

RESEARCH

Open Access

# Clinical predictors of inflammatory bowel disease in a genetically well-defined Caucasian population

Ziad M Kanaan<sup>1</sup>, Maurice R Eichenberger<sup>1</sup>, Surriya Ahmad<sup>1</sup>, Clayton Weller<sup>1</sup>, Henry Roberts<sup>1</sup>, Jianmin Pan<sup>2</sup>, Shesh N Rai<sup>2</sup>, Robert Petras<sup>3</sup>, E Brooks Weller Jr<sup>1</sup> and Susan Galandiuk<sup>1\*</sup>

## Abstract

**Background:** Crohn's disease (CD) and ulcerative colitis (UC), the two main types of inflammatory bowel disease (IBD), are multifactorial conditions of unknown etiology. The objective of this study is to examine the combined gene-environment interactions influencing IBD susceptibility in a well-defined Caucasian cohort in rural mid-America.

**Methods:** Patients were diagnosed to have CD or UC using conventional radiologic, endoscopic, and/or histopathologic findings. Histological diagnosis was made by a single specialist gastrointestinal pathologist with a particular interest in IBD. Information regarding cigarette smoke exposure was obtained by administration of the Behavioral Risk Factor Surveillance System Survey (BRFSS) to all patients. Genomic DNA was extracted from peripheral blood leukocytes, and polymerase chain reaction (PCR) amplification and genotyping were performed for 11 Single Nucleotide Polymorphisms (SNP) in *NOD2*, *IL23r*, *OCTN1* genes along with *IGR*.

**Results:** Our cohort consists of 1196 patients: 435 controls, 485 CD patients, and 276 UC patients. Only patients with genotype data for at least 7 of 11 SNPs were included in our data analysis. The control groups for all 11 SNPs were in Hardy-Weinberg Equilibrium. In genotype-association SNP analysis, all *NOD2* SNPs (rs5743293, rs2066844, rs2066845) and the *IL23r* SNP (rs11465804) showed a significant association to IBD ( $p < 0.03$ ). A multiple gene-interaction analysis showed an association between *NOD2* and *IL23r* with UC ( $p = 0.04$ ). There were no associations between any *OCTN1* and *IGR* SNPs and IBD in this cohort. A multivariable logistic regression analysis showed that female gender, "current" or "former" smoking status, family history of IBD, and *NOD2* SNP minor alleles were associated with CD.

**Conclusion:** IBD remains to be challenging to properly diagnose, characterize, and treat. Our study proposes a combined genetic, phenotypic, and environmental approach in an attempt to better understand IBD. Previously demonstrated associations between *OCTN1* and *IGR* and IBD were not confirmed.

**Keywords:** Inflammatory bowel disease, Crohn's disease, Ulcerative Colitis, SNP, *NOD2*, *IL23r*, *OCTN1*, *IGR*

## Background

Inflammatory bowel disease (IBD) is a chronic inflammatory disorder of the gastrointestinal (GI) tract, comprised of Crohn's disease (CD) and ulcerative colitis (UC). Over the past century, Northern Europe and

North America have witnessed a significant rise in incidences of IBD [1]. The inflammatory response in CD patients is described by transmural inflammation in any portion of the gastrointestinal tract while that of UC is usually limited to the mucosa and submucosa of the colon and rectum [2]. Although the exact pathogenesis is not completely known in IBD, our current understanding suggests a disease etiology dependent upon a multifaceted interaction between genetic, environmental, and clinical factors[3].

\* Correspondence: s0gala01@exchange.louisville.edu

<sup>1</sup>Department of Surgery, The Price Institute of Surgical Research and the Section of Colorectal Surgery, University of Louisville School of Medicine, 550 S. Jackson St, Louisville, KY 40292, USA

Full list of author information is available at the end of the article

Evidence of genetic factors implicated in this condition is supported by increased rates of IBD in monozygotic twins, and the ethnic differences in IBD frequency [4]. In addition, different races and ethnic groups have different degrees of susceptibility to IBD [5]. A number of studies have shown an association between IBD susceptibility and the nucleotide-binding oligomerization domain 2 gene (*NOD2*, also known as *CARD15*), interleukin-23 receptor gene (*IL23r*), organic cation transporter novel type 1 gene (*OCTN1*), and the intergenic region (*IGR*) variants [6-9]. *NOD2* gene mutations have been extensively studied and described in the literature in terms of IBD susceptibility. Loss-of-function mutations in *NOD2* gene appear to be the most significant for development of IBD, although the mechanism that increases disease susceptibility is poorly understood. One report suggests that *NOD2* contributes a protective function in host defense that when compromised elicits a loss in immune bacterial recognition [10]. The *NOD2* gene on chromosome 16 has been specifically implicated in susceptibility to CD with 30-50% of Caucasian CD patients having variants of this gene [6]. The three *NOD2* variant single nucleotide polymorphisms (SNPs) (rs2066844, rs2066845, and rs5743293) have been strongly associated with clinical presentation of inflammatory bowel disease [1,11].

