(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2014-0119493
(43) 공개일자 2014년10월10일

(51) 국제특허분류(Int. Cl.)
C12Q 1/68 (2006.01) *C12N 15/11* (2006.01)
(21) 출원번호 10-2013-0035219
(22) 출원일자 2013년04월01일
심사청구일자 2013년04월01일

(71) 출원인
울산대학교 산학협력단
울산광역시 남구 대학로 93
(72) 발명자
송규영
서울특별시 강남구 압구정로 151 신현대아파트 114동 204호
양석균
서울 강남구 언주로30길 26, G동 1301호 (도곡동, 타워팰리스)
(74) 대리인
특허법인태백

전체 청구항 수 : 총 8 항

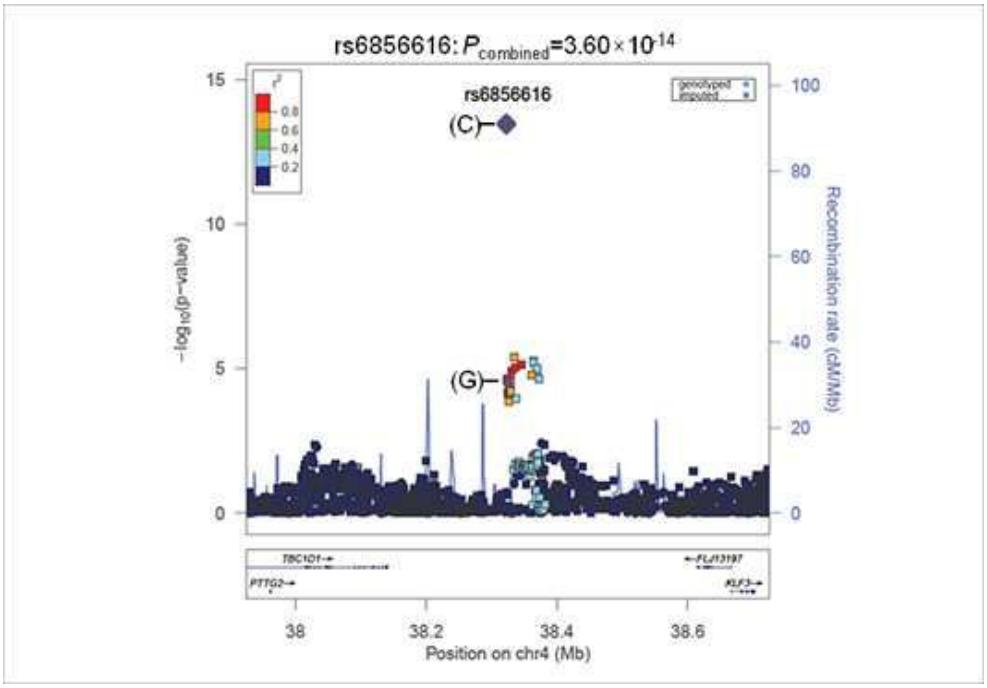
(54) 발명의 명칭 **크론병 감수성 진단용 폴리뉴클레오티드 마커 조성물**

(57) 요 약

본 발명은 크론병 감수성 진단을 위한 단일염기다형(Single-Nucleotide Polymorphism; SNP)을 포함하는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 제공하며, 보다 구체적으로는, 서열번호 1 내지 서열번호 4로 구성되는 군으로부터 1종 이상 선택되는 폴리뉴클레오티드로서, 27번째 염기(다형성부위)를 포함하는 10 내지 50개의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 포함하는 크론병 감수성 진단용 마커 조성물을 제공한다.

본 발명의 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드는 크론병의 유전적인 진단 표지로서 크론병의 진단, 크론병 위험 예측 등에 유용하게 사용될 수 있다.

대 표 도 - 도1



이 발명을 지원한 국가연구개발사업
과제고유번호 2010-0015648
부처명 교육과학기술부
연구관리전문기관 한국연구재단
연구사업명 도약연구지원사업
연구과제명 전유전체연관검색을 활용한 한국인 염증성장질환 유전자 발굴과 발병기전분석
기 여 율 1/1
주관기관 울산대학교 산학협력단
연구기간 2010.05.01 ~ 2013.04.30

특허청구의 범위

청구항 1

서열번호 1 내지 서열번호 4로 구성되는 군으로부터 1종 이상 선택되는 폴리뉴클레오티드로서, 27번째 염기(다형성부위)를 포함하는 10 내지 50개의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 포함하는 크론병 감수성 진단용 마커 조성물.

청구항 2

제1항에 있어서,

상기 서열번호 1의 27번째 염기(다형성부위)의 대립유전자형은 C/T; 서열번호 2의 27번째 염기(다형성부위)의 대립유전자형은 C/T; 서열번호 3의 27번째 염기(다형성부위)의 대립유전자형은 C/T; 및 서열번호 4의 27번째 염기(다형성부위)의 대립유전자형은 A/G인 것인, 크론병 감수성 진단용 마커 조성물.

청구항 3

제1항에 있어서,

상기 마커 조성물은, 한국인의 크론병 감수성을 진단하기 위해 사용되는 것인, 크론병 감수성 진단용 마커 조성물.

청구항 4

서열번호 1 내지 서열번호 4로 구성되는 군으로부터 1종 이상 선택되는 폴리뉴클레오티드로서, 27번째 염기(다형성부위)를 포함하는 10 내지 50개의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드로 이루어진 크론병 감수성 진단용 프로브.

청구항 5

제4항의 크론병 감수성 진단용 프로브가 집적된 크론병 감수성 진단용 DNA 마이크로어레이 칩.

청구항 6

서열번호 1 내지 서열번호 4로 구성되는 군으로부터 1종 이상 선택되는 폴리뉴클레오티드로서, 27번째 염기(다형성부위)를 포함하는 10 내지 50개의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 포함하는 크론병 감수성 진단용 키트.

