AN INVESTIGATION OF THE GENETIC DETERMINANTS OF SUSCEPTIBILITY AND DISEASE BEHAVIOUR IN EARLY ONSET INFLAMMATORY BOWEL DISEASE IN SCOTTISH CHILDREN

Richard Kay Russell

PhD thesis presented to The University of Edinburgh – 2007

Contents

Table of figures

Table of tables

Dedication

This thesis is dedicated to my family Jacqui, Stephanie and Michael, as well as my parents who have all supported me and given me the strength to complete this thesis. This thesis is also dedicated to the memory of the late Jean Russell.

List of Publications arising from this thesis

Permission from publishers and co-authors has been obtained for publications included in this thesis. (All papers included on CD attached to thesis)

Articles

1. Russell R K, Wilson D C, Satsangi J. Unravelling the complex genetics of inflammatory bowel disease. Archives of Disease in Childhood 2004; 89(7):598-603.

2. Russell R K, Satsangi J. IBD: A family affair. Best practice & Research Clinical Gastroenterology. 2004; 18(3):525-539.

3. Russell R K, Nimmo E R, Satsangi J. Molecular genetics of Crohn's disease. Current Opinion in Genetics & Development.2004; 14(3):264-270.

4. Russell RK, Wilson D C, Satsangi J. CARD 15 / NOD 2 in Paediatric Onset Crohn's Disease. Journal of Pediatric Gastroenterology & Nutrition 2004; 39(2)216.

5. Russell RK, Wilson D C, Satsangi J. More Paediatric NOD2/CARD 15 data. American Journal of Gastroenterology 2004; 99(7):1401.

6. Russell RK, Fardi R, Drummond HE, Wilson M, Wilson D C, Satsangi J. Perinatal passive smoke exposure may be more important than childhood exposure in the risk of developing childhood IBD Gut 2005;54(10):1500-1.

7. Russell RK, Drummond HE, Nimmo ER, Anderson N, Smith L, Wilson DC, Gillett PM, McGrogan P, Hassan K, Weaver LT, Bisset M, Mahdi G, Satsangi J. Genotype-phenotype analysis in childhood-onset Crohn's disease: NOD2/CARD15 variants consistently predict phenotypic characteristics of severe disease. Inflammatory Bowel Diseases 2005; 11(11):955-964.

8. Gaya DR, Russell RK, Nimmo ER, Satsangi J. New genes in inflammatory bowel disease: lessons for complex diseases? The Lancet 2006; 367: 1271-84.

9. Satsangi J, Noble C, Arnott IDR, Russell RK, Nimmo ER.DLG5 variants and susceptibility to inflammatory bowel disease in the Scottish population. Gut 2006; 55(7):1050.

10. Törkvist L, Noble CL, Lördal M , Sjöqvist U, Lindforss U, Nimmo ER, Russell RK, Löfberg R, Satsangi J. Contribution Of Card15 Variants In Determining Susceptibility To Crohn's Disease In Sweden. Scandinavian Journal of Gastroenterology 2006; 41(6):700-5.

11. Noble CL, Nimmo ER, Gaya DR, Russell RK, Satsangi J. Novel Susceptibility genes in Inflammatory Bowel Disease. World Journal of Gastroenterology 2006; 12(13):1191-1199.

12. Russell RK, Drummond HE, Nimmo ER, Anderson N, Noble CL, Wilson DC, Gillett PM, McGrogan P, Hassan K, Weaver LT, Bisset WM, Mahdi G, Satsangi J . Analysis of the influence of OCTN1/2 variants within the IBD5 locus on disease susceptibility and growth indices in early onset inflammatory bowel disease. Gut 2006; 55(8):1114-1123.

13. Törkvist L, Noble CL, Lördal M , Sjöqvist U, Lindforss U, Nimmo ER, Russell RK, Löfberg R, Satsangi J. The Contribution of the IBD5 Locus to Crohn's Disease in the Swedish Population. Scandinavian Journal of Gastroenterology 2007; 42(2):200-206.

14. Van Limbergen J, Russell RK, Nimmo ER, Ho GT, Arnott IDR, Wilson DC, Satsangi J. Genetics of the innate immune response in inflammatory bowel disease. Inflammatory Bowel Diseases 2007; 13(3):338-55.

15. Russell RK, Drummond HE, Nimmo ER, Anderson N, Wilson DC, Gillett PM, McGrogan P, Hassan K, Weaver LT, Bisset M, Mahdi G, Satsangi J.

The contribution of the DLG5 113A variant in early-onset inflammatory bowel disease. Journal of Pediatrics 2007; 150(3):268-73.

List of published abstracts

1. Russell RK, Drummond H, Smith L, Anderson N, Nimmo ER, Wilson DC, Gillett PM, McGrogan P, Hassan K, Weaver LT, Bisset M, Mahdi G, Satsangi J. Do NOD2/CARD15 variants explain the high prevalence of early-onset Crohn's disease in Scotland? Alimentary Pharmacology and Therapeutics 2005; 21(2):200

2. Russell RK, Fardi RV, Drummond H, Wilson M, Satsangi J, Wilson DC. Parental smoking during pregnancy and an atopic background predispose to paediatric inflammatory bowel disease. Gut 2005; 54 (Supplement 2):A2

3. Russell RK, Drummond H, Smith L, Anderson N, Nimmo ER, Wilson DC, Gillett PM, McGrogan P, Hassan K, Weaver LT, Bisset M, Mahdi G, Satsangi J. Do Nod2/Card15 Mutations Explain The High Prevalence Of Early Onset Inflammatory Bowel Disease? Archives of Disease in Childhood 2005; 90(Supplement 2): A15

4. Russell RK, Drummond H, Smith L, Anderson N, Nimmo ER, Wilson DC, Gillett PM, McGrogan P, Hassan K, Weaver LT, Bisset M, Mahdi G, Satsangi J. OCTN 1/2 variants are associated with disease susceptibility and phenotype in early onset Inflammatory Bowel Disease (IBD). Gastroenterology 2005; 128(4): A-38.

5. Russell RK, Drummond H, Nimmo ER, Wilson DC, McGrogan P, Hassan K, Mahdi G, Satsangi J. The DLG5-113a Mutation Is Associated With Susceptibility To Early Onset Inflammatory Bowel Disease And Demonstrates A Complex Genotype Phenotype Relationship. Journal of Pediatric Gastroenterology & Nutrition 2005; 40(5): 641-42.

6. Russell RK, Ip B, Aldhous MC, Drummond HE, Arnott IDR, McDougall M, Wilson DC, Gillett PM, McGrogan P, Hassan K, Weaver LT, Bisset M,

Mahdi G, Satsangi J. ASCA positivity independently predicts Oral Crohn's disease in a large early onset IBD population. Gut 2006; 55(suppl_2):A35.

7. Russell RK, Nimmo ER, Wilson DC, Gillett PM, Bisset WM, McGrogan P, Weaver LT, Satsangi J. Sequencing Of The Entire NOD2/CARD15 Gene: Analysis Of The Contribution Of Potential Disease Causing Mutations In More Than 500 Crohn's Disease (CD) Patients. Journal of Pediatric Gastroenterology & Nutrition 2006; 42:E13.

8. Russell RK, Van Limbergen JE et al. Disease phenotype but not genotype differs in children and adults with IBD at diagnosis: Analysis of more than 1500 patients from a homogenous patient from a homogenous patient cohort. Journal of Pediatric Gastroenterology & Nutrition 2007; 44(6):e17.

Declaration of Originality

I declare that all the work in this thesis is unless otherwise indicated is entirely my own performed in the gastrointestinal laboratory between August 2002-2006. This work has not been submitted for any other professional degree or professional qualification.

The experimental work and subsequent analysis was all carried out by Richard Russell with the exception of:

1. Additional help with performance of the ASCA ELISA determination and analysis was given by Brian Ip.
2. Dr Niall Anderson gave assistance with the Transmission disequilibrium testing analysis for NOD2/CARD15, IBD5 and DLG 113 and the Log-Linear model analysis for IBD 5.
3. The TaqMan genotyping was performed by the staff of the Wellcome trust clinical research facility, Western General Hospital, Edinburgh.
4. DNA sequencing was performed in the MRC Human Genetics Unit, University of Edinburgh.
5. The genotype data from the Scottish adult IBD population and healthy controls was performed by other members of the research team. Specifically for these populations, Dr Ian Arnott generated the NOD2/CARD15 data, Dr Colin Noble the IBD5 and DLG5 data and Dr Lucy Walker provided the ASCA data for healthy adult controls.

This work was funded by a University of Edinburgh Medical Faculty Fellowship and additional funding for the project was received as part of a programme grant from the Wellcome Trust (072789/Z/03/Z).

Acknowledgements

This thesis has only been possible because of the help of a large number of people who have all given unconditional support and encouragement and to whom I am indebted.

I would primarily like to acknowledge the help and support given to me by both of my supervisors Jack Satsangi and David Wilson. Jack has opened my mind to endless stream of ideas and given me the chance to share in his unrivalled body of knowledge and experience in the world of Inflammatory Bowel Disease genetics. David has supported me emotionally and intellectually over the past 8 years before, during and after the thesis. I thank both my supervisors also for their friendship and extreme patience.

Peter Gillett has bent over backwards to help me in anyway possible and has offered tremendous personal support and encouragement throughout which has been of great help.

I would like to thank all of the members of the Scottish Paediatric Gastroenterology and Nutrition Group: Paraic McGrogan, Kamal Hassan, Lawrence Weaver, Simon Ling, Michael Bisset and Gamal Mahdi for 100% personal support and enrolment of patients into the study.

From the gastrointestinal unit at the Western General Hospital special mention goes to Hazel Drummond for extreme patience and advice in all matters relating to databases; Elaine Nimmo for patience in the lab and always having an answer when things seem to be going wrong; Linda Smith for patient collection, Norman Anderson for teaching me the tricks of the trade and Colin Noble, Charlie Lees and Gwo-Tzer Ho for sharing their experience and offering advice from the front line. I also thank Johan Van Limbergen for believing in the work done so far and for taking the research onto the next stage. I would also like to acknowledge the help of all patients and parents who participated in the study together with the specialist nurses, dieticians and secretaries in each of the teaching hospitals as well as the paediatricians, practice nurses and GP's throughout Scotland whose support for the study was invaluable.

Abbreviations

ANCA :	Anti-Neutrophil Cytoplasmic Antibodies
ASCA:	Anti-Saccharomyces Cerevisiae Antibodies
BMI:	Body mass index
CARD:	Caspase Activating Recruitment Domain
CD:	Crohn's disease
CDAI:	Crohn's disease activity index
CRP:	C- reactive protein
DLG5:	Discs, large homolog 5 (Drosophila)
DepCat:	Deprivation category
DNA:	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide triphosphate
EDTA:	Ethylenediaminetetraacetic acid
ESR:	Erythrocyte sedimentation rate
HC:	Healthy Controls
IBD:	Inflammatory Bowel Disease
IC:	Indeterminate Colitis
IQR:	Inter Quartile Range
LD:	Linkage disequilibrium
MDP:	Muramyl Dipeptide
mM:	Millimolar
NF-κβ:	Nuclear factor kappa-beta
NOD:	Nucleotide Oligomerisation Domain
OCTN:	Organic Cation Transporter
OR:	Odds Ratio

PCR:	Polymerase chain reaction
RFLP:	Restriction Fragment Length Polymorphism
RR:	Relative risk
TDT:	Transmission disequilibrium testing
TE:	10mM Tris pH 7.6
SNP:	Single Nucleotide Polymorphism
UC:	Ulcerative colitis
UK:	United Kingdom
WTCRF:	Welcome Trust Clinical Research Facility

Abstract

A series of investigations examining the importance of genetic factors in the development of the inflammatory bowel diseases (IBD) namely Crohn's disease (CD), Ulcerative Colitis (UC) and Indeterminate Colitis (IC) has been undertaken in Scottish children. This has been performed by collection of clinical details and DNA from children with IBD, then analysing the contribution of various candidate genes to both disease susceptibility and disease phenotype. In order to carry out these studies the presenting features of a large cohort of children from across Scotland with IBD diagnosed at less than 16 years were collected, both by examination of hospital case records and by patient interview and questionnaire. For each patient a detailed analysis was made of disease phenotype at presentation including detailed examination of disease location, disease behaviour and growth parameters.

A repository of clinical material (DNA, plasma and lymphocytes) was collected from children to accompany the detailed clinical parameters allowing genotype-phenotype analysis at a later stage. Additionally, DNA was also collected from parents where possible to facilitate family based association analysis of candidate genes by transmission disequilibrium testing. A previous DNA repository of healthy Scottish controls had been collected previously and the data generated was available for use in this study. The phenotypic data was collected using an established phenotypic classification (the Vienna classification) used in adult studies as well as a personally devised paediatric phenotypic classification designed for use in this thesis.

Firstly, the contribution of the three common mutations within the NOD2/CARD15 gene (R702W, G908R and Leu1007finsC) was analysed in 247 children with IBD. The Leu1007finsC variant was associated with Crohn's disease by case-control ($p = 0.01$) and TDT analysis ($p = 0.006$). Genotype phenotype analysis demonstrated NOD2/CARD15 variants were strongly associated with several markers of disease severity in CD most notably with an increased need for surgery on multifactorial analysis. Then to examine the further contribution of other mutations within the whole NOD2/CARD15 gene, the 12 exons of the gene were sequenced in 24 paediatric CD patients, to identify any additional SNPs that may have conferred an increased susceptibility to CD. Two mutations (V955I, M863V) identified in

sequencing were genotyped in a large patient cohort, but were not found to confer increased disease susceptibility.

Next, the contribution of IBD5 locus was analysed in 299 children with IBD studying 5 SNPs, including mutations in the proposed candidate genes OCTN 1 and 2. Allele frequencies of OCTN1/2 variants were significantly higher in IBD/CD cases ($p<0.04$). The homozygous mutant OCTN1/2 haplotype was increased in IBD and UC patients ($p = 0.02$ for both) compared to healthy controls. OCTN1/2 variants however were not independent of the background IBD5 risk haplotype in conferring disease susceptibility. Genotype- phenotype analysis demonstrated association of the risk haplotype with both lower weight and body mass index centiles at diagnosis as analysed by multifactorial analysis.

The contribution of the 113 G/A mutation within the discs, large homolog 5 (DLG5) gene was examined in 296 children with IBD. TDT analysis demonstrated a significant association with IBD ($p<0.05$). Genotype-phenotype analysis demonstrated associations with higher social class, male sex and taller children. Finally, the Anti-Saccharomyces cerevisiae antibodies (ASCA) status of 301 IBD patients was determined. CD patients had a higher prevalence of ASCA antibodies compared to UC patients and healthy controls ($p<0.001$ for both). A positive ASCA antibody was more common in CD patients with markers of more severe disease and on multifactorial analysis in patients with CD involvement of the oral cavity ($p = 0.001$).

In summary, the candidate genes examined thus far in children with IBD in Scotland have demonstrated a relatively minor contribution to disease susceptibility but have been demonstrated to be associated with specific disease phenotypes in patients with Crohn's disease. The use of a novel paediatric phenotypic classification in this thesis has allowed description of these novel genotype-phenotype associations.

1 Introduction

1.1 Background

Crohn's disease (CD [MIM 266600]) , Ulcerative Colitis (UC[MIM 191390]) and Indeterminate Colitis (IC), the chronic inflammatory bowel diseases (IBD), are now common causes of chronic gastrointestinal disease in the developed world, with a reported prevalence rates of 1 in 250 in Northern European populations.[1] Moreover, whereas incidence rates in adults appeared to have stabilised in recent years in the United Kingdom, compelling data both from within Scotland, and elsewhere in the United Kingdom, suggest that the incidence of early onset cases in children continues to rise.[2-4]

IBD is a major cause of morbidity costing an estimated £720 million per year for patients in the United Kingdom.[5] The major burden of this disease impacts on children and young adults with a peak age of onset in the third decade of life. In childhood, IBD most commonly presents just before the start of the teenage years impacting on emotional and physical development affecting linear growth, education and future employment prospects. Treatment of the disease is often difficult and there is little evidence to base treatment on from appropriately designed trials in childhood with most of the "evidence" distilled from adult studies. An increased understanding of disease pathogenesis in childhood could lead to more appropriately designed treatment options not only for children but also for adults with IBD.

1.2 Incidence studies

This rising incidence of IBD has been demonstrated in many childhood populations across Europe and North America.[4;6-11] There has been a near 3 fold increased incidence in Scotland initially demonstrated in 2 retrospective studies,[2;6] and clarified in a UK wide prospective study that surveyed more than 700 new incident cases diagnosed at under 16 years of age over a 13 month period.[4] The incidence of IBD in this survey carried out by Sawczenko et al. between 1998 and 1999 in Scotland was found to be 6.5 cases per 100,000 population having increased from 2.6 per 100,000 in a previous study performed by Barton et al. in 1968. This Scottish incidence

compares to a UK average of 5.2 cases per 100,000 in 1998-99 study. The incidence figure in Scotland therefore is the highest in the United Kingdom and one of the highest rates worldwide.[11]

Interestingly across Scotland there is a significantly higher incidence of Crohn's disease in the North of Scotland compared to the South of Scotland. In a study of 580 incident IBD cases studied over the period 1981-95, Armitage et al. demonstrated that the incidence of Crohn's disease in the North of Scotland was significantly higher than in the South, 3.1 compared to 2.1 cases per 100,000 population respectively ($p<0.001$).[12] Additionally, the same study demonstrated the relative risk of developing IBD and CD analysed by Poisson regression analysis was significantly increased in areas of relative affluence, compared to areas of socio-economic deprivation ($p=0.02$ and $p=0.03$ for IBD and CD respectively).

IBD is particularly common in young people and has a major effect on quality of life with profound effects on growth, pubertal and sexual maturity, education and employment as well as psychosocial integration.[13] Growth impairment is a particular feature of Crohn's disease in children that has effects lasting into adult life with up to 35% of subjects showing evidence of permanent growth failure.[8]

1.3 Disease definitions

Crohn's disease is characterised by patchy transmural inflammation affecting any part of the gastrointestinal tract from mouth to anus and over time almost inevitably progresses, often leading to stricturing or fistulising complications.[14] UC is characterised by continuous mucosal inflammation that always involves the rectum extending to involve a variable extent of the colon proximally. In UK children presenting with IBD around 60% will have CD and 30 % will have a diagnosis of UC.[4] In around 10% of cases where IBD it is limited to the colon it proves impossible to distinguish between CD and UC, even after colectomy, and a diagnosis of IC is made. Extra-intestinal manifestations commonly involving the skin, joints, and the eyes occur in all forms of IBD.

1.4 Environmental studies

Epidemiological studies have suggested a role for many different environmental factors. A role for many environmental agents has been suggested to account for the rapid rise in the incidence of IBD previously described. Many researchers have noted the rapid rise in Crohn's disease and similar autoimmune diseases has paralleled a decrease in infectious diseases over the later half of the last century.[15] Attempts to identify infective agents involved in disease pathogenesis have therefore attracted a great deal of attention in recent years. This has lead investigators to speculate a link in the increasing incidence of IBD with domestic hygiene,[16] refrigeration[17] and vaccination[18] among many others. The specific role of *mycobacterium paratuberculosis* in the pathogenesis of Crohn's disease remains a subject of impassioned debate, but conclusive evidence for a clear role in disease aetiology is awaited.[19] The controversies regarding the Measles Mumps and Rubella (MMR) vaccine also been the subject of widespread media and medical interest. However, after a thorough review of the literature by means of a Cochrane review there seems to be little to base any role in disease aetiology for MMR on.[20] Robust evidence using meta-analyses data support a role for only two environmental in IBD pathogenesis, namely cigarette smoking and appendectomy.[21-23]

1.4.1 Smoking

Cigarette smoking has contrasting effects in CD and UC; in CD smoking increases disease susceptibility but conversely is protective against the development of UC.[24] As the average age at diagnosis of IBD in childhood is around 11 years it is not personal but passive smoking that is likely to be more relevant to disease pathogenesis in the childhood population. Two studies support a role for passive smoking during pregnancy and at birth as a risk factor for the development of CD in childhood[25;26] with no evidence seen in a third study.[18] Additionally further studies have demonstrated passive smoke exposure during childhood does not affect the risk of developing IBD in childhood,[18;25] but has been linked to an increased risk of CD and decreased risk of UC in adulthood.[27;28]

1.4.2 Appendectomy

The role of appendectomy and a reduced risk of Ulcerative Colitis has been replicated in many studies.[23;29;30] In the largest study, Andersson et al. examined more than 200,000 patients who had undergone appendectomy in Sweden between 1964-93.[29] This study demonstrated a reduced risk of subsequent ulcerative colitis for patients who had proven appendicitis but not for patients who had their appendix removed for other indications. The greatest effect was demonstrated for patients who had appendicitis and appendectomy under the age of 20 years.

The most compelling evidence for the involvement of microbial agents implicates the gut flora. Animal models of colitis – genetically engineered, chemically induced, or spontaneous – all require the presence of gut flora in order for IBD to become manifest.[31] These data complement clinical studies with a resolution of Crohn's disease with faecal diversion, and increasing evidence that antibiotic and probiotic therapy may attenuate disease.[32-34] Diet including breast-feeding, childhood deprivation, and passive smoking all require to be evaluated in large well designed prospective studies of early onset inflammatory bowel disease populations.

1.5 Disease heterogeneity

Other epidemiological data provide evidence for heterogeneity within IBD, not only by age of onset but there is also evidence of widespread disease heterogeneity in disease phenotype (disease location, need for surgery, disease behaviour and extraintestinal manifestations). The sub-classification of disease by phenotypic characteristics remains controversial,[35;36] with a notable lack of an agreed phenotypic classification within childhood IBD itself.[37]

1.5.1 Children vs. adult studies

This disease heterogeneity is perhaps best illustrated by examining the differences between disease presenting in childhood compared to adulthood. Crohn's disease in childhood is a male dominated disease manifest as a high proportion of disease in the

proximal gastrointestinal tract, contrasting with the adult onset female dominated disease with disease more commonly in the distal gastrointestinal tract (see chapter 3 for a more detailed discussion).[7;38] In ulcerative colitis, extensive colonic disease is the norm in childhood (80-90% in prospective studies) [39-41] but the exception at presentation in adult series (see chapter 3 for a more detailed discussion). [42;43]

Inflammatory bowel disease(s) can be therefore divided into subgroups dependant on disease location in the gastrointestinal tract and represents a number of closely related multifactorial diseases. The similarities and differences highlighted by epidemiological data could be explained by a combination of shared environmental and/or genetic factors.[44]

1.6 Evidence for genetic predisposition in inflammatory bowel disease

1.6.1 Ethnicity

Initial evidence for a genetic predisposition to the development of IBD stems from the observation of differences in the incidence of IBD between different ethnic groups. There is well documented evidence that Caucasians in Europe and North America have the highest rates of IBD followed by Afro- Americans and then Asians.[11] Among Caucasians the Ashkenazi Jewish population have the highest prevalence rates of both sporadic and familial disease.[45] The prevalence of IBD does however increase in a "low risk" population that then moves into an area of higher IBD prevalence demonstrating the key role of the environment. Probert et al. demonstrated this by studying the prevalence of UC in Asian migrants in the UK, located in Leicester. They demonstrated increased rates of UC in Asians compared with rates in their country of origin, and rates of UC in Hindu's and Sikhs that actually exceeded that of the local population in Leicester.[46]

1.6.2 Family studies

It is well recognised from many studies that the greatest risk factor for the development of IBD is having an affected family member,[47] with approximately 5-10% of all affected individuals with IBD reporting a positive family history: first

degree relatives, especially siblings are at greatest risk but the risk also extends to more distant relatives.[48] Quantification of this risk depends on a combination of ethnicity, the type of IBD - CD or UC and the exact relationship between the person at risk and the proband with IBD.

The increased risk to relatives is difficult to quantify with precision despite many published studies because of the selection bias inherent within many of these studies. In a multi-centre study of Italian IBD patients Meucci et al. demonstrated teaching hospital populations will tend to have over represented IBD patients with a positive family history compared to studies conducted from other hospitals.[49] Studies do not always report results in a similar manner, some will report data for all relatives or first degree relatives only and assign this risk to either risk of UC, CD and/or IBD. The studies therefore likely to give the most accurate figures for risk to relatives are population studies that incorporate all spectrums of disease severity. The length of follow up is also critically important as studies with a longer follow up period are more likely to include more family members with IBD.

1.6.3 First degree relatives

The first degree relatives of a patient with IBD have been studied both collectively and individually more than any other relatives.[48] Studies have mainly consisted of selected hospital populations and only very few have covered whole populations. In each study the risk of CD and UC has been considered separately for each relative and some studies have given a relative risk of CD/UC/IBD compared to the population that the study comes from. Table 1-1 highlights studies examining these risks in probands with Crohn's disease.

Table 1-1: Studies of first degree relatives in a proband with CD

Study Population	Number of patients	% of 1st degree CD relatives	% of 1st degree IBD relatives	RR of CD in 1st degree relative	RR of UC in 1st degree relative
Canadian[50]	1000	8.7%	not stated	not stated	not stated
Welsh[51]	139	5.0%	9.3%	x13	not stated
Swedish[52]	1048	6.9%	8.9%	x21	x6
USA[53]	522	15.1%	16.7%	not stated	not stated
Finnish[54]	257	10.9%	15.6%	not stated	not stated
UK[55]	433	not done	11.5%	x15 (for IBD)	not stated
USA[45]	258	7.4%	14.0%	x5-8* (for IBD)	not stated
UK[56]	424	9.4%	10.4%	x17-35	x3
USA[57]	554	12.2%	not stated	not stated	not stated
Belgian[58]	640	13.6%	14.5%	x12* (for IBD)	not stated
Danish[47]	133	2.2%	5.2%	x10*	x4*
Dutch[59]	400	8%	9.5%	not stated	not stated
USA[60]	80	16.2%	22.5%	not stated	not stated
French[61]	1316	7.5%	8.4%	not stated	not stated

*Age adjusted figures

RR = relative risk

The studies listed in table 1-1 are all of selected hospital populations except the Swedish study conducted by Monsen et al. which was an unselected population study.[52] Also, where studied it is clear the risk of developing both types of IBD is increased when the proband has CD including the risk of developing CD in first degree relatives which is higher than that of UC.[47;52;56] In studies examining UC probands similar findings are demonstrated with lower numbers of affected relatives in population based studies and increased risk of both types of IBD but with a greater risk of UC cf. CD.[45;47;53;54;56;60;62]

1.6.4 Age adjusted studies

Several studies listed have also used age adjusted figures for calculating IBD risk in first degree relatives.[45;47;58;60] The age adjusted relatives risks in all studies except Orholm's Danish study are based on calculations using Strömgren's methodology.[63] This is a risk assessment based on patients living to age 70 years. Combining the results of these studies together gives a lifetime risk for first degree relatives of a CD proband of 4.8- 5.2% for non-Jews and 7.8% for Jews of developing IBD.[45;58;60] The equivalent figures for first degree relatives with a UC proband are 1.6% for Non Jews and 5.2% for Jews.[45]

1.6.5 Sibling studies

Within family studies results from many studies have consistently shown siblings to be at the highest risk of developing IBD compared to other first degree relatives.[45;50;51;58;59;64;65] These studies clearly demonstrate that the risk of CD in a sibling of a CD proband is significantly higher than the average relative risk for other first degree relatives. Of all first degree relatives, offspring are the hardest to present accurate figures for as many may not have reached at age where they will manifest symptoms of IBD at the time of a particular study. This is reflected in the wide relative risks for a CD proband (2-30)[55;56;66] and for a UC proband (2-15).[56;66]

1.6.6 Parental studies

In the relatively unusual situation of a child being born to parents who both have CD it is clear there is a very high risk of the children developing IBD. Laharie et al. studied 30 IBD couples from Belgium and demonstrated an increasing probability of acquiring IBD with age such that by 28 years of age 33% of children will have developed IBD.[67] Bennett et al. studied 19 American IBD couples and demonstrated a similarly high risk in offspring from UC and "mixed" IBD couples having children.[68]

In the more common situation where one parent, rather than two, has IBD the risks to offspring are substantially less. As studies have shown in other first degree relatives, the child is at greatest risk of developing the same type of IBD as their parent but is also at increased risk of developing the opposite type of IBD. In a population study examining all patients discharged across Demark with IBD between 1977 and 1992 Orholm et al. demonstrated in offspring of UC parents that 6.2% developed IBD compared to 9.2% of offspring whose parents had Crohn's disease.[66]

Parents of children with IBD have been consistently found to be at the lowest risk of developing IBD compared to the other two groups of first degree relatives.[48] The relative risks for a parent of a child with CD and UC for developing IBD are similar 12-16 for CD and 8-19 for UC.[55;56]

1.6.7 Other relatives

The risk to second degree relatives (aunts, uncles, nieces, nephews and grandparents) has not been as extensively studied. The risk seems to be increased only for the same type of IBD as the proband with little or no increased risk of the opposite IBD phenotype as measured against prevalence rates in the general population.[47;69] Relative risks and age adjusted risks for this group have not been published. A third degree relative (cousin) is also at increased risk of IBD compared to the general population but the risk is considerably less than the risk to first degree relatives. In a recent population study from Iceland only third degree relatives of CD patients were shown to be at increased risk of disease but not UC relatives.[69]

1.6.8 Multiplex studies

In the situation where there is more than one affected first degree relative with IBD in the family the risk to other relatives increases e.g. Orholm et al. demonstrated in UC families where at least one first degree relative had UC the relative risk of developing UC was doubled compared to the situation where no first degree relative had UC.[47]

1.7 Twin studies

Twin studies have been a powerful tool used to identify the different contributions of genes and environment to complex polygenic diseases including IBD. Twins are in general brought up in the same environment so differences in disease prevalence between monozygotic and dizygotic pairs are used to estimate the relative contribution of genes and environment to a disease developing.

The early literature published on this subject contained many individual case reports of disease concordance in both CD and UC monozygotic twin pairs and there was even a case report of monozygotic triplets all with CD.[70] The true rate of twin concordance could only however be accurately estimated by large population studies.

1.7.1 Population studies

Tysk et al. published the first population survey of twins and concordance for IBD in 1988.[71] The cases were selected from the Swedish twin registry and were then matched with hospital inpatients records. They identified 80 twin pairs with IBD. The twins were diagnosed with IBD at an average age of 30 years old, with a follow up period of around 20 years. This original cohort was followed up 15 years later and 3 further cases of IBD were diagnosed.[72] All of the "new" cases had been symptomatic at the time of the original survey but did not fulfil the diagnostic criteria. In the monozygotic CD concordant pairs these authors were also able to demonstrate that 77% of pairs were concordant for disease location and that 67% of pairs were concordant for age at diagnosis within 2 years of each other.

A second population study was performed in Denmark that produced similar results to the Swedish study.[73] Orholm et al. studied 103 twin pairs again from a national registry using a postal questionnaire to identify affected twin pairs. The twins were diagnosed at a mean age of 23 with IBD with an average follow up of 8 years. In CD concordant pairs 2/5 were concordant for disease location and 2/5 were diagnosed within 2 years of each other. The authors were also able to calculate relative risks in this study: a monozygotic co-twin of a CD proband had a relative risk for developing CD of 667 and a UC proband had a relative risk of 71 for developing UC.

The third large published study was undertaken in the UK by Thompson et al. and examined 143 twin pairs.[74] In this study, patients from an IBD support group, the National Association for Colitis and Crohn's disease (NACC) were sent a questionnaire to identify IBD twin pairs. The selective nature of this study could potentially be a source of bias. In contrast to the other two studies the results showed similar concordance rates for CD and UC in monozygotic twin pairs at 20% and 16% respectively. The concordance rate for CD is much lower than that of the two population studies. In a later extension of this study population, a total of 249 twin pairs were studied. The extension study did show similar concordance rates to the two Scandinavian population studies.[75] Interestingly in this follow up study the concordance rate in dizygotic twins with CD was double that of siblings but similar concordance rates for dizygotic twins and siblings were found in UC. No information was given on concordance rates for disease characteristics. A summary of published twin studies appears in table 1-2.

Table 1-2: Summary data from the 3 published IBD twin studies

Population	CD MZ	CD DZ	UC/MZ	UC/DZ	Total
Danish[73]	5/10	0/27	3/21	3/45	103
British[74]	5/25	3/46	6/38	1/34	143
Swedish[71]	9/18	1/26	3/16	0/20	80
European	19/53	4/99	12/75	4/99	326

MZ=monozygotic; DZ = dizygotic

1.7.2 Concordance rates

Combining the results of these studies makes the overall concordance rates for CD 36% and 4% (for monozygotic and dizygotic pairs respectively), with the corresponding figures for UC 16% and 4%. In a similar manner to family studies, IBD co-twins are also at increased risk of the opposite IBD phenotype to their twin. The finding of twin pairs one with CD and the other with UC has been described for one set of monozygotic twins and in at least 5 sets of dizygotic twins.[73;75]

The concordance rates in twin studies allow an estimate to be made of the relative contribution of genetics to overall disease aetiology, an estimate that is expressed as the coefficient of heritability. The derived co-efficient of heritability in Crohn's disease is equivalent to that reported for type 1 diabetes, multiple sclerosis and asthma.[76]

This evidence from twin studies therefore confirms a significant role for genetics in disease development in IBD with a stronger genetic influence in CD compared with UC. There is however evidence of shared genetic susceptibility with both diseases occurring in some sets of twin pairs.

1.8 Familial Phenotypes

The similarity or difference between family members with IBD has been compared in familial (anyone with another family member with IBD) against sporadic disease (no family history of IBD) and in multiplex family studies where many members of the same family group have IBD. There are high rates of concordance for disease type in family studies with concordance rates between 78-100% in CD[49;58;77;78] and 81-86% for UC.[49;62;77;78] The difficulty in extending these studies to compare disease characteristics is that there were no standard definitions of disease phenotype agreed upon when many of the studies were performed making direct comparisons between studies difficult. The adoption of guidelines for CD phenotyping had the advantage of setting a phenotypic benchmark allowing comparison between studies,[35;36] but unfortunately most of the reported studies were performed prior to these classifications being adopted.

The phenotype characteristics of CD that have been studied are: age at disease onset, location of disease, disease behaviour, extraintestinal manifestations and the need for surgery.

1.8.1 Age at disease onset

The age at disease onset/diagnosis has been compared between familial and sporadic disease and been found to significantly younger in familial disease in some studies[45;52;62;65;79] but an equal number of studies show no difference.[49;50;54;58;61;77] Polito et al. in a retrospective study of 552 patients with Crohn's disease demonstrated stratification for age at diagnosis (<20 years at diagnosis compared to > 40 years at diagnosis) demonstrated age specific phenotypes. Younger age at diagnosis was associated with more ileal disease, more stricturing disease together with a greater frequency of surgery for intractable disease.[79] Small bowel disease became less frequent with older age at disease diagnosis contrasting with colonic disease location and inflammatory disease behaviour that became more common with older age.

1.9 Evidence for genetic anticipation

Following on from studies suggesting a younger age at diagnosis in familial cases of IBD researchers including Polito et al. went onto explore the evidence for genetic anticipation in IBD. The term "anticipation" describes a genetic phenomenon in which disease severity increases and age of onset decreases in subsequent generations affected by the disease in question. Anticipation has been clearly demonstrated in a range of monogenic neurological disorders including Huntingdon's chorea, Friedrich's Ataxia, Fragile X syndrome and myotonic dystrophy.[80] The common mechanism unifying these diseases is expansion of nucleotide triplet repeats, with the more repeats that are present the earlier the disease presents and in turn the more severe the disease is. If anticipation was demonstrated in IBD populations this would giving an insight into the genetic mechanisms underpinning IBD.

Polito et al. in a follow on study examined 59 parent child parent IBD pairs and demonstrated an average difference of 15 years earlier diagnosis in the children compared to their parents, with 91% of children diagnosed at younger age than parents.[81] The children also showed more extensive disease than their parents in most cases. Numerous replication studies followed, all universally showing a younger age at symptom onset/ diagnosis in children with IBD compared to their parents (see table 1-3 for study details).

Table 1 3: Studies supporting evidence for genetic "anticipation" in IBD

Author	Type of IBD	Number of parent/child pairs	% of children diagnosed first	Age difference (years)
Polito[79]	CD	59	91%	15
Satsangi[64]	IBD	77	90%	16
Grandbastien[82]	CD	57	84%	16
Peeters[58]	CD	31	not stated	17
Heresbach[83]	CD	61	90%	14
Lee[84]	IBD	137	89%	18
Hampe[85]	IBD	99	not stated	19
Freeman[50]	CD	46	100%	not stated
Faybush[86]	IBD	136	76%	16
Annese[78]	IBD	36	93%	23

Heresbach et al. were also able to demonstrate "anticipation" in not only 61 child-parent pairs but also in 17 aunt/uncle nephew/niece pair's showing a mean difference of age at diagnosis of 14 and 16 years in favour of the younger generation respectively.[83] The age difference at diagnosis effect in their study was significantly more marked in father child pairs at 21 years compared with 12 years in mother child pairs.

However before this apparent "anticipation" in IBD is accepted, the studies describing this phenomenon merit further scrutiny. Firstly, from first principles IBD is not the same as other diseases where anticipation has been previously described. The diseases in which anticipation has been demonstrated are *monogenic diseases* where as IBD from all other evidence from family and twin studies are likely to be polygenic *diseases*. After Polito's initial publication describing anticipation two groups independently criticised the data. Inskip et al. and Frisch et al. both ascribed the apparent anticipation to an effect of follow up time bias based on a retrospective study design.[87;88] Follow up bias describes children who are unaffected by IBD at the time of a cross sectional survey but who later develop disease with failure to take this into account skewing the results toward apparent anticipation.

To analyse whether anticipation is truly present, IBD birth cohorts have been followed up to remove any effect generated by follow up time bias. Picco analysed 928 consecutive patients with Crohn's disease diagnosed at the Mayo clinic: in analysis unadjusted for follow up time bias, he demonstrated age of diagnosis decreased by 5 years in each successive 10 year birth cohorts giving preliminary evidence of anticipation.[89] However, by adjusting the analysis to only include patients who were 40 at the time of study entry and had a diagnosis before age 40 this effect was reduced to around 1 year difference in successive cohorts. They suggested anticipation does not occur in IBD and is merely caused by failure to make statistical corrections to the data.[89] Hampe et al. analysed 2007 IBD patients with sporadic disease and 472 patients with familial disease and after adjustment for confounders between parental and child diagnosis no significant difference in age at diagnosis was demonstrtaed.[85] In a similar study from St Marks hospital Lee et al. examined 137 parent-child pairs and compared them with 214 patients with sporadic

IBD.[90] Comparison between the two cohorts using regression lines based on date of birth and diagnosis demonstrated no difference between the two groups despite an apparent age difference between parents and children analysed of 17 years. The difference in these results can be explained by ascertainment bias rather than anticipation. Faybush et al. replicated the findings of previous studies for 136 parent child pairs and also for 142 aunt/uncle nephew/niece pairs.[86] Uniquely they were able to show a difference in the mean age of onset of symptoms to be doubled in 40 grandparents – grandchild pairs compared to parent child pairs. They concluded from this again this apparent anticipation was due to follow up time bias and possible temporal changes effecting many generations at once.

1.10 Disease concordance for disease phenotypes

There is evidence of concordance not only of disease type but also for disease location in within CD families. In a study of 35 Italian families Annese et al. demonstrated 46% concordance for disease location.[78] Colombel et al. demonstrated concordance for disease location in 56% of patients in a study of 72 French families.[65] Satsangi et al. demonstrated a high degree of concordance for disease location in both 39 parent-child pairs and 82 sibling pairs with concordance rates of 75% and 80% respectively in a UK study.[64] Peeters et al. replicated this high concordance for disease location in sibling pairs in a Belgian CD study.[58] In a study of 60 CD families from the USA concordance for disease location was seen in 86% of those studied.[57] Lower disease concordance rates of only 16% for disease location were described in a small study of 25 UK families.[84]

In studies that have examined other Crohn's disease characteristics varying rates of concordance have been demonstrated: disease behaviour 17-82%,[57;61;78;84] extraintestinal manifestations 67-80% and rates of surgery 29-56%.[64;78]

Satsangi et al. demonstrated concordance for disease extent in 53% of 37 parent child and 68% of 69 sibling pairs with Ulcerative colitis.[64] In the same study of ulcerative colitis relative pairs there were similarly high degrees of concordance for

extraintestinal manifestations at 74% and 89% and in need for surgery at 77% and 67% respectively.[64]

1.10.1 Early onset disease

In early onset disease there is a stronger familial and therefore perhaps genetic contribution to disease development. In studying early onset populations the power to detect a difference between cases and controls for a given sample size will therefore be increased.[91] In a study of families with type 2 diabetes, Frayling et al. demonstrated that diabetes diagnosed under the age of 55 years had a much stronger genetic influence compared to those diagnosed over the age of 55.[92] In this study an increased number of genetic loci were identified in families with an earlier age of onset despite there being a smaller total number of families studied (245 in the early onset group compared to 328 in the later onset).

In families with IBD, Farmer et al. studied 838 patients with onset of IBD under the age of 21 years and found a positive family history in 35% of cases, higher than that recorded in most other series (table 1-1).[53] Comparing familial with sporadic disease familial disease tends to be associated with an earlier age of onset compared to sporadic disease.[48] Monsen et al. demonstrated a younger age at diagnosis in familial CD compared to sporadic CD at ages 25 and 33 years respectively,[52] with similar data produced by the same author in a second study of patients with UC.[62] The same trends in the data reported by Monsen et al. have been described by several other groups.[45;65;79] There is some evidence to indicate that disease starting earlier in life has a stronger genetic influence in patients with IBD in studies examining the IBD1 locus.[93;94]

Additionally, there are obvious advantages for family association studies as parents are more likely to be available for recruitment if younger patients are studied. Using a family based association approach in turn avoids some of the problems of population stratification encountered in case-control studies and reduces risk of population heterogeneity when comparing different study populations.[95] Furthermore the childhood IBD population is often considered to be free of the strong influences of the environment present in adult patients e.g. tobacco.[21;22]

1.11 Genetic models of disease in IBD

From the evidence presented thus far it is clear that both genes and environment are important in the pathogenesis of IBD. There is clearly a genetic risk for both types of IBD but it appears CD carries a higher risk than UC. The data discussed for twin and family studies however are not explained by simple Mendelian genetics. The risk to first degree relatives of developing disease is between 5-10% arguing against either an autosomal dominant or autosomal recessive inheritance pattern in the majority of patients. Exceptions to this would appear to exist in the rare situations where both parents have IBD with rates approaching up to 50% in the offspring suggestive of a dominant inheritance model.[67;68] Orholm et al. formally tested a model of autosomal inheritance patterns in familial disease in 637 Danish patients and their relatives using complex segregation analysis.[96] The results generated suggested an autosomal recessive pattern of inheritance was present in only 7% of familial disease in Crohn's disease and proposed an autosomal dominant model was present in 10% of familial cases of Ulcerative Colitis. In response to the publication of Orholm's data, Farrall argued that the presence of increased risk of both CD and UC in relatives and the increased risk when more than one relative had IBD, strongly argued against a recessive model of inheritance.[97]

The data suggest therefore that a complex pattern of inheritance is present in IBD in common with other common complex diseases, with the rates of inheritance described similar to those in multiple sclerosis, type 1 diabetes and schizophrenia and much higher than rates of inheritance in the common malignancies.[98] Ahmad et al. have proposed a disease model for IBD that describes CD and UC as related polygenic diseases that may share common susceptibility genes but also have disease specific susceptibility genes.[76] In addition, genes are likely to influence not only disease susceptibility but also disease phenotype.

Nonetheless it seems likely that the phenotypic heterogeneity reflects extensive genetic heterogeneity. The term "Crohn's disease" can be divided into subgroups dependant on disease location in the gastrointestinal tract and represents a number of closely related multifactorial diseases, sharing some but not all genetic and environmental determinants of susceptibility and behaviour.[44]

1.12 Findings genes involved in IBD

At the outset of this thesis, two complementary methods were traditionally employed by investigators searching for susceptibility genes in complex diseases using either candidate gene analysis or genome wide analysis.[99] The genome wide analysis until recently involved non-parametric linkage analysis of informative SNPs in relative pairs with IBD. More recently genome wide association studies involving hundreds of thousands of SNPs in patients and healthy controls have been utilised to identify further IBD susceptibility genes (see chapter 9). Other methods such as yeast two-hybrid studies and microarray studies and have also been used to identify further genes important in IBD susceptibility (these are also discussed in detail in chapter 9).[100]

1.12.1 Candidate gene studies

The analysis of candidate genes relies on an understanding of disease pathophysiology. In inflammatory bowel disease for example, the immuno-pathology of the disease has led to examination of genes involved in the regulation of the immune system, genes involved in the maintenance of mucosal integrity, and genes involved in cell-cell interactions.[101] The genes in question have then been further analysed by genetic "association" studies where the frequency of allelic variants of one or more polymorphisms of the genes in question are studied in patients with inflammatory bowel disease and compared with allelic frequencies in a well-matched control populations.[95] A significant distortion of frequencies between cases and controls under comparison would provoke further investigation of the gene of interest. *Direct* association describes the situation where association studies implicate a polymorphism that is likely to result in an amino acid change, in contrast to *indirect* association where the polymorphism is not likely to result in an amino-acid change and may or may not be in linkage with the causal variant. Many candidate genes have been subject to analysis in inflammatory bowel disease – notably the genes of the HLA system, genes involved in the regulation of cytokine production, mucin synthesis, and other aspects of epithelial barrier function.[102]

Candidate gene studies however are limited by the fact they rely on an assumed knowledge about the pathogenesis of IBD. The selection of genes for study is based on an *a priori* hypothesis about disease causality. A good case can usually be made to study many of the 30,000 genes on the human genome so, in the context of a disease that is complex and polygenic, candidate gene analysis based merely on a "scientific hunch" is a rather naive and approach that will be rarely successful.

1.12.2 Genome wide scanning

However, in contrast to candidate gene analysis it is the complementary technique of genome wide scanning (GWS) in which success has become most apparent in narrowing down the number of candidate genes for analysis and providing new insights into disease pathogenesis.[103] In GWS, genetic (non-parametric) "linkage" in relative pairs is used to identify broad regions of the genome that predispose to disease susceptibility and may at a later stage be shown to contain a disease specific gene.[104] They therefore reduce the potential number of candidate genes but do not identify these genes specifically. This technique is superior to candidate gene studies alone as it allows researchers the advantage of being able to identify genes in a "hypothesis free" manner, rather than selecting candidate genes on the basis of preconceived scientific ideology. This technique has been successfully applied in IBD to identify genes important both in disease susceptibility and disease phenotype.[101]

The technique of GWS evolved with the development of a linkage map of the human genome, involving informative microsatellite markers providing a framework for the systematic analysis of the human genome in both single gene disorders, and complex diseases.[105] Studies in complex diseases have required access to large numbers of multiply affected families (typically sibling pairs), semi-automated technology for genotyping, and particularly the evolution of statistical techniques for data analysis produced after running the scan. The technique of GWS has been applied by investigators in many common disorders – encompassing metabolic, respiratory, cardiovascular, endocrine, and neuropsychiatric disease.[104]

The candidate genes studies performed to date have been prioritised on the basis of functional and pathobiological relevance of the genes in question, in addition to the results from these linkage studies. In each disorder, a number of linkages with regions throughout the genome have been described. However, proceeding from the initial observation of linkage through replication to gene identification had not been achieved by many groups of investigators. Progress in gene identification following on from GWS in IBD relative pairs has been greater than in many similar complex polygenic diseases.[101] Several chromosomal regions along the genome have been replicated, with sufficient strength to satisfy stringent criteria laid down by the statistical geneticist as demonstrated by significant logarithm of odds (LOD) scores (see table 1-4).[76]

Table 1-3: The Lander and Kruglyak criteria

Criteria*	**LOD score**	**p value**
Suggestive linkage	2.2	$7x10^{-4}$
Significant linkage	3.6	$2x10^{-5}$
Highly significant linkage	5.4	$3x10^{-7}$

*These criteria apply only to analysis of genome wide scans using sibling pair studies, other relative pairs will have different significance thresholds; for confirmed linkage to be achieved a replication study in a separate population from the initial cohort with a nominal p value of < 0.01 would be required.

1.12.3 IBD loci identified in genome wide scanning

The success of the GWS in CD, in common with other complex diseases has been related to the size of the population sample in the scan and to a lesser extent the number of microsatellite markers used. There have been eleven scans in IBD patients since 1996 and a much larger number of replication studies. These scans have predominantly been performed in CD patients and a summary of the GWS for these patients appears in table 1-5.

Table 1-4: Results of Genome wide scans in Crohn's disease

Study Population	CD relative pairs	Number of microsatellite markers	highest LOD score	Other suggestive loci
European[106]	110	270	1.3 (16q12)	nil
UK[107]	81	260	5.47(12q13)	3p21,7q22
European[108]	162	358	2.3(10)	16q, 12q
USA[109]	65	350	2.8 (14q11)	5q
USA[110]	175	377	2.17 (1p36)	3q,16q,3p26,16#
USA[111]	127	751	3.6(14q11)	18q22
Canadian[112]	131	312	3.9(5q31-33)	19p13*
Canadian[113]	105	307	1.9(11q)	16q#, 17p,11p
USA[114]	108	358	2.26(6p13)	5q31,15q
Belgian[115]	106	323	2.97(1p13)	6q16,14q11,4
UK[116]	137	404	2.1(3q)	X

* 3p and 6p achieved suggestive linkage in this study for IBD but the contribution of CD to the LOD score was not stated.

Linkage on chromosome 16 distinct from IBD1 region

Thus far, numerous susceptibility loci have been implicated in inflammatory bowel disease with varying degrees of replication and statistical support demonstrated for 9 of such loci, termed IBD 1-9.[117] Each of these loci represent chromosomal regions each of which contain many potential IBD susceptibility genes.[101] Whilst some loci appear specific to Crohn's disease (e.g. IBD1 on 16q- OMIM 266600)[118] or ulcerative colitis (e.g. IBD2 on 12q- OMIM 601458),[119] others seem to confer susceptibility to IBD overall (e.g. IBD3 on 6p- OMIM 604519)[120] supporting the previously discussed hypothesis that Crohn's disease and ulcerative colitis are related polygenic disorders sharing some, but not all, susceptibility loci.

1.12.3.1 IBD1

Hugot et al. used a two stage approach to demonstrate the first CD susceptibility locus on chromosome 16q12 subsequently named IBD1 (MIM 605956).[106] These investigators performed initial genotyping in 40 IBD sibling pairs from 25 French families using a total of 270 microsatellite marker SNPs situated at regular intervals throughout the genome. This first stage of the investigation identified significant linkage (p value of<0.01) for three areas on chromosome 16 and one area on chromosome 1. In the second stage of the investigation a separate group of 53 families containing 70 sibling pairs provided supportive evidence for linkage with the region on chromosome 16 only, with a LOD score of 1.3.

Despite this relatively low LOD score in the initial study, this locus was then successfully replicated in different populations throughout the world. Ohmen et al. examined the IBD 1 locus in a replication study using 17 microsatellite markers across the region of the locus. In this study of 80 North American families (48 CD families with 75 sibpairs) the highest LOD score achieved for one of these markers was 2.41.[121] In another North American replication study, Brant et al. demonstrated linkage at the IBD1 locus in 148 relative pairs with a highest LOD score of 2.5 in both Jewish and non-Jewish CD patients.[122] Further support for this locus was provided by linkage analysis in a cohort of 58 Italian families that demonstrated linkage in CD patients and uniquely linkage in UC patients with LOD scores of 1.97

and 2.44 respectively at a microsatellite marker located within the IBD1 locus.[123] Cavanaugh et al. studied 54 multiplex IBD families and demonstrated a remarkably high LOD score of 6.3 for the IBD1 locus in these Australian CD patients. The largest replication study contained 581 sibling pairs carried out by the international IBD consortium using a combination of previously reported studies, generating a maximum LOD score for the locus of 5.79 in CD patietns.[118] The international IBD consortium study reported no significant linkage at the IBD1 locus for UC patients.

1.12.3.2 IBD2

In the only UK genome wide scan, Satsangi et al. demonstrated an IBD susceptibility locus on chromosome 12p13 named IBD2 (MIM 601458).[107] This scan was performed in 160 IBD families with a total of 186 sibling pairs with the strongest evidence for linkage found for a region spanning 41 centimorgans on the long arm of chromosome 12. Five adjacent microsatellite markers within in the region showed distortion of allele sharing, with the most significant evidence for linkage being seen at the marker D12S83 with a highly significant LOD score of 5.47. The same GWS also provided provisional evidence for disease susceptibility on areas of chromosomes 7 and 3. The IBD2 locus was also identified in a second genome wide scan performed by Duerr et al. in a North American IBD population (122 families with 208 relative pairs) achieving a LOD score of 2.0.[109] A combination of a number of international studies (pooled data on 613 families) has failed to show strong support for linkage at this locus for Crohn's disease.[118] In this IBD consortium study that presented convincing evidence for IBD 1 locus, the IBD 2 locus achieved a LOD score of only 1.2 for CD.[118]

It is now generally accepted that the IBD2 locus may be more strongly involved in UC than CD.[119] Parkes et al. in a large study of 581 relative pairs, demonstrated a LOD score of 3.91 in the 138 UC relative pairs studied compared with a LOD score of 1.66 for the 252 CD pairs and a LOD score of 1.29 for the 191 mixed IBD pairs. Further analysis of this by locus in 904 IBD relative pairs by Achkar et al, has demonstrated the highest LOD score in patients with extensive ulcerative colitis

(LOD 3.27) suggesting this locus influences *both* disease susceptibility and disease phenotype in UC patients.[124]

1.12.3.3 IBD3

This locus is located at chromosome 6p21 and incorporates the major histocompatibility complex, has been designated IBD 3 (MIM 604519). Prior to the application of genome wide scanning, the major histocompatibility complex was the candidate gene region subject to most attention in inflammatory bowel disease**.** Allelic association studies had demonstrated association of HLA class II variants with susceptibility to both ulcerative colitis and Crohn's disease.[125] In the two meta-analyses that have been performed of GWS in the IBD population the IBD 3 locus was ranked as the most consistent CD susceptibility locus in both analyses.[113;120] Van Heel et al. analysed a total of 1952 IBD relative pairs (1068 CD, 457 UC and 427 mixed) from 10 GWS and demonstrated IBD3 met genome wide significance for IBD only.[120] In a smaller meta-analysis of 709 IBD families containing 822 sibling pairs, Williams et al. analysed their own GWS data together with data from four other GWS.[113] They demonstrated IBD3 linkage from their meta-analysis for IBD and additionally CD.

