

ABSTRACT

Crohn's disease (CD) is one of the major type of inflammatory bowel diseases (IBD) that affects the gastrointestinal tract. There were a lot of susceptible genes being published based on different population studies as well as meta-analysis projects. In this study, we aimed to investigate the association of some genes, *NOD1*, *CXCL16*, *STAT6*, *TLR4*, *ATG16L1* and *IRGM* had been found with association to the disease on non-Malaysian population. Overall, 85 CD DNA samples and 250 control healthy samples were collected from University Malaya Medical Center (UMMC). DNA were extracted and analyzed by polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) or TaqMan assay amplification. Results obtained showed that *NOD1* ($P_{G/A}=0.5179$, $OR_{G/A}=1.135$, 95% $CI_{G/A}=0.786 - 1.637$), *STAT6* ($P_{A/G}=0.531$, $OR_{A/G}=0.884$, 95% $CI_{A/G}=0.621 - 1.258$) and *IRGM* variants ($P_{rs4958847}=0.154$, $P_{rs11747270}=0.09359$, $P_{rs72553867}=0.8844$) had no association with CD. *CXCL16* ($P_{C/T}=0.04819$, $OR_{C/T}=1.431$, 95% $CI_{C/T}=1.009 - 2.031$), *TLR4* ($P_{C/T}=0.007$, $OR_{C/T}=0.437$, 95% $CI_{C/T}=0.242 - 0.789$) and *ATG16L1* variant rs2241880 ($P_{A/G}=0.01462$, $OR_{A/G}=0.645$, 95% $CI_{A/G}=0.454 - 0.917$) were found to be associated with Malaysian patients. For stratified analysis, *TLR4* ($P=0.0002399$) and *ATG16L1* rs2241880 variant ($P=0.03619$) were significantly associated with Malay while *ATG16L1* variant 6754677 ($P=6.3 \times 10^{-3}$) was associated with Indian patients. In conclusion, *CXCL16*, *TLR4* and *ATG16L1* but not *NOD1*, *STAT6* and *IRGM* were found to have strong association with CD in the Malaysian population.

ABSTRAK

Penyakit Crohn (CD) merupakan salah satu kelas daripada penyakit debu inflamatori (IBD) yang menjangkiti seluruh trek gastrousus. Pelbagai gen telah ditemui sebagai faktor penyakit CD melalui pelbagai kajian yang dijalankan atas populasi tertentu ataupun kajian meta-analisis. Dalam kajian ini beberapa gen telah dipilih iaitu *NOD1*, *CXCL16*, *STAT6*, *TLR4*, *ATG16L1* dan *IRGM* telah diistihar berhubung dengan penyakit dalam populasi asing. Varian yang terkandung dalam gen-gen diperiksa dengan 85 sampel pesakit CD dan 250 sampel kawalan yang dikumpul dari Pusat Perubatan Universiti Malaya (PPUM). DNA diekstrak sebelum dianalisis melalui reaksi rantai polimerase- polimorfisme panjang berkas restriksi (PCR-RFLP) atau amplifikasi asai TaqMan. Hasil menunjukkan tiada hubungan jelas untuk *NOD1* ($P_{G/A}=0.5179$, $OR_{G/A}=1.135$, 95% $CI_{G/A}=0.786 - 1.637$), *STAT6* ($P_{A/G}=0.531$, $OR_{A/G}=0.884$, 95% $CI_{A/G}=0.621 - 1.258$) dan *IRGM* ($P_{rs4958847}=0.154$, $P_{rs11747270}=0.09359$, $P_{rs72553867}=0.8844$) dengan CD. Hasil daripada *CXCL16* ($P_{C/T}=0.04819$, $OR_{C/T}=1.431$, 95% $CI_{C/T}=1.009 - 2.031$), *TLR4* ($P_{C/T}=0.007$, $OR_{C/T}=0.437$, 95% $CI_{C/T}=0.242 - 0.789$) dan varian *ATG16L1* rs2241880 ($P_{A/G}=0.01462$, $OR_{A/G}=0.645$, 95% $CI_{A/G}=0.454 - 0.917$) didapati berkait dengan CD di Malaysia. Dalam analisis berasaskan kaum, *TLR4* ($P=0.0002399$) dan varian *ATG16L1* rs2241880 ($P=0.03619$) didapati berkait rapat dengan kaum Melayu manakala varian *ATG16L1* 6754677 ($P=6.3 \times 10^{-3}$) adalah dengan kaum India. Sebagai kesimpulan, *CXCL16*, *TLR4* dan *ATG16L1* didapati berhubung dengan CD di Malaysia.

