

Gene-centric association mapping of chromosome 3p implicates MST1 in IBD pathogenesis

P Goyette¹, C Lefebvre¹, A Ng², SR Brant³, JH Cho⁴, RH Duerr⁵, MS Silverberg⁶, KD Taylor⁷, A Latiano⁸, G Aumais⁹, C Deslandres¹⁰, G Jobin⁹, V Annesse⁸, MJ Daly^{11,12}, RJ Xavier^{2,12} and JD Rioux^{1,11}

Association mapping and candidate gene studies within inflammatory bowel diseases (IBD) linkage regions, as well as genome-wide association studies in Crohn's disease (CD) have led to the discovery of multiple risk genes, but these explain only a fraction of the genetic susceptibility observed in IBD. We have thus been pursuing a region on chromosome 3p21–22 showing linkage to CD and ulcerative colitis (UC) using a gene-centric association mapping approach. We identified 12 functional candidate genes by searching for literature citations with relevant keywords and for gene expression patterns consistent with immune/intestinal function. We then performed an association study composed of a screening phase, where tagging single nucleotide polymorphisms (SNPs) were evaluated in 1,020 IBD patients, and an independent replication phase in 745 IBD patients. These analyses identified and replicated significant association with IBD for four SNPs within a 1.2 Mb linkage disequilibrium region. We then identified a non-synonymous coding variant (rs3197999, R689C) in the macrophage-stimulating 1 (MST1) gene (*P*-value 3.62×10^{-6}) that accounts for the association signal, and shows association with both CD and UC. MST1 encodes macrophage-stimulating protein (MSP), a protein regulating the innate immune responses to bacterial ligands. R689C is predicted to interfere with MSP binding to its receptor, suggesting a role for this gene in the pathogenesis of IBD.

INTRODUCTION

Crohn's disease (CD) and ulcerative colitis (UC) are two related but distinct common inflammatory diseases of the gastrointestinal tract, collectively known as the inflammatory bowel diseases (IBD). The epidemiological data on IBD suggest a strong genetic contribution to disease pathogenesis and genetic studies indicate that CD and UC have disease-specific as well as shared susceptibility genes.^{1–3} Genome-wide linkage mapping studies in IBD have identified and confirmed several genomic intervals conferring risk to either CD, UC, or both.^{4–11} Association mapping, as well as candidate gene association studies, within linkage regions have led to confirmed associations between CD and coding variants in the NOD2 gene on chromosome

16q12 and multiple associated alleles forming a risk haplotype known as IBD5 on chromosome 5q31.^{8,10,12–14} More recently, genome-wide association (GWA) studies of CD have identified several novel CD-associated risk loci, including IL-23R, ATG16L1, PTGER4, IRGM, and NKX2-3.^{15–19} Because of the unbiased nature of GWA studies, these have identified processes involved in the pathogenesis of CD, which had not been previously implicated. Given that all of the GWA studies to date have been performed specifically on CD, it is not surprising that, with the exception of IL-23R, most of the association findings have been CD-specific risk factors. A general conclusion that can be drawn from the genetic studies of IBD performed to date is that the confirmed associated alleles only modestly modify an

¹Department of Medicine, Université de Montréal and the Montreal Heart Institute, Research Center, Montréal, Québec, Canada. ²Gastrointestinal Unit and Center for Computational and Integrative Biology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA. ³Department of Medicine, Johns Hopkins University, Harvey M and Lyn P Meyerhoff Inflammatory Bowel Disease Center, Baltimore, Maryland, USA. ⁴Division of Gastroenterology, Department of Medicine, Yale University, New Haven, Connecticut, USA. ⁵Division of Gastroenterology, Department of Medicine, Hepatology and Nutrition, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. ⁶Mount Sinai Hospital IBD Centre, University of Toronto, Toronto, Ontario, Canada. ⁷Medical Genetics Institute and IBD Center, Cedars-Sinai Medical Center, Los Angeles, California, USA. ⁸Gastrointestinal & Endoscopy Units, CSS-RCCS Hospital, Viale Cappuccini, San Giovanni Rotondo, Italy. ⁹Department of Gastroenterology, Hôpital Maisonneuve-Rosemont, Montréal, Québec, Canada. ¹⁰Department of Gastroenterology, Hôpital Sainte-Justine, Montréal, Québec, Canada. ¹¹The Broad Institute of MIT and Harvard and Harvard, Cambridge, Massachusetts, USA. ¹²Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA. Correspondence: JD Rioux (rioux@inflammgen.org)

Received 9 October 2007; accepted 21 November 2007; published online 16 January 2008. doi:10.1038/mi.2007.15

Table 1 Summary of association results for 12 candidate genes on chromosome 3p

Locus	SNPs tested ^a	Associated SNP	IBD			CD	UC
			Screen <i>P</i> -value ^b	Replication <i>P</i> -value ^c	Combined <i>P</i> -value ^b	Combined <i>P</i> -value ^b	Combined <i>P</i> -value ^b
MLH1	5/8	—	—	—	—	—	—
MYD88	5/7	—	—	—	—	—	—
LTF	5/7	—	—	—	—	—	—
GPX1	4/5	rs9818758	0.0098	0.0021	0.00013	0.00065	0.0089
		rs11716445	0.0078	0.0036	0.00016	0.0054	0.0041
MST1	3/6	rs9858213	0.0046	3.59×10 ⁻⁵	1.48×10 ⁻⁶	6.27×10 ⁻⁵	0.0013
TRIP	5/10	rs9858428	0.00092	0.0018	1.02×10 ⁻⁵	0.00055	0.0011
GNAI2	8/9	rs2298952	0.0044	NR	—	—	—
CISH	7/8	—	—	—	—	—	—
TLR9	5/5	—	—	—	—	—	—
PRKCD	12/15	—	—	—	—	—	—
IL17RB	9/10	—	—	—	—	—	—
IL17RD	20/20	—	—	—	—	—	—

CD, Crohn's disease; IBD, inflammatory bowel diseases; NR, non-replicating; SNP, single nucleotide polymorphism; UC, ulcerative colitis.

“—” results not significant at *P*-value threshold 0.01.

^aSuccessful SNP assays/total SNPs selected to tag locus.

^bTwo-tailed *P*-values.

