CHAPTER 1

INTRODUCTION

1.1 Inflammatory Bowel Disease (IBD)

Inflammatory bowel disease is the common name for two diseases: Crohn's disease (CD) and ulcerative colitis (UC). Both diseases are so similar that sometimes they are indistinguishable and this situation is named as "indeterminate colitis" (Plevy & Reguiero, 2005). Compared to the Asians, this disease is found to have a high prevalence in Western countries and populations, especially among the Jewish, followed by Caucasians and African-American.

The main difference between CD and UC is the affected areas. The former affects almost every possible site of the gastrointestinal tract while the latter affects only the large intestine and rectum. Other than this, other clinical and pathological characteristics are still available to assist in the differentiation between these two diseases (Table 1-1). Other than the intestines, IBD can also affect the eyes, skin, biliary system, ears, joints and bile duct by inducing related disease (Ghanchi & Rembacken, 2003; Weinstock, 2004).

Occurrence of IBD is equal for both genders. However, men is shown to have a higher risk than women in getting UC but the opposite is shown for the CD (Reddy & Wolf, 2004). Monozygotic twins deserve a higher risk compared to dizygotic twins with undisclosed reason (Plevy & Reguiero, 2005). In fertility issues, IBD patients encounter risks in both genders where the male would face a decrease in sperm count, sperm motility and morphological abnormality while the female would experience potential pregnancy difficulties such as preterm birth, low birth weight, morbid offspring and potential abortion (Bröms et al., 2012; Dotan et al., 2012; Hudson et al., 1997; Reddy & Wolf, 2004).

Table 1-1 Comparison between Ulcerative Colitis and Crohn's Disease

	Ulcerative Colitis	**Crohn's Disease**
Macroscopic features		
Rectal involvement	√	
Continuous involvement	√	
Skip regions		√
Terminal ileal ulceration		√
Stricturing		√
Aphthous ulcerations		√
Fistulization / anal fissuring		√
Longitudinal ulcerations		√
Microscopic features		
Crytitis, crypt branching and mucin depletion	√	
Uninterrupted disease	√	
Paneth cell hyperplasia	√	
Patchy involvement		√
Transmural inflammation		√
Granulomas		√
Neuronal hypertrophy		√

Adopted from Weinstock (2004)

The age of IBD patients ranges from 20 to 40, but most patients are in the second and third decade of their lives. Young IBD patients especially children are due to family disease history while the elders are most likely related to environmental factors (Weinstock, 2004). A previous study has shown that the mortality rate of UK-based IBD cohort was higher compared to IBD-free control (Card et al., 2003).

Previous studies have shown that microorganisms play an important role in IBD development. Besides that, habit such as smoking was found to protect against UC but

contribute to the risk of getting CD (Palmon & Mayer, 2006; Thomas et al., 1998) while diet taken daily would relate to IBD occurrence (Andersen et al., 2012). For genetic factors, *NOD2* (nucleotide-binding oligomerization domain-containing 2) was the first gene discovered to be associated with IBD (Hugot et al., 2001; Ogura et al., 2001). This gene was found in 2001 and at the same time it was disclosed to relate only with CD but not with ulcerative colitis. From immunology perspective, regulation of T-cells, dysfunction of cytokine and epithelial are thought to induce the disease (Weinstock, 2004).

In short, environment, genetics and immunology play important roles in this multifactorial disease but the actual etiology of this disease still remain unknown and are yet to be investigated.

1.2 Crohn's Disease (CD)

Crohn's disase (CD) is one of two forms of inflammatory bowel disease (IBD) whereby the other form is ulcerative colitis (UC). They shared similar characteristics and being hard to clearly distinguish until the CD was separated out from the big family after being described in 1932 by Dr Burill B Crohn and his colleagues with 14 cases (Crohn et al., 1932; Palmon & Mayer, 2006). Unlike UC with inflammation that only occurs at the intestine, any part of the gastrointestinal tract of CD patients could be affected by inflammation.

In general, the main symptoms found in CD patients include diarrhea, weight loss and abdominal discomfort. With endoscopy diagnosis, patients can observe inflammatory response along their gastrointestinal tract. Most patients diagnosed with CD are in their 20s or 30s and females have a higher risk compared to the males.

Imbalance in diet would also contribute to the risk (Amre et al., 2007). For smokers, their risk would become higher than nonsmokers, the cigarette quitters also deserved a 65% of risk decrease. For the environmental factor, CD was found to be related to the traffic-related pollutant but it is age-related but only to the young (Kaplan et al., 2010). Besides that, vaccination and breastfeeding were also considered as the risk factors previously but more studies are required for future investigation (Carbonnel et al., 2009).

For children patients, the onset age is around 10 years old (Amre et al., 2006). Young patients' growth rate might be retarded as much as 30% and they may experience a delayed puberty (Buller, 1997). A previous study conducted on relationship between infection-related exposure and CD development among children has shown that children with familial IBD history and owning a pet during young would have a higher risk of getting IBD (Amre et al., 2006). Amre et al. also suggested that those who have better hygiene by owning personal towel and having less siblings could also have their risk reduced.

CD is more common among Westerners compared to Asians. Jewish and Caucasians have a higher susceptibility among Western populations. Although Asian countries such as China, Korea, Japan and Taiwan had fewer reports regarding CD, the trend is increasing recently (Loftus, 2004; Thia et al., 2008). There are some suggestions for the causes of the disease including genetic, environment as well as living style.

For genetics, nucleotide-binding oligomerization domain-containing 2 (*NOD2*) is the first gene found to be related to CD (Hugot et al., 2001; Ogura et al., 2001), the gene contains a C-terminal leucine-rich repeat (LRR), a central nucleotide-binding site

(NBS) and two caspase activation and recruitment domain (CARD), which is important in bacterial invasion defending by recognizing the muramyl dipeptide (MDP) of becteria's peptidoglycan (Inohara, 2003).

There are three *NOD2* mutations being identified: Arg702Trp, Gly908Arg and Leu1007fs. All of them are located within the LRR region: former two mutations involved amino acid substitution while the last one involves a frameshift mutation. Inohara (2003) has shown that the former two mutations could retain little ability to detect MDP while it is totally loss for the mutation Leu1007fs. This shows the mutation would attenuate the transcripted NOD2 protein and affect the immune response. All three mutations especially Leu1007fs have been shown to suppress the interleukin-10 (*IL-10*) production by decreasing the transcription of *IL-10* in mice model (Noguchi et al., 2009). *IL-10* was known to act as an inflammation suppressor and this would lead to attenuated inflammation control.

Cho (2008) showed the mutations of *NOD2* gene are commonly found in the European cohort compared to the African and they are absent in Asian cohort such as Korean, China and Japan. CD was found severe among Caucasians while the trend of disease among countries are rising (Economou & Pappas, 2008) especially in the Western populations. Asian countries have fewer incidents compared to the west, but the cases are rising in Asia as reported by our local researchers (Goh & Xiao, 2009).

In Malaysia, the Indian population has a higher susceptibility on CD compared to the Malays and Chinese (Hilmi et al., 2006). Investigation of *NOD2* was done in 2009 (Chua et al.) but with no association established. Therefore it is interesting to

further investigate other susceptible genes in order to understand more about this disease among the Malaysian population.

CHAPTER 2

LITERATURE

REVIEW

2.1 Nucleotide-binding Oligomerization Domain-containing 1 (*NOD1*)

NOD1 (Gene ID: 10392) functions to recognize bacteria particles. It is located on chromosome 7p14. *NOD1* has been discovered to relate with activation of NF-kB and caspase-9 by Inohara et al. (1999). The same study also showed NOD1 can bind to caspase-9 and is important to induce apotosis, while NF-kB is important in inducing immune response.

NOD1 belongs to NOD-like receptor (NLR) family and is related to *NOD2* which is in the same family. Both *NOD1* and *NOD2* share similar structures but with only one caspase activation and recruitment domain (CARD) found in *NOD1* (Carneiro et al., 2004) instead of two in *NOD2*. *NOD1* is able to recognize γ-D-glutamyl-meso-diaminopimelic acid (iE-DAP) in peptidoglycan to trigger innate immune system. iE-DAP is unable to be detected by *NOD2* protein as iE-DAP only exists commonly in gram negative cells (Chamaillard et al., 2003).

Ferrero and Philpott (2005) have shown that the gene also acts on non-invasive bacteria such as *Helicobacter pylori* which the gene variations were related to asthma and IBD (Hysi, 2005; McGovern, 2005), Furthermore, Chen et al. (2008) showed that the deficiency of *NOD1* can lead to tumor development in the colon of mice which is induced by inflammation. It also serves as a backup system to trigger immune system in cases that the invaders have bypassed the Toll-like receptor (TLR) that recognizes lipopolysaccharides (LPS) of gram negative bacteria (Kim et al., 2004).

Since this gene shares many characteristics with *NOD2*, it grabs the attention of studying its disease susceptibility upon any form of mutation. Inohara et al. (1999) showed that that mutation of this gene further inhibits the activity of NF-kB. It has been

proven that *NOD1* is involved in asthma (Hysi, 2005), intestine protection (Chen et al., 2008; Kara et al., 2010) and gastric cancer (Kupcinskas et al., 2011; Wang et al., 2012) and is related to bacteria attachment and recognition.

For the initial study, SNP E266K (rs2075820) was chosen to be studied for its correlation and the CD susceptibility in the Malaysian population. The SNP is involved in the conversion of glutamic acid (E) to lysine (K) at codon-266 of NOD1 protein. It was shown to be significant for increasing the susceptibility of gastric mucosal inflammation patients towards *Helicobacter pylori* (Kara et al., 2010). There are still many other bacteria species that could be recognized by *NOD1*, such as *L. monocytogenes*, *P. aeruginosa*, *E. coli*, *C. jejuni*, *S. flecneri* and *S. pneumonia* (Kersse et al., 2011). Resulting from that, weak defense may occur without proper mechanism and finally can lead to inflammation.

2.2 Chemokine (C-X-C motif) Ligand 16 (*CXCL16*)

CXCL16 (Gene ID: 58191) encodes chemokine (C-X-C motif) ligand 16 protein and locates on chromosome 17p13. It contains a CXC chemokine domain, transmembrane domain, mucin-like stalk and cytoplasmic tail with tyrosine phosphorylation site. As a member of the CXC chemokine family, it interacts with the CXC chemokine receptors 6 (CXCR6) during the inflammatory process. It is able to recruit CXCR6 expressing cell since it can exist as transmembrane form chemokine. CXCR6 can be found on naïve CD8 cells, intraepithelial lymphocytes (IELs), natural killer T cells (NKT), activated CD8 and CD4 T cells. Three different names have been used to assign to it: STRL33, Bonzo and TYMSTR, which were clarified later that they belong to the same receptor (Matloubian et al., 2000; Wilbanks et al., 2001).

The chemokine family played an important role in connecting innate immune system and adaptive immune system. During immune response, it induces chemotactic respond towards activated CD8 cells. By upregulation of *CXCL16*, CD8 cell will be activated by dendrite cells through inflammatory stimulation (Matloubian et al., 2000). Large quantity of CXCL16 is believed to be involved in the bacterial phagocytosis (Shimaoka et al., 2003).

CXCL16 is made by dendrite cells and cells of the splenic red pulp and expressed in macrophages (Wilbanks et al., 2001), Scholz et al. (2007) showed that CXCL16 is constantly expressed in human and murine keratinocyte cell with participation of ADAM-10. ADAM-10 is a protease of disintegrin and metalloproteinase (ADAM) family, which is responsible in the cleaving of membrane bound form of CXCL16 to yield soluble-CXCL16 and further induce the migration of activated T cells (Abel et al., 2004; Gough et al., 2004). The soluble-CXCL16 (sol-CXCL16) is found to be potentially inducing NF-kB and TNF-α which are important in innate immune response and inflammatory response (Chandrasekar, 2004). The membrane-bound form was believed to help in the adhesion of bacteria towards phagocytosis while soluble form acts to attract CXCR6-positive T cells (Nakase et al., 2012).

While studying its role in IBD, Diegelmann et al. (2010) and Lehrke et al. (2008) have shown the level of CXCL16 was increased in IBD patients compared to the control group and the latter study showed highest CXCL16 level was found in CD patients. Seiderer et al. (2008) studied the Ala181Val (rs2277680) polymorphism in Caucasians IBD patients and they have concluded that this polymorphism might contribute to the severe disease phenotype in CD patient and young patient's onset.

2.3 Signal Transducer and Activator of Transcription 6 (*STAT6*)

STAT6 (Gene ID: 6778) is located on chromosome 12q13 and was discovered by Hamlin et al. (1999). This gene is activated by interleukin-4 (*IL-4)*. It takes part in responding pathway lead by *IL*-4 and IL-*13*. Binding of IL4 and IL-13 to its receptor–IL4-receptor (IL4-R) will induce phosphorylation and dimerization of STAT6, where homodimers formed is translocated into the nucleus and bind to the transcription site of DNA to activate genes involved in Th2 response (Elo et al., 2010). Th2 is responsible for humoral immunity and phagocyte-independent response. Failing of Th2 response might lead to continuous inflammation as Th2 plays another role to inhibit macrophage and Th1 cell that participate in phagocyte-dependent host respond before the response becomes too vigorous and harms host, leading to severe inflammation response (Romagnani, 1999).

Besides that, *STAT6* is also involved in the polarization of Th2 and CD4 lymphocyte, activate mast cell and B-cell promotion. *STAT6* deficiency mice had shown a lack of response from IL-4 signal pathway, such as impaired MHC class II expression, blockage of Th2 cytokine production by parasite, vanishing of immunoglobulin respond as well as T and B cells' proliferate respond (Schindler et al., 2007; Takeda et al., 1997; Wurster et al., 2000). *STAT6* also interacts with some other genes in Th2 regulation, such as *RUNX1*. *RUNX1* plays an important role in regulating Th2 cell differentiation by inhibiting the cell differentiation by binding to the IL-4 silencer region (Naoe et al., 2007) and inhibits the regulation of *GATA3* (Komine et al., 2003). *GATA3* is key transcription factors that regulates expression of Th2 cytokine genes as activation of *STAT6* in Th2 respond require *GATA3* expression.

The study of this gene in IBD was reported by several studies: Zhang et al. (2003) have shown that a 3'-untranslated region (3'UTR) SNP, G2964A (rs324015), was related to the susceptibility in familial CD phenotype while Klein et al. (2005) have shown that the same SNP is related to the susceptibility of CD for Caucasian German origin.

2.4 Toll-Like Receptor 4 (*TLR4*)

TLR4 (Gene ID: 7099) is located on chromosome 9q33 and encodes for toll-like receptor 4 protein, a conserved protein that serves as the main component of the innate immune system. It plays a role in innate immune system by recognizing the lipopolysaccharides (LPS) of gram-negative bacteria. Single nucleotide polymorphisms in this gene are believed to be associated with the disruption of transcripted protein structure. In the signaling pathway, this protein requires other accessory proteins, such as MD2, LPS-binding protein and CD14, a deficiency of any of these accessory proteins in mice will cause poor response of TLR4 (Philpott & Girardin, 2004). Poor response to LPS will lead to low expression as LPS-binding proteins must be present to initiate the reaction (Lavelle et al., 2010). As a product of this signaling pathway, NF-kB will be activated and further triggers the cytokine production.

D299G (rs4986790) and T399I (rs4986791) are another two cosegregate mutations that have been reported to be associated with CD. D299G is a non-synonymous mutation where aspartic acid is replaced by glycine at the 299 location of the protein sequence (Asp299Gly) while T399I is another replacement of threonine by isoleucine at the position of 399. Both mutations are found to lead to weakened TLR4 response to the initially inhaled LPS. Later, several studies have shown that TLR4 is related to other diseases as well, such as renal disease, obesity and diabetes, IBD,

cardiovascular disease, rheumatoid arthritis and Alzheimer's disease (O'Neill et al., 2009).

T399I is found to be related to the early disease onset in Caucasians (Zouiten-Mekki et al., 2009) but a previous research showed that it was significant in Caucasians UC patients (Török et al., 2004). Although the D299G has higher value of research according to few studies, no study has been reported regarding the relationship between T399I and Asian population.

2.5 Autophagy Related 16-Like 1 (*ATG16L1*)

Autophagy related 16-like 1 gene (*ATG16L1*, Gene ID: 55054) is located in chromosome 2q37. Cadwell et al. (2008) showed that ATG16L1 complex (ATG12-ATG5-ATG16L1) is an autophagy protein in human Paneth cell and as well as the intestinal epithelium of mice. In 2003, mouse ATG16L1 protein was discovered and identified together with its isoforms (α, β and γ) after comparing the novel protein found with yeast (Mizushima, 2003). One year later, Zheng et al. (2004) found human *ATG16L1* through large scale sequencing of human fetal brain cDNA library. Human *ATG16L1* span more than 43.9kD and consists of 19 exons with four isoforms.

Saitoh et al. (2008) have shown the deficiency of *ATG16L1* led to the death of mice model which carried deficient *ATG16L1* copy after being exposed to dextran sulphate sodium (DSS), an induced experimental model of colitis. IL1-β, a type of inflammatory cytokines, was detected in high level for the deficient model compared to the wild, suggesting this gene plays an important role in inflammation.

Hampe et al. (2007) first reported the relationship between this gene and CD through his genome wide scanning study on Germany CD cohort and UK CD cohort. The study showed one SNP in exon 9, rs2241880 possess significant effect on CD development and this finding was later supported by Cummings et al. (2007) who conducted the research with the same SNP on UK cohort. In 2008, Kuballa et al. showed CD patients who carried the mutant SNP have shown an impairment in the *Salmonella* capture in autophagosomes. Therefore, this might explain the important role of this gene in the autophagy process since the development of the CD is thought to be due to the failure in microorganisms cleansing.

A recent paper from Van Limbergen et al. (2012) revealed some novel SNPs to provide more possible investigations on this gene. Among the results, some have shown to have complete linkage disequilibrium with rs2241880, therefore they were excluded from the candidate list. Among those that were independent from rs2241880, rs6758317 and rs6754677 were chosen. The former is located in intron 2 while the latter is located in intron 14 of *ATG16L1* gene. Since there is no similar research conducted on the South East Asian population, this gene and three SNPs, rs2241880, rs6758317 and rs6754677 were chosen to be studied in Malaysian cohort.

2.6 Immunity-Related GTPase family M (*IRGM*)

The study of *IRGM* has started since 2009 for its susceptibility investigation to CD. This gene encodes immunity-related GTPase family M protein (Gene ID: 345611) which is believed to be related to the innate immune system of the human body towards invaded pathogens.

Singh (2006) reported that this gene participated in the autophagy process. He inhibited ATG7, one of the autophagy factors in human macrophage cell line to generate a decrease in IFN-γ-induced autophagy. After the cell line was treated with IRGM, the autophagy response restored. *IRGM* also showed participation in LC3-I to LC3-II conversion, which could be terminated by a knockdown of *IRGM*. Endogenous microtubule-associated protein light chain (LC3) exists in two forms: LC3-I found in cytosol and LC3-II found in membrane-associate form. The matured latter form directly participated in the formation of autophagosome.

Recently, Grégoire et al. (2011) reported the novel role of *IRGM*, they found that *IRGM* is one of the most targeted protein by some viruses from *Paramyxoviridae*, *Flaviviridae*, *Orthomyxoviridae*, *Retroviridaea* and *Togaviridae* viral families. The down regulation of *IRGM* using siRNA resulted in more than 65% of viral product increase in measles virus, 70% in hepatitis C virus and 30% in HIV-1, thus showing the importance of *IRGM* in autophagy process among viruses.

Although the disruption mechanism of SNP on *IRGM* protein remains undisclosed, a few SNPs have been revealed to relate with CD susceptibility. Parkes et al. (2007) and Barrett et al. (2008) discovered rs4958847 and rs11747270 located in the intron region of *IRGM b* were associated with CD in corresponded genome wide screening. Their significances were confirmed once again by studies conducted recently by Peter et al. (2011) in Jewish and Weersma et al. (2009) in Dutch.

Recently, Moon et al. (2012) compared *IRGM* sequence among 12 CD patients and 12 controls and manage to detect some other SNPs, amongst include rs72553867. This SNP involved a missense mutation of substitution of threonine by lysine and

discovered with significant association in Korean cohort. However, its effect on protein structure remains unknown. Since it was reported in the Asian cohort, it was also chosen to investigate on Malaysian data.

2.7 Objectives

This research aims to achieve the following objectives:

1. To generate genetic data for patients with Crohn's disease in Malaysia.
2. To investigate the association between candidates genes and Crohn's disease in the Malaysian population.

The null hypothesis apply in this research is "The chosen SNP has no association with CD population in Malaysia".

CHAPTER 3

MATERIALS AND METHODS

3.1 Materials

3.1.1 Agarose Gel

Two percent (w/v) agarose gel was prepared by mixing 2g of agarose powder (Promega) in 100ml 1X TBE and heated in a microwave oven (Sanyo). One microliter EtBr was added and mixed well before gel casting. Appropriate amount of agarose powder was used to prepare particular percentage of agarose gel.

3.1.2 1X Tris-Borate-EDTA (TBE) buffer

One hundred milliliter of 10X TBE stock (1st BASE) was mixed with 900ml of distilled water and 1X TBE buffer was ready for gel casting purpose and as electrophoresis buffer.

3.2 Methods

3.2.1 Blood samples

A total of 85 CD patients and 250 healthy individuals were recruited in this research and the details are shown in Table 3-1. All bloods samples were obtained from the University Malaya Medical Centre (UMMC), Kuala Lumpur, Malaysia. This study was approved by UMMC Ethics Review Board (approval number: 472.55) and informed consents were obtained from all participants.

Table 3-1 Composition of CD and control samples

Race	CD (%)	Controls (%)
Malay	19 (22.4)	74 (29.6)
Chinese	25 (29.4)	86 (34.4)
Indian	34 (40.0)	90 (36.0)
Others	7 (8.2)	0 (0)
Total	85 (100)	250 (100)

3.2.2 DNA Extraction

All collected blood samples were extracted using QIAamp DNA Mini Kit (Qiagen). Twenty microliter QIAGEN Protease was pipetted into a 1.5ml microcentrifuge tube before 200µl blood was added into the tube. Two hundred microliter Buffer AL provided in kit was added to the sample and mixed well by vortexing for 15 seconds. The mixture was then incubated for 10 minutes at 56^{o}C. Following that, 200µl ethanol was added to the samples and mixed well by vortex.

The mixture was then transferred to a QIAamp Mini spin column as provided without wetting the rim, followed by centrifugation at 6000xg for 1 minute. QIAamp Mini spin column was removed and placed in a new 2ml collection tube. Five hundred milliliter of Buffer AW1 was added to a QIAamp Mini spin column without wetting the rim and centrifuged at 6000xg for 1 minute. The flow through was discarded and the spin column was transferred to a new 2ml collection tube. Five hundred microliter Buffer AW2 was added into the spin column and centrifuged again at full speed for three minutes.

The QIAamp Mini spin column was the placed in a clean 1.5ml microcentrifuge tube. One hundred microliter double-distilled water was added into the QIAamp Mini spin column and incubated for one minute at room temperature. The spin column was

then centrifuged at 6000xg for one minute to elute the DNA, extracted DNA was then quantified.

3.2.3 DNA Quantification

DNA quantification was done using Implen NanoPhotometer. The OD readings at 260nm and 280nm wavelength were recorded. High purity DNA was obtained when the OD_{260}/OD_{280} reading is in a range of 1.8 – 2.0.

3.2.4 Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP)

3.2.4.1 Polymerase Chain Reaction (PCR) Optimization

Primer pairs obtained from various literatures were used in this study to perform PCR. The general protocol is showed in Table 3-2. For annealing temperature, the range was set from 60°C to 70°C. PCR product obtained was then electrophoresed using 2% (w/v) agarose gel and viewed under UV transilluminator. The temperature with the best product yielded was chosen for subsequent screening.

Table 3-2 PCR protocol applied in PCR reaction

Process		Temperature (°C)	Time
Initial denaturation		95.0	12 minutes
PCR (35 cycles)	Denaturation	94.0	30seconds
	Annealing	60-70	30seconds
	Extension	72.0	30seconds
Final extension		72.0	7 minutes
Soaking		15.0	∞

3.2.4.2 Polymerase Chain Reaction (PCR) Amplification

A single PCR reaction containing 18µl of PCR master mix as shown in Table 3-3 was added with 2µl of DNA template to give a final reaction volume of 20µl. A healthy individual DNA was used as positive control while distilled water was served as DNA blank. The primers involved for each gene amplification are listed in Table 3-4 with the sources as well. The preparation of bulk volume master mix was done on ice to maintain at low temperature. After 18µl of single reaction mix was aliquoted into each 0.2ml reaction tube, DNA templates, positive control DNA and distilled water were added accordingly to separate tubes. PCR reactions were carried out using Bio-Rad C1000 Thermal Cycler. The condition of PCR cycle is shown in Table 3-5.

