아주 유용한 단일염기다형으로 선발되어 임상실험을 거치게 될 것이며 임상실험의 결과가 계속적으로 진행되어 그 결과가 좋을 때는 경제적으로 아주 유용한 가치를 지니는 유전지표로 실제 임상에 사용할 수 있게 될 것이다.

[0005] 이러한 단일염기다형의 발견을 위한 종래 기술로는 제한효소를 이용한 제한 단편화 길이 다형성(Restriction Fragment Length Polymorphism; RFLP), 대립형질 특이적 혼성화(Allele-specific hybridization)를 이용한 TaqManTM 프로브 방법, 용융온도(melting temperature; Tm)를 이용한 역동적 대립형질 특이적 혼성화(Dynamic Allele-Specific Hybridization; DASH)방법, 중합반응을 이용한 pyrosequencing등이 있다. 최근, 마이크로어레이

이 기술은 단일 염기 신장(Single Base Extension)이라는 원리를 이용한 프로브들을 조그만 기관위에 집적화시켜 심은 후, 목적 DNA들과의 혼성화와 신장(Extension)을 시킴으로써 단일염기다형을 발견해 내고 있다.

[0006] 한편, 위암의 발생률과 사망률은 감소하고 있으나, 아직도 전 세계에서 두 번째로 많이 발생하는 악성 종양이다. 우리나라의 경우, 남성과 여성에서의 암 발생률은 위암이 1위를 차지하고 사망률의 경우 2위를 차지한다. 위암과 깊게 관련되는 *Helicobacter pylori* (*H. pylori*) 감염은 사람에서 가장 흔한 만성 감염으로서 전 세계 인구의 50% 이상이 감염되어 있다. 우리나라의 *H. pylori* 감염률은 전 국민에서 46.6%, 16세 이상의 성인에서는 69.4%, 40대 에서는 78.5% 라는 높은 감염률을 보인다. 위암은 로렌 분류법(Lauren classification)에 의해 장형 위암(Intestinal type of gastric cancer) 과 미만형 위암(Diffuse type of gastric cancer)로 분류된다. 장형 위암은 *H. pylori* 감염과 관련되어 만성 위염(chronic gastritis), 위축성 위염(atrophic gastritis), 장상피화생(intestinal metaplasia), 이형성증(displasia), 장형 위암(Intestinal type of gastric cancer)의 순서로 진행된다. 반면 미만형 위암은 정상적인 점막으로부터 *H. pylori* 감염과 관계없이 유발 될 수 있다. 우리나라에서 *H. pylori* 감염률이 높다는 것은 장형 위암에 대한 위험이 높다는 것을 의미한다.

[0007] 초기 위암은 5년 생존률이 95~100%로 높지만 증상을 나타내지 않는 경우가 80% 를 차지한다. 이는 조기 진단의 중요성과 함께 적절한 조기 진단법의 개발이 필요함을 뜻한다. *H. pylori* 감염된 100명 중 1~2명에서만 위암이 발생하기 때문에 *H. pylori* 감염만으로 위암 발생 위험군으로 판단 내리는 것은 적합하지 않다. 위암의 전 단계에 속하는 만성 위염 환자 사이에서 장형 위암이 발생할 위험성이 있는 사람을 미리 예측하는 것과 위암 발생에 관여하는 인자를 규명하는 것은 국민 보건에 중요하며 위암을 예방하기 위한 검사의 필수 요건이다. 장형 위암에 특이적으로 나타날 수 있는 단일염기다형을 찾아내는 것은 장형 위암을 조기에 진단할 수 있으며, 더 나아가 병태생리학적 경로를 밝히고 치료 방법을 개발하는데까지 이용될 수 있을 것이다.

발명의 내용

해결 하고자하는 과제

[0008] 본 발명자들은 한국인의 장형 위암 환자에서 특이적으로 나타나는 단일염기다형을 찾아내고자 연구를 거듭한 결과, 1번 염색체상의 유전자, 특히, ZMPSTE24, GPR37L1, LMX1A, SRGAP2, FAM5C, EDEM3 유전자로부터 장형 위암 환자에 특이적으로 나타나는 단일염기다형 부위를 발견하여 본 발명을 완성하게 되었다.

[0009] 따라서, 본 발명은 장형 위암과 관련된 단일염기다형으로서 ZMPSTE24, GPR37L1, LMX1A, SRGAP2, FAM5C, EDEM3 유전자로부터 유래된 단일염기다형을 포함하는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 제공하는 것을 목적으로 한다.

[0010] 또한, 본 발명은 상기 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드로 이루어진 장형 위암 진단을 위한 증폭용 프라이머 또는 장형 위암 진단용 프로브를 제공하는 것을 목적으로 한다.

[0011] 또한, 본 발명은 상기 프로브를 포함하는 마이크로어레이를 제공하는 것을 목적으로 한다.

[0012] 또한, 본 발명은 상기 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 이용한 장형 위암 진단 방법을 제공하는 것을 목적으로 한다.

과제 해결수단

[0013] 본 발명의 일 태양에 따라, 서열번호 1의 DNA 서열로 구성된 폴리뉴클레오티드에서, 501번째 염기(다형성 부위)가 A이고, 상기 501번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (a); 서열번호 2의 DNA 서열로 구성된 폴리뉴클레오티드에서, 101번째 염기(다형성 부위)가 G이고, 상기 101번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (b); 서열번호 3의 DNA 서열로 구성된 폴리뉴클레오티드에서, 401번째 염기(다형성 부위)가 G이고, 상기 401번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (c); 서열번호 4의 DNA 서열로 구성된 폴리뉴클레오티드에서, 501번째 염기(다형성 부위)가 G이고, 상기 501번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (d); 서열번호 5의 DNA 서열로 구성된 폴리뉴클레오티드에서, 301번째 염기(다형성 부위)가 C이고, 상기 301번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (e); 서열번호 6의 DNA 서열로 구성된 폴리뉴클레오티드에서, 501번째 염기(다형성 부위)가 G이고, 상기 501번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (f); 서열번호 7의 DNA 서열로 구성된 폴리뉴클레오티드에서, 262

번째 염기(다형성 부위)가 G이고, 상기 262번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (g); 서열번호 8의 DNA 서열로 구성된 폴리뉴클레오티드에서, 2187번째 염기(다형성 부위)가 G이고, 상기 2187번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (h); 서열번호 9의 DNA 서열로 구성된 폴리뉴클레오티드에서, 301번째 염기(다형성 부위)가 G이고, 상기 301번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (i); 및 서열번호 10의 DNA 서열로 구성된 폴리뉴클레오티드에서, 201번째 염기(다형성 부위)가 A이고, 상기 201번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (j)로 이루어진 군으로부터 하나 이상 선택된 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드가 제공된다.

[0014] 또한, 본 발명의 다른 태양에 따라, 상기 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드로 이루어진, 장형 위암 진단을 위한 증폭용 프라이머가 제공된다.

[0015] 또한, 본 발명의 다른 태양에 따라, 상기 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드로 이루어진, 장형 위암 진단용 프로브 및 이를 포함한 마이크로어레이가 제공된다.

[0016] 또한, 본 발명의 다른 태양에 따라, 상기 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 포함하는 장형 위암 진단용 키트가 제공된다.

[0017] 또한, 본 발명의 다른 태양에 따라, 장형 위암 진단에 필요한 정보를 제공하기 위하여, (i) 검체로부터 핵산 시료를 얻는 단계; 및 (ii) 상기 핵산 시료로부터, 서열번호 1 내지 10의 폴리뉴클레오티드로부터 하나 이상 선택된 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드의 다형성 부위의 염기 서열을 분석하는 단계 (단, 상기 다형성 부위는 서열번호 1의 DNA 서열의 경우 501번째 염기, 서열번호 2의 DNA 서열의 경우 101번째 염기, 서열번호 3의 DNA 서열의 경우 401번째 염기, 상기 다형성 부위는 서열번호 4의 DNA 서열의 경우 501번째 염기, 상기 다형성 부위는 서열번호 5의 DNA 서열의 경우 301번째 염기, 상기 다형성 부위는 서열번호 6의 DNA 서열의 경우 501번째 염기, 상기 다형성 부위는 서열번호 7의 DNA 서열의 경우 262번째 염기, 상기 다형성 부위는 서열번호 8의 DNA 서열의 경우 2187번째 염기, 상기 다형성 부위는 서열번호 9의 DNA 서열의 경우 301번째 염기, 서열번호 10의 DNA 서열의 경우 201번째 염기를 각각 의미한다)를 포함하는, 검체 중의 단일염기다형을 검출하는 방법이 제공된다.

효과

[0018] 본 발명에 따른 폴리뉴클레오티드 (a) 내지 (j)로 이루어진 군으로부터 하나 이상 선택된 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드는 장형 위암의 진단에 유용하게 사용될 수 있다.

[0019] 또한, 상기 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드로 이루어진 프라이머 또는 프로브, 및 상기 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 포함하는 키트는 장형 위암의 진단에 유용하게 사용될 수 있다.

[0020] 또한 본 발명의 분석방법은 장형 위암 진단에 필요한 정보를 제공하는데 유용하게 사용될 수 있다.

발명의 실시를 위한 구체적인 내용

[0021] 본 발명은 장형 위암과 관련된 것으로 새롭게 밝혀진 서열번호 1 내지 10의 서열로부터 유래하는 폴리뉴클레오티드 (a) 내지 (j)로 이루어진 군으로부터 하나 이상 선택된 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 제공한다. 즉, 본 발명은 서열번호 1의 DNA 서열로 구성된 폴리뉴클레오티드에서, 501번째 염기(다형성 부위)가 A이고, 상기 501번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (a); 서열번호 2의 DNA 서열로 구성된 폴리뉴클레오티드에서, 101번째 염기(다형성 부위)가 G이고, 상기 101번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (b); 서열번호 3의 DNA 서열로 구성된 폴리뉴클레오티드에서, 401번째 염기(다형성 부위)가 G이고, 상기 401번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (c); 서열번호 4의 DNA 서열로 구성된 폴리뉴클레오티드에서, 501번째 염기(다형성 부위)가 G이고, 상기 501번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (d); 서열번호 5의 DNA 서열로 구성된 폴리뉴클레오티드에서, 301번째 염기(다형성 부위)가 C이고, 상기 301번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (e); 서열번호 6의 DNA 서열로 구성된 폴리뉴클레오티드에서, 501번째 염기(다형성 부위)가 G이고, 상기 501번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (f); 서열번호 7의 DNA 서열로 구성된 폴리뉴클레오티드에서, 262번째 염기(다형성 부위)가 G이고, 상기 262번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴

리뉴클레오티드 (g); 서열번호 8의 DNA 서열로 구성된 폴리뉴클레오티드에서, 2187번째 염기(다형성 부위)가 G이고, 상기 2187번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (h); 서열번호 9의 DNA 서열로 구성된 폴리뉴클레오티드에서, 301번째 염기(다형성 부위)가 G이고, 상기 301번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (i); 및 서열번호 10의 DNA 서열로 구성된 폴리뉴클레오티드에서, 201번째 염기(다형성 부위)가 A이고, 상기 201번째 염기를 포함하는 10개 이상의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 (j)로 이루어진 군으로부터 하나 이상 선택된 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 제공한다.