Environmental factors are likely to contribute to IBD susceptibility; however, they are poorly understood and insufficiently described. The recent alarming rise in IBD incidence in the US points to changes in environmental factors rather than drastic alterations in allele frequency, since genetic remodeling could not occur over such a brief time [12]. It seems plausible that environmental influences play as significant a role as their genetic counterparts and their combined actions dramatically affect disease presentation [13].

The objective of this study is to examine the gene-environment interaction which may influence the causation of IBD. We aimed to:

1. Study the incidence of SNP variants in *NOD2*, *IL23r*, *OCTN1*, and the *IGR* genes in a unique well-characterized rural Caucasian IBD population.
2. Examine haplotype frequencies in *NOD2*, *IL23r*, and *IGR* genes in both CD and UC patients.
3. Explore the SNPs' interaction, their possible effect on IBD susceptibility, and report a possible signature interaction model that could differentiate CD and UC patients.
4. Report risk factors that could serve as clinical predictors of IBD susceptibility

## Results

### Population Description

For the purposes of our genetic and clinical comparisons, we only examined the Caucasian subset in order

to maintain statistical and descriptive homogeneity. Our patient population was composed of 1196 unrelated patients (63% women): 435 controls, 485 CD patients, and 276 UC patients.

The mean age at time of IBD diagnosis was 31 years. Twenty-seven percent of patients reported a family history of IBD. The smoking status of the population was defined as follows; 21% as "current" smokers, 30% as "former" smokers and 49% as "never" smokers. Our CD patients were described using the Vienna classification system (Table 1). Sixty-five percent of UC patients had pan-colitis, 22% had left-sided colitis and 13% isolated proctosigmoiditis.

### Analysis of Clinical Variables

Univariate analysis of UC vs. CD examined previously described relevant clinical factors associated with IBD. Gender, smoking status, number of IBD-associated operations, and family history of IBD were strong indicators of CD ( $p < 0.002$ ). Age at diagnosis above 27 years old was more predictive of UC ( $p = 0.003$ ).

### SNP and Haplotype Analysis

For SNP analysis, we included only data for patients successfully genotyped for at least 7 out of 11 SNPs. Thus, there were 909 patients remaining for analysis, as follows: 303 controls, 366 CD patients, and 223 UC patients. The control groups for all 11 SNPs were in HWE. Univariate analysis was performed on all 11 SNPs. In genotype-association SNP analysis, only the *NOD2* SNPs (rs5743293, rs2066844, rs2066845) and *IL23r* SNP-rs11465804 showed a significant association with IBD ( $p < 0.025$ ) (Table 2). The minor alleles in *NOD2* SNPs (rs5743293 and rs2066845) in addition to *IL23r* SNP-rs11465804, were associated with CD ( $p \leq$

**Table 1 Characteristics of CD patients using Vienna Classification**

Characteristic	CD Patients*
<b>Vienna Age</b> <i>N (%)</i>	
A1 (< 40 yrs)	372 (77)
A2 (> 40 yrs)	111 (23)
<b>Vienna Behavior</b> <i>N (%)</i>	
B1 (Non-stricturing, non-penetrating)	110 (23)
B2 (Stricturing)	131 (27)
B3 (Penetrating)	240 (50)
<b>Vienna Location</b> <i>N (%)</i>	
L1 (Terminal ileum)	129 (27)
L2 (Colon)	146 (31)
L3 (Ileocolon)	169 (35)
L4 (Upper GI)	32 (7)

\*Based on available clinical data

**Table 2 Univariate Analysis of Allelic Association in SNPs**

Single Nucleotide Polymorphism	Total # of Alleles (Total = 1644) (N, %)	Control (Total = 582) (N, %)	Crohn's Disease (Total = 654) (N, %)	Ulcerative Colitis (Total = 408) (N, %)	p-value
<b><i>NOD2</i> - rs5743293</b>					< 0.001 <sup>1</sup>
C	59 (4)	8 (1)	43 (7)	8 (2)	< 0.0001 <sup>2</sup>
N	1499 (96)	546 (99)	561 (93)	392 (98)	0.509 <sup>3</sup>
					< 0.001 <sup>4</sup>
<b><i>NOD2</i> - rs2066844</b>					< 0.001 <sup>1</sup>
C	1377 (93)	466 (97)	545 (89)	366 (93)	< 0.001 <sup>2</sup>
T	109 (7)	14 (3)	67 (11)	28 (7)	0.004 <sup>3</sup>
					0.042 <sup>4</sup>
<b><i>NOD2</i> - rs2066845</b>					< 0.001 <sup>1</sup>
G	1558 (98)	560 (99)	601 (96)	397 (99)	< 0.001 <sup>2</sup>
C	38 (2)	6 (1)	27 (4)	5 (1)	1.000 <sup>3</sup>
					0.006 <sup>4</sup>
<b><i>IL-23r</i> - rs11465804</b>					0.022 <sup>1</sup>
G	67 (4)	30 (5)	16 (3)	21 (5)	0.010 <sup>2</sup>
T	1545 (96)	534 (95)	630 (97)	381 (95)	0.948 <sup>3</sup>
					0.019 <sup>4</sup>

