



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2018-0037755
(43) 공개일자 2018년04월13일

(51) 국제특허분류(Int. Cl.)

C12Q 1/68 (2018.01)

(52) CPC특허분류

C12Q 1/6883 (2013.01)

C12Q 2600/156 (2013.01)

(21) 출원번호 10-2016-0128332

(22) 출원일자 2016년10월05일

심사청구일자 2016년10월05일

(71) 출원인

울산대학교 산학협력단

울산광역시 남구 대학로 93(무거동)

재단법인 아산사회복지재단

서울특별시 송파구 올림픽로43길 88 (풍납동)

(72) 발명자

송규영

서울특별시 강남구 압구정로 151, 114동 204호 (압구정동, 현대아파트)

양석균

서울특별시 강남구 연주로30길 26, G동 810호 (도곡동, 타워팰리스)

(74) 대리인

특허법인태백

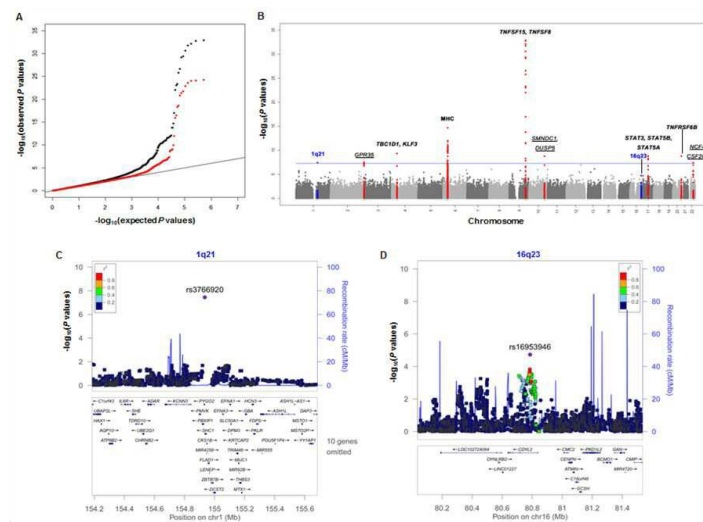
전체 청구항 수 : 총 11 항

(54) 발명의 명칭 한국인의 염증성장질환 위험 예측용 단일염기다형성 마커 조성물 및 이를 이용한 염증성장질환 위험 예측 방법

(57) 요약

본 발명은 한국인의 염증성장질환 위험 예측용 단일염기다형성 마커 조성물 및 이를 이용한 염증성장질환 위험 예측 방법에 관한 것이다. 본 발명자들은 1,505명의 IBD 환자 및 4,041명의 대조군으로 이루어진 한국인 개체군에서의 IBD에 대한 GWAS를 수행하였고, 전장 유전체 상 IBD에 대한 감수성에 있어 유의성을 보이는 3개의 새로운 유전좌위를 밝혀냈다. 또한, 본 발명자들은 28개의 알려진 IBD 유전좌위($P < 2.16 \times 10^{-4}$)가 한국인에서도 연관성이 있는 것을 확인하였다. 상기 결과들은 다른 개체군의 유전형을 분석하는데 있어 비교 분석에 활용될 수 있다.

대표도 - 도1



이 발명을 지원한 국가연구개발사업

과제고유번호 2014R1A2A1A09005824

부처명 미래창조과학부

연구관리전문기관 한국연구재단

연구사업명 (도전)(융합)도약연구지원사업

연구과제명 전유전체연관검색 데이터 분석을 이용한 한국인 염증성장질환 유전자발굴

기 여 율 70/100

주관기관 울산대학교 산학협력단

연구기간 2014.05.01 ~ 2017.04.30

이 발명을 지원한 국가연구개발사업

과제고유번호 A120176

부처명 보건복지부

연구관리전문기관 보건산업진흥원

연구사업명 질환극복기술개발

연구과제명 한국인 크론병의 발병기전, 임상적 및 유전적 특성 규명과 새로운 바이오마커 및 치료제

개발

기 여 율 30/100

주관기관 서울아산병원

연구기간 2012.05.01 ~ 2018.03.31

명세서

청구범위

청구항 1

서열번호 1의 101번째 뉴클레오타이드, 서열번호 2의 101번째 뉴클레오타이드 및 서열번호 3의 101번째 뉴클레오타이드로 이루어진 군에서 선택된 어느 하나 이상의 단일염기다형성(SNP) 마커를 포함하는 10-100개의 연속적인 DNA 서열로 구성되는 폴리뉴클레오타이드 또는 그의 상보적 폴리뉴클레오타이드를 포함하는 아시아인의 염증성 장질환(inflammatory bowel disease; IBD) 발병 위험 예측용 조성물.

청구항 2

제1항에 있어서, 상기 서열번호 1의 101번째 뉴클레오타이드의 대립유전자형은 A/G이고, 상기 서열번호 2의 101번째 뉴클레오타이드의 대립유전자형은 C/T이고, 상기 서열번호 3의 101번째 뉴클레오타이드의 대립유전자형은 A/G인 것을 특징으로 하는 아시아인의 염증성 장질환 발병 위험 예측용 조성물.

청구항 3

제1항에 있어서, 상기 아시아인은 한국인인 것을 특징으로 하는 아시아인의 염증성 장질환 발병 위험 예측용 조성물.

청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 염증성 장질환은 크론병(Crohn's disease; CD) 또는 궤양성대장염(ulcerative colitis; UC)인 것을 특징으로 하는 아시아인의 염증성 장질환 발병 위험 예측용 조성물.

청구항 5

분리된 시료로부터 DNA를 추출하는 단계;

상기 추출한 DNA로부터 서열번호 1의 101번째 뉴클레오타이드, 서열번호 2의 101번째 뉴클레오타이드 및 서열번호 3의 101번째 뉴클레오타이드로 이루어진 군에서 선택된 어느 하나 이상의 유전자형을 확인하는 단계; 및

상기 확인된 유전자형으로 염증성 장질환 발병 위험을 예측하는 단계를 포함하는 아시아인의 염증성 장질환 발병 위험 예측 정보를 제공하는 방법.

청구항 6

제5항에 있어서, 상기 염증성 장질환 발병 위험을 예측하는 단계는 상기 서열번호 1의 101번째 뉴클레오타이드의 유전자형이 A, 상기 서열번호 2의 101번째 뉴클레오타이드의 유전자형이 C 또는 상기 서열번호 3의 101번째 뉴클레오타이드의 유전자형이 G인 경우, 염증성 장질환 발병 위험이 높은 것으로 예측하는 것을 특징으로 하는 아시아인의 염증성 장질환 발병 위험 예측 정보를 제공하는 방법.

청구항 7

제5항에 있어서, 상기 아시아인은 한국인인 것을 특징으로 하는 아시아인의 염증성 장질환 발병 위험 예측 정보를 제공하는 방법.

청구항 8

제5항 내지 제7항 중 어느 한 항에 있어서, 상기 염증성 장질환은 크론병(Crohn's disease; CD) 또는 궤양성대장염(ulcerative colitis; UC)인 것을 특징으로 하는 아시아인의 염증성 장질환 발병 위험 예측 정보를 제공하는 방법.