Individual GWS have also suggested a role for IBD3. Hampe et al. in a large European GWS of 353 relative pairs demonstrated a peak LOD score of 4.2 for IBD in the IBD3 locus with an equal contribution from all types of IBD.[108] The individual LOD scores generated in other GWS have been relatively modest : Rioux et al demonstrated a LOD score of 2.3 for IBD patients and Barmada et al. a LOD of 2.26 for CD.[112;114] A LOD score of 3 for this locus was demonstrated in a UK replication study of 284 siblings pairs performed by Dechario et al, again with an equal contribution from all types of IBD.[126] Further linkage analyses had also suggested an important contribution of the histocompatibility complex to susceptibility to both Crohn's disease and ulcerative colitis.[127]

Interestingly in a study of 428 relative pairs Fisher et al. analysed the IBD3 locus by sex of the patient and demonstrated that IBD 3 may be a male sex specific locus.

Stratification for sex achieving a LOD score of 5.91 in males with IBD with equal contributions from UC and CD was demonstrated in Fisher's study.[128] Phenotypic analysis of the IBD3 region suggests it predisposes to a "colonic" IBD phenotype showing association with extensive and severe ulcerative colitis ,[127;129-131] and colonic Crohn's disease.[131-133]

The contrasting phenotypic predictors of the IBD1 and IBD3 loci determinants have been examined by Ahmad et al.[132] These authors have used the data generated to propose a molecular classification of Crohn's disease contrasting the population attributable risk of the IBD1 and IBD3 loci in Crohn's disease of the ileum (40%, and 19%, respectively), with colonic Crohn's disease (0%, and 39%, respectively).

1.12.3.4 IBD4

IBD 4 (MIM 606675) is located on chromosome 14q11 and matches the criteria for confirmed linkage in CD only. The IBD4 locus was first identified in the GWS by Ma et al. in North American CD relative pairs with a LOD score of 2.8.[109] Replication of the initial study was provided in another North American CD population by Duerr et al. in a GWS of 127 relative pairs from 62 CD families that generated a LOD score for the IBD4 locus of 2.8.[111] In a third GWS scan performed in the Flemish population analysis of 149 relative pairs with IBD, the IBD4 locus achieved the highest LOD score in this population but did not reach the criteria for significance set out by Lander and Kruglyak.[115] In a large international study of 792 IBD families consisting of 892 relative pairs mean allele sharing was increased in markers spanning the IBD4 locus.[134] The allele sharing was higher in those sibling pairs in whom at least one sibling was a smoker compared to pairs where neither sibling smoked.

1.12.3.5 IBD5

The IBD5 (MIM 606348) locus on chromosome 5q31-33 was first described in a relatively small CD only GWS of 65 sibling pairs achieving a LOD score of 2.2 in

this scan.[109] In a larger Canadian GWS containing 158 IBD sibling pairs (104 sib pairs with CD) Rioux et al. demonstrated a LOD score of 3 for this locus.[112] However, further stratification of patients by age at disease onset demonstrated significant linkage in the subpopulation of patients with onset of disease below the age of 16 years achieving a maximum LOD score of 3.9. Fine mapping of this area performed by the same group in 256 IBD trios has identified a single, highly conserved 250-kb haplotype of 11 SNPs spanning a cytokine gene cluster that is associated with Crohn's disease.[135] Further, more detailed discussion of this locus, can be found in chapter 6 where it is analysed in detail in Scottish early onset patients.

1.12.3.6 IBD6

IBD6 (MIM 606674) on 19p13 awaits replication as a confirmed IBD locus. This locus was identified in Rioux et al. in the genome wide scan of the Canadian IBD population. In this analysis of 158 sibling pairs the IBD6 locus achieved of LOD score of 4.6 for IBD.[112] This locus however does not strictly fulfil the Lander and Kruglyak criteria, as independent replication has not been achieved after this initial GWS. In an attempted replication study Low et al. studied marker SNPs at the IBD6 locus in 284 IBD sibling pairs but achieved a LOD score of only 1.59 for CD.[136] Van heel et al. performed a GWS in 137 relative pairs with CD and subsequently stratified the scan results based on markers at the IBD1 and IBD5 loci.[116] The results did not identify the IBD6 locus in the initial scan but after stratification for NOD2/CARD15 status (IBD 1 gene), CD patients who were not carriers of NOD2/CARD15 variant alleles demonstrated a LOD score of 2.9 for the IBD6 locus. In the same analysis, and in contrast to the findings for IBD1, patients who possessed markers on the IBD5 risk haplotype demonstrated a LOD score of 2.4 for IBD6 providing provisional evidence for epistasis between the IBD5 and IBD6 loci.

1.12.3.7 IBD7

The IBD7 locus (MIM 605225) is located at chromosome 1p36. This was suggested as an IBD locus by Cho et al. in GWS performed on 297 IBD relative pairs from 174 families from North America, 37% of whom were of Ashkenazi Jewish origin.[110] In analysis of all subjects included in the scan a LOD score of 2.65 was seen at the IBD7 locus, with a LOD score of 2.17 for CD. In a separate extension study by the same group of investigators analysis of markers at this locus was performed in a small Chaldean population originally of Iraqi origin.[137] In total, 13 family members with IBD were studied (11 had UC) with the maximal linkage marker at the IBD7 locus generating a LOD score of 3.

1.12.3.8 IBD 8

Initial genome wide scans identified the IBD1 locus on chromosome 16 as conferring susceptibility to IBD1,[106;118] but subsequent investigation has suggested the possibility of a second IBD locus on chromosome 16 that has been designated IBD8 (MIM 606668). In an extension study of the chromosome 16 region performed by Hampe et al. in 522 IBD relative pairs, a triple peaked linkage curve was seen on chromosome 16.[138] Once the contribution of IBD1 was excluded, linkage peaks were still seen for the 2 other CD peaks located at 16p (LOD 2.7) and 16q (LOD 3.1). This finding is yet to be replicated in other IBD datasets.

1.12.3.9 IBD 9

In the GWS performed by Satsangi et al, a linkage region was identified on chromosome 3p26 as well as regions on chromosome 12 and 7, with this area subsequently designated IBD9 (MIM 608448). In Satsangi's scan of 186 IBD sibling pairs suggestive linkage was demonstrated for the IBD9 locus with a LOD score of 2.69.[107] In a follow up extension study, Hampe et al. studied markers at the IBD9 locus in 353 sibling pairs demonstrating a LOD score of 1.65 for IBD predominantly from a CD linkage peak.[139] TDT analysis of one of the markers in the IBD9 locus

demonstrated significant over transmission in IBD trios. In a similarly designed replication study, Duerr et al. demonstrated significant over transmission of markers at this locus in 324 IBD sibling pairs with a maximum LOD score of 3.78.[140]

1.12.3.10 Genome wide scans in children with IBD

Although children have been part of many of the GWS, none to date have been exclusively paediatric.[103] The most exciting data relating to paediatrics were presented in the Canadian GWS which stratified patients by age at diagnosis and demonstrated CD patients under the age of 16 had the highest LOD score (3.9 cf. 3.0).[112]

Brant et al. studied markers on the IBD1 locus in two groups of patients, one group of early onset patient with severe disease, (age less than 22 years at diagnosis taking immunosuppressants or those who had surgery) and compared them to a group with neither of these characteristics.[94] The first group had a significantly higher LOD score for the 16 markers analysed at the IBD1 locus with a LOD score of 3.84. In a similar study by Akolkar et al. of marker SNPs at the IBD1 locus there was initially minimal evidence of linkage in the whole cohort (LOD 1.63) of 53 Ashkenazi Jewish families studied, but when the analysis was stratified to only include patients diagnosed under the age of 21 years the LOD score increased to 3.02.[93]

1.12.4 GWS- applications and limitations

The 11 GWS scans that have been performed over the past 10 years in IBD patients have produced exciting data but this enthusiasm has been tainted by the lack of subsequent gene identification for the majority of these loci. Aside from the notable success of transition from linkage to gene identification at the IBD1 locus, none of the other loci have provided conclusive evidence resulting in the *causal* candidate gene within the locus being identified. Gene identification has initially been hampered by lack of replication of some loci in extension studies. The size of an individual GWS is limited usually by the number of sibling pairs available in a specific population so these scans have been based on pragmatic numbers available

for recruitment rather than based on pre-calculated power calculations. This has meant some of the areas identified on scans may be false positives which may explain the lack of replication of some loci.

The relatively small numbers of relative pairs in some of these scans has meant statistically meaningful results have only been demonstrated by combining studies. Two groups have done this by performing meta-analyses of genome wide scans.[113;116] The combination of data after a loci has been identified has also been a very powerful way of making best use of results e.g. the international IBD consortium collaboration on the IBD 1 and IBD2 data.[118] The relative lack of power in these individual studies also makes the identification of genes of modest effect difficult using this method of gene identification.

Even when the genetic data for the loci in question has been proven robustly, there have then been difficulties narrowing down the regions of linkage identified on GWS. This problem of gene identification is highlighted by the IBD5 locus where tight linkage disequilibrium over a gene dense region has made gene identification very difficult even when the studies have involved several thousand patients.[141] The scans performed to data have also been CD dominated so this method has not been able thus far to identify genes important in UC susceptibility.

The study of patients, families and populations with IBD has therefore highlighted a key role for genetics in the pathogenesis of IBD. GWS have moved IBD genetics onto a stage of hypothesis driven candidate gene studies but this has not yet resulted in widespread success in identifying many of the candidate genes contained within these IBD loci. The role of the individual candidate genes and their influence in the Scottish childhood population is described in the subsequent chapters. The exciting data evolving from GWAS looks to hold more promise in moving on from identifying regions of interest to gene identification.

1.12.5 Thesis plan

The genetic background to inflammatory bowel disease underpinning this thesis has been presented in this chapter including where relevant, studies that have involved

children. The nine IBD loci identified are exciting step forward in the further understanding of the genetics of IBD but widespread identification of genes involved in IBD has yet to materialise from these studies.

The hypothesis proposed in this thesis is that childhood IBD therefore characterises a disease group with a stronger genetic influence and represents a group with a distinct disease phenotype. Apriori this hypothesis has been tested by the examination of candidate genes implicated in adult studies by first examining their contribution to IBD susceptibility in Scottish children and then analysed in detailed genotype phenotype studies that focussed on associations that may be present only in the paediatric population including post hoc analysis of several factors that were identified as areas of interest as the thesis was being performed.

To test this hypothesis, a cohort of patients with early onset IBD, diagnosed with the disease at less than 16 years of age were collected from across Scotland. Blood was collected for subsequent DNA extraction from patients with IBD and their parents. Data was collected by questionnaire and examination of a patient's case notes to rigorously characterise the patient's disease phenotype both at disease diagnosis and later follow up. Statistical analysis was performed by transmission disequilibrium testing (TDT) in the family trios as well as case control analysis. The controls were collected from parallel studies performed in the Scottish adult IBD population. In comparison of disease phenotype, phenotypic classification used for defining adult Crohn's disease (Vienna) was used with an additional classification for use in children's studies was designed specifically for use in this thesis. The Montreal classification was used for direct comparison between the childhood and adult populations.

In chapter 3 the similarities and differences between childhood onset and adult onset IBD phenotype were analysed, justifying why independent studies in the childhood onset population were important.

Chapter 4 analysed the role of the most widely implicated gene to date (NOD2/CARD15) in the Scottish early onset IBD population. The specific contribution of the common three disease susceptibility mutations implicated in previous studies were analysed in 247 Scottish IBD patients aged <16 years at

diagnosis and 414 parents by intrafamilial studies. The NOD2/CARD15 variants were also analysed in a case control study using 245 healthy controls, before genotype phenotype analysis was undertaken.

In chapter 5 a comprehensive analysis of the whole NOD2/CARD15 gene was performed after sequencing the 12 exons of the gene in 24 early onset patients with Crohn's disease. Then, 2 SNPs were analysed in greater detail assessing their contribution to disease susceptibility in a larger of population of more than 500 Scottish Crohn's disease patients consisting of both children and adults, which were then compared to results in 253 healthy controls.

The role of a second candidate gene (OCTN) within the IBD5 haplotype was analysed in Chapter 6. The role of this gene together with 3 SNPs from within the IBD5 haplotype were analysed in 299 childhood onset IBD patients assessing the contribution to disease susceptibility and phenotype. The controversies surrounding the role of OCTN genes in disease susceptibility were then discussed examining whether they are truly independent of the background IBD5 haplotype.

The role of a single SNP 113 G/A within the DLG5 gene, another putative IBD susceptibility gene was discussed in chapter 7. The SNP was assessed for both its contribution to disease susceptibility and phenotype in 296 IBD patients, before discussing the potential controversies surrounding this genes role in disease susceptibility from all studies to date as well as exploring potential genotype phenotype interactions.

Chapter 8 switches focus from genes to serological markers. The frequency of Anti-Saccharomyces cerevisiae antibodies (ASCA) antibodies in 301 children was analysed and frequencies compared between the 197 CD, 76 UC and 28 IC patients as well with 78 healthy population controls. Further analysis of ASCA status and disease phenotype was then performed in patients with CD at diagnosis.

In chapter 9 the implications of the data presented throughout the thesis are discussed. The results are placed in the context of current work performed during and after the completion of the experimental work of the thesis. The way forward to identifying new genes in the Scottish IBD population and a plan for continuation of the work is then discussed.

2 Patients, materials and methods

2.1 Definitions

2.1.1 "Early onset" patients

The age at which childhood ends and adulthood begins represents a continuum and the age limits that define childhood or meet a definition of early onset IBD are somewhat arbitrary. The definitions used for early onset patients in IBD have varied between studies and have included <16 years,[112] <18 years, <19 years,[142] <20 years,[79] <21years,[53] < 22 years[94] and even less than 40 years![36] The only scientifically defined age cut off is at less than 16 years and was derived from the study of the IBD5 locus by Rioux et al (see chapter 1) in the Canadian IBD population. This study demonstrated a bimodal age distribution of patients with IBD and used statistical evidence from Gaussian distribution curves to set this age limit.[112] For the purposes of this thesis we have targeted children with an age at diagnosis of IBD at less than 16 years at diagnosis. This decision was based pragmatically on the best available evidence and also represented an age at which it could be estimated a critical number of children could be recruited to allow for meaningful genetic analysis within the allocated study period.

2.1.2 Disease definition

The diagnosis of IBD was based on standard criteria as set out by Lennard-Jones.[143] After enteric infection was excluded the criteria were used to classify a patient as having either ulcerative colitis or Crohn's disease. A checklist based on these criteria was constructed to ensure each patient entering the study definitely fulfilled the criteria for having Crohn's disease (appendix 1). Using this checklist patients were defined as fulfilling the criteria for Crohn's disease if they met the definitions in three or more section criteria. If a given patient scored less than 3 in three *separate* section criteria then he/she was considered not to have Crohn's disease. The only exceptions to this standard were patients who had isolated colonic Crohn's disease who had granulomas and a pancolitis with inflammatory type disease behaviour; using these criteria a score of only 1 or 2 may be generated by such patients but the

presence of epitheliod granulomas in these circumstances was deemed enough to classify the patient as having Crohn's disease, as noted in the classification itself.

Patients were diagnosed with ulcerative colitis when they had uniform inflammation that commenced in the rectum and continued for a varying extent more proximally. In addition to chronic inflammation on histology the patients had evidence of ulceration, cryptitis, crypt abscess formation together with goblet cell/mucin depletion without any epitheliod granulomata.

A patient was categorised as having "indeterminate colitis" if definite evidence of chronic inflammatory bowel disease affecting the colon only was present, but the patient remained unclassifiable as either CD or UC, after considering all clinical, radiological, endoscopic and pathological findings.

2.2 Study participants

2.2.1 Patient and parents

Patients and parents were recruited from the three tertiary paediatric gastroenterology centres in Scotland based in the specialist childrens' hospitals in Edinburgh, Glasgow and Aberdeen between 2002 and 2005. Further study participants who were diagnosed with IBD under the age of 16 but who had transitioned to adult services at the Western General Hospital in Edinburgh were also recruited to the study. Information about the purpose of the study was supplied to all patients and parents prior to approaching them for study enrolment (appendices 2 and 3 respectively). In those patients and families who agreed to take part written consent was then obtained from patients and parents prior to participation in the study (appendices 4 and 5 respectively). The information sheets and consent forms shown in the appendices are examples of the type of forms used. The exact details of each form were all slightly different having been adapted to a style acceptable to the local research ethics committee.

2.2.2 Blood collection

Each patient had up to 10 millilitres (mls) of blood collected into EDTA and lithium heparin tubes. The blood was usually collected when the patients were having blood taken as part of their on going clinical care as requested by their supervising clinician and not merely for participation in the research study. A topical local anaesthetic cream [Eutectic Mixture of Lidocaine and Prilocaine (EMLA® Astra-Zeneca) or Amethocaine base 4.0%w/w (Ametop® Smith & Nephew)] was applied if requested by the patient or family, prior to venesection. Parents each gave 10mls of whole blood collected into an EDTA tube collected in a Vacutainer® (Nu-care) system. In the case of "parents" who were not the genetic parents of the child, as disclosed in confidence by the guardian to the researcher, "sham" blood taking was undertaken by placing a plaster on the carer's arm, to avoid disclosure of any potentially sensitive information to the child in the course of the family agreeing to take part in the study.

2.2.3 Healthy control recruitment

Healthy adult controls were recruited from healthy volunteers and from volunteers recruited from the Scottish blood transfusion service. The number of controls used in each of the genetic studies varied and the exact number used in the study of each of the candidate genes is listed individually in the subsequent chapters.

2.2.4 Ethical approval

Ethical approval was given for all participating centres (Edinburgh, Lothian Research Ethics Committee (LREC) Reference Number: LREC/2002/6/18; Aberdeen, Grampian Research Ethics Committee (GREC) reference number: GREC 03/0273; Glasgow, Yorkhill Research Ethics Committee (YREC) reference number: YREC number P12/03).

2.2.5 Data collection

Retrospective data were collected on patients using case note review and a patient /parent questionnaire). All patients with IBD completed a questionnaire, collecting

information on patient and parental smoking details, ethnicity and a detailed family history of IBD and other diseases (see appendix 6). Additional clinical data was collected on patient demographics, age at IBD symptom onset and diagnosis, medications, extraintestinal manifestations, history of atopic disease and need for surgery after the patient questionnaire was combined with examination of patient case notes.

2.3 Disease phenotype

All patients recruited to the study were phenotyped personally or with data entry assistance from Hazel Drummond (database manger). Proforma sheets were designed to collect the data in a consistent manner and to further reduce the chances of inter-observer error for Crohn's disease and Ulcerative/Indeterminate Colitis (Appendices 7 and 8 respectively). The phenotypic criteria used were the Vienna classification for patients with Crohn's disease [36] and a paediatric specific phenotypic classification for use in all types of IBD. The Montreal classification was used to directly compare childhood and adult onset IBD patients in chapter 3 only. Phenotypic information was collected at diagnosis and reassessed in 2 yearly intervals for CD patients for the first 10 years post diagnosis and then collected at 5 year intervals subsequently. In patients with UC/IC phenotypic details were collected at diagnosis and the time of last follow.

2.3.1 The Vienna classification

A summary of the classification appears in table 2-1.

2.3.1.1 Disease location

This classification divided disease location into four different disease locations prefixed by the letter L for location; L1 was used to represent Ileal disease and was defined as disease in the distal third of small bowel, with or without caecal involvement; L2 was used to represent colonic disease activity only and disease could be present nowhere else within the GI tract; L3 Ileo-colonic represented a combination of the first two disease location categories and was defined by disease in the Ileum i.e. the distal third of the small bowel plus at any location within the colon; L4 represented disease in the upper gastrointestinal tract anywhere proximal to terminal ileum with the exception of the mouth, with or without ileal and/or colonic involvement. Disease locations were exclusive i.e. each patient could only have a disease location in one category and once a patient was classified in the L4 location they could not move from that category at further disease follow up. Patients who had oral or perianal Crohn's disease, or a combination of both in the absence of disease activity elsewhere in the GI tract, did not fit into this classification system. For the purposes of analysis they were classified as disease location "none" and were analysed as a 5^{th} disease location not forming part of the original classification.

Disease extent was defined as the maximum disease involvement prior to the first surgical resection. The minimum criteria for involvement of a disease location were ulceration or the presence of an aphthous lesion. If only erythema and/or oedema were present that was not deemed enough to signify disease at that site.

2.3.1.2 Disease behaviour

The classification divided disease behaviour into three different categories prefixed by the letter B. B1 represented inflammatory disease behaviour and was represented by non-stricturing, non-penetrating disease as defined in the B2 and B3 behaviour

categories; B2 was used to represent stricturing disease that was defined as a constant luminal narrowing on radiology, at the time of endoscopy or surgery with a pre-stenotic dilatation or if accompanied by obstructive signs and symptoms clinically; B3 represented penetrating disease that represented intra-abdominal or perianal fistulas, or inflammatory masses plus or minus abscess formation at any time excluding complications arising in the immediate post operative period.

Like disease location each behaviour category can never be down graded which meant that patients who are in the B3 category cannot move from that category. In a patient who has stricturing and penetrating disease, classification is to the category B3.

2.3.1.3 Age at diagnosis

The age (A) category simply refers to the patient's age at diagnosis; either as A1 < 40 years or A2 > 40 years old. All of the patients in this study were in the A1 category so analysis was performed using age at diagnosis as a variable.

Table 2-1: Summary of the Vienna classification of Crohn's Disease

Disease Location	L1	Terminal ileum
	L2	Colon
	L3	Ileo-colonic
	L4	Upper GI tract
Disease Behaviour	B1	Non-stricturing, non penetrating
	B2	Stricturing
	B3	Penetrating
Age at Diagnosis	A1	<40 years
	A2	>40 years

2.3.2 Proposed paediatric phenotypic classification

In the absence of a recognised paediatric classification a paediatric phenotypic classification was devised for use in this thesis. In patients with Crohn's disease, location was categorised and analysed according to individual disease locations, regardless of disease involvement elsewhere in the gastrointestinal tract in contrast to the hierarchical model used in the Vienna classification. This model was adopted to fully explore potential relationships between disease location and a patient's genotype.

2.3.3 Crohn's disease

2.3.3.1 Disease Location

This phenotype therefore recorded the presence or absence of evidence for CD activity in each anatomical location (based on findings from endoscopy, biopsy, surgery or barium follow through) within the gastrointestinal tract, namely: oral, oesophageal, gastric antrum and body, duodenal, jejunal, ileal, caecal, ascending, transverse, descending and sigmoid colon together with rectum and peri-anal disease (table 2-2). In addition to these general disease definitions some further specific definitions were made for specific disease sites.

1. Oral Crohn's disease was defined by macroscopic changes (mucosal tags, deep linear ulceration, cobblestoning, lip swelling and fissuring) after examination by a paediatric dentist/oral medicine specialist only. The presence of epitheliod granulomata on buccal biopsy was also classified as oral Crohn's disease. The term oral CD was preferred to the term oro-facial granulomatosis.

2. Patients were diagnosed with Crohn's disease of the upper GI tract (oesophagus, stomach, duodenum or jejunum) when biopsies from any of these sites confirmed the presence of epitheliod granulomas (not merely in the presence of chronic inflammation) and/or when features of macroscopic disease were present. The presence of a chronic gastritis in the absence of other features of Crohn's disease was *not* sufficient to indicate disease involvement of the gastric body/antrum.

3. Perianal disease was defined by the presence of fissures, fistulae, abscess formation or perianal ulceration but not by the presence of skin tags alone.

4. The presence of granulomata was assessed by endoscopic biopsy at diagnosis or at the time of surgical resection.

2.3.3.2 Disease behaviour

The definitions for disease behaviour were the same as those used in the Vienna classification.[36] The reason why a patient was analysed as penetrating disease was recorded but not analysed separately.

Table 2-2: Summary of paediatric classification of Crohn's Disease

Disease Location	L1	Oral
	L2	Oesophageal
	L3	Gastric antrum
	L4	Gastric body
	L5	Duodenal
	L6	Jejunal
	L7	Ileal
	L8	Caecal
	L9	Ascending colon
	L10	Transverse colon
	L11	Descending colon
	L12	Sigmoid colon
	L13	Rectum
	L14	Perianal
Disease Behaviour	B1	Inflammatory
	B2	Stricturing
	B3	Penetrating

2.3.4 Ulcerative colitis

2.3.4.1 Disease location

No widely recognised disease classification for patients with Ulcerative colitis was widely accepted at the time of commencing the scientific work of this thesis in September 2002, either in children or in adults. For the purposes of the study disease extent was dived into 6 categories depending on the maximum disease extent defined by either macroscopic appearance or histology (rectum, sigmoid, descending, transverse, ascending and pancolitis when inflammation involved the whole colon) as listed in table 2-3. It was also noted whether the patient had a full assessment i.e. Colonoscopy or a limited assessment i.e. sigmoidoscopy and a subjective assessment of whether the examination was adequate i.e. good bowel preparation or not. Details of where the examination stopped were also recorded i.e. in patients undergoing colonoscopy whether the examiner reached the terminal ileum or whether the examination stopped prior to this at a specific colonic location.

Table 2-3: Summary of paediatric classification of Ulcerative colitis

Disease location	L1	Rectum
	L2	Sigmoid colon
	L3	Descending colon
	L4	Transverse colon
	L5	Ascending colon
	L6	Pancolitis

2.3.5 Surgery

Surgery was defined as any operative intervention for IBD except for examination under anaesthesia without any other surgical intervention. This definition was further adapted during the course of the study dividing surgical procedures for Crohn's disease into resections, drainage procedures or other types of operations.

2.3.6 Growth data

The basic anthropometric information collected in all patients were weight (in kilograms) and height (in metres). All patients had their height measured on a wall mounted stadiometer and weight measured on regularly calibrated scales. A body mass index (BMI) was calculated thereafter by dividing the patients weight by their height squared. The data were then plotted on a standard UK centile chart using the UK 1990 population data (© Child Growth Foundation 1996) and one of ten centile bands was then allocated. The centile band allocated was one of the following: $<0.4^{th}$, 0.4^{th}-2^{nd}, 2^{nd}-9^{th}, 9^{th}-25^{th} , 25^{th}-50^{th}, 50^{th}-75^{th}, 75^{th}-91^{st}, 91^{st} -98^{th},98^{th}-99.6^{th} and $>99.6^{th}$.

During the course of the study the anthropometric data collected were also defined using z-scores (British 1990 Growth Reference for height, weight and BMI ©Child Growth Foundation, 2 Mayfield Avenue, London W4 1PW). A *z score* for an item indicates how far and in what direction, that item deviates from its distribution's mean, expressed in units of its distribution's standard deviation prefixed by a + or -. A correlation of the patient's growth with their pubertal status was also recorded. The collection of pubertal staging however was abandoned during the study as it became apparent this data was poorly and inconsistently recorded in the patient's case notes, making it unsuitable for further detailed analysis.

2.3.7 Haematological and biochemical disease markers

Data were collected, using proforma sheets, for all patients who had blood tests performed at the time of diagnosis and at designated times of follow up i.e. 2 yearly for patients with CD and at diagnosis and last follow up for patients with UC/IC

(appendix 9). The values used to signify abnormal blood tests were as follows: Albumin <35 g/l, Erythrocyte sedimentation rate (ESR) > 10mm in the first hour. C-reactive protein (CRP) was recorded as normal or high (the specific units used varied between various laboratories used by the participating centres in the study). Abnormal values for haemoglobin concentration were age dependent: age 1-3 years <110 g/l, 3-6years <117g/l, >6 years <120g/l and platelets >400 $x10^9$/l. The normal ranges for vitamin B_{12} were 193–982 nanograms/l^{-1} and for serum folate 3–7μg/l^{-1}. Faecal Calprotectin was considered abnormal at > 50 μg/g.

2.4 Patients

2.4.1 Crohn's disease

The location of Crohn's disease in individual disease locations are listed in table 2-4. This table demonstrates that ileal and colonic disease locations are the commonest disease sites at disease presentation in children at 64% and 75-78% respectively. In patients with oral disease (15%) there were an equal number of cases diagnosed by expert clinical opinion and by buccal biopsy. The number of patients with oesophageal Crohn's disease is notably small at <3%. The oesophagus was the commonest disease site not to be biopsied and so this may be an underestimate of the true prevalence of oesophageal Crohn's disease.

Gastric Crohn's disease is a relatively common disease site with similar disease prevalence in both the gastric body and antrum at around 25% for each site. The higher prevalence in the antrum reflected the fact this site was preferred for gastric biopsies and that body biopsies were not always taken. Duodenal and jejunal CD were equally common at around 15%

Ileal disease entailed disease involving the terminal ileum in the majority of cases. This was demonstrated equally by changes found at endoscopy and on barium studies. The colon was the commonest site of disease and despite surveying regular sites throughout the colon disease prevalence in all sites was equally common at 75-78% of samples taken. At 2 year follow up there was no significant change in disease location at all sites.

2.4.2 Vienna classification

The location and behaviour of CD according to the Vienna classification are listed in table 2-5. In contrast to the findings of individual disease sites, ileal and colonic disease categories were the *least* common category in this disease classification, with the presence of isolated ileal disease in less than 5%. The L4 category of upper GI disease was the most common (46%), being at least twice as common as any of the other disease sites. It should also be noted that in 6% of patients with a diagnosis of CD the patients did not have a disease site that was recognized by the classification.

Children less than 8 years at diagnosis were proportionally more likely to have isolated colonic disease (L2) than those children diagnosed at more than 8 years of age 37.2% vs. 20.0%, p=0.02 OR=2.37[1.14-4.93].

At diagnosis, the majority of patients had inflammatory disease behaviour (81%), but there were a notable number of patients (14%) with penetrating disease behaviour, predominantly reflecting patients with peri-anal abscesses and fistulae.

At 2 year follow up, disease location as defined by the Vienna classification had not significantly changed compared to location at diagnosis. Disease behaviour was progressing with more stricturing and penetrating disease. Penetrating disease was present in 26% of patients at 2 years post diagnosis compared to 14% at diagnosis (p=0.02, OR= 2.09 [1.13-3.87]).

2.4.3 Ulcerative colitis

The disease locations for Ulcerative colitis at disease diagnosis and time of last follow up are shown in table 2-6. The median time of follow up was 2.33 years (IQR 1.33-4.17 years). Data on disease location was available for 96 children at diagnosis and 71 at follow up. Pancolitis was the commonest disease phenotype at diagnosis and follow up with prevalence close to 60%. Proctitis was the least common disease extent found in less than 2% of cases at follow up. There was no significant change in the distribution of disease location at the time of last follow up. There were 18 children diagnosed at <8 years of whom 8 (44%) had a pancolitis at diagnosis.

2.4.4 Indeterminate colitis

The disease locations for patients with indeterminate colitis are shown in table 2-7. The median time of follow up was 2.33 years (IQR 1.33-4.17 years). Data on disease location was available for 37 children at diagnosis and 21 at follow up. As for patients with ulcerative colitis, pancolitis was the commonest disease extent affecting 66.7% of patients at follow up. The median age at diagnosis was 9.9 years (IQR 7.7-11.9years) and median time of follow up was 2.2 years (IQR 1.6-3.9 years).There were 10 children diagnosed at <8 years of whom 8 (60%) had a pancolitis at diagnosis.

Table 2-4: CD Disease location at diagnosis and 2 year follow up

Disease Location	Diagnosis	2 year follow up
Oral	20/152 (13.6%)	14/98 (14.3%)
Oesophageal	4/142 (2.8%)	1/90 (1.1%)
Gastric antrum	39/145(26.9%)	12/84 (14.3%)
Gastric body	30/138 (21.7%)	18/94 (19.1%)
Duodenal	23/137 (16.8%)	9/88 (10.2%)
Jejunal	22/136 (16.2%)	14/90 (15.6%)
Ileal	91/142 (64.0%)	63/94 (67.0%)
Caecal	91/119 (76.5%)	58/79 (73.4%)
Ascending	92/122 (75.4%)	56/74 (75.7%)
Transverse	99/129 (76.7%)	60/79 (75.9%)
Descending	104/134 (77.6%)	61/81 (75.3%)
Sigmoid	115/147 (78.2%)	67/92 (72.8%)
Rectum	117/151 (77.5%)	67/93 (72.0%)
Perianal	66/158 (41.8%)	37/100 (37.0%)

Table 2-5: CD Vienna classification at diagnosis and 2 year follow up

Disease Location	**At Diagnosis**	**2 year follow up**
Terminal Ileum (L1)	8 (4.8%)	5 (4.8%)
Colon (L2)	38 (22.8%)	28 (26.9%)
Ileocolon (L3)	34 (20.3%)	22 (21.2%)
Upper gastrointestinal (L4)	77 (46.1%)	45 (43.3%)
None*	10 (6.0%)	4 (3.8%)
Disease Behaviour		
Inflammatory (B1)	136 (81.4%)	66 (63.5%)
Stricturing (B2)	7 (4.2%)	11 (10.6%)
Penetrating (B3)	24 (14.4%)	27 (25.9%)

Table 2-6: UC location at diagnosis and last follow up

Disease location	**Diagnosis**	**Last follow up**
Rectum	4 (4.2%)	1 (1.4%)
Sigmoid colon	12 (12.5%)	7 (9.9%)
Descending colon	9 (9.4%)	5 (7.0%)
Transverse colon	12 (12.5%)	10 (14.1%)
Ascending colon	3 (3.1%)	6 (8.5%)
Pancolitis	56 (58.3%)	42 (59.1%)

Table 2-7 IC location at diagnosis and last follow up

Disease location	**Diagnosis**	**Last follow up**
Rectum	5 (13.5%)	0
Sigmoid colon	3 (8.1%)	1 (4.7%)
Descending colon	2 (5.4%)	1 (4.7%)
Transverse colon	4 (10.8%)	2 (9.5%)
Ascending colon	2 (5.4%)	2 (9.5%)
Pancolitis	19 (51.4%)	14 (66.7%)
Unknown	2 (5.4%)	1 (4.7%)

2.4.5 Growth

The mean z scores for weight, height and BMI for 235 children with Crohn's disease at diagnosis were -0.67 (SD 1.43), - 0.51(SD 1.62) and -0.66 (SD 1.55) respectively. The equivalent figures for 85 children with ulcerative colitis were -0.07 (SD 1.15), -0.37 (SD 2.92) and -0.02(SD 1.07) respectively. When comparing results for CD with UC using a t-test the respect values for weight, height and BMI were ($p<0.0001$ 95% CI [-0.91 to -0.29], $p=0.7$[-0.84, 0.57] and $p<0.0001$ (-0.95, -0.31].

2.4.6 Surgery

Of the 256 CD patients with surgical details available 64 (25%) had at least one surgical intervention. Of those needing surgery, drainage operations were the most common in 30 (47%) patients, followed by resection in 25 (39%) with the remaining patients having combined operations or other surgical procedures. The median time to the first surgical resection was 1.92 years (IQR 0.33-4.54 years). The median time of follow up of these patients as a whole was 4.75 years (IQR 2.08-10.17 years).

In patients with ulcerative colitis details of surgery at follow up were available in 70 patients who had been followed up for a median time of 2.33 years (IQR 1.33-4.17 years). In this group of patients 15 (21.4%) had a colectomy with a median time to surgery from diagnosis of 1.5 years (IQR 0.7-3.1 years). In the 21 patients followed up with a diagnosis of IC only 1 had a colectomy.

2.5 Database

All study data were entered into a Microsoft Access 2000© database. Anonymised data were stored in this password-secure Access database, which was designed and maintained for the purpose of this study and the parallel genetic studies in adult IBD patients. The early-onset database was personally conceived and constructed for the thesis incorporating all of the data collected in the course of the study. All data were entered personally or by Hazel Drummond (database manager). All data were backed up on compact disc weekly, or after a period of large data entry before being stored

in secure fire-resistant container. A dedicated computer solely used for research studies without connection to the internet was used for storage of the study data.

2.6 Statistics

2.6.1 General statistics

Minitab statistical software version 13 (Minitab ltd, Coventry, UK) and GraphPad Instat (version 3.06 for Windows 98, GraphPad Software, San Diego California USA, www.graphpad.com) was used to analyse genotype- phenotype associations using Chi squared or Fishers exact test where appropriate for unifactorial analysis. Non-parametric data were analysed using the Mann Whitney U test in Minitab. In chapter 8 statistical analyses was performed using SPSS© (version 11.5, SPSS Inc. Headquarters, 233 S. Wacker Drive, 11th floor, Chicago, Illinois 60606). Multifactorial analysis was performed using Minitab software.

2.6.2 PedCheck

The PedCheck programme (Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA) was used to check for genotyping error and potential cases of non-paternity.[144] Any family trios with inconsistent results incompatible with normal patterns of inheritance were excluded from further genetic analyses.

2.6.3 Population attributable risk

The population attributable risk percentage (PAR %) was defined as the excess rate of disease in individuals with a mutation compared to those without. As described by Schlesselman, the PAR% is calculated from the attributable risk as a function of the prevalence in the exposed population, divided by incidence of IBD in the population.[145] To calculate this, the prevalence of CD was estimated at 100/100,000 and the frequency of all alleles in the control population was assumed to reflect that of the general population.

2.6.4 Transmission disequilibrium testing

Transmission disequilibrium testing analysis was performed using the TRANSMIT package (Version 2.5, David Clayton, Cambridge, UK).[146] This method makes full use of the data available using inferred genotypes if genotypic data from both parents are not available. Assistance with data analysis with the TRANSMIT programme was provided by Dr Niall Anderson (Senior lecturer in statistical genetics, Edinburgh University).

2.7 Cellular methods

2.7.1 DNA extraction

Genomic DNA was extracted from blood using a modified salting-out technique (appendix 10).[147] The optical density of the resultant DNA was then measured using a Gene Quant Pro instrument to estimate the actual DNA concentration. The solution was then re-suspended in an appropriate volume of 1xTE (10mM Tris (pH8.0), 1m EDTA (pH8.0)) to give a final DNA concentration of 100ng/µl.

2.7.2 Lymphocyte and Plasma extraction

Peripheral blood lymphocytes were extracted from the lithium heparin samples collected from patients with IBD. The lymphocytes then underwent subsequent cryopreservation and remain in this state currently. Plasma was isolated at the time of peripheral blood lymphocytes extraction and samples were stored at -80^{o}C until use. The lymphocyte extraction and separation of plasma was carried out by the staff in the Wellcome Trust clinical research facility, University of Edinburgh.

2.7.3 PCR

All standard Polymerase chain reactions (PCR) were run on a Techne Touchgene Gradient machine. The reagents used and reaction conditions are detailed for the individual SNPs in subsequent chapters.

2.7.4 DNA sequencing

In Chapter 5 primers for sequencing the NOD2/CARD15 gene were designed using data from Ensembl (www.ensembl.org). Direct sequencing was performed on the 7900 HT sequence detection system (Applied Biosystems, Foster City, Ca, USA) by the Technical Services section of the MRC Human Genetics Unit, Edinburgh. DNA sequences were analysed using Sequencher v 4.5 (Gene Codes Corporation, Ann Arbor, MI, USA). An example of the sequence produced for analysis is shown in figure 2-1.

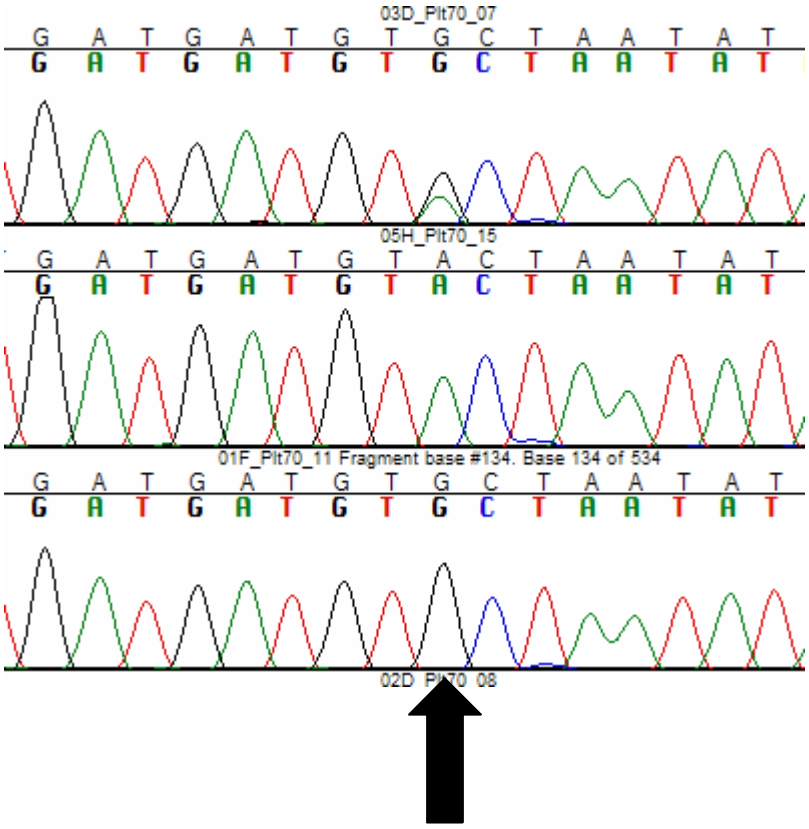


Figure 2-1: An example of genomic sequence from Sequencher

The arrow marks sequence demonstrating heterozygote, homozygous wild type and homozygous mutant respectively, reading from top to bottom.

2.7.5 TaqMan

The majority of genotyping was performed using TaqMan® (© Biosystems) and was performed in the Welcome Trust Clinical Research Facility (WTCRF) located within the Western General Hospital, Edinburgh. This PCR-based assay uses laser scanning technology that excites fluorescent dyes present in specifically designed probes. The system includes a built-in thermal cycler, a laser to induce fluorescence, CCD (charge-coupled device) detector, real-time sequence detection software, and TaqMan® reagents for the fluorogenic 5' nuclease assay. The cycle-by-cycle detection of the increase in the amount of PCR product is quantified in real time as the special probes, "reporter dye", fluoresces when the "quencher" is removed from the fragment during the PCR extension cycle.

2.7.5.1 SNP design

TaqMan genotyping was ordered using primers from an Assay on Demand Assay for SNPs that had previously been genotyped or, the target sequence for each SNP was submitted to the Assay by Design Service for Custom SNP Genotyping Assays. The Assays from these services consist of a mix of unlabelled primers and TaqMan Minor Groove Binder probes. Genotyping was performed using DNA plated in 384 (4x96) well-plates, using the TaqMan polymerase chain reaction-based method. . In each plate, negative control wells are used to check the quality of reaction.

2.7.5.2 PCR reaction

The final volume PCR reaction was 5 micro litres (µl) using 20 nanograms of genomic DNA, 2.5 µl of TaqMan Master Mix and 0.125µl of 40x Assay By design Genotyping Assay Mix, or 0.25µl of 20x if using Assay On Demand Genotyping Assay. The cycling parameters were as follows: 95° for 10 minutes, followed by 40 cycles of denaturation at 92° for 15 seconds and annealing/extension at 60° for 1 minute. PCR plates were then read on ABI PRISM 7900HT (Applied Biosystems)

instrument with Sequence Detection System (Applied Biosystems) version 2.1 software.

2.7.5.3 Allele discrimination

Allelic discrimination using this chemistry is based on the design of two TaqMan probes, specific for the wild type allele and the mutant allele separated using laser scanning technology. Each of the two probes is labelled with a different fluorescent tag 6-carboxy-fluorescein (FAM) and VIC®. Each probe is designed with the gene mutation affecting the middle part of the probe sequence. The binding efficiency of the wild type TaqMan probe to the mutant allele and vice versa is low, due to the mismatch within the TaqMan probe and the target sequence; therefore, mismatched binding is highly reduced.

2.7.5.4 Results

Allocation of allele status as homozygous wild type, heterozygote or homozygous mutant was made by personnel in the WTCRF blinded to the study aims. The results from TaqMan studies were then collected onto an Excel 2000© spreadsheet before being entered onto the Access 2000© study database, with results matched using pre-assigned unique study numbers. An example of the raw data plot generated from a TaqMan assay is shown in figure 2-2.

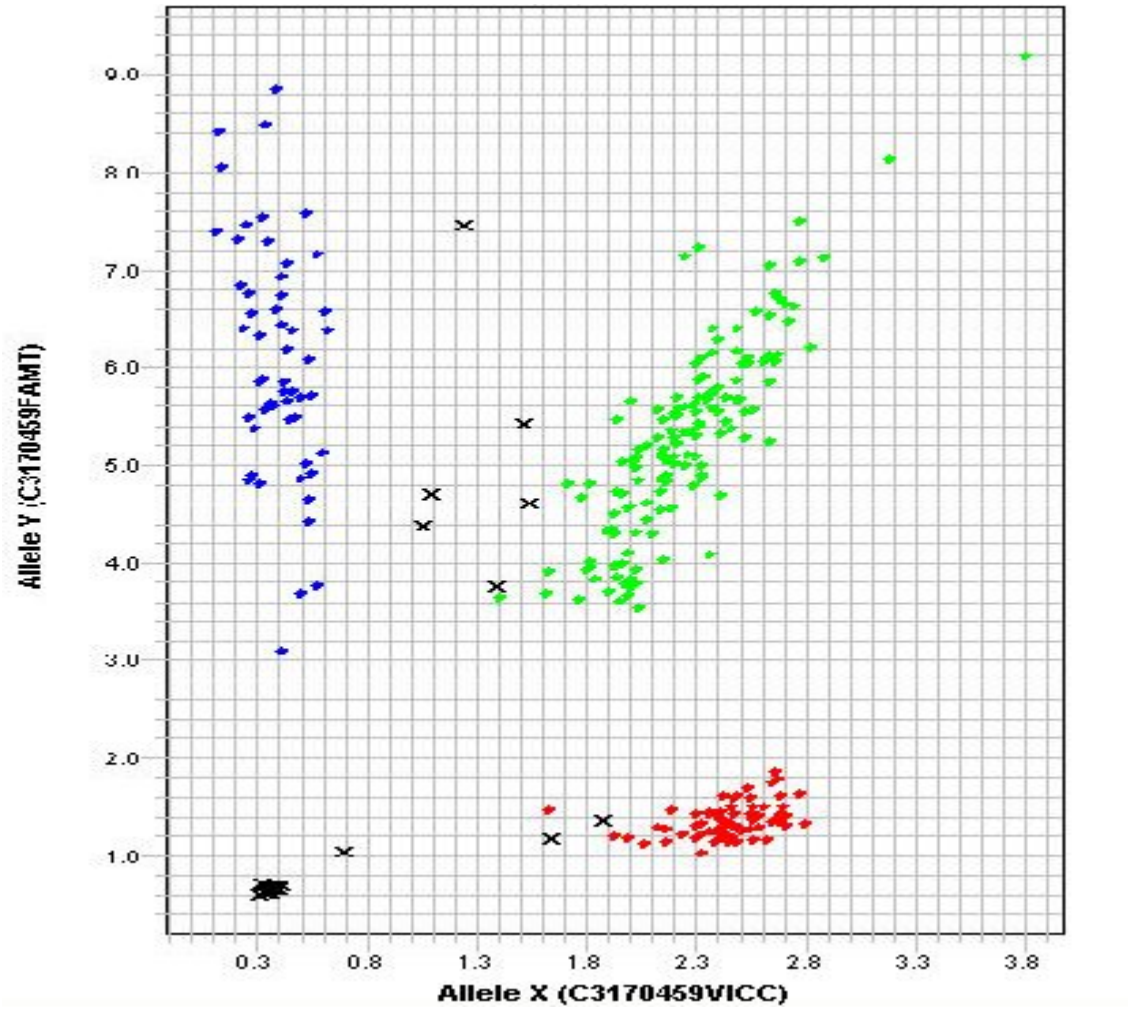


Figure 2-2: Example of allelic discrimination using TaqMan

Blue and red diamonds represent homozygotes binding to FAM and VIC respectively. The green diamonds represent heterozygotes. The black crosses represent samples where no allelic status can be assigned.

2.8 Primers

The primers used in analysis of the NOD2/CARD15 gene, IBD5 haplotype and DLG5 gene are listed in the relevant chapters.

2.9 Haplotypes

Haplotypes were calculated using Haploview version 3.2.

2.10 Gels used for PCR reactions

2.10.1 Preparation

Multipurpose agarose (2.25g) was added to 150 millilitres (ml) of half strength TBE into a 250 ml beaker. This was then microwaved on full power for between 2 and 3 minutes agitating half way through until all of the agarose had dissolved into solution. The agarose was then cooled to around 50°C by running under a cold water tap, before 10 microlitres (µl) of ethidium bromide was added and then agitated further to ensure even distribution. The agarose solution was then poured into a transparent gel-casting tray, which was sealed at each end using autoclave tape. Four combs containing either 12 or 16 wells were then inserted into the gel. The gel was then left to cool and solidify at room temperature for at least 30 minutes.

2.10.2 Electrophoresis

The gel was removed from its tray and case removed before being placed into an electrophoresis tank filled with half strength TBE. Before loading the PCR mix in question 5 µl of orange G loading buffer had been added to 20 µl of PCR mix. The mixture was then pipetted into the wells in the gel before running the electrophoresis. The running time varied dependent on optimal conditions for the PCR reaction in question but a standard running time of 30 minutes and 150 watts of power was used for most gels.

2.10.3 Interpretation

After the gel had completed electrophoresis it was then removed from the tank and placed in a covered ultraviolet hood. Images were then recorded using a digital camera image loaded onto a designated computer. The results were then read for each sample manually and recorded onto the Access© database. Each gel was run with controls for homozygous wild type and mutant as well as a heterozygote. If any of the controls failed then the gel was not used for genotype analysis.

3 Phenotypic characteristics of childhood onset IBD: Comparison with adult onset disease using the Montreal classification

3.1 Summary

Aims: To describe phenotypic characteristics of childhood-onset compared to adult-onset IBD in Scottish patients.

Setting: 353 children with IBD (256 CD, 97 UC) were recruited from the three paediatric gastroenterology centres in Scotland. 1234 adults (617 CD, 664 UC) patients were recruited from the Western General Hospital, Edinburgh.

Methods: All patients were phenotyped using the Montreal classification. Phenotypic comparisons were then made between childhood and adult populations using the Montreal classification.

Results: Unless stated $p \leq 0.0001$. Childhood onset IBD was characterised by more familial disease (34% vs. 20% OR=2.1 [1.6-2.8]) and less smoking (1% vs. 18% OR=0.05[0.02-0.14]). Childhood onset CD featured a male predominance (60% vs. 38%, OR=2.5 [1.8-3.3]). More extensive Crohn's disease was present at diagnosis in children with more combined colonic/upper GI and more combined ileo-colonic/upper GI (L2+L4, 18% vs. 2% OR=14.3 [6.6-30.8] and L3+L4, 37% vs. 4% OR=16.2[9.6-27.3] respectively). There was also less isolated ileal and less isolated colonic disease (L1, 4% vs. 30% OR=0.09[0.04-0.17] and L2, 16% vs. 37% OR=0.3[0.21-0.46] respectively). There was less stricturing and penetrating disease at diagnosis (B2, 4% vs. 9% p=0.007 OR=0.38[0.18-0.79] and B3 p=0.02 OR=0.45[0.2-0.9] respectively). There were also less extra-intestinal manifestations (18% vs. 28%, p=0.004 OR=0.58[0.4-0.85]. Childhood onset UC was characterised by more extensive colitis (E3, 74% vs. 41%, OR= 4.1[2.5-6.7] and conversely less proctitis (E1, 4% vs. 21%, p=0.0001, OR=0.2(0.06-0.5).

Conclusions: Childhood IBD has distinct phenotypic characteristics in comparison to adult IBD. Children with CD were more likely to be male, have more extensive disease (more upper tract disease, less isolated ileal and colonic disease) and less stricturing/penetrating disease at diagnosis. Children with UC have more extensive disease at diagnosis. The reasons for the phenotypic differences will be explored by examining patient genotypes between the two populations in subsequent chapters.

3.2 Introduction

The only established phenotypic classification available at the start of the thesis was the Vienna classification, an established phenotypic classification for adults with Crohn's disease.[36] No phenotypic classification was available for children, and additionally there was no consensus classification available for Ulcerative Colitis in adults or children. [36] In devising the Vienna phenotypic classification for Crohn's disease, a working party was convened in Vienna in 1998 and then after reviewing the evidence from five IBD populations in Europe and North America; their joint findings were subsequently published in 2000. The consensus that emerged based a phenotypic classification of Crohn's disease on: age at diagnosis, disease location and disease behaviour as described in detail in chapter 2.

Subsequently, the classification system was refined taking into consideration some of the difficulties encountered in applying the Vienna classification to Crohn's disease populations while performing clinical studies. A working party met in Montreal in 2005 and published their revised findings in the same year.[35] This classification provided a disease categorization that could be used for patients with both Crohn's disease and Ulcerative colitis.

Could the adult classification therefore simply be applied to children? This is addressed in this study by first examining disease phenotype in children and then addressing whether the IBD phenotype is different between childhood and adulthood populations.

3.3 Materials and Methods

Patients were recruited from paediatric gastroenterology centres in Scotland and compared to adult onset patients recruited from the Western General Hospital, Edinburgh.

3.3.1 Montreal classification

In 2005 a second "adult" phenotypic classification was proposed providing a disease classification for both Crohn's disease and Ulcerative Colitis (tables 3-1 and 3-2 respectively). This newer *Montreal* classification addressed the shortcomings of the Vienna classification which included specific recognition of a separate early onset disease category (<17 years).[35] The outline of the classification provided in the following text will concentrate only on the modifications to the Vienna criteria.

3.3.1.1 Disease location

In addition to the four disease locations described in the Vienna classification (ileal, colonic, ileo-colonic and upper GI, L1-4) the Montreal classification added three further subcategories; these were the first three categories (L1-L3) with the addition of L4 where the disease sites coexisted, meaning the L4 disease location was no longer an exclusive disease location allowing for more in depth phenotypic analysis.

In patients with Ulcerative colitis disease extent was divided into three categories. The first category (E1) was for patients with a proctitis i.e. inflammation limited to the rectum, the second (E2) was for patients with left sided disease distal to the splenic flexure and the last category (E3) described patients with extensive disease proximal to the splenic flexure.

