UNIVERSIDAD DE SALAMANCA

Departamento de Medicina

Facultad de Medicina

DOCTORAL THESIS

ANALYSIS OF GENETIC POLYMORPHISMS INVOLVED IN APOPTOSIS AND AUTOPHAGY PATHWAYS IN CROHN'S DISEASE

Alejandra Fernández Pordomingo

2014

INDEX

1. INTRODUCTION

Inflammatory bowel disease (IBD) is defined as a heterogeneous group of conditions of the gastrointestinal tract which is characterized by chronic inflammation. It comprises three main entities: Crohn's disease (CD), ulcerative colitis (UC) and indeterminate colitis. This disease is characterized by its chronic evolution, with periods of activity or outbreaks and stages of remission or stillness. CD can affect any part of the digestive tract, and the inflammation is typically transmural, which frequently triggers the creation of stenosis and fistulas. UC only affects the colon (although there may be a slight «backwash» ileitis), and it is limited to the mucosa and superficial submucosal layers[1]. In CD there is a large heterogeneity regarding the form of presentation, extraintestinal manifestations and location; as well as its behavior and the response to treatment of both entities, CD and UC[2].

The etiology of IBD is unknown, but it appears that all the variables are caused by a common mechanism. It is a disorder of multifactorial origin in which, for some individuals who are genetically predisposed, exposure to certain environmental factors (diet, tobacco, gut flora, geographical location) leads to an altered response of the innate and adaptive immunity, with the release of proinflammatory molecules that cause intestinal lesions due to the alteration of certain phenomena of apoptosis and repair[3]. The disease may be caused by environmental factors and maintained by an anomalous immune response to said factors which is genetically determined. The gut flora plays an essential role in its development[4]. The inflammatory process perpetuates due to the interaction between the cells of the intestinal epithelium and the competent cells of the immune system[5]. Among the different environmental factors that have been analyzed, the only ones that have shown a certain relation with IBD are tobacco, as a factor for a poor prognosis in CD, and appendectomy and tobacco, as protective factors in UC. However, these parameters do not account for the variability regarding the incidence and prevalence of these conditions. Therefore, the rates of appendectomies have decreased in developed countries, but the incidence of UC has remained stable, while the incidence of CD is low in populations with an important rate of tobacco consumption, such as Asia and Africa, and high in populations where the rate of smokers is low, such as Sweden and Canada[6].

The hypothesis that genetic factors may play some role in the etiopathogenesis of IBD comes from studies in which a higher prevalence of the disease was found among Caucasian and Jewish groups[7], and studies in which a certain family aggregation and higher incidence of the disease in monozygotic twins has been observed. In this regard, the prevalence of IBD among the Jewish population is 2 to 4 times higher than in any other ethnic group, and it is higher in Ashkenazi Jews than in Sephardi Jews or in Oriental Jews, with no influence of their geographic distribution[3,8]. The risk of having any of these diseases in monozygotic twins is around 20-50% for CD and 14-19% for UC, whereas this risk decreases in dizygotic twins and comes down to 0-7% for both

diseases[9]. The main risk factor for IBD is having a relative with the same disease: the risk of developing UC and CD in first-degree consanguineous relatives of patients with IBD is 1.6% and 5.2%, respectively. This means that the risk of developing CD is 12 to 15 times higher than in the general population in a comparable age range[10].

IBD is a polygenic disease that does not follow a Mendelian inheritance pattern, which makes it difficult to find specific susceptibility genes. Its origin cannot be explained by high penetrance mutations, but by a variety of mutations, some of them rare and some common, which have a slight influence on the risk of developing the disease. The studies with candidate genes, carried out in case-control cohort studies or case-parents triads analyze a specific gene with potential interest. The studies analyze the allele frequencies (in the case of cases and controls) or the transmission of a single-nucleotide polymorphism (SNP) to the offspring (in the case of triads), so that the difference between patients and controls or in the transmission to affected children reveal the involvement of the analyzed gene in the pathogenesis of the disease[11].

Over the last years, thanks to the sequencing of the human genome and the creation of public databases of SNPs, great advances have been made in this group of diseases thanks to the design of genome-wide association studies (GWAS) based on large populations. The advantage of this kind of studies with regard to the previous ones with candidate genes is that they are «hypothesis-free». Therefore, a total of 163 *loci*/susceptibility genes have been identified, which represents a significantly higher figure than in other complex diseases. Out of the 163 identified *loci*, 110 grant susceptibility for both entities, 30 are specific for CD and 23 for UC, which proves an ever-growing proportion of *loci* which are common for both entities and less specific *loci* for each one of them[12].

With regard to the different treatments used in IBD, it is possible that in the future, the application of genetics in this field (pharmacogenetics) will make it possible for practitioners to predict the response to the drug and to prevent its toxicity before starting the treatment. Some genes have been involved in the effectiveness and secondary effects of some drugs, although there is a need for wide and prospective studies with uniform criteria regarding the phenotype of patients which correlate these findings with the clinical practice.

2. HYPOTHESIS

In IBD, and particularly in CD, there has been an involvement of some genetic alterations that have an influence on the recognition of bacteria of the intestinal lumen (*NOD2/CARD15*), of microbioma and of mechanisms of the immune response. In this last group, phenomena of autophagy (*ATG16L1*) and apoptosis (*p53, BcL2, BAX*) play an essential role, and their function is fundamental for the maintenance of homeostasis of the «intestinal wall».

These genetic alterations have an influence not only on the appearance of the disease, but also on the different forms of presentation.

On the other hand, there are genetic alterations, some of which have not been identified, which regulate the response to the different therapies that are used in this disease.

In this regard, no studies have been presented in our population, and therefore we consider it necessary to carry out case-control studies of the genes that regulate these aspects in our patients.

3. OBJECTIVES

To study certain selected polymorphisms of *NOD2*, *ATG16L1*, *p53*, *BCL2* and *BAX*, related with innate immune response, processes of autophagy and apoptosis in a sample of patients with Crohn's disease from our population.

- o To analyze whether there is a relation between both variables and a higher susceptibility to suffer from this disease, its behavior, the need for surgery and the response to drugs.

- o To study whether the association of different mutated alleles is related with a higher susceptibility to suffer from this disease.

- o To study whether the association of different mutated alleles is related with the evolution of the disease.

4. MATERIAL AND METHODS

Observational case-control study.

4.1. Population

4.1.1. Patients

The study included 102 patients with Crohn's Disease who were randomly selected and had been monitored in our IBD service.