Table 3-3 Single reaction PCR mix

Components	Volume (µl)
10X Dream Taq buffer	2.0
50µM forward primer	0.15
50µM reverse primer	0.15
10mM each dNTP mix	0.15
5U/µl Dream Taq Polymerase	0.15
sdH_2O	15.4
Total	18.0

Table 3-4 Primer pairs used in PCR reaction

Gene (SNP)	Primers (5'-3')	Reference
NOD1 (rs2075820)	F: TGA GAC CAT CTT CAT CCT GG R: CTT CCC ACT GAG CAG GTT G	Molnar et al. (2007)
CXCL16 (rs2277680)	F: ACT CGT CCC AAT GAA ACC AC R: CCA CAG CT T CAT CTC CCA CT	Lundberg et al. (2005)
STAT6 (rs324015)	F: GAA GTT CAG GCT CTG AGA GAC R: AGA ATG GGC GGA GAA GCC T	Klein et al. (2005)
TLR4 (rs4986791)	F: GGT TGC TGT TCT CAA AGT GAT TTT GGG AGA A R: GGA AAT CCA GAT GTT CTA GTT GTT CTA AGC C	Török et al. (2004)

Table 3-5 PCR settings applied in PCR reaction

Process		Temperature (^{o}C)	Time
Initial denaturation		95.0	12 minutes
PCR (35 cycles)	Denaturation	94.0	30 seconds
	Annealing	64^{o}C (rs2075820, rs2277680) 67.5^{o}C (rs324015) 60^{o}C (rs4986791)	30 seconds
	Extension	72.0	30 seconds
Final extension		72.0	7 minutes
Soaking		15.0	∞

3.2.4.3 PCR Product Agarose Gel Electrophoresis

A 2% (w/v) agarose gel was used to separate the PCR products, 4µl PCR product was mixed with 1µl 6X loading dye (Fermentas) and loaded into the wells of gel that had stained with EtBr. A 50bp or 100bp DNA ladder (Thermo Scientific) was used as DNA marker for size reference. After electrophoresis at 140V for 15 minutes, the gel was viewed under UV transilluminator and the PCR products were then subjected to restriction enzyme digestion.

3.2.4.4 Restriction Enzyme (RE) Digestion

A single master mix of restriction enzyme reaction is shown in Table 3-6. All reactions were incubated at 37^{o}C overnight before analyzed using agarose gel electrophoresis. Restriction enzyme used for each SNP is shown in Table 3-7 as well as the sequence recognized and the product sizes before digestion.

Table 3-6 Single reaction mix used in restriction enzyme digestion

Solutions	Volume (µl)
PCR products	3.0
10X Restriction enzyme buffer	1.0
10U/µl Restriction enzyme	0.5
Distilled water	11.0
Total	15.5

Table 3-7 Restriction enzyme used for each SNP and its recognition sequence

Gene (SNP)	Restriction Enzyme	PCR Product Size (bp)	Sequence Recognized
NOD1 (rs2075820)	*Ava*I	379	5'...C↓Y C G R G...3' 3'...G R G C Y↑C...5'
CXCL16 (rs2277680)	*Alu*I	499	5'...A G↓C T...3' 3'...T C↑G A...5'
STAT6 (rs324015)	*Hin1*I	159	5'...G R↓C G Y C...3' 3'...C Y G C↑R G...5'
TLR4 (rs4986791)	*Hinf*I	124	5'...G↓A N T C...3' 3'...C T N A↑G...5'

3.2.4.5 Agarose Gel Electrophoresis of Restriction Fragments

Seven microliter of the digested product was mixed with 1.5µl of 6X loading dye (Fermentas) and loaded into the wells of gel stained with EtBr. A DNA ladder 100bp (Thermo Scientific) was used as DNA marker for subsequent reference. Digested products of *NOD1* (rs2075820) and *CXCL16* (rs2277680) were electrophoresed on a 2% (w/v) agarose gel while *STAT6* (rs324015) and *TLR4* (rs4986791) digested products were analyzed on a 3.5% (w/v) agarose gel to determine the presence of restricted fragments. After electrophoresis at 140V for 20 minutes, the gel was viewed under UV transluminator and the genotype for each sample was recorded. Table 3-8 shows the details of digested products using different restriction enzymes.

Table 3-8 Interpretation of details after enzyme digestion

Gene (SNP)	Restriction Enzyme	PCR Product Size (bp)	Product Size (bp)	Genotype
NOD1 (rs2075820)	*Ava*I	379	209, 170	Homozygous Wild (G)
			379	Homozygous Mutant (A)
			379, 209, 170	Heterozygous
CXCL16 (rs2277680)	*Alu*I	499	428, 49, 16b, 6	Homozygous Wild (C)
			477, 16, 6	Homozygous Mutant (T)
			477, 428, 49, 16, 6	Heterozygous
STAT6 (rs324015)	*Hin1*I	159	159	Homozygous Wild (A)
			139, 21	Homozygous Mutant (G)
			159, 139, 21	Heterozygous
TLR4 (rs4986791)	*Hinf*I	124	124	Homozygous Wild (C)
			95, 29	Homozygous Mutant (T)
			124, 95, 29	Heterozygous

3.2.4.6 PCR Product Sequencing

In order to examine the amplification of precise PCR products, four sequencing primer sets were designed covered the four SNPs along with their PCR primers. The sequences were shown in Table 3-9. Samples from each SNP covered homozygous and heterozygous genotype were chosen randomly to be sequenced. Master mix was prepared according to Table 3-3 and PCR protocol applied is shown in Table 3-10.

Table 3-9 Primers designed for PCR product sequencing

Gene (SNP)	**Primers (5'-3')**
NOD1 (rs2075820)	F: GTG ACT CCA AGT TCG TGC TG R: CCT ATG CCA GGA GGA TGT TCC
CXCL16 (rs2277680)	F: ATG CTT ACT CGG GGA TTG TGG C R: GCT ATC ATC CCC CAA ACA CTG TCC
STAT6 (rs324015)	F: ACC TGC TCT GGA CAC TTG CT R: AGC ATA GGA GGG GTC TGG AC
TLR4 (rs4986791)	F: CCC TGG TGA GTG TGA CTA T R: GTG AAG ATA TCT GGA AGG AAG

Table 3-10 PCR setting applied in sequencing primers

Process		**Temperature (^{o}C)**	**Time**
Initial denaturation		95.0	12 minutes
PCR (35 cycles)	Denaturation	94.0	30 seconds
	Annealing	65^{o}C (rs2075820) 67^{o}C (rs2277680) 68^{o}C (rs324015) 60^{o}C (rs4986791)	30 seconds
	Extension	72.0	30 seconds
Final extension		72.0	7 minutes
Soaking		15.0	∞

A 2% (w/v) agarose gel was used to separate the PCR product, 4µl of the PCR product was mixed with 1µl of 6X loading dye (Fermentas) and loaded into the wells of gel stained with EtBr. A 50bp or 100bp DNA ladder (Thermo Scientific) was used as DNA marker to determine the fragment size. After electrophoresis at 140V for 15 minutes, the gel was viewed under a UV transilluminator.

The PCR products for sequencing were purified using QIAquick PCR Purification Kit (Qiagen). Eighty microliter of the Buffer PB provided was mixed with 16µl of PCR product and the mixture was transferred to a QIAquick spin column. The mixture was then centrifuged at 17,900xg for 1 minute. The flow-through was discarded

and 0.75ml of Buffer PE was added and centrifuged again at 17,900xg for 1 minute. The flow-through was discarded and the QIAquick column was placed in a clean 1.5ml microcentrifuge tube and 50µl of sterile distilled water was added to the spin column, incubated at room temperature for 1 minute and centrifuged for 1 minute at 17,900xg. The solution containing DNA collected was sent for sequencing.

3.2.5 TaqMan SNP Genotyping Assay

3.2.5.1 TaqMan Assay Amplification

Variants located at *ATG16L1* (rs2241880, rs6758317 and rs6754677) and *IRGM* (rs4958847, rs11747270 and rs72553867) were genotyped using TaqMan probes (Applied Biosystem). Single reaction mix was prepared based on the protocol provided as shown in Table 3-11.

Table 3-11 Single reaction mix prepared for TaqMan amplification

Components	**Volume (µl)**
TaqMan® GTXpress™ Master Mix (2X)	5.0
TaqMan genotyping assay mix (20X)	0.5
Distilled water	3.5
Sample DNA (10ng/µl)	1.0
Total	10.0

One microliter of sample DNA was added to 9µl of Taqman reaction mixture and DNA blank was also prepared using distilled water. The plate was then sealed tightly with MicroAmp Optical Film. The plate was centrifuged briefly to eliminate air bubbles trapped and subjected to PCR (7500 Fast Real-Time PCR System, Applied Biosystem). The PCR protocol is showed in Table 3-12.

Table 3-12 PCR protocol used in TaqMan amplification

		Temperature (^{o}C)	Time
Pre-PCR Read		25.0	60 seconds
Holding		95.0	20 seconds
PCR (40 cycles)	Denaturation	95.0	3 seconds
	Annealing/ Extension	60.0	30 seconds
Post-PCR Read		25.0	60 seconds

Each of amplifications was analyzed using 7500 Software v2.0.4 (Applied Biosystem). The reporter dyes used in this research were VIC and 6-carboxyfluorescein (FAM) while the nonfluorescent quencher-dihydrocyclopyrroloindole tripeptide minor groove binder (NFQ-MGB) as the quencher. The details of allele reported by each dye are shown in Table 3-13.

Table 3-13 Bases assigned to different reporter dye

Gene	**Fluorophore**	
ATG16L1	**VIC (Allele 1)**	**FAM (Allele 2)**
rs2241880	A	G
rs6758317	C	T
rs6754677	A	G
IRGM	**VIC (Allele 1)**	**FAM (Allele 2)**
rs4958847	A	G
rs11747270	A	G
rs72553867	A	C

Source: http://www.appliedbiosystems.com.sg

The result was prepared in allelic discrimination plot. Each genotype detected was recorded and analysis was done by obtaining the genotype and allele frequency.

3.2.6 Statistical Analysis

Both the genotypic and allelic count was obtained for each SNP. As part of case-control study, comparison between patient and control data was done by Fisher exact test to investigate any significance. The odd ratios (OR) and 95% confidence interval (CI) were also calculated. Hardy-Weinberg equilibrium was checked for every SNP on control cohort. Any significant SNP found would undergo genotype association test.

3.2.6.1 Fisher Exact Test

This test was first described by Fisher (1922) from Pearson χ^2 test, differ from Pearson, it is applicable for small sample size. In conventional rule, from expected number perspective, Pearson χ^2 test is usually apply to those having at least five samples while Fisher exact test can apply to sample size less than five. Other than this, Pearson χ^2 test only demonstrates an approximate true *P*-value in small sample study and this can be solved by Fisher test as it gives an exact probability by calculating the actual chance.

Since there are small values obtained during this research, therefore this test, instead of Pearson χ^2 test, was applied in this study. Since null hypothesis of this test stated that first variable is independent from the second variable, thus null hypothesis was defined as “The SNP has no association in CD population”, where first variable was allele types and second variable would be the CD population.

The computation of *P*-value is showed in Appendix A, *P*-value is obtained by calculating the value of probability of each event in order to produce an outcome. It is simpler to be presented in formula where *a*, *b*, *c* and *d* are observed values and *n* is the number of total sample.

$$p = \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{a!\,b!\,c!\,d!\,n!}$$

However, Fisher exact test is advised only applicable to a 2x2 contingency table although there was a study showed that it could be extended to a particular table size (Mehta & Patel, 1983). Freeman-Halton extension of the Fisher exact test can be considered to analyze bigger table size, with sample size less than 300 (Freeman & Halton, 1951). In this research, it was used to calculate the significance of allele distribution among Malaysian as well as races with a significant level of P=0.05. Analysis of bigger table would be done whenever necessary.

3.2.6.2 Odd Ratios (OR)

The odd ratios (OR) is commonly used in case-control studies due to its convenience and estimation for relationship between two binary data with presence of confidence interval (CI) (Bland & Altman, 2000). "Odd" means a chance for an event to occur under particular condition. The ratio is computed by getting the number of interested event (*P*) divided by the number of uninterested event (*Q*). An OR=1 indicates the event has the equal chance to happen in two studied groups. However, the OR>1 showed event *P* is more likely to happen and event *Q* is more preferable if the OR<1. In this research, an allele with an OR>1 suggests it is a risk allele to the disease

while OR<1 suggests it conferred lesser risk to the disease. Calculation of OR is showed in Appendix B.

3.2.6.3 Confidence Interval (CI)

A 95% confidence interval (CI) means in a 100 samples, 95 of them would have the desired parameter. It is a quality control parameter and can be narrowed down by having lower percentage of accuracy desired, low standard error and large sample size. In this research, a CI range value that exclude 1 indicate the OR ratio is significant which with *P*=0.05 as border (Sim & Reid, 1999). If OR=1 is included within its 95% range, then there is no significant association discovered. Calculation of CI is showed in Appendix B.

3.2.6.4 Hardy-Weinberg Equilibrium

Hardy-Weinberg equilibrium was described in 1908 to explain the balance of genetic in nature (Dorak, 2012). Without condition factors such as non-random mating, genetic shifting, gene mutation and population migration, the inheritance of traits are thought to be balance from one generation to next generation. As this equilibrium is hard to be obeyed since random mating is hardly to be carried out by men, instead it could be used to study the effect of those factors in a real population.

In this research, Hardy-Weinberg equilibrium is used to examine the genetically balance of samples selected. This was to ensure no bias in selection which would lead to error in data collection. In order to obtain expected population for each data set, the following formula was applied:

$$p^2+2pq+q^2 = 1$$

where p represents wild type allele frequency and q represents mutant type allele frequency. The computation of expected population frequency is showed in Appendix C. If the expected frequency is close or exactly the same with observed, the samples then announced in equilibrium or else experienced disequilibrium which could be contributed by a few situations. Analysis of Hardy-Weinberg equilibrium only applied on control cohort that with random selection while patient cohort is not suitable due to specific selection for analysis purpose. In order to determine the significance of difference between observed and expected genotype frequency, the level of difference was calculated using Arlequin 3.0 software and a significant level of 0.05 was applied.

3.2.6.5 Genotype Association Test

There are three types of common models that known to apply in genotype data analysis: dominant, recessive and additive (Lewis, 2002). In dominant model, since a single copy out of two will increase the risk or induce the disease, the genotypes with mutant type (*mtwt*+*mtmt*) alleles are compared with homozygous wild (*wtwt*) during analysis. For recessive model, the homozygous mutant (*mtmt*) is compared with genotypes with wild type allele (*wtwt*+*mtwt*) since two copies are required in order to express its effect (Lewis & Knight, 2012). For additive model, it is to compare homozygous mutant (*mtmt*) with homozygous wild (*wtwt*) due to the gradually increase of risk with one or more copy an individual has carried. The details were showed in Appendix D.

The genotypes data were compared by implying Fisher exact test to determine existence of significant association. This test was only applied for those SNPs found with significant association with. Suitable model is chosen by considering result from every cohort with significant *P*-value.

Figure 3-1 showed the overview of methodology in this study arranged in flow chart.

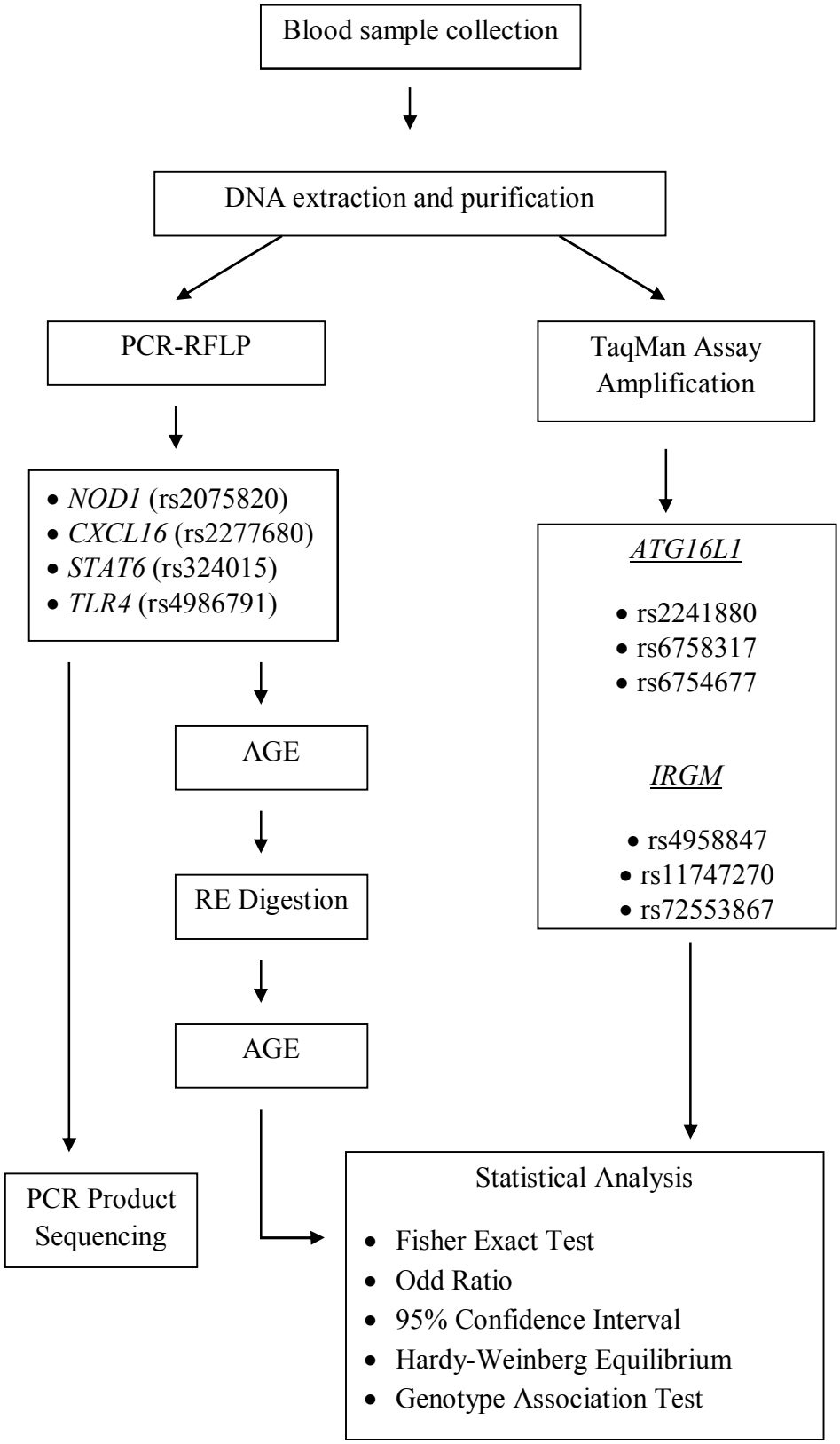


Figure 3-1 Overview of Study Methodology

CHAPTER 4

RESULTS

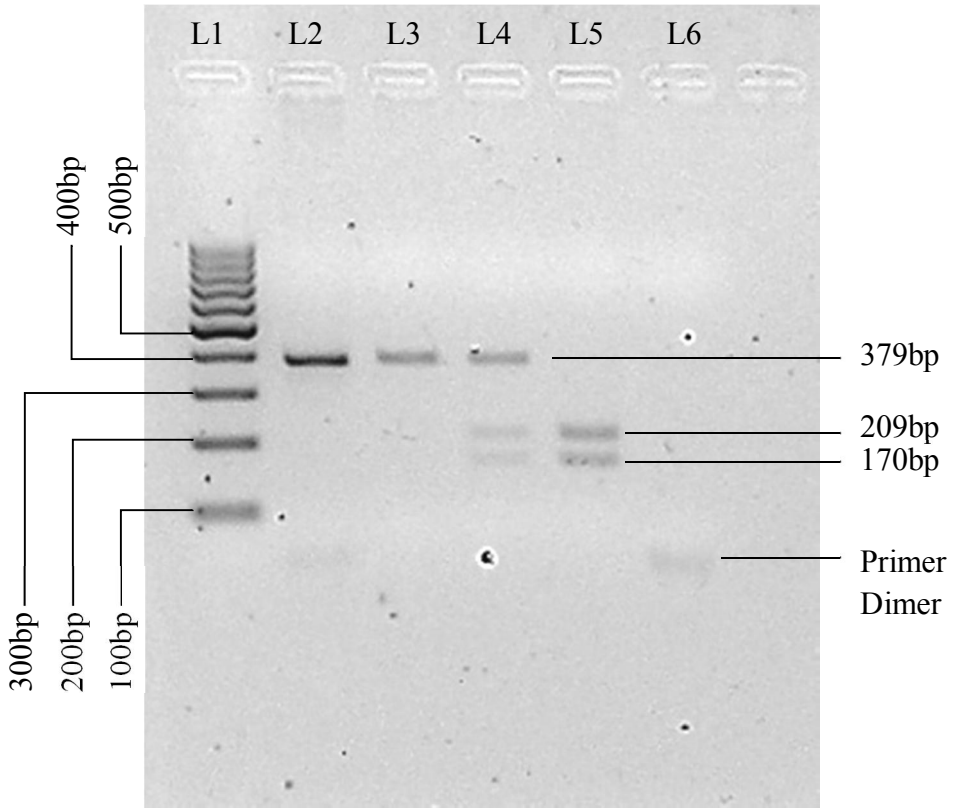
4.1 Hardy-Weinberg Equilibrium Analysis

All sample cohorts were in equilibrium after analyzing with Hardy-Weinberg equation using Arlequin 3.0 except *TLR4* rs4986791 and *IRGM* rs4958847 had disequilibrium with Malaysian cohort (Appendix F) while *ATG16L1* rs6754677 was disequilibrium with Malay cohort (Appendix G).

4.2 Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP)

4.2.1 Analysis of *NOD1* rs2075820

After the digestion of the PCR amplicon using *Ava*I, the sizes of restricted fragments was tallied with what was reported previously (Molnar et al., 2007). As shown in Figure 4-1, homozygous mutant type A allele could not be digested while the homozygous wild type G allele could be digested. From the result, the 379bp PCR amplicon was digested and resulted in 209bp and 170bp fragments for wild type genotype.



L1: 100bp DNA ladder
L2: Positive control (control DNA)
L3: Homozygous mutant type A allele (379bp)
L4: Heterozygous A/G (from top to bottom – 379bp, 209bp and 170bp)
L5: Homozygous wild type G allele (from top to bottom – 209bp and 170bp)
L6: Non-template control (DNA blank)

Figure 4-1 Gel electrophoresis of digested *NOD1* rs2075820 PCR amplicon

Figure 4-2 shows the electropherogram of a heterozygous A/G genotype. Homozygous genotype results are not shown since only a single peak was present at the SNP site (Appendix E). Double peaks which consist of A (bottom) and G (top) were observed for rs2075820 SNP. This indicates the correct allele variation was detected. The sequence of PCR amplicon with wild type G allele is shown in Figure 4-3. In Figure 4-3, the sequence can be digested by *Ava*I at the restriction site 5'-CCCGAG-3'.

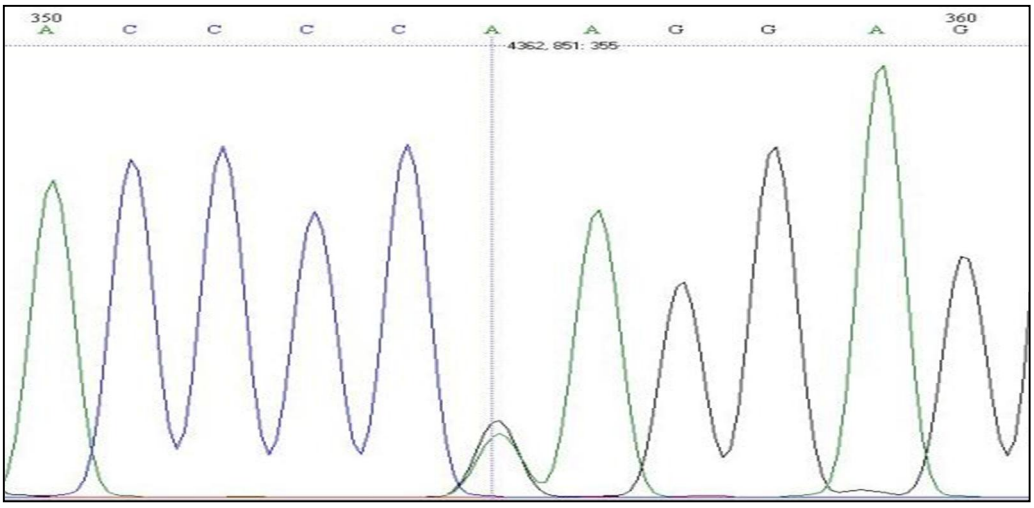


Figure 4-2 Sequencing result for *NOD1* rs2075820 heterozygous genotype

TGAGACCATCTTCATCCTGGGTGATGCTGGGGTGGGCAAGTCCATGCTGCTAC

AGCGGCTGCAGAGCCTCTGGGCCACGGGCCGGCTAGACGCAGGGGTCAAATTC

TTCTTCCACTTTCGCTGCCGCATGTTCAGCTGCTTCAAGGAAAGTGACAGGCT

GTGTCTGCAGGACCTGCTCTTCAAGCACTACTGCTACCCAGAGCGGGACC↓CC

GAGGAGGTGTTTGCCTTCCTGCTGCGCTTCCCCCACGTGGCCCTCTTCACCTT

CGATGGCCTGGACGAGCTGCACTCGGACTTGGACCTGAGCCGCGTGCCTGACA

GCTCCTGCCCCTGGGAGCCTGCCCACCCCCTGGTCTTGCTGGCCAACCTGCTC

AGTGGGAAG

Legend

Underlined nucleotides: PCR primer sequence

Bolded nucleotide: SNP site

Arrow: Restriction site

Figure 4-3 Wild type PCR amplicon sequence of *NOD1* rs2075820

Data from both patient and control groups were calculated and shown in Table 4-1. Since the 95% CI with OR=1 was in their range for both alleles, there was no effect of this variant on the disease. Fisher exact test of P=0.5179 showed no significant association for the SNP with CD, thus null hypothesis was not rejected. Overall, the controls have higher homozygous genotype (39.2% and 12.0% for wild and mutant respectively) compared to patients which have higher heterozygous genotype (55.3%).

