



## Research Paper

# The *NOD2*-Smoking Interaction in Crohn's Disease is likely Specific to the 1007 fs Mutation and may be Explained by Age at Diagnosis: A Meta-Analysis and Case-Only Study



M. Ellen Kuenzig<sup>a,b,d,i</sup>, Jeff Yim<sup>e,i</sup>, Stephanie Coward<sup>a,b,d,i</sup>, Bertus Eksteen<sup>a,i</sup>, Cynthia H. Seow<sup>a,b,i</sup>, Cheryl Barnabe<sup>a,b,d,i</sup>, Herman W. Barkema<sup>b,d,f,i</sup>, Mark S. Silverberg<sup>g,i</sup>, Peter L. Lakatos<sup>c,i</sup>, Paul L. Beck<sup>a,i</sup>, Richard Fedorak<sup>h,i</sup>, Levinus A. Dieleman<sup>h,i</sup>, Karen Madsen<sup>h,i</sup>, Remo Panaccione<sup>a,i</sup>, Subrata Ghosh<sup>a,f,i</sup>, Gilaad G. Kaplan<sup>a,b,d,f,i,\*</sup>

<sup>a</sup> Department of Medicine, University of Calgary, Calgary, Alberta, Canada

<sup>b</sup> Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada

<sup>c</sup> McGill University, Montreal General Hospital, Montreal, Quebec, Canada

<sup>d</sup> O'Brien Institute for Public Health, University of Calgary, Calgary, Alberta, Canada

<sup>e</sup> Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada

<sup>f</sup> Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, Canada

<sup>g</sup> Zane Cohen Centre for Digestive Diseases, Division of Gastroenterology, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada

<sup>h</sup> Department of Medicine, Division of Gastroenterology and CEGIR, University of Alberta, Edmonton, Canada

<sup>i</sup> Alberta Inflammatory Bowel Disease Consortium, University of Calgary, Calgary, Alberta, Canada

## ARTICLE INFO

## Article history:

Received 9 May 2017

Received in revised form 13 June 2017

Accepted 14 June 2017

Available online 16 June 2017

## Keywords:

Crohn's disease

*NOD2*

Cigarette smoking

Gene-environment interactions

Age

## ABSTRACT

**Background:** *NOD2* and smoking are risk factors for Crohn's disease. We meta-analyzed *NOD2*-smoking interactions in Crohn's disease (Phase 1), then explored the effect of age at diagnosis on *NOD2*-smoking interactions (Phase 2).

**Methods:** *Phase 1:* MEDLINE and EMBASE were searched for studies ( $n = 18$ ) providing data on *NOD2* and smoking in Crohn's disease. *NOD2*-smoking interactions were estimated using odds ratios (ORs) and 95% confidence intervals (CIs) calculated using random effects models. *Phase 2:* A case-only study compared the proportion of smokers and carriers of the 1007 fs variant across ages at diagnosis ( $\leq 16$ , 17–40,  $> 40$  years).

**Findings:** *Phase 1:* Having ever smoked was less common among carriers of the 1007 fs variant of *NOD2* (OR 0.74, 95%CI:0.66–0.83). There was no interaction between smoking and the G908R (OR 0.96, 95%CI:0.82–1.13) or the R702W variant (OR 0.89, 95%CI:0.76–1.05). *Phase 2:* The proportion of patients ( $n = 627$ ) carrying the 1007 fs variant decreased with age at diagnosis ( $\leq 16$  years: 15%; 17–40: 12%;  $> 40$ : 3%;  $p = 0.003$ ). Smoking was more common in older patients ( $\leq 16$  years: 4%; 17–40: 48%;  $> 40$ : 71%;  $p < 0.001$ ).

**Interpretation:** The negative *NOD2*-smoking interaction in Crohn's disease is specific to the 1007 fs variant. However, opposing rates of this variant and smoking across age at diagnosis may explain this negative interaction.

© 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Crohn's disease is a chronic inflammatory disease of the gastrointestinal tract believed to result from the interaction between genetic

and environmental factors, leading to an inappropriate immune response to intestinal microbes (Knights et al., 2013). However, identifying reproducible gene-environment interactions in patients with Crohn's disease remains elusive. Previous studies examining the effects of *NOD2* (nucleotide binding oligomerization domain containing 2) on Crohn's disease focus on three single nucleotide polymorphisms (SNPs): 1007fs, G908R, and R702W. Individuals with the 1007 fs variant have a 4-fold increased odds of developing Crohn's disease, while the G908R variant triples and the R702W variant doubles the odds of Crohn's disease (Economou et al., 2004). Active and ex-smokers both have an approximately 2-fold increase in the odds of developing Crohn's disease (Mahid et al., 2006).

**Abbreviations:** CI, confidence interval; *NOD2*, nucleotide binding oligomerization domain containing 2; OR, odds ratio; PRISMA, Preferred Reporting Items for Systematic Review and Meta-Analyses; SNP, single nucleotide polymorphism.

\* Corresponding author at: Department of Medicine, University of Calgary, 3280 Hospital Drive NW, 6D56, Calgary, AB T2N 4N1, Canada.

E-mail address: [gkaplan@ucalgary.ca](mailto:gkaplan@ucalgary.ca) (G.G. Kaplan).

Although *NOD2* and smoking both independently increase the risk of Crohn's disease, an interactive effect is not consistently demonstrated. Some studies demonstrated a negative interaction between *NOD2* and smoking (de Diego et al., 2006; Helbig et al., 2012) while others failed to demonstrate a significant interaction (Mardini et al., 2005; Mendoza et al., 2003). In part, this may be explained by age-specific effects of genetic and environmental factors. Genetic susceptibility may play a larger role in patients with early-onset Crohn's disease (i.e., diagnosis before age 16) compared to those diagnosed later in life (i.e., after age 40) (de Ridder et al., 2007). In contrast, patients with late-onset Crohn's disease may have a greater cumulative exposure to environmental factors (e.g., longer history of smoking) than those with early-onset Crohn's disease.

The aim of this study was to systematically review and summarize the current knowledge of *NOD2*-smoking interactions in Crohn's disease, including SNP-specific *NOD2*-smoking interactions, and carry out a case-only study to explore the role of age at diagnosis on the association between *NOD2* and smoking.