ACKNOWLEDGEMENT

At first I would like to express my sincere gratitude to my supervisors, Prof. Dr. Chua Kek Heng and Assoc. Prof. Dr. Ng Ching Ching for their continuous support throughout this research study. Their continuous guidance helped me a lot during the progress of research and writing of thesis. Without their precious advices I would not be to have a deep understanding on the subject of this research and I could not have imagined to have such better supervisors in my master study.

Secondly, special thank is dedicated to Mr. Kee Boon Pin for his technical guidance which helped me to become skilled on equipment operation. I would also like to thank to my fellow labmates for their helpful opinions while having discussions.

Last but not least, I would like to thank my family for being supportive and patient to me at the first place throughout my life.

CONTENTS

	Page
Title Page	i
Abstract	ii
Acknowledgements	iv
Contents	v
List of Figures	ix
List of Tables	xi
List of Symbol and Abbreviations	xv
List of Appendices	xix
Chapter 1: Introduction	1
1.1 Inflammatory Bowel Disease (IBD)	2
1.2 Crohn's Disease (CD)	4
Chapter 2: Literature Review	8
2.1 Nucleotide-binding Oligomerization Domain-containing 1 (<i>NOD1</i>)	9
2.2 Chemokine (C-X-C motif) Ligand 16 (<i>CXCL16</i>)	10
2.3 Signal Transducer and Activator of Transcription 6 (<i>STAT6</i>)	12
2.4 Toll-Like Receptor 4 (<i>TLR4</i>)	13
2.5 Autophagy Related 16-Like 1 (<i>ATG16L1</i>)	14
2.6 Immunity-Related GTPase family M (<i>IRGM</i>)	15

2.7	Objectives	17
Chapter 3: Materials and Methods		18
3.1	Materials	19
3.1.1	Agarose Gel	19
3.1.2	1X Tris-Borate-EDTA (TBE) buffer	19
3.2	Methods	19
3.2.1	Blood samples	19
3.2.2	DNA Extraction	20
3.2.3	DNA Quantification	21
3.2.4	Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP)	21
3.2.4.1	Polymerase Chain Reaction (PCR) Optimization	21
3.2.4.2	Polymerase Chain Reaction (PCR) Amplification	22
3.2.4.3	PCR Product Agarose Gel Electrophoresis	23
3.2.4.4	Restriction Enzyme (RE) Digestion	23
3.2.4.5	Agarose Gel Electrophoresis of Restriction Fragments	24
3.2.4.6	PCR Product Sequencing	25
3.2.5	TaqMan SNP Genotyping Assay	27
3.2.5.1	TaqMan Assay Amplification	27
3.2.6	Statistical Analysis	29
3.2.6.1	Fisher Exact Test	29

3.2.6.2	Odd Ratio (OR)	30
3.2.6.3	Confidence Interval (CI)	31
3.2.6.4	Hardy-Weinberg Equilibrium	31
3.2.6.5	Genotype Association Test	32
Chapter 4:	Results	35
4.1	Hardy-Weinberg equilibrium analysis	36
4.2	Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP)	36
4.2.1	Analysis of <i>NOD1</i> variant (rs2075820)	36
4.2.2	Analysis of <i>CXCL16</i> variant (rs2277680)	42
4.2.3	Analysis of <i>STAT6</i> variant (rs324015)	49
4.2.4	Analysis of <i>TLR4</i> variant (rs4986791)	54
4.3	TaqMan Assay Amplification	60
4.3.1	Analysis of <i>ATG16L1</i> variants	60
4.3.1.1	Analysis of rs2241880	60
4.3.1.2	Analysis of rs6758317	64
4.3.1.3	Analysis of rs6754677	66
4.3.2	Analysis of <i>IRGM</i> variants	69
4.3.2.1	Analysis of rs4958847	69
4.3.2.2	Analysis of rs11747270	72
4.3.2.3	Analysis of rs72553867	74

Chapter 5: Discussion	76
5.1 Analysis of <i>NOD1</i> gene variant	77
5.2 Analysis of <i>CXCL16</i> gene variant	78
5.3 Analysis of <i>STAT6</i> gene variant	80
5.4 Analysis of <i>TLR4</i> gene variant	82
5.5 Analysis of <i>ATG16L1</i> gene variants	84
5.6 Analysis of <i>IRGM</i> gene variants	89
5.7 Sample Quality	92
5.8 Limitations	93
5.9 Recommendations	94
Chapter 6: Conclusion	95
Appendices	97
Reference	117