<sup>1</sup> Global (Fisher's Exact Test)

<sup>2</sup> Control versus CD

<sup>3</sup> Control versus UC

<sup>4</sup> CD versus UC

0.01), but not UC. The *NOD2* SNP (rs2066844) minor allele was associated with both CD and UC ( $p < 0.005$ ), but was more significantly associated with CD. We also examined the genetic interaction between *NOD2* and *IL23r* genes for UC and CD; this interaction was predictive for UC ( $p = 0.04$ ), but not for CD ( $p = 0.63$ ).

The overall haplotype test was significant ( $p < 0.0001$ ) for *NOD2*, with significant differences in the frequencies of rare haplotypes, with both C-C-G ( $p = 0.008$ ) and C-C-C ( $p = 0.018$ ) being more associated with CD than with UC (Table 3). With *IL23r* gene, the C-A-C-G-G haplotype was associated with CD ( $p = 0.009$ ), while the T-A-C-G-G haplotype was associated with UC ( $p = 0.023$ ) (Table 4). *IIGR* gene analysis showed there was no overall haplotype association with CD or UC ( $p = 0.98$  and  $0.063$ ).

#### Multivariate Analysis of Genetic and Clinical Predictors (CD vs. UC)

We used a multivariable logistic regression model comprising the most significant clinical predictors and the above mentioned SNP variants implicated in IBD (Table 5). The *IL23r* SNP-rs11465804 was not included in the regression analysis since its correlative affect in IBD was not as significant as its *NOD2* counterparts. Female gender, current or former smoking status, IBD-associated surgery, family history of IBD, and frequency of all *NOD2* SNP minor alleles were descriptive predictors of CD. Age at time of diagnosis  $\geq 27$  years was predictive of UC diagnosis.

## Discussion

### I. Current Diagnosis of IBD

IBD is a complex genetic disorder based on a multifaceted interaction between specific genes and environmental factors. The definitive diagnosis and correct identification of IBD remains difficult. IBD diagnostic markers may be lacking in the earliest stages of disease [14]. Radiologic, endoscopic, and histological diagnostic approaches are commonly employed in identification of IBD; however, the inter-observer variability between pathologists calls into question their true clinical value. Distinction between CD and UC, particularly in the case of colonic IBD is important since surgical management varies in each disease, however, in many instances the pathological features overlap making differentiation problematic [15]. Particularly in CD where variability of histological presentation makes accurate diagnosis a challenging exercise, pathologists often differ on their individual assessments [16]. The apparent complexity of successful IBD diagnosis with current techniques demonstrates the immense value of developing a genetic and clinical screening tool.

### II. Clinical and Environmental Predictors of IBD

There are several established clinical descriptors associated with IBD. Our analysis identified age at diagnosis, gender, smoking status, and family history of IBD as being the most important clinical predictors to

**Table 3 NOD2 Haplotype Association Analysis**

SNPs	Haplotype	Haplotype Frequencies				p-value (CD vs. UC)
		Control	CD	UC	CD and UC combined	
rs574329 - rs2066844	C-C	1.34	6.46	1.97	4.75	< 0.0001
rs574329 - rs2066845	C-G	1.47	5.3	1.97	4.02	0.004
<b>NOD-2 gene</b>	C-C	0	1.18	0	0.74	0.027
rs2066844- rs2066845	C-C	1.13	3.54	1.09	2.64	0.009
rs574329 - rs2066844- rs2066845	C-C-G	1.35	5.27	1.97	4.01	0.008
	C-C-C	0	1.19	0	0.75	0.018

distinguish UC from CD. Previous reports indicate that these factors are significantly related to IBD, but failed to adequately describe their effects.

**A. Age at Time of Diagnosis**

Loftus *et al.* described a systematic review of North Americans with CD and UC in respect to patient age at time of diagnosis. The mean age for CD diagnosis ranged from 33 - 45 years [17] while UC patients were diagnosed 5 - 10 years [18,19] later. In concordance with these reports, the median age at time of diagnosis in our cohort was found to be 27 years. The age at time of diagnosis of  $\geq 27$  years age was found to be more predictive of UC patients ( $p = 0.04$ ).

