Gastroenterology 2017;153:550–565

Genetic Factors Interact With Tobacco Smoke to Modify Risk for Inflammatory Bowel Disease in Humans and Mice

Pankaj Yadav,[1,*] **David Ellinghaus**,[2,*] Gaëlle Rémy,[3] Sandra Freitag-Wolf,[1] Anabelle Cesaro,[3] Frauke Degenhardt,[2] Gabrielle Boucher,[4] Myriam Delacre,[3] The International IBD Genetics Consortium, Laurent Peyrin-Biroulet,[5] Muriel Pichavant,[3] John D. Rioux,[4,6] Philippe Gosset,[3] Andre Franke,[2] L. Philip Schumm,[7] Michael Krawczak,[1] Mathias Chamaillard,[3,§] Astrid Dempfle,[1,§] and Vibeke Andersen[8,9,10,§]

[1]Institute of Medical Informatics and Statistics, Christian-Albrechts-University of Kiel, Kiel, Germany; [2]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany; [3]University of Lille, Centre National De La Recherche Scientifique, Institut National de la Santé et de la Recherche Médicale, Le Centre Hospitalier Régional Universitaire Lille, Institut Pasteur de Lille, Centre d'Infection et d'Immunité de Lille, Lille, France; [4]Research Center, Montreal Heart Institute, Montreal, Quebec, Canada; [5]Department of Gastroenterology and INSERM Unité 954, Nancy University Hospital, Lorraine University, Vandoeuvre-lès-Nancy, France; [6]Faculté de Médecine, Université de Montréal, Montreal, Quebec, Canada; [7]Department of Public Health Sciences, University of Chicago, Chicago, Illinois; [8]Molecular Diagnostic and Clinical Research Unit, Institut for Regional Sundhedsforskning, Center Sønderjylland, University of Southern Denmark, Odense, Denmark; [9]Institute of Molecular Medicine, University of Southern Denmark, Odense, Denmark; and [10]Laboratory Center, Hospital of Southern Jutland, Aabenraa, Denmark

See Covering the Cover synopsis on page 337.

BACKGROUND & AIMS: The role of tobacco smoke in the etiology of inflammatory bowel disease (IBD) is unclear. We investigated interactions between genes and smoking (gene–smoking interactions) that affect risk for Crohn's disease (CD) and ulcerative colitis (UC) in a case-only study of patients and in mouse models of IBD. **METHODS:** We used 55 Immunochip-wide datasets that included 19,735 IBD cases (10,856 CD cases and 8879 UC cases) of known smoking status. We performed 3 meta-analyses each for CD, UC, and IBD (CD and UC combined), comparing data for never vs ever smokers, never vs current smokers, and never vs former smokers. We studied the effects of exposure to cigarette smoke in *Il10*$^{-/-}$ and *Nod2*$^{-/-}$ mice, as well as in Balb/c mice without disruption of these genes (wild-type mice). Mice were exposed to the smoke of 5 cigarettes per day, 5 days a week, for 8 weeks, in a ventilated smoking chamber, or ambient air (controls). Intestines were collected and analyzed histologically and by reverse transcription polymerase chain reaction. **RESULTS:** We identified 64 single nucleotide polymorphisms (SNPs) for which the association between the SNP and IBD were modified by smoking behavior (meta-analysis Wald test $P < 5.0 \times 10^{-5}$; heterogeneity Cochrane Q test $P > .05$). Twenty of these variants were located within the HLA region at 6p21. Analysis of classical HLA alleles (imputed from SNP genotypes) revealed an interaction with smoking. We replicated the interaction of a variant in *NOD2* with current smoking in relation to the risk for CD (frameshift variant fs1007insC; rs5743293). We identified 2 variants in the same genomic region (rs2270368 and rs17221417) that interact with smoking in relation to CD risk. Approximately 45% of the SNPs that interact with smoking were in close vicinity (≤1 Mb) to SNPs previously associated with IBD; many were located near or within genes that regulate mucosal barrier function and immune tolerance. Smoking modified the disease risk of some variants in opposite directions for CD vs UC. Exposure of Interleukin 10 (il10)-deficient mice to cigarette smoke accelerated development of colitis and increased expression of interferon gamma in the small intestine compared to wild-type mice exposed to smoke. NOD2-deficient mice exposed to cigarette smoke developed ileitis, characterized by increased expression of interferon gamma, compared to wild-type mice exposed to smoke. **CONCLUSIONS:** In an analysis of 55 Immunochip-wide datasets, we identified 64 SNPs whose association with risk for IBD is modified by tobacco smoking. Gene–smoking interactions were confirmed in mice with disruption of *Il10* and *Nod2*—variants of these genes have been associated with risk for IBD. Our findings from mice and humans revealed that the effects of smoking on risk for IBD depend on genetic variants.

Keywords: Animal Model; Nicotine; Inflammation; Gene–Environment Interaction.

Inflammatory bowel diseases (IBD), including Crohn's disease (CD; MIM 266600) and ulcerative colitis (UC; MIM 191390), are chronic lifelong illnesses of early onset that seriously impede the quality of life of patients and their families. IBD is affecting >2.5 million people in Europe (approximately 0.5%) and is becoming increasingly

***Authors share co-first authorship; §Authors share co-senior authorship.**

Abbreviations used in this paper: **CD, Crohn's disease; CI, confidence interval; G×E, gene–environment interaction; GWAS, genome-wide association study; HLA, human leukocyte antigen; IBD, inflammatory bowel diseases; IIBDGC, International Inflammatory Bowel Disease Genetics Consortium; *Il10*, interleukin 10; *Nod2*, nucleotide-binding oligomerization domain 2 protein; OR, odds ratio; SNP, single nucleotide polymorphism; UC, ulcerative colitis; WT, wild-type.**

Most current article

0016-5085/$36.00
http://dx.doi.org/10.1053/j.gastro.2017.05.010

EDITOR'S NOTES

BACKGROUND AND CONTEXT

Current smoking has been found to predispose to Crohn's disease but may protect against ulcerative colitis, whereas former smoking has been associated with risk of both Crohn's disease and ulcerative colitis through yet poorly understood mechanisms.

NEW FINDINGS

Using gene-smoking interaction analyses of nearly 20,000 patients with inflammatory bowel disease (IBD), 64 genetic variants were identified whose association with risk for IBD is modified by tobacco smoking. Gene–smoking interactions were confirmed in mice with disruption of Il10 and Nod2.

LIMITATIONS

This study did not measure the functional implications of most of the involved genetic loci.

IMPACT

The findings from mice and humans revealed that tobacco smoking modifies the IBD risk associated with specific genetic variants.

frequent in Asia and in developing countries.[1] The etiology of IBD involves both genetic and environmental factors, but the biological mechanisms of IBD development are still poorly understood. In particular, little is known about the possible role of gene−environment interaction (G×E) in IBD. In consequence, despite the many genotype−phenotype associations that have been identified in past genome-wide association studies (GWAS), >70% of the heritability of IBD is still unaccounted for.[1–3]

Smoking is the only well-established environmental risk factor for IBD.[4–6] Early case-control studies revealed an increased risk for both CD and UC in former smokers, whereas current smoking seems to predispose to CD, but to protect against UC.[6,7] This differential effect on risk was recently confirmed in a large prospective study of 229,111 women from the US Nurses' Health Study,[6] where the CD hazard ratio was found to be 1.35 for former and 1.90 for current smokers, using never smokers as a reference. By contrast, the UC hazard ratio was 1.56 for former, but 0.86 for current smokers. However, with a 95% confidence interval (CI) ranging from 0.61 to 1.20, the apparent UC protective effect of current smoking was not statistically significant.

The etiologic role of smoking in IBD is not yet fully understood, mainly because of the complex chemical composition of tobacco smoke.[8] Many candidate mechanisms appear worth consideration, including epigenetic changes that alter gene expression relevant to the innate and adaptive immune responses.[8] Smoking also induces compositional changes of the gut microbiota, which provides a plausible link to disease etiology as well.[9,10] Other possible mechanisms involve the post-translational modification of key proteins by constituents of tobacco smoke that activates the immune response and induces inflammation. For example, smoking has been found to induce citrullination of various proteins.[11] Citrullination affects the 3-dimensional structure of proteins in such a way that the latter may unfold and interior domains become exposed that can subsequently act as antigens. In rheumatoid arthritis, for example, smoking has been identified as an environmental trigger of anti-citrulline immunity in individuals with particular *HLA-DRB1* "shared epitope" alleles, a mechanism that might also explain why UC risk stays high even decades after smoking cessation.[6,12]

G×E studies are one way to unravel the biological mechanisms of disease development. As yet, however, only a few studies of interactions between genes and smoking (gene−smoking interactions) have been conducted in the context of IBD.[13–15] One of these studies reported a statistically significant interaction between *NOD2* gene variant 1007fs, predisposing to CD, and both ever and current smoking.[13] Two other small studies observed a significantly higher risk of CD for smokers among GG homozygotes for single nucleotide polymorphism (SNP) rs2241880 in the *ATG16L1* gene and among CC (wild-type [WT]) homozygotes for SNP rs1343151 in the *IL23R* gene.[14,15]

So far, gene−smoking interactions in IBD have not been investigated at a genome-wide level. Using the genotype data available from the International Inflammatory Bowel Disease Genetics Consortium (IIBDGC), we investigated whether the relative IBD risk of smokers is modified by any of the genetic variants included on the Illumina Immunochip (Illumina, San Diego, CA) itself, or by variants in the HLA region that can be imputed from Immunochip data using publicly available databases. For that purpose, we adopted a 2-tiered approach, including the verification in control individuals of the gene−smoking independence assumption implicit to the case-only design (stage I), followed by a case-only analysis to identify gene−smoking interactions (stage II). The epidemiological findings were complemented by functional studies of mice deficient for 2 of the genes identified as potential G×E partners and that encode for interleukin 10 (referred as *Il10*) and nucleotide-binding oligomerization domain 2 protein (referred as *Nod2*).

Materials and Methods

Inflammatory Bowel Diseases Immunochip Dataset

All DNA samples used in the present study were collected through the IIBDGC and originated from 48 sites in 17 countries in Europe, North America, and Australia.[2] Genotyping with the Immunochip custom genotyping array (Illumina, San Diego, CA) was performed in 34 batches in 11 different centers, as described elsewhere.[2] After quality control,[16] genotype data for a total of 132,890 SNPs with minor allele frequency >1% were tested for an interaction with smoking. For SNPs identified as potential G×E partners, additional quality control was carried out by visual inspection of the corresponding cluster plots.