3.3.1.2 Disease behaviour

The classification proposes waiting 5 years post diagnosis before allocating a specific disease behaviour. This is because of the dynamic nature of Crohn's disease demonstrated in several adult studies.[148-151] There is also a suggestion of 10 years post diagnosis for studies examining genotype-phenotype relationships! However, within the proposed framework leeway is also given, in that investigators are free to adapt a time line for disease behaviour that is suitable to the individual study population.

Recognition is also given to the fact that enteric and perianal fistulae may represent different disease phenotypes. Thus perianal disease is removed from the B3 category and instead added as a suffix "p" to any of the disease categories, B1-3. Perianal disease behaviour is defined as the presence of perianal abscesses or fistulae. There are no published studies suggesting the best time line for changes in disease behaviour in paediatric CD. So, within the present study behaviour is only analysed at the time of diagnosis, when comparing disease in children compared to adults.

3.3.1.3 Age at disease diagnosis

There were now three age categories instead of the two in the Vienna classification. The original <40 years was divided into an age <17 years at diagnosis (A1) and a 17-40 years category (A2). The >40 years category remained and was known as A3 in the new classification.

3.3.2 Parental smoking

Details were also collected on parental smoking and in a limited number cases smoking data was available from control children and their parents. Pairs of cases (IBD) and controls were matched for sex, age (within 2 years) and geographical location (attending same GP practice).

Table 3-1: The Crohn's disease Montreal classification

Age at diagnosis			
A1	16 y or younger		
A2	17 – 40 years		
A3	Over 40 years		
Disease location			
L1	Terminal Ileum	L1 + L4	TI + Upper GI
L2	Colon	L2 + L4	Colon + Upper GI
L3	Ileocolon	L3+L4	Ileocolon + Upper GI
L4	Upper GI		
Disease behaviour			
B1	Non-stricturing, non-penetrating	B1p	Non-stricturing, non-penetrating + perianal
B2	Stricturing	B2p	Stricturing + Perianal
B3	Penetrating	B3p	Penetrating + Perianal

Table 3-2: The Ulcerative colitis Montreal classification

Disease extent	
E1	Proctitis
E2	Left sided (distal)
E3	Extensive (pancolitis)

3.4 Results

3.4.1 Childhood compared to adult disease

Data on 353 children with IBD (256 CD, 97 UC) were compared to 1234 adults (617 CD, 664 UC) with the results presented in table 3-3. The median age at diagnosis in children with IBD was 11.4 years (IQR 9.0-13.1 years) and for adults was 32.1 years (IQR 24.5-46.2 years).

3.4.2 Time to diagnosis

The time from the onset of symptoms to diagnosis of was available for 337 children (242 CD, 94UC) and 1042 adults (494 CD, 548UC). The median time to diagnosis was faster in patients with ulcerative colitis compared to Crohn's disease in both children (0.33 years vs. 0.58 years, p=0.0009) and adults (0.25 years vs. 0.50 years, p<0.0001). In adults with ulcerative colitis the median time to diagnosis was faster than children (p=0.01) but there was no difference between children and adults with Crohn's disease (p=0.26).

3.4.3 Sex

In children with CD there was a clear male predominance of disease with the opposite being true in adult onset disease (60% vs. 38%, p<0.0001, OR=2.5[1.8-3.3]). In patients with UC the disease was equally likely to occur in both sexes both in child and adulthood.

3.4.4 Smoking

Smoking was uncommon in children with IBD in contrast to adults (1% vs. 18%, p<0.0001, OR=0.05[0.02-0.14]). In adults, smoking was more common in those with a diagnosis of CD compared to those with UC (27.2% vs. 10.6%, p<0.0001, OR = 3.13 [2.29-4.29]).

In total, 62 pairs of paediatric IBD cases and paediatric controls were compared. Parental smoking during pregnancy and around the time of birth was more common

in parents of IBD cases compared with control parents (54% vs. 29%, p=0.01; OR=2.87 [1.23–6.66]). Maternal smoking during pregnancy and at birth was also more common in the parents of IBD cases and CD cases than in parental controls (23% vs. 6.2%, p=0.04; OR=4.46[1.16–17.1]) and (27.8% vs. 8.3%, p=0.03; OR=4.23[1.05–16.97]). There was no significant effect seen when paternal smoking in pregnancy and at birth was analysed in IBD cases versus controls (p=0.27).

3.4.5 Family history

A family history of IBD was more common in children than in adults (33.9% vs. 19.6%, p<0.0001, OR=2.1[1.6-2.8]).

3.4.6 Disease Location

There was more extensive "panenteric" disease present in children compared to adults demonstrated by significantly more upper GI tract disease in children in combination with colonic or ileo-colonic disease than was present in adult patients, (L2+L4, 18% vs. 2%, p<0.0001 OR=14.3[6.6-30.8] and L3+L4, 37% vs. 4%, p<0.0001 OR=16.2[9.6-27.3]respectively). This data was complemented by there being significantly less isolated ileal disease and less isolated colonic disease in children compared to adults (L1, 4% vs. 30%, p<0.0001, OR=0.09[0.04-0.17], L2 16% vs. 37%, p<0.0001 OR=0.3[0.21-0.46] respectively. Isolated upper GI tract disease was uncommon in both children and adults. There were 12/52 (23.1%) patients with isolated colonic disease (L2) under 8 years of age compared to 29/218 (13.3%) diagnosed at 8 years and older showing a strong trend towards statistical significance (p=0.07, OR=1.96 [0.92-4.16]).

In patients with a diagnosis of UC (table 3-4) extensive colitis was more common in children: (E3, 74% vs. 41%, p<0.0001 OR=4.1[2.5-6.7]). Contrastingly left sided disease and proctitis were less common (E2, 21.9% vs.38.7%, p=0.001 OR=0.44 [0.27-0.74] and E1, 4% vs. 21%, p<0.0001 OR=0.2[0.06-0.5] respectively).

3.4.7 Disease behaviour

Inflammatory disease combined with perianal disease (B1p) was more common in children compared to adults (11.5% vs. 6.9%, p=0.03 OR=1.77 [1.05-2.97] with inflammatory disease alone (B1) showing a strong trend towards significance (80.2% vs. 73.8%, p=0.05 OR=1.43 [0.99–2.07]). Contrastingly stricturing (B2) and penetrating disease (B3) were less common in children at presentation compared to adults (B2, 4% vs. 9%, p=0.007 OR=0.38[0.18-0.79] and B3 4% vs. 8%, p=0.02 OR=0.45[0.2-0.9]).

3.4.8 EIMS

Extra intestinal Manifestations (EIM's) of IBD (skin, liver, eye and joint EIM's combined) were less common in children compared to adults (18% vs. 28%, p=0.004 OR=0.58[0.4-0.85]) at presentation.

Table 3-3: The presenting CD phenotype in children compared to adults

	Children	**Adults**	**p value**	**Odds Ratio**
Male: female	155/101	236/381	<0.0001	2.47 (1.83-3.33)
Smoker	5/252 (2.0%)	145/534 (27.1%)	<0.0001	0.05 (0.02-0.13)
Family history	89/250 (35.6%)	126/588 (21.4%)	<0.0001	2.03 (1.46-2.81)
Disease location				
L1	9/249 (3.6%)	158/526 (30.0%)	<0.0001	0.09 (0.04-0.17)
L2	39/249 (15.7%)	195/526 (37.1%)	<0.0001	0.32 (0.21-0.46)
L3	53/249 (21.3%)	109/526 (20.7%)	0.85	
L4	3/249 (1.2%)	15/526 (2.8%)	0.15	
L1+L4	6/249 (2.4%)	22/526 (4.2%)	0.21	
L2+L4	45/249 (18.1%)	8/526 (1.5%)	<0.0001	14.28 (6.62-30.83)
L3+L4	94/249 (37.7%)	19/526 (3.6%)	<0.0001	16.18 (9.57-27.35)
Disease behaviour				
B1	202/252(80.2%)	366/496 (73.8%)	0.05	1.43 (0.99–2.07)
B1p	29/252 (11.5%)	34/496 (6.9%)	0.03	1.77 (1.05-2.97)
B2	9/252 (3.6%)	44/496 (8.9%)	0.007	0.38 (0.18-0.79)
B2p	2/252 (0.8%)	1/496 (0.1%)	0.22	
B3	10/252 (4.0%)	42/496 (43.7%)	0.02	0.45 (0.22-0.91)
B3p	0	9/496 (1.8%)	-	
Bp	31/252 (12.3%)	44/496 (8.9%)	0.13	

Table 3-4: The presenting UC phenotype in children compared to adults

	Children	**Adults**	**p value**	**Odds Ratio**
Male: female	48/49	329/335	0.99	
Smoker	2/96 (2.1%)	69/649 (10.6%)	0.007	0.18 (0.04-0.74)
Family history	28/95(29.4%)	116/649 (17.9%)	0.007	1.92 (1.18-3.12)
Disease extent				
E3	71/96 (74.0%)	232/569 (40.8%)	<0.0001	4.13 (2.54-6.70)
E2	21/96 (21.9%)	220/569 (38.7%)	0.001	0.44 (0.27-0.74)
E1	4/96 (4.1%)	117/569 (20.6%)	0.0001	0.17 (0.06-0.47)

3.5 Discussion

This most notable finding of this study is that it has demonstrated clear differences between children and adults with IBD in terms of disease phenotype. There are clear sex differences between children and adults with CD, being a male predominant disease in childhood and a female predominant disease in adulthood as has been confirmed again in this study. Kugathasan et al demonstrated in a North American CD population (<18 years at diagnosis) 62% of 129 patients studied being male.[40] Sawczenko et al. described the same male predominance of 62% in a study of 431 UK CD patients (diagnosed <16 years).[39] A number of further paediatric studies have also replicated this finding.[7;9;152;153] The only study to report the opposite trend was from Canada where a *"paediatric"* population (under the age of 20 at diagnosis) demonstrated a female predominance in 57% of the 224 cases described.[38]

The transition between male and female dominated Crohn's disease seems to happen in most studies in the late teenage years causing speculation about both environmental and genetic influences at this age. The increased use of the oral contraceptive pill at this age has been suggested as a possible contributory environmental agent.[11] Godet et al. performed a meta-analysis combining the results of two cohort studies (30,379 unexposed and 30,673 exposed patients to oral contraceptive pill) and seven case-control studies involving (482 CD, 237 UC, and 3198 controls) after excluding studies of poor methodological quality.[154] The results demonstrated a modest increased risk for the development of both CD and UC but only CD being statistically significant (relative risk (RR) =1.44 (95% CI 1.12- 1.86 and RR=1.29 (95% CI 0.94-1.77) for CD and UC respectively.

This age/sex effect could also be influenced by genetic factors. Fisher et al. analysed the IBD3 locus by sex of the patient and demonstrated IBD3 is male sex specific locus.[128] There also appears to be a male specific effect in studies of the DLG5 gene, as discussed in more detail in chapter 7.[155;156] Interest in the role of the X chromosome in IBD susceptibility has been generated from an association with Turner's syndrome and IBD.[157]

There also appear to be distinct differences in disease location between children and adults, with children having more extensive disease. This study has highlighted that when the Montreal classification is applied to childhood and adult cohorts stark contrasts exist in disease location, with less isolated ileal and less isolated colonic disease but more upper GI tract disease in children combined with other disease sites compared to adults. In a large study in the Canadian CD population, Freeman demonstrated upper GI tract disease as defined by the Vienna classification was significantly more common in a cohort of 224 children (age <20years) compared to a cohort of 622 adults (20-39 years old) at 18% vs. 10 % respectively.[38]

It is unclear how much of this recognized difference in disease phenotype between children and adults results from differences in investigations used to diagnose IBD and how much represents a *true* phenotypic difference. Children in Europe at IBD diagnosis are now recommended to be investigated by upper GI endoscopy, ileo-colonoscopy and small bowel follow through.[158] Although these criteria defined by a working group of European paediatric gastroenterology, Hepatology and nutrition society (ESPGHAN) that met over several years in Porto and have only recently been published, this model of investigation has been employed by many paediatric gastroenterologists over many years, prior to the publication of these "official" guidelines. By contrast, adult patients will often undergo a more limited examination, which often omits an upper GI endoscopy,[159] despite published recommendations.[160]

Thus the well recognized increase in upper gastrointestinal disease in paediatric patients may be partly artifactual due to: differences in the investigation of patients, publication bias with a limited amount of published data available or may represent a true disease distinction. The true differences can only be addressed in population wide prospective studies employing the same methods of diagnostic investigations in patients of all ages. This study has demonstrated however that the differences in phenotype are not explained in patients with Crohn's disease by a difference in time from symptom onset to diagnosis.

This study data is limited in the fact comparisons are only made at disease diagnosis, but other studies have demonstrated disease location is a relatively stable phenotype with little change over time. Louis et al. followed 125 adult patients with Crohn's

disease and found at 10 years following diagnosis only 15% had changed disease location (location was defined using the Vienna classification).[150] In a prospective study of Norwegian adult patients followed up at 5 years post diagnosis, Henriksen et al. demonstrated disease location defined using the Vienna classification had changed in 13.5% of patients.[161]

In this current study inflammatory disease was more common in children compared to adults at diagnosis. There are no prospective studies in the paediatric population that have studied the evolution of disease behaviour over an extended period of time in paediatric Crohn's disease. In this study, a significant increase in the frequency of penetrating disease was demonstrated at 2 years post diagnosis (chapter 2). The study population examined in this thesis would need to be followed up for at least 5 years and probably 10 years to get a true indication of the rate of disease progression in paediatric Crohn's disease based on published data from adult studies.

Interestingly, in a study of disease behaviour in 2000 CD patients Cosnes et al. demonstrated 40% had penetrating disease as defined by the Vienna criteria at 5 years and 70% by 20 years.[14] Contrasting with the findings of the present study where adults had a higher rate of penetrating disease at diagnosis, the authors also noted penetrating disease was more common in disease diagnosed at a younger age (<40 years at diagnosis) although this population was not stratified into further age groups. Louis et al. studied 297 CD patients again using the Vienna classification and demonstrated disease behaviour had progressed significantly by 1 year post diagnosis.[150] This study did not show any effect of age at diagnosis.

It is now well recognized that rates of pancolitis, in patients with Ulcerative colitis, demonstrated in prospective childhood studies (80-90%)[39-41] are far higher than those documented from prospective adult studies (24% pancolitis[42] or 33% *extensive colitis*[43]). This study provides further replication of this data although the data was not collected prospectively it is the first study to directly compare a child and adult population using the Montreal classification. Sawczenko et al. demonstrated a pancolitis rate of 81% in 211 UK UC patients and Kugathasan et al, a rate of 90% in 60 UC patients from Wisconsin.[39;40] In the study from Heyman et al it was noted the very youngest children (diagnosed < 5 years of age) had the highest rates of

pancolitis at 96%.[41] These data suggest for Ulcerative colitis at least, that a true difference in phenotype *really does* exist.

There is also proxy evidence that UC starting in childhood is more severe disease phenotype. Eaden et al. in a meta-analysis of 116 studies examining UC and colo-rectal cancer demonstrated a doubling in the rate of colorectal cancer complicating UC, independent of the increased risk conferred from pancolitis with a risk of 15.7% at 30 years post diagnosis in children compared to 8.7% in adults.[162]

Within childhood onset disease itself, very early onset IBD (age<8years old) appears to be a colonic predominant disease. This current study has mirrored the data for colonic Crohn's disease, demonstrating that disease at diagnosis in children less than 8 years, defined using the Montreal classification was more common than in children more than 8 years old at diagnosis (although did not quite reach statistical significance). In published studies to date almost all under 5 year olds will have colonic only disease, with a higher proportion of UC and IC than children diagnosed at an older age.[41;163] In a very large study of nearly 1400 North American early onset patients, Heyman et al demonstrated by multivariate analysis that a colonic predominant phenotype exists in IBD diagnosed under the age of 8 years.[41] In parallel studies from Europe, Meinzer et al. demonstrated that the acquisition of ileal Crohn's disease is an age dependent phenomenon that is rare at age 8 years and becomes increasingly more common year on year towards the age of 16 years.[164] This suggests a colonic predominant phenotype in children who are diagnosed in early childhood which then changes throughout the childhood years as children grow older.

Using the Montreal classification does impose limitations in the anatomical sites of disease that can be compared. As the *individual* disease locations in this study for Crohn's disease (chapter 2) fit generally with those published by other groups they would appear to be typical of a paediatric onset Crohn's disease population. Therefore if a more extensive disease location categorisation was used the distinct differences between childhood and adult disease may become more apparent. In the UK prospective survey performed by Sawczenko et al. of 167 CD patients with details of disease phenotype at diagnosis all the disease location categories had very

similar proportions to those reported in this study.[39] The only exception was a decrease in the frequency of disease moving distally through the colon culminating in rectal disease in 49% of patients.

It would seem that clearly the adoption of an adult classification for the many reasons outlined in this chapter would not be appropriate for children, so a separate classification is needed. This question was previously highlighted by a working group at the first world congress of paediatric gastroenterology that met in Boston, USA in 2000 when it was conceded that there was no recognised disease classification system for children with IBD but that one was needed.[37] The hierarchical disease classification adopted in the Vienna classification would mean most paediatric patients, based on data from previous studies would have a disease location that fell into the upper GI (L4) category, potentially disguising any underlying genotype phenotype relationships.[36] Although the Montreal classification has improved on the "benchmark" classification proposed by the working group in Vienna it still does not address some of the specific issues relating to disease phenotype in children. However despite these limitations the Vienna and Montreal classifications allow uniformity of reporting for genotype–phenotype analysis in paediatric genetic studies in the interim period prior to the development of a widely accepted paediatric phenotypic classification.

In conclusion, disease phenotype in children differs markedly from patients who are diagnosed in adulthood. The possibility that this can be explained by a patient's genotype will be explored in subsequent chapters.

4 NOD2/CARD15: The influence of the three common variants on disease susceptibility and phenotype

4.1 Summary

Aims: To investigate the contribution of the three common NOD2/CARD15 mutations - R702W, G908R and Leu1007finsC to disease susceptibility and disease phenotype in the Scottish early-onset IBD population.

Subjects: 906 individuals including 247 Scottish IBD patients aged <16 years at diagnosis, 414 parents and 245 healthy controls.

Methods: Genotyping for the R702W and Leu1007finsC was performed by TaqMan and G908R by Restriction Fragment Length Polymorphism (RFLP). Transmission disequilibrium testing (TDT), case-control analysis and detailed genotype-phenotype analysis were then performed.

Results: The Leu1007finsC variant was associated with susceptibility to CD by case-control analysis (4.2% vs. 1.4% respectively, p=0.01) and TDT analysis (χ^2=4.42, p=0.006). The Population Attributable Risk (PAR) for the 3 NOD2/CARD15 mutations was 7.9%.Unifactorial analysis of NOD2/CARD15 carriage demonstrated carriage of NOD2/CARD15 variants was associated with, at diagnosis: decreased albumin (31.0% vs. 9.0%, p=0.001) and raised CRP (25% vs. 9.5%, p=0.04) and at follow up with: need for surgery (39.5% vs. 12.8%, p=0.0002) jejunal involvement (50% vs. 18.4%, p=0.01) jejunal and ileal involvement (50% vs. 10.7%, p=0.009), raised CRP (57.1% and 12.8%, p=0.0009), lower weight/height centile (75.0% vs. 20.2 %, p=0.03, 50.0% vs. 16.0 %, p=0.001 respectively) and stricturing disease (45.5% vs. 19.4%, p<0.05). Multifactorial analysis demonstrated variant carriage was associated with the need for surgery (p=0.004, OR 4.9 [1.5-14.7]).

Conclusions: The three common NOD2/CARD15 variants have a relatively small contribution to CD susceptibility in Scottish children but a major impact disease on phenotype. In particular NOD2/CARD15 variants are strongly associated with several markers of disease severity in paediatric CD, most notably with the need for surgery.

4.2 Introduction

4.2.1 IBD1

The aetiology of IBD is complex with strong supporting evidence for the importance of interaction of environmental agents and genetic factors in disease pathogenesis.[76] Genome wide scanning has led to the identification of several confirmed loci (IBD 1-9) as discussed in detail in chapter 1;[165] of these loci further data suggest that IBD1 and IBD5 loci may be particularly important in early onset disease.[99] As discussed in Chapter 1 direct comparison in non-Jewish and Jewish North American populations suggested that the IBD1 locus may confer a high risk particularly in early onset cases compared to late-onset disease in the same population.[93;94]

4.2.2 NOD2/CARD15

4.2.2.1 Gene Discovery

One of the major successes in IBD genetics in the 21st century has been the narrowing of the IBD1 region, leading to gene identification. The first susceptibility gene for CD was identified in 2001, on chromosome 16 and was initially named Nucleotide Oligomerisation Domain (NOD)2 and later renamed Caspase Activating Recruitment Domain (CARD) 15 (MIM 605956).[166;167] Hugot et al. in their seminal publication in *Nature* narrowed the region of linkage in the IBD1 locus using a collection of 26 densely packed microsatellite markers spaced at 1 centimorgan intervals in the pericentromic region of chromosome 16. An area within this region between 2 of these markers had the highest LOD score of 3.49 and the investigators then went on to construct a physical map of this narrowed region. Within the region one marker SNP (D16S3136) located within the NOD2 gene demonstrated borderline association with CD in TDT analysis of 108 family trios. Association with a second allele within the NOD2 gene was replicated in a second set of 76 families. Further analysis of the gene was then undertaken by genotyping 11 SNPs located throughout the gene together with 2 marker SNPs identified in construction of the physical map of the region, in 235 CD families. Several SNPs (numbered 8, 12 and

13 in this original publication) then demonstrated significant association when analysed by disequilibrium testing.

From the results of the case-control analysis (457 CD patients, 159 UC and 103 healthy controls) three common polymorphisms were identified by Hugot et al. that conferred increased susceptibility to CD: two missense mutations Arg702Trp and Gly908Arg and a frameshift mutation Leu1007fsincC that were named SNP 8, 12 and 13 respectively (see figure 1). The relative risk of CD increased for heterozygotes by a factor of 2-3 whereas homozygotes or compound heterozygotes had a 28-40 times increased risk.[166] There was no increased risk of Ulcerative colitis in patients possessing these NOD2/CARD15 mutations.

4.2.2.2 Concurrent replication

Ogura et al. chose NOD2/CARD15 as a positional candidate based on its position within the IBD1 locus and as a functional candidate gene based on the description of NOD proteins as being important in the recognition of bacteria.[168] In the first stage of their study, Ogura et al. sequenced the NOD2/CARD15 gene in 12 patients with Crohn's disease and 4 healthy controls. In three of the 12 patients they identified a frameshift mutation that truncated the NOD2/CARD15 protein from 1040 amino acids to 1007 (this was SNP 13 described in the Hugot et al publication).[167] In further analysis they then went onto further genotyping of this mutation in family trios containing an affected child with CD, demonstrating preferential transmission by TDT analysis. This finding was then replicated in 416 CD cases and 287 population controls demonstrating the NOD2/CARD15 frameshift mutation was associated with increased susceptibility to CD with a similar effect seen in both Jewish and non-Jewish patients.

4.2.2.3 Association with disease phenotype

In subsequent studies by Ahmad et al. it was suggested that *earlier onset* disease was characterised by a higher rate of NOD2/CARD15 mutations especially those carrying

double dose mutations. These patients carrying double dose mutations were diagnosed with Crohn's disease at age 23 years compared to those without double dose mutations who were diagnosed at 28 years of age (a statistically significant difference).[132] This study also described an association between NOD2/CARD15 variants and ileal Crohn's disease.

In a summary of the early NOD2/CARD15 publications Hugot described patients who were homozygous for the three common NOD2/CARD15 mutations as representing 34% of those with an age of onset of CD at less then 10 years, compared with only 3% who had an onset of 40 years or more,[169] justifying the need to conduct studies of these NOD2 /CARD15 mutations in exclusive paediatric populations.

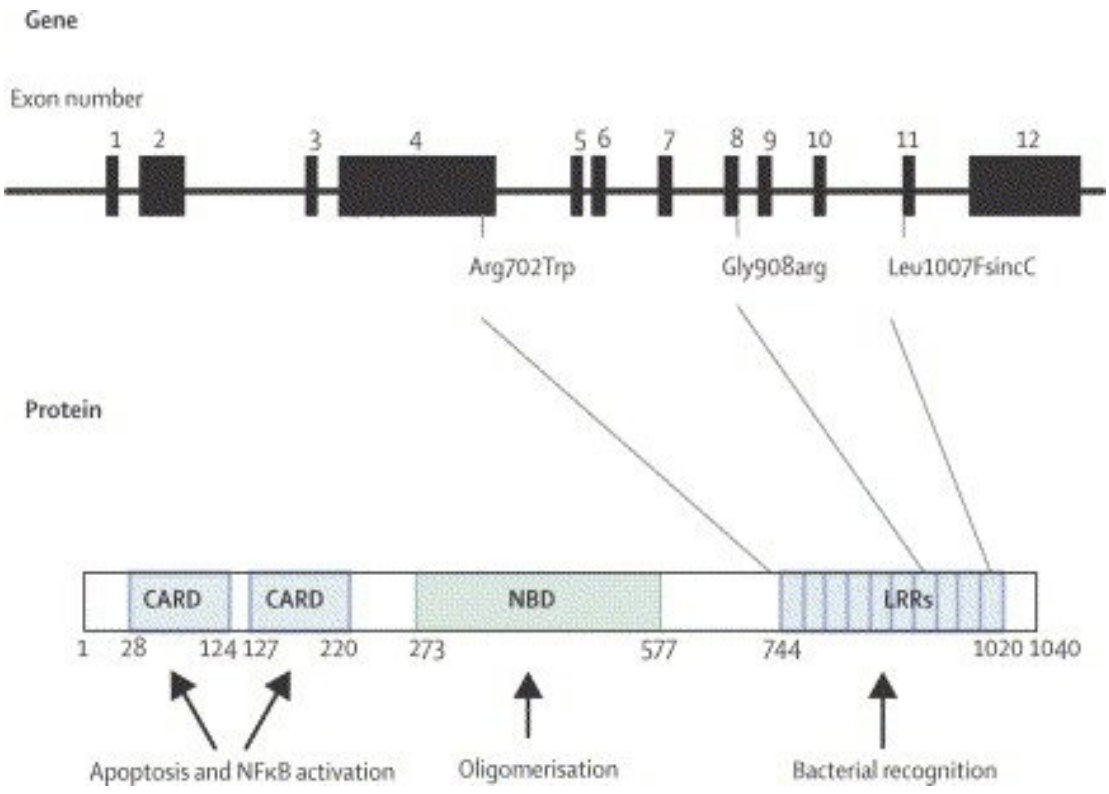


Figure 4-1: Structure of the NOD2/CARD15 gene and protein

4.2.2.4 Expression

In original experiments using a HEK293T kidney cell line Ogura et al, described NOD2/CARD15 as being expressed exclusively in monocytes, unlike family members NOD1 and APAF -1 which were widely expressed in most adult tissues.[170] However, subsequent studies from the same research group and replicated by others clearly show that NOD2/CARD15 is also expressed in primary epithelial cells in culture and intestinal epithelial cells (IEC) in the small and large bowel.[171;172] Lala et al. demonstrated NOD2/CARD15 mRNA is expressed in Paneth cells both in healthy controls and patients with CD, but with higher concentrations found in CD patients.[173]

Paneth cells are specialised epithelial cells located in the crypts of the small intestine and secrete a number of anti-bacterial substances in response to Muramyl dipeptide (MDP) a peptidoglycan motif commonly found in bacterial cell walls, and other bacterial products, most notably they secrete the defensins.[174]

Expression of NOD2/CARD15 is upregulated by Tumour Necrosis Factor-alpha (TNF-α) and interferon- gamma (IFN-γ), but IFN-γ has no effect in the absence of TNF-α.[172] Both cytokines act directly on NF-κβ making it more response to Lipopolysaccharide (LPS). The TNF-α response is dependent on NF-κβ binding sites in the NOD2/CARD15 promoter region and deletion of two of these binding sites results in diminished TNF-α activity.

4.2.2.5 Structure

The NOD2/CARD15 protein contains 1040 amino acid residues, separated into 3 functional domains: two N-terminal CARDs which are involved in protein – protein interactions, a central nucleotide binding domain (NBD) which mediates self oligomerisation and at the C-terminus a leucine rich repeat region (LRR), which interacts with MDP (figure 4-1).[132;175]

NOD2/CARD15 is a member of a family of intracellular cytosolic proteins NOD-like receptors that act as pattern-recognition receptors (PRR's) which recognise microbial components, also called pathogen-associated molecular patterns (PAMP's).[176] Other

family members are NOD1/CARD4 (MIM 605980) and Apoptotic protease activating factor (APAF) 1 (MIM 602233). Despite functional similarities, NOD1/CARD4 mutations have not been consistently been shown to be involved in susceptibility to CD.[177]

The NOD family share structural homology with R proteins in plants and Toll like receptors in mammals which form part of the innate immune response to bacterial products. NOD2/CARD15 shares sequence homology in the NBD and LRR domains with a wider family of around 20 proteins that are involved in apoptosis and inflammation collectively known as the CARD transcription enhancer, R [purine]-binding, pyrin, lots of leucine repeats (CATERPILLER) family.[178]

4.2.2.6 Function

NOD2/CARD15 is involved in the activation of nuclear factor-kappa beta (NF-κβ) within the cell after bacterial products bind to the LRR.[170] Specifically, it is the MDP moiety, the minimal essential structure of bacterial peptidoglycan (PGN) that is recognised by NOD2/CARD15 pathway.[175;179] The NOD proteins combine with the serine threonine kinase receptor-interacting CLARP- associated kinase (RICK) to activate NF-κβ.[180] Wild type NOD2/CARD15 can be activated by either Lipopolysaccharide (LPS) or peptidoglycan (PGN) in transfected cells to activate NF-κβ.[181] The cell surface receptor, TLR2 is activated by peptidoglycan in the extra-cellular space, resulting in NF-κβ activation.[182]

All three of the common NOD2/CARD15 mutations identified result in the same functional consequences, with marked hypo-responsiveness of NF-κβ to LPS or PGN compared to wild type NOD2/CARD15.[181] The R702W and G908R mutations show diminished activity of NF-κβ in response to LPS/PGN stimulation whereas the Leu1007fs mutation shows no activity at all. These experiments suggest a uniform loss of function as their common mode of action which specifically, for the Leu1007fsincC mutation in Caco2 cell line results in an inability to clear Salmonella typhimurium from cells in contrast to wild type NOD 2/CARD15.[183]

In this study we have investigated the contribution of the three common NOD2/CARD15 variants in the Scottish paediatric IBD population and examined genotype-phenotype relationships.

4.3 Patients and Methods

4.3.1 Patients

Two hundred and forty seven paediatric patients with IBD were studied. There were 167 children with CD, 60 with UC and with 20 IC.

4.3.2 Parents and controls

414 parents (232 mothers and 182 fathers) were analysed resulting in 180 complete family trios (73% of total). A total of 245 healthy Scottish adult controls were available for case-control analysis, with the detailed characteristics of this group reported previously by Arnott et al.[184]

Of the 167 CD patients in the study 91(54.5%) were male, 163(97.3%) were Caucasian, 57(36.1%) had a positive family history and 3 (1.9%) were current smokers. The median age at CD diagnosis was 11.5 years.

4.3.3 Disease Phenotype

CD disease location and behaviour using the Vienna classification are shown in table 4-1. Data using the paediatric phenotype developed for use in this thesis are shown in table 4-3. The data shown are for 167 CD patients analysed at diagnosis and 104 patients at 2 year follow up, unless stated in the tables.

4.3.4 Blood parameters

The results of blood results for patients at diagnosis and 2 year follow up are shown in table 4-2.

Table 4-1: Vienna classification of CD patients

Disease location	**At Diagnosis**	**2 year follow up**
Terminal Ileum (L1)	8 (4.8%)	5 (4.8%)
Colon (L2)	38 (22.8%)	28 (26.9%)
Ileocolon (L3)	34 (20.3%)	22 (21.2%)
Upper gastrointestinal (L4)	77 (46.1%)	45 (43.3%)
None*	10 (6.0%)	4 (3.8%)
Disease behaviour		
Inflammatory (B1)	136 (81.4%)	66 (63.5%)
Stricturing (B2)	7 (4.2%)	11 (10.6%)
Penetrating (B3)	24 (14.4%)	27 (25.9%)

Table 4-2: Blood results in CD patients at diagnosis and follow up

Blood Parameters	**At diagnosis**	**2 year follow up**
Raised ESR	121/137 (88.3%)	41/63 (65.1%)
Thrombocytosis	100/152 (65.8%)	26/75 (34.7%)
Raised CRP	76/118 (64.4%)	14/53 (26.4%)
Hypoalbuminaemia	71/138 (51.4%)	9/68 (13.2%)
Anaemia	112/164 (68.3%)	49/84 (58.3%)

Table 4-3: Paediatric disease location for CD patients

Disease Location	At diagnosis	2 year follow up
Oral	20/152 (13.6%)	14/98 (14.3%)
Oesophageal	4/142 (2.8%)	1/90 (1.1%)
Gastric antrum	39/145(26.9%)	12/84 (14.3%)
Gastric body	30/138 (21.7%)	18/94 (19.1%)
Duodenal	23/137 (16.8%)	9/88 (10.2%)
Jejunal	22/136 (16.2%)	14/90 (15.6%)
Ileal	91/142 (64.0%)	63/94 (67.0%)
Caecal	91/119 (76.5%)	58/79 (73.4%)
Ascending	92/122 (75.4%)	56/74 (75.7%)
Transverse	99/129 (76.7%)	60/79 (75.9%)
Descending	104/134 (77.6%)	61/81 (75.3%)
Sigmoid	115/147 (78.2%)	67/92 (72.8%)
Rectum	117/151 (77.5%)	67/93 (72.0%)
Perianal	66/158 (41.8%)	37/100 (37.0%)

4.3.5 Genotyping

The NOD2/CARD15 mutations R702W and 1007fsInsC were genotyped using TaqMan (© Biosystems) using assay on demand methodology. The G908R was analysed using Restriction Fragment Length Polymorphism (RFLP). The primers used for the genotyping are shown in table 4-4.

Table 4-4: Primers used to genotype common 3 NOD2/CARD15 mutations

SNP	Primers
R702W (rs2066844)	Forward: CCAGACATCTGAGAAGGCCCTGCTC Reverse: GGCGCCAGGCCTGTGCCCGCTGGTG
G908R (rs2066845)	Forward: AGGCCACTCTGGGATTGAG Reverse: GTGATCACCCAAGGCTTCCAG
1007fsInsC (AJ303140)	Forward: GCTGAGTGCCAGACATCTGAGA Reverse: CACCAGCGGGCACAGG

4.3.5.1 G908R RFLP

The PCR reaction was carried out in final volume of 25µl with the reaction conditions as follows for a total of 60 reactions: 150µl of 10x PCR buffer (160mM $(NH_4)_2SO_4$, 670mM Tris- HCl, 0.1% Tween-20), 150µl of each primer (10mM), 45µl 50mM $MgCl_2$, 15µl 25mM dNTPs, 19.5µl Taq polymerase (5U/µl) (Bioline) and 970µl of water. 1µl (100nanograms) of DNA was added to each 25µl reaction.

The reaction was then run on a programme of 5 minutes 94°C, 30 seconds 92°C, 30 seconds 60°C, 45seconds at 72°C and 35 cycles of 5minutes at 72°C.

The PCR reaction was digested overnight at 37°C after the addition of 1 unit of the restriction enzyme *Cfo*I and 3 µl of manufacturers' restriction buffer plus 2 µl of water.

The PCR product was then run on a 4% NuSieve 3:1 agarose gel (6g of gel added to 150 of 1 x TBE) with other methods as described as in the standard gel preparation listed in chapter 2. The PCR reactions were loaded onto the gel with a control wild type homozygote and control heterozygote. No homozygous mutant was available to run on the gels. If the controls were not identified when viewing the gel then the results were declared void and re-run. The samples were then were then subject to gel electrophoresis for 25 minutes (150 Volts, 150 milliamperes, 150 Watts).

Each variant allele was analysed for association with CD by case-control study and transmission disequilibrium testing. Genotype-phenotype analysis was performed by correlating NOD2/CARD15 carriage rates with disease phenotype, both at diagnosis and at 2 year follow-up. All data derived from patients and controls were consistent with Hardy-Weinberg equilibrium.

4.4 Results

4.4.1 Disease susceptibility

4.4.1.1 Allele frequencies

The results are summarised in table 4-5. Allelic frequencies for Leu1007fsInsC were higher in CD patients compared to both UC patients (4.2% vs. 0%, p=0.02 OR=∞) and healthy controls (4.2% vs. 1.4%, p=0.01 OR 3.0 [1.2-7.6]). Allelic variants of G908R were significantly higher in CD patients compared to healthy controls (2.2% vs. 0.2%, p=0.006 OR 10.5(1.3-85.5]). No significant difference was demonstrated in the allele frequencies of R702W in CD patients compared to either UC patients or healthy controls.

4.4.1.2 Carriage rates

The results are summarised in table 4-6. In total, one or more of NOD2/CARD15 allelic variants were present in 33 (19.8%) of the 167 CD patients, with 27 (16.2%) simple heterozygotes and 6 (3.6%) double dose carriers (2 homozygotes, 4 compound heterozygotes). This carriage rate in CD patients was higher than UC patients (19.8% vs. 8.3%, p=0.04 OR 2.7[1.0-7.3]) and demonstrated a trend towards higher carriage than in healthy adult controls (19.8% vs. 13.1%, p=0.07).

Table 4-5: NOD2/CARD15 variant allele frequencies in cases and controls

	R702W	**G908R**	**Leu1007fsInsC**	**Combined**
CD (n=167)	6.2%	2.2%	4.2%	12.6%
UC (n=60)	4.2%	0%	0%	4.2%
p-value*	0.41	0.11	0.02	0.01
Odds Ratio*	1.52 (0.55-4.2)	∞	∞	3.31 (1.27-8.61)
HC (n=245)	5.5%	0.2%	1.4%	7.1%
p-value**	0.61	0.006	0.01	0.007
Odds Ratio**	1.17 (0.63-2.16)	10.47(1.28-85.48)	3.02 (1.21-7.56)	1.91 (1.18-3.09)

* CD vs. UC; ** CD vs. HC

Table 4-6 NOD2/CARD15 variant carriage rates in cases and controls

	R702W	**G908R**	**Leu1007fsInsC**	**Combined**
CD (n=167)	11.3%	3.0%	8.4%	19.8%
UC (n=60)	8.3%	0%	0%	8.3%
p-value*	0.72	0.18	0.02	0.04
Odds Ratio*	1.44 (0.46-4.54)	∞	∞	2.71(1.00-7.30)
HC (n=245)	9.8%	0.4%	2.9%	13.1%
p-value**	0.63	0.03	0.01	0.07
Odds Ratio**	1.18 (0.60-2.30)	7.53 (0.87-65.06)	3.11 (1.23-7.88)	1.64 (0.96-2.79)

* CD vs. UC; ** CD vs. HC

4.4.1.3 Transmission disequilibrium testing

The results are summarised in table 4-7. Transmission disequilibrium testing demonstrated significant over transmission of the Leu1007fsInsC allele (χ^2=4.4, p=0.006). The R702W and G908R alleles were not significantly higher in CD patients when analysed by TDT analysis. The TDT analysis for the 3 common NOD2/CARD15 variants was performed in a total of 156 families of children with a diagnosis of Crohn's disease.

The overall population attributable risk for NOD2/CARD15 carriage in Crohn's disease patients was 7.9%.

Table 4-7: TDT results for 3 common NOD2/CARD15 variants

Mutation	Observed[1]	Expected[2]	Variance[3]	Chi-square	p value
R702W	18	15.63	6.50	0.86	0.42
G908R	7	6.5	2.98	0.08	0.72
1007fsInsC	14	9.80	3.96	4.42	0.006

1. Observed transmissions to affected offspring.

2. Expected transmissions to affected offspring based on Mendelian Inheritance patterns.

3. Variance of observed–expected transmissions.

4.4.2 Genotype- phenotype analysis: Unifactorial analysis

4.4.2.1 Age at diagnosis

The median age at diagnosis was 11.0 years in carriers of NOD2/CARD15 variant alleles and 11.3 years in non-carriers (p=0.38).

4.4.2.2 Family history

A history of IBD in first degree relatives was more common in children possessing a mutant copy of the G908R allele (9.1% vs. 1.4%, p=0.002, OR=7.0 [1.7-29.1]) but not for the other two NOD2/CARD15 alleles studied. Of those with a positive family history, carriage of one or more of the 3 NOD2/CARD15 variants was higher but did not reach statistical significance (26.3% vs. 16.8%, p=0.15).

Table 4-8: Genotype – phenotype analysis using the Vienna Classification

	*At diagnosis**	*p-value*	*2 year follow up*	*p-value*
Disease location				
Colonic	15.8% vs. 20.9%	0.48	10.7% vs. 26.7%	0.08
Ileal	50% vs. 18.2%	0.05	60% vs. 20.4%	0.07
Ileo-colonic	17.6% vs. 20.3%	0.73	22.7% vs. 22.2%	0.96
Upper GI	20.1% vs. 18.9%	0.76	26.7% vs. 19.0%	0.35
None	10% vs. 20.4%	0.42	0% vs. 23%	1
Disease behaviour				
Inflammatory	16.9% vs. 32.2%	0.05	16.7% vs. 31.6%	0.07
Stricturing	42.9% vs. 18.7%	0.14	45.5% vs. 19.4%	0.05
Penetrating	29.2% vs. 18.1%	0.21	25.9% vs. 20.8%	0.58

* Comparison of carriage of one or more NOD2/CARD15 variants in each parameter under consideration vs. carriage of those without the parameter under consideration.

Table 4-9: disease location and NOD2/CARD15 carriage

	At diagnosis	*p-value*	*2 year follow up*	*p-value*
Disease location				
Oral	5.0% vs. 22.7%	0.08	7.1% vs. 25.0%	0.18
Oesophageal	0% vs. 21.0%	0.58	0% vs. 23.6%	1
Gastric antrum	12.8% vs. 21.7%	0.23	33.3% vs. 23.6%	0.48
Gastric body	16.7% vs. 21.3%	0.57	22.2% vs. 21.0%	1
Duodenal	17.4% vs. 21.9%	0.78	22.2% vs. 24.0%	1
Jejunal	36.3% vs. 17.5%	0.04	50% vs. 18.4%	0.01
Ileal	26.4% vs.11.8%	0.04	30.2% vs. 12.9%	0.06
Caecal	19.8% vs. 21.4%	0.84	25.9% vs. 23.8%	0.85
Rectum	19.6% vs. 20.6%	0.90	22.4% vs. 26.9%	0.64
Perianal	21.2% vs. 19.6%	0.80	21.6% vs. 23.8%	0.80

a. Only 2 colonic locations have been included as results for all colonic locations demonstrated similar results.

Table 4-10: Blood parameters and NOD2/CARD15 carriage

	At diagnosis	*p-value*	*2 year follow up*	*p-value*
Variable				
ESR>10mm/hr	20.6% vs. 25.0%	0.74	18.1% vs. 26.8%	0.54
Thrombocytosis	24.0% vs. 13.5%	0.12	26.5% vs. 23.1%	0.74
Anaemia	21.4% vs. 17.3%	0.54	31.4% vs. 18.4%	0.17
Raised CRP	25% vs. 9.5%	0.04	57.1% vs. 12.8%	0.0009
Albumin<35g/l	31.0% vs. 9.0%	0.001	33.3 vs. 22.0%	0.43

4.4.2.3 Disease location

Using the Vienna classification (table 4-8), carriage of one or more NOD2/CARD15 variant alleles was more common at diagnosis (50.0% vs. 18.2%, p=0.05) and a trend for the same association was also present at 2 year follow (60.0% vs. 21.1%, p=0.07) in patients with ileal disease using the Vienna classification. NOD2/CARD15 carriage was also more common in those with ileal disease using the paediatric disease classification (table 4-9) at diagnosis (26.4% vs. 11.8%, p=0.04 OR=2.7 [1.0-7.1]) and approached significance at 2 year follow up (30.2% vs. 12.9%, p=0.06).

At disease diagnosis, the allele frequency of the R702W allele was more common in patients with jejunal disease activity compared to those with no jejunal activity (13.6% vs. 4.6%, p=0.02 OR=3.2 [1.1-9.7]). Carriage rates of one or more allelic variants of NOD2/CARD15 were higher in patients with jejunal disease at diagnosis and 2 year follow up compared to those with no jejunal disease (36.3% vs. 17.5%, p=0.04 OR 2.7[1.0-7.3] and 50% vs. 18.4%, p=0.01 OR=4.4[1.3-14.7] respectively). In patients who had both jejunal and ileal disease (n=14) carriage rates of one or more NOD2/CARD15 variant alleles were higher than in patients with no disease at either site, at diagnosis (42.9% vs. 10.9%, p=0.007 OR=6.2 [1.5-25.1]) and at 2 year follow up (50% vs. 10.7%, p=0.003 OR=9.7[1.9-49.1]).

No association was found between other disease locations, using either disease classification systems and NOD2/CARD15 mutations at diagnosis or follow up.

4.4.2.4 Disease behaviour

Possession of R702W mutants was less common in patients with purely inflammatory disease behaviour, compared to patients with complicated (stricturing or penetrating) disease behaviour (allele frequency 4.6% vs. 14.0%, p=0.01 OR=0.3[0.1-0.8]). Carriage of any of the NOD2/CARD15 variant alleles was also less common in patients with inflammatory disease at diagnosis, compared to patients with complicated disease (17.0% vs. 33.3%, p=0.04 OR=0.4[0.2-1.0]) and showed a trend towards significance at 2 year follow up (16.7% vs.31.6%, p=0.07).

The presence of stricturing disease behaviour at diagnosis showed a trend towards an increase in carriers of NOD2/CARD15 variant alleles (42.9% vs. 18.7%, p=0.14), and became significant in this group by 2 year follow up (45.5% vs. 19.4%, p=0.05 respectively).

4.4.2.5 Need for Surgery

In total 27% of patients in the study needed surgery for complications of Crohn's disease. The median time from diagnosis to surgery was 1.83 years (Inter Quartile Range=0.3-4.4). Patients possessing the R702W or Leu1007fsInsC variants or possessing any one or more of the three NOD2/CARD15 variants were more likely to have a need for surgery for intestinal complications of CD compared to those who did not need surgery (11.5% vs. 4.3%, p=0.02 OR=2.8[1.1-7.6], 10.0% vs. 2.3%, p=0.003 OR=4.8[1.6-14.7] and 39.5% vs. 12.8%, p=0.0002 OR=4.45 [2.0-10.1] respectively).

4.4.2.6 Blood parameters

The results are listed in table 4-10. Carriers of NOD2/CARD15 variant alleles were more likely to have a raised CRP at diagnosis and at 2 year follow up (25% vs. 9.5%, p=0.04 OR=3.2[1.0-10.0] and 57.1% and 12.8%, p=0.0009 OR=9.1[2.2-37.3] respectively) compared to patients who had normal values. At diagnosis, carriers of NOD2/CARD15 variants were more likely to have a low albumin (31.0% vs. 9.0%, p=0.001 OR=4.6[1.7-12.1]). No association was found between NOD2/CARD15 carriage and abnormal ESR, platelets or haemoglobin at diagnosis or 2 year follow up.

4.4.2.7 Anthropometry

No significant relationship was found between weight, height or body mass index (BMI) centile at diagnosis using standard UK centile charts (© Child Growth

Foundation,1996). At 2 year follow up NOD2/CARD15 carriers were more likely to be in the lowest weight centile (<0.4th) compared to those not in this centile band (75.0% vs. 20.2 %, p=0.03 OR=3.7[1.8-7.5]) and the (2-9th) height centile compared to those not in this centile band (50.0% vs. 16.0 %, p=0.001 OR=5.2[1.7-16.0]). No relationship was demonstrated with BMI centile at follow up.

4.4.2.8 Other variables

There were 68.5% of CD patients who had granulomata present at diagnosis. Variant alleles of G908R and 1007fsInsC were both significantly lower in the patients with histological evidence of granulomas compared to those with no evidence of granulomas (0.4% vs. 6.1% p=0.001 OR=0.1[0.0-0.6] and 3.0% vs. 9.1% p=0.03 OR=0.3[0.1-0.9] respectively). Carriage of one or more NOD2/CARD15 variants demonstrated the same relationship (13.8% vs. 30%, p=0.01 OR=0.4[0.2-0.8] respectively).

In CD patients with extraintestinal manifestations (EIM's) 10.1% had joint EIM's and 12.0% skin EIM's. No associations were found between NOD2/CARD15 mutations with the presence or absence of extra-intestinal manifestations.

4.4.3 Multifactorial Analysis

Multifactorial analysis using binary logistic regression controlling for factors implicated on unifactorial analysis- disease location, disease behaviour, weight centile, granulomata, jejunal disease and albumin demonstrated NOD2/CARD15 carriage was an independent risk factor for surgery for IBD (p=0.004, OR 4.9 95% CI:1.7-14.7).

4.5 Discussion

This study reports the contribution of the three common NOD2/CARD15 mutations in a large Scottish paediatric cohort and, although, a series of studies has confirmed the high incidence of early-onset of IBD in Scotland,[2;4;6] this study demonstrates a notably low variant allele frequencies of NOD2/CARD15 variants compared to the majority of published adult[185] and paediatric series (table 4-11).

In contrast to the wealth of published data in adults, there are only a small number of published studies looking at NOD2/CARD15 mutations in exclusively paediatric IBD populations (table 4-11).[186-194] The published studies performed to date have been of varying quality with four containing 100 or less Crohn's disease patients[186;187;189;191] and three containing no population controls.[187;188;191] The major strengths of this study, in comparison to the majority of published studies are, the large number of early onset patients, a largely homogenous study population and rigorous disease phenotyping.

Despite the methodological problems noted, the majority of paediatric studies have replicated the association of mutations within the NOD2/CARD15 gene and susceptibility to Crohn's disease.[186;189-194] These studies in general have demonstrated similar allele frequencies to corresponding adult studies in the same population; the exception is one small study of 65 German patients from Saxony published by Sun et al. that demonstrated notably high carriage rates with two thirds of the patients having at least one NOD2/CARD15 mutation.[191] The conclusions in this small selected population have to be treated with great caution however, as no population controls were analysed in the study.[195]

Table 4-11: Allele frequencies of common 3 NOD2/CARD15 variants in paediatric studies

Study location	**Number of CD patients** [a]	**Leu1007fsinsC**	**R702W**	**G908R**
This study	167	4.2%	6.4%	2.2%
Germany[191]	55	25.4%*	14%	5%
USA[193]	163	13.1%*	6.6%	6.0%
USA[190]	101	17%*	6%	11%
Canada[192]	167	9.6%**	8.4%	5.6%
Israel[189]	93	9.7%***	8.6%	21.5%
Sweden[187]	58	1.7%****	2.6%	0%
Italian[194]	134	7.1%****	9.0%	7.8%
Italian[186]	54	7.4%****	8.3%	8.3%

a. Refers to the number of Caucasian patients in each study

The age at diagnosis was <18 years in the studies listed except for the Scottish patients that were diagnosed at < 16 years, Swedish <17 years and the second study from the USA where age at disease onset was not stated.

Chi- square p value of allele frequency of 1007fsInsC in Scottish CD patients compared to allele frequency in other populations: * $p<0.00001$, ** $p=0.006$, *** $p=0.08$, **** p = not significant.

The Scottish population had a statistically lower allele frequency of the 1007fsInsC in comparison to four of the other populations studied (German, USA and Canadian) [190-193] and shows a strong trend towards significantly lower allele frequency in comparison to the 5th study population (Israeli).[188] The Scottish and Swedish (a similar high disease incidence Northern European population) demonstrated no significant statistical difference between them. It is interesting to note that the two Italian paediatric populations,[186;194] did not show any statistical difference when comparing the allele frequency of Leu1007fsinsC but did show a statistically higher carriage of NOD2/CARD15 variants overall ($p<0.001$).

Overall allele frequencies are significantly lower in the Scottish paediatric CD population compared to the CD patients studied in the index populations (mixed child and adult) of Northern Europe and North America ($p<0.00001$).[166;167] This observation provides a critical confirmation in the paediatric CD population of the evidence for genetic heterogeneity in the contribution of NOD2/CARD15 gene that was initially reported in the adult CD populations by Arnott et al.[184] Thus, the low prevalence of NOD2/CARD15 variant alleles present in Northern latitudes of Europe in adults (Scotland, Norway, Iceland, Finland, Ireland, Denmark and Sweden),[184;196-201] is very consistent with the findings of this study and the smaller Swedish paediatric study (figure 4-2).[187] This data contrasts with the higher allele frequency present in Southern European populations (France, Belgium, Germany, Spain, Italy, Hungary and Holland).[115;166;198;202-204] This data have been confirmed in an independent European study looking at NOD2/CARD15 data in several different European studies that confirmed a statistically lower ($p< 0.001$) in patients from Scandinavia compared to patients in other European countries (Scottish patients were not included in the study).[205]

The widespread heterogeneity in NOD2/CARD15 carriage rates in adult and childhood CD patients extends throughout the world outside of Europe and North America. There is an absence of NOD2/CARD15 carriage reported in four adult Asian populations,[206-209] and a low carriage rate has been reported in other non-Caucasian childhood Crohn's disease populations.[210] This heterogeneity is reflected not only in allele frequency in cases but also in controls.[211] The IBD International

Genetics Consortium has published data on the allele frequency of the common 3 NOD2/CARD15 mutations in more than 3,500 healthy controls. This has demonstrated some of the lowest allele frequencies in *control* populations are Northern European from Finland and Scotland (together with Australia) and the highest frequencies are found in populations from Belgium and Canada.

A recent meta–analysis of 42 NOD2/CARD15 case control studies gives an increased relative risk (RR) of CD of 2.4 for heterozygotes and 17 for homozygotes in patients possessing variant alleles of these three common mutations.[185] The overall population attributable risk for the 3 common NOD2/CARD15 mutations was between 12.7%-21.8% in the studies included in the meta-analysis. The PAR in this study was notably lower at 7.9%.