^cOne-tailed *P*-value.

individual's risk and that no single study or approach is sufficient to identify all risk factors.

We therefore have been pursuing a gene-centric association mapping approach of a confirmed common IBD locus (i.e., conferring risk to both CD and UC) on chromosome 3p21–22. This region was first identified in a genome-wide linkage scan of British families⁴ and then replicated with the findings from multiple genome-wide linkage scans.^{8,11,20} Subsequent association mapping studies using microsatellite markers provided additional evidence of association and some localization information.^{20,21} In an attempt to further narrow down the genomic interval for our search, we have targeted our association study to a 30 Mb overlap region between the reported linked regions. This region contains over 400 annotated loci, several of which may represent interesting candidates for IBD. We selected 12 candidate genes with strong prior probability by searching for literature citations with relevant keywords (inflammation, immunity, intestinal disease and so on) and by searching publicly available data sets for gene expression patterns consistent with genes having a role in immune and/or intestinal tissues. We then performed a two-stage association study composed of a screening phase in a set of 1,020 patients with IBD of European descent and (2) a follow-up independent replication phase in 745 patients with IBD from the North American NIDDK IBD Genetics Consortium. We report the identification of four single nucleotide polymorphisms (SNPs) showing significant evidence of association with IBD and replication in an independent cohort, all located within a single 1.2 Mb genetic interval. We then narrowed the associated interval to a 336 kb region and identified a non-synonymous coding variant (R689C) in the macrophage-stimulating 1 gene (MST1), which is a common

risk factor for IBD. Macrophage-stimulating protein (MSP), the protein encoded by MST1, and its receptor MST1R have been reported to be involved in macrophage chemotaxis and activation,^{22,23} as well as regulation of inflammatory responses (such as NO and IL12p40 production) following pro-inflammatory signals.^{24–27} Taken together, these data suggest a role for MST1 in IBD pathogenesis.

RESULTS

Gene-centric association study

Linkage evidence from multiple genome-wide and targeted studies indicates the presence of an IBD susceptibility locus (conferring risk to both CD and UC) in the chromosomal region 3p21–22. We selected a region for our association mapping project, based on the maximal overlap between previously reported linkage intervals. Since this approach defined a 30 Mb region containing over 400 annotated loci, we set out to identify potential functional candidates for inclusion in this project, based on published gene functions and expression patterns (see Methods). Genes were selected based on their reported roles in inflammation, immunity, bowel dysfunction, or intestinal inflammation in experimental models, as well as on expression patterns compatible with a role in immune or intestinal tissues. With this strategy, 12 candidate genes were selected for the screening phase (Table 1).

Eighty-nine SNPs tagging the common genetic variation surrounding these 12 candidate genes were typed in 215 father–mother-affected offspring trios and in 805 patients with IBD and 412 healthy controls, all of European ancestry. Allelic tests of association of the individual SNPs identified five SNPs with suggestive evidence of association ($P < 0.01$) in the combined screening cohorts (Table 1). As illustrated in Figure 1, four of

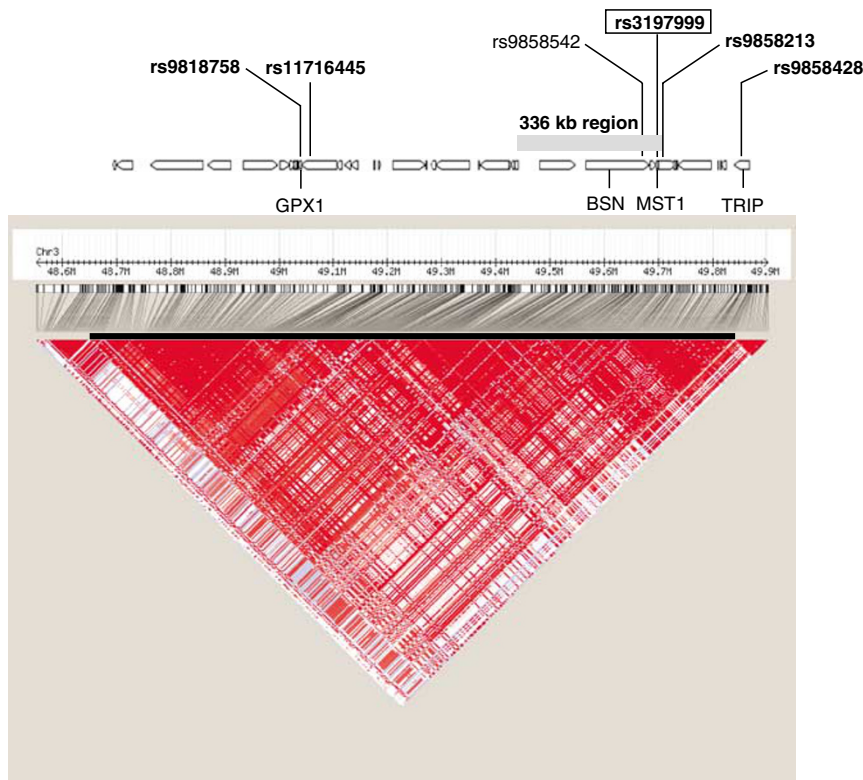


Figure 1 Association signal and linkage disequilibrium patterns across the 1.2 Mb region of LD. A candidate gene approach to association mapping of the linked region on chromosome 3p21–22 suggests a putative association signal located in a 1.2 Mb region of LD containing over 30 genes. The LD plot of pairwise D' values, derived from HapMap project data, indicates strong LD across the region (lower panel). Loci in the region tagged by the original study design are indicated on the gene track. The associated region of 336 kb, as defined by the correlation neighborhood at $r^2 > 0.8$, along with the replicated SNPs in the region are mapped against the gene track. Four of the five SNPs identified in the screening stage of the study (rs9818758, rs11716445, rs9858213, and rs9858428) are shown in bold; the fifth SNP (rs2298952) is located outside the region shown. The MST1 coding SNP (rs3197999) is boxed and the reported WTCCC SNP (rs9858542, in BSN) is in normal type. LD, linkage disequilibrium; MST1, macrophage-stimulating 1 gene; SNP, single nucleotide polymorphism.

these five SNPs were located within a 1.2 Mb region (position 48700711–49869034) of high linkage disequilibrium (LD). The fifth SNP is located 392 kb away from the interval described above.