Table 4-1 *P*-value, OR And 95% CI obtained for *NOD1* rs2075820

SNP	Genotype Count	Patients (n=85) (%)		Controls (n=250) (%)	
NOD1 rs2075820	GG	33 (38.8)		98 (39.2)	
	GA	47 (55.3)		122 (48.8)	
	AA	5 (5.9)		30 (12.0)	
	Allelic Count	**Patients (n=85) (%)**	**Controls (n=250) (%)**	**Odd Ratio (95% Confidence Interval)**	***P*-value**
	Wild type (G)	113 (66.5)	318 (63.6)	1.135 (0.786 - 1.637)	0.5179
	Mutant type (A)	57 (33.5)	182 (36.4)	0.881 (0.611 - 1.272)	

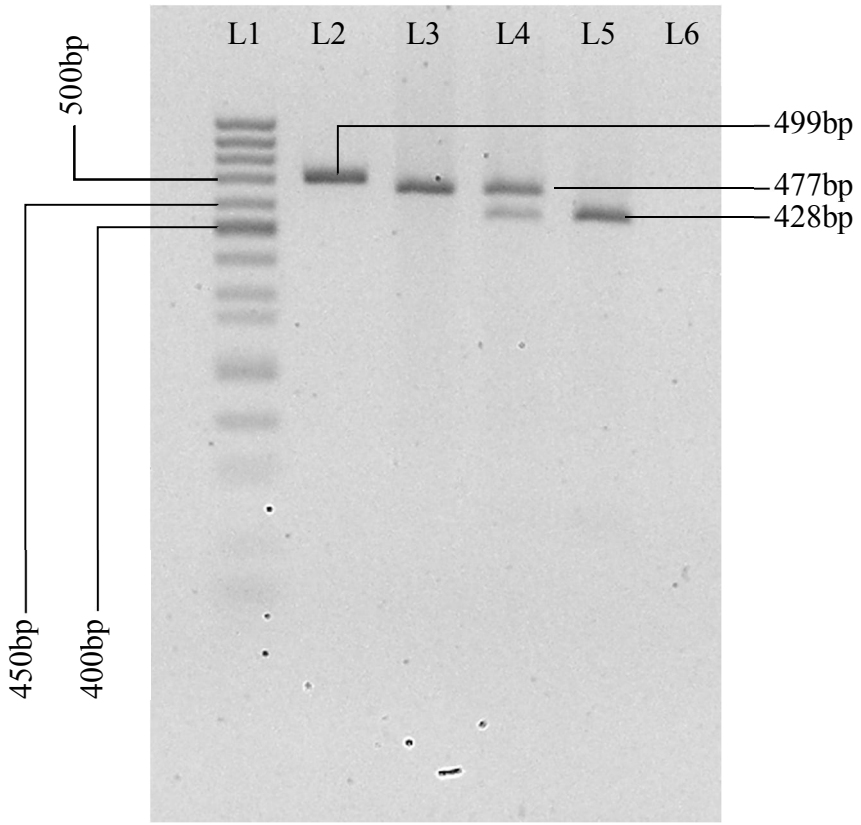
Table 4-2 shows the statistical analysis of individual race control cohort. All the three races of patient samples harbored higher wild type G allele (Malay=71.1%, Chinese=72%, Indian=63.2%) compared to control samples (Malay=68.9%, Chinese=68%, Indian=55%). No significant association was found for all the three races as $P>0.05$ after Fisher exact test. Overall, the SNP rs2075820 did not confer risk to CD in the Malaysian population.

Table 4-2 Data and statistical analysis of *NOD1* rs2075820 for different races

Race		Patients		Controls		*P*-value	OR (95% CI)
Malay	**Genotype**	**Count**	**%**	**Count**	**%**		
	GG	10	52.6	36	48.6	0.9334	1.172 (0.4274 – 3.218)
	GA	7	36.8	30	40.5		0.8556 (0.3020 – 2.424)
	AA	2	10.5	8	10.8		0.9706 (0.1885 – 4.996)
	Allele	**Count**	**%**	**Count**	**%**		
	G	27	71.1	102	68.9	0.8464	1.107 (0.5061 – 2.422)
	A	11	28.9	46	31.1		0.9034 (0.4130 – 1.976)
Chinese	**Genotype**	**Count**	**%**	**Count**	**%**		
	GG	12	48.0	38	44.2	0.8853	1.166 (0.4776 – 2.847)
	GA	12	48.0	41	47.7		1.013 (0.4154 – 2.471)
	AA	1	4.0	7	8.1		0.4702 (0.0551 – 4.015)
	Allele	**Count**	**%**	**Count**	**%**		
	G	36	72.0	117	68.0	0.7287	1.209 (0.6029 – 2.423)
	A	14	28.0	55	32.0		0.8273 (0.4127 – 1.659)
Indian	**Genotype**	**Count**	**%**	**Count**	**%**		
	GG	10	29.4	24	26.7	0.1207	1.146 (0.4785 – 2.744)
	GA	23	67.6	51	56.7		1.599 (0.6967 – 3.669)
	AA	1	2.94	15	16.7		0.1515 (0.0192 – 1.195)
	Allele	**Count**	**%**	**Count**	**%**		
	G	43	63.2	99	55.0	0.2537	1.407 (0.7929 – 2.498)
	A	25	36.8	81	45.0		0.7106 (0.4003 – 1.261)
Others	**Genotype**	**Count**	**%**	**Count**	**%**		
	GG	1	14.3	0	0	Not Calculated	Not Calculated
	GA	5	71.4	0	0		
	AA	1	14.3	0	0		
	Allele	**Count**	**%**	**Count**	**%**		
	G	7	50.0	0	0	Not Calculated	Not Calculated
	A	7	50.0	0	0		

4.2.2 Analysis of *CXCL16* rs2277680

The digestion of *CXCL16* wild type C allele amplicon yielded four products but only one band (428bp) could be seen. Other bands (6bp, 16bp and 49bp) were unable to be observed on the gel. For amplicon carried the mutant type T allele, three products formed but only a 477bp fragment could be observed while the rest (6bp and 16bp) could not be observed on the gel (Figure 4-4).



L1: 50bp DNA ladder
L2: Positive control (control DNA)
L3: Homozygous mutant type T allele (477bp)
L4: Heterozygous C/T (from top to bottom – 477bp and 428bp)
L5: Homozygous wild type C allele (from top to bottom – 428bp)
L6: Non-template control (DNA blank)

Figure 4-4 Gel electrophoresis of digested *CXCL16* rs2277680 PCR amplicon

Figure 4-5 shows the variant site of heterozygous sample consisted of double peak of wild type C allele (top) and mutant type T allele (bottom). This indicates the correct site was amplified, the sequence containing wild type C allele is shown in Figure 4-6. In general, there are three restriction sites for the wild type C allele while

the mutant type T allele only contains two restriction sites. Short fragments (6bp, 16bp and 49bp) were unable to be observed on the gel. Therefore, only two bands were seen for wild type allele and one for mutant type allele. PCR amplicon sequence shown in Figure 4-6 can be digested into four fragments as it contains three restriction sites 5'-AGCT-3'.

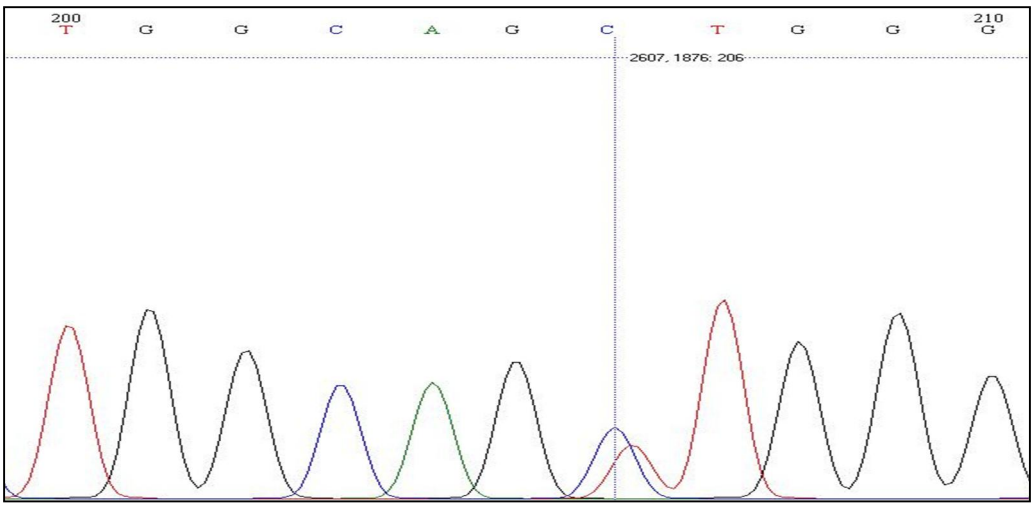


Figure 4-5 Sequencing result for *CXCL16* rs2277680 heterozygous genotype

ACTCGTCCCAATGAAACCACCATTCACACTGCGGGCCACAGTCTGGCAG↓**C**TGG

GCCTGAGGCTGGGGAGAACCAGAAGCAGCCGGAAAAAAATGCTGGTCCCACAGC

CAGGACATCAGCCACAGTGCCAGTCCTGTGCCTCCTGGCCATCATCTTCATCCT

CACCGCAGCCCTTTCCTATGTGCTGTGCAAGAGGAGGAGGGGGCAGTCACCGCA

GTCCTCTCCAGGTAAACTGGACCTTTGAACCCTCCCCCCAGGGACCCAGAGCCT

TCCTGACAGGCATCCTGACCCAGTGGCTCTCAGGTGGCTGGAAACCTGGCCTTG

GCACACAAGTATCAAGGTTGTCTGCACAAGGTTGACTCTGGGCCCTCTGAGGAC

AGTCAGTCGGGGACTGTAGACAGTGGAGAGGGAAGGAAGCCACCACAGGCACAA

GGACTAGTGCTTGGAGCAACATCATGGAGAAGATGGCTAGAGCAAG↓CTAGTGG

GAGATGAAG↓CTGTGG

Legend

Underlined nucleotides: PCR primer sequence

Bolded nucleotide: SNP site

Arrows: Restriction sites

Figure 4-6 Wild type PCR amplicon sequence of *CXCL16* rs2277680

According to Table 4-3, this SNP was found to be significantly associated with CD, with *P*=0.04819, in which the mutant type T allele significantly conferred protection against CD by having OR=0.699. In genotype count, the frequency of patients bearing homozygous mutant genotype (27.1%) was lower when compared to the control cohort (35.6%).

Table 4-3 *P*-value, OR And 95% CI obtained for *CXCL16* rs2277680

SNP	**Genotype Count**	**Patients (n=85) (%)**		**Controls (n=250) (%)**	
CXCL16 rs2277680	CC	21 (24.7)		39 (15.6)	
	CT	41 (48.2)		122 (48.8)	
	TT	23 (27.1)		89 (35.6)	
	Allelic Count	**Patients (n=85) (%)**	**Controls (n=250) (%)**	**Odd Ratio (95% Confidence Interval)**	***P*-value**
	Wild type (C)	83 (48.8)	200 (40.0)	1.431 (1.009 - 2.031)	0.04819
	Mutant type (T)	87 (51.2)	300 (60.0)	0.699 (0.493 - 0.992)	

Fisher exact analysis of individual race in Table 4-4 did not show significant result. This proposed that the SNP was only significant to Malaysian pooled cohort who has treated as one population and no association could be established between any individual race with the disease. As observed Malay and Chinese patient cohort harbored more wild type C allele (55.3% and 40% respectively) compared to control cohort (37.8% and 34.3% respectively). For Indian samples, they had higher frequency of wild type allele in the controls (47.2% vs. patient 45.6%).

Table 4-4 Data and statistical analysis of *CXCL16* rs2277680 for different races

Race		Patients		Controls		*P*-value	OR (95% CI)
Malay	**Genotype**	**Count**	**%**	**Count**	**%**		
	CC	6	31.6	8	10.8	0.08873	3.333 (0.925 – 12.01)
	CT	9	47.4	40	54.1		0.765 (0.2786 – 2.100)
	TT	4	21.1	26	35.1		0.4923 (0.148 – 1.637)
	Allele	**Count**	**%**	**Count**	**%**		
	C	21	55.3	56	37.8	0.06492	2.029 (0.9872 – 4.172)
	T	17	44.7	92	62.2		0.493 (0.2397 – 1.013)
Chinese	**Genotype**	**Count**	**%**	**Count**	**%**		
	CC	5	20.0	9	10.5	0.4142	2.139 (0.645 – 7.093)
	CT	10	40.0	41	47.7		0.7317 (0.296 – 1.809)
	TT	10	40.0	36	41.9		0.9259 (0.374 – 2.295)
	Allele	**Count**	**%**	**Count**	**%**		
	C	20	40.0	59	34.3	0.5035	1.277 (0.6682 – 2.440)
	T	30	60.0	113	65.7		0.7832 (0.410 – 1.497)
Indian	**Genotype**	**Count**	**%**	**Count**	**%**		
	CC	6	17.6	22	24.4	0.6086	0.662 (0.2426 – 1.808)
	CT	19	55.9	41	45.6		1.514 (0.6843 – 3.349)
	TT	9	26.5	27	30.0		0.84 (0.3466 – 2.036)
	Allele	**Count**	**%**	**Count**	**%**		
	C	31	45.6	85	47.2	0.8868	0.9364 (0.535 – 1.639)
	T	37	54.4	95	52.8		1.068 (0.6101 – 1.869)
Others	**Genotype**	**Count**	**%**	**Count**	**%**		
	CC	4	57.1	0	0	Not Calculated	Not Calculated
	CT	3	42.9	0	0		
	TT	0	0.0	0	0		
	Allele	**Count**	**%**	**Count**	**%**		
	C	11	78.6	0	0	Not Calculated	Not Calculated
	T	3	21.4	0	0		

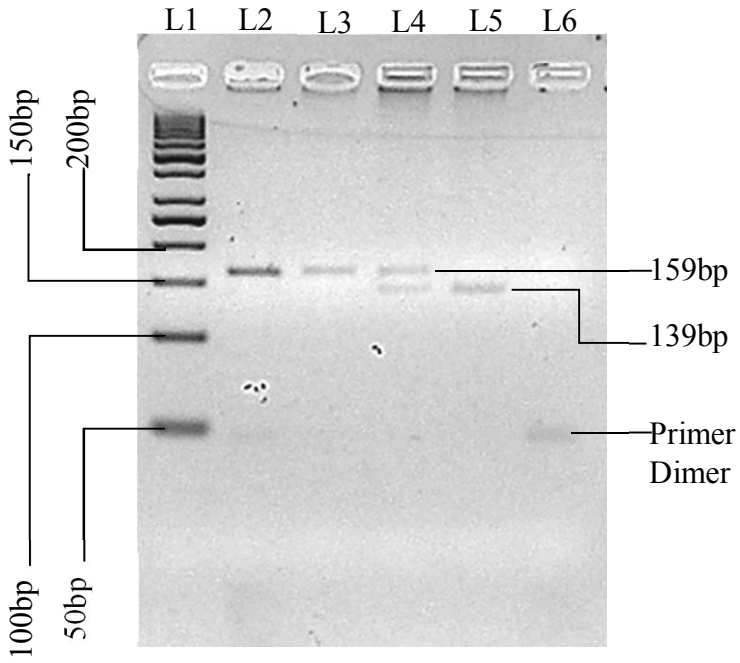
As this SNP was found to be significant with the Malaysian CD, the genotype frequency was subjected genotype association test. Table 4-5 shows the analysis result, where additive model with *P*=0.0448 was applied to the effect of this SNP.

Table 4-5 Genotype association analysis (additive model) for Malaysian *CXCL16* rs2277680

Cohort	**Genotype**		***P*-value**
	TT	**CC**	
CD	23	21	0.0448
Control	89	39	

4.2.3 Analysis of *STAT6* rs324015

After being digested with *Hin1*I, the 159bp wild type A allele amplicon remained intact while mutant type G allele amplicon was cleaved and only the 139bp band was observed since the 20bp digested product was too small to be seen on the gel (Figure 4-7).



L1: 50bp DNA ladder

L2: Positive control (control DNA)

L3: Homozygous wild type A allele (159bp)

L4: Heterozygous A/G (from top to bottom – 159bp and 139bp)

L5: Homozygous mutant type G allele (from top to bottom – 139bp)

L6: Non-template control (DNA blank)

Figure 4-7 Gel electrophoresis of digested *STAT6* rs324015 PCR amplicon

Figure 4-8 shows the electropherogram result. Nucleotide site denoted with 'R' indicates the existence of heterozygous genotype with both A and G alleles, which was the target SNP site. The higher peak is wild type A allele while the lower peak is mutant type G allele. Sequence of amplicon with mutant type G allele is shown in Figure 4-9. There was a modified nucleotide in PCR primer (italicized in Figure 4-9) which helped to form the restriction site for enzyme digestion. This amplicon could be then digested since sequence 5'-GACGCC-3' existed.

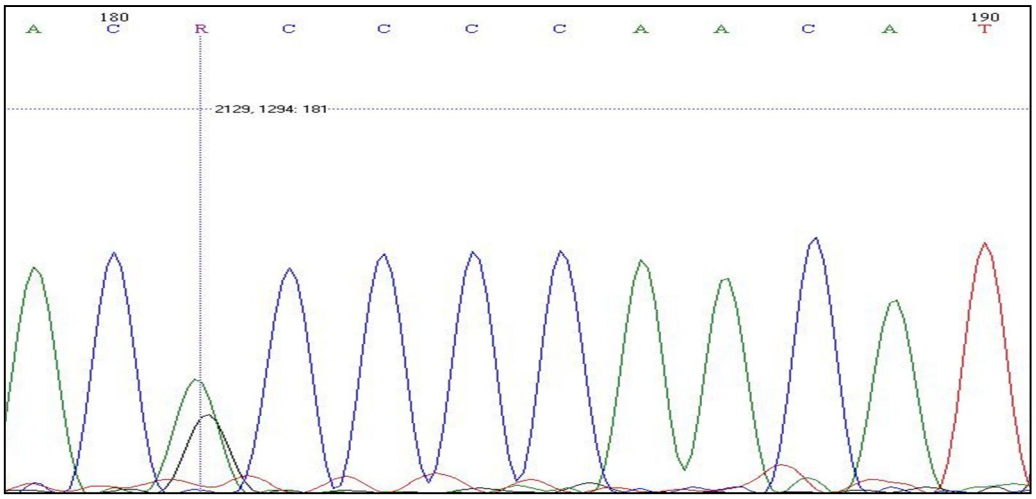


Figure 4-8 Sequencing result for *STAT6* rs324015 heterozygous genotype

GAAGTTCAGGCTCTGAGA*G*A↓C**G**CCCCAACATGCCTGCACCTGCAGCGCG

CACACGCACGCACACACACATACAGAGCTCTCTGAGGGTGATGGGGCTGA

GCAGGAGGGGGGCTGGGTAAGAGCACAGGTTAGGGCATGGAAGGCTTCTC

CGCCCATTCT

Legend

Underlined nucleotides: PCR primer sequence

Bolded nucleotide: SNP site

Italicized nucleotide: Modified nucleotide for enzyme restriction

Arrow: Restriction site

Figure 4-9 Mutant type PCR amplicon sequence of *STAT6* rs324015

Analysis of allelic count for rs324015 is shown in Table 4-6. Fisher exact test showed no significant association between the two alleles and the disease (P=0.5310). Both alleles did not have effect on the onset of the disease as shown by the OR analysis since 95% CI for both ORs included 1. For genotype count analysis, 21.2% controls carried homozygous wild type A allele (16.5% in patients) and 32.8% carried homozygous mutant type G allele (34.1% in patients).

Table 4-6 *P*-value, OR And 95% CI obtained for *STAT6* (rs324015)

SNP	**Genotype Count**	**Patients (n=85) (%)**		**Controls (n=250) (%)**	
STAT6 rs324015	AA	14 (16.5)		53 (21.2)	
	AG	42 (49.4)		115 (46.0)	
	GG	29 (34.1)		82 (32.8)	
	Allelic Count	**Patients (n=85) (%)**	**Controls (n=250) (%)**	**Odd Ratio (95% Confidence Interval)**	***P*-value**
	Wild type (A)	70 (41.2)	221 (44.2)	0.884 (0.621 - 1.258)	0.5310
	Mutant type (G)	100 (58.8)	279 (55.8)	1.132 (0.795 - 1.610)	

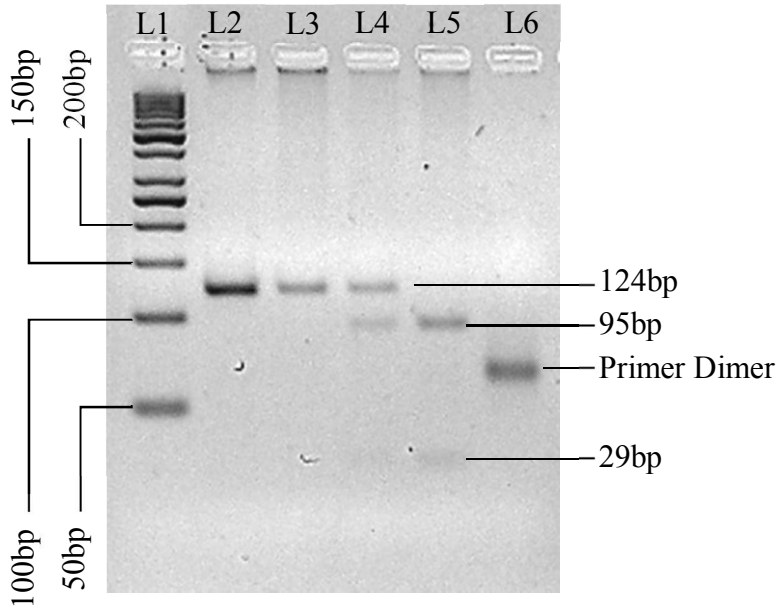
In the association analysis of individual race, no significant result was found as shown in Table 4-7. Malay, Chinese and Indian control cohorts gave *P*-values of 0.4596, 0.7491 and 0.3112 respectively. This concluded that SNP rs324015 had no association with the Malaysian CD as well as CD in individual race.

Table 4-7 Data and statistical analysis of *STAT6* rs324015 for different races

Race		Patients		Controls		*P*-value	OR (95% CI)
Malay	**Genotype**	**Count**	**%**	**Count**	**%**		
	AA	1	5.3	16	21.6	0.2017	0.2014 (0.025 – 1.626)
	AG	11	57.9	31	41.9		0.8477 (0.351 – 2.046)
	GG	7	36.8	27	36.5		1.015 (0.3569 – 2.889)
	Allele	**Count**	**%**	**Count**	**%**		
	A	13	34.2	63	42.6	0.3636	0.7016 (0.333 – 1.478)
	G	25	65.8	85	57.4		1.425 (0.6765 – 3.003)
Chinese	**Genotype**	**Count**	**%**	**Count**	**%**		
	AA	6	24.0	24	27.9	0.4556	0.816 (0.2907 – 2.289)
	AG	14	56.0	36	41.9		1.768 (0.7198 – 4.341)
	GG	5	20.0	26	30.2		0.577 (0.1954 – 1.703)
	Allele	**Count**	**%**	**Count**	**%**		
	A	26	52.0	84	48.8	0.7491	1.135 (0.6043 – 2.131)
	G	24	48.0	88	51.2		0.881 (0.4692 – 1.655)
Indian	**Genotype**	**Count**	**%**	**Count**	**%**		
	AA	4	11.8	13	14.4	0.6986	0.790 (0.2385 – 2.615)
	AG	16	47.1	48	53.3		0.778 (0.3528 – 1.715)
	GG	14	41.2	29	32.2		1.472 (0.6527 – 3.322)
	Allele	**Count**	**%**	**Count**	**%**		
	A	23	33.8	74	41.1	0.3112	0.732 (0.4084 – 1.312)
	G	45	66.2	106	58.9		1.366 (0.762 – 2.448)
Others	**Genotype**	**Count**	**%**	**Count**	**%**		
	AA	3		0	0	Not Calculated	Not Calculated
	AG	2		0	0		
	GG	2		0	0		
	Allele	**Count**	**%**	**Count**	**%**		
	A	8		0	0	Not Calculated	Not Calculated
	G	6		0	0		

4.2.4 Analysis of *TLR4* rs4986791

The *Hinf*I did not digest wild type C allele amplicon and hence a 124bp band was observed on the agarose gel. However, two fragments of 95bp and 29bp were seen for mutant type T allele amplicon as the result of *Hinf*I digestion. However, only the 95bp fragment was visualized on the gel while the short digested product of 29bp was not obvious.