Only samples with known smoking status were included in our study. We confined our meta-analyses to those IIBDGC centers that provided at least 10 samples with either CD or UC in each of the 3 smoking categories (never, current, or former), which yielded a total of 19,735 cases (10,856 CD, 8879 UC; Table 1). Ten of the participating centers also had genotype

Table 1. Overview of Case-Only and Control Data Used for Single Nucleotide Polymorphisms–Smoking Interaction Meta-analyses

Study center	CD			UC			Controls, n
	Cases, n	Current smokers, %	Former smokers, %	Cases, n	Current smokers, %	Former smokers, %	
USA, Los Angeles	1451	11.3	8.3	791	7.4	16.6	NA
Italy, Florence	1068	37.3	13.3	765	12.4	26.3	NA
Belgium, Leuven	908	37.4	7.5	516	21.7	29.8	340
Germany, Kiel	714	31.4	16.1	692	12.9	28.0	2490
UK, Newcastle	655	22.9	25.8	553	6.5	30.2	NA
UK, Exeter	428	38.3	19.6	663	15.7	38.5	NA
USA, Pittsburgh	620	28.4	8.1	449	7.8	19.1	312
Australia, Brisbane	435	44.4	7.8	447	19.9	28.6	528
New Zealand, Christchurch	435	25.5	23.4	425	13.2	37.4	NA
UK, Edinburgh	339	24.2	33.3	399	9.8	43.6	NA
Belgium, Liege	349	50.1	5.4	255	13.7	18.0	122
Canada, Toronto[a]	298	15.8	7.7	294	9.2	18.4	112
Canada, Montreal	293	30.0	8.9	202	13.9	35.1	258
UK, Cambridge	348	31.9	16.7	133	9.0	24.1	NA
Sweden, Örebro	293	33.8	16.4	151	16.6	29.1	NA
Lithuania	112	22.3	19.6	297	12.8	26.9	NA
UK, Torbay	111	31.5	36.9	294	6.5	39.1	NA
UK, Oxford	125	28.8	19.2	257	9.3	39.3	NA
UK, London	362	31.5	26.0	NA	NA	NA	NA
Canada, Toronto[b]	195	26.2	9.2	148	7.4	26.4	NA
Australia, Fremantle	178	24.7	27.5	165	12.7	35.8	NA
USA, Yale	182	17.0	11.5	149	10.1	16.8	322
UK, Dundee	108	42.6	17.6	216	18.5	38.4	NA
USA, Chicago	155	23.2	9.0	128	8.6	19.5	92
Norway	NA	NA	NA	272	11.0	31.6	NA
Sweden, Karolinska	210	23.3	32.9	NA	NA	NA	NA
Belgium, Brussels	192	40.6	15.1	NA	NA	NA	NA
Germany, Munich	NA	NA	NA	168	11.3	24.4	311
Netherlands, Groningen	152	42.1	14.5	NA	NA	NA	NA
Australia, Adelaide	92	33.7	15.2	50	22.0	20.0	NA
Australia, Townsville	48	31.2	39.6	NA	NA	NA	NA
Total	10,856			8879			4887

NA, not available
[a]National Institute of Diabetes, Digestive and Kidney Diseases, IBD Genetics Consortium.
[b]University of Toronto.

data from controls available (N = 8143), 60% of which (n = 4887) were of known smoking status.

To study the specific role of the HLA region, we used imputed classical HLA alleles for the 19,735 cases of interest from a previous IIBDGC study[16] (see Supplementary Methods).

Animal Studies

All animal studies were approved by the local investigational review board (AF 16/20090) in an accredited establishment at the Institute Pasteur de Lille (no. B59-108) according to governmental guidelines no. 86/609/CEE. Age- and sex-matched *Nod2*-deficient (*Nod2*$^{-/-}$), interleukin 10-deficient (*Il10*$^{-/-}$) and control Balb/c mice (without disruption of these genes; WT mice) had free access to a standard laboratory chow diet in a half-day light cycle exposure and temperature-controlled environment. 3R4F research cigarettes were purchased from the University of Kentucky. Eight- to 10-week-old mice were exposed to the smoke of 5 cigarettes per day, 5 days a week, for 8 weeks, in a ventilated smoking chamber (InExpose System; Emka, Scireq, Canada). The control group was exposed to ambient air.

Formalin-fixed, paraffin-embedded colon specimens were blindly scored for inflammation by 2 investigators (see Supplementary Methods). Relative messenger RNA levels were determined in colon samples according to standard methods using Actb as an internal reference gene (see Supplementary Methods).

Statistical Analysis

All statistical analyses of the human data were performed with either PLINK[17] or the R software (v. 3.2.1), as appropriate. The statistical significance of pairwise SNP–smoking interactions was assessed by logistic regression analysis as implemented in PLINK,[17] following a case-only approach.[18] We employed an additive allelic model of the genotype–phenotype relationship and encoded individual SNP genotypes (G) by allele counts. Genotypes were treated as predictor variables,

whereas the binary smoking status (E, see Equation ') was treated as the response variable, that is.

$$logit\{P(E = 1)\} = \theta_0 + \theta \bullet G \quad (1)$$

Following Piegorsch et al,[18] we do not include any additional predictor variables, such as age or sex, into the model.

Any significant association between G and E that occurs in cases points toward G×E at the population level, provided that the 2 assumptions underlying the case-only design are met, namely that the disease is sufficiently rare (ie, prevalence <5%) and G and E are uncorrelated in the general population.[18] Note that the case-only approach does not involve any further assumptions. The case-only paradigm is exemplified in Supplementary Table 10 for SNPs rs17221417 and rs2270368 from the *NOD2* gene region. Conceptually, multiplicative interaction between G and E is defined as the extent to which the true joint effect of G and E differs from the product of the 2 individual effects. From a case-only study, the genotypic odds ratio (OR) for exposure, that is, the odds of E given the presence of G divided by the odds of E given the absence of G, can be derived by taking the antilog of the θ estimate (Equation 1). One premise of the case-only design is that this OR can be interpreted as the multiplicative interaction between G and E on disease risk.

We performed a 2-tiered G×E study separately for CD, UC, and IBD (ie, CD and UC combined). In stage I, the validity of the G−E independence assumption underlying the case-only design was assessed for all 132,890 SNPs. To this end, the logistic regression model of the Equation was fitted to the available control data. In total, 15,196 SNPs were removed because of a nominally significant violation of the G−E independence assumption in controls (meta-analysis $P < .05$) for at least 1 of the 3 smoking contrasts never vs ever, never vs current or never vs former, leaving 117,694 SNPs for stage II.

In stage II, case-only analyses of gene−smoking interaction were carried out for all 117,694 SNPs and for 11,248 variants in the HLA region separately for each participating center. Population stratification correction was performed in individual study centers, following a recently proposed genomic control-based approach for case-only studies[19] (see Supplementary Methods).

The case-only analyses were carried out in triplicate, each time considering 1 of the 3 smoking contrasts—never vs ever, never vs current, and never vs former (Table 1). For meta-analysis, fixed- and random-effect models were fitted to the results using PLINK. An SNP was considered worth further consideration if the meta-analysis gene−smoking interaction (Wald) test yielded $P < 5.0 \times 10^{-5}$ and the heterogeneity (Cochrane Q) test yielded $P > .05$. Note that these criteria were not meant to control the family-wise error rate, that is, define a threshold for genome-wide statistical significance. In recognition of many previous human studies, and particularly our own mouse data, the present study was not geared toward disproving a genome-wide lack of gene−smoking interaction for IBD (eg, the "global null hypothesis"), but rather served to identify the strongest candidate genes for G×E. Thus, the significance thresholds employed here served as sensible filters to prioritize nominally significant findings. We also calculated the false discovery rate[20] for each SNP to control the estimated proportion of false-positive results among the identified potential G×E. We adopted 0.1 as a common threshold for the false discovery rate in subsequent considerations. At the chosen significance level of $\alpha = 5.0 \times 10^{-5}$, the meta-analysis of the never vs ever IBD cohorts (n = 19,735) had approximately 90% power[21] to detect even a small G×E effect (OR, 1.15), assuming a risk allele frequency of 0.15, a smoking frequency of 0.4, a genetic OR of 1.2 (as observed, on average, for IBD[1]), and a smoking OR of 1.4.[6] Expectedly,[21–23] a case-control analysis including all cases plus the available 4887 controls with known smoking status would have yielded dramatically smaller power of only 3%, assuming the same interaction effect and leaving all other parameters unchanged. Because smoking rates differed between centers, we also investigated, for each of the 64 SNPs identified as potential G×E partners, the relationship between the center-specific interaction ORs and smoking rates, using Spearman correlation coefficient.

To assess whether a given region harbored multiple independent gene−smoking interactions, regions with more than 1 SNP with $P < 5.0 \times 10^{-5}$ were scrutinized further. All analyses were repeated including into the respective statistical model the SNP with the smallest gene−smoking interaction P value within a given 1-Mb region (henceforth called the "top SNP") as a mandatory predictor. SNPs with a nominally significant Wald test ($P < .05$) in the conditional analysis were deemed independent gene−smoking interaction partners.

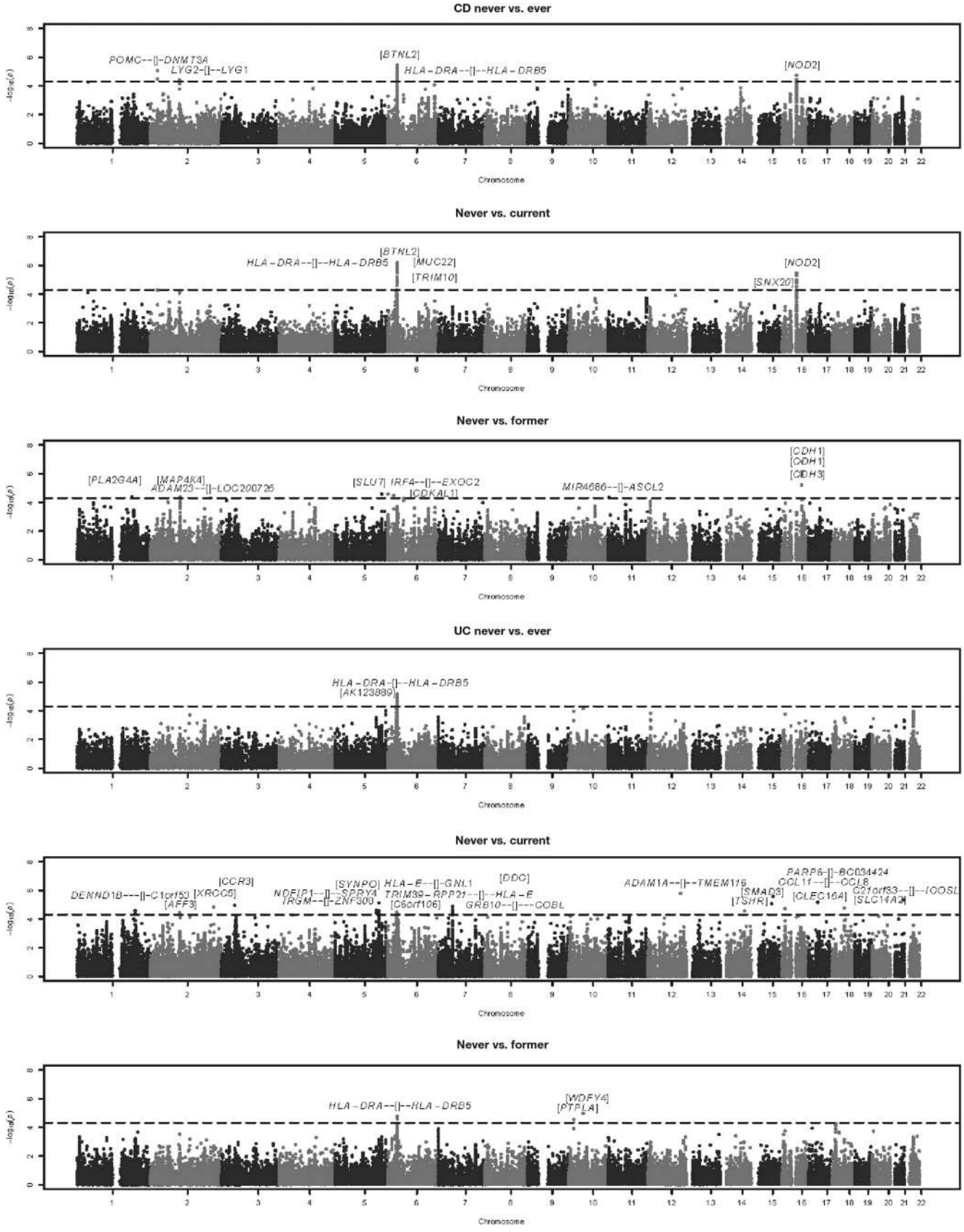
To assess whether the gene−smoking interactions identified in our study overlapped or coincided with previously reported IBD associations,[1,24] pair-wise linkage disequilibrium was estimated in the available control samples (n = 8143) irrespective of whether smoking information was also available or not. To this end, r^2 was computed in each center between pairs of SNPs no more than 1 Mb apart, where one showed an interaction in our study and one had been identified as a genetic main effect in a previous GWAS,[1,24] followed by the calculation of a sample size−weighted average of the center-wise r^2 values.