In all cases, the diagnosis of CD was established through clinical, radiological and endoscopic criteria (Lennard-Jones criteria). The reach and behavior of the disease were defined with an endoscopic test (ileocolonoscopy with gastroscopy in some cases) and radiological techniques (intestinal MRI, GI series, ultrasound, CT scan and capsule endoscopy), and they were divided according to the Montreal 2005 classification.

4.1.2. Controls

101 individuals without IBD from Salamanca which had undergone a study of the polymorphisms of the genes involved in this study, and whose data were included in the files of the Department of Genetics and Molecular Biology of the University of Salamanca.

4.2. Sample collection and processing

A sample of peripheral blood was taken from all patients, with their informed consent.

The DNA samples were obtained from peripheral blood mononuclear cells. All the patients and controls signed an informed consent form before the collection of the sample.

4.3. Clinical data collection

The records of all patients were reviewed and complemented with personal or telephone interviews in cases in which the information was incomplete. All the data were included in an Excel spreadsheet for the subsequent statistical analysis.

4.4. Clinical variables

➤ **AGE:** We recorded the age of each patient of the sample at diagnosis. In the statistical analysis, the patients were classified into three groups, according to the age at diagnosis and the Montreal classification: A1 ≤ 16 years; A2 > 16 and ≤ 40; and A3 > 40.

➤ **SEX:** Man or woman.

➤ **FAMILY HISTORY OF IBD:** We recorded whether the patients had first or second degree relatives with a history of IBD.

➤ **SMOKING HABITS:** Two categories were defined: smokers and non-smokers. The first one included all the patients who had smoked at some point, even though they may not have been active smokers at diagnosis, without other considerations, given the size of the sample.

➤ **PHENOTYPE OF THE DISEASE:** The patients were divided according to the Montreal classification which was described before.

➤ **EXTRAINTESTINAL MANIFESTATIONS:** A difference was established between patients who showed them and those who did not have them. We took into account arthropathy, erythema nodosum, pyoderma gangrenosum and uveitis.

➤ **PERIANAL DISEASE AT DIAGNOSIS AND/OR ALONG THE EVOLUTION:** Perianal disease (PAD) was considered an independent entity from fistulizing forms, and we recorded the presence of fissures, fistulas or abscesses. The assessment of the perianal disease was carried out with examination under anesthesia (EUA), endoanal ultrasound (EU) and pelvic magnetic resonance imaging (MRI).

➤ **ADMISSION CAUSED BY AN OUTBREAK OF THE DISEASE OR PROBLEMS ASSOCIATED TO IT:** We analyzed whether the patient had to be hospitalized along the evolution of the disease and the number of admissions. Also, we considered whether any of these admissions had been caused by the appearance of any of the following complications: megacolon, digestive hemorrhage, perforation or intra-abdominal abscess.

➤ **USE OF CORTICOIDS IN THE FIRST OUTBREAK, DEVELOPMENT OF CORTICODEPENDENCE OR CORTICORESISTANCE:**

a. Use of steroid treatment in the first outbreak of the disease.
b. Appearance of corticodependence, defined as the reappearance of the symptoms when the dose is reduced or the need for two cycles of corticoids in a period of six months, or for three cycles in one year.
c. Corticoresistance was defined as the lack of response to a treatment with adequate doses of corticoids for a pre-established period of time before the treatment of between one and two weeks, depending on the severity of the symptoms.

➢ **USE OF IMMUNOSUPPRESSANTS OR BIOPHARMACEUTICALS:**

We analyzed whether the patient, in the course of the disease, had received a treatment with immunomodulatory drugs: azathioprine, 6-mercaptopurine or methotrexate, or with biopharmaceuticals: infliximab or adalimumab.

For each of them, we recorded:

- Date when the drug was first administered.
- Dose received.
- Indication: remission induction; maintenance; prevention of postoperative recurrence; fistulizing disease; extraintestinal manifestations.
- Clinical response: improvement of the symptoms or closure of less than 50% of the fistulas with/without a concomitant treatment.
- Clinical remission: Disappearance of symptoms or closure of at least 50% of the fistulas without a concomitant treatment.
- Appearance of adverse effects, with mention to whether the event made it necessary to remove the drug.

➢ **SURGERY:**
a. We evaluated whether the patient had required surgery or not with regard to IBD.
b. We evaluated whether there had been
   i. an intestinal resection.
   ii. In the case of PAD, resection of fistulas and/or drainage of abscesses with suture.
c. We also analyzed whether the patient had undergone more than one operation.

4.5.   Methodology of the genetic studies

4.5.1.   Methodology of the polymorphism analysis

The following polymorphisms were analyzed with TaqMan probes in a StepOnePlus thermocycler from Applied Biosystems, with prior sequence-specific amplification of the genes with PCR (polymerase chain reaction).

GENE	SNP
NOD2/CARD15	rs2066844
ATG16L1	rs2241880
BAX	rs4645878
BCL-2	rs2279115
P53	rs1042522

The study of the polymorphism of the gene p53 (rs1042522) was carried out with PCR-RFLP. To do so, a fragment of 291 bp of exon 4 was amplified with PCR (polymerase chain reaction) and it was afterwards guided with restriction endonuclease.

4.6. Statistical methods

The description of qualitative variables was made with percentages, and continuous variables were expressed with the basic descriptors of average and standard deviation. In order to analyze the association between genes and cases and controls, as well as the association between clinical variables, the chi-square test for contingency tables was used. Logistic regression was used for the analysis of risk of the polymorphism, with an assessment of the odds ratio and the 95% confidence intervals.

Gene-gene interaction was analyzed with classification and regression trees (CART). This analysis is a process of recursive partitioning of the total sample based on the associations of the predictors with the answer variable. The division is always

considered binary. The best predictor for each partition is that which presents the highest value for the likelihood-ratio test or p-value associated to it. The process of partition ends when no association is found between the predictors and the variable of interest, or when the size of the sample is too small. The minimum sample size that was selected as a threshold to finish the partition process was 20 individuals. In order to prevent an increase in type I errors due to the large number of comparisons that are done in this analysis, we have used the penalized model described in Sall[13] (2002), which is not as restrictive as the model proposed by Bonferroni and which is usually applied in these procedures. In order to assess the adjustment of the model we used McFadden's pseudo R^2 and deviance. The validation of the classification trees was carried out with two procedures: 5-fold cross validation, in which the original sample is divided into five sub-samples in order to validate the adjustment of the resulting tree; and a procedure to validate the predictive power of the model. This last method consists of the random division of the sample into two sub-samples; the first one is used to create the tree (training sample) and the second one is used to validate its power of classification (validation sample). The validation sub-sample was created randomly, and it is made up of 25% of all the observations. This analysis was also used to research the gene-environment interactions in the explanation of the structure of clinical variables that determine the disease, such as age at diagnosis, extraintestinal manifestations, location and phenotype, presence of perianal disease, need for surgery, need for corticoids in the first outbreak, need for azathioprine, need for TNF inhibitors, and need for infliximab. Tobacco was considered an environmental variable.