L1: 50bp DNA ladder
L2: Positive control (control DNA)
L3: Homozygous wild type C allele (124bp)
L4: Heterozygous C/T (from top to bottom – 124bp, 95bp and 29bp)
L5: Homozygous mutant type T allele (from top to bottom – 95bp and 29bp)
L6: Non-template control (DNA blank)

Figure 4-10 Gel electrophoresis of digested *TLR4* rs4986791 PCR amplicon

As shown in Figure 4-11, the two peaks indicated the co-existence of wild type C allele (blue arrow) and mutant type T allele (red arrow) in a heterozygous sample. The sequence of amplicon with mutant allele is shown in Figure 4-12. There was a modified nucleotide in the PCR primer (italicized in Figure 4-12) which served as a restriction site for enzyme digestion as the genome sequence has no restriction site available. The presence of sequence 5'-GAATC-3' in the amplicon allows the digestion of T allele site by *Hinf*I.

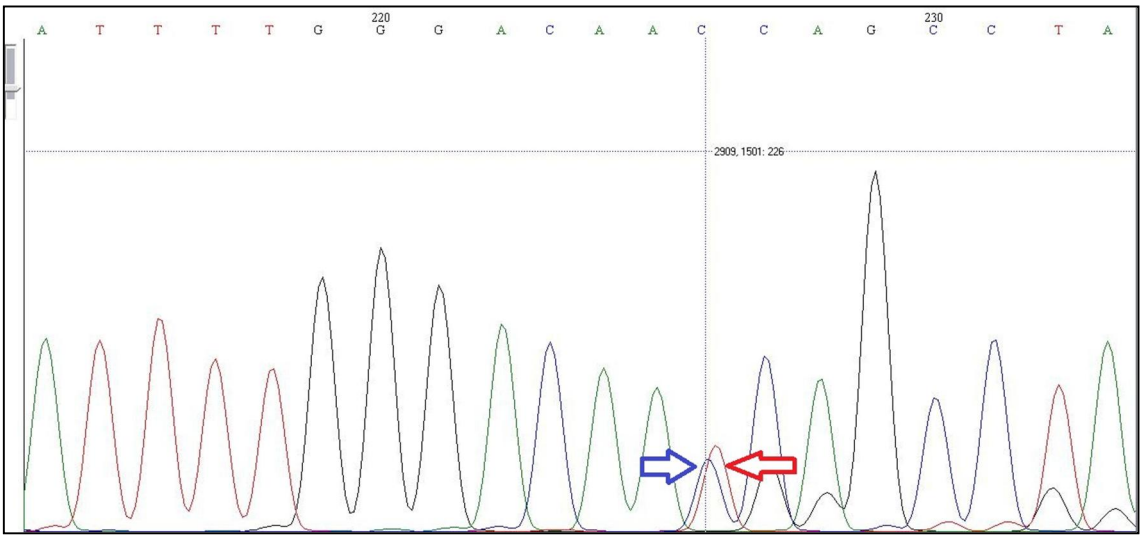


Figure 4-11 Sequencing result for *TLR4* rs4986791 heterozygous genotype

GGTTGCTGTTCTCAAAGTGATTTTGGGA*G*↓AA**T**CAGCCTAAAGTATTTAGAT

CTGAGCTTCAATGGTGTTATTACCATGAGTTCAAACTTCTTGGGCTTAGAAC

AACTAGAACATCTGGATTTCC

Legend

Underlined nucleotides: PCR primer sequence

Bolded nucleotide: SNP site

Italicized nucleotide: Modified nucleotide for enzyme restriction

Arrow: Restriction site

Figure 4-12 Mutant type PCR amplicon sequence of *TLR4* rs4986791

Fisher exact test for rs4986791 gave a *P*-value of 0.006886 which demonstrated significant association between this SNP and the Malaysian CD. As shown in Table 4-8, wild type C allele conferred protective effect against CD with an OR<1 while mutant type T allele increased the risk by having OR>1. In genotype analysis, control cohort carried more homozygous genotype (90% for wild and 1.6% for mutant) than the patient cohort which obtained higher frequency of heterozygous genotype (22.4% vs. control 8.4%). However, the Hardy-Weinberg disequilibrium of samples (*P*=0.00521), as shown in Table 4-9, might deny the reliability of this result and further analysis is required to investigate the deviation from the expected frequency.

Table 4-8 *P*-value, OR And 95% CI obtained for *TLR4* rs4986791

SNP	**Genotype Count**	**Patients (n=85) (%)**		**Controls (n=250) (%)**	
TLR4 rs4986791	CC	65 (76.5)		225 (90.0)	
	CT	19 (22.4)		21 (8.4)	
	TT	1 (1.2)		4 (1.6)	
	Allelic Count	**Patients (n=85) (%)**	**Controls (n=250) (%)**	**Odd Ratio (95% Confidence Interval)**	***P*-value**
	Wild type (C)	149 (87.6)	471 (94.2)	0.437 (0.242 - 0.789)	0.006886
	Mutant type (T)	21 (12.4)	29 (5.8)	2.289 (1.268 - 4.134)	

Table 4-9 Hardy-Weinberg equilibrium of control samples in *TLR4* rs4986791

SNP	**Genotype Count**	**Observed Frequency**	**Expected Frequency**	***P*-value**
TLR4 rs4986791	Homozygous Wild	0.900	0.887	0.00521
	Heterozygous	0.084	0.109	
	Homozygous Mutant	0.016	0.003	

The stratified analysis of rs4986791 among the three races of control samples uncovered a significant association of CD only with Malay population (P<0.0005) on genotype as well as allele analysis but neither Chinese (no result due to monomorphic) nor Indian (P=0.3382) (Table 4-10). This showed the significance of rs4986791 to CD was related with Malay population. Besides, P-value for genotype analysis within Malays also yielded a P<0.0005 which indicated appearance of any mutant type T allele was crucial in CD development for them.

Table 4-10 Data and statistical analysis of *TLR4* rs4986791 for different races

Race		Patients		Controls		*P*-value	OR (95% CI)
Malay	**Genotype**	**Count**	**%**	**Count**	**%**		
	CC	12	63.2	72	97.3	$1.47x10^{-4}$	0.0476 (0.0088 – 0.257)
	CT	7	36.8	2	2.7		21 (3.8899 – 113.37)
	TT	0	0.0	0	0.0		Not Calculated
	Allele	**Count**	**%**	**Count**	**%**		
	C	31	81.6	146	98.6	$2.4 x10^{-4}$	0.061 (0.012 – 0.306)
	T	7	18.4	2	1.4		16.48 (3.267 – 83.18)
Chinese	**Genotype**	**Count**	**%**	**Count**	**%**		
	CC	25	100.0	86	100.0	Not Calculated	Not Calculated
	CT	0	0.0	0	0.0		Not Calculated
	TT	0	0.0	0	0.0		Not Calculated
	Allele	**Count**	**%**	**Count**	**%**		
	C	50	100.0	172	100.0	Not Calculated	Not Calculated
	T	0	0.0	0	0.0		Not Calculated
Indian	**Genotype**	**Count**	**%**	**Count**	**%**		
	CC	21	61.8	67	74.4	0.2830	0.555 (0.240 – 1.282)
	CT	12	35.3	19	21.1		2.038 (0.857 – 4.849)
	TT	1	2.9	4	4.4		0.652 (0.0702 – 6.046)
	Allele	**Count**	**%**	**Count**	**%**		
	C	54	79.4	153	85.0	0.3382	0.681 (0.333 – 1.393)
	T	14	20.6	27	15.0		1.469 (0.7178 – 3.006)
Others	**Genotype**	**Count**	**%**	**Count**	**%**		
	CC	7	100.0	0	0	Not Calculated	Not Calculated
	CT	0	0.0	0	0		
	TT	0	0.0	0	0		
	Allele	**Count**	**%**	**Count**	**%**		
	C	14	100.0	0	0	Not Calculated	Not Calculated
	T	0	0.0	0	0		

Since the Malaysian genotype frequency has significant disequilibrium of Hardy-Weinberg, thus its genotype association test was not conducted. Race analysis showed that the distribution of *TLR4* rs4986791 in the Malay cohort was significantly associated with CD. As shown in Table 4-11, the mutant type T allele was found with dominant effect with a P=1.47x10^{-4}.

Table 4-11 Genotype association analysis (dominant model) for Malay *TLR4* rs4986791

Cohort	**Genotype**		***P*-value**
	CT + TT	**CC**	
CD	7	12	1.47x10^{-4}
Control	2	72	

4.3 TaqMan Assay Amplification

4.3.1 Analysis of *ATG16L1* Variants

4.3.1.1 Analysis of rs2241880

One of the variant of *ATG16L1*, rs2241880 obtained significant result of *P*=0.01816 in pooled Malaysian cohort and thus showed that this SNP is associated with CD (Table 4-12). The wild type A allele conferred protective effect against CD by having an OR value of 0.645 while the mutant type G allele exerted greater risk to carrierby having an OR value of1.550. Among the three genotypes, 20% of the patients harbored homozygous mutant G genotype while 24.7% with homozygous wild A genotype, meanwhile controls harbored only 13.2% and 39.2% respectively.

Table 4-12 *P*-value, OR And 95% CI obtained for *ATG16L1* rs2241880

SNP	**Genotype Count**	**Patients (n=85) (%)**		**Controls (n=250) (%)**	
ATG16L1 rs2241880	AA	21 (24.7)		98 (39.2)	
	AG	47 (55.3)		119 (47.6)	
	GG	17 (20.0)		33 (13.2)	
	Allelic Count	**Patients (n=85) (%)**	**Controls (n=250) (%)**	**Odd Ratio (95% Confidence Interval)**	***P*-value**
	Wild type (A)	89 (52.4)	315 (63.0)	0.645 (0.454 - 0.917)	0.01462
	Mutant type (G)	81 (47.6)	185 (37.0)	1.550 (1.090 - 2.202)	

Fisher exact analysis of the individual race cohort is shown in Table 4-13 whereby only the Malay cohort was found to be significantly associated with CD. All three races were found to have high frequency of mutant type G allele in the patients (Malay=50%, Chinese=30%, Indian=57.4%) compared to the controls (Malay=31.1%, Chinese=29.7%, Indian=48.9%). Malay population gave a *P*=0.03619 after allele analysis with Fisher exact test and its genotype frequency also fulfilled Hardy-

Weinberg equilibrium. Both Chinese and Indian populations showed insignificant results for Fisher exact test (P=1.0 and P=0.2566 respectively).This suggested that the significant association between rs2241880 and Malays might contributed to its significant association with the Malaysian cohort.

Table 4-13 Data and statistical analysis of *ATG16L1* rs2241880 for different races

Race		Patients		Controls		*P*-value	OR (95% CI)
Malay	**Genotype**	**Count**	**%**	**Count**	**%**		
	AA	4	21.1	34	45.9	0.07415	0.314 (0.0951 – 1.035)
	AG	11	57.9	34	45.9		1.618 (0.5839 – 4.482)
	GG	4	21.1	6	8.1		3.022 (0.7579 – 12.05)
	Allele	**Count**	**%**	**Count**	**%**		
	A	19	50.0	102	68.9	0.03619	0.451 (0.2184 – 0.931)
	G	19	50.0	46	31.1		2.217 (1.074 – 4.578)
Chinese	**Genotype**	**Count**	**%**	**Count**	**%**		
	AA	11	44.0	41	47.7	0.9364	0.862 (0.3521 – 2.113)
	AG	13	52.0	39	45.3		1.306 (0.535 – 3.186)
	GG	1	4.0	6	7.0		0.556 (0.0637 – 4.845)
	Allele	**Count**	**%**	**Count**	**%**		
	A	35	70.0	121	70.3	1.0	0.984 (0.4944 – 1.956)
	G	15	30.0	51	29.7		1.017 (0.5112 – 2.023)
Indian	**Genotype**	**Count**	**%**	**Count**	**%**		
	AA	4	11.8	23	25.6	0.2656	0.388 (0.1235 – 1.222)
	AG	21	61.8	46	51.1		1.545 (0.6903 – 3.459)
	GG	9	26.5	21	23.3		1.183 (0.4785 – 2.924)
	Allele	**Count**	**%**	**Count**	**%**		
	A	29	42.6	92	51.1	0.2566	0.711 (0.4053 – 1.248)
	G	39	57.4	88	48.9		1.406 (0.8011 – 2.468)
Others	**Genotype**	**Count**	**%**	**Count**	**%**		
	AA	2	28.6	0	0	Not Calculated	Not Calculated
	AG	2	28.6	0	0		
	GG	3	42.9	0	0		
	Allele	**Count**	**%**	**Count**	**%**		
	A	6	42.9	0	0	Not Calculated	Not Calculated
	G	8	57.1	0	0		

Tables 4-14 and 4-15 showed genotype analysis for Malaysian sample and Malay cohort respectively. Since the Malaysian population has significant outcomes for both dominant and additive models, thus the model of Malay population was required to verify the results. As the latter showed association with additive model, therefore this SNP was proposed to have an additive effect.

Table 4-14 Genotype association analysis (dominant and additive model) for Malaysian *ATG16L1* rs2241880

Cohort	**Genotype**		***P*-value**
	AG + GG	**AA**	
CD	64	21	0.01808
Control	152	98	
Cohort	**Genotype**		***P*-value**
	GG	**AA**	
CD	17	21	0.02639
Control	33	98	

Table 4-15 Genotype association analysis (additive model) for Malay *ATG16L1* rs2241880

Cohort	**Genotype**		***P*-value**
	GG	**AA**	
CD	4	4	0.04712
Control	6	34	

4.3.1.2 Analysis of rs6758317

The second SNP, rs675831 in *ATG16L1* did not show significant association with CD in the Malaysian population by having a *P*=0.3763 (Table 4-16) after Fisher exact analysis. Its alleles had no significant effect on CD onset with both 95% CI covered OR=1 in their value range. In genotype analysis, patients harbored higher percentage of homozygous mutant T (4.7%) and heterozygous CT (35.3%) genotypes while controls carried higher frequency of homozygous wild CC genotype (66%).

Table 4-16 *P*-value, OR And 95% CI obtained for *ATG16L1* rs6758317

SNP	**Genotype Count**	**Patients (n=85) (%)**		**Controls (n=250) (%)**	
ATG16L1 rs6758317	CC	51 (60.0)		165 (66.0)	
	CT	30 (35.3)		74 (29.6)	
	TT	4 (4.7)		11 (4.4)	
	Allelic Count	**Patients (n=85) (%)**	**Controls (n=250) (%)**	**Odd Ratio (95% Confidence Interval)**	***P*-value**
	Wild type (C)	132 (77.6)	404 (80.8)	0.825 (0.540 - 1.261)	0.4373
	Mutant type (T)	38 (22.4)	96 (19.2)	1.212 (0.793 - 1.851)	

Analysis of rs6758317 in three races in Malaysia did not obtain significant value. As shown in Table 4-17, Fisher exact analysis gave *P*=0.3506, *P*=0.6383 and *P*=1.0 for Malay, Chinese and Indian populations respectively. In conclusion, no association was established between the disease and this SNP in the three main races in Malaysia.

Table 4-17 Data and statistical analysis of *ATG16L1* rs6758317 for different races

Race		Patients		Controls		*P*-value	OR (95% CI)
Malay	**Genotype**	**Count**	**%**	**Count**	**%**		
	CC	11	57.9	50	67.6	0.34	0.66 (0.235 – 1.853)
	CT	7	36.8	23	31.1		1.294 (0.4508 – 3.712)
	TT	1	5.3	1	1.4		4.056 (0.2419 – 68.0)
	Allele	**Count**	**%**	**Count**	**%**		
	C	29	76.3	123	83.1	0.3506	0.655 (0.2764 – 1.552)
	T	9	23.7	25	16.9		1.527 (0.6444 – 3.618)
Chinese	**Genotype**	**Count**	**%**	**Count**	**%**		
	CC	17	68.0	67	77.9	0.2798	0.6026 (0.2256 – 1.61)
	CT	8	32.0	16	18.6		2.059 (0.757 – 5.6)
	TT	0	0	3	3.5		Not Calculated
	Allele	**Count**	**%**	**Count**	**%**		
	C	42	84.0	150	87.2	0.6383	0.77 (0.3198 – 1.854)
	T	8	16.0	22	12.8		1.3 (0.5394 to 3.127)
Indian	**Genotype**	**Count**	**%**	**Count**	**%**		
	CC	18	52.9	48	53.3	1.0	0.984 (0.4465 – 2.170)
	CT	13	38.2	35	38.9		0.973 (0.4322 – 2.189)
	TT	3	8.8	7	7.8		1.148 (0.279 – 4.719)
	Allele	**Count**	**%**	**Count**	**%**		
	C	49	72.1	131	72.8	1.0	0.9646 (0.5173 – 1.8)
	T	19	27.9	49	27.2		1.037 (0.5559 – 1.933)
Others	**Genotype**	**Count**	**%**	**Count**	**%**		
	CC	5	35.7	0	0	Not Calculated	Not Calculated
	CT	2	64.3	0	0		
	TT	0	0	0	0		
	Allele	**Count**	**%**	**Count**	**%**		
	C	12	85.7	0	0	Not Calculated	Not Calculated
	T	2	14.3	0	0		

4.3.1.3 Analysis of rs6754677

The Fisher exact test on rs6754677 in Table 4-18 showed no association of this variant with CD in the Malaysian population (*P*=0.11). The genotype count also showed higher frequency of mutant type A allele in the patients (57.6% vs. control 50.4%). Both alleles did not show obvious effect on CD with OR=1 included in both allele 95% CI range. Hardy-Weinberg equilibrium analysis showed that the samples collected were in equilibrium with insignificant *P*-value obtained (*P*=0.5294).

Table 4-18 *P*-value, OR And 95% CI obtained for *ATG16L1* rs6754677

SNP	Genotype Count	Patients (n=85) (%)		Controls (n=250) (%)	
ATG16L1 rs6754677	GG	15 (17.6)		64 (25.6)	
	GA	42 (49.4)		120 (48.0)	
	AA	28 (32.9)		66 (26.4)	
	Allelic Count	**Patients (n=85) (%)**	**Controls (n=250) (%)**	**Odd Ratio (95% Confidence Interval)**	***P*-value**
	Wild type (G)	72 (42.4)	248 (49.6)	0.747 (0.526 - 1.061)	0.110
	Mutant type (A)	98 (57.6)	252 (50.4)	1.340 (0.943 - 1.903)	

Among the three main races in Malaysia, significant association with CD was observed in Indians ($P=6.3\times10^{-3}$) but not in Malays (*P*=0.58) and Chinese (*P*=0.63) (Table 4-19). The Indians also have significant genotypes difference where homozygous wild type G was common in the controls while homozygous mutant A was common in CD individuals (*P*<0.003). However, Malay control cohort displayed disequilibrium from Hardy-Weinberg equation (*P*=0.04933) (Appendix G), this led to anticipation of false negative result for the Malays cohort. Indian and Chinese control cohorts were found to fulfill the Hardy-Weinberg equilibrium with *P*>0.05 (Appendices H & I).

Table 4-19 Data and statistical analysis of *ATG16L1* rs6754677 for different races

Race		Patients		Controls		*P*-value	OR (95% CI)
Malay	**Genotype**	**Count**	**%**	**Count**	**%**		
	GG	4	21.1	6	8.1	0.2206	3.022 (0.7579 – 12.05)
	GA	8	42.1	43	58.1		0.524 (0.1888 – 1.456)
	AA	7	36.8	25	33.8		1.143 (0.4004 – 3.265)
	Allele	**Count**	**%**	**Count**	**%**		
	G	16	42.1	55	37.2	0.7085	1.23 (0.5955 – 2.54)
	A	22	57.9	93	62.8		0.813 (0.3938 – 1.679)
Chinese	**Genotype**	**Count**	**%**	**Count**	**%**		
	GG	7	28.0	25	29.1	0.4927	0.949 (0.3528 – 2.552)
	GA	14	56.0	38	44.2		1.608 (0.6555 – 3.943)
	AA	4	16.0	23	26.7		0.522 (0.1617 – 1.683)
	Allele	**Count**	**%**	**Count**	**%**		
	G	28	56.0	88	51.2	0.63	1.215 (0.6448 – 2.289)
	A	22	44.0	84	48.8		0.823 (0.4369 – 1.551)
Indian	**Genotype**	**Count**	**%**	**Count**	**%**		
	GG	3	8.8	33	36.7	2.61×10^{-3}	0.167 (0.0474 – 0.59)
	GA	17	50.0	39	43.3		1.308 (0.5929 – 2.884)
	AA	14	41.2	18	20.0		2.8 (1.1892 – 6.592)
	Allele	**Count**	**%**	**Count**	**%**		
	G	23	33.8	105	58.3	6.3×10^{-3}	0.191 (0.0944 – 0.385)
	A	45	66.2	75	41.7		2.74 (1.529 – 4.908)
Others	**Genotype**	**Count**	**%**	**Count**	**%**		
	GG	1	14.3	0	0	Not Calculated	Not Calculated
	GA	3	42.9	0	0		
	AA	3	42.9	0	0		
	Allele	**Count**	**%**	**Count**	**%**		
	G	5	35.7	0	0	Not Calculated	Not Calculated
	A	9	64.3	0	0		

The CD in Indian population showed significant association with alleles of rs6754677, thus its genotype association was investigated. Significant associations were obtained for all three models. As demonstrated in Table 4-20, the *P*-values for dominant, recessive and additive models were 0.003371, 0.02177 and 0.001501 respectively. As both dominant and recessive model demonstrated significant allele effect, thus the additive model was proposed to be the most suitable model as single allele could not have both dominant and recessive properties.

Table 4-20 Genotype association analysis (dominant, recessive and additive models) for Indian *ATG16L1* rs6754677

Cohort	**Genotype**		***P*-value**
	GA + AA	**GG**	
CD	31	3	3.371×10^{-3}
Control	57	33	
Cohort	**Genotype**		***P*-value**
	AA	**GG + GA**	
CD	14	20	2.177×10^{-2}
Control	18	72	
Cohort	**Genotype**		***P*-value**
	AA	**GG**	
CD	14	3	1.501×10^{-3}
Control	18	33	

4.3.2 Analysis of *IRGM* Variants

4.3.2.1 Analysis of rs4958847

Table 4-21 demonstrated the result of Fisher exact test for rs4958847. *P*-value of 0.1540 suggested no significant association between this SNP and the disease in the Malaysian population. Both wild type and mutant type alleles did not show significant effect on CD with 95% CI included OR=1. Patients were found to harbor higher frequency of wild type allele (60%) as compared to the controls who carried higher frequency of mutant type allele (46.4%). Genotype analysis showed that patients carried 35.3% of homozygous wild genotype and 15.3% homozygous mutant genotype in relative to 32.4% and 25.2% of the respective genotypes in the controls. The Hardy-Weinberg equilibrium analysis in Table 4-22 showed significant disequilibrium occurred in the samples (*P*=0.02232) and this called for Hardy-Weinberg equilibrium analysis of control cohorts in each race to further investigate the source of disequilibrium.

Table 4-21 *P*-value, OR And 95% CI obtained for *IRGM* rs4958847

SNP	**Genotype Count**	**Patients (n=85) (%)**		**Controls (n=250) (%)**	
IRGM rs4958847	GG	30 (35.3)		81 (32.4)	
	GA	42 (49.4)		106 (42.4)	
	AA	13 (15.3)		63 (25.2)	
	Allelic Count	**Patients (n=85) (%)**	**Controls (n=250) (%)**	**Odd Ratio (95% Confidence Interval)**	***P*-value**
	Wild type (G)	102 (60.0)	268 (53.6)	1.299 (0.912 - 1.849)	0.1540
	Mutant type (A)	68 (40.0)	232 (46.4)	0.770 (0.541 - 1.097)	

Table 4-22 Hardy-Weinberg equilibrium of control samples for *IRGM* rs4958847

SNP	Genotype Count	Observed Frequency	Expected Frequency	*P*-value
IRGM rs4958847	Homozygous Wild	0.324	0.287	0.02232
	Heterozygous	0.424	0.497	
	Homozygous Mutant	0.252	0.215	

Table 4-23 showed Fisher exact analysis done within Malay, Chinese and Indian populations independently. None of these races had significant association with the disease as P=0.3404, P=1.0 and P=0.3847 was found for Malays, Chinese and Indians respectively. Although Indians were found significant (P=0.034) for their genotypic differences, yet it did not affect subsequent alleles frequency analysis.