[0022] 상기 서열번호 1 내지 10의 DNA 서열은 만성 위염 환자군과 장형 위암 환자군에서 통계학적으로 유의성 있는 차이(유의수준 p value < 0.05)를 보이는 단일염기다형으로서, 이들은 1번 염색체 상에 위치하는 ZMPSTE24, GPR37L1, LMX1A, SRGAP2, FAM5C, EDEM3 유전자에 존재한다. ZMPSTE24 유전자는 zinc metalloproteinase를 암호화하며, 이 유전자 상의 돌연변이는 mandibuloacral dysplasia 와 제한 피부병증(restrictive dermopathy)과 연관되는 것으로 보고된 바 있다. GPR37L1 (G protein-coupled receptor 37 like 1) 유전자는 그 기능이 아직 밝혀져 있지 않다. LMX1A (LIM homeobox transcription factor 1, alpha) 유전자는 인슐린 프로모터에 붙어 인슐린의 전사를 자극하는 것으로 알려져 있다. SRGAP2 (SLIT-ROBO Rho GTPase activating protein 2) 유전자는 ARHGAP 관련 단백질(family protein)을 암호화하며 흑색종(melanoma), 생식세포 종양(germ cell tumors), 연골육종(chondrosarcoma) 및 망막모세포종(retinoblastoma)에 발현되는 것이 보고된 바 있다. FAM5C (family with sequence similarity 5, member C) 유전자는 심근경색(myocardial infarction)과의 연관성이 보고된 바 있고, EDEM3 (ER degradation enhancer, mannosidase alpha-like 3) 유전자는 당단백(glycoprotein)에 의한 ERAD(endoplasmic reticulum-associated degradation) 및 만노오즈 트리밍(mannose trimming)을 증진시킨다고 보고된 바 있다.

[0023] 본 발명자들은 장형 위암과 관련된 후보 유전자군으로서 1번 염색체 상에 위치하는 ZMPSTE24, GPR37L1, LMX1A, SRGAP2, FAM5C, EDEM3 유전자를 선정하여, 장형 위암 진단에 있어서 분자 수준에서의 진단 지표를 개발하고자 다양한 연구를 수행하였다. 본 발명자들은 한국인 만성 위염 환자 24명과 장형 위암 환자 21명을 대상으로 유전자형 분석 및 통계학적 분석을 수행하였으며, 놀랍게도 수많은 단일염기다형 중 극히 일부의 단일염기다형만이 만성 위염 환자군과 장형 위암 환자군에서 통계적으로 유의성 있는 차이를 나타낸다는 것을 발견하였다. 특히, 서열번호 1 내지 10의 단일염기다형은 하디-와인버그 평형 (Hardy-Weinberg Equilibrium) 및 로지스틱 회귀분석 결과, 각각의 다형성 부위에 위험대립인자가 존재함을 확인하였다.

[0024] 본 발명은 상기 폴리뉴클레오티드 (a) 내지 (j)로 이루어진 군으로부터 하나 이상 선택된 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드와 혼성화하는 폴리뉴클레오티드를 포함하며, 상기 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드의 길이는 10 내지 100 뉴클레오티드, 바람직하게는 20 내지 60 뉴클레오티드, 더욱 바람직하게는 40 내지 60 뉴클레오티드이다.

[0025] 본 발명에 따른 상기 폴리뉴클레오티드 (a) 내지 (j)로 이루어진 군으로부터 하나 이상 선택된 폴리뉴클레오티드는 다형성 서열이다. 다형성 서열 (polymorphic sequence)이란 뉴클레오티드 서열 중에 단일염기다형을 나타내는 다형성 부위 (polymorphic site)를 포함하는 서열을 말한다. 다형성 부위 (polymorphic site)란 다형성 서열 중 단일염기다형이 일어나는 부위를 말한다. 상기 폴리뉴클레오티드는 DNA 또는 RNA가 될 수 있다.

[0026] 본 발명은 또한, 상기 폴리뉴클레오티드 (a) 내지 (j)로 이루어진 군으로부터 하나 이상 선택된 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드로 이루어진, 장형 위암 진단을 위한 증폭용 대립형질 특이적 프라이머를 포함한다. 여기서 "프라이머(primer)"란 적절한 버퍼 중의 적절한 조건 (예를 들면, 4개의 다른 뉴클레오티드 트리포스페이트(ATP,GTP,CTP,TTP) 및 DNA 폴리머라제) 및 적당한 온도 하에서 주형-지시 DNA 합성의 시작점으로서 작용할 수 있는 단일가닥 올리고뉴클레오티드를 말한다. 상기 프라이머의 적절한 길이는 사용 목적에 따라 달라질 수 있으나, 통상 10 내지 100, 바람직하게는 10 내지 80 뉴클레오티드이다. 짧은 프라이머 분자는 일반적으로 주형과 안정한 혼성체를 형성하기 위해서는 더 낮은 온도를 필요로 한다. 프라이머 서열은 주형과 완전하게 상보적일 필요는 없으나, 주형과 혼성화 할 정도로 충분히 상보적이어야 한다. 상기 프라이머는 3'-말단이 서열번호 1 내지 10의 단일염기다형과 정렬하는 것이 바람직하다. 상기 프라이머는 다형성 부위를 포함하는 표적 DNA에 혼성화하고, 상기 프라이머가 완전한 상동성을 보이는 대립형질 형태의 DNA에서 증폭을 개시한다. 이 프라이머는 반대편에 혼성화하는 제2 프라이머와 쌍을 이루어 사용된다. 증폭에 의하여 두 개의 프라이머로부터 산물이 증폭되고, 증폭된 산물만을 특이적으로 염색하여 시각화한다. 이는 특정 대립형질 형태가 존재한다는 것을 의미한다. 여기서 사용되는 대립형질 특이적 폴리뉴클레오티드들이 다른 변성화 과정을 거치게 되면,

Infinium™ Assay I 뿐만 아니라 다른 방법들에 의해서도 사용될 수 있다.

- [0027] 본 발명은 폴리뉴클레오티드 (a) 내지 (j)로 이루어진 군으로부터 하나 이상 선택된 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드로 이루어진, 장형 위암 진단용 프로브 및 이를 포함하는 장형 위암 진단용 마이크로어레이를 포함한다. 또한, 본 발명은 폴리뉴클레오티드 (a) 내지 (j)로 이루어진 군으로부터 하나 이상 선택된 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 포함하는, 장형 위암 진단용 키트를 포함한다.
- [0028] 본 발명은 장형 위암 진단에 필요한 정보를 제공하기 위하여, (i) 검체로부터 핵산 시료를 얻는 단계; 및 (ii) 상기 핵산 시료로부터, 서열번호 1 내지 10의 폴리뉴클레오티드로부터 하나 이상 선택된 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드의 다형성 부위의 염기 서열을 분석하는 단계 (단, 상기 다형성 부위는 서열번호 1의 DNA 서열의 경우 501번째 염기, 서열번호 2의 DNA 서열의 경우 101번째 염기, 서열번호 3의 DNA 서열의 경우 401번째 염기, 상기 다형성 부위는 서열번호 4의 DNA 서열의 경우 501번째 염기, 상기 다형성 부위는 서열번호 5의 DNA 서열의 경우 301번째 염기, 상기 다형성 부위는 서열번호 6의 DNA 서열의 경우 501번째 염기, 상기 다형성 부위는 서열번호 7의 DNA 서열의 경우 262번째 염기, 상기 다형성 부위는 서열번호 8의 DNA 서열의 경우 2187번째 염기, 상기 다형성 부위는 서열번호 9의 DNA 서열의 경우 301번째 염기, 서열번호 10의 DNA 서열의 경우 201번째 염기를 각각 의미한다)를 포함하는, 검체 중의 단일염기다형을 검출하는 방법을 포함한다.
- [0029] 상기 검체는 혈액, 조직, 세포 등을 포함하며, 핵산 시료는 통상의 방법에 따라 얻을 수 있다. 예를 들어 혈액에 경우 다음과 같이 핵산 시료를 얻을 수 있다. 혈액을 원심분리 시험관으로 옮기고 낮은 농도의 염 성분 완충용액(low-salt buffer solution(10mM Tris-HCl [pH 7.6], 10mM KCl, 10mM MgCl₂, 및 2mM EDTA))을 첨가한다. 그리고 Nonidet P-40을 첨가한 다음, 얻어진 용액을 잘 섞어준 후, 상온에서 10분간 약 2,200 rpm으로 원심분리한다. 이때 얻어진 덩어리는 고농도의 염 성분 완충용액(high-salt buffer solution(10mM Tris-HCl [pH 7.6], 10mM KCl, 10mM MgCl₂, 0.4M NaCl, 및 2mM EDTA))으로 재현탁한다. 게놈 DNA는 50ul의 10% SDS와 섞은 후, 약 55°C에서 밤새 반응시킨다. 반응 후에, 각 게놈 DNA들은 1.5ml 용량의 원심분리 시험관으로 옮겨, 6M NaCl을 0.3ml 첨가한다. 게놈 DNA들을 잘 섞어준 후, 약 12,000rpm에서 10분간 원심분리한다. 원심분리 후, 게놈 DNA가 들어있는 상층을 모아서 새로운 시험관으로 옮기고, 상온에서 동량의 100% 에탄올을 첨가한다. 게놈 DNA들이 침전될 때까지 잘 섞어 주고, 침전이 되면 70% 에탄올로 게놈 DNA들을 세척한다. 이 과정이 끝나면 시험관 뚜껑을 열어, 용액성분이 마르도록 놓아둔다. 그리고 TE 완충용액(pH 8.0)을 넣어 게놈 DNA를 재현탁함으로써, 핵산 시료를 얻을 수 있다.
- [0030] 여기에서, 상기 다형성 부위의 염기 서열을 분석하는 단계는 서열번호 1 내지 10의 폴리뉴클레오티드로 이루어진 군으로부터 선택된 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드 서열로서 다형성 부위를 포함하는 서열(단, 상기 다형성 부위는 서열번호 1의 DNA 서열의 경우 501번째 염기, 서열번호 2의 DNA 서열의 경우 101번째 염기, 서열번호 3의 DNA 서열의 경우 401번째 염기, 상기 다형성 부위는 서열번호 4의 DNA 서열의 경우 501번째 염기, 상기 다형성 부위는 서열번호 5의 DNA 서열의 경우 301번째 염기, 상기 다형성 부위는 서열번호 6의 DNA 서열의 경우 501번째 염기, 상기 다형성 부위는 서열번호 7의 DNA 서열의 경우 262번째 염기, 상기 다형성 부위는 서열번호 8의 DNA 서열의 경우 2187번째 염기, 상기 다형성 부위는 서열번호 9의 DNA 서열의 경우 301번째 염기, 서열번호 10의 DNA 서열의 경우 201번째 염기를 각각 의미한다)을 주형으로 하는 프라이머 또는 프로브를 이용하여 수행될 수 있다. 상기 프라이머 또는 프로브는 상기에서 설명한 바와 같다.
- [0031] 또한, 상기 다형성 부위의 염기 서열을 분석하는 단계는 폴리뉴클레오티드 (a) 내지 (j)로 이루어진 군으로부터 선택된 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드가 고정된 마이크로어레이에 상기 핵산 시료를 혼성화시키는 단계 (단, 상기 폴리뉴클레오티드 (a) 내지 (j)는 상기에서 정의한 바와 같다); 및 얻어진 혼성화 결과를 분석하는 단계를 포함하여 수행될 수 있으며, 상기 마이크로어레이에 고정되는 폴리뉴클레오티드 또는 그의 상보적 뉴클레오티드의 DNA 서열의 길이는 각각 10 내지 100 뉴클레오티드일 수 있다.
- [0032] 본 발명의 검체 중의 단일염기다형을 검출하는 방법에 따라 얻어진 결과는 장형 위암 진단에 필요한 정보를 제공할 수 있으며 예를 들어, 서열번호 1 내지 10의 다형성 부위의 염기가 각각 A, G, G, G, C, G, G, G, G, A (위험 대립인자)인 것이 하나 이상 존재하는 경우, 장형 위암에 걸릴 확률이 높은 위험군에 속하는 것으로 판정할 수 있다. 상기 위험 대립 인자를 갖는 핵산 서열이 하나의 검체에서 많이 검출되면 될수록 위험군에서 속하는 확률은 더 높은 것으로 판단할 수 있다.
- [0033] 또한, 상기 다형성 부위의 염기 서열을 분석하는 단계는 서열번호 1 내지 10의 DNA 서열로 이루어진 군으로부터 선택된 DNA 서열의 다형성 부위의 유전자형(genotype) 분석을 포함할 수 있다. 즉, 서열번호 1 내지 10의 다형