The software that was used for the descriptive analysis, logistic regression and contingency tables was *IBM SPSS version 19*. In the analysis of classification trees we used the algorithm implemented in *SAS JMP version 7*.

The significance levels used in all the analyses were the usual 0.05 and 0.01.

5. RESULTS

5.1. Clinical variables

The average age at diagnosis is 31.69±14.024 years, with a predominance of the 16-40 years range (Montreal A2; 72.5%). The distribution regarding sex is homogeneous, with a slight predominance of women (43.1% M; 56.9% W). Percentages regarding all other variables are shown in Table 1.

Table 1: *Percentages of the different clinical variables analyzed in the general sample.*

	N	%
Sex	V 44	43.1%
	M 58	56.9%
Tobacco	**53**	**51.96%**
AR IBD (1st degree)	**14**	**13.6 %**
L1	**40**	**39.2%**
L2	**18**	**17.6%**
L3	**42**	**41.2%**
± L4	**2**	**2%**
B1	**56**	**54.9%**
B2	**25**	**24.5%**
B3	**21**	**20.6%**
Extraintestinal Manifestations	**42**	**41.2%**
Perianal (PAD)	**37**	**36.3%**
GCS 1st outbreak	**39**	**38.2%**

GCs	**93**	**91.2%**	
AZA	**67**	**65.7%**	
Response to AZA	**50**	**49%**	**74.6% OF PATIENTS TREATED**
Remission with AZA	**22**	**21.6%**	**32.8% OF PATIENTS TREATED**
TNF inhibitors	**34**	**33.3%**	
IFX	**30**	**29.4%**	
Response to IFX	**22**	**21.5%**	**73.3% OF PATIENTS TREATED**
Remission with IFX	**11**	**10.8%**	**36.7% OF PATIENTS TREATED**
Surgery	**53**	**52%**	
Intestinal resection	**40**	**39.2%**	

5.2. Association between clinical variables

There is a tendency to statistical significance regarding the need of corticoids in the first outbreak when the age at diagnosis goes down (p=0.087). The use of azathioprine (AZA) was essentially required in inflammatory (62.5%) or fistulizing forms (81%) with ileal (65%) or ileocolonic location (78.6%) (p=0.015). There is statistical significance in the administration of AZA (p=0.002) and TNF inhibitors (p=0.005) as patients are younger, and this may be explained by the fact that the forms of presentation of the disease at these ages are more aggressive. Age at diagnosis was not associated with the response or the remission obtained with AZA or TNF inhibitors. As expected, a higher percentage of patients who required surgery corresponds with fistulizing (85.7%) and stenosing forms (60%) (p=0.000). There is a trend that reveals a higher probability of surgery when the location is ileal or ileocolonic, which means that ileal involvement would be a determining factor for surgery (p=0.132). Also, these locations are associated with the need of IMM or TNF inhibitors (p=0.01). Surgery was not required when there was response to or remission with these drugs.

No association was found between the polymorphisms of our study and the analyzed clinical variables.

5.3. Associations of polymorphisms with the control group

NOD2(rs2066844): The **homozygous TT** genotype is the most common one among Crohn's Disease patients (98.8%), and CC is the most common one in the control group

(100%). The **T allele** is more common in Crohn's patients than in the control group (92.8% vs. 7.2%).

P53 (Pro/Arg): The **homozygous CC (HN=*Pro/Pro*)** genotype is the most common one among Crohn's patients (67.6%), and the heterozygous CG genotype is the most common one in the control group (61.3%). **The C allele (*Pro*)** is more common in Crohn's patients than in the control group (52.5% vs. 47.5%), N.S.).

ATG16L1 (rs2241880): The **homozygous GG** genotype is the most common one among Crohn's patients (63.5%), and the heterozygous AG genotype is the most common one in the control group (58.3%). The **G allele** is more common in Crohn's patients than in the control group (53.7% vs. 46.3%, N.S.).

BAX (rs4645878): The **heterozygous GA** genotype is the most common one in Crohn's patients (57.4%), and the homozygous GG genotype is the most common one in the control group (51.1%, N.S.). The **allele A** is more common in Crohn's patients than in the control group (53.6% vs. 46.4%).

BCL2 (rs2279115): The **homozygous AA** genotype is the most common one among Crohn's patients (65.8%), and the homozygous CC genotype is the most common one in the control group (56.9%, N.S.). The **allele A** is more common in Crohn's patients than in the control group (55.3% vs. 44.7%, N.S.).

POLYMORPHISMS	DOMINANT ALLELES IN CD	DOMINANT ALLELES IN CONTROLS	ALLELES CD/CONTROLS
NOD2(rs2066844)	TT (98.8%) OR=43.09(CI 95% 5.30;349.7)	CC (100%) OR=1	T (92.8% vs. 7.2%) OR=24.11 (CI 95% 14.35; 40.53)
P53 (Pro/Arg)	CC(HN=*Pro/Pro*) (67.6%)	CG (61.3%) P=0.013	C (*Pro*) (52.5% vs. 47.5%)
ATG16L1 (rs2241880)	GG (63.5%) P=0.025	AG (58.3%)	G (53.7% vs. 46.3%) P=0.08
BAX (rs4645878)	GA (57.4%)	GG (51.1%) P=0.523	A (53.6% vs. 46.4%)