청구항 9

서열번호 1의 101번째 뉴클레오타이드, 서열번호 2의 101번째 뉴클레오타이드 및 서열번호 3의 101번째 뉴클레오타이드

드로 이루어진 군에서 선택된 어느 하나 이상의 단일염기다형성(SNP) 마커를 포함하는 10-100개의 연속적인 DNA 서열로 구성되는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드에 특이적으로 결합하는 프로브 또는 상기 폴리뉴클레오티드를 증폭하기 위한 프라이머를 포함하는 아시아인의 염증성 장질환 발병 위험 예측 키트.

청구항 10

제9항에 있어서, 상기 아시아인은 한국인인 것을 특징으로 하는 아시아인의 염증성 장질환 발병 위험 예측 키트.

청구항 11

제9항 또는 제10항에 있어서, 상기 염증성 장질환은 크론병(Crohn's disease; CD) 또는 궤양성대장염(ulcerative colitis; UC)인 것을 특징으로 하는 아시아인의 염증성 장질환 발병 위험 예측 키트.

발명의 설명

기술 분야

[0001] 본 발명은 한국인의 염증성장질환 위험 예측용 단일염기다형성 마커 조성물 및 이를 이용한 염증성장질환 위험 예측 방법에 관한 것이다.

배경 기술

[0002] 아시아에서 염증성장질환(inflammatory bowel disease; IBD)의 발생률(incidence) 및 유병률(prevalence)은 급격하게 증가하고 있는데, 이는 환경 변화에 기인한 것으로 추정된다. 여러 연구 결과에 따르면, IBD가 발병한 성별 비율, 병변의 위치 및 임상 경과와 같은 인구통계학적 및 표현형적 특징에 있어 아시아인 및 유럽인들 간에는 차이가 나타나는데, 이러한 결과를 토대로 다양한 개체군에서 유전적 및 환경적 영향을 모두 연구해야 할 필요성이 강조되고 있다. 동아시아인에서의 IBD 전장 유전체 연관성 분석(genome-wide association studies; GWAS), 대규모 유럽인 조상 연구 및 횡단 인종 연관성 분석 등 많은 연구에도 불구하고, IBD와 관련된 유전형을 명확히 이해하기에는 부족한 점이 많아 추가적인 연구가 필요한 실정이다.

선행기술문헌

비특허문헌

[0003] (비특허문헌 0001) Gut. 2014 Jan; 63(1):80-7

발명의 내용

해결하려는 과제

[0004] 본 발명의 목적은 rs3766920, rs16953946 및 rs3731257으로 이루어진 군에서 선택된 어느 하나 이상의 단일염기다형성 마커를 포함하는 아시아인의 염증성 장질환(inflammatory bowel disease; IBD) 발병 위험 예측용 조성물을 제공하는데 있다.

[0005] 또한, 본 발명의 다른 목적은 rs3766920, rs16953946 및 rs3731257으로 이루어진 군에서 선택된 어느 하나 이상의 단일염기다형성 마커의 대립유전자형을 확인하는 단계를 포함하는 아시아인의 염증성 장질환 발병 위험 예측 정보를 제공하는 방법을 제공하는데 있다.

[0006] 또한, 본 발명의 또 다른 목적은 rs3766920, rs16953946 및 rs3731257으로 이루어진 군에서 선택된 어느 하나 이상의 단일염기다형성 마커를 포함하는 폴리뉴클레오티드에 특이적으로 결합하는 프로브 또는 상기 폴리뉴클레오티드를 증폭하기 위한 프라이머를 포함하는 아시아인의 염증성 장질환 발병 위험 예측 키트를 제공하는데 있다.

과제의 해결 수단

[0007] 상기 과제의 해결을 위하여, 본 발명은 서열번호 1의 101번째 뉴클레오티드, 서열번호 2의 101번째 뉴클레오티

드 및 서열번호 3의 101번째 뉴클레오티드로 이루어진 군에서 선택된 어느 하나 이상의 단일염기다형성(SNP) 마커를 포함하는 10-100개의 연속적인 DNA 서열로 구성되는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 포함하는 아시아인의 염증성 장질환(inflammatory bowel disease; IBD) 발병 위험 예측용 조성물을 제공한다.

[0008] 또한, 본 발명은 분리된 시료로부터 DNA를 추출하는 단계; 상기 추출한 DNA로부터 서열번호 1의 101번째 뉴클레오티드, 서열번호 2의 101번째 뉴클레오티드 및 서열번호 3의 101번째 뉴클레오티드로 이루어진 군에서 선택된 어느 하나 이상의 유전자형을 확인하는 단계; 및 상기 확인된 유전자형으로 염증성 장질환 발병 위험을 예측하는 단계를 포함하는 아시아인의 염증성 장질환 발병 위험 예측 정보를 제공하는 방법을 제공한다.

[0009] 또한, 본 발명은 서열번호 1의 101번째 뉴클레오티드, 서열번호 2의 101번째 뉴클레오티드 및 서열번호 3의 101번째 뉴클레오티드로 이루어진 군에서 선택된 어느 하나 이상의 단일염기다형성(SNP) 마커를 포함하는 10-100개의 연속적인 DNA 서열로 구성되는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드에 특이적으로 결합하는 프로브 또는 상기 폴리뉴클레오티드를 증폭하기 위한 프라이머를 포함하는 아시아인의 염증성 장질환 발병 위험 예측 키트를 제공한다.

발명의 효과

[0010] 본 발명은 한국인의 염증성장질환 위험 예측용 단일염기다형성 마커 조성물 및 이를 이용한 염증성장질환 위험 예측 방법에 관한 것이다. 본 발명자들은 1,505명의 IBD 환자 및 4,041명의 대조군으로 이루어진 한국인 개체군에서의 IBD에 대한 GWAS를 수행하였고, 전장 유전체 상 IBD에 대한 감수성에 있어 유의성을 보이는 3개의 새로운 유전자좌위를 밝혀냈다. 또한, 본 발명자들은 28개의 알려진 IBD 유전자좌위($P < 2.16 \times 10^{-4}$)가 한국인에서도 연관성이 있는 것을 확인하였다. 상기 결과들은 다른 개체군의 유전형을 분석하는데 있어 비교 분석에 활용될 수 있다.

도면의 간단한 설명

[0011] 도 1은 발굴 단계에서 IBD에 대한 GWAS 결과를 나타낸다. (A) Quantile-quantile plot, $-\log_{10} P$ 수치는 예측된 null distribution에 대해 플롯하였다. 검은색 점은 전체 전장 유전체 통계를 나타내고(522,285 SNPs, $\lambda_{GC} = 1.106$), 붉은 색 점은 MHC 및 *TNFSF15* loci 바깥 부분의 SNPs를 나타낸다(516,236 SNPs, $\lambda_{GC} = 1.09$). (B) Manhattan plot, 새로운 유전자좌(loci)에 위치한 SNPs는 파란색으로 표시하였고, 전장 유전체 유의성 수준의 SNPs(파란 선, $P < 9.57 \times 10^{-8}$)는 붉은색으로 표시하였다. (C-E) 1q21에 위치한 rs3766920 (C), 16q23에 위치한 rs16953946 (D) 및 9p21에 위치한 rs3731257 (E)의 위치 연관성 플롯을 나타낸다. 원은 유전형 분석된 SNPs를 나타내고, 네모는 겹쳐져 추정된 SNPs를 나타낸다. 발굴 단계에서 가장 강하게 연관되어 있는 SNP는 작은 보라색 원으로 표시하였다. 선두 SNP 및 다른 SNPs 사이의 연관 불평형(linkage disequilibrium; LD)(r^2 수치)은 색깔을 사용하여 표시하였다. 분석된 유전자좌들의 관련 위치 및 전사 방향은 도면 아랫부분에 나타났다. 1000 Genomes Project (Nov 2014)로부터 얻은 아시아인 샘플의 추정 재조합 비율은 위치 LD 구조를 반영하여 플롯하였다. 플롯들은 LocusZoom (<http://csg.sph.umich.edu/locuszoom/>)을 사용하여 작성하였다.