성 부위의 유전자형이 A/A와 A/G (서열번호 1), G/G와 A/G (서열번호 2), G/G (서열번호 3), G/G (서열번호 4), C/C와 A/C (서열번호 5), G/G와 A/G (서열번호 6), G/G와 A/G (서열번호 7), G/G와 A/G (서열번호 8), G/G (서열번호 9) 및 A/A (서열번호 10) 가 하나 이상 존재하는 경우 장형 위암에 걸릴 위험도가 높은 것으로 판단할 수 있다. 상기 위험 유전자형을 갖는 핵산 서열이 하나의 검체에서 많이 검출되면 될수록 위험군에서 속하는 확률은 더 높은 것으로 판단할 수 있다.

[0034] 이하, 본 발명을 실시예를 통하여 더욱 상세히 설명한다. 그러나 이들 실시예는 본 발명을 예시하기 위한 것으로, 본 발명의 범위가 이들 실시예에 한정되는 것은 아니다.

[0035] 실시예

[0036] 1. 피검자 선정

[0037] 본 실험은 한국 아주대 병원으로부터 임상적으로 장형 위암으로 진단받은 21명의 환자와 만성 위염으로 진단받은 24명의 환자들로부터 동의를 받아서 실험과 연관성 분석이 수행되었다. 45명의 환자 모두가 *Helicobacter pylori*에 감염된 환자이다. 사용된 군집의 크기는 다음 표 1과 같다.

표 1

그룹	정상인	환자군	총합계
조사인원	24	21	45

[0039] 2. 게놈 DNA의 준비

[0040] 게놈 DNA(genomic DNA; gDNA)를 준비하기 위하여 장형 위암 환자와 만성 위염 환자들로부터 혈액을 채취하였다. 전체 정맥혈의 5ml은 각 지원자들의 전주정맥으로부터 얻었고 15% EDTA가 들어있는 Vacutainer tube로 옮겨졌다. 게놈 DNA를 추출하기 위하여, 혈액을 원심분리 실험관으로 옮기고 낮은 농도의 염 성분 완충용액 (low-salt buffer solution(10mM Tris-HCl [pH 7.6], 10mM KCl, 10mM MgCl₂, and 2mM EDTA))을 5ml 첨가 하였다. 그리고 Nonidet P-40을 첨가하였다. 얻어진 용액을 잘 섞어준 후, 상온에서 10분간 2,200 rpm으로 원심분리 하였다. 이 때 얻어진 덩어리는 고농도의 염 성분 완충용액(high-salt buffer solution(10mM Tris-HCl [pH 7.6], 10mM KCl, 10mM MgCl₂, 0.4M NaCl, and 2mM EDTA))으로 재현탁 하였다. 이 게놈 DNA는 50ul의 10% SDS와 섞은 후, 55℃에서 밤새 반응시켰다.

[0041] 반응 후에, 각 게놈 DNA들은 1.5ml 용량의 원심분리 실험관으로 옮겨, 6M NaCl을 0.3ml 첨가하였다. 게놈 DNA를 잘 섞어준 후, 12,000rpm에서 10분간 원심분리하였다. 원심분리 후, 게놈 DNA가 들어있는 상층을 모아서 새로운 시험관으로 옮기고, 상온에서 동량의 100% 에탄올을 첨가하였다. 게놈 DNA들이 침전될 때까지 잘 섞어 주고, 침전이 되면 70% 에탄올로 게놈 DNA들을 세척하였다. 이 과정이 끝나면 실험관 뚜껑을 열어, 용액성분이 마르도록 놓아둔다. 그리고 TE 완충용액(pH 8.0)을 넣어 게놈 DNA를 재현탁하였다.

[0042] 준비된 게놈 DNA는 Infinium II Assay를 위해 Quant-iTTM PicoGreen dsDNA reagent (Molecular Probes, Invitrogen)을 사용하여 750ng/15uL의 농도로 조정하였다.

[0043] 3. 전체 게놈의 유전자형 분석 - Infinium II Assay

[0044] 전체 게놈의 증폭(Whole-genome amplification(WGA)) 게놈 DNA (750ng/15uL)를 상온(22℃)에서 녹인 후, 0.8mL micotiter plate의 각 well에 옮겼다. 0.1N NaOH를 15uL 첨가하고 상온에서 10분간 반응시켰다. 반응이 끝나면 270uL의 MP1 용액과 300uL의 AMM 용액을 차례로 첨가한 후, 뒤집어가며 잘 섞어준다. 280g에서 원심분리하여 37℃ 오븐에서 20시간 동안 반응시켰다.

[0045] 증폭된 게놈 DNA의 단편화와 정제 전체 게놈 DNA 증폭 단계가 끝나면 50g에서 1분간 원심분리 한 후, 150uL 씩 3개의 여분의 well로 옮겼다. 결과적으로 4개의 well이 같은 게놈 DNA가 된다. 각 well에 FRG 용액을 50uL 첨가한 후, 1600rpm으로 1분 동안 와동(vortex)하였다. 상온에서 50g로 원심분리한 다음, 37℃에서 1시간 동안 반응시켰다. 게놈 DNA를 50g에서 원심분리 하고, PA-1 용액을 100uL 첨가하고 1600rpm으로 와동하여, 상온에서 50g

로 원심분리 한 후, 37°C에서 5분 동안 방치하였다. 상온에서 50g로 1분 동안 원심분리하고, 100% 아이소프로판올을 300uL 첨가하여 4°C에서 30분 동안 반응시켰다. 4°C에서 3000g로 20분 동안 원심분리하고 게놈 DNA 덩어리를 확인한 다음, microtiter plate를 뒤집어 상층을 제거하였다. 상온에서 1시간 동안 게놈 DNA를 말린 다음, RA1 용액을 42uL 넣고, 48°C 오븐에서 1시간 동안 반응시켰다.

[0046] *Sentrix^R BeadChip*의 준비와 활성화 증폭된 게놈 DNA를 RA1 용액에서 재현탁하는 동안, (*Sentrix^R*) *BeadChip*을 준비하였다. *BeadChip*을 꺼내어 게놈 DNA를 첨가하기 전까지 PB2 용액이 든 혼성화 챔버(Hybridization chamber)에 넣어 두었다.

[0047] 혼성화 증폭된 게놈 DNA의 재현탁 (48°C, 1시간 반응)이 끝나면, 1800rpm으로 1분 동안 와동하고 280g에서 원심분리한 다음, 95°C heat block에서 20분 동안 증폭된 게놈 DNA를 변성화 한 후, 280g로 원심분리하였다. 4개의 well로 나누어진 게놈 DNA를 하나의 well로 모으고, 280g로 원심분리하였다. 혼성화 챔버(Hybridization Chamber)로부터 *BeadChip Chamber*를 꺼내어 게놈 DNA를 넣어준 다음, *BeadChip Chamber*를 다시 혼성화 챔버(Hybridization Chamber)에 넣고, 48°C에서 18시간 동안 적정속도로 흔들어가면 반응시켰다.

[0048] 단일 염기 신장(*Single Base Extension; SBE*)에 의한 염색 18시간의 혼성화가 끝나면, *BeadChip* 을 꺼내어 WB1 용액과 PB1 용액에 차례로 세척한 후, 150uL의 RA1 용액을 흘려 보냈다. XC1 용액 450uL와 XC2 용액 450uL, TEM 용액 200uL에 차례로 각각 10분, 5분, 15분 동안 반응시키고, 450uL의 95% 포름아마이드/1mM EDTA를 흘려 보냈다. 450uL의 XC3 용액을 흘려보내고, Illuminas 프로토콜에 따라 LTM과 ATM 용액으로 게놈 DNA를 염색하였다. *BeadChip chamber*를 분해하여 *BeadChip*을 PB1 용액과, XC4 용액을 이용하여 차례로 세척하고, vacuum desiccator를 이용하여 말린다. 마지막 단계로 Illumina *BeadArray reader*을 사용하여 스캔함으로써 각 *BeadChip*에 대한 이미지 파일을 저장하였다.