**B. Gender**

In a separate report, Loftus and Sandborn described the influence of gender on disease occurrence for CD and UC [20]. They reported an increased incidence of UC among men, and CD among women [20]. Our analysis concurs with this; the Caucasians in this study showed increased diagnosis of CD in women and UC in men ( $p = 0.014$ ). The female predominance in CD could possibly be attributed to unspecified hormonal interactions [1].

**C. Family History of IBD**

The positive correlation between family history and IBD incidence seems plausible; this is possibly due to a combination of underlying genetic and environmental

factors. This was shown in a study done at a University Hospital in Finland where a comparison among IBD patients revealed 16% of CD patients and 14% of UC patients had at least one family member afflicted with IBD [21,22]. Our data strongly support these conclusions; i.e. a family history of IBD being more predictive of predisposition to CD than to UC ( $p < 0.0001$ ).

**D. Number of Surgeries**

In addition to independent factors of gender, age, and family history, treatment approach may also be descriptive in IBD severity and progression. A North American population based study reported that 43% of patients with CD required major surgery involving incision, excision, and intestinal anastomosis compared to 40% for all UC hospitalizations [21]. Our results showed similar findings; CD patients were more likely to require surgery than their UC counterparts ( $p = 0.015$ ).

**E. Smoking Status**

Smoking was originally recognized as a risk factor for IBD susceptibility over 25 years ago [23,24]. A meta-analysis performed by Mahid *et al.* showed that 12 of 13 reports found that "current" smoking status could be protective against UC development. In contrast, "former" smoker status was found to be associated with an increased risk for the development of UC. A parallel analysis of Crohn's patients revealed 6 of 9 reports

**Table 4 IL23r Haplotype Association Analysis**

SNPs	Haplotype	Frequencies (%)				p-value (CD vs. UC)
		Control	CD	UC	CD and UC combined	
rs10489629 - rs11465804	G-G	3.91	2.25	4.52	3.16	0.048
rs1004819 - rs10489629 - rs11465804	C-G-G	3.94	2.06	4.36	2.95	0.045
rs1004819 - rs2201841 - rs11465804 - rs11209026	T-C-G-G	0.01	0.02	1.1	0.45	0.014
	C-C-G-G	0	0.353	0	0.002	0.010
rs1004819 - rs10489629 - rs11465804 - rs11209026	C-A-G-G	0.4	0.78	0	0.1	0.016
	T-A-G-G	0.03	0.006	1.34	0.61	0.028
rs1004819 - rs10489629 - rs2201841 - rs11465804	C-G-T-G	3.92	2.14	4.37	3.02	0.049
	C-A-C-T	4.97	4.07	7.19	5.1	0.048
rs1004819 - rs10489629 - rs2201841 - rs11465804 - rs11209026	T-A-C-G-G	0.01	0.02	1.12	0.41	0.023
	C-A-C-G-G	0	0.36	0	0.008	0.009

**Table 5 Multivariable Logistic Regression Analysis for UC vs. CD**

Genetic and Clinical Variables		Ulcerative Colitis vs. Crohn's Disease		
		Estimate*	p-value	OR (95% CI)
<i>Age at time of diagnosis</i>	≥ 27	0.32	<b>0.04</b>	1.38 (1.02-1.87)
	< 27	0		
<i>Gender</i>	Female	-0.38	<b>0.01</b>	0.68 (0.50-0.92)
	Male	0		
<i>Smoker Status</i>	Current	-1.63	<b>&lt; 0.0001</b>	0.20 (0.13-0.30)
	Former	-0.81	<b>&lt; 0.0001</b>	0.45 (0.32-0.63)
	Never	0	<b>&lt; 0.0001</b>	
<i>Number of Operations</i>	Yes	-0.63	<b>0.002</b>	0.53 (0.36-0.79)
	No	0		
<i>Family History IBD</i>	Yes	-0.88	<b>&lt; 0.0001</b>	0.42 (0.30-0.57)
	No	0		
<i>NOD2 - rs5743293</i>	C	-1.72	<b>0.0002</b>	0.18 (0.07-0.44)
	N	0		
<i>NOD2 - rs2066844</i>	T	-0.60	<b>0.03</b>	0.55 (0.33-0.93)
	C	0		
<i>NOD2 - rs2066845</i>	C	-1.30	<b>0.02</b>	0.271 (0.09-0.84)
	G	0		

\* A positive estimate indicates the variable being more predictive of ulcerative colitis while a negative estimate is more predictive of Crohn's disease.

indicating an increased disease risk associated with having “ever smoked” [3]. Analysis of our patient cohort, “current” and “former” smoker statuses were both more associated with development of CD than UC ( $p < 0.001$ ).