## 2. Materials and Methods

### 2.1. Phase 1: Systematic Review and Meta-Analysis of *NOD2*-Smoking Interactions

Because of small sample sizes, most prior studies on *NOD2*-smoking interactions pooled the three most common *NOD2* variants (1007fs, G908R, R702W), treating each SNP as an equivalent risk allele for the *NOD2*-smoking interactions. Consequently, we conducted a systematic review and meta-analysis to separately evaluate the interaction between SNP-specific *NOD2* variants and smoking in patients with Crohn's disease. The systematic review and meta-analysis was based on a pre-determined protocol (Supplementary Table 1) and is reported in accordance with the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) checklist (Moher et al., 2009).

#### 2.1.1. Study Identification and Selection.

MEDLINE (1946–November 2015) and EMBASE (1974–November 2015) were searched to identify observational studies that examined both cigarette smoking and *NOD2* variants in relation to any Crohn's disease-related outcome (i.e., disease onset, prognosis, or phenotype). The search strategy is provided in Supplementary Table 1. Studies were included if they provided data on the interaction between a *NOD2* variant and smoking status in patients with Crohn's disease. Only studies reporting on each SNP independently were included. Studies reporting on the independent effects of cigarette smoking and *NOD2* were included if data on the interaction between *NOD2* and having ever smoked could be obtained from study authors. Additionally, authors of studies combining all SNPs in a single analysis were contacted to provide SNP-specific data and/or additional data on smoking behaviors if not included in their original publication.

References of included studies and relevant review articles were hand searched. Conference proceedings from major gastroenterological meetings (e.g., Digestive Diseases Week 2009–2015, American College of Gastroenterology Annual Scientific Meeting 2010–2014, and Congress of the European Crohn's and Colitis Organization 2011–2014) are indexed in EMBASE and were reviewed during the primary database search. When multiple manuscripts provided data on the same cohort of patients, the study providing the most complete data was included in the meta-analysis; if comparable data was presented, the most recent manuscript was included. The search was not limited by language, geography, or date of publication.

All records identified from the database search were independently screened by two study authors (MEK and SC). The full-text of selected abstracts was reviewed by two study authors (MEK and JY). Disagreements were resolved in consultation with a third author (GGK).

#### 2.1.2. Data Extraction

Two authors (MEK and JY) independently extracted data from eligible studies using a piloted data extraction form, including: study design, quality, and location; variants of *NOD2*; definition of smoking and source of smoking information (questionnaire, chart review, or interview); source of cases; sample size; prevalence of smoking; *NOD2* allele frequency; and the interaction between each *NOD2* variant and smoking, either as contingency tables or odds ratios (ORs) and their corresponding 95% confidence intervals (CIs).

#### 2.1.3. Outcomes

Case-only studies are commonly used to study gene-environment interactions (Thomas, 2010). We used all cases from identified studies to examine *NOD2*-smoking interactions in patients with Crohn's disease. We examined the association between ever smoking (current and former smoking) among individuals with a *NOD2* variant as compared to patients who were wild type. The analysis was based on smoking status at the time of diagnosis or at study entry when smoking 'at diagnosis' was unavailable. Analyses were conducted for the three most common *NOD2* variants (1007fs, G908R, and R702W) independently.

The following a priori sensitivity analyses were conducted: (1) removing studies with unclear definitions of smoking; (2) limiting analyses to studies defining smoking at diagnosis; and (3) limiting to studies where all participants were non-Jewish and White.

#### 2.1.4. Study Quality

The quality of included studies was assessed using a modified version of the Newcastle–Ottawa Quality Assessment Scale that incorporated aspects of study quality relevant to case-only studies of gene-environment interactions (Supplementary Table 1) (Wells et al., 2017). This included the methods used to select cases and misclassification bias (e.g., ascertainment of smoking status). The risk of bias due to residual confounding could not be assessed because the assessment of *NOD2*-smoking interactions was not the primary objective of most included studies. Study quality was evaluated independently by two authors (MEK and JY). Disagreements were resolved by consensus and in consultation with a third author (GGK).

#### 2.1.5. Statistical Analysis

Data were analyzed using the meta package in R version 3.3.2 (Schwarzer, 2017; R Core Team, 2016). ORs and corresponding 95% CIs were calculated from contingency tables for each study based on cases of Crohn's disease only, then pooled using random effects models to account for expected heterogeneity across studies. The estimated ORs reflect the odds of being a smoker among individuals with a *NOD2* mutation relative to the odds of being a smoker among patients without a *NOD2* mutation.

Heterogeneity was assessed using the  $I^2$  measure and the Cochran Q-statistic;  $p < 0.10$  was considered statistically significant. Variance between studies ( $\tau^2$ ) was quantified using the DerSimonian-Laird estimator. Publication bias was assessed using a visual inspection of funnel plots. The Egger test was used to assess funnel plot asymmetry (Egger et al., 1997).

### 2.2. Phase 2: Effect of Age at Diagnosis on *NOD2*-Smoking Interaction

We evaluated the hypothesis that the negative interaction between the 1007 fs variant in *NOD2* and cigarette smoking may be explained by the opposing prevalence of these factors across different ages at diagnosis. Based on the results of the meta-analysis we only evaluated the interaction between the 1007 fs variant of *NOD2* and smoking.

We conducted a case-only study in a cohort of patients with Crohn's disease enrolled in the Alberta IBD Consortium. The Alberta IBD Consortium recruited patients from tertiary care centers in Calgary, Alberta ( $n = 324$ ) and Edmonton, Alberta, Canada ( $n = 303$ ) between 2007 and 2014. All genotyped patients with a completed

environmental questionnaire were included in the analysis. Medical chart reviews were completed on each patient with Crohn's disease using a standardized electronic data extraction form to confirm the diagnosis of Crohn's disease and establish the age at diagnosis. Diagnosis of Crohn's disease was confirmed using standard clinical, endoscopic, radiologic, and histologic criteria (Lennard-Jones, 1989). A subset of chart reviews ( $n = 20$ ) was completed in duplicate to ensure consistency across reviewers. Age at diagnosis was defined a priori according to the Montreal Classification: A1 ( $\leq 16$  years of age); A2 (17 to 40 years); and A3 ( $>40$  years) (Silverberg et al., 2005; Satsangi, 2006). In order to account for population stratification, only non-Jewish Caucasian patients were included in the analysis.