Table 4-23 Data and statistical analysis of *IRGM* rs4958847 for different races

Race		Patients		Controls		*P*-value	OR (95% CI)
Malay	**Genotype**	**Count**	**%**	**Count**	**%**		
	GG	11	57.9	33	44.6	0.68	1.708 (0.6163 – 4.735)
	GA	6	31.6	30	40.5		0.677 (0.2315 – 1.979)
	AA	2	10.5	11	14.9		0.674 (0.1362 – 3.334)
	Allele	**Count**	**%**	**Count**	**%**		
	G	28	73.7	96	64.9	0.3404	1.571 (0.71 – 3.475)
	A	10	26.3	52	35.1		0.659 (0.2971 – 1.463)
Chinese	**Genotype**	**Count**	**%**	**Count**	**%**		
	GG	3	12.0	13	15.1	0.8199	0.766 (0.1999 – 2.933)
	GA	13	52.0	38	44.2		1.368 (0.5605 – 3.341)
	AA	9	36.0	35	40.7		0.82 (0.3256 – 2.063)
	Allele	**Count**	**%**	**Count**	**%**		
	G	19	38.0	64	37.2	1.0	1.034 (0.5403 – 1.98)
	A	31	62.0	108	62.8		0.967 (0.5051 – 1.851)
Indian	**Genotype**	**Count**	**%**	**Count**	**%**		
	GG	12	35.3	35	38.9	0.03402	0.857 (0.377 – 1.948)
	GA	21	61.8	38	42.2		2.211 (0.985 – 4.961)
	AA	1	2.9	17	18.9		0.130 (0.0166 – 1.019)
	Allele	**Count**	**%**	**Count**	**%**		
	G	45	66.2	108	60.0	0.3847	1.304 (0.7272 – 2.339)
	A	23	33.8	72	40.0		0.767 (0.4275 – 1.375)
Others	**Genotype**	**Count**	**%**	**Count**	**%**		
	GG	4	57.1	0	0	Not Calculated	Not Calculated
	GA	2	28.6	0	0		
	AA	1	14.3	0	0		
	Allele	**Count**	**%**	**Count**	**%**		
	G	10	71.4	0	0	Not Calculated	Not Calculated
	A	4	28.6	0	0		

4.3.2.2 Analysis of rs11747270

The second variant of *IRGM* (rs11747270) showed insignificant result with *P*=0.09359 for Fisher exact test (Table 4-24). This indicated that there was no association between this SNP and CD in the Malaysian population. Patients was found to carry higher frequency of wild type A allele (70.6%) while the controls contained higher frequency of mutant type G allele (36.8%). Both homozygous wild genotype and heterozygous genotype were inherited by 47.1% patients respectively. The controls only had 41.2% and 44% of homozygous wild genotype and heterozygous genotype inherited.

Table 4-24 *P*-value, OR And 95% CI obtained for *IRGM* rs11747270

SNP	**Genotype Count**	**Patients (n=85) (%)**		**Controls (n=250) (%)**	
IRGM rs11747270	AA	40 (47.1)		103 (41.2)	
	AG	40 (47.1)		110 (44.0)	
	GG	5 (5.9)		37 (14.8)	
	Allelic Count	**Patients (n=85) (%)**	**Controls (n=250) (%)**	**Odd Ratio (95% Confidence Interval)**	***P*-value**
	Wild type (A)	120 (70.6)	316 (63.2)	1.398 (0.959 - 2.037)	0.09359
	Mutant type (G)	50 (29.4)	184 (36.8)	0.716 (0.491 - 1.043)	

Data in Table 4-25 showed no significant association found between the three races with CD in Fisher exact test. The *P*-values for the three races were 0.6861, 1.0 and 0.06604 for Malays, Chinese and Indians respectively. Although Indians were associated significantly (*P*=0.04223) for their genotypic differences, yet it did not affect subsequent alleles frequencies analysis. Therefore rs11747270 had no association with the Malaysian CD as well as CD in each race.

Table 4-25 Data and statistical analysis of *IRGM* rs11747270 for different races

Race		Patients		Controls		*P*-value	OR (95% CI)
Malay	**Genotype**	**Count**	**%**	**Count**	**%**		
	AA	11	57.9	38	51.3	0.9161	1.303 (0.4704 – 3.606)
	AG	7	36.8	31	41.9		0.809 (0.2859 – 2.29)
	GG	1	5.3	5	6.8		0.767 (0.0842 – 6.981)
	Allele	**Count**	**%**	**Count**	**%**		
	A	29	76.3	107	72.3	0.6861	1.234 (0.5384 – 2.832)
	G	9	23.7	41	27.7		0.81 (0.3532 – 1.857)
Chinese	**Genotype**	**Count**	**%**	**Count**	**%**		
	AA	5	20.0	24	27.9	0.3264	0.646 (0.2177 – 1.916)
	AG	17	68.0	43	50.0		2.125 (0.8295 – 5.444)
	GG	3	12.0	19	22.1		0.481 (0.1298 – 1.781)
	Allele	**Count**	**%**	**Count**	**%**		
	A	27	54.0	91	52.9	1.0	1.045 (0.5556 – 1.965)
	G	23	46.0	81	47.1		0.957 (0.5089 – 1.8)
Indian	**Genotype**	**Count**	**%**	**Count**	**%**		
	AA	19	55.9	41	45.6	0.04223	1.514 (0.6843 – 3.349)
	AG	15	44.1	36	40.0		1.184 (0.5334 – 2.629)
	GG	0	0	13	14.4		Not Calculated
	Allele	**Count**	**%**	**Count**	**%**		
	A	53	77.9	118	65.6	0.06604	1.857 (0.9687 – 3.558)
	G	15	22.1	62	34.4		0.539 (0.281 – 1.032)
Others	**Genotype**	**Count**	**%**	**Count**	**%**		
	AA	5	71.4	0	0	Not Calculated	Not Calculated
	AG	1	14.3	0	0		
	GG	1	14.3	0	0		
	Allele	**Count**	**%**	**Count**	**%**		
	A	11	78.6	0	0	Not Calculated	Not Calculated
	G	3	21.4	0	0		

4.3.2.3 Analysis of rs72553867

Table 4-26 showed the Fisher exact test analysis for alleles of rs72553867. There was no significant association observed between this SNP and the Malaysian CD with 92.4% of wild type C allele found in patient samples while 20.4% of mutant type A allele found in control samples. Genotype analysis showed that both cohorts had no homozygous mutant genotype. Patients harbored higher percentage of heterozygous genotype (21.2%) while the controls harbored greater number of homozygous wild genotype (79.6%).

Table 4-26 *P*-value, OR And 95% CI obtained for *IRGM* rs72553867

SNP	**Genotype Count**	**Patients (n=85) (%)**		**Controls (n=250) (%)**	
IRGM rs72553867	CC	67 (78.8)		199 (79.6)	
	CA	18 (21.2)		51 (20.4)	
	AA	0 (0)		0 (0)	
	Allelic Count	**Patients (n=85) (%)**	**Controls (n=250) (%)**	**Odd Ratio (95% Confidence Interval)**	***P*-value**
	Wild type (C)	152 (92.4)	449 (79.6)	0.959 (0.544 to 1.693)	0.8844
	Mutant type (A)	18 (10.6)	51 (20.4)	1.043 (0.591 to 1.840)	

Table 4-27 showed the analysis of race stratified data of rs72553867 for Fisher exact test and Hardy-Weinberg equilibrium. In Fisher exact test, no race was found to be significantly related with CD as they all had *P*-values above 0.05 (Malay=0.4730, Chinese=1.0, Indian=0.1042). This verified that no association could be established between each race with CD.

Table 4-27 Data and statistical analysis of *IRGM* rs72553867 for different races

Race		Patients		Controls		*P*-value	OR (95% CI)
Malay	**Genotype**	**Count**	**%**	**Count**	**%**		
	CC	18	94.7	62	83.8	0.2914	3.484 (0.4239 – 28.63)
	CA	1	5.3	12	16.2		0.287 (0.0349 – 2.359)
	AA	0	0	0	0		Not Calculated
	Allele	**Count**	**%**	**Count**	**%**		
	C	37	97.4	136	91.9	0.4730	3.265 (0.4111 – 25.93)
	A	1	2.6	12	8.1		0.306 (0.0386 – 2.433)
Chinese	**Genotype**	**Count**	**%**	**Count**	**%**		
	CC	17	68.0	57	66.3	1.0	1.081 (0.4173 – 2.8)
	CA	8	32.0	29	33.7		0.925 (0.357 – 2.396)
	AA	0	0	0	0		Not Calculated
	Allele	**Count**	**%**	**Count**	**%**		
	C	42	84.0	143	83.1	1.0	1.065 (0.4528 – 2.504)
	A	8	16.0	29	16.9		0.939 (0.3994 – 2.208)
Indian	**Genotype**	**Count**	**%**	**Count**	**%**		
	CC	26	76.5	80	88.9	0.09217	0.406 (0.1451 – 1.138)
	CA	8	23.5	10	11.1		2.462 (0.8791 – 6.892)
	AA	0	0	0	0		Not Calculated
	Allele	**Count**	**%**	**Count**	**%**		
	C	60	88.2	170	94.4	0.1042	0.441 (0.1664 – 1.17)
	A	8	11.8	10	5.6		2.267 (0.8548 – 6.011)
Others	**Genotype**	**Count**	**%**	**Count**	**%**		
	CC	6	85.7	0	0	Not Calculated	Not Calculated
	CA	1	14.3	0	0		
	AA	0	0	0	0		
	Allele	**Count**	**%**	**Count**	**%**		
	C	13	92.9	0	0	Not Calculated	Not Calculated
	A	1	7.1	0	0		

CHAPTER 5

DISCUSSION

5.1 Analysis of *NOD1* rs2075820

As *NOD1* was found to have similar structure with *NOD2* (Carneiro et al., 2004) and it is able to trigger innate immune system by detecting iE-DAP (Chamaillard et al., 2003), it is therefore associated with CD development. However, this SNP showed no significant association with Malaysian CD cohort. Analysis by stratifying samples according to individual race carried out later also did not produce significant result.

While comparing Malaysia results with other countries, it is tallied with some studies conducted previously in Western countries. Zouali et al. (2003) showed that this SNP did not related to 381 IBD families which consisted of France, Sweden, Belgium, Spain, Denmark, Italy and Ireland with CD. Published data for Turkish demonstrated that the allele frequency of the mutant type A allele was 0.486 and 0.552 for patients and controls respectively which possessed no significant result (Özen et al., 2006). However, Tanabe et al. (2011) showed that although higher mutant type allele frequency was found in Japanese, they had no significant relationship with CD.

Data from National Center for Biotechnology Information (NCBI) showed the chosen SNP is located in NACHT domain, which consists of 171 amino acids in *NOD1*. NACHT is a conserved domain that exists in apoptosis proteins and found to interact with a number of other domains such as the CARD, WD40 repeat and LRR (Koonin & Aravind, 2000) and acts to hydrolyze ATP or GTP. Proell et al. (2008) has shown that the secondary structure of NACHT is conserved and could be connected to the LRR with a winged helix domain and superhelical domain. However, study done by Molnar et al. (2007) claimed that the mutation can cause the decreasing of helix-formatting potential. Interestingly, none of the three *NOD2* significant mutations were located within NACHT domain.

According to some other studies, Kara et al. (2010) reported that the carrier of E266K had a greater risk towards diseases induced by CagA-positive *Helicobacter pylori* such as antral atrophy and antral intestinal metaplasia. The E266K has also been related to interleukin-8 (IL-8) and *H. pylori* in a study by Hofner et al. (2007) on examination of *H. pylori*-induced duodenal ulcer and gastritis. Although there was no strong evidence showing the effect of *H. pylori* to CD onset but IL-8 has been proven to increase upon the presence of *H. pylori* (Fischer et al., 2001). Experiment on IL-8 and E266K focusing their polymorphisms had come out that both would increase theirs expression in inflamed tissue of CD patients.

As summary, it could be postulated that the occurrence of mutant type A allele is common among Malaysian with no increased risk or susceptibility towards CD and the mechanism is yet to be known. Current study shows that E266K does not affect Malaysian and normal individual that found with E266K might have the normal protein function or protein structure retained.

5.2 Analysis of *CXCL16* rs2277680

Current study disclosed the significance of rs2277680 of *CXCL16* in Malaysian with P=0.04819 and OR>1 where greater risk was possessed by the wild type C allele carrier. With the support of Hardy-Weinberg equilibrium analysis, this result was believed as true positive. Our results also showed that the gene pool in Malaysia which consisted of three main races would have increased risk compared to stratified race data which has no significant result. However, Seiderer et al. (2008) has reported a contrary result where the mutant type allele was determined as the risk allele in Caucasians instead of the wild type allele as for the Malaysian cohort. As validated by the Fisher *P*-value, this SNP was proposed with additive effect. It is again contrary with data from

Seiderer et al., their result did not show any significant association after being analyzed. These contrary results are believed to be due to population differences. With additive model, the absence of risk allele indicated lesser risk while the presence of single or two risk alleles possess single or double risk.

The SNP rs2277680 is located within exon 4 of *CXCL16* and involved an amino acid substitution from alanine to valine (Lundberg et al., 2005). It is believed that protein with wild type allele would have altered properties and lead to a risk in CD development. The effect might also lead to impairment in interaction with CXCR6 during the inflammatory process or affecting the cleavage to form soluble-CXCL16 which responds to innate immune system and inflammatory response (Chandrasekar, 2004).

Among the CXC ligand chemokines, only a few were being investigated on their association with CD. For example, *CXCL9* among Caucasian IBD patients was studied and significant association was found with rs2276886 polymorphisms (77147452-G/A) (Lacher et al., 2007) with risk effect of the wild type allele where patients with wild type allele were found to have an early disease onset. In the study conducted by Mrowicki et al. (2011), they showed that the SNP rs1801157 of *CXCL12* with G/A polymorphism had no significant association with Polish IBD.

Meanwhile, most CXC ligand family member has been proven to rise in expression level when being examined using CD samples. The chosen gene, *CXCL16*, was reported by Diegelmann et al. (2010) to have the highest expression level in intestines, spleen, liver and kidney of mice model, and also human colorectal cancer–derived intestinal epithelial cell lines. *CXCL16* was also found to be expressed in

inflamed biopsy samples and CD active patients. This confirmed that *CXCL16* is responsible for the inflammatory response in the innate immune system. By focusing to this respond, it becomes a potential treatment for IBD. Uza et al. (2011) demonstrated testing of anti-mouse *CXCL16* monoclonal antibody (mAb) in mice model with DSS-induced colitis while goat immunoglobulin G (IgG) as the control. The result showed that mAb model yielded better results and the mice body weight was retained much more with mAb treatment. This provided patients with another choice of treatment.

As our study showed significant result, similar to the results obtained from Caucasians as reported earlier, further functional study of the SNP should be carried out in order to fully understand its relationship with CD.

5.3 Analysis of *STAT6* rs324015

The SNP, rs324015 in *STAT6* was found not contributing significant relationship to the CD of Malaysian as well as individual race. In another study by Bhanoori et al. (2007) on South Indian woman, the same SNP was found to be related with endometriosis. Gao et al. (2000) has also shown that the *STAT6* was the susceptibility gene for Japanese asthma patients instead of CD as wild type allele had higher frequency in Japanese population. His study demonstrated a higher frequency of wild type A allele compared to the mutant type G allele (67% and 33% respectively) in this population. This suggests that the SNP may be related to diseases other than CD in Asians.

Caucasian was found susceptible to high risk of CD with the polymorphism of *STAT6*. However, in 2003, Xia et al. conducted a SNP investigation in 183 Dutch Caucasians CD patients and 173 controls, their result showed 80% and 75% of controls and patients carried the mutant type G allele respectively and no significant association could be established. The only significant result found was with female in the same study, who carried higher frequency of mutant type G allele compared to the controls (P=0.037). In the same year, Jong et al. conducted the same gene polymorphism study with 110 Netherlands Caucasian CD and 106 controls, they also could not establish the association of this SNP with CD in the population (P=0.19). These diverged results showed that geological factor might need to be taken into account for CD investigation.

The 3'-untranslated region (3'UTR) is important for translation purposes as well as localization and stability of mRNA. Mazumder et al. (2003) suggested that 3'UTR may play a part in regulatory mechanism. Chatterjee and Pal (2009) also reviewed that mutation in 3'UTR could affect stop codon, polyA tail and 3'UTR secondary structure and any alteration could contribute to diseases, for example, aniridia that affects the iris. Furthermore, *STAT6* is found to relate with Th2-initiation response to inhibit Th1 response (Elo et al., 2010), hence mutation in the gene might attenuate the dimerization of *STAT6* to further transmit the signal for Th2 response.

Although rs324015 is located within the 3'UTR, its mutation did not contribute much effect to Malaysian population. At the same time, this SNP also seems to have a mysterious effect that leads to poor immune response and trigger onset of CD as shown by Klein et al. (2005) on German Caucasians. Study by Dimon et al. (2001) on the

entire *STAT6* gene including 1.5 kb of 5' promoter region, all 23 exons and 3'UTR, they found no significant SNP to be related to IBD. This result is tally with Malaysian result.

5.4 Analysis of *TLR4* rs4986791

A significant result (*P*=0.006886) was obtained when comparing *TLR4* rs4986791 allelic difference between control and patient group. At the same time mutant allele carrier displayed 2.3 times of greater risk on CD (OR=2.289) compared to wild type allele carrier (OR=0.437). However, disequilibrium from Hardy-Weinberg analysis suggests a conflict result that cannot be established and therefore a large sample size is required for further confirmation. Genotype association analysis showed that this SNP was related only with Malay population with dominant effect. Significant value found proposed the presence of risk allele (mutant type T allele) would induce CD onset. As shown by recruited data, no homozygous mutant genotype was found within Malay samples, all nine samples with mutant type T allele presented were heterozygotes.

There are two tallied reports suggested that *TLR4* is important for the defense system and is related to the inflammatory response. In year 2000, Cario & Daniel experimented the expression of *TLR4* in CD patients by using TLR4 antisera and high expression of *TLR4* was found in epithelial cells of CD patients, while in inactive CD patients, only a small amount of *TLR4* expression was detected in the small intestinal and colonic epithelial cells. Before them, Frolova et al. (2008) reported the *TLR4* expression had increased in 13 CD active patients and 6 inactive CD samples compared to 24 healthy controls.

According to a few studies that conducted to investigate the protection effect of *TLR4* on *Helicobacter pylori* and IBD, the chosen SNP rs4986791 was existed frequently in duodenal ulcer and gastric cancer patients with *H. pylori* infection (Trejo-de la O et al., 2008). This variant also escalated the expression of TNF-α and *IL-10* that are associated with inflammation while weakening the expression of *IL-8*. The expression of *TLR4* also showed to be decreased in 10-fold when the chosen polymorphism existed (Prohinar et al., 2010).

All these findings and results concluded that *TLR4* polymorphism disrupts the normal function of *TLR4* protein as its variants are located on the LRR domain of the protein, which plays an important role as substrate binding site and protein-protein interactions. The LRR region is a common structure among TLRs member that has different functions (Fukata et al., 2009; Lavelle et al., 2010). Interestingly, three reported significant variant sites of *NOD2* are located on the LRR region. This might indicate that the effect of variants to the protein structure for both *NOD2* and *TLR4*. Therefore it could be postulated that rs4986791 affects the confirmation of mature protein which leads to malfunction in protein interaction.

However, Taiwanese researchers Hsiao et al. (2007) had reported that Taiwanese CD is not related with this SNP as well as the Koreans (Kim et al., 2012; Ye et al., 2009). Therefore Malay population in Malaysia is the only proposed with significant mutation lead by this SNP. The CD onset could be contributed by this population since it is the only result that found significant in this current study.

5.5 Analysis of *ATG16L1* Gene Variants

The first *ATG16L1* SNP, rs2241880 was discovered to be significant for Malaysian CD patients who harbored higher mutant type G allele. The significant association result has led to subsequent study on individual race, significant result was then found in the Malay population. Both sample cohorts of Malaysian and Malay were in Hardy-Weinberg equilibrium.

The results obtained from the second SNP of *ATG16L1*, rs6758317 were not significant regardless of Malaysian or individual race. The third SNP chosen, rs6754677, has association with Indian CD. These two SNPs have little information as they were only discovered recently by Van Limbergen et al. (2012). Malay population was shown to have Hardy-Weinberg disequilibrium for rs6754677. The disequilibrium might be due to non-random sample selection or the conditions of population equilibrium (non-random mating, genetic shifting, gene mutation and population migration) were disobeyed.

In genotype association analysis, rs2241880 and rs6754677 were found with additive effect. In order to validate the association model, data from Caucasian population (Csöngei, 2010) and genome wide study from Hampe et al. (2007) which showed significant association results was being analyzed, recessive and additive model were discovered for former Caucasians study while all three models were found for Hampe et al. study (Appendix J). Combination of these data along with Malaysian data showed that additive effect appeared as the most common model. Meanwhile for rs6754677, no model has been proposed so far, thus the additive model proposed in this study needs to be further investigated and justified.

Based on published data from 2007 to present, rs2241880 has been known to be associated with CD in many populations such as Germany, Netherlands, Caucasians, Hungarian, Scottish, Spanish, UK, New Zealand and Dutch. The details were listed in Table 5-1. As shown in the table, all populations have mutant type G allele as the risk allele. Most western populations had high frequency of mutant type G allele (>50%) in CD cohort which contributed to significant risk compared to normal cohort. Asians such as Japanese and Korean showed no relation with this SNP as reported by Yamazaki et al. (2007) and Yang et al. (2009).

Table 5-1 Risk allele frequency, *P*-value and OR value for various populations of *ATG16L1* (rs2241880)

Population (Reference)	**Risk Allele**	**Risk Allele Frequency in CD (%)**	***P*-value**	**OR Value**
Caucasians (Csöngei, 2010)	G	58.3	4.0×10^{-3}	1.69
Dutch (Weersma et al., 2008)	G	64.2	1.7×10^{-3}	1.36
Germany (Glas et al., 2008)	G	59.0	3.7×10^{-6}	-
Hungarian (Lakatos et al., 2008)	G	58.1	3.7×10^{-2}	1.39
Netherlands (Büning et al., 2007)	G	63.1	2.0×10^{-2}	1.43
New Zealand (Roberts et al., 2007)	G	58.0	1.0×10^{-4}	1.41
Scottish (Van Limbergen et al., 2008)	G	60.7	1.0×10^{-2}	1.32
Spanish (Márquez et al., 2009)	G	59.0	8.0×10^{-3}	1.28
Spanish (Palomino-Morales et al., 2009)	G	63.0	6.5×10^{-9}	1.62
UK (Prescott et al., 2007)	G	58.1	2.4×10^{-6}	-
Japanese (Yamazaki et al., 2007)	G	23.9	0.21	-
Korean (Yang et al., 2009)	G	32.1	0.501	1.124
Current study (Malaysian, 2013)	G	47.6	1.43×10^{-2}	1.55

The SNP rs2241880 is located at the 300th amino acid position of N-terminus of WD-repeat domain of exon 9, which is evolutionarily conserved (Figure 5-1). There are 32 β-strands in WD repeat of human *ATG16L1* and eight bladed β-propellers are formed from these strands. A change from polar threonine to non-polar alanine in this SNP leads to an unknown interaction. Although there is no experiment support, the mutation is believed to lead to a failure in the autophagy process. This SNP is also reported to be significant with CD association in our Malay population (*P*=0.03619).

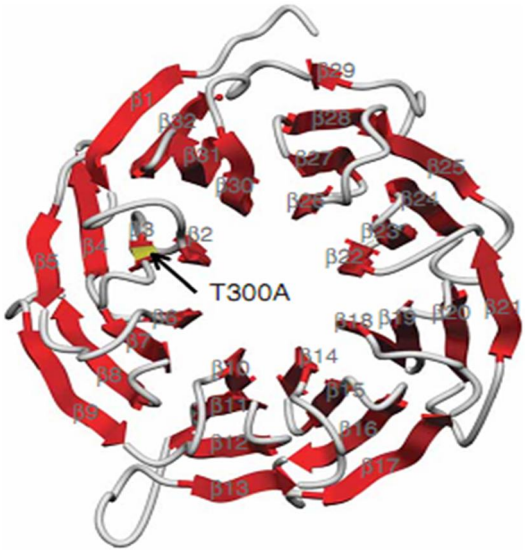


Figure 5-1 Location of rs2241880 (T300A) within WD repeat of *ATG16L1*

Adopted from Hampe et al. (2007)

A general WD repeat is known to consist of 44-60 residues sequence which contains GH dipeptide at N-terminus and WD (tryptophan-aspartic acid) dipeptide at C-terminus. The conserved core sequence between GH-WD is accord to the regular expression (Smith et al., 1999). WD-repeat is known to be crucial in some genes to assist in signal transduction, cytoskeletal assembly, cell cycle regulation and programmed cell apoptosis (Li & Roberts, 2001). Therefore other than CD, this region is also related with other diseases. Cockayne syndrome (CS) that involves slow development in new-birth was due to five WD-repeat deletion of CSA gene which leads to disruption in transcription interaction.