청구항 7

검체로부터 핵산 시료를 얻는 단계; 및 상기 시료에서 크론병 발병 위험을 나타내는 유전자 서열의 존재를 검출하는 단계를 포함하며, 상기 유전자 서열은 서열번호 1 내지 서열번호 4으로 구성되는 군으로부터 1종 이상 선택되는 서열에서, 27번째 염기(다형성부위)를 포함하는 10 내지 50개의 연속 DNA 서열 또는 그의 상보적 서열인 것인 크론병 감수성 진단을 위한 정보를 제공하는 방법.

청구항 8

제7항에 있어서,

상기 27번째 염기가,

서열번호 1의 경우 C, 서열번호 2의 경우 T, 서열번호 3의 경우 T, 및 서열번호 4의 경우 G인 것인 크론병 감수성 진단을 위한 정보를 제공하는 방법.

명 세 서

기 술 분 야

[0001] 본 발명은 크론병 감수성 진단용 폴리뉴클레오티드 마커 조성물 및 이를 이용한 크론병 감수성의 진단 방법에 관한 것이다.

배 경 기 술

[0002] 크론병은 염증성 장 질환의 일종으로 만성 재발성 질환이다. 이는 인종별로 유태인과 코카시안에서 발생이 많고 동양인에는 상대적으로 발병 빈도는 드물다. 그런데, 최근에는 남유럽과 우리나라를 포함하는 아시아 국가, 그리고 다른 개발도상국에서도 발병률이 증가하고 있다.

[0003] 특히 크론병은 젊은 사람(10대 후반~20대 초반)에게 복통, 설사, 체중 감소, 발열 등이 지속적으로 보여질 때에 의심해야 할 질환이며, 진단은 임상증상, X선 조영, 및 내시경 검사 또는 병리학적 검사 등을 조합해서 행해지고 있다. 그러나, 이들 방법은 경험과 숙련된 기술을 요하며, 특히 발병 초기에 있어서, 크론병과 궤양성 대장염의 감별이 곤란한 증례가 많다. 한편, 진단의 중심적 방법인 소화관 조영검사나 내시경 검사는, 환자의 육체적 혹은 정신적 고통을 주는 결점이 있고, 그러므로, 크론병을 간편하고 정밀도 좋게 판별할 수 있는 진단 방법이 요구되고 있다.

[0004] 즉, 크론병으로 많은 환자들이 고통 받고 있지만 이 질병의 명확한 원인이나 치료 방법은 제시되지 않았으며, 많은 임상 연구에도 불구하고 전문가의 판단이나 의견으로밖에 답변이 어려운 문제들이 많다. 따라서 환자 진료나 치료 방법에 차이가 있을 수 있으며, 이러한 차이를 줄이기 위해서는 크론병 위험을 정확하게 진단하는 것이 요구되고 있다. 이러한 점을 해결하기 위하여 한국등록특허 10-1159626호는 크론병 진단용 항체를 개시하고 있으나, 유전적 정보에 기초하여 한국인 특이적으로 크론병을 예측하는 기술에 대해서는 전혀 개시하고 있지 않다.

발명의 내용

해결하려는 과제

[0005] 따라서 본 발명은 한국인을 대상으로 크론병 감수성과 유의한 관련성을 가지는 단일염기다형(Single-Nucleotide Polymorphism; SNP)을 규명하고, 이를 크론병 감수성 진단의 임상적 표지 인자로 사용하고자 한다.

과제의 해결 수단

[0006] 상기 과제의 해결을 위하여 본 발명은 서열번호 1 내지 서열번호 4로 구성되는 군으로부터 1종 이상 선택되는 폴리뉴클레오티드로서, 27번째 염기(다형성부위)를 포함하는 10 내지 50개의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 포함하는 크론병 감수성 진단용 마커 조성물을 제공한다.

[0007] 또한 본 발명은 서열번호 1 내지 서열번호 4로 구성되는 군으로부터 1종 이상 선택되는 폴리뉴클레오티드로서, 27번째 염기(다형성부위)를 포함하는 10 내지 50개의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드로 이루어진 크론병 감수성 진단용 프로브 및 상기 프로브가 집적된 크론병 감수성 진단용 DNA 마이크로어레이 칩을 제공한다.

[0008] 또한 본 발명은 서열번호 1 내지 서열번호 4로 구성되는 군으로부터 1종 이상 선택되는 폴리뉴클레오티드로서, 27번째 염기(다형성부위)를 포함하는 10 내지 50개의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 포함하는 크론병 감수성 진단용 키트를 제공한다.

[0009] 또한 본 발명은 검체로부터 핵산 시료를 얻는 단계; 및 상기 시료에서 크론병 발병 위험을 나타내는 유전자 서열의 존재를 검출하는 단계를 포함하며, 상기 유전자 서열은 서열번호 1 내지 서열번호 4으로 구성되는 군으로부터 1종 이상 선택되는 서열에서, 27번째 염기(다형성부위)를 포함하는 10 내지 50개의 연속 DNA 서열 또는 그의 상보적 서열인 것인 크론병 감수성 진단을 위한 정보를 제공하는 방법을 제공한다.

발명의 효과

[0010] 본 발명의 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드는 크론병의 유전적인 진단 표지로서 크론병의 진단, 크론병 위험 예측 등에 유용하게 사용될 수 있는 바, 본 발명에 의하면, 염증성 대장염의 임상적 처치에 앞서 크론병 감수성이 높은 환자를 조기에 신속히 분류할 수 있어, 환자 맞춤형 치료를 효율적으로 수행할 수 있다.

도면의 간단한 설명

[0011] 도 1은 4p14의 rs6856616에 대한 위치적 연관 플롯(regional association plot)을 나타낸다. 상기 도에서 (G)는 GWAS의 뻣og_{10}P값을 나타내고, (C)는 종합된 결과를 나타낸다.

도 2는 10q25의 rs11195128에 대한 위치적 연관 플롯(regional association plot)을 나타낸다. 상기 도에서 (G)는 GWAS의 뻣og_{10}P값을 나타내고, (C)는 종합된 결과를 나타낸다.