### III. Genetic Predictors of IBD

*NOD2* gene is identified to be in the *IBD1* locus on chromosome 16 as a CD susceptibility gene. Hugot *et al.* mapped the *IBD1* linkage locus and identified *NOD2* as the underlying IBD susceptibility gene in 2001[25]. Thirty non-conservative polymorphisms have been identified within this gene, but three single SNPs (rs2066844, rs2066845, and rs5743293) account for approximately 82% of the mutated alleles [26]. Replication studies have confirmed that these three SNPs are independently associated with disease [11,27,28].

The receptor for the proinflammatory cytokine, IL-23r, also appears to play a key role in the progression of the chronic inflammation found in CD by driving inflammation through its role in the T-helper 17 (Th17) response. Activated myeloid and T cells express the IL-23 receptor and the level of its expression may affect the way Th17 cells create excess mucosal inflammation [9].

#### A. *NOD2* and *IL23r* SNPs

In our patient population, all *NOD2* variants and one *IL23r* variant were associated with IBD. After multivariable logistic regression analysis, only the *NOD2* SNPs (rs2066844, rs2066845, and rs5743293) remained highly

significant predictors of disease ( $p < 0.001$ ), and in this context all were descriptive of CD, in agreement with previous reports. A meta-analysis conducted by Economou *et al.* found the same three *NOD2* variants to be distinguishing risk factors for CD. According to their analyses, SNP rs5743293 carries a four-fold increase, rs2066845 a three-fold increase, and rs2066844 a two-fold increase of developing CD in non-Jewish descent Caucasians. They further reported that the combination of two of these *NOD2* variants resulted in an increased odds of developing CD of 17 times the normal population [29]. In contrast to *NOD2* SNPs, we did not find the same genetic importance associated with *IL23r* polymorphisms in terms of describing and characterizing IBD. Although only one *IL23r* SNP was marginally significant in our univariate analysis, another variation of this allele yielded notable results when present with a *NOD2* variant. A combination of *NOD2* rs2066844 and *IL23r* rs2201841 was descriptive of patients with UC compared to controls ( $p = 0.04$ ). No such association was found for CD patients.

#### B. *Haplotype Analysis for NOD2 and IL23r*

Another potential differentiator between CD and UC is haplotype analysis. Previous reports implicated numerous *NOD2* and *IL23r* haplotype sequences associated to CD [30,31]. The exact mechanism for disease pathogenesis in terms of these *NOD2* and *IL23r* SNPs remains unclear, but an understanding of how allelic combinations alter the course of disease may prove useful. In

our results, we found two haplotype sequences in *NOD2* associated with CD: C-C-C and C-C-G (rs5743293-rs2066844-rs2066845). With respect to the *IL23r* gene, one haplotype was associated with CD and one with UC: Haplotypes C-A-C-G-G with CD and T-A-C-G-G with UC (rs1004819-rs10489629-rs2201841-rs11465804-rs11209026). Although potentially useful as genetic descriptors of CD and UC, the rarity of these haplotypes makes their clinical use unlikely.

## Conclusion

Inflammatory bowel disease remains difficult to properly characterize, diagnose, and treat. Understanding the combined interactions between clinical, environmental, and genetic factors could serve as key in identifying IBD predictors as well as in properly differentiating between UC and CD. Our report implicates “female gender”, “current” or “former” smoking status, positive family history of IBD, and *NOD2* SNP minor alleles to be associated with CD. Clinically, this will help physicians identify patients with high risk for IBD development, differentiate CD from UC patients, and eventually help better tailor their medical/surgical treatment approach.

## Methods

### Population Selection and Classification

This study was approved by the University of Louisville Institutional Review Board (IRB). Written informed consent was obtained from all subjects. Genetic and clinical information was stored in a password-protected, prospectively maintained HIPAA compliant database. Patients were derived from a clinically well-described university-based IBD practice [4] and from a small relatively rural geographic area consisting of the state of Kentucky and southern Indiana. Our cohort is composed of 1337 patients (37% males and 63% females). It is comprised of unrelated individuals of various races (90% Caucasian, 7% African American, and 3% Asians). For the purpose of this study, we decided to study IBD in a uniform well-described group of patients, specifically Caucasians ( $n = 1196$ ). These patients were not studied for IBD before. There is a profound environmental effect in our studied population as they are derived from one of the areas in the United States with the highest rates of adult cigarette smoking. Additionally, thirty seven percent of all IBD patients in our cohort have a positive family history of IBD, defined as a first or second degree relative with IBD, as compared to only 9% of controls. Perianal CD (PCD) occurred in 147 (46%) of CD patients. All patients were diagnosed to have IBD, using conventional radiologic, endoscopic, and/or histopathologic findings. In cases of IBD colitis, the diagnosis was confirmed by a single specialist gastrointestinal pathologist with a particular interest in