In order to identify SNPs that show interaction with smoking in opposite direction in CD and UD, we searched for SNPs with a meta-analysis gene–smoking interaction (Wald test) $P < .01$ and a heterogeneity (Cochrane Q test) $P > .05$ for which the case-only G×E OR for one and the same risk allele was reversed between CD and UC (ie, OR <1 in CD and OR >1 in UC, or vice versa).

We illustrated the validity of the case-only approach by performing case-control analyses of CD for 2 selected SNPs from the *NOD2* gene region (see Supplementary Methods). These calculations naturally had to be confined to the centers that provided controls with smoking information. We also carried out stratified analyses of the genetic main effects of the 2 SNPs in never smokers and current smokers. We confined the genome-wide analysis to a case-only approach because this has much higher power than a case-control approach, as was noted already.

The mice data were analyzed statistically using a Kruskal-Wallis test or 2-way analysis of variance as implemented in GraphPad Prism 5, version 5.02 (GraphPad, La Jolla, CA). Statistical significance was defined as $P < .05$; measurements were summarized as mean ± SEM.

Functional annotation of the interacting SNPs and gene prioritization, pathway and tissue/cell type enrichment analysis, and regional linkage disequilibrium plots and annotation of association boundaries were performed as described in the Supplementary Methods using publicly accessible databases.


CD never vs. ever
POMC--[]-DNMT3A
LYG2-[]--LYG1
[BTNL2]
HLA-DRA--[]--HLA-DRB5
[NOD2]
Never vs. current
HLA-DRA--[]--HLA-DRB5
[BTNL2]
[MUC22]
[TRIM10]
[NOD2]
[SNX20]
Never vs. former
[PLA2G4A]
[MAP4K4]
ADAM23--[]-LOC200726
[SLU7] IRF4--[]--EXOC2
[CDKAL1]
MIR4686--[]-ASCL2
[CDH1]
[CDH1]
[CDH3]
UC never vs. ever
HLA-DRA-[]--HLA-DRB5
[AK123889]
Never vs. current
DENND1B---[]-C1orf150
[AFF3]
[XRCC5]
[CCR3]
NDFIP1--[]--SPRY4
IRGM--[]-ZNF300
[SYNPO] HLA-E--[]-GNL1
TRIM39-RPP21--[]--HLA-E
[C6orf106]
[DDC]
GRB10--[]---COBL
ADAM1A--[]--TMEM116
[TSHR]
[SMAD3]
PARP6-[]-BC034424
CCL11--[]--CCL8
[CLEC16A]
C21orf33--[]--ICOSL
[SLC14A2]
Never vs. former
HLA-DRA--[]--HLA-DRB5
[WDFY4]
[PTPLA]
-log10(p)
Chromosome

Results

Single Nucleotide Polymorphism−Smoking Interaction

Three Immunochip-wide meta-analyses of the interaction between smoking behavior (contrasts never vs ever, never vs current, and never vs former) and genotype (117,694 SNPs complying with the G−E independence assumption in controls, minor allele frequency >1%) were performed separately for CD, UC, and IBD (ie, CD and UC combined). Manhattan plots of the meta-analyses results highlighted several potentially interacting loci (Figure 1 and Supplementary Figure 1). Study-wide λ values, calculated to adjust for potential population stratification, were found to be small to moderate, with a maximum of 1.15 obtained in the "USA, Los Angeles" CD cohort (Supplementary Table 1 and Supplementary Figure 2).

With the never vs ever contrast, 46 interacting SNPs for CD (in 5 genomic regions), 53 interacting SNPs for UC (in 1 genomic region), and 65 interacting SNPs for IBD (in 3 genomic regions) initially fulfilled our filter criteria (Supplementary Table 2). When conditioning upon the genotypes of the region-specific top SNPs, 1 (CD), 1 (UC), and 3 (IBD) additional SNPs were found to exhibit residual SNP−smoking interaction of nominal significance (Supplementary Table 3), thereby indicating potentially independent interaction signals from the same genomic regions. Overall, a total of 6 SNPs (5 top, 1 additional independent), 2 SNPs (1, 1), and 6 SNPs (3, 3) were identified as interacting with smoking for CD, UC, and IBD risk, respectively, in the never vs ever smoker meta-analyses. Similarly, meta-analyses with the never vs current and never vs former smoker contrasts identified an additional 18 interacting SNPs (13 top, 5 additional independent) for CD, 24 SNPs (21, 3) for UC and 23 SNPs (18, 5) for IBD (Supplementary Tables 4–7). In summary, considering at least 1 of the 3 smoking contrasts and after adjustment for possible population stratification, 19 SNPs for CD, 25 SNPs for UC, and 25 SNPs for IBD were identified as interacting with smoking according to our filter criteria (Wald test, $P < 5.0 \times 10^{-5}$; Cochrane Q test, $P > .05$). Because 2 SNPs (rs9268923 and rs117782746) were overlapping between CD and IBD and 3 SNPs (rs3129890, rs7747521, and rs116883185) were overlapping between UC and IBD, the total number of unique SNPs was 64 (Figure 2, Table 2; for regional linkage disequilibrium plots, see Supplementary Figures 3–11). Some 52 of these 64 SNPs yielded a false discovery rate <0.1. Interestingly, the largest number of interacting markers was identified with the never vs current smoker contrast for UC (*middle panel* in Figure 2), but with the never vs former smoker contrast for CD (*bottom panel*). No correlation became apparent for any SNP between the center-specific smoking rates and the center-specific interaction ORs (Supplementary Figure 12). For the *NOD2* risk locus, we replicated the interaction between tobacco smoke exposure and the frameshift polymorphism fs1007insC (rs5743293; $P = 4.5 \times 10^{-3}$ for never vs current smoker contrast in CD) and identified 2 suggestive independent SNP-smoking interactions in the same gene region (rs2270368, $P = 2.9 \times 10^{-5}$, never vs current; rs17221417, $P = 3.3 \times 10^{-6}$; never vs current) by means of conditional analysis (Supplementary Table 8).

In addition to the candidate variants in Table 2, seven nominally significant interactions (Wald test, $P < .01$) were found to be of opposite direction for CD and UC (Table 3), which implies that one and the same allele of each of these SNPs increases the risk of CD in smokers, but at the same time protects smokers against UC.

By the time of our study, a total of 238 IBD-associated SNPs had been identified in GWAS.[1,24] To assess their possible overlap with gene−smoking interactions, we quantified the level of linkage disequilibrium between the 64 unique SNPs identified in our gene−smoking interaction study with those 229 IBD-associated SNPs for which we had genotype data available. Some 29 interacting SNPs were found to be located within 1 Mb of a GWAS-identified SNP. However, only 4 pairs of SNPs (IBD-associated, smoking interacting) were found to be in moderate linkage disequilibrium (eg, $r^2 > .05$; Supplementary Table 9).

To further illustrate the validity of the case-only approach, we performed smoking-stratified case-control analyses, considering never vs current smoker contrast, for 2 selected SNPs from the *NOD2* gene region (Supplementary Table 10). In the case of rs2270368, no association with CD was evident in current smokers (OR, 0.96; 95% CI, 0.78−1.16), but a protective effect of the minor allele emerged in never smokers (OR, 0.77; 95% CI, 0.69−0.85). Similarly, a smaller CD risk was found to be associated with rs17221417 in current smokers (OR, 1.28; 95% CI, 1.07−1.52) than in never smokers (OR, 1.74; 95% CI, 1.58−1.91).

HLA−Smoking Interaction

In our focused analyses of classical HLA alleles, a gene−smoking interaction meeting our filtering criteria was observed for 4 alleles in CD, 1 allele in UC, and 5 alleles in IBD (Table 4). Overall, unique alleles were identified with suggestive evidence of gene−smoking interaction with at least 1 smoking-status contrast. All of these alleles were

Figure 1. Manhattan plots of 6 Immunochip-wide meta-analyses highlighting potentially smoking-interacting loci for CD and UC. *Panels 1−3 (from top)* refer to 3 different smoking contrasts for CD: never vs ever, never vs current and never vs former. Similarly, *panels 4−6 (from top)* refer to 3 different smoking contrasts for UC. The *horizontal dashed line* indicates the threshold ($P = 5.0 \times 10^{-5}$) for suggestive evidence of gene−smoking interaction. All SNPs included in the analyses complied with the G−E independence assumption underlying the case-only design, leaving 117,694 SNPs for case-only meta-analysis. Top SNPs from Table 2 and the gene context (see Table 2 legend for gene context definition) are shown above the suggestive threshold line (in *bold*). Manhattan plots for IBD (ie, CD and UC combined) are provided as Supplementary Figure 1.