A standardized environmental questionnaire was used to assess smoking status of patients and included information on frequency, timing, quitting, and duration of cigarette use. A research coordinator facilitated the administration of the environmental questionnaire to ensure accuracy and completeness of data. Determination of smoking status was blinded to *NOD2* genotype. Ever smokers had smoked at least one cigarette per day for at least one year prior to being diagnosed and never smokers were lifelong non-smokers prior to diagnosis. Previous studies have shown retrospective recall of smoking habits have been demonstrated to be valid (sensitivity: 92%; specificity: 98%) (Wong et al., 2012).

DNA was extracted from venous blood of patients using the Qiagen (Germantown, MD, USA) DNAeasy kit. Samples were then sent to BGI-Shenzhen for processing. The 1007fs variant of *NOD2* was genotyped using a Goldengate platform (Illumina). DNA extraction and genotyping was conducted without knowledge of the patient's smoking status. Carriers of the 1007fs variant were defined as individuals who were either homozygous or heterozygous for this variant of the *NOD2* gene.

All statistical analyses were conducted using SAS Version 9.4 (Cary, NC, USA). We used the Cochran-Armitage test for trend to determine if the proportion of patients carrying the 1007fs variant and having ever smoked changed with age at diagnosis, as well as to test for differences in the city of residence and sex of patients across age groups. Age groups were determined a priori and based on the Montreal Classification (Silverberg et al., 2005; Satsangi, 2006).

To ensure the interaction between the 1007fs variant and smoking was consistent with the findings of the meta-analysis, we calculated the OR and 95% CI to estimate this interaction. This odds ratio compared the odds of carrying the 1007fs SNP among ever smokers to the odds of carrying the 1007 fs SNP (homozygotes for the risk allele and heterozygotes) among never smokers.

This study was approved by the Conjoint Research Ethics Boards at the University of Calgary and the University of Alberta. Signed informed consent was obtained from all study participants.

### 3. Results

#### 3.1. Phase 1: Systematic Review and Meta-Analysis of *NOD2*-Smoking Interaction

The literature search yielded 248 records; 170 remained after removing duplicates. An additional 53 studies were identified after reviewing references of included studies and relevant review articles. Four studies were eligible for inclusion (Supplementary Fig. 1) (de Diego et al., 2006; Helbig et al., 2012; Mendoza et al., 2003; Mardini et al., 2005). An additional 14 studies were included after obtaining SNP-specific data from study authors (Ananthkrishnan et al., 2014; Cleynen et al., 2013; Doecke et al., 2015; Giachino et al., 2004; Henckaerts et al., 2009; Karban et al., 2011; Laghi et al., 2005; Lakatos et al., 2005; Latiano, 2008; Nagy et al., 2005; Protic et al., 2008; Renda et al., 2008; Sabate et al., 2008; Walker et al., 2004). One study providing SNP-specific data was excluded from the analysis due to small cell sizes (Lauriola et al., 2011). In total, 18 studies (9064 patients with Crohn's disease) were included in the meta-analysis. Reasons for study

exclusion are outlined in Supplementary Table 2. The characteristics of included studies are provided in Table 1. The risk of bias in included studies is provided in Supplementary Table 3.

The 1007fs variant negatively interacted with having ever smoked cigarettes (pooled OR 0.74, 95%CI 0.66 to 0.83, 18 studies; heterogeneity:  $I^2 = 0\%$ ,  $\tau^2 = 0$ ,  $p = 0.49$ ; Fig. 1). Neither the G908R variant (pooled OR 0.96, 95%CI 0.82 to 1.13, 17 studies; heterogeneity:  $I^2 = 0\%$ ,  $\tau^2 = 0$ ,  $p = 0.54$ ; Fig. 2) nor the R702W variant (OR 0.89, 95%CI 0.76 to 1.05, 18 studies; heterogeneity:  $I^2 = 31\%$ ,  $\tau^2 = 0.03$ ,  $p = 0.10$ ; Fig. 3) interacted with cigarette smoking. Results were consistent in sensitivity analyses excluding studies with unclear definitions of smoking (Supplementary Table 4) and when limiting analyses to studies where smoking status was defined at diagnosis (Supplementary Table 5) and all participants were non-Jewish and White (Supplementary Table 6).

#### 3.1.1. Publication Bias

There was no evidence of publication bias for the interaction between *NOD2* and smoking, regardless of the SNP being analyzed (1007fs:  $p = 0.70$ ; G908R:  $p = 0.22$ ; R702W:  $p = 0.75$ ) (Supplementary Fig. 2A–C).

#### 3.2. Phase 2: Effect of Age at Diagnosis on *NOD2*-Smoking Interaction

We enrolled 627 patients with Crohn's disease who provided data on smoking status at diagnosis, 1007 fs *NOD2* genotype, and age at diagnosis (Table 2). The proportion of patients having ever smoked significantly increased with increasing age at diagnosis ( $p < 0.001$  for trend). Only 4% (4/91) of patients diagnosed under the age of 17 had ever smoked, compared to 48% (203/426) of those diagnosed between 17 and 40 and 71% (78/110) of patients diagnosed after 40 years of age (Table 2).

The proportion of patients carrying the 1007fs variant decreased significantly with increasing age at diagnosis ( $p = 0.003$  for trend). Fifteen percent (14/91) of patients diagnosed before age 17 and 12% of patients diagnosed between 17 and 40 were carriers of this variant, whereas 3% of individuals diagnosed after 40 were carriers (3/110) (Table 2).

The odds of carrying a 1007fs variant were significantly lower among individuals who had ever smoked cigarettes compared to individuals who had never smoked (OR 0.49, 95% CI 0.27 to 0.86).

### 4. Discussion

We systematically reviewed the interaction between *NOD2* and cigarette smoking in Crohn's disease by analyzing 18 studies with  $>9000$  patients, including previously unpublished data from 14 studies. Our pooled analyses demonstrated that the negative association between *NOD2* and cigarette smoking was specific to the 1007 fs variant. Moreover, our case-only study showed that the prevalence of both the 1007fs mutation and cigarette smoking vary across ages at diagnosis such that with advancing age, the prevalence of the 1007 fs variant decreases and exposure to cigarette smoke increases. These findings may explain inconsistencies between studies and the elusiveness of identifying reproducible gene-environment interactions.