To confirm the associations observed in this screening phase, we typed the five SNPs in an independent set of 350 patients with IBD and 342 controls as well as a cohort of 395 IBD families. Genotyping assays were successfully designed for all five putatively associated SNPs. All four of the SNPs located within the 1.2 Mb region had evidence of replication ($P < 0.01$) in the combined replication cohorts whereas the fifth SNP outside this region did not replicate (**Table 1**). Specifically, the strongest association signal observed was to the MST1 tagging SNP rs9858213 (replication $P = 3.59 \times 10^{-5}$; combined $P = 1.48 \times 10^{-6}$; odds ratio (OR) 1.33 (95% confidence interval (CI) 1.18–1.49). Although this SNP was selected to tag some of the regional variation around MST1, it is located approximately 5.5 kb centromeric of MST1. The other three replicated SNPs had combined P -values ranging from 0.0016 to 1.02×10^{-5} . Limiting the analysis to the CD and UC phenotypes did not improve the evidence of association for any of these SNPs, indicating that the identified association signals are for the general IBD phenotype.

Identification of association to a missense mutation in MST1

Given that there are multiple SNPs that have significant evidence of association in the screen and replication within a region of 1.2 Mb with high LD, it was important to determine whether there were single or multiple independent association signals in the region. We therefore used the combined screening and replication cohorts to have maximal statistical power and performed conditional logistic regression analyses of the associated SNPs. Conditioning the data upon rs9858213, the primary association signal tagging MST1, did not identify additional secondary signals in the region (all conditional $P > 0.01$), indicating that the multiple associated SNPs are likely detecting the same correlated primary signal within this region.

In an attempt to delineate the minimal genomic interval defined by our associated signal, and thus potential genes implicated by this association, we evaluated the correlation neighborhood around rs9858213, as defined by all correlated SNPs at a correlation coefficient (r^2) value of 0.8 or greater, in the HapMap data set. More specifically, inclusion of all highly correlated SNPs ($r^2 > 0.8$) to rs9858213 defined an associated genomic interval of 336 kb (**Figure 1**). The same 336 kb region was also identified at a lower r^2 threshold of 0.5, further strengthening the likelihood of an IBD susceptibility gene within this interval. This

<i>H. sapiens</i> /1-869	THNCWVLEGI I I PNRVCARS	R	WPAVFTRVSVFVDWIHKVMRLG
<i>M. mulata</i> /1-869	THNCWVLEGI I I PNRVCARS	R	WPAVFTRVSVFVDWIHKVMRLA
<i>C. familiaris</i> /1869	THDCWVLEGI I I PNRVCARP	R	WPAIFMRVSVFTDWIHKVMQLG
<i>B. taurus</i> /1-869	THDCWVLQGI I I PNRVCARP	R	WPAVFMRVSVFVDWIHKVMRLG
<i>R. norvegicus</i> /1-869	THDCWVLQGLI I PNRVCARP	R	WPAIFTRVSVFVDWINKVVQLE
<i>M. musculus</i> /1-869	THDCWVLQGLI I PNRVCARP	R	WPAIFTRVSVFVDWINKVMQLE
<i>G. gallus</i> /1-869	TADCWVLEGVITPSRVCART	D	QPALFIRVSLYVDWIHKVMRMV
<i>X. laevis</i> /1-869	THDCLVLEGVIVPARGCGKK	N	QPAIFTRVSVYVDWINKVMKMV
<i>D. rerio</i> /1-869	NSDCWVLEGV I I PMRRCGHA	G	QPNIFIRVSVYVDWIKKVMEMT
<i>S. scrofa</i> /1-869	EKDKYILQGVTSWGLGCALP	N	KPGVYVRVSRFVTWIEEIMRRN
		*	

689

Figure 2 Sequence conservation of the arginine residue at position 689 of MSP across different species. Amino-acid sequence is shown as a single-letter code (uppercase). The amino acid modified by the associated coding variant at position 689 (R689C) is in bold and demonstrates conservation of the arginine across mammalian species. MSP, macrophage-stimulating protein.

336 kb associated interval encompasses 10 genes, two of which (GPX1 and MST1) were selected in our original study design. The association signal for rs9858213 is also highly correlated ($r^2 = 0.95$) with a signal recently reported by the Wellcome Trust Case Control Consortium (WTCCC) in the same region, providing independent replication evidence of our observation.^{19,28} Specifically, the authors report association with a synonymous coding variant (rs9858542, OR 1.17; 95% CI 1.04–1.31) of the BSN locus, within a 1 Mb genomic region of LD, but do not provide any additional fine mapping to narrow down the associated interval. An evaluation of this SNP in our patient cohorts identifies an association signal (combined $P = 6.66 \times 10^{-6}$; OR 1.29; 95% CI 1.15–1.49) that is almost equivalent to the rs9858213 signal described above. To have greater statistical power to evaluate the contribution of this region to IBD susceptibility, we were interested in combining our association results with those from the WTCCC for rs9858542. Since there is extensive overlap between our replication cohort and the samples used in the NIDDK GWA scan, these cannot be considered as independent cohorts, and therefore we only included our current results and the WTCCC data set in the combined analysis. Combining our data with the published WTCCC data unambiguously establishes this region as a risk factor of genome-wide significance for IBD with an OR of 1.24 (95% CI 1.17–1.32; $P = 8.45 \times 10^{-13}$).

Having defined a region with unambiguous association to IBD, we then attempted to identify the primary causal allele underlying this association signal. For this, we investigated all SNPs highly correlated ($r^2 > 0.9$) to both, our association signal and the WTCCC marker. Using this approach, we identified a single non-synonymous coding variant in MST1 (rs3197999) with correlation coefficients of 0.90 and 0.95 to rs9858213 and rs9858542, respectively. This variant causes an arginine-to-cysteine amino-acid substitution at position 689 (R689C) located in exon 18 of the gene. Typing and association testing of this variant in the combined screening and replication cohorts showed significant evidence of association (combined $P = 7.25 \times 10^{-6}$). This missense coding variant (OR 1.30; 95% CI 1.16–1.46) and the rs9858213 SNP (OR 1.33; 95% CI 1.18–1.49) had equivalent strength of association. Reciprocal conditional analysis of rs3197999 confirms it as part of the same association signal as rs9858213 (data not shown) and therefore implicates MST1 as a potential risk factor for IBD.