[0049] 유전자형 분석 이미지 파일로부터 유전자형의 분석은 *Bead Studio Version 3.2 software*를 사용하여 각 단일염기다형에 대한 유전자형을 정하였다.

[0050] 4. 통계학적 분석

[0051] 단일염기다형들을 구성하는 대립인자들에 대한 집단내의 유전학적 분포여부에 대하여 하아디-와인버그 평형 검정을 하였다. 여기에서 단일염기다형의 기준이 되는 HWE p value > 0.05와 작은 대립인자 빈도(Minor Allele Frequency; MAF) > 0.1의 기준을 적용하였다. 장형 위암 환자군과 만성 위염 환자군간의 출현빈도를 이용하여 질병에 대한 각 단일염기다형들의 연관성분석은 유의수준을 < 0.05로 정하였으며, SAS software version 9.1.3를 사용하여 분석하였다. 연관성분석은 대립인자들 간의 장형 위암 환자군과 만성 위염 환자군에서의 출현빈도로 나누어 유의성 있는 차이를 비교하였으며, 유의수준을 < 0.05로 하였다. 오즈 비(Odds Ratio)는 로지스틱 회귀모형을 통하여 산출하였으며 우성 (Dominant), 열성 (Recessive) 및 공우성(Co-Dominant) 모형을 적용하였다. 각 모형에 따른 질병과의 유의성을 분석함에 있어 다수 대립인자를 A1이라 하고 소수 대립인자를 A2라 하면 우성 (Dominant) 모형은 A1A1의 빈도와 A1A2와 A2A2의 빈도에 대한 합을 비교하는 것이고, 열성 (Recessive) 형질 모형은 A1A1과 A1A2의 합과 A2A2의 빈도를 비교하는 것이고, 공우성 (Co-Dominant) 형질 모형은 A1A1, A1A2 및 A2A2에 대한 각각의 빈도를 비교하는 것이다.

[0052] 5. 결과

[0053] (1) 장형 위암 연관 단일염기다형 선별

[0054] 하기 표 2는 51만개의 단일염기다형들 중 장형 위암과 강하게 연관된 SNP 집합을 선별한 결과 중 가장 높은 점수를 나타낸 6개 집합을 나타낸 것이다.

표 2

단일염기다형 집합	Accuracy	Sensitivity	Specificity
rs12117796 rs16841280 rs2333715 rs3820380 rs10158553	97.78	100	95.83
rs12117796 rs16841280 rs971081 rs3820380 rs10158553	97.78	100	95.83

rs12117796 rs16841280 rs6691829 rs3820380 rs10158553	97.78	100	95.83
rs12117796 rs16841280 rs3860304 rs3820380 rs10158553	97.78	100	95.83
rs12117796 rs16841280 rs1171386 rs3820380 rs10158553	97.78	100	95.83
rs12117796 rs16841280 rs3795593 rs3820380 rs10158553	97.78	100	95.83

[0056] 표 2는 51만개의 단일염기다형들 중 장형 위암과 강하게 연관된 단일염기다형 집합을 찾기 위해 χ^2 value에 의해 상위 500개의 단일염기다형만을 추출한 후 forward selection 방법을 이용하여 장형 위암과 연관된 단일염기다형 집합을 추출한 결과 중 상위 6개 집합이다. forward selection 방법의 단계는 다음과 같다. 먼저 n 개의 단일염기다형에 대해 하나의 단일염기다형으로 구성된 n 개의 집합에 대한 인식률을 구하고 가장 높은 인식률을 보이는 집합 s_1 을 선택하고 그 다음으로, 전단계에서 선택된 집합 s_i 과 s_j 에 포함되지 않은 각 단일염기다형 s_j 에 대해 $s_i \cup \{s_j\}$ 를 구성하여 그 인식률을 구하고 가장 높은 인식률을 보이는 집합을 s_{i+1} 로 선택한다. i 가 n 보다 작을 경우 두번째 단계를 반복하여 수행하고 i 가 n 일 경우 s_k ($1 \leq k \leq n$) 중에서 가장 높은 인식률을 보이는 집합을 결과로 출력한다. 이 실험을 위한 인식기로는 결정 트리(decision tree)를 사용하였다. 그리고 자료의 수가 너무 작은 관계로 인식률을 평가함에 있어 인식기의 훈련과 평가를 위한 자료를 따로 나누어 사용하는 것이 불가능하므로 LOOCV 방법을 이용하였다. LOOCV 방법은 주어진 자료 중에서 각 자료 P_i 를 제외한 자료로 인식기를 훈련시키고 남은 자료 P_i 에 대해 훈련된 인식기를 이용하여 환자여부를 판별하고 그 결과의 맞고 틀림을 평가하는 과정을 모든 자료에 대해 수행하여 인식기가 환자여부를 맞게 판별한 수와 틀리게 판별한 수를 이용하여 인식률을 산정하는 것을 말한다. 표 2에 있어서, 각 칼럼이 의미하는 바는 다음과 같다.

- [0057] · 단일염기다형 집합은 forward selection 방법에 의해 추출된 단일염기다형의 집합이다.
- [0058] · Accuracy는 모든 샘플에 대해 바르게 판별한 비율을 말한다.
- [0059] · Sensitivity는 환자를 환자로 바르게 판별한 비율을 말한다.
- [0060] · Specificity는 정상을 정상으로 바르게 판별한 비율을 말한다.

[0061] (2) ZMPSTE24, GPR37L1, LMX1A, SRGAP2, FAM5C, EDEM3 유전자에 존재하는 유의성을 나타내는 단일염기다형

[0062] 하기 표 3는 상기 만성 위염 환자군과 장형 위암 환자군이 참여하여 장형 위암과의 연관성을 분석한 결과 통계학적으로 유의성을 나타낸 서열번호 1 내지 10의 다형성 서열의 특성을 나타낸 것이다.

표 3

서열번호	Reference SNP ID	대립인자	위치	유전자	SNP역할	아미노산 변화
1	rs12117796	G>A	40509484	ZMPSTE24	intron	변화없음
2	rs16841280	A>G	163458326	LMX1A	intron	변화없음
3	rs3795593	G>A	200358669	GPR37L1	5'UTR	변화없음
4	rs3820380	G>A	200363438	GPR37L1	intron	변화없음
5	rs10158553	A>C	204687870	SRGAP2	intron	변화없음
6	rs1171386	A>G	188578948	FAM5C	intron	변화없음
7	rs3860304	A>G	188501670	FAM5C	intron	변화없음
8	rs6691829	A>G	188456354	FAM5C	intron	변화없음
9	rs971081	G>A	182983080	EDEM3	intron	변화없음
10	rs2333715	A>G	182962655	EDEM3	intron	변화없음

[0064] · 서열번호는 각 단일염기다형을 나타내는 서열번호이다.

[0065] · Reference SNP ID는 표준단일염기다형의 명칭 (Reference SNP ID; rs)을 의미하며, 이는 인간게놈프로젝트 후, 미국 국립보건원 산하 생물공학정보연구소 (NCBI)의 데이터베이스에서 각 단일염기다형을 구분하기 위하여 붙인 이름이다.

- [0066] · 대립인자는 각 단일염기다형에서의 Major allele과 Minor allele을 나타낸다.
- [0067] · 위치는 염색체상에서 표준염기서열의 위치를 의미한다(NCBI Genome build 36.3).
- [0068] · SNP 역할은 상기 단일염기다형들이 유전자상에서 어디에 위치하고 있는지를 나타내주고 있다. 따라서 각각의 단일염기다형들이 유전자의 발현과 기능에 있어서 어떻게 작용하는지를 알 수 있다.
- [0069] · 아미노산의 변화는 상기 단일염기다형들이 유전자가 발현되어 단백질을 생성할 때, 그 단백질을 이루는 아미노산에 변화를 주고 있는지에 대한 설명이다.

[0070] (3) 유전자형(genotype) 및 위험대립인자 분석

[0071] 하기 표 4는 표 2에서 추출된 장형 위암 연관 단일염기다형 집합에 속하는 10 개의 단일염기다형(서열번호 1 내지 10)의 DNA 서열의 대립인자 빈도와 각 유전자형에 대한 만성 위염 환자군과 장형 위암 환자군 각각의 빈도를 나타내며 표 5는 표 4의 유전자형 빈도를 토대로 진행한 하아디-와인버그 평형 검정(Hardy-Weinberg Equilibrium Test) 결과와 장형 위암 연관성 분석[즉, 케이스-컨트롤 연관 분석(Case-control association study)] 결과로서, 하기 연관성 분석에 의해 장형 위암에 영향을 미칠 수 있는 단일염기다형들과 이들의 유전자형을 확인할 수 있다.

표 4

대립인자형 빈도 및 유전자형 빈도

서열 번호	Reference SNP ID	대립인자형 빈도		유전자형	만성위염군		장형위암군	
		Major allele	Minor allele		유전자형빈도(%)	유전자형빈도(%)		
1	rs12117796	0.811(G)	0.189(A)	GG	21 (87.5)	10 (47.6)		
				AG	3 (12.5)	8 (38.1)		
				AA	0 (0.0)	3 (14.3)		
2	rs16841280	0.856(A)	0.144(G)	AA	23 (95.8)	9 (42.9)		
				AG	1 (4.2)	12 (57.1)		
				GG	0 (0.0)	0 (0.0)		
3	rs3795593	0.722(G)	0.278(A)	GG	7 (29.2)	16 (76.2)		
				AG	14 (58.3)	5 (23.8)		
				AA	3 (12.5)	0 (0.0)		
4	rs3820380	0.722(G)	0.278(A)	GG	7 (29.2)	16 (76.2)		
				AG	14 (58.3)	5 (23.8)		
				AA	3 (12.5)	0 (0.0)		
5	rs10158553	0.678(A)	0.322(C)	AA	16 (66.7)	6 (28.6)		
				AC	8 (33.3)	9 (42.9)		
				CC	0 (0.0)	6 (28.6)		
6	rs1171386	0.756(A)	0.244(G)	AA	19 (79.2)	6 (28.6)		
				AG	5 (20.8)	13 (61.9)		
				GG	0 (0.0)	2 (9.5)		
7	rs3860304	0.756(A)	0.244(G)	AA	19 (79.2)	6 (28.6)		
				AG	5 (20.8)	13 (61.9)		
				GG	0 (0.0)	2 (9.5)		
8	rs6691829	0.756(A)	0.244(G)	AA	19 (79.2)	6 (28.6)		
				AG	5 (20.8)	13 (61.9)		
				GG	0 (0.0)	2 (9.5)		
9	rs971081	0.511(G)	0.489(A)	GG	2 (8.3)	10 (47.6)		
				AG	13 (54.2)	9 (42.9)		
				AA	9 (37.5)	2 (9.5)		
10	rs2333715	0.511(A)	0.489(G)	AA	2 (8.3)	10 (47.6)		
				AG	13 (54.2)	9 (42.9)		
				GG	9 (37.5)	2 (9.5)		