도 2는 (유럽인에서의 소수 형질을 기초로) 유럽인 및 한국인에서 이전에 보고된 178개의 크론병(Crohn's disease; CD) 및 궤양성대장염(ulcerative colitis; UC) 위험 변이체들에 대한 오즈 비(odds ration; OR)를 비교한 결과이다. 각각의 점은 CD 및 UC의 각 SNP에 대한 OR(로그 좌표 상)을 나타낸다. 색깔은 한국인에서의 CD 또는 UC에 대한 연관성 P 수치 범위를 표시한다. 붉은색 선은 선형 회귀로서, 한국인에서의 로그 값(ORs) 변이의 역수를 가중치로 하였다. 선형모델의 상관계수(correlation coefficient) 및 P 수치는 오른쪽 아래 구석에 나타났다.

발명을 실시하기 위한 구체적인 내용

[0012] 본 발명은 서열번호 1의 101번째 뉴클레오티드, 서열번호 2의 101번째 뉴클레오티드 및 서열번호 3의 101번째 뉴클레오티드로 이루어진 군에서 선택된 어느 하나 이상의 단일염기다형성(SNP) 마커를 포함하는 10-100개의 연속적인 DNA 서열로 구성되는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 포함하는 아시아인의 염증성 장질환(inflammatory bowel disease; IBD) 발병 위험 예측용 조성물을 제공한다.

[0013] 상세하게는, 상기 서열번호 1의 101번째 뉴클레오티드의 대립유전자형은 A/G이고, 상기 서열번호 2의 101번째

뉴클레오타이드의 대립유전자형은 C/T이고, 상기 서열번호 3의 101번째 뉴클레오타이드의 대립유전자형은 A/G일 수 있다.

- [0014] 바람직하게는, 상기 아시아인은 한국인일 수 있으나, 이에 제한되는 것은 아니다.
- [0015] 바람직하게는, 상기 염증성 장질환은 크론병(Crohn's disease; CD) 또는 궤양성대장염(ulcerative colitis; UC)일 수 있으나, 이에 제한되는 것은 아니다.
- [0016] 상기 서열번호 1의 101번째 뉴클레오타이드를 포함하는 폴리뉴클레오타이드에서 다형성 SNP 변이는 rs3766920로 표시하여 기재할 수 있다. 상기 rs3766920는 1q21 상의 *PYGO2-SHC* 내에 존재하며, 대립유전자형은 A/G이다. 상기 서열번호 1의 101번째 뉴클레오타이드는 다중염기서열 분석 방식에 따라 작성하였는데, A 또는 G일 수 있으므로 "r"이라고 기재하였다.
- [0017] 상기 서열번호 2의 101번째 뉴클레오타이드를 포함하는 폴리뉴클레오타이드에서 다형성 SNP 변이는 rs16953946로 표시하여 기재할 수 있다. 상기 rs16953946는 16q23 상의 *CDYL2* 내에 존재하며, 대립유전자형은 C/T이다. 상기 서열번호 2의 101번째 뉴클레오타이드는 다중염기서열 분석 방식에 따라 작성하였는데, C 또는 T일 수 있으므로 "y"라고 기재하였다.
- [0018] 상기 서열번호 3의 101번째 뉴클레오타이드를 포함하는 폴리뉴클레오타이드에서 다형성 SNP 변이는 rs3731257로 표시하여 기재할 수 있다. 상기 rs3731257는 9p21 상의 *CDKN2A-AS1-CDKN2B* 내에 존재하며, 대립유전자형은 A/G이다. 상기 서열번호 3의 101번째 뉴클레오타이드는 다중염기서열 분석 방식에 따라 작성하였는데, A 또는 G일 수 있으므로 "r"이라고 기재하였다.
- [0019] 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 '단일염기다형성', 즉 SNP(single nucleotide polymorphism)라고 한다.
- [0020] 본 발명의 "연관불평형(Linkage Disequilibrium; LD)"이란 집단유전학에서 반드시 동일 염색체상에 존재하는 두 개 이상의 유전자좌 (loci)에서의 대립유전자의 비-무작위적 연관 (non-random association)을 의미한다.
- [0021] 또한, 본 발명은 분리된 시료로부터 DNA를 추출하는 단계; 상기 추출한 DNA로부터 서열번호 1의 101번째 뉴클레오타이드, 서열번호 2의 101번째 뉴클레오타이드 및 서열번호 3의 101번째 뉴클레오타이드로 이루어진 군에서 선택된 어느 하나 이상의 유전자형을 확인하는 단계; 및 상기 확인된 유전자형으로 염증성 장질환 발병 위험을 예측하는 단계를 포함하는 아시아인의 염증성 장질환 발병 위험 예측 정보를 제공하는 방법을 제공한다.
- [0022] 상세하게는, 상기 염증성 장질환 발병 위험을 예측하는 단계는 상기 서열번호 1의 101번째 뉴클레오타이드의 유전자형이 A, 상기 서열번호 2의 101번째 뉴클레오타이드의 유전자형이 C 또는 상기 서열번호 3의 101번째 뉴클레오타이드의 유전자형이 G인 경우, 염증성 장질환 발병 위험이 높은 것으로 예측할 수 있다.
- [0023] 바람직하게는, 상기 아시아인은 한국인일 수 있으나, 이에 제한되는 것은 아니다.
- [0024] 바람직하게는, 상기 염증성 장질환은 크론병(Crohn's disease; CD) 또는 궤양성대장염(ulcerative colitis; UC)일 수 있으나, 이에 제한되는 것은 아니다.
- [0025] 상기 분리된 시료로부터 DNA를 추출하는 단계에서, DNA는 대상의 혈액, 피부세포, 점막 세포 및 모발 등 모든 세포로부터 분리될 수 있다. 해당 세포로부터 DNA를 추출하는 방법은 특별히 한정되지 않으며, 당업계에 공지된 기술 또는 시판되고 있는 DNA 추출용 키트를 사용할 수 있다.
- [0026] 상기 유전자형을 확인하는 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 공지된 방법 중 선택된 어느 하나 이상을 사용할 수 있다.
- [0027] 또한, 본 발명은 서열번호 1의 101번째 뉴클레오타이드, 서열번호 2의 101번째 뉴클레오타이드 및 서열번호 3의 101번째 뉴클레오타이드로 이루어진 군에서 선택된 어느 하나 이상의 단일염기다형성(SNP) 마커를 포함하는 10-100개

의 연속적인 DNA 서열로 구성되는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드에 특이적으로 결합하는 프로브 또는 상기 폴리뉴클레오티드를 증폭하기 위한 프라이머를 포함하는 아시아인의 염증성 장질환 발병 위험 예측 키트를 제공한다. 상기 키트는 PCR 키트 또는 DNA 칩 키트가 바람직하지만, 이에 제한되는 것은 아니다.

[0028] 바람직하게는, 상기 아시아인은 한국인일 수 있으나, 이에 제한되는 것은 아니다.

[0029] 바람직하게는, 상기 염증성 장질환은 크론병(Crohn's disease; CD) 또는 궤양성대장염(ulcerative colitis; UC)일 수 있으나, 이에 제한되는 것은 아니다.