도 3은 11q13의 rs11235667 및 rs11235604에 대한 위치적 연관 플롯(regional association plot)을 나타낸다. 상기 도에서 (G)는 GWAS의 뻣og_{10}P값을 나타내고, (C)는 종합된 결과를 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0012] 본 발명은 서열번호 1 내지 서열번호 4로 구성되는 군으로부터 1종 이상 선택되는 폴리뉴클레오티드로서, 27번째 염기(다형성부위)를 포함하는 10 내지 50개의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 포함하는 크론병 감수성 진단용 마커 조성물을 제공한다.

[0013] 본 발명의 한 구체예에서, 상기 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드의 길이는 8 내지 52 뉴클레오티드, 10 내지 50 뉴클레오티드 또는 20 내지 40 뉴클레오티드이나 이에 제한되는 것은 아니다.

[0014] 본 발명의 다형성(polymorphism)은 단일 유전자 좌위에 두 가지 이상의 대립 유전자가 존재하는 경우를 의미하며, '다형성부위(polymorphic site)'란 상기 대립 유전자가 존재하는 유전자 좌위를 의미한다. 다형성 부위 중에서, 사람에 따라 단일 염기만이 상이한 것을 '단일 뉴클레오타이드 다형성', 즉 SNP(single nucleotide polymorphism)라고 한다.

[0015] 본 발명의 다른 구체예에서, 상기 서열번호 1의 27번째 염기(다형성부위)의 대립유전자형은 C/T; 서열번호 2의 27번째 염기(다형성부위)의 대립유전자형은 C/T; 서열번호 3의 27번째 염기(다형성부위)의 대립유전자형은 C/T; 및 서열번호 4의 27번째 염기(다형성부위)의 대립유전자형은 A/G 일 수 있다.

[0016] 본 발명의 또 다른 구체예에서, 상기 마커 조성물은, 한국인의 크론병 감수성을 진단하기 위해 사용될 수 있다.

[0017] 본 발명의 발명자들은, 한국인 특이적으로 나타나는 크론병 감수성에 관련하는 유전자 마커를 발굴하고자 하였고, 그 결과, 서열번호 1 내지 서열번호 4로 나타나는 SNP가 크론병 감수성과 유의하다는 것을 밝혀내고 본 발명을 완성하게 되었다(이하 실시예 참조).

[0018] 본 발명의 또 다른 구체예에서, 상기 서열번호 1 내지 4의 SNP의 구체적인 염기서열은 아래 표 1과 같다.

표 1

[0019]

서열번호	rs 번호 (ncbi gene bank)	5'-부터의 염기서열
1	rs6856616	TTTTATAAGCATTCCCAAGTGAATCT[C/T]AATTTACCTACTGAAAAGCAATAAT
2	rs11195128	AGTCAGAAAAGGTCAACACAGAATAG[C/T]GATATAATCAAAATTACCCTCTACG
3	rs11235604	TTTACCCAACAAGGGCCAAGCAGGCG[C/T]GGGTGTCCCAGGAGCTGAAGAAGGC
4	rs11235667	TGACTAGTTTGGTGAACCAACACTAC[A/G]GAATAAAACAGGCCTCAAGGCAAAA

[0020] 상기 각각의 폴리뉴클레오티드에서 SNP의 위치는 27번째 염기로서, 주로 나타날 수 있는 대립유전자형을 괄호 안에 나타내었다.

[0021] 상기 서열번호 1 내지 서열번호 4로 구성되는 군으로부터 1종 이상 선택되는 폴리뉴클레오티드로서, 27번째 염기(다형성부위)를 포함하는 10 내지 50개의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드는 크론병 감수성 진단에 사용될 수 있다.

[0022] 본 발명은 또한, 크론병 감수성 진단에 사용될 수 있는 상기 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드로 이루어진 크론병 감수성 진단용 프로브 및 이것이 집적된 크론병 감수성 진단용 DNA 마이크로어레이 칩을 제공한다.

[0023] 본 발명의 크론병 감수성 진단용 DNA 마이크로어레이 칩은 당업자에게 알려진 방법으로 제작할 수 있다. 보다

구체적으로는, 상기 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드로 이루어진 크론병 감수성 진단용 프로브를 탐침 DNA 분자로 이용하여 DNA 칩의 기판상에 고정화시키기 위해 파이조일렉트릭(piezoelectric) 방식을 이용한 마이크로피펫팅(micropipetting)법 또는 핀(pin) 형태의 스폿터(spotter)를 이용한 방법 등을 사용할 수 있으나, 이에 제한되는 것은 아니다.

[0024] 또한 상기 DNA 마이크로어레이 칩의 기판은 아미노-실란(amino-silane), 폴리-L-라이신(poly-L-lysine) 및 알데히드(aldehyde)로 이루어진 군에서 선택되는 하나의 활성기가 코팅된 것이 바람직하나 이에 제한되는 것은 아니다. 또한, 상기 기판은 슬라이드 글라스, 플라스틱, 금속, 실리콘, 나일론 막, 및 니트로셀룰로스 막으로 이루어진 군에서 선택될 수 있으나 이에 제한되는 것은 아니다.

[0025] 본 발명은 또한 상기 서열번호 1 내지 서열번호 4로 구성되는 군으로부터 1종 이상 선택되는 폴리뉴클레오티드로서, 27번째 염기(다형성부위)를 포함하는 10 내지 50개의 연속 DNA 서열로 구성되는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 포함하는 크론병 감수성 진단용 키트를 제공한다.

[0026] 본 발명은 또한 검체로부터 핵산 시료를 얻는 단계; 및 상기 시료에서 크론병 발병 위험을 나타내는 유전자 서열의 존재를 검출하는 단계를 포함하는 크론병 감수성 평가 방법을 제공한다. 이때 상기 유전자 서열은 서열번호 1 내지 서열번호 4으로 구성되는 군으로부터 1종 이상 선택되는 서열에서, 27번째 염기(다형성부위)를 포함하는 10 내지 50개의 연속 DNA 서열 또는 그의 상보적 서열을 사용할 수 있다.

[0027] 본 발명의 다른 구체예에서, 상기 27번째 염기로는,

[0028] 서열번호 1의 경우 C, 서열번호 2의 경우 T, 서열번호 3의 경우 T, 및 서열번호 4의 경우 G를 사용할 수 있다.