Comparisons between paediatric and adult genetic studies (direct or indirect) have been performed on the 3 common NOD2/CARD15 mutations. No studies have demonstrated a statistically higher frequency of NOD2/CARD15 mutations in the paediatric population, with the data in this current study almost identical to those of Arnott et al. in studies in the Scottish adult CD population.[184] Two studies, one Italian[186] and one Israeli[188] have come close to statistical significance comparing NOD2/CARD15 carriage in children vs. adults ($p=0.06$ for both respectively) but further studies have confirmed no significant difference between adult and paediatric populations.[212] Thus the notion that early onset patients have a higher genetic contribution to developing IBD has not been proven in this current study or so far in other published studies of the NOD2/CARD15 gene comparing child-adult study populations.

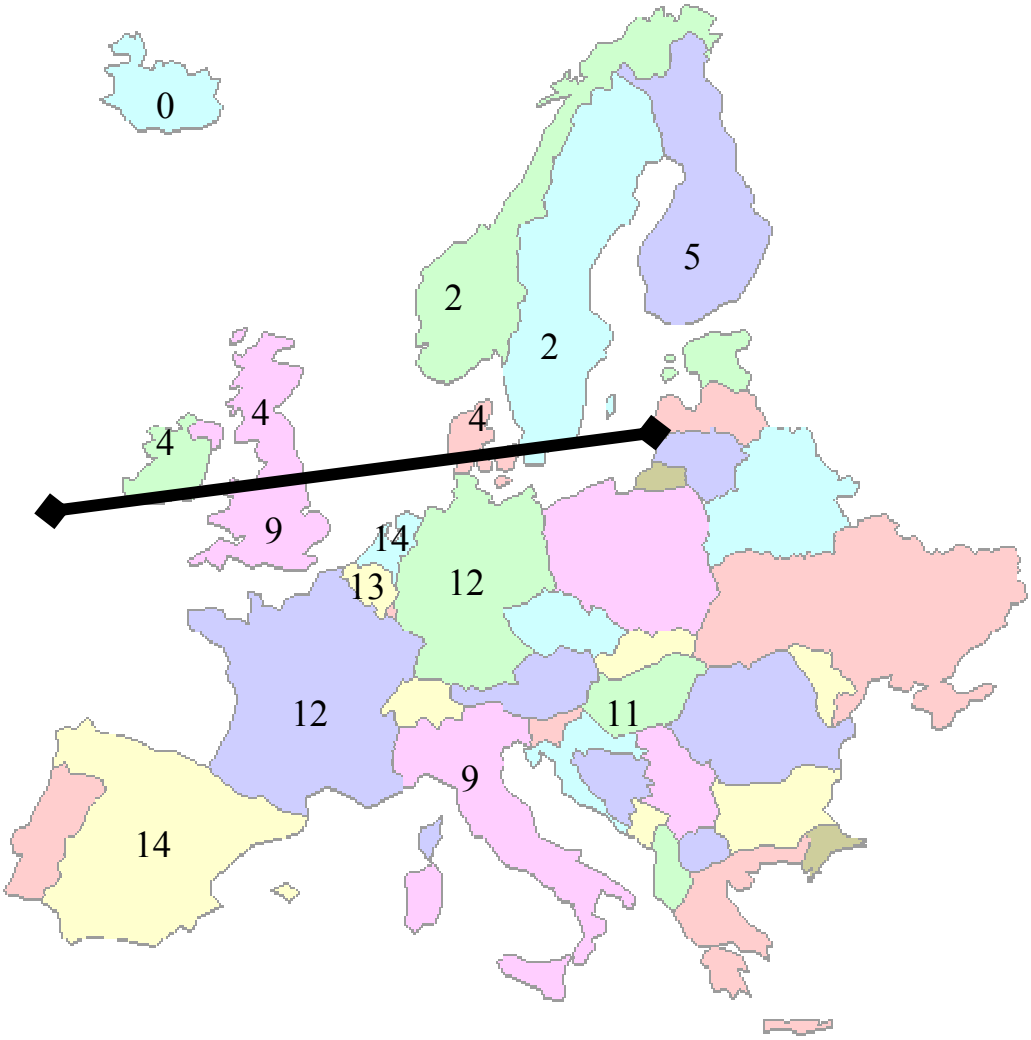


Figure 4-2: Genetic heterogeneity across Europe

The numbers in the figure correspond to the allele frequency of the 1007fs NOD2CARD15 mutation in countries across Europe. The black line illustrates the divide between the low allele frequencies of northern Europe compared to the higher frequencies of Southern Europe.

This illustration was later adapted for publication in the Lancet (Gaya DR, Russell RK, Nimmo ER, Satsangi J. New genes in inflammatory bowel disease: lessons for complex diseases? The Lancet 2006;367:1271-84).

The relative contribution of each of the three common alleles in this dataset to disease susceptibility is consistent with the reported genotype relative risk data reported in most adult populations studied i.e. the Leu1007fsInsC confers the greatest risk and R702W the least.[185] In an analysis of 42 case control studies from across the world Economou et al. demonstrated odds ratios for risk of CD of 2.20, 2.99, and 4.09 for R702W, G908R and Leu1007fsInsC respectively in non-Jewish Caucasians. Notwithstanding the relatively low allelic frequencies in cases and controls in this current study the 1007fsInsC mutation was associated with susceptibility to CD both in the case control study and by TDT analysis; G908R variants were associated with CD susceptibility in case-control analysis only, with no role demonstrated for the R702W mutation.

The case control data in childhood mirror the data in the adult population in Scotland where susceptibility was associated with Leu1007fsIsnC variants (p=0.003) and G908R variants (p=0.008) but not R702W(p=0.27).[184] The TDT data mirror findings both in a North American paediatric study and UK adult study.[193;213] The carriage rates of mutant NOD2/CARD15 alleles in Scottish paediatric CD patients overall were similar to the Scottish adult CD population published by Arnott et al (19.7% vs. 22.8% p =0.47).[184]

Thus the effect of the common NOD2/CARD15 mutations the Scottish paediatric IBD population on susceptibility, is definite but small, and these variants clearly do not explain the high incidence of early onset disease in Scotland.

The rigorous clinical phenotyping was a particular strength of this study and has allowed demonstration of several novel genotype phenotype relationships most notably a number of factors associated with Crohn's disease severity.

Surgery for IBD in this study was more common in those carrying NOD2/CARD15 variants on unifactorial (39.5% vs. 12.8%, p=0.0002) and multifactorial analysis (p=0.004, OR 4.92 (1.65-14.66). This finding replicates data from published paediatric[193] and adult series,[200;202;203;214;215] that all demonstrate possession of NOD2/CARD15 mutant alleles to be associated with an increased need for surgery in Crohn's disease. In a very similar study to this one, Kugathasan et al. demonstrated by regression analysis in a population of 186 paediatric CD patients that patients

carrying the 3 common NOD2/CARD15 mutations were at a significantly higher risk of having earlier surgery for their Crohn's disease.

In this present study, linked to the need for surgery and disease severity was an association between those carrying variants of NOD2/CARD15 disease and complicated (stricturing/penetrating) disease at diagnosis (33.3% vs. 17.0%, p=0.04) and specifically an association with stricturing disease at 2 year follow up (45.5% vs. 19.4%, p<0.05). The association of NOD2/CARD15 mutations and stricturing disease behaviour has been demonstrated in paediatric,[191;193] and several adult series,[200;202;203;216-219] including the meta-analysis of 16 case-control studies that demonstrated increased risk for stenosing disease for NOD2/CARD15 carriers [OR = 1.94, (95% CI: 1.61-2.34)].[185]

Several paediatric studies have demonstrated the presence of common abnormalities (anaemia, thrombocytosis, low albumin and raised CRP/ESR) in routine blood tests in children with IBD and CD at diagnosis.[220-223] The present study is the first to demonstrate a relationship between abnormalities in these commonly performed blood markers and carriage of NOD2/CARD15 variants. A raised CRP at diagnosis and 2 year follow up was more common in NOD2/CARD15 carriers (25% vs. 9.5%, p=0.04 and 57.1% and 12.8%, p=0.0009 respectively). CRP is a very useful, easily available, marker of disease activity in patients with CD that correlates with other markers of disease activity including the Crohn's disease activity index (CDAI) and faecal calprotectin.[224]

Similarly, an association between low serum albumin (<35 g/l) and NOD2/CARD15 carriage at CD diagnosis (31.0% vs. 9.0%, p=0.001) was demonstrated. Both of these findings give strength to the case of NOD2/CARD15 carriage being associated with a severe disease phenotype in paediatric patients. This finding of an association with a severe disease phenotype is supported by the previously mentioned data on risk of surgery in this study and a North American study as well as data from a small German study also showing association of NOD2/CARD15 mutations with several markers of severe disease.[225]

Growth failure is a recognised feature of severe Crohn's disease in children.[189] An association between NOD2/CARD15 carriage and a low weight centile (<0.4th

centile) as well as low height centile (2-9th centile) at 2 year follow up. This replicates the findings of Tomer et al,[190] who demonstrated increased NOD2/CARD15 carriage at diagnosis with a lower weight centile (<5th centile; 44% vs. 15%, p=0.003) in a population of 101 children with CD from the New York area, albeit our findings were at follow up and not diagnosis. These associations however, were not replicated in a second study performed by Wine et al. who studied 93 Israeli children with CD, perhaps because they used a more rigorous measure of growth using z-scores rather than centiles.[189]

A novel association in this study was demonstrated with NOD2/CARD15 carriage and jejunal CD involvement both at diagnosis and 2 year follow up (36.4% vs. 17.5%, p=0.04 and 50% vs. 18.4%, p=0.01 respectively). In addition it was demonstrated NOD2/CARD15 carriage was very high in both jejunal and ileal disease combined at diagnosis (42.6% vs. 10.9%, p=0.007) and follow up (50% vs. 10.7%, p=0.009). This demonstrates the limits of analysis using the Vienna classification and the potential benefits of using a more extensive phenotypic classification dividing disease location into more than four locations as described in this thesis. The association could also be explained by the marked phenotypic differences in disease distribution between paediatric and adult Crohn's patients, with a greater proportion of proximal disease in children.[79] This is illustrated by comparing Scottish children with Upper GI disease (as defined by the Vienna classification) (46.1%) with Scottish adults (9.8%), p<0.00001.[184]

The strongest and most consistent phenotypic association of NOD2/CARD15 genotype-phenotype studies has been with ileal disease location.[132;190;193;197;198;215;226;227] This well replicated association with NOD2/CARD15 variants and ileal disease that has been demonstrated in this study although only on unifactorial analysis. The relationship of NOD2/CARD15 variants with ileal disease was also demonstrated on meta-analysis of 16 case control studies that included genotype- phenotype data with an odds ratio of 2.5.[185]

Leading on from these consistent phenotypic associations, data from Wehkamp et al. has demonstrated that patients with ileal Crohn's disease have an α-defensin deficiency, most pronounced in NOD2/CARD15 mutants.[228] In his study of 45 CD

patients (24 NOD2/CARD15 mutants and 21 with wild-type) and 12 healthy controls, α-defensin production but not the production of other anti-bacterial products was decreased in patients with ileal CD as assessed by real time-PCR of biopsies. There was a 50% reduction of α-defensins in NOD2/CARD15 wild type patients compared to controls with a yet more pronounced effect in the NOD2/CARD15 mutants. In the same study, immunohistochemistry demonstrated that Paneth cells were the sole producers of α-defensins in the ileal biopsies. This provides a coherent hypothesis whereby Paneth cells (which express both NOD2/CARD15 and α-defensins) may be the common link for the occurrence of Crohn's disease of the ileum. This finding is further supported by microarray data from the NOD2/CARD15 knockout mouse model, where expression of cryptdins (the murine counterpart of defensins) is strongly influenced by genetic variants in NOD2/CARD15.[229]

Epitheliod granulomas are often considered to be a pathognomonic feature of Crohn's disease. This is the first study to suggest that possessing NOD2/CARD15 mutant alleles are *protective* against granulomata formation. A *positive* association between mutations of the R702W allele and granulomata formation has been demonstrated in one previous study.[218] A study specifically aimed to examine the relationship between granulomas and NOD2/CARD15 mutations in surgical resection specimens demonstrated no relationship.[230] The findings of this study in contrast to other studies, may partly be explained by the greater tendency to form granulomas in younger patients.[230]

In conclusion, this study has demonstrated that the common three NOD2/CARD15 mutations do make a contribution in susceptibility to Crohn's disease in the high incidence early onset Scottish Crohn's disease population but that the overall contribution is small (PAR 7.9%). The genetic susceptibility in this population should be further investigated by examining all mutations in the NOD2/CARD15 gene (see chapter 5) to exclude an effect from other mutations outside the common 3 mutations, before going onto to study other candidate genes located within the other IBD susceptibility loci (IBD2-9). In this study NOD2/CARD15 carriage was consistently associated with several markers of severe disease including the need for

surgery. Replication of these findings may allow for early stratification of treatment for Crohn's disease in children.

5 Searching for population specific variants in the NOD2/CARD15 gene

5.1 Summary

Aim: To sequence the whole NOD2/CARD15 gene, assessing the contribution to disease susceptibility of other mutations outside of the common three previously analysed (R702W, G908R and 1007finsC).

Subjects: 24 paediatric CD patients for stage 1 then 1478 subjects for stage 2 (320 IBD patients <16 years [210 CD, 80 UC and 30 IC], 343 adult CD patients, 542 parents and 273 populations controls).

Methods: Primers were designed to sequence the 12 exons of the NOD2/CARD15 gene in the 1st stage of the study. Two of the SNPs identified by sequencing were then genotyped in the 2nd stage of the study using the TaqMan system. The two SNPs were selected based on patterns of linkage disequilibrium plus the location of the SNPs within the NOD2/CARD15 gene.

Results: Sequencing identified 18 SNPs in total. Two non-synonymous coding mutations V955I and M863V were then genotyped in 1478 subjects. Carriage of M863V variants was not significantly higher in CD patients vs. healthy controls (1.35% vs. 0.37%, p=0.27). Carriage of V955I variants was higher in IC patients (n=29) compared to healthy controls (n=253) (41.4% vs. 16.2%, p=0.001 OR= 3.65 [1.62-8.21), but not compared to CD or UC patients (12.6% and 15.8% respectively). TDT analysis for V955I was negative.

Conclusions: Sequencing of the NOD2/CARD15 gene and subsequent SNP analysis did not identify any new mutations in the NOD2/CARD15 gene that are associated with increased disease susceptibility in patients with CD. Candidate genes other than NOD2/CARD15 should be investigated to explain the genetic contribution in patients with CD within the Scottish Paediatric population.

5.2 Introduction

The NOD2/CARD15 gene was the first gene described conferring increased susceptibility to Crohn's disease.[166;167] Since the two initial studies, three mutations within the NOD2/CARD15 gene (a frameshift Leu1007fsinsC and two missense mutations R702W and 908R) have been associated with increased susceptibility to Crohn's disease in most Caucasian populations studied.[185]

The contribution of these mutations to disease susceptibility in the Scottish paediatric population is relatively modest, with an increased risk conferred by 2 of the 3 common NOD2/CARD15 mutations (no increased risk for R702W) as described in chapter 4. The PAR in the Scottish paediatric and adult population is lower than most other populations studied at 8% and 11% respectively. Despite the small contribution to disease susceptibility NOD2/CARD15 variants are strong predictors of disease phenotype in children as demonstrated by an increased need for surgery in the Scottish paediatric study (chapter 4) replicating results from an North American Paediatric cohort.[193]

It is clear however that these three "common" NOD2/CARD15 mutations do not account for all of the contribution to disease susceptibility at the IBD1 locus. In a genome wide scan stratified for the common 3 mutations Shaw et al. demonstrated that less than half the genetic contribution at this locus was accounted for by these mutations, leading to a search for other mutations.[231] These rarer NOD2/CARD15 mutations have been identified in a large number of European patients. In this European study 50% of CD patients carried at least 1 NOD2/CARD15 mutation and 17% carried two**.** There were 27 rare mutations identified representing 19% of disease-causing mutations in total. The 3 common polymorphisms (R702W, G908R, and 1007fs) represented 32%, 18%, and 31% of disease-causing mutations, respectively. In combination, 93% of all mutations were located in the Leucine Rich Region (LRR) of the gene. The LRR domain is critically important for interacting with luminal bacteria.[232]

Population specific variants have been identified in the Jewish population. Sugimura et al. sequenced the NOD2/CARD15 gene in 12 Ashkenazi Jewish, Crohn's disease

patients and described a novel conserved variant (IVS8+158; JW1) which was found in Jewish but not non-Jewish patients.[233] JW1 is a C to T substitution in the intron 8 splicing region that it is almost always accompanied by the SNP P268S; the functional consequence of this variation has not been identified. Karban et al. could not replicate increased JW1 frequency in a population of Israeli Ashkenazi Jewish patients with CD but their study used UC patients as the controls instead of normal population controls.[234] A second study in European based Jewish patients by Tukel et al. also failed to replicate the JW1 association.[235]

The apriori aim of this chapter was to sequence all 12 exons of the NOD2/CARD15 gene to identify any potential novel disease causing mutations, focussing on mutations situated in the LRR then analyse the contribution of any potential mutations in a larger cohort of Scottish IBD patients comprising both children and adults.

5.3 Materials and Methods

Patients were recruited as described in chapter 2. Patient data and disease phenotyping was performed as previously described.

5.3.1 Patients

The patients included in this study are summarised in table 5-1. There were 320 IBD patients <16 years [210 CD, 80 UC and 30 IBDU] (cohort 1) and 343 adult CD patients (cohort 2) were studied together with additional samples from 542 parents (305 mothers, 237 fathers). There were a total of 221 (72.4%) completed family trios.

5.3.2 Controls

A total of 253 healthy populations controls (101 local healthy control volunteers recruited directly by the gastrointestinal unit and 152 Scottish blood donors) were used for case-control analysis.

Table 5-1: Demographics of patients involved in the study

	Cohort 1	**Cohort 2**
Sex (M/F)	175/145	137/206
Median Age at diagnosis(y)	11.08 IQR (8.55-12.92)	28.25 IQR(21.00-42.00)
Current Smoker	7/315 (2.2%)	94/333 (28.2%)
Family history of IBD	107/315 (34.0%)	66/343 (19.2%)
Caucasian (%)	313 (97.8%)	329/331 (99.4%)

Cohort 1 320 IBD patients <16 years at diagnosis; Cohort 2 343 CD patients from adult cohort

5.3.3 Sequencing of the NOD2/CARD15 gene

In the first stage of investigation, a total of 24 patients (<16 years at diagnosis) with CD were studied having been selected randomly from the patients collected as part of the wider genetics study. Primers were designed to sequence the exons of the NOD2/CARD15 gene together with intron/exon boundaries. To guide primer construction, the following methodology for primer design was used:

1. The initial base sequence of the intron/exon, within the NOD2/CARD15 gene, was found by searching in the Ensembl database (www.ensembl.org).
2. Primer construction was then performed in the Sigma genosys oligo website www.sigma-genosys.co.uk/oligo_custinfo.cfm, using a customer ID and PIN.
3. An initial (forward) primer was not selected from first 50-60 bases of an intron leading into an exon to avoid incorporating potential splice sites.
4. Whenever possible a primer was selected from a position 80-120 bases from the start of the exon being sequenced.
5. Each primer was 18-25 bases long (17 was the absolute minimum used).
6. A 50% guanine (G)/cytosine (C) ratio was desirable, but not essential.
7. Each primer finished with a guanine (G) or cytosine (C) base.
8. An annealing temperature of around 60°C (Tm) was aimed for.
9. Either a weak or no secondary structure was to be present.
10. There was no primer/dimer interaction.
11. The initial (forward) primer base structure was read backwards i.e. right to left and then the base structure reversed i.e. thymine (T) replaced by adenine (A) and guanine (G) by cytosine (C) and vice versa.
12. The second (reverse) primer was selected at around 600 bases from the site of the first (forward) primer site, with a minimum distance of 300 bases and maximum of 1000 bases.
13. If an exon was large enough to need more than 1 primer pair combination, the 2nd forward primer was started at around 120 bases before the site of the first reverse primer.

The primers documented in table 5-2 are those that were successfully used to sequence DNA. They do not represent all of the primer combinations tried as some were designed and failed to work, at which point a new set of primers were designed and tried.

The basic PCR reaction used was: 25µl of 10x PCR buffer (160mM (NH4)2SO4, 670mM Tris-HCI (pH8.8 at 25°C), 0.1% Tween-20), 25µl of each primer (10uM), 50 mM Magnesium Chloride 7.5-15µl , 25uM dNTP 2.5-5µl, Taq polymerase 2.5-5µl (5U/µl, Bioline) and 156µl of water. This was mixed and 20ul aliquots used per reaction. Into the 20µl of each reaction 1µl (100nanograms) of DNA was then added. A temperature gradient was then performed on the PCR machine with annealing temperatures ranging from 53°C to 65°C. The PCR product was then run on a standard gel and the optimum annealing temperature estimated at the point where a single band appeared at its brightest on the gel

After the optimum annealing temperature was obtained the basic PCR reaction was run again at the appropriate temperature using the 24 samples of DNA selected for analysis. After the PCR had been performed 5µl was removed from 3 of the 24 samples and run on a standard gel to ensure the reaction had been successful. If a product was obtained the sample was then sent for sequencing.

Direct sequencing was performed and the samples run on the 7900 HT sequence detection system (Applied Biosystems, Foster City, Ca, USA) by the Technical Services section of the MRC Human Genetics Unit, Edinburgh. DNA sequence was read using Sequencher v4.5 (Gene Codes Corporation, Ann Arbor, MI, USA).

Table 5-2: Primers used in the sequencing of the NOD2/CARD15 gene

Exon	Forward primer	Reverse primer
1.	TTGTGCCAGAATTGCTTG	AAGGGTAGAATAAGCTCTGGG
2.	TCTGAGGCTAGAACCATGG	TGAGGACAAATCAGTCTTGG
3.	GACCCTTTATTCTGGATGGAAG	CGGTACAGATAATGAGAGTTTGG
4(1)	TGCTCTCCTATCCCTTCAG	TCAGAGAAGCCCTTGAGG
4(2)	GAAGTACATCCGCACCGAG	GAAGGCTGCTGTGATCTG
4(3)	CCAGGCAACTCACCAATG	AAGGGAAGGGATCTGGG
5,6	CACTTCAGGGATGAATGAAAG	GCATTAGAGAACCCCTGC
7.	GTCTTCAATGCTTTCTTCCTG	TCTTGTCAAATGGACTCCAG
8.	AAGTCTGTAATGTAAAGCCAC	CCCAGCTCCTCCCTCTTC
9.	GAGCACCGCAATCAATTAG	CACTCAATCATCCACCTTTG
10.	TTCTTTATCCATGAGTTTGGG	CTTTATTGGTTACCTTCACTTC
11.	GAAGAGAGACGGTTACATTTCAC	CATTCTTCAACCACATCCC
12(1)	TAAAAACAGCCCTGACTTCC	AATTGTCTTGGGGAACAAAC
12(2)	ATTCAGAATATTAGTGACCTCAGC	ATGTTGGTCAGGTTGGTC

5.3.4 Genotyping

TaqMan was used to genotype the 2 SNPs: V955I and M863V using assay on demand and assay by design methodology respectively. TDT analyses in this study were performed using Family Based Association Tests (FBAT) software (version 1.7.3, available from http://www.biostat.harvard.edu/~fbat/fbat.htm). The results were reported using the methodology proposed by Kazeem & Farrall.[236]

5.4 Results

Sequencing of 48 chromosomes in 24 patients identified 18 SNPs in the 24 patients studied. The results are presented in table 5-3.The frequency of these SNPs in each of the 24 patients is listed in table 5-4.

In order to identify any novel potential disease causing mutations, SNPs were selected for further analysis if they caused an amino acid change in the LRR region of the protein. Four SNPs were identified but as two of these SNPs had been previously genotyped in chapter 4 the remaining two [Methionine863Valine (M863V) and Valine955Isoleucine (V955I)] were subject to further analysis in a larger population of patients to assess any contribution towards increased disease susceptibility.

An analysis of linkage disequilibrium between all 18 SNPs identified within the NOD2/CARD15 gene was performed (figure 5-1) using Haploview version 3.2 (www.broad.mit.edu/mpg/haploview). The 2 SNPs selected for further analysis were not in strong linkage disequilibrium (pair wise D prime score 0.34).

Table 5-3: SNPs identified in sequencing of the NOD2/CARD15 gene

SNP	SNP description	rs number	Exon number / base change
1	upstream	rs5743264	Exon 1 promoter T/C
2	5' UTR(-59)	rs5743266	Exon 1 G/A
3	intronic	rs2076753	Exon 2 promoter G/T*
4	Ser178Ser	rs2067085	Exon 2 C/G
5	Ala211Ala	rs5743269	Exon 3 C/T
6	Pro268Ser (SNP 5)	rs2066842	Exon 4(1) C/T
7	Arg459Arg (SNP 6)	rs2066843	Exon 4(2) C/T
8	Arg587Arg (SNP 7)	rs1861759	Exon 4(2) T/G
9		**	Exon 4(2) C/T
10	Arg702Trp (SNP 8)	rs2066844	Exon 4 (3)C/T
11		**	Exon 4 end + 10 bases A/C
12	Met863Val	**	Exon 6 A/G
13	Gly908Arg (SNP12)	rs2066845	Exon 8 G/C
14	Val955Ile	rs5743291	Exon 9 G/A
15		**	Exon 10 (post exonic +1) T/A
16	Leu1007fsincC (SNP13)	rs2066847	Exon 11 c+/-
17	3' UTR	rs3135499	Exon 12(1) A/C
18	3' UTR	rs3135500	Exon 12(2) G/A

* This is described in Lesage[219] (and here) as an Exon 2 promoter mutation, numbered from the ATG in the 2nd exon that was thought to be the start of the NOD2/CARD15. It's actually in intron one/two.

** No rs number listed in Ensembl

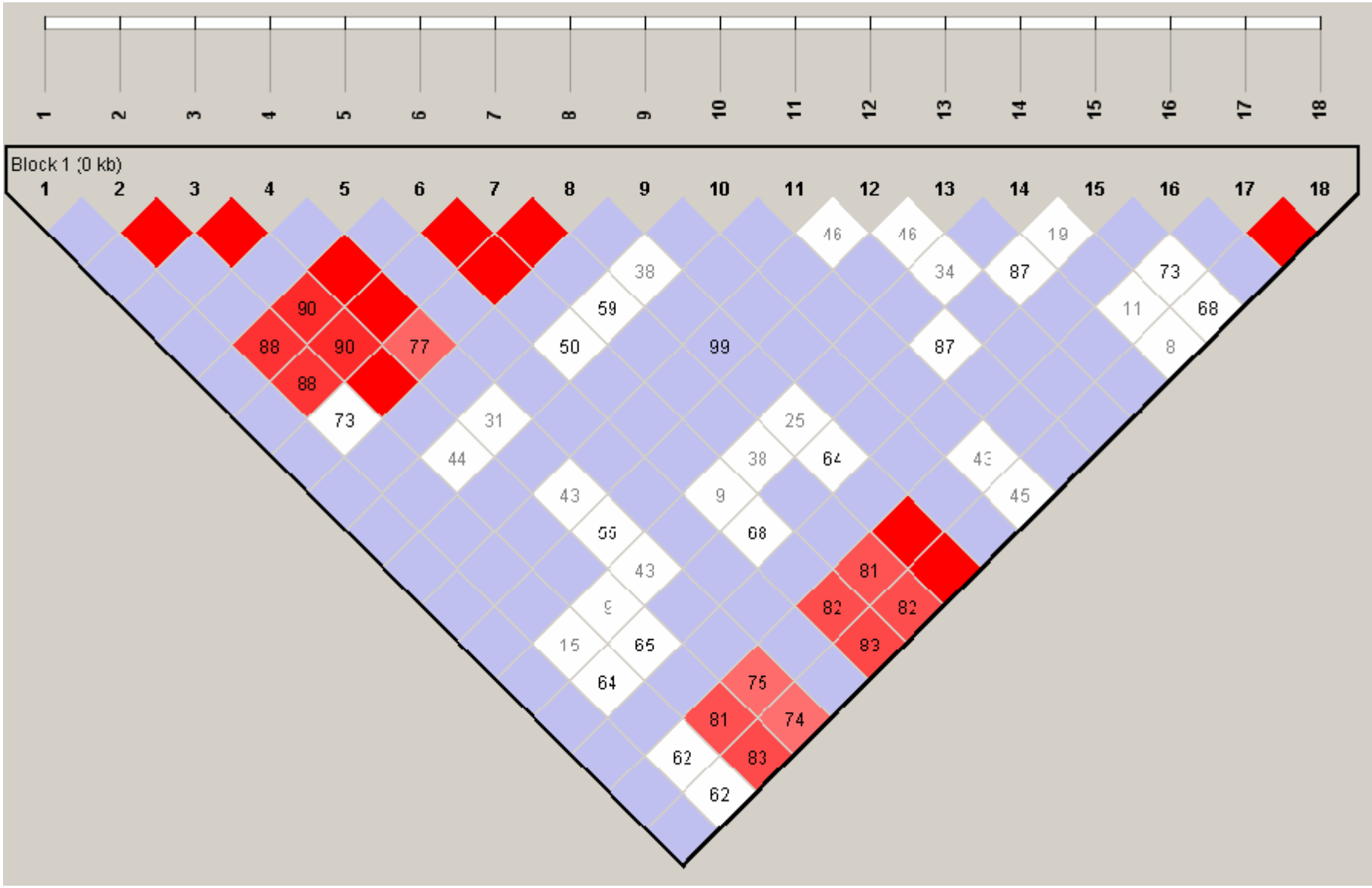
Table 5-4: Patient results of SNPs detected in sequencing

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
-	-	GG	CC	CT	CC	CC	TT	CC	CC	AA	AA	GG	GG	AA		AA	-
TT	GA	TG	CC	CC	TT	TT	TT	CC	CT	AA	AA	GG	GG	TT		AA	GG
TT	GG	GG	CG	CC	CC	CC	GG	CC	CC	AA	AA	GG	GG	TT		CC	AA
TT	GG	GG	CG	CC	CC	CC	TG	CC	CC	AA	AA	GG	GA	TA		AC	GA
TT	GG	GG	CG	CC	CC	CC	TG	CC	CC	AA	AA	GG	GA	TT		CC	AA
TT	GA	TG	CC	CC	CT	CT	TT	CC	CC	AA	AA	GG	GG	TA		AA	GG
TT	GA	TG	CC	CC	CT	CT	TT	CC	CT	AA	AA	GG	GG	TA		AA	GG
TT	GA	TG	CC	CC	-	CT	TT	CC	CT	AA	AA	GG	GG	TA		AA	GG
TT	AA	TT	CC	CC	TT	-	-	-	CC	AA	AA	GG	GA	TA		AA	GG
TT	GG	GG	CG	CC	CC	CC	TG	CC	CC	AA	AA	GG	-	TA		AC	GA
-	-	GG	GG	CC	CC	CC	GG	CT	CT	AA	AA	GG	-	TT		CC	AA
TT	GA	GG	CG	CC	CC	CC	GG	CC	CC	AC	AA	GG	-	TA		CC	AA

TT	GA	TG	CC	CC	CT	CT	TG	CC	CC	AA	AA	GG	GG	TT		AC	GA
TT	GA	TG	CC	CC	CT	CT	TT	CC	CC	AA	AA	GG	GG	TA		AA	GG
TT	GA	TG	CG	CC	CC	CC	TG	CC	CC	AA	AA	GG	GG	TA		AC	GA
TT	GA	TG	CG	CC	CT	CT	TG	CC	CC	AA	AA	GG	GG	TT		CC	AA
TT	GG	GG	CG	CC	CC	CC	TG	CC	CC	AA	-	GG	GA	TA		AC	GA
TT	GA	TG	CC	CC	CT	CT	TT	CC	CC	AA	AA	GG	GG	TA		AA	GG
TT	GA	TG	CG	CC	CT	CT	TG	CC	CC	AA	AA	GG	GG	TT	C+	AC	GA
TT	GA	TG	CG	CC	CT	CT	TT	CC	CC	AA	AA	GC	GC	TA		AA	GG
TC	AA	TT	CC	CC	TT	TT	TT	CC	CT	AA	AA	GG	GG	TT		AA	GG
TT	GA	TG	CG	CC	CT	CT	TT	CC	CC	AA	AA	GG	GG	TA		AA	GG
TT	AA	TT	CC	CC	TT	TT	TT	CC	CC	AA	AG	GC	GC	TT		AA	GG
-	-	TG	CG	CC	CT	CT	TG	CC	CC	AC	AG	GG	GG	TT		AC	GA

The numbers of the SNPs (corresponding to those listed in table 5-3) are placed horizontally across the top of the table and the results for each of the 24 patients studied are then listed vertically; – indicates that sequence was not available for that SNP for the patient in question.

Figure 5-1 : Linkage equilibrium across the NOD2/CARD15 gene



5.4.1 Case-control analysis

The results for case control analysis are presented in table 5-5.

5.4.1.1 V955I

There was no significant difference in carriage rates of the V955I variant in patients with Crohn's disease or Ulcerative colitis compared to healthy controls (12.6% vs. 16.2%, p=0.18 and 15.8% vs. 16.2%, p=0.86 respectively). The carriage rate of V955I variants was significantly higher in patients with IC (n=29) compared with healthy controls (n=253) (41.4% vs. 16.2%, p=0.001 OR= 3.65 [1.62-8.21).

5.4.1.2 M863V

Carriage of M863V variants was higher in patients with Crohn's disease compared to healthy controls but this was not significant (1.35% vs. 0.37%, p=0.27). No variants were detected in patients with UC or IC.

5.4.2 TDT analysis

TDT analysis was only performed for the G955A SNP, given the relative rarity of the M863V SNP. TDT analysis did not demonstrate distortion of transmission of G955A variants in patients with IBD, Crohn's disease or ulcerative colitis (table 5-6).

Table 5-5: Comparison of carriage rates by case-control analysis in the different disease groups

Disease group	M863V carriage	p value cf. HC	V955I variant carriage	p value cf. HC
HC	1/267 (0.37%)		41/253 (16.2%)	
CD children	2/207 (0.97%)	0.58	29/202 (14.4%)	0.29
CD adults	5/313 (1.6%)	0.22	36/305 (11.8%)	0.13
Both CD	7/520(1.35%)	0.27	64/506 (12.6%)	0.18
UC	0/80	-	12/78 (15.8%)	0.86
IC	0/30	-	12/29 (41.4%)	0.001

The numbers of carriers of the variant form of the SNP analysed is compared to the number of patients successfully genotyped.

HC=healthy controls; CD=Crohn's disease; UC=ulcerative colitis; IC=indeterminate colitis

Table 5-6: TDT results

Phenotype	**Family Based Association Tests (FBAT)**		**Heterozygous transmissions**		**Kazeem & Farrall[236]**	
	χ^2	p-value	Allele 1	Allele 2	OR	95% CI
IBD	0.023	0.879	25	21	1.19	(0.67-2.13)
CD	0.182	0.670	14	11	1.27	(0.58-2.80)
UC	0.000	1.000	7	6	1.17	(0.39-3.47)

5.5 Discussion

The data contained in this chapter demonstrates that there is unlikely to be a major contribution of other coding NOD2/CARD15 variants to CD susceptibility in the Scottish paediatric CD population. This is an important finding given the relatively small contribution of the three common NOD2/CARD15 mutations in the Scottish population that has been demonstrated in children (chapter 4) and previously by Arnott et al. in Scottish adults with Crohn's disease.[184]

There is clear evidence in most Western Crohn's disease populations that the three common mutations within the NOD2/CARD15 gene contribute to Crohn's disease, but no other major disease associated mutations have been identified.[101] A meta – analysis of 29 case-control studies has shown an increased relative risk for developing CD of 3 fold and 20 fold for heterozygous and homozygous mutant carriers respectively.[185] In the meta-analysis the Leu1007fsinsC mutation made the most significant contribution to increased disease susceptibility in comparison to the other two common disease associated SNPs.[185] This finding has been replicated in a pooled analysis of more than 7000 Crohn's disease patients.[237]

The results from gene sequencing in the current study are similar to those found in previous studies. Lesage et al. were the first group to sequence the entire NOD2/CARD15 gene in a heterogeneous European cohort of 51 Crohn's disease patients including patients from both Northern and Southern Europe.[219] They identified 67 mutations of which 31 were potential disease causing mutations with 93% of the 31 coding for amino acid changes in the LRR region of the NOD2/CARD15 protein. Of the mutations identified 9 had an allele frequency of >5% and 8 of the 9 were more frequent in a CD group of 453 adults compared with 103 French controls (mostly spouses of patients). Three (of the nine) were the common variants (R702W, G908R and 1007finsC) and all of the other 6 were identified in the present study (SNPs 3, 4,6,7,8 and 14 listed in Table 5-3). This replication of data from the Lesage study validates the findings of the current study demonstrating the techniques were appropriate to identify any common NOD2/CARD15 variant that may have been present in the Scottish population. The

Lesage study however was limited by the heterogeneous mixture of CD patients (8 different European countries were involved) and by the choice of controls.

As part of the sequencing study performed by Lesage et al both of the SNPs subject to detailed analysis within this present study were identified. The allele frequency of the V955I mutation was lower in CD patients than controls (63/906 (7.0%) vs. 21/206(10.2%), p=0.11) with comparative figures from this present study remarkably similar (65/1012 (6.4%) vs. 43/506(8.5%), p=0.14). The allele frequency of the M863V allele was rare in the Lesage study found in 4/906 (0.04%) alleles compared to 7/1040 (0.06%) in the current study. The SNP was not found in any UC patients in either study and was found in only 1 healthy control in the Scottish study.

The current study was designed to identify any population specific NOD2/CARD15 mutations in the Scottish population. None were identified, but given the similar allele frequencies in other Northern European populations the negative data in the current study are supported by Medici et al. who sequenced the NOD2/CARD15 gene in 91 Norwegian Crohn's disease patients identifying 23 SNPs across the gene.[238] They did not identify *any* NOD2/CARD15 mutations in this population that accounted for increased disease susceptibility, *including* the common three NOD2/CARD15 variants. However, it is important to note that the case-control study may have been underpowered as it was carried out in a relatively small cohort of 150 patients and 236 population controls. The authors of this study then went on to demonstrate the allele frequencies of the common NOD2/CARD15 mutations were less common in their Northern European (Norwegian) compared to Southern (German) of 309 CD patients as has been discussed in detail in chapter 4.

This study was aiming to identify novel variants only present in an early onset population but did not identify any. In one previous study, the NOD2/CARD15 gene was sequenced in an exclusively paediatric population (<18years old at diagnosis) of 55 Crohn's disease patients from Saxony, Germany.[191] This study identified 14 SNPs across the NOD2/CARD15 gene but only analysed the Leu1007finsC SNP further by case control analysis comparing allele frequencies with 101 healthy controls. In this small study they demonstrated extremely high rates of NOD2/CARD15 carriage, with 65% of the patients studied carrying at least one NOD2/CARD15 mutation.

This suggests that in early onset populations the population the patients are from, rather than the fact they the disease has an earlier age of onset is the predominant factor governing the NOD2/CARD15 mutations identified by sequencing.

The current study suggested for the first time that V955I variants increase susceptibility to developing IC. No analysis was carried out in patients with IC in the Lesage study or any other NOD2/CARD15 sequencing studies. The V955I mutation was selected for further study because it was identified as a non-synonymous coding SNP. A study by King et al. has subsequently suggested that V955I was not likely to be a disease causing mutation based on sequencing of the NOD2/CARD15 gene in species other than humans. They suggested it was unlikely to be a non-synonymous coding SNP because of a lack of evolutionary conservation of this SNP.[239]

This current study and thesis is one of very few that has studied candidate genes in patients with IC with most groups choosing to exclude patients with this diagnosis from genetic analysis. Paediatric IBD populations may be particularly appropriate to study these patients, as proportionately they contain a larger number of IC patients than published adult series contributing up to 30% of all IBD presenting in childhood.[4;240] Carvalho et al. described a series of 250 children presenting to the John Hopkins hospital 30% of whom had an initial diagnosis of IC.[240] The research focus thus far in this group of patients has been limited to the study of serological rather than genetic markers (see chapter 8 for more detailed discussion of studies).[241;242] This may in part relate to the small number of patients with this diagnosis and controversy about the exact disease definition.[243] This is the first study to demonstrate a higher carriage of any NOD2/CARD15 mutation in patients with IC but this is preliminary data as the number of patients analysed is small (n=29) and would need to be replicated by other groups.

In conclusion, despite sequencing all exons of the NOD2/CARD15 no new mutations that contributed to increased disease susceptibility were identified. The major genetic determinants of disease susceptibility of the Scottish population and indeed other Northern European IBD populations remain to be answered.

6 The role of the IBD5 haplotype and OCTN genes

6.1 Summary

Aims: To study the role of the SNPs within the IBD5 locus looking at effects on disease susceptibility and phenotype.

Subjects: 299 children (200 CD, 74 UC and 25 IC) together with 502 parents (for TDT) and 256 healthy controls.

Methods: Genotyping for the OCTN1/2 variants (SLC22A4 1672C→T, SLC22A5 -207G→C) and IBD5 marker SNPs (IGR2096a_1, IGR2198a_1, and IGR2230a_1) was performed by TaqMan.

Results: All SNPs were in strong linkage disequilibrium (D'>0.94). TDT analysis demonstrated association of the OCTN1 variant with IBD (p=0.01) and CD (p=0.04). Allele frequencies of the OCTN1/2 variants were significantly higher in IBD/CD cases ($0.04 < p > 0.01$). The homozygous mutant OCTN1/2 haplotype was significantly increased in IBD (24.3% vs. 16.1%, p=0.02) and UC patients (28.2% vs. 16.1%, p=0.02) compared with controls.

Unifactorial analysis in CD patients demonstrated carriage of the TC haplotype (a combination of the variant alleles of OCTN1/2) was associated with lower weight, height and BMI centile (<9th centile) at diagnosis (87.9% vs. 67.3%, p=0.002 OR=3.52[1.51-8.22], 84.1% vs. 68.4%, p<0.05 OR=2.44[1.00-5.99], 79.6% vs. 61.1%, p=0.02 OR=2.49[1.14-5.44] respectively) and lower weight centile at follow up (87.5% vs. 64.6% p=0.03, OR=3.83 [1.03-14.24]). Multifactorial binary logistic regression analysis confirmed association of the TC haplotype with lower weight centile at diagnosis (p=0.02, OR 3.41[1.20-9.66]).

Conclusions: Variants within the IBD5 haplotype are associated with increased disease susceptibility in IBD patients and with decreased growth parameters in CD patients. The OCTN1/2 variants remain potential positional candidate genes, but require further analysis.

6.2 Introduction

IBD starting in childhood can have profound effects on a child's growth, pubertal development and education. Growth failure is a common feature of childhood IBD particularly in Crohn's disease.[8] Several factors have been linked to growth failure in CD including: delay to diagnosis,[244] jejunal inflammation,[39] disease severity,[189] age at diagnosis[245] and genotype.[190]

The IBD5 locus (5q31-33) was first identified in Genome Wide Scan of North American CD patients as detailed in Chapter 1.[109;112] To summarise, Rioux et al. stratified the genetic data from the Canadian genome wide scan by age, and demonstrated the highest LOD score of 3.9 was found in CD patients diagnosed under 16 years of age.[112]

In a more detailed study of the IBD5 locus, Daly and colleagues reported strong linkage disequilibrium across the region, and derived a risk haplotype for Crohn's disease that was represented by 11 marker SNPs in separate haplotype blocks that spanned the whole 250 kb interval of the locus.[246] Heterozygotes for the IBD5 risk haplotype had a 2 fold increased risk of CD and homozygotes a 6 fold increase, but with no increased risk of UC.[135] Several European studies have replicated the association of IBD5 variants with susceptibility to adult onset CD,[247-250] and additionally, one study has demonstrated an association with UC.[247] Genotype-phenotype studies in adult-onset disease have shown association with perianal CD[249] and earlier age of disease onset.[248] IBD5 epistasis has been demonstrated with the IBD6 locus[116] and with NOD2/CARD15, for both CD[248] and UC.[247;251]

Peltekova et al. suggested two variants within the IBD5 interval are independently associated with CD; variant alleles of the OCTN1 gene (SLC22A4 C/**T,** missense mutation) and OCTN2 (SLC22A5 -207 G/**C,** promoter mutation).[252] Both of these genes have been suggested to play a role in carnitine transport but critical expression and functional data in IBD patients are lacking. In the initial publication from Peltekova and colleagues, the resulting 2 allele risk (**TC**) haplotype was independently associated with susceptibility to CD when CD patients and controls who were homozygous wild type for marker SNP IGR2078a_1 were compared.[252]

In this study apriori the contribution to disease susceptibility and phenotype of 3 marker SNPs on the IBD5 haplotype together with the OCTN1/2 variants and the TC haplotype within the Scottish paediatric IBD population has been assessed. Post hoc, further detailed analysis was performed of the relationship between growth parameters and the five IBD5 SNPs as well as the TC haplotype. In addition, the study has specifically addressed whether the contribution of the OCTN1/2 genes to disease susceptibility is independent of other determinants within the IBD5 locus.

6.3 Patients and Methods

6.3.1 Patients

Two-hundred and ninety nine patients with IBD diagnosed at less than age 16 years were studied. Two hundred patients had an established diagnosis of Crohn's disease, 74 ulcerative colitis and 25 patients had indeterminate colitis.

6.3.2 Parents and controls

Genotyping data from 502 parents were used to construct family trios for TDT analysis (71% of patients had complete family trios). Data from 256 healthy adult controls was also available for case-control analysis with the initial data in the healthy controls analysed by Noble et al.[253]

6.3.3 Disease phenotype

CD phenotype was assessed using the Vienna classification and a more extensive paediatric classification.[36] Specifically relevant to the present study , peri-anal disease was defined by the presence of anal fissures, peri-anal abscesses, fistulae and peri-anal ulcers, and to allow for comparison with the results published by other groups that have analysed the IBD5 haplotype,[249;254] the definition was adapted to use the above definition of perianal disease minus anal fissures.

6.3.4 Data collection

Disease phenotype was determined using a combination of patient questionnaire, interview and retrospective case note review as previously described in chapter 2.[255] The phenotypic characteristics of the CD patients are presented in tables 6-1 and 6-2 and UC patients in table 6-3. The phenotypic data for CD patients at diagnosis are based on 197 patients unless otherwise stated. In this CD patient group there were 121 males, the median age at diagnosis was 11.2 years (IQR 8.6-13.0), 97% were Caucasian, 2% smoked and 34% had a positive family of IBD. The UC data refer to 69 patients at disease diagnosis unless otherwise stated.

Table 6-1: disease location and behaviour of CD patients at diagnosis

Disease location	
Terminal Ileum (L1)	9 (4.6%)
Colon (L2)	48 (24.4%)
Ileocolon (L3)	38 (19.2%)
Upper gastrointestinal (L4)	93 (47.2%)
None	9 (4.6%)
Disease behaviour	
Inflammatory (B1)	163 (82.7%)
Stricturing (B2)	7 (3.6%)
Penetrating (B3)	27 (13.7%)
Disease Location	
Oral	27/182 (14.8%)
Oesophageal	7/170 (4.1%)
Gastric antrum	49/168 (29.2%)
Duodenal	26/167 (15.6%)
Jejunal	29/165 (17.6%)
Ileal	111/173 (64.2%)
Caecal	108/142 (76.0%)
Sigmoid	137/174 (78.7%)
Rectum	139/182 (76.4%)
Perianal*	80/190 (42.1%)
Perianal**	20/190 (10.5%)

*defined by the presence of fissures, perianal ulcers, abscesses or fistulae but not by the presence of skin tags alone; ** not by the presence of skin tags or fissures.

Table 6-2: Anthropometry and atopic diseases in CD patients at diagnosis

Anthropometry	**Number with variable**
Weight <9th Centile	71/185 (38.4%)
Weight <25th Centile	104/185 (56.2%)
Height <9th Centile	46/178 (25.8%)
Height <25th Centile	83/178 (46.6%)
BMI <9th Centile	66/178 (37.1%)
BMI <25th Centile	92/178 (51.7%)
Mean Weight z score with SD	-0.68 +/- 1.55
Mean Height z score with SD	-0.48 +/- 2.01
Mean BMI z score with SD	0.66 +/- 1.60
Atopic Diseases	
Asthma	58/198 (29.3%)
Eczema	68/198 (34.3%)
Hayfever	41/198 (20.7%)
All 3 diseases	15/198 (7.9%)

Table 6-3: Demographics of ulcerative colitis patients at diagnosis

Sex (M/F)	32/37
Median Age at diagnosis(*y*)	10.7 (IQR 8.5-12.9)
Current Smoker	1(1.4%)
Family history	20 (29.0%)
Caucasian (*%*)	66 (95.6%)
Disease extent	
Pancolitis	36 (52.2%)
Extensive colitis	13 (18.8%)
Left-sided colitis	16 (23.2%)
Proctitis	4 (5.8%)
Extraintestinal manifestations (*%*)	7/67 (10.4%)
Anthropometric Measurements	
Weight <9th centile	11/65 (16.9%)
Weight <25th centile	20/65 (30.8%)
Height <9th centile	9/59 (15.2%)
Height <25th centile	23/59 (39.0%)
BMI <9th centile	6/59 (10.2%)
BMI <25th centile	19/59 (15.2%)

6.3.5 Genotyping

Three marker SNPs on the IBD5 haplotype were typed IGR2096a_1, IGR2198a_1 and IGR2230a_1. The rs1050152 polymorphism of the OCTN1 gene (SCL22A4 exon 9 1672C→T) and the rs26313667 (SLC22A5 promoter, -207G→C) polymorphism of the OCTN2 gene were also typed. Genotyping was performed using TaqMan with all SNPs analysed using assay on demand methodology. The primers used for genotyping the SNPs are listed in table 6-4. The relative positions of these SNPs within the IBD5 locus is shown in figure 6-1.

The three NOD2/CARD15 mutations (Leu1007InsC, G908R and R702W) were genotyped as previously described in chapter 4.

6.3.6 Statistical analysis

Each of the 5 SNPs was analysed individually using allele frequencies and carriage rates for association with IBD, CD and UC. As the total number of patients with IC was small these data were included in IBD analysis overall but were not analysed as an individual disease group. Multifactorial analysis was performed using binary logistic regression analysis. All SNPs in patients and controls were in Hardy-Weinberg equilibrium.

Table 6-4: Primers used to examine each of the 5 SNPs examined in this study

IGR Number	**TaqMan Primers**
IGR2096	Forward: TCTGAGACAGGAGCCACTAGAG Reverse: CACAGCATCCAGAGTGATCCT
IGR2198	Forward: GGGTTGCATGAGCATTAAGTTTCAA Reverse: CCACATCAAGGATAAGACTGCTAAATACT
IGR2222	Forward: GCGGCTGGCCTTACATAGG Reverse: CCGCTCTGCCTGCCA
IGR2230	Forward: GCAGGCAGAACAGCCATACT Reverse: GGCCACAGAACTTTCATTAAAGTAGGA
IGR3002	Forward: TTTACAGGTGCTTACAACAGAATG Reverse: TAGTCTGACTGTCCTGATTGGAATC

Figure 6-1: Relative position of SNPs within the haplotype blocks

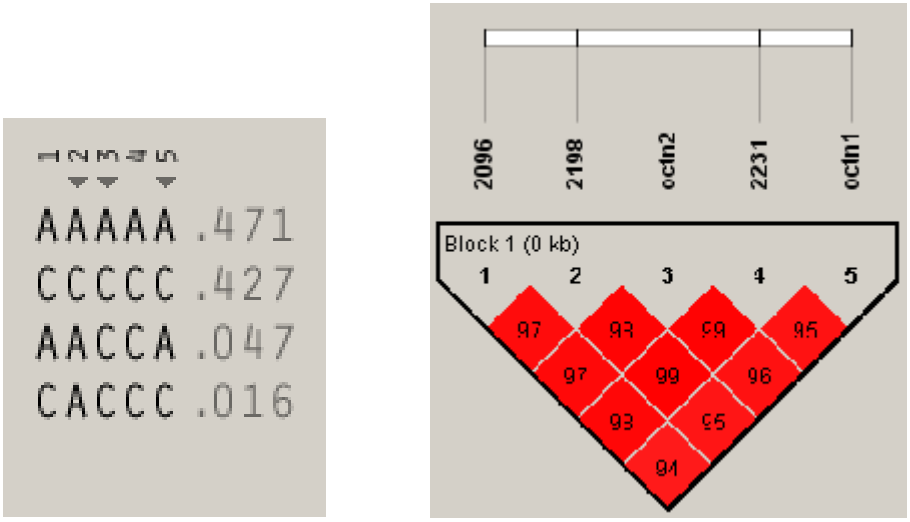
Figure 6-1: The relative position within the haplotype blocks of the IBD5 locus (as described by Daly et al, Nature Genetics 2001;29(2):229-32) of the 5 SNPs within the haplotype examined in this study. Also shown is the marker SNP IGR 2078a_1 examined in the study by Peltekova et al, Nature Genetics 2004;36(5):471-75. The corresponding rs numbers are given below each SNP examined in this study. In Daly's original study 11 separate haplotype blocks were described the numbers shown here (4-7) are the ones referred to in the original publication. IGR 2222a_1 refers to the OCTN2 variant (-207G→C) and IGR 3002a_1 refers to the OCTN1 variant (1672C→T).

6.4 Results

6.4.1 Linkage disequilibrium

Strong linkage disequilibrium (LD) was seen between the 3 IBD 5 haplotype SNPs and the two OCTN variants, as shown below in figure 6-2. As shown in the figure the pair wise D prime scores for all 5 markers was 0.94 or above.

Figure 6-2: linkage disequilibrium across the IBD5 Haplotype



A = Wild type, C= Mutant

Figure 6-2: The D prime scores of the 5 SNPs studied across the IBD5 haplotype are listed on the left in red with corresponding SNPs listed above. The most common haplotype frequencies are listed on the left.

6.4.2 Transmission Disequilibrium Testing

The Transmission Disequilibrium Testing results for IBD5 SNPs and OCTN variants in inflammatory bowel disease patients are shown in table 6-5. The TDT results were based on 293 patients with inflammatory bowel disease together with one or both parents (the numbers of informative trios were 271,197 and 74 for IBD, CD and UC respectively). The IBD5 Haplotype in this analysis refers to the homozygosity for mutants of IGR2096a_1, IGR2198a_1, and IGR2230a_1 combined. The TC haplotype in the table refers to homozygosity for the mutants of OCTN1/2 combined.