[0030] 상기 "프로브"는 mRNA와 특이적으로 결합을 이룰 수 있는 짧은 수 염기 내지 길게는 수백 염기에 해당하는 RNA 또는 DNA 등의 핵산 단편을 의미하며 라벨링되어 있어서 특정 mRNA의 존재 유무, 발현양을 확인할 수 있다. 프로브는 올리고뉴클레오티드(oligonucleotide) 프로브, 단쇄 DNA(single strand DNA) 프로브, 이중쇄 DNA(double strand DNA)프로브, RNA 프로브 등의 형태로 제작될 수 있다. 적절한 프로브의 선택 및 혼성화 조건은 당해 기술 분야에 공지된 기술에 따라 적절히 선택할 수 있다.

[0031] 상기 "프라이머"는 짧은 자유 3'-말단 수산화기(free 3' hydroxyl group)를 가지는 핵산 서열로 상보적인 템플레이트(template)와 염기쌍을 형성할 수 있고 템플레이트 가닥 복사를 위한 시작 지점으로서 작용하는 짧은 핵산 서열을 말한다. 프라이머는 적절한 완충용액 및 온도에서 중합반응을 위한 시약(즉, DNA 폴리머라제 또는 역전사효소) 및 상이한 4 가지의 뉴클레오사이드 트리포스페이트의 존재 하에서 DNA 합성을 개시할 수 있다. PCR 조건, 센스 및 안티센스 프라이머의 길이는 당업계에 공지된 기술에 따라 적절히 선택될 수 있다.

[0032] 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.

[0033] <실험예>

[0034] 하기의 실험예들은 본 발명에 따른 각각의 실시예에 공통적으로 적용되는 실험예를 제공하기 위한 것이다.

[0035] 1. 실험 대상

[0036] 총 3,494명의 IBD 환자(1,915명의 CD 환자 및 1,579명의 UC 환자) 및 7,532명의 건강한 대조군이 본 발명에 포함되었다. 본 발명에 포함된 개체의 임상적 특징은 표 1에 나타났다. 본 발명에 참가한 모든 참가자는 한국인이었다. 모든 IBD 환자는 아산병원 IBD 클리닉으로부터 모집하였다. IBD 진단은 기존의 임상적 판단, 방사선 검사, 내시경 검사 및 병리조직학적 검사 기준에 기초하였다. 불확정 대장염(indeterminate colitis)을 가진 환자는 본 발명에서 제외하였다. CD 환자들은 약간 개량한 몬트리얼 분류법(Montreal classification)을 이용하여 하위 분류하였다. 진단 시 나이(A1: 16살 이하; A2: 17 내지 40살; A3: 40살 이상), 병변의 위치(L1: 회장 말단(terminal ileum); L2: 결장(colon); L3: 회결장(ileocolon)) 및 병변의 발현 형태(B1: 염증형(inflammatory); B2: 협착형(stricturing); B3: 관통형(penetrating)). 몬트리얼 분류 시스템 중 상부 위장관(upper gastrointestinal) 및 항문 주변 병변(perianal disease modifier)은 적용하지 않았다. UC 환자들은 몬트리얼 분류법(Montreal classification)을 기초로, 최대 내시경 분포 범위(maximal endoscopic extent)에 따라 다음의 3개 카테고리 분류하였다. 궤양성 직장염(ulcerative proctitis), 좌측 결장염(left-sided colitis) 및 전결장염(extensive colitis).

[0037] 한국인에서 IBD 위험 연관성 유전자좌를 확인하고, GWAS에 의한 발굴 분석의 정확도를 높이기 위해서, 본 발명자들은 이전에 보고한 GWAS에서 사용된 915명의 IBD 환자(532명의 CD 환자, 383명의 UC 환자) 및 745명의 대조군 샘플과 함께, 국립보건연구원의 개체군 데이터로부터 590명의 IBD 환자(390명의 CD 환자, 200명의 UC 환자) 및 3,296명의 대조군 세트를 추가하여 조합하였다. 이전 및 현재 발굴 GWAS 패널, 독립적인 재현 패널 및 사용된 유전형 플랫폼은 표 1에 나타났다. 재현 코호트는 1,989명의 IBD 환자(993명의 CD 환자, 996명의 UC 환자) 및 3,491명의 건강한 대조군(국립보건연구원으로부터 제공받은 2,496명의 대조군 및 H-D Shin으로부터 제공받은 995명의 대조군)으로 구성되어 있다.

[0038] 본 발명은 아산병원 임상시험심사위원회의 승인을 받았고, 모든 샘플은 아산병원 임상시험심사위원회가 승인한 프로토콜 하에서 참가자로부터 사전 동의를 얻어 확보하였다.

	GWAS				Replication				Total			
	IBD	CD	UC	Control	IBD	CD	UC	Control	IBD	CD	UC	Control
No. of samples	1,505	922	583	4,041	1,989	993	996	3,491	3,494	1,915	1,579	7,532
Male (%)	976 (64.9)	651 (70.6)	325 (55.7)	1,602 (39.6)	1,275 (64.1)	723 (72.8)	552 (55.4)	1,771 (50.7)	2,251 (64.4)	1,374 (71.7)	877 (55.5)	3,373 (44.8)
Mean age at sampling (years)	31.1 ± 13.6	25.5 ± 9.1	40.1 ± 14.7	NA	34.9 ± 13.6	28.5 ± 9.4	41.3 ± 14.2	NA	33.3 ± 13.7	27.0 ± 9.4	40.8 ± 14.4	NA
Mean age at diagnosis (years)	27.6 ± 12.6	22.3 ± 8.2	36.0 ± 13.8		31.1 ± 12.8	25.4 ± 8.7	36.8 ± 13.7		29.6 ± 12.9	23.9 ± 8.6	36.5 ± 13.7	
Age group at diagnosis (%)												
≤ 16	302 (20.1)	249 (27.0)	53 (9.1)	NA	114 (5.7)	80 (8.1)	34 (3.4)	NA	416 (11.9)	329 (17.2)	87 (5.5)	NA
17–40	917 (60.9)	634 (68.8)	283 (48.5)	NA	1,419 (71.3)	841 (84.7)	578 (58.0)	NA	2,336 (66.9)	1,475 (77.0)	861 (54.5)	NA
≥ 40	286 (19.0)	39 (4.2)	247 (42.4)	NA	456 (22.9)	72 (7.3)	384 (38.6)	NA	742 (21.2)	111 (5.8)	631 (40.0)	NA
Location, no. (%)												
Ileum		161 (19.4)				228 (23.0)			389 (21.4)			
Colon		45 (5.4)				36 (3.6)			81 (4.5)			
Ileocolon		623 (75.2)				726 (73.3)			1,349 (74.2)			
NA		93				3			96			
Extent, no. (%)												
Proctitis			160 (29.4)				260 (26.3)				420 (27.4)	
Left-sided colitis			176 (32.4)				321 (32.5)				497 (32.4)	
Extensive colitis			208 (38.2)				407 (41.2)				615 (40.1)	
NA			39				8				47	
Behavior, no. (%)												
Inflammatory		325 (39.2)				453 (45.6)			778 (42.7)			
Structuring		164 (19.8)				186 (18.7)			350 (19.2)			
Penetrating		341 (41.1)				354 (35.6)			695 (38.1)			
NA		92							92			
Perianal fistula, no. (%)												
No		329 (39.6)				456 (45.9)			785 (43.1)			
Yes		501 (60.4)				537 (54.1)			1,038 (56.9)			
NA		92							92			
GWAS set 1												
Illumina chip				Omni-Quad								
No. of samples	915	532	383	745								
GWAS set 2												
Illumina chip				Omni-Quad								
No. of samples	500	390	200	3,296								
Replication set												
Genotyping technology												
No. of samples												

Sequenom PLEX, TaqMan SNP genotyping assay
1,989 3,491*

UC, ulcerative colitis; KNIH, Korean National Institute of Health

All patients with IBD were recruited at the Asan Medical Center (AMC) in Seoul.