BCL2 (rs2279115):	AA (65.8%)	CC (56.9%)	A (55.3% vs. 44.7%)
		P=0.087	P=0.058

5.4. Study of the interrelation between genes

The analysis includes all the genes that were studied in their dominant or recessive forms, according to the significance and magnitude of the odds-ratio that was obtained in the individual analysis of each gene. Therefore, in genes *NOD2*, *p53* and *BAX*, we have selected the recessive form (or model), and for the genes *BCL2* and *ATG16L1*, the dominant form (or model) was selected. The resulting tree can be seen here:
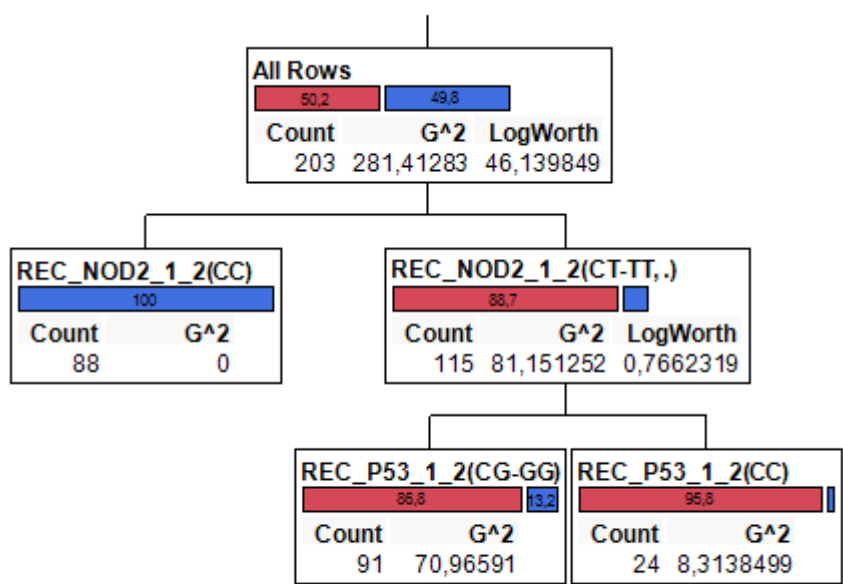
At the start, all patients are divided into case and control groups. The first partition is produced by the recessive model of NOD2, so that individuals with a CC genotype are all controls, while 88.7% of individuals with a CT or TT genotype are cases. This model with a single gene presents R^2=0.708, with great stability, because it barely varies compared with the results obtained through 5-fold cross validation (R^2=0.71). The percentage of well-classified patients is 93%.

In the second segmentation, no genes were found to be significant, although p53 is the one with the lowest p-value (p=0.17). Therefore, we carried out a second partition to explore the results. R^2 increases very slightly: R^2=0.714. The percentages of cases are 95.8% for the CC allele, and 86.8% for the CG/GG alleles. The percentage of well-

classified patients is the same as in the partition without p53. In the general model, and after all the estimates have been taken into account, sensitivity is 100% and specificity is 87%; the positive predictive value is 89% and the negative predictive value is 100%.

6. DISCUSSION

6.1. CLINICAL ASPECTS

The sample, which was randomly chosen from more than 400 patients who had been admitted in our department with CD, is similar to other populations regarding age, sex and all other clinical variables[14,15]. As regards age groups, there is a predominance of patients diagnosed from age 16 to age 40 (Montreal A2; 72.5%). The distribution of sex is mostly homogeneous, with a slight predominance of women (43.1% M; 56.9% W).

There is a tendency that shows a higher need for glucocorticoids (GCs) in the first outbreak when age at diagnosis is lower, but it is not statistically significant (p=0.087); the use of GCs was less common in patients over 40 years old. AZA was mainly used for inflammatory (62.5%) or fistulizing forms (81%), located in the ileum (65%) or ileocolon (78.6%) (p=0.015). There is a statistically significant association between the use of AZA (p=0.002) and TNF inhibitors (p=0.005) and a lower age of patients, which would be explained by the fact that the forms of presentation are more aggressive in this age group. In recent years, the use of immunomodulators and TNF inhibitors is increasingly common. However, the latest publications show a slightly elevated risk of lymphomas in patients treated with immunomodulators, which means that the current trend is to avoid the use of AZA in young patients and patients over 65 years old[16]. Age at diagnosis was not significantly associated with response to or remission with AZA or TNF inhibitors.

As expected, the higher percentage of patients who required surgery tallies with fistulizing (85.7%) and stenosing forms (60%) (p=0.000). There is a tendency towards a higher need for surgery when there is an ileal (57.5%) or ileocolonic location (57.1%). Therefore, an ileal location is a determining factor for surgery (p=0.132). Likewise, given the fact that one of the indications for surgery is the lack of response to the pharmacological treatment, patients who required IMM or TNF inhibitors were operated on more frequently (p=0.01). Surgery was not necessary whenever there was response or remission with these drugs (p=0.038 for response to AZA; p=0.02 for remission with IFX).

In our sample, tobacco consumption was not a behavioral or treatment determining factor, but the size of the sample makes it impossible to draw any conclusion in this regard. Tobacco is a proven agent of aggressive behavior and failure of treatment[17], and the percentages of these subgroups in our series also point in that same direction.

6.2. STUDY OF POLYMORPHISMS

The analysis of GWAS in IBD has detected multiple associations of certain SNPs and

associations of these polymorphisms with other diseases that can be classified as autoimmune. Some of these SNPs grant susceptibility to suffer from the disease, and others make it possible to develop certain phenotypes of the disease. More than 100 SNPs associated with IBD have been detected, 30 of which are specific for CD[18]. However, the conclusions that have been presented so far, if we accept the results obtained with *NOD2* (but not without certain reservations), have no clinical use and have not given a satisfactory answer to the expectations created by the hypotheses on which they were maintained. The obtained results often do not repeat themselves in other studies, and in any case, the strength of association between genotype and phenotype has generally been low, with areas under the curve below 0.70. For this reason, studies such as ours are justified. Our work studies SNPs of genes that could have an influence on the processes of autophagy and apoptosis in smaller and more homogeneous populations[19].

NOD2 (rs2066844)

Many genes have been involved so far in the prognosis of CD, and several attempts have been made to classify the disease according to genetic profiles. In this regard, *NOD2/CARD15* seems to be not only a susceptibility gene, but also a modifier of the disease[11]. Among all the published studies on the clinical relevance of mutations of the *NOD2/CARD15* gene, some of them provide information regarding the location of the disease, and particularly regarding the association with the location in the ileum and the upper intestinal tract[20], although some of them reveal the lack of a colonic location. Other studies show data on the relation between the variants of *NOD2/CARD15* with stenosing or penetrating forms of the disease, as well as on the relation between an earlier onset age and a higher need for surgery. These variants are not only associated to a first intestinal resection, but also to the need for new surgeries in patients with CD[21]. Patients who are heterozygous and homozygous for mutations of the different variables of *NOD2* show an increased risk of a more aggressive evolution of the disease[22]. Some studies have even found an association between *NOD2/CARD15* and symptoms of acute intestinal obstruction[23].