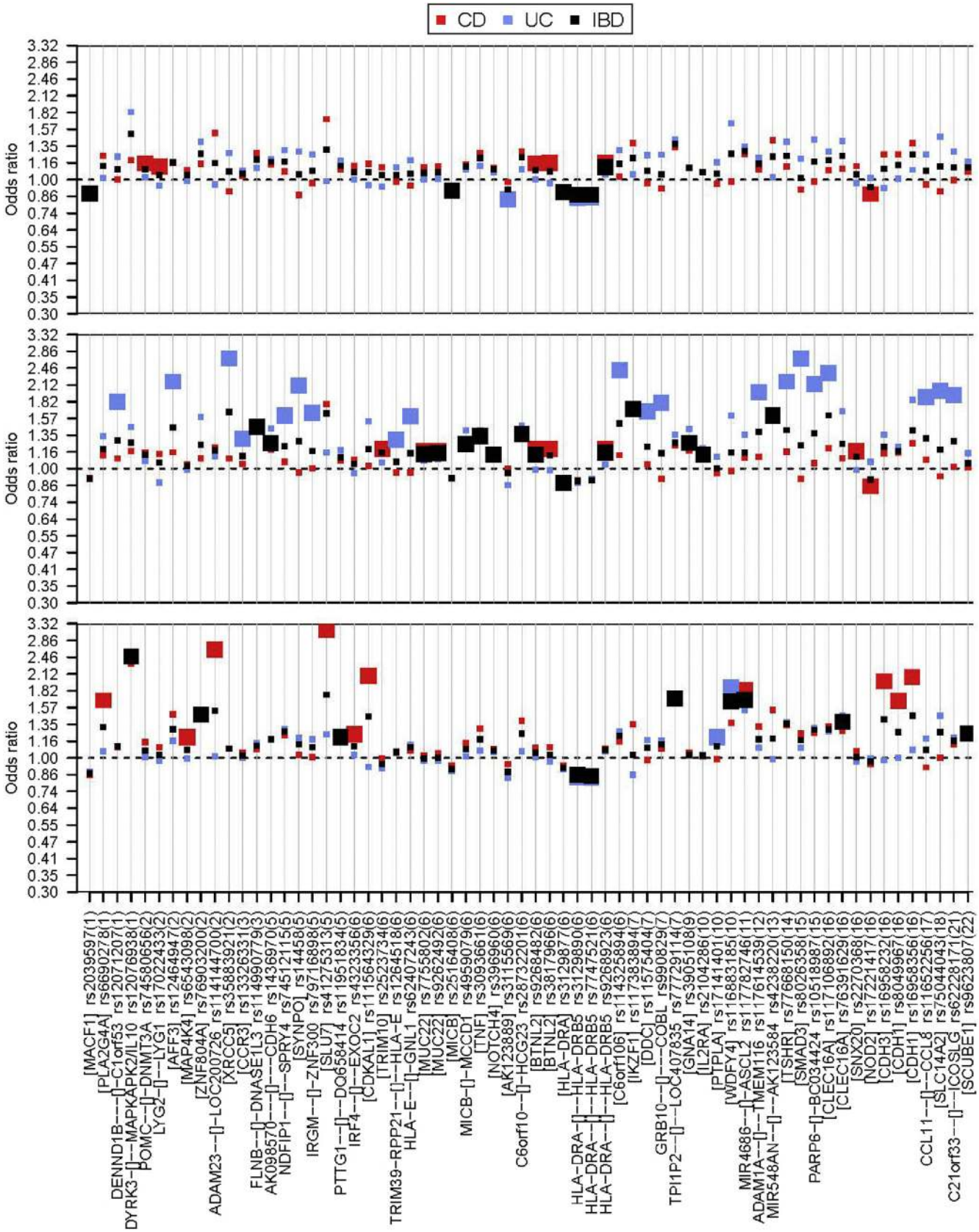


Figure 2. Visualization of SNPs identified as interacting with smoking (Table 2). The odds ratio is shown for all markers with suggestive gene–smoking interaction ($P < 5.0 \times 10^{-5}$); gene context (*see Table 2 legend* for gene context definition) is provided with each rs-number and chromosome numbers are given in *parentheses*. The 3 *panels* refer to different smoking contrasts, namely never vs ever (*top*), never vs current (*middle*) and never vs former (*bottom*). The *square color* refers to the IBD type (CD: *red*, UC: *blue*, IBD: *black*), *large squares* mark meta-analysis (Wald test) $P < 5.0 \times 10^{-5}$.

Table 2. Single Nucleotide Polymorphisms With a Suggestive Gene–Smoking Interaction (Meta-Analysis Wald Test $P < 5.0 \times 10^{-5}$ and Heterogeneity Cochrane Q Test $P_h > .05$) for at Least 1 Smoking Contrast (Never vs Ever, Never vs Current, or Never vs Former)

				Never (n = 6052) vs ever (n = 4804) cases					Never (n = 6052) vs current (n = 3177) cases					Never (n = 6052) vs former (n = 1627) cases)				
Chr:Pos	Gene context[a]	Alleles (A/B)[b]	SNP[c]	OR[d] (95% CI)	*P* value[e]	FDR	P_h value[f]	N_S	OR[d] (95% CI)	*P* value[e]	FDR	P_h value[f]	N_S	OR[d] (95% CI)	*P* value[e]	FDR	P_h value[f]	N_S
CD																		
1:186946010	[PLA2G4A]	A/G	rs6690278	1.24 (1.03–1.50)	.026	1.00	.49	28	1.13 (0.90–1.41)	.284	0.985	.72	28	**1.67 (1.31–2.14)**	**4.0×10^{-5}**	0.324	.96	28
2:25451789	POMC–[]-DNMT3A	A/G	rs74580656	**1.16 (1.08–1.23)**	**8.3×10^{-6}**	0.051	.10	29	1.16 (1.08–1.25)	5.1×10^{-5}	0.105	.23	29	1.15 (1.05–1.26)	3.2×10^{-3}	0.610	.17	29
2:99873658	LYG2-[]–LYG1	A/T	rs17022433	**1.13 (1.06–1.19)**	**3.9×10^{-5}**	0.115	.90	29	1.14 (1.07–1.21)	5.8×10^{-5}	0.112	.81	29	1.10 (1.02–1.20)	.018	0.742	.20	29
2:102444956	[MAP4K4]	G/A	rs6543098	1.09 (1.03–1.16)	5.5×10^{-3}	0.895	.44	29	1.04 (0.97–1.12)	.254	0.980	.46	29	**1.20 (1.10–1.31)**	**4.0×10^{-5}**	0.324	.95	29
2:207500875	ADAM23–[]-LOC200726	A/G	rs114144700	1.52 (1.00–2.31)	.050	1.00	.95	10	1.21 (0.76–1.93)	.419	0.989	1.00	9	**2.64 (1.65–4.21)**	**4.8×10^{-5}**	0.324	.91	13
5:159840935	[SLU7]	A/T	rs41275313	1.72 (1.17–2.52)	5.8×10^{-3}	0.900	.98	16	1.79 (1.11–2.87)	.017	0.879	.96	13	**3.14 (1.84–5.35)**	**2.6×10^{-5}**	0.324	.98	12
6:451373	IRF4–[]–EXOC2	A/G	rs4323356	1.13 (1.05–1.21)	5.2×10^{-4}	0.466	.89	29	1.09 (1.01–1.18)	.037	0.904	.62	29	**1.24 (1.12–1.36)**	**2.6×10^{-5}**	0.324	.75	29
6:20607724	[CDKAL1]	G/A	rs111564329	1.16 (0.89–1.51)	.287	1.00	.57	25	1.03 (0.71–1.49)	.894	0.999	048	18	**2.08 (1.47–2.94)**	**3.2×10^{-5}**	0.324	.57	23
6:30129676	[TRIM10]	C/A	rs2523734	1.12 (1.03–1.20)	4.9×10^{-3}	0.863	.92	29	**1.19 (1.10–1.30)**	**4.9×10^{-5}**[g]	0.105	084	29	1.00 (0.89–1.12)	.964	0.999	.82	29
6:30982209	[MUC22]	G/A	rs7755802	1.12 (1.06–1.18)	1.3×10^{-4}	0.229	.57	29	**1.17 (1.10–1.25)**	**1.4×10^{-6}**	0.007	.21	29	1.02 (0.94–1.11)	.622	0.965	.43	29
6:32367777	[BTNL2]	T/A	rs9268482[h]	**1.16 (1.09–1.23)**	**5.4×10^{-6}**	0.049	.39	29	**1.20 (1.12–1.29)**	**6.1×10^{-7}**	0.006	.38	29	1.09 (1.00–1.20)	.058	0.813	.70	29
6:32367847	[BTNL2]	G/A	rs3817966[h]	**1.16 (1.09–1.23)**	**3.6×10^{-6}**	0.049	.56	29	**1.19 (1.11–1.28)**	**7.5×10^{-7}**	0.006	.54	29	1.10 (1.00–1.20)	.039	0.812	.68	29
6:32432835	HLA-DRA–[]–HLA-DRB5	A/G	rs9268923	**1.16 (1.08–1.24)**	**1.2×10^{-5}**[g]	0.060	.83	28	**1.20 (1.11–1.29)**	**1.7×10^{-6}**[g]	0.007	.92	28	1.09 (0.99–1.20)	.083	0.826	.90	28
11:2282206	MIR4686–[]-ASCL2	A/C	rs117782746	1.26 (1.01–1.57)	.041	1.00	.87	27	1.11 (0.85–1.44)	.437	0.989	.92	24	**1.84 (1.37–2.46)**	**4.3×10^{-5}**	0.324	.95	27
16:50714335	[SNX20]	G/A	rs2270368	1.13 (1.06–1.20)	3.3×10^{-4}	0.399	.06	29	**1.17 (1.09–1.26)**	**2.9×10^{-5}**[g]	0.067	.17	29	1.06 (0.97–1.17)	.218	0.878	.33	29
16:50739582	[NOD2]	C/G	rs17221417	**0.88 (0.83–0.93)**	**1.8×10^{-5}**	0.072	.22	29	**0.85 (0.80–0.91)**	**3.3×10^{-6}**	0.011	.12	29	0.94 (0.87–1.03)	.172	0.861	.42	29
16:68681383	[CDH3]	G/A	rs16958232	1.25 (1.00–1.56)	.049	1.00	.58	26	1.15 (0.88–1.49)	.313	0.987	.41	25	**1.98 (1.47–2.67)**	**5.9×10^{-6}**[g]	0.175	.63	24
16:68822479	[CDH1]	A/G	rs8049967	1.25 (1.08–1.45)	2.7×10^{-3}	0.677	.50	29	1.16 (0.98–1.37)	.093	0.947	.76	29	**1.66 (1.36–2.03)**	**7.4×10^{-7}**[g]	0.043	.36	29
16:68824965	[CDH1]	G/A	rs16958356	1.39 (1.14–1.69)	1.0×10^{-3}	0.547	.92	29	1.26 (1.00–1.58)	.049	0.908	.97	27	**2.06 (1.58–2.70)**	**1.3×10^{-7}**	0.016	.90	27
				Never (n = 5239) vs ever (n = 3640) cases					Never (n = 5239) vs current (n = 1080) cases					Never (n = 5239) vs former (n = 2560) cases				
Chr:Pos	Gene context[a]	Alleles (A/B)[b]	SNP[c]	OR[d] (95% CI)	*P* value[e]	FDR	P_h value[f]	N_S	OR[d] (95% CI)	*P* value[e]	FDR	P_h value[f]	N_S	OR[d] (95% CI)	*P* value[e]	FDR	P_h value[f]	N_S
UC																		
1:197865112	DENND1B—[]-C1orf53	C/A	rs12071207	1.23 (1.01–1.50)	0.041	0.930	.67	26	**1.82 (1.38–2.41)**	**2.8×10^{-5}**	0.102	.81	25	1.11 (0.88–1.40)	.380	0.970	.53	26
2:100746909	[AFF3]	A/T	rs12464947	1.17 (0.92–1.49)	0.195	0.979	.85	23	**2.18 (1.50–3.16)**	**3.8×10^{-5}**	0.102	.95	20	1.16 (0.89–1.53)	.270	0.962	.89	22
2:217005480	[XRCC5]	G/A	rs35883921	1.27 (0.93–1.74)	0.136	0.977	.69	17	**2.69 (1.72–4.20)**	**1.5×10^{-5}**	0.102	.99	16	1.09 (0.72–1.65)	.688	0.983	.78	13
3:46298561	[CCR3]	A/G	rs13326331	1.08 (1.00–1.17)	0.045	0.943	.91	26	**1.30 (1.16–1.46)**	**1.1×10^{-5}**	0.102	.66	26	1.00 (0.91–1.09)	.982	0.999	.89	26
5:141611379	NDFIP1–[]–SPRY4	A/G	rs74512115	1.31 (1.12–1.52)	4.9×10^{-4}	0.264	.16	26	**1.61 (1.29–2.00)**	**2.4×10^{-5}**	0.102	.44	26	1.29 (1.09–1.53)	3.2×10^{-3}	0.784	.06	26
5:150038672	[SYNPO]	G/A	rs14458	1.28 (1.02–1.61)	0.030	0.906	.65	25	**2.10 (1.52–2.91)**	**7.2×10^{-6}**	0.102	.00	24	1.20 (0.93–1.54)	.164	0.962	.86	24
5:150272060	IRGM–[]-ZNF300	A/T	rs79716898	1.25 (1.06–1.47)	7.2×10^{-3}	0.780	.26	26	**1.65 (1.31–2.09)**	**2.7×10^{-5}**[g]	0.102	.72	26	1.18 (0.98–1.43)	.076	0.962	.38	26
6:30410206	TRIM39-RPP21–[]–HLA-E	A/G	rs1264518	1.12 (1.03–1.21)	9.1×10^{-3}	0.812	.11	26	**1.30 (1.14–1.47)**	**4.5×10^{-5}**[g]	0.102	.25	26	1.06 (0.97–1.17)	.202	0.962	.10	26
6:30505000	HLA-E–[]-GNL1	G/A	rs62407243	1.19 (1.03–1.39)	0.022	0.892	.92	26	**1.60 (1.28–1.99)**	**3.3×10^{-5}**	0.102	.99	26	1.12 (0.95–1.33)	.174	0.962	.96	26