TDT analysis demonstrated the IGR2198a_1 allele was associated with susceptibility to IBD ($p=0.01$). Homozygosity for the mutant IBD5 haplotype was associated with susceptibility to CD ($p<0.05$). The OCTN1 variant was associated with susceptibility to both IBD and CD ($p=0.01$ and $p=0.04$, respectively). The homozygous OCTN1/2 mutant haplotype was associated with susceptibility to IBD ($p=0.01$).

Table 6-5: TDT analysis

SNP	**IBD**		**CD**		**UC**	
	χ^2	**p value**	χ^2	**p value**	χ^2	**p value**
IGR2230a_1	2.42	0.12	0.87	0.35	2.00	0.16
IGR2198a_1	5.57	0.02	2.24	0.15	4.03	0.06
IGR2096a_1	2.34	0.16	1.11	0.32	1.36	0.24
IBD5 haplotype	10.59	0.05	10.13	<0.05	4.07	0.42
OCTN1	7.54	0.01	4.21	0.04	3.54	0.06
OCTN2	1.57	0.23	0.47	0.52	1.54	0.18
TC Haplotype	12.92	0.01	6.59	0.11	7.72	0.21

6.4.3 Case-control analysis

6.4.3.1 IBD

Allele frequencies and carriage rates for variants in the IBD5 haplotype in IBD patients compared with healthy controls are listed in table 6-6. A total of 299 IBD patients were compared to 256 healthy controls. The allele frequencies were significantly higher in IBD patients for all 3 IBD5 SNPs and for the OCTN1/2 variants than in healthy controls (p value 0.01-0.04). Homozygosity rates for the IBD5 variant alleles IGR2096a_1 and IGR2198a_1 as well as homozygosity rates for the mutant TC haplotype in patients with IBD were significantly higher when compared with healthy controls (24.4% vs. 15.2%, p=0.008 OR 1.79 [1.16-2.78], 22.8% vs. 15.2%, p=0.03, OR=1.64 [1.06-2.54], 24.3% vs. 16.1%, p=0.02 OR=1.67 [1.08-2.58] respectively).

Case control analysis in IBD patients using carriage rates demonstrated all 5 SNPs were associated with susceptibility to IBD. Homozygous carriage of IGR2096a_1, IGR2198a_1 and the TC haplotype were associated with susceptibility to IBD. The results for heterozygous carriage all failed to achieve significance and are not listed in the table.

Table 6-6: case-control analysis in IBD patients

Allelic Frequency	**IBD vs. HC**	**p value**	**Odds Ratio**
IGR2096a_1	49.6% vs. 42.0%	0.01	1.36 (1.07-1.73)
IGR2198a_1	48.4% vs. 41.0%	0.01	1.35 (1.07-1.72)
IGR2230a_1	53.8% vs. 47.5%	0.04	1.29 (1.01-1.64)
OCTN 1	50.0% vs. 42.9%	0.02	1.33 (1.05-1.69)
OCTN 2	54.5% vs. 47.9%	0.03	1.30 (1.03-1.66)
Homozygous Carriage			
IGR2096a_1	24.4% vs. 15.2%	0.008	1.79 (1.16-2.78)
IGR2198a_1	22.8% vs. 15.2%	0.03	1.64 (1.06-2.54)
IGR2230a_1	27.0% vs. 21.2%	0.12	1.38 (0.92-2.05)
TC Haplotype	24.3% vs. 16.1%	0.02	1.67 (1.08-2.58)
Combined IBD5 carriage	21.3% vs. 14.7%	0.06	1.57 (0.99-2.49)

6.4.3.2 CD

The allele frequencies and carriage rates for variants in the IBD5 haplotype in Crohn's disease patients compared with healthy controls are shown in table 6-7. A total of 200 CD patients were compared with 256 healthy controls. The frequencies of variant alleles in CD patients compared to healthy controls were significantly higher for IGR2096a_1 (49.2% vs. 42.0%, p=0.04 OR=1.34 [1.02-1.75]), IGR2198a_1 (47.7% vs. 41.0%, p=0.04 OR=1.31 [1.01-1.71] and OCTN2 (54.8% vs. 47.9%, p=0.04 OR=1.32 [1.01-1.72]). Homozygous carriage of mutant IGR2096a_1 was higher in CD patients compared to controls (23.6% vs. 15.2%, p=0.03, OR=1.72 [1.06-2.79]).The results for heterozygous carriage all failed to achieve significance and are not listed in the table.

Table 6-7: Case control analysis in Crohn's disease patients

Allelic Frequency	**CD vs. HC**	**p value**	**Odds Ratio**
IGR2096a_1	49.2% vs. 42.0%	0.04	1.34 (1.02-1.75)
IGR2198a_1	47.7% vs. 41.0%	0.04	1.31 (1.01-1.71)
IGR2230a_1	53.8% vs. 47.5%	0.06	1.29 (0.99-1.69)
OCTN 1	49.5% vs. 42.9%	0.05	1.30 (1.00-1.70)
OCTN 2	54.8% vs. 47.9%	0.04	1.32 (1.01-1.72)
Homozygous Carriage			
IGR2096a_1	23.6% vs. 15.2%	0.03	1.72 (1.06-2.79)
IGR2198a_1	20.8% vs. 15.2%	0.12	1.46 (0.90-2.38)
IGR2230a_1	26.9% vs. 21.2%	0.16	1.37 (0.88-2.14)
TC Haplotype	22.2% vs. 16.1%	0.11	1.48 (0.91-2.40)
Combined IBD5 haplotype	19.2% vs. 14.7%	0.42	1.38 (0.82-2.32)

6.4.3.3 UC

The allele frequencies and carriage rates for variants in the IBD5 haplotype in 74 Ulcerative Colitis patients were compared with 256 healthy controls with the results listed in table 6-8.

The frequency of the IGR2198a_1 variant allele was higher in UC patients than controls (50.7% vs. 41.0%, p=0.04 OR=1.48 [1.02-2.15]). Homozygosity for the variant alleles IGR2096a_1 (27.9% vs. 15.2%, p=0.01 OR=2.16 [1.15-4.05]) and IGR2198a_1 (25.7% vs. 15.2%, p=0.04 OR=1.92 [1.03-3.58]) were higher in UC patients than controls. Homozygous carriage of the mutant TC haplotype and of a combined IBD5 haplotype (homozygosity for all 3 IBD5 marker SNPs variants) was also higher in UC patients than controls (28.2% vs. 16.1%, p=0.02 OR=2.04 [1.10-3.78] and 26.6% vs. 14.7%, p=0.02 OR=2.10 [1.09-4.05] respectively). The results for heterozygous carriage all failed to achieve significance and are not listed.

Table 6-8: Case control analysis in UC patients

Allelic Frequency	**UC vs. HC**	**P value**	**Odds ratio**
IGR2096a_1	50.7% vs. 42.0%	0.07	1.42 (0.97-2.08)
IGR2198a_1	50.7% vs. 41.0%	0.04	1.48 (1.02-2.15)
IGR2230a_1	52.9% vs. 47.5%	0.25	1.24 (0.85-1.81)
OCTN1	52.1% vs. 42.9%	0.05	1.45 (1.00-2.10)
OCTN2	54.1% vs. 47.9%	0.18	1.28 (0.89-1.86)
Homozygous Carriage			
IGR2096a_1	27.9% vs. 15.2%	0.01	2.16 (1.15-4.05)
IGR2198a_1	26.8% vs. 15.2%	0.04	1.92 (1.03-3.58)
IGR2230a_1	27.5% vs. 21.2%	0.26	1.41 (0.77-2.60)
TC Haplotype	28.2% vs. 16.1%	0.02	2.04 (1.10-3.78)
Combined IBD5 Haplotype	26.6% vs. 14.7%	0.02	2.10 (1.09-4.05)

6.4.4 OCTN genes, independence within the IBD5 haplotype?

The association between the OCTN1/2 TC Haplotype and disease susceptibility was not independent of other determinants within the IBD5 Haplotype as shown in table 6-9. There was no independent association between the TC haplotype and susceptibility to IBD, CD or UC in individuals lacking the extended IBD5 risk haplotype markers (IGR2198a_1, IGR2230a_1 and IGR2096a_1). The data shown in table 6-9 do not show a significant difference between IBD, CD or UC cases compared with controls in carriage of the variant alleles representing the TC haplotype on a background of having no IGR2198a_1 risk alleles. Thus for Crohn's disease only 4 out of 47 patients carried the TC haplotype but no IGR2198a_1 variants compared with 10 out of 82 IGR2198a_1 variant negative controls (p=0.57). The same lack of association was also seen with the TC haplotype if either of the 2 other variants (IGR2230a_1, and IGR2096a_1) were analysed similarly, or if all 3 IBD5 SNPs were combined. These results are not listed in the table.

No independent effect was demonstrated if the OCTN variants 1 and 2 were considered individually rather than when combined as the TC haplotype.

Table 6-9: Comparison of patients and controls who are homozygous wild type for the IGR2198a_1 IBD5 haplotype marker SNP but have a risk allele in the TC haplotype

	Number of subjects with no IBD 5 risk alleles.	**No. of patients with TC haplotype but no IBD5 risk alleles.**
Controls	82	10
IBD patients	72	5
χ^2 p value comparing IBD patients to controls		0.27
CD patients	47	4
χ^2 p value comparing CD patients to controls		0.57
UC patients	18	0
χ^2 p value comparing UC patients to controls		0.20

6.4.5 Epistasis with NOD2/CARD15 mutations

There was no evidence of epistasis between IBD or CD patients with regard to the TC haplotype and patients carrying the common NOD2/CARD15 mutations (Leu1007InsC, G908R and R702W) in NOD2/CARD15 carriers vs. non-carriers for IBD and CD (70.3% vs. 76.8% p=0.39 and 71.0% vs. 76.9% p=0.49 respectively) as demonstrated in table 6-10.

Table 6-10: TC haplotype carriage stratified by NOD2/CARD15 carriage

IBD PATIENTS (N=244)	**NOD2/CARD15 CARRIER (N=37)**	**NOD2/CARD15 NON-CARRIER (N=207)**	**P VALUE COMPARING CARRIERS VS. NON-CARRIERS**
TC Haplotype	26 (70.3%)	159 (76.8%)	0.39
TC homozygous haplotype	5 (13.5%)	55 (26.6%)	0.09
CD patients (n=165)	NOD2/CARD15 carrier (n=31)	NOD2/CARD15 Non-carrier (n=134)	
TC Haplotype	22 (71.0%)	103 (76.9%)	0.49
TC homozygous haplotype	4 (12.9)	32 (23.9%)	0.18

6.4.6 Genotype-phenotype analysis

6.4.6.1 Unifactorial analysis

For clarity, the results presented in the genotype-phenotype analysis describe associations with the TC haplotype only, with the exception of growth parameters. The phenotypic associations with the TC haplotype generally reflected the genotype-phenotype results for all SNPs analysed across the IBD5 haplotype (all SNPs were in close LD). The results presented in subsequent tables relate clinical features of the Crohn's disease patients included in the study comparing carriage of the TC haplotype and homozygous TC haplotype showing the % carriage of each phenotypic variable compared to the % carriage of those without the phenotypic variable. All data are for phenotypic variables at diagnosis and p values for the comparison between the variables are listed in the table. The Odds ratios with 95% CI for significant results are then given in the text.

6.4.6.1.1 Disease Location

The results for disease location and behaviour are listed in table 6-11. Homozygous carriage of the TC haplotype was associated with gastric antral Crohn's disease, when comparison was made with patients assessed and found to have no antral disease at diagnosis (36.7% vs. 14.2%, p=0.006 OR=3.51 [1.40-8.83]).

6.4.6.1.2 Disease Behaviour and need for Surgery

No association was seen between carriage of the TC haplotype and disease behaviour. In this study 26.9% of patients underwent abdominal surgery for Crohn's disease during the period of analysis. Carriage of the TC haplotype was no higher in those patients needing surgery than those who did not (77.6% vs. 73.9% p=0.61).

Table 6-11 : Genotype-phenotype analysis of disease location and behaviour

Disease location	**TC haplotype**	**p value**	**Homozygous TC**	**p value**
Colonic	77.8% vs. 75.9%	0.80	28.9% vs. 21.1%	0.28
Ileal	77.8% vs. 76.3%	0.92	44.4% vs. 21.9%	0.21
Ileo-colonic	67.6% vs. 78.4%	0.18	17.6% vs. 24.3%	0.40
Upper GI	78.3% vs. 74.7%	0.57	20.5% vs. 25.3%	0.45
None	85.7% vs. 76.0%	0.55	14.3% vs. 23.4%	1
Disease behaviour				
Inflammatory	74.8% vs. 83.9%	0.28	22.4% vs. 25.8%	0.68
Stricturing	83.3% vs. 76.2%	0.68	16.7% vs. 23.3%	1
Penetrating	84.0% vs. 75.2%	0.33	28.0% vs. 22.2%	0.52
Disease location				
Oral	79.2% vs. 75.3%	0.68	20.8% vs. 18.1%	0.75
Oesophageal	83.3% vs. 75.3%	0.65	50% vs. 15.8%	0.06
Gastric antrum	86.4% vs. 71.7%	0.05	36.7% vs. 14.2%	0.006
Duodenal	86.9% vs. 74.6%	0.19	26.1% vs. 17.5%	0.33
Jejunal	78.6% vs. 76.8%	0.84	17.8% vs. 18.2%	0.96
Ileal	75.0% vs. 78.2%	0.65	23.0% vs. 18.2%	0.48
Caecal	75.2% vs. 79.3%	0.65	18.8% vs. 17.2%	0.84
Sigmoid	75.8% vs. 81.8%	0.46	22.6% vs. 18.2%	0.58
Rectum	77.6% vs. 73.7%	0.61	22.4% vs. 23.7%	0.86
Perianal*	78.9% vs. 74.0%	0.46	22.5% vs. 22.0%	0.93
Perianal**	83.3% vs. 75.3%	0.59	27.8% vs. 22.2%	0.45

* defined by the presence of fissures, perianal ulcers, abscesses or fistulae.
** defined by the presence perianal ulcers, abscesses or fistulae

6.4.6.1.3 Growth parameters

A comparison was made between lowest 3 (<9th) centile groupings and the remaining 7 (>9th) centiles for weight, height and BMI (table 6-12). Carriage of the TC haplotype was more common in patients with lower weight, height and BMI compared to those in higher centiles at diagnosis of Crohn's disease (87.9% vs. 67.3%, p=0.002 OR=3.52 [1.51-8.22], 84.1% vs. 68.4%, p<0.05 OR=2.44[1.00-5.99] and 79.6% vs. 61.1%, p=0.02 OR=2.49 [1.14-5.44] respectively).

If the analysis was done using median z scores comparing homozygous carriers of the TC haplotype with non-carriers a significant difference was demonstrated for BMI (-1.32 vs. -0.51 p=0.03) and a strong trend toward significance for weight (-1.16 vs. -0.49, p=0.05) but not for height (p=0.89). At 2 year follow up the TC haplotype was more common in lower compared to higher weight centiles (87.5% vs. 64.6% p=0.03, OR=3.83 [1.03-14.24]).

Table 6-12: Genotype-phenotype analysis of Anthropometric parameters

Centile	TC haplotype	p value	Homozygous TC	p value
Weight <9th centile	87.9% vs. 67.3%	0.002	28.8% vs. 16.8%	0.07
Height <9th centile	84.1% vs. 68.4%	<0.05	25.0% vs. 21.4%	0.62
BMI <9th centile	79.6% vs. 61.1%	0.02	27.8% vs. 16.8%	0.11

6.4.6.1.4 Asthma, eczema, hayfever and atopy

Homozygosity for the TC haplotype was less frequently seen in patients with asthma compared to those without (11.5% vs. 26.9%, p=0.02 OR 0.35[0.14-0.90]), less common in patients with hayfever than those without (7.5% vs. 27.0% p=0.009 OR=0.22[0.06-0.75]) and less common in patients who had all 3 atopic diseases combined (asthma, eczema and hayfever) compared to those without (0% vs. 25.3%, p=0.02) (table 6-13).

Table 6-13: Genotype phenotype analysis of associated atopic disease

Atopic disease	**TC haplotype**	**p value**	**Homozygous TC**	**p value**
Eczema	69.8% vs. 77.3%	0.27	15.9% vs. 26.1%	0.12
Asthma	73.1% vs. 75.4%	0.74	11.5% vs. 26.9%	0.02
Hayfever	72.5% vs. 75.9%	0.66	7.5% vs. 27%	0.009
All atopy	60.0% vs. 76.3%	0.16	0% vs. 25.3%	0.02
No atopy	75.6% vs. 74.5%	0.86	29.1% vs. 17.7%	0.07

6.4.6.1.5 Age at diagnosis

The median age at diagnosis of carriers of the TC haplotype was 11.6 years [vs. 10.9 years for non-carriers (p=0.12)] and for TC homozygotes 12.1 years [vs. 11.3 years in non-carriers (p=0.06)].

6.4.6.1.6 Other phenotypic variables

No association was seen with the TC haplotype and family history of IBD, extra-intestinal manifestations granulomas, or with blood test abnormalities (FBC, ESR, CRP, or albumin concentration).

6.4.6.1.7 Ulcerative colitis

Carriers of the TC haplotype were more common in patients in lower height centiles at diagnosis (<25th centile compared to >25th) (85.7% vs. 55.9%, p=0.02 OR 4.74[1.17-19.16] and showed a trend towards significance for weight centile (83.3% vs. 59.5%, p=0.07 OR=3.40[0.85-13.57]). Genotype-phenotype analysis in this group was limited by the small number of patients available for analysis (n=69). No relationships were found between the TC haplotype and age at diagnosis, extraintestinal manifestations, IBD family history, need for surgery and extent of disease.

6.4.6.2 Multifactorial analysis

Using multifactorial binary logistic regression considering weight centile as either less than or greater than the 9th centile, disease location and behaviour as defined by the Vienna classification, family history of IBD, erythema nodosum and gastric antral Crohn's disease a low weight centile was predicted by carriage of the TC haplotype (p=0.02, OR 3.41[1.20-9.66]) and the presence of antral Crohn's disease (p=0.003, OR= 4.02 [1.58-10.21]).

6.4.6.3 Log-Linear model analysis

Using Log-Linear model analysis to compare propensity across all 11 weight centiles groupings i.e. not dichotomised into above/below 9th Centile, a very strong trend towards higher carriage of the TC haplotype at lower weight centiles and lower TC haplotype carriage at higher weight centiles (p=0.05) at CD diagnosis was apparent (table 6-14). Homozygous carriers of the extended IBD5 risk haplotype (patients who were homozygous mutants for all 3 IBD5 haplotype markers IGR2096a_1, IGR2198a_1, and IGR2230a_1) were more likely to lie within lower BMI centiles and less likely to be within the highest BMI centiles in a logistic regression model (p=0.04).

Table 6-14 : Log linear model analysis

Marker/ Factor	Weight Centiles	Height Centiles	BMI Centiles
IGR 2096	0.525	0.713	0.300
IGR 2198	0.375	0.411	0.251
IGR 2230	0.180	0.300	0.194
IBD5 Haplotype	0.187	0.869	0.044
OCTN1	0.154	0.456	0.282
OCTN2	0.223	0.436	0.245
TC Haplotype	0.051	0.137	0.198

6.5 Discussion

This study demonstrates a definite role for variants within the IBD5 locus in determining susceptibility to both early-onset Crohn's disease and reports initial evidence of association in Ulcerative colitis. In addition, the first consistent genotype-phenotype data demonstrating these variants influence growth in both CD and UC. Additionally the study demonstrated that the OCTN1/2 variants do not act independently of variants within this region in determining disease susceptibility, emphasising the complexity of the IBD5 locus to disease susceptibility.

The most unique strengths of this study lies in the robust phenotypic data that accompany the genetic data and in the size of the childhood population studied. Such detailed growth parameters have not been compared in such a large childhood population in the published literature before. A very strong association between the IBD5 haplotype (and TC haplotype) for low weight and body mass index at diagnosis of Crohn's disease is evident, with a less significant association demonstrated with height centile. The relationship has been demonstrated on unifactorial and multifactorial analysis for weight centile (with the TC haplotype) and for homozygosity of the extended IBD5 haplotype with BMI. Although the relationship was stronger in Crohn's disease the same relationship was also strongly suggested in patients with UC. Given the relatively small numbers of UC patients the study may be underpowered to fully assess this relationship.

The relative weakness of the data presented is that there is a relatively short follow up period for most of the patients in the study as most were enrolled at, or soon after diagnosis. This means the effect of this relationship over time could not be tested to either support or refute these initial findings. It would be interesting to see whether the variants within the IBD5 locus do have a lasting influence on growth and whether they have an influence on adult anthropometry too, not just at diagnosis in children.

This data does represent some of the strongest to date relating genotype to abnormal growth variables in inflammatory bowel disease. Previous paediatric studies have tried to correlate patient genotype with growth parameters but overall results have been not been consistent between populations. Levine et al. in a study of 87 children

with Crohn's disease suggested variant alleles of the -238 promoter polymorphism in the TNF-α gene *protect* patients from height retardation.[245] The children carrying TNF-α -238 variant alleles had significantly higher mean height z scores compared to those carrying wild type alleles. Tomer et al. demonstrated a relationship between carriage of NOD2/CARD15 variant alleles and weight below the 5th centile (but not for height, BMI was not examined) in 87 Crohn's disease at diagnosis.[190] The data in Tomer's study were not subjected to multifactorial analysis; the findings have been confounded by the strong relationship between NOD2/CARD15 carriers with ileal disease which in turn is associated with lower weight at diagnosis.[256] Two subsequent publications analysing 93 and 65 CD patients respectively have failed to replicate these findings between carriage of NOD2/CARD15 variant alleles and growth parameters.[189;225] The lack of agreement between some of these studies and this present study results from the fact differences "centile" cut offs are used in different IBD populations e.g.3rd, <5th and <9th centile. Harmonisation between studies would be improved if z-scores were used universally to examine growth parameters. After this became apparent the growth data collected for the study was refined to include z-scores following the example of Levine et al.

As detailed in chapter 4 the relationship of NOD2/CARD15 genotype with growth parameters in Scottish children with CD demonstrated no relationship at diagnosis, although at follow up carriers of NOD2/CARD15 variant alleles were more common in a low centile band for height (2-9th) and for weight (<0.4th) only. This relationship was present on unifactorial analysis only and growth z-scores were not analysed in NOD2/CARD15 genotype-phenotype analysis.

There is no universally accepted definition of growth failure or good evidence of how it should be treated.[257] Growth failure nevertheless remains a common problem especially in paediatric Crohn's disease that can persist into adult life.[258] Markowitz et al. studied 48 young adults with IBD and found permanent growth failure to be present in 19-35% of patients, depending on the method used to define growth failure. There is also evidence from prospective paediatric studies that the growth failure is "pre-programmed" strongly suggesting a genetic influence.[259] Motil et al. prospectively followed growth in 69 patients with IBD over 3 years and

demonstrated that those with growth failure at the start of the study were the most likely to have growth failure at the end of the study, despite interventions including surgery. Although the data presented in this study do not demonstrate genotype predicts "*growth failure"* because the strongest growth associations are with weight and BMI not height, the data do demonstrate genetic markers may have the potential to be used to predict growth problems and in the longer term even perhaps determine treatment strategies. To further test this association longer term follow up would be needed in our cohort or analysis of growth parameters and IBD5 variants would need to be assessed in an adult cohort.

The association of markers in the IBD5 haplotype with low anthropometric centiles signifying a more severe disease phenotype are entirely consistent with published findings in Scottish adults with CD.[253] Noble et al. demonstrated possession of variant alleles of the TC haplotype in 374 Scottish adults with CD was associated with disease progression at follow up and on multifactorial analysis the need for surgery, both phenotypic variables representing a more severe disease phenotype. This cohort would have been very useful to test our hypothesis with growth parameters but this cohort as with most adult cohorts did not have their growth parameters recorded. This is one of the many reasons why studying childhood cohorts with IBD in addition to adult studies is important as genotypes affecting growth will be ignored by adult gastroenterologists. The association of IBD5 variants and more severe Crohn's disease has also been demonstrated by investigators out with the Scottish adult population.[260]

This study describes a novel association between the TC haplotype and gastric antral Crohn's disease. Previous investigators have linked possession of the TC haplotype with ileal,[261] colonic[262] and peri-anal [254] Crohn's disease but none of these studies have considered a relationship with Crohn's disease of the proximal gastrointestinal tract. The establishment of such a relationship gives some justification for using a phenotypic classification of Crohn's disease location in this thesis that is more extensive than the Vienna classification. The relationship of the antral disease in turn with low weight centile at diagnosis on multifactorial analysis is intriguing in light of previous prospective paediatric studies establishing a link between jejunal CD and

low weight centile at diagnosis.[39] The relationship with antral disease would be considered more robust if it was present at disease follow up and if the findings could be validated by other groups. The difficulty for validation is that adult studies are unlikely to look for antral disease as most adult gastroenterologists will not perform an upper GI endoscopy that is needed to define disease at this site.

The association between the IBD5 haplotype and susceptibility to CD has now been clearly demonstrated in adult studies[247-250;253;262] but this data is the first from an exclusive early-onset population to replicate the findings in Rioux's initial genome wide scan.[112] These data also represent the first paediatric study to demonstrate IBD5 variants increase susceptibility to UC. Giallourakis et al. have previously demonstrated increased susceptibility to UC in 187 German trios by TDT analysis.[247] It is noteworthy that the index study published by Rioux et al. from Canada may have been underpowered to assess association between IBD5 and UC as it contained only 20 families with UC.[112] This current dataset is considerably larger therefore than the index study, nevertheless is relatively small for a case-control association study, containing only 74 UC patients and therefore the data have to be interpreted carefully because of the small number of patients compared to controls. More recently the TC haplotype has been associated with CD susceptibility in 3 adult studies in Canadian, Scottish and German populations. [252;253;262] Waller et al. in a UK case control study of 512 UC patients and 750 controls have also demonstrated an association between the TC haplotype and UC.[263]

Despite demonstrating an association between the OCTN1/2 variants and disease susceptibility, the data do not support these variants as independent predictors of disease when the IBD5 risk markers are removed. This finding is contrary to the findings of Peltekova's study,[252] but consistent with all subsequent publications.[253;262;263] These apparent inconsistencies are likely to reflect the haplotype block structure of the region as described by Daly et al (figure 6-1).[246] The published studies to date outwith the studies in the Scottish IBD population have only studied one haplotype marker IGR2078a_1 (haplotype block 4) and examined the relationship to (or lack of it) with the OCTN1/2 genes (haplotype block 7).[252;262] This study has examined markers in block 4 (IGR2096a_1), block 5 (IGR2198a_1)

and block 7 (IGR2230a_1) representing an extended haplotype across the region. Fischer at al. have examined data from over 1200 case-controls from four European populations and demonstrated lack of independence of the OCTN1/2 variants from the background IBD5 haplotype.[264] Thus the apparent independent effect reported in Peltekova's data may reflect an effect of recombination events between haplotype block 4 and 7 in the Canadian population. Silverberg et al. have further analysed the IBD5 locus in a cohort of 1879 offspring with IBD from the North American genetics consortium study.[141] This study demonstrated that the IRF1, PDLIM, and P4HA2 genes within the IBD5 locus are as likely to be the causal genes from within IBD5 as OCTN1 (this study rejected a role for OCTN2).

Given this uncertainty about the identity of the causal gene within the IBD5 locus it remains pertinent to examine the role of other genes within the region: Interleukin 4 [265-267], Interleukin-13 [268] as well as Interleukins 3 and 5 together with GMCSF are all plausible candidate genes within the linkage interval.[269;270]

Given the data published in the Canadian genome wide scan demonstrated markers in the IBD5 locus had the highest frequency of the risk haplotype in patients with Crohn's disease diagnosed under the age of 16 it is pertinent to see if this is true of this and other paediatric studies.[112] This study together with a study by Babusukumar et al. from within the American early onset IBD populations have both demonstrated markers from within IBD5 haplotype are associated with increased susceptibility to CD,[271] however this risk does not seem to be higher than that of the corresponding adult IBD populations.[253;261] No association with IBD5 mutations and IBD susceptibility has been demonstrated in a smaller Italian study and a very small Hungarian study but both of these negative studies were probably underpowered.[194;272] No genotype-phenotype associations were demonstrated in the American paediatric study.

Direct comparison between this study and the study performed in the Scottish adult IBD population by Noble et al. has revealed very similar results. The genetic contribution to disease susceptibility in both populations is similar with a modest odds ratio in both studies, both studies have associated IBD5 variants with markers of disease severity and neither has supported an independent role for the OCTN

genes within the IBD5 locus. Similar findings have been found in adult and childhood populations in North America. It seems therefore that although the findings of Rioux initial GWS were correct in predicting a role for the IBD5 locus in childhood onset disease this risk has yet to be proven to be dependent on age at disease diagnosis.

The protective association of the homozygous TC haplotype with asthma, hayfever and atopy would be consistent with the evidence from genome wide scans that genes within the 5q31-33 region are not only important in IBD but also in asthma and atopy.[273-276] The contrasting findings that the homozygous haplotype may be a risk factor for IBD while being protective for asthma are consistent with the contrasting effects of the NOD1/CARD4 gene shown for the same 2 diseases.[277;278] Candidate genes associated with asthma and atopy within the 5q31-33 region include IL-12B gene[279]and IL-13.[280]

In summary this study has demonstrated evidence for genetic association between the OCTN1/2 variants within the IBD5 locus and susceptibility to IBD, CD and UC. Of note this study also describes novel and consistent phenotypic associations with low anthropometric centiles. However, the effect of the OCTN1/2 variants is not independent of the extended IBD5 region and further investigation of these and other candidate genes within the IBD5 linkage interval are necessary. This study has demonstrated similar results in the Scottish childhood and adult IBD populations.

7 The role of Drosophila Discs Large Homologue 5 (DLG5) in disease susceptibility and phenotype

7.1 Summary

Aims: To study allelic frequencies of the Drosophila Discs Large Homologue 5 (DLG5) gene 113G/A variant in childhood-onset IBD and additionally, match genotype with phenotypic characteristics at disease presentation.

Subjects: 296 children (197 CD, 73 UC, and 26 IC) with IBD were studied, together with 260 healthy controls. TDT analysis was performed on 270 IBD trios.

Methods: The DLG5 113G/A variant was genotyped by TaqMan®. Full clinical phenotypes including growth and social class (DepCat score) data were also collected.

Results: TDT analysis demonstrated a significant association of this variant with IBD ($p = 0.045$). On unifactorial analysis, 113 G/A variant carriage was associated with: **1)** higher social class (DepCat 1 compared to 2-7 and 1-2 compared to 3-7) (66.7% vs. 22.6% p=0.0005 OR 6.84[1.99-23.55] and 37.2% vs. 22.2% p=0.03, OR 2.08[1.04-4.17] respectively) **2)** higher height centile (>75th centile vs. <75th centile) (42.9% vs. 23.1%, p=0.01 OR 2.50[1.18-5.28] **3)** male sex in CD (29.3% vs. 16.9%, p=0.04 OR 2.04[1.01-4.11]. Multifactorial analysis demonstrated that higher social class (DepCat 1) was independently associated with carriage of variants of 113 G/A (p=0.001, OR=6.92[2.24-21.33]).

Conclusions: The DLG5 113 G/A mutation is associated with susceptibility to IBD. We have identified several phenotypic characteristics that are more common in carriers of variant alleles.

7.2 Introduction

A strong genetic basis for inflammatory bowel disease has been suggested by evidence from: twin studies, prevalence rates of IBD in families of an IBD proband, and differing prevalence rates of IBD between ethnic groups as discussed in detail in chapter 1.[45] Additionally several genes have been associated with susceptibility to inflammatory bowel disease including NOD2/CARD15[166;167] as discussed in chapters 4 and 5, IBD5(OCTN1/2)[252] as discussed in chapter 6, as well variants within the MDR1,[281] and the NOD1/CARD4 genes.[278] The disease model that has evolved over the past decade is that ulcerative colitis and Crohn's disease are related polygenic diseases sharing some but not all susceptibility genes.[76] Although the studies presented in chapters 4 and 6 demonstrate NOD2/CARD15 and IBD5 variants are strong determinants of disease phenotype they do not account for most of the genetic susceptibility risk in the Scottish childhood IBD population.

7.2.1 Deprivation and IBD

A rising incidence of childhood onset inflammatory bowel disease in Scotland has been documented over the past 3 decades by observers within Scotland [3;6] and independently by the British Paediatric Surveillance Unit.[4] Furthermore, within the Scottish population Armitage et al. have noted an association between susceptibility to early-onset Crohn's disease both with Northern latitudes and with relative affluence.[12] In a study of 580 incident IBD cases over the period 1981-95, the incidence of Crohn's disease in the North of Scotland was significantly higher than in the South, 3.1 compared to 2.1 cases per 100,000 population ($p<0.001$).[12] Additionally, the relative risk of developing IBD and CD, analysed by Poisson regression analysis was significantly increased in areas of relative affluence, compared to areas of socio-economic deprivation ($p=0.02$ and $p=0.03$ respectively).

The deprivation category (DepCat) score, has been used and validated, in several Scottish studies as a marker of socioeconomic status.[282] The DepCat score is used to assess social class and is scored 1-7, where 1 equates to the least deprivation and 7

the maximum deprivation. Although this measure of affluence has not previously been studied extensively in IBD , associations with other indirect measures of childhood deprivation in patients with IBD such as hygiene have been proposed.[16]

In other childhood studies, lower DepCat scores have been associated with increased rates of both dental caries and *Helicobacter pylori* infection.[283;284] Radford et al. demonstrated dental caries was more common in Scottish 1 year olds from the lowest DepCat categories (6 and 7) compared to higher scores (1-5). Malcolm et al. presented data showing *Helicobacter pylori* carriage was significantly higher in children from DepCat categories 6 and 7(34%) compared to DepCat 3-5(22%) and in turn higher than categories 1 and 2 (16%).

7.2.2 The DLG5 gene

Hampe et al. performed a genome-wide scan in an European population of 353 IBD sibling pairs (162 CD, 114 UC, 77 mixed) that included UK IBD patients, and identified an area on Chromosome 10 that achieved the criteria for "suggestive linkage" with a LOD score of 2.3 for IBD.[108;285] Despite its relatively modest LOD score the chromosome 10 locus was the highest ranking chromosomal region in this GWS leading the authors to pursue further investigation that eventually lead to candidate gene identification within the locus.

Stoll et al, (in a similar manner to Hugot et al. had used to identify NOD2/CARD15 at the IBD1 locus), used a positional cloning approach to narrow the linkage region in the chromosome 10 locus.[166;286] Fine mapping of the linkage region led to the suggestion that variants of the Drosophila, discs large homolog 5 (DLG 5) gene situated within this locus may be associated with increased susceptibility to Crohn's disease.[286] Stoll's group identified four common haplotypes within the DLG5 gene in the German IBD population. Haplotype D was significantly over-transmitted in IBD patients and the haplotype was identified by a 113G/A coding SNP. The 113G/A variant (a missense mutation from within exon 30 of the DLG5 gene) which encodes for the amino acid substitution at position 30 from Arginine to glutamine (R30Q) was associated with increased susceptibility to IBD and Crohn's disease.[286] Haplotype A was tagged by 8 marker SNPs including the marker SNP DLGe_26 was

a protective haplotype against IBD and CD, thus this SNP and the other 7 marker SNPs were significantly under-transmitted in IBD trios in this study.

Initially Stoll's group found the 113G/A variant was significantly over-transmitted in TDT analysis of 457 German trios with IBD (302 CD and 155 UC), p=0.004 for IBD and p=0.04 for CD. They then replicated these findings in a European case control cohort (538 CD and 548 controls) p=0.001 for IBD. They were not however able to replicate these finding in a separate cohort of 485 IBD trios (reported in the same publication) where they were only able to demonstrate a trend towards significance for IBD and CD (p=0.09 and p=0.07 respectively). Evidence of epistasis between the 113G/A variant of DLG5 and NOD2/CARD15 variants was also observed in the CD cohort in Stoll's study.

DLG5 is a member of the membrane-associated Guanylate kinase (MAGUK) family.[287] MAGUKs are known to form scaffolds for proteins involved in intracellular signal transduction. DLG5 is important in maintaining the epithelial structure, and the genetic variants in DLG5 could therefore interfere with the epithelial barrier function and in turn lead to increased permeability of the mucosa in the gastrointestinal tract.[287;288] In silico analysis suggests that the 113G/A (R30Q) variant may impair DLG5 scaffolding function, but as yet no expression or functional studies in IBD patients have been conducted.[286]

This is the first study of the DLG5 gene in childhood onset inflammatory bowel disease. A priori we examined the contribution of the DLG5 113G/A mutation with respect to disease susceptibility and phenotype. The results demonstrate a statistically significant but overall relatively modest increase in disease susceptibility for carriers of the variant alleles of DLG5 113 G/A. We identified several phenotypic factors associated with variant carriage viz. sex, height and social class.

7.3 Patients and methods

7.3.1 Patients

Two hundred and ninety six patients (197 CD, 73 UC, and 26 IC) with IBD were studied. The demographics of the 197 Crohn's disease patients are presented in table 7-1. There were 116 males, the median age at diagnosis was 11.21 years (IQR 8.61-13.02), 2.6% were smokers, 98% were Caucasian and 35% had a positive family history of IBD. A total of 17/189 (9.0%) had joint extraintestinal manifestations and 21/189 (11.1%) had Erythema Nodosum. Examining growth centiles (75-91st) at diagnosis there were 8/186 (4.3%) in this category for weight, 21/179 (11.7%) for height and 15/179 (8.4%) for BMI.

The demographic data for the 73 ulcerative colitis patients studied are shown in Table 7-2.

7.3.2 Parents and controls

In 69% of IBD patients family trios were available for TDT analysis. There were also data from 260 adult healthy controls (129 males and 131 females) that had been previously genotyped and reported by Noble et al. available for case-control analysis.[289]

Table 7-1: Demographics of Crohn's disease patients

Disease location	
Terminal Ileum (L1)	9 (4.6%)
Colon (L2)	48 (24.3%)
Ileocolonic (L3)	38 (19.3%)
Upper gastrointestinal (L4)	93 (47.2%)
None	9 (4.6%)
Disease behaviour	
Inflammatory (B1)	163 (82.7%)
Stricturing (B2)	7(3.6%)
Penetrating (B3)	27(13.7%)
Disease location	
Oral	28/183 (15.3%)
Oesophageal	7/170 (4.1%)
Gastric antrum	49/168 (29.2%)
Gastric body	32/167 (19.2%)
Duodenal	26/167 (15.6%)
Jejunal	29/165 (17.6%)
Ileal	111/173 (64.1%)
Caecal	108/142 (76.1%)
Sigmoid	137/174 (78.7%)
Rectum	149/192 (77.6%)
Perianal	80/190 (42.1%)

Table 7-2: Demographics of ulcerative colitis patients

Sex (M/F)	35/38
Median Age at diagnosis(*y*)	10.58 (IQR 8.42-12.83)
Current Smoker	1 (1.4%)
Family history	21 (28.7%)
Caucasian (*%*)	71 (97.3%)
Disease extent	
Extensive Colitis	52 (71.2%)
Left-sided colitis	17 (23.3%)
Proctitis	4 (5.5%)
Extraintestinal manifestations	
Joints	3 (4.3%)
Erythema Nodosum	4 (5.5%)
Anthropometric Measurements	
Weight 75-91st centile	7/64 (10.9%)
Height 75-91st centile	5/58 (8.6%)
BMI 75-91st centile	7/58 (12.1%)

7.3.3 Deprivation status

Socioeconomic status based on the patient's postcode was allocated using the Carstairs score. This score is a measure of socioeconomic deprivation derived from levels of male unemployment, head of the household's social class, the level of overcrowding in households, and car ownership generated from census data.[282] Data used to calculate the scores for this study were based on the UK 2001 census data. The Carstairs score was then assigned a numerical deprivation category (DepCat score) based on the postcode, scored 1-7; where a score of 1 is associated with no deprivation and a score of 7 equals maximum deprivation. The data for the DepCat score were not calculated as part of this study; the DepCat scores were merely matched with the patient postcode. The DepCat scores and matching postcodes were obtained from the Edinburgh University data library and were supplied by Robin Rice, librarian. In cases where a patient's postcode fell between two different DepCat categories, no score was allocated. Thus, this score is derived from a person's postcode and reflects the socioeconomic status of the area in which a patient lives rather than an individual case-specific score. The methodology used to define North and South Scotland was also based on a patients' postcodes.[12]

7.3.4 Genotyping

The SNP 113G/A (rs1248696) and DLGe_26 (rs2289311) were genotyped using the TaqMan system assay on demand methodology with primers shown in table 7-3. The genotyping methods used for the 3 NOD2/CARD15(Gly908Arg, Arg702Trp and Leu1007fsinsC) mutations have been described previously in chapter 4.

Table 7-3 : DLG5 SNP primers

SNP	**TaqMan Primers**
113G/A rs1248696	Forward: GCAGCTGAATGGAGAGGTTCT Reverse: CCCCTTCCACAGGCACTAC
DLGe_26 rs2289311	Forward: GGGGACAATGCTGGGCAGGGCCATC Reverse: AGACCCTGCTGCCCTGCCCTTCCTG

Table 7-4: Transmission Disequilibrium Testing results for the DLG5-113 mutant allele

	Observed	**Expected**	**Variation (O-E)**	χ^2	**p value**
IBD	68	58.543	22.272	4.01	0.045
CD	48	42.453	15.991	1.92	0.178
UC	20	16.09	6.4409	2.37	0.105

7.4 Results

The DLGe_26 (rs2289311) was not satisfactorily assayed with > 45% failure rate, so further analysis was not undertaken and thus the results presented are for the DLG 113G/A SNP only.

7.4.1 TDT Analysis

The results are presented in table 7-4. TDT analysis based on 270 families demonstrated an association of allelic variants of the 113G/A allele with IBD (χ^2=4.01, p=0.045) but not Crohn's disease or ulcerative colitis.

7.4.2 Case-control analysis

The results are presented in Table 7-5 for 296 IBD patients, 197 CD and 73 UC patients compared to 260 healthy controls. There was no significant difference in overall allele frequencies or carriage rates of 113G/A variant alleles in patients compared with healthy controls. The 113G/A SNP was in Hardy-Weinberg equilibrium in patients and controls.

Table 7-5: Case control analysis

DLG5 113A	**Controls**	**IBD p value***	**CD p value***	**UC p value***
Allelic Frequency	13.2%	12.9% p=0.92	12.4% p=0.75	13.7% p=0.87
Carrier Frequency	24.6%	25.0% p=0.92	23.2% p=0.73	27.8% p=0.58
Heterozygosity Rates	23.0%	23.9% p=0.81	22.8% p=0.96	27.3% p=0.44
Homozygosity Rates	1.5%	1.0%	1.0%	0%

*p value compared to healthy controls

Table 7-6: DepCat scores in IBD patients and DLG5 carriage

DepCat score	**Number of patients**	**DLG5 carriage**
1	12	66.7%
2	31	25.8%
3	56	12.5%
4	72	23.6%
5	45	33.3%
6	27	18.5%
7	21	23.8%
Unclassified	32	28.1%

7.4.3 Genotype-phenotype analysis: unifactorial analysis

7.4.3.1 Association with affluence

The results are presented in table 7-6 matching a patients DepCat score with carriage of DLG5 variants. A higher rate of carriage of the 113G/A variant was present in the patients with DepCat score 1 (n=12) compared to those with DepCat score 2-7 (n=252) (66.7% vs. 22.6% p=0.0005 OR 6.84[1.99-23.55]) and in comparing allele frequency between the two groups (33.3% vs. 13.3% p=0.006 OR=3.26(1.34-7.92). A higher carriage of 113G/A variants was also demonstrated if patients with DepCat scores 1 and 2 were analysed together (n=43), and compared to patients with DepCat scores 3-7(n=221) (37.2% vs. 22.2% p=0.03 OR 2.08[1.04-4.17] respectively). All patients in DepCat 1 were Caucasian and consisted of 7 males and 5 females. There was no significant difference between Northern (n=35) and Southern latitudes (n=259) in the carriage of variant 113G/A alleles (34.4% vs. 23.9%, p=0.20 respectively).

7.4.3.2 Growth centiles

Carriage of variants of the 113G/A allele were associated with the higher height centiles (>75th centile) for IBD and CD (42.9% vs. 23.1%, p=0.01 OR 2.50[1.18-5.28] and 40.7% vs. 18.9%, p=0.01 OR 2.95[1.23-7.04] respectively). Higher carriage of 113G/A variants was highest in the 75-91st centile for weight and height at diagnosis for patients with IBD compared to those in other centiles (56.2% vs. 23.4% p=0.004, OR 4.20[1.49-11.81] and 48.0% vs. 22.4% p=0.006 OR 3.19[1.36-7.47] respectively).

The same centile associations (75-91st) were also demonstrated for CD patients (62.5% vs. 21.8% p=0.008, OR 6.01[1.37-26.29] and 50% vs. 21.8% p=0.002 OR 4.20[1.6-11.00] respectively). There was no association demonstrated with Body Mass Index centiles or if z scores were used instead of centiles.

7.4.3.3 Sex

In males with Crohn's disease the carriage rate of 113G/A variant alleles was higher than in females with Crohn's disease (29.3% vs. 16.9%, p=0.04 OR 2.04[1.01-4.11] and there was a trend towards higher allele frequency (15.1% vs. 9.0%, p=0.07 OR 1.79 [0.94-3.40]). There was no difference between allele frequency and carriage rates of 113G/A variants in patients with UC or IBD.

7.4.3.4 Other variables

The carriage of 113G/A variant alleles was uncommon in patients with joint disease or joint symptoms but did not reach statistical significance (13.0% vs. 25.9%, p=0.21 OR 0.42[0.12-1.49] respectively). There was no difference in the median ages at diagnosis of carriers of 113G/A variant alleles vs. non-carriers for patients with CD at 11.4 years vs. 11.2 years (p=0.91) and with IBD at 11.3 years vs. 10.9 years (p=0.73). There was no difference in allele frequency or carriage rates of this 113G/A DLG5 variant in disease location or behaviour in patients with either Crohn's disease or Ulcerative colitis at diagnosis.

7.4.4 Multifactorial analysis

Binary logistic regression analysis combining family history of IBD, sex, weight, height and BMI centile (as either 75^{th} –91^{st} or other), social class (as either DepCat score 1 or 2-7) together with joint disease in all patients with IBD was performed. Carriage of the 113G/A variant was associated with DepCat score 1 taking into account all of the above factors (p=0.001, OR=6.92[2.24-21.33]). If the same model was run combining DepCat 1 and 2 together compared to 3-7 and with all other factors the association between higher socio-economic class and 113G/A carriage was confirmed (p=0.02, OR=2.49 [1.13-5.47]).

7.4.5 DLG5-NOD2/CARD15 interaction

There was no difference in carriage of the DLG5 113G/A variants when patients were stratified based on carriage of at least one of the three common CARD15 variants (Gly908Arg, Arg702Trp and Leu1007fsinsC). The carriage of CARD15 variant alleles compared to non-carriage in patients with IBD or CD patients was (14.0% vs. 20.5%, p=0.30 OR 0.63 [0.26-1.53] and 14.7% vs. 20.5% p=0.43 OR 0.66 [0.23-1.88] respectively).

7.5 Discussion

The overall association of the DLG5 113G/A variant alleles with early-onset IBD in the Scottish population appears relatively modest. Carriage of variants was more common in males, taller children and those in the highest social classes. These findings are interesting given that no clear genotype-phenotype relationship has been published previously but require replication in a larger numbers of patients in other populations before they can be accepted as true findings.[262;289]

An association between higher social class (DepCat 1) and early onset IBD in Scotland by Poisson regression analysis in a study of 580 IBD patients has been demonstrated by Armitage et al; these data clearly demonstrated that the relative risk of IBD and CD in Scottish children is decreased in areas of socio-economic deprivation.[12] The association of Crohn's disease with affluence has also been demonstrated in other childhood populations outside Scotland.[290] Blanchard et al. demonstrated an association between affluence also using postcodes and census data in a childhood CD population from Manitoba, Canada. In a follow up study also based in Manitoba, Green at al. demonstrated an association of IBD with higher social class and a negative correlation with rates of enteric infections.[291]

The reason for these associations with affluence remain unclear but the role of water quality,[16] diet (cold-chain hypothesis), hygiene and infection rates [17] are potentially plausible. Gent et al. studied 133 patients with CD and matched them with population controls demonstrating CD patients were significantly more likely to have a hot water tap and a separate bathroom. No association was seen in the 231 UC patients and population controls when studying these measures of hygiene reported in the same study. Hugot et al. proposed the rise in Crohn's disease in the 20th century was due to the increased use of refrigeration.[17] He proposed that psychotropic bacteria that exist particularly at refrigeration temperatures e.g. listeria, yersinia were responsible for interaction with peyers patches and thus involved in Crohn's disease aetiology. These observations are yet to be replicated.

The most provocative data from this study proposes that individuals within these affluent regions may also possess a risk genotype. The association with relative

affluence in the present study is present when DLG5 113G/A carriage rate is compared between patients with DepCat score 1 and DepCat 2-7 and when patients resident within DepCat 1 and 2 are compared to those with DepCat scores 3-7. Moreover the association holds on multifactorial analysis. The major limiting factor in this data however is the small number of patients in DepCat 1 alone at 12. The lack of a trend in the data over the social classes outside DepCat 1 suggest that this finding only relates to the most affluent of children or it is a false positive result created by the relatively small numbers of patients in this group. The data need to be replicated by other groups before broader conclusions can be drawn from analysis of this small patient group.

Assuming these results were found to be true the potential mechanisms whereby DepCat 1 children should differ from children with other DepCat scores and how this influences the development of IBD remain uncertain and a matter for investigation, but it is interesting to note some of the differences highlighted from published studies. Children from lower DepCat scores have higher rates of dental caries, are more likely to be colonized with yeasts, and have higher rates of Helicobacter pylori infection.[283;284] Breastfeeding incidence and duration are highest in DepCat 1 and 2 together in one study,[292] and they are higher in DepCat 1 compared with 2 and other categories in a second study.[293] Strange et al. demonstrated the highest breast feeding rates in Lothian in the period 2002-03 at 6-8 weeks of age in mothers who were DepCat 1 and 2 at a rate of 70% higher than all other DepCat categories. Smith et al. demonstrated breast feeding rates in Argyle and Clyde in the period 1995-99 were significantly higher in DepCat 1 at 51% compared in a stepwise decreasing rate to other DepCat scores with DepCat 7 demonstrating a breastfeeding rate of only 17%. These data present a possible mechanism of how DepCat score, environment, and genotype could all interact.

Growth and development are critical aspects of childhood onset inflammatory bowel disease that are often overlooked in adult practice.[294] Many series have now reported linear growth failure in series of children developing Crohn's disease in childhood.[295] Growth failure is a severe manifestation of CD that has been robustly linked to carriage of disease associated variants from within the IBD5 locus as discussed in

chapter 6. In the present study the association of DLG5 variants is not with growth retardation but rather with *higher* height and weight centiles leading to the hypothesis that this variant may act in the opposite manner, and in fact confer a protective effect against growth failure. A similar protective role has been previously suggested for the –238 TNF-α variant in a paediatric Crohn's disease population.[245] Levine et al. have described an association between the -238 TNF-α polymorphism and a higher height z-score in Israeli children with CD. Thus, the anthropometric data complement the associations described with affluence providing indirect support for the concept that this growth variant leads to disease expression in children lacking socio-economic deprivation.

The 113G/A variants were commoner in males compared to female children with Crohn's disease in this study. It has been reported that the DLG5 113G/A variant is more common in males.[156] Friedrichs et al. combined data from Stoll's original study and added the data from the Italian and Canadian populations (but not the UK) described by Daly et al. The overall odds ratio for risk of CD for the 113G/A variant in this combined study group was 1.52 but when calculated for men and women separately gave OR of 2.49 and 1.01 respectively. This effect is driven by a significantly lower carriage rate of the 113G/A in male control populations. In a separate study, Biank et al. studied 281 children with CD and demonstrated a significant negative association between DLG5 113G/A carriage and female sex and subsequent to this finding demonstrated a significant difference in carriage between males and females. Thus, it seems entirely plausible that studies in early onset populations may be more likely to give positive results when studying DLG5 variations because of the male predominance in childhood onset Crohn's disease in contrast with the female predominance in adult populations.[8;11]

What then of the overall contribution of this DLG5 variant to disease susceptibility? The results of this study and how it compares to other published TDT studies of the DLG5 gene are shown in Table 7-7. The TDT analysis in this study (p=0.045) reveals similar data to those of Stoll et al. in Germany and subsequent publications, which have suggested a weak overall association of variant alleles of DLG5 113G/A with IBD.[286;296] In Stoll's original study, initial TDT analysis was performed on 457

IBD trios (302 with CD) and demonstrated association for IBD (p=0.004) and CD (p=0.04).[286] In a replication cohort of 485 IBD trios (reported in the same publication) Stoll and colleagues were only able to demonstrate a trend towards significance for IBD and CD (p=0.09 and p=0.07 respectively). Daly and colleagues have reported TDT data on the 113G/A variant in two further populations; one reporting a significant association for IBD in a UK population of 124 IBD patients (p=0.02), and in the second no evidence of association in a Canadian IBD population of 182 patients.[296] No p value results were reported in Daly's TDT analysis for patients with Crohn's disease. In the Flemish population, in contrast to results published by other investigators, Vermeire et al. reported significant *undertransmission* of 113G/A variants in TDT analysis of 373 IBD trios.[254] Thus of the disease populations studied a modest effect on IBD susceptibility has been seen in 3 independent populations out of 6 studied and only one study[286] has demonstrated significant association with Crohn's disease. Given this modest effect this present study was probably underpowered to fully explore any association with Crohn's disease and was certainly underpowered for ulcerative colitis.