[0072]

- [0073] 표 4에 있어서, 각 칼럼이 의미하는 바는 다음과 같다.
- [0074] · 서열번호는 각 단일염기다형을 나타내는 서열번호이다.
- [0075] · Reference SNP ID는 표준단일염기다형의 명칭 (Reference SNP ID; rs)을 의미하며, 이는 인간게놈프로젝트 후, 미국 국립보건원 산하 생물공학정보연구소 (NCBI)의 데이터베이스에서 각 단일염기다형을 구분하기 위하여 붙인 이름이다.
- [0076] · 대립인자형 빈도(allele frequency) 란에서는 각각의 단일염기다형을 구성하는 두 개의 뉴클레오티드 염기를 표시하며, 집단 내에서 관찰되는 각각의 대립인자들의 빈도를 나타내고 있다. 여기서 대립인자란 단일염기다형 부위에 실제로 존재하고 있는 뉴클레오티드의 염기서열을 뜻하며, 하나의 단일염기다형은 그 특성상 두 개의 대립인자를 가질 수 있다. 이들 대립인자 빈도의 고저에 따라 큰 대립인자(Major allele)와 작은 대립인자(Minor allele)로 나뉜다. 특히 작은 대립인자의 빈도(Minor allele frequency; MAF)는 0.1을 기준으로 하여 단일염기다형과 단일염기변이(point mutation)을 나누는 유전학적 기준으로 사용된다. 이러한 기준은 다형성 자체의 안정성을 뜻하기 때문에(즉, 단일염기다형이 일어나는 부위는 일정하다라는 뜻), 유전적 표지인자로서의 사용 가능성을 대변해 준다.
- [0077] · 유전자형은 각 단일염기다형을 형성하는 각각의 대립인자들로 이루어지며, 단일염기다형 부위가 존재하는 유전자좌(locus)에 어떠한 대립인자들이 있는지를 나타내 주고 있다.
- [0078] · 유전자형 빈도(genotype frequency) 란은 단일염기다형을 형성하는 각각의 대립인자들로 이루어지며, 단일염기다형 부위가 존재하는 유전자좌(locus)에 어떠한 대립인자들이 있는지를 나타내 주고 있다. 여기서 큰 대립인자들로만 구성되어 있는 유전자형을 큰 동형유전자형(Major homotype), 큰 대립인자와 작은 대립인자로 이루어진 유전자형을 이형유전자형(Heterotype), 그리고 작은 대립인자들로만 이루어진 유전자형을 작은 동형유전자형(Minor homotype)이라고 한다. 숫자들은 각각의 유전자형에 해당하는 사람들의 수를 만성 위염 환자군과 장형 위암 환자군으로 나누어 나타낸다. 이를 관측치라고 할 수 있다.

표 5

케이스-컨트롤 연관 분석

서열 번호	Reference SNP ID	HWE	분석모델	OR (95% CI)	Logit p value	Exact p value	위험 대립인자
1	rs12117796	0.175	대립인자	7.499(1.977-28.445)	0.0031	9.784E-04	A
			공우성	6.548(1.630-26.309)	0.0081	0.0019	
			우성	7.700(1.749-33.898)	0.0069	0.0043	
			열성	>999.999(<0.001->999.999)	0.9748	0.0937	
2	rs16841280	0.257	대립인자	18.798(2.323-152.091)	0.0060	3.230E-04	G
			공우성	30.664(3.464-271.433)	0.0021	9.663E-05	
			우성	30.664(3.464-271.433)	0.0021	9.663E-05	
			열성	-	-	-	
3	rs3795593	0.726	대립인자	0.189(0.063-0.566)	0.0029	0.0012	G
			공우성	0.141(0.039-0.506)	0.0026	7.554E-04	
			우성	0.129(0.034-0.489)	0.0026	0.0017	
			열성	<0.001(<0.001->999.999)	0.9753	0.1426	
4	rs3820380	0.726	대립인자	0.189(0.063-0.566)	0.0029	0.0012	G
			공우성	0.141(0.039-0.506)	0.0026	7.554E-04	
			우성	0.129(0.034-0.489)	0.0026	0.0017	
			열성	<0.001(<0.001->999.999)	0.9753	0.1426	
5	rs10158553	0.365	대립인자	5.000(1.894-13.198)	0.0012	6.135E-04	C
			공우성	4.923(1.671-14.500)	0.0038	4.807E-04	
			우성	5.000(1.402-17.830)	0.0131	0.0097	
			열성	>999.999(<0.001->999.999)	0.9638	0.0067	
6	rs1171386	0.578	대립인자	5.848(1.923-17.786)	0.0019	8.182E-04	G
			공우성	8.766(2.318-33.153)	0.0014	4.021E-04	
			우성	9.500(2.423-37.248)	0.0012	7.276E-04	
			열성	>999.999(<0.001->999.999)	0.9795	0.2121	
7	rs3860304	0.578	대립인자	5.848(1.923-17.786)	0.0019	8.182E-04	G
			공우성	8.766(2.318-33.153)	0.0014	4.021E-04	
			우성	9.500(2.423-37.248)	0.0012	7.276E-04	
			열성	>999.999(<0.001->999.999)	0.9795	0.2121	
8	rs6691829	0.578	대립인자	5.848(1.923-17.786)	0.0019	8.182E-04	G
			공우성	8.766(2.318-33.153)	0.0014	4.021E-04	
			우성	9.500(2.423-37.248)	0.0012	7.276E-04	
			열성	>999.999(<0.001->999.999)	0.9795	0.2121	
9	rs971081	0.884	대립인자	0.246(0.102-0.594)	0.0018	0.0011	G
			공우성	0.206(0.069-0.611)	0.0044	4.785E-04	
			우성	0.100(0.019-0.538)	0.0073	0.0034	
			열성	0.175(0.033-0.937)	0.0417	0.0271	
10	rs2333715	0.884	대립인자	0.246(0.102-0.594)	0.0018	0.0011	A
			공우성	0.206(0.069-0.611)	0.0044	4.785E-04	
			우성	0.100(0.019-0.538)	0.0073	0.0034	
			열성	0.175(0.033-0.937)	0.0417	0.0271	

[0079]

[0080]

표 5는 하나의 단일염기다형을 이루는 두 개의 대립인자와 이로 인하여 형성되는 3가지 형태의 유전자형에 대한 만성 위염 환자군-장형 위암 환자군 간의 빈도차이를 비교하여 어떠한 대립인자가 또는 어떠한 유전자형이 질병과 연관되어 있는지를 분석한 결과이다. 이러한 연관성 분석은 대립인자형(allele model), 공동-우성 유전자형 (Co-dominant model), 우성 유전자형(Dominant model), 그리고 열성 유전자형(Recessive model)이라는 네 가지 항목에서 실시하였다. 표 5에 있어서 각 칼럼이 의미하는 바는 다음과 같다.

[0081]

· 서열번호는 각각의 단일염기다형을 나타내는 번호이다.

[0082]

· Reference SNP ID는 표준단일염기다형의 명칭 (Reference SNP ID; rs)을 의미하며, 이는 인간게놈프로젝트 후, 미국 국립보건원 산하 생물공학정보연구소 (NCBI)의 데이터베이스에서 각 단일염기다형을 구분하기 위하여 붙인 이름이다.

[0083]

· HWE는 하아디-와인버그 평형 (Hardy-Weinberg Equilibrium)의 상태를 나타내는 것이다. 카이제곱 (df=1) 검정에서 chi-value = 3.84(p-value=0.05, df=1)을 기준으로, 3.84보다 큰 경우에는 하아디-와인버그 평형 HWE (Hardy-Weinberg Equilibrium)으로 판단하고, 3.84보다 작은 경우에는 하아디-와인버그 비평형 (Hardy-Weinberg Disequilibrium)으로 판단하였다. 이는 만성 위염 환자군과 장형 위암 환자군 모두에서 관찰된 각 단일염기다형들의 유전적 표지인자로서의 기능과 이들로 이루어지는 유전자형의 분포가 일반적인 것들인지를 나타낸다.

[0084]

· 오즈 비율 (odds ratio, OR)은 장형 위암 환자 중에서 위험 대립인자를 가질 오즈에 대한 만성 위염 환자군 중에서 위험 대립인자를 가질 오즈의 비율을 나타낸다. 95% CI는 오즈 비율의 95% 신뢰구간을 나타내는 것으로, (하한 신뢰구간, 상한 신뢰구간)을 나타낸다. 신뢰구간은 1을 포함하는 경우에는 질병과의 연관성을 유의하다고 판단할 수 없다. 오즈비율이 1보다 크고 95%신뢰구간에 1이 포함되어 있지 않은 경우 소수 대립인자가 질환군에서 더 높은 빈도를 가지며, 오즈비율이 1보다 작은 경우에는 다수 대립인자가 질환군에서 더 높은 빈도를 가진다.

다.