[0029] 검체로부터 핵산 시료를 얻는 단계에서, 핵산은 진단 대상의 혈액, 피부세포, 점막 세포 및 모발 등 모든 세포로부터 분리될 수 있다. 상기 핵산은 DNA 또는 RNA일 수 있으며, 바람직하게는 DNA일 수 있다. 해당 세포로부터 핵산을 추출하는 방법은 특별히 한정되지 않으며, 당업계에 공지된 기술 또는 시판되고 있는 핵산 추출용 키트를 사용할 수 있다. 예를 들면, 페놀/클로로포름 추출법 및 프로테아제 K 처리방법을 사용할 수 있다.

[0030] DNA 또는 RNA 추출용 키트는 Qiagen, Inc, 및 Stratagene에서 구입할 수 있다. 상기에서 RNA를 추출하여 사용하는 경우에는 역전사를 통해 cDNA를 제조하여 사용할 수 있다.

[0031] 시료에서 크론병 발병 위험을 나타내는 유전자 서열의 존재를 검출하는 단계에서는, 유전자 서열 분석이 수행될 수 있다. 서열분석은 당업계에 공지된 방법을 모두 이용할 수 있으며, 구체적으로는 이에 제한되는 것은 아니나, 자동염기서열분석기를 사용하거나, 파이로시퀀싱(pyrosequencing), PCR-RELP법(restriction fragment length polymorphism), PCRSSCP법(single strand conformation polymorphism), PCR-SSO법(specific sequence oligonucleotide), PCR-SSO법과 도트 하이브리드화법을 조합한 ASO(allele specific oligonucleotide) 하이브리드화법, TaqMan-PCR법, MALDI-TOF/MS법, RCA법(rolling circle amplification), HRM(high resolution melting)법, 프라이머 신장법, 서던 블롯 하이브리드화법 및 도트 하이브리드화법 등의 공지의 방법 중 선택된 어느 하나 이상을 사용할 수 있다.

[0032] 검체의 시료로부터 상기 크론병 발병 위험을 나타내는 유전자 서열의 존재가 확인된다면, 상기 검체는 크론병이 발병할 위험이 높다고 진단할 수 있다.

[0033] 이하, 본 발명의 이해를 돕기 위하여 실시예를 들어 상세하게 설명하기로 한다. 다만 하기의 실시예는 본 발명의 내용을 예시하는 것일 뿐 본 발명의 범위가 하기 실시예에 한정되는 것은 아니다. 본 발명의 실시예는 당업계에서 평균적인 지식을 가진 자에게 본 발명을 보다 완전하게 설명하기 위해 제공되는 것이다.

[0034] **<실험예>**

[0035] 하기의 실험예들은 본 발명에 따른 각각의 실시예에 공통적으로 적용되는 실험예를 제공하기 위한 것이다.

[0036] **1. 실험 대상**

[0037] 2,311 크론병(Crohn's disease, CD) 환자군 및 2,442 대조군 환자군을 대상으로 환자-대조군 연구를 수행하였다. 모든 대상 환자는 한국인이었다.

[0038] 2,311 크론병 환자 중, 1,519명의 환자는 아산병원의 IBD 클리닉에서 모집되었고, 792명은 크론병에 대한 한국

리서치 네트워크에서 모집되었다. 크론병의 진단은 표준 임상, 방사선, 내시경, 병리조직학적 기준에 따라 수행되었다. 중기 대장염의 환자는 대상에서 제외되었다.

[0039] 대조군 환자 중, 복제 코호트 1을 위해, 울산대학병원 및 아산병원에서 732명의 건강한 지원자를 모집하였고, 복제 코호트 2를 위해, 서울대학병원의 건강검진센터에 내원한 720명의 건강한 지원자를 모집하였으며, 경상 국립대학병원에서 257명의 건강한 지원자를 모집하였다. 모든 샘플은 임상시험심사위원회에 의해 승인받은 프로토콜에 따라 수집하였다.

[0040] **2. Genotyping 및 분석**

[0041] 533 크론병 환자를 Illumina OmniExpress array를 사용하여 유전자형을 분석하였다. 대조를 위해, Illumina Omni1-Quad array에서 분석된 800명의 공지된 유전자형 데이터를 사용하였다. Y 및 미토콘드리아 염색체에의 SNP는 분석 대상에서 제외되었다. 품질 관리를 위해, call rate가 95% 미만, minor allele frequency(MAF)가 < 0.01, 대조군에서 Hardy-Weinberg equilibrium에 의한 significant deviation이 $P < 1\times10^{-5}$ 인 SNP는 대상에서 제외되었다. 유사하게, 이후 유전자형 비율이 96% 이하인 샘플은 제거되었다. 533 케이스 및 800 대조군 샘플 중 OmniExpress 및 Omni1-Quad array 모두에서 공통된 상염색체의 558,194 SNP 및 X 염색체의 13,044 SNP만이 이후 추가 품질관리의 대상이 되었다.

[0042] 1,333 GWAS 샘플의 잠재적 유전자 근친도(potential genetic relatedness)를 pair-wise identity에 기초하여 PLINK 1.07 software를 사용하여 분석하였다. 규명된 첫번째 상대적 쌍 각각에 대하여, 낮은 유전자형 call rate를 갖는 샘플은 제거되었다. 총 66 샘플(1 케이스 및 65 대조군)이 샘플 복제 및/또는 유전자 근친도에 의해 제거되었다. 계속해서, PCA(주성분분석, principal-component analysis)에 기초한 방법을 사용하여 software package EIGENSTRAT 3.0 에서의 population outliers 및 stratification을 검출하였다. 염색체 6의 HLA 영역, 염색체 8 및 5의 자리바꿈, 염색체 11의 두 영역을 포함하는 long-range LD의 다섯개의 구별되는 영역을 갖는 모든 SNP는 PCA에서 제외되었다. 초기 PCA를 위해, 모든 1,267 샘플(532 케이스 및 735 대조군)이 International HapMap Project의 194 참고 샘플과 함께 분석되었고, 그 결과, 두 가외치(outlier)가 검출되었다. SNP 및 샘플 품질관리 분석 결과, 532 케이스 및 733 대조군의 571,238 SNP의 유전자형 데이터를 GWAS 분석에 사용하였다.