The associated R689C variant and MST1 function

The MST1 locus encodes the MSP, a plasminogen-related growth factor containing four Kringle domains and a C-terminal serine protease-like homology domain. This domain of MSP has been suggested to lack enzymatic activity due to amino-acid substitutions in the active domain; however, it has been demonstrated to contain the high-affinity binding site for its receptor MST1R.^{29,30} The MSP protein was originally identified as a serum protein produced by the liver that could elicit macrophage chemotaxis and activation.^{22,23} MSP is secreted as a biologically inactive single-chain precursor (pro-MSP) by the liver, and can be activated through proteolytic cleavage by serine proteases in areas of blood coagulation and active inflammation.³¹ Although the specific activating protease is unknown, a role for macrophage cell surface proteases has been suggested from *in vitro* studies with cultured peritoneal macrophages and isolated macrophage membranes.³² MSP is the only known ligand for the MST1R, also known as RON receptor, a disulphide-linked heterodimer composed of an extracellular α -chain and a transmembrane β -chain with intrinsic kinase activity, expressed predominantly on tissue-resident (differentiated) macrophages and epithelial cells.³³ Signaling through the MSP/MST1R has been shown to inhibit lipopolysaccharide or cytokine-mediated production of inflammatory mediators by macrophages,^{24–27} through the inhibition of NF- κ B signaling, while mice deficient in MST1R show increased susceptibility to septic shock following lipopolysaccharide challenge.^{34,35} We have also observed, from publicly available microarray data sets, that MST1 gene expression is elevated during intestinal enterocyte differentiation in Caco-2 cells and in the biopsies from the noninflamed ileum of CD patients (data not shown).

An alignment of MSP protein sequences from different species shows that this non-conservative amino-acid substitution affects a residue that is conserved in mammals (**Figure 2**). By examining the MSP crystal structure data (PDB ID: 2ASU),³⁶ we note that R689 is located on the surface of the MSP β -chain and within a positively charged patch (L13 loop) that consists of a cluster of arginine residues including R683, which has been shown to be essential for MST1R binding³⁷ (**Figure 3**). The L13 loop borders the S1 specificity pocket (involving Y661, Q522, and Q568) and together form the putative high-affinity binding surface for MST1R, resembling the structurally determined

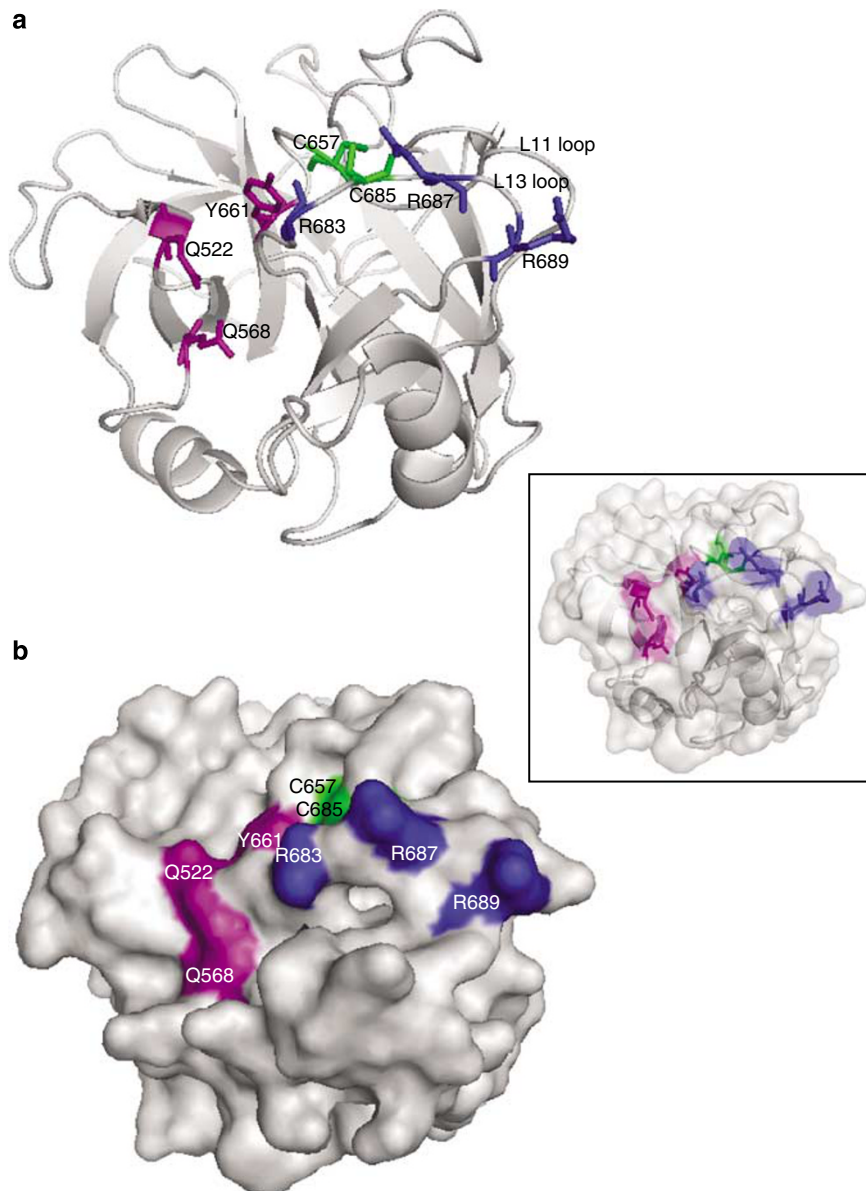


Figure 3 Three-dimensional representation of MSP. **(a)** Ribbon diagram and **(b)** molecular surface representations of the MSP β -chain. A merged superimposed image of the two representations is shown in the inset. The amino-acid residues discussed in the text are highlighted in strong colors and labeled. The residues (Y661, Q522, and Q568) corresponding to the “catalytic triad” of *bona fide* serine proteinases are colored red. The cluster of arginine residues (R689, R687, and R683) located within a positively charged patch on the L13 loop that forms a continuous region with the S1 specificity pocket, generating the putative primary (high-affinity) binding surface for its receptor, is colored blue. The two cysteine residues (C657 on loop L11 and C685 on loop L13) connected by a disulfide bond that lies in close proximity to the cluster of arginine residues, are colored green. The side chains of the highlighted residues are displayed in the ribbon diagram as sticks. The structural representations were constructed with PyMOL⁵⁸ and the data were obtained from the Protein Data Bank (pdb: 2ASU) containing three-dimensional structure coordinates from X-ray crystallographic analysis of 225 residues of the MSP β -chain, four residues of the α -chain, and 154 water molecules. MSP, macrophage-stimulating protein.