- [0085] · 대립인자형에 대한 연관성 분석에서 Logit_p value는 오즈비에 대한 통계적 유의성을 나타낸다. 이는 오즈비 만큼 만성 위염 환자군과 장형 위암 환자군간에 차이에 대한 유의성을 뜻한다. 여기서 유의성있는 차이가 있다고 판단하는 기준은 Logit_p value가 0.05 이하의 값을 나타낼 경우이다. 이러한 값은 로지스틱 회귀(logistic regression) 분석을 통하여 얻었다.
- [0086] · 대립인자형에 대한 연관성 분석에서 Exact p value 또한 Logit_p value 처럼 만성 위염 환자군과 장형 위암 환자군간에 차이에 대한 유의성을 뜻한다. 이 역시 p value의 값이 0.05 이하일 때, 유의성있는 차이를 인정한다. Exact p value는 피셔의 정밀 검정(Fisher's exact test)라는 통계 방식을 사용하여 얻어진 값인데, 이는 만성 위염 환자군과 장형 위암 환자군의 비교시 필요되는 기대빈도가 5 이하의 수가 있을 때, 정밀한 검정을 하기 위해 개발된 통계기법이다. 따라서 표 5에서 보여주는 데이터에는 매우 적합한 검정법이라 할 수 있다.
- [0087] · 위험 대립 인자 (risk allele)는 p-value에 유의성이 있는 경우 오즈비가 1 보가 크면 소수 대립인자, 오즈비가 1 보다 작으면 다수 대립인자로 하였다.
- [0088] · 각 네가지 분석모델에 따른 질병과의 유의성을 분석함에 있어 다수 대립인자를 A1이라 하고 소수 대립인자를 A2라 하면 우성 (Dominant) 모형은 A1A1의 빈도와 A1A2와 A2A2의 빈도에 대한 함을 비교하는 것이고, 열성 (Recessive) 형질 모형은 A1A1과 A1A2의 함과 A2A2의 빈도를 비교하는 것이고, 공우성 (Co-Dominant) 형질 모형은 A1A1, A1A2 및 A2A2에 대한 각각의 빈도를 비교하는 것이다.
- [0089] 대립인자 (Allele)에 대한 연관성 분석은 대립형질들이 만성 위염 환자군과 장형 위암 환자군 사이에서 출현되는 빈도수의 비율을 통하여 위험 대립인자가 장형 위암에 어느 정도의 연관성 (위험도)를 가지고 있는지 확인할 수 있다. 95% 신뢰구간에 1을 포함하고 있을 경우 유의성이 없기 때문에, 상기 표 5로부터 확인할 수 있는 바와 같이, 서열번호 1 내지 10의 다형성 부위의 뉴클레오티드 서열이 각각 A, G, G, G, C, G, G, G, G, A(위험 대립인자)인 것이 하나 이상 존재하는 경우, 장형 위암에 걸릴 가능성이 높은 것으로 판단할 수 있다
- [0090] 또한, 서열번호 1 내지 10은 대립인자와 공우성과 우성 모델에서 유의성을 보이며, 서열번호 1 내지 2, 서열번호 5 내지 8은 오즈비가 > 1로 유전자형(genotype)이 소수대립인자(Minor Allele)를 가지고 있는 사람이 다수대립인자(Major Allele)를 가지고 있는 사람보다 장형 위암에 걸릴 위험이 더 높은 것으로 나타났고 또한 유전자형이 소수 대립인자를 하나라도 가지고 있는 사람이 다수대립인자 둘로 구성된 사람보다 장형 위암에 걸릴 위험이 높은 것으로 나타났다. 반면 서열번호 3 내지 4, 서열번호 9 내지 10은 오즈비가 < 1로 유전자형(genotype)이 다수대립인자(Minor Allele)를 가지고 있는 사람이 소수대립인자(Major Allele)를 가지고 있는 사람보다 장형 위암에 걸릴 위험이 더 높은 것으로 나타났고 또한 유전자형이 다수 대립인자 둘을 가지고 있는 사람이 소수 대립인자를 하나라도 가진 사람보다 장형 위암에 걸릴 위험이 높은 것으로 나타났다.
- [0091] 상기 결과로부터, 서열번호 1 내지 10의 다형성 부위가 각각 유전자형이 A/A와 A/G (서열번호 1), G/G와 A/G (서열번호 2), G/G (서열번호 3), G/G (서열번호 4), C/C와 A/C (서열번호 5), G/G와 A/G (서열번호 6), G/G와 A/G (서열번호 7), G/G와 A/G (서열번호 8), G/G (서열번호 9) 및 A/A (서열번호 10)가 하나 이상 존재하는 경우 장형 위암에 걸릴 위험도가 높은 것으로 판단할 수 있다. 또한 서열번호 1의 경우는 A, 서열번호 2는 G, 서열번호 3은 G, 서열번호 4는 G, 서열번호 5는 C, 서열번호 6은 G, 서열번호 7은 G, 서열번호 8은 G, 서열번호 9는 G, 서열번호 10은 A가 각각 만성 위염군에 비해 장형 위암군에서 유의적으로 높은 출현빈도로 나타났으므로, 이는 각각의 대립인자들이 장형 위암에 감수성(susceptibility)을 가진다고 예측할 수 있다. 상기 위험 대립 인자를 갖는 핵산 서열이 하나의 검체에서 많이 검출되면 될수록 위험군에 속하는 확률은 더 높은 것으로 판단할 수 있다. 반대로, 서열번호 1의 경우는 G, 서열번호 2는 A, 서열번호 3은 A, 서열번호 4는 A, 서열번호 5는 A, 서열번호 6은 A, 서열번호 7은 A, 서열번호 8은 A, 서열번호 9는 A, 서열번호 10은 G가 각각 장형 위암군에 비해 만성 위염군에서 유의적으로 높은 출현빈도로 나타났으므로, 이는 각각의 대립인자를 가진 사람들은 장형 위암에 저항성을 가지고 있다고 예측할 수 있다. 따라서 본 발명에 있어서 서열번호 1 내지 10의 단일염기다형은 장형 위암과 관련된 진단에 효과적으로 이용될 수 있다. 구체적으로는 장형 위암의 진단을 위한 프라이머 또는 프로브로의 목표 부위(target site)로서 사용될 수 있다.
- [0092] 상기와 같은 분석의 결과를 볼 때, 각 대립인자는 각 개체에 있어서 동형접합자 (homozygote) 또는 이형접합자 (heterozygote)의 형태로 존재할 수 있다. 멘델의 유전법칙과 하디-와인버그 (Hardy-Weinberg) 법칙에 의하면, 집단을 구성하는 대립인자의 조성은 일정한 빈도로 유지되며, 통계적으로 유의한 경우 생물학적 기능상의 의미를 부여할 수 있다. 본 발명의 단일염기다형은 장형 위암 환자와 만성 위염 환자 사이에서 통계적으로 유의한

수준으로 차이를 나타내며 출현되어, 이들 단일염기다형들은 장형 위암의 진단에 사용될 수 있음을 알 수 있다.

서열 목록

- <110> College of Medicine Pochon CHA University Industry-Academic Cooperation Foundation
- <120> Polynucleotides derived from chromosome 1 comprising single nucleotide polymorphism, microarrays and diagnostic kits comprising the same, and analytic methods using the same

- <130> PN0243

- <150> KR10-2008-0122469
- <151> 2008-12-04

- <160> 10

- <170> Kopatent In 1.71

- <210> 1
- <211> 701
- <212> DNA
- <213> Homo sapiens

```

<400> 1
tcttttcccc ctgatttttc cctcctcctac atctcctatt cctcttctct tccatactc 60

tcaagtgctt cacaacacaa aactacaac gcagtcaaaa tggctctgtc agggccaggc 120

acagtggctt atgcctgtaa tcttggcact ttgggaagcc aagggtggag gattgcttaa 180

gcccaagagt ttaagaccag cctgggtaac acagtgagac tctgtctcta caataataa 240

tttttaaaaa acccaaatga aaaaaccaa agtggctctgt tcagtgacc cagaaattat 300

catctatcag ctcttcagct ccactgatga ggtttatatt attttagaga tagggcctcc 360

ctgtgttgcc caggctggag tacagtggct atttcacagg catgatcata gtgcactaca 420

gctttgagct cctgggctca agcagtcctc ctgccttggc ctctgagta gctgggacta 480

caggtgtgtg acaccatgcc yaagctgagg tttatatttag tctgaattac ataatgctca 540
    
```


ttgtttcatt ttccattgat gggatcaaga tcagaaaatg ggcgctagt ggatttacag 600

glatctattg taiggcaaat taaagtaaga tgaaaaagag cagagaaaat gctctaaaaa 660

tgcttaagga agtgctatct caaccacatt gtacaattat g 701

<210> 2
 <211> 201
 <212> DNA
 <213> Homo sapiens

<400> 2
 taggttcaaa tgaataaggt ttgagtgtgt ttgtcctaatt ttgggaagta acttaagtgg 60

atcgaataacc atctctcttt ggagtagtcc aattctgcta yattaattac tagtcttttg 120

aaggataaat tatctgagaa aattaaattg tttctaggca agggttggaa gacaaagggt 180

taattatctc tccaaattga a 201

<210> 3
 <211> 801
 <212> DNA
 <213> Homo sapiens

<400> 3
 ctacgaccta gtaagtgttc agtaagtgc agtgatggtt attgctgagt tttgaatgga 60

ggagctgcct tagaatcagg agacctgcgc cccagttccc attctgcccc acctccctgt 120

gtcaccctag gcaggccaca tttctcctcct agtttcaggg gctgaagca gatgccctct 180

tagggcccca cagcctgaca tgctgtaggg cctgagaagc ggcgtcgtgg aggacgagat 240

gtgtgagggc agcaaagagt gctatgtgtc cagcagaggg cctgccccg cctgtggccg 300

gaggctggga gggagggcag gcgagtgatg ccagacgcct gactggaggc ggatccagcc 360

ggccagctgc ctctctggag cccagctctt gggccccctg yactcacctg ctcttctgg 420

gctggctgtc tctgctcat ccagccatgc ggtggctgtg gcccttgct gtctctcttg 480
 ctgtgatfff ggctgtgggg ctaagcaggg tctctggggg tgccccctg cacctgggca 540
 ggcacagagc cgagaccag gagcagcaga gccgatccaa gaggggcacc gaggatgagg 600
 aggccaaggg cgtgcagcag tatgtccctg aggagtgggc ggaglacccc cggcccattc 660
 accctgctgg cctgcagcca accaagccct tggtagccac cagccctaac cccggcaagg 720
 atgggggcac ccagacagt gggcaggaac tgaggggcaa tctgacagga gcaccagggc 780
 agaggctaca gatccagaac c 801