[0043] **3. Imputation**

[0044] imputation analysis를 위해, 입력되지 않은 유전자형이 IMPUTE (v2.0)를 통해 1,000 게놈 프로젝트 데이터베이스의 Asian reference panel (JPT + CHB)을 사용하여 GWAS 샘플에 귀속되었다. 귀속된 유전자형 중, genotype probability가 < 90%, 귀속 확실성(imputation certainty)이 < 80%인 SNP, MAF < 5% 및 손실율이 > 10%인 유전자형은 추가 분석에서 제외되었다. 총 5,093,133의 귀속된 SNP가 품질 관리를 통과하였고, 571,238의 유전자형 분석된 SNP와 함께 연관도 분석에 사용되었다.

[0045] **4. 복제**

[0046] 평가를 위한 유전자형 분석은 matrix-assisted laser desorption/ionization time-of-flight MassARRAY platform (Sequenom, San Diego, CA, USA) 상에서 iPLEX로 수행되었다. rs773024, rs9654389, rs1393820, rs753988 및 rs6475004의 평가는 TaqMan SNP genotyping assay를 사용하여 수행하였다.

[0047] **5. 분산 및 위험도 분석**

[0048] 크론병 위험 형질 각각에 의해 설명되는 유전자 분산도가 So HC, Gui AH, Cherny SS, et al. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genet Epidemiol 2011;35:310-7 에 나타난 알고리즘에 의해 liability threshold model 하에서 측정되었다.

[0049] 상기 모델은 latent continuous liability을 제시하는데, 일반적으로 평균 0 및 분산 1 로 분포하는 것으로 추정된다. 본 발명의 발명자들은 한국인의 크론병 발병률을 0.0112%로 추정하였다. 단일 변종에 의해 설명되는 유전 정도를 계산하기 위해, 형질 빈도(allele frequency), 교차비(odds ratio, OR), 및 질병 출현율(disease prevalence)을 사용하였다. 또한 본 발명의 발명자들은 유전적 위험도(genetic risk score)를 샘플 위험 형질수(simple risk allele count)를 사용하여 계산하였다. 대상 개체는 9 위험 변종에 기초하여 그들의 유전적 위험도에 따라 분류되었고, 크론병과의 연관도는 로지스틱 회귀분석법을 사용하여 측정되었다.

[0050] **6. 통계학적 분석**

[0051] GWAS 스테이지에서, 단일 마커 연관 분석은 Cochran-Armitage trend test로 수행되었다. X 염색체에 있는 SNP는 로지스틱 회귀분석모델을 사용하여 성별을 공변량(covariate)으로 고려하여 분석하였다. genome-wide 연관도 및 인구집단 층화의 잠재적 영향에 대한 전체 유의도를 평가하기 위하여, quantile-quantile 플롯을 R(2.13.1)을 사용하여 생성하였다. 인구집단 층화(population stratification)의 잠재적 영향은 또한 genomic control inflation factor를 계산하여 평가하였다. $?\log_{10}P$ 의 Manhattan plot은 Haploview (v4.2)를 사용하여 생성되었다. 복제 분석은 두개의 후속 샘플을 각각 따로 분석한 후, 모든 환자 및 대조군의 결합 샘플을 함께 분석하여 수행되었다. 결합 샘플의 연관 분석은 Cochran-Mantel-Haenszel stratification analysis을 사용하여 수행하였다. Breslow-Day 테스트가 다른 샘플 코호트에 대한 OR간의 이질성을 평가하기 위해 사용되었다. 본 발명의 발명자들은 MHC 영역에서 관찰된 복수의 연관성에 대한 독립도를 계층적 조건부 로지스틱 회귀분석으로 측정하였다.

[0052] 이전에 보고된 코카시안 인종의 140 크론병 감수성 위치를 검출하기 위한 GWAS 샘플의 검증력 분석은 Quanto software package (Version 1.2.4, http://hydra.usc.edu/gxe/)를 사용하여 수행되었다. 각각의 보고된 SNP 중, 설정 P 값 0.05에서의 검출력은 보고된 OR 및 한국 인구에서의 대립유전자 빈도에 기초하여 계산되었다.

[0053] **7. Cis-eQTL 및 생물정보학적 분석**

[0054] cis-eQTL (expression quantitative trait locus) 분석을 위해, Genevar (GENe Expression VARiation) eQTL 브라우저의 공개 데이터를 적용하였다. 세 신규 위치에서의 SNP간의 cis 연관성 및 근접 유전자의 발현은 1차 섬유아세포, T-세포, 피부, 지방 및 림프아구 세포주에서 측정되었다.

[0055] 기능적 변이체(functional variant)의 진화적 보전(evolutionarily conservation) 예측을 위해, PhastCons46wayPlacental 및 PhyloP 스코어(± 5000 bp window)가 UCSC Genome Browser로부터 획득되었다. 아미노산 치환이 단백질 기능에 미치는 가능한 영향 예측이 SIFT (sorting intolerant from tolerant) (http://sift.jcvi.org) 및 PolyPhen-2 (polymorphism phenotyping v2) (http://genetics.bwh.harvard.edu/pph2) 알고리즘을 사용하여 수행되었다. "유해한(Deleterious)" 및/또는 "손상 가능성(Probably/Possible damaging)" 의 SIFT 및 PolyPhen-2 예측은 단백질 기능에의 효과를 예측하기 위해 판단되었다. 유전자간 영역의 분석을 위해, UCSC Genome Browser 의 ENCODE 프로젝트 데이터를 ChiP sequencing (ChiP-seq), 히스톤 변경 데이터, 및 전사 인자 결합부위의 SNP 효과를 조사하기 위해 사용하였다.