Since rs2241880 was widely investigated, Van Limbergen et al. (2012) initiated a study on the gene in order to disclose more rare non-synonymous SNP without considering the functional region of the gene. As a result, a number of SNPs was discovered which is located from 5' region to 3'UTR. Among them, rs6758317 (intron 2) and rs6754677 (intron 14) had a $P<10^{-4}$ and $P=0.03$ respectively in the study,

compared to our current study that has *P*=0.375 for the former and *P*=0.102 for the latter. Although intron is to be sliced away after transcription, its role in *ATG16L1* need to be further studied to discover its significance contribution as rs6754677 was found to be associated with unknown disruption mechanism.

Some novel characteristics of *ATG16L1* were observed in three published reports in year 2010. First, Cooney et al. reported the susceptibility variants in *NOD2* and *ATG16L1*, showed disability in autophagy reduction in dendritic cells from CD patients and proposed both genes were linked in the same pathway. Second, Homer et al. reported human macrophage, dendritic cell and epithelial *NOD2* cell were dependent on *ATG16L1* and *NOD2* variation to show different responses to muramyl dipeptide (MDP). Third, Travassos et al. revealed the autophagy was directed by *NOD1* and *NOD2* by recruiting *ATG16L1*. Leu1007insC (NOD2fs) was shown to generate truncated protein and paralyzed *ATG16L1* recruitment. All these findings show that the *NOD2-ATG16L1* axis were probably the critical pathway which is related to the onset of CD by dysregulate the autophagy process.

In summary, this gene requires more studies to be conducted to disclose the gene's role in the autophagy process, as it has been widely recognized to be associated with CD. Overall, significant association between *ATG16L1* and CD in Malaysia has been established.

5.6 Analysis of *IRGM* Gene Variants

Three SNPs in *IRGM* were analyzed and none of them gave significant result. The first SNP, rs4958847 has no significant association with CD for Malaysian as well as individual race cohorts. However, disequilibrium was found for Hardy-Weinberg analysis of this SNP. As subsequent investigation, Hardy-Weinberg equilibrium analysis of control samples of three races was done. Surprisingly no disequilibrium was found from each individual race. This shows the mixing of all genomes from the three populations is the main factor of sample disequilibrium that leads to deviation in corresponding analysis (Lewis, 2002).

Second and third SNPs of *IRGM*, rs11747270 and rs72663867 were having insignificant results on Malaysian and race cohort as well. Compared to the former SNPs, the rs72663867 was only being investigated on Korean population, which proved to be significant. As reported by Moon et al. (2012), this SNP is found to be associated with the Korean CD cohort (P=0.012). However, no significant association result was discovered in our current study.

Although *IRGM* was found to have no association with Malaysian cohort, it has been associated with a few populations such as Canadian, Italian, Spanish and Korean on various SNPs. The details are listed in Table 5-2. Both Korean and Malaysian did not pose significant result for rs4958847 while P<0.05 was found for the Westerner. Most populations for example Dutch, Belgium and New Zealand did not harbor more than 50% of the risk allele for rs4958847 in their CD cohort.

Table 5-2 Risk allele frequency, *P*-value and OR value for various populations in *IRGM* variants

Population (Reference)	***IRGM* Variants (rs#)**	**Risk Allele Frequency in CD (%)**	***P*-value**	**OR Value**
Ashkenazi Jewish (Peter et al., 2011)	rs11747270	21.0	1.2×10^{-3}	1.48
Canadian (Murdoch et al., 2012)	rs11747270	16.5	1.1×10^{-2}	2.01
Dutch-Belgian (Weersma et al., 2009)	rs4958847	16.0	2.26×10^{-5}	1.44
European (Parkes et al., 2007)	rs4958847	12.5	3.77×10^{-9}	1.36
Italian and Caucasians (Latiano et al., 2009)	rs4958847	78.0	2.5×10^{-3}	1.39
Spanish (Palomino-Morales et al., 2009)	rs4958847	77.8	0.02	0.79
New Zealand Caucasians (Roberts et al., 2008)	rs4958847	16.0	2.2×10^{-3}	1.767
Korean (Moon et al., 2012)	rs4958847	38.4	0.37	0.90
	rs72553867	15.1	0.012	0.67
Current study (Malaysia, 2013)	rs4958847	40.0	0.147	0.7701
	rs11747270	29.4	0.081	0.7156
	rs72553867	10.6	0.885	1.0426

In Asia, Japanese was the first to report no significant association between their populations and rs4958847 (Aizawa et al., 2008) in terms of CD susceptibility, followed by Korean in 2012 with insignificant result as well. Similarly, the African American was found with no association as well (Wang et al., 2009), which is contradictory with Western populations such as those listed in the Table 5-2.

Meanwhile for second SNP, the rs11747270, recent studies revealed its association with Ashkenazi Jewish and Canadian CD patients (Murdoch et al., 2012; Peter et al., 2011). Both populations have been identified to have high susceptibility previously and they demonstrated similar outcomes for this SNP ($P_{\text{Ashkenazi Jewish}} = 1.2 \times 10^{-3}$, $P_{\text{Canadian}} = 1.1 \times 10^{-2}$). However, local data with P=0.08089 declared no association

(Table 5-2). Instead of having risk effect, mutant type G allele in Malaysian possesses less risk (OR=0.7156) compared to previous Western populations that possess risk effect ($OR_{Ashkenazi\ Jewish}$ =1.48, $OR_{Canadian}$=2.01).

The SNP rs11747270 is located at 280bp upstream of 4th exon of *IRGM* and close to the splice-acceptor region, which is an intron region. It is critical in assembly of spliceosome as this would determine the spliced isoform of transcripted gene (Bekpen et al., 2010). The previous SNP rs4958847 is also found to locate in the intron region of IRGM *b* (Barrett et al., 2008). Generally, intron has effect on expression rate instead of transcipted protein structure. However, Lu et al. (2012) demonstrated that SNP located in intron region could be associated with disease onset. Therefore the intron region of *IRGM* needs to be investigated to discover its function.

Moon et al. (2012) reported that rs72553867 is associated with Korean CD patients (*P*=0.012; OR=0.67), but no significant result was shown in our study (*P*=0.885; OR=1.043). The SNP function is also different between the two populations where it has protective effect for the Koreans while as a risk factor for Malaysians (Table 5-2). However, more investigation is required for CD pathogenesis of this SNP in Malaysians using a larger sample size.

5.7 Sample Quality

In current study, the call rate for each sample is 100% for all investigated SNPs. Every single sample resulted in clear respond for all SNP screenings. On the other hand the disequilibrium that has occurred in Hardy-Weinberg equilibrium analysis could be due to the heterogeneous population that formed the pooled Malaysian cohort or particular reason such as having SNP within conservative region.

However, according to statistical power analysis done by using Quanto 1.2.4 on sample size, although most significant SNPs were having equilibrium samples, but they were showed to have lower power as with smaller sample size that would increase risk for type II error.

According to Quanto 1.2.4, for each CD and control group, different sample size was required for each chosen SNP. This is hardly being solved since CD patient sample is less available in Malaysia. This could be addressed by participation of every medical institution in Malaysia in order to obtain more samples for investigation.

5.8 Limitations

There were a few limitations in this research. The first was the sample size of the participants. There were only 85 CD and 250 healthy samples available. Ignoring control samples that can simply be obtained from any healthy individual, the patient samples were too little compared to other research conducted previously. The number of samples for each race should be increased as there were less than 50. This would lead to less reliable result while analysis within race was performed. Although this issue could be satisfied by the computation of data by more stringent statistical software, larger samples size should be acquired to be representative of the actual Malaysian population.

The second limitation was the source of patient samples and their representativeness. As mentioned in Chapter 3, all samples in this research were collected in UMMC, which is less representative since this disease was widespread throughout this country. In order to solve this matter, every medical center available in Malaysia could participate in similar research by providing patients' sample under proper legal process. By having representative samples from each state of Malaysia, the data obtained could be represented as all Malaysian.

The third limitation was related to the low genotype frequency found. Both *TLR4* SNP (rs4986791) and one of *IRGM*'s SNP (rs72553867) did not yield sufficient homozygous mutant genotype that could undergo Fisher exact analysis. This abused the power of statistic and less reliable data would be obtained. This is because an event with low occurrences could not be compared in statistic hence SNP with low genotype frequency raised the interest of studying their genetic profile.

5.9 Recommendations

Since there were SNPs found to be associated with CD, the progress of pathogenesis study can be initiated. The study can be performed from various prospects in order to understand more about the possible effects and outcomes it introduces. In addition, the genetic data generated from the study can also serve as a reference for future investigation of the same SNP or gene in various populations or possible treatment for CD patients.

Genetic profile of each race could be studied as well in order to understand more on their genes polymorphisms. This is because every population may inherit different gene profile from their ancestors and pass on to their offspring hence causing various results. Moreover, Malaysia consists of three main populations with some other native residents. Since Hardy-Weinberg equilibrium is hard to imply in reality due to man-made factors, investigation of original gene pool would become worthy as it could be conducted solely with better data interpretation.

For SNPs that are screened using PCR-RFLP method, other screening methods such as sequencing and microarray are better options as they produce more reliable outcome with less error.

CHAPTER 6

CONCLUSION

6.1 Conclusions

This was the first fruitful report of an investigation of the six genes along with their susceptible SNPs in Malaysia and is a good reference for future research. A complete data set was generated for each gene on Malaysians. As result, there are in total three genes that are found to be significant to Malaysian: *CXCL16*, *TLR4* and *ATG16L1*. No association could be established for *NOD1*, *STAT6* and *IRGM*.

Alongside this research, the Malay population is found to relate with *TLR4* and *ATG16L1* variants that disobey the conclusion made previously stating that the Indians may have the highest susceptibility. Meanwhile Indians were susceptible only to *ATG16L1*.

.

APPENDICES

Appendix A: Calculation in Fisher Exact Test

2×2 Contingency Table

Variable (Rows)	Variable (Columns)		Row Totals
	C1	C2	
R1	*a*	*b*	*a* + *b*
R2	*c*	*d*	*c* + *d*
Column Totals	*a* + *c*	*b* + *d*	*n* (Grand Totals)

$$p = \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{a!\,b!\,c!\,d!\,n!}$$

Appendix B: Calculation for Odd Ratio (OR) and 95% Confidence Interval (CI)

2×2 Contingency Table

Variable (Rows)	Variable (Columns)		Row Totals
	C1	**C2**	
R1	*a*	*b*	*a* + *b*
R2	*c*	*d*	*c* + *d*
Column Totals	*a* + *c*	*b* + *d*	*N* (Grand Totals)

$$\text{OR} = \frac{(a/b)}{(c/d)}$$

$$= \frac{ad}{bc}$$

$$95\%\ \text{CI} = \exp\left[\ln(\text{OR}) \pm 1.96\sqrt{\left(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}\right)}\right]$$

Appendix C: Calculation for Hardy-Weinberg Equilibrium

In order to apply Hardy-Weinberg equation, value of p and q must first be calculated.

Let observed value of:

p^2 = AA genotype

$2pq$ = AB genotype

q^2 = BB genotype

By applying equation ($p^2+2pq+q^2 = 1$), we could obtain the following data:

Frequency of A = $p^2 + 0.5\ (2pq)$

Frequency of B = 1- p

Therefore, expected value of:

p^2 = (Frequency of A)^2

$2pq$ = 2 (Frequency of A) (Frequency of B)

q^2 = (Frequency of B)^2

Appendix D: Data Distribution for Genotype Association Model Analysis

2×3 Contingency Table

Cohort	Genotype		
	wt/wt*	***wt/mt****	***mt/mt***
Cases	*a*	*b*	*c*
Controls	*d*	*e*	*f*

**wt*: wild type allele **mt*: mutant type allele

1. Dominant model

Cohort	Genotype	
	(wt/mt)+(mt/mt)	***wt/wt***
Cases	*b*+ *c*	*a*
Controls	*e* + *f*	*d*

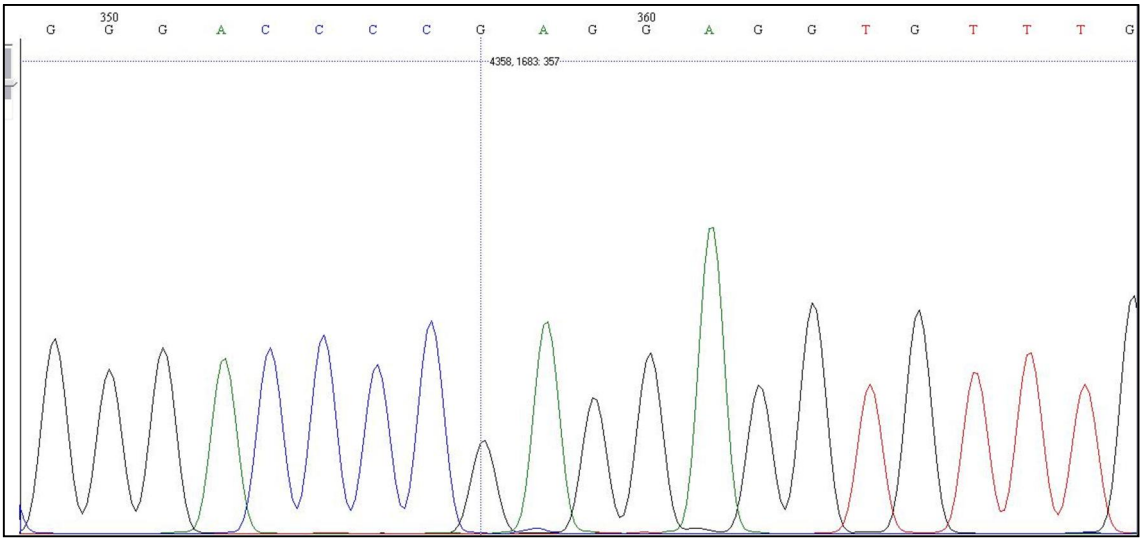
2. Recessive model

Cohort	Genotype	
	mt/mt	***(wt/wt)+ (wt/mt)***
Cases	*c*	*a* + *b*
Controls	*f*	*d* + *e*

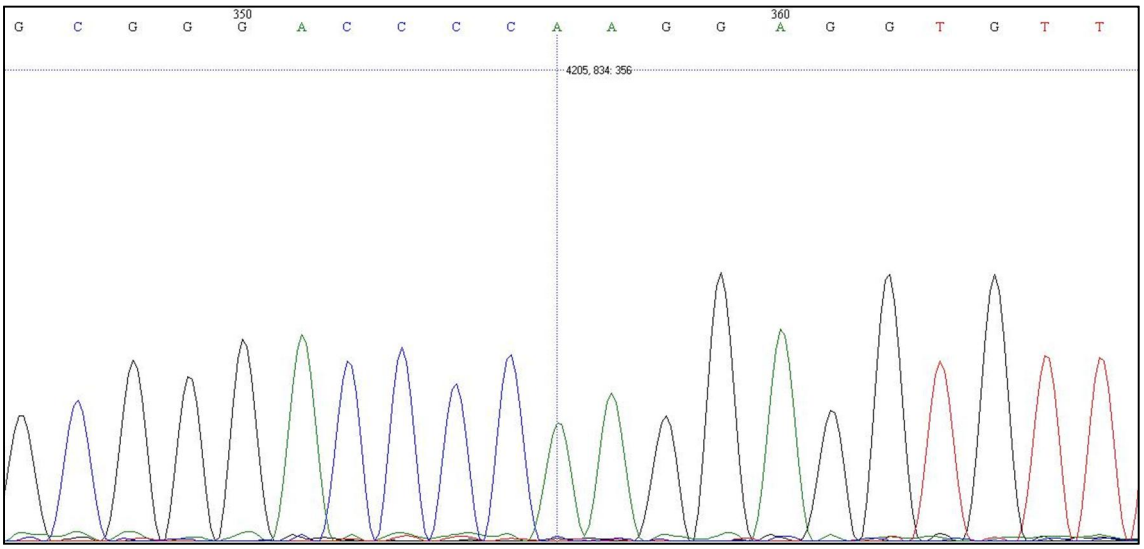
3. Additive model

Cohort	Genotype	
	mt/mt	***wt/wt***
Cases	*c*	*a*
Controls	*f*	*d*

Appendix E: Electropherogram of Sequenced Homozygous Genotypes for *NOD1* (rs2075820)



Homozygous wild genotype (GG) of *NOD1* (rs2075820)



Homozygous mutant genotype (AA) of *NOD1* (rs2075820)

Appendix F: Hardy-Weinberg Equilibrium Analysis for Malaysian Data by Arlequin 3.0

```
////////////////////////////////////////////////////////////////////
RUN NUMBER 1 (10/01/13 at 12:07:16)
////////////////////////////////////////////////////////////////////

 #[WARNING # 1] : bad value assigned to keyword in project file
 #[WARNING # 2] : invalid recessive allele, switching off recessive data


Project information:
--------------------
       NbSamples      =  2
       DataType       =  STANDARD
       GenotypicData  =  1
       GameticPhase   =  0
       RecessiveData  =  0

==============================
Settings used for Calculations
==============================

 General settings:
 -----------------
        Deletion Weight               = 1
        Transition Weight Weight      = 1
        Tranversion Weight Weight     = 1
        Epsilon Value                 = 1e-07
        Significant digits for output  = 5
        Use original haplotype definition
        Alllowed level of missing data = 0.05
```

```
 Active Tasks:
 -------------

    Hardy-Weinberg equilibrium test:
    --------------------------------
        No. of steps in Markov chain      = 1000000
        No. of Dememorisation Steps       = 100000
        Required precision on Probability = 0
        Test association at the Locus level


The computation of AMOVA with the option "CONVENTIONAL FST" is
not possible with unknown gametic phase data.
SKIPPING COMPUTATION OF GENETIC STRUCTURE

 #[WARNING # 3] : unable to compute Genetic Structure with current settings

================================================================================
== ANALYSES AT THE INTRA-POPULATION LEVEL
================================================================================


================================================================================
== Sample :    CD
================================================================================


================================
== Hardy-Weinberg equilibrium : (CD)
================================

Reference: Guo, S. and Thompson, E. 1992.
           Levene H. (1949).
```

Exact test using a Markov chain (for all Loci):
Forecasted chain length   :1000000
Dememorization steps   :100000

Locus	#Genot	Obs.Het.	Exp.Het.	P-value	s.d.	Steps done
1	85	0.55294	0.44838	0.04900	0.00023	1001000
2	85	0.48235	0.50268	0.82838	0.00037	1001000
3	85	0.49412	0.48730	1.00000	0.00000	1001000
4	85	0.22353	0.21782	1.00000	0.00000	1001000
5	85	0.55294	0.50184	0.38760	0.00050	1001000
6	85	0.35294	0.34918	1.00000	0.00000	1001000
7	85	0.49412	0.49119	1.00000	0.00000	1001000
8	85	0.49412	0.48284	1.00000	0.00000	1001000
9	85	0.47059	0.41768	0.29972	0.00048	1001000
10	85	0.21176	0.19046	0.58880	0.00051	1001000

==========
== Sample :   Control
==========

==========
== Hardy-Weinberg equilibrium : (Control)
==========

Reference: *Guo, S. and Thompson, E. 1992.*
Levene H. (1949).

Exact test using a Markov chain (for all Loci):
Forecasted chain length   :1000000
Dememorization steps   :100000

Locus	#Genot	Obs.Het.	Exp.Het.	P-value	s.d.	Steps done
1	250	0.48800	0.46394	0.49615	0.00045	1001000
2	250	0.48400	0.48015	1.00000	0.00000	1001000
3	250	0.46000	0.49426	0.30497	0.00048	1001000
4	250	0.08400	0.10949	0.00514	0.00007	1001000
5	250	0.47600	0.46713	0.78934	0.00041	1001000
6	250	0.29600	0.31089	0.42073	0.00048	1001000
7	250	0.48000	0.50097	0.52848	0.00049	1001000
8	250	0.42400	0.49840	0.02232	0.00013	1001000
9	250	0.44000	0.46608	0.41777	0.00049	1001000
10	250	0.20400	0.18356	0.08780	0.00027	1001000

```
================================================================================
== Summary of computations done within populations
================================================================================


////////////////////////////////////////////////////////////////
END OF RUN NUMBER 1 (10/01/13 at 12:07:22))
Total computing time for this run : 0h 0m 5s 709 ms
////////////////////////////////////////////////////////////////
```

Legend

Locus 1: *NOD1* (rs2075820); Locus 2: *CXCL16* (rs2277680); Locus 3: *STAT6* (rs324015); Locus 4: *TLR4* (rs4986791);

Locus 5: *ATG16L1* (rs2241880); Locus 6: *ATG16L1* (rs6758317); Locus 7: *ATG16L1* (rs6754677); Locus 8: *IRGM* (rs4958847);

Locus 9: *IRGM* (11747270); Locus 10: *IRGM* (rs72553867)

Appendix G: Hardy-Weinberg Equilibrium Analysis for Malay Control Cohort Data by Arlequin 3.0

```
//////////////////////////////////////////////////////////////////
RUN NUMBER 1 (20/06/13 at 14:01:35)
//////////////////////////////////////////////////////////////////

 #[WARNING # 7] : bad value assigned to keyword in project file
 #[WARNING # 8] : invalid recessive allele, switching off recessive data


Project information:
--------------------
       NbSamples      =  1
       DataType       =  STANDARD
       GenotypicData  =  1
       GameticPhase   =  0
       RecessiveData  =  0

==============================
Settings used for Calculations
==============================

 General settings:
 -----------------
       Deletion Weight               = 1
       Transition Weight Weight      = 1
       Tranversion Weight Weight     = 1
       Epsilon Value                 = 1e-07
       Significant digits for output  = 5
       Use original haplotype definition
       Alllowed level of missing data = 0.05
```

```
Active Tasks:
 -------------

    Hardy-Weinberg equilibrium test:
    --------------------------------
        No. of steps in Markov chain      = 1000000
        No. of Dememorisation Steps       = 100000
        Required precision on Probability = 0
        Test association at the Locus level


The computation of AMOVA with the option "CONVENTIONAL FST" is
not possible with unknown gametic phase data.
SKIPPING COMPUTATION OF GENETIC STRUCTURE

 #[WARNING # 9] : unable to compute Genetic Structure with current settings

================================================================================
== ANALYSES AT THE INTRA-POPULATION LEVEL
================================================================================


================================================================================
== Sample :    Control
================================================================================


================================
== Hardy-Weinberg equilibrium : (Control)
================================

Reference: Guo, S. and Thompson, E. 1992.
           Levene H. (1949).

Exact test using a Markov chain (for all Loci):
Forecasted chain length     :1000000
Dememorization steps        :100000
```

```
-------------------------------------------------------------------------
Locus  #Genot       Obs.Het.     Exp.Het.   P-value     s.d.  Steps done
-------------------------------------------------------------------------
    1      74       0.40541      0.43133    0.59803  0.00050     1001000
    2      74       0.54054      0.47362    0.32074  0.00050     1001000
    3      74       0.41892      0.49228    0.23928  0.00042     1001000
    4      74       0.02703      0.02684    1.00000  0.00000     1001000
    5      74       0.45946      0.43133    0.78445  0.00041     1001000
    6      74       0.31081      0.28268    0.67796  0.00046     1001000
    7      74       0.58108      0.47022    0.04933  0.00022     1001000
    8      74       0.40541      0.45891    0.32139  0.00045     1001000
    9      74       0.41892      0.40329    1.00000  0.00000     1001000
   10      74       0.16216      0.15003    1.00000  0.00000     1001000
-------------------------------------------------------------------------

=========================================================================
== Summary of computations done within populations
=========================================================================


/////////////////////////////////////////////////////////////////
END OF RUN NUMBER 1 (20/06/13 at 14:01:37))
Total computing time for this run : 0h 0m 2s 792 ms
/////////////////////////////////////////////////////////////////
```

Legend

Locus 1: *NOD1* (rs2075820); Locus 2: *CXCL16* (rs2277680); Locus 3: *STAT6* (rs324015); Locus 4: *TLR4* (rs4986791);

Locus 5: *ATG16L1* (rs2241880); Locus 6: *ATG16L1* (rs6758317); Locus 7: *ATG16L1* (rs6754677); Locus 8: *IRGM* (rs4958847);

Locus 9: *IRGM* (11747270); Locus 10: *IRGM* (rs72553867)