Our systematic review demonstrated that the majority of *NOD2*-smoking studies were underpowered to detect an interaction; for example, 10 identified studies included fewer than 300 participants (de Diego et al., 2006; Giachino et al., 2004; Laghi et al., 2005; Mardini et al., 2005; Mendoza et al., 2003; Protic et al., 2008; Renda et al., 2008; Sabate et al., 2008; Walker et al., 2004; Nagy et al., 2005). Because of these small samples, the three most common *NOD2* variants (1007fs, G908R, and R702W) were often combined. However, this meta-analysis confirms that the *NOD2*-smoking interaction in Crohn's disease is SNP-specific. Thus, combining SNPs to improve power may not be methodologically sound.

The biological plausibility of a negative association between smoking and the 1007 fs variant of *NOD2* is unclear. However, the

**Table 1**  
Characteristics of included studies.

Study	Country	Age at diagnosis; mean (sd), years	<i>NOD2</i> genotype frequencies <sup>a</sup>	Smoking status <sup>b</sup>	Timing of smoking	Sample size
Ananthakrishnan et al. (2014)	USA	Median (Q1,Q3): 24 (18, 33)	Homozygous: 7% Heterozygous: 26% Wild type: 67%	Current: 8% Former: 30% Never: 62%	At time of study	697 <sup>c</sup>
Cleynen et al. (2013)	Europe (7 countries)	26 (11)	Homozygous: 10% Heterozygous: 29% Wild type: 62%	Current: 21% Former: 27% Never: 52%	At time of study	1528 <sup>c</sup>
de Diego et al. (2006)	Spain	Ever smokers: 33(12) Never smokers: 34 (16) Range: 12 to 77 years	Carriers: 28% Wild type: 72%	Ever: 70% Never: 30%	At diagnosis	178
Doecke et al. (2015)	Australia and New Zealand	<16: 10% 16–40: 77% >40: 13%	Homozygous: 9% Heterozygous: 24% Wild type: 67%	Current: 44% Former: 46% Never: 10%	At diagnosis	675 <sup>d</sup>
Giachino et al. (2004)	Italy	<40: 77% >40: 23%	Carrier: 33% Wild type: 67%	Current: 35% Former: 15% Never: 50%	At time of study	184 <sup>c</sup>
Helbig et al. (2012)	Germany	26 (10)	Carrier: 46% Wild type: 54%	Current (diagnosis): 41% Non-current (diagnosis): 59% Ever (study): 58% Never (study): 42%	At diagnosis (N = 1283) At time of study (N = 1636)	1636
Henckaerts et al. (2009)	Belgium	Median (Q1, Q3): 24 (19, 32) ≤16: 18% 17–40: 71% >40: 11%	Carriers: 45% Wild type: 55%	Current: 64% Former: 8% Never: 27%	At diagnosis	755 <sup>c,e</sup>
Karban et al. (2011)	Israel	25 (12)	Homozygous: 6% Heterozygous: 26% Wild type: 68%	Current: 24% Former: 11% Never: 65%	At time of study	453 <sup>f</sup>
Laghi et al. (2005)	Italy	35 (13)	Carrier: 33% Wild type: 67%	Ever: 53% Never: 47%	At diagnosis	239
Lakatos et al. (2005)	Hungary	37 (9)	Homozygous: 10% Heterozygous: 25% Wild type: 65%	Current: 32% Former: 9% Never: 58%	At time of study	527
Latiano (2008)	Italy	29 (15)	Homozygous: 8% Heterozygous: 26% Wild type: 66%	Current: 33% Former: 12% Never: 55%	At time of study	763 <sup>c</sup>

(continued on next page)

Table 1 (continued)

Study	Country	Age at diagnosis; mean (sd), years	<i>NOD2</i> genotype frequencies <sup>a</sup>	Smoking status <sup>b</sup>	Timing of smoking	Sample size
Mardini et al. (2005)	USA	Median (range): 22 (11 to 41) years	G908R Heterozygous: 8% G908R Wild type: 92% 1007fs Homozygous: 2% 1007fs Heterozygous: 12% 1007fs Wild type: 86% R702W Homozygous: 1% R702W Heterozygous: 17% R702W Wild type: 82%	Ever: 54% Never: 46%	Unclear	202
Mendoza et al. (2003)	Spain	<40 years: 78% ≥40 years: 22%	Carrier: 33% Wild type: 67%	Smokers: 54% Non-smokers <sup>g</sup> : 46%	Unclear	204
Nagy et al. (2005)	Hungary	<i>NOD2</i> carrier: 30 <i>NOD2</i> wild type: 33 <sup>h</sup>	Homozygous: 4% Heterozygous: 30% Wild type: 66%	Current: 50% Former: 38% Never: 12%	At time of study	74
Protic et al. (2008)	Serbia	Median (range): 26 (6, 59) ≤16 years: 12% 17–40 years: 72% >40 years: 16%	Homozygous: 2% Heterozygous: 34% Wild type: 65%	Current: 22% Former: 1% Never: 77%	At time of study	131
Renda et al. (2008)	Italy	30 (12)	Homozygous: 6% Heterozygous: 24% Wild type: 69%	Current: 37% Former: 14% Never: 49%	At diagnosis	182 <sup>c</sup>
Sabate et al. (2008)	France	40 (14)	Homozygous: 5% Heterozygous: 30% Wild type: 64%	Current: 42% Former: 6% Never: 52%	At diagnosis	239 <sup>c</sup>
Walker et al. (2004)	Scotland	28 (14) Range: 5 to 76	Homozygous: 2% Heterozygous: 22% Wild type: 76%	Current: 22% Former: 47% Never: 30%	At time of study	228 <sup>c</sup>

<sup>a</sup> Patients homozygous for a *NOD2* mutation, either carrying two copies of the same mutation or one copy of a mutation and a copy of a different mutation (i.e., compound heterozygotes). Heterozygotes had one copy of a risk allele and one copy of the wild type allele. Wild type individuals had two copies of the normal allele. Carriers had at least one copy of the risk allele (i.e., homozygous or heterozygous). Genotype frequencies are expressed for all SNPs in the *NOD2* gene unless otherwise noted.

<sup>b</sup> Current smokers are patients that are smoking cigarettes at present. Former smokers had smoked in the past, but were no longer smoking at the time that smoking status was determined. Never smokers were lifetime non-smokers. Ever smokers were patients that had smoked at some point (i.e., includes current and former smokers).