[0056] **<실시예 1> 한국인을 대상으로 한 뉴클레오티드 마커 확립**

[0057] 품질 검사 및 Imputation 결과, 521명의 크론병 환자군 및 732 명의 대조군(복제 코호트 1)의 85 부위의 102 SNP가 추가 분석에 사용되었다. 이중 98 SNP가 성공적으로 유전자형이 분석되었고, 이전에 보고된 두 유전자 및 새로운 위치의 유전자가 Bonferroni correction 후 평가 분석에서 통계적 유의값을 보였다(*TNFSF15* locus의 rs6478109 (combined $P = 2.14 \times 10^{-44}$), *BTNL2*((butyrophilinlike protein 2) 인트론 1의 rs10947261 (combined $P = 2.67 \times 10^{-10}$), 및 4p14의 rs6856616 (combined $P = 1.78 \times 10^{-8}$)

[0058] 또한 크론병의 새로운 감수성 위치(susceptibility loci)를 확인하기 위해, 1,258 크론병 환자군 및 977 대조군(복제 코호트 2)의 25 위치의 27 SNP를 복제하여 확인하였다.

[0059] 종합적인 확인 결과, 6 위치(3 신규 + 3 종래 알려진 위치)의 9 SNP(**rs76418789, rs6856616, rs10947261, rs9271366, rs751728, rs2149085, rs394522, rs11195128, rs11235667**)가 유전적 중요한 연관도($P < 5 \times 10^{-8}$)를 갖는 것으로 확인되었다. 이중 신규 3 위치의 4 SNP를 이하 표 2에 나타내었다.

표 2

SNP	Risk allele	[a]Chr. location	Gene(s) or locus	Study	No. of samples Case	No. of samples Control	[b]RAF Case	[b]RAF Control	[c]P-value	[d]OR	95% [d]CI	[f]P-BD
rs6856616	C	Chr4: 38,325,036	*TBC1D1-KLF3*	GWAS	530	733	0.31	0.24	3.19E-05	1.45	(1.21-1.72)	
				Sample 1	519	730	0.29	0.23	1.95E-04	1.42	(1.18-1.71)	
				Sample 2	1,227	960	0.30	0.23	5.81E-07	1.41	(1.23-1.62)	
				[e]Combined	2,276	2,423	0.30	0.23	3.60E-14	1.43	(1.31-1.57)	9.39E-01
rs11195128	T	Chr10: 112,186,148	*SMNDC1-DUSP5*	GWAS	532	729	0.22	0.14	6.11E-07	1.69	(1.37-2.08)	
				Sample 1	514	724	0.21	0.17	3.40E-02	1.25	(1.02-1.54)	
				Sample 2	1,251	971	0.19	0.14	7.63E-05	1.38	(1.18-1.62)	
				[e]Combined	2,297	2,424	0.20	0.15	1.55E-10	1.42	(1.28-1.58)	1.13E-01
rs11235604	T	Chr11: 72,533,536	*ATG16L2*	GWAS	523	729	0.13	0.09	1.17E-04	1.63	(1.27-2.10)	
				Sample 1	486	714	0.14	0.10	3.02E-03	1.46	(1.13-1.87)	
				Sample 2	1180	957	0.14	0.09	7.04E-07	1.64	(1.35-2.00)	
				[e]Combined	2189	2400	0.14	0.09	2.44E-12	1.61	(1.41-1.83)	7.71E-01
rs11235667	G	Chr11: 72,863,697	*ATG16L2-FCHSD2*	GWAS	532	733	0.14	0.09	4.55E-05	1.66	(1.30-2.13)	
				Sample 1	512	731	0.14	0.10	4.14E-03	1.41	(1.11-1.79)	
				Sample 2	1,243	975	0.12	0.09	1.90E-03	1.36	(1.12-1.64)	
				[e]Combined	2,287	2,439	0.13	0.10	7.15E-09	1.46	(1.28-1.66)	4.04E-01

[0060]

[0061] a: 크로모좀; b: Risk Allele Frequency; c: Cochrane-Armitage trend test를 통해 얻어진 P 값; d: OR(odd ratio), CI(confidence interval); e: Cochran-Mantel-Haenszel (CMH) test statistic에 의해 계산된 종합된 P 값(Pcmh) 및 종합된 OR 값; f: OR 의 이종성을 위한 Breslow-Day (BD) test의 asymptotic P 값

[0062] 이중 가장 강한 연관 신호를 보이는 신규 SNP는 4p14의치의 rs6856616 (OR= 1.43, 95% CI = 1.31-1.57, combined $P = 3.60 \times 10^{-14}$)로 확인되었다(표 2 및 도 1 참조). 이는 *TBC1D1* (TBC1 domain family, member 1) (~184 kb telomeric) 및 *KLF3* (Kruppel-like factor 3) (~341 kb centromeric) 사이에 위치하였다. *TBC1D1* 는 비만 조절과 골격근의 글루코오스 대사와 연관된 유전자이고, *KLF3* 은 전사 억제, 조혈, B 세포 분화, 및 IL-17 매개 지방형성 억제에 중요한 역할을 수행하는 유전자이다.

[0063] 또한 10q25 위치의 rs11195128(OR = 1.42, 95% CI = 1.28-1.58, combined $P = 1.55 \times 10^{-10}$) 또한 연관성이 높은 새로운 위치로 확인되었다(표 2 및 도 2 참조). 이는 구축된 코딩 유전자가 없는 분위이나, 신규 안티센스 RNA RP11-525A16.1 의 다운스트림에 위치한다. *SMNDC1* (survival motor neuron domain containing 1) 및 *DUSP5* (dual specificity phosphatase 5) 유전자가 각각 rs11195128과 121.4 kb centromeric 및 71.5 kb telomeric 에 위치한다. *SMNDC1* 은 스플레오솜 복합체를 구성하며, 이의 과발현은 세포사멸을 야기시킨다. *DUSP5* 는 세포외 신호 조절 키나아제 1/2 (ERK1/2)- 특이적 포스파타아제로, 흉선 세포 및 T 세포 에서 모두 발현되며, IL-2, IL-7 및 IL-15와 함께 TCR 신호에 의해 활성화된다. 또한 이는 M-CSF 신호전달, 골수성 세포 분화, T-세포 분화와 관련된다.