interaction between hepatocyte growth factor (scatter factor), which is structurally related to MSP, and its receptor, MET.^{38,39} Furthermore, the proximity of R689C to a conserved cysteine residue (C685) that establishes a critical disulfide bond with C657 may introduce aberrant bond formations and disrupt the structural integrity of the adjacent receptor-binding surface. Hence, the strategic location of the R689C variant will likely have a substantial impact on receptor binding or activation, the functional implication of which is an active focus of our ongoing studies.

DISCUSSION

Over the last 20 years, genome-wide linkage scans of both CD and UC have identified several chromosomal regions conferring susceptibility to these diseases.^{4–10} Two main strategies have been used to identify susceptibility loci within linkage intervals: (1) association fine mapping of the entire linked interval, or (2) positional candidate gene (gene-centric) association studies. Both of these approaches were independently used to identify successfully NOD2 variants within the IBD1 linkage peak in the 16q12 region.^{8,10,12–14} Recent GWA studies have

identified and replicated multiple novel susceptibility genes for CD, and promise to identify many more genes in the near future.^{15–19} Interestingly, to date, few of these novel susceptibility loci can explain the linkage peaks previously defined from the genome-wide linkage analyses, suggesting that complementary approaches will be necessary to identify the full complement of susceptibility genes for CD and UC.

The current study has adopted a positional candidate gene association strategy similar to the one used in the identification of NOD2, for examining the confirmed chromosome 3p21–22 IBD-specific linkage peak, although testing 12 different genes within the linked region. This approach takes advantage of the power of association testing while limiting the extent of statistical testing by focusing on a genomic region that has been implicated by multiple linkage studies and by targeting a limited set of genes with higher prior probability of being true IBD loci.

The results of the current study identify and replicate significant evidence of association to four SNPs located in a 1.2 Mb region of high LD. Conditional logistic regression analysis of these associated SNPs defines a single association signal in the region. An analysis of the minimal associated interval for this signal, based on the evaluation of the correlation neighborhood at an $r^2 > 0.8$ around the top associated signal (rs9858213), defines a narrower region of 336 kb containing only 10 genes. In a recently reported GWA study from the WTCCC, a single synonymous SNP (rs9858542) located within the BSN gene was significantly associated with CD. Testing of this SNP in our own screening and replication cohorts identifies an association signal that is equivalent in strength and that is highly correlated to rs9858213, with combined results providing overwhelming evidence that this region contains a gene for IBD (P -value $< 10^{-12}$). A second recent GWA study⁴⁰ also reports a significant association signal within the same region for marker rs11718165, an intronic SNP also located in BSN. The authors report several associated SNPs across a 1.2 Mb region but do not narrow down the association further. Our current study, however, further identifies a single non-synonymous arginine-to-cysteine coding variant at base pair position 689 within exon 18 of the MST1 gene that can account for the association to this region. While the association results and the identification of a coding variant suggest a potential role for MST1 in susceptibility to both CD and UC, we cannot, however, formally exclude other variants in this region that have an impact on disease susceptibility.

The R689C variant is located within an important region that forms part of the putative high-affinity receptor-binding surface with the S1 specificity pocket located in the serine protease homology domain of MSP. Several aspects of MSP activation and signaling make it an attractive functional candidate for IBD. First, proteolytic activation of MSP has been proposed to occur in areas of active inflammation through cell surface proteases located on macrophages.^{31,32} Second, while MSP was originally identified as a protein that could elicit macrophage chemotaxis and activation, signaling through the MSP/MST1R has recently been shown to inhibit lipopolysaccharide or cytokine-mediated production of inflammatory mediators (such as NO, COX2, PGE2, and IL-12p40) by macrophages, through the inhibition

of NF- κ B signaling.^{24–27} Interestingly, increased NF- κ B signaling activity, as well as increased levels of NO and expression of IL-12p40, have all been observed in the inflamed mucosa of patients with IBD.^{41–43} Finally, mice deficient in MST1R show increased susceptibility to septic shock following lipopolysaccharide challenge, highlighting the importance of the MSP/MST1R signaling pathway in controlling inflammation triggered from innate immune responses.^{34,35} This dampening effect of the MSP/MST1R pathway on macrophage activation in response to microbial ligands may be of particular interest in the intestinal epithelium where the innate immune system is under constant challenge, due to close contact with a variety of microorganisms. The loss of macrophage hyporesponsiveness or anergy has been proposed as a potential mechanism for triggering local inflammation and ultimately aspects of adaptive immunity implicated in IBD.^{44–46} The capacity of MSP to control IL-12p40 gene expression, a subunit of the heterodimeric IL-23, is also interesting in the context of IBD given the recent observations showing that IL23 regulation plays a central role in the control of intestinal inflammation^{47,48} and that neutralizing anti-IL-12p40 antibodies prevent the development of disease in murine models of IBD.^{49,50} Finally, a recent report from the NIDDK IBD Genetic Consortium has identified a variant in a subunit of the IL23 receptor (IL-23R) that is associated with reduced risk of IBD, highlighting the importance of the IL23 signaling pathway in this disease.