The samples in the GWAS set 1 were used in the previous GWAS on CD and UC separately (ref 5, 6). The patients with IBD were recruited at the First Medical Center (PWHC) in Beijing.

Of 5,491 controls in the replication cohort, 2,496 controls were from the National Biobank of Korea supported by the Ministry of Health, Welfare and Family Affairs and 995 controls were from Prof. H-D Sain

2. 유전형 분석 및 확장된 GWAS 분석의 품질 관리

Illumina OmniExpressExome arrays를 사용하여 전체 595명의 IBD 환자 개체를 새롭게 유전형 분석하였다. 비교를 위해, 본 발명자들은 이전에 발표한 GWAS에 기재되어 공개적으로 구할 수 있는 800명의 데이터와, 추가적인 3,351명의 정상 대조군 개체로부터 얻은 데이터를 조합하였다. 모든 대조군 개체는 동일한 Illumina Omni1-Quad arrays를 사용하여 유전형 분석하였다. IBD 케이스에 있어서, 본 발명자들은 이전에 발표한 GWAS에 기재된 538명의 CD 환자 및 397명의 UC 환자로부터 얻은 샘플과, 추가적인 595명의 IBD 환자(395명의 CD 환자 및 200명의 UC 환자)로부터 얻은 샘플을 조합하였다. CD 및 UC 케이스는 Illumina OmniExpress arrays를 사용한 반면, IBD 케이스는 Illumina OmniExpressExome arrays를 사용하여 본 발명에서 새롭게 유전형 분석하였다. 전체적으로, 발굴 코호트는 1,530명의 IBD 케이스 및 4,151명의 대조군으로 구성되었다. 품질 관리는 각각의 데이터세트 상에서 별도로 수행되었다. 복제수 변이-관련 SNPs 뿐만 아니라 X, Y 및 미토콘드리아 염색체 상의 모든 SNPs는 제외하였다. 일부 품질 관리에 있어서, 만약 유전자형 분석 성공률(call rate) <98%, 소수 형질 빈도(minor

allele frequency; MAF) < 0.01 및 대조군에서 Hardy-Weinberg equilibrium으로부터의 유의성 편차가 $P < 1 \times 10^{-5}$ 이면, SNPs는 제외하였다. 유사하게, 본 발명자들은 유전형 분석을 $< 96\%$ 인 모든 샘플을 추후 분석에서 제거하였다. 본 발명자들은 PLINK 1.9 software (<http://pngu.mgh.harvard.edu/~purcell/plink/>)를 사용한 인종 분석을 토대로 5,681명 샘플의 유전적 연관성을 확인하였다. 1차 또는 2차 관련 샘플($PI_{HAT} > 0.2$, $IBS > 0.8$) 각각의 세트에 대하여, 더 낮은 유전형 분석 성공률(call rate)을 가진 샘플은 제거하였다. 따라서, 샘플 중복 및/또는 유전적 연관성으로 인해, 총 106명의 샘플이 제거되었다(15명 환자 케이스 및 91명 대조군). 그 후, 본 발명자들은 2세트의 IBD 케이스 및 2 세트의 대조군을 조합하였다. 조합된 샘플 세트에서 다시 품질 관리를 수행하였다. 조합된 샘플 세트에 있어서, 유전자형 분석 성공률(call rate) $< 98\%$, 대조군에서 Hardy-Weinberg equilibrium을 보이는 마커의 P 수치 $< 1.0 \times 10^{-5}$, MAF < 0.01 또는 케이스 및 대조군 사이의 차등적 결측치(differential missingness)가 $P < 1.0 \times 10^{-5}$ 인 경우, 분석에서 제외하였다. 샘플 중복 및/또는 유전적 연관성으로 인해, 추가적으로 25명의 샘플이 제거되었다(9명 환자 케이스 및 16명 대조군). 상기와 같이 SNP 및 샘플 품질 관리 분석을 수행한 후, 1,506명 환자 케이스 및 4,044명 대조군에서의 522,285 SNPs 유전형 데이터를 GWAS 분석에 적용하였다.

[0042] 다음으로, 본 발명자들은 software package EIGENSTRAT 3.0으로 개체군 이상치(outliers) 및 층화(stratification)를 검출하기 위해서, 주성분 분석(principal-component analysis; PCA)-기반 방법을 사용하였다. 우선, International HapMap Project로부터 얻은 194개의 대조 샘플과 함께, 5,550명의 샘플(1,506명 IBD 환자 및 4,044명 대조군)을 분석하였다. 4개의 개체군 이상치(1명 환자 케이스 및 3명 대조군)가 확인되었고, 제거하였다. 상기와 같은 SNP 및 샘플 품질 관리 분석 후, 본 발명자들은 최종적으로 1,505명 환자 케이스 및 4,041명 대조군에서의 522,285 SNPs 유전형 데이터를 GWAS 분석에 사용하였다.

[0043] 3. 결측치 추정(Imputation)

[0044] 본 발명자들은 상기에 기재한 품질 관리를 통과한 유전형 분석된 SNPs를 사용하여, 5,546명 샘플(1,505명 IBD 환자 및 4,041명 대조군)의 결측치를 추정하였다. 결측치 추정(Imputation)은 IMPUTE version 2.0 (https://mathgen.stats.ox.ac.uk/impute/impute_v2.html) 및 the Asian reference data (JPT + CHB) from the 1,000 Genomes Project databases (Feb2012release) (<http://www.1000genomes.org/>)를 사용하여 수행하였다. 낮은 결측치 추정 정도(사후 확률 점수 $< 80\%$), MAF $< 1\%$ 또는 결측율 $> 10\%$ 를 나타내는 SNPs는 추후 분석에서 제외하였다. 총 4,148,401개의 결측치 추정된 SNPs는 품질 관리를 통과하였고, 연관 분석을 위해 522,285개의 유전형 분석된 SNPs와 조합하였다. 다량의 반복적인 시험 분석의 부담을 피하기 위해서, 결측치 추정된 SNPs는 발굴 및 재현 분석에는 모두 포함시키지 않았다. 결측치 추정된 데이터는 발굴 단계에서 확인된 유전자좌의 위치 플롯에 있어서, 연관성 패턴을 분석하기 위해서만 사용되었다.

[0045] 4. 재현을 위한 SNP 선별

[0046] 전장 유전체 발굴 분석 후, 독립적인 코호트에서 재현을 위해 $P < 10^{-4}$ 를 나타내는 SNPs를 선별하였다. 새로운 감수성 유전자좌를 밝혀내기 위해서, 본 발명자들은 연관성 증거를 나타내는 > 1 SNP를 가진 영역을 선택하였고, 이전에 보고된 GWAS에서 IBD 위험 변이체로 알려진 적 없는, 낮은 연관 불평형(linkage disequilibrium; LD)을 가진 SNPs를 선택하였다. 1,989명의 IBD 환자 및 3,491명의 대조군으로 이루어진 독립적인 세트(재현 코호트)에서의 검증을 위해, $P < 10^{-4}$ 의 연관성을 보이는 60개의 유전자좌로부터 총 60개의 SNPs를 선별하였다.