[0064] 11q13, *FCHSD2* 유전자의 ~10.5 kb 부위의 rs11235667(OR = 1.46, 95% CI = 1.28-1.66, combined $P = 7.15 \times 10^{-9}$) 또한 연관성이 높은 새로운 위치로 확인되었다(표 2 및 도 3 참조). rs11235667은 *FCHSD2, ATG16L2* 및 *STARD10* 를 포함하는 ~430 kb의 LD (연관불균형, linkage disequilibrium) 영역에 존재한다. *FCHSD2* (FCH 및 double SH3 domains)은 두 SH3 도메인 및 C-말단의 PDZ-결합 모티프와 연결된 이중나선 도메인을 가지며, epithelial junction의 구아닐레이트 키나아제와 상호작용한다. *ATG16L2* (autophagy related 16-like 2) 는 크

론병 위험 유전자로 알려져 있는 *ATG16L1* 와 동종이다. *ATG16L1*는 자가포식 및 항균 펩타이드를 포함하는 파네스 세포 분비물와 관련이 있다. *ATG16L1* 및 *ATG16L2* 모두 7 WD 반복체 및 이중나선 구조를 갖는다.

[0065] 본 발명의 발명자들은 이후 한국인의 *FCHSD2*, *ATG16L2* 및 *STARD10* 에서 밀접한 SNP를 확인하였다. 이는 *ATG16L2*에 존재하는 rs11235604 이고, rs11235667과 함께(r^2 = 0.96) LD 영역에 존재하였으며, 강한 유의성과 크론병과의 강한 연관성을 보였다(OR = 1.61, combined P = 2.44 x 10^{-12}, 표 2, 도 3 참조). 변이체는 이중나선 영역의 N-말단 220 번째 위치의 아미노산 치환을 유발한다 (염기성 아르기닌을 비극성 트립토판으로 치환). UCSC Genome Browser database 에서 종간 rs11235604의 보존을 측정하기 위해 검색하였고, PhyloP 및 PhastCons 스코어는 각각 0.12 및 0.13으로 나타나, 상기 변이는 진화적으로 보존됨을 알 수 있었다. SIFT 및 PolyPhen-2을 사용한 rs11235604 의 *In silico* 평가는 각각 deleterious, benign의 결과를 보였다. 상기 결과는, 11q13과 연관된 rs11235604는 질병과 연관이 있는 유전 변이(causal variant)임을 강력히 시사한다.

[0066] **<실시예 2> 코카시안 인종의 데이터와 비교**

[0067] 상기 새로운 SNP가 코카시안 인종에 대해서도 적용되는지 확인하기 위해, 상기 GWAS dataset에서의 새로운 세 위치를 IIBDGC (CD dataset: 21102463, UC dataset: 21297633)로부터 조사하였다. 상기 dataset은 5,956 크론병(CD) 환자군, 6,945 UC, 및 21,770 대조군으로 구성되었다. 이하 표 3을 참조하면, 상기 세 위치 중 두 SNP, 4p14의 rs6856616(OR = 1.15, P = 3.47 x 10^{-4}) 및 10q25의 rs11195128(OR = 1.09, P = 7.16 x 10^{-4})가 UC 및 IBD와 함께 코카시안 인종의 크론병에도 일관된 연관성을 보였다.

표 3

[0068]

				IIBDGC											
				[b]RAF (IBD)				[b]RAF (CD)				[b]RAF (UC)			
SNP	Risk allele	[a]Chr. location	Gene(s) or locus	Case (12,882)	Control (21,770)	[c]OR	[d]P value	Case (5,956)	Control (14,927)	[c]OR	[d]P value	Case (6,945)	Control (13,727)	[c]OR	[d]P value
rs6856616	C	Chr4: 38,325,036	*TBC1D1-KLF3*	0.07	0.06	1.12	2.36E-04	0.08	0.06	1.15	3.47E-04	0.07	0.06	1.11	7.57E-03
rs11195128	T	Chr10: 112,186,148	*SMNDC1-DUSP5*	0.34	0.33	1.08	5.32E-05	0.35	0.33	1.09	7.16E-04	0.34	0.33	1.07	7.14E-03

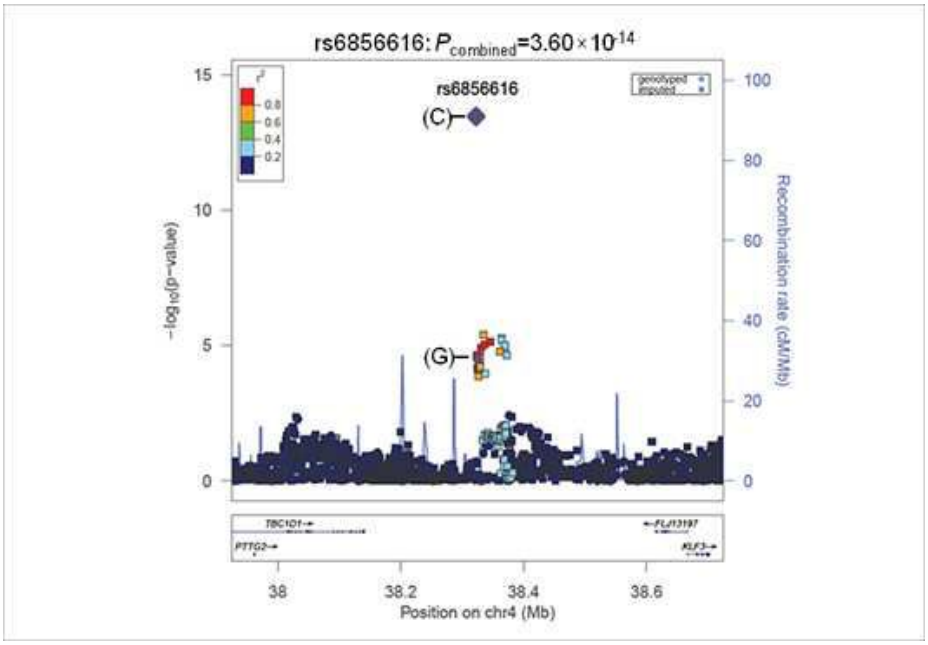
[0069] a: 크로모좀; b: Risk Allele Frequency; c: OR(odd ratio), d: Cochran-Mantel-Haenszel (CMH) test statistic에 의해 계산된 종합된 P 값(Pcmh)

[0070] 반면, 11q13의 rs11235667은 코카시안 인종에서 monomorphic 한 형태로 나타났으며, 이와 관련된 10 SNP 모두 연관도가 적은 것으로 나타나, 11q13와의 관련성은 코카시안 인종에서 적은 것으로 나타났다.