<sup>c</sup> *NOD2*-smoking data available on a subset of patients included in the study.

<sup>d</sup> Data obtained from study authors corresponded to the cohort from Brisbane, Australia (dataset 1). There were 675 patients in this dataset.

<sup>e</sup> Study was limited to patients without perianal disease at diagnosis.

<sup>f</sup> Data obtained from study authors corresponded to an updated (larger) cohort of patients.

<sup>g</sup> Smoking defined in the manuscript as 'smokers' and 'non-smokers' and was unclear if 'smokers' were ever smokers (i.e., current and former smokers) or current smokers only. Study was excluded in sensitivity analysis.

<sup>h</sup> Unclear if the age of study participants was presented as a mean or median.



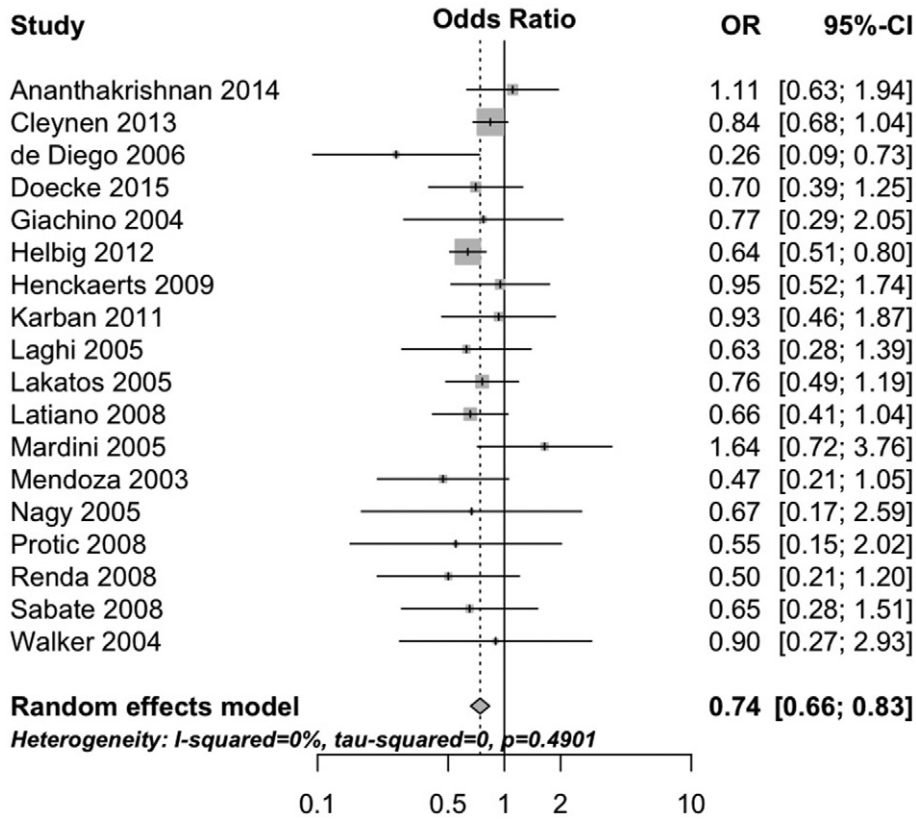


Fig. 1. Forest plot depicting the interaction between 1007fs variant of *NOD2* and cigarette smoking in patients with Crohn's disease.

complex interplay between the different *NOD2* variants and cigarette smoking, and their effects on dendritic cells may underlie the negative interaction observed between smoking and the 1007 fs variant but not

the *NOD2* variants (Butler et al., 2007; van Heel et al., 2005; Ueno et al., 2014; Kramer et al., 2006). The 1007fs variant of *NOD2* results in a protein that is truncated in the region of *NOD2* responsible for

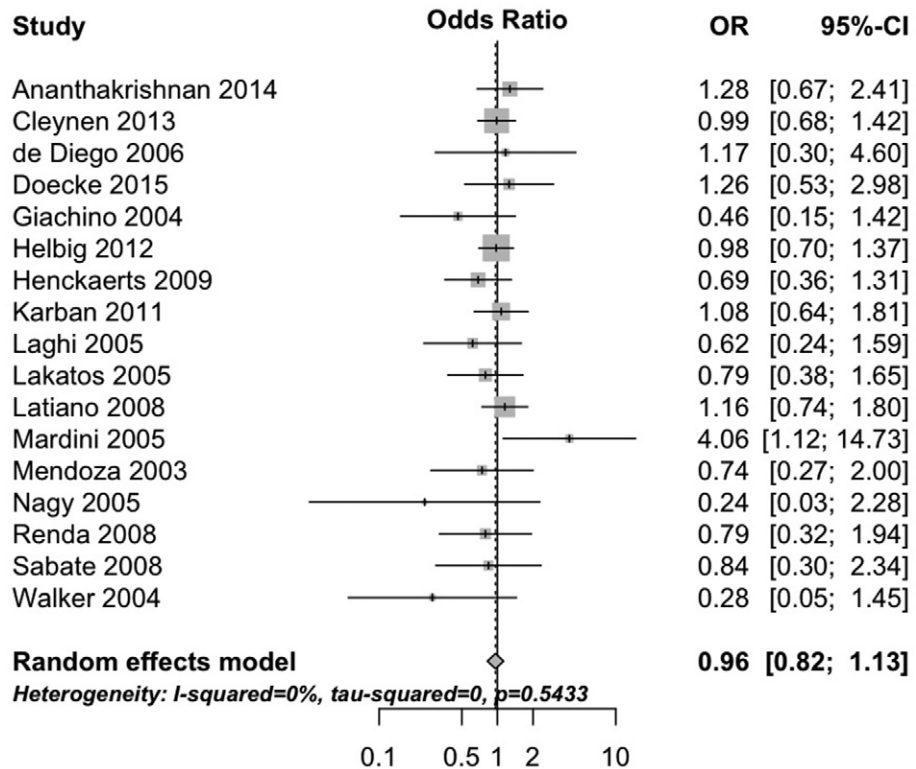


Fig. 2. Forest plot depicting the interaction between G908R variant of *NOD2* and cigarette smoking in patients with Crohn's disease.