[0047] 재현 코호트의 유전형 분석은 Sequenom iPLEX system (Analytical Genetics Technology Centre, Princess Margaret Hospital/ University Health Network in Toronto, Canada) 또는 TaqMan genotyping technology (7900HT Fast Real Time PCR System, Applied Biosystems)를 사용하여 수행하였다. rs3766920 및 rs4821558에 대한 검증은 TaqMan SNP genotyping assays를 사용하여 수행하였다. 본 발명자들은 유전자형 분석 성공률(call rate) $< 97.5\%$ 또는 Hardy-Weinberg equilibrium 편차 $P < 0.05$ 를 보이는 SNPs는 제외하였다. 분석 실패로 인하여 본 발명자들의 샘플에서 재현이 성공적으로 수행되지 않은 rs7335629에 대해서는, Japanese CD 발굴 분석 결과 및 본 발명자들의 발굴 분석 결과 사이의 메타-분석을 수행하였다.

[0048] 5. 통계 분석

[0049] IBD에 대한 전장 유전체 연관성 분석을 수행하였다. 연관성 분석은 Cochran-Armitage trend test를 사용하여 수행하였다. 전장 유전체 연관성의 전체적인 유의성 및 개체군 층화(stratification)의 잠재적 영향을 확인하기 위해, R(3.2.0) (<http://www.r-project.org/>)을 이용하여 quantile-quantile plot을 작성하였다. 또한, 개체군

층화(stratification)의 영향은 genomic control inflation factor($\lambda_{GC}=1.106$) 및 1000개의 케이스와 1000개의 대조군에 대한 genomic inflation factor($\lambda_{GC1000}=1.048$)를 계산하여 확인하였다. 실질적인 원인 변이체(causal variants)에 있어 다유전자형 구조 및 LD는 λ_{GC} 에 영향을 줄 수 있으므로, 본 발명자들은 stringent LD pruning($r^2 < 0.1$, $\lambda_{GC}=1.059$ after pruning) 후에 λ_{GC} 를 평가하였다. 또한, 로지스틱 회귀분석(logistic regression)에 있어 10개의 주성분을 포함한 후($\lambda_{GC}=1.102$) 및 선형 혼합 모델(linear mixed model)에 적용한 후($\lambda_{GC}=1.006$), λ_{GC} 를 평가하였다. $-\log_{10} P$ 의 Manhattan plot은 R(3.2.0) (<http://www.r-project.org/>)을 사용하여 작성하였다. 전장 유전체 유의성 유전자좌에서 독립적인 연관성을 확인하기 위해서, 조건부 회귀 분석(Conditional regression analysis)을 수행하였다. 재현 샘플들에서 유전형-표현형 연관성을 시험하기 위해서, Cochran-Armitage trend tests를 사용하였다. 조합된 GWAS 및 재현 샘플들의 최종 분석은 Cochran-Mantel-Haenszel test를 사용하여 수행하였다. 또한, 다른 코호트들로부터 샘플에 대한 오즈비(odds ratios, ORs) 사이의 이질성에 대한 유의성을 평가하기 위해서, Breslow-Day tests를 수행하였다.

[0050] 본 발명자들은 Bonferroni correction을 이용하여 본 발명에서 정한 $P < 9.57 \times 10^{-8}$ (0.05/522,285)의 전장 유전체 유의성 수준(genome-wide significance threshold)을 사용하였다. 유럽인 개체군에서 이미 보고된 IBD 유전자좌를 재현하기 위해서, 본 발명자들은 $P < 2.16 \times 10^{-4}$ (0.05/231, 상기 231은 유럽인 개체군에서 IBD와 연관된 것으로 알려진 200개의 유전자좌 유래 독립적인 SNPs 수)의 유의성 수준을 적용하였고, 본 발명의 데이터셋에 있어서 CD, UC 또는 IBD에 대한 연관성 중 어느 것이 상기 유의성 수준을 초과하는지 확인하였다. 유럽인 개체군에서 이전에 보고된 200개의 IBD 위험성 유전자좌를 검출하기 위한 본 발명의 GWAS 샘플 검정력 분석(Power analysis)은 Quanto software version 1.2.4 (<http://hydra.usc.edu/gxe/>)을 이용하여 수행하였다. 보고된 각각의 SNP에 있어서, 명목상 2.16×10^{-4} (0.05/231)의 P 수치에서 검출하기 위한 검정력은 보고된 오즈 비(OR) 및 한국인 개체군에서의 형질 빈도에 기초하여 계산되었다. 또한, 본 발명자들은 1000 Genomes databases (Feb2012release) (<http://www.1000genomes.org/>)를 사용하여, 아시아인 및 유럽인 개체군 간에 새로운 유전자좌 3개에 대한 LD 구조를 비교하였다.

[0051] 6. CD 및 UC 사이의 유전적 관련성

[0052] CD 및 UC 사이에 공유하는 유전적 변이의 비율은 genome-wide complex trait analysis tool (GCTA) 내의 이변량 선형 혼합-효과 모델(bivariate linear mixed-effects model)를 사용하여 측정하였다. 이변량 분석에 있어서, 대조군은 CD 및 UC 케이스 수에 대한 비율에 따라 나누었다.

[0053] <실시예>

[0054] 본 발명자들은 한국인 개체에서 IBD와 관련된 GWAS를 수행하였는데, 발굴 단계에서 1,505명 환자 케이스(CD: 922; UC: 583) 및 4,041명의 대조군이 포함되었고, 재현 코호트에서 추가적으로 1,989명의 환자 케이스(CD: 993; UC: 996) 및 3,491명의 대조군이 포함되었다(표 1). 모든 개체들은 표준 절차에 따라 IBD로 진단되었다. 엄격한 품질 관리 후, 본 발명자들은 로지스틱 회귀분석(logistic regression)의 경향성 분석을 이용하여 발굴 그룹에서 522,285 SNPs를 분석하였다. quantile-quantile plots은 정규 분포를 나타냈다(도 1). genomic inflation factor λ_{GC} 는 1.106 이었다(선형 혼합 모델에 적용시 1.006). 발굴 GWAS에서는 높은 전장 유전체 유의성을 나타내는 9개의 유전자좌(loci)를 확인하였는데(522,285 SNPs에 대해 Bonferroni correction을 이용하여 $P < 9.57 \times 10^{-8}$ 로 정함), 2개의 새로운 유전자좌(*PYGO2-SHC1*, *NCF4-CSF2RB*) 및 7개의 보고된 IBD 위험 유전자좌(*TNFSF15-TNFSF8*, *MHC*, *TBC1D1-KLF3*, *STAT3-STAT5B-STAT5A*, *TNFRSF6B*, *SMNDC1-DUSP5*, *GPR35*)가 포함되었다. 추가적으로 58개의 유전자좌가 연관성이 있는 것으로 나타났다($P < 1 \times 10^{-4}$). 상기 2개의 새로운 유전자좌 및 58개의 가능성 있는 유전자좌를 검증하기 위해서, 본 발명자들은 각 유전자좌에서 가장 유의성이 높은 SNP를 선별하고, 재현 코호트에서 IBD와의 관련성을 확인하였다. 결론적으로, 본 발명자들은 발굴 및 재현 분석을 조합하여 전장 유전체 유의성이 있는 3개의 새로운 IBD 유전자좌를 밝혀냈다: 1q21의 *PYGO2-SHC* 부위 내 rs3766920 (combined $P = 1.96 \times 10^{-9}$), 16q23의 *CDYL2* 부위 내 rs16953946 (combined $P = 2.48 \times 10^{-8}$) 및 9p21의 *CDKN2A-ASI-CDKN2B* 부위 내 rs3731257 (combined $P = 5.45 \times 10^{-8}$) (표 2 및 도 1). likelihood-modeling 분석 결과, 새롭게 밝혀진 유전자좌는 CD 및 UC 모두 연관성이 있었다. 각 유전자좌의 조건부 분석 결과, 더 이상 독립적이고 유의성 있는 추가적인 전장-유전체 신호는 나타나지 않았다. 3개의 유전자좌 모두는 가능성 있는 후보 유전