Appendix H: Hardy-Weinberg Equilibrium Analysis for Chinese Control Cohort Data by Arlequin 3.0

```
//////////////////////////////////////////////////////////////////////
RUN NUMBER 1 (20/06/13 at 14:00:21)
//////////////////////////////////////////////////////////////////////

 #[WARNING # 1] : bad value assigned to keyword in project file
 #[WARNING # 2] : invalid recessive allele, switching off recessive data


Project information:
--------------------
       NbSamples      =  1
       DataType       =  STANDARD
       GenotypicData  =  1
       GameticPhase   =  0
       RecessiveData  =  0

==============================
Settings used for Calculations
==============================

 General settings:
 -----------------
        Deletion Weight                = 1
        Transition Weight Weight       = 1
        Tranversion Weight Weight      = 1
        Epsilon Value                  = 1e-07
        Significant digits for output  = 5
        Use original haplotype definition
        Alllowed level of missing data = 0.05
```

```
Active Tasks:
 -------------

    Hardy-Weinberg equilibrium test:
    --------------------------------
        No. of steps in Markov chain      = 1000000
        No. of Dememorisation Steps       = 100000
        Required precision on Probability = 0
        Test association at the Locus level



The computation of AMOVA with the option "CONVENTIONAL FST" is
not possible with unknown gametic phase data.
SKIPPING COMPUTATION OF GENETIC STRUCTURE

 #[WARNING # 3] : unable to compute Genetic Structure with current settings

================================================================================
== ANALYSES AT THE INTRA-POPULATION LEVEL
================================================================================

================================================================================
== Sample :    Control
================================================================================

================================
== Hardy-Weinberg equilibrium : (Control)
================================

Reference: Guo, S. and Thompson, E. 1992.
           Levene H. (1949).

Exact test using a Markov chain (for all Loci):
Forecasted chain length     :1000000
Dememorization steps        :100000
```

```
-------------------------------------------------------------------------
Locus  #Genot       Obs.Het.     Exp.Het.   P-value     s.d.  Steps done
-------------------------------------------------------------------------
    1      86       0.47674      0.43758    0.46393  0.00049     1001000
    2      86       0.47674      0.45335    0.81070  0.00039     1001000
    3      86       0.41860      0.50265    0.13759  0.00033     1001000
    4   This locus is monomorphic: no test done.
    5      86       0.45349      0.41962    0.60317  0.00050     1001000
    6      86       0.18605      0.22440    0.13120  0.00034     1001000
    7      86       0.44186      0.50265    0.28577  0.00045     1001000
    8      86       0.44186      0.47001    0.64563  0.00046     1001000
    9      86       0.50000      0.50122    1.00000  0.00000     1001000
   10      86       0.33721      0.28199    0.11433  0.00034     1001000
-------------------------------------------------------------------------


=========================================================================
== Summary of computations done within populations
=========================================================================


/////////////////////////////////////////////////////////////////
END OF RUN NUMBER 1 (20/06/13 at 14:00:23))
Total computing time for this run : 0h 0m 2s 511 ms
/////////////////////////////////////////////////////////////////
```

Legend

Locus 1: *NOD1* (rs2075820); Locus 2: *CXCL16* (rs2277680); Locus 3: *STAT6* (rs324015); Locus 4: *TLR4* (rs4986791);

Locus 5: *ATG16L1* (rs2241880); Locus 6: *ATG16L1* (rs6758317); Locus 7: *ATG16L1* (rs6754677); Locus 8: *IRGM* (rs4958847);

Locus 9: *IRGM* (11747270); Locus 10: *IRGM* (rs72553867)

Appendix I: Hardy-Weinberg Equilibrium Analysis for Indian Control Cohort Data by Arlequin 3.0

```
//////////////////////////////////////////////////////////////////
RUN NUMBER 1 (20/06/13 at 14:01:04)
//////////////////////////////////////////////////////////////////

 #[WARNING # 4] : bad value assigned to keyword in project file
 #[WARNING # 5] : invalid recessive allele, switching off recessive data


Project information:
--------------------
       NbSamples      =  1
       DataType       =  STANDARD
       GenotypicData  =  1
       GameticPhase   =  0
       RecessiveData  =  0

==============================
Settings used for Calculations
==============================

 General settings:
 -----------------
        Deletion Weight                = 1
        Transition Weight Weight       = 1
        Tranversion Weight Weight      = 1
        Epsilon Value                  = 1e-07
        Significant digits for output  = 5
        Use original haplotype definition
        Alllowed level of missing data = 0.05
```

```
Active Tasks:
 -------------

    Hardy-Weinberg equilibrium test:
    --------------------------------
        No. of steps in Markov chain      = 1000000
        No. of Dememorisation Steps       = 100000
        Required precision on Probability = 0
        Test association at the Locus level


The computation of AMOVA with the option "CONVENTIONAL FST" is
not possible with unknown gametic phase data.
SKIPPING COMPUTATION OF GENETIC STRUCTURE

 #[WARNING # 6] : unable to compute Genetic Structure with current settings

================================================================================
== ANALYSES AT THE INTRA-POPULATION LEVEL
================================================================================

================================================================================
== Sample :    Control
================================================================================

================================
== Hardy-Weinberg equilibrium : (Control)
================================

Reference: Guo, S. and Thompson, E. 1992.
           Levene H. (1949).

Exact test using a Markov chain (for all Loci):
Forecasted chain length     :1000000
Dememorization steps        :100000
```

```
-------------------------------------------------------------------------
Locus  #Genot       Obs.Het.     Exp.Het.   P-value     s.d.  Steps done
-------------------------------------------------------------------------
    1      90       0.56667      0.49777    0.20639  0.00042     1001000
    2      90       0.44444      0.50056    0.30032  0.00044     1001000
    3      90       0.53333      0.48690    0.39313  0.00049     1001000
    4      90       0.21111      0.25642    0.10372  0.00032     1001000
    5      90       0.51111      0.50255    1.00000  0.00000     1001000
    6      90       0.38889      0.39845    0.79587  0.00040     1001000
    7      90       0.43333      0.48883    0.28801  0.00043     1001000
    8      90       0.42222      0.48268    0.27393  0.00044     1001000
    9      90       0.40000      0.45413    0.34681  0.00047     1001000
   10      90       0.11111      0.10552    1.00000  0.00000     1001000
-------------------------------------------------------------------------


=========================================================================
== Summary of computations done within populations
=========================================================================

/////////////////////////////////////////////////////////////////
END OF RUN NUMBER 1 (20/06/13 at 14:01:07))
Total computing time for this run : 0h 0m 2s 792 ms
/////////////////////////////////////////////////////////////////
```

Legend

Locus 1: *NOD1* (rs2075820); Locus 2: *CXCL16* (rs2277680); Locus 3: *STAT6* (rs324015); Locus 4: *TLR4* (rs4986791);

Locus 5: *ATG16L1* (rs2241880); Locus 6: *ATG16L1* (rs6758317); Locus 7: *ATG16L1* (rs6754677); Locus 8: *IRGM* (rs4958847);

Locus 9: *IRGM* (11747270); Locus 10: *IRGM* (rs72553867)

Appendix J: Genotype association analysis for *ATG16L1* (rs2241880)

Yang et al. (2007)	**Genotype**			***Yamazaki et al. (2009)***	**Genotype**		
	GG*	**GA**	**AA**		**GG**	**GA**	**AA**
CD	43	156	178	**CD**	23	184	274
Control	40	146	186	**Control**	32	167	238
Model (*P*-value)				**Model (*P*-value)**			
Dominant	0.501			**Dominant**	0.465		
Recessive	0.888			**Recessive**	0.125		
Additive	0.715			**Additive**	0.118		
Hampe et al. (2007)	**Genotype**			***Csongei et al. (2010)***	**Genotype**		
	GG	**GA**	**AA**		**GG**	**GA**	**AA**
CD	175	232	73	**CD**	108	151	56
Control	219	435	185	**Control**	79	163	72
Model (*P*-value)				**Model (*P*-value)**			
Dominant	2.58×10^{-3}			**Dominant**	0.114		
Recessive	7.7×10^{-5}			**Recessive**	0.0145		
Additive	3.38×10^{-5}			**Additive**	0.0161		

*G – risk allele

REFERENCE

Abel, S., Hundhausen, C., Mentlein, R., Schulte, A., Berkhout, T.A., Broadway, N., . . . Ludwig, A. (2004). The Transmembrane CXC-Chemokine Ligand 16 Is Induced by IFN-gamma and TNF-alpha and Shed by the Activity of the Disintegrin-Like Metalloproteinase ADAM10. *The Journal of Immunology,* **172**: 6362-6372.

Aizawa, H., Kinouchi, Y., Negoro, K., Kakuta, Y., Nomura, E., Seiichi, . . . Shimosegawa, T. (2008). Lack of Association Between IRGM and Crohn's Disease in Japanese Patients Demonstrates Population-Specific Differences in Inflammatory Bowel Disease. *Gastroenterology,* **134**(4): A-460.

Amre, D.K., D'Souza, S., Morgan, K., Seidman, G., Lambrette, P., Grimard, G., . . . Seidman, E.G. (2007). Imbalances in Dietary Consumption of Fatty Acids, Vegetables, and Fruits Are Associated With Risk for Crohn's Disease in Children. *The American Journal of Gastroenterology,* **102**(9): 2016-2025. doi: 10.1111/j.1572-0241.2007.01411.x

Amre, D.K., Lambrette, P., Law, L., Krupoves, A., Chotard, V., Costea, F., . . . Seidman, E.G. (2006). Investigating the Hygiene Hypothesis as a Risk Factor in Pediatric Onset Crohn's Disease: A Case-Control Study. *The American Journal of Gastroenterology,* **101**(5): 1005-1011. doi: 10.1111/j.1572-0241.2006.00526.x

Andersen, V., Olsen, A., Carbonnel, F., Tjønneland, A., & Vogel, U. (2012). Diet and risk of inflammatory bowel disease. *Digestive and Liver Disease,* **44**(3): 185-194. doi: 10.1016/j.dld.2011.10.001

Barrett, J.C., Hansoul, S., Nicolae, D.L., Cho, J.H., Duerr, R.H., Rioux, J.D., . . . Daly, M.J. (2008). Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nature Genetics,* **40**(8): 955-962. doi: 10.1038/ng.175

Bekpen, C., Xavier, R.J., & Eichler, E.E. (2010). Human IRGM gene "to be or not to be". *Seminars in Immunopathology,* **32**(4): 437-444. doi: 10.1007/s00281-010-0224-x

Bhanoori, M., Deenadayal, M., Kennedy, S., & Shivaji, S. (2007). The G2964A 3'-untranslated region polymorphism of the signal transducer and activator of transcription 6 gene is associated with endometriosis in South Indian women. *Human Reproduction,* **22**(4): 1026-1030. doi: 10.1093/humrep/del486

Bland, J.M., & Altman, D.G. (2000). Statistics notes: the odds ratio. *BMJ: British Medical Journal,* **320**(7247): 1468.

Bröms, G., Granath, F., Linder, M., Stephansson, O., Elmberg, M., & Kieler, H. (2012). Complications From Inflammatory Bowel Disease During Pregnancy and Delivery. *Clinical Gastroenterology and Hepatology*. doi: 10.1016/j.cgh.2012.08.018

Buller, H.A. (1997). Problems in diagnosis of IBD in children. *Neth J Med,* **50**(2): S8-11.

Büning, C., Durmus, T., Molnar, T., de Jong, D.J., Drenth, J.P.H., Fiedler, T., . . . Witt, H. (2007). A study in three European IBD cohorts confirms that the ATG16L1 c.898A>G (p.Thr300Ala) variant is a susceptibility factor for Crohn's disease. *Journal of Crohn's and Colitis,* **1**(2): 70-76. doi: 10.1016/j.crohns.2007.08.001

Cadwell, K., Liu, J.Y., Brown, S.L., Miyoshi, H., Loh, J., Lennerz, J.K., . . . Virgin Iv, H.W. (2008). A key role for autophagy and the autophagy gene Atg16l1 in mouse and human intestinal Paneth cells. *Nature,* **456**(7219): 259-263. doi: 10.1038/nature07416

Carbonnel, F., Jantchou, P., Monnet, E., & Cosnes, J. (2009). Environmental risk factors in Crohn's disease and ulcerative colitis: an update. *Gastroentérologie Clinique et Biologique,* **33, Suppl. 3**: S145—S157.

Card, T., Hubbard, R., & Logan, R.F.A. (2003). Mortality in Inflammatory Bowel Disease: A Population-Based Cohort Study. *Gastroenterology,* **125**: 1583–1590. doi: 10.1053/j.gastro.2003.09.029

Cario, E., & Podolsky, D.K. (2000). Differential Alteration in Intestinal Epithelial Cell Expression of Toll-Like Receptor 3 (TLR3) and TLR4 in Inflammatory Bowel Disease. *Infection and Immunity,* **68**(12): 7010-7017.

Carneiro, L.A.M., Travassos, L.H., & Philpott, D.J. (2004). Innate immune recognition of microbes through Nod1 and Nod2: implications for disease. *Microbes and Infection,* **6**(6): 609-616. doi: 10.1016/j.micinf.2004.01.012

Chamaillard, M., Hashimoto, M., Horie, Y., Qiu, J.M., Saab, L., Ogura, Y., . . . Inohara, N. (2003). An essential role for NOD1 in host recognition of bacterial peptidoglycan containing diaminopimelic acid. *Nature Immunology,* **4**(7): 702-707.

Chandrasekar, B. (2004). CXCL16 Signals via Gi, Phosphatidylinositol 3-Kinase, Akt, I B Kinase, and Nuclear Factor- B and Induces Cell-Cell Adhesion and Aortic Smooth Muscle Cell Proliferation. *Journal of Biological Chemistry,* **279**(5): 3188-3196. doi: 10.1074/jbc.M311660200

Chatterjee, S., & Pal, J.K. (2009). Role of 5′- and 3′-untranslated regions of mRNAs in human diseases. *Biology of the Cell,* **101**(5): 251-262. doi: 10.1042/bc20080104

Chen, G.Y., Shaw, M.H., Redondo, G., & Nunez, G. (2008). The Innate Immune Receptor Nod1 Protects the Intestine from Inflammation-Induced Tumorigenesis. *Cancer Research,* **68**(24): 10060-10067. doi: 10.1158/0008-5472.can-08-2061

Cho, J.H. (2008). The genetics and immunopathogenesis of inflammatory bowel disease. *Nature Reviews Immunology,* **8**(6): 458-466. doi: 10.1038/nri2340

Chua, K.H., Ida Hilmi, Ng, C.C., Eng, T.L., Shanthi Palaniappa, Lee, W.S., & Goh, K.-L. (2009). Identification of NOD2/CARD15 mutations in Malaysian patients with Crohn's disease *Journal of Digestive Diseases,* **10**: 124–130. doi: 10.1111/j.1751-2980.2009.00374x

Cooney, R., Baker, J., Brain, O., Danis, B., Pichulik, T., Allan, P., . . . Simmons, A. (2010). NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. *Nature Medicine,* **16**(1): 90-97. doi: 10.1038/nm.2069

Crohn, B.B., Ginzburg, L., & Oppenheimer, G.D. (1932). Regional ileitis: A pathologic and clinical entity. *JAMA: The Journal of the American Medical Association,* **99**(16): 1323-1329. doi: 10.1001/jama.1932.02740680019005

Csöngei, V. (2010). Interaction of the major inflammatory bowel disease susceptibility alleles in Crohn's disease patients. *World Journal of Gastroenterology,* **16**(2): 176. doi: 10.3748/wjg.v16.i2.176

Cummings, J.R.F., Cooney, R., Pathan, S., Anderson, C.A., Barrett, J.C., Beckly, J., . . . Jewell, D.P. (2007). Confirmation of the role ofATG16l1 as a Crohn's disease susceptibility gene. *Inflammatory Bowel Diseases,* **13**(8): 941-946. doi: 10.1002/ibd.20162

Diegelmann, J., Seiderer, J., Niess, J.-H., Haller, D., Göke, B., Reinecker, H.-C., & Brand, S. (2010). Expression and regulation of the chemokine CXCL16 in Crohn's disease and models of intestinal inflammation. *Inflammatory Bowel Diseases,* **16**(11): 1871-1881. doi: 10.1002/ibd.21306

Dimon, C., Allen, M., Heel, D.A.V., Mackay, I., Lench, N.J., Jewell, D.P., & Carey, A.H. (2001). Family based association studies of STAT6, a positional candidate gene for IBD. *Gastroenterology,* **120**(5): A458.

Dorak, M.T. (2012). Basic Population Genetics. Retrieved 17 June 2013, from http://dorakmt.tripod.com/genetics/popgen.html

Dotan, I., Alper, A., Rachmilewitz, D., Israeli, E., Odes, S., Chermesh, I., . . . Reif, S. (2012). Maternal inflammatory bowel disease has short and long-term effects on the health of their offspring: A multicenter study in Israel. *Journal of Crohn's and Colitis*. doi: 10.1016/j.crohns.2012.08.012

Economou, M., & Pappas, G. (2008). New global map of Crohn's disease: Genetic, environmental, and socioeconomic correlations. *Inflammatory Bowel Diseases,* **14**(5): 709-720. doi: 10.1002/ibd.20352

Elo, L.L., Järvenpää, H., Tuomela, S., Raghav, S., Ahlfors, H., Laurila, K., . . . Lahesmaa, R. (2010). Genome-wide Profiling of Interleukin-4 and STAT6 Transcription Factor Regulation of Human Th2 Cell Programming. *Immunity,* **32**(6): 852-862. doi: 10.1016/j.immuni.2010.06.011

Excoffier, Laval, L.G., & Schneider, S. (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis., **1**: 47-50.

Ferrero, R.L., & Philpott, D.J. (2005). Nod1 signaling to extracellular mucosal bacteria in epithelial cells. *International Congress Series,* **1285**: 68-77. doi: 10.1016/j.ics.2005.07.097

Fischer, W., Püls, J., Buhrdorf, R., Gebert, B., Odenbreit, S., & Haas, R. (2001). Systematic mutagenesis of the Helicobacter pylori cag pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8. *Molecular Microbiology,* **42**(5): 1337-1348. doi: 10.1046/j.1365-2958.2001.02714.x

Fisher, R.A. (1922). On the Interpretation of $\chi 2$ from Contingency Tables, and the Calculation of P. *Journal of the Royal Statistical Society,* **85**(1): 87-94.

Freeman, G.H., & Halton, J.H. (1951). Note on an Exact Treatment of Contingency, Goodness of Fit and Other Problems of Significance. *Biometrika,* **38**(1/2): 141-149. doi: 10.2307/2332323

Frolova, L., Drastich, P., Rossmann, P., Klimesova, K., & Tlaskalova-Hogenova, H. (2008). Expression of Toll-like Receptor 2 (TLR2), TLR4, and CD14 in Biopsy Samples of Patients With Inflammatory Bowel Diseases: Upregulated Expression of TLR2 in Terminal Ileum of Patients With Ulcerative Colitis. *Journal of Histochemistry and Cytochemistry,* **56**(3): 267-274. doi: 10.1369/jhc.7A7303.2007

Fukata, M., Vamadevan, A.S., & Abreu, M.T. (2009). Toll-like receptors (TLRs) and Nod-like receptors (NLRs) in inflammatory disorders. *Seminars in Immunology,* **21**(4): 242-253. doi: 10.1016/j.smim.2009.06.005

Gao, P.S., Mao, X.Q., Roberts, M.H., Arinobu, Y., Akaiwa, M., Enomoto, T., . . . Hopkin, J.M. (2000). Variants ofSTAT6(signal transducerand activator of transcription 6) in atopic asthma. *J Med Genet,* **37**: 380-382.

Gaudermon. (2002). Sample size requirements for matched case-control studies of gene-environment interaction. *Statistics in Medicine,* **21**: 35-50.

Ghanchi, F.D., & Rembacken, B.J. (2003). Inflammatory bowel disease and the eye. *Survey of Ophthalmology,* **48**(6): 663-676. doi: 10.1016/j.survophthal.2003.08.004

Glas, J., Konrad, A., Schmechel, S., Dambacher, J., Seiderer, J., Schroff, F., . . . Brand, S. (2008). TheATG16L1Gene Variants rs2241879 and rs2241880 (T300A) Are Strongly Associated With Susceptibility to Crohn's Disease in the German Population. *The American Journal of Gastroenterology,* **103**(3): 682-691. doi: 10.1111/j.1572-0241.2007.01694.x

Goh, K.L., & Xiao, S.-D. (2009). Inflammatory bowel disease: A survey of the epidemiology in Asia. *Journal of Digestive Diseases,* **10**(1): 1-6. doi: 10.1111/j.1751-2980.2008.00355.x

Gough, P.J., Garton, K.J., Wille, P.T., Rychlewski, M., Dempsey, P.J., & Raines, E.W. (2004). A Disintegrin and Metalloproteinase 10-Mediated Cleavage and Shedding Regulates the Cell Surface Expression of CXC Chemokine Ligand 16. *The Journal of Immunology,* **172**: 3678-3685.

Grégoire, I.P., Richetta, C., Meyniel-Schicklin, L., Borel, S., Pradezynski, F., Diaz, O., . . . Faure, M. (2011). IRGM Is a Common Target of RNA Viruses that Subvert the Autophagy Network. *PLoS Pathogens,* **7**(12): e1002422. doi: 10.1371/journal.ppat.1002422

Hamlin PJ, Komolmit P, Bransfield K, Jones PF, Smith NR, Aldersley MA, et al. Identification of Multiple Candidate Genes for IBD Susceptibility Using High-Density Transcript Mapping in the IBD2 Locus on Chromosome 12q. Gastroenterology. 1999; 117: 1029-31.

Hampe, J., Franke, A., Rosenstiel, P., Till, A., Teuber, M., Huse, K., . . . Schreiber, S. (2007). A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nature Genetics,* **39**(2): 207-211. doi: 10.1038/ng1954

Hilmi, I., Tan, Y.M., & Goh, K.L. (2006). Crohn's disease in adults: Observations in a multiracial Asian population. *World J Gastroenterol* **12**(9): 1435-1438.

Hofner, P., Gyulai, Z., Kiss, Z.F., Tiszai, A., Tiszlavicz, L., Tóth, G., . . . Mándi, Y. (2007). Genetic Polymorphisms of NOD1 and IL-8, but not Polymorphisms of TLR4 Genes, Are Associated with Helicobacter pylori-Induced Duodenal Ulcer and Gastritis. *Helicobacter,* **12**(2): 124-131. doi: 10.1111/j.1523-5378.2007.00481.x

Homer, C.R., Richmond, A.L., Rebert, N.A., Achkar, J.P., & McDonald, C. (2010). ATG16L1 and NOD2 Interact in an Autophagy-Dependent Antibacterial Pathway Implicated in Crohn's Disease Pathogenesis. *Gastroenterology,* **139**(5): 1630-1641.e1632. doi: 10.1053/j.gastro.2010.07.006

Hsiao, C.H., Wei, S.C., Wong, J.M., Lai, H.S., Chang, M.H., & Ni, Y.H. (2007). Pediatric Crohn Disease: Clinical and Genetic Characteristics in Taiwan. *Journal of Pediatric Gastroenterology and Nutrition,* **44**: 342–346.

Hudson, M., Flett, G., Sinclair, T.S., Templeton, P.W.B.A., & Mowat, N.A.G. (1997). Fertility and pregnancy in inflammatory bowel disease *International Journal of Gynecology & Obstetrics,* **58**: 229-237.

Hugot, J.P., Chamaillard, M., Zouali, H., Lesage, S., Cézard, J.P., Belaiche, J., . . . Thomas, G. (2001). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature,* **411**: 599-603.

Hysi, P. (2005). NOD1 variation, immunoglobulin E and asthma. *Human Molecular Genetics,* **14**(7): 935-941. doi: 10.1093/hmg/ddi087

Inohara, N. (2003). Host Recognition of Bacterial Muramyl Dipeptide Mediated through NOD2. IMPLICATIONS FOR CROHN'S DISEASE. *Journal of Biological Chemistry,* **278**(8): 5509-5512. doi: 10.1074/jbc.C200673200

Inohara, N., Koseki, T., Peso, L.d., Hu, Y., Yee, C., Chen, S., . . . Nunez, G. (1999). Nod1, an Apaf-1-like Activator of Caspase-9 and Nuclear Factor-kB. *The Journal of Biological Chemistry,* **274**(21): 14560-14567.