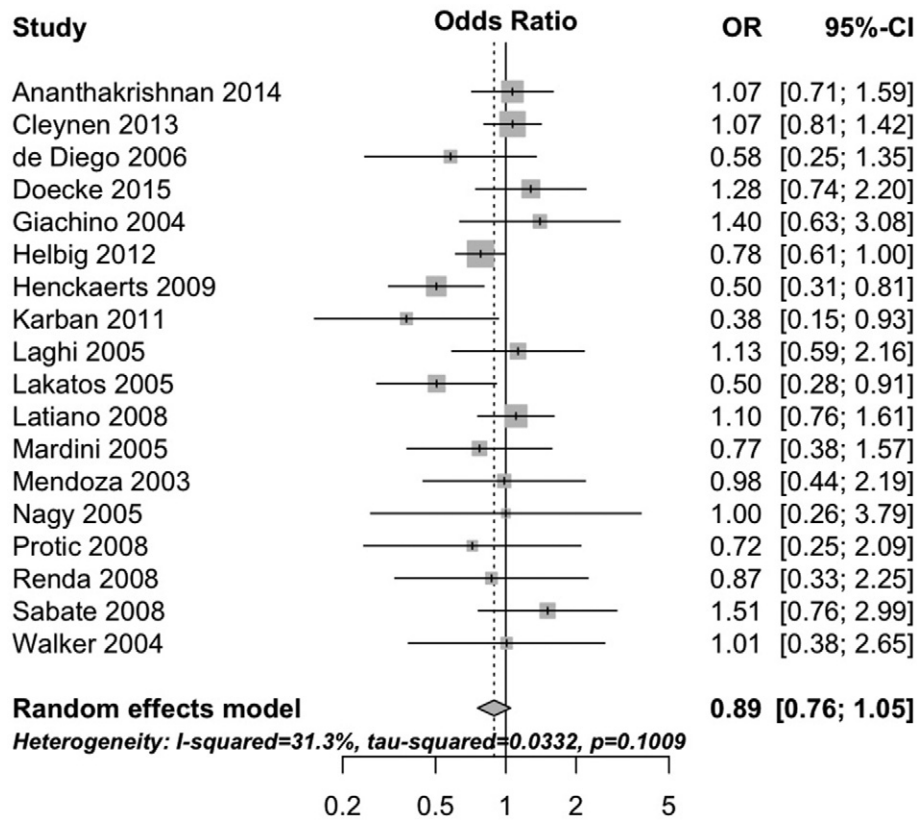


Fig. 3. Forest plot depicting the interaction between R702W variant of *NOD2* and cigarette smoking in patients with Crohn's disease.

intracellular detection of peptidoglycans on bacterial cell walls. The G908R and R702W variants are single base substitutions that may also alter the structure of *NOD2*. However, the functional impacts of each variant are not known.

Alternatively, the negative relationship between *NOD2* and smoking may be explained by a methodological phenomenon. Our case-only study examined the impact of age of onset on the interaction between *NOD2* and cigarette smoking in patients with Crohn's disease. Twenty percent of *NOD2* carriers were diagnosed prior to age 17, whereas only 3% were diagnosed after age 40. In contrast, nearly three-quarters of patients over the age of 40 at diagnosis had a history of smoking compared to only 4% of those diagnosed under age 17. This inverse relationship between *NOD2* and smoking, across ages at diagnosis results in minimal overlap of the two factors and may misleadingly appear as a

gene-environment interaction. However, the negative interaction between *NOD2* and smoking persists after adjusting for age at diagnosis (Helbig et al., 2012). Supplementary Fig. 3 explains the bias introduced by combining patients with varying ages of diagnosis, resulting in a negative interaction.

Consistent with prior studies, we analyzed the *NOD2*-smoking interaction in Crohn's disease using a case-only study design (Helbig et al., 2012). The case-only design increases resource and statistical efficiency to detect gene-environment interactions (Thomas, 2010). However, there are some assumptions of case-only designs including: (1) the absence of population stratification; and (2) the independence of the genetic and environmental risk factor in the source population for the cases (Liu et al., 2012; Pierce and Ahsan, 2010; Wang and Lee, 2008). Bias due to population stratification (i.e., confounding by ethnicity) arises when studies combine different ethnic groups with variable allele frequencies and prevalence of exposure (Wang and Lee, 2008). Sensitivity analyses limited to studies where all patients were non-Jewish and White were consistent with our main analyses. However, we were not able to assess the interaction between *NOD2* and smoking in other ethnicities (e.g., African Americans) due to the paucity of studies. Secondly, we assumed gene-environment independence due to a prior study that found no association between *NOD2* and smoking in their control population (Helbig et al., 2012) and *NOD2* has not been identified in genome-wide association studies of smoking behaviors (Furberg et al., 2010).

The results of this study point to the importance of evaluating genetic and environmental factors in sub-phenotypes of Crohn's disease. We focused on age at diagnosis. However, the findings of *NOD2*-smoking interaction studies may be influenced by other disease phenotypes such as disease location or disease behavior. Both *NOD2* and smoking are associated with ileal Crohn's disease. In addition, ileal and colonic Crohn's disease have been shown to be genetically distinct phenotypes (Cleynen et al., 2016). Thus, our study provides one example of how phenotypic characteristics may influence gene-environment interactions.

Table 2  
Characteristics of included patients.

Characteristic	Total	Age at diagnosis			p-value
		A1 (≤16)	A2 (17–40)	A3 (>40)	
Total (N)	627	91	426	110	
City					
Calgary	324 (52%)	45 (49%)	218 (51%)	61 (55%)	0.87
Edmonton	303 (48%)	46 (51%)	208 (49%)	49 (45%)	
Sex					
Male	260 (41%)	50 (55%)	155 (36%)	55 (50%)	0.68
Female	367 (59%)	41 (45%)	271 (64%)	55 (50%)	
<i>NOD2</i> (1007fs), N (%)					
Wild type	558 (89%)	77 (85%)	374 (88%)	107 (97%)	0.003
Carrier <sup>a</sup>	69 (11%)	14 (15%)	52 (12%)	3 (3%)	
Smoking status, N (%)					
Ever	285 (46%)	4 (4%)	203 (48%)	78 (71%)	<0.0001
Never	342 (54%)	87 (96%)	223 (52%)	32 (29%)	

Abbreviations: Q1: 1st quartile (25th percentile); Q3: 3rd quartile (75th percentile).