Table 2. Continued

Chr:Pos	Gene context[a]	Alleles (A/B)[b]	SNP[c]	Never (n = 5239) vs ever (n = 3640) cases					Never (n = 5239) vs current (n = 1080) cases					Never (n = 5239) vs former (n = 2560) cases				
				OR[d] (95% CI)	*P* value[e]	FDR	P_h value[f]	N_S	OR[d] (95% CI)	*P* value[e]	FDR	P_h value[f]	N_S	OR[d] (95% CI)	*P* value[e]	FDR	P_h value[f]	N_S
6:32224139	[AK123889]	A/G	rs3115569	**0.84 (0.78–0.91)**	**9.0×10^{-6}**[g]	0.063	.94	26	0.86 (0.76–0.97)	.017	0.397	.77	26	0.84 (0.77–0.91)	7.1×10^{-5}	0.307	.91	26
6:32414273	HLA-DRA-[]-HLA-DRB5	G/A	rs3129890	**0.85 (0.79–0.91)**	**7.3×10^{-6}**	0.063	.50	26	0.88 (0.79–0.98)	.025	0.444	.72	26	**0.84 (0.78–0.91)**	**2.6×10^{-5}**	0.302	.44	26
6:32431105	HLA-DRA–[]–HLA-DRB5	G/A	rs7747521	**0.86 (0.80–0.92)**	**2.1×10^{-5}**	0.067	.55	26	0.92 (0.82–1.02)	.112	0.650	.85	26	**0.84 (0.77–0.91)**	**1.7×10^{-5}**	0.302	.46	26
6:34561587	[C6orf106]	A/T	rs114325894	1.30 (0.96–1.77)	.090	0.977	.31	20	**2.42 (1.58–3.70)**	**4.4×10^{-5}**	0.102	.99	16	1.27 (0.88–1.82)	.197	0.962	.32	19
7:50565963	[DDC]	G/A	rs11575404	1.25 (1.06–1.47)	8.2×10^{-3}	0.798	.44	26	**1.67 (1.33–2.11)**	**1.3×10^{-5}**	0.102	.69	26	1.18 (0.98–1.41)	.088	0.962	.63	26
7:50909698	GRB10–[]—COBL	G/A	rs990829	1.25 (1.02–1.52)	.027	0.892	.29	25	**1.81 (1.37–2.37)**	**2.2×10^{-5}**[g]	0.102	.88	24	1.16 (0.92–1.47)	.195	0.962	.40	25
10:17642806	[PTPLA]	A/G	rs17141401	1.17 (1.08–1.26)	1.1×10^{-4}	0.155	.67	26	1.11 (0.98–1.25)	.087	0.607	.94	26	**1.20 (1.10–1.31)**	**2.9×10^{-5}**	0.302	.32	26
10:50003599	[WDFY4]	A/G	rs116883185	1.66 (1.29–2.12)	6.8×10^{-5}	0.137	.81	19	1.61 (1.09–2.38)	.017	0.392	.92	19	**1.88 (1.42–2.49)**	**1.0×10^{-5}**	0.302	.76	16
12:112357984	ADAM1A–[]–TMEM116	A/G	rs117614539	1.22 (1.00–1.48)	.055	0.964	.57	24	**1.98 (1.50–2.63)**	**1.6×10^{-6}**	0.102	.98	23	1.09 (0.87–1.38)	.454	0.973	.42	25
14:81500600	[TSHR]	A/G	rs77668150	1.40 (1.10–1.78)	5.6×10^{-3}	0.747	.95	25	**2.18 (1.51–3.13)**	**2.7×10^{-5}**	0.102	.96	24	1.38 (1.06–1.79)	.018	0.953	1.00	25
15:67444393	[SMAD3]	A/G	rs8026358	1.21 (0.91–1.60)	.197	0.979	.97	23	**2.68 (1.74–4.14)**	**8.4×10^{-6}**	0.102	.99	19	1.09 (0.79–1.51)	.592	0.980	.98	22
15:72565787	PARP6-[]-BC034424	A/G	rs10518987	1.43 (1.14–1.79)	2.1×10^{-3}	0.579	.76	24	**2.12 (1.55–2.91)**	**2.7×10^{-6}**	0.102	.99	23	1.31 (1.01–1.69)	.041	0.962	.88	25
16:11112283	[CLEC16A]	G/A	rs117106892	1.29 (0.97–1.70)	.077	0.977	.50	21	**2.36 (1.59–3.51)**	**1.9×10^{-5}**	0.102	.95	21	1.27 (0.90–1.79)	.176	0.962	.88	19
17:32632378	CCL11–[]–CCL8	A/G	rs11652256	1.26 (1.03–1.53)	.022	0.892	.59	26	**1.90 (1.44–2.51)**	**6.7×10^{-6}**	0.102	.86	26	1.19 (0.95–1.49)	.129	0.962	.92	26
18:42839184	[SLC14A2]	T/A	rs75044043	1.47 (1.19–1.81)	3.1×10^{-4}	0.221	.59	26	**2.01 (1.46–2.77)**	**1.8×10^{-5}**	0.102	.93	25	1.46 (1.16–1.84)	1.1×10^{-3}	0.693	.88	26
21:45597657	C21orf33–[]–ICOSLG	A/G	rs62228171	1.28 (1.04–1.58)	.017	0.892	.14	24	**1.93 (1.46–2.56)**	**4.6×10^{-6}**	0.102	.56	25	1.20 (0.95–1.51)	.136	0.962	.47	24
IBD																		
1:39749884	[MACF1]	A/C	rs2039597	0.88 (0.83–0.93)	1.5×10^{-5}	0.034	.92	55	0.92 (0.85–0.98)	.017	0.451	.49	55	0.87 (0.81–0.94)	1.8×10^{-4}	0.229	.98	55
1:206827199	DYRK3-[]-MAPKAPK2/IL10	C/A	rs12076938	1.51 (1.10–2.08)	.012	0.750	.58	17	1.26 (0.79–2.02)	.331	0.834	.89	15	2.48 (1.65–3.72)	1.2×10^{-5}	0.090	.87	15
2:185601641	[ZNF804A]	G/A	rs76903200	1.26 (1.09–1.46)	2.2×10^{-3}	0.519	.90	52	1.24 (1.02–1.50)	.028	0.512	.97	48	1.47 (1.23–1.77)	3.2×10^{-5}	0.162	1.00	53
3:58169958	FLNB–[]-DNASE1L3	G/A	rs114990779	1.20 (1.04–1.38)	.014	0.774	.97	53	1.45 (1.22–1.73)	3.3×10^{-5}	0.101	.94	51	1.11 (0.92–1.35)	.270	0.841	1.00	51
5:30639738	AK098570—[]—CDH6	C/A	rs1436970	1.17 (1.08–1.28)	2.5×10^{-4}	0.205	.87	55	1.26 (1.13–1.40)	2.0×10^{-5}	0.095	.87	55	1.18 (1.06–1.32)	2.3×10^{-3}	0.473	.54	55
5:159880356	PTTG1–[]–DQ658414	A/G	rs11951834	1.13 (1.06–1.21)	3.2×10^{-4}	0.223	.02	55	1.11 (1.01–1.21)	.022	0.484	.18	55	1.21 (1.11–1.31)	2.0×10^{-5}	0.125	.35	55
6:30986015	[MUC22]	G/A	rs9262492	1.07 (1.03–1.12)	1.9×10^{-3}	0.491	.36	55	1.15 (1.09–1.21)	6.7×10^{-7}	0.030	.71	55	1.00 (0.95–1.06)	.902	0.988	.20	55
6:31463491	[MICB]	A/G	rs2516408	0.91 (0.87–0.95)	2.4×10^{-5}[g]	0.051	.82	55	0.92 (0.87–0.97)	4.0×10^{-3}	0.337	.54	55	0.91 (0.86–0.96)	6.5×10^{-4}	0.340	.87	55
6:31488879	MICB-[]-MCCD1	A/G	rs4959079	1.13 (1.03–1.23)	7.8×10^{-3}	0.692	.98	55	1.25 (1.12–1.39)	4.5×10^{-5}[g]	0.101	1.00	55	1.08 (0.96–1.21)	.202	0.805	.81	55
6:31543758	[TNF]	A/G	rs3093661	1.22 (1.09–1.35)	3.1×10^{-4}	0.219	.85	55	1.34 (1.18–1.52)	6.6×10^{-6}[g]	0.070	.97	54	1.18 (1.03–1.36)	.016	0.585	.97	55

Chr:Pos	Gene context[a]	Alleles (A/B)[b]	SNP[c]	Never (n = 11,291) vs ever (n = 8444) cases					Never (n = 11,291) vs current (n = 4257) cases					Never (n = 11,291) vs former (n = 4187)				
				OR[d] (95% CI)	*P* value[e]	FDR	P_h value[f]	N_S	OR[d] (95% CI)	*P* value[e]	FDR	P_h value[f]	N_S	OR[d] (95% CI)	*P* value[e]	FDR	P_h value[f]	N_S