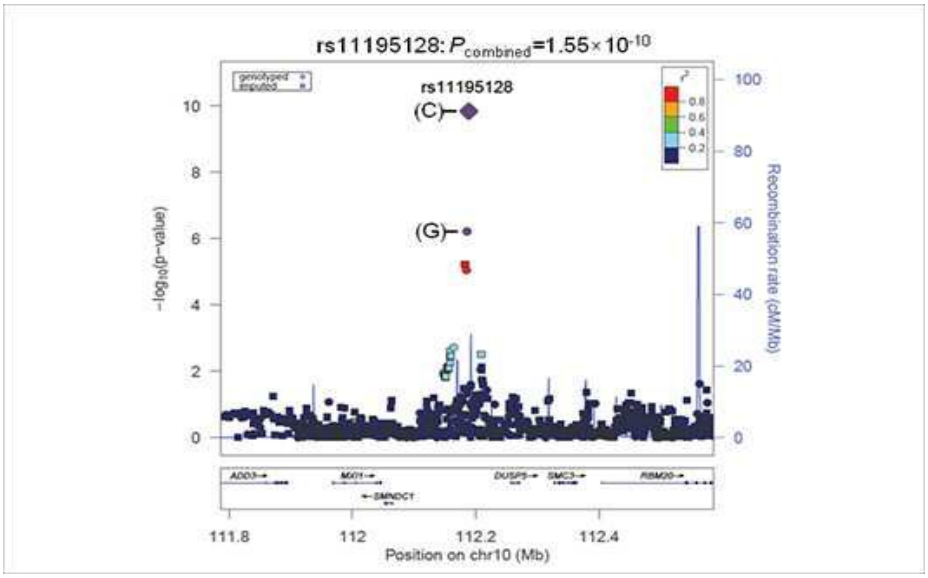
[0071] 이상으로 본 발명의 특정한 부분을 상세히 기술하였는 바, 당업계의 통상의 지식을 가진 자에게 있어서, 이러한 구체적 기술은 단지 바람직한 실시예일 뿐이며, 이에 의해 본 발명의 범위가 제한되는 것이 아닌 점은 명백할 것이다. 따라서, 본 발명의 실질적인 범위는 첨부된 청구항들과 그것들의 등가물에 의하여 정의된다고 할 것이다.

도면

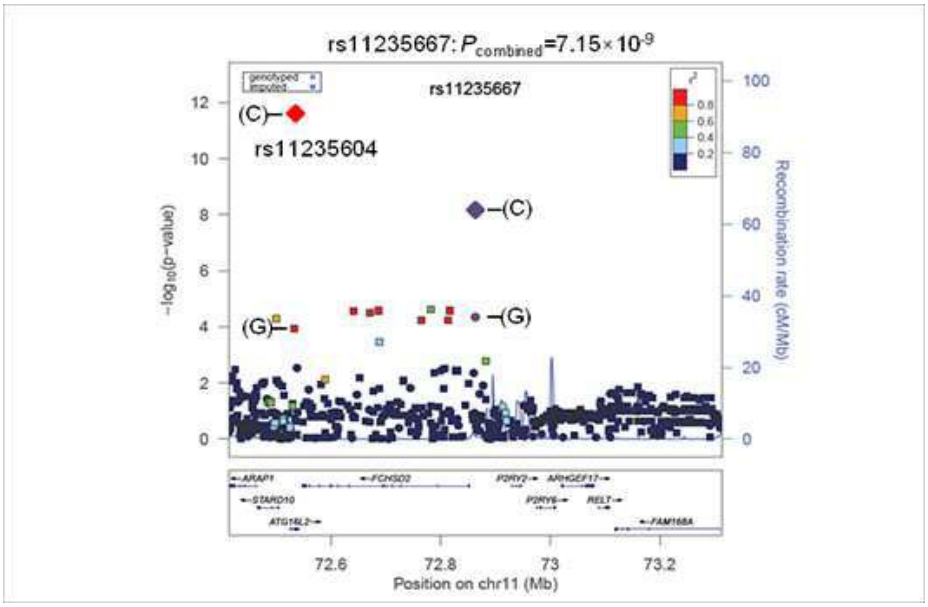
도면1


rs6856616: $P_{combined}$=3.60×10^{-14}
rs6856616
(C)
(G)
genotyped
imputed
−log10(p−value)
Recombination rate (cM/Mb)
Position on chr4 (Mb)
38
38.2
38.4
38.6

도면2


rs11195128: $P_{combined}$=1.55×10^{-10}
rs11195128
(C)
(G)
genotyped
imputed
−log10(p−value)
Recombination rate (cM/Mb)
Position on chr10 (Mb)
111.8
112
112.2
112.4

도면3



서 열 목 록

<110> University of Ulsan Foundation For Industry Cooperation

<120> Polynucleotide Marker Composition for Diagnosis of Susceptibility to Crohn's Disease

<130> ADP-2013-0128

<160> 4

<170> KopatentIn 2.0

<210> 1

<211> 52

<212> DNA

<213> Homo sapiens

<400> 1

ttttataagc attcccaagt gaatctyaat ttacctactg aaaagcaata at 52

<210> 2

<211> 52

<212> DNA

<213> Homo sapiens

<400> 2

agtcagaaaa ggtcaacaca gaatagygat ataatcaaaa ttaccctcta cg 52

<210> 3

<211> 52

<212> DNA

<213> Homo sapiens

<400> 3

tttacccaac aagggccaag caggcgyggg tgtcccagga gctgaagaag gc 52

<210> 4

<211> 52

<212> DNA

<213> Homo sapiens

<400> 4

tgactagttt ggtgaaccaa cactacrgaa taaaacaggc ctcaaggcaa aa 52