<sup>a</sup> Carriers of the 1007 fs variant (heterozygous or homozygous) were compared to wild type.

Our study provides the most comprehensive review of *NOD2*-smoking interactions in Crohn's disease, includes previously unpublished data, and evaluates the impact of age at diagnosis on this interaction. Nonetheless, limitations should be considered. Firstly, the association between *NOD2* and smoking was not adjusted for potential confounders (e.g., age, family history). Also, the quality of the meta-analysis was dependent on the quality of the individual studies that were included. With regard to the case-only study we were limited by sample size, which impeded our ability to calculate age-specific ORs for the interaction between *NOD2* and smoking and to investigate the role of sub-phenotypes of Crohn's disease (i.e., disease location and behavior).

Gene-environment interactions for Crohn's disease are difficult to identify. Our meta-analysis demonstrated that gene-environment interaction studies need to be designed to assess SNP-specific effects, such that only 100 fs variant of *NOD2* interacted with smoking. Moreover, our case-only study showed that the negative interaction between the 1007fs variant of *NOD2* and smoking may be influenced by the contrasting prevalence of *NOD2* and smoking across ages at diagnosis with Crohn's disease. This is one example of how gene-environment interactions may depend on sub-phenotypes of Crohn's disease. Thus, we suggest that future gene-environment studies should be powered to assess SNP-specific interactions and be designed to evaluate interactions in specific phenotypes of patients with Crohn's disease.

#### Funding Sources

This project was funded by Alberta Innovates - Health Solutions and a CIHR Team Grant (Health Challenges in Chronic Inflammation). Funders had no role in the study design, data analysis, and interpretation of the study results.

#### Conflict of Interest

The study authors have no competing interests to declare.

#### Author Contributions

**Guarantor of the Article:** Gilaad G. Kaplan, MD MPH.

**Specific Author Contributions:** Study concept and design: MEK, BE, CHS, CB, GGK. Data acquisition: MEK, JY, SC, PLL, GGK. Interpretation of the data: MEK, HWB, BE, CHS, CB, RP, SG, MSS, PLL, GGK. Statistical analysis: MEK, GGK. Drafting of the manuscript: MEK, BE, CHS, CB, GGK. Critical revision of the manuscript for important intellectual content: HWB, RP, SG, MSS, PLL, JY, SC, PLB, RF, LAD, KM. Final approval of the manuscript: MEK, JY, SC, BE, CHS, CB, HWB, MSS, PLL, PLB, RF, LAD, KM, RP, SG, GGK.

#### Acknowledgements

The study authors would like to thank the authors of the following studies for providing additional data: Ananthakrishnan et al. (2014) (Ashwin Ananthakrishnan); Cleynen et al. (2013) (Isabelle Cleynen); Cucchiara et al. (2007) (Anna Latiano, Vito Annese); Doecke et al. (2015) (James Doecke, Graham Radford-Smith); Fowler et al. (2014) (Sharyle Fowler, Vijay Yajnik); Giachino et al. (2004) (Mario De Marchi); Henckaerts et al. (2009) (Isabelle Cleynen, Liesbet Henckaerts); Karban et al. (2011) (Amir Karban); Laghi et al. (2005) (Alberto Malesci); Lakatos et al. (2005) (Peter Lakatos); Latiano (2008) (Anna Latiano, Vito Annese); Lauriola et al. (2011) (Mattia Lauriola); Maconi et al. (2009) (Giovanni Maconi); Protic et al. (2008) (Marijana Protic); Nagy et al. (2005) (Zsuzsanna Nagy); Renda et al. (2008) (Mario Cottone); Walker et al. (2004) (Marian Aldhous, Hazel E. Drummond, Jack Satsangi). We acknowledge the support of the Alberta IBD Consortium with data derived from patient registries from the Intestinal Inflammation Tissue Bank (IITB) at the University of

Calgary and the Center of Excellence for Gastrointestinal Inflammation and Immunity Research (CEGIIR) at the University of Alberta. Gil Kaplan holds a CIHR Embedded Clinician Research Award and an Alberta Innovates Population Health Award.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.ebiom.2017.06.012>.