Table 2. Continued

Chr:Pos	Gene context[a]	Alleles (A/B)[b]	SNP[c]	Never (n = 11,291) vs ever (n = 8444) cases					Never (n = 11,291) vs current (n = 4257) cases					Never (n = 11,291) vs former (n = 4187)				
				OR[d] (95% CI)	*P* value[e]	FDR	P_h value[f]	N_S	OR[d] (95% CI)	*P* value[e]	FDR	P_h value[f]	N_S	OR[d] (95% CI)	*P* value[e]	FDR	P_h value[f]	N_S
6:32191581	[NOTCH4]	A/T	rs396960	1.09 (1.04–1.15)	2.2×10^{-4}	0.195	.72	55	1.14 (1.07–1.21)	2.6×10^{-5}[g]	0.101	.45	55	1.07 (1.00–1.13)	.039	0.647	.47	55
6:32350454	C6org10–[]–HCG23	A/G	rs28732201	1.22 (1.09–1.37)	8.2×10^{-4}	0.339	.30	54	1.36 (1.18–1.57)	1.6×10^{-5}[g]	0.086	.90	52	1.24 (1.07–1.45)	5.7×10^{-3}	0.507	.86	53
6:32408597	[HLA-DRA]	A/G	rs3129877	0.90 (0.85–0.94)	3.9×10^{-6}[g]	0.012	.55	55	0.88 (0.83–0.93)	2.9×10^{-5}	0.101	.08	55	0.92 (0.86–0.97)	3.8×10^{-3}	0.487	.99	55
6:32414273	HLA-DRA–[]–HLA-DRB5	G/A	rs3129890	0.87 (0.83–0.92)	2.9×10^{-8}	0.001	.10	55	0.90 (0.84–0.95)	3.7×10^{-4}	0.192	.43	55	0.86 (0.81–0.91)	7.4×10^{-7}	0.010	.25	55
6:32431105	HLA-DRA–[]–HLA-DRB5	G/A	rs7747521	0.87 (0.83–0.92)	3.3×10^{-8}	0.001	.14	55	0.90 (0.85–0.96)	1.2×10^{-3}	0.258	.29	55	0.85 (0.80–0.91)	2.3×10^{-7}	0.006	.55	55
6:32432835	HLA-DRA–[]–HLA-DRB5	A/G	rs9268923	1.12 (1.06–1.17)	3.8×10^{-5}[g]	0.071	.66	52	1.15 (1.08–1.23)	9.3×10^{-6}	0.073	.93	52	1.07 (1.00–1.15)	.041	0.654	.72	52
7:50464756	[IKZF1]	A/C	rs117383894	1.21 (0.99–1.49)	.068	0.939	.35	35	1.70 (1.32–2.19)	3.7×10^{-5}	0.101	.82	30	1.03 (0.75–1.40)	.867	0.983	.65	31
7:128729699	TPI1P2–[]–LOC407835	C/A	rs77729114	1.38 (1.12–1.69)	2.2×10^{-3}	0.519	.87	43	1.26 (0.96–1.65)	.093	0.643	.98	40	1.70 (1.34–2.16)	1.1×10^{-5}	0.090	.98	46
9:80226683	[GNA14]	A/G	rs3905108	1.11 (1.02–1.22)	.021	0.819	.48	55	1.26 (1.13–1.41)	4.5×10^{-5}	0.101	.81	55	1.03 (0.92–1.16)	.626	0.947	.90	55
10:6099045	[IL2RA]	G/A	rs2104286	1.07 (1.02–1.12)	5.8×10^{-3}	0.630	.75	55	1.14 (1.07–1.21)	3.5×10^{-5}	0.101	.85	55	1.02 (0.96–1.08)	.509	0.926	.63	55
10:50003599	[WDFY4]	A/G	rs116883185	1.26 (1.06–1.49)	8.2×10^{-3}	0.707	.32	41	1.16 (0.92–1.45)	.205	0.760	.54	40	1.65 (1.33–2.05)	5.6×10^{-6}	0.051	.95	38
11:2282206	MIR4686–[]–ASCL2	A/C	rs117782746	1.29 (1.09–1.53)	3.1×10^{-3}	0.538	.99	46	1.16 (0.92–1.46)	.207	0.761	.99	39	1.68 (1.36–2.07)	1.1×10^{-6}	0.013	1.00	46
13:100108807	MIR548AN–[]–AK123584	G/A	rs4238220	1.23 (1.03–1.47)	.022	0.819	.65	48	1.61 (1.29–2.01)	2.7×10^{-5}	0.101	.70	43	1.19 (0.94–1.52)	.154	0.776	.96	43
16:11269399	[CLEC16A]	C/A	rs76391629	1.24 (1.10–1.40)	6.2×10^{-4}	0.300	.35	53	1.25 (1.07–1.46)	4.9×10^{-3}	0.344	.69	53	1.38 (1.19–1.61)	3.5×10^{-5}	0.169	.95	52
22:43704052	[SCUBE1]	C/A	rs9623807	1.12 (1.04–1.20)	2.5×10^{-3}	0.520	.58	55	1.05 (0.96–1.16)	.280	0.808	.50	55	1.24 (1.14–1.36)	1.6×10^{-6}	0.018	.93	55

NOTE. Bold type indicates meta-analysis Wald test $P < 5.0 \times 10^{-5}$.
CI, confidence interval; Chr:Pos, chromosome number, base-pair position; FDR, false discovery rate; calculated by the Benjamini-Hochberg procedure as implemented in the *p.adjust* tool of R; N_S, number of study centers providing data for meta-analyses of the respective SNP.
[a]Gene(s) spanning or flanking (<1 Mb) the interacting SNP, brackets indicate the position of the SNP, dashes indicate distance to flanking gene (-, >1 kb; –, >10 kb; —, >100 kb).
[b]Minor/major alleles.
[c]All listed SNPs complied with the G–E independence assumption ($P \geq .05$) in healthy controls.
[d]OR, genotypic odds ratio for exposure (see Methods).
[e]*P* value from a fixed-effects inverse-variance meta-analysis, based on center-specific Wald tests. Before meta-analysis, *P* values were individually adjusted for possible population stratification following a genomic control approach.
[f]P_h, heterogeneity *P* value from a Cochrane Q test.
[g]Secondary signals with unconditioned *P* values (see Supplementary Tables 3, 5, and 7 for *P* values of conditional analyses).
[h]rs9268482 showed the most significant interaction ($P = 6.1 \times 10^{-7}$) in the never vs current smoker analysis, whereas rs3817966 showed the most significant interaction ($P = 3.6 \times 10^{-6}$) in the never vs ever analysis.

Table 3. Differential Gene–Smoking Interaction in Crohn's Disease and Ulcerative Colitis

				CD			UC			
Smoking contrast	Chr	SNP	Minor allele	OR[a] (95% CI)	*P* value[b]	P_h value[c]	OR[a] (95% CI)	*P* value[b]	P_h value[c]	*P* value[d]
Never vs ever	6	rs3127599	A	1.10 (1.04–1.17)	1.5×10^{-3}	.88	0.91 (0.85–0.97)	7.1×10^{-3}	.19	3.6×10^{-5}
	11	rs117833518	A	0.81 (0.70–0.95)	8.4×10^{-3}	.49	1.33 (1.12–1.58)	1.2×10^{-3}	.86	3.0×10^{-5}
	19	rs2230330	A	0.58 (0.38–0.87)	8.7×10^{-3}	.76	1.84 (1.18–2.85)	6.8×10^{-3}	.63	1.6×10^{-4}
Never vs current	6	rs176095	G	1.12 (1.04–1.21)	3.3×10^{-3}	.33	0.83 (0.74–0.94)	4.3×10^{-3}	.94	7.1×10^{-5}
	14	rs10400765	G	0.90 (0.83–0.97)	7.3×10^{-3}	.75	1.19 (1.06–1.34)	3.4×10^{-3}	.41	8.6×10^{-5}
	16	rs9940076	A	1.09 (1.02–1.17)	7.3×10^{-3}	.92	0.85 (0.77–0.94)	1.6×10^{-3}	.58	4.1×10^{-5}
Never vs former	1	rs6682359	A	1.12 (1.03–1.22)	5.8×10^{-3}	.94	0.91 (0.85–0.98)	9.0×10^{-3}	.56	1.5×10^{-4}

Chr, chromosome.
[a]Genotypic odds ratio for exposure (see Methods).
[b]*P* value from a fixed-effects inverse-variance meta-analysis, based on center-specific Wald tests. Before meta-analysis, *P* values were individually adjusted for possible population stratification, following a genomic control approach.
[c]Heterogeneity (across study center) *P* value from a Cochrane Q test.
[d]*P* value from a heterogeneity (Cochrane Q) test of OR differences in the CD and UC.

Table 4. HLA alleles Involved in Gene–Smoking Interaction

		Never vs ever			Never vs current			Never vs former		
IBD type	HLA allele	OR (95% CI)	*P* value	P_h value	OR (95% CI)	*P* value	P_h value	OR (95% CI)	*P* value	P_h value
CD	HLA-B*57	1.31 (1.14–1.51)	1.2×10^{-4}	.77	1.39 (1.19–1.62)	2.7×10^{-5}	.70	1.33 (1.09–1.63)	5.1×10^{-3}	.97
	HLA-B*57:01	1.32 (1.15–1.52)	1.1×10^{-4}	.70	1.40 (1.19–1.63)	2.9×10^{-5}	.62	1.34 (1.09–1.65)	4.7×10^{-3}	.97
	HLA-DQA1*02:01	1.21 (1.12–1.31)	1.6×10^{-6}	.55	1.24 (1.14–1.36)	6.6×10^{-7}	.78	1.18 (1.05–1.32)	4.2×10^{-3}	.38
	HLA-DRB1*07:01	1.20 (1.11–1.30)	2.5×10^{-6}	.52	1.24 (1.14–1.35)	1.3×10^{-6}	.74	1.18 (1.05–1.32)	4.0×10^{-3}	.34
UC	HLA-DRB3*91:01	0.81 (0.74–0.89)	1.8×10^{-5}	.50	0.77 (0.66–0.90)	7.9×10^{-4}	.84	0.84 (0.76–0.94)	1.7×10^{-3}	.45
IBD	HLA-B*57	1.25 (1.12–1.40)	9.4×10^{-5}	.84	1.39 (1.22–1.59)	1.4×10^{-6}	.98	1.22 (1.06–1.41)	6.9×10^{-3}	.95
	HLA-B*57:01	1.25 (1.12–1.40)	1.0×10^{-4}	.79	1.40 (1.22–1.61)	1.1×10^{-6}	.97	1.22 (1.05–1.41)	8.8×10^{-3}	.94
	HLA-DQA1*02:01	1.15 (1.08–1.23)	5.5×10^{-6}	.45	1.23 (1.14–1.32)	1.0×10^{-7}	091	1.10 (1.02–1.20)	.015	.24
	HLA-DQB1*02:02	1.14 (1.06–1.22)	3.7×10^{-4}	.37	1.21 (1.10–1.31)	2.6×10^{-5}	.64	1.09 (1.00–1.20)	.058	.45
	HLA-DRB1*07:01	1.15 (1.08–1.22)	7.3×10^{-6}	.46	1.22 (1.13–1.31)	2.3×10^{-7}	.90	1.11 (1.02–1.20)	.013	.23

For details, see legend to Table 2.
P value from a fixed-effects inverse-variance meta-analysis, based on center-specific Wald tests. Before meta-analysis, *P* values were individually adjusted for possible population stratification following a genomic control approach.
CI, confidence interval; OR, genotypic odds ratio for exposure (see Methods); Ph, heterogeneity *P* value from a Cochrane Q test.

found to comply with the G−E independence assumption in controls.