Our results are coherent with those of previous studies with regard to the high penetrance of the disease on homozygous patients for the risk allele T of this SNP. In our sample, having the T allele for SNP rs2066844 of the *NOD2* gene was strongly associated with the risk of developing CD, compared with the control group. The T allele is more common in Crohn's patients than in the control group (92.8% vs. 7.2%), where the most common allele is C [OR: 24.12; (95% CI:14.354;40.53)], regardless of the phenotype. This association is much higher than what was observed in other published series, and this could be due either to the size of the sample or to the characteristics of our population. In the co-dominant model, the homozygous genotype TT was the most common one among Crohn's patients (98.8%), with OR 43.091 (95% CI -5.309;349.781); and in the control group it was the CC genotype (100%), with OR=1. This association is stronger in the dominant group. To sum up, we

have not found any homozygous CC case with CD in the co-dominant or the recessive models, and only one member of the control group presented a homozygous TT genotype in the co-dominant group. Therefore, we may declare that the C allele is a protective factor for CD.

The high penetrance of the T allele in our sample of CD patients would account for the fact that no relation was found with any of the clinical variables that were analyzed, not even with location or behavior. The differences in this regard described in other studies are difficult to explain: they are probably due to the small size of the sample in some of them, the variability of the disease between different populations, and even to inter-observer variability. It would also be necessary to discuss whether there really is a relation between some of the variants of the gene and the stenosing forms and a higher need for surgery, whether the results of the studies only reflect the high percentage of patients with ileal CD who develop stenosis and therefore require surgery, or whether this phenomenon may be determined by other genetic aspects of the patient.

P53 (Pro/Arg)

p53 is a tumor suppressor gene, and the SNP Arg>Pro of codon 72 (R72P) has been extensively analyzed in studies which focus on this field, because this domain is essential in the apoptotic response and the inhibition of carcinogenesis[24]. However, little is still known about the role that it plays in IBD and its clinical relevance. A study evaluated the association between this SNP and the risk of developing UC in a population in the north of Iran [n=190 (115 men; 75 women); mean age 32±8.6 years] and also found a significantly higher frequency of the *Pro* (C) allele in the population with UC, compared with the control group (p<0.0001). However, no association was found between being a carrier of the *Pro* allele and a higher risk of extensive disease (p<0.45), left-sided colitis (p<0.38) or an earlier onset of the disease (p<0.33).

The association between this polymorphism and CD seems to be particularly interesting for the identification of genetic markers that involve a potential malignant transformation of the disease. Mutations of *p53* have been found in fluids from the colon of patients with UC[25]. Also, alterations of the p53 protein have been found in patients with colorectal cancer associated to UC[26].

In our sample, the C allele (*Pro*) is more common in Crohn's patients than in the control group (52.5% vs. 47.5%), without statistically significant differences. The homozygous CC (HN=*Pro/Pro*) genotype for p53 (Pro/Arg) has been the most common one among our Crohn's patients (67.6%), and the heterozygous CG genotype was more common in the control group (61.3%; p=0.013). These results are significantly repeated in the co-dominant and recessive models (p=0.044; p=0.034). The interest of our results lies on the fact that, until now, there were only a few studies that had identified

p53 as a susceptibility gene for the development of CD. Probably, if the size of the sample were larger, these differences would be clearer and it could be appropriate to attempt to reproduce these results in other populations. When analyzing whether this polymorphism was associated to differences regarding the age of presentation, sex or other clinical variables, we did not find statistically significant data.

ATG16L1 (rs2241880)

In 2007, a series of GWAS identified a SNP of *ATG16L1* associated to a higher risk of developing CD: rs2241880[27,28]. This polymorphism induces a change of the A allele for the G allele in position 300 (T300A), and the GG genotype is associated with this disease, particularly when it has an ileal location. The prevalence of this SNP is relatively high in the general population: 58.1% in CD patients, compared with 51.3% in the control group. A German study that included 768 patients with CD confirmed a strong association of this SNP with the risk of having the disease (p=3.7 x 10^{-6}) and also found a significant association with 8 other SNPs (rs13412102, rs12471449, rs6431660, rs1441090, rs2289472, rs2241879, rs3792106, rs4663396), with p-values between 4.1 x 10^{-2} to 3.6 x 10^{-6}. In these 9 SPNs of *ATG16L1*, minor alleles were less common in patients with CD, with OR<1.0[29].

According to our results, the G allele is more common in Crohn's patients than in the control group (53.7% vs. 46.3%, p=0.08). Being a carrier of the A allele suggests a protective effect against CD, as in the study mentioned above[29]. The homozygous GG genotype is the most common one in Crohn's patients (63.5%; p=0.025), and the heterozygous AG genotype is more common in the control group (58.3%). The association between GG and Crohn's disease in this SNP was significant, similarly to what has been described in other series, both in the co-dominant (p=0.007) and the dominant model (p=0.01).

No differences have been found between the different allele variables and the phenotype of the disease, nor with sex or with the age group at diagnosis. The study by Glas et al. did not find a relation between the different genotypes of rs2241880 (T300A) and age at diagnosis; behavior of the disease; presence of extraintestinal manifestations; appearance of stenosis, fistulas or abscesses; or use of IMM[29]. Another British study on a wider population did not find a relation between rs2241880 and a certain phenotype of the disease, either[30], which suggests that the potential effect of this SNP on the risk of developing an ileal disease is much weaker than with certain variants of *NOD2/CARD15*, such as p.Leu1007fsX1008[31].

BAX (rs4645878)

The data we had on *BAX* and *BCL2* in IBD are related to the expression of those proteins in the lamina propria T cells of CD, and they showed a decrease in the expression of BAX and an increase in the BCL-XL/BAX ratio, which would involve a resistance to the apoptosis mechanisms and would ultimately contribute to the chronification of the inflammation[32]. Likewise, a decreased expression of BAX in cases of inflammation of the colonic epithelium has been found in UC, although unlike in CD cases, the BAX/BCL2 system does not appear to be involved in triggering the apoptosis of epithelial cells of the colonic mucosa[33]. However, although our population shows the same data that we already knew from previous studies and different GWAS on *NOD2*, we have not found other studies with candidate genes in IBD which study these polymorphisms of *BAX* (rs4645878) and *BCL2* (rs2279115). Studies on *BAX* (rs4645878) have been carried out for the analysis of hematologic diseases[34] and lung cancer[35], and there are studies on *BCL2* (rs2279115) for hematologic diseases[36] and different kinds of cancer: lung[35]; breast[37], thyroid[38], kidney[39]; and prostate[40].