Our identification of a missense coding variant in MST1 within the narrow interval defined by our association signal and the reported role of MSP in the regulation of innate immune responses implicate the role of MST1 gene, and its product MSP, in the pathogenesis of IBD. Interestingly, MST1 is one of a growing number of loci involved in the regulation of innate and adaptive immune responses, including NOD2, IL-23R, IRGM, and ATG16L1, which have been identified to have an impact on IBD. Further studies evaluating the functional significance of this coding variant on innate immune responses, on the MSP/MST1R signaling pathway in IBD, and on its interaction with IL-23R signaling will be essential for better understanding the mechanisms leading to IBD. It is interesting that both the MST1 and IL-23R variants are among the few susceptibility genes that have been shown to impact on both CD and UC susceptibility. Their functional proximity in the regulation of IL23 signaling may point to common pathological mechanisms of intestinal inflammation postulated by epidemiological studies of IBD.

METHODS

Subject ascertainment and diagnostic classification. The screening cohort consists of a combined cohort of Canadian and Italian IBD trios and tetrads (249 father–mother-affected offspring trios), collected from multiple sites in the province of Quebec and at the S. Giovanni Rotondo “CSS” (SGRC) Hospital in Italy, and of an Italian IBD case–control cohort (816 cases, 421 controls) also collected at the SGRC hospital. This cohort has previously been used and characterized in several association reports from our group,^{51–53} and additional cohort information can be found in **Supplementary Table S1** online. The diagnosis of IBD required (a) one or more symptoms of diarrhea, rectal bleeding, abdominal pain, fever, or complicated perianal disease, (b) occurrence of symptoms on two or more occasions separated by at least 8 weeks or ongoing symptoms of at least 6 weeks duration, and (c) objective

evidence of inflammation from radiologic, endoscopic, and histologic evaluation. All affected subjects fulfill clinical criteria for IBD^{54,55} and gave consent to use protocols approved by the institutional review boards of each institution contributing samples.

The replication cohort consisted of the combination of an IBD case-control cohort (353 cases and 344 controls) and of an IBD trios cohort (401 trios) collected from the North American NIDDK IBD Genetics Consortium (IBDGC) as described previously.^{15,16} Additional information on this cohort can be found in **Supplementary Tables S1** and **S2** online. Cases and geographically matched controls, as well as nuclear families, were ascertained through the Cedars-Sinai Medical Center, Johns Hopkins University, University of Chicago, University of Montreal, University of Pittsburgh, and the University of Toronto Genetics Research Centers (GRCs). In all cases, informed consent was obtained using protocols approved by the local institutional review board in all participating institutions.

Candidate gene selection. The genomic region included in this analysis was selected based on an evaluation of the maximal overlap between previously reported linkage signals for IBD on chromosome 3p.^{4,8,11,20} Using this approach, a 30 Mb region containing over 400 annotated loci was targeted (position 32185205–64772789).

Potential functional candidates were identified by searching publicly available online resources, such as PubMed, OMIM, Unigene and SymAtlas from the Genomic Institute of the Novartis Research Foundation, with a set of predefined search criteria. Specifically, the PubMed and OMIM databases were queried for genes having a link (1) to inflammation (over/underexpression in inflamed tissues, and pro-/anti-inflammatory role), (2) to immunity (regulation of immune responses or immune cell proliferation), (3) to the development of a bowel disease (malformation, bowel cancer, and intestinal dysfunction), and (4) to experimental models of intestinal inflammation (genetic mouse model or activation/inhibition of experimental models). The Unigene and SymAtlas databases were also queried for genes showing expression patterns compatible with a role in immune and/or intestinal tissues (tissue-specific expression patterns or marked overexpression in tissues of immune or intestinal origin). Genes were given a score of one for each positive attribute and then ranked based on the sum of these scores. The top 12 ranking genes were selected for association analysis (**Table 1**).

SNP selection. Screening phase. SNPs were selected from the HapMap database (www.hapmap.org) to capture the common genetic variation (allele frequency >5%) in and around each of the 12 candidate genes (the coding region plus 10–20 kb on either side). Specifically, using the Tagger algorithm,⁵⁶ we selected a set of 110 SNPs that tag all of the common genetic variation in these gene regions with a minimum pair-wise correlation coefficient (r^2) of 0.8.

Replication phase. The five SNPs showing evidence of association (P -value < 0.01 in the screening phase) were selected for replication; all five were successfully designed and tested on the replication cohort.

Genotyping and quality control analyses. Genotyping for the screening and replication phases of the project were performed in-house (The Laboratory for Genetics and Genomic Medicine of Inflammation, www.inflamngen.org) using primer extension chemistry and mass spectrometric analysis (iPlex assay, Sequenom, San Diego, CA) on the Sequenom MassArray.

Quality filtering of both samples and SNPs was performed prior to association testing to insure robust data quality for the screening phase. First, all SNPs with less than 75% genotyping call rates were excluded from the analysis. Then, an evaluation of empirical data quality distributions was performed and families and SNPs with an excess of Mendel errors (in the familial cohort) were removed. Finally, only DNAs and SNPs showing a minimum genotyping call rate of greater than 80% were included in the final analyses. Quality filtering brought the overall genotyping rate to 96.7% (ranging from 82.5 to 99.9%). Following these quality control steps, the data set consisted of 89 SNPs tested on 805 cases and 412 controls and 215 trios (for a total of 1020 IBD cases, see **Supplementary Table S1** online for further cohort information). Five SNPs showing

evidence of association in the combined screening cohort ($P < 0.01$) were then tested for replication in the combined NIDDK IBD case-control and trio replication cohort. Quality control thresholds used were the same as for the screening cohort, except for the final DNA and SNP quality threshold, which were set at 90%. This quality filtering brought the overall genotyping to 99.6% (ranging from 98.4 to 100%). Following quality control, all five putatively associated SNPs were evaluated in 350 cases and 342 controls and on 395 trios (for a total of 745 IBD cases, see **Supplementary Table S1** online for further cohort information). An evaluation of allelic frequencies for the different markers did not identify any significant differences between the different cohorts tested (data not shown).

Association testing. Association testing of single SNPs was performed by the transmission disequilibrium test, for trios, and by standard χ^2 test carried out on a 2×2 contingency table, for case-control cohorts, as implemented in the Haploview analysis software.⁵⁷ Combination of the results derived from the case-control and trios-tetrads in the screening and replication cohorts, as well as combination of the results from the screen and replication, was achieved by the calculation of a combined z -score as described previously.⁵¹ Association results (**Table 1**) are reported as two-tailed nominal significance P -values for the screening, fine mapping and combined cohorts, and one-tailed P -values for replication. The significance threshold for follow-up replication was set at a P -value < 0.01. ORs and 95% CI are reported for the combined data set. Conditional analyses between the different SNPs showing association were performed in the combined cohorts by logistic regression using the WHAP analysis software. The threshold for significant independent association signal was also set at a P -value < 0.01.