<210> 4
 <211> 701
 <212> DNA
 <213> Homo sapiens

<400> 4
 gttgattggg aatgtgatga cacgtaaagt gcccgacca taggaaatgc tcattaaatg 60
 gagttgttac cattattata attccttct tctttccttc cttccttct tctttccttc 120
 tttccttct tctttccttc cttgtttcca cagggttttg ctcgtcacc caggctggag 180
 tgcagtgtcg tgatcttggc tcactgcagc ctcagcctcc taggctcaag ccatcctct 240
 acctaaagttt cccgagtagt tgggaccaca ggcgtgcctc accacgacag gctatffff 300
 tgtatffffa gtagagacag ggttccatca tgttgcccag gctggccttg aactcctggg 360
 ctcaagtgat ctgctgcct cggcctctca aagtgctggg attacagtg tgagtcaccg 420
 caccggccc attattataa tttctatgct taccctcaa ttcattgtct aactgagcca 480
 gaagccaggt gtcttctct rtcccatag attttgacca agtcttctc tcttctctca 540
 catcctgccc acaggtctcc tctctgggag tcacgaactt cagcctctgt gcctgggca 600

ttgaccgctt ccacgtggcc accagcacc tgcccaaggt gaggcccatc gagcgggtgcc 660

aatccatcct ggccaagttg gctgtcatct gggtagggctc c 701

- <210> 5
- <211> 601
- <212> DNA
- <213> Homo sapiens

<400> 5
 ggtctcagcc tactggatga tgagctccag ggagctgtcc ttgtgcatgt gacttagagc 60

tttggttttc ccccttgcat gagtggtgaa aagctgtaga atagcattcc ctctcttcta 120

aagtcaggca gaacctccgc cctcacttct caaacatcaa accactgcag gctctggtec 180

agtgccggg gcacagagac ccctacagcc ctggccctg agtcccagcc cagtgggtea 240

gagtcttga ggggtgtggc tcacggggac aggtgccttc ctccagtttc ctccaagata 300

ktctccaac ctaccacct gtttcttgig cactcaacat ttattagcag cctctggaaa 360

aaccacaaga ttagtcaggg aacattacta ccgcagactg tgtgggaggc tagaagcacc 420

tgaggaaca gagtcagcaa ccttgggttg agaggacct tctaaccgct ggattcacac 480

ccacctgga ggctcatggc ctctctctc cttctttgg gaagtgagag aagtgccagc 540

cccgggagct aaggcagtaa ggaggtgagt gagggtggta gaaggggaga tagagcacta 600

a 601

- <210> 6
- <211> 1001
- <212> DNA
- <213> Homo sapiens

<400> 6

agacacctcc ctttatgatg tcaatctaag ggctgagcta agggaaataa atgtgattat 60

tgtaagtaac aatacttctg atctcagtta tataaactac atgtacattt ttaggaaaaa 120

aatgaataaa tatagatggt taaacatatg aagtatatct caaatgataa atgtttgtat 180

ttagctgtgg cctagaataa atacagaatt ttcttatggg ccaaagacag aaaacggtgt 240

taagatcaag tgatgaaaag ctttgaatac aagtttaagg aggtagactt cattcagtag 300

gattcgaaga actatgggg ctccttaaac agaactgagg catagccatt aatgtatct 360

aggaaaaata tctaccatca atcaaagtac aggttgaagt taagatgaac aggcgatgta 420

tcaattattg atttattaaa ctgagggaaa ctgctgcatg tataatgagg aagaaccacc 480

cagggttaga ttttgggaac rgaggaagag tgaattaaaa agttgacaga aaaataaaat 540

gagttaggtt tgaagagta aatattttaa attctgtggc agatatactt ggaagaaggt 600

ggctgtttaa ggctagagtt tgagagagaa gttaagacta cagatgaata aatgcagaaa 660

ctttgcacag taaagtgaac acaaaaacca attagcctga ttttgaatct tgattctgct 720

atgtactatg gtgagctact tgggaaagtc agttatcctt tccaaaactc agcttactta 780

actgtaaaat tgaaaaataa taagcattct gcacaggtat tataaaggaa ggatgacata 840

taaattaagt gcatatttga atgccttgta tatagtacat gctcaacaat catttttact 900

tttattatta tcattacaat aataaaaacc attactagta ttattatatt taatctctaa 960

tttgtagag gtgtagatt aaagaaaatt ggtaacatag a 1001

<210> 7
 <211> 1371
 <212> DNA
 <213> Homo sapiens

<400> 7
 ccaaacaaaa gggttgctga aagaatgaaa acagcaagtt tgacaattat gtgctagatg 60
 agtatccgat aatatgtttg tcggaggatc acaagttctt aatgtctaac ttcaatgaca 120
 tggatTTTTg tcttttggac attgaaaagt tcagcttaat tgtaatacac tttacgtttt 180
 acgtttgaga accaattgct actacctaaa tcctacatag atttcagtgt ttttcacaca 240
 gtgtattaac acacaatgtt tyaatatata ctatgcaggg aaattctgaa tttattacat 300
 atatatacat acctgtatat gcatacacat gctttatctg aattttaaag gtgcagatga 360
 aaactagtgc ttigtcatat ttccatttgg atattctatt ctatggtaaa aatttctgag 420
 taaatgcaat ttctttcttt ttatttattt atttatttat ttttgagacg gagtctcgct 480
 ctgtcgccea ggctggagtg cagtggcggg aacctccgct cactgcaagc tccgcctccc 540
 gggttcacgc cattcttctg ccacagcctc cggagtagct ggaactactg gcgcccgcga 600
 ccacgccccg ctaatttata tatatatata tatatatatt tagtagagac ggggtttcac 660
 cgtgttagcc aggatagtct cgatctctg ccctcatgat ccgcccacct cggccttcca 720
 tagtgcctggg attacaggcg tgagccaccg tgcccggccc tgagtaaaat acagtttcat 780
 tttatgcccc tcaattgttt gttttccata aattaatttt attttgacc atatgagtgt 840
 tatgctaaac ttigttagaa agtcaataaa atctggagaa actgttctgc aaaagctaaa 900
 tatctgacat tcagtaacta cgtgtttaaa gtgaccaata tcccaggcac tttttgaggt 960
 acttgagaaa taictgtgac caaaaataaa accccttgcc cttttggcat atactctcc 1020
 tttgttttac tatattcata agctagacct ttccacctgc agataagtta aaaaaaaga 1080
 ttttaagtgg aaaattgaat aatttgaat attattaatg caggagagga gtcactcaca 1140

atttttgtgg acaagcggaa gttgagcaaa cgagctgaag gaagtgattc caccaccaat 1200
 agctcttcgg tcaactctgga gacgctacat cagctagccg cttcttattt cattgacagg 1260
 gacagcacc ctcggagact tcaccacatt caaattgcat ccaactgccat aaaggttaaga 1320
 atgagtttac gctaaaatta tcaccttcca ttgtttgttt cctctatata c 1371

 <210> 8
 <211> 7297
 <212> DNA
 <213> Homo sapiens

 <400> 8
 aaaatcattt ttaattgatg tgtatTTTTat tgtaattga agaattataa ttgtatacat 60

 ttatgaggta acaactggatg ttattatata tgtatataat gaagaatgag taattcaaac 120

 taattaacat atctaccacc tcaaatactt atTTTTTTat ggtgagaata ttgaaatat 180

 attagaaata taitctgttg gcaatTTTga aacatacatt ttattaacaa gtcaacatgc 240

 tatgctatag atctcaaac ctatTTTTcc ggtctaactg aaactTTgca acatttgatg 300

 atcacatcct cataacctcc cccacctcc ttgtccctc tattcacaa tgtctgcttc 360

 cattatTTTg gttgTTTTa atgctctatt ttgatcaat aaatcttaag ggagcctaca 420

 ttactaact ctccctaata ctatttacc tttcatttg taatttTgt tatttccaac 480

 aacaacaaca aaagcaaaa atTTTgtTga agccatcatt gTgTTTTctc ttgtctggaa 540

 catgTTTTgt caccttcaca cagtactagg ataattcctt tgggaccaac tccattcagg 600

 actgccagtg ctctcattat ttacaacggc aatgcaaata ggggattcag gcacaactaa 660

 gagccaagaa taaagcagac cttagcagtaa agaatcacgg agcaagcata tttactata 720

 aatgaaaaaa attggacaga taaaactgga attaataatt aatccttcaa tgcaattcac 780

agacatttac ccataagaaa caacagcaaa caggaaacca tgacctctc aaaaagacag 840
 aacaaaaatt gaggactga cccaacaag atgctgattt gtgagctctc tgaccaagaa 900
 ttcaaacag cagttttcag gaaactcagt gatctctaag ataacacaga aaagcaattc 960
 tgaaatttat tagagaaatt taacaaagat attaaaataa taaaaaatca aacagaaatc 1020
 atggaataga gaaatacatt tcctgaactg aacaattcat tagaggtctc aacagtagaa 1080
 tggaccaagc agagaaaaga atcagtgagc taaaagactg gctatttgaa aatacatagt 1140
 cagggccagg aatggtggct cacgcctgta atcccagcac tttgggagc cgaggtgggt 1200
 ggatcacgag gtcaagagat cgagaccatc ctggccaaca tggtgaaacc ctgtctctac 1260
 taaaaataca aaaattagct ggatgtggtg gcggatgtct gtagtcccag ctactcagga 1320
 agctgaggca ggagaatcgc ttgaacctgg gaggtggagg tttcagtgag ccaagatctc 1380
 accactgcac tccagcctgg gtgacagagt gagactccaa aaaaaaatt aaaataataa 1440
 taataataat aaattttaaa aaaaggaaa tacacagtca gagggaaaca aacaagcaaa 1500
 caaatgaaaa acatgaaaag gaacaaagac cacctataag atacagaaaa ttacctcaa 1560
 agaccaatt gaagaattat tggattcaa aagtgagctg agcaagatca aggagtggaa 1620
 attttagtca aataaataat gaaaaactgt gcaaagcttg agaaatattc aggtatagaa 1680
 cgatctggga atactagtca gattcaatcc cgataagact acctccaagg catataataa 1740
 ccaaactccc taagatcaat gaaaaagaga gaatcctaaa agcagcaaga gaaaagaagc 1800
 aaaaaggatg agttcatgtc ctttacaggg acatgggtga agctggaaac catcattctc 1860
 agctaactat cgtaagatca gaaaaccaa cactgcatgt tctcaatcat atgtgggagt 1920