#### References

- Ananthakrishnan, A.N., et al., 2014. Differential effect of genetic burden on disease phenotypes in Crohn's disease and ulcerative colitis: analysis of a North American cohort. *Am. J. Gastroenterol.* 109 (3), 395–400.
- Butler, M., et al., 2007. *NOD2* activity modulates the phenotype of LPS-stimulated dendritic cells to promote the development of T-helper type 2-like lymphocytes – possible implications for *NOD2*-associated Crohn's disease. *J. Crohn's Colitis* 1 (2), 106–115.
- Cleynen, I., et al., 2013. Genetic factors conferring an increased susceptibility to develop Crohn's disease also influence disease phenotype: results from the IBDchip European Project. *Gut* 62 (11), 1556–1565.
- Cleynen, I., et al., 2016. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *Lancet* 387 (10014), 156–167.
- Cucchiara, S., et al., 2007. Role of *CARD15*, *DLG5* and *OCTN* genes polymorphisms in children with inflammatory bowel diseases. *World J. Gastroenterol.* 13 (8), 1221–1229.
- de Diego, C., et al., 2006. Influence of smoking habits and *CARD15* mutations on the onset of Crohn's disease. *Scand. J. Gastroenterol.* 41 (10), 1209–1211.
- de Ridder, L., et al., 2007. Genetic susceptibility has a more important role in pediatric-onset Crohn's disease than in adult-onset Crohn's disease. *Inflamm. Bowel Dis.* 13 (9), 1083–1092.
- Doecke, J.D., et al., 2015. Smoking behaviour modifies *IL23r*-associated disease risk in patients with Crohn's disease. *J. Gastroenterol. Hepatol.* 30 (2), 299–307.
- Economou, M., et al., 2004. Differential effects of *NOD2* variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis. *Am. J. Gastroenterol.* 99 (12), 2393–2404.
- Egger, M., et al., 1997. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315 (7109), 629–634.
- Fowler, S.A., et al., 2014. *SMAD3* gene variant is a risk factor for recurrent surgery in patients with Crohn's disease. *J. Crohn's Colitis* 8 (8), 845–851.
- Furberg, H., et al., 2010. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* 42 (5), 441–447.
- Giachino, D., et al., 2004. Analysis of the *CARD15* variants *R702W*, *G908R* and *L1007fs* in Italian IBD patients. *Eur. J. Hum. Genet.* 12 (3), 206–212.
- Helbig, K.L., et al., 2012. 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.* 13 (1), 14.
- Henckaerts, L., et al., 2009. Genetic risk profiling and prediction of disease course in Crohn's disease patients. *Clin. Gastroenterol. Hepatol.* 7 (9), 972–980.e2.
- Karban, A., et al., 2011. Non-Jewish Israeli IBD patients have significantly higher Glutathione S-Transferase *GSTT1*-null frequency. *Dig. Dis. Sci.* 56 (7), 2081–2087.
- Knights, D., Lassen, K.G., Xavier, R.J., 2013. Advances in inflammatory bowel disease pathogenesis: linking host genetics and the microbiome. *Gut* 62 (10), 1505–1510.
- Kramer, M., et al., 2006. Impaired dendritic cell function in Crohn's disease patients with *NOD2* 3020insC mutation. *J. Leukoc. Biol.* 79 (4), 860–866.
- Laghi, L., et al., 2005. Carriage of *CARD15* variants and smoking as risk factors for resective surgery in patients with Crohn's ileal disease. *Aliment. Pharmacol. Ther.* 22 (6), 557–564.
- Lakatos, P.L., et al., 2005. Toll-like receptor 4 and *NOD2/CARD15* mutations in Hungarian patients with Crohn's disease: phenotype-genotype correlations. *World J. Gastroenterol.* 11 (10), 1489–1495.
- Latiano, A., 2008. Replication of interleukin 23 receptor and autophagy-related 16-like 1 association in adult- and pediatric-onset inflammatory bowel disease in Italy. *World J. Gastroenterol.* 14 (29), 4643–4649.
- Lauriola, M., et al., 2011. *IL23R*, *NOD2/CARD15*, *ATG16L1* and *PHOX2B* polymorphisms in a group of patients with Crohn's disease and correlation with sub-phenotypes. *Int. J. Mol. Med.* 27 (3), 469–477.
- Lennard-Jones, J.E., 1989. Classification of inflammatory bowel disease. *Scand. J. Gastroenterol.* 24 (s170), 2–6.
- Liu, C.-Y., et al., 2012. Design and analysis issues in gene and environment studies. *Environ. Health* 11 (1), 93.
- Maconi, G., et al., 2009. *CARD15* gene variants and risk of reoperation in Crohn's disease patients. *Am. J. Gastroenterol.* 104 (10), 2483–2491.
- Mahid, S.S., et al., 2006. Smoking and inflammatory bowel disease: a meta-analysis. *Mayo Clin. Proc.* 81 (11), 1462–1471.
- Mardini, H.E., et al., 2005. Gastrointestinal Crohn's disease is associated with *NOD2/CARD15* gene polymorphisms, particularly *L1007P* homozygosity. *Dig. Dis. Sci.* 50 (12), 2316–2322.
- Mendoza, J.L., et al., 2003. Prevalence of mutations of the *NOD2/CARD15* gene and relation to phenotype in Spanish patients with Crohn disease. *Scand. J. Gastroenterol.* 38 (12), 1235–1240.
- Moher, D., et al., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 6 (7) e1000097-6.



- Nagy, Z., et al., 2005. Crohn's disease is associated with polymorphism of CARD15/NOD2 gene in a Hungarian population. *Ann. N. Y. Acad. Sci.* 1051 (1), 45–51.
- Pierce, B.L., Ahsan, H., 2010. Case-only genome-wide interaction study of disease risk, prognosis and treatment. *Genet. Epidemiol.* 34 (1), 7–15.
- Protic, M.B., et al., 2008. CARD15 gene polymorphisms in Serbian patients with Crohn's disease: genotype–phenotype analysis. *Eur. J. Gastroenterol. Hepatol.* 20 (10), 978–984.
- R. Core Team, 2016. R: A Language and Environment for Statistical Computing Available at: <http://www.R-project.org>.
- Renda, M.C., et al., 2008. The role of CARD15 mutations and smoking in the course of Crohn's disease in a Mediterranean area. *Am. J. Gastroenterol.* 103 (3), 649–655.
- Sabate, J.M., et al., 2008. The V249I polymorphism of the CX3CR1 gene is associated with fibrostenotic disease behavior in patients with Crohn's disease. *Eur. J. Gastroenterol. Hepatol.* 20 (8), 748–755.
- Satsangi, J., 2006. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut* 55 (6), 749–753.
- Schwarzer, G., 2017. meta: General Package for Meta-Analysis. R package version 4.1–0. Available at: <http://CRAN.R-project.org/package=meta>.
- Silverberg, M.S., et al., 2005. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a working party of the 2005 Montreal World Congress of Gastroenterology. *Can. J. Gastroenterol.* 19 (Suppl A), 5A–36A.
- Thomas, D., 2010. Methods for investigating gene-environment interactions in candidate pathway and genome-wide association studies. *Annu. Rev. Public Health* 31 (1), 21–36.
- Ueno, A., et al., 2014. Opposing effects of smoking in ulcerative Colitis and Crohn's disease may be explained by differential effects on dendritic cells. *Inflamm. Bowel Dis.* 20 (5), 800–810.
- van Heel, D.A., et al., 2005. Muramyl dipeptide and toll-like receptor sensitivity in NOD2-associated Crohn's disease. *Lancet* 365 (9473), 1794–1796.
- Walker, L.J., et al., 2004. Anti-*Saccharomyces cerevisiae* antibodies (ASCA) in Crohn's disease are associated with disease severity but not NOD2/CARD15 mutations. *Clin. Exp. Immunol.* 135 (3), 490–496.
- Wang, L.Y., Lee, W.C., 2008. Population stratification bias in the case-only study for gene-environment interactions. *Am. J. Epidemiol.* 168 (2), 197–201.
- Wells, G.A., et al., 2017. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. *ohri.ca*. Available at: [http://www.ohri.ca/programs/clinical\\_epidemiology/oxford.asp](http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp) (Accessed January 30, 2017).
- Wong, S.L., et al., 2012. Assessment of validity of self-reported smoking status. *Health Rep.* 23 (1), 47–53.