SUPPLEMENTARY MATERIAL is linked to the online version of the paper at <http://www.nature.com/mi>

ACKNOWLEDGMENTS

The North American IBD Genetics Consortium is funded by multiple grants from the National Institutes of Diabetes, Digestive and Kidney diseases (NIDDK): DK62431 (S.R.B.), DK62422/DK62429 (J.H.C.), DK62420 (R.H.D.), DK62432 (J.D.R.), DK62423 (M.S.S.), and DK62413 (K.D.T.). The current work was funded by a NIDDK Grant (DK064869) to J.D.R. and an NIAID Grant (AI062773) to R.J.X. A.N. is the recipient of a fellowship from the Crohn's and Colitis Foundation of America. We thank Catherine Labbe for her contribution to the genotyping performed as part of this project.

DISCLOSURE

The authors declared no conflict of interest.

© 2008 Society for Mucosal Immunology

REFERENCES

- Binder, V. Genetic epidemiology in inflammatory bowel disease. *Dig. Dis.* **16**, 351–355 (1998).
- Orholm, M., Munkholm, P., Langholz, E., Nielsen, O.H., Sorensen, T.I. & Binder, V. Familial occurrence of inflammatory bowel disease. *N. Engl. J. Med.* **324**, 84–88 (1991).
- Probert, C.S., Jayanthi, V., Hughes, A.O., Thompson, J.R., Wicks, A.C. & Mayberry, J.F. Prevalence and family risk of ulcerative colitis and Crohn's disease: an epidemiological study among Europeans and south Asians in Leicestershire. *Gut* **34**, 1547–1551 (1993).
- Satsangi, J. *et al.* Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nat. Genet.* **14**, 199–202 (1996).
- Hampe, J. *et al.* A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. *Am. J. Hum. Genet.* **64**, 808–816 (1999).
- Cho, J.H. *et al.* Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q, and 4q: evidence for epistasis between 1p and IBD1. *Proc. Natl. Acad. Sci. USA* **95**, 7502–7507 (1998).
- Ma, Y. *et al.* A genome-wide search identifies potential new susceptibility loci for Crohn's disease. *Inflamm. Bowel Dis.* **5**, 271–278 (1999).