자들 내에 있다. 녹아웃 마우스를 이용한 실험 결과, *Shc1*은 흥선 T 세포 발달의 후기 단계 및 말초 T 세포의 존적 기전에 있어서 중요한 역할을 하는 것으로 나타났고, *Cdkn2a*는 대식세포(macrophage) 활성화 및 JAK2-STAT1 기작을 통한 분극화를 조절하는데 중요한 역할을 하며, *CDYL2* 내에 위치한 변이(rs4613079, $r^2 < 0.2$)는 아프리카계 미국인에서 인터루킨-10(interleukin-10; IL-10) 수준과 연관성이 있는 것으로 나타났다. 본 발명자들은 상기 3개의 새로운 유전자좌가 유럽인 개체에서는 왜 유의성을 나타내지 않는지 확인하였다. 유럽인에 있어서, rs3766920은 단형성(monomorphic)이었다. 다른 2개의 유전자좌는 다형성이었고, 이에 본 발명자들은 위치적 LD 구조를 분석하였다. rs3731257에 대해서, 유럽인에서는 SNPs 부근에 있어 상기 SNP의 LD는 약했다. 아시아인에서 rs3731257에 대해 $r^2 > 0.5$ 인 12개의 SNPs 중, 11개의 SNPs가 유럽인보다 r^2 가 낮았다($P = 0.006$ for testing differential LD using 2-sided exact binomial test). 상기 결과는 LD 차이가 개체군 사이의 SNP 연관성 차이를 설명할 수도 있다는 것을 나타낸다. rs16953946에 대해서, 유럽인들에서는 LD가 일정한 수준으로 낮게 나타나지는 않았는데(17 SNPs 중 11개에 있어서 $r^2_{EUR} < r^2_{ASN}$, differential LD $P > 0.3$), 이는 빈도 또는 LD 차이와는 다른 인자들이 유효 크기에 있어서 이질성(heterogeneity)을 야기할 수 있다는 것을 나타낸다.

표 2

Chr	SNP	Position (hg19)	Candidate gene(s)	Risk allele	Study	No. of cases	No. of controls	RAF in cases	RAF in controls	OR (95% CI)	P*	P _{BD} ^Δ	LR phenotype [‡]
1q21	rs3766920	154,934,963	<i>PYGO2, SHCI</i>	A	GWAS	1,505	4,039	0.11	0.08	1.48 (1.29–1.70)	3.56×10^{-3}	0.084	IBD_U
					Replication	1,977	3,485	0.09	0.08	1.24 (1.08–1.43)	0.002		
					Combined [‡]	3,482	7,524	0.10	0.08	1.35 (1.22–1.49)	1.96×10^{-2}		
16q23	rs16953946	80,785,448	<i>CDYL2</i>	C	GWAS	1,505	4,040	0.20	0.17	1.26 (1.13–1.40)	1.85×10^{-3}	0.484	IBD_U
					Replication	1,981	3,485	0.20	0.17	1.20 (1.09–1.33)	2.90×10^{-4}		
					Combined [‡]	3,486	7,525	0.20	0.17	1.23 (1.14–1.32)	2.48×10^{-3}		
9p21	rs3731257	21,966,221	<i>CDKN2A-AS1, CDKN2A, CDKN2B-AS1, CDKN2B</i>	G	GWAS	1,504	4,041	0.51	0.46	1.21 (1.11–1.31)	9.01×10^{-6}	0.309	IBD_U
					Replication	1,982	3,453	0.50	0.46	1.14 (1.05–1.23)	0.001		
					Combined [‡]	3,486	7,494	0.50	0.46	1.17 (1.11–1.24)	5.45×10^{-3}		

Chr, chromosome; CI, confidence interval; IBD, inflammatory bowel disease; OR, odds ratio; Position, chromosome position; RAF, risk allele frequency.

*P-value was calculated using the Cochran-Armitage trend test.

^ΔP_{BD} was calculated using the Breslow-Day test.

[‡]The most strongly associated phenotype was assigned using likelihood ratio modeling. IBD unsaturated (IBD_U) loci were associated with both Crohn's disease and ulcerative colitis with no evidence of differences in effect size.

[‡]Combined P-value was calculated using the Cochran-Mantel-Haenszel test.

[0055]

[0056]

한국인 데이터를 사용하여, 본 발명자들은 이전에 보고된 200개의 '유럽인' IBD-연관 유전자좌(231개 독립적인 SNPs)를 확인하였다. 본 발명의 개체군에서는 161개의 유전자좌에서 178개의 SNPs 데이터를 얻을 수 있었다. 이들 중, 25개의 유전자좌에서 총 26개의 SNPs가 재현되었는데, 이는 아마도 본 발명의 코호트 샘플 규모를 4개 유전자좌에서 > 80% 검정력을 갖도록 평가하였기 때문인 것으로 보인다. 178개 SNPs 전체에 대한 아시아인 및 유럽인 사이의 유효 크기 비교 결과, 이전 연구 결과와 일치하게, CD 및 UC 모두 유효성의 방향에 있어 양성 연관성을 나타냈다(도 2; $r^2 = 0.40$ and $P = 3.56 \times 10^{-21}$ for CD; $r^2 = 0.53$ and $P = 1.10 \times 10^{-30}$ for UC). 아시아인에서는 최초로 밝혀낸 것으로 유럽인에서는 재현되지 않은, 아시아인-특이적인 상기 3개의 유전자좌는 본 연구를 통해서도 재현되었다.