Jong, D.J.d., Franke, B., Naber, A.H.J., Willemen, J.J.H.T., Heister, A.J., Brunner, H.G., . . . Hol, F.A. (2003). No evidence for involvement of IL-4R and CD11B from the IBD1 region and STAT6 in the IBD2 region in Crohn's disease. *European Journal of Human Genetics,* **11**(11): 884-887. doi: 10.1038/sj.ejhg.5201058

Kaplan, G.G., Hubbard, J., Korzenik, J., Sands, B.E., Panaccione, R., Ghosh, S., . . . Villeneuve, P.J. (2010). The Inflammatory Bowel Diseases and Ambient Air

Pollution: A Novel Association. *The American Journal of Gastroenterology,* **105**(11): 2412-2419. doi: 10.1038/ajg.2010.252

Kara, B., Akkiz, H., Doran, F., Bayram, S., Erken, E., Gumurdullu, Y., & Sandikci, M. (2010). The significance of E266K polymorphism in the NOD1 gene on Helicobacter Pylori infection: an effective force on pathogenesis? *Clinical and Experimental Medicine,* **10**(2): 107-112. doi: 10.1007/s10238-009-0077-6

Kersse, K., Bertrand, M.J.M., Lamkanfi, M., & Vandenabeele, P. (2011). NOD-like receptors and the innate immune system: Coping with danger, damage and death. *Cytokine & Growth Factor Reviews,* **22**(5-6): 257-276. doi: 10.1016/j.cytogfr.2011.09.003

Kim, E.J., Chung, W.C., Lee, K.-M., Paik, C.N., Jung, S.H., Lee, B.I., . . . Choi, K.Y. (2012). Association Between Toll-Like Receptors/CD14Gene Polymorphisms and Inflammatory Bowel Disease in Korean Population. *Journal of Korean Medical Science,* **27**(1): 72. doi: 10.3346/jkms.2011.27.1.72

Kim, J.G., Lee, S.J., & Kagnoff, M.F. (2004). Nod1 Is an Essential Signal Transducer in Intestinal Epithelial Cells Infected with Bacteria That Avoid Recognition by Toll-Like Receptors. *Infection and Immunity,* **72**(3): 1487-1495. doi: 10.1128/iai.72.3.1487-1495.2004

Klein, W., Tromm, A., Folwaczny, C., Hagedorn, M., Duerig, N., Epplen, J., . . . Griga, T. (2005). The G2964A polymorphism of the STAT6 gene in inflammatory bowel disease. *Digestive and Liver Disease,* **37**(3): 159-161. doi: 10.1016/j.dld.2004.10.011

Komine, O., Hayashi, K., Natsume, W., Watanabe, T., Seki, Y., Seki, N., . . . Satake, M. (2003). The Runx1 Transcription Factor Inhibits the Differentiation of Naive CD4+ T Cells into the Th2 Lineage by Repressing GATA3 Expression. *Journal of Experimental Medicine,* **198**(1): 51-61. doi: 10.1084/jem.20021200

Koonin, E.V., & Aravind, L. (2000). The NACHT family - a new group of predicted NTPases implicated in apoptosis and MHC transcription activation. *Trends in Biochemical Sciences,* **25**(5): 223-224.

Kuballa, P., Huett, A., Rioux, J.D., Daly, M.J., & Xavier, R.J. (2008). Impaired Autophagy of an Intracellular Pathogen Induced by a Crohn's Disease Associated ATG16L1 Variant. *PLoS ONE,* **3**(10): e3391. doi: 10.1371/journal.pone.0003391

Kupcinskas, J., Wex, T., Bornschein, J., Selgrad, M., Leja, M., Juozaityte, E., . . . Malfertheiner, P. (2011). Lack of association between gene polymorphisms of Angiotensin converting enzyme, *Nod-like receptor 1*, *Toll-like receptor* 4, *FAS/FASL* and the presence of *Helicobacter pylori*-induced premalignant gastric lesions and gastric cancer in Caucasians. *BMC Medical Genetics,* **12**: 112.

Lacher, M., Kappler, R., Berkholz, S., Baurecht, H., von Schweinitz, D., & Koletzko, S. (2007). Association of a CXCL9 polymorphism with pediatric Crohn's disease. *Biochemical and Biophysical Research Communications,* **363**(3): 701-707. doi: 10.1016/j.bbrc.2007.09.020

Lakatos, P.L., Szamosi, T., Szilvasi, A., Molnar, E., Lakatos, L., Kovacs, A., . . . Andrikovics, H. (2008). ATG16L1 and IL23 receptor (IL23R) genes are associated with disease susceptibility in Hungarian CD patients. *Digestive and Liver Disease,* **40**(11): 867-873. doi: 10.1016/j.dld.2008.03.022

Latiano, A., Palmieri, O., Cucchiara, S., Castro, M., D'Incà, R., Guariso, G., . . . Annese, V. (2009). Polymorphism of the IRGM Gene Might Predispose to Fistulizing Behavior in Crohn's Disease. *The American Journal of Gastroenterology,* **104**(1): 110-116. doi: 10.1038/ajg.2008.3

Lavelle, E.C., Murphy, C., O'Neill, L.A.J., & Creagh, E.M. (2010). The role of TLRs, NLRs, and RLRs in mucosal innate immunity and homeostasis. *Mucosal Immunology,* **3**(1): 17-28. doi: 10.1038/mi.2009.124

Lehrke, M., Konrad, A., Schachinger, V., Tillack, C., Seibold, F., Stark, R., . . . Broedl, U.C. (2008). CXCL16 is a surrogate marker of inflammatory bowel disease. *Scandinavian Journal of Gastroenterology,* **43**(3): 283-288. doi: 10.1080/00365520701679249

Lewis, C.M. (2002). Genetic association studies: Design, analysis and interpretation. *Briefings in Bioinformatics,* **3**(2): 146-153. doi: 10.1093/bib/3.2.146

Lewis, C.M., & Knight, J. (2012). Introduction to Genetic Association Studies. *Cold Spring Harbor Protocols,* **2012**(3): pdb.top068163. doi: 10.1101/pdb.top068163

Li, D., & Roberts, R. (2001). Human Genome and Diseases:¶WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. *Cellular and Molecular Life Sciences CMLS,* **58**(14): 2085-2097. doi: 10.1007/PL00000838

Loftus, E.V. (2004). Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. *Gastroenterology,* **126**(6): 1504-1517. doi: 10.1053/j.gastro.2004.01.063

Lu, R., Gao, X., Chen, Y., Ni, J., Yu, Y., Li, S., & Guo, L. (2012). Association of an NFKB1 intron SNP (rs4648068) with gastric cancer patients in the Han Chinese population. *BMC Gastroenterol,* **12**: 87. doi: 10.1186/1471-230x-12-87

Lundberg, G.A., Kellin, A., Negard, A.S., Lundman, P., Tornvall, P., Dimmeler, S., . . . Eriksson, P. (2005). Severity of coronary artery stenosis is associated with a polymorphism in the CXCL16/SR-PSOX gene. *Journal of Internal Medicine,* **257**: 415-422.

Márquez, A., Núñez, C., Martínez, A., Mendoza, J.L., Taxonera, C., Fernández-Arquero, M., . . . Urcelay, E. (2009). Role ofATG16L1Thr300Ala polymorphism in inflammatory bowel disease: A Study in the Spanish population and a meta-analysis. *Inflammatory Bowel Diseases,* **15**(11): 1697-1704. doi: 10.1002/ibd.21001

Matloubian, M., David, A., Engel, S., Ryan, J.E., & Cyster, J.G. (2000). A transmembrane CXC chemokine is a ligand for HIV-coreceptor Bonzo. *Nature Immunology,* **1**(4): 298-304.

Mazumder, B., Seshadri, V., & Fox, P.L. (2003). Translational control by the 3′-UTR: the ends specify the means. *Trends in Biochemical Sciences,* **28**(2): 91-98. doi: 10.1016/s0968-0004(03)00002-1

McGovern, D.P.B. (2005). Association between a complex insertion/deletion polymorphism in NOD1 (CARD4) and susceptibility to inflammatory bowel disease. *Human Molecular Genetics,* **14**(10): 1245-1250. doi: 10.1093/hmg/ddi135

Mehta, C.R., & Patel, N.R. (1983). A Network Algorithm for Performing Fisher's Exact Test in r × c Contingency Tables. *Journal of the American Statistical Association,* **78**(382): 427-434. doi: 10.2307/2288652

Mizushima, N. (2003). Mouse Apg16L, a novel WD-repeat protein, targets to the autophagic isolation membrane with the Apg12-Apg5 conjugate. *Journal of Cell Science,* **116**(9): 1679-1688. doi: 10.1242/jcs.00381

Molnar, T., Hofner, P., Nagy, F., Lakatos, P.L., Fischer, S., Lakatos, L., . . . Lonovics, J. (2007). NOD1 gene E266K polymorphism is associated with disease susceptibility but not with disease phenotype or NOD2/CARD15 in Hungarian patients with Crohn's disease. *Digestive and Liver Disease,* **39**(12): 1064-1070. doi: 10.1016/j.dld.2007.09.003

Moon, C.M., Shin, D.-J., Kim, S.W., Son, N.-H., Park, A., Park, B., . . . Cheon, J.H. (2012). Associations between genetic variants in the IRGM gene and inflammatory bowel diseases in the korean population. *Inflammatory Bowel Diseases*: n/a-n/a. doi: 10.1002/ibd.22972

Mrowicki, J., Przybyłowska, K., Dziki, Ł., Sygut, A., Wiśniewska-Jarosińska, M., Chojnacki, J., . . . Majsterek, I. (2011). Association of The -801G/A Polymorphism of *CXCL12* Gene with the Risk of Inflammatory Bowel Diseases Development in a Polish Population. *Polski Przegląd Chirurgiczny,* **83**(6): 334–338.

Murdoch, T.B., Bernstein, C.N., El-Gabalawy, H., Stempak, J.M., Sargent, M., Elias, B., . . . Silverberg, M.S. (2012). Prevalence of genetic variants associated with inflammatory bowel disease in a healthy First Nations cohort. *CMAJ,* **184**(8): E435-441. doi: 10.1503/cmaj.110613

Nakase, H., Matsuura, M., Mikami, S., Uza, N., & Chiba, T. (2012). Role of the CXC12-CXCR4 Axis and CXCL16 in Inflammatory Bowel Disease. *Intestinal Research,* **10**(2): 125. doi: 10.5217/ir.2012.10.2.125

Naoe, Y., Setoguchi, R., Akiyama, K., Muroi, S., Kuroda, M., Hatam, F., . . . Taniuchi, I. (2007). Repression of interleukin-4 in T helper type 1 cells by Runx/Cbf binding to the Il4 silencer. *Journal of Experimental Medicine,* **204**(8): 1749-1755. doi: 10.1084/jem.20062456

NCBI. National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov.

Noguchi, E., Homma, Y., Kang, X., Netea, M.G., & Ma, X. (2009). A Crohn's disease–associated NOD2 mutation suppresses transcription of human IL10 by inhibiting activity of the nuclear ribonucleoprotein hnRNP-A1. *Nature Immunology,* **10**(5): 471-479. doi: 10.1038/ni.1722

O'Neill, L.A.J., Bryant, C.E., & Doyle, S.L. (2009). Therapeutic Targeting of Toll-Like Receptors for Infectious and Inflammatory Diseases and Cancer. *Pharmacological Reviews,* **61**(2): 177-197. doi: 10.1124/pr.109.001073

Ogura, Y., Bonen, D.K., Inohara, N., Nicolae, D.L., Chen, F.F., Ramos, R., . . . Cho, J.H. (2001). A frameshift mutation in *NOD2* associated with susceptibility to Crohn's disease. *Nature,* **411**: 603-606.

Özen, S.C., Dağlı, Ü., Kiliç, M.Y., Törüner, M., Çelik, Y., Özkan, M., . . . Bozdayı, A.M. (2006). NOD2/CARD15, NOD1/CARD4, and ICAM-1 gene polymorphisms in Turkish patients with inflammatory bowel disease. *Journal of Gastroenterology,* **41**(4): 304-310. doi: 10.1007/s00535-005-1780-z

Palmon, R., & Mayer, L.F. (2006). Chapter 52 - Inflammatory Bowel Diseases: Ulcerative Colitis and Crohn's Disease. In R. R. Noel & R. M. Ian (Eds.), *The Autoimmune Diseases (Fourth Edition)* (pp. 713-XXIII). St. Louis: Academic Press.

Palomino-Morales, R.J., Oliver, J., Gómez-García, M., López-Nevot, M.A., Rodrigo, L., Nieto, A., . . . Martín, J. (2009). Association of ATG16L1 and IRGM genes polymorphisms with inflammatory bowel disease: a meta-analysis approach. *Genes and Immunity,* **10**(4): 356-364. doi: 10.1038/gene.2009.25

Parkes, M., Barrett, J.C., Prescott, N.J., Tremelling, M., Anderson, C.A., Fisher, S.A., . . . Mathew, C.G. (2007). Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nature Genetics,* **39**(7): 830-832. doi: 10.1038/ng2061

Peter, I., Mitchell, A.A., Ozelius, L., Erazo, M., Hu, J., Doheny, D., . . . Desnick, R.J. (2011). Evaluation of 22 genetic variants with Crohn's Disease risk in the Ashkenazi Jewish population: a case-control study. *BMC Medical Genetics,* **12**(1): 63. doi: 10.1186/1471-2350-12-63

Philpott, D., & Girardin, S. (2004). The role of Toll-like receptors and Nod proteins in bacterial infection. *Molecular Immunology,* **41**(11): 1099-1108. doi: 10.1016/j.molimm.2004.06.012

Plevy, S.E., & Reguiero, M. (2005). Chapter 46 - Autoimmunity — Inflammatory Bowel Disease. In T. L. Michael & W. T. Angus (Eds.), *Measuring Immunity* (pp. 525-542). London: Academic Press.

Prescott, N.J., Fisher, S.A., Franke, A., Hampe, J., Onnie, C.M., Soars, D., . . . Mathew, C.G. (2007). A Nonsynonymous SNP in ATG16L1 Predisposes to Ileal Crohn's Disease and Is Independent of CARD15 and IBD5. *Gastroenterology,* **132**(5): 1665-1671. doi: 10.1053/j.gastro.2007.03.034

Proell, M., Riedl, S.J., Fritz, J.H., Rojas, A.M., & Schwarzenbacher, R. (2008). The Nod-Like Receptor (NLR) Family: A Tale of Similarities and Differences. *PLoS ONE,* **3**(4): e2119. doi: 10.1371/journal.pone.0002119

Prohinar, P., Rallabhandi, P., Weiss, J.P., & Gioannini, T.L. (2010). Expression of Functional D299G.T399I Polymorphic Variant of TLR4 Depends More on Coexpression of MD-2 Than Does Wild-Type TLR4. *The Journal of Immunology,* **184**(8): 4362-4367. doi: 10.4049/jimmunol.0903142

Reddy, S.I., & Wolf, J.L. (2004). 40 - Gender Considerations in inflammatory Bowel Disease. In J. L. Marianne & M. D. Mda2 - Marianne J. Legato (Eds.), *Principles of Gender-Specific Medicine* (pp. 428-437). San Diego: Academic Press.

Roberts, R.L., Gearry, R.B., Hollis-Moffatt, J.E., Miller, A.L., Reid, J., Abkevich, V., . . . Kennedy, M.A. (2007). IL23R R381Q and ATG16L1 T300A Are Strongly Associated With Crohn's Disease in a Study of New Zealand Caucasians With Inflammatory Bowel Disease. *The American Journal of Gastroenterology,* **102**(12): 2754-2761. doi: 10.1111/j.1572-0241.2007.01525.x

Roberts, R.L., Hollis-Moffatt, J.E., Gearry, R.B., Kennedy, M.A., Barclay, M.L., & Merriman, T.R. (2008). Confirmation of association of IRGM and NCF4 with ileal Crohn's disease in a population-based cohort. *Genes and Immunity,* **9**(6): 561-565. doi: 10.1038/gene.2008.49

Romagnani, S. (1999). Th1/Th2 Cells *Inflammatory Bowel Diseases,* **5**(4): 285-294.

Saitoh, T., Fujita, N., Jang, M.H., Uematsu, S., Yang, B.-G., Satoh, T., . . . Akira, S. (2008). Loss of the autophagy protein Atg16L1 enhances endotoxin-induced IL-1β production. *Nature,* **456**(7219): 264-268. doi: 10.1038/nature07383

Schindler, C., Levy, D.E., & Decker, T. (2007). JAK-STAT Signaling: From Interferons to Cytokines. *Journal of Biological Chemistry,* **282**(28): 20059-20063. doi: 10.1074/jbc.R700016200

Scholz, F., Schulte, A., Adamski, F., Hundhausen, C., Mittag, J., Schwarz, A., . . . Ludwig, A. (2007). Constitutive Expression and Regulated Release of the Transmembrane Chemokine CXCL16 in Human and Murine Skin. *Journal of Investigative Dermatology,* **127**(6): 1444-1455. doi: 10.1038/sj.jid.5700751

Seiderer, J., Dambacher, J., Leistner, D., Tillack, C., Glas, J., Niess, J.-H., . . . Brand, S. (2008). Genotype–phenotype analysis of the CXCL16 p.Ala181Val polymorphism in inflammatory bowel disease. *Clinical Immunology,* **127**(1): 49-55. doi: 10.1016/j.clim.2007.11.016

Shimaoka, T., Nakayama, T., Kume, N., Takahashi, S., Yamaguchi, J., Minami, M., . . . Yonehara, S. (2003). Cutting Edge: SR-PSOX/CXC Chemokine Ligand 16 Mediates Bacterial Phagocytosis by APCs Through its Chemokine Domain1. *The Journal of Immunology,* **171**: 1647-1651.

Sim, J., & Reid, N. (1999). Statistical inference by confidence intervals: issues of interpretation and utilization. *Physical Therapy,* **79**(2): 186-195.

Singh, S.B. (2006). Human IRGM Induces Autophagy to Eliminate Intracellular Mycobacteria. *Science,* **313**(5792): 1438-1441. doi: 10.1126/science.1129577

Smith, T.F., Gaitatzes, C., Saxena, K., & Neer, E.J. (1999). The WD repeat: a common architecture for diverse functions. *Trends Biochem Sci,* **24**(5): 181-185.

Takeda, K., Kishimoto, T., & Akira, S. (1997). STAT6: its role in interleukin 4-mediated biological functions. *Journal of Molecular Medicine,* **75**: 317-326.

Tanabe, T., Yamaguchi, N., Eishi, Y., & Fujita, Y. (2011). The roles of NOD like receptors in inflammation are different between Japanese and Caucasian. *Inflammation and Regeneration,* **31**(2): 196-201.

Thia, K.T., Loftus, J.E.V., Sandborn, W.J., & Yang, S.-K. (2008). An Update on the Epidemiology of Inflammatory Bowel Disease in Asia. *The American Journal of Gastroenterology,* **103**(12): 3167-3182. doi: 10.1111/j.1572-0241.2008.02158.x

Thomas, G.A.O., Rhodes, J., & Green, J.T. (1998). Inflammatory Bowel Disease and Smoking-A Review *The American Journal of Gastroenterology,* **93**(2): 144-149.

Török, H.P., Glas, J., Tonenchi, L., Mussack, T., & Folwaczny, C. (2004). Polymorphisms of the lipopolysaccharide-signaling complex in inflammatory bowel disease: association of a mutation in the Toll-like receptor 4 gene with ulcerative colitis. *Clinical Immunology,* **112**(1): 85-91. doi: 10.1016/j.clim.2004.03.002

Travassos, L.H., Carneiro, L.A.M., Ramjeet, M., Hussey, S., Kim, Y.-G., Magalhães, J.G., . . . Philpott, D.J. (2010). Nod1 and Nod2 direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry. *Nature Immunology,* **11**(1): 55-62. doi: 10.1038/ni.1823

Trejo-de la O, A., Torres, J., Pérez-Rodríguez, M., Camorlinga-Ponce, M., Luna, L.F., Abdo-Francis, J.M., . . . Maldonado-Bernal, C. (2008). TLR4 single-nucleotide polymorphisms alter mucosal cytokine and chemokine patterns in Mexican patients with Helicobacter pylori-associated gastroduodenal diseases. *Clinical Immunology,* **129**(2): 333-340. doi: 10.1016/j.clim.2008.07.009

Uza, N., Nakase, H., Yamamoto, S., Yoshino, T., Takeda, Y., Ueno, S., . . . Chiba, T. (2011). SR-PSOX/CXCL16 plays a critical role in the progression of colonic inflammation. *Gut,* **60**(11): 1494-1505. doi: 10.1136/gut.2010.221879

Van Limbergen, J., Kabakchiev, B., Stempak, J.M., Schumm, P., Xu, W., Henderson, P., . . . Silverberg, M. (2012). Hypothesis-free analysis of ATG16L1 demonstrates gene-wide extent of association with Crohn's disease susceptibility. *Gut*. doi: 10.1136/gutjnl-2012-303165

Van Limbergen, J., Russell, R.K., Nimmo, E.R., Drummond, H.E., Smith, L., Anderson, N.H., . . . Satsangi, J. (2008). 393 The Contribution of Germline Variation in the Autophagy Genes ATG16l1 and Irgm to Childhood Onset Inflammatory Bowel Disease. *Gastroenterology,* **134**(4): A-53. doi: 10.1016/s0016-5085(08)60256-5

Wang, M.-H., Okazaki, T., Isaacs, K.L., Lewis, J.D., Smoot, D.T., Valentine, J.F., . . . Brant, S.R. (2009). In African Americans, NOD2, SLC22A5 and ATG16L1 But Not Il23R or IRGM Are Genetic Risk Factors for Crohn's Disease. *Gastroenterology,* **136**(5): A38.

Wang, P., Zhang, L., Jiang, J.M., Ma, D., Tao, H.X., Yuan, S.L., . . . Liu, C.J. (2012). Association of NOD1 and NOD2 genes polymorphisms withHelicobacter pylorirelated gastric cancer in a Chinese population. *World Journal of Gastroenterology,* **18**(17): 2112. doi: 10.3748/wjg.v18.i17.2112

Weersma, R.K., Stokkers, P.C.F., Cleynen, I., Wolfkamp, S.C.S., Henckaerts, L., Schreiber, S., . . . Vermeire, S. (2009). Confirmation of Multiple Crohn's Disease Susceptibility Loci in a Large Dutch–Belgian Cohort. *The American Journal of Gastroenterology,* **104**(3): 630-638. doi: 10.1038/ajg.2008.112

Weersma, R.K., Zhernakova, A., Nolte, I.M., Lefebvre, C., Rioux, J.D., Mulder, F., . . . Dijkstra, G. (2008). ATG16L1andIL23RAre Associated With Inflammatory Bowel Diseases but Not With Celiac Disease in The Netherlands. *The American Journal of Gastroenterology,* **103**(3): 621-627. doi: 10.1111/j.1572-0241.2007.01660.x

Weinstock, J.V. (2004). 51 - Inflammatory Bowel Disease. In S. Yehuda & R. R. Noel (Eds.), *Infection and Autoimmunity* (pp. 649-673). Amsterdam: Elsevier.

Wilbanks, A., Zondlo, S.C., Murphy, K., Mak, S., Soler, D., Langdon, P., . . . Briskin, M. (2001). Expression Cloning of the STRL33/BONZO/TYMSTR Ligand Reveals Elements of CC, CXC, and CX3C Chemokines. *The Journal of Immunology,* **166**: 5145-5154.

Wurster, A.L., Tanaka, T., & Grusby, M.J. (2000). The biology of Stat4 and Stat6. *Oncogene,* **19**: 2577-2584.

Xia, B., Crusius, J.B.A., Wu, J., Zwiers, A., Bodegraven, A.A.V., & Peña, A.S. (2003). Signal transducer and activator of transcription 6 gene G2964A polymorphism and inflammatory bowel disease. *Clin Exp Immunol,* **131**: 446–450.

Yamazaki, K., Onouchi, Y., Takazoe, M., Kubo, M., Nakamura, Y., & Hata, A. (2007). Association analysis of genetic variants in IL23R, ATG16L1 and 5p13.1 loci with Crohn's disease in Japanese patients. *Journal of Human Genetics,* **52**(7): 575-583. doi: 10.1007/s10038-007-0156-z

Yang, S.-K., Park, M., Lim, J., Park, S.H., Ye, B.D., Lee, I., & Song, K. (2009). Contribution ofIL23Rbut notATG16L1to Crohn's disease susceptibility in Koreans. *Inflammatory Bowel Diseases,* **15**(9): 1385-1390. doi: 10.1002/ibd.20921

Ye, B.D., Yang, S.K., Song, K., Yang, D.H., Yoon, S.M., Kim, K.J., . . . Kim, J.H. (2009). Association of Toll-Like Receptor Gene with Crohn's Disease in Koreans. *The Korean Journal of Gastroenterology,* **54**(6): 377. doi: 10.4166/kjg.2009.54.6.377

Zhang, W.J., Thompson, J.L., Poritz, L.S., Galka, E., & Kokun, W.A. (2003). Association analysis of a SNP polymorphism within the *3'UTR* of the *STAT6* gene in familial Crohn's disease. *Gastroenterology,* **124**(4): A374.

Zheng, H., Ji, C., Li, J., Jiang, H., Ren, M., Lu, Q., . . . Xie, Y. (2004). Cloning and Analysis of Human Apg16L. *DNA Sequence - The Journal of Sequencing and Mapping,* **15**(4): 303-305. doi: 10.1080/10425170400004104

Zouali, H., Lesage, S., Merlin, F., J-P Cézard and the EPWG-IBD group, J-F Colombel and the EPIMAD group, J Belaiche and the GETAID group, . . . Hugot, J.-P. (2003). CARD4/NOD1 is not involved in inflammatory bowel disease. *Gut,* **52**: 71-74.

Zouiten-Mekki, L., Kharrat, M., Karoui, S., Serghimi, M., Fekih, M., Matri, S., . . . Chaabouni, H. (2009). Tolllike receptor 4 (TLR4) polymorphisms in Tunisian patients with Crohn's disease: genotype-phenotype correlation. *BMC Gastroenterology,* **9**(1): 62. doi: 10.1186/1471-230x-9-62