IBD. CD patients and their clinical characteristics, in terms of age at diagnosis, disease location, and behavior were classified according to Vienna classification [32]. Disease location in UC was classified as rectosigmoid, left-sided, or Pancolitis. Comparator patients to function as controls included patients from the same geographic area seen for non-inflammatory, non-neoplastic disorders including hemorrhoidal disease and screening colonoscopy.

### Characterization of Environmental Risk

This included smoking status and it was verified by utilizing a validated Behavioral Risk Factor Surveillance Survey (BRFSS) that was prospectively administered. [24] “*Current smoking*” was defined as having smoked 100 cigarettes in one’s lifetime and smoking every day or some days over the past six months.<sup>24</sup> “*Former smoking*” was defined as having smoked 100 cigarettes in one’s lifetime but currently no longer smoking at all. [24] A “*non-smoker*” is defined as someone who has not smoked 100 cigarettes in his or her lifetime and who does not currently smoke.<sup>24</sup> The category “*ever smoking*” includes individuals from both the current and former smoking groups, while those who “*never smoked*” includes only non-smokers. [24]

### Characterization of Clinical Data

Clinical data included gender, family history of IBD and cancer, age at time of diagnosis, presence of extra intestinal or perianal disease, number of IBD-associated surgeries (as a surrogate index for disease severity), and recurrence in CD.

### Characterization of *NOD2*, *IL23R*, *OCTN*, and *IGR* Polymorphisms

Peripheral blood was obtained and genomic DNA was then extracted with a Puregene<sup>®</sup> DNA extraction kit (Gentra Systems Inc., Minneapolis, MN). Polymerase chain reaction (PCR) amplification and genotyping were performed on an ABI prism 7300 sequence detection system (Applied Biosystems<sup>®</sup>, Foster City, CA). TaqMan<sup>®</sup> SNP-specific PCR primers and fluorogenic probes were obtained from by Applied Biosystems (Applied Biosystems<sup>®</sup>, Foster City, CA). The fluorogenic minor groove binder (MGB) TaqMan<sup>®</sup> probes were labeled with a reporter dye, either FAM ([5-( $\gamma$ )-carboxyfluorescein]) or VICs (a proprietary fluorescent dye produced by Applied Biosystems) specific for the wildtype and variant alleles of each of the following SNPs[33,34] (additional file 1):

1- **Three *NOD2* SNPs:** rs5743293, rs2066844, rs2066845.

2- **Five *IL23r* SNPs:** rs1004819, rs10489629, rs2201841, rs11465804, and rs11209026.

3- **Two IGR SNPs:** rs2522057 and rs7705189.

4- **One OCTN SNP:** rs1050152.

### Statistical Methods

Descriptive statistics related to genetic and clinical characteristics were produced for the entire cohort. Categorical variables were compared using the Pearson Chi-square test (or Fisher's Exact test) for contingency tables [34]. The *t*-test or Wilcoxon rank sum test was used to test the cohort for continuous variables. We also fit the univariable and multivariable logistic regression models for the probabilities of patients in "CD" or "UC" groups about their possible predictors [35]. Furthermore, we examined haplotype associations between CD and UC.

We performed univariate analysis of our control population compared against CD and UC as well as CD against UC patients. Based on this analysis, we selected the following most significant IBD clinical predictors: age at time of diagnosis, gender, smoking status, IBD-associated surgery, and family history of IBD as descriptive of the IBD subtype. The age at time of diagnosis was dichotomized as  $< 27$  or  $\geq 27$  years (as the median age at time of diagnosis was 27). Using these clinical criteria, we performed multivariable logistic regression analysis combining both significant genetic and clinical predictors.

We explored the genotype association, allelic association, and Hardy-Weinberg Equilibrium (HWE) test for 11 SNPs using the Pearson Chi-square test. For allelic association analysis of the 11 SNPs, we used the binary logistic regression method to estimate their odds ratios and 95% confidence intervals for patients with CD or UC as compared to control group. Odds ratios and 95% confidence intervals were estimated using binary logistic regression. Since all the studied genes had already been identified as associated genes for IBD, CD and UC susceptibility, no multiple testing corrections were required. Haplotype analysis was conducted for all the SNPs of *NOD2*, *IL23r* and *IGR*. All calculations were performed with Statistical Analysis Software (SAS) program [34,36]. A *p*-value less than or equal to 0.05 was set to be significant.