Functional Annotation of Interacting Variants

We functionally annotated the 64 SNPs identified as potentially interacting with smoking for at least 1 of 3 smoking contrasts (see Supplementary Methods). For 37 of the interaction signals (58%), the lead SNP mapped within the transcript of a known gene while 27 signals were located in intergenic regions (Supplementary Table 11). The interacting SNPs included 1 coding missense variant (rs41275313) in the *SLU7* gene. However, all coding SNPs were predicted to be benign (Supplementary Table 12). Non-coding SNPs rs76903200 (*ZNF804A*; intronic), rs79716898 (intergenic between *IRGM* and *ZNF300*), and rs62407243 (intergenic between *HLA-E* and *GNL1*) were found to be potentially deleterious based on the prediction of reduced organismal fitness (Combined Annotation Dependent Depletion scores; Supplementary Table 12). Among coding SNPs that were in strong linkage disequilibrium ($r^2 > 0.8$ in the 1000 Genomes European samples; n = 376 variants) with the 64 lead SNPs (see Methods), we identified another 5 synonymous SNPs at the *ZNF300*, *IRGM*, *UHRF1BP1*, and *NOD2* (2×) gene loci and 1 non-synonymous SNP at the *GRAMD2* gene (Supplementary Table 13).

We also examined which of the 64 lead SNPs and of the SNPs in strong linkage disequilibrium ($r^2 > 0.8$) with the lead SNPs mapped to eQTLs from GTEx and Geuvadis eQTL studies,[25] or to functional annotations of the noncoding genome in different cell types provided by the Roadmap Epigenomics[26] and ENCODE[27] projects, using web tool HaploReg[28] (Supplementary Table 14; see Methods). The results are summarized in Supplementary Figure 13.

Pathway and Cell Type Enrichment Analyses

To ascertain whether genes at the putatively interacting loci were highly expressed in certain tissue/cell types, we conducted pathway and tissue/cell type enrichment analyses using DEPICT[29] with 77,840 microarray gene expression profiles from human, rat, and mouse, and 209 tissue/cell type annotations[30] (see Supplementary Methods). From DEPICT, we identified 24 gene sets (Supplementary Table 15) and 20 tissues (Supplementary Table 16) with significant enrichment of genes within the suggested interacting loci (false discovery rate <0.01). The results point mainly toward perturbation of immune response pathways in blood.

Loss of Nod2 *and* Il10 *Renders Mice Susceptible to Intestinal Inflammation in Response to Cigarette Smoke Exposure*

The possible impact of cigarette smoke on intestinal homeostasis was next evaluated in mice that were deficient for either *Il10* or *Nod2*, two proteins encoded by genes that were identified as potential smoking−interacting genes in our human data (Table 2). While no signs of disease were apparent in WT mice after 8 weeks of cigarette smoke exposure, smoking $Il10^{-/-}$ mice experienced greater body-weight loss (Figure 3*A*), increased disease activity index (Figure 3*B*, significant gene−smoking interaction; $P = .01$ at day 62), and accelerated development of rectal prolapses (Figure 3*C*). Consequently, cigarette smoking markedly reduced the colon length of $Il10^{-/-}$ mice compared to

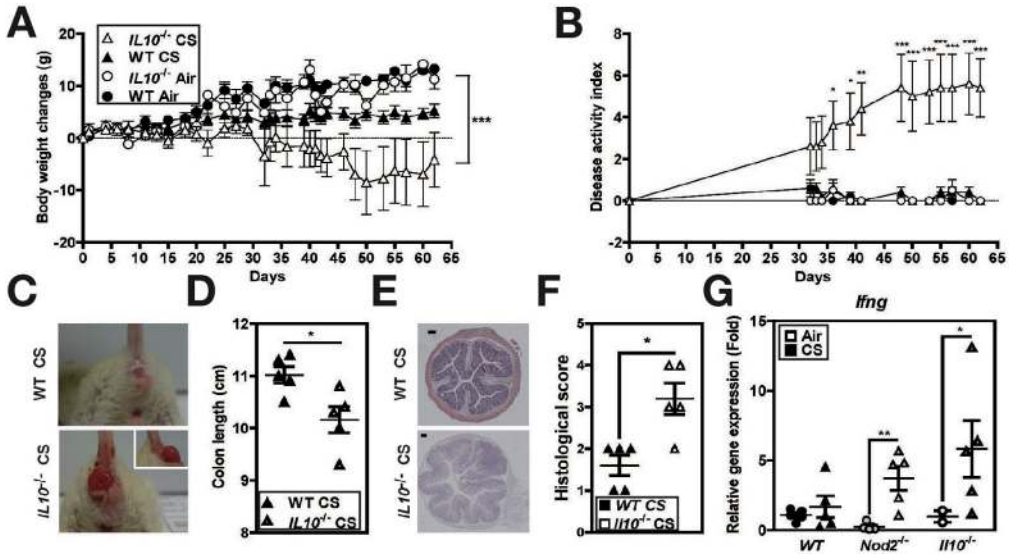


Figure 3. Impact of interleukin-10 and Nod2 deficiency on the risk of intestinal inflammation in cigarette smoke-exposed mice. WT and $Il10^{-/-}$ mice were exposed to cigarette smoke (CS), 5 days a week, during a period of 8 weeks. (*A*) Changes in body weight. (*B*) Disease Activity Index. (*C*) Prolapsus apparition under CS exposure in $Il10^{-/-}$ mice. (*D*) Colon length. (*E*) Representative H&E stainings. (*F*) Histologic score. (*G*) Relative expression level of *Ifng* gene in the ileal tissue of WT, *Nod2*-deficient and *Il10*-deficent mice that were either exposed to CS or not. *$P < .05$; ***$P < .001$ (Kruskal-Wallis test). Data are representative of 2 independent experiments.

similarly treated WT animals (Figure 3*D*). While histologic analysis failed to reveal any signs of inflammation in WT mice, a greater infiltration of immune cells within the colonic mucosa (Figures 3*E* and *F*) and an enhanced expression of *Ifng* transcripts in the ileum were observed in smoking *Il10*$^{-/-}$ mice (Figure 3*G*). Likewise, *Nod2*$^{-/-}$ mice also showed higher expression of *Ifng* in their ileum, but lacked signs of colitis and prolapse development (Figure 3*G* and Supplementary Figure 14). Taken together, our data suggest a strong protective role of both Il10 and Nod2 on intestinal homeostasis in response to tobacco smoking.

Discussion

We performed an extensive meta-analysis to investigate possible gene−smoking interactions in relation to the risk for CD, UC, or IBD. To this end, we used the existing Immunochip-wide data collated by the IIBDGC. Our analysis identified 19 SNPs for CD, 25 SNPs for UC, and 25 SNPs for IBD, that potentially interact with regard to disease risk considering at least 1 of 3 smoking contrasts (never vs ever, never vs current, or never vs former). Interestingly, the largest number of interacting SNPs was identified with the never vs current contrast for UC, but with the never vs former contrast for CD.

Our findings are highly relevant to advancing the understanding of IBD etiology for various reasons. First, the discrepancies observed between the 3 smoking contrasts suggest that the precise mechanism by which the smoking-induced disease risk of an individual is modified by their genetic makeup differs between past and current smokers. This disparity has not been considered in previous epidemiologic studies.[7,31–34] Second, we were able to show, for the first time, that some of the modification of smoking-induced IBD risk is brought about by more than one genetic factor located in the HLA region. Third, a clear-cut dependence upon smoking behavior became apparent in the HLA region for CD risk, but not UC risk. Such a differential role of G×E in the 2 IBD sub-entities may be a key to understanding why the smoking-induced risk for UC may still increase with time even decades after smoking cessation.[6] Finally, the scope and nature of gene−smoking interaction in IBD may be exemplary for other diseases. For example, in line with our own results, *BTNL2* and *HLA-DRB5* were recently identified as candidate interaction signals as well in a rheumatoid arthritis SNP-smoking interaction study.[35] Some 29 of the 64 unique interacting SNPs (45%) were found to lie in close vicinity (≤1Mb) to genes that were previously identified as being disease-associated in GWAS, including *IL10* (Table 2).[1,24] However, in view of the general lack of strong linkage disequilibrium between interacting and IBD-associated SNPs, we conclude that the respective association and interaction signals may highlight different genetic effects. Even if 2 functionally relevant variants lie in the same gene or functional unit, the ensuing disease risk can still be modified by smoking for 1 variant but not for the other. Along the same line, our focused analysis of the HLA region revealed that only a subset of the IBD-predisposing alleles[16] was found to interact with smoking as well. We also identified 7 SNPs that seemed to interact with smoking in opposite directions with regard to CD and UC. Because statistical interaction can be viewed from different angles, this difference can mean 1 of 2 things—either a genetic mechanism predisposing to 1 of the 2 sub-entities is rendered protective against the other in the presence of smoking, or the effect of smoking in relation to one sub-entity is reversed in comparison to the other by that mechanism. Simply put, a certain genotype may simultaneously render smoking a risk factor for CD and a protective factor against UC.

We used web-based computational tools for evaluating the potential functional consequences of the interacting SNPs. Only a few of the SNPs were found to have a known effect rendering a firm biological interpretation of the results difficult (Supplementary Tables 12–16). However, many of the interacting SNPs are located near or within genes that may be involved in, or interfere with, mucosal barrier function (eg, *NOD2, IRGM, CDH1*, and *GPSM3*[36]) or the adaptive immune response (eg, *IL2RA*, *CCL11*, *CCL8*, *MICB*, *IL10*, and the HLA region). Several SNPs in the HLA region were also found to interact with smoking in relation to either CD, UC, or IBD (Table 2), including 1 SNP (rs3129890) that had previously been found to be associated with a high risk of rheumatoid arthritis among smokers.[35] Moreover, we identified 6 HLA alleles with suggestive gene−smoking interaction (namely *HLA-DRB3*91:01, HLA-B*57, HLA-B*57:01, HLA-DQA1*02:01, HLA-DQB1*02:02*, and *HLA-DRB1*07:01*). Of these 6 alleles, 4 were identified previously to have a main effect[16] either on CD risk (*HLA-B*57:01, HLA-DQA1*02:01*, and *HLA-DRB1*07:01*) or on UC risk (*HLA-DRB1*07:01*, *HLA-DQA1*02:01*, and *HLA-DQB1*02:02*). The overlap between genetic interaction and main effect signals suggests that perturbation of the adaptive immune response may be one important mechanism by which smoking differentially confers risk to either CD or UC. Also of interest in this regard are the 7 SNPs that interact with smoking in opposite directions in CD and UC (Table 3), which included 1 SNP (rs176095) that has been found to be associated with atopic dermatitis[37] and asthma[38] before. This SNP is located on chromosome 6 near the *GPSM3* gene that regulates monocytes function and inhibits NLRP3-coupled inflammasome activation.[36] NLRP3 is a member of the NOD-like receptor family of intracellular sensors of danger signals, such as pathogen-associated molecular patterns, that controls IL1α response to cigarette smoke exposure in mice.[39]

Our epidemiologic results were exemplarily corroborated by the observation that *Nod2*- and *Il10*-deficient mice that were experimentally exposed to cigarette smoke had a greater risk of ileitis than similarly treated WT mice. One possible explanation for this difference could be that cigarette smoke exposure compromises the barrier function of the small intestine more effectively as a result of lower *NOD2* gene expression and a consequent reduction of chemokine and antimicrobial peptides secretion.[40] Likewise, long-term exposure to cigarette smoke decreases the number of Foxp3+ cells and the expression of IL-10 which, in combination, represses interferon-gamma production.[41]

Equally important is that a greater risk of colitis was observed in *Il10*-deficient mice that were exposed to cigarette smoke but not in WT and *Nod2*-deficient mice, suggesting a differential role of *Il10* on disease location in response to cigarette smoke.