The case-control data in this current study do not support Stoll's initial findings of a significant difference between cases and controls in possession of variants of the 113G/A allele, this is despite demonstrating very similar allele frequencies of 113G/A variants in this study cases compared to Stoll's, 13.0% and 13.2% respectively. The results from all case control studies are presented in table 7-8.[286] Other case-control studies have yielded inconsistent results.[262;286;289;296] Noble et al. demonstrated no increased risk in 374 Scottish adult CD patients carrying 113G/A variants using a similar group of controls than used in this present study. Torok et al. studied the 113G/A variant in 625 German CD patients and demonstrated no association. Daly et al. performed two case control studies in three different populations.[296] In the first study a combined Canadian/Italian population of 339 IBD (207 CD and 90 UC) patients and 207 controls a significant association was demonstrated with 113G/A with both CD *and* UC. However, in a much larger study in a UK population of 689 IBD patients (353 CD 336 UC) and 493 controls published in the same paper no association was demonstrated. Thus there has been a wide variation in the allele frequency of variants of the DLG5 113A allele in the

different populations studied in both cases and controls. A significant association has only been demonstrated in 2 out of the 6 populations studied.

Table 7-7: TDT analysis in published studies

Study population	**Numbers studied**	**IBD p-value**	**CD p-value**
This study	270	0.045	0.18
German[286]	457	0.004	0.04
German/UK[286]	485	0.09	0.07
Canada [296]	182	NS*	not given
UK [296]	124	0.017	not given
Pooled UK/Canada [296]	306	0.018	not given
Belgium[254]	373	0.01**	0.006**

* Exact value not stated; ** significant *undertransmission* of 113A

Table 7-8: published case control studies

Study Population	**No of cases**[1]	**Allele frequency IBD cases**	**No of controls**	**Allele frequency controls**	**Resultant p value**
This study	296	77/592 (13.0%)	260**	68/520 (13.1%)	ns
European* [286]	525	139/1050 (13.2%)	515	93/1030 (9.0%)	0.001
Canadian/Italian [296]	332	73/664 (11.0%)	202	24/404 (5.9%)	0.003
UK[296]	689	128/1378 (9.3%)	493	96/986 (9.7%)	ns
Scottish (adults)[289]	652	140/1087# (11.4%)	255**	67/510# (13.1%)	ns (0.30)
German [262]	970	196/1940 (10.1%)	972	210/1944 (10.8%)	ns

* Cases were European, controls were German only;** These were the same group of control patients

Personal communication CL noble

1. All numbers of patients or controls quoted are taken directly from the results tables in the relevant publications rather than the numbers quoted in other areas of the publications which often are not the same.

Indeed, in the published studies to date it would seem that heterogeneity amongst controls more than cases in the frequency of 113G/A variant alleles may help explain the major differences between study populations.[297] Tenesa et al. performed a meta-analysis of adult data from the studies of Stoll, Noble and Daly and demonstrated statistically significant heterogeneity between the different study populations more so among population controls than among cases. The heterogeneity described between different study populations matches that shown in the Scottish population for other IBD genes NOD2/CARD15 as discussed in chapter 4 and TLR4.[184;255;298]

The analysis in this study (with positive TDT analysis yet a negative case-control analysis) and the conflicting data from the several other series now reported are nonetheless consistent with the hypothesis that DLG5 may represent a gene of relatively modest effect in IBD, as evidenced by the relatively low odds ratio of 1.62 in the original study and lower still values in replication studies e.g. OR= 1.3 in Daly's study.[286;296] The different results between populations overall may be explained by a combination of factors- heterogeneity between geographically and ethnically distinct populations; as well as heterogeneity within subgroups of IBD patients with specific differences in the sex and age of cases. It is interesting to note that in this study of the Scottish childhood population using TDT analysis, a method that protects against the problems of population stratification to an extent, a positive association with IBD has been demonstrated when case-control results have been negative.[95] An adequately powered case-control study to prove beyond doubt a definite role for DLG5 variants in IBD aetiology may necessitate studies of greater size than any yet carried out in Crohn's disease or Ulcerative colitis, requiring several thousand patients and controls.[296] This strategy is similar to that applied to many other polygenic diseases, in which genetic determinants of low genotype relative risk have been identified.[91]

The results of this study are therefore consistent with those published to date with a modest effect on disease susceptibility and association with male sex. The associations of variant carriage with affluence and protection from growth retardation, while awaiting replication in other study populations need to be interpreted with relative caution given the small numbers of patients analysed.

8 The role of Anti-Saccharomyces cerevisiae antibody in predicting disease susceptibility and phenotype

8.1 Summary

Aims: To determine Anti-Saccharomyces cerevisiae antibodies (ASCA) status and its relationship with disease phenotype in childhood onset IBD.

Subjects: 301 Scottish early onset IBD patients [197 CD, 76 UC and 28 IC].

Methods: ASCA status (IgA, IgG) was determined using an ELISA kit. ASCA status was then matched with Crohn's disease (CD) phenotype at diagnosis.

Results: CD patients had a higher prevalence of ASCA antibodies compared to UC patients and healthy controls (41.6% vs. 15.8% Odds Ratio (OR) =3.80[1.93-7.50] and 41.6% vs. 7.7% OR= 8.56 [3.55-20.62] respectively).

Unifactorial analysis of disease phenotype in CD patients demonstrated a positive ASCA was associated with oral (68.0% vs. 38.6% OR 3.39[1.38-8.34]) and perianal (50.6% vs. 35.2% OR 1.89[1.04-3.44]) disease location as well as markers of disease severity: raised CRP (48.9% vs. 24.5% OR 2.95[1.36-6.37]), hypoalbuminaemia (51.8% vs. 27.0% OR 2.28[1.19-4.37]), surgery (55.1 vs. 36.6% OR 2.11[1.10-4.06]) and the presence of granulomata (47.7% vs. 28.8% OR 2.25[1.13-4.48]). On a multifactorial analysis the presence of oral disease (adjusted p = 0.001, OR 22.22[3.41-142.86]) and hypoalbuminaemia (adjusted p = 0.01 OR 4.78[1.40-16.39]) were found to be independently associated with ASCA status.

Conclusions: CD patients had a higher prevalence of ASCA than other IBD patients. ASCA status defines CD patients with a specific phenotype, showing association with markers of disease severity and oral CD involvement.

.

8.2 Introduction

The data on the presence of specific serologic markers in IBD patients, including children date back many years.[299;300] In 1988, Main et al. reported serum titres of both IgA and IgG antibodies directed against the oligomannan component of *Saccharomyces cerevisiae* (ASCA) were higher amongst patients with CD compared to controls.[301] In subsequent paediatric studies, patients with CD have been found to have a higher prevalence of ASCA antibodies (50-70%) compared to UC patients (10-15%) and healthy controls (<5%).[302;303] ASCA has been linked to various Crohn's disease phenotypes including disease location, disease behaviour and need for surgery.[214;304-308] The presence of ASCA antibodies seems to pre-date clinical symptoms and their presence has been demonstrated before the diagnosis of CD.[309] Whether the development of ASCA antibodies represents an adaptive immune response to an unknown environmental exposure or represents a genetic predisposition to developing Crohn's disease remains contentious.[310]

An association between ASCA status and the presence of NOD2/CARD15 variants also been inconsistently replicated with a number of studies supporting [214;311;312] or failing to support this association.[216;305;308] No associations have been suggested with other IBD genes. ASCA and p-ANCA are well established serological markers in IBD having associations with CD and UC respectively.[313]

Apriori this study aimed to determine the prevalence of ASCA antibodies in the cohort of patients with early onset IBD and assess relationships with disease phenotype. Post hoc analysis also examined the interaction of ASCA status with IBD candidate genes examined in earlier chapters of this thesis NOD2/CARD15 (chapter 4), IBD5/ OCTN (chapter 6) and DLG5 (chapter 7).

8.3 Patients and Methods

8.3.1 Patients and controls

There were 301 IBD (197 CD, 76 UC, and 28 IC) patients analysed. Of the 197 CD patients analysed there were 115 males (58.4%), the median age at diagnosis was 11.25 years (IQR 8.75-13.00) and 35.1% had a positive family history of IBD.

Data previously analysed by Walker et al. examining ASCA status in 78 Scottish adult blood donors who acted as healthy controls, were also available to use in the study.[305]

8.3.2 Crohn's Disease Phenotype

The presenting phenotype of patients with CD is shown in table 8-1.Of particular note in the analysis of this study oral Crohn's disease was defined by macroscopic changes (mucosal tags, deep linear ulceration, cobblestoning, lip swelling and fissuring) after examination by a paediatric dentist or oral medicine specialist and/or the presence of granulomata on buccal biopsy. The presence of oral Crohn's disease as stated by the patients physician was not taken as acceptable evidence of the presence of disease, based on findings from previous studies that have demonstrated a physicians assessment is unreliable.[314]

Table 8-1 Demographics of CD Patients

Disease location	
Ileal	9/192 (4.7%)
Ileocolonic	36/192 (18.8%)
Colonic	48/192 (25.0%)
Upper GI	90/192 (46.9%)
None	9/192 (4.7%)
Disease behaviour	
Inflammatory	160/193 (82.9%)
Stricturing	7/193 (3.6%)
Penetrating	26/193 (13.5%)
Disease location	
Oral	25/178 (14.0%)
Jejunal	28/161 (17.4%)
Perianal	77/185 (41.6%)
Disease parameters	
Raised CRP	90/139 (64.7%)
Albumin≤35g/l	85/159 (53.5%)
ESR >10	141/152 (92.8%)
Platelets >400	117/176 (66.5%)
Anaemia	87/178 (48.9%)
IBD Surgery	49/196 (25.0%)
Granulomata	32/184 (71.7%)

8.3.3 ASCA enzyme-linked immunosorbent assay (ELISA) method

Blood samples were collected from subjects when recruited as part of the early onset genetics study. Plasma was isolated at the time of lymphocyte extraction and samples were stored at -80^{o}C until analysis. The ASCA combi kit® (Generic Assays, Dahlewitz, Germany) was used to measure ASCA IgG and IgA in patient samples, according to the manufacturer's instructions. The optical density (OD) of the developed assay was measured using a spectrophotometric Enzyme-Linked Immunosorbent Assay (ELISA) reader at 450nm wavelength.

Results were interpreted by calculating the binding index (BI) with the formula: BI= $OD_{sample}/OD_{cut\text{-}off\ control}$. The cut-off control, positive and negative control samples were provided in the kit. A BI of ≥ 1 was considered to be a positive result, and <1 was considered to be a negative result. In keeping with standard ELISA practice each patient sample was analysed in duplicate.

8.3.4 Statistical Analysis

The χ^2 test was used for analysis of selected discrete variables. A multifactorial analysis of ASCA status compared with all variables with $p<0.25$ was carried out using a binary logistic regression analysis. The variable was considered independent if a result of $p<0.05$ was generated. Data were analysed using SPSS® Statistical Software, version 11.5.

8.4 Results

8.4.1 ASCA prevalence in patients and controls

CD patients had a higher ASCA positivity rate at 41.6% compared to UC patients (15.8%, p=0.0001 OR=3.80[1.93-7.50]), IC patients (0% p< 0.00001 OR=∞) and healthy controls (7.7%, p<0.0001 OR 8.56[3.55-20.62]). These results are summarised in table 8-2.

The sensitivity of ASCA in predicting CD was 41.6% with a specificity of 88.6%, positive predictive value of 86.3% and negative predictive value of 47.0%. Using likelihood ratios, patients who were ASCA positive were approximately 7 times more likely to have CD than UC.

T**able 8-2**: ASCA prevalence in patients and controls

Study Population	**ASCA positive**	**ASCA negative**	**P value and odds ratio***
CD (n=197)	82 (41.6%)	115 (58.4%)	-
UC (n=76)	12 (15.8%)	64 (84.2%)	p= 0.0001, OR=3.80[1.93-7.50]
IC (n=28)	0 (0%)	28 (100%)	p< 0.00001, OR=∞
HC (n=78)	6 (7.7%)	72 (92.3%)	p< 0.00001, OR= 8.56[3.55-20.62]

* compared to Crohn's disease

8.4.2 Unifactorial Analysis in Crohn's Disease patients

8.4.2.1 Age at diagnosis

There was no significant difference in age at diagnosis in patients who were ASCA positive compared to those who were ASCA negative 11.7 years vs. 11.4 years respectively (p=0.28).

8.4.2.2 Disease location and behaviour

The results summarising ASCA status and disease phenotype are listed in table 8-3. ASCA status was compared in all locations throughout the GI tract using a "paediatric" phenotype as well as using the locations defined in the Vienna classification. Only selected locations that reached statistical significance are presented in table 8-3 for the paediatric phenotype although all locations were analysed. Significant associations were seen with oral and perianal disease (68.0% vs. 38.6%, p= 0.008 OR=3.39[1.38-8.34] and 50.6% vs. 35.2%, p=0.04 OR=1.89[1.04-3.44] respectively).

In patients who did not have a disease location fitting into the Vienna classification classified as "none" a higher ASCA positivity rate was demonstrated (88.9% vs. 40.4% p=0.02 OR=11.76[1.44-100.00]). As described in the methods section, these are patients with oral or peri-anal disease or a combination of both (see chapter 2 for further details). No associations were seen between disease behaviour at diagnosis and ASCA status.

8.4.2.3 Blood results

A positive ASCA was more common in patients with hypoalbuminaemia (<35g/l) (51.8% vs. 29.0% p=0.01, OR=2.28[1.19-4.37] and raised CRP (48.9% vs. 24.5% p=0.006, OR=2.95[1.36-6.37]). No relationship was seen with raised ESR, thrombocytosis or anaemia (table 8-4).

8.4.2.4 Surgery

A positive ACSA status was present in Crohn's disease patients undergoing surgery (55.1% vs. 36.6% p=0.03 OR=2.11[1.10-4.06]).

8.4.2.5 Other variables

Patients with granulomata present at diagnosis had a higher rate of positivity than those without granulomata (47.7% vs. 28.8% p=0.02 OR=2.25[1.13-4.48]). No associations were seen with growth parameters, extra-intestinal manifestations or the presence of atopic disease.

Table 8-3: Univariate analysis of Crohn's disease phenotype

Disease parameter	**ASCA +ve With Variable**	**ASCA +ve Without Variable**	**p-Value, with Odds Ratio and 95% Confidence Interval**
Disease location			
Ileal	5 (55.6%)	77 (42.1%)	0.43, OR=1.72 (0.45-6.62)
Ileocolonic	12 (33.3%)	70 (44.9%)	0.21, OR=0.61 (0.29-1.32)
Colonic	19 (39.6%)	63 (43.8%)	0.61, OR=0.84(0.43-1.64)
Upper GI	38 (42.2%)	44 (43.1%)	0.90, OR=0.96(0.54-1.71)
None	8 (88.9%)	74 (40.4%)	0.02, OR=11.76(1.44-100.00)
Disease behaviour			
Inflammatory	66 (41.3%)	16(48.5%)	0.44, OR=0.75(0.35-1.48)
Stricturing	1(14.3%)	81(43.5%)	0.12, OR=0.22 (0.03-1.83)
Penetrating	15 (57.7%)	67 (40.1%)	0.09, OR=2.04 (0.88-4.70)
Disease location			
Oral	17 (68.0%)	59 (38.6%)	0.008, OR=3.39 (1.38-8.34)
Jejunal	15 (53.6%)	53 (39.8%)	0.19, OR=1.74 (0.77-3.95)
Perianal	39 (50.6%)	38 (35.2%)	0.04, OR=1.89 (1.04-3.44)

Table 8-4: ASCA and disease parameters

Disease parameter	**ASCA +ve With Variable**	**ASCA +ve Without Variable**	**p-Value, with Odds Ratio and 95% Confidence Interval**
Raised CRP	44 (48.9%)	12 (24.5%)	0.006, OR=2.95(1.36-6.37)
Albumin≤35	44 (51.8%)	20 (29.0%)	0.01, OR=2.28(1.19-4.37)
Raised ESR	60 (42.6%)	2 (18.2%)	0.14, OR=3.29(0.69-15.87)
Raised Platelets	53 (45.3%)	19 (32.2%)	0.10, OR=1.74(0.90-3.36)
Anaemia	40 (46.0%)	33 (36.3%)	0.25, OR=1.50(0.82-2.73)
IBD Surgery	27 (55.1%)	54 (36.6%)	0.03, OR=2.11(1.10-4.06)
Granulomas	63 (47.7%)	15 (28.8%)	0.02, OR=2.25(1.13-4.48)

8.4.3 Multifactorial Analysis

The results of this analysis are presented in table 8-5. A multifactorial binary logistic regression analysis was performed for ASCA status on the all variables with a p-value of <0.25. Of these the presence of oral disease and hypoalbuminaemia were related to a positive ASCA status (adjusted p=0.001, OR 22.2[3.4-142.8] and adjusted p=0.01, OR 4.78 [1.40-16.39] respectively).

Table 8-5: Multifactorial binary logistic regression analysis

Variable	**Unifactorial Analysis**			**Multifactorial Analysis**		
	p-Value	**OR**	**C.I. (95%)**	**p-Value***	**OR***	**C.I. (95%)***
Oral Disease	0.008	3.39	1.38-8.34	0.001	22.22	3.41-142.86
Jejunal Disease	0.185	1.74	0.77-3.95	0.109	2.96	0.79-11.11
Perianal Disease	0.036	1.89	1.04-3.44	0.597	1.29	0.50-3.33
BMI <9th Centile	0.237	1.45	0.78-2.71	0.184	1.99	0.72-5.49
Raised CRP	0.006	2.95	1.36-6.37	0.754	1.24	0.32-4.78
Albumin <35	0.013	2.28	1.19-4.37	0.012	4.78	1.40-16.39
Platelets	0.097	1.74	0.90-3.36	0.746	1.24	0.33-4.67
Granulomata	0.021	2.25	1.13-4.48	0.297	1.88	0.58-6.10

* Adjusted

8.4.4 ASCA and interaction with IBD genes

The carriage rate of the 3 common variants of NOD2/CARD15 (R702W, G908R and 1007finsC) was not significantly different in ASCA positive and negative patients (21.5% vs. 22.3% respectively, p=0.90). There was no evidence of any association with DLG5 113A variants or the TC haplotype in ASCA positive and negative patients (27.8% vs. 20.4%, p=0.23 and 77.9% vs. 68.3% p =0.15 respectively).

8.4.5 Childhood vs. adult data

The data for CD patients in this study were compared to that previously published by Walker et al that provided data the Scottish adult cohort for ASCA status. [305] The ASCA positivity rate in children with CD from this current study 82/197 (41.6%) was significantly lower when compared to adults with CD 81/143 (57%) (p=0.006, OR=0.55[0.35-0.84]). There was no significant difference ASCA positivity between children and adults with UC 12/76 (15.8%) compared to 14/75 (18.7%) p=0.64.

8.5 Discussion

In this study the previously described association of increased ASCA positivity in CD patients has been replicated in the largest paediatric IBD cohort studied to date. Additionally, using rigorous phenotypic analysis an association of ASCA status has been demonstrated with markers of disease severity and a novel association has been described with oral Crohn's disease.

The incidence of oral Crohn's disease is higher in Scottish children than in children from other parts of the UK.[39] It is interesting therefore that in this study a positive ASCA status acted as an independent risk factor for oral involvement in Scottish CD patients. No association has been demonstrated with oral disease involvement in the genetic studies described in this thesis in previous chapters (NOD2/CARD15, OCTN1/2, and DLG5). In this study a positive ASCA serology predicted oral involvement in CD patients with a sensitivity of 68.0% and specificity of 61.4% compared to CD patients without oral involvement.

In the only other published study examining ASCA status and oral CD, Savage et al. demonstrated higher salivary ASCA IgA in 26 patients with oral CD compared to 22 CD patients with no oral involvement ($p<0.02$).[315] This finding was independent of intestinal inflammation elsewhere in the gastrointestinal tract in the 26 patients with oral CD (12 of the patients had inflammatory changes on colonoscopy, 14 did not). In contrast to the findings in this study Savage et al. described a raised serum ASCA in CD patients with inflammation elsewhere in the GI tract but not in those patients with oral inflammation only. The findings of this present study are however more robust, bearing in mind Savage et al. examined a much smaller number of patients overall and the findings of the study were only subject to unifactorial analysis.

Previous studies in adult CD cohorts have found that positive ASCA status is found predominantly in those with more 'proximal' disease e.g. small intestine.[305;316] Walker et al. in a previous study of 143 Scottish adult CD patients demonstrated ASCA was associated with gastro-duodenal and small bowel CD compared to colonic CD. In a study of 100 CD patients Quinton et al. demonstrated ASCA was more common in CD patients with small bowel involvement. ASCA status has been

linked with ileal disease, [151;214;306;316] and the upper GI disease(L4) location of the Vienna classification by other groups.[305;317] Lack of replication of association with oral CD involvement as demonstrated in this study could be accounted for by the under-representation of patients with oral CD in other studies (it is not recognised as a disease location in the Vienna classification).[36] It is also possible that the contribution of oral CD to ASCA status was missed due to difficulties in the diagnosis of oral involvement if expert opinion was not sought.[314] In a prospective study of 48 CD patients Harty et al. demonstrated paediatric dentists were experts in diagnosing oral manifestations of CD but paediatric gastroenterologists missed the diagnosis in around half of cases. This current study has addressed the problems highlighted by previous studies by using an extensive description of disease phenotype, using only well defined cases of oral involvement by selecting cases where granulomata have been demonstrated on buccal biopsy or diagnosis has been confirmed by a paediatric dentist only, in addition to using a large study population.

ASCA status in CD patients has been associated with increased disease activity (measured by Paediatric Crohn's Disease Activity Index [PCDAI]) and disease severity.[305;318] In the present study ASCA status is a marker of severe CD with unifactorial associations with raised CRP, low albumin and surgery (replicating findings for NOD2/CARD15 mutations, chapter 4) and additionally in this study with hypoalbuminaemia on multifactorial analysis. The association between positive ASCA status and the need for surgery[304-306;308;319;320] and NOD2/CARD15 status and need for surgery[193;321] has also been demonstrated by other groups. Other groups have clearly demonstrated ASCA positivity is associated with more rapid disease complications with higher rates of stricturing and penetrating disease independent of any NOD2/CARD15 associations.[322;323] In a small study of 34 UC patients undergoing pouch surgery post operative complications were more common in patients who were ASCA positive.[324]

Several antibodies to other antigens have been described to be higher in patient with CD : OmpC (antibodies to the outer-membrane porin C of E.Coli), I2 (antibodies against a pseudomonas fluorescens-associated sequence) and cBir1(anti-flagellin antibodies).[325;326] Most recently a number of antiglycan antibodies have also been

described to occur more frequently in CD patients.[327] Positivity to not only ASCA but to the other microbial antigens has been shown to be a predictor of disease progression in a recent large paediatric study.[328] Dubinsky et al. studied 196 paediatric Crohn's disease patients and determined their anti-I2, anti-OmpC, anti-CBir1 as well as their ASCA status. Progression of disease behaviour to penetrating or stricturing type was predicted by positivity to one or more serum markers with greatest risk in those patients who were positive to all four markers.

Routine blood tests (FBC,ESR,CRP and albumin) will be abnormal in most children with IBD and so performing serological markers has little additional diagnostic benefit.[223;329] Beattie et al. demonstrated in 39 IBD patients and 37 controls at least one of these routine blood tests was likely to be abnormal. Khan et al. studied 90 IBD patients and studied ASCA and p-ANCA in addition to the previously mentioned routine blood tests and found little additional diagnostic benefit when serology was performed.

The results of this study demonstrated similar sensitivity and specificity to ASCA assays used in other published paediatric and adult studies as highlighted in table 8-6. However the lack of sensitivity demonstrated in this and other studies with increased ASCA prevalence found in other diseases like Behçet's disease, primary biliary cirrhosis, autoimmune hepatitis, and coeliac disease, has prevented ASCA status - becoming a useful *diagnostic* tool in UK clinical practice.[330]

Thus although ASCA has questionable clinical benefit as a diagnostic tool it does seem to characterise a specific more severe disease phenotype and thus could be used clinically in stratifying treatments in future studies in patients with Crohn's disease.

Why should ASCA be higher in patients with oral Crohn's disease in particular? It has recently been suggested that that the immunogen for the production of ASCA is *Candida albicans*.[331] Candida colonises the entire gastrointestinal tract but initial exposure will happen first in the mouth and the proximal gastrointestinal tract. With the similarities described between disease phenotypes predicted by NOD2/CARD15 and ASCA it could be argued they are representing the same disease mechanism. NOD2/CARD15 mutations represent a defect in the innate immune system triggered by specific ligands on bacterial cell walls.[101] ASCA may represent a specific,

genetically determined breakdown in a mechanism to innate tolerance to yeast or fungal molecular patterns, to which the intestinal epithelium is exposed. Evidence for ASCA representing a defect in the innate immune system has been demonstrated by its association with lower levels of mannose binding lectin(MBL), a component of the innate immune system.[332] Seibold et al. determined ASCA status plus screening for MBL variants in 58 CD patients, 18 UC patients and 47 healthy controls. ASCA positive patients were more likely to have MBL mutations and additionally were found to have lower serum MBL levels compared to ASCA negative patients.

Other investigators have demonstrated an association between ASCA status and the presence of NOD2/CARD15 variants[214;311;312] with demonstration of a gene dosing effect in one study.[311] This relationship however has not been universally replicated and as well as this study other groups including Scottish adult patients have been unable to replicate these findings.[216;305;308] The most recent study to address this issue has demonstrated patients and unaffected relatives who possess NOD2/CARD15 mutations are more likely to demonstrate positivity to one or more of the panel of microbial antigens.[333] The lack of replication in this study may in part be explained by the lower carriage of the NOD2/CARD15 in the Scottish population,[255] "under powering" investigation of any such relationship.

The role of ASCA serology in defining patients who are more likely to have CD than UC is demonstrated robustly in this study. The more interesting data however were with regard to the small group of patients (n=28) with IC none of whom had a positive ASCA, replicating previous paediatric data (n=7 patients) published by Gupta et al.[320] This data also replicates studies from Joosens et al. in a larger adult series of 97 IC patients that demonstrated patients who are ASCA negative tend to retain their diagnosis of IC over time, with a median follow up time of ten years in Joosens study.[242] Indeed in a follow up study of the same patient group if a wider panel of antibodies are determined (anti-I2 and anti-OmpC) a significant number of patients appear to be negative to all antibodies and remain so over an extended follow up period.[241]

Table 8-6: Sensitivity and specificity of ASCA in previous publications

Study	**Sensitivity**	**Specificity**	**CD Population Description**
Present study	41.6%	88.6%	Scottish paediatric cohort (n = 197)
Canani et al 2004[318]	58.8%	92.5%	American paediatric cohort (n=102)
Zholudev et al 2004[306]	44%	98-100%*	American paediatric cohort (n=81)
Ruemmele et al 1998[302]	55%	100%	Canadian paediatric cohort (n=130)
Walker et al 2004 [305]	57%	87%	Scottish adult cohort (n= 143)
Vermeire et al 2001[334]	41-76%**	86-98%**	Belgium adult cohort (n=100)
Quinton et al 1998[335]	57%	97%	French adult cohort (n=100)
Koutroubakis et al 2001[336]	39%	89%	Greek adult cohort (n=56)

*The specificity figures quoted was for ASCA IgA and ASCA IgG respectively.

** >1 form of ASCA assay was used in this study.

In conclusion ASCA prevalence is increased in patients with CD compared to patients with other types of IBD and healthy controls. Detailed phenotypic examination has allowed demonstrated of novel phenotypic association with oral CD on multifactorial analysis. The challenge now facing researchers and clinicians is how to incorporate these findings into the current model of disease pathogenesis and to explore the clinical applications of ASCA serology in treating patients with IBD.

9 Implications, further and future work

9.1 Implications

This thesis set out to investigate the contribution of several genes and their effect on disease susceptibility and phenotype in children with IBD. I was successful in collecting a suitable cohort of children to study the genetics of inflammatory bowel disease and collecting one of the largest such cohorts available in the world for such studies in children. After the childhood cohort was collected along with parental DNA the major genes implicated in disease susceptibility at the start of, and during the course of, the thesis were studied. The contribution of the NOD2/CARD15 gene which has been implicated in Crohn's disease populations throughout the Western world was thoroughly assessed and this gene does not explain the major genetic contribution to disease susceptibility in Scottish children. Subsequent studies of IBD5 and DLG5 did demonstrate a role in increasing susceptibility to IBD but their overall contribution was in a similar manner to NOD2/CARD15, relatively modest.

9.2 Regression analysis of relative contribution of CD susceptibility genes in Scottish childhood onset IBD

To clarify the exact contribution each gene made to disease susceptibility in the Scottish early onset population a regression analysis was undertaken. Applying a forward-Likelihood Ratio Stepwise regression model, genotypes and/or carriage status of the candidate genes studied in this thesis were analysed, namely: the 3 common NOD2/CARD15 polymorphisms (R702W, G908R and Leu1007fsinsC), OCTN1/SLC22A4 1672 C/T, OCTN2/SLC22A5 –207 G/C as well as 3 SNPs in the extended IBD5 haplotype (IGR2096a_1, IGR2198a_1, and IGR2230a_1), DLG5 113 G/A, together with a NOD1/CARD4 SNP (NOD1+32656)[337] and several TNF-α promoter SNPs (–1031T/C/-863C/A/-857C/T/-806C/T)[338] studied subsequent to the conclusion of this thesis. All SNPs were analysed in the Scottish early onset population for their relative contribution to CD susceptibility. Only one, the NOD2/CARD15 Leu1007fsinsC contributed significantly to CD susceptibility (p=0.009 OR 7.71 CI 1.65-35.98). This clarifies that further work is needed, continuing on from the findings of this thesis to determine the major genetic

influences in the Scottish early onset population that determine susceptibility to Crohn's disease.

The data contained within the thesis has explained very little of the genetic contribution to either Ulcerative colitis or indeterminate colitis. This is as a result of several inter-related factors: Crohn's disease from family studies has the largest genetic contribution to disease susceptibility compared to the other two types of IBD; [48] the number of patients in childhood IBD populations is proportionately less than the number of patients with Crohn's disease with double the number of CD patients compared to UC and 6 times more CD than IC. To explore susceptibility in these groups thoroughly, a much larger cohort of patients with these disease types would have to be collected.

Despite these limitations novel associations have been identified, homozygous carriage of the TC haplotype was found to be more common in UC patients, a result consistent with the fact associations with IBD5 and UC have been demonstrated by several other research groups.[247;263;339]

The only association found in the thesis for patients with indeterminate colitis was increased susceptibility in patients carrying V955I variants of the NOD2/CARD15 gene. This association has not been examined by any other research group to date and has to be interpreted cautiously given that that were only a small number of patients (n=28) in the IC group.

The successful replication of the three candidate genes within this thesis (NOD2/CARD15, IBD5 SNPs and DLG5) has been based on calculations using standard statistical values with a p value of <0.05. All three candidate genes have achieved this level of significance but the weakness of these findings was they were based on pragmatic numbers of patients recruited to the study rather than working to previously calculated power calculations. This traditional statistical threshold is therefore scientifically lenient in studies that are examining the genetics of complex chronic diseases.

The "winners curse" has been demonstrated many times in such studies where a candidate gene identified in an index study has not been replicated in larger subsequent studies, suggesting the initial finding was a false positive thus

demonstrating the weakness of employing traditional statistical modelling.[340] To avoid such problems Wacholder et al. proposed analysing candidate gene studies using the *false positive report probability*, where in addition to the p-value generated in the study consideration is also placed on the prior probability of genetic association and a power calculation.[341] This will become an even more important consideration as candidate genes generated by genome wide association studies (see below) are described where the generated odds ration for candidate genes in some studies has been less than 1.5. The strength of replication of candidate genes within this thesis is that the results have been replicated in many different populations,[342] and as such are likely to represent true genetic associations rather than the fact they reached the "traditional" statistical threshold when analysed in the Scottish childhood IBD population.

9.3 Novel methods for identification of candidate genes

9.3.1 Genome wide association studies

Most recently, catalysed in part by technological advances allowing cheaper genotyping, and the completion of the HapMap project, genome wide association studies have been used to identify IBD genes.[343] In these studies a large number of SNPs are studied across the whole genome in disease cases and compared to healthy controls providing a collection of hypothesis free candidate genes.[344] The first such IBD study to be performed was in the Japanese IBD population where Yamazaki et al. studied 80,000 SNPs across the genome. They reported an association between CD and the TNFSF15 (Tumour Necrosis Factor Superfamily, member 15) on 9q33.[345] They then replicated their findings in cohorts of European IBD patients.

Duerr et al. performed a genome-wide association study testing 308,332 SNPs in 547 ileal CD patients and 548 controls of non-Jewish European ancestry. This study implicated germline variations of interleukin-23 Receptor (IL23R) on Chromosome 1p31.[346]. Replication was obtained in a Jewish-ancestry ileal CD case-control cohort and by transmission disequilibrium testing in 883 families with IBD, the later demonstrating association for all types of CD as well as UC.

Although the IL23R gene data were published first the full GWAS that the IL23R data were derived from, with results for other candidate genes were published subsequently.[347] Rioux et al. analysed 304,413 SNPs in the patient group described in the IL23R study with ileal CD containing subjects with both Jewish and non-Jewish ancestry. This studies most significant results suggested a role for an intergenic area on 10q and for the ATG16L1 gene (see below).

In a German-British collaboration, Hampe et al. studied 19,779 SNPs in 735 Crohn's disease patients and 368 controls.[348] They demonstrated an association with a coding variant of ATG16L1 (Autophagy-related 16-like 1) gene, on 2q37.1 and replicated their findings in a family based association study and in an independent case control cohort.

In a further GWAS study from Belgium, Libioulle et al analysed 311,882 SNPs in 579 Belgian CD patients and 928 French and Belgian controls.[349] This study replicated the findings for NOD2/CARD15 and IL23R but described a novel CD susceptibility locus at 5q13 at the site of a gene desert. This large 250kb area is devoid of genes and separated into 5 haplotype blocks with the strongest signal located in block 3. The association with CD was replicated in an independent case-control cohort and within family trios. The authors presented some provisional data to suggest that a gene located at the edge of the gene desert was the important candidate gene from within this susceptibility locus namely the prostaglandin receptor EP4 gene.

The Wellcome trust case consortium have performed a GWAS of 7 common diseases of which Crohn's disease was one. [350] This study has replicated the findings of previous studies and presented evidence for novel candidate genes. This study examined 500,568 SNPs in 16,179 subjects of whom 2000 had Crohn's disease (collected from 5 UK centres- Cambridge, Oxford, London, Newcastle and Edinburgh) and 3,000 of whom were healthy UK population controls. This study replicated data for candidate genes identified by GWS (NOD2/CARD15 and IBD5)[106;109] and data from published GWAS (IL23R, ATG16L1, the 10q21 intergenic region and the 5q13 gene desert).[347-349;351]

The study also presented preliminary data on four novel candidate genes immunity related guanosine triphosphatase (IRGM), bassoon (BSN) gene, a cluster of SNPs around chromosome 10q24.2 and an area on chromosome 18p11upstream of the protein tyrosine phosphatase, nonreceptor type 2 (PTPN2) gene. Interestingly the PTPN2 gene was also identified in patients with type 1 diabetes and to a lesser extent in patients with Rheumatoid arthritis. In a subsequent publication, Parkes et al. examined the data on the four novel regions identified in the scan and examined them in independent case control studies.[352] These extension studies have provided further data supporting a role for IRGM and PTPN2 genes, as well as several other genes including IL12B a key player in the IL23/IL12 pathway.

While these studies represent a great advance in identifying genes important in IBD and other common disease they will not identify rare variants. Future efforts to capture all variants could be addressed by performing whole genome resequencing in cases and controls, but currently these studies are technologically and financially unachievable.[353]

9.3.2 Gene wide association studies

During the lifetime of the thesis researchers have moved away in the majority of cases from studying single or several SNPs within genes, to haplotype-tagging based approaches to comprehensively assess the gene-wide contribution to the disease being studied. The use of a gene-wide haplotype-tagging strategy has already proven successful in the analysis of other IBD candidate genes, in which equivocal evidence had been provided by single SNP analyses in different populations. This method of gene assessment has become feasible only since the full publication of the HapMap project data.[343]

Applying this approach to capture the haplotypic variations of the Multi Drug Resistance 1 (MDR1) gene, Ho et al. observed a highly significant association in the Scottish adult population between the common haplotypes of MDR1 and UC ($p=4.22 \times 10(-7)$) but not CD ($p=0.22$).[354] These data provided really clear evidence of an important contribution to susceptibility and phenotype, in the face of inconsistent reports from other populations, and have been supported by publication of a

subsequent MDR1 meta-analysis.[355] Ho and colleagues also confidently refuted the role of the pregnane X receptor gene (PXR/NR1I2) in IBD susceptibility in the same population, after application of a gene-wide haplotype tagging approach.[356]

In a further application of this technique in the study population collected for this thesis, Van Limbergen et al. has used this approach to move on from single SNP analysis in the NOD1/CARD4 gene,[337] to comprehensively refute a role for this gene in the Scottish early onset population.[357] McGovern et al. initially suggested the association of the deletion variant of a complex intronic insertion*2/deletion*1 polymorphism (32656) of NOD1/CARD4 with susceptibility to IBD using a combination of TDT and case-control analysis.[358] However, well-designed replication studies involving large numbers of IBD patients in four different populations have been negative, casting doubt on the contribution of this NOD1/CARD4 SNP and gene.[337;359;360] In support of these negative susceptibility studies, a gene wide association study was performed in 356 Scottish children with IBD using 9 marker SNPs based on HapMap data, with no association demonstrated using three complementary methods of statistical analysis namely, single and multiple marker case-control analysis, TDT analysis and log-likelihood analysis.

9.3.3 Yeast two hybrid studies

The yeast two-hybrid is a disease model that investigates protein-protein interactions and has also been employed to try and identify new genes important in IBD.[100] When investigating these 'protein- protein' interactions, a single "bait" protein is used to search for interaction with a library of proteins fused to the activation "prey" domain.[361] In studies of IBD genes an intestinal cell library has been used. This yeast two hybrid model has been used successfully in IBD, identifying interactions with the NOD2/CARD15 gene. Barnich et al. identified the GRIM19 protein interacting with NOD2/CARD15,[362] while Yamamoto-Furusho et al. identified Centaurin [beta]1 that interacts with both NOD1/CARD4 and NOD2/CARD15.[363]

9.3.4 Microarray Studies

Gene expression technology using gene microarrays allows a comprehensive picture of gene expression at the tissue and cellular level, thus helping understand the underlying physiological and pathological processes in IBD patients.[364] In these studies RNA is isolated from the tissue sample or cells that are to be examined e.g. intestinal tissue, and this is then used to generate cRNA or cDNA. These samples are then mounted on a "chip" and the arrays are fluorescently labelled so that they can be analysed by confocal laser scanning.[364]

In one of the first IBD micro-array studies, Lawrance et al. examined global gene expression profiles of inflamed colonic tissue in IBD patients.[365] They identified several hundred genes with altered expression both increased and decreased, that had not previously been linked to IBD. The results demonstrated distinct gene profile differences between patients with CD and UC. In further microarray studies, again using biopsy samples, Wu et al. have demonstrated different gene expression profiles when comparing inflamed and non-inflamed IBD samples, different gene profiles between non-inflamed IBD samples and healthy controls and again demonstrated distinct gene expression patterns between inflamed CD and UC samples.[366] In further microarray studies in the healthy adult colon, the gene expression profiles differ markedly between the left and right side of the colon, this effect is primarily driven by genes involved in developmental pathways (Noble CL, unpublished work).

9.4 A Northern European founder gene?

It is possible that any gene identified in the Scottish childhood (or adult) IBD populations in Scotland may be important not only for Scottish patients but also patients from other Northern European countries. Clear heterogeneity exists in carriage of NOD2/CARD15 variants in Northern Europe compared to Southern Europe where carriage rates are significantly higher.[101] It is intriguing, and indeed, may seem superficially paradoxical that the countries of Northern Europe where the incidence of Crohn's disease and Inflammatory bowel disease in children are highest,[9;10;12;367] have the lowest carriage of the commonest disease susceptibility

gene identified thus far- NOD2/CARD15. However, this means the identification of a "Northern European" founder gene currently remains a distinct possibility.

9.4.1 Genetic studies in the Swedish IBD population

In collaborative work with Swedish colleagues, (Dr L Torkvist, Karolinska institute, Stockholm) the contribution of genes in a second Northern European IBD population has been studied. The contribution of the three common NOD2/CARD15 mutations to disease susceptibility was studied in 178 Swedish adult CD patients and 143 healthy controls.[201] The results produced were very similar to the Scottish IBD population with a population attributable risk (PAR) of 11% matching the equivalent figure in the Scottish adult population and similar to the Scottish childhood population figure of 8%.[184;201]

Beyond studies of the NOD2/CARD15 gene the role of two further genes have been analysed in the Scottish and Swedish populations. Variants of OCTN1/2 together with IBD5 marker SNPs were studied in 178 CD patients and 143 controls and demonstrated susceptibility to CD, again with a similar degree of magnitude to that of the Scottish population.[253;368] The NOD1/CARD4 +32656 insertion/deletion SNP has been implicated in IBD susceptibility by McGovern et al,[278] has demonstrated no contributions to disease susceptibility respectively in 1791 IBD patients compared to 1649 healthy controls in a study that included both Scottish and Swedish adults with IBD and 313 Scottish children with IBD.[337]

Thus the data in two Northern European IBD cohorts are remarkably similar and have yet to explain the majority of genetic risk within either of these two Northern European populations.

9.4.2 Searching for a major Northern European susceptibility gene

The lack of genome wide scans in Northern European populations (with the exception of a small UC dominated Finnish scan)[369] means other methods of identifying candidate genes will need to be employed. Since commencing this thesis the methods available for identifying IBD genes have progressed rapidly, as already

discussed. So, rather than conducting a genome wide scan in a large Northern European population of relative pairs, as would have seemed logical at the start of the thesis, it would now be technically possible, and scientifically more robust to perform a genome wide association scan using Northern European populations of IBD patients and population controls. A genome wide association scan would seem to offer the best chance of novel gene identification in Northern Europe based on both current scientific knowledge and on practical numbers of subjects available to study.

An alternative approach would be to perform a genome wide association scan in an exclusive early onset population. This would mean collaboration between several groups of investigators in order to get an appropriately powered study, especially if the aim was to identify novel UC genes. An appropriately powered study to identify genes of modest effect would need to study at least 1000-2000 children with IBD plus controls.

9.5 Childhood vs. adult studies

One of the major hypotheses explored in this thesis was looking for a genetic difference between the childhood and adult IBD populations. There was preliminary evidence from family studies that disease starting in childhood had a stronger genetic contribution,[48] and evidence for the first two genes studied in the thesis IBD1 and IBD5, that children would be expected to have a higher mutation frequency.[94;112] This hypothesis has not been supported by the data in this thesis with respect to NOD2/CARD15 as no difference has demonstrated between adult and paediatric CD populations in Scotland (table 9-1). The same lack of effect is demonstrated for IBD5 and DLG5 by comparison of published figures for the childhood and adult populations in Scotland. It is interesting to note however that the ASCA prevalence was significantly *lower* in the childhood CD population although this serological marker clearly does not definitely reflect a genetic susceptibility towards disease development.

The similarities in genotype contrast starkly with disease phenotype which differs markedly between both populations. Childhood onset CD is a male dominated disease, more likely to result in Crohn's disease of the upper GI tract. Children also very clearly have more extensive ulcerative colitis than adults. Further examination of these differences should be explored by examining further candidate genes in both populations, standardising methods of investigation and comparing disease phenotypes in both populations over time.

As Rioux et al demonstrated for the analysis of the genome wide scan investigating the IBD5 locus it is possible to stratify the results of a scan by age at diagnosis.[112] The same technique could be applied to the GWAS that have been performed to date either using results from an individual scan or combining the results into a meta-analysis.

Table 9-1: Childhood vs. adult studies

Genetic/serological marker	**Carriage rate of variant alleles**	**p value children vs. adults**
NOD2/CARD15[184;255]	19.8% vs. 22.6%	p=0.48
IBD5[253;370]	82.7% vs. 76.8%	p=0.14
DLG5[289;371]	23.3% vs. 18.4%	p=0.15
ASCA[305;372]	42% vs. 57%	p=0.006

9.5.1 Future genetic studies and further work

This thesis has produced exciting data based on studies of the NOD2/CARD15, IBD5 and DLG5 genes in the Scottish early onset population. The relatively low numbers of patients in the thesis means all of the data would need to be replicated in other populations before the findings could be considered scientifically robust, this is despite the fact that the collection of patients presented in this thesis is one of the largest such collections collected in the world to date.

Further work to fully assess the genetic contribution to IBD could focus on the genes of the innate immune system given the rapid progress in understanding the importance of innate immune defects in IBD development,[102] performed after the detailed studies of the NOD2/CARD15 gene e.g. NOD1/CARD4, RIPK2/RICK and the TLR genes.

As mentioned further candidate genes could be identified by employing complementary gene finding strategies e.g. genome wide association studies, yeast two hybrid analysis and microarray studies. These human studies can be complemented by animal models of IBD using knock in and knock out mice to study the role of specific candidate genes.

Whatever methods of gene detection are employed it is clear collaboration between interested groups will be needed to study further candidate genes in detail. In the era of GWAS this means studying 2000-3000 patients and so will only happen in paediatric IBD studies with international collaboration between at least 4-6 national projects.

9.5.2 Disease site in Crohn's disease and relationship to genetic/serological markers

A summary of the relationship between the Crohn's disease locations within the gastrointestinal tract and the genes/serological markers studied in this thesis is shown in figure 9-1. The proposed phenotypic classification for use in paediatric genetic studies allows the description of both well recognised and novel genotype-phenotype associations.

Thus patients who carried one or more of the common variants of NOD2/CARD15 (Chapter 4) were more likely to have disease located in the small bowel, both jejunal (novel) and ileal (well recognised) but particularly with both sites combined (novel). Similarly Crohn's disease patients who possessed the TC haplotype, within the IBD5 locus (Chapter 6), were more likely to have disease in the gastric antrum, another novel observation. No relationship with disease location was demonstrated for DLG5 variants described in chapter 7. Patients with positive ASCA serology were more likely to have oral Crohn's as described in chapter 8.

Further studies in the Scottish early onset population by co-workers have added to this description of disease markers linked to specific sites of disease. Van Limbergen has demonstrated that the complex insertion deletion of the NOD1/CARD4 maybe associated with disease located in the gastric body ,[337] and in other work lead by the same author TNF-α promoter mutations have been linked with colonic disease location.[338]

9.5.3 Genetic associations with disease location, true finding or false association?

The description of association between sites of disease activity for most of the genes studied in this thesis, begs the question as to whether these are true findings or simply represent statistical flaws from multiple testing. The NOD2/CARD15 association with ileal disease is widely described and likely to be true, having been described in many studies and in a subsequent meta-analysis.[185] This is further supported with the description of the association of NOD2/CARD15 mutations with Paneth cells and defensin production which are located within the ileum.[228] Whether the association is only with ileal disease or whether it extends to jejunal disease as described in chapter 4 remains to be seen, but given the functional similarities between the ileum and jejunum this remains a plausible association that needs to be explored in other CD populations. The relative inaccessibility of the jejunum to biopsy means relying on assessing disease presence by the rather "blunt" method of small bowel follow through and perhaps this can only be fully explored when capsule

endoscopy that will pick up more subtle lesions, is used commonly in clinical practice.[373]

The association of IBD5 variants with gastric antral disease is a unique finding of this thesis in contrast to the well replicated evidence of NOD2/CARD15 mutations with ileal Crohn's disease. Paradoxically, the IBD5 association was described on multifactorial analysis (p=0.003, OR= 4.02 [1.58-10.21]) in contrast to the NOD2/CARD15 that was only described in this thesis, at least, by unifactorial analysis so *statistically* speaking the association is actually more robust. As the exact gene within the IBD5 locus is yet to be determined is harder to speculate at present as to the physiological basis of any such association.

In a similar manner to IBD5 the association of positive ASCA serology with oral Crohn's disease location is described on multifactorial analysis. ASCA serology has been associated with a variety of disease locations most notably with small bowel disease in several studies. Few investigators have described an association with oral Crohn's disease as patients with oral Crohn's disease were not analysed as part of their studies. The one publication that did investigate this association was a small study of 26 patients that did describe a positive correlation between a positive salivary IgA ASCA and oral CD.[374]

There are some limitations to the data that have been presented here because of the multiple testing of different disease sites with the various candidate genes. The robustness of these findings are discussed individually but without replication these could still be chance findings. Applying corrections for multiple testing like a Bonferroni correction has been considered too strict for genotype-phenotype analysis by many investigators because of the interdependency of the variables being examined. In the thesis to fully explore possible relationships no corrections were made for multiple testing initially but any findings on unifactorial analysis were further examined by multifactorial analysis. In reality the true validity of the findings will only become apparent when they are examined in independent groups of patients with childhood onset IBD.

Thus the associations of genes with specific disease locations form an emerging body of evidence that demonstrates Crohn's disease has many sub-phenotypes

characterised by specific genes or serological markers. These findings need to be further explored in independent cohorts with accompanying rigorous phenotypic data to allow these interesting data to be confirmed or refuted.

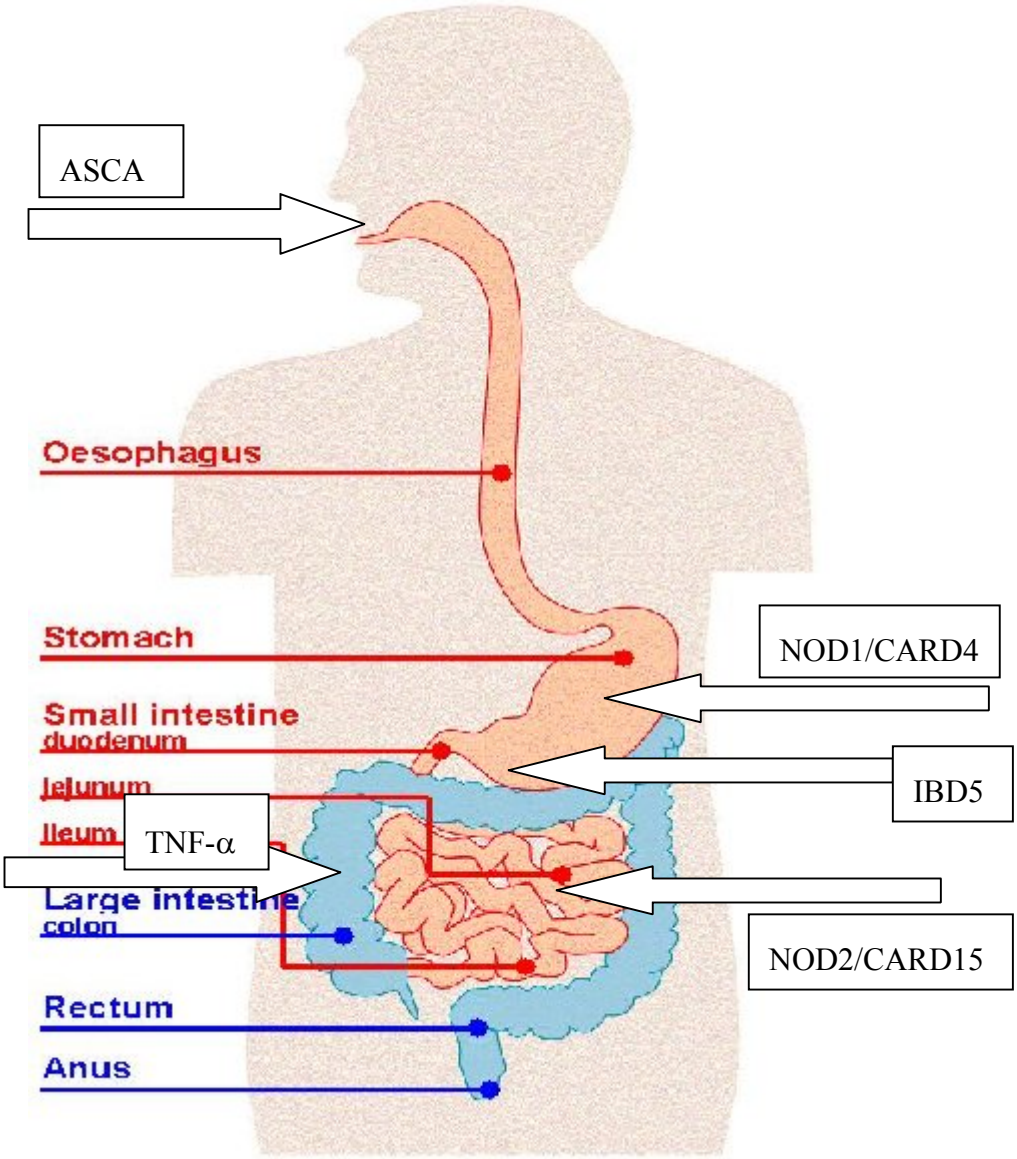


Figure 9-1: The association of various genetic and serological markers with disease location

9.6 Moving towards a uniform paediatric phenotype

There is as yet no recognised disease classification system for children with IBD.[37] Childhood onset CD does not fit well into the adult phenotypic classifications devised to date. The hierarchical disease classification designed for use in adults (Vienna) has resulted in most paediatric patients with a disease location that fell into the upper GI (L4) category potentially disguising any underlying genotype phenotype relationships.[36] The shortcomings of the Vienna classification have to some extent been addressed in the newer Montreal classification, which includes adoption of a non-hierarchal system and specific recognition of a separate early onset disease category (<17 years).[35] Despite the relative shortcomings of these phenotypic classifications when applied to paediatric CD populations, they have set standard definitions for disease location and disease behaviour making preliminary steps towards uniformity of reporting for genotype–phenotype analysis in paediatric genetic studies.

These classification systems however do not incorporate one of the major phenotypic variables present in childhood studies, namely growth. There is a strong argument to be made for an agreed and validated paediatric phenotypic classification, especially for Crohn's disease. The lack of an agreed system means in practice that many genetic studies in children do not report disease phenotype or report it in a variable and inconsistent manner making replication of the findings between different studies difficult.