### Additional material

**Additional file 1: The four studied genes (*NOD2*, *IL-23r*, *OCTN1*, and *IGR*) along with the corresponding SNPs.** A list of the studied Single Nucleotide Polymorphisms (SNPs) in each gene of interest along with their reference numbers.

### Acknowledgements

This publication is made possible by the John and Caroline Price Family Trust and a gracious gift from Ms. Sara Shallenberger Brown. Additionally, this work was supported by NIH/NIEHS grant 1P30ES014443-01A1 to The

Center for Environmental Genomics and Integrative Biology (CEGIB). These funding sources had no role in the production of this manuscript.

### Author details

<sup>1</sup>Department of Surgery, The Price Institute of Surgical Research and the Section of Colorectal Surgery, University of Louisville School of Medicine, 550 S. Jackson St, Louisville, KY 40292, USA. <sup>2</sup>The Department of Bioinformatics and Biostatistics, University of Louisville School of Medicine, 505 S. Hancock St., Louisville, KY 40202, USA. <sup>3</sup>AmeriPath Institute of Gastrointestinal Pathology and Digestive Disease, Oakwood village, OH 44146, USA.

### Authors' contributions

SG, ZK, and MRE completed the study design. ZK, SA, and SG drafted the manuscript. SG revised the manuscript for intellectual content. CW and HR completed the genetic analysis. EBW completed the acquisition of the clinical data. JP and SNR completed statistical data analysis. RP completed the histopathological review for enrolled inflammatory bowel disease patients. In addition, all authors have seen and approved the final version of this manuscript for submission.

### Competing interests

The authors declare that they have no competing interests.

Received: 19 September 2011 Accepted: 23 January 2012

Published: 23 January 2012

### References

- Loftus EV Jr: Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. *Gastroenterology* 2004, **126**:1504-1517.
- Martins NB, Peppercorn MA: Inflammatory bowel disease. *Am J Manag Care* 2004, **10**:544-552.
- Mahid SS, Minor KS, Soto RE, Hornung CA, Galandiuk S: Smoking and inflammatory bowel disease: a meta-analysis. *Mayo Clin Proc* 2006, **81**:1462-1471.
- Crawford NP, Colliver DW, Eichenberger MR, Funke AA, Kolodko V, Cobbs GA, et al: CARD15 genotype-phenotype relationships in a small inflammatory bowel disease population with severe disease affection status. *Dig Dis Sci* 2007, **52**:2716-2724.
- Mahid SS, Mulhall AM, Gholson RD, Eichenberger MR, Galandiuk S: Inflammatory bowel disease and African Americans: a systematic review. *Inflamm Bowel Dis* 2008, **14**:960-967.
- Alvarez-Lobos M, Arostegui JI, Sans M, Tassies D, Plaza S, Delgado S, et al: Crohn's disease patients carrying Nod2/CARD15 gene variants have an increased and early need for first surgery due to stricturing disease and higher rate of surgical recurrence. *Ann Surg* 2005, **242**:693-700.
- Armuzzi A, Ahmad T, Ling KL, de SA, Cullen S, van HD, et al: Genotype-phenotype analysis of the Crohn's disease susceptibility haplotype on chromosome 5q31. *Gut* 2003, **52**:1133-1139.
- Babusukumar U, Wang T, McGuire E, Broeckel U, Kugathasan S: Contribution of OCTN variants within the IBD5 locus to pediatric onset Crohn's disease. *Am J Gastroenterol* 2006, **101**:1354-1361.
- Lin Z, Poritz L, Franke A, Li TY, Ruether A, Byrnes KA, et al: Genetic association of nonsynonymous variants of the IL23R with familial and sporadic inflammatory bowel disease in women. *Dig Dis Sci* 2010, **55**:739-746.
- Inohara , Chamaillard , McDonald C, Nunez G: NOD-LRR proteins: role in host-microbial interactions and inflammatory disease. *Annu Rev Biochem* 2005, **74**:355-383.
- Hampe J, Grebe J, Nikolaus S, Solberg C, Croucher PJ, Mascheretti S, et al: Association of NOD2 (CARD 15) genotype with clinical course of Crohn's disease: a cohort study. *Lancet* 2002, **359**:1661-1665.
- Danese S, Sans M, Fiocchi C: Inflammatory bowel disease: the role of environmental factors. *Autoimmun Rev* 2004, **3**:394-400.
- Fiocchi C: Inflammatory bowel disease: etiology and pathogenesis. *Gastroenterology* 1998, **115**:182-205.
- Geboes K: Pathology of inflammatory bowel diseases (IBD): variability with time and treatment. *Colorectal Dis* 2001, **3**:2-12.
- Yantiss RK, Odze RD: Diagnostic difficulties in inflammatory bowel disease pathology. *Histopathology* 2006, **48**:116-132.