In our sample, the most common genotype for Crohn's patients is GA (57.4%), and the most common one in the control group is GG (51.1%). The A allele is more common in Crohn's patients than in the control group (53.6% vs. 46.4%), although the percentages are similar and therefore the differences are not statistically significant (p=0.523). Consequently, in our study, being a carrier of the G allele seems to represent a certain protection against the development of CD, compared with being a carrier for the A allele, and no statistical significance has been found in any of the inheritance models. However, this is the first study in our population which attempts to identify *BAX* and *BCL2* as susceptibility genes for having CD, and maybe a study on a larger population would provide more encouraging results in this sense.

BCL2 (rs2279115)

As we have said before, BAX/BCL2 is a system that interacts in apoptosis processes. An increase in the BAX/BCL2 ratio has been proven to promote the programmed cell death of CD4 cells.

Although its association is not statistically significant (p=0.058), the A allele is more common in our Crohn's patients than in the control group (55.3% vs. 44.7%). The homozygous AA genotype is more common among CD patients (65.8%), and the homozygous CC genotype is more common in the control group (56.9%), but once again, the data are not statistically significant.

When calculating OR for the co-dominant, dominant and recessive models, we have observed that the presence of the homozygous AA genotype increases the probability of developing Crohn's disease, compared with the homozygous CC genotype or the heterozygous CA genotype, and the result is statistically significant (p=0.036).

6.3. GENE-GENE INTERACTIONS

We have selected the CART statistical model instead of MARS or logistic regression (LR) for the study of the interaction between genes, because it is the most appropriate model for the statistical analysis of samples like ours with regard to the number of subjects and polymorphisms that are studied.

As we already pointed out in the results section: «*The analysis includes all the genes that were studied in their dominant or recessive forms, according to the significance and magnitude of the odds-ratio that was obtained in the individual analysis of each gene. Therefore, in genes NOD2, p53 and BAX, we have selected the recessive form (or model), and for the genes BCL2 and ATG16L1, the dominant form (or model) was selected*».

In the analysis of interactions, 88.7% of the individuals with the CT or TT genotype for *NOD2* are cases, and 95.8% of patients with the CC genotype for *p53* finally developed the disease, although this last result was not statistically significant (p=0.17). Therefore, the CT/TT genotypes of *NOD2* and the CC genotype of *p53* can classify CD patients with a 94% rate of accuracy, compared with controls (Se=100%; Sp=86%; PPV=92%; NPV=100%).

There is very little scientific evidence with regard to gene-gene interaction in IBD (epistasis). Two recent studies have described an interaction between *NOD2/CARD9* and *REL* in UC[41], and between the three more widely analyzed mutations of *NOD2* (1007finsC, R702W and G908R) and three other genes associated to IBD (*IL23R*, *DLG5* and *OCTN1*)[42]. This last study observed a higher frequency of the three NOD2 mutations among the population with IBD, and confirmed a significant association with CD, but not with UC. In the study of gene-gene interaction, the authors used a statistical method similar to that of our study, which includes the studied genes in their dominant or recessive forms, depending on the significance and the magnitude of the odds ratio obtained in the individual analysis of each gene; and they studied the interaction between the SNPs of *NOD2*, and between these SNPs and other genes. The results revealed a significant association between the different polymorphisms in CD, but not in UC. Another study used a statistical analysis method similar to *CART*, which is logistic regression (LR), and analyzed a predictive model which is the result of the combination of the 71 risk alleles that have been identified in CD through the different GWAS that have been carried out so far. The study also analyzed, with the LR method, the interaction between *NOD2*, *ATG16L1*, *IL10/IL0*, *C13orf31* and *chr*21q; and it found that the combination of all 71 SNPs (in their dominant mode) have a moderate-to-good predictive power for the risk of developing the disease (AUC=0.75), which is higher with a larger amount of risk alleles (AUC=0.77; p<0.0001)[43]. The results of our study reinforce the hypothesis that multiple genes, each of whom provides a small contribution, are involved in the susceptibility to suffer from CD. Also, and like in the

studies that have been mentioned before, individuals who are carriers of a larger amount of risk alleles have a higher risk of developing the disease[44,43].

Once that these genetic interactions have been identified, it is necessary to prove the relevance of these findings at a biological level. According to the statistical model that was used in our study (*CART*), the first two partitions are due to *NOD2* and *p53*, which is explained by the involvement of these genes in the response mechanisms of the host to different microorganisms and in the regulation of the apoptosis mechanisms. In our study, *NOD2* was the first gene that was associated with CD, and the different SNPs affect the Nod domain, which is in charge of the recognition of microorganisms. *p53* is a tumor suppressor gene and it is one of the main regulators of the integrity of the genome and the cell cycle through the different apoptotic pathways. *BAX* and *ATG16L1* are genes related to apoptosis mechanisms and autophagy, respectively, and they are essential elements in the innate immune response to intracellular bacteria. We also know that there is an interaction between *NOD2* and *ATG16L1* in the bacterial sequestration inside autophagosomes[45], although some studies have not been able to reproduce this interaction[46].

More recent studies consider that the defective processing of intracellular bacteria is a determining factor in the pathogenesis of CD. In this sense, the processes of autophagy and programmed cell death (apoptosis) play an essential role. This accounts for the fact that the genes which are involved in these pathways (*ATG16L1, p53, BAX, BCL-2*) are subject to many studies such as this one.

6.4.   STUDY OF CLINICAL VARIABLES AND POLYMORPHISMS

No statistically significant association was found between the polymorphisms that were studied here and the analyzed clinical variables. In previous studies, the attempts to correlate a genotype and phenotype had resulted in some associations in cohort studies, but the results could not be consistently reproduced in wider populations, with the exception of the association, in CD[47], between *NOD2* and an ileal location with stenosing/fistulizing behavior; and, in UC[48], between *HLA DRB*0103* and the extensive and severe forms of the disease. This is probably due to differences in the design of the studies with regard to the type of population, classification criteria, criteria regarding the definition of the studied variables, duration of the disease, or size of the population/analyzed sample. Also, CD is an entity that may vary along time with regard to its behavior, thus the importance of classifying the patients according to stable criteria, such as their genetic factors[17].