[0057]

CD 또는 UC 위험 형질 각각에 대해 나타나는 표현형 변이 백분율을 측정하기 위해서, 본 발명자들은 3개의 새로운 유전자좌(3 SNPs) 및 28개의 알려진 유전자좌(29 SNPs) 뿐만 아니라 이전에 보고된 CD 또는 UC 위험 유전자

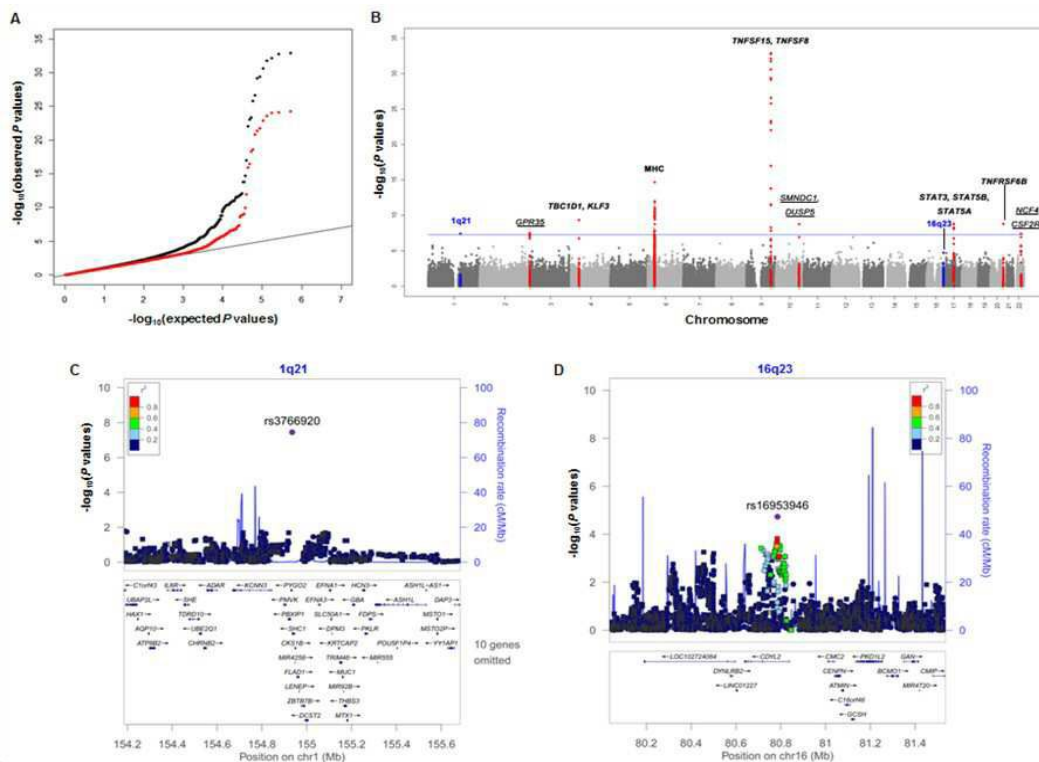
좌도 사용하였는데, 총 CD에 대하여 30 SNPs 및 UC에 대하여 30 SNPs이다(CD 및 UC 모두 해당하는 19 SNPs가 포함). 한국인 개체에서 CD 유병률(prevalence)이 0.0112%인 것을 고려하면, 30개의 유전자좌는 전체 CD 유전적 위험에 대해 6.82%에 해당되며, UC 유병률이 0.0309%인 것을 고려하면, 30개의 유전자좌는 전체 UC 유전적 위험에 대해 5.53%에 해당된다. 이변량 선형 혼합 모델(bivariate linear mixed model)에 적용하여 CD 및 UC 사이의 유전적 관련성을 측정하면, CD 및 UC 사이의 위험에 대한 유전적 연관성은 0.47이었다(표준오차, 0.08). 상기 결과는 백인 데이터(0.62; 표준 오차, 0.042) 보다는 낮게 측정되었는데, 이는 백인 및 아시아인 사이의 IBD에 대한 표현형 차이로 인한 것으로 추정된다.

[0058] 결론적으로, 본 발명자들은 IBD 발병 기작에 관여하는 T 세포 생물학, JAK-STAT 신호전달 및 IL10 기작과 연관되어 있는 3개의 새로운 유전자좌를 밝혀냈다. 또한, 본 발명자들은 이전에 보고된 28개의 유전자좌로부터 29개의 SNPs를 재현하였다. 따라서, 본 발명은 백인 및 아시아인에서 나타나는 IBD와 연관된 공통 기작 뿐만 아니라 상이한 기작을 밝히는데도 활용될 수 있다.

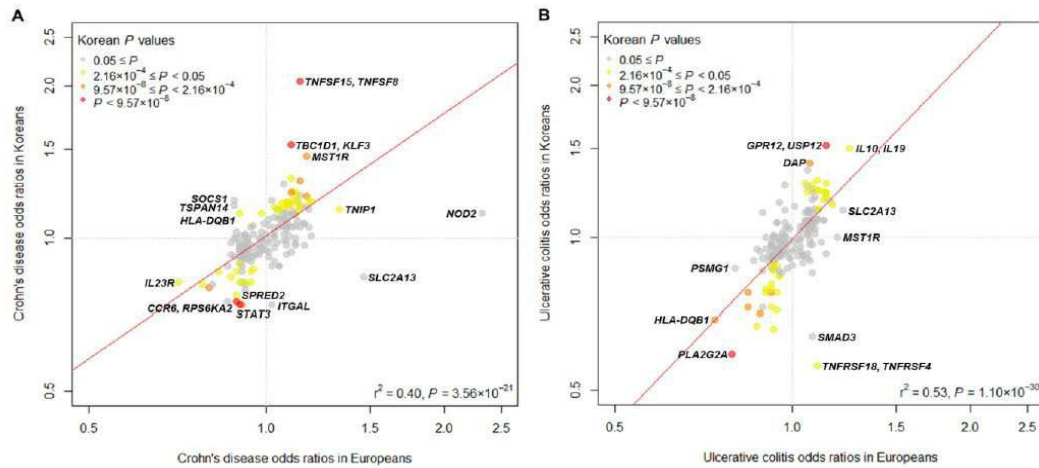
[0059] 이상으로 본 발명의 특정한 부분을 상세히 기술하였는 바, 당업계의 통상의 지식을 가진 자에게 있어서, 이러한 구체적 기술은 단지 바람직한 실시예일 뿐이며, 이에 의해 본 발명의 범위가 제한되는 것이 아닌 점은 명백할 것이다. 따라서, 본 발명의 실질적인 범위는 첨부된 청구항들과 그것들의 등가물에 의하여 정의된다고 할 것이다.

도면

도면1



도면2



서열 목록

- <110> University of Ulsan Foundation For Industry Cooperation
THE ASAN FOUNDATION
- <120> Composition of single nucleotide polymorphism markers for
predicting risk of inflammatory bowel disease in Koreans and
method for predicting risk of inflammatory bowel disease using
the same
- <130> ADP-2016-0316
- <160> 3
- <170> KopatentIn 2.0
- <210> 1
- <211> 201
- <212> DNA
- <213> Homo sapiens
- <400> 1

gaaaacggtc aactctccct tatccacaag gccaaaagag ggggggccgg ctgtgtgtgg 60

tgtggccacc ccgggcctag gctgggccgg gcctgtagaa rgtactcagg aactgcaggt 120

tttgctttc ccttccatga aactgacagg cattcgggct gacttctgcc tctcagtttc 180

gcggtaagag acccaagttt t 201

<210> 2

<211> 201

<212> DNA

<213> Homo sapiens

<400> 2

aaaggcaagc agcacgtggc agctaaggag agctttcaac agttttacag agctgagaag 60

acaaaagttg tagcaagagc ctatcaaggc aaacaagact ygaggatgga aaagctagga 120

cagaagagaa acagagaact aagacactat gcacgtgttt ttccttcaaa gtgctcaatg 180

attcttagtc aaagaaggat a 201

<210> 3

<211> 201

<212> DNA

<213> Homo sapiens

<400> 3

agtggcacag tagtgagctc gcagggccca ttcctatata gaaacacaga aaaataagca 60

aaagctgcc a gaattgactt cgtgggaact ctgggaaaca rtaagaaagt ttacagcaa 120

ccaagtgagt gctgaaccaa ggcaaaggta actgaaacgc aatgatagag aagaggggca 180

gggaagtgtt gggtagacaa a 201