We employed the powerful albeit rarely used case-only design that relies on 2 key assumptions, namely that the disease is sufficiently rare (ie, prevalence <5%) in the general population and that G and E are uncorrelated in the general population. Case-only studies offer a number of methodologic advantages compared to traditional case-control studies, including higher per-sample power and better exposure data quality.[18,42,43] The samples available to us provided 90% power for a small interaction effect (OR, 1.15) in a case-only analysis, but only 3% power in a case-control analysis. However, because the validity of the results of case-only analyses depends on the validity of the G−E independence assumption, the latter must be assessed empirically, for example, in control data from GWAS. This requirement cannot be obviated because many genetic variants are known to be associated with smoking behavior at the population level.[44] Indeed, in our study, 15,196 SNPs violated the G−E independence assumption with at least 1 of the 3 smoking contrasts considered (never vs ever, never vs current, or never vs former). Only one of these, a synonymous SNP (rs1051730) in the nicotinic receptor gene *CHRNA3* at 15q25, has been identified so far to be associated with smoking behavior.[45] For all other G−E associations notified in the controls of our study, the underlying reason remains unclear because none of the respective SNPs coincide with any previously identified association with smoking behavior.[44,45]

One limitation of our study is that the smoking data were abstracted locally from existing clinical and/or research records, which may have introduced some variability across centers in the way the data were initially recorded. To mitigate this, we focused on a clear-defined classification of smoking behavior as current (smoking within the past 3 months), ever (current or ex-smoker), or never as of the date of diagnosis (cases) or recruitment (controls). We also ascertained the year in which the subjects first started and finally stopped smoking, as applicable, and used this information in conjunction with the year of diagnosis to verify that each center was applying the smoking definition correctly. Because our study employed meta-analysis techniques to evaluate gene−smoking interactions at the center level, any remaining measurement error would have resulted only in a loss of power but not in an increased type 1 error rate. Moreover, the center-specific interaction estimates were not found to be related to the center-specific smoking rates for any of the interacting SNPs, an observation that reinforces our notion that differential assessment of smoking behavior was unlikely to affect the validity of our results. This notwithstanding, without accurate information on the actual number of cigarettes smoked (which would have been much more difficult to obtain), we were unable to account for potential dose-dependent effects of smoking.

In summary, our genome-wide study of G×E in IBD identified 64 SNPs with strong evidence for a complex modifying role in the smoking-related etiology of IBD. Functional studies in mice lend additional experimental support to these epidemiologic findings by highlighting a direct effect of *Il10* and *Nod2* on disease risk in response to smoking. Our study sheds new light on the role of smoking as an important component of IBD pathogenesis interacting with the genetic background of at-risk individuals.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at http://dx.doi.org/10.1053/j.gastro.2017.05.010.

References

1. **Liu JZ, Sommeren van S**, Huang H, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet 2015;47:979–986.
2. **Jostins L, Ripke S**, Weersma RK, et al. Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 2012;491:119–124.
3. **Franke A, McGovern DPB, Barrett JC**, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet 2010;42:1118–1125.
4. Andersen V, Vogel U. Systematic review: interactions between aspirin, and other nonsteroidal anti-inflammatory drugs, and polymorphisms in relation to colorectal cancer. Aliment Pharmacol Ther 2014;40:147–159.
5. Andersen V, Holst R, Vogel U. Systematic review: diet–gene interactions and the risk of colorectal cancer. Aliment Pharmacol Ther 2013;37:383–391.
6. Higuchi LM, Khalili H, Chan AT, et al. A prospective study of cigarette smoking and the risk of inflammatory bowel disease in women. Am J Gastroenterol 2012;107:1399–1406.
7. Cosnes J. Tobacco and IBD: relevance in the understanding of disease mechanisms and clinical practice. Best Pract Res Clin Gastroenterol 2004;18:481–496.
8. Parkes GC, Whelan K, Lindsay JO, et al. Smoking in inflammatory bowel disease: impact on disease course and insights into the aetiology of its effect. J Crohns Colitis 2014;8:717–725.
9. Biedermann L, Brülisauer K, Zeitz J, et al. Smoking cessation alters intestinal microbiota: insights from quantitative investigations on human fecal samples using FISH. Inflamm Bowel Dis 2014;20:1496–1501.
10. Benjamin JL, Hedin CRH, Koutsoumpas A, et al. Smokers with active Crohn's disease have a clinically relevant dysbiosis of the gastrointestinal microbiota. Inflamm Bowel Dis 2012;18:1092–1100.
11. Klareskog L, Stolt P, Lundberg K, et al. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to

autoantigens modified by citrullination. Arthritis Rheum 2006;54:38–46.
12. Wagner CA, Sokolove J, Lahey LJ, et al. Identification of anticitrullinated protein antibody reactivities in a subset of anti-CCP-negative rheumatoid arthritis: association with cigarette smoking and HLA-DRB1 "shared epitope" alleles. Ann Rheum Dis 2015;74:579–586.
13. Helbig KL, Nothnagel M, Hampe J, et al. A case-only study of gene–environment interaction between genetic susceptibility variants in NOD2 and cigarette smoking in Crohn's disease aetiology. BMC Med Genet 2012;13:14.
14. Fowler EV, Doecke J, Simms LA, et al. ATG16L1 T300A shows strong associations with disease subgroups in a large Australian IBD population: further support for significant disease heterogeneity. Am J Gastroenterol 2008; 103:2519–2526.
15. Doecke JD, Simms LA, Zhao ZZ, et al. Smoking behaviour modifies IL23r-associated disease risk in patients with Crohn's disease. J Gastroenterol Hepatol 2015; 30:299–307.
16. **Goyette P**, **Boucher G**, Mallon D, et al. High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. Nat Genet 2015;47:172–179.
17. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81:559–575.
18. Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. Stat Med 1994;13:153–162.
19. Yadav P, Freitag-Wolf S, Lieb W, et al. Allowing for population stratification in case-only studies of gene–environment interaction, using genomic control. Hum Genet 2015;134:1117–1125.
20. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 1995;57:289–300.
21. Gauderman WJ. Sample size requirements for matched case-control studies of gene-environment interaction. Stat Med 2002;21:35–50.
22. Yang Q, Khoury MJ, Flanders WD. Sample size requirements in case-only designs to detect gene–environment interaction. Am J Epidemiol 1997;146:713–720.
23. Kraft P, Yen YC, Stram DO, et al. Exploiting gene–environment interaction to detect genetic associations. Hum Hered 2007;63:111–119.
24. Ellinghaus D, Jostins L, Spain SL, et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nat Genet 2016;48:510–518.
25. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015;348:648–660.
26. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015;518:317–330.
27. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57–74.
28. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 2012;40:D930–D934.
29. Pers TH, Karjalainen JM, Chan Y, et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun 2015;6:5890.
30. **Fehrmann RSN**, **Karjalainen JM**, Krajewska M, et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. Nat Genet 2015;47:115–125.
31. Mahid SS, Minor KS, Soto RE, et al. Smoking and inflammatory bowel disease: a meta-analysis. Mayo Clin Proc 2006;81:1462–1471.
32. García Rodríguez LA, González-Pérez A, Johansson S, et al. Risk factors for inflammatory bowel disease in the general population. Aliment Pharmacol Ther 2005; 22:309–315.
33. Birrenbach T, Böcker U. Inflammatory bowel disease and smoking: a review of epidemiology, pathophysiology, and therapeutic implications. Inflamm Bowel Dis 2004; 10:848–859.
34. Thomas GA, Rhodes J, Green JT, et al. Role of smoking in inflammatory bowel disease: implications for therapy. Postgrad Med J 2000;76:273–279.
35. **Jiang X, Källberg H**, Chen Z, et al. An immunochip-based interaction study of contrasting interaction effects with smoking in ACPA-positive versus ACPA-negative rheumatoid arthritis. Rheumatology (Oxford) 2016;55:149–155.
36. Giguère PM, Gall BJ, Ezekwe EAD, et al. G Protein signaling modulator-3 inhibits the inflammasome activity of NLRP3. J Biol Chem 2014;289:33245–33257.
37. **Weidinger S**, **Willis-Owen SAG**, **Kamatani Y**, et al. A genome-wide association study of atopic dermatitis identifies loci with overlapping effects on asthma and psoriasis. Hum Mol Genet 2013;22:4841–4856.
38. Li L, Kabesch M, Bouzigon E, et al. Using eQTL weights to improve power for genome-wide association studies: a genetic study of childhood asthma. Front Genet 2013; 4:103.
39. Pauwels NS, Bracke KR, Dupont LL, et al. Role of IL-1α and the Nlrp3/caspase-1/IL-1β axis in cigarette smoke-induced pulmonary inflammation and COPD. Eur Respir J 2011;38:1019–1028.
40. Aldhous MC, Soo K, Stark LA, et al. Cigarette smoke extract (CSE) delays NOD2 expression and affects NOD2/RIPK2 interactions in intestinal epithelial cells. PLoS One 2011;6:e24715.
41. Duan M-C, Zhang J-Q, Liang Y, et al. Infiltration of IL-17-producing T cells and Treg cells in a mouse model of smoke-induced emphysema. Inflammation 2016; 39:1334–1344.
42. **Gatto NM**, **Campbell UB**, Rundle AG, et al. Further development of the case-only design for assessing gene–environment interaction: evaluation of and adjustment for bias. Int J Epidemiol 2004;33:1014–1024.

43. Yadav P, Freitag-Wolf S, Lieb W, et al. The role of linkage disequilibrium in case-only studies of gene–environment interactions. Hum Genet 2015;134:89–96.
44. **Wain LV**, **Shrine N**, Miller S, et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. Lancet Respir Med 2015;3:769–781.
45. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat Genet 2010;42:441–447.

Author names in bold designate shared co-first authorship.

Received September 16, 2016. Accepted May 8, 2017.

Reprint requests
Address requests for reprints to: Proff Vibeke Andersen, Laboratory Center, Hospital of Southern Jutland, Kresten Philipsens Vej 15, 6200 Aabenraa, Denmark. e-mail: fag.biblioteket@rsyd.dk.

Acknowledgments
The authors thank all of the patients for their participation in the local studies. The International IBD Genetics Consortium is gratefully acknowledged for providing access to the genotype and smoking data; a full list of members and affiliations of the International IBD Genetics Consortium is provide in the Supplementary Material.

Conflicts of interest
The authors disclose no conflicts.

Funding
This project was supported by the Deutsche Forschungsgemeinschaft through Excellence Cluster 306 "Inflammation at Interfaces," Research and Training Group 1743 "Genes, Environment and Inflammation," FOR 2107 and the PopGen Biobank. Additional support was provided by the German Federal Ministry of Education and Research within the framework of the e:Med research and funding concept (sysINFLAME grant 01ZX1306A). L. Philip Schumm was supported by IBD Genetics Consortium Data Coordinating Center grant (National Institute of Diabetes and Digestive and Kidney Diseases grant 5U01DK062429) and Vibeke Andersen by "Knud og Edith Eriksens Mindefond." The mice studies were supported by the Fondation pour la Recherche Médicale (grant number DEQ20130326475) to Mathias Chamaillard and by the Conseil Régional Nord-Pas-de-Calais (Smo'X project, grant number 13001437) to Philippe Gosset.