16. Lessells AM, Beck JS, Burnett RA, Howatson SR, Lee FD, McLaren KM, et al: **Observer variability in the histopathological reporting of abnormal rectal biopsy specimens.** *J Clin Pathol* 1994, **47**:48-52.
17. Loftus EV, Schoenfeld P, Sandborn WJ: **The epidemiology and natural history of Crohn's disease in population-based patient cohorts from North America: a systematic review.** *Aliment Pharmacol Ther* 2002, **16**:51-60.
18. Bjornsson S, Johannsson JH: **Inflammatory bowel disease in Iceland, 1990-1994: a prospective, nationwide, epidemiological study.** *Eur J Gastroenterol Hepatol* 2000, **12**:31-38.
19. Loftus EV, Silverstein MD, Sandborn WJ, Tremaine WJ, Harmsen WS, Zinsmeister AR: **Ulcerative colitis in Olmsted County, Minnesota, 1940-1993: incidence, prevalence, and survival.** *Gut* 2000, **46**:336-343.
20. Loftus EV, Sandborn WJ: **Epidemiology of inflammatory bowel disease.** *Gastroenterol Clin North Am* 2002, **31**:1-20.
21. Bernstein CN, Nabalamba A: **Hospitalization, surgery, and readmission rates of IBD in Canada: a population-based study.** *Am J Gastroenterol* 2006, **101**:110-118.
22. Halme L, Turunen U, Helio T, Paavola P, Walle T, Miettinen A, et al: **Familial and sporadic inflammatory bowel disease: comparison of clinical features and serological markers in a genetically homogeneous population.** *Scand J Gastroenterol* 2002, **37**:692-698.
23. Harries AD, Baird A, Rhodes J: **Non-smoking: a feature of ulcerative colitis.** *Br Med J (Clin Res Ed)* 1982, **284**:706.
24. Somerville KW, Logan RF, Edmond M, Langman MJ: **Smoking and Crohn's disease.** *Br Med J (Clin Res Ed)* 1984, **289**:954-956.
25. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, et al: **Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease.** *Nature* 2001, **411**:599-603.
26. Lesage S, Zouali H, Cezard JP, Colombel JF, Belaiche J, Almer S, et al: **CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease.** *Am J Hum Genet* 2002, **70**:845-857.
27. Cuthbert AP, Fisher SA, Mirza MM, King K, Hampe J, Croucher PJ, et al: **The contribution of NOD2 gene mutations to the risk and site of disease in inflammatory bowel disease.** *Gastroenterology* 2002, **122**:867-874.
28. Vermeire S, Wild G, Kocher K, Cousineau J, Dufresne L, Bitton A, et al: **CARD15 genetic variation in a Quebec population: prevalence, genotype-phenotype relationship, and haplotype structure.** *Am J Hum Genet* 2002, **71**:74-83.
29. Economou M, Trikalinos TA, Loizou KT, Tsianos EV, Ioannidis JP: **Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis.** *Am J Gastroenterol* 2004, **99**:2393-2404.
30. Okazaki T, Wang MH, Rawsthorne P, Sargent M, Datta LW, Shugart YY, et al: **Contributions of IBD5, IL23R, ATG16L1, and NOD2 to Crohn's disease risk in a population-based case-control study: evidence of gene-gene interactions.** *Inflamm Bowel Dis* 2008, **14**:1528-1541.
31. Sugimura K, Taylor KD, Lin YC, Hang T, Wang D, Tang YM, et al: **A novel NOD2/CARD15 haplotype conferring risk for Crohn disease in Ashkenazi Jews.** *Am J Hum Genet* 2003, **72**:509-518.
32. Gasche C, Scholmerich J, Brynskov J, D'Haens G, Hanauer SB, Irvine EJ, et al: **A simple classification of Crohn's disease: report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998.** *Inflamm Bowel Dis* 2000, **6**:8-15.
33. **Applied Biosystems.** [http://www.appliedbiosystems.com], Accessed July, 2010.
34. Fleiss JL, Levin BA, Paik MC: *Statistical methods for rates and proportions.* 3 edition. Hoboken, NJ: Wiley-Interscience; 2003.
35. Agresti A: *Categorical data analysis.* 2 edition. New York: Wiley; 2002.
36. **The SAS System V9.** Cary, NC, SAS Institute Inc; 2003.

doi:10.1186/1477-5751-11-7

**Cite this article as:** Kanaan et al.: Clinical predictors of inflammatory bowel disease in a genetically well-defined Caucasian population. *Journal of Negative Results in BioMedicine* 2012 **11**:7.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