tgaacaatga gatcacacgg acacagggag aggaacatca cacatcaggg cctgtcaggg 1980

gttgggggct aggggagggg taacattagg agaatacct aatgttggg acgggttgat 2040

gggtgtagca aaccacatg gcccgtgtat acctatgtaa caaaactgca cgttctgcac 2100

atgtaacgca gaacttaaag tataatTTTT tttaaaaaag aagcaataa cataaaagac 2160

ataaaaggag ctccaatttg tctggcraca aacttctcaa tggcattttc aatgtgttca 2220

aagataaaaa aaattgcaac tcaaaaatat tgtattaggc aaaatcatcc ttaaacatg 2280

aaggagagat aaactatttc ccaataaac aaaaatggag agaatttacc accaccagac 2340

tcactttaca agaatgata aaggaagctc ttcagtctga aagtaaaaac atacaaatat 2400

gccaaaaaaa agttttcaag gcataaatcc cactggtaaa attaagtaca aggacaaact 2460

cagaatactc tgtcacttta attttgggtg gcaattctct tgtaacctca tatgaagccc 2520

aaaagtaaaa tctgtcaaaa acagtaatag ctatagcaac tttctaagag atagtataaa 2580

aatctgtaaa cagaggagtt cgagaccagc ctggccaaca tggtgaaacc ctgtctctac 2640

tacaaacaca aaaattagtc gggcatgggtg gcagatgcct gtaatcccaa ctacttgggt 2700

ggctgaagca agagaatgc tgaacccgg gaggcagagg ttgcagtgag ctgagatgcg 2760

accactacac tccagcctag gcgacagagt gagactccat ctcaaaaaa caacaacaa 2820

aaaaagtaaa ctggggcaac taaaagtaaa catgtgggtg agtggagtta aagtgtaaaa 2880

ttttttgtgt gctttttgtc tttgttgtt ctgttcttag tttgttatcc aaggtaagtt 2940

gtcttctctt tagcataact tactatatct ataagatgat ttttgaagc ctcataactt 3000

gagataaatt catgaaaaat gaaaagcaac aaattaaaac aactaccaa tgaaaatcac 3060

ttaatcacia agaaagacag taagaagaga agaagagact aaaacgaacc agaaaacaag 3120

ccacaaaatg gcagtagtgt gcacttattt atcaataata aactgaatg taaatgatct 3180

cacttcccca attataacgt gtatagtggc tgaatgaata aagaagtaag acccaactat 3240

atgttgccctt caaaaaatcc acttaatgia gaaagataca catagactga gcgtgaaggg 3300

gtggaaaaag atactcgatg caactggaaa acaaaggagc aggagtagct atacctatat 3360

cagacaaaat agattaaaaa tctaagattt aaaaagacat aggaggtcaa tataaaatga 3420

taaaggggtc aatacagcaa gataatataa caattatcaa tagttttatg cccaacaaca 3480

gagtttcccta taagtatata aagcacacat tagtagagct aaaaatatat ataagactat 3540

aataaaataa cagcagggga atttaacacc ctcgatcag taatgaacat atcatccaga 3600

cagaaaaatca agaaagaaag agcacagtta aactacgcac tagatctaata gggccctaga 3660

tgacattaac aaaacatttc acccaactgc tgcagaatat acattctttc cttcagtgca 3720

tggaacattc ttcagaatag gtcatatcitt aagcaacaaa aaagtctgaa taaattaaaa 3780

aataaaaatc ttatcaagta acttttctga ccacaatgga ataaaaacca caaatgaaa 3840

acaagaggaa ccttgaaaaa tacacaaaca catagaata aaacatccta ctggacagct 3900

aatgagtcaa ttaaaaaaat aagaaggcaa ttaaaaaaaa caaatgaaaa caaaaataca 3960

acaaaccaa acttgagata caataacatc tatatcaagc gggatgttta tagtaataaa 4020

tgcttaaatt aaaagagtat aaagacttca aataacctaa caatgaacct caacgaacga 4080

gaaaagcaag aataaaccaa accccaatt aatagaagaa ataataaaaa tcacagcaga 4140

tatgaataaa gttgagacia aaaaatacat aagatcaaca aaaggaaaag ctgcatgttt 4200

gaaaagatat taacaaacc ttagccagat taagaataaa ggaagaagag gcaaataaat 4260

aaaataagaa atgaaaaagg gaaataaatg aaatTTtatac tagaaaaata cataagatca 4320
 atgatagaaa aagtTgggtt tttgaaaaga taatcaaaac tgacaaattc tttgtctaag 4380
 aaaaaaaga cctaattaaa taaaatcaga aatgaaaaag gagacataac aactgagaac 4440
 ccagaagcac aaagaatcgt tagagactat tTtgaacaac tgcatgccaa caaattggaa 4500
 aactTggaag aaatgaataa attcctggac acacacaact taccaagatt gaacaatgaa 4560
 gaaataaaaa aaaaattaac aagccaaaaa tcaatatggc tatcaaagag ataattcaaa 4620
 gtcttccatc aaaaTgaagc tcaggatcat TtgcttTact gctaactcta ccaaaattta 4680
 aagaagaacc aatatcagtt ctacccaaat tctccaaaaa tttggaagga aggtaatact 4740
 ccaaactcat tatatgaggt cagcattacc ctgatactaa aaccagacaa aaacatgaca 4800
 agaaaccaat aagaacacta cagatcaata tTactgatga gcagagatgc aaagataatc 4860
 aacaaaaac tagcaaacca aattcaacaa cacattaaca tgatcaagtg aaattcatcc 4920
 tagggataca aagatggTtc aacacatgca aatcaacaaa cgtgatgcat cacattagca 4980
 gaatcaagaa aataaacata tcattatttc aatagatgcc aaaaaagcac ttgataaaat 5040
 Tgaacaacag ttcatgataa aaactctcat caaaatgggt atagagagaa tacacttcaa 5100
 aattataaag gccatgtatg acaaaccac agtTaaact gtgctgaatg agaaaaataa 5160
 tttttctta tgagaaaggc cttttcttta aagactggaa caagataaga atgcttactt 5220
 tcaccacttt tattcaacat gacatggaaa gtctctcca gaacaagtag ggaaaagaac 5280
 taataaaaa gcataaaaac Tggaaagaag aagtcaaatt agcgtTgttt gtagacaaca 5340
 caattcata ctagaaaaa acctcaataa aaacaatgtt aaaattgata aaaacattca 5400
 gTaaagttgc gggatacaaa ataaactaca gaaatcaaca aataatctga aaaaacaaga 5460

aagcaacctc attfactatg gctacaactg tataaaatac ctcaaagcaa tcaaaccaaa 5520

taatggaaag atctatataa gaaaaactat gaaattctta tgcaaaaagt acaagaaaac 5580

acacaaataa tgtaaagata tttcatatgt atggattgga atatttaatg ttgttaaait 5640

gacaatacta ctaaagcaat ttacggattc agcacaatcc ccatcaaac agtggcattc 5700

ttcacagaaa tagaaaaaa aaattataaa atatgtgcag gacatcaaaa gataccctaaa 5760

taactaaagt aatcatgagc aaaaaagaac aaagctagaa tcatcacacc acctgacttc 5820

aaatttacta caaagctatc acaatcaaaa cagcatagta ctgggataaa agcagacaca 5880

tagaccaatg gaatagaata gagaaccata tatatatcca cacatttaca gccaaactcaa 5940

ctttgacaaa gactacacat acgttgggga aaagaaaatc ttttaataa atattgctgg 6000

aaaaaatgga taattttatg cagaaaaaat ctgaccct atctctcacc atacacaaaa 6060

atcaaataaa aataaattat agacttaaat ctaaaactta aaaatataaa actaaaaacg 6120

aaattattta ggaaacacta caggacattg gtctgggcaa agattttttt gtttgtaaaa 6180

cctcaacatt acagacaact aaagcaaaaa atagacaaat ggaattatat caagctaaaa 6240

agcctctgta cagtaaaca tcaagcaagt gaagagataa cccacagaat gggagaaaat 6300

atttacaac cattcatctg tccaggatt aataaccaga atacataagg agctcaaca 6360

actaaataac caaaacaaaa tactaaaaac caaaactaat aaccaatta aagactgggc 6420

agaatatctg aatacatatt tctcataaga agacataaag atggcaaaca gatatatgaa 6480

ataatgctca acatcactaa ttatcagaga aatgcaagtc aaaactacaa tttaaaatgg 6540

gttgtatcaa aagacaggca ataacaggtg ctggtgagga tgcagagaaa ggggaacct 6600

tctacactgt tggaggaaat atagattagt agagccacta aggagaactc tatggagttt 6660

gctcaaacac ctaaagtaga agtaccatat tattaatctc actactgaat atatatata 6720

aataaaggaa atcagtatat tgaatatata gctgcaactcc cttatttatt gcagcactat 6780

tccaataac caatacga ttcaatttaa ctgccatca gtggatgaat agattaagag 6840

aatgtgttat atatacaca tggaatatta ttcagccata aaatagaatg aaatcttgc 6900

atctgcaaca acatggatgg aacaagaggc ctttatatta agtgaataa gtcaatcaca 6960

gaaagacaaa tattctaacc cacatgttcc cactcatatg tggaagctaa aaaaagtaga 7020

actcatgaag atagagagta gaacggtggt ttccagaggc caggaagggt agtgggggct 7080

gggagaaaat gcagggaaaa taaaagaata aaaatgtatt tattaagtgt gcttaaaaat 7140

cgcaaagata gtaaatttta tatgtatggt ttaaccaagt aaaaataaaa caatcataac 7200

aaaactcaca aaacactata aaattaaaat ttagataact tcattgtaga aataagtgaa 7260

gtgaaaataa gcactaatga gaatcatggt gaaaaag 7297

<210> 9
 <211> 601
 <212> DNA
 <213> Homo sapiens

<400> 9
 atatttcag tgggaccata cgtgtccact gtaatgta ttttaaaatg ttcaactatt 60

ttacaatgtg agtgactctt gctgtgtact tcttctatat tcaatctgtg tgaggctctg 120

aaccaagat ttagtctaga atttacactt cctcttttca agttctcaac taacatgaca 180

cagttctaaa atattaaaga gaatacagtt gccttgtatt tgttatttga agtcagttgg 240

aaatTTTTT cgggtcccta cactcagaaa cacttttagc tttctgtagg gtaaaagtgt 300

rtaggttttg gaggtagaag ggcttgagtt caaatcctgt ctcagctatt tacaggttgg 360

accctaaaca ggttatTTaa cccatctgag cagtgtcttt ctatgtaaaa tgagaataat 420

tgcaaagagt tattgtagac tatagatttt tttttaaaga acaggaacca atttctttaa 480

ttttaatac attcaccttt tttatatgtc ttgtgcctgg tacataaatg cttagcaaat 540

ttttttgaaa gaactgaatt gaattcttgt gaagatgaag taaaataatg tttttttttt 600

t 601

<210> 10
 <211> 673
 <212> DNA
 <213> Homo sapiens

<400> 10
 gtatctttta taacaatttt gttctctgga cagtttgaac tagtaagtac aagaataggg 60

aagtagacca gccttagaag aaattagggg tagtttagaa tttgtgaaaa accagictat 120

attatatatt aactagtcca ccatagagaa aaaacggatt gttatggtat acatttgaaa 180

ggaaattaaa atttcattta ygggttttaa gagtttgaa atatagagtg taccaccttg 240

ctggcaggct aacaacacaa gggaaatttt ttcaaagggt tatttcttta ttaagtggaa 300

gacaacccaaa taggccacga gcggtagctc atgctgttaa tcccagcact ttgtgaggcc 360

aaggcgagtg gatcacttga ggccagaagt ttgagaccag cctcaccaac atggcgtaac 420

cccatctcta ctaaaaatac aaaaatcagc caggcatggg tgcgtgtgcc tgtaatccca 480

gtacctggg aggctgaggc acgaatctct tgaactggg aggtggaggt tgcagtgagc 540

tgagatagcg ccacagcact ccagcctgtg tgacagagca agacctttac tattacaata 600

tgaatctgga ttaactagac tcaaactgag cagacactga ttaaaattag aatatttttc 660

agcaatccaa ata 673