We have not found a relation between the different polymorphisms and the need for or response to therapies that are commonly used in CD. In this sense, few studies have been carried out so far which have found a significant relation between genetic

variables and the response to different drugs, or their effectiveness and toxicity. Up to now, we can only talk about a clinical application in the field of pharmacogenetics with regard to the activity of the TPMT enzyme and to the risk of toxicity with AZA[49,50,51].

7. CONCLUSIONS

- The clinical and therapeutic aspects of our series are similar to those described in other populations with Crohn's disease. Highlights the lack of influence of tobacco on the behavior of the disease and a higher use of immunomodulators, although remission rates achieved with these drugs and with anti-TNF are similar to expectations. Half of our patients required surgery, which is in line with previous studies.

- Our study confirms that *NOD2* (rs2066844) and *ATG16L1* (rs2241880) are susceptibility genes for Crohn's disease, and it has identified *p53* (Pro/Arg) as a risk *locus* to develop the disease.

- The prevalence of *NOD2* (rs2066844) was higher than what was described in other populations. With regard to the results on *BCL-2* and *BAX*, it would be necessary to carry out more studies with larger cohorts which may confirm the results that are suggested in our sample.

- No association has been found between the different SNPs and the behavior of the disease, nor with the response to the different therapies that have been used.

- Given the extremely high penetrance observed in the TT allele of *NOD2* (98.8%, OR=43.091), we could include this polymorphism in the diagnostic algorithm of Crohn's disease in cases of uncertain diagnosis. The absence of this polymorphism would mean an unlikely diagnosis of Crohn's disease.

- It is essential to analyze whether these results are confirmed in a larger study of our population.

8. REFERENCES

1. Shanahan, F. Crohn´s disease. <u>Lancet.</u> 2002 Jan 5; 359(9300):62-9.

2. Lakatos, L; Pandur, T; David, G. Association of extraintestinal manifestations of inflammatory bowel disease in a province of western Hungary with disease phenotype: results of a 25-year follow-up study. <u>World J Gastroenterol.</u> 2003 Oct;9(10):2300-7.

3. Cho, JH. The genetics and immunopathogenesis of inflammatory bowel disease. <u>Nat Rev Immunol.</u> 2008 Jun;8(6):458-66. doi: 10.1038/nri2340.

4. Guarner, F; Malagelada, J.R. Role of bacteria in experimental colitis. <u>Best Pract Res Clin Gastroenterol.</u> 2003 Oct;17(5):793-804.

5. Achkar, J.P; Duerr, R. The expanding universe of inflammatory bowel disease genetics. <u>Curr Opin Gastroenterol.</u> 2008 Jul;24(4):429-34. doi: 10.1097/MOG.0b013e3283009c92.

6. Cosnes, J. Epidemiology and natural history of inflammatory bowel diseases. <u>Gastroenterology.</u> 2011 May;140(6):1785-94. doi: 10.1053/j.gastro.2011.01.055.

7. Shugart, Y.Y. An SNP linkage scan identifies significant Crohn's disease loci on chromosomes 13q13.3 and, in Jewish families, on 1p35.2 and 3q29. <u>Genes Immun.</u> 2008 Mar;9(2):161-7. doi: 10.1038/sj.gene.6364460. Epub 2008 Jan 31.

8. Roth, M.P. Familial empiric risk estimates of inflammatory bowel disease in Ashkenazi Jews. <u>Gastroenterology.</u> 1989 Apr;96(4):1016-20.

9. Halme, L. Family and twin studies in inflammatory bowel disease. <u>World J Gastroenterol.</u> 2006 Jun 21;12(23):3668-72.

10. Simmons, J.D. Vitamin D receptor gene polymorphism: association with Crohn's disease susceptibility. <u>Gut.</u> 2000 Aug;47(2):211-4.

11. Tsianos, E.V. Role of genetics in the diagnosis and prognosis of Crohn's disease. <u>World J Gastroenterol.</u> 2012 Jan 14;18(2):105-18. doi: 10.3748/wjg.v18.i2.105.

12. Jostins, L. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. <u>Nature.</u> 2012 Nov 1;491(7422):119-24. doi: 10.1038/nature11582.

13. Sall, J. MonteCarlo calibration of distributions of partition statistics. MonteCarlo calibration of distributions of partition statistics. Technical report, SAS Inst., 2002.

14. Loftus, E.V. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. Gastroenterology. 2004 May;126(6):1504-17.

15. Lowe, A.M. Epidemiology of Crohn's disease in Quebec, Canada. Inflamm Bowel Dis. 2009 Mar;15(3):429-35. doi: 10.1002/ibd.20756.

16. Khan, N. Risk of lymphoma in patients with ulcerative colitis treated with thiopurines: a nationwide retrospective cohort study. Gastroenterology. 2013 Nov;145(5):1007-1015.e3. doi: 10.1053/j.gastro.2013.07.035. Epub 2013 Jul 25.

17. Álvarez-Lobos, M. Combined type-1 plasminogen activator inhibitor and NOD2/CARD15 genotyping predicts complicated Crohn's disease behaviour. Aliment Pharmacol Ther. 2007 Feb 15;25(4):429-40.

18. Jostins, L. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012 Nov 1;491(7422):119-24. doi: 10.1038/nature11582.

19. Cleynen, I. Genetic factors conferring an increased susceptibility to develop Crohn's disease also influence disease phenotype: results from the IBDchip European Project. Gut. 2013 Nov;62(11):1556-65. doi: 10.1136/gutjnl-2011-300777. Epub 2012 Dec 21.

20. Mardini, H.E. Gastroduodenal Crohn's disease is associated with NOD2/CARD15 gene polymorphisms, particularly L1007P homozygosity. Dig Dis Sci. 2005 Dec;50(12):2316-22.

21. Radford-Smith,G. Associations between NOD2/CARD15 genotype and phenotype in Crohn's disease--Are we there yet?. World J Gastroenterol. 2006 Nov 28;12(44):7097-103.

22. Adler, J. The prognostic power of the NOD2 genotype for complicated Crohn's disease: a meta-analysis. Am J Gastroenterol. 2011 Apr;106(4):699-712. doi: 10.1038/ajg.2011.19. Epub 2011 Feb 22.

23. Guagnozzi, D. Acute intestinal obstruction and NOD2/CARD15 mutations among Italian Crohn's disease patients. Eur Rev Med Pharmacol Sci. 2004 Sep-Oct;8(5):179-85.