LIST OF FIGURES

	Page
Figure 3-1 Overview of study methodology	34
Figure 4-1 Gel electrophoresis of digested <i>NOD1</i> rs2075820 PCR amplicon	37
Figure 4-2 Sequencing result for <i>NOD1</i> rs2075820 heterozygous genotype	38
Figure 4-3 Wild type PCR amplicon sequence of <i>NOD1</i> rs2075820	39
Figure 4-4 Gel electrophoresis of digested <i>CXCL16</i> rs2277680 PCR amplicon	43
Figure 4-5 Sequencing result for <i>CXCL16</i> rs2277680 heterozygous genotype	44
Figure 4-6 Wild type PCR amplicon sequence of <i>CXCL16</i> rs2277680	45
Figure 4-7 Gel electrophoresis of digested <i>STAT6</i> rs324015 PCR amplicon	49
Figure 4-8 Sequencing result for <i>STAT6</i> rs324015 heterozygous genotype	50
Figure 4-9 Mutant type PCR amplicon sequence of <i>STAT6</i> rs324015	51
Figure 4-10 Gel electrophoresis of digested <i>TLR4</i> rs4986791 PCR amplicon	54

Figure 4-11	Sequencing result for <i>TLR4</i> rs4986791 heterozygous genotype	55
Figure 4-12	Mutant type PCR amplicon sequence of <i>TLR4</i> rs4986791	55
Figure 5-1	Location of rs2241880 (T300A) within WD repeat of <i>ATG16L1</i>	87

LIST OF TABLES

	Page
Table 1-1 Comparison between Ulcerative Colitis and Crohn's Disease	3
Table 3-1 Composition of CD and control samples	20
Table 3-2 PCR protocol applied in PCR reaction	21
Table 3-3 Single reaction PCR mix	22
Table 3-4 Primer pairs used in PCR reaction	22
Table 3-5 PCR settings applied in PCR reaction	23
Table 3-6 Single reaction mix used in restriction enzyme digestion	24
Table 3-7 Restriction enzyme used for each SNP and its recognition sequence	24
Table 3-8 Interpretation of details after enzyme digestion	25
Table 3-9 Primers designed for PCR product sequencing	26
Table 3-10 PCR setting applied in sequencing primers	26
Table 3-11 Single reaction mix prepared for TaqMan amplification	27
Table 3-12 PCR protocol used in TaqMan amplification	28
Table 3-13 Bases assigned to different reporter dye	28
Table 4-1 <i>P</i> -value, OR And 95% CI obtained for <i>NOD1</i> rs2075820	40
Table 4-2 Data and statistical analysis of <i>NOD1</i> rs2075820 for different races	41

Table 4-3	<i>P</i> -value, OR And 95% CI obtained for <i>CXCL16</i> rs2277680	46
Table 4-4	Data and statistical analysis of <i>CXCL16</i> rs2277680 for different races	47
Table 4-5	Genotype association analysis (additive model) for Malaysian <i>CXCL16</i> rs2277680	48
Table 4-6	<i>P</i> -value, OR And 95% CI obtained for <i>STAT6</i> (rs324015)	52
Table 4-7	Data and statistical analysis of <i>STAT6</i> rs324015 for different races	53
Table 4-8	<i>P</i> -value, OR And 95% CI obtained for <i>TLR4</i> rs4986791	56
Table 4-9	Hardy-Weinberg equilibrium of control samples in <i>TLR4</i> rs4986791	56
Table 4-10	Data and statistical analysis of <i>TLR4</i> rs4986791 for different races	58
Table 4-11	Genotype association analysis (dominant model) for Malay <i>TLR4</i> rs4986791	59
Table 4-12	<i>P</i> -value, OR And 95% CI obtained for <i>ATG16L1</i> rs2241880	60
Table 4-13	Data and statistical analysis of <i>ATG16L1</i> rs2241880 for different races	62

Table 4-14	Genotype association analysis (dominant and additive model) for Malaysian <i>ATG16L1</i> rs2241880	63
Table 4-15	Genotype association analysis (additive model) for Malay <i>ATG16L1</i> rs2241880	63
Table 4-16	<i>P</i> -value, OR And 95% CI obtained for <i>ATG16L1</i> rs6758317	64
Table 4-17	Data and statistical analysis of <i>ATG16L1</i> rs6758317 for different races	65
Table 4-18	<i>P</i> -value, OR And 95% CI obtained for <i>ATG16L1</i> rs6754677	66
Table 4-19	Data and statistical analysis of <i>ATG16L1</i> rs6754677 for different races	67
Table 4-20	Genotype association analysis (dominant, recessive and additive model) for Indian <i>ATG16L1</i> rs6754677	68
Table 4-21	<i>P</i> -value, OR And 95% CI obtained for <i>IRGM</i> rs4958847	69
Table 4-22	Hardy-Weinberg equilibrium of control samples in <i>IRGM</i> rs4958847	70
Table 4-23	Data and statistical analysis of <i>IRGM</i> rs4958847 for different races	71
Table 4-24	<i>P</i> -value, OR And 95% CI obtained for <i>IRGM</i> rs11747270	72