The first steps towards a paediatric phenotypic classification have been taken in Europe at least with agreement on a good practice for investigation of new patients with IBD (the Porto criteria)[158]. All patients are recommended to be investigated with an upper GI endoscopy, ileo-colonoscopy and a barium follow through. In this thesis the studies were reported using an adult classification system[35;36] plus a specifically devised paediatric classification that is more extensive but as yet remains unvalidated in other paediatric studies.[255;370;371]

9.6.1 A framework for developing a paediatric disease phenotypic classification

9.6.1.1 IBD

Before designing a phenotypic framework for children it would have to be agreed what a "child" is. There are, as previously mentioned in chapter 2, many different age definitions, but in the absence of strong evidence it would seem sensible to fall in line with the Montreal classification and use age <17 years at diagnosis. The question as to whether within childhood age should be further subdivided is not directly addressed in this thesis. However, given the strong evidence produced by Heyman et al.[41] and Meinzer et al. [164] complemented by the data in this thesis about a different disease phenotype at age eight and below, it would seem dividing age at diagnosis into two age groups, 0-8 years and 9-16 years is reasonable.

All patients with IBD should then have some general phenotypic parameters noted. The classification should incorporate growth parameters using z- scores for weight, height and BMI. An assessment of puberty should be made using a standard pubertal classification system, although we failed to do this well in the thesis simply because the data was not recorded. Surgery should be recorded as any operation, excluding an EUA but the type of surgery performed should also be noted e.g. drainage, resection etc.

9.6.1.2 CD specific

A basis for a paediatric CD phenotype based on this thesis would involve defining Crohn's disease sites throughout the GI tract individually based on evidence from endoscopy, at the time of surgery, biopsy or barium follow through and would be independent of disease activity elsewhere in the GI tract. The disease behaviour classifications used in the Montreal classification would be satisfactory. This phenotypic classification is vindicated by the description of disease correlations with genetic markers that would have been missed by only sticking to either the Vienna or Montreal classifications. These novel correlations (as mentioned) include the

association of IBD5 variants with indices of growth in children with CD[370] and the association of NOD2/CARD15 mutations with jejunal disease location.[370]

9.6.1.3 UC specific

In assessing disease location it seems sensible to include the three disease locations in the Montreal classification of proctitis, left sided disease and extensive disease but also include a fourth category of pancolitis with inflammation of the entire colon given that this is such a common phenotype in children.

The thesis has not been able to assess the minimum period over which to re-assess changes in disease phenotype, given that IBD and especially Crohn's disease is dynamic over time. This study adopted a 2 yearly assessment in Crohn's disease patients, but the time interval which will adequately reflect changes in disease phenotype over time has not yet been well studied in paediatric patients. This question can only be addressed by long term follow up of the patients in this study and is beyond the time scale of this thesis.

No disease classification would be perfect but any classification adopted by the paediatric IBD community could be modified once used in clinical studies in a similar manner to the Vienna classification being modified into the Montreal classification.

9.6.2 Clinical implications

The translation of basic scientific into clinical practice is the biggest challenge facing clinician scientists involved in IBD genetics, none more so than those studying disease with onset in childhood. At present the aspiration of using genetics to influence diagnosis or treatment in IBD has not yet been fulfilled. The use of a good clinical history and a simple set of blood investigations is enough to screen children with suspected IBD in most cases. Patients are keen for translation of IBD genetics into clinical practice with 90% of 250 IBD patients interviewed expressing enthusiasm for genetic testing of both themselves and their children.[375] A genetic panel in combination with known environmental factors has been proposed by some groups but is not likely to be widely useful in clinical practice.[376] Attempts at

predicting response to medications used in IBD have generally been disappointing with negative or unreplicated results being common place thus far.

The work described within this thesis has definitely added to the data and evidence for stratifying Crohn's disease into different subtypes. NOD2/CARD15 variants were clearly associated in this thesis with markers of severe disease most notably with a defined clinical outcome-surgery. This finding has also been consistently replicated by other investigators. The association with surgery has been very clearly demonstrated in another paediatric study in the USA in a similar study group size to the Scottish paediatric study and with remarkably similar results.[193] These findings of association with NOD2/CARD15 variants and increased need for surgery have also been replicated in adult studies.[215;321;377]

The TC haplotype in the IBD5 locus also demonstrated association with a more severe disease phenotype manifest as lower weight/BMI at diagnosis and follow up. These data demonstrating association with growth parameters have not been replicated, however adult studies have replicated the association of IBD5 locus variants, with a more severe disease phenotype.[253;260] So, these mutations could potentially, at least, be used to stratify treatments in paediatric Crohn's disease studies with those carrying variant alleles being given immunosuppressants/biological treatments earlier in the disease course.

In summary, this thesis marks the end of the beginning in developing an understanding of the genes that underlie genetic susceptibility and disease phenotype in IBD patients in the Scottish childhood population. A fuller understanding of the genetic contribution to disease development will be gained by further studies examining other candidate genes recognised by the growing array of methods being used for gene identification followed by studies confirming changes in gene expression and function. Genome wide association studies look to be the most exciting method for identifying novel genes at present. Evermore studies in single disease populations are becoming less common and as further groups undertake such research studies in the childhood population only by collaboration between research groups will knowledge of importance to all children with IBD be truly fulfilled.

10 Appendices

10.1 Appendix 1: Lennard-Jones Criteria-check list

Study Number ☐

Chronic granulomatous lip	**1**	**Yes / No / Unknown**
Pyloroduodenal disease		**Yes / No / Unknown**
Small bowel disease		**Yes / No / Unknown**
Chronic anal lesion		**Yes / No / Unknown**
Discontinuous lesions	**2**	**Yes / No**
Fissuring ulcers	**3**	**Yes / No**
Abscess		**Yes / No**
Fistula		**Yes / No**
Strictures	**4**	**Yes / No**
Aphthoid ulcers	**5**	**Yes / No / Unknown**
Lymphoid aggregates		**Yes / No / Unknown**
Colonic mucin retention	**6**	**Yes / No / Unknown**
Granulomata	**7**	**0 / 1-2 / 2-5 /6-10 / > 10**

Number of section criteria met ☐

Total number of criteria met ☐

10.2 Appendix 2: Patient information leaflet

The Genetic Influence In the development of Crohn's disease and colitis in children

We invite you to participate in research which we think may be important. The information which follows tells you about it. Try to make sure you understand what will happen to you if you decide to take part. Whether or not you do take part is entirely your choice. Please discuss this with your family and ask any questions you want to about the research and we will try our best to answer them.

We would like to understand more about why people especially children develop Crohn's disease and colitis. These diseases are becoming more common in children especially teenagers and we want to try and work out why this is. We know that having a relative (either a parent or brother or sister) with Crohn's disease or colitis, gives you a higher chance of getting the disease. We want to examine the things that are passed onto you by your parents (your genes) and how these may result in the development of Crohn's disease or colitis.

You are going to undergo tests on your blood having been seen by one of the doctors in the clinic.

What we are asking

We would like to ask your permission to collect some extra specimens (two to four teaspoons of blood) which we can use for research in our laboratories. We are also asking both of your parents for blood samples too.

What are we going to do with them?

We are going to use these specimens to help research into your genes and the part they play in the development of Crohn's disease and colitis. Some of these specimens will be tested in the laboratory soon after being taken while others may be stored in a freezer before testing. Any specimens stored will only be used for studies which examine your genes and the development of Crohn's disease and colitis.

All our research is intended to increase understanding of how and why these diseases occur, how treatments work and how we may be able to improve treatments in the future.

We also plan to study how genetic make-up may lead to children getting Crohn's or colitis and whether this make-up can alter how effective the treatments are.

How will this affect me?

We will always collect specimens needed for your planned tests before any research specimens. If there are any problems during the collection of these specimens i.e. in collecting blood then the doctors will not collect specimens for research.

Taking extra blood will not cause you any problems. The research itself is not relevant to your care. If in the future we would like to perform research that may be directly relevant to you we will contact you to see if you agree.

Do the specimens have my name on them?

The people in the laboratory will not know your name as your specimens will have a number. Only Dr Russell or ward or outpatient staff supervised by him will be able to know this and link it to specimens collected.

What if I don't want to be included or change my mind?

You are entirely free to decide not to participate or drop out at any time and this will not affect your care in any way. If a specimen has been collected and you change your mind it will be destroyed at your request.

10.3 Appendix 3: Parent information sheet

Parent/Carer's Invitation to participate in a research study: The Genetic Influence In the development of Crohn's disease and colitis in children

We invite your child to participate in research which we think may be important. The information which follows tells you about it. It is important that you understand what is in this leaflet. It says what will happen if you take part and what the risks might be. Try to make sure you know what will happen to you if you decide to take part. Whether or not you do take part is entirely your choice. Please ask any questions you want to about the research and we will try our best to answer them.

The Department of Paediatric Gastroenterology at Royal Hospital for Sick Children together with the Department of Gastroenterology at the Western General Hospital and other paediatric gastroenterology units in Scotland have a major role in leading research into diseases of the bowel in children. Diseases which we see often in the clinics and have on-going research interest include inflammatory bowel disease namely Crohn's disease and ulcerative colitis.

Your child is due to undergo tests involving collection of a blood sample as part of their on going care.

What we are asking

We would like your permission to collect extra blood (up to 20 ml = 4 teaspoons) which we will use for ongoing projects within our laboratories. Some of these specimens may be stored in a freezer for future projects. If specimens are stored they will only be used for studies relating to genetics and Crohn's disease and ulcerative colitis. We will also collect and store important and relevant clinical details (e.g. age, sex, medication, disease, findings from the test) for analysis with the specimens.

How will this affect my child?

Priority will always be given to collecting specimens needed for our planned tests. If there are any difficulties during the collection of these specimens e.g. in collecting blood then the doctors will not collect specimens for research. Taking extra blood will not mean any significant extra risk or harm to your child.

What research is being done?

Broadly speaking we are planning to study how the genetic make-up of children with inflammatory bowel disease (Crohn's disease or Colitis) may lead to them getting the disease.

The research included within this application will not have any impact on the care of your child (including any direct benefit or harm). All our research is aimed to increase understanding of how and why diseases occur, how treatments work and how we may be able to improve treatments in the future. If any studies may affect the care of individual children then a separate consent form will be produced.

Can I help too?

Yes you can by giving blood too. We would like wherever possible to collect blood from both parents as well as your child to help get maximum information from your child's blood sample. Your blood sample will expand the potential information we will gain from taking your children's blood. We appreciate your help with the study and so will arrange to take blood samples at a time and place that is convenient to you.

Feedback As the information found is not of any clinical relevance to your child we will not routinely be feeding back any information to you directly. We will however feed the collective results back at local support group meetings (CICRA).

Is this anonymous?

While collection of these specimens will not be anonymous only doctors or nurses involved in his/her clinical care will have access to his/her name. This allows us to go back to the hospital notes if we need to identify important and relevant information later on. Within the laboratory anonymity will be maintained by allocating a number (rather than his/her name) to the specimens. Only Dr Russell or other staff from the wards/clinics under his supervision will be able to link his/her name to the specimen.

What happens if I or my child are not keen or change our minds?

You and your child are entirely at liberty to decide not to participate or drop out at any time and this will not affect his/her care in any way. If after specimens have been collected you change your minds contact us and the specimens will be discarded if they have not already been used.

Is this ethical?

We have gained approval from the Lothian Regional Ethics Committee for the collection and storage of these specimens and data and each individual projects.

Any other Questions?

Any other questions can be addressed by writing to Dr Russell at the Dept of Paediatric Gastroenterology, Sciennes Road , Edinburgh or by contacting his secretary (0131 536 0615).

We also have an independent adviser Dr Steve Cunningham, a consultant paediatrician who can answer any concerns you have. Dr Cunningham is based at the Royal Hospital for Sick Children in Edinburgh but is not a member of the research team.

10.4 Appendix 4: Patient consent form

CHILD'S INVITATION TO PARTICIPATE IN A RESEARCH STUDY: CONSENT

THE GENETIC INFLUENCE IN THE DEVELOPMENT OF CROHN'S DISEASE AND COLITIS IN CHILDREN

Name of Patient/Volunteer:

Address:

- ***I HAVE READ AND KEPT THE INFORMATION SHEET…***
- ***SPOKEN TO THE DOCTOR AND MY FAMILY AND ASKED ANY QUESTIONS I WOULD LIKE…***
- ***I REALISE I DO NOT HAVE TO DO THIS …***
- ***I REALISE ANY BLOOD SAMPLE STORED WILL ONLY BE USED TO STUDY THE EFFECTS OF GENES ON CROHNS DISEASE AND COLITIS …***
- ***AND CAN STOP OR WITHDRAW AT ANY STAGE WITHOUT AFFECTING MY TREATMENT….***
- ***I AM HAPPY TO GO AHEAD.***

Signed ……

Name ……

Witnessed ……

Signature ……

Date ……

10.5 Appendix 5: example of parent consent form

PARENT/ GUARDIAN WRITTEN CONSENT FORM:

Title of research proposal: THE GENETIC INFLUENCE IN THE DEVELOPMENT OF CROHN'S DISEASE AND COLITIS IN CHILDREN

GREC Number: GREC 03/0273

Name of Patient/Volunteer:

Address:

• The study organisers have invited my child to take part in this research. **Yes/No**

• I understand what is in the leaflet about the research. I have a copy of the leaflet to keep. **Yes/No**

• I have had the chance to talk and ask questions about the study. **Yes/No**

• I know what my child's part will be in the study and I know how long it will take. **Yes/No**

• I know my child will gain no direct benefit from taking part in the study. **Yes/No**

• I know how the study may affect my child. I have been told if there are possible risks. **Yes/No**

• I understand that my child should not actively take part in more than 1 research study at a time. **Yes/No**

• I know that the local Grampian Research Ethics Committee has seen and agreed to this study. **Yes/No**

• I understand that personal information is strictly confidential: I know the only people who may see information about my part in the study are the research team or an official representative of the organisation which funded the research. **Yes/No**

• I freely consent for my child to be a subject in the study. No-one has put pressure on me. **Yes/No**

• I know that my child can stop taking part in the study at any time. **Yes/No**

• I know if my child does not take part he/she will still be able to have their normal treatment. **Yes/No**

• I understand that the sample may be stored and held for future research aimed at understanding the genetic influence on inflammatory bowel disease (Crohn's disease and inflammatory bowel disease) only, and not for other studies without my permission. **Yes/No**

• I know that if there are any problems, I can contact: **Yes/No**

Dr Russell Tel. No. 0131 536 0615

Parent/Carer's Signature:.....

Date:.....

The following should be signed by the Clinician/Investigator responsible for obtaining consent

As the Clinician/Investigator responsible for this research or a designated deputy, I confirm that I have explained to the patient/volunteer named above the nature and purpose of the research to be undertaken.

Clinician's Name: .....

Clinician's Signature:.....

Date: .....

10.6 Appendix 6: Patient/parent questionnaire

Genetics of Inflammatory Bowel Disease

Confidential

QUESTIONNAIRE FOR PATIENTS WITH IBD

Patient Identification Number **Sex: Male / Female**

Date of Birth: Hospital ID:

Ethnic origin

White European	□	Hispanic	□
Jewish	□	Afro-Caribbean	□
Japanese	□	Asian	□
Other/Unknown	□		

Postcode:

Name of present hospital consultant: Hospital:

Diagnosis: Ulcerative Colitis/ Crohn's disease / indeterminate Colitis/OFG

Date at symptom onset (month/year)

Date of diagnosis: (month/year)

Age at diagnosis:

Year of diagnosis

Years of disease at assessment:

School year:

Other medical problems:

SMOKING

1. Patient

Does patient smoke cigarettes? YES NO

If Yes, How many cigarettes a day?

What age did you start smoking?

If No, Have you ever smoked? YES NO

How many cigarettes a day?

What age did you start smoking? (Year) What age did you stop? (Year)

2. Mother:

Did you smoke during pregnancy? YES NO

If Yes, How many cigarettes a day?

Where you smoking when patient born? YES NO

If Yes, How many cigarettes a day?

Do you currently smoke cigarettes? YES NO

If Yes, How many cigarettes a day?

What age did you start smoking?

If No, Have you ever smoked? YES NO

How many cigarettes a day?

3. Father:

Did you smoke during pregnancy? YES NO

If Yes, How many cigarettes a day?

Where you smoking when patient born? YES NO

If Yes, How many cigarettes a day?

Do you currently smoke cigarettes? YES NO

If Yes, How many cigarettes a day?

What age did you start smoking?

If No, Have you ever smoked? YES NO

How many cigarettes a day?

Pregnancy and Breast-feeding

Birth weight kg

Complications in pregnancy

Was patient breast-fed at all? YES NO

If so for how long? (Months)

Vaccinations

Have you given your child the following vaccinations? (Tick all that apply)

DPT HIB MMR

Hep B Men c

Family History

Do you have?

Coeliac disease? YES NO

Colon cancer? YES NO

Autistic spectrum disorder? YES NO

Asthma/eczema/hay fever/food allergy? YES NO

Does your mother or father have? (List affected family members/age of onset)

Crohn's disease? YES NO

Ulcerative colitis? YES NO

Coeliac disease? YES NO

Colon cancer? YES NO

Autistic spectrum disorder? YES NO

Brothers and sisters

1. In total, how many brothers and sisters do you have?

Sibling rank

Have you ever shared a bedroom ? YES NO

Do your brother or sisters have?(List affected family members/age of onset)

Crohn's disease? YES NO

Ulcerative colitis? YES NO

Coeliac disease? YES NO

Colon cancer? YES NO

Autistic spectrum disorder? YES NO

Do your grandparents, aunts, uncles, or cousins have? (List affected family members and age of onset)

Crohn's disease? YES NO

Ulcerative colitis? YES NO

Coeliac disease? YES NO

Colon cancer? YES NO

Autistic spectrum disorder? YES NO

Swimming pools

Do you use a swimming pool? Frequently/rarely/never

Family estimate of number of disease relapses:

Surgery

Has patient had any operations for IBD? YES NO

If yes, please list below, with year if possible.

Date of surgery Type of operation

Have you had your tonsils removed?	YES	NO
If yes, when? Age/year		
Have you had your appendix removed?	YES	NO
If yes, when? Age/year		
Any other surgery?	YES	NO
Have you had any broken bones?	YES	NO
If yes, how many?		

Medications

Have you been treated with the following?

Steroids	YES / NO / don't know
Nutrition	YES / NO / don't know
Azathioprine	YES / NO / don't know
6MP	YES / NO / don't know
Infliximab?	YES / NO / don't know
Methotrexate	YES / NO / don't know

Number of steroid courses

Family estimate

Charted

10.7 Appendix 7: Crohn's disease phenotypic data collection form

Study Number ☐

Crohn's disease site **Presentation / Follow-up** **Year**

Site	**Behaviour**
Oral	Inflammatory B1
Oesophagus	Stricturing B2
Gastric body	Penetrating B3
Gastric antrum	**Give details below**
Duodenal	Strictures
Jejunal	Abscesses
Ileal TI	Fistulae
Caecal	Inflammatory mass
Ascending	Perianal ulcer
Transverse	
Descending	
Sigmoid	
Rectal	
Perianal	

Investigations

DEXA	yes / no	
EIM's	yes / no	give details

10.8 Appendix 8: UC/IC phenotypic data collection form

Study Number ☐

Presentation / Follow-up Date of Last Follow-up Clinic:

Date of last investigation:

Extent of UC

Examination satisfactory / unsatisfactory

Examination type colonoscopy / sigmoidoscopy

Examination stopped at

Ulceration yes / no

Crypt abscesses yes / no

Cryptitis yes / no

Goblet cell depletion yes / no

Disease years ☐

Extent changed yes / no

Colectomy yes / no

Comments

DEXA yes / no

EIM's yes / no give details

10.9 Appendix 9 : Laboratory data collection form

Study Number | |

Presentation / Follow-up Year

ESR	
Hb	
PLATELETS	
CRP	
ALBUMIN	
CALPROTECTIN	
VITAMIN B12	
FOLATE	

For CRP please choose from - normal or raised.

Clinic Date: Age at Diagnosis:

Height : Ht centile/Z-score

Weight: Wt centile/Z-score

BMI: BMI centile/Z-score

Puberty progressing normally / delayed / not assessed / not appropriate

10.10 Appendix 10: The salting out method

All steps were performed having taken universal precautions before handling blood and all stages were then performed in an appropriate "hood" within the laboratory.

1. 10 mls of whole blood collected previously into an EDTA tube were added into a 50ml conical tube.
2. 40 mls of red cell lysis buffer (RCLB) were then added to the conical tube.
3. The resulting mixture was then mixed and re-suspended for 5 minutes.
4. The solution was centrifuged at 3000 rpm for 10 minutes.
5. The supernatant was then removed leaving a pellet in the "cone" of the tube.
6. The pellet was re-suspended with a further 40 mls of RCLB and then re-centrifuged at the same speed for a further 5 minutes.
7. The resulting "white blood cell" pellet was then further re-suspended in 3 mls of nuclear lysis buffer and sodium dodecyl sulphate.
8. To this solution 1ml of 6M sodium chloride and 3mls of chloroform were then added.
9. The resulting mixture was then aggressively agitated to allow mixing.
10. The solution was then centrifuged at 3000 rpm for 20 minutes.
11. The solution then separated into 3 layers. The middle layer was then carefully pipetted out to avoid mixing with the other 2 layers.
12. The layer was added to 20 mls of 100% ethanol resulting in precipitation of DNA.
13. This DNA pellet was removed from the alcohol and dried in room air for 5 minutes.
14. The DNA was transferred into 0.5 mls of TE and stored at 4ºC until dissolves.

Bibliography

1. Lapidus A. The changing epidemiology of inflammatory bowel diseases. *Acta Gastroenterologica Belgica* 2001;**64**:155-9.

2. Barton JR, Gillon S, Ferguson A. Incidence of inflammatory bowel disease in Scottish children between 1968 and 1983; marginal fall in ulcerative colitis, three-fold rise in Crohn's disease. *Gut.* 1989;**30**:618-22.

3. Armitage E, Drummond HE, Wilson DC, Ghosh S. Increasing incidence of both juvenile-onset Crohn's disease and ulcerative colitis in Scotland. *European Journal of Gastroenterology & Hepatology.* 2001;**13**:1439-47.

4. Sawczenko A, Sandhu BK, Logan RF, Jenkins H, Taylor CJ, Mian S *et al*. Prospective survey of childhood inflammatory bowel disease in the British Isles. *Lancet.* 2001;**357**:1093-4.

5. Williams JG, Roberts SE, Ali MF, Cheung WY, Cohen DR, Demery G *et al*. Gastroenterology services in the UK. The burden of disease, and the organisation and delivery of services for gastrointestinal and liver disorders: a review of the evidence. *Gut* 2007;**56**:1-113.

6. Armitage E, Drummond H, Ghosh S, Ferguson A. Incidence of juvenile-onset Crohn's disease in Scotland. *Lancet.* 1999;**353**:1496-7.

7. Auvin S, Molinie F, Gower-Rousseau C, Brazier F, Merle V, Grandbastien B *et al*. Incidence, Clinical Presentation and Location at Diagnosis of Pediatric Inflammatory Bowel Disease: A Prospective Population-Based Study in Northern France (1988-1999). *Journal of Pediatric Gastroenterology & Nutrition* 2005;**41**:49-55.

8. Griffiths AM. Specificities of inflammatory bowel disease in childhood. *Best Practice & Research in Clinical Gastroenterology.* 2004;**18**:509-23.

9. Hildebrand H, Finkel Y, Grahnquist L, Lindholm J, Ekbom A, Askling J. Changing pattern of paediatric inflammatory bowel disease in northern Stockholm 1990-2001. *Gut.* 2003;**52**:1432-4.

10. Kildebo S, Nordgaard K, Aronsen O, Breckan R, Burhol PG, Jorde R. The incidence of ulcerative colitis in Northern Norway from 1983 to 1986. The

Northern Norwegian Gastroenterology Society. *Scandinavian Journal of Gastroenterology* 1990;**25**:890-6.

11. Loftus EV, Jr. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. *Gastroenterology* 2004;**126**:1504-17.

12. Armitage EL, Aldhous MC, Anderson N, Drummond HE, Riemersma RA, Ghosh S *et al*. Incidence of juvenile-onset Crohn's disease in Scotland: association with northern latitude and affluence. *Gastroenterology.* 2004;**127**:1051-7.

13. Kim SC,.Ferry GD. Inflammatory bowel diseases in pediatric and adolescent patients: clinical, therapeutic, and psychosocial considerations. *Gastroenterology.* 2004;**126**:1550-60.

14. Cosnes J, Cattan S, Blain A, Beaugerie L, Carbonnel F, Parc R *et al*. Long-term evolution of disease behavior of Crohn's disease. *Inflammatory Bowel Diseases.* 2002;**8**:244-50.

15. Bach JF. The effect of infections on susceptibility to autoimmune and allergic diseases. *New England Journal of Medicine* 2002;**347**:911-20.

16. Gent AE, Hellier MD, Grace RH, Swarbrick ET, Coggon D. Inflammatory bowel disease and domestic hygiene in infancy. *Lancet* 1994;**343**:766-7.

17. Hugot JP, Alberti C, Berrebi D, Bingen E, CA C. Crohn's disease: the cold chain hypothesis. *Lancet* 2003;**362**:2012-5.

18. Baron S, Turck D, Leplat C, Merle V, Gower-Rousseau C, Marti R *et al*. Environmental risk factors in paediatric inflammatory bowel diseases: a population based case control study. *Gut* 2005;**54**:357-63.

19. Braun J. In search of desire: mimicry, MAP, and Crohn's disease. *Gastroenterology* 2006;**131**:312-4.

20. Demicheli V, Jefferson T, Rivetti A, Price D, Demicheli V. Vaccines for measles, mumps and rubella in children. *Cochrane.Database.Syst.Rev.* 2005;CD004407.

21. Calkins BM. A meta-analysis of the role of smoking in inflammatory bowel disease. *Digestive Diseases & Sciences* 1989;**34**:1841-54.

22. Mahid SS, Minor KS, Soto RE, Hornung CA, Galandiuk S. Smoking and inflammatory bowel disease: a meta-analysis. *Mayo Clin.Proc.* 2006;**81**:1462-71.

23. Koutroubakis IE, Vlachonikolis IG, Kouroumalis EA. Role of appendicitis and appendectomy in the pathogenesis of ulcerative colitis: a critical review. *Inflammatory Bowel Diseases* 2002;**8**:277-86.

24. Rubin DT,.Hanauer SB. Smoking and inflammatory bowel disease. *European Journal of Gastroenterology & Hepatology* 2000;**12**:855-62.

25. Lashner BA, Shaheen NJ, Hanauer SB, Kirschner BS. Passive smoking is associated with an increased risk of developing inflammatory bowel disease in children. *American Journal of Gastroenterology* 1993;**88**:356-9.

26. Russell RK, Farhadi R, Wilson M, Drummond H, Satsangi J, Wilson DC. Perinatal passive smoke exposure may be more important than childhood exposure in the risk of developing childhood IBD. *Gut* 2005;**54**:1500-1.

27. Persson PG, Ahlbom A, Hellers G. Inflammatory bowel disease and tobacco smoke--a case-control study. *Gut* 1990;**31**:1377-81.

28. Sandler RS, Sandler DP, McDonnell CW, Wurzelmann JI. Childhood exposure to environmental tobacco smoke and the risk of ulcerative colitis. *American Journal of Epidemiology* 1992;**135**:603-8.

29. Andersson RE, Olaison G, Tysk C, Ekbom A. Appendectomy and protection against ulcerative colitis. *New England Journal of Medicine* 2001;**344**:808-14.

30. Feeney MA, Murphy F, Clegg AJ, Trebble TM, Sharer NM, Snook JA. A case-control study of childhood environmental risk factors for the development of inflammatory bowel disease. *Eur.J Gastroenterol.Hepatol.* 2002;**14**:529-34.

31. Sellon RK, Tonkonogy S, Schultz M, Dieleman LA, Grenther W, Balish E *et al*. Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-10-deficient mice. *Infection & Immunity* 1998;**66**:5224-31.

32. Winslet MC, Allan A, Poxon V, Youngs D, Keighley MR. Faecal diversion for Crohn's colitis: a model to study the role of the faecal stream in the inflammatory process. *Gut* 1994;**35**:236-42.

33. Guslandi M. Antibiotics for inflammatory bowel disease: do they work? *European Journal of Gastroenterology & Hepatology* 2005;**17**:145-7.

34. Fiocchi C. Probiotics in inflammatory bowel disease: yet another mechanism of action? *Gastroenterology* 2006;**131**:2009-12.

35. Silverberg MS, Satsangi J, Ahmad T, Arnott I.D.R., Bernstein CN, Brant SR *et al*. Towards an integrated clinical, molecular and serological classification of inflammatory bowel disease: Report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. *Canadian Journal of Gastroenterology* 2005;**19**:5-36.

36. Gasche C, Scholmerich J, Brynskov J, D'Haens G, Hanauer SB, Irvine EJ *et al*. A simple classification of Crohn's disease: report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998. *Inflammatory Bowel Diseases* 2000;**6**:8-15.

37. Murch SH, Baldassano R, Buller H, Chin S, Griffiths AM, Hildebrand H *et al*. Inflammatory bowel disease: Working Group report of the second World Congress of Pediatric Gastroenterology, Hepatology, and Nutrition. *J Pediatr.Gastroenterol.Nutr* 2004;**39 Suppl 2**:S647-S654.

38. Freeman HJ. Comparison of longstanding pediatric-onset and adult-onset Crohn's disease. *Journal of Pediatric Gastroenterology & Nutrition* 2004;**39**:183-6.

39. Sawczenko A,.Sandhu BK. Presenting features of inflammatory bowel disease in Great Britain and Ireland. *Archives of Disease in Childhood.* 2003;**88**:995-1000.

40. Kugathasan S, Judd RH, Hoffmann RG, Heikenen J, Telega G, Khan F *et al*. Epidemiologic and clinical characteristics of children with newly diagnosed inflammatory bowel disease in Wisconsin: a statewide population-based study. *Journal of Pediatrics.* 2003;**143**:525-31.

41. Heyman MB, Kirschner BS, Gold BD, Ferry G, Baldassano R, Cohen SA *et al*. Children with early-onset inflammatory bowel disease (IBD): Analysis of a pediatric IBD consortium registry. *The Journal of Pediatrics* 2005;**146**:35-40.

42. Witte J, Shivananda S, Lennard-Jones JE, Beltrami M, Politi P, Bonanomi A *et al*. Disease outcome in inflammatory bowel disease: mortality, morbidity and therapeutic management of a 796-person inception cohort in the European Collaborative Study on Inflammatory Bowel Disease (EC-IBD). *Scandinavian Journal of Gastroenterology* 2000;**35**:1272-7.

43. Henriksen M, Jahnsen J, Lygren I, Sauar J, Kjellevold O, Schulz T *et al*. Ulcerative colitis and clinical course: results of a 5-year population-based follow-up study (the IBSEN study). *Inflammatory Bowel Diseases* 2006;**12**:543-50.

44. Arnott ID,.Satsangi J. Crohn's disease or Crohn's diseases? *Gut* 2003;**52**:460-1.

45. Yang H, McElree C, Roth MP, Shanahan F, Targan SR, Rotter JI. Familial empirical risks for inflammatory bowel disease: differences between Jews and non-Jews. *Gut* 1993;**34**:517-24.

46. Probert CS, Jayanthi V, Pinder D, Wicks AC, Mayberry JF. Epidemiological study of ulcerative proctocolitis in Indian migrants and the indigenous population of Leicestershire. *Gut* 1992;**33**:687-93.

47. Orholm M, Munkholm P, Langholz E, Nielsen OH, Sorensen IA, Binder V. Familial occurrence of inflammatory bowel disease. *New England Journal of Medicine* 1991;**324**:84-8.

48. Russell RK,.Satsangi J. IBD: a family affair. *Best Practice & Research Clinical Gastroenterology* 2004;**18**:525-39.

49. Meucci G, Vecchi M, Torgano G, Arrigoni M, Prada A, Rocca F *et al*. Familial aggregation of inflammatory bowel disease in northern Italy: a multicenter study. *Gastroenterology* 1992;**103**:514-9.

50. Freeman HJ. Familial Crohn's disease in single or multiple first-degree relatives. *Journal of Clinical Gastroenterology* 2002;**35**:9-13.

51. Mayberry JF, Rhodes J, Newcombe RG. Familial prevalence of inflammatory bowel disease in relatives of patients with Crohn's disease. *British Medical Journal* 1980;**280**:84.

52. Monsen U, Bernell O, Johansson C, Hellers G. Prevalence of inflammatory bowel disease among relatives of patients with Crohn's disease. *Scandinavian Journal of Gastroenterology* 1991;**26**:302-6.

53. Farmer RG, Michener WM, Mortimer EA. Studies of family history among patients with inflammatory bowel disease. *Clinics in Gastroenterology* 1980;**9**:271-7.

54. Halme L, Turunen U, Helio T, Paavola P, Walle T, Miettinen A *et al*. Familial and sporadic inflammatory bowel disease: comparison of clinical

features and serological markers in a genetically homogeneous population. *Scandinavian Journal of Gastroenterology* 2002;**37**:692-8.

55. Satsangi J, Rosenberg WMC, Jewell DP. The Prevalence of Inflammatory Bowel-Disease in Relatives of Patients with Crohns-Disease. *European Journal of Gastroenterology & Hepatology* 1994;**6**:413-6.

56. Probert CS, Jayanthi V, Hughes AO, Thompson JR, Wicks AC, Mayberry JF. Prevalence and family risk of ulcerative colitis and Crohn's disease: an epidemiological study among Europeans and south Asians in Leicestershire. *Gut* 1993;**34**:1547-51.

57. Bayless TM, Tokayer AZ, Polito JM, Quaskey SA, Mellits ED, Harris ML. Crohn's disease: concordance for site and clinical type in affected family members--potential hereditary influences. *Gastroenterology* 1996;**111**:573-9.

58. Peeters M, Nevens H, Baert F, Hiele M, de Meyer AM, Vlietinck R *et al*. Familial aggregation in Crohn's disease: increased age-adjusted risk and concordance in clinical characteristics. *Gastroenterology* 1996;**111**:597-603.

59. Weterman IT,.Pena AS. Familial incidence of Crohn's disease in The Netherlands and a review of the literature. *Gastroenterology* 1984;**86**:449-52.

60. Roth MP, Petersen GM, McElree C, Vadheim CM, Panish JF, Rotter JI. Familial empiric risk estimates of inflammatory bowel disease in Ashkenazi Jews. *Gastroenterology* 1989;**96**:1016-20.

61. Carbonnel F, Macaigne G, Beaugerie L, Gendre JP, Cosnes J. Crohn's disease severity in familial and sporadic cases. *Gut* 1999;**44**:91-5.

62. Monsen U, Brostrom O, Nordenvall B, Sorstad J, Hellers G. Prevalence of inflammatory bowel disease among relatives of patients with ulcerative colitis. *Scandinavian Journal of Gastroenterology* 1987;**22**:214-8.

63. Tillil H,.Kobberling J. Age-corrected empirical genetic risk estimates for first-degree relatives of IDDM patients. *Diabetes* 1987;**36**:93-9.

64. Satsangi J, Grootscholten C, Holt H, Jewell DP. Clinical patterns of familial inflammatory bowel disease. *Gut* 1996;**38**:738-41.

65. Colombel JF, Grandbastien B, Gower-Rousseau C, Plegat S, Evrard JP, Dupas JL *et al*. Clinical characteristics of Crohn's disease in 72 families. *Gastroenterology* 1996;**111**:604-7.

66. Orholm M, Fonager K, Sorensen HT. Risk of ulcerative colitis and Crohn's disease among offspring of patients with chronic inflammatory bowel disease. *American Journal of Gastroenterology* 1999;**94**:3236-8.

67. Laharie D, Debeugny S, Peeters M, Van Gossum A, Gower-Rousseau C, Belaiche J *et al*. Inflammatory bowel disease in spouses and their offspring. *Gastroenterology* 2001;**120**:816-9.

68. Bennett RA, Rubin PH, Present DH. Frequency of inflammatory bowel disease in offspring of couples both presenting with inflammatory bowel disease. *Gastroenterology* 1991;**100**:1638-43.

69. Reynisdottir I, Gudbjartsson DF, Johannsson JH, Manolescu I, Kristjansson K, Stefansson K *et al*. A genetic contribution to inflammatory bowel disease in Iceland: a genealogic approach. *Clinical Gastroenterology and Hepatology* 2004;**2**:806-12.

70. Purrmann J, Bertrams J, Borchard F, Miller B, Cleveland S, Stolze T *et al*. Monozygotic triplets with Crohn's disease of the colon. *Gastroenterology* 1986;**91**:1553-9.

71. Tysk C, Lindberg E, Jarnerot G, Floderus-Myrhed B. Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut* 1988;**29**:990-6.

72. Halfvarson J, Bodin L, Tysk C, Lindberg E, Jarnerot G. Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. *Gastroenterology* 2003;**124**:1767-73.

73. Orholm M, Binder V, Sorensen TI, Rasmussen LP, Kyvik KO. Concordance of inflammatory bowel disease among Danish twins. Results of a nationwide study. *Scandinavian Journal of Gastroenterology* 2000;**35**:1075-81.

74. Thompson NP, Driscoll R, Pounder RE, Wakefield AJ. Genetics versus environment in inflammatory bowel disease: results of a British twin study. *British Medical Journal* 1996;**312**:95-6.

75. Subhani J, Montgomery SM, Pounder RE, Wakefield AJ. Concordance rates of twins and siblings in inflammatory bowel disease (IBD). *Gut* 1998;**42**:TW159.

76. Ahmad T, Satsangi J, McGovern D, Bunce M, Jewell DP. Review article: the genetics of inflammatory bowel disease. *Alimentary Pharmacology & Therapeutics* 2001;**15**:731-48.

77. Lashner BA, Evans AA, Kirsner JB, Hanauer SB. Prevalence and incidence of inflammatory bowel disease in family members. *Gastroenterology* 1986;**91**:1396-400.

78. Annese V, Andreoli A, Astegiano M, Campieri M, Caprilli R, Cucchiara S *et al*. Clinical features in familial cases of Crohn's disease and ulcerative colitis in Italy: a GISC study. Italian Study Group for the Disease of Colon and Rectum. *American Journal of Gastroenterology* 2001;**96**:2939-45.

79. Polito JM, Childs B, Mellits ED, Tokayer AZ, Harris ML, Bayless TM. Crohn's disease: influence of age at diagnosis on site and clinical type of disease. *Gastroenterology* 1996;**111**:580-6.

80. Tsuji S. Molecular genetics of triplet repeats: unstable expansion of triplet repeats as a new mechanism for neurodegenerative diseases. *Internal Medicine* 1997;**36**:3-8.

81. Polito JMI. Preliminary evidence for genetic anticipation in Crohn's disease. *Lancet* 1996;**347**:798-800.

82. Grandbastien B, Peeters M, Franchimont D, Gower-Rousseau C, Speckel D, Rutgeerts P *et al*. Anticipation in familial Crohn's disease. *Gut* 1998;**42**:170-4.

83. Heresbach D, Gulwani-Akolkar B, Lesser M, Akolkar PN, Lin XY, Heresbach-Le Berre N *et al*. Anticipation in Crohn's disease may be influenced by gender and ethnicity of the transmitting parent. *American Journal of Gastroenterology* 1998;**93**:2368-72.

84. Lee JC,.Lennard-Jones JE. Inflammatory bowel disease in 67 families each with three or more affected first-degree relatives. *Gastroenterology* 1996;**111**:587-96.

85. Hampe J, Heymann K, Kruis W, Raedler A, Folsch UR, Schreiber S. Anticipation in inflammatory bowel disease: a phenomenon caused by an accumulation of confounders. *American Journal of Medical Genetics* 2000;**92**:178-83.

86. Faybush EM, Blanchard JF, Rawsthorne P, Bernstein CN. Generational differences in the age at diagnosis with IBD: genetic anticipation, bias, or temporal effects. *American Journal of Gastroenterology* 2002;**97**:636-40.

87. Inskip H, Coggon D, Osmond C. Follow-up time bias and Crohn's disease. *Lancet* 1996;**347**:1551.

88. Frisch M, Olsen J, Andersen PK. Follow-up time bias and Crohn's disease. *Lancet* 1996;**347**:1551.

89. Picco MFM. Methodologic Pitfalls in the Determination of Genetic Anticipation: The Case of Crohn Disease. *Annals of Internal Medicine* 2001;**134**:1124-9.

90. Lee JC, Bridger S, McGregor C, Macpherson AJ, Jones JE. Why children with inflammatory bowel disease are diagnosed at a younger age than their affected parent. *Gut* 1999;**44**:808-11.

91. Hattersley AT,.McCarthy MI. What makes a good genetic association study? *Lancet* 2005;**366**:1315-23.

92. Frayling TM, Wiltshire S, Hitman GA, Walker M, Levy JC, Sampson M *et al*. Young-Onset Type 2 Diabetes Families Are the Major Contributors to Genetic Loci in the Diabetes UK Warren 2 Genome Scan and Identify Putative Novel Loci on Chromosomes 8q21, 21q22, and 22q11. *Diabetes* 2003;**52**:1857-63.

93. Akolkar PN, Gulwani-Akolkar B, Lin XY, Zhou Z, Daly M, Katz S *et al*. The IBD1 locus for susceptibility to Crohn's disease has a greater impact in Ashkenazi Jews with early onset disease. *American Journal of Gastroenterology.* 2001;**96**:1127-32.

94. Brant SR, Panhuysen CI, Bailey-Wilson JE, Rohal PM, Lee S, Mann J *et al*. Linkage heterogeneity for the IBD1 locus in Crohn's disease pedigrees by disease onset and severity. *Gastroenterology* 2000;**119**:1483-90.

95. Cordell HJ,.Clayton DG. Genetic association studies. *Lancet* 2005;**366**:1121-31.

96. Orholm M, Iselius L, Sorensen TI, Munkholm P, Langholz E, Binder V. Investigation of inheritance of chronic inflammatory bowel diseases by complex segregation analysis. *British Medical Journal* 1993;**306**:20-4.

97. Farrall M. Inheritance of chronic inflammatory bowel disease. *British Medical Journal* 1993;**306**:651.

98. Hopper JL, Bishop DT, Easton DF. Population-based family studies in genetic epidemiology. *Lancet.* 2005;**366**:1397-406.

99. Russell RK, Wilson DC, Satsangi J. Unravelling the complex genetics of inflammatory bowel disease. *Arch Dis Child* 2004;**89**:598-603.

100. Noble C, Nimmo E, Gaya D, Russell RK, Satsangi J. Novel susceptibility genes in inflammatory bowel disease. *World Journal of Gastroenterology* 2006;**12**:1991-9.

101. Gaya DR, Russell RK, Nimmo ER, Satsangi J. New genes in inflammatory bowel disease: lessons for complex diseases? *The Lancet* 2006;**367**:1271-84.

102. Van Limbergen J, Russell RK, Nimmo ER, Ho GT, Arnott IDR, Wilson DC *et al*. Genetics of the innate immune response in inflammatory bowel disease. *Inflammatory Bowel Diseases* 2007;**13**:338-55.

103. Russell RK, Nimmo ER, Satsangi J. Molecular genetics of Crohn's disease. *Current Opinion in Genetics & Development* 2004;**14**:264-70.

104. Dawn TM,.Barrett JH. Genetic linkage studies. *Lancet.* 2005;**366**:1036-44.

105. Weissenbach J. The Human Genome Project: from mapping to sequencing. *Clinical Chemistry & Laboratory Medicine* 1998;**36**:511-4.

106. Hugot JP, Laurent-Puig P, Gower-Rousseau C, Olson JM, Lee JC, Beaugerie L *et al*. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* 1996;**379**:821-3.

107. Satsangi J, Parkes M, Louis E, Hashimoto L, Kato N, Welsh K *et al*. Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nature Genetics* 1996;**14**:199-202.

108. Hampe J, Schreiber S, Shaw SH, Lau KF, Bridger S, Macpherson AJ *et al*. A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. *American Journal of Human Genetics* 1999;**64**:808-16.

109. Ma Y, Ohmen JD, Li Z, Bentley LG, McElree C, Pressman S *et al*. A genome-wide search identifies potential new susceptibility loci for Crohn's disease. *Inflammatory Bowel Diseases.* 1999;**5**:271-8.

110. Cho JH, Nicolae DL, Gold LH, Fields CT, LaBuda MC, Rohal PM *et al*. Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q, and 4q: evidence for epistasis between 1p and IBD1. *Proceedings of the National Academy of Sciences of the United States of America* 1998;**95**:7502-7.

111. Duerr RH, Barmada MM, Zhang L, Pfutzer R, Weeks DE. High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11-12. *American Journal of Human Genetics.* 2000;**66**:1857-62.

112. Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, McLeod RS, Griffiths AM *et al*. Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *American Journal of Human Genetics.* 2000;**66**:1863-70.

113. Williams CN, Kocher K, Lander ES, Daly MJ, Rioux JD. Using a genome-wide scan and meta-analysis to identify a novel IBD locus and confirm previously identified IBD loci. *Inflammatory Bowel Diseases.* 2002;**8**:375-81.

114. Barmada MM, Brant SR, Nicolae DL, Achkar JP, Panhuysen CI, Bayless TM *et al*. A genome scan in 260 inflammatory bowel disease-affected relative pairs. *Inflammatory Bowel Diseases* 2004;**10**:15-22.

115. Vermeire S, Rutgeerts P, Van Steen K, Joossens S, Claessens G, Pierik M *et al*. Genome wide scan in a Flemish inflammatory bowel disease population: support for the IBD4 locus, population heterogeneity, and epistasis. *Gut* 2004;**53**:980-6.

116. Van Heel DA, Dechairo BM, Dawson G, McGovern DP, Negoro K, Carey AH *et al*. The IBD6 Crohn's disease locus demonstrates complex interactions with CARD15 and IBD5 disease-associated variants. *Human Molecular Genetics.* 2003;**12**:2569-75.

117. McKusick-Nathans Institute for Genetic Medicine, Johns Hopkins University Baltimore MD and National Center for Biotechnology Information National Library of Medicine Bethesda MD. Online Mendelian Inheritance in Man, OMIM (TM). World Wide Web URL: http://www.ncbi.nlm.nih.gov/omim/ . 2000.

118. Cavanaugh J. International collaboration provides convincing linkage replication in complex disease through analysis of a large pooled data set: Crohn disease and chromosome 16. *American Journal of Human Genetics* 2001;**68**:1165-71.

119. Parkes M, Barmada MM, Satsangi J, Weeks DE, Jewell DP, Duerr RH. The IBD2 locus shows linkage heterogeneity between ulcerative colitis and Crohn disease. *American Journal of Human Genetics.* 2000;**67**:1605-10.

120. Van Heel DA, Fisher SA, Kirby A, Daly MJ, Rioux JD, Lewis CM. Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs. *Hum.Mol.Genet.* 2004;**13**:763-70.

121. Ohmen JD, Yang HY, Yamamoto KK, Zhao HY, Ma Y, Bentley LG *et al*. Susceptibility locus for inflammatory bowel disease on chromosome 16 has a role in Crohn's disease, but not in ulcerative colitis. *Hum.Mol.Genet.* 1996;**5**:1679-83.

122. Brant SR, Fu Y, Fields CT, Baltazar R, Ravenhill G, Pickles MR *et al*. American families with Crohn's disease have strong evidence for linkage to chromosome 16 but not chromosome 12. *Gastroenterology* 1998;**115**:1056-61.

123. Annese V, Latiano A, Bovio P, Forabosco P, Piepoli A, Lombardi G *et al*. Genetic analysis in Italian families with inflammatory bowel disease supports linkage to the IBD1 locus--a GISC study. *European Journal of Human Genetics.* 1999;**7**:567-73.

124. Achkar JP, Dassopoulos T, Silverberg MS, Tuvlin JA, Duerr RH, Brant SR *et al*. Phenotype-stratified genetic linkage study demonstrates that IBD2 is an extensive ulcerative colitis locus. *American Journal of Gastroenterology* 2006;**101**:572-80.

125. Stokkers PC, Reitsma PH, Tytgat GN, van Deventer SJ. HLA-DR and -DQ phenotypes in inflammatory bowel disease: a meta-analysis. *Gut* 1999;**45**:395-401.

126. Dechairo B, Dimon C, van Heel D, Mackay I, Edwards M, Scambler P *et al*. Replication and extension studies of inflammatory bowel disease susceptibility regions confirm linkage to chromosome 6p (IBD3). *European Journal of Human Genetics.* 2001;**9**:627-33.

127. Yang H, Plevy SE, Taylor K, Tyan D, Fischel-Ghodsian N, McElree C *et al*. Linkage of Crohn's disease to the major histocompatibility complex region is detected by multiple non-parametric analyses. *Gut* 1999;**44**:519-26.

128. Fisher SA, Hampe J, Macpherson AJ, Forbes A, Lennard-Jones JE, Schreiber S *et al*. Sex stratification of an inflammatory bowel disease genome search shows male-specific linkage to the HLA region of chromosome 6. *European Journal of Human Genetics.* 2002;**10**:259-65.

129. Roussomoustakaki M, Satsangi J, Welsh K, Louis E, Fanning G, Targan S *et al*. Genetic markers may predict disease behavior in patients with ulcerative colitis. *Gastroenterology* 1997;**112**:1845-53.

130. Yamamoto-Furusho JK, Uscanga LF, Vargas-Alarcon G, Ruiz-Morales JA, Higuera L, Cutino T *et al*. Clinical and genetic heterogeneity in Mexican patients with ulcerative colitis. *Human Immunology* 2003;**64**:119-23.

131. Trachtenberg EA, Yang H, Hayes E, Vinson M, Lin C, Targan SR *et al*. HLA class II haplotype associations with inflammatory bowel disease in Jewish (Ashkenazi) and non-Jewish caucasian populations. *Human Immunology* 2000;**61**:326-33.

132. Ahmad T, Armuzzi A, Bunce M, Mulcahy-Hawes K, Marshall SE, Orchard TR *et al*. The molecular classification of the clinical manifestations of Crohn's disease. *Gastroenterology* 2002;**122**:854-66.

133. Silverberg MS, Mirea L, Bull SB, Murphy JE, Steinhart AH, Greenberg GR *et al*. A population- and family-based study of Canadian families reveals association of HLA DRB1*0103 with colonic involvement in inflammatory bowel disease. *Inflammatory Bowel Diseases* 2003;**9**:1-9.

134. Pierik M, Yang H, Barmada MM, Cavanaugh JA, Annese V, Brant SR *et al*. The IBD international genetics consortium provides further evidence for linkage to IBD4 and shows gene-environment interaction. *Inflammatory Bowel Diseases* 2005;**11**:1-7.

135. Rioux JD, Daly MJ, Silverberg MS, Lindblad K, Steinhart H, Cohen Z *et al*. Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nature Genetics* 2001;**29**:223-8.

136. Low JH, Williams FA, Yang X, Cullen S, Colley J, Ling KL *et al*. Inflammatory bowel disease is linked to 19p13 and associated with ICAM-1. *Inflammatory Bowel Diseases* 2004;**10**:173-81.

137. Cho JH, Nicolae DL, Ramos R, Fields CT, Rabenau K, Corradino S *et al*. Linkage and linkage disequilibrium in chromosome band 1p36 in American Chaldeans with inflammatory bowel disease. *Human Molecular Genetics.* 2000;**9**:1425-32.

138. Hampe J, Frenzel H, Mirza MM, Croucher PJ, Cuthbert A, Mascheretti S *et al*. Evidence for a NOD2-independent susceptibility locus for inflammatory bowel disease on chromosome 16p. *Proceedings of the National Academy of Sciences of the United States of America.* 2002;**99**:321-6.

139. Hampe J, Lynch NJ, Daniels S, Bridger S, Macpherson AJ, Stokkers P *et al*. Fine mapping of the chromosome 3p susceptibility locus in inflammatory bowel disease. *Gut.* 2001;**48**:191-7.

140. Duerr RH, Barmada MM, Zhang L, Achkar JP, Cho JH, Hanauer SB *et al*. Evidence for an inflammatory bowel disease locus on chromosome 3p26: linkage, transmission/disequilibrium and partitioning of linkage. *Human Molecular Genetics.* 2002;**11**:2599-606.

141. Silverberg MS, Duerr RH, Brant SR, Bromfield G, Datta LW, Jani N *et al*. Refined genomic localization and ethnic differences observed for the IBD5 association with Crohn's disease. *European Journal of Human Genetics* 2007;**15**:328-35.

142. Singer HC, Anderson JG, Frischer H, Kirsner JB. Familial aspects of inflammatory bowel disease. *Gastroenterology* 1971;**61**:423-30.

143. Lennard-Jones JE. Classification of inflammatory bowel disease. *Scandinavian Journal of Gastroenterology* 1989;**170**:2-6.

144. O'Connell JR,.Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *American Journal of Human Genetics* 1998;**63**:259-66.

145. Denman DW,.Schlesselman JJ. Interval Estimation of the Attributable Risk for Multiple Exposure Levels in Case-Control Studies. *Biometrics* 1983;**39**:185-92.

146. Clayton D. A generalization of the transmission/disequillibrium test for uncertain- haplotype transmission. *American Journal of Human Genetics* 1999;**65**:1170-7.

147. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research* 1988;**16**:1215.

148. Freeman HJ. Application of the Vienna Classification for Crohn's disease to a single clinician database of 877 patients. *Canadian Journal of Gastroenterology*. 2001;**15**:89-93.

149. Sachar DB, Bodian CA, Goldstein ES, Present DH, Bayless TM, Picco M *et al*. Is perianal Crohn's disease associated with intestinal fistulization? *American Journal of Gastroenterology*. 2005;**100**:1547-9.

150. Louis E, Collard A, Oger AF, Degroote E, Aboul Nasr El Yafi FA, Belaiche J. Behaviour of Crohn's disease according to the Vienna classification: changing pattern over the course of the disease. *Gut* 2001;**49**:777-82.

151. Smith BR, Arnott ID, Drummond HE, Nimmo ER, Satsangi J. Disease location, anti-Saccharomyces cerevisiae antibody, and NOD2/CARD15 genotype influence the progression of disease behavior in Crohn's disease. *Inflammatory Bowel Diseases* 2004;**10**:521-8.

152. Bernstein CN, Blanchard JF, Rawsthorne P, Wajda A. Epidemiology of Crohn's disease and ulcerative colitis in a central Canadian province: a population-based study. *American Journal of Epidemiology.* 1999;**149**:916-24.

153. Lindberg E, Lindquist B, Holmquist L, Hildebrand H. Inflammatory bowel disease in children and adolescents in Sweden, 1984-1995. *Journal of Pediatric Gastroenterology & Nutrition.* 2000;**30**:259-64.