8. Rioux, J.D. *et al.* Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am. J. Hum. Genet.* **66**, 1863–1870 (2000).
9. Duerr, R.H., Barmada, M.M., Zhang, L., Pfutzer, R. & Weeks, D.E. High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11–12. *Am. J. Hum. Genet.* **66**, 1857–1862 (2000).
10. Hugot, J.P. *et al.* Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* **379**, 821–823 (1996).
11. Paavola, P. *et al.* Genetic analysis in Finnish families with inflammatory bowel disease supports linkage to chromosome 3p21. *Eur. J. Hum. Genet.* **9**, 328–334 (2001).
12. Hugot, J.P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603 (2001).
13. Ogura, Y. *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**, 603–606 (2001).
14. Rioux, J.D. *et al.* Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat. Genet.* **29**, 223–228 (2001).
15. Duerr, R.H. *et al.* A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* **314**, 1461–1463 (2006).
16. Rioux, J.D. *et al.* Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat. Genet.* **39**, 596–604 (2007).
17. Libioulle, C. *et al.* Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet.* **3**, e58 (2007).
18. Hampe, J. *et al.* A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat. Genet.* **39**, 207–211 (2007).
19. Parkes, M. *et al.* Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat. Genet.* **39**, 830–832 (2007).
20. Hampe, J. *et al.* Fine mapping of the chromosome 3p susceptibility locus in inflammatory bowel disease. *Gut* **48**, 191–197 (2001).
21. Duerr, R.H. *et al.* Evidence for an inflammatory bowel disease locus on chromosome 3p26: linkage, transmission/disequilibrium and partitioning of linkage. *Hum. Mol. Genet.* **11**, 2599–2606 (2002).
22. Leonard, E.J. & Skeel, A.H. Isolation of macrophage stimulating protein (MSP) from human serum. *Exp. Cell Res.* **114**, 117–126 (1978).
23. Leonard, E.J. & Skeel, A. A serum protein that stimulates macrophage movement, chemotaxis and spreading. *Exp. Cell Res.* **102**, 434–438 (1976).
24. Morrison, A.C., Wilson, C.B., Ray, M. & Correll, P.H. Macrophage-stimulating protein, the ligand for the stem cell-derived tyrosine kinase/ RON receptor tyrosine kinase, inhibits IL-12 production by primary peritoneal macrophages stimulated with IFN- γ and lipopolysaccharide. *J. Immunol.* **172**, 1825–1832 (2004).
25. Chen, Y.Q., Fisher, J.H. & Wang, M.H. Activation of the RON receptor tyrosine kinase inhibits inducible nitric oxide synthase (iNOS) expression by murine peritoneal exudate macrophages: phosphatidylinositol-3 kinase is required for RON-mediated inhibition of iNOS expression. *J. Immunol.* **161**, 4950–4959 (1998).
26. Zhou, Y.Q., Chen, Y.Q., Fisher, J.H. & Wang, M.H. Activation of the RON receptor tyrosine kinase by macrophage-stimulating protein inhibits inducible cyclooxygenase-2 expression in murine macrophages. *J. Biol. Chem.* **277**, 38104–38110 (2002).
27. Liu, Q.P., Fruit, K., Ward, J. & Correll, P.H. Negative regulation of macrophage activation in response to IFN- γ and lipopolysaccharide by the STK/RON receptor tyrosine kinase. *J. Immunol.* **163**, 6606–6613 (1999).
28. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
29. Han, S., Stuart, L.A. & Degen, S.J. Characterization of the DNF15S2 locus on human chromosome 3: identification of a gene coding for four kringle domains with homology to hepatocyte growth factor. *Biochemistry* **30**, 9768–9780 (1991).
30. Yoshimura, T., Yuhki, N., Wang, M.H., Skeel, A. & Leonard, E.J. Cloning, sequencing, and expression of human macrophage stimulating protein (MSP, MST1) confirms MSP as a member of the family of kringle proteins and locates the MSP gene on chromosome 3. *J. Biol. Chem.* **268**, 15461–15468 (1993).
31. Wang, M.H., Gonias, S.L., Skeel, A., Wolf, B.B., Yoshimura, T. & Leonard, E.J. Proteolytic activation of single-chain precursor macrophage-stimulating protein by nerve growth factor- γ and epidermal growth factor-binding protein, members of the kallikrein family. *J. Biol. Chem.* **269**, 13806–13810 (1994).
32. Wang, M.H., Skeel, A. & Leonard, E.J. Proteolytic cleavage and activation of pro-macrophage-stimulating protein by resident peritoneal macrophage membrane proteases. *J. Clin. Invest.* **97**, 720–727 (1996).
33. Iwama, A. *et al.* Terminal differentiation of murine resident peritoneal macrophages is characterized by expression of the STK protein tyrosine kinase, a receptor for macrophage-stimulating protein. *Blood* **86**, 3394–3403 (1995).
34. Waltz, S.E. *et al.* Ron-mediated cytoplasmic signaling is dispensable for viability but is required to limit inflammatory responses. *J. Clin. Invest.* **108**, 567–576 (2001).
35. Correll, P.H., Iwama, A., Tondat, S., Mayrhofer, G., Suda, T. & Bernstein, A. Deregulated inflammatory response in mice lacking the STK/RON receptor tyrosine kinase. *Genes Funct.* **1**, 69–83 (1997).
36. Carafoli, F., Chirgadze, D.Y., Blundell, T.L. & Gherardi, E. Crystal structure of the beta-chain of human hepatocyte growth factor-like/macrophage stimulating protein. *FEBS J.* **272**, 5799–5807 (2005).
37. Danilkovitch, A., Miller, M. & Leonard, E.J. Interaction of macrophage-stimulating protein with its receptor. Residues critical for beta chain binding and evidence for independent alpha chain binding. *J. Biol. Chem.* **274**, 29937–29943 (1999).
38. Kirchhofer, D. *et al.* Structural and functional basis of the serine protease-like hepatocyte growth factor beta-chain in Met binding and signaling. *J. Biol. Chem.* **279**, 39915–39924 (2004).
39. Stamos, J., Lazarus, R.A., Yao, X., Kirchhofer, D. & Wiesmann, C. Crystal structure of the HGF beta-chain in complex with the SemA domain of the Met receptor. *EMBO J.* **23**, 2325–2335 (2004).
40. Raelson, J.V. *et al.* Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci. *Proc. Natl. Acad. Sci. USA* **104**, 14747–14752 (2007).
41. Schreiber, S., Nikolaus, S. & Hampe, J. Activation of nuclear factor kappa B inflammatory bowel disease. *Gut* **42**, 477–484 (1998).
42. Kimura, H. *et al.* Increased nitric oxide production and inducible nitric oxide synthase activity in colonic mucosa of patients with active ulcerative colitis and Crohn's disease. *Dig. Dis. Sci.* **42**, 1047–1054 (1997).
43. Omata, F., Birkenbach, M., Matsuzaki, S., Christ, A.D. & Blumberg, R.S. The expression of IL-12 p40 and its homologue, Epstein-Barr virus-induced gene 3, in inflammatory bowel disease. *Inflamm. Bowel Dis.* **7**, 215–220 (2001).
44. Smith, P.D., Ochsenbauer-Jambor, C. & Smythies, L.E. Intestinal macrophages: unique effector cells of the innate immune system. *Immunol. Rev.* **206**, 149–159 (2005).
45. Smythies, L.E. *et al.* Human intestinal macrophages display profound inflammatory anergy despite avid phagocytic and bacteriocidal activity. *J. Clin. Invest.* **115**, 66–75 (2005).
46. Takeda, K. *et al.* Enhanced Th1 activity and development of chronic enterocolitis in mice devoid of Stat3 in macrophages and neutrophils. *Immunity* **10**, 39–49 (1999).
47. Fuss, I.J. *et al.* Both IL-12p70 and IL-23 are synthesized during active Crohn's disease and are down-regulated by treatment with anti-IL-12 p40 monoclonal antibody. *Inflamm. Bowel Dis.* **12**, 9–15 (2006).
48. Becker, C. *et al.* Constitutive p40 promoter activation and IL-23 production in the terminal ileum mediated by dendritic cells. *J. Clin. Invest.* **112**, 693–706 (2003).
49. Hue, S. *et al.* Interleukin-23 drives innate and T cell-mediated intestinal inflammation. *J. Exp. Med.* **203**, 2473–2483 (2006).
50. Kullberg, M.C. *et al.* IL-23 plays a key role in *Helicobacter hepaticus*-induced T cell-dependent colitis. *J. Exp. Med.* **203**, 2485–2494 (2006).
51. Tello-Ruiz, M.K. *et al.* Haplotype-based association analysis of 56 functional candidate genes in the IBD6 locus on chromosome 19. *Eur. J. Hum. Genet.* **14**, 780–790 (2006).
52. van Bodegraven, A.A. *et al.* Genetic variation in myosin IXB is associated with ulcerative colitis. *Gastroenterology* **131**, 1768–1774 (2006).
53. De Jager, P.L. *et al.* The role of the Toll receptor pathway in susceptibility to inflammatory bowel diseases. *Genes Immun.* **8**, 387–397 (2007).
54. Lennard-Jones, J.E. Classification of inflammatory bowel disease. *Scand. J. Gastroenterol.* **170** (Suppl.), 2–6 (1989) discussion 16–9.
55. Podolsky, D.K. Inflammatory bowel disease (1). *N. Engl. J. Med.* **325**, 928–937 (1991).
56. de Bakker, P.I., Yelensky, R., Pe'er, I., Gabriel, S.B., Daly, M.J. & Altshuler, D. Efficiency and power in genetic association studies. *Nat. Genet.* **37**, 1217–1223 (2005).
57. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
58. DeLano, W.L. *The PyMOL Molecular Graphics System* (DeLano Scientific, Palo Alto, CA, 2002).