Table 4-25	Data and statistical analysis of <i>IRGM</i> rs11747270 for different races	73
Table 4-26	<i>P</i> -value, OR And 95% CI obtained for <i>IRGM</i> rs72553867	74
Table 4-27	Data and statistical analysis of <i>IRGM</i> rs72553867 for different races	75
Table 5-1	Risk allele frequency, <i>P</i> -value and OR value for various populations of <i>ATG16L1</i> (rs2241880)	86
Table 5-2	Risk allele frequency, <i>P</i> -value and OR value for various populations in <i>IRGM</i> variants	90

LIST OF SYMBOLS AND ABBREVIATIONS

3'UTR	3'-untranslated region
ADAM	Disintegrin and metalloproteinase
AGE	Agarose gel electrophoresis
ATG16L1	Autophagy Related 16-Like 1
ATP	Adenosine triphosphate
bp	base pair
CARD	Caspase Activation and Recruitment Domain
CD	Crohn's disease
cDNA	Complementary deoxyribonucleic acid
CI	Confidence interval
CXCL4	Chemokine (C-X-C motif) Ligand 4
CXCL9	Chemokine (C-X-C motif) Ligand 9
CXCL12	Chemokine (C-X-C motif) Ligand 12
CXCL16	Chemokine (C-X-C motif) Ligand 16
CXCR6	CXC chemokine receptors 6
DNA	Deoxyribonucleic acid
DSS	dextran sulphate sodium
EtBr	Ethidium Bromide
GTP	Guanosine triphosphate

IBD	Inflammatory bowel disease
iE-DAP	γ -D-glutamyl-meso-diaminopimelic acid
IELs	Intraepithelial lymphocytes
IFN- γ	Interferon-gamma
IL-1 β	Interleukin-1-beta
IL-4	Interleukin-4
IL4-R	Interleukin-4-receptor
IL-8	Interleukin-8
IL-10	Interleukin-10
IL-13	Interleukin-13
IRGM	Immunity-Related GTPase family M
LPS	Lipopolysaccharides
LRR	Leucine-Rich Repeat
MDP	Muramyl dipeptide
MHC	Major histocompatibility complex
ml	milliliter
μ l	microliter
NACHT	NAIP (NLR family, apoptosis inhibitory protein), CIITA (class II, major histocompatibility complex, transactivator), HET-E and TP1(telomerase-associated proteins)

NBS	Nucleotide-Binding Site
NCBI	National Center for Biotechnology Information
NF- κ B	Nuclear Factor kappa-B
NKT	Natural killer T cells
NLR	NOD-like receptor
NOD1	Nucleotide-binding Oligomerization Domain-containing 1
NOD2	Nucleotide-binding Oligomerization Domain-containing 2
OD	Optical density
OR	Odd ratio
PCR	Polymerase Chain Reaction
RE	Restriction Enzyme
RFLP	Restriction Fragment Length Polymorphism
siRNA	Small interfering RNA
SNP	Single Nucleotide Polymorphism
STAT6	Signal Transducer and Activator of Transcription 6
TBE	Tri-Borate-EDTA
Th2	T helper cell 2
TLR	Toll-like receptor
TLR4	Toll-Like Receptor 4
TNF- α	Tumour Neurosis Factor-alpha

UC	Ulcerative colitis
UMMC	University Malaya Medical Centre
UV	Ultraviolet
WD	tryptophan (W) – aspartic acid (D)

LIST OF APPENDICES

	Page
Appendix A: Calculation in Fisher Exact Test	98
Appendix B: Calculation for Odd Ratio (OR) and 95% Confidence Interval (CI)	99
Appendix C: Calculation for Hardy-Weinberg Equilibrium	100
Appendix D: Data Distribution for Genotype Association Model Analysis	101
Appendix E: Electropherogram of Sequenced Homozygous Genotypes for <i>NOD1</i> (rs2075820)	102
Appendix F: Hardy-Weinberg Equilibrium Analysis for Malaysian Data by Arlequin 3.0	103
Appendix G: Hardy-Weinberg Equilibrium Analysis for Malay Control Cohort Data by Arlequin 3.0	107
Appendix H: Hardy-Weinberg Equilibrium Analysis for Chinese Control Cohort Data by Arlequin 3.0	110
Appendix I: Hardy-Weinberg Equilibrium Analysis for Indian Control Cohort Data by Arlequin 3.0	113
Appendix J: Genotype association analysis for <i>ATG16L1</i> (rs2241880)	116