154. Godet PG, May GR, Sutherland LR. Meta-analysis of the role of oral contraceptive agents in inflammatory bowel disease. *Gut* 1995;**37**:668-73.

155. Biank V, Friedrichs F, Babusukumar U, Wang T, Stoll M, Broeckel U *et al*. DLG5 R30Q Variant Is a Female-Specific Protective Factor in Pediatric Onset Crohn's Disease. *The American Journal of Gastroenterology* 2007;**102**:391-8.

156. Friedrichs F, Brescianini S, Annese V, Latiano A, Berger K, Kugathasan S *et al*. Evidence of transmission ratio distortion of DLG5 R30Q variant in general and implication of an association with Crohn disease in men. *Human Genetics* 2006;**119**:305-11.

157. Hayward PA, Satsangi J, Jewell DP. Inflammatory bowel disease and the X chromosome. *Q J Med* 1996;**89**:713-8.

158. IBD Working Group of the European Society for Paediatric Gastroenterology HaN. Inflammatory Bowel Disease in Children and Adolescents: Recommendations for Diagnosis-The Porto Criteria. *Journal of Pediatric Gastroenterology & Nutrition* 2005;**41**:1-7.

159. Shaoul R, Karban A, Weiss B, Reif S, Wasserman D, Pacht A *et al*. NOD2/CARD15 mutations and presence of granulomas in pediatric and adult Crohn's disease. *Inflammatory Bowel Diseases* 2004;**10**:709-14.

160. Hommes DW,.van Deventer SJ. Endoscopy in inflammatory bowel diseases. *Gastroenterology.* 2004;**126**:1561-73.

161. Henriksen M, Jahnsen J, Lygren I, Aadland E, Schulz T, Vatn MH *et al*. Clinical course in Crohn's disease: Results of a five-year population-based follow-up study (the IBSEN study). *Scandinavian Journal of Gastroenterology* 2007;**42**:602-10.

162. Eaden JA, Abrams KR, Mayberry JF. The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut* 2001;**48**:526-35.

163. Mamula P, Telega GW, Markowitz JE, Brown KA, Russo PA, Piccoli DA *et al*. Inflammatory bowel disease in children 5 years of age and younger. *American Journal of Gastroenterology* 2002;**97**:2005-10.

164. Meinzer U, Idestrom M, Alberti C, Peuchmaur M, Belarbi N, Bellaiche M *et al*. Ileal involvement is age dependent in pediatric Crohn's disease. *Inflammatory Bowel Diseases.* 2005;**11**:639-44.

165. Ahmad T, Tamboli CP, Jewell D, Colombel JF. Clinical relevance of advances in genetics and pharmacogenetics of IBD. *Gastroenterology.* 2004;**126**:1533-49.

166. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J *et al*. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001;**411**:599-603.

167. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R *et al*. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001;**411**:603-6.

168. Inohara N, Ogura Y, Chen FF, Muto A, Nunez G. Human Nod1 confers responsiveness to bacterial lipopolysaccharides. *Journal of Biological Chemistry.* 2001;**276**:2551-4.

169. Hugot JPZ. Lessons to be learned from the NOD2 gene in Crohn's disease. *European Journal of Gastroenterology & Hepatology* 2003;**15**:593-7.

170. Ogura Y, Inohara N, Benito A, Chen FF, Yamaoka S, Nunez G. Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-kappaB. *Journal of Biological Chemistry.* 2001;**276**:4812-8.

171. Gutierrez O, Pipaon C, Inohara N, Fontalba A, Ogura Y, Prosper F *et al*. Induction of Nod2 in myelomonocytic and intestinal epithelial cells via nuclear factor-kappa B activation. *Journal of Biological Chemistry* 2002;**277**:41701-5.

172. Rosenstiel P, Fantini M, Brautigam K, Kuhbacher T, Waetzig GH, Seegert D *et al*. TNF-alpha and IFN-gamma regulate the expression of the NOD2 (CARD15) gene in human intestinal epithelial cells. *Gastroenterology.* 2003;**124**:1001-9.

173. Lala S, Ogura Y, Osborne C, Hor SY, Bromfield A, Davies S *et al*. Crohn's disease and the NOD2 gene: A role for Paneth cells. *Gastroenterology* 2003;**125**:47-57.

174. Ayabe T, Satchell DP, Wilson CL, Parks WC, Selsted ME, Ouellette AJ. Secretion of microbicidal alpha-defensins by intestinal Paneth cells in response to bacteria. *Nature Immunology* 2000;**1**:113-8.

175. Girardin SE, Boneca IG, Viala J, Chamaillard M, Labigne A, Thomas G *et al*. Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *Journal of Biological Chemistry* 2003;**278**:8869-72.

176. Inohara N, Ogura Y, Nunez G. Nods: a family of cytosolic proteins that regulate the host response to pathogens. *Current Opinion in Microbiology* 2002;**5**:76-80.

177. Zouali H, Lesage S, Merlin F, Cezard JP, Colombel JF, Belaiche J *et al*. CARD4/NOD1 is not involved in inflammatory bowel disease. *Gut* 2003;**52**:71-4.

178. Albrecht M, Domingues FS, Schreiber S, Lengauer T. Structural localization of disease-associated sequence variations in the NACHT and LRR domains of PYPAF1 and NOD2. *FEBS Letters.* 2003;**554**:520-8.

179. Inohara N, Ogura Y, Fontalba A, Gutierrez O, Pons F, Crespo J *et al*. Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. *Journal of Biological Chemistry* 2003;**278**:5509-12.

180. Kobayashi K, Inohara N, Hernandez LD, Galan JE, Nunez G, Janeway CA *et al*. RICK/Rip2/CARDIAK mediates signalling for receptors of the innate and adaptive immune systems. *Nature* 2002;**416**:194-9.

181. Bonen DK, Ogura Y, Nicolae DL, Inohara N, Saab L, Tanabe T *et al*. Crohn's disease-associated NOD2 variants share a signaling defect in response to lipopolysaccharide and peptidoglycan. *Gastroenterology.* 2003;**124**:140-6.

182. Netea MG, Kullberg BJ, de Jong DJ, Franke B, Sprong T, Naber TH *et al*. NOD2 mediates anti-inflammatory signals induced by TLR2 ligands: implications for Crohn's disease. *European Journal of Immunology.* 2004;**34**:2052-9.

183. Hisamatsu T, Suzuki M, Reinecker HC, Nadeau WJ, McCormick BA, Podolsky DK. CARD15/NOD2 functions as an antibacterial factor in human intestinal epithelial cells. *Gastroenterology.* 2003;**124**:993-1000.

184. Arnott ID, Nimmo ER, Drummond HE, Fennell J, Smith BR, MacKinlay E *et al*. NOD2/CARD15, TLR4 and CD14 mutations in Scottish and Irish Crohn's

disease patients: evidence for genetic heterogeneity within Europe? *Genes & Immunity* 2004;**5**:417-25.

185. Economou M, Trikalinos TA, Loizou KT, Tsianos EV, Ioannidis JPA. Differential Effects of NOD2 Variants on Crohn's Disease Risk and Phenotype in Diverse Populations: A Metaanalysis. *The American Journal of Gastroenterology* 2004;**99**:2393-404.

186. Ferraris A, Knafelz D, Torres B, Fortina P, Castro M, Dallapiccola B. Analysis of CARD15 gene variants in Italian pediatric patients with inflammatory bowel diseases. *Journal of Pediatrics* 2005;**147**:272-3.

187. Idestrom M, Rubio C, Granath F, Finkel Y, Hugot JP. CARD15 Mutations Are Rare in Swedish Pediatric Crohn Disease. *Journal of Pediatric Gastroenterology & Nutrition* 2005;**40**:456-60.

188. Weiss B, Shamir R, Bujanover Y, Waterman M, Hartman C, Fradkin A *et al*. NOD2/CARD15 mutation analysis and genotype-phenotype correlation in Jewish pediatric patients compared with adults with Crohn's disease. *Journal of Pediatrics.* 2004;**145**:208-12.

189. Wine E, Reif SS, Leshinsky-Silver E, Weiss B, Shaoul RR, Shamir R *et al*. Pediatric Crohn's disease and growth retardation: the role of genotype, phenotype, and disease severity. *Pediatrics* 2004;**114**:1281-6.

190. Tomer G, Ceballos C, Concepcion E, Benkov KJ. NOD2/CARD15 variants are associated with lower weight at diagnosis in children with Crohn's disease. *American Journal of Gastroenterology* 2003;**98**:2479-84.

191. Sun L, Roesler J., Rosen-Wolff A, Winkler U., Koch R., Thurigen A. *et al*. CARD15 Genotype and Phenotype Analysis in 55 Pediatric Patients With Crohn Disease From Saxony, Germany. *Journal of Pediatric Gastroenterology & Nutrition* 2003;**37**:492-7.

192. Griffiths AM, Sherman P, Thijs G, Ray P, Silverberg MS. Prevalence of NOD2/CARD15 Mutations in Pediatric IBD. *Gastroenterology* 2003;**124**:A-367.

193. Kugathasan S, Collins N, Maresso K, Hoffmann RG, Stephens M, Werlin SL *et al*. CARD15 gene mutations and risk for early surgery in pediatric-onset Crohn's disease. *Clinical Gastroenterology & Hepatology* 2004;**2**:1003-9.

194. Ferraris A, Torres B, Knafelz D, Barabino A, Lionetti P, de Angelis GL *et al*. Relationship between CARD15, SLC22A4/5, and DLG5 polymorphisms and

early-onset inflammatory bowel diseases: an Italian multicentric study. *Inflammatory Bowel Diseases* 2006;**12**:355-61.

195. Russell RK, Wilson DC, Satsangi J. CARD 15 / NOD 2 in Paediatric Onset Crohn's Disease. *Journal of Pediatric Gastroenterology & Nutrition* 2004;**39**:216.

196. Vind I, Vieira A, Hougs L, Tavares L, Riis L, Skytt AP *et al*. NOD2/CARD15 Gene Polymorphisms in Crohn's Disease: A Genotype-Phenotype Analysis in Danish and Portuguese Patients and Controls. *Digestion* 2005;**72**:156-63.

197. Bairead E, Harmon DL, Curtis AM, Kelly Y, O'Leary C, Gardner M *et al*. Association of NOD2 with Crohn's disease in a homogenous Irish population. *European Journal of Human Genetics* 2003;**11**:237-44.

198. Hampe J, Grebe J, Nikolaus S, Solberg C, Croucher PJ, Mascheretti S *et al*. Association of NOD2 (CARD 15) genotype with clinical course of Crohn's disease: a cohort study. *Lancet.* 2002;**359**:1661-5.

199. Thjodleifsson B, Sigthorsson G, Cariglia N, Reynisdottir I, Gudbjartsson DF, Kristjansson K *et al*. Subclinical intestinal inflammation: an inherited abnormality in Crohn's disease relatives? *Gastroenterology* 2003;**124**:1728-37.

200. Helio T, Halme L, Lappalainen M, Fodstad H, Paavola-Sakki P, Turunen U *et al*. CARD15/NOD2 gene variants are associated with familially occurring and complicated forms of Crohn's disease. *Gut* 2003;**52**:558-62.

201. Torkvist L, Noble CL, Lordal M, Sjoqvist U, Lindforss U, Nimmo ER *et al*. Contribution of CARD15 variants in determining susceptibility to Crohn's disease in Sweden. *Scandinavian Journal of Gastroenterology.* 2006;**41**:700-5.

202. Buning C, Molnar T, Nagy F, Lonovics J, Weltrich R, Bochow B *et al*. NOD2/CARD15 gene polymorphism in patients with inflammatory bowel disease: is Hungary different? *World Journal of Gastroenterology* 2005;**11**:407-11.

203. Mendoza JL, Murillo LS, Fernandez L, Pena AS, Lana R, Urcelay E *et al*. Prevalence of mutations of the NOD2/CARD15 gene and relation to phenotype in Spanish patients with Crohn disease. *Scandinavian Journal of Gastroenterology.* 2003;**38**:1235-40.

204. Linde K, Boor PP, Houwing-Duistermaat JJ, Kuipers EJ, Wilson JH, de Rooij FW. CARD15 and Crohn's disease: healthy homozygous carriers of the 3020insC frameshift mutation. *American Journal of Gastroenterology.* 2003;**98**:613-7.

205. Riis L, Vind I, Vermeire S, Wolters F, Katsanos K, Politi P *et al*. The prevalence of genetic and serologic markers in an unselected European population-based cohort of IBD patients. *Inflammatory Bowel Diseases* 2007;**13**:24-32.

206. Croucher PJ, Mascheretti S, Hampe J, Huse K, Frenzel H, Stoll M *et al*. Haplotype structure and association to Crohn's disease of CARD15 mutations in two ethnically divergent populations. *European Journal of Human Genetics* 2003;**11**:6-16.

207. Leong RWA. NOD2/CARD15 gene polymorphisms and Crohn's disease in the Chinese population. *Alimentary Pharmacology & Therapeutics* 2003;**17**:1465-70.

208. Inoue N, Tamura K, Kinouchi Y, Fukuda Y, Takahashi S, Ogura Y *et al*. Lack of common NOD2 variants in Japanese patients with Crohn's disease. *Gastroenterology.* 2002;**123**:86-91.

209. Yamazaki K, Takazoe M, Tanaka T, Kazumori T, Nakamura Y. Absence of mutation in the NOD2/CARD15 gene among 483 Japanese patients with Crohn's disease. *Journal of Human Genetics* 2002;**47**:469-72.

210. Kugathasan S, Loizides A, Babusukumar U, McGuire E, Wang T, Hooper P *et al*. Comparative phenotypic and CARD15 mutational analysis among African American, Hispanic, and White children with Crohn's disease. *Inflammatory Bowel Diseases.* 2005;**11**:631-8.

211. Hugot JP, Zaccaria I, Cavanaugh J, Yang H, Vermeire S, Lappalainen M *et al*. Prevalence of CARD15/NOD2 Mutations in Caucasian Healthy People. *American Journal of Gastroenterology* 2007;**102**:1259-67.

212. Leshinsky-Silver E, Karban A, Buzhakor E, Fridlander M, Yakir B, Eliakim R *et al*. Is age of onset of Crohn's disease governed by mutations in NOD2/caspase recruitment domains 15 and Toll-like receptor 4? Evaluation of a pediatric cohort. *Pediatr Res* 2005;**58**:499-504.

213. Hampe J, Cuthbert A, Croucher PJ, Mirza MM, Mascheretti S, Fisher S *et al*. Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet* 2001;**357**:1925-8.

214. Annese V, Lombardi G, Perri F, D'Inca R, Ardizzone S, Riegler G *et al*. Variants of CARD15 are associated with an aggressive clinical course of Crohn's disease--an IG-IBD study. *American Journal of Gastroenterology* 2005;**100**:84-92.

215. Buning C, Genschel J, Buhner S, Kruger S, Kling K, Dignass A *et al*. Mutations in the NOD2/CARD15 gene in Crohn's disease are associated with ileocaecal resection and are a risk factor for reoperation. *Alimentary Pharmacology & Therapeutics.* 2004;**19**:1073-8.

216. Abreu MT, Taylor KD, Lin YC, Hang T, Gaiennie J, Landers CJ *et al*. Mutations in NOD2 are associated with fibrostenosing disease in patients with Crohn's disease. *Gastroenterology* 2002;**123**:679-88.

217. Brant SR, Picco MF, Achkar JP, Bayless TM, Kane SV, Brzezinski A *et al*. Defining complex contributions of NOD2/CARD15 gene mutations, age at onset, and tobacco use on Crohn's disease phenotypes. *Inflammatory Bowel Diseases.* 2003;**9**:281-9.

218. Heresbach D, Gicquel-Douabin V, Birebent B, D'halluin PN, Heresbach-Le Berre N, Dreano S *et al*. NOD2/CARD15 gene polymorphisms in Crohn's disease: a genotype- phenotype analysis. *European Journal of Gastroenterology & Hepatology.* 2004;**16**:55-62.

219. Lesage S, Zouali H, Cezard JP, Colombel JF, Belaiche J, Almer S *et al*. CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *American Journal of Human Genetics* 2002;**70**:845-57.

220. Weinstein TA, Levine M, Pettei MJ, Gold DM, Kessler BH, Levine JJ. Age and family history at presentation of pediatric inflammatory bowel disease. *Journal of Pediatric Gastroenterology & Nutrition* 2003;**37**:609-13.

221. Cabrera-Abreu JC, Davies P, Matek Z, Murphy MS. Performance of blood tests in diagnosis of inflammatory bowel disease in a specialist clinic. *Archives of Disease in Childhood.* 2004;**89**:69-71.

222. Russell RK, Read JR, Rogers P, Gillett PM, Satsangi J, Wilson DC. Factors affecting the time to IBD diagnosis in children: Implications for practice. *Gut* 2005;**54**:359.

223. Beattie RM, Walker-Smith JA, Murch SH. Indications for investigation of chronic gastrointestinal symptoms. *Archives of Disease in Childhood* 1995;**73**:354-5.

224. Vermeire S, Van Assche G, Rutgeerts P. C-reactive protein as a marker for inflammatory bowel disease. *Inflammatory Bowel Diseases.* 2004;**10**:661-5.

225. Roesler J, Thurigen A, Sun L, Koch R, Winkler U, Laass MW *et al*. Influence of CARD15 Mutations on Disease Activity and Response to Therapy in 65 Pediatric Crohn Patients from Saxony, Germany. *Journal of Pediatric Gastroenterology & Nutrition* 2005;**41**:27-32.

226. Newman B, Silverberg MS, Gu X, Zhang Q, Lazaro A, Steinhart AH *et al*. CARD15 and HLA DRB1 alleles influence susceptibility and disease localization in Crohn's disease. *American Journal of Gastroenterology.* 2004;**99**:306-15.

227. Cuthbert AP, Fisher SA, Mirza MM, King K, Hampe J, Croucher PJ *et al*. The contribution of NOD2 gene mutations to the risk and site of disease in inflammatory bowel disease. *Gastroenterology.* 2002;**122**:867-74.

228. Wehkamp J, Harder J, Weichenthal M, Schwab M, Schaffeler E, Schlee M *et al*. NOD2 (CARD15) mutations in Crohn's disease are associated with diminished mucosal alpha-defensin expression. *Gut* 2004;**53**:1658-64.

229. Kobayashi KS, Chamaillard M, Ogura Y, Henegariu O, Inohara N, Nunez G *et al*. Nod2-Dependent Regulation of Innate and Adaptive Immunity in the Intestinal Tract. *Science* 2005;**307**:731-4.

230. Pierik M, De Hertogh G, Vermeire S, Van Assche G, Van Eyken P, Joossens S *et al*. Epithelioid granulomas, pattern recognition receptors, and phenotypes of Crohn's disease. *Gut* 2005;**54**:223-7.

231. Shaw SH, Hampe J, White R, Mathew CG, Curran ME, Schreiber S. Stratification by CARD15 variant genotype in a genome-wide search for inflammatory bowel disease susceptibility loci. *Human Genetics.* 2003;**113**:514-21.

232. Tanabe T, Chamaillard M, Ogura Y, Zhu L, Qiu S, Masumoto J *et al*. Regulatory regions and critical residues of NOD2 involved in muramyl dipeptide recognition. *EMBO J.* 2004;**23**:1587-97.

233. Sugimura K, Taylor KD, Lin YC, Hang T, Wang D, Tang YM *et al*. A novel NOD2/CARD15 haplotype conferring risk for Crohn disease in Ashkenazi Jews. *American Journal of Human Genetics* 2003;**72**:509-18.

234. Karban A,.Eliakim R. Failure to replicate the association of the novel NOD2/CARD15 haplotype (S268P-JW1) to Crohn's disease in the Jewish Israeli population. *Gastroenterology.* 2004;**126**:624-5.

235. Tukel T, Shalata A, Present D, Rachmilewitz D, Mayer L, Grant D *et al*. Crohn disease: frequency and nature of CARD15 mutations in Ashkenazi and Sephardi/Oriental Jewish families. *American Journal of Human Genetics.* 2004;**74**:623-36.

236. Kazeem GR,.Farrall M. Integrating case-control and TDT studies. *Annals of Human Genetics.* 2005;**69**:329-35.

237. Oostenbrug LE, Nolte IM, Oosterom E, van der SG, Te Meerman GJ, van Dullemen HM *et al*. CARD15 in inflammatory bowel disease and Crohn's disease phenotypes: An association study and pooled analysis. *Digestive & Liver Disease* 2006;**38**:834-45.

238. Medici V, Mascheretti S, Croucher PJ, Stoll M, Hampe J, Grebe J *et al*. Extreme heterogeneity in CARD15 and DLG5 Crohn disease-associated polymorphisms between German and Norwegian populations. *European Journal of Human Genetics* 2006;**14**:459-68.

239. King K, Sheikh MF, Cuthbert AP, Fisher SA, Onnie CM, Mirza MM *et al*. Mutation, selection, and evolution of the Crohn disease susceptibility gene CARD15. *Human Mutatation* 2006;**27**:44-54.

240. Carvalho RS, Abadom V, Dilworth HP, Thompson R, Oliva-Hemker M, Cuffari C. Indeterminate colitis: a significant subgroup of pediatric IBD. *Inflammatory Bowel Diseases* 2006;**12**:258-62.

241. Joossens S, Colombel JF, Landers C, Poulain D, Geboes K, Bossuyt X *et al*. Anti-outer membrane of porin C and anti-I2 antibodies in indeterminate colitis. *Gut* 2006;**55**:1667-9.

242. Joossens S, Reinisch W, Vermeire S, Sendid B, Poulain D, Peeters M *et al*. The value of serologic markers in indeterminate colitis: a prospective follow-up study. *Gastroenterology* 2002;**122**:1242-7.

243. Guindi M,.Riddell RH. Indeterminate colitis. *Journal of Clinical Pathology* 2004;**57**:1233-44.

244. Spray C, Debelle GD, Murphy MS. Current diagnosis, management and morbidity in paediatric inflammatory bowel disease. *Acta Paediatrica.* 2001;**90**:400-5.

245. Levine A, Shamir R, Wine E, Weiss B, Karban A, Shaoul RR *et al*. TNF Promoter Polymorphisms and Modulation of Growth Retardation and Disease Severity in Pediatric Crohn's Disease. *American Journal of Gastroenterology* 2005;**100**:1598-604.

246. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nature Genetics* 2001;**29**:229-32.

247. Giallourakis C, Stoll M, Miller K, Hampe J, Lander ES, Daly MJ *et al*. IBD5 is a general risk factor for inflammatory bowel disease: replication of association with Crohn disease and identification of a novel association with ulcerative colitis. *American Journal of Human Genetics.* 2003;**73**:205-11.

248. Mirza MM, Fisher SA, King K, Cuthbert AP, Hampe J, Sanderson J *et al*. Genetic evidence for interaction of the 5q31 cytokine locus and the CARD15 gene in Crohn disease. *American Journal of Human Genetics.* 2003;**72**:1018-22.

249. Armuzzi A, Ahmad T, Ling KL, de Silva A, Cullen S, van Heel D *et al*. Genotype-phenotype analysis of the Crohn's disease susceptibility haplotype on chromosome 5q31. *Gut.* 2003;**52**:1133-9.

250. Negoro K, McGovern DP, Kinouchi Y, Takahashi S, Lench NJ, Shimosegawa T *et al*. Analysis of the IBD5 locus and potential gene-gene interactions in Crohn's disease. *Gut* 2003;**52**:541-6.

251. McGovern DP, Van Heel DA, Negoro K, Ahmad T, Jewell DP. Further evidence of IBD5/CARD15 (NOD2) epistasis in the susceptibility to ulcerative colitis. *American Journal of Human Genetics.* 2003;**73**:1465-6.

252. Peltekova VD, Wintle RF, Rubin LA, Amos CI, Huang Q, Gu X *et al*. Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nature Genetics* 2004;**36**:471-5.

253. Noble CL, Nimmo ER, Drummond H, Ho GT, Tenesa A, Smith L *et al*. The Contribution of OCTN1/2 Variants Within the IBD5 Locus to Disease Susceptibility and Severity in Crohn's Disease. *Gastroenterology* 2005;**129**:1854-64.

254. Vermeire S, Pierik M, Hlavaty T, Claessens G, van Schuerbeeck N, Joossens S *et al*. Association of organic cation transporter risk haplotype with perianal penetrating Crohn's disease but not with susceptibility to IBD. *Gastroenterology* 2005;**129**:1845-53.

255. Russell RK, Drummond HE, Nimmo ER, Anderson N, Smith L, Wilson DC *et al*. Genotype-phenotype analysis in childhood-onset Crohn's disease: NOD2/CARD15 variants consistently predict phenotypic characteristics of severe disease. *Inflammatory Bowel Diseases* 2005;**11**:955-64.

256. Russell RK, Satsangi J, Wilson DC. More Pediatric NOD2/CARD Data. *American Journal of Gastroenterology.* 2004;**99**:1401.

257. Newby EA, Sawczenko A, Thomas AG, Wilson D. Interventions for growth failure in childhood Crohn's disease. *The Cochrane Database of Systematic Reviews* 2005;CD003873.

258. Markowitz J, Grancher K, Rosa J, Aiges H, Daum F. Growth failure in pediatric inflammatory bowel disease. *Journal of Pediatric Gastroenterology & Nutrition* 1993;**16**:373-80.

259. Motil KJ, Grand RJ, Davis-Kraft L, Ferlic LL, Smith EO. Growth failure in children with inflammatory bowel disease: a prospective study. *Gastroenterology* 1993;**105**:681-91.

260. Brescianini S, Trinh T, Stoll M, Schreiber S, Rioux JD, Daly MJ. IBD5 is associated with an extensive complicated Crohn's disease feature: implications from genotype-phenotype analysis. *Gut* 2007;**56**:149-50.

261. Newman B, Gu X, Wintle R, Cescon D, Yazdanpanah M, Liu X *et al*. A risk haplotype in the solute carrier family 22A4/22A5 gene cluster influences phenotypic expression of Crohn's disease. *Gastroenterology* 2005;**128**:260-9.

262. Torok HP, Glas J, Tonenchi L, Lohse P, Muller-Myhsok B, Limbersky O *et al*. Polymorphisms in the DLG5 and OCTN cation transporter genes in Crohn's disease. *Gut* 2005;**54**:1421-7.

263. Waller S, Tremelling M, Bredin F, Godfrey L, Howson J, Parkes M. Evidence for association of OCTN genes and IBD5 with ulcerative colitis. *Gut* 2006;**55**:809-14.

264. Fisher SA, Hampe J, Onnie CM, Daly MJ, Curley C, Purcell S *et al*. Direct or indirect association in a complex disease: the role of SLC22A4 and SLC22A5 functional variants in Crohn disease. *Human Mutation.* 2006;**27**:778-85.

265. Klein W, Tromm A, Griga T, Fricke H, Folwaczny C, Hocke M *et al*. Interleukin-4 and interleukin-4 receptor gene polymorphisms in inflammatory bowel diseases. *Genes & Immunity* 2001;**2**:287-9.

266. Aithal GP, Day CP, Leathart J, Daly AK, Hudson M. Association of single nucleotide polymorphisms in the interleukin-4 gene and interleukin-4 receptor gene with Crohn's disease in a British population. *Genes & Immunity* 2001;**2**:44-7.

267. Karttunnen R, Breese EJ, Walker-Smith JA, MacDonald TT. Decreased mucosal interleukin-4 (IL-4) production in gut inflammation. *Journal of Clinical Pathology* 1994;**47**:1015-8.

268. Kadivar K, Ruchelli ED, Markowitz JE, Defelice ML, Strogatz ML, Kanzaria MM *et al*. Intestinal interleukin-13 in pediatric inflammatory bowel disease patients. *Inflammatory Bowel Diseases* 2004;**10**:593-8.

269. Carvalho AT, Elia CC, de Souza HS, Elias PR, Pontes EL, Lukashok HP *et al*. Immunohistochemical study of intestinal eosinophils in inflammatory bowel disease. *Journal of Clinical Gastroenterology* 2003;**36**:120-5.

270. Noguchi M, Hiwatashi N, Liu ZX, Toyota T. Increased secretion of granulocyte-macrophage colony-stimulating factor in mucosal lesions of inflammatory bowel disease. *Digestion* 2001;**63**:32-6.

271. Babusukumar U, Wang T, McGuire E, Broeckel U, Kugathasan S. Contribution of OCTN variants within the IBD5 locus to pediatric onset Crohn's disease. *American Journal of Gastroenterology* 2006;**101**:1354-61.

272. Bene J, Magyari L, Talian G, Komlosi K, Gasztonyi B, Tari B *et al*. Prevalence of SLC22A4, SLC22A5 and CARD15 gene mutations in Hungarian pediatric patients with Crohn's disease. *World Journal of Gastroenterology*. 2006;**12**:5550-3.

273. The Collaborative Study on the Genetics of Asthma. A genome-wide search for asthma susceptibility loci in ethnically diverse populations. *Nature Genetics* 1997;**15**:389-92.

274. Xu J, Meyers DA, Ober C, Blumenthal MN, Mellen B, Barnes KC *et al*. Genomewide screen and identification of gene-gene interactions for asthma-susceptibility loci in three U.S. populations: collaborative study on the genetics of asthma. *American Journal of Human Genetics* 2001;**68**:1437-46.

275. Holloway JW, Lonjou C, Beghe B, Peng Q, Gaunt TR, Gomes I *et al*. Linkage analysis of the 5q31-33 candidate region for asthma in 240 UK families. *Genes & Immunity* 2001;**2**:20-4.

276. Hizawa N, Freidhoff LR, Ehrlich E, Chiu YF, Duffy DL, Schou C *et al*. Genetic influences of chromosomes 5q31-q33 and 11q13 on specific IgE responsiveness to common inhaled allergens among African American families. Collaborative Study on the Genetics of Asthma (CSGA). *Journal of Allergy and Clinical Immunology* 1998;**102**:449-53.

277. Hysi P, Kabesch M, Moffatt MF, Schedel M, Carr D, Zhang Y *et al*. NOD1 variation, immunoglobulin E and asthma. *Hum.Mol.Genet.* 2005;**14**:935-41.

278. McGovern DPB, Hysi P, Ahmad T, van Heel DA, Moffatt MF, Carey A *et al*. Association between a complex insertion/deletion polymorphism in NOD1(CARD4) and susceptibility to inflammatory bowel disease. *Hum.Mol.Genet.* 2005;**14**:1245-50.

279. Randolph AG, Lange C, Silverman EK, Lazarus R, Silverman ES, Raby B *et al*. The IL12B gene is associated with asthma. *American Journal of Human Genetics* 2004;**75**:709-15.

280. Howard TD, Whittaker PA, Zaiman AL, Koppelman GH, Xu J, Hanley MT *et al*. Identification and Association of Polymorphisms in the Interleukin-13 Gene with Asthma and Atopy in a Dutch Population. *Am.J.Respir.Cell Mol.Biol.* 2001;**25**:377-84.

281. Brant SR, Panhuysen CI, Nicolae D, Reddy DM, Bonen DK, Karaliukas R *et al*. MDR1 Ala893 polymorphism is associated with inflammatory bowel disease. *American Journal of Human Genetics* 2003;**73**:1282-92.

282. Carstairs V,.Morris R. Deprivation and Health in Scotland. *Health Bulletin (Edinburgh)* 1990;**48**:162-75.

283. Radford JR, Ballantyne HM, Nugent Z, Beighton D, Robertson M, Longbottom C *et al*. Caries-associated micro-organisms in infants from different socio-economic backgrounds in Scotland. *Journal of Dentistry* 2000;**28**:307-12.

284. Malcolm CA, MacKay WG, Shepherd A, Weaver LT. Helicobacter pylori in children is strongly associated with poverty. *Scottish Medical Journal* 2004;**49**:136-8.

285. Lander E,.Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* 1995;**11**:241-7.

286. Stoll M, Corneliussen B, Costello CM, Waetzig GH, Mellgard B, Koch WA *et al*. Genetic variation in DLG5 is associated with inflammatory bowel disease. *Nature Genetics* 2004;**36**:476-80.

287. Friedrichs F,.Stoll M. Role of discs large homolog 5. *World Journal of Gastroenterology* 2006;**12**:3651-6.

288. Nakamura H, Sudo T, Tsuiki H, Miyake H, Morisaki T, Sasaki J *et al*. Identification of a novel human homolog of the Drosophila dlg, P-dlg, specifically expressed in the gland tissues and interacting with p55. *FEBS Letters* 1998;**433**:63-7.

289. Noble CL, Nimmo ER, Drummond H, Smith L, Arnott ID, Satsangi J. DLG5 variants do not influence susceptibility to inflammatory bowel disease in the Scottish population. *Gut* 2005;**54**:1416-20.

290. Blanchard JF, Bernstein CN, Wajda A, Rawsthorne P. Small-area variations and sociodemographic correlates for the incidence of Crohn's disease and ulcerative colitis. *American Journal of Epidemiology* 2001;**154**:328-35.

291. Green C, Elliott L, Beaudoin C, Bernstein CN. A population-based ecologic study of inflammatory bowel disease: searching for etiologic clues. *American Journal of Epidemiology.* 2006;**164**:615-23.

292. Gavin S, Strange E Delaney K Houston J. Breastfeeding in lothian 2004. http: and www.nhslothian.scot.nhs.uk/publications/health_plan_2005/pdf/breastfeeding.pdf. 1-9. 2005.

293. Smith C, Burlison A. Closing the Gap: Tackling Inequalities in Health in Argyll & Clyde. www.achb.scot.nhs.uk/reports/closing%20the%20gap.pdf. 2001.

294. Barton JR,.Ferguson A. Failure to record variables of growth and development in children with inflammatory bowel disease. *BMJ.* 1989;**298**:865-6.

295. Kleinman RE, Baldassano RN, Caplan A, Griffiths AM, Heyman MB, Issenman RM *et al*. Nutrition support for pediatric patients with inflammatory bowel disease: a clinical report of the North American Society for Pediatric Gastroenterology, Hepatology And Nutrition. *Journal of Pediatric Gastroenterology & Nutrition.* 2004;**39**:15-27.

296. Daly MJ, Pearce AV, Farwell L, Fisher SA, Latiano A, Prescott NJ *et al*. Association of DLG5 R30Q variant with inflammatory bowel disease. *European Journal of Human Genetics* 2005;**13**:835-9.

297. Tenesa A, Noble CL, Satsangi J, Dunlop M. Association of DLG5 and inflammatory bowel disease across populations. *European Journal of Human Genetics* 2005;**13**:835-9.

298. Arnott ID, Ho GT, Nimmo ER, Satsangi J. Toll-like receptor 4 gene in IBD: further evidence for genetic heterogeneity in Europe. *Gut* 2005;**54**:308.

299. Broberger O,.Perlmann P. Demonstration of an epithelial antigen in colon by means of fluorescent antibodies from children with ulcerative colitis. *Journal of Experimental Medicine* 1962;**115**:13-26.

300. Perlmann P,.Broberger O. Auto-antibodies against antigen derived from colon in the microsomes of regional colonic lymph glands in human ulcerative colitis. *Nature* 1960;**188**:749-51.

301. Main J, McKenzie H, Yeaman GR, Kerr MA, Robson D, Pennington CR *et al*. Antibody to Saccharomyces cerevisiae (bakers' yeast) in Crohn's disease. *British Medical Journal* 1988;**297**:1105-6.

302. Ruemmele FM, Targan SR, Levy G, Dubinsky M, Braun J, Seidman EG. Diagnostic accuracy of serological assays in pediatric inflammatory bowel disease. *Gastroenterology* 1998;**115**:822-9.

303. Hoffenberg EJ, Fidanza S, Sauaia A. Serologic testing for inflammatory bowel disease. *Journal of Pediatrics* 1999;**134**:447-52.

304. Forcione DG, Rosen MJ, Kisiel JB, Sands BE. Anti-Saccharomyces cerevisiae antibody (ASCA) positivity is associated with increased risk for early surgery in Crohn's disease. *Gut* 2004;**53**:1117-22.

305. Walker LJ, Aldhous MC, Drummond HE, Smith BR, Nimmo ER, Arnott ID *et al*. Anti-Saccharomyces cerevisiae antibodies (ASCA) in Crohn's disease are associated with disease severity but not NOD2/CARD15 mutations. *Clinical & Experimental Immunology* 2004;**135**:490-6.

306. Zholudev A, Zurakowski D, Young W, Leichtner A, Bousvaros A. Serologic testing with ANCA, ASCA, and anti-OmpC in children and young adults with Crohn's disease and ulcerative colitis: diagnostic value and correlation with disease phenotype. *American Journal of Gastroenterology* 2004;**99**:2235-41.

307. Louis E, Michel V, Hugot JP, Reenaers C, Fontaine F, Delforge M *et al*. Early development of stricturing or penetrating pattern in Crohn's disease is influenced by disease location, number of flares, and smoking but not by NOD2/CARD15 genotype. *Gut* 2003;**52**:552-7.

308. Arnott ID, Landers CJ, Nimmo EJ, Drummond HE, Smith BK, Targan SR *et al*. Sero-reactivity to microbial components in Crohn's disease is associated with disease severity and progression, but not NOD2/CARD15 genotype. *American Journal of Gastroenterology* 2004;**99**:2376-84.

309. Israeli E, Grotto I, Gilburd B, Balicer RD, Goldin E, Wiik A *et al*. Anti-Saccharomyces cerevisiae and antineutrophil cytoplasmic antibodies as predictors of inflammatory bowel disease. *Gut* 2005;**54**:1232-6.

310. Seibold F. ASCA: genetic marker, predictor of disease, or marker of a response to an environmental antigen? *Gut* 2005;**54**:1212-3.

311. Cruyssen BV, Peeters H, Hoffman IE, Laukens D, Coucke P, Marichal D *et al*. CARD15 polymorphisms are associated with anti-Saccharomyces cerevisiae antibodies in Caucasian Crohn's disease patients. *Clinical and Experimental Immunology* 2005;**140**:354-9.

312. Esters N, Pierik M, van Steen K, Vermeire S, Claessens G, Joossens S *et al*. Transmission of CARD15 (NOD2) variants within families of patients with inflammatory bowel disease. *American Journal of Gastroenterology.* 2004;**99**:299-305.

313. Peeters M, Joossens S, Vermeire S, Vlietinck R, Bossuyt X, Rutgeerts P. Diagnostic value of anti-Saccharomyces cerevisiae and antineutrophil cytoplasmic autoantibodies in inflammatory bowel disease. *American Journal of Gastroenterology* 2001;**96**:730-4.

314. Harty S, Fleming P, Rowland M, Crushell E, McDermott M, Drumm B *et al*. A prospective study of the oral manifestations of Crohn's disease. *Clinical Gastroenterology and Hepatology.* 2005;**3**:886-91.

315. Savage NW, Barnard K, Shirlaw PJ, Rahman D, Mistry M, Escudier MP *et al*. Serum and salivary IgA antibody responses to Saccharomyces cerevisiae, Candida albicans and Streptococcus mutans in orofacial granulomatosis and Crohn's disease. *Clinical & Experimental Immunology.* 2004;**135**:483-9.

316. Quinton JF, Sendid B, Reumaux D, Duthilleul P, Cortot A, Grandbastien B *et al*. Anti-Saccharomyces cerevisiae mannan antibodies combined with antineutrophil cytoplasmic autoantibodies in inflammatory bowel disease: prevalence and diagnostic role. *Gut* 1998;**42**:788-91.

317. Klebl FH, Bataille F, Bertea CR, Herfarth H, Hofstadter F, Scholmerich J *et al*. Association of perinuclear antineutrophil cytoplasmic antibodies and anti-Saccharomyces cerevisiae antibodies with Vienna classification subtypes of Crohn's disease. *Inflammatory Bowel Diseases* 2003;**9**:302-7.

318. Canani RB, Romano MT, Greco L, Terrin G, Sferlazzas C, Barabino A *et al*. Effects of disease activity on anti-Saccharomyces cerevisiae antibodies: implications for diagnosis and follow-up of children with Crohn's disease. *Inflammatory Bowel Diseases* 2004;**10**:234-9.

319. Gupta N, Cohen SA, Bostrom AG, Kirschner BS, Baldassano RN, Winter HS *et al*. Risk factors for initial surgery in pediatric patients with Crohn's disease. *Gastroenterology.* 2006;**130**:1069-77.

320. Gupta SK, Fitzgerald JF, Croffie JM, Pfefferkorn MD, Molleston JP, Corkins MR. Comparison of serological markers of inflammatory bowel disease with clinical diagnosis in children. *Inflammatory Bowel Diseases* 2004;**10**:240-4.

321. Alvarez-Lobos M, Arostegui JI, Sans M, Tassies D, Plaza S, Delgado S *et al*. Crohn's Disease Patients Carrying Nod2/CARD15 Gene Variants Have an Increased and Early Need for First Surgery due to Stricturing Disease and Higher Rate of Surgical Recurrence. *Annals of Surgery* 2005;**242**:693-700.

322. Amre DK, Lu SE, Costea F, Seidman EG. Utility of serological markers in predicting the early occurrence of complications and surgery in pediatric Crohn's disease patients. *American Journal of Gastroenterology.* 2006;**101**:645-52.

323. Dassopoulos T, Frangakis C, Cruz-Correa M, Talor MV, Burek CL, Datta L *et al*. Antibodies to saccharomyces cerevisiae in Crohn's disease: higher titers are associated with a greater frequency of mutant NOD2/CARD15 alleles and with a higher probability of complicated disease. *Inflammatory Bowel Diseases.* 2007;**13**:143-51.

324. Dendrinos KG, Becker JM, Stucchi AF, Saubermann LJ, LaMorte W, Farraye FA. Anti-Saccharomyces cerevisiae antibodies are associated with the development of postoperative fistulas following ileal pouch-anal anastomosis. *Journal of Gastrointestinal Surgery.* 2006;**10**:1060-4.

325. Targan SR, Landers CJ, Yang H, Lodes MJ, Cong Y, Papadakis KA *et al*. Antibodies to CBir1 Flagellin Define a Unique Response That Is Associated Independently With Complicated Crohn's Disease. *Gastroenterology* 2005;**128**:2020-8.

326. Mow WS, Vasiliauskas EA, Lin YC, Fleshner PR, Papadakis KA, Taylor KD *et al*. Association of antibody responses to microbial antigens and complications of small bowel Crohn's disease. *Gastroenterology* 2004;**126**:414-24.

327. Dotan I, Fishman S, Dgani Y, Schwartz M, Karban A, Lerner A *et al*. Antibodies against laminaribioside and chitobioside are novel serologic markers in Crohn's disease. *Gastroenterology* 2006;**131**:366-78.

328. Dubinsky MC, Lin YC, Dutridge D, Picornell Y, Landers CJ, Farrior S *et al*. Serum immune responses predict rapid disease progression among children

with Crohn's disease: immune responses predict disease progression. *American Journal of Gastroenterology* 2006;**101**:360-7.

329. Khan K, Schwarzenberg SJ, Sharp H, Greenwood D, Weisdorf-Schindele S. Role of serology and routine laboratory tests in childhood inflammatory bowel disease. *Inflammatory Bowel Diseases* 2002;**8**:325-9.

330. Damoiseaux JG, Bouten B, Linders AM, Austen J, Roozendaal C, Russel MG *et al*. Diagnostic value of anti-Saccharomyces cerevisiae and antineutrophil cytoplasmic antibodies for inflammatory bowel disease: high prevalence in patients with celiac disease. *Journal of Clinical Immunology* 2002;**22**:281-8.

331. Standaert-Vitse A, Jouault T, Vandewalle P, Mille C, Seddik M, Sendid B *et al*. Candida albicans is an immunogen for anti-Saccharomyces cerevisiae antibody markers of Crohn's disease. *Gastroenterology* 2006;**130**:1764-75.

332. Seibold F, Konrad A, Flogerzi B, Seibold-Schmid B, Arni S, Juliger S *et al*. Genetic variants of the mannan-binding lectin are associated with immune reactivity to mannans in Crohn's disease. *Gastroenterology* 2004;**127**:1076-84.

333. Devlin SM, Yang H, Ippoliti A, Taylor KD, Landers CJ, Su X *et al*. NOD2 Variants and Antibody Response to Microbial Antigens in Crohn's Disease Patients and their Unaffected Relatives. *Gastroenterology* 2006;**132**:576-86.

334. Vermeire S, Joossens S, Peeters M, Monsuur F, Marien G, Bossuyt X *et al*. Comparative study of ASCA (Anti-Saccharomyces cerevisiae antibody) assays in inflammatory bowel disease. *Gastroenterology.* 2001;**120**:827-33.

335. Quinton JF, Sendid B, Reumaux D, Duthilleul P, Cortot A, Grandbastien B *et al*. Anti-Saccharomyces cerevisiae mannan antibodies combined with antineutrophil cytoplasmic autoantibodies in inflammatory bowel disease: prevalence and diagnostic role.[see comment]. *Gut.* 1998;**42**:788-91.

336. Koutroubakis IE, Petinaki E, Mouzas IA, Vlachonikolis IG, Anagnostopoulou E, Castanas E *et al*. Anti-Saccharomyces cerevisiae mannan antibodies and antineutrophil cytoplasmic autoantibodies in Greek patients with inflammatory bowel disease. *American Journal of Gastroenterology* 2001;**96**:449-54.

337. Van Limbergen J, Russell RK, Nimmo ER, Torkvist L, Lees CW, Drummond HE *et al*. Contribution of the NOD1/CARD4 insertion/deletion polymorphism +32656 to inflammatory bowel disease in Northern Europe. *Inflammatory Bowel Diseases* 2007;**13**:882-9.

338. Van Limbergen J, Russell RK, Nimmo ER, Drummond HE, Anderson N, Wilson DC *et al*. Tumour necrosis factor alpha promoter polymorphisms influence disease phenotype and severity in childhood inflammatory bowel disease. *Gastroenterology* 2006;**130**:A-3(19).

339. Palmieri O, Latiano A, VALVANO R, D'Inca R, Vecchi M, STURNIOLO GC *et al*. Variants of OCTN1-2 cation transporter genes are associated with both Crohn's disease and ulcerative colitis. *Alimentary Pharmacology & Therapeutics* 2006;**23**:497-506.

340. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genetics.* 2003;**33**:177-82.

341. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *Journal of the National Cancer Institute.* 2004;**96**:434-42.

342. Van Limbergen J, Russell RK, Nimmo ER, Satsangi J. The Genetics of Inflammatory Bowel Disease. *The American Journal of Gastroenterology* 2007;**102**:2820-31.

343. The International HapMap Consortium. The International HapMap Project. *Nature* 2003;**426**:789-96.

344. Hirschhorn JN,.Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* 2005;**6**:95-108.

345. Yamazaki K, McGovern D, Ragoussis J, Paolucci M, Butler H, Jewell D *et al*. Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. *Hum.Mol.Genet.* 2005;**14**:3499-506.

346. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ *et al*. A Genome-Wide Association Study Identifies IL23R as an Inflammatory Bowel Disease Gene. *Science* 2006;**314**:1461-3.

347. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A *et al*. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nature Genetics* 2007;**30**:596-604.

348. Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K *et al*. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nature Genetics* 2007;**39**:207-11.

349. Libioulle C, Louis E, Hansoul S, Sandor C, Farnir F, Franchimont D *et al*. Novel Crohn Disease Locus Identified by Genome-Wide Association Maps to a Gene Desert on 5p13.1 and Modulates Expression of PTGER4. *PLoS.Genetics.* 2007;**3**:e58.

350. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;**447**:661-78.

351. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ *et al*. A Genome-Wide Association Study Identifies IL23R as an Inflammatory Bowel Disease Gene. *Science* 2006;**314**:1461-3.

352. Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA *et al*. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nature Genetics.* 2007;**39**:830-2.

353. Palmer LJ,.Cardon LR. Shaking the tree: mapping complex disease genes with linkage disequilibrium. *Lancet.* 2005;**366**:1223-34.

354. Ho GT, Soranzo N, Nimmo ER, Tenesa A, Goldstein DB, Satsangi J. ABCB1/MDR1 gene determines susceptibility and phenotype in ulcerative colitis: discrimination of critical variants using a gene-wide haplotype tagging approach. *Hum.Mol.Genet.* 2006;**15**:797-805.

355. Annese V, Valvano MR, Palmieri O, Latiano A, Bossa F, Andriulli A. Multidrug resistance 1 gene in inflammatory bowel disease: a meta-analysis. *World Journal of Gastroenterology* 2006;**12**:3636-44.

356. Ho GT, Soranzo N, Tate SK, Drummond H, Nimmo ER, Tenesa A *et al*. Lack of association of the pregnane X receptor (PXR/NR1I2) gene with inflammatory bowel disease: parallel allelic association study and gene wide haplotype analysis. *Gut* 2006;**55**:1676-7.

357. Van Limbergen J, Nimmo ER, Russell RK, Drummond HE, Smith L, Anderson NH *et al*. Investigation of NOD1/CARD4 variation in Inflammatory Bowel Disease using a haplotype-tagging strategy. *Human Molecular Genetics* 2007;**16**:2175-86.

358. McGovern DPB, Hysi P, Ahmad T, van Heel DA, Moffatt MF, Carey A *et al*. Association between a complex insertion/deletion polymorphism in NOD1 (CARD4) and susceptibility to inflammatory bowel disease. *Hum.Mol.Genet.* 2005;**14**:1245-50.

359. Tremelling M, Hancock L, Bredin F, Sharpstone D, Bingham SA, Parkes M. Complex insertion/deletion polymorphism in NOD1 (CARD4) is not associated with inflammatory bowel disease susceptibility in East Anglia panel. *Inflammatory Bowel Diseases.* 2006;**12**:967-71.

360. Franke A, Ruether A, Wedemeyer N, Karlsen TH, Nebel A, Schreiber S. No association between the functional CARD4 insertion/deletion polymorphism and inflammatory bowel diseases in the German population. *Gut* 2006;**55**:1679-80.

361. Fields S,.Song O. A novel genetic system to detect protein-protein interactions. *Nature* 1989;**340**:245-6.

362. Barnich N, Hisamatsu T, Aguirre JE, Xavier R, Reinecker HC, Podolsky DK. GRIM-19 interacts with nucleotide oligomerization domain 2 and serves as downstream effector of anti-bacterial function in intestinal epithelial cells. *Journal of Biological Chemistry* 2005;**280**:19021-6.

363. Yamamoto-Furusho JK, Barnich N, Xavier R, Hisamatsu T, Podolsky DK. Centaurin beta1 Down-regulates Nucleotide-binding Oligomerization Domains 1- and 2-dependent NF-{kappa}B Activation. *J.Biol.Chem.* 2006;**281**:36060-70.

364. Stoughton RB. Applications of DNA microarrays in biology. *Annual Review of Biochemistry.* 2005;**74**:53-82.

365. Lawrance IC, Fiocchi C, Chakravarti S. Ulcerative colitis and Crohn's disease: distinctive gene expression profiles and novel susceptibility candidate genes. *Human Molecular Genetics.* 2001;**10**:445-56.

366. Wu F, Dassopoulos T, Cope L, Maitra A, Brant SR, Harris ML *et al*. Genome-wide gene expression differences in Crohn's disease and ulcerative colitis from endoscopic pinch biopsies: Insights into distinctive pathogenesis. *Inflammatory Bowel Diseases* 2007;**13**:807-21.

367. Turunen P, Kolho KL, Auvinen A, Iltanen S, Huhtala H, Ashorn M. Incidence of Inflammatory Bowel Disease in Finnish Children, 1987-2003. *Inflammatory Bowel Diseases* 2006;**12**:677-83.

368. Torkvist L, Noble CL, Lordal M, Sjoqvist U, Lindforss U, Nimmo ER *et al*. Contribution of the IBD5 Locus to Crohn's Disease in the Swedish Population. *Scandinavian Journal of Gastroenterology.* 2007;**42**:200-6.

369. Paavola-Sakki P, Ollikainen V, Helio T, Halme L, Turunen U, Lahermo P *et al*. Genome-wide search in Finnish families with inflammatory bowel disease provides evidence for novel susceptibility loci. *European Journal of Human Genetics.* 2003;**11**:112-20.

370. Russell RK, Drummond HE, Nimmo ER, Anderson NH, Noble CL, Wilson DC *et al*. Analysis of the influence of OCTN1/2 variants within the IBD5 locus on disease susceptibility and growth indices in early onset inflammatory bowel disease. *Gut* 2006;**55**:1114-23.

371. Russell RK, Drummond HE, Nimmo ER, Anderson N, Wilson DC, Gillett PM *et al*. The contribution of the DLG5 113A variant in early-onset inflammatory bowel disease. *Journal of Pediatrics* 2007;**150**:268-73.

372. Russell RK, Ip B, Aldhous MC, Drummond HE, Arnott IDR, McDougall M *et al*. ASCA positivity independently predicts oral Crohn's disease in a large early onset inflammatory bowel disease population. *Gut* 2006;**55**:A10.

373. Triester SL, Leighton JA, Leontiadis GI, Gurudu SR, Fleischer DE, Hara AK *et al*. A meta-analysis of the yield of capsule endoscopy compared to other diagnostic modalities in patients with non-stricturing small bowel Crohn's disease. *American Journal of Gastroenterology* 2006;**101**:954-64.

374. Savage NW, Barnard K, Shirlaw PJ, Rahman D, Mistry M, Escudier MP *et al*. Serum and salivary IgA antibody responses to Saccharomyces cerevisiae, Candida albicans and Streptococcus mutans in orofacial granulomatosis and Crohn's disease. *Clinical & Experimental Immunology* 2004;**135**:483-9.

375. Lal S, Appelton J, Mascarenhas J, Stempak JM, Esplen MJ, Silverberg MS. Attitudes toward genetic testing in patients with inflammatory bowel disease. *European Journal of Gastroenterology & Hepatology* 2007;**19**:321-7.

376. McGovern DP, Butler H, Ahmad T, Paolucci M, Van Heel DA, Negoro K *et al*. TUCAN (CARD8) genetic variants and inflammatory bowel disease. *Gastroenterology* 2006;**131**:1190-6.

377. Seiderer J, Schnitzler F, Brand S, Staudinger T, Pfennig S, Herrmann K *et al*. Homozygosity for the CARD15 frameshift mutation 1007fs is predictive of early onset of Crohn's disease with ileal stenosis, entero-enteral fistulas, and frequent need for surgical intervention with high risk of re-stenosis. *Scandinavian Journal of Gastroenterology* 2006;**41**:1421-32.