24. Sakamuro, D. The polyproline region of p53 is required to activate apoptosis but not growth arrest. Oncogene. 1997 Aug 18;15(8):887-98.

25. Lang, S.M. Molecular screening of patients with long standing extensive ulcerative colitis: detection of p53 and Ki-ras mutations by single strand conformation polymorphism analysis and differential hybridisation in colonic lavage fluid. Gut. 1999 Jun;44(6):822-5.

26. Fujii, S. Usefulness of analysis of p53 alteration and observation of surface microstructure for diagnosis of ulcerative colitis-associated colorectal neoplasia. J Exp Clin Cancer Res. 2003 Mar;22(1):107-15.

27. Imielinski, M. Common variants at five new loci associated with early-onset inflammatory bowel disease. Nat Genet. 2009 Dec;41(12):1335-40. doi: 10.1038/ng.489. Epub 2009 Nov 15.

28. Franke, A. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet. 2010 Dec;42(12):1118-25. doi: 10.1038/ng.717.

29. Glas, J. The ATG16L1 gene variants rs2241879 and rs2241880 (T300A) are strongly associated with susceptibility to Crohn's disease in the German population. Am J Gastroenterol. 2008 Mar;103(3):682-91. Epub 2007 Dec 20.

30. Cummings, J.R. Confirmation of the role of ATG16L1 as a Crohn's disease susceptibility gene. Inflamm Bowel Dis. 2007 Aug;13(8):941-6.

31. Seiderer, J. Homozygosity for the CARD15 frameshift mutation 1007fs is predictive of early onset of Crohn's disease with ileal stenosis, entero-enteral fistulas, and frequent need for surgical intervention with high risk of re-stenosis. Scand J Gastroenterol. 2006 Dec;41(12):1421-32.

32. Itoh, J. Decreased Bax expression by mucosal T cells favours resistance to apoptosis in Crohn's disease. Gut. 2001 Jul;49(1):35-41.

33. Iimura, M. Bax is downregulated in inflamed colonic mucosa of ulcerative colitis. Gut. 2000 Aug;47(2):228-35.

34. Zintzaras, E. Synopsis and synthesis of candidate-gene association studies in chronic lymphocytic leukemia: the CUMAGAS-CLL information system. Am J Epidemiol. 2009 Sep 15;170(6):671-8. doi: 10.1093/aje/kwp201. Epub 2009 Aug 21.

35. Gu, S. Association of CASP3 polymorphism with hematologic toxicity in patients with advanced non-small-cell lung carcinoma treated with platinum-based chemotherapy. Cancer Sci. 2012 Aug;103(8):1451-9. doi: 10.1111/j.1349-7006.2012.02323.x. Epub 2012 Jun 18.

36. Park, Y.H. Interaction between BCL2 and interleukin-10 gene polymorphisms alter outcomes of diffuse large B-cell lymphoma following rituximab plus CHOP chemotherapy. Clin Cancer Res. 2009 Mar 15;15(6):2107-15. doi: 10.1158/1078-0432.CCR-08-1588. Epub 2009 Mar 10.

37. Searle, C.J. A BCL2 promoter polymorphism rs2279115 is not associated with BCL2 protein expression or patient survival in breast cancer patients. Springerplus. 2012 Oct 23;1:38. doi: 10.1186/2193-1801-1-38. eCollection 2012.

38. Eun, Y.G. A Polymorphism (rs1801018, Thr7Thr) of BCL2 is Associated with Papillary Thyroid Cancer in Korean Population. Clin Exp Otorhinolaryngol. 2011 Sep;4(3):149-54. doi: 10.3342/ceo.2011.4.3.149. Epub 2011 Sep 6.

39. Hirata, H. The bcl2 -938CC genotype has poor prognosis and lower survival in renal cáncer. J Urol. 2009 Aug;182(2):721-7. doi: 10.1016/j.juro.2009.03.081. Epub 2009 Jun 18.

40. Hirata, H. Bcl2 -938C/A polymorphism carries increased risk of biochemical recurrence after radical prostatectomy. J Urol. 2009 Apr;181(4):1907-12. doi: 10.1016/j.juro.2008.11.093. Epub 2009 Feb 23.

41. McGovern, D.P. Genome-wide association identifies multiple ulcerative colitis susceptibility loci. Nat Genet. 2010 Apr;42(4):332-7. doi: 10.1038/ng.549. Epub 2010 Mar 14.

42. Lin,Z. NOD2 Mutations Affect Muramyl Dipeptide Stimulation of Human B Lymphocytes and Interact with Other IBD-Associated Genes. Dig Dis Sci. 2013 Sep;58(9):2599-607. doi: 10.1007/s10620-013-2696-8. Epub 2013 May 26.

43. Wang, M.H. A novel approach to detect cumulative genetic effects and genetic interactions in Crohn's disease. Inflamm Bowel Dis. 2013 Aug;19(9):1799-808. doi: 10.1097/MIB.0b013e31828706a0.

44. Weersma, R.K. Molecular prediction of disease risk and severity in a large Dutch Crohn's disease cohort. Gut. 2009 Mar;58(3):388-95. doi: 10.1136/gut.2007.144865. Epub 2008 Sep 29.

45. Hampe, J. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. Nat Genet. 2007 Feb;39(2):207-11. Epub 2006 Dec 31.

46. Glas, J. The ATG16L1 gene variants rs2241879 and rs2241880 (T300A) are strongly associated with susceptibility to Crohn's disease in the German population. Am J Gastroenterol. 2008 Mar;103(3):682-91. Epub 2007 Dec 20.

47. Ahmad, T. The molecular classification of the clinical manifestations of Crohn's disease. Gastroenterology. 2002 Apr;122(4):854-66.

48. Thompson, A.I. Genetics of ulcerative colitis. Inflamm Bowel Dis. 2011 Mar;17(3):831-48. doi: 10.1002/ibd.21375. Epub 2010 Nov 12.

49. Schwab, M. Azathioprine therapy and adverse drug reactions in patients with inflammatory bowel disease: impact of thiopurine S-methyltransferase polymorphism. Pharmacogenetics. 2002 Aug;12(6):429-36.

50. Srimartpirom, S. Thiopurine S-methyltransferase genetic polymorphism in the Thai population. Br J Clin Pharmacol. 2004 Jul;58(1):66-70.

51. Lennard, L. TPMT in the treatment of Crohn's disease with azathioprine. Gut. 2002 Aug;51(2):143-6.